GENETICS

A long noncoding RNA promotes parasite differentiation in African trypanosomes

Fabien Guegan[1]*, K. Shanmugha Rajan[2]†, Fábio Bento[1]‡, Daniel Pinto-Neves[1]§, Mariana Sequeira[1], Natalia Gumińska[3], Seweryn Mroczek[3,4], Andrzej Dziembowski[3]‡, Smadar Cohen-Chalamish[2], Tirza Doniger[2], Beathrice Galili[2], Antonio M. Estévez[5], Cedric Notredame[6,7], Shulamit Michaeli[2], Luisa M. Figueiredo[1]*

The parasite *Trypanosoma brucei* causes African sleeping sickness that is fatal to patients if untreated. Parasite differentiation from a replicative slender form into a quiescent stumpy form promotes host survival and parasite transmission. Long noncoding RNAs (lncRNAs) are known to regulate cell differentiation in other eukaryotes. To determine whether lncRNAs are also involved in parasite differentiation, we used RNA sequencing to survey the *T. brucei* genome, identifying 1428 previously uncharacterized lncRNA genes. We find that *grumpy* lncRNA is a key regulator that promotes parasite differentiation into the quiescent stumpy form. This function is promoted by a small nucleolar RNA encoded within the *grumpy* lncRNA. *snoGRUMPY* binds to messenger RNAs of at least two stumpy regulatory genes, promoting their expression. *grumpy* overexpression reduces parasitemia in infected mice. Our analyses suggest that *T. brucei* lncRNAs modulate parasite-host interactions and provide a mechanism by which *grumpy* regulates cell differentiation in trypanosomes.

INTRODUCTION

Trypanosoma brucei is a unicellular kinetoplastid parasite that cycles between an insect (tsetse fly) and a mammalian host. Once *T. brucei* reproduces to a critical density in mammalian blood, a quorum-sensing mechanism is activated, and parasites differentiate into a quiescent, nondividing stumpy form (*1*). This shift in parasite life-style limits parasite population size and extends host survival, making it a promising step for therapeutic intervention. The stumpy form also facilitates transmission to the tsetse fly vector and development into insect procyclic forms (*2*). In *T. brucei*, parasite density is sensed via the stumpy induction factor (SIF) (*1*) and the SIF signaling pathway, which promotes gene expression as well as morphological and metabolic changes associated with the stumpy form (*3*). To date, 43 genes have been shown to function in the SIF signaling pathway, playing roles that range from signal transduction to signal response (*3*). RNA-binding protein 7A (RBP7A) and RBP7B, proteins that are predicted to bind RNA, are necessary for SIF-induced stumpy formation (*3*). RBP7A/B null mutant parasites are unresponsive to the SIF signal and are unable to differentiate into stumpy forms. RBP7 genes are therefore crucial regulators of parasite differentiation, yet their mode of action and target genes are unknown (*4*, *5*).

In mammals, transcription is pervasive, with 80% of genomic DNA being actively transcribed in different cell types or physiological situations, yet only a small fraction (less than 3%) of the genome encodes protein-coding genes (*6*). The vast majority of the mammalian genome encodes noncoding DNA sequences such as introns, pseudo-genes, transposons, telomeres, or noncoding RNAs (ncRNAs). ncRNAs are divided into two groups, small ncRNAs [<200 nucleotides (nt)] and long ncRNAs (lncRNAs) (>200 nt). Small ncRNAs include RNAs that regulate splicing, small nuclear RNAs (snRNAs), tRNAs that help decode mRNA sequences into proteins, small interfering RNAs and microRNAs that regulate mRNA stability and translation, small nucleolar RNAs (snoRNAs) that regulate ribosomal RNA (rRNA) processing and modification, and others (*7*). *T. brucei* has all of the canonical small ncRNAs described above except for microRNAs (*8*).

In eukaryotes, lncRNAs resemble mRNAs as they harbor a 5′ cap and are polyadenylated (*9*). Their abundance is comparable to that of protein-coding genes (*10*). lncRNAs function in many cellular pathways (*11*–*13*), including cell differentiation (*14*, *15*). This family of RNAs can regulate cell fate choice by either promoting or inhibiting differentiation. For example, skin stem cells express the ANCR (antidifferentiation ncRNA) and TINCR (terminal differentiation ncRNA) lncRNAs, which function antagonistically. While ANCR suppresses the epidermal differentiation pathway and maintains the stem cell compartment, TINCR promotes epidermal terminal differentiation (*16*, *17*). lncRNAs also regulate antigenic variation of *Plasmodium falciparum* (*18*, *19*) and the host cell response during *Toxoplasma gondii* infection (*20*), indicating that they are important players in parasite infection.

