Universidade de Lisboa

Faculdade de Medicina de Lisboa

ROLE OF HISTONE H1 IN CHROMATIN AND GENE EXPRESSION IN THE AFRICAN TRYPANOSOME: BROAD SKILLS, SPECIFIC FUNCTIONS?

Ana Pena

Tese especialmente elaborada para a obtenção do grau de Doutor em Ciências Biomédicas na Especialidade de Microbiologia e Parasitologia

2015

Universidade de Lisboa

Faculdade de Medicina de Lisboa

ROLE OF HISTONE H1 IN CHROMATIN AND GENE EXPRESSION IN THE AFRICAN TRYPANOSOME: BROAD SKILLS, SPECIFIC FUNCTIONS?

Ana Pena

Orientador: Dra. Luísa Figueiredo
Co-orientador: Dr. Christian Janzen

Tese especialmente elaborada para a obtenção do grau de Doutor em Ciências Biomédicas na Especialidade de Microbiologia e Parasitologia

JÚRI

Presidente:
Faculdade de Medicina – Universidade de Lisboa
Professor Doutor José Melo Cristino

Vogais:
Universidade de Würzburg, Alemanha
Professor Tim Nicolai Siegel, PhD

The Rockefeller University, EUA
Professora F. Nina Papavasiliou, Associate Professor

InstiMedicina Molecular – Faculdade de Medicina – Universidade de Lisboa
Doutor Rui Miguel Prudêncio Cunha Pignatelli
Doutora Liliana Mâncio Silva
Doutora Luísa Miranda Figueiredo (*orientadora*)

Faculdade de Medicina – Universidade de Lisboa
Professor Doutor Sérgio Alexandre Fernandes de Almeida

A impressão desta dissertação foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 22 de Setembro de 2015

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I thank my supervisor Luísa Figueiredo for leading me in this big adventure and challenge that is doing a PhD. Thank you for accepting me in your lab, for being a close supervisor, attentively guiding me through the way. I am grateful to your encouragement and advises, scientific and for 'life' which made me grow as a better scientist and person.

I also thank my co-supervisor Christian Janzen, for always being available for scientific discussions and for having patience for the bureaucratic issues.

To my thesis committee: Dr. Miguel Godinho, Dr. Maria Mota and Dr. Ana Pamplona for the scientific input and for 'teasing me' in good ways. I am special grateful to Dr. Ana Pamplona. It was by your hand that I engaged in the world of Science and started the way that leaded me to the present moment in my career.

Words are very little to thank all the friends I have made at IMM. Your words, smiles, enthusiasm were my source of courage and positive thinking each day. I specially thank to my colleagues (present and former): Leonor, Idálio, Daniel, Mafalda, Sandra, Filipa, Fabien, Xico, Fábio, Margarida and Helena. Our lab is unique, and only we know why! A warm embrace also to our nice neighbours from Miguel Prudêncio lab, particularly to my dear and old friend Jorge Santos.

To my flatmate Eliana Real, for all of our intellectually elevated conversations, very futile but healthy gossiping and most of all for being such a good and present friend.

To Maria and Idálio, for your help during the thesis writing and especially to Idálio for the support, care and understanding.

Finally, to my family and friends 'since ever' I am more grateful than words could say. Thank you for being proud of me, for believing me and for giving me your love.

ABBREVIATIONS

3'UTR	3' untranslated region
4sU	4-thiouridine
4TU	4-thiouracil
5'UTR	5' untranslated region
aa	amino acid(s)
AAT	animal African trypanosomiasis
ActD	actinomycin D
Amp^R	ampicilin resistance gene
BARP	brucei alanine-rich protein
base J	β-D-glucosyl-hydroxymethyluracil
BES	bloodstream expression site
Br-UTP	5-bromouridine-5'-triphophate
BSD^R	blasticidin resistance gene
BSF	bloodstream form
CDS	coding sequence
ChIP	chromatin immunoprecipitation
CITFA	class I transcription factor A
COIII	cytochrome oxidase subunit III gene
DCV	Vybrant® DyeCycle™ Violet
dsRNA	double-stranded RNA
EP	procyclin with Glu-Pro repeats
ESAG	expression site-associated gene
ESB	expression site body
FACS	fluorescence activated cell sorting
FAIRE	formaldehyde-assisted isolation of regulatory elements
FDR	false discovery rate

GPEET	procyclin with Gly-Pro-Glu-Glu-Thr repeats
GPI	glycosylphosphatidylinositol
GRESAG	gene related with ESAG
GSEA	gene set enrichment analysis
H1	histone H1
HAT	human African trypanosomiasis
HMGB	high-mobility group box protein
HR	homologous recombination
IC_{50}	half maximal inhibitory concentration
kDNA	kinetoplast DNA
LUC	luciferase gene
MACS	magnetic-activated cell sorting
MES	metacyclic expression site
MMS	methyl methanesulphonate
MNase	Micrococcal nuclease
mRNA	messenger RNA
MVSG	metacyclic variant surface glycoprotein
ND7	NADH dehydrogenase subunit 7 gene
PAG	procyclin-associated gene
PF	procyclic form
PI	propidium iodide
Pol I	RNA polymerase I
Pol II	RNA polymerase II
Pol III	RNA polymerase III
ppm	parts per million
PTEN	phosphatase and tensin homolog
PTM	post-translational modification
PTU	polycistronic transcription unit

PUR^{R}	puromycin resistance gene
qPCR	quantitative real-time PCR
Rag2$^{-/-}\gamma c^{-/-}$	recombinase activating gene 2 and common cytokine receptor γ chain double knock-out
RBP	RNA-binding protein
rDNA	ribosomal DNA
RNAi	RNA interference
RNA-Seq	RNA sequencing
rRNA	ribosomal RNA
siRNA	small interfering RNA
SL	spliced leader
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SSR	strand-switch region
TbASF1A	*Trypanosoma brucei* anti-silencing factor 1A
TbBDF3	*Trypanosoma brucei* bromodomain factor 3
TbCAF-1b	*Trypanosoma brucei* chromatin assembly factor 1b
TbDAC1	*Trypanosoma brucei* histone deacetylase 1
TbDAC3	*Trypanosoma brucei* histone deacetylase 1
TbDOT1A	*Trypanosoma brucei* disruptor of telomeric silencing 1A
TbDOT1B	*Trypanosoma brucei* disruptor of telomeric silencing 1B
TbELP3b	*Trypanosoma brucei* elongator protein 3b
TbFACT	*Trypanosoma brucei* facilitates chromatin transcription complex
TbHAT2	*Trypanosoma brucei* histone acetyltransferase 2
TbHAT3	*Trypanosoma brucei* histone acetyltransferase 3
TbISWI	*Trypanosoma brucei* imitation switch
TbKu70/Ku80	*Trypanosoma brucei* Ku heterodimer (70 kDa and 80 kDa proteins)
TbMCM-BP	*Trypanosoma brucei* mini-chromosome maintenance-binding protein

TbNLP	*Trypanosoma brucei* nucleoplasmin-like protein
TbNUP-1	*Trypanosoma brucei* lamin-like protein
TbORC1/CDC6	*Trypanosoma brucei* origin replication complex 1/cell division cycle 6-like protein
TbRAP1	*Trypanosoma brucei* repressor/activator of protein 1
TbRPB5z	*Trypanosoma brucei* RNA Pol I RPB5 variant subunit
TbRPB6z	*Trypanosoma brucei* RNA Pol I RPB6 variant subunit
TbSIR2rp1	*Trypanosoma brucei* Silent information regulator 2-related protein 1
TbTIF2	*Trypanosoma brucei* TRF-interacting factor 2
TbTRF	*Trypanosoma brucei* telomeric repeat-binding factor
TbU3	*Trypanosoma brucei* high-affinity uracil transporter 3
TDB	trypanosome dilution buffer
TDP-1	trypanosome DNA binding protein
TPE	telomere position effect
tRNA	transfer RNA
TSS	transcription start site
TTS	transcription termination site
UTP	uridine-5'-triphosphate
VSG	variant surface glycoprotein
β-tub	β-tubulin gene

RESUMO

O *Trypanosoma brucei* é um protozoário unicelular que vive na corrente sanguínea e espaços intersticiais de um hospedeiro mamífero e que é transmitido pela mosca tsétsé, o seu vetor de transmissão. Em humanos, o *T. brucei* causa a doença do sono. O *T brucei* tem um ciclo de vida complexo no qual o parasita passa por vários estadios de desenvolvimento. Dentro do mamífero, o parasita vive sob a forma de tripomastigotas sanguíneos (*bloodstream forms*; BSFs), e após ser transmitido para a mosca tsé-tsé, diferencia-se em tripomastigotas procíclicos (*procyclic forms*; PFs). O parasita *T. brucei* é um organismo que divergiu cedo na evolução e que possui muitas características pouco usuais para um eucariota. A sua transcrição é maioritariamente constitutiva e executada pela RNA polimerase II (Pol II)., enquanto que a RNA polimerase I (Pol I) transcreve não só os RNAs ribossomais, mas também genes codificantes de proteínas. Entre os genes transcritos por Pol I encontram-se os genes que codificam as glicoproteínas variantes de superfície (*variant surface glycoproteins*; VSGs) e as prociclinas. As VSGs e as prociclinas são proteínas muito abundantes que revestem a superfície dos BSFs e dos PFs, respetivamente, e cuja expressão é estritamente regulada ao longo do desenvolvimento do parasita.

Nos BFSs, os genes das VSGs são transcritos em locais de expressão subteloméricos (*bloodstream expression sites*; BESs). Existem cerca de 15 BESs no genoma, contudo apenas um está transcricionalmente ativo em determinado momento. Os restantes BESs necessitam de ser mantidos silenciados por forma a manter a expressão monoalélica de um único BES. Para evitar o reconhecimento pelo sistema imunitário do hospedeiro, periodicamente os parasitas mudam o gene *VSG* expresso, através de um mecanismo denominado de variação antigénica. Nos BSFs o loci das prociclinas também está parcialmente reprimido por forma a evitar a substituição das VSGs por prociclinas à superfície. Existem evidências de que em *T. brucei* alguns fatores epigenéticos regulam a ativação ou o silenciamento transcricional de Pol I nos BESs e loci das prociclinas. Para além disso, existem algumas variantes de histonas e modificações pós-traducionais de histonas que estão associadas com os locais de iniciação e terminação da transcrição por Pol II.

No núcleo de células eucarióticas o DNA encontra-se enrolado em volta de um octâmero de histonas formando o nucleosoma, a unidade básica da cromatina. A histona H1 (H1) liga-se ao DNA que entra e sai do nucleosoma, estabilizando a sua estrutura e contribuindo para o estabelecimento de estruturas de cromatina de ordem mais elevada. Em *T. brucei* a H1 é codificada por pelo menos 5 genes que, de acordo com a sua sequência proteica esperada, podem agrupar-se em 3 classes. No meu trabalho de doutoramento, tive por objetivo investigar as funções da H1 na estrutura da cromatina, na expressão génica e na transcrição de *T. brucei*. Para além disso, também foi estudada a importância da H1 nas infeções *in vivo* de *T. brucei* no hospedeiro mamífero.

Foram realizados estudos funcionais em que se depletaram todas as classes de H1 por RNA de interferência (RNAi). Na presença de tetraciclina, RNAs de dupla-cadeia são produzidos que induzem a depleção dos RNAs mensageiros (mRNAs) de H1. Usando imunoprecipitação de chromatina (*chromatin immunoprecipitation*; ChIP) associada a histona H3 e *formaldehyde-assisted isolation of regulatory elements* (FAIRE), detetou-se uma descondensação global de cromatina ao longo do genoma nos parasitas depletados de H1. No entanto, esta descondensação foi mais proeminente nos promotores de BES silenciados e nos loci de prociclinas, apesar de mais ligeiramente nestes últimos. Curiosamente, a sequenciação de RNA (RNA-Seq) revelou que a abertura global da cromatina devida à depleção de H1 não estava associada a alterações generalizadas de expressão génica mas pelo contrário à desrepressão de um subgrupo específico de genes. De cerca de 9000 genes, apenas 26 passarm a ser significativamente sobreexpressos, maioritariamente genes de BESs silenciados (inclusive VSGs silenciados) e genes de loci de prociclinas. Eu adaptei e apliquei a marcação metabólica de RNA com 4-tiouridina (4sU) em *T. brucei* para purificar e quantificar os transcritos nascentes produzidos nos loci transcritos por Pol I e Pol II. A marcação com 4sU traz vantagens em relação à técnica classicamente utilizada de *nuclear run–on*, por permitir a medição da transcrição *de novo* em células não perturbadas. A quantificação de RNA nascente marcado com 4sU demonstrou que a H1 é necessária para reprimir a transcrição de Pol I nos BESs silenciados e nos loci de prociclinas, mas aparentemente não o é para os genes transcritos por Pol II.

No seu conjunto, os nossos resultados suportam um modelo em que a H1 induz a compactação global de cromatina, mas desempenha funções mais específicas no que respeita à expressão génica, nomeadamente a regulação transcricional de genes silenciados de Pol I. Este modelo é consistente com a noção de que a transcrição policistrónica por Pol II, que ocorre na maior parte do genoma de tripanossomas, é constitutiva e a maioria dos genes são controlados ao nível pós-transcricional. Pelo contrário, existem evidências de que os genes transcritos por Pol I são regulados ao nível transcricional, consistente com a função da H1 no silenciamento da iniciação/elongamento da transcrição por Pol I.

Outro fenótipo observado em parasitas depletados de H1 foi um aumento da resistência a danos no DNA induzidos por metanossulfonato de metilo (MMS). Esta observação poderá indicar que a H1 desempenha funções mais gerais ao suprimir a recombinação homóloga e, talvez, a instabilidade genómica.

Interessantemente, os parasitas de *T. brucei* depletados em H1 revelaram uma aparente perda de fitness durante infeções *in vivo*. Em murganhos infetados com parasitas depletados em H1, a progressão da parasitémia (percentagem de parasitas no sangue) foi mais lenta e os murganhos sucumbiram mais tarde. Curiosamente, este fenótipo foi completamente abolido em murganhos severamente imuno-comprometidos que não possuem linfócitos T, B e NK, demonstrando que a fitness de parasitas depletados em H1 é perturbada porque, de alguma forma, estes parasitas tornam-se mais suscetíveis ao sistema imunitário do hospedeiro.

Coletivamente os nossos resultados mostram que a H1 de *T. brucei*, um parasita que divergiu cedo na evolução, tem capacidades abrangentes porque é necessária para compactar a cromatina globalmente ao longo do genoma, mas em termos de transcrição desempenha funções específicas porque inibe a transcrição de um grupo restrito de genes, nomeadamente de loci silenciados transcritos por Pol I. A regulação de um grupo específio de genes mediada por H1 foi também observada noutros eucariotas superiores e inferiores, o que sugere que apesar de ser a família mais divergente de histonas, as suas funções foram conservadas ao longo da evolução.

SUMMARY

Trypanosoma brucei is a unicellular protozoan parasite that lives in the bloodstream and interstitial spaces of a mammalian host and that is transmitted by the tsetse, an insect vector. In humans, it causes sleeping sickness. *T. brucei* has a complex life cycle that progresses through different developmental stages. Inside the mammal, the parasite lives as bloodstream forms (BSFs), which differentiates differentiates into procyclic forms (PFs) upon transmission to the insect vector. *T. brucei* is an evolutionarily early-branching organism that possesses many unusual features for an eukaryote. Most of its transcription is constitutive and driven by RNA polymerase II (Pol II) and RNA polymerase I (Pol I) transcribes not only ribosomal RNA genes, but also protein-coding genes, including genes encoding variant surface glycoproteins (VSGs) and procyclins. These are highly abundant proteins that cover the surface of the BSFs and the PFs, respectively, and whose expression is tightly developmentally regulated.

In BSFs, VSGs are transcribed in subtelomeric bloodstream expression sites (BESs). There are ~15 BESs in the genome, however only one is fully transcriptionally active at a time. The remaining BESs need to be kept silent to maintain monoallelic expression of BES. To evade the host immune system, parasites switch the expressed *VSG* from time to time by a mechanism called antigenic variation. In addition, in BSFs procyclin loci are also partially repressed to prevent illegitimate VSG replacement for procyclins. Some epigenetic factors have been found to regulate Pol I transcriptional activation/silencing at BESs and procyclin loci. Besides, a number of histone variants and histone modifications seem to be associated with Pol II transcription initiation and termination.

Inside the eukaryotic nucleus, DNA is wrapped around an octamer of histones that forms the nucleosome, the basic unit of chromatin. Linker histone or histone H1 (H1) binds to the DNA entering and exiting the nucleosome, stabilizing its structure and contributing to the establishment of higher-order chromatin structures. In *T. brucei*, H1 is encoded by at least five genes and, based on the predicted protein sequence, they can be split in three classes. In my PhD work I aimed at analyzing the roles of H1 in

chromatin structure, genome-wide gene expression and transcription in *T. brucei*. In addition, the importance of H1 for *T. brucei* mammalian infections *in vivo* was also addressed.

Functional studies were performed by simultaneously depleting all classes of H1 by RNA interference (RNAi). Using chromatin immunoprecipitation (ChIP) of histone H3 and formaldehyde-assisted isolation of regulatory elements (FAIRE), in H1-depleted parasites we detected a global chromatin opening across the genome, even though this happened at silent BESs promoters more prominently and procyclin loci, to a lesser extent. Curiously, RNA sequencing (RNA-Seq) revealed that global chromatin opening due to H1 depletion is not associated with widespread alterations in gene expression but rather with derepression of a specific subset of genes. Out of ~9,000 genes, only 26 were significantly upregulated, mostly from silent BESs (including silent *VSG*s) and procyclin loci genes. I adapted and used 4-thiouridine (4sU) metabolic labeling of RNA in *T. brucei* to purify and quantify nascent transcripts emanating from Pol I and Pol II-transcribed loci. 4sU-labeling has the advantage over classical nuclear run-on of measuring transcription *de novo* in unperturbed cells. Quantification of 4sU-labeled nascent RNA showed that H1 is necessary to repress Pol I transcription at silent BESs and procyclin loci but apparently not at Pol II genes.

Altogether our data strongly supports a model in which H1 induces global chromatin compaction, but in what concerns gene expression, H1 plays more specific functions, namely on transcriptional regulation of silent Pol I transcribed genes. This is consistent with the long-stand notion that polycistronic transcription by Pol II is constitutive and most genes are controlled essentially at the post-transcriptional level. On the other hand, Pol I genes are known to be regulated at the transcriptional level, consistent with a role of H1 in silencing Pol I transcription initiation/elongation.

An additional phenotype observed in H1-depleted parasites was its increased resistance to methyl methanoesulfonate (MMS)-induced DNA damage. This might indicate that H1 also plays a more general role such as suppressing homologous recombination and perhaps genomic instability.

Interestingly, H1-depleted *T. brucei* parasites revealed an apparent loss-of-fitness during *in vivo* infections. In mice infected with H1-depleted parasites, progression of parasitaemia (percentage of parasites in the blood) was delayed and mice succumbed later in infection. Curiously, this phenotype was completely abolished in highly immune-compromised mice that lack T, B and NK lymphocytes, demonstrating that fitness of H1-depleted parasites inside the host is impaired because, somehow, these parasites are more susceptible to host immune system.

Altogether our results show that H1 of the early divergent *T. brucei* has 'broad skills' because it compacts chromatin throughout the genome, but it plays 'specific functions' at the transcription level because it represses transcription of a limited cohort of genes, namely silent Pol I loci. H1-mediated regulation of a specific set of genes was also observed in other higher and lower eukaryotes, which suggests that despite being the most divergent family of histones, its functions were conserved throughout evolution.

1 INTRODUCTION

1.1 TRYPANOSOMA BRUCEI

1.1.1 EARLY EVOLUTIONARY ORIGINS

If origin of Life is one of the biggest mysteries in Biology, deciphering the evolution of the first eukaryotic lineages is not a minor enigma. At the early branches of the eukaryotic tree of life, many groups of protists have daunted phylogenetic positioning, diverging incredibly from the plant, fungi and animal branches. Despite its apparent simplicity as organisms, about 1,000 million years of life on Earth (Eme *et al.*, 2014) set the grounds for a tremendous diversification in lifestyles and the emergence of exceptional biological features among them. An emblematic example is the supergroup Excavata, phylum Euglenozoa to which the parasite *Trypanosoma brucei* belongs. Given its unique and divergent properties, the root of eukaryotic tree was even proposed to lie within Euglenozoa or between it and all other eukaryotes, which would place *T. brucei* as one of the earliest branching organisms. Independently of the veracity of this hypothesis, trypanosomes such as *T. brucei* stand as evolutionary highly distant organisms from the well-known eukaryotic groups of animals, fungi and plants (Cavalier-Smith, 2010).

T. brucei is a unicellular flagellate that parasitizes the bloodstream and interstitial spaces of various mammalian hosts following transmission by flies of the genus *Glossina,* also known as 'tsetse' flies. Within Euglenozoa, trypanosomes belong to a very particular clade, the Kinetoplastida, marked by the presence of special mitochondrial DNA, the kinetoplast or kinetoplast DNA (kDNA). This is a very large, usually disc-shaped mass consisting of an extensive network of interlocked circular DNA molecules, located adjacent to the basal body at the base of the flagellum. As its name indicates, kDNA arose uniquely in these organisms (Adl *et al.*, 2012) (Fig. 1).

Trypanosomatida, such as *T. brucei*, offer a rare opportunity to study the origin and evolution of parasitism. This is because trypanosomatids were the only kinetoplastids acquiring an obligatory parasitic lifestyle, developing specific

adaptations to succeed in a lifetime co-existence with their hosts (Lukes *et al.*, 2014). Researcher's attention has focused particularly on trypanosomatids that represent a burden for public health and agricultural economy, namely species that infect humans and livestock from the Trypanosoma and Leishmania genus. These trypanosomatid species share their life between a variety of mammalian hosts and insect vectors and might develop intracellularly in host cells at some stage of its life cycle, such as *T. cruzi* and *Leishmania* spp., or, as in the case of *T. brucei*, live exclusively as extracellular parasites.

My PhD work focused on the study of *Trypanosoma brucei*. *T. brucei* holds a number of unusual, some unique, biological characteristics, which can go from transcription and nucleus organization to cell cycle, metabolism and surface proteins. While some are shared by most trypanosomatids, others, such as antigenic variation of its protective coat of variant surface glycoproteins (VSGs), are exclusive to African trypanosomes and arose as extremely efficient adaptations to life inside its hosts. The availability of a fully sequenced genome (Berriman *et al.*, 2005) and of a variety of molecular tools, including an RNA interference (RNAi) system (Ngo *et al.*, 1998), has made of *Trypanosoma brucei* a favorite model to gain insights not only about the biology, evolution and disease mechanisms in trypanosomatids, but also for fundamental questions in eukaryotic cell biology. Indeed, many important discoveries in eukaryotic molecular biology were made in cultured *T. brucei* such as RNA editing (Benne *et al.*, 1986), *trans*-splicing (Boothroyd & Cross, 1982) and glycosylphosphatidylinositol (GPI) anchors (Ferguson *et al.*, 1988); also, the hypermodified DNA base J (Gommers-Ampt *et al.*, 1993), the glycosomes (Opperdoes & Borst, 1977) and, quite recently, an unconventional set of kinetochore proteins (Akiyoshi 2014) were first identified in *T. brucei*. Besides, this parasite has been used as a privileged model organism for studying the eukaryotic flagella (Baron *et al.*, 2007, Branche *et al.*, 2006).

1.1.2 AFRICAN TRYPANOSOMIASIS

African trypanosomiasis encompasses a variety of human and animal pathologies caused by infection with different *Trypanosoma* species, after transmission through the bite of an infected tsetse fly (Gibson, 2007). It affects mainly sub-Saharan Africa, where suitable environmental conditions exist for the presence of the tsetse fly vector (*Glossina* spp.) (WHO, 2013). Among African trypanosomes, only *Trypanosoma brucei* is pathogenic for humans, causing human African trypanosomiasis (HAT), more specifically two of its subspecies: *T. b. gambiense* and *T. b. rhodesiense*. A third *T. brucei* subspecies, *T. b. brucei* is non-infective for humans, but parasitizes domestic and wild animals, causing a disease called Nagana. Because of its proximity to the human infective subspecies, *T. b. brucei* has been widely used as an experimental model of HAT in murine and rat animal models (Giroud *et al.*, 2009, Mogk *et al.*, 2014).

Clinically, HAT develops in two stages. In the first or early stage, parasites live in the lymphatic system and bloodstream. After a variable period, a late or second stage starts when the trypanosomes invade the central nervous system. This late stage is accompanied by progressive neurological damage leading to neuropsychiatric, motor and sensory disturbances that eventually result into coma and death (WHO, 2013, Brun *et al.*, 2010). The disease owes its name of 'sleeping sickness' due to the sleep disorders long known to characterize most late-stage HAT patients. Sleep abnormalities include a deregulation of the circadian rhythm of the sleep/wake cycle characterized by periods of nocturnal insomnia and daytime sleep, episodes of uncontrollable sleep, and fragmentation of the normal sleep structure (Blum *et al.*, 2006, Buguet *et al.*, 2005).

HAT is usually considered lethal if left untreated; however sporadic cases of natural progression to asymptomatic carriage or even apparent spontaneous resolution of the infection have been reported for gambiense HAT, resembling the trypanotolerance phenomena described for some African cattle species (Jamonneau *et al.*, 2012, Guirnalda *et al.*, 2007).

Although the pathologies caused by gambiense and rhodesiense subspecies are included under the general term of human African trypanosomiasis, or sleeping

sickness, they have different epidemiology and clinical progression. *T. b. gambiense* is found in Western and Central Africa and occurs mainly in humans, accounting for the vast majority of HAT cases, about 98% in the last decade (WHO, 2013). It is characterized by a chronic progression with an average duration of almost 3 years (Seyffer *et al.*, 2012). By contrast, *T. b. rhodesiense* is found in Eastern and Southern Africa and generally causes an acute diseases that rapidly progresses into second-stage, leading to death within 6 months (Oguchi *et al.*, 2012). However, it is less adapted to human beings, affecting mostly animals (livestock and wildlife), which act as dangerous reservoirs of the parasite (WHO, 2013).

The resurgence of HAT epidemics in the 1980's and 1990's alerted the international community for the risks of neglecting this disease. Since then, an international coordinated effort lead by the World Health Organization reinforced control and surveillance programs which lead to a drop in HAT cases since 2009 to less than 10,000 new cases per year (WHO, 2013). Nevertheless, it is believed that a substantial number of cases remain unreported and more recent estimates predict that 70 million people in tropical Africa are at risk for HAT (WHO, 2013).

As a whole, African trypanosomiasis represents a serious health and socio-economic problem. Besides the severe life threat posed by HAT, animal African trypanosomiasis (AAT), or Nagana causes around 3 million cattle deaths and total annual losses of US$ 4.75 billion. AAT is an incredibly heavy burden particularly in rural areas where livestock production is the main livelihood, perpetrating underdevelopment and poverty in these regions (FAO, 2014).

In humans, a major obstacle for eradication of HAT is the insufficiency of the existent treatments. No prophylactic treatment or vaccine is available and the drugs used to treat early and late-stage HAT are in their majority not accessible orally, often very toxic, and sometimes ineffective (Kennedy, 2013). While some new promising drugs are on the pipeline (e.g. fexinidazole and oxaborole SCYX-7158) (http://www.dndi.org) (Kennedy, 2013), investment on the search for more effective treatments has been set as a priority. That, together with continued surveillance, improved diagnostics and vector control policies will be vital to meet World Health

Organization's target to eliminate HAT as a public health problem by 2020 (WHO, 2015).

1.1.3 A CHALLENGING LIFE - FROM MAMMAL TO FLY

As for the other trypanosomes, *Trypanosoma brucei* is a heteroxenous parasite i.e. requires more than one obligatory host to complete its life cycle: a mammalian host and the blood-feeding tsetse fly vector. Infection in the mammal starts with a bite of an infected tsetse that inoculates the infective cell cycle-arrested metacyclic stage into the mammal bloodstream and draining lymphatics. *T. brucei* then differentiates to a long slender bloodstream form (BSF) that divides asexually by binary fission. In the mammal, parasites can live extracellularly in the blood or in the extravascular fluids and interstitial spaces of organs and tissues. At high parasite density, a yet unidentified parasite-released factor induces differentiation of long slender forms to short stumpy BSFs, passing through an intermediate form (Vassella *et al.*, 1997, Reuner *et al.*, 1997). Stumpy forms are non-dividing cells and the only life cycle stage transmissible to the tsetse, which can occur during a blood meal of the fly.

Once in the fly midgut, the stumpy forms differentiate into procyclic forms (PFs), which re-enter the cell cycle, actively multiplying and colonizing the fly gut. PFs undergo a complex differentiation process while migrating to the fly proventriculus. During this process each PF gives rise by an asymmetric cell division to one long and one short epimastigote; the long epimastigotes are not able to proceed in development while the short epimastigotes migrate to the tsetse salivary glands. Here, they attach to epithelial cells and start to multiply. Attached epimastigotes are the only life cycle stage known to perform meiosis (Peacock *et al.*, 2011). Finally, in the salivary glands, epimastigotes differentiate to metacyclic forms, ready to be transmitted back to the mammalian host by a tsetse bite (Sharma *et al.*, 2009) (Fig. 1).

Throughout its life cycle, *T. brucei* exhibits a series of specific adaptations that allow it to cope with the completely different environments it encounters in the mammal and the fly. The several life cycle stages reflect the alterations undergone by the parasite to be best adapted to the host's environment while balancing survival,

proliferation and transmissibility. The process of differentiating from one life cycle stage to the other is complex and implicates deep changes in cell morphology and ultrastructure, metabolism, cell cycle and cell surface proteins, accompanied by a strong reprogramming in gene expression (Kramer, 2012).

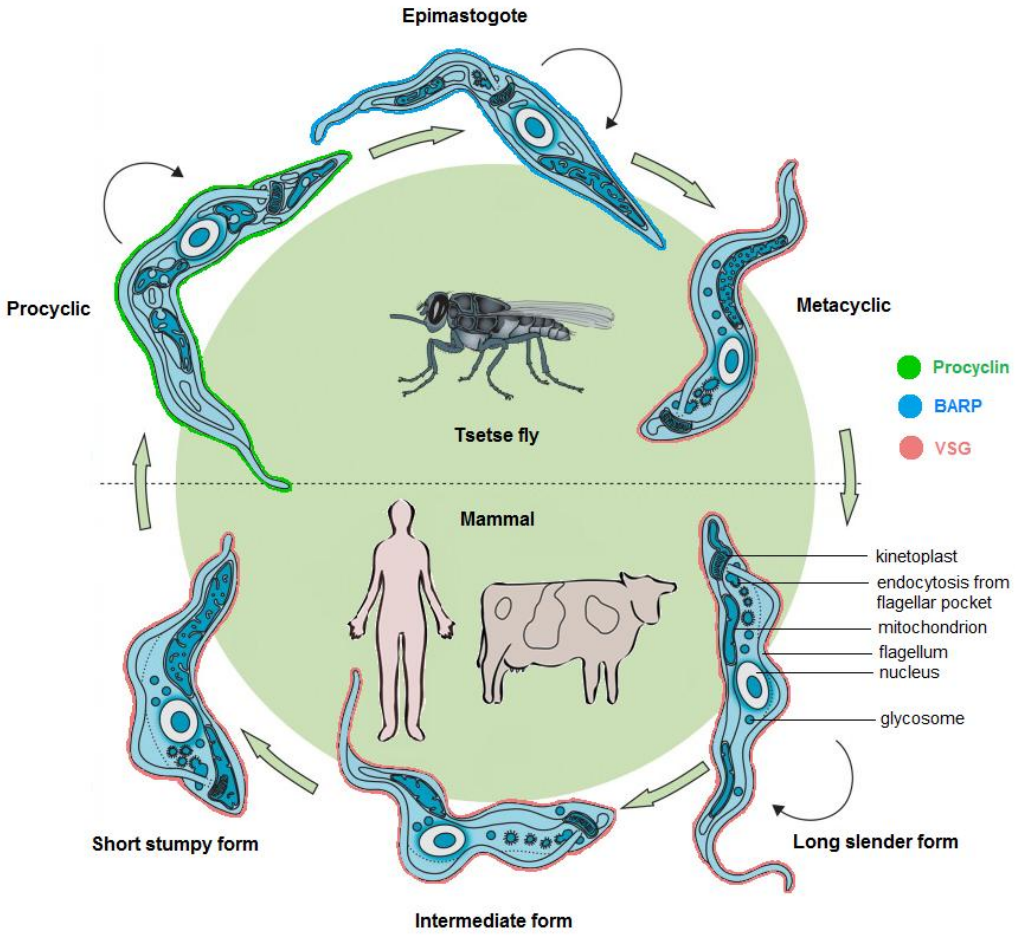


Figure 1. Life cycle of African trypanosomes. In the bloodstream of the mammalian host, parasites exist as a polymorphic population of bloodstream forms (BSFs) consisting of dividing (black arrows) slender forms and cell cycle arrested intermediate and stumpy forms. In the tsetse fly vector, after entering the midgut, stumpy forms differentiate to the procyclic forms (PFs), which further develop to the migrating epimastigotes. In the salivary glands, the latter transform into the infective metacyclic forms, which are injected during the next blood meal of the fly into the mammalian host (adapted from Brun *et al.*, 2010).

So far, the best-studied differentiation pathway in *T. brucei*, and in kinetoplastids in general, is the differentiation of BSFs to PFs. Five recent studies brought to light the genome-wide changes in gene expression between these two developmental stages identifying a significant percentage of genes that are differentially expressed between them (6–40 % of all genes analyzed, depending on

the technology used) (Jensen *et al.*, 2009, Kabani *et al.*, 2009, Nilsson *et al.*, 2010, Siegel *et al.*, 2010, Veitch *et al.*, 2010, Queiroz *et al.*, 2009). This reflects the extraordinary parasite capacity of rapidly remodeling its gene expression pattern upon sensing the environmental changes that take place when switching hosts.

One of the fundamental changes occurring when *T. brucei* alternates between the mammal and the insect is the dramatic change of its surface protein composition. Inside the fly, PFs are coated with highly acidic and repetitive proteins, the procyclins (GPEET and EP procyclins), (Roditi *et al.*, 1989, Clayton & Mowatt, 1989, Vassella *et al.*, 2001) which are later exchanged for a coat of brucei alanine-rich protein (BARP) coat in epimastigotes (Urwyler *et al.*, 2007). Further differentiation into the metacyclic stage replaces BARPs for a dense coat of metacyclic variant surface glycoproteins (MVSGs). After transmission to the mammal, a similar type of VSG (Cross, 1975) is maintained as the major surface protein in the BSF slender and stumpy forms of *T. brucei* (Vickerman, 1985). When stumpy forms differentiate to PFs, the VSG coat is shed and replaced by procyclins (Roditi *et al.*, 1989) (Fig. 1).

As extracellular parasites, one of the major challenges that *T. brucei* is confronted with are the host's natural defenses. The cell surface of the parasite lies at the host-parasite interface, serving as a first line of defense against the attacks mounted by the host. In PFs, the procyclin coat is crucial for parasite survival in the tsetse midgut (Ruepp *et al.*, 1997, Acosta-Serrano *et al.*, 2001). On the other hand, in BSFs the very dense coat of VSGs is crucial for protecting the parasite against the mammalian immune system. VSGs are molecules of approximately 55-65 kDa (Cross, 1975), which are present at the parasite cell surface as homodimers and are attached to the membrane by a GPI moiety (Ferguson *et al.*, 1988). A stunning number of VSG homodimers are tightly and orderly packed at the parasite surface: ~10^7 VSG molecules per parasite (Jackson *et al.*, 1985), which account for about 90% of its surface proteins (Grunfelder *et al.*, 2002). Although the VSG family displays extensive sequence variation, the secondary and tertiary structure of these molecules is highly conserved (Blum *et al.*, 1993, Carrington & Boothroyd, 1996), maybe to ensure that a dense VSG monolayer is always formed.

The VSG coat is a remarkable protective barrier against the innate and the adaptive immune system of the mammal. VSG protects parasites against recognition and lysis by the alternative complement pathway (Ferrante & Allison, 1983) and from anti-VSG antibody-mediated phagocytosis (Guirnalda *et al.*, 2007, Pan *et al.*, 2006). At low antibody titres, protection against antibody-mediated killing results in part from clearance of antibody-VSG complexes from the parasite surface, facilitated by the high rates of endocytosis and surface recycling existent in *T. brucei* (Barry, 1979, Natesan *et al.*, 2007) and the hydrodynamic forces generated by its flagellar movement (Engstler *et al.*, 2007). Additionally, the dense VSG coat is probably shielding many other cell-surface molecules from immune recognition such as invariant or less variable protein domains such as the VSG C-terminal domain (Schwede *et al.*, 2011). There is also evidence that VSG shed from the membrane contributes to immune suppression of both B- and T-cell activation (Mansfield & Paulnock, 2005).

Even though all these phenomena cooperate to defend the parasite against the host immune response, *T. brucei* accomplishes immune evasion primarily through another exceptional strategy: antigenic variation of the VSG. As it will be discussed next, antigenic variation is as far as we know the leading process that allows the *T. brucei* parasite to thrive against the immune attack and persist in the bloodstream of the mammal for prolonged periods.

1.1.4 ANTIGENIC VARIATION - THE POWER OF CHANGING

Paradoxically, although the VSG coat is essential for immune evasion, it is extremely immunogenic. Most VSGs possess an N-terminal ‘variable’ domain, with high sequence divergence and which contains the epitopes exposed at the surface (Miller *et al.*, 1984, Cross, 1984). The VSG C-terminal domain is generally more conserved and encodes the subsurface GPI-anchor signal (Cross, 1984, Carrington *et al.*, 1991). The VSG epitopes strongly elicit an antibody response in the mammal (Morrison *et al.*, 1982, Black *et al.*, 2010). The parasite has the capacity to evade antibody-mediated killing mostly via antigenic variation of VSGs.

Typically during a *T. brucei* infection, VSG epitopes are recognized and activate host B-cells to produce VSG-specific antibodies against the 'major', predominantly expressed VSG in the population. The anti-VSG antibodies produced are largely responsible for clearing most of the parasites, leading to a drop in parasitaemia, i.e. in parasite density in the blood (Black *et al.*, 2010). A few parasites in the population have switched their VSG and represent 'minor' VSG variants that have not been recognized yet by the immune system; as a result, these parasites will successfully replicate and substitute the previous parasite population. As prior VSG variants are recognized by the immune system and cleared, newly switched variants emerge, giving rise to the characteristic periodic waves of parasitaemia, with several infection peaks which reflect the episodes of parasite clearance and relapse, described for the first time in humans more than a century ago (Ross & Thomson, 1910) (Fig. 2).

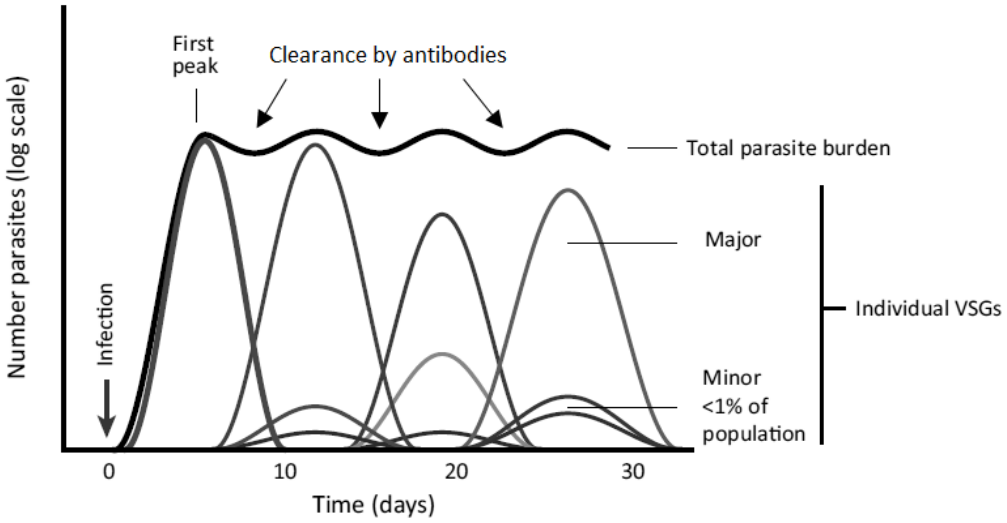


Figure 2. Antigenic variation during *T. brucei* infection. During infection the overall parasitaemia, i.e. 'total parasite burden' in the blood varies in a wave-like manner (upper black line). Peaks correspond to maximal parasite density and valleys represent massive parasite clearance primarily by the action of VSG-specific antibodies produced by the host. A small number of VSG variants ('major') dominate the infections peaks, but the entire number of variants expressed at any time is greater and there may be many VSGs present in low abundance ('minor'). Except for the 'total parasite burden', each line corresponds to different VSGs and peaks only once. The first peak is frequently dominated by the VSG most expressed in the initial population (day 0) (adapted from McCulloch & Field, 2015).

Antigenic variation relies on two key features: VSG monoallelic expression and VSG switching. Monoallelic expression consists in the expression of a single *VSG* gene

out of a genomic repertoire of ~2,000 *VSG* genes encoding different variants of the protein (Cross *et al.*, 2014). Therefore, each parasite expresses only one VSG variant at the surface at a time. The active *VSG* gene is expressed from a specialized subtelomeric locus called a bloodstream expression site (BES). Of the ~15 BESs present in the genome (Hertz-Fowler *et al.*, 2008), only one BES is transcriptionally active, while the others are silent.

VSG switching can occur via different molecular mechanisms. The most frequent mechanism is the activation of a silent *VSG* by homologous recombination (HR) into the active BES (Robinson *et al.*, 1999). Recombination is the only way to access the *VSG* repertoire of ~2000 genes that lie mainly in subtelomeric non-BES loci The most frequent HR switching mechanism is the so-called duplicative gene conversion. This involves duplication and insertion of an inactive telomeric or array *VSG* into the active BES, replacing the previously active *VSG* gene. A new *VSG* might also be assembled by segmental gene conversion in which segments of *VSG* genes and pseudogenes are copied and recombine to generate novel functional 'mosaic' *VSGs* that will replace the previously active *VSG*. In addition, HR can also mediate a telomere exchange in which reciprocal crossover between two telomeres (no DNA sequence is lost) and their associated *VSGs* takes place.

VSG switching can also be recombination-independent. This less common mechanism consists of a transcriptional activation of a different BES and silencing of the previous one with no involvement of DNA rearrangements by recombination. This is termed *in situ* switch (McCulloch *et al.*, 2015, Rudenko, 2011). In contrast with BSFs, *in situ* transcriptional activation is considered to be the only route by which *MVSGs* are activated in metacyclic parasites (Barry & McCulloch, 2001).

Curiously, trypanosome infections typically show hierarchical expression of individual *VSGs* in which some variants appear early in infection while others appear progressively later (Gray, 1965, Miller & Turner, 1981, Morrison *et al.*, 2005, Marcello & Barry, 2007). This sequential expression seems to depend on the differential activation probabilities of each VSG, which are dictated in part by the type of genetic locus they occupy. During infection, the first *VSGs* to be switched on are intact

telomeric *VSGs* resident in BESs and minichromosomes followed by inactive *VSGs* from subtelomeric arrays, and subsequently by 'mosaic' *VSGs* (Robinson *et al.*, 1999, Morrison *et al.*, 2005, Aitcheson *et al.*, 2005, Marcello & Barry, 2007, Lythgoe *et al.*, 2007, Hall *et al.*, 2013, McCulloch *et al.*, 2015).

We know that one infection peak can actually be composed by mixtures of subpopulations expressing more than one VSG variant (Miller & Turner, 1981, Hajduk & Vickerman, 1981, Robinson *et al.*, 1999, Barry & McCulloch, 2001) (Fig. 2). Recent studies observed that although a small number of VSG variants dominate in each peak of infection, there was a higher than expected diversity of total VSG variants being expressed at any given time (an average of 20-30 VSG variants/peak). As a result, probably most of the intact *VSG* repertoire is exhausted rather early in infection and recombinatorial mechanisms such as mosaic formation should be critical to expand the *VSG* repertoire. Perhaps the assembly of mosaic *VSGs* can even be the predominant switch mode in nature, where infections are sustained for long periods (Hall *et al.*, 2013, Mugnier *et al.*, 2015).

It is important to note that VSG switching does not depend on the immune response to occur because *T. brucei* still switches VSG in axenic cultures (Doyle *et al.*, 1980). Nevertheless, switching rates are noticeably higher (10^{-2}-10^{-5} switches/trypanosome/generation) in strains recently isolated from nature or fly-transmitted which are 'pleomorphic', i.e. that retain the capacity to differentiate into stumpy forms and are tsetse transmissible (Turner & Barry, 1989, Turner, 1997). By contrast, in laboratory-adapted 'monomorphic' strains, which have lost the ability to naturally differentiate into the stumpy stage and complete the life cycle, switching frequencies are typically low (10^{-4}-10^{-7} switches/trypanosome/generation) (Lamont *et al.*, 1986, Aitcheson *et al.*, 2005, Boothroyd *et al.*, 2009). This supports the view that VSG switching is a stochastic process in which antibodies act chiefly a selective force, rather than a trigger.

Antigenic variation is essential for parasite survival. However, it needs to be balanced in order ensure infection chronicity. While VSG switching needs to occur at a frequency that ensures infection persistence, the appearance of new VSG variants

should be kept under certain limits to avoid *VSG* repertoire exhaustion or host immune system overwhelming (Gjini *et al.*, 2010). The interplay between antigenic variation and stumpy differentiation seems critical for this balance. Stumpy forms are the only transmissible stage to the tsetse fly and differentiate at high parasite densities apparently through a quorum-sensing-like mechanism (Vassella *et al.*, 1997, Reuner *et al.*, 1997). Although slender forms establish the first BSF population within the mammal, stumpy forms predominate in chronic infections (MacGregor *et al.*, 2011). Since stumpy forms do not divide and do not switch VSG they contribute to prevent population overgrow and limit the frequency of emergence of new VSG variants, optimizing the balance between transmission probability, parasite virulence and chronicity of infection (Matthews *et al.*, 2015, MacGregor *et al.*, 2011).

1.2 CHROMATIN

1.2.1 GENOME ORGANIZATION

Trypanosoma brucei is a diploid organism with a nuclear genome of approximately 35 Mb/haploid DNA content. Despite its small genome, the karyotype in *T. brucei* is remarkable, consisting of more than 120 chromosomes (Berriman *et al.*, 2005) divided into three different classes according to their size: 11 megabase chromosomes (1-6 Mb), 1-5 intermediate chromosomes (200-900 kb) and ~100 minichromosomes (30-150 kb) (Melville *et al.*, 2000, Wickstead *et al.*, 2004, Cross *et al.*, 2014). The core of minichromosomes consists mostly of repetitive palindromes of 177-bp repeats (Wickstead *et al.*, 2004). Intermediate chromosome structure is not entirely known but it shares with minichromosomes a large core of 177-bp repeats (Wickstead *et al.*, 2004). Some strains of *T. brucei* also contain nuclear extrachromosomal circular DNAs, termed NlaIII repeat-elements (Alsford *et al.*, 2003) of unknown function.

About 15% of the total DNA in trypanosomes corresponds to the kDNA which consists on a intertwined network of thousands of double-stranded circular DNA molecules which include large maxicircles (~23 kb) (Sloof *et al.*, 1992) and smaller minicircles (~1 kb) (Chen & Donelson, 1980). Similarly to mitochondrial DNA in other

eukaryotes, maxicircles encode ribosomal RNA (rRNA) genes plus genes mostly coding for subunits of the mitochondrial membrane complexes involved in oxidative phosphorylation. The kinetoplast has a special replication and segregation mechanism, prior to nuclear DNA replication and division, respectively (Jensen & Englund, 2012, Woodward & Gull, 1990). Importantly, many of its maxicircle transcripts require a process of RNA editing in order to mature into functional and translatable messenger RNAs (mRNAs). This process is characterized by, sometimes extensive, insertions and/or deletions of uridine (U) nucleotides into the pre-mRNAs (Goringer, 2012).

Gene density in trypanosomes is extremely high, with the 26-megabase genome encoding ~9,000 putative genes encompassing all known housekeeping genes (Berriman *et al.*, 2005). Trypanosome genome is organized in a very unusual way for eukaryotes. The majority of protein-coding genes are densely packed in large directional clusters in which genes are organized in a head-to-tail orientation. Most of these clusters are transcribed polycistronically by RNA polymerase II (Pol II). Hence, each of them is termed a polycistronic transcription unit (PTU) (Imboden *et al.*, 1987, Muhich & Boothroyd, 1988) (see 1.3.1 Transcriptional and post-transcriptional regulation). Such type of organization resembles that of prokaryotic operons, except for the fact that in trypanosomes genes belonging to the same PTU are generally not functionally related (Berriman *et al.*, 2005). Although polycistrons exist among other eukaryotes such as *Caenorhabditis elegans*, in trypanosomes they are remarkable for encompassing nearly all protein-coding genes in a genome-wide fashion: about 150 PTUs are predicted to exist in the housekeeping regions of *T. brucei* genome where some can harbor $>$ 100 genes (Siegel *et al.*, 2009, Kolev *et al.*, 2010).

1.2.2 THE LARGE VSG REPERTOIRE

Remarkably, the large family of *VSGs* is composed of ~2,000 genes and occupies a vast extension of the genome. Hundreds of silent *VSG* copies are organized in tandem arrays at proximal subtelomeric positions in megabase chromosomes or as single genes in minichromosomes (Williams *et al.*, 1982, Berriman *et al.*, 2005, Cross *et al.*, 2014). Besides, a small part of the *VSG*s resides in BESs, which are located at

megabase or intermediate chromosomes (Becker *et al.*, 2004, Hertz-Fowler *et al.*, 2008) (Fig. 3).

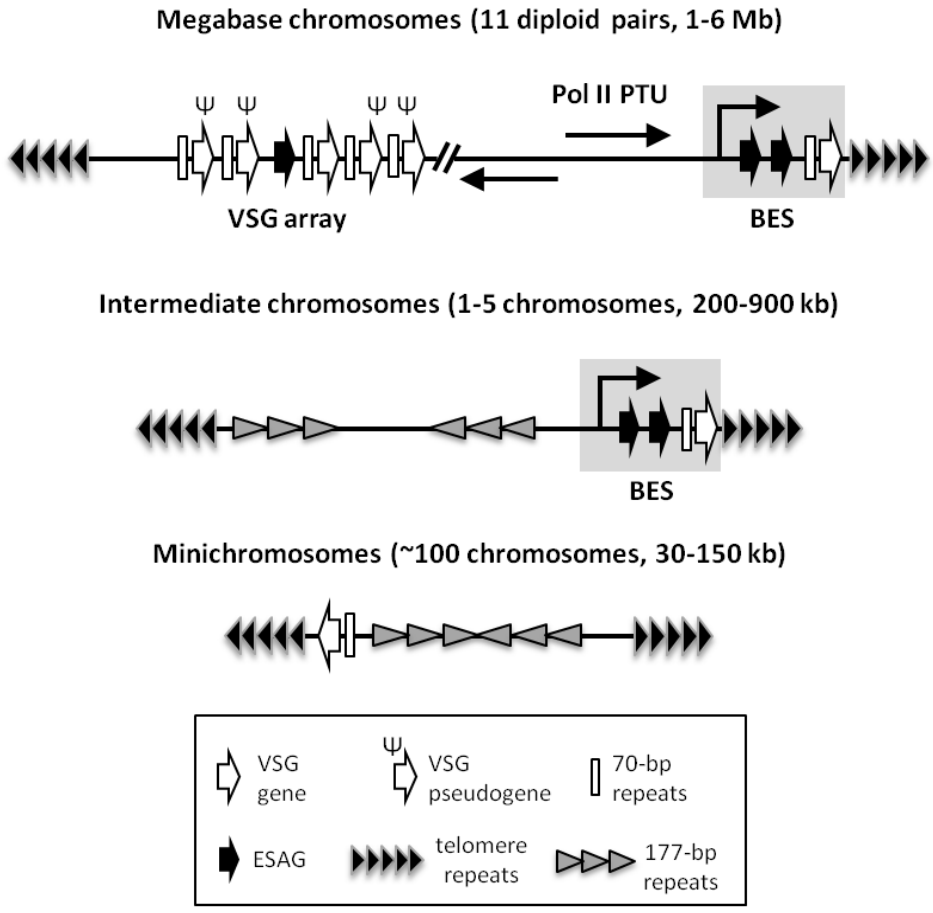


Figure 3. Distribution of *VSG* genes in the genome of *T. brucei*. Megabase chromosomes encode all housekeeping genes, which are expressed in polycistronic transcription units (PTUs) by RNA polymerase II (Pol II) (see 1.3.1 Transcriptional and post-transcriptional regulation). Hundreds of *VSG* copies are located proximal to subtelomeres in *VSG* arrays and include *VSG* genes, pseudogenes and incomplete sequences. *VSGs* are transcribed in bloodstream expression sites (BESs) which reside at subtelomeres in megabase and intermediate chromosomes. A BES contains several expression site-associated genes (*ESAG*s) and a single functional *VSG* at its end. *ESAGs* are interspersed through other genomic locations such as *VSG* arrays. Minichromosomes can also contain *VSGs* at subtelomeres. Intermediate and minichromosomes contain stretches of 177-bp repeats, which in minichromosomes form a palindromic core. Typically, *VSGs* are flanked upstream by 70-bp repeats, important for homologous recombination-dependent *VSG* switching (see 1.3.2 Control of bloodstream expression sites). Arrows indicate direction of transcription. The diagram is not drawn to scale. For simplicity, genes are single-oriented but they display both orientations in the genome and can reside at both subtelomeric ends.

In the genome of *T. brucei VSGs* can be grouped into four categories: 'functional' (encodes all recognizable features of known functional *VSGs*), 'atypical' (complete genes possibly encoding proteins with inconsistent VSG folding or post-translational modification), 'pseudogene' (with frameshifts and/or in-frame stop

codons), and 'incomplete' containing N-terminal or C-terminal fragments (Berriman *et al.*, 2005). Whereas putative functional *VSGs* represent a minority (13-20% in the strain Lister 427), pseudo and incomplete *VSGs* predominate (Berriman *et al.*, 2005, Marcello & Barry, 2007). Among this large family, there appears to exist small *VSG* subfamilies with high sequence identity at the N-terminal or C-terminal which are thought be important to provide interacting partners for mosaic *VSG* formation (Marcello & Barry, 2007). Part of the *VSG* archive is not unique (10% in *T. b. brucei* strain Lister 427), existing as two copies or more (Cross *et al.*, 2014). Besides, *MVSGs* represent a very small part of the repertoire. In Lister 427 there are six *MVSGs* (Cross *et al.*, 2014), with sequence identities consistent with those expressed in metacyclic parasites differentiated *in vitro* (Kolev *et al.*, 2012).

The *VSG* repertoire is significantly divergent between species and strains and even between different life cycle stages or laboratory growth conditions. For examples, most *VSG* gene sequences in strains Lister 427 and TREU 927/4 are very distant. The closest match between both has less than 50% coding sequence (CDS) identity. An important observation was that the *VSG* archive also diverges considerably within the same strain (Lister 427) propagated as different life cycle forms (BSFs *vs* PFs) or cultured in distinct laboratories, exhibiting loss or duplication of *VSG* genes (Cross *et al.*, 2014). Besides providing us with a comprehension of the diversity of the *VSG* family, such data draws attention for the importance of knowing the 'VSGnome' repertoire of the strain/isolate being studied.

Similarly to *T. brucei*, many other pathogens such as *Plasmodium falciparum* (malaria parasite) (Scherf *et al.*, 2008) and *Borrelia* spp. (Lyme disease bacteria) (Barbour *et al.*, 2000, Zhang *et al.*, 1997) express their variant surface antigens at subtelomeres. This highlights the existence of preferential conditions at these genomic locations for the function of such genes. Since subtelomeres are prone to ectopic recombination it seems likely that these are privileged sites for *VSG* duplication and recombination, and therefore for *VSG* repertoire diversification; besides, it has become apparent that the characteristic reversible silencing events associated with the telomere probably sustain monoallelic *VSG* expression and provide transcriptional insulation of silent *VSG* copies, essential to antigenic variation (Barry *et al.*, 2003). The

relationship between telomeric location and *VSG* silencing will be discussed in a subsequent section (see 1.3.2. Control of bloodstream expression sites).

