Universidad de Granada

Instituto de Parasitología y
Biomedicina "López-Neyra"

The transcriptional activator SNF2PH regulates VSG and promotes infective form surface protein expression in *Trypanosoma brucei*

Andreu Saura Rebollar

Doctoral dissertation

**PhD program in
Biochemestry and Molecular
Biology**

Universidad de Granada

**Instituto de Parasitología y
Biomedicina "López-Neyra" CSIC**

The transcriptional activator SNF2PH regulates VSG and promotes infective form surface protein expression in *Trypanosoma brucei*

Andreu Saura Rebollar

Doctoral dissertation

**PhD program in
Biochemestry and Molecular
Biology**

El activador SNF2PH regula la transcripción de la VSG y promueve la expresión de proteínas de superficie de la forma infectiva de *Trypanosoma brucei*

The transcriptional activator SNF2PH regulates VSG and promotes infective form surface protein expression in *Trypanosoma brucei*

El Doctorando/ The Doctoral candidate, Andreu **Saura Rebollar**

y el director de tesis /Thesis supervisor, **Miguel A. Navarro Carretero**

Garantizamos, al firmar esta tesis doctoral, que trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de los autores ha ser citados cuando se han utilizado sus resultados o publicaciones

/

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor and as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when the results or publications have been used) have been respected

Granada a 30 de septiembre de 2019 / September 30th, 2019

Director de la Tesis/Thesis Supervisor   Doctorando/ Doctoral Candidate

Miguel A. Navarro Carretero   Andreu Saura Rebollar

Profesor de Investigación del CSIC

Front and reverse cover: Artwork by Andreu Saura. *No vull tornar-me a aixecar* (2017).
Acrylic on wood panel. Flyes designed by an unkown artist.

Agradecimientos/ Agraïments/ Greetings

Els petits canvis són poderosos

Capità Enciam (1994)

A mons pares

Index

Abbreviation list

Abbreviation list

5'-AMP	Adenosine 5'-monophosphate
aES	Active Expression Site
AMPKα1	AMP-Activated Protein Kinase
ASF1	Anti-Silencing Function *protein* 1
BC	Basic Copies
BES	Bloodstream form Expression Site
BF	Bloodstream Form
CAF1	Chromatin Assembly Factor 1
CHD1L	Chromodomain Helicase DNA Binding Protein 1 Like
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP-sequencing
CITFA	Class I Transcription Factor A
Co-IP	Co-Immunoprecipitation
DAC	Histone Deacetylase
DAPI	4', 6-Ddiamidino-2-Phenylindole
DEG	Differentially Expressed Genes
DNAse	Deoxyribonuclease
Dpi	Days Post Infection
DRALI	Double Renilla Active site Luciferase Inactive site
EP1-2	EP procyclin
ES	Expression Site
ESAGs	Expression Site-Associated Genes
ESB	Expression Site Body
ESPM	Expression Site Promoter Mapping
FACS	Fluorescence-Activated Cell Sorting
FACT	FAcilitates Chromosome Transcription
FDR	False Discovery Rate
FLuc	Firefly Luciferase

GFP	Green Fluorescent Protein
GPEET2	Gly-Pro-Glu-Glu-Thr containing procyclin
H3K4me3	Histone 3 trimethylated at Lys4
HA	Haemagglutinin
HMT	Histone Methyl Transferases
HSF	Highly SUMOylation Focus
iES	Inactive Expression Site
IF	Immunofluorescence
ISGs	Invariant Surface Glycoproteins
IP	Immunoprecipitation
ISWI	Imitation SWItch
kDa	Kilodalton
LC-MS/MS	Liquid Chromatography coupled to tandem Mass Spectrometry
logCPM	Log for Count Per Million
LogFC	Log Fold Change
mRNA	Messenger RNA
NSD3	Nuclear receptor Binding SET Domain Protein 3
NUP1	Nucleoporin 1
ORF	Open Reading Frame
PAD	Protein Associated with Differentiation
PAG	Procyclin Associated Genes
PCNA	Proliferative Cell Nuclear Antigen
PF	Procyclic Form
PH	Plant Homeodomain
PIAS	Protein Inhibitor of Activated STAT-1
PLA	Proximity Ligation Assay
pseVSG	pseudo VSG
PTM	Post Translational Modification
qPCR	Quantitative PCR
rDNA	Ribosomal DNA

RLuc	Renilla Luciferase
RNAi	Interference RNA
RNAse	Ribonuclease
RNA-seq	RNA-sequencing
RPA135	RPAI second largest subunit
RPAI	RNA Polymerase I Largest subunit
RPB1	RNA Polymerase II Largest subunit
RPB5z	RNA Polymerase II 5z subunit
RPMK	Reads per Kilobase Million
rRNA	Ribosomal RNA
RT-qPCR	Quantitative reverse transcription PCR
SENPs	SEntrin Specific Proteases
SIR	Silent Information Regulator
SIZ1	SAP and mIZ-finger domain
SL	Spliced leader
SMCA	SWI/SNF-related Matrix-associated Actin dependent Regulator Chromatin group A
SNF2	Sucrose Nonfermenting Protein 2
SSR	Switch Strand Region
SUMO	Small Ubiquitin-like MOdifier
TF	Transcription factor
THT-2	Hexose Transporter isoform
TIS	Transcription Initiation Site
VSG	Variant Surface Glycoprotein
VSG-ES	VSG Expression Site
WT	Wild Type

1. Summary

1. Summary

Trypanosoma brucei is an extracellular parasite belonging to *Trypanosomatidae* family (Protist kingdom). In the mammalian host, the bloodstream form of *T. brucei* causes sleeping sickness or Human African Trypanosomiasis (HAT) and *nagana* in cattle. The parasite undergoes morphological changes between the insect and the mammalian host over the course of its life cycle. The mammalian bloodstream form expresses the Variant Surface Glycoproteins (VSG) required for antigenic variation. Changes in the expression of different types of VSGs allow the parasite to elude the mammalian host adaptive immune response, leading to a persistent infection. In the insect host, the VSG is replaced by a different glycoprotein family named procyclins when the parasite differentiates to procyclic form in the tsetse fly. During the bloodstream form of the parasite, a single VSG is expressed from a telomeric locus known as the VSG-Expression Site, (VSG-ES). The transcription of this locus is mediated by RNA Polymerase I (RNA Pol I) in a nuclear body called the Expression Site Body (ESB), which is located outside the nucleolus where this polymerase normally resides. In the genome of *T. brucei* there are about 1000 genes for different types of VSG genes although they can only be expressed from a VSG-ES of approximately 15 different telomeric loci, of which only one is transcribed at any given time. The expression of a VSG on the surface of the parasite can be changed to another either by recombination of another VSG gene in the active VSG-ES, or by a *in situ* transcriptional switch to another of the 15 VSG-ESs by epigenetic mechanisms not very known. During differentiation to the procyclic form, no VSG is expressed and the RNA pol I in the nucleolus transcribes the genes of the procyclins, while the telomeric VSG-ES is repressed.

Previous published work from our group described SUMOylation (Small Ubiquitin-like MOdifier, SUMO) as a Post Translational Modification (PTM) that regulates VSG expression (Lopez-Farfan et al., 2014). SUMOylation is a large and reversible post-translational modification (PTM) that regulates many critical processes in eukaryotes, including transcription, DNA replication, repair and protein signaling. Transcription factors are well known SUMO targets, whose activity can be modulated in both gene silencing and activation. In bloodstream forms trypanosomes, SUMO-conjugated proteins are enriched at the highly-SUMOylated focus (HSF), which co-localized to the nuclear body ESB and are also abundant in the active VSG-ES chromatin (Lopez-Farfan et al., 2014). Thus, it is possible that SUMO acts as an epigenetic chromatin mark to positive regulate VSG expression.

In this PhD thesis, we first identify major SUMO-conjugated proteins in the mammalian infective form by proteomic analyses as a starting point to uncover new factors related to VSG

expression. This study yielded to the identification of several SUMO targets, some related to VSG expression and probably linked to developmental regulation, in addition to proteins involved in well-known process regulated by SUMO as DNA repair, replication, etc.

Out of the SUMO-modified proteins, we identified a transcriptional activator (Tb927.3.2140) which contains a Sucrose Non Fermentative-2 (Snf2) chromatin remodeler domain and a Plant Homeodomain (PH); we thus named SNF2PH. The SNF2 domain functions by modulating chromatin accessibility, while the PH domain is involved in recognizing epigenetic chromatin marks involved in gene expression regulation of development. Our data suggested that the PH domain of SUMOylated SNF2PH changes the protein interaction interface that reads histone tail modifications and thus regulates VSG-ES transcription by increasing chromatin accessibility. Importantly, ectopic expression of SNF2PH in the insect form, where it is normally downregulated, triggers the transcription of bloodstream stage-specific genes of surface proteins, while the expression of a truncated version of the SNF2PH (SNF2ΔPH) that lacked the PH domain did not induce the transcription of these genes. Altogether, these data suggest that SNF2PH SUMOylation positively regulates VSG monoallelic transcription while the PH domain is required for the expression of bloodstream-specific surface protein genes. Thus, SNF2PH functions as a positive transcriptional activator, and is required to ensure coordinated expression of surface stage-specific proteins. Altogether our results suggest that SNF2PH is a central regulator of VSG monoallelic expression and maintenance infective stage-specific surface protein integrity, acting as a major regulator of pathogenicity.

Analysis of gene expression in bloodstream form cell lines where SNF2PH was depleted using RNA interference detect silencing of the active VSG-ES and the activation of the expression of the developmentally regulated genes markers, such as PADs (Protein associated with the differentiation) and procyclins, suggesting a role in development. In addition, of SNF2PH protein levels in stumpy-like cells obtained after AMPKα1 activation by AMP together with detection of SNF2PH downregulation in stumpy form where AMPKα1 activation occurs, suggests AMPK pathway negatively regulates SNF2PH expression by an undefined mechanism.

We propose a model whereby SNF2PH has distinct activities in a PTM-dependent manner, since SNF2PH chromatin remodeling protein requires SUMO modification to maintain active monoallelic VSG expression while the PH domain functions to ensure coordinated transcription of surface stage-specific protein genes and thus adaptation to the mammalian host. This finding shows a novel link between the epigenetic inheritance of a single active allele, among a multiallelic VSG family by the PTM of SNF2PH, and subnuclear body dynamics by SUMOylation.

Summary

All of which provide a new framework to unravel the molecular mechanisms underlying inheritance of monoallelic expression of the VSG.

Resumen

Trypanosoma brucei es un protozoo parásito extracelular perteneciente a la familia *Trypanosomatidae* (Reino Protista). En su hospedador mamífero, la forma sanguínea de *T. brucei* causa la enfermedad del sueño o tripanosomiasis africana humana y el *nagana* en ganado. El parásito sufre cambios morfológicos entre el insecto vector (*Glossina sp.*) y el huésped mamífero al largo del su ciclo de vida. En la superficie de la forma sanguínea del parásito expresa la glicoproteína variable de superficie (VSG, *Variant Surface Glycoprotein*). Cambios en la expresión de diferentes tipos de VSG por un proceso conocido como variación antigénica permite al parásito evadir la respuesta inmunitaria adaptativa del hospedador mamífero lo que le proporciona una infección persistente. En cambio, la VSG es remplazada por otra familia de glicoproteínas conocida como prociclinas cuando el parásito se diferencia a la forma procíclica del insecto vector, la mosca tsetsé. En la forma sanguínea del parásito se expresa una única VSG en la superficie, cuyo gen se transcribe desde un locus telomérico conocido como el sitio de expresión de la VSG (VSG-*Expression site*, VSG-ES). La transcripción de este *locus* esta mediada por la RNA Polimerasa I (RNA Pol I) en un cuerpo nuclear denominado cuerpo del sitio de expresión (*Expression Site Body*, ESB), el cual está localizado fuera del nucléolo donde normalmente reside esta polimerasa. En el genoma de *T. brucei* existen alrededor de 1000 genes para distintos tipos de VSG aunque únicamente se pueden expresar desde un sitio de expresión telomérico de entre los aproximadamente 15 *loci* teloméricos, de los cuales únicamente se transcribe uno en un momento concreto. La expresión de una VSG en la superficie del parasito puede cambiar a otra mediante recombinación de otro gen VSG en el VSG-ES activo, o bien por un mecanismo de cambio transcripcional *in situ* a otro de los 15 VSG-ES mediante mecanismos epigenéticos no bien conocidos. Durante la diferenciación a la forma procíclica, no se expresa ninguna VSG ya que el VSG-ES telomérico es reprimido y la RNA pol I, en el nucléolo transcribe los genes de las prociclinas.

En un trabajo publicado anteriormente por nuestro grupo se describió a la SUMOilación (*Small Ubiquitin-like MOdifier*, SUMO) como una modificación postraduccional (PTM) que regula la expresión de la VSG (Lopez-Farfan et al., 2014). La SUMOilación de proteínas es una modificación postraduccional (PTM) reversible de aproximadamente 11 kDa, que regula muchos procesos críticos en eucariotas como son la transcripción, la replicación del DNA, la reparación y la señalización de proteínas. Los factores de transcripción son dianas de SUMO bien conocidas cuya actividad puede ser modulada tanto para su inhibición como para su

activación. En la forma sanguínea de tripanosoma, las proteínas conjugadas con SUMO se encuentran enriquecidas en el núcleo en un foco altamente SUMOilado (*Highly SUMOylated Focus*, HSF), que colocaliza parcialmente con el cuerpo nuclear ESB y también abundan en la cromatina del VSG-ES activo (Lopez-Farfan et al., 2014). Por lo tanto, es posible que SUMO actúe como una marca epigenética de la cromatina para regular positivamente la expresión de la VSG.

En esta tesis doctoral, primero identificamos las principales proteínas conjugadas con SUMO en la forma sanguínea mediante análisis proteómicos como punto de partida para descubrir nuevos factores relacionados con la expresión de la VSG. Este estudio dio lugar a la identificación de varios sustratos de SUMO, algunos relacionados con la expresión de la VSG y probablemente vinculados a la regulación del desarrollo, además de proteínas involucradas en procesos conocidos regulados por SUMO como la reparación de DNA, replicación, etc.

De entre las proteínas modificadas con SUMO, se identificó un factor activador de la transcripción (Tb927.3.2140) que contiene un dominio de remodelación de cromatina perteneciente a la familia Snf2 (*Sucrose Non Fermentative*), junto con un dominio PH (*Plant Homeodomain*), el cual nombramos SNF2PH. El dominio SNF2 funciona modulando la accesibilidad de la cromatina, mientras que el dominio PH está involucrado en el reconocimiento de marcas epigenéticas involucradas en la regulación de la expresión génica en el desarrollo. Los datos obtenidos sugieren que el dominio PH de SNF2PH modificado con SUMO, cambia la interfaz de interacción con otras proteínas y probablemente mediante la lectura de modificaciones en las histonas regula la transcripción del VSG-ES activo al aumentar la accesibilidad de la cromatina. Es importante destacar que la expresión ectópica de SNF2PH en la forma procíclica del insecto vector, donde normalmente no se expresa significativamente, desencadena la transcripción de genes de las proteínas de superficie específicas de la forma sanguínea, mientras que la expresión de una versión truncada de SNF2PH (SNF2ΔPH) que carecía del dominio PH no indujo la transcripción de estos genes. En conjunto, estos datos sugieren que la SUMOilación de SNF2PH regula positivamente la transcripción monoalélica de la VSG, mientras que el dominio PH es necesario para la expresión de proteínas de superficie específicas de la forma sanguínea. Por tanto, SNF2PH funciona como un activador transcripcional positivo, integrando la expresión de proteínas de superficie de la forma sanguínea infecciosa como la regulación de la expresión monoalélica de la VSG, actuando como un regulador principal de la patogenicidad.

Summary

Análisis de la expresión génica en la forma sanguínea donde los niveles de SNF2PH se redujeron utilizando interferencia de RNA nos permitió detectar un silenciamiento del VSG-ES activo y la activación de la expresión de marcadores de desarrollo, tales como PADs (*Protein Associated with Differentiation*) y las prociclinas, sugiriendo una función en el desarrollo. Ademas, la reducción de los niveles de SNF2PH observada en células *stumpy-like* obtenidas por activación de AMPKα1 con AMP junto con la regulación negativa de SNF2PH en forma *stumpy* (forma preadaptada a la transición hacia la forma procíclica del insecto vector) donde ocurre la activación de AMPKα1, sugiere que la vía AMPKα1 regula negativamente la expresión de SNF2PH por un mecanismo indefinido.

Proponemos un modelo mediante el cual SNF2PH tiene actividades distintas de una manera dependiente de PTM, ya que la proteína de remodelación de la cromatina SNF2PH requiere la modificación SUMO para mantener la transcripción activa que regula la expresión monoalélica de la VSG, mientras que el dominio PH funciona para garantizar la transcripción coordinada de genes de proteínas específicas de superficie la etapa sanguíneos y, por lo tanto, la adaptación a El huésped mamífero. Este hallazgo muestra un nuevo vínculo entre la herencia epigenética de un solo alelo activo, entre una familia VSG multialélica por el PTM de SNF2PH, y la dinámica del cuerpo subnuclear por SUMOylation. Todo lo cual proporciona un nuevo marco para desentrañar los mecanismos moleculares subyacentes a la herencia de la expresión monoalélica de VSG.

2. Introduction

2. Introduction

2.1- Trypanosoma brucei

Trypanosoma brucei (*T. brucei*) is an extracellular parasite, belonging to *Trypanosomatidae* family (Protist kingdom), which early branched from the main eukaryote linages diverged from the *Eukarya* domain, whose genome has been already sequenced (Berriman et al., 2005) and is the main causative agent of the sleeping sickness. *T. brucei* belongs to the kinetoplastea class, which is characterized by a single mitochondria containing a multicopy, complex DNA known as Kinetoplast, associated with the basal body of the flagellum. Many of its organelles and structures are present in a single copy and located from the anterior to the posterior part of the cell **(Fig 01)**.

1. Nucleus
2. Kinetoplast
3. Flagellar pocket
4. Basal body and probasal body
5. Axoneme and paraflagellar rod
6. Golgi
7. Lysosome
8. Endosomes
9. Mitochondrion
10. Microtubule cytoskeleton
11. Glycosomes
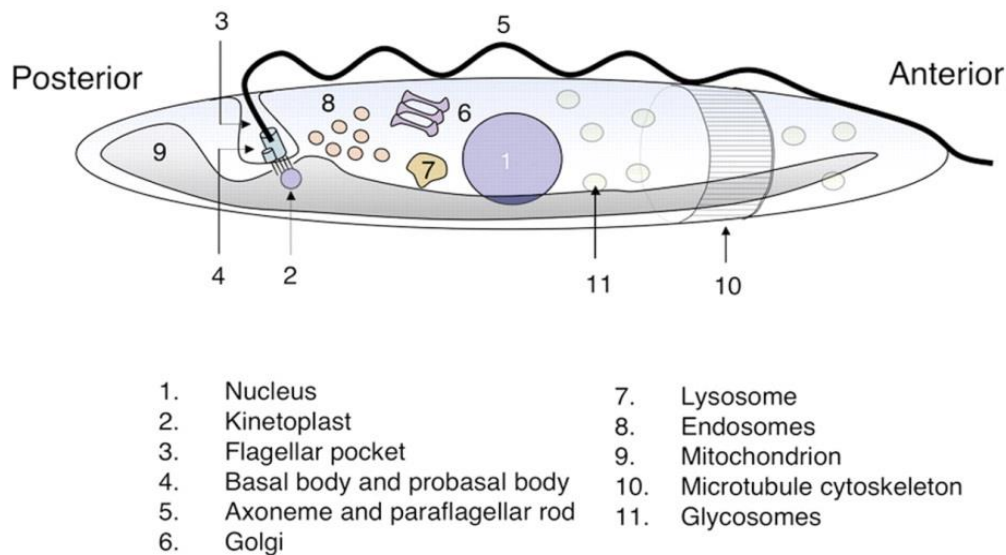
Figure 01. Celular strcutureof bloodfstream form *Trypanosoma brucei*. The developmental cell biology of *Trypanosoma brucei*. Keith R. Matthews. Journal of Cell Science (2005).

Under the membrane there is a network of microtubules that forms the cytoskeleton and the entire surface of the parasite is covered by a dense layer constituted by the Variable Surface Glycoprotein (VSG).

2.1.1 -Sleeping sickness

T. brucei is the agent responsible for sleeping sickness in humans and *nagana* in animal infections. The two subspecies involved in the disease in humans comprises *T. brucei rhodesiense* (Eastern and Southern Africa), responsible for 5% of reported cases, leading to acute infection, and *T. brucei gambiense* (West and Center Africa), responsible for 95% of reported cases, causing chronic infection. *T. brucei brucei* causes *nagana* which severely

affects the livestock in Sub-Saharan Africa. This last subspecies is the most studied and the most used for laboratory studies.

After the bite of an infected Tsetse fly, the trypanosomes spread through the hemolymphatic system causing discomfort symptoms, lack of energy and intermittent fevers (Stage 1 or haemo-lymphatic). When the parasite crosses the blood-brain barrier, it is found in the cerebrospinal fluid, rising to neurological and endocrine disorders, driving to a coma and death if untreated (Stage 2 or meningo-encephalitic). The disease occurs in 36 countries in Sub-Saharian Africa where the vector (*Glossina sp*.) is present, which constitutes the main transmission media of the disease. In 2009, the number of new cases dropped below 10000 (Simarro et al., 2011). This dropping trend continued until 2014 when only 3796 cases were reported with less than 15000 estimated cases according to World Health Organization (WHO). Recent monitoring studies reported 2164 cases in 2016, leading to only 1447 reported cases in 2017 according to passive surveillance (Franco et al., 2017). Nowadays, the indicators of elimination are on track for the 2020 goal, were expected to be eliminated as a public health problem.

Diseases caused by trypanosomatid parasites have been classified as neglected tropical diseases (NTD), since the pharmaceutics industry has not dedicated much attention and the available chemotherapeutics are deficient due to their toxicity and development of resistance, beside NTDs lack of effective vaccines. Currently, available drugs were designed more than 60 years ago with unknown mechanism of action. Pentamidine and Suramine are used for the Hemolymphatic phase, whereas the Melarsoprol was used for the Neurological phase. This last one is based on arsenic derivate that is extremely toxic, which has been replaced by Eflornithine and Nifurtimox as less toxic alternatives (Franco et al., 2014; Simarro et al., 2011). Further on, combination of Nifurtimox- Eflornithine (Nifurtimox- Eflornithine Combination Therapy or NECT) has become the first line treatment for stage 2 for the Human African Trypanosiomasis (HAT) which yielded encouraging data on efficacy and side-effect profile (Priotto et al., 2009), (Opigo and Woodrow, 2009). Disease control is based on the use of these obsolete drugs and vector control, nevertheless pharmaceutical companies started to develop new drugs for the treatment of these diseases, given the potential market in poor countries. Nowadays, Fexinidazole consists in the first all-oral treatment for both stages 1 and 2 that requires one daily dose of pills for ten days and cures 91% of people with severe sleeping sickness (Chappuis, 2018) (Pollastri, 2018).

2.1.2- Life cycle and VSG expression

Trypanosoma brucei undergoes morphological changes between the insect and the mammalian host over the course of its lifecycle which includes three main stages: bloodstream (mammalian host), procyclic (midgut of the Tsetse *Glossina* genus) and metacyclic (Tsetse salivary glands). In both bloodstream and metacyclic, parasites are coated with a variant surface glycoprotein (VSG), which is expressed from one of many telomeric expression sites (ES). This is necessary for antigenic variation, enabling a persistent infection of host adaptive immunity leading to a chronic infection. In the insect host, the VSG is replaced by a dense coat of procyclins that are expressed from chromosome-internal loci when the parasite differentiates to procyclic form in the Tsetse fly. Importantly, both VSG and procyclins genes are transcribed by RNA pol I. Metacyclic ESs consists in monocistronic transcription units containing the VSG gene, with its promoter 5kb upstream. For the other side, bloodstream form ES are polycistronic transcription units containing 10 ES-associated genes (ESAGs), of known and unknown functions, with a promoter 40-60kb upstream (Dreesen et al., 2006) **(Fig 02)**.
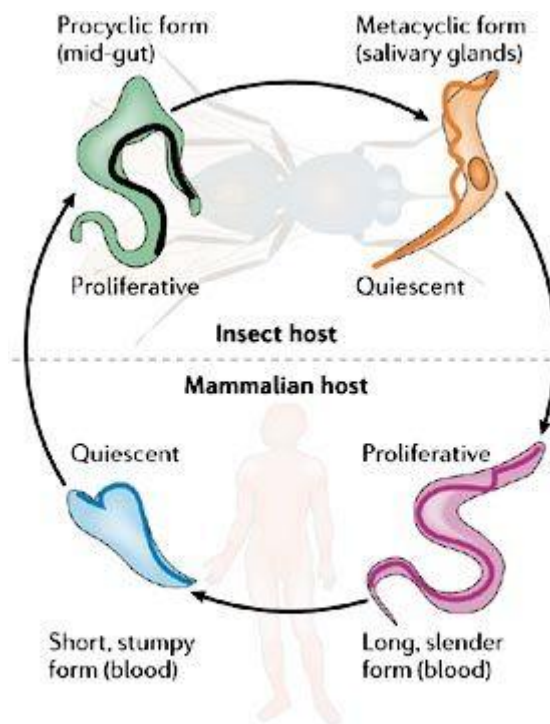
Figure 02. *Trypanosoma brucei* cell cycle. *Dreesen O. (2006) Nature.*

During the bloodstream form of the parasite, a single VSG is expressed in a nuclear body called Expression site body (ESB) which colocalizes with RNA polymerase I, required for VSG expression. There are about ~1000 VSG genes, expressed from approximately 15 telomeric loci, in which only a single VSG is expressed. One VSG can switch to another one by recombination, in situ switching or epigenetic mechanisms. When the parasite differentiates into the procyclic form, no VSG is expressed and the RNA polymerase I is recruited into the nucleus, allowing the transcription of the procyclin genes, whereas VSG-ES are repressed (Navarro et al., 2007) **(Fig 03)**.
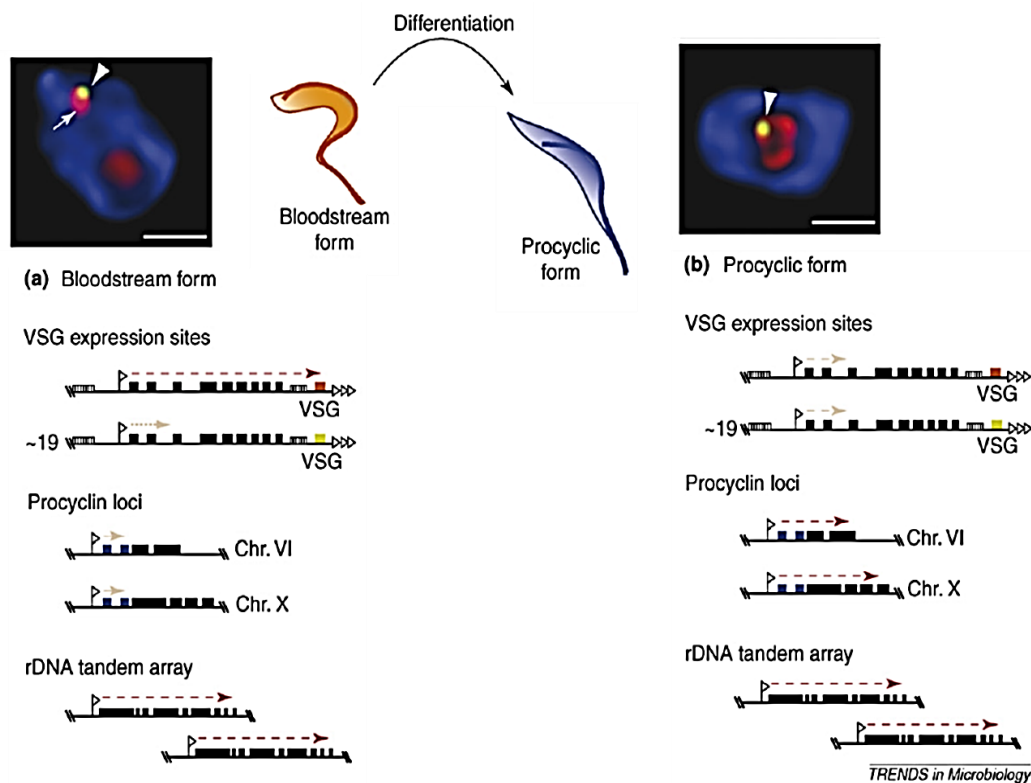
Figure 03. Genomic organization of RNA polymerase I transcribed loci and expression pattern in two developmental forms of *Trypanosoma brucei*. Adapted from *Navarro et al. (2007). Trends in Microbiology.*

T. brucei pathogenicity is due to a sophisticated strategy of antigenic variation of the VSG (Donelson, 2003; Pays et al., 2004; Taylor and Rudenko, 2006) since the parasite is continuously exposed to the host's immune response. There are about ~10^7 molecules of a single VSG type constituting the dense glycoprotein layer in the bloodstream form of the parasite, which changes with a certain frequency in the population (10^{-2}-10^{-5}), allowing the parasite to escape from host-specific antibodies against the predominant VSG, enabling a persistent infection by parasites exhibiting a de novo expressed VSG **(Fig. 04)**.

Thus, antigenic variation facilitates the establishment of chronic infections and hinders the development of effective vaccines against this parasite. The genomic organization of *T. brucei* reflects the importance of antigenic variation in the survival of the parasite. The trypanosomes are diploid organisms whose genome consists of 11 pairs of megabasic chromosomes (1 to 6 Mb), ~5 intermediate chromosomes (200 to 900 Kb) and ~1000 minichromosomes (50 to 150 Kb) (Wickstead et al., 2004). The haploid genome is ~35 Mb, with a variation up to 25% between different isolates or strains, which reflects its genomic plasticity (El-Sayed et al., 2000). The megabasic chromosomes contain all the maintenance related genes, and they are diploid except at their ends (Taylor and Rudenko, 2006).The intermediate chromosomes and the minichromosomes share a common structure consisting in a palindromic central zone and a non-repetitive subtelomeric region on both ends, followed by telomeric repeats (Wickstead et al., 2004). The subtelomeric regions of both types of chromosomes contain VSG genes and present a high homology degree with the subtelomeric zones of the megachromosomes, which allow the recombination of silenced VSG from the minichromosomes into the megachromosome expression sites. The intermediate and minichromosomes constitute more than 10% of the nuclear genome and exist as a genetic reservoir to increase the VSG repertoire.
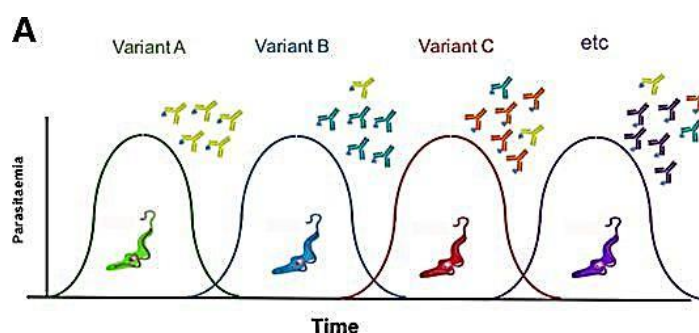
Figure 04. Trypanosome infection profile. Progressive waves of parasitemia are composed of trypanosome populations with antigenically distinct VSG coats. Adapted from *Richard McCulloch et al. (2017). Emerging topics in Life Sciences.*

There are more than a thousand VSG and pseudo VSG gene copies in the *T. brucei* genome (Berriman et al., 2005), but approximately 20 are found in regions known as VSG expression sites (VSG-ES) (Navarro and Cross, 1996; Pays and Nolan, 1998). All the copies out of the VSG expression sites are known as basic copies and are arranged in tandem in subtelomeric region from the three chromosome types. The VSG-ESs are located in subtelomeric regions of megabasic and intermediate chromosomes and consist in polycistronic units from 40 to 60 Kb, which extend from the promoter to the VSG gene and also contain pseudo VSG and Expression Site Associated Genes (ESAGs), most of them with unknown function **(Fig 05)**.
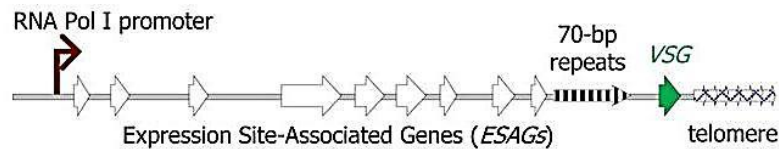
Figure 05. Structure of a VSG-ES. Adapted from *Richard McCulloch et al. (2017). Emerging topics in Life*

During the bloodstream form of the parasite, only one of the 20 VSG-ESs is active in a given time, known as active VSG-ES (Borst, 2002). The monoallelic expression is crucial for the evasion of the immune response; nevertheless, the molecular mechanisms that regulate this process are mostly unknown.

2.1.3 - Mechanisms of antigenic variation

Different molecular mechanisms have been identified in *T. brucei* antigenic variation **(Fig 06.)**. The most frequent mechanism involves homologous recombination (Robinson et al., 1999). Within this type, changes have been described by *A) gene conversion* (a donor VSG gene is duplicated to the active VSG-ES) (Robinson et al., 1999), by *B) telomeric exchange* when two VSGs are exchanged (Aitcheson et al., 2005), and by *C) telomeric conversion* (the donor is a telomere with its associated VSG gene) (Kooter et al., 1988). Finally, the *D) transcriptional switch* mechanism does not require changes in DNA, and involves *in situ* activation of a previously inactive VSG-ES to an active VGS-ES (Cross et al., 1998). The VSG gene is preceded by a series of ~70 bp repeats **(Fig 05.)** and it is believed that these repetitions should be important either for the control and the high recombination degree in this locus; however the mechanisms involved are still unknown.
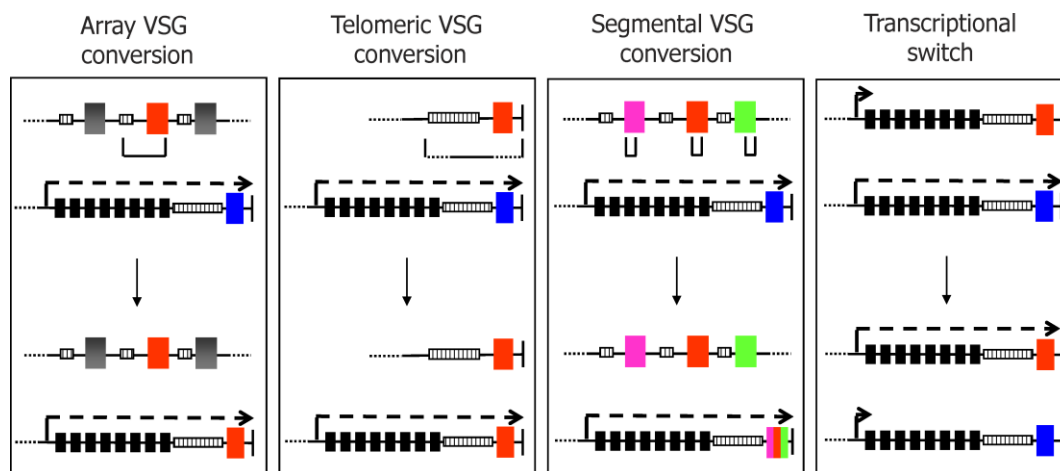
Fig 6. Mechanisms of *VSG* Switching during Antigenic Variation in T. brucei. *Stockdale et al. (2008). Plos Biology.* Blue box: VSG gene prior to switch. Red box: silent VSG gene. Pink, red or green boxes: multiple, normally non-functional VSGs.

40

2.2 - Gene expression in *T. brucei*

The nuclear genome of *T. brucei* is transcribed by three RNA polymerases (RNAPI, II and III), whose activities are analogous to the classic RNAPs present in higher eukaryotes but with slight differences. In trypanosomes, RNAPI transcribes ribosomal RNAs (rRNAs) and the coding genes for VSG in the bloodstream form, whereas in procyclic form transcribes procyclin (Gunzl et al., 2003; Navarro and Gull, 2001; Rudenko et al., 1989). RNAPIII transcribes small nuclear RNAs (snRNAs), RNAs involved in translation (tRNA and 5S rRNAs), RNA involved in other cellular processes (7SL RNA) and also the U-snRNAs (U2, U3 and U4), that are transcribed by RNAPII in other eukaryotes (Gunzl et al., 1995). Nevertheless, the activity of the RNAPII is conserved, which is responsible of the messenger RNA transcription (mRNAs).

2.2.1. Polycistronic transcription

Most of the *T. brucei* genes are organized in large gene clusters oriented in the same direction in a polycistronic form (Berriman et al., 2005), generating long transcripts that contain dozens of genes that they need to be processed before being translated. This organization is similar to the bacterial operons, especially because encoding DNA sequences are never interrupted by introns (Clayton, 2002). Unlikely bacterial operons, the *T. brucei* polycistronic units contain genes without an apparent functional relationship with a differential post-transcriptional regulation (Colasante et al., 2007). The processing of polycistronic transcripts into monocistrons requires two coupled steps: trans-splicing and polyadenilation (LeBowitz et al., 1993; Matthews et al., 1994). The transplicing consists in the addition of a 39 nucleotide sequence at the 5'-end of all mRNA, known as "Spliced Leader" (SL) RNA (De Lange and Borst, 1982). There are 200 copies of the SL gene in the genome that are monocistronically transcribed (Kooter et al., 1984).The SL RNA provides a cap structure to al mRNA, which protects them from degradation **(Fig. 07).** The polyadenylation occurs as a trans-splicing dependent manner of the subsequent gene (LeBowitz et al., 1993; Matthews et al., 1994).
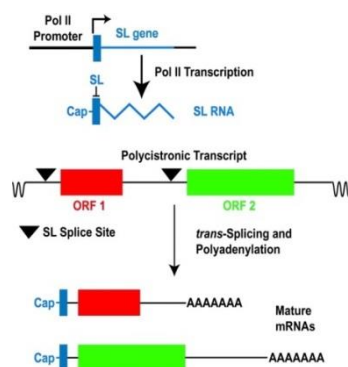
Figure 07. Transcription of the SL gene. *Scott et al. (2003). PNAS.*

2.2.2-Promoters and transcriptional regulation

Initially, it has been proposed that the regulation of the transcription start site by the RNAPII was practically non-existent in trypanosomatids (Clayton, 2002) due to its difficulty to identify promoter regions and homologous transcription factors from those that are described in eukaryotes. *In silico* analysis of *T. brucei* genome databases identified some proteins involved in the regulation of gene expression, however functional studies revealed that trypanosomatids possesses more transcription factors than were initially estimated (Martinez-Calvillo et al., 2010). As previously described, most of the genes in *T. brucei* are transcribed in a polycistronic way. Within each polycistronic transcription unit (PTU), genes could be transcribed in a divergent or convergent way from the same strands. The regions that separate the PTUs are known as Switch Strand Regions (SSR) (Siegel et al., 2009). Studies in *Leishmania* demonstrate that transcription mediated by RNAPII starts between two divergent PTUs and ends in an SSR between two convergent policystronic units (Martinez-Calvillo et al., 2003), suggesting that divergent SSR regions are possible transcription initiation sites; however the precise sequences for both transcription initiation and termination are still unknown (Martinez-Calvillo et al., 2010). Nevertheless, the only RNAPII promoter characterized so far in *T. brucei* is the SL RNA gene, which lacks of TATA box; however it has been identified several orthologous transcription factors that interact specifically with this promoter, being essential for its transcription (Schimanski et al., 2006; Schimanski et al., 2005). Some studies suggest that mediated RNAPII transcription may be initiated from a non-promoter region and the open chromatin structure is enough for the transcription initiation (McAndrew et al., 1998). Recently, it has been identified GT-rich promoter motifs that can drive transcription and promote the deposition of the histone variant H2AZ in a genomic context-dependent manner. But, *in vivo* reporter assays revealed promoter activity from different transcription start regions (TSR), suggesting the presence of specific DNA elements within them that are able to induce transcription (Wedel et al., 2017).

T. brucei presents four canonical histones: H2A, H2B, H3 and H4, and four histone variants: H2AZ, H2BV, H3V and H4V. ChIP-seq analysis of H4K10ac, H2AZ, H2BV and the BDF3 factor showed an enrichment in transcription initiation sites (divergent SSR), whereas H3V and H4V are enriched in transcription termination sites (convergent SSR)(Siegel et al., 2009) **(Fig.08)**. Several studies reported that H3V regulates RNA pol II transcription termination together with base J, both found at transcription termination sites (TTS), suggesting an important role in telomeric gene repression (Reynolds et al., 2016; Schulz et al., 2016). Additionally, it has also been reported the inhibition activity of histone H1 in RNA pol I transcription (Pena et al.,

2014), whereas it depletion leads to a de-repression of silent ES promoters (Povelones et al., 2012).
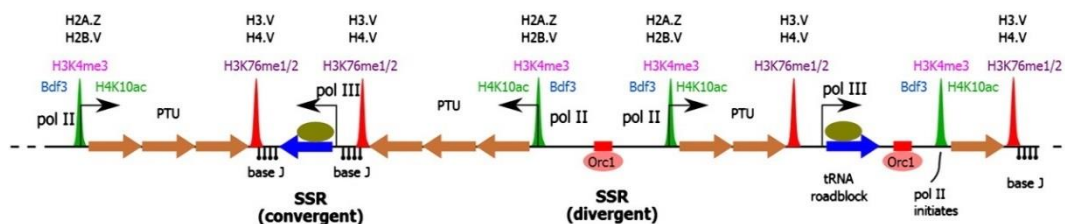
Figure 08. Pol II transcription initiates from weakly defined promoters in divergent SSR and some PTUs. Initiation locus are enriched for TbBDF3, H4K10ac, H3K4me3, H2AZ and H2BV. The epigenetic signals demarcate transcription units and regulate gene expression. Adapted from *Johannes et al. (2014). Gene Regulatory Mechanisms.*

A transcriptional activity of a *T. brucei* SSR in mammalian cells suggested that promoters in *T. brucei* may be similar to other eukaryotes, despite not being identified yet. Overall, the results lead to a divergent mechanism of transcription initiation mediated by RNAPII in trypanosomatids compared to those found in higher eukaryotes. For the RNAPIII, there are promoters that share similar structures to other eukayotes, including adjacent regulatory elements (Nakaar et al., 1994; Nakaar et al., 1997).The tRNA promoters contain two intergenic boxes (A and B) that act in *cis* like other eukaryotes.

The sequences for RNAPI promoters in *T. brucei* have also been characterized (Palenchar and Bellofatto, 2006). The rDNA promoter consists of two centred elements (domains I and II) and one distant (domain III), which resembles the typical RNAPI promoters in eukaryotes (Janz and Clayton, 1994). The procyclin promoter is similar to the rDNA promoter and is constituted by three regulatory elements. In contrast, the VSG-ES promoter requires a CA dinucleotide sequence at the start of the coding sequence (CA comprises -1/+1 positions) in addition to other regulatory sequences in positions -60 and -36 (Janz and Clayton, 1994). The differences observed in different RNAPI promoters suggest that the complex recruitment is carried out by a set of different transcription factors. There have been several attempts to characterize RNAPI complexes and other transcription factors responsible for the activity in different promoters (Gunzl et al., 2003; Nguyen et al., 2007; Nguyen et al., 2006; Walgraffe et al., 2005). Nevertheless, no specific factors have found so far for the transcription regulation of coding genes during different stages of the cell cycle, involving VSG and procyclin genes.

2.2.3 Regulation of VSG monoallelic expression

Coding genes for VSG and procyclin proteins are regulated at transcriptional level, such as the SL gene (Lustig et al., 2007). Regulation of VSG transcription is complex, as far as it is not only repressed in the procyclic form, but all the promoters at the active expression sites are repressed in bloodstream form, excepting one (Kooter and Borst, 1984). This event is known as monoallelic exclusion, which guarantees the expression of a single gene belonging to a large family of genes preserving a high homology. A proposed mechanism for the allelic exclusion arise from the identification of VEX1 (*VSG Exclusion-1*) by using a genetic screen, that restricts expression and prevents the simultaneous establishment of more than one active VSG gene (Glover et al., 2016). This described mechanism allows trypanosomes to express just one VSG gene in a given time. In *T. brucei*, only one VSG is expressed out from the 1000 VSG during the infective form and is transcribed from one of the ~20 expression sites (ES), but the mechanism of monoallelic expression control and VSG-silencing is still unknown despite the different proposed models.

In procyclic form, the mechanism of VSG-ES suppression remains unclear and differs from the bloodstream form (Navarro et al., 1999) .Though, it seems that there is a sequence specificity at promoter level, since a ribosomal promoter inserted in a VSG-ES is de-repressed while a promoter of a VSG-ES remains repressed (Horn and Cross, 1997). However, the enhanced activity from a ribosomal promoter in the procyclic form difficulties the interpretation. It has been proposed the possibility of a ~70 bp sequences close to the promoter, may play a role in monoallelic regulation (Vanhamme et al., 1995). However, deletion studies covering ~1000 bp up and downstream to the promoter where not able to find any possible regulatory sequences (Navarro and Cross, 1998), and neither no arrangements have been observed in DNA associated with changes during the transcriptional state (Navarro and Cross, 1996).

The VSG-ES promoters preserve a high homology degree and no differences are detected between active and inactive promoters (Gottesdiener et al., 1991). Nevertheless, it has been found a thymine modification present in inactive promoters, known as base J, important for the maintenance of the inactive state. Currently, the mechanisms that ensure monoallelic VSG expression remain unclear. However, several models have been proposed. The most relevant is based on nuclear compartmentation, telomeric silencing and regulation at the RNA elongation-maduration level (Navarro et al., 2007; Pays et al., 2004). Other studies suggested that both chromatin structure and epigenetic regulation play an important role. The proposed models are not mutually excluding and probably the monoallelic expression involves different levels of epigenetic regulation.

2.2.3.1 - Nuclear compartmentation

One of the accepted models is based on nuclear compartmentation. Transcription of procyclin and rDNA genes occurs in the nucleus, the main nuclear compartment harbouring the RNAPI in eukaryotic cells, responsible for ribosome biogenesis. However, the transcription of the active VSG-ES is brought out in an extranuclear body called Expression Site Body (ESB) (Landeira and Navarro, 2007; Navarro and Gull, 2001), which contains the RNAPI. This extranuclear body is uniquely associated with the active VSG-ES, while a partially active VSG is located in the periphery of the nucleus, similar to the rDNA and procyclin promoters (Navarro et al., 2007) **(Fig. 09)**. This suggests that the recruitment of the active expression site to the ESB is crucial for the monoallelic expression. It is remarkable that the transcriptional state acquired through the allelic exclusion mechanism must be inheritable, since the expressed VSG-ES at a given moment must be maintained during successive generations. It has recently shown that the active VSG-locus presents a delay during the separation of sister chromatids and it is only associated to the ESB during mitosis. This delay is dependent of the Cohesin Complex, since the depletion of any of the subunits of the complex causes a premature separation of the chromatids, leading to an increase in the switch frequency form a previously inactivated VSG (Landeira et al., 2009).
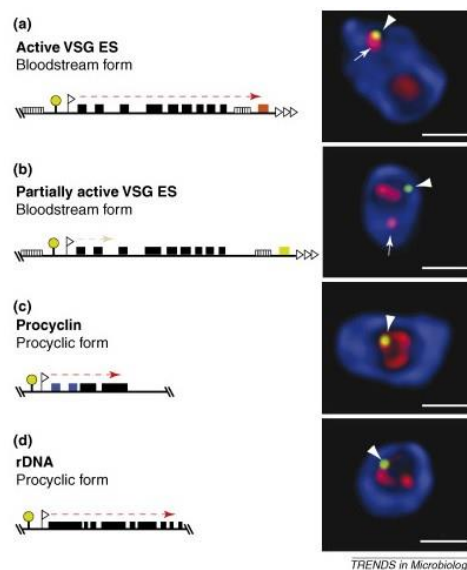
Figure 09. Subnuclear location of RNA pol I promoters and location of RNA pol I transcribed loci. Active VSG-ES (a) Partially active VSG-ES (b), procyclin (c) and rDNA locus were GFP-lacI tagged. Different GFP-lacI-tagged cell lines were labelled by double immunofluorescence with anti-pol I polyclonal Ab (red), anti-GFP monoclonal Ab (green) and DAPI (blue). The VSG-ES is indicated with an arrowhead, whereas the ESB is indicated with an arrow. Adapted from *Navarro et al. (2007). Trends in Microbiology.*

2.2.3.2 - Telomeric silencing

It has been proposed that the telomere structure may have an important role in regulation of VSG expression, since it is exclusively expressed from a subtelomeric locus (De Lange and Borst, 1982; Horn and Cross, 1995). It has been observed that the TbRAP1 binding protein, an important component of the telomeric complex, affects VSG-ES silencing. Depletion of TbRAPI causes the loss of telomere repression that contains VSG-ESs and leads to the formation of more than one ESB (Yang et al., 2009). Although the derepression is not specific, it suggests that the chromatin structure in the telomere constitutes an important component in the maintenance of a single ESB. In contrast, at the insect stage, subtelomeric silencing by TbRAP1 has been involved in changes in chromatin structure at the subtelomeric VSG loci, which contributes to its strong silencing effect on VSGs (Pandya et al., 2013). Further studies demonstrated the role of a non-coding RNA in antigenic variation which provides a link between telomeric silencing and subtelomere integrity through TbRAP1 regulated transcription (Nanavaty et al., 2017). In addition, it has also been characterized a telomere DNA-binding factor (TbTRF) that plays a role in VSG switching regulation and its depletion leads to an increase in switching events, indicating that the telomere binding activity is also important for VSG switching regulation (Jehi et al., 2014). Recent data suggested that there is a different telomere complex composition during both developmental stages of the parasites, in which the telomere associated protein (TelAP1) forms a complex with TbTRF and TbRAP1 to influence ES silencing kinetics during developmental differentiation (Reis et al., 2018).

2.2.3.3 - Regulation of elongation

The biggest difference between the active and the inactive VSG-ES is the ability of the first one to recruit elongation and processing factors to an RNA pol I transcribed site. Possibly such factors that are not required for rDNA transcription, could reside in the ESB to ensure efficient transcription of the active VSG-ES (Daniels et al., 2010; Pays et al., 2004). RT-qPCR experiments indicate that some inactive promoters remain moderately active but its transcription is aborted, without reaching the VSG, suggesting the existence of transcription control at elongation level (Vanhamme et al., 2000). This could be explained due to a differential nuclear position of the VSG-ESs depending on their degree of activation. Whereas the inactive VSG-ES is located randomly in the nucleus (Perez-Morga et al., 2001), a partially active VSG-ES is located in the periphery of the nucleolus where it is accessible to the RNA pol I machinery, but not to other elongation and processing factors that allow the complete transcription of the active VSG in the ESB (Navarro et al., 2007). Nevertheless, given the low transcription level of

these inactive promoters (~10%) and the number of VSG-ES promoters in the genome (~20), the described transcription of the inactive VSG-ES does not seem significant.

2.2.3.4 - Chromatin structure and epigenetic regulation

Histones are extremely divergent from eukaryotes, however it have been identified post-transcriptional modifications such as methylation and acetylation (Janzen et al., 2006; Mandava et al., 2007) that together with chromatin remodelling enzymes, begin to reveal new chromatin-related functions in this parasite (Rudenko, 2010). Histone modification and its variants have an evolutionary conserved role in transcriptional control throughout the eukaryotic lineage, including trypanosomes (Siegel et al., 2009). ChIP analysis described that the active VSG-ES presents fewer nucleosomes in comparison with the inactive, as well as the rDNA transcribed regions, which are depleted of nucleosomes. In contrast, the ribosomal spacer and the non-transcribed DNA repeats possess a high density of Histone H3, used as a nucleosome marker (Figueiredo and Cross, 2010; Stanne and Rudenko, 2010). Post-transcriptional histone modifications are important in both regulation of chromatin structure and gene expression. Concerning this, it was found that the deletion of the histone methyltransferase enzyme DOT1 (Disruptor of telomeric silencing 1) is responsible for the trimethylation of the histone H3 in *T. brucei*, leading to a 10-fold increase of several inactive VSGs transcripts (Scott et al., 1997). However, only one VSG is transcribed to a normal level (Figueiredo et al., 2008b), suggesting that other factors are involved to ensure the expression of a single VSG. In addition, *T. brucei* presents 3 acetyltransferases (HAT1-3)(Siegel et al., 2008), and depletion of HAT1 causes loss of repression of a telomeric reporter gene without affecting the transcription of VSG-ES promoters (Kawahara et al., 2008). Controversially, the histone deacetylase TbHDAC1 antagonizes the telomere repression in bloodstream form cells whereas TbHDAC3 is required for VSG-ES promoter silencing (Wang et al., 2010).

DNA methylation is one of the best characterized epigenetic markers in higher eukaryotes, bacteria and archaea. This modification is considered present in *T. brucei* due to the discovery of the putative gene of a DNA methyltransferase (Militello et al., 2008), but its role in the regulation of gene expression is still not clear. Also, other proteins involved in epigenetic regulation have been identified a related to the control of the monoallelic expression in *T. brucei*. Within them, there is an ISWI (Imitation switch) factor homologue, member of the SNF2 superfamily of chromatin remodelers, whose depletion leads to an increase in the transcription of the inactive VSG-ES promoters, without increasing the transcription of inactive VSGs (Hughes et al., 2007). Likewise, the FACT (Facilitates Chromosome transcription) chromatin remodelling complex has also been involved in maintaining the silencing of the

inactive VSG-ES promoters. Interestingly, the FACT subunit TbSpt16 was found enriched in the promoter region of the inactive VSG, but not in the active promoter in the bloodstream form (Denninger et al., 2010), and its depletion results in a decrease in histones at silent VSG-ESs (Denninger and Rudenko, 2014), highlighting its importance in chromatin repression at silent ES promoters. In addition, the described TbTDP1, an HMG box protein, is found to be enriched at the active ES (Narayanan and Rudenko, 2013). A property of this family member is to facilitate chromatin decondensation, thus making chromatin more accessible to regulatory factors and facilitating the recruitment of transcription activators (Štros, 2010).

Overall, these results suggest that the chromatin modification is an important component in inactive VSG-ES silencing, together with other epigenetic factors responsible for a complete activation of a single VSG-ES **(Fig 010)**. Nevertheless, it is still necessary to identify which factors are involved in marking the active VSG-ES as a part of the mechanism of monoallelic exclusion.
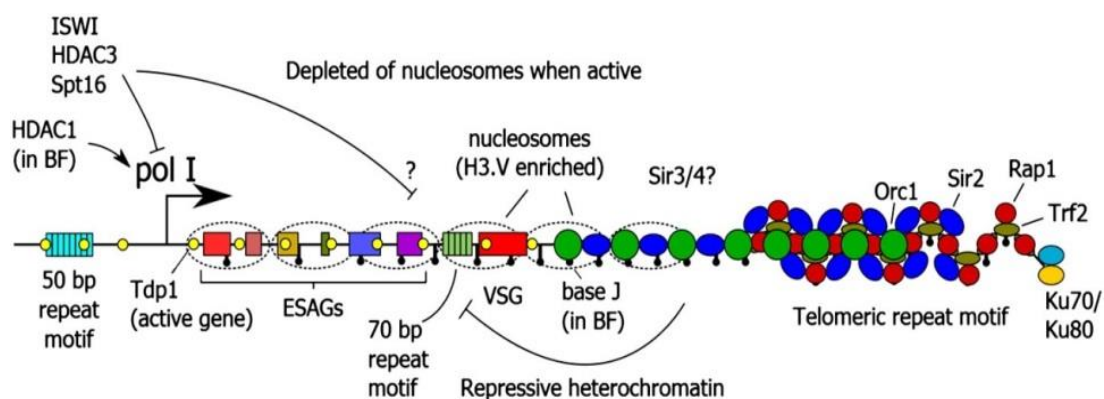
Figure 010.Epigenetic marks define the VSG-Es transcriptional state. A repressive chromatin structure is formed by TbTRF2 and TbRAP1 (which may recruit Sir2) as well as TbORC1, propagating to sub-telomeric regions. It is not known whether other proteins fulfil the roles of yeast Sir3 and Sir4, for which orthologues are absent in *T. brucei*. Base J is present at an increasing density towards the telomere termini, and is required for ES silencing. Nucleosomes present on a silent ES are enriched for the transcriptional terminating variant H3.V, and are depleted on an active ES. The HMG box protein TbTDP1 is present on the active ES, and is associated with chromatin decondensation. The histone deacetylaseTbHDAC3 and the chromatin remodeller TbISWI are required for efficient ES silencing, and TbHDAC1 is required for activated expression. Adapted from *Johannes et al. (2014). Gene Regulatory Mechanisms.*

2.3- SUMO-mediated regulation of gene expression

SUMOylation is a post-translational modification which involves the covalent binding of a similar peptide to ubiquitin called SUMO (Small Ubiquitin-like Modifier) to certain proteins to modulate its function. Since its discovery, SUMOylation has been involved in various cellular processes in higher eukaryotes since its discovery. The first SUMO identified gene was SMT3 in *S. cerevisiae*, which binds to the activating protein RanGAP1 to modulate its cellular location (Mahajan et al., 1997; Matunis et al., 1996). SUMOylation has an important role in celular processes such as transcriptional regulation, DNA damage, epigenetic control, nucleus-cytoplasm transport, protein localization and stability, stress response and cell cycle progression. At molecular level, SUMOylation alters the surface of the protein affecting its interaction with other macromolecules and interfering or promoting protein-protein interactions in different ways (Geiss-Friedlander and Melchior, 2007). So, post-translational modification by SUMO can regulate protein activity in different ways.

2.3.1 - The SUMO family

SUMO is a protein of ~12 kDa, with similar structure to ubiquitin, but only shares a 20% of identity in the aminoacid sequence. The majority of eukaryotes have a single SUMO gene, such as yeast, *C. elegans* and *D. melanogaster*, called SMT3, while plants and vertebrates present several SUMO genes (Novatchkova et al., 2004; Wang et al., 2008). In humans have been described 4 SUMO genes with different functions and location, called SUMO1-4 (Owerbach et al., 2005; Zhang et al., 2008). All SUMO proteins in eukaryotes are translated as immature precursors that must be processed by specific proteases to generate the mature form that has a C-terminal motif (glycine-glycine), which is necessary for its conjugation to other proteins **(Fig.011).** Generally, only a small fraction of a certain protein is found to be SUMOylated. This modification is reversed by the action of cysteine proteases, responsible for deSUMOylation. In yeast, these enzymes called UlpI and Ulp2 (Ulp, Ubiquitin-like protein-specific proteases) (Gong and Yeh, 2006). In humans, there are six Ulp homologs known as SENPs (Sentrin-Specific Proteases). In addition to their deSUMOylase activity, these enzymes are responsible for the maturation of the SUMO precursor through its C-terminal hydrolase activity **(Fig 011)**. In *T. brucei* only one SUMO orthologous gene has been identified (Liao et al., 2010), which is essential for cell cycle progression and viability in both procyclic and bloodstream form stages (Obado et al., 2011). A large scale proteomic analysis of the possible SUMO substrates was first characterized in *T. cruzi*, in which 236 proteins were identified involved in various biological processes (Bayona et al., 2011). All these data together suggest that SUMOylation constitutes an essential process in trypanosomes and other eukaryotes. Nevertheless, the SUMO

conjugation mechanism and the possible substrates and involving processes have been poorly investigated in *T. brucei*. To date, proteomic analysis of SUMO substrates in procyclic form led to the identification of 45 proteins involved in several important cellular processes, among them epigenetic regulation of gene expression (Iribarren et al., 2015a). Recently, it has been proposed the involvement of SUMO chain formation in chromatin organization and for the proper telomere positioning in *T. brucei*, remarking the importance of this PTM in polymeric structures through protein interaction (Iribarren et al., 2018). The identification of novel SUMO substrates in *T. brucei* still constitutes an evolving field since the development of powerful proteomic approaches that allows unravelling the function of SUMO modified targets.

2.3.2- Enzymatic mechanism of SUMOylation

SUMOylation process requires an enzymatic cascade that involves three steps. First, the mature protein SUMO is activated at the C-terminal end by the E1 activation enzyme, constituted by the heterodimer AOS1-UBA2 in yeast (Johnson et al., 1997). In this first step the E1 enzyme uses ATP to form a SUMO-adenylate bond between the carboxyl-terminal glycine residue of SUMO and an internal cysteine residue of the UBA2 subunit of the enzyme. Second, the activated SUMO protein is transferred to the E2 conjugating enzyme UBC9, to form a thioester bond between a cysteine residue of UBC9 and the C-terminal carboxyl group of SUMO (Johnson et al., 1997; Saitoh et al., 1998). Finally, the SUMO group is transferred from UBC9 to the substrate by generation of an isopeptide bond between the C-terminal glycine residue of SUMO and the amino group of a lysine of the substrate. This last step is generally mediated by SUMO E3 ligase enzymes, which determines the substrate specificity and catalyzes the transfer of SUMO from UBC9 conjugating enzyme (Johnson and Gupta, 2001; Takahashi et al., 2001; Tozluoglu et al., 2010)**(Fig. 011)**.
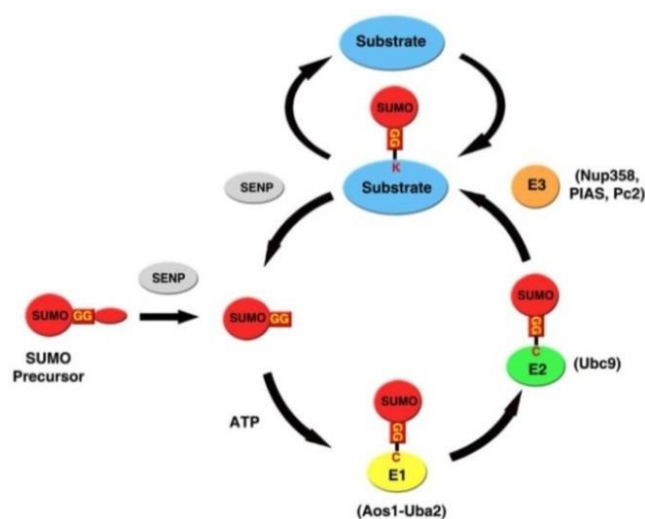
Figure 011. The enzymatic mechanism of SUMOylation . *Wan et al. Curr Protein Pept Sci.*

The transfer of SUMO from the E2 conjugation enzyme to the substrate can be SUMO E3 ligase independent, but the reaction occurs with very low efficiency, so it requires the action of E3 ligase enzymes to enhance it (Johnson and Gupta, 2001; Takahashi et al., 2001; Tozluoglu et al., 2010).

SUMO conjugation to the substrates requires the presence of a consensus sequence. Most of the proteins modified by SUMO contain the lysine residue within a Ψ-K-X (D/E) motif, where Ψ is a hydrophobic amino acid, K is the modified lysine by SUMO and X can be either glutamic (E) or aspartic acid (D) (Sampson et al., 2001). Recognition of the consensus sequence is only possible if it is localized in a structure-free region. (Ulrich, 2009), since the E2 conjugation enzyme does not recognize the consensus sites if they are in stable helix structures.

As previously discussed, the SUMO conjugation mechanism together with the specificity and the enzymes involved in the SUMOylation cascade have been poorly investigated in trypanosomatids. Nevertheless, it has been reconstructed the SUMOylation system *in vitro* with the recombinant enzymes E1 (TbAos1/TbUba2) and E2 (TbUbc9) of the SUMOylation pathway in *T. brucei* by sequence analysis and GST-pulldown (Ye et al., 2015). Additionally, it has also been developed a bacterial strain engineered to produce SUMOylated proteins using the *T. brucei* enzymatic system, which provides a useful tool for target validation and other functional studies of SUMOylated proteins in trypanosomatids (Iribarren et al., 2015b). In *T. cruzi* ,it has been identified potential SUMO orthologous components by *in silico* analysis, including 4 possible SUMO E3 ligases SP-RING type (Bayona et al., 2011), that constitutes the main group of SUMO E3 ligases by the presence of a conserved SP-motif, essential for this function. However, none of these enzymes has been characterized at molecular level. In contrast, in *T. brucei*, it has been identified one member of the homologous PIAS (Protein Inhibitor of activated STAT) in *S. cerevisiae* (TbSIZ1/PIAS) by two-hybrid screen analysis in yeast, which is involved in SUMOylation of the active VSG-ES chromatin. Functional analysis revealed that TbSIZ1 depletion reduces SUMO-conjugated proteins leading to a decreased RNA pol I recruitment to the VSG-ES (López-Farfán et al., 2014), suggesting that SUMOylation is involved in the expression of the active VSG-ES.

2.3.3 - SUMOylation and transcriptional regulation
The most common group of substrates modified by SUMO are transcription factors, whose transcriptional activity can be altered positively or negatively as a consequence of SUMOylation (Lyst and Stancheva, 2007). The SUMOylation has generally been associated with transcriptional repression. Some transcriptional correpressors have been described, such as

histone deacetylases (HDACs) complexes that are preferably associated with SUMOylated transcription factors (Garcia-Dominguez and Reyes, 2009). In *S. cerevisiae*, the main 4 histones are SUMOylated and are involved in transcriptional repression (Nathan et al., 2006). Among other examples, the human histone H4 is also SUMOylated and associated with a negative regulation of transcription (Shiio and Eisenman, 2003). However, the SUMOylation of transcription factors has also been related to an increase in transcriptional activity (Goodson et al., 2001). ChIP analysis in yeast found that SUMOylated proteins in constitutively expressed chromatin-associated genes and not in repressed genes (Rosonina et al., 2010). However, SUMOylated proteins were only detected while activated in genes that are expressed in inducible form. According to the proposed model by the authors, SUMOylation allows to regulate transcription levels and facilitate their silencing when the activating signal disappears (Rosonina et al., 2010). In addition to transcription factors, large-scale proteomics studies have identified other components of the transcription machinery as SUMO substrates in yeast, mammals and plants (Makhnevych et al., 2009; Miller et al., 2010; Rosas-Acosta et al., 2005). These include subunits of the general transcription factors (GTFs) and mediators and subunits of the RNA Polymerases, for example, the SUMOylation of the RPB1 subunit in response to UV damage (Chen et al., 2009). There are also chromatin remodelling factors as the subunits of the SWI/SNF complex, DNA topoisomerase II (Bachant et al., 2002), histones and subunits of the nuclear complex, such as cohesins and other associated proteins (Stead et al., 2003). Interestingly, further studies focused in cell-specific SUMOylated targets during mouse spermatogenesis identified proteins with a unique role in spermatogenesis and during meiotic progression, which proposes SUMO involvement in developmental regulated processes (Xiao et al., 2016). Similar to the SUMOylated substrates identified previously in other models, these targets included proteins involved in transcriptional regulation, stress response, nuclear-cytoplasmatic transport, cell-cycle control and other processes, which appear differentially expressed between both developmental stages spermatocytes and spermatids studied **(Table 01)**. Remarkably, they included several proteins implicated in regulation of chromatin remodelling during meiosis, as previously described, such as SWI/SNF-Related regulator of chromatin 4 and 5 (SMARCA4 and SMARCA5) This example provides a new insight of SUMO to functionally regulate developmental proteins.

Stress-related, heat shock proteins	Accession number [*]	MW	Number of unique peptides	
			spermatocytes	spermatids
heat shock 70 kDa protein 4	gi|112293266 (+3)	94 kDa	18	11
inducible heat shock protein 70	gi|118490060 (+4)	70 kDa		7
stress-70 protein, mitochondria –	gi|162461907 (+5)	73 kDa	5	4
heat shock 70 kDa protein 1-like	gi|124339838 (+3)	71 kDa		2
heat shock protein 105 kDa	gi|114145505 (+5)	96 kDa		3
glutathione S-transferase P 1	gi|10092608 (+5)	24 kDa	2	
1-Cys peroxiredoxin protein 2	gi|3789944 (+3)	25 kDa	2	
dnaJ homolog subfamily B member 1 (Heat shock 40 kDa protein 1)	gi|9055242 (+2)	38 kDa	2	
protein DJ-1	gi|55741460 (+1)	20 kDa	2	
Ephx1 protein	gi|34784388 (+4)	51 kDa	2	
DNA breaks, chromatin remodeling			spermatocytes	spermatids
mediator of DNA damage checkpoint protein 1 (MDC1)	gi|132626693 (+5)	185 kDa	8	
SWI/SNF related regulator of chromatin SMARCA5	gi|148678936 (+1)	116 kDa	5	
SWI/SNF related regulator of chromatin SMARCA4	gi|148693261 (+8)	185 kDa	2	
matrin-3	gi|25141233 (+4)	95 kDa	3	3
ruvB-like	gi|9790083 (+1)	50 kDa	3	
ruvB-like 2	gi|6755382 (+1)	51 kDa	2	
AT rich interactive domain 2 (ARID, RFX-like)	gi|262231796	196 kDa	3	
damage specific DNA binding protein 1	gi|148709424 (+5)	108 kDa	3	
poly [ADP-ribose] polymerase 1	gi|20806109 (+4)	113 kDa	2	
SUMO/ubiquitin pathway			spermatocytes	spermatids
tripartite motif protein 28 (TRIM28, KAP1) SUMO ligase	gi|148706135 (+3)	89 kDa	7	2

Table 01. Identification of SUMOylated proteins unique to spermatocytes or spermatid via tandem MS in mice. Adapted from *Xiao et al. (2016) Reproduction Research.*

These studies are interesting since several of these factors have been associated to the regulation of gene expression in *T. brucei* and more specifically to the regulation of the monoallelic VSG expression, as cited previously. Therefore, SUMOylation of some or several of these proteins could play an important role in transcription regulation in *T. brucei*. To date it has been possible to identify 53 site-specific SUMOylation site from 45 targets in the *T. brucei* procyclic form **(Table 2.),** most of them homologues from the identified targets in other organisms. Nevertheless, there is still a gap in bloodstream form modified targets, in which this PTM may play an important role during development leading to differential patterns of identified partners among these two developmental stages.

Gene Identification	Description
Tb927.7.2830	Histone H2A
Tb927.10.10460	Histone H2B
Tb927.1.2430	Histone H3
Tb927.10.15350	Histone H3V (h3vaR)
Tb927.5.4170	Histone H4
Tb927.4.1330	DNA topoisomerase IB, large subunit
Tb927.9.5190	Proliferative cell nuclear antigen (PCNA)
Tb927.2.4390	Endo/exonuclease Mre11 (MRE11)
Tb927.4.1270	RuvB-like DNA helicase
Tb927.10.740	Structural maintenance of chromosome 4 (SMC4)
Tb927.8.3680	Kinetoplastid kinetochore protein 4
Tb927.1.1170	DNA-directed RNA polymerase subunit RBP12
Tb927.4.1630	Ribosomal RNA processing protein 6 (RRP6)
Tb927.5.1200	Exosome component CSL4 (CSL4)
Tb927.11.16600	Exosome-associated protein 2 (EAP2)
Tb927.11.370	Repressor activator protein 1 (RAP1)
Tb927.9.11070	TbSIZ1
Tb927.3.3590	U3 small nucleolar ribonucleoprotein protein MPP10
Tb927.10.14330	Hypothetical protein/U3 Small Nucleolar RNA-Associated Protein 14
Tb927.5.2880	Chaperone protein DNAj
Tb927.11.13090	Elongation factor 1 gamma
Tb927.9.7590	60S ribosomal protein L11
Tb927.10.13720	RNA-binding protein, putative (RBP29)
Tb927.11.6350	AAA ATPase
Tb10.v4.0062	Variant surface glycoprotein (VSG, pseudogene)
Tb927.9.340	Variant surface glycoprotein (VSG, pseudogene)
Tb927.4.2740	Hypothetical protein/P25-alpha
Tb927.3.2350	Hypothetical protein
Tb927.11.11840	Hypothetical protein
Tb927.8.6130	Hypothetical protein
Tb927.10.11200	Hypothetical protein
Tb927.11.2100	Hypothetical protein
Tb927.9.1410	Hypothetical protein
Tb927.9.10680	Hypothetical protein
Tb927.3.5370	Hypothetical protein
Tb927.7.2640	Hypothetical protein
Tb927.3.4140	Hypothetical protein
Tb927.11.4880	Hypothetical protein
Tb927.11.5200	Hypothetical protein
Tb927.6.1070	Hypothetical protein
Tb927.8.2830	Hypothetical protein
Tb927.11.5230	Hypothetical protein
Tb927.9.1410	Hypothetical protein
Tb927.9.13320	Hypothetical protein
Tb927.10.14520	Hypothetical protein

Table 02. Identification of SUMOylated proteins in procyclic form *T. brucei* Adapted from *Iribarren et al. (2015)*

Cellular Microbiology.

2.4- Role of SUMOylation in VSG expression

As previously commented, kinetoplastids possess a SUMO conjugation system (Bayona et al., 2011; Garcı et al., 2003) in which *T. brucei* plays an essential role in several molecular events, such as normal cell cycle progression (Liao et al., 2010), localization of nuclear factors for appropriate chromosome segregation (Obado et al., 2011) and its response by oxidative stress (Klein et al., 2013), among others. The role of SUMO in *T. brucei* pathogenesis is increasingly evident, as far as it has been demonstrated the involvement of the SUMO ligase TbSIZ1 in regulation of VSG expression (López-Farfán et al., 2014) together with the involvement of SUMO polymeric chains in chromatin organization in procyclic forms (Iribarren et al., 2018). Importantly, in *T. brucei* bloodstream forms, SUMO-conjugated proteins are predominantly enriched in a nuclear Highly SUMOylated Focus (HSF)(López-Farfán et al., 2014),whereas in the procyclic insect form, where no VSG is expressed, SUMO-conjugated nuclear proteins are dispersed in small foci in the nucleus **(Fig 012.)**
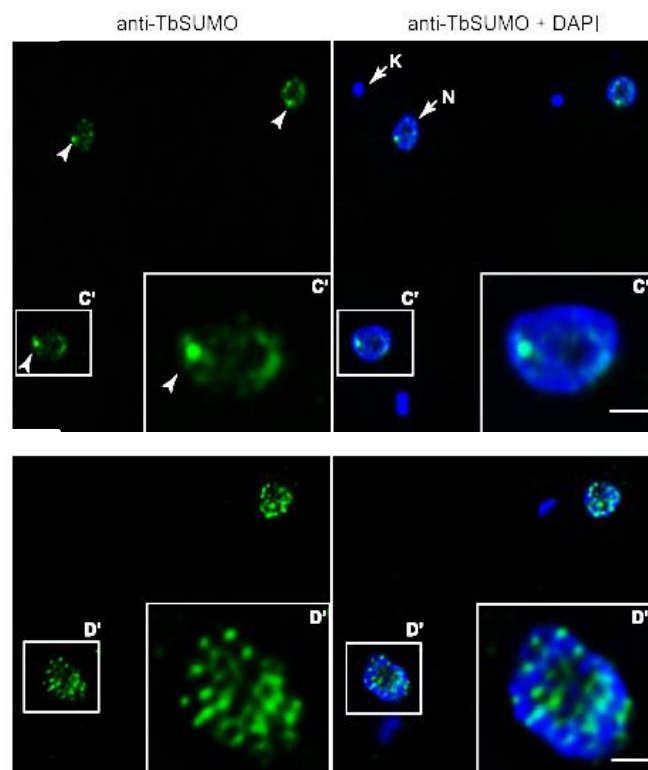
Figure 012. Expression pattern of SUMOylated proteins in *T. brucei*. In the bloodstream form nucleus, SUMO conjugated proteins are diffusely distributed in the nucleoplasm including a Highly SUMOylated Focus (HSF) (Arrowhead, upper panels). In the nucleus of procyclic form cells, SUMO conjugated proteins are diffusely distributed in many foci. 3D- IF analysis using anti-TbSUMO mAb (green) and DNA was stained with DAPI (Blue). Adapted from *López-Farfán et al. (2014). PLOS Pathogens.*

It has been demonstrated that SUMO-conjugated proteins are associated with the ESB and the active VSG-ES. Double immunofluorescence (IF) analysis using the anti-TbSUMO mAb and the anti-RPA1 antiserum showed that the HSF partially colocalized with the nuclear body ESB **(Fig 013 A)**. Additionally, the GFP-LacI tagged VSG-ES partially colocalized with the HSF by IF using the anti-TbSUMO mAb **(Fig 013 B)**. Furthermore, ChIP analysis revealed that SUMO-conjugated proteins are a distinct feature of the active VSG-ES, and probably works as an epigenetic mark to positive regulate VSG expression. However, to date non SUMO-target transcription factor have been described to regulate VSG transcription.

In this study, we aim to identify novel uncharacterized factors, specifically those involved in transcription that would help us to understand the possible roles of SUMO through the different developmental stages in *T. brucei*.
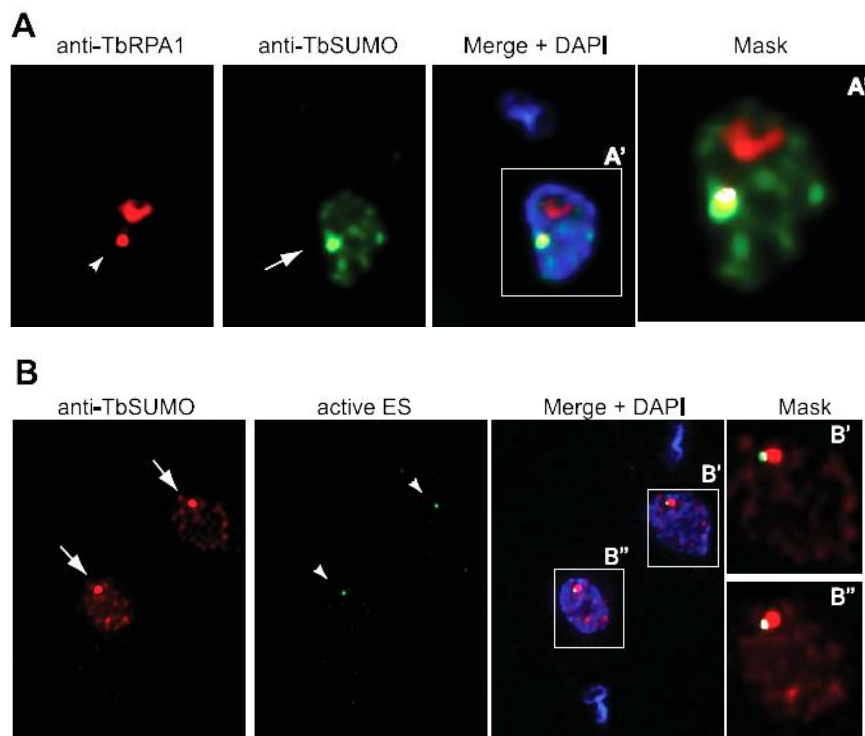
Figure 013. The Highly SUMOylated Focus (HSF) associates with the Expression Site Body (ESB) and the active VSG-ES. (A) SUMOylated proteins in the HSF partially colocalize with the RNA Pol I in the ES. 3D- IF analysis in bloodstream form cells using anti-RPAI antiserum (red), anti-TbSUMO mAb (green) and DAPI staining (blue). The ESB (arrowhead) and the HSF (arrow) are indicated. **(B)** The HSF associates with the active VSG-ES. IF analysis was performed in a cell line expressing the GFP-LacI tagged VSG-ES using the anti-TbSUMO mAb (red), rabbit anti-GFP antiserum (green) and DAPI stain (Blue). The HSF (arrow) and the GFP dot (arrowhead) are shown. Adapted from *López-Farfán et al. (2014). PLOS Pathogens.*

3. Objectives

3. Objectives

3.1- Identification of SUMO-conjugated substrates to unravel Expression Site Body (ESB) related factors and/or involved in VSG expression.

