

University of Nottingham
UK | CHINA | MALAYSIA

Exploiting trypanosomes as models for pathogenicity and core biology research

Thesis submitted to the University of Nottingham towards the degree of Doctor of Philosophy

July 2020

Georgina Awuah-Mensah

Mphil, BSc (Hons)

Contents

Abstract **8**

Declaration **10**

Acknowledgement **11**

Abbreviations **12**

1 BACKGROUND **15**

1.1 Model systems for studying eukaryotic biology 15

1.1.1 Unicellular models for common biology 15

1.1.2 Multicellular models 18

1.1.3 Eukaryotic diversity and model systems 19

1.1.4 Biology from non-traditional model systems 21

1.1.5 Emerging models 22

1.2 African Trypanosomes 24

1.2.1 Importance 24

1.2.2 African trypanosome biology 25

1.3 Trypanosomes as a model 30

1.3.1 Axenic culture 31

1.3.2 Genetic modification in Trypanosomes 32

1.3.3 Current limitations 38

1.4 Aims 39

2 INTRODUCTION I **40**

2.1 Kinesin motors 40

2.1.1 Structure of kinesins 40

2.1.2 Processivity along microtubules 42

2.1.3 Architecture of kinesin motors 44

2.2 Kinesin families across eukaryotes 45

2.2.1 Nomenclature 45

2.2.2 Distribution of kinesin families 46

2.3 Functional differentiation of kinesins 48
2.3.1 Function of the various families 48
2.3.2 Differing functions within families 49
2.4 *T. brucei* as a choice model organism for studying kinesins 51
2.4.1 Kinesin-1 51
2.4.2 Kinesin-3 52
2.4.3 Kinesin-2 52
2.4.4 Kinesin-13 53
2.4.5 Kinesin-14 53
2.5 Aim I 53
3 MATERIALS AND METHODS 55
3.1 Cell lines and culture 55
3.2 Genome-wide RNAi targeting fragments 55
3.3 Plasmids for transgenic modification 56
3.3.1 *T. brucei* endogenous locus tagging and RNAi plasmid construction 56
3.4 Genotypic analysis of genetically-modified strains 59
3.5 Preparation of *T. congolense* DRiF-Seq samples 60
3.6 Sequencing of DRiF-Seq samples 60
3.7 Processing of DRiF-Seq data 60
3.8 Quantitative enrichment analysis 63
3.9 Genetic modification of trypanosomes 63
3.9.1 *T. brucei* 63
3.9.2 *T. congolense* 64
3.10 Subcellular localisation of proteins 65
3.11 Measurement of bioluminescence 66
3.11.1 *In vitro* measurement 66

3.11.2 *In vivo* assessment 66
3.12 Flow cytometry 67
3.13 Double-strand break assay 68
3.14 Estimating level of protein in whole cell lysates 68
3.15 RNAi induction assay and growth analysis 69
3.16 Statistical analysis 70
4 RESULTS 1 71
4.1 Characterisation of *T. brucei* kinesins 71
4.1.1 *T. brucei* kinesins localise to different subcellular locations associated with their possible function 72
4.2 Effect of depletion of *T. brucei* kinesins 76
4.2.1 Depletion of TbKIN1, 2A and TbKIFC1 78
4.2.2 Depletion of TbKIN3 produces growth and phenotypic defect 80
4.2.3 TbKIN13-1 knockdown produced a defect in mitosis 81
4.3 Kinesin-1 is not essential to bloodstream-form *T. brucei* in culture 83
5 DISCUSSION I 85
5.1 Can *T. brucei* be used as a model to study kinesins? 85
5.1.1 Comparison of *T. brucei* kinesins to common model organisms 85
5.1.2 *T. brucei* kinesin candidates for investigating eukaryotic kinesins 86
5.2 Proposed future studies 87
5.2.1 Contribution of the motor domain to kinesin function 87
5.2.2 A new approach to study functional differentiation of kinesins in *T. brucei* 88
6 INTRODUCTION II 89
6.1 Drug resistance in *T. congolense* 90
6.2 *T. congolense* as a new model for studying animal African trypanosomiasis 91
6.2.1 Genome 91
6.2.2 Axenic culture 91

6.2.3 Genetic modification in *T. congolense* 92

6.3 Aim II 95

7 RESULTS 2 97

7.1 Genetic modification of *T. congolense* 97

7.1.1 *T. congolense* single marker cell line 97

7.1.2 Transcription of transgenes from rDNA promoters at endogenous loci 99

7.2 Genetically-modified *T. congolense* is infective to mice 101

7.3 Creating tools for inducible expression 102

7.3.1 Selection of potentially silent loci for regulated transgenesis 102

7.3.2 All minichromosomal loci support transgenesis 104

7.3.3 Minichromosomal loci support regulatable expression 106

7.4 Transgenes at the minichromosomal loci are stable with respect to mitotic growth 108

8 RESULTS 3 110

8.1 Design and use of *T. congolense* RNAi constructs 110

8.1.1 Design of constructs for inducible RNA interference in *T. congolense* 110

8.1.2 RNAi depletes exogenous GFP in *T. congolense* 113

8.2 Relative level of RNAi-targeting appears independent of expression level 114

8.3 RNAi reduces both protein and mRNA levels of non-essential genes in *T. congolense* 117

8.3.1 RNAi-induced depletion reduces level of protein 117

8.3.2 RNAi-induction reduces mRNA levels of endogenous gene in *T. congolense* . 118

8.4 RNAi directed against transcripts of clathrin heavy chain and tubulin phenocopies *T. brucei* 118

8.5 RNAi targeted to KKIP1 and putative expressed VSG in *T. congolense* does not phenocopy *T. brucei* 121

8.6 The proportion of cells that produce morphological defects decrease when RNAi lines are kept in continuous culture 124

9 RESULTS 4 126

9.1 Ways to increase number of transformants in trypanosomes 126

9.2 Design of construct for inducible homing endonuclease cleavage 128

9.3 Inducing endonuclease transcription results in loss of resistant marker 129

9.4 Induction of DSB significantly increases number of transformants in *T. congolense* 133

9.5 Transfection efficiency of TcoTTS cell line is stable to freeze-thaw 135

9.6 Optimisation of number of transformants per transfection 136

10 RESULT 5 138

10.1 A quantitative genome-wide RNAi screen in *T. congolense* 138

10.1.1 Plasmid library 139

10.1.2 Bacterial library 140

10.1.3 *T. congolense* genome-wide RNAi library 143

10.2 *T. congolense* DRiF-Seq 144

10.2.1 *T. congolense* DRiF-Seq can differentiate between genes that produce an effect on knockdown by RNAi and those that do not 145

10.2.2 Independent *T. congolense* DRiF-Seq libraries demonstrate good agreement in gene fitness profiles 151

10.2.3 There is a strong positive correlation between effect sizes of genes in both libraries 153

10.2.4 A set of genes in the libraries is associated with an apparent gain of fitness 155

10.2.5 Quantitative enrichment analysis 158

10.2.6 Differential outcome between individual gene knockdown and outcome in DRiF-Seq library 166

10.2.7 Investigating differential biology: *T. congolense* x *T. brucei* 170

10.3 Isometamidium resistance and mode of action studies 177

10.3.1 Screening the *T. congolense* genome for isometamidium sensitivity/resistance 180

11 DISCUSSION II 182

11.1 Transgenesis in bloodstream-form *T. congolense* 182

11.2 RNAi penetrance in *T. congolense* 183

11.2.1 Refractory cells 183

11.2.2 Threshold effect on the cell population 184

11.3 Optimal conditions to increase number of transformants 185

11.4 Genome-wide *T. congolense* DRiF-Seq 187

11.4.1 Low RNAi penetrance in *T. congolense* 187

11.4.2 Differential biology between *T. congolense* and *T. brucei* 188

11.4.3 Quorum sensing 189

11.4.4 Application of *T. congolense* genome-wide technology 190

12 REFERENCE 193

Abstract

Model systems are commonly used to investigate biological processes and understand molecular mechanisms. The choice of a model cell or organism is informed by the scope of the question being asked. Decades of research using predominantly animal and fungal models form the bulk of our knowledge of eukaryotic processes. The diversity of eukaryotic organisms, however, indicates that this may not be representative of what exists in more diverged groups, and key processes like cell division vary across the larger group. In the past, non-traditional models including *Tetrahymena* have been applied to answer biological questions because of peculiar traits that make them suitable for investigation. African trypanosomes offer an excellent opportunity as model organisms because of their divergence from common models and importance to human health and livelihood. Particularly *Trypanosoma brucei*, with several genetic tools available to study them, has become an interesting model to study processes like cell division, cytoskeleton and antigenic variation. Here, I present work exploring the application of two species of African trypanosomes as models to investigate common biology of eukaryotic kinesins and specific biology of animal African trypanosomiasis.

The genetic tractability of *T. brucei* and its cytoskeletal structure, which is predominantly made up of tubulin, makes it a good model to study functional differentiation between kinesin families. I show cellular localisation and characterise the effect of RNAi depletion of representative kinesin families. The data suggest that Kinesin-2, -3 and 13-1 perform similar roles in *T. brucei* as seen in other systems, and can therefore be candidates for investigation. On the contrary, loss of both alleles of *T. brucei* Kinesin-1 does not affect *in vitro* growth, suggesting possible functional redundancy in trypanosomes.

Trypanosoma congolense is the most significant cause of animal African trypanosomiasis, a wasting disease of livestock that results in huge economic losses to agriculture particularly in Africa. Disease presentation differs from *T. brucei,* and the lack of genetic tools makes it difficult to study the biology of *T. congolense*, specifically its pathogenesis in animal trypanosomiasis. I show the progressive

development and application of genetic tools for endogenous locus tagging, RNA interference and high-complexity mutant library production. I demonstrate that loci on the minichromosome can be targeted for transgene integration, give good expression and are regulatable. Genes targeted by RNAi could only be depleted to about 50% of original levels, with similar proportion of cells expressing mutant phenotype, suggesting that *T. congolense* has lower RNAi penetrance than *T. brucei.* The development and optimization of a system that uses a homing endonuclease to increase DNA breaks at integration sites consistently increased the number of positive transformants to ~ 50,000 per transfection. This *T. congolense* line maintained stable transfection efficiency even after cycles of freeze-thawing.

Finally, I demonstrate for the first time, a genome-wide technology in *T. congolense.* Using all the tools developed, I create two RNAi libraries containing $> 5x10^5$ independent transformants. The method of library generation by direct RNAi fragment sequencing (DRiF-Seq) allowed detection of very small changes in gene fitness cost due to RNAi. This enabled fitness screen of > 9000 genes in *T. congolense* even at a low RNAi penetrance. Analysis of differential biology between *T. congolense* and *T. brucei* gives some glimpse into similarities and differences in processes like endocytosis and quorum sensing. I also demonstrate application of the *T. congolense* RNAi libraries in a drug screen to investigate isometamidium mode of action and resistance.

Declaration

I declare that this thesis is the result of my own work, except where included data is explicitly mentioned, which has been undertaken during my period of registration for this degree at The University of Nottingham.

Acknowledgement

First and foremost, I would like to thank God for the opportunity to meet and work with all the people who supported me from start to finish. My sincere appreciation goes to Dr Catarina Gadelha who secured funding to support my PhD and also for her guidance, encouragement and input in the project.

A special thanks to my supervisor, Dr Bill Wickstead for the invaluable training and support these past years. I lost count of the many times I barged into your office, lab book in hand, seeking your opinion on one thing or the other. Thank you for your patience and kindness. I also appreciate the opportunity I had to present my work in the US.

Thank you to the members of Wickstead, Gadelha, Huett, Friel and Allers labs for your aid every now and then and for sharing supplies and equipment. Simon, Jen and Lorenzo, thank you for all the times you helped me out in the lab and Tom for the long lunch chats. Thank you!

This work was funded by the University of Nottingham and Sir Halley Stewart Trust.

Abbreviations

AAT	animal African trypanosomiasis
BARP	brucei alanine rich protein
BSF	bloodstream form
Cas	CRISPR-associated protein
CDK	cyclin-dependent kinase
CDS	coding DNA sequence
CRISPR	clustered regularly interspaced short palindromic repeats
DPS	downstream processing unit
DRiF-Seq	direct RNAi fragment sequencing
DSB	double strand break
dsRNA	double-stranded RNA
ESP	enriched-in-surfeome protein
FDR	false discovery rate
GARP	glutamate and alanine rich protein
gDNA	genomic DNA
GFP	green fluorescent protein
GoF	gain of fitness
gRNA	guide RNA
HAT	human African trypanosomiasis
HEK	human embryonic kidney
HMI	Hirumi's Modified Iscove's medium
HYG	Hygromycin resistant marker

IC	intermediate chromosome
IFT	intraflagellar transport
IMDM	Iscove's Modified Dulbecco's Medium
ISM	isometamidium
ITR	intergenic region
kDa	kilodalton
KHC	kinesin heavy chain
KIF	kinesin family
KLC	kinesin light chain
KKIP	KKT-interacting protein
LoF	loss of fitness
Mb	megabase
MBC	megabase-sized chromosome
MC	minichromosome
mVSG	metacyclic variant surface glycoprotein
NEO	Neomycin selection marker
NGS	next generation sequencing
PARP	procyclic acidic repetitive protein
PCF	procyclic form
PPDK	pyruvate phosphate dikinase
RBP	RNA-binding proteins
rDNA	ribosomal DNA
RIT-seq	RNAi target sequencing

RNAi	RNA interference
T7RNAP	T7 RNA polymerase
TcBSF	*Trypanosoma congolense* bloodstream form medium
TetO	tetracycline operator
TetR	tetracycline repressor protein
UPS	upstream processing unit
UTR	untranslated region
VSG	variant surface glycoprotein

1 BACKGROUND

1.1 Model systems for studying eukaryotic biology

The study of model systems has a hugely important place in biology. The concept of using such models is based primarily on the evolutionary principle that all organisms share a common ancestry. As such, we can infer understanding of cellular processes in other organisms from our understanding of similar processes studied in a model organism. While this is still a fundamental principle, biologists are also faced with the dilemma of a very large diversity in organisms. For example, within just the eukaryotic domain, heterotrophs are classified alongside autotrophs and mixotrophs that switch between different energy sources; highly varied genome sizes with some single-cell protists having 200 times the size of the human genome; and free-living organisms classified with parasitic and symbiotic organisms. At the molecular level, research continues to reveal diversity in even common biological processes that had been previously accepted as canonical. The very nature of this diversity means that the choice of a research model requires thought of what the question is we are asking. Conceptually, biological models can be grouped in two main clusters based on whether they are used to investigate “common” or “specific” biology. Models of common biology are primarily used to understand molecular mechanisms for functions that are generally shared across eukaryotes; whereas specific models are used to investigate aspects of biology which differ, for example in pathogenicity and disease.

1.1.1 Unicellular models for common biology

Unicellular models have historically supported our understanding of common eukaryotic biology. All eukaryotes share the general arrangement and functional compartmentalization of a cell, including membrane-bound organelles such as the nucleus, Golgi apparatus, endoplasmic reticulum, mitochondria and vesicles. There is a striking overlap in cellular mechanisms and processes essential for life with shared similarities between the control of cell division, cellular transport and motility.

Metabolic pathways, DNA replication, transcription, RNA processing and translation are also generally shared.

Yeasts

There are several reasons why yeasts, particularly *Saccharomyces cerevisiae* (budding yeast) and *Schizosaccharomyces pombe* (fission yeast), have been preferred model organisms for many investigations. *S. cerevisiae* is historically one of the most-studied unicellular models; see (Botstein and Fink, 1988, 2011; Rosas-Murrieta et al., 2015). It is fast growing, dividing every 90 – 140 minutes, which means investigations can traverse several generations within a few days. It is cheap and easy to culture, surviving over wide temperature and pH conditions. A most desirable trait of yeasts as experimental organisms is the ease of generating spontaneous mutations. Another attribute of yeasts enabling them to become a forerunner as choice model organism is that, despite belonging to a different clade of the eukaryotic group, it is "surprisingly" closely related to animals. Although yeasts and humans are separated by around 1 billion years of evolution, evidence indicates a good conservation of much gene function (Kachroo et al., 2015). Both *S. cerevisiae* and *S. pombe* are normally haploid and can be grown exclusively by mitotic division or stimulated to undergo meiosis, making studies of meiosis and segregation process possible.

Yeasts have therefore served as excellent models for systems such as the eukaryotic cell cycle and the signal recognition particle (SRP) pathway. Cyclin-dependent kinases (CDKs) were first discovered in fission yeast as essential for cell cycle progression (Nurse and Thuriaux, 1980). Homologues of the pioneer CDK, *cdc2*, were identified in human cells and shown to have the ability to complement yeast mutants (Lee and Nurse, 1987), supporting the functional conservation between them.

Similarly, the SRP pathway appears universally conserved in eukaryotes and is the main route for targeting nascent membrane and secretory proteins to the ER in eukaryotes and ensuring their correct localisation. Much of our understanding of the *in vivo* mechanism by which SRP specifically and efficiently targets proteins for translocation is derived from studies in yeast [reviewed in (Aviram and Schuldiner, 2017)].

Another complex cellular machinery extensively studied in yeast is the kinetochore. The kinetochore is a complex of ~30 proteins assembled at chromosome centromeres, and acts as an attachment point for the mitotic spindle during chromosome segregation in cell division. Given the importance of chromosome segregation to genome stability, elucidating kinetochore structure is of key importance. Human kinetochores are large macromolecular machines difficult to isolate and purify, and some of the first sub-structural studies came from the far simpler yeast kinetochore (Gonen et al., 2012). Many of the answers to both structure and function of kinetochore components were derived from work in yeast [reviewed in (Kern and Cheeseman, 2012)]. However, the kinetochore is also an excellent example of where comparisons between model systems can be misleading with respect to the apparent universality of eukaryotic biology. Although yeast and human kinetochores share much of their organisation, studies in other eukaryotes show that there is much more diversity in the way to build a kinetochore (Akiyoshi and Gull, 2014; D'Archivio and Wickstead, 2017; Hooff et al., 2017).

While yeasts continue to have a premium place in research, biological differences in the eukaryotic tree of life limit their universal use. For instance, yeasts would be an unsuitable model to study cell motility, for the obvious reason of lacking a cilium or flagellum. Neither would they be suitable models for the study of the canonical process of phagocytosis, because yeasts are osmotroph organisms. Lastly, in the area of molecular technology, the lack of a complete RNA interference (RNAi) machinery does not allow us to exploit the machinery for gene functional studies.

Mammalian cells

Another example of an important model are the many different culturable mammalian cell lines. The importance of mammalian cells in our understanding of infection and disease cannot be underestimated. HeLa, Chinese hamster ovary (CHO), baby hamster kidney (BHK) and Madin-Darby Canine Kidney (MDCK) cells are just a few of those *in vitro* systems currently available. Mammalian (non-human) lines are often chosen as models to investigate human physiology and disease, due to similar/shared biology. For example, CHO cells are important in investigating diseases

such as visceral leishmaniasis, and also for large-scale production of therapeutic recombinant proteins. The advent of the CRISPR/Cas and siRNA technologies have increased our toolkit to genetically manipulate these lines and their applications. Many mammalian cells also have the ability to differentiate *in vitro*, and this has fueled the recently growing field of organoids for use in research (Rossi et al., 2018). On the contrary, the most amenable mammalian cell lines are often cancerous or transformed lines and mutations gathered by these cells can alter normal cellular biology.

1.1.2 Multicellular models

Organoids are great tools to capture some of the key features of a real organ, but are still artificial systems. Whole organisms, on the other hand, are made up of various cell types that have differentiated in their natural environment, working together as a whole. Studies with multicellular models enables us to see the effect of various internal and environmental stimuli in different cell types. They are also essential to study animal developmental biology. In this regard, *Caenorhabditis elegans, Drosophila,* zebrafish and mice in particular are important models. Equally, *Arabidopsis thaliana* makes a great model for plant development.

Caenorhabditis elegans

One of the most revolutionary contributions of *C. elegans* to research was the discovery of RNA interference (RNAi) in this worm. Genes are easily knocked-down through RNAi by feeding the worm bacteria expressing plasmids that contain a gene of interest (Fire et al., 1998). RNAi has since been discovered and exploited for targeted depletion to study gene function in other organisms. But *C. elegans* mutants are still important models for studying, for example, the neuromuscular system. Worm mutants for neuromuscular disorders are in fact sometimes used by researchers as a first pass for testing potential drug treatment [reviewed in (Sleigh and Sattelle, 2010)].

Drosophila melanogaster

D. melanogaster is frequently used in forward genetics studies to trace the genetic basis of observable phenotypes. The fact that this fly has a small genome made up of 4 chromosome makes it relatively easy to trace mutations back to specific genes. The fly shares ~75% of genome sequence similarity with humans, providing a simple system to elucidate and better understand aspects of human physiology and disease. For example, our understanding of the molecular mechanism behind the circadian rhythm is heavily based on experiments using *D. melanogaster* [reviewed in (Dubowy and Sehgal, 2017; Peschel and Helfrich-Förster, 2011)].

Danio rerio

While zebrafish are important in developmental biology, they are also the organism in which scientists discovered morpholino, an oligomer of DNA bases linked to a methylenemorpholine backbone. With the ability to block access to the RNA, morpholinos have been used as molecular tools for knocking down genes to investigate function (Nasevicius and Ekker, 2000).

Mus musculus

Mice bear much genetic and physiological similarity to humans, and are useful for elucidating complex systemic interactions and therapeutic interventions. The ability to create knockout mice is a valuable research tool for investigating gene function. Cancer, diabetes, obesity, Parkinson's and heart disease have all been studied using mouse models.

1.1.3 Eukaryotic diversity and model systems

Eukaryotes represent a highly diverse domain of life. The group exhibit a high level of genomic variation which has no established relationship to their size (Lynch and Conery, 2003). The recovery of phylogenetic relationships between the deep classes of eukaryotes is a challenging task. The phylogenetic tree of eukaryotes has seen very dramatic changes as more genome sequences become available. Many modern studies agree on 6 supergroups comprising the Amoebozoa, Excavata, Opisthokonta,

Archaeplastida, Rhizaria and Chromalveolates. The proposed 5-clade tree, which is now the most widely accepted interpretation brings Rhizaria and most members of Chromalveolates into a single clade named SAR (Figure 1.1) (Burki, 2014; Roger and Simpson, 2009).

The diversity of eukaryotes means that using a model from one or two groups alone to generalise cellular processes is an inadequate representation of the existing diversity. The most common eukaryotic models belong to only one of the major eukaryotic groups – Opisthokonts (animals and fungi) – with contributions also from Archaeplastida (plants and algae).

This narrow selection of representative models severely under-represents the diversity of eukaryotic life. It has become evidential by the increasing differences identified in key biological processes that extrapolations from models must be done cautiously. Additionally, it has created a view in the scientific community that anything that departs from biological processes in Opisthokonts is “non-canonical”.

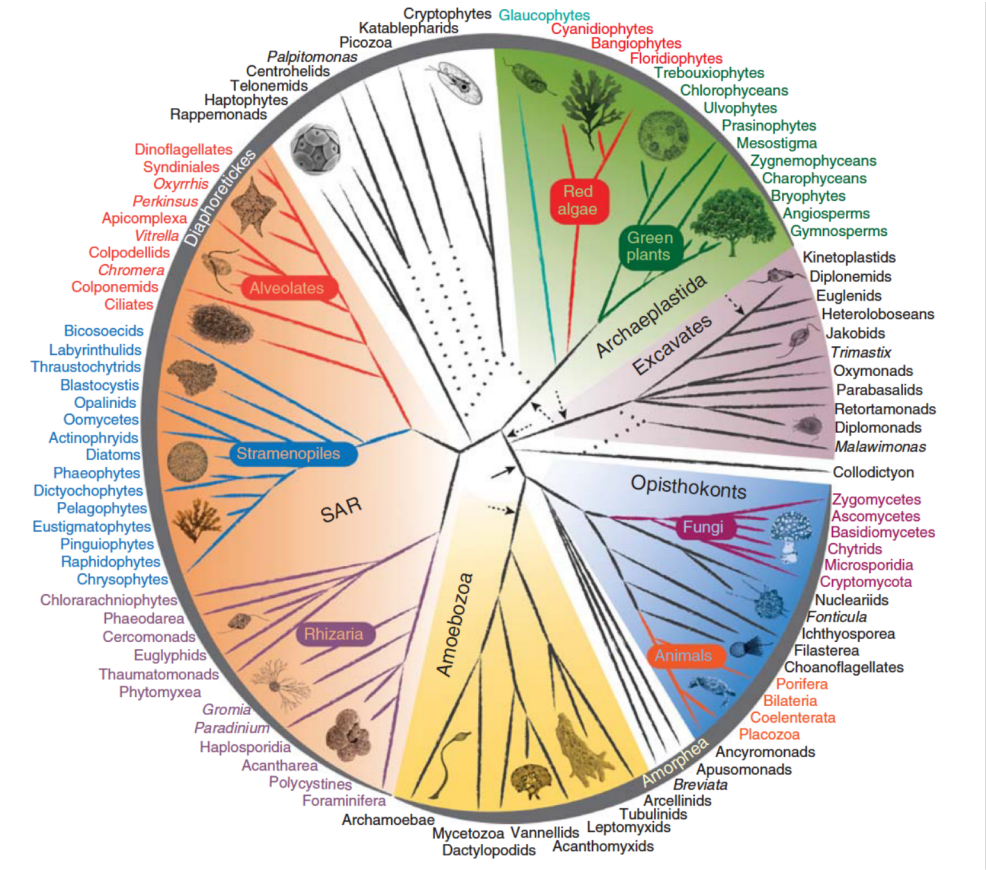


Figure 1.1: A phylogenetic tree showing the current understanding of diversity of eukaryotes. Five major clades are highlighted in colour(Archaeplastida, Excavates, Opisthokonts, Amoebozoa, SAR) and a number of lineages yet to be placed [Reproduced from (Burki, 2014)]

1.1.4 Biology from non-traditional model systems

Traditional model organisms, on the basis of their characteristics and multiple advantages, lend themselves to a variety of research investigating common biology. In contrast, special features associated with certain organisms make them ideal for investigating particular research questions. For example, research involving telomeres and the molecular basis underlining how they function to preserve chromosome length was investigated using *Tetrahymena* as a model. Unlike traditional models, *Tetrahymena* contains ~275 transcriptionally active chromosomes, each with a ploidy of 45-50 (Ruehle et al., 2016). The fact that several chromosomes are present enabled identification of a repetitive TTGGGG sequence that marks the chromosomes ends and aided the first discovery of telomeres (Blackburn and Gall, 1978). Similar repetitive sequences were later found in other organisms as a way of protecting their chromosome (Greider and Szostak, 2006). Work in *Tetrahymena* also discovered the

enzyme telomerase (Greider and Blackburn, 1985), and the involvement of telomeres in cell aging and chromosome healing (Blackburn et al., 1983).

RNA editing in eukaryotes was first discovered in trypanosomes [reviewed in (Read et al., 2016)]. This unicellular excavate has been the model to understand the concept and process of RNA editing, which involves site-specific post-transcriptional modification (other than splicing and polyadenylation) of the RNA sequence. The process involves nucleotide insertions and modifications such as deamination of bases which alters the resultant amino acid sequence of the protein. The first description in trypanosomes was the insertion of 4 uridine residues to mitochondrial mRNA encoding cytochrome c oxidase. This insertion repairs a frameshift in the gene (Benne et al., 1986).

To study specific biology, it is imperative to adopt a non-canonical model with close biological semblance to the organism of interest. The apicomplexan group consist of a wide range of organisms including some of the most important causes of parasitic disease – *Plasmodium, Toxoplasma, Babesia* and *Cryptosporidia*. While they share some specific biology (e.g. apicoplast used to bore through host cells), each parasite presents unique features which influence disease presentation and severity. *T. gondii* for instance infects and form cysts in the host brain and is frequently associated with neurological disorders. *Plasmodium spp*. on the other hand, primarily infects blood where invasion and destruction of red blood cells results in anaemia.

1.1.5 Emerging models

For many biological questions, the traditional models do a great job. As the scope of modern investigations become broader, the most appropriate models to address research questions may well be a less-studied organism that belongs to groups known to have 'non-canonical' biological processes. There is therefore the need to expand our range of models to encompass a wider set of organisms and diversity.

Genetic tractability is one of the important elements that is facilitating the renaissance of non-traditional models. As an organism becomes culturable, and its genome becomes 'malleable' (among other things), it becomes easier and more desirable to

exploit its unique biology to fill in the missing gaps in our understanding. In this regard, CRISPR/Cas has given popularity to previously unexplored organisms.

Ideal characteristics of an eukaryote to be a new model would be:

1. The organism must occupy a biological niche – a scientific question it can be used to investigate.
2. It must be cheap and fast to grow, to ensure quick turnover.
3. Its genome must be sequenced.
4. Axenic/ex-vivo culture. An ideal new model should have a means of culturing continuously *in vitro.* Growth should be supported in defined or semi-defined medium and allow generation and sustaining of a clonal cell line. For an organism with multiple stages in its lifecycle, being able to culture more than one stage is desirable.
5. Recapitulate infection biology. When a disease-causing organism is taken from its natural host into *in vitro* culture, it should retain infectivity as evidence that it has not lost important characteristics unique to infection.
6. Genetic modification. The ability to modify an organism's genome is an extremely favourable trait. Ideally, both forward and reverse genetics should be available for investigating gene function.
7. Genome-wide technology. The availability of a genome-wide technology in the tools available for any model permits more complex investigations. The scope of research possibilities become wider when mutants can easily be generated at large scale.
8. Sex/cross. The ability to mate offers the opportunity to investigate the process of sexual reproduction.

An emerging model does not need to have all the listed traits but the more of it present, the more versatile and useful it will be.

In what follows, I will make a case for trypanosomes as an important model for studying many aspects of eukaryotic biology. I will describe features that make them suitable for investigating some common biology of eukaryotes, in particular the family

of motor proteins kinesins, as well as discussing the need for new models within the trypanosome lineage.

1.2 African Trypanosomes

1.2.1 Importance

Trypanosomes are kinetoplastids and they include multiple disease-causing species. African trypanosomes are significant causes of human and livestock diseases. Two subspecies of *Trypanosoma brucei, T. b. gambiense* and *T. b. rhodesiense* (Figure 1.2) cause human African trypanosomiasis (HAT). Disease burden due to morbidity and mortality is estimated at 1.6 million disability-adjusted life years in some of the poorest communities (WHO, 2014). The ability of these sub-species to cause HAT is due to an inherent resistance to human trypanolytic factors (Pays and Vanhollebeke, 2009). In addition to causing HAT, *T. brucei* causes animal African trypanosomiasis (AAT). There are other species that also cause AAT, and these are frequently associated with specific animal hosts. For instance, *T. suis* and *T. simiae* commonly infect pigs, while *T. evansi* infect camels. Three species: *T. congolense, T. vivax and T. brucei* infect cattle (Figure 1.2). Cattle AAT is a wasting disease also known as nagana. It renders huge economic losses to the African region as a result of agricultural impact. Beef production losses alone are estimated to be $1 – 1.2 billion per year (UN, 2014). In addition, economic strain is also imposed by chemotherapy treatments that are continuously administered to livestock (UN, 2014). Of the economically important species, *T. congolense* is responsible for the vast majority of cattle infections (Muhanguzi et al., 2017; Sheferaw et al., 2019; Tchamdja et al., 2017).

Most trypanosomes are transmitted by biting insects. Exceptions are an aquatic clade which is transmitted by leeches and infect fish, amphibians and aquatic reptiles, and *T. equiperdum,* transmitted through direct transfer in equines. For those transmitted by insects, the *T. cruzi* clade (Figure 1.2)*,* causative agents of Chagas disease, is spread by triatomine bugs in Central and South America. The *T. brucei* clade of African trypanosomes is primarily transmitted by tsetse flies (*Glossina spp.*) in sub-Saharan

Africa. The closely-related protozoan *Leishmania spp.* also causes significant human disease, and is transmitted by female phlebotomine sandflies.

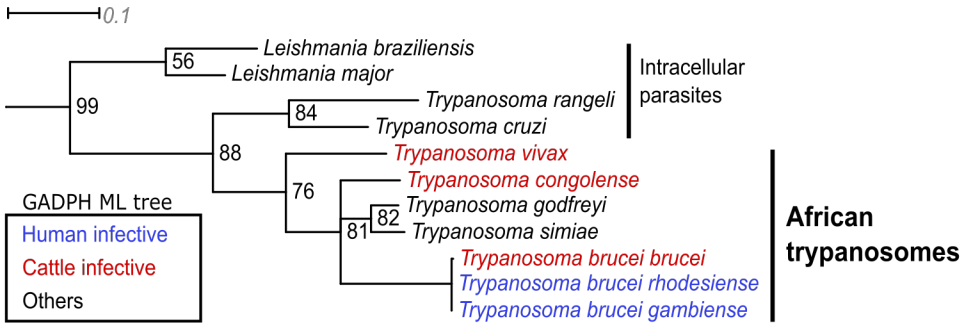


Figure 1.2: The phylogeny of the Trypanosomatida. Shown is a maximum likelihood tree based on alignment of GAPDH genes. Numbers by nodes are support values from 100 pseudo-replicates of the analysis. Courtesy of Bill Wickstead (unpublished).

1.2.2 African trypanosome biology

Lifecycle

The lifecycle of *T. brucei* is shown in Figure 1.3. Cyclical transmission is mediated by the tsetse fly vector as it takes a blood meal from an infected host. The lifecycles of *T. congolense* and *T. vivax* are similar but differs in development within the tsetse host (see below). In addition to cyclical tsetse fly transmission, *T. vivax* is also transmitted mechanically by haematophagous insects like the Tabanid fly (Desquesnes and Dia, 2003) which has caused it to move beyond the confines of the "tsetse fly belt" of Africa into South America (Jones and Dávila, 2001).

When a tsetse fly ingests mammalian blood with the parasites, *T. brucei* and *T. congolense* move to the fly midgut where they differentiate into proliferative procyclic forms and then to proventricular forms before migrating to the mouthpart (Figure 1.3). In the mouthpart they colonize different areas – *T. brucei* migrates to the salivary gland, while *T. congolense* goes to the proboscis (Peacock et al., 2012; Vickerman, 1985). *T. vivax,* on the other hand, has a truncated lifecycle in the insect host: it does not go to the midgut but remains yet in the proboscis (Gardiner, 1989). As such, *T. vivax* does not have a procyclic stage. Cells in the insect mouth differentiate into dividing epimastigotes and then non-dividing metacyclics prior to transmission to mammalian host (Figure 1.3). Metacyclics become slender, proliferative

trypomastigotes in the bloodstream of infected animals. In response to density-dependent quorum sensing, slender bloodstream forms of *T. brucei* and perhaps *T. vivax* differentiate into stumpy cells which are taken up by the insect for the cycle to continue (Figure 1.3).

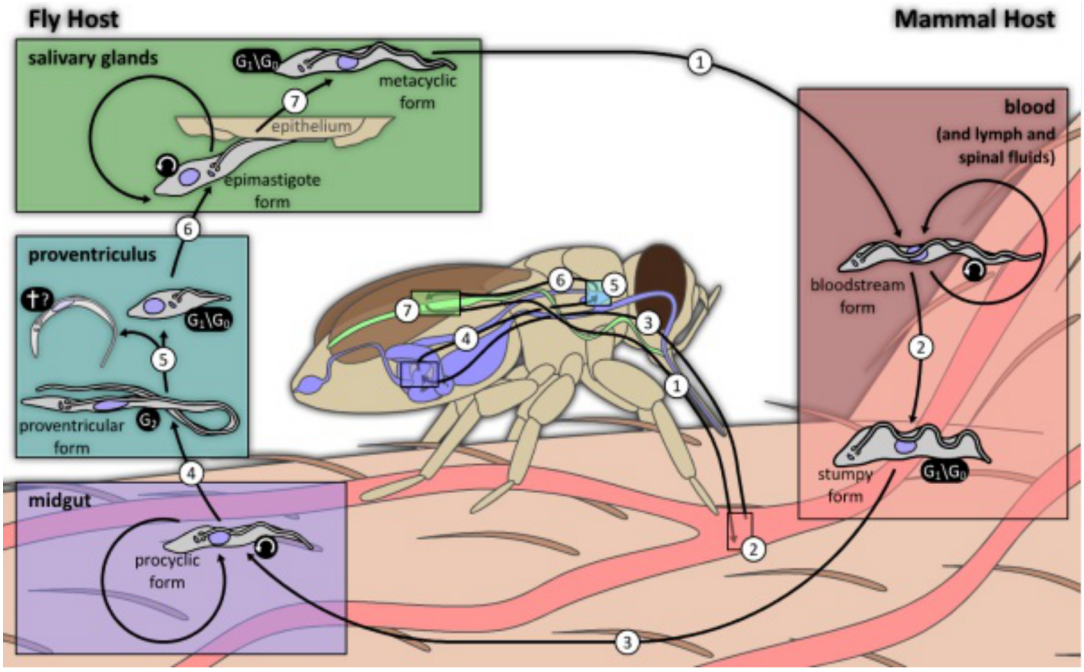


Figure 1.3: An outline of the life cycle of *Trypanosoma brucei* through the mammalian and Tsetse hosts [Taken from (Wheeler, 2014)].

Major surface proteins

Trypanosomes undergo various changes during their lifecycle, including adaptation of the surface coats at each stage in response to the environment (Figure 1.4). Bloodstream-form *T. brucei* is covered by a thick monolayer of ~10^7 copies of a single variant surface glycoprotein (VSG) of ~60 kDa in size (Cross, 1975). The VSG coat changes periodically in the process of antigenic variation to create "waves of parasitaemia" as an immune invasion strategy to ensure parasite survival in the host. VSG first appears on metacyclic cells. There are 5 metacyclic VSG expression sites identified in the *T. brucei* laboratory strain Lister 427 (Cross et al., 2014). *T. brucei* VSG is structured with 350-400 amino acid residues forming a hypervariable N-terminal domain, and 40-80 residues making up a conserved C-terminal domain which is GPI-anchored to the plasma membrane (Ferguson et al., 1988; Turner et al., 1988).

Frequent recombination from a repertoire of ~3000 VSG genes introduces significantly different surface coat variation (Hall et al., 2013; Marcello and Barry, 2007).

Similarly, both metacyclic and bloodstream-form *T. congolense* are covered by VSG (Rovis et al., 1978) but bear some differences. First, *T. congolense* VSG is smaller (~48 kDa) (Fish et al., 1989). And, while it has the N-terminal signal peptide sequence and C-terminal GPI insertion sequence, the C-terminal domain is not conserved as in *T. brucei. T. vivax* VSG is smaller still, ~40 kDa, and appears to have very little carbohydrate decoration (Gardiner et al., 1996).

A cell surface phylome of VSGs of *T. brucei*, *T. vivax*, and *T. congolense*, produced from genome sequence analysis, reveals that VSG from the different parasites arise from different lineages (Jackson et al., 2012). *T. congolense* VSG repertoire is derived from multiple ancestral lineages and is more heterologous than the *T. brucei* repertoire which is of more recent origin. *T. brucei* VSGs consist of a- and b- type VSG (Jackson et al., 2012). *T. congolense* has no a-VSG but has two b-VSG subfamilies which are structurally more diverse. *T. vivax* is even more diverse, with 4 VSG families: a- and b-VSGs plus two subfamilies unique to *T. vivax*, representing ancestral lineages that have been lost in the other two species. Predicted VSG pseudogenes are fewer in *T. congolense* (21-30%) and *T. vivax* (16-27%) compared to 70% in *T. brucei.* While *T. brucei* VSGs share a common C-terminal domain, *T. congolense* has 15-20 different such domains unrelated to that of *T. brucei* sequence (Jackson et al., 2012).

In the insect midgut stage, *T. brucei* is covered by a procyclin coat (Figure 1.4), which is composed of four different proteins GPEET, EP1, EP2 and EP3, named on the bases of repeats in their amino acid sequence (Bütikofer et al., 1997; Roditi et al., 1998). *T. brucei* and *T. congolense* epimastigotes are specifically coated by alanine-rich protein BARP (Urwyler et al., 2007) and glutamine and alanine-rich glycoprotein GARP (Bayne et al., 1993; Beecroft et al., 1993), respectively (Figure 1.4).

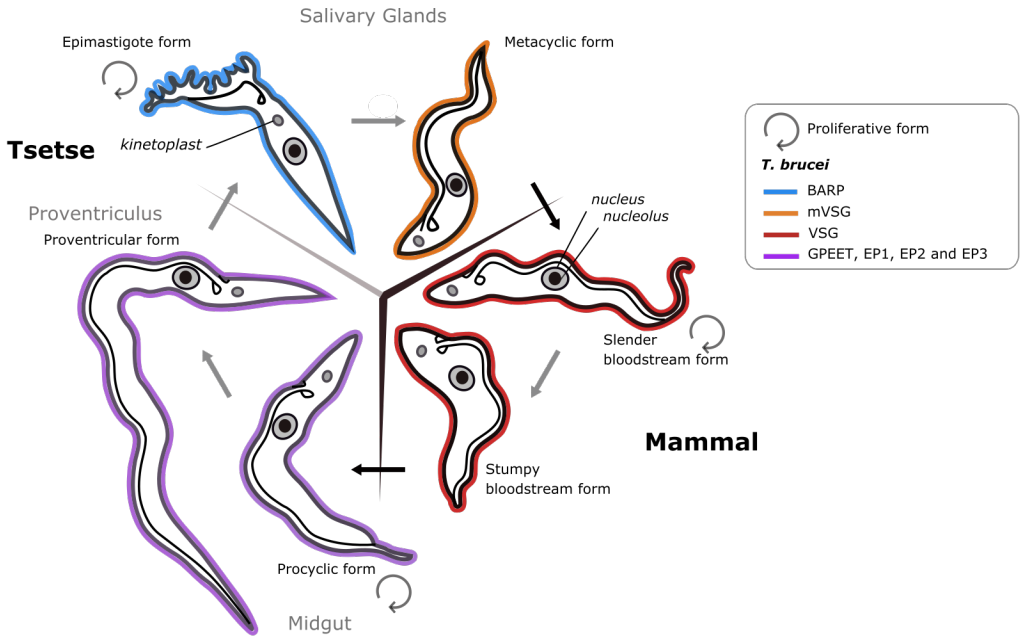


Figure 1.4: Schematic representation of the changing surface coat of *T. brucei* through the lifecycle. Black arrows indicate transmission between tsetse and human host (Courtesy of Jennifer McDonald).

Genomic differences

The diploid *T. brucei* genome consists of 11 pairs of megabase-sized chromosomes (MBCs), ~5 intermediate chromosomes (ICs) and ~100 minichromosomes (MCs) (Ersfeld et al., 1999). Sequencing the genome of *T. brucei* (strain TREU 927) has shown MBCs to contain 9068 predicted genes (Berriman et al., 2005) . Housekeeping genes are located on MBCs, whereas ICs and MCs contain a collection of silent VSG genes and regions of 177 bp repeats. *T. brucei* MCs are transcriptionally silent. A draft genome sequence for *T. congolense* (strain IL3000) and *T. vivax* (strain Y486) (Jackson et al., 2012) have 11 assembled pseudochromosomes mapped to the *T. brucei* MBCs, showing high level of gene synteny. *T. congolense* minichromosomes are smaller than those of *T. brucei* (25-50 kb), with a 369 bp repeat region (Gibson and Borst, 1986). Transcription status of MCs are unknown. The average length of an ORF is similar between *T. brucei* and *T. congolense,* but *T. congolense* have longer intergenic regions.

Disease pathology, treatment and drug resistance

Different species of African trypanosomes produce distinct infections in mammals. *T. congolense* remain in the vascular system, adhering to erythrocytes and vascular

walls (Banks, 1979; Hemphill et al., 1994). In addition, *T. congolense* is known to sequester in the vascular walls of the brain, while *T. vivax* remain evenly distributed in the blood vessels. *T. brucei* and *T. vivax* are able to migrate out of the bloodstream into tissues, lymph nodes and interstitial spaces (Kennedy, 2013), which explains why *T. brucei* infections are marked by neurological disorders both in HAT and AAT.

Without treatment, trypanosomiasis can be fatal, but available drugs for treatment of trypanosomiasis are old and have high toxicity (Table 1.1). In response to instituted WHO control programmes, the number of HAT cases dropped by 82% between 1998 (37,991 cases) and 2011 (6,743 cases) (Simarro et al., 2012). Hence, HAT was targeted for elimination as a public health threat in 2020, and eradication by 2030 (Holmes, 2014; Savioli, 2012). Elimination targets have since been surpassed at each surveillance, and fewer cases than anticipated are reported (Barrett, 2018; Franco et al., 2017; Francoid et al., 2018). The introduction of a new treatment drug, Fexinidazole (Kaiser et al., 2011; Torreele et al., 2010) gives more hope for eradication.

For AAT there are 6 commercial chemotherapy treatments currently available (Table 1.1). Diminazene aceturate and isometamidium chloride are the most heavily relied on, with isometamidium being the only therapeutic drug used in the field. Resistance to both drugs have been reported across the African continent. Most reports of AAT drug resistance involve *T. congolense* infections (Afewerk et al., 2000; Moti et al., 2012; Mulugeta et al., 1997), which could reflect the fact that *T. congolense* is the most prevalent species (Tchamdja et al., 2017).

Hidden differences

Most of the knowledge we have about African trypanosomes is drawn from *T. brucei*. Yet, there are hidden differences or speculated similarities among the African trypanosomes, and these remain largely unknown because of the paucity of data on *T. congolense* and *T. vivax*. The matter of quorum sensing is one such area of interest because of its importance in pathogenicity. In *T. brucei,* quorum sensing is a density-dependent response to a parasite-derived signalling factor that causes differentiation of proliferative slender bloodstream forms to non-proliferative stumpy forms for transmission (Vassella et al., 1997). In *T. congolense* and *T. vivax*, no clear heterogenic

forms have been reported in mammalian host, although orthologues of some of the genes involved in stumpy differentiation are present. Both species have been shown to arrest at the G1 phase of the cell cycle, suggesting that a non-proliferative form exists in the blood (Shapiro et al., 1984). Finally, *T. congolense* conditioned medium can induce stumpy formation in *T. brucei* (Silvester et al., 2017).

Drug	Species	Toxicity
HAT treatment		
Suramin	*T. brucei* (early stage)	Possible renal failure, skin lesions, anaphylactic shock, bone marrow toxicity, and neurological complications
Pentamidine	*T. brucei* (early stage)	Hyperglycaemia, hypoglycaemia, hypotension, gastrointestinal features
Melarsoprol	*T. brucei* (late stage)	Extremely painful injections, post-treatment reactive encephalopathy
Eflornithine-nifurtimox	*T. brucei* (late stage)	Eflornithine is associated with bone marrow toxicity, alopecia, seizures, and gastrointestinal symptoms
AAT treatment		
Diminazene aceturate	*T. congolense, T. vivax*	Toxic to horses, donkeys, dogs and camels.
Ethidium bromide/chloride	*T. congolense, T. vivax*	Toxic to horses. Potentially carcinogenic
Isometamidium chloride	*T. congolense, T. vivax*	Toxic at high doses. Highly irritant.
Quinapyramine sulphate	*T. brucei, T. congolense, T. vivax*	Toxic at high doses. Fast resistance acquisition
Suramin sodium	*T. brucei*	Severe necrosis at injection site
Melarsamine dihydrochloride	*T. brucei*	Minor tissue reaction

Table 1.1: Trypanosomiasis treatment drugs and toxicity (Fairlamb, 2003; Giordani et al., 2016)

1.3 Trypanosomes as a model

Trypanosomes (Supergroup Excavata) are highly diverged from model systems belonging to Opisthokont or Archaeplastida. This divergence serves as an advantage to the adoption of trypanosomes as research models, as they provide further insight to the very diverse eukaryotic biology. They grow fast dividing every 8 hours, with established system for *in vitro* culture (see below), and some laboratory strains are able to recapitulate infection biology. The availability of sequenced genomes of

African trypanosomes, and advanced genetic technology in *T. brucei* are also advantageous.

1.3.1 Axenic culture

One of the conditions for using any organism as a model for research is the availability of axenic culture. *T. brucei* was put ahead of the other African trypanosomes as the first to be successfully cultured *in vitro.* Two stages of the parasite have so far been adapted to long-term axenic culture – the insect midgut procyclic stage and the mammalian bloodstream stage. Long-term axenic culture was first established for procyclics. Successful attempts defined HX25 (Cross and Manning, 1973) and Cunningham's medium (SM), a modified version of Grace's insect medium which replicates the amino acid composition of tsetse fly haemolymph (Cunningham and Honigberg, 1977). HX25 did not support high cell density, but a modified version, SDM-79 gave increased cell density by decreasing glucose concentration while increasing that of pyruvate, folic acid and amino acids which are the main source of energy in procyclics, and finally the addition of 15% foetal bovine serum (FBS) (Brun and Schönenberger, 1979).

Initial cultures of bloodstream forms required the use of mouse fibroblasts or bovine aortic endothelial cells as feeder cells (Hirumi et al., 1977b, 1977a). Feeder cells were the source of cystine, an essential amino acid for *T. brucei* but also toxic. Later innovation created a system to generate cystine in culture by adding cysteine along with β-mercaptoethanol and bathocuproine to the medium. Cystine is reduced to provide an alternative source of cysteine*,* discounting the need for feeder cells (Duszenko et al., 1985). Following these reports, the much widely used bloodstream culture medium, HMI-9 (Iscove's modified Dulbecco's modified Eagle's Minimal Essential Medium (IMDM) supplemented with bathocuproine, cysteine, cytosine, hypoxanthine, β-mercaptoethanol, uracil and thymidine), was established (Hirumi and Hirumi, 1989).

By dropping the temperature from 37 ^{o}C to 28 ^{o}C and adding 3 mM citrate and or cis-aconitate, bloodstream-form *T. brucei* can be differentiated into procyclic forms (Brun

and Schönenberger, 1981). *In vitro* attempts to differentiate procyclic to epimastigote to metacyclic forms by manipulating culture constituents often resulted in failure (Cunningham, 1977; Cunningham et al., 1981). Hirumi and colleagues report a successful differentiation from procyclics to metacyclics in continuous culture with feeder cells. Epimastigote and metacyclic forms appeared in culture along with procyclics at 28^{o}C (Hirumi et al., 1992). However, the reason for the observed differentiation was not clarified, and the study does not make any distinction whether observed metacyclics are truly differentiated *in vitro* instead of previous infective forms simply surviving at low temperature. Thus, there is no clear means of differentiating *T. brucei* procyclics to epimastigote and metacyclic forms by manipulating culture conditions.

However, it is possible to do so by genetic manipulation. Recently, RNA-binding proteins (RBPs) have been associated with different life stages and differentiation. RBP6 is enriched in insect proventricular trypanosomes, and its overexpression results in production of epimastigotes and metacyclic forms in culture (Kolev et al., 2012, 2014). Similarly, RBP10 overexpression in procyclics produced bloodstream-stage *T. brucei* (Mugo and Clayton, 2017; Wurst et al., 2012). At present, RBP-induced differentiation requires a genetic manipulation step and may still be a while away from being established and widely accepted as a means of differentiating *T. brucei.*

The establishment of long-term axenic culture of other African trypanosomes has lagged behind those of *T. brucei*. Culture of *T. vivax* is still a challenge, with no defined axenic culture established for bloodstream forms, which are mostly passaged in mice. Good advances have been achieved in *T. congolense in vitro* culture (see section 6.2.2 for review).

1.3.2 Genetic modification in Trypanosomes

Direct modification of the genome

Direct modification of the *T. brucei* genome has been an important key factor driving increased use of *T. brucei* in research. The most exploited way of doing this for the past few decades has been by piggybacking on the homologous recombination

machinery for integration of transgenes. Homologous recombination is a means by which cells repair double strand breaks during replication. As linearised DNA constructs are introduced into the cell by electroporation, the cellular machinery of dividing cells is tricked into incorporating the construct bearing sequence homology with the target region into the desired locus. By introducing a linear construct of the neomycin selection marker, *NEO,* flanked by upstream and downstream sequences targeting the tubulin locus, researchers demonstrated the insertion of *NEO* at the tubulin gene cluster in procyclic *T. brucei* (ten Asbroek et al., 1990) and then in bloodstream cells (ten Asbroek et al., 1990).

Efficiency of incorporating transgenes is influenced by the sequence homology as well as the length of the homologous region (Barnes and McCulloch, 2007; Bell and McCulloch, 2003). Endogenous gene targeting has been used to knockout genes to analyse gene function, and also to tag genes with fluorescent protein or epitope tag (Daniels et al., 2012) for protein localisation or purification.

CRISPR/Cas9

Clustered regularly interspaced short palindromic repeats (CRISPR) DNA sequences and Cas (CRISPR-associated) protein system of adaptable immune defence against foreign nucleic acids in prokaryotes is emerging as another technology for modifying the genome of trypanosomatids (reviewed in Lander and Chiurillo, 2019). Cas proteins process long precursor RNA into ~40 base nucleotide CRISPR RNA (crRNA) that serve to guide it to target nucleic acids for cleavage. The discovery that engineered guide RNA (gRNA) sequences acting like crRNAs direct *Streptococcus pyogenes* Cas9 protein activity *in vitro* to induce double strand break established its potential application in genome editing (Jinek et al., 2012). While the research community have put a lot more effort in developing CRISPR/Cas9 toolkit for other trypanomatids (reviewed in Lander and Chiurillo, 2019), its application in *T. brucei* is quite recent, perhaps due to the vast tools already available for this species. A CRISPR/Cas9 tool kit was first reported in *T. brucei* in 2017 (Beneke et al., 2017). A year later, an inducible version of the system was produced which also promises to produce higher efficiency of gene targeting than what is seen with the use of homing endonucleases (Rico et al., 2018). With this high

efficiency, CRISPR/Cas9 may very well be a landmark for even more complex genetic manipulation which is currently not attainable.

Inducible gene expression systems

To enable the controlled expression of proteins, including those toxic to cells and also for mutational analysis, a prokaryotic inducible expression system was adapted in *T. brucei.* Wirtz and Clayton (1995) developed the first of such systems. They first introduced a coding sequence of the *Escherichia coli* tetracycline repressor (TetR) at the parasite tubulin locus. By introducing a construct with a tetracycline operator (TetO) placed closely downstream of a procyclin promoter, transcription of the gene of interest can be turned on or off by the presence or absence of tetracycline in the culture. When integrated genes are placed under the control of a bacteriophage T7 promoter in *T. brucei* cells that express T7 RNA polymerase, a higher amount of protein was produced which could also be tightly regulated by the tetracycline operator and repressor system (Wirtz et al., 1998, 1999a). A modification of the system expressed both T7 RNA polymerase and TetR from a single DNA construct (Poon et al., 2012). Using an externally acquired promoter for inducible transcription makes it applicable in different stages of the parasite lifecycle, avoiding the possibility of endogenous machinery hijacking control of promoter activity at a different parasite stage (Biebinger et al., 1996). In a similar manner, the rDNA promoter was also modified for tetracycline inducibility (Alsford et al., 2005). Other more recent inducible systems are the vanillic acid (Sunter, 2016) and cumate- (Li et al., 2017) inducible expression systems. Both use the same ON-OFF principle mediated by the presence or absence of vanillic acid or cumate respectively.

Efficient regulation of gene expression by any inducible system depends on the integration locus. Different integration loci are known to have position effect and may result in leaky expression in the “off” state. An actively transcribed locus like the tubulin locus is a poor choice for externally controlled inducible expression. To have a tightly controlled “off” state would require the prevention of readthrough of integrated DNA from endogenous transcription. The region upstream of the promoter of the 18S rRNA (White et al., 1986) had become the preferred locus for integration

of DNA constructs for inducible expression, but has to be targeted in a manner that allows transgenes to run in opposite to the direction of endogenous transcription to avoid readthrough (Wirtz and Clayton, 1995). Characterisation of the minichromosomes of *T. brucei* as a transcriptionally silent and repetitive regions (Alsford et al., 2001; Zomerdijk et al., 1992) presented a better option for efficient inducible regulation to circumvent the problem of readthrough. Wickstead *et al* (2002) demonstrated ectopic expression of transgenes from minichromosome loci (177 bp repeat and silent VSG loci) and revealed that integration at the rDNA spacer region provided very poor regulation in at least a subset of clones produced.

RNA interference

RNA interference (RNAi) is a mechanism of regulating gene expression by controlling mRNA levels. Three pathways involving small interfering RNA (siRNA), microRNA (miRNA) or Piwi-interacting RNA (piRNA) are responsible for regulating RNA levels in cells. Both miRNA and siRNA are involved in post-transcriptional gene silencing, while piRNA regulates gene expression at the level of transcription. Experimental application of RNAi takes advantage of the siRNA pathway to deplete target protein levels without knocking out the gene. Double-stranded RNA introduced into the cell becomes a template for the siRNA pathway. The ribonuclease Dicer cleaves the RNA into ~20bp siRNA fragments. Single-stranded siRNA, together with the protein Argonaute, form the RNA-induced silencing complex (RISC). RISC is then targeted to RNA complementary to the sequence of the siRNA and activates its cleavage (Ipsaro and Joshua-tor, 2015; Wilson and Doudna, 2013). RNAi silencing can be achieved experimentally in mammalian cells by introducing synthetic siRNAs (Zou et al., 2003), pol III transcribed shRNAs (a stem loop of ~50 nucleotides and 2 nucleotides of 3'-overhang at one end) (Brummelkamp et al., 2002; Paddison et al., 2002) or miRNAs.

A functional RNAi machinery improves the genetic tractability of *T. brucei.* Gene knockdown by RNAi is well established in this species (Ngo et al., 1998), and does not require direct introduction of dsRNA. Instead, a gene fragment can be placed between two opposing inducible promoters to produce a head-to-head dsRNA *in vivo* (Figure 1.5), or by using a DNA plasmid that produces hairpin RNA. Coupling inducible systems

with RNAi enables the experimental control of expression of proteins (Alibu et al., 2005; Bastin et al., 2000; Shi et al., 2000) allowing functional analysis of individual genes.

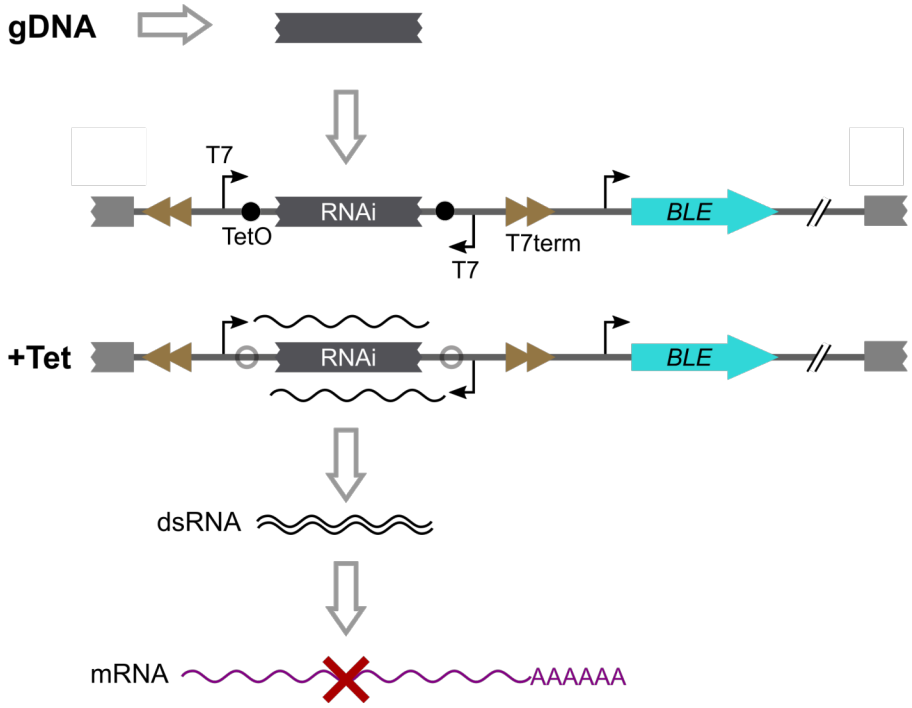


Figure 1.5: Schematic representation of a common system for RNA interference in *T. brucei.* Genomic DNA (gDNA) is introduced between two opposing T7 in the tetracycline inducible expression system. The presence of tetracycline (+Tet) releases the repressor protein from the operator (TetO) to allow transcription of a dsRNA which activates the siRNA pathway.

Genome-wide screening

Application of RNAi for genome-wide screening among trypanosomes and related species has to date only been possible in *T. brucei.* The first genome-wide RNAi screen was performed in procyclic forms to identify genes that play a role in glycosylation of the procyclin surface coat (Morris et al., 2002). Randomly sheared (by sonication) genomic DNA fragments were ligated into a tetracycline-inducible head-to-head RNAi construct and transfected into procyclic cells at an efficiency to provide ~ 5x genome coverage. Cells were selected for resistance to concanavalin A (conA) binding after RNAi induction. In this way, they identified a cell line that had silenced the hexokinase gene, highlighting that glycolysis is necessary for the glycosylation of procyclin.

Genome-wide RNAi-mediated loss-of-function screens have been used to identify proteins involved in trypanocidal drug mechanism of action and resistance (Baker et

al., 2011; Schumann Burkard et al., 2011). Selecting for cells resistant to eflornithine and melarsoprol in a genome-wide screen led to the identification of the amino acid transporter AAT6, and the adenosine transporter TbAT1 respectively. These gene products are involved in the drug uptake, and hence knockdown confers drug resistance (Schumann Burkard et al., 2011). A similar screen with eflornithine and nifurtimox confirmed the AAT6 transporter and found resistance to nifurtimox was due to silencing of the enzyme nitroreductase (Baker et al., 2011).

The era of next generation sequencing (NGS) introduced a method that would further increase the complexity of trypanosome research. For instance, NGS for transcriptome analysis (RNA-seq) (reviewed in Siegel *et al.*, 2011) has been used to analyse transcriptome differences between different trypanosome life stages (Silvester et al., 2018). In a revolutionary adaptation of genome-wide RNAi library, Alsford *et al.* used a high-throughput approach to conduct genome-wide RNAi target sequencing (RIT-Seq) and generated in the process fitness readout for almost all core genes (~7500) in *T. brucei* (Alsford et al., 2011). The RIT-seq technique, similar to previous libraries, inserted randomly fragmented genomic DNA into an RNAi construct for integration into both bloodstream, procyclic and differentiating cells. Genomic DNA from tetracycline-induced and non-induced cultures was extracted, RNAi target amplified, fragmented and then sequenced by NGS. Reads were mapped to the genome and the gene abundance of the induced verses the non-induced was used to determine the fitness cost of each gene. The RIT-seq data highlighted essential genes like those involved in transcription, translation and flagellar function, as well as those involved in differentiation like glycolytic proteins. Adaptations of RIT-seq has since been used to study several trypanocidal drugs (eflornithine, suramin, nifurtimox, pentamidine, melarsoprol, isometamidium) which confirms previous studies and elucidates the multi-resistance facilitated by knockdown of certain transporters while identifying other mechanisms of resistance (Alsford et al., 2012; Baker et al., 2015). These data facilitate our understanding of the biology of *T. brucei*, important aspects of disease treatment and control, and encourages further use of this model.

