THE UNIVERSITY
of EDINBURGH

Chromatin organisation and transcription regulation in *Trypanosoma brucei*

Desislava Plamenova Staneva

Thesis presented for the degree of

Doctor of Philosophy

The University of Edinburgh

2020

Declaration of originality

I declare that this thesis was written by me and the research presented here is the result of my own work, unless otherwise indicated. This thesis has not been submitted for any other degree from this or another university.

Desislava Plamenova Staneva

2020

Lay summary

In various environmental conditions and at different stages of development, organisms turn some of their genes “on” and other genes - “off”. This process also enables cells that make up different organs in a multicellular organism to perform distinct functions, such as gas exchange by the lung and detoxification by the liver. In organisms other than bacteria and related microbes, DNA is located in a cellular compartment called the nucleus where it is wrapped around spheres made of proteins called histones. Specific chemical modifications on the histones recruit other factors that turn genes “on” or “off”.

The aim of this project was to characterise the mechanism of gene regulation in *Trypanosoma brucei*, a single-celled parasite which causes sleeping sickness in Africa. The approach taken was to identify and add fluorescent tags to proteins potentially involved in adding, binding or removing chemical modifications on the histones. The tags were used to visualise under a microscope which candidate proteins are found in the nucleus and may therefore interact with trypanosome DNA to activate or repress genes. In subsequent experiments, antibodies against the tags were used to pull the nuclear proteins out of the cell and examine their associated DNA regions and protein interacting partners.

Overall, 34 proteins were tagged and 8 of them were found to associate with DNA regions which regulate gene activation and repression. Additionally, two of these proteins were found to be part of the same protein complex. These results represent an initial stage of characterisation of gene regulation in an evolutionarily ancient organism.

Abstract

In eukaryotic nuclei, DNA is packaged into chromatin by association with histone proteins. Specific histone post-translational modifications result in the formation of transcriptionally active and silent chromatin domains. The aim of this project was to characterise chromatin organisation and assess the role of transcription regulation in *Trypanosoma brucei*, the causative agent of sleeping sickness in Africa. The approach taken was endogenous fluorescent tagging of putative readers, writers and erasers of histone modifications in *T. brucei* and subsequent use of the tags to localise the candidate proteins in the cell and to identify their genomic associations and protein interaction networks. Chromatin immunoprecipitation followed by sequencing showed that many of the nuclear proteins associate with RNAPII transcription start regions (TSRs). Whereas most proteins were enriched broadly over those regions, Chromo1 and SET27 displayed sharp overlapping peak profiles. Moreover, co-immunoprecipitation and mass spectrometry revealed that Chromo1 and SET27 are likely part of the same complex together with four uncharacterised proteins and JBP2, an enzyme involved in the synthesis of the DNA modification base J. Overall, this work provides the basis for investigating the role of chromatin factors in trypanosome transcription regulation.

Table of contents

List of abbreviations

5mC	5-methylcytosine
ABC	Ammonium bicarbonate
ACN	Acetonitrile
Ago1	Protein argonaute 1
ApoL1	Apolipoprotein L1
ARE	AU-rich element
ATF	Adipose tissue form
ATP	Adenosine triphosphate
BARP	*brucei* alanine-rich protein
BES	Bloodstream expression site
BF	Bloodstream form
bp	Base pair
BSA	Bovine serum albumin
bsr	Blasticidin resistance
cAMP	Cyclic adenosine monophosphate
CDS	Coding sequence
CENP-A	Centromere protein A
ChIP	Chromatin immunoprecipitation
CIR147	Chromosome internal repeats 147

CNS	Central nervous system
CRISPR	Clustered regularly interspaced short palindromic repeats
CTD	C-terminal domain
DABCO	1, 4-diazabicyclo[2.2.2]octane
DAPI	4′,6-diamidino-2-phenylindole
DCL	Dicer-like
DMP	Dimethyl pimelimidate
DNA	Deoxyribonucleic acid
DOT1	Disruptor of telomeric silencing 1
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Egtazic acid
ES	Expression site
ESAG	Expression site-associated gene
ESB	Expression site body
EP	Glutamine-Proline rich procyclin
eYFP	Enhanced yellow fluorescent protein
FACS	Fluorescence-activated cell sorting
FASP	Filtered-aided sample preparation
G1	Gap 1 phase
G2	Gap 2 phase

GFP	Green fluorescent protein
GPEET	Glycine-Proline-Glutamine-Glutamine-Threonine rich procyclin
GPI	Glycosylphosphatidylinositol
H3K4me3	Histone H3 lysine 4 trimethylation
HAT	Histone acetyltransferase or Human African trypanosomiasis
HDAC	Histone deacetylase
HOMedU	Hydroxymethyldeoxyuridine
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidase
IAA	Iodoacetamide
IP	Immunoprecipitation
JBP1	J binding protein 1
JBP2	J biosynthesis protein 2
Jmj	Jumonji domain
K	Kinetoplast
kb	Kilobase
kDa	Kilodalton
kDNA	Kinetoplast DNA
KKT	Kinetoplastid kinetochore protein
KMT	Lysine methyltransferase
LB	Luria broth
LC-MS/MS	Liquid chromatography with tandem mass spectrometry

LSD1	Lysine-specific histone demethylase 1
MBT	Malignant brain tumour
MES	Metacyclic expression site or 2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
N	Nucleus
NECT	Nifurtimox–eflornithine combination therapy
ORF	Open reading frame
PAD	Protein associated with differentiation
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PF	Procyclic form
PHD	Plant homeodomain
PTM	Post-translational modification
PTU	Polycistronic transcription unit
RHS	Retrotransposon hot spot
RBP	RNA binding protein
RNA	Ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
rRNA	Ribosomal RNA

S	Synthesis phase
SDS	Sodium dodecyl sulfate
SET	Suppressor of variegation 3-9, enhancer of zeste and trithorax
SIF	Stumpy induction factor
Sir2	Silent information regulator 2
siRNA	Small interfering RNA
SL	Spliced leader
SLACS	Spliced leader-associated conserved sequence
SRA	Serum resistance-associated
SSR	Strand-switch region
SWI/SNF	Switch/sucrose non-fermentable
TAC	tripartite attachment complex
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TbHpHbR	Trypanosome haptoglobin-haemoglobin receptor
TE	Tris-EDTA
TF	Transcription factor
TFA	Trifluoroacetic acid
TgsGP	*T. b. gambiense*-specific glycoprotein
TH	Thymidine hydroxylase
TLF	Trypanosome lytic factor
TSR	Transcription start region

TTR	Transcription termination region
tRNA	Transfer RNA
UTR	Untranslated region
VSG	Variant surface glycoprotein
WHO	World Health Organisation
wt	Wild type
YFP	Yellow fluorescent protein
Znf	Zinc finger

CHAPTER 1

Introduction

1.1 *Trypanosoma brucei*

Trypanosoma brucei is a flagellated unicellular parasite that causes African trypanosomiasis, also known as sleeping sickness in humans and nagana in animals. It was discovered as the causative agent of cattle trypanosomiasis in the 19th century by Sir David Bruce (Bruce, 1895; Steverding, 2008), after whom the parasite is named.

T. brucei belongs to the order Kinetoplastida together with other important pathogens, including *Trypanosoma cruzi* which causes Chagas disease in Latin America and *Leishmania* species causing leishmaniasis worldwide. Kinetoplastids were previously classified as members of the Excavata supergroup of eukaryotes (Walker et al., 2011) but this grouping was based on morphology and not monophyly. More recently, the excavate lineage was split into three, with kinetoplastids classified as part of the group of discobids (Figure 1.1) (Keeling and Burki, 2019). As illustrated in Figure 1.1, kinetoplastids are evolutionarily distant from well-studied model organisms, such as budding and fission yeasts, and as a consequence many aspects of their biology differ from the common eukaryotic norm. For instance, a characteristic feature of these organisms is the presence of a single large mitochondrion in the cell where mitochondrial DNA (termed “kinetoplast DNA” or “kDNA”) is organised as a network of catenated maxi- and mini-circles, the former encoding rRNAs and respiratory chain proteins and the latter - guide RNAs for editing of mitochondrial protein transcripts (Cavalcanti and De Souza, 2018; Lukeš et al., 2002; Shapiro and Englund, 1995).

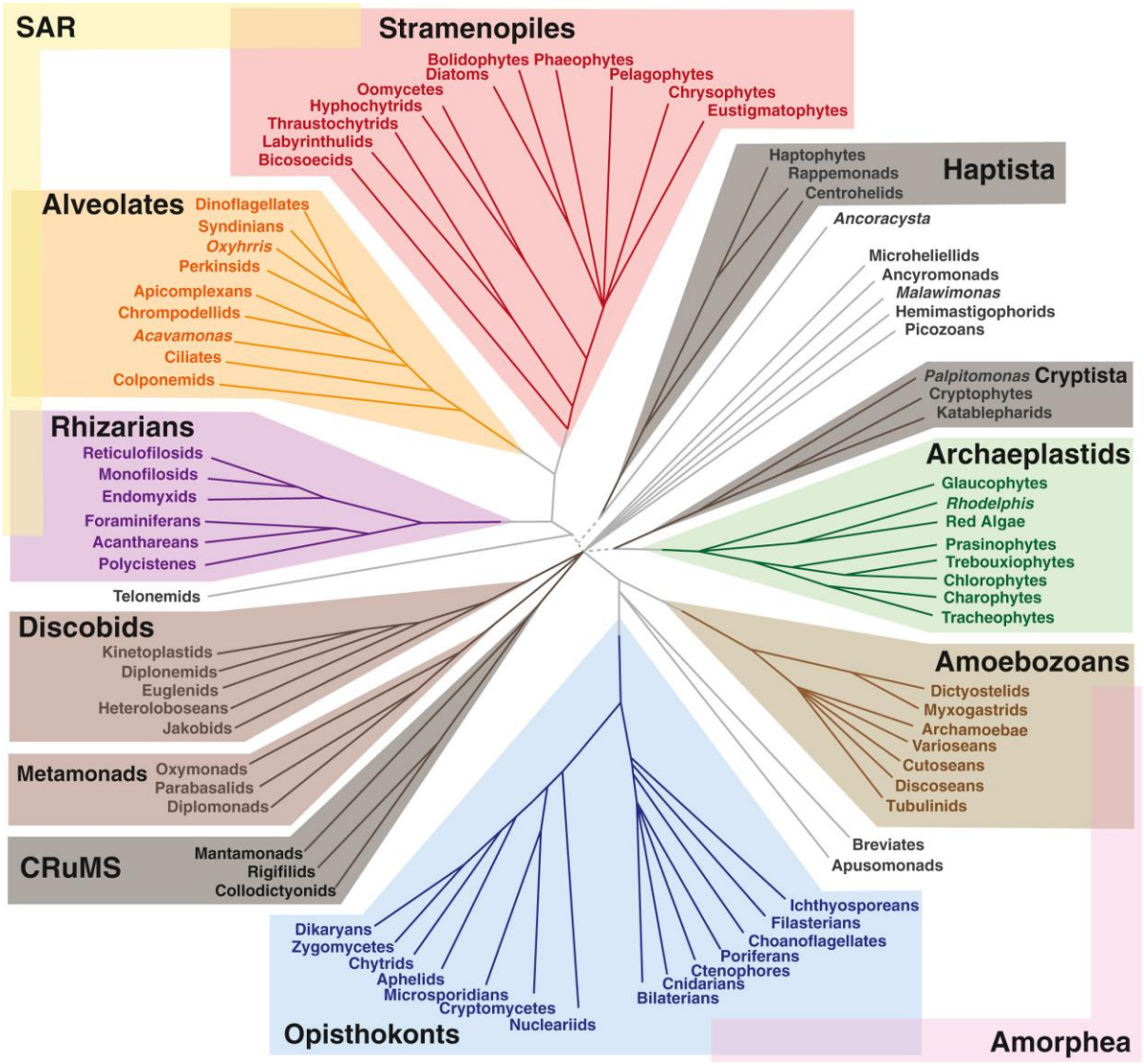


Figure 1.1 Evolutionary tree of eukaryotes. Most commonly studied model organisms are within the Opisthokonta group whereas kinetoplastids are part of the evolutionarily distant group of Discobids. Figure from Keeling and Burki (2019).

1.1.1 African trypanosomiasis

The three subspecies of *T. brucei* (*T. b. gambiense*, *T. b. rhodesiense* and *T. b. brucei*) are endemic to sub-Saharan Africa where they are transmitted to humans and animals by the tsetse fly (genus *Glossina*) (Brun et al., 2010). The people most vulnerable to infection live in rural areas where they rely primarily on agriculture, hunting and fishing for their income (WHO, 2019). Thus, African trypanosomiasis is exerting not just a health burden but also economic hardship on affected communities.

Only *T. b. gambiense* and *T. b. rhodesiense* are human-infective, the former being present in western and central Africa and the latter – in eastern and southern Africa (Figure 1.2). *T. b. gambiense* causes a chronic form of the disease and is responsible for 98% of the sleeping sickness cases whereas *T. b. rhodesiense* causes an acute infection that accounts for 2% of the cases (WHO, 2019). Sleeping sickness progresses in two stages with characteristic symptoms and localisation of parasites in the body. In the early form of the disease, known as the haemolymphatic stage, trypanosomes multiply in the blood, lymph, interstitial spaces as well as the adipose tissue (Trindade et al., 2016) and the skin (Caljon et al., 2016; Capewell et al., 2016). The symptoms at this stage include headache, pruritus and bouts of fever (Büscher et al., 2017). In the second (meningoencephalitic or neurological) stage of sleeping sickness, trypanosomes cross the blood-brain barrier and invade the central nervous system (CNS) which gives rise to neurological symptoms including confusion, poor coordination and the sleep-wake cycle disturbances after which the disease is named (Büscher et al., 2017).

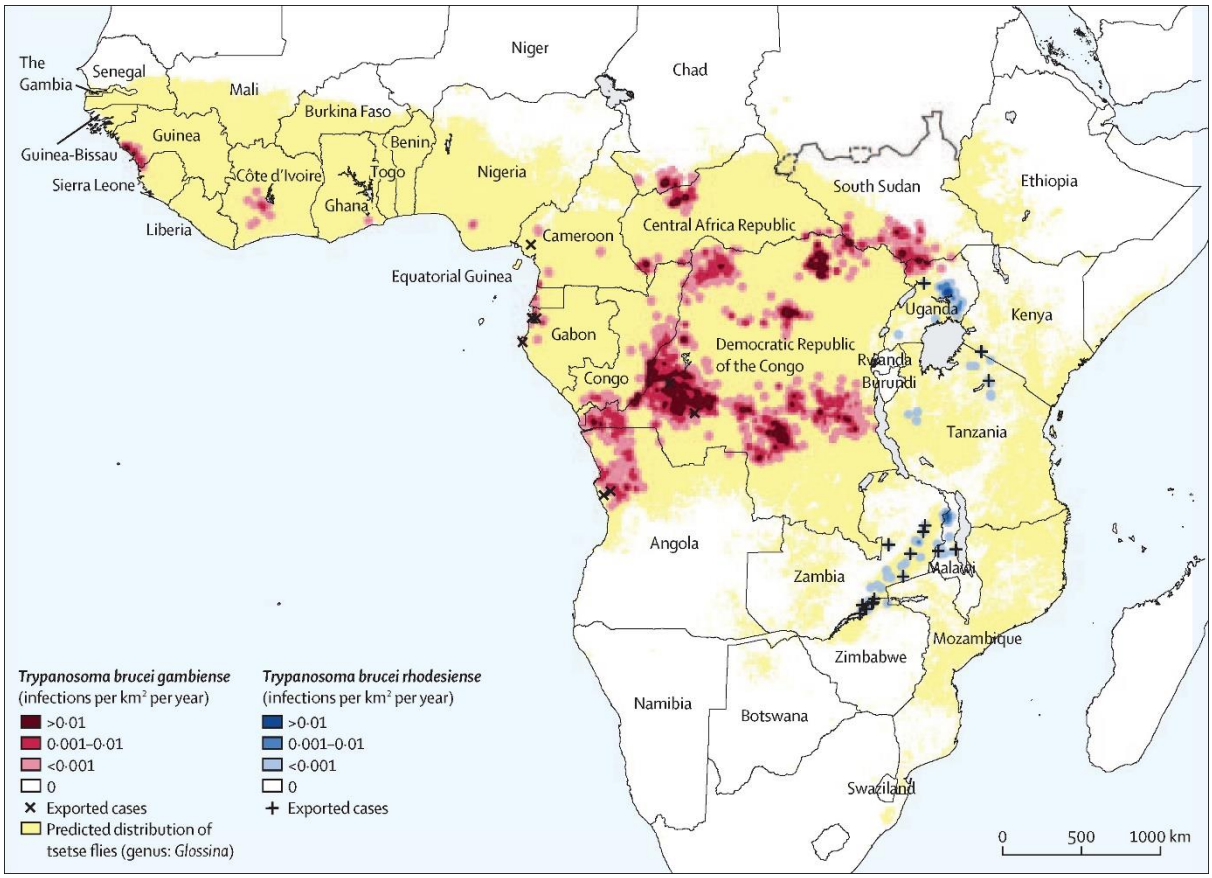


Figure 1.2 Distribution of sleeping sickness cases caused by *T. b. gambiense* and *T. b. rhodesiense* during the 2010-2014 period. Figure from Büscher et al. (2017).

Without treatment, sleeping sickness is almost always fatal, thus early diagnosis of the disease is paramount. Ideally, trypanosome infections should be confirmed microscopically in a blood smear, lymph node aspirate or, during the second stage of the disease, cerebrospinal fluid. This applies particularly to *T. b. rhodesiense* infections because of the high level of parasitaemia caused by this subspecies (Kennedy, 2019). In the case of *T. b. gambiense*, the parasitaemia is cyclical and thus direct observation of a trypanosome infection in a patient is more difficult and alternative serological methods have been used (Kennedy, 2019).

Distinguishing between the early and late form of the disease is also highly important because of the toxic nature of drugs used to treat the neurological stage. Currently, first-line treatment for early stage *T. b. gambiense* infection is intramuscular or intravenous administration of pentamidine which disrupts mitochondrial function of the parasite (Lanteri et al., 2008). Early stage *T. b. rhodesiense* infection is treated with intravenous administration of suramin which may act by inhibiting glycolysis (Morgan et al., 2011) on which the parasites are entirely dependent for energy in the mammalian bloodstream (Priest and Hajduk, 1994). While both pentamidine and suramin are generally well tolerated by patients, the drugs currently in use for late stage sleeping sickness are highly toxic. Melarsoprol is an arsenic drug used to combat CNS infection with both *T. b. rhodesiense* and *T. b. gambiense* and causes post-treatment reactive encephalopathy in about 10% of cases, half of which are fatal (Brun et al., 2010). Second stage *T. b. gambiense* infection is also treated with nifurtimox–eflornithine combination therapy (NECT), with bone marrow toxicity and the potential for septicaemia being the major adverse effects of this treatment (Field et al., 2017; Kennedy, 2013). Melarsoprol may act by inhibiting mitosis (Thomas et al., 2018), nifurtimox – through oxidative stress (Denise and Barrett, 2001) or the production of cytotoxic nitrile metabolites (Hall et al., 2011) and eflornithine – by inhibiting ornithine decarboxylase (Bacchi et al., 1980). All of the treatments discussed here have been in use for more than a decade (some for over 50 years) and there is a growing concern about trypanosome drug resistance. In particular, resistance to both pentamidine and melarsopol caused by mutations in the transporters involved in their uptake has been documented (Baker et al., 2013). New drugs are currently being developed, including the orally administered fexinidazole which was shown to be effective in 91% of patients with a late stage sleeping sickness caused by *T. b. gambiense* in a randomised trial conducted in the Democratic Republic of Congo (Mesu et al., 2018). Fexinidazole acts by inhibiting DNA synthesis and was recently approved for treatment of both early and late stage sleeping sickness (Deeks, 2019).

Disease prevention is done predominantly via vector population control as vaccination is deemed unfeasible due to the parasite's ability to change its surface antigen presentation (see section 1.1.3).

Currently, the prevalence of sleeping sickness is declining (Figure 1.3), with only 1446 new cases reported in 2017, and the disease is targeted by the World Health Organisation (WHO) for elimination by 2020 (WHO, 2019). This may be problematic for *T. b. rhodesiense* as it predominantly infects animals which can act as a reservoir for the disease (Simarro et al., 2014).

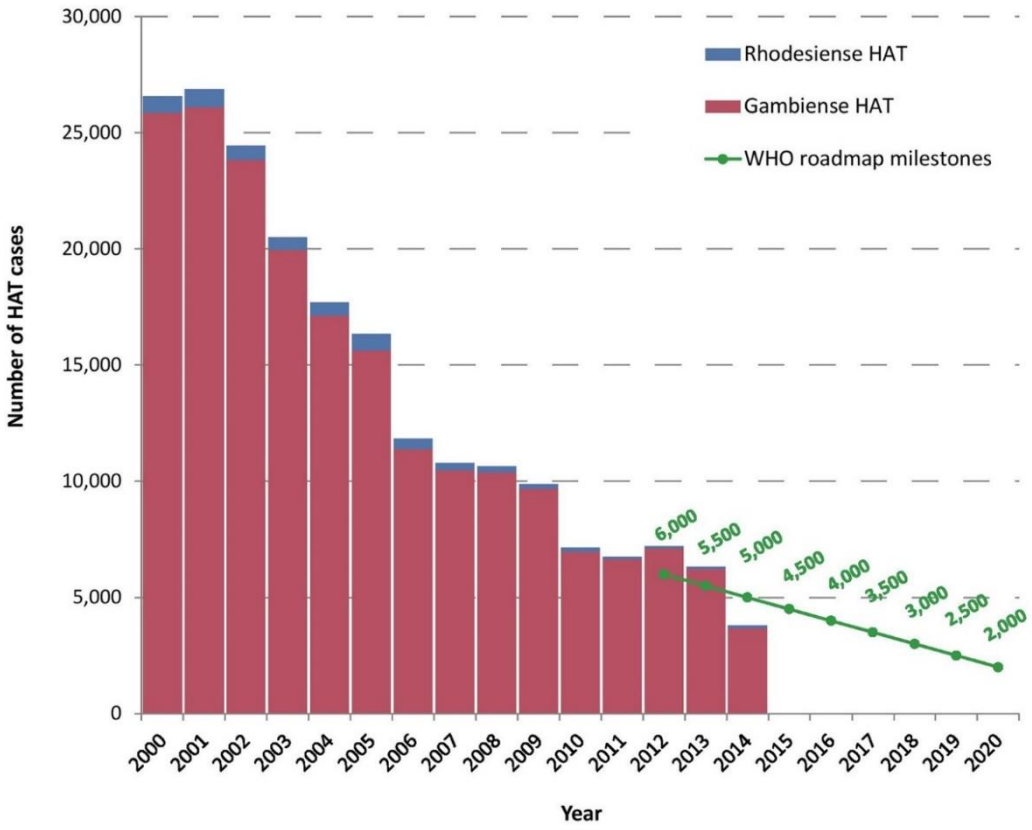


Figure 1.3 Sleeping sickness cases caused by *T. b. rhodesiense* and *T. b. gambiense* for the period 2000-2014 with projected cases for the period 2015-2020. HAT – human African trypanosomiasis. Figure from Franco et al. (2017).

Animal African trypanosomiasis, also known as nagana, is caused predominantly by *Trypanosoma congolense* and *Trypanosoma vivax*, with *T. brucei* accounting for a minor proportion of cases (Morrison et al., 2016). It is characterised by anaemia, emaciation, infertility and loss of productivity (Giordani et al., 2016). While human sleeping sickness is on track to be eliminated in the near future, nagana remains one of the most significant livestock infections in sub-Saharan Africa, hindering agricultural development in rural areas (Morrison et al., 2016).

Humans and some primates have innate immunity against several trypanosome species, including *T. b. brucei*, owing to trypanosome lytic factors (TLFs) present in their serum. TLFs are high-density lipoprotein particles containing apolipoprotein L1 (ApoL1), which kills the parasites by forming ionic pores in their endolysosomal membranes (Perez-Morga, 2005; Vanhamme et al., 2003). ApoL1 is internalised primarily via the trypanosome haptoglobin-haemoglobin receptor (TbHpHbR) (Vanhollebeke et al., 2008). The initial association with endosomes is critical for ApoL1 function because membrane insertion of this protein is dependent on acidic pH (Hager, 1994; Lecordier et al., 2009). *T. b. rhodesiense* and *T. b gambiense* have evolved different mechanisms for escaping the trypanolytic action of ApoL1. *T. b. rhodesiense* expresses a serum resistance-associated (SRA) protein which is targeted to the endolysosomal system of the parasite where it neutralizes ApoL1 via strong interaction with the C-terminal helix of ApoL1 (Lecordier et al., 2009; Van Xong et al., 1998; Vanhamme et al., 2003). The *T. b. gambiense* subspecies is divided into two genetically distinct groups which differ in their resistance mechanisms to human serum. Group 1 resistance is due to an amino acid substitution in TbHpHbR which reduces ApoL1 uptake (Dejesus et al., 2013). Additionally, group 1 resistance is conferred by *T. b. gambiense*-specific glycoprotein (TgsGP) which induces stiffening of endolysosomal membranes (Uzureau et al., 2013), presumably inhibiting ApoL1 insertion. Group 2 *T. b. gambiense* exhibit variable resistance to human serum. They have been shown to internalize TLF and therefore their resistance could be due

to neutralisation or countering the effects of ApoL1 as opposed to impaired uptake (Capewell et al., 2011).

1.1.2 Trypanosome cell biology

The trypanosome cell (Figure 1.4) is spindle-shaped with a single flagellum emerging from an invagination of the cell membrane termed the flagellar pocket which is located in the posterior end of the cell and is the only known site of exo- and endocytosis (Overath and Engstler, 2004). The flagellum is attached lengthwise until the anterior end of the parasite and extends beyond the cell body. Cell morphology is maintained by a subpellicular array of microtubules, oriented with their plus ends towards the posterior and with their minus ends towards the anterior of the cell (Robinson, 1995). Many trypanosome organelles are highly positioned and single-copy: the nucleus, mitochondrion (including the kinetoplast), Golgi, terminal lysosome and flagellum, and thus their accurate segregation during cell division is of vital importance.

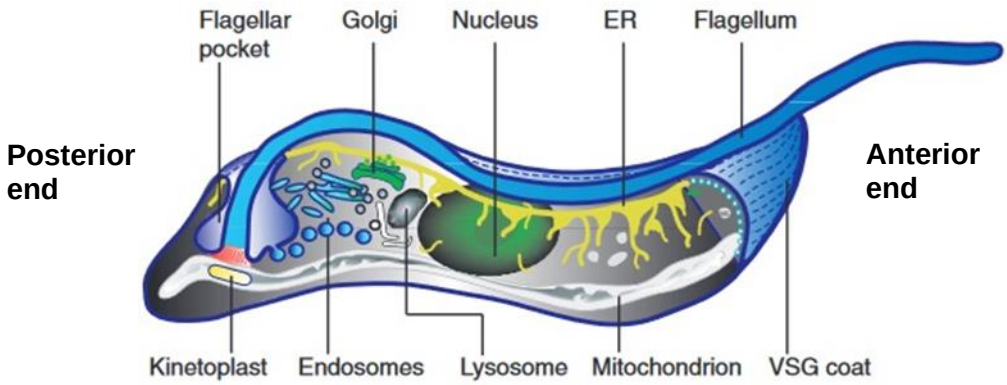


Figure 1.4 *T. brucei* cell architecture. The positions of major organelles are highlighted. The flagellar pocket and the kinetoplast are at the posterior end whereas the flagellum extends beyond the cell body from the anterior end. Figure from Grunfelder et al. (2003).

In laboratory culture, *T. brucei* undergoes propulsive movement interspersed with periods of tumbling or irregular movement (Langousis and Hill, 2014). When the culture media viscosity was designed to match that of blood, the proportion of trypanosomes undergoing propulsive movement increased, demonstrating that the movement of bloodstream form *T. brucei* is fine tuned to its host environment (Heddergott et al., 2012). In the fly host trypanosomes exhibit social motility, a form of en masse movement in response to a particular stimulus. Trypanosome social motility depends on flagellar cAMP signalling which is in turn required for progression of *T. brucei* through the fly host tissues and completion of its life cycle (Shaw et al., 2019).

The basal body of the flagellum is linked directly to the kinetoplast genome via the tripartite attachment complex (TAC) which traverses the double mitochondrial membrane (Ogbadoyi et al., 2003). This physical link between the basal body and the kinetoplast provides a mechanistic explanation for their coordinate replication and division (Robinson and Gull, 1991). Duplication of the basal body and nucleation of a daughter flagellum are the earliest detectable cytological events during the trypanosome cell cycle (Woodward and Gull, 1990). *T. brucei* undergoes a closed mitosis, with the mitotic spindle forming inside the nucleus (Ogbadoyi et al., 2000). No appreciable condensation of chromosomes has been observed during trypanosome cell division. The kinetoplast replicates before the nuclear genome and is segregated together with the basal body prior to the onset of mitosis (Sherwin et al., 1989). This allows the stages of the cell cycle to be distinguished (microscopically or via FACS sorting) by DNA staining (Figure 1.5). In G1 phase, cells have one kinetoplast and one nucleus which is termed 1K1N configuration. As the cell cycle progresses, the basal body duplicates, driving the segregation of the kinetoplast. The kinetoplast starts dividing (has elongated morphology) during the nuclear S phase. In G2, the kinetoplast has divided and the cells are in a 2K1N configuration. Finally, post-mitotic cells have two kinetoplasts and two nuclei (2K2N).

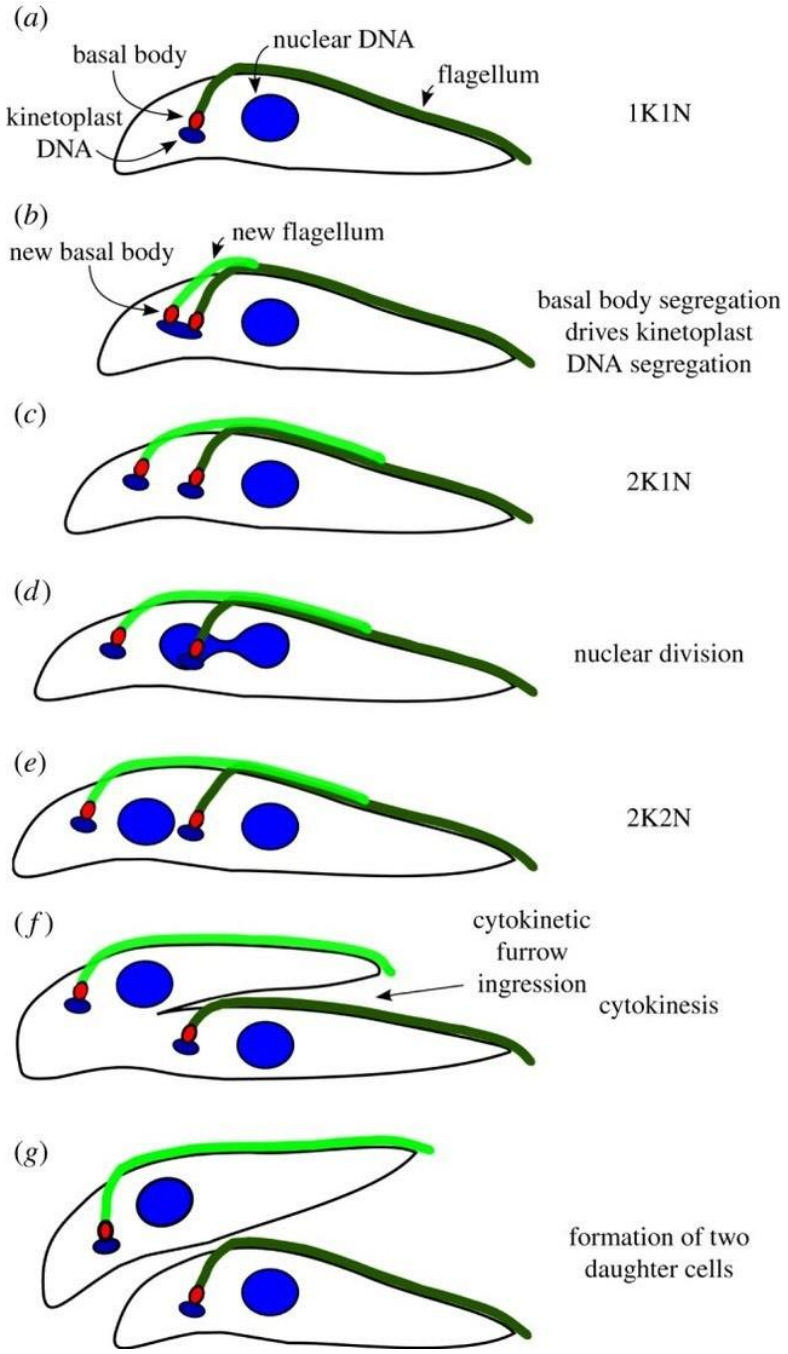


Figure 1.5 *T. brucei* cell cycle. (a) G1 cells have a single kinetoplast and nucleus (1K1N). (b) Coordinate division of the basal body and the kinetoplast. (c) The kinetoplast has divided during the nuclear S phase (2K1N). (d) The nuclear genome enters mitosis. (e) Post-mitotic cells have two kinetoplasts and two nuclei (2K2N). (f) Cytokinesis proceeds via cytokinesis furrow ingression between the old and new flagellum. (g) Both daughter cells have inherited a single kinetoplast, nucleus and flagellum. Figure from Akiyoshi and Gull (2013).

1.1.3 Life cycle

During its life cycle, *T. brucei* alternates between a mammalian host and the tsetse fly vector (Figure 1.6). The within and between host transitions are characterised by specific morphological and surface proteome changes as well as metabolic adaptations that allow the parasites to survive in the varying host environments they encounter.

There are two morphotypes during the developmental cycle of the parasite – a trypomastigote and an epimastigote form. *T. brucei* spends the majority of its life cycle as a trypomastigote which is characterised by posterior localisation of the basal body and the kinetoplast. The latter are more central in the cell and anterior to the nucleus in the epimastigote form (Langousis and Hill, 2014).

The parasite is extracellular at all developmental stages. The long slender form trypanosomes proliferate in the mammalian bloodstream until they reach a critical density when they differentiate to the division-arrested short stumpy form. This process is triggered by the accumulation of the stumpy induction factor (SIF), an extracellular signal internalised via an oligopeptide transporter related to G protein-coupled receptors (Rojas et al., 2019). Accumulation of non-dividing stumpy forms limits the parasite burden and prolongs host survival as well as adapting the parasite for transmission to the tsetse (MacGregor et al., 2012). Recently, it has been shown that the skin and adipose tissue are additional reservoirs of *T. brucei* in the mammalian host (Caljon et al., 2016; Capewell et al., 2016; Trindade et al., 2016). Adipose tissue forms (ATFs) were found to be transcriptionally and metabolically distinct from bloodstream form (BF) parasites but were also capable of differentiating to the stumpy form.

Stumpy cells are ingested by the tsetse during a blood meal and, once in the fly midgut, they sense a drop in temperature and the presence of *cis*-aconitate and citrate, which triggers their differentiation to the proliferating procyclic form (PF) (Engstler, 2004). Proteases in the fly midgut may also contribute to the production of the differentiation signal. Proteins associated with differentiation (PAD) are carboxylate transporter protein family members required for the response to *cis*-aconitate and citrate (Dean et al., 2009). PAD2 expression is upregulated upon the decrease in temperature in the fly whereas PAD1 is a stumpy-specific surface protein.

Procyclic cells migrate from the tsetse midgut to the fly proventriculus where they undergo an asymmetric cell division which generates one short and one long epimastigote form. Short epimastigotes attach to the salivary gland epithelium of the fly and divide symmetrically until eventually an asymmetric division gives rise to the metacyclic form which detaches from the epithelium and can infect the mammalian host, thus completing the life cycle of *T. brucei* (Rotureau et al., 2012). The fly salivary glands are also the site of haploid gamete production via meiosis and their subsequent fusion facilitated by flagellar interactions (Peacock et al., 2014).

Laboratory cell lines of bloodstream form trypanosomes can be pleomorphic or monomorphic depending on their differentiation capacity. Pleomorphic cells can differentiate to the stumpy form whereas monomorphic cells have reduced this ability and exist only in the slender form.

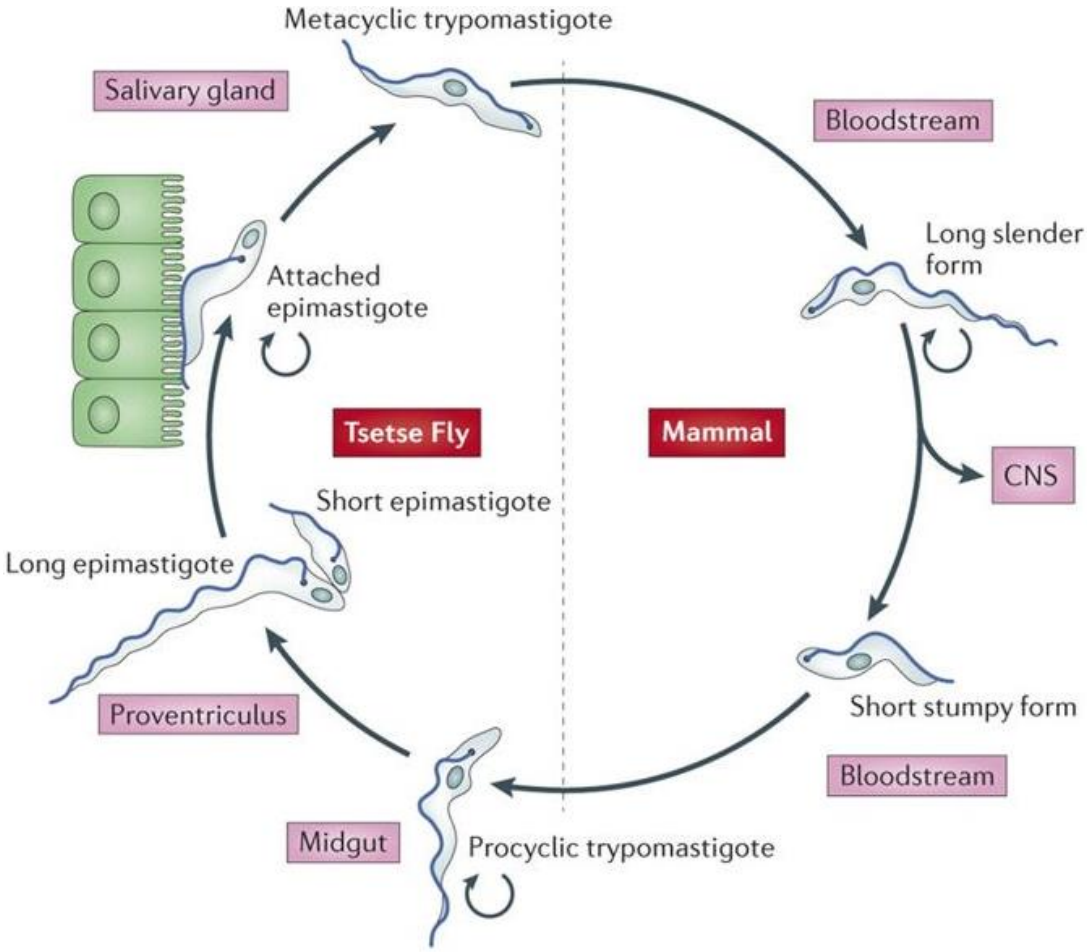


Figure 1.6 Life cycle of *T. brucei*. Major parasite developmental stages are depicted. Dividing forms are marked with circular arrows. Figure from Langousis and Hill (2014).

In the mammalian host, *T. brucei* is covered by a dense surface coat made of ~10 million copies of a glycosylphosphatidylinositol (GPI)-anchored variant surface glycoprotein (VSG). VSGs are homodimers, with each monomer consisting of a variable N-terminal domain which is membrane-distal and exposed, and a conserved C-terminal domain which links the VSG to the plasma membrane (Carrington et al., 1991; Schwede et al., 2011). The VSG coat protects trypanosomes by shielding invariant surface antigens from the mammalian immune system (Ziegelbauer and Overath, 1993). Additionally, a periodic change of the surface coat, termed

antigenic variation, helps *T. brucei* evade the host immune response and maintain a chronic infection (Vickerman, 1978). Normally, only a single VSG is expressed from an archive consisting of ~2000 genes (Horn, 2014). *T. brucei* periodically switches to express a new VSG to which no antibodies have been produced, leading to waves of parasitaemia (Figure 1.7). Up to ~100 different VSG variants can be present in the population at peak parasitaemia (Mugnier et al., 2015). VSG expression is discussed in more detail in section 1.3.2. The VSG coat is exceptionally fluid due to attachment via a GPI anchor as opposed to a transmembrane domain. This fluidity facilitates coat recycling through the endocytic system which is important for rapid VSG turnover during switching as well as clearing of surface-bound antibodies present at low concentrations (Engstler et al., 2007; Natesan et al., 2011; Pal et al., 2003; Seyfang et al., 1990).

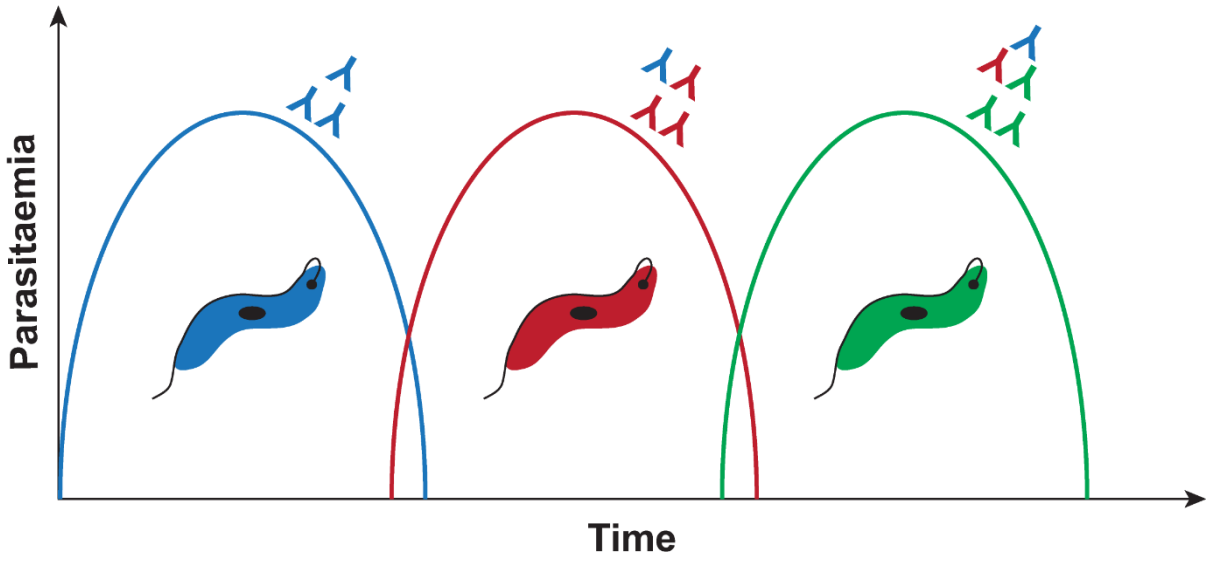


Figure 1.7 *T. brucei* antigenic variation leads to relapsing parasitaemia. During a trypanosome infection, some parasites switch to express new VSGs (depicted by the different colours) that are not immediately recognised by the immune system. Antibodies are depicted as Y-shaped structures.

In the fly midgut, procyclic cells have a different glycoprotein surface coat made of EP (containing Glu-Pro repeats) and GPEET (containing Gly-Pro-Glu-Glu-Thr repeats) procyclins (Roditi and Liniger, 2002). Later during the tsetse infection, *T. brucei* loses the GPEET procyclins (Vassella et al., 2000) whereas the EP coat persists until the epimastigote stage in the salivary glands when the parasites are covered by a *brucei* alanine-rich protein (BARP) (Urwyler et al., 2007). Upon differentiation of the epimastigotes to the metacyclic form, the BARP coat is replaced by VSG.

Apart from changes in the surface coat, trypanosomes undergo significant metabolic reprogramming during their developmental cycle. In the mammalian bloodstream, *T. brucei* generates ATP exclusively via glycolysis (Haanstra et al., 2011), with ATFs having upregulated expression of enzymes involved in fatty acid oxidation (Trindade et al., 2016). Kinetoplastids and their sister group Diplonemida possess unique peroxisome-like organelles called glycosomes which sequester most of the glycolytic enzymes (Allmann and Bringaud, 2017; Opperdoes and Borst, 1977). In the fly, glucose is replaced by proline as the main energy source (Bursell, 1963). Proline is actively imported (L'Hostis et al., 1993) and metabolised in the parasite mitochondrion via oxidative phosphorylation (van Hellemond et al., 2005). Consequently, procyclic cells possess a more active and elaborate mitochondrion than slender bloodstream forms (Vickerman, 1985). In preparation for transition from the mammalian to the fly host, stumpy forms activate their mitochondrion, as evidenced by the upregulation of mitochondrial respiration genes (Brown et al., 1973; Michelotti and Hajduk, 1987; Van Grinsven et al., 2009) and the expansion and elaboration of the mitochondrial ultrastructure (Silvester et al., 2017; Vickerman, 1965).

1.1.4 Genome organisation

The genome of *T. brucei* is ~35 Mb and is organized into 11 megabase (1-6 Mb), 1-5 intermediate (200-900 kb) and about 100 minichromosomes (50-150 kb) (El-Sayed et al., 2000) (Figure 1.8). The megabase chromosomes are diploid whereas the intermediate and minichromosomes are probably haploid (Ersfeld, 2011). The chromosomes are linear and end with typical telomeric repeats (TTAGGG). Trypanosome telomere length increases by ~ 8 bp per cell division owing to constitutive telomerase activity (Horn et al., 2000). This increase is combined with an occasional loss of large portion of the telomeric repeats (Bernards et al., 1983; Pays et al., 1983). Thus, telomere length can differ significantly between individual parasites in the population (3-20 kb in the Lister 427 strain) (Dreesen and Cross, 2008). A significant proportion of the trypanosome genome (~15%) is dedicated to providing a large reservoir of VSG genes. The diploid megabase chromosomes harbour about 9000 genes, including housekeeping genes as well as a major proportion of the ~2000 VSGs, which are located in long subtelomeric arrays (Berriman et al., 2005). A smaller subset of VSG genes and VSG gene fragments is also found on the intermediate and minichromosomes.

The vast majority of *T. brucei* genes are intronless, with the exception of a poly(A) polymerase and a DNA/RNA helicase (Jaé et al., 2010; Mair et al., 2000; Siegel et al., 2010). Most genes are transcribed in polycistronic units, with some units containing more than 100 genes. The housekeeping genes are predicted to be expressed from ~150 transcription units, with a mean of 153 kb and 55 ORFs (Daniels et al., 2010). The single active VSG is transcribed from one of several specialised subtelomeric regions termed VSG expression sites (ESs) found on the megabase and intermediate chromosomes. Bloodstream parasites express VSG from one of ~15 polycistronic bloodstream expression sites (BESs) whereas metacyclic forms in the fly salivary glands express VSG from one of 5 monocistronic metacyclic expression sites (MESs) (Alarcon et al., 1994; Graham and Barry, 1995; Hertz-Fowler et al., 2008; Kolev et al., 2017). Trypanosome gene expression is discussed in more detail in section 1.3.

