Unexpected diversity in eukaryotic transcription revealed by the retrotransposon hotspot family of *Trypanosoma brucei*

Francesca Florini[1,2], Arunasalam Naguleswaran[1], Walid H. Gharib[3], Frédéric Bringaud[4] and Isabel Roditi[1,*]

[1]Institute of Cell Biology, University of Bern, Bern, Switzerland, [2]Graduate School of Cellular and Biomedical Science, University of Bern, Bern, Switzerland, [3]Interfaculty Bioinformatics Unit, University of Bern, Switzerland and [4]Laboratoire de Microbiologie Fondamentale et Pathogénicité (MFP), UMR 5234 CNRS, Université de Bordeaux, France

ABSTRACT

The path from DNA to RNA to protein in eukaryotes is guided by a series of factors linking transcription, mRNA export and translation. Many of these are conserved from yeast to humans. Trypanosomatids, which diverged early in the eukaryotic lineage, exhibit unusual features such as polycistronic transcription and trans-splicing of all messenger RNAs. They possess basal transcription factors, but lack recognisable orthologues of many factors required for transcription elongation and mRNA export. We show that retrotransposon hotspot (RHS) proteins fulfil some of these functions and that their depletion globally impairs nascent RNA synthesis by RNA polymerase II. Three sub-families are part of a coordinated process in which RHS6 is most closely associated with chromatin, RHS4 is part of the Pol II complex and RHS2 connects transcription with the translation machinery. In summary, our results show that the components of eukaryotic transcription are far from being universal, and reveal unsuspected plasticity in the course of evolution.

INTRODUCTION

Trypanosoma brucei spp are unicellular eukaryotes that are pathogenic for humans and domestic animals. They are members of the family of *Trypanosomatidae*, which includes other parasites such as *T. cruzi*, the causative agent of Chagas' Disease, and various species of *Leishmania*. Most of these parasites have a life cycle that is split between a mammalian host and an insect host which, in the case of *Trypanosoma brucei brucei*, is the tsetse fly. In the mammal, bloodstream form trypanosomes evade the host immune system by frequent changes of their variant surface glycoprotein (VSG) coat ([1](#)). This is shed when the parasites are ingested by a tsetse fly and differentiate to procyclic forms in the fly midgut. Trypanosomatids diverged early from other eukaryotes and have several unusual characteristics, such as a highly condensed mitochondrial genome (the kinetoplast), RNA editing via guide RNAs, and universal trans-splicing of nuclear-encoded mRNAs ([2](#)). Moreover, African trypanosomes harbour the only RNA polymerase I (Pol I) known to transcribe mRNAs as well as ribosomal RNAs. The most prominent of these are the mRNAs coding for VSG in bloodstream forms and the procyclin coat of the tsetse midgut forms ([3](#)). Transcription of these genes by Pol I ensures higher expression, since Pol I is known to initiate transcription ten times faster than RNA polymerase II (Pol II) ([4](#)).

Transcription by Pol II is also unusual, sharing some features with prokaryotes and others with eukaryotes. In archaea and bacteria, several genes, which are normally functionally related to each other, are transcribed as a long transcript starting from a single promoter and translated directly from this polycistronic transcript ([5](#)). At the other extreme, the majority of eukaryotic genes are monocistronic, with each having its own promoter. There are exceptions to this: some nematodes, including *C. elegans,* are known to have a portion of their genomes organized in operons ([6–8](#)), and some genes in *Drosophila* and other organisms are transcribed as dicistrons ([9](#)). Trypanosomatids, though, are the only eukaryotes in which all genes transcribed by Pol II are organized in long polycistronic units containing up to 100 genes ([10,11](#)). The polycistronic RNAs are co-transcriptionally processed, giving rise to monocistronic mRNAs that are capped with a 39-nt spliced leader (SL) and polyadenylated. Trans-splicing of the SL and polyadenylation are both guided by pyrimidine-rich tracts followed by an AG dinucleotide ([12–16](#)) which resemble the

*To whom correspondence should be addressed. Tel: +41 31 631 4647; Fax: +4131 631 4684; Email: isabel.roditi@izb.unibe.ch

signals for cis-splicing in other eukaryotes. There are approximately 100 tandem copies of the SL gene (17). It is synthesized as a precursor, and is the only Pol II-transcribed sequence with a promoter in front of each copy (18,19). For protein-coding gene arrays, transcription seems to start over broad regions of around 10 kb (20,21). Genome-wide studies demonstrated that the boundaries of different polycistronic units in *T. b. brucei* are marked by distinct histone variants and different histone modifications. The histone variants H2AZ and H2BV and the modifications H3K4me3 and H3K10ac are enriched at transcription start sites, while the histone variants H3V and H4V are more abundant at transcription termination sites (22–25). Trypanosomatids also contain a modified nucleotide, base J (26), that is enriched at the end of polycistronic transcription units and promotes termination by Pol II (27).

The divergence of trypanosomatids from other eukaryotes explains the difficulties encountered in identifying the canonical transcription factors that are normally present in eukaryotes. Until recently, it was thought that these were completely absent, but it is now accepted that most of the basal transcription factors are present, although not particularly well conserved (17,28–30). Trypanosomatids also lack recognisable orthologues of a number of factors required for transcription elongation (31,32) and mRNA export (33). In yeast and humans, transcription and mRNA export are connected via the TREX complex (34), but, in trypanosomatids, no components of this complex have been identified apart from TbSub2 (33,35). It was recently shown, however, that mRNA export occurs cotranscriptionally (36). Another peculiarity of trypanosomatid Pol II is that the major subunit, RPB1, lacks the characteristic heptapeptide repeats found in the C-terminal domain (CTD) in other eukaryotes (37). Differential phosphorylation of these repeats is required for the CTD to recruit different factors during the transcriptional cycle (38–41). Despite lacking the heptapeptide repeats, it has recently been demonstrated that C-terminus of *T. brucei* Pol II is essential for transcription (37) and that it contains several phosphorylation sites (42), some of which are required for its activity (43).

In an attempt to identify trypanosome-specific transcription factors associated with Pol II, two groups performed tandem affinity purification with different subunits as bait. In both cases, members of the retrotransposon hotspot 4 (RHS) family were identified in the core complex (44,45). More recently, it was demonstrated that TbRRM1, a nucleoprotein that modulates chromatin structure in *T. b. brucei* interacted with several RHS (46). RHS were first identified in 2002 as a multigene family clustered at subtelomeres (47). Their name derives from a retrotransposon insertion site in the 5′ portion of the coding region, which is predicted to disrupt the function of ∼60% of the genes. Bringaud *et al.* identified 52 RHS (pseudo)genes in *T. b. brucei* (TREU927) and classified them into 6 sub-families, RHS1 to RHS6, based on the sequence similarities of their C-termini. The only recognizable feature of RHS proteins is an ATP/GTP binding domain. Following the completion of the *T. b. brucei* genome sequence, more RHS genes have been annotated—118 at present—and a seventh sub-family (RHS7) was added. RHS proteins are expressed in blood-

stream and procyclic forms. Five sub-families were localized to the nucleus, while RHS2 showed a perinuclear localization (47). Based on their known interactions with Pol II and a chromatin-associated protein, we hypothesized that RHS proteins could be trypanosome-specific factors playing a role in transcription. In the present work, we analyse the RHS2, RHS4 and RHS6 sub-families. We show that all three sub-families associate with regions of the genome that are transcribed by Pol II. Furthermore, depletion of RHS reduces nascent RNA synthesis and leads to accumulation of RNA in the nucleus, suggesting that they play a role in Pol II transcription elongation and mRNA export.

MATERIALS AND METHODS

Cell culture

Procyclic forms of *T. brucei brucei* EATRO 1125, initially derived from AnTat 1.1 bloodstream forms (48), and the derivative for inducible RNAi, AnTat1.1 90-13 (49), were used for the experiments performed in this study. Procyclic forms were cultured in SDM-79 (50) containing 10% foetal bovine serum (FBS) and 20 mM glycerol at 27°C. Parasites were maintained at cell densities between 10^6 and 10^7 cells ml^{-1}.

Stable transfections of the parasites were performed by electroporation with 10 μg linearized plasmids as previously described (51). Selection of clones was achieved by limiting dilution in medium supplemented with antibiotics (phleomycin at 1.5 μg ml^{-1} for RNAi constructs; puromycin at 1 μg ml^{-1} for HA-tagging constructs). For induction of RNAi, 1 μg ml^{-1} tetracycline was added to the culture medium.

Primers and constructs

Synthetic restriction sites are underlined.

- RHS2RNAiFW: GGGGGATCCAATTGAGAGTGTGGGCGAAC
- RHS2RNAiRV: GGGCTCGAGATTACCACTTGCCAGAACGG
- RHS4RNAiFW: GGGGGATCCCAGTTGAGACTCATTGGGCA
- RHS4RNAiRV: GGGCTCGAGCCTCACCTCCAGCTCTATCG
- RHS6RNAiFW: GGATCCCTGATGCAATTGCTGAGGAT
- RHS6RNAiRV: CTCGAGTATTCGCCACTTCTCTTGCC
- RHS6-HA-ApaI: TAGGGCCCAGTGTCTTGAGGCATGTAGAGG
- RHS6-HA-NotI: TAGCGGCCGCAATTCGTTATTCGCCACTTCTCTTGC
- RHS4-HA-Not: TAGCGGCCGCAACGCGTCTTCACCTCCTTCCAC
- RHS4-HA-Apa: TAGGGCCCTCGAGTTTTTAAAGTTATCG
- RHS2-HA-Apa: TAGGGCCCAGGACGGGACTACTTACATG
- RHS2-HA-Not: TAGCGGCCGCAAGGCAGCGGGGCCACCAGCAATAG

- SL oligo (FISH): Cy5-CAATATAGTACAGAAACTGT
- Oligo d(T) probe (FISH): Cy3-TTTTTTTTTTTTTTTTTTTTT
- SL probe (Northern blot): CAATATAGTACAGAAACTGTTCTAATAATAGCGTT (52)
- 18S probe (Northern blot): GTTCGTCTTGGTGCGGTCTA
- qPCR oligos:
- Procycl_Pro_67s: GAGCTTAATGTCCTTTTCA
- Procycl_Pro_191as: CAAAACAACCATATCACTTC
- QChIP RLP10 F: AAGAGCATGCCAGCAAATC
- QChIP RLP10 R: GGGTCAGCGATATACTCCGT
- QChIP alpha tub F: CCGTCACGTGTAAGATGAGC
- QChIP alpha tub R: ACTCCACACAGCGGAAGAG
- 18S_up_377s: CCATGCTCTCTCGTGTGTGTA (53)
- 18S_up_265as: TTCCTCAAGGCGTCACTCTATC (53)

All plasmid inserts were generated by PCR using genomic DNA from *T. b. brucei* AnTat 1.1 as a template. RNAi constructs were generated by cloning the PCR insert into the stem loop vector pSL-Comp1 (54). HA-tagging constructs were derivatives of the cassette-type construct pC-PTP-NEO (55).