1.3.3 Current limitations

In a lot of ways *T. brucei* fits the idea of a model described previously but does not match all the characteristics. For instance, in terms of biological niche *T. brucei* covers very well the HAT niche but does not cover the range of trypanosome infections generally. There is a need for a model that better addresses AAT. Secondly, the commonly used laboratory strains Lister 427 (L427) and EATRO1125 differ in animal model infectivity. L427 is a monomorphic strain, meaning that bloodstream forms continue to divide as long slender *in vitro* without responding to density-dependent quorum sensing (Hirumi et al., 1980). While this is an advantage for *in vitro* work, the same cannot be said for L427 animal infections. *In vivo,* L427 divide exponentially in the host bloodstream until animal death. It does not differentiate to stumpy form in response to increasing parasite density and does not mimic the characteristic wave of parasitaemia associated with chronic infection in natural hosts.

EATRO1125 on the other hand is a pleomorphic strain which shows varied morphology in cultured bloodstream stage – from slender to 'intermediate' and stumpy forms. Pleomorphism is generally characteristic of a strain that has not been adapted to *in vitro* culture for too long, and a sign that strains are closer to the wild type and may provide a more natural infection model. Hence, for animal studies, EATRO1125 is the preferred strain. Mouse infection models with EATRO1125 mimics the disease more closely, showing waves of parasitaemia and stumpy formation (Reuner et al., 1997). While this provides a disease model for research, the fact that mice are not the natural host of African trypanosomes poses a limitation to what can be investigated. For example, differences in immune response does not allow detailed infection-based immunity studies (Greca and Magez, 2011). A cattle disease model would be a better choice.

Lastly, if we were to overcome the above possible limitations, there is still the problem of inability to make crosses in trypanosomes. Meiosis in *T. brucei* occur in the salivary gland of the tsetse fly (Peacock et al., 2011) which is out of reach to allow for investigation that require crossing. In models where crosses are commonly used, such as in yeasts, one of the major applications is for the purpose of generating a desired

progeny carrying the combination of genetic modifications from the parents. The relationship between gene cohorts can be studied in a single cell line by first making crosses to obtain a mutant carrying the genotype of interest. This is very useful because there are few drugs that can be used for selection of transformants in direct modification of yeasts and each direct modification require a different drug to select transformants. Generating mutants by crossing avoids the need for additional drug selection. However, the amount of genetic technology available for *T. brucei* allows us to achieve similar outcome without the need to cross. For example, we can use RNAi to knockdown multiple genes in the same cell line with a single transfection by piecing together fragments of the genes to use as the RNAi target.

1.4 Aims

In this thesis I am going to address two applications of trypanosomes as a model; presented in two parts. In Part I, I will look at the use of *T. brucei* for studying common biology. In particular, this work will study kinesins and whether *T. brucei* can serve as a more manipulable system than animal cells or yeasts. In Part II, I am going to present *T. congolense* as a new specific biology model. Both parts of the thesis will have an introduction chapter that explains specific concepts required to understand the section.

2 INTRODUCTION I

2.1 Kinesin motors

Cytoskeletal motors are important macromolecules involved in many cellular processes that are crucial for eukaryotic life. Their diversity of roles include cell motility, signalling, intracellular transport, protein synthesis and nuclear division. There are three superfamilies of cytoskeletal motors: Myosins, Dyneins and Kinesins. The function of these motors, either working alone, together or in antagonism, contributes to order within cells. Each motor requires molecular tracks on which to move or interact. Myosins use actin filaments as tracks, while dyneins and kinesins both interact with microtubules [reviewed in (Vale, 2003)]. Of the three superfamilies, kinesins are the only ones known to be ubiquitous across eukaryotes. Myosins were lost in certain members of the Excavata (e.g. *Trichomonas* and *Giardia*) and Archaeplastida (e.g. *Arabidopsis*) clades; and dyneins are also absent in some Archaeplastida and Amoebozoa (Matsuzaki et al., 2004; Richards and Cavalier-Smith, 2005; Wickstead and Gull, 2007, 2011).

2.1.1 Structure of kinesins

Kinesin motors are ATPases: they convert chemical energy through the hydrolysis of ATP into mechanical energy to perform some form of work. Depending on the type of kinesin involved, mechanical energy generated can be used to move along the microtubule tracks or to depolymerize microtubules. The general structure of a kinesin can be divided into 3 main sections – the motor domain, the neck and the tail regions.

The motor domain is a highly conserved region comprising of approximately 350 amino acid residues. It is the catalytic core of kinesins – the region that contains the ATP and microtubule binding sites. Its solved structures (Figure 2.1A) suggest that it is composed of eight centrally placed ß-pleated sheets flanked on each side by 3 α-helices. A highly conserved motif (GQTGSGKT) forms a phosphate-binding loop (P-loop) located on one side of ß-pleated sheets. The P-loop forms the region of

interaction with ATP (Endow et al., 2010; Kirchner et al., 1999; Kull et al., 1996; Sablin et al., 1996; Scarabelli and Grant, 2013). Lying adjacent to the P-loop is another loop, L9 (switch I). Conformational changes in L9 are thought to trigger ATP hydrolysis (Hirokawa and Noda, 2008).

The microtubule binding region lies on the opposite side to the P-loop. It involves a number of structural elements at the motor domain. The core region of kinesin interaction with microtubules occurs at loop 12 with remaining interactions occurring in alpha-5, loops 7/8, loop 11/α-4 loop (Marx et al., 2009; Sosa et al., 1997; Woehlke et al., 1997). This cluster is considered the switch II complex of kinesins. Similar to switch I, conformational changes in this region determines the degree of interaction with microtubules. From crystal structures, conformational change of switch II is coupled to that of switch I and is thought to influence the movement of kinesin.

Most kinesins have a flexible neck-linker after the motor domain. The neck-linker plays a role in directionality and processivity along the microtubule. It consist of 14 to 18 amino acid residues (Hariharan and Hancock, 2009) and its length is thought to influence the level of processivity of the kinesin (Düselder et al., 2012; Shastry and Hancock, 2011). As ATP is hydrolysed to propel the kinesin forward, conformational changes occur in the neck which reinforce the directionality of kinesin movement (Shastry and Hancock, 2010).

The tail is a variable coiled-coil region important for protein oligomerization. Kinesins may exist in the monomeric form, but the formation of dimers plays an activating role for some kinesins. A kinesin dimer has two heads which enables it to 'take steps' along the microtubule by alternatively attaching and detaching individual heads. The formation of dimers through the interaction of the tail has been shown to increase the velocity of movement of *Caenorhabditis elegans* Unc104 (Tomishige et al., 2002). Beyond self-interaction, some tails also associate with other proteins to interact with structures within the cell. For instance, the tail of kinesin heavy chains (KHC) in mammalian cells associates with different kinesin light chains (KLC) in order to interact with distinct vesicles or organelles (Figure 2.1B). The tail may also have other details that are specific to kinesin types. In some instances, it contains flexible "hinge regions"

which allows the tail to fold upon the motor domain to inhibit microtubule-binding (Hackney and Stock, 2000).

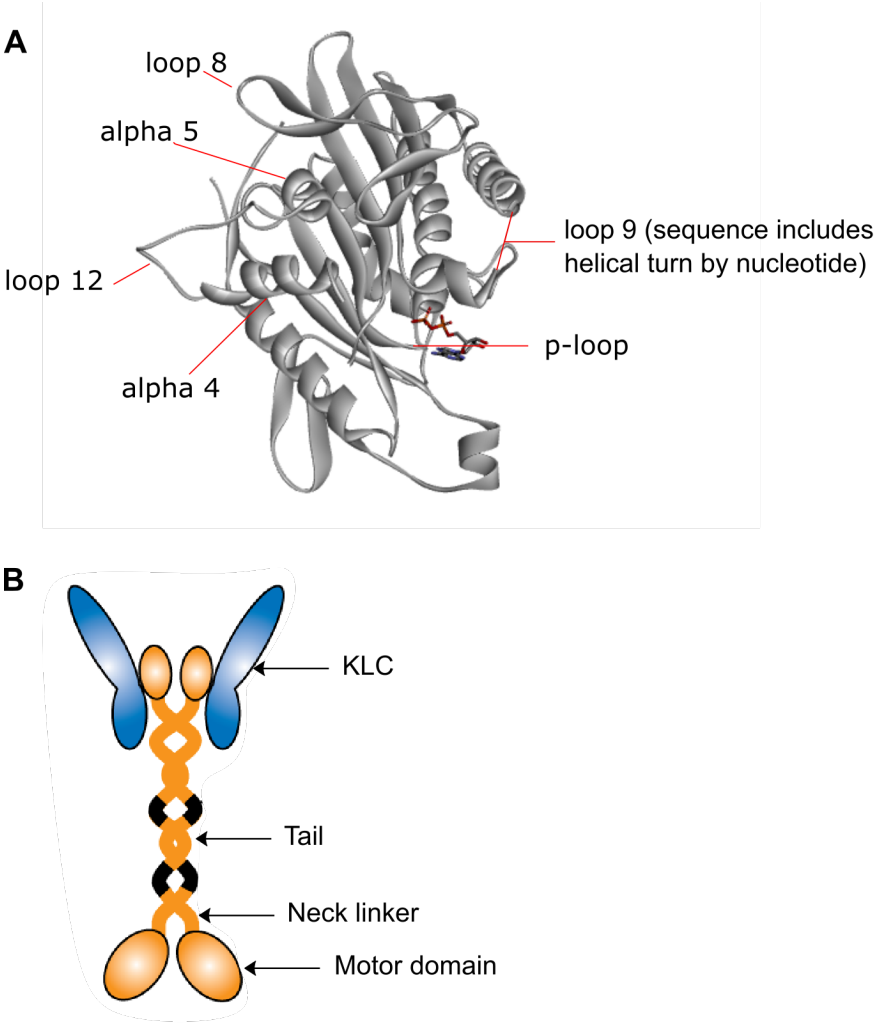


Figure 2.1: Features of the conventional kinesin, KHC. A. Motor domain structure of Kif5 (Courtesy: Hannah Belsham). **B.** Schematic of KHC showing the different structural features [Adapted from (Brownhill et al., 2009)]. KLC, Kinesin light chain.

2.1.2 Processivity along microtubules

The "hand-over-hand" model is the most accepted model for movement of most kinesin families (Asbury, 2003). Experimental evidence over the years support that kinesins with two motor domains "walk" by an alternating movement of the two heads (motor domains) as they bind and unbind microtubule. The release of a bound head is thought to depend on the binding of the other head. In this way, one kinesin head always remains bound to microtubules while the kinesin moves an approximate distance of 8 nm with each step (Kaseda et al., 2003) The distance moved correlates

with the distance between tubulin dimers of the microtubule. When the stepping distance was measured for a single Cy3-labelled kinesin head, it showed that each head steps over approximately twice the distance a kinesin moves each time and then waits a turn before taking the next step (Yildiz et al., 2004). Thus, with each step, the trailing head of the kinesin bypasses a tubulin dimer which is occupied by the leading head to occupy the next dimer where it binds and becomes the new leading head (Figure 2.2). By measuring their stepwise motion, kinesin heads were shown to alternate between two different values with each sequential step suggestive of an alternation between two conformations (Asbury, 2003). Hence the model is often referred to as the "asymmetrical" hand-over-hand model. The fast stepping of the head is believed to minimize the risk of detaching from the microtubule.

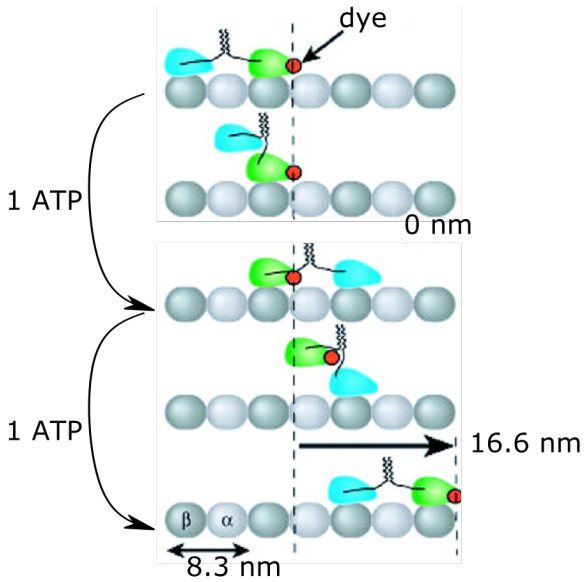


Figure 2.2: Hand over hand movement of Kinesin 1 member [Adapted from (Yildiz et al., 2004)]

Prior to ATP-binding, the neck linker exists in a disordered state referred to as the "undocked" conformation. As ATP binds to the head, there is a re-orientation within the switch I, switch II and P-loop subdomains which elicits a conformational change a distance away from the nucleotide-binding pocket. These changes create a cavity between the P-loop and the microtubule binding domain. The neck linker attains an ordered conformation and extends towards the plus end of the microtubule as it docks on the motor domain. This causes a highly conserved residue of the neck linker (Ile 325) to become inserted in the cavity created. With the hydrolysis of ATP, the neck

linker reverts to the undocked conformation, and the trailing motor domain detach from the microtubule and move forward searching for the next available tubulin binding space (Atherton et al., 2014; Wang et al., 2015). The ADP-bound head has the weakest binding affinity for microtubules (Cross, 2004). The minus-end-directed pull of the neck linker attached to the leading strand loosens the affinity of nucleotide binding and the subsequent release of ADP from the binding pocket of the leading head allowing the cycle to repeat itself. The reliance of nucleotide-binding affinity on direction of internal load ensures the cooperative movement of the two heads (Uemura and Ishiwata, 2003).

2.1.3 Architecture of kinesin motors

Architecture in this context refers to the placement of the motor domain relative to the tail. Based on architecture, there are three broad groups of kinesins: N-, C- and M-kinesins (Hirokawa et al., 1998).

The N-kinesins have their motor domain located at the N-terminus of the tail and they move towards the plus end of the microtubule. This represents the stereotypical architecture found in most kinesins. *Drosophila* KHC is an example of a kinesin with an N-terminally placed motor domain. They function as translocators, moving organelles, vesicles and mRNA within the cell.

The motor domain of C-kinesins on the other hand is located at the C-terminus and these kinesins move towards the minus end of microtubules. Only one group – the largest single kinesin family – is known to have this architecture. *Drosophila* Ncd belongs to this group. The Ncd motor domain is located at the C-terminus relative to the tail and the kinesin displays a minus end directionality. Ncd has a nuclear localisation during interphase, and is required for spindle cross-linking during mitosis or meiosis (Hatsumi and Endow, 1992).

M-kinesins have a centrally-placed motor domain. Unlike the other two groups, M-kinesins studied to date are unable to 'walk' on the microtubule, but are able to recognize both ends of a microtubule. Functionally, currently identified M-kinesins are microtubule depolymerizers (Hirokawa and Takemura, 2004; Miki et al., 2001). Mitotic

centromere-associated kinesin (MCAK) belongs to this group. They function by depolymerising microtubules at both ends. MCAK moves diffusively on the microtubule lattice without the use of ATP, which is instead used in the depolymerisation of microtubules.

2.2 Kinesin families across eukaryotes

2.2.1 Nomenclature

At the time of the discovery of various kinesin proteins there was no classification system in place, which meant that similar kinesins were given different names unrelated to each other. Several classifications have since been proposed based usually on phylogenetic analysis of kinesins sampled at the time. One of the early classifications used the kinesin motor domain sequence of 34 kinesins from 15 protists, fungi, animals and plants species, by Goodson *et al* (1994). They classified kinesins into 5 subfamilies, A to E. This was followed by a similar classification based on 56 kinesins. This classification supported the first five subfamilies and added three more to make a total of 8 subfamilies. Subfamilies were either named after one of the members (e.g. BimC subfamily) or a characteristic feature of the group (e.g. C-terminal subfamily) (Moore and Endow, 1996). Then Miki *et al.* used near complete kinesin sets from mouse and human to define 11 classes of N-kinesins, 2 of C-kinesin classes and 1 M-kinesin (Miki *et al.*, 2001).

The latest unifying system of classification places kinesins into 14 families based on the sequence of their motor domains and adopted Arabic numbering for naming families preceded by “Kinesin-”. Initial application resulted in Kinesin-1 to Kinesin-14 families. Sub-families are identified by adding an uppercase alphabet to the family name e.g. Kinesin-14A and Kinesin-14B (Lawrence et al., 2004). For a kinesin to be identified as a family or subfamily, it must contain sequences from a minimum of two kingdoms. Kinesins that do not fall into any family are termed ‘Orphan Kinesins’ (Goldstein, 1993). Individually assigned kinesin names are still maintained alongside the new nomenclature.

The first phylogenetic evidence based on complete repertoires across diverse eukaryotes (Wickstead *et al.,* 2006) united Kinesin-4 and Kinesin-10 into one family (Kinesin-4/10). The same study found no evidence for a monophyletic Kinesin-11 group, split Kinesin-12 into two families assigning them Kinesin-15 and Kinesin-16 in accordance with current nomenclature, and discovered a new Kinesin-17 family. Using a larger data set, three new Kinesin families were later discovered and named Kinesin-18, 19 and 20, putting the total number of Kinesin families at 17 (Wickstead et al., 2010a). The changes in the nomenclature described above are summarised in Figure 2.3.

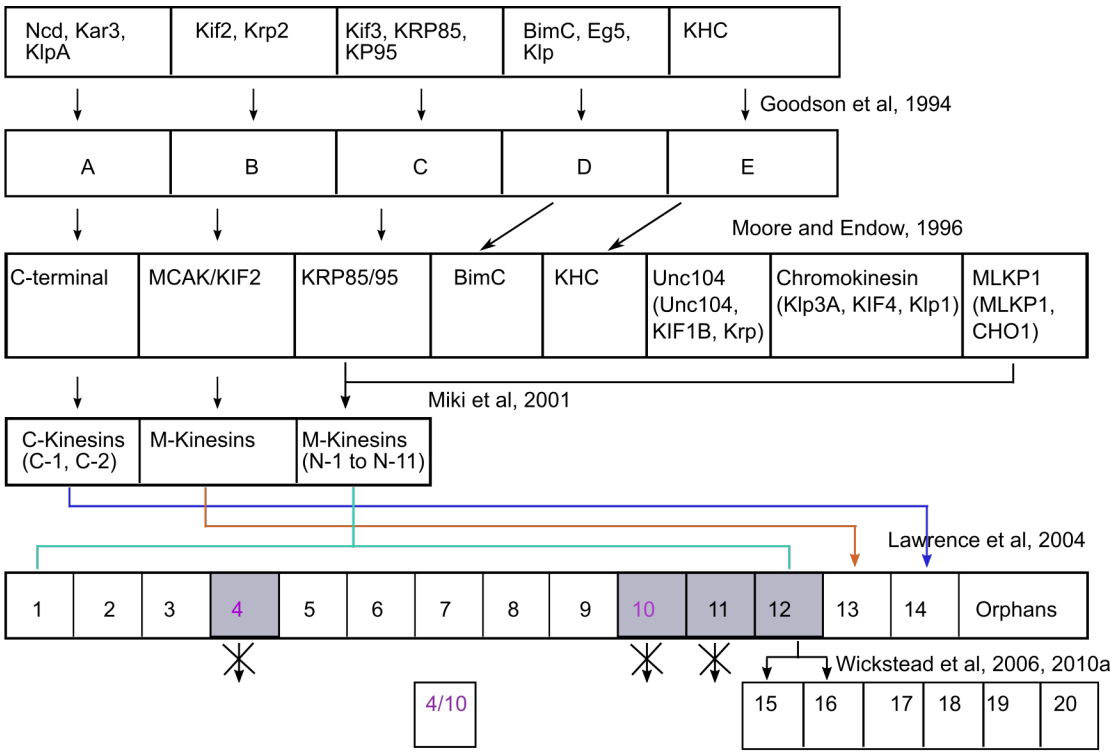


Figure 2.3: Historic classification of kinesin families and relationship between systems. Filled squares represent family numbers that have been re-grouped or are not supported in most recent phylogenies hence those numbers do not currently refer to any individual kinesin family.

2.2.2 Distribution of kinesin families

The number and distribution of Kinesin families varies greatly among eukaryotes (Figure 2.4). *S. cerevisiae* has a total of 6 kinesins, humans 45, and *Arabidopsis* 61 (Hildebrandt and Hoyt, 2000; Hirokawa et al., 2009; Julie Lee and Liu, 2004; Reddy and Day, 2001). Evidence presented from phylogenetic analysis sampled across

representative eukaryotic group members established that variation exists not only in number but also types of kinesin families (Figure 2.4). It was also clear that no single organism expresses paralogues of all 17 families. The least represented family was Kinesin-20, while Kinesin-14 was found in nearly all organisms. The number of families represented were as low as two in *Theileria annulata* (Alveolata) to as high as 16 in *Phytophthora sojae* (Stramenopiles) (Wickstead and Gull, 2006; Wickstead et al., 2010a). It seems however that the total number of kinesins encoded by an organism does not necessarily correlate to the number of families present in that organism. *Homo sapiens* and *Tetrahymena thermophila* exemplify this: 14 kinesin families are encoded in each of their genomes, however the latter encode 32 more kinesins than the former.

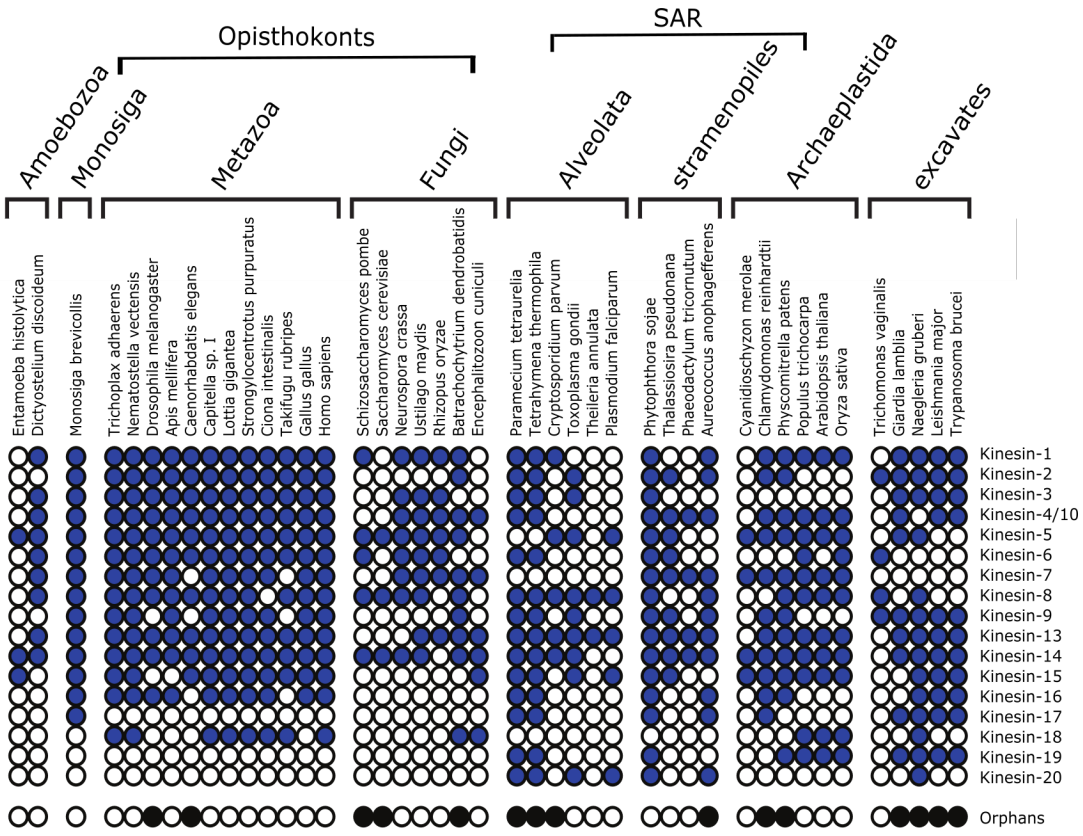


Figure 2.4: Distribution of kinesin families across eukaryotes. Presence of a paralogue of a kinesin family (blue) or an orphan (black) is indicated by filled dots. Open dots indicate absence of the kinesin. [Modified from (Wickstead et al., 2010a)]

The Last Eukaryotic Common Ancestor (LECA) is a eukaryotic evolutionary concept that refers to the common ancestor from which all eukaryotes diverged. The sheer number of kinesin subfamilies within the families, as well as ungrouped kinesins,

represent a large number of paralogues that suggest a complex evolutionary diversification of kinesins since LECA. It is possible to reconstruct the evolutionary history of kinesins by tracing back to LECA. Mapping of kinesin paralogues to alternative evolutionary trees suggest that the minimal repertoire in LECA were Kinesin-1, -2, -3, -4/10, -5, -8, -13, -14, -17 and Kinesin-9 (9A and 9B subfamilies) (Wickstead et al., 2010a). This goes to show that a large proportion of the observed diversity in the kinesin superfamily was already established in LECA.

2.3 Functional differentiation of kinesins

2.3.1 Function of the various families

Based on the nature of their interaction with microtubules, kinesin families can be classified as having two broad functions – translocation of cargo along microtubules and depolymerisation of microtubule. In a very general sense, different families perform different cellular functions. Most families are known to function mainly as translocators, except for Kinesin-13 family members which consist of depolymerisers, and Kinesin-8 which has demonstrated the ability to be both a translocator and a depolymeriser (Table 2.1).

Depolymerisers in the Kinesin-13 family play important roles in cellular process that require regulation of microtubule lengths, such as in cell division and interphase microtubule dynamics. Unlike other families, this group acts on both ends of the microtubule but is unable to move along the tracks in the same manner as translocators. Using a single-molecule microscopy assay to study MCAK, it was demonstrated that movement to the microtubule ends is achieved through lattice diffusion (Helenius et al., 2006). These data showed that transient interactions occur between MCAK and the microtubule which did not require the use of ATP, and that ATP is only required for the depolymerisation activity of either end. This form of movement could be responsible for the ability of kinesin-13 to target both ends of the microtubule.

The Kinesin-8 family contain members that are also able to depolymerise microtubules but the difference here is that they undergo the conventional plus-end directed motility characteristic of most kinesins. Hence, its depolymerisation activity occurs at the plus-end (Varga et al., 2006). *S. cerevisiae* Kinesin-8 also plays a role in cell division (Cottingham and Hoyt, 1997; Cottingham et al., 1999). The distribution of these two families raises one of the fundamental questions of functional differentiation involving redundancy. Most organisms encode at least one type of Kinesin-13, which is expected given this family's role in cell division. However, there is no Kinesin-13 member encoded in the yeast genome, which suggests that they rely on a different motor for microtubule depolymerisation, or do not use kinesins at all for this purpose. If it were the case that yeast Kinesin-8 substituted for Kinesin-13, the two families could be considered mutually exclusive. But it is noteworthy that both families also occur together in other organisms and may not represent redundancy. In addition, Kinesin-8 members do not have some of the amino acid motifs considered essential in the depolymerisation activity of the Kinesin-13 members, suggesting a possible difference in mechanism of microtubule depolymerisation (Walczak et al., 2013).

2.3.2 Differing functions within families

Functional differentiation in kinesins goes beyond the broader family groups. Members of the same family can vary both in biochemistry and cellular function. An eukaryote may carry several kinesins belonging to the same family occurring as isoforms of the same gene. Human and mouse have three isoforms of Kinesin-1 (KIF5A, B and C). While KIF5B is expressed in all cell types, KIF5A and C are expressed in neurons where they are involved in axonal transport. All three isoforms are transporters of cargo, but the kind of cargo they carry differ. Knocking out KIF5A in mice causes epilepsy and lethality due to impaired transport of $GABA_A$ receptor. Neither the presence of KIF5B and KIF5C could rescue this phenotype (Nakajima et al., 2012) emphasising that they do not carry the same cargo.

Another example of this is seen in *T. brucei* which has an expanded repertoire of the Kinesin-13 types arising from 7 different genes in the genome. Tagging of all seven kinesins showed that they localise to different areas within the cell. Out of this repertoire, only one – TbKIN13-1 – is found in the nucleus and plays a role in mitosis (Chan and Ersfeld, 2010; Wickstead et al., 2010b). TbKIN13-2, which localises to the flagellum, plays a role in flagellar length control (Chan and Ersfeld, 2010).

The differences that exist in function across the different families and even within families is not well understood. It is unclear what the molecular basis for these observed differences is, but some have suggested the motor domain to be the decision-making point. There is evidence that some kinesins are more processive than others (Scarabelli et al., 2015). The difference in processivity between Kinesin-1 and Kinesin-3 for instance have been associated with certain residue changes in the motor domain (Sablin and Fletterick, 2004; Scarabelli et al., 2015).

Directionality	Families	Known function(s)
Plus-end directed	Kinesin-1	Vesicle, organelle and mRNA transport
	Kinesin-2	Vesicle, Melanosome, Intraflagellar transport
	Kinesin-3	Vesicle and organelle transport
	Kinesin-4/10	Chromosome movement, organelle transport
	Kinesin-5	Microtubule plus end depolymeriser. Enhances microtubule polymerisation
	Kinesin-6	Spindle assembly, cytokinesis
	Kinesin-7	Kinetochore-microtubule attachment
	Kinesin-8	Microtubule plus end depolymeriser, mitochondrion transport, chromosome positioning, microtubule polymeriser
	Kinesin-9	9A – Central pair 9B – flagellar function
Minus-end directed	Kinesin-14	Microtubule minus end depolymeriser, spindle pole organisation
No directional movement on lattice	Kinesin-13	Microtubule plus and minus end depolymeriser, Chromosome segregation

Table 2.1: Known functions for kinesin families.

2.4 *T. brucei* as a choice model organism for studying kinesins

T. brucei is potentially a good model to study kinesins because it has a complex cytoskeleton dominated by tubulin and which is involved in multiple functions: chromosome and mitochondrial segregation, Golgi organisation, cell shape maintenance and cell motility (mediated by the parasite's single flagellum). Additionally, *T. brucei* has a large number of kinesin families, represented by 47 kinesin-like proteins (Berriman et al., 2005): one each of Kinesin-1, -3, -17, two of Kinesin-2, -4/10, -9, -14, -15, -19, seven Kinesin-13, three kinesin-16, 16 X paralogue group members, and 6 ungrouped kinesins. No member of kinesin -4, -5, -6, -7 and -8 families are present (Figure 2.4). All the absent kinesins are involved in mitosis in other organisms (Wickstead and Gull, 2006; Wickstead et al., 2010a).

Lastly, genetic tractability of *T. brucei* makes it a good model for functional studies (see Section 1.3.2). Endogenous locus tagging and targeted protein depletion by RNAi are two well-established techniques in *T. brucei* that can be applied to study kinesin function, especially if that is associated with a growth defect in trypanosomes.

T. brucei Kinesin-1, -2, -3, -13 and -14 were identified as the most favourable candidates to work with: Kinesin-1, -3, and -14 because they are present as single copy genes; inclusion of Kinesin-2 because its members have a very distinct role in intraflagellar transport in all organisms tested so far (Prevo et al., 2017); and Kinesin-13 because it is the family we probably know more about in *T. brucei* (see Section 2.4.4).

2.4.1 Kinesin-1

At the time of writing this thesis, there was no published data on the role of Kinesin-1 (Tb927.1.1350) or -3 (Tb927.11.2490) in *T. brucei.* RIT-seq data suggest depletion of either kinesin affects parasite development (Alsford et al., 2011). Kinesin-1 family members have a wide range of functions in other organisms including transport of mitochondria, lysosomes and endoplasmic reticulum. Disruption of KIF5B in mice cells affected lysosome dispersion and caused an abnormal aggregation of mitochondria at the perinuclear position (Tanaka et al., 1998). In a human cell line, KIF5B was also

shown to participate in the spatial organization of the endoplasmic reticulum and Golgi apparatus as well as trafficking outside the ER (Gupta et al., 2008). It is involved in the transport of mRNA (Brendza et al., 2000), intermediate filaments (Prahlad et al., 1998) and also plays a role in the secretory and endocytic pathways. In metazoans, Kinesin-1 members avidly partake in axonal transport, and its depletion results in neuronal defect and lethality in *Drosophila* (Saxton et al., 1991) and mice (Nakajima et al., 2012).

2.4.2 Kinesin-3

Kinesin-3 members function as vesicle and organelle transporters. It is a well-documented transporter of pre-synaptic vesicles in neurons [reviewed in (Hirokawa and Noda, 2008)]. Disruption of neuron-specific KIF1A in mice decreased pre-synaptic vesicles and resulted in the death of mice (Yonekawa et al., 1998). In the fungus, *Ustilago maydis,* both Kinesin-1 and -3 cooperate in hyphal growth via vesicle transport. Knockout of either or both kinesins resulted in short hyphal growth (Schuchardt et al., 2005). Kinesin-3 was shown to co-localise with mitochondria *in vivo* and also to transport mitochondria along microtubules (Nangaku et al., 1994). Kinesin-3 transports early endosomes in fungi (Seidel et al., 2013; Steinberg, 2015; Wedlich-Söldner et al., 2002). *Dictyostelium* Unc104/KIF1A knockout lines had an estimated 62% reduction in organelle movement compared to control, indicating the motor to be involved in organelle transport in this organism (Pollock et al., 1999).

2.4.3 Kinesin-2

Kinesin-2 plays a key role in moving particles in cilia and flagella, in a process called intraflagellar transport (IFT). *C. elegans* uses two different Kinesin-2 motors: heterotrimeric kinesin-II and homodimeric OSM-3 to build a cilium. Both kinesins work redundantly to build the middle segment of cilia, and OSM-3 individually builds the distal segment (Snow et al., 2004). The genome of *T. brucei* encodes two Kinesin-2 proteins TbKIN2A (Tb11.01.5490) and TbKIN2B (Tb927.5.2090). Work from our lab shows that there is redundancy in the roles played by both kinesins (Stretton, 2016).

Knockdown of TbKIN2A is associated with some phenotypic defect, but it is only when both genes are silenced that a strong growth defect is observed.

2.4.4 Kinesin-13

T. brucei has an expanded repertoire of Kinesin-13, with 7 sub-types (TbKIN13-1 to TbKIN13-7) distributed in three cellular locations. TbKIN13-1 (Tb927.9.3650) is found in the nucleus, TbKIN13-2 (Tb927.11.4660) is enriched at the distal tip of the flagellum, and the remaining five are in the cell body. Two different studies have confirmed TbKIN13-1 as the only subtype associated with mitosis (Chan et al., 2010; Wickstead et al., 2010b), specifically with the mitotic spindle. Knockdown by RNAi caused mitotic spindle defects and chromosome segregation impairment, resulting in an accumulation of anuclear cells, and confirming the mitotic role of TbKIN13-1.

2.4.5 Kinesin-14

Kinesin-14 family members occur as dimers, with a C-terminally placed motor domain. Members of the family can bind to, crosslink or slide microtubules. They are involved in spindle assembly [reviewed in (Braun et al., 2016; She and Yang, 2017)]. Interestingly, work on *T. brucei* Kinesin-14, TbKIFC1 (Tb927.10.14890), suggests a novel role of transporting acidocalcisomes in a retrograde manner in procyclic cells (Dutoya et al., 2001).

2.5 Aim I

In Part I, I aim to investigate whether *T. brucei* can be used as a more manipulatable model than current models such as mammalian cells, to investigate kinesin biology across eukaryotes. As explained in Section 2.4, *T. brucei* has particular characteristics that favour its use in studying kinesin biology. From the point of view of understanding kinesins in general, we must first characterise the families in *T. brucei* – whether their function is similar or different from studies elsewhere will be useful to our understanding of eukaryotic kinesins. Furthermore, the application of genetic technology such as RNAi to investigate kinesins in *T. brucei* requires that there is a

detectable growth defect associated with protein depletion, which is possible when the gene is essential. It is important then to test the essentiality of *T. brucei* kinesins and their associated phenotypes. To this end, the specific objectives are:

1. To endogenously tag and determine the cellular localisation of *T. brucei* Kinesin-1, -2, -3 and -13 (TbKIN13-1) and -14 (TbKIFC1)
2. Perform RNAi against the kinesins to determine whether they have an associated fitness cost.
3. To assess whether the characterised kinesins will be usable to screen general kinesin function.

3 MATERIALS AND METHODS

3.1 Cell lines and culture

The single marker bloodstream-form line SmOxB427 (Poon et al., 2012), derived from *Trypanosoma brucei* Lister 427 strain, was the parental cell line for all *T. brucei* transfections. SmOxB427 expresses T7 RNA polymerase and the tetracycline repressor, and it was cultured in HMI-9 at 37 °C, 5 % CO_2 supplemented with 15 % foetal bovine serum (Hirumi and Hirumi, 1989) and 1 μg/ml puromycin (the latter to maintain the SmOx plasmid).

The bloodstream-form *T. congolense* IL3000 strain, derived from infected mouse blood in Theo Baltz lab (kindly donated by Mike Barrett), was cultured in TcBSF1 without red blood cells (Coustou et al., 2010), supplemented with 15 % goat serum (GIBCO) at 34 °C.

3.2 Genome-wide RNAi targeting fragments

T. congolense IL3000 genomic DNA (20 μg) was sheared by sonication in SSC buffer (150 mM NaCl, 15 mM sodium citrate, pH 7.0) on ice. This was followed by isopropanol precipitation before size-selecting sheared fragments between 400-500 bp in length through agarose gel excision. DNA fragments were purified from agarose, end-repaired using NEBNext End Repair Module (NEB#: E6050S) at 20 °C for 2 h before attaching linkers Sfi_vectorette and Sbf_linker (see Table 3.1). Linkers were generated by annealing custom primers at 50 μM in 100 mM NaCl, 10 mM Tris HCl, pH 7.2 solution, starting at 85 °C for 2 min and then decreasing temperature over 20 min to 55 °C. Linkers were incubated with end-repaired fragments at 16 °C for 2 h with 20x molar excess of linkers to fragments and a 4:1 molar ratio between Sfi_vectorette and Sbf_linker, respectively. To remove excess linkers, linked fragments were size-selected by gel electrophoresis and amount of amplifiable, linked fragments was estimated by quantitative PCR using primers 28s and 2fr (see Table 3.1) in the presence of 0.5× EvaGreen (Biotium). Amplified fragments were digested with SfiI and

SbfI before cloning into p3T7-TcoV construct (see Supplementary File S1) to generate a genome-wide RNAi plasmid library. A fraction of the plasmid-insert ligation mix was transformed into XL1-Blue strain of competent *E. coli* for routine estimation of transformation efficiency. For high efficiency transformation, the remaining ligation mix was transformed into NEB 10-beta electrocompetent *E. coli.*

3.3 Plasmids for transgenic modification

3.3.1 *T. brucei* endogenous locus tagging and RNAi plasmid construction

A plasmid construct used for tagging the N-terminal end of *T. brucei* KKT2 (kinetoplastid kinetochore protein 2) with mStrawberry (mSt), pEnNmSt-KKT2-N (D'Archivio and Wickstead, 2017) was modified to generate pEnNmSt-0 (Figure 3.1, Supplementary File S1) which was used for mSt-tagging of kinesins using the strategy in Figure 3.2. The mSt CDS was first amplified from pEnNmSt-KKT2-N. The plasmid (pEnNmSt-KKT2-N) was then digested with SpeI and BamHI to remove mSt CDS along with the *KKT2* fragments before ligating the amplified mSt CDS into the vector to generate pEnNmSt-0 (Figure 3.2).

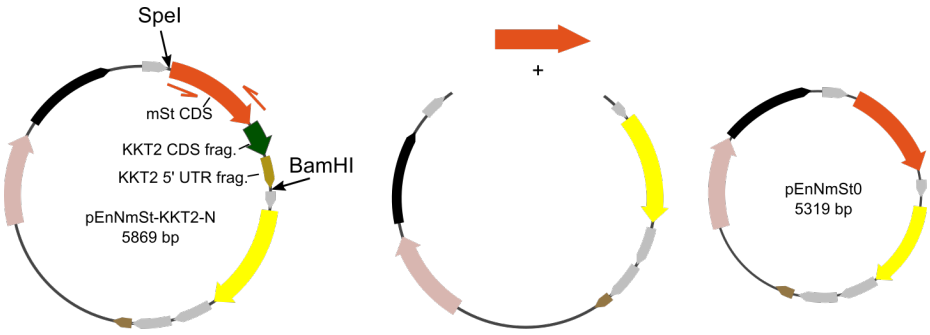


Figure 3.1: Schematic for creating pEnNmSt0

Kinesin ORF sequences Tb927.1.1350 (Kinesin-1), Tb927.11.13920 (Kinesin-2), Tb927.11.2490 (Kinesin-3), Tb927.9.3650 (Kinesin-13-1), Tb927.10.14890 (Kinesin-14) were retrieved from www.tritrypdb.org. For endogenous locus tagging of each kinesin, ~250 bp upstream (5' UTR region) and downstream (N-terminal end of protein) of the ATG were amplified from *T. brucei* Lister 427 genomic DNA as targeting sequences. Primers were designed to include XbaI, BamHI and NotI restriction sites at

the ends of the amplicons such that ligating them together created a unique NotI restriction site between them for linearization, and different sites (XbaI and BamHI) at opposite ends to ligate with the right ends of the plasmid backbone (Table 3.1). Amplicons were digested and purified prior to ligation. Plasmid pEnNmSt-0 was also digested with XbaI and BamHI. Target sequences were inserted into linearized plasmid immediately downstream of the mSt CDS to create plasmids for tagging the N-terminal end of the kinesins (see Figure 3.2). Plasmids were transformed into XL-1 blue strain of competent *E. coli* cells before plating on LB agar (LB broth, 1.5 % agar) supplemented with 100 µg/ml ampicillin. After overnight incubation at 37 °C, transformed colonies were selected and grown in LB broth for plasmid isolation using standard kits (Qiagen miniprep). Correct insertion of targeting sequence was confirmed by test digestion and DNA sequencing (data not shown).

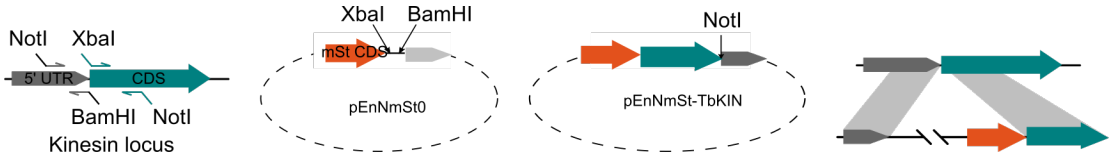


Figure 3.2: Schematic of endogenous locus tagging

For inducible RNAi, ~600 bp fragment of the motor domain sequence of each kinesin was amplified from the *T. brucei* Lister 427 genome using primers specified in Table 3.1. Amplicons of *TbKIN1*, *3* and *13-1* were cloned into the p2T7-177 plasmid which integrates into the 177 bp repeat region of *T. brucei* minichromosomes (Wickstead et al., 2002). The plasmid has opposing T7 promoters to create head-to-head RNAs and contain a bleomycin drug selection marker. Cloning was performed as above.

A parental plasmid for hairpin RNAi constructs, pBSG0, was made by modifying an inducible overexpression plasmid, pV4d-Dex-mStKIN17_427 (Sarah Whipple and Catarina Gadelha, unpublished results). The plasmid was digested with HindIII and SbfI to remove mSt-tagged kinesin before introducing a multi-cloning site (Figure 3.3A). To make target RNA hairpins, the same region of the motor domain was amplified using primers (Table 3.1) that generated a longer (~350 bp) and shorter fragment (~300 bp). The fragments were ligated such that they were complementary to each other separated by ~50 bp of a non-complimentary region on the longer fragment (Figure

3.3B). Hairpin fragments of *TbKIN2A*, 3*, 13-1* and *TbKIFC1* were cloned into BamHI and XbaI sites of pBSG0.

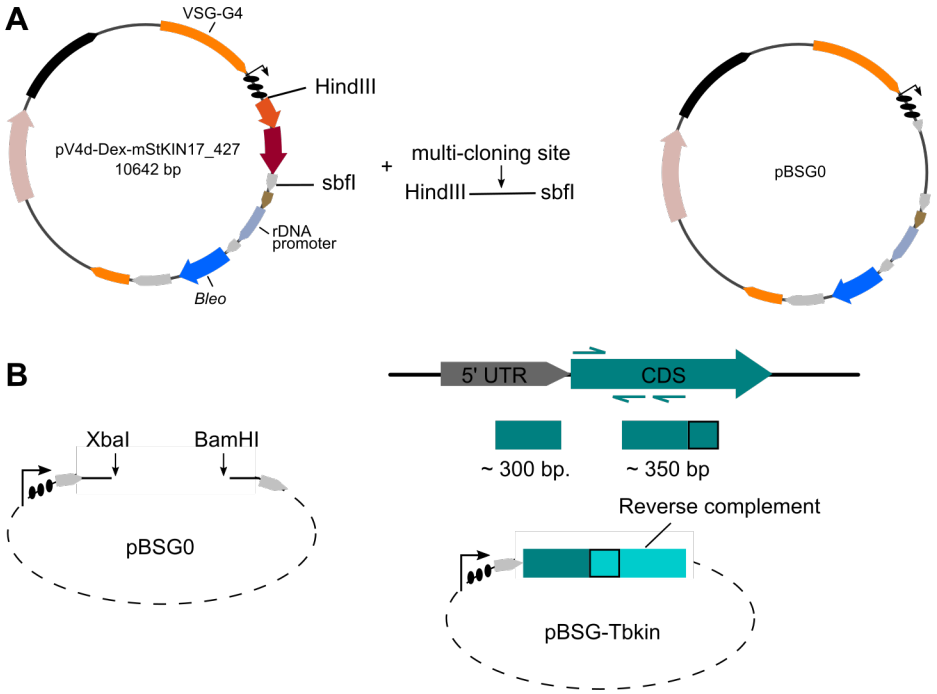


Figure 3.3: Schematic of making hairpin RNAi constructs. A. Shows generation of pBSG0. Multi-cloning site (HindII, EcoRV, XbaI, BamHI, SbfI) **B.** Generation of hairpin RNAi fragments and RNAi construct (pBSG-Tbkin). Short RNAi fragments were amplified using the same forward primer with a reverse primer located internal of the long RNAi fragment. Short RNAi fragments ligated to reverse complement of long fragments generate complementary regions separated by a noncomplementary region (black square) which forms the hairpin loop.

Target/name	Primer sequence
Genome-wide primers	
Sfi_vectorette	F: CTCTCCCTTCTCGAATCGTAACCGTTCGGCCAGTGAGGCCT R: [Phos]AGGCCTCACTGGCCTGCAGCCAAGGTATGGAGAAGGGAGAG
Sbf_linker	Sbf-link1-F: TTACACGACCACGTGATGACTCCCTGCAGG Sbf-link1-R: CCTGCAGGGAGTCATCACGTGGTCGT
28s	CTTCTCGAATCGTAACCGTTC
2fr	ACGACCACGTGATGACTCC
Endogenous locus tagging	
Tb927.1.1350	TAGCGGCCGCCACAAGTCGTATACCTTCC
5’ UTR	TATGGATCCGGGGTGGTCAAATGATT
Tb927.1.1350	TATCTAGAATGCAAACATCGCGACC
N- terminal CDS	TAGCGGCCGCATTGATCAACTGCAGTGTCT
Tb927.11.13920	TAGCGGCCGCGTGCATGGTTGGGTAA
5’ UTR	TATGGATCCGTTAAACCCACTCCTGAAT
Tb927.11.13920	TATCTAGAATGTCTAAGAAACACCAAGG
N- terminal CDS	TAGCGGCCGCACCTCCTGCAGAAAGA
Tb927.11.2490	TAGCGGCCGCACTTTGTGATTTCAGTTCGT
5’ UTR	TATGGATCCTCCCAATACGAATATATGCG

Tb927.11.2490 N-terminal CDS	TATCTAGAATGTCGAAGCCACGCGAT TAGCGGCCGCCTGTCGTTGCTGTCAAAAG
Tb927.9.3650 5’ UTR	TAGCGGCCGCCTTGTTGGGTGTGAACA TATGGATCCTCCTTTCCAACTACCTATAGA
Tb927.9.3650 N- terminal CDS	TATCTAGAATGGCGAAGTGGGAATTAA TAGCGGCCGCAAGCGACTCTTCCTCC
Tb927.10.14890 5’ UTR	TAGCGGCCGCTCCCTATCTTATCCTACTTTCT TATGGATCCCTTTCTGCAGCACAATTAAG
Tb927.10.14890 N- terminal CDS	TATCTAGAATGTCTGTGGAACAACC TAGCGGCCGCGAAAGCATGACAGCAAAC
Target	**Primer sequence**
Tb927.1.1350	TATGGATCCGTGCCGTCAATCAAATCTT ATTAAGCTTTTGTGAAGGACTACAGCA
Tb927.11.2490	TATGGATCCCGTGATGTAGTGGTTGAC ATTAAGCTTACCTAAAACCGTCAGAGAT
Tb927.9.3650	TATGGATCCCTGGCATCTATGCTCTTG ATTAAGCTTGAGACAGCACCAATCATC
Hairpin RNAi	
Tb927.11.13920	GTAAAATTCACAAGCTTGATATCGTTACGTCAGCAACAAGAACG TCCTGTTCAAAGTCAAACTGT
Tb927.11.13920	CATGCCTGCAGGATCTACGTCAGCAACAAGAACG ACAGTTTGACTTTGAACAGGAAGTGCGATCGACTGCTATT
Tb927.11.2490	GTAAAATTCACAAGCTTGATATCGTTGGTTGCGAAGGAAGTG GGTCATTCATATTTGTTGCTGC
Tb927.11.2490	CATGCCTGCAGGATCTGGTTGCGAAGGAAGTG GCAGCAACAAATATGAATGACCATACTCTCCTCAGAATCCACTTG
Tb927.9.3650	GTAAAATTCACAAGCTTGATATCGTTCCCACGAAGCAGATGT AACTTACCGAAGAGCTCCT
Tb927.9.3650	CATGCCTCCAGGATCTCCCACGAAGCAGATGT AGGAGCTCTTCGGTAAGTTCTCAAGAATTGCGTGCG
Tb927.10.14890	GTAAAATTCACAAGCTTGATATCGTTGCGAGATGATCCGAGA GTCGAGACGCTCACTTC
Tb927.10.14890	CATGCCTGCAGGATCTGCGAGATGATCCGAGA GAAGTGAGCGTCTCGACTTGTTCGGACAGTGATATACA

Table 3.1: Primers

3.4 Genotypic analysis of genetically-modified strains

For diagnostic PCR to analyse integration of DNA construct, ~10^5 trypanosome cells were harvested by centrifugation at 1000 *xg* for 2 min and washed in phosphate-buffered saline (PBS) (137 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) supplemented with 50 mM glucose. Cells were lysed in 50 mM NaOH, 2 mM EDTA, heated at 95 ^oC for 10 min and neutralised by addition of Tris-HCl pH 6.8 to 75 mM final concentration.

3.5 Preparation of *T. congolense* DRiF-Seq samples

For Direct-RNAi Fragment Sequencing of bloodstream-form *T. congolense* following RNAi library induction, ~50 million cells were harvested by centrifugation (1200 x*g* for 10 minutes), washed twice in 2 ml TNEP (2 mM Tris-HCl, 15 mM $NaPO_3$, 10 mM EDTA pH 8), and cell pellets frozen at -80 ^{o}C until further processing. Duplicate samples were taken at each sampling point but only one set was processed further. Genomic DNA was extracted from frozen pellets using the Qiagen DNeasy blood & tissue kit (Cat No. 69504). RNAi fragments were amplified from genomic DNA by qPCR both to quantify the total number of RNAi fragments represented in the sample and ensure the reaction did not enter plateau phase. qPCR reactions used primers 2fw and 137 and contained 0.5x Evagreen dye (Biotium). A second round of qPCR was performed to add indices to the amplified RNAi fragments for next generation sequencing.

Primer	Sequence
2fw	CCGGCCGCTCTAGAACTA
137	CCTCGAGGTCGACGGTAT

3.6 Sequencing of DRiF-Seq samples

Sequencing was performed by University of Nottingham Deep Seq facility. Fragment-length distributions were analysed using a TapeStation 4200 and High Sensitivity D1000 ScreenTape Assay (Agilent). Samples were pooled in equimolar amounts and final pool quantification performed using the KAPA Library Quantification Kit for Illumina (Roche). The pool was sequenced using a NextSeq 500 System (Illumina) and high-output 150 cycle kit v2.5 (Illumina), providing 45-56 million pairs of 75-bp paired-end reads passing filter per sample.

3.7 Processing of DRiF-Seq data

Processing of DRiF-Seq data was performed by Dr Bill Wickstead (University of Nottingham) and is presented here for information.

The first 2 bases in read 2 that are common to all pairs were removed and reads were trimmed for quality using Trim Galore v0.4.4 (www.bioinformatics.babraham.ac.uk; -q 28 --paired --retain_unpaired --stringency 20 -r1 30 -r2 30 --length 25) and aligned

the TriTrypDB *T. congolense* IL3000_2019 long-read assemblies (v46) using bowtie2 v2.3.4 (Langmead and Salzberg, 2012) allowing only concordant mappings of 150-750 bp fragments (--no-mixed --no-discordant -I 150 -X 750). Mappings to forward or reverse strand with respect to read 1 (FLAG=99 or 163, respectively) were extracted from SAM files. All mapped pairs starting and ending at the same genome position and in the same orientation were considered to represent the same RNAi fragment in each DRiF-Seq library.

To reduce variation due to sampling, only fragments represented by ≥128 pairs of reads across all samples for a given library were considered (438775 and 429773 fragments in library 1 and 2, respectively). Since DRiF-Seq libraries are collections of random fragments, it is not possible to use true biological replication to estimate fragment variance (<1% of fragments are common between the 2 independent libraries, and these are mostly from repeated genomic regions). Additionally, technical replication of sequencing (or resampling methods) capture only variance introduced due to sampling, not all the other sources of variation in the experiments. For these reasons, approaches such as Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK; (Li et al., 2014)) are not good ways to analyse DRiF-Seq (or RIT-Seq) data. However, since a DRiF-Seq experiment contains a large number of fragments which will have no loss-of-fitness (e.g. non-genic fragments or those targetting non-essential genes), the distribution of the data is an effective means to estimate overall experimental variance (*a priori* noise estimation). Only two assumptions were made in making these estimates: i) that the convolution of the variation from all sources will be near Gaussian, and ii) that variation will be dependent on fragment abundance. Both of these assumptions are very well supported by the distribution of the data.

All statistical treatment was performed in R using custom scripts. To avoid infinities, counts were increased by 0.5 and mean $\log_2$ abundance (A) and $\log_2$ fold-change (M) calculated for each time-point against non-induced control. To estimate dispersion at different fragment abundance, data were separated into 16 abundance quantiles and the distribution of M values in each fitted to an unconstrained Gaussian by non-linear

least squares method. Fitted distribution means and log-transformed standard deviations were then fitted to second-order polynomials to model dispersion against A. The process is illustrated in Figure 3.4.

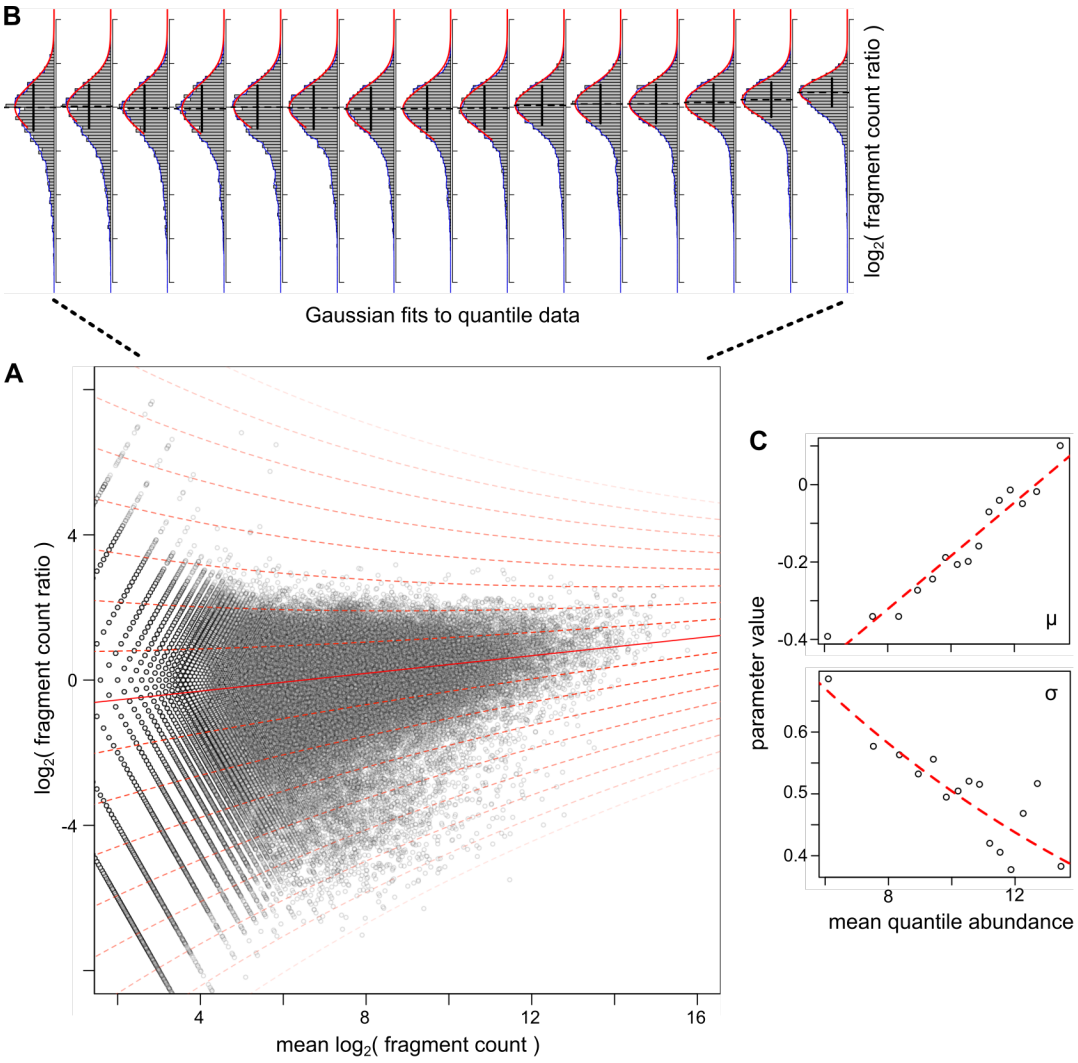


Figure 3.4: Statistical treatment of DRiF-Seq data. A. Shows a plot of mean $\log_2$ fragment abundance against $\log_2$ fold-change of any induction time point against the non-induced **B.** Plot shows estimation of dispersion by fitting the distribution of $\log_2$ fold-change values to a Gaussian using the non-linear least square method. **C.** Plot shows a model of dispersion against fragment abundance using fitted distribution means (μ) and log-transformed standard deviations (σ).

Z values and corresponding p-values were calculated from individual M values given the modelled mean and standard deviations for the specified A. P-values were adjusted for multiple comparisons by controlling for false discovery rate using the method of Benjamini & Hochberg (Benjamini and Hochberg, 1995). Statistical treatment at the gene level was as above, except counts with fragment counts combined. Fragments were mapped to gene features based on the TriTrypDB v46

T. congolense IL3000_2019 annotation (CDS only). An overlap of ≥150 bp was necessary to associate a fragment with a feature. Fragments with mixed association (192 and 191 for library 1 and 2, respectively) were discarded.

3.8 Quantitative enrichment analysis

Quantitative enrichment analysis (QEA) has advantages over simple gene set enrichment analysis as it accounts for the effect size and effect direction associated with a set of genes rather than only set membership at an arbitrary threshold. QEA was performed by calculating the mean value of a metric (e.g. mean Z or mean ΔZ) for genes associated with a set and comparing to the value associated with all genes (the null set). To test for significance of enrichment, 10^6 Monte Carlo permutations of the data were made by randomly sampling a set of equal size to the test set from the null set and measuring the same test metric. P-values were adjusted for multiple comparisons using the method of Benjamini & Hochberg (Benjamini and Hochberg, 1995).

Since gene annotation in *T. congolense* lags behind *T. brucei*, gene ontology information and also gene ID appearance in the literature was taken from the *T. brucei* 927 genome (TriTrypDB v46). Data were mapped onto *T. congolense* based on genes that were reciprocal best hits across the predicted proteomes at e-value ≤10^{-5}. This maps both strict 1-to-1 orthologues and also only the best hit for any 1-to-n orthologues between the species (6630 of 10052 predicted *T. congolense* proteins).

3.9 Genetic modification of trypanosomes

3.9.1 *T. brucei*

5-10 µg DNA was used to transfect 2.5-3x10^7 *T. brucei* cells. These were harvested, washed and resuspended in 90 µl Tb-BSF buffer (90 mM sodium phosphate, 5 mM potassium chloride, 0.15 mM calcium chloride, 50 mM HEPES, pH 7.3) (Schumann Burkard et al., 2011). Plasmids were digested with NotI restriction enzyme, purified by anion exchange (Qiagen plasmid purification kit), air-dried, resuspended in 30 µl Tb-

BSF buffer, and added to the cells in a 2 mm electroporation cuvette. This was allowed to incubate for 5 mins before electroporation using Amaxa Nucleofector (Lonza) program Z-001. Cells were quickly transferred to a culture flask containing 25 ml pre-warmed medium and incubated for 4 to 6 h to allow recovery. After recovery, transformants were selected for by adding appropriate drugs (2 µg/ml G418 or Phleomycin) and plated in 96-well plates. After 6 days, selected clones were verified for the presence of tagged protein by Western blot. For RNAi transfections, Western blot was performed after 24 h induction using 1 µg/ml tetracycline.

3.9.2 *T. congolense*

Transfection of *T. congolense* followed a protocol similar to transfection of *T. brucei*. Unless otherwise stated, 10 µg of linearized plasmid, purified as above, was transfected into 2-3x10^7 cells resuspended in sterile Tb-BSF buffer with the use of Amaxa Nucleofector (Lonza), program Z-001. After cells had recovered for 8-16 h, transformants were selected using appropriate drug (0.5, 0.2 and 0.5 µg/ml of Puromycin, G418 and Hygromycin B respectively) and plated across 96-wells. High dilutions of cells were supplemented with 20 % conditioned medium.