- Affinity purification of His-HA-TbSUMO conjugated substrates.
- Identification of specific SUMOylation sites in TbSUMO conjugated substrates.
- Initial characterization of the selected substrates to unravel of a possible function linked to VSG expression.

3.2- Detailed functional characterization of at least one factor whose function regulates VSG expression.

The selected factor will be functionally characterized using a wide range of available molecular tools in Trypanosoma brucei.

- Structural analysis of conserved domains.
- Expression analysis at the different developmental stages in both laboratory-induced forms and isolated infective forms in experimental animal models.
- Analysis of the subcellular location with other available markers, especially those that may suggest a possible functional relation with the VSG.
- Characterization of a possible function into VSG regulation by interference RNA (RNAi) and VSG expression analysis in reporter cell lines.
- Interference RNA to analyze the lack of function of the selected factor into the global gene expression, particularly in differentiation.

Objetivos

3.1- Identificación de proteínas modificadas por SUMO con el fin de descubrir nuevos factores asociados al ESB (Expression Site Body) y/o involucrados en la regulación de la expresión de la VSG.

- Purificación por afinidad de sustratos conjugados a His-HA-TbSUMO.
- Identificación de sitios específicos de SUMOilación en los sustratos.
- Caracterización inicial de una posible función de los sustratos seleccionados en la expresión de la VSG.

3.2- Caracterización funcional detallada de al menos un factor cuya función regule la expresión de la VSG.

El factor seleccionado será caracterizado funcionalmente utilizando las numerosas herramientas moleculares disponibles en Tripanosoma brucei

- Análisis estructural de dominios conservados.
- Análisis de expresión en las diferentes formas del desarrollo del parásito, tanto formas inducidas en el laboratorio como formas silvestres aisladas de infecciones en animal de experimentación.
- Análisis de la localización subcelular con respecto a otros marcadores disponibles, en especial aquellos que puedan sugerir una posible relación funcional con la VSG.
- Estudio de su posible SUMOilación utilizando diferentes aproximaciones.
- Caracterización una posible función en la regulación de la expresión de la VSG, utilizando Interferencia de RNA (RNAi) y análisis de expresión de la VSG (mRNA y proteína) en líneas reporteras.
- Interferencia de RNA para analizar la falta de función del factor seleccionado en la expresión génica global, en particular una posible función en la diferenciación del parásito.

4. Results

4. Results

4.1 - Identification of SUMO-conjugated substrates to unravel Expression Site Body (ESB) related factors and/or involved in VSG expression.

Transcription factors are well known SUMO targets, whose activity could be modulated in a positive or negative manner, thus SUMOylation functions in both gene silencing and gene activation mechanisms (Rosonina et al., 2017). As previously discussed in the introduction, SUMO-conjugated proteins are predominantly enriched in a nuclear Highly SUMOylated Focus (HSF) (López-Farfán et al., 2014), that partially colocalizes with the ESB in the bloodstream form (BF) of *T.brucei*. Thus, we sought to identify main SUMO-conjugated proteins in the infective form of the parasite by proteomic analyses of SUMO affinity-purified proteins. In order to do that, we applied a first approach based on the identification of His-HA-tagged SUMO substrates, followed by a site-specific identification of the SUMO acceptor site in the identified targets and its possible function related to VSG expression.

4.1.1 -Indentification of His-HA-TbSUMO substrates

Proteomic identification of SUMO substrates has become a complex issue due to methodological and experimental constraints. To accomplish this in the BF we first generated a cell line that constitutively expresses an 8xHis-HA-tagged SUMO version by double replacement of endogenous alleles, as previously described in the insect procyclic form (Iribarren et al., 2015a). This cell line displayed no difference in proliferation under regular culture conditions compared to wild type (WT) parasites, suggesting that SUMO function is not affected (data not shown) (Iribarren et al., 2015a). To isolate SUMO conjugates from transfected cells, we directly resuspended the parasite pellet in 6 M urea to avoid deconjugating peptidase activity as well as co-purification of non-covalent SUMO interacting proteins. Cell extracts were later subjected to Ni^{2+}-NTA chromatography coupled to HA-affinity enrichment and, after confirming the presence of SUMO conjugates in the eluates by Western blot (WB), samples were analyzed by MS/MS proteomics. Using this approach, from three independent experiments, we were able to identify 194 potentially SUMOylated proteins that were not present in control purifications, with a few reproducible candidates (8 proteins from experimental replicates 2 and 3) **(Table 1 and Appendix I Table S1 (A,B,C))**. The lack of reproducibility stressed the need to utilize a different strategy that leads to a specific enrichment of SUMOylated proteins allowing a confident large-scale identification of SUMO substrates.

Gene ID	MW	Description	N2	N3
Tb927.6.4430	36 kDa	Homoserine kinase	1	1
Tb927.10.1610	16 kDa	Hypothetical protein, conserved	1	1
Tb927.11.4380	72 kDa	ATP-dependent DEAD/H RNA helicase, putative	1	2
Tb927.10.13900	39 kDa	UDP-galactose transporter, putative	**1**	2
Tb927.8.3940	63 kDa	Hypothetical protein, conserved	1	1
Tb927.3.2140	**105 kDa**	**Transcription activator, putative**	**1**	**2**
Tb927.11.9890	63 kDa	Signal recognition particle receptor alpha subunit, putative	1	1
Tb927.2.2400	85 kDa	Glycosyltransferase (GlcNAc)	1	1

Table 1. Reproducible SUMOylated substrates from two independent proteomic analysis after a tandem histidine/HA affinity purification. Number of peptides for each experimental replicate are shown in columns N2 and N3 for a proteins and peptide threshold of 95% and 0.5% FDR, respectively, with 1 minimum number of peptide displayed as exclusive peptide count (Scaffold parameters).

4.1.2- Site-specific identification of SUMOylation sites in BF cell line expressing His-HA-TbSUMO[T106K]

Next, we applied a novel approach that specifically enriches SUMOylated peptides and identifies the modified Lys in each SUMO target (Tammsalu et al., 2015);(Iribarren et al., 2015a). To do so, we generated a new cell line expressing, in a tetracycline-inducible manner, the His-HA-TbSUMO gene harbouring T106K mutation, previous to the di-Gly motif (His-HA-TbSUMO[T106K]). The introduction of this mutation generates a T-shape peptide after Lys-C digestion allowing the enrichment of SUMOylated peptides using a specific anti-GG antibody and the unambiguous identification of the modified Lys residues by MS. Cells expressing His-HA-*Tb*SUMO[T106K] were able to conjugate SUMO substrates as the non-tagged cell line **(Figure 1A)** and showed the typical SUMOylation pattern, with a nuclear HSF, by indirect immunofluorescence analysis (López-Farfán et al., 2014) **(Figure 1B)**. Considering the limitations of BF culture regarding cell density, large mass of His-HA-*Tb*SUMO[T106K] expressing parasites were obtained from in Wistar infected rats and SUMOylated proteins were purified by nickel-affinity chromatography under denaturing conditions **(Figures 1C and D)**. Final eluates were subjected to Lys-C digestion and further anti-GG enrichment of SUMOylated peptides, prior to MS/MS identification.

As displayed in **Table 2**, with this approach we were able to identify 45 lysine-modified residues in 37 proteins. Most of the identified proteins showed one SUMOylation site, whereas 6 of them (DNA Topo 1B, SMC-4, Polyubiquitin, rRNA processing protein, DnaJ domain containing protein and hypothetical protein Tb927.9.13320) showed two sites, and the hypothetical protein Tb927.10.12030 displayed three sites. Globally, 23 out of 37 proteins have experimentally assigned or predicted function, whereas the remaining 14 correspond to

hypothetical with unknown function products. When comparing our data with the previously identified proteins in PF, we found 15 SUMO-conjugated substrates shared among these two developmental stages **(Table 3)** suggesting the existence of conserved processes regulated by SUMO in both BF and PF developmental stages.
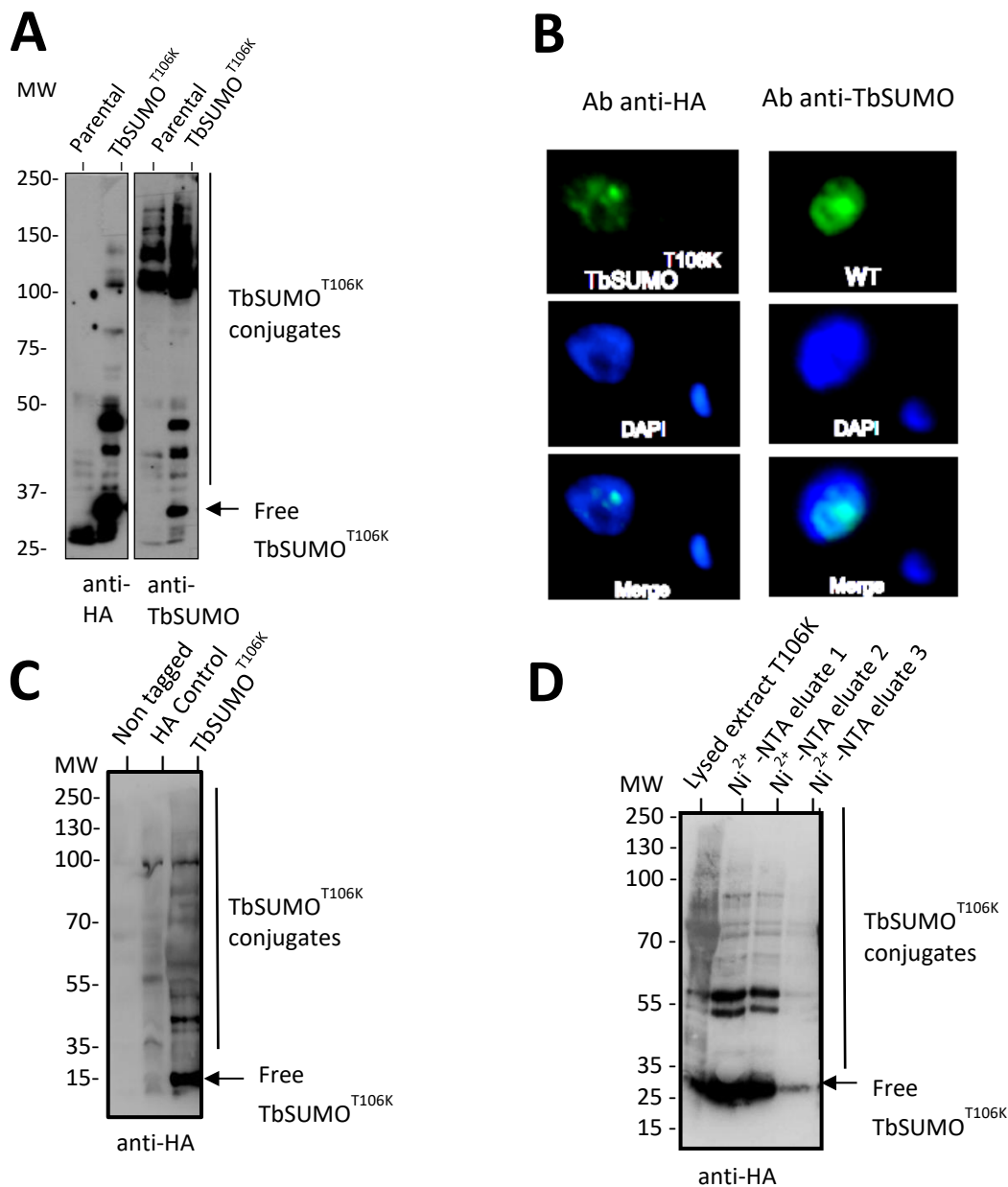
Figure 1. **Bloodstream form expression of His-HA-TbSUMO^{T106K}. A.** Expressing His-HA-*Tb*SUMO^{T106K} parasites are able to conjugate substrates. Protein cell extracts were prepared with 20 mM NEM and separated in a 10% acyilamide gel (5×10^6 cells/lane). Parental and 24 h doxycycline induced proteins (*Tb*SUMO^{T106K}) were incubated against anti-HA mAb. The same blot was also proved with anti-TbSUMO mAb (1C9H8) to detect endogenous SUMO conjugates. **B.** Immunofluorescence analysis of parasites expressing His-HA-TbSUMOT106K (24 h after induction) stained using an anti-HA antibody displayed a similar nuclear pattern including the High SUMOylated Focus (HSF) detected by the monoclonal antibody against TbSUMO in wild type cells as described previously (Lopez-Farfan, et al. 2014). DNA content was stained with DAPI (blue). Merged signal image is shown. **C.** Expression pattern of His-HA-*Tb*SUMO^{T106K} conjugates (*Tb*SUMO^{T106K}) from whole cells extracts after isolation from Wistar IGS rats (Charles River UK). Controls for parental (Non tagged) and HA epitope (HA control) were included. Samples were prepared under denaturing conditions and loaded in a 4-20% precast polyacrylamide gel (BioRad). **D.** Western blot analysis of representative eluted fractions (Eluates 1-3) of His-HA-*Tb*SUMO^{T106K} conjugates after Ni^{2+}NTA purification in expressing parasites isolated from Wistar IGS rat and lysed under denaturing conditions (Lysed extract T106K).

Gene ID	Description	SUMOylated lysine	SUMOylated peptide	Concensus motif
Tb927.7.2830	Histone H2A	5	MATP**K**QAVKKASKGGSSRSV	
Tb927.7.6360	Histone H2A variant (h2aZ)	32	GVAMSPEQASALTGG**K**LGGKAVGPAHGKGKG	
Tb927.10.15350	H3V (h3vaR)	31	SVASRPIQAVARAPV**K**KVENTPPQKRHHRWR	
Tb927.4.1330	DNA topoisomerase IB, large subunit	20	QKPKSGEGKGKKVAV**K**DEEVNGKRVVVKKED	+
		47	KKEDMTEEKIKKVVI**K**EEENELEMVAAGMGP	
Tb927.1.1170	DNA-directed RNA polymerase subunit RBP12 (RBP12)	7	MLSYTV**K**EEVKDEKLPGANNFA	+
Tb927.3.3220	RNA polymerase-associated protein CTR9	838	LEDFKELHGHRVPQV**K**NENEGFAESPAPWFS	+
Tb927.9.10680	RNA polymerase III RPC4	171	AEPKHEAEFSVEGDV**K**VPVAETGNDGIAFLK	Reversed
Tb927.9.5190	Proliferative cell nuclear antigen (PCNA)	195	ALLRASHAPTVDPRSK**G**ESDVKTEDEEADAC	+
Tb927.10.740	Structural maintenance of chromosome 4, putative (SMC4)	623	QIGNRMETPFTSPTP**K**AKRLFDLITPVNDRF	+
		1302	ERKRKRTGSTGDVQI**K**VEDEVACNNEAADIL	
Tb927.2.3480	Transcription elongation factor s-II (TFIIS2-2)	117	KAQAGEANERRVRGV**K**TEFADGGEQKAVEAP	+
Tb927.5.3210	Small ubiquitin-related modifier, putative (TbSUMO)	65	KSRTALKKLIDTYCK**K**QGISRNSVRFLFDGT	
Tb927.11.9920	Polyubiquitin	48	KEGIPPDQQRLIFAG**K**QLEEGRTLADYNIQK	
		63	KQLEEGRTLADYNIQ**K**ESTLHLVLRLRGGMQ	
Tb927.10.13720	RNA-binding protein, putative (RBP29)	313	GGEDTEPEKSATLPM**K**TESDSVSCPMEVNCN	+
Tb927.9.6870	RNA-binding protein, putative (RBSR1)	23	NIEKRATKKELLEFL**K**PMEEHIEDSWLARRP	
Tb927.9.15060	rRNA processing protein, putative	160	ARIEERAHRRAMKDQ**K**KYGKEVQAEVLRQRA	Reversed
		161	ARIEERAHRRAMKDQK**K**YGKEVQAEVLRQRA	
Tb927.10.1630	ATP-binding cassette sub-family e member 1	187	KMRVLLKPQYVDQVP**K**VTKGKVGDLLTKADE	
Tb927.10.14520	Basal body protein	371	VPAKDPGTAPSAVVA**K**SEKEEPPAAKTPRPT	+
Tb927.9.13880	Kinetoplastid membrane protein KMP-11	76	KFNKKMREHSEHFKA**K**FAELLEQQKNAQFPG	
Tb927.10.6720	Plasma-membrane choline transporter	16	MMQGFPSLGEKDPAA**K**PPAEGKPTSASGEKQ	
Tb927.9.9400	Ceramide phosphorylethanolamine synthase	312	SREVUEDGVPVAIVI**K**NEEMMNFDGKS	
Tb927.5.630	Acidic phosphatase	391	KDRKEDVASGSVHQG**K**GMPSTNIDPF	
Tb927.9.1560	DnaJ domain containing protein	407	KPTLKPSVSKSSISK**K**APTTSGSAKKAPTKR	
		416	KSSISKKAPTTSGSA**K**KAPTKRVVKPVKKAG	
Tb927.5.370	75 kDa invariant surface glycoprotein, putative (ISG75)	514	DDINIEEGGAKSKNT**K**TAAGLDSDI	
Tb927.3.2350	Hypothetical protein	35	SRINQALELATKVTV**K**EEPDTENGGGKAAII	+
Tb927.11.5230	Hypothetical protein	24	AAIYARAEEAQNAPV**K**LQPVPAFELYKAARE	
Tb927.9.13320	Hypothetical protein	484	VADGLAAVGLPEIRK**K**TDEEGEAEAEVAVKP	+
		498	KKTDEEGEAEAEVAV**K**PEPVVSLDG	
Tb927.10.12030	Hypothetical protein	100	GKGGTNKTSHDFELT**K**ALAKDGSIHREKRRL	Reversed
		139	ERFHKGNAHYSVSDS**K**PTKRRKRDHSVDKDA	
		167	KDAVVEAEFGSQSIN**K**KERSVKASRKDSNRE	
Tb927.4.2620	Hypothetical protein	235	ALTSAEERRHAEVDV**K**PNIYPTAT	Reversed
Tb927.3.4140	Hypothetical protein	455	ASIRQQMKNKKNAEI**K**VEEIEDVAVDDGAAA	+
Tb927.9.1410	Hypothetical protein	228	APEPLQCETEAVGNI**K**EELCKLQDSTAEVRL	+
Tb927.10.14190	Hypothetical protein	332	VVGRKRERRSCESGI**K**IEDHNNETVVEVGEL	+
Tb927.11.11840	Hypothetical protein	69	YCSRAPTEEKVVAQV**K**EEDGRGTGSSSRRNI	+
Tb927.10.15760	Hypothetical protein	2	M**K**SEGTPIPTRTPKSGK	+
Tb927.2.4520	Hypothetical protein	354	EDRYMFVVDQGNNTQ**K**EILDGIRAYIGGEIE	
Tb927.11.4920	Hypothetical protein	138	KEKDFVHRYAEAKMK**K**SIPREESDDVAKKVA	
Tb927.3.5370	Hypothetical protein	11	MPFWKYFVIT**K**PEVDDDDLPLAKHRI	+
Tb927.7.5660	Hypothetical protein	535	GPFTAHLLDPMPKFI**K**EETDTSKPAGLTKPR	+

Table 2. Site-specific identification of SUMOylated lysines and SUMOylated proteins in BF *T. brucei.* The modified Lys residues appear in bold.

Gene ID	Description	SUMOylated lysine BF	SUMOylated lysine PF*
Tb927.7.2830	Histone H2A	5	5, 9, 10, 129
Tb927.10.15350	H3V (h3vaR)	31	32
Tb927.4.1330	DNA topoisomerase IB, large subunit	20, 47	13, 20, 47, 75
Tb927.1.1170	DNA-directed RNA polymerase subunit RBP12 (RBP12)	7	7
Tb927.9.5190	Proliferative cell nuclear antigen (PCNA)	195	201
Tb927.10.740	Structural maintenance of chromosome 4, putative (SMC4)	623, 1302	1302
Tb927.10.13720	RNA-binding protein, putative (RBP29)	313	313
Tb927.10.14520	Basal body protein	371	371
Tb927.3.2350	Hypothetical protein	35	35
Tb927.11.5230	Hypothetical protein	24	24
Tb927.9.13320	Hypothetical protein	484, 498	498
Tb927.3.4140	Hypothetical protein	455	455
Tb927.9.1410	Hypothetical protein	228	271
Tb927.11.11840	Hypothetical protein	69	69
Tb927.3.5370	Hypothetical protein	11	132

Table 3. Comparative identification of specific lysine acceptor sites in BF and PF developmental stages. *Data from Iribarren *et al* (2015).

From the SUMO modified lysine residues identified, 21 of them were found in the sequence context that matches either consensus or reversed consensus motif. Analysis of SUMOylation sites with *pLogo* server (https://plogo.uconn.edu/) (O'Shea et al., 2013) led to 19 out of the 45 mapped lysine acceptor residues in which statistical significant positions contained V at position -1 (P-value=0.033) and E at position +2 (P-value=$4.223e^{-9}$) for a fixed K at position 0 **(Figure 2)**. Interestingly, there was a considerably enrichment of I at -1 compared with V preference at that position in PF data (log-odd: 3.026).
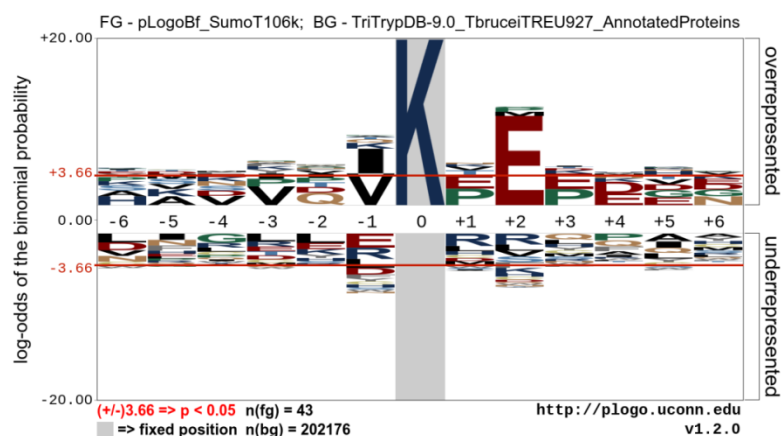
Figure 2 Analysis of SUMOylation sites. Representation of over- and underrepresented residues are shown above and below the X axis, respectively. Bonferroni-corrected statistical significance values are represented in horizontal ruler lines above and below the X axis. The Y axis corresponds to the logarithmic binomial probability of the statistical significance. Red horizontal lines correspond to threshold values of ±3.66, considering p<0.05. Legend to the figure: n(fg): number of aligned sequences in the input data set (foreground). n: TriTrypDB-9.0_Tbrucei927 protein data set used as background data.

4.1.3 -SUMO target validation

 A challenge intrinsic for SUMO target proteins identification is the sub-stoichiometric level and labile nature of this post-translational modification. Thus, we decided to validate some of the targets using a SUMOylation trypanosome-specific system by expressing essential pathway subunits in bacteria (Iribarren et al., 2015b). This system involves the co-expression in *E. coli* of the target of interest together with all trypanosome SUMOylation enzymes (*Tb*E1a/*Tb*E1b and *Tb*E2) including TbSUMO. The selected candidates, RNA polymerase-associated protein CTR9 (Tb927.3.3220) and the Transcription elongation factor s-II (TFIIS2-2) (Tb927.2.3480), were expressed in bacteria containing 3 copies of epitope tag Flag (3xFlag) as recombinant proteins together with the complete trypanosome SUMOylation system, or along without the system subunits as negative controls. After 5 h of induction with 1mM IPTG, whole bacterial extracts were analysed by WB using anti-Flag antibodies. As shown in **Figure 3A and B**, CTR9 and TFIIS2-2 were successfully expressed in *E.coli* and and only when the target was co-expressed with the whole trypanosome SUMOylation machinery, additional slower migrating bands were detected, suggesting larger bands correspond to mono- and multi-/poly-SUMOylated forms of each protein appeared.

Figure 3. *In bacteria* SUMOylation of CTR9 and TFIIS2-2. Validation of lysine acceptor sites for CTR9 (**A.**) and TFIIS2-2 (**B.**) by *in vitro* SUMOylation assay in bacteria. Anti-Flag Western blot analysis performed on soluble cells extracts from induced *E.coli* cultures transformed with pET28-CTR9-3xFlag (CTR9) and pET28-TFIIS2-2-3xFlag (TbTFIIS2) alone (Lane 1 in **A**. and **B**., respectively) or in the complete SUMOylation system background (pCDFDuet-1-*Tb*SUMO-*Tb*E2 plus pACYCDuet-1-*Tb*E1a-*Tb*E1b) for CTR9 (CTR9-SUMO-E1-E2, lane 2 in (**A.**) and TbTFIIS2 (TbTFIIS2-SUMO-E1-E2, lane 2 in **B**). Anti-Flag signal intensity was scanned using an LI-COR Odyssey and analyzed using ImageQuant software.

4.1.4 - *In situ* analysis of SUMO mediated interactions

To investigate if SUMO could act either as a modifier or as a scaffold protein inside the HSF, we performed a Proximity Ligation Assay (PLA) (O-link Bioscience) between SUMO and five substrates identified in the site-specific proteomic experiment: H3V, DNA topoisomerase IB, CTR9, Tb927.10.14190 and TFIIS2-2. PLA methodology utilizes target specific primary antibodies (raised in different species) and the corresponding oligonucleotide-labeled secondary antibodies (PLA probes). When the PLA probes are in close proximity, they support ligation and amplification of a product, which can be detected by a complementary fluorescently labeled oligonucleotide.

To this end, we generated transgenic cell lines to N-terminally tag the proteins with a 3xHA epitope and their expression was analyzed after 24 h of doxycycline induction by IFI. As shown in **Figure 4**, all targets exhibited a mainly nuclear localization in agreement with their described function.
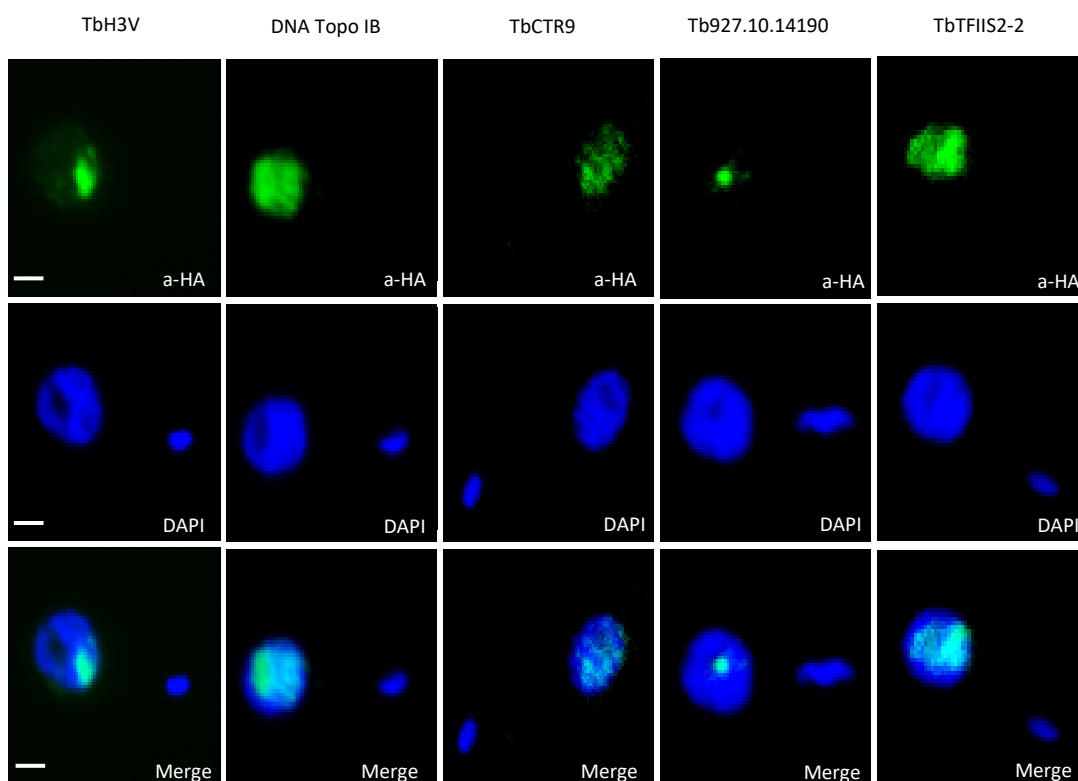
Figure 4. N-Terminal 3-HA tagging of representative SUMO conjugated proteins. Immunofluorescence analysis using anti-HA mAb of 3HA tagged SUMO conjugated partners corresponding to H3V (Tb927.10.15350), DNA topoisomerase 1B (Tb927.4.1330), CTR9 (Tb927.3.3220), Hypothetical protein (Tb927.10.14190) and TFIIS2-2 (Tb927.2.3420) after 24 h of Doxycycline induction. DNA content was stained with DAPI (blue). Merged anti-HA-DAPI images are shown. Scale bar, 1µm.

When we evaluated PLA reactions for each selected target using anti-SUMO and anti-HA, positive signals were detected in the nucli of all of them **(Figure 5A)**. Interaction signals were observed in a percentage of the population (from 20-55% of the cells depending on the cell line, **Figure 5A**) in agreement with the high dynamic range of this PTM. All together, these results imply the existence of multiple SUMOylated proteins and/or SUMO-interacting proteins in a particular focus within the nucleus.

Cell Line	anti-HA (%)	anti-TbSUMO (%)	anti-HA/TbSUMO (%)
H3V	1.21±0.05	0.17±0.17	20.16±0.88
DNA Topo IB	2.49±0.51	3.73±1.56	56.10±4.25
CTR9	5.21±4.33	3.71±2.08	46.55±4.91
Tb927.10.14190	5.33±1.76	2.30±1.63	32.73±3.46
TFIIS2-2	2.07±1.16	3.73±1.56	25.85±1.93
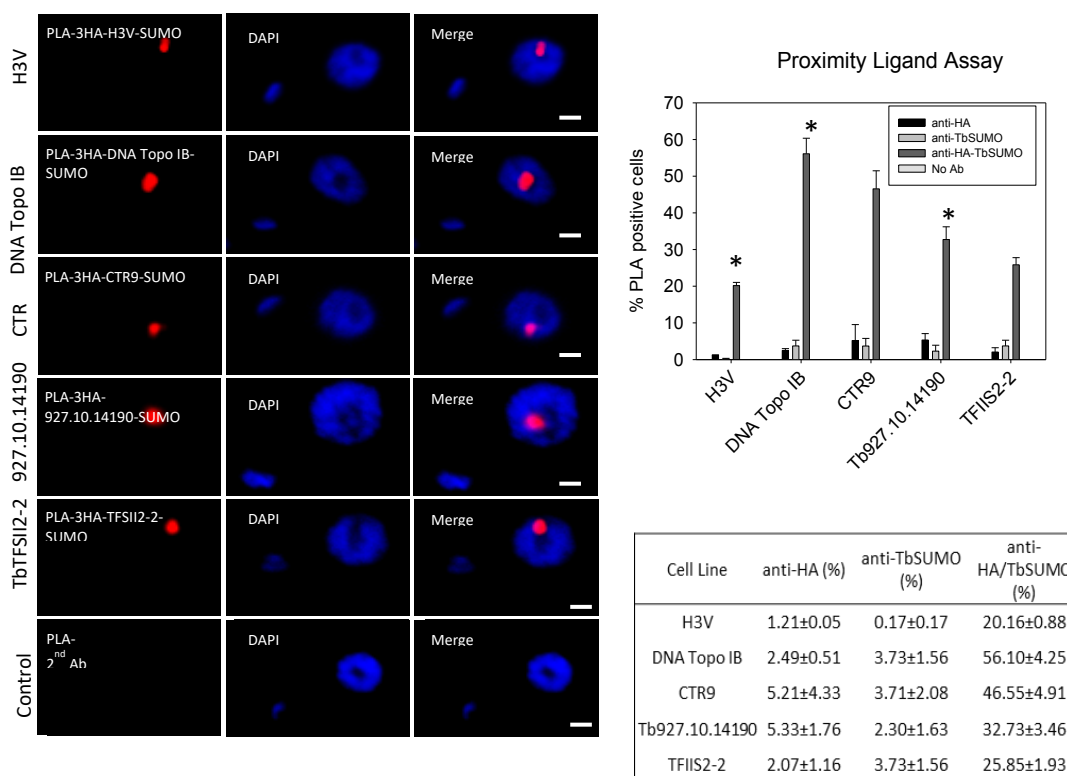
Figure 5. (A) *In situ* **analysis of SUMO mediated interactors.** Cells expressing 3HA tagged proteins were detected by *in situ* Ligand Proximity Assay (PLA) in BF using rabbit anti-HA polyclonal antibody (Abcam) and mouse anti-TbSUMO antibody after 24 h of doxycycline induction. Control reaction was performed without primary antibodies. DNA was DAPI stained. Histogram shows the percentage of PLA-positive cells for the TbSUMO/HA reaction. The % (±standard deviation) of positive PLA signals including false positives from each antibody incubated alone were quantified in 100x objective. At least 2 replicates per cell line were analyzed. Statistical analysis (Student T test) *p<0.05 (referred to anti-HA positive PLA). Scale bar, 1μm.

4.1.5 - Functional analysis of selected SUMO substrates

To investigate a potential functional relevance of SUMO conjugated proteins identified above in VSG expression, we depleted, by inducible RNA interference (RNAi), two of the proteins since tackling the whole list was out of the scope of this approach. We selected factors related to transcription such as, TFSII2-2 and the histone H3V and analyzed phenotypes after 24h of RNAi **(Figure 5B)**. Relative transcript levels by RT-qPCR analysis of cells depleted of TFSII2-2 and H3V indicated a reduced expression of active VSG221 (P<0.001 and P<0.05, respectively). Interestingly, the mRNA of a VSG located in an inactive telomeric expression site, the VSG121, was significantly increased (P<0.01 and P<0.001, respectively). The VSG121 messenger is normally detected from few cells (aprox. < 0.5%) in any population of VSG221 expressors that have undergone *in situ* transcriptional switch to the previously inactive VSG121 expression site. These results suggest the VSG switching frequency was increased after TFSII2-2 and H3V depletion. A control gene named C1, transcribed by RNA pol II, remained unmodified in any of the depletion experiments suggesting cell general fitness had not been severely affected. However, 18S ribosomal rDNA expression was considerably reduced after H3V depletion (P<0.05) suggesting this histone variant is required for RNA pol I transcription.
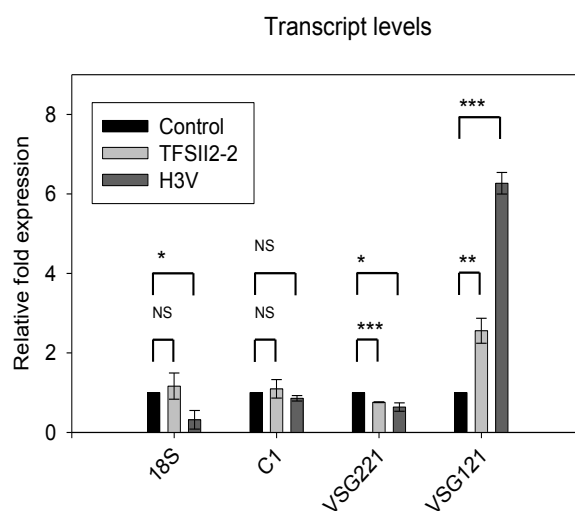
Figure 5. (B) Depletion of TFIIS2-2 and H3V deregulates both active and inactive VSG-ES. Relative transcript levels upon 24h of RNAi shows deregulation of active VSG221 and inactive VSG121. Only 18S ribosomal rDNA is altered after H3V depletion. C1 (AN1-like zinc finger, Chr. X) is represented as a control locus for qPCR valdation. Results represent the average from three independent clones and data normalized with U2 mRNA (mean±SEM). Statistical analysis (Student T-test): p<0.05, **p<0.01, ***p<0.01, NS: Non-significant.

4.2 -Transcriptional activator SNF2PH as a novel SUMO target

The main scope of this work is to unravel some factors associated to the ESB and/or related to VSG expression. Some of the previous described factors obtained during the site-specific identification constitute reliable SUMO substrates; nevertheless, their VSG-related function remains unclear, despite that they probably are associated to the ESB. To solve that, we returned to a previous proteomic screening of SUMO modified proteins by using the cell line expressing an 8xHis and HA-tagged TbSUMO version. As stated previously, this method is less-specific at the same time that the background increases; however, it expands the range of proteins that might retain a VSG-related function. Thus, the 8xHis and HA-tagged TbSUMO approach led us to identify among other proteins **(Appendix I Table S1)**, an uncharacterized protein annotated in the TriTrypDB database as Transcription Activator SNF2PH (Tb927.3.2140, length 948aa), which was reproducible in two independent proteomic screenings **(Appendix I Table S1B & C)**. Structural domain analysis suggests is a member of the Snf2 (Sucrose Nonfermenting Protein 2), superfamily of chromatin remodelling factors (Flaus et al., 2006; Prasad and Ekwall, 2013) using ATP as substrate (Hu et al., 2013). In yeast, Snf2 either activates or represses transcription depending on the transcriptional context by dissociating basic transcription factors from different promoters (Whitehouse et al., 2003). Chromatin remodelers function by reconfiguration on the surface of the nucleosomal substrates, deforms nucleosomal DNA at the site of binding and translocates the DNA over the surface of nucleosomes (Farnung et al., 2017; Liu et al., 2017; Whitehouse et al., 2003). Interestingly, trypanosome SNF2PH contains a Plant Homeodomain (PHD) that is absent from other known trypanosome chromatin remodelers. The PHD finger is a conserved domain that binds H3 tails modified by PTMs in higher eukaryotes, thereby conferring specificity (Musselman and Kutateladze, 2011). PHD reads unmodified H3 tails as well as H3 trimethylated at Lys4 (H3K4me3) (Pena et al., 2006) or acetylated at Lys9 and Lys14 (Musselman and Kutateladze, 2011) but also possesses nucleic acid binding activity (Weaver et al., 2018), regulating nucleosome and DNA translocation to regulate transcriptional activity. Importantly, trypanosome histones are highly divergent, such that the precise recognition determinants for SNF2PH are unlikely to be the same as for metazoa. In addition, PHD fingers are important for maintaining gene expression during development (Wysocka et al., 2006). The function of SUMOylation in PHD binding to histones may constitute a regulatory link to chromatin remodelling activity regarding developmental gene expression in *T. brucei.*

4.2.1 - Sequence analysis of SNF2PH

In silico analysis of SNF2PH compared to other Snf-2 like family members revealed a significant conservation of the characteristic ATP binding SNF2_N domain and the helicase **(Figures 6A and B)**. Trypanosome SNF2PH showed high sequence homology to the *H. sapiens* HsSMCA5 and *M. musculus* MsSMCA1 (ISWI subfamily), in addition of to the chromodomain-helicase-DNA 1-like domain included in MsCHD1L and the *S. cerevisiae* ScCHD1L. A distinct feature that differentiates SNF2PH from other chromatin remodellers described in trypanosomes is the presence of a Plant Homeodomain (PHD). Alignment of SNF2 PHD with the PHD contained in the well-characterized MmHMT3 and HsNSD3 showed high conservation **(Fig 6C)**.
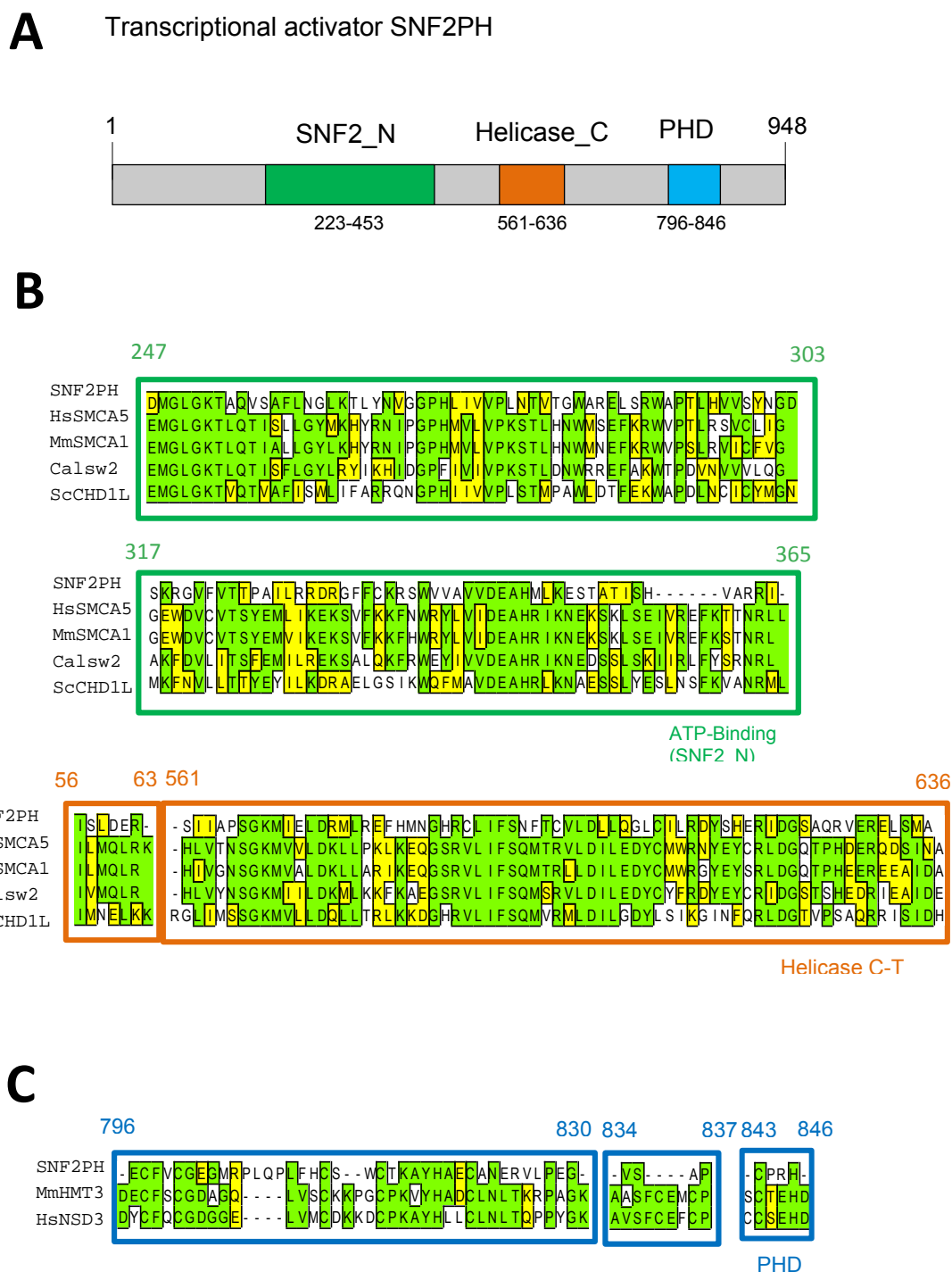
Figure 6. **The SNF2 N and helicase domains are conserved in trypanosome SNF2PH whilst PH domain is normally included in methyltransferases. (A)** Schematic representation of the conserved regions of SNF2PH in other organisms. SNF2 N (green), Helicase C (orange) and PHD3 (blue). **(B)** Sequence alignment of SNF2 N (ATP-binding site) and Helicase C-terminal domains of SNF2PH (Tb927.3.2140), *H. sapiens* SMCA5 (O60264.1), *Mus musculus* SMCA1 (Q6PGB8.1), *M. musculus* CHD1L (Q9CXF7.1), *C. albicans* Isw2 (Q5A310.1) and *S. cerevisiae* CHD1L (P32657.1). **(C)** SNF2PH contains a PHD3 domain characteristic of HMTases. Sequence alignment of the PHD3 domain of *T. brucei* SNF2PH, *M. musculus* HMT3 (O88491) and *Homo sapiens* NSD3 (Q9BZ95.1).

4.2.2 - SNF2PH chromatin remodeler is developmentally regulated is associated with the nuclear body ESB

In order to investigate SNF2PH protein expression, we generated a monoclonal antibody (mAb) against *T. brucei* SNF2PH (11C10E4). Western blot analysis of total protein extracts showed SNF2PF protein level is developmentally regulated, higher in the bloodstream form compared with insect procyclic form **(Fig 7A & Appendix II Figure S1)**. To determine the specificity of the anti-SNF2PH antibody, we knocked down protein expression by RNAi and analysed whole cell extracts by Western blot. As expected, the SNF2PH protein intensity was markedly reduced in induced RNAi cells, while no other signal was detected by the mAb, confirming specificity **(Figure 7B).**

Subcellular localization of SNF2PH by 3D-deconvolution immunofluorescence microscopy (3D-IF) using the anti-SNF2PH mAb showed a predominant nuclear localization throughout the cell cycle, with a disperse distribution in the nucleoplasm including punctuate structures and enrichment at the nucleolus periphery **(Fig 7C)**. To investigate whether SNF2PH associates with the active VSG-ES locus, we used a GFP-lacI targeted VSG promoter cell line, allowing the nuclear position of the active VSG-ES locus to be visualised (Navarro and Gull, 2001). Thus, we detected 38.8% (n=55) co-localization with the GFP-tagged active ES by double 3D-IF, using an anti-GFP rabbit antiserum (Invitrogen) and 11C10E4 mAb anti-SNF2PH **(Fig 7D)**, suggesting SNF2PH partially colocalizes with the active VSG-ES (Pearson's correlation coefficient).

To investigate possible SNF2PH subnuclear association with the high SUMOylated focus (HSF), we labelled cells with an anti-TbSUMO mAb (López-Farfán et al., 2014) and a rabbit polyclonal antiserum against SNF2PH (see **Appendix Methods**). 3D-IF analysis again showed a partial colocalization between SNF2PH and HSF (53.7% of the cells (n=67)) **(Fig 7E)**, suggesting a variable association in nuclei of SNF2PH with the HSF, which we assume is due to the highly dynamic nature of SUMOylation. Similar colocalization in 53.49% of the cells was observed between SNF2PH and the ESB **(Fig 7F)**, revealed as the extranucleolar pol I signal visualized with a YFP-tagged RPB5z (specific RNA pol I subunit 5z (Landeira et al., 2009)).
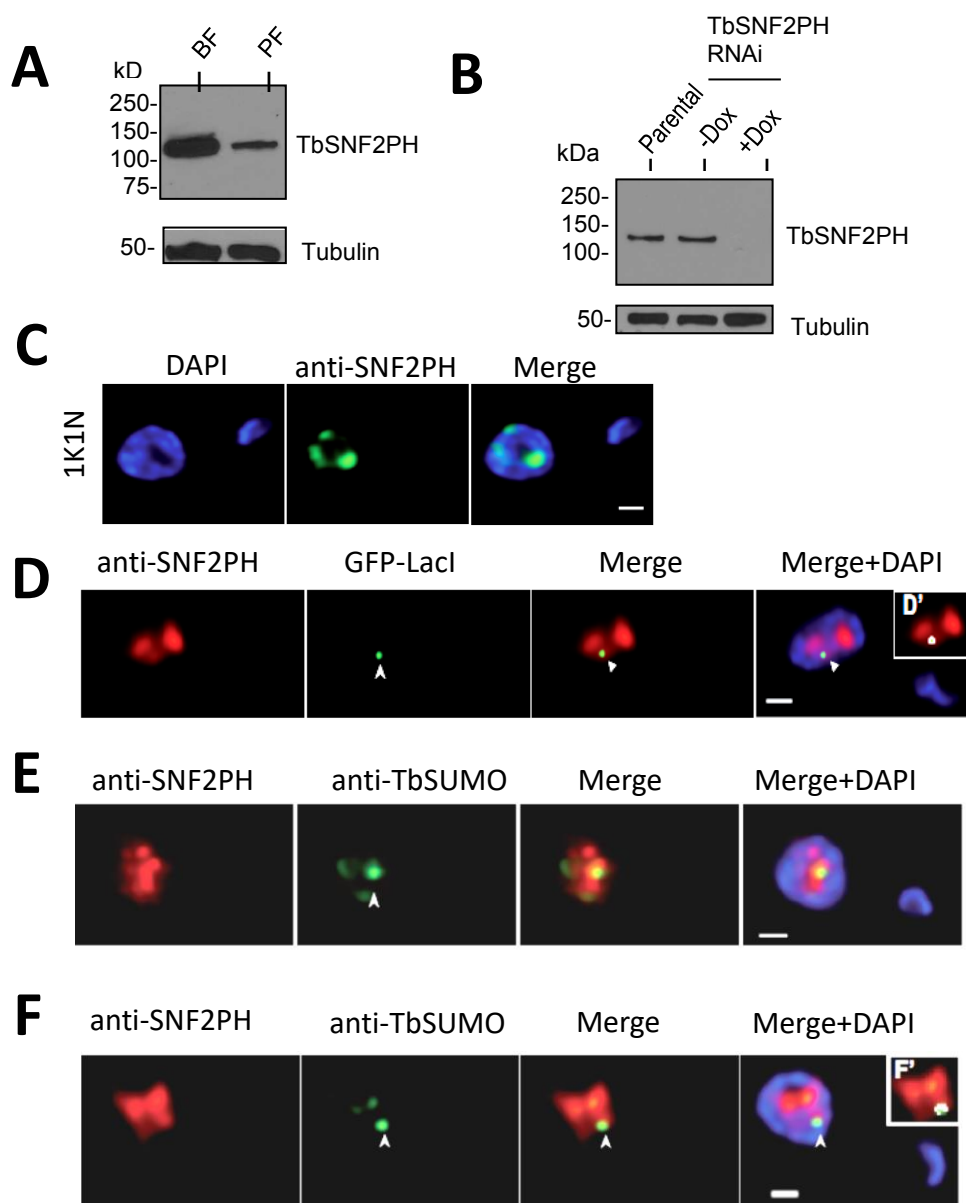
Figure 7. SNF2PH is a developmentally regulated protein associated with the expression site body (ESB) and highly SUMOylated focus (HSF). (A) SNF2PH is differentially expressed in *T. brucei* developmental stages. **(B)** Knockdown of SNF2PH in bloodstream form leads to protein depletion after 24h. (5x10⁶ cells/lane); parental, uninduced (dox-) and induced (dox+). Total cell extracts were analyzed by Western blotting using the anti-SNF2PH mAb. **(C)** SNF2PH is diffusely distributed in the nucleoplasm with certain enrichment in the nucleolus. Panels show DAPI and green channels after IF with the anti-SNF2PH mAb (11C10E4). Scale bar, 1 μm. **(D)** SNF2PH associates with the active VSG-ES. A cell line with a GFP-LacI tag in the active VSG ES (Navarro and Gull, 2001) was subjected to double 3D-IF using anti-SNF2PH mAb (red), anti-GFP rabbit antiserum (green) and DAPI staining. Maximum intensity projections of deconvolved slices containing the GFP signal are shown (arrowhead). (D') Inset shows a higher magnification of the nucleus including anti-SNFPH and anti-GFP fluorescence signals colocalization mask (white). Scale bar, 1 μm. **(E)** Colocalization analysis of SNF2PH with the Highly SUMOylated focus (HSF). SNF2PH associates with the HSF (arrowhead) in bloodstream form nucleus. Indirect 3D-IF analyses were carried out using the rabbit anti-SNF2PH antiserum (red) and the anti-TbSUMO mAb 1C9H8 (green) (López-Farfán et al., 2014). Scale bar, 1 μm. **(F)** SNF2PH partially colocalizes with YFP-tagged TbRPB5z in the ESB. A cell line expressing an N-terminal YFP-RPB5z fusion (Landeira et al., 2009), an RNA pol I-specific subunit, was analyzed by double 3D-IF with anti-SNF2PH mAb (red) and rabbit anti-GFP antiserum (green). Deconvolved slices containing both SNF2PH and the extra-nucleolar ESB (arrowhead) are represented as maximum intensity projections. (F') Inset shows a higher magnification of the nucleus including anti-SNFPH and anti-GFP fluorescence signals colocalization mask (white). Scale bar, 1 μm.

4.2.3- SNF2PH is SUMO conjugated

Partial colocalization of SNF2PH with the HSF, and identification from a SUMO -enriched extract, suggests modification by TbSUMO among other SUMOylated proteins present. To investigate whether SNF2PH is a "bona fide" SUMO target, we carried out immunoprecipitation (IP) assays utilizing the anti-TbSUMO mAb (López-Farfán et al., 2014) and anti-SNF2PH antiserum under denaturing conditions in order to identify only SUMO covalently bound to targets. IP experiments showed that SNF2PH is SUMOylated when analyzed by Western blotting using the anti-TbSUMO mAb on a SNF2PH immunopreciptated extract **(Fig 8A).** The reciprocal experiment using an anti-SNF2PH antiserum on a TbSUMO immunoprecipitated extract reproducibly detected SNF2PH conjugated to TbSUMO **(Fig 8B)**. While these immunoprecipitations demonstrated that SNF2PH is a SUMOyolated protein, is unknown if nuclear conjugation with SUMO is associated with dispersed nuclear foci or localisation to a specific nuclear site. To investigate this possibility, we performed Proximity Ligation Assays (PLA). After a first IF experiment using anti-SUMO mAb and antiSHF2PH Rb antiserum, the assay detected a positive amplification dot, suggesting that SNF2PH is SUMOylated in situ in both the nucleolus and nuclear periphery in one (84.12%±0.25%) or two dots (15.88%±0.25%) **(Appendix II S2A and S2B Figs).**

The low signal of SNF2PH antibody in TbSUMO immunopecipitations is due to the fact that SUMOylation is a highly dynamic process and only a small percentage of SNF2PH molecules are SUMOylated in a given time This has also been reported for TbRPAI (RNA Polymerase I largest subunit) (López-Farfán et al., 2014) and other SUMO targets described for *T. brucei* (Iribarren et al., 2015a).

Due to the highly dynamic SUMOylation process, an alternative approach was required to investigate SUMO conjugation in detail, based on the expression and purification of recombinant tagged proteins in bacteria, which allowed an *in vivo* reaction in a heterologous system lacking SUMO. Thus, we used a bacterial strain engineered to express the complete *T. brucei* SUMOylation system together with the target protein (Iribarren et al., 2015b). To facilitate soluble expression of the recombinant SNF2PH protein we evaluated two different constructs containing SNF2PH N-terminal or C-terminal domains (SNF2PHN and SNF2PHC, respectively). Next we co-expressed processed TbSUMO (already exposing the diGly motif) with both activating enzyme subunits (TbE1a/TbE1b) and the conjugating enzyme (TbE2) using two different Duet® vectors, while a compatible pET28 vector was used to co-express the SNF2PH fragments harbouring a tri-Flag tag that allows the detection of the SUMOylated forms by Western blot analysis directly on cell lysates without any enrichment step. As shown in

Figure 8C, SNF2PHN appears as a single band migrating at the expected size when expressed alone in *E. coli* (lane 1) or when co-expressed with the partially reconstituted system (lane 2 and 3). However, when co-expressed with the complete SUMOylation system, additional slower-migrating bands can be detected (lane 4) suggesting that this domain can be SUMO conjugated. In contrast, the C-terminal domain of SNF2PH is not a target of SUMOylation since in all cases was visualized as a single protein band at the expected molecular weight **(Fig 8D)**.

To validate heterologous SUMOylation of SNF2PHN, we performed *in vitro* deconjugation reactions using the specific *T. brucei* SUMO isopeptidase TbSENP, obtained as a recombinant protein. As shown in **Figure 8E**, the additional slowly migrating bands observed when SNF2PHN was co-expressed with the *T. brucei* SUMOylation system in bacteria (lane 1) completely disappeared upon treatment of cell lysates with TbSENP (lane 2), and the deconjugation ability of TbSENP was specifically inhibited by the addition of 20 mM NEM (lane 3). To investigate the nature of SUMOylation of SNF2PHN, we compared the patterns obtained in bacteria when replacing wild-type SUMO with a variant that is unable to form SUMO chains **(Fig 8F)**. In the latter case, a doublet near the 55 kDa marker can be detected suggesting that there are at least two major sites for SUMOylation in SNF2PHN. Taking into account that when SUMO branching is allowed one of the bands remained almost constant while the other disappeared with a concomitant appearance of high molecular weight adducts, we speculate that SNF2PH possess two major sites, one being mono SUMOylated and the other likely being poly-SUMOylated.
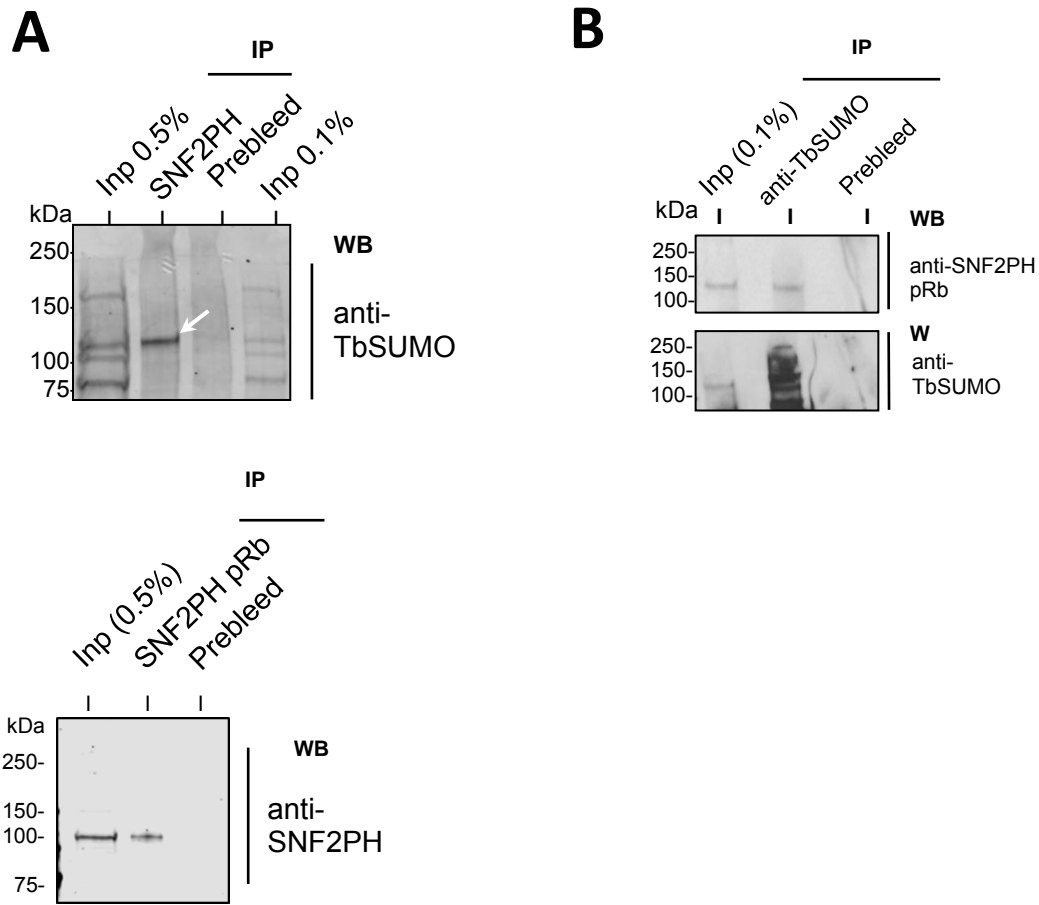
Figure 8. Coimmunoprecipitation and *in vitro* SUMOylation assay using a heterologous system demonstrate SNF2PH is a SUMO target. (A) Immunoprecipitation (IP) of bloodstram SUMOylated proteins revealed that SNF2PH is SUMOylated. A nuclear fraction was lysed in urea-containing buffer and proteins were immunoprecipitated with rabbit anti-SNF2PH antiserum or unspecific antiserum (prebleed) and probed with anti-TbSUMO mAb 1C9H8 (arrow). As a control IP samples were reprobed with SNF2PH antiserum (right). **(B)** A reciprocal IP experiment was performed using anti-TbSUMO mAb and probed with SNF2PH antiserum. As a control, the blot was reprobed with anti-TbSUMO mAb (lower panel). Inp: Input, IP (0.5%).
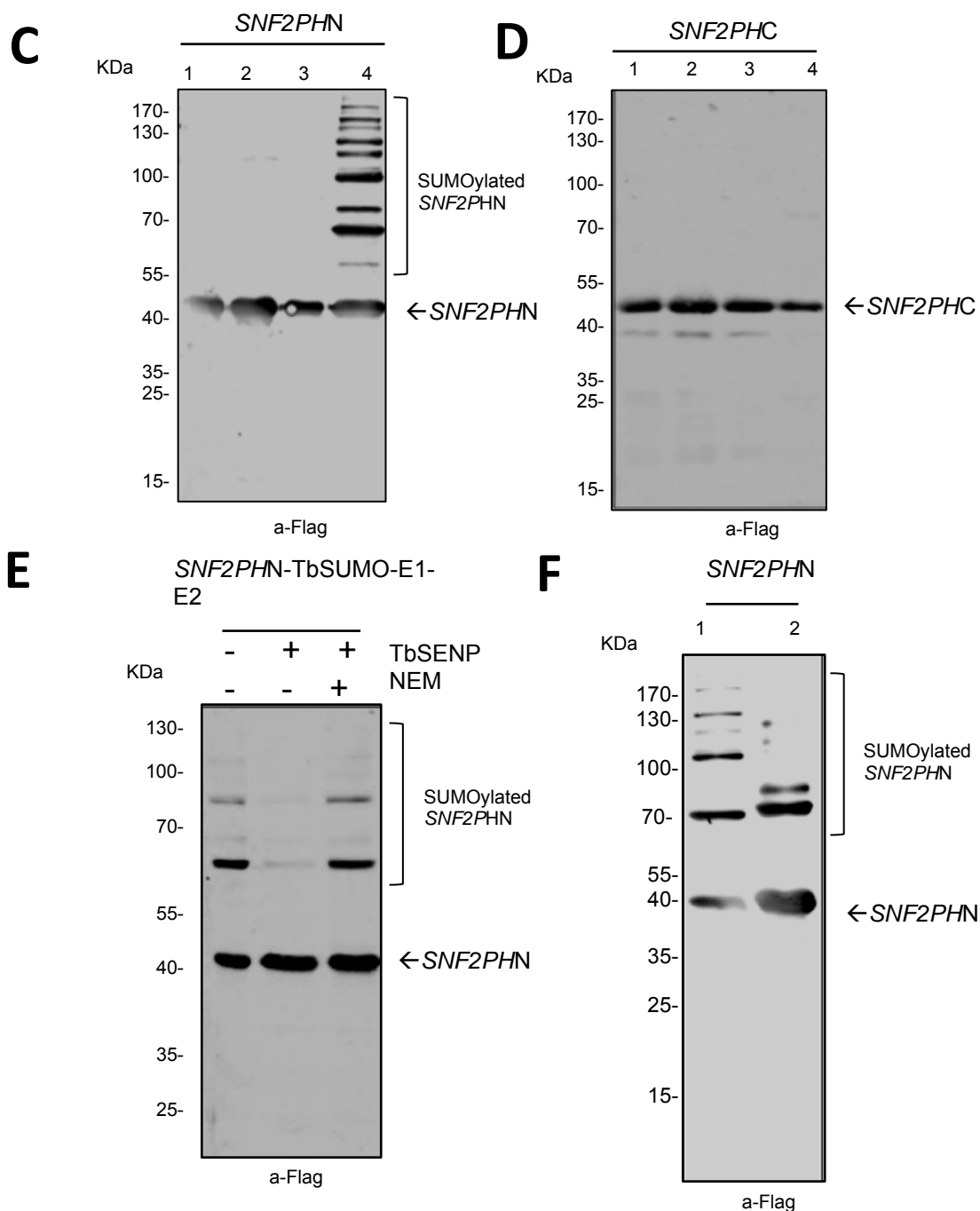
Figure 8. (C) Anti-Flag Western blot analysis of SNF2PHN performed on soluble cell extracts from induced cultures of *E. coli* transformed with pET28-SNF2PHN-3xFlag alone (lane 1), or in the background of an incomplete (lane 2, pACYCDuet-1-TbE1a-TbE1b; lane 3, pCDFDuet-1-TbSUMO-TbE2) or a complete (lane 4, pCDFDuet-1-TbSUMO-TbE2 plus pACYCDuet-1-TbE1a-TbE1b) SUMOylation system. **(D)** Similar samples as described in **(C)** were analyzed for SNF2PHC. **(E)** Cell lysates of E. coli heterologously expressing SNF2PH and the complete *T. brucei* SUMOylation system were incubated at 28°C in the absence (-) or presence (+) of recombinant TbSENP. The deconjugation activity of TbSENP was specifically inhibited by the addition of 20 mM NEM. Reaction mixtures were analyzed by Western blot using anti-Flag monoclonal antibodies. **(F)** Western blot analysis of SUMOylated SNF2PHN pattern performed on soluble cell extracts from a complete bacterial SUMOylation system using a wild type version of SUMO (lane 1) or a Lys-deficient version of SUMO (TbSUMO K9R) unable to form chains (lane 2).

4.2.4 - SNF2PH is highly enriched in the chromatin upstream of the active VSG-ES promoter while reduced at inactive ones

To study SNF2PH occupancy within the VSG-ES locus, we performed chromatin immunoprecipitation (ChIP) using anti-SNF2PH antiserum and chromatin isolated from a VSG tagged promoter cell line DRALI (Double Renilla Active site Luciferase Inactive site). To overcome the problem of high sequence similarity between multiple VSG-ESs, we developed a dual-reporter cell-line which contains the Renilla luciferase (*RLuc*) gene inserted 400 bp downstream of the active VSG-ES promoter, and the Firefly luciferase (*FLuc*) gene downstream of an inactive VSG-ES promoter **(Fig 9A)**. The reporter genes in the DRALI cell line allowed us to determine SNF2PH occupancy at the region downstream of the promoter in either active or inactive VSG-ESs. First, we analyzed occupancy of SNF2PH by ChIP and quantitative PCR (qPCR), which detect significant SNF2PH enrichment at the RLuc gene downstream of the active VSG-ES promoter (p<0.001) as well as at the active VSG221 gene located in the telomere of BES1 (p<0.01). However, FLuc located downstream of an inactive VSG-ES promoter **(Fig 9A)** was not significantly detected **(Fig 9B)**. SNF2PH enrichment was also not detected at VSG genes known to be located at silent telomeric ES position in this strain (Hertz-Fowler et al., 2008), such as VSG121 (VSG in BES3), VSGVO2 (BES2) and VSGJS (BES13) **(Fig 9B)**. Altogether, the active VSG221 gene immunoprecipitated more efficiently than all inactive VSG telomeric loci analyzed, suggesting SNF2PH associates preferentially with the active ES telomere. Additionally, SNF2PH was detected at RNA pol I-transcribed loci, such as the rDNA (p<0.01) and EP procyclin (p<0.05) promoters. However, SNF2PH was most enriched in the active VSG-ES chromatin compared with EP and rDNA promoters.

Occupancy of SNF2PH at the two promoters of the surface glycoprotein genes characteristic of mammalian and insect forms (VSG-ES and EP) implicates SNF2PH in regulation of developmental gene expression. Enrichment was also detected for the splice leader (SL) promoter (p<0.05) and coding regions. To investigate this possibility, we analyzed potential changes in the occupancy of TbRPB1 (largest subunit of the RNA pol II polymerase) on these loci upon SNF2PH depletion. Analyses by qPCR of Immunoprecipitated DNA using anti-TbRPB1 affinity purified antiserum, did not show significant changes in RNA pol II enrichment in both spliced leader promoter (SL Pro) and coding region (SL coding) in a SNF2PH depleted cell line **(Appendix II S3 Fig)**. These data suggest that the major activity of SNF2PH is not on RNA pol II transcription.

In eukaryotes, chromatin remodelers are detected at RNA pol II promoters and play important roles in their activity (Lai and Wade, 2011);(Cairns, 2009). We investigated the presence of

SNF2PH in chromatin across the genome, aside from the multiallelic VSG-ESs, to identify additional genes targeted by this protein. We compared quantitative enrichment profiles with ChIP-seq peak distribution and considered 0-mismatch error to avoid bias in polymorphic sequences within repetitive chromosomal loci, leading to defined peaks (q value<0.05) **(Appendix I Table S4, Fig 9C)**. As demonstrated by quantitative ChIP, the site of enrichment corresponded to developmentally regulated loci EP and GPEET2 procyclin, and 18S ribosomal DNA and SL RNA-related sequences. Interestingly, SNF2PH localizes at H3.V, a histone variant recently associated with VSG monoallelic expression (Müller et al., 2018) and its own coding sequence. SNF2PH was also significantly enriched in a substantial number of tRNAs gene arrays located in chromosomes XI, X, VIII, VII, V, IV, and III **(Fig 9C)**, interestingly tRNA clusters are known to function as a chromatin insulators in eukaryotes from yeast to human, reviewed in (Van Bortle et al., 2014).

Figure 9B, shows SNF2PH is enriched in the active VSG-ES (BES1) locus by ChIP qPCR analysis using unique sequences as the Rluc reporter inserted downstream of the promoter and the VSG221 gene at the telomeric end. However, we wish to investigate in detail a possible SNF2PH occupancy at the promoter adjacent area, nevertheless highly homologues sequences shared among most of the VSG-ESs (BES1 to BES17 in Hertz-Fowler et al., 2008) prevent this analysis by a simple ChIP-seq alignment. We have previously reported polymorphism in the sequence at particular regions located at the core promoter and upstream the promoter region, referred as ES promoter PCR fragments 1 and 4 (ESPM1 and ESPM4) (schematically represented in **Fig 9D**) (López-Farfán et al., 2014). These minor polymorphisms in the sequences allowed us to differentiate among different BES promoter regions. In particular, PCR fragment ESPM 4 and 1 yielded 14 different sequences when genomic DNA was used as template (Figure S5 in (López-Farfán et al., 2014), providing considerable covering of most of the BESs (Hertz-Fowler et al., 2008)). ChIP-seq data was generated from the immunoprecipitated chromatin with the anti-SNF2PH antiserum which was then PCR amplified with ESPM1 and 4 PCR primers, and the products were deep-sequenced (see Material and Methods). Reads were aligned on BES promoter sequences index file built by combining in a single lane the sequences from the ESPM1 and 4 together with the sequences from the corresponding BESs (BES1, 2, 3, 4, 7, 7dw, 10, 10dw, 12, 13, 15, 15dw, 17, 17dw described previously (Hertz-Fowler et al., 2008)). Next, using Bowtie software, alignment of the reads were assigned to the BES promoter index file, and the number of reads aligning to each BES was represented in **Figure 9D**. This data showed SNF2PH is enriched to the active BES1, at the ESPM1, which correspond to the core promoter of the active BES1 (VSG-ES221). SNF2PH was

detected to a lesser extent in other BES promoters suggesting that SNF2PH is controlling inactive promoters as well **(Fig 9D).** Interestingly, the increase of read count was at the ESPM1 fragment region where the actual ES promoter is located, suggesting SN2PH is associated with the active core ES promoter rather than the upstream promoter region**.** Together these data indicate that SNF2PH is located in several BES promoters, however it is most enriched at the active VSG-ES promoter (ESPM1 of the BES1) **(Fig 9E)**.

In *T. brucei*, SUMOylated chromatin-associated proteins are particularly enriched upstream of the promoter region, suggesting the presence of structural components of chromatin susceptible to this modification (López-Farfán et al., 2014). Considering the presence of SUMO-enriched chromatin associated proteins upstream of the active VSG-ES (López-Farfán et al., 2014), we performed a sequential ChIP analysis to discriminate the number of sequences relative to the active VSG-ES, which appear to be SUMOylated from the overall SNF2PH immunoprecipitated sequences. For this, we carried out a ChIP assay from input DNA which was previously immunoprecipitated using the mAb against TbSUMO to ensure enrichment of SUMOylated active VSG sequences. Then, a second round of immunoprecipitation was achieved using the polyclonal antibody against SNF2PH and a nonspecific serum as a negative control. After cloning and sequencing the immunoprecipitated genomic fragments, we quantified the number of sequences corresponding to the active and other telomeric inactive sites **(Appendix II S4 Fig)**. Surprisingly, we detected 11 sequences corresponding to the active VSG-ES in contrast to four sequences in the control sample (n=50). This result suggests that SNF2PH is enriched at sequences upstream the active VSG-ES promoter.

A

DRALI cell line

B
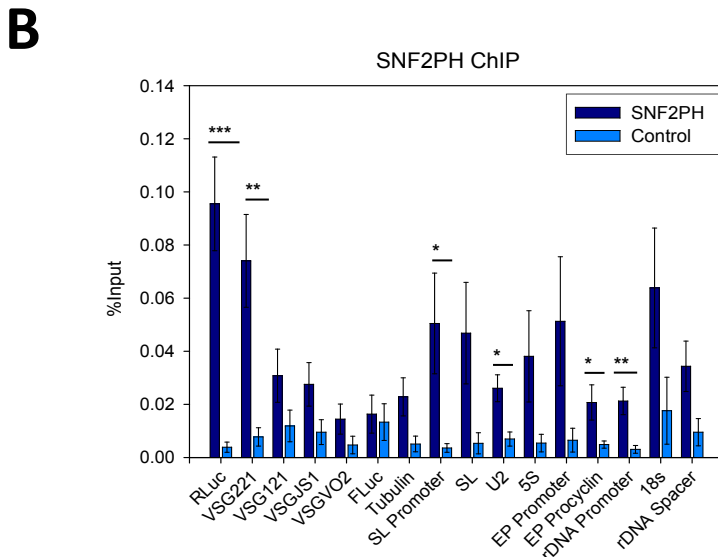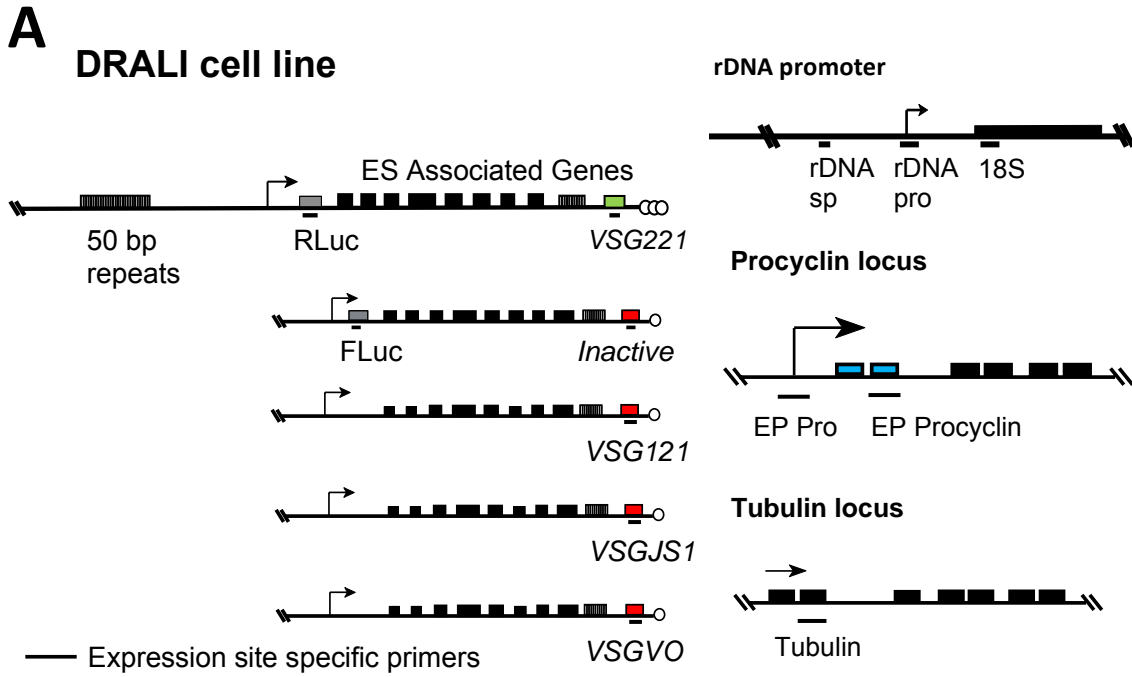
Figure 9. SNF2PH is enriched upstream of the active VSG-ES chromatin while in silent promoters is also detected. (A) Schematic representation of loci of interest in DRALI, the dual-reporter cell line (not to scale). Two reporters were inserted, Renilla-luciferase gene (RLuc) downstream of the Active VSG221-ES (BES1) promoter and firefly Luciferase gene (FLuc) downstream of an Inactive VSG-ES (DRALI). Few other inactive VSGs known to be telomeric in this strain are also represented (VSG121 (BES6), VSGJS1 (BES13) and VSGVO2 (BES2)). Schematic representations for other chromosomal loci (rDNA promoter, Procyclin and Tubulin) are shown. Color code: grey (reporters), green (active VSG-ES), red (Inactive VSG-ESs), blue (procyclin locus). Arrow (promoters). **(B)** Chromatin at the active VSG-ES is enriched for SNF2PH. Chromatin immunoprecipitation (ChIP) analysis by quantitative PCR of reporter sequences inserted downstream of the VSE-ES promoters suggests SNF2PH is highly enriched in the active 221VSG-ES (RLuc) compared to an inactive VSG-ES promoter (FLuc) (***$p<0.001$). SNF2PH enrichment on the active telomeric VSG221 compared to inactive VSGs was also significant (*$p<0.05$ -**$p<0.01$). SNF2PH occupancy was detected at the splice leader promoter (SL promoter, Pol II-transcribed) and EP Procyclin ($p<0.05$), as well as the rDNA promoter ($p<0.01$). (Student's t-test) *$p<0.05$ **$p<0.01$, ***$p<0.001$). ChIP analyses are shown as the average of at least three independent experiments with standard error of the mean (SEM). Data are represented as percent of input immunoprecipitated (% input).