To date, only 95 putative lncRNA genes have been annotated in *T. brucei*, all with unknown functions (*21*). Compared to 9598 *T. brucei* protein-coding genes (*22*), this small number prompted us to analyze the noncoding repertoire of *T. brucei*. Here, we find 1428 additional *T. brucei* lncRNAs, including *grumpy*. We uncover a mechanism whereby *grumpy* encodes a snoRNA that regulates parasite differentiation.

[1]Instituto de Medicina Molecular–Joao Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal. [2]The Mina and Everard Goodman Faculty of Life Sciences and Advanced Materials and Nanotechnology Institute, Bar-Ilan University, Ramat-Gan 52900, Israel. [3]Laboratory of RNA Biology, International Institute of Molecular and Cell Biology, Warsaw, Poland. [4]Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Warsaw, Poland. [5]Instituto de Parasitologia y Biomedicina 'Lopez-Neyra,' IPBLN-CSIC, Parque Tecnológico de Ciencias de la Salud, Avda. del Conocimiento 17, 18016 Armilla, Granada, Spain. [6]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain. [7]Universitat Pompeu Fabra (UPF), Barcelona, Spain.
*Corresponding author. Email: gueganfabien@gmail.com (F.G.); lmf@medicina.ulisboa.pt (L.M.F.)
†Present address: Department of Chemical and Structural Biology, The Weizmann Institute of Science, Rehovot 760001, Israel.
‡Present address: Institute of Molecular Biology and Institute of Developmental Biology and Neurobiology, Johannes Gutenberg Universität, 55128 Mainz, Germany.
§Present address: SGS Portugal S.A., Polo Tecnológico de Lisboa, R. Cesina Adães Bermudes Lote 11 N° 1, 1600-604 Lisboa, Portugal.

RESULTS

We used a combination of strand-specific and paired-end RNA sequencing (RNA-seq), in silico analysis, and database integration to reannotate the lncRNA gene repertoire of *T. brucei* (Fig. 1A and figs. S1 to S5). We identified 1428 previously uncharacterized transcripts that are longer than 200 nt, have a low coding potential score, few ribosomal interactions, and which do not encode any unique peptides (tables S1 and S2 and figs. S2 to S5). These putative lncRNAs are scattered in a mostly intergenic fashion throughout the 11 chromosomes of the *T. brucei* genome (fig. S6). They are shorter, less expressed, and less GC-rich than *T. brucei* mRNAs (fig. S7), but they otherwise harbor regular mRNA trans-splicing/polyadenylation motifs (fig. S8). We confirmed the full-length sequences from the polyadenylate [poly(A)] tail to the splice leader sequence of 1016 of 1428 lncRNAs using Nanopore direct RNA-seq (Fig. 1B, table S3, and fig. S9). Nanopore sequencing also revealed that poly(A) tails of lncRNA tend to be longer (most frequently ~112 nt) than the poly(A) tails of other transcripts (~95 nt) (Fig. 1C). This opens up an avenue for future investigation of polyadenylation and processing of lncRNAs in trypanosomes. We detected some of these transcripts either in the nucleus or in the cytoplasm at various stages of the *T. brucei* life cycle (Fig. 1D). Notably, 25% of the newly identified lncRNA are differentially expressed between mammalian bloodstream form (BSF) and insect procyclic form (fig. S10 and table S4), a greater percentage than protein-coding transcripts (only 16% respond to parasite transition). The *T. brucei* lncRNA gene repertoire is substantial (11% of total genes) and shows a highly dynamic expression pattern during the parasite life cycle.

We analyzed output from an RNA interference (RNAi) screen (*23*) to test our hypothesis that lncRNAs are involved in parasite differentiation. We found a total of 399 lncRNA genes that appear to be required for differentiation to occur (Fig. 1E and table S5), consistent with our expectation that *T. brucei* lncRNAs regulate parasite transition and adaptation between mammalian and insect vector hosts. In other eukaryotes, lncRNAs have been reported to regulate cell differentiation by modulating the expression of their neighboring genes (*14*). We found 19 *T. brucei* lncRNAs genes located immediately upstream or downstream of 18 of the 43 SIF pathway genes (table S6). The lncRNA Ksplice-3137a, which we named *grumpy* (for "regulator of growth and stump formation), is located upstream of RBP7A and RBP7B (Fig. 1A), which are both required for SIF-induced stumpy formation (*3*). *grumpy*'s pattern of expression is similar to that of RBP7, which is transcribed both in the BSF and the procyclic forms (PCF) of *T. brucei* (Fig. 2A). However, unlike RBP7 transcripts, *grumpy* does not stably interact with *T. brucei* ribosomes (Fig. 2A) and does not produce detectable peptides (table S2).