Transfection efficiency was determined by estimating the total number of transformants per transfection. By assuming a Poisson distribution, the probability of any number of cells being in any well is given by equation (1)

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad (1)$$

where λ is the average number of cells/clones in the volume put into each well. Hence, the probability of a well being empty, P(0) is:

$$P(0) = e^{-\lambda} \qquad (2)$$

Rearranging equation (2) allows us to estimate the average number of cells per well (λ) from the overall number of positive wells (N_{+ve}) to obtain equation (3)

$$\lambda = -In\ (1 - \frac{N_{+ve}}{N_{well}}) \qquad (3)$$

where N_{well} is the total number of wells on the plate. Therefore, an estimate of the overall number of cells on a plate (i.e. the total number of clones per transfection or transfection efficiency) is given by equation (4):

$$\lambda N_{well} \quad (4)$$

The 90% confidence interval (range) was calculated for the total number of transformants estimated.

Plasmids, parental cell lines and selective drug concentrations are shown in Table 3.2.

Plasmid	Parental background	Selective drug
pTcoSM	IL3000	Puromycin
pTcoLUC1	TcoSM or IL3000	G418
pTcoLUC2	TcoSM or IL3000	G418
p2T7-TV-GFP	TcoSM	G418
p2T7-TcoV-GFP	TcoSM	G418
p2T7-TcoV-GFP-T	TcoSM	G418
pGad11-TcoDEAH	TcoSM	Hygromycin B
p3T7-TcoV-GFP	TcoSM	G418
p3T7-TcoV-TUBA	TcoSM	G418
p3T7-TcoV-CLH	TcoSM	G418
p3T7-TcoV-VSG	TcoSM	G418
p3T7-TcoV-KKIP1	TcoSM	G418
pTV-TTS	TcoSM	Hygromycin B
pTco-TTS	TcoSM	Hygromycin B
pTcoV-Lib	TcoTTS	G418

Table 3.2. Transfected plasmids, parental lines and selection drugs used in generating transgenic lines.

3.10 Subcellular localisation of proteins

To localise proteins at the subcellular level, cells were harvested at mid-log phase by centrifugation, washed twice in PBS and resuspended in the same buffer. Bloodstream-form *T. congolense* adheres to glass slides and therefore plain glass slides were used for microscopy. Cells at ~ $2x10^7$ cells/ml were allowed to adhere for ~5 mins. Bloodstream-form *T. brucei* was attach to glutaraldehyde (GA)-derivatized slides. Approximately 10^7 cells/ml were placed on the derivatised slides and allowed to adhere for 2 mins. Both *T. brucei* and *T. congolense* cells were fixed with 2 % formaldehyde (FA) for ~ 5 mins. Slides were washed in PBS to remove FA and any cells that had not adhered before permeabilization in methanol at -20 ^oC for 10 mins. This

was followed by rehydration in PBS before mounting in a solution of DNA stain and photostabilizing agent (1 % wt/vol 1,4-diazabicyclo[2.2.2]octane, 90 % vol/vol glycerol, 50 mM sodium phosphate, pH 8.0, and 200 ng/ml 4',6-diamidino-2-phenylindole). Images were captured using a Retiga R1 CCD camera (Qimaging). All images were taken with no prior illumination. Images for level comparison of fluorescence signal intensity were usually taken with the same exposure settings. Image acquisition was controlled by µManager open source software (Edelstein et al., 2014). Images were processed using Fiji image processing package (Schindelin et al., 2012).

3.11 Measurement of bioluminescence

3.11.1 *In vitro* measurement

In vitro luciferase expression in cells was monitored using a luciferase assay kit (Roche Molecular Biochemicals) following manufacturer instructions. A volume of 50 µl cell lysis buffer was added to an equal volume of parasite suspensions containing $5x10^4$ and $5x10^5$ cells for each line in 96-well microplates. Lysis was allowed to occur for about 10 mins before addition of luciferin-containing substrate. The plate was immediately transferred to a luminometer to measure light emission. Luminescence was expressed in arbitrary light units.

3.11.2 *In vivo* assessment

3 female BALB/c mice were infected with 10^6 parasites (IL3000-LUC2 line) per mouse using the intraperitoneal route. First peak of parasitemia was observed 4 days after infection. Cells were taken from the second wave of parasitemia. About 10^4 parasites were transferred to 13 new recipient mice using intravenous injections. Images of the bioluminescence given off from infected mice were taken at day 6 of infection after inoculating them with luciferin. Emitted photons were captured using a charge couple device (CCD) camera (IVIS Imaging System Lumina, Caliper, Villepinte, France) set in high resolution (medium binning) mode.

3.12 Flow cytometry

Procedures for flow cytometry for bloodstream-form *T. congolense* had not been established. Therefore, some fixative methods were tested. Three fixatives were investigated to find a good method to preserve membrane integrity and kept GFP stable for long enough to measure levels (~1-2hours). Cells were fixed in 1 % FA with either 4 % BSA, 0.3 M glucose in Tris-HCl or 0.2 % GA. Propidium iodide (PI) penetrates cells with compromised membrane at the time of fixation. As a means of measuring membrane integrity, PI was added to fixatives at 10 µg/ml. Proportion of PI-positive cells in each sample was estimated by fluorescence microscopy. After 1.2 h of fixation, <10 % of cells fixed with (1 % FA + 0.2 % GA) were PI-positive, while all other fixatives had >90 % of cells with compromised membrane integrity (Figure 3.5A). At 2.4 h, PI-positive cells had increased (~18 %) suggesting PI was still getting into cells but at a much lower rate, which is compatible to obtain accurate fluorescence level in cells. Flow cytometry data show that only 6 % of cells fixed with 1 % FA have the forward-to-side scatter ratio that would be expected for cells with intact membrane (gated region) < 30 minutes of sample fixation. While (1 % FA + 0.2 % GA) had 57 % of cells with intact membrane (Figure 3.5B) within 30 minutes after fixation. All cells were therefore fixed in 1 % FA + 0.2 % GA prior to flow cytometry.

Cells (~$3x10^6$) were harvested by centrifuging and fixed in 120 µl fixative (1 % FA, 0.2 % GA, 10 µg/ml PI in PBS) for 10 minutes. Fixative was then diluted with 500 µl PBS. Cells were run on the FC500 flow cytometer (Beckman Coulter) to measure green fluorescence and red fluorescence as a proxy for GFP and PI respectively. A total of 20,000 events were recorded for each sample within a gating for the cell size. A 'no GFP' control was included in each run. In addition, a PI-positive control was included by permeabilizing the cell membrane by adding 1 % Triton X100 to the fixative.

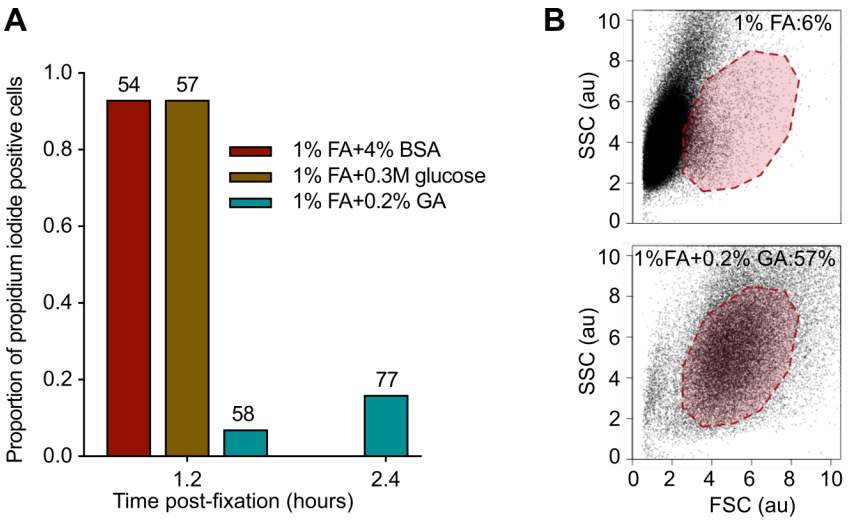


Figure 3.5: Addition of glutaraldehyde improves fixation of *T. congolense*. A. Plot of proportion of propidium iodide positive cells at different times post fixation with different fixative solutions. Numbers on top of bars represent total number of cells counted. **B.** Plot of forward (FSC) vs side-scatter (SSC) of cells fixed with either 1 % formaldehyde (FA) only or 1 % formaldehyde + 0.2 % glutaraldehyde (GA). Red highlight indicates gated region equivalent to live cells with intact membrane.

3.13 Double-strand break assay

Non-induced (0 h) and 48 h tetracycline-induced TcoTTS and TVTTS clones were seeded into 96-well plates at an expected 1.5 cells/well without Hygromycin B. After allowing cells to grow out (11-15 generations), the number of positive wells was counted before sub-culturing into two other plates with or without Hygromycin (2 µg/ml) as test and control respectively. After ~3 generations, test and control plates were compared in order to estimate the number of dead wells on the test plate (with Hygromycin) compared to control plate (without Hygromycin). The loss of Hygromycin resistance was used as a proxy for I-SceI endonuclease restriction activity. I-SceI restriction activity was quantified by modelling the percentage of cutting using Poisson distribution to model number of cells/well, where λ equals positive wells and Binomial distribution to model resistant and sensitive cells.

3.14 Estimating level of protein in whole cell lysates

Approximately 2-3x10^7 cells were harvested by centrifugation 1000 *xg*, washed in PBS, resuspended in Laemmli buffer at 100 °C and heated for 5 mins at the same temperature (100 °C) to obtain cell lysates. Trypanosome proteins were separated

using reducing Sodium Dodecyl Sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Electrophoresis was run in Tris-glycine buffer using 7.5 % or 4-20 % polyacrylamide gel (Biorad). Separated proteins were electrotransferred by immersion onto a nitrocellulose membrane (GE healthcare life sciences). Transfer was run at 40 mA for 14 h in transfer buffer (25 mM Tris, 192 mM glycine, 0.02 % SDS, 10 % methanol). Membranes were stained with Ponceau S (for loading control), scanned and destained by washing with TBS-T (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1 % Tween-20). Membranes were blocked with 5 % (w/v) milk in TBS-T and probed with either 2 µg/ml 1C51 anti-mCherry mouse monoclonal antibody (Abcam), 0.5 µg/ml anti-mCherry rabbit polyclonal antibodies (Abcam), or 1 µg/ml anti-GFP two-monoclonal mouse antibodies (Roche) in 1 % (w/v) milk in TBS-T. Antibody incubations were incubated with rocking agitation for about 1.5 h. Membranes were washed in TBS-T for 1 min and 3x for 10 mins to remove unbound antibodies. Membranes were further incubated with goat anti-rabbit IgG or goat anti-mouse IgG HRP conjugate (Sigma) secondary antibodies, diluted 1:40,000 or 1:20,000 respectively in 1 % milk. Incubation and washing were performed as above. Antibodies were detected using western lightning enhanced chemiluminescence reagent (GE healthcare life sciences), membranes were either exposed to photographic film, or captured by exposure to UV light using the fusion FX spectra (Vilber Lourmat).

3.15 RNAi induction assay and growth analysis

T. brucei cells were seeded at 10^5 cells/ml while *T. congolense* cells were seeded at ~$2x10^5$ cells/ml. RNAi lines were grown with (induced) or without (non-induced) 1 µg/ml tetracycline. At 24 h time points, cultures were observed with the microscope for morphological changes and cell densities estimated. Slides were prepared for each time point and processed as described in Section 3.10. Cultures were maintained at the starting concentration by sub-culturing. Tetracycline was constantly added to induced cell lines at each sub-culture over the period of induction.

3.16 Statistical analysis

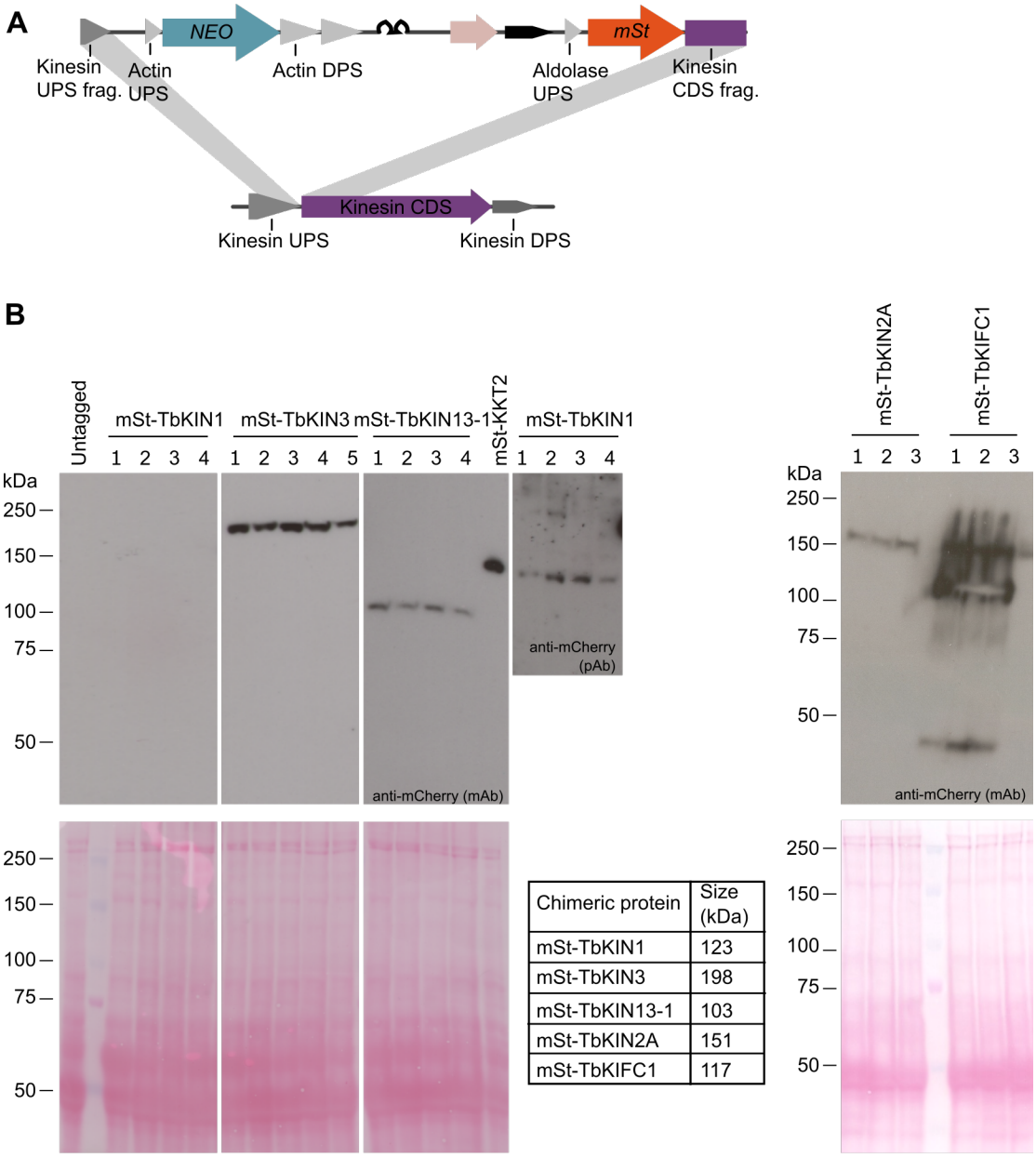
Descriptive statistics including mean, median and range were used to describe data. The student's t test and one-way ANOVA were used to compare means. For multiple hypothesis testing, the fisher's exact test with Bonferroni correction was used. Differences between groups were considered to be significant at p-value of < 0.05. Analysis was performed with GraphPad prism 7 software. For statistical analysis of DRiF-Seq data see Section 3.7.

4 RESULTS 1

4.1 Characterisation of *T. brucei* kinesins

As discussed in the introduction (Section 2.5), to investigate whether *T. brucei* can be a more manipulatable system than current models for the study of kinesins, it is important to characterise the families that are present in *T. brucei.* I first selected 5 kinesin families for the following reasons: Kinesin-1 (Tb927.1.1350), -3 (Tb927.11.2490) and -14 [Tb927.10.14890; TbKIFC1 (Dutoya et al., 2001)] because they are present as single copy genes; Kinesin-2 (Tb927.11.13920) in other systems plays a role in intraflagellar transport; and Kinesin-13 [Tb927.9.3650; TbKIN13-1 (Wickstead et al., 2010b)] is the family we know most about in *T brucei.* Following from the nomenclature of TbKIN13-1, I have called Kinesin-1, TbKIN1, Kinesin-3, TbKIN3 and the selected Kinesin-2, TbKIN2A.

In other to localise these kinesins *in vivo*, I tagged them at the N- terminal end with a fluorescent marker to allow detection of the protein. A linearised endogenous locus tagging construct enabled insertion of an mStrawberry coding sequence at the 5'-end of the coding sequence of each kinesin (Figure 4.1A), generating 5 different lines expressing mSt-tagged TbKIN1, 2A, 3, 13-1 or TbKIFC1. Probing whole cell lysates of selected clones with an anti-mCherry monoclonal antibody (which recognises mStrawberry) shows the expected molecular weight of the chimeric protein for all kinesins (Figure 4.1B), indicative that all generated lines expressed a tagged kinesin as expected. All four mSt-TbKIN1 clones presented very low protein levels; to detect the tagged protein in these cell lysates, the nitrocellulose membrane was reprobed with a polyclonal antibody and exposed for several hours onto x-ray film (Figure 4.1B).



Chimeric protein	Size (kDa)
mSt-TbKIN1	123
mSt-TbKIN3	198
mSt-TbKIN13-1	103
mSt-TbKIN2A	151
mSt-TbKIFC1	117

Figure 4.1: Tagging of selected kinesins. A. Schematic showing integration of a construct to N-terminally tag TbKIN1, 2A, 3, 13-1 and TbKIFC1. UPS/DPS = upstream/downstream processing sequences. **B.** Western blot of whole cell lysates of selected clones detecting mStrawbery (mSt) using mouse anti-mCherry monoclonal or polyclonal. mSt-KKT2 was used as a tagged control. Ponceau staining is used as loading control.

4.1.1 *T. brucei* kinesins localise to different subcellular locations associated with their possible function

To determine the subcellular localisation of selected kinesins, I used native fluorescence microscopy to visualise mStrawberry.

In agreement with the low levels of mSt-TbKIN1 seen in Western blots, I could not visualise mSt-TbKIN1 signal above background by native fluorescence microscopy (Figure 4.2A), and was therefore unable to determine its localisation. Localisation data was later published as part of the Tryptag project (www.tryptag.org), where TbKIN1 was tagged at the C-terminus with mNeonGreen (mN) in procyclic-form *T. brucei* (Shaner et al., 2013). Tryptag is a project aimed at providing the localisation of all proteins encoded within the trypanosome genome (Dean et al., 2017). With their TbKIN1-tagged line, the researchers were able to detect signal above background. Their data show a punctate cytoplasmic distribution for TbKIN1-mN in procyclic cells (Figure 4.2B), while I observed no signal above background in bloodstream forms (Figure 4.2A).

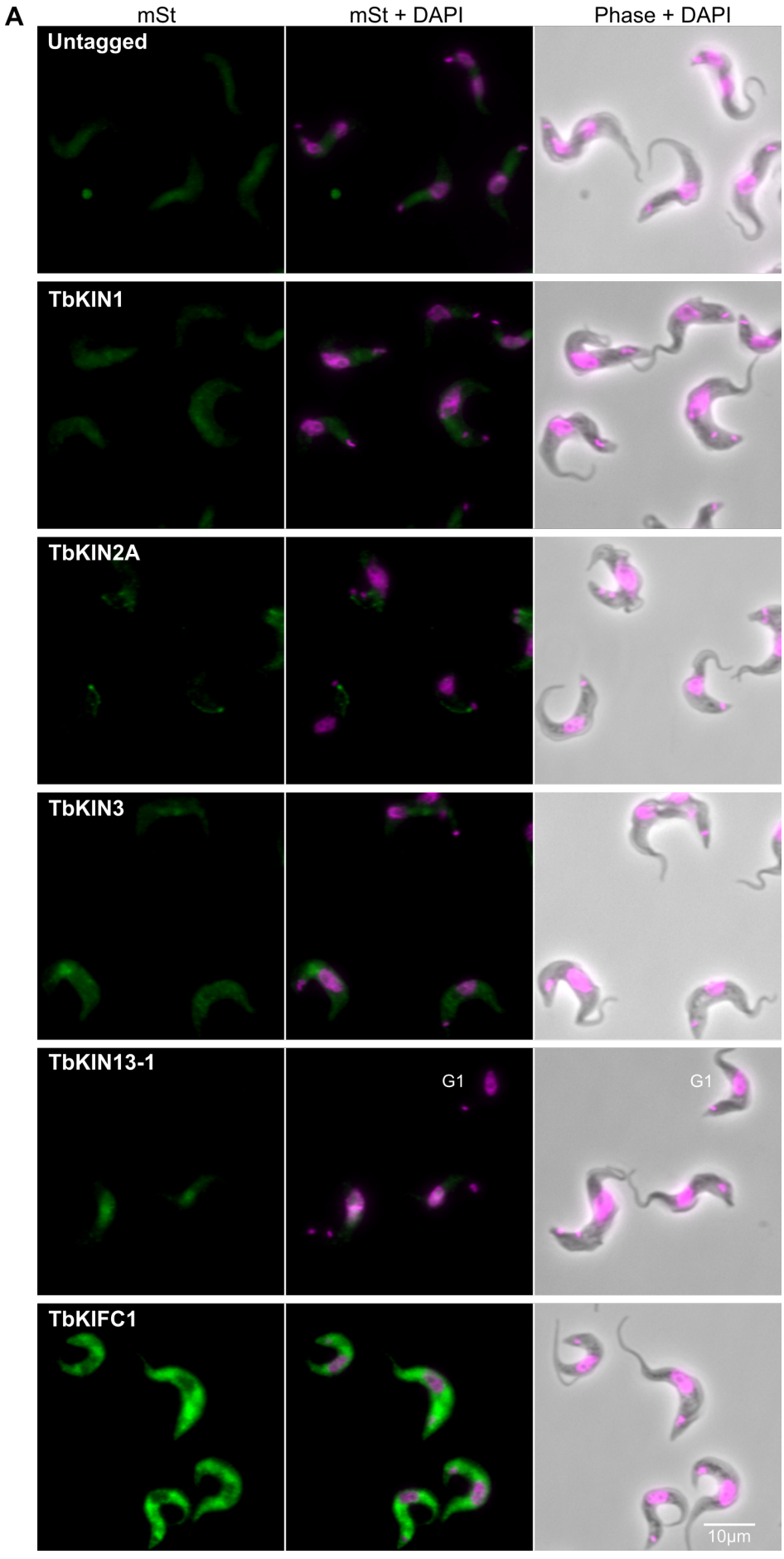
Native fluorescence of mSt-TbKIN2A shows flagellar localisation (Figure 4.2A). The kinesin appears in a punctate pattern along the flagellum and is enriched at the base of the flagellum. Localisation of TbKIN2A is consistent with an involvement in intraflagellar transport.

mSt-TbKIN3 is enriched at the region between the kinetoplast and the nucleus (Figure 4.2A). In *T. brucei* this region contains the endosomes, lysosome, Golgi and extensions of the endoplasmic reticulum. These are key organelles of the endomembrane network and in *T. brucei* are located close to the flagellar pocket, the only region where active endocytosis and secretion are known to occur in this parasite. In other systems, Kinesin-3 plays a role in endocytosis by transporting endocytic vesicles. The localisation of mSt-TbKIN3 is therefore consistent with a possible involvement in endocytosis.

Among the Kinesin-13 members in *T. brucei,* TbKIN13-1 is the only one that localises to the nucleus (Chan et al., 2010; Wickstead et al., 2010b). In agreement with published literature, mSt-TbKIN13-1 localises to the nucleus of the bloodstream-form cell (Figure 4.2A). Analysis of fluorescence signal from cells at different cell cycle stages (based on the number of kinetoplasts and nuclei, as well as the nuclear division process) showed that the mSt-TbKIN13-1 was undetectable in G1 cells. Fluorescence signal was detectable from G2 through metaphase and cytokinesis (Figure 4.3). In

metaphase, mSt-TbKIN13-1 was found to associate with the mitotic spindle as the nucleus is pulled apart (Figure 4.3). This shows that TbKIN13-1 has a multi-localisation at different cell cycle stages and associates with the mitotic spindle as previously reported (Wickstead et al., 2010b).

Fluorescence signal shows cytoplasmic localisation of TbKIFC1, consistent with published localisation in *T. brucei* procyclic forms where they are reported to associate with acidic vesicles in the cytoplasm (Dutoya et al., 2001).


A
mSt
mSt + DAPI
Phase + DAPI
Untagged
TbKIN1
TbKIN2A
TbKIN3
TbKIN13-1
G1
G1
TbKIFC1
10µm

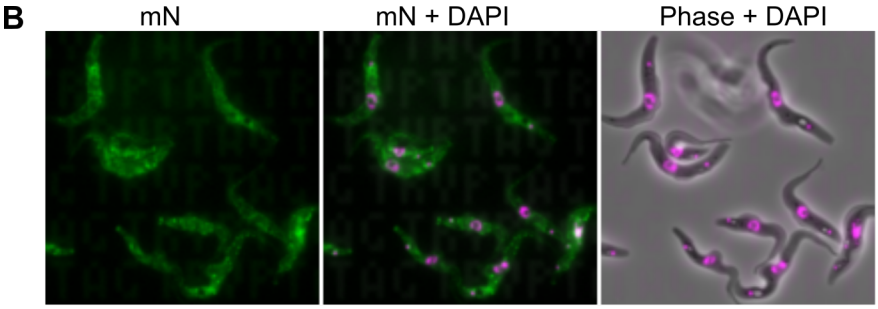


Figure 4.2: Localisation of *T. brucei* kinesins. A. Native fluorescence microscopy of mSt-tagged kinesins showing the different localisations of each kinesin. Images are representative of population. **B.** Fluorescence micrographs of mNeonGreen-tagged TbKIN1 (Adapted from www.tryptag.org)

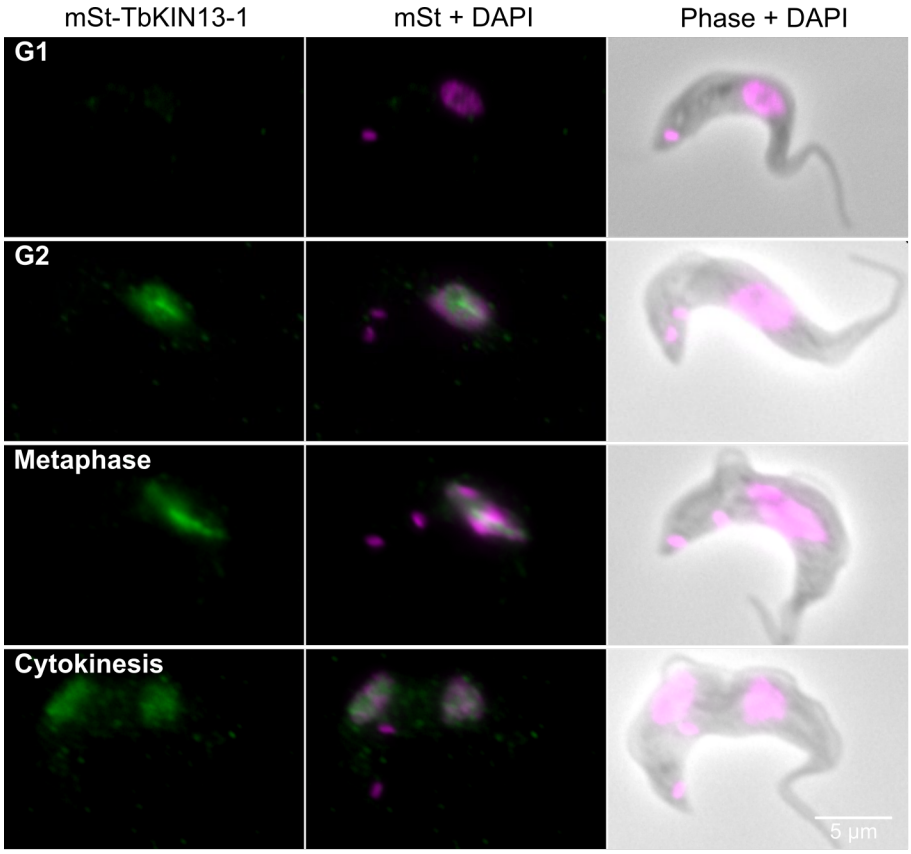


Figure 4.3: Micrograph of mSt-TbKIN13-1 localisation at different stages through the cell cycle.

4.2 Effect of depletion of *T. brucei* kinesins

To characterise effect of loss of kinesins, I targeted them for RNAi-mediated depletion. Constructs for RNAi in *T. brucei* can be designed to generate a double-stranded RNA of a sense and antisense strand from the same fragment (head-to-head) or by generating a single strand containing both the sense and antisense strand (hairpin). Although not clearly proven, it is suggested that because intramolecular folding is

quicker, using hairpin RNA generates more double-stranded RNAs, producing a greater knockdown and an increased RNAi penetrance. I designed constructs to produce both head-to-head and hairpin RNAs. Head-to-head constructs placed opposing T7 promoters at each side of a target fragment, and integrated at the 177 bp-repeat sequence of the minichromosomes (Figure 4.4A).

Head-to-head constructs were used to generate RNAi targeting TbKIN1, 3 and 13-1. Hairpin RNAi fragments were also designed to have a short DNA fragment separating RNAi fragments complementary to each other, and integrated at the minichromosomal VSG-G4 locus. Hairpin constructs were used to generate RNAi targeting TbKIN1, 2A, 3, 13-1 and TbKIFC1. All RNAi constructs were transfected onto the respective mSt-tagged cell line (to enable monitoring of kinesin depletion through mStrawberry level).

I found no evidence of leakiness after transfecting, selecting and analysing RNAi cell lines, as the levels of tagged proteins in parental (tagged) transgenic cell lines and non-induced (tagged +RNAi) cell lines were indistinguishable. The data show knockdown of all kinesins by 24 to 48 h post RNAi induction with tetracycline. Knockdown of TbKIN1, TbKIFC1 and TbKIN2A were below the level of detection (Figure 4.4B). Detectable amount of protein was present after TbKIN3 and TbKIN13-1 knockdown, but the levels were substantially reduced. However, switching to constructs that produce a double-stranded hairpin RNA made no difference to the levels remaining after knockdown (Figure 4.4C).

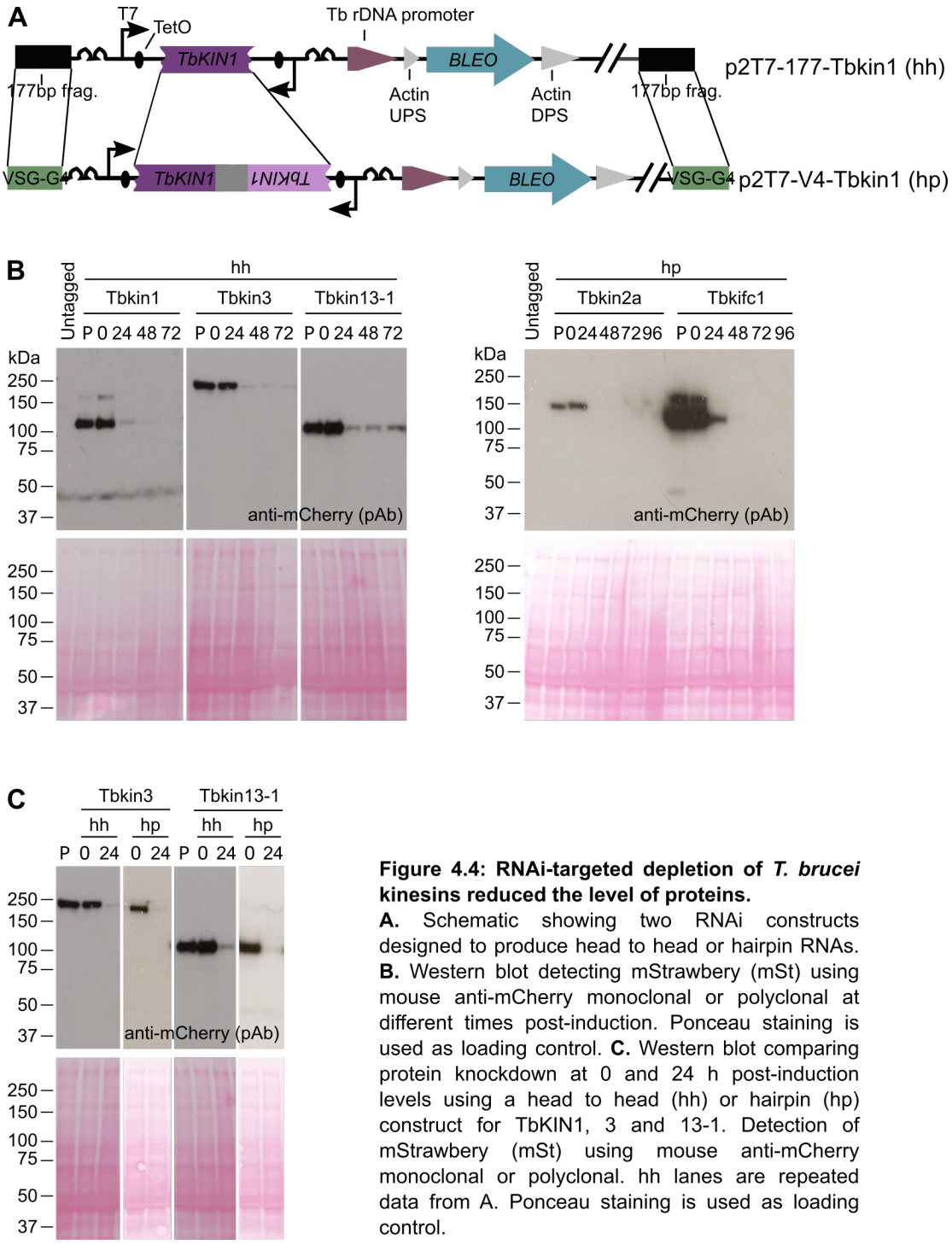


Figure 4.4: RNAi-targeted depletion of *T. brucei* kinesins reduced the level of proteins.
A. Schematic showing two RNAi constructs designed to produce head to head or hairpin RNAs. **B.** Western blot detecting mStrawbery (mSt) using mouse anti-mCherry monoclonal or polyclonal at different times post-induction. Ponceau staining is used as loading control. **C.** Western blot comparing protein knockdown at 0 and 24 h post-induction levels using a head to head (hh) or hairpin (hp) construct for TbKIN1, 3 and 13-1. Detection of mStrawbery (mSt) using mouse anti-mCherry monoclonal or polyclonal. hh lanes are repeated data from A. Ponceau staining is used as loading control.

4.2.1 Depletion of TbKIN1, 2A and TbKIFC1

To determine whether the kinesins targeted are required for normal growth of bloodstream-form cells, I compared the growth rate in culture between the parental, non-induced and tetracycline-induced lines. Cells were also observed for phenotypic changes. The data show that RNAi-mediated depletion of TbKIN1 or TbKIFC1 did not

produce an observable growth or other observable phenotypic defect in cells (Figure 4.5A). However, RNAi-targeted depletion does not completely remove protein so the lack of defect could be because of the residual level. In procyclic *T. brucei,* knockdown of TbKIFC1 also resulted in no detectable growth defect in culture, but did show lower Ca^{2+} content in acidocalcisomes and a dysfunctional release of Ca^{2+} (Dutoya et al., 2001). This was not investigated in current mutants, and TbKIFC1 was not pursued further.

TbKIN2A showed a small defect in population growth and there was some observed phenotypic defect in culture (Figure 4.5B). The progression of *T. brucei* through the cell cycle can be monitored using the duplication of the kinetoplast and nuclear segregation. A trypanosome in G1 has one kinetoplast (K) and one nucleus (N). As the cell cycle progresses, the kinetoplast first replicates and divides to give a cell with 2K1N. During mitosis, replication of the nucleus generates a cell with 2K2N. A 2K2N cell finally divides into two daughter cells of 1K1N configuration. When this progression goes wrong, it may reflect in the KN configurations. Depletion of TbKIN2A resulted in an increase in abnormal KN configurations, ~40 % at 72 h induction. Within this abnormal KN population was a subset of 2K2N cells that were unable to go through cytokinesis and restarted a new cycle, creating an integer K to N ratio (xKxN/2xKxN). Most of the cells however had a non-integer K to N ratio, suggesting incorrect division and resulting in the abnormal distribution of KN in daughter cells. Beside its canonical role in motility, proper functioning of the flagellum is important for cell division in *T. brucei.* As shown by the several studies in both insect and bloodstream forms, gene silencing that affects flagellum structure or function ultimately results in failure of cytokinesis [reviewed in (Langousis and Hill, 2014; Ralston and Hill, 2008; Vaughan and Gull, 2008)]. The lack of growth defect is likely due to the presence of a second Kinesin-2 in *T. brucei* (TbKIN2B, Tb927.5.2090) acting redundant to TbKIN2A (Stretton, 2016).

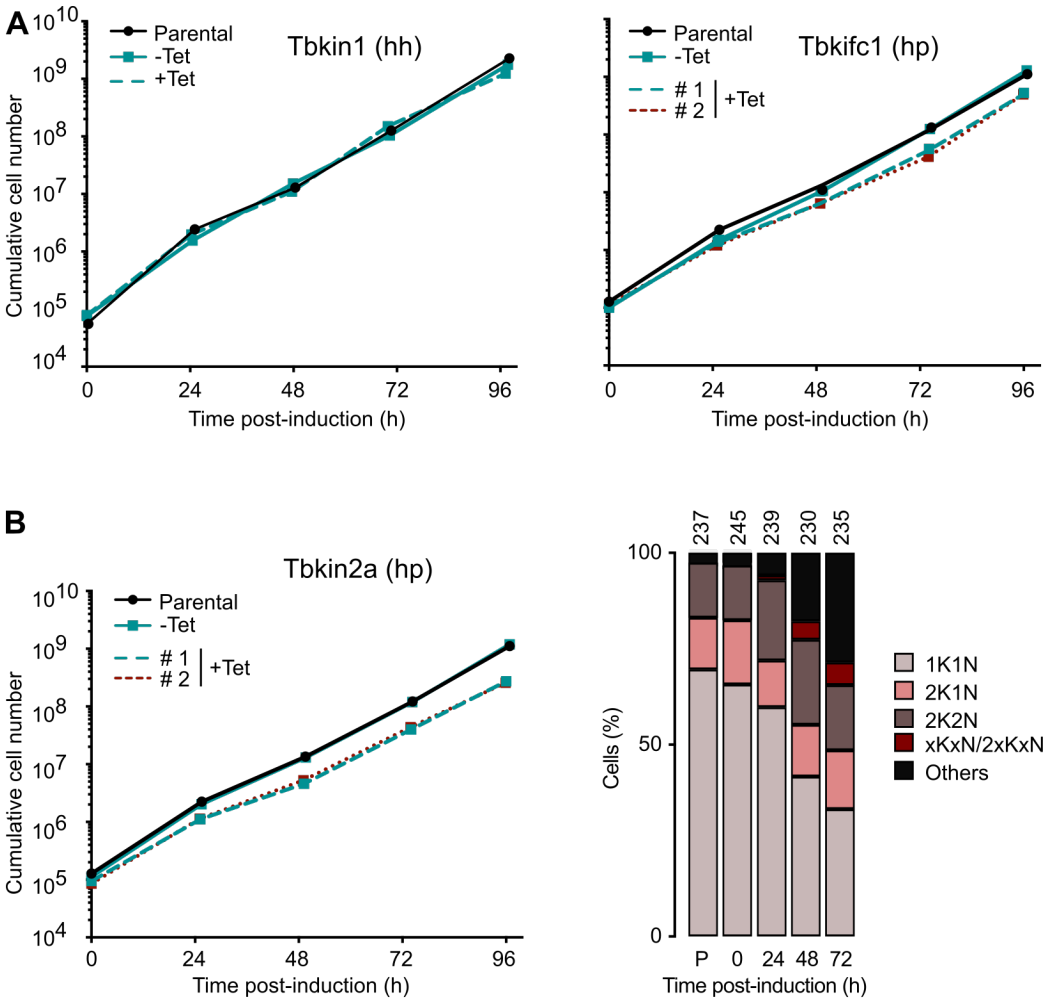


Figure 4.5: Knockdown of TbKIN1, 2A and TbKIFC1 produced little or no growth defect. A. Growth curve of *T. brucei* cells following RNAi-induction of Tbkin1 and Tbkifc1. Tbkifc1 shows 2 technical replicates of induction of a single clone. **B.** Growth curve and quantification of different KN configurations of parental and Tbkin2a non-induced and induced RNAi lines. Image of Tbkin2a (hp) shows 2 technical replicates of induction of a single clone; numbers above bars are total number of cells counted; xKxN/2xKxN, an abnormal KN configuration where the K to N ratio is an integer; Others, abnormal KN configurations where the K to N ratio is a non-integer.

4.2.2 Depletion of TbKIN3 produces growth and phenotypic defect

Depletion of TbKIN3 resulted in a substantial reduction of cell numbers from 24 h post-induction (Figure 4.6A). The level of reduction is consistent with the RIT-seq data which detected a high fitness cost with knockdown of this kinesin. Similarly, loss of KIF1A or KHC function in mice (Yonekawa et al., 1998) and *Drosophila* (Saxton et al., 1991) respectively was associated with lethality. An increasing proportion of cells presented with enlarged flagellar pocket and rounding of the cell body (Figure 4.6B) up to 48 h (Figure 4.6C). Enlargement of the pocket has been reported before for

clathrin knockdown in *T. brucei*. This phenotype was then termed “big eye”, and is associated with the inhibition of endocytosis due to the absence of clathrin (Allen et al., 2003). Clathrin knockdown has no effect on exocytosis; thus, the continuous build-up of secretion and recycling vesicles in the flagellar pocket causes the enlargement of the organelle and consequent distortion of the cell. At 48 h post RNAi induction, 40 % of cells in the population showed large pockets and abnormal body shape. Flagellum detachment was also observed in ~14 % of these defectuous cells. The results from protein ablation support the localisation data, and suggest that TbKIN3 is an essential kinesin involved in endocytosis.

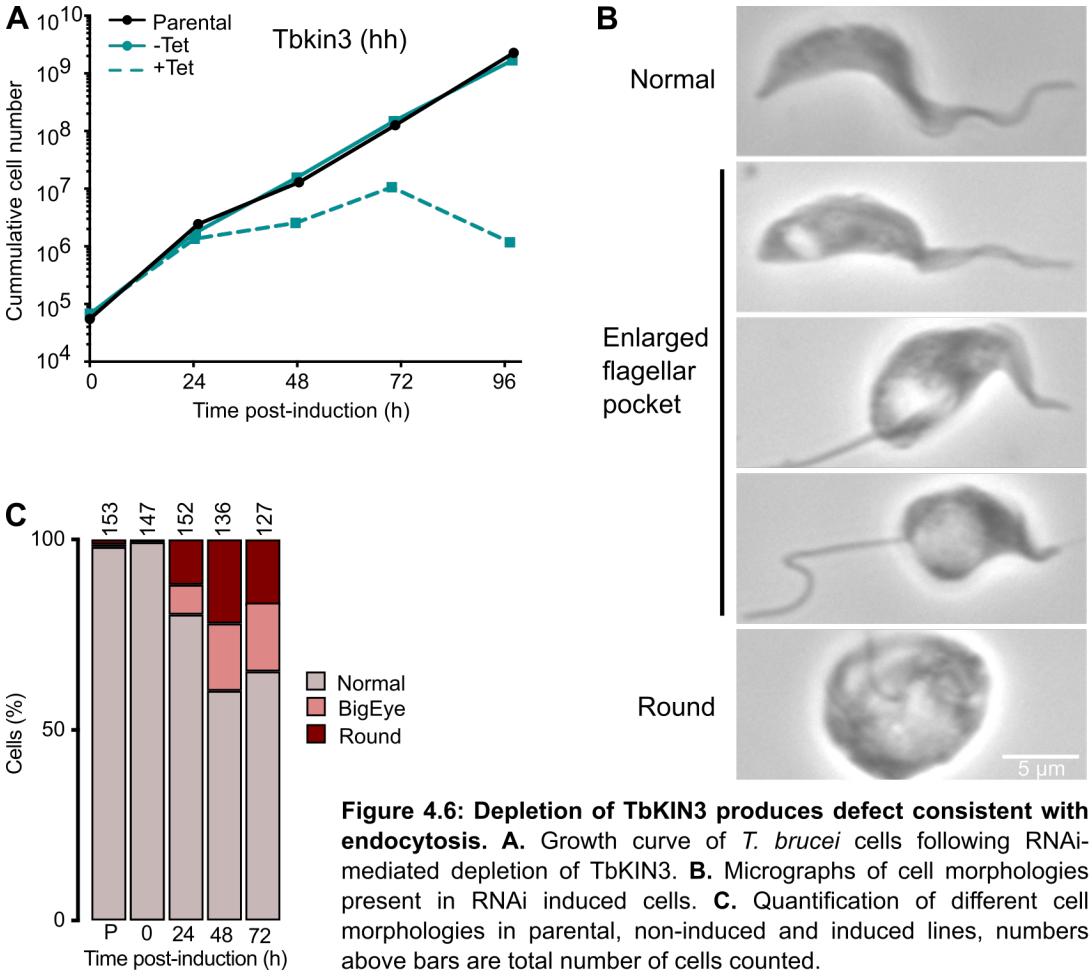


Figure 4.6: Depletion of TbKIN3 produces defect consistent with endocytosis. A. Growth curve of *T. brucei* cells following RNAi-mediated depletion of TbKIN3. **B.** Micrographs of cell morphologies present in RNAi induced cells. **C.** Quantification of different cell morphologies in parental, non-induced and induced lines, numbers above bars are total number of cells counted.

4.2.3 TbKIN13-1 knockdown produced a defect in mitosis

RNAi-mediated knockdown of TbKIN13-1 has been demonstrated in previous studies to exert an influence on mitosis, specifically affecting mitotic spindle dynamics and chromosome segregation (Chan et al., 2010; Wickstead et al., 2010b). This leads to a

build-up of anucleate (1K0N) cells both in procyclic and bloodstream forms of the parasite, although the phenotype was more prominent in bloodstream form (Chan et al., 2010; Wickstead et al., 2010b).

In this study, the cell line containing construct for RNAi-mediated depletion designed to produce head-to-head double-stranded RNA showed no growth defect at the population level (Figure 4.7A), and no significant increase in cells with abnormal KN numbers ($p > 0.05$). Nor did it record the characteristic build-up of anucleate cells as previously reported (Figure 4.7B). In the previous study with bloodstream forms, knockdown produced undetectable level of the protein by immunoblot after 48 h of RNAi induction (Chan et al., 2010).

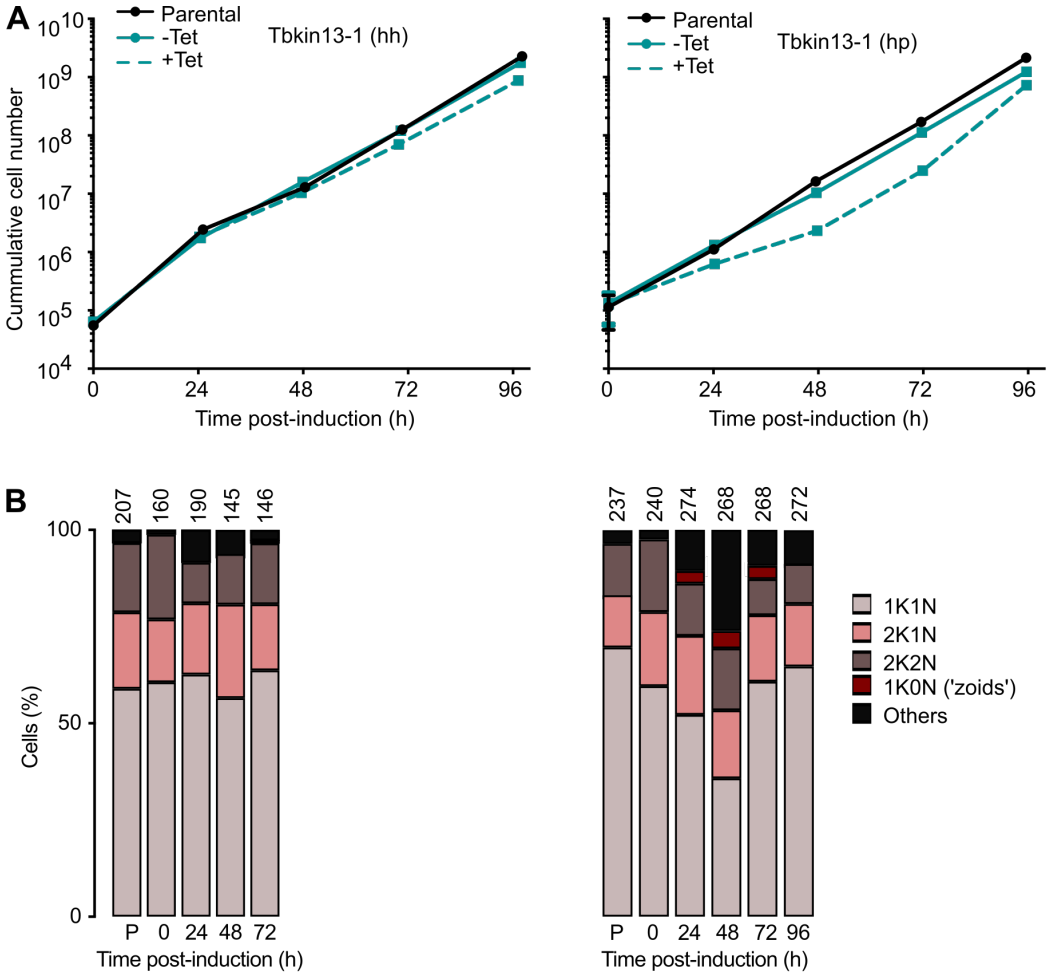


Figure 4.7: Hairpin construct produces marginal defect in RNAi-targeted depletion of TbKIN13-1. A. Growth curve and quantification of KN configurations of parental and Tbkin13-1 non-induced and induced RNAi lines following integration with a head to head construct, numbers above bars are total number of cells counted. **B.** Growth curve and quantification of KN configurations of parental and Tbkin13-1 non-induced and induced RNAi lines following integration with hairpin construct, numbers above bars are total number of cells counted.

In an attempt to improve RNAi penetrance, a hairpin construct was also transfected into the parental (tagged) cell line, but there was no distinguishable difference in protein depletion using either a head-to-head or hairpin construct (Figure 4.4C). However, hairpin RNAi gave a growth defect and an increase in the proportion of cells with abnormal KN configuration by 48 h induction ($p < 0.0001$; x^2) (Figure 4.7B). Interestingly, I found only a small increase in 1K0N cells (~2 %) compared to the reported 25 % (Chan and Ersfeld, 2010), and ~5 % increase in cells that did not undergo cytokinesis (xKxN/2xKxN). Where cytokinesis did occur, an increase in other abnormal KN arrangements (14 %), which is also lower than the published 35 % (Chan and Ersfeld, 2010), suggest either incorrect division or failed nuclear segregation. Together, these data support what is already known for TbKIN13-1: that knockdown affects mitosis. However, defects observed here even with hairpin construct was less than what was reported (Chan et al., 2010).

4.3 Kinesin-1 is not essential to bloodstream-form *T. brucei* in culture

Historically, Kinesin -1 has been seen as the "classical" kinesin because it was the first to be identified. Depletion of KHC in *Drosophila* and KIF5A in mice results in neuronal defect (Nakajima et al., 2012; Saxton et al., 1991). Depletion of KIF5B also in mice cells affected lysosome dispersion and caused aggregation of mitochondria at the perinuclear position. In *T. brucei,* knockdown of TbKIN1 was associated with a significant loss of fitness in bloodstream forms by RIT-seq (Alsford et al., 2011). However, re-analysis of the RIT-seq raw data with more sensitive and appropriate methods (i.e. a posteriori noise estimation) suggests that this is likely to be a false positive (Bill Wickstead, personal communication).

In my experiments, depletion of TbKIN1 produced no detectable growth or phenotypic defect. It remains unclear if residual protein level following gene silencing is sufficient to prevent cellular defects, or TbKIN1 is redundant in *T. brucei.* To investigate whether TbKIN1 is essential, I knocked out both alleles of *TbKIN1* by replacing them with two different drug markers (Figure 4.8A). Clones were obtained after each round of single and double knockouts, and verified by diagnostic PCR. PCR results showed that all

double knockouts have entirely lost *TbKIN1* (Figure 4.8B). However, the growth rates of hemi- and double knockouts were not different from each other or the parental (p = 0.98, one-way ANOVA) (Figure 4.8C), indicating that *TbKIN1* is non-essential in *T. brucei.*

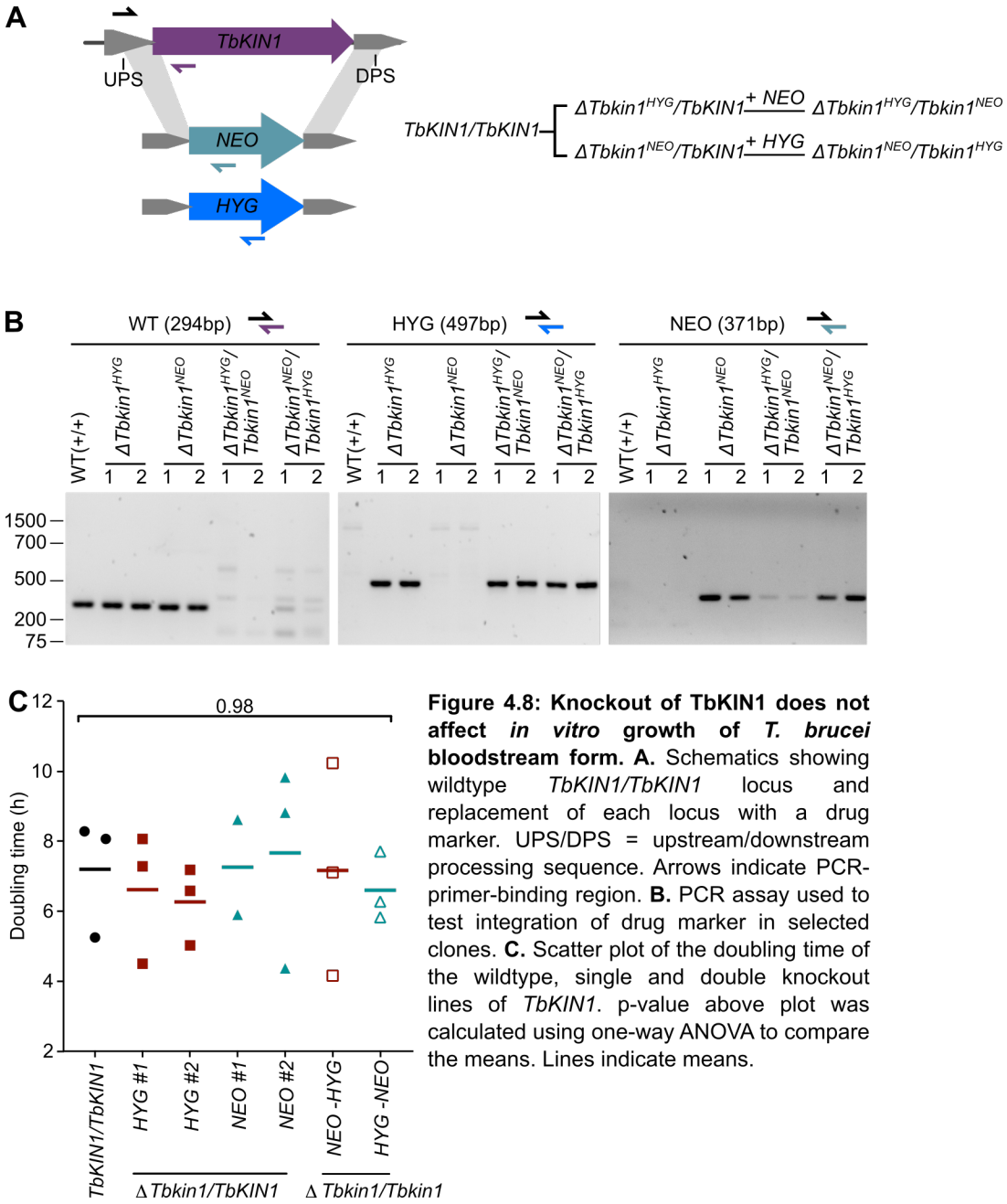


Figure 4.8: Knockout of TbKIN1 does not affect *in vitro* growth of *T. brucei* bloodstream form. A. Schematics showing wildtype *TbKIN1/TbKIN1* locus and replacement of each locus with a drug marker. UPS/DPS = upstream/downstream processing sequence. Arrows indicate PCR-primer-binding region. **B.** PCR assay used to test integration of drug marker in selected clones. **C.** Scatter plot of the doubling time of the wildtype, single and double knockout lines of *TbKIN1*. p-value above plot was calculated using one-way ANOVA to compare the means. Lines indicate means.

5 DISCUSSION I

T. brucei has a complex cytoskeleton dominated by tubulin, making it a great model to study kinesins. The advancement of genetic tools in *T. brucei* makes it easy to perform various functional studies, and in Part I of this thesis, I reported on the localisation and gene silencing of *TbKIN1*, *2A*, *3, 13-1* and *TbKIFC1* by RNA interference (RNAi) as well as knockout of *TbKIN1.*

5.1 Can *T. brucei* be used as a model to study kinesins?

5.1.1 Comparison of *T. brucei* kinesins to common model organisms

Cellular localisation and knockdown data indicate that Kinesin-2, -3 and 13-1 in *T. brucei* perform roles similar to what is known about these families in other organisms. There is therefore conservation of function between related model organisms (fungi and animals) and the more divergent *T. brucei.*

Yet, I was unable to reproduce the growth defect previously reported for knockdown of TbKIN13-1 in bloodstream-form cells (Chan, Matthews and Ersfeldk, 2010). That was independent of RNAi construct type (hairpin or head-to-head) and construct integration site – in my hands, cell growth defect was milder than the one observed by Chan and colleagues. However, my individual RNAi studies agree very well with the results from competition studies produced in the Wickstead lab through genome-wide RNAi libraries.

While there is conservation of function of some families in *T. brucei* we also find examples of families that are different from the common models. Cellular localisation and the absence of observable growth defect when ablating Kinesin-14 agree with published data (Dutoya et al., 2001), which show Kinesin-14 to play a novel role transporting the organelle acidocalcisomes

The well-established function of Kinesin-1 as a transporter has largely come from investigating multicellular organisms such as fruit flies and mice. In mammals, isoforms of Kinesin-1 are expressed in different cell types and transport different

cargo. In macrophages, Kif5B plays a role in receptor-mediated phagocytosis by assisting recycling of receptors from the phagosome and delivery of Rab11-containing membrane stores to effect pseudopods (Silver and Harrison, 2011). Kif5A, on the other hand, functions in the nervous system, transporting neuronal vesicles (Nakajima et al., 2012). While this evolutional adaptation to multiple roles fits why Kinesin-1 will be important in a multicellular organism, in a unicellular organism, Kinesin-1 may have no specialised cargo to transport. It may be that it works alongside another transporter like Kinesin-3 to transport the same kind of cargo, which introduces redundancy. Indeed, in *Dictyostelium*, knockout of one (*Ddkif5*) of two Kinesin- 1 family members does not affect cell growth (Iwai et al., 2004) and the role of *Ddkif3* that appears to be essential could not be determined due to repeated failure to knockout the gene (Röhlk et al., 2008).

However, finding that bloodstream-form *T. brucei* survive *in vitro* at the loss of Kinesin-1 does not rule out the possibility that this kinesin may be required at a different lifecycle stage or only in the natural host.

5.1.2 *T. brucei* kinesin candidates for investigating eukaryotic kinesins

I have shown that at least 3 *T. brucei* kinesins (Kinesin-2, 3 and 13-1) produce measurable growth defect that make them suitable candidates for further investigation in understanding functional differentiation.

Kinesin-3 is represented by a single kinesin gene in *T. brucei* and produced a strong growth phenotype on knockdown. The genome of *T. brucei* encodes two Kinesin-2 proteins, TbKIN2A (Tb11.01.5490) and TbKIN2B (Tb927.5.2090), that may act redundant to each other. Although knockdown of TbKIN2A produced a measurable defect, it has been shown that when both are knocked down there is stronger growth impairment (Stretton, 2016). Similarly, knockdown of the only Kinesin-13 member associated with the mitotic spindle produced a mild defect. One way of intensifying the observed growth defect is to introduce an agent that affects microtubule dynamics. Nocodazole is a chemical agent known to bind to β-tubulin and disrupts microtubule polymerisation. Because of its ability to destabilize microtubules, it has

several applications in cell biology research to manipulate microtubule dynamics. It has been used to synchronise cell division for analysis of DNA content (Li et al., 2016; Yiangou et al., 2019) and to induce Golgi ministacks by disrupting trafficking between Golgi and the endoplasmic reticulum (Fourriere et al., 2016). A low concentration of nocodazole, enough to only slightly destabilize microtubules, may be enough to intensify the growth defect from knocking down TbKIN13-1, which would enable more complex and insightful experiments, e.g. those to detect rescue from RNAi effect.

5.2 Proposed future studies

5.2.1 Contribution of the motor domain to kinesin function

In understanding the role of the conserved motor domain to functional differentiation, it is important to investigate how the motor domain generally affect the function of a kinesin. One way to assess this is by swapping the motors of different kinesins to see whether it will make any difference to function within cells. In the past when others have experimentally fused parts of different kinesins, assessment of the chimeric kinesins has been by comparison of their biochemical properties to the wildtype counterpart (Shastry and Hancock, 2010; Thiede et al., 2013). While chimeric kinesins tend to have properties that lie between that of the parents, nothing is known about their functional competence within the cell. A *T. brucei* model offers the opportunity to undertake such an experiment.

Although Kinesin-1 was not identified as a candidate in *T. brucei* for investigating kinesin function because of non-essentiality *in vitro,* it could perhaps be used in a motor domain switch experiment. While assessing the effect of motor domain on kinesins associated with a defect is the first goal, another interesting question to ask is what will happen if we were to switch the motor of a non-essential kinesin with that of an essential kinesin? Would a chimeric *T. brucei* Kinesin-1 having a Kinesin-3 motor for example attempt to rescue the phenotype produced on knockdown of Kinesin-3?

5.2.2 A new approach to study functional differentiation of kinesins in *T. brucei*

The evolution of gene function remains important in understanding the diversity of present-day genes. One way to understand mechanism of how ancient genes diversified to perform various function within the same organism, or as orthologues in different organisms, is the reconstruction of ancestral genes. Ancestral sequence reconstruction is a method of studying ancient genes by phylogenetically inferring their sequence using present-day sequences, *de novo* synthesising them and studying their function either *in vitro* or in a model system. In this way, areas of the protein more likely to affect its function become the focus (Randall et al., 2016; Thornton, 2004). For kinesins, this area will be the conserved motor domain. We propose a similar approach to study functional diversity in eukaryotic kinesins using the ancestral gene, specifically of the motor domain sequence, of kinesins. Additionally, the inclusion of a consensus motor domain sequence is proposed. As the name suggest, this calculates the most frequent residue at each position in a multiple sequence alignment of the motor domain across representative eukaryotes. Such a study will use both the consensus and ancestral sequences of a kinesin family as a representation of the eukaryotic superfamily. The consensus represents the present-day eukaryotic sequence and the ancestral represents what existed in the past.