84
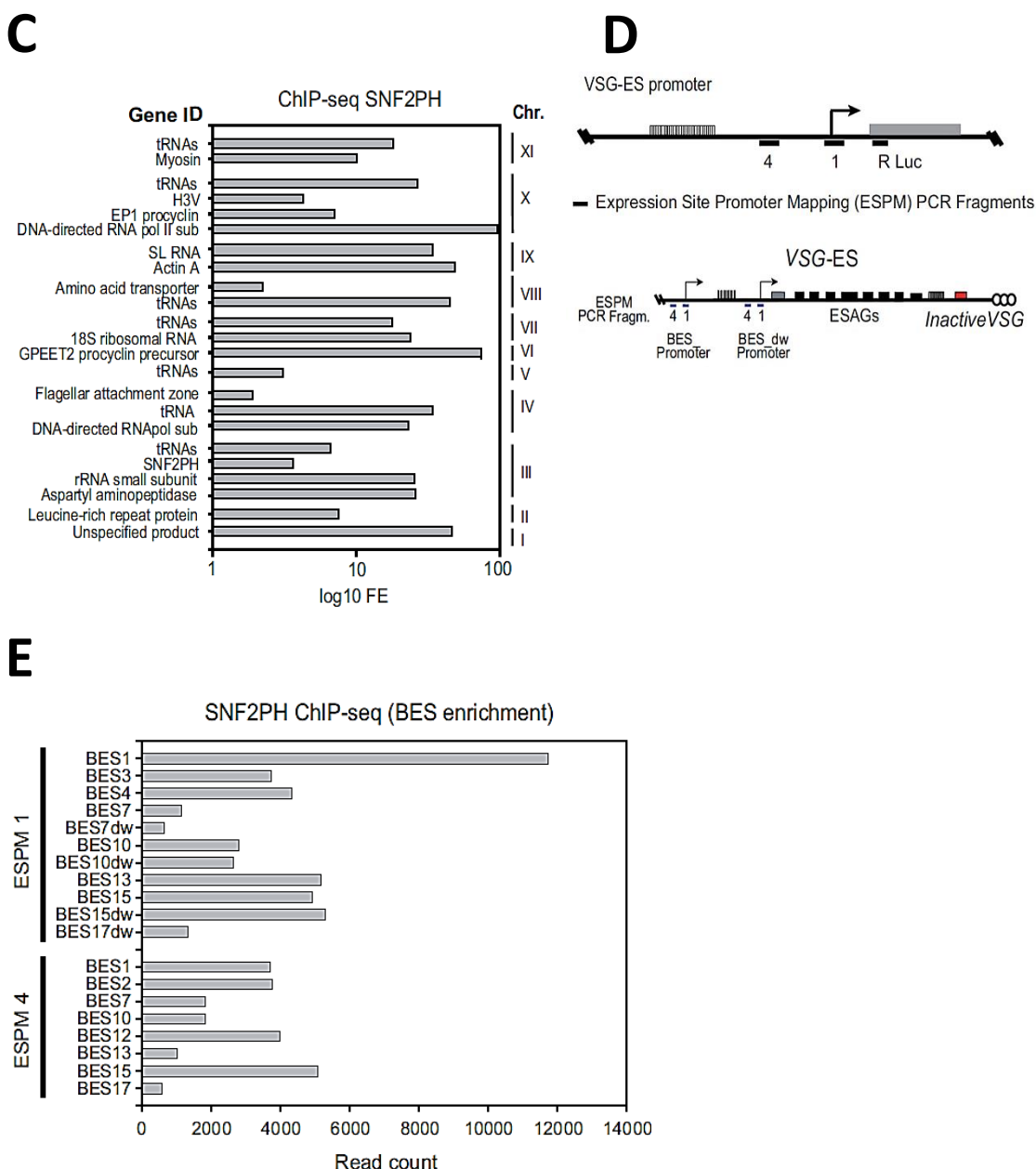
Figure 9. (C) Distribution of SNF2PH on internal chromosomal positions. ChIP-seq analysis using the SNF2PH antiserum and *T. b. brucei* 427 genomic library (v4). Histogram illustrates peak enrichment of representative genes expressed as log10 fold enrichment (FE). This global analysis confirmed previous ChIP data locating SNF2PH on developmentally regulated genes (EP and GPEET), RNA pol I driven rDNA (ribRNAs) and the SL cluster of small RNAs that are trans-spliced in every mRNA. Interestingly, beside those essential genes for cell growth, SNF2PH occupies few other coding genes; noteworthy is H3V protein recently linked to the regulation of VSG monoallelic expression (Müller et al., 2018). In addition, SNF2PH was consistently enriched in tRNA gene clusters in 7 chromosomes. Due to highly homologous sequences among ESAGs, all ES related sequences (ESAGs genes, VSG-BC, etc.) located in chromosomal internal positions were excluded of this graph, since we cannot rule out whether the ChIP-seq reads came from the active VSG-ES Schema of VSG-promoter region mapping indicating ESPM fragments amplified by qPCR. **(D)** Schema of VSG-promoter region indicating the location of ESPM PCR fragments amplified (upper panel). Detailed schema of the promoter region showing both upstream and downstream (dw) BES from the tandem repeated promoters ESPM 1 and 4 (lower panel). **(E)** Chromatin at the core promoter of the active VSG-ES is highly enriched in SNF2PH. ChIP-seq data using SNF2PH antiserum reveals a higher number of reads corresponding to the sequence polymorphism of the BES1 at the PCR fragment 1, (ESPM1) mapping at the VSG-ES promoter **(Fig 9D)** described before in (López-Farfán et al., 2014). dw, downstream promoter.

4.2.5 - SUMOylation functions to recruit SNF2PH at the active VSG-ES

SUMOylated chromatin-associated proteins are enriched upstream of the promoter region in *T. brucei* (López-Farfán et al., 2014). We assessed the importance of SUMOylation on SNF2PH targeting by mutating lysine residue 2 to alanine (K2A). K2 was chosen as it was predicted by in silico analysis (SUMO V2.0 Webserver (http://sumosp.biocuckoo.org/) and is contained within the N-terminal region modified using the *in bacteria* SUMOylation system **(Fig 8)**. SNF2PH_K2A was expressed with a HA tag from the endogenous locus by recombination. We were not able to obtain second allele replacement, suggesting that HA-SNF2PH_K2A expression ablates cell viability. Next, we performed a chromatin IP using ChIP grade rabbit antiserum against HA epitope after 48h expression of HA-SNF2PH_K2A, using the parental cell line as a mock control **(Fig 10A)**. The wild-type expression of HA-tagged SNF2PH showed a similar occupancy to the SNF2PH, enriched upstream of the active VSG-ES promoter. However, the chromatin location of the HA-SNF2PH_K2A mutant was reduced along the active VSG221 up to 4.6-fold compared with the HA-SNF2PH WT control.

In order to determine whether K2A mutation also affects to SNF2PH occupancy at the promoter region, we constructed a ChIP-seq index library including all the BES sequences from PCR regions ESPM1 and 4 **(Fig 9D)** and map the reads PCR amplified from DNA immunoprecipitated using antiHA antibodies and chromatin from cell lines expressing HA-SNF2PH and HA-SNF2PH K2A **(Fig 10B)**. As controls we used gene Control 1, a known promoter in chromosome 7, SSR7 (Strand Switch Region, SSR) and an RNA pol III (U2) were included after input normalization. All these controls constitute defined consensus regions described for *T. brucei*. Sequence alignments using Bowtie1 and 0 mismatch error yielded a number of reads aligned in BES ESPM1/4 libraries at the active promoter in BES1 and BES2 (highly sequence homologue) that were reduced for the cell line expressing the mutant K2A, to approximately ~0.5–fold. Conversely, at inactive promoters of BES the number of reads was increased (BES12) or not significantly changed (BES7 and BES13). This result is consistent with mutation K2A reduced protein SUMOylation, which decreases the occupancy of SNF2PH in the active promoter chromatin **(Fig 10B)**.

Additionally, we specifically examined the number of reads along the selected telomeric BES from the HA-SNF2PH generated library corresponding to the ESPM-4 amplified products yielding to ~12-fold enrichment in VSG221 sequences compared with input but was below ~5 fold for other telomeric BES **(Fig 10C)**.

We also assessed the effect of the K2A mutation in SUMO conjugation by PLA detection using anti-SUMO mAb and the rabbit anti-HA antiserum. The signal for 3HA-SNF2PH (69.20%±4.77) was considerably greater than for the HA-K2A-SNF2PH mutant (26.49%±15.02, n=2), suggesting that mutation in lysine residue 2 is required for SNF2PH SUMOylation *in situ* **(Fig 10D)**. We conclude that SNF2PH is mainly located at the active VSG-ES and that this position requires SUMO modification.
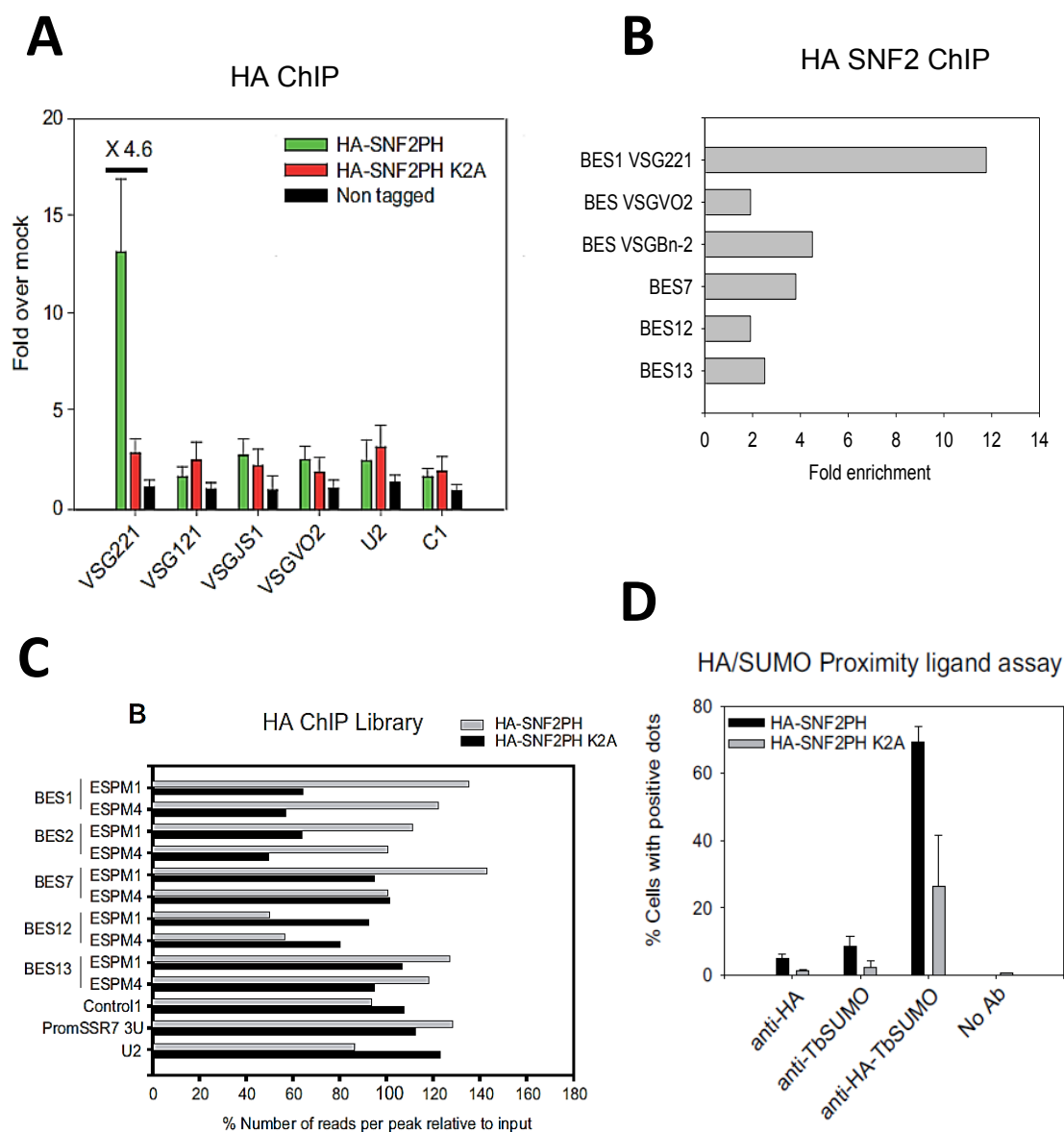
Figure 10. Expression of a SUMO-deficient mutant reduces SNF2PH occupancy at the active VSG-ES. **(A)** ChIP analysis of SNF2PH using rabbit anti-HA tag antibody detects differences between WT and K2A mutant SNF2PH upstream of the active ES. Data from three independent clones are represented as fold enrichment. A non-tagged cell line (Single Marker) was included as a negative HA control. **(B)** Percent reads per peak relative to input from the HA-SNF2PH and HA-SNF2PH K2A generated ChIP libraries. Control 1; C1, PromSSR7 3U; Switch Strand Region promoter, Chr.7. **(C)** Number of reads corresponding to HA-SNF2PH immunoprecipitated DNA at different telomeric BES from ESPM-4 amplified products, expressed as fold enrichment over input. **(D)** Percent cells with positive amplification signal in PLA assay is reduced in situ after mutation of K2. Percentage of positive cells: 69.2±4.7% (SNF2PH), 26.4±15.0% (HA-SNF2PH-K2A) (n=2). Error bars represent means ± SEM.

4.2.6- SNF2PH is a transcriptional activator that regulates VSG expression

To investigate SNF2PH function, we knocked down SNF2PH and analyzed the impact on cells using the dual-reporter cell line (DRALI). First, Western blot analysis demonstrated depletion of the protein after 24h RNAi induction **(Fig 5B, Appendix II Figure S5A)**. Importantly, significant reduction to proliferation was observed after 72h of protein depletion, suggesting SNF2PH is essential for normal fitness **(Appendix II S5B Fig)**. RT-qPCR analysis after 48h of SNF2PH RNAi induction showed a significantly reduced level of the RLuc and VSG221 mRNAs (p<0.05), without a significant effect on RNA pol II- or pol III-transcribed loci C1 and U2, respectively. No reduction was detected in the mature or pre-spliced rDNA+ 780 **(Fig 11A)**, suggesting a SNF2PH specificity in VSG expression. Next, we analyzed TbRPAI occupancy in the VSG-ES chromatin in cells depleted of SNF2PH to address whether SNF2PH is required for active VSG-ES transcription. ChIP-qPCR analysis using anti-RPAI antiserum allowed comparing the occupancy of RNA pol I in chromatin from SNF2PH depleted cells with the parental cell line **(Fig 11B)**. As previously described, TbRPAI appears to be enriched along the active VSG-ES. Upon SNF2PH depletion, we detected lower levels of TbRPAI recruitment to the active VSG-ES compared with the parental cell line. TbRPAI occupancy decreased significantly in the telomeric active VSG221 (p<0.05). This reduction was also detected in the RLuc gene located downstream of the active promoter (2.94- fold). A lower decrease was observed in the rDNA promoter and ribosomal 18S locus however, this reduction in these loci was not significant. These results are consistent with to the lack of reduction in 18s and rDNA transcripts, as shown in **Figure 11A**., and suggests that SNF2PH is specifically involved in the recruitment of TbRPAI to the active VSG-ES chromatin.

While a reduction of the reporter at the active VSG-ES promoter was detected upon RNAi in three independent clones, we found that Fluc activity from inactive VSG-ES promoters was clone dependent **(Appendix II Fig S5C)**. Relative expression of FLuc transcripts correlates with the FLuc expression level **(Appendix II Fig S5D)**. We performed RNA-seq analysis on paired groups of individual clones **(Appendix I Table S5, Appendix II Fig S5E)**, which showed variability in the VSG that is derepressed. Some, like VSG427-15 (BES10) and VSG-14 (BES8), but not all were upregulated, while the active BES1 telomeric VSG221 gene was consistently downregulated, suggesting that where SNF2PH was depleted, random derepression of inactive VSG-ES promoters occurred. Thus, SNF2PH depletion induced derepression of a cluster of silent BES but not all, similar to recently reported for H3V-H4V KO (Müller et al., 2018). As the PH domain is known to bind Histone 3 tails, the H3V Knock Out (KO) (Müller et al., 2018) and SNF2PH depletion phenotypes may be related.

Next, we analyze whether VSG protein levels were also reduced in a SNF2PH depleted cell line by quantitative western blot analysis of total protein extracts. Analysis of VSG221 protein levels in three independent SNF2PH RNAi clones **(Fig 11C)**, showed a significant downregulation of VSG221 expression compare to the parental cell line, using anti-tubulin as a loading control ($p < 0.05$) **(Fig 11D)**. A decrease of VSG221-expressing cell surface was also detected after SNF2PH RNAi by fluorescence-activated cell sorting, (FACS) **(Fig 11E)**. These results indicate that SNF2PH functions as positive transcription factor for VSG expression.

Finally, to identify factors associated with SNF2PH mediating VSG expression we ectopically expressed triple-HA tagged SNF2PH in procyclic forms followed by affinity purification to identify protein interactions by LC-MS/MS **(Appendix I S2 Table)**. Among the identified proteins we found mRNA splicing factor TbPRP9 and nucleosome assembly protein, AGC kinase 1 (AEK1), a kinase essential for the bloodstream form stage (Jensen et al., 2016) and a TP-dependent RNA helicase SUB2, (Tb927.10.540). Importantly, several previously identified VSG transcription factors, including Spt16 included in FACT complex (Facilitates Chromosome Transcription) (Nguyen et al., 2012), the proliferative cell nuclear antigen (PCNA) and a subunit from the Class I transcription factor A (Nguyen et al., 2013) were also identified. Furthermore, RNA pol I subunit RPA135 and the RNA pol II RPB1 were also found to co-purify with SNF2PH suggesting possibly a transient interaction with RNA polymerases subunits, as previously described for yeast SNF2. Together these data suggest that SNF2PH occupies a central position in VSG transcription regulation and likely interactions with RNA pol II transcription.
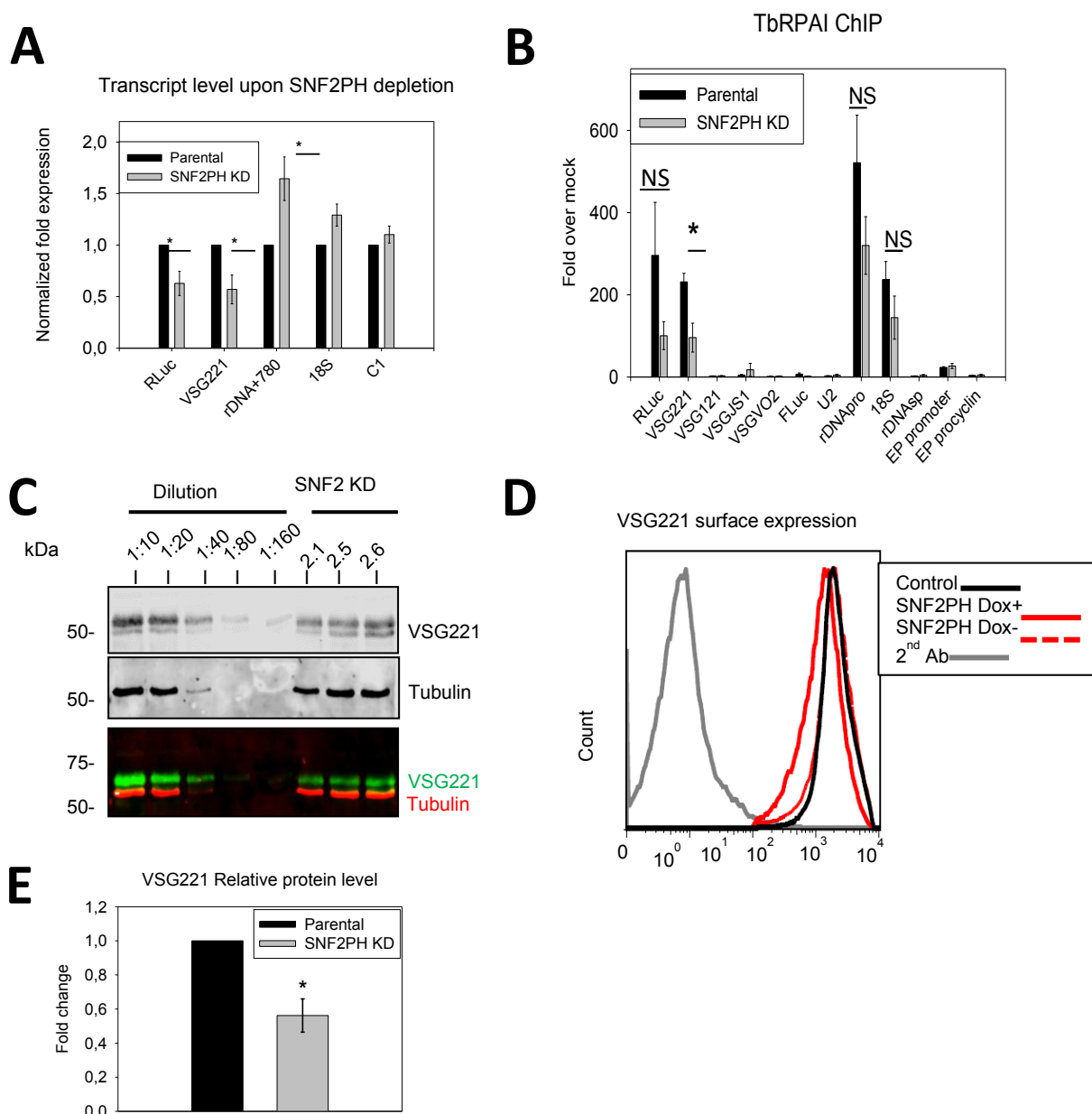
Figure 11. SNF2PH depletion results in reduction of the active VSG expression. **(A)** Reduced VSG-ES transcripts upon SNF2PH 48h RNAi. Quantitative RT-qPCR analysis indicates 43% reduction of VSG221 mRNA, validated by RLuc reporter (*p<0.05). No significant changes in ribosomal RNA transcripts were detected (18s and rDNA+780). **(B)** Reduced RNA pol I occupancy in the active VSG-ES upon SNF2PH depletion. TbRPAI analysis was carried out in 48h RNAi induced cell lines and the parental cell line (DRALI). Statistical analysis shows a significant reduction of TbRPAI occupancy levels between parental and SNF2PH depleted cells at the active VSG-ES (*p<0.05). Data from three independent clones with standard error (SE) are represented as fold over non-specific antiserum. NS; Nonsignificant **(C)** Quantitative Western blots of VSG221 expression in three independent SNF2PH RNAi clones using IR fluorescence. Anti-VSG221 and tubulin antibodies were incubated with the same blot and developed using goat anti-rabbit IgG 800 and anti-mouse IgG 700 Dylight (Thermo-Fisher). A standard curve based on tubulin normalized anti-VSG221 signal intensity was generated using different concentrations of parental cell extracts (R^2=0.99). **(D)** Quantitation of VSG221 expression relative to the parental cell line. Relative VSG221 protein levels appears to be reduced compared with the parental cell line (*p<0.05). **(E)** FACS analysis of VSG221 expressing cells shows a decreased active VSG221 population upon SNF2PH depletion (continuous red line) performed on DRALI cell line. Control: DRALI cell line population expressing VSG221. SNF2PH-dox+: induced SNF2PH depleted cells. SNF2PH-dox-: uninduced SNF2PH depleted cells. Secondary Ab: population incubated with secondary antibodies as a negative control. All data are reported as the mean ± SEM (n=3). *p<0.05 using two-tailed t-test for paired observations.

4.2.7- SNF2PH is required for the maintenance of the bloodstream stage expression profile

The previous results suggest that SNF2PH regulates VSG expression however also interacts with a RNA pol II subunit and RNA binding proteins proposing SNF2PH may regulates the expression of additional genes. In addition, SNF2PH protein levels in the bloodstream form are decreased in the insect form suggesting a possible function in development. To investigate that possibility, we knocked down SNF2PH in bloodstream trypanosomes. Transcript levels for EP procyclin ($p<0.01$), PAD1 and PAD2 were significantly upregulated ($p<0.05$) whereas MyoB levels decreased ($p<0.01$) **(Fig 12A)**. Relative mRNA levels of ribosomal 18S and C1 were unchanged. An increase in EP procyclin and PAD1/2 transcription followed by a reduction in MyoB, similar to the transcript profile of the stumpy form, a transition to the insect stage (Kabani et al., 2009);(Barquilla et al., 2012). Furthermore, we investigated how SNF2PH influences EP procyclin replacement dynamics during *in vitro* differentiation from the BF to the PF. We triggered differentiation with 3 mM cis-aconitate **(Fig 12B)**. Under these conditions, SNF2PH-depleted cells expressed higher procyclin transcripts at 4h of induction, whereas VSG221 mRNA levels were reduced compared to the parental cell line. Surface protein analyses of EP procyclin and VSG221 in SNF2PH depleted cells by IF, found a ~2-fold faster replacement of VSG by EP Procyclin **(Appendix II S6A&B Figs)**.
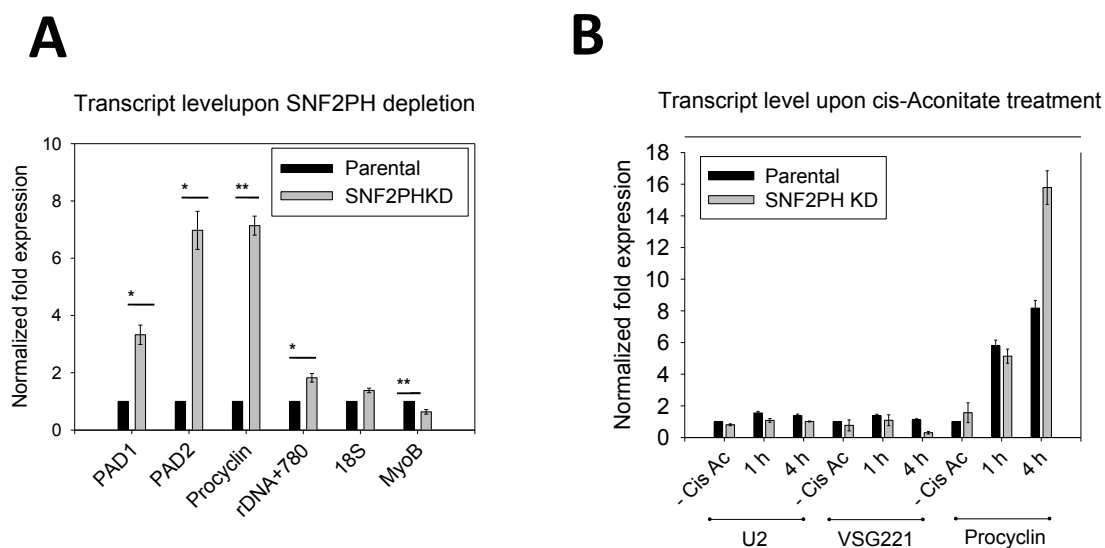
Figure 12. SNF2PH ensures VSG expression opposing differentiation. (A) Depletion of SNF2PH (48h) results in an increase of Procyclin and Proteins Associated with Differentiation (PADs) transcripts. Procyclin, PAD1 and PAD2 mRNAs are upregulated when SNF2PH is depleted. Results are the average from three independent clones and data normalized with U2 mRNA. Error bars represent means ± SEM. (*$p<0.05$, **$p<0.01$) using two-tailed Student's t-test for paired observations. **(B)** SNF2PH KD cells differentiate to procyclics more efficiently. The procyclin transcript is increased in SNF2PH KD induced cells during in vitro differentiation compared with parental cell line. Parental and SNF2PH depleted cells were treated with 3mM cis-aconitate and temperature shift for 4 hours. Quantitative PCR data from two independent clones were normalized against C1 (RNA pol II-transcribed) as a housekeeping gene. Error bars represent means ± standard deviation (SD).

To investigate additional genes regulated by SNF2PH we carried out RNA-seq upon knockdown in the bloodstream form. Eighteen representative differentially expressed genes (DEG) with a false discovery rate (FDR) <0.05 were found out of 8673 genes from two experimental replicates and 221 transcripts with a p<0.05 **(Fig 12 C, Appendix I S3A Table)**. Among them, clear upregulation of genes related to the procyclic form, including procyclin genes (EP1-2, GPEET2), followed by procyclin-associated genes (PAG1, 4 and 5) and expression site-associated genes (ESAG2 and 8), some of them consistent with previous analysis in procylic forms (Queiroz et al., 2009). We also found upregulation of protein associated with differentiation 1, 2 (PAD1 and PAD2), zing finger family ZC3H18 and the receptor-type adenylate cyclase GRESAG4-related transcripts. Conversely, genes with normally higher expression in bloodstream form were downregulated, including the alternative oxidase and glucose transporter THT-2 **(Appendix I S3B Table)**.

Given that SNF2PH expression is reduced in the procyclic stage **(Fig 7A)**, we sought to investigate the gain of function phenotype induced by SNF2PH ectopic expression in the procyclic form. Interestingly, we detected an increase in inactive telomeric VSG121, VSGBn2 and VSGJS1 transcripts from two independent clones in the procyclic form where no VSG is normally expressed **(Fig 12D)**.

To determine possible changes to global gene expression induced by SNF2PH ectopic expression, we used RNA-seq and compared the overexpressor and parental procyclic cell line **(Fig 12E)**. We found increased expression of ESAG3 from several BES, ranging from 2.5 to 5.2-log FC, while the promoter adjacent transferrin-binding protein ESAG6/7 increased 1.7 to 3.2-log FC **(Fig 12E and Appendix I, Table S6A)**. Of the total mapped reads, 55.08% were linked to telomeric bloodstream ESs (BESs), including BES1 (BES40 ~5.2- log FC), BES4 (BES28/98 ~4.5-log FC), BES2 (BES129/126 ~3.8- log FC) and BES13 (BES56/153/51/4 ~2.6- log FC. Importantly, no expression of VSG basic copies genes, located within internal chromosomal arrays, was detected, ruling out global chromatin deregulation. These results show that, in procyclic forms where no VSG is expressed, ectopic expression of SNF2PH induced upregulation of telomeric VSG-ES (BES) transcripts, suggesting that SNF2PH functions as a central regulator of the BES. Thus, in BF, where SNF2PH is highly expressed, this chromatin remodeler likely functions to promote and maintain the expression of VSG-ES (BES) **(Fig 12E and Appendix I, Table S6A)**.

Next, we analyzed the relevance of the SNF2PH Plant Homeodomain and overexpressed a truncated form of SNF2PH lacking the PH domain (SNF2ΔPH) **(Fig 12F)**. RNA-seq analysis detected 737 genes (FDR<0.05 and p<0.01, out of 7918 genes) differentially expressed after

ectopic expression of SNF2ΔPH versus 118 genes (FDR<0.05 and p<0.001, out of 8451 genes) for SNF2PH full-length, illustrated in **Appendix II S6C & D Figs**. Most significantly, induction of BES-related genes did not occur in cells overexpressing the SNF2ΔPH mutant lacking the PH domain **(Fig 12F)**, suggesting that the PH domain is required to provide SNF2PH specificity to bind and recognizes VSG-ES (BES) chromatin **(Fig 12F and Appendix I, Table S6B)**.

Significantly, beyond impacts to BES expression, invariant surface glycoprotein ISG65 genes, transcribed exclusively in the bloodstream form, were also significantly increased up to 6-fold in procyclics ectopically expressing SNF2PH compared to the parental procyclic cell line **(Appendix I, Table S6A)**. Importantly, ISG gene transcripts were not detected in cell lines expressing SNF2ΔPH **(Appendix I, Table S6B)**. As VSG is transcribed by the RNA pol I and ISG65 by RNA pol II, this suggests that SNF2PH acts beyond the telomeric BES, and may act as a global transcriptional activator of surface protein genes irrespective of the RNA polymerase involved.

Altogether, these results suggest the PH domain is essential to direct SNF2PH transcriptional activation function to specific bloodstream form surface proteins genes. In metazoan, the transcription factors containing homeodomain determine cell fates during development, and this is the chromatin-binding domain necessary to regulate downstream target genes. This result provides an important contribution of the plant homeodomain in recognition of epigenetically labeled chromatin to regulate gene expression in trypanosomes, an early-branching eukaryote.

C

SNF2PH KD

D

SNF2PH ectopic expression in PF

E

SNF2PH FL (Full-length)

F

SNF2ΔPH
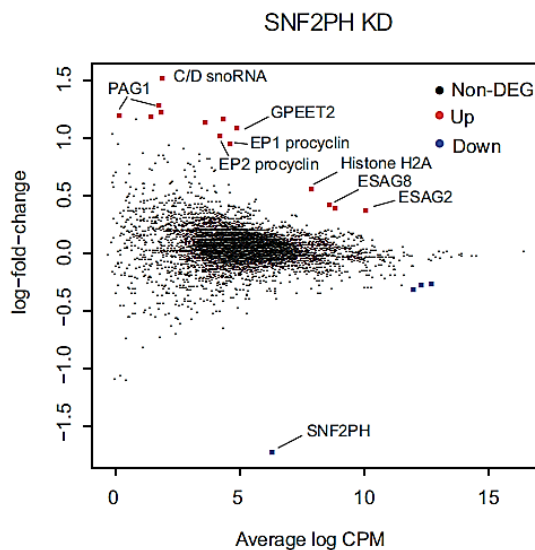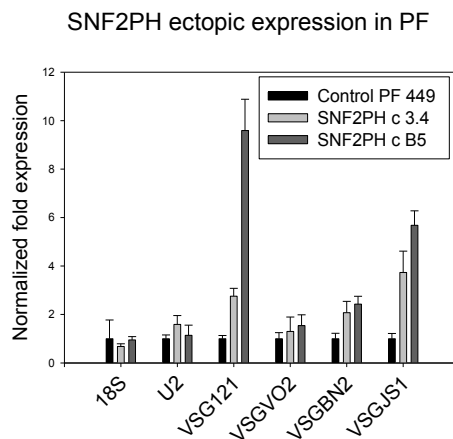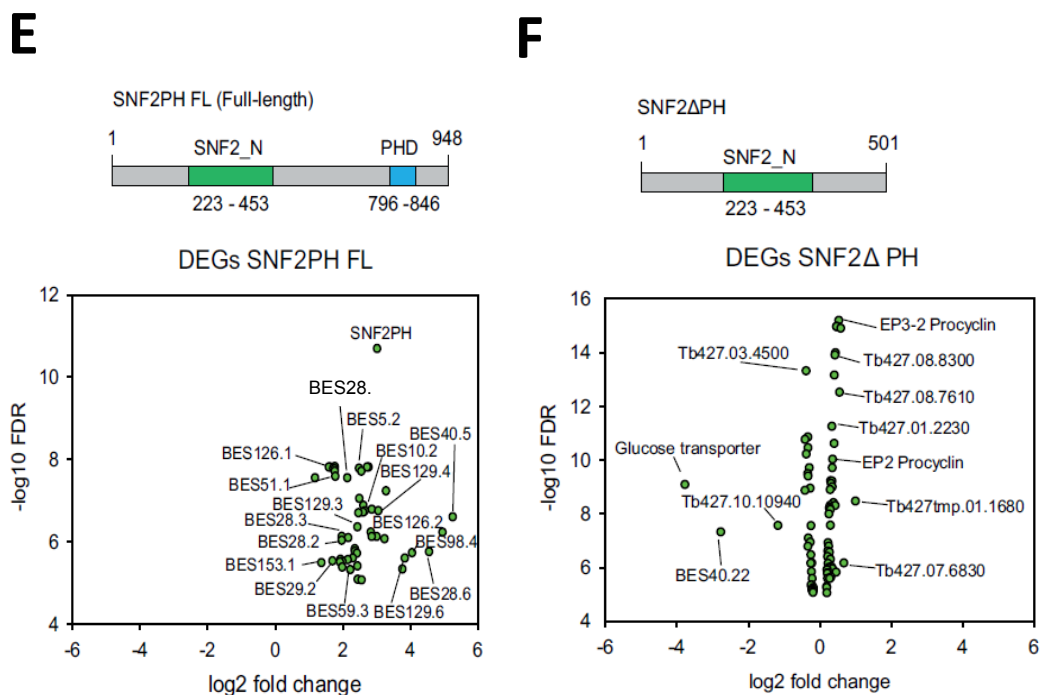
Figure 12. (C) Scatter plot for differentially expressed genes upon SNF2PH depletion (FDR<0.05). Non-DEG: Non-differentially expressed genes. Up: upregulated genes. Down: downregulated genes. **(D)** Ectopic expression of SNF2PH in procyclic form upregulates telomeric VSG mRNAs. Histogram showing the relative expression of mRNA measured by qRT-PCR of VSG genes located in different telomeric VSG-ESs (BES) genes after 48h of induction of SNF2PH ectopic expression in procyclic form. This analysis included mRNAs from telomeric VSG genes (BF stage-specific) not expressed normally in the insect procyclic form, including the VSG121 (BES3), VSGJS1 (BES13) VSGBR2 (BES15) and VSGVO2 (BES2). Data from two independent clones and parental controls are represented as normalized fold expression relative to C1 (RNA pol II transcribed) as a housekeeping gene. Error bars represent means ± SD from technical replicates for each independent clone. **(E)** Ectopic expression of SNF2PH in procyclic induces expression of bloodstream form surface proteins. Scatter plot for differentially expressed genes (DEG) in RNA-seq analysis shows upregulation of telomeric BESs. A significant increase of BES associated genes (ESAGs) linked to telomeric BESs was detected after full-length SNF2PH ectopic expression in the procyclic form. Data from at least two experimental replicates are represented as log2 fold change (FC) for genes with FDR<0.05 and p<0.001 after correction with the uninduced procyclic cell line. Relative mRNA levels of procyclic cells after 48h of SNF2PH ectopic expression increase telomeric BESs, including VSG221-ES (BES1; TAR40), VSG121-ES (BES3; TAR15), VSGJS1-ES (BES13; TAR56) and VSGVO2-ES (BES2; TAR129), VSGR2-ES ES (BES15; TAR126), see (Hertz-Fowler et al., 2008) for detailed BES and TAR nomenclature **(Appendix I, Table S6A)**. Ectopic expression of SNF2PH full-length also induced invariant surface glycoprotein 65 (ISG65) genes **(Appendix I, Table S6A)**. Data from two independent clones and parental controls are represented as normalized fold expression relative to C1 (RNA pol II transcribed) as a housekeeping gene. **(F)** Expression of bloodstream form surface proteins requires SNF2PH plant homeodomain. EP Procyclins and other insect-stage specific markers (FDR<0.05, p<0.001) are expressed in the SNF2PH mutant form lacking the Plant Homeodomain (SNF2ΔPH). No BF surface proteins like VSG-ES (BES related or ESAGs) neither invariant surface glycoproteins (ISG65) genes were induced under ectopic expression of the SNF2ΔPH mutant.

4.2.8 - SNF2PH is down-regulated in quiescent stumpy forms

The stumpy form of pleomorphic trypanosomes is pre-adapted to the metabolism required by the procyclic form for survival within the insect vector. Given upregulation of stumpy form markers in SNF2PH knockdown cells, we examined SNF2PH dynamics during the stumpy to procyclic form transition. Recent evidences suggest that the AMP-dependent kinase, AMPKα1 is a key regulator of the development of quiescence in bloodstream form trypanosomes (Saldivia et al., 2016) reviwed in (Silvester et al., 2017). Upon AMPKα1 activation, stumpy-like differentiation was induced in a monomorphic cell line. Quantitative Western blots and RT-qPCR analyses of monomorphic cells treated with an AMP analog showed a significant decreased in SNF2PH proteins levels and transcripts **(Figure 13A)**. Reduction of SNF2PH protein levels in stumpy-like cells obtained after AMPKα1 activation by AMP suggests AMPK pathway negatively regulates SNF2PH expression.

We also assessed the transcriptional profile in stumpy-like forms induced by AMP treatment and compared with SNF2PH knockdown. RT-qPCR confirmed a transcriptional profile characteristic of the stumpy form in untreated SNF2PH knockdown versus AMP treated cells, in which relative mRNA levels for SNF2PH transcripts were downregulated (p<0.01) **(Figure 13B)**. Interestingly, stumpy form-like transcriptome changes were more prominent in cells treated with 5'-AMP and depleted for SNF2PH compared with 5'-AMP alone and 5'-AMP untreated knockdown cells.

An important question, however, is whether SNF2PH downregulation occurs naturally during *in vivo* differentiation of a wild-type pleomorphic strain. Bloodstream pleomorphic trypanosomes undergo differentiation from the proliferative to the quiescent stumpy form throughout mice infection. SNF2PH protein levels analysis in the pleomorphic AnTAT 90.13 strain, decreased at 4-5 days post infection **(Figure 13B)**, whereas AMPKα1 was fully activated as previously described (Saldivia et al., 2016). Taken altogether, these data suggest that SNF2PH is negatively regulated during stumpy transition and mechanistically linked to AMPKα1 activation by an undefined mechanism.
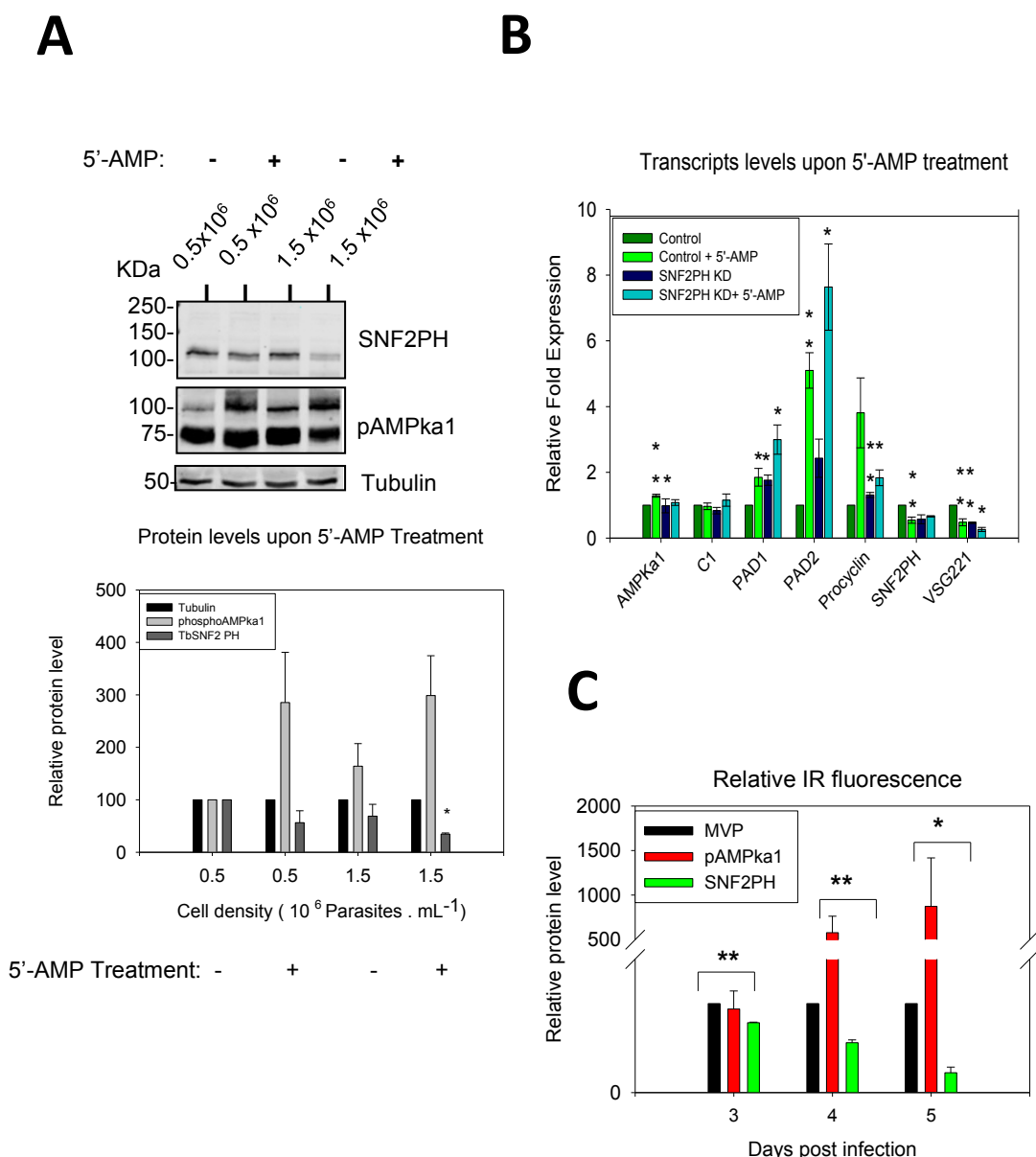
Figure 13. Downregulation of SNF2PH during stumpy form development. (A) (Upper panel) SNF2PH protein levels are reduced upon AMPK activation. Representative Western blots of cells treated with 1 μM AMP analogue for 18 hours compared with an untreated control at two different densities. (Lower panel) Histogram showing relative proteinlevels for phosphorylated AMPKα1 for the western blot shown in upper panel. (Student's T test): *p<0.001. **(B)** AMP analogue treatment positively regulates stumpy form gene expression of PAD genes in SNF2PH depleted cells. Relative transcript levels after treating control and SNF2PH depleted cells (24h) with or without 1μM AMP for 18 hours. Genes associated with stumpy form differentiation are represented as PAD1 and 2, procyclin and VSG221 and normalized with U2 as a housekeeping gene from three independent replicates. (Student's T-test): *p<0.05, **p<0.01. **(C)** SNF2PH is downregulated during in vivo stumpy form transition. Histogram of relative protein levels for SNF2PH and phosphorylated AMPKα1 from quantitative western blot analyses. Data from in vivo infections represent the mean of at least 2 mice with standard errors of the mean (SEM). Loading information: Dpi: days post infection. Ms: Mouse. (Student's T-test): *p<0.01, **p<0.001.

5. Discussion

5. Discussion

Antigenic variation of the variant surface glycoprotein gene (VSG) in *Trypanosoma brucei* undergoes a complex transcription regulation process leading to monoallelic expression of a single VSG at any given time. In the bloodstream form, only one of about 15 telomeric VSG-ESs is transcribed resulting in the expression on the surface of a dense coat of a single VSG. The active transcriptional state is maintaining throughout many generations without changes in the promoter sequences, suggesting trypanosome uses epigenetic mechanisms as post-translational modification (PTM) of proteins to mark the active VSG-ES chromatin. Several PTMs have been associated with VSG regulation, although most of it functioned to maintain the silencing state of inactive VSG-ESs, reviewed in (Figueiredo et al., 2009). Importantly, protein SUMOylation is one the few PTM described that functions as a positive epigenetic regulation, and is mediated by the SUMO E3 ligase TbSIZ1/PIAS1, required for SUMOylation along the active VSG-ES chromatin (López-Farfán et al., 2014). SUMO constitutes a remarkable group of PTMs as protein targets are modified by a large 12kDa polypeptide that significantly alters protein structure. SUMO is involved in regulation of complex biological processes such as multiple proteins in the same nuclear structure or gene cluster transcription, or responds to external stress or internal stimuli to regulate signalling pathways, and monitoring other PTM modifications as protein phosphorylation (Zhao, 2018). Recently, it became evident, for example, that crosstalking between phosphorylation and SUMOylation impacts transcription activity as well as protein-protein interactions (Rosonina et al., 2017). It seems now clear that protein covalently targeted by SUMO provides a great modification of the protein interaction interface in such a way that either increases the recruitment of PTM on enzymatic complexes to modify chromatin structure at a given promoter region, or by regulating the enzymatic activity of chromatin modifying enzymes, altering chromatin structure, leading to repression or activation of gene expression.

In this work, we decided to uncover novel proteins associated with the nuclear body ESB and likely related to antigenic variation by performing MS/MS proteomic analysis of SUMOylated proteins isolated from infective bloodstream form (BF) extracts. Different proteomic approaches allowed us to identify SUMO-targets developmentally regulated in the BF, included in the HSF within the ESB and associated with the active VSG-ES (López-Farfán et al., 2014). Many of them involved in several cellular processes in the bloodstream form of the parasite as described previously in the insect procyclic form (Iribarren et al., 2015a). The identification of SUMO conjugates is a challenging process considering the usual low proportion of the

modified forms of a given protein and the susceptibility of deconjugation by specific isopeptidases. In this work, we first attempted to perform a large-scale identification of SUMO targets, using protein extracts from the bloodstream form of the parasite, by affinity purification of His/HA-SUMO conjugated proteins but did not achieve the high standard required. Thus, independent purifications and MS/MS analyses identified proteins listed in **Table S1A**, **B**, **C**, which showed low reproducibility and contained contaminant proteins, probably due to the lack of a specific enrichment step. Nevertheless, this approach allowed us to identify several targets that should be considered as genuine substrates, since they were previously described as SUMO conjugated in other organisms such as yeast and mammals. Therefore, many of these proteins ought to be considered the function of SUMO modification in each target for further investigation. In fact, some of the identified proteins constitute well-known protein targets of SUMO in other trypanosomes and several eukaryotic organisms including: DEAD/H helicases, KU70, GTP-binding, eIF4E, SMC family members, histones, DNA repair proteins, among others (Iribarren et al., 2015a);(Hendriks and Vertegaal, 2016) .

In addition to SUMO-conjugated identification of proteins, we unequivocally identify the targets and most importantly the SUMOylation site/s in order to fully advance in future functional analyses. Thus, we performed site-specific proteomics by generating His-HA-TbSUMOT^{106K} expressing parasites followed by specific enrichment of SUMOylated peptides with anti-GG antibodies, as previously described (Iribarren et al., 2015a);(Tammsalu et al., 2015). Following this approach, we were able to unambiguously identify a subset of 37 confident SUMOylated proteins and 45 modified Lys residues.

This data provides the first list of SUMOylated proteins in the infective stage of the parasite. We used a similar approach to identify SUMO-conjugated proteins, one of which was previously described in the insect stage, named Procycliclic Form (Iribarren et al., 2015a). Thus, it is feasible to compare the protein profiles in the two developmental stages in order to identify proteins that utilize SUMO at least as one as the PTMs, which may be involved in the complex parasite development (Rico et al., 2017).

Eukaryotic organisms present a consensus SUMOylation site (ψ-K-X-E; ψ: aliphatic aminoacid; X: any aminoacid) that seems to be conserved in a considerable number of targets in the protozoa parasite *T. brucei*. Interestingly, in the PF developmental stage, the aliphatic residue in position -1 is preferably V, whereas in the infective BF, both V and I appeared in similar frequency. This data shows a higher flexibility in the consensus SUMOylation site at this position in the bloodstream form of the parasite.

Most of the proteins modified by SUMO, identified in the infective form, are involved in nuclear processes. These proteins include histones, PCNA, DNA topoisomerase 1B, RBP12 and SMC4, which strengthens SUMO relevance as a PTM that regulates fundamental processes. Many of the substrates identified here in the BF **(Table 2)**, were also found to be SUMOylated in the PF stage (Iribarren et al., 2015a), suggesting a conserved function by this PTM, regarding processes such as chromatin remodeling and DNA/RNA metabolism **(Table 3)**. It is noteworthy that most of these conserved SUMO targets are SUMOylated on particular lysines, highlighting the importance of specific residues on their regulation. Differential SUMOylation sites in *Drosophila melanogaster* Ultraspirace (Usp), a protein involved in development, suggests Usp modulates its activity by changing molecular interactions (Bielska et al., 2012).

Interestingly, histone H2A, H3V, DNA topoisomerase IB, PCNA, SMC4, and the hypothetical protein Tb927.9.13320 constitute SUMO substrates that presented distinct developmentally regulated modifications in the number and/or position of the acceptor lysines between BF and PF **(Table 3)**. Additional lysine acceptor sites observed in the infective form of the parasite suggest a stage-specific SUMO-mediated modification which most likely acquires new functions linked to pathogenicity. This is in agreement with SUMO being essential for maintaining cell homeostasis when the parasite encounters environmental stress such as: osmotic stress, hypoxia, heat shock, and nutrient stress (Enserink, 2015), most of which occurs when parasites infect a mammalian host through the bite of the insect vector. Thus, trypanosomes fine tune nuclear targets, described above, probably remodel gene expression and activate and/or repress chromatin, similarly to other eukaryotes (Niskanen and Palvimo, 2017). In mice development, SUMO-specific protease 2 (SENP2) regulates epigenetic gene expression by modification of the PTM status of Polycomb Repressive Complex 1 (PRC1), changing promoter affinity. Depletion of SENP2 down-regulates expression of Polycomb regulated genes. It seems that SUMOylation increases binding affinity of Pc2/CBX4 (PRC1 subunit) to H3K27me3. SENP2 specifically controls SUMOylation status of Pc2/CBX4, which facilitates PRC1 binding to H3K27me3 leading to transcriptional repression (Kang et al., 2010).

Most importantly, latest reviews remark on the importance of SUMO as a key molecular PTM in the regulation of transcription factors and chromatin-associated proteins involved in differentiation and developmental processes (Deyrieux and Wilson, 2017);(Monribot-Villanueva et al., 2016).

Suppression of chromosomal rearrangements is controlled by SUMOylation in yeast, in particular by Mms21, a SUMO ligase that maintains genome integrity and one of its mayor substrates are the SMC-family proteins (Albuquerque et al., 2013). Unfortunately, it is not currently known how SUMO functions in SMC proteins. We found SMC4 a subunit of the

condensin complex (Structural Maintenance of Chromosome family), as a reproducible SUMO target that interestingly, contains a stage-specific SUMOylation site in the infective developmental form (K623) **(Table 3)**. This result raises the possibility that distinct SUMOylation sites of a subunit of the same protein complex may lead to different biological outcomes in each developmental stage.

Additionally, the crosstalk between SUMO and other PTMs is a very interesting feature to be considered as molecular mechanism to regulate protein targets. Histone 2A, for example, appeared to be highly acetylated in both developmental stages of *T. brucei* (Moretti et al., 2018) on some of the lysine residues identified as SUMOylated by site-specific proteomics. Overexpression and mislocalization of histone H3 variant (CENP-A in humans and Cse4 in yeast) leads to aneuploidy in yeast and is detected in human cancers. In yeast, the centromeric histone variant Cse4, is regulated by SUMOylation mainly by Siz1 and Siz2 E3 ligases at the lysine 65 (K65) and the expression of cse4 K65R mutant leads to increased stability and mislocalization of cse4 K65R. All of which suggest that SUMO regulates protein stability and genomic localization (Ohkuni et al., 2018).

It has been reported that *T. brucei* bromodomain factor 2 (TbBDF2) is able to recognize hyperacetylated N-T of H2AZ, essential for cell growth (Yang et al., 2017), but the role of SUMO on histone regulation has not been elucidated yet. Acetylation-SUMOylation crosstalk is also observed in reader protein TRIM24 and histone H3 epigenetic modulation (Appikonda et al., 2018). Considering these facts, we cannot rule out a fine-tuned epigenetic regulation network of histones that might be differentially expressed in *T. brucei* life stages.

Interestingly, polyubiquitin was also identified as SUMO substrate in BF. Although the crosstalk between these two related modifiers has been described in other organisms (Lamoliatte et al., 2017), this is the first report on SUMOylation of ubiquitin (Ub) in *T. brucei*, revealing novel connections of these two PTMs in this organism. SUMO and Ub can interact in different ways; by competing for the same Lys residue on a substrate, regulating the enzymes of the pathways required for the modification or participating on polyubiquitination of different substrates mainly for proteasomal degradation, among others (Hunter and Sun, 2008). Hybrid chains have been described in humans, comprising of ubiquitination of SUMO and SUMOylation of Ub (Hendriks and Vertegaal, 2016). In this way, the finding of SUMOylated ubiquitin and SUMO itself in our site-specific proteomic study confirms the existence of SUMO-SUMO and SUMO-Ub chains in BF parasites, suggesting a complex crosstalk between these pathways.

From the differentially SUMOylated substrates identified in BF, we highlighted TFIIS2-2, a non-characterized trypanosome homolog of transcription elongation factor (TFIIS). To date, two

TFIIS homologues have been identified (TbTFIIS1 and TbTFIIS1-2), with a cooperating activity to promote transcriptional regulation necessary for mRNA transcription (Uzureau et al., 2008). Recent work demonstrated TFIIS2-2 interaction with trimethylated histones H4K17me3 and H3K32me3 with its PWWP domain (Wang et al., 2019), reinforcing its function as a histone PTM reader involved in transcriptional regulation. Another interesting finding is the SUMOylation of the RNA polymerase associated subunit CTR9, an essential factor that regulates several genes involved in the control of gene expression (Ouna et al., 2012). This protein is one of the core components of PAF complex, implicated in the transition from transcriptional initiation to elongation (Yoo et al., 2013). Further validation of *in vitro* SUMOylation for TFIIS2-2 and CTR9 indicates that the activities of these proteins might be influenced by this PTM, and future mutational analysis will help to uncover its biological significance.

One of the most remarkable aspects of this work is the observation that five different nuclear proteins (as judged by IFI, See **Figure 4**) can be recruited to a single intense focus on the basis of their modification or association with SUMO (as demonstrated by PLA analysis, See **Figure 5A**). We believe that, similar to the role described for SUMO in the formation of PML nuclear bodies (Matunis et al., 2006), *Tb*SUMO could have a role as scaffold or "glue" protein mediating interactions between SUMO moieties of SUMOylated proteins and SUMO binding motifs on SUMO interacting proteins. Also noteworthy is the association link between SUMO PTM and VSG expression, determined by functional depletion of the previously validated substrates. As showed in **Figure 5B**, TFIIS2-2 depletion led to a deregulation of VSG-ES, suggesting the importance of SUMO modification of this particular homolog as a requirement for its regulatory role associated with VSG expression. These experimental evidences describe for the first time an association of HSF related proteins and VSG regulation, remarking the importance of this PTM in both antigenic expression and DNA accessibility.

The enrichment of SUMOylated nuclear factors in the HSF is a premiere for its functionally related role in the ESB. Regarding to this, it is noteworthy to remark the variable number of spots observed per single PLA positive cell expressing a selected TbSUMO partner, which not only approves a differential protein dynamics along cell cycle but also its probable association to the HSF in a given time or in multiple foci, concerning other regulatory functions. The prevalence of a single spot in pre-mitotic cells would suggest a greater fraction of the SUMOylated partner residing the ESB, whereas a multiple localized pattern would conceive SUMO modification in the relocation of subnuclear compartments as well as stabilizing protein interactions (Sharrocks, 2006). Altogether, these results provide important new insights into the role of SUMO in mediating the assembly of subnuclear structures, such as the HSF, and in

regulating the distribution of proteins between the HSF and the nucleoplasm, while at the same time reinforcing its function as an epigenetic modification that regulates VSG expression.

In conclusion, the use of the powerful tool of specific enrichment of SUMO modified peptides using the diGly-Lys specific antibody provided a first described list of SUMOylated proteins with their acceptor site, firstly described in BF. Even though they constitute reliable targets, the number of identified lysine acceptor sites is limited due to the experimental constrains, as previously commented. Then, we also considered the former list obtained from 8xHis and HA-tagged SUMO version together with the lysine modified candidate list in order to increase the range of SUMO modified targets that appear to be masked by the astringency of the site-specific identification method. Out of the selected proteins with remarkable reproducibility in two independent assays **(Table S1 a, b, c),** we decided to functionally characterise SNF2PH (previously annotated as Transcription Activator).

It is widely known how SUMO post translational modification of proteins modulate transcription by activation or repression of an extensive range of transcription factors to adjust and fine-tune regulation of gene expression (Girdwood et al., 2004; Lyst and Stancheva, 2007). Several chromatin remodelers and histones PTM epigenetic factors have been widely investigated to address a possible function in VSG transcription regulation and thus, in antigenic variation of the parasite. However, lack of function analysis of chromatin remodelers and TFs analysed so far, consistently showed a similar phenotype, inactive/silent VSG-ES promoter become derepressed. Related to the SNF2 subfamily, TbISWI knockdown results in increase transcription of silent VSG ES promoters, leading to derepression of the region just downstream of *silent* promoters (Hughes et al., 2007; Stanne et al., 2011). Besides, NLP knockdown results in 50-fold derepression of the silent VSG-ES promoters. However, NLP depletion also induced expression of VSG Basic Copy gene arrays and minichromosome VSGs, which do not contain any upstream promoter associated; suggesting NLP is a global chromatin assembly factor (Narayanan et al., 2011). Other TF associated with VSG expression is CITFA that binds to all RNA pol I promoters, and is essential for RNA pol I transcription in the parasite (Nguyen et al., 2012). Further, depletion of TDP1 results in up to 40-90% reduction in VSG and rRNA transcripts, suggesting is a basal TF of RNA pol I transcription (Narayanan and Rudenko, 2013). In addition, the chromatin remodeler FACT function in the maintenance of repressed chromatin stated at silent VSG ES promoters (Denninger and Rudenko, 2014). SWI2/SNF2-like protein (JBP2) regulates the de novo site-specific localization of J, a modified thymine base synthesis in bloodstream form (Kieft et al., 2007). *T. brucei* bromodomain containing chromatin remodeler factors were functionally investigated using a mammalian bromodomain

inhibitor called I-BET151. However, I-BET151 inhibition of family bromodomain members hampered function assignment of each specific factor in the group. Most importantly, inhibition by I-BET151 induced no only derepression of silent telomeric ES VSGs, but also VSG basic copies suggesting a global effect of chromatin packing (Schulz et al., 2015). In addition, the I-BET151 inhibitor probably induced lack of function of essential bromodomain chromatin assembly proteins and trypanosome cells viability induces upregulation of silent genes, including VSGs. Furthermore, it remains unknown where bromodomain proteins located in the chromatin, or what is the signalling mechanism used to induce developmental regulation state.

Many studies have focused in trypanosome chromatin remodeling (ChR) factors, notwithstanding no one of previously described ChR includes a conserved domain involved in development, as the plant homeodomain (PHD) included in SNF2PH we characterise in this work. The PHD is a domain involved in regulation of developmental gene expression in eukaryotes (Sanchez and Zhou, 2011). PHD fingers can recognize the unmodified and modified histone H3 tail. In other eukaryotes, PHD finger is contained in many central regulatory proteins such as, transcription co-activators p300 and CBP, the co-repressor TIF1, the Trithorax-group, Mi-2 complex (histone deacetylase), the Polycomb-like protein, etc. PHD-fingers functions as epigenome readers controlling gene expression through molecular recruitment of chromatin regulators and transcription factors (Musselman and Kutateladze, 2011).

In trypanosomes, the PH domain provides to SNF2PH of a homeodomain previously described as an epigenetic reader that function in developmental gene expression, concurrent with a chromatin remodeler domain that regulates chromatin accessibility and nucleosome positioning, linking two fundamental function in a single protein, no described previously in protozoan parasites. In this work, we characterize SNF2PH as a novel TF involved in the modification of chromatin structure in two manners, firstly allowing access to transcriptional activators to the highly enriched SNF2PH VSG-ES active promoter chromatin, while in silent VSG-ES promoters, a slight but significant SNF2PH occupancy showed an opposite function reducing chromatin accessibility to inactive VSG-ES promoters. This is suggested by SNF2PH depletion experiments in the dual reporter cell line **(Fig 9)**. In addition, expression of developmentally regulated genes was affected upon SNF2PH depletion by upregulating the expression of genes that normally are repressed, as procyclin and PAD genes. These results suggest two molecular mechanisms operating in the regulation of VSG-ES, as suggested elsewhere (Navarro et al., 2007). SNF2PH is highly expressed in the infective bloodstream form contrary to the insect procyclic form, what could explain the positive function in the

bloodstream form to maintain this developmental stage. Domain-structure comparative analysis of trypanosome SNF2PH identified conserved chromatin remodeling factors from yeast, SNF2H (Sucrose Nonfermenting Protein 2 Homolog) to human SMCA5 (SWI/SNF-related regulator of chromatin 5) well known chromatin remodeling factors that function as mayor regulators of developmental gene expression (He et al., 2016; Jin et al., 2015).

However, in the bloodstream form, SNF2PH function in a dual opposite manner in the active versus silent VSG-ES promoters. Most likely, a highly dynamic protein modification, as PTMs, may operate to modify SNF2PH function by prompting high chromatin accessibility in the active VSG-ES, while reducing accessibility in silent VSG-ES promoters. An increasing number of TFs and histone modifier enzymes are known to be recruited to promoters in a SUMO-dependent manner, as histone deacetylases, histone demethylase, histone methyltransferase, as chromatin remodeling factors and chromatin-associated proteins (see for review (Ouyang et al., 2009)). Thus, protein SUMOylation plays a central role in histone modifications and chromatin remodeling to regulate gene expression. Most of previous studies suggested that protein SUMOylation repress access of TFs to chromatin, however recent analyses suggest that SUMOylation on the chromatin also regulate highly expressed genes (Rosonina et al., 2017; Wotton et al., 2017). We previously described that chromatin SUMOylation is a distinct feature of the active VSG-ES locus by ChIP. SUMOylation of chromatin-associated proteins is required for the active transcriptional state of the VSG-ES and essential for efficient recruitment of RNA pol I to the VSG-ES promoter (López-Farfán et al., 2014).

In this work, we demonstrate that SNF2PH is a SUMO-conjugated protein, and ChIP sequencing associates SUMO-SNF2PH with chromatin from the active VSG-ES promoter to the telomeric VSG gene. Therefore, SUMOylated SNF2PH function as a positive TF for the monoallelicaly expressed VSG. Conversely, no-SUMOylated SNF2PH was located mainly in silent VSG promoters. Previously described SUMO function together with our data suggest SUMO-target SNF2PH functions recruiting to developmentally regulated promoters, co-regulators, histone modifier enzymes or specific transcription factors, acting as a TF factor for VSG expression and bloodstream form specific genes. The analysis of the amplified products from ESPM-4 immunoprecipitated genomic DNA from sequential ChIP assay using the antiserum against SNF2PH, yielded a large number of the active VSG promoter sequences compared with the control (prebleed after a previous enrichment in TbSUMO immunoprecipitated sequences, suggesting that only the SUMOylated SNF2PH is located upstream the active VSG-ES promoter region) **(Appendix II S4 Fig)**. Additionally, ESPM-4 amplified genomic sequences obtained from the ChIP library data using the rabbit antiserum against HA epitope in HA-SNF2PH expressing

cells defined SNF2PH main occupancy at the active VSG-ES. Furthermore, the lack of enrichment in the active VSG promoter in ChIP assays of the K2A mutant suggests that SUMOylation in lysine residue 2 altered the established occupancy of SNF2PH throughout the active VSG-ES, which may suggest TbSUMO PTM as a requirement for SNF2PH recruitment at the active promoter. This evidence is supported by a reduction of number reads linked to the active VSG221 in the HA-K2A Immunoprecipitated DNA, confirming that lysine residue 2 modification is necessary for SNF2PH occupancy upstream ES promoter sequences. PLA data if only shows a single dot, o two dots in G2, also suggest the SNF2PH-SUMOylated is at a single nuclear site, as the HSF, previously associated with the active VSG-ES chromatin. Alternatively, SUMO modification may increase the protein stability of SNF2PH in bloodstream trypanosomes, as occurs in a Arabidopsis SWI/SNF protein BRM, which is SUMO modified to allow protein stabilisation in root development (Zhang et al., 2017). In addition, the SNF2_N containing domain SMCA4 and SMCA5 proteins are also SUMOylated during mouse spermatogenesis, as cell-specific epigenetic mark involved in chromatin remodeling during meiosis (Yuxuan et al., 2016). Further, *Drosophila* BRAHMA and human SMCA4 homologues, regulates developmental transcriptional changes in a SUMO dependent manner (Monribot-Villanueva et al., 2016).

SNF2PH Plant Homeodomain shows high homology to the nuclear receptor binding SET domain, a histone methyltransferases (HMTases), and acting as a bookmark reader of histone modifications. Thus, lysine methylation in H3 by lysine-HMTases regulates the function leading to activation or repression of gene expression (Morishita and di Luccio, 2011). The PH domain containing protein ING2 binds histone H3 trimethylated at lysine 4 (H3K4me3), promoting active chromatin state (Pena et al., 2006). In other protozoan parasites, as *Toxoplasma gondii*, SWI2/SNF2 brahma-like (TgBRG1) contains a conserved DEXDc helicase domain as trypanosome SNF2PH. The Histone lysine methyltransferase SET1 (TgSET1) has a PHD finger suggesting a H3 PTM reader function (Jeffers et al., 2017). However, SNF2PH contains both the DEXDc helicase and PH domains. In the genome of *P. falciparum,* a protozoa parasite that also undergoes antigenic variation, there is a Snf-2 family member containing a SMARCA-related domain and interestingly a PH domain (PFF1185w), suggesting a related function. In *T. brucei*, HMTases play a function in VSG expression by the Disruptor Of Telomeric silencing 1B (DOT1B) throughout dimethylation of lysine 76 on histone 3 (H3K79). Lack of DOT1B function induced inactive VSG-ESs derepression, suggesting histone methylation is required for VSG telomeric silencing (Figueiredo et al., 2008a). Further studies showed a similar VSG-ESs silent derepression effect close to the promoters supporting histone deacetylase by DAC1-DAC3 are

required for silencing at inactive VSG-ES promoters (Wang et al., 2010). Our data suggest a dual function for SNF2PH as a dynamic modifier of chromatin architecture and as an epigenetic mark reader of developmentally regulated genes. Thus, SNF2PH shows an evolutionary acquisition of two distinct functional domains within a single transcription factor in this early branched eukaryote.

ChIP experiments using the antiserum anti-SNF2PH showed enrichment at the active VSG-ES and consequent downstream sequences as revealed by *RLuc* reporter and active VSG specific sequences, while no significant enrichment was observed in silent VSG-ESs. This data was also supported with an increased number of sequences relative to BES1 compared with other telomeric loci observed in ChIP-seq analysis, suggesting SNF2PH functions as a positive regulator of VSG-ES transcription. In addition of SNF2PH positive function in VSG-ES expression, the enrichment in procyclin promoter of SNF2PH while procyclin and PADs genes are upregulated after SNF2PH depletion, suggests a negative function in the other developmental form. However, SNF2PH function in the rDNA promoter is not clear since RNA transcripts (18s RNA and unspliced RNA) are not significantly changed upon protein depletion, while TbRPAI occupancy is not significantly reduced at ribosomal locus **(Fig 11B)**. Despite of SNF2PH recruitment at SL promoter **(Appendix II S3 Fig)**, it is problematic to demonstrate any function since occupancy of TbRPB1 was not affected in SNF2PH depleted cell line. SNF2PH maybe also associated with RNA pol II promoters allowing accessibility or recruitment of other factors, but its function seems not to be essential to drive productive RNA pol II transcription. Interestingly, ChIP-seq analysis revealed SNF2PH occupancy at the H3.V locus, which suggests that SNF2PH regulates H3.V expression, a histone variant crucial for regulating VSG expression (Müller et al., 2018). Furthermore, SNF2PH is enriched in tRNAs clusters on several chromosomes **(Fig 9C)**. tRNAs can act as insulators to prevent the spread of silencing in *S. cerevisiae* (Donze et al., 1999);(Dhillon et al., 2009) and in mammals to prevent enhancers from activating promoters (Ebersole et al., 2011);(Raab et al., 2012). Although, no evidence of such insulator function has been described in trypanosomes, the presence of SNF2PH at tRNA loci suggests a tentative model whereby SNF2PH/tRNAs could act as a global chromatin organiser to maintain bloodstream stage-specific gene expression.

Affinity purification of 3HA full-length version of SNF2PH led us to identify SHF2PH associated with proteins functionally related to chromatin **(Appendix I Table S2)**. LC-MS/MS analysis showed a major abundance of the hypothetical protein Tb927.3.1300, a non-well characterized cytoplasmatic factor related to posttranscriptional regulation of gene expression. Other relevant partner is the EF-Hand domain pair (Tb927.7.6810), which shares

homology with serine/threonine protein phosphatase 2A regulatory subunit B. EF-Hand domain containing proteins appear to be related to inositol triphosphate (IP3) signalling pathway in *T. brucei* (King-Keller et al., 2015), important in differentiation and host-parasite adaptation by downstream effectors released from phosphatidylinositol bisphosphate (PIP2), that include diacylglycerol (DAG). Another putative mediator that allows gene transcription is the Zing finger protein family member ZC3H22 (Tb927.7.2680), which is found to be implicated in posttranscriptional regulation and may act as a repressor RNA binding protein (Erben et al., 2014), highlighting another regulation mechanism given by this interaction. Furthermore, relevant VSG regulators factors may be related to SNF2PH complex network, including the subunit Spt16 from Facilitates Chromatin Transcription Complex (FACT), acting as a histone chaperone to facilitate transcription. Spt16 subunit is enriched on silent VSG-ES promoters and its depletion leads to its derepression in both developmental stages (Denninger et al., 2010). Interestingly, we also appreciate the presence of Proliferative Cell Nuclear Antigen (PCNA) essential element of the replication machinery (Valenciano et al., 2015), suggesting SNF2PH function to modulate proliferation and cell cycle progression of the bloodstream form. Other interesting factor associated with SNF2PH is the Class I transcription factor A, subunit 4 (CITFA-4), a conserved subunit of a RNA pol I promoter binding complex, necessary for RNA pol I transcription of ribosomal genes and the active VSG-ES (Nguyen et al., 2012). Finally, SNF2PH seems to be related to RNA pol I second largest subunit (TbRPA135), required for RNA pol I assembly and mediation of its specific interaction (Daniels et al., 2012), and the largest RNA pol II subunit (TbRPB1). This latter consistent with SNF2PH occupancy at pol II-transcribed loci **(Fig. 9B)**. This leads us to speculate that SNF2PH-pol II association facilitates accessibility of transcription factors to regulate RNA pol II transcription.

RNAseq analysis upon SNF2PH depletion yielded differentially expressed genes related to the differentiation to the procyclic form, from surface proteins to other genes related to the glycolysis metabolism, DNA interaction and other developmental regulated genes, which enhances the previous differentiating-like phenotype **(Appendix I Table S3)**. In addition, RT-qPCR data confirmed a reduction of the active VSG-ES transcription followed by an increase in Procyclin transcripts, suggesting that SNF2PH regulates differentiation by opposite transcriptional control of these two developmentally regulated genes. In summary, these results altogether suggest SNF2PH chromatin remodeling functions to preserve the infective form differentiation state. Interestingly, the ectopic expression of SNF2PH protein in procyclic form conferred bloodstream form characteristics, such as an atopic expression of bloodstream form telomeric VSGs, ISGs and ESAGs, reinforcing the role of the PH domain in promoting VSG

expression in an established developmental stage. Simultaneously, SNF2PH could be conceived as an important regulator factor to keep host pathogenicity and virulence by maintaining the proliferative bloodstream form fate.

The preadaptation to stumpy form transition requires AMPKα1 activation preceding SNF2PH inactivation which implies a coordinated dual regulation with opposite functions depending on the parasite environment. The activation of the energy sensor AMPK induces a similar phenotype observed after SNF2PH depletion in bloodstream form *T. brucei*, which upregulates PADs expression and could be enhanced after 5'-AMP treatment. Analysis of SNF2PH regulation during *in vivo* differentiation of mice-isolated pleomorphic cells suggested that SNF2PH is downregulated after AMPKα1 activation during the infection course. Whereas AMPKα1 activation promotes differentiation, reduced SNF2PH expression in stumpy forms observed in wild-type pleomorphic trypanosomes during mice infection confirmed the biological relevance. This result rule out a possible SIF-independent induction of differentiation by VSG-ES transcription attenuation described previously (Zimmermann et al., 2017), in wich cells are comitted to an open interval to progress the expression of a new VSG given by ES attenuation as a consequence of an early developmental transition.