T. brucei centromeres have been mapped by topoisomerase II activity to AT-rich repetitive regions stretching over several kb on the megabase chromosomes (Obado et al., 2007). These repeats are ~147 bp long, and are thus termed chromosome internal repeats 147 (CIR147) (Patrick et al., 2009). They are usually flanked by rRNA genes and retrotransposons. Eukaryotic centromeres typically contain CENP-A, a histone H3 variant which specifies the site of kinetochore assembly (Buscaino et al., 2010; Earnshaw and Rothfield, 1985). Kinetoplastids lack CENP-A and the ~50 kinetochore proteins conserved from yeast to human, and instead possess a set of about 20 unconventional kinetoplastid kinetochore proteins (KKTs) that do not have a homolog in other eukaryotes (Akiyoshi and Gull, 2014). The attachments of megabase chromosomes to the spindle are mediated by KKT4 (Llauró et al., 2018).

No site-specific accumulation of topoisomerase II activity has been detected on the intermediate and minichromosomes which may not possess canonical centromeres. Additionally, the number of spindle microtubules was found to be insufficient for one microtubule to attach to each chromosome (Wickstead et al., 2004). Intermediate chromosomes are composed of a non-repetitive core and 177 bp subtelomeric repeats and contain VSG ESs (Figure 1.8), suggesting these chromosomes may have originated via subtelomeric breakage of megabase chromosomes (Berriman et al., 2002). Minichromosomes have a large palindromic core made of 177 bp repeats with a single inversion point in the centre and harbour VSG genes in their subtelomeric regions (Figure 1.8). They are partitioned faithfully during cell division via lateral attachments to the spindle microtubules prior to the segregation of the megabase chromosomes (Alsford et al., 2001; Ersfeld and Gull, 1997). Surprisingly, cohesin, which is normally required for keeping the sister chromatids together during cell division, was found to be dispensable during minichromosome segregation (Bessat and Ersfeld, 2009b).

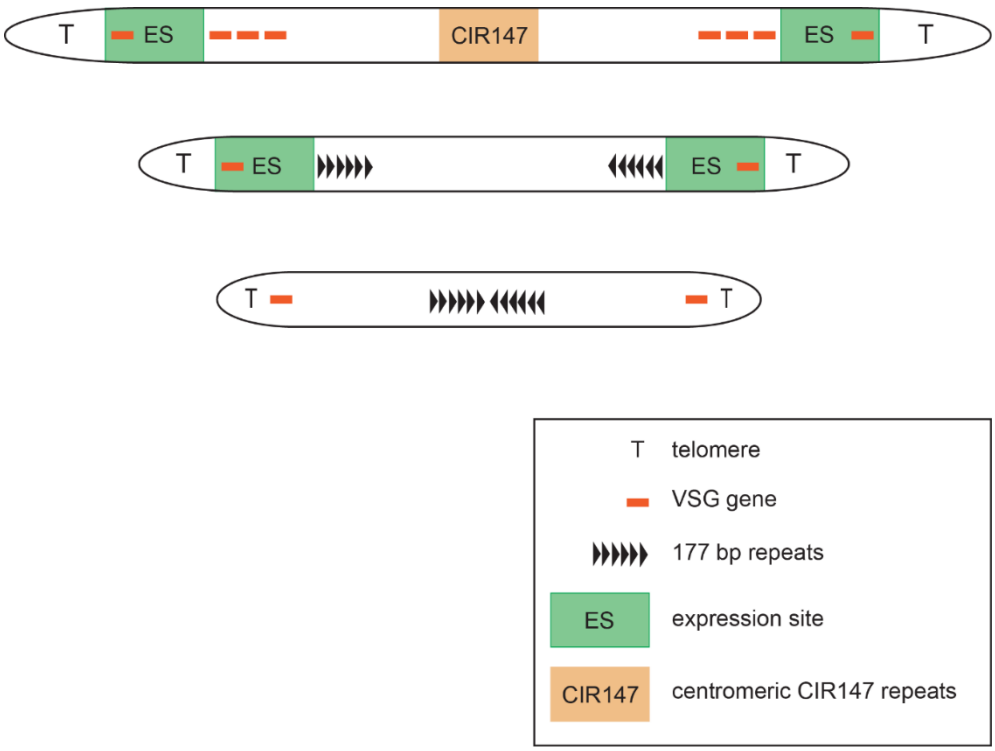


Figure 1.8 Trypanosome chromosome organisation. From top to bottom: megabase, intermediate and minichromosomes (not drawn to scale). All chromosome types end with canonical telomeric repeats. Centromeric CIR147 repeats have been found only on the megabase chromosomes. VSGs are transcribed from subtelomeric ESs (BESs in the bloodstream form and MESs in the metacyclic form) on the megabase and intermediate chromosomes. Most of the silent VSG genes are located in subtelomeric arrays on the megabase chromosomes. 177 bp repeats are found in the subtelomeres of intermediate and the core of minichromosomes. Minichromosomes harbour an additional reservoir of silent VSG genes.

1.2 Chromatin

In eukaryotic nuclei, DNA is packaged into chromatin by association with histone proteins (Li and Reinberg, 2011). This reduces significantly the length of DNA and allows eukaryotic genomes to fit into relatively small nuclei. Additionally, chromatin regulates gene expression, minimizes DNA damage and genomic instability, and facilitates chromosome segregation during cell division.

The fundamental unit of chromatin is the nucleosome which consists of ~147 bp of DNA wrapped around a histone octamer composed of one histone H3/H4 tetramer and two histone H2A/H2B dimers (Kornberg, 1974; Luger et al., 1997). One or more of these core histones can be substituted for a histone variant which gives nucleosomes new functional properties (Weber and Henikoff, 2014). Histone H1 binds to linker DNA between neighbouring nucleosomes and facilitates higher order compaction of chromatin.

1.2.1 Active and silent chromatin

The two main states of chromatin are euchromatin which is found mainly in the nuclear interior and heterochromatin which is located predominantly in the nuclear periphery. An exception from this architecture are the inverted nuclei found in the rod cells of the retina in some nocturnal mammals (Solovei et al., 2009). In these inverted nuclei, heterochromatin is found in the nuclear centre whereas euchromatin is localised in the nuclear periphery. Euchromatin and heterochromatin were initially distinguished microscopically by Emil Heitz in 1928 as differentially stained chromosomal regions during the cell cycle (Passarge, 1979). Euchromatin is more open and accessible to the transcription machinery whereas heterochromatin is highly condensed and inaccessible (Allshire and Madhani, 2018). Heterochromatic regions can be further subdivided into constitutive heterochromatin which is always repressive and facultative heterochromatin which can decondense when triggered by specific environmental and developmental stimuli. Constitutive heterochromatin is found at

centromeres, telomeres and transposable elements, and is involved in the faithful segregation of chromosomes during cell division as well as the suppression of transposons. Facultative heterochromatin is found, for example, on the inactive X chromosome of female mammals.

Chromatin organisation into transcriptionally active and repressive domains depends on the presence of histone variants, DNA methylation and post-translational modifications (PTMs) of the histones. For instance, the H2AZ histone variant is thought to destabilise nucleosomes near transcription start sites (TSSs), thus facilitating the opening of chromatin and transcription initiation (Henikoff and Smith, 2015). Conversely, methylation of cytosine residues in DNA represses gene expression (Jaenisch and Bird, 2003). The effect of histone PTMs is discussed in more detail in section 1.2.2.

Additionally, RNA interference (RNAi) is known to provide a link between transcriptional and post-transcriptional silencing. In many eukaryotes, siRNA producing loci, such as retrotransposable elements, recruit the RNAi machinery which stimulates the formation of heterochromatin leading to transcriptional silencing of these regions (Allshire and Madhani, 2018).

1.2.2 Readers, writers and erasers of histone modifications

Histones are globular proteins with unstructured N-terminal tails protruding from the nucleosome core. The tails are rich in basic amino acids and are subject to various post-translational modifications (Table 1.1) which are deposited by “writers”, interpreted by “readers” and removed by ”erasers” (Torres and Fujimori, 2015) (Figure 1.9). The globular C-terminal domains of histones can also be post-translationally modified. Reader, writer and eraser activities may be contained within the same protein or protein complex which could, for example, aid heterochromatin spreading (reader + writer activity). This project focused on acetylation and methylation which are the best characterised histone PTMs.

Table 1.1 Overview of histone post-translational modifications

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	**K**-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	**K**-me1 **K**-me2 **K**-me3	Transcription, Repair
Methylation (arginines)	**R**-me1 **R**-me2a **R**-me2s	Transcription
Phosphorylation	**S**-ph **T**-ph	Transcription, Repair, Condensation
Ubiquitylation	**K**-ub	Transcription, Repair
Sumoylation	**K**-su	Transcription
ADP ribosylation	**E**-ar	Transcription
Deimination	**R > Cit**	Transcription
Proline Isomerization	**P-cis > P-trans**	Transcription

Some PTMs, such as phosphorylation, can occur only once per amino acid residue whereas other PTMs, such as methylation, can occur mutiple times per amino acid residue. Table from Kouzarides (2007).

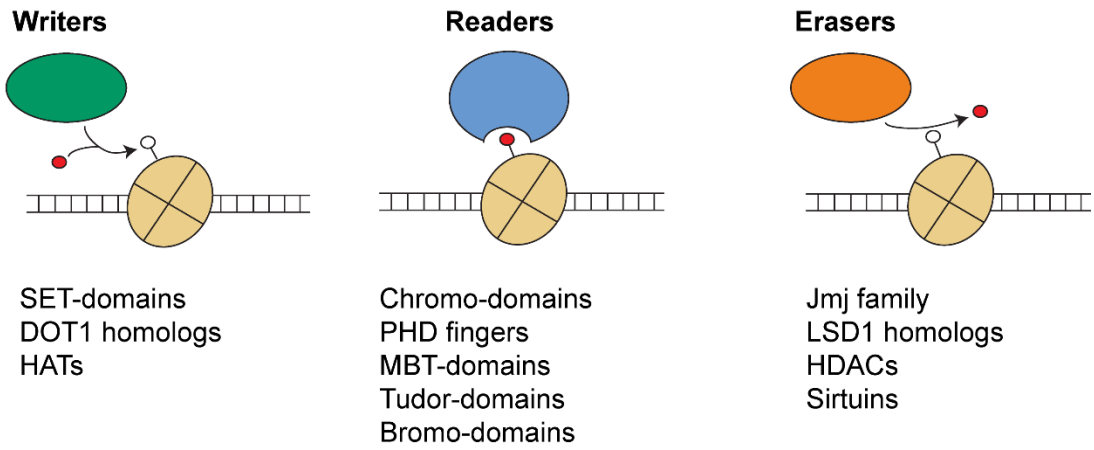


Figure 1.9 Histone writers, readers and erasers. Histone tails protrude from the nucleosome core and can be modified by writer enzymes. The resulting modifications recruit readers which mediate the cellular response. The removal of the histone modifications is catalysed by erasers. Examples of protein classes mediating each activity are included below the schematics.

Histone acetyl transferases (HATs) are a diverse class of enzymes that use acetyl-CoA as a substrate to acetylate lysine residues of the histone tails (Lee and Workman, 2007). The acetyl groups neutralize the positive charges of the histones, thus decreasing the nucleosome affinity for the negatively charged DNA and causing chromatin decompaction (Hong et al., 1993). Bromo-domain proteins bind to acetylated histones and therefore are typically present in active chromatin. Acetyl marks are removed by several classes of histone deacetylases (HDACs) and silent information regulator 2 (Sir2) proteins, or sirtuins, leading to tighter binding between DNA and histones and suppression of gene expression (Haberland et al., 2009; Schwer and Verdin, 2008).

Lysine residues can also be mono-, di- or tri-methylated by lysine methyltransferases (KMTs) which use S-adenosyl methionine as a methyl group donor. Most KMTs contain SET (suppressor of variegation 3-9, enhancer of zeste and trithorax) catalytic domains, with the exception of DOT1 which methylates H3K79. Depending on the residue that is modified, histone methylation can result in the activation or repression of gene expression. For instance, tri-methylation of K4 of histone H3 (H3K4me3) is a hallmark of active transcription (Santos-Rosa et al., 2002). Methylation of H3K36 and H3K79 have also been associated with actively transcribed genomic regions. H3K36me2/me3 residues recruit an HDAC complex which removes acetyl groups from histones and promotes nucleosome reassembly behind RNA Pol II (RNAPII) (Smolle et al., 2013). The mechanism of action of H3K79 methylation is less well understood (Farooq et al., 2016). H3K9, H3K27 and H4K20 methylations are normally found in silent chromatin regions. Di- and tri-methylated H3K9 residues recruit heterochromatin protein-1 (HP1), leading to the formation of constitutive heterochromatin (Allshire and Madhani, 2018) whereas methylation of H3K27 is found in facultative chromatin (Jamieson et al., 2016). Although, in many eukaryotes H3K9 methylation specifies heterochromatin formation, there are known exceptions. For example, budding yeast heterochromatin is formed upon histone deacetylation by sirtuin proteins (Kueng et al., 2013). Chromo-domains, malignant brain tumour (MBT) and plant homeodomain (PHD) fingers recognize methylated

lysines whereas Tudor-domains bind to methylated lysines and arginines (Taverna et al., 2007). Lysine demethylases fall into two categories: Jumonji (Jmj) family and LSD1 homologs (Cloos et al., 2008).

Individual histone PTMs do not act in isolation but rather function in the context of other histone modifications present on the same or on neighbouring nucleosomes. Effector proteins or protein complexes often contain multiple binding domains for the same or different histone marks (Ruthenburg et al., 2007). Significantly, bivalent nucleosomes have been described which contain both active and repressive marks (e.g. H3K4me3 and H3K27me3). These are found mainly at developmentally regulated genes, keeping the latter silent but poised for activation (Bernstein et al., 2006; Fisher and Fisher, 2011; Mikkelsen et al., 2007). All of these factors contribute to the formation of a complex chromatin landscape that, together with post-transcriptional mechanisms, can account for the elaborate gene expression patterns in eukaryotes.

1.2.3 Specifics of trypanosome chromatin

Similarly to other eukaryotes, electron microscopy of nuclei from *T. brucei* has revealed electron-dense “heterochromatic” regions close to the nuclear periphery and electron-lucent “euchromatic” regions in the nuclear interior and in the vicinity of nuclear pores (Daniels et al., 2010; Ogbadoyi et al., 2000). However, it is unclear whether these microscopically distinct regions correspond functionally to transcriptionally active and silent genomic regions. Trypanosomes have a high gene density and most of the megabase chromosomes are considered to be transcriptionally active, with the exception of the silent VSG arrays. On the other hand, the minichromosomes are highly repetitive and contain only silent VSG genes, so are likely to be heterochromatic. VSG ESs can be in an active or repressed state and might therefore represent examples of trypanosome facultative heterochromatin.

Although histones are some of the most well-conserved proteins among eukaryotes, trypanosome histones differ significantly from the common eukaryotic norm (Mandava et al., 2007; Thatcher and Gorovsky, 1994) (Figure 1.10). Notably, *T. brucei* lacks the H3K9 residue and therefore employs a distinct mechanism for heterochromatin formation.

A survey of trypanosome histone modifications using Edman degradation and mass spectrometry revealed an absence of some well-conserved PTMs and a presence of trypanosome-specific marks (Janzen et al., 2006a; Mandava et al., 2007). The C-terminal region of H2A is heavily acetylated in *T. brucei*. The function of these acetyl marks is unclear, but they may contribute to a more open chromatin structure. H2B is the least conserved trypanosome histone and some *T. brucei*-specific marks have been detected on it: A1 methylation and K4, K12 and K16 acetylation. Additionally, the trypanosome H2B lack K120, a conserved lysine residue, ubiquitination of which in many organisms is required for H3K4 and H3K79 methylation (Briggs et al., 2002; Sun and Allis, 2002). A few residues of the N-terminal region of histone H3 are conserved in *T. brucei*, including K4 which, as in other eukaryotes, can be methylated and the H3K4me3 mark has been found at active RNAPII TSRs (Wright et al., 2010). One of the best characterized trypanosome modifications is methylation of H3K76 (the homolog of H3K79) by DOT1A and DOT1B which are responsible for di- and tri-methylation of this residue, respectively (Janzen et al., 2006b). In *T. brucei*, H3K76 methylation is involved in the regulation of the cell cycle and VSG switching (Figueiredo et al., 2008; Janzen et al., 2006b). H4 is the most conserved *T. brucei* histone. Acetylation of H4K10 (homolog of H4K12) was found to be enriched at RNAPII TSRs and is potentially required for transcription initiation (Siegel et al., 2009). A similar set of histone modifications was identified more recently in *Trypanosoma cruzi* (De Jesus et al., 2016).

Apart from the core histones, trypanosomes have four histone variants – H2AZ and H2BV which are found near RNAPII transcription start regions (TSRs), and H3V and H4V which are found at RNAPII transcription termination regions (TTRs) (Siegel et al., 2009).

The trypanosome linker histone H1 lacks the central globular domain thought to mediate nucleosome interactions and instead consists of a single domain corresponding to the C-terminal portion of H1 in other eukaryotes (Povelones et al., 2012). Knock down experiments showed that in *T. brucei* H1 is required for chromatin compaction and VSG silencing (Povelones et al., 2012).

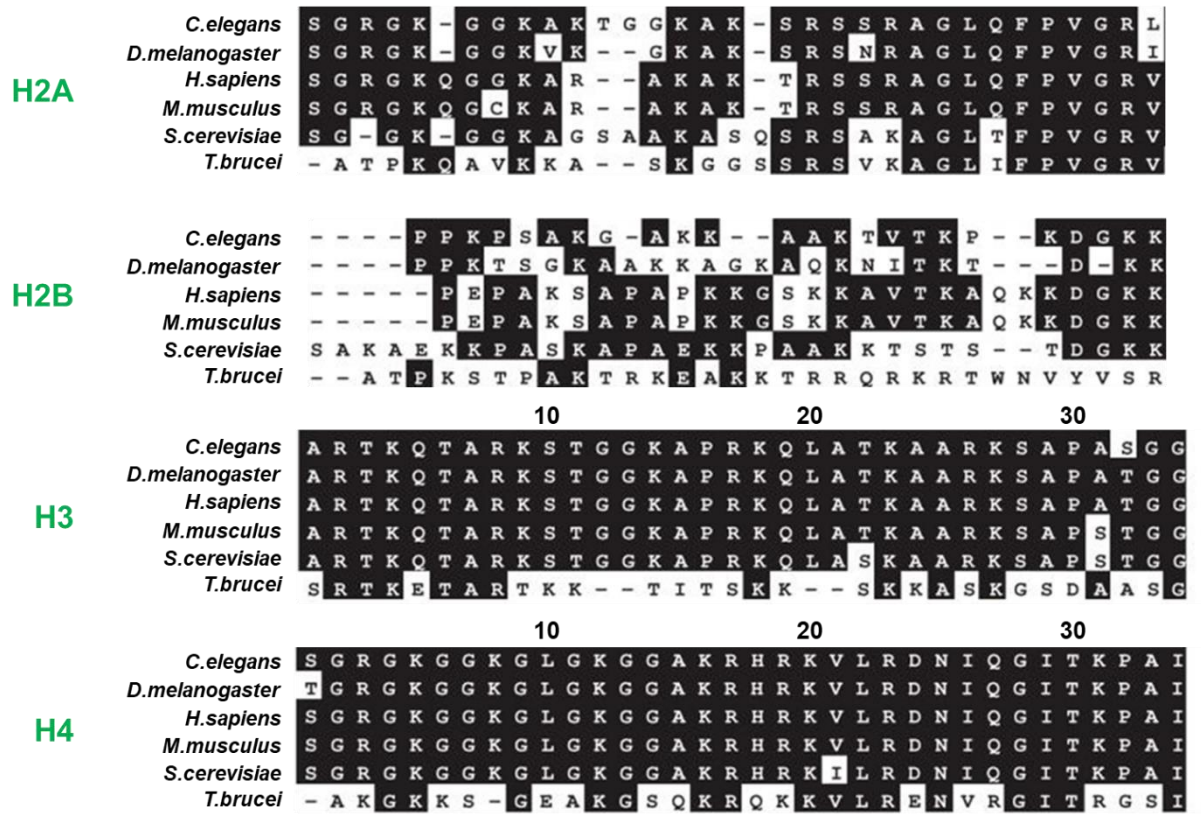


Figure 1.10 Alignment of the N-terminal tails of core histones in *T. brucei* and some Opisthokonts. Identical residues are shaded. The sequence ruler is included only when the sequences are highly concordant and is based on the *H. sapiens* positions. Figure modified from Mandava et al. (2007).

There are two classes of trypanosome histone acetyltransferases: the MYST family and two homologs of the human ELP3 (ELP3a and ELP3b) (Figueiredo et al., 2009; Kawahara et al., 2008; Siegel et al., 2007). The trypanosome genome also encodes class I and class II HDACs as well as SIR2-related deacetylases (Alsford et al., 2007; García-Salcedo et al., 2003; Ingram and Horn, 2002). *T. brucei* has ~30 SET-domain putative lysine methyltransferases as well as two DOT1 homologs (DOT1A and DOT1B). Lysine methyl marks are removed by JmjC demethylases whereas LSD1 homologs have not been found (Figueiredo et al., 2009). Additionally, *T. brucei* has several candidate arginine methyltransferases. Putative readers of histone PTMs in *T. brucei* include proteins containing Bromo-, Chromo-, PHD finger and Tudor domains.

The RNAi machinery of *T. brucei* consists of a single Argonaute protein (Ago1) and two Dicer-like enzymes (DCL1 and DCL2) (Ngô et al., 1998; Patrick et al., 2009; Shi et al., 2006). Most trypanosome siRNAs are produced from two classes of retrotransposons (ingi and SLACS), followed by centromeric CIR147 repeats, long inverted repeats and convergent transcription units (Tschudi et al., 2012). These observations suggest that, similarly to other eukaryotes, RNAi is protecting the trypanosome genome from the spread of transposable elements and is potentially involved in centromeric heterochromatin formation.

T. brucei encodes a putative DNA methyltransferase which was identified based on sequence similarity to a bacterial DNA methyltransferase gene and was found to be expressed in both bloodstream and procyclic parasites (Militello et al., 2008). The same study found modest amount of 5-methylcytosine (5mC) at both developmental stages of *T. brucei*. However, an unpublished bisulfite sequencing data from the Allshire lab did not identify any genomic loci that are highly methylated in the population of BF or PF parasites. This suggested that either there is no DNA methylation or low levels of 5mC are present stochastically across the trypanosome genome.

Bloodstream form trypanosomes have a modified thymidine residue, termed base J, which is bulkier than 5-mC (Gommers-Ampt et al., 1993) and is found at transcription termination sites and telomeric regions (Schulz et al., 2016). The absence of base J in the procyclic form was demonstrated by an anti-J dot-blot test (Van Leeuwen et al., 1997). Base J is found in kinetoplastids and a few other closely related organisms (Borst and Sabatini, 2008). It is synthesized by J binding protein 1 (JBP1), J biosynthesis protein 2 (JBP2) and a ß-glucosyltransferase (Figure 1.11). While both JBP1 and JBP2 are thymidine hydroxylases catalysing the first step of base J production, JBP1 was found to be involved in J maintenance (Cross et al., 2002) and JBP2 – in de novo J synthesis (DiPaolo et al., 2005; Kieft et al., 2007). It was previously thought that base J could be involved in gene silencing because of its association with silent telomeric VSG genes. This hypothesis was disproven upon the observation that JBP1 single and JBP1/JBP2 double knockouts resulted in a marked decrease or absence of base J, respectively, but that did not lead to cell lethality or derepression of silent VSG genes (Borst and Sabatini, 2008). A second idea, based on analogy with 5mC function, was that base J represses homologous recombination between repetitive sequences but this hypothesis was also rejected (Cross et al., 2002). Thus, at present, the precise function of base J is unclear.

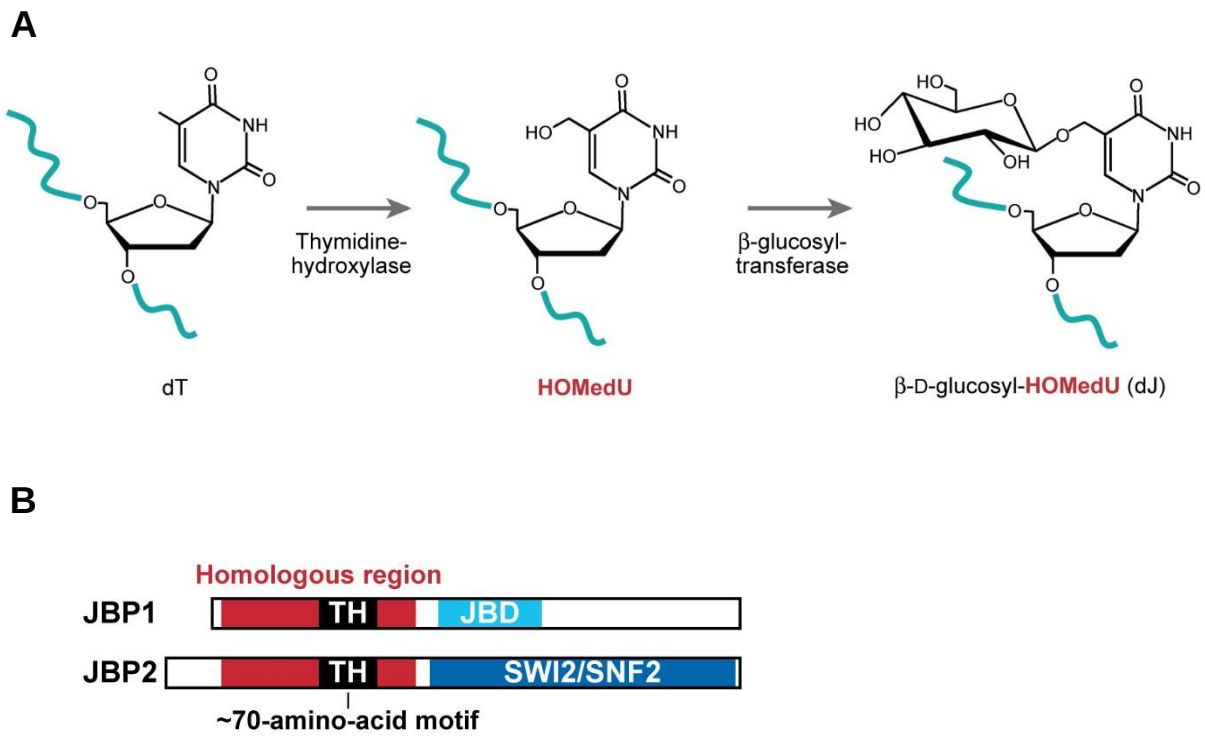


Figure 1.11 Base J synthesis.

(A) Synthesis mechanism. The first step of base J synthesis is oxidation of a thymidine residue to hydroxymethyldeoxyuridine (HOMedU). In the second step, HOMedU is glucosylated to form base J.

(B) Domain architecture of JBP1 and JBP2. The shared homologous region of JBP1 and JBP2 is shaded in red. TH – thymidine hydroxylase catalytic domain; JBD – J-binding domain; SWI2/SNF2 – family of ATPase/DNA helicases.

Figure from Borst and Sabatini (2008).

1.3 Trypanosome gene expression

Most genes in *T. brucei* are transcribed from polycistronic transcription units (PTUs) which, unlike bacterial operons, often contain functionally unrelated genes. mRNAs in a polycistron are resolved by trans-splicing of an m^7G-capped 39-nt spliced leader (SL) sequence at the 5′ end of the mRNA and polyadenylation at the 3′ end (Günzl, 2010).

1.3.1 RNAPII transcription

Trypanosome RNAPII transcription start regions (TSRs) are not well characterized with the exception of the SL TSR which is defined by a loose consensus motif (Gilinger, 2001). Due to the high demand for SL RNA, *T. brucei* harbours ~200 tandem copies of the SL gene which are one of the few trypanosome genes transcribed monocistronically. Within a polycistronic RNAPII PTU, genes are arranged in a head-to-tail fashion and are transcribed from the same DNA strand. Transcription usually initiates from broad (~10 kb) GT-rich divergent strand-switch regions (SSRs) and terminates at convergent strand-switch regions (Daniels et al., 2010; Wedel et al., 2017) (Figure 1.12). TSRs are marked by the presence of H2AZ, H2BV, H3K4me3, H4K10ac and TTRs – by H3V, H4V and base J (Schulz et al., 2016; Siegel et al., 2009; Wright et al., 2010). RNAPII transcription can also terminate at non-SSRs upon encounter of RNAPI or RNAPIII transcribed genes (e.g. rRNAs and tRNAs) (Marchetti, 1998; Siegel et al., 2009).

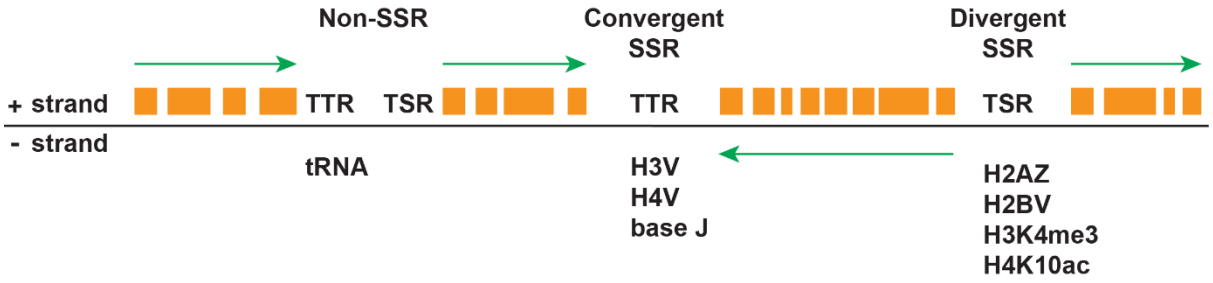


Figure 1.12 RNAPII PTU organization. Divergent SSRs are probable transcription start regions and convergent SSRs – probable termination regions. The figure shows four PTUs with individual genes depicted as orange boxes. Transcription direction within each PTU is marked by green arrows. Specific marks of TSRs and TTRs are listed below these regions.

The trypanosome RNAPII has 12 subunits which are homologous to those found in other eukaryotes (Das et al., 2008). The basal transcription factors include conserved and divergent proteins: TBP(TRF4), TFIIA, TFIIB, TFIIH, Mediator complex, a PAF complex and two proteins specific to kinetoplastids (Clayton, 2014; Lecordier et al., 2007; Lee et al., 2010; Ouna et al., 2012; Ruan et al., 2004; Schimanski et al., 2005; Srivastava et al., 2017). The C-terminal domain (CTD) of the largest RNAPII subunit in *T. brucei* is phosphorylated and is required for mRNA production, similarly to other eukaryotes (Das et al., 2017). However, the trypanosome RNAPII CTD lacks the classic heptad repeats and the significance of the phosphorylated residues for the regulation of mRNA synthesis is unclear.

1.3.2 RNAPI transcription

In *T. brucei*, RNAPI transcribes not only rRNA genes but also VSG and procyclin mRNAs (Gunzl et al., 2003). The reason for this could be the high demand for producing surface proteins as RNAPI transcription can give ~10 times greater mRNA output than RNAPII transcription (Clayton, 2014).

Procyclin genes are transcribed from PTUs found in the core of megabase chromosomes 6 and 10 (Haenni et al., 2006). In contrast, VSGs are transcribed from expression sites found in the subtelomeres of megabase and intermediate chromosomes. The single active VSG in the bloodstream form is transcribed from a BES which localises to a special subnuclear compartment – the expression site body (ESB) (Navarro and Gull, 2001). ESB is a second, non-nucleolar focus of RNAPI in the trypanosome nucleus. BESs are large polycistronic transcription units (~50 kb) that include several expression site-associated genes (ESAGs) and a single telomere-proximal VSG gene (Figure 1.13). The function of most ESAGs is unknown, with the exception of ESAG4 which encodes an adenylate cyclase (Salmon et al., 2012), ESAG6 and ESAG7 which together encode a transferrin uptake receptor (Steverding, 1995) and ESAG10 which codes for a folate receptor (Dewar et al., 2016). Between the ESAGs and the VSG gene there is a stretch of AT-rich 70 bp repeats.

Figure 1.13 BES organisation. A BES is a subtelomeric polycistronic unit consisting of ESAGs, 70 bp repeats and a telomere-proximal VSG gene. The active BES is transcribed by RNAPI.

VSG switching can occur via two mechanisms. During an *in situ* (transcriptional) switch, one ES is silenced while another ES is activated (Van der Ploeg et al., 1984). More commonly, switching occurs by recombination of a new VSG into the active BES (Robinson et al., 1999). Recombinational switches are facilitated by homology of regions within the VSG gene or proximal to it (such as the 70 bp repeats) as well as the inherent instability of the subtelomeric regions which are hot spots for recombination (Horn, 2014). Additionally, fragments of several VSG genes can recombine together into the active BES to form a mosaic VSG, further expanding the repertoire of trypanosome surface proteins used to evade the host immune system.

The order in which VSGs are expressed during infection is semi-predictable, as demonstrated by the preferential expression of certain VSGs at the initial peaks of parasitaemia and throughout the later stages of a chronic infection (Morrison et al., 2005). During the course of an infection, *in situ* switches are initially favoured followed by recombinational switches, with generation of mosaic VSGs dominating the late stages of trypanosomiasis. Additionally, parasites with shorter VSGs were found to dominate early in infection (Liu et al., 2018). This could be explained by the lower metabolic costs required to produce shorter VSGs leading to faster trypanosome growth rates.

1.3.3 Post-transcriptional gene regulation

It is thought that *T. brucei* regulates its gene expression predominantly post-transcriptionally via control of RNA stability and translation. The reason for this is two-fold: firstly, most of the trypanosome genes are transcribed from polycistronic units which contain functionally unrelated genes whose expression needs to be regulated separately. A notable exception are the VSG genes whose expression is regulated transcriptionally. Secondly, RNA binding proteins (RBPs) play an important role in trypanosome post-transcriptional gene regulation via binding to AU-rich elements (AREs) present in the 3′ UTRs of target RNAs (Clayton, 2002; Mayho et al., 2006). For example, it has been shown that RBP6 is sufficient for the procyclic

to metacyclic differentiation (Kolev et al., 2012), and RBP10 – for promoting the procyclic to bloodstream transition (Mugo and Clayton, 2017). However, it is unclear how the expression of these RBPs is itself regulated. Additionally, it has recently been demonstrated that depletion of retrotransposon hot spot (RHS) genes reduces nascent RNAPII transcription and leads to accumulation of mRNAs in the nucleus, suggesting that RHS genes are involved in the regulation of RNAPII transcription elongation and mRNA export (Florini et al., 2018). These findings indicate that there is a transcriptional component to trypanosome gene regulation.

1.4 Project aims

The aim of this project was to assess what role histone modification readers, writers and erasers play in trypanosome chromatin organisation and transcription regulation. To address this question, the following approaches were used:

(i) Identification, tagging and localisation of putative readers, writers and erasers of histone modifications in *T. brucei*

(ii) Chromatin immunoprecipitation and sequencing (ChIP-seq) of the nuclear candidate proteins to identify associated genomic regions

(iii) Co-immunoprecipitations followed by mass spectrometry to characterise candidate protein interaction networks

CHAPTER 2

Materials and methods

2.1 Trypanosome methods

2.1.1 Trypanosome cell culture

Lister 427 bloodstream and procyclic cell lines were used throughout this project. The potential history of these cell lines was recorded by George Cross: http://tryps.rockefeller.edu/trypsru2_pedigrees.html.

Bloodstream form monomorphic cell lines were grown at 37^{o}C and 5% CO_2 in HMI-9 medium (Appendix A) supplemented with 10% Fetal Calf Serum (Gibco), 1% Penicillin-Streptomycin (Gibco) and selective drug(s) (when appropriate) (Hirumi and Hirumi, 1989). Cells cultures were maintained below 3 x 10^6 cells/ml.

Procyclic form cell lines were grown at 27^{o}C in SDM-79 medium (Appendix A) supplemented with 10% Fetal Calf Serum (Gibco), 1% Penicillin-Streptomycin (Gibco), 0.1% hemin and selective drugs (when appropriate) (Brun, 1979). Cell cultures were maintained between 5 x 10^5 and 3 x 10^7 cells/ml.

Glycerol stocks were made by mixing 500 µl trypanosome cell culture with 500 µl of the appropriate medium containing 28% glycerol (final glycerol concentration was 14%). Cryovials were incubated overnight at -80^{o}C in a polystyrene container (to prevent ice crystal formation) before being moved to a box in a -80^{o}C freezer or in liquid nitrogen for long-term storage.

Glycerol stocks of bloodstream cells were thawed by adding the contents of the cryovial to 5 ml HMI-9 medium without selective drugs and centrifuging at 1000 g for 10 min to remove

glycerol. The cell pellet was resuspended in 5 ml HMI-9 medium and incubated overnight at 37°C and 5% CO_2 before adding selective drug(s) (if appropriate).

Glycerol stocks of procyclic cells were thawed by adding the contents of the cryovial to 5 ml SDM-79 medium without removing glycerol and incubating overnight at 27°C before adding selective drug(s) (if appropriate).

Table 1 Cell lines with drug resistances used in this study

Cell lines	Selective drug	Concentration
BF 427 cell lines with YFP-tagged proteins	blasticidin	5 µg/ml
PF 427 cell lines with YFP-tagged proteins	blasticidin	100 µg/ml

2.1.2 Immunofluorescence

1 x 10^6 parasites were harvested by centrifugation at 4000 g for 5 min. The cells were washed once with 500 µl cold 1X PBS and centrifuged again at 4000 g for 5 min. The cells were fixed by resuspending the pellets in 125 µl cold 1X PBS + 125 µl 8% paraformaldehyde and incubating on ice for 10 min. The fixed samples were centrifuged at 4000 g for 5 min and the cells were resuspended in 130 µl 0.1 M glycine (diluted in 1X PBS). The cells were incubated overnight at 4°C, centrifuged at 4000 g for 5 min and resuspended in 200 µl 1X PBS. Wells were drawn on polylysine slides using a hydrophobic pen. 10 µl cells were added to each well and the slides were incubated for 1 h at room temperature in a humid chamber to let the parasites stick to the wells. The excess PBS was removed using a vacuum pump and 0.1% Triton X-100 (diluted in 1X PBS) was added to the wells for 2 min in order to permeabilise the cells. The slides were washed with 1X PBS for 5 min on a lab rocker. The samples were

blocked in 2% BSA (diluted in 1X PBS) for 45 min at 37ºC inside a humid chamber. Excess blocking solution was removed with a vacuum pump and the cells were incubated with a 1:500 dilution (in 2% BSA) of the primary antibody, rabbit anti-GFP (Thermo Fisher Scientific), for 45 min at 37ºC inside a humid chamber. The slides were then washed 3 x 5 min with 1X PBS on a lab rocker. The cells were incubated with 1:1000 dilution (in 2% BSA) of the secondary antibody, Alexa Fluor 568 goat anti-rabbit (Thermo Fisher Scientific) for 45 min at 37ºC inside a humid chamber. From here on, the slides were kept in the dark as much as possible to prevent photobleaching of the secondary antibody signal. The slides were washed 2 x 5 min with 1X PBS on a lab rocker, incubated with 1µg/ml DAPI for 2 min and washed again 2 x 5 min with 1X PBS. Coverslips were mounted onto the slides by adding a few drops of mowiol + DABCO (Appendix A) and sealing with nail polish. The slides were left to dry overnight at room temperature in the dark before visualisation under the microscope (Zeiss Axio Imager).

2.1.3 Trypanosome transfections

Approximately 5 x 10^7 bloodstream or procyclic cells were harvested per transfection by centrifugation at 1000 g for 10 min. The cells were washed once with 5 ml media (HMI-9 or SDM-79) and pelleted again by centrifugation at 1000 g for 10 min before resuspending them in 100 µl ice-cold transfection buffer, Amaxa or homemade (Appendix A), and transferring them to an electroporation cuvette (Ingenio). 10-20 µl containing 100 ng - 5 µg DNA were then added to the cuvette. The cells were mixed with the DNA by flicking and electroporated in Amaxa Nucleofector II (Lonza) using the X-001 programme for bloodstream cells and the Z-001 programme for procyclic cells. A “no DNA” mock transfection was performed in parallel as a control.

Electroporated bloodstream cells were added to 30 ml HMI-9 medium and two 10-fold serial dilutions were performed. 1 ml of transfected cells were plated per well on 24-well plates (1 plate per dilution) and incubated at 37ºC and 5% CO_2 for 6 h before adding 1 ml media

containing 2X concentration of the selective drug per well. The mock-transfected cells were dead by day 5 post-transfection and then genuine transfectants were moved to flasks.

Electroporated procyclic cells were added to 6 ml SDM-79 medium and incubated overnight at 27ºC before being mixed with 6 ml media containing 2X concentration of the selective drug. Two 24-well plates per transfection were set up as follows:

Plate 1

row A – 2 ml cell culture (+1X final drug concentration)

row B – 1.5 ml SDM-79 + 1X drug

row C – 1.5 ml SDM-79 + 1X drug

row D – 1.0 ml SDM-79 + 1X drug

Plate 2

row A – 1.0 ml SDM-79 + 1X drug

row B – 0.5 ml SDM-79 + 1X drug

Serial dilutions were performed by transferring 500 µl from well A1 (row A, column 1) on plate 1 to well B1 (row B, column 1) on plate 1. The dilutions were continued in the same fashion until row B on plate 2. The mock-transfected cells were dead by day 7-9 post-transfection and then the genuine transfectants were moved to flasks.

2.2 Standard DNA methods

2.2.1 Ethanol precipitation

Various DNA samples required cleaning of impurities and/or concentration. This was done via ethanol precipitation. 0.1 V of 3 M sodium acetate pH 5.2 and 2.5 V of ice-cold 100% ethanol were added to the DNA. The samples were precipitated at -80ºC for 1 h or at -20ºC overnight. The samples were then centrifuged at 16 000 g for 15 min at 4ºC. The supernatant was discarded, and the DNA was washed once with ice-cold 70% ethanol. The samples were centrifuged again at 16 000 g for 15 min at 4ºC, the supernatant was discarded, and the DNA was air-dried for 10 min at room temperature. Finally, the DNA was resuspended in the desired volume of distilled water and the resulting concentration was measured by NanoDrop (Thermo Fisher Scientific).

2.2.2 Genomic DNA extraction

Approximately 4 x 10^7 cells from bloodstream or 1 x 10^8 cells from procyclic cultures were harvested by centrifugation at 1000 g for 10 min. The cells were then washed once with 500 µl 1X PBS and pelleted again by centrifugation at 1000 g for 10 min. Genomic DNA was extracted using DNeasy Blood & Tissue kit (Qiagen) following the manufacturer's instructions. The DNA samples were ethanol precipitated and the amount of extracted DNA was quantified using NanoDrop (Thermo Fisher Scientific).

2.2.3 Polymerase chain reaction (PCR)

PCR amplifications of DNA were carried out in 0.2 ml PCR tubes (STARLAB) in an MJ Research PTC-225 Gradient Thermal Cycler (Marshall Scientific). 10-200 ng DNA was used as template. All primers (Appendix B) were purchased from Integrated DNA Technologies. PCR amplifications aiming to confirm the presence of a DNA fragment in a sample were

carried out with Taq DNA Polymerase (Roche) whereas PCRs for cloning purposes were done using high-fidelity DNA polymerases - Phusion High-Fidelity DNA Polymerase (New England Biolabs) or Platinum Pfx DNA Polymerase (Thermo Fisher Scientific).

2.2.4 DNA agarose gel electrophoresis

DNA fragments in a sample were separated by size via agarose gel electrophoresis. Usually, 1% agarose was used except for resolving very large (> 2kb) or very small (< 500 bp) fragments for which 0.8% and 2% agarose solutions were used, respectively. The agarose powder was dissolved by boiling in 1X TBE (Appendix A) buffer except for DNA gel extractions when 1X TAE (Appendix A) buffer was used. The solution was cooled down to approximately 60^{o}C before adding 0.3 µg/ml ethidium bromide (Sigma-Aldrich). The gel was poured into a cast containing a gel comb and left to solidify, then transferred to a gel tank filled with 1X TBE or 1X TAE buffer, as appropriate. DNA samples were resuspended in loading dye (NEB) and loaded in the gel wells alongside a DNA ladder that allowed estimation of the fragment sizes. The gels were run at 100-120 V for time sufficient to resolve the desired fragments.

2.2.5 DNA gel extraction

When a PCR reaction resulted in amplification of multiple fragments in addition to the desired one, the correct amplicon was purified from the agarose gel using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) following the manufacturer's instructions. The eluted DNA was ethanol precipitated and the resulting sample concentration was measured by NanoDrop (Thermo Fisher Scientific).

2.2.6 Plasmid restriction digests

Plasmids were digested with restriction enzymes from NEB following the manufacturer's instructions. Usually, a single or a double digest was performed in CutSmart Buffer (NEB) at 37°C for 1 h followed by enzyme inactivation at 80°C for 20 min.

2.2.7 Bacterial transformations

DH5α (NEB) or XL10-Gold (Agilent technologies) competent *E. coli* cells were stored at -80°C and thawed on ice immediately before transformation. Approximately 1-5 µl of a plasmid ligation reaction or 0.5-1 µl of a plasmid preparation were added to 30-50 µl competent cells and the mixture was incubated on ice for 30 min. The cells were then heat-shocked at 42°C for 40 s and put on ice for 5 min. 950 µl of LB medium (Appendix A) was added to the cells which were then incubated at 37°C for 1 h with shaking to allow them to recover. Finally, the cells were plated onto LB selective plates containing 100µg/ml ampicillin (all plasmids used in this study had an ampicillin resistance gene) and incubated at 37°C overnight. The following morning the plates were taken out of the incubator and the success of the transformations was assessed by the presence of colonies on the plates.

2.2.8 Plasmid minipreps

Single colonies from bacterial transformation plates were grown in 5 ml LB at 37°C overnight with shaking. The plasmid isolation was done using either a QIAprep Spin Miniprep Kit (QIAGEN) or a Monarch Plasmid Miniprep Kit (NEB) following the manufacturers' instructions. The amount of DNA in the plasmid preparations was quantified by NanoDrop Spectrophotometer (Thermo Fisher Scientific).

2.3 Standard protein methods

2.3.1 Protein extraction

4 x 10^7 bloodstream or procyclic cells were harvested by centrifugation at 1000 g for 10 min. The cells were washed once with 500 µl 1X PBS and resuspended in 75 µl 1X PBS + 25 µl 4X NuPAGE LDS Sample Buffer (Thermo Fisher Scientific). The samples were then boiled at 95^{o}C for 5 min.