Rescue experiments could be set up by switching a kinesin wildtype motor with its an ancestral or consensus motor. This is where the technology of *T. brucei* becomes really useful. Using RNAi, we can selectively knockdown the wildtype kinesin and investigate whether any of the hybrid kinesins will attempt to rescue the expected RNAi phenotype.

Secondly, by comparing the consensus with the ancestral sequence, we can identify motor domain residues that have been conserved over years of evolution and diversification, with the understanding that conserved residues would be important to kinesin function. The desired outcome will be to determine the effect identified residues have on the function of the different kinesin families. A system like this will be immensely useful in investigating functional differentiation among the various kinesin families in eukaryotes.

6 INTRODUCTION II

Animal African trypanosomiasis (AAT) is a complex disease caused by different trypanosomes, largely *T. brucei, T. congolense* and *T. vivax* (see Section 1.2). *T. congolense* and *T vivax* cause more cases of AAT than *T. brucei,* but most investigations on trypanosomiasis have used *T. brucei.* Initial emphasis on human disease (HAT) prioritized *T. brucei* as the primary model for studying trypanosome infections in general, but on the basis of importance, both *T. congolense* and *T. vivax* would be more suitable models for AAT. In some areas, *T. congolense* is the highest cause of AAT, sometimes responsible for >50% of infection (Muhanguzi et al., 2017; Sheferaw et al., 2019; Tchamdja et al., 2017). More importantly, *T. congolense* is different from *T. brucei* in its life cycle, disease presentation and response to drugs; there are reports of increasing drug resistance particularly in *T. congolense* (see Section 6.1). Tailoring and improving AAT interventions will require better understanding of the most important causative agents, or at least a representative organism chosen with the intention of investigating AAT.

T. congolense satisfies a lot of traits necessary for a new model organism. It occupies a biological niche as a better representative of AAT (see above). In addition to this, it is cheap to grow and multiplies at a reasonably fast rate. In our hands, bloodstream-form *T. congolense* strain IL3000 has an average doubling time of 12 hours. The parasite genome has been sequenced (see Section 6.2.1) and axenic culture has been established (see Section 6.2.2). However, genetic technology for working with the infective, bloodstream form of *T. congolense* is non-existent (see Section 6.2.3). In order for *T. congolense* to be set up as a model for AAT, there is a need to create genetic tools.

The ability to extrapolate our extensive knowledge of genetic manipulation from *T. brucei* would greatly facilitate the establishment of *T. congolense* as an AAT model. With this in mind, it is worth noting that *T. vivax* does things a bit more differently to *T. brucei* than *T. congolense. T. vivax* shows a clear difference in disease spread by its adaptation to non-tsetse fly transmission. The extent of this difference is not known hence extrapolating knowledge from *T. brucei* to *T. vivax* may not be as

straightforward. While we cannot be certain how much knowledge from *T. brucei* is applicable to *T. congolense* we at least know that transmission of these two parasites bear more similarity. In this section, I present data aimed at improving the genetic tractability of *T. congolense* as a new model for investigating AAT.

6.1 Drug resistance in *T. congolense*

The threat of drug resistance in African trypanosomes is one very important reason to focus on *T. congolense*. There are three main drugs for treating AAT in the field – homidium chloride, isometamidium chloride (ISM) and diminazene aceturate (Table 1.1). Of these, ISM is the only prophylaxis and the most important treatment for *T. congolense* infections. ISM resistance has been reported in *T. congolense* (Clausen et al., 1992; Moloo and Kutuza, 1990) and is increasing (Mungube et al., 2012; Tchamdja et al., 2017). Further to this, the identification of *T. congolense* isolates with resistance to all 3 treatment options (Afewerk et al., 2000; Mulugeta et al., 1997) poses even a greater threat to the future of AAT treatment. In some reports, the mode of resistance to AAT drugs in *T. brucei* has been linked to transporters and efflux pumps (Bridges et al., 2007; Carter et al., 1995; Mäser et al., 1999). It is unclear whether this is the same for *T. congolense.* However, it has been reported that although *T. brucei* aminopurine transporter (TbAT1) mediates resistance to diminazene drugs, its closest paralogue in *T. congolense* (TcoAT1) is not involved in diminazene uptake and is therefore unlikely to play a role in drug resistance (Munday et al., 2013). The advantage here of adapting *T. congolense* as a model is that any drug studies will be directly applicable to the most threatening AAT agent with regard to drug resistance, and an intervention that primarily considers *T. congolense* infections may be a quick way to bring AAT under some control.

6.2 *T. congolense* as a new model for studying animal African trypanosomiasis

6.2.1 Genome

The genome of *T. congolense* IL3000 strain has been sequenced by partial genome shotgun sequencing (Jackson et al., 2012). This reference IL3000 genome is available at TriTrypDB (www.tritrypdb.org), although gene annotation remains poor due to the small number of investigations with *T. congolense*. A PacBio assembly of the same IL3000 strain (Abbas et al., 2018) provides additional information on the genome structure of *T. congolense*.

6.2.2 Axenic culture

Unlike *T. brucei, in vitro* culture has been established for all lifecycle stages of *T. congolense*, enabling complete differentiation to successive stages in the cycle without genetic manipulation. Epimastigote differentiated from procyclic forms (PCF) were successfully cultured in MEM with glutamine supplemented with 20% foetal bovine serum (FBS). After a few weeks, this generated metacyclic forms (Gray et al., 1984) that could be maintained in MEM supplemented with 20% FBS in the presence of bovine aorta endothelial cell monolayers (Gray et al., 1985). Bloodstream forms (BSF) were also adapted to culture and could differentiate to insect PCF (Hirumi and Hirumi, 1991; Ross et al., 1985), but it was only recently that a more defined, long-term culture was established for BSF (Coustou et al., 2010). In their work, Coustou and colleagues defined a set of culture media for the various stages of the parasite from 6 different savannah strains. Using several strains was especially important because early studies had established that strains of *T. congolense* behaved quite differently from each other even in the composition of medium that supported their growth, and the savannah strain IL3000 proved to be the most malleable (Frame et al., 1991; Hirumi and Hirumi, 1991). It is still unclear why this difference exists between strains.

6.2.3 Genetic modification in *T. congolense*

Some genetic modification has been performed on *T. congolense* in the past. While this is in no way extensive, some of this pioneer work has provided key knowledge needed to develop specific tools and expand the range of modification and research application of *T. congolense.*

Exogenous gene expression in T. congolense

As the only African trypanosome whose complete lifecycle can be reproduced *in vitro*, *T. congolense* was initially recognized as an attractive model to investigate stage-specific gene expression. Downey and Donelson performed a promoter-search study after a failed attempt to express luciferase in *T. congolense* under the control of *T. brucei* promoter and UTRs (Downey and Donelson, 1999a). This led them to identify a transcription initiation point upstream of the rDNA and a 377 bp fragment that transcended the initiation point that successfully drove the transcription of transgenic luciferase (up to 35000-fold above background) in a transient transfection. The promoter test construct used (Figure 6.1A) contained *T. congolense GARP* Upstream (UPS) and downstream (DPS) processing sequences.

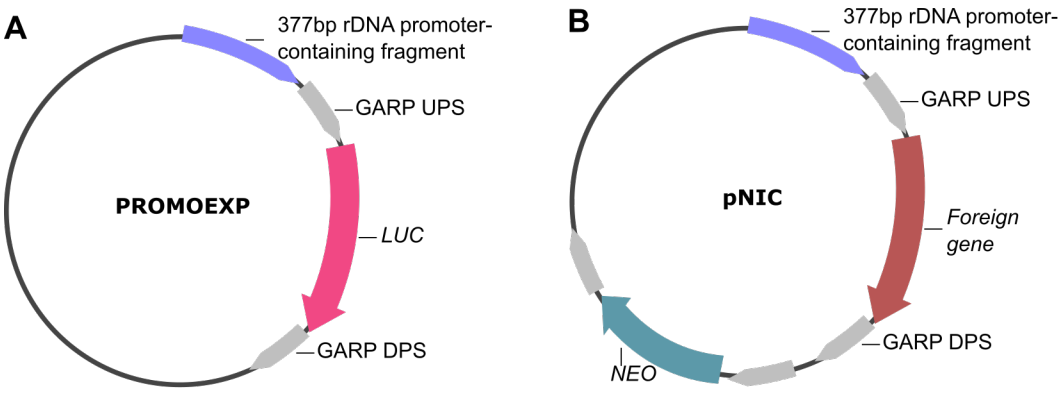


Figure 6.1: Constructs used in genetic modification of *T. congolense*. A. Promoter test construct used in transient transfection. **B.** Construct integrated into genome in a stable transfection.

In an initial attempt to transfect PCF cells, the methodology of electroporation designed for *T. brucei* was directly transferred to *T. congolense.* Owing to the success of transient transfection, the same group proceeded to show that, as with *T. brucei,* by applying selective drug pressure we can select for transfectants that have integrated a construct into their genome (Downey and Donelson, 1999b). Similar to

Figure 6.1A, *GARP* processing sequences flank the gene of interest for integration at the *GARP* locus along with *NEO* for selection (Figure 6.1B). Even with a high number of cells (~$3x10^8$) and large amount of DNA (100µg), the process was hugely inefficient. The researchers reported ~10% of cells surviving electroporation, but provided no information on how many cells successfully integrated the DNA construct. This plasmid was used to express luciferase, GFP and *T. brucei* PARP, but an attempt to overexpress *T. congolense* cysteine protease failed – the construct was reportedly present, but there was no detectable increase in protease level or activity. It was suggested that the rDNA promoter may not be strong enough for overexpression, or expression may be under the strict control of endogenous factors.

Inducible gene expression and depletion

The expression of exogenous genes could not be regulated in *T. congolense* and there was no system to deplete endogenous genes. Once again, drawing from the experience of *T. brucei,* the tetracycline-inducible system and RNA interference (RNAi) were exploited in *T. congolense. T. brucei* plasmids pLEW13 and pLEW29 (Wirtz et al., 1999b) integrating at the tubulin locus were used to create tetracycline-inducible gene expression systems in *T. congolense* PCF TRUM183 (TRUM183: 29/13 cell line) (Inoue et al., 2002) and IL3000 (IL3000: 13-29 cell line) (Coustou et al., 2010). Both plasmids heterologously target *T. brucei* sequences for integration in *T. congolense* and also possess *T. brucei* processing sequences. Stable integration requires flanking sequences to direct homologous recombination into a targeted locus (ten Asbroek et al., 1993). *T. brucei* sequences with sufficient homology can direct integration to *T. congolense*. While the original *T. brucei* RNA Pol I large subunit gene sequence in pLEW29 failed to integrate in *T. congolense,* targeting the tubulin locus was successful. UPS and DPS flanking the expressed gene influence efficiency of gene expression. As a result, even when *T. brucei* processing sequences work in *T. congolense*, as with pLEW13 and 29 constructs, inefficient expression may occur. Also, integrating at an actively transcribed locus could explain the poor gene regulation subsequently observed in RNAi studies. By using a *T. brucei* RNAi construct targeting the depletion of α-tubulin ($p2T7^{Ti}$/ *αTUB*) (LaCount et al., 2000), Inoue and colleagues witnessed that

after transient transfection, in the absence of tetracycline induction, 20% of cells showed the characteristic "FAT" phenotype of tubulin depletion. When RNAi was induced, this increased to 50% FAT cells after 48h in the presence of tetracycline (Inoue et al., 2002). Poor regulation (maximum of 4-fold) resulting in some basal transcription in the absence of RNAi induction was also reported for IL3000: 13-29 line (Coustou et al., 2010).

Genetic modification of different life stages

As direct genetic modification was demonstrated in PCFs, some studies attempted to produce transgenic lines of other lifecycle stages either by direct transfection or by capitalizing on the ability to sequentially differentiate *in vitro* (through the process of metacyclogenesis). Transgenic expression of eGFP was made directly in procyclic or epimastigote forms and then differentiated into transgenic epimastigotes and metacyclics *in vitro* (Sakurai et al., 2009; Suganuma et al., 2012, 2013).

Development of a robust long-term axenic culture for BSFs set the stage for direct transfection procedures and transgenic modification of *T. congolense* BSFs. Coustou and colleagues were the first to attempt direct transfection of BSFs (Coustou et al., 2010). In their work, they also improved transfection success rate for PCFs, as no efficiency was seen for several strains, and for the reference IL3000 success was < 30%. Success in this case refers to obtaining at least one single independent clone, which was less than a third of the time. The best conditions using Nucleofection technology (Lonza) improved IL3000 transfection success in PCF to 75%, but the same could not be said for BSFs (Table 6.1). From 13 attempts, only 3 were successful. Because of the consistent failure, they were unable to generate an inducible IL3000 BSF cell line by direct transfection, and could not perform direct RNAi on BSF *T. congolense.* Coustou and colleague, however, managed to differentiate transgenic strains through the complete life cycle and observed corresponding phenotypes of GFP fluorescence, luciferase activity and tubulin depletion phenotype, and obtained a differentiated BSF of the procyclic IL3000:13-29 inducible line. Transfections reported here were all made by the heterologous integration of plasmids containing *T. brucei* integration (tubulin) and processing sequences (Table 6.1). Heterologous use of

T. brucei signals may not be efficient, and could explain why inducible expression of luciferase in their IL3000:13-29 showed only 4-fold increase above background.

While some improvement has been made in genetic modification of *T. congolense,* there are issues with efficiency marked by frequent failures of transfection, inconsistency of targeting, poor tetracycline regulation and a lack of BSF-specific tools. Consistent success of transfection is a first step yet to be achieved. Without it, even single gene studies are arduous and large-scale genome-wide screens virtually impossible.

	Parental line	Amaxa solution/ program	Construct transfected	Successful/total assay
PCF	IL3000	Parasitic 2/X-001	pLEW13 pLEW20c-GFP	1/1 1/1
	IL3000:13	Parasitic 2/X-001	pLEW29c	1/2
	IL3000:13-29	Parasitic 2/X-001	pLEW20c-GFP pLEW20c-Rluc p2T7^{Ti}/*αTUB*	1/1 1/1 1/2
BSF	IL3000	Parasitic 2/U-033	pLEW13 pLEW20c-Rluc	1/1 0/1
	IL3000:13	Parasitic 2/X-001, U-033	pLEW29c	0/7
	IL3000:13:29*		pLEW20c-Rluc	1/1

Table 6.1: Transfection success rates [reproduced from (Coustou et al., 2010)]. * Differentiated from PCF to BSF. Constructs with c indicate that original integration sequence had been replaced with tubulin sequence (of *T. brucei*).

6.3 Aim II

In Part II, I aim to present *T. congolense* as a new specific biology model. As emphasised earlier, *T. congolense* is a very suitable model for AAT, but has severe limitation in the area of genetic technology. Being able to improve our ability to genetically manipulate this trypanosome will encourage its use as a more appropriate model to tackle AAT. Here, I address the following specific objectives in *T. congolense*:

1. Create tools for inducible transgene expression
2. Develop a system for specific silencing of genes
3. Develop a system for high efficiency transgenesis
4. Generate a genome-wide RNAi library to screen gene fitness cost

5. Conduct a genome-wide screen for isometamidium resistance in *T. congolense*

7 RESULTS 2

7.1 Genetic modification of *T. congolense*

7.1.1 *T. congolense* single marker cell line

In the last two decades, *T. brucei* research has been greatly helped by a tetracycline inducible system (Wirtz and Clayton, 1995; Wirtz et al., 1999a). Precise control of exogenous gene expression is mediated by expressing the bacterial Tetracycline repressor protein (TetR) in *T. brucei*. By placing an operator sequence (TetO) immediately downstream of a strong promoter, gene expression can be regulated by tetracycline. TetR detaches from the TetO sequence in the presence of tetracycline, giving the polymerase access to the promoter to drive transcription of downstream genes. While there are other systems for controlling exogenous gene expression in *T. brucei,* the tetracycline system remains highly favoured. Unlike prokaryotic systems, tetracycline is well tolerated by trypanosomes and can be used in high concentrations without toxicity to cells. It has the highest inducibility fold change recorded for any inducible system. Against the vanillic acid induction system and under the control of the same promoter, tetracycline induction produced approximately 14-fold more protein than vanillin induction (Sunter, 2016). Another system described in *T. brucei,* the IPTG induction system (Tschopp, Charrière and Schneider, 2011), has not been widely used or characterised. To date, there is only a single report of its use in tRNA expression. In this report, the authors noted expression of protein in the absence of induction but did not provide further details on quantitation. The lack of data to support its efficient application in expressing transgenes except for tRNA has limited its application.

In *T. brucei,* it is a common occurrence to place transcription of exogenous genes under the control of bacteriophage T7-promoter. While this is not the only promoter sequence available, it is highly favoured for its small size, the fact that it is a foreign promoter, and for its strength. The T7-promoter is only 18 bp in length (rDNA promoter core is 290 bp (White et al., 1986)). A short promoter length makes it easy

to insert into a DNA construct. Using a promoter which is not recognized by endogenous polymerase has an extra advantage, as its activity can be controlled independently, without triggering processes that may affect endogenous promoters. Finally, the T7 promoter is one of the strongest promoters discovered. When very high transcript or protein levels is desired, a T7 promoter is preferred. Driving transcription from the T7 promoter requires expression of T7 RNA polymerase (T7RNAP).

To create a tetracycline inducible cell line in *T. congolense* (TcoSM), we developed a construct to integrate T7RNAP and TetR at the tubulin locus (Figure 7.1A).

Tubulin is essential and continuously transcribed by the cell. Integration at this locus therefore supports constitutive expression of T7RNAP and TetR. TcoSM was made by Bill Wickstead and is described here because it is the parental line used for most subsequent genetic modification.

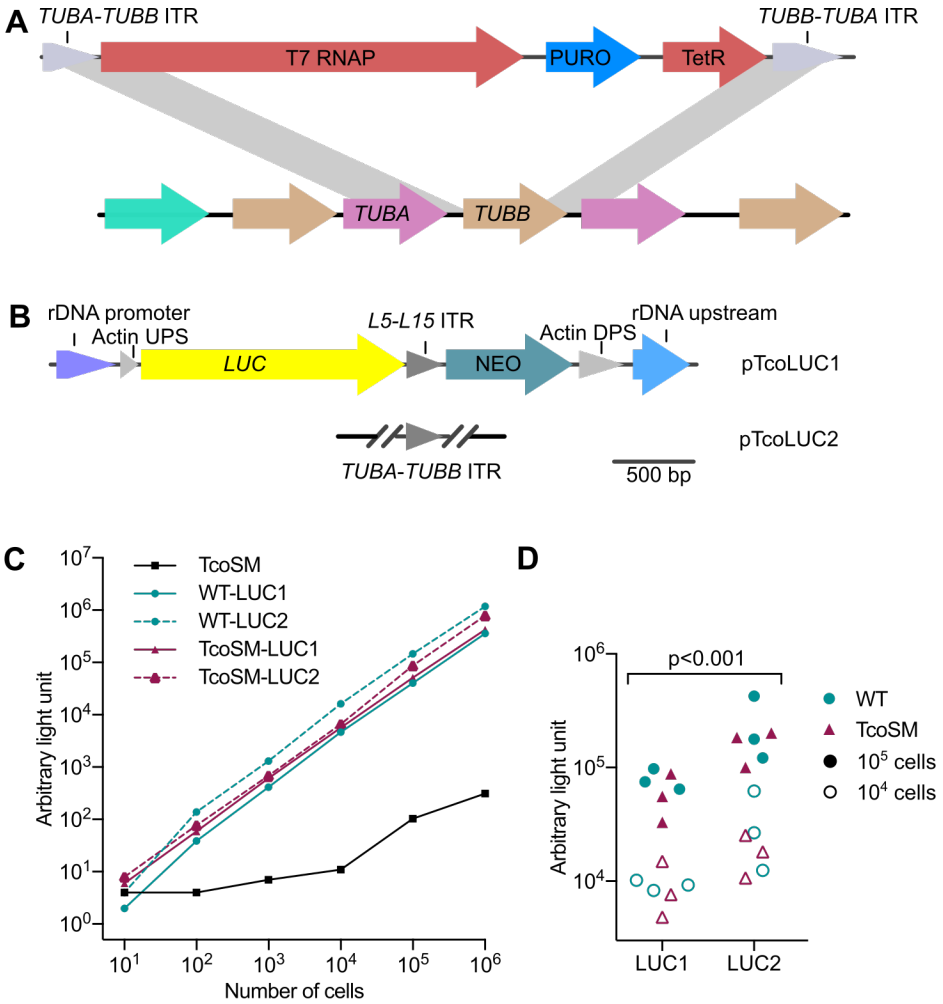


Figure 7.1: Transgenesis of IL3000. A. Schematic showing integration of a construct to constitutively express T7 RNA polymerase (T7RNAP) and tetracycline repressor protein (TetR) from the tubulin locus to generate cell line, TcoSM. ITR = intergenic region. **B.** Schematic of two luciferase constructs. The Ribosomal protein L5-L15 ITR in pTcoLUC1 is replaced by *TUBA-TUBB* ITR in pTcoLUC2. UPS/DPS = upstream/downstream processing sequences, respectively. **C.** Serial dilutions of transgenic strains were measured *in vitro* for bioluminescence (a.u.). Different strains are named for their parental line, Wildtype (WT) or single-marker (TcoSM) and the construct integrated, pTcoLUC1 (LUC1) or pTcoLUC2 (LUC2). Technical replicates from 3 independent clones of each line are shown along with mean and standard error across the clones. **D.** Bioluminescence in LUC1 verses LUC2 strains. The amount of arbitrary light unit per cell of LUC2 cells was higher than LUC1 cells. t-test was performed on log-transformed a.u. per cell.

7.1.2 Transcription of transgenes from rDNA promoters at endogenous loci

The rDNA promoter is a strong endogenous promoter. In *T. congolense,* the promoter has been used to successfully drive transcription of exogenous genes without integration into the genome (Downey and Donelson, 1999b, 1999a). The silent rDNA spacer region is also an important site for transgene integration in *T. brucei.* The rDNA spacer sequence in *T. congolense* is very different from the one in *T. brucei.* A failed

attempt to integrate into this region in *T. congolense* using *T. brucei* sequence to direct integration (Inoue et al., 2002) supports the fact that the two sequences bear little similarity to each other. Like most transgenesis attempted in *T. congolense*, the process was justified on the basis that the two "closely"-related species might possess enough similarity to allow recognition of sequences and homologous recombination to take place. This has previously worked for the tubulin array sequence, but the *T. brucei* rDNA spacer could not be used to derive stable transfectants (Downey and Donelson, 1999a). Replacing with *T. congolense*-specific rDNA sequence, however, allows correct integration (Inoue et al., 2002). Until now, this was the only report of integration and transgenesis at the rDNA spacer.

Two near-identical constructs (pTcoLUC1 and pTcoLUC2) were generated. They both integrate at the rDNA locus, and luciferase expression is driven by the endogenous rDNA promoter targeted. Where pTcoLUC1 and pTcoLUC2 differ is in the sequence downstream of the luciferase ORF, because alongside promoter strength, processing sequences downstream of a gene influence the level of protein produced: pTcoLUC1 contains the intergenic region (ITR) between the genes encoding ribosomal protein L5 (TcIL3000_9_6340) and ribosomal protein L15 (TcIL3000.9.6540); whereas pTcoLUC2 contains the *TUBA*-*TUBB* ITR (Figure 7.1B). The L5-L15 ITR was selected because ribosomal proteins are highly expressed. Both pTcoLUC1 and pTcoLUC2 constructs described above were made by Sarah Whipple from the Gadelha lab (University of Nottingham).

I transfected both constructs (pTcoLUC1 and 2) into either wildtype IL3000 (WT) or the single marker line (TcoSM); 4 transfections in total. I obtained transformants for all transfections. To test for the expression of luciferase by the rDNA promoter, I assayed *in vitro* luciferase activity by measuring bioluminescence released as a result of oxidization of luciferin. Luciferase expression in all instances was >1000-fold above background (Figure 7.1C). Amount of luciferase expressed in these lines is similar to what was reported in luciferase-expressing *T. vivax* cells (D'Archivio et al., 2013). The *T. vivax* luciferase line also has luciferase integrated at the rDNA locus and under the control of an rDNA promoter. Linearity of luciferase expression occurred over at least

5 log-fold of magnitude which is about a fold higher than reported in *T. vivax* (>3 log-fold). Within this range, luciferase level gives a good estimate of the number of cells present. The *in vitro* assay has a limit of detection of ~100 cells, about 3x lower than *T. vivax,* suggesting that *T. congolense* luciferase expression compares favourably with the published *T. vivax* line.

To examine the influence of downstream processing sequence (DPS) on protein expression level, I compared *in vitro* expression of luciferase between the two classes of transgenic lines (LUC1 and LUC2), Figure 7.1D. I found *TUBA*-*TUBB* ITR (LUC2) gives 2x higher protein levels than L5-L15 ITR (LUC1), implying that DPS significantly influence the level of protein produced. This is worth considering in the design of constructs for transgene expression in *T. congolense.*

7.2 Genetically-modified *T. congolense* is infective to mice

The ability of the lab-adapted *T. congolense* IL3000 cells to infect animals is not known. Long-term *in vitro* culture alters the biology of cells and may result in loss of infectivity or altered pathogenicity of parasites, which limits the use of the system for *in vivo* studies. This is true for *T. brucei*, where the monomorphic Lister 427 strain has altered pathogenicity and no longer produces the expected disease presentation in mice. To test whether our transgenic *T. congolense* is infective, we collaborated with Gavin Wright and Delphine Autheman (Wellcome Sanger Institute). Data presented in this section were generated in their lab and are presented here as they relate to the utility of modifications made for work *in vivo.*

Three female BALB/c mice were initially infected with 10^6 parasites of WT-LUC2 per mouse by the intraperitoneal route. An initial wave of parasitaemia was observed at day 4 post-infection. This was followed by a second wave of parasitaemia a week later. The waves of parasitaemia observed are characteristic of active infection. From the infected mice, 10^4 parasites were transferred into 13 new recipient mice using IV injections which also established infection. Images of infection in 4 of the mice are presented in Figure 7.2.

While the above data show that WT-LUC2 is able to infect mice, we did not infect mice with TcoSM-LUC2. However, TcoSM was derived from IL3000 and if one round of genetic modification did not prevent infectivity, it is unlikely that two modifications would. If the argument is made that WT-LUC2 is likely to retain infectivity because it is directly descended from the wildtype strain, whereas TcoSM-LUC2 is a derivative of a modified strain, the definition of "wildtype" must be considered. Wildtype as used here refers to a strain in which no modification has been made to the genome. In the strict sense, a strain that has been passaged *in vitro* over years and become lab-adapted most likely show some deviation from true wildtype. If the process of lab adaptation did not inhibit animal infectivity of the strain, it is likely that TcoSM- derivatives like TcoSM-LUC2 would retain infectivity.

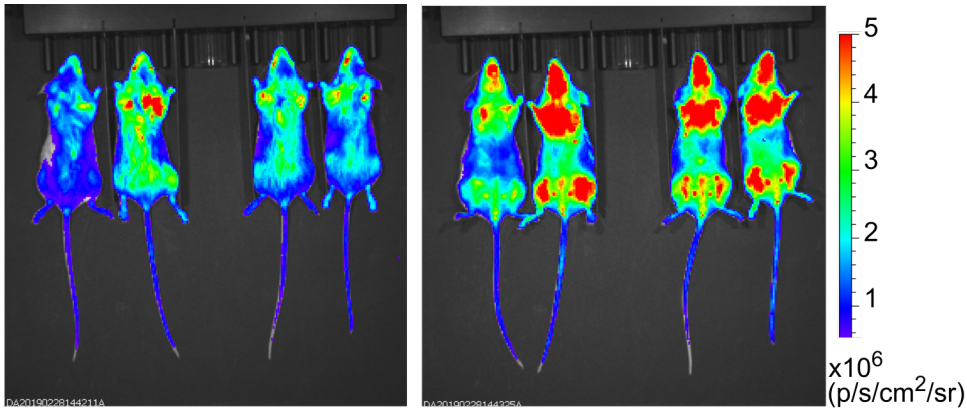


Figure 7.2: Genetic modification of IL3000 does not prevent infectivity. Bioluminescence imaging of ventral and dorsal views of four mice infected with WT-LUC2 at day 6 of post-infection. Data provided by Gavin Wright and Delphine Autheman (Wellcome Sanger institute).

7.3 Creating tools for inducible expression

7.3.1 Selection of potentially silent loci for regulated transgenesis

Actively-transcribed sites are generally not optimal for regulatable transgene expression, as integration at these sites present the risk of disrupting endogenous transcription. Inducible transgenesis from actively-transcribed loci has the added likelihood of ineffective down-regulation, as readthrough may occur even in the non-induced state. Therefore, integration at a transcriptionally silent locus, in the absence of native transcription, enables better control. Minichromosomes are a potentially

good site for transgenesis. No active transcription occurs in this part of the *T. brucei* genome, and it is possible that minichromosomes are also silent in *T. congolense.* Minichromosomes in *T. brucei* are also stable during mitosis, replicating with fidelity (Wickstead et al., 2003a). To date these aspects of chromosome biology have not been tested in *T. congolense.*

To investigate potential loci for transgenesis, I selected 3 sites on *T. congolense* minichromosomes. These sites were:

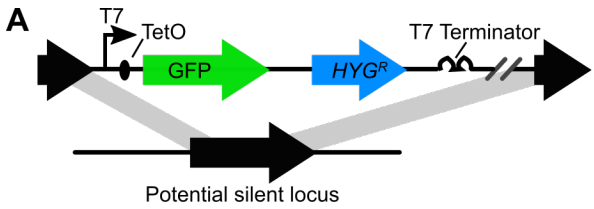
1. **A minichromosomal *VSG*** – *T. brucei* maintains a reservoir of inactive *VSG* on the minichromosomes that are used in antigenic variation. From the recent PacBio assembly of the *T. congolense* IL3000 genome (Abbas et al., 2018). I selected an apparently single *VSG* found in a sub-telomeric location on the minichromosome contig, called the *silent VSG* herein.
2. **369 bp repeat sequence** is a core of tandem repeats homologous to the *T. brucei* 177 bp repeat. There are ~10,000 copies of the repeat sequence which provides a large copy number for integration sites. The homologous 177 bp repeat has been shown to be a good site for integration and regulation (Wickstead et al., 2002).
3. **Family of genes encoding putative DEAH helicase** – *T. congolense* minichromosomes contain a family of ~20 copies of putative *DEAH* helicase sequences located at the subtelomeres. This occurrence is unique; the same is not seen on *T. brucei* minichromosomes. With no orthologue to compare to, we know nothing about the genes on minichromosomes predicted to encode DEAH helicases (herein called *DEAH* helicases). While their location on subtelomeres favours the assumption that they are not transcribed, the subtelomeres in *T. congolense* have not yet been shown to be silent. The minichromosomal loci predicted to encode DEAH helicases (herein referred to as *DEAH* loci) were selected to include an intermediate copy number of loci and also to investigate transcription status of the site.

7.3.2 All minichromosomal loci support transgenesis

To investigate loci for transgenesis, I targeted an inducible T7-driven GFP-reporter construct (Figure 7.3A) to all three selected loci listed above. The construct was designed to express GFP and Hygromycin marker under the control of a single tetracycline-inducible T7 promoter. The presence of a family of putative *DEAH* helicases at the subtelomeres of minichromosomes is not seen *in T. brucei.* Moreover, the transcriptional status of *T. congolense* subtelomeres is unknown. To investigate the possibility of active transcription of *DEAH* helicases at the subtelomeres, the construct was integrated in both the sense (*DEAH*-F) and antisense (*DEAH*-R) direction at a *DEAH* locus. As a positive control cell line, a GFP-expression construct lacking a promoter, targeted to the *TUB* locus (*GFP::TUB*), was also transfected. The latter, without a promoter, relied on RNA Pol II readthrough for expression of GFP.

Positive transformants were selected by growing cells in the presence of induction and drug selection (1 µg/ml tetracycline, 0.5 µg/ml hygromycin) for 10 days. All targeted sites yielded positive transformants showing that loci on the minichromosomes in *T. congolense* support transgene integration and transcription. However, efficiency of transfection varied between loci (Figure 7.3B). Targeting the silent *VSG* and 369 bp repeat gave ~1000 independent clones per transfection which was indistinguishable from efficiencies at endogenous pol II transcribed region.

A



B

Target sequence	Copy number	Transformation efficiency (90% confidence interval)
Silent VSG	1	540 – 982
DEAH-F	~20	37 – 89
DEAH-R	~20	33 – 83
369bp repeat	~10 000	692 – 1188

Figure 7.3: Minichromosomal loci vary in efficiency for transgenesis. A. Schematic of the inducible GFP reporter used for integration at potential silent loci on the minichromosome. **B.** A table of selected minichromosomal loci showing the number of copies present and the number of independent clones from targeting each site with the same amount of DNA. silent VSG = inactive VSG, DEAH = putative DEAH RNA helicase loci with construct integrated in either the sense (F) or antisense direction (R).

Surprisingly, the *DEAH* loci was targeted with ~10-fold lower efficiency than the silent *VSG* locus. It did not matter in what direction the construct was integrated: both sense and antisense integrations gave the same efficiency. To confirm that the low efficiency when targeting *DEAH* loci was not due to an experimental error, I repeated transfection twice with a modified construct. I removed TetO from the construct targeting *DEAH* in the sense direction to generate a construct for T7-driven constitutive expression of GFP. Two independent transfections with the modified construct gave similar efficiencies (0-150 independent clones) to *DEAH-F* and *DEAH-R.* This shows that transfection efficiency when targeting *DEAH* loci in *T. congolense* is low. A possible explanation could be that they are active loci and do not easily allow integration as that could disrupt endogenous transcription. Or integration was successful but readthrough across endogenous genes may have occurred from the T7-driven transcription to produce toxic proteins that killed transformants. However, the latter explanation seems unlikely for two reasons: First, when designing the construct, a T7 terminator was incorporated downstream of the *HYG* marker to stop the promoter from transcribing endogenous genes. Secondly, even if high level of transcription from the T7 promoter managed to proceed beyond the terminator, one would expect that toxic products would only be produced where readthrough is in the

same direction as the genes, and in the opposite direction no readthrough would occur to produce toxic proteins. Therefore, only one of the constructs should have a low transfection efficiency. However, both *DEAH*-F and -R show the same outcome of poor transfection efficiency.

Recording the same transfection efficiency by targeting one single locus (silent *VSG*) and ~10,000 loci (369 bp repeat) suggest here that, as for *T. brucei* (Wickstead et al., 2003b), transformation efficiency is independent of target copy number.

7.3.3 Minichromosomal loci support regulatable expression

Transformants derived from all three target sites above showed detectable GFP expression when cultured in the presence of tetracycline and assessed by native fluorescence microscopy (Figure 7.4A). *GFP::TUB* control also displayed detectable fluorescence, as expected.

To test regulation, cellular green fluorescence level, with or without tetracycline, were measured by flow cytometry. At least 2 randomly-selected clones and also the population of transformants were tested. The population was included to detect any disparity between the general behaviour of the population and individual clones. *GFP::TUB* was included in the analysis. Figure 7.4B shows the relative levels of green fluorescence in cells in which different loci were targeted. Median fluorescence of parental cells (background, no GFP) was ~250 arbitrary units (a.u.). GFP expression by RNA Pol II readthrough from the tubulin locus (*GFP::TUB*) raised fluorescence levels by ~3-fold above background.

In the absence of tetracycline, all clones in 2 of the 3 loci (silent *VSG* and 369 bp repeat) were indistinguishable from background (Figure 7.4B). This is not to say that GFP is not expressed when the system is down-regulated (OFF state), but rather that any transcription taking place cannot be measured above background. As described in Methods (Figure 3.4), using glutaraldehyde in fixation increases background fluorescence level. This will also explain why no variation can be seen between the different clones in the down-regulated (OFF) state.

Addition of tetracycline caused > 15-fold increase in median fluorescence above background from all three sites (Figure 7.4B).

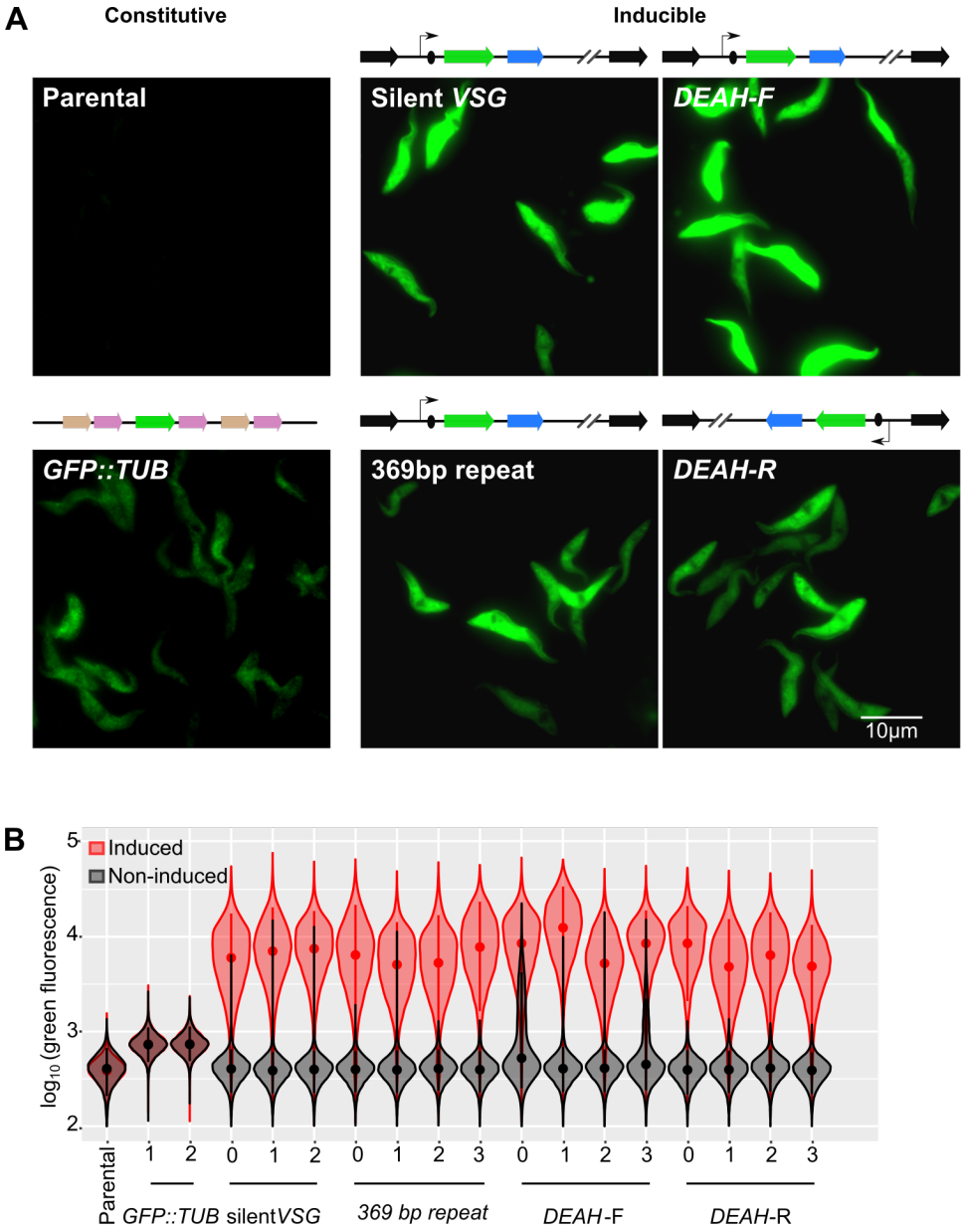


Figure 7.4: Minichromosomal loci support regulated transgene expression. A. Micrograph showing green fluorescence of different cell lines. Fluorescence from the parental (untransformed) line and constitutive expression by Pol II from the tubulin locus (*GFP::TUB*) are also shown. All images captured with same settings, but inducible constructs are shown at 5x less brightness of other lines to reduce saturation of signal. **B.** Violin plot of green fluorescence levels of cells with or without 1 µg/ml tetracycline. > 25,000 cells measured per sample. Dots indicate median fluorescence; lines are 95 % confidence interval. Numbers on the x-axis refer to different clones of the same line with 0 being the population

Median fluorescence of the lowest T7-driven GFP expression is ~5-fold above that of RNA Pol II, in agreement with T7 promoter driving higher levels at transcription. No

evidence of fluorescence level variation between clones was observed when targeting the silent *VSG* or the 369 bp repeat locus. Clone-to-clone variation was observed when targeting the minichromosomal 177 bp repeat in *T. brucei* (Wickstead et al., 2002). Since transfection efficiency is the same when targeting either silent *VSG* or 369 bp repeat, the silent *VSG* was preferred as the integration loci for subsequent integrations.

When targeting *DEAH* in the sense direction (*DEAH*-F), 1 of 3 clones produced green fluorescence at a level considerably above background without induction. The population (0) shows an even greater (~2-fold) increase, suggesting that this may be true for more transformants. When targeted in the anti-sense direction (*DEAH*-R) however, the down-regulated state is within background level for all clones and the population. This behaviour suggests that active transcription occurs at some of the *DEAH* loci. In the tetracycline-induced state, there is also some variation between clones, with at least a 4-fold fluorescence difference between the highest (DEAH-F #1) and lowest expression levels (DEAH-F #2) (Figure 7.4B).

While clonal variation and transcription leakiness are not desirable features of a 'good' integration site for transgene expression, integration at *DEAH*-F gives some of the highest GFP protein level observed here. If the specific *DEAH*-F locus that produced such phenotype were to be characterised, it may prove itself a useful one for applications that require high protein levels.

7.4 Transgenes at the minichromosomal loci are stable with respect to mitotic growth

Minichromosomes in *T. brucei* are segregated with high fidelity. The rate of loss is estimated at < 1 loss per 1000 cell generations (Wickstead et al., 2003a). Their stability during mitosis makes the minichromosomes preferable sites for integration. I therefore aimed to determine the stability of *T. congolense* minichromosomes carrying transgenic genes. Clones and populations targeting the three different sites (plus control) were cultured with tetracycline with selection or without hygromycin selection for 3 weeks (~42 generations) before measuring green fluorescence by flow

cytometry. Keeping hygromycin selective pressure allows only resistant cells (and, by extension, GFP-expressing cells) to survive. When selective pressure is removed, if minichromosome inheritance is unstable, then cells not expressing GFP may build up in the population, which will in turn reduce cell median fluorescence detected by flow cytometry. These data are presented in Figure 7.5. Median fluorescence was unchanged in the absence of selection at the silent *VSG* and *DEAH* loci. Constructs integrated at the 369 bp repeat loci showed at least 1 of 4 clones having ~20 % decrease in median fluorescence level in the absence of hygromycin selection, indicating that a subset of cells did not inherit the minichromosome carrying the *HYG* marker and *GFP*. This is equivalent to ~0.5 % loss in median fluorescence intensity per generation.

Drawing from all the data presented in this and the previous sections, it can be concluded that the silent *VSG* and 369 bp repeat are suitable sites for integration, giving reasonable transfection efficiency and support of stable and regulatable transgenesis.

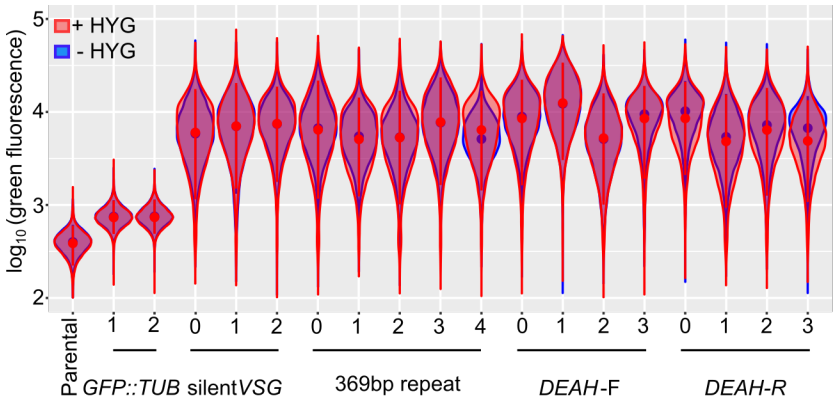


Figure 7.5: Minichromosome transgenic loci are stable with respect to mitotic growth. Violin plot of green fluorescence levels of cells cultured in the presence of 1 µg/ml tetracycline with and without drug selection (HYG) over 3 weeks (~42 generations). Dots indicate median fluorescence; lines are 95 % confidence interval.

8 RESULTS 3

8.1 Design and use of *T. congolense* RNAi constructs

8.1.1 Design of constructs for inducible RNA interference in *T. congolense*

In the last two decades, reverse genetic studies in *T. brucei* have been helped by the ability to use endogenous RNA interference (RNAi) to alter mRNA levels for target genes. RNAi approaches to genetic studies in trypanosomes require introduction of a fragment of gene homologous to the gene of interest. Constructs are either designed to enable transcription in both directions by using opposing promoters to create double-stranded RNA (dsRNA) or by placing a stuffer between a sense and antisense sequence of the fragment to create a stem-loop structure post-transcription. The presence of dsRNA triggers post-transcriptional gene silencing of the gene of interest. Changes observed when a gene is depleted may give insight into their function within the cell.

To investigate RNAi in *T. congolense* IL3000 cells we designed and created RNAi constructs that combine parts of *T. brucei* plasmids with *T. congolense* DNA (Figure 8.1A). The design of RNAi constructs was informed by a number of factors.

Integration sequences – From the integration loci tested in Section 7.3, the *silent VSG* showed high transfection efficiency, efficient regulation of transgenes and little variation between clones. The same integration sequence was incorporated in the design of all RNAi plasmids made in this section.

RNAi cassette – The RNAi cassette consists primarily of two opposing T7 promoters between which the fragment of the target gene is inserted. Again, data presented in Section 7.3 and reports from literature (Coustou et al., 2010, 2012) prove the T7 promoter works in *T. congolense.* I have shown that this promoter expresses at least 5-fold above endogenous Pol II (Section 7.3). So I selected the T7 promoter to drive transcription. The promoter in the RNAi cassette was taken directly from the *T. brucei* p2T7 plasmid which has the arrangement of an RNAi fragment placed between two

opposing T7 promoters. There is therefore no *T. brucei* gene fragment in this region. I then placed either a fragment of the sequence encoding GFP or a *T. congolense* gene fragment between the opposing promoters.

Drug cassette – The Wickstead and Gadelha labs have performed several endogenous locus tagging in bloodstream form *T. congolense.* In these tests, and from my own observations, using *NEO* for selection gives slightly better (~2x) transfection efficiency than *HYG* (personal communication). The reasons behind this are unclear, but as efficiency is important for downstream application, I chose to use *NEO* for selection of transformants.

UPS/DPS – The selection of upstream and downstream processing sequences (UPS/DPS) once again is based on the outcome of the endo-tag in Section 7, it is clear that actin UPS/DPS from *T. brucei* works sufficiently to select transformants with either Pol II or T7 transcription. Admittedly, the processing activity of *T. brucei* UPS/DPS may not be as efficient in *T. congolense* as *T. congolense* UPS/DPS, but since proven I chose to use that of *T. brucei* in my constructs.

Promoter for selection – The most convenient promoter to use would be the endogenous Pol I rDNA promoter from *T. brucei.* It would be useful had this promoter worked in *T. congolense* because the system has already been set up to use it in inducible expression and could easily be adapted. However, the *T. brucei* rDNA promoter is very different to the equivalent *T. congolense* promoter. In fact, the two promoter sequences are so different they cannot be meaningfully aligned. Attempts to integrate sequences to *T. congolense* rDNA promoter using *T. brucei* rDNA failed (Inoue et al., 2002), likely due to integration problems given the different sequences rather than function. As such we do not know if the *T. brucei* rDNA promoter is functional in *T. congolense.* So, I tested 3 promoter configurations:

- *T. brucei* rDNA
- *T. congolense* rDNA
- T7

Terminator – The terminator sequence for transcription of rDNA has not been defined in *T. congolense* so we do not know where transcription stops. I included a configuration with a T7 promoter which is terminated by a downstream T7 terminator in the drug cassette.

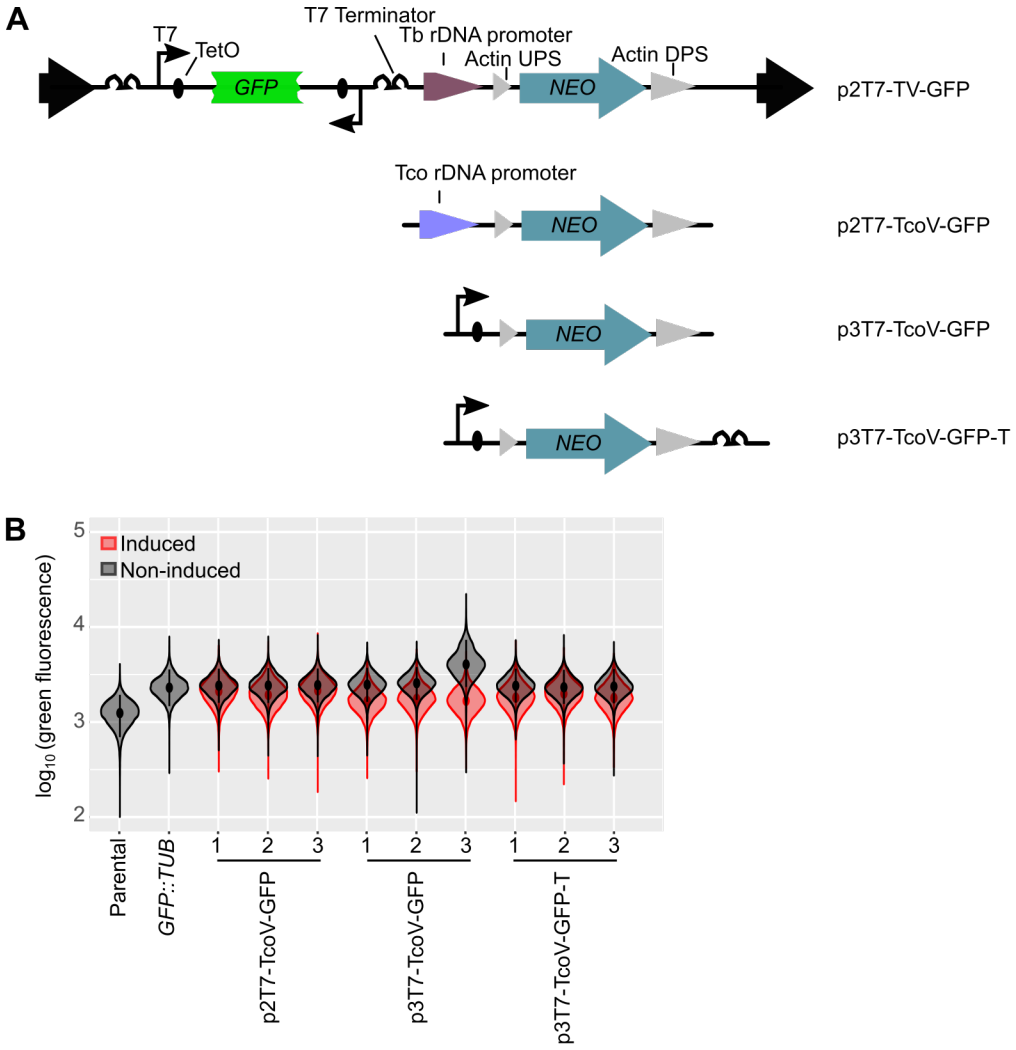


Figure 8.1: Design and testing of different RNAi constructs in *T. congolense*. A. Schematics of 4 RNAi constructs designed with different arrangement of drug cassette transfected into *T. congolense*. Constructs have the same arrangement of GFP cassette. **B.** Violin plot of green fluorescence of *T. congolense* cells; untagged (parental), GFP constitutively expressed at the tubulin locus (GFP::TUB) and RNAi cells with or without 1 µg/ml tetracycline. > 25,000 cells per sample measured by flow cytometry. Dots indicate median fluorescence; lines indicate 5 to 95 percentile. Numbers on x-axis refer to different clones.

Transfecting the *GFP* RNAi construct where selection is driven by *T. brucei* rDNA promoter (p2T7-TV-GFP), did not generate stable transformants. The same construct was used to target RNAi to α-tubulin, clathrin heavy chain, putative expressed VSG. Transfection of all three constructs also gave no transformants. In contrast,

transfection of the other 3 RNAi constructs (p2T7-TcoV-GFP, p3T7-TcoV-GFP and p3T7-TcoV-GFP-T) yielded >2000 positive transformants in each instance. As expected, the number of transformants were at least 2x higher with *NEO* selection marker than with *HYG* marker (1000 positive transformants) as was seen with the GFP expression construct in Section 7.3 where integration is also targeted to the *silent VSG* locus.

Given the transfection efficiencies when targeting the *silent VSG* locus (Section 7.3) and data from other RNAi plasmids, the failure to obtain transformants when selection is driven by *T. brucei* rDNA promoter strongly suggests that *T. brucei* rDNA promoter is not functional in *T. congolense.* As seen previously, both T7 and *T. congolense* rDNA promoter function at the *silent VSG* locus.

8.1.2 RNAi depletes exogenous GFP in *T. congolense*

To test whether the integrated RNAi constructs can cause depletion of exogenous GFP expressed in *T. congolense* cells, I selected 3 transformants each from the RNAi transfection that gave transformants (p2T7-TcoV-GFP, p3T7-TcoV-GFP and p3T7-TcoV-GFP-T), induced by adding 1 µg/ml of tetracycline for 72 h and measured green fluorescence of the induced and non-induced lines by flow cytometry.

In the absence of an RNAi construct (GFP::TUB), the level of green fluorescence level is ~1500 (Figure 8.1B) which is only 500 units above background (~1000) in this experiment. At 0 h, the levels of green fluorescence in all but one of the RNAi cell lines are consistent with the level in GFP::TUB showing no evidence of leakiness. Clone 3 of p3T7-TcoV-GFP RNAi lines have a higher level of green fluorescence without induction (~2500) suggesting a possible duplication of the *GFP*. After 72 h of induction, RNAi against GFP directed with the three different regimes on average reduced green fluorescence to about 1200.

RNAi lines transfected with p3T7-TcoV-GFP on average gave a slightly greater reduction in green fluorescence on induction although there is no evidence that they are any different from the other RNAi lines. Clone 3 of p3T7-TcoV-GFP expressed about twice the level of green fluorescence expected in a *GFP::TUB* line. Interestingly,

fluorescence level in the induced line still came down to the same level as the other clones suggesting that there is a minimum level to which knockdown can be achieved and that is independent of the expression levels and therefore highly expressed genes might have a higher level of knockdown.

These outcomes mean that all RNAi constructs are similar in depleting levels of transgenes. Because p3T7-TcoV-GFP gave a marginally higher average reduction, this construct was used downstream.

8.2 Relative level of RNAi-targeting appears independent of expression level

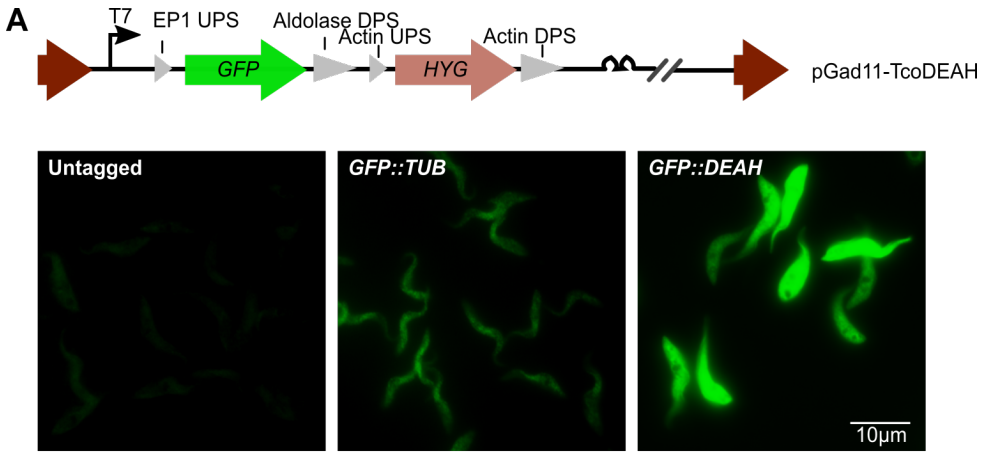
As RNAi is a templated process wherein product from degraded mRNA can act as template for further knockdown, it is unclear how knockdown responds to expression level. Three responses are possible:

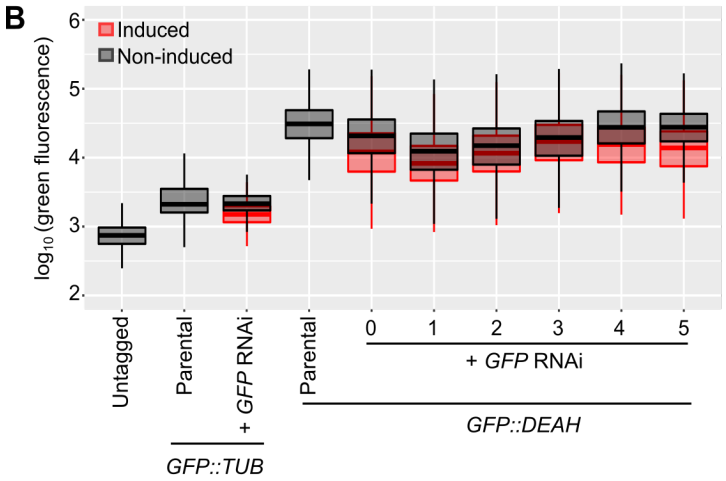
1. High expression of the target mRNA leads to a low proportion of mRNA removed as new mRNA saturates the RNAi system.
2. Same proportion of target mRNA is removed regardless of the amount of mRNA at the start because the RNAi system reaches equilibrium.
3. High proportion of mRNA is removed due to an increased available template. In this case, the more of the target mRNA present, the greater the level of knockdown.

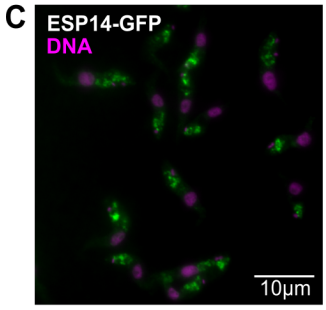
One clone from Section 8.1 suggested that point 3 might be true. To test whether RNAi-targeting is dependent on expression level, I increased the level of cytoplasmic GFP expressed in the parental cell line and targeted RNAi in a low and high-expressed GFP line.

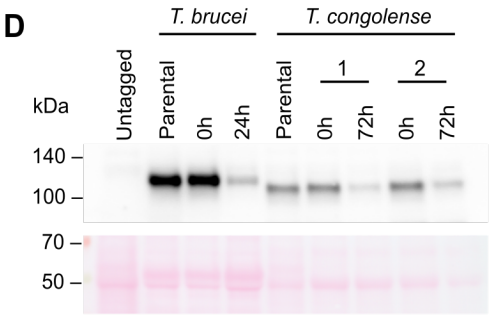
Data from Section 7.3 shows that locus encoding putative DEAH helicases on the minichromosomes (*DEAH* locus) can be used for high level expression of exogenous genes. Therefore, to generate a cell line that expresses a higher level of cytoplasmic GFP, I constitutively expressed GFP from the *DEAH* locus under control of a T7 promoter by creating the construct pGad11-TcoDEAH (Figure 8.2A) and transfecting into IL3000 single marker line.

After integrating the RNAi construct p3T7-TcoV-GFP into the parental expressing GFP from the *DEAH* locus (*GFP::DEAH*), I induced RNAi for 72 h and measured green fluorescence of induced and non-induced cells by flow cytometry. *GFP::DEAH* gave ~13x more fluorescence than *GFP::TUB* in the absence of an RNAi construct (Figure 8.2B). The introduction of RNAi construct into *GFP::DEAH* decreased fluorescence level of non-induced cells to 11500 (10000-12000) suggesting that the RNAi construct may be leaky in these lines. Additionally, it suggests that there is an epigenetic factor that causes clones arising from the parental to grow out into cells with different distribution of fluorescence from the parental and each other. This presents as variation in mean fluorescent levels between clones even though they were genetically identical at the start (see also *GFP::DEAH* microscopy – Figure 8.2A). RNAi-targeted depletion of *GFP* in *GFP::TUB* decreased levels ~40 % (from mean fluorescence of 1600 to 1000) which is similar to previous observation (see Section 8.1). I observed ~50 % reduction of green fluorescence in the population (clone 0) and 3 out of 5 clones tested and ≤ 10 % decrease in 2 of 5 clones.


A
T7
EP1 UPS
Aldolase DPS
Actin UPS
Actin DPS
GFP
HYG
pGad11-TcoDEAH
Untagged
GFP::TUB
GFP::DEAH
10µm


B
Induced
Non-induced
log10 (green fluorescence)
6
5
4
3
2
Untagged
Parental
+ GFP RNAi
GFP::TUB
Parental
0
1
2
3
4
5
+ GFP RNAi
GFP::DEAH


C
ESP14-GFP
DNA
10µm


D
T. brucei
T. congolense
kDa
Untagged
Parental
0h
24h
Parental
1
2
0h
72h
0h
72h
140
100
70
50

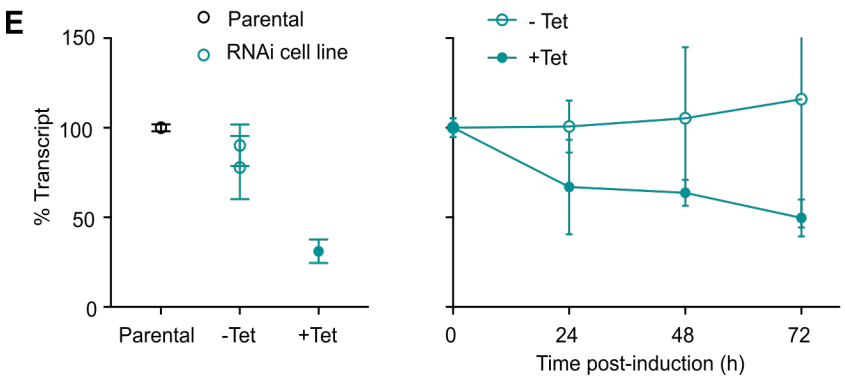

E
Parental
RNAi cell line
% Transcript
150
100
50
0
Parental
-Tet
+Tet
- Tet
+Tet
0
24
48
72
Time post-induction (h)

Figure 8.2: RNA interference reduces both mRNA and protein level in bloodstream form *T. congolense.* **A.** Schematic of construct (pGad11-TcoDEAH) to constitutively express GFP from the putative DEAH helicase loci on the minichromosomes (*DEAH* loci). Fluorescence micrographs of *GFP::TUB* cell line and *GFP::DEAH* (generated from transfection of pGad11-TcoDEAH) lines expressing different levels of cytoplasmic GFP. **B.** Tukey box plot of green fluorescence in T. congolense RNAi lines with or without 1 µg/ml tetracycline (> 25,000 cells per sample measured by flow cytometry). Horizontal lines indicate median fluorescence; box and vertical lines indicate represent quartiles and 5 to 95 percentile respectively. Numbers on the x-axis refer to different clones of line with 0 being the population. **C.** Micrograph showing native fluorescence of GFP-tagged ESP14, cells are counterstained with 4′,6-diamidino-2-phenylindole (DAPI) to show DNA. **D.** Western blot detecting GFP using mouse anti-GFP (2x monoclonal) at different hours post-induction. **E.** Plots show Pyruvate phosphate dikinase (PPDK) mRNA abundance at different hours post-induction of RNAi. Data in (E) provided by Pieter Steketee and Liam Morrison.