Northern and western blot analysis

Total RNA isolation with guanidine thiocyanate and Northern blot analysis were performed according to standard procedures (56,57). 10 μg of RNA were loaded in each lane. Radioactively labeled probes were prepared using a Megaprime DNA labelling kit (Amersham Biosciences) according to manufacturer's instructions. A 5′-labeled antisense oligonucleotide probe for the 18S RNA was used as loading control (58). Blots were hybridized at 65° and washed in 0.2× SSC/0.1% SDS for PCR-based probes, or in 1× SSC/0.1% SDS for 18S. For the SL probe, hybridization was performed at 37° and washed in 4× SSC/0.1% SDS at RT. The blots were exposed to Phosphorimager screens and scanned with Typhoon FLA 7000 (GE Healthcare Life Sciences). All signals were normalized in Fiji.

For Western blot analysis protein samples (2×10^6 cell equivalents per lane) were separated on 12% SDS-polyacrylamide gels and transferred to Immobilon-P membrane (Millipore). Membranes were incubated overnight with primary antibodies: anti-RHS2, anti-RHS4 and anti-RHS6 (Rat, 1:10 000) (47), anti-HA 3F10 (Rat, 1:2500, Sigma-Aldrich), anti-Alba3 (Rabbit, 1:250) (54). Secondary antibodies were used 1:5000: swine anti-rabbit HRP (Dako Denmark), donkey anti-rat HRP (Invitrogen). The blots were incubated with Super Signal West Pico PLUS Chemiluminescent substrate (Thermo Scientific) and detected with an Amersham Imager 600.

Immunofluorescence and fluorescence *in situ* hybridization (FISH)

For immunofluorescence, cells were harvested by centrifugation, washed once in PBS, fixed in 4% paraformaldehyde in PBS and permeabilized with Triton X-100 (0.1% in PBS). After blocking in PBS/3% BSA, cells were incubated with primary antibody (1:1000 for anti-RHS antibodies, 1:250 for anti-HA) for 1 h at room temperature, washed 3 times in PBS and subsequently incubated with secondary antibody (1:2000, Alexa Fluor 488 donkey anti-rat, ThermoFisher) for 1 h at room temperature. Cells were then washed 3 times in PBS and stained with Hoechst 33342 prior to embedding in Mowiol (Sigma Aldrich).

RNA FISH was performed according to Cassola and colleagues (59): cells were permeabilized and simultaneously blocked for 30 min in PBS/0.5% saponin/2% BSA, and prehybridized for 1 h in hybridization solution (2% BSA, 5x Denhardt′s solution, 4× SSC, 5% dextran sulphate, 35% formamide, 0.5 μg μl⁻¹ tRNA). Cy3-labelled oligo-d(T)$_{20}$ and Cy5-labelled SL-oligo (2 ng μl⁻¹ in hybridization solution) were hybridized overnight at room temperature. Cells were then washed with decreasing concentrations of SSC (from 4× to 1×), and nuclei were stained with Hoechst 33342.

To block transcription, cells were treated with 10 μg ml⁻¹ Actinomycin D (Sigma Aldrich) for 2 h. In parallel, a control sample was treated with the same volume of DMSO for the same length of time. Images were captured with a Leica DM 5500 B microscope and analysed using LAS AF software (Leica) and Fiji.

Flow cytometry

2×10^6 cells were collected, washed once in cold PBS and resuspended in 300 μl PBS. Cells were then fixed by drop-wise addition of 700 μl of ice-cold EtOH. After an incubation on ice for 1 h, cells were harvested again by centrifugation and resuspended in 500 μl of PBS containing 50 μg ml⁻¹ propidium iodide (Sigma Aldrich), 0.1 mg ml⁻¹ RNase A and 0.05% Triton X-100. Cells were incubated at 37°C for 30 min and 10^4 cells per sample were analysed with a Novo-Cyte flow cytometer (ACEA Biosciences).

Co-immunoprecipitation and mass spectrometry

For isolation of proteins interacting with RHS2-HA or RHS6-HA, 2×10^8 cells were harvested by centrifugation at 2000 rpm for 5 min and washed once in PBS. Cells were then cross-linked by addition of 900 μl 0.1% paraformaldehyde for 8 min, after which the reaction was blocked by addition of 100 μl 1.25 M glycine. After 5 min, cells were pelleted, washed once in PBS and resuspended in lysis buffer (20 mM Tris, 140 mM KCl,1.8 mM MgCl₂, 0.1% NP40, 10% glycerol, supplemented with EDTA-free protease inhibitor cocktail (Roche)). Cells were sonicated 3 times for 10 s with a Branson Digital Sonifier at 10% amplitude, interspersed with 30-second intervals on ice. The sonicate was placed at 4° for 30 min, after which it was incubated overnight with 25 μl Pierce Anti-HA Magnetic Beads (ThermoScientific), previously washed 3 times in 1 ml lysis buffer. For RNase and DNase controls, lysates and beads were supplemented with RNase A or DNase I at a final concentration of 100 μg ml⁻¹ during the overnight incubation. An untagged culture was always used in parallel as a negative control. Protein cross-linking was reversed by incubation of the sample in Laemmli Buffer for 20 min at 95°C. Samples were loaded on an SDS-polyacrylamide gel and run until they

had just entered the 12% resolving gel. Proteins were then gel-extracted, subjected to trypsin digestion and analysed by LC–tandem MS. Protein identification was performed at the Proteomics and Mass Spectrometry facility at University of Bern, CH.

Chromatin immunoprecipitation (ChIP) and RNA immunoprecipitation (RIP)

ChIP experiments were performed as described previously (60,61) with some modifications. 10^8 parasites were washed once in PBS and fixed with 1% paraformaldehyde (w/v) in PBS for 8 min at room temperature; the reaction was stopped by adding glycine to a final concentration of 125 mM for 5 min. Cells were then washed once in PBS and resuspended in lysis buffer (50 mMTris-HCl, pH 8, 10 mM EDTA, 1% SDS, and EDTA-free protease inhibitor cocktail (Roche)). Subsequently, the lysate was sonicated with a Bioruptor (Diagenode) using the following settings: Power 'L', ON/OFF (Sec) 30/90 for 7 min, to produce fragments <1 kb. The soluble chromatin was diluted 1:100 in RIPA buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, 140 mM NaCl) and used for immunoprecipitation. Anti-RHS (47) and anti-RPB1 antibodies immobilized on Dynabeads Sheep Anti-Rat IgG (Invitrogen) were used for immunoprecipitation. For the Pol I control, a monoclonal Anti-RPA1 antibody (62) was immobilized on Dynabeads Pan Mouse IgG (Invitrogen). For RNase treatment, lysates and beads were supplemented with RNase A at a final concentration of 100 μg ml^{-1} during the overnight incubation. DNA was eluted from the beads by digestion with Proteinase K at 68°C for 2 h, purified by phenol-chloroform extraction and ethanol precipitation, and quantified using a Qubit dsDNA HS Assay kit (Invitrogen). For qPCR, 1 ng DNA was added to each reaction; qPCR reactions were performed in triplicate. qPCR was performed using MESA Green qPCR MasterMix Plus for SYBR Assay (Eurogentec) in the ABI Prism 7000 sequence detection system (Applied Biosystems). The data were analyzed using 7000 System SDS software v1.2 (Applied Biosystems). Primers used for qPCR are listed above. Illumina sequencing was performed at Fasteris (Geneva) using the Hiseq system.

RNA IP experiments were performed using cells expressing RHS2-HA and RHS6-HA as described previously (63). For isolation of RNA bound to RHS, anti-HA magnetic beads (Pierce) were used, and the RNA-protein complexes were resuspended in 20 mM Tris–HCl (pH 7.5), 5 mM EDTA, 50mMNaCl, 0.1% SDS, and 50 μg ml^{-1} proteinase K and incubated at 68°C for 40 min. RNA was purified by phenol-chloroform extraction and ethanol precipitation, subjected to DNase treatment, and quantified using a Qubit RNA HS Assay kit (Invitrogen).

5-Ethynyl uridine incorporation and processing for GRO-Seq

Procyclic forms at a density of $7–8 \times 10^6$ ml^{-1} were pulsed for 10 min with 200 μM 5-ethynyl uridine (5-EU; Jena Biosciences). Cells were spun down and nuclei were isolated as described by Murphy *et al.* (64) with some modifications. Briefly, cell pellets were resuspended in ice-cold buffer

(0.5 M sucrose, 50 mM KCl, 5 mM $MgCl_2$, 50 mM Tris–HCl, pH 7.4) and incubated on ice. After 5 min, NP-40 was added to final concentration of 0.1%, the sample was vortexed for 10 s and nuclei were pelleted by centrifugation at 3300g at 4°C for 3 min. Total RNA was isolated from nuclear pellets and stored at –70°C. The amount of RNA was measured using a Qubit RNA HS assay kit (Invitrogen). Copper-dependent click-it chemistry was used to label nascent RNAs that incorporated 5-EU with biotin as described (65), starting with 2 mg total RNA. Biotin-ligated RNA was fragmented to ∼250 bp using a magnesium-dependent RNA fragmentation module (New England Biolabs) and the biotinylated RNA fragments were purified using Dynabeads™ M280 streptavidin (ThermoFisher Scientific) and quantitated using a Qubit RNA HS assay kit. Purified nascent RNAs were used for cDNA library preparation using Illumina Truseq RNA sample preparation kit and sequenced at Fasteris (Geneva) as described above.

NGS sequencing and bioinformatics

Illumina reads were mapped and analysed using the Galaxy suite version 18.05 (66), as described in Naguleswaran *et al.* (67). For Figures 4A and 4B, the Chip-seq reads were mapped against release 8.1 of *T. b. brucei* TREU927 genome downloaded from TritrypDB (http://tritrypdb.org/tritrypdb/). The BWA software (Li H., 2013, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1) with default parameters was launched in paired-end and single-end modes with respect to the replicate sequencing libraries. We discarded alignment matches shorter than 19 bp and allowed gaps shorter than 100 bp. The algorithm uses a scoring scheme (Z-dropoff) which avoids unnecessary extension of poor matching regions and reduces poor alignments in good matching regions by dynamically calculating the alignment score. the full list of the default parameters can be found on the software website (http://bio-bwa.sourceforge.net/bwa.shtml). The output of BWA was sorted and indexed using SAMtools (68). Peak calling was performed using MACS2 software (69) on RHS2, RHS4, RHS6 and Pol II using the replicates for each run. Reads from duplicate controls (DNA extracted from the same chromatin samples used for ChIP) were used as input.