2.3.2 Western blot analysis

Whole cell protein extracts were separated by polyacrylamide gel electrophoresis (PAGE). Protein from 4 x 10^6 cells was loaded per lane in NuPAGE Bis-Tris Mini Gels (Thermo Fisher Scientific). Samples were run for 30 – 60 min in a Mini Gel Tank (Thermo Fisher Scientific) in 1X NuPAGE MES Running Buffer or 1X NuPAGE MOPS Running Buffer (Thermo Fisher Scientific) depending on the protein sizes that had to be resolved. For separation of smaller proteins, shorter run times and MES buffer was used. Conversely, larger proteins were separated with longer run times and MOPS buffer. Following PAGE, proteins were transferred onto nitrocellulose membranes in a Mini Blot Module (Thermo Fisher Scientific) at 20 V for 1 h. The success of the transfer was assessed by staining the membranes with Ponceau S (Sigma-Aldrich). Membranes were then blocked in 5% milk powder in PBS-T (PBS + 0.05% tween) at room temperature for 30 min. After blocking, membranes were incubated with a 1:1000 dilution (in 5% milk in PBS-T) of the primary antibody, mouse anti-GFP (Roche), at 4^{o}C overnight on a lab rocker. The membranes were washed 3 x 10 min with PBS-T on a lab shaker and incubated with a 1:2500 dilution (in 5% milk in PBS-T) of the secondary antibody, HRP-conjugated anti-mouse, at room temperature for 1 h on a lab shaker. The membranes were washed again 3 x 10 min with PBS-T. Finally, proteins were detected with Amersham

ECL Prime Western Blotting Detection Reagent (GE Healthcare) following manufacturer's instructions and visualised using Amersham Hyperfilm ECL (GE Healthcare).

2.4 Chromatin immunoprecipitation and sequencing (ChIP-seq)

2.4.1 ChIP

4 x 10^8 cells were harvested per ChIP sample by centrifugation at 1500 g for 10 min. The cell pellets were resuspended in 40 ml HMI-9 or SDM-79 media, as appropriate, and then 4 ml formaldehyde solution (Appendix A) was added to each sample. The cells were fixed with 0.8% final formaldehyde concentration for 20 min at room temperature. The fixation was stopped by the addition of 2.5 ml 2 M glycine, followed by centrifugation at 1500 g for 20 min at 4ºC. The resulting cell pellets were washed once with 10 ml cold 1X PBS and centrifuged at 3500 g for 20 min at 4ºC. The next steps involving Lysis buffers 1 and 2 were performed without resuspending the pellets by pipetting to avoid cells sticking to the plastic tips. Vigorous shaking of the samples was used instead. The pellets were first resuspended in 2.5 ml Lysis buffer 1 (Appendix A), supplemented just before use with EDTA-free protease inhibitor tablets (Roche). The samples were incubated at 4ºC for 10 min on a lab rocker and then centrifuged at 3500 g for 20 min at 4ºC. The resulting pellets were resuspended in 2.5 ml Lysis buffer 2 (Appendix A), supplemented just before use with EDTA-free protease inhibitor tablets (Roche) and incubated at room temperature for 10 min on a lab rocker. The samples were then centrifuged at 3500 g for 20 min at 4ºC. The pellets were resuspended in 300 µl Lysis buffer 3 (Appendix A), supplemented just before use with SDS (0.2% final concentration) and EDTA-free protease inhibitor tablets (Roche), and transferred to microfuge tubes. The chromatin was fragmented by sonication at a high setting in a Bioruptor (Diagenode) sonicator at 5°C for 30 cycles (30 s ON / 30 s OFF per cycle). After sonication, 900 µl Lysis buffer 3, supplemented with protease inhibitors but not SDS, was added to each sample. The samples were then centrifuged at 16000 g for 10 min at 4°C and the resulting supernatants were transferred to

new microfuge tubes. At this stage, 10 µl ‘’input’’ was taken from each sample and frozen overnight at -80°C. The rest of the sample, the ‘’IP’’, was incubated rotating overnight at 4°C with 10 µg rabbit anti-GFP antibody (Thermo Fisher Scientific) and 50 µl Protein G Dynabeads (washed beforehand 2x with Lysis buffer 3). Thus, the antibody/bead crosslinking and the IP steps were combined together.

The following day, the beads were collected using a magnet and the supernatant was discarded. The beads were first washed with 900 µl Lysis buffer 3 on a magnet. The beads were then washed with 900 µl Wash buffer 1 (Appendix A) rotating for 10 min at 4°C, followed by a wash with 900 µl Wash buffer 2 (Appendix A) rotating for 10 min at 4°C. The last wash was with 900 µl TE buffer (Appendix A) on a magnet and during that wash the beads were transferred to a new microfuge tube. This was done because DNA sticks to the plastic of microfuge tubes and consequently the washes cannot efficiently remove unspecific DNA from the sample. 200 µl Elution buffer (Appendix A) was added to the washed beads (IPs) as well as the inputs and the samples were then incubated overnight at 65°C with shaking at 1000 rpm, combining together the elution and reverse crosslinking steps.

The next day, the beads were separated from the eluate on a magnet and discarded. 10 µl DNAse-free RNAse (Roche) was added per eluate which was then incubated at 37°C for 2 h shaking at 1000 rpm. Then, 8 µl Proteinase K (10 mg/ml) was added per sample which was subsequently incubated at 55°C for 2 h shaking at 1000 rpm. Finally, the DNA was purified using a QIAquick PCR Purification Kit (Qiagen) and quantified by a Qubit fluorometer using Qubit dsDNA HS Assay Kit (Invitrogen).

2.4.2 Library preparation

Library preparation of ChIP DNA was done in Eppendorf RNA/DNA LoBind Microfuge Tubes (Sigma-Aldrich) except the PCR step. 0.5-20 ng DNA was blunt ended at room temperature for 45 min using a Quick Blunting Kit from NEB. DNA was then purified using AMPure beads

(Beckman Coulter) with a bead:DNA sample ratio of 1.6:1. DNA was A-tailed using Klenow Fragment (3'→5' exo-) from NEB for 30 min at 37°C. Samples were incubated at 75°C for 5 min to heat-inactivate the enzyme and then placed on ice for 5 min. NEXTflex (Bioo Scientific) adapters containing barcodes were ligated immediately after A-tailing for 25 min at room temperature using a Quick Ligation Kit from NEB. Samples were purified 2x using a 1:1 ratio with AMPure beads. DNA was PCR amplified (16-18 cycles) and subsequently subjected to size selection with AMPure beads in several rounds. First, a “right-hand” purification was done with a bead:DNA sample ratio of 0.65:1 and the supernatant resulting after binding was taken for a 0.7:1 bead:sample AMPure purification. Finally, two 1:1 AMPure purifications were performed. The resulting library was quantified by Qubit fluorometer using Qubit dsDNA HS Assay Kit (Invitrogen). Fragment size distribution was assessed by 2100 Bioanalyzer Instrument (Agilent Technologies) using a Bioanalyzer High Sensitivity DNA Analysis kit (Agilent). Ideally, the libraries had a peak at 300 bp (~ 150 bp fragment DNA + ~ 150 bp adapter sequence). Libraries were sequenced on Illumina MiniSeq (Allshire lab), HiSeq 4000 (Edinburgh Genomics) or NextSeq (Western General Hospital, Edinburgh). In all cases, 75 bp paired-end sequencing was performed.

2.4.3 ChIP-seq data analysis

FASTQ files (Cock et al., 2010) were processed to bigWig format (Kent et al., 2010) for visualisation in a genome browser as follows. The sequencing reads were first subjected to quality control using FastQC (Andrews, 2018). Duplicates were then removed via pyFastqDuplicateRemover (Webb et al., 2018) and the reads were subsequently aligned to the Tb427v9.2 genome (Müller et al., 2018) using Bowtie2 (Langmead and Salzberg, 2012). The resulting SAM files (Li et al., 2009) were converted to BAM format (Li et al., 2009), then sorted and indexed. IPs were normalised to the respective inputs and to the library size via bamCompare (Ramírez et al., 2016). The output of bamCompare were bigWig files (Kent et

al., 2010) which were visualised using Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

Metagene plots were generated as follows. Chromo1 peaks were called by MACS2 (Feng et al., 2012). 20 kb regions around Chromo1 peak summits were divided into 50 bp windows. For each protein, the reads in each window were counted using pyReadCounters (Webb et al., 2018). The reads around all Chromo1 peaks were added up, normalised to library size and input and represented as a density plot, centred around the Chromo1 peak summits. The average metagene plots were generated analogously, except that the reads around Chromo1 peaks were averaged before plotting.

Heatmaps were generated as follows. Chromo1 peaks were called by MACS2 (Feng et al., 2012). 20 kb regions around Chromo1 peak summits were divided into 50 bp windows. For each protein, the reads in each window were counted using pyReadCounters (Webb et al., 2018). The reads around individual Chromo1 peaks were normalised to library size and input and represented as a density plot, centred around the Chromo1 peak summits. Each heatmap was generated as an average of all replicates for the particular protein.

2.5 Protein immunoprecipitation (IP) and mass spectrometry analysis

2.5.1 Crosslinking of antibody to magnetic beads

All crosslinking steps were performed at room temperature. 1 ml Protein G Dynabeads (Thermo Fisher Scientific) were washed 2 x 1 ml Lysis buffer MS (Appendix A). 1 vial of mouse anti-GFP antibody (Roche) was resuspended in 1ml Lysis buffer MS and added to the beads. The beads were rotated for 30 min. The beads were washes 2 x 1 ml Lysis buffer MS on magnet and added to 10 ml Borate buffer (Appendix A) supplemented with 50 mg DMP (Thermo Fisher Scientific). Only DMP opened within one month of the experiment was used. The beads were rotated for 30 min, washed once with 10 ml Lysis buffer MS and resuspended

in 2 ml Lysis buffer MS. The antibody-crosslinked beads could then be stored at 4°C for several months.

2.5.2 Protein IP

4 x 10^8 cells were harvested per IP by centrifugation at 1500 g for 10 min. The cells were washed once with cold PBS and centrifuged at 1500 g for 10 min. The resulting pellets were resuspended in 200 µl Lysis buffer MS, supplemented just before use with EDTA-free protease inhibitor tablets (Roche), and incubated on ice for 15 min. The samples were sonicated at a high setting in a Bioruptor (Diagenode) sonicator at 5°C for 3 cycles (12 s ON / 12 s OFF per cycle). The chromatin and soluble fractions were separated by centrifugation at 16000 g for 5 min at 4°C. The soluble fraction was taken for the IP. 120 µl of antibody-crosslinked beads were added per sample and the IPs were rotated for 1 h at 4°C. The samples were washed 3 x 1 ml Lysis buffer MS on magnet and the protein was eluted with 40 µl RapiGest surfactant (Waters) at 55°C for 15 min. Only RapiGest resuspended within one week of the experiment in 50 mM Tris-HCl pH 8.0 was used.

2.5.3 FASP digestion

DTT was added to each sample to a final concentration of 25 mM. The samples were incubated at 95°C for 5 min and then let stay at room temperature until cooled. The samples were mixed with 100 µl 8 M Urea in 0.1 M Tris-HCl pH 8.0 and transferred to Vivakon spin (filter) column 30K cartridge. The columns were centrifuged at 12000 g for 10-15 min. 100 µl 0.05 M IAA (in 8 M Urea in 0.1 M Tris-HCl pH 8.0) was added to each column. The samples were shaken at 600 rpm for 1 min at room temperature and then incubated in the dark for 20 min at room temperature. The columns were centrifuged at 12000 g for 10-15 min and then 100 µl 8 M Urea in 0.1 M Tris-HCl pH 8.0 was added to each column. The columns were centrifuged at 12000 g for 10-15 min and then washed 2 x 100 µl ABC buffer (Appendix A). Only fresh ABC buffer made within one week of the experiment was used. The ABC buffer

was prepared with MS grade water and the pH was assessed with pH indicator paper. Each sample was digested with 0.5 µg MS Grade Pierce Trypsin Protease (Thermo Fisher Scientific) in 100 µl ABC buffer overnight at 37°C.

2.5.4 Stage tip preparation

After overnight trypsin digestion, the columns were centrifuged at 12000 g for 15 min, then washed with 100 µl ABC buffer and centrifuged again at 12000 g for 15 min. The samples were acidified to pH 2-3 (assessed with pH indicator paper) with ~ 10 µl 10% TFA.

Stage tips (Rappsilber et al., 2007) were prepared as follows. Three C-18 paper discs were cut out and placed in a tip. The stage tips were conditioned by sequential passing though centrifugation at 1000 g of 50 µl methanol, 50 µl 80% ACN in 0.1% TFA and 50 µl 0.1% TFA, in this order. The protein samples were then loaded onto the tips at 800 g for 30 min. The stage tips were washed with 75 µl 0.1% TFA and frozen at -20°C. Dr Tania Auchynnikava eluted the protein from the stage tips and run the mass spectrometry analysis.

CHAPTER 3

Tagging and localisation of putative readers, writers and erasers of histone post-translational modifications in *T. brucei*

3.1 Introduction

The interplay between reading, writing and erasing of specific histone post-translational modifications facilitates, among other things, the organisation of chromatin into active and silent domains which may be constitutive or change depending on the cell type, developmental stage and environmental stimuli sensed by the organism. These chromatin regulation activities have been attributed to particular protein domains. For example, a histone methyl mark could be deposited by a SET-domain protein, read by a Chromo-domain protein and removed by a Jmj-domain protein.

Proteins containing domains associated with reading, writing and erasing of histone PTMs can be found in the trypanosome genome but their functions are largely unknown. Together with Dr Roberta Carloni, we initiated characterisation of these proteins by endogenous ORF tagging followed by examination of their cellular localisation. The goal was to narrow down the list of candidate proteins to those present in the nucleus which may be involved in trypanosome chromatin organisation and gene expression regulation. We decided to focus on proteins likely to be involved in depositing, binding or removing acetyl and methyl marks as the functions of these histone PTMs are well characterised in other eukaryotes.

3.2 Selection and tagging of the candidate proteins

We identified 71 putative readers, writers and erasers of histone acetyl and methyl marks based on their predicted protein domains (Table 3.1). Of these, 68 proteins were successfully tagged with YFP and localised microscopically in bloodstream form parasites: 16 were found to be nuclear, 29 were exclusively cytoplasmic, and 23 localised both in the nucleus and in the cytoplasm.

Ago1, the only Argonaute protein in *T. brucei*, was also tagged in addition to the putative chromatin modifiers because of the link between RNAi and heterochromatin formation in other eukaryotes (Allshire and Madhani, 2018).

When available, existing protein names from the TriTryp database (www.tritrypdb.org) were used. Proteins annotated as “hypothetical” in TriTrypDB were given names in this project. To avoid ambiguity, the unique gene codes for the candidate proteins discussed in this thesis can be found in Tables 3.2 and 6.1.

N-terminal tagging was chosen to avoid interference with the 3′ UTRs which contain sequences known to control mRNA stability (Clayton, 2019). The pPOTv4 plasmid (Dean et al., 2015) used for tagging contained, in this order, actin (act) 5′ UTR, blasticidin S deaminase (bsr), aldolase (ald) 3′ UTR, actin (act) 5′ UTR, glycine-serine linker (GS10), enhanced YFP tag (eYFP), glycine-serine linker (GS10) and PFR2 3′ UTR (Figure 3.1A). The GS10 linkers on both sides of the eYFP tag could assist the folding of fusion proteins. The tagging constructs were made by fusion PCR of three fragments: a ~500 bp fragment homologous to end of the 5′ UTR of each gene, a region of the pPOTv4 plasmid containing the blasticidin resistance cassette (bsr) and the eYFP tag, and a ~500 bp fragment homologous to the beginning of the coding sequence of each gene (Figure 3.1A). The long regions of homology to the 5′ UTR and the CDS of each gene were required for efficient homologous recombination and construct integration. Tagging constructs (~2.7 kb) were transfected into bloodstream slender

(monomorphic) parasites of the Lister 427 strain. The semi-clonal populations (designated as "clones" in this thesis for simplicity) obtained after transfection and blasticidin selection were tested for correct integration of the tagging construct in the genome by PCR and expression of the tagged protein by western blot analysis (Appendix C). Typically, four clones per construct were picked after the mock-transfected cells had died. Tagging of a particular protein was deemed successful when the construct integration was correct and/or the tagged protein was detected by western blot analysis. More emphasis was given to the westerns, although it was taken into account that some proteins may not be expressed in all developmental forms of the parasite.

Table 3.1 Tagging summary

	Protein class	Histone mark	Members identified	Tagged successfully	Nuclear	Cytoplasmic	Nuclear and cytoplasmic
Readers	Bromo-domain	ac	7	7	6	-	1
	PHD-domain	me	5	4	1	-	3
	Tudor-domain	me	1	1	-	1	-
	Chromo-domain	me	1	1	1	-	-
	PWWP	me	1	1	1	-	-
	Znf-CW	me	1	1	1	-	-
Writers	HAT	ac	11	10	4	2	4
	SET-domain	me	30	30	-	21	9
	DOT	me	3	2	-	-	2
Eraser	HDAC	ac	7	7	1	3	3
	JmjC	me	4	4	1	2	1
Total:			**71**	**68**	**16**	**29**	**23**

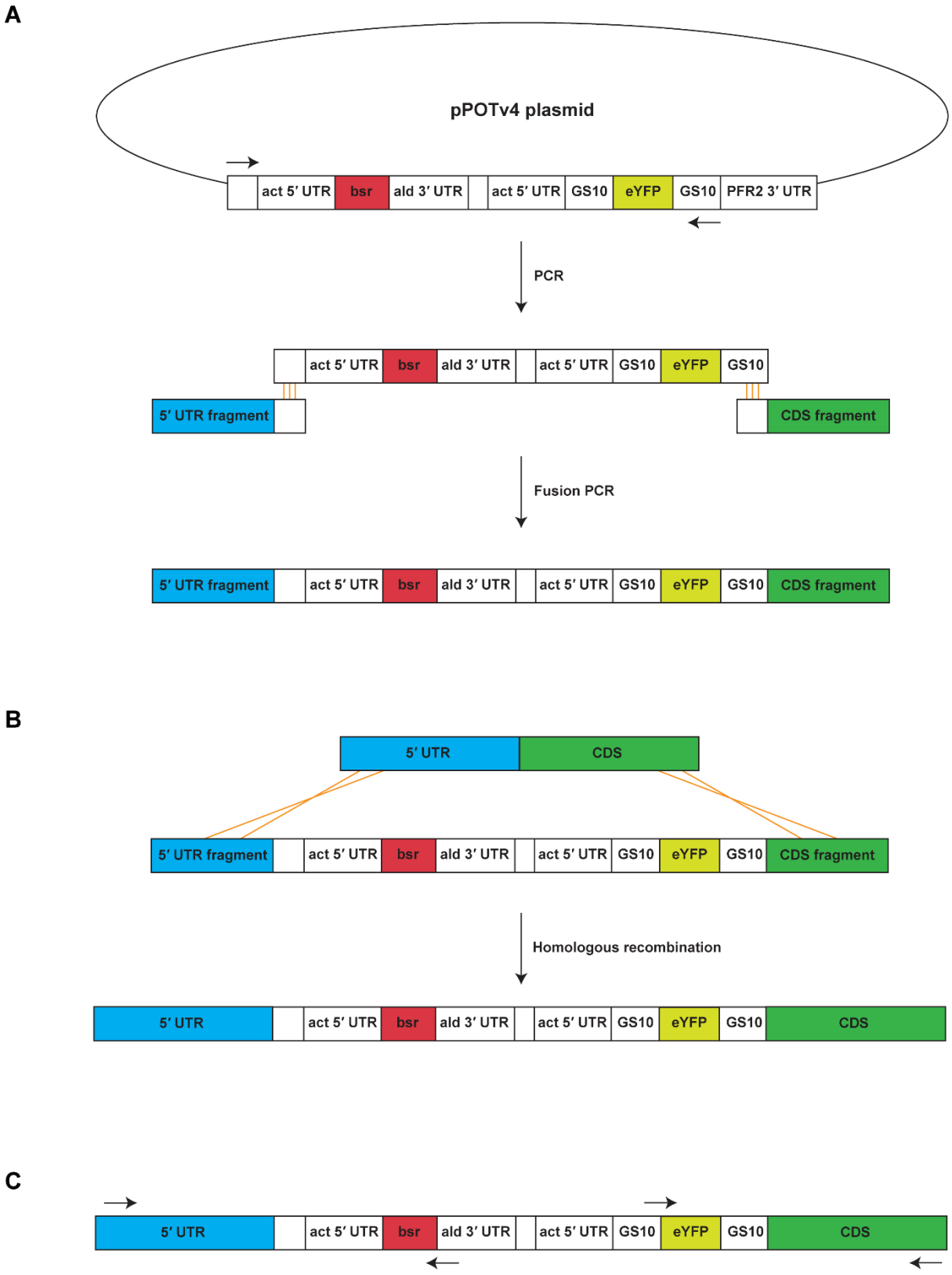


Figure 3.1 Tagging strategy.

Legend continued on the next page.

Figure 3.1 continued

(A) Preparation of the tagging constructs by fusion PCR. A fragment containing the bsr gene and the eYFP tag was PCR amplified from the pPOTv4 plasmid (the positions of the forward and reverse primers are shown as arrows). A ~500 bp fragment was PCR amplified from the end of the 5′ UTR of each gene, with a 3′ overhang (shown as a white box) being added by the reverse primer. A ~500 bp fragment was PCR amplified from the beginning of the CDS of each gene, with a 5′ overhang (shown as a white box) being added by the forward primer. The overhangs were complementary to the fragment from the pPOTv4 plasmid and were utilised to combine the three fragments together by fusion PCR.

(B) Tagging construct integration via homologous recombination using homologies within the 5′ UTR and CDS regions.

(C) PCR validation for correct integration of the tagging constructs after transfection and blasticidin selection. Two primer pairs were used (their positions are shown as arrows): the left primer pair tested for correct integration at the 5′ UTR junction and the right primer pair – for correct integration at the CDS junction of the candidate protein genes.

The tagging strategy is illustrated below for SET27 (Tb927.9.13470). Multiple PCR reactions were run in order to amplify sufficient amounts of the pPOTv4, 5′ UTR and CDS fragments (Figure 3.2A). The three types of fragments were joined together via several fusion PCR reactions (Figure 3.2B). The fusion PCR always generated many unspecific bands. Thus, the fusion construct was always gel extracted prior to transfection. Genomic DNA and protein were extracted from the four SET27 clones chosen after blasticidin selection to test for correct construct integration (Figure 3.2C) and YFP-tagged protein expression (Figure 3.2D), respectively.

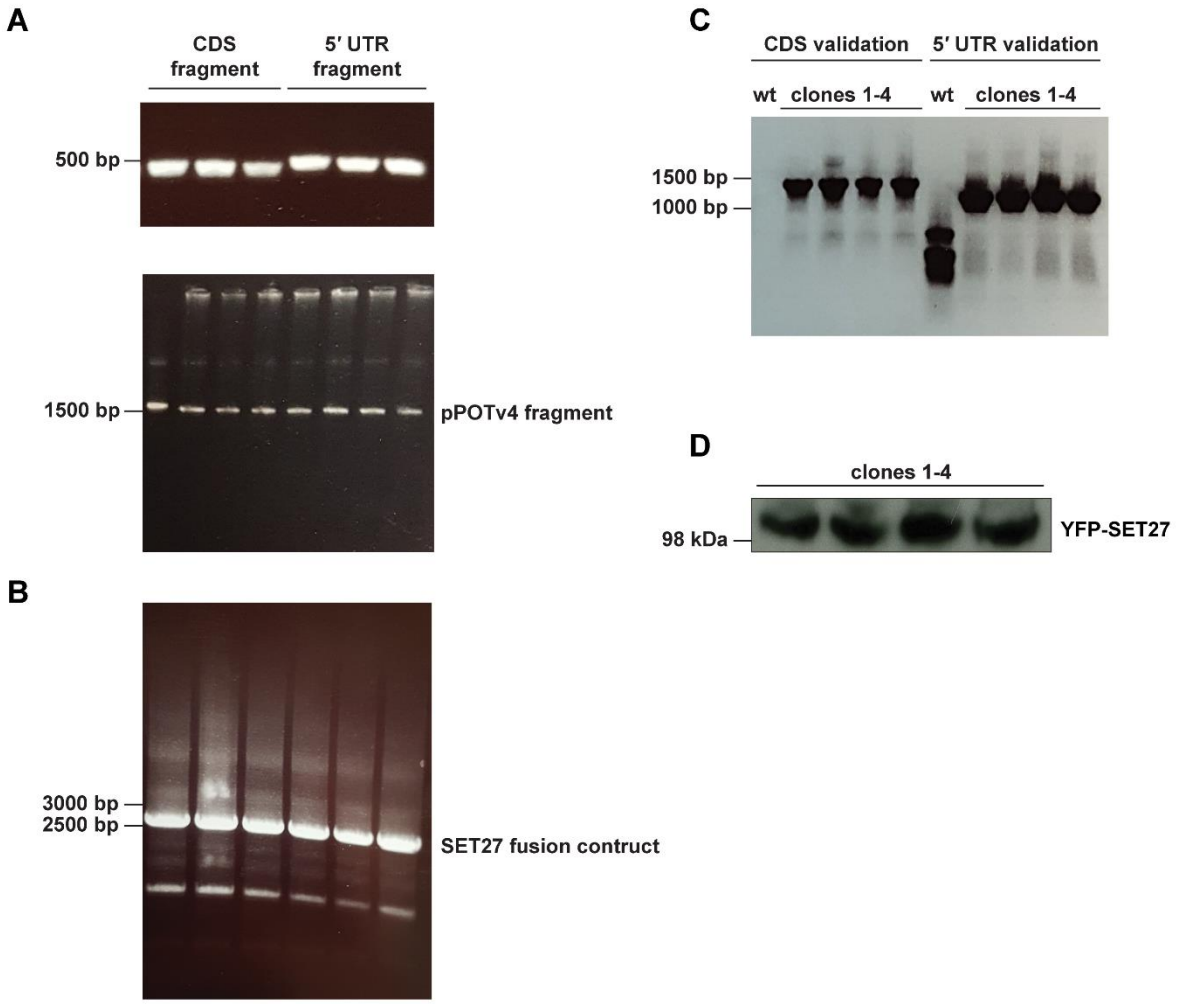


Figure 3.2 SET27 tagging.

(A) PCR amplification of the SET27 CDS (464 bp) and 5′ UTR (505 bp) fragments and the pPOTv4 fragment (1694 bp).

(B) Production of the SET27 tagging construct (2663 bp) via fusion PCR.

(C) PCR validation of correct tagging construct integration. Genomic DNA from wt untagged cells was used as a negative control. Genomic DNA from the four SET27 clones selected showed correct integration of the tagging construct at the CDS (1237 bp) and 5′ UTR (1055 bp) junctions.

(D) The presence of YFP-tagged SET27 (112.8 kDa) was detected in all four clones.

3.3 Localisation of the candidate proteins

The YFP tags were used for immunolocalisation of the candidate proteins (Figure 3.3). Proteins with known localisation, such as the kinetochore protein KKT2 (Akiyoshi and Gull, 2014; Llauró et al., 2018), were used as controls. As expected, KKT2 formed discrete puncta on opposite poles of the two daughter nuclei during cell division (Figure 3.3). From this section onwards, unless otherwise indicated, I present data only for the proteins I tagged and subsequently worked on during this project.

Table 3.2 summarises the candidate protein localisation in the bloodstream form identified in this study and compares it to the localisation in the procyclic form as determined in the TrypTag project (www.tryptag.org), the aim of which is to tag and localise all *T. brucei* proteins except VSGs (Dean et al., 2017). In TrypTag, proteins were tagged with mNeonGreen (Shaner et al., 2013) which is approximately the same size as YFP but is brighter and protein localisation can be visualised microscopically without antibody staining. As of 5 January 2020, data for 20 out of my 34 candidates (59%) were available on TrypTag, and most of these were in agreement with my observations. However, a different localisation in the PF was observed for Tudor1, HDAC2, Sir2rp3 and LCM1. For two of these proteins, the images and their descriptions on TrypTag did not match: Tudor1 was classified as a nuclear and cytoplasmic protein but the images shown on the website support an exclusively cytoplasmic localisation; Sir2rp3 was described as a cytoplasmic protein but the available images show that it is also present in the nucleus. Most candidates were tagged N-terminally in the TrypTag project, with some genes (PWWP1, Znf-CW1, SET5, SET21, SET24, SET26 SET30, HDAC2, HDAC3, Sir2rp3, Jmj1) also being tagged on the C terminus. In many of the latter cases, changing the position of the tag also resulted in an altered cellular localisation, highlighting the fact that tagging can lead to protein mislocalisation.

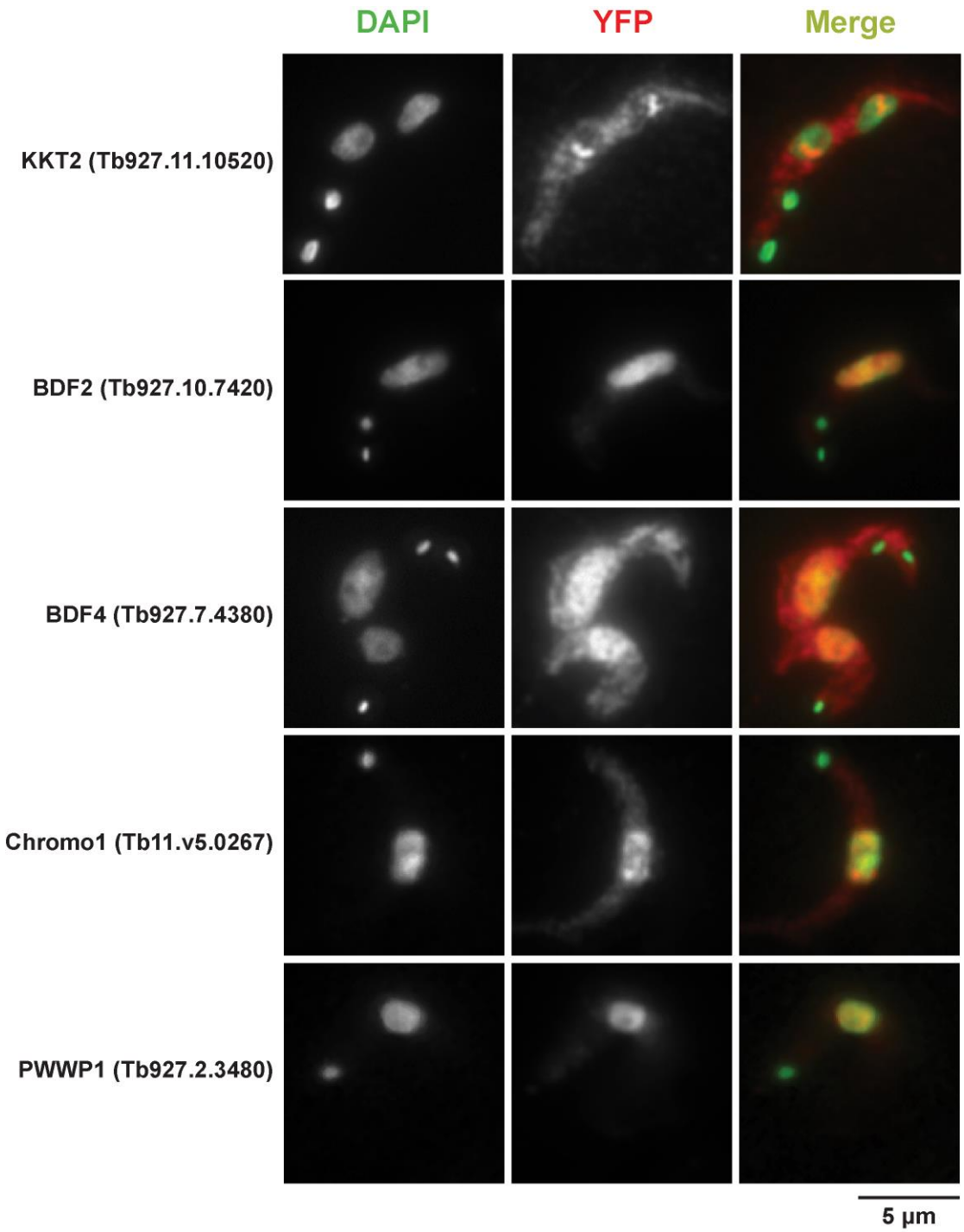


Figure 3.3 Localisation of the candidate proteins. Candidate proteins were immunolocalised in bloodstream form trypanosomes using an anti-GFP primary and a red fluorescently labelled secondary antibody. Nuclear and kinetoplast DNA were stained with DAPI. The previously characterised kinetochore protein KKT2 was used as a control.