These data reveal that what determines the level of RNAi effect in *T. congolense* is more complex than just protein levels. At least when driving expression from the T7 promoter it appears the amount of protein produced is higher than pol II-driven expression, but it does not result in a strong RNAi effect. However, with Pol II-driven expression it did seem that when there are higher transcript levels RNAi could knockdown protein levels proportionally further (Section 8.1.2).

8.3 RNAi reduces both protein and mRNA levels of non-essential genes in *T. congolense*

8.3.1 RNAi-induced depletion reduces level of protein

Data presented this far has focused on RNAi-mediated depletion of exogenous gene. To test whether RNAi-targeting also caused reduction of endogenous protein level in *T. congolense,* we targeted RNAi to genes likely to be non-essential in *T. congolense*. We then assessed protein and target mRNA levels to investigate level of depletion.

ESP14 (Tb927.7.470) is a membrane-associated protein that localises to the flagellar pocket in *T. brucei* (Gadelha et al., 2015). Knockdown of *T. brucei ESP14* does not produce any observable defect in cells even though levels are highly reduced by ≥ 80 % at 24 h of induction (Whipple, 2017). I targeted RNAi to the orthologue of *ESP14* in *T. congolense* (TcIL3000_7_180). Sarah Whipple tagged the *T. congolense ESP14* with GFP at the C-terminal end (Figure 8.2C). This *T. congolense ESP14-GFP* line was kindly donated by Catarina Gadelha and was used as the parental line for transfection of the RNAi construct against *ESP14*. Levels of protein were monitored using antibodies

against GFP. I compared depletion in *T. congolense* with a similar line generated in *T. brucei* where ESP14 is tagged at the N-terminal end with GFP (*T. brucei* cell line was provided by Sarah Whipple). Level of GFP-ESP14 is 3x higher in intensity in *T. brucei* (Figure 8.2D). After 72 h induction, intensity of GFP detection in *T. congolense* cells decreased to the same level as 24 h induction in *T. brucei*. This means RNAi can effectively reduce endogenous protein levels in *T. congolense.*

8.3.2 RNAi-induction reduces mRNA levels of endogenous gene in *T. congolense*

The primary target of double-stranded RNA is the mRNA transcript of genes that bear sequence homology to the RNAi fragment. To test the levels of reduction of transcripts, our collaborators Pieter Steketee and Liam Morrison (Roslin Institute), induced depletion of the glycolytic enzyme, Pyruvate phosphate dikinase (PPDK) by RNAi and measured the amount of *PPDK* transcript present. Transcript level dropped to ~50 % of the non-induced by 24 h and stayed at the same level for subsequent time points up to 72 h (Figure 8.2E). This amount of reduction in transcript level agrees with the reduction seen at the protein level (Section 8.1).

RNAi-mediated reduction of transcript and protein levels in *T. congolense* appears lower than *T. brucei* where usually 80-90 % reduction is seen within 24 h of induction.

8.4 RNAi directed against transcripts of clathrin heavy chain and tubulin phenocopies *T. brucei*

Data presented in Section 8.2 show that RNAi can be used to reduce both protein and mRNA levels but also suggest that there may be a lower penetrance of RNAi in *T. congolense* compared to *T. brucei.* Knockdown reduces both protein and mRNA levels of non-essential genes to ~50 %. Previous reports have shown a defect when targeting tubulin in *T. congolense* (Coustou et al., 2010; Inoue et al., 2002). To test if the levels of RNAi-mediated depletion observed can cause similar defect, I targeted RNAi against genes encoding α-tubulin and clathrin heavy chain.

In the genome of both *T. brucei* and *T. congolense,* the genes encoding α- and β-tubulin are arrayed in a tandem repeat. Depletion of α-tubulin produces a phenotype

known as FAT which results in compromised integrity of the cytoskeleton and stalled cytokinesis (Ngo et al., 1998). Affected cells become multinucleated and lose their normal slender morphology to take on a spherical shape. The FAT morphology was also observed in *T. congolense* when RNAi was targeted to α-tubulin (Coustou et al., 2010; Inoue et al., 2002).

Clathrin plays a major role in endocytosis. It is made up of 2 components – heavy and light chains. Knockdown of clathrin heavy chain in bloodstream form *T. brucei* results in enlargement of the flagellar pocket. This morphology is referred to as a BigEye (Allen et al., 2003). Induction of RNAi against α-tubulin and clathrin heavy chain phenocopy morphological defects seen in *T. brucei* (Figure 8.3A).

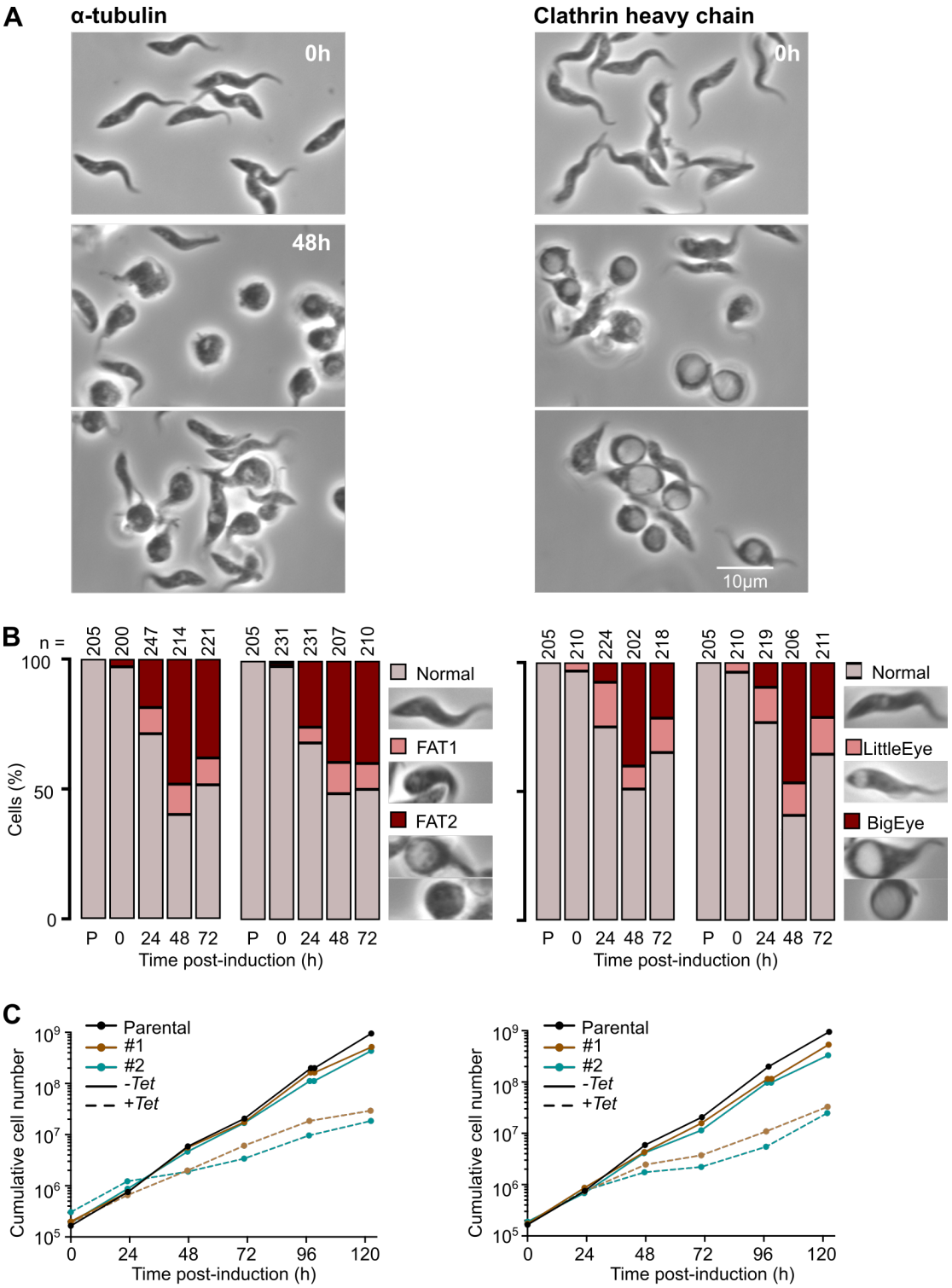


Figure 8.3: RNAi directed against essential genes produce defect in *T. congolense*.
A. Micrographs of RNAi cells with or without 1 µg/ml tetracycline. **B.** Quantification of different cell morphologies of two clones in parental, non-induced and induced cell lines, numbers above bars are total number of cells counted. **C.** Cumulative growth curve of *T. congolense* cells following RNAi-induction.

I observed the FAT and BigEye morphology in *T. congolense* RNAi cells. Without induction, ~3-4 % of cells had a detectable morphological defect. This may be due to leakiness of the RNAi or the presence of tetracycline in the serum used in the culture media. Following 24 h of induction, this number increased to ~30 % and progressed to 50-60 % at 48 h and did not increase at further time points.

The proportion of cells producing the defect is less in *T. congolense* than in *T. brucei.* More than 80 % of cells produce FAT morphology within 12 h of induction of RNAi against α-tubulin in *T. brucei* (LaCount et al., 2000). This again suggests RNAi penetrance may be lower in *T. congolense* than in *T. brucei*.

A clear growth defect in RNAi line was detectable beyond 24 h of induction *T. congolense*. At this point, the number of cells in culture becomes affected after two rounds of cell cycle as an increasing proportion of cells do not divide. In *T. brucei* growth defect was observable after 16 h of inducing RNAi against clathrin heavy chain (Allen et al., 2003). In *T. brucei* bloodstream cells where RNAi is targeted against tubulin and clathrin, cells decline after a relatively short time but recover after a few days around day 4 and the cells that grow out are refractory to RNAi. However, when targeting the same genes in *T. congolense* the cell numbers never decline but continue to grow but slower than non-RNAi cells over the period of 5 days observed. The data suggest that with every new generation of the *T. congolense* RNAi cells there is a proportion that is RNAi-refractory and a proportion that is affected by RNAi regardless of their parentage. Therefore, the progeny of RNAi-refractory cells may become sensitive to RNAi in the next generation.

8.5 RNAi targeted to KKIP1 and putative expressed VSG in *T. congolense* does not phenocopy *T. brucei*

To further test the utility of RNAi to investigate gene function, genes that are likely to be essential in *T. congolense,* but whose role has not previously been tested were targeted by RNAi.

KKIP1 is a kinetochore-associated protein. Depletion in *T. brucei* affects DNA segregation resulting in accumulation of cells with a large unsegregated nuclei. Onset of defect occurs very quickly such that within 24 h about 70% of cells present with large unsegregated nuclei (D'Archivio and Wickstead, 2017).

Variant surface glycoprotein (VSG) is the major surface glycoprotein of the bloodstream form trypanosome. In a strategy adapted to evade host immune system, trypanosomes switch their expressed VSG by deactivating the old and reactivating a new *silent VSG* or by a recombination event where a pseudo-VSG replaces the expressed *VSG* at the active expression site. Depletion of expressed *VSG* in *T. brucei* produces a characteristic arrest of cells at the pre-cytokinesis state leading to a build-up of cells with 2 kinetoplasts and 2 nuclei (Cheung et al., 2016; Sheader et al., 2005).

There are two *T. congolense* IL3000 lines – an *in vivo* line and an *in vitro* line. The VSG expressed in the *in vivo* line (TcIL3000_0_28420) is not the same as the one in the *in vitro* line. We do not know the actual VSG expressed in the IL3000 grown our lab. We took a *de novo* assembly of RNA-Seq data of the IL3000 line from Liam Morrison. It must be noted that the VSG expressed in our line may not even be present in the assembly. In this assembly, we identified a very high copy transcript that suggest it is the major expressed VSG in the line in Liam Morrison's lab, Roslin institute (Personal communication-Liam Morrison, Bill Wickstead). Because the IL3000 line in our lab was derived from the one in Liam Morrison's lab, at the time of doing this experiment it was believed that the identified VSG was the putative expressed VSG in the *T. congolense* IL3000 as grown in our lab (see also Section 10.2.6). The *de novo* RNA-Seq data is not yet published but data was made available by Liam Morrison.

I targeted RNAi to *T. congolense KKIP1* twice using two different RNAi fragments. I also targeted RNAi to the putative expressed VSG in our line based on the sequence from the *de novo* assembly of RNA-Seq data.

PCR assay confirmed all clones have integrated the RNAi construct at the correct locus (Figure 8.4). Surprisingly, RNAi induction against KKIP1 produced no detectable morphological or growth defect in cells after 72 h of induction. Given its role in *T. brucei,* it is very unlikely that *KKIP1* is non-essential in *T. congolense*. It is therefore

likely that KKIP1 depletion is insufficient to produce a strong morphological and detectable growth defect.

I identified that all selected clones of *VSG* RNAi lines showed no defect after 48 h of induction suggesting that cells have either switched their VSG or VSG levels do not drop low enough to cause defect. A repeat transfection and further testing of 10 more clones produced varying outcomes. At 72 h induction, 5 clones appeared to have some morphological defect, but this could not be repeated.

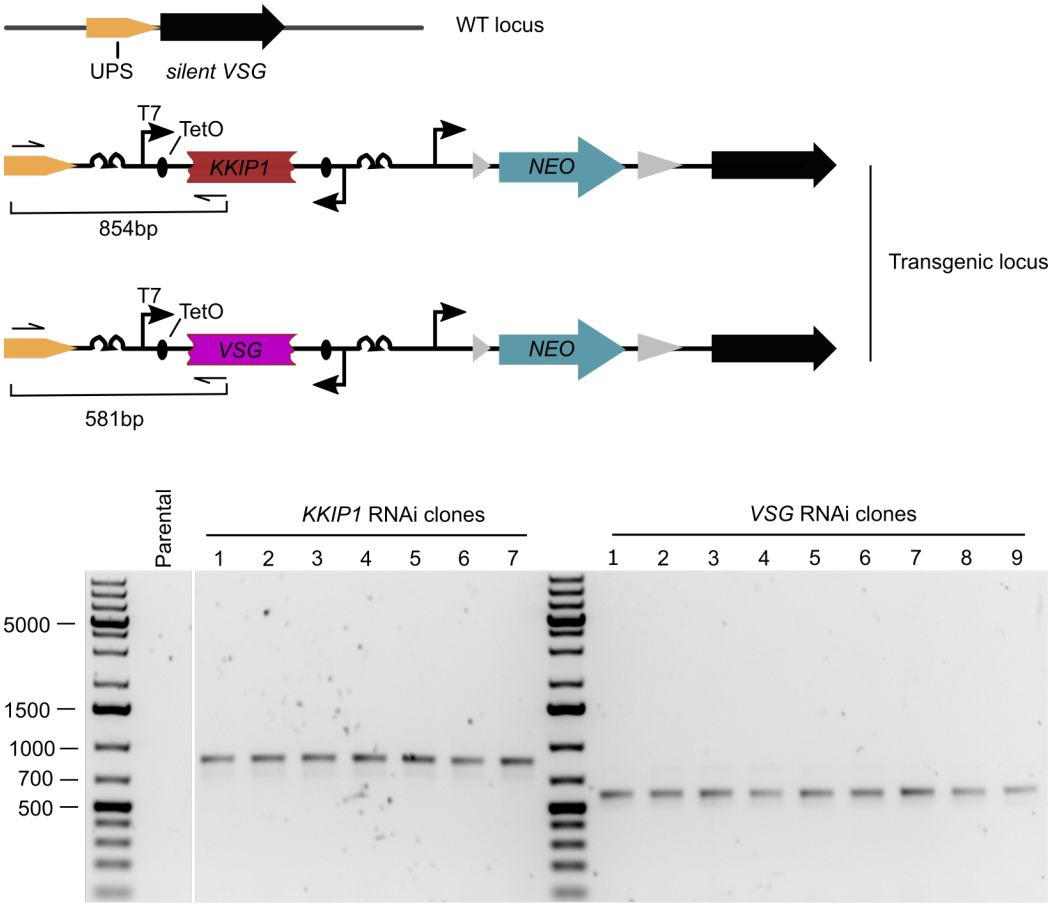


Figure 8.4: RNAi directed against KKIP1 and putatively expressed VSG does not phenocopy *T. brucei*. Schematic showing wildtype silent *VSG* locus (WT) and transgenic locus after integration of RNAi construct against *KKIP1* and putatively active *VSG*. UPS, upstream processing sequence. Arrows indicate PCR-primer-binding region. PCR assay used to test integration of RNAi construct in selected clones.

8.6 The proportion of cells that produce morphological defects decrease when RNAi lines are kept in continuous culture

In *T. brucei,* population of cells refractory to RNAi are known to grow out and take over the culture after a few days of induction (Allen et al., 2003; Wang et al., 2000).

To test stability of RNAi lines, I induced depletion of clathrin heavy chain in lines maintained non-induced in culture over different lengths of time (1 and 8 weeks) or kept frozen after a week in culture and then thawed (1 week-thawed). The thawed cells stayed in culture an additional 5 days before induction.

There was no difference between the number of cells with morphological defect in non- induced lines of 1 week, 8 weeks or 1 week-thawed (1.4-5.9 %; Figure 8.5).

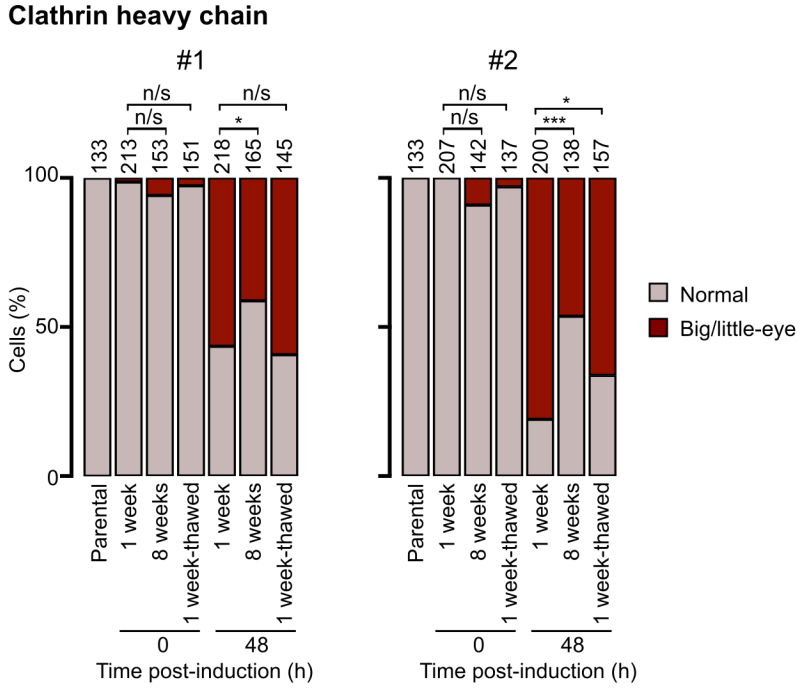


Figure 8.5: Continuous culture of *T. congolense* RNAi lines decreases proportion of cells that produce morphological defect on induction. Quantification of cell morphologies after 48 h RNAi induction against clathrin heavy chain. RNAi lines were maintained in culture for 1 week, 8 weeks or frozen after 1 week in culture and thawed. Cells stayed in culture additional 5 days before induction (1 week-thawed). p-value was calculated using Fisher's exact test with Bonferroni correction (n/s = non-significant, *= p<0.05, *** = p<0.0001)

After 48 h induction, the proportion of cells that produced morphological defect were not different between 1 week and 1 week-thawed lines (56-59 %). However, 8 weeks of continuous culture, showed a decrease of approximately one-third of cells producing morphological defect (41-46 %, $p < 0.0152$; Fisher's exact test).

This data suggest RNAi lines are more stable if frozen than continuously maintained in culture. When RNAi cell lines are generated, they must be tested immediately or frozen down until later since freezing the cells may preserve the proportion that would produce morphological defect upon induction.

9 RESULTS 4

9.1 Ways to increase number of transformants in trypanosomes

Introduction of exogenous DNA into cells by the process of transfection is a core technique in several areas of research. The ability to transfect cells allows the study of gene expression, protein function and interaction, to name a few. To be able to study *T. congolense* in the same manner as *T. brucei,* it is important that certain genetic technology milestones are achieved for this particular protozoan.

A first milestone is consistent transfection success. Prior to the current study, most transfections in *T. congolense* have been performed using the insect stage (procyclic form), whilst great difficulty in transfecting bloodstream-form *T. congolense* has been common in the field (Coustou et al., 2010). In the previous chapters I described how we have addressed this and turned bloodstream-form *T. congolense* transfection into a routine, consistent procedure in our labs and those of collaborators.

Along with being able to transfect the infective, disease-relevant form of the parasite, high transfection efficiency (TE) is also desirable – the more transformants one obtains from a single experiment, the more complex experiments can be performed. This milestone has also been achieved for *T. congolense* in our labs, with routine TE being 1000-2000 transformants per transfection.

However, TE yielding ~100,000 transformants would enable high-throughput experiments at the whole genome level. This efficiency was once reported for *T. brucei* (Glover and Horn, 2009) and facilitated the development of the first genome-scale RNAi library in trypanosome (RIT-seq) (Alsford et al., 2011). With our current TE of bloodstream-form *T. congolense* between 1000-2000 transformants, a 5x library coverage of the parasite genome (~100,000 transformants) will require ≥ 50 transfections. While not impossible, increasing TE in *T. congolense* by ~10x would ease and streamline the process, making it more generically applicable.

Three general approaches can be considered to increase number of transformants in trypanosomes:

Increase input – Two inputs go into the transfection process – DNA and cells. The number of transformants may be increased by increasing either or both. By increasing DNA mass, more DNA becomes available in the reaction for introduction into cells. Increasing the number of cells will also increase number of transformants. However, at some point the reaction is expected to saturate. Large cell numbers are also limiting in terms of volume of culture required. We can therefore explore an optimal combination of DNA mass and cell number to increase transformants to some extent.

Increase DNA entry into cells – This can be achieved by modifying the transfection protocol which involves altering the transfection buffer and/or electroporation parameters. A handful of buffers have been successfully used to transfect trypanosomes, e.g. Zimmermann post-fusion medium (ZPFM) (Carruthers et al., 1993), cytomix (Van Den Hoff et al., 1992), *T. brucei* BSF transfection buffer (Schumann Burkard et al., 2011) and human T-cell Nucleofector solution (Lonza). The composition of the T-cell solution is proprietary and reliance on it increases the cost of transfections. Of the three remaining buffers, in the lab we routinely use the *T. brucei* BSF transfection buffer with an Amaxa 2b nucleofector apparatus (Lonza) to transfect *T. brucei*. The nucleofection system has been shown to give higher TE in *T. brucei* than other electroporation methods (Burkard et al., 2007; MacGregor et al., 2013). Some exploration of the nucleofection parameters has taken place by us (Bill Wickstead, personal communication) and others (D'Archivio et al., 2011), following which, the program Z-001 (highest voltage and shortest pulse available) was selected by us as yielding highest TE.

Increase DNA recombination – Recombination events can be increased either by increasing number of cells in the recombinogenic phase of the cell cycle and/or by introducing DNA double-strand breaks (DSB). Hydroxyurea (HU) is an S-phase poison that inhibits DNA replication and stalls cells at the G1/S boundary. After 12 h (procyclic form) or 6 h (bloodstream form) of HU treatment, > 90 % of *T. brucei* cells show S phase markers (Chowdhury et al., 2008; Forsythe et al., 2009). For a few hours after removing HU, there is a synchronised progression of most cells through the next stages of the cell cycle, which provides an advantage to improve transfection. By

transfecting HU-treated cells within 1 h of HU removal, a 5 to 8-fold increase in TE was observed in *T. cruzi* (Olmo et al., 2018). Benz and colleagues described an optimised centrifugal counter-flow elutriation protocol that can be used to obtain synchronised populations of *T. brucei*, with G1 cells making > 90 % of the population (Benz et al., 2017). G1 cells are then separated on the basis of size. While this may be a better way of getting most cells in the population in the recombinogenic phase (whilst avoiding chemical poisoning), no data are currently available (to our knowledge) of TE using cultures synchronized with this protocol. Neither is it known the potential artefacts of elutriation due to extensive cell manipulation outside culturing conditions.

DNA DSB have been used in *T. brucei* to greatly increase transfection efficiency by either the use of a homing endonuclease (I-SceI) or CRISPR-Cas9. Expression of I-SceI in *T. brucei* can be used to create DSB at specific sites by introducing the 18bp restriction sequence at the target site (Glover et al., 2008). By making I-SceI expression regulated, the precise introduction of DSB can be exploited to increase recombination events at a target integration site. This method has been reported to improve *T. brucei* transfection efficiency up to 250-fold (Glover and Horn, 2009). A tetracycline-inducible CRISPR-Cas9-targeted gene editing tool has been developed for *T. brucei.* After an initial transfection to express guide RNA in trypanosome cells, the system was then used to achieve ~ 80-95 % efficiency without drug selection. This is the highest reported efficiency of gene targeting in *T. brucei* (Rico et al., 2018) but the system has not yet been tested in plasmid library experiments.

9.2 Design of construct for inducible homing endonuclease cleavage

CRISPR-Cas9 has the potential to give the highest increase in number of transformants. But CRISPR has not yet been developed in *T. congolense.* In contrast, I-SceI homing endonuclease has been tested in *T. brucei* both for increasing TE and enabling library production. It is a simpler system, requiring fewer transfections and components to set up. I-SceI homing endonuclease also offers greater specificity in landing site and should give sufficient increase on its own.

To set up a system for inducible homing endonuclease cleavage in *T. congolense*, we designed constructs that combine part of an existing *T. brucei* construct with *T. congolense*-specific sequences (Figure 9.1).

1. Inducible promoter – Both T7 and *T. brucei* rDNA promoters were used in generating two constructs. The strong T7 promoter (pTco-TTS) was chosen because it has been extensively tested, but is at times prone to leakiness. A second construct was made using the *T. brucei* rDNA promoter (pTV-TTS). This promoter is weak, perhaps not even functional, but considered here because we expect only little I-SceI activity to be necessary to generate DSB for increased TE.
2. Integration site – The integration site selected was that used successfully for RNAi – silent *VSG* (Chapter 8- Section 8.1.1). The concept is that the RNAi vector replaces the I-SceI construct at this integration site.
3. Drug cassette – I used *HYG* as selection marker, transcribed by a constitutive T7 promoter in both constructs. *HYG* is placed downstream of the *I-SceI* to prevent readthrough and unintended expression of the *I-SceI*.
4. The I-SceI restriction site was placed upstream of the *I-SceI* so that when it cuts, both the *HYG* and *I-SceI* will be lost as telomere-proximal.

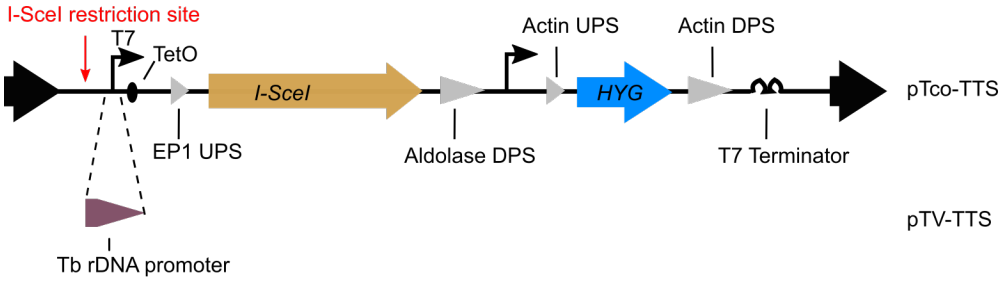


Figure 9.1: Constructs for establishing I-SceI homing endonuclease system in *T. congolense*. Schematics of 2 constructs designed with different inducible promoters (T7 or *T. brucei* rDNA) for expressing I-SceI in *T. congolense*. UPS, upstream processing sequence; DPS, downstream processing sequence.

9.3 Inducing endonuclease transcription results in loss of resistant marker

To test for DSB, I transfected each construct (pTco-TTS and pTV-TTS) into TcoSM cells (Figure 9.2A). Following selection of transformants, I cultured 3 clones of TcoTTS (T7

promoter) and 1 clone of TVTTS (*T. brucei* rDNA promoter) with or without tetracycline for 48 h in the absence of hygromycin. After 48 h, all cultures were plated (at 1.5 cells per well in a 96-well plate) without hygromycin or tetracycline. When positive wells appear in plates, these were mirrored into new plates containing fresh medium supplemented or not with hygromycin, so to identify those clones who had lost HYG resistance (Figure 9.2B).

All wells grew when subbed into fresh plates without hygromycin. However, without induction, ≤ 5 % of wells died in the presence of hygromycin (Table 9.1). When TcoTTS were induced, 54 – 85 % of wells died in the presence of hygromycin. Induced TVTTS had significantly lower dead wells (8 %). In addition to wells where all cells died, there was a population of wells that still grew in the presence of hygromycin, but at a lower rate than the control, herein referred to as "strugglers". "Strugglers" would be expected in some wells because not all wells will be clonal, which would mean these could be mixed populations of hygromycin sensitive (HYG^S) and resistant (HYG^R) cells. But "strugglers" could also be produced if cells were losing the *HYG* during grow out.

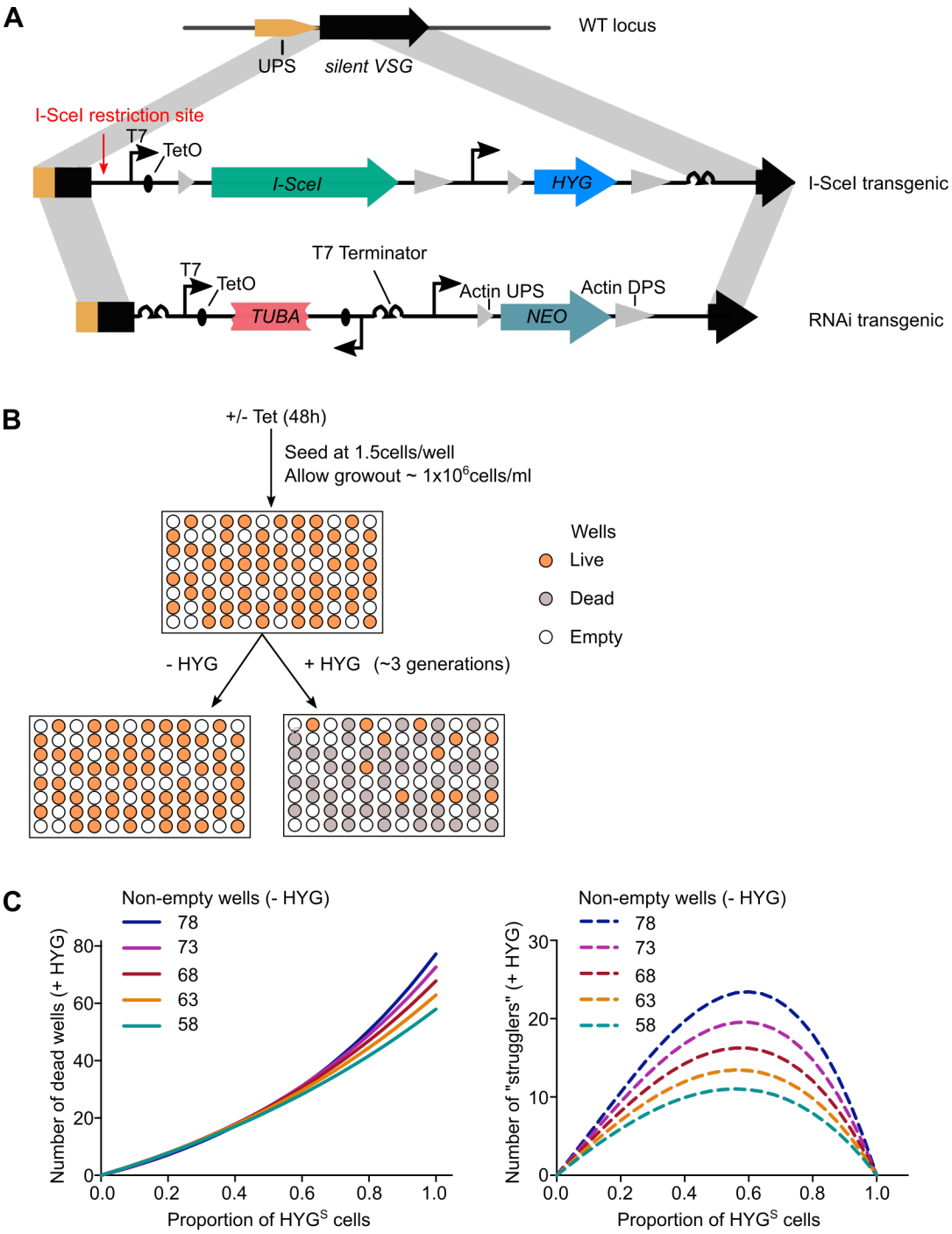


Figure 9.2: Induction of I-SceI expression causes a loss of hygromycin resistance in TcoTTS/TV-TTS lines. A. Schematic of intended integrations at the silent *VSG* locus, first by pTco-TTS/pTV-TTS and with TUBA RNAi construct **B.** Schematic of experimental setup to determine the proprotion of cells that lose hygromycin sensitivity, with or without tetracycline induction of I-SceI. Live wells, contain live cells after seeding; empty wells, contain no cells when seeding; dead wells, live wells that were killed by hygromycin (HYG). **C.** Simulation of the proportion of hygromycin sensitive (HYG^S) and hygromycin resistant (HYG^R) cells after 48 h (+/- Tet). Simulation assumes a Poisson distribution of cells in wells and binomial distribution for proportion of HYG^S and HYG^R cells/well. Left plot shows how curves look like for a range of 58-78 dead wells seeded at 1.5 cells/well. Right plot is the corresponding number of "strugglers" (wells with 50 % or less cells than the control) expected for the simulated dead wells.

Cell line	Non-empty well	- Tet		Non-empty well	+ Tet		Estimate of loss	Expected "strugglers" of +HYG	
		-HYG	+HYG		-HYG	+HYG		-Tet	+Tet
TcoTTS#1	67	0/2	6/13	71	0/2	60/1	85 %	4	12
TcoTTS#2	61	0/0	3/8	58	0/1	36/11	78 %	2	7
TcoTTS#3	68	0/2	2/11	59	0/2	32/10	73 %	2	8
TVTTS	71	0/0	0/6	76	0/0	6/5	12 %	0	5

Table 9.1: Different well categories recorded from testing hygromycin resistance in TcoTTS/TV-TTS line (+/- Tet). Total number of live wells from seeding cells after 48 h +/- Tet (Non-empty wells), Numbers represent wells that are dead/"strugglers" of the different cell lines +/- hygromycin. Wells were seeded at 1.5 cells/well and allowed to grow out (~$1x10^6$ cells/ml) before duplicating the plate (+/- hygromycin) for another 2 days (3 generations). "Strugglers", wells with ≤ 50 % density of control. Estimate of loss and expected "strugglers" are based on simulation in Figure 9.2B.

We modelled the distribution of wells with HYG^R, HYG^S and mix of $HYG^{R/S}$ at different rate of loss at different plate densities (Figure 9.2C, Supplementary File S2). The number of cells/well was modelled as a Poisson distribution with λ estimates from the observed number of non-empty wells (- HYG), whereas the number of cells that were resistant or sensitive was modelled by the Binomial distribution. Any mixed well with ≥ 50 % HYG^S cells in the model was considered a "struggler".

For both induced and non-induced cultures of TcoTTS clones #1, 2 and 3, in the presence of HYG there were ≥ 3x more "strugglers" experimentally than expected by the model to correspond with the number of actual dead wells counted. This is likely the result of mixed wells in the induced set. *HYG* loss post-cloning could also be caused by chromosomal loss during grow out in both induced and non-induced cultures.

From the modelling, induction of I-SceI expression in cells causes loss of downstream drug marker in ~16% of TVTTS and 60-90% of TcoTTS cells. The different inducible promoters correlate with the proportion of cells that became HYG^S after 48 h of induction. The stronger T7 promoter generated a greater proportion of HYG^S wells than the weaker *T. brucei* rDNA promoter.

9.4 Induction of DSB significantly increases number of transformants in *T. congolense*

To test if induction of I-SceI expression increases transfection efficiency in *T. congolense*, TcoTTS cells cultured for 16 h with or without tetracycline were transfected with an RNAi construct targeting the alpha tubulin ORF (see Section 8.4 and Figure 8.3). Tubulin knockdown was used to allow me to screen for inducibility of RNAi by the onset of identifiable morphological change in transfected cells. Three clones were tested.

Tetracycline induction of I-SceI expression and consequent DSB increased the number of transformants from ~1500 (1100-1800) to 11950 (9450-14700) (Figure 9.3A). At ~10,000 independent transformants per transfection, the system is very feasible for generating a genome-wide library of clones. A 5x genome-wide coverage of *T. congolense* is now possible with only 10 transfections.

The RNAi construct is intended to replace the I-SceI at the silent *VSG* locus. In the PacBio assembly of the *T. congolense* genome (Abbas et al., 2018), there is only one silent *VSG* locus present. If indeed the case, then there should only be an RNAi transgenic locus present after transfecting the RNAi construct. However, if there are multiple loci, then there will be an RNAi transgenic locus and either the I-SceI transgenic or WT locus. To test if RNAi constructs were integrated at the I-SceI transgenic site, I selected 10 positive wells from each transfection (the one with and the one without tetracycline induction of I-SceI expression) and I designed a multiplex PCR strategy that distinguishes between loci in these clones (Figure 9.3B). Following PCR, gel electrophoresis was used to resolve amplicons (Figure 9.3C). A ~630 bp band was present in all samples likely representing non-specific binding. Parental control (TcoSM) produces one band above 1 kb indicative of wildtype (WT) locus, as expected (Figure 9.3C). TcoTTS shows two bands (Figure 9.3C): a WT and a modified locus (I-SceI transgenic locus), suggesting there are at least 2 (or more) silent *VSG* loci present (not one as predicted by the PacBio assembly of the *T. congolense* genome sequence). Both transfection populations and all 20 RNAi lines have both a WT locus and a RNAi transgenic locus (Figure 9.3C). Without tetracycline induction (Figure 9.3C), 4/10 lines

have the DNA construct integrated at the I-SceI transgenic locus (samples 1, 2, 5 and 7). Samples 3, 4, 8 and 9 show strong (whereas samples 6 and 10 show some) indication of the presence of WT, I-SceI and RNAi loci, suggesting there are at least 3 silent *VSG* loci in *T. congolense* IL3000 cells grown in our labs.

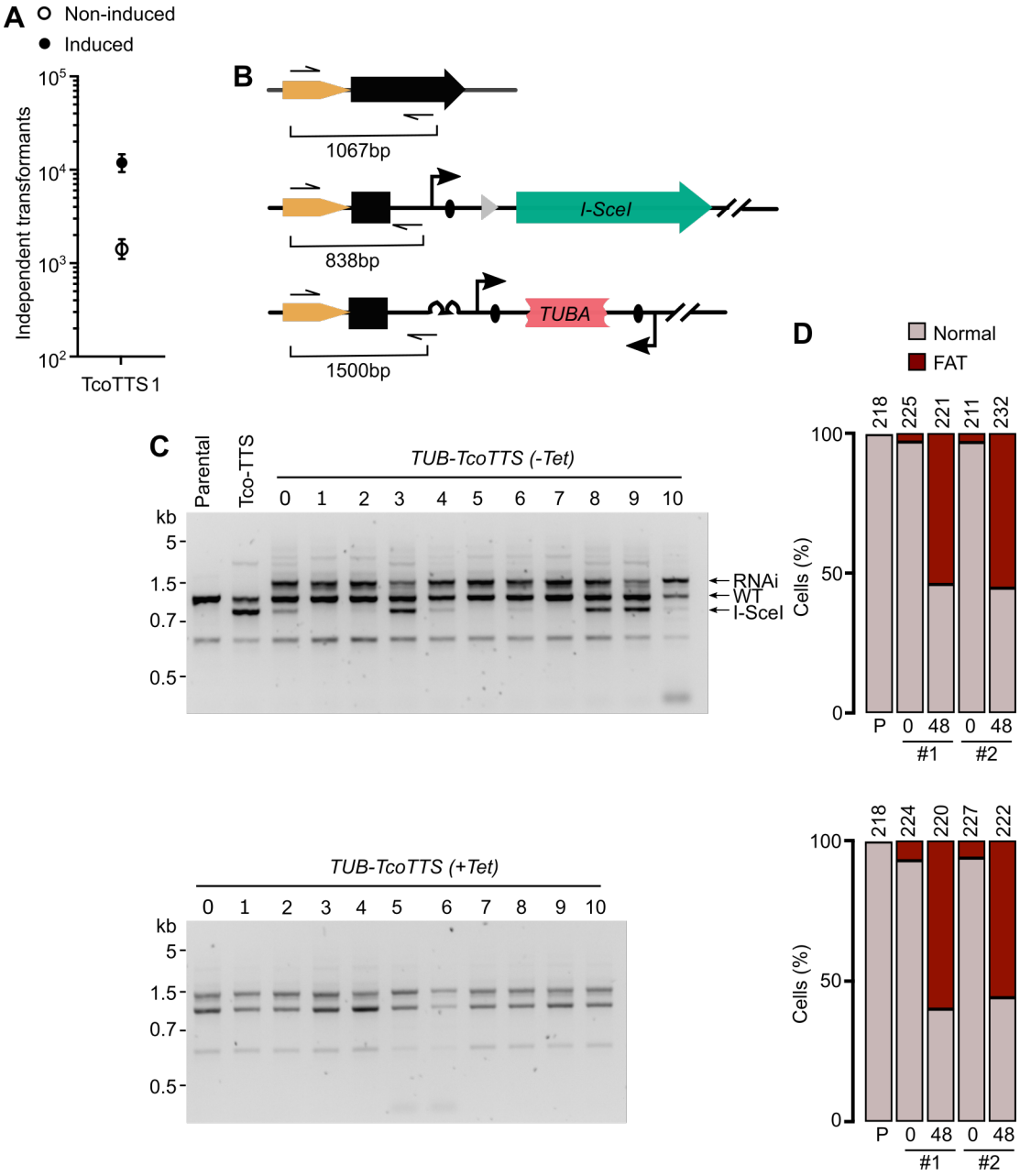


Figure 9.3: Induction of I-SceI expression in TcoTTS line increases transformants and maintains RNAi inducibility. A. Plot showing number of transformants resulting from transfecting TcoTTS (+/- Tet for 16 h) with a *TUBA* RNAi construct. Independent transformants were estimated by limiting dilution. **B.** Schematic showing PCR strategy to distinguish different loci **C.** PCR assay used to test modification of the silent *VSG* loci in parental, TcoTTS and 10 RNAi clones arising from induced or non-induced TcoTTS. **D.** Quantification of cell morphologies of parental and 2 induced/non-induced RNAi clones arising from TcoTTS (+/- Tet), numbers above bars are total number of cells counted.

When tetracycline is added before transfection (Figure 9.3C), 10/10 samples lost the I-SceI transgenic locus. The RNAi construct replaced the I-SceI transgenic locus in all instances, suggesting a directed integration at the site of inducible DSB break. Coupled with the increase in transfection efficiency (mentioned above), the data suggest increased DSB at the target site and a directed integration of RNAi construct at the I-SceI transgenic locus. Use of TcoTTS for integration has no effect on RNAi inducibility, as 2/2 lines produced +/- Tet expected FAT phenotype at 48 h (Figure 9.3D).

9.5 Transfection efficiency of TcoTTS cell line is stable to freeze-thaw

Low levels of I-SceI induction would lead to the loss of I-SceI cutting ability and reduce transfection efficiency. The *T. brucei* Sce* cell line used in RIT-seq is very unstable; it appears to have an intrinsic tendency to lose I-SceI and lower TE over time in culture (Glover et al., 2015). Therefore, *T. brucei* Sce* cells have to be generated fresh every time it is needed for a RNAi library experiment (Glover et al., 2015).

To test stability of TcoTTS lines, I induced and transfected cells that have been kept in culture, or frozen and later thawed, for induction of I-SceI expression and transfection. Clones were either kept in continuous culture for about 3 weeks (30-35 generations) prior to transfection, or frozen after 1 week in culture and then thawed and maintained for another week in culture (21 generations in total) before transfection. Three clones were tested to validate clone-clone variation. After 17 h +/- Tet, cells were transfected with *TUBA* RNAi construct (p3T7-TcoV-TUBA, see Section 8.4).

Consistently, all TcoTTS induced (DSB) cultures prior to transfection gave ~10,000 transformants with 10 µg input DNA (Figure 9.4). No difference was observed between cells that had only been cultured for 1 week prior to freezing and those in continuous culture. In some instances, transfection of non-induced lines produced similar number of transformants as induced lines. This was observed both in cultured and frozen/thawed cells, indicating that it may not be due to freezing/thawing. The cause behind observation is not known, but it is not believed to affect the usability of these cells for high efficiency transfection, so was not investigated further. In summary, the data show TcoTTS cells are stable to freeze-thaw and maintain full DSB

capacity for at least up to 35 generations in culture (Figure 9.4). Unlike the *T. brucei* line used in the RITseq experiment, freezing TcoTTS does not reduce transfection efficiency.

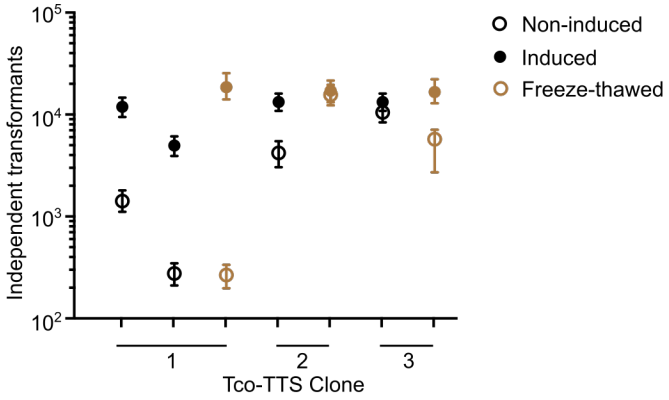


Figure 9.4: Little variation in transfection efficiency of induced TcoTTS cells either in culture or freeze-thawed. Plot showing number of transformants transfecting TcoTTS (+/- Tet for 16 h) using *TUBA* RNAi construct. TcoTTS were maintained in culture (total of 30-35 generations post-selection of transformants) or frozen after a week in culture, thawed and cultured for another week (total 21 generations in culture post-selection of transformants). Independent transformants were estimated by limiting dilution. First data point for TcoTTS clone 1 is a repeat data (Figure 9.3A).

9.6 Optimisation of number of transformants per transfection

To this point, no test adjustment has been made to the input of transfections. Both number of cells and DNA amount are expected to change the number of transformants (see Section 9.1). Therefore, to optimize the system, I tested transfection using different amounts of DNA and cell numbers. After 17 h +/- Tet, TcoTTS was transfected with *TUBA* RNAi construct (p3T7-TcoV-TUBA).

The number of transformants increased approximately linearly depending on amount of DNA up to 20 µg (19625-30550; 90 % confidence interval), reaching saturation between 20 and 40 µg (Figure 9.5). A similar saturation was observed in *T. brucei* 427-TTS between 10 and 20 µg (Simon D'Archivio, personal communication). Cell numbers also linearly increased the yield of transformants when transfected with 10 µg of DNA (23075-35500; 90 % confidence interval). It is likely that increasing cell input further will increase yield, but also increase the volumes of culture required (potentially a limitation). Plotting the same data as transformants per µg of DNA shows a steady decrease in transformants from using 5 µg (2390) to 40 µg (429) of DNA. For every

doubling of input DNA there is an approximate loss of 25 % transformants per µg DNA. The data shows that doubling the number of input cells accounts for the loss per µg DNA. Doubling the number of cells reduced transfection efficiency by 8-15 %. Since I am interested in increasing the number of transformants per transfection and not necessarily transfection efficiency, 20 µg DNA with $5x10^7$ cells was selected as optimal. The combination of the optimal DNA mass with increased cell numbers appears to be additive, giving ~45,000 (36400-56500; 90 % confidence interval) transformants per transfection (Figure 9.5). This number of transformants per transfection means that I will require fewer transfections and can increase genome-wide coverage. With these parameters, and only 5 individual transfections, I can obtain a 10x genome-wide library coverage.

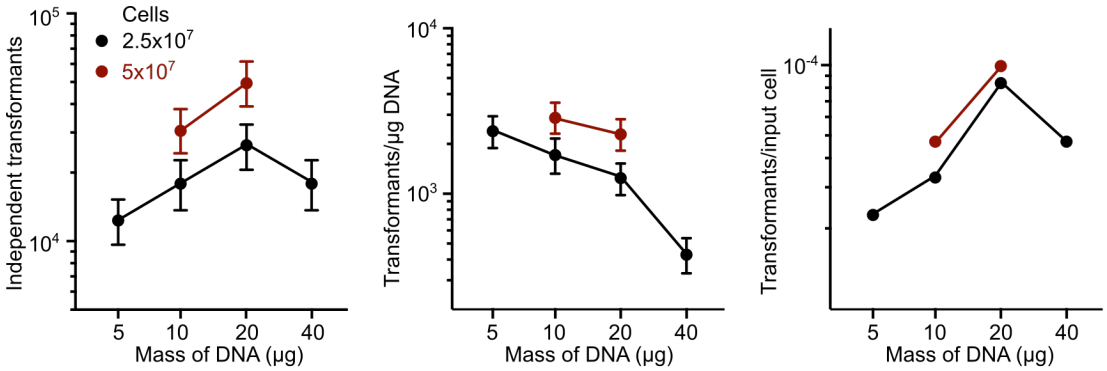


Figure 9.5: Cell number and DNA amount additively increase number of transformants. Plot shows independent transformants obtained using different DNA masses and cell numbers of TcoTTS line. The same data is presented as number of transformants per µg of input DNA and transformants per input cell. Independent transformants were estimated by limiting dilution. Bars = 90% confidence interval for estimates in individual transfections.

10 RESULT 5

10.1 A quantitative genome-wide RNAi screen in *T. congolense*

RNAi is a powerful tool for functional characterization of genes in *T. brucei.* Using the technology to characterise one gene at a time is limiting, but by combining RNAi with genome-wide technology and Next Generation DNA Sequencing (NGS), RIT-seq was established as a fitness screen for most *T. brucei* genes (Alsford et al., 2011). RIT-seq has contributed immensely to our current knowledge of *T. brucei* by providing knockdown profiles of ~7500 non-redundant protein-coding genes in the genome. This genome-wide library has had many research applications, including deciphering pathways involved in drug resistance (Alsford et al., 2012, 2014; Baker et al., 2015; Stortz et al., 2017).

The RIT-seq method does very well in applications where strong positive selection can be used to isolate rare RNAi mutants that have growth advantage under specific experimental conditions. It is therefore a good method for identifying genes associated with a strong loss of fitness (LoF) due to RNAi. However, it performs poorly in quantifying the effect of RNAi against individual genes. This is largely because variation within the data is high, making the method noisy and therefore of low sensitivity.

The Wickstead and Gadelha labs have developed a new method that allows direct RNAi fragment sequencing (DRiF-seq) and robust quantification of LoF at genome-scale. DRiF-Seq differs from RIT-seq in generation of the fragment library, recipient cells, site of integration of RNAi construct, method of sequencing and method of data analysis (D'Archivio, Brusini et al, manuscript in preparation). DRiF-Seq enables direct mapping of RNAi target fragment to specific genes which allows monitoring of the growth of 100,000s of RNAi mutants as individual clones while accounting for sources of variation through "model agnostic", a posteriori noise estimation. Therefore, for a situation like in *T. congolense* where I expect a relatively low RNAi effect on growth due to penetrance issues (demonstrated in Chapter 8), I need a method like DRiF- Seq which is less noisy and has a higher sensitivity, with an estimated positive predictive

value ~20-fold that of RIT-Seq. In addition, the DRiF-Seq method is easily transferable which makes its application in *T. congolense* compatible without major modifications to the method.

To this point, I have established and tested a functional RNAi construct in *T. congolense* (Chapter 8) and developed the technology to generate high numbers of transformants that are required for transfection at genome-scale (Chapter 9). Here, I intend to show genome-wide RNAi data of quantitative LoF screen in *T. congolense* using DRiF-Seq, and its application to identify *T. congolense* genes involved in the mode of action of and resistance to isometamidium. The text that follows herein describes the process of generating a *T. congolense* gDNA library compatible with DRiF-Seq (Figure 10.1).

10.1.1 Plasmid library

In order to generate a *T. congolense* genome-wide RNAi plasmid library compatible with DRiF-Seq method I needed an RNAi plasmid (Figure 10.1A) and fragments of the whole *T. congolense* genome (Figure 10.1B).

- RNAi fragments – To generate targeting RNAi fragments of the whole genome of *T. congolense* IL3000 (of which around one-third is annotated as coding sequences, or CDS), I isolated genomic DNA and sheared it by sonication (Figure 10.1B). Sheared DNA was run on a gel and ~450 bp fragments were excised. A 450 bp fragment size was chosen because this is an appropriate length for RNAi and is also within the optimum range for NGS. Selected fragments were end-repaired to achieve a blunt end before ligation of two different linkers that encode unique restriction sites (SbfI and SfiI) to each end (Figure 10.1B). Using 8-cutter restriction enzymes like SbfI and SfiI ensures that very few RNAi fragments are lost due to cutting within, because of the low frequency of finding 8-bp recognition sites within the genome (1 in every 65 kb). Attachment of these linkers made the RNAi fragments compatible with DRiF-Seq and enabled qPCR amplification of the fragments from primers

complementary to the linkers at both ends. Fragments were then digested with SbfI and SfiI prior to ligation into an RNAi construct (see below).

- RNAi construct – p3T7-TcoV was selected as the most tested RNAi construct arrangement that gave the highest RNAi penetrance (Section 8.1). This construct was used as the base construct for all RNAi screens presented in Chapter 8 (*TUBA, CHC, ESP14, GFP*). Data have been shown on quantification of mRNA, protein and phenotypic characteristics using this construct. p3T7-TcoV has also been used by our collaborators who found it to give similar knockdown as it does in my hands. I modified p3T7-TcoV to introduce primer-annealing regions (for downstream amplification of RNAi targeting fragment) and unique SfiI and NsiI restriction sites flanking a stuffer in between the opposing T7 promoters to make the plasmid compatible with DRiF-Seq (Figure 10.1A). When digested with SfiI and NsiI (the latter produces cohesive ends compatible with SbfI restriction site), the cut vector enables ligation with the digested fragments described above (Figure 10.1B), generating the RNAi plasmid library p3T7-TcoVLib.

10.1.2 Bacterial library

In order to avoid sub-sampling and ensure targeting of the whole genome with losses within an acceptable error range, it is critical to determine and maintain complexity of the intermediate libraries at each step above a threshold that minimizes loss of genes throughout the process up to the final *T. congolense* library.

We took a bottom-up approach by first working out the desired complexity of the *T. congolense* library and used that to determine the complexity of the preceding bacterial library. The probability of hitting any target gene in the trypanosome library is affected by the length of the gene and the number of targeting fragments. To ensure a desired complexity of the trypanosome library, there must be a reasonable number of RNAi fragments targeting any single gene in the genome, referred to as the genic coverage. The more the number of fragments, the higher the chance of hitting the gene, and the confidence of hitting any gene increases towards infinity. For example,

to achieve a 1x genic coverage, we have to successfully target 1 RNAi fragment to each of the ~10,000 CDS in the *T. congolense* genome. If we used exactly 10,000 RNAi fragments, statistically, we will lose a third of the genes purely by chance (~3300 genes). To improve our chances of targeting 1 RNAi fragment to each of the genes, we must increase the targeting fragments by three times (30,000). So for every 1x genic coverage of trypanosome library, we will require 30,000 RNAi fragments. Consequently, we will require 150,000 and 300,000 RNAi fragments for a 5x and 10x genic coverage of trypanosome library respectively.

As the confidence of hitting any gene increases infinitely with an infinite number of fragments, we considered experimental constraints such as the size of culture flask and medium to determine a minimal acceptable loss of genes in the trypanosome library. Lorenzo Brusini and Bill Wickstead simulated the effects of library size on predicted gene coverage in a library of randomly-isolated 500 bp fragments from the trypanosome genome by assuming a Poisson distribution (Brusini, 2019). Based on this simulation, the confidence of hitting a 1 kb gene with ≥2 fragments at 5x genic coverage is estimated at ~96%. At 10x coverage this increases to >99%. Therefore, we chose a range of 5x genic coverage as the minimum acceptable coverage for the trypanosome library, and 10x coverage as the maximum limit for a single library based on experimental constraints. This would comprise of ~150,000 - 300,000 independent clones each carrying a 500 bp fragment. We chose to make 2 trypanosome libraries, each with a 5x genic coverage.

To prevent sub-sampling prior to transformation into trypanosomes, the minimum acceptable complexity of the preceding bacterial library was set at 10 times ($1.5x10^6$ colony forming units or cfu) the desired number of independent clones for the *T. congolense* library. In order to achieve this complexity, high-efficiency electrocompetent *E. coli* was transformed with the plasmid-insert ligation reaction, and a fraction of the transformation was plated to estimate efficiency (Figure 10.1B). The remainder of the transformation was cultured in broth for plasmid isolation (Figure 10.1B). Bacteria transformation yielded ~$6x10^7$ cfu, which is ~30x above the

desired minimum (see above). Plasmid isolation and restriction digest confirmed a distribution around 450 bp of insert fragment in the plasmid (Figure 10.1B).

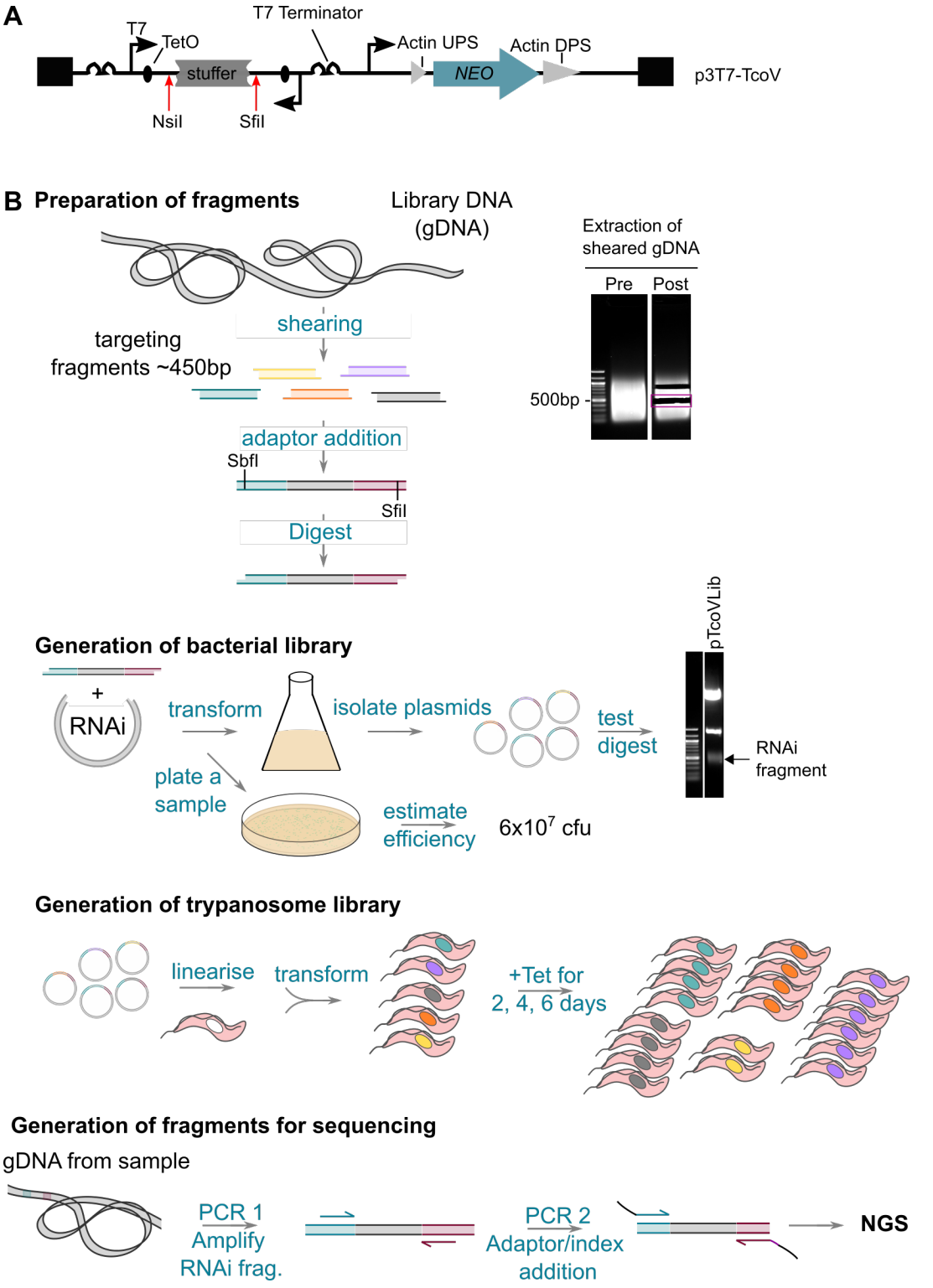


Figure 10.1: Schematic of genome-wide RNAi library production process. A. Schematic of the base library construct (p3T7-TcoV) designed with a stuffer flanked by SfiI and NsiI restriction sites to ligate RNAi fragments. UPS, upstream processing sequence; DPS, downstream processing sequence. **B.** Randomly sheared genomic DNA (gDNA) fragment (~450 bp selected, red box), are linked with adaptors, ligated into plasmid (p3T7-TcoV) and transformed into bacteria to generate bacterial library before plasmid isolation (p3T7-TcoVLib) – test digested to confirm RNAi fragments, linearization of plasmid library and transfection into trypanosomes. Following generation of trypanosome libraries, tetracycline induction (+ Tet) either comes at a fitness cost (yellow), fitness gain (purple) or at no cost to parasites (turquoise/orange/grey). Amplification of whole targeting fragments and direct NGS provide direct counts of fragment representation in the library.