1.2.3 CHROMATIN ORGANIZATION IN EUKARYOTES

Organization of DNA into chromatin is a hallmark of the eukaryotic nucleus. Chromatin is necessary for packaging the DNA molecules inside the physical limits of the nucleus and plays important roles in essential processes such as DNA replication, DNA damage repair, RNA processing and transcription. The basic unit of chromatin is the nucleosome: ~145–147 bp DNA superhelix wrapped around an octamer composed of two copies each of the histone proteins H2A, H2B, H3 and H4 (Luger *et al.*, 1997). Interacting with the nucleosome core, there is usually a distinct histone, histone H1 (H1) or 'linker histone', which binds externally to the nucleosomal DNA and the 'linker DNA' i.e. the short DNA segments that interconnect nucleosomes, and that altogether with the nucleosome core forms the chromatosome (Simpson, 1978, Bharath *et al.*, 2003). Histones are small proteins (~13 kDa) rich in positively charged, basic amino acids (aa) (e.g. arginine and lysine) that facilitate interaction with the negatively charged DNA molecules.

The long-held notion is that chromatin can be organized in a series of increasingly complex and compact conformations which go from 'primary' to higher-order 'secondary' and 'tertiary' structures. Primary structure of chromatin consists in the linear DNA molecules arranged in long nucleosomal arrays often referred as 'beads-on-a-string'. A secondary structure is defined as that arising from the folding and condensation of individual nucleosomal arrays into chromatin fiber, such as the often mentioned '30-nm fiber', which is driven by short-range interactions between neighboring nucleosomes where histone H1 seems to be a key player (Robinson & Rhodes, 2006). Succeeding interactions between fiber leads to large-scale chromatin assemblages, designated as tertiary structure, which ultimately culminates in the fully condensed chromosomes typically observed during metaphase (Luger *et al.*, 2012). Despite the well-established importance of higher-order structures for chromatin function, the conformation of such structures, particularly that of the 30-nm fiber, still

remains unsolved and, more recently, its existence *in vivo* has even been questioned (Maeshima *et al.*, 2010, Nishino *et al.*, 2012).

Chromatin can be organized and regulate short-range DNA domains, for example the local structure of an active promoter, or larger DNA domains (up to Mb of length) generating local structures at the micro-scale. For instance, chromatin can organize into nuclear bodies, chromatin domains or territories, which might involve several chromatin modifications and binding proteins and anchoring to nuclear structures like the nuclear lamina (Pombo & Dillon, 2015). The nucleolus is a hallmark example of a nuclear body of eukaryotes: it is the subnuclear compartment where transcription of ribosomal DNA (rDNA) by RNA polymerase I (Pol I), rRNA processing and ribosome assembly takes place (Lam & Trinkle-Mulcahy, 2015).

Chromatin domains can be subdivided in heterochromatin and euchromatin. In a simplified definition, heterochromatin is more packed (visible as electron-dense domains), gene-poor and generally less transcriptionally active. By contrast, euchromatin typically consists of open chromatin regions that are gene-rich and transcriptionally active. Transposable elements and repetitive DNA regions such as the centromere and the telomere are usually organized in constitutive heterochromatin, which suppresses transcription and/or DNA recombination at these loci, maintaining genome integrity. On the other hand, facultative heterochromatin has the potential to interchange with euchromatin, i.e. to reversible convert to more decondensed chromatin and transcriptionally active state, for instance in the context of monoallelic gene expression (Trojer & Reinberg, 2007). In most cell types repressive heterochromatin is organized around the nucleolus or tethered to the nuclear envelop, whereas euchromatin is found at more central positions in the nucleus (Misteli, 2005, Padeken & Heun, 2014, Pombo & Dillon, 2015).

Different structural and functional states of chromatin can be defined by DNA modifications and variations in nucleosome composition specified by different histone variants or histone covalent post-translational modifications (PTMs) (e.g. methylation, acetylation, phosphorylation), especially present at their N-terminal tails (Luger & Richmond, 1998). Given the ever-increasing number of histone PTMs identified, the

theoretical possible combinations of nucleosome components are astonishing and have the capacity to 'code' for a large number of chromatin states. PTMs are added or removed from histones by several histone-modifying enzymes, which 'write' or "erase" the code. Other chromatin-binding proteins can "read" specific histone PTMs and recruit several other factors such chromatin remodelers and chromatin architectural proteins which trigger a set of downstream events that include changes in nucleosome positioning, DNA accessibility and recruitment of other nuclear machinery e.g. transcription factors (Strahl & Allis, 2000, Lee *et al.*, 2010, Yun *et al.*, 2011).

Together with histone PTMs, histone variants contribute to alter nucleosome stability, chromosome structure and gene expression. Histone variants are non-allelic variants with a distinct amino acid sequence from their canonical counterparts and their deposition onto the DNA is not replication-dependent. Variants might substitute missing histones or be specifically recruited to defined genomic locations to serve specific functions (Luger *et al.*, 2012).

The action of all these players results in a pallet of dynamic chromatin states which can influence for instance, DNA replication, RNA processing and transcription, for instance, working as epigenetic on/off transcriptional switches (Preuss & Pikaard, 2007, Venkatesh & Workman, 2015).

1.2.4 CHROMATIN ORGANIZATION IN *T. BRUCEI*

Alike other lower eukaryotes, *T. brucei* mitosis is 'closed', which means that nuclear envelope and nucleolus do not disassemble (Vickerman & Preston, 1970, Ogbadoyi *et al.*, 2000). Interestingly, even though all chromosome classes segregate via a mitotic spindle (Ersfeld *et al.*, 1998), it was recently found that *T. brucei* uses unconventional kinetochores which appear to have emerged exclusively in kinetoplastids early in evolution (Akiyoshi & Gull, 2014).

The main nuclear architectural features typical of eukaryotes are also observed in *T. brucei* such as a double bilayered nuclear envelope, nuclear pore complexes with a conserved composition (DeGrasse *et al.*, 2009), a nucleolus and domains of

euchromatin and heterochromatin, the latter predominant at the nuclear periphery (Ersfeld, 2011). While the typical nuclear lamina found in metazoans seems to be absent from the parasite (Field *et al.*, 2012), one protein with lamin-like functions, NUP-1, has been identified in *T. brucei* (Rout & Field, 2001, DuBois *et al.*, 2012).and shown to be important for nucleus structural integrity and telomere positioning at the nuclear periphery (DuBois *et al.*, 2012).

Contrary to what generally occurs, in trypanosomatids chromosomes do not visibly condense during metaphase and a 30-nm fiber has never been observed (Hecker & Gander, 1985, Burri *et al.*, 1995). Overall trypanosome DNA is less compacted within the nucleus when compared with higher eukaryotes (Hecker & Gander, 1985) and even between two stages of the parasite life cycle, with chromatin being more compact in BSFs than in PFs (Schlimme *et al.*, 1993). This has been suggested as an important adaptation to accommodate the rapid changes in gene expression required throughout life cycle development of the parasite (Hecker *et al.*, 1994, Belli, 2000). Curiously, the less compacted state of trypanosome chromatin has been attributed the presence of divergent histones and differential expression of histone variants between BSFs and PFs, namely that of H1 proteins (Schlimme *et al.*, 1993, Burri *et al.*, 1994).

A paradigmatic example of nuclear subcompartmentalization in *T. brucei* is the so-called expression site body (ESB). This is an extranucleolar Pol I body, functionally distinct from the nucleolus and that appears to be solely dedicated to the transcription of the active BES in BSFs (Navarro & Gull, 2001, Chaves *et al.*, 1998). Telomeres also occupy different nuclear positions throughout development of *T. brucei*. While in the bloodstream slenders, at interphase, telomeres preferentially localize at inner regions of the nucleus in the transmissible, non-dividing stumpy and in the insect PF stage, telomeres cluster close to the nuclear periphery and appear to be associated with heterochromatin (Ogbadoyi *et al.*, 2000, Perez-Morga *et al.*, 2001). Thus, telomeres are closer to the peripheral heterochromatic regions in those life cycle stages in which active BES is almost (Amiguet-Vercher *et al.*, 2004) or completely inactive (stumpy and PF, respectively). While telomeric repeats are probably packed into constitutive heterochromatin, the BESs at the subtelomeres probably correspond to facultative

heterochromatin whose transcriptional status is exchangeable during an *in situ* switch and or throughout *T. brucei* life cycle progression.

1.2.5 EPIGENETIC REGULATORS IN *T. BRUCEI*

There is compelling evidence that epigenetic mechanisms underlie essential processes in *T. brucei* such as the cell cycle, cell differentiation and antigenic variation (Figueiredo *et al.*, 2009). An 'epigenetic' phenomenon can be defined as a heritable change in genome function that does not depend on changes in DNA sequence. Epigenetic changes are linked with modifications on chromatin composition and structure that that generally derive from different histone PTMs, histone variants or covalent DNA modifications (Deans & Maggert, 2015). Subsequently, many of the epigenetic regulators that have been implicated in *T. brucei* biology are briefly described.

Histones. *T. brucei* contains the four canonical core histones (H2A, H2B, H3 and H4) and histone H1 (Alsford & Horn, 2004). According with its classification as canonical histones, mRNAs of the major histones of *T. brucei* oscillate throughout the cell cycle, peaking at the S-phase (Ersfeld *et al.*, 1996, Archer *et al.*, 2011). Though histones are some of the most conserved proteins throughout evolution, trypanosomatid histones are among the most divergent histones known, particularly within the unstructured N-terminal tails (Thatcher & Gorovsky, 1994).

Histone PTMs described in *T. brucei* are rather unusual. Core histones lack many PTMs that are well-conserved across eukaryotes and on the other hand, some histone PTMs appear exclusive of the parasite. Examples of the latter are a complex pattern of acetylations on multiple lysines of H2A C-terminus and abundant levels of methylated alanines at H2A, H2B and H4 N-termini (Janzen *et al.*, 2006a, Mandava *et al.*, 2007). Nevertheless, some PTMs conserved in other eukaryotes are also found in *T. brucei* such as the tri-methylation of H3 lysine 4 (H3K4) and mono-, di- or tri-methylation of H3K76 (K79 in other species) (Janzen *et al.*, 2006b); in addition, acetylation of H4 in the lysine residues K4, K10 and K14, might be the corresponding

homologues to acetylations conserved in other eukaryotes at lysine residues in nearby positions (K5, K12 and K16) (Mandava *et al.*, 2007).

Histone variants in *T brucei* encompass four variants of core histones: a homologue of H2A.Z and the three parasite-specific histone variants H3.V, H4.V and H2B.V (Lowell & Cross, 2004, Lowell *et al.*, 2005). While H2A.Z and H2B.V are essential for parasite viability, H3.V and H4.V are not (Lowell & Cross, 2004, Lowell *et al.*, 2005, Siegel *et al.*, 2009). *T. brucei* also contains sequence variants of H1 which display a few aa differences (Gruter & Betschart, 2001), but it is unknown whether these are functional distinct.

Histone-modifying enzymes have also been characterized in *T. brucei.* Histone deacetylases include four homologues of class I and class II deacetylases TbDAC1-2 and TbDAC3-4, respectively (Ingram & Horn, 2002, Wang *et al.*, 2010) and class III silent information regulator 2 (SIR2)-related proteins (TbSIR2rp1-3) (Alsford *et al.*, 2007, Garcia-Salcedo *et al.*, 2003). Histone deacetylase activity has been confirmed in TbDAC1, TbDAC3 and TbSIR2rp1 (Garcia-Salcedo *et al.*, 2003, Wang *et al.*, 2010). *T. brucei* also possesses histone acetyltransferases from the MYST-family (TbHAT1-3) and putative histone acetyltransferases related with Elongator protein 3 (ELP3), TbELP3a and TbELP3b (Alsford & Horn, 2011). TbHAT2 and TbHAT3 are known to catalyze H4 lysine acetylation (Siegel *et al.*, 2008, Kawahara *et al.*, 2008). Two lysine methyltransferase orthologs of disruptor of telomeric silencing 1 (DOT1), TbDOT1A and TbDOT1B, were characterized in the parasite: TbDOT1A is responsible for the mono- and di-methylation of H3K76 while TbDOT1B catalyzes its tri-methylation (Janzen *et al.*, 2006b).

Histone chaperones regulate the supply, deposition or displacement of histones from nucleosomes, being key players in the multiple processes where nucleosome reorganization is required. Some histone chaperones also may couple nucleosome assembly with crucial DNA metabolic processes such as DNA replication (Gurard-Levin *et al.*, 2014). *T. brucei* possesses four types of histone chaperones: anti-silencing factor 1A (TbASF1A), chromatin assembly factor 1 subunit b (TbCAF-1b), facilitates chromatin transcription complex (TbFACT) and nucleoplasmin-like protein

(TbNLP) (Alsford & Horn, 2012, Denninger *et al.*, 2010, Denninger & Rudenko, 2014, Narayanan *et al.*, 2011).

Chromatin-remodeling enzymes comprise different subfamilies of multiprotein complexes with a core ATPase which catalyze a wide range of changes in the conformation, composition and position of nucleosomes, providing a means of producing rapid chromatin rearrangements (Narlikar *et al.*, 2013). *T. brucei* possesses a novel member of the imitation switch (ISWI) family of SWI2/SNF2-related chromatin-remodeling enzymes (TbISWI). TbISWI is a nuclear protein and associates with chromatin at several loci across the genome (Hughes *et al.*, 2007, Stanne *et al.*, 2011, Stanne *et al.*, 2015).

Chromatin architectural proteins can be histones or non-histone proteins that interact with and modulate chromatin structure acting as DNA benders, bridgers or wrappers (Luijsterburg *et al.*, 2008). High-mobility group box (HMGB) proteins are ubiquitous chromatin architectural proteins that usually decompact chromatin and can facilitate transcription, DNA replication, DNA recombination or repair (Stros, 2010). HMGBs can interact with nucleosomes, displace histone H1 and facilitate the recruitment of other nuclear regulatory factors (Stros, 2010). A HMGB protein has been characterized in *T. brucei*: the trypanosome DNA binding protein (TDP-1), which predominantly associates with open chromatin at Pol I loci (Narayanan & Rudenko, 2013).

Telomeric proteins form a complex that protects telomeres and mediates the assembly of a specialized heterochromatin at telomeres, which typically spreads onto subtelomeric regions (Blasco, 2007). A number of telomere-associated proteins have been characterized in *T. brucei*: the Ku DNA-end binding heterodimer (TbKu70/Ku80), (Conway *et al.*, 2002, Janzen *et al.*, 2004), the above-mentioned histone deacetylase SIR2rp1 (Alsford *et al.*, 2007, Garcia-Salcedo *et al.*, 2003), an homologue of repressor/activator protein 1 (TbRAP1) (Yang *et al.*, 2009), a telomeric repeat-binding factor (TbTRF) and TRF-interacting factor 2 (TbTIF2) (Yang *et al.*, 2009, Jehi *et al.*, 2014b, Jehi *et al.*, 2014a, Li *et al.*, 2005). In addition, similarly to other eukaryotes, the origin replication complex of *T. brucei,* which consists on the origin replication complex

1/cell division cycle 6-like protein (TbORC1/CDC6), also associates with telomeric repeats (Benmerzouga *et al.*, 2013).

DNA covalent modifications so far known in *T. brucei* are rather atypical. Contrarily to most eukaryotes, DNA methylation (5−methylcytosine) levels are very low in *T. brucei* (Militello *et al.*, 2008). Besides, about 1% of thymidine residues in the nuclear genome are replaced by a bizarre DNA modification unique to trypanosomatids and the related *Euglena* algae: β-D-glucosyl-hydroxymethyluracil, also termed base J (Gommers-Ampt *et al.*, 1993, van Leeuwen *et al.*, 1998, Dooijes *et al.*, 2000). Curiously, base J appears to be essential for *T. cruzi* or *Leishmania* spp. (Genest *et al.*, 2005, Ekanayake *et al.*, 2011) but it is dispensable for *T. brucei* (Cliffe *et al.*, 2009). Regardless of recent advances about the role of base J in transcription initiation and termination in *T. cruzi* and *Leishmania* spp., respectively (Ekanayake & Sabatini, 2011, van Luenen *et al.*, 2012, Reynolds *et al.*, 2014), the functions of this DNA modification in *T. brucei* remain largely a mystery.

Reverse genetics has revealed that many of the referred epigenetic factors are important regulators of gene expression in trypanosomatids (Figueiredo *et al.*, 2009). Not surprisingly, some of these factors have been implicated in several functions beyond transcriptional control, such as the cell cycle (e.g. TbDOT1A, TbCAF-1b), DNA repair and homologous recombination (e.g. TbSIR2rp1, TbHAT1), subtelomere/telomere integrity (e.g. TbKu70/Ku80, TbTIF2, TbTRF) and parasite differentiation (e.g. TbDOT1B). These topics are out of the scope of this thesis and below I will only focus on the functions of these factors for gene expression in *T. brucei.*

Most of what is known so far about chromatin organization and epigenetics in *T. brucei* is related to *VSG* genes (Figueiredo *et al.*, 2009, Navarro *et al.*, 2007). The focus on these genes is not only due to the importance of antigenic variation for *T. brucei* but also because BESs are controlled at the level of transcription, contrasting with gene expression for most of *T. brucei* genome, which is probably regulated mostly at the post-transcriptional level. The relevance of epigenetics for the control of *VSG* and

procyclin gene expression will be highlighted in following sections (see 1.3.2 Control of bloodstream expression sites and 1.3.3 Control of procyclin loci).

1.3 GENE EXPRESSION IN *T. BRUCEI*

1.3.1 TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL REGULATION

Organization of transcription units

As mentioned earlier (see 1.2.1 Genome organization), in trypanosomatids large clusters of unrelated genes are co-transcribed by Pol II from the same coding strand as a single PTU (Berriman *et al.*, 2005, Imboden *et al.*, 1987, Muhich & Boothroyd, 1988). Typically, the regions between PTUs are strand-switch regions (SSRs) because they separate PTUs with opposite coding strands. Experimental evidence supports that Pol II transcription initiates bidirectionally at divergent SSRs and terminates at convergent SSRs (Martinez-Calvillo *et al.*, 2003, El-Sayed *et al.*, 2003, Kolev *et al.*, 2010). Pol II transcription can also initiate internally within transcripton units and at other non-SSRs, for instance where a change of RNA polymerase type likely occurs (Kolev *et al.*, 2010, Siegel *et al.*, 2009). Regions between Pol II transcription units often contain transfer RNAs (tRNAs) and other RNA polymerase III (Pol III)-transcribed small RNAs. Besides, in *T. brucei* the Pol I-transcribed rDNA units and procyclin loci (see below, RNA polymerases) also lie between Pol II transcription units ((TriTrypDB 9.0) (Fig. 4).

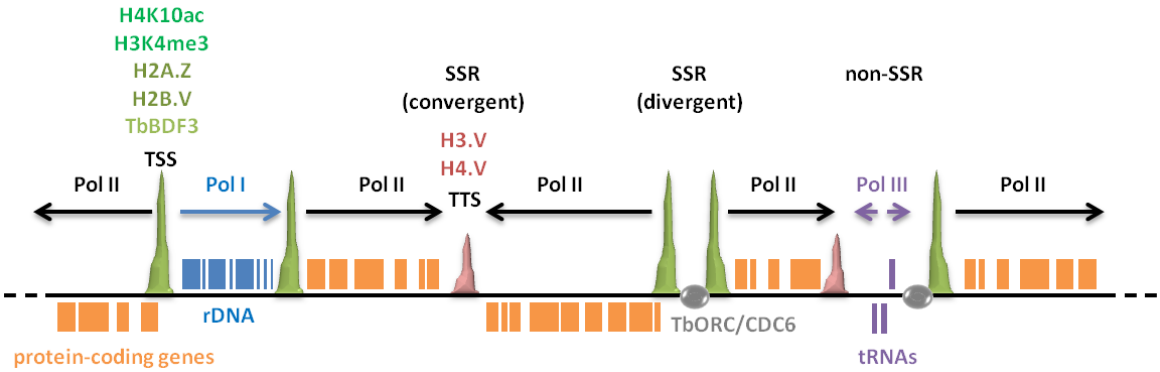


Figure 4. (previous page) Transcription units and flanking epigenetic signals in *T. brucei.* Most of *T. brucei* genome is organized in polycistronic transcription units (PTUs) transcribed by Pol II. Neighboring PTUs are usually separated by strand-switch regions (SSR). In SSRs, Pol II transcription initiates at divergent SSRs and terminates at convergent SSRs. Transcription might also initiate at non-SSRs. Arrows indicate direction of transcription. Regions in-between Pol II units often contain Pol III-transcribed genes such as tRNAs and the Pol I-transcribed rDNA units and procyclin loci. Predicted transcription start sites (TSS) are enriched for acetylated H4K10 (H4K10ac), tri-methylated H3K4 (H3K4me3), H2A.Z, H2B.V and TbBDF3. Putative Pol II transcription termination sites (TTSs) are enriched for H3.V and H4.V. In addition, the origin replication complex protein TbORC1/CDC6 binds frequently at the boundaries of PTUs, mostly at predicted TSSs. This a representative diagram, not drawn to scale.

RNA polymerases

The three eukaryotic RNA polymerases have been biochemically identified in trypanosomatids (Martinez-Calvillo *et al.*, 2010, Bindereif, 2012). Remarkably, *T. brucei* Pol I is the only known multifunctional Pol I among eukaryotes that, besides transcribing the rDNA unit (18S, 5.8S and 28S-like genes), also transcribes three other protein-coding loci: the *VSG* bloodstream expression sites (BESs), the procyclin loci and the *MVSG* expression sites (MESs) (Kooter & Borst, 1984, Ginger *et al.*, 2002, Gunzl *et al.*, 2003). Besides, ~10 rDNA transcription units are predicted in *T. brucei* (Alsford *et al.*, 2005b), a small number compared with other eukaryotes (e.g. more than a hundred in yeast and mammals). It is also unknown whether only half the rDNA genes are transcriptionally active as was shown in other eukaryotes (Dammann *et al.*, 1993, Grummt & Langst, 2013). *T. brucei* Pol II transcribes most mRNAs, the spliced leader (SL) RNA gene required for mRNA processing by *trans*-splicing (see below, mRNA splicing and polyadenylation) and small nucleolar RNAs (snoRNAs) (Gilinger & Bellofatto, 2001, Dunbar *et al.*, 2000), whereas Pol III transcribes tRNAs, 5S rRNA, small nuclear RNAs (snRNAs) and the 7SL RNA, similarly to the other eukaryotes (Nakaar *et al.*, 1994, Michaeli *et al.*, 1992, Fantoni *et al.*, 1994).

RNA polymerase I transcription of rDNA genes is the most active transcription in a eukaryotic cell, generally accounting for over 50% of the total transcription activity. The high rates of rDNA transcription seem to arise from the high transcription initiation rates of Pol I which can be achieved by ~150 Pol I enzymes simultaneously transcribing an rDNA gene (Albert *et al.*, 2012). Therefore it seems reasonable that *T.*

brucei had evolved the unique capacity to highjack the highly efficient Pol I enzyme to ensure the production of the high number of VSG molecules it requires (~10^7 VSG surface copies per bloodstream parasite (Jackson *et al.*, 1985)) from only a single active gene.

The *T. brucei* multifunctional Pol I might result from specific differences in its structure and molecular partners (e.g. transcription factors) even though all Pol I machinery deviations identified so far, such as the trypanosome-specific Pol I subunit variants TbRPB5z, TbRPB6z and TbRPB10z (Nguyen *et al.*, 2006, Devaux *et al.*, 2007), are shared by other trypanosomatids (e.g. *T. cruzi*) that lack *VSG* or procyclin genes. In addition, the promoters of rDNA, BES, procyclin loci and MES are structurally different (Vanhamme *et al.*, 1995b, Ginger *et al.*, 2002, Pham *et al.*, 1996, Laufer & Gunzl, 2001, Brown *et al.*, 1992), suggesting that they might recruit different regulatory factors. The compartmentalization of the transcriptionally active BES in the ESB also supports this hypothesis. The putative histone acetyltransferase TbELP3b specifically localizes at the nucleolus and is excluded from the ESB. Accordingly, it suppresses rDNA transcription but does not affect transcription in BES (Alsford & Horn, 2011). In recent years, a multisubunit complex was found to act as a general transcription factor of Pol I, conserved across trypanosomatids. The multi-subunit complex, termed class I transcription factor A (CITFA), binds to *VSG*, procyclin and rDNA promoters and it is necessary for transcription initiation in all of them (Brandenburg *et al.*, 2007, Nguyen *et al.*, 2012, Nguyen *et al.*, 2014).

Epigenetic regulation of transcription

Unlike many eukaryotes, in which Pol II regulation is considerably well understood, in trypanosomes it remains essentially a mystery. No Pol II promoters or associated transcription factors have been found except for the SL RNA gene promoter (Gunzl *et al.*, 1997, Das *et al.*, 2005, Schimanski *et al.*, 2005). A milestone in trypanosome research was the finding that several epigenetic marks associate with putative Pol II transcription start sites (TSSs) and transcription termination sites (TTSs) (Respuela *et al.*, 2008, Thomas *et al.*, 2009, Siegel *et al.*, 2009, Wright *et al.*, 2010). In *T. brucei*, chromatin immunoprecipitation (ChIP) sequencing revealed that H2A.Z, H2B.V

acetylated H4K10, tri-methylated H3K4 and a putative acetyl-binding protein, the bromodomain factor 3 (TbBDF3) are enriched at predicted transcription start sites (TSSs) whereas H3.V and H4.V preferentially locate at potential TTSs (Siegel *et al.*, 2009, Wright *et al.*, 2010). (Fig. 4). High-throughput RNA sequencing (RNA-Seq) of a library enriched in precursor transcripts was used to map the TSSs in *T. brucei* (Kolev *et al.*, 2010), demonstrating that, indeed, transcription initiates at those regions (SSRs and non-SSRs) where an enrichment for H2A.Z, H2B.V, acetylated H4K10, tri-methylated H3K4 and TbBDF3 was previously detected (Siegel *et al.*, 2009). H2A.Z and H2B.V specifically associate with each other (Lowell *et al.*, 2005) and nucleosomes containing H2A.Z-H2B.V dimers associate less with H3 and H4, which is indicative of destabilized nucleosomes. This suggests that H2A.Z/H2B.V-enriched chromatin is less compact and therefore more accessible to the Pol II transcription machinery (Siegel *et al.*, 2009). These data suggests that epigenetic marking has been adopted as an evolutionarily ancient means of controlling Pol II transcription. Nevertheless, we still ignore how these marks functionally interact and regulate Pol II activity.

Others have suggested that Pol II transcription initiation might be unusually unspecific and that the presence of a more open chromatin is sufficient for its engagement and transcription initiation (McAndrew *et al.*, 1998). Chromatin is devoid of nucleosomes at the highly expressed SL RNA gene arrays transcribed by Pol II (see below, mRNA splicing and polyadenylation) when compared with its associated retrotransposon SLACS and with other Pol II protein-coding genes; nucleosomes are also highly depleted at the promoter regions of Pol I transcribed rDNA units and BESs, when compared with its respective protein-coding genes (Hitchcock *et al.*, 2007, Stanne & Rudenko, 2010, Figueiredo & Cross, 2010). Thus, nucleosome occupancy is likely playing a role in the control of transcription initiation/elongation by the two types of RNA polymerases in trypanosomatids, as has been reported in other eukaryotes.

In several eukaryotes an association between promoter regions and/or the transcriptional activity and replication origins has been observed (Mechali *et al.*, 2013). Interestingly, replication is also functionally coupled to transcription in *T. brucei* (Tiengwe *et al.*, 2012, Calderano *et al.*, 2015). The origin replication complex protein

TbORC1/CDC6 of *T. brucei* binds at the boundaries of PTUs and upon its depletion mRNA levels increase globally throughout the genome (Tiengwe *et al.*, 2012). In addition, H3K76 mono- and di-methylated is relatively enriched at some PTU boundaries, possibly marking replication origins (Gassen *et al.*, 2012).

mRNA splicing and polyadenylation

Besides its genome-wide polycistronic transcription, trypanosomes also differ remarkably from other eukaryotes regarding mRNA maturation. The polycistronic precursor transcripts are processed into mature monocistronic mRNAs via *trans*-splicing, which adds a m^7G-capped 39-nucleotide Spliced Leader RNA to the 5' untranslated region (5'UTR), and by polyadenylation of the 3' untranslated region (3'UTR) (Boothroyd & Cross, 1982, Ullu & Tschudi, 1991). *Trans*-splicing and polyadenylation occur cotranscriptionally and the *trans*-splicing of the downstream mRNA is coupled with polyadenylation of the upstream mRNA (Huang & van der Ploeg, 1991, LeBowitz *et al.*, 1993, Ullu *et al.*, 1993). Contrary to other eukaryotes where mRNA capping machinery directly interacts with Pol II, *trans*-splicing provides a means of uncoupling Pol II from the 5' m^7G capping of mRNAs, which is essential to allow *T. brucei* Pol I to produce functional mRNAs (Rudenko *et al.*, 1991, Zomerdijk *et al.*, 1991). *Trans*-splicing of trypanosome mRNAs contrasts with mRNA maturation via *cis*-splicing, which typically occurs in eukaryotes and excises non-coding introns from the pre-mRNAs. In *T.brucei*, however, *cis*-spliced introns seem to be restricted to only two genes, the poly(A) polymerase (Mair *et al.*, 2000) and a putative RNA helicase gene (Berriman *et al.*, 2005, Jae *et al.*, 2010, Siegel *et al.*, 2010).

Post-transcriptional regulation

Within a PTU, the steady-state mRNA abundances between genes can vary at least 50-fold (Siegel *et al.*, 2010) and also differ considerably throughout the life cycle of the parasite. Recent genome-wide studies have compared the transcript levels between BSF and PF stages using different technical approaches (microarrays, digital

gene expression and RNA-Seq) and *T. brucei* cell-lines (monomorphic and pleomorphic). Collectively, these studies have exposed an extensive number of genes (6-40% of genes) whose mRNA levels significantly change between the two parasite stages (Jensen *et al.*, 2009, Kabani *et al.*, 2009, Nilsson *et al.*, 2010, Siegel *et al.*, 2010, Veitch *et al.*, 2010, Queiroz *et al.*, 2009). In the absence of transcriptional control in, post-transcriptional control of mRNA levels is expected to be the major source of gene expression regulation in trypanosomes, and has been interrogated for a long time (Kramer, 2012). However, so far, this veil has been lifted considerably for only a few genes. Paradigmatic examples are the switch from low to high mRNA levels from procyclin and phosphoglycerate kinase B genes upon differentiation of BSFs to PFs. The developmental regulation of these transcripts results from the control of mRNA half-live mediated by U-rich regulatory elements at the 3'UTR, which destabilize both mRNAs in BSFs (Hotz *et al.*, 1997, Blattner & Clayton, 1995).

RNA-Seq based genome-wide approaches have been valuable tools in gaining a comprehensive view on the dominant post-transcriptional mechanisms shaping *T. brucei* transcriptome. Several studies have gathered evidence for the importance of alternative splicing and polyadenylation (Kolev *et al.*, 2010, Nilsson *et al.*, 2010, Siegel *et al.*, 2010) and control of mRNA half-live and decay rates (Fadda *et al.*, 2014, Manful *et al.*, 2011) in regulation of mRNA levels. More recently, widespread control at the level of translation is also supported by two studies where RNA translation was examined by ribosome profiling in BSFs and PFs (Jensen *et al.*, 2014, Vasquez *et al.*, 2014). Finally, *trans*-acting RNA-binding proteins have been emerging as crucial players in remodeling gene expression patterns during *T. brucei* life cycle (Kolev *et al.*, 2014). A major breakthrough in trypanosome research came from the observation that the overexpression of a single *T. brucei* RNA-binding protein (RBP), TbRBP6, is sufficient to trigger the entire differentiation pathway of BSFs to metacyclic parasites in axenic cultures (Kolev *et al.*, 2012). In addition, several other studies have identified subsets of target mRNAs potentially regulated by RPBs throughout the life cycle development in *T. brucei* (Najafabadi *et al.*, 2013, Walrad *et al.*, 2012, Wurst *et al.*, 2012).

Even though most of the genes are expected to be controlled mainly post-transcriptionally, remarkable exceptions to this are the Pol I specialized loci from

where the genes coding for *VSGs* (the BESs), procyclin and *MVSGs* (the MESs) are expressed, which are controlled are at the level of transcription (Rudenko, 2010, Park *et al.*, 2012, Barry & McCulloch, 2001). How gene expression is regulated at these loci will be explored below.

1.3.2 CONTROL OF BLOODSTREAM EXPRESSION SITES

Bloodstream expression sites

Bloodstream expression sites are ~45-60 kb polycistronic units that contain several expression site-associated genes (*ESAGs*) and a single functional *VSG* as the last gene located ~0.2-1.5 kb away from the telomere (Hertz-Fowler *et al.*, 2008). The *VSG* unit is typically a 3-4 kb cassette flanked upstream by tracts of 70-bp repeats which vary in length and can go up to several kbs (Hertz-Fowler *et al.*, 2008). The *VSG* also contains a conserved region from the end of the C-terminal region to elements in the 3'UTR. Both 70-bp repeats together and the conserved *VSG* 3' regions provide boundaries of homology for switching via gene conversion with an inactive *VSG* (Michels et al., 1983), even though recombination can also occur further upstream within an ESAG (McCulloch *et al.*, 1997) or other BES sequences (Hertz-Fowler *et al.*, 2008, Boothroyd *et al.*, 2009).

BESs promoters are very short and composed of two elements residing within 67 bp upstream of the TSS (Vanhamme *et al.*, 1995b, Pham *et al.*, 1996). Half of the BESs contain an additional promoter ~13 kb upstream, which drives the expression of ESAG10 (Hertz-Fowler *et al.*, 2008). The upstream region of BESs invariably contains long arrays of 50-bp repeats (Zomerdijk *et al.*, 1990) (Fig. 5).

Each BES encodes a distinct combination of *ESAGs*, which comprise polymorphic variants of *ESAG1-12* (Hertz-Fowler *et al.*, 2008). Some *ESAGs* have been characterized and encode proteins important for a successful infection inside the mammal. For instance, *ESAG6* and *ESAG7* encode the subunits of the transferrin receptor involved in the uptake of mammalian transferrin, the only iron source for BSF

parasites (Bitter *et al.*, 1998). Studying the *ESAGs* expressed from the BES is challenged by the fact that many *ESAGs* are members of large families of similar but not identical genes dispersed throughout the genome (Berriman *et al.*, 2005). Since different BESs contain different polymorphic versions of the same *ESAGs* it has been debated whether different ESAG variants might favor *T. brucei* infection in specific mammalian hosts (Gerrits *et al.*, 2002, Salmon *et al.*, 2005).

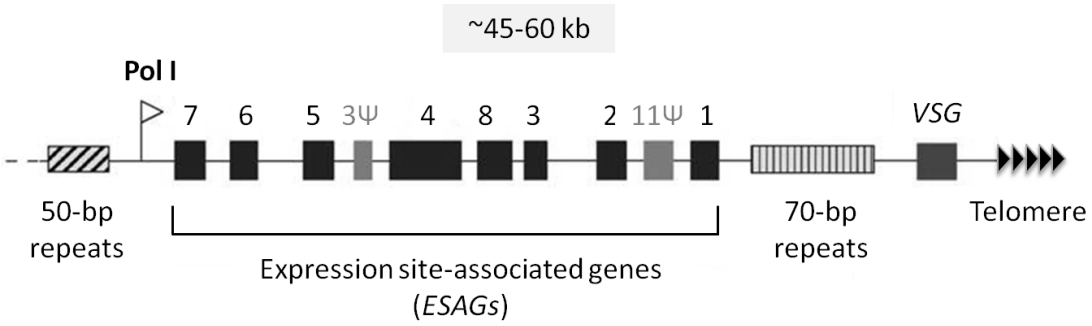


Figure 5. A generic bloodstream expression site (BES). BESs have a typical size of ~45-60 kb, localize at the subtelomeres and are transcribed polycistronically by Pol I (Pol I promoter, flag). Several expression site-associated genes (*ESAGs*) or pseudogenes (Ψ) are encoded within a BES. The diagram depicts the most common set of *ESAGs*, but the number and order of *ESAGs* varies between BESs. The last gene in a BES is always a 'functional' *VSG* that resides ~0.2-1.5 kb upstream of the telomere. Different DNA repeats are associated with the BES (hatched boxes): most *VSGs* are flanked upstream by a stretch of 70-bp; upstream of the BES promoter there are long tracts of 50-bp repeats; telomere repeats locate downstream of the BES and mark the chromosome end (adapted from McCulloch & Horn, 2009).

Similarly to *VSGs*, *MVSGs* activated in the metacyclic parasites are transcribed by Pol I from a subtelomeric expression site. The metacyclic expression site (MES) probably evolved from BESs even though important differences distinguish both. MES are short, ~1-5 kb monocistronic units which lack *ESAGs* and encode only a *MSVSG* gene (Graham & Barry, 1995, Alarcon *et al.*, 1994, Ginger *et al.*, 2002). In fact, they represent the only Pol I monocistronic units with protein-coding genes known in *T. brucei.* Also, BES and MES promoters are structurally similar but functionally different (Ginger *et al.*, 2002). A recent study confirmed former evidence that even though several *MVSGs* are expressed in a metacyclic population, only one *MVSG* is activated in each individual metacyclic parasite, similarly to *VSG* activation in BSFs (Ramey-Butler *et al.*, 2015). However, in contrast with *VSGs*, *MVSGs* are switched on only by *in situ*

transcriptional activation, which appears to be controlled at the level of transcription initiation (Graham & Barry, 1995, Pedram & Donelson, 1999, Ginger *et al.*, 2002).

Transcriptional regulation of bloodstream expression sites

How monoallelic expression of the BES is achieved has been a subject of thorough investigation. The parasite needs to maintain transcription active at one BES while silencing it in the others and the active/silent transcriptional states need to be rapidly reversible to ensure successful *in situ* switching. Besides, upon differentiation of BSFs to PFs, all BESs are silenced and procyclin loci need to be fully activated.

It has become clear that trypanosomes and other protozoan parasites such as *P. falciparum* and *Toxoplasma gondii* (toxoplasmosis parasite) use epigenetic regulation as a common means for controlling gene expression (Croken *et al.*, 2012). *P. falciparum*, for instance, employs epigenetic regulation to control antigenic variation of *var* genes that are essential for its virulence (Scherf *et al.*, 2008). Similarly, in trypanosomes, there is compelling evidence that control of BES expression and VSG switching is a complex interplay of chromatin-associated factors (Figueiredo 2009, Glover 2013): the chromosomal location proximal to telomeres, the nuclear positioning of BESs and its epigenetic environment.

Telomere-proximal location is thought to contribute to BES regulation. Such is supported by several findings. Telomeric heterochromatin has the ability to transcriptionally silence genes positioned nearby, an effect termed telomere position effect (TPE) which extends from yeast to higher eukaryotes (Gottschling *et al.*, 1990, Tham & Zakian, 2002, Misteli, 2014). TPE silencing occurs in *T. brucei* and is capable of silencing transcription from Pol I promoters inserted proximal to telomeres (Horn & Cross, 1995, Horn & Cross, 1997). Also, the telomere-associated proteins TbRAP1 and TbORC1/CDC6 are necessary for BES silencing and expression of only one VSG at the surface (Yang *et al.*, 2009, Benmerzouga *et al.*, 2013). However, telomere-independent repression must be operating in BESs as well. For instance, telomeric silencing extends only to 2 kb upstream of *de novo* telomeres in PFs (Glover & Horn, 2006) and BES-

associated silencing is preserved even when the telomere is lost (Glover *et al.*, 2007). Besides, other proteins which are important for telomere maintenance such as the complex Ku70/Ku80 (Conway *et al.*, 2002, Janzen *et al.*, 2004), and telomeric silencing, namely the histone acetyltransferase TbHAT1 (Kawahara *et al.*, 2008) and TbSIR2rp1 (Alsford *et al.*, 2007) and the histone deacetylase TbDAC1 (Wang *et al.*, 2010), are not required for BES silencing.

Nuclear positioning importance for BES monoallelic expression is underscored by the finding of the ESB. In BSFs the active BES specifically locates and is transcribed at the ESB while silent BESs are excluded from it (Navarro & Gull, 2001). Differentiation of BSFs to PFs is accompanied by ESB loss and rapid repositioning of the active BES promoter close to the nuclear envelope (Landeira & Navarro, 2007, Navarro & Gull, 2001). It is unknown what functionally distinguishes Pol I transcription at the ESB from the nucleolus. The ESB might provide specific factors required for BES Pol I transcription and RNA processing. There is evidence that the active BES is marked by a specific epigenetic environment that distinguishes it from the equally Pol I-transcribed rDNA loci, for instance: high levels of the SUMOylation PTM (Lopez-Farfan *et al.*, 2014) and low levels/absence of the putative histone acetyltransferase TbELP3b (Alsford & Horn, 2011).

Epigenetic regulation is certainly controlling expression at BESs. Consistent with the different nuclear territories occupied by active and silent BESs, chromatin at active BES is nucleosome-depleted and displays a more accessible when compared with silent BESs (Figueiredo & Cross, 2010, Stanne & Rudenko, 2010). Derepression of BES promoter-proximal regions occurs upon depletion of histone H3 (Alsford & Horn, 2012), histone H1 (Povelones *et al.*, 2012), the histone chaperones TbFACT (Denninger *et al.*, 2010, Denninger & Rudenko, 2014), TbNLP (Narayanan *et al.*, 2011), TbASF1A and TbCAF-1b (Alsford & Horn, 2012) and histone deacetylase TbDAC3 (Wang *et al.*, 2010). The histone methyltrasferase TbDOT1B (Figueiredo *et al.*, 2008), the chromatin remodeler TbISWI (Stanne *et al.*, 2011, Hughes *et al.*, 2007), the telomeric protein TbRAP1 (Yang *et al.*, 2009) and the lamin-like TbNUP-1 (DuBois *et al.*, 2012) were found to repress the entire BES locus including its terminal *VSGs* (Table 1). Depletion or deletion of each of these factors, though, leads to a significant but weak *VSG*

derepression, still several orders of magnitude (~10^2-10^3) lower than the expression of the active *VSG.* This indicates that the repressive epigenetic environment at silent BESs is robustly sustained by several non-redundant mechanisms, with distinct ranges of activity.

Based on the available evidence, a current model proposes that net transcription at BES results from a balance between two major forces: i) a repressive epigenetic gradient spreading from the telomere which gradually attenuates Pol I transcription elongation and/or RNA processing; ii) the recruitment of transcription factors and Pol I machinery to BES promoter which drives transcription initiation (Gunzl *et al.*, 2014)

Transcription attenuation towards the telomere is supported by several evidences: despite *VSG* monoallelic expression, transcription is actually initiated at several silent BESs simultaneously, which was observed in cell populations (Ansorge *et al.*, 1999, Vanhamme *et al.*, 2000, Pays *et al.*, 1990) and in single-cells of BSFs and PFs (Kassem *et al.*, 2014); RNA processing (processing, export to cytoplasm and stability) is more efficient for transcripts from active than silent BESs (Vanhamme *et al.*, 2000); several chromatin-related factors silence BES promoter regions but not the terminal *VSGs* (Table 1); transcripts from promoter-proximal regions are more abundant than from promoter-distal regions of silent BESs; TbRAP1–mediated silencing is weaker at BES promoter-proximal regions that at terminal ones (Yang *et al.*, 2009); shut down of BES during slender to stumpy differentiation was associated with gradual Pol I stalling towards the telomere (Amiguet-Vercher *et al.*, 2004); finally, in a model of induced ectopic *VSG* expression, BES reactivation occurs gradually from the promoter towards the telomere (Batram *et al.*, 2014).

Transcription initiation control at BES promoters is sustained by more recent findings. Even when silent BESs are derepressed, transcripts from promoter-proximal regions are still 100 to 1000-fold lower when compared with the active BES (Yang *et al.*, 2009, Figueiredo *et al.*, 2008), indicating that additional mechanisms prevent them from full activation. Importantly, subunits of the Pol I transcription factor CITFA, predominately bind at the promoter of active BES in comparison with silent BESs, and

this correlates with higher levels of Pol I occupancy and promoter-proximal primary transcripts (Nguyen *et al.*, 2014). Collectively, this suggests the existence of higher rates of transcription initiation at the active BES promoter. On the contrary, at the promoters of silent BESs, access to CITFA and Pol I is somehow constrained and by consequence transcription initiation is reduced (Nguyen *et al.*, 2014).

Some epigenetic factors are probably controlling primarily transcription initiation at BES promoters. These include the histone chaperone TbFACT and the chromatin architectural protein TDP-1. TbFACT (TbSPT16 subunit) is highly enriched at the silent BES promoters (Denninger *et al.*, 2010) and it is necessary for chromatin condensation and transcriptional repression at its promoter regions (Denninger & Rudenko, 2014) (Table 1). Contrary to most known BES regulators, which function on BES silencing, TDP-1 regulates the active BES (Narayanan & Rudenko, 2013). TDP-1 is specifically enriched at active Pol I loci (active BES and the rDNA loci) particularly at promoter regions, in inverse correlation with nucleosome density. Depletion of TDP-1 drives chromatin condensation and shut down of transcription at these Pol I loci (Narayanan & Rudenko, 2013) (Table 1). In light of these observations, TDP-1 was proposed to 'replace' the nucleosomes at active BES, creating and/or maintaining an open chromatin structure that allows maximal rates of Pol I transcription initiation/elongation.

Cell cycle link with transcriptional regulation of BES is evidenced by a number of observations, even though we ignore how replication occurs in the BESs or other *VSG* loci. TbORC1/CDC6 and the mini-chromosome maintenance-binding protein (TbMCM-BP) which make part of the pre-replication complex, are necessary for fully silencing BES-linked *VSGs*, *MVSGs* and procyclin genes (Tiengwe *et al.*, 2012, Benmerzouga *et al.*, 2013, Kim *et al.*, 2013); cohesin, that maintains cohesion between sister chromatids during mitosis, is critical for the inheritance of the transcriptionally active BES (Landeira *et al.*, 2009). Also, the histone chaperones TbFACT and TbCAF-1b specifically repress BESs at the S- and G2/M-phase of the cell cycle (Denninger *et al.*, 2010, Alsford & Horn, 2012) (Table 1).

Table 1. Chromatin-associated factors involved in BES and procyclin expression in *T. brucei* bloodstream forms.

Functional category	Factor	Necessary for BES silent state?	Necessary for BES active state?	Necessary for procyclin silent state?	Other related roles	References
Histones	**H3**	YES: PRO	n.d.	n.d.	n.d.	(Alsford & Horn, 2012)
	H1	YES: PRO	n.d.	n.d.	Chromatin compaction particularly at silent BES promoters	(Povelones *et al.*, 2012)
Histone modifying-enzymes	**DOT1B**	YES: PRO;*VSG*	NO	n.d.	Rapid *VSG in situ* switching	(Figueiredo *et al.*, 2008)
	DAC3	YES: PRO	n.d.	n.d.	-	(Wang *et al.*, 2010)
Histone chaperones	**FACT**	YES: PRO	YES: *VSG*	n.d.	Maintaining chromatin compaction at silent BES promoters/chromatin opening at active BES; BES silencing is G2/M-dependent	(Denninger & Rudenko, 2014, Denninger *et al.*, 2010)
	NLP	YES: PRO	YES: PRO;*VSG*	YES: PRO; *EP1*	Binds equally to active and silent BESs; interacts with TbISWI	(Narayanan *et al.*, 2011, Stanne *et al.*, 2015)
	ASF1A	YES: PRO	n.d.	n.d.	BES silencing at all cell cycle stages	(Alsford & Horn, 2012)
	CAF-1b	YES: PRO	n.d.	n.d.	BES silencing is S- and G2/M-dependent	
Chromatin remodeler	**ISWI**	YES: PRO > *VSG*	NO	YES: PRO	Binds equally to active and silent BESs:interacts with TbNLP	(Hughes *et al.*, 2007, Stanne *et al.*, 2011, Stanne *et al.*, 2015)
Telomeric protein	**RAP1**	YES: PRO < *VSG*	NO	n.d.	Repression of *MVSGs*	(Yang *et al.*, 2009, Pandya *et al.*, 2013)
Lamin-like protein	**NUP-1**	YES: PRO < *VSG*	n.d.	YES: *EP, GPEET*	Positioning of ESB to nuclear periphery during PF differentiation; Repression of *VSG* switching	(DuBois *et al.*, 2012)
Chromatin architectural protein	**TDP-1**	n.d.	YES: *VSG*	n.d.	Maintaining chromatin open at active Pol I units (BES and rDNA)	(Narayanan & Rudenko, 2013)
Pol I transcription factor	**CITFA**	n.d.	YES: *VSG*	n.d.	High CITFA occupancy at active Pol I promoters (BES and rDNA)	(Nguyen *et al.*, 2014)
Replication -related	**ORC1/CDC6**	YES: *VSG*	NO	n.d.	Repression of *VSG* switching	(Benmerzouga *et al.*, 2013)
	MCM-BP	YES: PRO < *VSG*	NO	YES: *EP1*	Repression of mini-chromosome *VSG*	(Kim *et al.*, 2013)
	Cohesin complex	YES: *VSG*	YES: *VSG*	n.d.	Inheritance of active BES; Repression of *VSG in situ* switching	(Landeira *et al.*, 2009)

n.d., not determined; PRO, at promoter region; PRO > VSG, silencing effect is stronger at BES promoter region than at *VSG*; PRO < VSG silencing effect is stronger at *VSG* than at BES promoter

Some kind of monitoring of the number of VSG molecules also seems to be connected with the cell cycle. When VSG synthesis is suppressed by RNAi-mediated knockdown, there is a general protein translation block (Smith *et al.*, 2009) and parasites do not proceed to cytokinesis (Sheader *et al.*, 2005). This might suggest that *T. brucei* evolved a VSG 'counting' mechanism that monitors the number of VSG transcripts and/or proteins, turning it into a decisive checkpoint for parasite replication and survival.

VSG switching is also regulated by epigenetic factors involved in BESs transcriptional control such as: histone H1 (Povelones *et al.*, 2012), TbNUP-1 (DuBois *et al.*, 2012) and TbORC1/CDC6 (Benmerzouga *et al.*, 2013) in which all repress *VSG* switching; TbDOT1B which is required for fast transcriptional *in situ* switching (Figueiredo *et al.*, 2008); histone acetyltransferase TbHAT3, which inhibits BES-linked *VSG* recombination (Glover & Horn, 2014). Telomeres are propitious sites for double-strand DNA breaks (DSBs) and ectopic HR that trigger *VSG* switching (Boothroyd *et al.*, 2009, Li, 2015). Consistent with this, the telomere-binding proteins TbTIF2 and TbTRF suppress *VSG* switching due to, at least for TbTIF2, its ability to restrain the amount of DSBs at telomeres (Jehi *et al.*, 2014a, Jehi *et al.*, 2014b).

Despite the recent advances, there are still many questions unanswered on the control of VSG monoallelic expression: What are the mechanisms behind differential regulation of active/silent BES promoters? Are epigenetic factors and chromatin structure driving the transcriptional status of BES or are these a downstream effect of BES transcriptional state? Given that downregulation of active BES is associated with derepression of silent BESs (upon depletion of TbFACT and TbNLP, Table 1) or with ectopic expression of a second *VSG* (Batram *et al.*, 2014) is there any cross-talk between silent and active BESs? Do parasites 'count' the number of VSG molecules?

1.3.3 CONTROL OF PROCYCLIN LOCI

Procyclin loci

When BSFs differentiate to PFs, the VSG coat is shed and replaced within hours by a high number of copies (10^6-10^7) of other surface GPI-anchored glycoproteins, the procyclins (Roditi *et al.*, 1989, Clayton & Mowatt, 1989). Procyclin proteins can be classified as EP or GPEET, according to the presence of internal dipeptide Glu-Pro (EP) repeats or pentapeptide Gly-Pro-Glu-Glu-Thr (GPEET) repeats in their C-terminal domain (Vassella *et al.*, 2001).

Similarly to *VSGs*, procyclin genes are expressed from polycistronic units transcribed by Pol I (Gunzl *et al.*, 2003). However, in contrast with BESs, procyclin loci are short polycistronic units of ~6-9 kbs located at chromosomal internal positions flanked by Pol II PTUs (Rudenko *et al.*, 1990, Liniger *et al.*, 2001). There are four procyclin loci located in the two alleles of chromosome 6 and 10. Each procyclin locus encodes two procyclin genes in tandem, followed by one or more procyclin-associated genes (*PAGs*), which potentially encode membrane or secreted proteins (Roditi *et al.*, 1998, Haenni *et al.*, 2006). *T. brucei* Lister 427 strain has four EP isoforms and a GPEET isoform (Acosta-Serrano *et al.*, 1999). The procyclin loci on chromosome 10 encode *EP1-EP2* procyclin genes followed by several *PAG* genes, except for *PAG3*. In several strains, such as Lister 427, the two allelic loci on chromosome 10 are heterozygous, encoding distinct *PAGs* (Koenig-Martin *et al.*, 1992, Liniger et al., 2001) On chromosome 6, procyclin loci are identical and have four genes consisting on *GPEET-EP3-PAG3* and a gene related to ESAG2 (*GRESAG2*) (Roditi *et al.*, 1998). In *T. brucei* Lister 427 genome this locus is preceded by additional units encoding only *EP3-PAG3* and belong to a Pol II-transcribed unit (TriTrypDB 9.0) (Fig. 6).

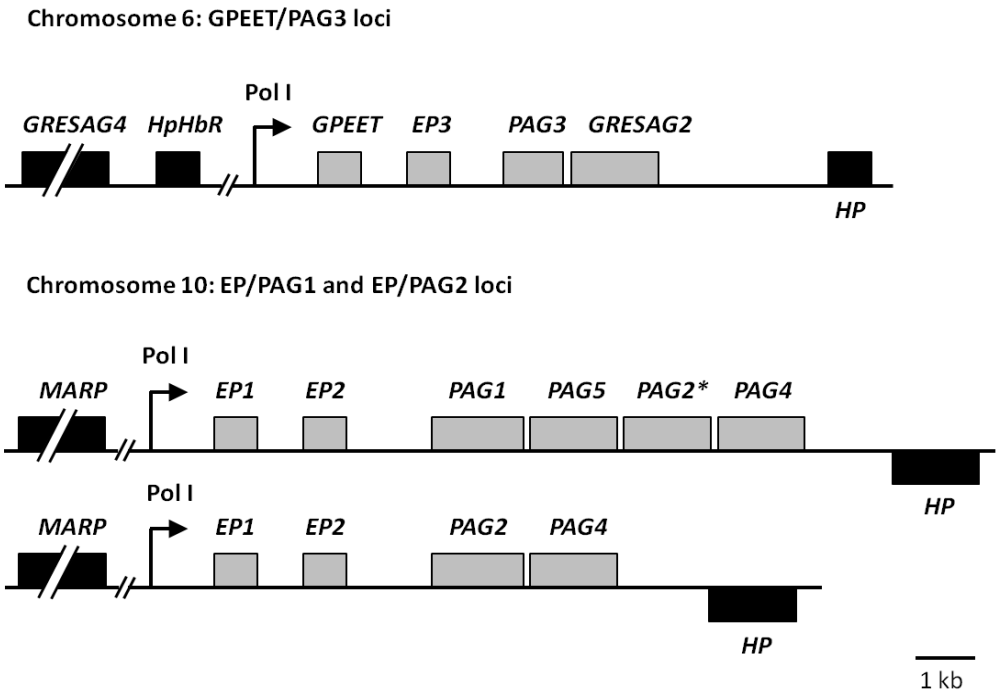


Figure 6. Genomic context of procyclin loci in *T. brucei*. Four procyclin loci exist on the two alleles of chromosome 6 and 10. Arrows indicate Pol I transcription start sites at procyclin promoters. Grey boxes indicate coding sequences of procyclin loci genes. Black boxes indicate open reading frames of Pol II-transcribed genes. Flanking genes of procyclin loci are based on the annotated genome of *T. brucei* TREU 927 and Lister 427 (TriTrypDB 9.0). Organization of genes at procyclin loci located on chromosome 10 is based on previous genome analysis of *T. brucei* Lister 427 (Koenig-Martin *et al.*, 1992, Liniger et al., 2001) and AnTat 1.1 (Haenni *et al.*, 2009). *EP*, gene encoding a procyclin with dipeptide repeats; *GPEET*, gene encoding a procyclin with pentapeptide repeats; *GRESAG*, gene related to ESAG; *HP*, gene encoding hypothetical protein; HpHbR, gene encoding the haptoglobin-hemoglobin receptor; *MARP*, gene encoding a microtubule-associated repetitive protein; *PAG*, procyclin-associated gene; The figure is drawn to scale (adapated from Haenni et al., 2009).