The pairwise Spearman correlations in Figure 4A were computed using the R package GGally (B. Schloerke *et al.*, 2018, GGally: Extension to 'ggplot2'. R package version 1.4.0). To generate the heatmap in Figure 4B, alignment coverage was normalized to Reads Per Kilobase Million (RPKM) using deepTools (70) with the option normalize using RPKM and converted to BigWig format. The normalized alignment as well as the peak enriched intervals (narrow peaks outputs) were used as input files in R Bioconductor package EnrichedHeatmap (Zuguang Gu, 2017, EnrichedHeatmap: Making Enriched Heatmaps. R package) was used to compute the heatmap.

Raw read files are deposited at the European Nucleotide Archives (ENA) http://www.ebi.ac.uk/ena as study PR-JEB29185.
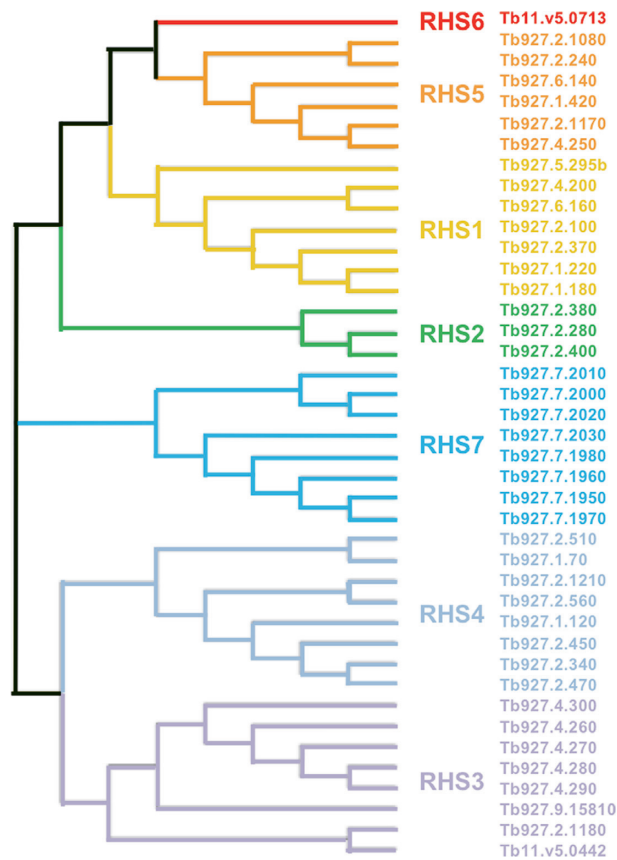
Figure 1. RHS are grouped into seven sub-families neighbour-joining phylogenetic tree, without distance correction, of protein-coding RHS in *T. b. brucei* TREU927. The tree was obtained with the Multiple Sequence Alignment tool Clustal Omega. Gene IDs and annotations are from Tritrypdb (http://tritrypdb.org/tritrypdb/).

RESULTS

RHS are required for growth and viability

Of the 118 genes annotated as RHS genes in the genome of *T. b. brucei* TREU927, 78 are considered to be pseudogenes, either because of the insertion of a retrotransposon or because the open reading frame is truncated. Figure 1 shows the 40 'full length' RHS grouped into seven sub-families based on similarities in their C-termini. RHS are highly expressed by different life-cycle stages (67), with the exception of the RHS7 sub-family which is only expressed in the tsetse salivary glands (AN and IR, unpublished). We chose to focus on three sub-families—RHS2 because of the unusual localization of these proteins compared to other RHS, RHS4 because of their association with Pol II (44,45) and RHS6 because it was annotated as the single member of its sub-family. Bringaud *et al.* reported that RHS2 was perinuclear (47). In our hands, and using the same polyclonal antisera, but a different stock of *T. brucei*, RHS2 was mainly in the cytoplasm (Figure 2A). RHS4 and RHS6 were both nuclear (Figure 2A), as reported previously (47).

Since nothing is known about the function of RHS proteins, we performed RNAi using tetracycline inducible constructs that were targeted against specific regions in the coding region of each sub-family; these were designed to knock down all members of a particular sub-family at the same time (Figure 2B). Depletion of both RHS2 and RHS6 resulted in growth arrest by day 3 post induction. The phenotype upon knockdown of RHS4 was even more extreme, resulting in the onset of cell death by day 4 and complete clearance of the culture by day 6. The efficiency of RNAi, and its impact on the expression of other RHS, was assessed by Northern and Western blot analyses on day 2 post induction (Figure 2B and Supplementary Figure S1), when the cells were still growing normally. Whereas the depletion of RHS2 and RHS4 were both very efficient, residual RHS6 protein was clearly visible. Knockdown of any one of these sub-families did not affect the expression of the others, indicating that there was no crosstalk or compensatory mechanisms (Supplementary Figure S1).

To investigate whether cells that had stopped dividing were stalled at a particular phase of the cell cycle, we took advantage of the fact that these can be distinguished based on the configuration of the nucleus and the kinetoplast DNA (kDNA) in the single mitochondrion. In G1/G0, cells contain one nucleus and one kDNA (1N1K), cells in S phase duplicate and segregate their kDNA ahead of their nuclear DNA (1N2K) and post-mitotic cells have the configuration 2N2K. A small proportion of cells in a culture contains kDNA but no nuclear DNA (0N1K), presumably due to a defect in DNA replication and /or nuclear segregation. These cells are termed zoids. When the three RNAi lines were analysed for cell-cycle defects, both RHS2 and RHS4 knockdowns showed a slight increase in the number of zoids (Figure 3A). This phenotype was much more pronounced in RHS 6 knockdown, however, reaching 25% by day 2. Propidium iodide staining showed an increase in cells with a nuclear content <2*n* (presumably corresponding to zoids), but no increase in cells with a nuclear content ≥4n (Figure 3B). This points to a defect in DNA replication rather than nuclear segregation.

Taken together, our RNAi data show that RHS2, RHS4 and RHS6 are all required for normal cell growth and viability; furthermore, loss of RHS6 causes a defect in DNA replication.

Proteins interacting with RHS2 and RHS6 point to roles in RNA metabolism

As mentioned above, several RHS sub-families coprecipitate with TbRRM1 (46) and RHS4 is part of the core complex of Pol II (44,45). To obtain more information about RHS protein complexes we tagged members of each of the three sub-families *in situ*. RHS2 and RHS6 tagged at the C-terminus with a haemagglutinin (HA) tag localized in the same way as the endogenous proteins (Supplementary Figure S2). RHS4 was consistently mislocalized to the cytoplasm, however, regardless of the tag or whether it was placed at the N- or C-terminus (Supplementary Figure S2). Similar problems with mislocalization were reported independently for RHS4 (Tryptag.org).