10.1.3 *T. congolense* genome-wide RNAi library

Prior to generation of the *T. congolense* library, the plasmids isolated from the bacterial library were digested with NotI to generate linear constructs for transfection.

It is expected that transformation using 20 µg DNA and $5x10^7$ TcoTTS cells after 16 h of induction of I-SceI expression would yield ~$4.5x10^4$ independent transformants (see Section 9.6). Using the same protocol in Section 9.6, I transfected a total of $4x10^8$ bloostream-form *T. congolense* cells with a total of 160 µg DNA, as 8 separate transfections (Figure 10.2A). Following nucleofection, cells were allowed to recover in fresh medium without drug selection as a single culture for 16 h, before dividing that equally into 2 separate cultures. Just before the split, a sample of the total transfection was taken to estimate by limiting dilution the number of independent transformants (Figure 10.2A). Selective drug was added to each culture over a period of 10 days to generate 2 independent *T. congolense* RNAi libraries with ~5x genic coverage per library. Estimate of independent transformants gave 525,000 (418,500-650,000; 90% CI) which is within the expected range. This total number of transformants is equivalent $2.6x10^5$ independent transformants per library. Generating two independent *T. congolense* libraries enables me to test reproducibility of RNAi effect across different, separate experiments.

In the absence of induction of RNAi, both libraries were similar in growth rate, with a generation time of 14-16 h. RNAi was induced by adding tetracycline (1µg/ml) to the culture medium over 6 days. Samples of ~$5x10^7$ cells (see mathematical reasoning below) were taken in duplicates for each library on days 0, 2, 4 and 6 post-RNAi induction (Figure 10.2B). In order to map changes in gene-specific fragments with a

suitable level of confidence (+/- √n), we assumed an arbitrary minimum of 100 copies per targeting fragment (+/-10%). At each time point I sampled ~250 copies per fragments (~5x10^7 cells). The actual number of copies of targeting fragment present were estimated by qPCR of genomic DNA isolated from experimental samples using barcoded primers (see Table 3.1). Following copy number estimation, amplified fragments were prepared for a High-Output NextSeq 550 sequencing run to provide 75 bp paired-end reads.

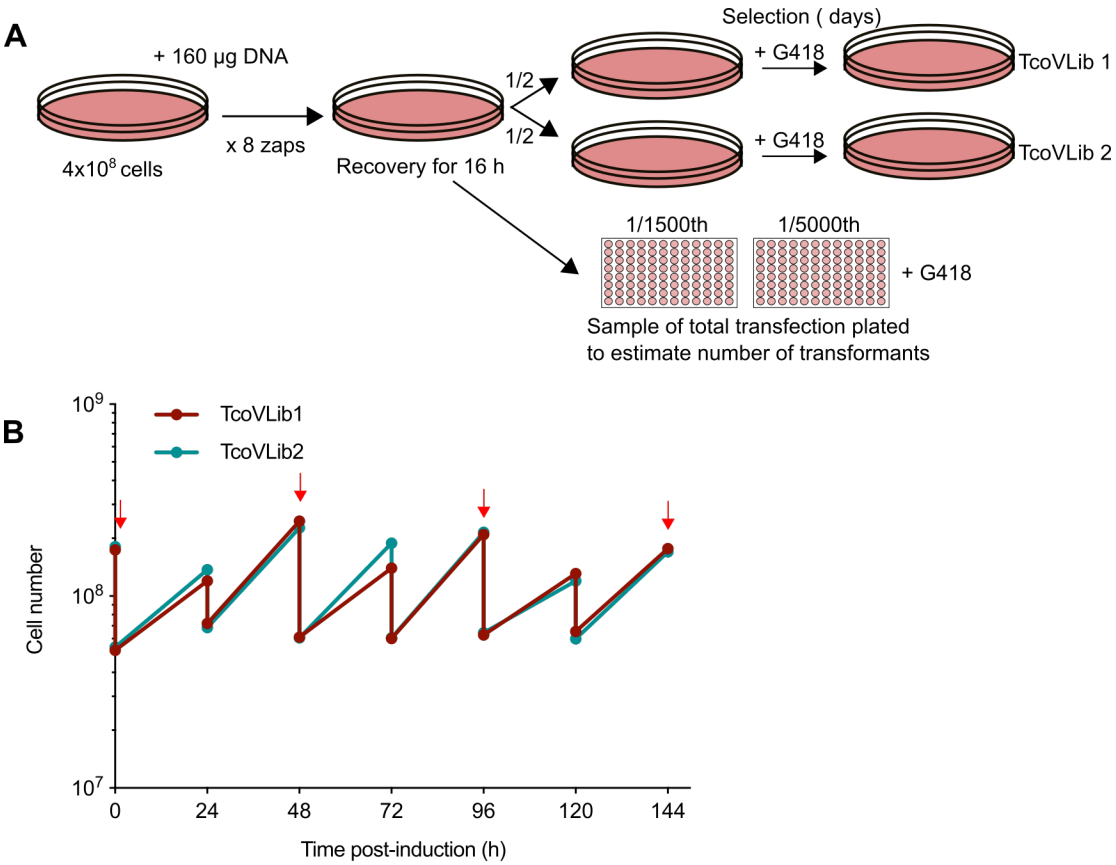


Figure 10.2: Transfection of *T. congolense* and RNAi induction. A. Schematic of trypanosome library generation process. After transfection and 16 h recovery, cells were split into two cultures before drug selection to generate two biological replicas of the trypanosome library. **B.** Growth of *T. congolense* RNAi libraries following RNAi induction. Red arrows, times at which samples were collected. y-axis represent total cell number in 100 ml volume per library.

10.2 *T. congolense* DRiF-Seq

NGS reads were matched to the PacBio assembly of the *T. congolense* IL3000 genome (Abbas et al., 2018). A total of 200 and 196 million pairs of reads passing quality filter were generated for Library 1 and 2 respectively, with each time point represented by

≥ 43 million pairs. Of the total reads, 193 and 189 million pairs map concordantly to the IL3000 PacBio assembly, representing 438,775 and 429,773 individual fragments with unique fragment boundaries and with an average of ≥ 32 counts per sample. 234,171 and 230,234 fragments overlap one of the 10,323 annotated CDS by ≥ 150 bp, and thus were considered to likely target these genes on RNAi induction, giving an average of ~14 fragments per 1 kb CDS per library. Based on the number of overlapping fragments, 9616 and 9619 genes were targeted for RNAi in Library 1 and 2 respectively. Only genes targeted by ≥2 RNAi fragments were included in analysis. Thus, each library covers about 93% of the genes in the genome assembly (9616 and 9619 genes) with 9452 genes present in both libraries (see Supplementary File S3 for DRiF-Seq data). There are therefore very few genes present in only one library (164, 167 genes). The 93% of annotated genes present per library is lower than the expected 96% for a 5x genic coverage (see Section 10.1.2) and this difference can be due to two possible explanations given below. Between the 2 libraries, there was no fragments targeting 540 genes. Of the 540 genes absent in either of the libraries, 365 are from contigs not assigned to chromosomes (2.6x the number expected by chance), suggesting at least some of these non-covered genes are due to errors in the assemblies for these regions. However, also in this set are genes mostly annotated as encoding histone H2A, H2B, H3 and H4 (11 of 14, 20 of 22, 11 of 12, and 12 of 20, respectively), which would be expected to be essential in *T. congolense* and have in fact a strong, early loss-of-fitness seen by DRiF-Seq in *T. brucei* (data not shown). In addition, genes encoding β-tubulin, Hsp90, dynamin and non-mitochondrial elongation factor Tu are in this set of genes with no mapping fragments. This strongly suggests that some fragments associated with strong knockdown effects are missing from the libraries due to selection prior to induction, most likely due to leakiness in transcription of the RNAi fragment.

10.2.1 *T. congolense* DRiF-Seq can differentiate between genes that produce an effect on knockdown by RNAi and those that do not

Mapping of RNAi fragments to annotated CDS sequences shows the distribution of targeting fragments along the length of a gene (Figure 10.3). Several target fragments

with different magnitudes of fold change in fragment abundance upon induction may be mapped to a particular gene.

By inducing RNAi and competing clones against each other, we will see a reduction in the number of cells for clones where knockdown of the gene causes a loss of fitness. Where several fragments target an essential gene over the induction period, it would be expected that all are reduced in abundance relative to fragments targeting non-essential genes. As proof of principle, we observed the behaviour of RNAi fragments mapping to the CDS encoding clathrin heavy chain, CHC (TcIL3000_10_5120) and the orthologue of the *T. brucei* ESP14, TcoESP14 (TcIL3000_7_180) over the period of induction (Figure 10.3). We compared $\log_2$ ratio of fragment abundance (+Tet/-Tet) over 6 days of induction. At day 2, very few genes (30 and 11 in Library 1 and 2 respectively) including *CHC* (reduced by 1.2-fold) are associated with a detectable change in ratio at false discovery rate (FDR) ≤ 0.01 (Figure 10.4A). At day 4 and 6 post-induction, there is a much greater change in fragment abundance for some genes – up to 4.3 and 7.3-fold respectively (Figure 10.4B and C). *CHC* fragment abundance reduced by 1.2- and 3.2-fold on day 2 and 4 respectively. Based on data from Section 8.4, the loss of *CHC* fragments is lower than the loss of cells observed on individual knockdown of *CHC* (1.7- and 10.2-fold on day 2 and 4 respectively). This is expected because loss of fragment is not the same as loss of cells for two reasons. Firstly, a cell that fails to go through cell division due to an RNAi effect might be in G2 and therefore has twice the amount of DNA. While RNAi effect will reduce the number of individual cells, the library will account for the fact that there is twice the amount of DNA present in a G2-phase cell. Therefore, there may be half the number of cells but twice as much DNA. Secondly, cells may die but their DNA may still remain present in the culture flask, and be isolated and counted in the library. The latter is not likely to be the case here because when sampling the trypanosome library cultures, cells were washed (in phosphate buffer saline) to remove soluble DNA of dead cells/debris before gDNA isolation, fragment amplification and sequencing.

The abundance of RNAi fragments targeting a non-essential gene, TcoESP14 (see Section 8.3) did not significantly change in either library (FDR > 0.1) over the period of

induction (Figure 10.4). This suggest that genes that show no effect due to RNAi in the library may not be required for *in vitro* growth of bloodstream-form *T. congolense*.

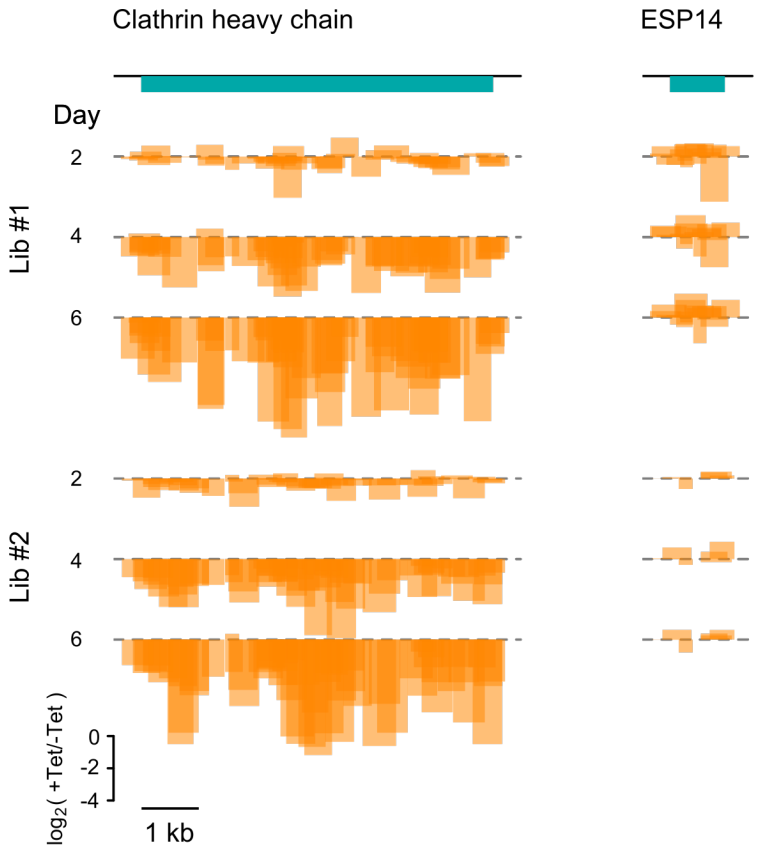
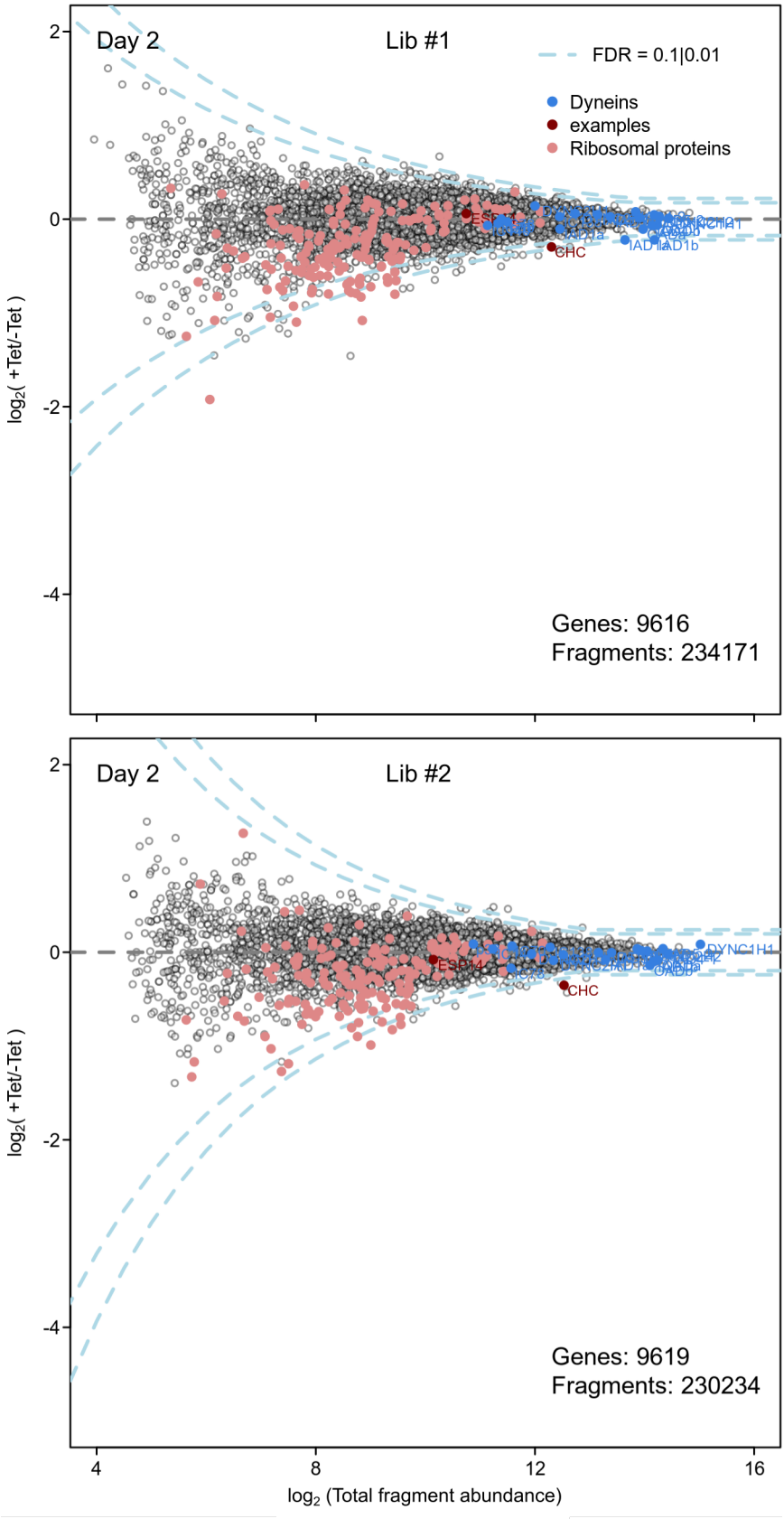
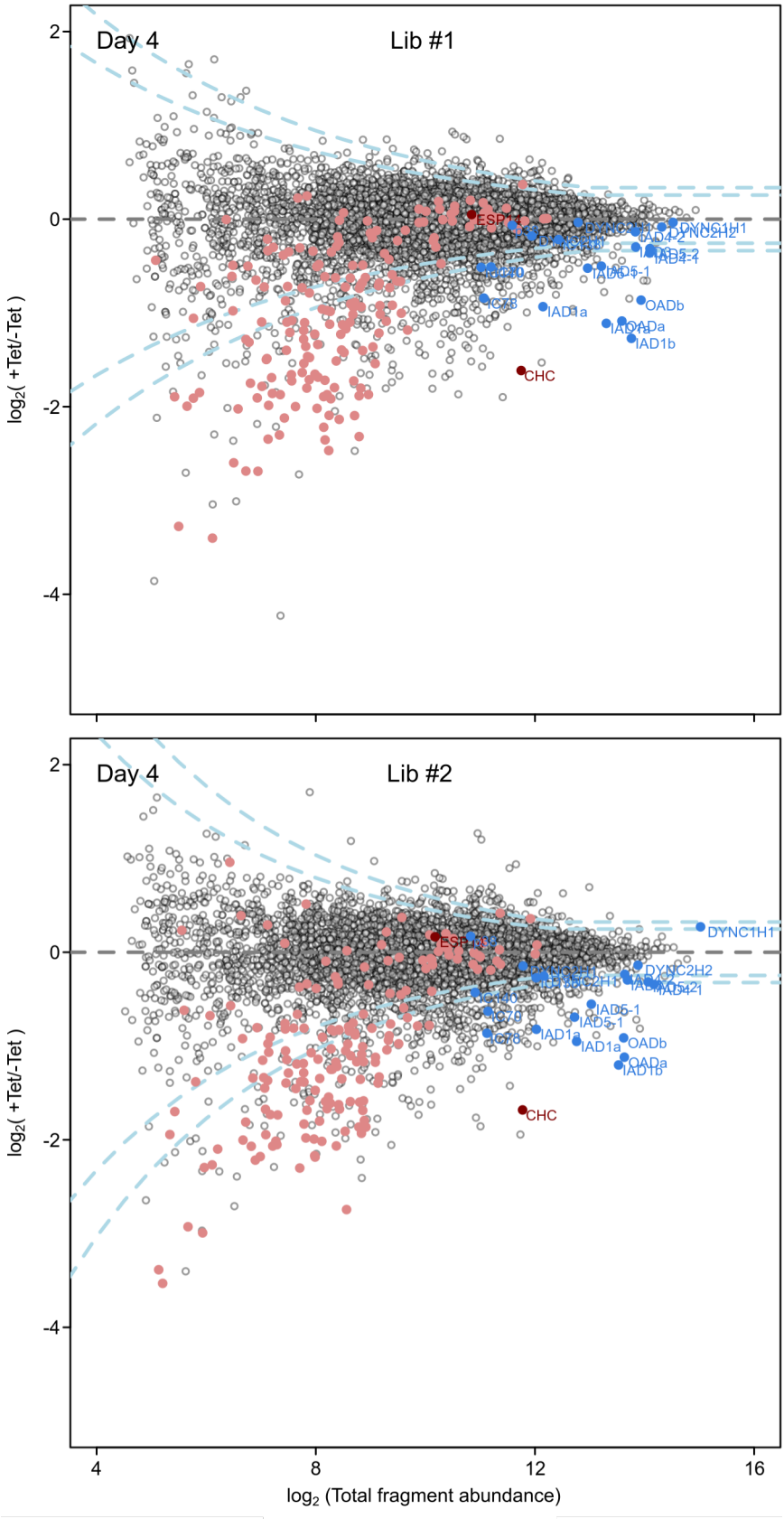


Figure 10.3: RNAi fragments mapped to the same gene vary in their magnitude of change on induction. Mapping of RNAi fragments (orange) to coding sequence (turquoise) of clathrin heavy chain and ESP14 at different days of induction of both libraries.


Day 2
Lib #1
FDR = 0.1|0.01
Dyneins
examples
Ribosomal proteins
CHC
IAD1b
Genes: 9616
Fragments: 234171
Day 2
Lib #2
DYNC1H1
CHC
Genes: 9619
Fragments: 230234
log2(+Tet/-Tet)
log2 (Total fragment abundance)
2
0
-2
-4
4
8
12
16


Day 4
Lib #1
Day 4
Lib #2
log2(+Tet/-Tet)
log2 (Total fragment abundance)
2
0
-2
-4
4
8
12
16
DYNC1H1
DYNC2H2
IAD4-2
IAD4-1
IAD5-1
IAD1a
OADb
OADa
IAD1b
CHC

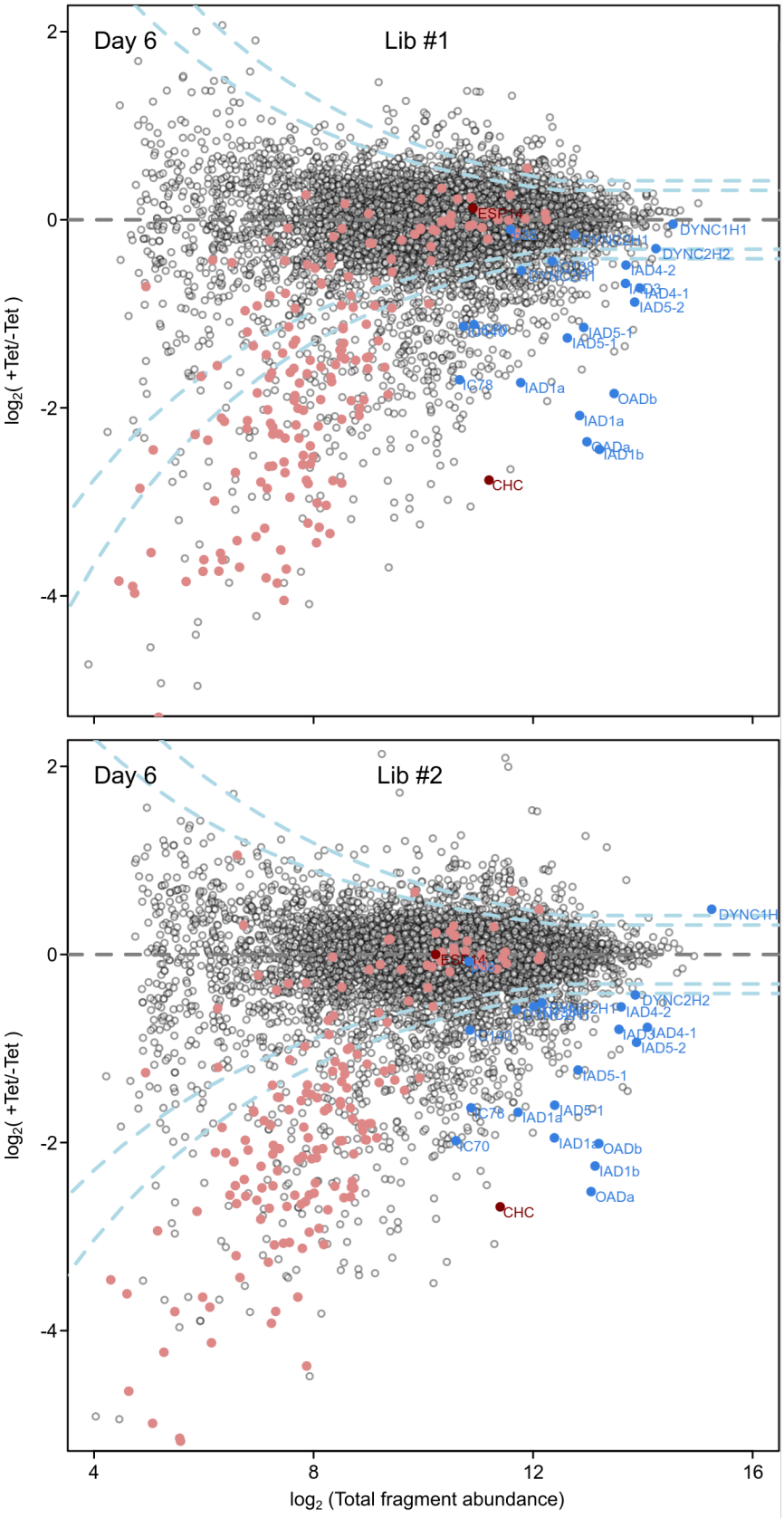
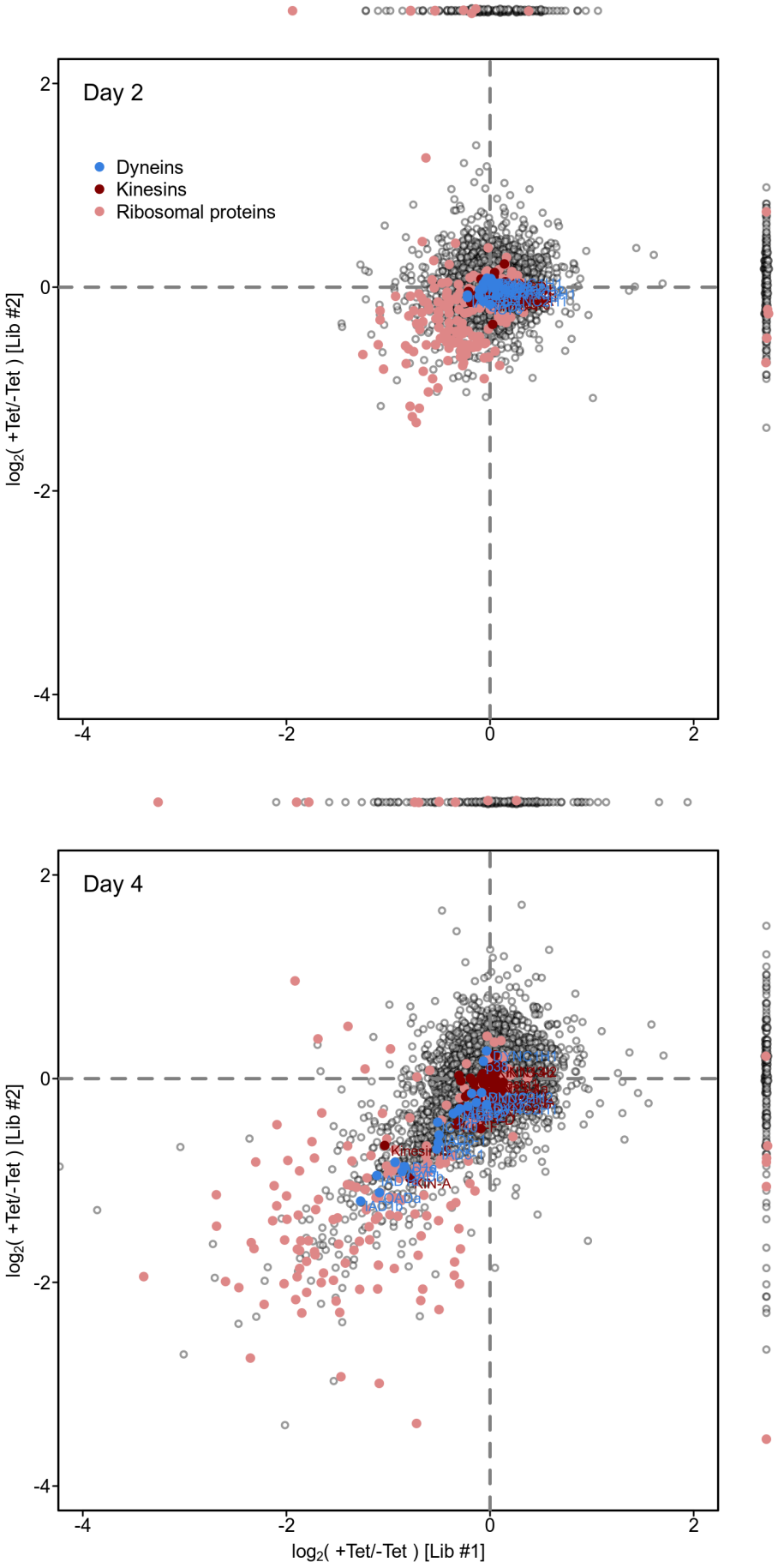


Figure 10.4: Variation in target fragment abundance between different genes on induction. Plots of $\log_2$ fold change of gene fragment abundance with (+Tet/-Tet) against total gene fragment abundance on day 2, 4 and 6 post-induction of trypanosome library 1 (Lib #1) and 2 (Lib #2). Circles represent targeted genes.

10.2.2 Independent *T. congolense* DRiF-Seq libraries demonstrate good agreement in gene fitness profiles

Being the first RNAi library in *T. congolense,* it is important to demonstrate robustness of the libraries which was the primary reason for creating two separate libraries. As a measure of robustness, we investigated the level of agreement between the two independent libraries. We anticipate that the majority of ribosomal proteins will be essential. Also, most ribosomal proteins are small in size, which will typically reflect as a low fragment abundance due to generally fewer targeting fragments of smaller genes. We found that ribosomal protein genes in both libraries are among the earliest responders to induction (FDR < 0.01). On day 2, 9 and 4 genes (in Library 1 and 2 respectively) encoding ribosomal proteins showed a significant fold change at FDR < 0.01 (Figure 10.5). At day 4 (87, 85; FDR<0.01) and day 6 (97, 111; FDR<0.01) there were more genes encoding ribosomal proteins that had a significantly negative fitness cost due to RNAi respectively (Figure 10.5).

Most dyneins play a role in flagellar beating. Few are also involved in retrograde transport along the flagellum as well as transport within the cytoplasm. RNAi against subunits of dynein complexes are therefore used as a way of evaluating the importance of flagellar beating in trypanosomes. Knockdown of dynein heavy and intermediate chains in *T. brucei* showed a strong LoF (Alsford et al., 2011). Here, the heavy and light chains of the axonemal dyneins have a relatively strong LoF effect (up to 6-fold by day 6) and is similar for the intermediate chains (up to 4.3-fold by day 6) that are also associated with the axonemes. The dyneins involved in intraflagellar transport (IFT) generally have a smaller effect (up to 1.7-fold by day 6). Cytoplasmic dynein components (DYNC1H1 and IC1) show no effect due to RNAi. These data match observation in RIT-seq where RNAi against cytoplasmic dynein 1 (DYNC1H1) did not cause a defect. The outcome of RNAi against axonemal dynein heavy chains (DYNC2H1 and DYNC2H2) is also in agreement with previous outcome in *T. brucei.* Knockdown of either DYNC2H1 or DYNC2H2 in bloodstream-form *T. brucei* inhibited IFT resulting in short and abnormal axoneme (Blisnick et al., 2014). These data strongly suggest that flagellar motility is essential in bloodstream-form *T. congolense* as it is in *T. brucei.*


Day 2
Dyneins
Kinesins
Ribosomal proteins
log2(+Tet/-Tet) [Lib #2]
Day 4
log2(+Tet/-Tet) [Lib #2]
log2(+Tet/-Tet) [Lib #1]
2
0
-2
-4

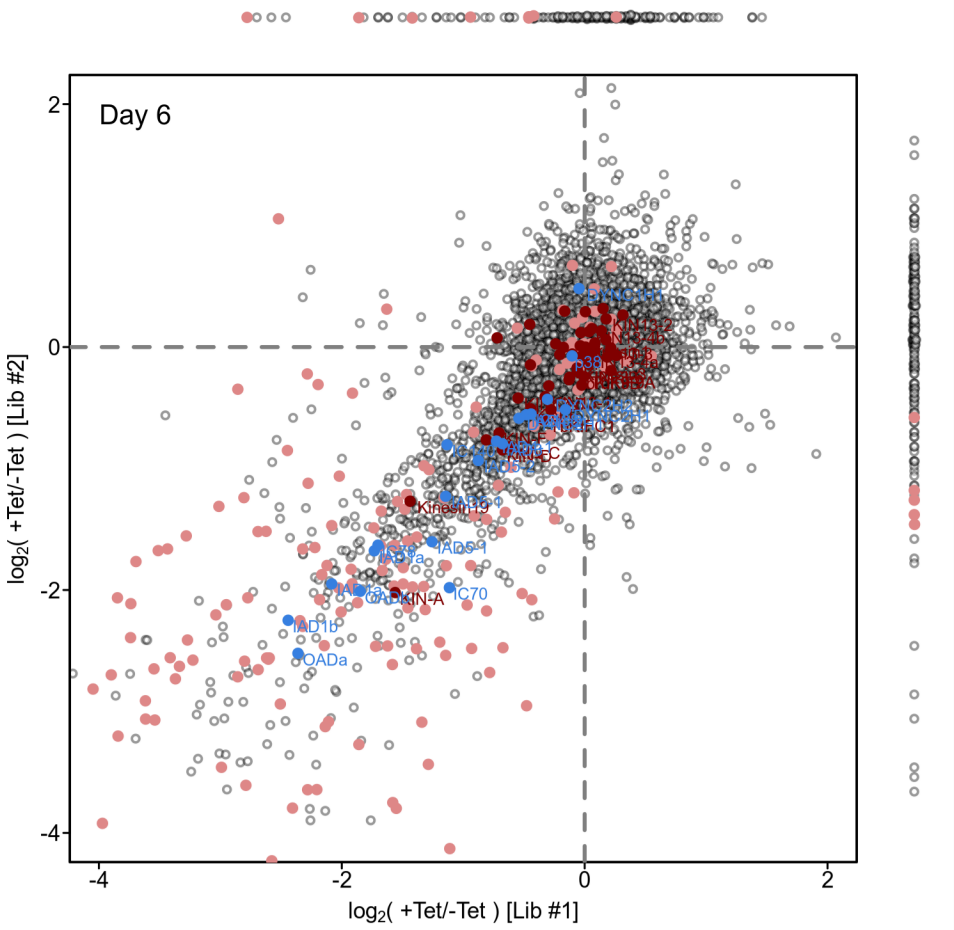


Figure 10.5: Gene fitness profiles are consistent between *T. congolense* DRIF-Seq libraries. Plots of $\log_2$ fold change (+Tet/-Tet) abundance of genes in library 2 (Lib #2) against library 1 (Lib #1). Circles outside of square enclosure are genes targeted present in Lib #1 (right) or Lib #2 (top) only.

10.2.3 There is a strong positive correlation between effect sizes of genes in both libraries

Generally, the effect sizes of knockdown of genes in Library 1 correlates with the effect sizes in Library 2 (r= 0.17, 0.61 and 0.74 for day 2, 4 and 6 respectively). The degree of positive correlation is weak on day 2, stronger on day 4 and even more so on day 6 post-induction. We can therefore plot the libraries against each other without changing the observed effect of genes (Figure 10.5).

An alternate representation of selected data from Figure 10.5 shows the spread of the effect sizes associated with the individual fragments of a given gene (Figure 10.6). Data are shown for three genes associated with loss of fitness – clathrin heavy chain

(TcIL3000_10_5120), 60S ribosomal protein L30 (TcIL3000_10_3430), and kinesin-19 (TcIL3000_6_3820); and two genes with no associated fitness cost – PPDK (TcIL3000.11.6820) and TcoESP14 (TcIL3000_7_180) in both libraries. In all instances, growth curves of individual libraries are closely matched, which supports the strong correlation between the independent libraries. Fold changes of different fragments are sometimes widespread in LoF genes, particularly on day 4 and 6 post-induction (up to a range of 3.7 to 32-fold reduction on day 6 of 60S ribosomal protein L30). As discussed in Section 10.2.1, the RNAi-mediated depletion of *CHC* produced a greater fold reduction when measured by growth of cells than of DNA. As expected, depletion of *TcoESP14* and *PPDK*, where individual knockdowns were not associated with LoF (see Section 8.3), produced no significant fold change (Figure 10.6).

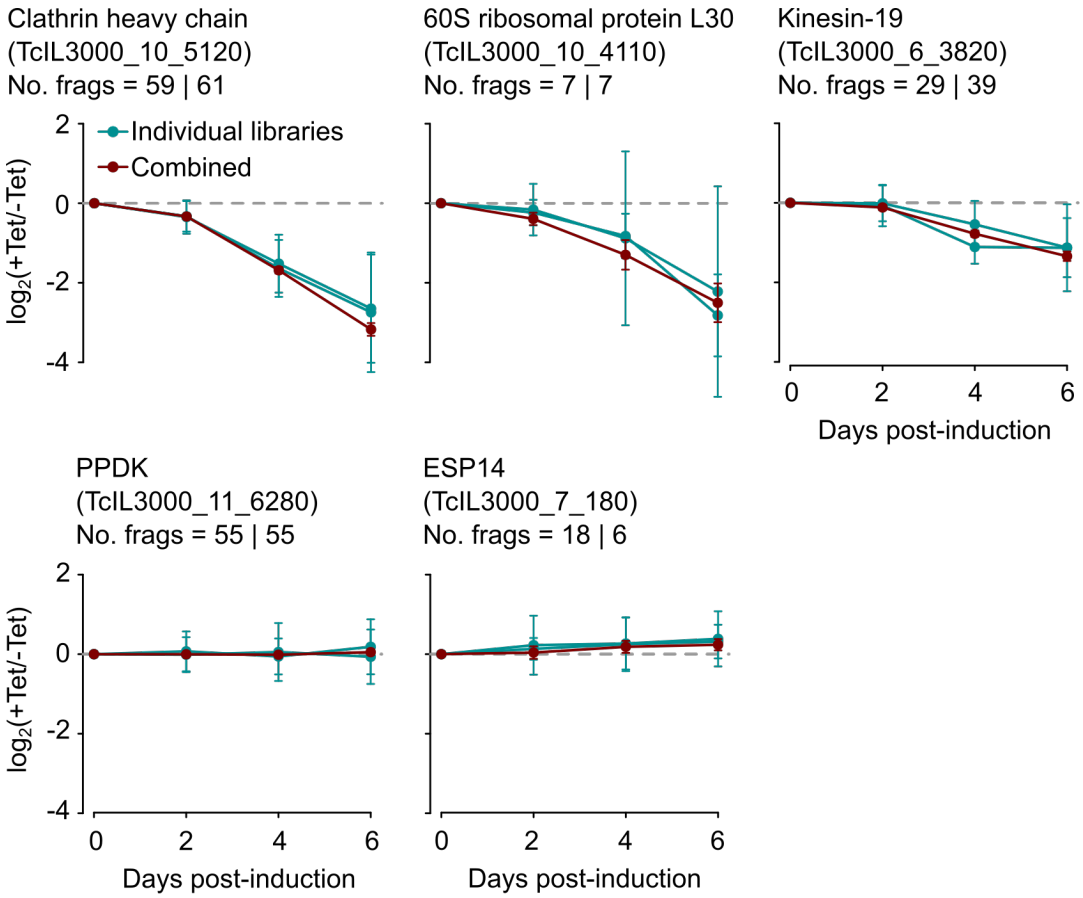


Figure 10.6: Plots showing the spread of different fragments (around the median) of selected genes in individual libraries and combined data from both libraries over 6 days of induction.

10.2.4 A set of genes in the libraries is associated with an apparent gain of fitness

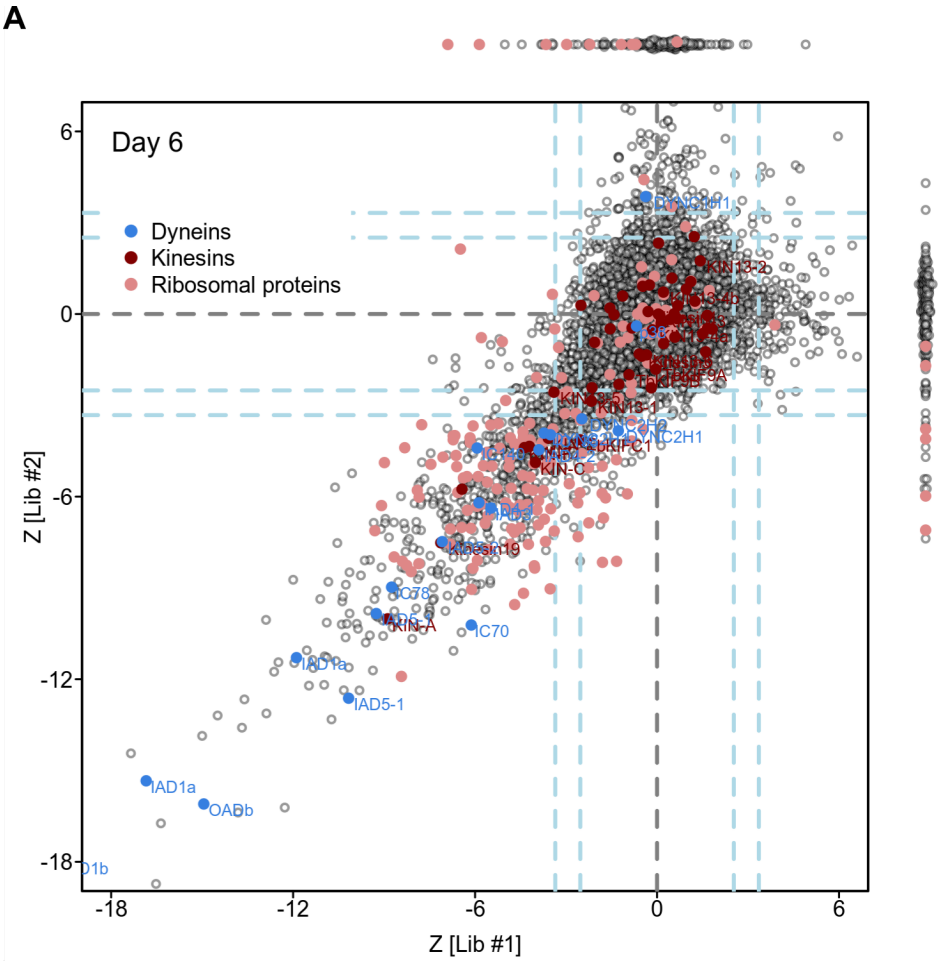
Up to this point, LoF measurements have been based on the magnitude of fold change of fragment abundance over the RNAi induction period. Based on fragment count, and the spread of the data, Z scores were calculated for any particular fragment abundance. Figure 10.7A shows the Z values of Library 1 against those of Library 2; once again, there is a strong positive correlation between libraries (r = 0.173, 0.621 and 0.729 for days 2, 4 and 6 respectively). There are 727 genes causing significant LoF at FDR≤0.01. Of this number, 503 show significant LoF in both libraries, while 82 and 142 are LoF genes in Library 1 or 2 respectively. Most importantly, there are a number of genes associated with a significant apparent gain of fitness (GoF). Knockdown of these genes produce a growth advantage, resulting in an increase in the number of cells carrying RNAi fragments of these genes. In *T. brucei*, knockdown of very few genes provide any kind of growth advantage. Therefore, I did not anticipate that there would be many genes that could significantly increase growth rate of bloodstream-form *T. congolense* in culture. And yet, our work shows 209 genes whose knockdown leads to significantly increased growth rate (FDR ≤ 0.01). Most GoF genes are associated with only one library (104 genes in Library 1 and 98 in Library 2, FDR ≤ 0.01). There is however a small number of genes associated with GoF in both libraries (7, FDR ≤ 0.01). The lack of commonality between the two GoF gene sets could be possibly interpreted as noise in the system. Perhaps of interest, GoF effect size is generally smaller (Z < 10) than that observed for LoF (Z = <22). But more importantly, the same is not seen for those genes associated with LoF, whereby both libraries share > 70% overlap.

Of the 7 GoF genes in common, there are three kinases, a synthase, a phosphatase, a zinc finger protein and a serine peptidase. Kinases are of particular interest because of the function associated with the *T. brucei* orthologues (Mony et al., 2014; Saldivia et al., 2016). In *T. brucei,* AMP-activated protein kinases (TbAMPKs) include α1 and α2 subunits that form two complexes with β and γ subunits. Actively-dividing bloodstream-form cells depleted of each TbAMPKα protein separately did not show significant changes in growth in culture, nor did they respond to treatment with AMP

or AMP analogues that typically induce a quiescent-like phenotype (Saldivia et al., 2016). However, when both the TbAMPKα1 and TbAMPKα2 subunits were depleted using RNAi, the parasite showed resistance to AMP-induced growth inhibition, suggesting that TbAMPKα1/2 are involved in positive regulation of differentiation into the quiescent-like form (Saldivia et al., 2016). The *T. congolense* orthologues of the AMPK α-subunits (TcIL3000.A.H_000761100, TcIL3000.A.H_000349400) do not produce neither a significant loss or gain of fitness in either library (Z= -1.42, -1.28 and Z= 2.24, 2.93 respectively at FDR ≤ 0.01), which agrees with the outcome of the individual RNAi in bloodstream-form *T. brucei* mentioned above. However, our data show that knockdown of *T. congolense* orthologues of the β (TcIL3000.A.H_000584600) and γ (TcIL3000.A.H_000349400) subunits of AMPK produces a significant GoF. Depletion of the γ-subunit produced a greater GoF in both libraries (Z=7.98, 7.06 at FDR ≤ 0.01) than the β-subunit (Z= 3.45, 4.9 at FDR ≤ 0.01). These data strongly suggest that, similar to *T brucei,* AMPK plays a role in the regulation of quiescence in *T. congolense*, and that the GoF associated with AMPK knockdown in *T. congolense* is most likely mediated by the β and γ-subunits. The lab is currently investigating these observations in greater detail.

Another area of interest of our labs is eukaryotic cell cycle control. We looked at genes involved in cell cycle regulation to identify those associated with significant LoF in *T. congolense* (Figure 10.7B). As expected, knockdown of a number of kinetoplastid kinetochore proteins [KKT (9)], KKT interacting proteins [KKIP (2)], CRKs (2) and cyclins (2) are associated with LoF. KKTs and KKIPs are components of the kinetochore. KKIP1, a Ndc80/Nuf2-like protein (D'Archivio and Wickstead, 2017), has recently been described as a connector between a stable kinetoplastid outer kinetochore complex (consisting of 9 KKIPs) and the inner kinetochore composed of KKTs (Brusini et al., 2019). Some KKT proteins are required for chromosome segregation (Akiyoshi and Gull, 2014) and their depletion impacts cell growth. Trypanosome-specific cdc2- related kinases (CRK) interact with different cyclins (CYC) for progression through the different cell cycle stages. Knockdown of some cyclins and CRKs have

been associated with cell cycle arrest and stage-specific enrichment [reviewed in (Li, 2012)].



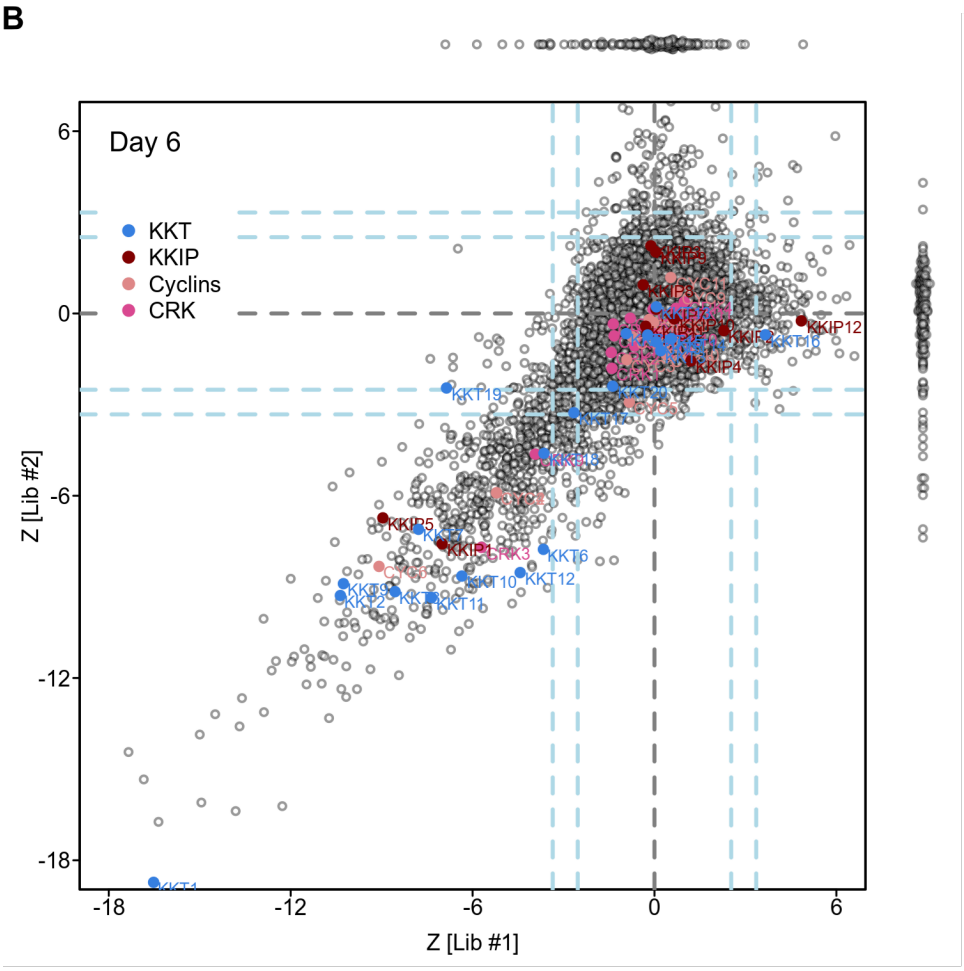


Figure 10.7: Knockdown of certain genes are associated with significant loss of fitness. Plot of Z-values for genes at day 6 post-induction, Lib #2 against Lib #1. **A.** Annotation of groups of genes (Dyneins, Kinesins, Ribosomal proteins) showing different fitness profiles. Blue dashed lines = false discovery rate (FDR) of 0.01 and 0.1. **B.** Annotation of groups of genes involved in cell cycle regulation (KKT, KKIP, Cyclins, CRK). Blue dashed lines = false discovery rate (FDR) at 0.01 and 0.1.

10.2.5 Quantitative enrichment analysis

To investigate fitness not of individual genes but whole biological processes and cellular compartments, we explored the DRiF-Seq data using Gene ontology (GO) annotations that provide vocabularies to describe genes. Quantitative enrichment analysis will enable us to see whether there are sets of genes associated with LoF. We analyzed both computed and curated GO terms that were significantly depleted in both libraries. Figure 10.8 shows a plot of the outcome. See Table 10.1 and 10.2 for mean and adjusted p-values. The p-values were adjusted for multiple comparisons by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). As previously discussed in Section 10.2.2, we expect sets of genes like dyneins and ribosomal

proteins to have a strong effect. The 18 genes in the set annotated with the GO component “dynein complex” have on average a very strong effect (Z = -7.96) which is unlikely to be by chance (adj. p < 1e-5). Also, the set of genes (43) annotated with GO component “ribosome” on average have a strong effect (Z = -4.85) also unlikely to be due to chance (adj. p < 1e-5). Genes annotated as “COPI vesicle coat” have a very strong LoF in *T. congolense* (Z = -9.89, adj p = 5.1e-3). There is also a LoF effect produced by the V0 and V1 domains of proton-transporting V-type ATPase (Z = -6.31 and -7.79 respectively). V-ATPases regulate the acidity of intracellular compartments. This homeostatic state is linked to the function of membrane trafficking (Forgac, 2007). Depletion of V-ATPases in *T. brucei* has been linked to resistance to isometamidium (Baker et al., 2015).

Using GO terms for quantitative enrichment has two important limitations. First, not all genes in the genome have been formally assigned a GO term. This includes many genes whose products have been characterised. For example, there is no GO term for AMP-activated kinase. Secondly, there are no GO terms associated with genes that have a GoF effect we observe in our data (see Section 10.2.4). We therefore considered an alternative way of quantitative enrichment – using co-publication of genes as an indicator of genes likely to be involved in a specific cellular process/compartment. This alternative method also allows us to define biological associations between GoF genes that were significantly enriched in both libraries (Figure 10.8, Table 10.3). Gene-publication associations were downloaded from TriTrypDB (v46) using data from *T. brucei* proteins as a reference. Among 43 publications associated with significant LoF ($p < 0.05$), the strongest hits (adj p <4e-5) were publications related to the flagellum, translation, transcription, RNA splicing and cytokinesis. Other significant hits were related to the proteasome and nuclear pore complex. The associated processes identified here are in agreement with the GO terms associated with significant LoF. I examined one of the strong hits (PMID 30030284, Z = -9.08, adj p <4e-5) which is a publication investigating the association between the distribution of axonemal proteins and flagellum motility. This paper shows that there is an inherent asymmetrical distribution of the proximal

(Tb927.8.4400, Tb927.7.5660) and distal (Tb927.5.1900 and Tb927.11.16090) outer dynein arm-docking complexes along the axonemes of *T. brucei* and *Leishmania mexicana.* This asymmetry is involved in regulation of flagellar beat, and is maintained through IFT. Knockdown of IFT46 (Tb927.6.3100) involved in retrograde transport changes the asymmetry of the docking complexes, resulting in cells with shorter flagella, cytokinesis defects, and reduced population growth (Edwards et al., 2018). Again, this strongly supports the idea that axonemal transport and, as a result, flagellar motility is important to *T. congolense.*

Not only did we see PMIDs associated with LoF, but there were also sets associated with apparent GoF (adj $p < 0.05$). Three publications look at genes involved in life cycle regulation and differentiation. The publication with the strongest hit (mean = 2.97, adj p = 5.8e-4, PMID 27732844) examines the role of AMP-activated protein kinases (AMPK) in the transition of *T. brucei* from the proliferative stage to the quiescent stage. As discussed in Section 10.2.4, activation of the AMPKα pathway induces transition to non-proliferative forms; and inhibition with an AMPK inhibitor abolishes this differentiation *in vivo*, enabling parasites to continue proliferation (Saldivia et al., 2016). These data strongly suggest similar cell cycle regulation between *T. congolense* and *T. brucei.*

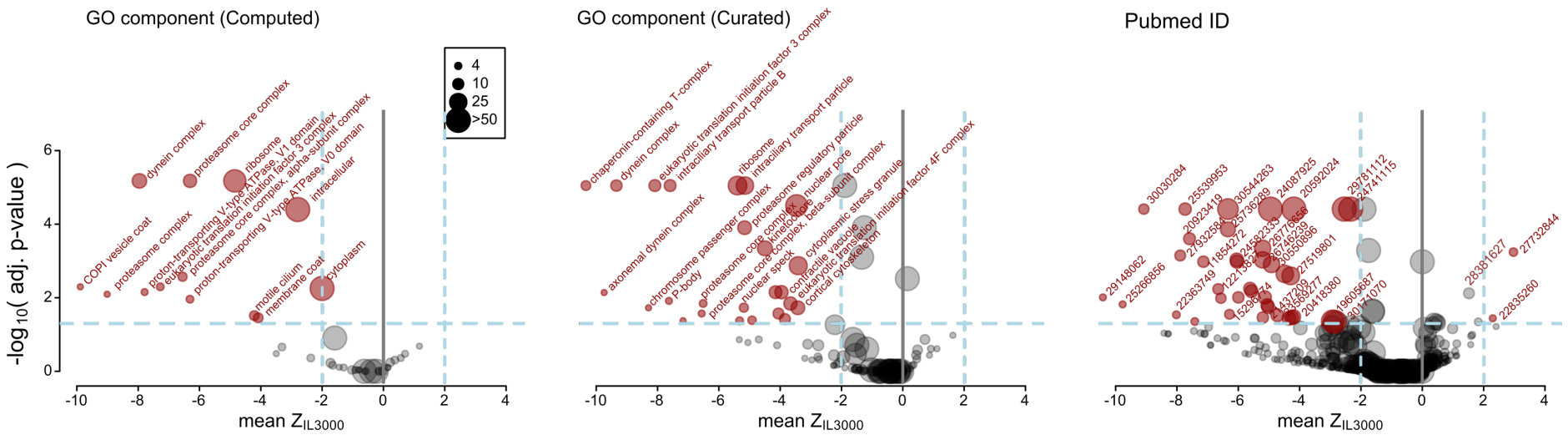


Figure 10.8: Quantitative enrichment analysis. Biological processes and locations enriched in *T. congolense* RNAi

GO ID	Term name	No. genes	Mean Z	Adj. p-val
GO:0030286	dynein complex	18	-7.96	<1e-5
GO:0005839	proteasome core complex	15	-6.31	<1e-5
GO:0005840	ribosome	43	-4.85	<1e-5
GO:0005622	intracellular	63	-2.79	4.0e-5
GO:0019773	proteasome core complex, alpha-subunit complex	7	-6.55	2.7e-3
GO:0030126	COPI vesicle coat	3	-9.89	5.1e-3
GO:0005852	eukaryotic translation initiation factor 3 complex	5	-7.28	5.1e-3
GO:0005737	cytoplasm	73	-2.00	5.6e-3
GO:0033180	proton-transporting V-type ATPase, V1 domain	4	-7.79	7.0e-3
GO:0000502	proteasome complex	3	-9.01	8.0e-3
GO:0033179	proton-transporting V-type ATPase, V0 domain	5	-6.31	1.1e-2
GO:0031514	motile cilium	8	-4.21	3.1e-2
GO:0030117	membrane coat	8	-4.08	3.5e-2

Table 10.1: Quantitative enrichment of computed GO components associated with > 2 genes in *T. congolense* genome-wide DRiF-Seq. A total of 39 terms are associated with > 2 genes. Only terms with adjusted p-value ≤ 0.05 are shown. Adjusted p-values < 1e-5 indicate sets where 0 of 10^6 permutations produced values as or more extreme than the mean Z.

GO ID	Term name	No. genes	Mean Z	Adj. p-val
GO:0005832	chaperonin-containing T-complex	8	-10.34	<1e-5
GO:0030286	dynein complex	11	-9.34	<1e-5
GO:0005852	eukaryotic translation initiation factor 3 complex	12	-8.09	<1e-5
GO:0030992	intraciliary transport particle B	12	-7.59	<1e-5
GO:0005840	ribosome	30	-5.39	<1e-5
GO:0030990	intraciliary transport particle	26	-5.15	<1e-5
GO:0005930	axoneme	274	-1.89	<1e-5
GO:0005643	nuclear pore	41	-3.47	3.1e-5
GO:0005838	proteasome regulatory particle	16	-5.16	1.3e-4
GO:0005634	nucleus	837	-1.26	1.3e-4
GO:0000776	kinetochore	19	-4.50	4.5e-4
GO:0005654	nucleoplasm	520	-1.34	7.9e-4
GO:0010494	cytoplasmic stress granule	28	-3.41	1.3e-3
GO:0000331	contractile vacuole	14	-4.15	7.1e-3
GO:0043234	protein complex	15	-3.96	7.1e-3
GO:0005858	axonemal dynein complex	3	-9.74	7.2e-3
GO:0000932	P-body	4	-7.63	1.2e-2
GO:0005839	proteasome core complex	5	-6.51	1.4e-2
GO:0016281	eukaryotic translation initiation factor 4F complex	15	-3.67	1.4e-2
GO:0016607	nuclear speck	7	-5.19	1.9e-2
GO:0032133	chromosome passenger complex	3	-8.29	1.9e-2
GO:0030863	cortical cytoskeleton	16	-3.42	1.9e-2

GO:0019774	proteasome core complex, beta-subunit complex	4	-6.56	2.7e-2
GO:0070565	telomere-telomerase complex	10	-4.06	2.7e-2
GO:0031514	motile cilium	10	-3.84	3.8e-2
GO:0005847	mRNA cleavage and polyadenylation specificity factor complex	6	-4.92	4.1e-2
GO:0097165	nuclear stress granule	3	-7.17	4.2e-2
GO:0044611	nuclear pore inner ring	5	-5.32	4.2e-2

Table 10.2: Quantitative enrichment of curated GO components (of 124) associated with > 2 genes in *T. congolense* genome-wide DRiF-Seq. Adjusted p-values < 1e-5 indicate sets where 0 of 10^6 permutations produced values as or more extreme than the mean Z.