Transcriptional and post-transcriptional regulation

EP and GPEET procyclins are differentially expressed during early to late PF transition, which happens *in vivo* when PFs migrate to the ectoperitrophic space in the fly midgut. In early PFs, both EP and, predominantly, GPEET procyclins are synthesized whereas in late PFs GPEET expression is repressed (Vassella *et al.*, 2000, Acosta-Serrano *et al.*, 2001, Vassella *et al.*, 2001). Therefore, in contrast with BESs, transcription can occur simultaneously from two or more procyclin expression sites (Mowatt & Clayton, 1987, Ruepp *et al.*, 1997). Upon differentiation to PFs, transcription at active BES is rapidly shut down and procyclins start to be highly

expressed within the first hours (Roditi *et al.*, 1989, Pays *et al.*, 1993). This tight control of procyclin expression is achieved at both the transcriptional and post-transcriptional levels. Transcription at the procyclin loci is controlled at the level of initiation and elongation. In BSFs, transcription initiation at procyclin promoters is ~10-fold lower than in PFs (Pays *et al.*, 1990). Transcription elongation is also inhibited along the procyclin units. In the first hours after inducing differentiation with several stimuli, transcription is upregulated strongly at promoter-distal genes such as *GRESAG2* (40-fold) than at the promoter-proximal genes such as the procyclin genes (6-fold) (Vanhamme *et al.*, 1995a).

Epigenetic regulation is also controlling expression of procyclin genes. In PFs the ESB disappears and, contrary to BES, the procyclin loci are transcribed from a peripheral region of the nucleolus, to a region of high Pol I transcriptional activity proximal to where rDNA units locate (Landeira & Navarro, 2007). If in one hand differential nuclear positioning seems to distinguish the activation of BESs and procylin loci (Landeira & Navarro, 2007, Navarro & Gull, 2001), in the other, many of the epigenetic factors involved in BES transcriptional silencing also act as repressors of the procyclin loci, suggesting a shared epigenetic mechanisms for Pol I silencing. These factors include TbNUP-1, TbNLP, TbISWI and TbMCM-BP (DuBois *et al.*, 2012, Narayanan *et al.*, 2011, Stanne *et al.*, 2011, Kim *et al.*, 2013) (Table 1). Involvement of the nuclear lamina (TbNUP-1) in both silencing of BESs and procyclin loci suggests a common repression mechanism via heterochromatin regions close to the nuclear periphery. Perhaps both active BES and procyclin loci are pulled away from repressive chromatin via recruitment to the ESB or the nucleolus in the BSFs and PFs, respectively.

Post-transcriptional regulation is critical to define procyclin mRNA abundance in BSFs and PFs. Elements at the 3'UTRs of both EP and GPEET procyclin mRNAs strongly destabilize and induce rapid degradation of the transcripts leading to ~11-fold downregulation of mRNA levels in BSFs (Hotz *et al.*, 1998). In particular, the presence of 26mer containing two U-rich tracts in the 3'UTRs is critical for the stability of procyclin mRNAs, diminishing mRNAs half-lives about 5-fold in BSFs (Hotz *et al.*, 1997). Such 3'UTR elements together with the CDS of procyclin transcripts also mediate

translation repression of procyclin mRNAs (Hotz *et al.*, 1997, Schurch *et al.*, 1997). Reduced translational efficiency of *EP*, *GPEET* and *PAG* mRNAs in BSFs was also recently corroborated by ribosome profiling (Vasquez *et al.*, 2014, Jensen *et al.*, 2014). Given the existence of detectable amounts of procyclin mRNAs in BSFs, translational repression should be responsible, at least in part, for preventing the inappropriate expression of procyclin proteins in the BSF stages. Altogether the several transcriptional and post-transcriptional levels of regulation were predicted to account for at least 1000-fold downregulation of procyclin expression in BSFs when compared with PFs (Hotz *et al.*, 1998).

In addition, repression of *GPEET* expression in late PFs is also determined post-transcriptionally (Vassella *et al.*, 2000). A *cis*-regulatory element in the 3'UTR of *GPEET* mRNA, termed the 'glycerol responsive element', destabilizes the transcript in late PFs in the fly gut and *in vitro* in response to high glucose or glycerol deprivation (Vassella *et al.*, 2000, Vassella *et al.*, 2004, Urwyler *et al.*, 2005). Besides, during transition from early to late PFs, *GPEET* mRNA stability decreases significantly and this correlates with an accumulation of alternatively processed *GPEET* transcripts e.g. lacking a poly(A) tail, with extended 3´UTRs or carrying oligo(U) tails (Knusel & Roditi, 2013).

1.4 HISTONE H1

1.4.1 STRUCTURAL AND FUNCTIONAL DIVERSITY

Histone H1 or linker histone, binds to the DNA that links nucleosomes, as well as DNA entering and exiting the nucleosome core (Noll & Kornberg, 1977, Allan *et al.*, 1980). The binding of H1 stabilizes nucleosomes and promotes the folding of chromatin into higher-order chromatin structures (Robinson & Rhodes, 2006). Histone H1 affects nucleosome spacing and mobility and can stabilize the position of nucleosomes by inhibiting spontaneous transitions in the nucleosomal structure and chromatin remodeling (Ramachandran *et al.*, 2003). The H1/core nucleosome ratio is variable and it can range from 0.5 in embryonic stem cells up to 1.3 in chicken erythrocytes (Woodcock *et al.*, 2006, Bates *et al.*, 1981).

Metazoan H1 exhibits a tripartite structure with a short flexible N-terminal tail, a globular domain containing a winged-helix fold and a long, extended, extremely lysine rich C-terminal tail. N- and C-terminal domains are less structured and so far, only the crystallographic structure of the globular domain of H1 has been determined (Ramakrishnan *et al.*, 1993). Chromatin reconstitution with histone H1 fragments consisting of the globular and C-terminal domain have shown that from the three structural domains, the C-terminal appears to be the critical one for chromatin folding (Allan *et al.*, 1980, Allan *et al.*, 1986). In lower eukaryotes, histone H1 does not have this classical structure. In the ciliate protozoa *Tetrahymena termophila*, H1 lacks the globular domain and is similar to the C-terminal domain of metazoan H1 (Hayashi *et al.*, 1987), whereas H1 of the yeast *Saccharomyces cerevisiae*, possesses two globular domains (Patterton *et al.*, 1998) (Fig.7).

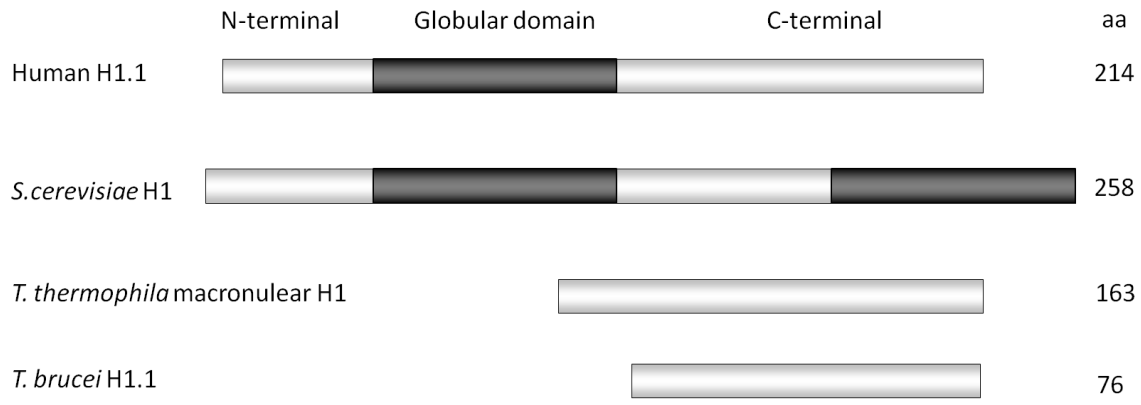


Figure 7. Structural diversity of histone H1 proteins. Diagram depicting the major structural domains of histone H1 in different eukaryotes. The number of amino acid (aa) residues in each H1 protein is indicated. Metazoans typically have several H1 variants, whereas other organisms such as *S. cerevisiae* and *T thermophila* only contain a single H1 protein. *T. brucei* encodes several H1 isoforms. H1 in metazoans (e.g. human H1.1) has a characteristic tripartite structure consisting of a short flexible N-terminal tail, a globular domain (dark grey) with a winged-helix fold, and a long, extremely lysine rich C-terminal tail. By contrast in *S. cerevisiae* H1 contains two globular domains. In lower eukaryotes such as *T. thermophila* and *T. brucei* (e.g. H1.1), H1 is a smaller protein that lacks a globular domain and resembles the C-terminus of H1 from metazoans.

Histone H1 is the least conserved of all histones and it is frequently encoded by a multigene family of highly heterogeneous members (Happel & Doenecke, 2009, Izzo *et al.*, 2008). In mouse and humans, histone H1 is encoded by multiple non-allelic

genes, which form several protein variants (11 variants in humans). Variants can be subgrouped according to their temporal and spatial expression. For instance, in mammals there are replication-dependent and replication-independent H1 variants. Besides, there are somatic H1 variants which are present in almost every cell and H1 subtypes which exist mainly in terminally differentiated cells or gametes (Happel & Doenecke, 2009). Different H1 protein subtypes can also differ in their chromatin-binding affinities, genome distribution and PTMs (Harshman *et al.*, 2013, Happel & Doenecke, 2009). In contrast with organisms that have a H1 gene family encoding several isoforms, other organisms have very few or only a single H1 gene such as *Saccharomyces cerevisiae* (Ushinsky *et al.*, 1997).

Among histones, histone H1 has been then least studied (Harshman *et al.*, 2013). Much of what we know about the functions of H1 comes from *in vitro* studies. These studies indicated that two principal functions of H1 histones are to stabilize the DNA entering and exiting the core particle and to facilitate the folding of nucleosome arrays and chromatin condensation (Ramakrishnan, 1997, Wolffe *et al.*, 1997, Bednar *et al.*, 1998, Thoma *et al.*, 1979). *In vitro* transcription analysis of chromatin reconstituted in the absence or presence of histone H1 also suggested that H1 acts primarily as a transcriptional repressor (Laybourn & Kadonaga, 1991, Shimamura *et al.*, 1989). More recently, the functions of H1 have been studied *in vivo* using gene knockdown/knockout or overexpression approaches. Both in higher and lower eukaryotes, H1 was confirmed to be a global regulator of chromatin architecture *in vivo* (Fan *et al.*, 2005, Masina *et al.*, 2007, Hashimoto *et al.*, 2010, Barra *et al.*, 2000). However, some studies also revealed that H1 is not a global transcription repressor, but rather regulates the transcription of specific and limited sets of genes by a yet unknown mechanism. Such is observed for the mammalian H1 with the typical tripartite structure, as well as for the unconventional H1 of yeast (Hellauer *et al.*, 2001), or the simpler H1 of *T. thermophila* (Shen & Gorovsky, 1996). Interestingly, in the case of mouse embryonic stem cells, when H1 is depleted only a specific set of genes is up or downregulated, even though chromatin gets globally less compacted (Fan *et al.*, 2005). In addition, a number of studies provide evidence that different H1 variants play different biological roles, including regulation of gene expression. For

instance, selective knockdown of each H1 variant in a human cell line alters the expression of distinct subsets of genes, and has different phenotypic outcomes at the cell cycle and cell survival levels (Sancho *et al.*, 2008).

H1 is also involved in other biological processes, including inhibition of DNA repair, telomere maintenance in yeast and mammalian cells (Downs *et al.*, 2003, Murga *et al.*, 2007), silencing of retrotransposons at the rDNA locus in *D. melanogaster* (Vujatovic *et al.*, 2012) and differentiation and virulence in *Leishmania major*, a trypanosomatid closely related with *T. brucei* (Smirlis *et al.*, 2006). Despite the conservation of some functions, it remains unclear why H1 is dispensable for growth in unicellular eukaryotes such as *S. cerevisiae* and *T. thermophila* (Shen *et al.*, 1995, Shen & Gorovsky, 1996, Patterton *et al.*, 1998), but is absolutely essential for mammalian embryonic development (Fan *et al.*, 2003, Zhang *et al.*, 2012) and *Drosophila melanogaster* development (Lu *et al.*, 2009). Besides, studies in the fungus *Ascobolus immersus* (Barra *et al.*, 2000) *and Saccharomyces cerevisiae* (Downs *et al.*, 2003) revealed that even though H1 is not essential for growth, it results in shortened life span, indicating that loss of H1 has important consequences even in these lower eukaryotes.

Studies in unperturbed chromatin *in vivo* have shown that rather than being a mere structural protein that statically binds to chromatin, H1 interacting with chromatin is very dynamic. These are highly mobile proteins which exchange continuously between chromatin regions (Misteli *et al.*, 2000). H1 dynamically interacts with other highly mobile chromatin architectural proteins, the HMGB proteins, and both can act as regulators of nucleosome accessibility and transcription. Similarly to H1, HMGBs can bind to linker DNA and with DNA entering/exiting the nucleosome (Thomas & Stott, 2012). In different cellular contexts, H1-HMGB interchanging drives opposite chromatin structure transitions, with H1 contributing to a more compacted chromatin and its replacement by HMGBs leading to a looser chromatin structure. Consequently, HGMBs can compete with H1 in order to activate particular transcriptional programs (Ogawa *et al.*, 1995, Travers, 2003, Catez *et al.*, 2004, Nalabothula *et al.*, 2014). In other cases, H1 and HMGB1 were observed to co-

act in the formation of heterochromatin, chromatin remodeling and transcriptional silencing (El Gazzar *et al.*, 2009).

Another interacting partner of H1 recently described is Phosphatase and TENsin homologue (PTEN), a multifunctional tumor suppressor. PTEN stabilizes H1 binding to chromatin and H1 occupancy at gene promoters. Physical interaction between PTEN and H1 is essential for global chromatin condensation that apparently underlies PTEN-mediated genome-wide transcriptional regulation (Chen *et al.*, 2014). This stands as an example in which H1 seems to act as an effector of gene regulatory signals, bridging control of chromatin structure and regulation of gene expression patterns.

The mechanism behind H1-mediated control of gene expression was recently studied in detail in human cells. Within the first minute of progesterone hormone stimulation, a coordinated recruitment of chromatin remodeling complexes, histone modifications and kinase activation signals leads to histone H1 displacement from promoter regions which is a prerequisite for the following transcriptional activation of hormone target genes (Vicent *et al.*, 2011). This study remarkably implies histone H1 as one of the elementary components regulating promoter chromatin architecture necessary in the initial steps of promoter activation.

Similarly to what is already known for core histones, the functional diversity of H1 is in part a result of the different of H1 variants and their PTMs. In human cells, H1.4 variant phosphorylated at specific amino acids is enriched at active Pol I and Pol II transcribed genes (Zheng *et al.*, 2010).

1.4.2 HISTONE H1 IN *T. BRUCEI*

T. brucei H1 proteins differ considerably from metazoan H1. *T. brucei* H1 are predicted to be small proteins (7-8 kDa) that contain a very lysine polypeptide sequence, similar to the typical C-terminal domain of metazoan H1 and lack a central globular domain (Burri *et al.*, 1995, Schlimme *et al.*, 1993). In the genome of *T. brucei* Lister 427, H1 proteins are predicted to be encoded by a multigene family with at least five genes, arranged in tandem and distributed in two separate clusters of the same

polycistronic unit, located on chromosome 11. Each gene encodes different H1 protein isoforms, or sequence variants, which can be sorted into three classes according to a few amino acid substitutions in the N-terminal (TriTrypDB 9.0) (Gruter & Betschart, 2001). Even among the *T. brucei* species, H1-coding genes vary significantly in number and sequence between subspecies and even strains (Gruter & Betschart, 2001).

The small molecular weight and high hydrophobicity of trypanosomatid H1 proteins initially precluded its identification by standard gel analysis. *T. brucei* H1 was first identified by separation on Triton acid-urea PAGE (according to its hydrophobicity) and two-dimensional gel electrophoresis (hydrophobicity/charge and molecular weight) optimized for high resolution. This analysis revealed the presence of five to six H1 protein variants and/or PTMs (Burri *et al.*, 1993, Schlimme *et al.*, 1993). Interestingly, the type and abundance of these H1 variants/PTMs varies between BSFs and PFs, with one of them being overexpressed in BSFs (Burri *et al.*, 1994, Burri *et al.*, 1995). The distinct set of H1 proteins in the two life cycle stages was proposed to contribute for the fact that chromatin in BSFs is more condensed than in PFs (Schlimme *et al.*, 1993).

Dissociation of H1 from the DNA fiber under high-salt concentrations *in vitro* showed that H1 is essential for the regular spacing of nucleosomes and salt-dependent chromatin condensation both in BSFs and PFs (Burri *et al.*, 1993, Burri *et al.*, 1994). *T. brucei* H1 is displaced from the DNA filament at lower salt concentrations compared with H1 of higher eukaryotes, indicating that *T. brucei* H1 interacts more weakly with chromatin (Burri *et al.*, 1993). Although this weak binding was initially proposed to explain why chromatin in *T. brucei* does not form the 30-nm fiber (Schlimme *et al.*, 1993), it was later shown that *T. brucei* H1 is capable of reconstituting dense chromatin tangles and a 30-nm fiber in mammalian nucleosome filaments *in vitro* (Burri *et al.*, 1995). Therefore, the less compacted chromatin of *T. brucei* cannot be explained alone by H1, but should result from the joined properties of *T. brucei* H1 and other chromatin components, including core histones, which bind weakly to chromatin when compared with their counterparts in higher eukaryotes (Bender *et al.*, 1992, Hecker *et al.*, 1994).

Very little is known about the functions of histone H1 in *T. brucei in vivo*. Until recently, the functional properties of *T. brucei* H1 had only been analyzed in *in vitro* systems using isolated nucleosome filaments and studying the effects of salt-dependent dissociation and reconstitution of H1 (Burri *et al.*, 1994, Burri *et al.*, 1995, Schlimme *et al.*, 1993). These studies have corroborated repeatedly the role of H1 in chromatin condensation and the presence of distinct H1 protein isoforms and/or PTMs in BSFs and PFs in *T. brucei*. Besides, they have also shown that *T. brucei* histone H1 phosphorylation is inversely correlated with chromatin condensation (Burri *et al.*, 1995). This is in line with similar observations done in *T. thermophila* (Lin *et al.*, 1991) but also, more recently, in higher eukaryotes (Lopez *et al.*, 2015) and sustain the hypothesis that H1 phosphorylation triggers transient chromatin decondensation states facilitating access, for instance, to RNA polymerases (Roth & Allis, 1992, Zheng *et al.*, 2010). Phosphorylation of *T. brucei* H1 is probably restricted to its N-terminus, which is the only containing serine and threonine residues, the two most common phosphorylation sites. Curiously, the three H1 classes of *T. brucei* vary in the number of serine and threonine residues at the N-terminus (Gruter & Betschart, 2001).

In parallel to my PhD work, another group addressed the roles of H1 in *T. brucei in vivo*. RNAi-mediated depletion of the several H1 isoforms lead to a reduction of heterochromatin domains inside the nucleus of BSFs, observed by electron microscopy, and to a more accessible chromatin observed upon digestion with Micrococcal Nuclease (MNase) (Povelones *et al.*, 2012). Interestingly, when H1 was depleted, derepression of a silent BES promoter was observed and *VSG* switching frequencies increased, suggesting a role of H1 in antigenic variation (Povelones *et al.*, 2012).

At the onset of this thesis, in 2011, no *in vivo* studies were available to address the role of H1 in *T. brucei*. Considering the unusual gene expression and genome organization of this parasite, I decided that for my PhD work it would be interesting to understand the relevance of H1 in the biology of this parasite, at the level of genome function and phenotype inside the host.

2 AIMS

In the present thesis I had four major aims: (i) to determine the importance of *T. brucei* histone H1 in overall chromatin structure; (ii) to study the importance of H1 in genome-wide gene expression regulation; (iii) to investigate the role of H1 in transcription initiation/elongation in *T. brucei*; (vi) to examine the general relevance of H1 for *T. brucei* infection in the mammal.

In addition, we also addressed the role of H1 in response to DNA damage, as a possible although indirect read-out for H1 involvement in HR. (see Results 3.1).

Since *T. brucei* H1 is encoded by a multigene family, complete H1-null mutants were difficult to generate. A loss-of-function approach was used in which RNAi-mediated knockdown of all H1 variants allowed a reduction in total H1 levels. Two H1-mutant clones were used throughout the work to study the effects of H1 depletion in the several topics referred above (see Results 3.1).

To test the role of H1 in VSG transcription, we adapted the method of 4-thiouridine (4sU)-metabolic labeling of nascent transcripts to *T. brucei* for the first time. Due to the potential promising applications of this method in the study of *T. brucei* transcription, part of my thesis was devoted to improve this method in *T. brucei* (see Results 3.3).

3 RESULTS

3.1 TRYPANOSOMA BRUCEI HISTONE H1 INHIBITS RNA POLYMERASE I TRANSCRIPTION

The majority of results presented below were published in the Molecular Microbiology peer-reviewed journal. The article in the publication format can be found in the Annexes.

Ana C. Pena[1], Mafalda R. Pimentel[1], Helena Manso[1], Rita Vaz-Drago[1], Daniel Pinto-Neves[1], Francisco Aresta-Branco[1], Filipa Rijo-Ferreira[1], Fabien Guegan[1], Luis Pedro Coelho[1†], Maria Carmo-Fonseca[1], Nuno L. Barbosa-Morais[1,2] and Luisa M. Figueiredo[1]

[1]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, Edifício Egas Moniz, 1649-028 Lisboa, Portugal.

[2]Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK.

†Present address: European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany.

3.1.1 AUTHOR CONTRIBUTIONS

Ana C. Pena and Luisa M. Figueiredo designed the experiments and wrote the manuscript. Ana C. Pena performed most of the experiments and data analysis and generated most of the figures and tables. Mafalda R. Pimentel performed the majority of the ChIP assays and contributed to essential background experiments. Helena Manso and Daniel Pinto-Neves analyzed RNA-Seq data, generated the heat map, volcano plot, GSEA plot, RNA-Seq *vs* qPCR correlation plot. Rita Vaz-Drago contributed

with expertise and critical advice and helped in experimental design of 4sU experiments. Franscisco Aresta-Branco performed the FAIRE assays and part of the ChIP assays. Filipa Rijo-Ferreira generated the H1 RNAi transgenic parasites, performed the Southern blots and generated part of the RNA-Seq cDNA libraries. Fabien Guegan generated part of the RNA-Seq cDNA libraries. Luis Pedro Coelho and Nuno Barbosa-Morais contributed with fundamental discussion and guidance to RNA-Seq analysis. Maria Carmo-Fonseca participated in discussion and experimental design of 4sU experiments. Luisa M. Figueiredo generated the H1 RNAi construct, mentored and supervised the work. All authors revised part or the entire manuscript.

3.1.2 RESULTS

3.1.2.1 *T. brucei* histone H1 family comprises three hypothetical protein classes

The current version of *T. brucei* genome (Berriman *et al.*, 2005) indicates that there are five unique H1 genes on chromosome 11. As previously suggested (Gruter & Betschart, 2001), these genes are organized in two clusters belonging to a single polycistronic unit and are interspersed by five non-histone genes. We showed by PCR and Southern blot that this gene organization is correct in the *T. brucei* Lister 427 strain, although other H1 alleles may be missing from the genome database (Fig. S1). Since each gene is predicted to encode a unique protein, we assigned them paralogue numbers from H1.1 to H1.5 according to the most recent nomenclature for histone variants (Talbert *et al.*, 2012) (Fig. 1A). Alignment of the five H1 protein sequences reveals that the first seven aa define three different types of N-terminal sequences (Fig. 1B) (Gruter & Betschart, 2001). We classified them accordingly and named them MAKTT (H1.1), MAKASA (H1.2, H1.3 and H1.5) and MNNTT (H1.4). The sequence of the 3'UTRs is almost identical between H1 classes (96% on average), whereas the 5'UTRs are less conserved and specific for each of the three classes of H1 genes (Fig. S2). Previous studies reported that one H1 protein variant/PTM was predominant in BSFs (identified by Triton acid-urea and SDS-tricine PAGE) (Burri *et al.*, 1994, Schlimme *et al.*, 1993) which likely corresponded to a MAKASA H1 (Burri *et al.*, 1995, Gruter & Betschart, 2001).

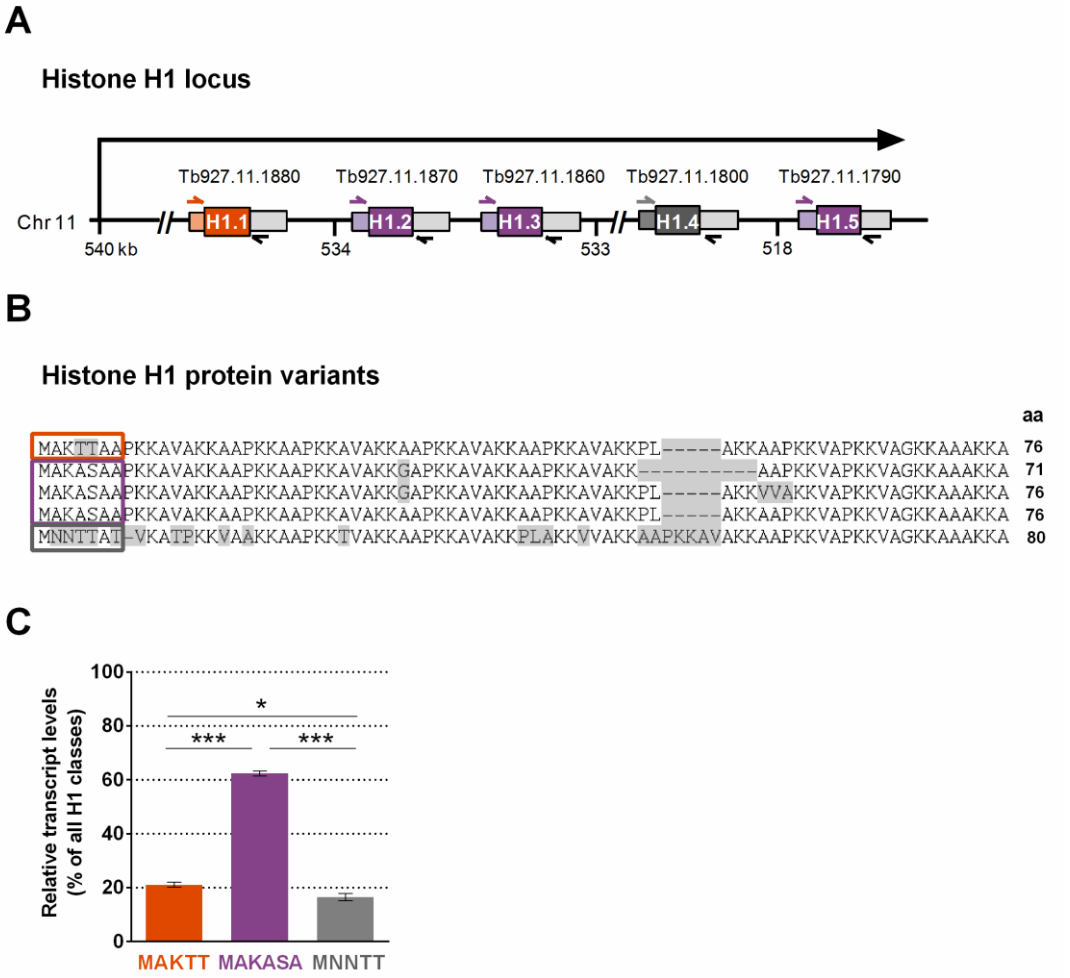


Figure 1. Histone H1 variants in *T. brucei.* (A) Diagram shows that in *T. brucei* H1 gene family is encoded by five genes (H1.1-H1.5) in two clusters ~14.5 kb apart of the same polycistronic unit in chromosome 11. Arrow indicates the transcription start site and direction of transcription. Primers used for qPCR are indicated by colour-code: orange – primer specific for MAKTT 5'UTR; purple – primer specific for MAKASA 5'UTR; grey – primer specific for MNNTT 5'UTR; black – primers that anneal with the 3'UTRs, which are conserved across all H1 genes. Coding sequences are coloured according to the 5'UTR and N terminal sequence (see panel B). Gene IDs correspond to *T. brucei* TREU 927, which has a more complete genome annotation (TriTrypDB 9.0). **(B)** Alignment of the predicted amino acid (aa) sequences of H1 variants. Divergent aa are shadowed in grey. Gaps (-) were introduced for best alignment. H1 variants were grouped into three classes according to the N-terminal sequences: MAKTT (orange), MAKASA (purple) and MNNTT (grey). **(C)** MAKASA is the most abundant histone H1 class at the RNA level. Transcript levels of each H1 class were quantified by qPCR and normalized to the total transcripts of all H1 classes. Eleven independent experiments were analyzed. Results are shown as mean ± SEM. Statistical significance was determined by non-parametric Mann-Whitney U tests. * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$.

To quantify the mRNA levels of H1 variants we performed quantitative real-time PCR (qPCR) with primer pairs specific for each class of linker histones (Fig. S3). Transcripts of class MAKASA are clearly the most abundant in BSF, 3-fold more relative to classes MAKTT and MNNTT ($P < 0.001$) (Fig. 1C). MNNTT and MAKTT have comparable transcript levels although MAKTT transcripts are slightly more abundant ($P < 0.05$). The observed transcript levels follow the gene copy number for each H1 class. According to the genome database MAKASA class contains 3 gene copies while MAKTT and MNNTT are encoded by a single gene. Assuming that transcription rate is the same for all H1 genes (inside the same polycistronic unit) this suggests that gene copy number is the main determinant of H1 mRNA levels in the BSF stage. Altogether, our results support the previous studies where one of the H1 protein variants was found to be more expressed, and show that the most abundant H1 in BSF of *T. brucei*, at the mRNA level, is class MAKASA.

3.1.2.2 Depletion of H1 causes no considerable changes in parasite growth in culture

To address the function of H1 in *T. brucei*, we generated RNAi cell-lines that allowed simultaneous inducible depletion of all classes of H1. The RNAi cell-lines were created by transfecting the PL1S cell-line with an H1 RNAi construct that contained a MAKASA sequence (H1.5 CDS and 3'UTR) under the control of two tetracycline-inducible opposing T7 RNA polymerase promoters that would generate double-stranded RNA (dsRNA) of the MAKASA sequence in the presence of tetracycline. The construct was designed to integrate by single cross-over in the rDNA spacer region (Fig. 2A). PL1S cell-line expresses T7 RNA polymerase constitutively (Alibu *et al.*, 2005), has BES2 active (*VSG9*) and bears a reporter luciferase gene (*LUC*) at silent BES1 (*VSG2*) (Yang *et al.*, 2009) (Fig. 4A). Because H1 genes are 89% identical, we expected that small interfering RNAs (siRNAs) generated from a dsRNA of a MAKASA gene would successfully knock-down all classes of H1 transcripts. Indeed, after inducing RNAi in two independent clones (C1 and C2), the mRNA levels of each of the three H1 classes were diminished to 39% on average during 6 days indicating that the knock-down was effective for all H1 classes in both clones (Fig. 2C, $P < 0.001$; Fig. S4A, $P < 0.001$). No

significant differences in H1 depletion levels were observed between day 2, 4 and 6 of RNAi induction. A clear reduction in H1 protein level was also observed after RNAi induction in a cell-line where one of the MAKASA variants is HA-tagged (Fig. 2C). The non-induced clones also showed a reduction of H1 transcripts (a decrease on average of 30% of the parental cell-line PL1S, $P < 0.05$) (Fig. S4B), which is likely due to the previously described leaky expression of the RNAi cassette (Alsford *et al.*, 2005a). In induced clones we observed that depletion of H1 resulted in a minor growth defect *in vitro*. The doubling times were slightly longer in C1 (7:16 h) and C2 (7:44 h) compared with the parental cell-line PL1S (6:32 h) ($P < 0.01$ and $P < 0.001$, respectively) (Fig. 3A). This is consistent with previous observations (Povelones *et al.*, 2012).

Because proper expression of histones is necessary for a normal cell cycle progression (Marzluff, 2005), we decided to characterize the effect of histone H1 depletion in *T. brucei* cell cycle. DNA was stained with propidium iodide (PI) in ethanol-fixed cells and the percentage of single cells (Fig. S5A) in each cell cycle stage was measured by fluorescence activated cell sorting (FACS) according to its DNA content (Fig. 3B). We did not observe any significant perturbations in cell cycle phases or in populations with abnormal ploidy (< 2C or > 4C) in both clone C1 and C2 during 6 days of RNAi induction (Fig. 3C). Similar results were obtained in live parasites stained with Vybrant® DyeCycle™ Violet (DCV) vital DNA dye (Fig. S5B). Given that there were no substantial cell cycle alterations, the small reduction in growth could be due to increased cell death induced by H1 depletion. We quantified dead cells by FACS in non-fixed parasites stained with PI. Because an intact cell membrane is not permeable to PI, only cells were cell membrane is compromised will be stained, that is, dead or dying cells. In PL1S parental cell-line there are always ~0.5% of dead cells in the population. Equivalent frequencies of dead cells were observed for clone C1. However, parasite populations from clone C2 have ~1.5% dead cells throughout the 6 days of RNAi, which probably explains why this clone has a longer doubling time. Although these mutants are not a complete knockout of the histone H1 gene family, our data suggest that H1 is not essential for growth of *T. brucei* in culture.

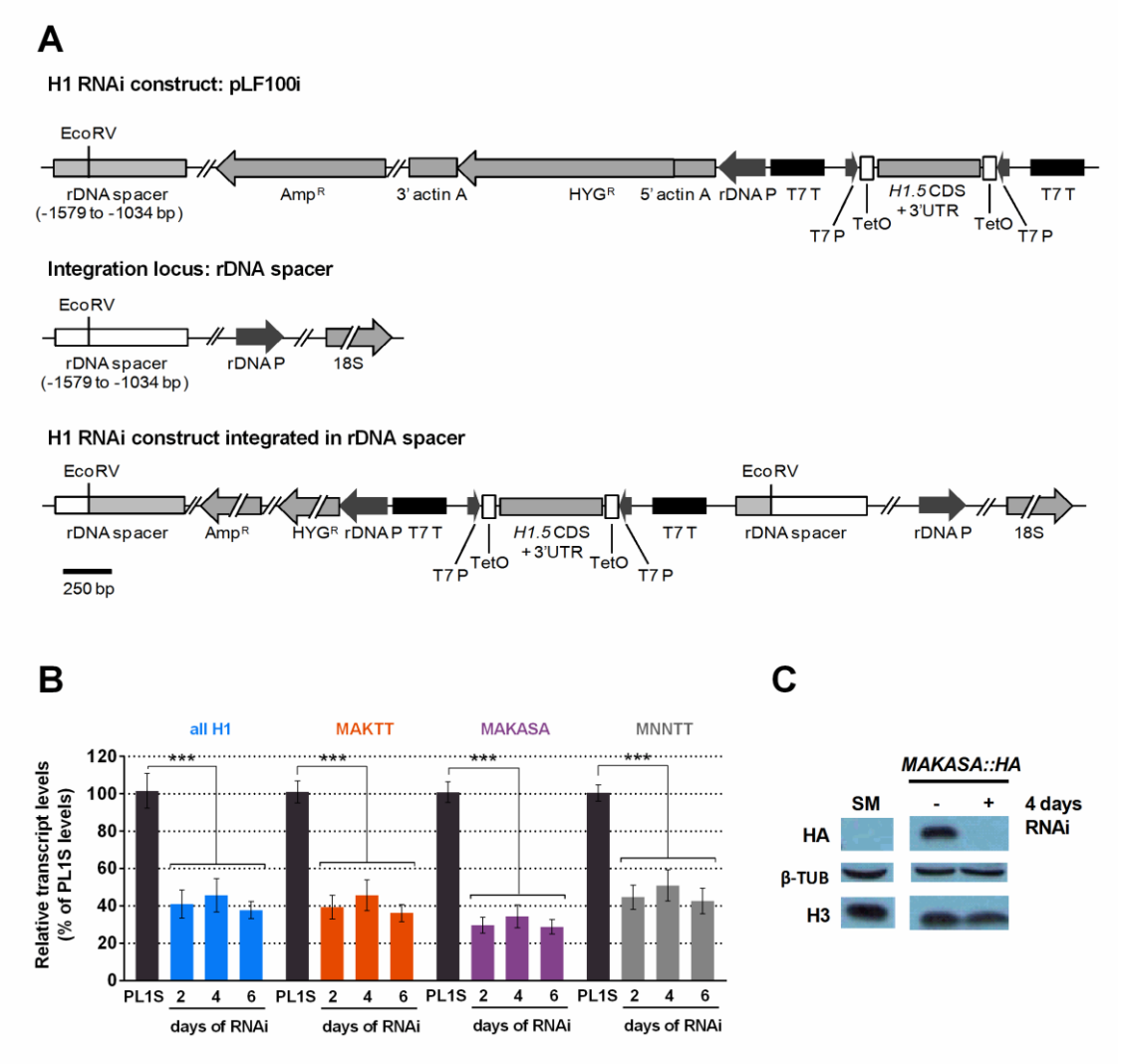


Figure 2. H1 RNAi depletes ~60% of the transcripts from all histone H1 classes. **(A)** Scheme showing the H1 RNAi construct, pLF100i, used to generate the transgenic H1 RNAi cell-line used in this study. Below is shown the integration region in the genome, the rDNA spacer, and pLF100i after integration. AmpR, ampicilin resistance gene; EcoRV, EcoRV restriction site; HYG^R, hygromycin resistance gene; rDNA P, rDNA promoter; T7 T, T7 terminator; T7 P, T7 promoter; TetO, tetracycline operator; 18S, 18S rDNA gene; 3' actin A, 3'UTR of actin A gene; 5' actin A, 5'UTR of actin A gene. **(B)** Efficiency of transcript depletion for all H1 classes and for each H1 class during RNAi induction for 2, 4 and 6 days. Transcript levels were measured by qPCR and are shown as an average of the two histone H1 RNAi clones relative to the parental cell-line PL1S. Three independent experiments were analyzed. Results are shown as mean ± SEM. Statistical significance was determined by non-parametric Mann-Whitney U tests. * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$ **(C)** Depletion of H1 MAKASA protein after four days of RNAi. Cell extracts were prepared after 4 days of RNAi for the MAKASA::HA cell-line LMF52.1. This cell-line was generated in the single-marker (SM) background, which was used as a negative control for HA detection. Total cell extracts were subjected to Western blotting with an anti-HA antibody to detect histone H1 MAKASA (HA). The levels of β-tubulin (β-TUB) and histone H3 proteins were detected as loading controls. Exposition time: 10 sec.

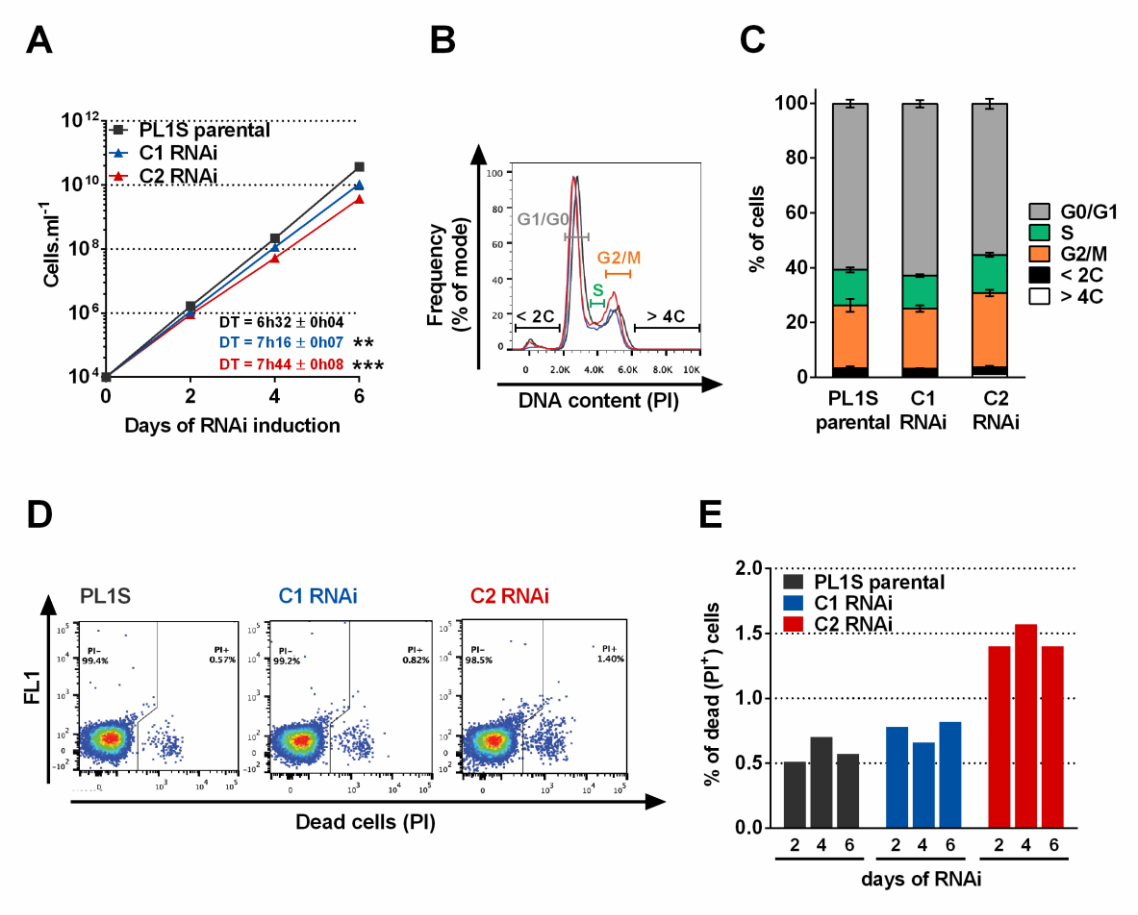


Figure 3. H1 depletion causes no major growth defect *in vitro*. (A) Cumulative growth in culture of BSFs of T. brucei upon H1 RNAi depletion during 2, 4 and 6 days. Mean doubling times (DT) are shown for each cell-line. Five to nine independent experiments were analyzed. Results are shown as mean ± SEM. Statistical significance was determined by a non-parametric Mann-Whitney U test. * P < 0.05, ** P < 0.01, ***P < 0.001 **(B)** Overlay of FACS histograms of cell cycle profiles of PL1S (grey), C1 RNAi (blue) and C2 RNAi (red) after 4 days of RNAi. DNA content was measured by propidium iodide (PI) staining in ethanol-fixed parasites. Frequency of parasites is normalized to mode (most frequent value of PI intensity). A representation of the gates used to quantify the proportion of each cell cycle stage is shown. < 2C, DNA content lower than the diploid chromosome number; > 4C, DNA content higher than the tetraploid chromosome number. **(C)** Depletion of H1 causes no significant changes in cell cycle. Proportion of *T. brucei* parasites in each cell cycle stage, determined by FACS as represented in (B). Measurements at day 2, 4 and 6 of RNAi were pooled from one independent experiment. Results are shown as mean ± SEM. **(D)** H1 depletion seems to induce *T. brucei* cell death. FACS dot plots showing PI intensity in the x-axis and FL1 intensity in the y-axis at day 6 of RNAi. Dead parasites were measured by PI staining in live parasites. The percentage of PI- and PI+ parasites are shown in each plot. **(E)** Percentage of dead, PI+ parasites, at 2, 4 and 6 days of RNAi. One independent experiment was analyzed.

3.1.2.3 H1 compacts chromatin at different levels across the genome

It has been previously demonstrated, using *in vitro* systems, that *T. brucei* H1 can condense chromatin (Burri *et al.*, 1994, Burri *et al.*, 1995). More recently,

Povelones *et al*. used electron-microscopy analysis and micrococcal sensitivity assays to show that H1 plays a role in heterochromatin formation *in vivo*. Here we investigated the role of H1 in chromatin condensation using an alternative approach: FAIRE (formaldehyde-assisted isolation of regulatory elements) of H1-depleted clones subjected to RNAi induction for 6 days. By performing a phenol-chloroform extraction on a cross-linked and sheared chromatin sample, FAIRE fractionates DNA that is preferentially less tightly associated to proteins (such as histones) (Giresi *et al.*, 2007). Quantitative PCR was used to quantify such enrichment and a plasmid spike with an ampicillin-resistance gene (Amp^R) was used as a normalizer for DNA input. Strikingly, silent BES promoter regions (97 bp downstream of transcription start site) were among the loci in which chromatin opened the most (Fig. 4). In both H1-depleted clones (C1 and C2), chromatin of the silent BES promoter regions opened on average 10-fold compared to the parental cell-line PL1S ($P < 0.001$) (Fig. 4B; Fig. S6A). The chromatin also opened at the luciferase gene (*LUC*), which is 1.2 kb downstream of a silent BES promoter, although in a slightly smaller degree (sevenfold relatively to PL1S; $P < 0.001$). In most other loci, loss of H1 resulted in a 1.8- to 4-fold increase in FAIRE enrichment (Fig. 4B), suggesting that chromatin becomes globally more accessible. These include silent *VSG*s from bloodstream and metacyclic expression sites (*VSG2*, *VSG3*, *VSG18*, *MVSG*; $P < 0.01$), the procyclin promoter region and procyclin *EP2* gene, typically transcribed in procyclic stages ($P < 0.01$), transcription start sites of Pol II polycistronic units containing either β-tubulin (*β-tub*-PR; $P < 0.01$) or structural maintenance of chromosome 3 gene (*SMC3*-PR; $P < 0.01$) and a number of Pol II-transcribed genes (very abundant *β-tub* and at lower levels *ISP*, *SMC3* and *PAG3*). Three loci showed a non-significant tendency for a more open chromatin conformation: the BES2 actively transcribed genes (blasticidin resistance gene, BSD^R, and *VSG9*) and the *18S* rDNA, all of which had been previously shown to have a very open, perhaps close to maximal, chromatin conformation (Figueiredo & Cross, 2010, Stanne & Rudenko, 2010).

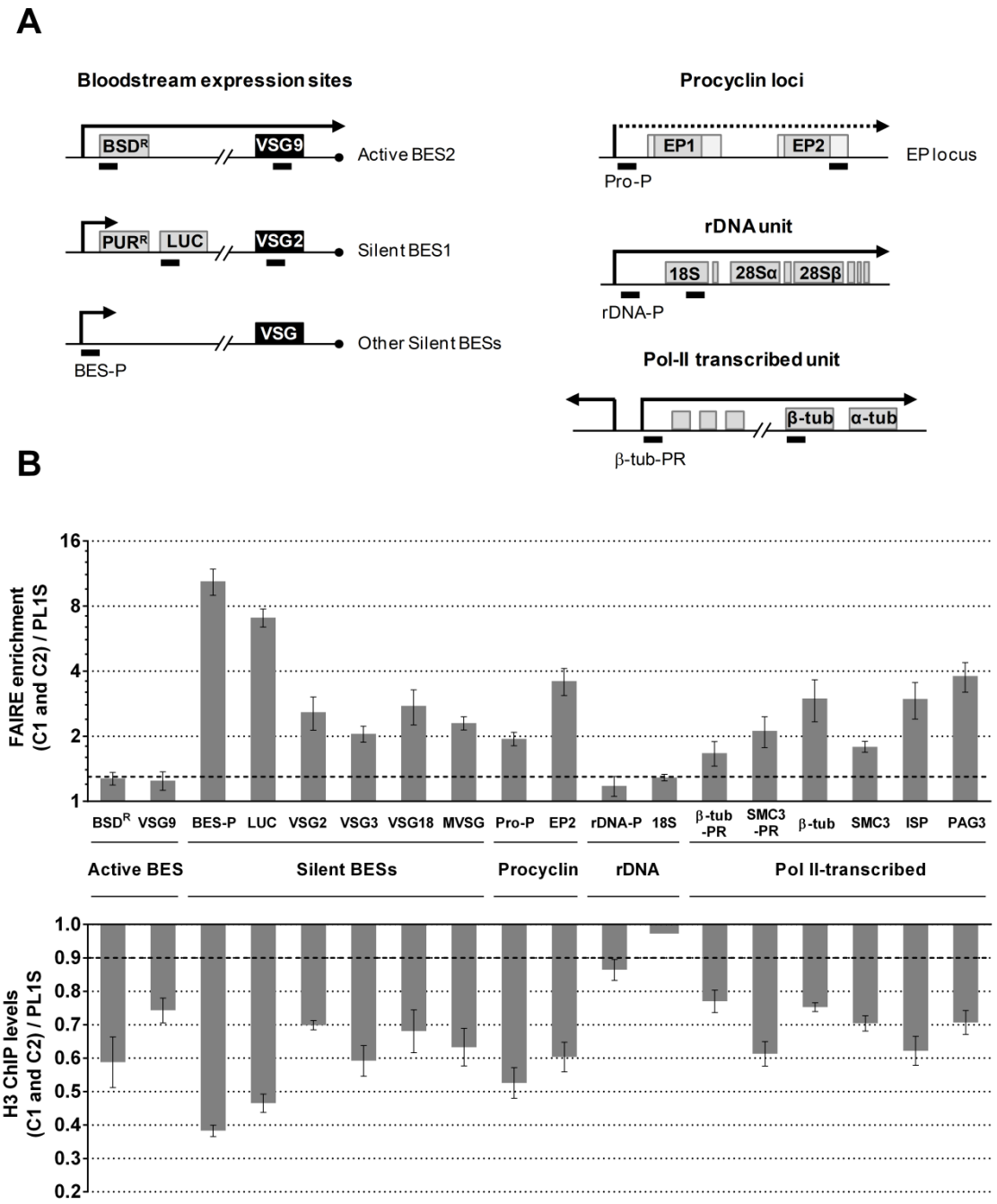


Figure 4. H1 compacts chromatin at different levels across genome. (A) Diagram indicates the amplicons (bars) amplified by qPCR for loci in BESs, procyclin loci, rDNA loci and a Pol II-transcribed polycistronic unit. Procyclin loci are partially transcribed in BSFs, as represented by a dashed arrow. Primers for EP2 procyclin are located in the 3'UTR (light grey). Chromatin opening was measured by FAIRE **(B)** and nucleosome occupancy was determined by histone H3 chromatin immunoprecipitation (ChIP) **(C)** at several gene loci in parental cell-line PL1S and histone H1 RNAi clones 6 days after induction. DNA isolated by FAIRE was quantified by qPCR and an enrichment corresponds to a more open chromatin. An AmpR gene contained in a plasmid spike was used as a normalizer for DNA input. For H3 ChIP, DNA was quantified by qPCR, compared to total input material and normalized to 18S rDNA. Results are expressed as fold-change relatively to parental cell-line PL1S. Note that FAIRE enrichment is represented in a logarithmic scale. Data for individual clones prior to normalization to the parental PL1S are shown in Fig. S5. Statistical significance was calculated using an empirical Bayes approach. Significant fold-change ($P < 0.05$) is shown above (B) and below (C) dashed line in FAIRE and ChIP plots respectively.

Figure 4. (continued from previous page) Two to three independent experiments were analyzed. Results are shown in mean ± SEM. BSD^R, blasticidin resistance gene; BES-P, promoter region of silent BESs; β-TUB, β-tubulin gene; β-TUB-PR, promoter region of β-TUB polycistronic unit; EP2, procyclin EP2 gene; ISP, inhibitor of serine peptidase gene; LUC, luciferase gene; MVSG, metacyclic VSG gene; Pro-P, procyclin promoter region; PURR, puromycin resistance gene; PAG3, procyclin associated gene 3; rDNA-P, rDNA promoter region; 18S, ribosomal 18S gene; SMC3, structural maintenance of chromosome 3 gene; SMC3-PR, promoter region of SMC3 polycistronic unit; VSG, variant surface glycoprotein gene.

The FAIRE results were further confirmed by chromatin immunoprecipitation (ChIP) of histone H3 (as a read out of nucleosome density). As previously shown (Figueiredo & Cross, 2010), FAIRE and H3 ChIP are highly consistent techniques and, globally, they reflect almost a mirror image from each other: for most loci, when chromatin becomes more open, we detect an increase in FAIRE enrichment and a decrease in H3 ChIP for both clones (Fig. 4B–C and Fig. S6). A comparison between the *P*-values of FAIRE and H3 ChIP experiments further confirms the consistency of the two techniques (Table S1). We observe, by ChIP, that most genes lost around 30% of nucleosomes. As expected, ChIP revealed that silent BES promoters are the loci that, upon loss of histone H1, lost more histone H3 (close to 60%). Interestingly, although FAIRE did not detect a very dramatic change in chromatin condensation of the procyclin promoter region, H3 ChIP detected a reduction of nucleosomes more pronounced than average (48%). Also, FAIRE did not detect a significant change in chromatin condensation of genes in active BES, but ChIP could detect a significant but modest loss of histone H3 at the promoter region and at the active *VSG* gene. As expected, a gene encoded by the mitochondrial genome (cytochrome *c* oxidase subunit III, *COIII*, whose genome is not organized around nucleosomes), showed only background levels of histone H3 (Fig. S6B, $P > 0.05$). Overall, these results confirmed that when H1 is depleted there is a global loss of histone H3 across multiple loci of the genome, except for silent BES promoters, which opened 10-fold more (60% loss of histone H3) and procyclin promoters to a lesser extent. Moreover, our data greatly strengthen FAIRE as a robust method to study global changes in chromatin condensation.

3.1.2.4 H1 regulates expression of Pol I-transcribed genes

Given the global role of H1 in chromatin condensation, we next tested how changes in chromatin structure affected expression genome-wide. We used RNA-Seq to compare the expression profile of H1-depleted clones (at day 4 of RNAi) and parental cell-line PL1S (Fig. 5, Fig. S7). H1 depletion resulted in significant changes (false discovery rate (FDR) adjusted *P*-value < 0.05) in the expression of 26 out of 8996 expressed genes (Fig. 5A). Interestingly, all were upregulated and all but one gene are transcribed by Pol I (Fig. 5A–C).