The tagged versions of RHS2 and RHS6 were used to immunoprecipitate complexes in order to identify interacting partners. All major interactions were confirmed to be DNA- and RNA-independent. The predominant components co-purifying with RHS2 were ribosomal proteins,
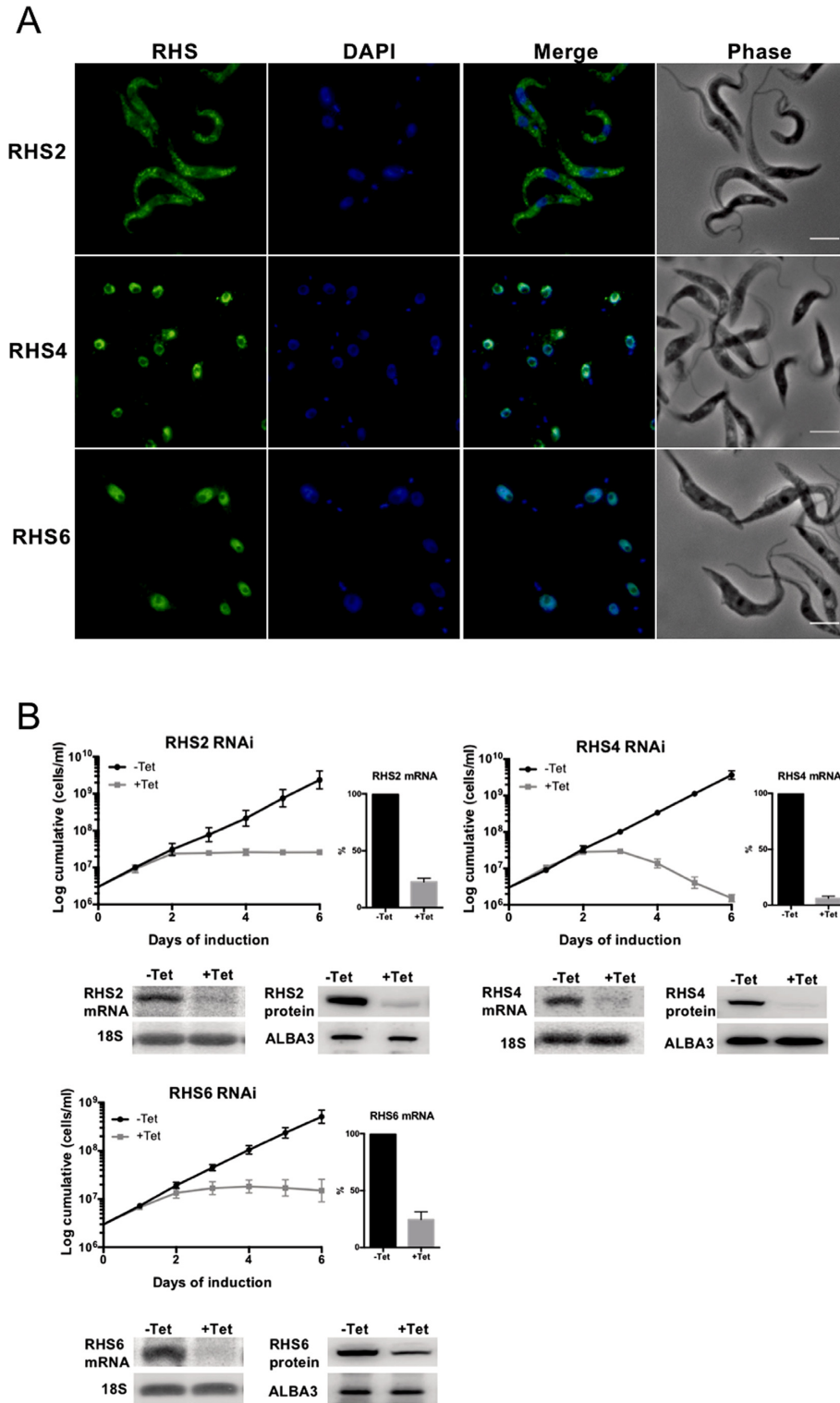
Figure 2. RHS are required for growth and viability of *T. brucei* procyclic forms (**A**) Immunolocalization of RHS proteins. Cells were incubated with anti-RHS2, anti-RHS4 or anti-RHS6 polyclonal antisera (first panel) and DAPI (second panel). The third and fourth panels show the merged fluorescence and phase contrast, respectively. Scale bar = 5 μm. (**B**) Effect of RHS RNAi on growth. The graphs display cumulative cell numbers from three biological replicates of each RNAi line. Uninduced cultures (–Tet) and cultures induced with tetracycline (+Tet) were monitored daily for 6 days. The efficiency of knockdown by RNAi was assessed by northern and western blot analysis on day 2 after induction of RNAi, and quantified with Fiji 2.0 in biological triplicates. Error bars = SD. 18S ribosomal RNA (northern blot) and Alba3 (Western blot) were used as loading controls. See also Supplementary Figure S1.
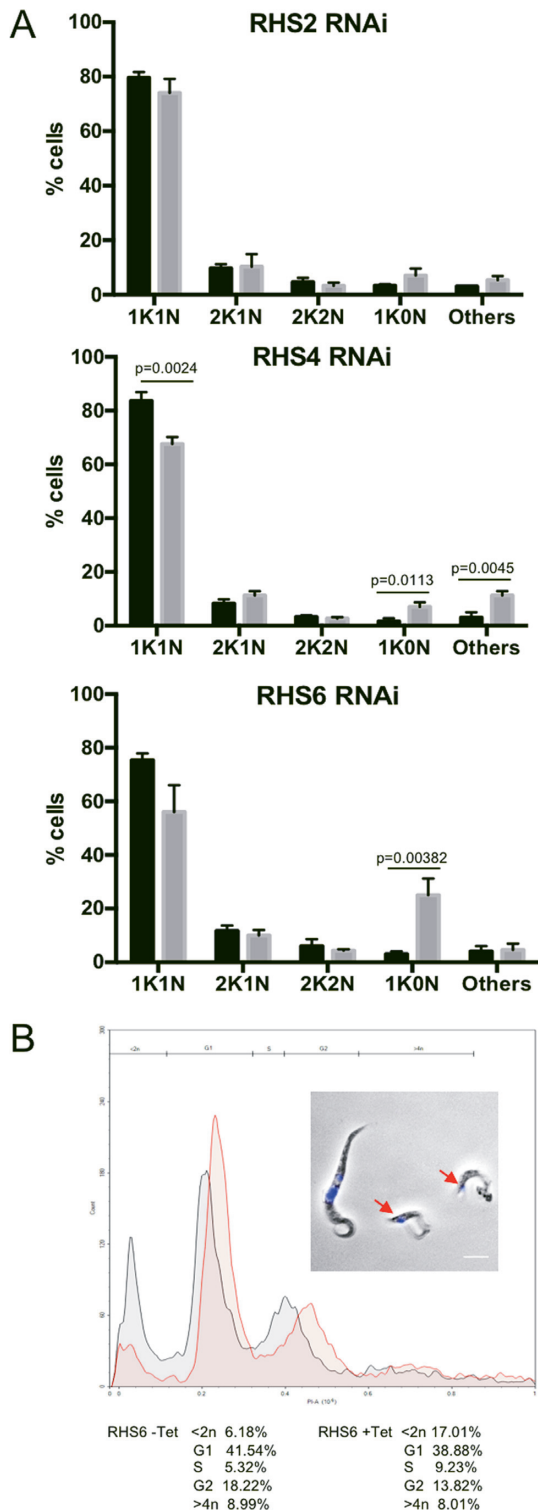
Figure 3. Depletion of RHS6 causes a defect in DNA replication (**A**) Effect of RHS RNAi on the cell cycle. Configurations of nuclear and kDNA were determined in uninduced (−Tet) and induced (+Tet) cells ($N = 200$ per sample). The analysis was performed at day 2 post induction. *P*-values are shown for unpaired, two-tailed *t*-tests. Error bars = SD. (**B**) DNA content and proportion of cells in different phases of the cell cycle. The figure shows one of three representative experiments. Cultures of the RHS6 RNAi cell line that were not induced (red), or induced for 2 days (grey), were analysed by flow cytometry following propidium iodide staining. The inset shows cells stained with DAPI. Arrowheads show two zoids (cells with kDNA, but no nucleus). Scale bar = 5 μm.

RNA binding proteins (e.g. puf6, DRBD2, polyadenylate binding proteins 1 and 2) and translation initiation factors (Tables 1A and Supplementary Table S1). The proteins present in the pulldown of RHS6 showed no consistent overlap with those obtained with RHS2. In addition to RHS6, members of the RHS1, 4 and 5 sub-families were present. The major category consisted of factors known to be involved in transcription elongation in other eukaryotes (Tables 1B and Supplementary Table S2), for example the two largest subunits of Pol II, Spt5 and 6, ISWI, FACT-Spt16 and several components of the PAF complex. Proteins involved in RNA processing (e.g. PTB2, PRP19, splicing auxiliary factors and symplekin) and RNA export were also present. A third category consisted of different subunits of the replication factor C complex, which may explain why knockdown of RHS6 results in a defect in DNA replication. Given that many of the proteins interacting with RHS2 and RHS6 possess RNA-binding domains, we performed RNA-immunoprecipitations (RIP). Although nucleic acids were present in both complexes after pulldown, these were lost when the RHS6 complex was treated with DNAse, demonstrating that RHS6 predominantly binds to DNA and not RNA (Supplementary Figure S3A). In contrast, the RHS2 complex contained both RNA and DNA. These data suggest that both RHS2 and RHS6 are involved in RNA metabolism, but at different levels: RHS6 is in a complex with proteins connected to transcription elongation and RNA processing, while RHS2, despite binding both DNA and RNA, mainly interacts with the translation machinery.

Chromatin immunoprecipitation reveals global association with RNA polymerase II transcription units

The results shown above strongly suggest that RHS proteins associate with chromatin. To identify DNA binding sites, we performed chromatin immunoprecipitation with antisera against RHS2, 4 and 6 followed by deep sequencing (ChIP-Seq). In parallel, we performed ChIP-Seq with antibodies directed against RPB1, the large subunit of Pol II. All three RHS sub-families showed the same distribution as Pol II (Figures 4A and B) with Pearson correlations ≥0.94, while correlations between the different RHS exceeded 0.99. Like Pol II, the RHS proteins covered almost the entire genome, but were depleted from regions known to be transcribed by Pol I such as the procyclin locus on chromosome 6 (Figure 4C). As a control, we performed ChIP with the anti-RBP1 antiserum or with a monoclonal antibody against the large subunit of Pol I, followed by qPCR. This confirmed the differential binding of the two polymerases to their respective transcription units (Supplementary Figure S3B). In addition, we found that RHS and Pol II were strongly depleted from silent regions of the genome. One example is a region on chromosome 4, encoding VSG and Expression Site Associated Genes (Figure 4D), which produces no steady state mRNA in this stage of the life cycle. The input genomic DNA gave good coverage of these regions, however, confirming that the lack of reads after ChIP was not due to a mapping artefact. Taken together, these data show that RHS are globally associated with active Pol II and suggest a potential role in transcription elongation.
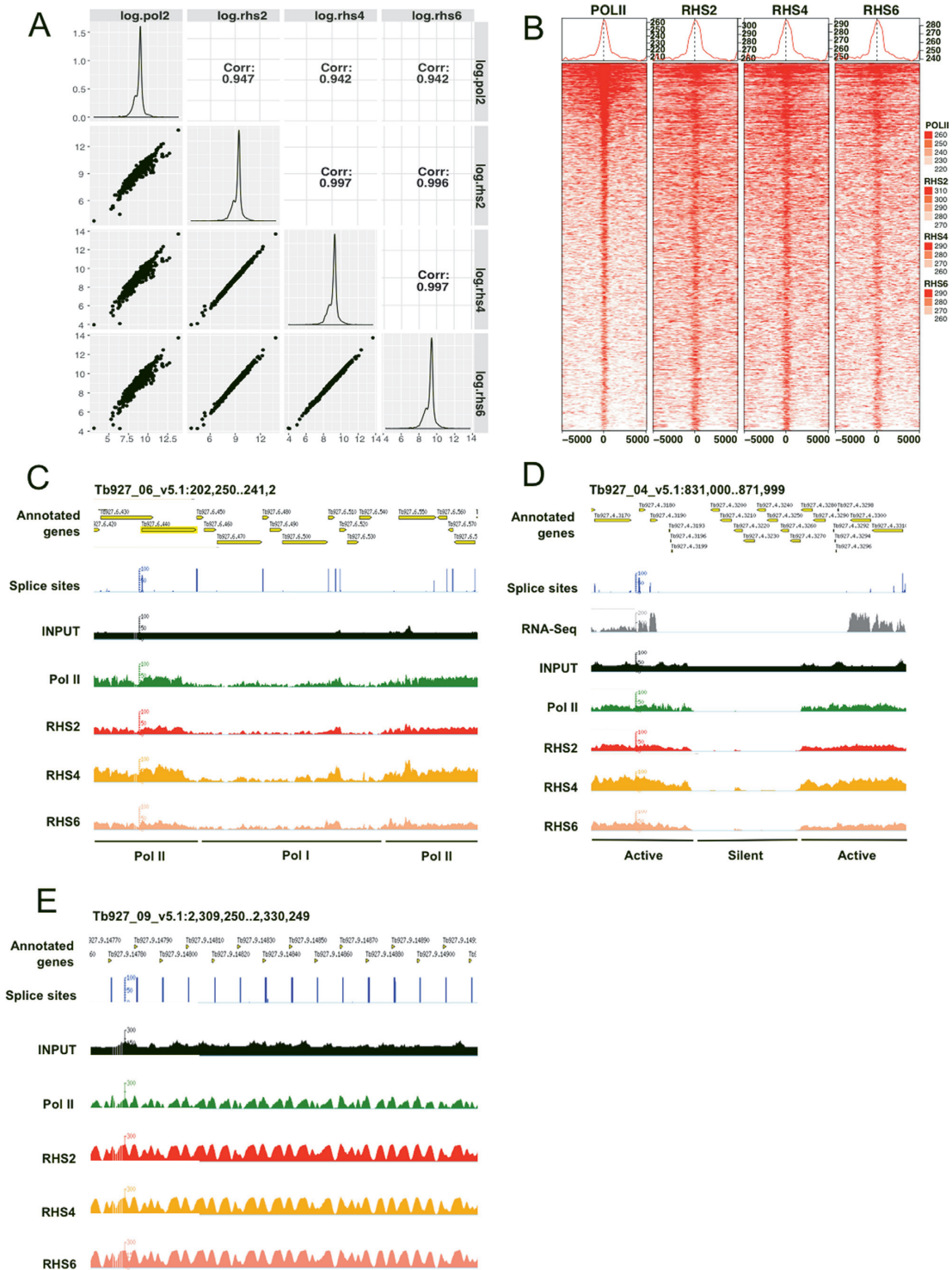
Figure 4. RHS2, RHS4 and RHS6 are globally associated with actively transcribing Pol II (**A**) Pairwise Spearman correlations of Pol II, RHS2, RHS4 and RHS6. Each dot represents the natural logarithm of the average read counts of a 10kb bin plotted pairwise between samples. The sub-figures over the diagonal (upper left to lower right corners) represent the distribution of Pol II, RHS2, RHS4 and RHS6, respectively. The correlation values corresponding to their mirror correlation graphs are above the diagonal. (**B**) Alignment of RHS2, 4 and 6 as a function of Pol II enrichment. On the x-axis the '0' corresponds to the summits of the peaks of Pol II ±5 kb. The colour values correspond to the coverage/depth (reads per kb per million reads mapped) over a certain position in a 10kb window. (**C**) RHS2, RHS4, RHS6 and Pol II occupancy across a region of chromosome 6 that includes a procyclin locus that is transcribed by Pol I (3). The flanking genes are transcribed by Pol II. Spliced leader addition sites are from Nilsson and colleagues (100). See Supplemental Figure S3b for additional controls. (**D**) RHS2, RHS4, RHS6 and Pol II occupancy across a region of chromosome 4 reveals that RHS proteins are associated with actively transcribed regions. Reads obtained from input chromatin are shown in black. Spliced leader addition sites are from Nilsson and colleagues (100). RNA-seq reads are from early procyclic forms (67). Genes are depicted as yellow bars and annotated according to TriTrypDB. (**E**) RHS2, RHS4, RHS6 and Pol II occupancy across part of the SL locus on chromosome 9. Reads obtained from input chromatin are shown in black. The 5′ ends of the spliced leader precursors are from reference 100.