Pubmed ID	Title	No. genes	Mean Z	Adj. p-val
Positive Z				
27732844	The AMPKalpha1 Pathway Positively Regulates the Developmental Transition from Proliferation to Quiescence in *Trypanosoma brucei*.	6	2.97	5.8e-4
28381627	Characterization of RBP9 and RBP10, two developmentally regulated RNA-binding proteins in *Trypanosoma brucei*.	10	1.52	7.6e-3
22835260	In silico analysis of trypanosomatids' helicases.	4	2.30	3.6e-2
25233456	Genome of the avirulent human-infective trypanosome--*Trypanosoma rangeli*.	5	1.82	4.5e-2
Negative Z				
30030284	Direction of flagellum beat propagation is controlled by proximal/distal outer dynein arm asymmetry.	9	-9.08	<4e-5
25539953	The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates--identification of conserved and divergent features based on orthologue analysis.	13	-7.73	<4e-5
30544263	Unexpected diversity in eukaryotic transcription revealed by the retrotransposon hotspot family of *Trypanosoma brucei*.	37	-6.34	<4e-5
24087925	Functional characterization of the trypanosome translational repressor SCD6.	55	-4.95	<4e-5
20592024	Analysis of spliceosomal proteins in Trypanosomatids reveals novel functions in mRNA processing.	69	-4.19	<4e-5

29781112	Identification of TOEFAZ1-interacting proteins reveals key regulators of *Trypanosoma brucei* cytokinesis.	437	-2.55	<4e-5
24741115	Proteomic analysis of intact flagella of procyclic *Trypanosoma brucei* cells identifies novel flagellar proteins with unique sub-localization and dynamics.	576	-2.34	<4e-5
18242729	Differential expression of glycosomal and mitochondrial proteins in the two major life-cycle stages of *Trypanosoma brucei*.	850	-1.90	<4e-5
25736289	A dynamic coordination of flagellum and cytoplasmic cytoskeleton assembly specifies cell morphogenesis in trypanosomes.	19	-6.34	1.4e-4
20923419	Biochemical analysis of PIFTC3, the *Trypanosoma brucei* orthologue of nematode DYF-13, reveals interactions with established and putative intraflagellar transport components.	12	-7.59	2.5e-4
26776656	The Flagellum attachment zone: 'The cellular ruler' of trypanosome morphology.	23	-5.21	4.5e-4
16302277	Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*.	788	-1.75	5.2e-4
27932584	The eIF3 complex of *Trypanosoma brucei*: composition conservation does not imply the conservation of structural assembly and subunits function.	10	-7.89	7.1e-4
16525475	Flagellar motility is required for the viability of the bloodstream trypanosome.	16	-6.05	9.3e-4
26746239	Flagellar pocket restructuring through the *Leishmania* life cycle involves a discrete flagellum attachment zone.	23	-5.21	9.3e-4
11854272	An easily dissociated 26 S proteasome catalyzes an essential ubiquitin-mediated protein degradation pathway in *Trypanosoma brucei*.	11	-7.14	1.0e-3
24582333	Discovery of unconventional kinetochores in kinetoplastids.	15	-6.04	1.0e-3
30550896	Cellular landmarks of *Trypanosoma brucei* and *Leishmania mexicana*.	23	-4.92	1.2e-3
19525551	Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor.	26	-4.48	2.3e-3
27519801	Cilium transition zone proteome reveals compartmentalization and differential dynamics of ciliopathy complexes.	27	-4.29	2.5e-3
19070634	Identification and characterization of nuclear non-canonical poly(A) polymerases from *Trypanosoma brucei*.	14	-5.62	5.7e-3

12213827	Functional characterization of the 11 non-ATPase subunit proteins in the trypanosome 19 S proteasomal regulatory complex.	10	-6.65	5.8e-3
16621069	Characterization of RNA polymerase II subunits of *Trypanosoma brucei*.	14	-5.57	6.6e-3
29148062	Flagellar incorporation of proteins follows at least two different routes in trypanosomes.	4	-10.41	9.8e-3
18713834	Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169.	11	-6.00	9.8e-3
18094047	Intraflagellar transport and functional analysis of genes required for flagellum formation in trypanosomes.	15	-5.14	9.8e-3
21363968	High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome.	9	-6.57	1.0e-2
25266856	FACT plays a major role in histone dynamics affecting VSG expression site control in *Trypanosoma brucei*.	4	-9.77	1.5e-2
16962183	Biochemical characterization of *Trypanosoma brucei* RNA polymerase II.	14	-5.06	1.6e-2
25972344	Assembly and maintenance of the flagellum attachment zone filament in *Trypanosoma brucei*.	14	-5.03	1.8e-2
29555687	Proteomic analysis of the cell cycle of procylic form *Trypanosoma brucei*.	665	-1.62	2.3e-2
21483801	The cell cycle regulated transcriptome of *Trypanosoma brucei*.	666	-1.62	2.4e-2
15296774	More than one way to build a flagellum: comparative genomics of parasitic protozoa.	8	-6.29	2.8e-2
22363749	The orthologue of Sjogren's syndrome nuclear autoantigen 1 (SSNA1) in *Trypanosoma brucei* is an immunogenic self-assembling molecule.	5	-8.02	2.9e-2
23569277	Evolution of modular intraflagellar transport from a coatomer-like progenitor.	14	-4.73	3.0e-2
23145003	Alterations in DRBD3 ribonucleoprotein complexes in response to stress in *Trypanosoma brucei*.	18	-4.21	3.3e-2
21437209	Identification of contractile vacuole proteins in *Trypanosoma cruzi*.	11	-5.20	3.5e-2
16246139	An in silico analysis of trypanosomatid RNA polymerases: insights into their unusual transcription.	17	-4.26	3.5e-2
20418380	TbPRMT6 is a type I protein arginine methyltransferase that contributes to cytokinesis in *Trypanosoma brucei*.	16	-4.35	3.6e-2
19605687	SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes.	43	-2.93	4.4e-2

16882690	Conserved and specific functions of axoneme components in trypanosome motility.	5	-7.42	4.4e-2
30171070	The CIF1 protein is a master orchestrator of trypanosome cytokinesis that recruits several cytokinesis regulators to the cytokinesis initiation site.	45	-2.85	4.5e-2

Table 10.3: Quantitative enrichment of 622 publications referencing > 2 genes in *T. congolense* genome-wide DRiF-Seq. References in publications were to *T. brucei* TREU927 gene IDs and have been translated based on 1-to-1 orthology. Only publications with adjusted p-value ≤ 0.05 are shown. Adjusted p-values < 4e-5 indicate sets where 0 of 10^6 permutations produced values as or more extreme than the mean Z.

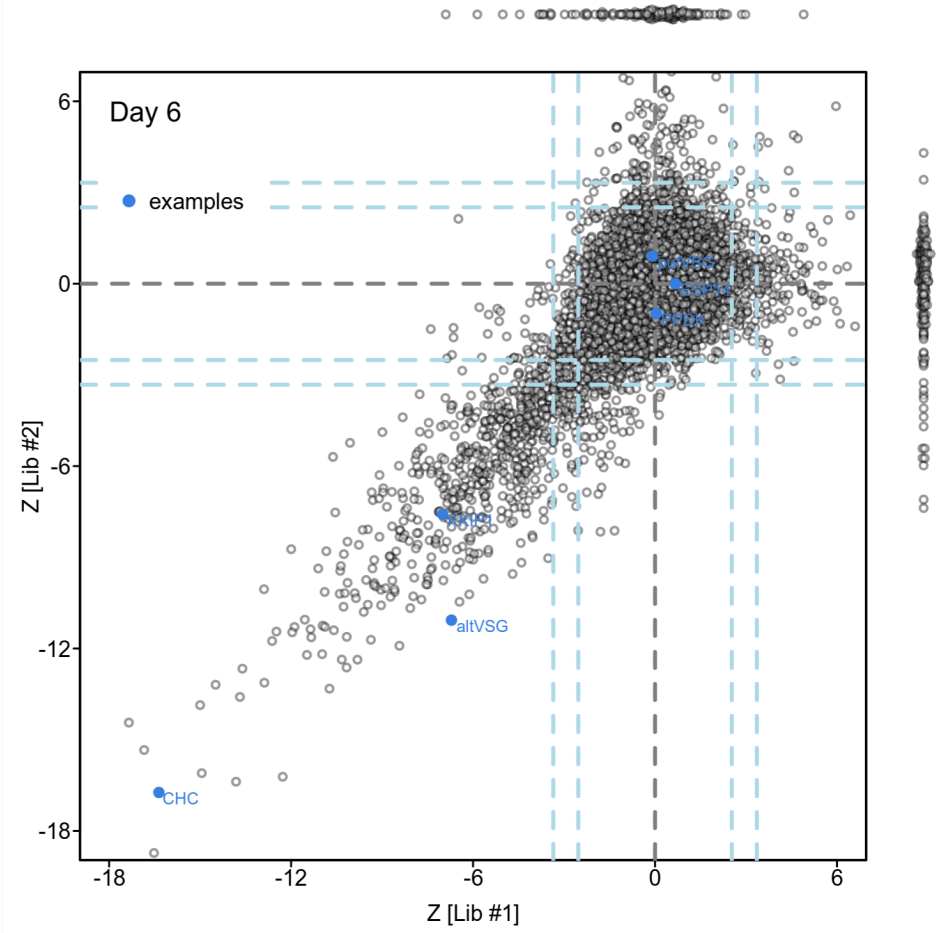
10.2.6 Differential outcome between individual gene knockdown and outcome in DRiF-Seq library

VSG

In Section 8.4, RNAi knockdown of the putative VSG (putVSG) did not show growth defect in culture. My data suggested that either this gene was not the major expressed *VSG* in our cell line, that depletion levels were not low enough to produce a sustained effect, or that depletion of the active *VSG* has no effect on growth of *T. congolense* in culture. To understand this further, we looked for fragment sequences targeting the *putVSG* in our libraries. RNAi against the *putVSG* had no effect on gene fragment abundance in *T. congolense* DRiF-Seq libraries. On the contrary, we found an *alternative VSG* (*altVSG*, TcIL3000.A.H_000310000 located at the telomere of chromosome 3) which showed strong LoF in both libraries (Figure 10.9A). Figure 10.9B shows mapping of individual fragments to the *putVSG* and *altVSG* in both libraries. A plot of how the different fragments are spread around the median over 6 days of induction is also shown in Figure 10.9C. The data clearly show that knockdown of *putVSG* in bloodstream-form *T. congolense* does not decrease fragment abundance. Conversely, knockdown of the *altVSG* shows a decrease in abundance of fragments mapping to its coding sequence, resulting in a significant LoF upon RNAi induction. The associated LoF suggests the *altVSG* may be the major VSG in this cell line. Total RNA sequencing of six *T. congolense* lines we have generated confirmed that the *altVSG* is the most highly expressed in all lines (data not shown), establishing the *altVSG* as the

major expressed VSG in *T. congolense* cells grown in our labs, including TcoTTS used to generate DRiF-Seq library experiments.

A



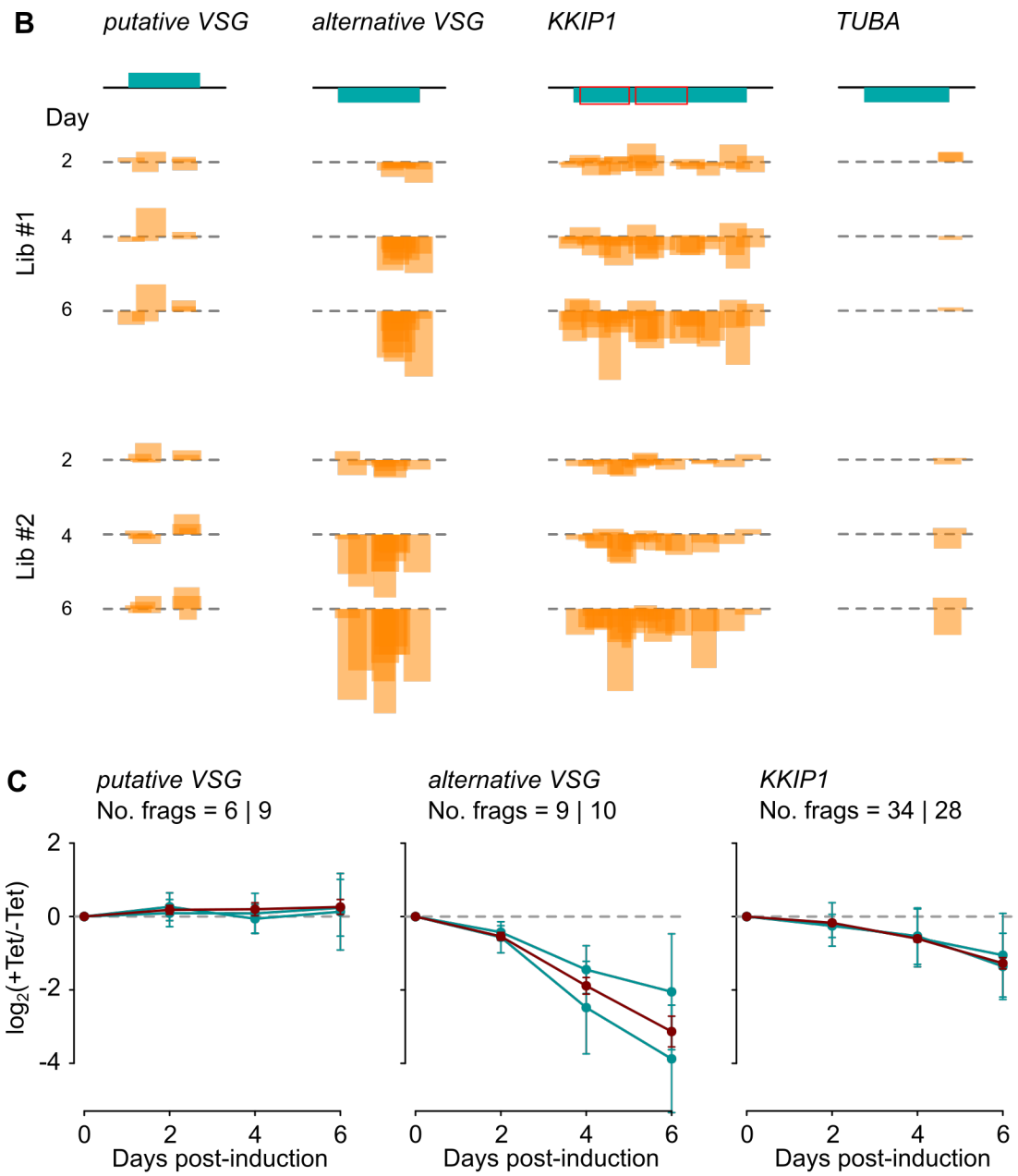


Figure 10.9: Differential outcome of individual gene knockdowns and knockdown in *T. congolense* libraries. A. Plot of day 6 Z values of genes in two *T. congolense* RNAi libraries. Annotations are genes that were previously knocked down or attempted to knock down individually. *altVSG*, alternative VSG; *putVSG*, putative VSG; *CHC*, clathrin heavy chain. Circles outside of square enclosure are genes present in Lib #1 (right) or Lib #2 (top) only. **B.** Mapping of RNAi fragments (orange) to coding sequence (turquoise) of genes at different days of induction of both libraries. **C.** Plots showing the spread of different fragments of genes in individual libraries and combined data from both libraries over 6 days of induction.

KKIP1

Individual RNAi against KKIP1 (using two different target sequences, red boxes in Figure 10.9B) did not produce a detectable morphological or growth defect in cells after 72 h of tetracycline induction (see Section 8.4). In DRiF-Seq however, RNAi

fragments targeting *KKIP1* produced a significant decrease in fragment abundance (Z = -6.952, FDR = 10e-9.7; Z = -7.572, FDR = 10e-11.6 for Lib 1 and 2 respectively) after 6 days of induction (Figure 10.9A). RNAi fragments targeting almost all ORF of *KKIP1* resulted in a decrease in fragment abundance on induction (Figure 10.9B). There was a gradual decrease in median fragment abundance – 1.2-fold on day 2, 1.6-fold on day 4 and 2.3-fold on day 6 (Figure 10.9C). This is exactly what one would expect from ablation of such an essential kinetochore component. The lack of detectable phenotype when individually ablating KKIP1 likely reflects clone-to-clone variation. If repeating that experiment, I would probably interrogate many more clones; although never as many as a library, in-competition experiment could do it once.

Tubulin

Individual knockdown of α-tubulin produced a phenotypic and strong growth defect in *T. congolense* (see Section 8.4). I was therefore expecting α-tubulin to produce a strong loss of fitness profile in my libraries. Interestingly, this was not the case; the α-tubulin gene did not show a significant decrease in abundance over the period of induction in the DRiF-Seq libraries. We found that in both libraries very few fragments targeted α-tubulin, and all fragments targeted a specific sequence at the C-terminal end of the gene (Figure 10.9B). Over the period of induction shown, there was little fold change (FDR > 0.5) in abundance of RNAi fragments targeting this region. It is unlikely that we lost all other RNAi fragments targeting other areas along the gene by chance. A more plausible explanation for this inconsistency between individual knockdown and library screen is leakiness of the DNA construct. A leaky construct could have led to loss of clones targeting α-tubulin at regions that produce strong loss of fitness prior to addition of tetracycline to the culture medium. This would explain why only clones where RNAi targeting does not produce a high effect are present at the time of sampling the libraries. If this assessment is right, it is possible that we are losing other genes with very strong fitness cost because of leakiness (See Section 10.2).

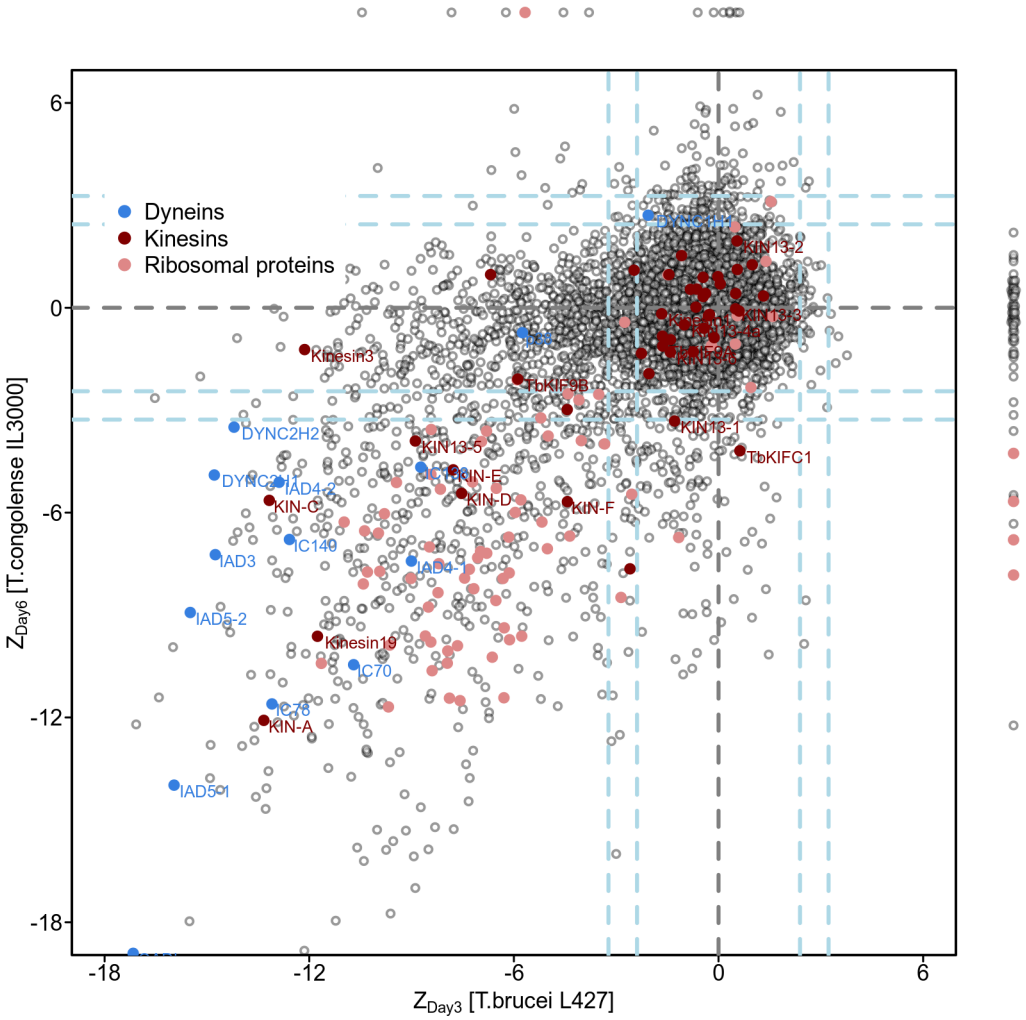
10.2.7 Investigating differential biology: *T. congolense* x *T. brucei*

For the first time we have the tools to investigate differences between *T. brucei* and *T. congolense* on a genome-wide scale. As part of a collaborative project between the Gadelha and Wickstead labs, we have previously generated a DRiF-Seq library of *T. brucei* Lister 427 strain (Brusini, 2019). To investigate RNAi-mediated fitness differences between *T. brucei* and *T. congolense,* we compared DRiF-Seq experiment end-points (when Z values are the strongest) between the two parasites (which represent Day 3 for *T. brucei* and Day 6 for *T. congolense*), and annotated genes encoding ribosomal proteins, molecular motors and cell cycle proteins (Figure 10.10).

We found that RNAi against axonemal dyneins is associated with strong LoF in both *T. congolense* and *T. brucei*, indicative of the importance of maintaining flagellar function to both parasites (Figure 10.10). On the other hand, cytoplasmic dynein 2 (DYNC2H1 and DYNC2H2, associated with intraflagellar transport) produced a stronger LoF in *T. brucei.* This may be because *T. brucei* has a longer flagellum that extends past the cell body for 1-2 μm, possibly requiring a greater investment in IFT (*T. congolense* flagellum does not extend beyond cell body anterior end). The difference in LoF could also be due to the differential RNAi penetrance between the two species. Quantitative enrichment analysis (see below; Figure 10.11) however, showed that genes associated with the GO term "dynein" do not produce differential effect in *T. congolense* and *T. brucei.* Thus, the observed difference in LoF of dyneins is more likely a result of differential RNAi penetrance. The fact that curated GO analysis shows enrichment of the 26 genes associated with GO term "intraciliary transport particle" in *T. brucei* ($\Delta Z = -2.46$) is unlikely to be due to chance (adj. $p = 3.6e\text{-}4$), and also indicate that *T. brucei* is more invested in IFT than *T. congolense.*

The kinesins have similar LoF profiles in both parasites. It is interesting though that Kinesin-3 – a fast plus-end translocator important for transport of vesicles – is associated with LoF only in *T. brucei.* In Section 4.2.2, RNAi against *TbKIN3* was associated with enlarged flagellar pocket characteristic of stalled endocytosis. High rate of endocytosis occurs in *T. brucei* as a result of VSG coat recycling on the cell

surface (Engstler et al., 2004; Seyfang et al., 1990). It is possible that knockdown of Kinesin-3 in *T. congolense* had no effect because VSG shedding on the surface of *T. congolense* is perhaps slower than in *T. brucei,* and may not need to use a fast motor like Kinesin-3 for the process of VSG recycling. However, there is no direct comparative data to prove or otherwise that there is any difference in endocytic rate between different species of the African trypanosomes. When investigated with differential quantitative enrichment analysis, we did not find kinesins generally to be different between the two parasites (Figure 10.11).



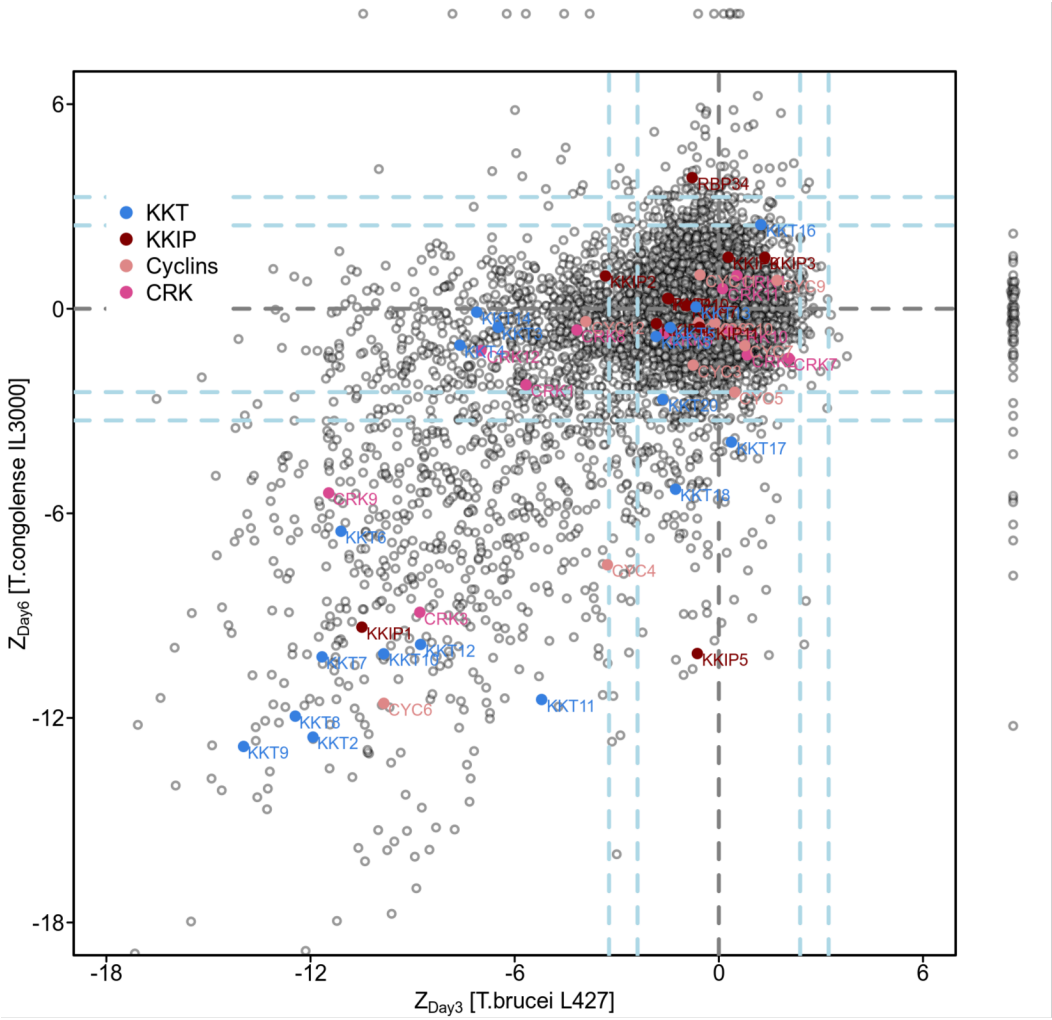


Figure 10.10: Analysis of differential biology between *T. congolense* and *T. brucei.* Plot of induction day 6 Z-values of genes of *T congolense* RNAi library against induction day 3 Z-values of genes of *T. brucei* Lister 427 RNAi library (Brusini, 2019) datasets. Annotation of groups of genes showing different fitness profiles. Circles outside of square enclosure are genes targeted in *T. brucei* (right) or *T. congolense* (top) only.

Differences may exist in cell cycle regulation of T. brucei and T. congolense

Most cell cycle proteins show similar LoF in both *T. brucei* and *T. congolense* (Figure 10.10). Few exceptions – KKIP5, CYC4 and KKT17 and KKT18 are associated with LoF in *T. congolense* only. At present, it is unclear why that is, and what it means for the cell cycle of this species. KKIP5, expressed from S-phase to metaphase, and lost at anaphase (D'Archivio and Wickstead, 2017), has recently been shown to help maintain genome integrity by involvement at the G2/M cell cycle checkpoint. In insect-form *T. brucei*, RNAi-mediated depletion of KKIP5 showed moderate growth defect, causing minimal DNA damage-induced metaphase arrest, which resulted in chromosome mis-

segregation and aneuploidy (Zhou et al., 2019). However, in bloodstream-form *T. brucei* DRiF-Seq, depletion of KKIP5 shows no obvious growth defect. CYC4 is also suggested to play a role in G1/S cell cycle checkpoint (Li and Wang, 2003). Taken together, the data suggest that there may be some fundamental differences in cell cycle regulation between *T. brucei* and *T. congolense*.

The number of genes associated with gain of fitness is greater in T. congolense than in T. brucei

Comparing our two-species DRiF-Seq datasets, there are fewer genes associated with GoF in *T. brucei* than *T. congolense*. Seven genes causing apparent GoF (at FDR < 0.01) when depleted in *T. congolense* were discussed in Section 10.2.4. In comparison, there was no gene associated with GoF in *T. brucei* at FDR < 0.01 (Figure 10.10). At an FDR of 0.1, only 3 genes are associated with GoF in *T. brucei.* Taken together, the data suggest that *in vitro* growth of *T. congolense* is directly influenced by several genes.

Comparative GO term enrichment T. congolense x T. brucei

We performed quantitative enrichment analysis of computed and curated GO terms, as well as gene sets appearing in publications, to look for differential effect of genes involved in specific biological processes (Figure 10.11). The off-diagonal distances of the Z-values and the Benjamini-Hochberg adjusted p-values are presented in Tables 10.4, 10.5 and 10.6. The data suggest there are more terms and publications enriched in *T. brucei* than *T. congolense* (Figure 10.11). The GO terms "COPII vesicle coat" (ΔZ = -5.81, adj. p = 3.6e-3) and "prefoldin complex" (ΔZ = -4.21, adj. p = 1.8e-2) are enriched in *T. brucei* (see Section 10.2.6), suggesting that COPII-mediated transport is more important for *T. brucei.* Only the 5 genes associated with the GO term "V-type ATPase, V0 domain" was more enriched in *T. congolense* (ΔZ = 2.83, adj. p = 3.6e-3).

The outcome of quantitative enrichment of genes occurring in publications supported the enriched GO terms. Publications that are more enriched in *T. brucei* (adj. p < 0.05) covered processes involving translation, RNA splicing, nuclear pore complex, IFT, endocytosis, secretory pathway and the proteasome (Table 10.6). One paper (PMID 19759175) that is highly enriched in *T. brucei* (ΔZ = -5.81) unlikely due to chance

(adj. p = 3.5e-3) examined the role of COPII machinery in ER exits of specifically GPI-anchored cargo in bloodstream forms. The paper showed that RNAi depletion of either of the paralogues of COPII subunits – TbSec23.1 (Tb927.8.3660), TbSec23.2 (Tb927.10.7740), TbSec24.1 (Tb927.3.1210) and TbSec24.2 (Tb927.3.5420) was lethal to cells. After 12 h of RNAi induction, a severe reduction in trafficking of VSG (5.5 and 4-fold) was observed in cells depleted of TbSec23.2 or TbSec24.1 respectively. A similar observation (~2.7-fold reduction) was made in the trafficking of another GPI-anchored protein. These data showed a preferential loading of GPI-anchored cargo into COPII vesicles (Sevova and Bangs, 2009). The fact that *T. brucei* VSG coat is very dense, and trafficking of VSG is selectively mediated by COPII, supports why genes associated with COPII transport are more enriched in *T. brucei*. This also concurs with the possibility that membrane turnover may be faster in *T. brucei* than in *T. congolense,* which agrees with kinesin-3 knockdown in *T. congolense* not being associated with LoF (see above; Figure 10.10). The highest hit enriched in *T. congolense* over *T. brucei* (PMID 23899193, ΔZ = 3.58, adj. p = 9.2e-4) is a publication that investigated the contribution of threonine to acetate and fatty acid production in insect-form *T. brucei*. The paper showed that threonine is the main carbon source for fatty acid synthesis in procyclics, contributing ~2.5 times higher than glucose (Millerioux et al., 2013). As the gene sets within this paper are those involved in threonine degradation, it suggests that *T. congolense* may be more invested in this pathway than *T. brucei* during the bloodstream stage and indicates a possible fundamental difference in metabolism.

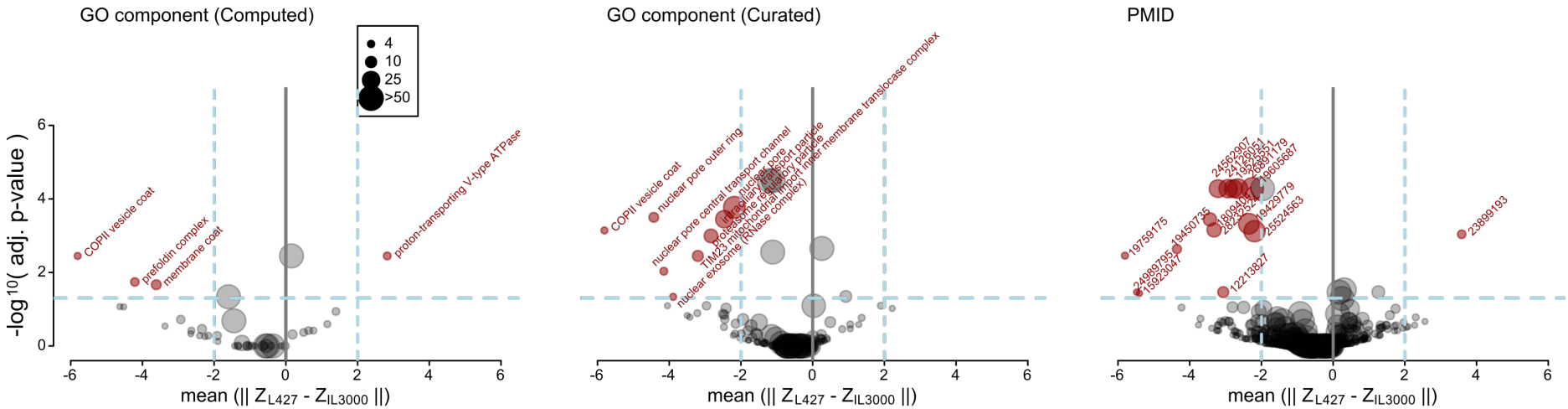


Figure 10.11: Quantitative enrichment analysis showing differential biological processes and locations enriched in *T. congolense* and *T. brucei* Lister 427 RNAi (Brusini, 2019)

GO ID	Term name	No. genes	ΔZ	Adj. p-val
Positive ΔZ				
GO:0033179	proton-transporting V-type ATPase, V0 domain	5	2.83	3.6e-3
Negative ΔZ				
GO:0030127	COPII vesicle coat	4	-5.81	3.6e-3
GO:0016272	prefoldin complex	6	-4.21	1.8e-2
GO:0030117	membrane coat	8	-3.61	2.2e-2

Table 10.4: Quantitative enrichment analysis for differential effects of computed GO components associated with > 2 genes in *T. congolense* and *T. brucei* DRiF-Seq. 'ΔZ' indicates the off-diagonal distance for the Z-values in *T. congolense* and *T. brucei*, $||Z_{L427} - Z_{IL3000}||$. Only terms with adjusted p- value ≤ 0.05 are shown.

GO ID	Term name	No. genes	ΔZ	Adj. p-val
GO:0005643	nuclear pore	41	-2.18	1.7e-4
GO:0031080	nuclear pore outer ring	8	-4.43	3.2e-4
GO:0030990	intraciliary transport particle	26	-2.46	3.6e-4
GO:0030127	COPII vesicle coat	4	-5.81	7.2e-4
GO:0005838	proteasome regulatory particle	16	-2.84	1.0e-3
GO:0005744	TIM23 mitochondrial import inner membrane translocase complex	10	-3.20	3.6e-3
GO:0044613	nuclear pore central transport channel	5	-4.15	9.3e-3
GO:0000176	nuclear exosome (RNase complex)	4	-3.89	4.6e-2

Table 10.5: Quantitative enrichment analysis for differential effects of curated GO components associated with > 2 genes in *T. congolense* and *T. brucei* DRiF-Seq. 'ΔZ' indicates the off-diagonal distance for the Z-values in *T. congolense* and *T. brucei*, $||Z_{L427} - Z_{IL3000}||$. Only terms with adjusted p- value ≤ 0.05 are shown.

Pubmed ID	Title	No. genes	ΔZ	Adj. p-val
Positive ΔZ				
23899193	The threonine degradation pathway of the *Trypanosoma brucei* procyclic form: the main carbon source for lipid biosynthesis is under metabolic control.	6	3.58	9.2e-4
19070634	Identification and characterization of nuclear non-canonical poly(A) polymerases from *Trypanosoma brucei*.	14	1.26	3.4e-2
Negative ΔZ				
24562907	Genetic validation of aminoacyl-tRNA synthetases as drug targets in *Trypanosoma brucei*.	27	-3.21	<5e-5

24126051	A multiple aminoacyl-tRNA synthetase complex that enhances tRNA-aminoacylation in African trypanosomes.	29	-2.93	<5e-5
19525551	Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor.	27	-2.78	<5e-5
26891179	Interactome mapping reveals the evolutionary history of the nuclear pore complex.	33	-2.66	<5e-5
19605687	SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes.	42	-2.25	<5e-5
20581293	The pre-mRNA splicing machinery of trypanosomes: complex or simplified?	73	-1.97	<5e-5
18094047	Intraflagellar transport and functional analysis of genes required for flagellum formation in trypanosomes.	15	-3.44	3.6e-4
19429779	Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors.	37	-2.35	4.7e-4
28232524	Lineage-specific proteins essential for endocytosis in trypanosomes.	17	-3.32	7.0e-4
25524563	The spliceosomal PRP19 complex of trypanosomes.	39	-2.19	7.6e-4
19450735	DRBD1 is the *Trypanosoma brucei* homologue of the spliceosome-associated protein 49.	7	-4.36	2.3e-3
19759175	Streamlined architecture and glycosylphosphatidylinositol-dependent trafficking in the early secretory pathway of African trypanosomes.	4	-5.81	3.5e-3
24989795	The intraflagellar transport dynein complex of trypanosomes is made of a heterodimer of dynein heavy chains and of light and intermediate chains of distinct functions.	3	-5.48	3.4e-2
12213827	Functional characterization of the 11 non-ATPase subunit proteins in the trypanosome 19 S proteasomal regulatory complex.	10	-3.06	3.4e-2
15923047	mRNA splicing in *Trypanosoma brucei*: branch-point mapping reveals differences from the canonical U2 snRNA-mediated recognition.	3	-5.40	3.7e-2

Table 10.6: Quantitative enrichment for differential effects of genes occurring in 622 publications referencing > 2 genes. References in publications were to *T. brucei* TREU927 gene IDs and have been translated based on 1-to-1 orthology. 'ΔZ' indicates the off-diagonal distance for the Z-values in *T. congolense* and *T. brucei*, $||Z_{L427} - Z_{IL3000}||$. Only publications with adjusted p-value ≤ 0.05 are shown. Adjusted p-values < 5e-5 indicate sets where 0 of 10^6 permutations produced values as or more extreme than the mean Z.

10.3 Isometamidium resistance and mode of action studies

Understanding the mechanism of drug mode of action and resistance helps the identification of novel drug targets and guide identification of new therapeutics to

minimize resistance. Effective combination therapy of existing drugs is one of the immediate impacts of having a good understanding of these mechanisms.

Isometamidium (ISM) is the most important veterinary drug for treating *T. congolense* infections either alone or as mixed infection with another *Trypanosoma* spp. As such, *T. congolense* resistance to ISM is a cause for much concern. There is some evidence suggesting a possible mode of action and resistance of ISM in *T. congolense.* ISM was found to accumulate in the kinetoplast in both *T. brucei* and *T. congolense* (Boibessot et al., 2002; Wilkes et al., 1997) suggesting the organelle's possible involvement in drug resistance or mode of action. ISM accumulation in the kinetoplast has been suggested to be an energy-driven process (Sutherland and Holmes, *1993*). Resistant *T. brucei* isolates were found to have reduced mitochondrial membrane potential leading to less accumulation of ISM in the kinetoplast (Eze et al., 2016). Therefore, ISM resistance could develop from any or a combination of:

1. decreased diffusion through the membrane of the kinetoplast.
2. alteration of a drug transporter.
3. an increase in efflux rate of ISM from the kinetoplast

When analysing DRiF-Seq data, we paid particular attention to genes previously reported to be involved in drug resistance. ISM resistance in *T. brucei* has been linked to the ATP synthase, where a point mutation in the γ-subunit is associated with multi-drug resistance (Gould and Schnaufer, 2014). RNAi-depletion of subunits of the V-type ATPase also identified its involvement in ISM resistance, particularly of the $V_{0\text{-}c}$ and $V_{0\text{-}d}$ subunits (Baker et al., 2015). DRiF-Seq data show that knockdown of most V_1 subunits of the V-Type ATPase produces a LoF effect in both species, but there is evidence of differential effect between the subunits (Figure 10.12). Depletion of the V_0 subunits was associated with greater LoF in *T. congolense* than in *T. brucei* (Figure 10.11). These data support the mechanistic differences between isometamidium resistance in the parasites.

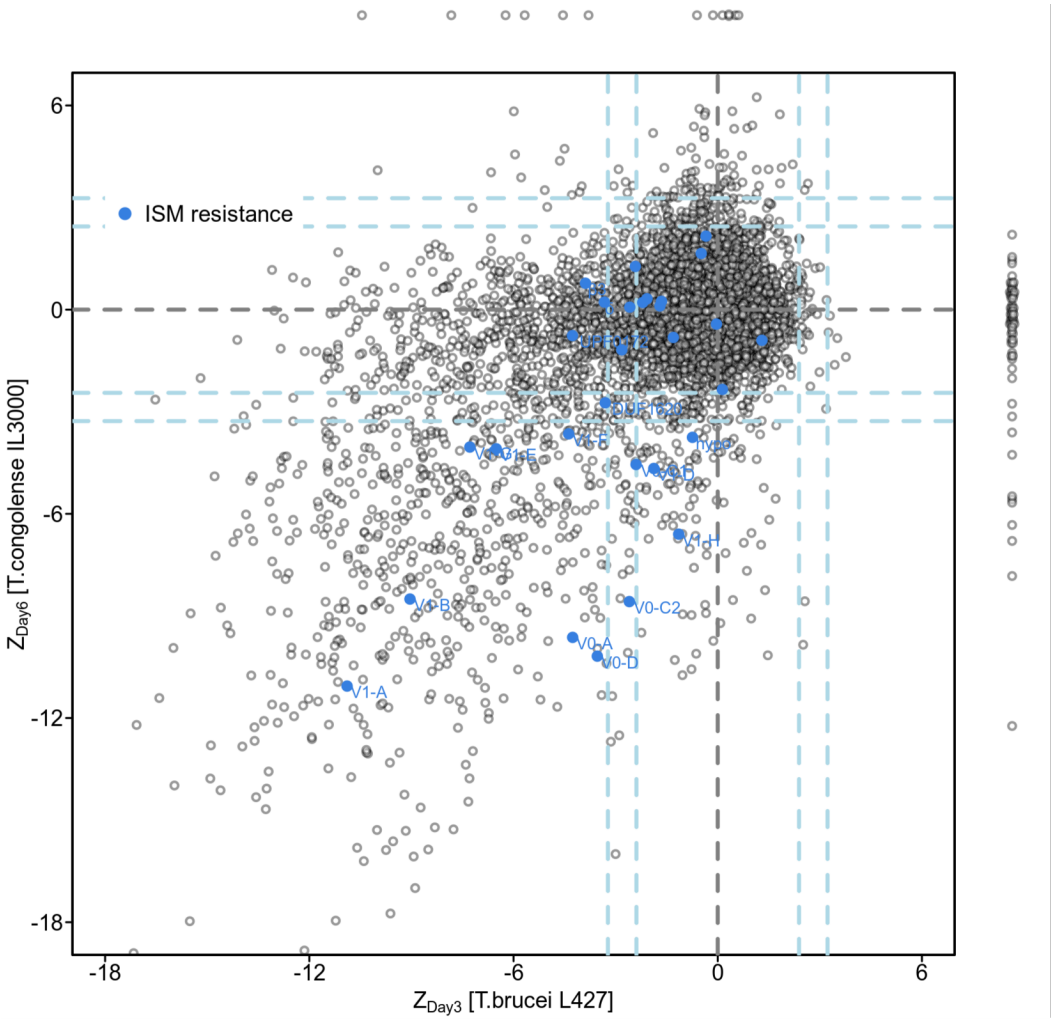


Figure 10.12: Differential response of genes associated with isometamidium resistance in *T. brucei*. Plot of induction day 6 Z-values of genes of *T. congolense* RNAi library against induction day 3 Z-values of genes of *T. brucei* Lister 427 RNAi library (Brusini, 2019) datasets. Annotation of orthologues of genes associated with ISM resistance in *T. brucei* shows their fitness profiles in both parasites. Circles outside of square enclosure are genes targeted in *T. brucei* (right) or *T. congolense* (top) only.

We used the DRiF-Seq method to interrogate resistance and mode of action of drugs in *T. congolense*. In this section, I describe how the *T. congolense* genome-wide RNAi library was used to study the mode of action (MoA) and resistance of ISM. Firstly, I assessed the response of bloodstream-form *T. congolense* IL3000 (TcoSM) to different concentrations of Veridium (Ceva Santé Animale, France) for different time periods, spanning nearly an order of magnitude either side of a previously-reported EC_{50} of ~1 nM at 48 h exposure to Veridium (Sahin et al., 2014). I tested drug concentrations from 0.13 nM to 6 nM at √10 dilutions and monitored growth over a period of 6 days, measuring the amount of inhibition against control (Figure 10.13A). There is a time

dependence to the response of cells to different drug concentrations. The effect of ISM on cells becomes considerably greater from day 3 onwards, which would prevent a reasonable length of time for an RNAi experiment. This means for a 6-day RNAi induction experiment in the presence of ISM, an impact on cells due to RNAi may only be detectable 3 days post-induction. Hence, there is effectively 3 days left in the experiment to measure a rescue or worsening impact as a result of RNAi. At 48 h, the EC_{50} of Veridium (ISM) was 1.5 nM (Figure 10.13B) which is not far from the published EC_{50}.

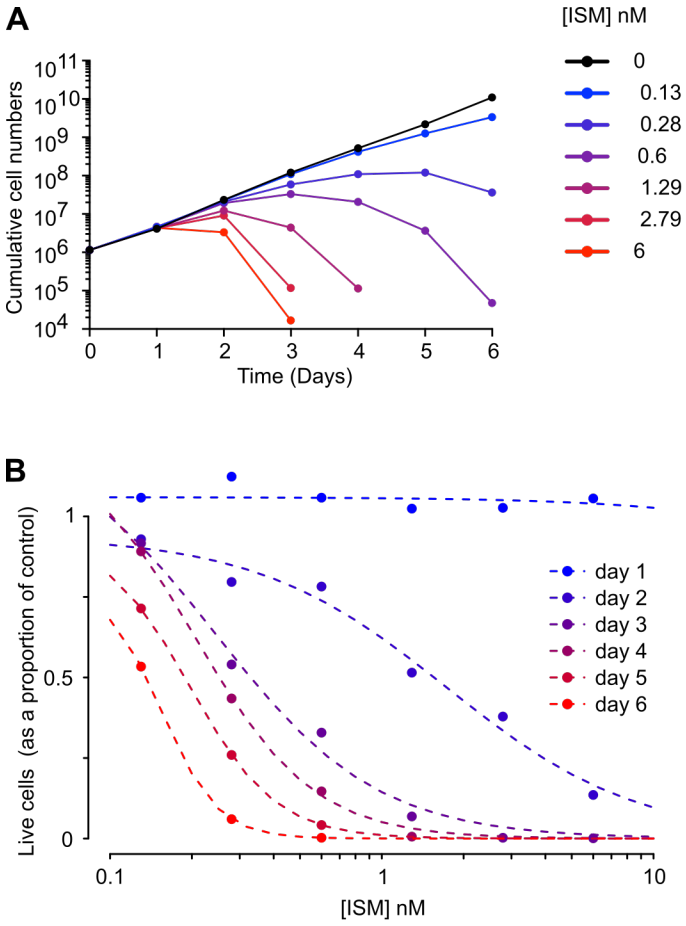


Figure 10.13: Response of trypanosome cells to different concentrations of isometamidium. **A.** Growth curve of *T. congolense* treated with different concentrations of ISM over a period of 6 days showing the time dependent response to ISM inhibition. **B.** An EC50 curve of *T. congolense* treated with the different concentrations of ISM over 6 days.

10.3.1 Screening the *T. congolense* genome for isometamidium sensitivity/resistance

To interrogate both resistance and sensitivity, I set up two DRiF-Seq experiments using the genome-wide RNAi Libraries 1 and 2 at two different concentrations of drug. A

low concentration of ISM (0.3 nM) which gives maximum sensitivity for detection of fold changes in both directions, i.e. either as gain or loss of fitness. This would allow detection of RNAi against genes that confer moderate resistance or increase sensitivity to ISM as a way of identifying genes involved in mode of action. A high concentration of ISM (1.2 nM) was also used to identify knockdowns that confer strong resistance. ISM was added at the same time as tetracycline to Libraries 1 and 2, which were kept as continuous cultures for 6 days (0.3 nM ISM) or 8 days (1.2 nM ISM) (Figure 10.14). Samples of ~250 copies per fragments (~$5x10^7$ cells) were taken on days 2, 4 and 6 for 0.3 nM ISM cultures, and on day 2 and 4 of 1.2 nM ISM cultures (day 8 samples were between $4x10^6$ – $1.7x10^7$ cells). gDNA preparation, NGS and analysis of this experiment are ongoing and do not form part of this thesis.

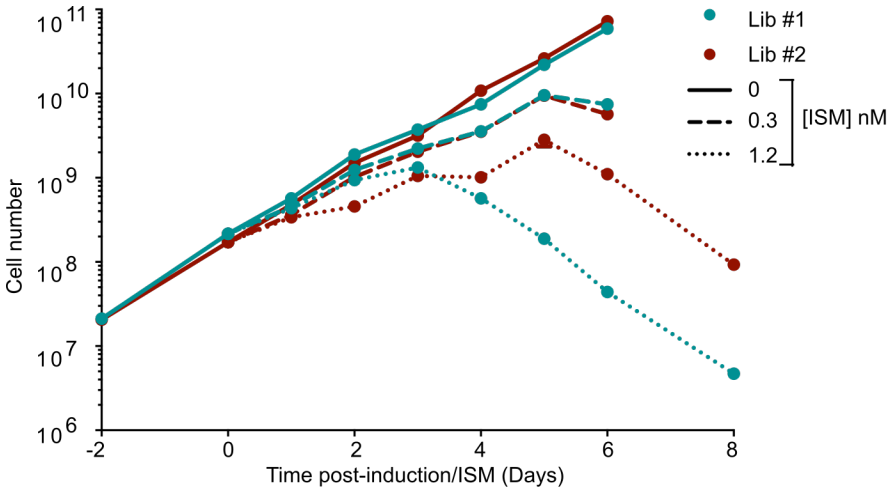


Figure 10.14: Screening of *T. congolense* libraries for genes involved in isometamidium mode of action and resistance. Growth curves of *T. congolense* RNAi Lib #1 and 2 treated with low (0.3 nM) and high (1.2 nM) concentrations of ISM with induction up to 8 days.

11 DISCUSSION II

The ability to study the biology of Animal African Trypanosomiasis (AAT) requires a suitable model organism selected for its representation and importance in the aetiology of the disease. *T. congolense* is potentially a good model for AAT, but as a model system it has had severe limitation in the area of genetic tools. When genetic technology becomes available for any organism, its genome becomes malleable. Among other things, it becomes easier and more desirable to exploit its unique biology to fill in the missing gaps in our understanding of life cycle processes, environmental adaptation, pathogenesis and virulence. For the first time, I have systematically developed tools for consistent high-efficiency transgene expression which, combined with inducible RNAi, enabled us to generate genome-wide library screens of gene fitness cost and drug studies in *T. congolense*.

11.1 Transgenesis in bloodstream-form *T. congolense*

Direct transgenesis of infective-form *T. congolense* was not easily achievable. Studies that used genetically-modified bloodstream-form (BSF) *T. congolense* either obtained the cells *in vitro* through differentiation of genetically-modified insect form through to BSF (Coustou et al., 2010) or to metacyclic form, then injected into mice for *in vivo* differentiation into BSF, which is then collected from mouse blood (Coustou et al., 2012). The authors provided no reason for the inability to transfect BSF, and this tedious, time-consuming process of differentiation from a modified fly stage became the default way of obtaining genetically-modified *T. congolense* BSF. In Chapter 7 I described how I developed and validated a direct method to genetically modify BSF *T. congolense* IL3000 strain. Importantly, despite genetic manipulation, the parasite remains infective to mouse models and produces the classic waves of parasitaemia characteristic of the disease.

11.2 RNAi penetrance in *T. congolense*

Here I have described the generation of RNAi tools and a means for directly transfecting DNA constructs in *T. congolense* BSF. I have also validated gene silencing in this lifecycle stage by demonstrating mRNA and protein depletion to 50% normal level, with relatively low number of cells (< 5%) showing the respective phenotype without RNAi induction (RNAi leakiness). When directly comparing knockdown effect of the same gene in the two species, the most effective RNAi-targeted depletion in *T. congolense* achieved less reduction in mRNA and milder phenotype than we observe in *T. brucei.* We therefore deduce that *T. congolense* has a lower penetrance of RNAi. Low penetrance be explained by two possible scenarios: the presence of a fraction of cells refractory to RNAi or the result of a threshold effect on the population.

11.2.1 Refractory cells

Within a population, there may be a proportion of cells that have found a way to break their RNAi machinery, such that they are not affected by RNAi induction. If 50% of cells were refractory, then the remaining 50% will be affected by RNAi. However, the scenario of having this large proportion of refractory does not fit the observation made, because once cells become refractory, they do not become sensitive again. Therefore, if half of the population was refractory, I will expect to see a quick recovery after induction as the affected cells die off. On the contrary, for the gene knockdowns presented here, I observe a similar proportion of cells (~50%) producing an abnormal phenotype at every generation. The implication is that every cell in the population could have a functional RNAi machinery, and whether a cell produced an abnormal phenotype in any generation is determined by some unknown factor that is probably epigenetic. Therefore, cells that were unaffected in one generation are likely to produce progeny that becomes affected in the next. Given that only half of the population are killed every 2 days, I could not observe the grow out of true refractory cells over the period of RNAi induction, and so cannot argue the absence of refractory cells but that the proportion present is less than 50%. I thus conclude that RNAi- resistant cells is not the true explanation for low RNAi penetrance.

11.2.2 Threshold effect on the cell population

Alternatively, the threshold effect considers what is happening at the level of protein in individual cells due to RNAi and its overall impact on a population. It assumes that there is a level of protein (hypothetically 50 molecules) below which any cell would become affected by RNAi, and that in parental cells with no integrated RNAi construct, the average level in the population is well above this threshold (100 molecules) with a range of 95 – 105 in a normal distribution (Figure 11.1). If we then assume that without tetracycline induction (0 h), the average protein level in RNAi cells is already down to 53 (58 – 48) as a result of leaky expression, then within this population, a small percentage of cells having < 50 molecules of protein will have an abnormal phenotype. These would represent the 5 % of the population that show abnormal phenotype without induction of RNAi against clathrin or tubulin. In this scenario, a 48 h induction period will only reduce the average level of protein to 50 (55 – 45) to tip 50% of the cells over the threshold (Figure 11.1). This would mean the actual RNAi induction produced little effect reducing the average protein by only 3% to cause 50% of cells in the population to be abnormal. On the contrary, I have shown that mRNA level in RNAi population without induction is higher (> 80%) than the 53% (i.e. 53 of 100 molecules) specified in this hypothesis, and tetracycline addition to the culture reduces the level by a significant amount. The observed leakiness at the mRNA level though suggest that a small number of cells have dropped considerably in protein level to produce the subpopulation with abnormal phenotype. Therefore, the threshold effect alone also does not explain the low RNAi penetrance in *T. congolense.* It will appear that what is happening is more complex and may be a combination of the two scenarios discussed plus the effect of an epigenetic factor.

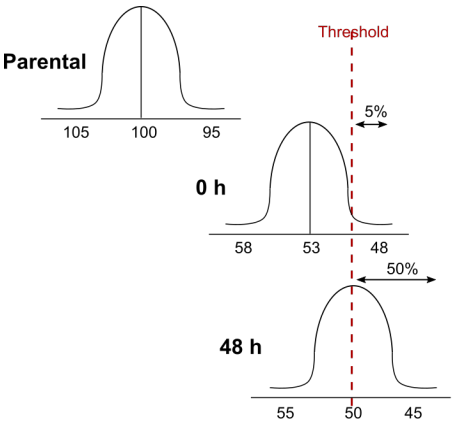


Figure 11.1: Schematic showing threshold effect on cell population

11.3 Optimal conditions to increase number of transformants

I have shown that we can use the I-SceI-mediated inducible double-strand break in *T. congolense* to increase stable transformants to > 50,000 per transfection. The number of transformants I obtained with BSF *T. congolense* is higher than what has been achieved in *T. brucei* Lister 427 using the same method. It would seem that, contrary to purported belief, BSF *T. congolense* is more efficient to transfect than *T. brucei.* In fact, we found that the limited dilution method following transfection appears to underestimate the number of transformants, since DRiF-Seq library experiments show a higher number of transformants (assessed on the basis of fragment counts). What this means is that there is something else going on during growth of *T. congolense* in wells that causes loss of some transformants. This phenomenon is similar to what occurs when culturing *T. congolense* at very low density. When there are very few cells present in medium, the culture usually takes longer than expected to grow, and if I was cloning in plates, I tend to have less than the expected number of positive wells. Adding some conditioned medium to low-density cultures seems to improve growth rate and increase the number of positive wells. This suggests that there is some growth stimulating factor secreted by cells into the medium. If this were so, then having fewer cells in culture (as in the case of cloning) would mean this growth factor will be highly diluted, affecting the survival rate of *T. congolense*.

I have asserted earlier that achieving certain milestones in numbers of transformants makes it possible for more complex experiments to be conducted (see Section 9.1). By virtue of increasing transformants to a level that allow genome-scale experiments, the complexity of research we can do with *T. congolense* has increased. DRiF-Seq transfection efficiency is so high that one could not only interrogate the entire genome, but also that against a set of genes. For instance, if wanting to study genes that affect kinases (~150 kinase genes vs ~10,000 genes in total), this would require $1.5x10^6$ transformants. Currently, this is achievable with 30 transfections of TcoTTS. A really exciting experiment, that of genome-squared (i.e genome vs. genome where genes are knocked down in all possible pairwise combinations) requires 10^8 transformants, which is not feasible with current technology. Such an experiment requires a few modifications and is the next exciting milestone that our labs could possibly attain. This milestone will require a 100-fold increase in the number of transformants. The current input of cells only has 1 in 10^4 cells transformed, and so there is an excess of cells that can still become transformed. Whether it is possible to transform more cells using a different technology like CRISPR/Cas9 depends on where the bottleneck of transformation lies. If the bottleneck is double-strand breaks, then a CRISPR/Cas9 technology may further increase transformants. As we have shown, I-SceI causes double-strand break in >70% of *T. congolense* cells, but CRISPR/Cas9 causes double-strand break in almost 100% of *T. brucei* cells expressing Cas9 (Rico et al., 2018). However, if the bottleneck is at the entry of DNA into cells, or the process of DNA recombination, then potentially CRISPR/Cas9 will not make any difference.

We find that *T. congolense* cell lines with inducible expression of I-SceI (TcoTTS) are stable in the same way as the analogous *T. brucei* lines generated in the Gadelha lab (427-TTS, AnTat-TTS). In both species, integration was targeted to loci on the minichromosome. Previously published *T. brucei* Sce* line, however, have to be generated each time the cell is required for an experiment, because of high instability and consequent drastic drop in transfection efficiency when freeze-thawed (Glover et al., 2015). Sce* drives construct integration to a ribosomal spacer. Being as the loci

targeted usually affect the stability of a transgene, this suggests that the differences in stability could be a result of the targeted loci.

11.4 Genome-wide *T. congolense* DRiF-Seq

I have applied the technology developed here to generate, for the first time, two genome-wide RNAi libraries to screen gene fitness cost in the most important animal trypanosome. I have also applied DRiF-Seq to interrogate mechanism of drug action and resistance. The number of possible applications our study has is huge, and colleagues have approached us with ideas for collaborative projects addressing basic parasite biology, vaccinology, and also as the means to identify resistance to new drugs currently in development.

11.4.1 Low RNAi penetrance in *T. congolense*

T. congolense shows a low penetrance of RNAi across all genes. This is true for individual gene knockdown as well as in populations of mutants. We do not know whether this is an intrinsic feature of the IL3000 strain or generally to do with the *T. congolense* species, or indeed if using a different DNA construct setup would improve penetrance. In *T. brucei,* designing the construct to produce a stem-loop RNAi fragment *in vivo* sometimes improves level of knockdown (Durand-Dubief et al., 2003). As an immediate attempt to improve penetrance, I did generate an RNAi construct that would produce a stem-loop targeting fragment in *T. congolense*, but this did not change RNAi penetrance (data not shown), and even if it had the construct would not be DRiF-Seq compatible.

DRiF-Seq could, in fact, be used to explore factors that affect penetrance, for instance, by using an individual RNAi line against an essential gene “X” as the recipient cells into which a genome-wide DRiF-Seq library is transfected. Then, upon tetracycline induction, individual RNAis that rescue or worsen the effect due to depletion of gene X mRNA could be identified by NGS and fragment read counts. Those genes that worsen the gene X knockdown effect might be genes that one could over-express to cause greater penetrance of RNAi.

Alternatively, a different strain of *T. congolense* may enable greater penetrance. In *T. brucei*, gene fitness cost has a ~20-fold difference between Lister 427 and EATRO1125 strains. Importantly, the DRiF-Seq technology is very robust and easily transferrable to other strains/species of African trypanosomes – so long as the strain/species in question is culturable *in vitro* and transfectable.

11.4.2 Differential biology between *T. congolense* and *T. brucei*

Even though the penetrance of RNAi is lower, the sensitivity of DRiF-Seq allows us to precisely and robustly quantify the effect of RNAi across the genome of *T. congolense.* Even more exciting is that, for the first time, we have the ability to make direct comparisons, at the genomic level, between species of African trypanosomes.

Overall core biology is similar between *T. congolense* and *T. brucei.* But there is clear evidence of differential biology between the two, some of which is already known, but most of which was revealed and dissected by DRiF-Seq. For example, the literature shows that hydrodynamic forces generated by the high motility of *T. brucei* cells drive the clearance of antibody-VSG complexes from the surface for endocytosis and recycling (Engstler et al., 2007). This may imply that as *T. congolense* has a slower motility than *T. brucei* (Bargul et al., 2016) there may be resulting difference in their endocytic rates. Indeed, our DRiF-Seq experiments reveal a clear signature for essentiality of endocytosis. Significantly, DRiF has pinpointed this difference to be mostly based on COPII-dependence in *T. brucei* specifically.

In the fly vector, *T. brucei* breaks down proline which is in abundance in haemolymph, to utilize it as the main carbon and energy source. In glucose-depleted medium, a similar metabolism takes place as amino acids serve as the source of carbon and energy. However, in culture medium where glucose is in excess, insect forms preferentially shift to glucose catabolism (Coustou et al., 2008; Lamour et al., 2005). When growing in such glucose-rich medium, threonine is the most consumed amino acid, serving as the main carbon source for lipid metabolism. In insect forms, threonine contributed 2.5 times more to lipid biosynthesis than glucose (Millerioux et al., 2013, 2018). In contrast, glucose contributes 1.4 times more acetate (a precursor

for lipid biosynthesis) than threonine does in BSF (Mazet et al., 2013). The comparison made between bloodstream-form *T. brucei* and *T. congolense* reveals that the latter has greater reliance on threonine as a carbon source than the former. This investment in threonine degradation is probably not related to culture medium composition, because the amount of threonine in HMI-9 (*T. brucei* BSF medium) is higher (0.798 mM) than the amount used when culturing bloodstream-form *T. congolense* (0.403 mM in TcBSF1 medium). Both media contain glucose, although HMI-9 has approximately twice more (25 mM) than TcBSF1. Hence *T. congolense* appears to prefer threonine as a carbon source.

Isometamidium is an important treatment drug and the only prophylaxis available for animal African trypanosomiasis. The drug is threatened by increasing resistance in the field. Mining the DRiF-Seq data for *T. congolense* orthologues of *T. brucei* genes associated with isometamidium resistance, supports the hypothesis that resistance mechanisms will be different between the two species. While V-type ATPases are involved in resistance in both instances, the differential effect of subunits depletion attest to a mechanistic difference in isometamidium resistance.

11.4.3 Quorum sensing

In addition to finding genes associated with fitness cost to *T. congolense,* DRiF-Seq identified genes that caused an apparent gain of fitness. The fold-change of knockdown of these genes suggest that they would otherwise have been missed in an individual RNAi. Thanks to the sensitivity of the DRiF-Seq method, we are able to gain insight into key biology aspects of, for example, quorum sensing in *T. congolense*.

In the mammalian host, *T. brucei* density-dependent differentiation of proliferative (slender) to transmissible, non-proliferative (stumpy) forms contributes to the chronicity of parasite infection (MacGregor et al., 2011, 2012). Stumpy formation is important in pathogenesis and disease transmission, and is mediated by quorum sensing. The quorum sensing signal – the stumpy induction factor or SIF – is transmitted to cells through the *Tb*GPR89 transporter on the surface membrane of slender forms, and triggers the process of differentiation into stumpy forms via the

SIF response pathway (Rojas et al., 2019). A genome-scale RNAi screen done on monomorphic (not sensitive to SIF) *T. brucei* cells and using an artificial inducer of stumpy formation, uncovered signalling components (kinases, phosphatases and RNA binding proteins) of the SIF response pathway (Mony et al., 2014). Density-dependent growth arrest has been observed in *T. congolense* too, but without the characteristic morphological transformation that takes place in *T. brucei* (Silvester et al., 2017). Additionally, the stumpy-specific cell surface marker PAD1 (Dean et al., 2009) remained unchanged in growth-arrested *T. congolense* (Silvester et al., 2018). Yet, the *T. congolense* orthologue of *TbHYP2* (a component of the SIF response) can rescue null mutants of *T. brucei*, suggesting some level of conservation of the quorum sensing signalling pathway between the two species (Silvester et al., 2017). mRNA profiling of cells in peak parasitaemia showed a distinct difference in gene regulation between *T. congolense* and *T. brucei* (Silvester et al., 2018), and the components that participate in *T. congolense* quorum sensing and the pathway remained then unknown.