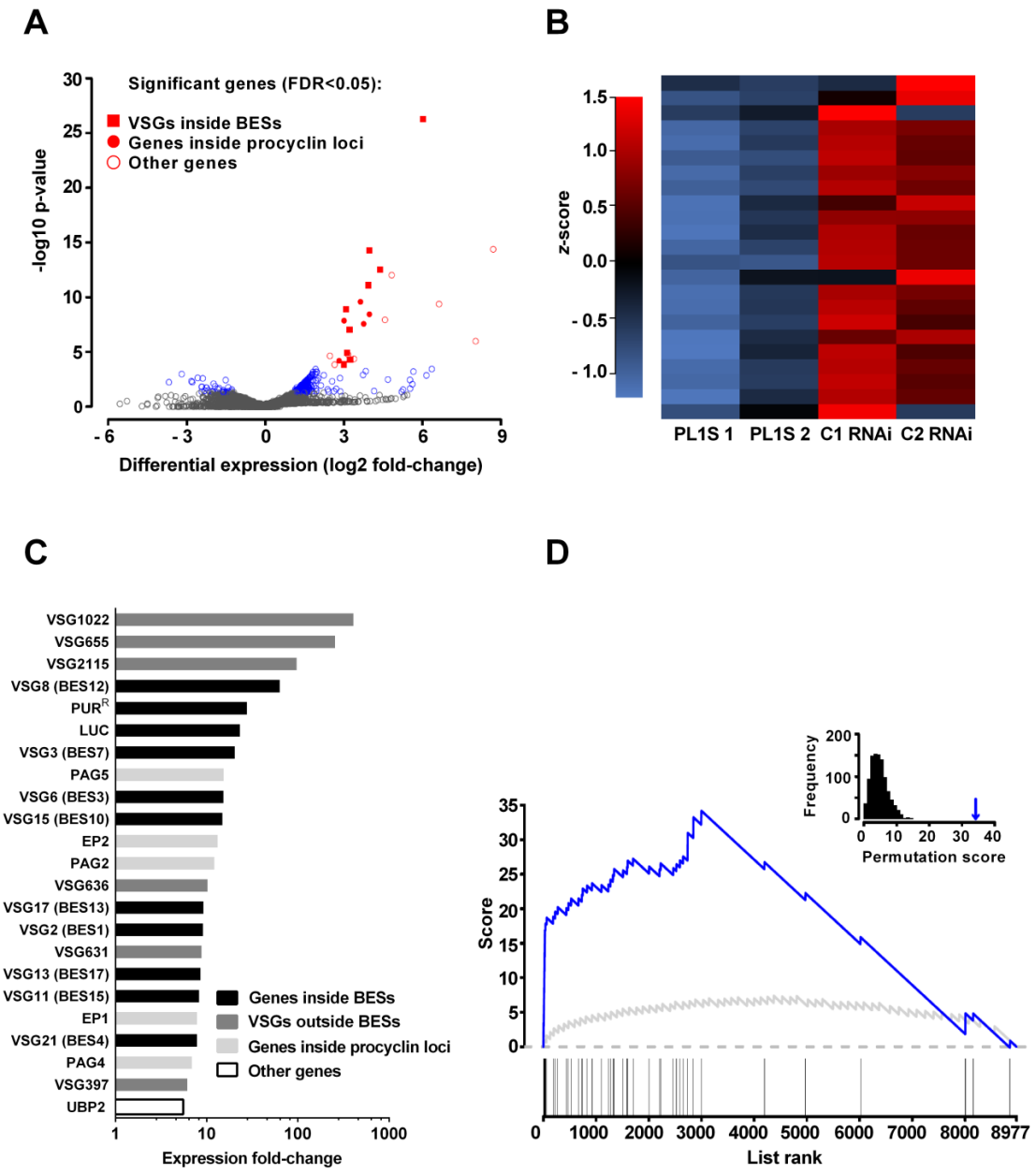


Figure 5. (previous page) H1 inhibits expression of RNA Pol I-transcribed genes. Genes with altered expression upon H1 depletion (4 days of RNAi) were identified by genome-wide RNA sequencing in BSFs of *T. brucei*. **(A)** Volcano plot analysis of differentially expressed genes between RNAi clones C1 and C2 relative to two biological replicates of the parental cell-line PL1S shows that H1 regulates a small set of genes, mainly located in procyclin loci (red filled circles) and bloodstream expression sites (BESs) (red filled squares). Each point represents a gene, with log2 fold-change (RNAi/PL1S) of gene expression across samples plotted on the x-axis and the corresponding statistical significance (−log10 P-value) plotted on the y-axis. Genes statistically significant (FDR-adjusted P < 0.05) are indicated in red; in grey are non-significant genes; in blue are those genes which would be significant if a less conservative statistical method had been used (FDR-adjusted P > 0.05 and P < 0.05). **(B)** Heatmap showing the z-scores calculated for the normalized DESeq expression levels of differentially expressed genes (FDR-adjusted P < 0.05) obtained by genome-wide RNA sequencing in BSFs of T. brucei. Genes are ordered by decreasing fold-change of expression, as in panel C. **(C)** Bar-plot showing the fold-change of expression of annotated differentially expressed genes in RNAi clones C1 and C2 relative to the parental cell-line PL1S. For VSG genes that are inside a BES, the respective BES number in shown in brackets. **(D)** Gene set enrichment analysis for genes located in BESs and procyclin loci. Blue line indicates the enrichment score distribution across genes ranked by decreasing statistical evidence of differential expression. Grey line represents 95th percentile of the enrichment scores obtained by 1000 random permutations of the gene ranks. The histogram at the top-right corner of the figure represents the distribution of maximum scores obtained by random permutation, with the arrow indicating the experimental enrichment score (i.e. the 'peak' in the main plot). The diagram at the bottom shows where the members of this gene set (i.e. genes located in BESs and procyclin loci) appear in the ranked list of genes. An FDR-adjusted P < 0.001 was obtained for this gene set, as described in Experimental procedures. Genes with undetermined fold-change due to too low unique read counts were excluded from the analysis.

Most of these genes (18) are BES-associated genes (*VSG*s and *ESAGs*) or genes located in procyclin loci (*EP* procyclin genes and *PAG*s), which are normally silent in this stage of the parasite life cycle. In H1-depleted mutants, BES-associated *VSG*s are on average 18-fold more expressed, with a maximal derepression within this subset close to 65-fold (*VSG8*) (Fig. 5C). Other eight *VSG*s annotated as outside a BES were also upregulated. Because we do not know the genomic location of these *VSG*s, we cannot evaluate whether there is a nearby promoter or polycistronic unit that could read through these *VSG*s. The transcript levels of 592 expressed *VSG*s did not change significantly, which is not surprising because in general *VSG* genes are located in tandem arrays that lack a functional promoter (Marcello & Barry, 2007). Most *VSG*s are not expressed and thus were not included in our analysis. H1 depletion also lead to an increased expression of the puromycin resistance (PUR^R) and the luciferase genes (Fig. 5C), both located immediately downstream of a silent BES promoter (BES1) (Fig. 3A), thus indicating that derepression occurred throughout the entire BES locus. However, as shown by the luciferase assay (Fig 6C), the derepression of silent BESs, is still far

from the full activation seen in active BES, indicating that silent BESs are only partially derepressed. Only one of the *ESAG*s was found differentially expressed (*ESAG2* from BES12) most probably because the repetitive nature of *ESAG*s prevented unambiguous read alignment and were thus not included in the analysis. Due to too low unique read counts, the fold change of *ESAG2* remained undetermined. Although *PAG*s are located in multiple copies in the genome, in Pol II transcription units (Kim *et al.*, 2013) or within the procyclin loci (Haenni *et al.*, 2006), the differentially expressed members of *PAG2*, *PAG4* and *PAG5* are specifically located in these latter loci. For example, although there are two *PAG2* genes, one in the procyclin locus and the other in a polycistronic unit on the same chromosome 10, the only gene that is significantly upregulated when H1 is depleted is the one within the procyclin locus. The results above strongly suggest that H1 predominantly inhibits expression of Pol I genes.

We used gene set enrichment analysis (GSEA) (Subramanian *et al.*, 2005) to investigate if specific gene sets are enriched or depleted among differentially expressed genes. We found that genes located in BESs and procyclin loci are very significantly enriched in our list of differentially expressed genes (FDR-adjusted $P < 0.001$), as illustrated by the large difference between the detected enrichment score distribution and the 95th percentile of the enrichment scores obtained by random permutations (blue and grey lines, respectively, in Fig. 5D) This enrichment is also observed if the two sets of genes are analyzed independently (Fig. S8). This analysis further indicates that H1 plays an important role in the expression of Pol I-dependent genes. RNA-Seq results were confirmed at the RNA level by qPCR (Fig. 6A–B; Fig. S9) and at the protein level by luciferase activity assay (Fig. 6C). First we checked the correlation between qPCR and RNA-Seq fold-changes of gene expression of 34 genes selected so that we would cover a wide range of expression levels. We observed a very high Pearson's correlation ($r = 0.90$) between the two types of experimental data, suggesting that qPCR and RNA-Seq are highly consistent (Fig. 6A). To further confirm the observation that Pol I, but not Pol II genes become derepressed when H1 is depleted, we measured by qPCR the transcript levels in the two RNAi clones induced for 2, 4 and 6 days (Fig. 6B, Fig. S10). These were normalized to *18S* transcript levels because these are highly expressed genes, depleted of H1 (Povelones *et al.*, 2012) and

with chromatin largely insensitive to H1 depletion (Fig. 4B). It should be noted that rRNA genes could not be included in our RNA-Seq analysis due to their repetitive nature.We confirmed that H1 depletion did not affect any of the tested Pol II transcribed genes nor the actively transcribed *VSG9*. As detected by RNA-Seq, *VSG*s in silent BESs (*VSG2* and *VSG3*) were significantly derepressed ($P < 0.001$) (Fig. 6B), although the level of derepression was slightly smaller by qPCR (5- to 12-fold) than RNA-Seq (9- to 21-fold). We also confirmed that procyclin *EP2* was fivefold derepressed. Another *VSG* in a silent BES (*VSG18*) and a *VSG* from a metacyclic expression site (*MVSG*; Tb427-639 (Cross *et al.*, 2014)) were found to be four- to fivefold derepressed by qPCR ($P < 0.001$) but were not detected as being differentially expressed in the RNA-Seq data. This inconsistency might be due to the fact that silent *VSG*s and *MVSG*s are expressed at a very low level in bloodstream forms (Donelson, 2003) and for this type of genes, statistical tests for differential expression based on RNA-Seq data are not very powerful. Finally, BES derepression was further confirmed at the protein level by measuring the activity of luciferase in the entire population. We observed an increase in luciferase activity of 7- to 10-fold ($P < 0.001$; Fig. 6C), thus confirming RNA-Seq and qPCR data that silent BES1 was derepressed ($P < 0.001$; Fig. 6B). Overall, qPCR and luciferase activity assays confirmed/validated the RNASeq data. We conclude that, like in genome-wide studies in mammalian cells and yeast (Hellauer *et al.*, 2001, Fan *et al.*, 2005), *T. brucei* H1 is not a global transcription repressor. It rather regulates a particular subset of genes, by repressing expression of Pol I-dependent genes. It is clear that the most significant subgroup of genes that is derepressed are genes from BESs or procyclin loci, confirming that H1 is essential for keeping these specific loci silent.

To examine whether the upregulation of silent *VSG* transcripts triggered by H1 depletion would induce loss of monoallelic expression of a single type of VSG protein at the parasite surface, we analyzed the expression of two silent VSGs, VSG2 and VSG13 at the cell surface by FACS (Fig.7). Live parasites of both C1 and C2 RNAi clones were stained with specific antibodies against VSG2 and VSG13 at day 2, 4 and 6 of RNAi induction. Dead parasites were excluded from the analysis by gating on the negative population for PI.

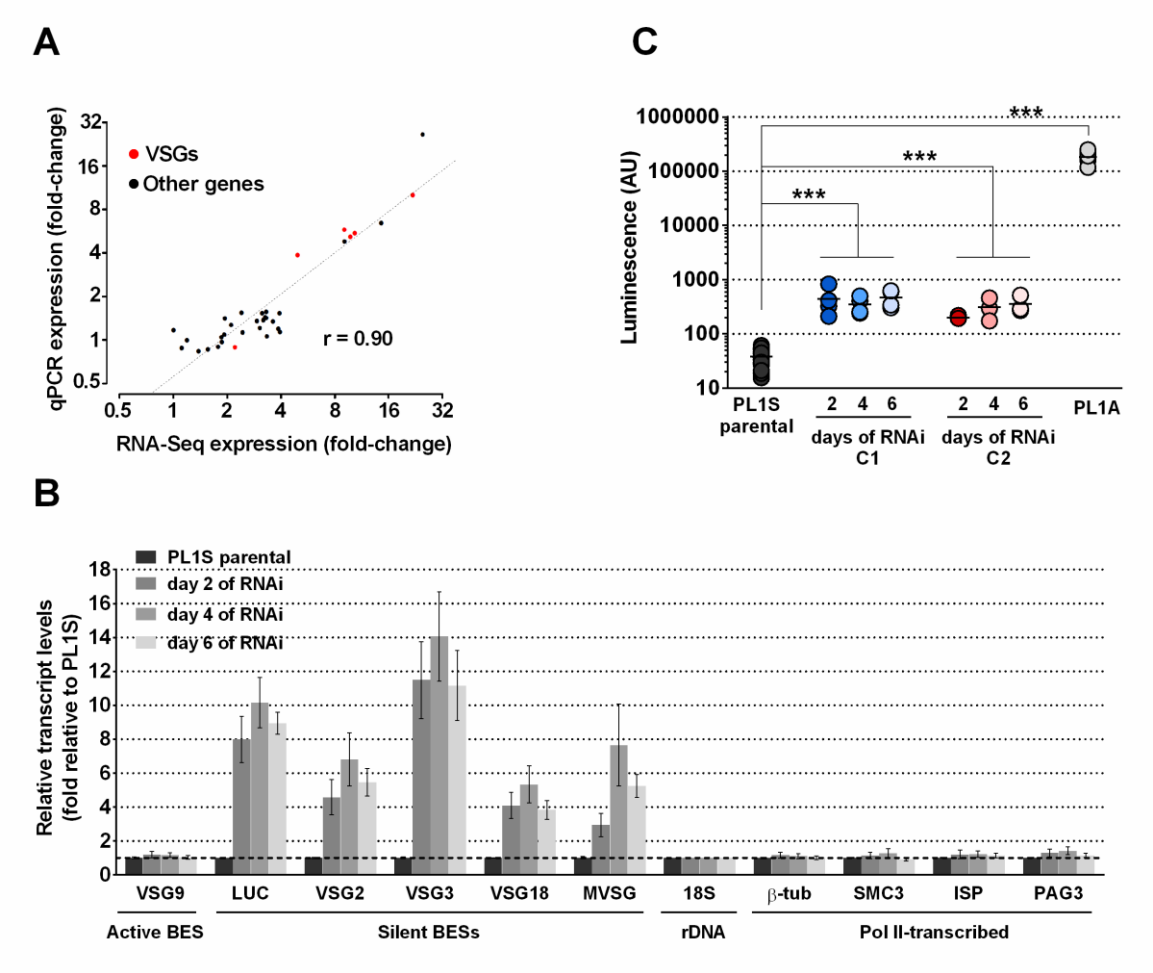


Figure 6. RNA-Seq data are validated by qPCR and luciferase assays. (A) Scatter plot of fold-changes of gene expression in RNAi clones C1 and C2 relative to the parental cell-line PL1S. Thirty-four genes with mean expression levels ranging from 12 to 512,586 RPKMs by RNA-Seq were randomly selected and quantified by qPCR. Pearson's correlation coefficients (r) are indicated. **(B)** Quantification of mRNA expression after 2, 4 and 6 days of RNAi induction. Transcript levels were measured by qPCR. Transcript levels are plotted as an average of the fold-change of RNAi clones C1 and C2 relatively to the parental cell-line PL1S. Significant ($P < 0.05$) fold-change of all RNAi days is shown above dashed line. Three independent experiments were analyzed. Results are shown as mean ± SEM. **(C)** Derepression of silent BESs was assessed by quantification of luminescence throughout 2, 4 and 6 days of RNAi induction. A luciferase reporter gene located downstream of the promoter of silent BES1 (see diagram Fig. 4A) allows the monitoring of its transcriptional activity by measuring the luminescence emitted (AU, arbitrary units). Horizontal line indicates average of luminescence levels for each experimental group. PL1A refers to a cell-line that is isogenic of PL1S, but in which BES1 is active and therefore luciferase gene is transcribed at maximal levels. Three to four independent experiments were analyzed. Statistical significance was calculated by non-parametric Mann-Whitney U tests.* $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$.

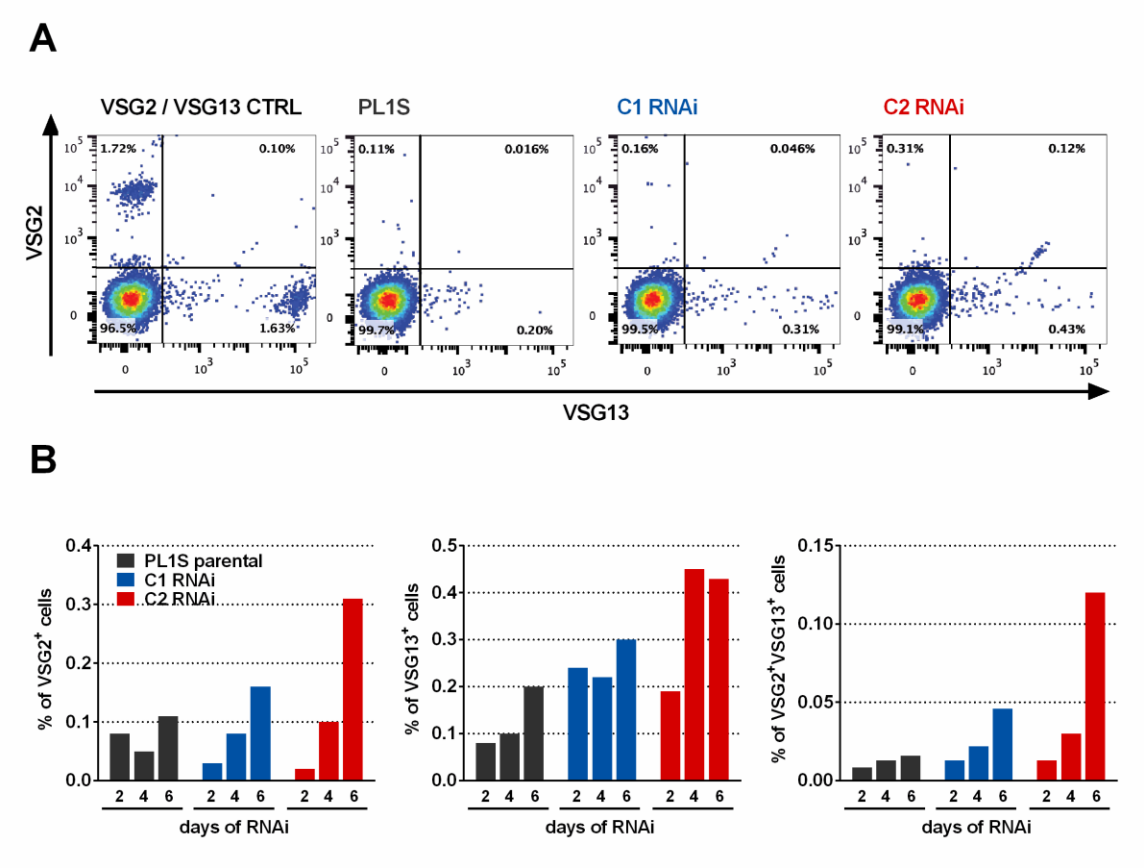


Figure 7. H1 depletion leads to the expression of more than one silent VSG at the surface. **(A)** FACS dot plots showing VSG2 and VSG13 expression at the cell surface after 6 days of RNAi. Live parasites were stained with anti-VSG2, anti-VSG13 and propidium iodide (PI). Cell populations were gated in the live, PI- cells as shown in Figure 3D. The first plot depicts a positive control were C1 RNAi parasites were mixed with ~1% of a VSG2-expressing parasite cell-line + ~1% of a VSG13-expressing cell-line. The four gates and respective percentages of cells per each are shown. **(B)** Percentages of VSG2+VSG13-, VSG2-VSG13+ and VSG2+VSG13+ populations after day 2, 4 and 6 of RNAi. Percentages were quantified by FACS analysis using the indicated gates, as represented in (A). One independent experiment was performed.

The percentage of both VSG2 and VSG13 expressing parasites (VSG2$^+$ and VSG13$^+$) is maintained low but increases throughout RNAi induction up to 2- to 3-fold (0.3-0.4% of total parasites) when compared with PL1S parental parasites (Fig. 7B). This indicates that upon H1 depletion the increase in mRNA levels of silent VSGs induced by H1 depletion leads to the abnormally high levels of parasites with a silent VSG at the surface. Importantly, the percentage of double-expressors (VSG2$^+$ and VSG13$^+$) also augments, particularly at day 6 of RNAi (10-fold compared with PL1S), indicating that part of the silent VSG expressors display more than one silent VSG at the surface (Fig. 7A, B). When compared with PL1S, the increase in parasites expressing

silent VSGs at the surface seems to come from parasites in which expression levels of silent VSGs are as high as in parasites where those VSGs are active (Fig. 7A). This suggest that depletion of histone H1 leads to an increase in VSG switching rates, as was reported before (Povelones *et al.*, 2012) and reinforces the implications of histone H1 for antigenic variation. Nevertheless, this experiment was only performed once. Therefore it should be repeated in order to verify the reproducibility of our observations and confirm our conclusions.

3.1.2.5 H1 inhibits transcription from silent BESs and procyclin sites

We next sought to understand how H1 depletion leads to an increase of steady-state mRNA levels of the silent *VSG*s and procyclins. Because H1 depletion leads to a more relaxed chromatin at Pol I promoter regions (especially at the silent BES promoters), we hypothesized that the increase in mRNA levels resulted from an increase in transcription rates at these loci. To test this hypothesis we used metabolic labeling of nascent transcripts with 4-thiouridine (4sU) (see Results 3.3.). This method allows the quantification of newly synthesized RNA in live cells, therefore the direct measurement of RNA transcription rates with minimal interference on gene expression and cell viability (Dolken *et al.*, 2008). After incubating *T. brucei* BSF parasites with 4sU, total RNA was extracted, thiol-specific biotinylated and subsequently purified on streptavidin-coated magnetic beads. qPCR was then used to quantify the newly transcribed RNAs purified. We analyzed the kinetics of 4sU incorporation into RNA in *T. brucei* during 2, 5, and 10 min of labeling. Since newly transcribed RNA contains higher amounts of unprocessed, primary transcripts, we determined the levels of nascent transcripts by quantifying the ratio of intergenic relative to coding sequence expression for *18S* and β-tubulin transcripts (Fig. 8A). As expected, we observed a rapid increase in the levels of 4sU-tagged RNA with increased duration of labeling (Fig. 8B). Based on this result and on previous reports (Rabani *et al.*, 2011, Windhager *et al.*, 2012), we chose 10 min as an appropriate 4sU-labeling time to estimate transcription rates. We next compared the levels of nascent transcripts of individual genes in a metabolically labeled sample relative to the total RNA. Although the yield of purification was generally low (0.015–0.025%), for all measured loci we observed that

the labeled samples had 7- to 20-fold more nascent transcripts than non-labeled samples (Fig. 8C), indicating that the levels of labeled transcripts are significantly enriched over the background. To test if the 4sU-labeled RNAs are a product of transcription, we repeated the labeling experiment in cells that were previously treated for 5 min with actinomycin D (ActD), a well-known transcription inhibitor. We observed that, upon transcription inhibition, the levels of 4sU-labeled RNAs recovered by MACS were dramatically reduced to the same low levels detected in the NO 4sU samples (Fig. 8D). Altogether these results further confirmed that 4sU-labeled RNAs are a reflection of the transcription happening inside live parasites.

4sU metabolic labeling was next used to test whether H1 depletion results in a higher transcription rate from BESs and procyclin promoters. For this, RNAi was first induced for 2 days in clones C1 and C2 and, at the end of this period, RNAi clones and parental PL1S parasites were incubated with 4sU. After 10 min the 4sU-labeling reaction was stopped and total RNA was extracted. The relative abundance of intergenic transcripts for *18S* and β-tubulin genes showed that purification of newly transcribed RNA was equally efficient in all cell-lines (Fig. 8E). Analysis of newly transcribed RNA revealed that H1 depletion indeed increases the transcription rate from silent BES promoters, resulting in higher abundance of precursor transcripts from the BES promoter region *BES-P* ($P < 0.001$) and the downstream luciferase gene ($P < 0.001$). When H1 is depleted, transcription rate is also higher from procyclin promoters, since we detected significantly more nascent transcripts from the procyclin promoter region ($P < 0.01$) and at the EP2 procyclin gene ($P < 0.001$) (Fig. 8F). These results show that when H1 is depleted, there seems to be a higher transcription rate of silent BESs and procyclin loci, most likely due to an increase of transcription initiation from the promoter. A Pol II gene (β-tubulin) showed no change in transcription rate, confirming the qPCR analysis of steady state RNA levels (Fig. 6B). Curiously, the BES promoter regions are the loci in which the levels of nascent transcripts have increased the most (sixfold), which is consistent with the fact that the most dramatic changes in chromatin condensation observed upon H1 depletion were at the silent BES promoters. This may suggest a more significant role for H1 at silent BESs. We also observed a slight increase in transcription rate at the BSD^R gene located downstream

of the active BES promoter, consistent with the decrease in H3 detected by ChIP. Taken together, these results not only support our RNA-Seq data but provide additional mechanistic insight, showing that histone H1 is an effective regulator of transcription at the Pol I-transcribed BESs and procyclin loci.

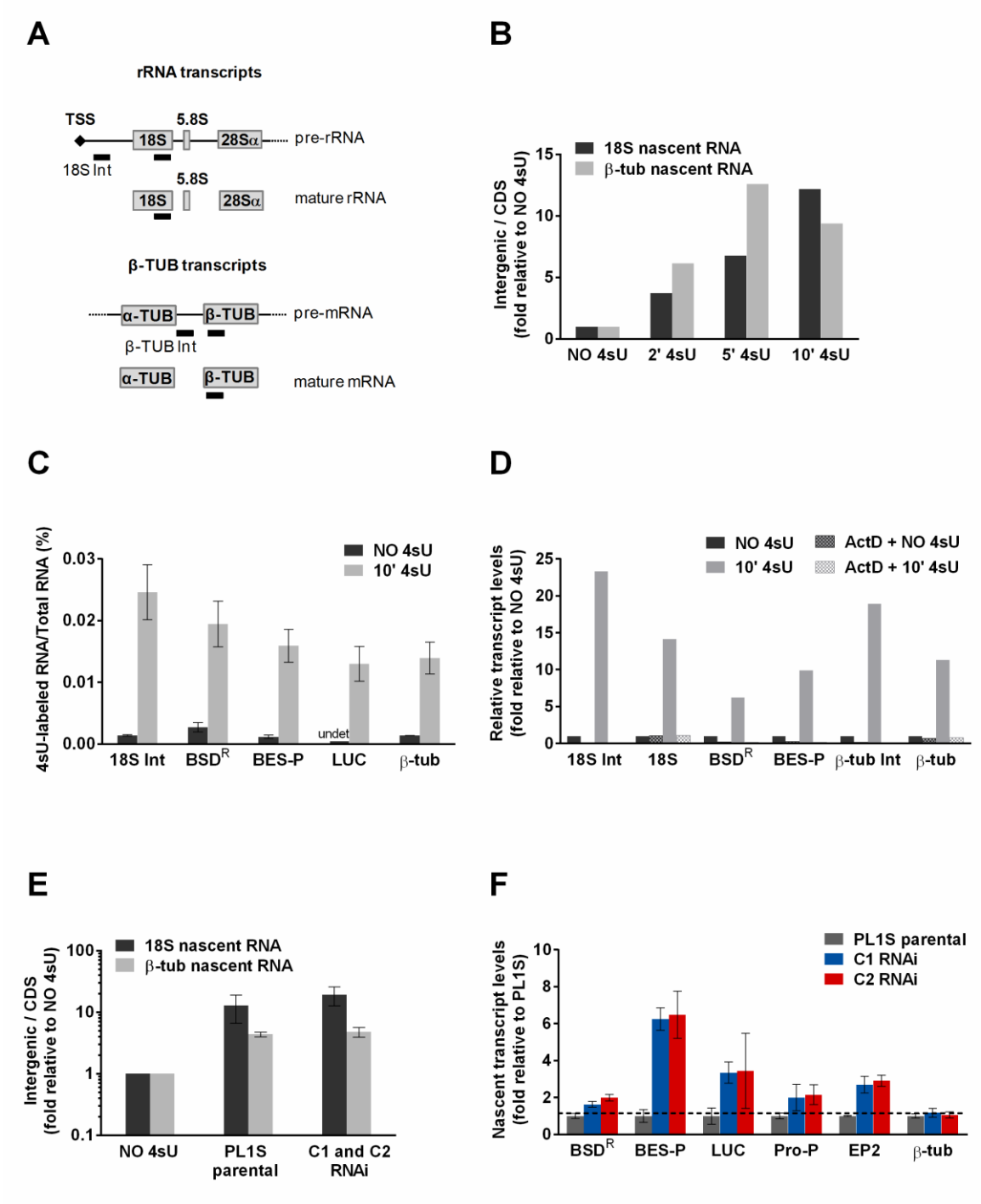


Figure 8. (previous page) H1 depletion results in increased transcription rate in silent BESs and procyclin loci. Nascent transcripts were labeled with 4sU for 10 min, biotinylated and streptavidin-purified. RNAs were quantified by qPCR, compared with non-labeled sample and normalized to 18S rRNA transcripts. **(A)** Diagram indicates the amplicons (bars) amplified by qPCR for precursor and mature rRNA and β-tubulin transcripts. **(B)** Kinetics of 4sU-labeling of newly transcribed RNA measured by the ratio of intergenic/CDS transcripts. **(C)** Percentage of 4sU-labeled RNA relative to total RNA for several loci with 10 min of 4sU-labeling. NO 4sU, n=2; 10' 4sU, n=7 **(D)** Relative transcript levels in 4sU-labeled RNA in the absence or presence of a 5 min treatment with actinomycin D (ActD), which inhibits transcription, prior to 4sU-labeling for 10 min. PL1S, C1 and C2, n=3-7 **(E)** Enrichment in nascent transcripts in PL1S and H1-depleted mutants with 10 min of 4sU-labeling. **(F)** Fold increase in nascent transcripts in H1-depleted mutants in comparison with PL1S parental cell-line. Results are shown as mean ± SEM. Two independent experiments were performed, n = 3–4 for each cell-line. Significant fold-change ($P < 0.05$) was calculated using an empirical Bayes approach and is shown above dashed line. Primers used to amplify 18S Int were the same used to amplify downstream promoter region of rDNA (rRNA-P). BES-P, promoter region of silent BESs; BSD^R, blasticidin resistance gene; β-tub, β-tubulin gene; β-tub Int, β-tubulin intergenic region; EP2, procyclin EP2 gene; LUC, luciferase gene; Pro-P, procyclin promoter region; TSS, transcription start site; undet, undetermined - transcript level lower than the qPCR limit of detection; 18S, ribosomal 18S gene; 18S Int, 18S intergenic region.

3.1.2.6 Depletion of H1 increases resistance to MMS-induced DNA damage

The fact that H1 is required for global chromatin compaction, but it has a very limited role in global transcription regulation, led us to hypothesize that H1 may play other functions in the cell. Povelones *et al.* have recently shown that H1 inhibits recombination of telomeric *VSG* genes, but not of a reporter gene *URA3* (Povelones *et al.*, 2012). In yeast, H1 suppresses HR thus inhibiting DNA repair and promoting genome stability (Downs *et al.*, 2003). To test if DNA repair is affected in *T. brucei*, we treated H1-depleted clones with a DNA damaging agent, methyl methanesulphonate (MMS), and measured the subsequent cell survival. MMS is an alkylating agent that stalls replication forks, a damage that can induce formation of DSBs and be repaired by HR (Lundin *et al.*, 2005, Nikolova *et al.*, 2010, Ma *et al.*, 2011).

Parasites in log-phase growth were exposed to MMS for 2 days, after which cell viability (metabolic function) was measured with Alamar Blue® (Raz *et al.*, 1997). RNAi induction started 2 days before MMS treatment and continued throughout MMS exposure. We observed that both H1-depleted clones were clearly more resistant to MMS-damage than the parental cell-line PL1S (Fig. 9A). Half maximal inhibitory

concentration (IC_{50}) determined from dose–response curve was on average 2.76 parts per million (ppm) for PL1S and significantly higher for C1 and C2, 4.45 and 4.59 ppm respectively ($P < 0.01$) (Fig. 9B). This result indicates that DNA repair is more effective when H1 levels are reduced. This could be because H1 limits access to DNA repair machinery and, as a result, suppresses HR. This is likely the case since HR-mediated repair dominates in *T. brucei* (Burton *et al.*, 2007, Glover *et al.*, 2008).

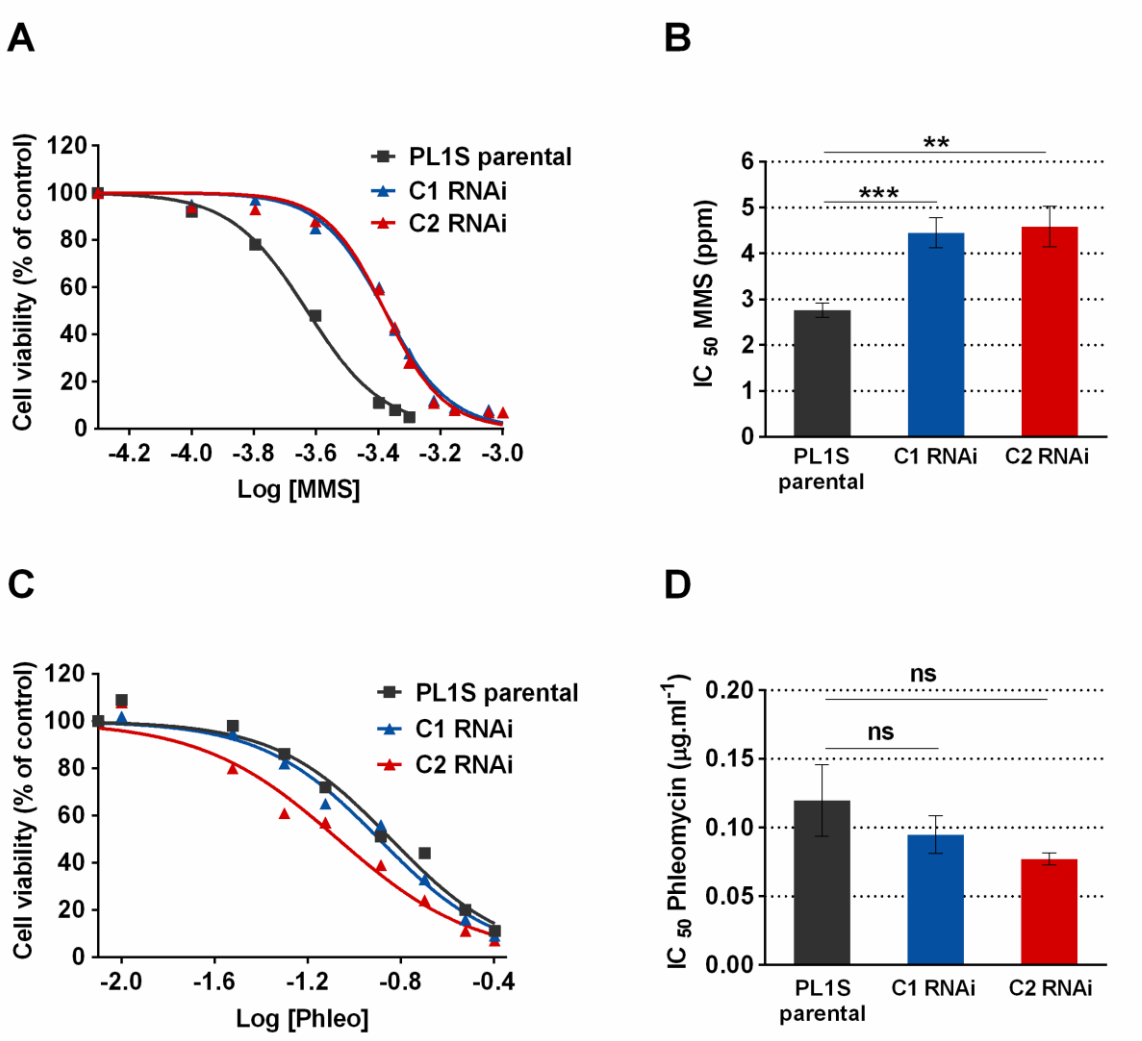


Figure 9. H1 depletion increases resistance to MMS-induced DNA damage. Dose–response curves of BSF of T. brucei to exposure to MMS **(A)** or phleomycin **(C)**, in parental and H1-depleted clones. Curves were obtained by fitting a non-linear regression with variable slope to the data. Cell viability was assessed by AlamarBlue® staining at the end of the treatment. IC50 values calculated from the corresponding regressions generated for MMS **(B)** and phleomycin **(D)** treatments. MMS concentration is expressed in parts per million (ppm) of volume and phleomycin is expressed in µg.ml-1. Seven to five independent experiments were analyzed for MMS treatment. Three to four independent experiments were analyzed for phleomycin treatment. Results are expressed in mean ± SEM. Statistical significance was determined by a non-parametric Mann-Whitney test. ns, non-significant P-value. * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$.

We also tested the sensitivity of H1-depleted parasites to another DNA damaging drug, phleomycin, which generates single-strand and double-strand DNA breaks (Giloni *et al.*, 1981). Sensitivity to phleomycin has been frequently used to study DNA repair in *T. brucei* (Proudfoot & McCulloch, 2006, Glover *et al.*, 2011). However, we did not observe significant changes in sensitivity to phleomycin and its respective IC_{50} values between PL1S parental cell-line and RNAi clones (Fig.9C-D). The distinct responses of H1-depleted parasites to MMS- and phelomycin-induced damage might be explained by the different DNA damage mechanisms and the extension of DNA lesions produced by the two drugs (Glover & Horn, 2012).

3.1.3 MATERIALS AND METHODS

CELL-LINES AND GROWTH MEDIUM

Trypanosoma brucei brucei bloodstream form (BSF) parasites (strain Lister 427, antigenic type MiTat 1.2, clone 221a) (Johnson & Cross, 1979) were cultured in HMI-11 as described in Hirumi and Hirumi (1989). Cells were diluted according to their doubling times (around 6-7 h) in order to maintain log-phase growth. Parasite growth was measured by counting cells in a Neubauer chamber.

PL1S cell-line was described in Yang *et al.* (2009). H1 RNAi cell-line clones (C1 and C2) were obtained by transfecting *T. brucei* PL1S cells with a H1 RNAi construct (pLF100i) (formerly generated by Luísa Figueiredo). To generate pLF100i, a MAKASA H1 sequence (H1.5) containing the CDS and the 3'UTR was PCR amplified from a cDNA library of *T. brucei* Lister 427 and inserted into a p2T7TA RNAi vector (Alibu *et al.*, 2005). In this vector the target sequence is cloned in-between two opposing T7 promoters that are inducible with tetracycline due to the presence of tetracycline operator sequences downstream of each one the promoters (Fig. 2A). H1 protein levels were measured in the cell-line LMF52.1, which has a single-marker background (Wirtz *et al.*, 1999), contains the H1 RNAi construct (pLF100i) and one H1 MAKASA gene endogenously HA-tagged at the C-terminal (MAKASA::HA) (formerly generated by

Luísa Figueiredo). RNAi was induced by adding 1 μg.ml^{-1} of tetracyclin (Fisher Scientific) to the medium for a maximum of 6 days.

RNA EXTRACTION AND REAL-TIME QUANTITATIVE PCR ANALYSIS

Parasites were harvested by centrifugation at 650 *g* for 10 min, 4°C and immediately resuspended in PureZOL (BioRad) or TRIzol (Invitrogen). RNA was isolated following the manufacturer's instructions and RNA quantity and quality was assessed on a NanoDrop 2000 (Thermo Scientific). The procedure used to isolate nascent RNA labelled with 4sU is described in a section below. cDNA was generated using a Superscript cDNA Synthesis Kit (Invitrogen), according to manufacturer's protocol. Quantitative PCR (qPCR) was performed using 1 × SYBR Green PCR Master Mix (Applied Biosystems). Negative controls lacking reverse transcriptase were confirmed by qPCR. Primer efficiencies were determined using standard curves with 3-logs coverage. Amplification reactions were performed in duplicates. Relative quantification was performed based on the CT (cycle threshold) value and the PCR efficiency-corrected method of Pfaffl (Pfaffl, 2001). DNA and transcript levels were normalized to those of 18S, which did not change significantly after histone H1 depletion. Primer sequences are listed in Table S2.

DETECTION OF H1 PROTEIN BY WESTERN BLOTTING

Depletion of H1 protein after H1 RNAi induction was detected using the MAKASA::HA LMF52.1 cell-line. Whole cell lysates were prepared by resuspending 1-2x10^{6} BSF parasites in 1x Laemmeli buffer at a concentration of 10^{5} cells.μl^{-1}, boiled for 10 min, loaded onto a Mini-Protean TGX pre-cast PAGE gels (BioRad) and separated by SDS-PAGE. Transfer to a nitrocellulose membrane was done with an iBlot Gel Transfer Device (Thermo Fischer Scientific) for 7 min, according to the manufacturer's protocol. Nitrocellulose membrane was blocked for 1 h at RT with PBS/0.1% Tween 20/5% milk, probed overnight at 4°C with polyclonal mouse anti-HA HA RU 1:5000 (Memorial Sloan-Kettering Cancer Center), monoclonal mouse anti-β-tubulin KMX1 1:1000 (kindly

given by Keith Gull) or polyclonal rabbit anti-H3 1:100 000 (custom-made in Pineda) as primary antibodies diluted in PBS/0.1% Tween20/3% milk. After washing the membrane 3x with PBS/0.1% Tween20 at RT, it was incubated with the secondary antibodies, sheep anti-mouse IgG-HRP 1:10 000 (GE Healthcare Life Sciences; NA931) or donkey anti-rabbit IgG-HRP 1:10 000 (GE Healthcare Life Sciences; NA934) diluted in PBS/0.1% Tween20/3% milk, for 1 h at RT. After washing 3x in washing buffer the blot was developed, protected from light, with Western Lightening ECL Plus (PerkinElmer) and light emission was detected in an X-ray film with 10 sec. exposure.

FLOW CYTOMETRY ANALYSIS OF CELL CYCLE

Fluorescence-activated cell sorting (FACS) analysis of cell cycle profiles with propidium iodide (PI) staining was done basically as described (Janzen *et al.*, 2006b). At day 2, 4, and 6 after H1 RNAi induction, $2x10^6$ of BSF parasites were collected from culture by centrifugation at 1300 *g*, 10 min 4°C, washed 1x with trypanosome dilution buffer (TDB) (5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20mM Na_2HPO_4, 2 mM NaH_2PO_4, 20 mM glucose, pH 7.7) (Cross, 1975), resuspended in 1 ml cold PBS/2mM EDTA, and fixed by adding drop-wise 2.5 ml cold 95% ethanol while gently vortexing. Cells were fixed and stored during 1-5 days at 4°C, after which they were washed 1x with 1 ml PBS/2mM EDTA at RT and resuspended in 1 ml of PBS/2mM EDTA plus 10 µg RNAse A (Carl Roth). Incubation proceeded at 37°C for 30 min. DNA was stained by adding 1 mg propidium iodide (PI) and incubated at RT for 30 min. Cells were kept on ice until analysis on a BD LSRFortessa (BD Biosciences) flow cytometer. Cell cycle analysis was completemented by staining live parasites with the vital dye Vybrant DyeCycle Violet Stain (Thermo Fischer Scientific). Briefly, $1x10^6$ BSF parasites grown in culture cells were resuspended in 1 ml 2% HMI-11 medium to which 0.5 µl of dye Vybrant DyeCycle Violet Stain was added per ml of cell suspension. Cells were incubated at 37°C for 10 min, protected from light and preceded to analysis in a BD LSRFortessa (BD Biosciences) flow cytometer. Data was analyzed on FlowJo v.X software.

FLOW CYTOMETRY ANALYSIS OF VSG SURFACE EXPRESSION

Surface expression of VSGs was analyzed by FACS in live BSFs essentially as described before (Figueiredo *et al.*, 2008). $0.5x10^6$ of BSF from culture were centrifuged in a previously chilled eppendorf at 2800 *g* for 4 min at 4°C. Cells were resuspended in 50 µl of cold HMI-11 to which has been added 1:500 rabbit anti-VSG2 Alexa Fluor 488 conjugated antibody and 1:5000 rabbit anti-VSG13 Alexa Fluor 647 conjugated antibody (described before (Figueiredo *et al.*, 2008) and kindly gifted by George Cross). Cells were incubated for 15 min at 4°C with gentle shaking and protected from light. Cells were washed 3x in 1.3 ml of cold HMI-11, collected by centrifugation at 2800 *g*, 4min, 4°C and resuspended in 500 µl of cold HMI-11. Cells were kept on ice until analysis in a BD LSRFortessa (BD Biosciences) flow cytometer. To stain dead cells, 1 min before FACS analysis, 1 mg of propidium iodide (PI) was added to each sample. PI is a fluorophore that intercalates into nucleic acids. Because cell membrane is compromised in dying/dead cells, PI can diffuse inside these cells and stain their DNA. Data was analyzed on FlowJo v.X software. Dead cells were gated on the PI positive population and discarded from the analysis of VSG expression.

FAIRE

Formaldehyde-assisted isolation of regulatory elements (FAIRE) was adapted to *T. brucei* as described previously (Figueiredo & Cross, 2010). 1 x 10^7 BSFs were collected, cross-linking was performed in 1.1% formaldehyde in HM1-11 medium for 10 min at RT and reaction was stopped by adding glycine to a final concentration of 110 mM. To control for DNA input, an external DNA spike was added to each sample prior to sonication. The spike consisted in the commercial vector pBluescript-SK (Promega) (70 ng/sample) that contained an *AmpR* gene. Chromatin was sonicated using a Bioruptor (Diagenode) for 10 cycles (30 sec on / 30 sec off) and subsequently submitted to two consecutive phenol-chloroform extractions. DNA was Na-Acetate/ethanol precipitated in the presence of 20 µg/ml of glycogen, overnight at -20°C, resuspended in elution buffer (Qiagen) and treated with 100 µg/ml of RNase at 37°C for 1-2 h. Finally, DNA samples were purified using a QIAquick PCR Purification Kit (Qiagen). To control for the

gene copy number, a 'total non-cross-linked DNA' control was included in which the same protocol was followed except that cross-linking was omitted. Quantification of DNA isolated by FAIRE and in the toal non-cross-linked DNA samples was performed by real-time qPCR as described above. FAIRE enrichment was calculated by normalizing the DNA levels in FAIRE samples with DNA levels in non-cross-linked samples and total DNA input was controlled for by normalizing to the *AmpR* levels from spike.

CHROMATIN IMMUNOPRECIPITATION

Chromatin immunoprecipitation (ChIP) was performed essentially as described (Siegel *et al.*, 2009). BSFs (4-6 x 10^7 parasites) were collected, cells were fixed with 1.1% formaldehyde in HM1-11 medium for 20 min at RT. Fixation was stopped by adding glycine to a final concentration of 110 mM. After cell lysis, chromatin was sheared by sonication using a Bioruptor (Diagenode) for 10 cycles (30 sec on / 30 sec off). Immunoprecipitation was performed overnight at 4°C with IgG-coated magnetic beads (Dynabeads Protein G, Thermo Fischer Scientific) coupled with 10 µg of custom-made monoclonal anti-H3 antibody (kindly provided by Christian Janzen). Beads were washed with cold RIPA buffer (50 mM HEPES-KOH, pKa 7.55, 500 mM LiCl, 1 mM EDTA, 1.0% NP-40, 0.7% Na-Deoxycholate) and eluted with elution buffer (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA, 1% SDS). Cross-linking was reverted at 65°C for 9-15 h. Samples were treated with RNAse A (80 µg/sample) for 37°C, 2 h followed by treatment with proteinase K (80 µg/sample) at 55°C for 2 h. Immunoprecipitated DNA was purified using a QIAquick Gel Extraction Kit (Qiagen) To control for the total input of chromatin in-between samples, a 'total input sample' was prepared in parallel in which the same protocol was performed except the immunoprecipitation step. DNA was analysed by real-time qPCR has described above. Levels of DNA isolated by ChIP were calculated by controlling for DNA levels in the total input sample and normalizing to 18S rDNA levels.

RNA-SEQ DATA ANALYSIS

Total RNA was extracted 4 days after induction of H1 depletion in vitro. Ribosomal RNA was depleted with RiboMinus kit (Invitrogen). PolyA-containing RNA was purified using Dynabeads oligo-(dT) (Invitrogen) and cDNA was synthesized using random hexamers according to Complete RNA-seq Library System kit (NuGEN) instructions. Samples were sequenced in an Illumina HiSeq2000 platform (EMBL). Sequence reads were 100 nucleotides long and paired-ended. Their quality was accessed using the FASTQC quality control tool (http://www.bioinformatics.babraham.ac.uk/projects/ fastqc/). Contiguous read segments for which the quality score at each base was greater than 28 (-h 28), and longer than 25 nucleotides (-l 25) were selected using the DynamicTrim and LengthSort Perl-based software respectively (Cock et al., 2010). The sequencing data were then aligned to a hybrid genome composed of the TREU 927/4 genome (TriTrypDB 4.1), in which the VSG coding regions in subtelomeric VSG arrays were removed and replaced with the Tb427 VSG coding regions as a separate chromosome. Alignments were obtained using bowtie 2 (Langmead *et al.*, 2009), allowing for one mismatch. Non-unique reads (i.e. aligning in different locations) were excluded and the number of reads mapping to each feature (gene) was measured with the htseq-count software (http://www-huber.embl.de/users/anders/HTSeq/ doc/count.html) using the 'intersection-strict' mode. These sequence data have been submitted to the ArrayExpress database (EBI-EMBL) under Accession No. E-MTAB-1715. Differential gene expression was analysed in R (v 3.0.2), using DESeq (v 1.14.0) (Anders & Huber, 2010) from Bioconductor (v 2.6). DESeq uses the negative binomial distribution for modelling read counts per genomic feature. We first estimated the relative library sizes with the function estimateSizeFactors: for a matrix of raw read counts (where samples are represented in columns and genes in rows), each column is divided by the geometric means of the rows and the median of these ratios is used as the size factor for that specific column. Then, we used the function estimateDispersions: for each condition, it first computes an empirical distribution value for each gene, then fits by regression a dispersion-mean relationship and finally chooses for each gene, from the empirical and the fitted value, a dispersion parameter that will be used in subsequent tests. The function nbinomTest was finally used to test for differences between the

base means of the two conditions (RNAi and PL1S in this case). Feature regions were defined as gene coding sequence (CDS) regions, except in cases where the number of uniquely mappable positions inside the CDS is smaller than 100, in which case UTRs (annotated by Nilsson et al., 2010) were also included. To determine uniquely mappable positions we extracted all 100bp sequences (pseudo-reads) appearing in the reference genome and mapped them back using the same software and parameters used previously. After filtering out non-unique alignments, positions where one pseudo-read was successfully mapped were considered mappable. We corrected for multiple comparisons by controlling the FDR, following the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995), which relies on the P-values being uniformly distributed under the null hypothesis consists of sorting the P-values in ascending order, and then dividing each observed P-value by its percentile rank to get an estimated FDR (Noble, 2009). Genes with a FDR-adjusted P-value < 0.05 were considered significant.

GENE SET ENRICHMENT ANALYSIS

Gene set enrichment analysis (GSEA) was performed as described before (Subramanian *et al.*, 2005). GSEA allows us to investigate if specific gene sets (i.e. groups of genes that share common biological properties of interest) show statistically significant, concordant differences between two biological states and are therefore correlated with the phenotypic distinction under study. Non-expressed genes were removed from the analysis and the remaining were ranked by differential expression P-value. For a given gene set S (in Fig. 5D 'Genes inside BESs and procyclin loci'; in Fig. S8 'Genes inside BESs', 'VSGs inside BESs', 'VSGs outside BESs' and 'Genes inside procyclin loci'), an enrichment score was calculated by walking down the ranked list of genes increasing a running-sum statistic when the gene is in S and decreasing it otherwise and extracting its maximum value at the end. Thus, a high enrichment score indicates a preference for the genes in S to be among the more significantly differentially expressed genes. Statistical significance of this enrichment was determined by comparing the observed maximum enrichment score for gene set S with the

distribution of maximum scores obtained by randomly permuting the gene ranks 1000 times (histogram in the top right corner), allowing the estimate of the associated FDR-adjusted P-value. That significance is also illustrated by comparing the distribution of scores across observed gene ranks (blue line) with a simulation of scores distribution if gene ranks were randomly selected (grey line). This simulation was generated by 1000 random permutations of the gene ranks and represented as the 95th percentile of the enrichment scores obtained.

LUCIFERASE ACTIVITY

Luciferase activity was measured with a Firefly Luciferase Kit (Biotium) from 1.5 × 106 parasites following the manufacturer's protocol. Luminescence was measured in an Infi- nite M200 plate reader (Tecan).

4_SU-LABELLING AND PURIFICATION OF NASCENT TRANSCRIPTS

For metabolic labelling of newly transcribed RNA, 1.5–1.85 x 10^8 BSFs were collected by centrifugation at 970 g for 10 min, at room temp. and, washed three times each with 25, 50 and 2 ml of trypanosome dilution buffer (TDB) (5 mM KCl, 80 mM NaCl, 1 mM MgSO4, 20 mM Na2HPO4, 2 mM NaH2PO4, 20 mM glucose, pH 7.7). Cells were resuspended in 2 ml of TDB and 4-thiouridine (4sU) (Sigma) was added to a final concentration of 500 µM for 2, 5 or 10 min at 37°C. In experiments where transcription was inhibited, actinomycin D (Sigma) was added to a final concentration of 20 µg ml^{-1}, 5 min before starting 4sU-labelling. A negative control sample, with parasites with no 4sU addition (NO 4sU sample), was also included for each experiment. Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's protocol and RNA was dissolved in RNasefree Tris-EDTA buffer (TE) (10 mM Tris-HCl pH 8, 1 mM EDTA). Sample quantity and quality was assessed in a NanoDrop2000 (Thermo Scientific). Biotinylation and purification of 4sU-labelled RNA was performed essentially as described before (Dolken *et al.*, 2008). Labelled RNA was biotinylated with EZ-Link Biotin-HPDP (Pierce), dissolved in dimethylformamide. Biotinylation

reaction was carried out in labelling buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) and 0.2 mg ml−1 Biotin-HPDP for 1.5 h at room temperature. An amount of 50–90 µg of total RNA was used for the biotinylation reaction. Unbound Biotin-HPDP was removed by chloroform/isoamyl alcohol (24:1) extraction. RNA was precipitated with a 1:10 volume of 5 M NaCl, an equal volume of isopropanol and 15 µg of GlicoBlue (Life Technologies) at 20000 g for 20 min, 4°C. The pellet was washed with an equal volume of 75% ethanol and resuspended in RNase-free TE. Biotinylated RNA was captured using µMACS streptavidin beads and columns (Miltenyi). Up to 58 µg of biotinylated RNA was incubated with 100 µl of beads for 15 min at room temperature, with rotation. Beads were magnetically fixed and washed 3× with washing buffer (100 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween20) at 65°C followed by 3× washes with washing buffer at room temperature. 4sU-labelled RNA was eluted in two rounds by adding 100 µl of freshly prepared 100 mM dithiothreitol (DTT). Eluted RNA was precipitated with a 1:10 volume of 5 M NaCl, 2.5× volume of 100% ethanol and 30 µg of GlicoBlue, at −80°C for at least 30 min. RNA was washed in 75% ethanol and recovered by centrifugation as described above and resuspended in RNase-free water. cDNA synthesis and qPCR analysis was performed as described in a previous section. Levels of nascent transcripts were assessed by calculating the ratio of intergenic/CDS levels for 18S and β-tub. In this case, an internal normalization for 18S was not needed since two types of transcripts are being compared inside the same sample. The fold-change relative to NO 4sU samples was then calculated and plotted. In the experiment where transcription was blocked by actinomycin D treatment, an equal amount of RNA was used to synthesize cDNA and no further normalization was applied. Transcript levels for each locus in the MACS-isolated fraction were plotted as foldchange to the NO 4sU sample. To compare nascent transcript levels in H1 RNAi clones relative to PL1S, qPCR levels were first normalized to those of 18S transcripts and afterwards fold-change was calculated for H1 RNAi clones at each locus, relative to its expression in PL1S.