We describe here a novel chromatin remodeling factor with a dual regulation function in either activation or repression in an opposite manner for developmentally regulated genes **(Fig 15)**. Our data suggest a model involving epigenetic regulation by a SUMOylated SNF2PH form enriched at the active VSG-ES chromatin maintaining a single VSG expression to allow antigenic variation work efficiently. In bloodstream form, TFs and/or histone modifier enzymes are recruited to the active VSG promoter in a SUMOylated SNF2PH-dependent manner. As a tentative model we propose that the PH domain of a SUMO-conjugated SNF2PH changes the protein interaction interface by reading a distinct histone mark (for example, recognising histone H3 trimethylated) and thus, regulates VSG-ES active transcription by increasing chromatin accessibility. Whilst, PH domain of unmodified SNF2PH may read different histone marks (for example, unmodified H3 tails) to repress silent VSG-ES transcription. In quiescent stumpy form, SNF2PH is negatively regulated by unknown events, wich ptomotes VSG-ES silencing and PADs/Procyclin transcription upregulation, suggesting a function in development.
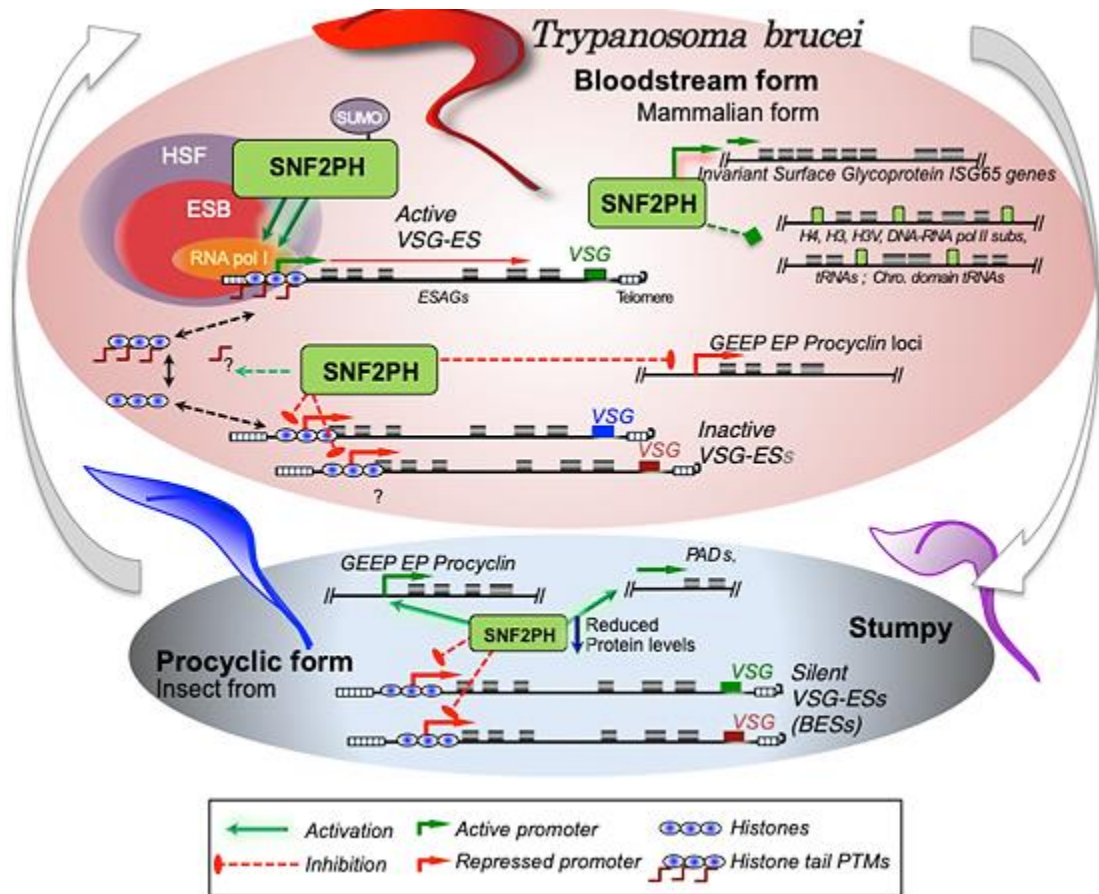
Figure 15. Trypanosome SNF2PH promotes VSG monoallelic control and bloodstream stage gene expression.
In the mammalian host (the proliferative bloodstream form), SUMO modified SNF2PH is recruited along the chromatin of active telomeric VSG-ES (BES1), particularly to the core promoter sequences **(Fig 9)**. Unmodified SNF2PH was also detected at silent BES promoters. SNF2PH occupancy at promoter chromatin facilitates recruitment of RNA polymerase I (TbRPA1) to the active VSG-ES, and high VSG transcripts levels **(Fig 11)**. Ectopic expression of SNF2PH in the insect form, where this protein is normally not expressed, increased the expression of bloodstream form specific protein surface genes, such as Invariant Surface Glycoprotein. ISGs (65kD), suggesting SNF2PH acts as a global regulator of bloodstream form developmentally regulated gene expression. In addition, other genes associated with VSG expression regulation, as the Histone variants H3V are also found to be bound to chromatin at the gene locus **(Fig 9C)**. Question marks (?) denote that possible molecular mechanisms occurring are unknown. In the Stumpy quiescent form, a transition form to the insect developmental stage, SNF2PH protein levels are downregulated and PADs, EP procyclins, GPEET2 genes characteristics of the insect form were upregulated, similar to what occurred in bloodstream form cell lines depleted of SNF2PH **(Fig 11)**. Induction of differentiation *in vitro* by AMP analog treatment from the proliferative bloodstream form to the stumpy-like form induced rapid SNF2PH reduction of the protein level and gene expression remodeling to this developmental form **(Fig 13)**. Similarly, in wild type pleomorphic strains, the SNF2PH protein was reduced during naturally occurring progression of differentiation in mice infection to acquire the stumpy quiescent form, as transition to the insect developmental stage.

6. Conclusions

6. Conclusions

1st - Affinity purification of SUMO conjugated proteins in the bloodstream form of the parasite allowed us to identify a preliminary list of SUMO of abundant substrates and its unambiguous identification of the lysine acceptor site.

2nd. - Initial functional studies of several SUMOylated proteins using RNA interference and intracellular localization suggest that some of them, such as TFIIS2.2, H3V and SNF2PH, regulate the expression of the VSG. Therefore, in addition to the biological processes regulated by SUMO described in eukaryotes, the identification of SUMO-conjugated proteins is an original unbiased approach to isolate new factors involved in the regulation of VSG expression in trypanosomes.

3rd - Out of the SUMO modified proteins from the list, we identified SNF2PH, which contains a chromatin remodeling domain (SNF_N), followed by a plant homeodomain (PH), involved in the recognition of epigenetic marks in the chromatin (particularly at histone 3) and in regulation of genetic expression during development. In trypanosomes, SNF2PH protein is highly expressed in the bloodstream form of the parasite and is SUMO-modified.

4th - SNF2PH is a SUMO-modified upregulated protein in the bloodstream form of the parasite and associates with the active VSG-ES promoter, whereas a less-SUMOylated SNF2PH is found at silent VSG promoters, suggesting that SUMO modification is important in targeting and maintenance of SNF2PH to the active VSG-ES, associated to the Expression Site Body (ESB). This data suggests that SUMO harbors an active role in SNF2PH targeting at the active VSG-ES.

5th – SNF2PH depletion by RNA interference (RNAi) decreases the expression of the Expression Site Associated Genes (ESAGs) and the VSG associated to the active expression site, whereas upregulates procyclic form markers such as EP procyclin and PADs genes. Overall, these results suggest that SNF2PH functions in order to maintain the infective bloodstream form stage of the parasite.

6th - Affinity purification of SNF2PH associated proteins in order to assess complex formation, leads to the identification of several factors with transcriptional related functions, including RNA polymerase subunits, transcription factors, and other factors involved in cell cycle regulation and DNA replication.

7th - Ectopic expression of SNF2PH in the procyclic form of the parasite, where it is expressed at reduced levels, induced the expression of bloodstream form-specific surface proteins, including silenced VSG-ES, ESAGs and invariable surface glycoproteins (ISG). However, ectopic expression of the SNF2PHdeltaPH mutant that lacks the PH domain did not produce this phenotype. These results suggest that SNF2PH functions as a transcriptional activator for the maintenance of the

expression profile of the membrane proteins of the bloodstream form and that the PH domain of the protein is necessary for this function.

8th - The quiescent form of the parasite or Stumpy form is a pre-adaptation to the form of the vector insect that requires the activation of AMPKα1. The expression levels of SNF2PH were significantly reduced in quiescent forms of the parasite obtained from both wild type strains in the course of an infection and in cells where activation of AMPKα1 was induced. It suggests that SNF2PH is negatively regulated by the AMPKα1 pathway by events of signaling still unknown.

9th - In summary, SNF2PH function is required to ensure coordinated expression of surface stage-specific proteins and thus adaptation to the mammalian host. Altogether our results suggest that SNF2PH is a central regulator of VSG monoallelic expression and maintenance infective stage-specific surface protein integrity.

Conclusiones

1ª - La purificación por afinidad de proteínas conjugadas con el péptido SUMO (*Small Ubiquitin-like MOdifier*) en la forma sanguínea del parásito nos permitió elaborar una lista de proteínas potencialmente modificadas por SUMO, así como determinar la lisina sustrato de SUMO utilizando una técnica de proteómica pionera que identifica el sitio modificado.

2ª - Los estudios funcionales iniciales de varias proteínas SUMOiladas utilizando la interferencia de RNA y la localización intracelular sugieren que algunos de ellos como TFIIS2.2, H3V y SNF2PH regulan la expresión de VSG. Por lo tanto, además de los procesos biológicos regulados por SUMO descritos en eucariotas, la identificación de proteínas conjugadas con SUMO es un enfoque no sesgado y original para aislar nuevos factores involucrados en la regulación de la expresión de la VSG en tripanosomas.

3ª - De entre los sustratos conjugados por SUMO, destacamos SNF2PH, un activador transcripcional que contiene un dominio de remodelación de cromatina (Snf2), y un dominio homeo, PH (*Plant Homeodomain*). En tripanosomas, la proteína SNF2PH se expresa predominantemente en la forma sanguínea del parásito y se encuentra modificada por SUMO.

4ª - La forma SUMOilada de SNF2PH se asocia con el promotor activo del sitio de expresión de la VSG (VSG-ES), sin embargo, la forma no SUMOilada se localiza en promotores VSG no activos, aunque en menor proporción. Estos resultados sugieren que SNF2PH vincula todos los *loci* de expresión de la VSG, mientras que la forma modificada por SUMO está enriquecida y se acopla específicamente en la cromatina del sitio de expresión activo de la VSG, reconociendo probablemente una marca epigenética específica y contribuyendo al mantenimiento del estado activo de la cromatina.

5ª - La interferencia de RNA (RNAi) de SNF2PH en la forma sanguínea reduce los mRNAs de los genes asociados (ESAGs) y la VSG del sitio de expresión activo. La disminución en la expresión de SNF2PH también aumentó los mRNAs de prociclina y PADs, característicos de la forma del insecto vector. En conjunto, estos datos sugieren que es necesario la expresión de altos niveles de SNF2PH para el mantenimiento de la forma sanguínea del parásito.

6ª - La purificación por afinidad de las proteínas asociadas a SNF2PH para evaluar la formación de complejos condujo a la identificación de varios factores con funciones relacionadas con la transcripción, incluidas las subunidades de la RNA polimerasa y factores de transcripción involucrados en la regulación del ciclo celular y la replicación del DNA.

7ª - La expresión ectópica de SNF2PH en la forma procíclica del parásito, donde se expresa en niveles muy reducidos, indujo la expresión de genes específicos de la forma sanguínea del parásito, incluyendo los VSG-ES silenciados, los ESAGs y las glicoproteínas invariables de superficie (ISG). Sin embargo, la expresión ectópica del mutante SNF2PHdeltaPH que carece del dominio PH, no produjo este fenotipo. Estos resultados sugieren que SNF2PH funciona como activador transcripcional para el mantenimiento del perfil de expresión de las proteínas de membrana de la

117

forma sanguínea del parásito, mientras que el dominio PH de la proteína es necesario para esta función.

8ª - La forma quiescente del parásito o forma *Stumpy* es una preadaptación hacia a la forma del insecto vector que requiere la activación de AMPKα1. Los niveles de expresión de SNF2PH se redujeron de forma significativa en formas quiescentes del parasito obtenidas tanto cepas salvajes en el curso de una infección como en células donde se indujo la activación de AMPKα1. Ello sugiere que SNF2PH se regula negativamente por la ruta AMPKα1 por eventos de señalización aún desconocidos.

9ª - En conjunto, los resultados mostrados en esta tesis sugieren que SNF2PH es necesario para coordinar la expresión en el desarrollo de las proteínas de superficie específicas de la forma sanguínea y, por tanto, la adaptación al huésped mamífero. Asimismo, SNF2PH es un regulador central de la expresión monoalélica de la VSG y del mantenimiento de la integridad de las proteínas de superficie específicas de la forma sanguínea infectiva del parásito.

7. Materials and Methods

7. Materials and Methods

7.1-Materials

7.1.1- Trypanosome cell lines

7.1.1.1- Single marker cell line (SM)
Monomorphic cell line (SM) from *Trypanosoma brucei brucei* (221a clone, derived from Lister427 (Wirtz *et al*. 1999).

7.1.1.2- DRALI cell line
The established dual-reporter DRALI cell line, in the 'single-marker' cell line contains a *Renilla* Luciferase (*RLuc*)-reporter plasmid (pMig104) integrated 405 bp downstream of the active ES promoter, and Firefly-Luciferase reporter (pMig126) *(FLuc)-* reporter integrated 405 pb downstream of an inactive ES promoter (BES15/TAR126 VSGbR-2/427-11). pMig104 contains as targeting sequences a region located downstream of the promoter, a highly conserved sequence among different *VSG*-ESs. Reporter activities analyses allow us to identify a clone that showed low reporter activities where the gene is inserted in the inactive ES promoter, in contrast to the reporter activities where the *RLuc* gene is inserted in the active *VSG*-ES promoter **(Figure 9A)**. We identified the site of insertion by PCR and sequencing of the flanking *RLuc* and *FLuc* regions from DRALI genomic DNAs using specific primers.

7.1.1.3- YFP::RPB5z cell line
Monomorphic cell line containing the YFP tagged small RNA pol I subunit 5Z (Landeira and Navarro, 2007).

7.1.1.4- GFP::αES (93) cell line
Monomorphic cell line with the active VSG-ES marked with 256 molecules of GFP-lacR bound to sequences upstream of the promoter to localize the position of this locus (Navarro and Gull, 2001).

7.1.1.5- 8xHis-HA TbSUMO cell line
Cell line which expressed a polyhistidine (8x)-tagged and hemagglutinin (HA) epitope tagged version of SUMO by replacement of the endogenous gene by two consecutive homologous recombination rounds. Generated cell lines were able to produce tagged SUMO conjugates as the parental cell line.

7.1.1.6 - His-HA-TbSUMO^{T106K} cell line
The established His-HA-TbSUMO^{T106K} cell line was obtained by transfection of the generated construct pLEW100v5-His-HA- TbSUMO^{T106K} as previously detailed (Iribarren et al., 2015a)

Vector linearization, transfection and selection procedures were performed as previously described (Wirtz et al., 1999).

7.1.1.7- Slender and stumpy AnTAT 90.13

AnTAT pleomorphic cell line was used to obtain different parasite life cycle forms during the 4th and 7th days post infection from BALB/c mice (Charles River) for slender and stumpy forms respectively, with approximately 10^4 total parasites. Blood samples were collected in heparin containing tubes and processed as described in Experimental procedures.

7.1.1.8 - Procyclic form 449

T. brucei 427 strain transfected with pHD449 plasmid, rising to a tetracycline repressor-expressing cell line.

7.1.2 –Oligonucleotides

Oligo name	Sequence (5'-3')	Gene	Restriction enzyme end	Application
SNF2PH-c2PrC_U	CG<u>GGATCC</u>GAGCTCTCGATGGCGCGTTTCAACGG	C-T SNF2PH	*Bam* HI	RNAi / His tag
SNF2PH-c2PrC_L	C<u>AAGCTT</u>GGTTAACGTGAGATTCACTGAAACTGCC	C-T SNF2PH	*Hind* III	RNAi / His tag
nSnf-2_NcoI_U	TCC<u>CCATGG</u>CTATGAAGTTTGAGAACCCCGTT	N-T SNF2PH	*Nco* I	Flag tag
nSnf-2_NcoI_L	AAC<u>CCATGG</u>AAATTGGACTCCATCGAACTGGTA	N-T SNF2PH	*Nco* I	Flag tag
cSnf-2_NcoI_ atg_U	TCC<u>CCATGG</u>CTATGCACCGCGACGGACTAACTGAA	C-T SNF2PH	*Nco* I	Flag tag
cSnf-2_NcoI_L	AAC<u>CCATGGG</u>TGAGATTCACTGAAACTGCC	C-T SNF2PH	*Nco* I	Flag tag
FwCTR9-Duet-NcoI	<u>CCATGG</u>CCATGCAATATATATCGGAAC	CTR9	*Nco* I	Flag tag
RvCTR9-Duet-NcoI	<u>CCATGG</u>ATATATTCCCATCATTAGCGG	CTR9	*Nco* I	Flag tag
FwTFIIS2-2-Duet-NcoI	<u>CCATGG</u>CCATGATTCCATCATTTGCTC	TFIIS2-2	*Nco* I	Flag tag
Rv TFIIS2-2-Duet-NcoI	<u>CCATGG</u>ACTGTTCAGCACCATCCAC	TFIIS2-2	*Nco* I	Flag tag
SNF2-haNter-U	GCCG<u>CTCAGC</u>TGGGATCCGGCATGAAGTTTGAGAACCCCGTT	N-T SNF2PH	*Blp* I	HA tag
SNF2-haNter-L	GCAAGCTT<u>CTCGAG</u>CCGCTTTCCTTCAGCAT	N-T SNF2PH	*Xho* I	HA tag
SNF2-PHD-U	AACG<u>CTCAGC</u>TGGGCATGAACAGGGAAGAACACTGGAGT	PHD	*Blp* I	HA tag
3'UTR-SNF2-PHD_U	GTTATCGCCACGTCATTCGTACTTA	3'UTR PHD	__	HA tag
5SNF2-New-U	GTTAACG<u>CTCAGC</u>TGGGCATGAAGTTTGAGAACCCCGTT	Full length SNF2PH	*Blp* I	HA tag
3SNF2_Bam-L	A<u>GGATCC</u>CTAGTGAGATTCACTGAAACT	Full length SNF2PH	*Bam* HI	HA /3 x HA tags
Hyg_L	CAATAGGTCAGGCTCTCGCTGAAT	Hygromycin	__	Diagnosis
Bla_L	CAGAGATGGGGATGCTGTTGATTG	Blasticidin	__	Diagnosis
3HA_U	AC<u>AAGCTT</u>GATGTACCCATACGATGTGCCGGATTACGCTGGGAGCTACCCATACGATGTGCCGGATTACGCTGGCTACCCATACGATGTGCCA	3 x HA	*Hind* III	3 x HA tag
H3v_Up	GGCAAGCTT<u>CAGCTG</u>ATGGCGCAAATGAAGAAAATAAC	H3V	*Pvu* II	3 x HA tag
H3v_Low	GCC<u>GGATCC</u>CTCGAGTTAGTTACGCTCGCCTCGGAG	H3V	*Bam* HI	3 x HA tag
TopoIB_Up	GGCAAGCTT<u>CAGCTG</u>ATGGGTAAGGCACAGAAGCCG	N-T TopoIB	*Pvu* II	3 x HA tag
TopoIB_Low	GCCGGATCC<u>CTCGAG</u>GCAGCGAATGGGATGCTTC	N-T TopoIB	*Xho* I	3 x HA tag
CTR9_Up	GGCAAGCTT<u>CAGCTG</u>ATGCAATATATATCGGAACCC	N-T CTR9	*Pvu* II	3 x HA tag
CTR9_Low	GCCGGATCC<u>CTCGAG</u>TGAACATTCAAGAGCTCTCTC	N-T CTR9	*Xho* I	3 x HA tag
Hypot_10.14190_Up	GGCAAGCTT<u>CAGCTG</u>ATGATTCATCCGAAGCTTATG	Tb927.10.14190	*Pvu* II	3 x HA tag
Hypot_10.14190_Low	GCC<u>GGATCC</u>CTCGAGTCAACAATGCCGTCGCTCGAT	Tb927.10.14190	*Bam* HI	3 x HA tag
TFIIS2-2_Up	GGCAAGCTT<u>CAGCTG</u>ATGATTCCATCATTTGCTCCA	TFIIS2-2	*Pvu* II	3 x HA tag
TFIIS2-2_Low	GCC<u>GGATCC</u>CTCGAGTCACTGTTCAGCACCATCCAC	TFIIS2-2	*Bam* HI	3 x HA tag

Table 4. Oligonucleotides used for cloning and diagnosis. Sequences from each oligonucleotide pair are shown with their respective enzyme restriction site at 5' end (underlined). N-T: Amino-terminal. C-T: Carboxi-terminal. HA: Hemagglutinin. Gene DB annotation: SNF2PH (Tb927.3.2140), CTR9 (Tb927.3.3220), TbTFIIS2-2 (Tb927.2.3420), H3V (Tb927.10.15350), DNA topoisomerase 1B (Tb927.4.1330), Hypothetical protein (Tb927.10.14190).

Gene or Region	Primer name	Sequence (5´-3`)
1	ESMP1_U	GGTGTGGCGGACGTCTCGAAC
	ESMP1_L	CCTCTAAAATACGCTCAGCCCGTCC
2	ESMP2_U	ATGAAGGTCTTGCGCTGTCT
	ESMP2_L	CGCCACACCTTGAGTCTGATATCC
3	ESMP3_U	TGTTGTTGCAGTATGTTTTTCTTC
	ESMP3_L	AGACAGCGCAAGACCTTCAT
4	ESMP4_U	CGGAGAATATTTCGGATGC
	ESMP4_L	AATCGTTACGGCCAAATTCA
5	ESMP5int_U	TGTTGTGCTTTCTAACACTTCCTT
	ESMP5_L	ATTCCTCCCCACAGAAAGA
6	ESMP5_U	TTTGTTTGTTTCATGTTTTTGTG
	ESMP5int_L	AAGGAAGTGTTAGAAAGCACAACA
7	ESMP7_U	TGGTTGCAGTTATGGAGCAG
	ESMP7_L	CCAACCCTCAAGCAGGATAA
VSG221	VSG221_U	AGCTAGACGACCAACCGAAGG
	VSG221_L	CGCTGGTGCCGCTCTCCTTTG
PseVSG	221pseudo_U	CAAGCATTACCAGAGAAGT
	221pseudo_L	CGTCATTCAGTTTCCTTATT
VSG121	VSG121_U	CCTGACATCGGACGGTAAC
	VSG121short_L	TGGTCGTATTTGCCTTCCTT
VSGJS1	VSGJS1_U	TTCTGCTTCTTTGCCCTTGT
	VSGJS1_L	AAAATGAAGCGGAAATGGTG
VSGVO2	VSGVO2 short_U	ACAGAATCGGCCACAGAAAG
	VSGVO1_L	CATTTCCGCGTTGTCTTGTA
BC118	BC118_U	CAGAAGCGCCAATACAACAA
	BC118_L	CGTTAGAAACCACGCCAGTT
BC222	BC222_U	CGGAAGAGACAATTCGAAGG
	BC222_L	CGCCATTGACATCCCTACTT
18S	18S_U	GACGTAATCTGCCGCCAAAAT
	18S_L	AACGCCATGGCAGTCCAGTAC
rDNA pro	rDNAprom_U	GTCAATACAACACACAATAGG
	rDNApro_L	CTTAACTGAGGAAGTGTCATA
rDNA sp	rDNAspacer_U	ATTTTCTCTACCCCTCTCTT
	rDNAspacer_L	ATCATCGTATCATTTTCATC
EP cds	EP3-2U	ATGGCACCTCGTTCCCTTTA
	EP3-2L	AGAATGCGGCAACGAGACCAA
EP pro	ProcyProm_U	AGTTTAAGATGTTCTCGTGAT
	ProcyProm_L	CTTTTTGGTGTAATTGAAGTC
FLuc	Luc_U	GTGTTGGGCGCGTTATTTAT
	Luc_L	CATCGACTGAAATCCCTGGT
RLuc	Ren_U	GATAACTGGTCCGCAGTGGT
	Ren_L	ACCAGATTTGCCTGATTTGC
SL	SL_U	CCGACACGTTTCTGGCACGACAG
	SL_L	TGCGTGTGTTGGCCCAGCTGCTAC
SL Pro	SL Pro_U	ATAATGCATGACCCCTTTGTTTCCCATAA
	SLPro_L	ATAGATCTCAGAAACTGTTCTAATAATAGCGTTAGTTG
Tubulin	Tub_U	AGGCAACGGGAGGTCGCTATG
	Tub_L	GGGATGGGATGATGGAGAAAG
Myosin B	MyoB_U	CTGCAGAACAAGCACGGCATT
	MyoB_L	ACGCTCAACAGTGGCAGTGAA
U2	U2_U	CTGCGTGATCTTTTGTTCCT
	U2_L	CGACTGCGTTTTCGTATTTT
C1	C1_U	TTGTGACGACGAGAGCAAAC
	C1_L	GAAGTGGTTGAACGCCAAAT
5S	5Snew_U	GACCATACTTGGCCGAATG
	5Snew_L	TACAACACCCCGGGTTCC
705	rDNA_705 U	ATTGCCGCTGCTTTTTACAC
	rDNA_890 L	TATCAGGTGCCAAGCCCTAC
PAD1	PAD1_U	TCATGGTTTCGCCATTCTCGTAACC

	PAD1_L	CTCAGCCACTTCTCTCCTACAACAC
PAD2	PAD2_U	AGGGTGATGCCAAAGAACAC
	PAD2_L	TACCCACACCGTTGAGAACA
AMPKa1	C1_U	GGATCCAGGCAAATGACCAAG
	C1_L	GTCCCTCTTCTACCTTTACAG
SNF2PH	Snf2_qPcr_U	AGTCCCGTGAAGGTCCTCTT
	Snf2_qPcr_L	CTTGTCGAACTGGTGGGTCT

Table 5. Primers used in ChIP-qPCR analysis and qRT-PCR.

7.1.3 -DNA Plasmids constructs

Plasmid	Origin	Description	Tag	Linearization	Selection marker
p2T7Bla	A. Estévez	Phleomycin marker from p2T7 (*LaCount et al.*) was sunstituted by Blasticidin used to generate an iRNA from a bidirectional T7 promoter.	—	*Not*I	Blasticidin (2.5µg/ml BF)
p2T7-cSNF2PH-Bla	This thesis	The PCR product corresponding to 1113-bp of C-T SNF2PH using the oligo pair **SNF2PH-c2PrC_U /SNF2PH-c2PrC_L** was cloned into the *Bam*HI / *Hind*III restriction sites of p2T7Bla.	—	*Not*I	Blasticidin (2.5µg/ml BF)
pET28a(+)-6xHis tag	Novagen	Expression plasmid in *E. coli* to produce an inducible 6 x Histidine fusion protein.	6xHis	—	Kanamycin (25µg/ml)
pET28a(+)-6xHis-cSNF2PH	This thesis	The PCR product corresponding to 1113-bp of C-T SNF2PH using the oligo pair **SNF2PH-c2PrC_U /SNF2PH-c2PrC_L** was cloned into the *Bam*HI / *Hind*III restriction sites of pET28a(+)-6xHis to allow the heterologous expression in *E.coli.*	6xHis	—	Kanamycin (25µg/ml)
pET28a(+)-3xFlag	P. Irribaren / V. Álvarez	Expression plasmid in *E. coli* to produce an inducible 3 x Flag fusion protein.	3xFlag	—	Ampicillin (100µg/ml)
pET28a(+)-3xFlag-SNF2PHN	This thesis	The PCR product corresponding to 735-bp of N-T SNF2PH using the oligo pair **nSnf-2_NcoI_U /nSnf-2_NcoI_L** was cloned into the *Nco*I restriction sites of pET28a(+)-3xFlag to allow the heteroelogous expression of the substrate protein in *E. coli* for the in *vitro* SUMOylation system.	3xFlag	—	Ampicillin (100µg/ml)
pET28a(+)-3xFlag-SNF2PHC	This thesis	The PCR product corresponding to 705-bp of C-T SNF2PH using the oligo pair **cSnf-2_NcoI_ atg_U/cSnf-2_NcoI_L** was cloned into the *Nco*I restriction sites of pET28a(+)-3xFlag.	3xFlag	—	Ampicillin (100µg/ml)
pET28a(+)-3xFlag-TFIIS2-2	This thesis	The PCR product corresponding to 1144-bp of TFIIS2-2 ORF using the oligo pair **FwTFIIS2-2-Duet-NcoI /Rv TFIIS2-2-Duet-NcoI** was cloned into the *Nco*I restriction sites of pET28a(+)-3xFlag	3xFlag	—	Ampicillin (100µg/ml)
pET28a(+)-3xFlag-CTR9	This thesis	The PCR product corresponding to 2704-bp of CTR9 ORF using the oligo pair **FwCTR9-Duet-Ncol /RvCTR9-Duet-Ncol** was cloned into the *Nco*I restriction sites of pET28a(+)-3xFlag.	3xFlag	—	Ampicillin (100µg/ml)
pTAP-HA-H2Bv-Hyg	C. Cordón	Expression plasmid in *T. brucei* targeted to the ribosomal spacer (ectopical expression) to express the H2Bv gene under an inducible T7 (procyclin) promoter.	HA	*Not*I	Hygromycin (2.5µg/ml BF; 50µg/ml PF)
pTAP1-RPB7cBla-Overex	D. López	Expression plasmid in *T. brucei* to express the RPB7 gene, used as a backbone for Blasticidine selection marker.	HA	*Not*I	Blasticidin (2.5µg/ml BF)
pTAP-HA-nSNF2PH-Hyg	This thesis	The N-T fragment of SNF2PH (1068-bp) was aplified using the oligo pair **SNF2-haNter-U/SNF2-haNter-L** and cloned into the *Blp*I/*Xho*I sites of pTAP-HA-H2Bv-Hyg and expressed in *T. brucei* by homologous recombination (readthrough) under an inducible T7 promoter.	HA	*Sex*AI	Hygromycin (2.5µg/ml BF)

Plasmid	Origin	Description	Tag	Linearization	Selection marker
pTAP-HA-nSNF2PH-Bla	This thesis	Expression plasmid containing the N-T fragment of SNF2PH (readthrough) with Blasticidine marker, obtained by cloning the HA SNF2 N-T gene (*Nsi*I/*Xho*I digested) derived from pTAP-HA-nSNF2PH-Hyg into the *Nsi*I/*Xho*I restriction sites of pTAP1-RPB7cBla-Overexp, used as backbone.	HA	*Sex*AI	Blasticidin (2.5µg/ml BF)
pTAP-HA-SNF2PH-FL-Hyg	This thesis	Plasmid generated to ectopically express the full length SNF2PH fragment under an inducible T7 promoter. The PCR product corresponding to the 2866-bp of the SNF2PH ORF was amplified using the oligo pair **5SNF2-New-U/3SNF2_Bam-L**, and cloned into the *Blp*I/*Bam*HI restiction sites of pTAP-HA-H2Bv-Hyg	HA	*Not*I	Hygromycin (2.5µg/ml BF; 50µg/ml PF)
pTAP-HA-PHD-Hyg	This thesis	Plasmid generated to ectopically express the PH domain of SNF2PH under an inducible T7 promoter. PCR product corresponding to the 747-bp fragment of PH domain was aplified using the oligo pair **SNF2-PHD-U/ 3SNF2_Bam-L** and cloned into the *Blp*I/*Bam*HI sites of pTAP-HA-H2Bv-Hyg.	HA	*Not*I	Hygromycin (2.5µg/ml BF)
p3HA-PHD-Hyg	This thesis	Plasmid generated to ectopically express the 3xHA tagged PH domain of SNF2PH by cloning the 3xHA PHD gene fragment, amplified with the oligo pair **3HA_U/3SNF2_Bam-L** into the *Hind*III/*Bam*HI restriction sites of pTAP-HA-PHD-Hyg.	3xHA	*Not*I	Hygromycin (2.5µg/ml BF)
p3HA-PHD-Bla	This thesis	Plasmid generated to ectopically express the 3xHA tagged PH domain with Blasticidine marker, by cloning the 3xHA PHD gene fragment (*Nsi*I/*Xho*I digested) derived from p3HA-PHD-Hyg into the *Nsi*I/*Xho*I restriction sites of pTAP-HA-nSNF2PH-Bla, used as backbone	3xHA	*Not*I	Blasticidin (2.5µg/ml BF)
p3HA-nSNF2PH-Hyg	This thesis	Plasmid generated to express the 3xHA tagged N-T fragment of SNF2PH (readthrough) under an inducible T7 promoter, by cloning the HA nSNF2PH insert (*Blp*I/*Xho*I digested) derived from pTAP-HA-nSNF2PH-Hyg into the *Blp*I/*Xho*I restriction sites of p3HA-PHD-Hyg, used as backbone.	3xHA	*Sex*AI	Hygromycin (2.5µg/ml BF)
p3HA-nSNF2PH-Bla	This thesis	Plasmid generated to express the 3xHA tagged N-T fragment of SNF2PH (readthrough) with Blasticidine marker, by cloning the 3xHA nSNF2PH insert (*Nsi*I/*Xho*I digested) derived from p3HA-nSNF2PH-Hyg into the *Nsi*I/*Xho*I restriction sites of p3HA-PHD-Bla, used as backbone.	3xHA	*Sex*AI	Blasticidin (2.5µg/ml BF)

Plasmid	Origin	Description	Tag	Linearization	Selection marker
p3HA-SNF2PH-FL-Hyg	This thesis	Plasmid generated to ectopically express the 3 x HA tagged full length SNF2PH fragment under an inducible T7 promoter, by cloning the full length SNF2PH fragment (*Pvu*II/*Bam*HI digested) derived from pTAP-HA-SNF2PH-FL-Hyg into the *Pvu*II/*Bam*HI restriction sites of p3HA-PHD-Hyg, used as backbone.	3xHA	*Not*I	Hygromycin (2.5μg/ml BF; 50μg/ml PF)
p3HA-K2A-nSNF2PH-Hyg	This thesis	Plasmid generated to express the 3xHA tagged N-T SNF2PH K2A mutant (readthrough) by cloning the 412-bp of the N-terminal ORF with K2A mutation into the *Pvu*II/*Sex*AI restriction sites of p3HA-nSNF2PH-Hyg, used as backbone.	3xHA	*Bsm*I	Hygromycin (2.5μg/ml BF)
p3HA-K2A-nSNF2PH-Bla	This thesis	Plasmid generated to express the 3xHA tagged N-T SNF2PH K2A mutant (readthrough) with Blasticidine marker, by cloning the N-T SNF2PH K2A insert (*Nsi*I/*Xho*I digested) derived from p3HA-K2A-nSNF2PH-Hyg into the *Nsi*I/*Xho*I restriction sites of p3HA-nSNF2PH-Bla, used as backbone.	3xHA	*Bsm*I	Blasticidin (2.5μg/ml BF)
p3HA-H3v-Hyg	This thesis	Plasmid generated to ectopically express the 3xHA tagged H3v by cloning the full-length sequence (420 bp), amplified with the oligo pair **H3v_Up /H3v_Low** and cloned into the *Pvu*II/*Bam*HI restriction sites of p3HA-PHD-Hyg.	3xHA	*Not*I	Hygromycin (2.5μg/ml BF)
p3HA-TopoIB-Hyg	This thesis	Plasmid generated to express the 3xHA tagged N-T Topo Isomerase IB (readthrough), by cloning the N-T sequence (555-bp) amplified with the oligo pair **TopoIB_Up /TopoIB_Low** and cloned into the *Pvu*II/*Xho*I restriction sites of p3HA-PHD-Hyg.	3xHA	*Not*I	Hygromycin (2.5μg/ml BF)
p3HA-CTR9-Hyg	This thesis	Plasmid generated to express the 3xHA tagged N-T CTR9 (readthrough), by cloning the N-T sequence (1065-bp) amplified with the oligo pair **CTR9_Up /CTR9_Low** and cloned into the *Pvu*II/*Xho*I restriction sites of p3HA-PHD-Hyg.	3xHA	*Sex*AI	Hygromycin (2.5μg/ml BF)
p3HA-Hypot_10.14190-Hyg	This thesis	Plasmid generated to ectopically express the 3xHA tagged Hypothetical protein Tb927.10.14190 by cloning the full-length sequence (1383-bp), amplified with the oligo pair **Hypot_10.14190_Up /Hypot_10.14190_Low** and cloned into the *Pvu*II/*Bam*HI restriction sites of p3HA-PHD-Hyg.	3xHA	*Not*I	Hygromycin (2.5μg/ml BF)

Plasmid	Origin	Description	Tag	Linearization	Selection marker
p3HA-TFIIS2-2-Hyg	This thesis	Plasmid generated to ectopically express the 3xHA tagged TFIIS2 by cloning the full-length sequence (1146-bp), amplified with the oligo pair **TFIIS2-2_Up /TFIIS2-2_Low** and cloned into the *Pvu*II/*Bam*HI restriction sites of p3HA-PHD-Hyg.	3xHA	*Bsm*I	Hygromycin (2.5µg/ml BF)
pLEW100v5-His-HA-TbSUMO	D. López/ P. Irribaren	Plasmid generated to ectopically express the Histidine HA tagged version of TbSUMO under an inducible T7 promoter.	8xHis HA	*Not*I	Phleomycin (2.5µg/ml BF)
pLEW100v5-His-HA-TbSUMOT106K	P. Irribaren / V. Álvarez	Plasmid generated to ectopically express the Histidine HA tagged T106K mutant version of TbSUMO under an inducible T7 promoter.	8xHis HA	*Not*I	Phleomycin (2.5µg/ml BF)

Table 6. Plasmid constructs generated for transfection in *T. brucei* and heterologous protein expression in *E. coli*. Brief descriptions of origin, cloning strategy, expressed epitope tag, enzyme used for linearizing and BF/PF drug concentrations for each generated construct are detailed above. ORF: Open Reading Frame. n: N-Terminal. c: C-Terminal. UTR: Untranslated Region.

7.1.4- Antibodies

7.1.4.1- SNF2PH monoclonal antibody
C-terminal fragment of SNF2PH (Tb927.3.2140) was amplified by PCR **(Appendix Methods)** and the PCR product was cloned into *Bam*HI and *Hind*III sites of pET28a vector (Novagen) expressed as a C-terminal His tag. The C-terminal Histidine tagged fragment of SNF2PH was inoculated into mice to generate anti-SNF2PH (11C10E4) monoclonal antibody (mAb), using standard procedures. Hybridomas were screened against the recombinant protein by ELISA and further confirmed by western blot analysis using trypanosome protein extracts that recognized the protein of the expected size. Hybridome cell line 11C10E4 was grown as ascites.

7.1.4.2 -SNF2PH polyclonal antibody
SNF2PH polyclonal antibody was obtained by affinity purification from rabbit antiserum after several inoculations of the recombinant protein using an *Aminolink* column (Pierce), following the manufacturer's instructions.

7.1.4.3 –TbSUMO monoclonal antibody
Anti-TbSUMO (1C9H8) monoclonal antibody was generated as previously described (López-Farfán et al., 2014).

7.1.4.4-TbRPAI polyclonal antibody
Anti-TbRPA1 antiserum was generated as previously described (Navarro and Gull, 2001).

7.1.4.5- Anti-tubulin KMX mAb
Anti-tubulin monoclonal antibody clone KMX was generated as previously described (Sasse and Gull, 1988).

7.1.4.6- Anti-MVP monoclonal antibody
Anti-MVP monoclonal antibody was generated as previously described (Barquilla et al., 2012).

7.1.4.7- Rabbit anti-VSG221 antiserum
Anti-VSG221 antiserum was was generated as previously described (Castillo-Acosta et al., 2015).

7.2-Methods[1]

7.2.1- *Trypanosoma* cell culture

The bloodstream form of *T. brucei* strains were cultured in HMI-9 medium supplemented with 10% fetal bovine serum at 37ºC with 5% of CO_2 in a humidified incubator. The procyclic form was cultured at 28ºC in SDM-79 medium at 28ºC without CO_2 in a non-humidified incubator. Cell density was checked by cell counter (Beckman).

7.2.2- Bloodstream form transfection

Bloodstream form culture ($1-2x10^7$ parasites) was harvested in a cell density of $0.5-1x10^6$ parasites/ml. Cells were washed with prewarmed Trypanosome Dilution Buffer (TDB) supplemented with 20 mM glucose and resuspended in 420 µl of Cytomix at 37ºC per transfection in a sterile cuvette including 10 µg of plasmidic DNA previously linearized. Transfections were performed in a *Gene pulser cuvette 0.2 cm* electroporator (BioRad) using *Square* parameters at 1,5kV, 25 µF and 0.2 ms. Electroporated cells were diluted in 12.5 ml of prewarmed HMI-9 and dispensed in a transparent flat bottom 24-well plate at 0.5 ml/well. After 18h, drug of selection was added at its suitable concentration for clone selecion. After 5-6 days, positive clones were transferred in a new flask containing fresh media supplemented with each respective drug.

7.2.3- Procyclic form transfection

Procyclic form culture ($1-2x10^7$ parasites) was harvested in a cell density of $5-15x10^6$ parasites/ml. Cells were washed and resuspended in Cells were washed with cold TDB and resuspended in 100 µl of ZPFM per transfection in a sterile cuvette including 10 µg of plasmidic DNA previously linearized. Transfections were performed in an *Amaxa Nucleofector* electroporator (BioRad) using the stablished parameters for the X-014 program. Electroporated cells were diluted in 10 ml of SDM-79 at 28ºC in a T25 flask. After 18h, cells were selected with the drugs of interest and subjected into serial 1:10 dilutions up to 1.100 and dispensed into a transparent flat bottom 96-well plate in a final volume of 100 µl /well. After 12-16 days, positive clones were transferred in a new flask containing the respective drugs.

7.2.4- Freezing and defrosting *T. brucei*

Cells were harvested in a density of $0.5-1x10^6$ and resuspended in 200 µl of HMI-9 or SDM-79 including 10% glycerol as a preservative for bloodstream and procyclic forms, respectively.

---

[1] The composition of buffers for all the described methods appear detailed in Appendix III

Vials would be preserved at -80ºC or in liquid nitrogen for a long term. For thawing cells, the content of the vial is diluted in 10 ml of prewarmed HMI-9 or SDM-79.

7.2.5- Animal infection and isolation of *T. brucei* in blood

In order to expand bloodstream form cultures to obtain large amounts of cellular mass needed for further analysis, Balb/c mice or Wistar rats were infected with $1x10^6$ or $5x10^7$ bloodstream form parasites, respectively. After 3 days post infection, the animals were sacrificed with CO_2 and the blood was extracted by heart puncture or maintained till 5 days in the case for stumpy formation in mice.

7.2.6- Isolation of *T. brucei* in blood

7.2.6.1- BALB/c mice

Parasites are isolated from blood cells by harvesting 3 min at 1400 rpm at room temperature. This step generates a white layer between the serum and red blood cells, containing parasites that are resuspended in 500 µl of Separation buffer (SB) supplemented with 10mM glucose (SB glucose) to further be applied onto a DEAE cellulose column previously equilibrated with SB-glucose. The eluted parasites were washed in 10 ml of Phosphate Buffered Saline 1x (PBS) including protease inhibitors (2x) previous to be harvested and stored at -80ºC.

7.2.6.2- Wistar rats

The extracted blood is diluted 1:1 with PBS and the sample was applied onto Ficoll Histopaque (Sigma Aldrich) at a ratio of 3 ml resine per 10 ml of diluted blood. The mixture was centrifuged without brake to avoid the rupture of the parasitic layer, and the serum layer formed over the parasitc ring is discarded. The parasitic ring corresponding to the fractioned parasites were washed and centrifuged to finally be resuspended in 5 ml of SB glucose at 4ºC in order to be applied through the DEAE column, previously equilibrated with SB glucose. Then, the fraction not retained through the column corresponding to the isolated parasites were centrifuged and stored at -80ºC including protease inhibitors.

7.2.7- 2D Inmunofluorescence

Bloodstream and procyclic fom cells were fixed with 4% of paraformaldehyde (PFA) (p/v) during 20 min, and then permeabilized with 1% NP-40 in PBS (V/V) for 1h at room temperature. After washing, cells were blocked with 5% of blocking reagent (Roche) in PBS-T during 30 min and incubated with the primary antibody in 0.5% of blocking reagent during 45 minutes at room temperature. After washing, secondary antibodies conjugated with Alexa Fluor 488 or 594 were incubated like the same procedure used for the primary antibodies.

Finally, the slides are dried with methanol and mounted using *VectaShield mounting media* with DAPI (4'6'-diamidino-2-phenylindole) (Vector Labs).

7.2.8- 3D Inmunofluorescence

Deconvolution 3D microscopy analysis was performed as previously described (Landeira *et al.* 2007). Stacks (0.1-µm z step) acquisition was performed with a microscope system (Cell R IX81; Olympus), 100x objective, NA 1.40, illumination system (MT20; Olympus), and camera (Orca CCD; Hamamatsu). Deconvolution of 3D images was performed using Huygens Essential software (version 2.9; Scientific Volume Imaging) using an experimentally calculated point-spread function with 0.2 µm TetraSpeck microspheres (Invitrogen). Colocalization analysis, pseudocoloring and maximum intensity projections were performed using *ImageJ* version 1.43 software (National Institutes of Health). For the colocalization mask, two points were considered as colocalized if the ratio of the intensity is strictly higher than the ratio setting value set up at 80% in all the analyses. Colocalization percentage Scoring was done by direct optical observation and questionable cells were analyzed by 2D or 3D digital imaging.

7.2.9- RNAi experiments

For RNAi experiments, *T. brucei brucei* Lister 427 SM ("single marker") bloodstream form cell line was used (Wirtz et al., 1994). RNAi constructs were made using the vector p2T7Bla to allow doxycycline-induced expression of dsRNA from a bidirectional T7 promoter in the DRALI bloodstream form cell line. Amplified fragments of the open reading frame of the gene of interest were cloned into the *Bam*HI and *Hind*III sites of p2T7Bla. RNAi stable cell lines were diluted daily to maintain the density of the cultures between 1×10^5-1×10^6 parasites/ml in a medium with selection drugs. dsRNA synthesis was induced by the addition of 1 µg/µl of tetracycline analogue Doxycycline (Sigma) in every dilution step. The growth rate was compared upon 72h with the parental cell line grown in parallel, and clones with a higher growth diminution above 20% were selected.

7.2.10- RNAseq analysis

Total RNA from at least two independent biological replicates of both SNF2PH knock down (BF) and ectopic expression after 48h of doxycycline induction were used to generate a library from poli (A)+ mRNA isolated fragments. Libraries were sequenced on an Illumina NextSeq 500 platform (150 cycles) in paired-end mode with a read length of 2x76 bp and sequence depth of approximately 50 million reads per sample. The miARma-Seq pipeline (Andrés-León et al., 2016) was used to analyze all transcriptomic data. In detail, this pipeline contains all needed software to automatically perform any kind of differential expression analysis. It uses fastqc to check the quality of the reads and aligned them using hisat2 on T. brucei TREU427 reference

genome (TritrypDB release 39). Subsequently, the aligned reads are quantified and summarized for each gene using featurecounts. Finally, gene counts are analyzed using the edgeR package from Bioconductor. In such a way, all samples were size corrected in order to be comparable and then normalized using the TMM method from the EdgeR package. TMM values for each gene were used for the differentially expression analysis. RPKM values for each gene were calculated from the normalized read counts values using the rpkm method from edgeR. Genes transcripts isolated form uninduced versus induced SNF2PH RNAi cells with a [log2FC] >=1 (log2 of Fold Change) and FDR <= 0.05) were considered as differentially expressed. Additionally, miARma-Seq (Andrés-León et al., 2016) generated a volcano plot to facilitate the identification of genes that felt higher variation in expression.

7.2.11- Isolation of Genomic DNA from *T. brucei*

To isolate genomic DNA, a Trypanosome culture (1 x10^8 total bloodstream or procyclic form cells) was centrifuged at 1400g during 10 min at 4ºC. Next, parasites are resuspended in 400 µl of EA buffer, incubated 5 min on ice and centrifuged at 8000 rpm during 15 min at 4ºC to discard the supernatant. The resulting pellet is resuspended again with 400 µl EB buffer with 1 µl of RNaseA (10 mg/ml) and incubated 15 min at 37ºC. Next, it was added 5 µl of Proteinase K (10 mg/ml, Sigma) and incubated at 55ºC during 2h-overnight. It is added a volume of phenol, mixing by inversion and centrifuging at 13000 rpms at RT. The supernatant is transferred in a new tube. The procedure is repeated again with a volume of phenol: chloroform (1:1) and a final volume of chloroform. Isolated DNA is precipitated with 2.5 and 0.1 volumes of ethanol and 3M acetic acid pH 5.0 respectively, and then washed with 70% ethanol. Finally, the pellet was air-dried and resuspended in 100 µl of 10 mM Tris HCl pH 8.0.

7.2.12- Isolation of total RNA from *T. brucei*

Bloodstream and procylic form cells were grown to a final density of 1 x10^6 in 50 ml (Approximately 5x10^7 total cells) and harvested 10 min at 1400 g. Cells were resuspended in 350 µl of buffer RLT followed by 350 µl of 70% ethanol. The sample (700 µl) was transferred to a RNasy spin column (Qiagen) placed in 2 ml collection tube and centrifuged 15 sec at 8000g. Next, the sample was washed by applying 350 µl of buffer RW1 to the column, centrifuging again and proceeding to DNase I digestion (Qiagen). DNaseI incubation mix is prepared with 10 µl of DNAseI stock solution in 70 µl of buffer RDD and the final volume was applied directly to the RNasy spin column membrane at 37ºC during 30 min. Succeeding DNAseI incubation, the column is washed two times with 500 µl buffer RPE, after the last wash, the sample was centrifuged 12 min at maximum speed to discard any residual volume. Finally, the column was placed into a new eppendorf tube and eluted with 30 µl of RNAse free water.

7.2.13- Isolation of plasmidic DNA from *E. coli*

Isolation of plasmidic DNA from DH5α *E.coli* strain was carried out by the Alkaline Lysis protocol (Sambrook J, 2001), using the buffers P1 (TrisHCl 50mM pH 8, EDTA 10mM, RNAse 100µg/ml), P2 (NaOH 200mM, SDS 1%) and P3 (potassium acetate 3M pH 5.5).

7.2.14- Cloning of genomic sequences

Regular techniques of gene cloning, including PCR amplification, enzymatic digestion, plasmid transformation and isolation were carried out according to Sambrook, Russell and Maniatis. (Sambrook J, 2001). All cloning procedures for all generated constructs are detailed in **Tables 4 & 6**.

7.2.15- cDNA synthesis

cDNA synthesis was performed using *qScript cDNA Supermix* (Quanta Biosciences). The reaction mix consist in 4 µl *qScript cDNA Supermix* (5x) with 1 µg of RNA in a final volume of 20 µl. The amplification program sets as follow: 5 min at 25ºC, 30 min at 42ºC and 5 min at 85ºC. For each quantitative PCR reaction mix it was used 1 µl of cDNA for a single copy gene or 1 µl of a 1:100 dilution for multicopy genes. Reactions were performed in duplicate in 96-well plates with a non retrotranscribed sample to exclude any contamination from genomic DNA. Relative expression levels were referred to a control (parental cell line) and normalized against a housekeeping gene (U2, Pol-III transcribed gene), using the software *BioRad CFX Manager* (Bio-Rad). Experimental condition was performed in triplicate and analyzed by Student's T-test using Sigma Plot software.

7.2.16- Quantitative PCR (qPCR)

Quantitative PCR was performed using the iQ SYBR Green Supermix (Bio Rad). The reaction mix consists in 5 µl of iQ SYBR Green Supermix (2x), 0.5 µl of each pair of ~100-150 mer oligonucleotids at 10 µM and 1 or 2 µl of sample depending if it is a RNA or a ChIP product, respectively, in a final volume of 10 µl. Duplicated reactions in 96-well plates were amplified in a BioRad CFX96 Cycler, using the conditions as follows: 3 min at 95ºC (previous denaturalization) followed by 32 cycles of 30 sec at 95ºC / 30 sec at 60ºC / 30 sec at 72ºC, with a final extension of 1 min at 72ºC. Fluorescent reads were measured during the extension step. Each pair of oligonucleotids was designed by the software *Primer3* and synthetized by Sigma Aldrich. To determine primer efficiency, it was generated a standard curve for serial input dilutions (1:50, 1:100 and 1:200) in order to extrapolate the amount of precipitated DNA corresponding to its abundance in a given locus. Each experimental assay was displayed in triplicate and analyzed by Student's T-test. Oligonucleotid pair sequences appear detailed in **Table 5.**

7.2.17- Relative quantification (qPCR)

Relative quantification of PCR fragments was performed by qPCR as described before with genomic DNA template and the quantification was performed using the $\triangle\triangle$CT method and *RLuc* single copy gene was used as the reference sample.

7.2.18- Chromatin Immunoprecipitation (ChIP)

7.2.18.1- Fixing cells

Bloodstream and procyclic fom cells were fixed with formaldehyde solution (Sigma) to a final concentration of 1% and gently mixed during crosslink. Then, it was added 2.5 M glycine 20 mM TrisHCl to a final concentration of 125mM. Fixed cells were centrifuged and washed with PBS1x at 4ºC. Pellet from centrifuged cells was resuspended in 10 ml of PBS1x including protease inhibitors (PIC1x) and centrifuged again to be stored at -80ºc.

7.2.18.2- Cell lysis

Cells were resuspended in 1 ml Lysis buffer per 10^8 cells and aliquot in 500 µl (containing 5 $\times 10^7$ cells) to shear the chromatin to average fragment size of 500-300bp in a Vibracell sonicator. After sonication, cells were centrifuged and carefully removed the supernatant containing shared chromatin.

7.2.18.3- Preparing Protein G and Sepharose 4-B beads

Protein G and Sepharose 4B beads (Sigma) were washed with TE buffer. Next, Protein G beads were blocked over night with 1mg/ml BSA in TE buffer.

7.2.18.4- Chromatin Immunoprecipitation

Thawed chromatin aliquots were diluted 1:5 in ChIP dilution, precleared with 150 µl Sepharose 4B beads during 1h at 4ºC on rotation and centrifuged at 1000g during 5 min at 4ºC. For a single IP, 2.5 ml corresponding to 5×10^7 cells were used to incubate with the antibody over night on rotation at 4ºC. Before that, a 10% of input sample was taken as a reference value. Next day, 40 µl of Protein G beads were added and incubated during 1h at 4ºC on rotation. Beads were centrifuged at 1000g during 2 min and washed in low-bind tubes with 1 ml of the following; 1x Low salt buffer, 1x High salt buffer, 1x LiCl buffer and 2x TE buffer. To elute, it was applied 500 µl of elution buffer and incubated 30 min at RT on rotation. After spinning down, the beads were removed, and the supernatant were transferred in a new low bind tube.

7.2.18.5- Reversal cross-link and DNA purification

Eluted chromatin was reverse cross-linked by adding 20 µl of NaCl to a final concentration of 200 mM over night at 65ºC in a rotatory oven. Then 5 µl of RNAse A were added at 10 mg/ml (incubated 1h at 37ºC) and 8 µl Proteinase K at 20 mg/ml (incubated 2h at 55ºC). Total DNA

was extracted by phenol-chloroform method and after ethanol precipitation; the pellet was resuspended in 50 μl of bidistilled water.

7.2.19- Generation of a ChIP-seq library

Immunoprecipitated DNA (~5 ng) from each condition was evaluated in a 2100 BioAnalyzer to assess fragmentation size and subjected to end-repair enzymatic plus dA-tailing treatments further to be ligated to adapters using the Illumina TruSeq DNA Sample preparation kit, following the manufacturer's instructions. Adapter-ligated libraries were enriched with 15 cycles of PCR using Illumina PE primers and purified with a double-sided SPRI size selection in a range below 300 bp. Libraries were sequenced in an Illumina NextSeq 500 platform leading to a 650000 reads per sample. The raw reads were processed using the miARma-Seq pipeline (Andrés-León et al., 2016) to measure quality, adapter sequence removal and read alignment. Briefly, this software first assessed the quality of the sequences using FASTQC tool kit (Andrews, 2010). After that, adapter sequence where removed using the cutadapt (Martin, 2011) utility. Once reads were processed, they were aligned against the *Trypanosoma brucei* Lister 427 genome obtained from TriTrypDB version 34 using the BWA aligner with default parameters (Li and Durbin, 2009). Later, final results obtained from miARma-seq were processed with macs (Feng et al., 2012). Therefore, each paired sample (-f BAMPE) chip (-t) was processed against the input sample (-c) to eliminate general peaks in both types of samples using as organism size 2.7x107 (-g). The correspondence of peaks between both types of samples (HA-SNF2PH and HA-SNF2PH K2A) and with gene sequences was carried out with the intersectBed script from bedtools (Quinlan and Hall, 2010) using the gene annotation provided by *TriTrypDB version 34*. Reads were mapped to the *T.brucei TREU427* annotated genes (TritrypDB release 39) and visualized in GB browser A starting pool of 8 amplicons 18S, U2, C1, prom SSR7, ESPM-1/4, VSG221 (BES1) and VSG121 (BES3)) were combined together in a single lane per condition (including respective inputs). Coding sequences for 18 and U2 were used as reference genes to evaluate the immunoprecipitation efficiency in each experimental case. Differential peak distribution was represented as fold enrichment relative to input and assessed by –*logP* value, considering a 0-nucleotide mismatch to discern telomeric sequences.

ChIP-seq analysis in Figure 9 E. To discriminate among the BES promoters we carried out ChIP-seq analysis using selected PCR ES Promoter Mapping (ESPM) regions known to have sequence polymorphisms among different BESs, as previously described (López-Farfán et al., 2014). A pool of amplicons of sequences of the promoter regions ESPM1 and ESPM4 (**Fig 9D**, defined in the primer list included in **Table 5**) containing the corresponding sequences from the BES1, BES2, BES3, BES4, BES7, BES7dw, BES10, BES10dw, BES12, BES13, BES15, BES 15dw, BES17,

BES17dw and 18S, U2, C1 as control were combined together in a single lane to build an index file (ebwt), next the alignment was done with bowtie -S -n0 command to consider a 0 nucleotide mismatch to distinguish among few nucleotide sequence differences in each BES, as described before (López-Farfán et al., 2014). The actual number of reads aligned on each BES in represented in the histogram of **Figure 9E.**

7.2.20- Sequential ChIP analysis

For sequential ChIP analysis, non-reversed Immunoprecipitated DNA was diluted 1:5 in ChIP dilution buffer to a final SDS concentration of 0.25% and incubated again with the antibodies of interest.

7.2.21- Luciferase assay

Luciferase assay were carried out using the *Luciferase Assay System* (Promega®) following the manufacturer's instructions. A bloodstream form culture (3×10^6 cells from a BSF culture) were collected, centrifuged (1400xg at 4ºC for 10 min), washed in 500 µl TDB plus 20mM glucose and centrifuged again. Pellets were resuspended in 50 µl of 1x Luciferase cell culture Lysis Reagent (provided in the kit). Renilla reagent was prepared with *Stop and Glo* buffer to 1x final concentration. For each measurement, 50 µl of *LARII* reagent was mixed with 10 µl each lysate for Luciferase reporter detection, then the remaining reacted lysate was incubated with Renilla reagent (1x) for Renilla reporter activity. Lectures were performed in a *FB 12* Single Tube Luminometer (Titertek-Berthold) with pre-stabilised parameters (2 sec of delay time/ 10 sec temp).

7.2.22- FACS (Fluorescent-activated cell sorting) analysis

SNF2PH RNAi bloodstream form induced cultures (1.5×10^7 cells) were collected, centrifuged (1400 xg at 4ºC for 10 min), washed in 5ml 1X PBS and centrifuged again. Pellets were fixed overnight in 1% paraformaldehyde at 4ºC. After incubation the cells were centrifuged and washed in 5ml 1X PBS and resuspended in 500µl 1X PBS. Diluted primary antibodies (in 0.5% blocking reagent) were added for 1h (anti-VSG221 1:3000), followed by Alexa Fluor 488 goat anti-rabbit (Invitrogen) secondary antibody incubation. Propidium iodide (PI) and RNase were added to 40 µg/ml and 10 µg/ml final concentrations, respectively, and incubated 30 min at RT. DNA content of 30.000 cells per sample was analysed with a *FACScan* flow cytometer using the *Cellquest* software (BD Biosciences). Gating was determined with control cells and the same values were used for all treated cells.

7.2.23- Proximity Ligand Assay (PLA)

The PLA assay was performed according to the manufacturer's protocol (Olink Bioscience, Uppsala, Sweden). Bloodstream form cells were dried on slides, fixed 20 min in 4% PFA and

permeabilized 20 min with 1% NP40. After blocking, primary antibodies (rabbit anti-SNF2PH affinity-purified polyclonal (1:1000) and mouse anti-TbSUMO 1C9H8 (1:2000)) were incubated for 1h. Then, the PLA probe solution containing the secondary antibodies conjugated with oligonucleotides (PLA probe MINUS and PLUS) were applied. Secondary species-specific antibodies conjugated with oligonucleotides would hybridize to the two PLA probes if they were in close proximity (<40nm). The slides were then incubated at room temperature for 30 min with the ligation solution together with the ligase that would join the two hybridized oligonucleotides to a closed circle. The amplification mix containing nucleotides and polymerase was applied to the slides. The oligonucleotide arm of one of the PLA probes acts as a primer for rolling circle amplification (RCA) reaction, generating a repeated sequence and extended product. Finally, the nuclei DNA were stained with DAPI and the slides were examined under fluorescent microscopy. As negative control and to discard cross-reaction signal, complete protocol was performed in parallel using primary antibodies separately.

7.2.24- Cell extracts and Immunoblots

Parasites from 20 ml cultures at density of ~1.5 x 10^6 bloodstream or procyclic form cells per ml were collected by centrifugation and washed once in TDB with 1X protease inhibitor cocktail (Roche) and 0-40mM N-ethylmaleimide (NEM, Sigma) to avoid the action of desumoylases. Pellets were resuspended in the same buffer at concentration of 1 x 10^6 cells per µl and further quantified by Bradford protein assay (BioRad). Whole cell lysates were prepared by diluting 1:1 in 2x *Laemmli* sample buffer and heated at 95ºC for 5 min. Samples were loaded (~5 x 10^6 cells per lane or 40 µg of proteins) in 4–20% precast polyacrylamide gels (Bio-Rad Laboratories) and analysed by Western blot using anti-SNF2PH mAb ascites (1:1000), anti-SNF2PH affinity-purified antiserum, monoclonal anti-HA (1:500) high affinity (clone 3F10, Roche Applied Science) and mouse monoclonal anti-Flag (1:5000, Sigma). Mouse monocolonal anti-TbSUMO mAb ascites (1:1000), anti-Tubulin mAb (1:5000), anti-MVP mAb (1:1000), rabbit anti-VSG221 antiserum (1:50000) and rabbit polyclonal anti-p-AMPK (1:1000) (Phospho-AMPKα (Thr172) (40H9) mAb #2535, Cell Signalling technologies) were used as described previously (Barquilla et al., 2012; Lopez-Farfan et al., 2014; Saldivia et al., 2016).

7.2.25- Quantitative WB

For quantitative Western blots total protein extracts from 5x10^5 parasites were loaded (with the exception of the parental dilution series) on a 10% acrylamide Bis-Tris gel. After electrophoresis proteins were transferred to a nitrocellulose membrane. Membrane was incubated with blocking buffer (PBS 1X 0.05% tween (PBS-T), 5% non-fat milk) prior to incubation with rabbit polyclonal antibodies against VSG-221 (1:50.000), SNF2PH (1:1000),

TbTubulin (1:5.000) and anti-TbMVP mAb (1:1000) during 1 hour at RT. To quantify phosphorylated levels, rabbit polyclonal anti-p-AMPK (1:1000) antibody was incubated overnight at 4º. After washing with PBS-T, secondary antibodies goat anti-mouse IgG (H+L) 700 and anti-rabbit IgG (H+L) 800 Dylight were applied for 60 minutes at RT (1:10.000, Thermo-Fisher). Membrane was scanned using an LI-COR Odyssey scanner and analysed using Odyssey IR imaging software 3.0.42. Scan settings were medium image quality, 169 µm resolution and intensity 2.5–5.0 for both channels. Antibody signals were analysed as integrated intensities of regions defined around the blots of interest in either channel. A standard curve based on TbTubulin-normalized anti-TbVSG221 signal intensity was generated using different concentrations of parental cell extracts (R2 = 0.99). The standard curve regression was used to determine VSG221 expression levels in SNF2PH-depleted cell lines. For both detection of AMPK phosphorylated levels and SNF2PH in *in vitro* and *in vivo* assays, an MVP-normalized anti-SNF2PH and/or anti-p-AMPK were used to quantify differences in signal intensity compared with the parental condition.

7.2.26- Expression and purification of recombinant proteins in *E. coli*

7.2.26.1- Sample preparation
For the expression of recombinant proteins, *E. coli* strain KRX was transformed with pET28a plasmid (which includes a 6xHistidine tag in its open reading frame), containing the Gene Of Interest (GOI) in frame. KRX cells were previously grown in a 40 ml of LB media (with kanamycin selection (25µg/ml)) over night at 37ºC to further be inoculated into 2 L (1:50 dilution) until the culture achieves a density of 0.4-0.5 at 600 nm. Then, the culture was induced with 0.5 mM IPTG and 0.01% Rhamnose during 4h at 37ºC (This condition might be variable depending of each protein feature to be expressed). After centrifugation (15 min at 8000 rpm and 4ºC), the remaining pellet was resuspended in approximately 5 ml of Lysis buffer to facilitate the sonication procedure. After sonication, the sample was centrifuged at 12500 rpm during 15 min and diluted 1:2 with Lysis buffer without detergents to allow a 0.5% of final concentration of them, while maintaining the urea concentration (for insoluble proteins). After this, the sample was filtered by 0.4 µM to discard impurities.

7.2.26.2- Column preparation and sample loading
To prepare the column, it was applied 0.5-1 ml of packed Ni-Sepharose FF (GE Healthcare®) (20% Slurry with ethanol as a preservative) into filter columns (BioRad®). Next, the column was washed with 10 column volumes (He et al.) of distilled water followed by 10 ml of lysis buffer to equilibrate it. The crude protein extract was passed through the column, transferring all the

content (protein extract + Ni-Sepharose FF) into a falcon tube to be incubated 1h at 30 min on rotation at RT.

7.2.26.3- *Wash and elution procedure of the recombinant protein*

After the incubation, the extract was passed through the column by gravity flow meanwhile the Ni-Sepharose FF was retained and packed again into the filter column. Next, the column was washed with 40 CV of Lysis buffer to discard nonspecific protein binding into the Ni-Sepharose FF. The elution step was performed in batch, applying 1 ml of Lysis buffer with increased concentration of imidazole (50-500 mM) to compete with nickel charges for histidine tagged protein release. Finally, the eluted protein was dialyzed over night against 5 L of PBS 1x at 4ºC with gentle mix. Then, the concentration of the dialyzed protein was quantified by Bradford at 595 nm and the quality was checked in a 15% polyacrylamide gel after *Coomassie* stain.

7.2.27- Antibody purification

The recombinant protein of interest was immobilized into an *Aminolink* column in coupling buffer, following the manufacturer's instructions (Pierce). The antiserum was centrifuged during 5 min at 2000g to discard any precipitate. Then, the supernatant was incubated 2h at 4ºC with the resin, previously equilibrated with coupling buffer and linked to the recombinant protein. Next, the sample was applied onto the column and allow to flow through the bed resin. Sample and resin were washed with wash/binding buffer by gravity flow until the optic density was under 0.01 at 280 nm. The antibodies were eluted by applying 1 ml of IgG elution buffer (0.15 mM Glycine buffer pH 2.5-3) and neutralized with 50 µl (1:20) of 1M Tris HCl pH 9.0.

7.2.28- *In vitro T. brucei* SUMOylation assay in bacteria

Cloning and purification of recombinant TbSUMO (Tb927.5.3210), TbE2 (Tb927.2.2460), TbE1a (Tb427tmp.02.5410), TbE1b (Tb927.5.3430) and TbSENP (Tb927.9.2220) strategy was established as previously described (Iribarren et al., 2015b). *In vivo* reconstituted SUMOylation system was performed in *Escherichia coli* BL21 (DE3) cells transformed with pCDFDuet-1-TbSUMO/TbE2, followed by pACYCDuet-1-TbE1a-TbE1b. Competent bacteria were transformed again with pET28a-SNF2PHN or pET28a-SNF2PHC. Assessment of SUMOylation reaction in cells containing three previous plasmids were induced grown in Luria-Bertani (LB) medium at 37ºC to an OD_{600nm} of 0.6, and induced with1 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) for 5hr at 37ºC with shaking (250 rpm). Harvested cells were resuspended in lysis buffer and sonicated. Centrifuged samples (70 µl) were resuspended in *Laemmli* sample buffer with 100 mM DTT and boiled for 5 min and analyzed by Western blot.

Deconjugation assays were carried out with 70 µl of supernatants and incubated for 2hr at 28ºC with 70 µl of TbSENP supernatant including or without 25mM N-ethylmaleimide. Separated proteins by SDS-PAGE (7.5 or 10% acrylamide) followed by *Coomassie* Blue staining were transferred to a nitrocellulose Hybond ECL membrane (GE Healthcare) and probed with anti-Flag M2 mouse monoclonal antibody (Sigma, Sant Louis, MO, USA) 1:5000. Horseradish peroxidase-conjugated goat anti-mouse secondary antibody (Sigma) 1:5000 was detected by Chemiluminescence using SuperSignal West Pico Chemiluminiscent Substrate (Pierce, Rockford, II, USA).

7.2.29 - Purification of complexes in *T. brucei*

7.2.29.1- HA Magnetic beads

A total cell mass of 4.0 x 10^{10} procyclic form cells, strain T.b. 449 expressing the 3HA tagged version of SNF2PH was induced with 1µg/ml Doxycycline during 48 h and harvested at 1400 g rpm during 10 min at 4ºC. The pellet is washed in 50 ml PBS 1x including protease inhibitors and subjected to cryogenic grinding, resulting in a lyophilized powder with all nuclear components, as previously described (Obado S.O., 2016). Immunoprecipitation assays were performed with 50 mg of lyophilized powder and resuspended immediately in 1 ml of the buffer of choice, followed by three cycles of sonication of 15 sec at 50 W and centrifuged at 20000g during 10 min at 4ºC. The supernatant containing the nuclear fraction was incubated with 10 µl (0.1 mg) of previously equilibrated HA-Magnetic beads and incubated during 2h at 4ºC on rotation and washed three times, preserving the same buffer conditions. HA-Magnetic beads were eluted at 99ºC during 5 min with 15 µl of NuPAGE SDS Sample buffer (Life technologies) with 1.5 µl of NU PAGE SDS Sample Reducing Agent (Life technologies) and denatured at 99ºC during 5 min prior to be analysed in a SDS-PAGE gel with silver stain.

7.2.29.2- Protein identification by liquid chromatography coupled to tandem mass spectrometry

Sample preparation for MS analysis come from 6 IP assays following the same protocol as described in the previous section but eluting with 50 µl of 2% SDS and 20 mM TrisHCl pH 8.0 with 20 min incubation at 72ºC and precipitated overnight with 100% Ethanol to a final volume of 500 µl. After centrifugation, the final sample was resuspended in 50 µl of 100% Ethanol. Samples were subjected to tryptic digest and reductive alkylation of Cys groups with 50 mM iodoacetamide and finally vacuum-dried to be dissolved in 1% acetic acid. Then, tryptic peptides mixtures were injected onto a C-18 reversed phase nano-column (100 mm ID, 12 cm, Teknokroma) and separated in a continuous acetonitrile gradient. Eluted peptides from the RP nano-column were fragmented in an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo

Scientific). For protein identification, the mass spectra were deconvoluted using MaxQuant version 1.5 searching the *T. brucei*427_927_Tritryp-3.1 annotate protein database (37220 proteins). Search engine was performed assuming the full trypsin digestion (*stricttrypsin*) in Mascot version 2.4.1. with pre-stablished parameters (Fragment Tolerance: 0.60 Da (Monoisotopic) Parent Tolerance: 10.0 PPM (Monoisotopic) Fixed Modifications: +57 on C (Carbamidomethyl) Variable Modifications: -17 on n (Gln->pyro-Glu), +16 on M (Oxidation), +32 on M (Dioxidation), +42 on n (Acetyl)). To visualize MS-spec data we used *Scaffold Proteome Software* version 4.4.6 with a 95% peptide threshold and 99% protein threshold and 2 peptides minimum. False discovery rates (FDR) of 0.7% were calculated for both peptide and protein levels.

7.2.29.3- Purification of SUMO conjugated proteins and further screening

7.2.29.3.1- Histidine coupled to HA Agarose purification for His-HA-tagged SUMO conjugates
To optimize the purification of SUMO conjugated proteins, we first developed a cell line which expressed a polyhistidine (8x) tag plus a HA epitope tagged version of SUMO by replacement of the endogenous gene by two consecutive homologous recombination rounds. Generated cell lines were able to produce tagged SUMO conjugates as the parental cell line. SUMO conjugates were purified from parasites resuspended in 6M urea by nickel affinity chromatography using denaturing conditions to discard the action of SUMO peptidases as well as avoiding noncovalent protein binding of SUMO interaction proteins. Urea concentration was reduced during the elution process of the conjugates and they were subjected to a second affinity purification step using anti-HA agarose beads. Although SUMO conjugates were recovered efficiently, the result showed a notable overlapping between the control and the experimental sample, rising to sample contamination due to an experimental inaccuracy. Nevertheless, the generation of this cell line demonstrated that SUMO functionality is not being affected by the addition of consecutive tags and, for the other side, this method provided a preliminary candidate list described in results.

7.2.29.3.2- Purification of TbSUMOT106K cell line conjugates
To elucidate the experimental constraints of the previous method using the HisHATbSUMO Strain, a new approach was applied in order to optimize the efficiency and detection of SUMO conjugates (Tammsalu et al., 2015), that allows a increment of Lysine residues in target proteins rising to an enrichment of SUMOylated peptides by an anti-diGlycine antibody (K-ε-GG). From the previously generated cell line HisHATbSUMO, the Threonine residue at 106 of the SUMO group was mutated into a Lysine, previous to the diGlycine motif, rising to the generation of the HisHATbSUMOT106K cell line, with the same conjugation ability that the

parental cell line. Parasites were lysed in denaturing buffer and SUMOylated proteins were purified by nickel affinity chromatography. Furthermore, samples were digested with Lys-C to expose a remnant diGlycine residue in the target protein containing the modified SUMO version. Peptides containing the diGly-Lys domain were captured by specific antibodies.

Detailed protocol is described as follows:

Isolated bloodstream parasites (5×10^{10}) from Wistar IGS rats (Charles River) were collected by centrifugation, washed with TDB supplemented with 20 mM NEM, immediately resuspended in lysis buffer for a final concentration of $\sim 4 \times 10^9$ parasites/ml and sonicated until loss of viscosity. Lysates were centrifuged for 30 min at 46.000 x g and filtered by 0.2 µm. Adjusted protein concentration for the cleared supernatant was adjusted up to 50 mg per ml Ni^{2+}-NTA resin (GE Healthcare), preequilibrated with lysis buffer, and incubated overnight at 4ºC with gentle stirring. Then, the resin was washed with 5 column volumes (CV) of Lysis buffer, followed by 10 CV of buffer 1, 10 CV of buffer 2 and 10 CV of buffer 1. Proteins were eluted three times with 2 CV of elution buffer and concentrated by acetone precipitation (80% v/v). Eluted fractions were analyzed by WB to validate the purification efficiency. Enrichment of His-HA-TbSUMOT106K conjugates by immunoaffinity purification with the K-ε-GG-specific antibody after Lys-C or Lys-C/Glu-C digestion was performed according to the published protocol (Tammsalu et al., 2015).

7.2.30- Protein identification (MS/MS) for TbSUMOT106K conjugates
Processed peptides were desalted using equilibrated C18-bonded silica StageTips (Pierce) and vacuum-dried before reconstituting with 0.1% (v/v) of trifluoroacetic acid (TFA). Then, the protein mixtures were separated using a linear acetonitrile gradient (5-90%) in an EASY-nLC 1000 Liquid chromatography system coupled to Q Exactive mass spectrometer via EASY-Spray ion source, including EASY-Spray analytical column (50 cm x 75 µm Ø) and Acclaim PepMap 100 precolumn (2 cm x 75 µm Ø) bypasses. For protein identification, the mass spectra were deconvoluted using MaxQuant version 1.3.0.5 searching the TriTrypDB_8.0_TbruceiLister427 annotate protein database. Search engine was performed in *Andromeda* version 1.5.6.0 with pre-stablished parameters (Fixed Modifications: Carbamidomethyl (C), Variable Modifications: Oxidation (M), Acetylation (protein N-term), Gly-gly (K, not C-term), Phosphorylation [18], Enzyme: Lys-C/P (3 max. missed cleavages), Lys-C/P and Glu-C (5 max. missed cleavages), Special amino acids: K, MS/MS Tolerance: 20 p.p.m, Main search: 6, First search: 20 p.p.m., Top MS/MS peaks per 100 Da: 10). MS-spec data was visualized with a false discovery rate

(FDR) of 0.01% for peptide and protein thresholds with 7 residues as a minimum peptide length.

7.2.31- Bioinformatic analysis (TbSUMO^{T106K} conjugates)

Analysis of consensus SUMO sequences were performed using *pLogo* web server and conserved residues were scaled relative to Bonferroni-corrected statistical significance values, using TriTrypDB-9.0_Tbrucei927 input data set as background. Venn diagram representation was carried out using the web tool *bioinformatics.psb.ugent.be/webtools/Venn* with the BF and PF protein lists as an input data.

7.2.32- Co-immunoprecipitation assays (Co-IP)

7.2.32.1- Starting up conditions for Immunoprecipitation of SNF2PH using denaturing conditions

Bloodstream form cultured cells ($3x10^9$) were resuspended in 1.2 ml Lysis buffer per $2.5x10^9$ cells/ml, followed by a sonication cycle till loose the viscosity (Supernatant 1), and centrifuged at 13500 rpm during 10 min. The extract was diluted 1:6 for a final concentration of urea at 1 M in a volume of 7.2 ml, including protease inhibitors and 20 mM NEM (Supernatant 2). Next, the extract was split into 1.5 ml tubes and added the corresponding detergents to optimize the condition for the IP. The assayed conditions in 1M urea were: 1) 0.1 % NP40, 2) 1 % NP40 + 0.5% Triton X-100, 3) 1 % NP40 + 0.1% SDS, 4) 1 % NP40 + 0.5% Lauryl-Sarcosine, 5) 0.1 % NP40 without urea, 6) No antibody control. For each IP, 60 µg of polyclonal antibody anti-SNF2PH (2 µl at 30 µg/µl) was incubated over night at 4ºC on rotation. Then, the protein extract was incubated with 40 µl of Protein G Sepharose, equilibrated previously with 1 ml of the buffer of choice, during 1h at 4ºC in order to capture the IgGs, followed by 6 washing steps in the same conditions than the incubation buffer. Finally, elution was performed by addition of sample buffer 6x and denaturing at 99ºC during 5 min to further be analyzed by WB using the polyclonal antibody against SNF2PH.

7.2.32.2- Immunoprecipitation assay to determine SNF2PH SUMOylation

Once established the optimal constraint for the polyclonal antibody against SNF2PH in denaturing conditions, the assay was scaled to a larger volume. For each IP, cells were lysed in 1.6 ml lysis buffer per $4x10^9$ cells. Then, the resulting nuclear extracts were carefully diluted 1:6 with dilution buffer, for a final concentration of 1 M urea in a final volume of 8.4 ml. Next, the nuclear extract was split into proportional volumes ($2x10^9$ cells) and incubated with 60 µg of anti-SNF2PH (6 µl at 30 µg/µl) and the same concentration of an unspecific antibody (negative antibody), over night at 4ºC. Finally, 50 µl of packed protein G Sepharose was used to

capture the IgGs and subjected to consecutive washing steps and eluted, to later be analyzed by WB.

7.2.33- Differentiation to procyclic form

Differentiation from slender to insect procyclic form was induced by 3 mM cis-Aconitate (Sigma Aldrich) , with a temperature shift from 37ºC to 28ºC and switching the medium to Differentiating Trypanosome Medium (DTM) as previously described (Barquilla et al., 2012). The Assessment of the differentiation process was monitored by a double IF to detect the expression of the surface glycoproteins using anti-procyclin and anti-VSG221 antibodies.