Table 1. RHS2 and RHS6 interact with proteins involved in RNA metabolism

Gene ID	Product	PMSS average ($N = 3$)
A. RHS2-associated proteins		
Tb927.2.280/380/400	RHS2	1985
Tb927.9.10770	Polyadenylate-binding protein 2	595.18
Tb927.3.5050	60S ribosomal protein L4	460.87
Tb927.9.6070	40S ribosomal protein S3	454.24
Tb927.2.5910	40S ribosomal protein S13	448.72
Tb927.10.14710	40S ribosomal protein S2,	432.82
Tb927.7.6090	Eukaryotic translation initiation factor 3, subunit A	430.40
Tb927.10.3940	40S ribosomal protein S3A	418.16
Tb927.11.6300	40S ribosomal protein S5	415.28
Tb927.11.9710	60S ribosomal protein L10a	392
Tb927.8.1340	60S ribosomal protein L7a	365.25
Tb927.11.14020	Nuclear RNA binding domain 2 (NRBD2)	321.01
Tb927.5.2570	Eukaryotic translation initiation factor 3, subunit B	251.97
Tb927.10.8290	Eukaryotic translation initiation factor 3 subunit C	207.45
Tb927.10.11760	Pumilio/PUF RNA binding protein 6	206.97
Tb927.6.4370	Eukaryotic translation initiation factor 3, subunit D	179.68
Tb927.11.550	SCD6	179.21
Tb927.6.4440	RNA-binding protein 42	159.09
Tb927.9.13990	DRBD2, RNA-binding protein	150
Tb927.9.9290	Polyadenylate-binding protein-1	144.07
Tb927.11.11590	Eukaryotic translation initiation factor 3, subunit E	99.48
Tb927.8.1190	Eukaryotic translation initiation factor 3, subunit H	98.69
Tb927.11.13250	Eukaryotic translation initiation factor 2 gamma	84.86
Tb927.8.1500	Hypothetical protein, mRNA binding	70.43
Tb927.9.8130	Nascent polypeptide associated complex	61.98
B. RHS6-associated proteins		
Tb11.v5.0713	RHS6	4459
Tb927.1.420/2.240/2.1080	RHS5	3218
Tb927.1.120/2.340	RHS4	676.5
Tb927.5.4420	Nucleolar RNA helicase II	570.1
Tb927.10.7440	MTR4 RNA helicase	519.8
Tb927.3.5620	FACT complex subunit SPT16	486.3
Tb927.11.13360	AAA ATPase	481.5
Tb927.2.5240	PRP19-like protein	382.3
Tb927.11.11550	DNA topoisomerase II alpha	313.83
Tb927.2.100/1.220/2.370	RHS1	287
Tb927.2.1810	Chromatin-remodeling complex ATPase chain ISWI	285.3
Tb927.11.5650/6.3980/10.7990/11.9550	Replication factor C subunits	213.4
Tb927.10.13720	RNA binding protein 29	174
Tb927.8.7490	Symplekin	172
Tb927.8.7400	RNA polymerase IIA largest subunit (RPB1)	160
Tb927.8.2640	Ubiquitin-activating enzyme E1	146.8
Tb927.4.3810	DNA-directed RNA Polymerase II subunit (RPB2)	116.4
Tb927.10.7060	NUP96	97.8
Tb927.2.5810	Spt6 (transcription elongation)	115.3
Tb927.11.14100	Polypyrimidine tract-binding II (PTB2/DRBD4)	268.64
Tb927.3.3220	CTR9 (PAF complex)	132.11
Tb927.10.3200	U2 splicing auxiliary factor	177.5
Tb927.2.5030	Spt5 (transcription elongation)	62.70
Tb927.11.10230	CDC73 (PAF complex)	47.3

List of proteins consistently identified in immunoprecipitations of (A) RHS2 and (B) RHS6 with ranking according to PMSS average. All proteins listed were identified in three biological replicates. For simplicity, we have not included the full set of ribosomal proteins. The complete lists of proteins are shown in Supplemental Tables S1 and S2.

As shown above, the RHS2 complex contains RNA. Two RHS4 sub-family members are part of the Pol II complex (44,45) and an association with RNA has been reported for one of them (71). To determine whether the binding to chromatin was mediated by RNA we subjected ChIP samples to RNase A treatment during the immunoprecipitation. This had little effect on the amount of DNA pulled down with RHS6, but reduced the amount associated with RHS2 and RHS4 to 30% and 25%, respectively (Supplementary Figure S3C). This implies that RHS6 is in close proximity to

the DNA and that RNA partially bridges the interaction of RHS2 and RHS4 with DNA.

Depletion of RHS causes a block in spliced leader expression and RNA export

Under normal conditions the majority of RHS2 is in the cytoplasm (Figure 2A), but the finding that it is also associated with nuclear DNA suggests that it might be shuttling between the two compartments. In line with this possibility, bioinformatic analysis of the RHS2 family identified a puta-

tive nuclear localization signal and a putative nuclear export signal in each of the three family members. It has previously been reported that active transcription is required for nucleocytoplasmic shuttling of certain proteins and that blocking transcription with Actinomycin D (ActD) leads to their accumulation in the nucleus (72–74). Consistent with this, incubating trypanosomes with ActD resulted in an increase in nuclear RHS2 within 2 h (Figures 5A and B). An increase in nuclear RHS2 was also observed when RHS4 was knocked down, suggesting that RHS4 depletion and Act D have similar effects. This is again consistent with RHS4 playing a role in transcription.

To test if RHS are involved in transporting mRNAs to the cytoplasm, we analysed the RNAi cell lines by fluorescence *in situ* hybridization (FISH). For this we used two probes: an oligo d(T) probe that binds the poly(A) tails and a spliced leader (SL) probe that recognizes the conserved 39 base sequence at the 5′ end of all mRNAs. It has been shown previously that blocking export by down-regulating Mex67 causes poly(A) RNA to accumulate in the nucleus (33,75). We replicated these findings and also showed that RNAs detected with the SL probe accumulate in the same way (Figure 5C). Depletion of RHS2, 4 or 6 resulted in accumulation of poly(A) RNA in the nucleus (Figures 5C and D) indicating that RNA export was impaired. This was most pronounced for RHS4, which is down-regulated more efficiently than the other RHS. In contrast to the Mex67 RNAi line, however, RHS-depleted cells with nuclear poly(A) RNA had a strongly decreased signal for SL (Figure 5C). This suggests that the RNAs accumulating in the nucleus are not mature (i.e. have not been trans-spliced). The multiple copies of the SL precursor are transcribed by Pol II (18). Moreover, RBP1 and RHS proteins are associated with this region of the genome (Figure 4E), so one possibility is that knocking down RHS would lead to a defect in their synthesis. Northern blot analysis confirmed that cells depleted of RHS4 had lower levels of the SL precursor and of spliced mRNAs (Figure 5E). Since treatment with ActD leads to increased nuclear RHS2, we analysed its effect on the distribution of poly(A) RNA. Within 2 h there was a clear increase in the amount of poly(A) RNA in the nucleus (Supplementary Figure S4). Taken together, these results show that blocking transcription results in nuclear accumulation of mRNA and that the same phenotype is observed when RHS proteins are depleted. This suggests that active transcription and mRNA export are interconnected, which is consistent with recent results from Goos *et al.* (36). In this respect trypanosomes resemble yeast and metazoa, which coordinate these two processes via transcription elongation complexes (76), although they use alternative factors to accomplish this.

RHS depletion causes a global reduction in RNA polymerase II transcription

If RHS are required for transcription elongation, we would predict that knocking them down would affect the synthesis of nascent RNAs. To test this, cultures were pulsed with 5-ethynyl uridine (5-EU) for 10 min. Cells were subsequently separated into nuclear and cytoplasmic fractions before extracting RNA, biotinylating the ethynyl groups,

and cleaving the RNA into fragments of ∼300 bases. Fragmentation is necessary to ensure that only newly synthesized RNAs, rather than entire polycistronic transcripts, are isolated. Biotinylated RNA was then captured with streptavidin beads. With this length of pulse, approximately equal amounts of incorporated 5-EU were in the nuclear and cytoplasmic fractions of wild type cells (Supplementary Figure S5A). When the RNAi lines were analysed by the same method, depletion of RHS resulted in reduced amounts of 5-EU-labelled nuclear RNA in all three cases (Figure 6A). To determine whether this effect was global or restricted to specific regions of the genome, the nascent RNAs were sequenced. Global run-on sequencing (GRO-Seq) has not been performed before in trypanosomes and at present there are no reliable methods for depleting ribosomal RNAs. To overcome this problem, we sequenced the entire biotinylated RNA obtained from the RNAi lines and mapped the reads to the coding regions of genes. In each case we obtained a minimum of 1.25 million mapped reads per sample. Biological replicates showed excellent correlations (>0.93; Supplementary Figure S5B). Moreover, the correlation between induced and uninduced RNAi lines within an experiment was >0.98 (Figure 6B). This indicates that the same regions are transcribed, albeit to a lesser extent when RHS were depleted. In summary, our results demonstrate that RHS proteins are globally involved in Pol II transcription in trypanosomes.