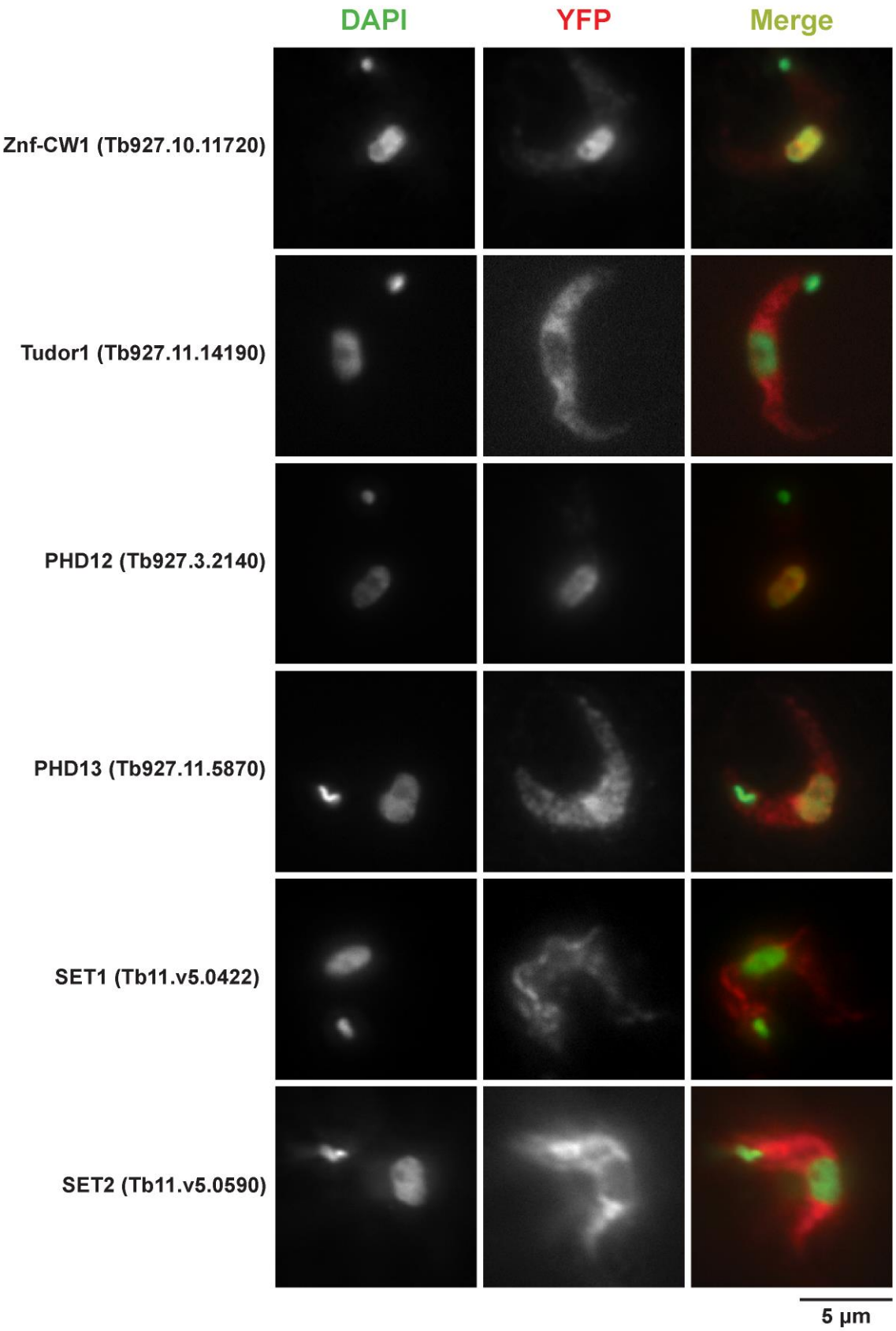


Figure 3.3 continued

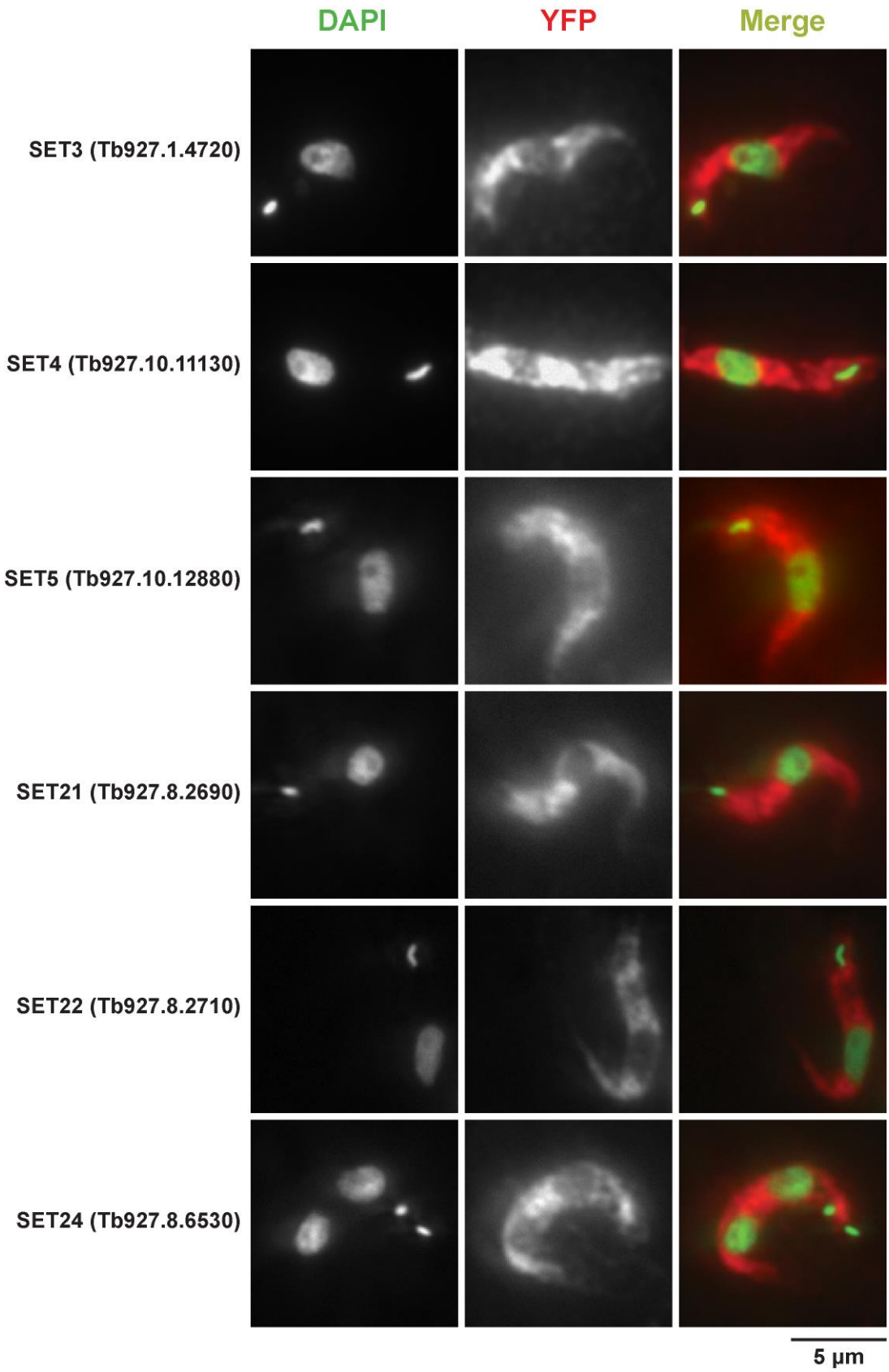


Figure 3.3 continued

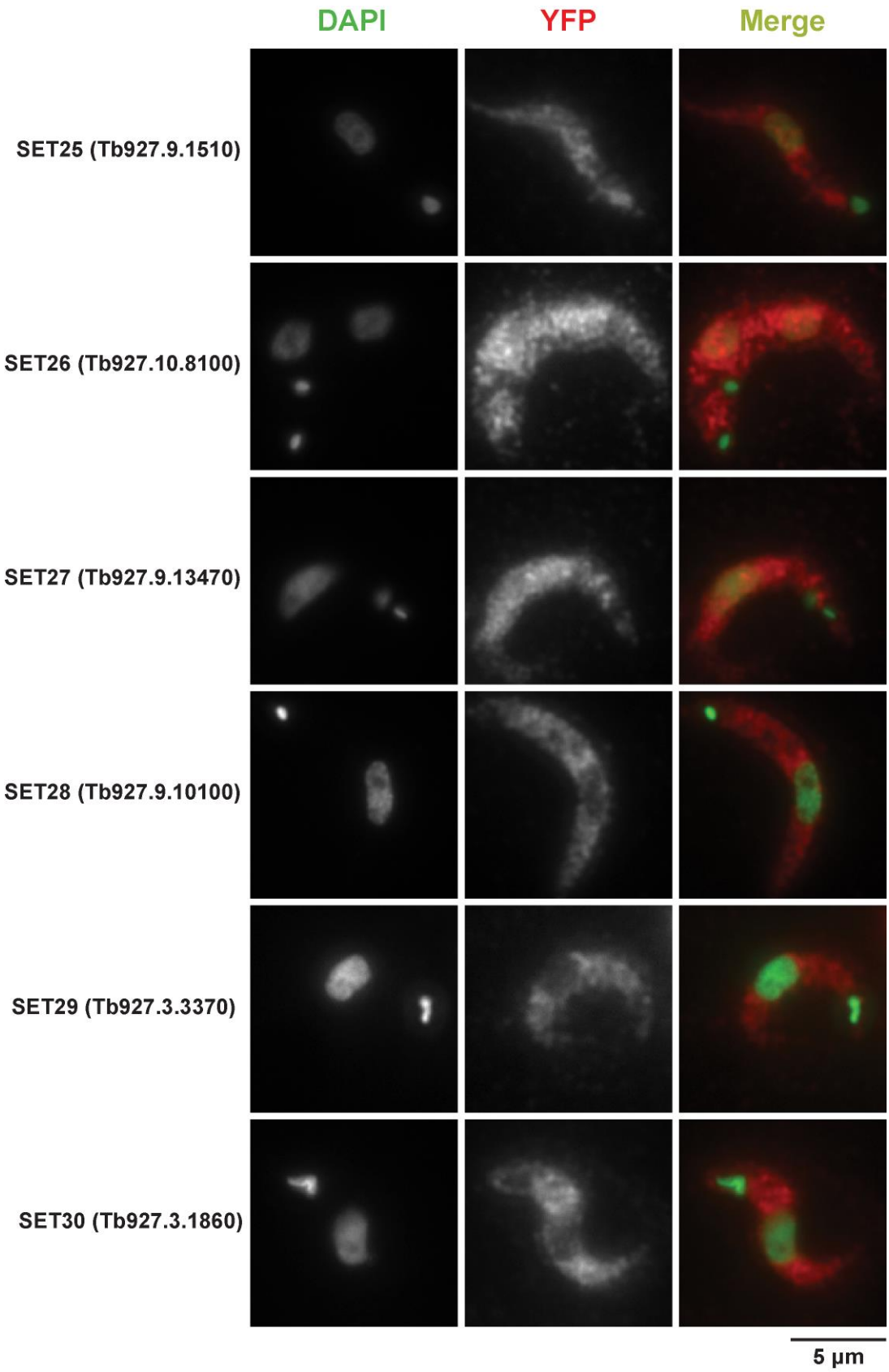


Figure 3.3 continued

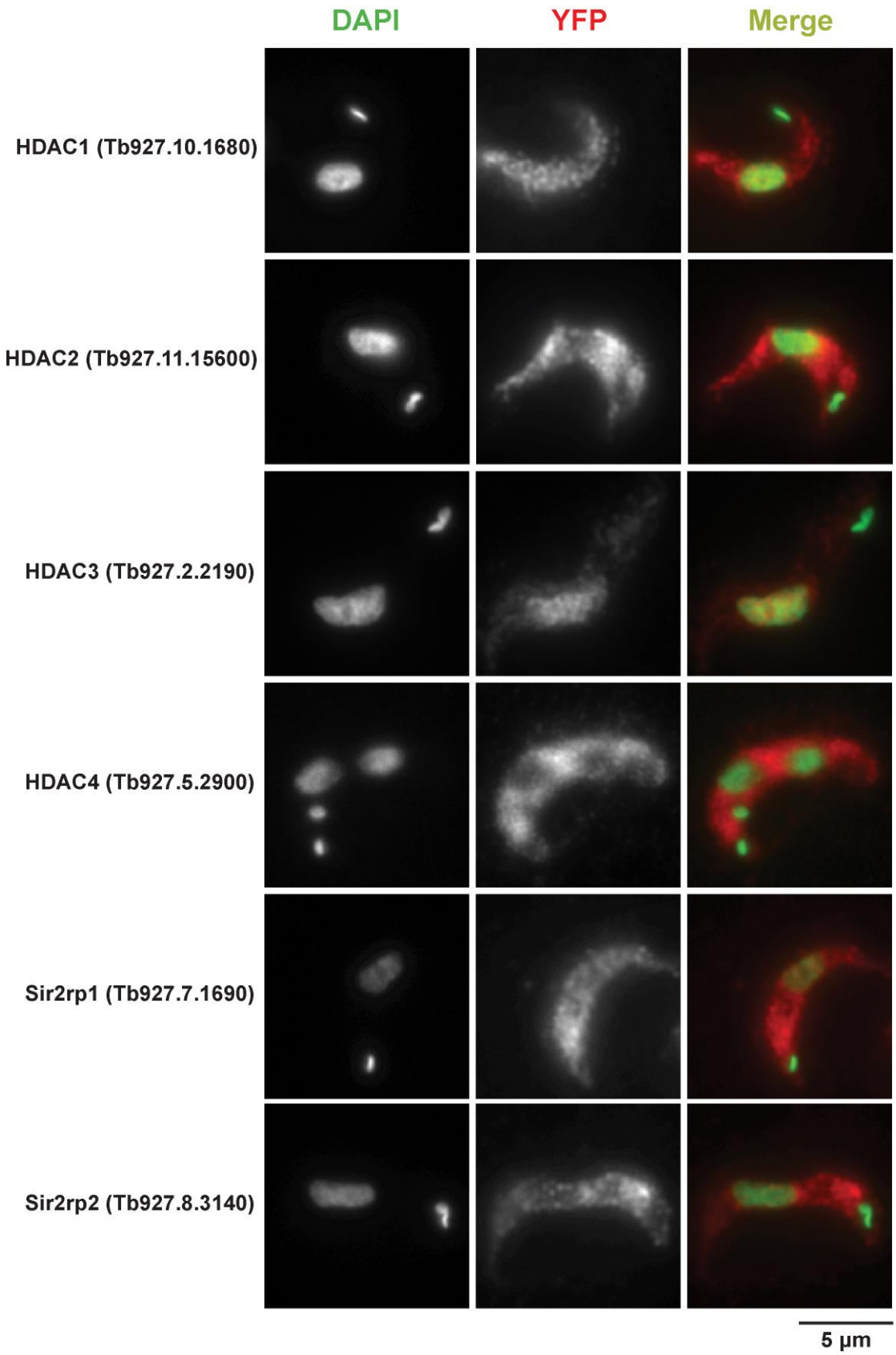


Figure 3.3 continued

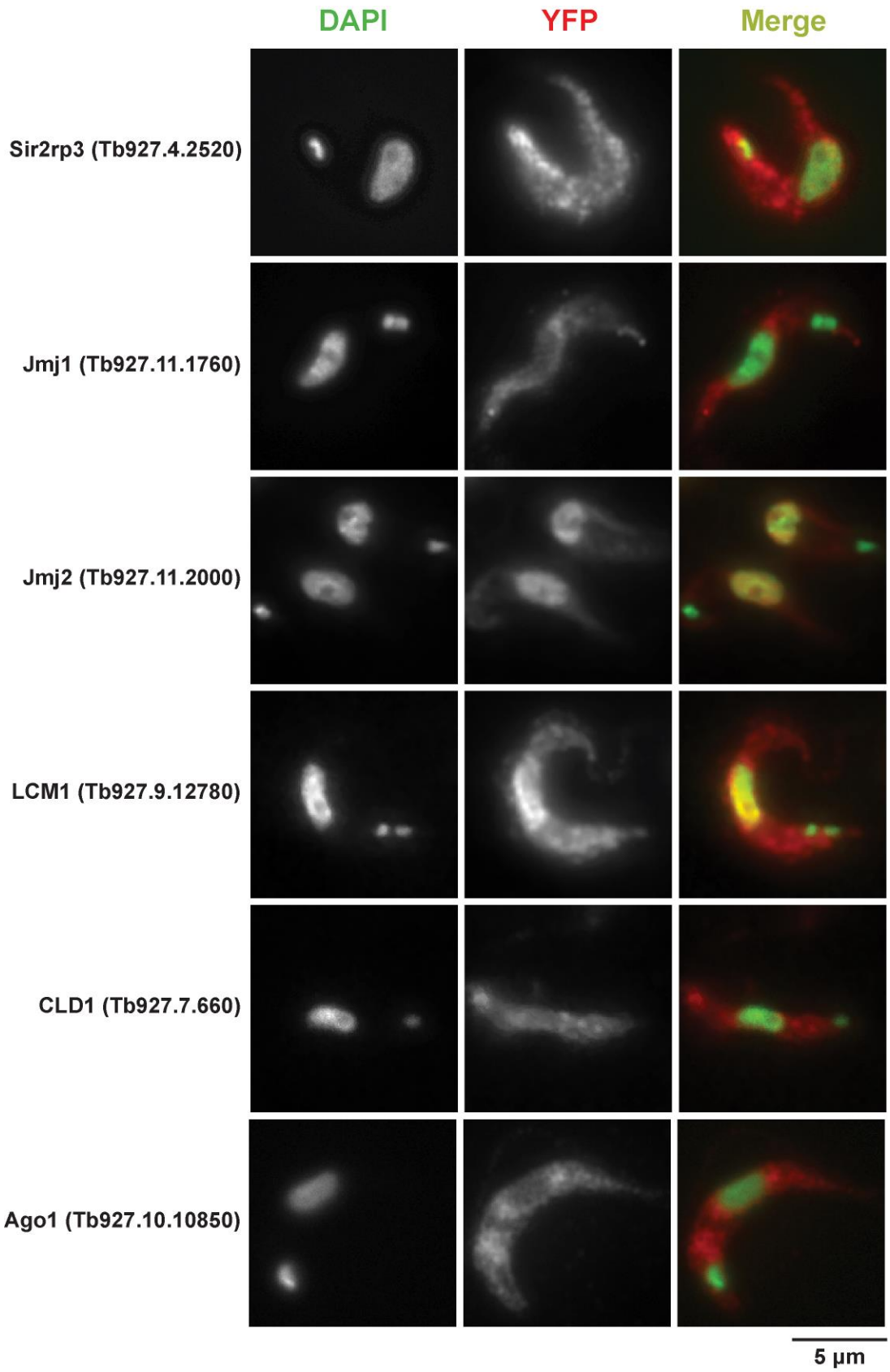


Figure 3.3 continued

Table 3.2 Comparison of candidate localisation with TrypTag data

	Protein name	Gene ID	Protein domains	Localisation in this study (BF)	TrypTag[1] localisation (PF)	Additional information (TriTrypDB[2])
Readers	BDF2	Tb927.10.7420	Bromo	nucleus	N-terminal tag: nucleus	
	BDF4	Tb927.7.4380	Bromo	nucleus and cytoplasm	-	
	PHD12	Tb927.3.2140	PHD	nucleus	-	Putative transcription activator
	PHD13	Tb927.11.5870	PHD	nucleus and cytoplasm	-	
	Tudor1	Tb927.11.14190	Tudor	cytoplasm	N-terminal tag: cytoplasm, flagellar cytoplasm, nuclear lumen*	
	Chromo1	Tb11.v5.0267	Chromo	nucleus	-	
	PWWP1	Tb927.2.3480	PWWP	nucleus	N-terminal tag and C-terminal tag: nucleoplasm	Putative transcription elongation factor s-II
	Znf-CW1	Tb927.10.11720	Znf-CW	nucleus	N-terminal tag and C-terminal tag: nucleoplasm	
Writers	SET1	Tb11.v5.0422	SET	cytoplasm	-	
	SET2	Tb11.v5.0590	SET	cytoplasm	-	
	SET3	Tb927.1.4720	SET	cytoplasm	C-terminal tag: cytoplasm	
	SET4	Tb927.10.11130	SET	cytoplasm	N-terminal tag: cytoplasm (points)	
	SET5	Tb927.10.12880	SET	cytoplasm	N-terminal tag: cytoplasm (reticulated); C-terminal tag: kinetoplast, mitochondrion	
	SET21	Tb927.8.2690	SET	cytoplasm	N-terminal tag: cytoplasm (reticulated); C-terminal tag: nucleoplasm, cytoplasm, flagellar cytoplasm	
	SET22	Tb927.8.2710	SET	cytoplasm	-	
	SET24	Tb927.8.6530	SET	cytoplasm	N-terminal tag: cytoplasm (reticulated); C-terminal tag: kinetoplast, mitochondrion	
	SET25	Tb927.9.1510	SET	nucleus and cytoplasm	-	
	SET26	Tb927.10.8100	SET	nucleus and cytoplasm	N-terminal tag and C-terminal tag: nucleoplasm, cytoplasm	
	SET27	Tb927.9.13470	SET	nucleus and cytoplasm	-	
	SET28	Tb927.9.10100	SET	cytoplasm	-	
	SET29	Tb927.3.3370	SET	cytoplasm	-	
	SET30	Tb927.3.1860	SET	cytoplasm	N-terminal tag: cytoplasm (points); C-terminal tag: kinetoplast, mitochondrion	
Erasers	HDAC1	Tb927.10.1680	-	nucleus and cytoplasm	C-terminal tag: cytoplasm, flagellar cytoplasm, nuclear lumen	Chromatin de-repression in BF
	HDAC2	Tb927.11.15600	-	cytoplasm	N-terminal tag: cytoplasm, flagellar cytoplasm, nuclear lumen; C-terminal tag: cytoplasm (reticulated)	
	HDAC3	Tb927.2.2190	-	nucleus	N-terminal tag: nucleolus (peripheral, points), nucleoplasm (points); C-terminal tag: nucleoplasm	VSG silencing
	HDAC4	Tb927.5.2900	-	cytoplasm	-	
	Sir2rp1	Tb927.7.1690	-	nucleus and cytoplasm	-	
	Sir2rp2	Tb927.8.3140	-	cytoplasm	-	
	Sir2rp3	Tb927.4.2520	-	nucleus and cytoplasm	N-terminal tag: cytoplasm, flagellar cytoplasm*; C-terminal tag: mitochondrion	
	Jmj1	Tb927.11.1760	Cupin (JmjC)	cytoplasm	N-terminal tag: cytoplasm; C-terminal tag: cytoplasm, flagellar cytoplasm, nucleoplasm	
	Jmj2	Tb927.11.2000	Cupin	nucleus	N-terminal tag: nucleolus, nucleus; C-terminal tag: nucleus	
	LCM1	Tb927.9.12780	Cupin	nucleus and cytoplasm	N-terminal tag: cytoplasm (patchy)	
	CLD1	Tb927.7.660	Cupin	cytoplasm	C-terminal tag: cytoplasm	
	Ago1	Tb927.10.10850	-	cytoplasm	C-terminal tag: cytoplasm	

[1], [2] - the data from www.tryptag.org and www.tritrypdb.org was accessed on 05 January 2020.

3.4 Discussion

This Chapter presented data on the tagging and cellular localisation of 8 putative readers, 14 putative writers and 11 putative erasers of trypanosome histone PTMs as well as data on Ago1 and control proteins. The candidates chosen were largely uncharacterised with a few exceptions. HDAC1 and HDAC3 are essential histone deacetylases, with HDAC1 opposing Sir2rp1-dependent telomeric silencing in BF only, whereas HDAC3 is required for VSG ES silencing in both BF and PF cells (Alsford et al., 2007; Wang et al., 2010).

This initial screen for protein localisation showed that 7 of the candidates are nuclear, 17 – cytoplasmic, and 9 – both nuclear and cytoplasmic. One surprising finding from this survey was that the majority of SET-domain proteins are cytoplasmic, and none are exclusively nuclear, suggesting that they might methylate non-histone substrates in the cytoplasm.

One issue with the immunolocalisation methodology used was cytoplasmic background from the secondary antibody. Thus, although proteins could be confidently classified as nuclear, cytoplasmic localisation was less certain. Since nuclear proteins were of particular interest, this drawback was deemed acceptable.

Comparison with the available TrypTag data showed that in most cases the protein localisation is identical in bloodstream and procyclic cells. The two clear differences observed (for HDAC2 and LCM) could be due to differential localisation of the candidates throughout the trypanosome life cycle. No cell cycle differences in localisation were observed during the screen in BF parasites.

Overall, in this Chapter, I laid the groundwork for the remainder of this project by tagging and localising proteins which are potentially involved in trypanosome chromatin organisation and transcription regulation. In subsequent experiments, the YFP tags were used to perform ChIP-seq analysis to determine protein distribution across the *T. brucei* genome, and affinity selections followed by mass spectrometry to identify protein interaction networks.

CHAPTER 4

Many of the nuclear putative chromatin regulators are concentrated at RNAPII transcription start regions

4.1 Introduction

T. brucei is thought to regulate its gene expression predominantly post-transcriptionally. However, electron micrographs have shown the presence of darkly stained regions in the nuclei of bloodstream and procyclic parasites (Daniels et al., 2010; Rout and Field, 2001) which may correspond to transcriptionally inactive heterochromatic regions. The extensive reservoir of transcriptionally silent VSGs as well as repetitive elements, such as the ingi and SLACS retrotransposons and the centromeric CIR147 repeats, are expected to be found in these regions.

T. brucei's histones are highly divergent from their counterparts in other eukaryotes. A few histone PTMs and histone variants have been implicated in the regulation of trypanosome transcription: H3K4me3 and H4K10ac as well as H2AZ and H2BV are found at RNAPII TSRs whereas H3V and H4V mark RNAPII termination regions (Siegel et al., 2009; Wright et al., 2010). In general, however, detailed characterisation of trypanosome histone PTMs and their regulators is lacking.

To determine if any of the putative readers, writers and erasers of histone marks identified in Chapter 3 are involved in chromatin organisation, the YFP-tagged candidate proteins with nuclear localisation were analysed by ChIP-seq to assess their association with and distribution across the genome. It was hypothesised that some proteins may be enriched at repetitive elements or silent VSG arrays which would indicate that they play a role in

trypanosome chromatin silencing. Conversely, candidates found at TSRs or over polycistronic transcription units could be exerting positive or negative regulation on transcription.

4.2 Candidate protein enrichment at RNAPII TSRs

ChIP-seq was performed on the nuclear and nuclear and cytoplasmic candidate proteins. The untagged Lister 427 cell line was used as a negative control and a YFP-tagged kinetochore protein (KKT2) was included as a positive control. ChIP-seq was also performed on Ago1 which microscopically appeared cytoplasmic but was suspected to interact with chromatin, analogously to its association with siRNA producing loci in other eukaryotes (Allshire and Madhani, 2018). Two ChIP-seq replicates were performed for each cell line except Chromo1 for which three replicates were performed. Thus, a total of 94 samples were sequenced (47 inputs and 47 ChIPs). The data obtained were subjected to sequence duplicate removal and were subsequently aligned to the Tb427v9.2 genome (Müller et al., 2018). The ChIP data was then normalised to input data and visualised in a genome browser. Previously published ChIP-seq data for H2AZ (Wedel et al., 2017) was used for comparison. The results showed that 8 proteins are concentrated at RNAPII TSRs, as marked by H2AZ histones (Figure 4.1A), whereas the other 13 proteins tested did not show enrichment at specific regions across the genome (Figure 4.1B). The 8 positive proteins did not display peaks at any other locations including the active VSG-2 ES or the silent VSG ESs. H2AZ ChIP-seq data showed enrichment not only at RNAPII TSRs but also over 7 out of the 11 megabase chromosome centromeres. However, no input data was available to normalise the H2AZ ChIP to, so enrichment over repetitive centromeric regions could represent an experimental artefact. KKT2, as expected, had peaks only over centromeric regions.

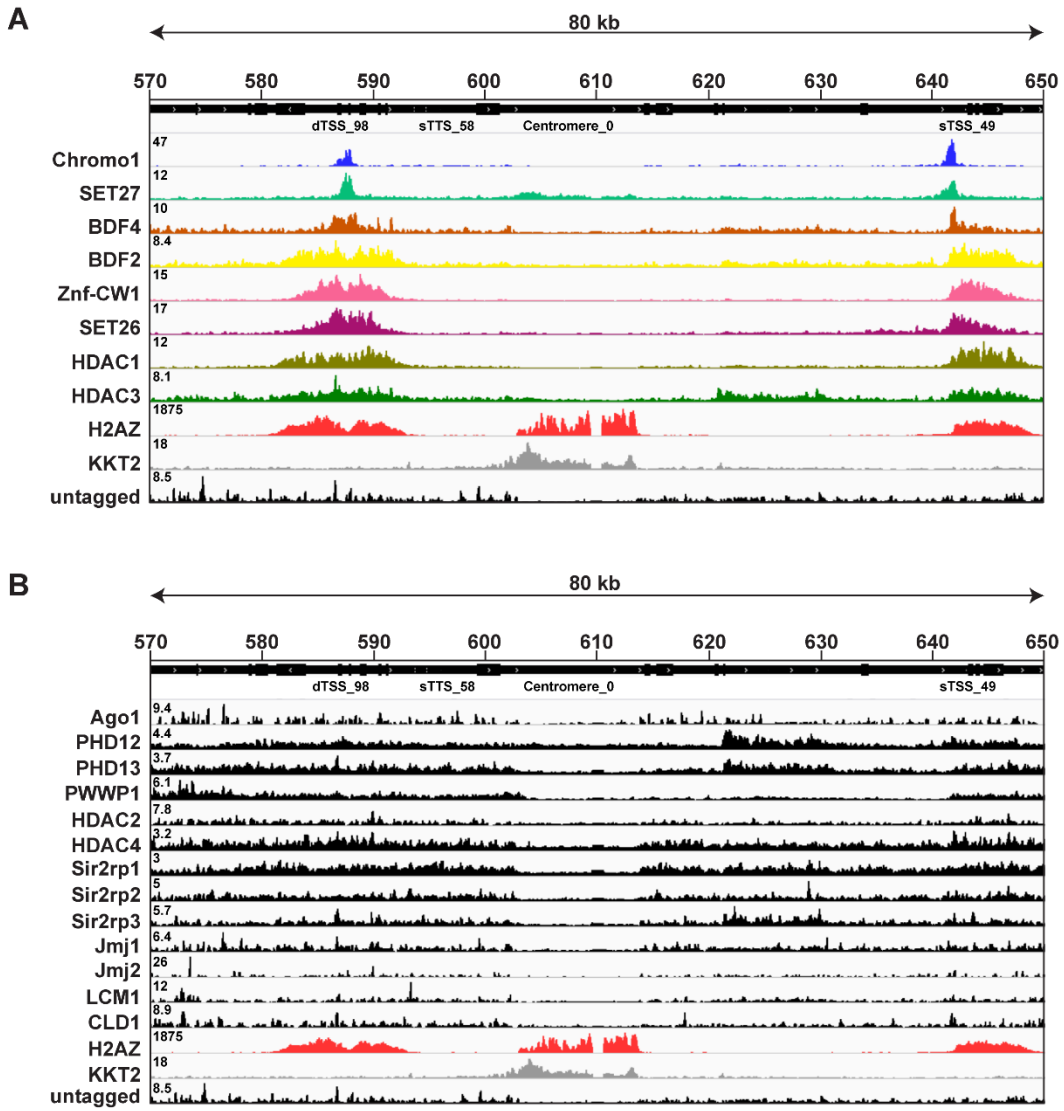


Figure 4.1 Candidate protein enrichment at RNAPII TSRs. The figure shows an 80 kb region on chromosome 1 encompassing a centromere and two TSRs (dTSS – divergent TSS; sTSS – single TSS). Only a single ChIP-seq replicate for each protein is shown. Each track was scaled separately to the highest peak summit (shown as a number in the top left corner of each track) in order to demonstrate the peak shapes better.

(A) Candidate proteins with peaks at RNAPII TSRs marked by H2AZ histones. H2AZ data from Wedel et al (2017) was used.

(B) Candidate proteins with no enrichment in the same region. The kinetochore protein KKT2 and the untagged Lister 427 cell line were included as a positive and a negative control, respectively.

4.3 Analysis of the candidate protein enrichment profiles

Although many of the candidate proteins analysed were concentrated at RNAPII TSRs, their enrichment profiles differed. Chromo1, SET27 and BDF4 displayed sharp peak profiles whereas BDF2, Znf-CW1, SET26, HDAC1 and HDAC3 were broadly enriched. These trends were analysed bioinformatically using Chromo1 as a reference (Figure 4.2). 177 Chromo1 peaks were called using MACS2 (Feng et al., 2012) combined with manual filtering of the false positives. Protein enrichment was then plotted in 50 bp windows 10 kb upstream and 10 kb downstream of the peak summits. The general peak profile for each protein was represented as a metagene plot whereas the enrichment around individual Chromo1 peaks was shown as a heatmap (Figure 4.3). The results of the analysis showed that Chromo1, SET27 and BDF4 have sharp peaks at all locations whereas the other candidates consistently displayed a broad enrichment around Chromo1.

The metagene plots in Figure 4.3 were generated by summing the reads around all Chromo1 peak summits, normalising them to library size and input and representing them as a density plot. Thus, more prominent peaks had a higher contribution to the metagene plot than less prominent ones. Average metagene plots were also generated by averaging the reads in each of the 50 bp windows before plotting them as a density around Chromo1 (Figure 4.4). The average metagene plots showed the same trends as seen in Figure 4.3 but were noisier because minor peaks contributed to the plots equally to major ones.

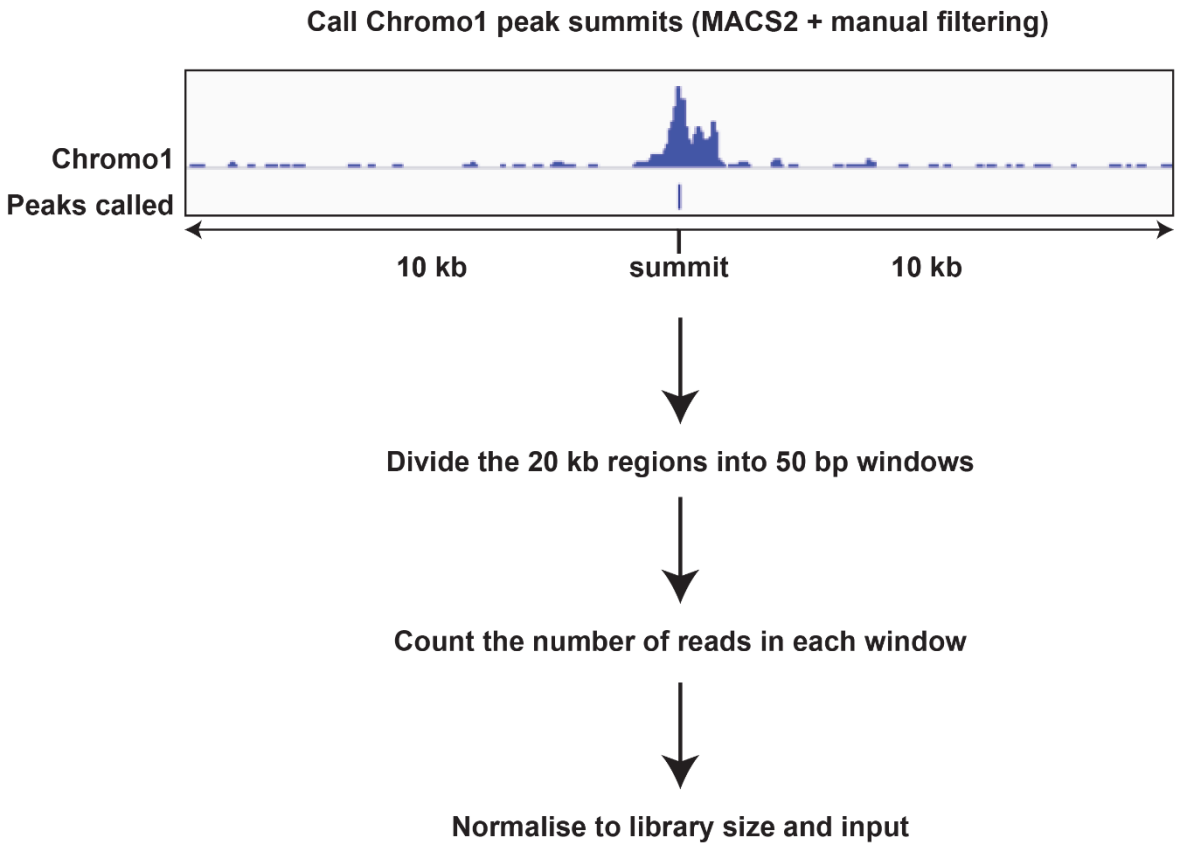


Figure 4.2. ChIP-seq analysis strategy. Chromo1 peak summits were called using MACS2 combined with manual filtering of the false positive calls. The 20 kb region around each summit was divided into 50 bp windows. For each candidate analysed, the reads in each window were counted and normalised to library size and input.

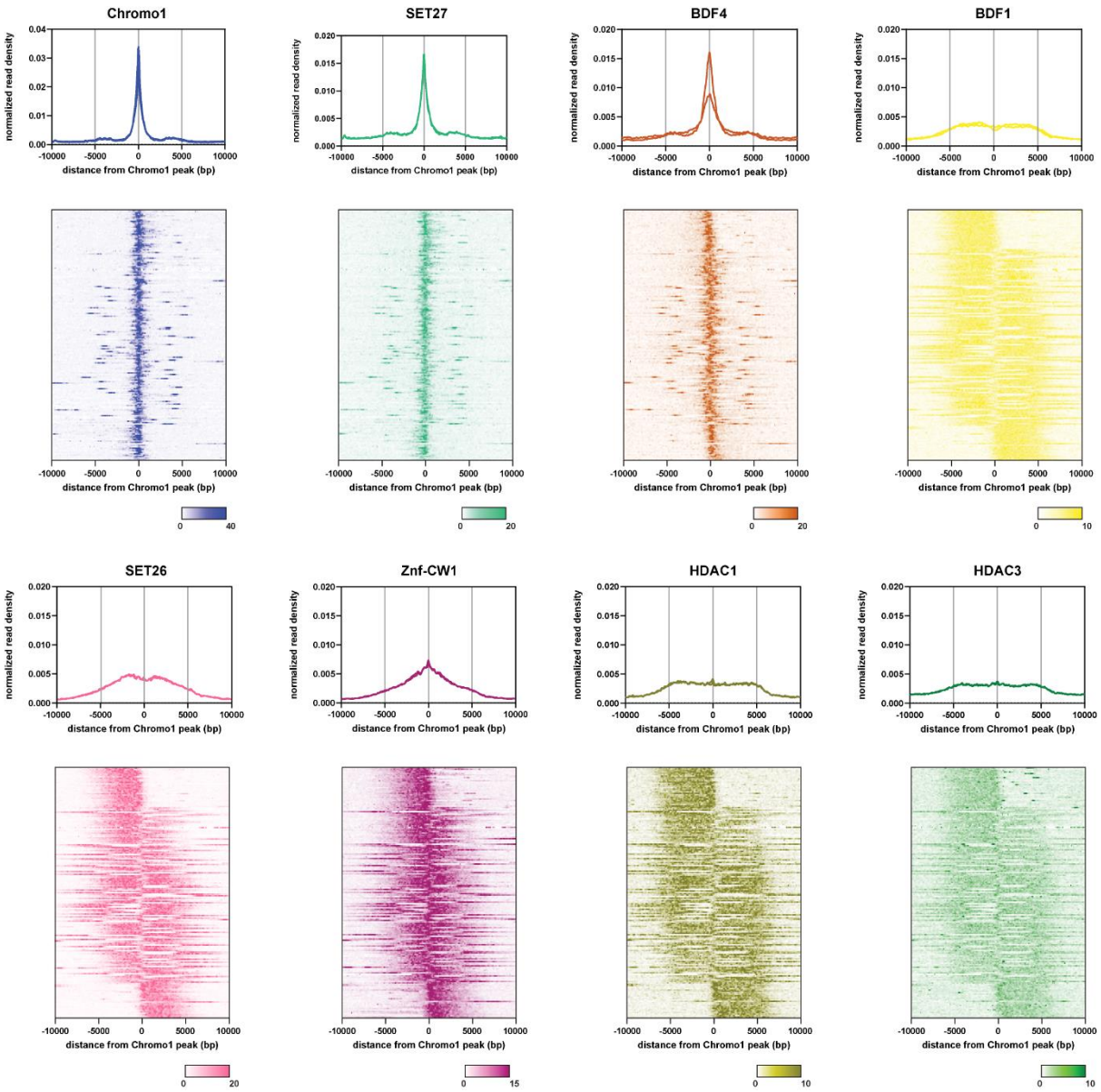


Figure 4.3 Enrichment profile analysis. For each protein, read density around all Chromo1 peaks was added and represented as a metagene plot whereas read density around individual Chromo1 peaks was shown as a heatmap. Going from top to bottom, each heatmap is showing enrichment around the 177 Chromo1 peaks in the same order. The scale bars at the bottom of each heatmap represent reads normalised to library size and input.

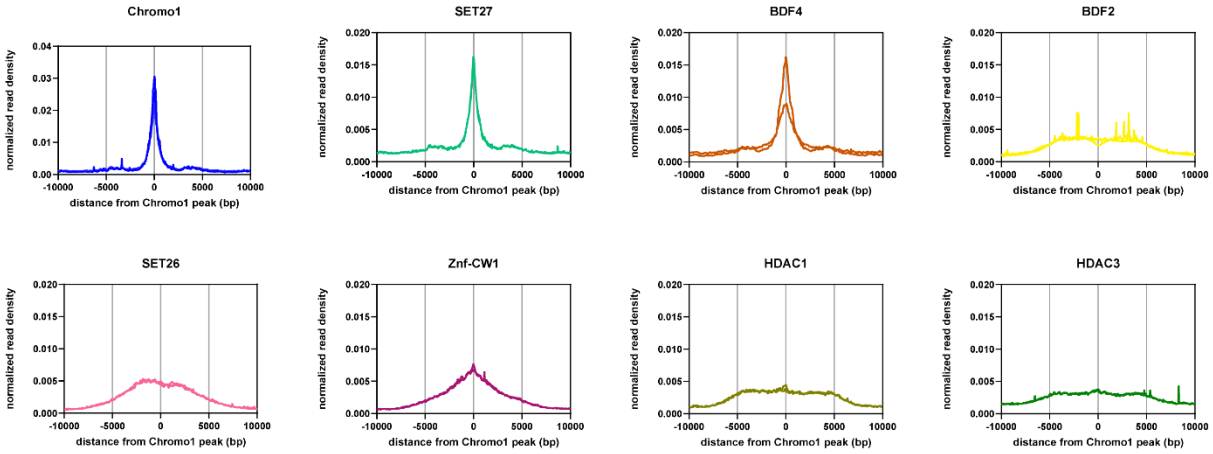


Figure 4.4 Average metagene plots. For each protein, reads in each 50 bp window around Chromo1 peak summits were added up and averaged before plotting as a density.

The candidate proteins that produced broad peaks at transcription start regions appeared to have enrichment in a specific direction around each Chromo1 peak summit. To investigate this further, the SET26 heatmap was sorted by directionality of its enrichment: the top of the heatmap shows enrichment in a 5′ direction, the middle – enrichment in both directions, and the bottom – enrichment in a 3′ direction from Chromo1 (Figure 4.5). The middle portion of the heatmap represents two types of profile – enrichment in both directions from a single Chromo1 peak or enrichment in opposite directions from two nearby Chromo1 peaks. The heatmaps of the other candidate proteins were sorted according to SET26 (Figure 4.3).

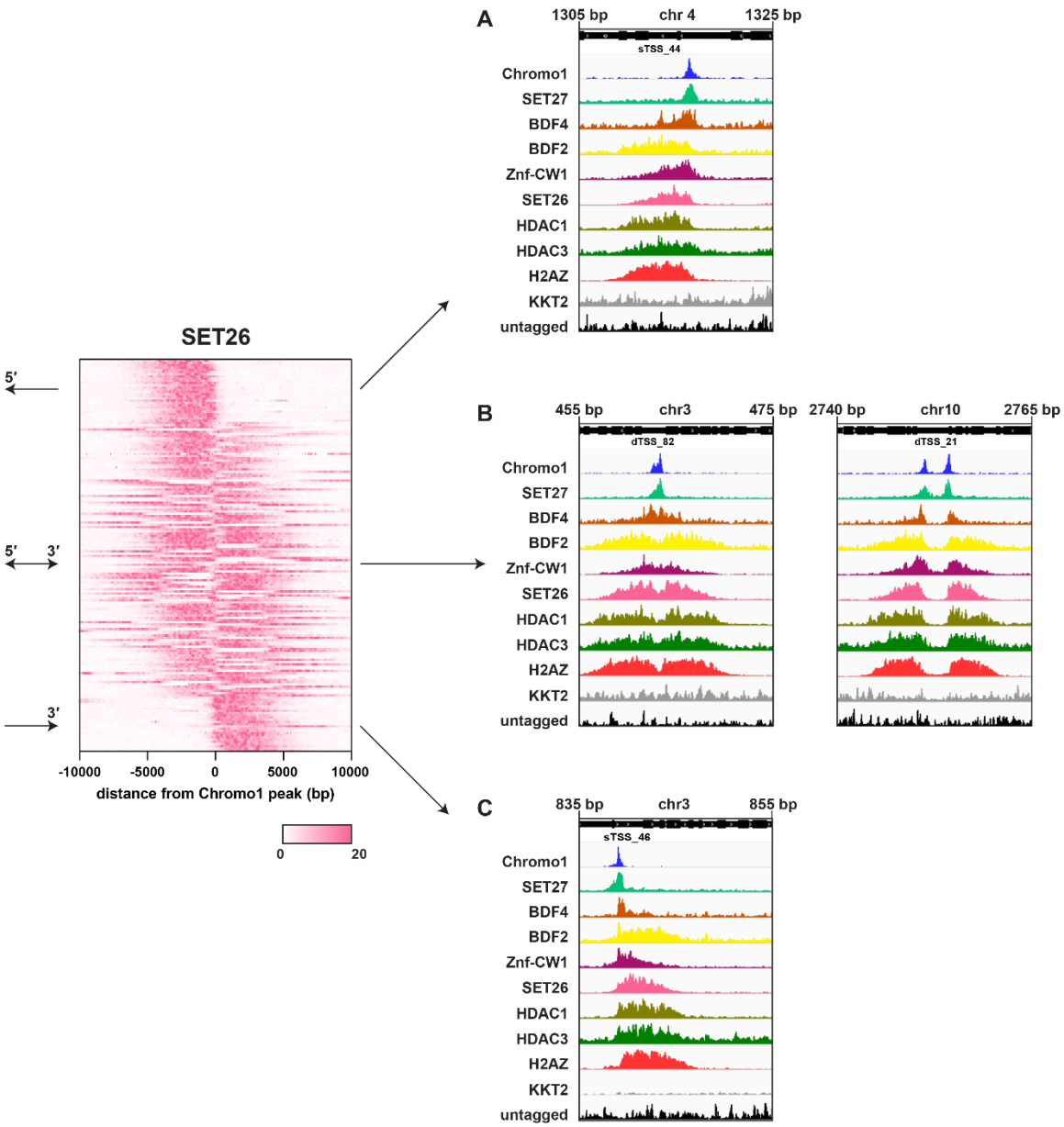


Figure 4.5 Heatmap sorting. The SET26 heatmap was sorted according to the direction of enrichment around Chromo1 peak summits. The panels on the right show example peaks from each region of the heatmap. All other heatmaps in Figure 4.3 are sorted according to SET26.

(A) Enrichment in a 5′ direction from Chromo1.

(B) Enrichment in both directions from Chromo1.

(C) Enrichment in a 3′ direction from Chromo1.

The heatmaps of the proteins with broad peaks showed identical patterns with respect to the direction of enrichment around Chromo1. This suggested that the direction of candidate enrichment might be correlated with the direction of RNAPII transcription. To address this, available RNA-seq data for slender form parasites (Naguleswaran et al., 2018) was analysed. The RNA reads were separated into those originating from the plus and the minus strand of DNA. The resulting data was plotted as heatmaps analogously to the candidate proteins (Figure 4.6). This revealed that the top part of the heatmaps represents two TSRs directing transcription along the minus stand of DNA (Figure 4.6A). Conversely, the bottom part of the heatmap represents two TSRs directing transcription along the plus strand of DNA (Figure 4.6B). The middle of the heatmaps represents two scenarios: a bi-directional TSR or two nearby TSRs that direct transcription in opposite directions on opposite DNA strands (Figure 4.6C). Overall, this analysis showed that the direction of enrichment of the candidates with broad peaks is indeed correlated with the direction of RNAPII transcription.

The enrichment levels of the candidate proteins at RNAPII TSRs were also quantified. Candidate protein reads were counted in non-overlapping 2 kb windows around each Chromo1 peak summit. In this analysis, Chromo1 peaks were used as a marker for transcription region locations. These reads were then normalised to library size and input and represented as box plots (Figure 4.7). On average, Chromo1 was the most highly enriched protein across RNAPII TSRs. Chromo1 also displayed higher variation in its enrichment compared to the other candidate proteins.

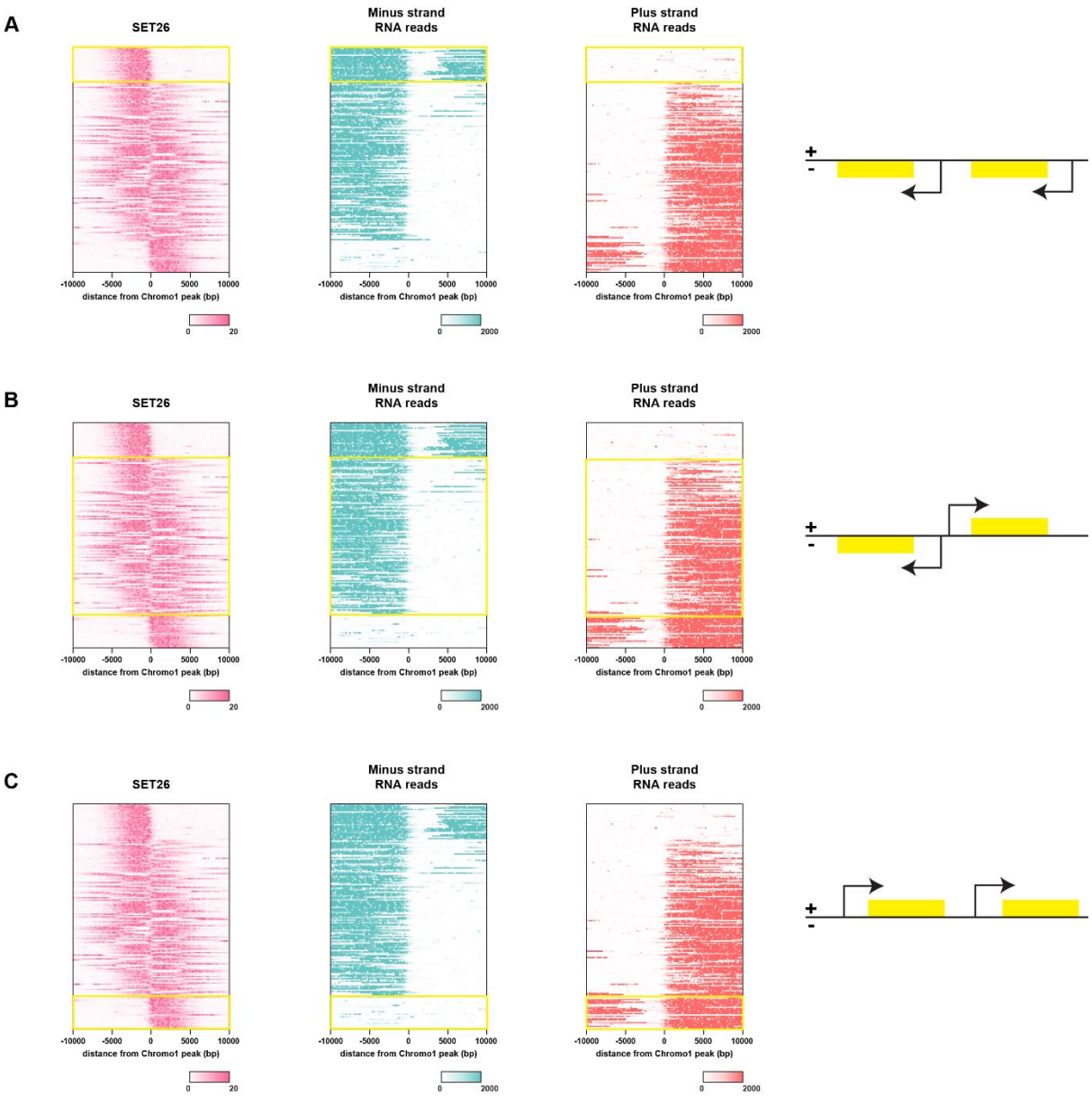


Figure 4.6 Candidate protein enrichment is correlated with the direction of RNAPII transcription. RNA-seq data from Naguleswaran et al. (2018) was used.

(A) The top part of the heatmaps represents transcription along the minus strand of DNA.

(B) The middle part of the heatmaps represents transcription going in opposite directions on the plus and minus strands of DNA.

(C) The bottom portion of the heatmaps shows transcription along the plus strand of DNA.

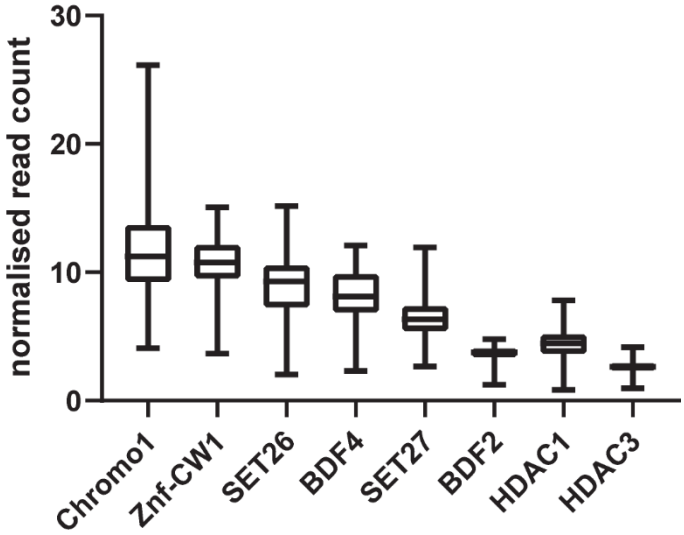


Figure 4.7 Candidate protein reads around RNAPII TSRs. The box plots represent the distribution of reads in 2 kb windows around Chromo1 peak summits which were used as RNAPII TSR markers. The boxes extend from the 25th to 75th percentiles. The line in each box shows the median value. The whiskers extend to the lowest and highest values.

4.4 Chromo1 binding sequence motif search

Chromo1, SET27 and BDF4 ChIP-seq profiles revealed sharp peaks at transcription start regions, which suggested that one or more of these candidate proteins might be recruited to specific genomic loci via a DNA sequence motif. To test this hypothesis, MEME Suite Motif Discovery software (http://meme-suite.org) was run on DNA sequences underlying Chromo1 peaks. Two window sizes were used – a 200 bp window encompassing the sequences immediately surrounding Chromo1 peak summits and a 1000 bp window including the sequences underlying the entire Chromo1 peaks. The highest-scoring motif for the 200 bp window was GT-rich (Figure 4.8A) consistent with previously published data about the nucleotide composition of trypanosome RNAPII TSRs (Wedel et al., 2017). The highest-scoring motif for the 1000 bp window was an A-rich sequence (Figure 4.8B).

The GT-rich motif was found at 68 sites across the 177 Chromo1 peaks whereas the A-rich motif was found at 258 sites. Additionally, the motifs were long and their positions were scattered throughout the 200 bp and 1000 bp windows rather than being concentrated immediately under the Chromo1 peak summits (Figure 4.8C and D). These observations suggest that the sequences found via MEME are not true binding motifs but rather result from skewed nucleotide composition at the TSRs.

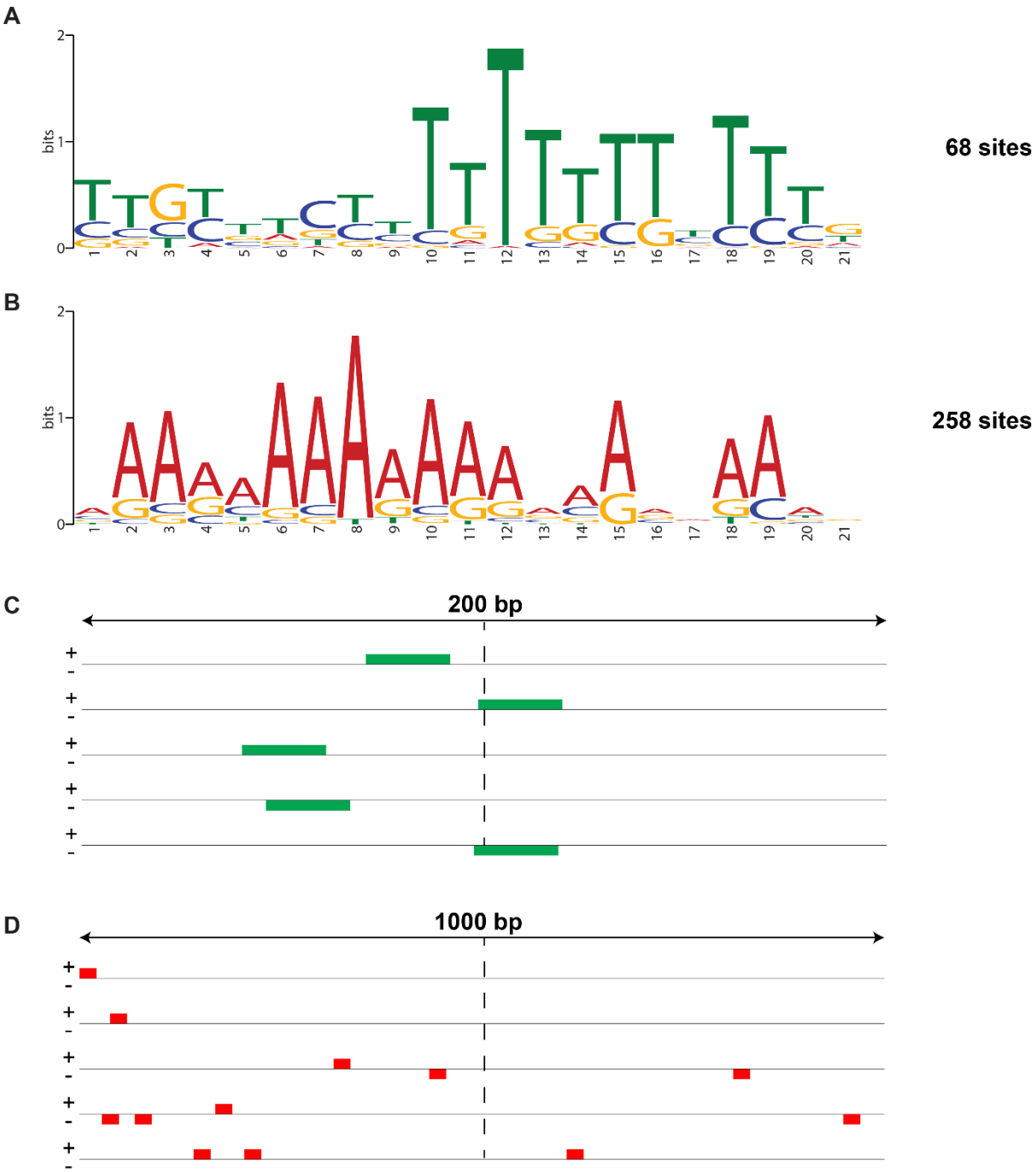


Figure 4.8 Chromo1 binding motif search.

(A) Best-scoring motif generated using a 200 bp window around Chromo1 peaks.

(B) Best-scoring motif generated using a 1000 bp window around Chromo1 peaks.

(C) and (D) Rectangles show the positions of the motifs from A and B, respectively, for five Chromo1 peaks. The dashed lines in the middle of the 200 bp and 1000 bp windows indicate the position of Chromo1 peak summits.

4.5 Discussion

ChIP-seq analysis of the nuclear and nuclear and cytoplasmic candidate proteins as well as Ago1 showed that 8 of them interact with chromatin. These included the putative histone methyl mark readers Chromo1 and Znf-CW1, acetyl mark readers BDF2 and BDF4, methyl mark writers SET26 and SET27 and acetyl mark erasers HDAC1 and HDAC3. Surprisingly, these proteins were always found together at RNAPII TSRs and at no other locations.

Chromo1, SET27 and BDF4 always displayed sharp peak profiles whereas the other proteins were consistently broadly enriched over all RNAPII transcription start regions. These distinct groupings suggest that the two sets of candidate proteins may be performing two different functions. Potentially, the proteins with sharp peaks are involved in RNAPII recruitment at specific genomic loci whereas the other proteins may play a role in transcription initiation or the early stages of transcription elongation. Evidence supporting the latter possibility came from analysis of available RNA-seq data which showed that the direction of enrichment of the candidate proteins with broad peaks is correlated with the direction of RNAPII transcription.

While H2AZ and H2BV histones as well as the H3K4me3 mark are distributed broadly throughout TSR regions, some of the putative chromatin regulators had sharp peaks in these locations, suggesting that these proteins might be recruited via DNA sequence motifs rather than histone binding. The search for a Chromo1 binding motif yielded a GT-rich and an A-rich sequence depending on the window length used. However, these sequences were distributed randomly and only under some Chromo1 peaks as opposed to being concentrated under all Chromo1 summits. These observations suggest that there is a not a real DNA motif but rather a preference for certain nucleotide sequences where Chromo1 binds.

Overall, in this Chapter, the list of candidate proteins potentially involved in trypanosome chromatin organisation was narrowed down further to 8 proteins, none of which were found to interact with silent genomic regions.

CHAPTER 5

Identification of protein interaction networks of the TSR-associated factors

5.1 Introduction

In Chapter 4, eight proteins were found to associate with RNAPII TSRs, however their role in RNAPII-directed transcription remains to be determined. The two distinct ChIP-seq profiles observed (sharp and broad) suggested that these proteins are performing at least two different functions, possibly related to transcription initiation and elongation. It is also possible that these TSR-associated factors might act as transcriptional activators or repressors. For example, Bromo-domain proteins are normally associated with actively transcribed regions whereas HDACs contribute to transcriptional repression. These two possibilities could not be distinguished based on the ChIP-seq data alone. Thus, the approach taken was affinity purification coupled with mass spectrometry in order to further characterise the candidate proteins and to get an indication of what their functions might be. All protein affinity selections in this Chapter were performed on bloodstream slender parasites.