7.2.34- AMP Analog treatment and obtaining *in vivo* stumpy forms

Parasites in culture at a low density (2x 10^5 cells /ml) were incubated with 8-pCT-2'-O-Me-5'-AMP (1µM) (c078; Biolog Life Science Institute) during 18hr as previously described (Saldivia et al., 2016). To avoid the AMPK activation caused by cell density, the control and treated cells were analyzed at the same cell density. Slender and stumpy forms of pleomorphic AnTat 90.13 were purified from Balb/c mice at 5 days post infection. Blood was collected in heparin containing tubes and centrifuged at 1000 g during 6 min at room temperature. Then cells were washed twice with separation buffer (SB) supplemented with 10 mM glucose. The buffy coat was resuspended in SB-glucose and passed through a diethylaminoethylcelullose DE52 (Whatman) anion exchange chromatographic column, and trypanosomes were collected in the flow-through fraction. Cells were maintained at 37ºC during the process.

8. Appendix

Appendix I: Supplemental tables

Table S1. SUMO-target proteins identified in independent proteomic analyses by tandem affinity purification (histidine/HA). Data from three independent proteomic purifications are represented in separate tables (a, b, c) after background subtraction (control) for a protein and peptide threshold of 95% and 0.5% FDR, respectively. Exclusive number of peptides for the identified substrate is shown.

Table S1. (A) SUMO-target proteins identified by LC-MS/MS in replicate 1.

Accession Number	MW	Identified Proteins (48/280)	HA	Control
Tb927.10.740	155 kDa	structural maintenance of chromosome 4, putative (SMC4)	18	0
Tb927.5.3210	12 kDa	small ubiquitin-related modifier, putative (TbSUMO)	8	0
Tb927.9.5190	32 kDa	proliferative cell nuclear antigen (PCNA), putative	8	0
Tb927.9.13320	55 kDa	hypothetical protein, conserved	7	0
Tb927.7.5310	73 kDa	hypothetical protein, conserved	4	0
Tb927.6.1070	102 kDa	hypothetical protein, conserved	4	0
Tb927.3.4140	62 kDa	hypothetical protein, conserved	3	0
Tb927.7.5090	102 kDa	hypothetical protein, conserved	3	0
Tb927.8.4990	70 kDa	69 kDa paraflagellar rod protein,PFR2 (PFR-B)	3	0
Tb927.11.9920	77 kDa	polyubiquitin, putative	2	0
Tb927.5.4220	11 kDa	histone H4, putative	2	0
Tb927.3.5370	34 kDa	hypothetical protein, conserved	2	0
Tb927.7.3950	107 kDa	RNA-editing 3' terminal uridylyl transferase 1,KRET1 (KRET1)	2	0
Tb927.4.1300	42 kDa	hypothetical protein, conserved	2	0
Tb927.11.3120	75 kDa	nucleolar GTP-binding protein 1 (NOG1)	2	0
Tb927.9.2490	67 kDa	nucleolar protein (NOP66)	2	0
Tb927.6.1250	57 kDa	hypothetical protein, conserved	2	0
Tb927.11.3300	108 kDa	hypothetical protein, conserved	2	0
Tb927.5.1900	77 kDa	hypothetical protein, conserved	2	0
Tb927.7.3820	72 kDa	hypothetical protein, conserved	1	0
Tb927.7.660	53 kDa	hypothetical protein, conserved	1	0
Tb927.8.5850	467 kDa	hypothetical protein, conserved	1	0
Tb927.5.2950	88 kDa	hypothetical protein, conserved	1	0
Tb927.10.16440	29 kDa	variant surface glycoprotein, fragment	1	0
Tb927.1.2470	15 kDa	histone H3, putative	1	0
Tb927.3.5030	80 kDa	KU70 protein (KU70)	1	0
Tb927.4.2080	105 kDa	C2 domain containing protein (CC2D)	1	0
Tb927.11.11840	14 kDa	hypothetical protein, conserved	1	0
Tb927.8.3680	70 kDa	hypothetical protein, conserved	1	0
Tb927.9.10680	17 kDa	hypothetical protein, conserved	1	0
Tb927.7.5530	42 kDa	hypothetical protein, conserved	1	0
Tb927.8.850	85 kDa	hypothetical protein, conserved	1	0
Tb927.3.1820	25 kDa	mitochondrial RNA binding complex 1 subunit (MRB1820)	1	0
Tb927.10.15660	86 kDa	hypothetical protein, conserved	1	0
Tb927.7.940	54 kDa	protein kinase C substrate protein, heavy chain, putative, glucosidase II beta subunit	1	0
Tb927.9.4450	78 kDa	hypothetical protein	1	0
Tb09.v4.0186	49 kDa	variant surface glycoprotein (VSG, pseudogene), putative	1	0
Tb927.7.4940-DECOY	?	Tb927.7.4940-DECOY oligopeptidase...	1	0
Tb927.4.2560	74 kDa	cardiolipin synthetase, putative	1	0
Tb927.10.1270	127 kDa	mismatch repair protein, putative	1	0
Tb927.10.9360	57 kDa	hypothetical protein, conserved	1	0
Tb927.11.4510	197 kDa	hypothetical protein, conserved	1	0
Tb927.8.1770	59 kDa	hypothetical protein, conserved	1	0
Tb927.8.5110	127 kDa	hypothetical protein, conserved	1	0
Tb927.9.13060	58 kDa	hypothetical protein, conserved	1	0
Tb927.10.4080	127 kDa	hypothetical protein, conserved	1	0
Tb927.7.4650	121 kDa	SNF2 DNA repair protein, putative	1	0
Tb927.5.1290	52 kDa	hypothetical protein, conserved	1	0

Table S1 . (B) SUMO- target proteins identified by LC - MS/MS in replicate 2.

Accession Number	MW	Identified Proteins (80/797)	HA	Control
Tb927.8.7590	140 kDa	receptor-type adenylate cyclase GRESAG 4	2	0
Tb927.2.2380	97 kDa	glycosyltransferase (GlcNAc), putative	1	0
Tb927.10.11780	131 kDa	hypothetical protein, conserved	1	0
Tb11.v5.0784	139 kDa	receptor-type adenylate cyclase GRESAG 4, putative	1	0
Tb927.9.15980	43 kDa	nucleoside transporter 1, putative	2	0
Tb927.1.480	161 kDa	leucine-rich repeat protein (LRRP)	1	0
Tb927.5.1450	134 kDa	ATP pyrophosphate-lyase, putative; adenylyl cyclase	1	0
Tb927.7.3860	27 kDa	PQ loop repeat, putative	1	0
Tb927.4.470	22 kDa	snoRNP protein gar1	1	0
Tb927.11.2510	146 kDa	NUC173 domain containing protein	1	0
Tb927.11.5020	187 kDa	hypothetical protein	1	0
gi|261327151	25 kDa	hypothetical protein	1	0
Tb927.6.3520	74 kDa	Uncharacterised ACR, YagE family COG1723	1	0
Tb927.7.1010	20 kDa	hypothetical protein, conserved	1	0
Tb927.11.4380	72 kDa	ATP-dependent DEAD/H RNA helicase, putative	1	0
Tb927.6.4430	36 kDa	homoserine kinase	1	0
Tb427tmp.01.2430	486 kDa	hypothetical protein, conserved	1	0
Tb927.5.2520	104 kDa	TPR repeat, putative	1	0
Tb927.9.10670	67 kDa	bardet-biedl syndrome 1 protein	1	0
Tb927.10.10770	67 kDa	Generative cell specific 1 protein, putative	1	0
Tb927.11.3130	71 kDa	glycosomal transporter (GAT2)	1	0
Tb927.7.530	171 kDa	FYVE zinc finger, putative	1	0
Tb927.3.2140	**105 kDa**	**transcription activator, putative**	**1**	**0**
Tb927.11.9890	63 kDa	signal recognition particle receptor alpha subunit, putative	1	0
TM35_000231200	10 kDa	hypothetical protein	1	0
Tb927.1.4520	83 kDa	Zinc finger, C3HC4 type (RING finger), putative	1	0
Tb927.3.2790	125 kDa	leucine-rich repeat protein (LRRP)	1	0
Tb927.5.520	56 kDa	stomatin-like protein, putative	1	0
Tb927.9.210	53 kDa	variant surface glycoprotein (VSG, pseudogene), putative	1	0
Tb927.10.13490	31 kDa	CAMK/CAMKL family protein kinase, putative	1	0
Tb927.5.3830	34 kDa	dihydroorotate dehydrogenase	1	0
Tb927.11.14780	46 kDa	phosphomannose isomerase	1	0
Tb927.2.2400	85 kDa	glycosyltransferase (GlcNAc)	1	0
Tb927.10.11020	105 kDa	DNA mismatch repair protein MSH2	1	0
Tb927.2.2650	368 kDa	hypothetical protein, conserved	1	0
Tb927.8.4350	114 kDa	hypothetical protein, conserved	1	0
Tb927.11.10190	133 kDa	telomerase reverse transcriptase, putative	1	0
Tb927.11.16340	125 kDa	Importin-beta N-terminal domain/HEAT-like repeat	1	0
Tb10.v4.0134	54 kDa	variant surface glycoprotein (VSG), putative	1	0
Tb927.2.3280	48 kDa	65 kDa invariant surface glycoprotein	1	0
Tb927.10.13900	39 kDa	UDP-galactose transporter, putative	**1**	0
Tb927.11.3550	32 kDa	XPA-interacting protein, putative	1	0
Tb927.10.9890	97 kDa	hypothetical protein, conserved	1	0
Tb927.10.1610	16 kDa	hypothetical protein, conserved	1	0
Tb927.6.430	138 kDa	receptor-type adenylate cyclase GRESAG 4, putative	1	0
Tb927.11.5740	109 kDa	formin, putative; formin-like protein	1	0
Tb927.10.13090	66 kDa	PhoD-like phosphatase, putative	1	0
Tb927.7.4130	256 kDa	hypothetical protein, conserved	1	0
Tb927.11.5070	93 kDa	GPI ethanolamine phosphate transferase 3, putative	1	0
Tb927.10.2880	304 kDa	Voltage-dependent calcium channel subunit, putative	1	0
Tb927.8.3940	63 kDa	hypothetical protein, conserved	1	0
Tb927.2.2920	49 kDa	UAA transporter family, putative	1	0
Tb927.1.4540	58 kDa	hypothetical protein, conserved	1	0
Tb927.3.1350	87 kDa	hypothetical protein, conserved	1	0
Tb927.3.1860	83 kDa	ATP-grasp domain containing protein	1	0
Tb927.6.1580	53 kDa	polynucleotide kinase 3'-phosphatase, putative	1	0
Tb927.6.4100	94 kDa	hypothetical protein, conserved	1	0
Tb927.7.1470	12 kDa	Mitochondrial ATP synthase subunit c-3	1	0
Tb927.7.2390	103 kDa	Tripartite attachment complex protein 102	1	0
Tb927.11.18060	28 kDa	variant surface glycoprotein (VSG), (fragment)]	1	0
Tb927.8.3840	80 kDa	hypothetical protein, conserved	1	0
Tb927.9.3860	80 kDa	Dyggve-Melchior-Clausen syndrome protein	1	0
Tb927.9.4850	53 kDa	leucine-rich repeat protein (LRRP), putative	1	0
Tb927.10.7580	111 kDa	hypothetical protein, conserved	1	0
Tb927.11.5020	187 kDa	hypothetical protein, conserved	1	0
Tb927.11.6580	77 kDa	hypothetical protein, conserved	1	0
Tb927.11.6830	109 kDa	Domain of unknown function (DUF2779)	1	0
Tb927.11.6920	173 kDa	hypothetical protein, conserved	1	0

Tb927.11.9750	73 kDa	Protein of unknown function (DUF498/DUF598)	1	0
Tb927.11.16210	81 kDa	cAMP response protein 1, putative	1	0
Tb927.10.15080	180 kDa	WD40 Repeat 1	1	0
Tb927.5.2930	43 kDa	Mitochondrial ATP synthase subunit	1	0
Tb927.6.1910	57 kDa	hypothetical protein	1	0
Tb927.7.1510	69 kDa	pseudouridylate synthase I	1	0
Tb927.7.5440	217 kDa	hypothetical protein	1	0
Tb927.8.4400	76 kDa	hypothetical protein	1	0
Tb927.7.4400	83 kDa	inositol polyphosphate kinase-like protein, putative	1	0
Tb927.5.2750	48 kDa	Alpha/beta hydrolase family, putative	1	0
Tb927.11.14250	59 kDa	t-complex protein 1 subunit epsilon	1	0
Tb927.11.7540	34 kDa	electron-transfer-flavoprotein alpha polypeptide	1	0

Table S1. (C) SUMO-target proteins identified by LC - MS/MS in replicate 3.

Accession Number	MW	Identified Proteins (66/243)	HA	Control
Tb927.8.5580	526 kDa	hypothetical protein, conserved	3	0
Tb927.4.4130	100 kDa	hypothetical protein, conserved	3	0
Tb927.10.4170	292 kDa	hypothetical protein, conserved	2	0
Tb927.10.8750	39 kDa	GTPase activating protein, putative	2	0
Tb927.11.11900	98 kDa	coatomer gamma subunit, putative	2	0
Tb927.3.4030	135 kDa	hypothetical protein, conserved	2	0
Tb927.10.13900	39 kDa	UDP-galactose transporter	2	0
Tb927.3.4020	256 kDa	phosphatidylinositol 4-kinase alpha, putative	2	0
Tb927.4.840	46 kDa	hypothetical protein, conserved	2	0
Tb927.2.3370	44 kDa	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative	2	0
Tb927.8.1130	158 kDa	protein phosphatase with EF-Hand domains, putative	2	0
Tb927.2.3320	48 kDa	65 kDa invariant surface glycoprotein	2	0
Tb927.2.4950	156 kDa	hypothetical protein, conserved	2	0
Tb927.4.3870	135 kDa	receptor-type adenylate cyclase GRESAG 4, putative	2	0
Tb927.3.2140	**105 kDa**	**transcription activator**	**2**	**0**
Tb927.10.12890	91 kDa	bifunctional aminoacyl-tRNA synthetase, putative	2	0
Tb927.11.4380	248 kDa	ATP-dependent RNA helicase, putative	2	0
Tb09.v4.0017 (+1)	31 kDa	rab1 (small gtp-binding protein rab1, putative)	2	0
Tb927.11.10670	134 kDa	glycosyl hydrolase, putative	2	0
Tb927.11.5060	120 kDa	hypothetical protein, conserved	2	0
Tb927.8.2450	34 kDa	hypothetical protein, conserved	2	0
Tb927.11.15470	71 kDa	methionyl-tRNA formyltransferase	2	0
Tb927.5.2420	103 kDa	hypothetical protein, conserved	2	0
Tb927.8.3190	33 kDa	hypothetical protein, conserved	2	0
Tb927.8.4500	85 kDa	eukaryotic translation initation factor 4 gamma, putative	2	0
Tb09.v4.0016	54 kDa	expression site- associated gene (ESAG) protein, putative	2	0
Tb927.3.4960	182 kDa	kinesin, putative	2	0
Tb927.10.6050	191 kDa	clathrin heavy chain (CHC)	2	0
Tb927.10.7350	198 kDa	hypothetical protein, conserved	2	0
Tb927.9.8040	98 kDa	hypothetical protein, conserved	2	0
Tb927.11.18230	50 kDa	variant surface glycoprotein (VSG), putative	2	0
Tb927.8.6970	74 kDa	3-methylcrotonyl-CoA carboxylase alpha subunit, putative	1	0
Tb927.6.3150	501 kDa	Hydin	1	0
Tb927.10.13280	286 kDa	hypothetical protein, conserved	1	0
Tb927.11.1160	123 kDa	hypothetical protein, conserved	1	0
Tb927.7.1570	156 kDa	hypothetical protein, conserved	1	0
Tb927.11.9890	64 kDa	signal recognition particle receptor alpha subunit, putative	1	0
Tb927.8.3940	61 kDa	hypothetical protein, conserved	1	0
Tb927.10.9490	45 kDa	expression site-associated gene 3 (ESAG3)-like protein	1	0
Tb927.5.1560	84 kDa	ATP-dependent DEAD/H RNA helicase, putative	1	0
Tb927.10.1890	178 kDa	calpain, putative, cysteine peptidase, Clan CA, family C2, putative	1	0
Tb927.2.2400	85 kDa	hypothetical protein, conserved	1	0
Tb927.4.5530	51 kDa	variant surface glycoprotein (VSG), putative	1	0
Tb927.11.7290	162 kDa	pantothenate kinase subunit, putative	1	0
Tb927.10.6300	50 kDa	hypothetical protein, conserved	1	0
Tb927.7.1780	26 kDa	Adenine phosphoribosyltransferase, putative	1	0
Tb927.9.16250	53 kDa	variant surface glycoprotein (VSG, atypical), putative	1	0
Tb927.5.2660	98 kDa	hypothetical protein, conserved	1	0
Tb927.3.3150	87 kDa	hypothetical protein, conserved	1	0
Tb927.4.1060	124 kDa	hypothetical protein, conserved	1	0
Tb927.6.4430	36 kDa	homoserine kinase, putative (HK)	1	0
Tb927.9.10530	119 kDa	hypothetical protein, conserved	1	0
Tb927.11.840	87 kDa	hypothetical protein, conserved	1	0
Tb927.8.8140	68 kDa	small GTP-binding rab protein, putative	1	0
Tb927.7.3330	503 kDa	hypothetical protein, conserved	1	0
Tb927.7.3770	61 kDa	hypothetical protein, conserved	1	0
Tb927.9.11410	15 kDa	unspecified product	1	0
Tb927.10.3970	55 kDa	hypothetical protein, conserved	1	0
Tb927.11.12630	57 kDa	hypothetical protein, conserved	1	0
Tb927.10.11420	81 kDa	hypothetical protein, conserved	1	0
Tb927.8.2890	46 kDa	hypothetical protein, conserved	1	0
Tb927.5.1320	111 kDa	hypothetical protein, conserved	1	0
Tb927.5.2920	52 kDa	hypothetical protein, conserved	1	0
Tb10.v4.0042	25 kDa	hypothetical protein, conserved	1	0
Tb927.10.1610	16 kDa	hypothetical protein, conserved	1	0
Tb927.3.2210	69 kDa	hypothetical protein, conserved	1	0

Table S2. Proteins associated with SNF2PH. Partners were selected by a high number of peptides identified by LC-MS/MS. Contaminant proteins from negative control sample were subtracted from the list.

Accession Number	Description	Number of peptides
Tb927.3.2140	Transcription activator, putative	68
Tb927.1.1380	Serine/threonine protein phosphatase 2A regulatory subunit, putative	14
Tb927.3.1300	Hypothetical protein, conserved (posttranscriptional regulation of gene expression)	12
Tb927.7.6810	EF-hand domain pair, putative	9
Tb927.3.5620	Facilitates chromatin transcription complex (FACT) subunit spt16	8
Tb927.07.2680	Zinc finger protein family member, putative (ZC3H22)	7
Tb927.10.4070	Hypothetical protein, conserved	6
Tb927.2.5240	pre-mRNA splicing factor 19 (TbPRP19)	4
Tb927.10.4440	Predicted SAP domain protein	4
Tb927.11.4910	Predicted ankyrin repeat family protein	4
Tb927.11.12790	Ribonucleoside-diphosphate reductase small chain (RNR2)	4
Tb927.11.5650	Replication factor C, subunit 1, putative, replication factor C	4
Tb927.3.2440	AGC essential kinase 1 (AEK1)	4
Tb927.8.1990	Peroxidoxin	4
Tb927.7.4810	HD domain containing protein, putative	3
Tb927.3.1240	Protein phosphatase 2A, putative	3
Tb927.9.9230	Hypothetical protein with HD Motif	3
Tb927.3.1900	Hypothetical protein, conserved (Midasin AAA ATPase)	2
Tb927.10.540	ATP-dependent RNA helicase SUB2, putative	2
Tb927.11.13090	Elongation factor 1 gamma, putative	2
Tb927.11.4900	WD40/YVTN repeat-like-containing protein	2
Tb927.9.5190	Proliferative cell nuclear antigen (PCNA), putative	2
Tb927.10.15180	Nucleosome assembly protein	2
Tb927.11.630	RNA polymerase I second largest subunit (RPA135)	2
Tb927.4.5020	RNA polymerase IIA largest subunit (RPB1)	2
Tb927.11.8310	Class I transcription factor A, subunit 4 (CITFA-4)	2

Table S3. Genes expressed after SNF2PH depletion. Representative values are expressed as a log fold change (logFC) and log for count per million (logCPM).

Table S3. (A) Displayed genes were identified from two biological replicates with a FDR<0.05.

Gene	Length	logFC	logCPM	FDR	Description
Tb427.03.2140	2847	-1.73	6.18	1.54E-45	Transcription activator
Tb427.07.2820	405	0.55	7.78	3.64E-05	histone 2A, putative
Tb427.06.510	345	1.08	4.77	1.38E-03	GPEET2 procyclin precursor
Tb427_09_v4.snoRNA.0026	96	1.52	1.77	4.20E-03	C/D snoRNA
Tb427.01.4710	537	1.16	4.24	8.90E-03	Hypothetical protein, conserved
Tb427.BES40.15	2031	0.42	8.50	8.90E-03	Expression site-associated gene 8 (ESAG8) protein, putative
Tb427.10.10910	2115	-0.32	11.87	1.17E-02	Unspecified product
Tb427.10.10240	1218	1.28	1.66	1.40E-02	Procyclin-associated gene 1 (PAG1) protein
Tb427.10.10260	426	0.95	4.52	1.94E-02	EP1 procyclin
Tb427.BES40.18	1413	0.37	9.94	1.94E-02	Expression site-associated gene 2 (ESAG2) protein, putative
Tb427.BES40.14	1893	0.39	8.73	1.94E-02	Expression site-associated gene 8 (ESAG8) protein, putative
Tb427.10.10230	1113	1.22	1.73	2.35E-02	Procyclin-associated gene 5 (PAG5) protein
Tb427.10.7160	621	1.20	0.06	2.46E-02	Procyclin-associated gene 1 (PAG1) protein
Tb427.10.10210	1122	1.17	1.30	2.67E-02	Procyclin-associated gene 4 (PAG4) protein
Tb427.10.10250	390	1.02	4.10	2.91E-02	EP2 procyclin
Tb427.10.10960	2115	-0.28	12.19	2.97E-02	Unspecified product
Tb427.10.10970	2115	-0.27	12.56	3.42E-02	heat shock protein, putative

Table S3. (B) Displayed genes were identified from two biological replicates with a FDR<0.03.

Gene ID	Length	LogFC	LogCPM	FDR	Gene Description
Tb427.03.2580	789	1,17	0,82	5,58E-02	Hypothetical protein
Tb427.10.480	1302	-1,10	0,45	5,76E-02	Hypothetical protein
Tb427.08.6650	2001	-1,09	-0,02	6,79E-02	RNA-binding protein, putative
Tb427.10.10890	21,15	-0,35	8,12	7,73E-02	Heat shock protein
Tb427.10.3400	1362	-1,06	0,18	8,37E-02	Hypothetical protein
Tb427.10.10940	2115	-0,34	8,23	8,38E-02	Heat shock protein
Tb427.10.10980	2115	-0,27	11,51	8,82E-02	Heat shock protein
Tb427.BES65.13	1530	1,04	-0,13	1,15E-01	Variant surface glycoprotein
Tb427.10.7090	990	-0,39	6,45	1,15E-01	Alternative oxidase
Tb427.10.10930	2115	-0,30	10,40	1,19E-01	Heat shock protein
Tb427.BES153.14	1530	1,02	-0,11	1,19E-01	Variant surface glycoprotein
Tb427.10.10920	2115	-0,30	10,17	1,36E-01	Unspecified product
Tb427.08.8330	2667	-0,34	7,30	1,36E-01	Calpain
Tb427.07.5930	1791	0,33	7,25	1,73E-01	Protein Associated with Differentiation (PAD1)
Tb427.10.10950	2115	-0,29	10,00	1,73E-01	Unspecified product
Tb427.03.1500	1101	0,47	5,65	2,05E-01	Variant surface glycoprotein (VSG)-related, putative
Tb427.07.5940	1818	0,32	7,28	2,10E-01	Protein Associated with Differentiation (PAD2)
Tb427.07.180	1314	0,33	7,09	2,10E-01	Trypanosomal VSG domain containing protein, putative
Tb427.10.10900	2115	-0,31	8,69	2,10E-01	Unspecified product
Tb427tmp.01.3810	480	0,95	0,12	2,15E-01	Hypothetical protein
Tb427.10.8490	1590	-0,31	7,90	2,81E-01	Glucose transporter
Tb427.10.11220	1365	0,45	5,60	2,98E-01	Procyclic form surface glycoprotein

Table S4. Quantitative enrichment profile of SNF2PH by ChIP-seq. Representative peaks identified considering 0-missmatch error and FDR<0.05 from significant background genes (not shown).

Chr	FE	GeneName	Description	Chromosome
Tb427_01_v4	45.849	Tb427.01.3760	unspecified product	Chr. I
Tb427_02_v4	7.48829	Tb427.02.1380	leucine-rich repeat protein	Chr. II
Tb427_03_v4	25.5176	Tb427.03.3410	aspartyl aminopeptidase	Chr. III
Tb427_03_v4	25.2224	Tb427.03.3447	rRNA small subunit	
Tb427_03_v4	3.62822	Tb427.03.2140	SNF2PH	
Tb427_03_v4	6.54444	Tb427.03.4393	tRNAs	
Tb427_04_v4	23.0135	Tb427.04.3510	DNA-directed RNApol sub	Chr. IV
Tb427_04_v4	33.9863	Tb427.04.3193	tRNA	
Tb427_04_v4	1.87372	Tb427.04.3740	flagellar attachment zone protein	
Tb427_05_v4	3.07073	Tb427.05.2765	tRNAs	Chr. V
Tb427_06_v4	73.4494	Tb427.06.510	GPEET2 procyclin precursor	Chr. VI
Tb427_07_v4	23.8057	Tb427.07.6881	18S ribosomal RNA	Chr. VII
Tb427_07_v4	17.6557	Tb427.07.6821	tRNAs	
Tb427_08_v4	44.4233	Tb427.08.2856	tRNAs	Chr. VIII
Tb427_08_v4	2.22606	Tb427.08.7740	amino acid transporter	
Tb427_09_v4	48.7038	Tb427tmp.211.0620	actin A	Chr. IX
Tb427_09_v4	33.8135	Tb427_09_SLRNA_0028	SL RNA	
Tb427_10_v5	95.4136	Tb427.10.13320	DNA-directed RNA pol II sub	Chr. X
Tb427_10_v5	6.96547	Tb427.10.10260	EP1 procyclin	
Tb427_10_v5	4.23325	Tb427.10.15350	H3V	
Tb427_10_v5	26.4951	Tb427_10_tRNA_Asn_1	tRNAs	
Tb427_11_01_v4	10.116	Tb427tmp.01.7990	myosin	Chr. XI
Tb427_11_01_v4	18.134	Tb427_11_tRNA_Ala_1	tRNAs	

Table S5. Telomeric BES expression by RNA-seq analysis of SNF2PH depleted cells in *T. brucei* bloodstream forms.
Expression analysis performed in paired group of individual clones showed variability in which VSG is expressed in a given time.

Gene	logFC	Chr	Description	VSG
Tb427.03.2140	**-1.047460197**	**Tb427_03_v4**	**Transcription activator putative**	
Tb427.BES64.2	0.946502267	Tb427_telo64_v1	variant surface glycoprotein (VSG)	BES8/TAR64 **VSG-14**
Tb427.BES40.22	-0.496052939	Tb427_telo40_v2	variant surface glycoprotein (VSG)	BES1/TAR40 **VSG221**
Tb427.BES29.9	-0.519139056	Tb427_telo29_v3	variant surface glycoprotein (VSG)	BES12/TAR29 **VSG1.8**
Tb427.BES126.15	-0.449408386	Tb427_telo126_v2	variant surface glycoprotein (VSG)	BES15/TAR126 **VSGbR-2**
Tb427.BES134.6	0.50671955	Tb427_telo134_v1	variant surface glycoprotein (VSG)	BES10/TAR134 **VSG427-15**
Tb427.BES56.13	-0.430680755	Tb427_telo56_v2	variant surfave glycoprotein (VSG)	BES13/TAR56 **VSGJS1**
Tb427.BES98.12	-0.128764621	Tb427_telo98_v2	variant surface glycoprotein (VSG)	BES5/TAR98 **VSG800**

Table S6. RNA-seq analysis of SNF2PH ectopic expression in *T. brucei* procyclic forms.

Table S6. (A) Representative differentially expressed genes (FDR<0.05) upon ectopic expression of full-length SNF2PH in procyclic forms.

Gene	logFC	FDR	Chr	Description	Associated VSG
Tb427.03.2140	3.00	2.03E-11	Tb427_03_v4	transcription activator putative	---
13J3.02	1.75	1.50E-08	TP13J3	folate transporter putative	---
H25N7.15	2.75	1.54E-08	TPH25N7	transferrin-binding protein putative	---
Tb427.BES40.3	2.71	1.54E-08	Tb427_telo40_v2	transferrin-binding protein putative	VSG 221
Tb427.BES126.1	1.61	1.54E-08	Tb427_telo126_v2	folate transporter putative	VSG bR2
Tb427.BES29.1	1.58	1.54E-08	Tb427_telo29_v3	folate transporter putative	VSG MITat 1.8
Tb427.BES5.2	2.47	1.65E-08	Tb427_telo3_v2	transferrin-binding protein putative	VSG T3
Tb427.BES56.1	1.71	1.65E-08	Tb427_telo56_v2	folate transporter putative	VSG JS1
Tb427.BES59.1	1.75	1.71E-08	Tb427_telo59_v2	folate transporter putative	VSG MITat 1.13
Tb427.BES98.2	2.54	1.95E-08	Tb427_telo98_v2	transferrin-binding protein putative	VSG 800
Tb427.BES15.1	1.75	1.95E-08	Tb427_telo15_v2	folate transporter putative	VSG 121
Tb427.BES51.1	1.77	2.59E-08	Tb427_telo51_v2	folate transporter putative	VSG Bn-2
Tb427.BES28.4	2.12	2.81E-08	Tb427_telo28_v2	transferrin-binding protein putative	VSG T3
Tb427.08.480	1.17	2.81E-08	Tb427_08_v4	phosphatidic acid phosphatase protein putative	---
N19B2.165	3.27	5.86E-08	TPN19B2	transferrin-binding protein putative	---
Tb427.BES10.1	2.48	8.91E-08	Tb427_telo10_v1	transferrin-binding protein putative	VSG MITat 1.19
Tb427.BES10.2	2.60	1.32E-07	Tb427_telo10_v1	transferrin-binding protein putative	VSG MITat 1.19
Tb427.BES122.2	2.85	1.64E-07	Tb427_telo122_v2	expression site-associated gene 6 (ESAG6) protein putative	---
Tb427.BES129.4	3.04	1.77E-07	Tb427_telo129_v2	transferrin-binding protein putative	VSG VO2
Tb427.BES153.3	2.64	1.80E-07	Tb427_telo153_v2	transferrin-binding protein putative	VSG 224
Tb427.BES56.2	2.59	1.93E-07	Tb427_telo56_v2	transferrin-binding protein putative	VSG JS1
Tb427.BES122.1	2.45	1.99E-07	Tb427_telo122_v2	transferrin-binding protein putative	---
Tb427.BES40.5	5.24	2.51E-07	Tb427_telo40_v2	expression site-associated gene 3 (ESAG3 pseudogene) putative	VSG 221
Tb427.BES129.3	2.42	4.39E-07	Tb427_telo129_v2	transferrin-binding protein putative	VSG VO2
Tb427.BES126.2	2.82	5.88E-07	Tb427_telo126_v2	transferrin-binding protein putative	VSG bR2
Tb427.02.3300	2.29	2.52E-06	Tb427_02_v4	65 kDa invariant surface glycoprotein	----
Tb427.02.3270	1.98	3.04E-06	Tb427_02_v4	65 kDa invariant surface glycoprotein	----
Tb427.02.3280	1.97	3.04E-06	Tb427_02_v4	65 kDa invariant surface glycoprotein	----

Table S6. (B) Differentially expressed genes (FDR<0.05) upon ectopic expression of SNF2PH truncated isoform in procyclic forms.

| Gene	logFC	FDR	Chr	Description
Tb427.03.2140	**3.29**	**1.95E-41**	**Tb427_03_v4**	**transcription activator putative**
Tb427.05.4260	3.57	9.16E-30	Tb427_05_v4	histone H4 putative
Tb427.05.440	0.72	8.09E-21	Tb427_05_v4	trans-sialidase putative
13J3.08	0.54	4.89E-18	TP13J3	hypothetical protein
Tb427.03.590	-0.70	4.89E-18	Tb427_03_v4	adenosine transporter putative
Tb427.07.190	0.54	5.44E-17	Tb427_07_v4	metallo- peptidase Clan MA(E) Family M3 putative
Tb427.06.520	0.52	6.35E-16	Tb427_06_v4	EP3-2 procyclin
Tb427.05.4020	0.49	1.08E-15	Tb427_05_v4	hypothetical protein
Tb427.10.1500	0.46	1.08E-15	Tb427_10_v5	methionyl-tRNA synthetase putative
Tb427.07.6850	0.46	1.08E-15	Tb427_07_v4	trans-sialidase
Tb427.08.7640	0.58	1.24E-15	Tb427_08_v4	amino acid transporter 1 putative
Tb427.08.8300	0.43	1.01E-14	Tb427_08_v4	amino acid transporter putative
Tb427tmp.02.3770	0.41	1.23E-14	Tb427_11_01_v4	hypothetical protein conserved
Tb427.03.4500	-0.38	4.76E-14	Tb427_03_v4	fumarate hydratase class I
Tb427tmp.02.0100	0.40	6.88E-14	Tb427_11_01_v4	carboxypeptidase putative
Tb427.08.7610	0.54	2.98E-13	Tb427_08_v4	amino acid transporter 1 putative
Tb427.01.2230	0.33	5.52E-12	Tb427_01_v4	unspecified product
Tb427.07.4570	-0.33	1.37E-11	Tb427_07_v4	inosine-guanine nucleoside hydrolase
Tb427tmp.02.3440	-0.41	1.70E-11	Tb427_11_01_v4	hypothetical protein conserved
Tb427.BES10.0	0.39	2.40E-11	Tb427_telo10_v1	beta lactamase
Tb427tmp.211.1800	-0.33	3.46E-11	Tb427_09_v4	hypothetical protein conserved
Tb427.05.470	-0.38	5.92E-11	Tb427_05_v4	monocarboxylate transporter-like protein
Tb427.10.10250	0.35	9.28E-11	Tb427_10_v5	EP2 procyclin
Tb427.07.7090	0.34	1.93E-10	Tb427_07_v4	hypothetical protein conserved
Tb427tmp.01.5310	-0.29	1.93E-10	Tb427_11_01_v4	receptor-type adenylate cyclase GRESAG 4 putative
Tb427.10.9080	-0.34	3.17E-10	Tb427_10_v5	pteridine transporter putative
Tb427.07.5970	-0.33	3.91E-10	Tb427_07_v4	hypothetical protein conserved
Tb427.08.1610	0.29	5.88E-10	Tb427_08_v4	MSP-B putative

Appendix II: Supplemental figures
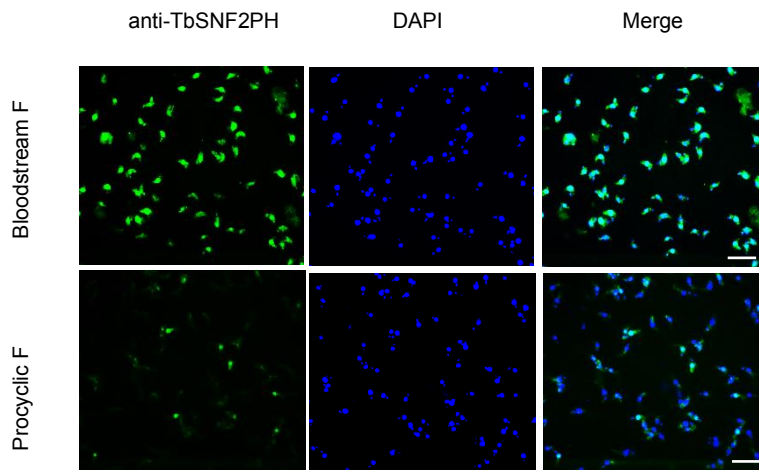
Figure S1. Study by IF of SNF2PH in BSF and PF. Panels show DAPI stain and FITC channel using the anti-SNF2PH mAb and an anti-Mouse IgG Alexa Fluor 488 secondary Ab. Scale bars, 15 µm. SNF2PH is differentially expressed in both developmental stages (p<0.001).
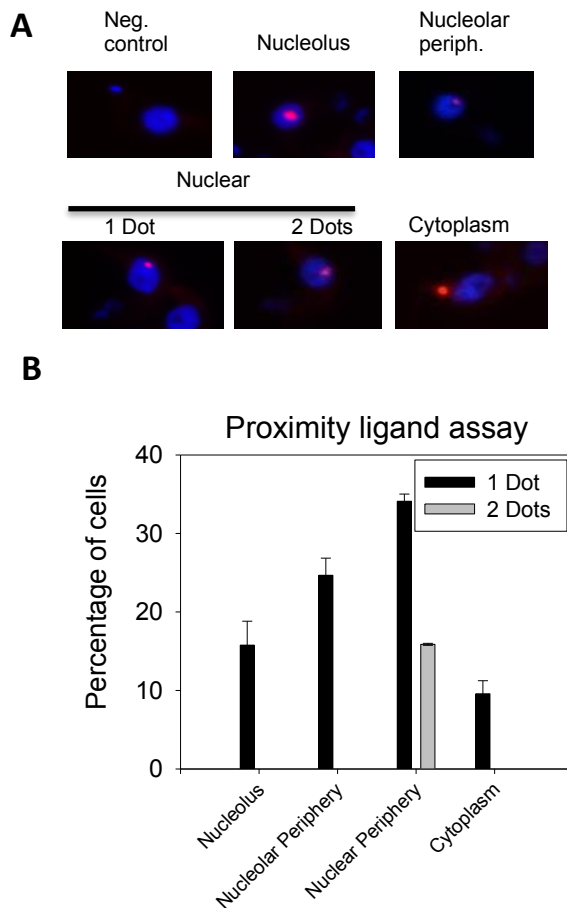

Figure S2. *In situ* detection of SUMOylated SNF2PH using a Proximity Ligation Assay (PLA). (A) SNF2PH SUMOylation was detected by *in situ* PLA (red dot) in bloodstream form cells using rabbit anti-SNF2PH antiserum and mouse anti-TbSUMO mAb. SNF2PH appears to be SUMOylated in both nucleolus and nuclear periphery (1 or 2 Dots). Dot in cytoplasm indicates an artefactual reaction. **(B)** Histogram showing the percentage of positive amplification signal detected in each compartment. Positive cells = 22.2±5.5% (n=2).
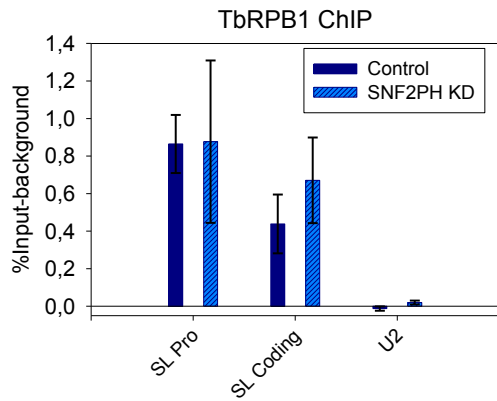
Figure S3. ChIP analysis in SNF2PH depleted cells. Partial reduction of SNF2PH (48h) and parental cell line (DRALI) shows that TbRPB1 is not reduced at both Spliced leader promoter (SL Pro) and coding region (SL Coding). U2 used as a Pol III negative locus. Depletion of SNF2PH does not affect to RNA Pol II recruitment at Pol II-related loci. Data from at least three independent assays. Standard errors are represented as a percentage of input-background after background subtraction of the pre-bleed antiserum.
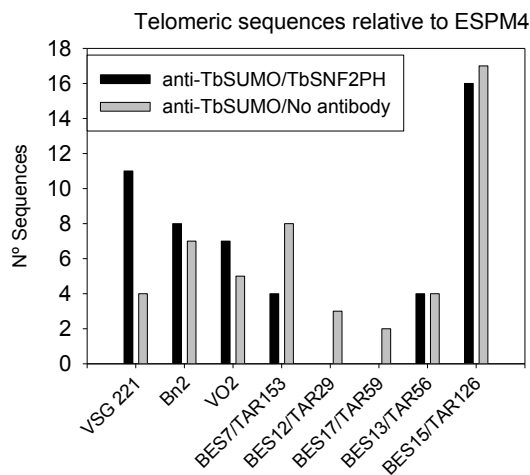

Figure S4. Analisys of ESPM-4 fragments form telomeric sequences occupy by SNF2PH. Histogram shows the number of cloned ESPM-4 fragments form telomeric sequences after SNF2PH immunoprecipitation, previously enriched with TbSUMO immunoprecipitated sequences. No antibody control was used to distinguish unspecific cloned sequences. (n=50 per assayed condition).
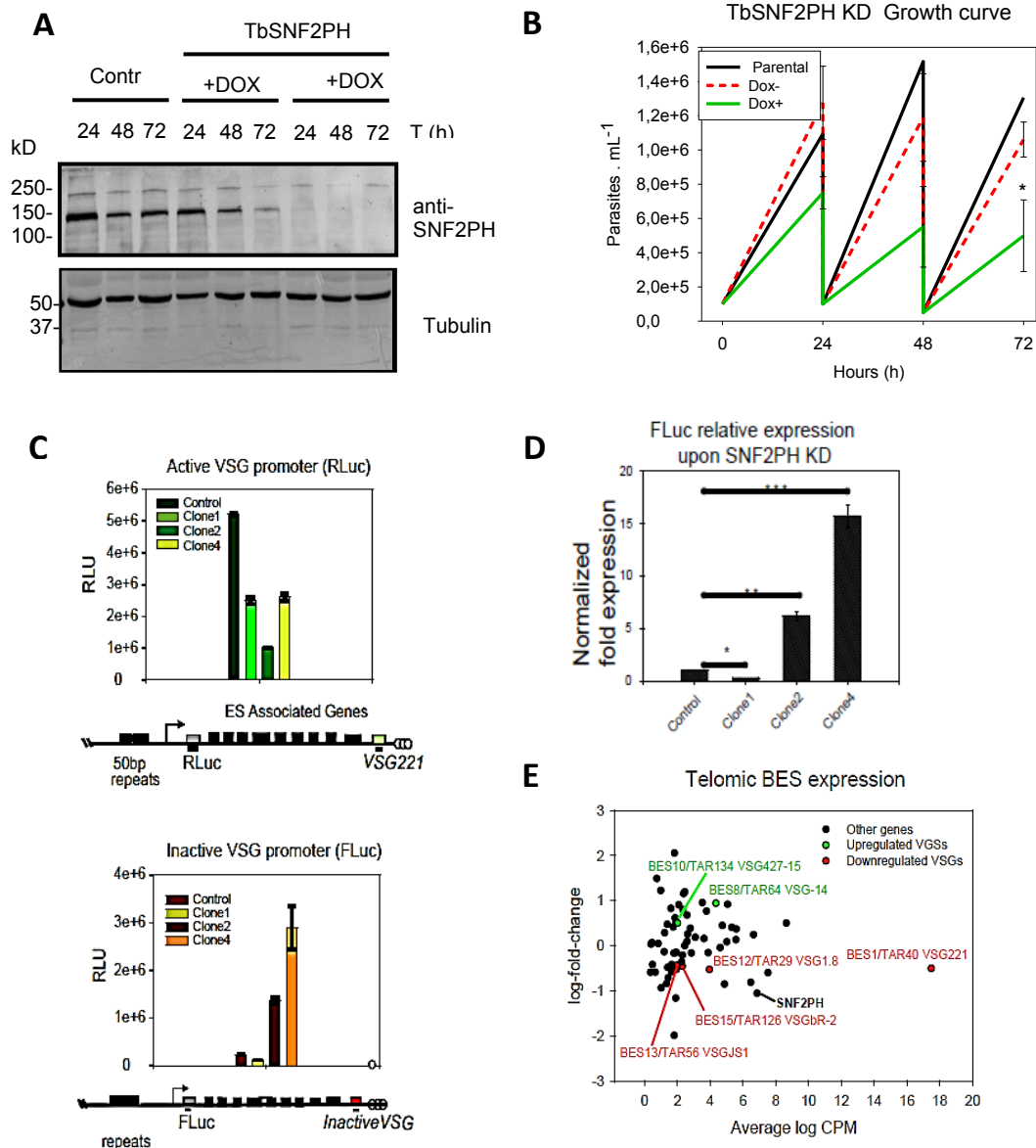
Figure S5. SNF2PH depletion features and expression profile. (A) Time curse knock down of SNF2PH in Bloodstream form. Rabbit anti-SNF2PH antiserum was used to monetarize protein depletion by western blot analysis. **(B)** SNF2PH depletion reduced growth of bloodstream form cells. Using inducible RNAi, two independent clones showed an abrupt growth arrest upon 72h of doxycycline induction. Black line: Parental cell line (Parental). Discontinuous red line: Uninduced SNF2PH RNAi (dox-). Continuous green line: Induced SNF2PH RNAi (dox+). (Student T-test *p<0.01). **(C)** Reduction of the active VSG221 expression is reported after 48h of RNAi induction in all clones analyzed, whereas inactive VSG expression is clone dependent. Triplicate fluorescent measures are represented as the average ±SEM from three independent SNF2PH RNAi clones for both Renilla (active VSG-ES promoter, upper panel) and Luciferase (inactive VSG-ES promoter, lower panel) reporters in a DRALI cell line. **(D)** SNF2PH KD reflects clone variability in inactive VSG transcription. Relative transcripts levels for FLuc reporter gene upstream an inactive VSG-ES assessed by RT-qPCR upon 48h of doxycycline induction. Data from three independent clones with standard error of the mean were normalized with U2 mRNA, transcribed by pol III. (Student's T-test: *p<0.05, **p<0.01, ***p<0.001). **(E)** Scatter plot showing telomeric BES expression by RNA-seq analysis of SNF2PH depleted cells. Inactive VSG-ESs upregulation occurs only in a subset of the inactive telomers, while other inactive BESs appear to be unaffected. Thus, BES deregulation seems to be variable within each of the clones analysed, consistent with different reporter activities detected by the VSG-ES promoters (Fig S5C). SNF2PH is included as KD control. Data from two independent clones are presented compared with two RNAs isolated from the parental cell line.
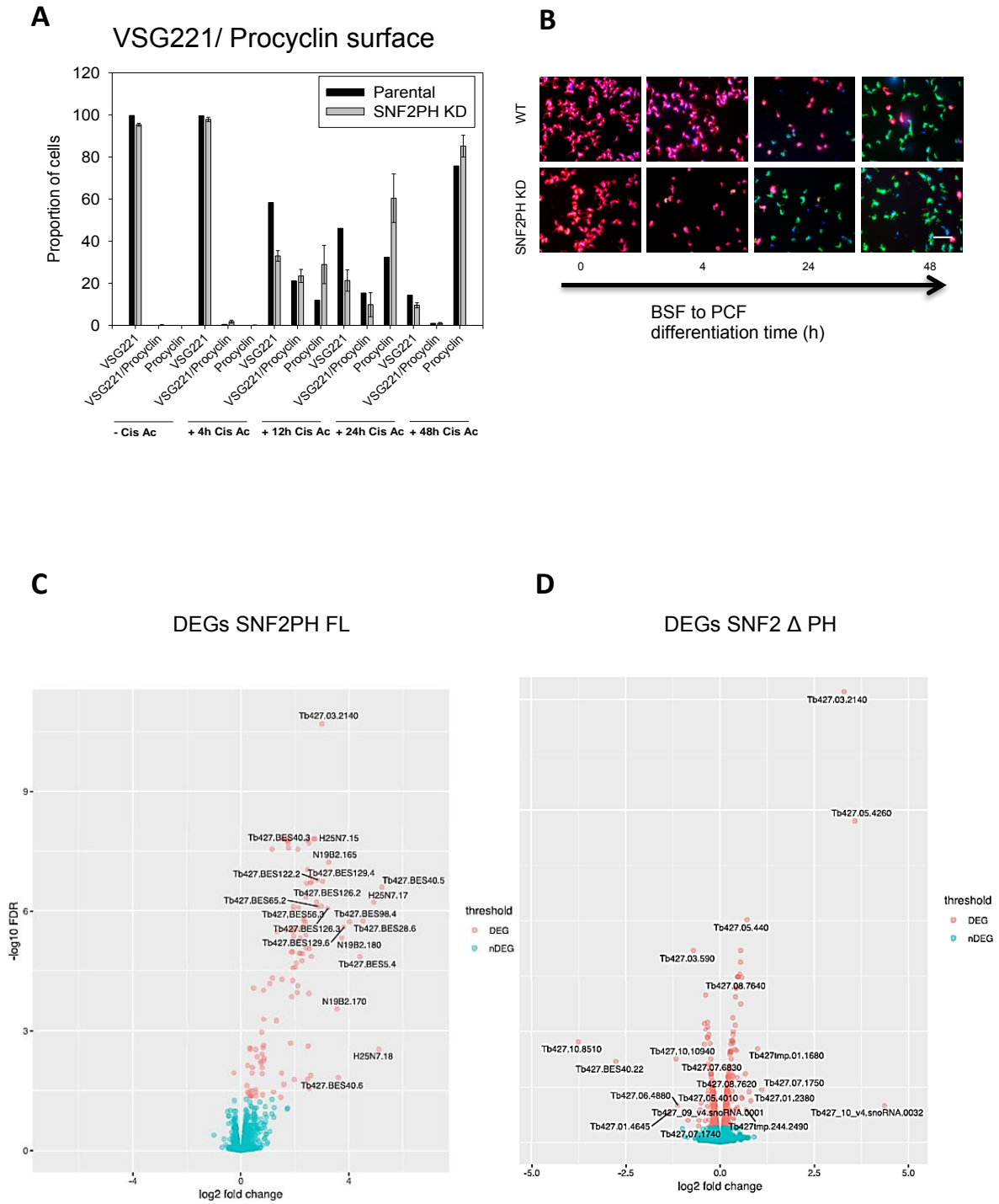
Figure S6. SNF2PH is required for maintenance of bloodstream form stage-specific expression profile. (A) Proportion of cells expressing VSG221 and Procyclin at several time-points during differentiation. Differentiation is accelerated when SNF2PH is depleted. Error bars represent means ± SD of two representative biological replicates. **(B)** Representative microscopy images at differentiation time-points. Double indirect 3D-IF was performed with rabbit anti-VSG221 (red), mouse anti-Procyclin (green) and DAPI staining (Blue) upon 3 mM of cis-Aconitate induction. Scale bar, 15 μm. **(C)** Volcano plot showing upregulated genes (117) versus downregulated (1) in SNF2PH. The log2 FC and FDR (-log10 FDR) values are shown. **(D)** Volcano plot showing most upregulated (397) and downregulated (340) genes in SNF2ΔPH considering FDR<0.05.

Appendix III: Buffer composition

Bloodstream form transfection

Isolation of *T. brucei* in blood

Cytomix

Reagent	Concentration
EGTA	2 mM
KCl	120 mM
$CaCl_2$	0.15 mM
[a]KPO_4	10 mM
HEPES	25 mM
$MgCl_2.6H_2O$	5 mM
Glucose	0.50%
BSA	100µg/ml
Hypoxanthine	1 mM
pH 7.6	

Separation buffer (SB)

Reagent	Concentration
NaCl	44 mM
Na_2HPO_4	57 mM
NaH_2PO_4	3 mM
Glucose	10 mM
pH 8.0	

Phosphate buffered saline (PBS)1x

Reagent	Concentration
NaCl	136.9mM
KCl	2.68mM
Na_2HPO_4	10.14mM
KH_2PO_4	76mM
pH 7.2	

Trypanosome Dilution Buffer (TDB)

Reagent	Concentration
Na_2HPO_4	20 mM
NaH_2PO_4	2 mM
NaCl	80 mM
$MgSO_4$	5 mM
Glucose	20 mM
pH 7.7	

Isolation of Genomic DNA from *T. brucei*

EA buffer

Reagent	Concentration
TrisHCl pH 8.0	10mM
NaCl	250mM
NP-40	0.5

Procyclic form transfection

ZPFM

Reagent	Concentration
NaCl	132 mM
KCl	8 mM
NaH_2PO_4	8 mM
KH_2PO_4	1.5 mM
$MgOAC.4H_2O$	1.5 mM
$CaCl_2$	90µM
pH 7.0	

EB buffer

Reagent	Concentration
TrisHCl pH 8.0	10mM
NaCl	250mM
EDTA pH 8.0	10 mM
SDS	0.5

Chromatin Immunoprecipitation (ChIP)

ChIP Lysis buffer

Reagent	Concentration
SDS	1.00%
EDTA pH 8.0	10 mM
TrisHCl pH 8.0	50 mM
PMSF	1 mM
DTT	1 mM

ChIP LiCl buffer

Reagent	Concentration
LiCl	0.25 M
NP-40	1.00%
Na-Deoxycholate	1.00%
EDTA pH 8.0	1 mM
TrisHCl pH 8.0	10 mM

ChIP Dilution buffer

Reagent	Concentration
SDS	0.01%
Triton X-100	1.10%
EDTA pH 8.0	1.2 mM
TrisHCl pH 8.0	16.7 mM
NaCl	150 mM
PMSF	1 mM
DTT	1 mM

ChIP TE buffer

Reagent	Concentration
TrisHCl pH 8.0	10 mM
EDTA pH 8.0	1 mM

ChIP Elution buffer

Reagent	Concentration
SDS	1.00%
NaHCO$_3$	0.1 M

ChIP Low Salt buffer

Reagent	Concentration
SDS	0.10%
Triton X-100	1.00%
EDTA pH 8.0	2 mM
TrisHCl pH 8.0	20 mM
NaCl	150 mM
pH 8.0	

ChIP High Salt buffer

Reagent	Concentration
SDS	0.10%
Triton X-100	1.00%
EDTA pH 8.0	2 mM
TrisHCl pH 8.0	20 mM
NaCl	500 mM
pH 8.0	

Purification of recombinant proteins in *E. coli*

Protein purification Lysis Buffer

Reagent	Concentration
Sodium phosphate	20 mM
NaCl	0.5 M
DTT	1mM
PMSF	1 mM
Leupeptin	1µg/ml
EGTA	1 mM
Lysozyme	100 mg/ml
Imidazole	20 mM
pH 7.4	
Lauryl Sarcosine	0.10%
Urea	8 M

SUMOylation assay in bacteria

In vitro SUMOylation Lysis buffer

Reagent	Concentration
NaCl	150 mM
TrisHCl	500 mM
Lysozyme	0.4 mg/ml
Triton X-100	0.10%
EDTA	10 mM
PMSF	1 mM

Purification of complexes in *T. brucei*

HA Magnetic beads

Reagent	Concentration
HEPES pH 7.4	20 mM
Sodium citrate	50 mM
MgCl$_2$	1 mM
CaCl$_2$	10 µM
CHAPS	0.10%
Protease inhibitors	2x

Purification of TbSUMO^{T106K} cell line conjugates

TbSUMO^{T106K} conjugates Lysis buffer

Reagent	Concentration
Urea	6 M
TrisHCl	50 mM
NaCl	500 mM
ß-mercaptoethanol	5 mM
pH 7.5	

TbSUMO^{T106K} conjugates Buffer 1

Reagent	Concentration
Urea	8 M
TrisHCl	10 mM
Na$_2$HPO$_4$	100 mM
NaH$_2$PO$_4$	100 mM
Imidazole	10 mM
ß-mercaptoethanol	5 mM
pH 8.0	

TbSUMO^{T106K} conjugates Buffer 2

Reagent	Concentration
Urea	8 M
TrisHCl	10 mM
Na$_2$HPO$_4$	100 mM
NaH$_2$PO$_4$	100 mM
Imidazole	10 mM
ß-mercaptoethanol	5 mM
pH 6.3	

TbSUMO^{T106K} conjugates Elution

Reagent	Concentration
Urea	8 M
TrisHCl	10 mM
Na$_2$HPO$_4$	100 mM
NaH$_2$PO$_4$	100 mM
Imidazole	200 mM
ß-mercaptoethanol	5 mM
pH 8.0	

Co-immunoprecipitation assays (Co-IP)

Lysis buffer SNF2PH IP

Reagent	Concentration
Urea	6 M
HEPES pH 7.5	50 mM
NaCl	500 mM
N-ethylmaleimide	20 mM
NP-40	0.50%

Lysis buffer to determine SNF2PH SUMOylation

Reagent	Concentration
HEPES pH 7.5	50 mM
NaCl	500 mM
N-ethylmaleimide	20 mM
NP-40	0.50%
Urea	6 M
Protease inhibitors	2x

Dilution buffer to determine SNF2PH SUMOylation

Reagent	Concentration
HEPES pH 7.5	50 mM
NaCl	500 mM
N-ethylmaleimide	10 mM
NP-40	1.00%
Lauryl Sarcosine	0.50%
EDTA	0.1 mM
Protease inhibitors	1x

Lysis buffer for Phosphorylated proteins

Reagent	Concentration
HEPES pH 7.4	20 mM
Sodium citrate	50 mM
$MgCl_2$	1 mM
$CaCl_2$	10 μM
NP-40	1.00%
Lauryl Sarcosine	0.50%
N-ethylmaleimide	10 mM
Protease inhibitors	2x
PhosphoStop	1 x

9. Appendix Methods

Appendix Methods of SNF2PH Protein tagging, purification and antibody production

AM I: Generation of a HA tagged cell line expressing SNF2PH

We first generate a single allele tagging of SNF2PH from the construct pTAP-HA-nSNF2PH-Hyg replacing a single allele by read-through insertion. Furthermore, we proceeded to replace the second copy of the endogenous SNF2PH gene by switching the resistance gene to Blasticidin. Double allele tagging considerably improved the monoallelic expression of SNF2PH. Positive clones obtained in the second round after Blasticidin selection (2.5µg/ml). Then they were checked by amplification of the upstream region containing the Hygromycin or Blasticidin resistance genes, and the respective 3' UTR SNF2PH gene in order to assess the appropriate replacement of both endogenous copy of the targeted gene **(S7A Fig)**. The generated monoclonal antibody mAb 11C10E4 allowed us to detect the HA tagged version of SNF2PH, also it was detected using the mAb against HA (Roche) **(S7B Fig)**. In order to improve SNF2PH detection, we decided to clone a 3 HA epitope upstream the N-terminus of SNF2PH protein, as well as to generate a powerful plasmid expressing tool for other genes of interest. For this purpose, we amplified a PCR product containing the remaining 2 HA epitopes and then subcloned in pTAP-HA -PHD-Hyg to finally obtain p3HA -PHD-Hyg. The construct was sequenced to ensure that the triple HA epitope was in frame under the T7 promoter. Next, we cloned into p3HA- PHD-Hyg the previous N-Terminus SNF2PH fragment to obtain the p3HA-nSNF2PH-Hyg. Then, the Hygromicin resistance gene was replaced by Blasticidin, yielding to p3HA-nSNF2PH-Bla with the aim of knocking the other allele copy. After the first round of selected clones for p3HA-nSNF2PH-Hyg, we performed a second electroporation to target to the second endogenous copy of SNF2PH to ensure a complete 3 HA tagged protein expression (see **Tables 4 and 6** for cloning details). From the selected clones to analyze, only two of them were positive for 3 HA tagging in both SNF2PH endogenous copies **(S7C Fig)**. Next, we confirmed the expression levels of these positive clones after 24h of Doxycycline induction **(S7D Fig).**
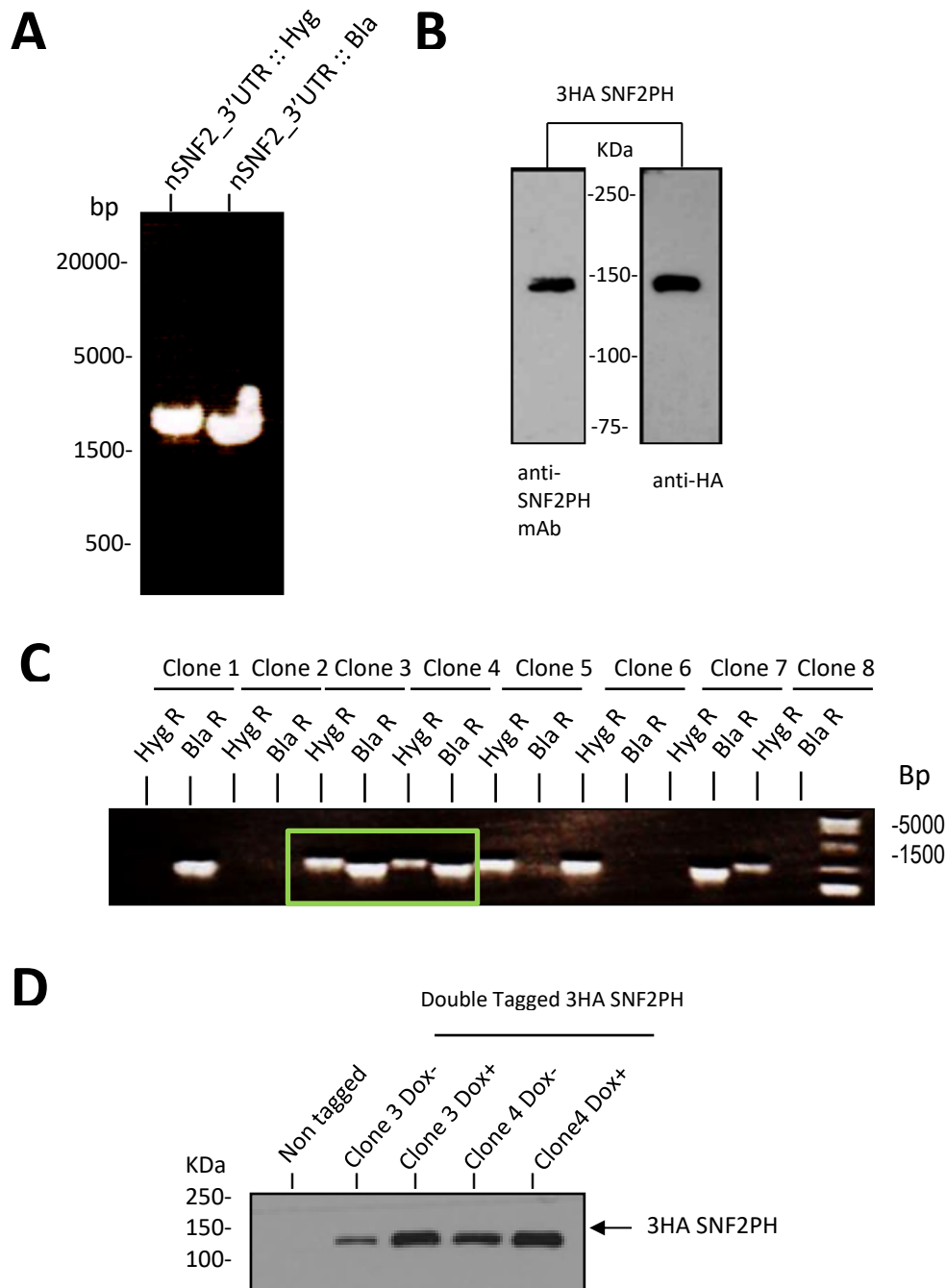
Figure S7. Cloning steps to generate a stable 3HA tagged SNF2PH cell line. (A) Diagnostic PCR from genomic DNA for the selected double allele tagged SNF2PH clones. Lanes correspond to the amplified region between SNF2PH 3'UTR and the Hygromicin and Blasticidin resistance gene, respectively. **(B)** Expression and detection of endogenous and tagged SNF2PH. Bloodstream form *T. brucei* cell line expressing HA epitope in both allele copies (3HA-SNF2PH), detected by both monoclonal antibody anti-SNF2PH or anti-HA epitope (~105kDa). **(C)** Diagnostic PCR of genomic DNA extracted from the clones during the second round of antibiotic selection. Each clone was analyzed by the amplified 3'UTR fragment of SNF2PH followed by the upstream region of the Hygromycin or Blasticidin resistance gene (Hyg R and Bla R, respectively). Clones 3 and 4 were positive for the 3 HA tagged version of SNF2PH in both allele copies (Green box). **(D)** WB analysis of 3HA SNF2PH expression in clones 3 and 4 induced with Doxyciclin (Dox) using the monoclonal antibody against HA epitope (Roche).

AM II: Purification of recombinant 6His-cSNF2PH in *E. coli* KRX cells

To generate polyclonal and monoclonal antibodies against SNF2PH, we expressed the C terminal domain of this protein into the plasmid pET28a (pET28a(+)- 6xHis-cSNF2PH, **Tables 4 and 6**) to avoid the restriction sites at the N terminus of the protein that did not let to clone the fragment. After the expression in KRX cells, we scaled the culture up to 2-3 L of LB media with kanamycin selection (25 μg/ml media), including 1 mM IPTG and 0.01% arabinose to allow the transcription from the KRX promoters. We previously stablished the optimal conditions for yielding enough amount of recombinant 6HiscSNF2PH by a pilot expression assay (4h at 37ºC). We also checked its solubility by detergent titration and urea addition, giving rise into a mostly insoluble fraction with a low representation of the soluble portion **(S8A Fig)**. We joined together fractions with a high representation of the purified protein with a relative molecular weight of 46.2 kDa and dialyzed against 1 L of PBS per ml of fraction at 4°C. We quantified final dialyzed fraction by Bradford and evaluated in an SDS-polyacrylamide gel testing the purity. Then, we performed successive rounds of 6His cSNF2PH protein purification in order to obtain enough amount of recombinant protein to be inoculated into both mice and rabbit for monoclonal and polyclonal antibody production, respectively, along with obtaining suficient antigen to be adsorved in ELISA plates.

AM III: Production of SNF2PH polyclonal and monoclonal antibodies

First, we produced a polyclonal antibody in mouse after five intraperitoneal inoculations (with uncomplete adjuvant) of the recombinant C-T domain of SNF2PH expressed in *E. coli* KRX cells. The final antiserum was titrated in a blot assay to determine the optimal working dilution. The polyclonal antibody was able detect a 105 KDa band corresponding to SNF2PH in subsequent serial dilutions **(S8B Fig)**. Next, we produced a monoclonal antibody by a fusion of mice spleen cells with a SP cell line. This allowed us to obtain hybridoma cells producing a specific antibody against a SNF2PH epitope depending of the antibody expression of the clone generated, evaluated by ELISA assay performed in 96-well plates with adsorbed 6His cSNF2PH protein. We started with a first lineage of clones (parents) that diverged in subclones with an epitope-specific antibody expression from the tested supernatant. Out of the obtained subclones, only one of them (mAb 11C10E4) succeed to produce a monoclonal population of antibodies allowing SNF2PH detection **(S8C Fig)**. Finally, we obtained ascites by growing hybridoma cells within the peritoneal cavity to produce a fluid (ascites) containing a higher concentration of SNF2PH antibody. This method provides higher yield of antibody than a cell culture hybridoma supernatant **(S8D Fig)**.
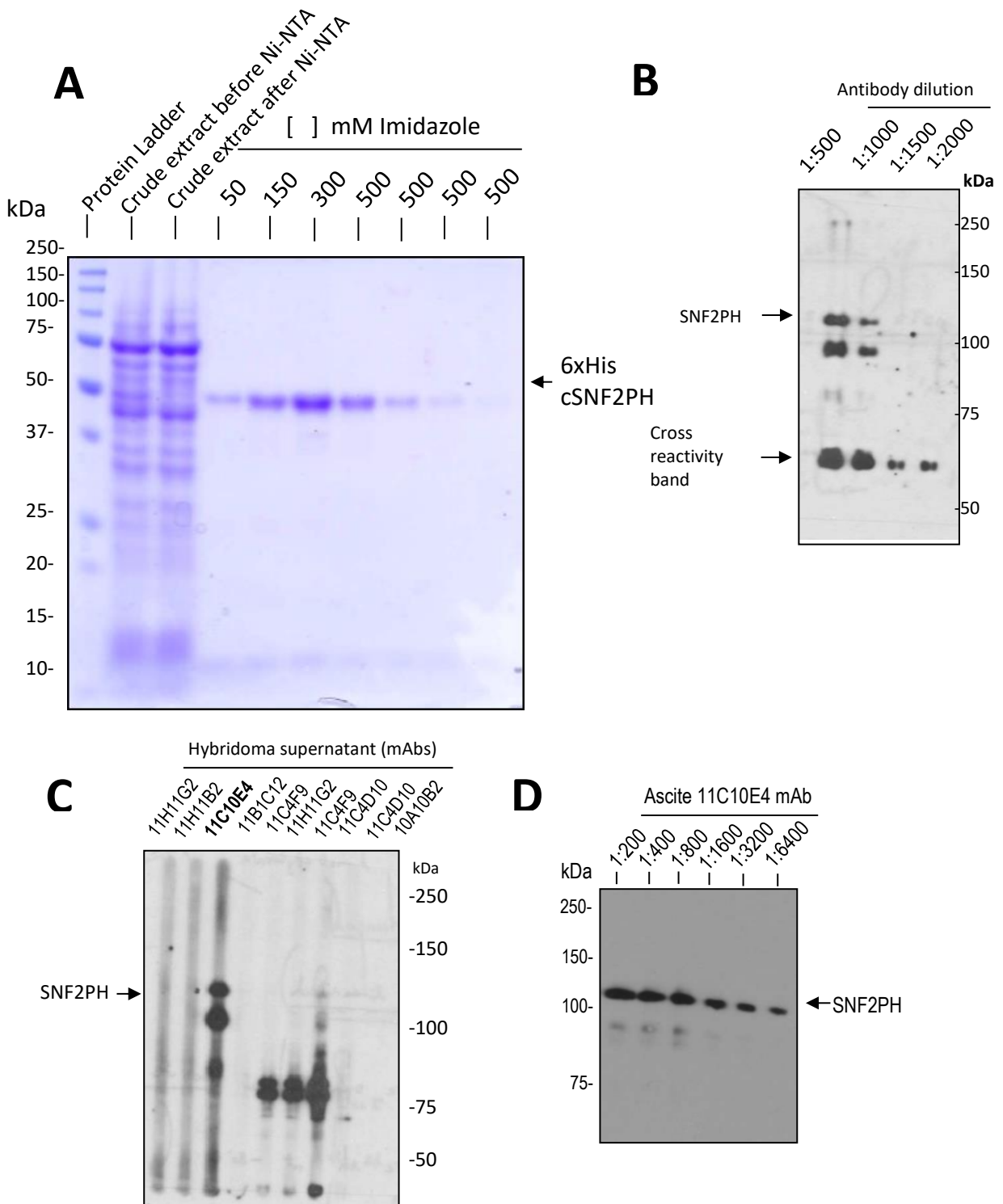
Figure S8. Recombinant 6HiscSNF2PH purification and monoclonal antibody production. (A) *Coomassie* stain of purified 6HiscSNF2PH fractions in a 15% polyacrylamide SDS-gel. Crude extracts from lysed bacterial cells were evaluated for 6His cSNF2PH expression (46.2 kDa) upon IPTG and arabinose induction before Ni^{2+}-NTA column purification (second lane). After Ni^{2+}-NTA column, the apparent expressed band disappeared, ensuring a truthful affinity union of the histidine charges and the Nickel ions (third lane). Then, we eluted using an increased imidazole gradient that competes with interaction His-Ni^{2+} interaction by a greater affinity for Nickel ions than histidine. We observed that the optimum imidazole concentration for His cSNF2PH elution was around 300 mM (sixth lane) yielding a notably purity by qualitative evaluation after *Coomassie* stain. **(B)** Titration of the polyclonal antibody produced in mice against SNF2PH. **(C)** WB analysis of each supernatant containing monoclonal antibodies produced by a subcloned hybridoma lineage. **(D)** Titration of TbSNF2PH mAb ascite in a preparative *T. brucei* protein extract. The yield of the obtained TbSNF2PH ascites increased up to 1:6400-fold dilutions.

Second, we decided to produce polyclonal antibodies in rabbit to improve the affinity or solving any handicap in any of the molecular approaches for the monoclonal antibody. After 5 inoculations of 500 µg recombinant protein per dose and rabbit, we isolated the whole antisera containing the unpurified polyclonal antibody. We also included a control of no immunized sera before initiating the immunizations (preimmune serum or prebleed). Titration of the 5th bleed showed a majority band corresponding to the endogenous SNF2PH in a total *Trypanosoma brucei* protein extract together with other unspecific bands **(S9A Fig)**. Next, we purified the polyclonal antibody in order to reduce the background and improve the specificity of the antibody. To address it, we use *Aminolink Plus Immobilization Kit* (Thermo Fisher) that allows the purification of antibodies by immobilizing them in a agarose beads which has been previously activated with aldehyde groups to enable a covalent coupling via primary amines (-NH$_2$). It was necessary to concentrate the recombinant protein into a minimal volume in order to facilitate the coupling with the agarose beads by using *Amicon Centrifugal Filter* units (Millipore) for concentration and purification of biological solutions. We applied 4 ml of dialyzed purified protein to obtain approximately 100µl of concentrated protein in successive rounds of centrifugation, yelding to 1.5 ml of concentrated protein at 2.17 µg/µl. After the coupling of the protein, we added the antiserum and eluted by pH change followed by a neutralization to avoid antibody denaturation. We quantified the eluted antibodies using Nanodrop Spectrophotometer with absorbance measured at 280 nm. We joined fractions with higher concentration and titrated by WB. **(S9B Fig)**. Finally, we stablished the optimal dilution for the antibody (1:1000) **(S9C Fig)**.
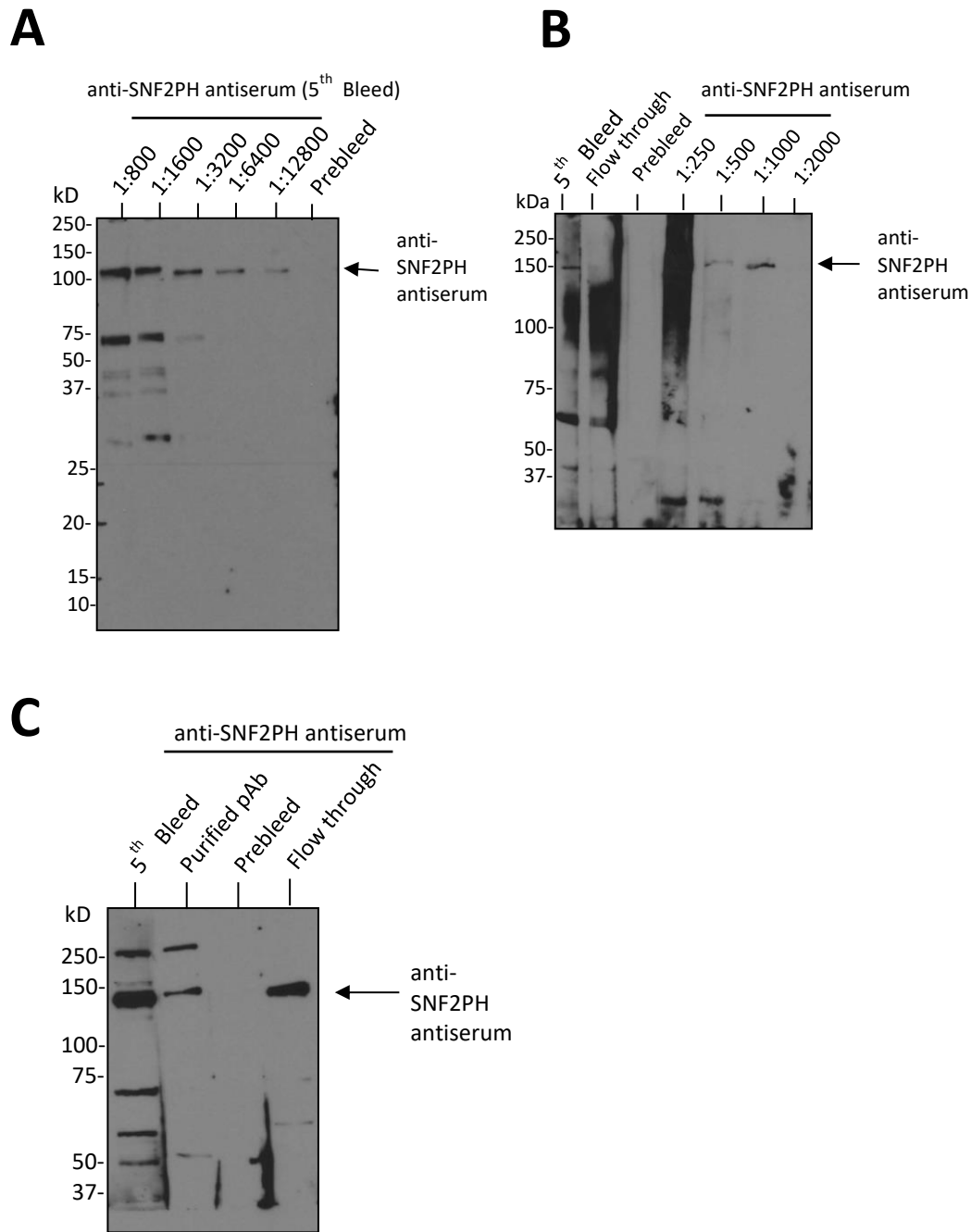
Figure S9. Production and purification of SNF2PH polyclonal antibody (A) Serial dilutions of the polyclonal antibody anti-SNF2PH produced in rabbit from the 5th bleed. The strength of the signal from the antisera against SNF2PH could be diluted up to 1:12800. **(B)** Titration of the joined purified fraction for the polyclonal antibody against SNF2PH. The Lanes of blot analysis show the signal for SNF2PH from the 5th bleed (5th bleed), the not bound antiserum fraction (Flow throught), no signal for prebleed antiserum (Prebleed), and serial dilutions of the purified fraction of anti-SNF2PH pRbAb (last 4 lanes), being 1:1000 an optimal dilution to work with. **(C)** Representative blot for purified anti-SNF2PH pRbAb at 1:1000 dilution. Although the binding of the antibody throughout the agarose beads is effective, there is an important loss of specific SNF2PH antibodies due to its excess in the flow through.