DISCUSSION

Trypanosoma brucei transcriptional regulation is divergent from most other eukaryotes and may take advantage of some species-specific factors to overcome these differences. We hypothesized that RHS proteins, which were previously found to associate with chromatin and Pol II (44–46), might play such a role in trypanosomes. These proteins have been neglected since their discovery, in part because of the lack of similarity to any other known proteins outside the genus *Trypanosoma* and in part because of the absence of informative functional domains. RHS mRNAs and proteins are extremely abundant and have often been identified in transcriptomic and proteomic studies, but tend to be excluded from analyses, either as probable contaminants or because of their daunting complexity (17,77,78). We show that members of the RHS2, RHS4 and RHS6 sub-families are required for the growth of trypanosomes and that each sub-family plays non-redundant and essential roles in transcription and mRNA export. Together they form part of a larger network in which RHS6 is most closely associated with the DNA, RHS4 is part of the core complex of Pol II (44,45) and RHS2 connects transcription with the translation machinery.

RHS2 and RHS6 interact with different sets of proteins, with the former associating mainly with cytosolic RNA-binding proteins and translation initiation factors, and the latter with all the other nuclear RHS (RHS1, RHS4 and RHS5), as well as factors involved in chromatin remodelling, and transcription elongation. Orthologues of known elongation factors in other eukaryotes co-precipitate with the RHS6 complex and show good overlap with proteins recently found to interact with Pol II (17). In addition to
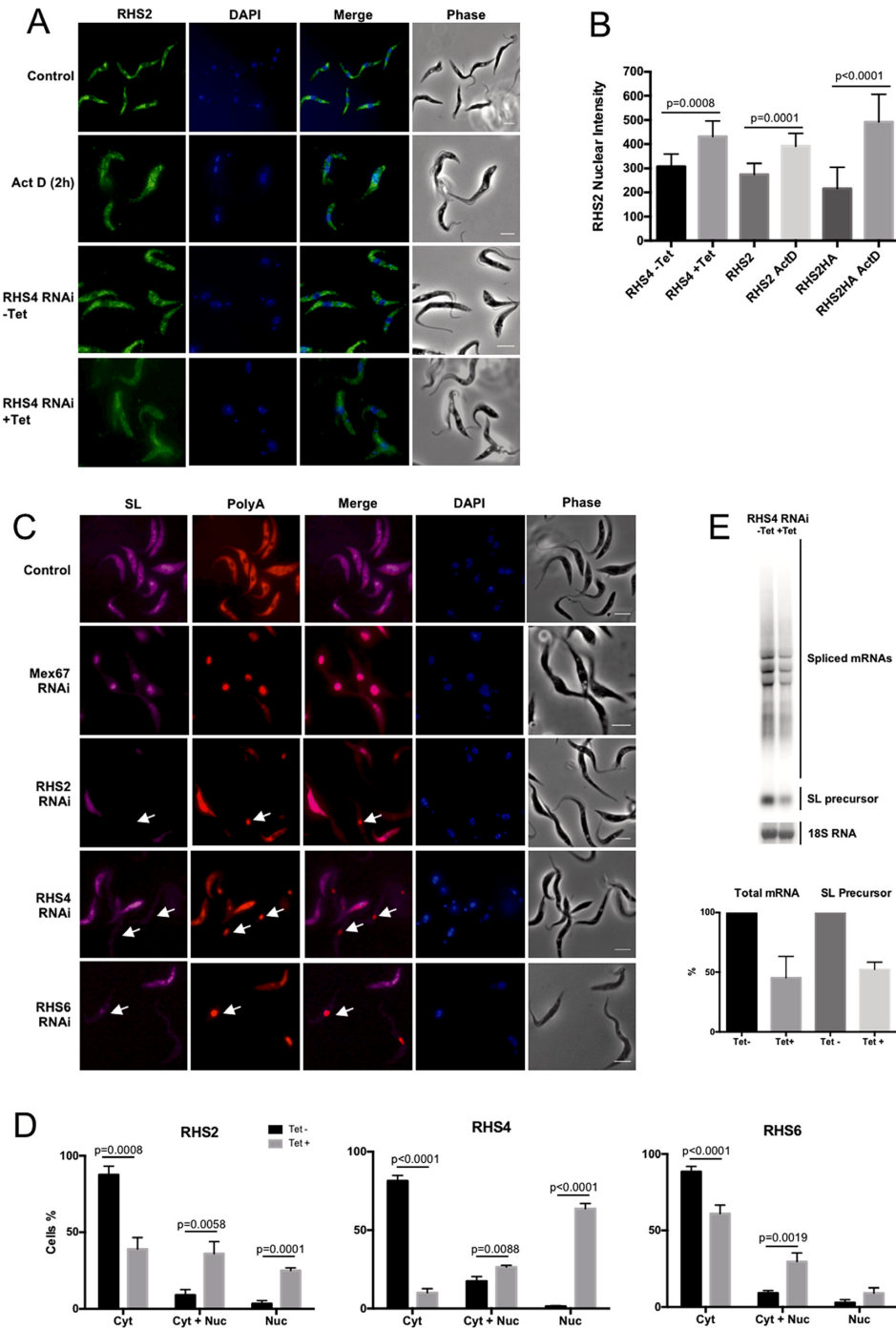
Figure 5. RHS depletion causes defects in RNA export and spliced leader expression (**A**) Redistribution of RHS2 in response to inhibition of Pol II or depletion of RHS4. Immunofluorescence with anti-RHS2 antisera shows that RHS2 is largely excluded from the nucleus of untreated control cells or the uninduced RHS4 RNAi cell line, but shows increased nuclear localization in cells treated with 10 μg ml^{-1} Actinomycin D (ActD) for 2 h or in cells from the RHS4 RNAi line 2 days post induction. Cells were co-stained with DAPI. Scale bar = 5μm. (**B**) Quantification of nuclear fluorescence signals with Fiji 2.0 using the cell lines described above. As an additional control, a cell line in which RHS2 is tagged with an HA epitope (RHS2HA) was either untreated, or treated with ActD for 2 h, and stained with an anti-HA antibody. P-values are shown for unpaired, two-tailed t-tests. Error bars = SD. (**C**) Fluorescence in situ hybridization (FISH) to detect spliced and polyadenylated RNAs. RNAi lines for Mex67, RHS2, RHS4 or RHS6 were induced with tetracycline for 2 days and hybridized with Cy3-oligo d(T) to detect poly(A) tails (PolyA) and a Cy5-labelled probe complementary to the spliced leader (SL). The latter also binds to the spliced leader precursor. Cells were stained in parallel with DAPI. Scale bar = 5 μm. The RHS4 RNAi cell line cultured in the absence of tetracycline served as the control. (**D**) Quantification of polyA signals with Fiji 2.0. Biological triplicates were performed for each cell line. Cyt: >70% of the signal is cytoplasmic. Nuc: >70% of the signal is nuclear. For each sample, n = 200 cells. P-values are shown for unpaired, two-tailed *t*-tests. Error bars = SD. White arrows point to examples of cells with nuclear accumulation of poly(A) but no detectable SL RNA. (**E**) Upper panel: Northern blot analysis of total RNA extracted from cells cultured without tetracycline (–Tet) or induced with tetracycline for 2 days (+Tet). An oligonucleotide complementary to the spliced leader, 5′-labelled with ^{32}P-phosphate, was used as a probe. Signals were detected using a Typhoon FLA 7000 (GE Healthcare). Lower panel: quantification of 3 biological replicates. Signals for mRNA and the SL precursor RNA were normalized to 18S rRNA. Error bars = SD.
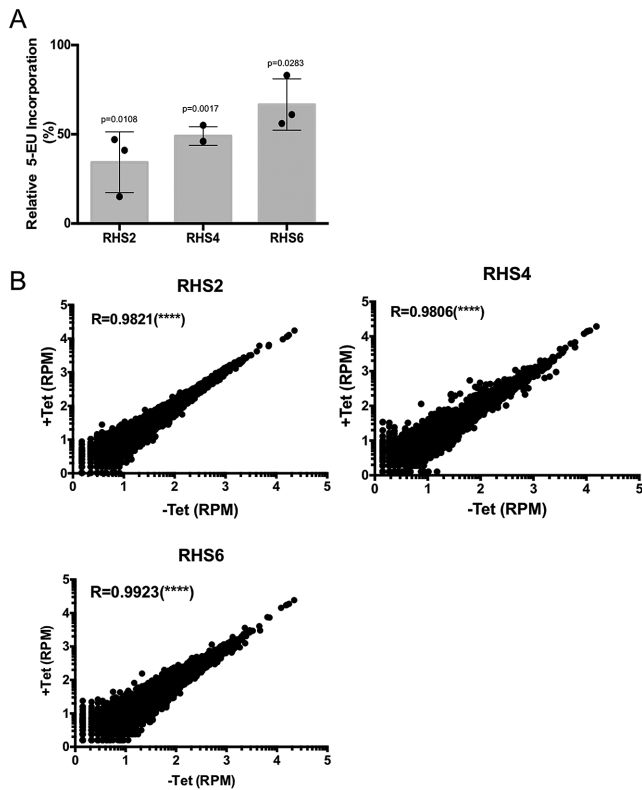
A

B

Figure 6. RHS are involved in global mRNA synthesis (**A**) Quantification of 5-ethynyl uridine (5-EU) incorporation. Each RNAi line was cultured in the presence or absence of tetracycline for 2 days and pulsed for 10 min with 5-EU ($n = 3$). Values are given as a percentage of the uninduced control. P-values are shown for paired, one-tailed t-tests. Error bars = SD. (**B**) Correlation between reads per million (RPM) for nascent mRNAs from RNAi lines cultured in the presence or absence of tetracycline for 2 days. R: Pearson correlation coefficient. Biological replicates were performed. See also Supplemental Figure S5.

the factors described above, several subunits of Replication Factor C, a protein complex known to be of major importance in DNA replication, were associated with RHS6. This might explain why RHS6 depletion compromises the replication of nuclear DNA resulting in the accumulation of zoids. It is difficult, however, to know whether this effect is direct or a secondary consequence of a transcriptional defect. A connection between transcription and replication has been demonstrated in several organisms (79–81). These two processes may also be interlinked in trypanosomes, as two initiators of replication, ORC1 and ORC6, bind at transcription start sites (TSS) and their depletion causes changes in mRNA abundance (82).