5.2 Chromo1 and SET27 display strong reciprocal interaction with each other

Affinity purification of YFP-Chromo1 and YFP-SET27 followed by liquid chromatography with tandem mass spectrometry (LC-MS/MS) showed that they interact strongly with each other, with a common set of four uncharacterised proteins (designated in this thesis as UP1-4) and the J biosynthesis protein JBP2 (Figure 5.1). This suggested that Chromo1 and SET27 may be part of the same protein complex which is described in more detail in Chapter 6. Additionally, Chromo1 co-purified with two RNAPII subunits (RPB1 and RPB3), transcription elongation factor s-II (TFIIS2-1) and repressor activator protein 1 (RAP1) which is involved in VSG silencing (Yang et al., 2009a). The SET27 affinity purification also contained TFIIS2-1, the cohesin subunit SMC3 (Bessat and Ersfeld, 2009a) and JBP1 which, in addition to JBP2, is involved in the synthesis of base J. The absence of significantly enriched histones in these preparations suggests that the interaction between Chromo1, SET27 and the other five putative complex members is not mediated by chromatin. The full list of significantly enriched proteins in this and other affinity purifications in this project can be found in Appendix D.

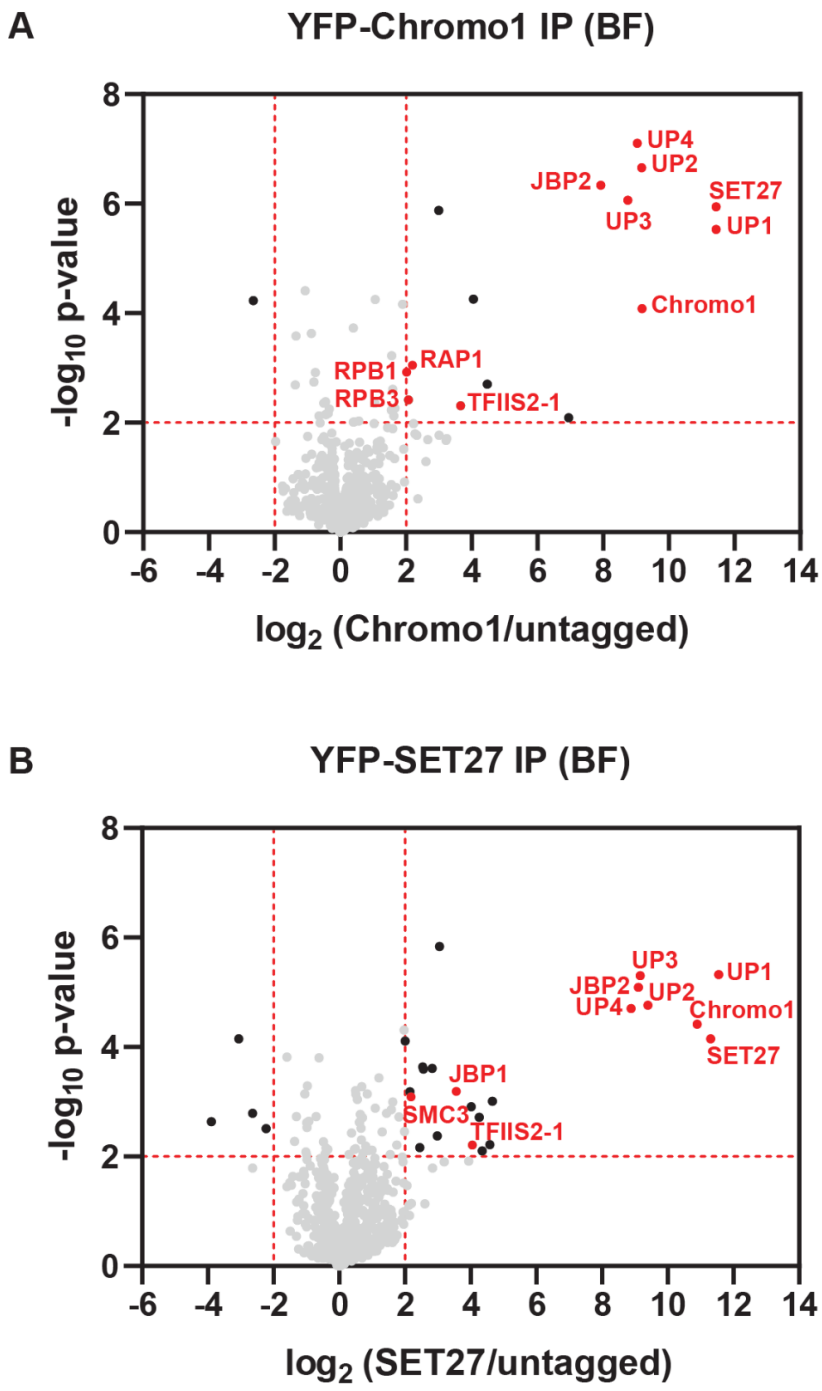


Figure 5.1 Chromo1 and SET27 co-purify each other in bloodstream form parasites.

(A) YFP-Chromo1 IP

(B) YFP-SET27 IP

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

5.3 Znf-CW1 and SET26 display strong reciprocal interaction with each other

Affinity purification of YFP-Znf-CW1 and YFP-SET26 followed by LC-MS/MS showed that these proteins interact strongly with each other. The presence of many histones and histone variants in these preparations indicates that this protein interaction might be indirect and mediated by chromatin. The FACT complex subunits spt16 and POB3 as well as the Bromo-domain factor BDF7 were also found in the two preparations. Additionally, both affinity purifications contained importins and the nucleoporin NUP59.

Actin and actin-related proteins, ruv-B and SWI/SNF helicases, BDF2 and BDF5 co-purified with Znf-CW1 only whereas three RNAPII subunits (RPB1, RPB2 and RPB3) were significantly enriched only in the SET26 affinity selection.

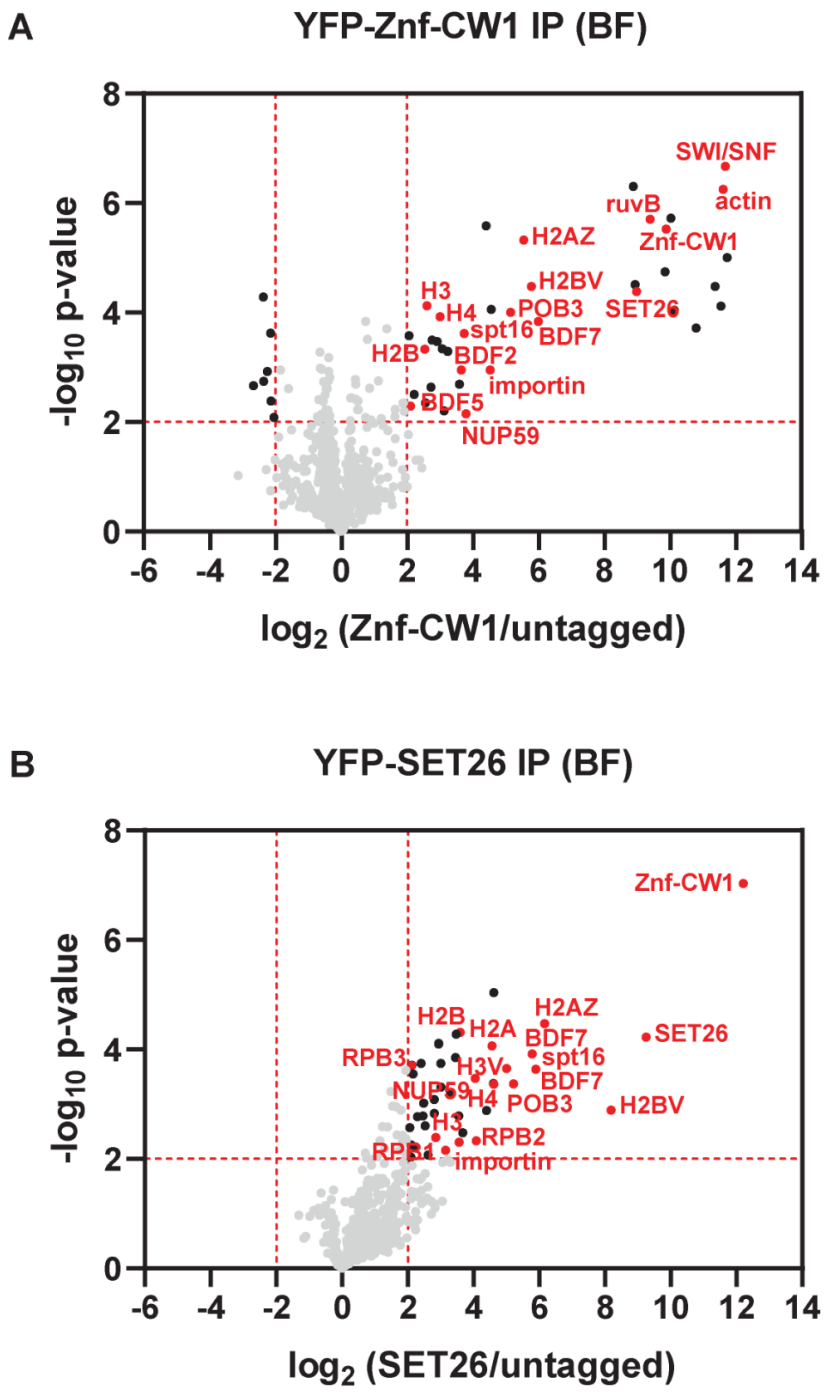


Figure 5.2 Znf-CW1 and SET26 co-purify each other in bloodstream form parasites.

(A) YFP-Znf-CW1 IP

(B) YFP-SET26 IP

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

5.4 BDF2 and HDAC3 display strong reciprocal interaction with each other

Affinity purification of YFP-BDF2 and YFP-HDAC3 coupled with LC-MS/MS showed that they interact strongly with each other. The telomere factors TRF, TRF-interacting factor 2 (TIF2) and telomere-associated protein 1 (TelAP1) as well as the TSR-associated histone variants H2AZ and H2BV and the microhomology-mediated end-joining DNA polymerase theta (POLQ) were enriched in both affinity purifications.

Two class I transcription factors A (CITFA-1 and CITFA-4) co-purified only with BDF2 whereas RAP1 and the cohesin subunits SMC1 and SMC3 were found only in the HDAC3 preparation. The BDF2 interaction with Znf-CW1 seen in section 5.3 was also confirmed and could explain the presence of actin, ruvB and SWI/SNF helicases in the BDF2 preparation.

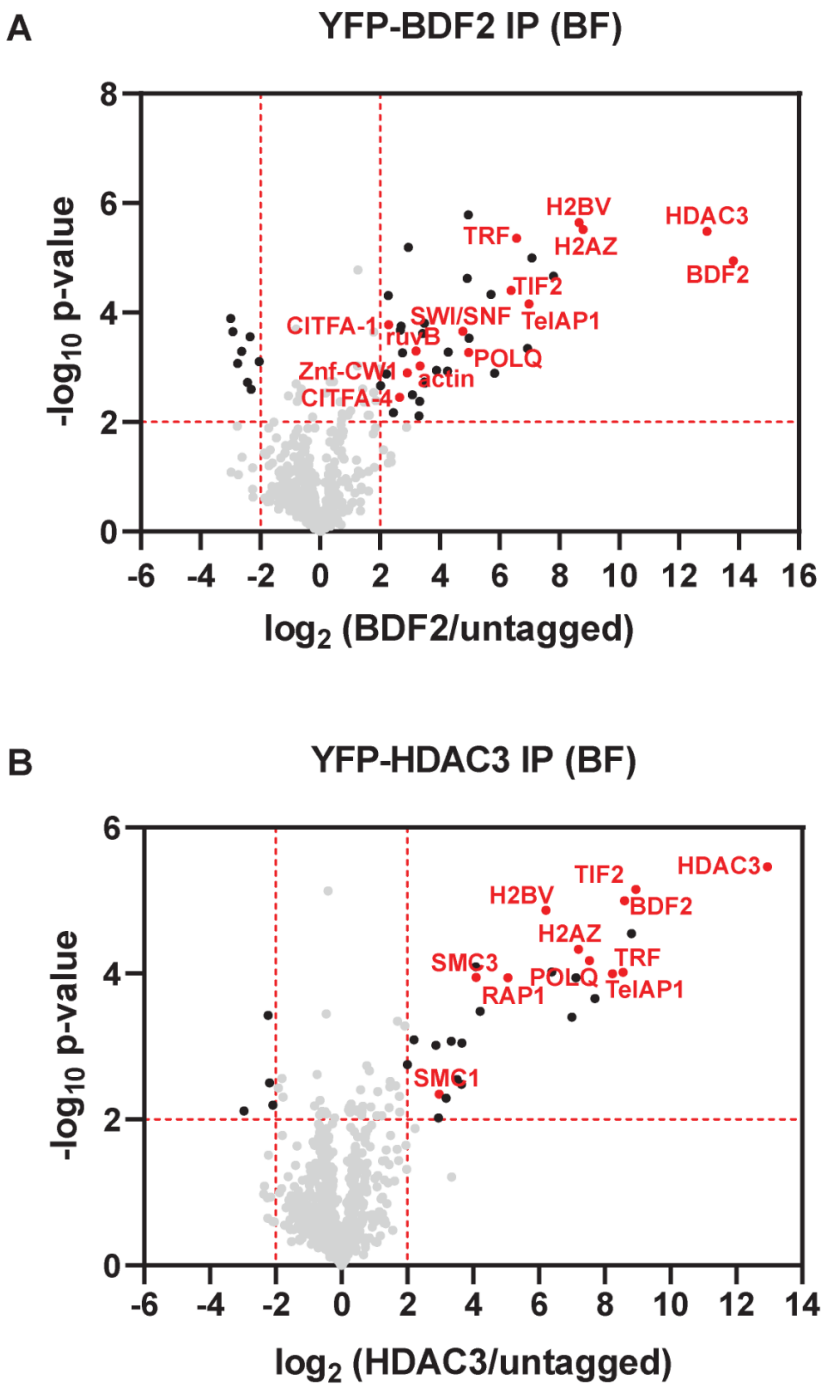


Figure 5.3 BDF2 and HDAC3 co-purify each other in bloodstream form parasites.

(A) YFP-BDF2 IP

(B) YFP-HDAC3 IP

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

5.5 BDF4 and BDF1 display strong reciprocal interaction with each other

Affinity purification of BDF4 coupled with LC-MS/MS showed that BDF4 interacts strongly with BDF1 and to a lesser degree with two other Bromo-domain proteins (BDF3 and BDF5).

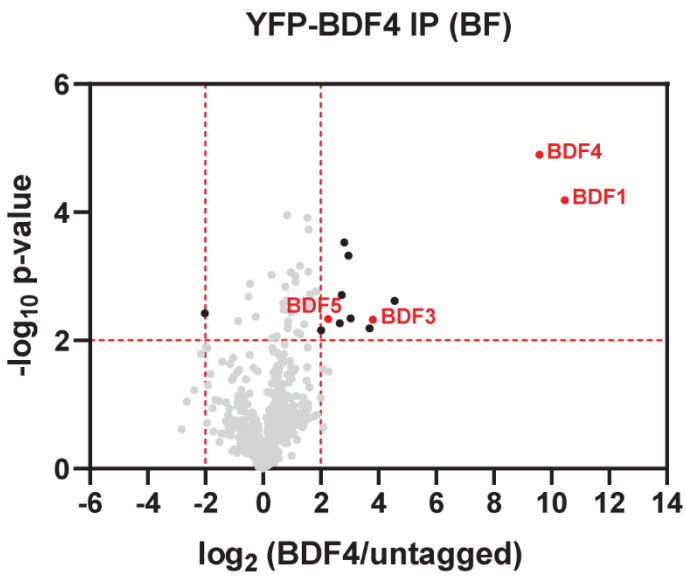


Figure 5.4 YFP-BDF4 IP.

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

5.6 HDAC1 does not interact with other candidate proteins

No other candidate proteins co-purified with YFP-HDAC1. Significantly enriched in this preparation were four uncharacterised proteins and a valosin-containing protein homolog.

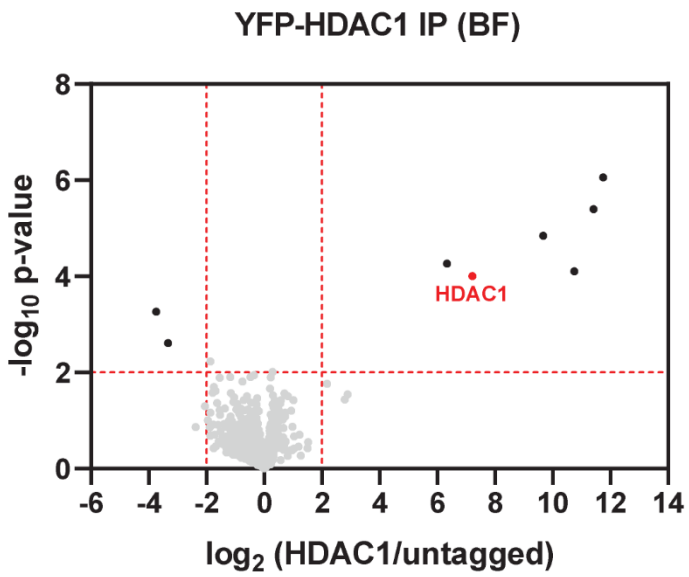


Figure 5.5 YFP-HDAC1 IP.

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

5.7 Discussion

Chromo1 and SET27 were found to strongly associate with each other and are likely part of the same protein complex (described in more detail in Chapter 6) together with JBP2 and 4 uncharacterised proteins. The presence of JBP2 was surprising because it is involved in the synthesis of base J which is found at telomeres and TTRs as opposed to transcription start regions (Schulz et al., 2016).

Znf-CW1 and SET26 were also found to associate strongly with each other. The presence of many histones and histone variants in these affinity purifications indicates that this protein-protein interaction, unlike the Chromo1-SET27 interaction, may be mediated by chromatin. Additionally, the enrichment of importins and the nucleoporin NUP59 in these preparations suggests possible association with the nuclear transport machinery. BDF7 as well as the two subunits of the FACT complex, POB3 and spt16, also co-purified with both Znf-CW1 and SET26. The FACT complex is known to facilitate RNAPII transcription elongation by acting as a histone chaperone that destabilises nucleosome structure in front of the transcribing polymerase and restores it behind (Belotserkovskaya, 2003; Formosa et al., 2002; Saunders, 2003). SET26 but not Znf-CW1 was found to associate with three RNAPII subunits whereas Znf-CW1 but not SET26 co-purified with Bromo-domain proteins BDF2 and BDF5, ruvB and SWI/SNF helicases as well as actin and actin-related proteins. In budding yeast, the SWI/SNF chromatin remodeller has been described as a transcriptional activator that works by destabilising nucleosomes and thus promoting an open chromatin structure (Hirschhorn et al., 1992). Actin and actin-related proteins have been identified as subunits of the SWI/SNF complex (Hargreaves and Crabtree, 2011). RuvB helicase is an ATPase that is involved in homologous recombination in bacteria (West, 2003). In eukaryotes, ruvB has been described as a chromatin decondenser (Magalska et al., 2014) and is also known as part of the SWR1 and INO80 chromatin remodeller complexes both of which also contain actin-related proteins (Bao and Shen, 2011). SWR1 facilitates the exchange of canonical H2A for the less stable

H2AZ histones found at TSRs and can thus act as a transcriptional activator whereas INO80 can regulate transcription positively or negatively. These data indicate that Znf-CW1 and SET26 interact with and possibly target transcriptional activators and repressors to TSR regions.

BDF2 and HDAC3 were found to associate strongly with each other as well as with three telomeric factors, consistent with the role of HDAC3 in silencing of ES-VSGs which are located in subtelomeric regions on the megabase and intermediate chromosomes (Wang et al., 2010). Additionally, HDAC3 co-purified with RAP1 which is also known to be an ES silencing factor (Yang et al., 2009b). The TSR-associated histone variants H2AZ and H2BV were also enriched in both preparations, indicating that the BDF2-HDAC3 interaction may be indirect. The BDF2 affinity purification confirmed the interaction with Znf-CW1 and its associated proteins (SWI/SNF, ruvB and actin). Interestingly, BDF2 also co-purified with two RNAPI transcription factors (CITFA-1 and CITFA-4) (Brandenburg et al., 2007), providing further evidence for the association of HDAC3 and BDF2 with VSG expression.

Finally, BDF4 was found to associate strongly with BDF1 but less strongly with BDF3 and BDF5 whereas HDAC1 did not show interactions with other candidate proteins.

Overall, the affinity selection experiments described in this Chapter showed that although 8 candidate proteins were enriched at RNAPII TSRs, they form distinct protein interaction networks. These networks gave indication of what putative complexes these TSR factors formed and what their functions might be.

CHAPTER 6

Characterisation of the putative Chromo1/SET27 complex in bloodstream and procyclic form parasites

6.1 Introduction

The results of the affinity selection experiments presented in Chapter 5 suggested that the association between Chromo1 and SET27 may occur in the context of a protein complex. The reason for this is two-fold. First, in both YFP-Chromo1 and YFP-SET27 purifications the same 7 proteins were most highly enriched in all of the replicate experiments, and second, the absence of histones in these preparations indicates that the interactions between the 7 proteins are not mediated by chromatin. Additionally, the ChIP-seq data from Chapter 4 showed that Chromo1 and SET27 had the same sharp peak profiles at TSRs in contrast to most of the candidate proteins which were broadly enriched in these regions. However, BDF4, which also displayed a very focused enrichment at TSRs, did not co-purify with Chromo1 and SET27.

The aim of the experiments presented in this Chapter was to characterise the composition of the putative protein complex in BF and PF parasites and to gain an insight into its function, particularly with respect to RNAPII transcription regulation.

6.2 Domain composition, cellular localisation and genomic association of the putative complex members

Chromo1 and SET27 were part of the initial screen for putative readers, writers and erasers of trypanosome histone PTMs because of their predicted protein domains (Chapter 3). Homology models of these domains by Dr Jeyaprakash Arulanandam showed that the single Chromo domain of Chromo1 is lacking three key aromatic residues important for binding of histone methyl marks (Figure 6.1A) whereas the single SET domain of SET27 is very similar to the SET domain found in a human H3K4 monomethyltransferase (Figure 6.1B). Analysis of the amino acid sequence of the uncharacterised proteins showed that UP3 has a weakly predicted Chromo domain whereas UP4 has a weakly predicted FYVE/PHD zinc finger domain, both domains being found in histone methyl mark readers. JBP2 was found to contain a DEAD-box helicase domain. The analysis did not yield information about protein domains in UP1 and UP2.

UP1-4 and JBP2 were YFP-tagged and immunolocalised in BF cells (Figure 6.2) utilising the same strategy as in the initial screen (Chapter 3). UP1-4 were found to be nuclear with dotted appearance whereas JBP2 was both nuclear and cytoplasmic.

To assess whether the putative protein complex has the same composition during the insect stage of *T. brucei*'s life cycle, Chromo1, SET27 and JBP2 were also YFP-tagged in the procyclic form. JBP2 was expressed at approximately the same level in both BF and PF cells (Appendix C) in contrast to previous studies in which JBP2 was found to be absent in PF parasites (Borst and Sabatini, 2008). Microscopy imaging without antibody staining showed that in the PF Chromo1 and SET27 are nuclear whereas JBP2 is cytoplasmic (Figure 6.3). In the TrypTag project, UP1 was found to be mitochondrial when tagged N-terminally and nuclear when tagged C-terminally whereas UP4 localised to the nucleus when tagged N- or C-terminally.

ChIP-seq of Chromo1 tagged in the procyclic form showed sharp peaks at TSR regions, analogously to the results obtained in the bloodstream form (Figure 6.4). None of the 7 proteins were found to be essential in an RNAi screen (Alsford et al., 2011). Thus, in this project it was attempted to generate Chromo1 and SET27 knockouts in BF cells via CRISPR. However only a single allele of each protein could be knocked out, suggesting that Chromo1 and SET27 are essential in the bloodstream form. Table 6.1 summarises the available information about the 7 putative complex members.

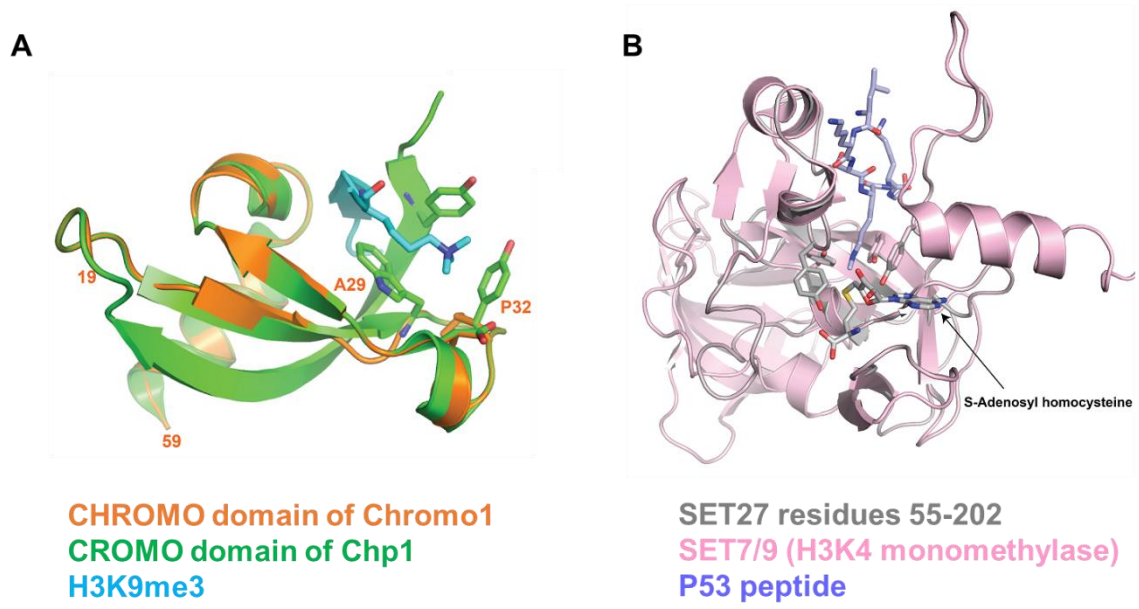


Figure 6.1 Homology models of Chromo1 and SET27 domains.

(A) The Chromo domain of Chromo1 (orange) was modelled to the Chromo domain of Chp1 (green). Note that the *T. brucei* Chromo1 lacks the three aromatic residues necessary for binding of methylated lysines.

(B) The SET domain of SET27 was modelled to SET7/9 (H3K4 monomethyltransferase). Homology models by Dr Jeyaprakash Arulanandam.

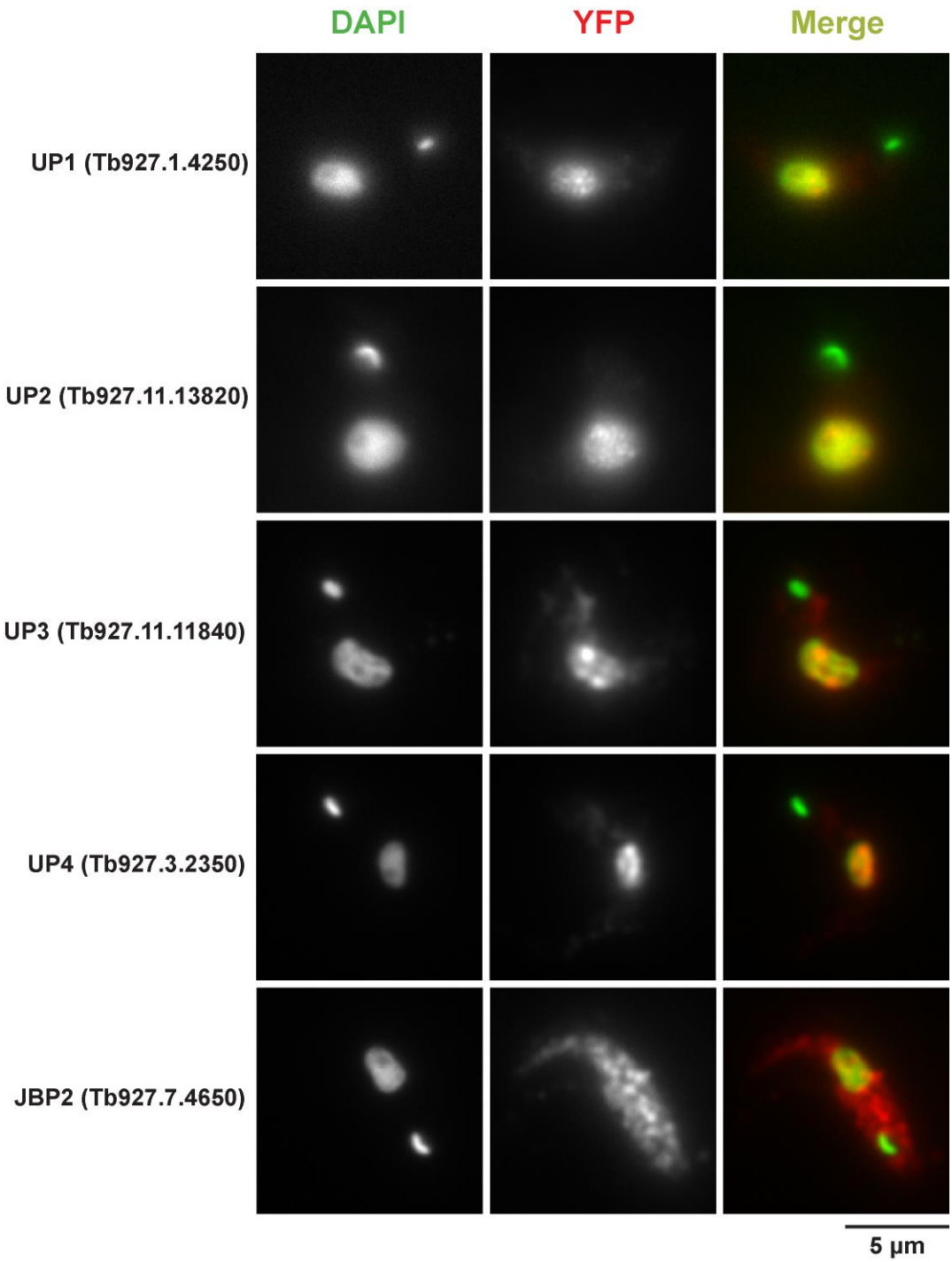


Figure 6.2 Immunolocalisation of UP1-4 and JBP2 in BF parasites. Immunolocalisation was performed using an anti-GFP primary and a red fluorescently labelled secondary antibody. Nuclear and kinetoplast DNA were stained with DAPI.

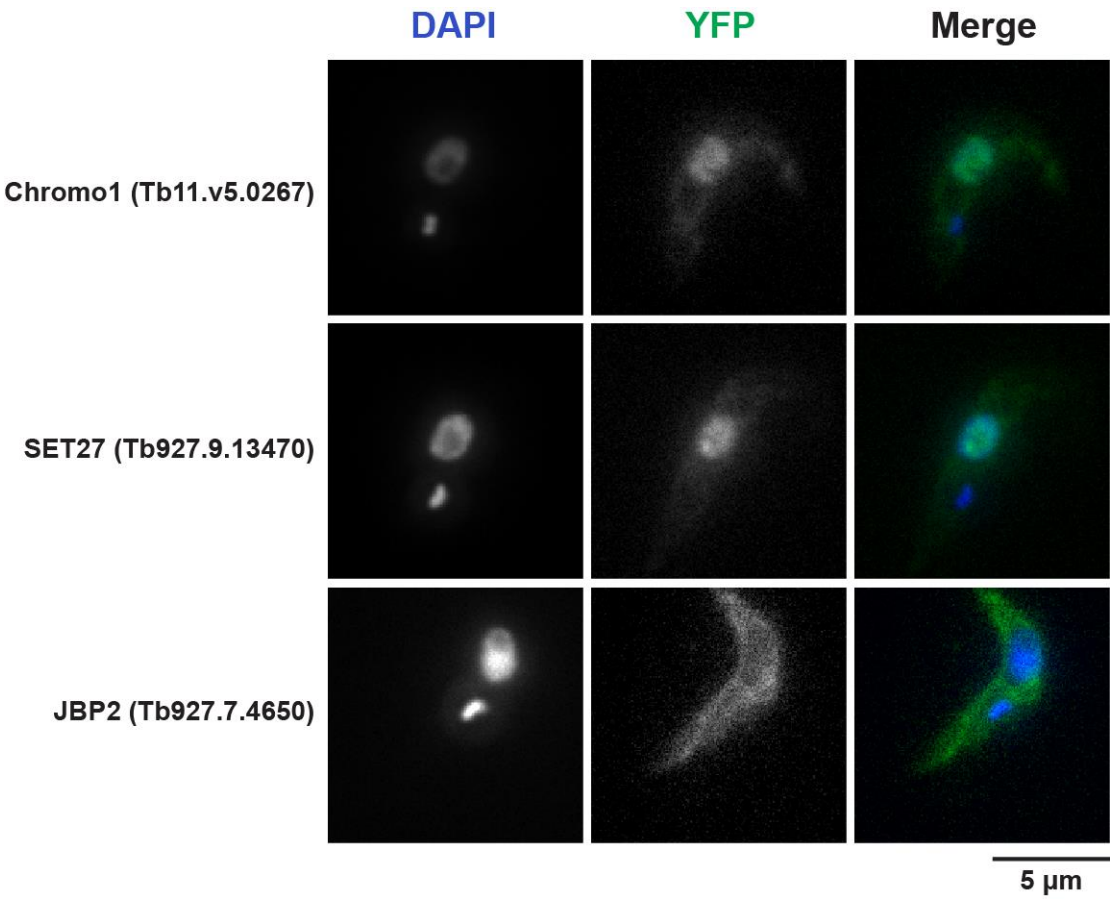


Figure 6.3 Localisation of Chromo1, SET27 and JBP2 in PF parasites. Proteins were localised via direct visualisation of the YFP tag without antibody staining. Nuclear and kinetoplast DNA were stained with DAPI.

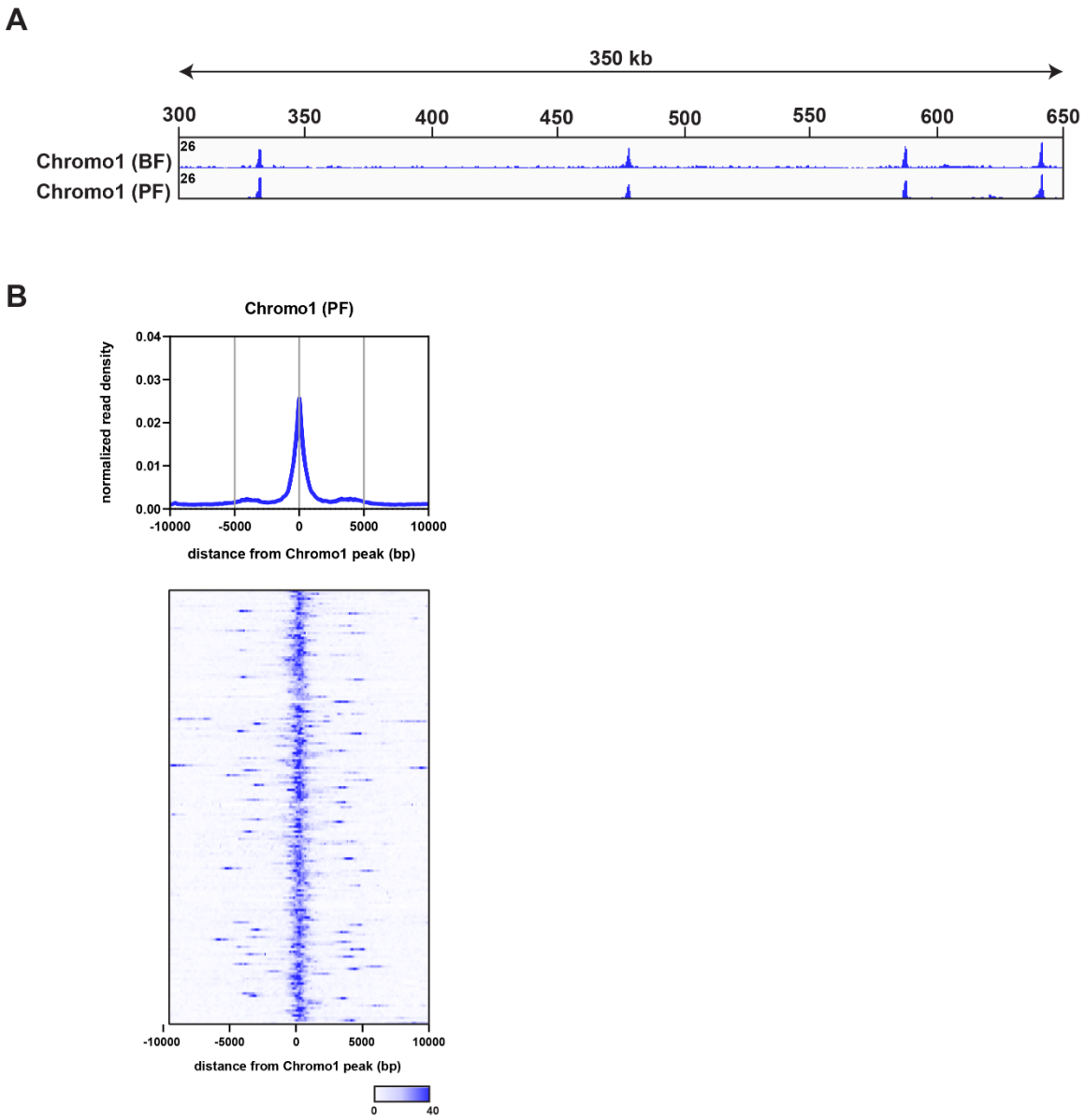


Figure 6.4 Chromo1 enrichment at RNAPII TSRs in BF and PF parasites.

(A) BF and PF Chromo1 enrichment over a 350 kb region on chromosome 1. Each track is scaled separately to the highest peak summit (shown as a number in the top left corner of each track) in order to demonstrate the peak shapes better.

(B) Metagene plot and heatmap plots showing Chromo1 enrichment in the PF.

Table 6.1 Summary of available information on the putative Chromo1/SET27 complex members

Protein name	Gene ID	Domains	Localisation in this study (BF)	Localisation in this study (PF)	TrypTag[1] localisation (PF)	RNAi phenotype (TriTrypDB[2])
Chromo1	Tb11.v5.0267	Chromo	nucleus	nucleus	-	growth defect
SET27	Tb927.9.13470	SET	nucleus and cytoplasm	nucleus	-	no growth defect
UP1	Tb927.1.4250	-	nucleus	not tagged in PF	N-terminal tag: mitochondrion; C-terminal tag: nucleoplasm	abnormal cell cycle
UP2	Tb927.11.13820	-	nucleus	not tagged in PF	-	mild growth defect
UP3	Tb927.11.11840	Chromo	nucleus	not tagged in PF	-	no growth defect
UP4	Tb927.3.2350	FYVE/PHD zinc finger	nucleus	not tagged in PF	N-terminal tag: nucleoplasm; C-terminal tag: nucleoplasm	growth defect
JBP2	Tb927.7.4650	DEAD-box helicase	nucleus and cytoplasm	cytoplasm	-	mild growth defect

[1], [2] - the data from www.tryptag.org and www.tritrypdb.org was accessed on 08 January 2020.

6.3 Complex composition in the bloodstream and procyclic form

Affinity selection of the tagged putative complex members in the bloodstream form showed that they all co-purify with each other. Additionally, they were the most enriched proteins in 5 out of the 7 affinity selections, as shown by their clustering in the top right corner of volcano plots (Figure 6.5).

UP2 and UP4 preparations had similar composition – both proteins co-purified with JBP1, NUP59, FACT complex subunits (spt16 and POB3), RNAPII subunits (RPBs), cohesin subunits (SMC1 and SMC3) as well as multiple histones and kinetochore proteins (KKTs).

UP1 co-purified with NUP59 as well as with a single histone (H4V) and a single kinetochore protein (KKT3), both of which were highly enriched in the UP2 and UP4 experiments.

A nucleolus protein (Tb927.10.3220) was enriched in the Chromo1, SET27, UP3 and JBP2 purifications. Additionally, one uncharacterised protein (Tb927.10.3070) was found to associate with SET27, UP3 and JBP2 whereas a second uncharacterised protein (Tb927.10.4800) co-purified with Chromo1, SET27, UP1, UP2 and UP3.

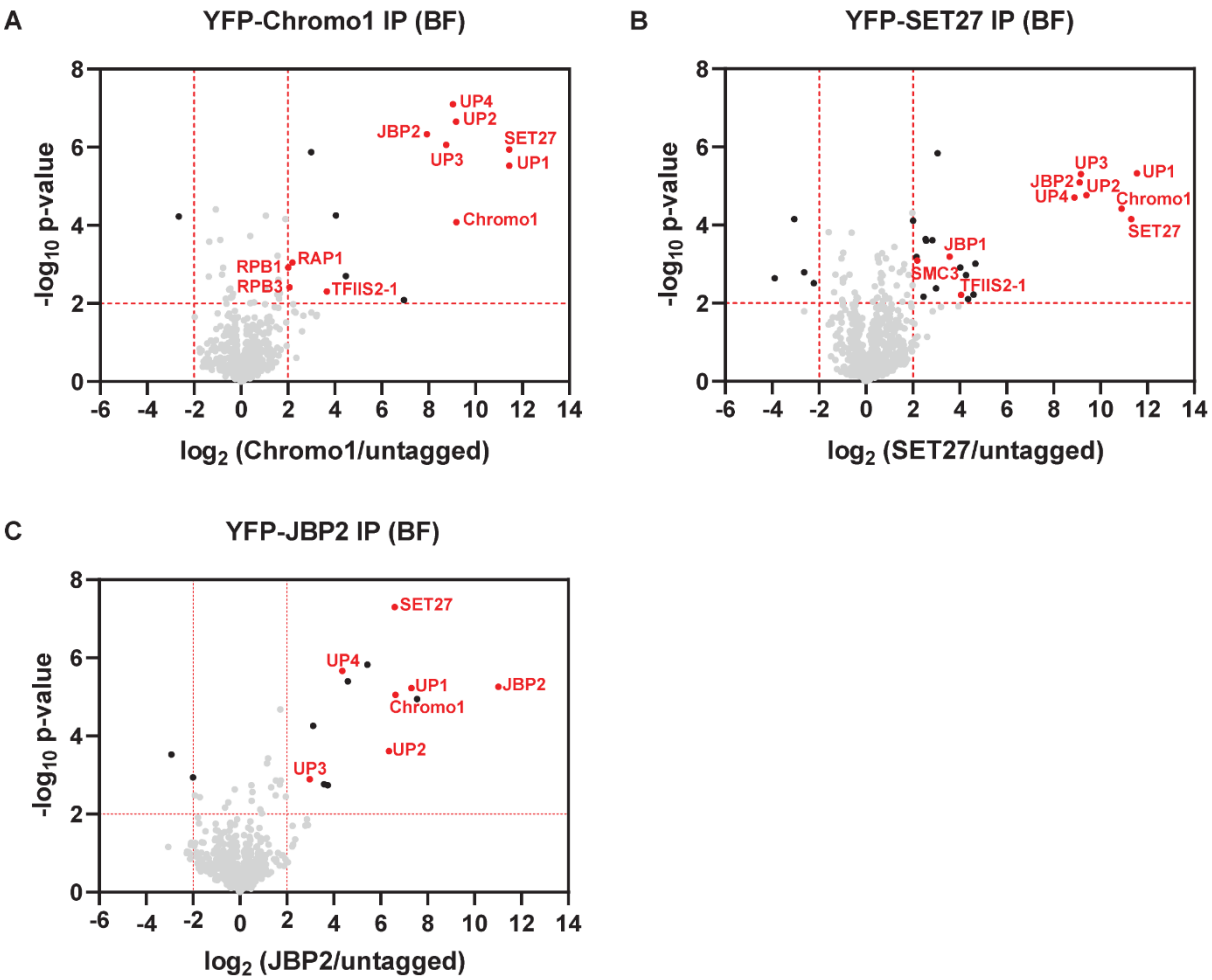


Figure 6.5 Affinity purification of the putative complex members in BF parasites.

(A) YFP-Chromo1 IP (panel copied from Figure 5.1)

(B) YFP-SET27 IP (panel copied from Figure 5.1)

(C) YFP-JBP2 IP

(D) YFP-UP1 IP

(E) YFP-UP2 IP

(F) YFP-UP3 IP

(G) YFP-UP4 IP

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

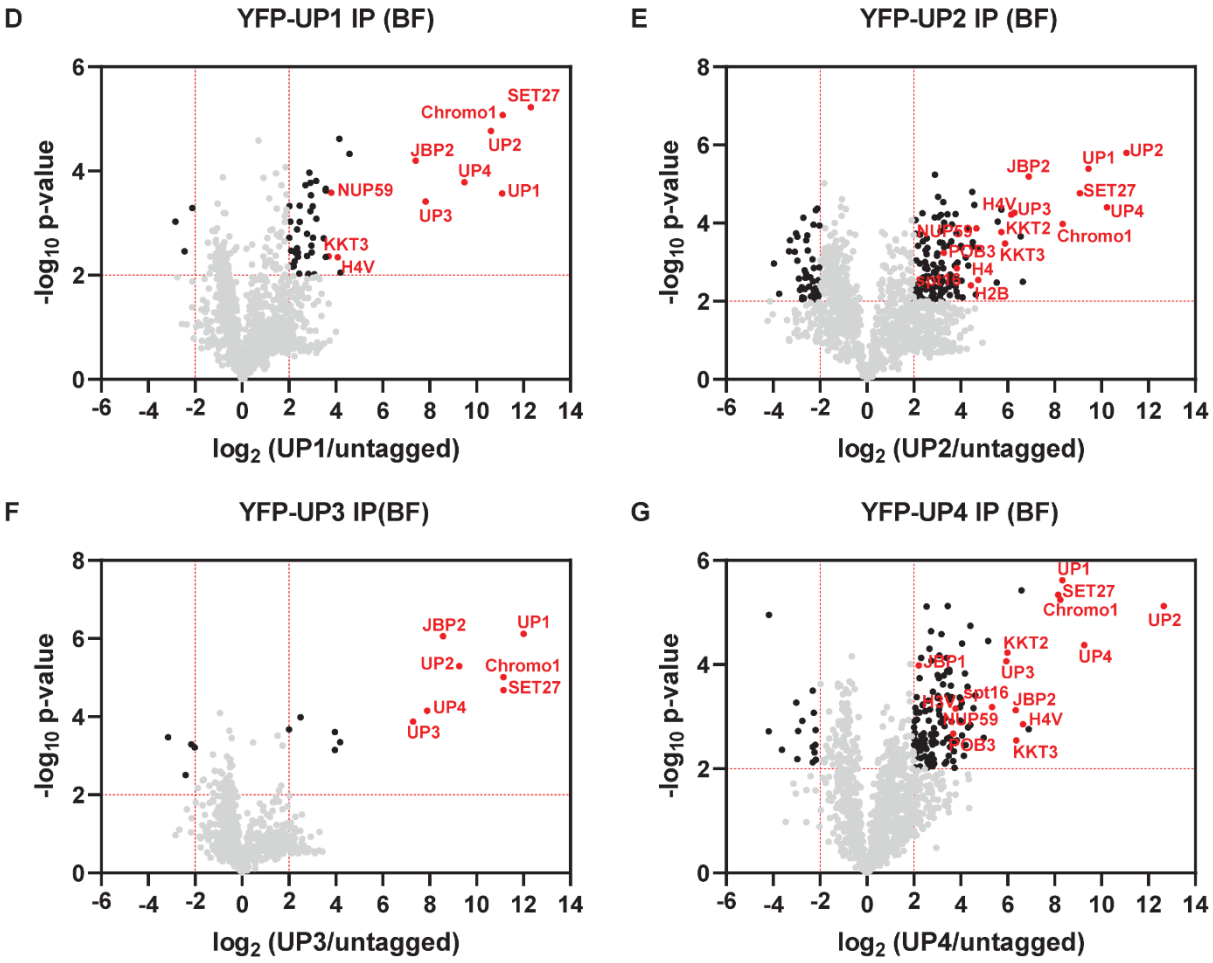


Figure 6.5 continued.