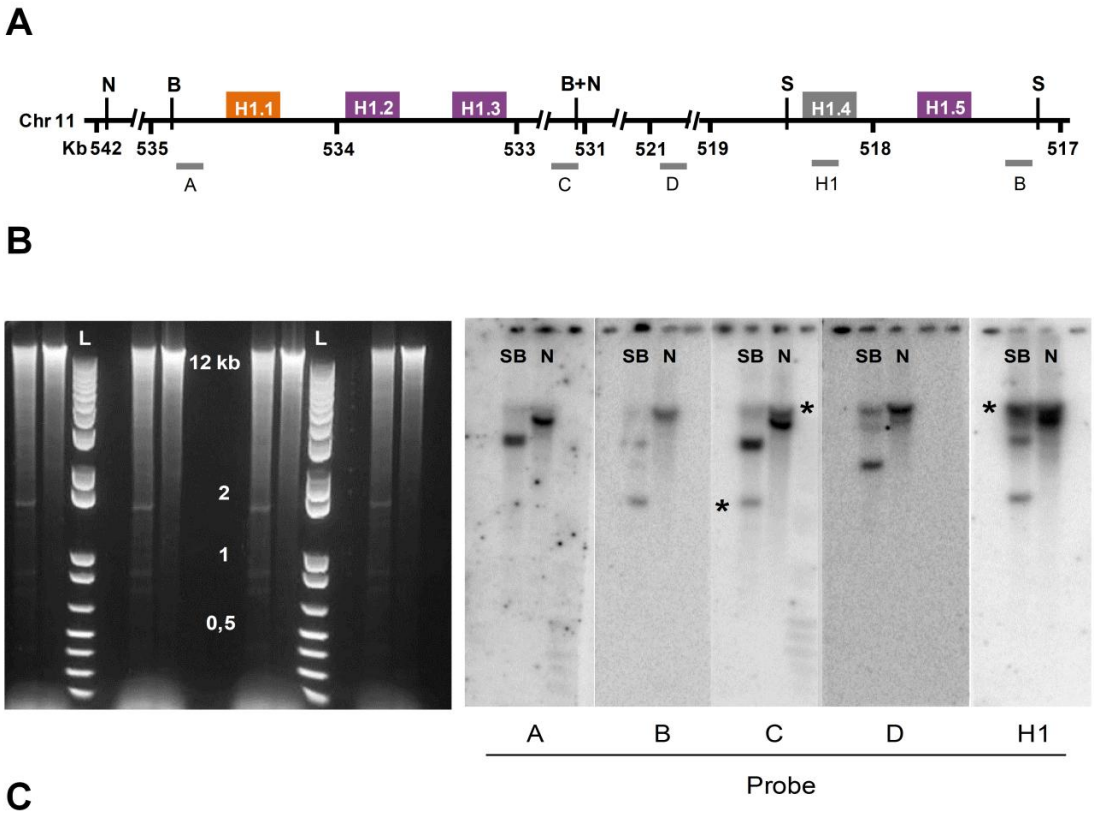
MMS AND PHLEOMYCIN-INDUCED DNA DAMAGE

Parasites in which RNAi has been pre-induced for 2 days with tetracycline (1 µg ml^{-1}), were incubated in a 96-well plate (200 µl per well) with sixfold range dilutions of methyl methanesulphonate (MMS; Sigma) or phleomycin (Cayla) in HMI-11, at a final concentration of 0.5–4 × 10^4 cells ml^{-1} . After 2 days, cell viability was measured with AlamarBlue® (Sigma). Ten microlitres of Alamar Blue® was added per well followed by incubation at 37°C, 4 h and fluorescence was measured (530ex/590em nm). RNAi induction with tetracycline and selection with Hygromycin was maintained during the 2-day exposure to MMS. MMS concentration is expressed in parts per million (ppm) of volume.

STATISTICAL ANALYSIS

To estimate statistical significance of data from FAIRE, ChIP and 4sU experiments calculated as a fold increase, we computed a log-normal posterior distribution of the fold-change, using an empirical Bayes approach. The statistical significance for remaining comparisons was given by non-parametric Mann–Whitney U tests. A *P*-value < 0.05 was considered significant. For the statistical analysis of RNA-Seq data, see RNA-Seq methods above.

3.1.4 SUPPLEMENTARY FIGURES AND TABLES



H1 locus	Primers	Amplicon (Kb)	N-terminal sequences identified
H1.1-1.3	Fwd 5'-ggaaagtagaaaggaaaataaaat-3' Rev 5'-cgttcgttgtacacaggaac-3'	1.65	5'- MAKTT, MAKASA, MAKASA -3'
H1.4-1.5	Fwd 5'-ggcggcaacttactgcag-3' Rev 5'-gtccgatgaacagatgcattc-3'	1.8	5'- MNNTT, MAKASA -3'
Putative new H1 locus	Fwd 5'-ggcggcaacttactgcag-3' Rev 5'-cgttcgttgtacacaggaac-3'	3	MNNTT

Figure S1. Genotyping of histone H1 loci. (A) Diagram of H1 loci with hybridization probes and restriction enzyme target sequences represented. H1 loci are colored according to the code associated with the 5'UTR-specific primers and protein N-terminal sequences defined for Figure 1. **(B)** Southern blot analysis from *T. brucei* Lister427 genomic DNA. DNA was double-digested with SmaI (S) and BamHI (B) (SB lane) and single digested with NheI (N lane). DNA restriction fragments were separated by standard agarose gel electrophoresis. DNA probes were made by PCR amplification of specific gene sequences (probe A GeneID: Tb927.11.1890; probe B GeneID: Tb927.11.1780; probe C GeneID: Tb927.11.1850; probe D GeneID: Tb927.11.1820 and probe H1 GeneID: Tb927.11.1800) and 32P-radiolabeled. Blots were hybridized and visualized by phosphor imaging. There were three unexpected bands (*) which suggest the existence of an additional H1 locus. **(C)** Summary table of PCR and sequencing results of each H1 loci. A putative new locus was identified by PCR amplification and sequencing, being composed by a unique histone H1 gene.

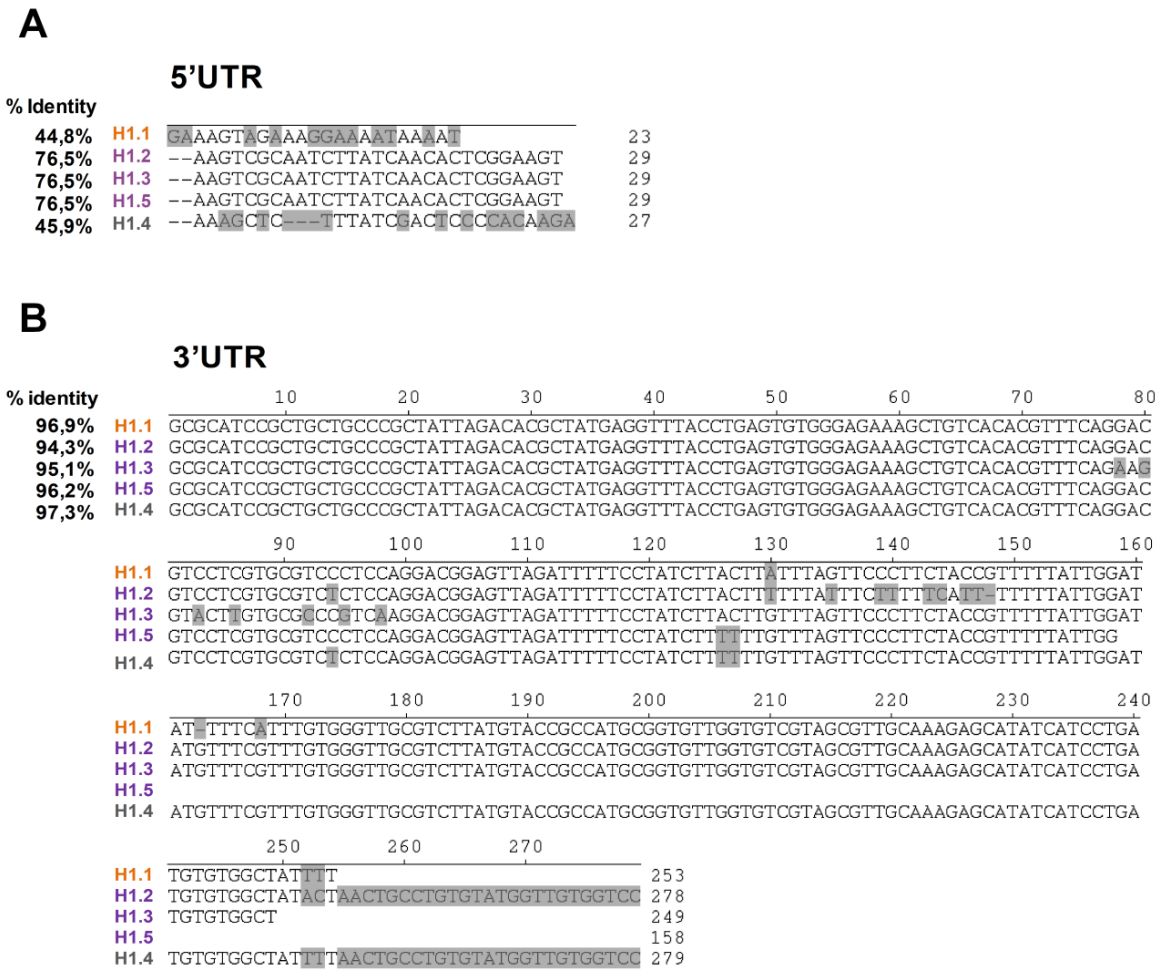


Figure S2. Each histone H1 class has a unique 5'UTR. (A) Alignment of the predominant 5'UTRs and **(B)** 3'UTRs of histone H1 variants of *T. brucei.* Average percentage of sequence identity between each variant is shown on the left and sequence length on the right of each sequence. Divergent nucleotides are indicated in gray. Gaps (-) were introduced for best alignment. Predominant UTRs in BSFs were derived from (Siegel *et al.*, 2010) and (Manful *et al.*, 2011).

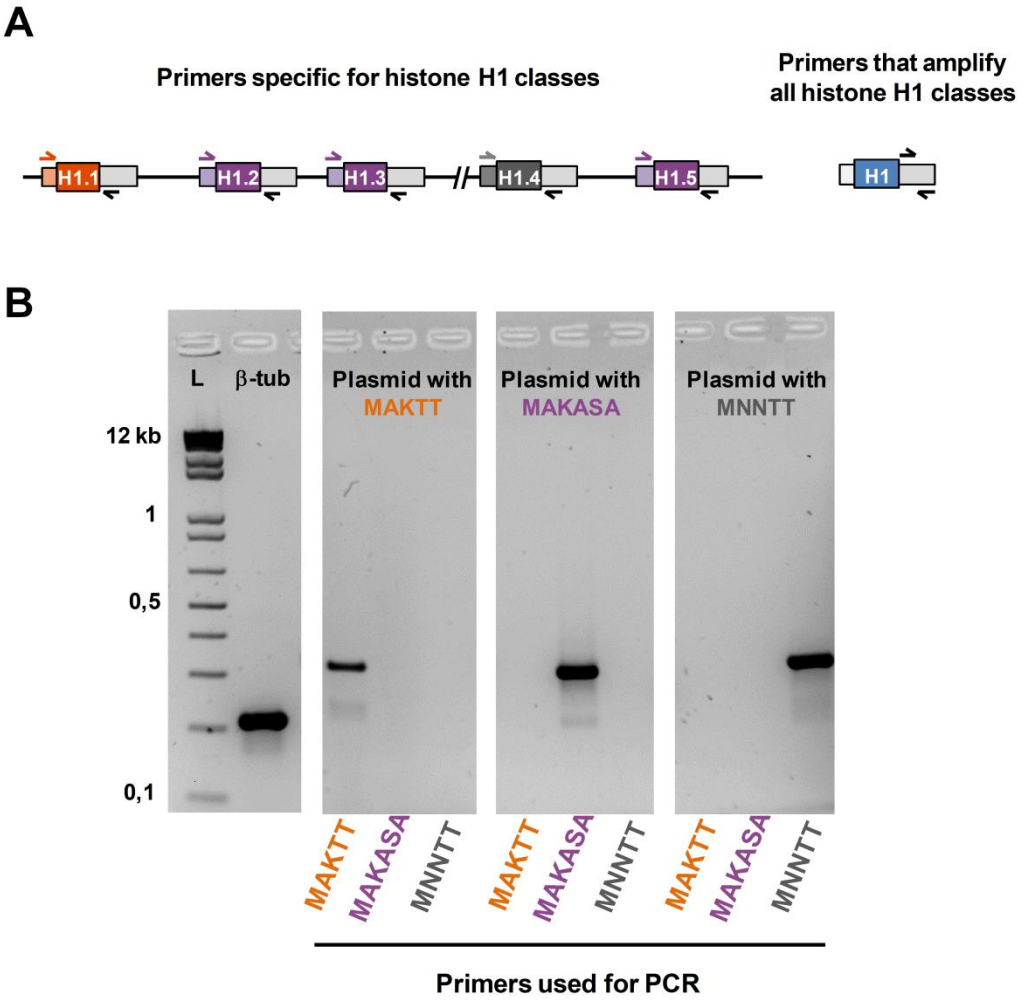


Figure S3. Specificity of histone H1 primers. (A) Class-specific amplification of histone H1 was obtained by using forward primers that annealed with the class-specific 5'UTR and a reverse primer that annealed with the conserved 3'UTR sequence common to all H1 classes. Amplification of all H1 classes was obtained by using a forward and a reverse primer that annealed both to 3'UTR sequences common to all H1 classes. **(B)** Class-specificity of the primers was confirmed by PCR using as a template plasmids containing, either MAKASA, MAKTT or MNNTT gene. Primers for a given class only amplify a fragment of the corresponding class.

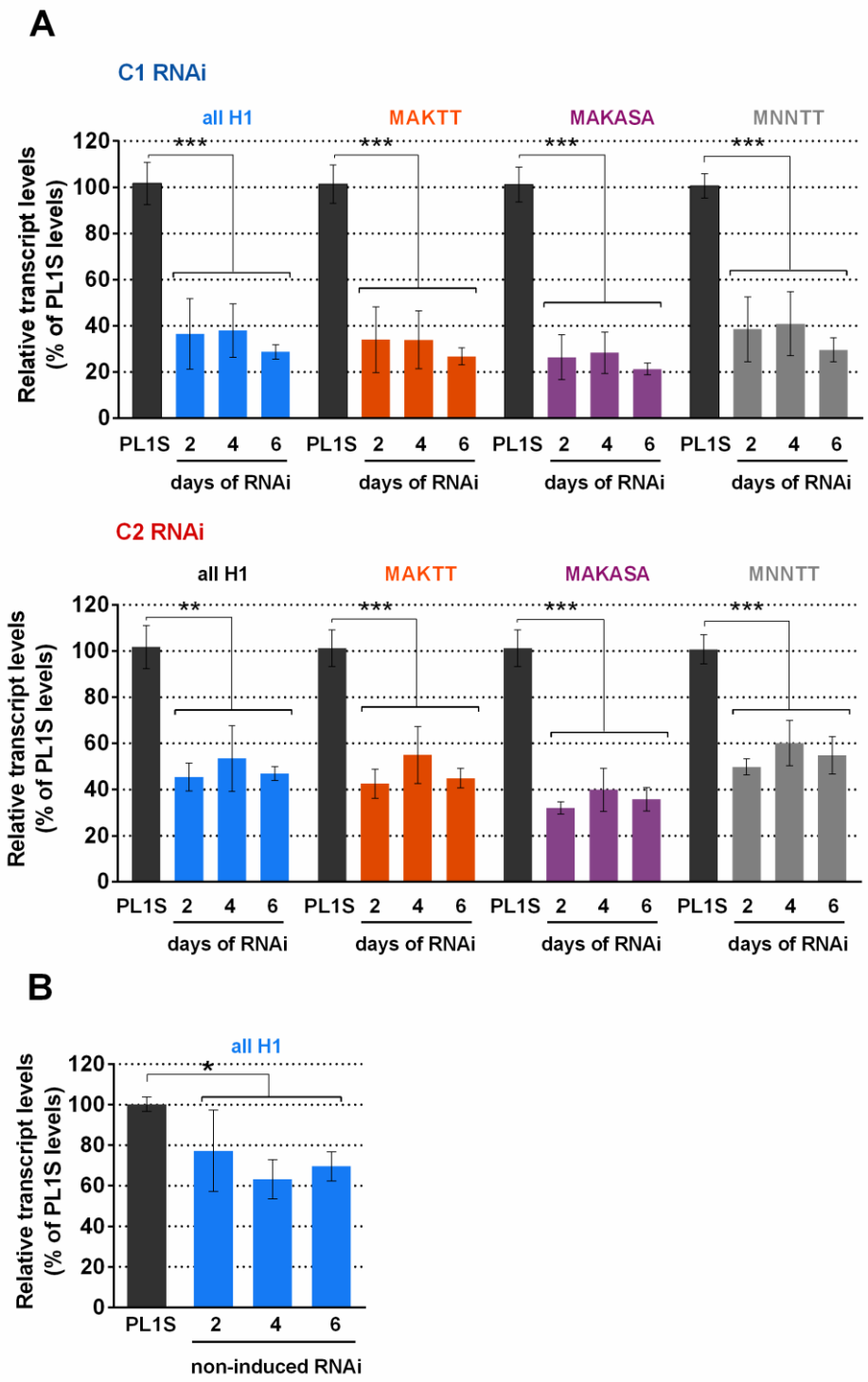


Figure S4. Efficiency of Histone H1 depletion. (A) Percentage of depletion of histone H1 classes during RNAi induction for 2, 4 and 6 days in RNAi clones C1 and C2 relatively to the parental cell-line PL1S. Primers for all H1 classes and primers specific for each H1 classes were used. Three independent experiments were analyzed. **(B)** Non-induced RNAi clones C1 and C2 show a reduction of histone H1 from 77% to 63% of PL1S levels. One and two independent experiments were analyzed for C1 and C2, respectively. Results are shown as mean ± SEM. Statistical significance was determined by a non-parametric Mann-Whitney test. * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$.

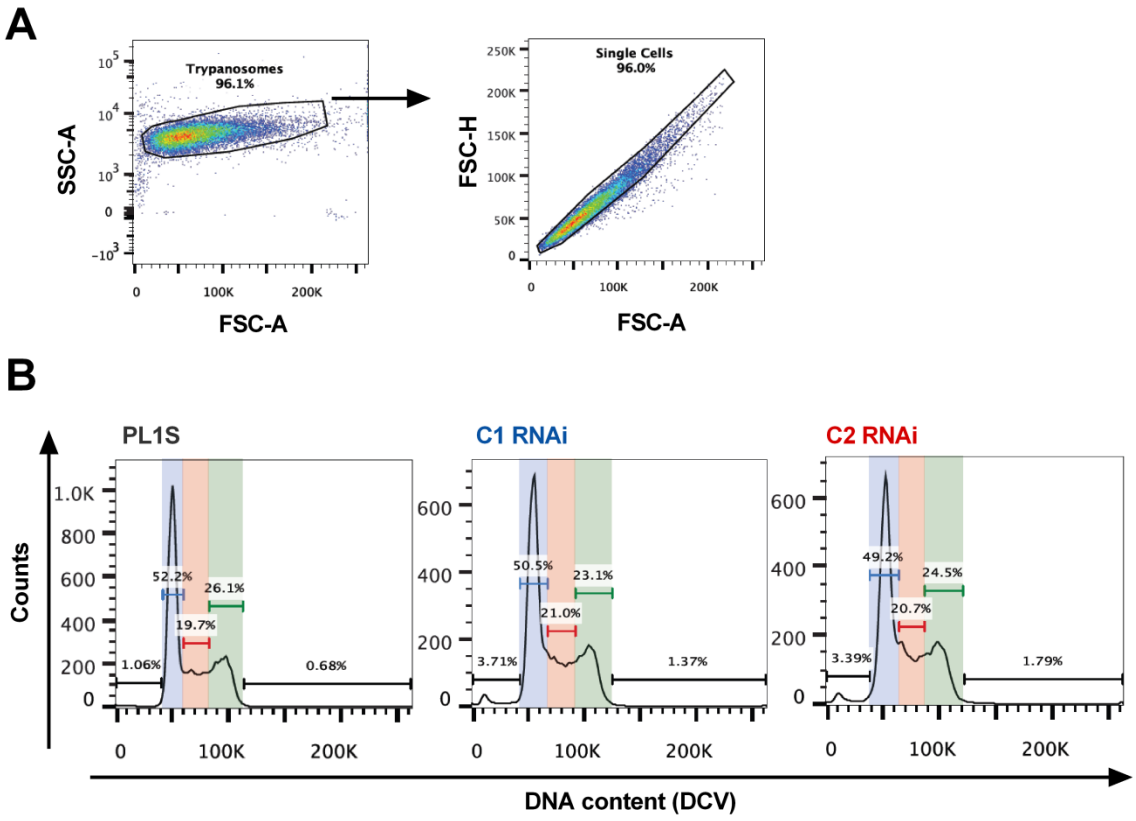


Figure S5. H1 depletion causes no major changes in cell cycle. (A) Gating strategy used in FACS to select trypanosomes. Trypanosomes were gated using SSC-A and FSC-A, and singlets were selected from the gated trypanosomes, using FSC-A and FSC-H. **(B)** FACS histograms of cell cycle profiles of PL1S, C1 RNAi and C2 RNAi after 6 days of RNAi. DNA content was measured by Vybrant® DyeCycle™ Violet (DCV) staining in live parasites. Frequency of parasites is normalized to mode (most frequent value of DCV intensity). Gates used to quantify the fraction of each cell cycle stage are shown. < 2C, DNA content lower than the diploid chromosome number; > 4C, DNA content higher than the tetraploid chromosome number. One independent experiment was analyzed.

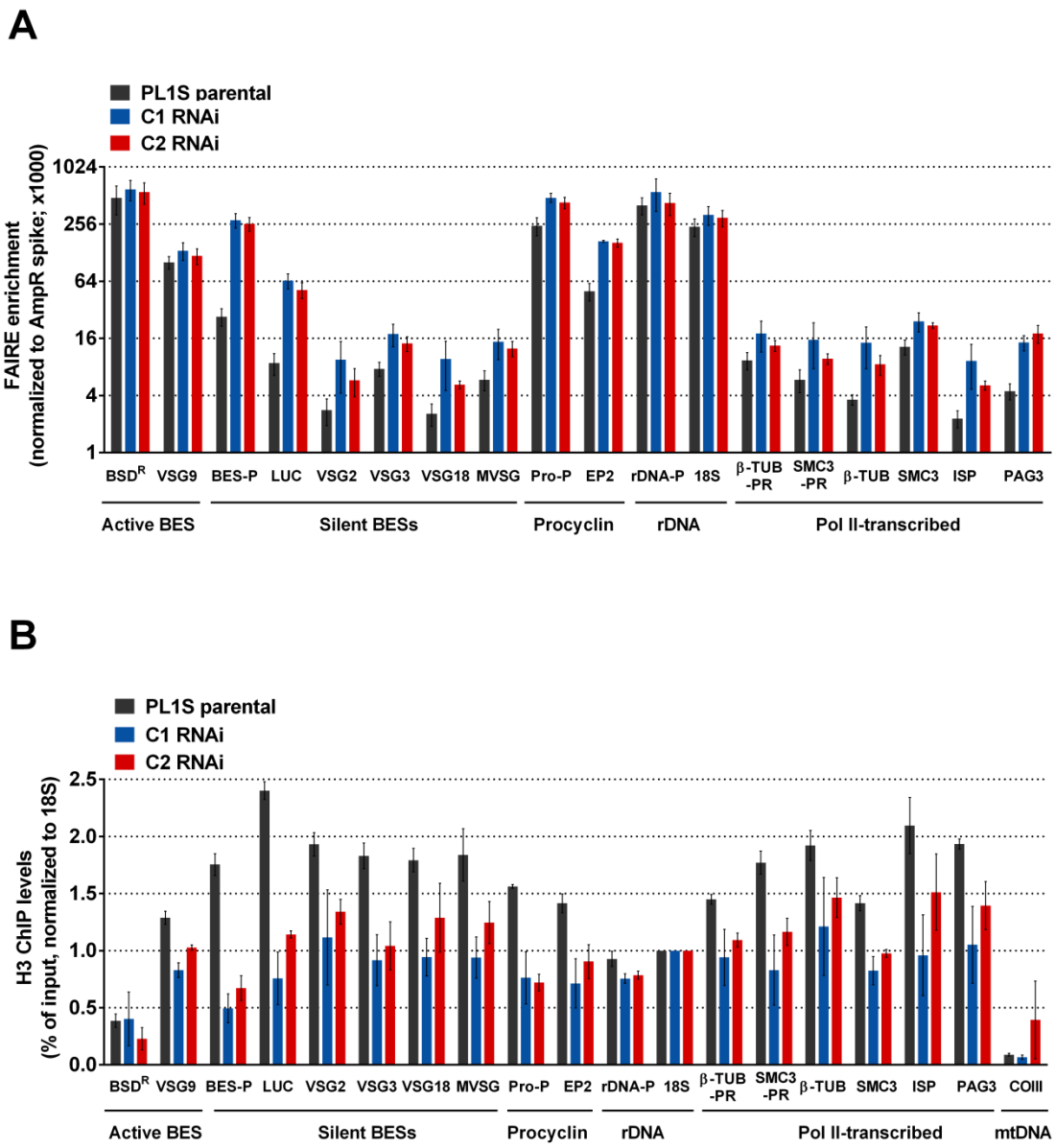


Figure S6. H1 compacts chromatin at different levels across genome. (A) FAIRE enrichment and **(B)** H3 ChIP levels in individual H1 RNAi clones C1 and C2 after 6 days of RNAi induction. For FAIRE an AmpR gene contained in a plasmid spike was used as a normalizer for DNA input. For H3 ChIP, DNA was quantified by qPCR, compared to the total input material and normalized to 18S DNA. Results are expressed as fold-change relatively to parental cell-line PL1S. Note that FAIRE enrichment is represented in a logarithmic scale. Two to three independent experiments were analyzed. Results are shown in mean ± SEM. BSD^R, blasticidin resistance gene; VSG, variant surface glycoprotein gene; BES-P, promoter region of silent BESs; LUC, luciferase gene; MVSG, metacyclic VSG gene; Pro-P, procyclin promoter region; EP2, procyclin EP2 gene; rDNA-P, rDNA promoter region; 18S, ribosomal 18S gene; β-tub, β-tubulin gene; SMC3, structural maintenance of chromosome 3 gene; β-tub-PR, promoter region of β-tub polycistronic unit; SMC3-PR, promoter region of SMC3 polycistronic unit; ISP, inhibitor of serine peptidase gene; PAG3, procyclin associated gene 3; COIII, cytochrome c oxidase subunit III gene; mtDNA, mitochondrial DNA.

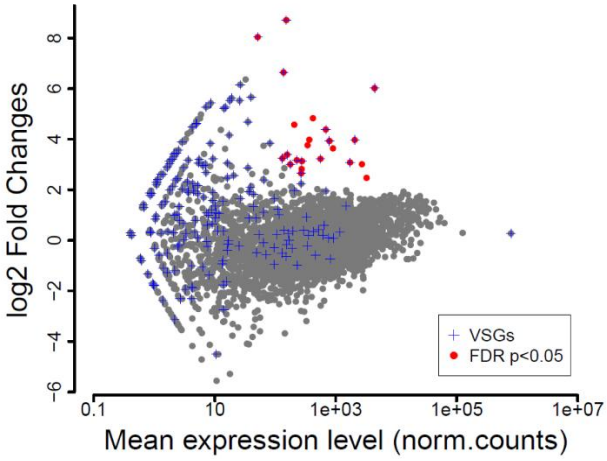


Figure S7. Gene expression MA-plot of H1-depleted clones relative to parental clone. Mean expression levels (normalized counts) across samples are plotted on the x-axis and the corresponding log2 fold-changes between H1 RNAi and PL1S samples are plotted on the y-axis. Each point represents a gene. Genes with false discovery rate (FDR) adjusted *P*-values<0.05 are considered statistical significant and are indicated in red; VSG genes are indicated in blue.

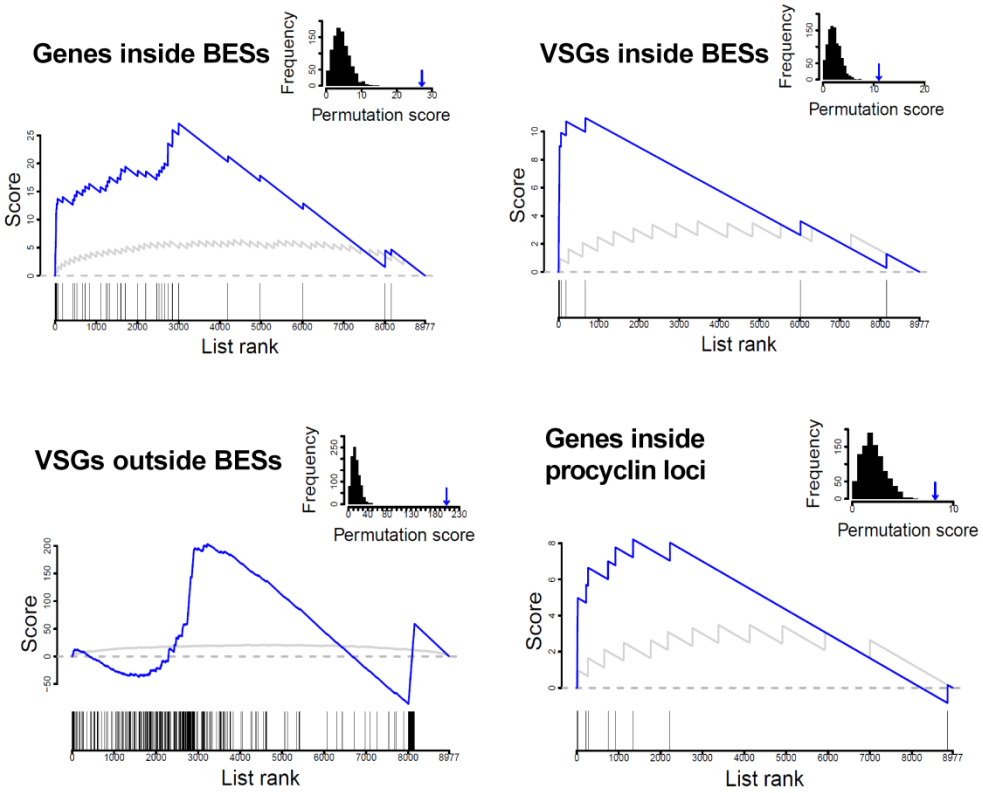


Figure S8. H1 regulates transcription of genes located in bloodstream expression sites and procyclin loci. Gene set enrichment analysis plots showing, for the subset of genes in each plot title, the enrichment score distributions across genes ranked by decreasing statistical evidence of differential expression (blue line). Light gray line represents 95th percentile of the enrichment scores obtained by 1,000 random permutations of the gene ranks. The histogram at the top-right corner of the figure represents the distribution of maximum scores obtained by random permutation, with the arrow indicating the experimental enrichment score (i.e. the "peak" in the main plot). The diagrams at the bottom show where the members of each gene set appear in the ranked list of genes.

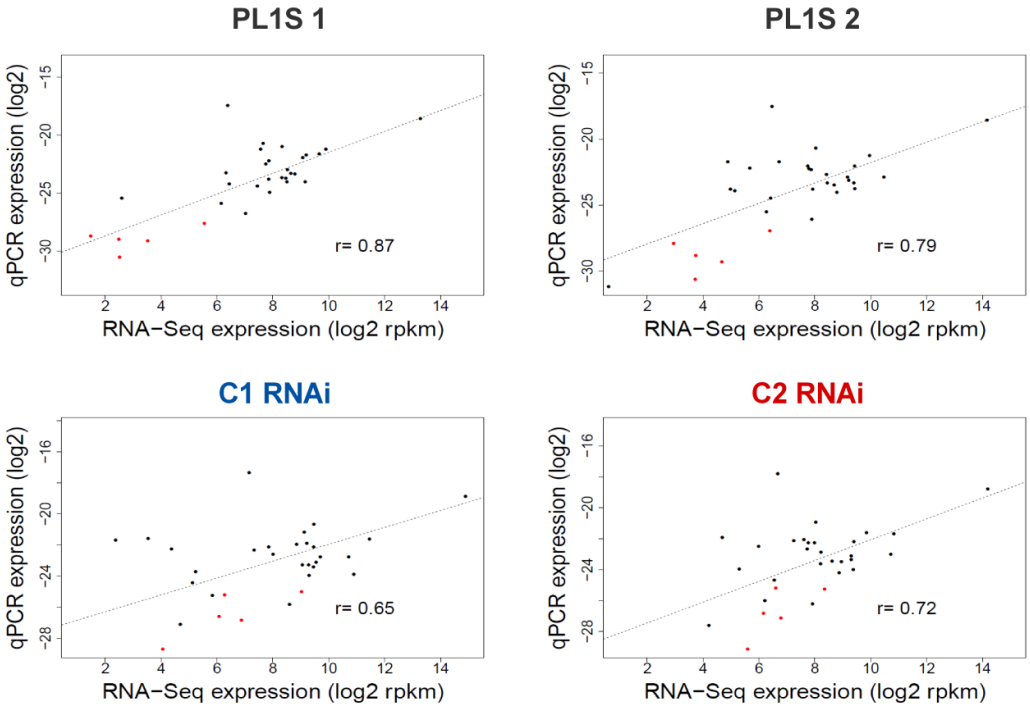


Figure S9. Validation of RNA-Seq results by qPCR. Scatter plots of log2 qPCR levels and log2 RNA-Seq RPKMs for each sample (PL1S and H1-depleted clones, C1 and C2). 34 genes with mean expression levels ranging from 12 to 512,586 RPKMs by RNA-Seq were randomly selected and quantified by qPCR. VSG genes are indicated in red. Pearson's correlation coefficients (r) are indicated.

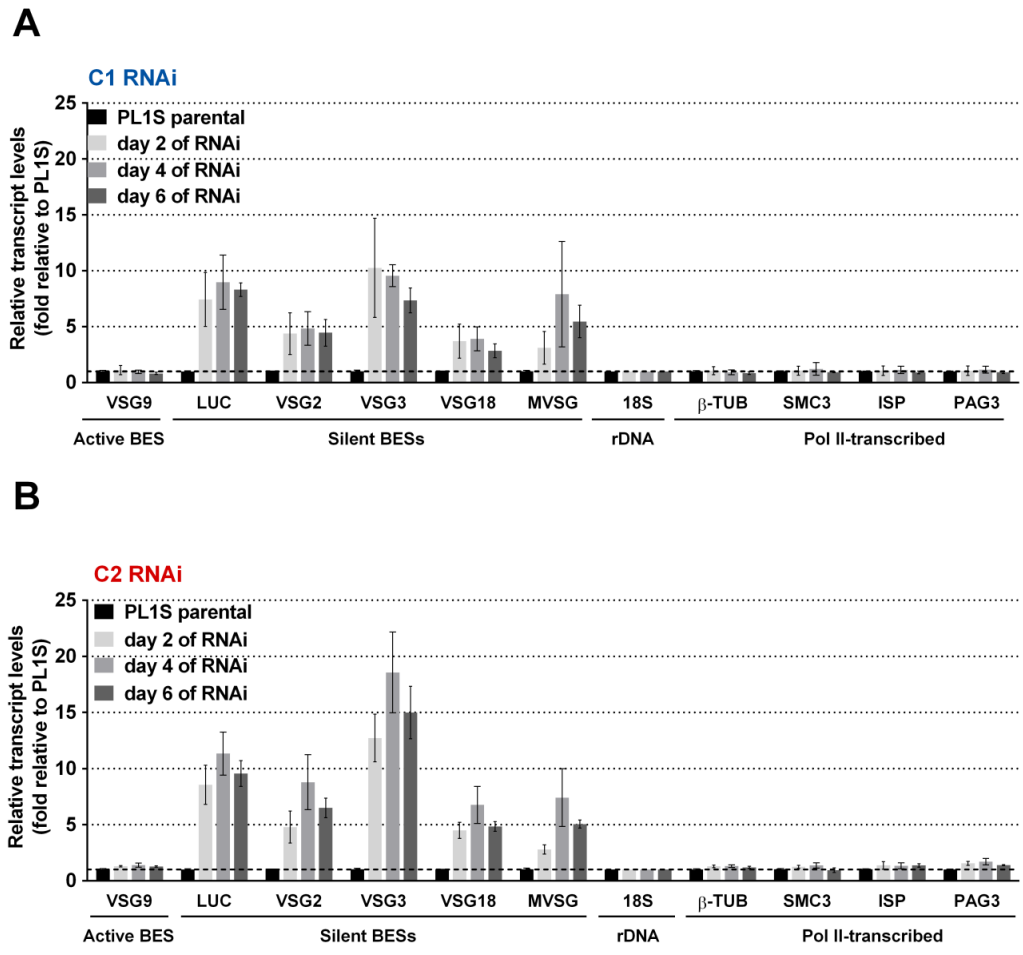


Figure S10. (previous page) Loss of histone H1 leads to VSG derepression. Quantification of gene expression 2, 4 and 6 days after RNAi induction in clone C1 (A) and clone C2 (B), relative to the parental-cell line PL1S. Silent BESs and VSGs become derepressed when histone H1 is depleted. Three independent experiments were analyzed. Results are shown as mean ± SEM. Statistical significance was calculated using a non-parametric Mann-Whitney test. Significant fold-changes ($P < 0.05$) are shown above dashed-line.

Table S1. Comparison of P-values for fold-change levels of H1-depleted clones relative to PL1S in FAIRE and ChIP experiments. Non-significant p-values (>0.05) are highlighted in bold.

	Fold relative to PL1S		P value	
Loci	**FAIRE**	**ChIP-H3**	**FAIRE**	**ChIP-H3**
BSD	1.28	0.59	**7.8155e-02**	4.3350e-02
VSG9	1.25	0.74	**1.1154e-01**	3.8466e-02
BES-P	10.42	0.38	2.3594e-45	1.2662e-08
LUC	7.08	0.47	1.1271e-32	4.3546e-07
VSG2	2.59	0.70	2.7651e-08	5.1371e-03
VSG3	2.05	0.59	8.0545e-06	3.8335e-04
VSG18	2.77	0.68	3.5050e-09	4.0048e-03
MVSG	2.30	0.63	2.3810e-07	1.6109e-03
Pro-P	1.95	0.53	3.1702e-05	6.1951e-05
EP2	3.60	0.60	2.1755e-14	1.7638e-03
rDNA-P	1.18	0.86	**1.9650e-01**	**2.3235e-01**
18S	1.29	1.00	**6.1862e-02**	-
β-tub-PR	1.67	0.77	1.9647e-03	3.4793e-02
SMC3-PR	2.12	0.61	1.3259e-05	7.5486e-04
β-tub	2.99	0.75	4.0836e-10	1.5674e-02
SMC3	1.79	0.70	2.2672e-04	1.2807e-02
ISP	2.98	0.62	1.9575e-10	6.7962e-04
PAG3	3.80	0.71	2.8251e-15	5.3581e-04

Table S2. Primers used for quantitative real-time PCR.

Primer name /GeneID	Used to validate RNA-Seq	Amplified region	Forward primer (5'-3')	Reverse primer (5'-3')
All H1		All classes of H1 genes	GGAGAGAGCTGTCACACG	ATGGCGGTACATAAGACGC
BES-P		Silent BESs promoter region (97bp downstream TSS)	AAAACCCTTAGCGTTACCAC	AAACAACACACCAGCCCTC
BSD^R		Blasticidin resistance gene	CGGCTACAATCAACAGCATC	ACGATACAAGTCAGGTTGCC
β-tub	✓	b-tubulin gene	TTCAGGCTGGCCAATGCG	TACGGAGTCCATTGTACCTG
β-tub-PR		b-tubulin promoter region	AGCCGCTACTCGATTTCACA	TCGCCATACCGACACAGC
COIII		Cytochrome c oxidase subunit III gene	GAGGGAACGGGAGAGGAACG	TGTTTCTGCAAACAATCTCATCTGG
EP1	✓	Procyclin EP2 gene (3'UTR)	AATGTCTTATTAACCATCGCCTG	AAAATTATGGAATACGCAACCG
EP2	✓	Procyclin EP2 gene (3'UTR)	GGTGGCCCCAGTTATTTCTTT	TATGCAAGTGTCCTGTCGCC
ISP		Inhibitor of serine peptidase gene	GTCGTGTGATGGAGGATGG	TGAAGTGCTTGGTCGGAAC
LUC	✓	Luciferase reporter gene	ATGTCCGTTCGGTTGGCAG	CATACTGTTGAGCAATTCACG
MAKTT		MAKTT class H1 gene	GGAAAGTAGAAAGGAAAATAAAAT	TCTAATAGCGGGCAGCAG
MAKASA		MNNTT class H1 gene	TCGCAATCTTATCAACACTCG	TCTAATAGCGGGCAGCAG
MNNTT		MAKASA class H1 gene	CTTTATCGACTCCCCACAAG	TCTAATAGCGGGCAGCAG
MVSG		Metacyclic variant surface glycoprotein 639 gene	TCGCACTTTCAGCTCTGCTC	GCCGACCACTCGCTGTCC
PAG3		Procyclin-associated gene 3	GCAAATGCTCCTCTTCTCC	TGGCTGCCGAAGCAAGCG
Pro-P		Procyclin promoter region (10bp downstream TSS)	AGTTTAAGATGTTCTCGTGAT	CTTTTTGGTGTAATTGAAGTC
rDNA-P = 18S Int		Ribosomal DNA promoter region (15bp downstream TSS)	CTGACCGCTCTCAGACCG	ACACACGTATTACACACACTC
SMC3	✓	Structural maintenance of chromosome 3 gene	GAGGCTGGATGATGAGAGG	ATTCAGCTTCACTGATGATGG
SMC3-PR		SMC3 promoter region	GGCGGTGTGGGTGTATCC	ATCGTCGCCCCTAAGTGC
VSG2	✓	Variant surface glycoprotein 2 gene	AGCAGCCAAGAGGTAACAGC	CAACTGCAGCTTGCAAGGAA
VSG3	✓	Variant surface glycoprotein 3 gene	GCTTATTTTGTGTCTGTCGC	GACGCAGCAGAATCAACAC
VSG9	✓	Variant surface glycoprotein 9 gene	ACTAAGCTCGTGGCGCAC	CGCGTAGTTGACGCATGAC
VSG13	✓	Variant surface glycoprotein 13 gene	ATAACGCATGGCCATCTTGAC	GTCGTTGCTGTGGATTGCTC
VSG18	✓	Variant surface glycoprotein 18 gene	ACAGACCGCCGACAGTATC	GTATCTTTGTAGGCCGCTGC
VSG21	✓	Variant surface glycoprotein 21 gene	CAGCGCAAGTACAGGACG	TGCTTCGTCGTCGCTTAC
18S		18S ribosomal DNA	ACGGAATGGCACCACAAGAC	GTCCGTTGACGGAATCAACC
Tb927.10.2080	✓	hypothetical protein, conserved	AGCTGCAGCCACAGATATTC	TTGTTGCATTGGTGGCGTAC
Tb927.10.2400	✓	hypothetical protein	TTGTGGGTGCCAGAACTTTG	TGTTTCTCGTGCATCCCTTC

Tb11.02.5800	✓	calmodulin, putative	ACGCGTGAGTATTTGAAGCG	TCGTAGTCTGCTTCAGCCTTC
Tb927.7.4520	✓	hypothetical protein, conserved	ACAAGAAAGCAACGGGGAAG	TGGTGAAGCTCTGCAAGAAC
Tb09.211.2140	✓	hypothetical protein, unlikely	CGTTCCTTTTATTTCGCTCACG	AGATGATAAAGCAATGGGAGCTC
Tb09.211.4750	✓	hypothetical protein, unlikely	GGTGGCATGGTTGGGTAAAG	TTCCGGGTGTATGCAATTGC
Tb11.03.0370	✓	hypothetical protein, conserved	AGCAGCCTCTTTTGTGCTTC	CGTGATACGCAAAAGGGAAAC
Tb927.4.3430	✓	hypothetical protein, conserved	ATCCATGGCTTTCGCAAACG	ACGACACATTTCTGGAACGC
Tb09.211.3190	✓	hypothetical protein, unlikely	TGGGCGGTGATTCTCTTAAT	TTTAACCGAGGTGTAACGTGAA
Tb09.160.0380	✓	hypothetical protein, conserved	TCAGTTGTTGCGGAATGTCC	AACATCACGCACGCCATTG
Tb927.10.10760	✓	hypothetical protein	TTCCACGAGAGCGAAAAGTG	TCGAGCATTGACAAGGAACC
Tb927.10.11060	✓	hypothetical protein, conserved	TATGTTCGTGAACCGCACTG	AAAAGTCATCGTCGCGACTC
Tb09.160.2440	✓	hypothetical protein, conserved	TTTTCTTCTGGCAGCATCGC	AACCCGCGCTGTAACAATTC
Tb927.8.2470	✓	hypothetical protein, conserved	TGTTCAGCGCGAGAAAGTAC	ACTTTGGGCTATGCTCCTTCTC
Tb11.03.0935	✓	RNA polymerase subunit, putative (RPB6z)	GCACGCCAAATTGTGAATGG	ATGAATAGGATCGACGGATGCC
Tb09.211.2150	✓	Nuclear poly(A)-binding protein 1 (PABP2), putative (PABP2)	AGACATGCGGAACAAACAGC	ATGCGCGTGTTCATGAAGTG
Tb11.01.6220	✓	procyclin-associated gene 4 (PAG4) protein, putative	TGGAAACGAAAAGGGCAAGG	ATGACTTTCGTTCGCCCATG
Tb927.10.2100	✓	elongation factor 1-alpha,EF-1-alpha (TEF1)	AACATGATCACCGGCACATC	TGTTGCAGCACACAACCATC
Tb09.244.2400	✓	BARP protein (BARP)	AGTGTACTGATACAGCGGAAGG	TGCCGCAACTTTCTAGCAAC
Tb927.10.12100	✓	RNA-binding protein, putative (RBP7B)	TTCTACGGTGATGTGCTGCA	TTGTGAAGTCCGAGGATGGC
Tb927.7.5940	✓	Protein Associated with Differentiation (TbPAD2)	ATCGCAGCGTTCTCAAATGG	TTGTTGCGTGTACCACTCAC
Tb11.02.0740	✓	60S ribosomal protein L44	GCGAAGACGACCAAGAAGATTG	TTGTCGTTCAGCTCGAAGTG
Tb11.42.0004	✓	hypothetical protein, conserved	TGGCTTGGTCCGCATTTTTG	GCCACAATGCTGTTGGAATG

3.2 HISTONE H1 IS IMPORTANT FOR T. BRUCEI FITNESS *IN VIVO*

Results from the first part of this section (Results 3.2.2.1. H1 depletion impairs progression of acute infection in mice) were published, together with the previous results section (Results 3.1. *Trypanosoma brucei* histone H1 inhibits RNA polymerase I) in the Molecular Microbiology peer-reviewed journal. The article in the publication format can be found in the Annexes.

Ana C. Pena[1] and Luisa M. Figueiredo[1]

[1]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, Edifício Egas Moniz, 1649-028 Lisboa, Portugal.

3.2.1 AUTHOR CONTRIBUTIONS

Ana C. Pena and Luisa M. Figueiredo designed the experiments and wrote the manuscript. Ana C. Pena did all the experiments, analyzed the data and produced the figures. Luisa M. Figueiredo, mentored and contributed to the experimental design of the work.

3.2.2 RESULTS

3.2.2.1 H1 depletion impairs progression of acute infection in mice

Given the involvement of *T. brucei* H1 in resistance to DNA damage, *VSG* transcription and global chromatin condensation, it might be surprising how H1-depleted parasites can grow so well in culture (Results 1.3.1., Fig. 1A). A 50% reduction in H1 levels in mice is lethal (Fan *et al.*, 2003). In contrast, histone H1 is not necessary for growth or survival *in vitro* in the unicellular eukaryotes *S. cerevisiae* and *T. thermophila* (Shen *et al.*, 1995, Shen & Gorovsky, 1996, Patterton *et al.*, 1998). On the other hand, in another kinetoplastid parasite, *Leishmania major,* overexpression of histone H1 strongly reduces infectivity in mice (Smirlis *et al.*, 2006).

Thus, we decided to investigate the effect of H1 depletion during an *in vivo* infection, a setting where the parasite is exposed to a much more challenging environment compared with culture conditions. With this objective, we established infections in mice. We inoculated C57BL/6 mice with 100 parasites of the parental cell-line, PL1S, or H1-depleted clones. In the latter, RNAi was pre-induced for two days with tetracycline in culture. As previously published, mice also received doxycycline, a tetracycline analogue, in water two days before infection and during entire infection (Lecordier *et al.*, 2005). Doxycycline had no effect on PL1S-infected mice (Fig. S1). Interestingly, mice infected with H1-depleted parasites survived longer ($P < 0.01$ for C1 and $P < 0.001$ for C2) (Fig. 1A). Indeed, whereas 55% of the PL1S-infected mice died in the first nine days of infection, the majority of C1 or C2-infected mice survived until day 13 and 15, respectively. Such prolonged life-span was associated with a delay in parasite appearance in the blood (Fig. 1B; Fig. S2). In fact, whereas parental parasites reached detectable parasitaemia within 4–7 days of infection, H1-depleted clones took 4–15 days. This is much later than what would be expected from the growth delay measured *in vitro* (detectable parasitaemia should be reached at day 4–5), which indicates that host factors must have contributed to this diminished parasitaemia.

It is interesting to note that the least virulent clone (C2) was the one with the most efficient H1 depletion ($P < 0.05$; Fig. 1A–C), suggestive of a dose-dependent response. Consistent with the fact that in 'non-induced' RNAi clones H1 transcripts are in fact 20% lower than in parental strain (Results 1.5., Fig. S4B), in mice we observed that clone C2 caused a slight increase in life-span ($P < 0.01$) (Fig. 1A and Fig. S1). These results clearly indicate that although H1-depleted cells have hardly any growth defect *in vitro*, H1 is essential for parasite fitness *in vivo*.

In around 50% of the mice infected with the parental cell-line PL1S and 40–75% of the mice infected with H1 depleted clones, two peaks of parasitaemia were detected (Fig. 1B; Fig. S2). In the majority of the remaining mice infected with H1-depleted parasites, a single peak of parasitaemia appeared later, after nine days of infection. In order to test if parasites detected in the second peak of H1-depleted clones were RNAi revertants, i.e. parasites that have lost the capacity of performing RNAi, we measured by qPCR the H1 transcript levels at 9–14 days post-infection. We

observed that H1 transcripts were still reduced to around 40% and 12%, confirming that parasites were still depleted of H1 ($P < 0.05$) (Fig. 1C), indicating that parasites that grow later in infection are not RNAi revertants.

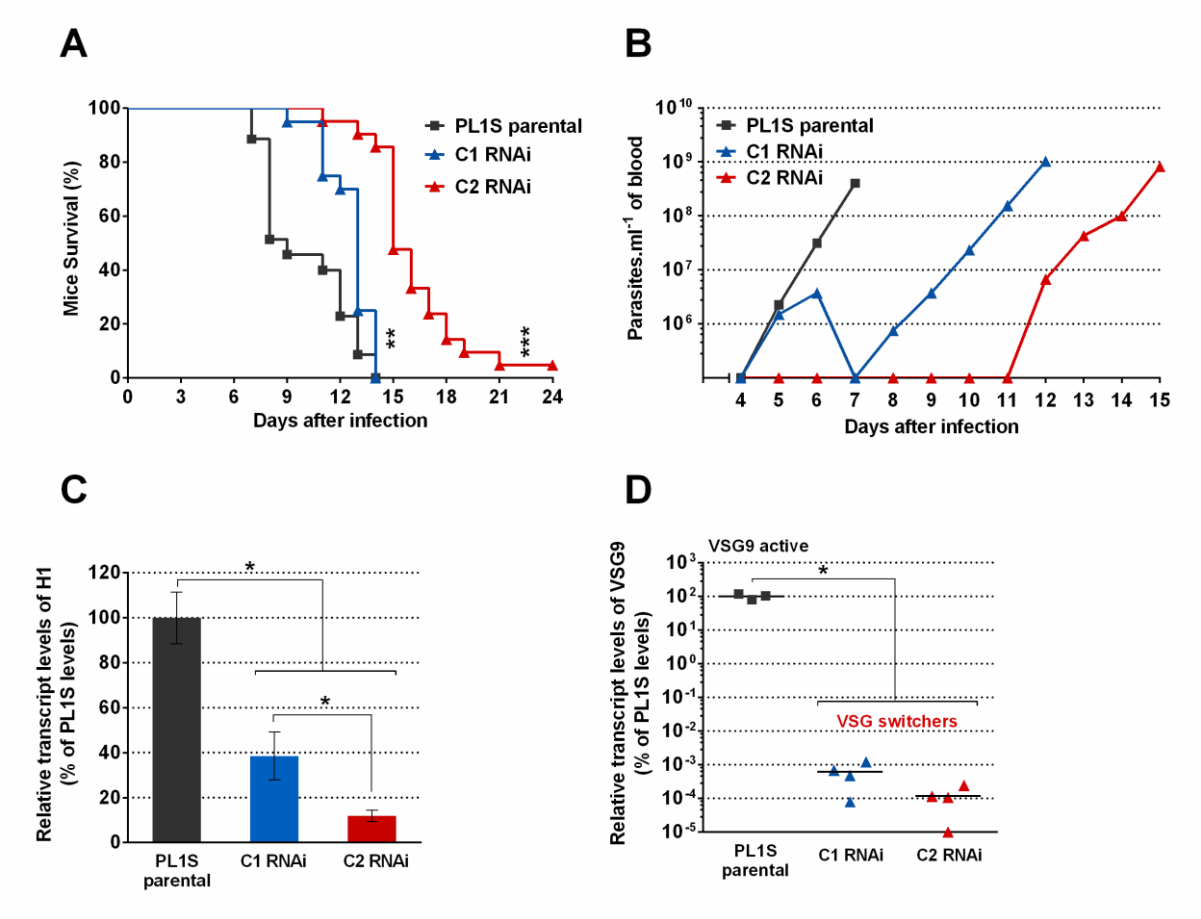


Figure 1. H1 depletion compromises T. brucei fitness *in vivo*. (A) Survival of mice infected with 100 parasites of parental cell-line PL1S or the two H1-depleted clones (C1 or C2). Three to four independent experiments were analyzed. Mice n = 20–35 per group. Survival curves were statistically compared by a Log-rank (Mantel-Cox) test. **(B)** Representative examples of parasitaemia in the blood of mice infected with T. brucei parasites depleted of H1. While PL1S parasites typically reached a peak in parasite number in 7–12 days, histone H1-depleted parasites had a delay of several days in parasitaemia. **(C)** Depletion of histone H1 in T. brucei parasites isolated from the blood of infected mice. PL1S parasites were collected at day 7 post-infection and C1 and C2 parasites at days 9–14 post-infection. Transcript levels of histone H1 were quantified with primers for all histone H1 classes by qPCR. Results are shown as percentage of transcript levels relative to PL1S levels. **(D)** Quantification of VSG9 expression in parasites isolated from the blood of mice (same parasites analyzed in (C)). C1 and C2 parasites underwent VSG switching and no longer express VSG9. Transcript levels were measured by qPCR. Mice n = 3–4 per group. Results are shown as mean ± SEM. Statistical significance was determined by non-parametric Mann-Whitney U tests. * P < 0.05, ** P < 0.01, ***P < 0.001.

Typically in a *T. brucei* infection, parasites in the second peak of parasitaemia express a *VSG* different from the one expressed in the initial parasite population. In

order to confirm that VSG switching took place, we collected C1 and C2 parasites from the blood on days 9–14 post-infection and tested whether they were still expressing the original *VSG9* or if this *VSG* had become silent. By qPCR we confirmed that in H1-depleted parasites *VSG9* was no longer transcribed confirming that these mutant parasites had switched VSGs ($P < 0.05$) (Fig. 1D). This included the H1-depleted parasites that generated a single delayed peak of parasitaemia, confirming that this corresponded indeed to a second peak. Overall, these results show that when H1 is depleted, parasite fitness becomes compromised, but parasites are still capable of switching their VSGs.

Even though depletion of H1 significantly disturbs infection *in vivo*, eventually H1-depleted parasites generate a second peak of parasitaemia, which reaches high parasite numbers (10^7-10^9 parasites.ml^{-1}) and results in mice death. Only one exception to this occurred in one (out of 21) mouse infected with clone C2, which never showed high parasitaemia and survived until the end of RNAi induction, 24 days after infection.

3.2.2.2 *In vivo*, H1 depleted parasites adapt and regain fitness

Given that H1 depletion is maintained until the end of infection (Fig. 1C), the observed re-establishment of parasite fitness in H1-depleted parasites could not be explained by the appearance of RNAi revertants. Intrigued by this observation, we asked whether fitness recovery in H1 mutants was underlined by transitory or stable modifications occurring in the parasite. To answer this question, we transferred *T. brucei* H1-depleted parasites from the first peak and the second peak of parasitaemia into naïve mice. Induction of RNAi was kept in these mice by administration of doxycycline in water. We observed that H1-depleted parasites transferred from the first peak replicated more slowly and lead to prolonged mouse survival (Fig. 2A-C), reproducing what was observed previously in an infection initiated by H1-depleted parasites coming from *in vitro* culture (Fig 1.A,B). In contrast, H1-depleted parasites transferred from the second peak rapidly replicated in the blood and lead to mice death within 8-9 days, similarly to the parental PL1S parasites (Fig. 2A,B,D). These results demonstrate that even though H1 depletion disrupts parasite fitness to the

host, ultimately H1-depleted parasites can undergo modifications that are stable and transmissible to the following generations and which compensate for the functions of histone H1 or the adverse effects of its depletion.