Further evidence that RHS are involved in transcription elongation comes from ChIP-seq, with excellent correlations between the binding sites for RHS and active Pol II. In other eukaryotes the histone modification H3K36me3 is a marker for elongation by Pol II (83–85). Unfortunately, this epitope is only partially conserved in trypanosomes and commercial antibodies directed against it are not specific (data not shown). Moreover, no recognisable orthologues of Set2 histone methyltransferase can be identified. Interestingly, we did not find evidence of Pol II accumulation at TSS. This may reflect the polycistronic nature of transcrip-

tion, rather than the individual regulation of promoters operating in the majority of eukaryotes. It is also worth noting that trypanosomes, like yeasts (86,87), do not encode recognisable orthologues of NELF subunits which pause Pol II at promoters in metazoa (88,89). The ChIP-seq experiments also indicate that silent regions of the genome are silent in every respect, with no Pol II binding, no trans-splicing and no RHS. It is possible that these regions are covered with other proteins to prevent access to them, or that some regions of the genome are sequestered in a particular domain of the nucleus that is less accessible.

Depletion of RHS caused a reduction in the incorporation of 5-EU into nascent RNA, prompting us to establish GRO-Seq to determine which RNAs were affected. In keeping with their distribution throughout the transcription units, RHS seem to be required for the production of all mRNAs transcribed by Pol II. Once again, this suggests that RHS are general facilitators of transcription. While we cannot exclude entirely that RHS might also be required for transcription initiation, their main roles appear to be in elongation.

Messenger RNA export is another process that is tightly connected to transcription in other organisms (90–93). This would also appear to be true of *T. brucei*, as depletion of any of the RHS resulted in accumulation of nuclear RNAs that were polyadenylated, but not spliced. This differed from the phenotype caused by the depletion of the export factor Mex67, which resulted in accumulation of mRNAs that were trans-spliced and polyadenylated. These disparities might be explained by the finding that the SL precursor, which is also transcribed by Pol II, was reduced after knockdown of RHS. In other eukaryotes, actively transcribing Pol II is localized close to nuclear pores to facilitate mRNA export (94–96). Consistent with this, we found that RNA export factors and nucleoporins were associated with RHS6. There are numerous examples of proteins involved in transcription that shuttle between the nucleus and the cytoplasm (72–74). We postulate that this might be the case for RHS2, which is predominantly cytoplasmic, but is also associated with chromatin.

Multiple copies of RHS genes have been annotated in all African trypanosomes (*T. brucei, T. congolense, T. vivax*) and in other members of the genus *Trypanosoma*, such as *T. grayi, T. theileri* and *T. cruzi*. *T. cruzi* RHS have been identified in the nuclear proteome and shown to be upregulated after heat shock, but nothing is known about their function (97,98). *Leishmania* spp., despite belonging to the trypanosomatids, have no RHS genes. A possible explanation could come from phylogenetic studies, which indicate that *Leishmania* and trypanosomes have quite distant evolutionary histories (99). One reason that the RHS family might have expanded is that they contain a hot spot for retrotransposon insertion, which creates nonsense mutations or frameshifts within the coding sequence and results in the production of truncated proteins. As a consequence, the parasites may use gene duplication as a means of preserving intact copies of the genes. The fact that RHS genes are largely sub-telomeric would aid recombination. Both gene duplication and retrotransposon insertion seem to happen independently in different trypanosome strains, as there is a lack of synteny in the distribution of RHS genes (Trit-

rypDB). All in all, this suggests that retrotransposon insertion is an undesirable event that might still occur.

In conclusion, our data show that RHS proteins are novel factors that are intimately associated with the production and export of Pol II transcripts in trypanosomes. This, in turn, demonstrates that the components of eukaryotic transcription are far from being universal, and indicates that studies of a diverse range of organisms would be beneficial to our understanding of how this process has evolved over time.

DATA AVAILABILITY

Raw read files are deposited at the European Nucleotide Archives (ENA) http://www.ebi.ac.uk/ena as study PR-JEB29185.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. McCulloch,R., Cobbold,C.A., Figueiredo,L., Jackson,A., Morrison,L.J., Mugnier,M.R., Papavasiliou,N., Schnaufer,A. and Matthews,K. (2017) Emerging challenges in understanding trypanosome antigenic variation. *Emerg. Top. Life Sci.*, **1**, 585–592.
2. Vickerman,K. (1994) The evolutionary expansion of the trypanosomatid flagellates. *Int. J. Parasitol.*, **24**, 1317–1331.
3. Günzl,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.
4. Clayton,C.E. (2014) Networks of gene expression regulation in *Trypanosoma brucei. Mol. Biochem. Parasitol.*, **195**, 96–106.
5. Osbourn,A.E. and Field,B. (2009) Operons. *Cell Mol. Life Sci.*, **66**, 3755–3775.
6. Blumenthal,T. (1998) Gene clusters and polycistronic transcription in eukaryotes. *Bioessays*, **20**, 480–487.
7. Spieth,J., Brooke,G., Kuersten,S., Lea,K. and Blumenthal,T. (1993) Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell*, **73**, 521–532.
8. Guiliano,D.B. and Blaxter,M.L. (2006) Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genet.*, **2**, e198.
9. Michalak,K., Orr,W.C. and Radyuk,S.N. (2008) Drosophila peroxiredoxin 5 is the second gene in a dicistronic operon. *Biochem. Biophys. Res. Commun.*, **368**, 273–278.
10. Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R. *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, **309**, 436–442.
11. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E. and Haas,B. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei. Science*, **309**, 416–422.
12. Vassella,E., Braun,R. and Roditi,I. (1994) Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site: a dual role for polypyrimidine tracts in trypanosomes? *Nucleic Acids Res.*, **22**, 1359–1364.
13. 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.
14. Ullu,E., Matthews,K.R. and Tschudi,C. (1993) Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol. Cell Biol.*, **13**, 720–725.
15. Schurch,N., Hehl,A., Vassella,E., Braun,R. and Roditi,I. (1994) Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. *Mol. Cell Biol.*, **14**, 3668–3675.
16. Hug,M., Hotz,H.R., Hartmann,C. and Clayton,C. (1994) Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. *Mol. Cell Biol.*, **14**, 7428–7435.
17. Srivastava,A., Badjatia,N., Lee,J.H., Hao,B. and Günzl,A. (2018) An RNA polymerase II-associated TFIIF-like complex is indispensable for SL RNA gene transcription in *Trypanosoma brucei. Nucleic Acids Res.*, **46**, 1695–1709.
18. Gilinger,G. and Bellofatto,V. (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res.*, **29**, 1556–1564.
19. Günzl,A., Ullu,E., Dorner,M., Fragoso,S.P., Hoffmann,K.F., Milner,J.D., Morita,Y., Nguu,E.K., Vanacova,S., Wunsch,S. *et al.* (1997) Transcription of the *Trypanosoma brucei* spliced leader RNA gene is dependent only on the presence of upstream regulatory elements. *Mol. Biochem. Parasitol.*, **85**, 67–76.
20. Kolev,N.G., Franklin,J.B., Carmi,S., Shi,H., Michaeli,S. and Tschudi,C. (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog.*, **6**, e1001090.
21. Wedel,C., Forstner,K.U., Derr,R. and Siegel,T.N. (2017) GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J.*, **36**, 2581–2594.
22. 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.
23. Wright,J.R., Siegel,T.N. and Cross,G.A. (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei. Mol. Biochem. Parasitol.*, **172**, 141–144.
24. Gassen,A., Brechtefeld,D., Schandry,N., Arteaga-Salas,J.M., Israel,L., Imhof,A. and Janzen,C.J. (2012) DOT1A-dependent H3K76 methylation is required for replication regulation in *Trypanosoma brucei. Nucleic Acids Res.*, **40**, 10302–10311.
25. Alsford,S. and Horn,D. (2012) Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. *Nucleic Acids Res.*, **40**, 10150–10160.
26. Borst,P. and Sabatini,R. (2008) Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.*, **62**, 235–251.
27. Reynolds,D.L., Hofmeister,B.T., Cliffe,L., Siegel,T.N., Anderson,B.A., Beverley,S.M., Schmitz,R.J. and Sabatini,R. (2016) Base J represses genes at the end of polycistronic gene clusters in

Leishmania major by promoting RNAP II termination. *Mol. Microbiol.*, **101**, 559–574.

28. Schimanski,B., Brandenburg,J., Nguyen,T.N., Caimano,M.J. and Günzl,A. (2006) A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. *Nucleic Acids Res.*, **34**, 1676–1684.

29. Schimanski,B., Nguyen,T.N. and Günzl,A. (2005) Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. *Mol. Cell Biol.*, **25**, 7303–7313.

30. Palenchar,J.B., Liu,W., Palenchar,P.M. and Bellofatto,V. (2006) A divergent transcription factor TFIIB in trypanosomes is required for RNA polymerase II-dependent spliced leader RNA transcription and cell viability. *Eukaryot. Cell*, **5**, 293–300.

31. 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.

32. Selth,L.A., Sigurdsson,S. and Svejstrup,J.Q. (2010) Transcript Elongation by RNA Polymerase II. *Annu. Rev. Biochem.*, **79**, 271–293.

33. Dostalova,A., Käser,S., Cristodero,M. and Schimanski,B. (2013) The nuclear mRNA export receptor Mex67-Mtr2 of *Trypanosoma brucei* contains a unique and essential zinc finger motif. *Mol. Microbiol.*, **88**, 728–739.

34. Katahira,J. (2012) mRNA export and the TREX complex. *Biochim. Biophys. Acta*, **1819**, 507–513.

35. Serpeloni,M., Moraes,C.B., Muniz,J.R., Motta,M.C., Ramos,A.S., Kessler,R.L., Inoue,A.H., daRocha,W.D., Yamada-Ogatta,S.F., Fragoso,S.P. et al. (2011) An essential nuclear protein in trypanosomes is a component of mRNA transcription/export pathway. *PLoS One*, **6**, e20730.

36. Goos,C., Dejung,M., Wehman,A.M., E,M.N., Schmidt,J., Sunter,J., Engstler,M., Butter,F. and Kramer,S. (2018) Trypanosomes can initiate nuclear export co-transcriptionally. *Nucleic Acids Res.*, doi:10.1093/nar/gky1136.

37. Das,A. and Bellofatto,V. (2009) The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in *Trypanosoma brucei*. *PLoS One*, **4**, e6959.

38. Buratowski,S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell*, **36**, 541–546.

39. Komarnitsky,P., Cho,E.J. and Buratowski,S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.*, **14**, 2452–2460.

40. Mayer,A., Heidemann,M., Lidschreiber,M., Schreieck,A., Sun,M., Hintermair,C., Kremmer,E., Eick,D. and Cramer,P. (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*, **336**, 1723–1725.

41. Cadena,D.L. and Dahmus,M.E. (1987) Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.*, **262**, 12468–12474.

42. Urbaniak,M.D., Martin,D.M. and Ferguson,M.A. (2013) Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*. *J. Proteome Res.*, **12**, 2233–2244.