YFP-Chromo1, YFP-SET27 and YFP-JBP2 were affinity purified from PF cells in order to identify their protein interactors at this developmental stage of *T. brucei* (Figure 6.6). Chromo1 co-purified with all putative complex members except JBP2. In contrast, Chromo1, along with SET27, UP1 and UP2, was significantly enriched in the JBP2 experiment. UP3 and UP4 were completely absent from the JBP2 preparation (they were not identified even among the nonsignificant proteins). SET27 was found to associate with all the other putative complex members, with JBP2 being significantly less enriched than the rest. Additionally, Chromo1 and SET27 co-purified with the RNAPII-associated protein LEO1 and the kinetochore-interacting protein KKIP4. SET27 also co-purified with the ISWI chromatin remodeller and the cohesin subunit SMC1.

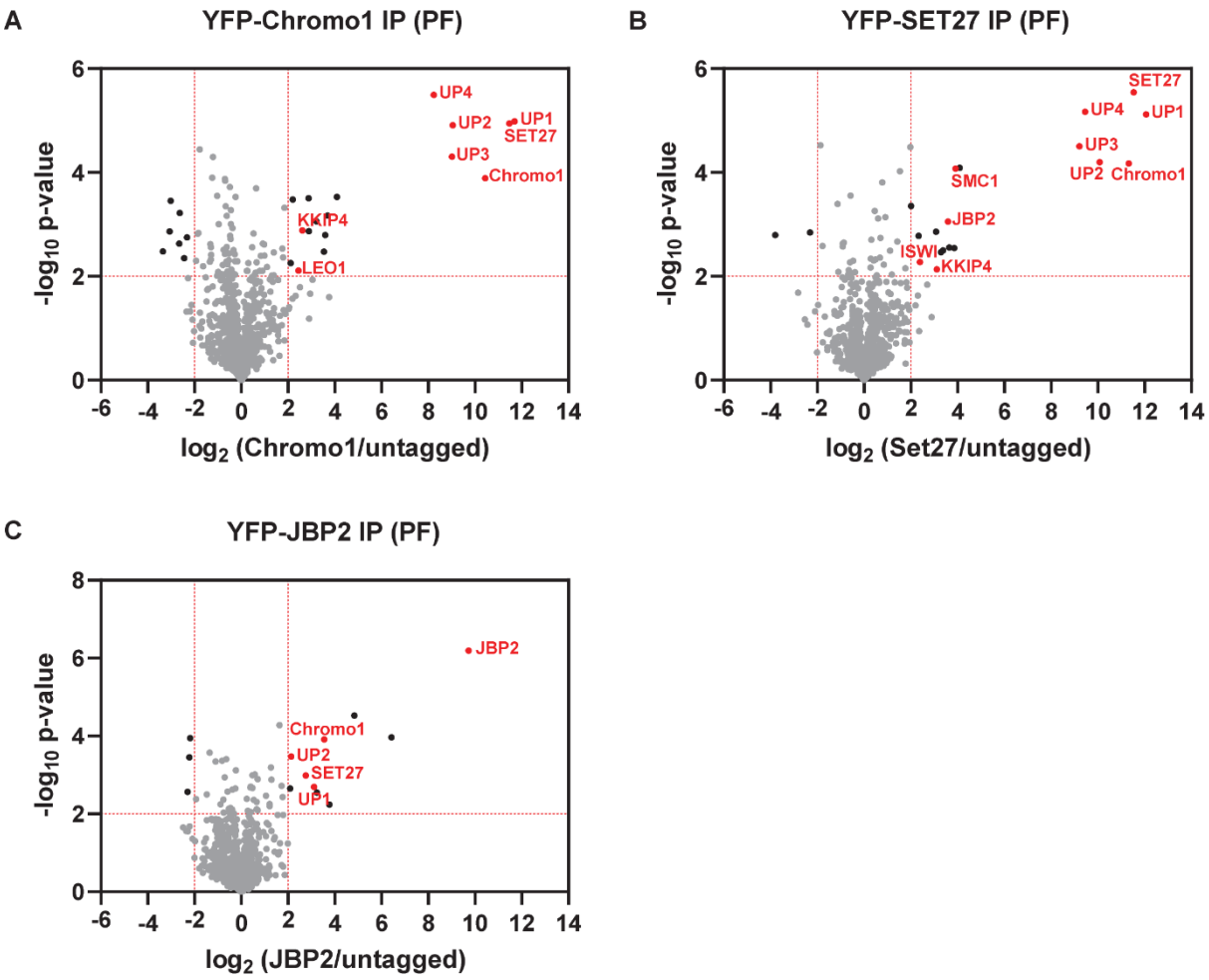


Figure 6.6 Affinity purification of the putative complex members in PF parasites.

(A) YFP-Chromo1 IP

(B) YFP-SET27 IP

(C) YFP-JBP2 IP

Cut-offs used for significance: $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2. Significantly enriched proteins of interest are marked in red. Data for each plot is based on three biological replicates.

6.4 Discussion

Mass spectrometry data presented in this Chapter provided further evidence supporting the idea that Chromo1 and SET27 form a complex together with UP1-4 and JBP2 in bloodstream form parasites. The absence of key residues necessary for histone methyl mark binding in the Chromo1 homology model indicates that it might not interact directly with histones. Instead, UP3 and/or UP4 might bind methylated histones owing to a weakly predicted Chromo and FYVE/PHD zinc finger domain, respectively, in their amino acid sequence. Additionally, the presence of multiple histones and histone variants in the UP2 and UP4 affinity selections suggests that these proteins might be the chromatin-interacting members of the complex. The similarity of SET27 to an H3K4 methyltransferase is consistent with the presence of H3K4me3 at trypanosome TSRs (Wright et al., 2010). The enrichment of RNAPII and FACT complex subunits in the UP2 and UP4 affinity purifications indicates potential involvement in RNAPII transcription regulation. Surprisingly, UP2 and UP4 also co-purified with multiple kinetochore proteins, and thus it would be interesting to investigate via ChIP-seq whether UP2 and UP4 associate with centromeric regions which are known to be transcribed in other eukaryotes (Djupedal et al., 2005; Saffery et al., 2003). The YFP-JBP2 affinity selection differed from the others in that the putative complex members did not cluster together in the top right corner of the volcano plot, suggesting that JBP2 may be a facultative subunit of this complex.

Mass spectrometry data from procyclic form parasites showed that JBP2 is weakly associated with only 4 of the other complex members. This is consistent with the changes in protein localisation observed: in the bloodstream form JBP2 and SET27 are nuclear and cytoplasmic whereas in the procyclic form SET27 is exclusively nuclear and JBP2 is found only in the cytoplasm. Although the putative complex has different composition in BF and PF cells, its function might be conserved during trypanosome development, as indicated by the identical ChIP-seq profile of Chromo1 at different stages of *T. brucei*'s life cycle.

CHAPTER 7

Discussion

The ability to survive various environmental conditions, progress through development and, in the case of multicellular organisms, form tissues, organs and systems specialised to perform different functions relies on precise regulation of gene expression. In eukaryotes, DNA exists in the context of chromatin, the latter functioning to make DNA more or less accessible to the transcription machinery, thus providing the first level of gene expression control. Gene activation and repression are influenced by largely conserved chemical modifications on DNA and histone proteins. However, as one of the earliest-branching eukaryotes, *T. brucei* represents a divergent system for the study of gene expression. Thus, by identifying the similarities and differences in the mechanisms for gene expression control between *T. brucei* and commonly studied model organisms within the Opisthokonta supergroup, both gene regulation principles fundamental to eukaryotes and specific to a particular supergroup could be uncovered. This study aimed to characterise proteins potentially involved in the deposition, binding or removal of trypanosome histone acetyl and methyl PTMs which are known to be important for chromatin organisation in other eukaryotes. However, it should be noted that the regulation of trypanosome chromatin may depend on different histone modifications which have not been investigated in this study.

In the first stage of the project, 34 candidate proteins were fluorescently tagged and immunolocalised in bloodstream form parasites. The localisations observed were largely consistent with data on procyclic cells from the TrypTag project (Dean et al., 2017). One surprising finding from the screen was that most of the 30 SET-domain candidate proteins were exclusively cytoplasmic, suggesting that they might methylate non-histone substrates, such as tubulin which is known to be modified by a histone methyltransferase that links chromatin and cytoskeletal organisation (Park et al., 2016). Alternatively, the cytoplasmic

SET-domain proteins might be secreted and act on host molecules, reminiscent of the mechanism by which many pathogens modulate their host activity, including chromatin organisation and gene expression (Silmon De Monerri and Kim, 2014).

Chromatin immunoprecipitation and sequencing of the nuclear and nuclear and cytoplasmic candidate proteins showed that 8 of them associate with RNAPII TSRs and none are enriched over other genomic regions. Analysis of the ChIP-seq profiles with respect to Chromo1 showed that Chromo1, SET27 and BDF4 have focused enrichment at all TSRs whereas SET26, BDF2, Znf-CW1, HDAC1 and HDAC3 are consistently broadly distributed around Chromo1. This observation suggests that these proteins perform at least two different functions at TSRs, possibly related to transcription initiation or elongation. The distribution of the broadly enriched proteins was correlated with the direction of RNAPII transcription, suggesting that they might travel for a short distance with the elongating polymerase before falling off. Among the identified TSR factors, some are expected to activate and others – to repress transcription. For example, BDF2 and BDF4 are putative Bromo-domain readers of acetylated histones, the latter being loosely associated with DNA and consequently found at transcriptionally active regions (Hebbes et al., 1988; Ridsdale et al., 1990). Conversely, HDAC1 and HDAC3 are putative erasers of histone acetyl modifications and are therefore likely to be associated with transcriptionally repressive chromatin. Since acetylation physically disrupts the electrostatic attraction between DNA and histones, the roles of BDFs and HDACs in chromatin organisation are likely to be conserved between *T. brucei* and other eukaryotes. Although the 8 TSR-enriched proteins were always found together, it cannot be concluded that this is true for individual parasites based on the ChIP-seq data obtained from cell populations. Exploring the possibility that different cells have a different combination of the candidate proteins in association with different TSRs would require optimisation of single-cell ChIP-seq methodology in *T. brucei*.

The effect of the candidate proteins on transcription could be tested by measuring the level of newly synthesised transcripts upon depleting or knocking out the TSR-associated factors. Nascent RNA-seq as opposed to steady-state RNA-seq should be used in order to avoid interference from post-transcriptional gene regulation mechanisms. The results interpretation following such knock out or knock down experiments could be complicated if the candidate proteins are essential, which was found to be the case for Chromo1 and SET27. Additionally, removal of a single protein from TSRs might not have an effect on transcription if there is redundancy of function among the TSR-associated proteins, and combinatorial experiments might be required instead.

Interaction partners of the TSR-associated factors were identified via affinity purifications followed by mass spectrometry. These experiments showed that although many candidate proteins are enriched over the same genomic regions, they do not all interact with each other and are likely part of different protein complexes that perform distinct functions at RNAPII TSRs. The putative complex identified with highest confidence was that of Chromo1 and SET27 with UP1-4 and JBP2. Both Chromo1 and SET27 displayed sharp peak profiles at TSRs in contrast to the other candidate proteins, and it would be interesting to investigate whether that is also the case for UP1-4 and JBP2. However, BDF4, which also had focused enrichment at TSRs, was not identified as an interacting partner of Chromo1 and SET27.

Two outstanding questions are how the putative Chromo1/SET27 complex is recruited to TSR regions and what its function is, particularly in relation to RNAPII transcription. No DNA sequence binding motif was identified for Chromo1, the location of which is coincident with SET27, so it is unlikely that the putative complex is recruited to TSRs via interaction with DNA. The other two possibilities are recruitment via binding to histones or other TSR-associated proteins. Chromo1, UP3 and UP4 have weakly predicted methyl mark binding domains in their amino acid sequences whereas the SET domain of SET27 was found to bear similarity to the SET domain of an H3K4 monomethyltransferase. These observations indicate that the

putative complex might contain several readers and a writer of a histone methyl PTM, such as the H3K4me3 mark which, analogously to other eukaryotes, is found at trypanosome TSRs (Wright et al., 2010). The SET domain of SET27 contains a tyrosine residue predicted to exert steric hindrance that would prevent the addition of more than one methyl group on histone lysines. Thus, if SET27 indeed methylates H3K4, it is unclear which methyltransferase(s) write(s) the di- and tri- methyl marks on that residue. *In vitro* methyltransferase assays using recombinant SET27 and synthetic H3K4me0, H3K4me1 and H3K4me2 peptides performed by Dr Sharon White gave a negative result (none of the peptide residues were methylated by SET27). However, multiple variables of the *in vitro* assays might require optimisation, including expression of recombinant protein from eukaryotic as opposed to bacterial cells and presence of other members of the complex. A more unbiased approach would be to purify the entire protein complex from trypanosome cells and attempt the *in vitro* methyltransferase assay with recombinant *T. brucei* nucleosomes. However, this methodology might be unsuccessful due to insufficient amount and/or insufficient purity of the protein complex.

The function of the putative Chromo1/SET27 complex at RNAPII TSRs is of particular interest because of the apparent dominance of post-transcriptional gene regulation mechanisms in *T. brucei*. This could be investigated by measuring the levels of newly synthesised transcripts in the absence of one or more of the 7 putative complex members. Chromo1 and SET27 were found to be essential which could be because they perform an important function or because their absence disrupts the integrity of the protein complex. Thus, knock downs of Chromo1 and SET27 or knock outs of other members of the complex, provided they are non-essential, could be utilised.

The presence of JBP2 in the TSR-associated complex was surprising because JBP2 is involved in the synthesis of the TTR marker base J. This suggests the interesting possibility that the function of the Chromo1/SET27 complex might be to terminate antisense transcripts, analogously to other eukaryotes where most TSRs are bi-directional but antisense transcripts

originating from those are rapidly degraded (Bresson and Tollervey, 2018). However, many TSRs in *T. brucei* generate stable transcripts in both directions, so it is unclear how the Chromo1/SET27 complex would terminate antisense transcripts from uni- but not from bi-directional TSRs. It is also possible that the function of JBP2 at RNAPII TSRs is independent of its role in base J synthesis.

Overall, the work presented in this thesis identified novel proteins associated with trypanosome RNAPII TSRs and initiated their characterisation. The presence of multiple factors near transcription start sites supports the idea of a complex regulatory landscape where transcriptional control is likely to play an important role alongside post-transcriptional mechanisms in the expression of *T. brucei* genes.

Acknowledgements

I would like to thank my supervisors Prof. Robin Allshire and Prof. Keith Matthews for always being open to discuss ideas throughout this project and for giving me freedom to explore research avenues I found interesting. I would also like to thank members of both labs for help with the experimental procedures. I received financial support for my PhD project from Wellcome.

References

Akiyoshi, B., and Gull, K. (2013). Evolutionary cell biology of chromosome segregation: insights from trypanosomes. Open Biology *3*, 130023.

Akiyoshi, B., and Gull, K. (2014). Discovery of Unconventional Kinetochores in Kinetoplastids. Cell *156*, 1247-1258.

Alarcon, C.M., Son, H.J., Hall, T., and Donelson, J.E. (1994). A monocistronic transcript for a trypanosome variant surface glycoprotein. Molecular and Cellular Biology *14*, 5579-5591.

Allmann, S., and Bringaud, F. (2017). Glycosomes: A comprehensive view of their metabolic roles in T. brucei. The International Journal of Biochemistry & Cell Biology *85*, 85-90.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nature Reviews Molecular Cell Biology *19*, 229-244.

Alsford, S., Kawahara, T., Isamah, C., and Horn, D. (2007). A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. Molecular Microbiology *63*, 724-736.

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 Research *21*, 915-924.

Alsford, S., Wickstead, B., Ersfeld, K., and Gull, K. (2001). Diversity and dynamics of the minichromosomal karyotype in Trypanosoma brucei. Molecular and Biochemical Parasitology *113*, 79-88.

Andrews, S. (2018). FastQC: a quality control tool for high throughput sequence data. Available online at: www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed 02 December 2019].

Bacchi, C., Nathan, H., Hutner, S., McCann, P., and Sjoerdsma, A. (1980). Polyamine metabolism: a potential therapeutic target in trypanosomes. Science *210*, 332-334.

Baker, N., De Koning, H.P., Mäser, P., and Horn, D. (2013). Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. Trends in Parasitology *29,* 110-118.

Bao, Y., and Shen, X. (2011). SnapShot: Chromatin Remodeling: INO80 and SWR1. Cell *144*, p158-158.e152.

Belotserkovskaya, R. (2003). FACT Facilitates Transcription-Dependent Nucleosome Alteration. Science *301*, 1090-1093.

Bernards, A., Michels, P.A.M., Lincke, C.R., and Borst, P. (1983). Growth of chromosome ends in multiplying trypanosomes. Nature *303*, 592-597.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315-326.

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.

Berriman, M., Hall, N., Sheader, K., Bringaud, F., Tiwari, B., Isobe, T., Bowman, S., Corton, C., Clark, L., Cross, G.A.M.*, et al.* (2002). The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. Molecular and Biochemical Parasitology *122*, 131-140.

Bessat, M., and Ersfeld, K. (2009a). Functional characterization of cohesin SMC3 and separase and their roles in the segregation of large and minichromosomes in Trypanosoma brucei. Molecular Microbiology *71*, 1371-1385.

Bessat, M., and Ersfeld, K. (2009b). Functional characterization of cohesin SMC3 and separase and their roles in the segregation of large and minichromosomes inTrypanosoma brucei. Molecular Microbiology *71*, 1371-1385.

Borst, P., and Sabatini, R. (2008). Base J: Discovery, Biosynthesis, and Possible Functions. Annual Review of Microbiology *62*, 235-251.

Brandenburg, J., Schimanski, B., Nogoceke, E., Nguyen, T.N., Padovan, J.C., Chait, B.T., Cross, G.A.M., and Günzl, A. (2007). Multifunctional class I transcription in Trypanosoma brucei depends on a novel protein complex. EMBO J *26*, 4856-4866.

Bresson, S., and Tollervey, D. (2018). Surveillance-ready transcription: nuclear RNA decay as a default fate. Open Biology *8*, 170270.

Briggs, S.D., Xiao, T., Sun, Z.-W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., and Strahl, B.D. (2002). Trans-histone regulatory pathway in chromatin. Nature *418*, 498-498.

Brown, R.C., Evans, D.A., and Vickerman, K. (1973). Changes in oxidative metabolism and ultrastructure accompanying differentiation of the mitochondrion in Trypanosoma brucei. International Journal for Parasitology *3*, 691-704.

Bruce, D. (1895). Preliminary report on the tsetse fly disease or nagana. Zululand (London: Durban, Bennett & Davis).

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

Brun, R., Blum, J., Chappuis, F., and Burri, C. (2010). Human African trypanosomiasis. The Lancet *375*, 148-159.

Bursell, E. (1963). Aspects of the metabolism of amino acids in the tsetse fly, Glossina (Diptera). Journal of Insect Physiology *9*, 439-452.

Buscaino, A., Allshire, R., and Pidoux, A. (2010). Building centromeres: home sweet home or a nomadic existence? Current Opinion in Genetics & Development *20*, 118-126.

Büscher, P., Cecchi, G., Jamonneau, V., and Priotto, G. (2017). Human African trypanosomiasis. The Lancet *390*, 2397-2409.

Caljon, G., Van Reet, N., De Trez, C., Vermeersch, M., Pérez-Morga, D., and Van Den Abbeele, J. (2016). The Dermis as a Delivery Site of Trypanosoma brucei for Tsetse Flies. PLOS Pathogens *12*, e1005744.

Capewell, P., Cren-Travaillé, C., Marchesi, F., Johnston, P., Clucas, C., Benson, R.A., Gorman, T.-A., Calvo-Alvarez, E., Crouzols, A., Jouvion, G*., et al.* (2016). The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. eLife *5*, e17716.

Capewell, P., Veitch, N.J., Turner, C.M.R., Raper, J., Berriman, M., Hajduk, S.L., and Macleod, A. (2011). Differences between Trypanosoma brucei gambiense Groups 1 and 2 in Their Resistance to Killing by Trypanolytic Factor 1. PLOS Neglected Tropical Diseases *5*, e1287.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., and Turner, M. (1991). Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. Journal of Molecular Biology *221*, 823-835.

Cavalcanti, D.P., and De Souza, W. (2018). The Kinetoplast of Trypanosomatids: From Early Studies of Electron Microscopy to Recent Advances in Atomic Force Microscopy. Scanning *2018*, 1-10.

Clayton, C. (2019). Regulation of gene expression in trypanosomatids: living with polycistronic transcription. Open Biology *9*, 190072.

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. EMBO J *21*, 1881-1888.

Clayton, C.E. (2014). Networks of gene expression regulation in Trypanosoma brucei. Molecular and Biochemical Parasitology *195*, 96-106.

Cloos, P.A.C., Christensen, J., Agger, K., and Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes & Development *22*, 1115-1140.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., and Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research *38*, 1767-1771.

Cross, M., Kieft, R., Sabatini, R., Dirks-Mulder, A., Chaves, I., and Borst, P. (2002). J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. Molecular microbiology *46*, 37-47.

Daniels, J.P., Gull, K., and Wickstead, B. (2010). Cell Biology of the Trypanosome Genome. Microbiology and Molecular Biology Reviews *74*, 552-569.

Das, A., Banday, M., and Bellofatto, V. (2008). RNA Polymerase Transcription Machinery in Trypanosomes: FIG. 1. Eukaryotic Cell *7*, 429-434.

Das, A., Banday, M., Fisher, M.A., Chang, Y.-J., Rosenfeld, J., and Bellofatto, V. (2017). An essential domain of an early-diverged RNA polymerase II functions to accurately decode a primitive chromatin landscape. Nucleic Acids Research *45*, 7886-7896.

De Jesus, T.C.L., Nunes, V.S., Lopes, M.D.C., Martil, D.E., Iwai, L.K., Moretti, N.S., Machado, F.C., De Lima-Stein, M.L., Thiemann, O.H., Elias, M.C*., et al.* (2016). Chromatin Proteomics Reveals Variable Histone Modifications during the Life Cycle of Trypanosoma cruzi. Journal of Proteome Research *15*, 2039-2051.

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., Wheeler, R.J., Hodkinson, I., Gluenz, E., and Gull, K. (2015). A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. Open Biology *5*, 140197-140197.

Dean, S., Sunter, J.D., and Wheeler, R.J. (2017). TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. Trends in Parasitology *33*, 80-82.

Deeks, E.D. (2019). Fexinidazole: First Global Approval. Drugs *79*, 215-220.

Dejesus, E., Kieft, R., Albright, B., Stephens, N.A., and Hajduk, S.L. (2013). A Single Amino Acid Substitution in the Group 1 Trypanosoma brucei gambiense Haptoglobin-Hemoglobin Receptor Abolishes TLF-1 Binding. PLOS Pathogens *9*, e1003317.

Denise, H., and Barrett, M.P. (2001). Uptake and mode of action of drugs used against sleeping sickness. Biochemical Pharmacology *61*, 1-5.

Dewar, S., Sienkiewicz, N., Ong, H.B., Wall, R.J., Horn, D., and Fairlamb, A.H. (2016). The Role of Folate Transport in Antifolate Drug Action inTrypanosoma brucei. Journal of Biological Chemistry *291*, 24768-24778.

DiPaolo, C., Kieft, R., Cross, M., and Sabatini, R. (2005). Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. Molecular cell *17*, 441-451.

Djupedal, I., Portoso, M., Spåhr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes & development *19*, 2301-2306.

Dreesen, O., and Cross, G.A.M. (2008). Telomere length in Trypanosoma brucei. Experimental parasitology *118*, 103-110.

Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma *91*, 313-321.

El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., and Donelson, J.E. (2000). The African trypanosome genome. International Journal for Parasitology *30*, 329-345.

Engstler, M. (2004). Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. Genes & Development *18*, 2798-2811.

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. (2011). Nuclear architecture, genome and chromatin organisation in Trypanosoma brucei. Research in Microbiology *162*, 626-636.

Ersfeld, K., and Gull, K. (1997). Partitioning of Large and Minichromosomes in Trypanosoma brucei. Science *276*, 611-614.

Farooq, Z., Banday, S., Pandita, T.K., and Altaf, M. (2016). The many faces of histone H3K79 methylation. Mutation Research/Reviews in Mutation Research *768*, 46-52.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nature Protocols *7*, 1728-1740.

Field, M.C., Horn, D., Fairlamb, A.H., Ferguson, M.A.J., Gray, D.W., Read, K.D., De Rycker, M., Torrie, L.S., Wyatt, P.G., Wyllie, S.*, et al.* (2017). Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. Nature Reviews Microbiology *15*, 217-231.

Figueiredo, L.M., Cross, G.A.M., and Janzen, C.J. (2009). Epigenetic regulation in African trypanosomes: a new kid on the block. Nature Reviews Microbiology *7*, 504-513.

Figueiredo, L.M., Janzen, C.J., and Cross, G.A.M. (2008). A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes. PLOS Biology *6*, e161.

Fisher, C.L., and Fisher, A.G. (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. Current Opinion in Genetics & Development *21*, 140-146.

Florini, F., Naguleswaran, A., Gharib, W.H., Bringaud, F., and Roditi, I. (2018). Unexpected diversity in eukaryotic transcription revealed by the retrotransposon hotspot family of Trypanosoma brucei. Nucleic acids research *47*, 1725-1739.

Formosa, T., Ruone, S., Adams, M.D., Olsen, A.E., Eriksson, P., Yu, Y.X., Rhoades, A.R., Kaufman, P.D., and Stillman, D.J. (2002). Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae cause dependence on the Hir/Hpc pathway: Polymerase passage may degrade chromatin structure. Genetics *162*, 1557-1571.

García-Salcedo, J.A., Gijón, P., Nolan, D.P., Tebabi, P., and Pays, E. (2003). A chromosomal SIR2 homologue with both histone NAD-dependent ADP-ribosyltransferase and deacetylase activities is involved in DNA repair in Trypanosoma brucei. EMBO J *22*, 5851-5862.

Gilinger, G. (2001). Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Research *29*, 1556-1564.

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.

Gommers-Ampt, J.H., Van Leeuwen, F., De Beer, A.L.J., Vliegenthart, J.F.G., Dizdaroglu, M., Kowalak, J.A., Crain, P.F., and Borst, P. (1993). β-d-glucosyl-hydroxymethyluracil: A novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell *75*, 1129-1136.

Graham, S.V., and Barry, J.D. (1995). Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of Trypanosoma brucei. Molecular and Cellular Biology *15*, 5945-5956.

Günzl, A. (2010). The Pre-mRNA Splicing Machinery of Trypanosomes: Complex or Simplified? Eukaryotic Cell *9*, 1159-1170.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.C., Chung, H.M., Lee, P.T., and Lee, M.G.S. (2003). RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma brucei. Eukaryotic Cell *2*, 542-551.

Haanstra, J.R., Kerkhoven, E.J., Van Tuijl, A., Blits, M., Wurst, M., Van Nuland, R., Albert, M.-A., Michels, P.A.M., Bouwman, J., Clayton, C.*, et al.* (2011). A domino effect in drug action: from metabolic assault towards parasite differentiation. Molecular Microbiology *79*, 94-108.

Haberland, M., Montgomery, R.L., and Olson, E.N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nature Reviews Genetics *10*, 32-42.

Haenni, S., Renggli, C.K., Fragoso, C.M., Oberle, M., and Roditi, I. (2006). The procyclin-associated genes of Trypanosoma brucei are not essential for cyclical transmission by tsetse. Molecular and Biochemical Parasitology *150*, 144-156.

Hager, K.M. (1994). Endocytosis of a cytotoxic human high density lipoprotein results in disruption of acidic intracellular vesicles and subsequent killing of African trypanosomes. The Journal of Cell Biology *126*, 155-167.

Hall, B.S., Bot, C., and Wilkinson, S.R. (2011). Nifurtimox Activation by Trypanosomal Type I Nitroreductases Generates Cytotoxic Nitrile Metabolites. Journal of Biological Chemistry *286*, 13088-13095.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Research *21*, 396-420.

Hebbes, T.R., Thorne, A.W., and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J *7*, 1395-1402.

Heddergott, N., Krüger, T., Babu, S.B., Wei, A., Stellamanns, E., Uppaluri, S., Pfohl, T., Stark, H., and Engstler, M. (2012). Trypanosome Motion Represents an Adaptation to the Crowded Environment of the Vertebrate Bloodstream. PLOS Pathogens *8*, e1003023.

Henikoff, S., and Smith, M.M. (2015). Histone Variants and Epigenetics. Cold Spring Harbor Perspectives in Biology *7*, a019364.

Hertz-Fowler, C., Figueiredo, L.M., Quail, M.A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I.*, et al.* (2008). Telomeric Expression Sites Are Highly Conserved in Trypanosoma brucei. PLOS ONE *3*, e3527.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes & Development *6*, 2288-2298.

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. The Journal of parasitology, 985-989.

Hong, L., Schroth, G., Matthews, H., Yau, P., and Bradbury, E. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4" tail" to DNA. Journal of Biological Chemistry *268*, 305-314.

Horn, D. (2014). Antigenic variation in African trypanosomes. Molecular and biochemical parasitology *195*, 123-129.

Horn, D., Spence, C., and Ingram, A.K. (2000). Telomere maintenance and length regulation in Trypanosoma brucei. EMBO J *19*, 2332-2339.

Ingram, A.K., and Horn, D. (2002). Histone deacetylases in Trypanosoma brucei: two are essential and another is required for normal cell cycle progression. Molecular Microbiology *45*, 89-97.

Jaé, N., Wang, P., Gu, T., Hühn, M., Palfi, Z., Urlaub, H., and Bindereif, A. (2010). Essential Role of a Trypanosome U4-Specific Sm Core Protein in Small Nuclear Ribonucleoprotein Assembly and Splicing. Eukaryotic Cell *9*, 379-386.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics *33*, 245-254.

Jamieson, K., Wiles, E.T., McNaught, K.J., Sidoli, S., Leggett, N., Shao, Y., Garcia, B.A., and Selker, E.U. (2016). Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. Genome Research *26*, 97-107.

Janzen, C.J., Fernandez, J.P., Deng, H., Diaz, R., Hake, S.B., and Cross, G.A.M. (2006a). Unusual histone modifications inTrypanosoma brucei. FEBS Letters *580*, 2306-2310.

Janzen, C.J., Hake, S.B., Lowell, J.E., and Cross, G.A.M. (2006b). Selective Di- or Trimethylation of Histone H3 Lysine 76 by Two DOT1 Homologs Is Important for Cell Cycle Regulation in Trypanosoma brucei. Molecular Cell *23*, 497-507.

Kawahara, T., Siegel, T.N., Ingram, A.K., Alsford, S., Cross, G.A.M., and Horn, D. (2008). Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. Molecular Microbiology *69*, 1054-1068.

Keeling, P.J., and Burki, F. (2019). Progress towards the Tree of Eukaryotes. Current Biology *29*, R808-R817.

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

Kennedy, P.G.E. (2019). Update on human African trypanosomiasis (sleeping sickness). Journal of Neurology *266*, 2334-2337.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204-2207.

Kieft, R., Brand, V., Ekanayake, D.K., Sweeney, K., DiPaolo, C., Reznikoff, W.S., and Sabatini, R. (2007). JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form Trypanosoma brucei. Molecular and biochemical parasitology *156*, 24-31.

Kolev, N.G., Günzl, A., and Tschudi, C. (2017). Metacyclic VSG expression site promoters are recognized by the same general transcription factor that is required for RNA polymerase I transcription of bloodstream expression sites. Molecular and Biochemical Parasitology *216*, 52-55.

Kolev, N.G., Ramey-Butler, K., Cross, G.A.M., Ullu, E., and Tschudi, C. (2012). Developmental Progression to Infectivity in Trypanosoma brucei Triggered by an RNA-Binding Protein. Science *338*, 1352-1353.

Kornberg, R.D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. Science *184*, 868-871.

Kueng, S., Oppikofer, M., and Gasser, S.M. (2013). SIR Proteins and the Assembly of Silent Chromatin in Budding Yeast. Annual Review of Genetics *47*, 275-306.

L'Hostis, C., Geindre, M., and Deshusses, J. (1993). Active transport of l-proline in the protozoan parasiteTrypanosoma brucei brucei. Biochemical Journal *291*, 297-301.

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

Langousis, G., and Hill, K.L. (2014). Motility and more: the flagellum of Trypanosoma brucei. Nature Reviews Microbiology *12*, 505-518.

Lanteri, C.A., Tidwell, R.R., and Meshnick, S.R. (2008). The Mitochondrion Is a Site of Trypanocidal Action of the Aromatic Diamidine DB75 in Bloodstream Forms of Trypanosoma brucei. Antimicrobial Agents and Chemotherapy *52*, 875-882.

Lecordier, L., Devaux, S., Uzureau, P., Dierick, J.F., Walgraffe, D., Poelvoorde, P., Pays, E., and Vanhamme, L. (2007). Characterization of a TFIIH homologue from Trypanosoma brucei. Molecular Microbiology *64*, 1164-1181.

Lecordier, L., Vanhollebeke, B., Poelvoorde, P., Tebabi, P., Paturiaux-Hanocq, F., Andris, F., Lins, L., and Pays, E. (2009). C-Terminal Mutants of Apolipoprotein L-I Efficiently Kill Both Trypanosoma brucei brucei and Trypanosoma brucei rhodesiense. PLOS Pathogens *5*, e1000685.

Lee, J.H., Cai, G., Panigrahi, A.K., Dunham-Ems, S., Nguyen, T.N., Radolf, J.D., Asturias, F.J., and Günzl, A. (2010). A TFIIH-Associated Mediator Head Is a Basal Factor of Small Nuclear Spliced Leader RNA Gene Transcription in Early-Diverged Trypanosomes. Molecular and Cellular Biology *30*, 5502-5513.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nature Reviews Molecular Cell Biology *8*, 284-295.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current Opinion in Genetics & Development *21*, 175-186.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, D., Albergante, L., Newman, T.J., and Horn, D. (2018). Faster growth with shorter antigens can explain a VSG hierarchy during African trypanosome infections: a feint attack by parasites. Scientific Reports *8*, 10922.

Llauró, A., Hayashi, H., Bailey, M.E., Wilson, A., Ludzia, P., Asbury, C.L., and Akiyoshi, B. (2018). The kinetoplastid kinetochore protein KKT4 is an unconventional microtubule tip–coupling protein. The Journal of Cell Biology, jcb.201711181.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251-260.

Lukeš, J., Lys Guilbride, D., Votýpka, J., Zíková, A., Benne, R., and Englund, P.T. (2002). Kinetoplast DNA Network: Evolution of an Improbable Structure. Eukaryotic Cell *1*, 495-502.

MacGregor, P., Szöőr, B., Savill, N.J., and Matthews, K.R. (2012). Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. Nature Reviews Microbiology *10*, 431-438.

Magalska, A., Anna, Moreno-Andrés, D., Zanini, F., Schooley, A., Sachdev, R., Schwarz, H., Madlung, J., and Antonin, W. (2014). RuvB-like ATPases Function in Chromatin Decondensation at the End of Mitosis. Developmental Cell *31*, 305-318.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I.*, et al.* (2000). A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. RNA *6*, 163-169.

Mandava, V., Fernandez, J.P., Deng, H., Janzen, C.J., Hake, S.B., and Cross, G.A.M. (2007). Histone modifications in Trypanosoma brucei. Molecular and biochemical parasitology *156*, 41-50.

Marchetti, M. (1998). Physical and transcriptional analysis of the Trypanosoma brucei genome reveals a typical eukaryotic arrangement with close interspersionof RNA polymerase II- and III-transcribed genes. Nucleic Acids Research *26*, 3591-3598.

Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S., and Matthews, K. (2006). Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in Trypanosoma brucei: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. Nucleic Acids Research *34*, 5312-5324.

Mesu, V.K.B.K., Kalonji, W.M., Bardonneau, C., Mordt, O.V., Blesson, S., Simon, F., Delhomme, S., Bernhard, S., Kuziena, W., Lubaki, J.-P.F.*, et al.* (2018). Oral fexinidazole for late-stage African Trypanosoma brucei gambiense trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial. The Lancet *391*, 144-154.

Michelotti, E., and Hajduk, S. (1987). Developmental regulation of trypanosome mitochondrial gene expression. Journal of Biological Chemistry *262*, 927-932.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Militello, K.T., Wang, P., Jayakar, S.K., Pietrasik, R.L., Dupont, C.D., Dodd, K., King, A.M., and Valenti, P.R. (2008). African Trypanosomes Contain 5-Methylcytosine in Nuclear DNA. Eukaryotic Cell *7*, 2012.

Morgan, H.P., McNae, I.W., Nowicki, M.W., Zhong, W., Michels, P.A.M., Auld, D.S., Fothergill-Gilmore, L.A., and Walkinshaw, M.D. (2011). The Trypanocidal Drug Suramin and Other Trypan Blue Mimetics Are Inhibitors of Pyruvate Kinases and Bind to the Adenosine Site. Journal of Biological Chemistry *286*, 31232-31240.

Morrison, L.J., Majiwa, P., Read, A.F., and Barry, J.D. (2005). Probabilistic order in antigenic variation of Trypanosoma brucei. International Journal for Parasitology *35*, 961-972.

Morrison, L.J., Vezza, L., Rowan, T., and Hope, J.C. (2016). Animal African Trypanosomiasis: Time to Increase Focus on Clinically Relevant Parasite and Host Species. Trends in parasitology *32*, 599-607.

Mugnier, M.R., Cross, G.A.M., and Papavasiliou, F.N. (2015). The in vivo dynamics of antigenic variation in Trypanosoma brucei. Science *347*, 1470-1473.

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

Müller, L.S.M., Cosentino, R.O., Förstner, K.U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C.J., Arampatzi, P., Vogel, J., Steinbiss, S.*, et al.* (2018). Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature *563*, 121-125.

Naguleswaran, A., Doiron, N., and Roditi, I. (2018). RNA-Seq analysis validates the use of culture-derived Trypanosoma brucei and provides new markers for mammalian and insect life-cycle stages. BMC Genomics *19*, 227.

Natesan, S.K.A., Black, A., Matthews, K.R., Mottram, J.C., and Field, M.C. (2011). Trypanosoma brucei brucei: Endocytic recycling is important for mouse infectivity. Experimental Parasitology *127*, 777-783.

Navarro, M., and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. Nature *414*, 759-763.

Ngô, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proceedings of the National Academy of Sciences *95*, 14687-14692.

Obado, S.O., Bot, C., Nilsson, D., Andersson, B., and Kelly, J.M. (2007). Repetitive DNA is associated with centromeric domains in Trypanosoma brucei but not Trypanosoma cruzi. Genome Biology *8*, R37.

Ogbadoyi, E., Ersfeld, K., Robinson, D., Sherwin, T., and Gull, K. (2000). Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. Chromosoma *108*, 501-513.

Ogbadoyi, E.O., Robinson, D.R., and Gull, K. (2003). A high-order trans-membrane structural linkage is responsible for mitochondrial genome positioning and segregation by flagellar basal bodies in trypanosomes. Mol Biol Cell *14*, 1769-1779.

Opperdoes, F.R., and Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei : The glycosome. FEBS Letters *80*, 360-364.

Ouna, B.A., Nyambega, B., Manful, T., Helbig, C., Males, M., Fadda, A., and Clayton, C. (2012). Depletion of Trypanosome CTR9 Leads to Gene Expression Defects. PLOS ONE *7*, e34256.

Overath, P., and Engstler, M. (2004). Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system. Molecular Microbiology *53*, 735-744.

Pal, A., Hall, B.S., Jeffries, T.R., and Field, M.C. (2003). Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in Trypanosoma brucei. Biochemical Journal *374*, 443-451.

Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.-C., Davis, I.J., Fahey, C.C., Hacker, K.E.*, et al.* (2016). Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell *166*, 950-962.

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am J Hum Genet *31*, 106-115.

Patrick, K.L., Shi, H., Kolev, N.G., Ersfeld, K., Tschudi, C., and Ullu, E. (2009). Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote Trypanosoma brucei. Proceedings of the National Academy of Sciences *106*, 17933-17938.

Pays, E., Laurent, M., Delinte, K., Van Meirvenne, N., and Steinert, M. (1983). Differential size variations between transcriptionally active and inactive telomeres of Trypanosoma brucei. Nucleic Acids Research *11*, 8137-8147.

Peacock, L., Bailey, M., Carrington, M., and Gibson, W. (2014). Meiosis and Haploid Gametes in the Pathogen Trypanosoma brucei. Current Biology *24*, 181-186.

Perez-Morga, D. (2005). Apolipoprotein L-I Promotes Trypanosome Lysis by Forming Pores in Lysosomal Membranes. Science *309*, 469-472.

Povelones, M.L., Gluenz, E., Dembek, M., Gull, K., and Rudenko, G. (2012). Histone H1 Plays a Role in Heterochromatin Formation and VSG Expression Site Silencing in Trypanosoma brucei. PLOS Pathogens *8*, e1003010.

Priest, J.W., and Hajduk, S.L. (1994). Developmental regulation of mitochondrial biogenesis inTrypanosoma brucei. Journal of Bioenergetics and Biomembranes *26*, 179-191.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research *44*, W160-W165.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature Protocols *2*, 1896-1906.

Ridsdale, J.A., Hendzel, M.J., Delcuve, G.P., and Davie, J.R. (1990). Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. Journal of Biological Chemistry *265*, 5150-5156.

Robinson, D.R. (1995). Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. The Journal of Cell Biology *128*, 1163-1172.

Robinson, D.R., and Gull, K. (1991). Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. Nature *352*, 731-733.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology *29*, 24-26.

Robinson, N.P., Burman, N., Melville, S.E., and Barry, J.D. (1999). Predominance of DuplicativeVSGGene Conversion in Antigenic Variation in African Trypanosomes. Molecular and Cellular Biology *19*, 5839-5846.

Roditi, I., and Liniger, M. (2002). Dressed for success: the surface coats of insect-borne protozoan parasites. Trends in Microbiology *10*, 128-134.

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.e316.

Rotureau, B., Subota, I., Buisson, J., and Bastin, P. (2012). A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly. Development *139*, 1842-1850.

Rout, M.P., and Field, M.C. (2001). Isolation and Characterization of Subnuclear Compartments from Trypanosoma brucei : IDENTIFICATION OF A MAJOR REPETITIVE NUCLEAR LAMINA COMPONENT. Journal of Biological Chemistry *276*, 38261-38271.

Ruan, J.-p., Arhin, G.K., Ullu, E., and Tschudi, C. (2004). Functional Characterization of a Trypanosoma brucei TATA-Binding Protein-Related Factor Points to a Universal Regulator of Transcription in Trypanosomes. Molecular and Cellular Biology *24*, 9610-9618.

Ruthenburg, A.J., Li, H., Patel, D.J., and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nature Reviews Molecular Cell Biology *8*, 983-994.

Saffery, R., Sumer, H., Hassan, S., Wong, L.H., Craig, J.M., Todokoro, K., Anderson, M., Stafford, A., and Choo, K.A. (2003). Transcription within a functional human centromere. Molecular cell *12*, 509-516.

Salmon, D., Vanwalleghem, G., Morias, Y., Denoeud, J., Krumbholz, C., Lhomme, F., Bachmaier, S., Kador, M., Gossmann, J., Dias, F.B.S*., et al.* (2012). Adenylate Cyclases of Trypanosoma brucei Inhibit the Innate Immune Response of the Host. Pathogens *337*, 463-466.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Saunders, A. (2003). Tracking FACT and the RNA Polymerase II Elongation Complex Through Chromatin in Vivo. Science *301*, 1094-1096.

Schimanski, B., Nguyen, T.N., and Günzl, A. (2005). Characterization of a Multisubunit Transcription Factor Complex Essential for Spliced-Leader RNA Gene Transcription in Trypanosoma brucei. Molecular and Cellular Biology *25*, 7303-7313.

Schulz, D., Zaringhalam, M., Papavasiliou, F.N., and Kim, H.-S. (2016). Base J and H3.V Regulate Transcriptional Termination in Trypanosoma brucei. PLOS Genetics *12*, e1005762.

Schwede, A., Jones, N., Engstler, M., and Carrington, M. (2011). The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Molecular and Biochemical Parasitology *175*, 201-204.

Schwer, B., and Verdin, E. (2008). Conserved Metabolic Regulatory Functions of Sirtuins. Cell Metabolism *7*, 104-112.

Seyfang, A., Mecke, D., and Duszenko, M. (1990). Degradation, Recycling, and Shedding ofTrypanosoma bruceiVariant Surface Glycoprotein. The Journal of Protozoology *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. Nature Methods *10*, 407-409.

Shapiro, T.A., and Englund, P.T. (1995). The Structure and Replication of Kinetoplast DNA. Annual Review of Microbiology *49*, 117-143.

Shaw, S., Demarco, S.F., Rehmann, R., Wenzler, T., Florini, F., Roditi, I., and Hill, K.L. (2019). Flagellar cAMP signaling controls trypanosome progression through host tissues. Nature Communications *10*.

Sherwin, T., Gull, K., and Vickerman, K. (1989). The cell division cycle of Trypanosoma brucei brucei: timing of event markers and cytoskeletal modulations. Philosophical Transactions of the Royal Society of London B, Biological Sciences *323*, 573-588.

Shi, H., Tschudi, C., and Ullu, E. (2006). An unusual Dicer-like1 protein fuels the RNA interference pathway in Trypanosoma brucei. RNA *12*, 2063-2072.

Siegel, T.N., Hekstra, D.R., Kemp, L.E., Figueiredo, L.M., Lowell, J.E., Fenyo, D., Wang, X., Dewell, S., and Cross, G.A.M. (2009). Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes & Development *23*, 1063-1076.

Siegel, T.N., Hekstra, D.R., Wang, X., Dewell, S., and Cross, G.A.M. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. Nucleic Acids Research *38*, 4946-4957.

Siegel, T.N., Kawahara, T., Degrasse, J.A., Janzen, C.J., Horn, D., and Cross, G.A.M. (2007). Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in Trypanosoma brucei. Molecular Microbiology *67*, 762-771.

Silmon De Monerri, N.C., and Kim, K. (2014). Pathogens Hijack the Epigenome. The American Journal of Pathology *184*, 897-911.

Silvester, E., McWilliam, K., and Matthews, K. (2017). The Cytological Events and Molecular Control of Life Cycle Development of Trypanosoma brucei in the Mammalian Bloodstream. Pathogens *6*, 29.