AM IV: Stablishing the optimal conditions for Immunoprecipitation using the polyclonal antibody anti-SNF2PH

The purpose of generating a polyclonal antibody against SNF2PH comes from the restriction of the monoclonal in its application for other molecular methods such as Immunoprecipitation (IP) and Chromatin Immunoprecipitation (ChIP) assays. In these last two techniques we did not succeed in a concluding result leading to a lack of reproducibility. On the other side, we face the problem of specificity of the animal species in terms of antibody cross reactivity when in the assay there are present antibodies from closer families. Having a polyclonal antibody against SNF2PH produced in rabbit gives us the chance of performing IP assays without any interference in the blot assay when mice or rat antibodies are used. For this reason, we sought for the optimal IP conditions for the polyclonal antibody against SNF2PH. Our starting point was the detergent titration in denaturing conditions to preserve the reversible link of the SUMO modification group. So, we choose NP-40 at two concentrations (0.1 and 1%) and combined, or not, with Triton X-100, SDS and Lauryl Sarcosine at 0.5% in a background of 1M urea. A *Trypanosoma* cell extract was lysed with 1% NP-40 and 6M of urea previously to be diluted and distributed at the assayed conditions. After an overnight incubation of the polyclonal antibody anti-SNF2PH in each case, the antibody was eluted with heat after serial washing steps. Eluted immunoprecipitated products were subjected to a WB analysis using the previously generated monoclonal antibody against SNF2PH. We also included the total protein extract as a starting point to evaluate the efficiency of the IP (Input) **(S10 Fig)**. Results showed an increased IP efficiency for polyclonal anti-SNF2PH antibody with 1% NP-40 and 0.5% Lauryl Sarcosine in denaturing conditions (1M urea). At this point, we set up the conditions for SUMO immunoprecipitation complexes in addition to other molecular approaches.

IP: anti-SNF2PH antiserum (c13)
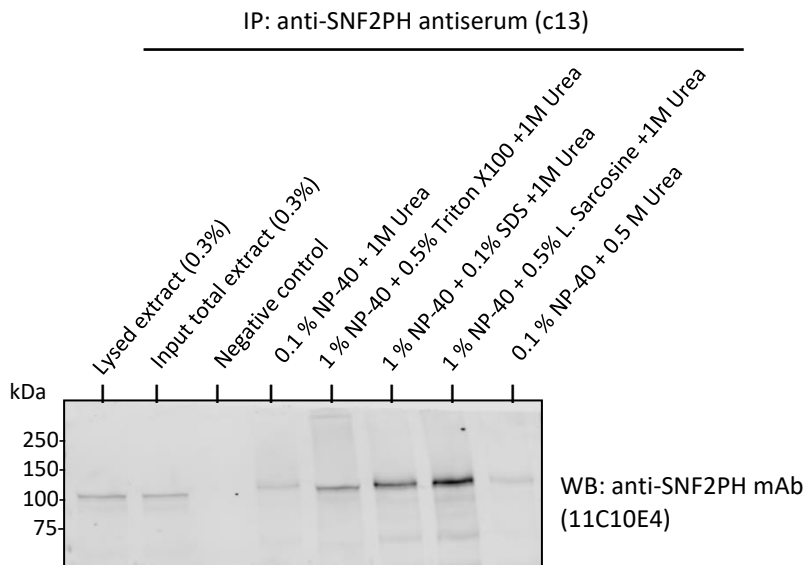

WB: anti-SNF2PH mAb
(11C10E4)

Figure S10. Setting up the conditions for anti-SNF2PH polyclonal antibody. Western blot assay against SNF2PH mAb reveals the finest efficiency for SNF2PH polyclonal antibody immunoprecipitation depending on denaturing conditions. Lanes represent different condition assayed; including an input of the whole lysed extract before and after being diluted (Lysed extract (0.3%) and Input total extract (0.3%), respectively). Less efficiency was reached in presence of NP-40 alone, whereas the combination with other detergents improved it. From the highest to lowest efficiency order; Lauryl Sarcosine, SDS and Triton X-100.

10. Appendix Publications

SUMOylated SNF2PH promotes variant surface glycoprotein expression in bloodstream trypanosomes

Andreu Saura[1], Paula A Iribarren[2], Domingo Rojas-Barros[1], Jean M Bart[1], Diana López-Farfán[1], Eduardo Andrés-León[1], Isabel Vidal-Cobo[1], Cordula Boehm[3], Vanina E Alvarez[2], Mark C Field[3,4] & Miguel Navarro[1,*] (iD)

Abstract

SUMOylation is a post-translational modification that positively regulates monoallelic expression of the trypanosome variant surface glycoprotein (VSG). The presence of a highly SUMOylated focus associated with the nuclear body, where the VSG gene is transcribed, further suggests an important role of SUMOylation in regulating VSG expression. Here, we show that SNF2PH, a SUMOylated plant homeodomain (PH)-transcription factor, is upregulated in the bloodstream form of the parasite and enriched at the active VSG telomere. SUMOylation promotes the recruitment of SNF2PH to the VSG promoter, where it is required to maintain RNA polymerase I and thus to regulate VSG transcript levels. Further, ectopic overexpression of SNF2PH in insect forms, but not of a mutant lacking the PH domain, induces the expression of bloodstream stage-specific surface proteins. These data suggest that SNF2PH SUMOylation positively regulates VSG monoallelic transcription, while the PH domain is required for the expression of bloodstream-specific surface proteins. Thus, SNF2PH functions as a positive activator, linking expression of infective form surface proteins and VSG regulation, thereby acting as a major regulator of pathogenicity.

Introduction

Antigenic variation, the major mechanism by which African trypanosomes evade the host immune response, is mediated by switching expression between immunologically distinct variant surface glycoprotein (VSG) genes [1]. The active VSG gene is transcribed polycistronically by RNA polymerase I, together with several expression site-associated genes (ESAGs), from a large telomeric locus (40–60 kb), known as a VSG expression site (VSG-ES), currently named bloodstream ES (BESs) [2]. In the mammalian bloodstream form (BF), where antigenic variation occurs, only one of ~15 VSG-ES genes is transcribed at a given time, resulting in monoallelic expression and a dense surface coat comprised of a single VSG type [3–5]. The active VSG-ES is recruited to a nuclear compartment, the expression site body (ESB), which facilitates monoallelic transcription [6–8]. Interestingly, small ubiquitin-like modifier (SUMO) post-transcriptionally modified proteins are associated with the ESB within a highly SUMOylated focus (HSF) [9]. However, in the insect or procyclic form, VSGs are not expressed and procyclin glycoproteins cover the parasite surface [10].

SUMOylation is a large and reversible post-translational modification (PTM) that regulates many critical processes, including transcription, protein–protein interactions, protein stability, nuclear localization, DNA repair, and signaling [11]. In *Trypanosoma brucei*, there is a single SUMO ortholog, which is essential for cell cycle progression of the procyclic form [12]. Proteomic analyses of SUMO substrates in this life stage identified 45 proteins involved in multiple cellular processes, including epigenetic regulation of gene expression [13]. Transcription factors are well known SUMO targets, whose activity can be modulated in both gene silencing and activation [14]. In *T. brucei* BF, SUMO-conjugated proteins were detected highly enriched in the nucleus in a single focus (HSF) associated with the ES body (ESB) and in the active VSG-ES chromatin, suggesting chromatin SUMOylation acts as a positive epigenetic mark to regulate VSG expression [9]. Chromatin SUMOylation to the active VSG-ES locus is required for efficient recruitment of RNA polymerase I in a SUMO E3 ligase (TbSIZ1/PIAS)-dependent manner, suggesting protein SUMOylation facilitates the accessibility of additional transcription factors [9].

1 Instituto de Parasitología y Biomedicina "López-Neyra", CSIC (IPBLN-CSIC), Granada, Spain
2 IIB-UNSAM, Buenos Aires, Argentina
3 School of Life Sciences, University of Dundee, Dundee, UK
4 Biology Centre, Institute of Parasitology, Czech Academy of Sciences, Ceske Budejovice, Czech Republic
*Corresponding author. Tel: +34 958181651; Fax: +34 958181633; E-mail: miguel.navarro@ipb.csic.es

Thus, we sought to identify major SUMO-conjugated proteins in the mammalian infective form and found a novel protein annotated as a transcription activator in the database (Tb927.3.2140). Structural conserved domain predictions suggest that Tb927.3.2140 is a member of the Snf2 (Sucrose Nonfermenting Protein 2) SF2 helicase-like superfamily 2 of chromatin remodelers [15–17] and also contains a plant homeodomain (PHD). Thus, we designate the protein SNF2PH.

Here, we show that SNF2PH is a developmentally regulated protein enriched at chromatin of the VSG-ES (BES) telomere, particularly at promoter regions when modified by SUMO. SNF2PH depletion leads to reduced VSG transcription and upregulation of developmental markers for the insect stage. ChIP-seq data suggest SNF2PH binds to selective regions in chromatin, in addition to the active VSG-ES, like developmentally regulated loci, rDNA, SL-RNA, and, interestingly, also to clusters of tRNA genes, which function as insulators for repressed and active chromatin domains in other eukaryotes. SNF2PH is strongly downregulated in quiescent (pre-adapted to host transition) trypanosomes generated in both pleomorphic (differentiation-competent) and monomorphic (by AMPKα1-activation) strains. Further, SNF2PH expression is negatively regulated in the insect procyclic form. Most importantly, overexpression of SNF2PH in the insect form triggers the expression of bloodstream stage-specific surface protein genes, suggesting a role as positive regulator of differentiation. Thus, SNF2PH links immune evasion with pathogenicity.

Results

Trypanosome SNF2PH is SNF2_N-related protein that contains an unusual plant homeodomain

SUMOylation is a hallmark of epigenetic VSG regulation at the level of chromatin and nuclear architecture [9]. The highly SUMOylated focus (HSF) detected by a specific mAb against TbSUMO in the nucleus of bloodstream form (BF) trypanosomes was recently associated with the nuclear body ESB [9], the site for VSG-ES monoallelic expression [6]. Recognition of HSF together with the detection of highly SUMOylated proteins at the active VSG-ES chromatin by ChIP analysis suggests that a number of SUMOylated proteins are mechanistically involved in regulation of VSG expression [9]. Therefore, identifying these proteins is a novel approach for the discovery of factors involved in VSG regulation. To identify abundant SUMOylated proteins, we performed a non-exhaustive proteomic analysis utilizing BF protein extracts from a cell line expressing an 8His-HA-tagged SUMO (see Materials and Methods). LC-MS/MS analyses of His-HA-affinity-purified extracts robustly identified several proteins (see Appendix Table S1). Particularly, interesting was Tb927.3.2140 (length 948 aa), a protein annotated in the TriTrypDB database as a transcription activator, which contains a conserved SNF2 family N-terminal domain.

Comparative analyses of Tb927.3.2140 at CDART [18] and the NCBI CDD domain database identified three conserved domains: DEXHc_Snf, e-value $9.4e^{-74}$, SF2_C_SNF, e-value $8.0e^{-50}$, PHD5_NSD, e-value $6.2e^{-14}$. Structural CD predictions suggest than Tb927.3.2140 is a member of the Snf2 family (Sucrose Nonfermenting Protein 2) from the SF2 helicase-like superfamily 2

of chromatin remodelers [15–17], which regulate DNA accessibility to facilitate central cellular processes as transcription, DNA repair, DNA replication and cell differentiation [15,16]. Next, searching for Tb927.3.2140 homologues using DELTA-BLAST against UniProtKB/SwissProt database, identified a protein member of the SWI/SNF family, SMARCA1 (e-value, $4e^{-157}$) (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 1) also known as the global transcription activator SNF2L1 (length, 1054 aa) (homonyms SWI; ISWI; SWI2; SNF2L; SNF2L1; SNF2LB; SNF2LT; hSNF2L; NURF140) all described to be involved in transcription for either gene activation or gene repression [16]. In addition to the SNF2 N domain, a conserved helicase C-terminal domain was also detected, known to function as a chaperon-like in the assembly of protein complexes. Interestingly, Tb927.3.2140 also contains a plant homeodomain (PHD) that is absent from other known trypanosome chromatin remodelers. The PHD is a conserved homeodomain involved in development [19] that binds H3 tails and reads unmodified H3 tails [20] as well as H3 trimethylated at Lys4 (H3K4me3) [21] or acetylated at Lys8 and Lys14 [22,23]. The PH domain is conserved in histone methyltransferases, including murine HMT3 and human NSD3 (Appendix Fig S1). Thus, we named this chromatin-remodeling factor trypanosome SNF2PH.

SNF2PH is developmentally regulated and associated with the ESB nuclear body

In order to investigate SNF2PH protein expression, we raised a monoclonal antibody (mAb) (11C10E4) against the recombinant protein expressed in bacteria. Western blot of total protein extracts and immunofluorescence (IF) showed that SNF2PH protein levels are developmentally regulated, with higher expression in the bloodstream compared to the insect form (Fig 1A and Appendix Fig S2). The specificity of mAb11C10E4 was demonstrated as SNF2PH levels in whole cell extracts were markedly reduced in RNAi cells (Fig 1B).

Subcellular localization of SNF2PH by 3D-deconvolution IF (3D-IF) microscopy with mAb11C10E4 showed a nuclear localization, with disperse distribution in the nucleoplasm, including puncta and enrichment at the nucleolar periphery (Fig 1C). To determine whether SNF2PH associates with the active VSG-ES locus, we used a GFP-lacI targeted VSG-promoter cell line [6]. We detected 38.8% (n = 55) colocalization (Pearson's correlation coefficient) with GFP-tagged active VSG-ES using an anti-GFP rabbit antiserum and mAb11C10E4 anti-SNF2PH (Fig 1D). Statistical analysis showed even lower association in 2K1N cells (S-G2) (Fig EV1A), suggesting that SNF2PH association with the active VSG-ES locus occurs in a cell cycle-dependent manner.

To investigate the association with the HSF, we stained cells with anti-TbSUMO mAb [9] and anti-SNF2PH antiserum (Materials and Methods). 3D-IF showed colocalization between SNF2PH and HSF in 53.7% of the cells (n = 67)) (Figs 1E and EV1B), likely due to the highly dynamic nature of protein SUMOylation. Similar colocalization in 53.49% of the cells was observed between SNF2PH and the ESB (Figs 1F and EV1C), indicated by extranucleolar pol I signal visualized with a YFP-tagged RPB5z (specific RNA pol I subunit 5z [24]) (Figs 1F and EV1C).
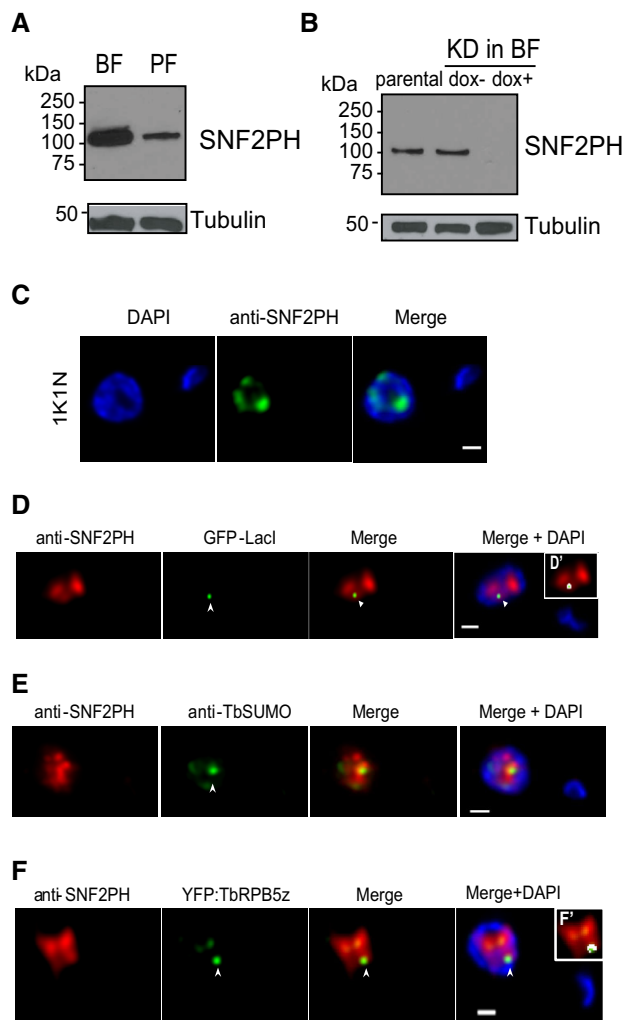
Figure 1. SNF2PH is a developmentally regulated protein associated with the expression site body (ESB) and highly SUMOylated focus (HSF).

A SNF2PH is differentially expressed in *T. brucei* developmental stages. The mammalian infective form, bloodstream form (BF), and the insect form, procyclic form (PF).

B Knockdown of SNF2PH by inducible RNA interference in bloodstream form leads to protein depletion after 24 h. (5×10^6 cells/lane): parental, uninduced (dox⁻), and induced (dox⁺). Total cell extracts were analyzed by Western blotting using the anti-SNF2PH mAb.

C SNF2PH is diffusely distributed in the nucleoplasm with certain enrichment in the nucleolus. Panels show DAPI and green channels after IF with the anti-SNF2PH mAb (11C10E4). Scale bar, 1 µm.

D SNF2PH associates with the active VSG-ES. A cell line with a GFP-LacI tag in the active VSG-ES [6] was subjected to double 3D-IF using anti-SNF2PH mAb (red), anti-GFP rabbit antiserum (green) and DAPI staining. Maximum intensity projections of deconvolved slices containing the GFP signal are shown (arrowhead). (D') Inset shows a higher magnification of the nucleus including anti-SNFPH and anti-GFP fluorescence signals colocalization mask (white). Scale bar, 1 µm.

E Colocalization analysis of SNF2PH with the Highly SUMOylated focus (HSF). SNF2PH associates with the HSF (arrowhead) in bloodstream form nucleus. Indirect 3D-IF analyses were carried out using the rabbit anti-SNF2PH antiserum (red) and the anti-TbSUMO mAb 1C9H8 (green) [9]. Scale bar, 1 µm.

F CSNF2PH partially colocalizes with YFP-tagged TbRPB5z in the ESB. A cell line expressing an N-terminal fusion of a Yellow Fluorescent Protein (YFG) with the RNA pol I-specific subunit RPB5z described previously [24] was used to analyze by double 3D-IF a possible association of SNF2PH with the ESB. The 3D-IF was performed using the anti-SNF2PH mAb (red) and rabbit anti-GFP antiserum (green) that recognizes the Yellow GFP variant. Deconvolved slices containing both SNF2PH and the extranucleolar ESB (arrowhead) are represented as maximum intensity projections. (F') Inset shows a higher magnification of the nucleus including anti-SNF2PH and anti-GFP fluorescence signals colocalization mask (white). Scale bar, 1 µm.

nucleolus and nuclear periphery in one ($84.12\% \pm 0.25\%$) or two puncta ($15.88\% \pm 0.25\%$) (Appendix Fig S3).

The low signal of SNF2PH antibody in TbSUMO IP experiments is likely a consequence of the dynamic nature of SUMOylation, yielding a small population of SUMOylated SNF2PH form at any given time; similar behavior has been demonstrated for TbRPAI (RNA Polymerase I largest subunit) [9] and additional SUMO proteins in *T. brucei* [25]. To determine which domains of SNF2PH are SUMOylated, we used an *E. coli* strain expressing the complete *T. brucei* SUMOylation system [13]. We evaluated two different constructs encompassing the SNF2PH N-terminal or C-terminal domain (SNF2PH-N and SNF2PH-C, respectively), bearing a Flag tag. We co-expressed SNF2PH-N and SNF2PH-C in *E. coli* with TbSUMO (already exposing the diGly motif) and both activating enzyme subunits (TbE1a/TbE1b) plus the conjugating enzyme (TbE2). SNF2PHN appears as a single band migrating at the expected position when expressed alone in *E. coli* (Fig 2C, lane 1) or when co-expressed with a partially reconstituted system (lanes 2 and 3). However, when co-expressed with the complete SUMOylation system, additional slower-migrating bands can be detected (lane 4), suggesting that the N-terminal domain of SNF2PH can be SUMO conjugated. In contrast, the C-terminal domain is not a target of SUMOylation since it is only detected as a single protein band at the expected position of the unmodified protein (Fig 2D).

To confirm heterologous SUMOylation of SNF2PHN, we performed *in vitro* deconjugation reactions using the specific *T. brucei* SUMO isopeptidase TbSENP. As shown in Fig 2E, the

SNF2PH is SUMO conjugated

To investigate whether SNF2PH is a *bona fide* SUMO-modified protein, we carried out immunoprecipitation (IP) utilizing the anti-TbSUMO mAb [9] and anti-SNF2PH antiserum under denaturing conditions to capture only proteins with covalent SUMO modifications. IP suggested that SNF2PH is SUMOylated when analyzed by Western blotting using the anti-TbSUMO mAb on a SNF2PH immunoprecipitate (Fig 2A). The reciprocal experiment, using anti-SNF2PH antiserum on a TbSUMO immunoprecipitate reproducibly detected SNF2PH conjugated to TbSUMO (Fig 2B).

While IP demonstrates that SNF2PH is SUMOylated, it is unknown whether nuclear conjugation with SUMO is associated with dispersed nuclear foci or localization to a specific subnuclear site. We performed Proximity Ligation assays (PLA) (O-link Bioscience), an IF method where a signal is produced only if two proteins, or a protein and its PTM, are within 40 nm. After a first IF experiment using anti-SUMO mAb and the SNF2PH antiserum, secondary species-specific antibodies conjugated with oligonucleotides are hybridized to the two PLA probes to produce a DNA by rolling circle replication. As the PLA assay detected positive amplification this suggests that SNF2PH is SUMOylated *in situ* in both the
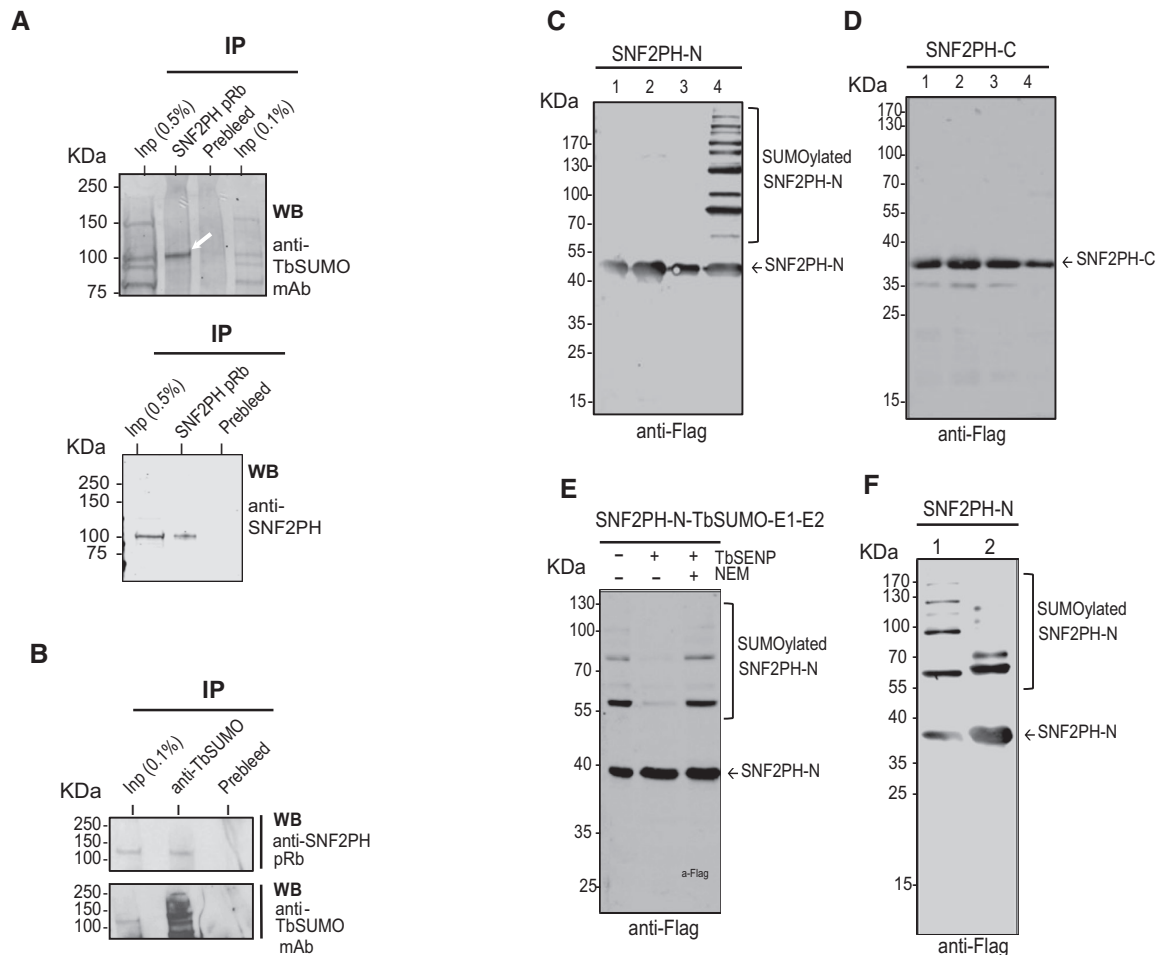
Figure 2. SNF2PH is SUMOylated *in vivo* in trypanosomes and *in vitro* using a SUMOylation heterologous system.

A Immunoprecipitation (IP) of bloodstream SUMOylated proteins revealed that SNF2PH is SUMOylated. A nuclear fraction was lysed in urea-containing buffer, and proteins were immunoprecipitated with rabbit anti-SNF2PH antiserum or unspecific antiserum (prebleed) and probed with anti-TbSUMO mAb 1C9H8 (arrow). As a control, IP samples were reprobed with SNF2PH antiserum (below).

B A reciprocal IP experiment was performed using anti-TbSUMO mAb and probed with SNF2PH antiserum. As a control, the blot was reprobed with anti-TbSUMO mAb (lower panel). Inp: Input, IP (0.5%).

C Anti-Flag Western blot analysis of SNF2PHN performed on soluble cell extracts from induced cultures of *E. coli* transformed with pET28-SNF2PHN-3xFlag alone (lane 1), or in the background of an incomplete (lane 2, pACYCDuet-1-TbE1a-TbE1b; lane 3, pCDFDuet-1-TbSUMO-TbE2) or a complete (lane 4, pCDFDuet-1-TbSUMO-TbE2 plus pACYCDuet-1-TbE1a-TbE1b) SUMOylation system.

D Similar samples as described in (C) were analyzed for SNF2PHC.

E Cell lysates of *E. coli* heterologously expressing SNF2PH and the complete *T. brucei* SUMOylation system were incubated at 28°C in the absence (−) or presence (+) of recombinant TbSENP. The deconjugation activity of TbSENP was specifically inhibited by the addition of 20 mM NEM. Reaction mixtures were analyzed by Western blot using anti-Flag monoclonal antibodies.

F Western blot analysis of SUMOylated SNF2PHN pattern performed on soluble cell extracts from a complete bacterial SUMOylation system using a wild type version of SUMO (lane 1) or a Lys-deficient version of SUMO (TbSUMO K9R) unable to form chains (lane 2).

additional slowly migrating bands observed when SNF2PHN was co-expressed with the *T. brucei* SUMOylation bacterial system (lane 1) completely disappear upon treatment of cell lysates with TbSENP (lane 2), and the deconjugation ability of TbSENP was specifically inhibited by addition of 20 mM NEM (lane 3). To investigate the nature of SUMOylation of SNF2PHN, we compared the patterns obtained in the bacterial system when replacing wild-type SUMO with a variant unable to form SUMO chains (Fig 2F). In the latter case, a doublet near the 55 kDa marker can be detected, suggesting that there are at least two major sites for SUMOylation in the SNF2PH N-terminus.

SNF2PH is highly enriched at active VSG-ES promoter chromatin

To investigate SNF2PH occupancy at VSG-ES loci, we performed chromatin IP (ChIP) using anti-SNF2PH antiserum in a promoter-tagged cell line. To overcome the problem of highly homologous sequences at the promoter region among the 15 telomeric VSG-ESs, we developed a tagged cell line (Dual-reporter Renilla Active Luciferase Inactive or DRALI) (loci of interest schematic in Fig 3A). The reporter genes in the DRALI cell line allowed us to determine SNF2PH occupancy at the region downstream of the promoter in either active or inactive VSG-ESs. First, we analyzed occupancy of

SNF2PH by ChIP and quantitative PCR (qPCR), which detect significant SNF2PH enrichment at the *RLuc* gene downstream of the active VSG-ES promoter (*P* < 0.001) as well as at the active VSG221 gene located in the telomere of BES1 (*P* < 0.01). However, *FLuc* located downstream of an inactive VSG-ES promoter (Fig 3A) was not significantly detected (Fig 3B). SNF2PH enrichment was also not
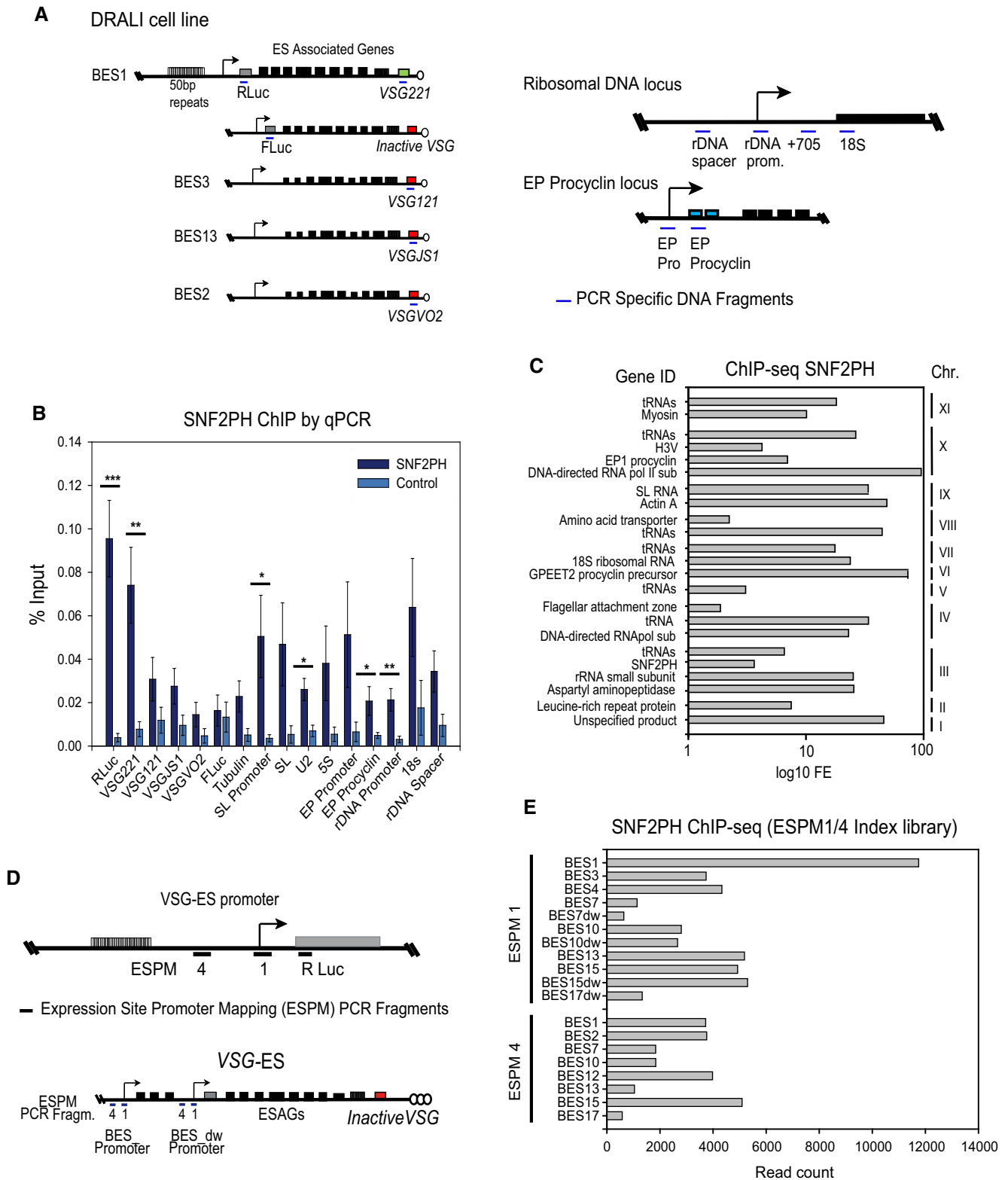

Figure 3.

◄

detected at *VSG* genes known to be located at silent telomeric ES position in this strain [2], such as *VSG121* (VSG in BES3), *VSGVO2* (BES2) and *VSGJS* (BES13) (Fig 3B). Altogether, the active *VSG221* gene immunoprecipitated more efficiently than all inactive *VSG* telomeric loci analyzed, suggesting SNF2PH associates preferentially with the active ES telomere. Additionally, SNF2PH was detected at other RNA pol I-transcribed loci, including rDNA and EP procyclin promoters. Occupancy of SNF2PH at the two promoters of the surface glycoprotein genes characteristic of mammalian and insect forms (VSG-ES and EP) implicates SNF2PH in regulation of developmental gene expression. Enrichment was also detected for the splice leader (SL) promoter ($P < 0.05$) and coding regions. However, SNF2PH was most enriched at the active VSG-ES chromatin compared with EP and rDNA promoters.

In eukaryotes, chromatin remodelers are detected at RNA pol II promoters and play important roles in their activity [26,27]. We investigated the presence of SNF2PH in chromatin across the genome, aside from the multiallelic VSG-ESs, to identify additional genes targeted by this protein. We compared quantitative enrichment profiles with ChIP-seq peak distribution and considered 0-mismatch error to avoid bias in polymorphic sequences within repetitive chromosomal loci, leading to defined peaks (q value < 0.05, Dataset EV1; Fig 3C). As demonstrated by quantitative ChIP, the site of enrichment corresponded to developmentally regulated loci EP and GPEET2 procyclin, and 18S ribosomal DNA and SL-RNA-related sequences. Interestingly, SNF2PH localizes at H3.V, a histone variant recently associated with VSG monoallelic expression [28] and its own coding sequence. SNF2PH was also significantly enriched at a substantial number of tRNAs gene arrays located in chromosomes XI, X, VIII, VII, V, IV, and III (Fig 3C); interestingly, tRNA clusters are known to function as a chromatin insulators in eukaryotes from yeast to human, reviewed in Ref. [29].

Figure 3B shows that SNF2PH is enriched at the active VSG-ES (BES1) locus by ChIP qPCR analysis using unique sequences like the Rluc reporter inserted downstream of the promoter and the

VSG221 gene at the telomeric end. However, we wished to investigate in detail a possible SNF2PH occupancy at the area adjacent to the promoter; nevertheless, highly homologues sequences shared among most of the VSG-ESs (BES1 to BES17 in ref. [2]) prevented this analysis using a simple ChIP-seq alignment. We have previously reported polymorphisms in the sequence at particular regions located at the core promoter and upstream the promoter region, referred to as ES promoter PCR fragments 1 and 4 (ESPM1 and ESPM4) (schematically represented in Fig 3D) [9]. These minor polymorphisms in the sequences allowed us to differentiate among different BES promoter regions. In particular, PCR fragment ESPM 4 and 1 yielded 14 different sequences when genomic DNA was used as template (Appendix Fig S5A in [9]) providing considerable covering of most of the BESs [2]. ChIP-seq data were generated from the immunoprecipitated chromatin with the anti-SNF2PH serum, which was then PCR amplified with ESPM1 and 4 PCR primers, and the products were deep-sequenced (see Materials and Methods). Reads were aligned to BES promoter sequences, and an index file was built by combining the sequences from the ESPM1 and 4 together in a single lane with the sequences from the corresponding BESs (BES1, 2, 3, 4, 7, 7dw, 10, 10dw, 12, 13, 15, 15dw, 17, 17dw described previously [2]). Next, using Bowtie software, the alignments of the reads were assigned to the BES promoter index file, and the number of reads aligning to each BES is shown in Fig 3D. These data showed that SNF2PH is enriched at the active BES1, at the ESPM1, which corresponds to the core promoter of the active BES1 (VSG-ES221). SNF2PH was detected to a lesser extent at other BES promoters, suggesting that SNF2PH is controlling inactive promoters as well (Fig 3D). Interestingly, the most prominent increase of read counts was found at the ESPM1 fragment region, where the actual ES promoter is located, suggesting SN2PH is associated with the active core ES promoter rather than the upstream promoter region (Dataset EV1). Together, these data indicate that SNF2PH is located at several BES promoters; however, it is most enriched at the active VSG-ES promoter (ESPM1 of the BES1) (Fig 3E).

SUMOylation functions to recruit SNF2PH at the active VSG-ES

SUMOylated chromatin-associated proteins are enriched upstream of the promoter region in *T. brucei* [9]. We assessed the importance of SUMOylation on SNF2PH targeting by mutating lysine residue 2 to alanine (K2A). K2 was selected as it was predicted as a modification site (SUMO V2.0 Webserver (http://sumosp.biocuckoo.org/) and is contained within the N-terminal region modified by the *in*

bacteria SUMOylation system (Fig 2). SNF2PH K2A was expressed with an HA tag from the endogenous locus (Fig EV2). We performed ChIP qPCR analysis using anti-HA rabbit antiserum after 48-h expression of HA-SNF2PH K2A, using the parental cell line as a control (Fig 4A). Expression of HA-tagged SNF2PH showed similar occupancy to SNF2PH, but by contrast the HA-SNF2PH K2A mutant was reduced in the active VSG221 telomeric locus (BES1) up to 4.6-fold compared to inactive VSG genes, VSG121, VSGJS1, VSGVO2
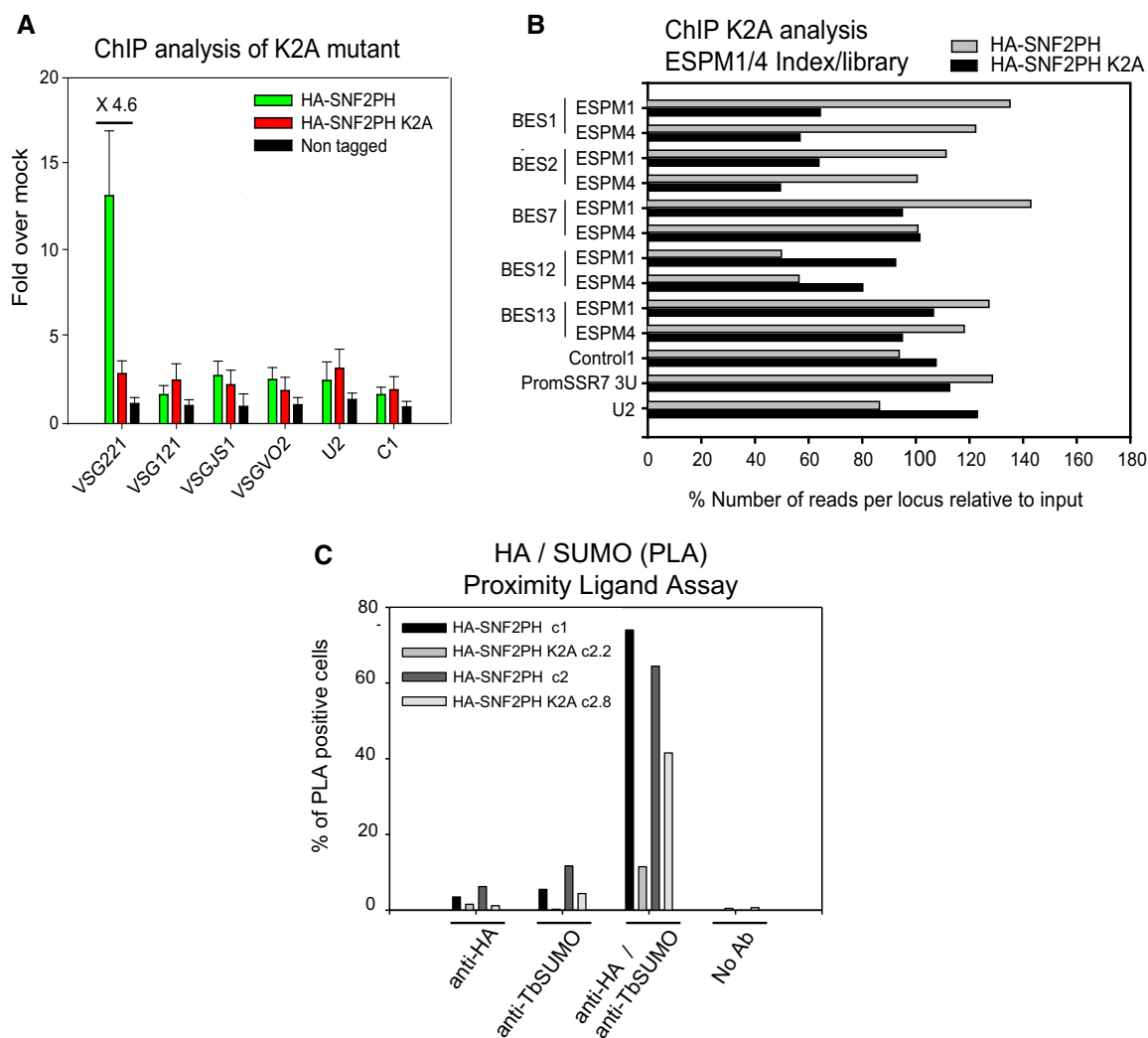
Figure 4. Expression of a SUMO-deficient mutant reduces SNF2PH occupancy at the active VSG-ES.

A ChIP analysis of mutant HA-SNF2PH K2A shows a reduction in the active telomeric VSG221 gene occupancy compare to the WT HA-SNF2PH. ChIP experiments were carried out using the anti-HA antibody and chromatin isolated from 3 independent clones expressing HA-SNF2PH K2A compared with a cell line expressing WT HA-SNF2PH. The mean of the ChIP analyses from three independent clones is represented as fold enrichment. A non-tagged cell line (Single Marker) was included as a negative HA control. Histogram shows the mean from three independent clones expressing the K2A mutant, and error bars represent means ± standard error of the mean (SEM).

B Reads per locus relative to input (%). ChIP analysis using anti-HA antibody in HA-SNF2PH and HA-SNF2PH K2A cell lines aligned with the ESPM1/4 index library containing the sequences from all the BES amplicons. ChIP analysis of the HA-SNF2PH versus the mutant HA-SNF2PH K2A generated ChIP libraries. Control 1; C1, PromSSR7 3U; Switch Strand Region promoter, Chr.7.

C *In situ* detection of SUMOylated SNF2PH using a Proximity Ligation Assay (PLA) is reduced in mutant SNF2PH K2A. Percentage of nuclei showing positive amplification signal in PLA assay analyzed as described before [9]. Histogram comparing the % PLA positive cells is represented separately for two independent clones overexpressing the HA-SNF2PH K2A mutant. Trypanosome transgenic cell lines usually vary considerably in protein expression level and kinetic, leading to a considerable degree of variability. Notwithstanding, both cell lines showed a decrease in PLA-positive nuclei upon expression of SNF2PH K2A. WT HA-SNF2PH c1 PLA-positive 73.98% (*n* = 246 total cells) versus HA-SNF2PH K2A Clone 2.2, positive PLA 11.47% (*n* = 514). WT HA-SNF2PH c2 PLA-positive cells 64.43% (*n* = 298) versus HA-SNF2PH K2A Clone 2.8 PLA-positive cells 41.39% (*n* = 387).

(located at the telomeric end of the BES3, BES13, and BES2, respectively [2]).

In order to determine whether K2A mutation also affects to SNF2PH occupancy at the promoter region, we constructed a ChIP-seq index library including all the BES sequences from PCR regions ESPM1 and 4 (Fig 3D) and map the reads PCR amplified from DNA immunoprecipitated using anti-HA antibodies and chromatin from cell lines expressing HA-SNF2PH and HA-SNF2PH K2A (Fig 4B). As controls we used gene Control 1, a promoter in chromosome 7, SSR7 (Strand Switch Region, SSR) (sequence defined by the prom SSR7 3U primers in Appendix Table S4) a RNA pol III (U2) were included after input normalization (Fig 4B and Dataset EV2). Sequence alignments using *Bowtie1* and 0 mismatch error yielded a number of reads aligned in BES ESPM1/4 libraries at the active promoter in BES1 and BES2 (highly sequence homologue) that were reduced for the cell line expressing the mutant K2A, to approximately ~0.5-fold. Conversely, at inactive promoters of BES the number of reads was increased (BES12) or no significantly changed (BES7 and BES13). This result is consistent with mutation K2A reduced protein SUMOylation, which decreases the occupancy of SNF2PH in the active promoter chromatin (Fig 4B).

We also assessed the effect of the K2A mutation on SUMO conjugation by PLA. The PLA signal for 3HA-SNF2PH (69.20% ± 4.77) was considerably greater than for the HA-SNF2PH K2A mutant (26.49% ± 15.02, $n = 2$), suggesting that K2 is required for efficient SUMOylation *in situ* (Fig 4C). We conclude that SNF2PH is mainly located at the active VSG-ES and that this specifically requires SUMO modification.

SNF2PH is a transcriptional activator that regulates VSG expression

To obtain direct evidence for function, we knocked down SNF2PH and analyzed the effect using the DRALI cell line. Western blot demonstrated depletion of the protein after 24 h RNAi (Fig 1B and Appendix Fig S4A) and a reduction to proliferation was also observed suggesting SNF2PH is essential for normal fitness (Appendix Fig S4B). RT–qPCR analysis after 48 h RNAi indicated significantly reduced levels of *RLuc* and *VSG221* mRNAs ($P < 0.05$), without changes to RNA pol II- or pol III-transcribed control loci C1 and U2, respectively. No reduction was detected in mature or pre-spliced rDNA + 780 RNAs (Fig 5A), suggesting SNF2PH depletion

decreases *VSG* expression specifically. Next, we analyzed TbRPAI (pol I largest subunit) occupancy in VSG-ES chromatin in cells depleted of SNF2PH by ChIP using anti-TbRPAI (Fig 5B). Upon SNF2PH depletion, we detected lower levels of TbRPAI recruitment to the active VSG-ES (BES1) including the active *VSG221* telomeric locus ($P < 0.05$). This also detected for the *RLuc* gene inserted downstream of the active promoter (2.94-fold). A lower decrease at the rDNA promoter and ribosomal 18S locus was not significant. Hence, SNF2PH is specifically involved in recruitment of the RNA polymerase to the active VSG-ES, suggesting SNF2PH is required for active VSG-ES transcription.

While a reduction of the reporter at the active VSG-ES promoter was detected upon RNAi in three independent clones, we found that *Fluc* activity from inactive VSG-ES promoters was clone dependent (Appendix Fig S4C). Relative expression of *FLuc* transcripts correlates with the FLuc expression level (Appendix Fig S4D). We performed RNA-seq analysis on paired groups of individual clones (Appendix Fig S4E and Dataset EV3), which showed variability in the VSG that is derepressed. Some, like VSG427-15 (BES10) and VSG-14 (BES8), but not all were upregulated, while the active BES1 telomeric *VSG221* gene was consistently downregulated, suggesting that where SNF2PH was depleted, random derepression of inactive VSG-ES promoters occurred. Thus, SNF2PH depletion induced derepression of a cluster of silent BES but not all, similar to recently reported for H3V-H4V KO [28]. As the PH domain is known to bind Histone 3 tails, the H3V KO [28] and SNF2PH depletion phenotypes may be related.

Next, we asked whether VSG protein expression was also reduced in SNF2PH depleted cells. VSG221 protein levels in three independent SNF2PH RNAi clones (Fig 5C) were significantly decreased ($P < 0.05$) (Fig 5D). Decreased VSG221 at the cell surface was also detected after SNF2PH RNAi by fluorescence-activated cell sorting (FACS) (Fig 5E). These results indicate that SNF2PH functions as positive transcription factor for VSG expression.

To identify factors associated with SNF2PH, we ectopically expressed triple-HA-tagged SNF2PH followed by affinity purification and LC-MSMS (Appendix Table S2 and Table EV1). Among the identified proteins, we found mRNA splicing factor TbPRP9 and nucleosome assembly protein, AGC kinase 1 (AEK1), a kinase essential for the bloodstream form stage [30] and a TP-dependent RNA helicase SUB2, (Tb927.10.540). Importantly, several previously identified VSG transcription factors, including Spt16 included in FACT complex (Facilitates Chromosome Transcription) [31], the

Figure 5. SNF2PH depletion results in a reduction of active VSG expression.

A   Reduced VSG-ES transcripts upon SNF2PH 48 h RNAi. Quantitative RT–qPCR analysis indicates 43% reduction of VSG221 mRNA, validated by RLuc reporter (*$P < 0.05$). No significant changes in ribosomal RNA transcripts were detected (18s and rDNA + 780).

B   Reduced RNA pol I occupancy in the active VSG-ES upon SNF2PH depletion. TbRPAI analysis was carried out in 48 h RNAi-induced cell lines and the parental cell line (DRALI). Statistical analysis shows a significant reduction of TbRPAI occupancy levels between parental and SNF2PH-depleted cells at the active VSG-ES (*$P < 0.05$). Data from three independent clones with standard error of the mean (SEM) are represented as fold over non-specific antiserum. NS; nonsignificant.

C   Quantitative Western blots of VSG221 expression in three independent SNF2PH RNAi clones using IR fluorescence. Anti-VSG221 and tubulin antibodies were incubated with the same blot and developed using goat anti-rabbit IgG 800 and anti-mouse IgG 700 Dylight (Thermo Fisher). A standard curve based on tubulin-normalized anti-VSG221 signal intensity was generated using different concentrations of parental cell extracts (R^2=0.99).

D   Quantitation of VSG221 expression relative to the parental cell line. Relative VSG221 protein levels appear to be reduced compared with the parental cell line (*$P < 0.05$).

E   FACS analysis of VSG221 expressing cells shows a decreased active VSG221 population upon SNF2PH depletion (continuous red line) performed on DRALI cell line. Control: DRALI cell line population expressing VSG221. SNF2PH-dox$^+$: induced SNF2PH-depleted cells. SNF2PH-dox$^-$: uninduced SNF2PH-depleted cells. Secondary Ab: population incubated with secondary antibodies as a negative control.

Data information: All data are reported as the mean ± SEM for biological replicates ($n = 3$). *$P < 0.05$ using two-tailed Student's *t*-test for paired observations.

A

Transcript level upon SNF2PH depletion

B

TbRPA1 ChIP

C

D

VSG221 Relative
Protein Level

E

VSG221
Surface Expression
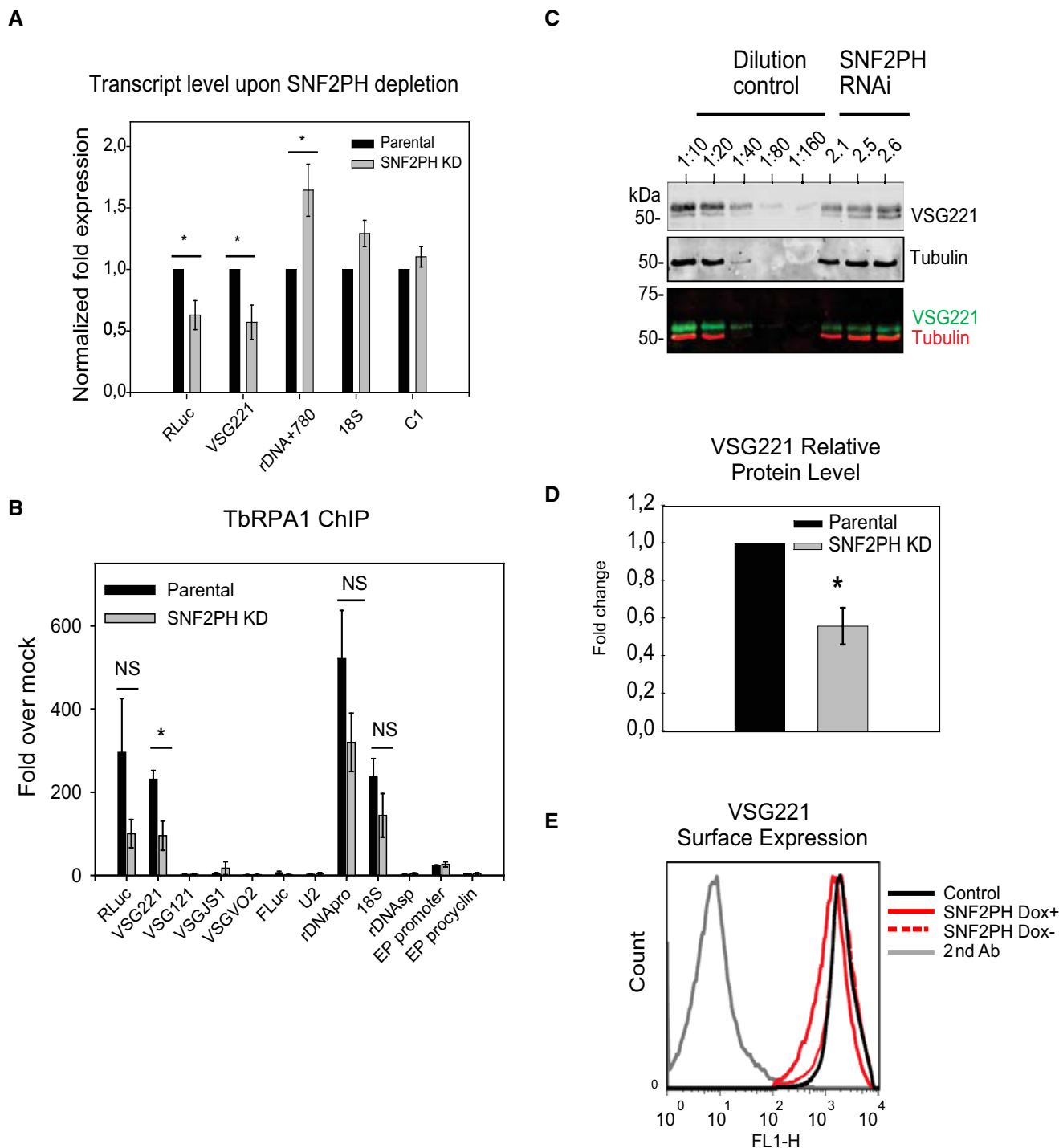
Figure 5.

proliferative cell nuclear antigen (PCNA) and a subunit from the Class I transcription factor A [32] were also identified (Appendix Table S2). Furthermore, RNA pol I subunit RPA135 and the RNA pol II RPB1 were also found to co-purify with SNF2PH suggesting possibly a transient interaction with RNA polymerases subunits, as previously described for yeast SNF2. Together, these data suggest that SNF2PH occupies a central position in VSG transcription regulation and links to RNA pol II transcription.

SNF2PH is required for the maintenance of the bloodstream stage expression profile

The above results suggest that SNF2PH regulates VSG expression however also interacts with a RNA pol II subunit and RNA binding proteins suggesting SNF2PH may regulates the expression of additional genes. In addition, SNF2PH protein levels in the bloodstream form are decreased in the insect form suggesting a possible function

in development. To investigate that possibility, we knocked down SNF2PH in bloodstream trypanosomes. Transcript levels for EP procyclin ($P < 0.01$), PAD1 and PAD2 (Protein Associated with Differentiation 1 and 2) were significantly upregulated ($P < 0.05$), whereas MyoB levels decreased ($P < 0.01$) (Fig 6A). Relative mRNA levels of ribosomal 18S and C1 were unchanged. An increase in EP procyclin and PAD1/2 transcription followed by a reduction in MyoB, similar to the transcript profile of the stumpy form, a transition to the insect stage [33,34].

Furthermore, we investigated whether SNF2PH depletion influences EP procyclin replacement dynamics during *in vitro* differentiation from the BF to the PF. We triggered differentiation with 3 mM *cis*-aconitate (Fig 6B). Under these conditions, SNF2PH-depleted cells expressed higher procyclin transcripts at 4 h of induction, whereas VSG221 mRNA levels were reduced compared to the parental cell line. Surface protein analyses of EP procyclin and VSG221 in SNF2PH-depleted cells by IF, found a ~2-fold faster replacement of VSG by EP procyclin (Fig EV3).

To investigate additional genes regulated by SNF2PH, we carried out RNA-seq upon knockdown in the bloodstream form (Fig 6C). Eighteen representative differentially expressed genes (DEG) with a false discovery rate (FDR) < 0.05 were found out of 8673 genes from two experimental replicates and 221 transcripts with a $P < 0.05$ (Dataset EV4). Among them, clear upregulation of genes related to the procyclic form, including procyclin genes (EP1-2, GPEET2), followed by procyclin-associated genes (PAG1, 4 and 5) and expression site-associated genes (ESAG2 and 8), some of them consistent with previous analysis in procyclic forms [33] (Appendix Table S3). We also found upregulation of protein associated with differentiation 1, 2 (PAD1 and PAD2), zing finger family ZC3H18, and the receptor-type adenylate cyclase GRESAG4-related transcripts. Conversely, genes with normally higher expression in bloodstream form were downregulated,

including the alternative oxidase and glucose transporter THT-2 (Dataset EV4).

Given that SNF2PH expression is clearly reduced in the procyclic stage (Fig 1A), we sought to investigate the gain-of-function phenotype induced by ectopic expression of SNF2PH in the procyclic form. Interestingly, we detected an increase in silent telomeric VSG121, VSGBn2, and VSGJS1 transcripts from two independent clones in the procyclic form where no VSG is normally expressed (Fig 6D).

To determine possible changes to global gene expression induced by SNF2PH overexpression, we used RNA-seq and compared the overexpressor and parental procyclic cell line (Fig 6E). We found increased expression of ESAG3 from several BES, ranging from 2.5 to 5.2- log FC, while the promoter adjacent transferrin-binding protein ESAG6/7 increased 1.7 to 3.2- log FC (Fig 6E and Dataset EV5). Of the total mapped reads, 55.08% were linked to telomeric bloodstream ESs (BESs), including BES1 (BES40 ~5.2- log FC), BES4 (BES28/98 ~4.5- log FC), BES2 (BES129/126 ~3.8- log FC), and BES13 (BES56/153/51/4 ~2.6- log FC) (Fig 6E and Dataset EV5). Importantly, no expression of VSG basic copies genes, located within internal chromosomal arrays, was detected, ruling out global chromatin deregulation. These results show that, in procyclic forms where no VSG is expressed, overexpression of SNF2PH induced upregulation of telomeric VSG-ES (BES) transcripts, suggesting that SNF2PH functions as a central regulator of the BES. Thus, in BF, where SNF2PH is highly expressed, this chromatin remodeler likely functions to promote and maintain the expression of VSG-ES (BES).

Next, we analyzed the relevance of the SNF2PH plant homeodomain and overexpressed a truncated form of SNF2PH lacking the PH domain (SNF2ΔPH) (Fig 6F). RNA-seq analysis detected 737 genes (FDR < 0.05 and $P < 0.01$, out of 7,918 genes) differentially expressed after overexpression of SNF2ΔPH versus 118 genes

---

Figure 6. SNF2PH is required for maintaining infective form surface protein expression.

A Depletion of SNF2PH (48 h) in the bloodstream form results in an increase of procyclin and proteins associated with differentiation (PADs) transcripts. Procyclin, PAD1, and PAD2 mRNAs are upregulated when SNF2PH is depleted. Results are the average from three independent clones and data normalized with U2 mRNA. Error bars represent means ± SEM. (*$P < 0.05$, **$P < 0.01$ using two-tailed Student's *t*-test for paired observations.

B SNF2PH KD cells differentiate to procyclics more efficiently. The procyclin transcript is increased in SNF2PH KD-depleted cells during *in vitro* differentiation compared with parental cell line. Parental and SNF2PH-depleted cells were treated with 3 mM cis-aconitate and temperature shift for 4 hours. Quantitative RT–PCR data from two independent clones were normalized against C1 (RNA pol II-transcribed) as a housekeeping gene. Error bars represent means ± standard deviation (SD).

C Scatter plot for differentially expressed genes (DEG) from RNA-seq analysis upon SNF2PH depletion (FDR < 0.05). Non-DEG: Non-differentially expressed genes. Up: upregulated genes. Down: downregulated genes.

D Ectopic expression of SNF2PH in procyclic form upregulates telomeric VSG mRNAs. Histogram showing the relative expression of mRNA measured by qRT–PCR of VSG genes located in different telomeric VSG-ESs (BES) genes after 48 h of induction of SNF2PH overexpression in procyclic form. This analysis included mRNAs from telomeric VSG genes (BF stage-specific) not expressed normally in the insect procyclic form, including the VSG121 (BES3), VSGJS1 (BES13) VSGBR2 (BES15), and VSGVO2 (BES2). Data from two independent clones and parental controls are represented as normalized fold expression relative to C1 (RNA pol II transcribed) as a housekeeping gene. Error bars represent means ± SD from technical replicates for each independent clone.

E Ectopic expression of SNF2PH in procyclic induces expression of bloodstream form surface proteins. Scatter plot for differentially expressed genes (DEG) in RNA-seq analysis shows upregulation of telomeric BESs. A significant increase of BES-associated genes (ESAGs) linked to telomeric BESs was detected after full-length SNF2PH overexpression in the procyclic form. Data from at least two experimental replicates are represented as log$_2$ fold change (FC) for genes with FDR < 0.05 and $P < 0.001$ after correction with the uninduced procyclic cell line. Relative mRNA levels of procyclic cells after 48 h of SNF2PH overexpression increase telomeric BESs, including VSG221-ES (BES1; TAR40), VSG121-ES (BES3; TAR15), VSGJS1-ES (BES13; TAR56), and VSGVO2-ES (BES2; TAR129), VSGR2-ES ES (BES15; TAR126), see (2) for detailed BES and TAR nomenclature (Dataset EV5). Ectopic expression of SNF2PH full-length also induced invariant surface glycoprotein 65 (ISG65) genes (Dataset EV5). Data from two independent clones and parental controls are represented as normalized fold expression relative to C1 (RNA pol II transcribed) as a housekeeping gene.

F Expression of bloodstream form surface proteins requires SNF2PH plant homeodomain. EP procyclins and other insect stage-specific markers (FDR < 0.05, $P < 0.001$) are expressed in the SNF2PH mutant form lacking of the plant homeodomain (SNF2ΔPH). No BF surface proteins like VSG-ES (BES related or ESAGs) neither invariant surface glycoproteins (ISG65) genes were induced under ectopic expression of the SNF2ΔPH mutant.

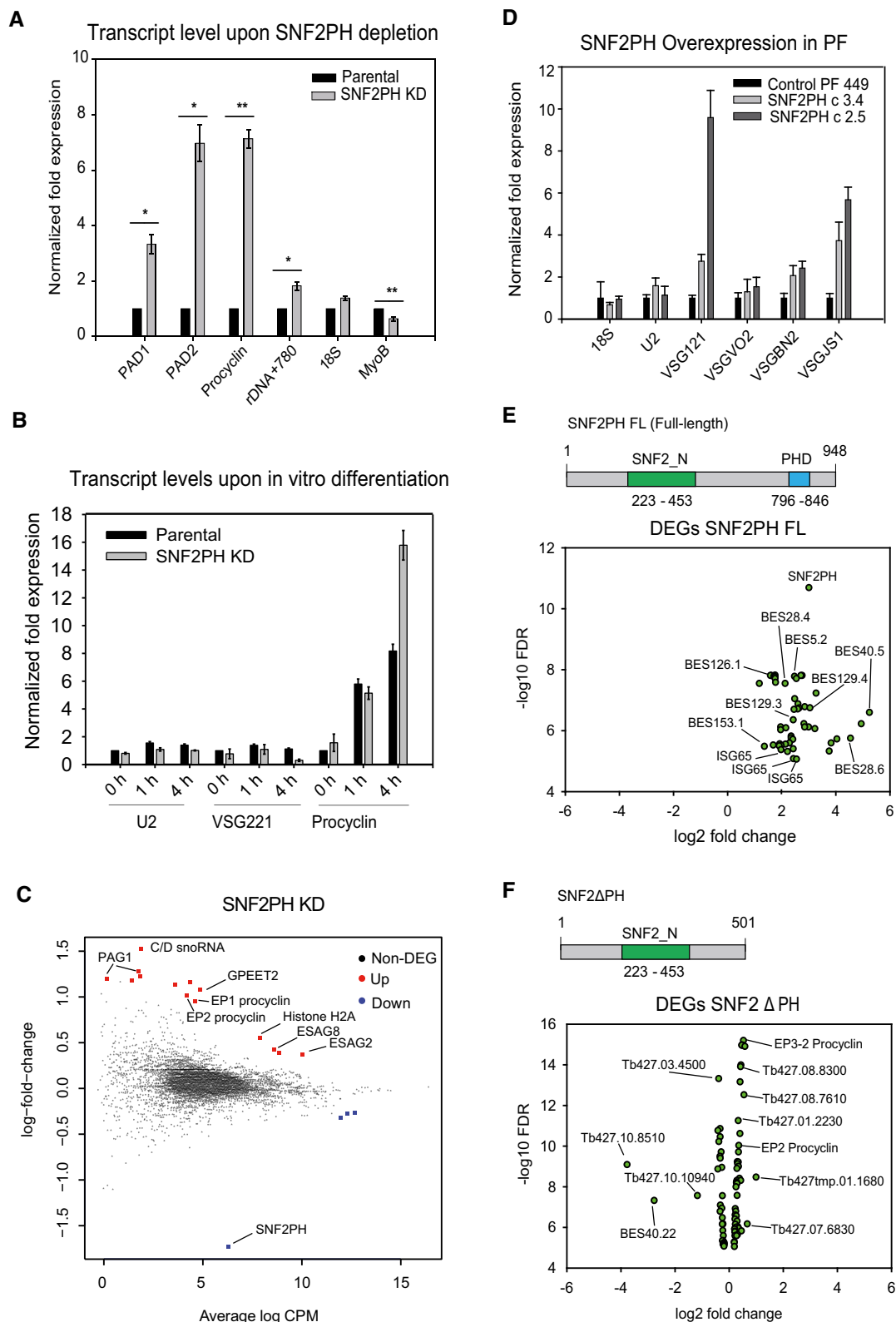**Figure 6.**

(FDR < 0.05 and *P* < 0.001, out of 8,451 genes) for SNF2PH full length, illustrated in Fig EV3C and D. Most significantly, induction of BES-related genes did not occur in cells overexpressing the

SNF2ΔPH mutant lacking the PH domain (Fig 6F), suggesting that the PH domain is required to provide SNF2PH specificity to bind and recognizes VSG-ES (BES) chromatin (Fig 6F and Dataset EV5).

Significantly, beyond impacts to BES expression, invariant surface glycoprotein ISG65 genes, transcribed exclusively in the bloodstream form, were also significantly increased up to 6-fold in procyclics ectopically expressing SNF2PH compared to the parental procyclic cell line (Fig 6E). Importantly, ISG gene transcripts were not detected in cell lines expressing SNF2ΔPH (Dataset EV5). We do not know whether ISG65 is expressed at the surface of these cells, but consider it unlikely as we previously demonstrated that ectopically expressed ISGs are rapidly degraded in insect stage cells [35], as well as with ectopic expressed mRNA VSG. As VSG is transcribed by the RNA pol I and ISG65 by RNA pol II, this suggests that SNF2PH acts beyond the telomeric BES, and may act as a global transcriptional activator of surface protein genes irrespective of the RNA polymerase involved.

These results suggest that the PH domain is essential to direct SNF2PH to specific bloodstream form surface proteins genes, promoting transcriptional activation. In metazoans, transcription factors containing homeodomains determine cell fates during development, and this is the chromatin-binding domain necessary to regulate downstream target genes. Our results provide an important contribution of the plant homeodomain in recognition epigenetic chromatin patterns to regulate gene expression in trypanosomes, an early-branching eukaryote.

SNF2PH is downregulated in quiescent stumpy forms

The stumpy form of pleomorphic trypanosomes is pre-adapted to the metabolism required by the procyclic form for survival within the insect vector. Given upregulation of stumpy form markers in SNF2PH knockdown cells, we examined SNF2PH dynamics during the stumpy to procyclic form transition. Recent evidences suggest that the AMP-dependent kinase, AMPKα1, is a key regulator of the development of quiescence in bloodstream form trypanosomes [42] reviewed in [43]. Upon AMPKα1 activation, stumpy-like differentiation was induced in a monomorphic cell line. Quantitative Western blots and RT–qPCR analyses of monomorphic cells treated with an AMP analog showed a significant decreased in SNF2PH proteins levels and transcripts (Appendix Fig S5A). Reduction of SNF2PH protein levels in stumpy-like cells obtained after AMPKα1 activation by AMP suggests AMPK pathway negatively regulates SNF2PH expression.

We also assessed the transcriptional profile in stumpy-like forms induced by AMP treatment and compared with SNF2PH knockdown. qRT–qPCR confirmed a transcriptional profile characteristic of the stumpy form in untreated SNF2PH knockdown versus AMP treated cells, in which relative mRNA levels for SNF2PH transcripts were downregulated ($P < 0.01$; Appendix Fig S5B). Interestingly, stumpy form-like transcriptome changes were more prominent in cells treated with 5′-AMP and depleted for SNF2PH compared with 5′-AMP alone and 5′-AMP untreated knockdown cells.

An important question, however, is whether SNF2PH downregulation occurs naturally during *in vivo* differentiation of a wild-type pleomorphic strain. Bloodstream pleomorphic trypanosomes undergo differentiation from the proliferative to the quiescent stumpy form throughout mice infection. SNF2PH protein levels analysis in the pleomorphic AnTAT 90.13 strain decreased at 4-5 days postinfection (Appendix Fig S5C), whereas AMPKα1 was

fully activated as previously described [42]. Taken altogether, these data suggest that SNF2PH is negatively regulated during transition to stumpy form and mechanistically linked to AMPKα1 activation by an undefined mechanism.

Discussion

Antigenic variation in African trypanosomes is mediated by complex regulatory mechanisms that secure monoallelic expression of a single VSG, the fundamental basis of immune evasion. The transcriptional state of VSG genes is maintained through several generations and is the product of epigenetic mechanisms and post-translational modifications (PTM) that mark the active VSG genes. Several PTMs are associated with silencing the inactive VSG-ESs [3], but SUMOylation is the only known modification associated with the active VSG-ES [9]. A concentration of SUMOylated proteins [9] is located adjacent to the expression site body where VSG gene transcription occurs. Additionally, several chromatin-associated SUMO-modified proteins are enriched at the active VSG-ES, and chromatin SUMOylation is also required for efficient RPAI recruitment [9].

Epigenetic factors are likely regulators of VSG transcription; silencing of TbISWI, HDACs, and additional chromatin remodelers such as ASF1, CAF1, and FACT, consistently results in similar phenotypes, i.e., the derepression of normally silent VSG-ES subtelomeric regions (reviewed in [4]). By contrast, pharmacological inhibition of bromodomain chromatin remodelers leads to the upregulation of both silent and basic copy VSG genes, suggesting a role in maintaining global chromatin organization, rather than specific action at the active ES [36]. Further, post-translational histone modifications are associated with the repression of silent VSG-ESs (see [37] for a review). Interestingly, proteins regulating nuclear architecture also alter monoallelic VSG expression, including the lamina proteins NUP-1 [38] and cohesin, where silencing leads to a ~10-fold increase in VSG switch frequency [24]. Deletion of histone variants H3.V and H4.V highly increases VSG switch frequency *via* homologous recombination, and the telomeric VSG-ES loci are confined to distinct compartments within the nucleus [28]. Altogether, these data support a model where epigenetically marked VSG-ES chromatin occupies a unique subnuclear compartment, facilitating monoallelic expression [6].

The product of Tb927.3.2140 is a SUMO-conjugated protein unregulated in the bloodstream form. Tb927.3.2140 possesses domains homologous with the SWI/SNF family of chromatin remodelers, e.g., SMARCA1, also known as the global transcription activator SNF2L1, and SNF2 helicase domains. Thus, we named the gene product SNF2PH. The SNF2 family N-terminal domain participates in many processes [16], while the chromatin SWI/SNF remodeler superfamily regulates either specific or global gene expression. SNF2PH also contains helicase domains homologous to transcriptional activators HsSMCA5 and MsSMCA1 of the ISWI subfamily (Appendix Fig S1). Interestingly, SNF2PH contains a homeodomain that interacts directly with histone tails to regulate development, the plant homeodomain (PHD), which is present in histone methyltransferases, such as MmHMT3 and HsNSD3 (Appendix Fig S1). The *H. sapiens* chromatin remodeler CHD4-based NuRD also contains a PHD [39] and was shown to act in

both a positive and negative manner in regulating cell-specific differentiation [16]. Multiple lines of evidence suggest that SNF2PH has an architecture typical of transcriptional regulators and chromatin readers and similarly acts in both, positive and negative manners, to regulate developmental stage-specific gene expression of surface proteins in trypanosomes.

SUMO modification acts as a switch for the localization of SNF2PH, and modified SNF2PH is recruited to the active VSG-ES promoter, while unmodified SNF2PH is present at silent VSG promoters. While we cannot be certain if this is cause or consequence, loss of function associated with a non-SUMOylated mutant of SNF2PH suggests an active role for SUMOylation in targeting. SNF2PH depletion leads to downregulation of the active VSG-ES, indicating a possible function as a positive regulator. This conclusion is supported by an increased frequency of sequences related to BES1 in the ChIP-seq data, as well as upregulation of procyclin and PAD genes after SNF2PH knockdown, suggesting a suppressive function in the insect stage. Interestingly, ChIP-seq analysis revealed SNF2PH occupancy at the H3.V locus, which suggests that SNF2PH regulates H3.V expression, a histone variant crucial for regulating VSG expression [28].

SNF2PH associates with several proteins, including histone chaperone Spt16 [40] and proliferative cell nuclear antigen (PCNA), essential for replication [41], suggesting a link between SNF2PH and cell cycle progression. Another significant interactor is the CITFA-4 subunit of the RNA pol I promoter binding complex [31]. SNF2PH also shows an association with RNA pol I subunit RPA135 and RNA pol II subunit RPB1, the latter consistent with SNF2PH occupancy at pol II-transcribed loci (Fig 3B). This prompts us to speculate that SNF2PH-pol II association facilitates accessibility of transcription factors to regulate RNA pol II transcription.

SNF2PH depletion leads to expression of insect form genes, which suggests a role in maintaining the BF chromatin state by opposing control of the two major developmentally regulated surface gene families, procyclins and VSGs. Importantly, overexpression of SNF2PH in the procyclic form leads to the expression of bloodstream form telomeric VSGs and ESAGs, requiring the PH domain to promote telomeric expression in a repressed VSG transcriptional background. Consequently, this domain is apparently required to maintain bloodstream form through developmental regulation [19].

The stumpy form is a pre-adaptation to the insect host and requires AMPKα1 activation [42]. 5′-AMP analog treatment leading to enhanced AMPKα1 activation in monomorphic cells resulted in SNF2PH downregulation. Whereas AMPKα1 activation promotes differentiation, reduced SNF2PH expression in stumpy forms observed in wild-type pleomorphic trypanosomes during mice infection confirmed the biological relevance. This result rules out a possible SIF-independent induction of differentiation by VSG-ES transcription attenuation described previously [44].

Interestingly, SNF2PH is enriched at tRNAs clusters on several chromosomes (Fig 3C). tRNAs can act as insulators to prevent the spread of silencing in *S. cerevisiae* [45,46] and in mammals to prevent enhancers from activating promoters [47,48]. We detected low colocalization of SNF2PH with GFP-tagged active VSG-ES (Fig 1D), and even lower association in 2K1N cells (G2) (Fig EV1A), which may suggest that the association of SNF2PH with the active VSG-ES locus occurs preferentially in G1. Although such insulator function has not been described in trypanosomes, the presence of SNF2PH at tRNA loci during interphase, as some insulator proteins [49], suggests a tentative model where SNF2PH/tRNAs could act as a chromatin domains organizer to maintain bloodstream stage-specific surface protein gene expression.

The ISG65 genes are transcribed by RNA pol II from polycistronic arrays located in core chromosomal regions, while VSG genes are transcribed by RNA pol I from the VSG-ES (BES) telomeric locus. Notwithstanding, the chromatin remodeler SNF2PH, utilizing the PH domain as epigenetic reader, recognizes both distinct gene families. This is suggested by SNF2PH ectopic expression in procyclic forms, which led to a developmental epigenetic reprogramming, similar to homeodomain proteins in other organisms. Furthermore, ISGs are transcribed at similar levels for all allelic variants, while the VSG is monoallelically transcribed at one out of 15 different telomeric BESs [2]. Interestingly, SNF2PH overexpressed in procyclic forms lacking SUMO induced then expression of all BES (Fig 6E), and monoallelic expression of VSG genes was not achieved. We speculate that SNF2PH SUMOylation is likely the modification that SNF2PH acquires at the nuclear body ESB, where VSG-ES transcription occurs [6,9], allowing SNF2PH to recognize and activate a single BES telomere among the VSG multiallelic gene family.

In sum, SNF2PH requires SUMO modification to function as a transcriptional activator of VSG-ES monoallelic expression, and the PH domain is required for this and for maintaining the mammalian infective form surface protein coat, ensuring continuous and proper VSG and ISG surface display, essential for pathogenicity.

Materials and Methods

Trypanosome strains and cell lines

Trypanosma brucei brucei, bloodstream form (Lister 427, antigenic type MiTat 1.2, clone 221a), 427 procyclic form and the pleomorphic AnTAT 90.13 were used in this study. The dual-reporter cell line, DRALI, contains the *Renilla* luciferase (*RLuc*) gene inserted 405 bp downstream of the active VSG-ES promoter and the *Firefly* luciferase (*FLuc)* gene downstream of an inactive VSG-ES promoter. The insertion site was checked by sequencing the flanking region from DRALI genomic DNA confirming *RLuc* inserted in the active VSG221-ES (BES1), whereas *FLuc* was inserted downstream of the inactive VSG-ES promoter BES15/TAR126 VSGbR-2/427-11. The VSG221-ES GFP-tagging and YFP-TbRPB5z fusion were previously described [24].

Recombinant proteins and monoclonal antibodies

C-terminal fragment of SNF2PH (Tb927.3.2140) was amplified by PCR, and the PCR product was cloned into *Bam*HI and *Hind*III sites of pET28a vector (Novagen) expressed as a C-terminal His tag (Appendix Tables S5 and S6). The recombinant protein was purified using NI Sepharose Fast flow 6 (GE Healthcare) and inoculated into mice to generate anti-SNF2PH (11C10E4) monoclonal antibody (mAb), using standard procedures. Hybridomas were screened against the recombinant protein by ELISA and further confirmed by Western blot analysis using trypanosome protein extracts that recognized the protein of the expected size.

Hybridoma cell line 11C10E4 was grown as ascites. SNF2PH polyclonal antibody was obtained by affinity purification from rabbit antiserum after several inoculations of the recombinant protein using an Aminolink column (Pierce), following the manufacturer's instructions. Anti-TbSUMO (1C9H8) monoclonal antibody was generated as previously described [9].

3xHA tagging of SNF2PH versions

A *T. brucei* bloodstream form cell line expressing a 3xHA-tagged SNF2PH was developed by replacing both copies of the endogenous gene. For the 3xHA K2A, one allele was replaced by the mutant version due to cell viability. Procyclic form cells carrying both 3xHA-tagged full-length and truncated SNF2PH isoforms were ectopically expressed from the ribosomal spacer. Cloning procedures are detailed in Appendix Tables S5 and S6.