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

44. Das,A., Li,H., Liu,T. and Bellofatto,V. (2006) Biochemical characterization of *Trypanosoma brucei* RNA polymerase II. *Mol. Biochem. Parasitol.*, **150**, 201–210.

45. Devaux,S., Lecordier,L., Uzureau,P., Walgraffe,D., Dierick,J.F., Poelvoorde,P., Pays,E. and Vanhamme,L. (2006) Characterization of RNA polymerase II subunits of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **148**, 60–68.

46. Naguleswaran,A., Gunasekera,K., Schimanski,B., Heller,M., Hemphill,A., Ochsenreiter,T. and Roditi,I. (2015) *Trypanosoma brucei* RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure. *MBio*, **6**, e00114.

47. Bringaud,F., Biteau,N., Melville,S.E., Hez,S., El-Sayed,N.M., Leech,V., Berriman,M., Hall,N., Donelson,J.E. and Baltz,T. (2002) A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic

48. subtelomeric regions of *Trypanosoma brucei*. *Eukaryot. Cell.*, **1**, 137–151.

48. Le Ray,D., Barry,J.D., Easton,C. and Vickerman,K. (1977) First tsetse fly transmission of the "AnTat" serodeme of *Trypanosoma brucei*. *Ann. Soc. Belg. Med. Trop.*, **57**, 369–381.

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

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

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

52. Kramer,S., Queiroz,R., Ellis,L., Webb,H., Hoheisel,J.D., Clayton,C. and Carrington,M. (2008) Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169. *J. Cell Sci.*, **121**, 3002–3014.

53. Kerry,L.E., Pegg,E.E., Cameron,D.P., Budzak,J., Poortinga,G., Hannan,K.M., Hannan,R.D. and Rudenko,G. (2017) Selective inhibition of RNA polymerase I transcription as a potential approach to treat African trypanosomiasis. *PLoS Negl. Trop. Dis.*, **11**, e0005432.

54. Mani,J., Güttinger,A., Schimanski,B., Heller,M., Acosta-Serrano,A., Pescher,P., Späth,G. and Roditi,I. (2011) Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS One*, **6**, e22463.

55. Schimanski,B., Nguyen,T.N. and Günzl,A. (2005) Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. *Eukaryot. Cell*, **4**, 1942–1950.

56. Chomczynski,P. and Sacchi,N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.

57. Roditi,I., Carrington,M. and Turner,M. (1987) Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature*, **325**, 272–274.

58. Flück,C., Salomone,J.Y., Kurath,U. and Roditi,I. (2003) Cycloheximide-mediated accumulation of transcripts from a procyclin expression site depends on the intergenic region. *Mol. Biochem. Parasitol.*, **127**, 93–97.

59. Cassola,A., De Gaudenzi,J.G. and Frasch,A.C. (2007) Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. *Mol. Microbiol.*, **65**, 655–670.

60. Lowell,J.E. and Cross,G.A. (2004) A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. *J. Cell Sci.*, **117**, 5937–5947.

61. Dahl,J.A. and Collas,P. (2007) A quick and quantitative chromatin immunoprecipitation assay for small cell samples. *Front. Biosci.*, **12**, 4925–4931.

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

63. Schumann Burkard,G., Käser,S., de Araujo,P.R., Schimanski,B., Naguleswaran,A., Knüsel,S., Heller,M. and Roditi,I. (2013) Nucleolar proteins regulate stage-specific gene expression and ribosomal RNA maturation in *Trypanosoma brucei*. *Mol. Microbiol.*, **88**, 827–840.

64. Murphy,N.B., Pays,A., Tebabi,P., Coquelet,H., Guyaux,M., Steinert,M. and Pays,E. (1987) *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. *J. Mol. Biol.*, **195**, 855–871.

65. Hong,V., Presolski,S.I., Ma,C. and Finn,M.G. (2009) Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed. Engl.*, **48**, 9879–9883.

66. Afgan,E., Baker,D., Batut,B., van den Beek,M., Bouvier,D., Cech,M., Chilton,J., Clements,D., Coraor,N., Gruning,B.A. et al. (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.*, **46**, W537–W544.

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

68. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup. (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

69. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.

70. Ramirez,F., Dundar,F., Diehl,S., Gruning,B.A. and Manke,T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.

71. Lueong,S., Merce,C., Fischer,B., Hoheisel,J.D. and Erben,E.D. (2016) Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome. *Mol. Microbiol.*, **100**, 457–471.

72. Inoue,A.H., Serpeloni,M., Hiraiwa,P.M., Yamada-Ogatta,S.F., Muniz,J.R., Motta,M.C., Vidal,N.M., Goldenberg,S. and Avila,A.R. (2014) Identification of a novel nucleocytoplasmic shuttling RNA helicase of trypanosomes. *PLoS One*, **9**, e109521.

73. Lee,S., Neumann,M., Stearman,R., Stauber,R., Pause,A., Pavlakis,G.N. and Klausner,R.D. (1999) Transcription-dependent nuclear-cytoplasmic trafficking is required for the function of the von Hippel-Lindau tumor suppressor protein. *Mol. Cell Biol.*, **19**, 1486–1497.

74. Smillie,D.A. and Sommerville,J. (2002) RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. *J. Cell Sci.*, **115**, 395–407.

75. Schwede,A., Manful,T., Jha,B.A., Helbig,C., Bercovich,N., Stewart,M. and Clayton,C. (2009) The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. *Nucleic Acids Res.*, **37**, 5511–5528.

76. Perales,R. and Bentley,D. (2009) "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell*, **36**, 178–191.

77. Christiano,R., Kolev,N.G., Shi,H., Ullu,E., Walther,T.C. and Tschudi,C. (2017) The proteome and transcriptome of the infectious metacyclic form of *Trypanosoma brucei* define quiescent cells primed for mammalian invasion. *Mol. Microbiol.*, **106**, 74–92.

78. Ambrosio,D.L., Badjatia,N. and Gunzl,A. (2015) The spliceosomal PRP19 complex of trypanosomes. *Mol. Microbiol.*, **95**, 885–901.

79. Saponaro,M., Kantidakis,T., Mitter,R., Kelly,G.P., Heron,M., Williams,H., Soding,J., Stewart,A. and Svejstrup,J.Q. (2014) RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell*, **157**, 1037–1049.

80. MacAlpine,D.M., Rodriguez,H.K. and Bell,S.P. (2004) Coordination of replication and transcription along a Drosophila chromosome. *Genes Dev.*, **18**, 3094–3105.

81. Ramamoorthy,M., Tadokoro,T., Rybanska,I., Ghosh,A.K., Wersto,R., May,A., Kulikowicz,T., Sykora,P., Croteau,D.L. and Bohr,V.A. (2012) RECQL5 cooperates with Topoisomerase II alpha in DNA decatenation and cell cycle progression. *Nucleic Acids Res.*, **40**, 1621–1635.

82. Tiengwe,C., Marcello,L., Farr,H., Gadelha,C., Burchmore,R., Barry,J.D., Bell,S.D. and McCulloch,R. (2012) Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One*, **7**, e32674.

83. Li,B., Carey,M. and Workman,J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707–719.

84. Krogan,N.J., Kim,M., Tong,A., Golshani,A., Cagney,G., Canadien,V., Richards,D.P., Beattie,B.K., Emili,A., Boone,C. *et al.* (2003) Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.*, **23**, 4207–4218.

85. Fong,N., Saldi,T., Sheridan,R.M., Cortazar,M.A. and Bentley,D.L. (2017) RNA Pol II dynamics modulate Co-transcriptional chromatin modification, CTD phosphorylation, and transcriptional direction. *Mol. Cell*, **66**, 546–557.

86. Hartzog,G.A. and Fu,J. (2013) The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta*, **1829**, 105–115.

87. Booth,G.T., Wang,I.X., Cheung,V.G. and Lis,J.T. (2016) Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Res.*, **26**, 799–811.

88. Vos,S.M., Farnung,L., Urlaub,H. and Cramer,P. (2018) Structure of paused transcription complex Pol II-DSIF-NELF. *Nature*, **560**, 601–606.

89. Yamaguchi,Y., Takagi,T., Wada,T., Yano,K., Furuya,A., Sugimoto,S., Hasegawa,J. and Handa,H. (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, **97**, 41–51.

90. Wickramasinghe,V.O. and Laskey,R.A. (2015) Control of mammalian gene expression by selective mRNA export. *Nat. Rev. Mol. Cell Biol.*, **16**, 431–442.

91. Rubin,J.D. and Taatjes,D.J. (2015) Molecular biology: mediating transcription and RNA export. *Nature*, **526**, 199–200.

92. Gwizdek,C., Iglesias,N., Rodriguez,M.S., Ossareh-Nazari,B., Hobeika,M., Divita,G., Stutz,F. and Dargemont,C. (2006) Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 16376–16381.

93. Garcia-Oliver,E., Garcia-Molinero,V. and Rodriguez-Navarro,S. (2012) mRNA export and gene expression: the SAGA-TREX-2 connection. *Biochim. Biophys. Acta*, **1819**, 555–565.

94. Ibarra,A. and Hetzer,M.W. (2015) Nuclear pore proteins and the control of genome functions. *Genes Dev.*, **29**, 337–349.

95. Dieppois,G. and Stutz,F. (2010) Connecting the transcription site to the nuclear pore: a multi-tether process that regulates gene expression. *J. Cell Sci.*, **123**, 1989–1999.

96. Kurshakova,M.M., Krasnov,A.N., Kopytova,D.V., Shidlovskii,Y.V., Nikolenko,J.V., Nabirochkina,E.N., Spehner,D., Schultz,P., Tora,L. and Georgieva,S.G. (2007) SAGA and a novel Drosophila export complex anchor efficient transcription and mRNA export to NPC. *EMBO J.*, **26**, 4956–4965.

97. dos Santos Ade,C. Jr., Kalume,D.E., Camargo,R., Gomez-Mendoza,D.P., Correa,J.R., Charneau,S., de Sousa,M.V., de Lima,B.D. and Ricart,C.A. (2015) Unveiling the *Trypanosoma cruzi* Nuclear Proteome. *PLoS One*, **10**, e0138667.

98. Perez-Morales,D., Lanz-Mendoza,H., Hurtado,G., Martinez-Espinosa,R. and Espinoza,B. (2012) Proteomic analysis of T*rypanosoma cruzi* epimastigotes subjected to heat shock. *J. Biomed. Biotechnol.*, **2012**, 902803.

99. Hamilton,P.B., Stevens,J.R., Gaunt,M.W., Gidley,J. and Gibson,W.C. (2004) Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *Int. J. Parasitol.*, **34**, 1393–1404.

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