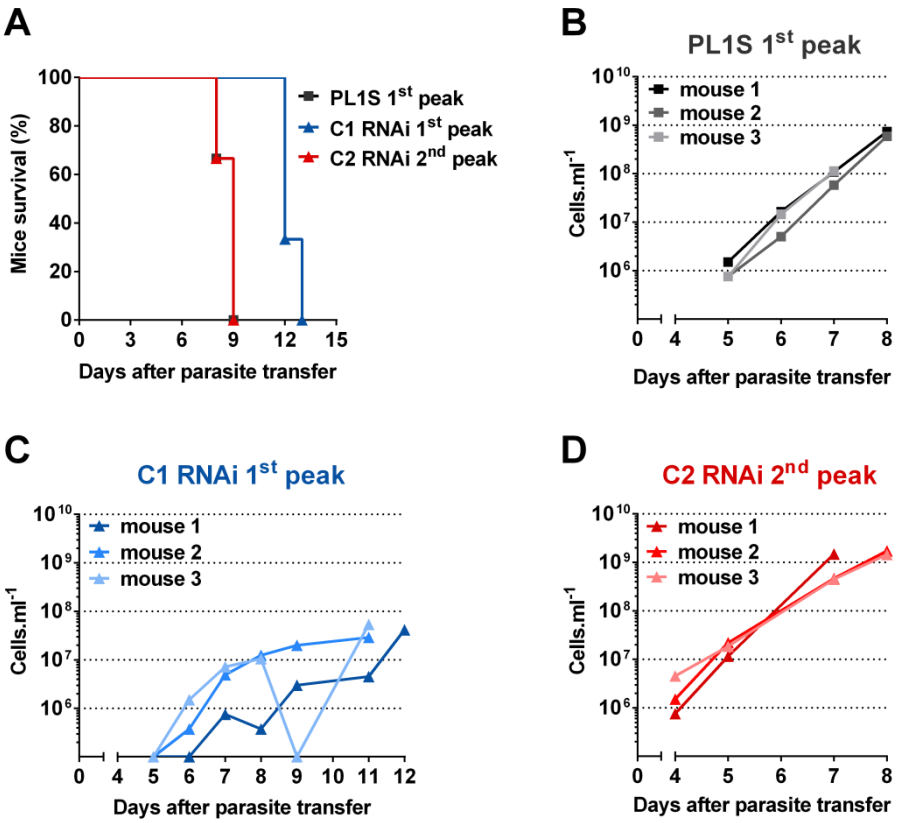


Figure 2. H1-depleted parasites can stably recover their fitness *in vivo*. *T. brucei* parasites at different peaks of parasitaemia were transferred from the blood of infected mice into naïve mice. **(A)** Survival of mice that received 100 parasites from the parental cell-line PL1S from the first peak of parasitaemia, the H1-depleted clone C1 from the first peak of parasitaemia or the H1-depleted clone C2 from the second peak of parasitaemia. One independent experiment was analyzed. Mice n = 3 per group. **(B)** Parasitaemia in the blood of mice that received PL1S parasites from the first peak of parasitaemia, **(C)** C1 parasites from the first peak of parasitaemia and **(D)** C2 parasites from the second peak of parasitaemia. After parasite transfer, PL1S parasites reached a peak in parasitaemia in 7-8 days, while C1 parasites from the first peak maintained a delay of several days in growth. On the contrary, C2 parasites from the second peak restored their replication rates to similar levels of those in PL1S parental cell-line.

3.2.2.3 H1 is important for *T. brucei* ability to evade the host immune system

We observed that appearance of H1-depleted parasites in the blood of mice is delayed up to 10 days relative to the predicted from their doubling times *in vitro* (Fig. 1B; Fig. S2). This suggests that when histone H1 levels are below normal, *T. brucei* parasites are less efficient in adapting to drastic changes in their environment, such as those encountered when entering the mouse host. An obvious difference between culture and mouse is the presence of immune system. Thus we investigated whether the immune system played a role in parasite growth delay. For this, we established infections in $Rag2^{-/-}\gamma c^{-/-}$ mice (also known as $Rag2^{-/-}Il2rg^{-/-}$), a mouse strain in which recombinase activating gene-2 (Rag2) and common cytokine receptor γ chain (γc) were knocked-out resulting in a severe immunodeficiency characterized by the complete absence of T, B and natural killer lymphocytes (Mazurier *et al.*, 1999). As a consequence, these mice have a wide-range of immune responses compromised: lack of adaptive immunity, lack of natural killer-mediated innate immune response and several other innate immunity defects (Shultz *et al.*, 2007).

As expected, due to the absence of major immune responses, PL1S parasites replicate rapidly and lead to mice death up to seven days earlier in $Rag2^{-/-}\gamma c^{-/-}$ mice when compared with wild-type mice ($P < 0.05$) (Fig. 3A). Strikingly, the delay in parasite appearance in the blood and the prolonged survival of mice, distinctive of an infection by H1-depleted parasites, was completely abolished in $Rag2^{-/-}\gamma c^{-/-}$ mice. While in wild-type mice infected with C2 parasites, 50% of the mice survived eight days longer than PL1S-infected mice ($P < 0.01$), $Rag2^{-/-}\gamma c^{-/-}$ mice infected with C2 died within the first 7-8 days of infection, as fast as when infected with PL1S parental parasites (*P*-value non-significant) (Fig. 3A). In agreement with this, parasitaemia in blood and parasite replication rates were identical in C2 and PL1S-infected $Rag2^{-/-}\gamma c^{-/-}$ mice. In these mice, both C2 and PL1S parasites generated only one peak of parasitaemia and replication rates were similar to those of parasites growing *in vitro* (Fig. 3D,E). This result shows that the mouse immune system is a limiting factor for parasite progression under subnormal H1 levels, and contradicts the contribution of other determinants such as parasite-intrinsic factors.

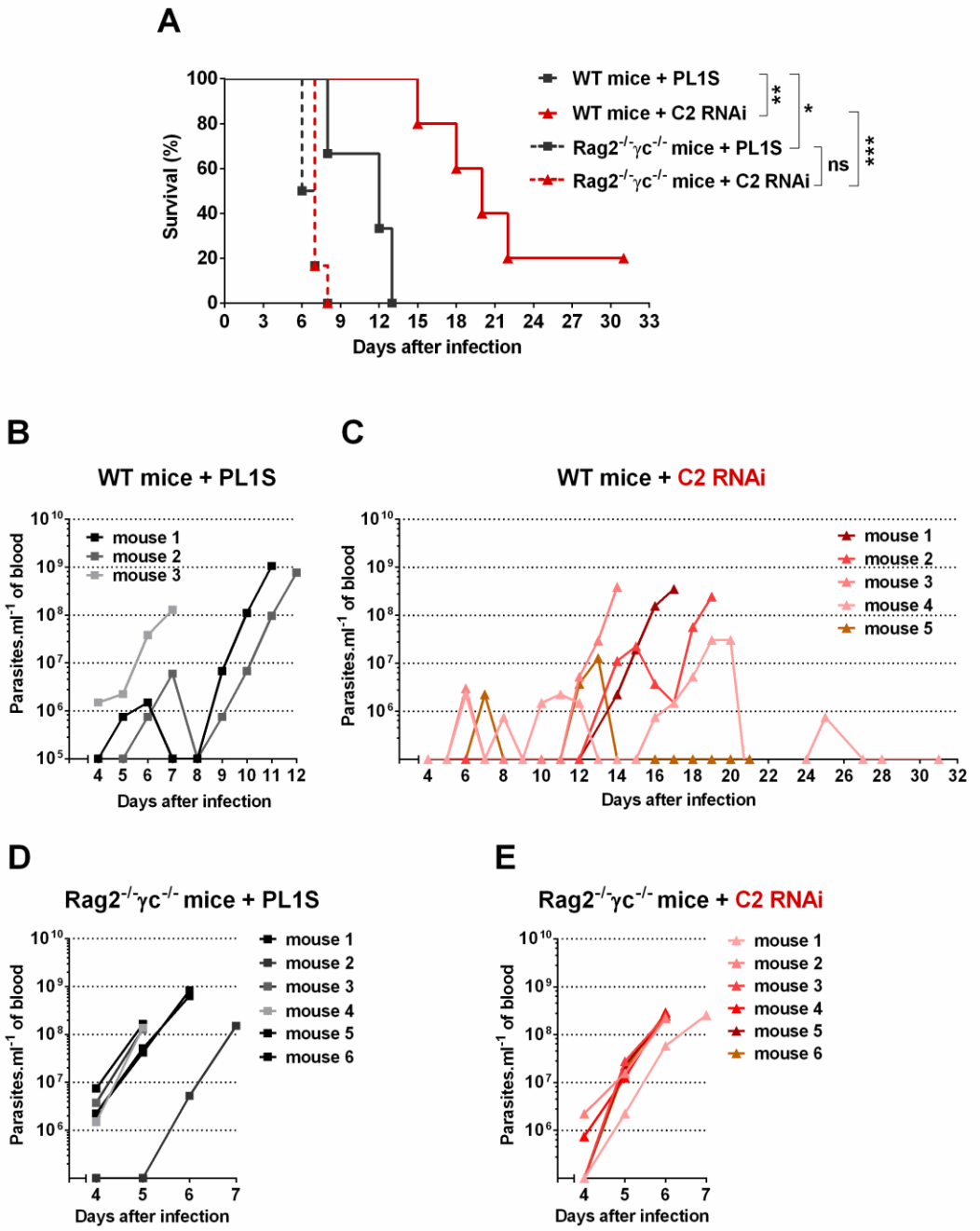


Figure 3. Histone H1 is important for *T. brucei* immune evasion. **(A)** Survival of wild-type and Rag2-/-γc-/- mice infected with 100 parasites of the parental cell-line PL1S or the H1-depleted clone C2. **(B)** Parasitaemia in the blood of wild-type mice infected with PL1S and **(C)** H1-depleted clone C2 parasites. **(D)** Parasitaemia in the blood of Rag2-/-γc-/-mice infected with PL1S and **(E)** H1-depleted clone C2 parasites. In Rag2-/-γc-/-mice the delay in parasitaemia observed between PL1S and C2 parasites is abolished and both parasite cell-lines grow at fast rates. One independent experiment was analyzed. Mice n = 3–6 per group. Survival curves were statistically compared by a Log-rank (Mantel-Cox) test. ns, non-significant P-value. * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$.

As described in Figure 1, infection in wild-type mice resulted in two peaks of parasitaemia in more than 50% of the mice infected either with C2 or PL1S parental cell-line (Fig. 3B,C). Interestingly, one mouse infected with C2 clone even exhibited five peaks of parasitaemia and survived, apparently healthy, for 87 days after which it was sacrificed (doxycycline treatment was stopped at day 31 after infection). It is therefore possible that C2 parasites were actually cleared by the immune system in this mouse.

Altogether, our results clearly demonstrate that the host immune system is a key factor behind disturbance of *T. brucei* fitness and adaptation to the host environment *in vivo*, induced by H1 depletion. Curiously, this implies that H1-depleted parasites are somehow more susceptible to recognition and/or destruction by the host immune system and consequently that histone H1 is important for *T. brucei* immune evasion, even if indirectly.

3.2.3 MATERIALS AND METHODS

MOUSE INFECTIONS

Inbred C57BL/6 wild-type mice (Charles River) and Rag2$^{-/-}\gamma c^{-/-}$ (kind gift of Bruno Silva-Santos laboratory) were housed in the pathogen-free facilities of the Instituto de Medicina Molecular (IMM). The animal facility and the experimental procedures complied with European Guideline 86/609/EC, followed the Federation of European Laboratory Animal Science Associations (FELASA) guidelines and recommendations concerning laboratory animal welfare and were approved by the Instituto de Medicina Molecular Animal Care and Ethics Committee (AEC_2011_006_LF_TBrucei _IMM). Mice were infected intraperitoneally with 100 parasites of *T. brucei* Lister 427 PL1S and RNAi clones C1 and C2 (with and without induction). RNAi induction was initiated in culture with 1 μg ml^{-1} tetracycline 2 days prior to mouse infection. RNAi was maintained *in vivo* in mice by watering them with 1 mg ml−1 of doxycycline (doxycycline hyclate, Sigma). Mice received doxycycline 2 days prior to infection and during the whole course of infection. Parasitaemia was monitored throughout

infection by collecting 1 µl of blood from the mouse tail, diluting it in 1:150 in HMI-11 medium and counting in a Neubauer chamber. In parasite transfer experiments, blood was collected by cardiac puncture from C1- or C2-infected mice, at the 1st peak or 2nd peak of parasitaemia, respectively, in 20 µl of heparin solution (50 UI/ml). Blood was diluted in HM1-11 media to a final concentration of 500 parasites/ml. Each naïve C57BL/6 mouse was infected intraperitoneally with 200 µl of this suspension, containing 100 parasites.

REAL-TIME QUANTITATIVE PCR ANALYSIS

For RNA extraction of parasites in the blood, between day 7 and 14 post-infection, 25–30 µl of blood from the tail of infected mice (parasitaemia ∼108 parasites ml−1 of blood) was collected in red blood cell lysis buffer (150 mM NH4Cl, 10 mM KHCO3, 1 mM EDTA.Na2, pH 7.4), washed in 1× TDB and extracted with PureZOL, according to manufacturer's instructions (Bio-Rad). cDNA synthesis and qPCR were performed as described above (see Results 3.1). Primers are listed in Table S2.

STATISTICAL ANALYSIS

Survival curves were compared by a Log-rank (Mantel-Cox) test. The statistical significance for remaining comparisons was given by Mann–Whitney U tests. A *P*-value < 0.05 was considered significant.

3.2.4 SUPPLEMENTARY FIGURES

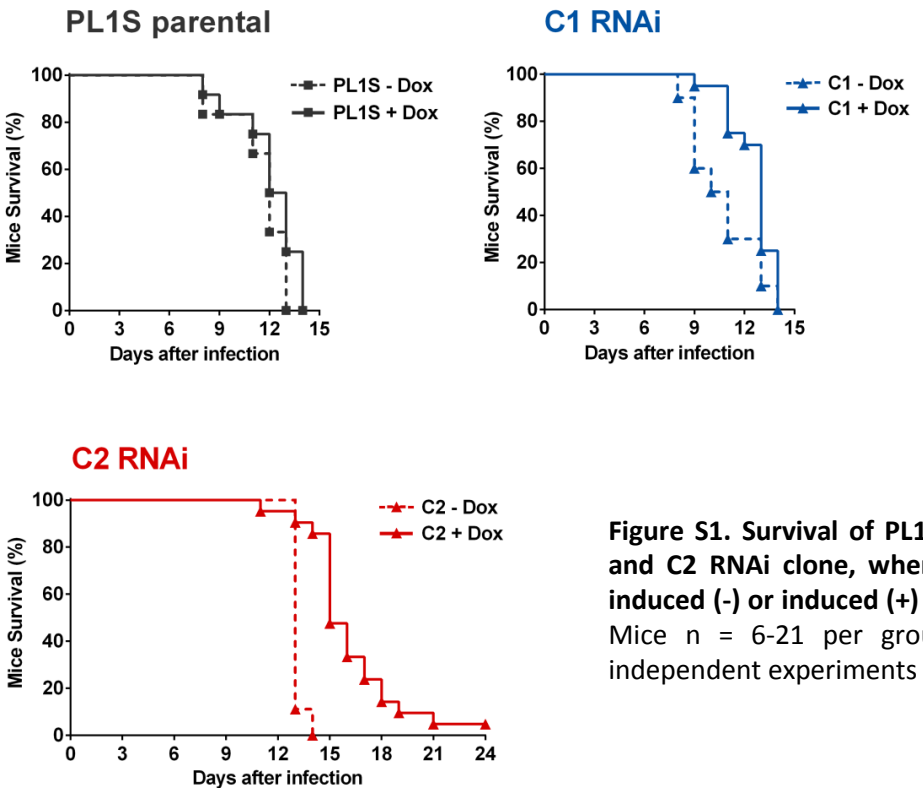


Figure S1. Survival of PL1S, C1 RNAi clone and C2 RNAi clone, when RNAi was non-induced (-) or induced (+) with doxycycline. Mice n = 6-21 per group. Two to four independent experiments were analyzed.

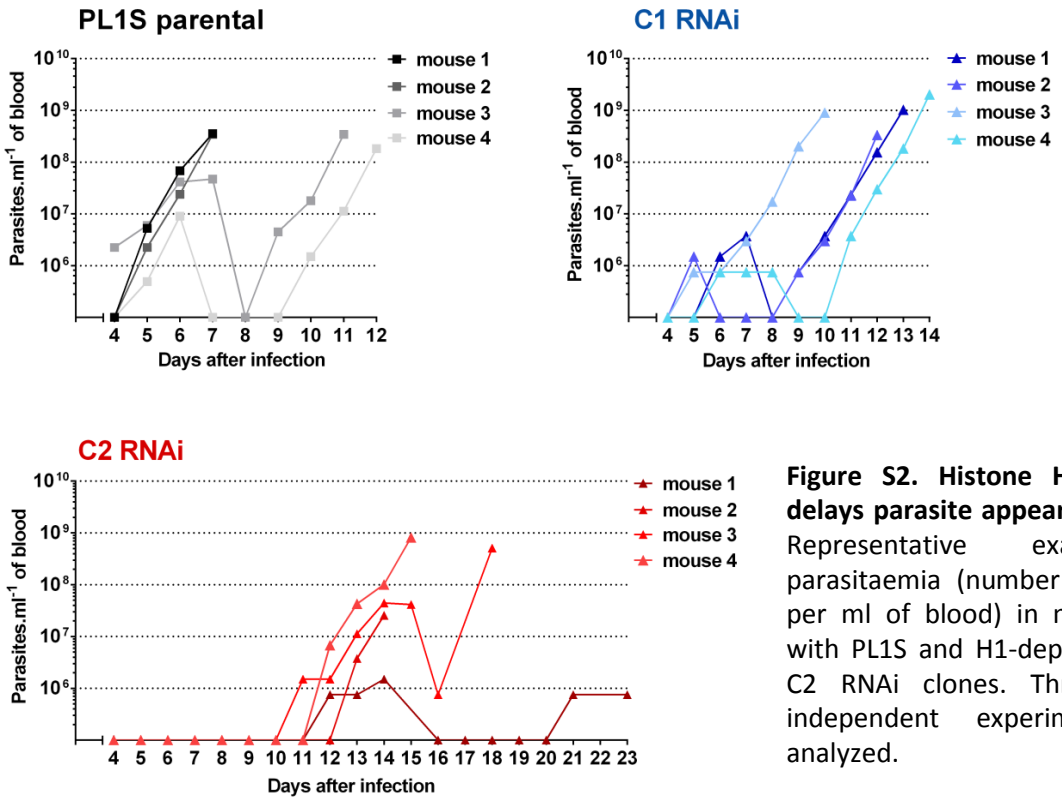


Figure S2. Histone H1 depletion delays parasite appearance *in vivo*. Representative example of parasitaemia (number of parasites per ml of blood) in mice infected with PL1S and H1-depleted C1 and C2 RNAi clones. Three to four independent experiments were analyzed.

3.3 NEW METHOD FOR STUDYING TRANSCRIPTION IN *T. BRUCEI* – IMPLEMENTATION OF 4-THIOURIDINE LABELING OF NASCENT TRANSCRIPTS

Ana C. Pena[1], Rita Vaz-Drago[1], Maria Carmo-Fonseca[1] and Luisa M. Figueiredo[1]

[1]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, Edifício Egas Moniz, 1649-028 Lisboa, Portugal.

3.3.1 AUTHOR CONTRIBUTIONS

Ana C. Pena and Luisa M. Figueiredo designed the experiments. Ana C. Pena performed all the experiments, data analysis and created all the figures and tables. Rita Vaz-Drago contributed with expertise and critical advice and helped in the design and performance of some experiments. Maria Carmo-Fonseca participated in discussion and experimental design of some experiments. Luisa M. Figueiredo mentored and supervised the work.

For decades nuclear run-on has been the gold standard method to analyze and quantify nascent RNAs. Typically, in nuclear run-on assays, nuclei are isolated and transcription is allowed to resume *in vitro* in the presence of radiolabeled nucleosides. Because run-on reaction times are short, only nascent RNAs will be radiolabeled, which are then detected and quantified by hybridization with DNA probes. In *T. brucei,* nuclear run-on assays were instrumental to show for the first time that VSG and procyclins are not transcribed by Pol II (Rudenko *et al.*, 1989, Kooter & Borst, 1984). However, despite recent improvements on its precision and genome coverage (Core *et al.*, 2008), nuclear run-on assays remain conditioned by the fact that transcription is measured *in vitro* and in isolated nuclei. In these conditions we cannot exclude, for instance, perturbations in RNA polymerase activity and deviations from the original *in*

vivo transcriptional status. Indeed, in *T. brucei* nuclear run-on assays, nascent transcript levels can vary significantly with different nuclei preparations (Kooter & Borst, 1984, Rudenko *et al.*, 1989).

In recent years, more sophisticated methods were developed to directly analyze nascent RNAs, based on RNA metabolic labeling. In these methodologies, endogenous RNA is labeled with a non-radioactive modified precursor of uridine-5'-triphosphate (UTP) [e.g., 4-thiouracil (4TU), 4-thiouridine (4sU), 5-etyniluridine and 5'-bromo-uridine] followed by separation of the tagged RNAs by affinity or immunopurification (Perez-Ortin *et al.*, 2013, Tani & Akimitsu, 2012). These new technologies have the advantage of addressing RNA dynamics in unperturbed cells since the uptake of the modified precursor is done though nucleotide or nucleoside cell membrane receptors with no need for nuclei isolation.

4sU labeling can also be used to measure RNA half-lives and decay kinetics without the need of using transcription-blocking drugs, such as ActD. This is because an important property of this method is that newly transcribed (labeled) and pre-existing (unlabeled) RNA fractions can be fractionated and characterized separately by genome-wide sequencing. Thus, contrary to nuclear-run on, newly transcribed RNA can be quantified directly with no interference from steady-state RNA levels. Because the modified UTP precursor will only be incorporated into RNA synthesized *de novo*, the labeled RNA represents a 'snapshot' of the combined effects of RNA transcription and degradation during the timeframe of labeling. Therefore, depending on the duration of labeling and the RNA fraction quantified (total, labeled, unlabeled), mathematical models can be applied to determine RNA transcription rates, RNA half-lives and decay kinetics with whole genome coverage (Tani & Akimitsu, 2012).

4sU is one of the modified nucleotides most used to metabolically label RNA. The technique has been applied in a broad range of organisms including mammals, insects (*Drosophila*), amphibians (*Xenopus*), plants (*Arabidopsis*) and yeast (Dolken *et al.*, 2008, Friedel *et al.*, 2009, Rabani *et al.*, 2011, Sabo *et al.*, 2014, Miller *et al.*, 2011, Sidaway-Lee *et al.*, 2014) 4sU is a naturally occurring modified uridine that offers the advantage of a rapid cellular uptake and RNA incorporation, high yield of 4sU-RNA

isolation and minimal interference in cell growth and mRNA expression in several cell-types (Cleary *et al.*, 2005, Kenzelmann *et al.*, 2007, Dolken *et al.*, 2008, Friedel *et al.*, 2009). In *T. brucei* 4sU is reported to be non toxic up to 1 mM during 72h (Ali *et al.*, 2013).Due to its incorporation efficiency, 4sU has the advantage of enabling very short labeling times (as little as 5 min) (Windhager *et al.*, 2012, Dolken *et al.*, 2008). When sufficiently short, the labeled-RNAs will mostly be inside the nucleus and not subjected to degradation, reflecting active transcription. Besides, short labeling pulses allow high temporal resolution and detection of short-term changes in RNA synthesis and degradation such as those occurring in primary signaling events. Taking advantage of this capacity, Dölken *et al.* identified for the first time a network of short-lived transcripts downregulated during fibroblast response to interferon (Dolken *et al.*, 2008).

Recent studies combined short 4sU-labeling times (10 min) with microarrays or RNA-Seq to determine with high temporal precision the contributions of RNA synthesis and decay to changes in total RNA levels during stress and immune responses in yeast and murine cells, respectively.(Rabani *et al.*, 2011, Miller *et al.*, 2011) The technique also gave a major contribution in disclosing gene sets directly activated or repressed by the transcription factor Myc (Sabo *et al.*, 2014). 4sU-labeling technology also offers the possibility to follow not only RNA transcription and decay but also RNA processing dynamics. Ultra-short and progressively 4sU-labeling times (from 5 to 60 min) combined with RNA-Seq revealed the kinetics of RNA splicing and non-coding RNA processing at nucleotide resolution, in human B-cells (Windhager *et al.*, 2012).

Metabolic labeling displays important advantages including the analysis of unperturbed cells, with higher temporal resolution and sensitivity and the capacity to perform a dynamic transcriptome analysis, distinguishing the contributions of transcription, splicing and RNA decay to total cellular RNA levels. Such a technique could be highly valuable to reveal many aspects of RNA dynamics in *T.brucei* whose study has been technically limited so far.

3.3.2 RESULTS

4sU metabolic labeling starts with 4sU incubation with cells for a given labeling duration. Cells rapidly uptake 4sU and phosphorylate it to 4sU-5'-triphosphate (4sUTP) which is then incorporated into newly transcribed RNA. Following isolation of total cellular RNA, the 4sU-labeled RNA fraction is thiol-specifically biotinylated generating a disulfide bond between biotin and the 4sU-RNA. Total RNA is then separated into labeled (newly transcribed) and unlabeled (pre-existing RNA) RNA fractions of high purity using magnetic-activated cell sorting (MACS) columns loaded with streptavidin-coated magnetic beads. The unlabeled fraction is collected from the flow-through and column washing steps while the labeled RNA is recovered from beads by adding a reducing agent (e.g. dithiothreitol) that cleaves the disulfide bond and releases the newly transcribed RNA from the beads (Fig. 1).

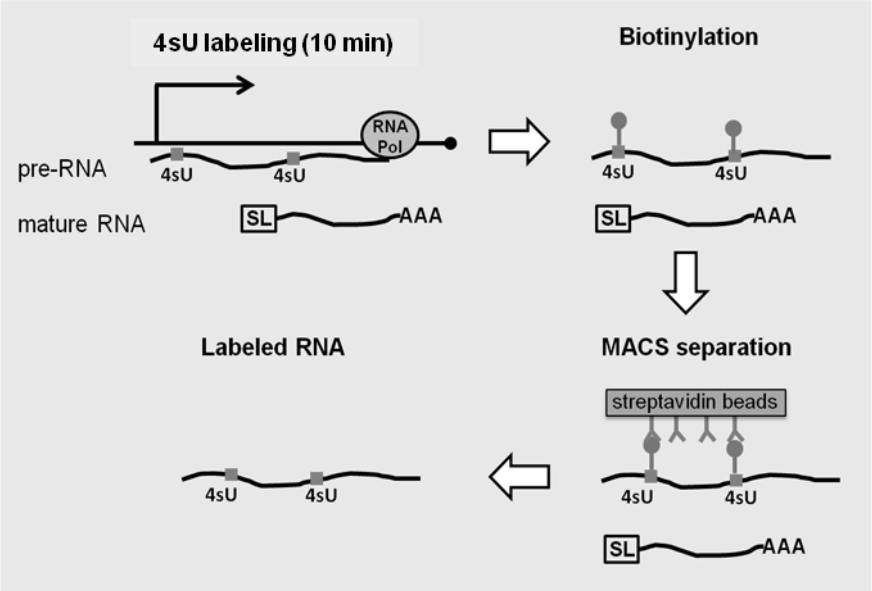


Figure 1. A standard protocol of 4-thiouridine labeling and purification of nascent RNA. Cells are incubated for a short time (e.g. 10 min) with 4-thiouridine (4sU), which is uptaken and incorporated mostly in nascent pre-RNAs These RNAs have not be processed yet into mature RNAs by the addition of the spliced leader (SL) and the polyA tail (AAA). A Pol I-transcribed loci is used as an example. After the labeling reaction, total RNA is extracted and biotinylated. During the biotinylation reaction, biotin molecules binds to thiol group of 4sU residues in a 1:1 stoichiometry. By magnetic-activated cell sorting (MACS) with streptavidin beads, biotin-4sU containing RNA molecules are purified from total RNA. Labeled RNA is then recovered by collecting the flow-through after addition of a reductive agent (e.g. dithiothreitol) that cleaved the disulfide bond between biotin and 4sU.

As explained above (see Results 3.1.2.5, Fig. 8F), we used 4sU metabolic labeling to show that *T. brucei* histone H1 inhibits transcription at BESs and procyclin loci. While testing 4sU-labeling in *T. brucei*, we observed that 4sU-labeled RNA was increasingly enriched in primary transcripts with increasing labeling times and no recovery of 4sU-labeled RNA was possible when transcription was blocked (see Results 3.1.2.5, Fig. 8B, D). These results validated the application of 4sU metabolic labeling to tag and purify newly transcribed RNAs in *T. brucei*.

The efficiency of recovery of nascent RNA in *T. brucei* was still far from that reported in mammalian cell-lines. Whereas in human B-cells, 10 min labeling time with 500 μM of 4sU leads to a recovery of around 1.5 % of total RNA (Windhager *et al.*, 2012), in *T. brucei* we recovered only 0.015–0.025% (even though the labeled RNA levels were clearly over the background corresponding to non-labeled RNA (see Results 3.1.2.5, Fig. 8C). To confirm that our protocol was working, we purified nascent RNA from B-cells. With a 15 min 4sU labeling, we were able to reproducibly obtain 1-2% of nascent RNA, confirming that our reagents and methodology was successful for mammalian cells. The purity and quantity of 4sU-labeled RNA is critical for instance to pursue genome-wide studies and use high-throughput technologies such as RNA-Seq. Bearing this in mind, we did various attempts to improve the yield of 4sU-labeled RNA recovered in *T. brucei.* We systematically tested several changes to the protocol and monitored potential problems that could be limiting the three major steps in the protocol: 4sU labeling; biotinylation; MACS separation. Table 1 summarizes all the attempts to improve the 4sU-labeling efficiency in *T. brucei*.

3.3.2.1 Labeling efficiency

We started by addressing potential problems in the labeling step, which is a result of 4sU uptake and RNA incorporation processes. Because the latter directly depends on 4sU concentration and transcriptional activity of cells (Radle *et al.*, 2013), we tested if labeling with two times more 4sU (1000 μM) increased the yield. Purification and quantifications were performed as described above. We observed no improvement of the yield when more 4sU was used.

Next we tested cell density. In our protocol labeling is performed in a small volume (2 ml) with parasites at a high cell density (~ $0.8x10^7$ cells.ml^{-1}). Because labeling times are very short (2, 5, 10 min), we did not expect major changes in transcription due to cell density. However, because parasites in culture at densities > $1.5x10^6$ cells.ml^{-1} arrest growth and undergo cell death after some hours, some transcriptional inhibition might have be occurring with the parasite densities used. Therefore, we repeated the 4sU labeling and purification protocol with parasites at a normal cell density ($0.9x10^6$ cells.ml^{-1}). We observed that the quantity of 4sU-labeled RNA recovered was not significantly improved with a lower parasite density.

3.3.2.2 4sU transport across the membrane

Next, we hypothesized that 4sU uptake could be limited by inefficient transport across the membrane. Trypanosomes possess both salvage and biosynthesis routes for pyrimidines (de Koning *et al.*, 2005). This means that *T. brucei* parasites can both synthetize *de novo* or highjack the host's nucleobases or nucleosides to produce their own pyrimidine nucleotides. The only pyrimidine transporter identified so far in *T. brucei* bloodstream forms, *T. brucei* uracil transporter 3 (TbU3), uptakes uracil with high affinity (K_m = 0.54 μM) and uridine with much lower affinity (K_m = 9500 μM), salvaging them into its own UTP production (Ali *et al.*, 2013). Since efficiency of uracil transport is greater than uridine, we hypothesized that labeling with 4-thiouracil (4TU) might improve the yield of labeled RNA. Contrary to 4sU, the transport of 4TU was previously studied in *T. brucei* BSFs and shown to occur with lower affinity than uracil but higher than uridine (K_i = 159 μM). Metabolic RNA labeling with 4TU can be achieved essentially following the same basic protocol as for 4sU labeling and it has been applied successfully in mice, *Drosophila* and in the protists *Toxoplasma gondii* and *Karenia brevis* (Gay *et al.*, 2014, Miller *et al.*, 2009, Cleary *et al.*, 2005, Morey & Van Dolah, 2013).

Table 1. Tests performed in order to improve yield of purification/recovery of 4sU-labeled RNA in *T. brucei*.

Protocol step	Test	Conditions	Labeling media	Modified precursor	Labeling time (min)	Number of samples	Results
4sU uptake/ incorporation	4sU	1000 μM 4sU	TDB	4sU	5	*n* = 1	same yield
	Cell density and labeling media	normal cell density	TDB / HMI-11	4sU	10	*n* = 2	same yield
	4TU and labeling media	500uM, 1000 μM 4TU, normal cell density	TDB / HMI-11	4TU	10	*n* = 3	same yield
	Permeabilization	0.4 μM digitonin, 10 min	Transcription buffer	4sU	5, 15, 30	*n* = 3	decreases yield
	Labeling media	HMI-11, high cell density	HMI-11	4sU	10	*n* = 1	same yield
	RNA editing	qPCR for ND7 pre/post-edited; ActD treatment	TDB	4sU	10	*n* = 1	no significant incorporation by RNA editing
Biotinylation	Protein contamination	Proteinase K treatment	HMI-11	4sU / 4TU	10	*n* = 2	same yield
		Coomassie staining	TDB / HMI-11	4sU	10	*n* = 6	no protein detected in RNA sample before biotinylation
		Western blotting anti-VSG	TDB / HMI-11	4sU	10	*n* = 4	no VSG detected in RNA sample before biotinylation
		Bradford	TDB / HMI-11	4sU / 4TU	10	*n* = 7	2-10 μg protein/100 μg RNA; probably inaccurate
	RNA degradation	Agarose gel	TDB / HM1-11	4sU / 4TU	10	*n* = 19	no RNA degradation
MACS separation	RNA degradation	Agarose gel	TDB / HM1-11	4sU / 4TU	2, 5, 10	*n* = 26	RNA appears intact

We tested labeling of *T. brucei* with 4TU at 500 μM and 1000 μM, for 10 min, with parasites at a normal cell density. Again, 4TU labeling did not improve the yield of nascent RNA recovery when compared with 4sU labeling.

Next we tested if cell membrane permeabilization would improve recovery of 4sU-labeled RNAs by increasing 4sU entry into the parasite. Basically, we followed the protocol described for labeling nascent RNAs with 5-bromouridine-5'-triphophate (Br-UTP) for immunofluorescence in *T. brucei* (Navarro & Gull, 2001), with the difference that trypanosomes were permeabilized with digitonin in conditions already tested in our lab. Instead of improving, the levels of 4sU-labeled RNAs were lower for several genes, when compared with the non-permeabilized trypanosomes, probably due to perturbation of cellular homeostasis as a result of membrane impairment.

A possible explanation for the reduced levels of 4sU or 4TU tagging could be competition by other routes of UTP metabolism during 4sU/4TU uptake or incorporation into RNA. During *in vitro* culture *T. brucei* is exposed to extremely nutrient-rich conditions. The HMI-11 media contains aspartate and glutamine, which can be uptaken by the parasite and diverted for *de novo* biosynthesis of UTP. Besides, uracil and uridine present in the serum can enter through TbU3 to the salvage pathway (Ong *et al.*, 2013), most likely dominating over 4sU and 4TU transport. To exclude the effect of these external sources, we decided to compare 4sU and 4TU-labeling of trypanosomes in HMI-11 media with labeling in trypanosome dilution buffer (TDB), a buffer which only contains salts and glucose (Cross, 1975). At normal cell densities labeling in TDB did not make any difference in the yields of both for 4sU or 4TU labeled-RNAs. The same was true for 4sU labeling in TDB at high cell densities.

3.3.2.3 4sU deviation by RNA editing

An additional hypothesis was that RNA editing competes with transcription for the 4sU nucleotide, which would contribute to poor yield of 4sU incorporation into nuclear transcripts. Using pairs of primers that amplify pre-edited and post-edited transcripts of the pan edited mitochondrial NADH dehydrogenase subunit 7 (ND7) gene (Goringer, 2012), we observed that reduction in both *ND7* transcript levels is

similar to nuclear transcripts upon actinomycin D treatment, suggesting that 4sU incorporation is mainly occurring during transcription (Fig. 2).

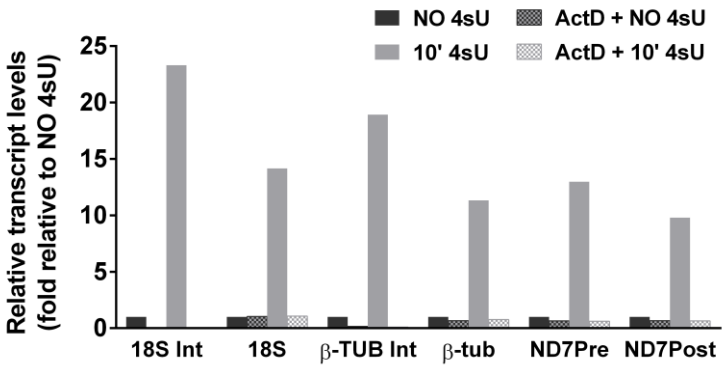


Figure 2. 4sU is mainly incorporated during transcription and not via RNA editing. Prior to 4sU-labeling, transcription was inhibited by actinomycin D (ActD) for 5 min. Relative transcript levels in 4sU-labeled RNA was measured as described before (see. 3.1. Results). Recovery of 4sU-labeled ND7 edited RNA is blocked with ActD, indicating that 4sU incorporation is mostly dependent on transcription and not on editing. One independent experiment was performed. β-tub, β-tubulin gene; β-tub Int, β-tubulin intergenic region; ND7Pre, ND7 RNA pre-edited; ND7Edit, ND7 RNA after editing; 18S, ribosomal 18S gene; 18S Int, 18S intergenic region.

In summary, we were not able to improve the yield of 4sU-labeled RNA despite our efforts to increase 4sU uptake or incorporation. Thus, we next addressed potential limitations in the downstream steps of the protocol.

3.3.2.4 Protein contamination

Biotinylation is carried out by biotin-HPDP Trypanosoma brucei, which adds a biotin residue to a thiol group by forming a disulfide bond between the two. If present, any other molecules that bear a thiol group will potentially compete with 4sU-RNA for the biotinylation reaction and decrease its efficiency. Proteins rich in thiol groups are the most probable contaminants of this kind when extracting RNA. To exclude the possibility of a protein contaminant, we used different methods to test the presence of proteins in total RNA samples prior to biotinylation. First, we analysed samples by sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE), stained by Coomassie Blue or subjected to Western blotting with an anti-VSG antibody. We did not detect any proteins in several RNA samples (*n* = 4-6) with any of the methods.

We further examined the presence of proteins with the Bradford method, a colorimetric assay that can quantify protein concentrations as low as ~1 μg.ml^{-1}. The assay detected 2 – 10 μg of protein / 100 μg RNA in all except one of the tested samples of total RNA before biotinylation (*n* = 7). However, we should be cautious when interpreting the protein levels given by Bradford. First, with all dye-based methods for protein quantification, one should use a protein standard that gives a similar color yield to that of the protein being assayed. Since we ignore which protein contaminants could be present in our RNA samples, we used a regularly used standard, bovine serum albumin, which typically has dye color development greater than most proteins, resulting in overestimation of the protein concentration.

Another important fact is that nucleic acids such as RNA can actually react with Bradford dye with 10-20% of color yields when compared with equivalent amounts of protein (Wenrich & Trumbo, 2012). In fact, that level of interference is in agreement with the 2-10% of protein mass calculated in our RNA samples. Taken together with the fact that staining of SDS-PAGE gel with Coomassie Blue (which in our lab was shown to detect down to 500 ng of protein) was not consistent with Bradford measurements, we consider Bradford quantifications as inaccurate, not confirming the existence of protein contamination in total RNA samples. In agreement with these results, treating RNA before biotinylation with proteinase K did not ameliorate the final yields of labeled RNA, labeled either with 4sU or 4TU. Overall these results suggest that no significant contamination of RNA with proteins was precluding biotinylation reaction.

3.3.2.5 RNA integrity

Finally, we tested if RNA was being degraded during the protocol. We monitored RNA integrity by agarose gel electrophoresis in RNA samples taken at the three major steps of the protocol: Total RNA, biotinylated RNA and MACS-separated RNA. In all the tested samples of total and biotinylated RNA (*n* = 19), the typical RNA band profile of *T. brucei* was visible. The three bands between 1-2 kb corresponding to *T. brucei* rRNAs 18S, 28Sα and 28Sβ, of decreasing molecular weights respectively,

were well-defined and represented the most abundant RNAs, as expected for intact RNA moieties. Less abundant transcripts of lower molecular weight corresponding to mRNAs and tRNAs were also visible (Fig. 3A) indicating RNA sample was of very good quality. Biotinylated RNA from B-cells processed in parallel with *T. brucei* RNA also displayed no considerable degradation, exhibiting the two intense bands of 28S and 18S rRNA, together with mRNAs and tRNAs (Fig. 3A).

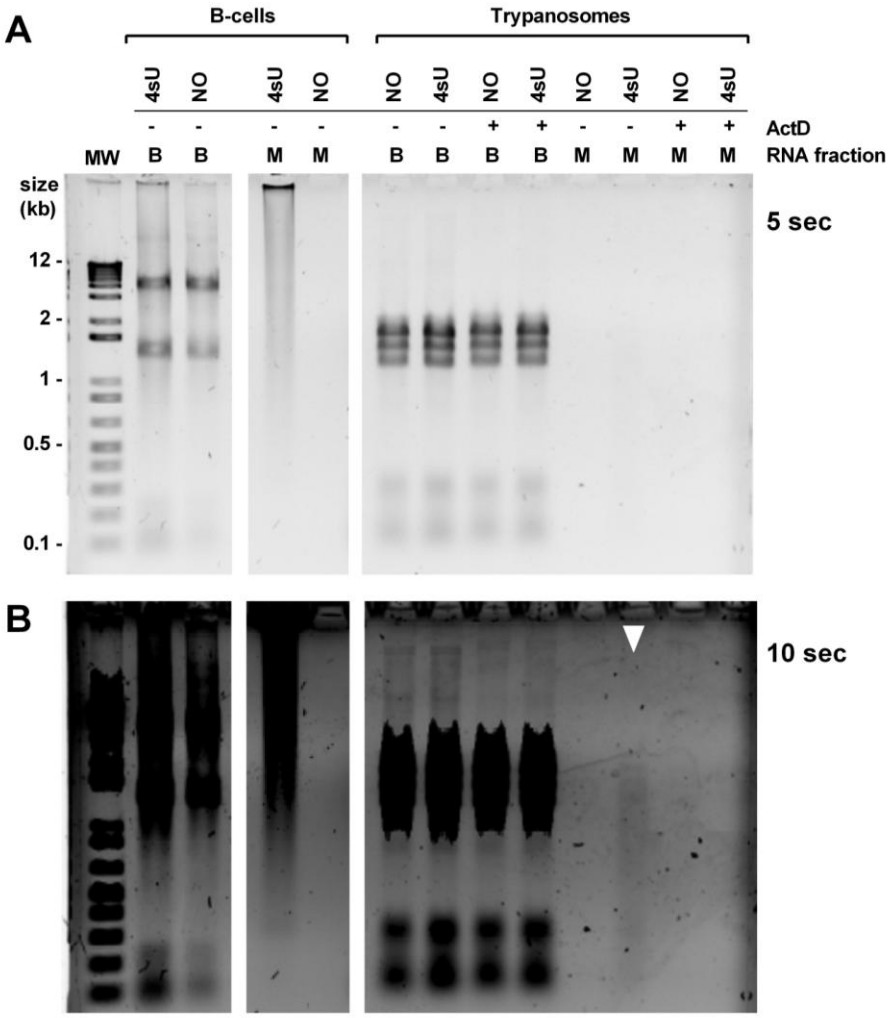


Figure 3. No RNA degradation occurred during 4sU-labeling protocol. Agarose gel electrophoresis analysis of RNA samples after biotinylation and separation by MACS. RNA was run in a native 1% agarose gel, stained with GelRed™ and detected under UV light. **(A)** RNA detection after 5 sec (B) and 10 sec of UV exposure. RNA from B-cells and trypanosomes not labeled (NO) and labeled with 4sU is shown; RNA fractions corresponding to total biotinylated RNA **(B)** and separated by MACS (M) were analyzed as well as RNA from cells (-) not treated or (+) treated with ActD 5 min prior to 4sU labeling. White arrow indicates the smear of nascent transcripts from *T. brucei.*

After purification by MACS, agarose gel electrophoresis was not sensitive enough to detect RNA in most of the samples (*n* = 26). However, in one of the experiments, a smear of nascent transcripts could be detected with prolonged UV exposure (Fig. 3B). The smear of newly transcribed RNAs contained significantly less mature rRNAs and transcripts in a higher range of molecular weights, as reported before (Radle *et al.*, 2013). As expected, nascent RNA was present in the purified RNA from 4sU-labeled trypanosomes but not in the RNA of trypanosomes not exposed to 4sU or where transcription had been blocked with ActD (Fig. 3B). Nascent transcripts purified from B-cells also distributed as a smear, validating the interpretation done for *T. brucei* RNA (Fig. 3A). However, the abundance of nascent RNAs is greater in B-cells than in trypanosomes, in agreement with the low yields of 4sU-labeled RNA in *T. brucei*. Altogether, the electrophoretic RNA profiles indicate that no significant RNA degradation occurred throughout the entire protocol of 4sU labeling.

In summary, despite our efforts we did not succeed in improving the recovery yield of 4sU labeled nascent RNA. However we can exclude the influence of the following factors: inefficient labeling due to low 4sU concentration, high cell-density or labeling media; poor 4sU transport across the membrane; deviation of 4sU by RNA editing; protein contaminants or RNA degradation throughout the protocol.

3.3.3 MATERIALS AND METHODS

CELL-LINES AND 4_SU-LABELING CONDITIONS

T. brucei BSF parasites were cultured as described before (see 3.1.3 Materials and Methods). 4sU-labeling and nascent RNA purification from B lymphocytes was used as a positive control of the protocol. B lymphocytes (immortalized lymphoblastoid cell-line GM16113; NIGMS Human Genetic Cell Repository, Coriell Institute for Medical Research) were cultured in RPMI 1640 medium (Gibco) supplemented with 18% heat-inactivated serum, 2mM non-essential amino acid solution (Gibco) and 2mM L-Glutamin (Gibco) at 37°C in 5% CO2, at a concentration of 1 x 10^6 cells/ml.

Metabolic labeling was essentially performed as described before (see 3.1.3. Materials and Methods) but different labeling conditions were tested: 4sU at a higher concentration of 1000 µM; labeling with 500 µM or 1000 µM of 4-thiouracil (4TU) (Sigma); labeling reactions carried out in HMI-11 media (Hirumi & Hirumi, 1989); labeling reaction performed at normal parasite densities of ~1x10^6 cells/ml in 120 ml of TDB (5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20mM Na_2HPO_4, 2 mM NaH_2PO_4, 20 mM glucose, pH 7.7) or HMI-11 media. In these reactions no TDB washes prior to labeling were done and 4sU or 4TU were added directly to parasites resuspended in TDB or parasites cultured in HM1-11 media. 4sU-labeling reaction in permeabilized parasites is described below.

4_SU-LABELING IN PERMEABILIZED PARASITES

For testing 4sU labeling in permeabilized cells we adapted a previously established protocol for labeling nascent RNA with Br-UTP in *T. brucei* (Navarro & Gull, 2001). 1 x 10^8 BSF parasites were collected as described before (see 3.1.3. Materials and Methods). After the first TDB wash, cells were resuspended in 50 ml of TDB with 0.4 µM digitonin and incubated at RT for 10 min. Parasites were collected by centrifugation at 970 *g* for 10 min at RT, washed in 2 ml of TDB and carefully resuspended in 2 ml of transcription buffer (50 mM HEPES pH7.9, 100 mM KCl 5 mM $MgCl_2$ and 0.5 mM EGTA; 5 mM DTT, 1x protease inhibitor cocktail (P8340, Sigma) and 80 U/ml RNAse inhibitor (Thermo Fischer Scientific)) to a final concentration of 5x10^7 cells/ml. Labeling reaction was started by adding a mix of ribonucleotide 5'-triphosphates to the following final concentrations: ATP, 2000 µM; CTP, 1000 µM; GTP, 1000 µM (Thermo Fish Scientific) and 4sU 500 µM (Sigma). To the non-permeabilized cell controls no rNTPs was added except 4sU. Labeling reaction was performed in transcription buffer with rNTPs at 33°C for 5, 15 or 30 min, after which cells were collected by centrifugation at 2500*g* for 5 min at 4°C and resuspended in 3 ml of TRIzol reagent (Thermo Fischer Scientific).

PROTEINASE K TREATMENT OF TOTAL RNA

1.4 x 10^8 BSFs were labeled with 500 µM 4sU or 1000 µM 4TU for 10 min at 37°C at normal cell densities (~1.2x10^6 cells/ml) in HMI-11. No TDB washes were performed and 4sU or 4TU were added directly to parasites in HMI-11. Cells were collected, resuspended in TRIzol and total RNA was extracted as described previously (see 3.1.3. Materials and Methods). Total RNA (91 µg/sample) was treated with 200 µg/ml proteinase K (Fischer Scientific) in TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) with 0.5% SDS in a final volume of 400 µl, for 2h at 37°C. Afterwards, RNA was extracted by a standard phenol:chloroform:isoamylic alcohol.extraction, followed by precipitation with 1:10 volume of 5 M NaCl and 1x volume of isopropanol and dissolution in TE. Biotinylation and capture of nascent RNAs by MACS was subsequently performed as already described (see 3.1.3. Materials and Methods).

DETECTION OF PROTEIN CONTAMINANTS BY WESTERN BLOTTING

To detect eventual contamination of RNA samples with VSG protein, total RNA samples (20-30 µg/sample) were suspended in 1x Laemmeli buffer at a final volume of 30 µl, boiled for 10 min and loaded onto a Mini-Protean TGX pre-cast PAGE gels (BioRad). SDS-PAGE and transfer to a nitrocellulose membrane was performed as described previously (see 3.1.3. Materials and Methods). Before incubation with antibodies, the nitrocellulose membrane was stained for 30 min at RT in a Coomassie Blue solution (40% distilled water, 10% acetic acid, 50% methanol, 0.25% by weight Coomassie Brilliant Blue R-250) and destained for 3:30 h in destaining solution (60% distilled water, 10% acetic acid, 40% methanol). Nitrocellulose membrane was blocked for 1 h at RT with PBS/0.1% Tween 20/5% milk, probed overnight at 4°C with the primary antibody, a rabbit polyclonal antibody anti-VSG1.13 not CRD-depleted (kindly provided by George Cross) 1:1000 in PBS/0.1% Tween20/3% milk. CRD stands for cross-reactive determinant and comprises a region of the C-terminal domain of VSG which displays immunological cross-reactivity. Antisera not CRD-depleted contains antibodies that recognize CRD and cross-react with several VSG variants (Barbet & McGuire, 1978). After washing the membrane 3x with PBS/0.1% Tween20 at RT, it was incubated with

secondary antibody donkey anti-rabbit-HRP (GE Healthcare Life Sciences; NA934) 1:10 000 in PBS/0.1% Tween20/3% milk for 1 h at RT. After washing 3x in washing buffer the blot was developed, protected from light, with Western Lightening ECL Plus (PerkinElmer; sensitivity of 1-10 pg of protein). Light emission was detected in a ChemiDoc XRS+ System (BioRad) for 5-30 min and afterwards in an X-ray film with 1 h of exposure.

BRADFORD QUANTIFICATION OF PROTEINS

Protein Assay Dye Reagent Concentrate (BioRad) was added to total RNA samples (1:4) or to bovine serum albumin standards (1:4) to reach a final volume of 50 µl. Samples were incubated for 5-10 min at RT, after which absorbance was read at 595 nm in 2 µl per sample in a NanoDrop2000 (Thermo Scientific). A standard curve for protein concentration was calculated using serial dilutions of bovine serum albumin prepared in TE buffer, the same solvent of RNA samples. Linear range of detection determined by standard curve was ~8-100 ng/µl of protein concentration.

ANALYSIS OF RNA INTEGRITY

To examine whether RNA degradation was occuring at any step of 4sU-labeling protocol, several samples of total, biotinylated and MACS-separated RNA were loaded into 1% agarose gels prepared in TBE buffer 0.5x (45 mM Tris-borate, 1 mM EDTA pH 8.0 in RNAse-free water) and run in the same buffer. Each sample loaded contained ~100-150 ng of RNA according to the quantity measured by NanoDrop 2000 (Thermo Scientific). Nucleic acids were detected by gel post-staining gel with GelRed (Biotium) for 20 min at RT and detected after UV exposure for 5-10 sec in ChemiDoc XRS+ System (BioRad).

REAL-TIME QUANTITATIVE PCR ANALYSIS OF RNA EDITED TRANSCRIPTS

To assess wether 4sU was being significantly incorporated via RNA editing we compared the levels of pre-edited and post-edited transcripts of *ND7,* measured by qPCR, before and after transcription inhibition in 4sU-labeled parasites. Transcription inhibition by ActD (Sigma) and 4sU-labeling (10 min) was done as described before (see 3.1.3. Materials and Methods). RNA extraction, cDNA synthesis and qPCR conditions were reported in a previous section (see 3.1.3. Materials and Methods). Transcript levels were normalized to NO 4sU control and are plotted as fold-change relative to this sample. An equal amount of RNA was used to synthesize cDNA and therefore no additional normalization was applied. Primers used to amplify pre-edited and after editing transcripts are shown in Table S2 in section 3.1.4 Supplementary figures and tables.

4 GENERAL DISCUSSION

4.1 H1 CONDENSES CHROMATIN GLOBALLY BUT HAS A MORE PRONOUNCED EFFECT AT SILENT POL I TRANSCRIPTION UNITS

The first section of results of my thesis work (see Results 3.1.) can be summarized in the following conclusions: (i) in *T.brucei* BSFs, H1 is necessary for widespread chromatin compaction with special impact at the chromatin of silent BESs promoters; (ii) H1 is not a global regulator of gene expression in *T. brucei* and (iii) H1 predominantly represses transcription of silent Pol I genes.

In this thesis, using FAIRE and ChIP of H3, we showed that depletion of H1 resulted in a global increase of chromatin accessibility (2.9-fold) and a 36% reduction of histone H3 in most loci, suggesting that H1 maintains chromatin compacted throughout the genome (Results 3.1., Figure 4A,B). This was expected in the context of what is known about the importance of H1 to global chromatin compaction in several eukaryotes (Barra *et al.*, 2000, Fan *et al.*, 2005, Masina *et al.*, 2007, Hashimoto *et al.*, 2010) and in *T. brucei* (Povelones *et al.*, 2012, Burri *et al.*, 1995). Indeed, in mammalian embryonic cells, a ~50% reduction in H1 levels leads to a general decondensation of polynucleosome fibers and reduction in nucleosome repeat length (Fan *et al.*, 2005). In agreement, in the kinetoplastid *L. major*, in which H1 histones are structurally similar to *T. brucei*, overexpression of H1 globally compacts chromatin (Masina *et al.*, 2007). Importantly, our data is consistent with (Povelones *et al.*, 2012*)* who showed that in *T. brucei* H1 depletion increases chromatin accessibility and reduces putative heterochromatin domains dispersed in the nucleus (Povelones *et al.*, 2012).

In our observations we detected loci in which chromatin was not significantly affected by H1 depletion: in the actively transcribed *VSG* and BSD^R genes (BES1) and rDNA *18S* gene, which probably already have a maximal chromatin decompaction and therefore chromatin remained essentially open in the presence of normal or reduced levels of H1. The most important exception to this general trend was the chromatin of silent BESs promoters, which became 10-fold more open and lost more than 60% of histone H3. Procyclin promoter regions showed an average decompaction by FAIRE,

but a significant drop in H3 content (48%), suggesting that the chromatin of these loci is also relatively sensitive to H1 loss. A similar trend at silent BESs promoters was observed by (Povelones *et al.*, 2012) using an MNase sensitivity assay which, similarly to FAIRE, addresses chromatin accessibility. ChIP also showed that silent BES promoters have more H1 than procyclin promoters (Povelones *et al.*, 2012), which may explain why depletion of H1 results in a more dramatic chromatin opening at the silent BESs than procyclin promoter regions.