RNAi experiments

SNF2PH RNAi construct was made using the p2T7Bla vector [50]. Since most of the RNAi constructs using this vector are leaky, comparative analyses always included addition of the dox induced (+) and uninduced (−) RNAi in the parental cell line (DRALI). Amplified PCR fragment corresponding to 1113-bp of C-term SNF2PH ORF was cloned into *Bam*HI and *Hind*III sites of p2T7Bla and transfected into the dual-reporter cell line DRALI (Appendix Tables S5 and S6). Synthesis of dsRNA was induced by adding 1 µg/ml of doxycycline. At least three independent clones were analyzed, and protein depletion was confirmed by Western blot using specific antibodies.

RT–qPCR

RNA isolation, cDNA synthesis, and qPCR were performed as previously described [9]. Relative expression levels were referred to a control (parental cell line) and normalized against a housekeeping gene (U2, pol III-transcribed gene), using the software Bio-Rad CFX Manager Software. Experimental conditions were performed in triplicate and analyzed by Student's *t*-test. A detailed primer list is found in Appendix Table S4.

RNA-seq analysis

Total RNA from at least two independent biological replicates of both SNF2PH knock down (BF) and overexpression (PF) after 48 h of doxycycline induction was used to generate a library from poli (A) + mRNA isolated fragments. Libraries were sequenced on an Illumina NextSeq 500 platform (150 cycles) in paired-end mode with a read length of 2×76 bp and sequence depth of approximately 50 million reads per sample. The miARma-Seq pipeline [51] was used to analyze all transcriptomic data. In detail, this pipeline contains all needed software to automatically perform any kind of differential expression analysis. It uses fastqc to check the quality of the reads and aligned them using hisat2 on *T. brucei* TREU427 reference genome (TritrypDB release 39). Subsequently, the aligned reads are quantified and summarized for each gene using featurecounts. Finally, gene counts are analyzed using the edgeR package from Bioconductor. In such a way, all samples were size corrected in order to be comparable and then normalized using the TMM

method from the EdgeR package. TMM values for each gene were used for the differentially expression analysis. RPKM values for each gene were calculated from the normalized read counts values using the rpkm method from edgeR. Genes transcripts isolated form uninduced versus induced SNF2PH RNAi cells with a [$\log_2$FC] ≥ 1 ($\log_2$ of Fold Change) and FDR ≤ 0.05) were considered as differentially expressed. Additionally, miARma-Seq [51] generated a volcano plot to facilitate the identification of genes that felt higher variation in expression.

Chromatin Immunoprecipitation (ChIP) and ChIP sequencing (ChIP-seq)

Bloodstream form *T. brucei* cultures were fixed and processed as previously described [9]. Pre-cleared chromatin (5×10^7 cells per IP) was incubated with each antibody (90 µg anti-SNF2PH, 6 µg anti-TbRPAI, 6 µg of rabbit anti-HA tag antibody (abcam), and 90 µg of an unspecific antiserum). The immunoprecipitated products were reverse crosslinked, and the extracted DNA was analyzed by quantitative PCR (qPCR). For ChIP-seq analysis, the protocol was scaled for a final concentration of ~5 ng of immunoprecipitated DNA. To compare the amount of DNA immunoprecipitated to the total input DNA, 10% of the pre-cleared chromatin saved as input was processed with the eluted immunoprecipitated products before the crosslink reversal step. Quantitative PCR was performed using SYBR green Supermix (Bio-Rad) in a CFX96 cycler (Bio-Rad). IP percentages were determined as previously described [9]. At least three independent experimental assays were displayed and analyzed by Student's *t*-test. A detailed primer list is detailed in Appendix Table S4.

Generation of the ChIP-seq library

ChIP-seq analysis in Fig 3D

Immunoprecipitated DNA (~5 ng) from each condition was evaluated in a 2100 BioAnalyzer to assess fragmentation size and subjected to end-repair enzymatic plus dA-tailing treatments further to be ligated to adapters using the Illumina TruSeq DNA Sample preparation kit, following the manufacturer's instructions. Adapter-ligated libraries were enriched with 15 cycles of PCR using Illumina PE primers and purified with a double-sided SPRI size selection in a range below 300 bp. Libraries were sequenced in an Illumina NextSeq 500 platform leading to a 650,000 reads per sample. The raw reads were processed using the miARma-Seq pipeline [51] to measure quality, adapter sequence removal and read alignment. Briefly, this software first assessed the quality of the sequences using FASTQC tool kit. After that, adapter sequences were removed using the cutadapt utility. Once reads were processed, they were aligned against the Trypanosoma brucei Lister 427 genome obtained from TriTrypDB version 34 using the BWA aligner with default parameters. Later, final results obtained from miARma-seq were processed with macs. Therefore, each paired sample (-f BAMPE) chip (-t) was processed against the input sample (-c) to eliminate general peaks in both types of samples using as organism size $2.7e^7$ (-g). The correspondence of peaks between both types of samples (HA-SNF2PH and HA-SNF2PH K2A) and with gene sequences was carried out with the intersectBed script from bedtools using the gene annotation provided by TriTrypDB version 34 and visualized in GB browser. A starting pool of 8 amplicons 18S, U2, C1, prom SSR7,

ESPM-1/4, VSG221 (BES1), and VSG121 (BES3)) were combined together in a single lane per condition (including respective inputs). Coding sequences for 18 and U2 were used as reference genes to evaluate the immunoprecipitation efficiency in each experimental case. Differential peak distribution was represented as fold enrichment relative to input and assessed by –logP value, considering a 0 nucleotide mismatch to discern telomeric sequences.

ChIP-seq analysis in Fig 3E

To discriminate among the BES promoters, we carried out ChIP-seq analysis using selected PCR ES Promoter Mapping (ESPM) regions known to have sequence polymorphisms among different BESs, as previously described [9]. A pool of amplicons of sequences of the promoter regions ESPM1 and ESPM4 (Fig 3D, defined with the primer include in Appendix Table S4) containing the corresponding sequences from the BES1, BES2, BES3, BES4, BES7, BES7dw, BES10, BES10dw, BES12, BES13, BES15, BES 15dw, BES17, BES17dw, and 18S, U2, C1 as control were combined together in a single lane to build an index file (ebwt), next the alignment was done with bowtie -S -n0 command to consider a 0 nucleotide mismatch to distinguish among few nucleotide sequence differences in each BES, as described before [9]. The actual number of reads aligned on each BES in represented in the histogram of Fig 3E.

Cell extracts and Immunoblots

Parasite cultures were collected by centrifugation and washed once in Trypanosome Dilution Buffer (TDB) with 1X protease inhibitor cocktail (Roche) and 20 mM *N*-ethylmaleimide (NEM) and pellets were processed as previously described [9]. Western blot membranes were incubated with anti-SNF2PH mAb ascites (1:1,000), anti-SNF2PH affinity-purified antiserum (1:1,000) and monoclonal anti-HA high affinity (1:500, clone 3F10, Roche Applied Science). Mouse monoclonal anti-TbSUMO mAb ascites (1:1,000), anti-Tubulin mAb (1:5,000), anti-MVP mAb (1:1,000), rabbit anti-VSG221 antiserum (1:50,000), and rabbit mAb Phospho-AMPKα (Thr172) (40H9) (1:1,000, Cell Signalling technologies) were used as described previously [9,34,42].

Quantitative western blots

Quantitative Western blots analyses were performed as previously described [9]. Membranes were incubated with anti-VSG221 (1:50,000), anti-SNF2PH affinity-purified antiserum (1:1,000), Tubulin (1:5,000), and anti-MVP mAb (1:1,000). A standard curve based on Tubulin-normalized anti-VSG221 signal intensity was generated using different concentrations of parental cell extracts (R^2 = 0.99). The standard curve regression was used to determine VSG221 expression levels in SNF2-depleted cell lines. For both detection of AMPK phosphorylated levels and SNF2PH in *in vitro* and *in vivo* assays, a MVP-normalized anti-SNF2PH and/or anti-p-AMPK were used to quantify differences in signal intensity compared with the parental condition.

Purification and identification of TbSUMO conjugates

To identify SUMO conjugates from a bloodstream form, *T. brucei* cell line expressing an 8xHis and HA-tagged version of SUMO (Tb927.5.3210) was developed by replacing both copies of the endogenous gene, see Appendix Tables S5 and S6. Purification of conjugates was performed in denaturing conditions by nickel affinity chromatography. After imidazole elution, the urea concentration was decreased and SUMOylated proteins were subjected to a second affinity purification step using an anti-HA agarose resin. Final conjugates were analyzed by mass spectrometry, and processing data was performed as previously described [34].

Purification of protein complexes and identification by Nano-LC-MS/MS

A total cell mass of 4.0×10^{10} procyclic form cells, strain *T. brucei* 449 expressing the 3HA-tagged version of TbSNF2PH was induced with 1 μg/ml doxycycline during 48 h and harvested at 1,400 *g* rpm during 10 min at 4°C. The pellet is washed in 50 ml PBS 1× including protease inhibitors and subjected to cryogenic grinding, resulting in a lyophilized powder with all nuclear components, as previously described [52]. Immunoprecipitation assays were performed with 50 mg of lyophilized powder and resuspended immediately in 1 ml of Lysis buffer (20 mM HEPES pH 7.4, 50 mM sodium citrate, 1 mM MgCl$_2$, 10 μM CaCl$_2$, 2× protease inhibitor cocktail (Roche), and 0.1% CHAPS), followed by three cycles of sonication of 15 s at 50 W and centrifuged at 20,000 *g* during 10 min at 4°C. The supernatant containing the nuclear fraction was incubated with 10 μl (0.1 mg) of previously equilibrated HA-Magnetic beads and incubated during 2 h at 4°C on rotation and washed three times, preserving the same buffer conditions. HA-Magnetic beads were eluted at 99°C during 5 min with 15 μl of NuPAGE SDS Sample buffer (Life technologies) with 1.5 μl of NU PAGE SDS Sample Reducing Agent (Life technologies) and denatured at 99°C during 5 min prior to being analyzed in an SDS–PAGE gel with silver stain. Sample preparation for MS analysis was eluted with 50 μl of 2% SDS and 20 mM Tris–HCl pH 8.0 at 72°C and precipitated with 100% ethanol. After centrifugation, the sample was subjected to tryptic digestion and reductive alkylation of Cys groups with 50 mM iodoacetamide and finally vacuum-dried to be dissolved in 1% acetic acid. Then, tryptic peptides mixtures were injected onto a C-18 reversed phase nano-column (100 mM ID, 12 cm, Teknokroma) and separated in a continuous acetonitrile gradient. Eluted peptides from the RP nano-column were fragmented in a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific). For protein identification, the mass spectra were deconvoluted using MaxQuant version 1.5 searching the T. brucei427_927_Tritryp-3.1 annotate protein database (37,220 proteins). Search engine analysis was performed assuming the full trypsin digestion (strict trypsin) in Mascot version 2.4.1. with pre-established parameters (Fragment Tolerance: 0.60 Da (Monoisotopic) Parent Tolerance: 10.0 PPM (Monoisotopic) Fixed Modifications: +57 on C (Carbamidomethyl) Variable Modifications: -17 on n (Gln->pyro-Glu), +16 on M (Oxidation), +32 on M (Dioxidation), +42 on n (Acetyl)). To visualize MS-spec data, we used Scaffold Proteome Software version 4.4.6 with a 0.5% peptide threshold and 5% protein threshold and 1 peptide minimum. False discovery rates (FDR) were calculated for both peptide and protein levels. A non-tagged cell line (procyclic form 449) was used to subtract contaminant proteins.

Immunoprecipitation in denaturing conditions to detect TbSUMO conjugates

For each immunoprecipitation (IP) experiment, 1.0×10^{10} bloodstream form (BF) cells were used. Cells were washed in TDB with 1X protease inhibitor cocktail (Roche) and 20 mM NEM. Pellets were resuspended at ~5.0×10^9 cells/ml in urea buffer (6M Urea, 50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM NEM, 0.5% NP-40, 2× protease inhibitor cocktail (Roche)) and sonicated until their viscosity was lost. The cell lysate was centrifuged at 20,000 g for 10 min at 4°C, and the supernatant with the nuclear enriched fraction was stored at −80°C. For immunoprecipitations, the nuclear extract was diluted 1:6 with dilution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1% NP-40, 0.5% Lauryl Sarcosine, 0.1 mM EDTA, 10 mM NEM, 1× protease inhibitor cocktail (Roche)) followed by overnight antibody incubation at 4°C on rotation. Antibody concentration for IP experiments was 180 µg/ml anti-SNF2PH rabbit antiserum, 800 µg/ml of anti-TbSUMO mAb 1C9H8, and 180 µg/ml of unspecific IgGs (prebleed antiserum). Previously equilibrated Protein G Sepharose beads (Sigma-Aldrich) were added to each diluted extract (~5.0×10^9 cells) containing the antibodies and incubated during 1 hr at 4°C on rotation to capture specific IgGs. Beads were washed five times for 5 min at 4°C on a rotating wheel with 1 ml of wash buffer (1 M urea, 50 mM HEPES pH 7.5, 500 mM NaCl, 1% NP-40, 0.5% Lauryl Sarcosine, 0.1 mM EDTA, 10 mM NEM, 1× protease inhibitor cocktail (Roche)). Then, beads were eluted with 2× Laemmli sample buffer and boiled at 99°C for 5 min. IP samples and inputs were subjected to SDS–PAGE and quantitative Western blotting using the appropriated antibodies.

3D-Immunofluorescence

Three-dimensional immunofluorescence (3D-IF) was carried out on cells in suspension as previously described [24]. Mouse anti-TbSUMO mAb 1:2,000, mouse anti-SNF2PH mAb 1:1,000, rabbit anti-SNF2PH affinity-purified antiserum 1:1,000, rabbit anti-VSG221 antiserum 1:50,000, mouse anti-Procyclin mAb 1:500 (MyBio-Source), and mouse anti-GFP mAb, 1:600 (Invitrogen) were used as primary antibodies. Alexa Fluor 488 and Alexa Fluor 594 goat anti-mouse or anti-rabbit (Invitrogen) were used as secondary antibodies. Cells were DAPI stained. Pseudocoloring, colocalization analysis, and maximum intensity projections were performed using ImageJ Fiji program version 1.51n software (National Institutes of Health), and one-way analysis of variance was used to compare the Pearson's coefficient value generated by the JACoP analysis plugin, available under ImageJ Fiji. For the colocalization mask, the plugin "Colocalization highlighter" was used where two points are considered as colocalized if their respective intensities are strictly higher than the threshold of their channels, which was set to 80% and if their ratio of intensity is higher than the ratio setting value of 80%.

In vitro Trypanosoma brucei SUMOylation assay in bacteria

In vitro reconstituted SUMOylation system was performed in *Escherichia coli* BL21 (DE3) cells transformed with pCDFDuet-1-TbSUMO/TbE2, followed by pACYCDuet-1-TbE1a-TbE1b. Competent bacteria were transformed again with pET28a(+)-3xFlag-SNF2PHN or pET28a(+)-3xFlag-SNF2PHC (See Appendix Tables S5 and S6 for cloning details). Assessment of SUMOylation reaction and TbSENP deconjugation assays were performed according to [13], and samples were analyzed by Western blot using an anti-Flag M2 mouse monoclonal antibody 1:5,000 (Sigma-Aldrich). Horseradish peroxidase-conjugated goat anti-mouse secondary antibody 1:5,000 (Sigma) was detected by Chemiluminescence using SuperSignal West Pico Chemiluminescent Substrate (Pierce).

Proximity ligation assay

The PLA assay was performed as previously described [9] by using the rabbit anti-SNF2PH affinity-purified antiserum (1:1,000) and mouse anti-TbSUMO 1C9H8 (1:2,000) as primary antibodies.

Luciferase assay

Luciferase assays were carried out using the Luciferase Assay System (Promega®) following the manufacturer's instructions from bloodstream form culture (3×10^6 cells) of control and SNF2PH depleted cells. Lectures were performed in a FB 12 Single Tube Luminometer (Titertek-Berthold) with pre-established parameters (2 s of delay time/10 s temp).

Fluorescent-activated cell sorting (FACS) analysis

SNF2PH RNAi bloodstream form induced cultures (1.5×10^7 cells) were collected and processed as previously described [9] using anti-VSG221 (1:3,000) as primary antibody. Alexa Fluor 488 goat anti-rabbit (Invitrogen) was used as secondary antibody.

Differentiation to procyclic form

Differentiation from slender to insect procyclic form was induced by 3 mM cis-Aconitate (Sigma-Aldrich), with a temperature shift from 37°C to 28°C and switching the medium to Differentiating Trypanosome Medium (DTM) as previously described [34]. The assessment of the differentiation process was monitored by a double IF to detect the expression of the surface glycoproteins using anti-procyclin and anti-VSG221 antibodies.

AMP Analog treatment and obtaining *in vivo* stumpy forms

Parasites in culture at a low density (2×10^5 cells/ml) were incubated with 8-pCT-2′-O-Me-5′-AMP (1 µM) (c078; Biolog Life Science Institute) during 18 h. To avoid the AMPK activation caused by cell density, the control and treated cells were analyzed at the same cell density. Slender and stumpy forms of pleomorphic AnTat 90.13 were purified from Balb/c mice at 3–5 days postinfection as previously described [42].

Ethics statement

Slender (Lister 427, antigenic type MiTat 1.2, clone 221a) and stumpy (pleomorphic AnTat 90.13) forms were isolated from Wistar and Balb/C mice rats, respectively, in compliance with policies approved by the Committee on Use and Care of Laboratory Animals of the Institute for Parasitology and Biomedicine López-Neyra, National Spanish Research Council (CSIC-IPBLN).

Statistical analysis

Statistical analysis was performed using two-tailed Student's *t*-test for paired observations using SigmaPlot Systat Software.

Data availability

Expanded View for this article is available online.

Acknowledgements

Author contributions

AS carried out and analyzed most experiments described in the manuscript. MN designed and directed research. PAI and VEA developed and performed the heterologous trypanosome SUMOylation system of SNF2PH fragments and wrote/interpreted these results. DR-B, DL-F, CB, and JMB performed distinct experiments included in this work. EA-L analyzed and helps to interpret the NGS data in RNA-seq and ChiP-seq and IV-C provided technical support. MN, VEA, and MCF supervised, edit, and founded the study. AS, MCF, and MN interpreted the data and wrote the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Cross GAM (1975) Identification, purification and properties of variant-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology* 71: 393–417

2. Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, Brooks K, Churcher C, Fahkro S, Goodhead I *et al* (2008) Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One* 3: e3527

3. Figueiredo LM, Cross GA, Janzen CJ (2009) Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Microbiol* 7: 504–513

4. Duraisingh MT, Horn D (2016) Epigenetic regulation of virulence gene expression in parasitic protozoa. *Cell Host Microbe* 19: 629–640

5. Cestari I, Stuart K (2018) Transcriptional regulation of telomeric expression sites and antigenic variation in trypanosomes. *Curr Genomics* 19: 119–132

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

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

8. McCulloch R, Navarro M (2016) The protozoan nucleus. *Mol Biochem Parasitol* 209: 76–87

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

10. Crozier TWM, Tinti M, Wheeler RJ, Ly T, Ferguson MAJ, Lamond AI (2018) Proteomic analysis of the cell cycle of procyclic form *Trypanosoma brucei*. *Mol Cell Proteomics* 17: 1184–1195

11. Zhao X (2018) SUMO-mediated regulation of nuclear functions and signaling processes. *Mol Cell* 71: 409–418

12. Liao S, Wang T, Fan K, Tu X (2010) The small ubiquitin-like modifier (SUMO) is essential in cell cycle regulation in Trypanosoma brucei. *Exp Cell Res* 316: 704–715

13. Iribarren PA, Berazategui MA, Cazzulo JJ, Alvarez VE (2015b) Biosynthesis of SUMOylated Proteins in Bacteria Using the Trypanosoma brucei Enzymatic System. *PLoS One* 10: e0134950

14. Rosonina E, Akhter A, Dou Y, Babu J, Sri Theivakadadcham VS (2017) Regulation of transcription factors by sumoylation. *Transcription* 8: 220–231

15. Giles KA, Gould CM, Du Q, Skvortsova K, Song JZ, Maddugoda MP, Achinger-Kawecka J, Stirzaker C, Clark SJ, Taberlay PC (2019) Integrated epigenomic analysis stratifies chromatin remodellers into distinct functional groups. *Epigenet Chromatin* 12: 12

16. Clapier CR, Iwasa J, Cairns BR, Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* 18: 407

17. Yan L, Wu H, Li X, Gao N, Chen Z (2019) Structures of the ISWI-nucleosome complex reveal a conserved mechanism of chromatin remodeling. *Nat Struct Mol Biol* 26: 258–266

18. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR *et al* (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45: D200–D203

19. Mouriz A, López-González L, Jarillo JA, Piñeiro M (2015) PHDs govern plant development. *Plant Signal Behav* 10: e993253

20. Musselman CA, Kutateladze TG (2011) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res* 39: 9061–9071

21. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P *et al* (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442: 86

22. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR *et al* (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442: 96–99

23. Sanchez R, Zhou M-M (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* 36: 364–372

24. Landeira D, Bart JM, Van Tyne D, Navarro M (2009) Cohesin regulates VSG monoallelic expression in trypanosomes. *J Cell Biol* 186: 243–254

25. Iribarren PA, Berazategui MA, Bayona JC, Almeida IC, Cazzulo JJ, Alvarez VE (2015a) Different proteomic strategies to identify genuine Small Ubiquitin-like MOdifier targets and their modification sites in *Trypanosoma brucei* procyclic forms. *Cell Microbiol* 17: 1413–1422

26. Lai AY, Wade PA (2011) Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* 11: 588–596

27. Cairns BR (2009) The logic of chromatin architecture and remodelling at promoters. *Nature* 461: 193–198

28. Muller LSM, Cosentino RO, Forstner KU, Guizetti J, Wedel C, Kaplan N, Janzen CJ, Arampatzi P, Vogel J, Steinbiss S *et al* (2018) Genome

organization and DNA accessibility control antigenic variation in trypa-
nosomes. *Nature* 563: 121−125

29. Van Bortle K, Nichols MH, Li L, Ong C-T, Takenaka N, Qin ZS, Corces VG
(2014) Insulator function and topological domain border strength scale
with architectural protein occupancy. *Genome Biol* 15: R82

30. Jensen BC, Booster N, Vidadala RSR, Maly DJ, Parsons M (2016) A novel
protein kinase is essential in bloodstream *Trypanosoma brucei*. *Int J
Parasitol* 46: 479−483

31. Nguyen TN, Nguyen BN, Lee JH, Panigrahi AK, Gunzl A (2012)
Characterization of a novel class I transcription factor A (CITFA) subunit
that is indispensable for transcription by the multifunctional RNA poly-
merase I of *Trypanosoma brucei*. *Eukaryot Cell* 11: 1573−1581

32. Nguyen TN, Muller LS, Park SH, Siegel TN, Gunzl A (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 Res 42: 3164−3176

33. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C (2009)
Transcriptome analysis of differentiating trypanosomes reveals the exis-
tence of multiple post-transcriptional regulons. *BMC Genom* 10: 495

34. Barquilla A, Saldivia M, Diaz R, Bart JM, Vidal I, Calvo E, Hall MN,
Navarro M (2012) Third target of rapamycin complex negatively regu-
lates development of quiescence in *Trypanosoma brucei*. *Proc Natl Acad
Sci USA* 109: 14399−14404

35. Leung KF, Riley F, Carrington M, Field MC (2011) Ubiquitylation as a
general mechanism for internalisation of trans-membrane domain
surface proteins in trypanosomes. *Eukaryot Cell* 10: 916−931

36. Schulz D, Mugnier MR, Paulsen EM, Kim HS, Chung CW, Tough DF, Rioja
I, Prinjha RK, Papavasiliou FN, Debler EW (2015) Bromodomain proteins
contribute to maintenance of bloodstream form stage identity in the
African Trypanosome. *PLoS Biol* 13: e1002316

37. Figueiredo LM, Janzen CJ, Cross GAM (2008) A histone methyltransferase
modulates antigenic variation in African Trypanosomes. *PLoS Biol* 6: e161

38. Dubois KN, Alsford S, Holden JM, Buisson J, Swiderski M, Bart JM,
Ratushny AV, Wan Y, Bastin P, Barry JD *et al* (2012) NUP-1 is a large
coiled-coil nucleoskeletal protein in trypanosomes with lamin-like func-
tions. *PLoS Biol* 10: e1001287

39. Watson AA, Mahajan P, Mertens HD, Deery MJ, Zhang W, Pham P, Du X,
Bartke T, Zhang W, Edlich C *et al* (2012) The PHD and chromo domains
regulate the ATPase activity of the human chromatin remodeler CHD4.
J Mol Biol 422: 3−17

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

41. Valenciano AL, Ramsey AC, Mackey ZB (2015) Deviating the level of
proliferating cell nuclear antigen in Trypanosoma brucei elicits distinct
mechanisms for inhibiting proliferation and cell cycle progression. *Cell
Cycle* 14: 674−688

42. Saldivia M, Ceballos-Perez G, Bart JM, Navarro M (2016) The AMPKal-
pha1 pathway positively regulates the developmental transition from
proliferation to quiescence in *Trypanosoma brucei*. *Cell Rep* 17: 660−670

43. Silvester E, McWilliam KR, Matthews KR (2017) The cytological events
and molecular control of life cycle development of trypanosoma brucei
in the mammalian bloodstream. *Pathogens* 6: E29

44. Zimmermann H, Subota I, Batram C, Kramer S, Janzen CJ, Jones NG,
Engstler M (2017) A quorum sensing-independent path to stumpy
development in *Trypanosoma brucei*. *PLoS Pathog* 13: e1006324

45. Donze D, Adams CR, Rine J, Kamakaka RT (1999) The boundaries of the
silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev* 13:
698−708

46. Dhillon N, Raab J, Guzzo J, Szyjka SJ, Gangadharan S, Aparicio OM,
Andrews B, Kamakaka RT (2009) DNA polymerase ε, acetylases and
remodellers cooperate to form a specialized chromatin structure at a
tRNA insulator. *EMBO J* 28: 2583−2600

47. Ebersole T, Kim JH, Samoshkin A, Kouprina N, Pavlicek A, White RJ, Lari-
onov V (2011) tRNA genes protect a reporter gene from epigenetic
silencing in mouse cells. *Cell Cycle* 10: 2779−2791

48. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, Haussler D,
Kamakaka RT (2012) Human tRNA genes function as chromatin insula-
tors. *EMBO J* 31: 330−350

49. Gurudatta BV, Yang J, Van Bortle K, Donlin-Asp PG, Corces VG (2013)
Dynamic changes in the genomic localization of DNA replication-related
element binding factor during the cell cycle. *Cell Cycle* 12: 1605−1615

50. Penate X, Lopez-Farfan D, Landeira D, Wentland A, Vidal I, Navarro M
(2009) RNA pol II subunit RPB7 is required for RNA pol I-mediated tran-
scription in *Trypanosoma brucei*. *EMBO Rep* 10: 252−257

51. Andres-Leon E, Nunez-Torres R, Rojas AM (2016) miARma-Seq: a compre-
hensive tool for miRNA, mRNA and circRNA analysis. *Sci Rep* 6: 25749

52. Obado SO, Field MC, Chait BT, Rout MP (2016) High-Efficiency Isolation
of Nuclear Envelope Protein Complexes from Trypanosomes. *Methods
Mol Biol* 1411: 67−80

Identification of SUMO conjugates in bloodstream trypanosomes and potential function in VSG expression

Andreu Saura[1], Paula Iribarren[2], Claudia Gómez-Liñán[1], Domingo Rojas-Barros[1], Sara Torres-Rusillo[1], Vanina Álvarez[2] and Miguel Navarro[1*]

[1]Instituto de Parasitología y Biomedicina "López-Neyra",CSIC (IPBLN-CSIC), 18016 Granada, Spain.

[2]IIB-INTECH, Buenos Aires, Argentina, Avenida 25 de Mayo y Francia, 1650 San Martin, Buenos Aires, Argentina.

[*]Correspondence should be addressed to. MN. Tel:34-958181651; Fax: 34-958181633.

E-mail: miguel.navarro@ipb.csic.es

Running title: **Tracing trypanosome SUMO-conjugates in search for VSG regulators**

Summary

Infective bloodstream form (BF) trypanosomes avoid the host immune response by switching the expression of Variant Surface Glycoproteins (VSGs). Transcription of the active telomeric VSG loci termed Expression Site (ES) occurs in a nuclear body, named ESB, that is associated with a highly SUMOylated focus (HSF). SUMOylation is a large PTM known to regulate many biological processes, as transcription and subcellular localization. Chromatin SUMOylation is a distinct feature of the active VSG-ES locus and is essential for efficient recruitment of RNA polymerase to the promoter, thus functions as a positive epigenetic mark in this protozoa parasite. To approach the complexity of the nuclear HSF, we first applied affinity purification of SUMO-conjugated proteins and later site-specific MS/MS proteomic strategies aimed to identify unequivocally SUMO substrates and the site of modification. This method allowed us to discover 45 modified lysine residues in 37 proteins, some exclusively present in the BF stage, suggesting SUMOylation possibly controls developmental regulation of target proteins. Additionally, proximity ligation assays, identified SUMO-conjugated a distinct nuclear localization site suggesting assembly of the HSF might involve protein-protein interactions mediated by SUMO. Ours results suggest a model whereby protein SUMOylation status underlies targets nuclear dynamics towards the HSF location. Finally, and as a proof of concept, RNAi depletion of selected SUMO targets such as H3V and TFIIS2-2, demonstrated regulated VSG expression. Altogether, our results suggest that SUMO is critical to nuclear allocate factors to the unique site of VSG transcription. In addition, SUMO it is a useful molecular bait to identify novel factors involved in VSG antigenic variation of the parasite, which we made available to the field for future studies.

Keywords

Trypanosoma brucei, Variant Surface Glycoprotein (VSG), Antigenic Variation, Post-Translational Protein Modification (PTM). Small Ubiquitin-Related Modifier (SUMO), High SUMOylation foci (HSF), SUMO-Conjugation, SUMOylation site-specific identification, Lysine cleavage site, Epigenetic.

Abbreviated Summary

In trypanosomes infective form surface, the VSG protein is exchanged to avoid the host immune response. VSG expression of the active telomeric ES occurs in a single nuclear body known as the ESB, which contains proteins modified by the SUMO peptide. The covalent addition of SUMO alters the interactive surface of target proteins leading to distinct protein-protein interactions. Hence, we identify SUMO conjugated nuclear factors and show that selected targets localize to a single site in the nucleus, further depletion of particular targets deregulated VSG expression. We present a list of SUMO-conjugated (and the K-modified) nuclear factors, which likely function as regulators of gene expression for future studies.

Introduction

Post-translational modifications (PTMs) are important routes to increase the versatility of proteins encoded in the genome. Beyond small-molecule modifications (like phosphorylation, alkylation, glycosylation, etc.) proteins can be modified with whole protein moieties. Among them, ubiquitination is the best characterized however, many other structurally related proteins have been described. The Small Ubiquitin-like modifier (SUMO) has been shown to modulate important chromatin and nuclear associated events such as chromatin remodeling, DNA repair, transcriptional regulation and telomere maintenance. The covalent addition of SUMO typically alters the interactive surface of the target protein leading to changes in its biological activity, stability, and subcellular distribution, among other outputs. This large PTM functions mainly by changing the targets protein-protein interaction and stability. SUMOylation is a reversible and highly dynamic modification and as a consequence, only a minor fraction of a target protein is normally found to be modified at a given time. Protein SUMOylation takes place through the concerted action of a cascade of enzymes. Firstly, the maturation step requires the action of a specific protease (Ulp1) to cleave a few aminoacids at the C-terminal end of SUMO which then exposes a conserved diglycine motif (diGly). This subsequently allows the ATP-dependent activation by the E1 activation enzyme which is followed by its transfer to the cysteine residue of the E2 conjugating enzyme (Ubc9) to finally be conjugated to a Lys residue on the target protein. SUMO conjugation to a substrate-specific lysine usually depends on E3-ligases, forming an isopeptide bond between the SUMO C-terminal glycine residue and the ε-

82 amino group of the lysine residue in the target protein. Remarkably, the modified lysine

83 residue is usually embedded in a SUMO consensus motif (ψ-K-X-E, were ψ represents

84 L, I or V) (Flotho and Melchior, 2013; Han, 2018; Wang and Dasso, 2009). It is well

85 known that SUMO pathway is highly conserved among eukaryotic organisms and

86 constitutes an essential system in many of them.

87 In *Trypanosoma brucei*, the etiological agent of sleeping sickness in humans and

88 nagana in livestock, SUMO has proved to be essential, since it is necessary for normal

89 cell cycle progression (Liao et al., 2010), required for centrosome segregation (Obado

90 et al., 2011) and involved in the oxidative stress response (Klein et al., 2013).

91 Furthermore, other Kinetoplastid parasites such as *T. cruzi* and *Leishmania* possess a

92 functional SUMOylation system (Bayona et al., 2011). The identification of SUMOylated

93 proteins by site-specific proteomic analysis in the insect stage of *T. brucei* known as

94 procyclic form (PF), has led to the confident identification of 45 SUMOylated substrates,

95 most of them related to nuclear remodeling events (Iribarren et al., 2015a). Recently,

96 SUMO chain formation was proposed to function in chromatin organization and

97 telomere positioning in the nucleus of the insect form of the parasite (Iribarren et al.,

98 2018). Most importantly, protein SUMOylation in trypanosome pathogenesis is

99 becoming increasingly evident, as in the infective stage of the parasite, known as

100 bloodstream form (BF), this PTM regulates the expression of the Variant Surface

101 Glycoprotein (VSG) (Lopez-Farfan et al., 2014). African trypanosomes elude the host

102 immune response by switching the expression between VSGs, only one of which is

103 expressed at any given time. The recruitment of a single VSG-Expression Site (VSG-

104 ES) telomeric locus to a discrete nuclear body (ESB) underlies the mechanism

4

105   responsible for VSG monoallelic expression (Navarro and Gull, 2001). A highly

106   SUMOylated Focus (HSF) was described as a novel nuclear structure that partially

107   colocalizes with the RNA polymerase I at the ESB (Lopez-Farfan et al., 2014).

108   Additionally, SUMOylation of chromatin-associated proteins is required for the active

109   transcriptional state of the VSG-ES and essential for efficient recruitment of RNA pol I to

110   the VSG-ES promoter. This mechanism employs a novel SIZ/PIAS SUMO E3 ligase

111   responsible for the VSG-ES chromatin SUMOylation, leading to a new positive

112   epigenetic mark (Lopez-Farfan et al., 2014).

113   Therefore, identifying SUMO-modified proteins in the infective bloodstream form of

114   the parasite constitutes a logical initial effort to discover novel factors involved in VSG

115   regulation. In this study, we performed a site-specific proteomic analysis using BF

116   protein extracts providing a list of novel SUMOylated substrates using the constitutive

117   expression of His-HA-*Tb*SUMO[T106K] (Iribarren et al., 2015a; Tammsalu et al., 2015) now

118   for the first time in the infective form of *T. brucei*. SUMO targets were validated for a

119   number of the proteins using a SUMOylation *ex vivo* assay established in bacteria by

120   expressing *T. brucei* pathway machinery. We also addressed a possible single-site

121   location in the nucleus of the modified substrate and for selected factors, we

122   investigated a possible function in VSG expression to confirm at least some of the

123   identified proteins are indeed involved in VSG regulation. In summary, we have

124   identified novel SUMO substrates and the Lys conjugated sites, which provide new

125   insights into the functions regulated by SUMO throughout different developmental

126   stages.

127

Results

Proteomic identification of SUMOylated proteins and SUMOylation sites in the
Bloodstream Form

The highly SUMOylated focus (HSF) detected in the nucleus of bloodstream form (BF) trypanosomes (Lopez-Farfan et al., 2014) colocalizes with the nuclear body ESB, the site for VSG monoallelic expression (Navarro and Gull, 2001). The high intensity SUMO signal detected by an anti-TbSUMO monoclonal antibody at the nuclear site where VSG transcription occurs suggests a considerable number of SUMOylated proteins might be involved in VSG expression. Thus, we used SUMO as bait to identify novel factors potentially associated with VSG expression utilizing proteomic MS/MS analyses of affinity-purified extracts. Proteomic identification of SUMO substrates has become a complex issue due to methodological and experimental constraints. To accomplish this in the BF we first generated a cell line that constitutively expresses an 8xHis-HA-tagged SUMO version by double replacement of endogenous alleles, as previously described in the insect procyclic form (PF) (Iribarren et al., 2015a). This cell line displayed no difference in proliferation under regular culture conditions compared to wild type (WT) parasites, suggesting that SUMO function is not affected (data not shown). To isolate SUMO conjugates from transfected cells we directly resuspended the parasite pellet in 6 M urea to avoid deconjugating peptidase activity as well as co-purification of non-covalent SUMO interacting proteins. Cell extracts were later subjected to Ni^{2+}-NTA chromatography coupled to HA-affinity enrichment and, after confirming the presence of SUMO conjugates in the eluates by Western Blot (WB), samples were analyzed by MS/MS proteomics. Using this approach, from three independent experiments, we were

151 able to identify 194 potentially SUMOylated proteins that were not present in control

152 purifications, with a few reproducible candidates (8 proteins from experimental

153 replicates 2 and 3) (Table S1 (A,B,C)). The lack of reproducibility stressed the need to

154 utilize a different strategy that leads to a specific enrichment of SUMOylated proteins

155 allowing a confident large-scale identification of SUMO substrates.

156 Next, we applied a novel approach that specifically enriches SUMOylated peptides

157 and identifies the modified Lys in each SUMO target (Iribarren et al., 2015a; Tammsalu

158 et al., 2015). To do so, we generated a cell line expressing, in a tetracycline inducible

159 manner, the His-HA-TbSUMO with the T106K mutation, previous to the di-Gly motif

160 (His-HA-TbSUMO^{T106K}). The introduction of this mutation generates a T-shape peptide

161 after Lys-C digestion allowing the enrichment of SUMOylated peptides using a specific

162 anti-GG antibody and the unambiguous identification of the modified Lys residues by

163 MS. Cells expressing His-HA-TbSUMO^{T106K} were able to conjugate SUMO substrates

164 as were the non-tagged cell line (Figure 1A) and showed the typical SUMOylation

165 pattern, with a nuclear HSF, by indirect immunofluorescence analysis (Lopez-Farfan et

166 al., 2014) (Figure 1B). Transgenic cells showed no differences in cell growth compared

167 with the parental cell line (Figure 1C). Considering the limitations of BF culture

168 regarding cell density, His-HA-TbSUMO^{T106K} expressing parasites were amplified in

169 Wistar rats and SUMOylated proteins were purified by nickel-affinity chromatography

170 under denaturing conditions (Figures 1D and E). Final eluates were subjected to Lys-C

171 digestion and further anti-GG enrichment of SUMOylated peptides, prior to MS/MS

172 identification.

173   As displayed in Table 1, following this approach we were able to identify 45 lysine-

174   modified residues in 37 proteins. Most of the identified proteins showed one

175   SUMOylation site, whereas 6 of them (DNA Topo 1B, SMC-4, Polyubiquitin, rRNA

176   processing protein, DnaJ domain containing protein and hypothetical protein

177   Tb927.9.13320) showed two sites, and one target (hypothetical protein

178   Tb927.10.12030), three sites. Globally, 23 out of 37 proteins have experimentally

179   assigned or predicted function, whereas the remaining 14 correspond to hypothetical

180   with unknown function products. When comparing our data with the previously identified

181   proteins in PF, we found 15 SUMO-conjugated substrates shared between these two

182   developmental stages (Table S2) suggesting the existence of conserved processes

183   regulated by SUMO in both BF and PF developmental stages.

184   From the SUMO modified lysine residues identified, 21 of them were found in the

185   sequence context that matches either consensus or reversed consensus motif. Analysis

186   of SUMOylation sites with *pLogo* server (https://plogo.uconn.edu/) (O'Shea et al., 2013)

187   led to 19 out of the 45 mapped lysine acceptor residues in which statistical significant

188   positions contained V at position -1 (P-value < 0.05) and E at position +2 (P-value <

189   0.001) for a fixed K at position 0 (Figure 2). Interestingly, there was a considerably

190   enrichment of I at -1 compared with V preference at that position in PF data (log-odd:

191   3.026).

192

193   *SUMO target validation*

194   A challenge intrinsic for SUMO target proteins identification is the sub-stoichiometric

195   level and labile nature of this post-translational modification. Thus, we decided to

validate some of the targets using a SUMOylation trypanosome-specific system by expressing essential pathway subunits in bacteria (Iribarren et al., 2015b). This system involves the co-expression in *E. coli* of the target of interest together with all trypanosome SUMOylation enzymes (TbE1a/TbE1b and TbE2), including TbSUMO. The selected candidates, RNA polymerase-associated protein CTR9 (Tb927.3.3220) and the Transcription elongation factor S-II (TFIIS2-2) (Tb927.2.3480), were expressed in bacteria containing 3 copies of epitope tag Flag (3xFlag) as recombinant proteins together with the complete trypanosome SUMOylation system, or along without the system subunits as negative controls. After 5 h of induction with 1mM IPTG, whole bacterial extracts were analysed by WB using anti-Flag antibodies. As shown in Figure 3A and B, CTR9 and TFIIS2-2 were successfully expressed in *E. coli* and only when the target was co-expressed with the whole trypanosome SUMOylation machinery, additional slower migrating bands were detected, suggesting larger bands correspond to mono- and multi-/poly-SUMOylated forms of each protein appeared.

In situ analysis of SUMO mediated interactions

To investigate if SUMO could act either as a modifier or as a scaffold protein inside the HSF, we performed a Proximity Ligation Assay (O-link Bioscience) between SUMO and five substrates identified in the site-specific proteomic experiment: H3V, DNA topoisomerase IB, CTR9, Tb927.10.14190 and TFIIS2-2.

PLA methodology utilizes target specific primary antibodies (raised in different species) and the corresponding oligonucleotide-labeled secondary antibodies (PLA probes). When the PLA probes are in close proximity, they support ligation and

219 amplification of a product, which can be detected by a complementary fluorescently

220 labeled oligonucleotide.

221 To this end, we generated transgenic cell lines to N-terminally tag the proteins with a

222 3xHA epitope and their expression was analyzed after 24 h of doxycycline induction by

223 IFI. As shown in Figure 4, all targets exhibited a mainly nuclear localization in

224 agreement with their described function.

225 When we evaluated PLA reactions for each selected target using anti-SUMO and

226 anti-HA, positive signals were detected in the nucleus of all of them (Figure 5).

227 Interaction signals were observed in a percentage of the population (from 20-55% of the

228 cells depending on the cell line, Figure 5) in agreement with the high dynamic range of

229 this PTM. All together, these results imply the existence of multiple SUMOylated

230 proteins and/or SUMO-interacting proteins in a particular focus within the nucleus.

231

232 *Functional analysis of selected SUMO substrates*

233 To investigate a potential functional relevance of SUMO conjugated proteins identified

234 above in VSG expression, we depleted, by inducible RNA interference (RNAi), two of

235 the proteins since tackling the whole list was out of the scope of this work. We selected

236 factors related to transcription such as, TFSII2-2 and the histone H3V and analyzed

237 phenotypes after 24h of RNAi (Figure 6A). Relative transcript levels by RT-qPCR

238 analysis of cells depleted of TFSII2-2 and H3V indicated a reduced expression of active

239 VSG221 ($P<0.001$ and $P<0.05$, respectively). Interestingly, the mRNA of a VSG located

240 in an inactive telomeric expression site, the VSG121, was significantly increased

241 ($P<0.01$ and $P<0.001$, respectively). The VSG121 messenger is normally detected from

242 few cells (aprox. < 0.5%) in any population of VSG221 expressors that have undergone

243 *in situ* transcriptional switch to the previously inactive VSG121 expression site. These

244 results suggest the VSG switching frequency was increased after TFSII2-2 and H3V

245 depletion. A control gene named C1, transcribed by RNA pol II, remained unmodified in

246 any of the depletion experiments suggesting cell general fitness had not been severely

247 affected. However, 18S ribosomal rDNA expression was considerably reduced after

248 H3V depletion (P<0.05) suggesting this histone variant is required for RNA pol I

249 transcription.

250 To determine if TFIIS2-2 depletion compromises VSG regulation, we carried out a RNA-

251 seq assay. First, cells depleted in TFIIS2-2 did not show any significant growth defect in

252 doubling time (Figure 6B) and TFIIS2-2 depletion was evaluated by western blot

253 analysis using the specific TFIIS2-2 mAb (Figure 6C).   RNA-seq analysis of TFIIS2-2

254 depleted cells led to the identification of 100 Differentially Expressed Genes (DEGs)

255 showing FDR<0.05 from 9313 background genes (Figure 6D &Table S3). Out of the

256 DEGs, we found an overrepresentation upregulation of Bloodstream form Expression

257 Sites (BESs) associated to different telomeric Expression Site Associated Genes

258 (ESAGs), including those related to the transferrin-binding protein and folate

259 transporter. Additionally, we found upregulation of PAGs and procyclin genes and other

260 DEGs that showed downregulation, involving some hypothetical proteins and the H3V,

261 proposing that TFIIS2-2 influences H3V regulation. All this data together suggests that

262 TFIIS2-2 regulates VSG expression by increasing transcriptional levels of previously

263 repressed inactive BES. Additionally, we analyzed switching events associated to H3V

264 depletion by double direct IF. After several passes, doxycycline induced parasites

265 represented the 0.9% of VSG221/VSG121 double expressions, whereas no proportion

266 of cells were found to be related to VSG121 neither switched off cells (Figure 6E). This

267 result suggest that a lack of H3V increases recombination leading to a transcriptional

268 profile of inactive BES expression, excluding switching events. The global deregulation

269 observed for both active and inactive VSG-ES after depletion of these two SUMO-

270 conjugated proteins and the PLA analysis showed in Figure 5, suggest that

271 posttranslational modification by SUMO may influence thefunction of these factors in

272 VSG regulation, providing specificity.

273

Discussion

275 By consecutively changing the variant surface glycoprotein (VSG) coat, the parasite

276 *Trypanosoma brucei* undergoes antigenic variation and eludes the host immune

277 response. Previous work has shown that VSG transcriptional regulation occurs

278 epigenetically, representing a genuine example of monoallelic expression of a multigene

279 family (reviewed in (Cestari and Stuart, 2018; Duraisingh and Horn, 2016)). We have

280 proposed a model whereby the recruitment of a single VSG Expression Site (VSG-ES)

281 telomeric locus by the unique nuclear Body (ESB) underlies the molecular mechanism

282 that restricts the monoallelic expression of the VSG (Navarro & Gull, 2001) (reviewed in

283 (Navarro et al., 2007) (McCulloch and Navarro, 2016)). Recent results convincingly

284 support the ES body/VSG-ES unique interaction model for monoallelic expression,

285 providing insights into the epigenetic inheritance mediated by nuclear architecture

286 (Landeira, et al. 2009). Cohesin complex is important in preserving VSG-ES epigenetic

287 active inherency, as the knockdown of Cohesin led to the premature separation of the

288   VSG-ES sister chromatids off the nuclear body ESB, triggering antigenic switching

289   (Landeira, et al. 2009). More recently, SUMOylation of chromatin was described as a

290   novel chromatin epigenetic mark that positively regulates VSG expression and stains a

291   focus in the nucleus (HSF) associated with the ESB (Lopez-Farfan, et al., 2014). Thus,

292   in this work we decided to uncover novel proteins associated with the nuclear body ESB

293   and likely related to antigenic variation by performing MS/MS proteomic analysis of

294   SUMOylated proteins isolated from infective bloodstream form (BF) extracts. Using this

295   alternative and novel approach allowed us to identify otherwise inscrutable factors

296   associated with the ESB and make them available to researchers in the field for

297   possible detailed studies in the future.

298   The identification of SUMO conjugates is a challenging process considering the

299   usually low proportion of the modified forms of a given protein and the susceptibility of

300   deconjugation by specific isopeptidases. Initial attempts to purify SUMO-conjugated

301   proteins in trypanosomes failed completely (Klein et al., 2013). In this work, we first

302   attempted to perform a large-scale identification of SUMO targets, using protein extracts

303   from the bloodstream form of the parasite, by affinity purification of His/HA-SUMO

304   conjugated proteins but did not achieve the high standard required. Thus, independent

305   purifications and MS/MS analyses identified proteins listed in Table S1A, B, C, which

306   showed low reproducibility and contained contaminant proteins, probably due to the lack

307   of a specific enrichment step. Nevertheless, this approach allowed us to identify several

308   targets (Table S1) that should be considered as genuine substrates, since they were

309   previously described as SUMO conjugated in other organisms such as yeast and

310   mammals. Therefore, many of these proteins ought to be considered the function of

311 SUMO modification in each target for further investigation. In fact, several trypanosome

312 orthologs for example, Tb927.10.740 (SMC4) and Tb927.9.5190 (PCNA) are currently

313 under study in our groups, and to date, co-IP experiments confirmed that both are

314 SUMOylated proteins in trypanosomes (Rojas-Barros *et al.*, & Iribarren*et al.*,

315 unpublished). Furthermore, Tb927.3.2140 (SNF2PH, formally known as Transcription

316 Activator) has been recently functionally characterized as a chromatin-remodeling

317 factor, whose locus reposition is regulated in a SUMO-dependent manner (Saura *et al.*,

318 EMBO Rep. *In Press*). This approach also identified well-known protein targets of

319 SUMO in other trypanosomes and several eukaryotic organisms including: DEAD/H

320 helicases, KU70, GTP-binding, eIF4E, SMC family members, histones, DNA repair

321 proteins, etc. (Hendriks and Vertegaal, 2016; Iribarren et al., 2015a).

322 In addition to SUMO-conjugated identification of proteins, we unequivocally identify

323 the targets and most importantly the SUMOylation site/s in order to fully advance in

324 future functional analyses. Thus, we performed site-specific proteomics by generating

325 His-HA-*Tb*SUMO^{T106K} expressing parasites followed by specific enrichment of

326 SUMOylated peptides with anti-GG antibodies, as previously described (Iribarren et al.,

327 2015a; Tammsalu et al., 2015). Following this approach we were able to unambiguously

328 identify a subset of 37 confident SUMOylated proteins and 45 modified Lys residues.

329 This data provides the first list of SUMOylated proteins in the infective stage of the

330 parasite. We used a similar approach to identify SUMO-conjugated proteins, one of

331 which was previously described in the insect stage, named Procycliclic Form (PF)

332 (Iribarren et al., 2015a). Thus, it is feasible to compare the protein profiles in the two

333 developmental stages in order to identify proteins that utilize SUMO at least as one as

334 the PTMs, which may be involved in the complex parasite development (Rico et al.,

335 2017).

336 Eukaryotic organisms present a consensus SUMOylation site (ψ-K-X-E; ψ: aliphatic

337 aminoacid; X: any aminoacid) that seems to be conserved in a considerable number of

338 targets in the protozoa parasite *T. brucei*. Interestingly, in the PF developmental stage,

339 the aliphatic residue in position -1 is preferably V, whereas in the infective BF, both V

340 and I appeared in similar frequency. This data shows a higher flexibility in the

341 consensus SUMOylation site at this position in the bloodstream form of the parasite.

342 Most of the proteins modified by SUMO, identified in the infective form, are involved

343 in nuclear processes. These proteins include: histones, PCNA, DNA topoisomerase 1B,

344 RBP12 and SMC4, which strengthens SUMO relevance as a PTM that regulates

345 fundamental processes. Many of the substrates identified here in the BF (Table 1), were

346 also found to be SUMOylated in the PF stage (Iribarren et al., 2015a), suggesting a

347 conserved function by this PTM, regarding processes such as chromatin remodeling

348 and DNA/RNA metabolism. It is noteworthy that most of these conserved SUMO targets

349 are SUMOylated on particular lysines, highlighting the importance of specific residues

350 on their regulation. Differential SUMOylation sites in *Drosophila melanogaster*

351 Ultraspirace (Usp), a protein involved in development, suggests Usp modulates its

352 activity by changing molecular interactions (Bielska et al., 2012).

353 Interestingly, histone H2A, H3V, DNA topoisomerase IB, PCNA, SMC4, and the

354 hypothetical protein Tb927.9.13320 constitute SUMO substrates that presented distinct

355 developmentally regulated modifications in the number and/or position of the acceptor

356 lysines between BF and PF (Table S2). Additional lysine acceptor sites observed in the

357 infective form of the parasite suggest a stage-specific SUMO-mediated modification

358 which most likely acquires new functions linked to pathogenicity. This is in agreement

359 with SUMO being essential for maintaining cell homeostasis when the parasite

360 encounters environmental stress such as: osmotic stress, hypoxia, heat shock, and

361 nutrient stress (Enserink, 2015), most of which occurs when parasites infect a

362 mammalian host through the bite of the insect vector. Thus, trypanosomes fine tune

363 nuclear targets, described above, probably remodel gene expression and activate

364 and/or repress chromatin, similarly to other eukaryotes (Niskanen and Palvimo, 2017).

365 In mice development, SUMO-specific protease 2 (SENP2) regulates epigenetic gene

366 expression by modification of the PTM status of Polycomb Repressive Complex 1

367 (PRC1), changing promoter affinity. Depletion of SENP2 down-regulates expression of

368 Polycomb regulated genes. It seems that SUMOylation increases binding affinity of

369 Pc2/CBX4 (PRC1 subunit) to H3K27me3. SENP2 specifically controls SUMOylation

370 status of Pc2/CBX4, which facilitates PRC1 binding to H3K27me3 leading to

371 transcriptional repression (Kang et al., 2010).

372 Most importantly, latest reviews remark on the importance of SUMO as a key

373 molecular PTM in the regulation of transcription factors and chromatin-associated

374 proteins involved in differentiation and developmental processes (Deyrieux and Wilson,

375 2017; Monribot-Villanueva et al., 2016).

376 Suppression of chromosomal rearrangements is controlled by SUMOylation in yeast,

377 in particular by Mms21, a SUMO ligase that maintains genome integrity and one of its

378 mayor substrates are the SMC-family proteins (Albuquerque et al., 2013). Unfortunately,

379 it is not currently known how SUMO functions in SMC proteins. We found SMC4 a

16

380   subunit of the condensin complex (Structural Maintenance of Chromosome family), as a

381   reproducible SUMO target that interestingly, contains a stage-specific SUMOylation site

382   in the infective developmental form (K623) (Table S2). This result raises the possibility

383   that distinct SUMOylation sites of a subunit of the same protein complex may lead to

384   different biological outcomes in each developmental stage.

385   Additionally, the crosstalk between SUMO and other PTMs is a very interesting

386   feature to be considered as molecular mechanism to regulate protein targets. Histone

387   2A, for example, appeared to be highly acetylated in both developmental stages of *T.*

388   *brucei* (Moretti et al., 2018) on some of the lysine residues identified as SUMOylated by

389   site-specific proteomics. Overexpression and mislocalization of histone H3 variant

390   (CENP-A in humans and Cse4 in yeast) leads to aneuploidy in yeast and is detected in

391   human cancers. In yeast, the centromeric histone variant Cse4, is regulated by

392   SUMOylation mainly by Siz1 and Siz2 E3 ligases at the lysine 65 (K65) and the

393   expression of cse4 K65R mutant leads to increased stability and mislocalization of cse4

394   K65R. All of which suggest that SUMO regulates protein stability and genomic

395   localization (Obkun et al., 2018).

396   It has been reported that *T. brucei* bromodomain factor 2 (TbBDF2) is able to

397   recognize hyperacetylated N-T of H2AZ, essential for cell growth (Yang et al., 2017),

398   but the role of SUMO on histone regulation has not been elucidated yet. Acetylation-

399   SUMOylation crosstalk is also observed in reader protein TRIM24 and histone H3

400   epigenetic modulation (Appikonda et al., 2018). Considering these facts, we cannot rule

401   out a fine-tuned epigenetic regulation network of histones that might be differentially

402   expressed in *T. brucei* life stages.

403   Interestingly, polyubiquitin was also identified as SUMO substrate in BF. Although the

404   crosstalk between these two related modifiers has been described in other organisms

405   (Lamoliatte et al., 2017), this is the first report on SUMOylation of ubiquitin (Ub) in *T.*

406   *brucei*, revealing novel connections of these two PTMs in this organism. SUMO and Ub

407   can interact in different ways; by competing for the same Lys residue on a substrate,

408   regulating the enzymes of the pathways required for the modification or participating on

409   polyubiquitination of different substrates mainly for proteasomal degradation, among

410   others (Hunter and Sun, 2008). Hybrid chains have been described in humans,

411   comprising of ubiquitination of SUMO and SUMOylation of Ub (Hendriks and Vertegaal,

412   2016). In this way, the finding of SUMOylated ubiquitin and SUMO itself in our site-

413   specific proteomic study confirms the existence of SUMO-SUMO and SUMO-Ub chains

414   in BF parasites, suggesting a complex crosstalk between these pathways.

415   From the differentially SUMOylated substrates identified in BF, we highlighted

416   TFIIS2-2, a non-characterized trypanosome homolog of transcription elongation factor

417   (TFIIS). To date, two TFIIS homologues have been identified (TbTFIIS1 and TbTFIIS1-

418   2), with a cooperating activity to promote transcriptional regulation necessary for mRNA

419   transcription (Uzureau et al., 2008). Recent work demonstrated TFIIS2-2 interaction

420   with trimethylated histones H4K17me3 and H3K32me3 with its PWWP domain (Wang

421   et al., 2019), reinforcing its function as a histone PTM reader involved in transcriptional

422   regulation. Another interesting finding is the SUMOylation of the RNA polymerase

423   associated subunit CTR9, an essential factor that regulates several genes involved in

424   the control of gene expression (Ouna et al., 2012). This protein is one of the core

425   components of the PAF complex, implicated in the transition from transcriptional

426 initiation to elongation (Yoo et al., 2013). Further validation of *in vitro* SUMOylation for

427 TFIIS2-2 and CTR9 indicates that the activities of these proteins might be influenced by

428 this PTM, and future mutational analysis will help to uncover its biological significance.

429   One of the most remarkable aspects of this work is the observation that five different

430 nuclear proteins (visualized by IFI, See Figure 4) can be recruited to a single intense

431 focus on the basis of their modification or association with SUMO (as demonstrated by

432 PLA analysis, See Figure 5). We believe that, similar to the role described for SUMO in

433 the formation of PML nuclear bodies (Matunis et al., 2006), *Tb*SUMO could have a role

434 as a scaffold or "glue" protein mediating interactions between SUMO moieties of

435 SUMOylated proteins and SUMO binding motifs on SUMO interacting proteins.

436   Also noteworthy is the association link between SUMO PTM and VSG expression,

437 determined by functional depletion of the previously validated substrates. As showed in

438 Figure 6, TFIIS2-2 depletion led to a deregulation of VSG-ES, suggesting the

439 importance of SUMO modification of this particular homolog as a requirement for its

440 regulatory role associated with VSG expression. Similarly, the H3V, previously found to

441 be enriched at telomeres (Lowell and Cross, 2004), showed significant inactive

442 expression upon depletion, probably due to an increased recombination since depletion

443 of H3V alone is not enough to induce switching itself, as previously reported (Muller et

444 al., 2018). This data agrees with the previously described role of H3V in maintaining

445 silencing of a subset of pol-I transcribed VSG genes (cite Schulz *et al.* 2016), while its

446 deletion increases its trancriptional silencing. These experimental evidences describe

447 for the first time an association of HSF related proteins and VSG regulation, remarking

448 the importance of this PTM in both antigenic expression and DNA accessibility.

449   Given the high complexity to understand the specific role of each identified target in

450   VSG expression; our purpose is to provide a snapshot of detailed SUMOylated

451   substrates to facilitate further related studies highlighting its relevance into the antigenic

452   variation field, as we previously demonstrated. Moreover, we also speculate that given

453   the prominence of some of the identified targets into the HSF throughout the nucleus

454   assessed by PLA, might be presumably involved in VSG expression.

455   Altogether, these results provide important new insights into the role of SUMO in

456   mediating the assembly of subnuclear structures, such as the HSF, and in regulating

457   the distribution of proteins between the HSF and the nucleoplasm, while at the same

458   time reinforcing its function as an epigenetic modification that regulates VSG

459   expression.

Experimental procedures

Ethics statement

Isolation of parasites from Wistar IGS rats was performed by trained technicians in the animal facility of the Institute of Parasitology and Biomedicine Lopez-Neyra (Spanish National Research Council) IPBLN-CSIC. These samples were obtained in compliance with policies approved by the Committee on Use and Care of Laboratory Animals of the IPBLN-CSIC.

Trypanosome cell lines

Trypanosoma brucei single marker (derived from *T. brucei* bloodstream form, Molteno Institute Trypanozoon antigenic type (MITat) 1.2; clone 221) cell line grown in HMI-9 medium supplemented of 10% inactivated Fetal Calf Serum was used for all experimental procedures. Plasmid construct pLEW100v5-His-HA-TbSUMO^{T106K} was previously described (Iribarren et al., 2015a). The His-HA-*Tb*SUMO^{T106K} transgenic bloodstream form cell line and any others generated in this work was obtained by transfection and selection procedures as previously described (Wirtz et al., 1999).

Plasmid Constructs

Full length sequences of CTR9 (Tb927.3.3220) and TbTFIIS2-2 (Tb927.2.3420) were amplified by PCR from genomic DNA in order to construct the plasmids for substrate expression, using the following primers: FwCTR9-Duet-NcoI (5'-CCATGGCCATGCAATATATATCGGAAC -3') RvCTR9-Duet-NcoI (5'-CCATGGATATATTCC-3') FwTFIIS2-2-Duet-NcoI (5'-

483   CCATGGCCATGATTCCATCATTTGCTC   -3')   Rv   TFIIS2-2-Duet-NcoI   (5'-

484   CCATGGACTGTTCAGCACCATCCAC -3'). Final amplification products were cloned

485   into the expression vector pET28a-3xFlag to express both N and C-terminal 3xFlag

486   fusion proteins through NcoI restriction sites. Full length or N-terminal PCR products for

487   HA epitope tagging of H3V (Tb927.10.15350), DNA topoisomerase 1B (Tb927.4.1330),

488   CTR9, Hypothetical protein (Tb927.10.14190) and TFIIS2-2 were cloned into the p3HA-

489   x-Hyg vector using the following set of primers (restriction sites and amplified PCR

490   products corresponding to the N-terminal or full-length appear specified as follows) :

491   H3v_Up (5'-GGCAAGCTTCAGCTGATGGCGCAAATGAAGAAAATAAC-3'), H3v_Low

492   (5'- GCCGGATCCCTCGAGTTAGTTACGCTCGCCTCGGAG-3') PvuII/BamHI (Full

493   length), TopoIB_Up (5'- GGCAAGCTTCAGCTGATGGGTAAGGCACAGAAGCCG-3'),

494   TopoIB_Low (5'-GCCGGATCCCTCGAGGCAGCGAATGGGATGCTTC-3) PvuII/XhoI

495   (N-terminal), CTR9_Up (5'- GGCAAGCTTCAGCTGATGCAATATATATCGGAACCC-3')

496   CTR9_Low (5'-GCCGGATCCCTCGAGTGAACATTCAAGAGCTCTCTC-3') PvuII/XhoI

497   (N-terminal),   Hypot_10.14190_Up   (5'-

498   GGCAAGCTTCAGCTGATGATTCATCCGAAGCTTATG-3') Hypot_10.14190_Low (5'-

499   GCCGGATCCCTCGAGTCAACAATGCCGTCGCTCGAT-3') PvuII/BamHI (Full length),

500   TFIIS2-2_Up (5'-GGCAAGCTTCAGCTGATGATTCCATCATTTGCTCCA-3') TFIIS2-

501   2_Low (5'-GCCGGATCCCTCGAGTCACTGTTCAGCACCATCCAC-3') PvuII/BamHI

502   (Full length). Final constructs were confirmed by sequencing.

503

504   *Indirect immunofluorescence*

505   Parasites were harvested by centrifugation at 1400g for 10 min at room temperature,

506   washed and dried on slide, and then fixed in 4% paraformaldehyde (PFA) for 20 min.

507   Slides were, washed with PBS, blocked with 0.5% blocking reagent (Roche) and

508   incubated with rat monoclonal anti-HA (1:500) high affinity (clone 3F10, Roche Applied

509   Science) diluted in 0.5% blocking reagent. Anti-Rat Alexa-Fluor 488 were used as

510   secondary antibody (Invitrogen). Cells were DAPI stained and visualized with Zeiss

511   fluorescence microscope Axio Imager, equipped with the AxioVision system. Images

512   were treated using ImageJ Fiji program version 1.51n software (National Institutes of

513   Health). Three-dimensional immunofluorescence (3D-IF) was carried out as previously

514   described (Landeira et al., 2009).

515

516   *Western blot analysis*

517   Parasites (~1.5 x 10^6 cells per ml) were collected by centrifugation and washed once in

518   Trypanosome Dilution Buffer (TDB) with 1X protease inhibitor cocktail (Roche) and 20

519   mM N-ethylmaleimide (NEM). Pellets were resuspended in the same buffer at a

520   concentration of 1 x 10^6 cells per µl and further quantified by Bradford protein assay

521   (BioRad). Whole cell lysates were prepared by diluting 1:1 in 2x Laemmli sample buffer

522   and heated at 99ºC for 5 min. Samples were loaded (~5 x 10^6 cells per lane or 40 µg of

523   proteins) in 4–20% precast polyacrylamide gels (Bio-Rad Laboratories) and analyzed by

524   Western blot using monoclonal anti-HA high affinity (1:500) (clone 3F10, Roche Applied

525   Science), mouse monoclonal anti-Flag M2 mouse monoclonal antibody (1:5000) (Sigma

526   Aldrich) and anti-*Tb*SUMO clone 1C9H8 mAb ascites (1:1000) as previously described

527   (Lopez-Farfan et al., 2014). Horseradish peroxidase-conjugated goat anti-rat secondary

528   antibody (Dako) (1:1000) was detected by chemiluminiscence using Lumi-Light Western

529   Blotting Substrate (Roche). Goat anti-Mouse IgG (H+L) DyLight 680 Conjugate (Thermo

530   Scientific) was scanned using an LI-COR Odyssey scanner and analyzed using

531   Odyssey IR imaging software 3.0.42.

532

533   *Isolation of parasites from blood*

534   Parasites expressing His-HA-*Tb*SUMO^{T106K} were propagated in Wistar IGS rats

535   (Charles River UK) during 3 days to a final burden of 1 x 10^8 parasites/ml. After

536   bleeding, total extracted blood was diluted 1:1 with Phosphate buffered saline (PBS),

537   distributed in Ficoll Histopaque (Sigma) containing tubes and harvested at 1000 x g

538   during 10 min at 4ºC. The supernatant containing the parasite ring phase was

539   transferred to a new tube containing Separation buffer (SB) and, after serum removal,

540   was then transferred into a pre-equilibrated DEAE column (Whatman DE52, pre-

541   swollen). After the addition of SB buffer into the DEAE column, the eluted parasites

542   were harvested at 1400 x g during 10 min at 4ºC along with protease inhibitors

543   supplemented with 20 mM NEM and immediately resuspended in Lysis buffer.

544

545   *Purification of TbSUMO^{T106K} cell line conjugates*

546   Isolated parasites (5x10^{10}) from Wistar IGS rats were collected by centrifugation,

547   washed with TDB supplemented with 20 mM NEM, immediately resuspended in lysis

548   buffer (6M Urea, 50 mM Tris-HCl, 500 mM NaCl, 5 mM ß-mercaptoethanol, pH 7.5) for

549   a final concentration of ~4x10^9 parasites/ml and sonicated until loss of viscosity. Lysates

550   were centrifuged for 30 min at 46.000 x g and filtered by 0.2 µm. Adjusted protein

551   concentration for the cleared supernatant was adjusted up to 50 mg per ml of Ni^{2+}-NTA

552   resin (GE Healthcare), pre-equilibrated with lysis buffer, and incubated overnight at 4ºC

553   with gentle stirring. The resin was then washed with 5 column volumes (CV) of Lysis

554   buffer, followed by 10 CV of buffer 1 (8M Urea, 10 mM mM Tris-HCl ,100 mM

555   Na_2HPO_4/NaH_2PO_4, 10 mM Imidazol, 5 mM ß-mercaptoethanol, pH 8.0), 10 CV of

556   buffer 2 (8M Urea, 10 mM mM Tris-HCl ,100 mM Na_2HPO_4/NaH_2PO_4, 10 mM Imidazol,

557   5 mM ß-mercaptoethanol, pH 6.3) and 10 CV of buffer 1. Proteins were eluted three

558   times with 2 CV of elution buffer (8M Urea, 10 mM mM Tris-HCl ,100 mM

559   Na_2HPO_4/NaH_2PO_4, 200 mM Imidazol, 5 mM ß-mercaptoethanol, pH 8.0) and

560   concentrated by acetone precipitation (80% v/v). Eluted fractions were analyzed by WB

561   to validate the purification efficiency. Enrichment of His-HA-*Tb*SUMO^{T106K} conjugates by

562   immunoaffinity purification with the K-ε-GG-specific antibody after Lys-C or Lys-C/Glu-C

563   digestion was performed according to the published protocol (Tammsalu et al., 2015).

564

565   *MS/MS proteomic analysis*

566   Processed peptides were desalted using equilibrated C18-bonded silica StageTips

567   (Pierce) and vacuum-dried before reconstituting with 0.1% (v/v) of trifluoroacetic acid

568   (TFA). The protein mixtures were then separated using a linear acetonitrile gradient (5-

569   90%) in an EASY-nLC 1000 Liquid chromatography system coupled to Q Exactive mass

570   spectrometer via EASY-Spray ion source, including EASY-Spray analytical column (50

571   cm x 75 µm Ø) and Acclaim PepMap 100 precolumn (2 cm x 75 µm Ø) bypasses. For

572   protein identification, the mass spectra were deconvoluted using MaxQuant version

573   1.3.0.5 searching the *TriTrypDB_8.0_TbruceiLister427* annotate protein database.

574 Search engine was performed in Andromeda version 1.5.6.0 with pre-established

575 parameters (Fixed Modifications: Carbamidomethyl (C), Variable Modifications:

576 Oxidation (M), Acetylation (protein N-term), Gly-gly (K, not C-term), Phosphorylation

577 (Bielska et al.), Enzyme: Lys-C/P (3 max. missed cleavages), Lys-C/P and Glu-C (5

578 max. missed cleavages), Special amino acids: K, MS/MS Tolerance: 20 p.p.m, Main

579 search: 6, First search: 20 p.p.m., Top MS/MS peaks per 100 Da: 10). MS-spec data

580 was visualized with a false discovery rate (FDR) of 0.01% for peptide and protein

581 thresholds with 7 residues as a minimum peptide length.

582

583 *Bioinformatic analysis*

584 Analysis of consensus SUMO sequences were performed using *pLogo* web server and

585 conserved residues were scaled relative to Bonferroni-corrected statistical significance

586 values, using *TriTrypDB-9.0_Tbrucei927* input data set as background.

587

588 *In vitro SUMOylation assay in bacteria*

589 SUMOylation assay *in bacteria* was performed as previously described (Iribarren et al.,

590 2015b). SUMOylation system was induced in *Escherichia coli* BL21 (DE3) cells

591 transformed with pCDFDuet-1-TbSUMO/TbE2, followed by pACYCDuet-1-TbE1a-

592 TbE1b. Competent bacteria were transformed again with pET28-CTR9-3xFlag or

593 pET28-TFIIS2-2-3xFlag. Cells containing the three plasmids were grown in Luria-

594 Bertani (LB) medium at 37ºC to an OD_{600nm} of 0.6, and induced with 1 mM isopropyl ß-

595 D-1-thiogalactopyranoside (IPTG) for 5 h at 37ºC with shaking (250 rpm). Harvested

596 cells were resuspended in lysis buffer (150 mM NaCl, 50 mM Tris HCl, 0.4 mg/ml

597 lysozyme, 0.1% Triton X-100, 10 mM ethylene diamine tetraacetic acid (EDTA), 1 mM

598 phenylsulfonyl fluoride (PMSF)-pH 7.6) and sonicated. Centrifuged samples (70 µl)

599 were resuspended in Laemmli sample buffer with 100 mM DTT, boiled for 5 min and

600 analyzed by Western blot.

601

602 *Proximity Ligation Assay*

603 The PLA assay was done according to the manufacturer's protocol (Olink Bioscience,

604 Uppsala, Sweden). Cells were dried on slides, fixed 20 min in 4% PFA and

605 permeabilized 20 min with 1% NP40. After blocking, primary antibodies (2 µg/ml rabbit

606 anti-HA tag antibody (abcam, ab9110) and mouse anti-TbSUMO 1C9H8 (1:2000)) were

607 incubated 1h. Then, the PLA probe solution containing the secondary antibodies

608 conjugated with oligonucleotides (PLA probe MINUS and PLUS) were applied.

609 Secondary species-specific antibodies conjugated with oligonucleotides would hybridize

610 to the two PLA probes if they were in close proximity (<40nm). The slides were then

611 incubated with the ligation solution together with the ligase which would join the two

612 hybridized oligonucleotides to a closed circle. The amplification mix containing

613 nucleotides and polymerase was applied to the slides. The oligonucleotide arm of one

614 of the PLA probes acts as a primer for rolling circle amplification (RCA) reaction,

615 generating a repeated sequence and extended product. Finally, the nuclei DNA were

616 stained with DAPI and the slides were examined under fluorescent microscopy. As

617 negative control and to discard cross-reaction signal, complete protocol was performed

618 in parallel using primary antibodies separately.

619

620 *RNAi experiments and RT-qPCR analysis*

621 TFIIS2-2 and H3V constructs for RNA interference experiments were made as

622 previously reported using the p2T7Bla vector (Penate et al., 2009). PCR fragments for

623 each substrate were generated with the primers described in *Plasmid Constructs*.

624 Amplified PCR fragments corresponding to full-length sequences of TFIIS2-2 (1146 bp)

625 and H3V (420) were cloned into BamHI and HindIII sites of p2T7Bla and transfected

626 into single marker cell line, used as reference control. Synthesis of dsRNA was induced

627 by adding 1µg/ml of doxycycline. At least three independent clones were analyzed.

628 Total RNA was isolated from $5x10^7$ parasites using the RNAsy Mini kit (Qiagen),

629 previously treated with *DNaseI* (Qiagen), following manufacturer's instructions. The

630 RNA concentration was verified by nanodrop at A_{260}/A_{280} ratio prior to cDNA synthesis

631 from 1 µg of RNA with the qScript cDNA Synthesis Kit (Quanta) following

632 manufacturer's instructions. To discard genomic DNA contamination, an untreated

633 reverse transcriptase sample was included. qPCR was performed using SYBR green

634 Supermix (BioRad) in a CFX96 cycler (BioRad) with 1 µl of cDNA of undiluted and 1:100

635 dilution of single or multicopy genes, respectively, including 5 µl of 2x SYBR mix and

636 500 nM of each pair of primers in a final volume of 10 µl. Duplicate reactions were

637 subjected to a previous denaturalization during 3 min at 95ºC, followed by 32 cycles of

638 30 sec at 95ºC / 30 sec at 60ºC / 30 sec at 72ºC, with a final extension of 1 min at 72ºC.

639 Fluorescent reads were measured during the extension step. Primer sequences

640 corresponding to 18S, C1, U2, VSG221 and VSG121 are detailed in (Lopez-Farfan et

641 al., 2014). Relative expression levels were referred to a control (parental cell line) and

642 normalized against a housekeeping gene (U2) using the software BioRad CFX Manager

643 Software. Experimental conditions were performed in triplicate and analyzed by

644 Student's T-test.

645

646 **Acknowledgements**

651

652 **Author Contributions**

653 Conceived and designed the experiments: MN, AS, VA.

654 Performed the experiments: AS, PI, CGL, STR.

655 Contributing reagents/materials/analysis tools: AS PI, DRB, STR, CGL.

656 MN, AS, VA, wrote the paper with input from all authors.

657

658 **Conflict of interest**

659 The authors declare no conflicts of interest.

660

661 **Funding information**

References

Albuquerque CP, Wang G, Lee NS, Kolodner RD, Putnam CD, et al. (2013) Distinct SUMO Ligases Cooperate with Esc2 and Slx5 to Suppress Duplication-Mediated Genome Rearrangements. PLoS Genet 9(8) e1003670.

Bayona, J.C., Nakayasu, E.S., Laverrière, M., Aguilar, C., Sobreira, T.J.P., Choi, H., Nesvizhskii, A.I., Almeida, I.C., Cazzulo, J.J., and Alvarez, V.E. (2011). SUMOylation Pathway in Trypanosoma cruzi: Functional Characterization and Proteomic Analysis of Target Proteins. Molecular & Cellular Proteomics 10.

Appikonda, S., Thakkar, K.N., Shah, P.K., Dent, S.Y.R., Andersen, J.N., and Barton, M.C. (2018). Cross-talk between chromatin acetylation and SUMOylation of tripartite motif–containing protein 24 (TRIM24) impacts cell adhesion. The Journal of Biological Chemistry 293, 7476-7485.

Bielska, K., Seliga, J., Wieczorek, E., Kędracka-Krok, S., Niedenthal, R., and Ożyhar, A. (2012). Alternative sumoylation sites in the Drosophila nuclear receptor Usp. The Journal of Steroid Biochemistry and Molecular Biology 132, 227-238.

Cestari, I., and Stuart, K. (2018). Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes. Curr Genomics 19, 119-132.

Deyrieux, A.F., and Wilson, V.G. (2017). Sumoylation in Development and Differentiation. In SUMO Regulation of Cellular Processes, V.G. Wilson, ed. (Cham: Springer International Publishing), pp. 197-214.

Duraisingh, M.T., and Horn, D. (2016). Epigenetic Regulation of Virulence Gene Expression in Parasitic Protozoa. Cell host & microbe 19, 629-640.

Enserink, J.M. (2015). Sumo and the cellular stress response. Cell Div 10, 4.

Han, Z., Feng, Y., Gu, B., Li, Y., Chen, H. (2018). "The post-translational modification, SUMOylation, and cancer (Review)". International Journal of Oncology 52, 1081-1094.

Hendriks, I.A., and Vertegaal, A.C. (2016). A comprehensive compilation of SUMO proteomics. Nat Rev Mol Cell Biol 17, 581-595.

Hunter, T., and Sun, H. (2008). Crosstalk between the SUMO and ubiquitin pathways. Ernst Schering Found Symp Proc, 1-16.