Simarro, P., Franco, J., Diarra, A., and Jannin, J. (2014). Epidemiology of human African trypanosomiasis. Clinical Epidemiology, 257.

Smolle, M., Workman, J.L., and Venkatesh, S. (2013). reSETting chromatin during transcription elongation. Epigenetics *8*, 10-15.

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009). Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. Cell *137*, 356-368.

Srivastava, A., Badjatia, N., Lee, J.H., Hao, B., and Günzl, A. (2017). An RNA polymerase II-associated TFIIF-like complex is indispensable for SL RNA gene transcription in Trypanosoma brucei. Nucleic Acids Research *46*, 1695-1709.

Steverding, D. (1995). Transferrin-binding protein complex is the receptor for transferrin uptake in Trypanosoma brucei. The Journal of Cell Biology *131*, 1173-1182.

Steverding, D. (2008). The history of African trypanosomiasis. Parasites & Vectors *1*, 3.

Sun, Z.-W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature Structural & Molecular Biology *14*, 1025-1040.

Thatcher, T.H., and Gorovsky, M.A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Research *22*, 174-179.

Thomas, J.A., Baker, N., Hutchinson, S., Dominicus, C., Trenaman, A., Glover, L., Alsford, S., and Horn, D. (2018). Insights into antitrypanosomal drug mode-of-action from cytology-based profiling. PLOS Neglected Tropical Diseases *12*, e0006980.

Torres, I.O., and Fujimori, D.G. (2015). Functional coupling between writers, erasers and readers of histone and DNA methylation. Current Opinion in Structural Biology *35*, 68-75.

Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Simon, Pinto, A., Jan*, et al.* (2016). Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. Cell Host & Microbe *19*, 837-848.

Tschudi, C., Shi, H., Franklin, J.B., and Ullu, E. (2012). Small interfering RNA-producing loci in the ancient parasitic eukaryote Trypanosoma brucei. BMC Genomics *13*, 427.

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. Molecular Microbiology *63*, 218-228.

Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homblé, F., Grélard, A., Zhendre, V., Nolan, D.P., Lins, L.*, et al.* (2013). Mechanism of Trypanosoma brucei gambiense resistance to human serum. Nature *501*, 430-434.

Van der Ploeg, L.H.T., Schwartz, D.C., Cantor, C.R., and Borst, P. (1984). Antigenic variation in trypanosoma brucei analyzed by electrophoretic separation of chromosome-sized DNA molecules. Cell *37*, 77-84.

Van Grinsven, K.W.A., Van Den Abbeele, J., Van Den Bossche, P., Van Hellemond, J.J., and Tielens, A.G.M. (2009). Adaptations in the Glucose Metabolism of ProcyclicTrypanosoma bruceiIsolates from Tsetse Flies and during Differentiation of Bloodstream Forms. Eukaryotic Cell *8*, 1307-1311.

van Hellemond, J.J., Opperdoes, F.R., and Tielens, A.G.M. (2005). The extraordinary mitochondrion and unusual citric acid cycle inTrypanosoma brucei. Biochemical Society Transactions *33*, 967-971.

Van Leeuwen, F., Wijsman, E.R., Kieft, R., Van Der Marel, G.A., Van Boom, J.H., and Borst, P. (1997). Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. Genes & Development *11*, 3232-3241.

Van Xong, H., Vanhamme, L., Chamekh, M., Chimfwembe, C.E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P., and Pays, E. (1998). A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in Trypanosoma rhodesiense. Cell *95*, 839-846.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A.*, et al.* (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature *422*, 83-87.

Vanhollebeke, B., De Muylder, G., Nielsen, M.J., Pays, A., Tebabi, P., Dieu, M., Raes, M., Moestrup, S.K., and Pays, E. (2008). A Haptoglobin-Hemoglobin Receptor Conveys Innate Immunity to Trypanosoma brucei in Humans. Science *320*, 677-681.

Vassella, E., Van Den Abbeele, J., Butikofer, P., Kunz Renggli, C.,´Furger, A., Brun, R., and Roditi, I. (2000). A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes & Development *14*, 615–626.

Vickerman, K. (1965). Polymorphism and Mitochondrial Activity In Sleeping Sickness Trypanosomes. Nature *208*, 762-766.

Vickerman, K. (1978). Antigenic variation in trypanosomes. Nature *273*, 613-617.

Vickerman, K. (1985). DEVELOPMENTAL CYCLES AND BIOLOGY OF PATHOGENIC TRYPANOSOMES. British Medical Bulletin *41*, 105-114.

Walker, G., Dorrell, R.G., Schlacht, A., and Dacks, J.B. (2011). Eukaryotic systematics: a user's guide for cell biologists and parasitologists. Parasitology *138*, 1638-1663.

Wang, Q.-P., Kawahara, T., and Horn, D. (2010). Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Molecular microbiology *77*, 1237-1245.

Webb, S., Kudla, G., and Granneman, S. (2018). The pyCRAC Manual version 1.3.2. Available online at: https://git.ecdf.ed.ac.uk/sgrannem/pycrac [Accessed at 02 December 2019].

Weber, C.M., and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. Genes & Development *28*, 672-682.

Wedel, C., Förstner, K.U., Derr, R., and Siegel, T.N. (2017). GT -rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. EMBO J *36*, 2581-2594.

West, S.C. (2003). Molecular views of recombination proteins and their control. Nature Reviews Molecular Cell Biology *4*, 435-445.

Wickstead, B., Ersfeld, K., and Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome research *14*, 1014-1024.

Woodward, R., and Gull, K. (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of Trypanosoma brucei. Journal of Cell Science *95*, 49-57.

Wright, J.R., Siegel, T.N., and Cross, G.A.M. (2010). Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in Trypanosoma brucei. Molecular and Biochemical Parasitology *172*, 141-144.

Yang, X., Figueiredo, L.M., Espinal, A., Okubo, E., and Li, B. (2009a). RAP1 Is Essential for Silencing Telomeric Variant Surface Glycoprotein Genes in Trypanosoma brucei. Cell *137*, 99-109.

Yang, X., Figueiredo, L.M., Espinal, A., Okubo, E., and Li, B. (2009b). RAP1 Is Essential for Silencing Telomeric Variant Surface Glycoprotein Genes in Trypanosoma brucei. Cell *137*, 99-109.

Ziegelbauer, K., and Overath, P. (1993). Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei. Infect Immun *61*, 4540-4545.

Appendix A

Chemical solutions

I. Trypanosome work

HMI-9 media

One bottle of HMI-9 powder (Life technologies)

45mM $NaHCO_3$

256μM β-mercaptoethanol

up to 4 litres with autoclaved dH2O

pH 7.5

The medium was filtered through a 0.22μm filter.

Before use, the media was supplemented with 10% heat-inactivated FCS (Gibco), 100U/ml Penicillin/ Streptomycin (Gibco) and selective drugs (when appropriate).

SDM-79 media

1 bottle of SDM-79 powder (Life technologies)

9g $NaHCO_3$

4.5 litres of autoclaved dH2O

pH 7.3

The medium was filtered through a 0.22μm filter.

Before use, the media was supplemented with 10% heat-inactivated FCS (Gibco), 100U/ml Penicillin/ Streptomycin (Gibco), 0.1% hemin and selective drugs (when appropriate).

1X PBS pH 7.4

Component	Concentration
NaCl	137 mM
KCl	2.7 mM
Na_2HPO_4	10 mM
KH_2PO_4	1.8 mM

Mowiol + DABCO

Component	Concentration
Mowiol 4-88 reagent (Calbiochem)	10% (w/v)
Glycerol	25% (v/v)
Tris-HCl pH 8.5	100 mM
DABCO	2.5% (v/v)

Homemade transfection buffer (Burkard et al., 2011)

Component	Concentration
NaPO4	90 mM
KCl	5 mM
$CaCl_2$	0.15 mM
HEPES-KOH pH 7.3	50 mM

II. ChIP

Formaldehyde solution

Component	Concentration
HEPES-KOH pH 7.5	50 mM
NaCl	100 mM
EDTA	1 mM
EGTA	0.5 mM
Formaldehyde	8% (v/v)

Lysis buffer 1

Component	Concentration
HEPES-KOH pH 7.5	50 mM
NaCl	140 mM
EDTA	1 mM
Glycerol	10% (v/v)
NP-40	0.5 % (v/v)
Triton X-100	0.25% (v/v)

Lysis buffer 2

Component	Concentration
Tris-HCl pH 8.0	10 mM
NaCl	200 mM
EDTA	1 mM
EGTA	0.5 mM

Lysis buffer 3

Component	Concentration
HEPES-KOH pH 7.5	50 mM
NaCl	140 mM
EDTA	1 mM
Triton X-100	1% (v/v)
sodium deoxycholate	0.1% (v/v)

Wash buffer 1

Component	Concentration
HEPES-KOH pH 7.5	50 mM
NaCl	500 mM
EDTA	1 mM
Triton X-100	1% (v/v)
sodium deoxycholate	0.1% (w/v)

Wash buffer 2

Component	Concentration
Tris-HCl pH 8.0	10 mM
LiCl	250 mM
NP-40	0.5% (v/v)
sodium deoxycholate	0.5% (w/v)
EDTA	1 mM

TE buffer

Component	Concentration
Tris-HCl pH 8.0	10 mM
EDTA	1 mM

Elution buffer

Component	Concentration
Tris-HCl pH 8.0	50 mM
EDTA	10 mM
SDS	1% (v/v)

III. Protein IPs

Lysis buffer MS

Component	Concentration
Tris-HCl pH 8.0	50 mM
KCl	150 mM
NP-40	0.2% (v/v)

Borate buffer pH 9.0

Component	Concentration
Boric acid	40 mM
Sodium tetraborate decahydrate	40 mM

ABC buffer pH 8.0

Component	Concentration
Ammonium bicarbonate	50 mM

IV. Miscellaneous

TBE buffer

Component	Concentration
Tris	89 mM
Acetic acid	89 mM
EDTA	2 mM

TAE buffer

Component	Concentration
Tris	40 mM
Acetic acid	20 mM
EDTA	1 mM

LB broth

Component	Concentration
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
H_2O	to 1 L

Appendix B

Primers used in this study

"F" denotes a forward primer, whereas "R" indicates a reverse primer.

Lowercase sequences align to plasmid DNA, whereas upper case sequences align to trypanosome DNA.

I. N-terminal YFP tagging

Primer name	Sequence
blast-eYFP F	gtataatgcagacctgctgc
blast eYFP R	actacccgatcctgatcc
Chromo1 CDS F	ggatcaggatcgggtagtATGTCCATCCACCAAACAAAAGTC
Chromo1 CDS R	TGACCACTCTCCCAAGCCTA
Chromo1 5′UTR F	GATGTATGGCGCAGAAGGGT
Chromo1 5′UTR R	gcagcaggtctgcattatacGAGTTCCGGTTGCAACTGATTG
PHD12 CDS F	ggatcaggatcgggtagtATGAAGTTTGAGAACCCCGTTG
PHD12 CDS R	TTCTCGGTCGGCATTGGAAA
PHD12 5′UTR F	ACGAGCGGGTTTAGTGAAGG
PHD12 5′UTR R	gcagcaggtctgcattatacCGGTGTGCCTGTGTGATAAA
PHD13 CDS F	ggatcaggatcgggtagtATGAGCACATCCCTCACGTC
PHD13 CDS R	GCACGACTTTCACCACGTTC
PHD13 5′UTR F	ACCCTCTTCAATGCACTCCG
PHD13 5′UTR R	gcagcaggtctgcattatacTGCAGACTACGGCAAAAGGA
PWWP1 CDS F	ggatcaggatcgggtagtATGATTCCATCATTTGCTCCAGG
PWWP1 CDS R	CCTTCGTCTTCTTTCCGCCT
PWWP1 5′UTR F	GGGCGTAATAACGAGGTGGA
PWWP1 5′UTR R	gcagcaggtctgcattatacTTTTTTTCTTTTCCCCCCGTGG

Primer name	Sequence
Tudor CDS F	ggatcaggatcgggtagtATGTCTTTTCTTGTGTATGCCG
Tudor CDS R	AGTAACCTGTGCATCGCCAA
Tudor 5′UTR F	CTTATCGGCTGCTTTCGCAC
Tudor 5′UTR R	gcagcaggtctgcattatacTTGAGTATGTCCTCTGTTACCCC
Znf-CW1 CDS F	ggatcaggatcgggtagtATGCTAACTCGAAGAGCCGC
Znf-CW1 CDS R	TCACCTCAGGTTCGCTTTCC
Znf-CW1 5′UTR F	TAACGACACTGCAAATCCCT
Znf-CW1 5′UTR R	gcagcaggtctgcattatacATATATATATATAACAAATATGCACAAC
BDF2 CDS F	ggatcaggatcgggtagtATGAGCAAGAACGAGCGAGATAC
BDF2 CDS R	CGCAGTGTTGATTCGTCGTC
BDF2 5′UTR F	TGTGGATTCAGTGGGTTGGG
BDF2 5′UTR R	gcagcaggtctgcattatacTGCTCTAACGCAGATGTTCG
BDF4 CDS F	ggatcaggatcgggtagtATGTCGGGGGGCACCTC
BDF4 CDS R	AAAACCTCCTCCATCAGCCG
BDF4 5′UTR F	TTACTGGGTTGTGTGAGCGG
BDF4 5′UTR R	gcagcaggtctgcattatacTGTGGAGGGACAACGAAGAG
SET1 CDS F	ggatcaggatcgggtagtATGCAAACAGAAGCCCTGC
SET1 CDS R	AGTAGACCGTGGTAACGGGA
SET1 5′UTR F	ACAGTACCCCGTTTGCTGAG
SET1 5′UTR R	gcagcaggtctgcattatacGCAACCAAAAAAAAAAAGGAGCTAC
SET2 CDS F	ggatcaggatcgggtagtATGGTGAGGCTTCGCTCG
SET2 CDS R	AAGGCTCCTGTTCACTGACG
SET2 5′UTR F	TGAGTCTCCGCATATGGCAC
SET2 5′UTR R	gcagcaggtctgcattatacGCAAGGGCTAACCGACCC
SET3 CDS F	ggatcaggatcgggtagtATGGAGGAGTTGCGTAAACG
SET3 CDS R	TGGCGGCAAAATCTCAGCTA
SET3 5′UTR F	ATTACATGGAGGGGATGCCG
SET3 5′UTR R	gcagcaggtctgcattatacGGCAACTGTTGTGCCCAC

Primer name	Sequence
SET4 CDS F	ggatcaggatcgggtagtATGTCATGGGTGGAAGACAAGTC
SET4 CDS R	TAAGCGTTACGTGAGCCTCG
SET4 5′UTR F	CCACTTTCCCTTAGCCCGAT
SET4 5′UTR R	gcagcaggtctgcattatacCCCAGGTGGGTTGGCAAC
SET5 CDS F	ggatcaggatcgggtagtATGCCGACTCAGAGCCCC
SET5 CDS R	GCCACAAAAGAGCTTCCACG
SET5 5′UTR F	GAGTGTGAGGAACGGAAGCA
SET5 5′UTR R	gcagcaggtctgcattatacCCTTTTCCACTTTCTCCCCC
SET21 CDS F	ggatcaggatcgggtagtATGCTCCAGCAATTATCCGC
SET21 CDS R	GCGTTGTTTCTCTGGCGAAG
SET21 5′UTR F	GATGTGGTCATGCGTGGTTG
SET21 5′UTR R	gcagcaggtctgcattatacAGCCCAATGATACTGGGCC
SET22 CDS F	ggatcaggatcgggtagtATGGCCACTGTCACGGAG
SET22 CDS R	GAGCCAGTTCGTTGGTGAGA
SET22 5′UTR F	CAGCACGCTTCATCCCTACA
SET22 5′UTR R	gcagcaggtctgcattatacTTTTGTTGCCAAATTACGTATGTAC
SET23 CDS F	ggatcaggatcgggtagtATGCCTGGGTGTCTTGAGG
SET23 CDS R	TTCGACAGCACATTCCGTGA
SET23 5′UTR F	TCACTACAAGGCCGGTTGTC
SET23 5′UTR R	gcagcaggtctgcattatacTATCCGGTCATATATATATATATAT
SET24 CDS F	ggatcaggatcgggtagtATGAACAGGGGGTCGGTG
SET24 CDS R	TCTTGAGGTTTTCCCTCGCC
SET24 5′UTR F	GAGGGGAAAGTGCTGCTCAT
SET24 5′UTR R	gcagcaggtctgcattatacTCGCTACGCCTTTCCACC
SET25 CDS F	ggatcaggatcgggtagtATGCCGGCCAAACCCAC
SET25 CDS R	CCAACACTGCTGTAACCGGA
SET25 5′UTR F	GGTGCAAAGGAATGACGGTG
SET25 5′UTR R	gcagcaggtctgcattatacCGCTATCGGTCGTGAAAACTAC

Primer name	Sequence
SET26 CDS F	ggatcaggatcgggtagtATGATATGCAAGGTTTTGGAGCG
SET26 CDS R	TACTCACACATCACGGTGCC
SET26 5′UTR F	GGTTACCCGAAAGCAGTGGA
SET26 5′UTR R	gcagcaggtctgcattatacGCACTGTGGCAACTAATCACG
SET27 CDS F	ggatcaggatcgggtagtATGGATGTGCATTCCCCGTTAC
SET27 CDS R	TCGTTCGCAAACCGACCATA
SET27 5′UTR F	TCTCGACTATGCGGTTCGTG
SET27 5′UTR R	gcagcaggtctgcattatacTCCACACGAGGAAACACTGC
SET28 CDS F	ggatcaggatcgggtagtATGGTAATGAAAGCCAATTTCTGC
SET28 CDS R	TTTGGTTGGCAAGAATGCCG
SET28 5′UTR F	AATGGTTTGCCGGGACATCT
SET28 5′UTR R	gcagcaggtctgcattatacTTCTCCTCGCTCTTAGTTTATCCG
SET29 CDS F	ggatcaggatcgggtagtATGTCTGAGGTGGTACCTACTCG
SET29 CDS R	AAAGCGGCAGGGACTTCTAC
SET29 5′UTR F	TGAGGCGCATGATCCACTTT
SET29 5′UTR R	gcagcaggtctgcattatacACCAACTTGGAGTACAAAGAAGAC
SET30 CDS F	ggatcaggatcgggtagtATGTACAGAATAGCAATTCGACG
SET30 CDS R	TTACCTCCAGGATGAGCGGA
SET30 5′UTR F	CACAATAAGTGCCATGCGGT
SET30 5′UTR R	gcagcaggtctgcattatacCCCTTTTCTTATACTCCTTCACG
Jmj1 CDS F	ggatcaggatcgggtagtATGCTGGAACTGGACGGAC
Jmj1 CDS R	CCCTCCAGCGTTCAATGACT
Jmj1 5′UTR F	ACCAGCATTGGGATCGGTTT
Jmj1 5′UTR R	gcagcaggtctgcattatacAGCCAACAGAAACTTGCAAAAATATATAC
Jmj2 CDS F	ggatcaggatcgggtagtATGGCGCAGAGCTGGTTC
Jmj2 CDS R	CGGACGATGGGACAAGTTCA
Jmj2 5′UTR F	TTGGCTGAAGAGCATGCAGA
Jmj2 5′UTR R	gcagcaggtctgcattatacGTGTGAAACACTTCAAAATAAAAATAAGAC

Primer name	Sequence
LCM1 CDS F	ggatcaggatcgggtagtATGTCCATTGCGGTAGAAGAGAG
LCM1 CDS R	CTTGGATTGCATAACGGCGG
LCM1 5′UTR F	ATCGAAGACACGACGGCTTT
LCM1 5′UTR R	gcagcaggtctgcattatacTAAGGTGTTTTGTGAGGTAGCAAC
CLD1 CDS F	ggatcaggatcgggtagtATGCCAAATGAAAAGGCGCAG
CLD1 CDS R	TGCATCCGATTCCGTCACAA
CLD1 5′UTR F	GATTTGCTTGGGTGGTGCTG
CLD1 5′UTR R	gcagcaggtctgcattatacCGCATGAACTTCCTGCAGTTTTTG
HDAC1 CDS F	ggatcaggatcgggtagtATGAATGAGGATGGTGATTGTCGG
HDAC1 CDS R	GATGGCCTGTACCCGGAAAA
HDAC1 5′UTR F	TCTTGTTTCCCTCTGCGTCC
HDAC1 5′UTR R	gcagcaggtctgcattatacTTGAGAAACTGTACGAACCTTCTAGG
HDAC2 CDS F	ggatcaggatcgggtagtATGGTTGTGGAGCCTCCTG
HDAC2 CDS R	TGACTCGTCGTGGAACACAG
HDAC2 5′UTR F	GCTACTGGTGGAGCATGTGT
HDAC2 5′UTR R	gcagcaggtctgcattatacAGCTGGGCATGCAATCAGT
HDAC3 CDS F	ggatcaggatcgggtagtATGGGCAAAGAAACGCGTG
HDAC3 CDS R	TTGCGGTGATGTCCAGAGTG
HDAC3 5′UTR F	CCTGCCATATTGTGGGGGTT
HDAC3 5′UTR R	gcagcaggtctgcattatacTGCGGTTGCGGTTCAAC
HDAC4 CDS F	ggatcaggatcgggtagtATGTCTGTGAGGGAACGG
HDAC4 CDS R	TGACACGTCCGGTAGAAACA
HDAC4 5′UTR F	ATAGACACATCGGCGGAGTT
HDAC4 5′UTR R	gcagcaggtctgcattatacACGATAGAGTTGACTAAGTAAACAC
Sir2rp1 CDS F	ggatcaggatcgggtagtATGACAGAACCGAAGTTAGCAAC
Sir2rp1 CDS R	CATGCCCTTCTTGGCGAGTA
Sir2rp1 5′UTR F	GCTGAAGAGGTGCCTACGTT
Sir2rp1 5′UTR R	gcagcaggtctgcattatacTTCACGTTTCCCTATACTAGTTGG

Primer name	Sequence
Sir2rp2 CDS F	ggatcaggatcgggtagtATGGCTGACCGCCTTGCT
Sir2rp2 CDS R	CTCGTAGTCGCCATCAGGTC
Sir2rp2 5′UTR F	TGCGCCCATGAGTTGTAGTT
Sir2rp2 5′UTR R	gcagcaggtctgcattatacATCACTCAACTGCACTAGCTTAAAC
Sir2rp3 CDS F	ggatcaggatcgggtagtATGAGGCGGCCCAATGGT
Sir2rp3 CDS R	TCTACGACTCCGCAGCATTC
Sir2rp3 5′UTR F	TGGATGAACCGACTGCACAA
Sir2rp3 5′UTR R	gcagcaggtctgcattatacTGTTGAACCTACAAACAGTCCTTC
Ago1 CDS F	ggatcaggatcgggtagtATGTCTGACTGGGAACGTGG
Ago1 CDS R	GGTTTGTCCATGCCTCCACT
Ago1 5′UTR F	TAGCAGCAAAATGATGGCAC
Ago1 5′UTR R	gcagcaggtctgcattatacTTTATTTAAATCCTTTTTATTAGGTTGC
KKT2 CDS F	ggatcaggatcgggtagtATGTTCAATGTCTCACCAGCG
KKT2 CDS R	CACGTACACTCGACCAAACC
KKT2 5′UTR F	TTACGAGGGGTGGTAGTGGT
KKT2 5′UTR R	gcagcaggtctgcattatacGACTGTGCCCCGTCCCTT
UP1 CDS F	ggatcaggatcgggtagtATGGTAGGAGGATATTTACGTATGC
UP1 CDS R	GTGATCAACAACACATACGGATG
UP1 5′UTR F	TGTCACAGATGGCCCATG
UP1 5′UTR R	gcagcaggtctgcattatacTCCCTATTATGCAATAAAGAGAAAGG
UP2 CDS F	ggatcaggatcgggtagtATGGATTTAGGGGAAGATTCGAC
UP2 CDS R	CTTCCTCGTTACCCACTTGC
UP2 5′UTR F	TCTCCGTTGTTGGCAACC
UP2 5′UTR R	gcagcaggtctgcattatacCACTAATACCTTCTTCACCAGTTAC
UP3 CDS F	ggatcaggatcgggtagtATGAACGAGCCTGATGCCAT
UP3 CDS R	CAAATAGTCAATGAGCACCTGTGG
UP3 5′UTR F	CTGAAGCGTATGGAGACTGGC
UP3 5′UTR R	gcagcaggtctgcattatacCACTTTGCCCTTCAATACACCTATAC

Primer name	Sequence
UP4 CDS F	ggatcaggatcgggtagtATGGATCCGGCGGTAGGG
UP4 CDS R	ACAGCAAGCGGGTTCCATG
UP4 5′UTR F	AGTCGTTTAGATGCTTGCGTTG
UP4 5′UTR R	gcagcaggtctgcattatacATGTTGAAGGGAGAAAGTCTACGAT
JBP2 CDS F	ggatcaggatcgggtagtATGCCTATGTTTATGGATGGGGC
JBP2 CDS R	GAAGCCTCTGTGTGAGAGCA
JBP2 5′UTR F	AGGACACACACGAAACCGAT
JBP2 5′UTR R	gcagcaggtctgcattatacTGCCGTAACTGGTAGCGTAAAG

II. Integration validation

Primer name	Sequence
eYFP F	taagcttgtgagcaagggcg
blast R	accatggttagccctccca
Chromo1 5′UTR F	CACCACGGTGCTTGTAGTC
Chromo1 CDS R	GGTACTGAGAGGAAGATTCCTGAA
PHD12 5′UTR F	GGTAAACTATAGCCATTGGGGAC
PHD12 CDS R	AATGTCGCTCGAACCGTTC
PHD13 5′UTR F	GGCAATAGTTTCAGGTGAAGGG
PHD13 CDS R	CAAAGGGTGCTCGATCAGGA
PWWP1 5′UTR F	GTCACACTACTGTAGACCGGTG
PWWP1 CDS R	CTCCTTCCTTGAGATGTGATGACG
Tudor 5′UTR F	CTGCCATATGCCACTCACCT
Tudor CDS R	CGAGATACAGTTTCACAGTCGG
Znf-CW1 5′UTR F	CGCTAAAGTAACCCCTAAGCTC
Znf-CW1 CDS R	GATGAACTGATCACCTCTCCCTC
BDF2 5′UTR F	CTAGTGGATCAACAATTCGAACCC
BDF2 CDS R	CTTGTTCCGTTTTTCTGCACG
BDF4 5′UTR F	CGTTTGTGGGATGATCTTTCCTTAC
BDF4 CDS R	CGCTCCGCTTCTACCTCAG
SET1 5′UTR F	CTCGTAATGGATGCGGACC
SET1 CDS R	CGTGTCCGCCTTAACGAC
SET2 5′UTR F	GATGCTGATCAGCTTTTGGCAC
SET2 CDS R	GCAGCCTGTGTTTCTGGC
SET3 5′UTR F	CGCTTGTTGCTGATGGATG
SET3 CDS R	CCATCGTCATTACCCTCACG
SET4 5′UTR F	CACCACCAACTTGGGGAGG
SET4 CDS R	CGCCACTGCACAAATCAGG

Primer name	Sequence
SET5 5′UTR F	GAGGAGAGAATGTGTGCGAAG
SET5 CDS R	GAGGTGGGAACACTCAACAG
SET21 5′UTR F	GCCTAGAAGGAAGGCGGTAG
SET21 CDS R	CCAGTTATGTTCACCTTCATCCG
SET22 5′UTR F	GGCAGATGGATGCTTCACCG
SET22 CDS R	GTGCATCCTATCTCCGAGATCG
SET23 5′UTR F	GGCTACTGCAACTGCCATG
SET23 CDS R	GACTATAACAGCCAGCTCCGG
SET24 5′UTR F	CCTGACTGGTCGCTCGTAC
SET24 CDS R	CCGCTGGATTACAACTGTGC
SET25 5′UTR F	GACAGTTGGATTCACTGCACTG
SET25 CDS R	GTCATAAACTTTGCACGACTTATCC
SET26 5′UTR F	GAAGGCATCCCTGCTGCTTC
SET26 CDS R	GCTTTGTTCGTGTTCTTTTGGAG
SET27 5′UTR F	CCTGAGTGGTGTGTGGTAACG
SET27 CDS R	GCACTAATGTTTGCCGTAACG
SET28 5′UTR F	GAACGTTTCCCGTTAGGTCAAAC
SET28 CDS R	CGTAGCAACAGCGGATATGG
SET29 5′UTR F	GAAAGGGGTGCCTACGTGTC
SET29 CDS R	GTCGGGCGCCTCTCTAAC
SET30 5′UTR F	GGTCTCACTGTCAATAAGAGGATG
SET30 CDS R	CGTACCTGATGCACTCAACAG
Jmj1 5′UTR F	CATATGCTTGTTGGTACGACCTTG
Jmj1 CDS R	ATCAGACTGCATGTGCCAGTC
Jmj2 5′UTR F	GAAGCTTGATCTGACTCTCAGGG
Jmj2 CDS R	TGGTGCACATCCATGCCTG
LCM1 5′UTR F	TGGAAGGGATGCACTCCTAG
LCM1 CDS R	GGTGCCATCTTTACAAGCATTG

Primer name	Sequence
CLD1 5′UTR F	TTCAGCCTCTTGGCGCAG
CLD1 CDS R	CTGGTGCCATTCTTACTGTCCT
HDAC1 5′UTR F	GAGGCAACTACTGTTTGCACTC
HDAC1 CDS R	CCGTTTCCAATGCCTACATCC
HDAC2 5′UTR F	GTGGTATGCGTGGATGTAGTTG
HDAC2 CDS R	GCGTAAACTCAAATAGCGCACG
HDAC3 5′UTR F	CGTCGTGACAGTTTTGTGACG
HDAC3 CDS R	CGACACCTGTACCTTCAGCTTC
HDAC4 5′UTR F	GGAAGAGCTCTACCGTGGAG
HDAC4 CDS R	CTCTAGGATCGTATGCCCATGC
Sir2rp1 5′UTR F	TCAACGGCTGGGTGACATC
Sir2rp1 CDS R	GTCTATGTTCTGCGTACAACAGC
Sir2rp2 5′UTR F	GGGCTGGATGACAAATGAGTG
Sir2rp2 CDS R	CCTGCACAACACGGTCAG
Sir2rp3 5′UTR F	GCGGTCTCCACTGCTGTTG
Sir2rp3 CDS R	GCTGCAGGGTATACGTTACCG
Ago1 5′UTR F	GTTGTCTGACTGGAAACTTACTCG
Ago1 CDS R	CCTCGGGTCGATGTACAGAG
KKT2 5′UTR F	GCCTGACCTAATCATTGTCGG
KKT2 CDS R	GACGATACTTTCAAGGCGCAC
UP1 5′UTR F	TGTTCGTGTCGTGGCTGATG
UP1 CDS R	GAGTCTGCTGTATATCCAATGTCC
UP2 5′UTR F	GGTGAAGTGTGAGTACGTGAG
UP2 CDS R	CCATCGTCATCTGGTTTGTGTG
UP3 5′UTR F	CGCCACACACATTATGAAGCG
UP3 CDS R	CCTTTTCTTCTGTAGGTGCACG
UP4 5′UTR F	TGACCGATAATGCATGTTCTGC
UP4 CDS R	TAGCAGGACAGCACTTCAGC

Primer name	Sequence
JBP2 5′UTR F	GTGGCGTAATTTGCACAACAAG
JBP2 CDS R	CAATGTTTACAGCTGCTTCAGC

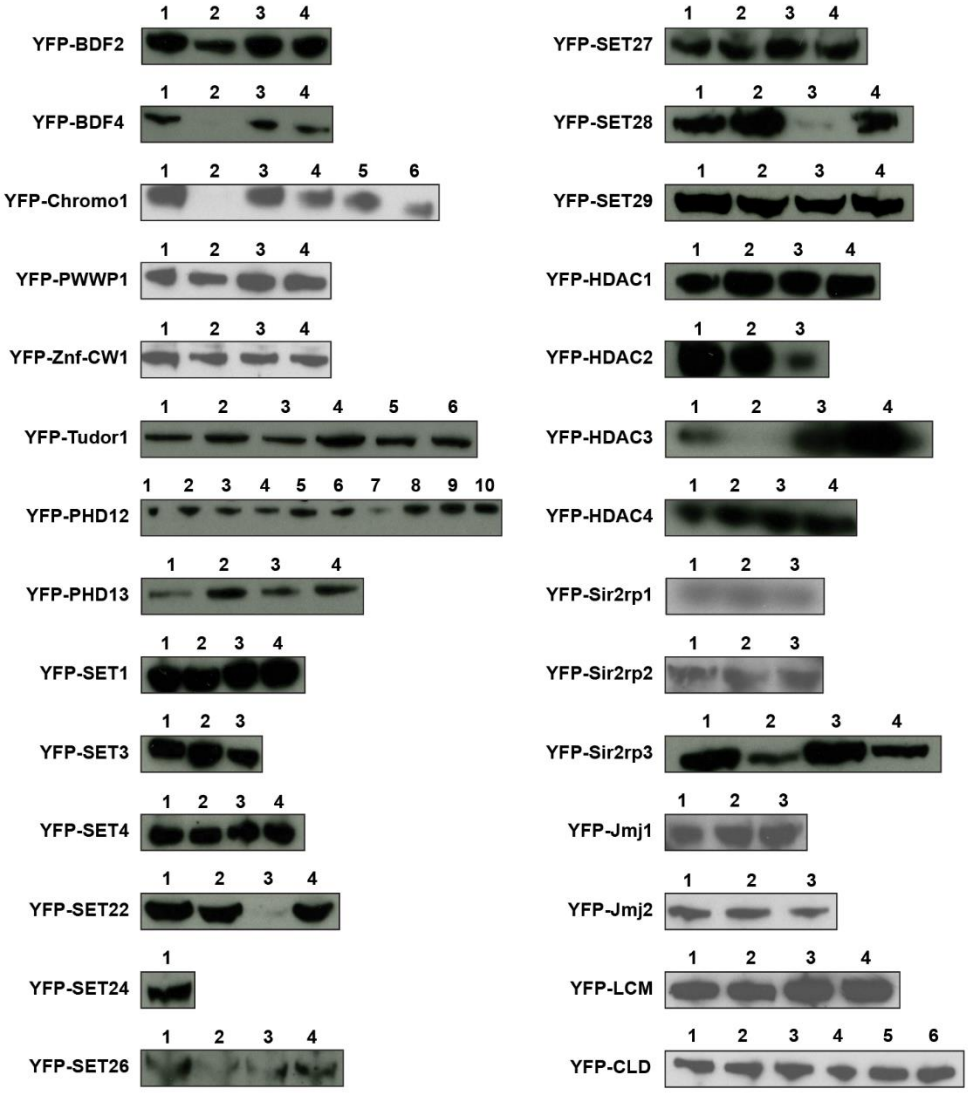
Appendix C

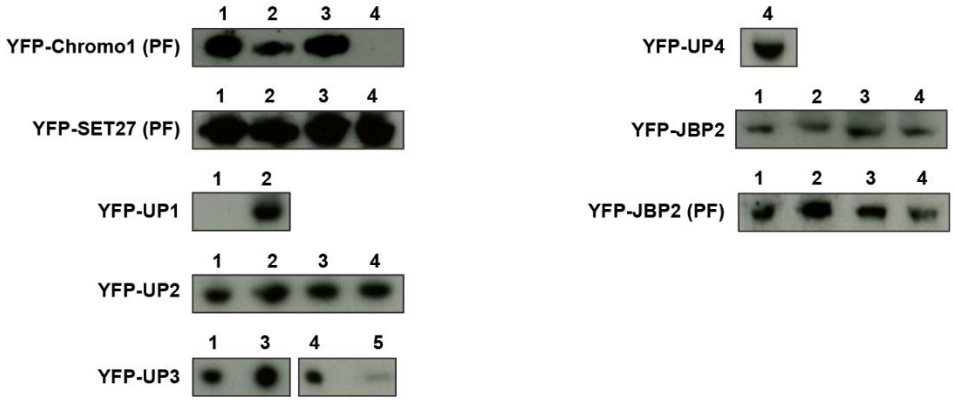
Clone tests

All proteins except the ones marked with “PF” were tagged in the bloodstream form.

I. Protein expression validation

SET2, SET5, SET21, SET25 and SET30 were not detected by western blot analysis. The other candidate proteins were detected in at least some of the clones tested (data below). The clone number is indicated above each lane.




1 2 3 4
YFP-Chromo1 (PF)
1 2 3 4
YFP-SET27 (PF)
1 2
YFP-UP1
1 2 3 4
YFP-UP2
1 3 4 5
YFP-UP3
4
YFP-UP4
1 2 3 4
YFP-JBP2
1 2 3 4
YFP-JBP2 (PF)

II. Tagging constructs integration validation

The table below summarises the PCR validations carried out to confirm correct construct integration at the 5′UTR and CDS junctions of each gene. The number of clones reflects the cell line stocks as present on 9 January 2020. Some clones were discarded if they did not express the tagged protein or if the PCR validations did not show correct integration. Additionally, some clones were lost during the project because they did not survive freeze-thawing.

Protein name	Gene code	Number of clones	5′ UTR PCR check	CDS PCR check
BDF2	Tb927.10.7420	4	unclear	correct
BDF4	Tb927.7.4380	3	correct	correct
Chromo1	Tb11.v5.0267	5	correct	unclear
PWWP1	Tb927.2.3480	4	correct	correct
Znf-CW1	Tb927.10.11720	4	correct	correct (unclear in clone 4)
Tudor	Tb927.11.14190	6	correct	correct
PHD12	Tb927.3.2140	7	correct	correct
PHD13	Tb927.11.5870	4	correct	correct
SET1	Tb11.v5.0422	4	correct	correct
SET2	Tb11.v5.0590	8	correct	correct
SET3	Tb927.1.4720	3	correct	correct
SET4	Tb927.10.11130	4	unclear	correct
SET5	Tb927.10.12880	8	correct (unclear in clones 1-4)	correct
SET21	Tb927.8.2690	4	correct	correct
SET22	Tb927.8.2710	4	correct	correct
SET24	Tb927.8.6530	4	correct	correct

Protein name	Gene code	Number of clones	5′ UTR PCR check	CDS PCR check
SET25	Tb927.9.1510	1	unclear	unclear
SET26	Tb927.10.8100	4	correct	correct
SET27	Tb927.9.13470	4	correct	correct
SET28	Tb927.9.10100	4	correct	correct
SET29	Tb927.3.3370	2	correct	correct
SET30	Tb927.3.1860	4	correct	correct
HDAC1	Tb927.10.1680	4	correct	correct
HDAC2	Tb927.11.15600	3	correct	correct (unclear in clone 1)
HDAC3	Tb927.2.2190	4	correct	correct
HDAC4	Tb927.5.2900	4	correct	correct
Sir2rp1	Tb927.7.1690	3	not detected	correct
Sir2rp2	Tb927.8.3140	3	correct	correct (unclear in clone 2)
Sir2rp3	Tb927.4.2520	4	correct	correct
Jmj1	Tb927.11.1760	3	correct	correct
Jmj2	Tb927.11.2000	3	correct	correct
LCM	Tb927.9.12780	4	correct	correct
CLD	Tb927.7.660	6	correct	correct
Chromo1 (PF)	Tb11.v5.0267	4	correct	correct
SET27 (PF)	Tb927.9.13470	4	correct	correct
UP1	Tb927.1.4250	1	correct	correct
UP2	Tb927.11.13820	4	correct	correct
UP3	Tb927.11.11840	4	correct	correct
UP4	Tb927.3.2350	1	correct	correct
JBP2	Tb927.7.4650	4	correct	correct
JBP2 (PF)	Tb927.7.4650	4	correct	correct

Appendix D

Significantly enriched proteins in the affinity purification experiments

The cut-offs used for significance are $p < 0.01$ and $\log_2$ (tagged/untagged) > 2 or < -2.

Chromo1 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**$-\log_{10}$ Student's t-test p-value**	**$\log_2$ (tagged/untagged)**
Q4GYF3	Tb927.1.4250	UP1 (hypothetical protein)	5.53089	11.4431
Q38D80	Tb927.9.13470	SET27	5.94395	11.4397
Q57X70	Tb11.v5.0267	Chromo1	4.08625	9.18821
Q382F5	Tb927.11.13820	UP2 (hypothetical protein)	6.65416	9.17537
Q582U8	Tb927.3.2350	UP4 (hypothetical protein)	7.10211	9.04339
Q382Z7	Tb927.11.11840	UP3 (hypothetical protein)	6.06473	8.75264
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	6.33932	7.9307
Q38FV1	Tb927.9.2070	hypothetical protein	2.09326	6.95152
Q38BC1	Tb927.10.4800	hypothetical protein	2.70589	4.47099
Q38BS4	Tb927.10.3220	nucleolus protein	4.25958	4.04605
Q586X9	Tb927.2.3580	transcription elongation factor s-II (TFIIS2-1)	2.30696	3.65927
Q57W25	Tb927.5.800	casein kinase I, isoform 2	5.87744	2.9952
Q387L4	Tb927.11.370	repressor activator protein 1 (RAP1)	3.04907	2.19994
Q57YU2	Tb927.8.7400	RNA polymerase IIA largest subunit (RPB1)	2.92416	2.01862
Q580S5	Tb927.3.5500	DNA-directed RNA polymerase II subunit 3 (RPB3)	2.41931	2.06975

SET27 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q4GYF3	Tb927.1.4250	UP1	5.32451	11.5549
Q38D80	Tb927.9.13470	SET27	4.15023	11.3024
Q57X70	Tb11.v5.0267	Chromo1	4.41402	10.8901
Q382F5	Tb927.11.13820	UP2	4.76568	9.39333
Q382Z7	Tb927.11.11840	UP3	5.30182	9.15674
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	5.09272	9.10164
Q582U8	Tb927.3.2350	UP4	4.70479	8.88102
Q38BT9	Tb927.10.3070	hypothetical protein	3.00745	4.65672
Q57Z72	Tb927.5.3210	small ubiquitin-related modifier	2.2191	4.57882
Q38EP6	Tb927.9.6920	hypothetical protein	2.10149	4.34594
Q38BC1	Tb927.10.4800	hypothetical protein	2.71533	4.25565
Q586X9	Tb927.2.3580	transcription elongation factor s-II (TFIIS2-1)	2.21306	4.05266
Q38BS4	Tb927.10.3220	nucleolus protein	2.90768	4.0165
P86938	Tb927.11.13640	J-binding protein 1 (JBP1)	3.18784	3.5602
Q57W25	Tb927.5.800	casein kinase I, isoform 2	5.83603	3.04821
Q389E6	Tb927.10.11960	hypothetical protein	2.37392	2.98484
Q57W24	Tb927.5.790	casein kinase I isoform 1	3.60939	2.82452
Q57VC7	Tb927.5.1360	nucleoside 2-deoxyribosyltransferase	3.59736	2.56284
Q586H7	Tb927.6.5050	V-type ATPase, C subunit	3.63701	2.53399
Q388E9	Tb927.10.14530	19S proteasome non-atpase subunit 8	2.16045	2.44652
Q57UB5	Tb927.5.3510	structural maintenance of chromosome 3 (SMC3)	3.09014	2.17858
Q38E40	Tb927.9.9550	hypothetical protein	3.18533	2.14501
Q38C86	Tb927.10.1550	proteasome regulatory non-ATP-ase subunit 5	4.10981	2.00096

Znf-CW1 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q57UN4	Tb927.7.4040	hypothetical protein	5.00494	11.7469
Q383K6	Tb927.11.10730	SWI/SNF-related helicase	6.66968	11.6905
Q580C6	Tb927.4.980	actin	6.24716	11.6278
Q389H2	Tb927.10.11690	YEATS family	4.11583	11.561
Q57ZL0	Tb927.3.3020	actin-like protein	4.47839	11.3822
Q584P9	Tb927.6.2570	SUMO-interacting motif-containing protein	3.71735	10.8057
Q38C43	Tb927.10.2000	actin-like protein	4.04326	10.1429
Q385E0	Tb927.11.6290	HIT zinc finger	5.72496	10.0378
Q389G9	Tb927.10.11720	Znf-CW1	5.52659	9.89842
Q57YH0	Tb927.8.600	Bucentaur or craniofacial development	4.74135	9.8613
Q583J3	Tb927.4.2000	ruvB-like DNA helicase	5.70316	9.40665
Q38AF4	Tb927.10.8100	SET26	4.38096	8.99975
Q385I5	Tb927.11.5830	YL1 nuclear protein	4.51071	8.95025
Q581V4	Tb927.4.1270	ruvB-like DNA helicase	6.30647	8.89021
Q385D4	Tb927.11.6350	AAA ATPase (BDF7)	3.83483	6.00911
Q384T0	Tb927.11.7350	Histone H2BV	4.47655	5.79935
Q582M3	Tb927.7.6360	Histone H2AZ	5.32221	5.55756
Q388G3	Tb927.10.14390	FACT complex subunit POB3	4.00616	5.16272
Q387D1	Tb927.11.1210	Domain of unknown function (DUF4470)	4.0592	4.56921
Q583S0	Tb927.6.2640	importin alpha subunit	2.95249	4.54393
Q383E9	Tb927.11.11290	heat shock protein 70	5.58338	4.41933
Q386P3	Tb927.11.2670	Nucleoporin NUP59	2.14711	3.80024
Q580R3	Tb927.3.5620	Facilitates chromatin transcription complex subunit spt16	3.61713	3.74787
Q38AM1	Tb927.10.7420	BDF2	2.9484	3.65793
Q38D21	Tb927.9.14430	Casein kinase II	2.69264	3.60069
Q385P5	Tb927.11.5230	hypothetical protein protein	3.28779	3.24957
Q57XG4	Tb927.8.2250	Fungal tRNA ligase phosphodiesterase domain containing protein	2.20582	3.13698
Q386Y7	Tb927.11.1670	cysteine desulfurase	3.33443	3.08449
Q57Z31	Tb927.5.4170	Histone H4	3.9249	3.01404
Q57YA3	Tb927.7.2820	Histone H2A	3.46827	2.92822
Q386Y4	Tb927.11.1700	hypothetical protein	3.49597	2.77325
Q4GYX7	Tb927.1.2430	Histone H3	4.12096	2.61517
Q4GYK7	Tb927.1.3670	expression site-associated gene 8 (ESAG8) protein	2.35331	2.57668
Q389T1	Tb927.10.10590	Histone H2B	3.33103	2.55487
Q586Q9	Tb927.2.4580	UNC119	2.50076	2.23221
Q382J7	Tb927.11.13400	BDF5	2.29002	2.11787
Q387T9	Tb927.10.15750	Tripartite attachment complex protein 197	3.57696	2.07161
Q57ZX9	Tb927.5.3030	Intraflagellar transport protein 121	2.63744	2.74248