The use of complementary methods for assessing chromatin condensation (ChIP and FAIRE) allowed us to corroborate former evidence and detect further changes in chromatin condensation with higher sensitivity. In particular, FAIRE enriches for DNA not bound to proteins, and consequently for nucleosome-depleted regions of chromatin (Giresi *et al.*, 2007). This obviates a limitation existent in the MNase sensitivity approach followed by Povelones *et al*. (2012), which quantifies mononucleosome DNA and does not determine the levels of DNA free of nucleosomes. This is probably one of the reasons why we were able to detect significant chromatin opening at several loci throughout the genome by FAIRE while Povelones *et al.* did not. Overall, our data solidly allows us to conclude that although *T. brucei* H1 has a global role in chromatin condensation, there are genomic sites, such as the BES promoter, in which H1 role seems to be more important or less redundant.

Consistent with an important function of H1 in the chromatin of BESs and somewhat procyclin promoter regions, in our RNA-Seq study almost all derepressed genes identified are from Pol I loci (*VSG*s, *EP* procyclin genes and *PAGs*) (Results 3.1., Figure 5C). To test if H1 acts at the transcriptional level, we chose to measure transcription rates by metabolically labeling nascent transcripts with 4sU nucleoside analog. The advantage of this approach over the classic nuclear run-on is that the assay is done in unperturbed cells, instead of permeabilized cells or isolated nuclei. Despite a low yield of recovery, we clearly show that a 4sU-labeled RNA sample is enriched in molecules of RNA that contain intergenic sequences, which are present in primary unprocessed transcripts (which we called nascent transcripts) (Results 3.1., Fig. 8B,C). Moreover, by blocking transcription with ActD we showed that the labeling is

dependent on transcription, suggesting that the transcript levels in 4sU-labeled RNAs are a reflection of transcription rates of these genes (Results 3.1, Fig. 8D).

We observed that depletion of H1 apparently leads to an increased transcription rate of silent procyclin loci and BESs, but no changes in transcription of Pol II loci (Results 3.1., Figure 8F). Because we detected BESs derepression with a pair of primers located immediately downstream of the silent BESs and procyclin promoters, we can conclude that probably H1 acts, at least in part, as an inhibitor of Pol I transcription initiation. In addition, we observed a stronger increase in nascent transcripts at silent BESs promoters when compared with proyclin loci promoters (Results 3.1., Fig. 8F). This suggests that H1 is particularly important for repressing transcription initiation at silent BESs, contributing to a less extent to transcriptional inhibition at procyclin loci. This agrees with our observations of a more modest impact of H1 depletion in chromatin decondensation at procyclin loci (Results 3.1., Fig. 3B). Also in support of this, it was shown by ChIP that chromatin of silenced procyclin promoters in BSFs and fully activated procyclin promoters in PFs contain comparable levels of H1, while H1 markedly binds more to silent BESs promoters than active BES (Povelones *et al.*, 2012).

Based on our results with 4sU-metabolic labeling, we propose a role for histone H1 in the inhibition of Pol I transcription initiation at silent BES promoters. Our results are consistent with a recently published work showing that BES monoallelic transcription is controlled at least partially at the level of transcription initiation, which is dependent on the presence of the CITFA transcription factor complex (Nguyen *et al.*, 2014). Additional evidence for our hypothesis could be obtained by determining Pol I occupancy at BES promoters before and after H1 depletion. We did several attempts to determine Pol I occupancy at promoter regions using ChIP in cell-lines where different Pol I subunit genes were tagged with different tags (TY1::RPB6z, HA::RPB6z or TY1::RPB5z), but for unknown reasons we were never able to consistently immunoprecipitate more DNA from Pol I loci than Pol II loci. TY1 and HA epitopes could be too small and/or become inaccessible to antibodies when embedded in the folded protein or in the Pol I complex. In fact, tandem affinity purification (Nguyen *et al.*, 2007) and all the reported ChIPs of RPB6z so far used cell-lines in which RPB6z was N-

terminally tagged with a PTP tag (Park *et al.*, 2011, Nguyen *et al.*, 2014), which bears two epitope domains and a protease cleavage site and is considerably larger than TY1 or HA.

Collectively our data indicates that H1 controls to some extent transcription initiation at Pol I promoters. To our knowledge this is the first mechanistic evidence in *T. brucei* on how depletion of a chromatin component results in an increase of steady-state transcripts by increasing transcription rate, particularly at BESs. It is interesting to notice that, upon H1 depletion, the FAIRE analysis revealed that chromatin of a silent BES opened 10-fold at the promoter, 7-fold at the luciferase gene (1.2 kb downstream of BES promoter) and 2.6-fold at *VSG2* (56.4 kb downstream BES promoter) (Results 3.1., Fig. 4B). Such gradual decreasing effect of chromatin decompaction with increasing proximity to telomeres was also detected by H3 ChIP and it suggests that changes in chromatin are not homogeneous throughout the entire BES. These chromatin changes are also reflected in the steady-state RNA levels, since by RNA-Seq it was detected that the luciferase gene was more derepressed than *VSG2*, which is at the telomeric end of the same BES (Results 3.1., Fig. 5C).

Such disparities of derepression at the beginning and end of a silent BES have been observed in other mutants for chromatin-associated factors. Namely, in mutants for histone H3, histone chaperones TbFACT (TbSPT16), TbNLP, TbASF1A and TbCAF-1b and histone deacetylase TbDAC3 the BES promoter region is derepressed, but not *VSGs* (Denninger & Rudenko, 2014, Alsford *et al.*, 2012, Denninger *et al.*, 2010, Wang *et al.*, 2010, Hughes *et al.*, 2007, Narayanan *et al.*, 2011). In contrast, depletion of TbRAP1 or deletion of TbMCM-BP results in a stronger derepression of *VSG* genes rather than silent BES promoters (Yang *et al.*, 2009, Kim *et al.*, 2013). Although it is not clear how these regulate *VSG* expression, they fit the current view that BES transcriptional silencing is the net result of two forces: i) promoter-proximal inhibition of Pol I transcription initiation/elongation; ii) a repressive gradient emanating from telomeres which prevents transcription elongation and/or RNA processing (Gunzl *et al.*, 2014). In light of this view, we postulate that histone H1 helps maintaining a compact chromatin structure throughout silent BESs, but with a predominant role at the promoter regions and transcription initiation. When H1 is depleted, chromatin of BES promoter region

opens dramatically, which probably facilitates the recruitment of transcription factors and Pol I machinery. At the telomeric end of BESs, H1 depletion also leads to opening of chromatin but not so dramatically, thus contributing at a lower extent to inhibition of transcription elongation and, perhaps, RNA splicing or polyadenylation.

Intriguingly, 9 out of the 25 upregulated loci in H1-depleted trypanosomes correspond to *VSG* sequences outside BESs: a metacyclic *VSG* (Tb427-397 (Cross *et al.*, 2014)), three complete *VSGs* located at arrays (Tb437-631/636/655; (Cross *et al.*, 2014)) and five partial *VSG* sequences probably located at arrays. Since no promoters were ever found at *VSG* arrays or minichromosomes, we ignore how *VSG* sequences located at these loci are transcribed. Derepression of a minichromosomal *VSG* was reported in parallel with derepression of silent BES-linked *VSGs* and procyclin genes in a TbMCM-BP knock-out strain, but no explanation for this was put forward by the authors (Kim *et al.*, 2013). Transcription also increases significantly in minichromosomes and *VSG* arrays in TbNLP and TbISWI depleted mutants when a BES promoter was inserted at these loci (Stanne *et al.*, 2011, Narayanan *et al.*, 2011). This suggests that these *VSG* loci, which are normally regarded as transcriptionally inactive can be transcribed when its chromatin is perturbed.

A possible explanation for our observations is an increase in pervasive transcription by Pol II. Pol II pervasively transcribes a large part of the non-coding genome in yeast and mammals (Jensen *et al.*, 2013). Assembly of pre-initiation complexes and recruitment of Pol II to cryptic transcription origins ('cryptic promoters') occurs typically at nucleosome-depleted regions. Pol II cryptic transcription increases when nucleosome occupancy is reduced or when nucleosome position is disturbed (Hennig & Fischer, 2013). This happens, for instance, in yeast mutants for chromatin proteins such as histone H3, FACT and ISWI (Kaplan *et al.*, 2003, Cheung *et al.*, 2008).

In *T. brucei*, the promoter-independent activity of Pol II and potentially higher transcription initiation at higher genome accessibility levels (McAndrew *et al.*, 1998) argues in favor of its DNA template unspecificity. Therefore, we hypothesize that a Pol II cryptic promoter-like activity increases upon chromatin opening at *VSG* arrays when

H1 content is reduced. If such hypothesis is correct, we expect that in H1 RNAi clones chromatin decondenses at *VSG* arrays. Besides, transcripts from other inactive and/or non-coding loci, such as intergenic regions or transposable elements, might also increase. Moreover, if Pol II is the polymerase involved, then the levels of cryptic transcripts should be sensitive to Pol II inhibitors and novel sites of transcription initiation may be present.

Silent BESs have high levels of H1 when compared with active Pol I loci such as rDNA or active BES (Povelones *et al.*, 2012). On the other hand, active BES is highly depleted of H1, in agreement with the observation that active BES is very nucleosome-depleted relative to silent BESs (Stanne & Rudenko, 2010, Figueiredo & Cross, 2010). Altogether, it seems clear that H1 is essential to establish a compacted chromatin at silent BESs and, conversely, its absence or reduced levels is most probably necessary to maintain an open and transcriptionally active chromatin at the active BES.

Therefore, and important question arises: How is histone H1 recruited to silent BESs and how is it kept away from the active BES? It is tempting to think that one of the factors influencing such distribution of H1 is TDP-1. TDP-1 is an HMGB protein enriched at active Pol I loci (rDNA and BES), particularly at promoter proximal regions TDP-1 distribution is inversely correlated with histone occupancy, including histone H1, and when it is depleted, chromatin at active BES and rDNA condenses and transcription is downregulated (Narayanan & Rudenko, 2013). Therefore, the presence of TDP-1 seems essential for excluding nucleosomes and maintain an open chromatin conformation at Pol I loci which facilitates transcription (Narayanan & Rudenko, 2013). In metazoans H1 and HMGB bind to linker DNA and the exchange of H1 by HMGB proteins has been shown to drive opposite chromatin transitions, with H1 associated with a compacted chromatin and its replacement by HMGB1 leading to a more open chromatin (Travers, 2003, Nightingale *et al.*, 1996).

We hypothesize that in *T. brucei* H1 and TDP-1 act in a similar fashion. Perhaps the inter-change between H1 and TDP-1 is one of the first steps in the establishment of opposed chromatin structures associated with BES activation and silencing. Together with HMGB proteins, the highly dynamic nature of histone H1 chromatin binding

(Misteli *et al.*, 2000, Lever *et al.*, 2000) might have made of this histone the perfect choice to direct the rapid transcriptional changes that are essential for *T. brucei* survival such as during and *in situ* switch of BES in BSFs or BES shutdown and procyclin loci full activation during differentiation to PFs. In this sense, H1 would act as an important ruler of facultative heterochromatin at silent BESs and procyclin loci.

One can argue that the structural divergence between metazoan H1 and *T. brucei* H1, which lacks a globular domain, stands against the proposal that H1 dynamics shares similarities between both organisms. However, remarkably, several mutational analysis of the C-terminal domain of metazoan H1 have shown that this domain is what basically determined the dynamics and high-affinity binding of H1 to chromatin (Hendzel *et al.*, 2004). Noticeably, it was recently shown in human cells that the C-terminal domain of H1 is the one functionally implied in the interaction with important gene expression regulators. C-terminal of H1 physically interacts with tumor suppressor PTEN and mediates the global chromatin condensation pattern thought to underlie the gene expression program directed by PTEN (Chen *et al.*, 2014). Because *T. brucei* H1 essentially resembles the basic, lysine rich C-terminal of higher eukaryotes, it seems reasonable to assume that it retains, in general, comparable binding kinetics and protein-interaction abilities. In support of this hypothesis, the heterologous addition of H1 from *T. brucei* and *L. major* recapitulates endogenous activity of mammalian H1 in chromatin condensation and cell cycle regulation (Smirlis *et al.*, 2006, Burri *et al.*, 1995)

Remarkably, in human cells, displacement of histone H1 is intrinsically necessary for the first steps of promoter transcriptional activation during hormone stimulation (Vicent *et al.*, 2011). Such displacement happens extremely fast, in the first minute of stimulation and is preceded by the coordinated recruitment of several chromatin remodeling complexes, histone modifying enzymes and kinases, and accompanied by H1 phosphorylation and other PTMs at core histones (Vicent *et al.*, 2011). Perhaps a similar multistep mechanism operates at BESs where displacement or deposition of histone H1 at promoters necessarily precludes activation or silencing of transcription, respectively.

We can speculate about the involvement in this process of some chromatin remodelers and histone chaperones already characterized in *T. brucei*. In *Drosophila*, the chromatin remodeler ISWI appears essential for the specific assembly of H1 into chromatin *in vivo* and its activity in global chromatin compaction (Corona *et al.*, 2007, Siriaco *et al.*, 2009). Reciprocally, in human cells, nucleosome rearrangement by SWI/SNF chromatin remodelers is modulated by H1 (Ramachandran *et al.*, 2003). *T. brucei* TbISWI (SWI/SNF family) is necessary for repressing transcription at procyclin loci and silent BESs, especially at the promoter regions of the later (Stanne & Rudenko, 2010, Hughes *et al.*, 2007). Therefore, TbISWI is a potential candidate involved in the assembly of H1 into silent Pol I promoters BESs. In addition, the *T. brucei* putative histone chaperone TbNLP might also regulate the targeted deposition of H1. TbNLP is a member of the nucleophosmin/nucleoplasmin family of nuclear chaperones and regulates transcription at active and silent BESs and procyclin loci (Narayanan *et al.*, 2011). In humans, a protein nucleophosmin is a chaperone of the linker histone and directly regulates H1 deposition into chromatin (Gadad *et al.*, 2011). Interestingly, in support of this hypothesis, TbISWI and TbNLP co-localize at both active and silent Pol I loci (rDNA, active/silent BESs and procyclin loci) in BSFs (Stanne *et al.*, 2011, Narayanan *et al.*, 2011, Stanne *et al.*, 2015) and were recently shown to interact and, together with two other chromatin-associated proteins, compose the *T. brucei* ISWI remodeling complex (Stanne *et al.*, 2015).

Although depletion of H1 caused global changes in chromatin condensation, by RNA-Seq we detected a very low number of genes with significantly altered expression levels (only 26 out of 8996 expressed genes). Similar trends have been observed in the past: only 26 of over 6000 genes were downregulated in H1 knock-out mutants of yeast, (Hellauer *et al.*, 2001); and 29 genes in mammalian cells showed an altered expression of more than 2-fold (Fan *et al.*, 2005). In contrast to these organisms, however, *T. brucei* H1 acts only as a negative-regulator of gene expression, since H1 depletion results in higher transcripts levels for all significant genes. The fact that H1 is important for global chromatin condensation but only regulates a specific set of genes both in lower eukaryotes, such as *T. brucei*, and in the evolutionarily distant mammals (Fan *et al.*, 2005), might indicate functional conservation of the mechanisms

underlying control of gene expression mediated by these histones. This also supports the use of less complex and amenable to genetic engineering organisms such as *T. brucei* as models to study H1 roles in gene expression.

The apparently specific role of histone H1 in *T. brucei* Pol I loci is likely due to the fact that Pol I genes are probably the only ones that are transcriptionally regulated in *T. brucei*. In fact, because most Pol II-dependent genes are organized in polycistronic units, their expression is believed to be constitutive and gene regulation of Pol II to occur mainly post-transcriptionally. The levels of H1 in Pol II-transcribed loci were similar to silent BESs, which, together with our observations, corroborates the hypothesis that chromatin compaction does not impact considerably on Pol II transcription (Povelones *et al.*, 2012). We cannot exclude that in H1 RNAi clones, the remaining levels of H1 (~40% at the mRNA level) (Results 3.1. Fig. 2B) are enough to avoid transcriptional alterations at Pol II-transcribed genes. Establishing a complete H1 knockout would be most appropriate to address this question. We attempted to do such an H1 KO but the existence of five H1 genes and eventually more, not annotated in the genome (Results 3.1., Fig. S1) has precluded this so far.

4.2 H1 PLAYS A ROLE IN DNA REPAIR

The noticeable discrepancy between a global role of H1 in chromatin compaction and the limited role in transcription regulation also observed in other organisms (Fan *et al.*, 2005, Hellauer *et al.*, 2001) suggests that *T. brucei* H1 plays important but yet unsolved roles that go beyond transcription regulation. Indeed, in *T. brucei*, H1 is inhibitory of DNA repair induced by MMS (Results 3.1., Fig.9). This could be due to the role of H1 in keeping chromatin closed and thus refractory to DNA repair machinery. Since HR-mediated DNA repair dominates in *T. brucei* (Burton *et al.*, 2007, Glover *et al.*, 2008), it is likely that H1 is a suppressor of HR in this parasite. These results are also consistent with the observations that H1 depleted cells undergo more frequent *VSG* gene conversion, a recombination-based switching mechanism (Povelones *et al.*, 2012).

In yeast, H1 suppresses HR-mediated DNA repair and recombination between repeated DNA regions such as telomeres and rDNA transcription units, maintaining genome stability and prolonged life-span (Downs *et al.*, 2003, Li *et al.*, 2008). Thus, in *T. brucei,* since H1 ubiquitously contributes to chromatin condensation we could hypothesize that is also plays a more general role in suppressing widespread illegitimate HR and maintaining genome stability. The role of H1 in HR inhibition could then be particularly relevant to prevent recombination at highly repeated regions such as BESs. The fact that we did not see any severe growth or cell cycle defects in H1-depleted cells (Results 3.1., Fig. 3) suggests that either enough H1 levels exist in depleted cells to avoid deleterious genome instability while facilitating DNA repair, or that H1 is not necessary for maintaining fundamental genome integrity. An H1 KO cell-line would clarify this issue. Also, to sort out if increased resistance to DNA damage in H1-depleted parasites is dependent on HR, we could examine the response to DNA damage in H1-depleted HR-mutant *T. brucei* parasites, such as mutants for RAD51 and RAD51-related proteins (McCulloch & Barry, 1999, Proudfoot & McCulloch, 2005).

Finally, we observed that depletion of H1 increases resistance to MMS-induced DNA damage but not to DNA damage induced by phleomycin. This might rely on the fact that phleomycin triggers a more extensive DNA damage than MMS. Phleomycin can create DSBs across the genome, whereas MMS stalls replication forks, which are only present at discrete regions of the genome at any given time during the S-phase of the cell cycle. Perhaps the widespread damage caused by phleomycin counteracts the beneficial effect of H1 depletion in facilitating the recruitment of DNA repair machinery. Indeed, a study in *T. brucei* shows that histone γH2A, a common marker of DNA damage in eukaryotes, is detected in restricted nuclear foci or in the entire nucleus, depending if the damage is induced by MMS or phleomycin, respectively (Glover & Horn, 2012).

In the future, the sensitivity of H1 mutants to other DNA-damaging agents that induce DSBs such as exposure to ionizing radiation should be tested. Also, HR could be tested by different and more direct methods. For instance by analyzing the frequency of transformants recovered after electroporation with a construct that integrates via cross-over.

4.3 H1 IS IMPORTANT FOR PARASITE FITNESS *IN VIVO*

Consistent with previous observations, we detected a significant but modest defect on cell growth in cultivated bloodstream parasites when ~60% of H1 is depleted in *T. brucei* (Results 3.1, Fig. 3A) (Povelones *et al.*, 2012). This suggests that, like in other lower eukaryotes (Shen *et al.*, 1995, Shen & Gorovsky, 1996, Patterton *et al.*, 1998, Barra *et al.*, 2000), H1 may not be essential for *T. brucei* survival in culture. An alternative explanation is that the remaining H1 (~40%) is sufficient for *T. brucei* survival. A full knock-out of the entire H1 gene family would allow us to discriminate between the two possibilities. With the advent of CRISPR/Cas9 system, which has been used to engineer parasites such as *P. falciparum* (Ghorbal *et al.*, 2014), knock-out of a multigene family in *T. brucei* may be foreseen in the near future.

When we infected mice with H1 depleted parasites, these mutant parasites were substantially less infective, allowing mice to survive up to twice longer (Results 3.2, Fig. 1A, B). This survival phenotype was associated with a lower parasitaemia in the first days of infection. The lower parasitaemia cannot be simply explained by the slower parasite growth rate detected *in vitro*. Indeed, based on *in vitro* growth rates, we estimated a delay in parasitaemia of up to one day, but *in vivo* the delay ranged from 1-10 days, which indicates that H1 is necessary for adapting to the host environment at the onset of infection. Consistent with a role of H1 in host adaptation, overexpression of H1 in *L. major* or *L. donovani* also results in a delay in differentiation of promastigotes into amastigotes (Smirlis *et al.*, 2006, Alexandratos *et al.*, 2013).

When a second population of parasites established (switchers) (Results 3.2, Fig. 1D), parasitaemia increased very rapidly, eventually killing the mouse (Results 3.2, Fig. 1B; Fig. S2). This second population of parasites did not inactivate or loose H1 RNAi cassette (Results 3.2, Fig. 1C). One possible explanation for this apparent gain-of-fitness is that, after 9-14 days of infection, parasites have activated a compensatory pathway for H1 function. This could involve, for instance, replacement by other small linker DNA-binding proteins such as HMGB proteins, which have been shown to interchange or associate with H1, either antagonizing or synergizing its chromatin interactions, respectively (Thomas & Stott, 2012, El Gazzar *et al.*, 2009). Interestingly, H1-depleted parasites that regained fitness did not regress in their phenotype when

establishing a novel infection in mice, indicating that the H1 putative compensatory mechanism is inheritable (Results 3.2, Fig. 2). A proteome or transcriptome analysis of these H1-depleted parasites could allow us to identify such compensatory mechanism.

When we asked why H1 depleted parasites took longer to reach high numbers in the bloodstream of the mammal, one of the most plausible hypothesis was that, somehow, these parasites were less resistant to the host immune system. Supporting this hypothesis was our former observation that monoallelic expression of *VSGs*, crucial for parasite immune evasion, was impaired by H1 depletion. Confirming our predictions, in the highly immune-compromised Rag2$^{-/-}\gamma c^{-/-}$ mice, H1-depleted parasites behaved essentially as the parental parasites containing normal H1 levels (Results, 3.2. Fig. 3). Importantly, this observation allowed us to rule out the possibility that severe parasite growth defects inexistent in culture could have arisen exclusively *in vivo*. On the contrary, our data strongly suggests that H1 is necessary for *T. brucei* full capacity of evading the host immune system. Future studies are needed to dissect the mechanism behind this. Rag2$^{-/-}\gamma c^{-/-}$ mice are severely compromised in several innate and acquired immunity components (Mazurier *et al.*, 1999, Shultz *et al.*, 2007). One could perform infections in mice with more restricted immune deficiencies to narrow down which components of the immune system does *T. brucei* fails to circumvent when H1 is depleted.

Different mechanisms could explain the reduced fitness of H1 depleted parasites in immune-competent animals. First, it is possible that deregulation of gene expression could have a negative impact on putative mechanisms that may allow trypanosomes to sense and resist to the host immune system. These genes could be among the 26 differentially expressed genes or other genes whose repetitive nature prevents unambiguous mapping of RNA-Seq reads and accurate expression analysis (*ESAGs* and rRNAs, for example). We should also consider that gene expression probably differs between parasites growing *in vivo* and *in vitro*. Therefore, we cannot exclude that *in vivo* H1 depletion could alter the expression of Pol II transcribed genes that were not detected in our RNA-Seq analysis done for parasites in culture.

Second, it is appealing to associate the transient loss-of-fitness in H1-depleted parasites with the observed alterations in *VSG* expression. Povelones et al. showed that deletion of H1 results in a slight increase in *VSG* switching in culture (Povelones *et al.*, 2012). We could speculate that derepression of several silent *VSGs* or increased *VSG* switching frequency could raise heterogeneity in the first peak of parasitaemia with abnormal exposure of multiple VSGs. Mixed VSG coats could enhance activation of the immune system and be more easily controlled by it. Indeed, the proportion of parasites expressing one or more silent VSGs at the surface seems to increase *in vitro* when H1 is depleted (Results 3.1, Fig. 7) even though such abnormal VSG expressors are present in low proportions (maximum 0.4%) in the overall population. However, this hypothesis is inconsistent with the observation that trypanosomes expressing a mosaic coat of two different VSGs are equally infective (Munoz-Jordan *et al.*, 1996), are cleared by the immune system at normal kinetics and actually escape early detection by host B-cells (Dubois *et al.*, 2005).

One should note that in these studies only two different VSGs were expressed simultaneously and constitutively at equal or comparable amounts. In H1-depleted trypanosomes, multiple silent VSGs are expressed and not in a constitutive way but rather dependent on H1 RNAi induction. The exposure of multiple VSGs at variable levels could have a different outcome. Besides, the appearance of multiple VSG expressors might be significantly higher *in vivo*. In the H1-depleted clone C2, which displayed the stronger loss-of-fitness, H1 levels were much lower *in vivo* (~10% of parental mRNA levels) than in culture (~40% of parental mRNA levels) (Results 3.2, Fig. 1C) and thus *in vivo* the H1-mutant phenotype may be more pronounced. Also, in mouse infections, trypanosomes are not subject to drug selection of a single BES. Therefore, we can speculate that derepression of silent VSGs and VSG switching frequencies are also increased in H1-depleted parasites growing *in vivo*.

In immune-competent mice, H1-depleted parasites display a delay in parasitemia detected already four days post-infection. Thus, the intervening host immune components must rapidly target these parasites after inoculation. VSG-specific antibodies only reach enough titres to induce parasite clearance around seven days after infection in parental parasites (Results 3.2, Fig. S2) and thus should not

explain alone the early impairment in parasitaemia of H1-depleted parasites. Considering the fast host response, innate immune effectors are likely candidates. One possibility is an increased recognition by the complement system. The host alternative complement pathway is activated and triggers cell lysis in *T. brucei* parasites that lack a VSG coat (Ferrante & Allison, 1983). A study proposes that even a minimal and experimentally undetectable perturbation of the VSG coat is enough to trigger clearance of trypanosomes in 12 h due to increased recognition by the complement or by another unrelated mouse lytic factor (Sheader *et al.*, 2005).

We might speculate that under increased H1 depletion *in vivo*, there are alterations in the VSG coat not detected in culture. When ectopic expression of a second *VSG* is induced, there is temporary silencing of the active *VSG* at the mRNA and protein levels (Batram *et al.*, 2014). It is not clear how this occurs, and if the additional VSG copies are sensed only at the RNA level or if VSG protein synthesis and trafficking also respond to excessive VSG levels. In light of these observations, perhaps a partial downregulation of the active *VSG* also happens in H1-depleted parasites due to simultaneous derepression of several silent *VSGs*. This reduction in VSG levels would be certainly small since low *VSG* mRNA levels rapidly induce cell cycle arrest and translation block (Sheader *et al.*, 2005, Smith *et al.*, 2009). If such defects were occurring in H1-depleted trypanosomes, parasite growth would be impaired in immune-compromised mice, which was not the case (Results 3.2, Fig. 3E). As proposed by others, a minor reduction in the number of surface VSG molecules could be sufficient to expose other membrane epitopes normally shielded by the VSGs, trigger complement activation and parasite lysis or phagocytosis by macrophages.

These hypotheses, though, remain speculative and need to be tested in the future. Transcriptome and/or proteome analysis would determine the relative levels of silent and active VSGs in H1-depleted trypanosomes growing *in vivo*. Infections in mice deficient in complement or other components of the immune system could potentially reveal the host effectors involved.

In summary, our results uncover the functional relevance of *T. brucei* H1 in the parasite ability to evade the host immune system. In addition, our data highlights the

differences that emerge when trypanosomes are living in culture or inside its mammalian host *in vivo*, pointing out the importance of adressing the host-pathogen interactions when assessing phenotypic effects in the parasite.

5 CONCLUDING REMARKS

In my PhD work I showed that in *T. brucei* H1 keeps chromatin globally compact, inhibits DNA repair and represses the transcription of Pol I silent genes, including *VSG*s and procyclins. My data suggests that histone H1 acts as an inhibitor of Pol I transcription initiation and elongation, probably by controlling chromatin accessibility to transcription factors and Pol I machinery. In light of what is known about the dynamics of H1 binding and control of transcriptional activation in higher eukaryotes we propose that H1 assembly or displacement is a prerequisite for transcriptional silencing or activation of Pol I in *T. brucei*. It would be interesting to study the role of H1 during differentiation of bloodstream forms in procyclic parasites, in which fast shutdown of active BES occurs, accompanied by full transcriptional activation of procyclin loci. According to the hypothesis that H1 acts in the first steps of BES transcriptional silencing, we would expect that H1 depletion would delay BES transcriptional repression during differentiation.

A future interesting direction is the study of the predicted H1 variants present in *T. brucei*. There is evidence in higher and lower eukaryotes that specific H1 variants occupy distinct genome locations and participate in the regulation of particular subsets of genes. Maybe a specific *T. brucei* H1 variant associates with and inhibits transcription at Pol I loci. H1 variants might also exhibit different PTMs namely different phosphorylation patterns since H1 isoforms differ in the number and position of serine/threonine residues at their N-terminal domain. H1 variants with particular phosphorylation patterns associate with Pol I or Pol II transcribed loci in human cells, which tempt us to hypothesize that a similar mechanism could also happen in *T. brucei.* Possible approaches to investigate the functional relevance of H1 variants in *T. brucei* could be by determining the genomic distribution of each H1 variant (e.g. ChIP sequencing) and by analysing the phenotype of mutant parasites in which each variant was deleted or overexpressed.

I also demonstrated that H1 is important for parasite fitness during mammalian infection *in vivo*. Although it is clear that parasite immune evasion is impaired when H1

is depleted, mechanistically we ignore how this occurs. It shows nevertheless that H1 is important for *T. brucei* adaptation to its natural host, and parasites seem to rapidly develop a compensatory mechanism to overcome the deleterious effects of losing H1. Understanding why H1-depleted parasites are less resistant to the mammalian immune system is a challenge ahead. In the future, this question should be addressed using a pleomorphic *T. brucei* cell-line, which more closely resembles natural *T. brucei* populations and allows us to examine the role of H1 in the differentiation of the transmissible stumpy stages.

Finally, during my PhD work, I adapted metabolic labeling of RNA with 4sU to *T. brucei*. Several attempts were done to improve the final 4sU-RNA yield, even though all of them failed. Additional tests could be performed such has RNA fragmentation before biotinylation, which was successfully applied for 4TU metabolic labeling (Gay et al., 2014) or using methanethiosulfonate instead of HPDP to covalently bind biotin, which strongly increases 4sU-RNA yields (Duffy *et al.*, 2015). If further attempts to improve the efficiency of 4sU-labeling continue to fail, alternative technologies for purification and quantification of nascent RNAs in *T. brucei* should be explored.

Metabolic labeling of RNA with 4sU has the important advantage of allowing the measurement of transcription *de novo* in virtually unperturbed cells, contrarily to nuclear run-on, which has been the gold-standard for measuring transcription in trypanosomes. 4sU-labeled RNAs are suitable for high-throughput outputs such as RNA-Seq.and are a potentially valuable tool to study with high resolution and whole genome coverage topics whose study has been precluded by technical limitations in *T. brucei*. For instance, one of the most important enigmas this method could solve is the exact position of Pol II transcription start sites across the genome, which has been precluded by the fact that no Pol II promoter motifs have been identified until now for the vast majority of *T. brucei* genes. This methodology could also confirm (or not) the dogma in the field that all Pol II-transcribed PTUs are transcribed in the same levels. Besides, in the context of my PhD work, it can be used to temporally follow Pol I transcriptional activation or silencing and identify Pol II cryptic transcription initiation.

Compared with Pol II transcription, much less in known about Pol I regulation in eukaryotes. My work contributes to this knowledge by directly showing the importance of histone H1 in the control of transcription levels at Pol I loci. In other organisms, H1-mediated control of specific genes has been observed but only for Pol II-transcribed genes. It would be interesting to determine if, in these organisms, H1 also conserves similar roles as those we found in *T. brucei* regarding the control of Pol I transcription.

In sum, collectively my data demonstrates that *T. brucei* H1 has 'broad skills' when it comes to chromatin condensation since it is necessary for chromatin compaction globally throughout the genome; but H1 also has 'specific functions' in what concerns gene expression and, it seems, immune evasion.

6 REFERENCES

Acosta-Serrano, A., R. N. Cole, A. Mehlert, M. G. Lee, M. A. Ferguson & P. T. Englund, (1999) The procyclin repertoire of Trypanosoma brucei. Identification and structural characterization of the Glu-Pro-rich polypeptides. *The Journal of biological chemistry* **274**: 29763-29771.

Acosta-Serrano, A., E. Vassella, M. Liniger, C. Kunz Renggli, R. Brun, I. Roditi & P. T. Englund, (2001) The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 1513-1518.

Adl, S. M., A. G. Simpson, C. E. Lane, J. Lukes, D. Bass, S. S. Bowser, M. W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. Le Gall, D. H. Lynn, H. McManus, E. A. Mitchell, S. E. Mozley-Stanridge, L. W. Parfrey, J. Pawlowski, S. Rueckert, R. S. Shadwick, C. L. Schoch, A. Smirnov & F. W. Spiegel, (2012) The revised classification of eukaryotes. *The Journal of eukaryotic microbiology* **59**: 429-493.

Aitcheson, N., S. Talbot, J. Shapiro, K. Hughes, C. Adkin, T. Butt, K. Sheader & G. Rudenko, (2005) VSG switching in Trypanosoma brucei: antigenic variation analysed using RNAi in the absence of immune selection. *Molecular microbiology* **57**: 1608-1622.

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

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

Albert, B., J. Perez-Fernandez, I. Leger-Silvestre & O. Gadal, (2012) Regulation of ribosomal RNA production by RNA polymerase I: does elongation come first? *Genetics research international* **2012**: 276948.

Alexandratos, A., J. Clos, M. Samiotaki, A. Efstathiou, G. Panayotou, K. Soteriadou & D. Smirlis, (2013) The loss of virulence of histone H1 overexpressing Leishmania donovani parasites is directly associated with a reduction of HSP83 rate of translation. *Molecular microbiology* **88**: 1015-1031.

Ali, J. A., D. J. Creek, K. Burgess, H. C. Allison, M. C. Field, P. Maser & H. P. De Koning, (2013) Pyrimidine salvage in Trypanosoma brucei bloodstream forms and the trypanocidal action of halogenated pyrimidines. *Molecular pharmacology* **83**: 439-453.

Alibu, V. P., L. Storm, S. Haile, C. Clayton & D. Horn, (2005) A doubly inducible system for RNA interference and rapid RNAi plasmid construction in Trypanosoma brucei. *Molecular and biochemical parasitology* **139**: 75-82.

Allan, J., P. G. Hartman, C. Crane-Robinson & F. X. Aviles, (1980) The structure of histone H1 and its location in chromatin. *Nature* **288**: 675-679.

Allan, J., T. Mitchell, N. Harborne, L. Bohm & C. Crane-Robinson, (1986) Roles of H1 domains in determining higher order chromatin structure and H1 location. *Journal of molecular biology* **187**: 591-601.

Alsford, N. S., M. Navarro, H. R. Jamnadass, H. Dunbar, M. Ackroyd, N. B. Murphy, K. Gull & K. Ersfeld, (2003) The identification of circular extrachromosomal DNA in the nuclear genome of Trypanosoma brucei. *Molecular microbiology* **47**: 277-289.

Alsford, S., K. duBois, D. Horn & M. C. Field, (2012) Epigenetic mechanisms, nuclear architecture and the control of gene expression in trypanosomes. *Expert Rev Mol Med* **14**: e13.

Alsford, S., L. Glover & D. Horn, (2005a) Multiplex analysis of RNA interference defects in Trypanosoma brucei. *Molecular and biochemical parasitology* **139**: 129-132.

Alsford, S. & D. Horn, (2004) Trypanosomatid histones. *Molecular microbiology* **53**: 365-372.

Alsford, S. & D. Horn, (2011) Elongator protein 3b negatively regulates ribosomal DNA transcription in african trypanosomes. *Molecular and cellular biology* **31**: 1822-1832.

Alsford, S. & D. Horn, (2012) Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. *Nucleic acids research*.

Alsford, S., T. Kawahara, L. Glover & D. Horn, (2005b) Tagging a T. brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Molecular and biochemical parasitology* **144**: 142-148.

Alsford, S., T. Kawahara, C. Isamah & D. Horn, (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.

Amiguet-Vercher, A., D. Perez-Morga, A. Pays, P. Poelvoorde, H. Van Xong, P. Tebabi, L. Vanhamme & E. Pays, (2004) Loss of the mono-allelic control of the VSG expression sites during the development of Trypanosoma brucei in the bloodstream. *Molecular microbiology* **51**: 1577-1588.

Anders, S. & W. Huber, (2010) Differential expression analysis for sequence count data. *Genome biology* **11**: R106.

Ansorge, I., D. Steverding, S. Melville, C. Hartmann & C. Clayton, (1999) Transcription of 'inactive' expression sites in African trypanosomes leads to expression of multiple transferrin receptor RNAs in bloodstream forms. *Molecular and biochemical parasitology* **101**: 81-94.

Archer, S. K., D. Inchaustegui, R. Queiroz & C. Clayton, (2011) The cell cycle regulated transcriptome of Trypanosoma brucei. *PloS one* **6**: e18425.

Barbet, A. F. & T. C. McGuire, (1978) Crossreacting determinants in variant-specific surface antigens of African trypanosomes. *Proceedings of the National Academy of Sciences of the United States of America* **75**: 1989-1993.

Barbour, A. G., C. J. Carter & C. D. Sohaskey, (2000) Surface protein variation by expression site switching in the relapsing fever agent Borrelia hermsii. *Infection and immunity* **68**: 7114-7121.

Baron, D. M., K. S. Ralston, Z. P. Kabututu & K. L. Hill, (2007) Functional genomics in Trypanosoma brucei identifies evolutionarily conserved components of motile flagella. *Journal of cell science* **120**: 478-491.

Barra, J. L., L. Rhounim, J. L. Rossignol & G. Faugeron, (2000) Histone H1 is dispensable for methylation-associated gene silencing in Ascobolus immersus and essential for long life span. *Molecular and cellular biology* **20**: 61-69.

Barry, J. D., (1979) Capping of variable antigen on Trypanosoma brucei, and its immunological and biological significance. *Journal of cell science* **37**: 287-302.

Barry, J. D., M. L. Ginger, P. Burton & R. McCulloch, (2003) Why are parasite contingency genes often associated with telomeres? *International journal for parasitology* **33**: 29-45.

Barry, J. D. & R. McCulloch, (2001) Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv Parasitol* **49**: 1-70.

Bates, D. L., P. J. Butler, E. C. Pearson & J. O. Thomas, (1981) Stability of the higher-order structure of chicken-erythrocyte chromatin in solution. *European journal of biochemistry / FEBS* **119**: 469-476.

Batram, C., N. G. Jones, C. J. Janzen, S. M. Markert & M. Engstler, (2014) Expression site attenuation mechanistically links antigenic variation and development in Trypanosoma brucei. *Elife* **3**: e02324.

Becker, M., N. Aitcheson, E. Byles, B. Wickstead, E. Louis & G. Rudenko, (2004) Isolation of the repertoire of VSG expression site containing telomeres of Trypanosoma brucei 427 using transformation-associated recombination in yeast. *Genome research* **14**: 2319-2329.

Bednar, J., R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster & C. L. Woodcock, (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin.

Proceedings of the National Academy of Sciences of the United States of America **95**: 14173-14178.

Belli, S. I., (2000) Chromatin remodelling during the life cycle of trypanosomatids. *International journal for parasitology* **30**: 679-687.

Bender, K., B. Betschart & H. Hecker, (1992) Histone-DNA interactions in the chromatin of procyclic Trypanosoma brucei brucei. *Parasitology research* **78**: 495-500.

Benjamini, Y. & Y. Hochberg, (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289-300.

Benmerzouga, I., J. Concepcion-Acevedo, H. S. Kim, A. V. Vandoros, G. A. Cross, M. M. Klingbeil & B. Li, (2013) Trypanosoma brucei Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. *Molecular microbiology* **87**: 196-210.

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

Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D. C. Bartholomeu, N. J. Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Bohme, L. Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T. J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P. J. Mooney, S. Moule, D. M. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C. S. Peacock, J. Peterson, M. A. Quail, E. Rabbinowitsch, M. A. Rajandream, C. Reitter, S. L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. Turner, A. Tait, A. R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, et al., (2005) The genome of the African trypanosome Trypanosoma brucei. *Science* **309**: 416-422.

Bharath, M. M., N. R. Chandra & M. R. Rao, (2003) Molecular modeling of the chromatosome particle. *Nucleic acids research* **31**: 4264-4274.

Bindereif, A., (2012) *RNA metabolism in trypanosomes*. Springer, New York.

Bitter, W., H. Gerrits, R. Kieft & P. Borst, (1998) The role of transferrin-receptor variation in the host range of Trypanosoma brucei. *Nature* **391**: 499-502.

Black, S. J., P. Guirnalda, D. Frenkel, C. Haynes & V. Bockstal, (2010) Induction and regulation of Trypanosoma brucei VSG-specific antibody responses. *Parasitology* **137**: 2041-2049.

Blasco, M. A., (2007) The epigenetic regulation of mammalian telomeres. *Nature reviews. Genetics* **8**: 299-309.

Blattner, J. & C. E. Clayton, (1995) The 3'-untranslated regions from the Trypanosoma brucei phosphoglycerate kinase-encoding genes mediate developmental regulation. *Gene* **162**: 153-156.

Blum, J., C. Schmid & C. Burri, (2006) Clinical aspects of 2541 patients with second stage human African trypanosomiasis. *Acta tropica* **97**: 55-64.

Blum, M. L., J. A. Down, A. M. Gurnett, M. Carrington, M. J. Turner & D. C. Wiley, (1993) A structural motif in the variant surface glycoproteins of Trypanosoma brucei. *Nature* **362**: 603-609.

Boothroyd, C. E., O. Dreesen, T. Leonova, K. I. Ly, L. M. Figueiredo, G. A. Cross & F. N. Papavasiliou, (2009) A yeast-endonuclease-generated DNA break induces antigenic switching in Trypanosoma brucei. *Nature* **459**: 278-281.

Boothroyd, J. C. & G. A. Cross, (1982) Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5' end. *Gene* **20**: 281-289.
Branche, C., L. Kohl, G. Toutirais, J. Buisson, J. Cosson & P. Bastin, (2006) Conserved and specific functions of axoneme components in trypanosome motility. *Journal of cell science* **119**: 3443-3455.
Brandenburg, J., B. Schimanski, E. Nogoceke, T. N. Nguyen, J. C. Padovan, B. T. Chait, G. A. Cross & A. Gunzl, (2007) Multifunctional class I transcription in Trypanosoma brucei depends on a novel protein complex. *The EMBO journal* **26**: 4856-4866.
Brown, S. D., J. Huang & L. H. Van der Ploeg, (1992) The promoter for the procyclic acidic repetitive protein (PARP) genes of Trypanosoma brucei shares features with RNA polymerase I promoters. *Molecular and cellular biology* **12**: 2644-2652.
Brun, R., J. Blum, F. Chappuis & C. Burri, (2010) Human African trypanosomiasis. *Lancet* **375**: 148-159.
Buguet, A., S. Bisser, T. Josenando, F. Chapotot & R. Cespuglio, (2005) Sleep structure: a new diagnostic tool for stage determination in sleeping sickness. *Acta tropica* **93**: 107-117.
Burri, M., W. Schlimme, B. Betschart & H. Hecker, (1994) Characterization of the histones of Trypanosoma brucei brucei bloodstream forms. *Acta tropica* **58**: 291-305.
Burri, M., W. Schlimme, B. Betschart, U. Kampfer, J. Schaller & H. Hecker, (1993) Biochemical and functional characterization of histone H1-like proteins in procyclic Trypanosoma brucei brucei. *Parasitology research* **79**: 649-659.
Burri, M., W. Schlimme, B. Betschart, H. Lindner, U. Kampfer, J. Schaller & H. Hecker, (1995) Partial amino acid sequence and functional aspects of histone H1 proteins in Trypanosoma brucei brucei. *Biology of the cell / under the auspices of the European Cell Biology Organization* **83**: 23-31.
Burton, P., D. J. McBride, J. M. Wilkes, J. D. Barry & R. McCulloch, (2007) Ku heterodimer-independent end joining in Trypanosoma brucei cell extracts relies upon sequence microhomology. *Eukaryotic cell* **6**: 1773-1781.
Calderano, S. G., W. C. Drosopoulos, M. M. Quaresma, C. A. Marques, S. Kosiyatrakul, R. McCulloch, C. L. Schildkraut & M. C. Elias, (2015) Single molecule analysis of Trypanosoma brucei DNA replication dynamics. *Nucleic acids research* **43**: 2655-2665.
Carrington, M. & J. Boothroyd, (1996) Implications of conserved structural motifs in disparate trypanosome surface proteins. *Molecular and biochemical parasitology* **81**: 119-126.
Carrington, M., N. Miller, M. Blum, I. Roditi, D. Wiley & M. Turner, (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.
Catez, F., H. Yang, K. J. Tracey, R. Reeves, T. Misteli & M. Bustin, (2004) Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. *Molecular and cellular biology* **24**: 4321-4328.
Cavalier-Smith, T., (2010) Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biology letters* **6**: 342-345.
Chaves, I., J. Zomerdijk, A. Dirks-Mulder, R. W. Dirks, A. K. Raap & P. Borst, (1998) Subnuclear localization of the active variant surface glycoprotein gene expression site in Trypanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 12328-12333.
Chen, K. K. & J. E. Donelson, (1980) Sequences of two kinetoplast DNA minicircles of Tryptanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 2445-2449.
Chen, Z. H., M. Zhu, J. Yang, H. Liang, J. He, S. He, P. Wang, X. Kang, M. A. McNutt, Y. Yin & W. H. Shen, (2014) PTEN Interacts with Histone H1 and Controls Chromatin Condensation. *Cell Rep* **8**: 2003-2014.

Cheung, V., G. Chua, N. N. Batada, C. R. Landry, S. W. Michnick, T. R. Hughes & F. Winston, (2008) Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. *PLoS biology* **6**: e277.

Clayton, C. E. & M. R. Mowatt, (1989) The procyclic acidic repetitive proteins of Trypanosoma brucei. Purification and post-translational modification. *The Journal of biological chemistry* **264**: 15088-15093.

Cleary, M. D., C. D. Meiering, E. Jan, R. Guymon & J. C. Boothroyd, (2005) Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nature biotechnology* **23**: 232-237.

Cliffe, L. J., R. Kieft, T. Southern, S. R. Birkeland, M. Marshall, K. Sweeney & R. Sabatini, (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic acids research* **37**: 1452-1462.

Conway, C., R. McCulloch, M. L. Ginger, N. P. Robinson, A. Browitt & J. D. Barry, (2002) Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. *The Journal of biological chemistry* **277**: 21269-21277.

Core, L. J., J. J. Waterfall & J. T. Lis, (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845-1848.

Corona, D. F., G. Siriaco, J. A. Armstrong, N. Snarskaya, S. A. McClymont, M. P. Scott & J. W. Tamkun, (2007) ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS biology* **5**: e232.

Croken, M. M., S. C. Nardelli & K. Kim, (2012) Chromatin modifications, epigenetics, and how protozoan parasites regulate their lives. *Trends in parasitology* **28**: 202-213.

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

Cross, G. A., (1984) Structure of the variant glycoproteins and surface coat of Trypanosoma brucei. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **307**: 3-12.

Cross, G. A., H. S. Kim & B. Wickstead, (2014) Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Molecular and biochemical parasitology* **195**: 59-73.

Dammann, R., R. Lucchini, T. Koller & J. M. Sogo, (1993) Chromatin structures and transcription of rDNA in yeast Saccharomyces cerevisiae. *Nucleic acids research* **21**: 2331-2338.

Das, A., Q. Zhang, J. B. Palenchar, B. Chatterjee, G. A. Cross & V. Bellofatto, (2005) Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Molecular and cellular biology* **25**: 7314-7322.

de Koning, H. P., D. J. Bridges & R. J. Burchmore, (2005) Purine and pyrimidine transport in pathogenic protozoa: from biology to therapy. *FEMS microbiology reviews* **29**: 987-1020.

Deans, C. & K. A. Maggert, (2015) What do you mean, "epigenetic"? *Genetics* **199**: 887-896.

DeGrasse, J. A., K. N. DuBois, D. Devos, T. N. Siegel, A. Sali, M. C. Field, M. P. Rout & B. T. Chait, (2009) Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Molecular & cellular proteomics : MCP* **8**: 2119-2130.

Denninger, V., A. Fullbrook, M. Bessat, K. Ersfeld & G. Rudenko, (2010) The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in Trypanosoma brucei. *Molecular microbiology* **78**: 459-474.

Denninger, V. & G. Rudenko, (2014) FACT plays a major role in histone dynamics affecting VSG expression site control in Trypanosoma brucei. *Molecular microbiology* **94**: 945-962.

Devaux, S., S. Kelly, L. Lecordier, B. Wickstead, D. Perez-Morga, E. Pays, L. Vanhamme & K. Gull, (2007) Diversification of function by different isoforms of conventionally shared RNA polymerase subunits. *Molecular biology of the cell* **18**: 1293-1301.

Dolken, L., Z. Ruzsics, B. Radle, C. C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal & U. H. Koszinowski, (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**: 1959-1972.

Donelson, J. E., (2003) Antigenic variation and the African trypanosome genome. *Acta tropica* **85**: 391-404.

Dooijes, D., I. Chaves, R. Kieft, A. Dirks-Mulder, W. Martin & P. Borst, (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of Euglena gracilis. *Nucleic acids research* **28**: 3017-3021.

Downs, J. A., E. Kosmidou, A. Morgan & S. P. Jackson, (2003) Suppression of homologous recombination by the Saccharomyces cerevisiae linker histone. *Molecular cell* **11**: 1685-1692.

Doyle, J. J., H. Hirumi, K. Hirumi, E. N. Lupton & G. A. Cross, (1980) Antigenic variation in clones of animal-infective Trypanosoma brucei derived and maintained in vitro. *Parasitology* **80**: 359-369.

DuBois, K. N., S. Alsford, J. M. Holden, J. Buisson, M. Swiderski, J. M. Bart, A. V. Ratushny, Y. Wan, P. Bastin, J. D. Barry, M. Navarro, D. Horn, J. D. Aitchison, M. P. Rout & M. C. Field, (2012) NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. *PLoS biology* **10**: e1001287.

Dubois, M. E., K. P. Demick & J. M. Mansfield, (2005) Trypanosomes expressing a mosaic variant surface glycoprotein coat escape early detection by the immune system. *Infection and immunity* **73**: 2690-2697.

Duffy, E. E., M. Rutenberg-Schoenberg, C. D. Stark, R. R. Kitchen, M. B. Gerstein & M. D. Simon, (2015) Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Molecular cell* **59**: 858-866.

Dunbar, D. A., A. A. Chen, S. Wormsley & S. J. Baserga, (2000) The genes for small nucleolar RNAs in Trypanosoma brucei are organized in clusters and are transcribed as a polycistronic RNA. *Nucleic acids research* **28**: 2855-2861.

Ekanayake, D. & R. Sabatini, (2011) Epigenetic regulation of polymerase II transcription initiation in Trypanosoma cruzi: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryotic cell* **10**: 1465-1472.

Ekanayake, D. K., T. Minning, B. Weatherly, K. Gunasekera, D. Nilsson, R. Tarleton, T. Ochsenreiter & R. Sabatini, (2011) Epigenetic regulation of transcription and virulence in Trypanosoma cruzi by O-linked thymine glucosylation of DNA. *Molecular and cellular biology* **31**: 1690-1700.

El-Sayed, N. M., E. Ghedin, J. Song, A. MacLeod, F. Bringaud, C. Larkin, D. Wanless, J. Peterson, L. Hou, S. Taylor, A. Tweedie, N. Biteau, H. G. Khalak, X. Lin, T. Mason, L. Hannick, E. Caler, G. Blandin, D. Bartholomeu, A. J. Simpson, S. Kaul, H. Zhao, G. Pai, S. Van Aken, T. Utterback, B. Haas, H. L. Koo, L. Umayam, B. Suh, C. Gerrard, V. Leech, R. Qi, S. Zhou, D. Schwartz, T. Feldblyum, S. Salzberg, A. Tait, C. M. Turner, E. Ullu, O. White, S. Melville, M. D. Adams, C. M. Fraser & J. E. Donelson, (2003) The sequence and analysis of Trypanosoma brucei chromosome II. *Nucleic acids research* **31**: 4856-4863.

El Gazzar, M., B. K. Yoza, X. Chen, B. A. Garcia, N. L. Young & C. E. McCall, (2009) Chromatin-specific remodeling by HMGB1 and linker histone H1 silences proinflammatory genes during endotoxin tolerance. *Molecular and cellular biology* **29**: 1959-1971.

Eme, L., S. C. Sharpe, M. W. Brown & A. J. Roger, (2014) On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb Perspect Biol* **6**.

Engstler, M., T. Pfohl, S. Herminghaus, M. Boshart, G. Wiegertjes, N. Heddergott & P. Overath, (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. *Res Microbiol* **162**: 626-636.

Ersfeld, K., K. Asbeck & K. Gull, (1998) Direct visualisation of individual gene organisation in Trypanosoma brucei by high-resolution in situ hybridisation. *Chromosoma* **107**: 237-240.

Ersfeld, K., R. Docherty, S. Alsford & K. Gull, (1996) A fluorescence in situ hybridisation study of the regulation of histone mRNA levels during the cell cycle of Trypanosoma brucei. *Molecular and biochemical parasitology* **81**: 201-209.

Fadda, A., M. Ryten, D. Droll, F. Rojas, V. Farber, J. R. Haanstra, C. Merce, B. M. Bakker, K. Matthews & C. Clayton, (2014) Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Molecular microbiology* **94**: 307-326.

Fan, Y., T. Nikitina, E. M. Morin-Kensicki, J. Zhao, T. R. Magnuson, C. L. Woodcock & A. I. Skoultchi, (2003) H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Molecular and cellular biology* **23**: 4559-4572.

Fan, Y., T. Nikitina, J. Zhao, T. J. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock & A. I. Skoultchi, (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**: 1199-1212.

Fantoni, A., A. O. Dare & C. Tschudi, (1994) RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. *Molecular and cellular biology* **14**: 2021-2028.

FAO, (2014) Programme Against African Trypanosomosis (PAAT). In: http://www.fao.org/ag/AGAINFO/programmes/en/paat/disease.html. Food and Agriculture Organization of the United Nations, pp.

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

Ferrante, A. & A. C. Allison, (1983) Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. *Parasite immunology* **5**: 491-498.

Field, M. C., D. Horn, S. Alsford, L. Koreny & M. P. Rout, (2012) Telomeres, tethers and trypanosomes. *Nucleus* **3**: 478-486.

Figueiredo, L. M. & G. A. Cross, (2010) Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. *Eukaryotic cell* **9**: 148-154.

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

Figueiredo, L. M., C. J. Janzen & G. A. Cross, (2008) A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS biology* **6**: e161.

Friedel, C. C., L. Dolken, Z. Ruzsics, U. H. Koszinowski & R. Zimmer, (2009) Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic acids research* **37**: e115.

Gadad, S. S., P. Senapati, S. H. Syed, R. E. Rajan, J. Shandilya, V. Swaminathan, S. Chatterjee, E. Colombo, S. Dimitrov, P. G. Pelicci, U. Ranga & T. K. Kundu, (2011) The multifunctional protein nucleophosmin (NPM1) is a human linker histone H1 chaperone. *Biochemistry* **50**: 2780-2789.

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

Gassen, A., D. Brechtefeld, N. Schandry, J. M. Arteaga-Salas, L. Israel, A. Imhof & C. J. Janzen, (2012) DOT1A-dependent H3K76 methylation is required for replication regulation in Trypanosoma brucei. *Nucleic acids research* **40**: 10302-10311.

Gay, L., K. V. Karfilis, M. R. Miller, C. Q. Doe & K. Stankunas, (2014) Applying thiouracil tagging to mouse transcriptome analysis. *Nat Protoc* **9**: 410-420.

Genest, P. A., B. ter Riet, C. Dumas, B. Papadopoulou, H. G. van Luenen & P. Borst, (2005) Formation of linear inverted repeat amplicons following targeting of an essential gene in Leishmania. *Nucleic acids research* **33**: 1699-1709.

Gerrits, H., R. Mussmann, W. Bitter, R. Kieft & P. Borst, (2002) The physiological significance of transferrin receptor variations in Trypanosoma brucei. *Molecular and biochemical parasitology* **119**: 237-247.