Iribarren, P.A., Berazategui, M.A., Bayona, J.C., Almeida, I.C., Cazzulo, J.J., and Alvarez, V.E. (2015a). Different proteomic strategies to identify genuine Small Ubiquitin-like MOdifier targets and their modification sites in Trypanosoma brucei procyclic forms. Cellular microbiology 17, 1413-1422.

Iribarren, P.A., Berazategui, M.A., Cazzulo, J.J., and Alvarez, V.E. (2015b). Biosynthesis of SUMOylated Proteins in Bacteria Using the Trypanosoma brucei Enzymatic System. PLoS One 10, e0134950.

Iribarren, P.A., Di Marzio, L.A., Berazategui, M.A., De Gaudenzi, J.G., and Alvarez, V.E. (2018). SUMO polymeric chains are involved in nuclear foci formation and

709 chromatin organization in Trypanosoma brucei procyclic forms. PLOS ONE *13,*
710 e0193528.

711 Klein, C.A., Droll, D., and Clayton, C. (2013). SUMOylation in Trypanosoma brucei.
712 PeerJ 1, e180-e180.

713 Lamoliatte, F., McManus, F.P., Maarifi, G., Chelbi-Alix, M.K., and Thibault, P. (2017).
714 Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using
715 sequential peptide immunopurification. Nat Commun 8, 14109.

716 Landeira, D., Bart, J.M., Van Tyne, D., and Navarro, M. (2009). Cohesin regulates VSG
717 monoallelic expression in trypanosomes. J Cell Biol 186, 243-254.

718 Lopez-Farfan, D., Bart, J.M., Rojas-Barros, D.I., and Navarro, M. (2014). SUMOylation
719 by the E3 ligase TbSIZ1/PIAS1 positively regulates VSG expression in
720 Trypanosoma brucei. PLoS Pathog 10, e1004545.

721 Matunis, M.J., Zhang, X.D., and Ellis, N.A. (2006). SUMO: the glue that binds. Dev Cell
722 11, 596-597.

723 McCulloch, R., and Navarro, M. (2016). The protozoan nucleus. Molecular and
724 biochemical parasitology 209, 76-87.

725 Monribot-Villanueva, J., Zurita, M., and Vázquez, M. (2016). Developmental
726 transcriptional regulation by SUMOylation, an evolving field. genesis 55, e23009.

727 Moretti, N.S., Cestari, I., Anupama, A., Stuart, K., and Schenkman, S. (2018).
728 Comparative Proteomic Analysis of Lysine Acetylation in Trypanosomes. Journal of
729 Proteome Research 17, 374-385.

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

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

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

738 Niskanen, E.A., and Palvimo, J.J. (2017). Chromatin SUMOylation in heat stress: To
739 protect, pause and organise? BioEssays 39, 1600263.

740 O'Shea, J.P., Chou, M.F., Quader, S.A., Ryan, J.K., Church, G.M., and Schwartz, D.
741 (2013). pLogo: a probabilistic approach to visualizing sequence motifs. Nature
742 Methods 10, 1211.

743 Obado, S.O., Bot, C., Echeverry, M.C., Bayona, J.C., Alvarez, V.E., Taylor, M.C., and
744 Kelly, J.M. (2011). Centromere-associated topoisomerase activity in bloodstream
745 form Trypanosoma brucei. Nucleic Acids Research 39, 1023-1033

746 Ouna, B.A., Nyambega, B., Manful, T., Helbig, C., Males, M., Fadda, A., and Clayton, C.
747 (2012). Depletion of Trypanosome CTR9 Leads to Gene Expression Defects. PLOS
748 ONE 7, e34256..

749 Penate, X., Lopez-Farfan, D., Landeira, D., Wentland, A., Vidal, I., and Navarro, M.
750   (2009). RNA pol II subunit RPB7 is required for RNA pol I-mediated transcription in
751   Trypanosoma brucei. EMBO reports 10, 252-257.

752 Rico, E., Ivens, A., Glover, L., Horn, D., and Matthews, K.R. (2017). Genome-wide RNAi
753   selection identifies a regulator of transmission stage-enriched gene families and
754   cell-type differentiation in Trypanosoma brucei. PLoS Pathog 13, e1006279.

755 Tammsalu, T., Matic, I., Jaffray, E.G., Ibrahim, A.F., Tatham, M.H., and Hay, R.T.
756   (2015). Proteome-wide identification of SUMO modification sites by mass
757   spectrometry. Nat Protoc 10, 1374-1388.

758 Uzureau, P., Daniels, J.-P., Walgraffe, D., Wickstead, B., Pays, E., Gull, K., and
759   Vanhamme, L. (2008). Identification and characterization of two trypanosome TFIIS
760   proteins exhibiting particular domain architectures and differential nuclear
761   localizations. Molecular Microbiology 69, 1121-1136.

762 Wang, R., Gao, J., Zhang, J., Zhang, X., Xu, C., Liao, S., and Tu, X. (2019). Solution
763   structure of TbTFIIS2-2 PWWP domain from Trypanosoma brucei and its binding to
764   H4K17me3 and H3K32me3. Biochem J 476, 421-431.

765 Wang, Y., and Dasso, M. (2009). SUMOylation and deSUMOylation at a glance. J Cell
766   Sci 122, 4249-4252.

767 Wirtz, E., Leal, S., Ochatt, C., and Cross, G.M. (1999). A tightly regulated inducible
768   expression system for conditional gene knock-outs and dominant-negative genetics
769   in Trypanosoma brucei. Molecular and Biochemical Parasitology 99, 89-101.

770 Yang, X., Wu, X., Zhang, J., Zhang, X., Xu, C., Liao, S., and Tu, X. (2017). Recognition
771   of hyperacetylated N-terminus of H2AZ by TbBDF2 from <em>Trypanosoma
772   brucei</em&gt. Biochemical Journal 474, 3817.

773 Yoo, H.S., Seo, J.H., and Yoo, J.Y. (2013). CTR9, a component of PAF complex,
774   controls elongation block at the c-Fos locus via signal-dependent regulation of
775   chromatin-bound NELF dissociation. PLoS One 8, e61055.

776

777

778

779 **Supporting Information**

780 Additional supporting information may be found online in the Supporting Information

781 section at the end of the article.

782

783

Table 1. Site-specific identification of SUMOylated lysines and SUMOylated proteins in

785 *T. brucei* bloodstream form. The modified Lys residues appear in bold.

786

Gene ID	Description	SUMOylated Lysine	SUMOylated peptide	Consensus motif
Tb927.7.2830	Histone H2A	5	MATP**K**QAVKKASKGGSSRSV	
Tb927.7.6360	Histone H2A variant (h2aZ)	32	GVAMSPEQASALTGG**K**LGGKAVGPAHGKGKG	
Tb927.10.15350	H3V (h3vaR)	31	SVASRPIQAVARAPV**K**KVENTPPQKRHHRWR	
Tb927.4.1330	DNA topoisomerase IB. large subunit	20	QKPKSGEGKGKKVAV**K**DEEVNGKRVVVKKED	+
		47	KKEDMTEEKIKKVVI**K**EEENELEMVAAGMGP	
Tb927.1.1170	DNA-directed RNA polymerase subunit RBP12 (RBP12)	7	MLSYTV**K**EEVKDEKLPGANNFA	+
Tb927.3.3220	RNA polymerase-associated protein CTR9	838	LEDFKELHGHRVPQV**K**NENEGFAESPAPWFS	+
Tb927.9.10680	RNA polymerase III RPC4	171	AEPKHEAEFSVEGDV**K**VPVAETGNDGIAFLK	Reversed
Tb927.9.5190	Proliferative cell nuclear antigen (PCNA)	195	ALLRASHAPTVDPRS**K**GESDVKTEDEEADAC	+
Tb927.10.740	Structural maintenance of chromosome 4. (SMC4)	623	QIGNRMETPFTSPTP**K**AKRLFDLITPVNDRF	+
		1302	ERKRKRTGSTGDVQI**K**VEDEVACNNEAADIL	
Tb927.2.3480	Transcription elongation factor s-II (TFIIS2-2)	117	KAQAGEANERRVRGV**K**TEFADGGEQKAVEAP	+
Tb927.5.3210	Small ubiquitin-related modifier. putative (TbSUMO)	65	KSRTALKKLIDTYCK**K**QGISRNSVRFLFDGT	
Tb927.11.9920	Polyubiquitin	48	KEGIPPDQQRLIFAG**K**QLEEGRTLADYNIQK	
		63	KQLEEGRTLADYNIQ**K**ESTLHLVLRLRGGMQ	
Tb927.10.13720	RNA-binding protein. putative (RBP29)	313	GGEDTEPEKSATLPM**K**TESDSVSCPMEVNCN	+
Tb927.9.6870	RNA-binding protein. putative (RBSR1)	23	NIEKRATKKELLEFL**K**PMEEHIEDSWLARRP	
Tb927.9.15060	rRNA processing protein. putative	160/161	ARIEERAHRRAMKDQ**KK**YGKEVQAEVLRQRA	Reversed
Tb927.10.1630	ATP-binding cassete sub-family e member 1	187	KMRVLLKPQYVDQVP**K**VTKGKVGDLLTKADE	
Tb927.10.14520	Basal body protein	371	VPAKDPGTAPSAVVA**K**SEKEEPPAAKTPRPT	+
Tb927.9.13880	Kinetoplastid membrane protein KMP-11	76	KFNKKMREHSEHFKA**K**FAELLEQQKNAQFPG	
Tb927.10.6720	Plasma-membrane choline transporter	16	MMQGFPSLGEKDPAA**K**PPAEGKPTSASGEKQ	
Tb927.9.9400	Ceramide phosphorylethanolamine synthase	312	SREVUEDGVPVAIVI**K**NEEMMNFDGKS	
Tb927.5.630	Acidic phosphatase	391	KDRKEDVASGSVHQG**K**GMPSTNIDPF	
Tb927.9.1560	DnaJ domain containing protein	407	KPTLKPSVSKSSISK**K**APTTSGSAKKAPTKR	
		416	KSSISKKAPTTSGSA**K**KAPTKRVVKPVKKAG	
Tb927.5.370	75 kDa invariant surface glycoprotein. putative (ISG75)	514	DDINIEEGGAKSKNT**K**TAAGLDSDI	
Tb927.3.2350	Hypothetical protein	35	SRINQALELATKVTV**K**EEPDTENGGGKAAII	+
Tb927.11.5230	Hypothetical protein	24	AAIYARAEEAQNAPV**K**LQPVPAFELYKAARE	
Tb927.9.13320	Hypothetical protein	484	VADGLAAVGLPEIRK**K**TDEEGEAEAEVAVKP	+
		498	KKTDEEGEAEAEVAV**K**PEPVVSLDG	
Tb927.10.12030	Hypothetical protein	100	GKGGTNKTSHDFELT**K**ALAKDGSIHREKRRL	Reversed
		139	ERFHKGNAHYSVSDS**K**PTKRRKRDHSVDKDA	
		167	KDAVVEAEFGSQSIN**K**KERSVKASRKDSNRE	
Tb927.4.2620	Hypothetical protein	235	ALTSAEERRHAEVDV**K**PNIYPTAT	Reversed
Tb927.3.4140	Hypothetical protein	455	ASIRQQMKNKKNAEI**K**VEEIEDVAVDDGAAA	+
Tb927.9.1410	Hypothetical protein	228	APEPLQCETEAVGNI**K**EELCKLQDSTAEVRL	+
Tb927.10.14190	Hypothetical protein	332	VVGRKRERRSCESGI**K**IEDHNNETVVEVGEL	+
Tb927.11.11840	Hypothetical protein	69	YCSRAPTEEKVVAQV**K**EEDGRGTGSSSRRNI	+
Tb927.10.15760	Hypothetical protein	2	M**K**SEGTPIPTRTPKSGK	+
Tb927.2.4520	Hypothetical protein	354	EDRYMFVVDQGNNTQ**K**EILDGIRAYIGGEIE	
Tb927.11.4920	Hypothetical protein	138	KEKDFVHRYAEAKMK**K**SIPREESDDVAKKVA	
Tb927.3.5370	Hypothetical protein	11	MPFWKYFVIT**K**PEVDDDDLPLAKHRI	+
Tb927.7.5660	Hypothetical protein	535	GPFTAHLLDPMPKFI**K**EETDTSKPAGLTKPR	+

787
788

Figure Legends

790 **Fig. 1.** Bloodstream form expression and localization of His-HA-TbSUMO[T106K] is similar

791 wildtype TbSUMO. A. Expressing His-HA-*Tb*SUMO[T106K] parasites are able to conjugate

792 substrates. Protein cell extracts were prepared with 20 mM NEM and separated in a

793 10% acyilamide gel (5×10^6 cells/lane). Parental and 24 h doxycycline induced proteins

794 (*Tb*SUMO[T106K]) were incubated against anti-HA mAb. The same blot was also probed

795 with anti-TbSUMO mAb (1C9H8) to detect endogenous SUMO conjugates. B.

796 Immunofluorescence analysis of parasites expressing His-HA-*Tb*SUMO[T106K] (24 h after

797 induction) stained using an anti-HA antibody displayed a similar nuclear pattern

798 including the High SUMOylated Focus (HSF) detected by the monoclonal antibody

799 against TbSUMO in wild type cells as described previously (Lopez-Farfan, et al. 2014).

800 DNA content was stained with DAPI (blue). Merged signal image is shown. **C.** Growth

801 curve of His-HA-*Tb*SUMO[T106K] strain with (Dox+) or without (Dox-) doxycycline

802 induction. **D.** Expression pattern of His-HA-*Tb*SUMO[T106K] conjugates (*Tb*SUMO[T106K])

803 from whole cell extracts after isolation from Wistar IGS rats (Charles River UK). Controls

804 for parental (Non tagged) and HA epitope (HA control) were included. Samples were

805 prepared under denaturing conditions and loaded in a 4-20% precast polyacrylamide

806 gel (BioRad). **E.** Western blot analysis of representative eluted fractions (Eluates 1-3) of

807 His-HA-*Tb*SUMO[T106K] conjugates after Ni^{2+}-NTA purification in expressing parasites

808 isolated from Wistar IGS rat and lysed under denaturing conditions (Lysed extract

809 T106K).

810

811 **Fig. 2.** Analysis of SUMOylation sites. Representation of over- and underrepresented

812 residues are shown above and below the X axis, respectively. Bonferroni-corrected

813 statistical significance values are represented in horizontal ruler lines above and below

814 the X axis. The Y axis corresponds to the logarithmic binomial probability of the

815 statistical significance. Red horizontal lines correspond to threshold values of ±3.66,

816 considering p<0.05. Figure legend: n(fg): number of aligned sequences in the input data

817 set (foreground). n(dg): TriTrypDB-9.0_Tbrucei927 protein data set used as background

818 data.

819

820 **Fig. 3.** *In bacteria* SUMOylation of CTR9 and TFIIS2-2 using trypanosome enzymatic

821 system*.* Validation of lysine acceptor sites for CTR9 (**A.**) and TFIIS2-2 (Han) by *in vitro*

822 SUMOylation assay in bacteria. Anti-Flag Western blot analysis performed on soluble

823 cell extracts from induced *E.coli* cultures transformed with pET28-CTR9-3xFlag (CTR9)

824 and pET28-TFIIS2-2-3xFlag (TbTFIIS2) alone (lane 1 in A. and B., respectively) or in

825 the complete SUMOylation system background (pCDFDuet-1-*Tb*SUMO-*Tb*E2 plus

826 pACYCDuet-1-*Tb*E1a-*Tb*E1b) for CTR9 (CTR9-SUMO-E1-E2, lane 2 in A.) and

827 TbTFIIS2 (TbTFIIS2-SUMO-E1-E2, lane 2 in B). Anti-Flag signal intensity was scanned

828 using an LI-COR Odyssey and analyzed using ImageQuant software.

829

830 **Fig. 4**. N-Terminal 3-HA tagging of representative SUMO conjugated proteins.

831 Immunofluorescence analysis using anti-HA mAb of 3HA tagged SUMO conjugated

832 partners corresponding to H3V (Tb327.10.15350), DNA topoisomerase 1B

833 (Tb927.4.1330), CTR9 (Tb927.3.3220), Hypothetical protein (Tb927.10.14190) and

834   TFIIS2-2 (Tb927.10.14190) after 24 h of Doxycycline induction. DNA content was

835   stained with DAPI (blue). Merged anti-HA-DAPI images are shown.

836

837   **Fig. 5.** *In situ* analysis of SUMO mediated interactors. Cells expressing 3HA tagged

838   proteins were detected by *in situ* Ligand Proximity Assay in BF using rabbit anti-HA

839   polyclonal antibody (Abcam) and mouse anti-TbSUMO antibody after 24 h of

840   Doxycycline induction. Control reactions were performed without primary antibodies.

841   DNA was DAPI stained. Histogram shows the percentage of PLA-positive cells for the

842   TbSUMO/HA reaction. The % (±standard deviation) of positive PLA signals including

843   false positives from each antibody incubated alone were quantified in 100x objective. At

844   least 2 replicates per cell line were analyzed. Statistical analysis (Student T test)

845   *p<0.05 (referred to anti-HA positive PLA).

846

847   **Fig. 6**. Depletion of TFIIS2-2 and H3V deregulates both active and inactive VSG-ES. **A.**

848   Relative transcript levels upon 24h of RNAi shows deregulation of active VSG221 and

849   inactive VSG121. Only 18S ribosomal rDNA is altered after H3V depletion. C1 (AN1-like

850   zinc finger, Chr. X) is represented as a control locus for qPCR valdation. Results

851   represent the average from three independent clones and data normalized with U2

852   mRNA (mean±SEM). Statistical analysis (Student T-test): p<0.05, **p<0.01, ***p<0.01,

853   NS: Non-significant. **B.** Growth curve of TFIIS2-2 depleted cells with (Dox+) or without

854   (Dox-) doxycycline induction. **C.** Control of TFIIS2-2 depletion by RNAi for the assay

855   shown in D. Cell extracts ($5x10^6$ cells/lane) were analyzed by Western blotting (upper

856   panel) using the anti-TFIIS2-2 mAb (7E10G10). Ponceau (lower panner) shows loading

857 control. Control and doxycycline induced (Dox+) and uninduced (Dox-) are shown. **D.**

858 Volcano plot showing overexpressed DEGs (FDR<0.05) after 24h TFIIS2-2 depletion in

859 bloodstream forms. **E.** Histogram illustrating the proportion of cells expressing VSG221,

860 VSG121, double expressers (Double VSG221/VSG121) and switched off cells (VSG221

861 off) by double direct IF after H3V depletion by RNAi. Statistical analysis (Student's T-

862 test) shows difference in the mean within the control and induced (Dox+) cells in both

863 proportion of cells expressing VSG221 and VSG221/VSG121 (p<0.05).

864

865 **ORCID:**

866 Miguel Navarro: https://orcid.org/0000-0003-2301-2699

867 Vanina Alvarez: https://orcid.org/0000-0002-9890-6980

868

A

Parental *Tb*SUMO Parental *Tb*SUMO^T106K ^T106K

250—
150—
100—
75—
50—
37—
25—
15—

*Tb*SUMO^T106K conjugates

*Tb*SUMO^T106K

Ab anti-HA   mAb anti-*Tb*SUMO

B

Ab anti-HA   Ab anti-TbSUMO

TbSUMO^T106K   WT

DAPI   DAPI

Merge   Merge

C

*Tb*SUMO^T106k Growth curve

Parasites . ml⁻¹

— Parental
— Dox-
— Dox+

Hours (h)

D

Non tagged   HA Control   *Tb*SUMO^T106K

250—
130—
100—
70—
55—
35—
15—

*Tb*SUMO^T106K conjugates

*Tb*SUMO^T106K

anti-HA

E

Lysed extract T106K   Ni²⁺-NTA Eluate 1   Ni²⁺-NTA Eluate 2   Ni²⁺-NTA Eluate 3

250—
130—
100—
70—
55—
35—
25—
15—

*Tb*SUMO^T106K conjugates

*Tb*SUMO^T106K

anti-HA

Saura et al. Figure 1

Saura et al. Figure 2

A
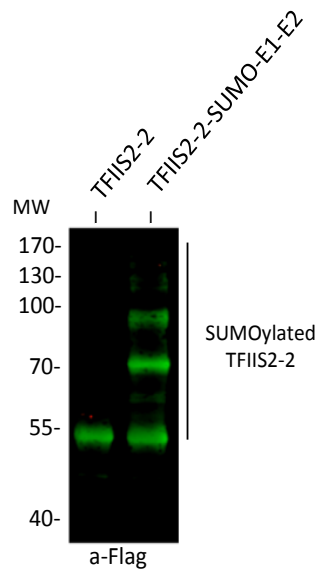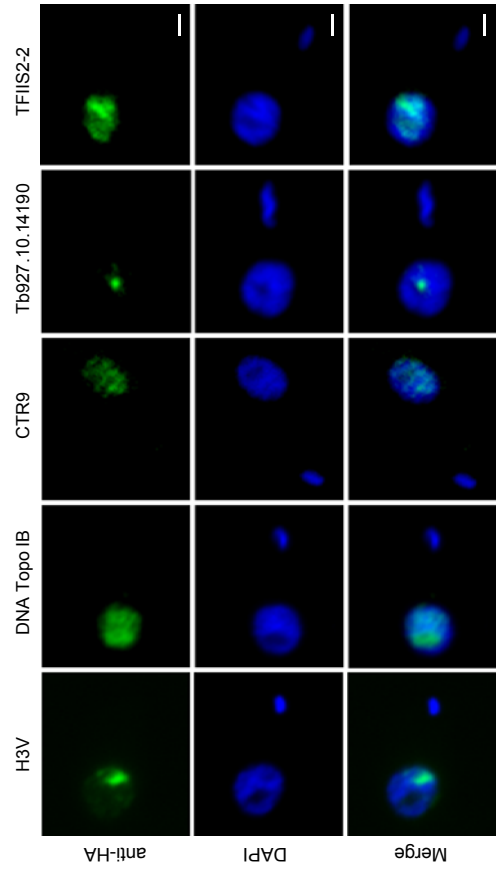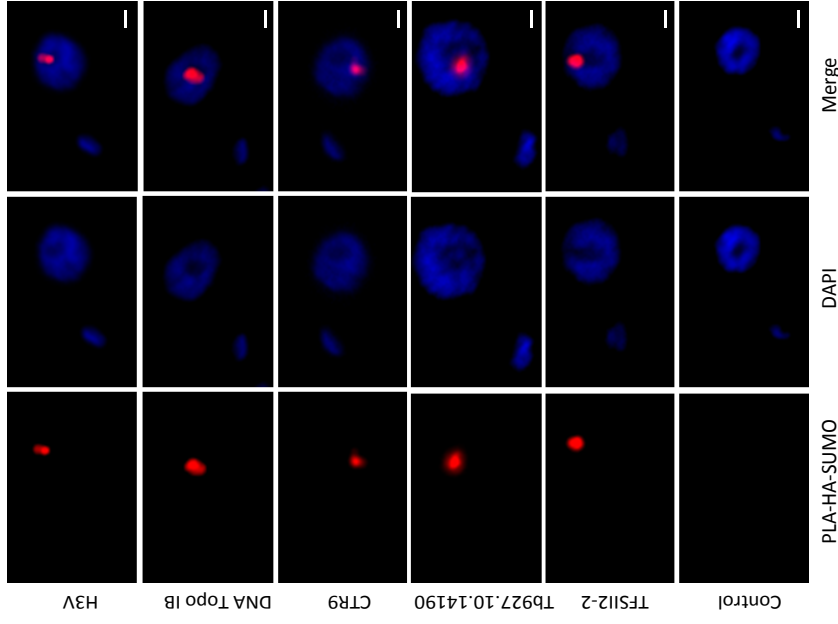
MW

170-
130-
100-
70-
55-

40-

35-

CTR9 | CTR9-SUMO-E1-E2

SUMOylated
CTR9

a-Flag

B

MW

170-
130-
100-

70-

55-

40-

TFIIS2-2 | TFIIS2-2-SUMO-E1-E2

SUMOylated
TFIIS2-2

a-Flag

Saura et al. Figure 3

Saura et al. Figure 5

Cell Line	a-HA (%)	a-TbSUMO (%)	a-HA/TbSUMO (%)
H3V	1.21±0.05	0.17±0.17	20.16±0.88
DNA Topo IB	2.49±0.51	3.73±1.56	56.10±4.25
CTR9	5.21±4.33	3.71±2.08	46.55±4.91
Tb927.10.14190	5.33±1.76	2.30±1.63	32.73±3.46
TFIIS2-2	2.07±1.16	3.73±1.56	25.85±1.93

Saura et al. Figure 6
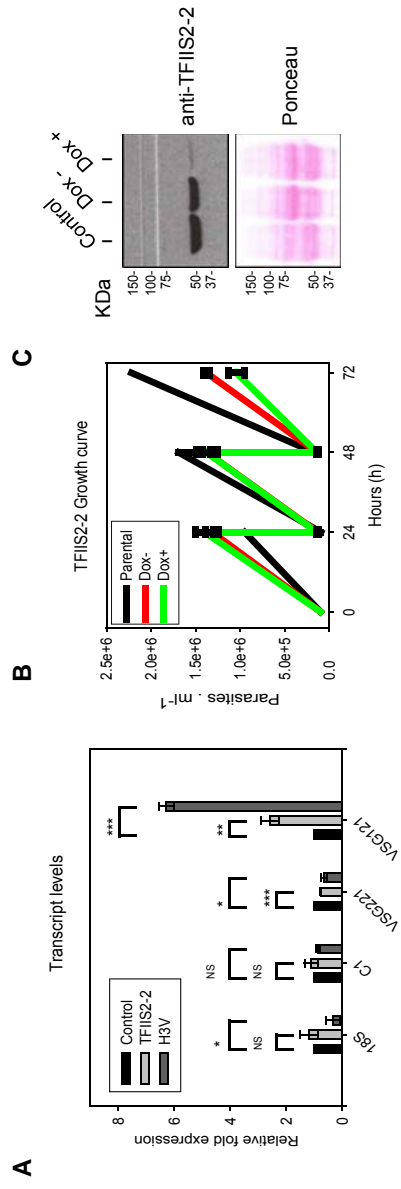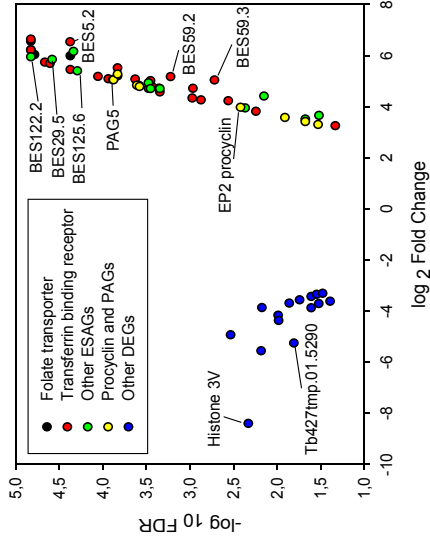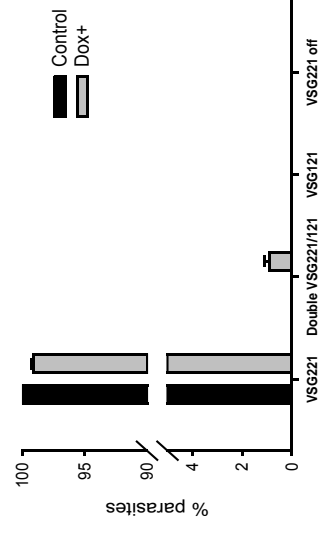
Supplemental tables

Table S1. Identification of putative SUMOylation substrates after tandem affinity HA/His purification. Data from three independent proteomic purifications are represented in separate tables **(A, B, C)** after background subtraction (control) for a protein and peptide threshold of 95% and 0.5% FDR, respectively. Exclusive number of peptides for the identified substrate is shown.

Table S1A

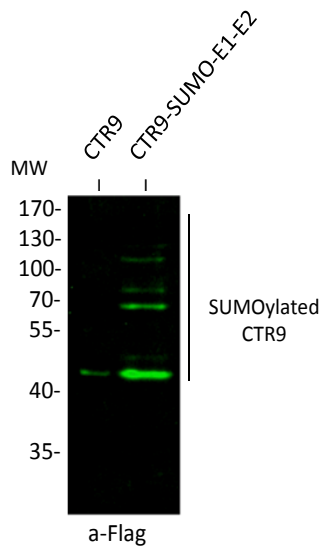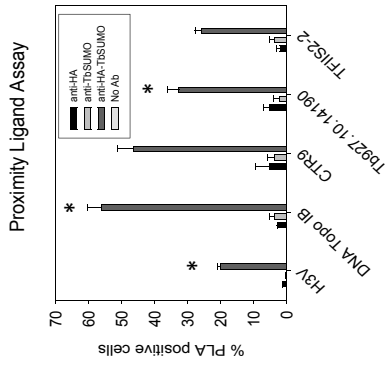| Accession Number	MW	Identified Proteins (48/280)	HA	Control
Tb927.10.740	155 kDa	structural maintenance of chromosome 4, putative (SMC4)	18	0
Tb927.5.3210	12 kDa	small ubiquitin-related modifier, putative (TbSUMO)	8	0
Tb927.9.5190	32 kDa	proliferative cell nuclear antigen (PCNA), putative	8	0
Tb927.9.13320	55 kDa	hypothetical protein, conserved	7	0
Tb927.7.5310	73 kDa	hypothetical protein, conserved	4	0
Tb927.6.1070	102 kDa	hypothetical protein, conserved	4	0
Tb927.3.4140	62 kDa	hypothetical protein, conserved	3	0
Tb927.7.5090	102 kDa	hypothetical protein, conserved	3	0
Tb927.8.4990	70 kDa	69 kDa paraflagellar rod protein,PFR2 (PFR-B)	3	0
Tb927.11.9920	77 kDa	polyubiquitin, putative	2	0
Tb927.5.4220	11 kDa	histone H4, putative	2	0
Tb927.3.5370	34 kDa	hypothetical protein, conserved	2	0
Tb927.7.3950	107 kDa	RNA-editing 3' terminal uridylyl transferase 1,KRET1 (KRET1)	2	0
Tb927.4.1300	42 kDa	hypothetical protein, conserved	2	0
Tb927.11.3120	75 kDa	nucleolar GTP-binding protein 1 (NOG1)	2	0
Tb927.9.2490	67 kDa	nucleolar protein (NOP66)	2	0
Tb927.6.1250	57 kDa	hypothetical protein, conserved	2	0
Tb927.11.3300	108 kDa	hypothetical protein, conserved	2	0
Tb927.5.1900	77 kDa	hypothetical protein, conserved	2	0
Tb927.7.3820	72 kDa	hypothetical protein, conserved	1	0
Tb927.7.660	53 kDa	hypothetical protein, conserved	1	0
Tb927.8.5850	467 kDa	hypothetical protein, conserved	1	0
Tb927.5.2950	88 kDa	hypothetical protein, conserved	1	0
Tb927.10.16440	29 kDa	variant surface glycoprotein, fragment	1	0
Tb927.1.2470	15 kDa	histone H3, putative	1	0
Tb927.3.5030	80 kDa	KU70 protein (KU70)	1	0
Tb927.4.2080	105 kDa	C2 domain containing protein (CC2D)	1	0
Tb927.11.11840	14 kDa	hypothetical protein, conserved	1	0
Tb927.8.3680	70 kDa	hypothetical protein, conserved	1	0
Tb927.9.10680	17 kDa	hypothetical protein, conserved	1	0
Tb927.7.5530	42 kDa	hypothetical protein, conserved	1	0
Tb927.8.850	85 kDa	hypothetical protein, conserved	1	0
Tb927.3.1820	25 kDa	mitochondrial RNA binding complex 1 subunit (MRB1820)	1	0
Tb927.10.15660	86 kDa	hypothetical protein, conserved	1	0
Tb927.7.940	54 kDa	protein kinase C substrate protein, putative,glucosidase II beta subunit	1	0
Tb927.9.4450	78 kDa	hypothetical protein	1	0
Tb927.7.4940	49 kDa	variant surface glycoprotein (VSG, pseudogene), putative	1	0
-DECOY	?	Tb927.7.4940-DECOY oligopeptidase...	1	0
Tb927.4.2560	74 kDa	cardiolipin synthetase, putative	1	0
Tb927.10.1270	127 kDa	mismatch repair protein, putative	1	0
Tb927.10.9360	57 kDa	hypothetical protein, conserved	1	0
Tb927.11.4510	197 kDa	hypothetical protein, conserved	1	0
Tb927.8.1770	59 kDa	hypothetical protein, conserved	1	0
Tb927.8.5110	127 kDa	hypothetical protein, conserved	1	0
Tb927.9.13060	58 kDa	hypothetical protein, conserved	1	0
Tb927.10.4080	127 kDa	hypothetical protein, conserved	1	0
Tb927.7.4650	121 kDa	SNF2 DNA repair protein, putative	1	0
Tb927.5.1290	52 kDa	hypothetical protein, conserved	1	0

Table S1B

Accession Number	MW	Identified Proteins (80/797)	HA	Control
Tb927.8.7590	140 kDa	receptor-type adenylate cyclase GRESAG 4	2	0
Tb927.2.2380	97 kDa	glycosyltransferase (GlcNAc), putative	1	0
Tb927.10.11780	131 kDa	hypothetical protein, conserved	1	0
Tb11.v5.0784	139 kDa	receptor-type adenylate cyclase GRESAG 4, putative	1	0
Tb927.9.15980	43 kDa	nucleoside transporter 1, putative	2	0
Tb927.1.480	161 kDa	leucine-rich repeat protein (LRRP)	1	0
Tb927.5.1450	134 kDa	ATP pyrophosphate-lyase, putative; adenylyl cyclase	1	0
Tb927.7.3860	27 kDa	PQ loop repeat, putative	1	0
Tb927.4.470	22 kDa	snoRNP protein gar1	1	0
Tb927.11.2510	146 kDa	NUC173 domain containing protein	1	0
Tb927.11.5020	187 kDa	hypothetical protein	1	0
gi|261327151	25 kDa	hypothetical protein	1	0
Tb927.6.3520	74 kDa	Uncharacterised ACR, YagE family COG1723	1	0
Tb927.7.1010	20 kDa	hypothetical protein, conserved	1	0
Tb927.11.4380	72 kDa	ATP-dependent DEAD/H RNA helicase, putative	1	0
Tb927.6.4430	36 kDa	homoserine kinase	1	0
Tb427tmp.01.2430	486 kDa	hypothetical protein, conserved	1	0
Tb927.5.2520	104 kDa	TPR repeat, putative	1	0
Tb927.9.10670	67 kDa	bardet-biedl syndrome 1 protein	1	0
Tb927.10.10770	67 kDa	Generative cell specific 1 protein, putative	1	0
Tb927.11.3130	71 kDa	glycosomal transporter (GAT2)	1	0
Tb927.7.530	171 kDa	FYVE zinc finger, putative	1	0
Tb927.3.2140	105 kDa	transcription activator, putative	1	0
Tb927.11.9890	63 kDa	signal recognition particle receptor alpha subunit, putative	1	0
TM35_000231200	10 kDa	hypothetical protein	1	0
Tb927.1.4520	83 kDa	Zinc finger, C3HC4 type (RING finger), putative	1	0
Tb927.3.2790	125 kDa	leucine-rich repeat protein (LRRP)	1	0
Tb927.5.520	56 kDa	stomatin-like protein, putative	1	0
Tb927.9.210	53 kDa	variant surface glycoprotein (VSG, pseudogene), putative	1	0
Tb927.10.13490	31 kDa	CAMK/CAMKL family protein kinase, putative	1	0
Tb927.5.3830	34 kDa	dihydroorotate dehydrogenase	1	0
Tb927.11.14780	46 kDa	phosphomannose isomerase	1	0
Tb927.2.2400	85 kDa	glycosyltransferase (GlcNAc)	1	0
Tb927.10.11020	105 kDa	DNA mismatch repair protein MSH2	1	0
Tb927.2.2650	368 kDa	hypothetical protein, conserved	1	0
Tb927.8.4350	114 kDa	hypothetical protein, conserved	1	0
Tb927.11.10190	133 kDa	telomerase reverse transcriptase, putative	1	0
Tb927.11.16340	125 kDa	Importin-beta N-terminal domain/HEAT-like repeat	1	0
Tb10.v4.0134	54 kDa	variant surface glycoprotein (VSG), putative	1	0
Tb927.2.3280	48 kDa	65 kDa invariant surface glycoprotein	1	0
Tb927.10.13900	39 kDa	UDP-galactose transporter, putative	**1**	0
Tb927.11.3550	32 kDa	XPA-interacting protein, putative	1	0
Tb927.10.9890	97 kDa	hypothetical protein, conserved	1	0
Tb927.10.1610	16 kDa	hypothetical protein, conserved	1	0
Tb927.6.430	138 kDa	receptor-type adenylate cyclase GRESAG 4, putative	1	0
Tb927.11.5740	109 kDa	formin, putative; formin-like protein	1	0
Tb927.10.13090	66 kDa	PhoD-like phosphatase, putative	1	0
Tb927.7.4130	256 kDa	hypothetical protein, conserved	1	0
Tb927.11.5070	93 kDa	GPI ethanolamine phosphate transferase 3, putative	1	0
Tb927.10.2880	304 kDa	Voltage-dependent calcium channel subunit, putative	1	0
Tb927.8.3940	63 kDa	hypothetical protein, conserved	1	0
Tb927.2.2920	49 kDa	UAA transporter family, putative	1	0
Tb927.1.4540	58 kDa	hypothetical protein, conserved	1	0
Tb927.3.1350	87 kDa	hypothetical protein, conserved	1	0
Tb927.3.1860	83 kDa	ATP-grasp domain containing protein	1	0
Tb927.6.1580	53 kDa	polynucleotide kinase 3'-phosphatase, putative	1	0
Tb927.6.4100	94 kDa	hypothetical protein, conserved	1	0
Tb927.7.1470	12 kDa	Mitochondrial ATP synthase subunit c-3	1	0
Tb927.7.2390	103 kDa	Tripartite attachment complex protein 102	1	0
Tb927.11.18060	28 kDa	variant surface glycoprotein (VSG), (fragment)]	1	0
Tb927.8.3840	80 kDa	hypothetical protein, conserved	1	0
Tb927.9.3860	80 kDa	Dyggve-Melchior-Clausen syndrome protein	1	0
Tb927.9.4850	53 kDa	leucine-rich repeat protein (LRRP), putative	1	0
Tb927.10.7580	111 kDa	hypothetical protein, conserved	1	0
Tb927.11.5020	187 kDa	hypothetical protein, conserved	1	0

Tb927.11.6580	77 kDa	hypothetical protein, conserved	1	0
Tb927.11.6830	109 kDa	Domain of unknown function(DUF2779)	1	0
Tb927.11.6920	173 kDa	hypothetical protein, conserved	1	0
Tb927.11.9750	73 kDa	Protein of unknown function (DUF498/DUF598)	1	0
Tb927.11.16210	81 kDa	cAMP response protein 1, putative	1	0
Tb927.10.15080	180 kDa	WD40 Repeat 1	1	0
Tb927.5.2930	43 kDa	Mitochondrial ATP synthase subunit	1	0
Tb927.6.1910	57 kDa	hypothetical protein	1	0
Tb927.7.1510	69 kDa	pseudouridylate synthase I	1	0
Tb927.7.5440	217 kDa	hypothetical protein	1	0
Tb927.8.4400	76 kDa	hypothetical protein	1	0
Tb927.7.4400	83 kDa	inositol polyphosphate kinase-like protein, putative	1	0
Tb927.5.2750	48 kDa	Alpha/beta hydrolase family, putative	1	0
Tb927.11.14250	59 kDa	t-complex protein 1 subunit epsilon	1	0
Tb927.11.7540	34 kDa	electron-transfer-flavoprotein alpha polypeptide	1	0

Table S1C

Accession Number	MW	Identified Proteins (66/243)	HA	Control
Tb927.8.5580	526 kDa	hypothetical protein, conserved	3	0
Tb927.4.4130	100 kDa	hypothetical protein, conserved	3	0
Tb927.10.4170	292 kDa	hypothetical protein, conserved	2	0
Tb927.10.8750	39 kDa	GTPase activating protein, putative	2	0
Tb927.11.11900	98 kDa	coatomer gamma subunit, putative	2	0
Tb927.3.4030	135 kDa	hypothetical protein, conserved	2	0
Tb927.10.13900	39 kDa	UDP-galactose transporter	2	0
Tb927.3.4020	256 kDa	phosphatidylinositol 4-kinase alpha, putative	2	0
Tb927.4.840	46 kDa	hypothetical protein, conserved	2	0
Tb927.2.3370	44 kDa	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative	2	0
Tb927.8.1130	158 kDa	protein phosphatase with EF-Hand domains, putative	2	0
Tb927.2.3320	48 kDa	65 kDa invariant surface glycoprotein	2	0
Tb927.2.4950	156 kDa	hypothetical protein, conserved	2	0
Tb927.4.3870	135 kDa	receptor-type adenylate cyclase GRESAG 4, putative	2	0
Tb927.3.2140	105 kDa	transcription activator	2	0
Tb927.10.12890	91 kDa	bifunctional aminoacyl-tRNA synthetase, putative	2	0
Tb927.11.4380	248 kDa	ATP-dependent RNA helicase, putative	2	0
Tb09.v4.0017 (+1)	31 kDa	rab1 (small gtp-binding protein rab1, putative)	2	0
Tb927.11.10670	134 kDa	glycosyl hydrolase, putative	2	0
Tb927.11.5060	120 kDa	hypothetical protein, conserved	2	0
Tb927.8.2450	34 kDa	hypothetical protein, conserved	2	0
Tb927.11.15470	71 kDa	methionyl-tRNA formyltransferase	2	0
Tb927.5.2420	103 kDa	hypothetical protein, conserved	2	0
Tb927.8.3190	33 kDa	hypothetical protein, conserved	2	0
Tb927.8.4500	85 kDa	eukaryotic translation initation factor 4 gamma, putative	2	0
Tb09.v4.0016	54 kDa	expression site- associated gene (ESAG) protein, putative	2	0
Tb927.3.4960	182 kDa	kinesin, putative	2	0
Tb927.10.6050	191 kDa	clathrin heavy chain (CHC)	2	0
Tb927.10.7350	198 kDa	hypothetical protein, conserved	2	0
Tb927.9.8040	98 kDa	hypothetical protein, conserved	2	0
Tb927.11.18230	50 kDa	variant surface glycoprotein (VSG), putative	2	0
Tb927.8.6970	74 kDa	3-methylcrotonyl-CoA carboxylase alpha subunit, putative	1	0
Tb927.6.3150	501 kDa	Hydin	1	0
Tb927.10.13280	286 kDa	hypothetical protein, conserved	1	0
Tb927.11.1160	123 kDa	hypothetical protein, conserved	1	0
Tb927.7.1570	156 kDa	hypothetical protein, conserved	1	0
Tb927.11.9890	64 kDa	signal recognition particle receptor alpha subunit, putative	1	0
Tb927.8.3940	61 kDa	hypothetical protein, conserved	1	0
Tb927.10.9490	45 kDa	expression site-associated gene 3 (ESAG3)-like protein	1	0
Tb927.5.1560	84 kDa	ATP-dependent DEAD/H RNA helicase, putative	1	0
Tb927.10.1890	178 kDa	calpain, putative,cysteine peptidase, Clan CA, family C2, putative	1	0
Tb927.2.2400	85 kDa	hypothetical protein, conserved	1	0
Tb927.4.5530	51 kDa	variant surface glycoprotein (VSG), putative	1	0
Tb927.11.7290	162 kDa	pantothenate kinase subunit, putative	1	0
Tb927.10.6300	50 kDa	hypothetical protein, conserved	1	0
Tb927.7.1780	26 kDa	Adenine phosphoribosyltransferase, putative	1	0
Tb927.9.16250	53 kDa	variant surface glycoprotein (VSG, atypical), putative	1	0
Tb927.5.2660	98 kDa	hypothetical protein, conserved	1	0
Tb927.3.3150	87 kDa	hypothetical protein, conserved	1	0
Tb927.4.1060	124 kDa	hypothetical protein, conserved	1	0
Tb927.6.4430	36 kDa	homoserine kinase, putative (HK)	1	0

Tb927.9.10530	119 kDa	hypothetical protein, conserved	1	0
Tb927.11.840	87 kDa	hypothetical protein, conserved	1	0
Tb927.8.8140	68 kDa	small GTP-binding rab protein, putative	1	0
Tb927.7.3330	503 kDa	hypothetical protein, conserved	1	0
Tb927.7.3770	61 kDa	hypothetical protein, conserved	1	0
Tb927.9.11410	15 kDa	unspecified product	1	0
Tb927.10.3970	55 kDa	hypothetical protein, conserved	1	0
Tb927.11.12630	57 kDa	hypothetical protein, conserved	1	0
Tb927.10.11420	81 kDa	hypothetical protein, conserved	1	0
Tb927.8.2890	46 kDa	hypothetical protein, conserved	1	0
Tb927.5.1320	111 kDa	hypothetical protein, conserved	1	0
Tb927.5.2920	52 kDa	hypothetical protein, conserved	1	0
Tb10.v4.0042	25 kDa	hypothetical protein, conserved	1	0
Tb927.10.1610	16 kDa	hypothetical protein, conserved	1	0
Tb927.3.2210	69 kDa	hypothetical protein, conserved	1	0

Table S2. Comparative identification of specific lysine acceptor sites in BF and PF developmental stages.

Gene ID	Description	SUMOylated lysine BF	SUMOylated lysine PF
Tb927.7.2830	Histone H2A	K(5)	K(5); K (9); K(10); K(129)
Tb927.10.15350	H3V (h3vaR)	K(31)	K(32)
Tb927.4.1330	DNA topoisomerase IB, large subunit	K(20); K(47)	K(13); K(20); K(47); K(75)
Tb927.1.1170	DNA-directed RNA polymerase subunit RBP12 (RBP12)	K(7)	K(7)
Tb927.9.5190	Proliferative cell nuclear antigen (PCNA)	K(195)	K(201)
Tb927.10.740	Structural maintenance of chromosome 4, putative (SMC4)	K(623); K(1302)	K(1302)
Tb927.10.13720	RNA-binding protein, putative (RBP29)	K(313)	K(313)
Tb927.10.14520	Basal body protein	K(371)	K(371)
Tb927.3.2350	Hypothetical protein	K(35)	K(35)
Tb927.11.5230	Hypothetical protein	K(24)	K(24)
Tb927.9.13320	Hypothetical protein	K(484); K(498)	K(498)
Tb927.3.4140	Hypothetical protein	K(455)	K(455)
Tb927.9.1410	Hypothetical protein	K(228)	K(271)
Tb927.11.11840	Hypothetical protein	K(69)	K(69)
Tb927.3.5370	Hypothetical protein	K(11)	K(132)

11. References

11. References

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

Albuquerque, C.P., Wang, G., Lee, N.S., Kolodner, R.D., Putnam, C.D., and Zhou, H. (2013). Distinct SUMO ligases cooperate with Esc2 and Slx5 to suppress duplication-mediated genome rearrangements. PLoS genetics *9*, e1003670-e1003670.

Andrés-León, E., Núñez-Torres, R., and Rojas, A.M. (2016). miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. Scientific Reports *6*, 25749.

Andrews, S. (2010). FastQC , A Quality Control tool for High Throughput Sequence Data. In http://wwwbioinformaticsbabrahamacuk/projects/fastqc/.

Appikonda, S., Thakkar, K.N., Shah, P.K., Dent, S.Y.R., Andersen, J.N., and Barton, M.C. (2018). Cross-talk between chromatin acetylation and SUMOylation of tripartite motif–containing protein 24 (TRIM24) impacts cell adhesion. The Journal of Biological Chemistry *293*, 7476-7485.

Bachant, J., Alcasabas, A., Blat, Y., Kleckner, N., and Elledge, S.J. (2002). The SUMO-1 isopeptidase Smt4 is linked to centromeric cohesion through SUMO-1 modification of DNA topoisomerase II. Mol Cell *9*, 1169-1182.

Barquilla, A., Saldivia, M., Diaz, R., Bart, J.M., Vidal, I., Calvo, E., Hall, M.N., and Navarro, M. (2012). Third target of rapamycin complex negatively regulates development of quiescence in Trypanosoma brucei. Proc Natl Acad Sci U S A *109*, 14399-14404.

Bayona, J.C., Nakayasu, E.S., Laverriere, M., Aguilar, C., Sobreira, T.J., Choi, H., Nesvizhskii, A.I., Almeida, I.C., Cazzulo, J.J., and Alvarez, V.E. (2011). SUMOylation pathway in Trypanosoma cruzi: Functional characterization and proteomic analysis of target proteins. Mol Cell Proteomics.

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

Bielska, K., Seliga, J., Wieczorek, E., Kędracka-Krok, S., Niedenthal, R., and Ożyhar, A. (2012). Alternative sumoylation sites in the Drosophila nuclear receptor Usp. The Journal of Steroid Biochemistry and Molecular Biology *132*, 227-238.

Borst, P. (2002). Antigenic variation and allelic exclusion. Cell *109*, 5-8.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

References

Castillo-Acosta, V.M., Ruiz-Pérez, L.M., Van Damme, E.J.M., Balzarini, J., and González-Pacanowska, D. (2015). Exposure of Trypanosoma brucei to an N-acetylglucosamine-Binding Lectin Induces VSG Switching and Glycosylation Defects Resulting in Reduced Infectivity. PLOS Neglected Tropical Diseases *9*, e0003612.

Chappuis, F. (2018). Oral fexinidazole for human African trypanosomiasis. The Lancet *391*, 100-102.

Chen, X., Ding, B., LeJeune, D., Ruggiero, C., and Li, S. (2009). Rpb1 sumoylation in response to UV radiation or transcriptional impairment in yeast. PLoS One *4*, e5267.

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

Colasante, C., Robles, A., Li, C.H., Schwede, A., Benz, C., Voncken, F., Guilbride, D.L., and Clayton, C. (2007). Regulated expression of glycosomal phosphoglycerate kinase in Trypanosoma brucei. Mol Biochem Parasitol *151*, 193-204.

Cross, M., Taylor, M.C., and Borst, P. (1998). Frequent loss of the active site during variant surface glycoprotein expression site switching in vitro in Trypanosoma brucei. Mol Cell Biol *18*, 198-205.

Daniels, J.-P., Gull, K., and Wickstead, B. (2012). The Trypanosomatid-Specific N Terminus of RPA2 Is Required for RNA Polymerase I Assembly, Localization, and Function. Eukaryotic Cell *11*, 662-672.

Daniels, J.P., Gull, K., and Wickstead, B. (2010). Cell biology of the trypanosome genome. Microbiol Mol Biol Rev *74*, 552-569.

De Lange, T., and Borst, P. (1982). Genomic environment of the expression-linked extra copies of genes for surface antigens of Trypanosoma brucei resembles the end of a chromosome. Nature *299*, 451-453.

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

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

Deyrieux, A.F., and Wilson, V.G. (2017). Sumoylation in Development and Differentiation. In SUMO Regulation of Cellular Processes, V.G. Wilson, ed. (Cham: Springer International Publishing), pp. 197-214.

Dhillon, N., Raab, J., Guzzo, J., Szyjka, S.J., Gangadharan, S., Aparicio, O.M., Andrews, B., and Kamakaka, R.T. (2009). DNA polymerase ε, acetylases and remodellers cooperate to form a specialized chromatin structure at a tRNA insulator. The EMBO Journal *28*, 2583-2600.

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

References

Donze, D., Adams, C.R., Rine, J., and Kamakaka, R.T. (1999). The boundaries of the silenced HMR domain in Saccharomyces cerevisiae. Genes & development *13*, 698-708.

Dreesen, O., Li, B., and Cross, G.A.M. (2006). Telomere structure and function in trypanosomes: a proposal. Nature Reviews Microbiology *5*, 70.

Ebersole, T., Kim, J.-H., Samoshkin, A., Kouprina, N., Pavlicek, A., White, R.J., and Larionov, V. (2011). tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. Cell cycle (Georgetown, Tex) *10*, 2779-2791.

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

Enserink, J.M. (2015). Sumo and the cellular stress response. Cell Div *10*, 4-4.

Erben, E.D., Fadda, A., Lueong, S., Hoheisel, J.D., and Clayton, C. (2014). A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. PLoS Pathog *10*, e1004178.

Farnung, L., Vos, S.M., Wigge, C., and Cramer, P. (2017). Nucleosome-Chd1 structure and implications for chromatin remodelling. Nature *550*, 539-542.

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

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

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

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

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

Flaus, A., Martin, D.M., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res *34*, 2887-2905.

Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Mattioli, R.C., and Argaw, D. (2017). Monitoring the elimination of human African trypanosomiasis: Update to 2014. PLOS Neglected Tropical Diseases *11*, e0005585.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clinical epidemiology *6*, 257-275.

Garcı, x, a-Estrada, C., Reguera, R.M., Villa, H., Requena, J.M., x, Müller, S., Pérez-Pertejo, Y., Balaña-Fouce, R.*, et al.* (2003). Identification of a gene in Leishmania infantum encoding a protein that contains a SP-RING/MIZ zinc finger domain. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression *1629*, 44-52.

References

Garcia-Dominguez, M., and Reyes, J.C. (2009). SUMO association with repressor complexes, emerging routes for transcriptional control. Biochim Biophys Acta *1789*, 451-459.

Geiss-Friedlander, R., and Melchior, F. (2007). Concepts in sumoylation: a decade on. Nat Rev Mol Cell Biol *8*, 947-956.

Girdwood, D.W., Tatham, M.H., and Hay, R.T. (2004). SUMO and transcriptional regulation. Semin Cell Dev Biol *15*, 201-210.

Glover, L., Hutchinson, S., Alsford, S., and Horn, D. (2016). VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. Proceedings of the National Academy of Sciences of the United States of America *113*, 7225-7230.

Gong, L., and Yeh, E.T. (2006). Characterization of a family of nucleolar SUMO-specific proteases with preference for SUMO-2 or SUMO-3. J Biol Chem *281*, 15869-15877.

Goodson, M.L., Hong, Y., Rogers, R., Matunis, M.J., Park-Sarge, O.K., and Sarge, K.D. (2001). Sumo-1 modification regulates the DNA binding activity of heat shock transcription factor 2, a promyelocytic leukemia nuclear body associated transcription factor. J Biol Chem *276*, 18513-18518.

Gottesdiener, K., Chung, H.M., Brown, S.D., Lee, M.G., and Van der Ploeg, L.H. (1991). Characterization of VSG gene expression site promoters and promoter-associated DNA rearrangement events. Mol Cell Biol *11*, 2467-2480.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.C., Chung, H.M., Lee, P.T., and Lee, M.G. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. Eukaryot Cell *2*, 542-551.

Gunzl, A., Tschudi, C., Nakaar, V., and Ullu, E. (1995). Accurate transcription of the Trypanosoma brucei U2 small nuclear RNA gene in a homologous extract. J Biol Chem *270*, 17287-17291.

He, S., Limi, S., McGreal, R.S., Xie, Q., Brennan, L.A., Kantorow, W.L., Kokavec, J., Majumdar, R., Hou, H., Edelmann, W.*, et al.* (2016). Chromatin remodeling enzyme Snf2h regulates embryonic lens differentiation and denucleation. Development *143*, 1937-1947.

Hendriks, I.A., and Vertegaal, A.C. (2016). A comprehensive compilation of SUMO proteomics. Nat Rev Mol Cell Biol *17*, 581-595.

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

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

Horn, D., and Cross, G.A. (1997). Position-dependent and promoter-specific regulation of gene expression in Trypanosoma brucei. Embo J *16*, 7422-7431.

Hu, Y., Zhu, N., Wang, X., Yi, Q., Zhu, D., Lai, Y., and Zhao, Y. (2013). Analysis of rice Snf2 family proteins and their potential roles in epigenetic regulation. Plant Physiology and Biochemistry *70*, 33-42.

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

Hunter, T., and Sun, H. (2008). Crosstalk between the SUMO and ubiquitin pathways. Ernst Schering Found Symp Proc, 1-16.

Iribarren, P.A., Berazategui, M.A., Bayona, J.C., Almeida, I.C., Cazzulo, J.J., and Alvarez, V.E. (2015a). Different proteomic strategies to identify genuine Small Ubiquitin-like MOdifier targets and their modification sites in Trypanosoma brucei procyclic forms. Cellular Microbiology *17*, 1413-1422.

Iribarren, P.A., Berazategui, M.A., Cazzulo, J.J., and Alvarez, V.E. (2015b). Biosynthesis of SUMOylated Proteins in Bacteria Using the Trypanosoma brucei Enzymatic System. PLOS ONE *10*, e0134950.

Iribarren, P.A., Di Marzio, L.A., Berazategui, M.A., De Gaudenzi, J.G., and Alvarez, V.E. (2018). SUMO polymeric chains are involved in nuclear foci formation and chromatin organization in Trypanosoma brucei procyclic forms. PloS one *13*, e0193528-e0193528.

Janz, L., and Clayton, C. (1994). The PARP and rRNA promoters of Trypanosoma brucei are composed of dissimilar sequence elements that are functionally interchangeable. Mol Cell Biol *14*, 5804-5811.

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

Jeffers, V., Yang, C., Huang, S., and Sullivan, W.J. (2017). Bromodomains in Protozoan Parasites: Evolution, Function, and Opportunities for Drug Development. Microbiology and Molecular Biology Reviews : MMBR *81*, e00047-00016.

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

Jensen, B.C., Booster, N., Vidadala, R.S.R., Maly, D.J., and Parsons, M. (2016). A novel protein kinase is essential in bloodstream Trypanosoma brucei. International Journal for Parasitology *46*, 479-483.

Jin, Q., Mao, X., Li, B., Guan, S., Yao, F., and Jin, F. (2015). Overexpression of SMARCA5 correlates with cell proliferation and migration in breast cancer. Tumor Biology *36*, 1895-1902.

Johnson, E.S., and Gupta, A.A. (2001). An E3-like factor that promotes SUMO conjugation to the yeast septins. Cell *106*, 735-744.

Johnson, E.S., Schwienhorst, I., Dohmen, R.J., and Blobel, G. (1997). The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. Embo J *16*, 5509-5519.

Kabani, S., Fenn, K., Ross, A., Ivens, A., Smith, T.K., Ghazal, P., and Matthews, K. (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*, 427-427.

Kang, X., Qi, Y., Zuo, Y., Wang, Q., Zou, Y., Schwartz, R.J., Cheng, J., and Yeh, E.T.H. (2010). SUMO-specific protease 2 is essential for suppression of polycomb group protein-mediated gene silencing during embryonic development. Molecular cell *38*, 191-201.

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

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

King-Keller, S., Moore, C.A., Docampo, R., and Moreno, S.N.J. (2015). Ca(2+) Regulation of Trypanosoma brucei Phosphoinositide Phospholipase C. Eukaryotic Cell *14*, 486-494.

Klein, C.A., Droll, D., and Clayton, C. (2013). SUMOylation in Trypanosoma brucei. PeerJ *1*, e180.

Kononenko, A.V., Bansal, R., Lee, N.C., Grimes, B.R., Masumoto, H., Earnshaw, W.C., Larionov, V., and Kouprina, N. (2014). A portable BRCA1-HAC (human artificial chromosome) module for analysis of BRCA1 tumor suppressor function. Nucleic Acids Res *42*.

Kooter, J.M., and Borst, P. (1984). Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucleic Acids Res *12*, 9457-9472.

Kooter, J.M., De Lange, T., and Borst, P. (1984). Discontinuous synthesis of mRNA in trypanosomes. Embo J *3*, 2387-2392.

Kooter, J.M., Winter, A.J., de Oliveira, C., Wagter, R., and Borst, P. (1988). Boundaries of telomere conversion in Trypanosoma brucei. Gene *69*, 1-11.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nature Reviews Cancer *11*, 588.

Lamoliatte, F., McManus, F.P., Maarifi, G., Chelbi-Alix, M.K., and Thibault, P. (2017). Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification. Nat Commun *8*, 14109.

Landeira, D., Bart, J.M., Van Tyne, D., and Navarro, M. (2009). Cohesin regulates VSG monoallelic expression in trypanosomes. J Cell Biol *186*, 243-254.

Landeira, D., and Navarro, M. (2007). Nuclear repositioning of the VSG promoter during developmental silencing in Trypanosoma brucei. J Cell Biol *176*, 133-139.

References

LeBowitz, J.H., Smith, H.Q., Rusche, L., and Beverley, S.M. (1993). Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes Dev *7*, 996-1007.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754-1760.

Liao, S., Wang, T., Fan, K., and Tu, X. (2010). The small ubiquitin-like modifier (SUMO) is essential in cell cycle regulation in Trypanosoma brucei. Exp Cell Res *316*, 704-715.

Liu, X., Li, M., Xia, X., Li, X., and Chen, Z. (2017). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. Nature *544*, 440-445.

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

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

Lustig, Y., Sheiner, L., Vagima, Y., Goldshmidt, H., Das, A., Bellofatto, V., and Michaeli, S. (2007). Spliced-leader RNA silencing: a novel stress-induced mechanism in Trypanosoma brucei. EMBO Rep *8*, 408-413.

Lyst, M.J., and Stancheva, I. (2007). A role for SUMO modification in transcriptional repression and activation. Biochem Soc Trans *35*, 1389-1392.

Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997). A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell *88*, 97-107.

Makhnevych, T., Sydorskyy, Y., Xin, X., Srikumar, T., Vizeacoumar, F.J., Jeram, S.M., Li, Z., Bahr, S., Andrews, B.J., Boone, C., *et al.* (2009). Global map of SUMO function revealed by protein-protein interaction and genetic networks. Mol Cell *33*, 124-135.

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

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 *17*, 3.

Martinez-Calvillo, S., Vizuet-de-Rueda, J.C., Florencio-Martinez, L.E., Manning-Cela, R.G., and Figueroa-Angulo, E.E. (2010). Gene expression in trypanosomatid parasites. J Biomed Biotechnol *2010*, 525241.

Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. (2003). Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. Mol Cell *11*, 1291-1299.

Matthews, K.R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev *8*, 491-501.

References

Matunis, M.J., Coutavas, E., and Blobel, G. (1996). A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. J Cell Biol *135*, 1457-1470.

Matunis, M.J., Zhang, X.D., and Ellis, N.A. (2006). SUMO: the glue that binds. Dev Cell *11*, 596-597.

McAndrew, M., Graham, S., Hartmann, C., and Clayton, C. (1998). Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. Exp Parasitol *90*, 65-76.

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

Miller, M.J., Barrett-Wilt, G.A., Hua, Z., and Vierstra, R.D. (2010). Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis. Proc Natl Acad Sci U S A *107*, 16512-16517.

Monribot-Villanueva, J., Zurita, M., and Vázquez, M. (2016). Developmental transcriptional regulation by SUMOylation, an evolving field. genesis *55*, e23009.

Moretti, N.S., Cestari, I., Anupama, A., Stuart, K., and Schenkman, S. (2018). Comparative Proteomic Analysis of Lysine Acetylation in Trypanosomes. Journal of Proteome Research *17*, 374-385.

Morishita, M., and di Luccio, E. (2011). Cancers and the NSD family of histone lysine methyltransferases. Biochim Biophys Acta *1816*, 158-163.

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

Musselman, C.A., and Kutateladze, T.G. (2011). Handpicking epigenetic marks with PHD fingers. Nucleic Acids Research *39*, 9061-9071.

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

Nakaar, V., Gunzl, A., Ullu, E., and Tschudi, C. (1997). Structure of the Trypanosoma brucei U6 snRNA gene promoter. Mol Biochem Parasitol *88*, 13-23.

Nanavaty, V., Sandhu, R., Jehi, S.E., Pandya, U.M., and Li, B. (2017). Trypanosoma brucei RAP1 maintains telomere and subtelomere integrity by suppressing TERRA and telomeric RNA:DNA hybrids. Nucleic acids research *45*, 5785-5796.

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

References

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

Nathan, D., Ingvarsdottir, K., Sterner, D.E., Bylebyl, G.R., Dokmanovic, M., Dorsey, J.A., Whelan, K.A., Krsmanovic, M., Lane, W.S., Meluh, P.B.*, et al.* (2006). Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. Genes Dev *20*, 966-976.

Navarro, M., and Cross, G.A. (1996). DNA rearrangements associated with multiple consecutive directed antigenic switches in Trypanosoma brucei. Mol Cell Biol *16*, 3615-3625.

Navarro, M., and Cross, G.A. (1998). In situ analysis of a variant surface glycoprotein expression-site promoter region in Trypanosoma brucei. Mol Biochem Parasitol *94*, 53-66.

Navarro, M., Cross, G.A., and Wirtz, E. (1999). Trypanosoma brucei variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. Embo J *18*, 2265-2272.

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

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

Nguyen, T.N., Müller, L.S.M., Park, S.H., Siegel, T.N., and Günzl, A. (2013). 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., Nguyen, B.N., Lee, J.H., Panigrahi, A.K., and Gunzl, A. (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. Eukaryot Cell *11*, 1573-1581.

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

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

Niskanen, E.A., and Palvimo, J.J. (2017). Chromatin SUMOylation in heat stress: To protect, pause and organise? BioEssays *39*, 1600263.

Novatchkova, M., Budhiraja, R., Coupland, G., Eisenhaber, F., and Bachmair, A. (2004). SUMO conjugation in plants. Planta *220*, 1-8.

O'Shea, J.P., Chou, M.F., Quader, S.A., Ryan, J.K., Church, G.M., and Schwartz, D. (2013). pLogo: a probabilistic approach to visualizing sequence motifs. Nature Methods *10*, 1211.

Obado S.O., F.M.C., Chait B.T., Rout M.P. (2016). High-Efficiency Isolation of Nuclear Envelope Protein Complexes from Trypanosomes. . The Nuclear Envelope Methods in Molecular Biology *1411*.

References

Obado, S.O., Bot, C., Echeverry, M.C., Bayona, J.C., Alvarez, V.E., Taylor, M.C., and Kelly, J.M. (2011). Centromere-associated topoisomerase activity in bloodstream form Trypanosoma brucei. Nucleic Acids Res *39*, 1023-1033.

Ohkuni, K., Levy-Myers, R., Warren, J., Au, W.-C., Takahashi, Y., Baker, R.E., and Basrai, M.A. (2018). N-terminal Sumoylation of Centromeric Histone H3 Variant Cse4 Regulates Its Proteolysis To Prevent Mislocalization to Non-centromeric Chromatin. G3: Genes|Genomes|Genetics *8*, 1215.

Opigo, J., and Woodrow, C. (2009). NECT trial: more than a small victory over sleeping sickness. The Lancet *374*, 7-9.

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

Ouyang, J., Valin, A., and Gill, G. (2009). Regulation of transcription factor activity by SUMO modification. Methods Mol Biol *497*, 141-152.

Owerbach, D., McKay, E.M., Yeh, E.T., Gabbay, K.H., and Bohren, K.M. (2005). A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. Biochem Biophys Res Commun *337*, 517-520.

Palenchar, J.B., and Bellofatto, V. (2006). Gene transcription in trypanosomes. Mol Biochem Parasitol *146*, 135-141.

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

Pays, E., and Nolan, D.P. (1998). Expression and function of surface proteins in Trypanosoma brucei. Mol Biochem Parasitol *91*, 3-36.

Pays, E., Vanhamme, L., and Perez-Morga, D. (2004). Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries. Curr Opin Microbiol *7*, 369-374.

Pena, A.C., Pimentel, M.R., Manso, H., Vaz-Drago, R., Pinto-Neves, D., Aresta-Branco, F., Rijo-Ferreira, F., Guegan, F., Pedro Coelho, L., Carmo-Fonseca, M*., et al.* (2014). Trypanosoma brucei histone H1 inhibits RNA polymerase I transcription and is important for parasite fitness in vivo. Molecular Microbiology *93*, 645-663.

Pena, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature *442*, 100-103.

Perez-Morga, D., Amiguet-Vercher, A., Vermijlen, D., and Pays, E. (2001). Organization of telomeres during the cell and life cycles of Trypanosoma brucei. J Eukaryot Microbiol *48*, 221-226.

Pollastri, M.P. (2018). Fexinidazole: A New Drug for African Sleeping Sickness on the Horizon. Trends in Parasitology *34*, 178-179.

<center>*References*</center>

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

Prasad, P., and Ekwall, K. (2013). A snapshot of Snf2 enzymes in fission yeast. Biochem Soc Trans *41*, 1640-1647.

Priotto, G., Kasparian, S., Mutombo, W., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., Baudin, E., Buard, V., Kazadi-Kyanza, S*., et al.* (2009). Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. The Lancet *374*, 56-64.

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

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Raab, J.R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P.A., Haussler, D., and Kamakaka, R.T. (2012). Human tRNA genes function as chromatin insulators. The EMBO journal *31*, 330-350.

Reis, H., Schwebs, M., Dietz, S., Janzen, C.J., and Butter, F. (2018). TelAP1 links telomere complexes with developmental expression site silencing in African trypanosomes. Nucleic Acids Research *46*, 2820-2833.

Reynolds, D., Hofmeister, B.T., Cliffe, L., Alabady, M., Siegel, T.N., Schmitz, R.J., and Sabatini, R. (2016). Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma brucei. PLoS genetics *12*, e1005758-e1005758.

Rico, E., Ivens, A., Glover, L., Horn, D., and Matthews, K.R. (2017). Genome-wide RNAi selection identifies a regulator of transmission stage-enriched gene families and cell-type differentiation in Trypanosoma brucei. PLOS Pathogens *13*, e1006279.

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

Rosas-Acosta, G., Russell, W.K., Deyrieux, A., Russell, D.H., and Wilson, V.G. (2005). A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. Mol Cell Proteomics *4*, 56-72.

Rosonina, E., Akhter, A., Dou, Y., Babu, J., and Sri Theivakadadcham, V.S. (2017). Regulation of transcription factors by sumoylation. Transcription, e1311829.

Rosonina, E., Duncan, S.M., and Manley, J.L. (2010). SUMO functions in constitutive transcription and during activation of inducible genes in yeast. Genes Dev *24*, 1242-1252.

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

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

Saitoh, H., Sparrow, D.B., Shiomi, T., Pu, R.T., Nishimoto, T., Mohun, T.J., and Dasso, M. (1998). Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2. Curr Biol *8*, 121-124.

Saldivia, M., Ceballos-Perez, G., Bart, J.M., and Navarro, M. (2016). The AMPKalpha1 Pathway Positively Regulates the Developmental Transition from Proliferation to Quiescence in Trypanosoma brucei. Cell Rep *17*, 660-670.

Sambrook J, R.D. (2001). Molecular Cloning: A Laboratory Manual, 3rd edn edn (Cold Spring Harbor, NY).

Sampson, D.A., Wang, M., and Matunis, M.J. (2001). The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. J Biol Chem *276*, 21664-21669.

Sanchez, R., and Zhou, M.-M. (2011). The PHD Finger: A Versatile Epigenome Reader. Trends in biochemical sciences *36*, 364-372.

Sasse, R., and Gull, K. (1988). Tubulin post-translational modifications and the construction of microtubular organelles in Trypanosoma brucei. Journal of Cell Science *90*, 577-589.

Schimanski, B., Brandenburg, J., Nguyen, T.N., Caimano, M.J., and Gunzl, A. (2006). A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in Trypanosoma brucei. Nucleic Acids Res *34*, 1676-1684.

Schimanski, B., Nguyen, T.N., and Gunzl, A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Mol Cell Biol *25*, 7303-7313.

Schulz, D., Mugnier, M.R., Paulsen, E.M., Kim, H.S., Chung, C.W., Tough, D.F., Rioja, I., Prinjha, R.K., Papavasiliou, F.N., and Debler, E.W. (2015). Bromodomain Proteins Contribute to Maintenance of Bloodstream Form Stage Identity in the African Trypanosome. PLoS Biol *13*, e1002316.

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

Scott, V., Sherwin, T., and Gull, K. (1997). gamma-tubulin in trypanosomes: molecular characterisation and localisation to multiple and diverse microtubule organising centres. J Cell Sci *110 (Pt 2)*, 157-168.

Sharrocks, A.D. (2006). PIAS proteins and transcriptional regulation--more than just SUMO E3 ligases? Genes Dev *20*, 754-758.

Shiio, Y., and Eisenman, R.N. (2003). Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A *100*, 13225-13230.

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

Siegel, T.N., Kawahara, T., Degrasse, J.A., Janzen, C.J., Horn, D., and Cross, G.A. (2008). Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in Trypanosoma brucei. Mol Microbiol *67*, 762-771.

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

Simarro, P.P., Diarra, A., Ruiz Postigo, J.A., Franco, J.R., and Jannin, J.G. (2011). The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward. PLoS Negl Trop Dis *5*, e1007.

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

Stanne, T.M., and Rudenko, G. (2010). Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. Eukaryot Cell *9*, 136-147.

Stead, K., Aguilar, C., Hartman, T., Drexel, M., Meluh, P., and Guacci, V. (2003). Pds5p regulates the maintenance of sister chromatid cohesion and is sumoylated to promote the dissolution of cohesion. J Cell Biol *163*, 729-741.

Štros, M. (2010). HMGB proteins: Interactions with DNA and chromatin. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1799*, 101-113.

Takahashi, Y., Kahyo, T., Toh, E.A., Yasuda, H., and Kikuchi, Y. (2001). Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates. J Biol Chem *276*, 48973-48977.

Tammsalu, T., Matic, I., Jaffray, E.G., Ibrahim, A.F., Tatham, M.H., and Hay, R.T. (2015). Proteome-wide identification of SUMO modification sites by mass spectrometry. Nat Protoc *10*, 1374-1388.

Taylor, J.E., and Rudenko, G. (2006). Switching trypanosome coats: what's in the wardrobe? Trends Genet *22*, 614-620.

Tozluoglu, M., Karaca, E., Nussinov, R., and Haliloglu, T. (2010). A mechanistic view of the role of E3 in sumoylation. PLoS Comput Biol *6*.

Ulrich, H.D. (2009). The SUMO system: an overview. Methods Mol Biol *497*, 3-16.

Uzureau, P., Daniels, J.-P., Walgraffe, D., Wickstead, B., Pays, E., Gull, K., and Vanhamme, L. (2008). Identification and characterization of two trypanosome TFIIS proteins exhibiting particular domain architectures and differential nuclear localizations. Molecular Microbiology *69*, 1121-1136.

References

Valenciano, A.L., Ramsey, A.C., and Mackey, Z.B. (2015). Deviating the level of proliferating cell nuclear antigen in Trypanosoma brucei elicits distinct mechanisms for inhibiting proliferation and cell cycle progression. Cell Cycle *14*, 674-688.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biology *15*, R82.

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

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

Walgraffe, D., Devaux, S., Lecordier, L., Dierick, J.F., Dieu, M., Van den Abbeele, J., Pays, E., and Vanhamme, L. (2005). Characterization of subunits of the RNA polymerase I complex in Trypanosoma brucei. Mol Biochem Parasitol *139*, 249-260.

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

Wang, R., Gao, J., Zhang, J., Zhang, X., Xu, C., Liao, S., and Tu, X. (2019). Solution structure of TbTFIIS2-2 PWWP domain from Trypanosoma brucei and its binding to H4K17me3 and H3K32me3. Biochem J *476*, 421-431.

Wang, Y., Ladunga, I., Miller, A.R., Horken, K.M., Plucinak, T., Weeks, D.P., and Bailey, C.P. (2008). The small ubiquitin-like modifier (SUMO) and SUMO-conjugating system of Chlamydomonas reinhardtii. Genetics *179*, 177-192.

Weaver, M.T., Morrison, A.E., and Musselman, A.C. (2018). Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. Molecules *23*.

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

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M.D., and Owen-Hughes, T. (2003). Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Mol Cell Biol *23*, 1935-1945. WHO (1998). Control and surveillance of African trypanosomiasis. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser *881*, I-VI, 1-114.

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

Wirtz, E., Hartmann, C., and Clayton, C. (1994). Gene expression mediated by bacteriophage T3 and T7 RNA polymerases in transgenic trypanosomes. Nucleic Acids Res *22*, 3887-3894.

References

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.M. (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.

Wotton, D., Pemberton, L.F., and Merrill-Schools, J. (2017). SUMO and Chromatin Remodeling. In SUMO Regulation of Cellular Processes, V.G. Wilson, ed. (Cham: Springer International Publishing), pp. 35-50.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86.

Xiao, Y., Pollack, D., Andrusier, M., Levy, A., Callaway, M., Nieves, E., Reddi, P., and Vigodner, M. (2016). Identification of cell-specific targets of sumoylation during mouse spermatogenesis. Reproduction (Cambridge, England) *151*, 149-166.

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

Yang, X., Wu, X., Zhang, J., Zhang, X., Xu, C., Liao, S., and Tu, X. (2017). Recognition of hyperacetylated N-terminus of H2AZ by TbBDF2 from Trypanosoma brucei. Biochemical Journal *474*, 3817.

Ye, K., Zhang, X., Ni, J., Liao, S., and Tu, X. (2015). Identification of enzymes involved in SUMOylation in Trypanosoma brucei. Scientific Reports *5*, 10097.

Yoo, H.S., Seo, J.H., and Yoo, J.Y. (2013). CTR9, a component of PAF complex, controls elongation block at the c-Fos locus via signal-dependent regulation of chromatin-bound NELF dissociation. PLoS One *8*, e61055.

Yuxuan, X., Daniel, P., Miriam, A., Avi, L., Myrasol, C., Edward, N., Prabhakara, R., and Margarita, V. (2016). Identification of cell-specific targets of sumoylation during mouse spermatogenesis. REPRODUCTION *151*, 149-166.

Zhang, J., Lai, J., Wang, F., Yang, S., He, Z., Jiang, J., Li, Q., Wu, Q., Liu, Y., Yu, M.*, et al.* (2017). A SUMO Ligase AtMMS21 Regulates the Stability of the Chromatin Remodeler BRAHMA in Root Development. Plant Physiology *173*, 1574.

Zhang, X.D., Goeres, J., Zhang, H., Yen, T.J., Porter, A.C., and Matunis, M.J. (2008). SUMO-2/3 modification and binding regulate the association of CENP-E with kinetochores and progression through mitosis. Mol Cell *29*, 729-741.

Zhao, X. (2018). SUMO-Mediated Regulation of Nuclear Functions and Signaling Processes. Mol Cell *71*, 409-418.

Zimmermann, H., Subota, I., Batram, C., Kramer, S., Janzen, C.J., Jones, N.G., and Engstler, M. (2017). A quorum sensing-independent path to stumpy development in Trypanosoma brucei. PLoS Pathog *13*, e1006324.

Universidad de Granada

Instituto de Parasitología y
Biomedicina "López-Neyra" CSIC