Our DRiF-Seq study provides further evidence that there is a system of quorum sensing in *T. congolense* that is somewhat similar to the one artificially induced in monomorphic *T. brucei.* The signature of genes identified with associated gain of fitness include kinases and phosphatases much like in monomorphic *T. brucei.* Further investigation into the identified components of *T. congolense* quorum sensing pathway is needed to gain further insight into the differences and similarities between the different pathways.

11.4.4 Application of *T. congolense* genome-wide technology

An immediate application of a *T. congolense* DRiF-seq library will be to use it to screen for mode of action and resistance of treatment drugs. We used our DRiF-Seq libraries to directly investigate the action of isometamidium in *T. congolense* as a proof of principle. By treating cells with pre-determined appropriate amounts of the drug, we were able to directly detect genes that influence cells to persist in the presence of isometamidium (data not shown in this thesis) as was achieved for *T. brucei* (Alsford

et al., 2012). Colleagues have previously generated a isometamidium-resistant line through gradually increasing the concentration of drug in the culture medium over a period of 5 months (Tihon et al., 2017). The authors then compared the genetic changes between parental, isometamidium-sensitive cells and the newly-generated resistant ones. The method is, however, time inefficient, does not produce consistent outcomes that can confidently explain how drug resistance arise, and does not give insight into drug mode of action. Screening the whole genome for determinants of resistance is quicker and the results likely more reliable; in addition, with the right adjustment of drug concentration it may be possible to obtain some insight into mode of action. DRiF-Seq libraries could become a really important means of screening new drugs to predict possible resistance.

Further modifications to genome-wide screens can be applied to several other research questions. Classical forward genetics approaches in yeast and fruit flies take advantage of dominant-negative mutants. Here I have shown that one can achieve an adequate level of transgene overexpression with the DNA constructs I created. Truncated protein overexpression (either as individual protein or as a library of dominant negative mutants) could easily be adopted in place of RNAi. This could prove to be a way of circumventing the issue of RNAi penetrance discussed earlier.

DRiF can also be used to study epistasis at the genome level much like how Lorenzo Brusini used in a combination of a dominant-negative expression mutant of KKIP1 and genome-wide RNAi depletion, to investigate modulators of KKIP1 in *T. brucei.* This modification of the method, called eDRiF (for epistatic DRiF), was used as a way of identifying regulators of the trypanosome kinetochore (Brusini, 2019). The tools produced here also enable us to apply a similar method to study epistatic interactions in *T. congolense*.

One of the attractions to adopting *T. congolense* as a model organism for trypanosome research is the possibility of conducting screens across the differentiation process of the parasite. Unlike *T. brucei, in vitro* differentiation of *T. congolense* through all the life stages have been demonstrated. Coupled with DRiF, we are in the position to directly study the biological process of differentiation.

It is a known fact that disease pathologies of *T. brucei* and *T. congolense* differ. *T. congolense* has the tendency to adhere to host endothelial cells. This phenomenon is observed also in culture, as *T. congolense* cells adhere to plastic and glass, while *T. brucei* remains in suspension without strongly interacting with vessel surfaces. We can finally use the developed technology to look for genes that influence adherence and hence underly the pathology of *T. congolense.*

12 REFERENCE

Abbas, A.H., Pereira, S.S., D'Archivio, S., Wickstead, B., Morrison, L.J., Hall, N., Hertz-Fowler, C., Darby, A.C., and Jackson, A.P. (2018). The structure of a conserved telomeric region associated with variant antigen loci in the blood parasite *Trypanosoma congolense*. Genome Biol. Evol. *10*, 2458–2473.

Afewerk, Y., Clausen, P.H., Abebe, G., Tilahun, G., and Mehlitz, D. (2000). Multiple-drug resistant *Trypanosoma congolense* populations in village cattle of Metekel district, north-west Ethiopia. Acta Trop. *76*, 231–238.

Akiyoshi, B., and Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. Cell *156*, 1247–1258.

Alibu, V.P., Storm, L., Haile, S., Clayton, C., and Horn, D. (2005). A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *139*, 75–82.

Allen, C.L., Goulding, D., and Field, M.C. (2003). Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. EMBO J. *22*, 4991–5002.

Alsford, S., Wickstead, B., Ersfeld, K., and Gull, K. (2001). Diversity and dynamics of the minichromosomal karyotype in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *113*, 79–88.

Alsford, S., Kawahara, T., Glover, L., and Horn, D. (2005). Tagging a *T. brucei* RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. Mol. Biochem. Parasitol. *144*, 142–148.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-fowler, C., and Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res. *21*, 915–924.

Alsford, S., Eckert, S., Baker, N., Glover, L., Sanchez-Flores, A., Leung, K.F., Turner, D.J., Field, M.C., Berriman, M., and Horn, D. (2012). High-throughput decoding of antitrypanosomal drug efficacy and resistance. Nature *482*, 232–236.

Alsford, S., Currier, R.B., Guerra-Assunção, J.A., Clark, T.G., and Horn, D. (2014). Cathepsin-L can resist lysis by human serum in *Trypanosoma brucei brucei*. PLoS Pathog. *10*, e1004130.

ten Asbroek, A.L.M.A., Marc, O., and Borst, P. (1990). Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. Nature *348*, 174–175.

ten Asbroek, A.L.M.A., Mol, C.A.A.M., Kieft, R., and Borst, P. (1993). Stable transformation of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *59*, 133–142.

Asbury, C.L. (2003). Kinesin moves by an asymmetric hand-over-hand mechanism. Science. *302*, 2130–2134.

Atherton, J., Farabella, I., Yu, I.M., Rosenfeld, S.S., Houdusse, A., Topf, M., and Moores,

C.A. (2014). Conserved mechanisms of microtubule-stimulated ADP release, ATP binding, and force generation in transport kinesins. Elife *3*, e03680.

Aviram, N., and Schuldiner, M. (2017). Targeting and translocation of proteins to the endoplasmic reticulum at a glance. J. Cell Sci. *130*, 4079–4085.

Baker, N., Alsford, S., and Horn, D. (2011). Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. Mol. Biochem. Parasitol. *176*, 55–57.

Baker, N., Hamilton, G., Wilkes, J.M., Hutchinson, S., Barrett, M.P., and Horn, D. (2015). Vacuolar ATPase depletion affects mitochondrial ATPase function, kinetoplast dependency, and drug sensitivity in trypanosomes. Proc. Natl. Acad. Sci. *112*, 9112–9117.

Banks, K.L. (1979). *In vitro* binding of *Trypanosoma congolense* to erythrocytes. J. Protozool *26*, 103–108.

Bargul, J.L., Jung, J., McOdimba, F.A., Omogo, C.O., Adung'a, V.O., Krüger, T., Masiga, D.K., and Engstler, M. (2016). Species-specific adaptations of trypanosome morphology and motility to the mammalian host. PLoS Pathog. *12*, e1005448.

Barnes, R.L., and McCulloch, R. (2007). *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. Nucleic Acids Res. *35*, 3478–3493.

Barrett, M.P. (2018). The elimination of human African trypanosomiasis is in sight: Report from the third WHO stakeholders meeting on elimination of gambiense human African trypanosomiasis. PLoS Negl. Trop. Dis. *12*, e0006925.

Bastin, P., Ellis, K., Kohl, L., and Gull, K. (2000). Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. J. Cell Sci. *113*, 3321–3328.

Bayne, R.A.L., Kilbride, E.A., Lainson, F.A., Tetley, L., and Barry, J.D. (1993). A major surface antigen of procyclic stage *Trypanosoma congolense*. Mol. Biochem. Parasitol. *61*, 295–310.

Beecroft, R.P., Roditi, I., and Pearson, T.W. (1993). Identification and characterization of an acidic major surface glycoprotein from procyclic stage *Trypanosoma congolense*. Mol. Biochem. Parasitol. *61*, 285–294.

Bell, J.S., and McCulloch, R. (2003). Mismatch Repair Regulates Homologous Recombination, but Has Little Influence on Antigenic Variation, in *Trypanosoma brucei*. J. Biol. Chem. *278*, 45182–45188.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., and Gluenz, E. (2017). A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. R. Soc. Open Sci. *4*, 170095.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate : A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Benne, R., Van Den Burg, J., Brakenhoff, J.P.J., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46*, 819–826.

Benz, C., Dondelinger, F., Mckean, P.G., and Urbaniak, M.D. (2017). Cell cycle synchronisation of *Trypanosoma brucei* by centrifugal counter-flow elutriation reveals the timing of nuclear and kinetoplast DNA replication. Sci. Rep. *7*, 17599.

Berriman, M., Ghedin, E., Hertz-fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al. (2005). The genome of the African trypanosome *Trypanosoma brucei*. Science. *309*, 416–422.

Biebinger, S., Rettenmaier, S., Flaspohler, J., Hartmann, C., Peña-Diaz, J., Elizabeth Wirtz, L., Hotz, H.R., Barry, J.D., and Clayton, C. (1996). The PARP promoter of *Trypanosoma brucei* is developmentally regulated in a chromosomal context. Nucleic Acids Res. *24*, 1202–1211.

Blackburn, E.H., and Gall, J.G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. J. Mol. Biol. *120*, 33–53.

Blackburn, E.H., Budarf, M.L., Challoner, P.B., Cherry, J.M., Howard, E.A., Katzen, A.L., Pan, W.C., and Ryan, T. (1983). DNA termini in ciliate macronuclei. Cold Spring Harb Symp Quant Biol *47*, 1195–1207.

Blackburn, E.H., Greider, C.W., and Szostak, J.W. (2006). Telomeres and telomerase: The path from maize, *Tetrahymena* and yeast to human cancer and aging. Nat. Med. *12*, 1133–1138.

Blisnick, T., Buisson, J., Absalon, S., Marie, A., Cayet, N., and Bastin, P. (2014). The intraflagellar transport dynein complex of trypanosomes is made of a heterodimer of dynein heavy chains and of light and intermediate chains of distinct functions. Mol. Biol. Cell *25*, 2620–2633.

Boibessot, I., Turner, C.M.R., Watson, D.G., Goldie, E., Connel, G., McIntosh, A., Grant, M.H., and Skellern, G.G. (2002). Metabolism and distribution of phenanthridine trypanocides in *Trypanosoma brucei*. Acta Trop. *84*, 219–228.

Botstein, D., and Fink, G.R. (1988). Yeast: An experimental organism for modern biology. Science. *240*, 1439–1442.

Botstein, D., and Fink, G.R. (2011). Yeast: An experimental organism for 21st century biology. Genetics *189*, 695–704.

Braun, M., Diez, S., and Lansky, Z. (2016). Cell Biology: Kinesin-14 backsteps to organize polymerizing microtubules. Curr. Biol. *26*, R1271–R1296.

Brendza, R.P., Serbus, L.R., Duffy, J.B., and Saxton, W.M. (2000). A function for kinesin I in the posterior transport of oskar mRNA and staufen protein. Science. *289*, 2120–2122.

Bridges, D.J., Gould, M.K., Nerima, B., Maser, P., Burchmore, R.J.S., and de Koning, H.P. (2007). Loss of the high-affinity pentamidine transporter is responsible for high levels

of cross-resistance between arsenical and diamidine drugs in African trypanosomes. Mol. Pharmacol. *71*, 1098–1108.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science. *296*, 550–553.

Brun, R., and Schönenberger (1979). Cultivation and *in vitro* cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium: short communication. Acta Trop. *36*, 289–292.

Brun, R., and Schönenberger, M. (1981). Stimulating effect of citrate and cis-aconitate on the transformation of *Trypanosoma brucei* bloodstream forms to procyclic forms *in vitro*. Z Parasitenkd *66*, 17–24.

Brusini, L. (2019). Phosphoregulation of divergent cell division in protozoan parasites. University of Nottingham.

Brusini, L., Archivio, S.D., Mcdonald, J., and Wickstead, B. (2019). Ndc80 / Nuf2-like protein KKIP1 connects a stable kinetoplastid outer kinetochore complex to the inner kinetochore and responds to metaphase tension. bioRxiv, doi.org/10.1101/764829

Burkard, G., Fragoso, C.M., and Roditi, I. (2007). Highly efficient stable transformation of bloodstream forms of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *153*, 220–223.

Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb. Perspect. Biol. *6*, 1–17.

Bütikofer, P., Ruepp, S., Boschung, M., and Roditi, I. (1997). "GPEET" procyclin is the major surface protein of procyclic culture forms of *Trypanosoma brucei brucei* strain 427. Biochem. J. *326*, 415–423.

Carruthers, V.B., Ploeg, L.H.T. Van Der, and Cross, G.A.M. (1993). DNA-mediated transformation of bloodstream-form *Trypanosoma brucei*. Nucleic Acids Res. *21*, 2537–2538.

Carter, N.S., Berger, B.J., and Fairlamb, A.H. (1995). Uptake of diamidine drugs by the P2 nucleoside transporter in melarsen-sensitive and -resistant *Trypanosoma brucei brucei*. J. Biol. Chem. *270*, 28153–28157.

Chan, K.Y., and Ersfeld, K. (2010). The role of the Kinesin-13 family protein TbKif13-2 in flagellar length control of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *174*, 137–140.

Chan, K.Y., Matthews, K.R., and Ersfeldk, K. (2010). Functional characterisation and drug target validation of a mitotic kinesin-13 in *Trypanosoma brucei*. PLoS Pathog. *6*, e1001050.

Cheung, J.L.Y., Wand, N. V, Ooi, C.P., Ridewood, S., Wheeler, R.J., and Rudenko, G. (2016). Blocking synthesis of the variant surface glycoprotein coat in *Trypanosoma brucei* leads to an increase in macrophage phagocytosis due to reduced clearance of surface coat antibodies. PLoS Pathog. *12*, e1006023.

Chowdhury, A.R., Zhao, Z., and Englund, P.T. (2008). Effect of hydroxyurea on procyclic *Trypanosoma brucei*: An unconventional mechanism for achieving synchronous

growth. Eukaryot. Cell *7*, 425–428.

Clausen, P.H., Sidibe, I., Kaboré, I., and Bauer, B. (1992). Development of multiple drug resistance of *Trypanosoma congolense* in Zebu cattle under high natural tsetse fly challenge in the pastoral zone of Samorogouan, Burkina Faso. Acta Trop. *51*, 229–236.

Clemmens, C.S., Morris, M.T., Lyda, T.A., Acosta-Serrano, A., and Morris, J.C. (2009). *Trypanosoma brucei* AMP-activated kinase subunit homologs influence surface molecule expression. Exp. Parasitol. *123*, 250–257.

Cottingham, F.R., and Hoyt, M.A. (1997). Mitotic spindle positioning in *Saccharomyces cerevisiae* is accomplished by antagonistically acting microtubule motor proteins. J. Cell Biol. *138*, 1041–1053.

Cottingham, F.R., Gheber, L., Miller, D.L., and Hoyt, M.A. (1999). Novel roles for *Saccharomyces cerevisiae* mitotic spindle motors. J. Cell Biol. *147*, 335–349.

Coustou, V., Biran, M., Breton, M., Guegan, F., Rivière, L., Plazolles, N., Nolan, D., Barrett, M.P., Franconi, J.M., and Bringaud, F. (2008). Glucose-induced remodeling of intermediary and energy metabolism in procyclic *Trypanosoma brucei*. J. Biol. Chem. *283*, 16343–16354.

Coustou, V., Guegan, F., Plazolles, N., and Baltz, T. (2010). Complete *in vitro* life cycle of *Trypanosoma congolense*: Development of genetic tools. PLoS Negl. Trop. Dis. *4*, e618.

Coustou, V., Plazolles, N., Guegan, F., and Baltz, T. (2012). Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cell. Microbiol. *14*, 431–445.

Cross, G.A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitology *71*, 393–417.

Cross, R.A. (2004). The kinetic mechanism of kinesin. Trends Biochem. Sci. *29*, 301–309.

Cross, G.A.M., and Manning, J.C. (1973). Cultivation of *Trypanosoma brucei* sspp. in semi-defined and defined media. Parasitology *67*, 315–331.

Cross, G.A.M., Kim, H.S., and Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. Mol. Biochem. Parasitol. *195*, 59–73.

Cunningham, I. (1977). New culture medium for maintenance of tsetse tissues and growth of trypanosomatids. J. Protozool. *24*, 325–329.

Cunningham, I., and Honigberg, B. (1977). Infectivity reacquisition by *Trypanosoma brucei brucei* cultivated with tsetse salivary glands. Science. *197*, 1279–1282.

Cunningham, I., Honigberg, B.M., and Taylor, A.M. (1981). Infectivity of monomorphic and pleomorphic *Trypanosoma brucei* stocks cultivated at 28°C with various tsetse fly tissues. J Parasitol *67*, 391–397.

D'Archivio, S., and Wickstead, B. (2017). Trypanosome outer kinetochore proteins

suggest conservation of chromosome segregation machinery across eukaryotes. J. Cell Biol. *216*, 379–391.

D'Archivio, S., Medina, M., Cosson, A., Chamond, N., Rotureau, B., Minoprio, P., and Goyard, S. (2011). Genetic engineering of *Trypanosoma (dutonella) vivax* and *in vitro* differentiation under axenic conditions. PLoS Negl. Trop. Dis. *5*, e1461.

D'Archivio, S., Cosson, A., Medina, M., Lang, T., Minoprio, P., and Goyard, S. (2013). Non-invasive in vivo study of the *Trypanosoma vivax* infectious process consolidates the brain commitment in late infections. PLoS Negl. Trop. Dis. *7*, e1976.

Daniels, J.P., Gull, K., and Wickstead, B. (2012). The trypanosomatid-specific N terminus of RPA2 is required for RNA polymerase I assembly, localization, and function. Eukaryot. Cell *11*, 662–672.

Dean, S., Marchetti, R., Kirk, K., and Matthews, K.R. (2009). A surface transporter family conveys the trypanosome differentiation signal. Nature *459*, 213–217.

Dean, S., Sunter, J.D., and Wheeler, R.J. (2017). TrypTag.org: A trypanosome genome-wide protein localisation resource. Trends Parasitol. *33*, 80–82.

Desquesnes, M., and Dia, M.L. (2003). *Trypanosoma vivax*: Mechanical transmission in cattle by one of the most common African tabanids, *Atylotus agrestis*. Exp. Parasitol. *103*, 35–43.

Downey, N., and Donelson, J.E. (1999a). Search for promoters for the GARP and rRNA genes of *Trypanosoma congolense*. Mol. Biochem. Parasitol. *104*, 25–38.

Downey, N., and Donelson, J.E. (1999b). Expression of foreign proteins in *Trypanosoma congolense*. Mol Biochem Parasitol *104*, 39–53.

Dubowy, C., and Sehgal, A. (2017). Circadian rhythms and sleep in *Drosophila melanogaster*. Genetics *205*, 1373–1397.

Durand-Dubief, M., Kohl, L., and Bastin, P. (2003). Efficiency and specificity of RNA interference generated by intra- and intermolecular double stranded RNA in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *129*, 11–21.

Düselder, A., Thiede, C., Schmidt, C.F., and Lakämper, S. (2012). Neck-linker length dependence of processive kinesin-5 motility. J. Mol. Biol. *423*, 159–168.

Duszenko, B.Y.M., Ferguson, M.A.J., Lamont, G.S., Rifkin, M.R., and Cross, G.A.M. (1985). Cysteine eliminates the feeder cell requirement for cultivation of *Trypanosoma brucei* bloodstream forms *in vitro*. J. Exp. Med. *162*, 1256–1263.

Dutoya, S., Gibert, S., Lemercier, G., Santarelli, X., Baltz, D., Baltz, T., and Bakalara, N. (2001). A Novel C-terminal kinesin is essential for maintaining functional acidocalcisomes in *Trypanosoma brucei*. J. Biol. Chem. *276*, 49117–49124.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. J. Biol. Methods *1*, e10.

Edwards, B.F.L., Wheeler, R.J., Barker, A.R., Moreira-Leite, F.F., Gull, K., and Sunter, J.D. (2018). Direction of flagellum beat propagation is controlled by proximal/distal

outer dynein arm asymmetry. Proc. Natl. Acad. Sci. U. S. A. *115*, E7341–E7350.

Endow, S.A., Kull, F.J., and Liu, H. (2010). Kinesins at a glance. J. Cell Sci. *123*, 3420–3423.

Engstler, M., Thilo, L., Weise, F., Grünfelder, C.G., Schwarz, H., Boshart, M., and Overath, P. (2004). Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in *Trypanosoma brucei*. J. Cell Sci. *117*, 1105–1115.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., and Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell *131*, 505–515.

Ersfeld, K., Melville, S.E., and Gull, K. (1999). Nuclear and genome organization of *Trypanosoma brucei*. Parasitol. Today *15*, 58–63.

Eze, A.A., Gould, M.K., Munday, J.C., Tagoe, D.N.A., Stelmanis, V., Schnaufer, A., and De Koning, H.P. (2016). Reduced mitochondrial membrane potential is a late adaptation of *Trypanosoma brucei brucei* to isometamidium preceded by mutations in the γ subunit of the F1Fo-ATPase. PLoS Negl. Trop. Dis. *10*, e0004791.

Fairlamb, A.H. (2003). Chemotherapy of human African trypanosomiasis: Current and future prospects. Trends Parasitol. *19*, 488–494.

Ferguson, M., Homans, S., Dwek, R., and Rademacher, T. (1988). Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. Science. *239*, 753–759.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.F., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature *391*, 806–811.

Fish, W.R., Muriuki, C.W., Muthiani, A.M., Grab, D.J., and Lonsdale-Eccles, J.D. (1989). Disulfide bond involvement in the maintenance of the cryptic nature of the cross-reacting determinant of metacyclic forms of *Trypanosoma congolense*. Biochemistry *28*, 5415–5421.

Forgac, M. (2007). Vacuolar ATPases: Rotary proton pumps in physiology and pathophysiology. Nat. Rev. Mol. Cell Biol. *8*, 917–929.

Forsythe, G.R., McCulloch, R., and Hammarton, T.C. (2009). Hydroxyurea-induced synchronisation of bloodstream stage *Trypanosoma brucei*. Mol. Biochem. Parasitol. *164*, 131–136.

Fourriere, L., Divoux, S., Roceri, M., Perez, F., and Boncompain, G. (2016). Microtubule-independent secretion requires functional maturation of Golgi elements. J. Cell Sci. *129*, 3238–3250.

Frame, I.A., Ross, C.A., and Luckins, A.G. (1991). Variability of *in vitro* culture characteristics, including metacyclic trypomastigote production, in different stocks of *Trypanosoma congolense*. Acta Trop. *50*, 135–140.

Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Mattioli, R.C., and Argaw, D. (2017). Monitoring the elimination of human African trypanosomiasis:

Update to 2014. PLoS Negl. Trop. Dis. *11*, e0005585.

Francoid, J.R., Cecchiid, G., Priottoid, G., Paone, M., Diarra, A., Grout, L., Simarro, P.P., Zhaoid, W., and Argaw, D. (2018). Monitoring the elimination of human African trypanosomiasis: Update to 2016. PLoS Negl. Trop. Dis. *12*, e0006890.

Gadelha, C., Zhang, W., Chamberlain, J.W., Chait, B.T., Wickstead, B., and Field, M.C. (2015). Architecture of a host-parasite interface: complex targeting mechanisms revealed through proteomics. Mol. Cell. Proteomics *14*, 1911–1926.

Gardiner, P.R. (1989). Recent studies of the biology of *Trypanosoma vivax*. Adv. Parasitol. *28*, 229–317.

Gardiner, P.R., Nene, V., Barry, M.M., Thatthi, R., Burleigh, B., and Clarke, M.W. (1996). Characterization of a small variable surface glycoprotein from *Trypanosoma vivax*. Mol. Biochem. Parasitol. *82*, 1–11.

Gibson, W.C., and Borst, P. (1986). Size-fractionation of the small chromosomes of *Trypanozoon* and *Nannomonas* trypanosomes by pulsed field gradient gel electrophoresis. Mol. Biochem. Parasitol. *18*, 127–140.

Giordani, F., Morrison, L.J., Rowan, T.G., De Koning, H.P., and Barrett, M.P. (2016). The animal trypanosomiases and their chemotherapy: A review. Parasitology *143*, 1862–1889.

Glover, L., and Horn, D. (2009). Site-specific DNA double-strand breaks greatly increase stable transformation efficiency in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *166*, 194–197.

Glover, L., McCulloch, R., and Horn, D. (2008). Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic Acids Res. *36*, 2608–2618.

Glover, L., Alsford, S., Baker, N., Turner, D.J., Sanchez-Flores, A., Hutchinson, S., Hertz-Fowler, C., Berriman, M., and Horn, D. (2015). Genome-scale RNAi screens for high-throughput phenotyping in bloodstream-form African trypanosomes. Nat. Protoc. *10*, 106–133.

Goldstein, L.S.B. (1993). The kinesin family functional redundancy. Annu. Rev. Genet *27*, 319–351.

Gonen, S., Akiyoshi, B., Iadanza, M.G., Shi, D., Duggan, N., Biggins, S., and Gonen, T. (2012). The structure of purified kinetochores reveals multiple microtubule-attachment sites. Nat. Struct. Mol. Biol. *19*, 925–929.

Goodson, H. V, Kang, S.J., and Endow, S.A. (1994). Molecular phylogeny of the kinesin family of microtubule motor proteins. J. Cell Sci. *107*, 1875–1884.

Gould, M.K., and Schnaufer, A. (2014). Independence from kinetoplast DNA maintenance and expression is associated with multidrug resistance in *Trypanosoma brucei in vitro*. Antimicrob. Agents Chemother. *58*, 2925–2928.

Gray, M.A., Ross, C.A., Taylor, A.M., Tetley, L., and Luckins, A.G. (1984). *In vitro* cultivation of *Trypanosoma congolense*: the production of infective forms from

metacyclic trypanosomes cultured on bovine endothelial cell monolayers. Acta Trop. *42*, 99–111.

Gray, M.A., Ross, C.A., Taylor, A.M., Tetley, L., and Luckins, A.G. (1985). *In vitro* cultivation of *Trypanosoma congolense*: The production of infective forms from metacyclic trypanosomes cultured on bovine endothelial cell monolayers. Acta Trop. *42*, 99–111.

Greca, F. La, and Magez, S. (2011). Vaccination against trypanosomiasis: Can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? Hum. Vaccin. *7*, 1225–1233.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. Cell *43*, 405–413.

Gupta, V., Palmer, K.J., Spence, P., Hudson, A., and Stephens, D.J. (2008). Kinesin-1 (uKHC/KIF5B) is required for bidirectional motility of ER exit sites and efficient ER-to-Golgi transport. Traffic *9*, 1850–1866.

Hackney, D.D., and Stock, M.F. (2000). Kinesin's IAK tail domain inhibits initial microtubule-stimulated ADP release. Nat. Cell Biol. *2*, 257–260.

Hall, J.P.J., Wang, H., and David Barry, J. (2013). Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation. PLoS Pathog. *9*, e1003502.

Hariharan, V., and Hancock, W.O. (2009). Insights into the mechanical properties of the Kinesin neck linker domain from sequence analysis and molecular dynamics simulations. Cell Mol. Bioeng. *2*, 177–189.

Hatsumi, M., and Endow, S.A. (1992). Mutants of the microtubule motor protein, nonclaret disjunctional, affect spindle structure and chromosome movement in meiosis and mitosis. J. Cell Sci. *101*, 547–559.

Helenius, J., Brouhard, G., Kalaidzidis, Y., Diez, S., and Howard, J. (2006). The depolymerizing kinesin MCAK uses lattice diffusion to rapidly target microtubule ends. Nature *441*, 115–119.

Hemphill, A., Frame, I., and Ross, C.A. (1994). The interaction of *Trypanosoma congolense* with endothelial cells. Parasitology *109*, 631–641.

Hildebrandt, E.R., and Hoyt, M.A. (2000). Mitotic motors in *Saccharomyces cerevisiae*. Biochim. Biophys. Acta *1496*, 99–116.

Hirokawa, N., and Noda, Y. (2008). Intracellular transport and kinesin superfamily proteins : Structure dynamics and function. Physiol. Rev. *88*, 1089–1118.

Hirokawa, N., and Takemura, R. (2004). Kinesin superfamily proteins and their various functions and dynamics. Exp. Cell Res. *301*, 50–59.

Hirokawa, N., Noda, Y., and Okada, Y. (1998). Kinesin and dynein superfamily proteins in organelle transport and cell division. Curr. Opin. Cell Biol. *10*, 60–73.

Hirokawa, N., Noda, Y., Tanaka, Y., and Niwa, S. (2009). Kinesin superfamily motor proteins and intracellular transport. Nat. Rev. Mol. Cell Biol. *10*, 682–696.

Hirumi, H., and Hirumi, K. (1989). Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. J. Parasitol. *75*, 985–989.

Hirumi, H., and Hirumi, K. (1991). *In vitro* cultivation of *Trypanosoma congolense* bloodstream forms in the absence of feeder cell layers. Parasitology *102*, 225–236.

Hirumi, H., Doyle, J.J., and Hirumi, K. (1977a). Cultivation of bloodstream *Trypanosoma brucei*. Bull. World Health Organ. *55*, 405–409.

Hirumi, H., Doyle, J.J., and Hirumi, K. (1977b). African trypanosomes: Cultivation of animal-infective *Trypanosoma brucei in vitro*. Science. *196*, 992–994.

Hirumi, H., Hirumi, K., Doyle, J.J., Cross, G.A.M., Hirumi, H., Hirumi, K., Doyle, J.J., Cross, G.A.M., and Cross, G.A.M. (1980). *In vitro* cloning of animal-infective bloodstream forms of *Trypanosoma brucei*. Parasitology *80*, 371–382.

Hirumi, H., Hirumi, K., Moloo, S.K., and Shaw, M.K. (1992). *Trypanosoma brucei brucei*: *In vitro* production of metacyclic forms. J. Protozool *39*, 619–627.

Van Den Hoff, M.J. b, Moorman, A.F.M., and Lamers, W.H. (1992). Electroporation in "intracellular" buffer increases cell survival. Nucleic Acids Res. *20*, 2902.

Holmes, P. (2014). First WHO meeting of stakeholders on elimination of gambiense human African trypanosomiasis. PLoS Negl. Trop. Dis. *8*, e3244.

Hooff, J.J., Tromer, E., Wijk, L.M., Snel, B., and Kops, G.J. (2017). Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. *18*, 1559–1571.

Inoue, N., Otsu, K., Ferraro, D.M., and Donelson, J.E. (2002). Tetracycline-regulated RNA interference in *Trypanosoma congolense*. Mol. Biochem. Parasitol. *120*, 309–313.

Ipsaro, J.J., and Joshua-tor, L. (2015). From guide to target : molecular insights into eukaryotic RNA-interference machinery. Nat. Struct. Mol. Biol. *22*, 20–28.

Iwai, S., Ishiji, A., Mabuchi, I., and Sutoh, K. (2004). A novel actin-bundling kinesin-related protein from *Dictyostelium discoideum*. J. Biol. Chem. *279*, 4696–4704.

Jackson, A.P., Berry, A., Aslett, M., Allison, H.C., Burton, P., Vavrova-Anderson, J., Brown, R., Browne, H., Corton, N., Hauser, H., et al. (2012). Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. Proc. Natl. Acad. Sci. *109*, 3416–3421.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. *337*, 816–822.

Jones, T.W., and Dávila, A.M. (2001). *Trypanosoma vivax*- out of Africa. Trends Parasitol. *17*, 99–101.

Julie Lee, Y.-R., and Liu, B. (2004). Cytoskeletal motors in *Arabidopsis*. Sixty-one kinesins and seventeen myosins. Plant Physiol. *136*, 3877–3883.

Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., and Marcotte,

E.M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science. *348*, 921–925.

Kaiser, M., Bray, M.A., Cal, M., Trunz, B.B., Torreele, E., and Brun, R. (2011). Antitrypanosomal activity of fexinidazole, a new oral nitroimidazole drug candidate for treatment of sleeping sickness. Antimicrob. Agents Chemother. *55*, 5602–5608.

Kalidas, S., Li, Q., and Phillips, M.A. (2011). A Gateway® compatible vector for gene silencing in bloodstream form *Trypanosoma brucei*. Mol. Biochem. Parasitol. *178*, 51–55.

Kaseda, K., Higuchi, H., and Hirose, K. (2003). Alternate fast and slow stepping of a heterodimeric kinesin molecule. Nat. Cell Biol. *5*, 1079–1082.

Kennedy, P.G.E. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). Lancet Neurol. *12*, 186–194.

Kern, D.M., and Cheeseman, I.M. (2012). Kinetochore structure: Pulling answers from yeast. Curr. Biol. *22*, R842–R844.

Kirchner, J., Woehlke, G., and Schliwa, M. (1999). Universal and unique features of kinesin motors: Insights from a comparison of fungal and animal conventional kinesins. Biol. Chem. *380*, 915–921.

Kolev, N.G., Ullu, E., and Tschudi, C. (2012). Developmental progression to triggered by an RNA-binding protein. Science. *338*, 1352–1354.

Kolev, N.G., Ullu, E., and Tschudi, C. (2014). The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*. Cell. Microbiol. *16*, 482–489.

Kull, F.J., Sablin, E.P., Lau, R., Fletterick, R.J., and Vale, R.D. (1996). Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. Nature *380*, 550–554.

LaCount, D.J., Bruse, S., Hill, K.L., and Donelson, J.E. (2000). Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. Mol. Biochem. Parasitol. *111*, 67–76.

Lamour, N., Rivière, L., Coustou, V., Coombs, G.H., Barrett, M.P., and Bringaud, F. (2005). Proline metabolism in procyclic *Trypanosoma brucei* is down-regulated in the presence of glucose. J. Biol. Chem. *280*, 11902–11910.

Lander, N., and Chiurillo, M.A. (2019). State-of-the-art CRISPR/Cas9 technology for genome editing in trypanosomatids. J Eukaryot Microbiol *66*, 981–991.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langousis, G., and Hill, K.L. (2014). Motility and more: The flagellum of *Trypanosoma brucei*. Nat. Rev. Microbiol. *12*, 505–518.

Lawrence, C.J., Dawe, R.K., Christie, K.R., Cleveland, D.W., Dawson, S.C., Endow, S.A., Goldstein, L.S.B., Goodson, H. V., Hirokawa, N., Howard, J., et al. (2004). A standardized kinesin nomenclature. J. Cell Biol. *167*, 19–22.

Lee, M.G., and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature *327*, 31–35.

Li, Z. (2012). Regulation of the cell division cycle in *Trypanosoma brucei*. Eukaryot. Cell *11*, 1180–1190.

Li, Z., and Wang, C.C. (2003). A PHO80-like cyclin and a B-type cyclin control the cell cycle of the procyclic form of *Trypanosoma brucei*. J. Biol. Chem. *278*, 20652–20658.

Li, C., Zhang, Y., Yang, Q., Ye, F., Sun, S.Y., Chen, E.S., and Liou, Y.C. (2016). NuSAP modulates the dynamics of kinetochore microtubules by attenuating MCAK depolymerisation activity. Sci. Rep. *6*, 1–12.

Li, F.-J., Xu, Z.-S., Mon Aye, H., Brasseur, A., Lun, Z.-R., Tan, K.S., and He, C.Y. (2017). An efficient cumate-inducible system for procyclic and bloodstream form *Trypanosoma brucei*. Mol. Biochem. Parasitol. *214*, 101–104.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. Science. *302*, 1401–1404.

MacGregor, P., Savill, N.J., Hall, D., and Matthews, K.R. (2011). Transmission stages dominate trypanosome within-host dynamics during chronic infections. Cell Host Microbe *9*, 310–318.

MacGregor, P., Szöo'R, B., Savill, N.J., and Matthews, K.R. (2012). Trypanosomal immune evasion, chronicity and transmission: An elegant balancing act. Nat. Rev. Microbiol. *10*, 431–438.

MacGregor, P., Rojas, F., Dean, S., and Matthews, K.R. (2013). Stable transformation of pleomorphic bloodstream form *Trypanosoma brucei*. Mol. Biochem. Parasitol. *190*, 60–62.

Marcello, L., and Barry, J.D. (2007). Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Res. *17*, 1344–1352.

Marx, A., Hoenger, A., and Mandelkow, E. (2009). Structures of kinesin motor proteins. Cell Motil. Cytoskeleton *66*, 958–966.

Mäser, P., Sütterlin, C., Kralli, A., and Kaminsky, R. (1999). A nucleoside transporter from *Trypanosoma brucei* involved in drug resistance. Science. *285*, 242–244.

Matsuzaki, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., et al. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature *428*, 653–657.

Mazet, M., Morand, P., Biran, M., Bouyssou, G., Courtois, P., Daulouède, S., Millerioux, Y., Franconi, J.M., Vincendeau, P., Moreau, P., et al. (2013). Revisiting the central metabolism of the bloodstream forms of *Trypanosoma brucei*: Production of acetate in the mitochondrion is essential for parasite viability. PLoS Negl. Trop. Dis. *7*, e2587.

Miki, H., Setou, M., Kaneshiro, K., and Hirokawa, N. (2001). All kinesin superfamily protein, KIF, genes in mouse and human. Proc. Natl. Acad. Sci. U. S. A. *98*, 7004–7011.

Millerioux, Y., Ebikeme, C., Biran, M., Morand, P., Bouyssou, G., Vincent, I.M., Mazet, M., Riviere, L., Franconi, J.M., Burchmore, R.J.S., et al. (2013). The threonine degradation pathway of the *Trypanosoma brucei* procyclic form: The main carbon source for lipid biosynthesis is under metabolic control. Mol. Microbiol. *90*, 114–129.

Millerioux, Y., Mazet, M., Bouyssou, G., Allmann, S., Kiema, T.R., Bertiaux, E., Fouillen, L., Thapa, C., Biran, M., Plazolles, N., et al. (2018). *De novo* biosynthesis of sterols and fatty acids in the *Trypanosoma brucei* procyclic form: Carbon source preferences and metabolic flux redistributions. PLoS Pathogens. *14*, e1007116.

Moloo, S.K., and Kutuza, S.B. (1990). Expression of resistance to isometamidium and diminazene in *Trypanosoma congolense* in Boran cattle infected by *Glossina morsitans centralis*. Acta Trop. *47*, 79–89.

Mony, B.M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D., and Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. Nature *505*, 681–685.

Moore, J.D., and Endow, S.A. (1996). Kinesin proteins: The phylum of motors for microtubule-based motility. BioEssays *18*, 207–219.

Morris, J.C., Wang, Z., Drew, M.E., and Englund, P.T. (2002). Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. EMBO J. *21*, 4429–4438.

Moti, Y., Fikru, R., Van Den Abbeele, J., Büscher, P., Van den Bossche, P., Duchateau, L., and Delespaux, V. (2012). Ghibe river basin in Ethiopia: Present situation of trypanocidal drug resistance in *Trypanosoma congolense* using tests in mice and PCR-RFLP. Vet. Parasitol. *189*, 197–203.

Mugo, E., and Clayton, C. (2017). Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*. PLoS Pathog *13*, e1006560.

Muhanguzi, D., Mugenyi, A., Bigirwa, G., Kamusiime, M., Kitibwa, A., Akurut, G.G., Ochwo, S., Amanyire, W., Okech, S.G., Hattendorf, J., et al. (2017). African animal trypanosomiasis as a constraint to livestock health and production in Karamoja region: A detailed qualitative and quantitative assessment. BMC Vet. Res. *13*, 355.

Mulugeta, W., Wilkes, J., Mulatub, W., Majiwa, P.A.O., Masake, R., and Peregrine, A.S. (1997). Long-term occurrence of *Trypanosoma congolense* resistant to diminazene, isometamidium and homidium in cattle at Ghibe, Ethiopia. Acta Trop. *64*, 205–217.

Munday, J.C., Rojas López, K.E., Eze, A.A., Delespaux, V., Van Den Abbeele, J., Rowan, T., Barrett, M.P., Morrison, L.J., and de Koning, H.P. (2013). Functional expression of TcoAT1 reveals it to be a P1-type nucleoside transporter with no capacity for diminazene uptake. Int. J. Parasitol. Drugs Drug Resist. *3*, 69–76.

Mungube, E.O., Vitouley, H.S., Allegye-Cudjoe, E., Diall, O., Boucoum, Z., Diarra, B., Sanogo, Y., Randolph, T., Bauer, B., Zessin, K.H., et al. (2012). Detection of multiple

drug-resistant *Trypanosoma congolense* populations in village cattle of south-east Mali. Parasites and Vectors *5*, 155.

Nakajima, K., Yin, X., Takei, Y., Seog, D.H., Homma, N., and Hirokawa, N. (2012). Molecular motor KIF5A is essential for GABAA receptor transport, and KIF5A deletion causes epilepsy. Neuron *76*, 945–961.

Nangaku, M., Sato-Yoshitake, R., Okada, Y., Noda, Y., Takemura, R., Yamazaki, H., and Hirokawa, N. (1994). KIF1B, a novel microtubule plus end-directed monomeric motor protein for transport of mitochondria. Cell *79*, 1209–1220.

Nasevicius, A., and Ekker, S.C. (2000). Effective targeted gene "knockdown" in zebrafish. Nat. Genet. *26*, 216–220.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. *95*, 14687–14692.

Nurse, P., and Thuriaux, P. (1980). Regulatory genes controlling mitosis in the fission yeast *Schizosaccharomyces pombe*. Genetics *96*, 627–637.

Olmo, F., Costa, F.C., Mann, G.S., Taylor, M.C., and Kelly, J.M. (2018). Optimising genetic transformation of *Trypanosoma cruzi* using hydroxyurea-induced cell-cycle synchronisation. Mol. Biochem. Parasitol. *226*, 34–36.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev. *16*, 948–958.

Pays, E., and Vanhollebeke, B. (2009). Human innate immunity against African trypanosomes. Curr. Opin. Immunol. *21*, 493–498.

Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M., and Gibson, W. (2011). Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. Proc. Natl. Acad. Sci. U. S. A. *108*, 3671–3676.

Peacock, L., Cook, S., Ferris, V., Bailey, M., and Gibson, W. (2012). The life cycle of *Trypanosoma (Nannomonas) congolense* in the tsetse fly. Parasites and Vectors *5*, 109.

Peschel, N., and Helfrich-Förster, C. (2011). Setting the clock - By nature: Circadian rhythm in the fruitfly *Drosophila melanogaster*. FEBS Lett. *585*, 1435–1442.

Pollock, N., De Hostos, E.L., Turck, C.W., and Vale, R.D. (1999). Reconstitution of membrane transport powered by a novel dimeric kinesin motor of the Unc104/KIF1a family purified from *Dictyostelium*. J. Cell Biol. *147*, 493–505.

Poon, S.K., Peacock, L., Gibson, W., Gull, K., and Kelly, S. (2012). A modular and optimized single marker system for generating *Trypanosoma brucei* cell lines expressing T7 RNA polymerase and the tetracycline repressor. Open Biol. *2*, 110037.

Prahlad, V., Yoon, M., Moir, R.D., Vale, R.D., and Goldman, R.D. (1998). Rapid movements of vimentin on microtubule tracks: Kinesin-dependent assembly of intermediate filament networks. J. Cell Biol. *143*, 159–170.

Prevo, B., Scholey, J.M., and Peterman, E.J.G. (2017). Intraflagellar transport:

Mechanisms of motor action, cooperation and cargo delivery. Physiol. Behav. *284*, 2905–2931.

Ralston, K.S., and Hill, K.L. (2008). The flagellum of *Trypanosoma brucei*: New tricks from an old dog. Int. J. Parasitol. *38*, 869–884.

Randall, R.N., Radford, C.E., Roof, K.A., Natarajan, D.K., and Gaucher, E.A. (2016). An experimental phylogeny to benchmark ancestral sequence reconstruction. Nat. Commun. *7*, 12847.

Read, L.K., Lukeš, J., and Hashimi, H. (2016). Trypanosome RNA editing: The complexity of getting U in and taking U out. Wiley Interdiscip. Rev. RNA *7*, 33–51.

Reddy, A.S., and Day, I.S. (2001). Kinesins in the *Arabidopsis* genome: A comparative analysis among eukaryotes. BMC Genomics *2*, 2.

Reuner, B., Vassella, E., Yutzy, B., and Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of *Trypanosoma brucei* bloodstream forms in culture. Mol. Biochem. Parasitol. *90*, 269–280.

Richards, T.A., and Cavalier-Smith, T. (2005). Myosin domain evolution and the primary divergence of eukaryotes. Nature *436*, 1113–1118.

Rico, E., Jeacock, L., Kovářová, J., and Horn, D. (2018). Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. Sci. Rep. *8*, 7960.

Roditi, I., Furger, A., Ruepp, S., Schürch, N., and Bütikofer, P. (1998). Unravelling the procyclin coat of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *91*, 117–130.

Roger, A.J., and Simpson, A.G.B. (2009). Evolution: Revisiting the root of the eukaryote tree. Curr. Biol. *19*, R165–R167.

Röhlk, C., Rohlfs, M., Leier, S., Schliwa, M., Liu, X., Parsch, J., and Woehlke, G. (2008). Properties of the Kinesin-1 motor DdKif3 from *Dictyostelium discoideum*. Eur. J. Cell Biol. *87*, 237–249.

Rojas, F., Silvester, E., Young, J., Milne, R., Tettey, M., Houston, D.R., Walkinshaw, M.D., Pérez-Pi, I., Auer, M., Denton, H., et al. (2019). Oligopeptide signaling through TbGPR89 drives Trypanosome quorum sensing. Cell *176*, 306-317.e16.

Rosas-Murrieta, N., Rojas-Sánchez, G., Reyes-Carmona, S., Martínez-Contreras, R., Martínez-Montiel, N., Millán-Pérez-Peña, L., and Herrera-Camacho, I. (2015). Study of cellular processes in higher eukaryotes using the yeast *Schizosaccharomyces pombe* as a model. In Microbiology in Agriculture and Human Health, pp. 107–132.

Ross, C.A., Gray, M.A., Taylor, A.M., and Luckins, A.G. (1985). *In vitro* cultivation of *Trypanosoma congolense*: establishment of infective mammalian forms in continuous culture after isolation from the blood of infected mice. Acta Trop. *42*, 113–122.

Rossi, G., Manfrin, A., and Lutolf, M.P. (2018). Progress and potential in organoid research. Nat. Rev. Genet. *19*, 671–687.

Rovis, L., Barbet, A.F., and Williams, R.O. (1978). Characterisation of the surface coat of *Trypanosoma congolense*. Nature *271*, 654–656.

Ruehle, M.D., Orias, E., and Pearson, C.G. (2016). *Tetrahymena* as a unicellular model eukaryote: Genetic and genomic tools. Genetics *203*, 649–665.

Sablin, E.P., and Fletterick, R.J. (2004). Coordination between motor domains in processive kinesins. J. Biol. Chem. *279*, 15707–15710.

Sablin, E.P., Kull, F.J., Cooke, R., Vale, R.D., and Fletterick, R.J. (1996). Crystal structure of the motor domain of the kinesin-related motor ncd. Nature *380*, 555–559.

Sahin, A., Asencio, C., Izotte, J., Pillay, D., Coustou, V., Karembe, H., and Baltz, T. (2014). The susceptibility of *Trypanosoma congolense* and *Trypanosoma brucei* to isometamidium chloride and its synthetic impurities. Vet. Parasitol. *203*, 270–275.

Sakurai, T., Tanaka, M., Kawazu, S. ichiro, and Inoue, N. (2009). Establishment of an in vitro transgene expression system in epimastigotes of *Trypanosoma congolense*. Parasitol. Int. *58*, 110–113.

Saldivia, M., Ceballos-Pérez, G., Bart, J.M., and Navarro, M. (2016). The AMPKα1 pathway positively regulates the developmental transition from proliferation to quiescence in *Trypanosoma brucei*. Cell Rep. *17*, 660–670.

Savioli, D.D. and L. (2012). Accelerating work to overcome the global impact of neglected tropical diseases: A roadmap for implementation. World Health Organisation.

Saxton, W.M., Hicks, J., Et Goldstein, L.S., and Raff, E.C. (1991). Kinesin heavy chain is essential for viability and neuromuscular functions in *Drosophila*, but mutants show no defects in mitosis. Cell *64*, 1093–1102.

Scarabelli, G., and Grant, B.J. (2013). Mapping the structural and dynamical features of kinesin motor domains. PLoS Comput. Biol. *9*, e1003329.

Scarabelli, G., Soppina, V., Yao, X.Q., Atherton, J., Moores, C.A., Verhey, K.J., and Grant, B.J. (2015). Mapping the processivity determinants of the Kinesin-3 motor domain. Biophys. J. *109*, 1537–1540.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schuchardt, I., Absmann, D., Thines, E., Schuberth, C., and Steinberg, G. (2005). Myosin-V, Kinesin-1, and Kinesin-3 cooperate in hyphal growth of the fungus *Ustilago maydis*. Mol. Biol. Cell *16*, 5191–5201.

Schumann Burkard, G., Jutzi, P., and Roditi, I. (2011). Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. Mol. Biochem. Parasitol. *175*, 91–94.

Seidel, C., Moreno-Velásquez, S.D., Riquelme, M., and Fischer, R. (2013). *Neurospora crassa* NKIN2, a kinesin-3 motor, transports early endosomes and is required for polarized growth. Eukaryot. Cell *12*, 1020–1032.

Sevova, E.S., and Bangs, J.D. (2009). Streamlined architecture and glycosylphosphatidylinositol- dependent trafficking in the early secretory pathway of

African trypanosomes. Mol. Biol. Cell *20*, 4739–4750.

Seyfang, A., Mecke, D., and Duszenko, M. (1990). Degradation, recycling, and shedding of *Trypanosoma brucei* variant surface glycoprotein. J. Protozool. *37*, 546–552.

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

Shapiro, S.Z., Naessens, J., Liesegang, B., Moloo, S.K., and Magondu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. Acta Trop. *41*, 313–323.

Shastry, S., and Hancock, W.O. (2010). Neck linker length determines the degree of processivity in Kinesin-1 and Kinesin-2 motors. Curr. Biol. *20*, 939–943.

Shastry, S., and Hancock, W.O. (2011). Interhead tension determines processivity across diverse N-terminal kinesins. Proc. Natl. Acad. Sci. *108*, 16253–16258.

She, Z.-Y., and Yang, W.-X. (2017). Molecular mechanisms of kinesin-14 motors in spindle assembly and chromosome segregation. J. Cell Sci. *130*, 2097–2110.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, K., and Rudenko, G. (2005). Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. Proc. Natl. Acad. Sci. U. S. A. *102*, 8716–8721.

Sheferaw, D., Abebe, R., Fekadu, A., Kassaye, S., Amenu, K., Data, D., Geresu, E., Olbamo, G., Anjulo, A., Yigebahal, Z., et al. (2019). Prevalence of bovine trypanosomosis and vector density in a dry season in Gamo-Gofa and Dawuro Zones, Southern Ethiopia. Vet. Parasitol. Reg. Stud. Reports *18*, 100343.

Shi, H., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C., and Ullu, E. (2000). Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. RNA *6*, 1069–1076.

Siegel, T.N., Gunasekera, K., Cross, G.A.M., and Ochsenreiter, T. (2011). Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. Rev. Trends Parasitol. *27*, 434–441.

Silver, K.E., and Harrison, R.E. (2011). Kinesin 5B is necessary for delivery of membrane and receptors during Fc γ R-mediated phagocytosis. J Immunol *186*, 816–825.

Silvester, E., Young, J., Ivens, A., and Matthews, K.R. (2017). Interspecies quorum sensing in co-infections can manipulate trypanosome transmission potential. Nat. Microbiol. *2*, 1471–1479.

Silvester, E., Ivens, A., and Matthews, K.R. (2018). A gene expression comparison of *Trypanosoma brucei* and *Trypanosoma congolense* in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. PLoS Negl. Trop. Dis. *12*, e0006863.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Ruiz-Postigo, J.A., Fèvre, E.M., Mattioli, R.C., and Jannin, J.G. (2012). Estimating and mapping the population at

risk of sleeping sickness. PLoS Negl. Trop. Dis. *6*, e1859.

Sleigh, J.N., and Sattelle, D.B. (2010). *C. elegans* models of neuromuscular diseases expedite translational research. Transl. Neurosci. *1*, 214–227.

Snow, J.J., Ou, G., Gunnarson, A.L., Walker, M.R.S., Zhou, H.M., Brust-Mascher, I., and Scholey, J.M. (2004). Two anterograde intraflagellar transport motors cooperate to build sensory cilia on *C. elegans* neurons. Nat. Cell Biol. *6*, 1109–1113.

Sosa, H., Dias, D.P., Hoenger, A., Whittaker, M., Wilson-Kubalek, E., Sablin, E., Fletterick, R.J., Vale, R.D., and Milligan, R. a (1997). A model for the microtubule-Ncd motor protein complex obtained by cryo-electron microscopy and image analysis. Cell *90*, 217–224.

Steinberg, G. (2015). Kinesin-3 in the basidiomycete *Ustilago maydis* transports organelles along the entire microtubule array. Fungal Genet. Biol. *74*, 59–61.

Stortz, J.A., Serafim, T.D., Alsford, S., Wilkes, J., Fernandez-Cortes, F., Hamilton, G., Briggs, E., Lemgruber, L., Horn, D., Mottram, J.C., et al. (2017). Genome-wide and protein kinase-focused RNAi screens reveal conserved and novel damage response pathways in *Trypanosoma brucei*. PLoS Pathog. *13*, e1006477.

Stretton, O. (2016). New approaches towards understanding trypanosome mitosis. University of Nottingham.

Suganuma, K., Yamasaki, S., Asada, M., Kawazu, S.I., and Inoue, N. (2012). The epimastigote stage-specific gene expression of CESP is tightly regulated by its 3' UTR. Mol. Biochem. Parasitol. *186*, 77–80.

Suganuma, K., Mochabo, K.M., Hakimi, H., Yamasaki, S., Yamagishi, J., Asada, M., Kawazu, S.I., and Inoue, N. (2013). Adenosine-uridine-rich element is one of the required cis-elements for epimastigote form stage-specific gene expression of the *congolense* epimastigote specific protein. Mol. Biochem. Parasitol. *191*, 36–43.

Sunter, J.D. (2016). A vanillic acid inducible expression system for *Trypanosoma brucei*. Mol. Biochem. Parasitol. *207*, 45–48.

Sutherland, I.A., and Holmes, P.H. (1993). Alterations in drug transport in resistant *Trypanosoma congolense*. Acta Trop. *54*, 271–278.

Tanaka, Y., Kanai, Y., Okada, Y., Nonaka, S., Takeda, S., Harada, A., and Hirokawa, N. (1998). Targeted disruption of mouse conventional kinesin heavy chain, kif5B, results in abnormal perinuclear clustering of mitochondria. Cell *93*, 1147–1158.

Tchamdja, E., Kulo, A.E., Vitouley, H.S., Batawui, K., Bankolé, A.A., Adomefa, K., Cecchi, G., Hoppenheit, A., Clausen, P.H., De Deken, R., et al. (2017). Cattle breeding, trypanosomosis prevalence and drug resistance in Northern Togo. Vet. Parasitol. *236*, 86–92.

Thiede, C., Lakämper, S., Wessel, A.D., Kramer, S., and Schmidt, C.F. (2013). A chimeric kinesin-1 head/kinesin-5 tail motor switches between diffusive and processive motility. Biophys. J. *104*, 432–441.

Thornton, J.W. (2004). Resurrecting ancient genes: Experimental analysis of extinct

molecules. Nat. Rev. *5*, 366–375.

Tihon, E., Imamura, H., Van den Broeck, F., Vermeiren, L., Dujardin, J.C., and Van Den Abbeele, J. (2017). Genomic analysis of Isometamidium Chloride resistance in *Trypanosoma congolense*. Int. J. Parasitol. Drugs Drug Resist. *7*, 350–361.

Tomishige, M., Klopfenstein, D.R., and Vale, R.D. (2002). Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization. Science. *297*, 2263–2266.

Torreele, E., Trunz, B.B., Tweats, D., Kaiser, M., Brun, R., Mazué, G., Bray, M.A., and Pécoul, B. (2010). Fexinidazole - a new oral nitroimidazole drug candidate entering clinical development for the treatment of sleeping sickness. PLoS Negl. Trop. Dis. *4*, e923.

Tschopp, F., Charrière, F., and Schneider, A. (2011). In vivo study in *Trypanosoma brucei* links mitochondrial transfer RNA import to mitochondrial protein import. EMBO Rep. *12*, 825–832.

Turner, C.M.R., Vickerman, K., Barry, J.D., and Maudlin, I. (1988). An estimate of the size of the metacyclic variable antigen repertoire of *Trypanosoma brucei rhodesiense*. Parasitology *97*, 269–276.

Uemura, S., and Ishiwata, S. (2003). Loading direction regulates the affinity of ADP for kinesin. Nat. Struct. Biol. *10*, 308–311.

UN (2014). Programme against African trypanosomosis. Http://Www.Fao.Org/Paat/the-Programme/the-Disease/En/.

Urwyler, S., Studer, E., Renggli, C.K., and Roditi, I. (2007). A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. Mol. Microbiol. *63*, 218–228.

Vale, R.D. (2003). The molecular motor toolbox for intracellular transport. Cell *112*, 467–480.

Varga, V., Helenius, J., Tanaka, K., Hyman, A.A., Tanaka, T.U., and Howard, J. (2006). Yeast kinesin-8 depolymerizes microtubules in a length-dependent manner. Nat Cell Biol *8*, 957–962.

Vassella, E., Reuner, B., Yutzy, B., and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J. Cell Sci. *110*, 2661–2671.

Vaughan, S., and Gull, K. (2008). The structural mechanics of cell division in *Trypanosoma brucei*. Biochem. Soc. Trans. *36*, 421–424.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. Br. Med. Bull. *41*, 105–114.

Walczak, C.E., Gayek, S., and Ohi, R. (2013). Microtubule-depolymerizing kinesins. Annu. Rev. Cell Dev. Biol *29*, 417–441.

Wang, W., Cao, L., Wang, C., Gigant, B., and Knossow, M. (2015). Kinesin, 30 years later: Recent insights from structural studies. Protein Sci. *24*, 1047–1056.

Wang, Z., Morris, J.C., Drew, M.E., and Englund, P.T. (2000). Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. J. Biol. Chem. *275*, 40174–40179.

Wedlich-Söldner, R., Straube, A., Friedrich, M.W., and Steinberg, G. (2002). A balance of KIF1A-like kinesin and dynein organizes early endosomes in the fungus *Ustilago maydis*. EMBO J. *21*, 2946–2957.

Whipple, S. (2017). Dissection of flagellar pocket function in Trypanosoma brucei. University of Nottingham.

White, T.C., Rudenko, G., and Borst, P. (1986). Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Res. *14*, 9471–9489.

WHO (2014). Human African Trypanosomes. Glob Heal Obs.

Wickstead, B., and Gull, K. (2006). A "holistic" kinesin phylogeny reveals new kinesin families and predicts protein functions. Mol. Biol. Cell *17*, 1734–1743.

Wickstead, B., and Gull, K. (2007). Dyneins across eukaryotes: A comparative genomic analysis. Traffic *8*, 1708–1721.

Wickstead, B., and Gull, K. (2011). The evolution of the cytoskeleton. J. Cell Biol. *194*, 513–525.

Wickstead, B., Ersfeld, K., and Gull, K. (2002). Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *125*, 211–216.

Wickstead, B., Ersfeld, K., and Gull, K. (2003a). The mitotic stability of the minichromosomes of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *132*, 97–100.

Wickstead, B., Ersfeld, K., and Gull, K. (2003b). The frequency of gene targeting in *Trypanosoma brucei* is independent of target site copy number. Nucleic Acids Res. *31*, 3993–4000.

Wickstead, B., Gull, K., and Richards, T.A. (2010a). Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. BMC Evol. Biol. *10*, 110.

Wickstead, B., Carrington, J.T., Gluenz, E., and Gull, K. (2010b). The expanded kinesin-13 repertoire of trypanosomes contains only one mitotic kinesin indicating multiple extra-nuclear roles. PLoS One *5*, e15020.

Wilkes, J.M., Mulugeta, W., Wells, C., and Peregrine, A.S. (1997). Modulation of mitochondrial electrical potential: a candidate mechanism for drug resistance in African trypanosomes. Biochem. J. *326*, 755–761.

Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. Annu Rev Biophys *42*, 217–239.

Wirtz, E., and Clayton, C. (1995). Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. Science. *268*, 1179–1183.

Wirtz, E., Hoek, M., and Cross, G.A.M. (1998). Regulated processive transcription of chromatin by T7 RNA polymerase in *Trypanosoma brucei*. Nucleic Acids Res. *26*, 4626–4634.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A.M. (1999a). A tightly regulated inducible expression system for dominant negative approaches in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *99*, 89–101.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A.M. (1999b). A tightly regulated inducible expression system for conditional gene knock-outs and dominant negative approaches in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *99*, 89–101.

Woehlke, G., Ruby, A.K., Hart, C.L., Ly, B., Hom-Booher, N., and Vale, R.D. (1997). Microtubule interaction site of the kinesin motor. Cell *90*, 207–216.

Wurst, M., Seliger, B., Jha, B.A., Klein, C., Queiroz, R., and Clayton, C. (2012). Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. Mol. Microbiol. *83*, 1048–1063.

Yiangou, L., Grandy, R.A., Morell, C.M., Tomaz, R.A., Osnato, A., Kadiwala, J., Muraro, D., Garcia-Bernardo, J., Nakanoh, S., Bernard, W.G., et al. (2019). Method to synchronize cell cycle of human pluripotent stem cells without affecting their fundamental characteristics. Stem Cell Reports *12*, 165–179.

Yildiz, A., Tomishige, M., Vale, R.D., and Selvin, P.R. (2004). Kinesin walks hand-over-hand. Science. *303*, 676–678.

Yonekawa, V., Harada, A., Okada, Y., Funakoshi, T., Kanai, Y., Takei, Y., Terada, S., Noda, T., and Hirokawa, N. (1998). Defect in synaptic vesicle precursor transport and neuronal cell death in KIF1A motor protein-deficient mice. J. Cell Biol. *141*, 431–441.

Zhou, Q., Pham, K.T.M., Hu, H., Kurasawa, Y., and Li, Z. (2019). A kinetochore-based ATM/ATR-independent DNA damage checkpoint maintains genomic integrity in trypanosomes. Nucleic Acids Res.

Zomerdijk, J.C., Kieft, R., and Borst, P. (1992). A ribosomal RNA gene promoter at the telomere of mini-chromosome in *Trypanosoma brucei*. Nucleic Acids Res. *20*, 2725–2734.

Zou, G.M., Wu, W., Chen, J., and Rowley, J.D. (2003). Duplexes of 21-nucleotide RNAs mediate RNA interference in differentiated mouse ES cells. Biol. Cell *95*, 365–371.