Ghorbal, M., M. Gorman, C. R. Macpherson, R. M. Martins, A. Scherf & J. J. Lopez-Rubio, (2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nature biotechnology* **32**: 819-821.

Gibson, W., (2007) Resolution of the species problem in African trypanosomes. *International journal for parasitology* **37**: 829-838.

Gilinger, G. & V. Bellofatto, (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic acids research* **29**: 1556-1564.

Giloni, L., M. Takeshita, F. Johnson, C. Iden & A. P. Grollman, (1981) Bleomycin-induced strand-scission of DNA. Mechanism of deoxyribose cleavage. *The Journal of biological chemistry* **256**: 8608-8615.

Ginger, M. L., P. A. Blundell, A. M. Lewis, A. Browitt, A. Gunzl & J. D. Barry, (2002) Ex vivo and in vitro identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. *Eukaryotic cell* **1**: 1000-1009.

Giresi, P. G., J. Kim, R. M. McDaniell, V. R. Iyer & J. D. Lieb, (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* **17**: 877-885.

Giroud, C., F. Ottones, V. Coustou, D. Dacheux, N. Biteau, B. Miezan, N. Van Reet, M. Carrington, F. Doua & T. Baltz, (2009) Murine Models for Trypanosoma brucei gambiense disease progression--from silent to chronic infections and early brain tropism. *PLoS neglected tropical diseases* **3**: e509.

Gjini, E., D. T. Haydon, J. D. Barry & C. A. Cobbold, (2010) Critical interplay between parasite differentiation, host immunity, and antigenic variation in trypanosome infections. *The American naturalist* **176**: 424-439.

Glover, L., S. Alsford, C. Beattie & D. Horn, (2007) Deletion of a trypanosome telomere leads to loss of silencing and progressive loss of terminal DNA in the absence of cell cycle arrest. *Nucleic acids research* **35**: 872-880.

Glover, L. & D. Horn, (2006) Repression of polymerase I-mediated gene expression at Trypanosoma brucei telomeres. *EMBO reports* **7**: 93-99.

Glover, L. & D. Horn, (2012) Trypanosomal histone gammaH2A and the DNA damage response. *Molecular and biochemical parasitology* **183**: 78-83.

Glover, L. & D. Horn, (2014) Locus-specific control of DNA resection and suppression of subtelomeric VSG recombination by HAT3 in the African trypanosome. *Nucleic acids research* **42**: 12600-12613.

Glover, L., J. Jun & D. Horn, (2011) Microhomology-mediated deletion and gene conversion in African trypanosomes. *Nucleic acids research* **39**: 1372-1380.

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

Gommers-Ampt, J. H., F. Van Leeuwen, A. L. de Beer, J. F. Vliegenthart, M. Dizdaroglu, J. A. Kowalak, P. F. Crain & P. Borst, (1993) beta-D-glucosyl-hydroxymethyluracil: a novel

modified base present in the DNA of the parasitic protozoan T. brucei. *Cell* **75**: 1129-1136.
Goringer, H. U., (2012) 'Gestalt,' composition and function of the Trypanosoma brucei editosome. *Annu Rev Microbiol* **66**: 65-82.
Gottschling, D. E., O. M. Aparicio, B. L. Billington & V. A. Zakian, (1990) Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. *Cell* **63**: 751-762.
Graham, S. V. & J. D. Barry, (1995) Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of Trypanosoma brucei. *Molecular and cellular biology* **15**: 5945-5956.
Gray, A. R., (1965) Antigenic variation in a strain of Trypanosoma brucei transmitted by Glossina morsitans and G. palpalis. *J Gen Microbiol* **41**: 195-214.
Grummt, I. & G. Langst, (2013) Epigenetic control of RNA polymerase I transcription in mammalian cells. *Biochimica et biophysica acta* **1829**: 393-404.
Grunfelder, C. G., M. Engstler, F. Weise, H. Schwarz, Y. D. Stierhof, M. Boshart & P. Overath, (2002) Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels. *Traffic* **3**: 547-559.
Gruter, E. & B. Betschart, (2001) Isolation, characterisation and organisation of histone H1 genes in African trypanosomes. *Parasitology research* **87**: 977-984.
Guirnalda, P., N. B. Murphy, D. Nolan & S. J. Black, (2007) Anti-Trypanosoma brucei activity in Cape buffalo serum during the cryptic phase of parasitemia is mediated by antibodies. *International journal for parasitology* **37**: 1391-1399.
Gunzl, A., T. Bruderer, G. Laufer, B. Schimanski, L. C. Tu, H. M. Chung, P. T. Lee & M. G. Lee, (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryotic cell* **2**: 542-551.
Gunzl, A., J. K. Kirkham, T. N. Nguyen, N. Badjatia & S. H. Park, (2014) Mono-allelic VSG expression by RNA polymerase I in Trypanosoma brucei: Expression site control from both ends? *Gene*.
Gunzl, A., E. Ullu, M. Dorner, S. P. Fragoso, K. F. Hoffmann, J. D. Milner, Y. Morita, E. K. Nguu, S. Vanacova, S. Wunsch, A. O. Dare, H. Kwon & C. Tschudi, (1997) Transcription of the Trypanosoma brucei spliced leader RNA gene is dependent only on the presence of upstream regulatory elements. *Molecular and biochemical parasitology* **85**: 67-76.
Gurard-Levin, Z. A., J. P. Quivy & G. Almouzni, (2014) Histone chaperones: assisting histone traffic and nucleosome dynamics. *Annual review of biochemistry* **83**: 487-517.
Haenni, S., C. K. Renggli, C. M. Fragoso, M. Oberle & I. Roditi, (2006) The procyclin-associated genes of Trypanosoma brucei are not essential for cyclical transmission by tsetse. *Molecular and biochemical parasitology* **150**: 144-156.
Haenni, S., E. Studer, G. S. Burkard & I. Roditi, (2009) Bidirectional silencing of RNA polymerase I transcription by a strand switch region in Trypanosoma brucei. *Nucleic acids research* **37**: 5007-5018.
Hajduk, S. L. & K. Vickerman, (1981) Antigenic Variation in Cyclically Transmitted Trypanosoma-Brucei - Variable Antigen Type Composition of the 1st Parasitemia in Mice Bitten by Trypanosome-Infected Glossina-Morsitans. *Parasitology* **83**: 609-621.
Hall, J. P., H. Wang & J. D. Barry, (2013) Mosaic VSGs and the scale of Trypanosoma brucei antigenic variation. *PLoS pathogens* **9**: e1003502.
Happel, N. & D. Doenecke, (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431**: 1-12.
Harshman, S. W., N. L. Young, M. R. Parthun & M. A. Freitas, (2013) H1 histones: current perspectives and challenges. *Nucleic acids research* **41**: 9593-9609.
Hashimoto, H., Y. Takami, E. Sonoda, T. Iwasaki, H. Iwano, M. Tachibana, S. Takeda, T. Nakayama, H. Kimura & Y. Shinkai, (2010) Histone H1 null vertebrate cells exhibit altered nucleosome architecture. *Nucleic acids research* **38**: 3533-3545.

Hayashi, T., H. Hayashi & K. Iwai, (1987) Tetrahymena histone H1. Isolation and amino acid sequence lacking the central hydrophobic domain conserved in other H1 histones. *Journal of biochemistry* **102**: 369-376.

Hecker, H., B. Betschart, K. Bender, M. Burri & W. Schlimme, (1994) The chromatin of trypanosomes. *International journal for parasitology* **24**: 809-819.

Hecker, H. & E. S. Gander, (1985) The compaction pattern of the chromatin of trypanosomes. *Biology of the cell / under the auspices of the European Cell Biology Organization* **53**: 199-208.

Hellauer, K., E. Sirard & B. Turcotte, (2001) Decreased expression of specific genes in yeast cells lacking histone H1. *The Journal of biological chemistry* **276**: 13587-13592.

Hendzel, M. J., M. A. Lever, E. Crawford & J. P. Th'ng, (2004) The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *The Journal of biological chemistry* **279**: 20028-20034.

Hennig, B. P. & T. Fischer, (2013) The great repression: chromatin and cryptic transcription. *Transcription* **4**: 97-101.

Hertz-Fowler, C., L. M. Figueiredo, M. A. Quail, M. Becker, A. Jackson, N. Bason, K. Brooks, C. Churcher, S. Fahkro, I. Goodhead, P. Heath, M. Kartvelishvili, K. Mungall, D. Harris, H. Hauser, M. Sanders, D. Saunders, K. Seeger, S. Sharp, J. E. Taylor, D. Walker, B. White, R. Young, G. A. Cross, G. Rudenko, J. D. Barry, E. J. Louis & M. Berriman, (2008) Telomeric expression sites are highly conserved in Trypanosoma brucei. *PloS one* **3**: e3527.

Hirumi, H. & K. Hirumi, (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* **75**: 985-989.

Hitchcock, R. A., S. Thomas, D. A. Campbell & N. R. Sturm, (2007) The promoter and transcribed regions of the Leishmania tarentolae spliced leader RNA gene array are devoid of nucleosomes. *BMC microbiology* **7**: 44.

Horn, D. & G. A. Cross, (1995) A developmentally regulated position effect at a telomeric locus in Trypanosoma brucei. *Cell* **83**: 555-561.

Horn, D. & G. A. Cross, (1997) Position-dependent and promoter-specific regulation of gene expression in Trypanosoma brucei. *The EMBO journal* **16**: 7422-7431.

Hotz, H. R., S. Biebinger, J. Flaspohler & C. Clayton, (1998) PARP gene expression: control at many levels. *Molecular and biochemical parasitology* **91**: 131-143.

Hotz, H. R., C. Hartmann, K. Huober, M. Hug & C. Clayton, (1997) Mechanisms of developmental regulation in Trypanosoma brucei: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic acids research* **25**: 3017-3026.

Huang, J. & L. H. van der Ploeg, (1991) Maturation of polycistronic pre-mRNA in Trypanosoma brucei: analysis of trans splicing and poly(A) addition at nascent RNA transcripts from the hsp70 locus. *Molecular and cellular biology* **11**: 3180-3190.

Hughes, K., M. Wand, L. Foulston, R. Young, K. Harley, S. Terry, K. Ersfeld & G. Rudenko, (2007) A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *The EMBO journal* **26**: 2400-2410.

Imboden, M. A., P. W. Laird, M. Affolter & T. Seebeck, (1987) Transcription of the intergenic regions of the tubulin gene cluster of Trypanosoma brucei: evidence for a polycistronic transcription unit in a eukaryote. *Nucleic acids research* **15**: 7357-7368.

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

Izzo, A., K. Kamieniarz & R. Schneider, (2008) The histone H1 family: specific members, specific functions? *Biological chemistry* **389**: 333-343.

Jackson, D. G., M. J. Owen & H. P. Voorheis, (1985) A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from Trypanosoma brucei. *The Biochemical journal* **230**: 195-202.

Jae, N., P. Wang, T. Gu, M. Huhn, Z. Palfi, H. Urlaub & A. Bindereif, (2010) Essential role of a trypanosome U4-specific Sm core protein in small nuclear ribonucleoprotein assembly and splicing. *Eukaryotic cell* **9**: 379-386.

Jamonneau, V., H. Ilboudo, J. Kabore, D. Kaba, M. Koffi, P. Solano, A. Garcia, D. Courtin, C. Laveissiere, K. Lingue, P. Buscher & B. Bucheton, (2012) Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal. *PLoS neglected tropical diseases* **6**: e1691.

Janzen, C. J., J. P. Fernandez, H. Deng, R. Diaz, S. B. Hake & G. A. Cross, (2006a) Unusual histone modifications in Trypanosoma brucei. *FEBS letters* **580**: 2306-2310.

Janzen, C. J., S. B. Hake, J. E. Lowell & G. A. Cross, (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.

Janzen, C. J., F. Lander, O. Dreesen & G. A. Cross, (2004) Telomere length regulation and transcriptional silencing in KU80-deficient Trypanosoma brucei. *Nucleic acids research* **32**: 6575-6584.

Jehi, S. E., X. Li, R. Sandhu, F. Ye, I. Benmerzouga, M. Zhang, Y. Zhao & B. Li, (2014a) Suppression of subtelomeric VSG switching by Trypanosoma brucei TRF requires its TTAGGG repeat-binding activity. *Nucleic acids research* **42**: 12899-12911.

Jehi, S. E., F. Wu & B. Li, (2014b) Trypanosoma brucei TIF2 suppresses VSG switching by maintaining subtelomere integrity. *Cell research*.

Jensen, B. C., G. Ramasamy, E. J. Vasconcelos, N. T. Ingolia, P. J. Myler & M. Parsons, (2014) Extensive stage-regulation of translation revealed by ribosome profiling of Trypanosoma brucei. *BMC genomics* **15**: 911.

Jensen, B. C., D. Sivam, C. T. Kifer, P. J. Myler & M. Parsons, (2009) Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. *BMC genomics* **10**: 482.

Jensen, R. E. & P. T. Englund, (2012) Network news: the replication of kinetoplast DNA. *Annu Rev Microbiol* **66**: 473-491.

Jensen, T. H., A. Jacquier & D. Libri, (2013) Dealing with pervasive transcription. *Molecular cell* **52**: 473-484.

Johnson, J. G. & G. A. Cross, (1979) Selective cleavage of variant surface glycoproteins from Trypanosoma brucei. *The Biochemical journal* **178**: 689-697.

Kabani, S., K. Fenn, A. Ross, A. Ivens, T. K. Smith, P. Ghazal & K. Matthews, (2009) Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. *BMC genomics* **10**: 427.

Kaplan, C. D., L. Laprade & F. Winston, (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**: 1096-1099.

Kassem, A., E. Pays & L. Vanhamme, (2014) Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in Trypanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 8943-8948.

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

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

Kenzelmann, M., S. Maertens, M. Hergenhahn, S. Kueffer, A. Hotz-Wagenblatt, L. Li, S. Wang, C. Ittrich, T. Lemberger, R. Arribas, S. Jonnakuty, M. C. Hollstein, W. Schmid, N. Gretz,

H. J. Grone & G. Schutz, (2007) Microarray analysis of newly synthesized RNA in cells and animals. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 6164-6169.
Kim, H. S., S. H. Park, A. Gunzl & G. A. Cross, (2013) MCM-BP is required for repression of life-cycle specific genes transcribed by RNA polymerase I in the mammalian infectious form of Trypanosoma brucei. *PloS one* **8**: e57001.
Knusel, S. & I. Roditi, (2013) Insights into the regulation of GPEET procyclin during differentiation from early to late procyclic forms of Trypanosoma brucei. *Molecular and biochemical parasitology* **191**: 66-74.
Koenig-Martin, E., M. Yamage & I. Roditi, (1992) A procyclin-associated gene in Trypanosoma brucei encodes a polypeptide related to ESAG 6 and 7 proteins. *Molecular and biochemical parasitology* **55**: 135-145.
Kolev, N. G., J. B. Franklin, S. Carmi, H. Shi, S. Michaeli & C. Tschudi, (2010) The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS pathogens* **6**: e1001090.
Kolev, N. G., K. Ramey-Butler, G. A. Cross, E. Ullu & C. Tschudi, (2012) Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. *Science* **338**: 1352-1353.
Kolev, N. G., E. Ullu & C. Tschudi, (2014) The emerging role of RNA-binding proteins in the life cycle of Trypanosoma brucei. *Cellular microbiology* **16**: 482-489.
Kooter, J. M. & P. Borst, (1984) Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic acids research* **12**: 9457-9472.
Kramer, S., (2012) Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. *Molecular and biochemical parasitology* **181**: 61-72.
Lam, Y. W. & L. Trinkle-Mulcahy, (2015) New insights into nucleolar structure and function. *F1000prime reports* **7**: 48.
Lamont, G. S., R. S. Tucker & G. A. Cross, (1986) Analysis of antigen switching rates in Trypanosoma brucei. *Parasitology* **92 (Pt 2)**: 355-367.
Landeira, D., J. M. Bart, D. Van Tyne & M. Navarro, (2009) Cohesin regulates VSG monoallelic expression in trypanosomes. *The Journal of cell biology* **186**: 243-254.
Landeira, D. & M. Navarro, (2007) Nuclear repositioning of the VSG promoter during developmental silencing in Trypanosoma brucei. *The Journal of cell biology* **176**: 133-139.
Langmead, B., C. Trapnell, M. Pop & S. L. Salzberg, (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.
Laufer, G. & A. Gunzl, (2001) In-vitro competition analysis of procyclin gene and variant surface glycoprotein gene expression site transcription in Trypanosoma brucei. *Molecular and biochemical parasitology* **113**: 55-65.
Laybourn, P. J. & J. T. Kadonaga, (1991) Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* **254**: 238-245.
LeBowitz, J. H., H. Q. Smith, L. Rusche & S. M. Beverley, (1993) Coupling of poly(A) site selection and trans-splicing in Leishmania. *Genes & development* **7**: 996-1007.
Lecordier, L., D. Walgraffe, S. Devaux, P. Poelvoorde, E. Pays & L. Vanhamme, (2005) Trypanosoma brucei RNA interference in the mammalian host. *Molecular and biochemical parasitology* **140**: 127-131.
Lee, J. S., E. Smith & A. Shilatifard, (2010) The language of histone crosstalk. *Cell* **142**: 682-685.
Lever, M. A., J. P. Th'ng, X. Sun & M. J. Hendzel, (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature* **408**: 873-876.
Li, B., (2015) DNA double-strand breaks and telomeres play important roles in trypanosoma brucei antigenic variation. *Eukaryotic cell* **14**: 196-205.

Li, B., A. Espinal & G. A. Cross, (2005) Trypanosome telomeres are protected by a homologue of mammalian TRF2. *Molecular and cellular biology* **25**: 5011-5021.

Li, C., J. E. Mueller, M. Elfline & M. Bryk, (2008) Linker histone H1 represses recombination at the ribosomal DNA locus in the budding yeast Saccharomyces cerevisiae. *Molecular microbiology* **67**: 906-919.

Lin, R., R. G. Cook & C. D. Allis, (1991) Proteolytic removal of core histone amino termini and dephosphorylation of histone H1 correlate with the formation of condensed chromatin and transcriptional silencing during Tetrahymena macronuclear development. *Genes & development* **5**: 1601-1610.

Liniger, M., K. Bodenmuller, E. Pays, S. Gallati & I. Roditi, (2001) Overlapping sense and antisense transcription units in Trypanosoma brucei. *Molecular microbiology* **40**: 869-878.

Lopez-Farfan, D., J. M. Bart, D. I. Rojas-Barros & M. Navarro, (2014) SUMOylation by the E3 ligase TbSIZ1/PIAS1 positively regulates VSG expression in Trypanosoma brucei. *PLoS pathogens* **10**: e1004545.

Lopez, R., B. Sarg, H. Lindner, S. Bartolome, I. Ponte, P. Suau & A. Roque, (2015) Linker histone partial phosphorylation: effects on secondary structure and chromatin condensation. *Nucleic acids research* **43**: 4463-4476.

Lowell, J. E. & G. A. Cross, (2004) A variant histone H3 is enriched at telomeres in Trypanosoma brucei. *Journal of cell science* **117**: 5937-5947.

Lowell, J. E., F. Kaiser, C. J. Janzen & G. A. Cross, (2005) Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in Trypanosoma brucei. *Journal of cell science* **118**: 5721-5730.

Lu, X., S. N. Wontakal, A. V. Emelyanov, P. Morcillo, A. Y. Konev, D. V. Fyodorov & A. I. Skoultchi, (2009) Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes & development* **23**: 452-465.

Luger, K., M. L. Dechassa & D. J. Tremethick, (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews. Molecular cell biology* **13**: 436-447.

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

Luger, K. & T. J. Richmond, (1998) The histone tails of the nucleosome. *Current opinion in genetics & development* **8**: 140-146.

Luijsterburg, M. S., M. F. White, R. van Driel & R. T. Dame, (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* **43**: 393-418.

Lukes, J., T. Skalicky, J. Tyc, J. Votypka & V. Yurchenko, (2014) Evolution of parasitism in kinetoplastid flagellates. *Molecular and biochemical parasitology* **195**: 115-122.

Lundin, C., M. North, K. Erixon, K. Walters, D. Jenssen, A. S. Goldman & T. Helleday, (2005) Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic acids research* **33**: 3799-3811.

Lythgoe, K. A., L. J. Morrison, A. F. Read & J. D. Barry, (2007) Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 8095-8100.

Ma, W., J. W. Westmoreland, D. A. Gordenin & M. A. Resnick, (2011) Alkylation base damage is converted into repairable double-strand breaks and complex intermediates in G2 cells lacking AP endonuclease. *PLoS genetics* **7**: e1002059.

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

Maeshima, K., S. Hihara & M. Eltsov, (2010) Chromatin structure: does the 30-nm fibre exist in vivo? *Current opinion in cell biology* **22**: 291-297.
Mair, G., H. Shi, H. Li, A. Djikeng, H. O. Aviles, J. R. Bishop, F. H. Falcone, C. Gavrilescu, J. L. Montgomery, M. I. Santori, L. S. Stern, Z. Wang, E. Ullu & C. Tschudi, (2000) A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* **6**: 163-169.
Mandava, V., J. P. Fernandez, H. Deng, C. J. Janzen, S. B. Hake & G. A. Cross, (2007) Histone modifications in Trypanosoma brucei. *Molecular and biochemical parasitology* **156**: 41-50.
Manful, T., A. Fadda & C. Clayton, (2011) The role of the 5'-3' exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA* **17**: 2039-2047.
Mansfield, J. M. & D. M. Paulnock, (2005) Regulation of innate and acquired immunity in African trypanosomiasis. *Parasite immunology* **27**: 361-371.
Marcello, L. & J. D. Barry, (2007) Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome research* **17**: 1344-1352.
Martinez-Calvillo, S., J. C. Vizuet-de-Rueda, L. E. Florencio-Martinez, R. G. Manning-Cela & E. E. Figueroa-Angulo, (2010) Gene expression in trypanosomatid parasites. *Journal of biomedicine & biotechnology* **2010**: 525241.
Martinez-Calvillo, S., S. Yan, D. Nguyen, M. Fox, K. Stuart & P. J. Myler, (2003) Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Molecular cell* **11**: 1291-1299.
Marzluff, W. F., (2005) Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Current opinion in cell biology* **17**: 274-280.
Masina, S., H. Zangger, D. Rivier & N. Fasel, (2007) Histone H1 regulates chromatin condensation in Leishmania parasites. *Experimental parasitology* **116**: 83-87.
Matthews, K. R., R. McCulloch & L. J. Morrison, (2015) The within-host dynamics of African trypanosome infections. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**.
Mazurier, F., A. Fontanellas, S. Salesse, L. Taine, S. Landriau, F. Moreau-Gaudry, J. Reiffers, B. Peault, J. P. Di Santo & H. de Verneuil, (1999) A novel immunodeficient mouse model--RAG2 x common cytokine receptor gamma chain double mutants--requiring exogenous cytokine administration for human hematopoietic stem cell engraftment. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* **19**: 533-541.
McAndrew, M., S. Graham, C. Hartmann & C. Clayton, (1998) Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Experimental parasitology* **90**: 65-76.
McCulloch, R. & J. D. Barry, (1999) A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. *Genes & development* **13**: 2875-2888.
McCulloch, R. & M. C. Field, (2015) Quantitative sequencing confirms VSG diversity as central to immune evasion by Trypanosoma brucei. *Trends in parasitology*.
McCulloch, R. & D. Horn, (2009) What has DNA sequencing revealed about the VSG expression sites of African trypanosomes? *Trends in parasitology* **25**: 359-363.
McCulloch, R., L. J. Morrison & J. P. Hall, (2015) DNA Recombination Strategies During Antigenic Variation in the African Trypanosome. *Microbiology spectrum* **3**.
McCulloch, R., G. Rudenko & P. Borst, (1997) Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Molecular and cellular biology* **17**: 833-843.
Mechali, M., K. Yoshida, P. Coulombe & P. Pasero, (2013) Genetic and epigenetic determinants of DNA replication origins, position and activation. *Current opinion in genetics & development* **23**: 124-131.

Melville, S. E., V. Leech, M. Navarro & G. A. Cross, (2000) The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. *Molecular and biochemical parasitology* **111**: 261-273.

Michaeli, S., D. Podell, N. Agabian & E. Ullu, (1992) The 7SL RNA homologue of Trypanosoma brucei is closely related to mammalian 7SL RNA. *Molecular and biochemical parasitology* **51**: 55-64.

Michels, P. A., A. Y. Liu, A. Bernards, P. Sloof, M. M. Van der Bijl, A. H. Schinkel, H. H. Menke, P. Borst, G. H. Veeneman, M. C. Tromp & J. H. Van Boom, (1983) Activation of the genes for variant surface glycoproteins 117 and 118 in Trypanosoma brucei. *Journal of molecular biology* **166**: 537-556.

Militello, K. T., P. Wang, S. K. Jayakar, R. L. Pietrasik, C. D. Dupont, K. Dodd, A. M. King & P. R. Valenti, (2008) African trypanosomes contain 5-methylcytosine in nuclear DNA. *Eukaryotic cell* **7**: 2012-2016.

Miller, C., B. Schwalb, K. Maier, D. Schulz, S. Dumcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dolken, D. E. Martin, A. Tresch & P. Cramer, (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**: 458.

Miller, E. N., L. M. Allan & M. J. Turner, (1984) Mapping of antigenic determinants within peptides of a variant surface glycoprotein of Trypanosoma brucei. *Molecular and biochemical parasitology* **13**: 309-322.

Miller, E. N. & M. J. Turner, (1981) Analysis of antigenic types appearing in first relapse populations of clones of Trypanosoma brucei. *Parasitology* **82**: 63-80.

Miller, M. R., K. J. Robinson, M. D. Cleary & C. Q. Doe, (2009) TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Methods* **6**: 439-441.

Misteli, T., (2005) Concepts in nuclear architecture. *Bioessays* **27**: 477-487.

Misteli, T., (2014) The long reach of telomeres. *Genes & development* **28**: 2445-2446.

Misteli, T., A. Gunjan, R. Hock, M. Bustin & D. T. Brown, (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature* **408**: 877-881.

Mogk, S., A. Meiwes, S. Shtopel, U. Schraermeyer, M. Lazarus, B. Kubata, H. Wolburg & M. Duszenko, (2014) Cyclical appearance of African trypanosomes in the cerebrospinal fluid: new insights in how trypanosomes enter the CNS. *PloS one* **9**: e91372.

Morey, J. S. & F. M. Van Dolah, (2013) Global analysis of mRNA half-lives and de novo transcription in a dinoflagellate, Karenia brevis. *PloS one* **8**: e66347.

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

Morrison, W. I., S. J. Black, J. Paris, C. A. Hinson & P. W. Wells, (1982) Protective immunity and specificity of antibody responses elicited in cattle by irradiated Trypanosoma brucei. *Parasite immunology* **4**: 395-407.

Mowatt, M. R. & C. E. Clayton, (1987) Developmental regulation of a novel repetitive protein of Trypanosoma brucei. *Molecular and cellular biology* **7**: 2838-2844.

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

Muhich, M. L. & J. C. Boothroyd, (1988) Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. *Molecular and cellular biology* **8**: 3837-3846.

Munoz-Jordan, J. L., K. P. Davies & G. A. Cross, (1996) Stable expression of mosaic coats of variant surface glycoproteins in Trypanosoma brucei. *Science* **272**: 1795-1797.

Murga, M., I. Jaco, Y. Fan, R. Soria, B. Martinez-Pastor, M. Cuadrado, S. M. Yang, M. A. Blasco, A. I. Skoultchi & O. Fernandez-Capetillo, (2007) Global chromatin compaction limits the strength of the DNA damage response. *The Journal of cell biology* **178**: 1101-1108.

Najafabadi, H. S., Z. Lu, C. MacPherson, V. Mehta, V. Adoue, T. Pastinen & R. Salavati, (2013) Global identification of conserved post-transcriptional regulatory programs in trypanosomatids. *Nucleic acids research* **41**: 8591-8600.

Nakaar, V., A. O. Dare, D. Hong, E. Ullu & C. Tschudi, (1994) Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. *Molecular and cellular biology* **14**: 6736-6742.

Nalabothula, N., G. McVicker, J. Maiorano, R. Martin, J. K. Pritchard & Y. N. Fondufe-Mittendorf, (2014) The chromatin architectural proteins HMGD1 and H1 bind reciprocally and have opposite effects on chromatin structure and gene regulation. *BMC genomics* **15**: 92.

Narayanan, M. S., M. Kushwaha, K. Ersfeld, A. Fullbrook, T. M. Stanne & G. Rudenko, (2011) NLP is a novel transcription regulator involved in VSG expression site control in Trypanosoma brucei. *Nucleic acids research* **39**: 2018-2031.

Narayanan, M. S. & G. Rudenko, (2013) TDP1 is an HMG chromatin protein facilitating RNA polymerase I transcription in African trypanosomes. *Nucleic acids research* **41**: 2981-2992.

Narlikar, G. J., R. Sundaramoorthy & T. Owen-Hughes, (2013) Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**: 490-503.

Natesan, S. K., L. Peacock, K. Matthews, W. Gibson & M. C. Field, (2007) Activation of endocytosis as an adaptation to the mammalian host by trypanosomes. *Eukaryotic cell* **6**: 2029-2037.

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

Navarro, M., X. Penate & D. Landeira, (2007) Nuclear architecture underlying gene expression in Trypanosoma brucei. *Trends Microbiol* **15**: 263-270.

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

Nguyen, T. N., L. S. Muller, S. H. Park, T. N. Siegel & A. Gunzl, (2014) Promoter occupancy of the basal class I transcription factor A differs strongly between active and silent VSG expression sites in Trypanosoma brucei. *Nucleic acids research* **42**: 3164-3176.

Nguyen, T. N., B. N. Nguyen, J. H. Lee, A. K. Panigrahi & A. Gunzl, (2012) Characterization of a novel class I transcription factor A (CITFA) subunit that is indispensable for transcription by the multifunctional RNA polymerase I of Trypanosoma brucei. *Eukaryotic cell* **11**: 1573-1581.

Nguyen, T. N., B. Schimanski & A. Gunzl, (2007) Active RNA polymerase I of Trypanosoma brucei harbors a novel subunit essential for transcription. *Molecular and cellular biology* **27**: 6254-6263.

Nguyen, T. N., B. Schimanski, A. Zahn, B. Klumpp & A. Gunzl, (2006) Purification of an eight subunit RNA polymerase I complex in Trypanosoma brucei. *Molecular and biochemical parasitology* **149**: 27-37.

Nightingale, K., S. Dimitrov, R. Reeves & A. P. Wolffe, (1996) Evidence for a shared structural role for HMG1 and linker histones B4 and H1 in organizing chromatin. *The EMBO journal* **15**: 548-561.

Nikolova, T., M. Ensminger, M. Lobrich & B. Kaina, (2010) Homologous recombination protects mammalian cells from replication-associated DNA double-strand breaks arising in response to methyl methanesulfonate. *DNA Repair (Amst)* **9**: 1050-1063.

Nilsson, D., K. Gunasekera, J. Mani, M. Osteras, L. Farinelli, L. Baerlocher, I. Roditi & T. Ochsenreiter, (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS pathogens* **6**: e1001037.

Nishino, Y., M. Eltsov, Y. Joti, K. Ito, H. Takata, Y. Takahashi, S. Hihara, A. S. Frangakis, N. Imamoto, T. Ishikawa & K. Maeshima, (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal* **31**: 1644-1653.

Noble, W. S., (2009) How does multiple testing correction work? *Nature biotechnology* **27**: 1135-1137.

Noll, M. & R. D. Kornberg, (1977) Action of micrococcal nuclease on chromatin and the location of histone H1. *Journal of molecular biology* **109**: 393-404.

Ogawa, Y., S. Aizawa, H. Shirakawa & M. Yoshida, (1995) Stimulation of transcription accompanying relaxation of chromatin structure in cells overexpressing high mobility group 1 protein. *The Journal of biological chemistry* **270**: 9272-9280.

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

Oguchi, Y., E. Kummer, F. Seyffer, M. Berynskyy, B. Anstett, R. Zahn, R. C. Wade, A. Mogk & B. Bukau, (2012) A tightly regulated molecular toggle controls AAA+ disaggregase. *Nature structural & molecular biology* **19**: 1338-1346.

Ong, H. B., N. Sienkiewicz, S. Wyllie, S. Patterson & A. H. Fairlamb, (2013) Trypanosoma brucei (UMP synthase null mutants) are avirulent in mice, but recover virulence upon prolonged culture in vitro while retaining pyrimidine auxotrophy. *Molecular microbiology* **90**: 443-455.

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

Padeken, J. & P. Heun, (2014) Nucleolus and nuclear periphery: velcro for heterochromatin. *Current opinion in cell biology* **28**: 54-60.

Pan, W., O. Ogunremi, G. Wei, M. Shi & H. Tabel, (2006) CR3 (CD11b/CD18) is the major macrophage receptor for IgM antibody-mediated phagocytosis of African trypanosomes: diverse effect on subsequent synthesis of tumor necrosis factor alpha and nitric oxide. *Microbes Infect* **8**: 1209-1218.

Pandya, U. M., R. Sandhu & B. Li, (2013) Silencing subtelomeric VSGs by Trypanosoma brucei RAP1 at the insect stage involves chromatin structure changes. *Nucleic acids research* **41**: 7673-7682.

Park, S. H., T. N. Nguyen & A. Gunzl, (2012) Development of an efficient in vitro transcription system for bloodstream form Trypanosoma brucei reveals life cycle-independent functionality of class I transcription factor A. *Molecular and biochemical parasitology* **181**: 29-36.

Park, S. H., T. N. Nguyen, J. K. Kirkham, J. H. Lee & A. Gunzl, (2011) Transcription by the multifunctional RNA polymerase I in Trypanosoma brucei functions independently of RPB7. *Molecular and biochemical parasitology* **180**: 35-42.

Patterton, H. G., C. C. Landel, D. Landsman, C. L. Peterson & R. T. Simpson, (1998) The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. *The Journal of biological chemistry* **273**: 7268-7276.

Pays, E., H. Coquelet, P. Tebabi, A. Pays, D. Jefferies, M. Steinert, E. Koenig, R. O. Williams & I. Roditi, (1990) Trypanosoma brucei: constitutive activity of the VSG and procyclin gene promoters. *The EMBO journal* **9**: 3145-3151.

Pays, E., J. Hanocq-Quertier, F. Hanocq, S. Van Assel, D. Nolan & S. Rolin, (1993) Abrupt RNA changes precede the first cell division during the differentiation of Trypanosoma brucei bloodstream forms into procyclic forms in vitro. *Molecular and biochemical parasitology* **61**: 107-114.

Peacock, L., V. Ferris, R. Sharma, J. Sunter, M. Bailey, M. Carrington & W. Gibson, (2011) Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly.

Proceedings of the National Academy of Sciences of the United States of America **108**: 3671-3676.

Pedram, M. & J. E. Donelson, (1999) The anatomy and transcription of a monocistronic expression site for a metacyclic variant surface glycoprotein gene in Trypanosoma brucei. *The Journal of biological chemistry* **274**: 16876-16883.

Perez-Morga, D., A. Amiguet-Vercher, D. Vermijlen & E. Pays, (2001) Organization of telomeres during the cell and life cycles of Trypanosoma brucei. *The Journal of eukaryotic microbiology* **48**: 221-226.

Perez-Ortin, J. E., P. Alepuz, S. Chavez & M. Choder, (2013) Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. *Journal of molecular biology* **425**: 3750-3775.

Pfaffl, M. W., (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research* **29**: e45.

Pham, V. P., C. C. Qi & K. M. Gottesdiener, (1996) A detailed mutational analysis of the VSG gene expression site promoter. *Molecular and biochemical parasitology* **75**: 241-254.

Pombo, A. & N. Dillon, (2015) Three-dimensional genome architecture: players and mechanisms. *Nature reviews. Molecular cell biology* **16**: 245-257.

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

Preuss, S. & C. S. Pikaard, (2007) rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. *Biochimica et biophysica acta* **1769**: 383-392.

Proudfoot, C. & R. McCulloch, (2005) Distinct roles for two RAD51-related genes in Trypanosoma brucei antigenic variation. *Nucleic acids research* **33**: 6906-6919.

Proudfoot, C. & R. McCulloch, (2006) Trypanosoma brucei DMC1 does not act in DNA recombination, repair or antigenic variation in bloodstream stage cells. *Molecular and biochemical parasitology* **145**: 245-253.

Queiroz, R., C. Benz, K. Fellenberg, J. D. Hoheisel & C. Clayton, (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC genomics* **10**: 495.

Rabani, M., J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit & A. Regev, (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature biotechnology* **29**: 436-442.

Radle, B., A. J. Rutkowski, Z. Ruzsics, C. C. Friedel, U. H. Koszinowski & L. Dolken, (2013) Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *Journal of visualized experiments : JoVE*.

Ramachandran, A., M. Omar, P. Cheslock & G. R. Schnitzler, (2003) Linker histone H1 modulates nucleosome remodeling by human SWI/SNF. *The Journal of biological chemistry* **278**: 48590-48601.

Ramakrishnan, V., (1997) Histone structure and the organization of the nucleosome. *Annual review of biophysics and biomolecular structure* **26**: 83-112.

Ramakrishnan, V., J. T. Finch, V. Graziano, P. L. Lee & R. M. Sweet, (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**: 219-223.

Ramey-Butler, K., E. Ullu, N. G. Kolev & C. Tschudi, (2015) Synchronous expression of individual metacyclic variant surface glycoprotein genes in Trypanosoma brucei. *Molecular and biochemical parasitology* **200**: 1-4.

Raz, B., M. Iten, Y. Grether-Buhler, R. Kaminsky & R. Brun, (1997) The Alamar Blue assay to determine drug sensitivity of African trypanosomes (T.b. rhodesiense and T.b. gambiense) in vitro. *Acta tropica* **68**: 139-147.

Respuela, P., M. Ferella, A. Rada-Iglesias & L. Aslund, (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in Trypanosoma cruzi. *The Journal of biological chemistry* **283**: 15884-15892.

Reuner, B., E. Vassella, B. Yutzy & M. Boshart, (1997) Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Molecular and biochemical parasitology* **90**: 269-280.

Reynolds, D., L. Cliffe, K. U. Forstner, C. C. Hon, T. N. Siegel & R. Sabatini, (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic acids research* **42**: 9717-9729.

Robinson, N. P., N. Burman, S. E. Melville & J. D. Barry, (1999) Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. *Molecular and cellular biology* **19**: 5839-5846.

Robinson, P. J. & D. Rhodes, (2006) Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Current opinion in structural biology* **16**: 336-343.

Roditi, I., A. Furger, S. Ruepp, N. Schurch & P. Butikofer, (1998) Unravelling the procyclin coat of Trypanosoma brucei. *Molecular and biochemical parasitology* **91**: 117-130.

Roditi, I., H. Schwarz, T. W. Pearson, R. P. Beecroft, M. K. Liu, J. P. Richardson, H. J. Buhring, J. Pleiss, R. Bulow, R. O. Williams & et al., (1989) Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of Trypanosoma brucei. *The Journal of cell biology* **108**: 737-746.

Ross, R. & D. Thomson, (1910) A Case of Sleeping Sickness showing Regular Periodical Increase of the Parasites Disclosed. *British medical journal* **1**: 1544-1545.

Roth, S. Y. & C. D. Allis, (1992) Chromatin condensation: does histone H1 dephosphorylation play a role? *Trends in biochemical sciences* **17**: 93-98.

Rout, M. P. & M. C. Field, (2001) Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component. *The Journal of biological chemistry* **276**: 38261-38271.

Rudenko, G., (2010) Epigenetics and transcriptional control in African trypanosomes. *Essays Biochem* **48**: 201-219.

Rudenko, G., (2011) African trypanosomes: the genome and adaptations for immune evasion. *Essays Biochem* **51**: 47-62.

Rudenko, G., D. Bishop, K. Gottesdiener & L. H. Van der Ploeg, (1989) Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form Trypanosoma brucei. *The EMBO journal* **8**: 4259-4263.

Rudenko, G., H. M. Chung, V. P. Pham & L. H. Van der Ploeg, (1991) RNA polymerase I can mediate expression of CAT and neo protein-coding genes in Trypanosoma brucei. *The EMBO journal* **10**: 3387-3397.

Rudenko, G., S. Le Blancq, J. Smith, M. G. Lee, A. Rattray & L. H. Van der Ploeg, (1990) Procyclic acidic repetitive protein (PARP) genes located in an unusually small alpha-amanitin-resistant transcription unit: PARP promoter activity assayed by transient DNA transfection of Trypanosoma brucei. *Molecular and cellular biology* **10**: 3492-3504.

Ruepp, S., A. Furger, U. Kurath, C. K. Renggli, A. Hemphill, R. Brun & I. Roditi, (1997) Survival of Trypanosoma brucei in the tsetse fly is enhanced by the expression of specific forms of procyclin. *The Journal of cell biology* **137**: 1369-1379.

Sabo, A., T. R. Kress, M. Pelizzola, S. de Pretis, M. M. Gorski, A. Tesi, M. J. Morelli, P. Bora, M. Doni, A. Verrecchia, C. Tonelli, G. Faga, V. Bianchi, A. Ronchi, D. Low, H. Muller, E. Guccione, S. Campaner & B. Amati, (2014) Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* **511**: 488-492.

Salmon, D., F. Paturiaux-Hanocq, P. Poelvoorde, L. Vanhamme & E. Pays, (2005) Trypanosoma brucei: growth differences in different mammalian sera are not due to the species-specificity of transferrin. *Experimental parasitology* **109**: 188-194.

Sancho, M., E. Diani, M. Beato & A. Jordan, (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS genetics* **4**: e1000227.

Scherf, A., J. J. Lopez-Rubio & L. Riviere, (2008) Antigenic variation in Plasmodium falciparum. *Annu Rev Microbiol* **62**: 445-470.

Schimanski, B., T. N. Nguyen & A. Gunzl, (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.

Schlimme, W., M. Burri, K. Bender, B. Betschart & H. Hecker, (1993) Trypanosoma brucei brucei: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. *Parasitology* **107 (Pt 3)**: 237-247.

Schurch, N., A. Furger, U. Kurath & I. Roditi, (1997) Contributions of the procyclin 3' untranslated region and coding region to the regulation of expression in bloodstream forms of Trypanosoma brucei. *Molecular and biochemical parasitology* **89**: 109-121.

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

Seyffer, F., E. Kummer, Y. Oguchi, J. Winkler, M. Kumar, R. Zahn, V. Sourjik, B. Bukau & A. Mogk, (2012) Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces. *Nature structural & molecular biology* **19**: 1347-1355.

Sharma, R., E. Gluenz, L. Peacock, W. Gibson, K. Gull & M. Carrington, (2009) The heart of darkness: growth and form of Trypanosoma brucei in the tsetse fly. *Trends in parasitology* **25**: 517-524.

Sheader, K., S. Vaughan, J. Minchin, K. Hughes, K. Gull & G. Rudenko, (2005) Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 8716-8721.

Shen, X. & M. A. Gorovsky, (1996) Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell* **86**: 475-483.

Shen, X., L. Yu, J. W. Weir & M. A. Gorovsky, (1995) Linker histones are not essential and affect chromatin condensation in vivo. *Cell* **82**: 47-56.

Shimamura, A., M. Sapp, A. Rodriguez-Campos & A. Worcel, (1989) Histone H1 represses transcription from minichromosomes assembled in vitro. *Molecular and cellular biology* **9**: 5573-5584.

Shultz, L. D., F. Ishikawa & D. L. Greiner, (2007) Humanized mice in translational biomedical research. *Nature reviews. Immunology* **7**: 118-130.

Sidaway-Lee, K., M. J. Costa, D. A. Rand, B. Finkenstadt & S. Penfield, (2014) Direct measurement of transcription rates reveals multiple mechanisms for configuration of the Arabidopsis ambient temperature response. *Genome biology* **15**: R45.

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

Siegel, T. N., D. R. Hekstra, X. Wang, S. Dewell & G. A. Cross, (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., T. Kawahara, J. A. Degrasse, C. J. Janzen, D. Horn & G. A. Cross, (2008) Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in Trypanosoma brucei. *Molecular microbiology* **67**: 762-771.

Simpson, R. T., (1978) Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* **17**: 5524-5531.

Siriaco, G., R. Deuring, M. Chioda, P. B. Becker & J. W. Tamkun, (2009) Drosophila ISWI regulates the association of histone H1 with interphase chromosomes in vivo. *Genetics* **182**: 661-669.

Sloof, P., A. de Haan, W. Eier, M. van Iersel, E. Boel, H. van Steeg & R. Benne, (1992) The nucleotide sequence of the variable region in Trypanosoma brucei completes the sequence analysis of the maxicircle component of mitochondrial kinetoplast DNA. *Molecular and biochemical parasitology* **56**: 289-299.

Smirlis, D., S. N. Bisti, E. Xingi, G. Konidou, M. Thiakaki & K. P. Soteriadou, (2006) Leishmania histone H1 overexpression delays parasite cell-cycle progression, parasite differentiation and reduces Leishmania infectivity in vivo. *Molecular microbiology* **60**: 1457-1473.

Smith, T. K., N. Vasileva, E. Gluenz, S. Terry, N. Portman, S. Kramer, M. Carrington, S. Michaeli, K. Gull & G. Rudenko, (2009) Blocking variant surface glycoprotein synthesis in Trypanosoma brucei triggers a general arrest in translation initiation. *PloS one* **4**: e7532.

Stanne, T., M. S. Narayanan, S. Ridewood, A. Ling, K. Witmer, M. Kushwaha, S. Wiesler, B. Wickstead, J. Wood & G. Rudenko, (2015) Identification of the ISWI chromatin remodeling complex of the early branching eukaryote Trypanosoma brucei. *The Journal of biological chemistry*.

Stanne, T. M., M. Kushwaha, M. Wand, J. E. Taylor & G. Rudenko, (2011) TbISWI regulates multiple polymerase I (Pol I)-transcribed loci and is present at Pol II transcription boundaries in Trypanosoma brucei. *Eukaryotic cell* **10**: 964-976.

Stanne, T. M. & G. Rudenko, (2010) Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. *Eukaryotic cell* **9**: 136-147.

Strahl, B. D. & C. D. Allis, (2000) The language of covalent histone modifications. *Nature* **403**: 41-45.

Stros, M., (2010) HMGB proteins: interactions with DNA and chromatin. *Biochimica et biophysica acta* **1799**: 101-113.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander & J. P. Mesirov, (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15545-15550.

Talbert, P. B., K. Ahmad, G. Almouzni, J. Ausio, F. Berger, P. L. Bhalla, W. M. Bonner, W. Z. Cande, B. P. Chadwick, S. W. Chan, G. A. Cross, L. Cui, S. I. Dimitrov, D. Doenecke, J. M. Eirin-Lopez, M. A. Gorovsky, S. B. Hake, B. A. Hamkalo, S. Holec, S. E. Jacobsen, K. Kamieniarz, S. Khochbin, A. G. Ladurner, D. Landsman, J. A. Latham, B. Loppin, H. S. Malik, W. F. Marzluff, J. R. Pehrson, J. Postberg, R. Schneider, M. B. Singh, M. M. Smith, E. Thompson, M. E. Torres-Padilla, D. J. Tremethick, B. M. Turner, J. H. Waterborg, H. Wollmann, R. Yelagandula, B. Zhu & S. Henikoff, (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenetics & chromatin* **5**: 7.

Tani, H. & N. Akimitsu, (2012) Genome-wide technology for determining RNA stability in mammalian cells: historical perspective and recent advantages based on modified nucleotide labeling. *RNA Biol* **9**: 1233-1238.

Tham, W. H. & V. A. Zakian, (2002) Transcriptional silencing at Saccharomyces telomeres: implications for other organisms. *Oncogene* **21**: 512-521.

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

Thoma, F., T. Koller & A. Klug, (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of cell biology* **83**: 403-427.

Thomas, J. O. & K. Stott, (2012) H1 and HMGB1: modulators of chromatin structure. *Biochemical Society transactions* **40**: 341-346.

Thomas, S., A. Green, N. R. Sturm, D. A. Campbell & P. J. Myler, (2009) Histone acetylations mark origins of polycistronic transcription in Leishmania major. *BMC genomics* **10**: 152.

Tiengwe, C., L. Marcello, H. Farr, N. Dickens, S. Kelly, M. Swiderski, D. Vaughan, K. Gull, J. D. Barry, S. D. Bell & R. McCulloch, (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of Trypanosoma brucei. *Cell Rep* **2**: 185-197.

Travers, A. A., (2003) Priming the nucleosome: a role for HMGB proteins? *EMBO reports* **4**: 131-136.

Trojer, P. & D. Reinberg, (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Molecular cell* **28**: 1-13.

Turner, C. M., (1997) The rate of antigenic variation in fly-transmitted and syringe-passaged infections of Trypanosoma brucei. *FEMS microbiology letters* **153**: 227-231.

Turner, C. M. & J. D. Barry, (1989) High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. *Parasitology* **99 Pt 1**: 67-75.

Ullu, E., K. R. Matthews & C. Tschudi, (1993) Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Molecular and cellular biology* **13**: 720-725.

Ullu, E. & C. Tschudi, (1991) Trans splicing in trypanosomes requires methylation of the 5' end of the spliced leader RNA. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 10074-10078.

Urwyler, S., E. Studer, C. K. Renggli & I. Roditi, (2007) A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei. *Molecular microbiology* **63**: 218-228.

Urwyler, S., E. Vassella, J. Van Den Abbeele, C. K. Renggli, P. Blundell, J. D. Barry & I. Roditi, (2005) Expression of procyclin mRNAs during cyclical transmission of Trypanosoma brucei. *PLoS pathogens* **1**: e22.

Ushinsky, S. C., H. Bussey, A. A. Ahmed, Y. Wang, J. Friesen, B. A. Williams & R. K. Storms, (1997) Histone H1 in Saccharomyces cerevisiae. *Yeast* **13**: 151-161.

van Leeuwen, F., M. C. Taylor, A. Mondragon, H. Moreau, W. Gibson, R. Kieft & P. Borst, (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 2366-2371.

van Luenen, H. G., C. Farris, S. Jan, P. A. Genest, P. Tripathi, A. Velds, R. M. Kerkhoven, M. Nieuwland, A. Haydock, G. Ramasamy, S. Vainio, T. Heidebrecht, A. Perrakis, L. Pagie, B. van Steensel, P. J. Myler & P. Borst, (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell* **150**: 909-921.

Vanhamme, L., M. Berberof, D. Le Ray & E. Pays, (1995a) Stimuli of differentiation regulate RNA elongation in the transcription units for the major stage-specific antigens of Trypanosoma brucei. *Nucleic acids research* **23**: 1862-1869.

Vanhamme, L., A. Pays, P. Tebabi, S. Alexandre & E. Pays, (1995b) Specific binding of proteins to the noncoding strand of a crucial element of the variant surface glycoprotein, procyclin, and ribosomal promoters of trypanosoma brucei. *Molecular and cellular biology* **15**: 5598-5606.

Vanhamme, L., P. Poelvoorde, A. Pays, P. Tebabi, H. Van Xong & E. Pays, (2000) Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. *Molecular microbiology* **36**: 328-340.

Vasquez, J. J., C. C. Hon, J. T. Vanselow, A. Schlosser & T. N. Siegel, (2014) Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. *Nucleic acids research* **42**: 3623-3637.

Vassella, E., A. Acosta-Serrano, E. Studer, S. H. Lee, P. T. Englund & I. Roditi, (2001) Multiple procyclin isoforms are expressed differentially during the development of insect forms of Trypanosoma brucei. *Journal of molecular biology* **312**: 597-607.

Vassella, E., J. V. Den Abbeele, P. Butikofer, C. K. Renggli, A. Furger, R. Brun & I. Roditi, (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.

Vassella, E., M. Probst, A. Schneider, E. Studer, C. K. Renggli & I. Roditi, (2004) Expression of a major surface protein of Trypanosoma brucei insect forms is controlled by the activity of mitochondrial enzymes. *Molecular biology of the cell* **15**: 3986-3993.

Vassella, E., B. Reuner, B. Yutzy & M. Boshart, (1997) Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *Journal of cell science* **110 (Pt 21)**: 2661-2671.

Veitch, N. J., P. C. Johnson, U. Trivedi, S. Terry, D. Wildridge & A. MacLeod, (2010) Digital gene expression analysis of two life cycle stages of the human-infective parasite, Trypanosoma brucei gambiense reveals differentially expressed clusters of co-regulated genes. *BMC genomics* **11**: 124.

Venkatesh, S. & J. L. Workman, (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nature reviews. Molecular cell biology* **16**: 178-189.

Vicent, G. P., A. S. Nacht, J. Font-Mateu, G. Castellano, L. Gaveglia, C. Ballare & M. Beato, (2011) Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. *Genes & development* **25**: 845-862.

Vickerman, K., (1985) Developmental cycles and biology of pathogenic trypanosomes. *British medical bulletin* **41**: 105-114.

Vickerman, K. & T. M. Preston, (1970) Spindle microtubules in the dividing nuclei of trypanosomes. *Journal of cell science* **6**: 365-383.

Vujatovic, O., K. Zaragoza, A. Vaquero, O. Reina, J. Bernues & F. Azorin, (2012) Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity. *Nucleic acids research* **40**: 5402-5414.

Walrad, P. B., P. Capewell, K. Fenn & K. R. Matthews, (2012) The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in Trypanosoma brucei. *Nucleic acids research* **40**: 2869-2883.

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

Wenrich, B. R. & T. A. Trumbo, (2012) Interaction of nucleic acids with Coomassie Blue G-250 in the Bradford assay. *Analytical biochemistry* **428**: 93-95.

WHO, (2013) Control and surveillance of human African trypanosomiasis: report of a WHO expert committee. In: WHO Tecnical Report Series. Geneva, Switzerland: World Health Organization, pp.

WHO, (2015) Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected tropical diseases

In. Geneva, Switzerland, pp.

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

Williams, R. O., J. R. Young & P. A. Majiwa, (1982) Genomic environment of T. brucei VSG genes: presence of a minichromosome. *Nature* **299**: 417-421.

Windhager, L., T. Bonfert, K. Burger, Z. Ruzsics, S. Krebs, S. Kaufmann, G. Malterer, A. L'Hernault, M. Schilhabel, S. Schreiber, P. Rosenstiel, R. Zimmer, D. Eick, C. C. Friedel & L. Dolken, (2012) Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome research* **22**: 2031-2042.

Wirtz, E., S. Leal, C. Ochatt & G. A. Cross, (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Molecular and biochemical parasitology* **99**: 89-101.

Wolffe, A. P., S. Khochbin & S. Dimitrov, (1997) What do linker histones do in chromatin? *Bioessays* **19**: 249-255.

Woodcock, C. L., A. I. Skoultchi & Y. Fan, (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **14**: 17-25.

Woodward, R. & K. Gull, (1990) Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of Trypanosoma brucei. *Journal of cell science* **95 (Pt 1)**: 49-57.

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

Wurst, M., B. Seliger, B. A. Jha, C. Klein, R. Queiroz & C. Clayton, (2012) Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in Trypanosoma brucei. *Molecular microbiology* **83**: 1048-1063.

Yang, X., L. M. Figueiredo, A. Espinal, E. Okubo & B. Li, (2009) RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei. *Cell* **137**: 99-109.

Yun, M., J. Wu, J. L. Workman & B. Li, (2011) Readers of histone modifications. *Cell research* **21**: 564-578.

Zhang, J. R., J. M. Hardham, A. G. Barbour & S. J. Norris, (1997) Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**: 275-285.

Zhang, Y., M. Cooke, S. Panjwani, K. Cao, B. Krauth, P. Y. Ho, M. Medrzycki, D. T. Berhe, C. Pan, T. C. McDevitt & Y. Fan, (2012) Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS genetics* **8**: e1002691.

Zheng, Y., S. John, J. J. Pesavento, J. R. Schultz-Norton, R. L. Schiltz, S. Baek, A. M. Nardulli, G. L. Hager, N. L. Kelleher & C. A. Mizzen, (2010) Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. *The Journal of cell biology* **189**: 407-415.

Zomerdijk, J. C., R. Kieft & P. Borst, (1991) Efficient production of functional mRNA mediated by RNA polymerase I in Trypanosoma brucei. *Nature* **353**: 772-775.

Zomerdijk, J. C., M. Ouellette, A. L. ten Asbroek, R. Kieft, A. M. Bommer, C. E. Clayton & P. Borst, (1990) The promoter for a variant surface glycoprotein gene expression site in Trypanosoma brucei. *The EMBO journal* **9**: 2791-2801.

7 ANNEXES

7.1 PUBLICATIONS