SET26 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q389G9	Tb927.10.11720	Znf-CW1	7.02945	12.2049
Q38AF4	Tb927.10.8100	SET26	4.22372	9.24843
Q384T0	Tb927.11.7350	Histone H2BV	2.89069	8.17785
Q582M3	Tb927.7.6360	Histone H2AZ	4.46897	6.16313
Q580R3	Tb927.3.5620	Facilitates chromatin transcription complex subunit spt16	3.63648	5.89062
Q385D4	Tb927.11.6350	AAA ATPase (BDF7)	3.91991	5.78368
Q388G3	Tb927.10.14390	FACT complex subunit POB3	3.37037	5.22017
Q387X7	Tb927.10.15350	Histone H3V	3.6525	5.00264
Q57Z31	Tb927.5.4170	Histone H4	3.36077	4.61553
Q383E9	Tb927.11.11290	heat shock protein 70	5.03661	4.61157
Q384B5	Tb927.11.9130	Replication factor A protein 1	3.37426	4.60043
Q57YA3	Tb927.7.2820	Histone H2A	4.0639	4.55846
Q386Y4	Tb927.11.1700	hypothetical protein	2.88207	4.39468
Q583H5	Tb927.4.3810	DNA-directed RNA polymerase II subunit 2 (RPB2)	2.33135	4.08113
Q386P3	Tb927.11.2670	Nucleoporin NUP59	3.46948	4.04731
Q38D92	Tb927.9.13320	hypothetical protein	2.47669	3.66993
Q389T1	Tb927.10.10590	Histone H2B	4.31282	3.59015
Q583S0	Tb927.6.2640	importin alpha subunit	2.30335	3.56287
Q38AE0	Tb927.10.8240	hypothetical protein	2.78086	3.54509
Q580N2	Tb927.3.4100	Pyruvate transporter	4.27969	3.47669
Q38A77	Tb927.10.8970	kinetoplast-associated protein 4 isoform 2	3.85018	3.44607
Q38E19	Tb927.9.9810	hypothetical protein	3.20778	3.29031
Q57YU2	Tb927.8.7400	RNA polymerase IIA largest subunit (RPB1)	2.15392	3.152
Q385C3	Tb927.11.6460	hypothetical protein	3.31061	3.00282
Q38FW0	Tb927.9.1980	hypothetical protein	3.74352	3.00098
Q38D21	Tb927.9.14430	Casein kinase II	4.09447	2.94132
Q57WL5	Tb927.8.8150	C2 domain containing protein	4.11178	2.93721
Q4GYX7	Tb927.1.2430	Histone H3	2.39318	2.84673
Q57ZS7	Tb927.5.2080	GMP reductase	2.82782	2.811
Q387Q3	Tb927.10.16120	inosine-5'-monophosphate dehydrogenase	3.08141	2.80256
Q388C7	Tb927.10.14770	Associated kinase of Tb14-3-3	2.07295	2.61634
Q388K3	Tb927.10.13960	paralyzed flagella protein 20	2.60087	2.52675
Q389A2	Tb927.10.12410	hypothetical protein	3.01927	2.48542
Q38DT1	Tb927.9.10920	kinetoplastid kinetochore protein 3 (KKT3)	2.78001	2.45683
Q584R4	Tb927.6.2420	p22 protein precursor	3.74551	2.4026
Q386D5	Tb927.11.3770	Dpy-30 motif containing protein	2.77261	2.29603
Q381T5	Tb927.11.15000	survival of motor neuron (SMN)-like protein	2.214	2.23543
Q38B98	Tb927.10.5030	ubiquitin/ribosomal protein S27a	3.55043	2.16014
Q580S5	Tb927.3.5500	DNA-directed RNA polymerase II subunit 3 (RPB3)	3.71295	2.12249
Q582W2	Tb927.3.2490	hypothetical protein	2.03244	2.09296
Q57X47	Tb927.5.4570	Flagellum adhesion protein 3	2.25438	2.09208
Q57YE9	Tb927.8.820	VID27 cytoplasmic protein	2.57063	2.0567

BDF2 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q38AM1	Tb927.10.7420	BDF2	4.94142	13.812
Q586J9	Tb927.2.2190	HDAC3	5.48101	12.925
Q582M3	Tb927.7.6360	Histone H2AZ	5.5163	8.77685
Q384T0	Tb927.11.7350	Histone H2BV	5.64211	8.65261
Q38E38	Tb927.9.9580	tubulin tyrosine ligase protein	4.66156	7.7985
Q582V9	Tb927.3.2460	hypothetical protein	4.9989	7.06729
Q383U2	Tb927.11.9870	Telomere-associated protein 1 (TelAP1)	4.1549	6.97577
Q586Z9	Tb927.6.4330	hypothetical protein	3.34214	6.9275
Q388W1	Tb927.10.12850	ttaggg binding factor (TRF)	5.35967	6.55814
Q57XY9	Tb927.3.1560	TRF-Interacting Factor 2 (TIF2)	4.40537	6.3738
Q38FD7	Tb927.9.4000	hypothetical protein	2.89011	5.81918
Q586Q9	Tb927.2.4580	UNC119	4.33193	5.70718
Q57UN4	Tb927.7.4040	hypothetical protein	3.53327	4.9661
Q385L3	Tb927.11.5550	DNA polymerase theta (POLQ)	3.27198	4.95997
Q57ZL0	Tb927.3.3020	actin-like protein	5.7851	4.94816
Q584P9	Tb927.6.2570	SUMO-interacting motif-containing protein	4.62419	4.90393
Q383K6	Tb927.11.10730	SWI/SNF-related helicase	3.65983	4.76426
Q389H2	Tb927.10.11690	YEATS family	3.27706	4.27104
Q38EB9	Tb927.9.8520	hypothetical protein	2.92994	4.2494
Q57YH0	Tb927.8.600	Bucentaur or craniofacial development	2.94405	3.88251
Q389D3	Tb927.10.12100	RNA-binding protein 7B	2.71219	3.49313
Q385T8	Tb927.11.4800	clathrin coat assembly protein	3.80316	3.47689
Q38BZ3	Tb927.10.2520	PrimPol-like protein 2	2.73316	3.47355
Q583J3	Tb927.4.2000	ruvB-like DNA helicase	3.61615	3.41918
Q580C6	Tb927.4.980	actin	3.02507	3.32672
Q38C43	Tb927.10.2000	actin-like protein	2.37492	3.31999
Q57WS6	Tb927.7.1240	Sphingosine kinase	2.11279	3.29644
Q581V4	Tb927.4.1270	ruvB-like DNA helicase	3.29543	3.20032
Q385E0	Tb927.11.6290	HIT zinc finger	2.49815	3.0698
Q57YW0	Tb927.8.3530	glycerol-3-phosphate dehydrogenase [NAD+], glycosomal	5.18765	2.94186
Q389G9	Tb927.10.11720	Znf-CW1	2.89967	2.89912
Q57ZS8	Tb927.5.2090	kinesin	3.26358	2.74008
Q57W25	Tb927.5.800	casein kinase I, isoform 2	3.7468	2.69126
Q389D9	Tb927.10.12030	hypothetical protein	3.67541	2.67071
Q384I8	Tb927.11.8310	class I transcription factor A, subunit 4 (CITFA-4)	2.451	2.64382
Q385A8	Tb927.11.6610	hypothetical protein	2.16774	2.44445
Q387B3	Tb927.11.1390	class I transcription factor A, subunit 1 (CITFA-1)	3.77834	2.28532
Q38DS7	Tb927.9.10960	ATP-dependent DEAD/H RNA helicase	4.30734	2.26913
Q57XM9	Tb927.7.2450	SUMO-interacting motif-containing protein	2.87997	2.2086
Q38D76	Tb927.9.13510	Poly(A)-specific ribonuclease PARN-3	2.66067	2.00854

HDAC3 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q586J9	Tb927.2.2190	HDAC3	5.46467	12.9502
Q57XY9	Tb927.3.1560	TRF-Interacting Factor 2 (TIF2)	5.1545	8.95426
Q586Z9	Tb927.6.4330	hypothetical protein	4.54883	8.81523
Q38AM1	Tb927.10.7420	BDF2	4.99548	8.60155
Q388W1	Tb927.10.12850	ttaggg binding factor (TRF)	4.01868	8.55698
Q383U2	Tb927.11.9870	Telomere-associated protein 1 (TelAP1)	3.9975	8.24039
Q38EB9	Tb927.9.8520	hypothetical protein	3.65687	7.71109
Q385L3	Tb927.11.5550	DNA polymerase theta (POLQ)	4.17948	7.53688
Q582M3	Tb927.7.6360	Histone H2AZ	4.33192	7.20786
Q582V9	Tb927.3.2460	hypothetical protein	3.9418	7.12454
Q38FD7	Tb927.9.4000	hypothetical protein	3.40403	7.00728
Q38BZ3	Tb927.10.2520	PrimPol-like protein 2	4.02346	6.39412
Q384T0	Tb927.11.7350	Histone H2BV	4.86712	6.21943
Q387L4	Tb927.11.370	repressor activator protein 1 (RAP1)	3.94259	5.0573
Q586Q9	Tb927.2.4580	UNC119	3.48225	4.21712
Q57UB5	Tb927.5.3510	structural maintenance of chromosome 3 (SMC3)	3.94742	4.09873
Q57W25	Tb927.5.800	casein kinase I, isoform 2	4.09069	4.08785
Q580Q4	Tb927.4.5310	Repressor of differentiation kinase 2	3.04617	3.66189
Q388S9	Tb927.10.13180	Nrap protein	2.48075	3.65437
Q4GZ82	Tb927.1.1370	rRNA biogenesis protein	2.54895	3.52187
Q38D76	Tb927.9.13510	Poly(A)-specific ribonuclease PARN-3	3.07138	3.34021
Q580T1	Tb927.3.5440	SNF2 DNA repair protein	2.29208	3.18214
Q38DK9	Tb927.9.11850	structural maintenance of chromosome 1 (SMC1)	2.34857	2.96831
Q385E2	Tb927.11.6270	inositol-1,4,5-trisphosphate (IP3) 5-phosphatase	2.02254	2.94661
Q57WS6	Tb927.7.1240	Sphingosine kinase	3.01619	2.87586
Q385P5	Tb927.11.5230	hypothetical protein	3.09052	2.2089
Q38DH5	Tb927.9.12300	replication factor C, subunit 3, putative	2.75055	2.00741

BDF4 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q38AE9	Tb927.10.8150	BDF1	4.188294496	10.46347491
Q57UR8	Tb927.7.4380	BDF4	4.899856567	9.580053965
Q38D21	Tb927.9.14430	Casein kinase II	2.61634521	4.563868841
Q383S2	Tb927.11.10070	BDF3	2.321038534	3.810703278
Q57YY8	Tb927.8.3250	dynein heavy chain	2.187183878	3.700003306
Q57V45	Tb927.3.3410	aspartyl aminopeptidase	2.343303395	3.043106079
Q383J7	Tb927.11.10820	Casein kinase II subunit beta	3.324763121	2.964817047
Q586Q9	Tb927.2.4580	UNC119	3.52554182	2.820368449
Q57VC7	Tb927.5.1360	nucleoside 2-deoxyribosyltransferase	2.708928634	2.726233164
Q38D92	Tb927.9.13320	hypothetical protein	2.267759854	2.656667074
Q382J7	Tb927.11.13400	BDF5	2.3335	2.2561
Q57UC1	Tb927.5.3450	eukaryotic translation initiation factor eIF2A	2.15945	2.01132

HDAC1 IP (BF) significantly enriched interactors (over untagged)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q38FV1	Tb927.9.2070	hypothetical protein	6.05488	11.7449
Q57TZ5	Tb927.7.1650	hypothetical protein	5.39599	11.4167
Q584W0	Tb927.6.3170	hypothetical protein	4.1002	10.7475
Q57WX2	Tb927.3.890	hypothetical protein	4.84634	9.67638
Q38C74	Tb927.10.1680	HDAC1	4.00622	7.22132
Q585H7	Tb927.4.3730	valosin-containing protein homolog	4.26412	6.34073

UP1 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q38D80	Tb927.9.13470	SET27	5.22095	12.3133
Q57X70	Tb11.v5.0267	Chromo1	5.07492	11.1154
Q4GYF3	Tb927.1.4250	UP1	3.56962	11.0978
Q382F5	Tb927.11.13820	UP2	4.76698	10.6183
Q582U8	Tb927.3.2350	UP4	3.78496	9.48828
Q382Z7	Tb927.11.11840	UP3	3.41063	7.83152
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	4.19571	7.40661
Q6Q833	unknown	Cytosolic leucyl aminopeptidase	4.32601	4.58063
Q57VD7	Tb927.5.1260	Sulfate transporter N-terminal domain with GLY motif	2.04922	4.19333
Q38BV7	Tb927.10.2880	Voltage-dependent calcium channel subunit	4.61807	4.1535
Q587H6	Tb927.2.2670	Histone H4V	2.3433	4.08871
Q386P3	Tb927.11.2670	Nucleoporin NUP59	3.58081	3.80248
Q38DT1	Tb927.9.10920	KKT3	2.36245	3.69693
Q38D98	Tb927.9.13240	hypothetical protein	3.65657	3.57762
Q38AJ5	Tb927.10.7690	Component of motile flagella 4	3.62145	3.57747
Q38BC1	Tb927.10.4800	hypothetical protein	2.3474	3.57494
Q4GYH2	Tb927.1.4050	ser/thr protein phosphatase	2.70693	3.48548
Q581R3	Tb927.4.1680	ZFP family member	3.08338	3.17001
Q387C9	Tb927.11.1230	hypothetical protein	3.80537	3.15644
Q38AJ4	Tb927.10.7700	ATP-binding cassette protein subfamily G	2.01695	3.0844
Q580T8	Tb927.3.5370	hypothetical protein	2.71627	3.03187
Q383I9	Tb927.11.10900	Component of motile flagella 9	3.32095	3.01354
Q38BK8	Tb927.10.3920	hypothetical protein	2.43833	2.97422
Q386J2	Tb927.11.3190	hypothetical protein	2.56917	2.95939
Q38EU8	Tb927.9.6310	ABC transporter	3.53322	2.94353
Q384A0	Tb927.11.9290	Flagellum attachment zone protein 20	3.22935	2.92033
Q57W25	Tb927.5.800	casein kinase I, isoform 2	3.77448	2.91854
Q388W2	Tb927.10.12840	mitochondrial carrier protein	3.96795	2.87731
Q386Q1	Tb927.11.2590	Flagellum attachment zone protein 12	2.02414	2.83273
Q38BB2	Tb927.10.4890	zinc-finger of a C2HC-type	2.79325	2.77737
Q582N5	Tb927.3.5020	Flagellar Member 6	2.36696	2.77578
Q57WR9	Tb927.7.1310	hypothetical protein	3.72993	2.7053
Q389M5	Tb927.10.11140	hypothetical protein	3.33023	2.46573
Q580Q1	Tb927.4.5340	Flagellum attachment zone protein 11	3.0201	2.45737
Q383Y5	Tb927.11.9440	Component of motile flagella 63	2.02543	2.43402
Q387T9	Tb927.10.15750	Tripartite attachment complex protein 197	2.35433	2.38792
Q384Y9	Tb927.11.6810	tubulin-tyrsoine ligase-like protein	2.4306	2.38668
Q38AU6	Tb927.10.6600	hypothetical protein	2.87595	2.3536
Q57W13	Tb927.8.4220	hypothetical protein	2.51882	2.35262
Q57XS0	Tb927.7.7000	hook complex protein	2.43208	2.28287
Q386T4	Tb927.11.2250	conserved protein, unknown function	2.26802	2.25032
Q57ZN8	Tb927.5.1690	hypothetical protein	2.17237	2.20315
Q587G2	Tb927.2.3020	Component of motile flagella 76	2.16063	2.14742
Q38A96	Tb927.10.8780	AAA domain containing protein	2.18827	2.13604
Q38F06	Tb927.9.5490	cation transporter	2.46722	2.09721
Q95PL4	unknown	multidrug resistance protein E	3.02632	2.06137
Q389Z2	Tb927.10.9840	chaperone protein DnaJ	3.32902	2.02277
Q57YE9	Tb927.8.820	VID27 cytoplasmic protein	2.71641	2.01695

UP2 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q382F5	Tb927.11.13820	UP2	5.79442	11.0665
Q582U8	Tb927.3.2350	UP4	4.40324	10.2328
Q4GYF3	Tb927.1.4250	UP1	5.38966	9.43585
Q38D80	Tb927.9.13470	SET27	4.76293	9.06901
Q57X70	Tb11.v5.0267	Chromo1	3.97649	8.339
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	5.19228	6.89575
Q57VD7	Tb927.5.1260	Sulfate transporter N-terminal domain with GLY motif	2.49465	6.63837
Q6Q833	unknown	Cytosolic leucyl aminopeptidase	3.6544	6.55041
Q382Z7	Tb927.11.11840	UP3	4.26211	6.28218
Q587H6	Tb927.2.2670	Histone H4V	4.21739	6.15441
Q38DT1	Tb927.9.10920	KKT3	3.47343	5.88261
Q383M7	Tb927.11.10520	KKT2	3.77198	5.72991
Q580T8	Tb927.3.5370	hypothetical protein	4.34004	5.72326
Q381K6	Tb927.11.15800	Cytokinesis initiation factor 1	4.0394	5.56799
Q26677	unknown	Flagellar calcium-binding protein TB-1.7G	2.47998	5.5266
Q383Y5	Tb927.11.9440	Component of motile flagella 63	3.33396	4.73828
Q57Z31	Tb927.5.4170	Histone H4	2.5465	4.73739
Q386P3	Tb927.11.2670	Nucleoporin NUP59	3.86537	4.66003
Q580B5	Tb927.4.870	dynein heavy chain	2.17212	4.64013
Q57TT0	Tb927.8.6830	kinesin	4.46155	4.57467
Q581R3	Tb927.4.1680	ZFP family member	3.51005	4.5018
Q384A0	Tb927.11.9290	Flagellum attachment zone protein 20	4.79768	4.48193
Q38AJ5	Tb927.10.7690	Component of motile flagella 4	3.3903	4.4792
Q389T1	Tb927.10.10590	Histone H2B	2.40209	4.43258
Q38BV7	Tb927.10.2880	Voltage-dependent calcium channel subunit	3.85415	4.33032
Q387E7	Tb927.11.1030	KKT7	2.91173	4.31726
Q38AX3	Tb927.10.6330	KKT1	3.11881	4.20842
Q57YA3	Tb927.7.2820	Histone H2A	2.09922	4.07256
Q38FW0	Tb927.9.1980	hypothetical protein	3.36766	4.02389
Q38D98	Tb927.9.13240	hypothetical protein	2.52044	4.02299
Q581S0	Tb927.4.1610	Tripartite attachment complex 40	3.41898	4.01079
Q57WX6	Tb927.3.930	dynein heavy chain	2.12186	3.87169
Q57W67	Tb927.8.3060	metallo-peptidase, Clan MF, Family M17	2.9884	3.86066
Q580R3	Tb927.3.5620	Facilitates chromatin transcription complex subunit spt16	2.84241	3.82469
Q57YY8	Tb927.8.3250	dynein heavy chain	2.56391	3.82188
Q387S9	Tb927.10.15850	Peroxisome biogenesis factor 12	2.24018	3.82072
Q38A04	Tb927.10.9720	RNA-editing-associated protein 1	3.73967	3.80872
Q4GYH2	Tb927.1.4050	ser/thr protein phosphatase	2.93372	3.78338
Q38DR1	Tb927.9.11140	hypothetical protein	4.23108	3.77187
Q38AJ4	Tb927.10.7700	ATP-binding cassette protein subfamily G	2.36676	3.76977
Q57WY5	Tb927.3.1020	Flagellum attachment zone protein 13	2.90707	3.7439
Q57U69	Tb927.5.3970	adenylate kinase	2.42205	3.64411
Q38FZ1	Tb927.9.1560	DnaJ domain containing protein	2.63277	3.63974
Q388V4	Tb927.10.12920	Flagellum attachment zone protein 18	3.96063	3.61645
Q580Y3	Tb927.8.3630	folate transporter	3.52676	3.5956
Q57XS0	Tb927.7.7000	hook complex protein	2.53628	3.55313
Q582N5	Tb927.3.5020	Flagellar Member 6	2.68204	3.53929
Q38EU8	Tb927.9.6310	ABC transporter	2.15738	3.5297
Q57WR9	Tb927.7.1310	hypothetical protein	2.37222	3.498
Q583L9	Tb927.6.3150	Hydin	3.42478	3.48424
Q387X7	Tb927.10.15350	Histone H3V	2.47892	3.46408
Q38DK9	Tb927.9.11850	structural maintenance of chromosome 1 (SMC1)	4.22275	3.42273
Q57UB5	Tb927.5.3510	structural maintenance of chromosome 3 (SMC3)	2.28828	3.32895
Q57Y68	Tb927.7.6610	hypothetical protein	3.91002	3.30217
Q381H7	Tb927.11.16090	Outer dynein arm docking complex protein 2	3.37634	3.28378
Q388Z4	Tb927.10.12490	kinesin	4.00989	3.28243
Q388G3	Tb927.10.14390	FACT complex subunit POB3	3.23755	3.27371
Q386J2	Tb927.11.3190	hypothetical protein	4.53935	3.24855
Q384T7	Tb927.11.7280	DNA-directed RNA polymerase II, subunit 9 (RPB9)	2.84572	3.24671
Q57VI3	Tb927.7.6280	Domain of unknown function (DUF3508)	3.48795	3.21732
Q586M2	Tb927.2.2510	hypothetical protein	2.53089	3.21484
Q57WH0	Tb927.7.3550	cytoskeleton associated protein	2.00533	3.21186
Q386T9	Tb927.11.2200	DNA-directed RNA polymerase II (RPB11)	2.53568	3.20711
Q385Y9	Tb927.11.4280	hypothetical protein	2.22384	3.2018
Q38FX8	Tb927.9.1710	hypothetical protein	2.6116	3.18024
Q57XC7	Tb927.8.2620	SUMO-interacting motif-containing protein	2.47393	3.1723
Q586D0	Tb927.2.5760	Flagellar Member 8	2.98058	3.17199
Q384T0	Tb927.11.7350	Histone H2BV	2.24285	3.15694
Q586Z9	Tb927.6.4330	hypothetical protein	4.20956	3.14999
Q57X85	Tb927.7.4690	hypothetical protein	2.45193	3.13593
Q38EE7	Tb927.9.8160	chaperone protein DnaJ	3.78677	3.12908
Q57Y65	Tb927.7.6640	hypothetical protein	2.17706	3.12278
Q57VT6	Tb927.7.900	hypothetical protein	2.15155	3.1209
Q389M5	Tb927.10.11140	hypothetical protein	2.35286	3.11251
Q57UX1	Tb927.8.4810	prohibitin 1	2.49432	3.10752
Q57X11	Tb927.8.5710	recombination initiation protein NBS1	3.50427	3.10545
Q386T4	Tb927.11.2250	hypothetical protein	4.14935	3.03731
Q38FX9	Tb927.9.1700	btb/poz domain containing protein	4.6686	3.03081
Q584J5	Tb927.4.5110	KKT8	2.8391	3.00299
Q382S5	Tb927.11.12560	hypothetical protein	2.83275	2.99863
Q38BK8	Tb927.10.3920	hypothetical protein	3.22303	2.99412
Q383I9	Tb927.11.10900	Component of motile flagella 9	3.16495	2.98693
Q38A96	Tb927.10.8780	AAA domain containing protein	2.78596	2.94916
Q386C6	Tb927.11.3880	actin	2.74896	2.92834

UP2 IP (BF) significantly enriched interactors (over untagged control) - continued				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q386Q1	Tb927.11.2590	Flagellum attachment zone protein 12	2.71943	2.9136
Q384B8	Tb927.11.9100	Domain of unknown function (DUF4586)	2.05727	2.9024
Q4GYE4	Tb927.1.4340	hypothetical protein	5.23694	2.89179
B3GVP4	unknown	VSG 427-8=VSG MITat 1.8	4.14501	2.88687
Q38BB2	Tb927.10.4890	zinc-finger of a C2HC-type	2.36917	2.88115
Q586Q4	Tb927.2.4520	hypothetical protein	2.54783	2.88041
Q580Q1	Tb927.4.5340	Flagellum attachment zone protein 11	2.36259	2.86048
Q38CI4	Tb927.10.560	40S ribosomal proteins S11	2.74136	2.84936
Q582U7	Tb927.3.2340	Peroxisome biogenesis protein 2	2.18448	2.84052
Q381S3	Tb927.11.15120	hypothetical protein	2.93598	2.82416
Q582M3	Tb927.7.6360	Histone H2AZ	2.25903	2.77984
Q580U8	Tb927.3.5270	DNA-directed RNA polymerases II subunit (RPB4)	2.12703	2.76284
Q38AM8	Tb927.10.7350	hypothetical protein	2.63186	2.65127
Q388E6	Tb927.10.14570	kinesin motor domain containing protein	2.011	2.6438
Q57ZN8	Tb927.5.1690	hypothetical protein	2.3709	2.61129
Q38A92	Tb927.10.8820	hook complex protein	2.82113	2.57494
Q580Y8	Tb927.8.3680	KKT4	3.36283	2.57251
Q38BC1	Tb927.10.4800	hypothetical protein	3.60889	2.54658
Q38E67	Tb927.9.9230	hypothetical protein	2.02218	2.54492
Q38EE1	Tb927.9.8240	hypothetical protein	3.20321	2.53871
Q383L7	Tb927.11.10620	hypothetical protein	2.90163	2.53326
P86938	Tb927.11.13640	J-binding protein 1 (JBP1)	2.63488	2.53216
Q38B41	Tb927.10.5630	hypothetical protein	2.29915	2.51424
Q57VP6	Tb927.8.1080	centrin	3.30648	2.49766
Q57XA3	Tb927.7.4870	hypothetical protein	3.0028	2.47192
Q382Q9	Tb927.11.12720	hypothetical protein	2.84744	2.47072
Q385K8	Tb927.11.5600	Archaic Translocase of outer membrane 14 kDa subunit	2.61065	2.44737
Q57YJ5	Tb927.8.5940	hypothetical protein	2.58303	2.43508
Q386I2	Tb927.11.3290	Tripartite attachment complex protein 166	2.60034	2.41968
Q38D28	Tb927.9.14290	Cytokinesis initiation factor 2	2.61055	2.3997
Q38C79	Tb927.10.1620	phosphoserine/threonine/tyrosine-binding protein	2.43146	2.39557
Q586H5	Tb927.6.5030	Component of motile flagella protein 46	2.44808	2.38433
Q57V00	Tb927.7.2110	KKT11	3.528	2.38368
Q57YU4	Tb927.8.7420	TerD domain containing protein	2.48737	2.36636
Q389U0	Tb927.10.10370	hypothetical protein	4.25275	2.36336
Q38FH5	Tb927.9.3560	hypothetical protein	2.06895	2.36123
Q4FKP3	Tb927.10.4310	prohibitin 2	2.54112	2.34555
Q57XW2	Tb927.3.1290	cullin 4B	2.51208	2.3094
Q384Q9	Tb927.11.7470	hypothetical protein	2.49797	2.28697
Q38FF7	Tb927.9.3760	Nucleoporin GLE2	3.7057	2.26391
Q57V66	Tb927.3.3200	Domain of unknown function (DUF4586)	2.29317	2.26136
Q38B80	Tb927.10.5230	ADP-ribosylation factor-like protein 13	2.13685	2.25931
Q38F31	Tb927.9.5220	conserved protein	2.96408	2.24629
Q57W72	Tb927.8.3010	hook complex protein	2.47276	2.24426
Q582V4	Tb927.3.2410	Peroxisome biogenesis factor 10	2.23194	2.22581
Q57YM4	Tb927.8.6230	hypothetical protein	2.04061	2.19427
Q38AP0	Tb927.10.7230	Flagellar Member 1	2.29404	2.17546
Q57VN9	Tb927.8.1150	KKT9	3.41616	2.17394
Q583W6	Tb927.4.2890	hypothetical protein	2.15338	2.126
Q57UW5	Tb927.8.4870	Component of motile flagella 6	2.00962	2.11748
Q580T2	Tb927.3.5430	hypothetical protein	3.0439	2.10735
Q57U64	Tb927.8.7790	zinc finger domain, LSD1 subclass	3.72446	2.08379
Q38AE6	Tb927.10.8180	hypothetical protein	4.07423	2.07314
Q384I0	Tb927.11.8390	hypothetical protein	2.02327	2.04373
Q386N3	Tb927.11.2770	hypothetical protein	2.41945	2.02705
Q57XM9	Tb927.7.2450	SUMO-interacting motif-containing protein	2.29054	2.00189

UP3 IP (BF) significantly enriched interactors (over untagged control)			$-\log_{10}$ Student's t-test p-value	$\log_2$ (tagged/untagged)
UniProt code	**TriTryp code**	**Description**		
Q4GYF3	Tb927.1.4250	UP1	6.11723	12.01
Q38D80	Tb927.9.13470	SET27	4.67356	11.155
Q57X70	Tb11.v5.0267	Chromo1	5.01216	11.1492
Q382F5	Tb927.11.13820	UP2	5.29297	9.26193
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	6.05653	8.57771
Q582U8	Tb927.3.2350	UP4	4.15155	7.89934
Q382Z7	Tb927.11.11840	UP3	3.86729	7.29641
Q38BS4	Tb927.10.3220	nucleolus protein	3.34187	4.18638
Q38BT9	Tb927.10.3070	hypothetical protein	3.14567	3.96175
Q38BC1	Tb927.10.4800	hypothetical protein	3.60454	3.95909
Q57WL5	Tb927.8.8150	C2 domain containing protein	3.98004	2.49688
Q57W25	Tb927.5.800	casein kinase I, isoform 2	3.6691	2.00944

UP4 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q382F5	Tb927.11.13820	UP2	5.12374	12.6508
Q582U8	Tb927.3.2350	UP4	4.37044	9.26017
Q4GYF3	Tb927.1.4250	UP1	5.61893	8.33027
Q57X70	Tb11.v5.0267	Chromo1	5.24134	8.24409
Q38D80	Tb927.9.13470	SET27	5.33522	8.15527
Q580T8	Tb927.3.5370	hypothetical protein	2.75592	6.89111
Q587H6	Tb927.2.2670	Histone H4V	2.85427	6.64952
Q6Q833	unknown	Cytosolic leucyl aminopeptidase	5.42113	6.58493
Q38DT1	Tb927.9.10920	KKT3	2.54149	6.36313
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	3.12381	6.34184
Q383M7	Tb927.11.10520	KKT2	4.22619	5.98738
Q382Z7	Tb927.11.11840	UP3	4.06131	5.93919
Q386P3	Tb927.11.2670	Nucleoporin NUP59	3.18287	5.32975
Q387E7	Tb927.11.1030	KKT7	4.45304	5.16458
Q57VD7	Tb927.5.1260	Sulfate transporter N-terminal domain with GLY motif	2.59296	4.97177
Q381U0	Tb927.11.14950	zinc finger protein 2	3.40909	4.61978
Q386Y4	Tb927.11.1700	hypothetical protein	3.16169	4.53993
Q57V00	Tb927.7.2110	KKT11	4.74217	4.40567
Q38A76	Tb927.10.8980	hypothetical protein	2.83586	4.40095
Q38FZ1	Tb927.9.1560	DnaJ domain containing protein	2.79216	4.30403
Q38E89	Tb927.9.8950	metallo- peptidase, Clan M- Family M48	3.57136	4.28239
Q57XG4	Tb927.8.2250	Fungal tRNA ligase phosphodiesterase domain containing protein	2.45051	4.20185
Q581S0	Tb927.4.1610	Tripartite attachment complex 40	3.82998	4.16823
Q38AX3	Tb927.10.6330	KKT1	2.24921	4.1283
Q38B56	Tb927.10.5480	60S ribosomal protein L24	4.4013	4.05029
Q57V66	Tb927.3.3200	Domain of unknown function (DUF4586)	2.63599	4.04718
Q57XF8	Tb927.8.2310	(H+)-ATPase G subunit	3.2614	4.03563
Q580R3	Tb927.3.5620	Facilitates chromatin transcription complex subunit spt16	3.32866	4.01882
Q57UB5	Tb927.5.3510	structural maintenance of chromosome 3 (SMC3)	2.92222	3.98553
Q38CG8	Tb927.10.720	Flagellum attachment zone protein 24	3.17951	3.95699
Q584J5	Tb927.4.5110	KKT8	3.3684	3.94628
Q38BV7	Tb927.10.2880	Voltage-dependent calcium channel subunit	2.54183	3.83234
Q387X7	Tb927.10.15350	Histone H3V	3.15485	3.76861
Q582U7	Tb927.3.2340	Peroxisome biogenesis protein 2	2.01564	3.72485
Q388K3	Tb927.10.13960	paralyzed flagella protein 20	2.33077	3.70703
Q580V3	Tb927.3.5220	hypothetical protein	2.14665	3.67733
Q388G3	Tb927.10.14390	FACT complex subunit POB3	2.67034	3.66906
Q388S2	Tb927.10.13250	hypothetical protein	2.5343	3.64925
Q38DN9	Tb927.9.11490	60S ribosomal protein L27a	2.48604	3.64224
Q57Z31	Tb927.5.4170	Histone H4	3.37477	3.61829
Q38DK9	Tb927.9.11850	structural maintenance of chromosome 1 (SMC1)	2.90066	3.60053
Q580Y8	Tb927.8.3680	KKT4	3.59373	3.57498
Q38A92	Tb927.10.8820	hook complex protein	2.41768	3.53761
Q384Y7	Tb927.11.6830	Domain of unknown function(DUF2779)	2.60012	3.52829
Q581B0	Tb927.8.3900	hypothetical protein	3.85654	3.50388
Q389S0	Tb927.10.10580	Histone H2B	3.89989	3.50026
Q57TT0	Tb927.8.6830	kinesin	3.32117	3.49855
Q386B1	Tb927.11.4030	hypothetical protein	2.24809	3.4811
Q388D4	Tb927.10.14700	hypothetical protein	5.12457	3.43337
Q585C6	Tb927.6.1210	KKT6	3.40277	3.42645
Q586U9	Tb927.2.4280	kinetoplastid-specific dual specificity phosphatase	3.87139	3.42052
Q38A04	Tb927.10.9720	RNA-editing-associated protein 1	2.57953	3.41831
Q389D9	Tb927.10.12030	hypothetical protein	4.12984	3.39346
Q57UC0	Tb927.5.3460	Flagellum attachment zone protein 16	2.13376	3.38419
Q38B88	Tb927.10.5140	Mitogen-activated protein kinase 6	2.09892	3.31585
Q386Q1	Tb927.11.2590	Flagellum attachment zone protein 12	2.8171	3.30506
Q38B18	Tb927.10.5860	hypothetical protein	2.95905	3.29446
Q38B68	Tb927.10.5350	dynein heavy chain	3.78648	3.27526
Q57X47	Tb927.5.4570	Flagellum adhesion protein 3	2.49959	3.239
Q580F6	Tb927.8.1680	KKT12	3.61701	3.23839
Q38BX8	Tb927.10.2670	hypothetical protein	2.95091	3.22556
Q384Y9	Tb927.11.6810	tubulin-tyrsoine ligase-like protein	4.58295	3.16925
Q383E9	Tb927.11.11290	heat shock protein 70	2.88909	3.1394
Q383P6	Tb927.11.10330	regulator of chromosome condensation 1-like protein	3.20062	3.08994
Q38BR6	Tb927.10.3310	Component of motile flagella 76b	4.17037	3.08308
Q57XU4	Tb927.7.7240	leucine-rich repeat protein (LRRP)	3.0372	3.08159
Q385G3	Tb927.11.6050	Flagellar-associated PapD-like	3.73873	3.06229
Q387B6	Tb927.11.1360	zinc-finger of a C2HC-type	3.12906	3.05748
Q38BB2	Tb927.10.4890	zinc-finger of a C2HC-type	3.39645	3.02259
Q38AG8	Tb927.10.7960	hypothetical protein	3.79617	3.00841
Q386J2	Tb927.11.3190	hypothetical protein	3.35924	3.00792
Q4GYJ5	Tb927.1.3820	ATP synthase subunit	2.49855	2.94254
Q57XY9	Tb927.3.1560	TRF-Interacting Factor 2	2.66516	2.94067
Q57VN9	Tb927.8.1150	KKT9	2.09685	2.86161
Q38DX7	Tb927.9.10350	Snf7	2.77962	2.85416
Q388E2	Tb927.10.14610	leucine-rich repeat protein (LRRP)	2.18323	2.84363
Q384B8	Tb927.11.9100	Domain of unknown function (DUF4586)	2.56171	2.8078
Q57YA3	Tb927.7.2820	Histone H2A	2.65002	2.80415
Q580D2	Tb927.4.1040	hypothetical protein	3.47945	2.80045
Q38DI2	Tb927.9.12170	leucine-rich repeat protein (LRRP)	2.11065	2.79942
Q581S1	Tb927.4.1600	AAA domain containing protein	2.0476	2.79503
P17545	unknown	DNA-directed RNA polymerase II subunit RPB1-B	2.46145	2.79142
Q38FE2	Tb927.9.3930	hypothetical protein	2.13453	2.75691
Q582N9	Tb927.3.4970	hypothetical protein	2.27455	2.75658
Q57XR0	Tb927.7.6900	Sister chromatid cohesion protein 1	2.5114	2.74704
Q38EU8	Tb927.9.6310	ABC transporter	4.07056	2.74136
Q57ZH0	Tb927.7.5190	hypothetical protein	2.6464	2.73139
Q38BT0	Tb927.10.3160	hypothetical protein	4.63982	2.73104
Q382L8	Tb927.11.13180	Present in the outer mitochondrial membrane proteome 10	2.66016	2.72331
Q38AY3	Tb927.10.6230	MORN repeat containing protein	3.13783	2.68292
Q382Q9	Tb927.11.12720	hypothetical protein	4.30305	2.65736
Q585L9	Tb927.6.1920	hypothetical protein	2.06179	2.6493

UP4 IP (BF) significantly enriched interactors (over untagged control) - continued				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q57ZR3	Tb927.5.1940	hypothetical protein	2.16095	2.64044
Q38DC0	Tb927.9.12990	hypothetical protein	2.92214	2.62969
Q57YR3	Tb927.8.7060	hypothetical protein	2.05764	2.62934
Q385P1	Tb927.11.5270	ubiquitin carboxyl-terminal hydrolase	3.30536	2.61905
Q57YB6	Tb927.7.3060	Tripartite Attachment Complex Protein 42	2.21702	2.60677
Q586Z9	Tb927.6.4330	hypothetical protein	2.11862	2.60475
Q57YU4	Tb927.8.7420	TerD domain containing protein	2.02282	2.58795
Q57W57	Tb927.5.1120	Phage tail fibre repeat	2.20607	2.58135
Q387G6	Tb927.11.840	sodium/hydrogen exchanger	2.13237	2.56998
Q38CF5	Tb927.10.850	hook complex protein	2.19531	2.56694
Q388G8	Tb927.10.14340	hypothetical protein	5.11497	2.53607
Q57XV3	Tb927.3.1200	hypothetical protein	2.37774	2.53054
Q385Y9	Tb927.11.4280	hypothetical protein	2.39934	2.50286
Q383J7	Tb927.11.10820	Casein kinase II subunit beta	3.10171	2.48472
Q38FB5	Tb927.9.4250	hypothetical protein	2.56263	2.46619
Q57Y65	Tb927.7.6640	hypothetical protein	2.23786	2.45738
Q57VT6	Tb927.7.900	hypothetical protein	2.48117	2.44676
Q38BV0	Tb927.10.2950	hypothetical protein	2.67553	2.41736
Q386T9	Tb927.11.2200	DNA-directed RNA polymerase II (RPB11)	2.52273	2.41572
Q388L3	Tb927.10.13860	GPI-anchor transamidase subunit 8 (GPI8)	2.4044	2.40205
Q385W2	Tb927.11.4550	Domain of unknown function (DUF4200)	3.23239	2.38746
Q57YN7	Tb927.8.6360	SPRY domain containing protein	3.21554	2.3694
Q388B1	Tb927.10.14950	Zinc finger CCCH domain-containing protein 40	2.56279	2.33777
Q38E83	Tb927.9.9060	Lsm12 protein	4.12741	2.30923
Q57YJ5	Tb927.8.5940	hypothetical protein	2.63581	2.302
Q583H5	Tb927.4.3810	DNA-directed RNA polymerase II subunit 2 (RPB2)	2.20242	2.29649
Q38A96	Tb927.10.8780	AAA domain containing protein	2.34879	2.26979
Q57YB7	Tb927.7.3070	UAA transporter family	2.47525	2.26311
Q57XU5	Tb927.7.7250	Ankyrin repeats (3 copies)	2.04203	2.25871
Q584K2	Tb927.4.5030	protein phosphatase 1	3.73734	2.24431
Q381D9	Tb927.11.16470	SUMO-interacting motif-containing protein	3.07945	2.23723
Q580Z9	Tb927.8.3790	paraflagellar rod component	2.42419	2.23019
P86938	Tb927.11.13640	J-binding protein 1 (JBP1)	3.98007	2.20626
Q384T0	Tb927.11.7350	Histone H2BV	2.48887	2.18045
Q57WE6	Tb927.7.3310	Radial spoke head 1 homolog	2.52389	2.17486
Q38AT9	Tb927.10.6670	dynein light chain	2.35856	2.14146
Q580X9	Tb927.8.3590	hypothetical protein	2.94055	2.12824
Q57WH0	Tb927.7.3550	cytoskeleton associated protein	3.3741	2.11994
Q386M3	Tb927.11.2880	kinesin	3.07633	2.09208
Q384T7	Tb927.11.7280	DNA-direcetd RNA polymerase II, subunit 9 (RPB9)	2.29297	2.08795
Q388H0	Tb927.10.14320	Flagellum attachment zone protein 9	2.02257	2.07634
Q583D4	Tb927.4.4490	Multidrug resistance protein E	2.01548	2.07212
Q382B6	Tb927.11.14210	conserved protein	2.21387	2.06744
Q585J9	Tb927.4.3490	DNA-directed RNA polymerases II and III subunit RPB6	3.16585	2.05721
Q38BC4	Tb927.10.4770	phosphatidylinositol-4-phosphate 5-kinase	2.68653	2.04154
Q386Y7	Tb927.11.1670	cysteine desulfurase	2.91115	2.02465
Q38B61	Tb927.10.5430	Flagellar C1a complex subunit C1a-32	2.57077	2.01609
Q38BI0	Tb927.10.4200	hypothetical protein	2.4594	2.00433
Q582M3	Tb927.7.6360	Histone H2AZ	2.79472	2.0039
Q381K6	Tb927.11.15800	Cytokinesis initiation factor 1	2.66271	2.00025

JBP2 IP (BF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	5.25552	11.0141
Q38BS4	Tb927.10.3220	nucleolus protein	4.94393	7.55358
Q4GYF3	Tb927.1.4250	UP1	5.22388	7.30672
Q57X70	Tb11.v5.0267	Chromo1	5.05146	6.629
Q38D80	Tb927.9.13470	SET27	7.29847	6.59239
Q382F5	Tb927.11.13820	UP2	3.60931	6.35231
Q38BT9	Tb927.10.3070	hypothetical protein	5.82335	5.42506
B3GVE0	Tb427.BES129.5	Antigen 5; Expression site-associated gene 5 (ESAG5) protein	5.396	4.58942
Q582U8	Tb927.3.2350	UP4	5.66093	4.35951
Q38CK3	Tb927.10.350	protein kinase PK4	2.73811	3.7404
Q382Z7	Tb927.11.11840	UP3	2.76527	3.58508
Q580S0	Tb927.3.5550	intraflagellar transport protein 27	4.25443	3.12221
D6XLL1	Tb927.8.4770	small GTP-binding protein Rab18	2.89017	2.9715

Chromo1 IP (PF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q4GYF3	Tb927.1.4250	UP1 (hypothetical protein)	4.98354	11.6932
Q38D80	Tb927.9.13470	SET27	4.94346	11.4728
Q57X70	Tb11.v5.0267	Chromo1	3.88754	10.4428
Q382F5	Tb927.11.13820	UP2 (hypothetical protein)	4.90783	9.05433
Q382Z7	Tb927.11.11840	UP3 (hypothetical protein)	4.30466	9.02126
Q582U8	Tb927.3.2350	UP4 (hypothetical protein)	5.49396	8.24288
Q586Q9	Tb927.2.4580	UNC119	3.52705	4.09544
Q387E2	Tb927.11.1080	Nucleolar protein 60	3.16545	3.69547
Q38CX9	Tb927.9.15290	CHAT domain containing protein	2.79381	3.59882
Q57W25	Tb927.5.800	casein kinase I, isoform 2	2.47271	3.54108
Q383Q5	Tb927.11.10240	ATP-dependent protease subunit HslV	3.0576	3.21513
Q38F03	Tb927.9.5520	ubiquitin carboxyl-terminal hydrolase	2.86507	2.89305
Q383P6	Tb927.11.10330	regulator of chromosome condensation 1-like protein	3.50459	2.88753
Q57YB8	Tb927.7.3080	Kinetochore interacting protein 4 (KKIP4)	2.8807	2.61397
Q38DC5	Tb927.9.12900	RNA polymerase-associated protein LEO1	2.10663	2.44683
Q587E7	Tb927.2.3370	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase	3.47591	2.20318
Q381Q0	Tb927.11.15350	RNA-binding protein	2.25431	2.113

SET27 IP (PF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
Q4GYF3	Tb927.1.4250	UP1 (hypothetical protein)	5.11774	12.0571
Q38D80	Tb927.9.13470	SET27	5.54302	11.5223
Q57X70	Tb11.v5.0267	Chromo1	4.17426	11.317
Q382F5	Tb927.11.13820	UP2 (hypothetical protein)	4.19839	10.0766
Q582U8	Tb927.3.2350	UP4 (hypothetical protein)	5.1661	9.44665
Q382Z7	Tb927.11.11840	UP3 (hypothetical protein)	4.50061	9.20653
Q586Q9	Tb927.2.4580	UNC119	4.08532	4.07789
Q38DK9	Tb927.9.11850	structural maintenance of chromosome 1 (SMC1)	4.0748	3.9085
Q38F03	Tb927.9.5520	ubiquitin carboxyl-terminal hydrolase	2.54232	3.85404
Q38EP6	Tb927.9.6920	hypothetical protein	2.55134	3.6399
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	3.05096	3.58027
Q38BJ2	Tb927.10.4080	hypothetical protein	2.49817	3.37049
Q57W25	Tb927.5.800	casein kinase I, isoform 2	2.46448	3.28149
Q57YB8	Tb927.7.3080	Kinetochore interacting protein 4 (KKIP4)	2.13365	3.10401
Q383T1	Tb927.11.9980	2-oxoglutarate dehydrogenase E1 component	2.85949	3.06799
Q585S8	Tb927.2.1810	chromatin-remodeling complex atpase chain ISWI	2.27408	2.38569
Q583Z3	Tb927.4.3160	dihydroxyacetone phosphate acyltransferase	2.77719	2.32668
Q383P6	Tb927.11.10330	regulator of chromosome condensation 1-like protein	3.35243	2.00515

JBP2 IP (PF) significantly enriched interactors (over untagged control)				
UniProt code	**TriTryp code**	**Description**	**-log$_{10}$ Student's t-test p-value**	**log$_2$ (tagged/untagged)**
D6XIZ6	Tb927.7.4650	J Biosynthesis Protein 2 (JBP2)	6.19273	9.72772
Q38BS4	Tb927.10.3220	nucleolus protein	3.96192	6.43148
Q38B40	Tb927.10.5640	Gem-associated protein 2	4.5201	4.83616
Q381T5	Tb927.11.15000	survival of motor neuron (SMN)-like protein	2.23751	3.76993
Q57X70	Tb11.v5.0267	Chromo1	3.90736	3.54498
Q38BT9	Tb927.10.3070	hypothetical protein, conserved	2.54363	3.23571
Q38D80	Tb927.9.13470	SET27	2.68779	3.11671
Q4GYF3	Tb927.1.4250	UP1	2.98071	2.76386
Q382F5	Tb927.11.13820	UP2	3.46696	2.1395
Q381P0	Tb927.11.15450	Basal body protein	2.65092	2.09708