To further characterize the *grumpy* transcript, we used a circular reverse transcription polymerase chain reaction (cRT-PCR) assay, in which *T. brucei* RNAs are circularized via their 5′ to 3′ end junctions, amplified, and sequenced. We used gene-specific primers to confirm that *grumpy* is a trans-spliced and polyadenylated lncRNA transcript expressed as, at least, five different isoforms, including the smallest [359 base pairs (bp)], the major (397 bp), and the longest forms (432 bp) (Fig. 2B). These findings are consistent with the Ksplice in silico analysis, which revealed 1 splice-acceptor site and 10 alternative polyadenylation sites for *grumpy* (Fig. 2B). We also used Nanopore direct RNA-seq to confirm the full-length sequence of grumpy lncRNA in BSF and procyclic form (Fig. 2C).

RT quantitative PCR (qPCR) analysis reveals that RBP7 family genes and *grumpy* lncRNA expression differ during parasite differentiation. These different gene expression behaviors are independent of whether transition from the slender to stumpy form is induced by a chemical compound [pCPT-cAMP (chlorophenylthio)adenosine 3′,5′-cyclic monophosphate sodium salt)] or produced by cell density (Fig. 3A and fig. S11). While RBP7B mRNA decreases during stumpy formation, RBP7A and *grumpy* are up-regulated. Moreover, RNA–fluorescence in situ hybridization (FISH) analysis revealed changes in the subcellular localization of *grumpy* during stumpy formation (Fig. 3B). Whereas in slender forms, *grumpy* localizes in three distinct nuclear foci (one in the nucleolus and two in the nucleoplasm), in stumpy forms, *grumpy* localizes in a single nucleolar focus (Fig. 3B). Moreover, we observed that the signal intensity of nucleolar grumpy increases with time during stumpy formation (Fig. 3C). Contrary to our initial hypothesis, these changes in subcellular localization suggest that *grumpy* may act through a trans-acting mechanism, which means that *grumpy* acts in a nuclear localization different from its transcription locus.

Subcellular localization of lncRNAs is often regulated through recognition of RNA motifs and interactions with RNA binding proteins (*24*, *25*). To study subcellular localization changes of *grumpy* during stumpy formation, pulldowns were performed in vitro using the stumpy form of nuclear lysates. Mass spectrometry (MS) analysis revealed seven RNA binding proteins that bind significantly to *grumpy* (Fig. 4A and table S7). The two proteins that showed the highest levels of association with *grumpy* are the heterogeneous nuclear ribonucleoprotein F/H (hnRNP F/H) and the double RNA binding domain protein 3 (DRBD3). hnRNP F/H and DRBD3 are two major factors that regulate RNA splicing and polyadenylation in *T. brucei* (*26*, *27*). hnRNP F/H is a nuclear factor highly expressed in BSFs that regulates both mRNA stability and trans-splicing (*26*). hnRNP F/H can also repress trans-splicing of specific mRNAs (*26*). DRBD3 can be found both in the cytoplasm and nucleus of the parasite where it performs distinct functions (*27*). DRBD3 acts as stabilizer of a subset of mRNAs encoding for developmentally regulated membrane proteins (*28*), ribosomal proteins, and translation factors (*29*). DRBD3 moves to the nucleus in response to environmental cues (*29*). We validated the interaction of DRBD3 with *grumpy* in vivo using RNA immunoprecipitation and qPCR assays (or RIP-qPCR) (Fig. 4B).

Given that *grumpy* is recognized by and interacts with splicing factors involved both in trans-splicing and RNA stability, we examined whether *grumpy* RNA is further processed and matured in stumpy-stage parasites. A total of 88 nt of *grumpy* encode for a snoRNA gene (snoRNA), originally named TB10CS'2C1 and which we call here *snoGRUMPY* (fig. S12) (*30*, *31*). Northern blot analysis using an antisense RNA probe targeting the snoRNA sequence reveals (i) that the *grumpy* lncRNA is processed into *snoGRUMPY* and (ii) that *snoGRUMPY* is more abundant in stumpy forms compared to the other life forms tested (Fig. 4C). Fractionation of whole-cell lysates in 10 to 30% sucrose gradients shows that *snoGRUMPY* is associated with smaller RNPs (fractions 1 to 3) and that *grumpy* lncRNA fractionates in a slightly heavier fraction (fraction 5) (fig. S13), suggesting that the lncRNA associates with protein(s) in addition to the core C/D small nucleolar RNP (snoRNP) that associates with *snoGRUMPY* (fig. S13).

snoGRUMPY maturation appears to depend on hnRNP F/H protein as parasites depleted of this splicing factor show lower levels of

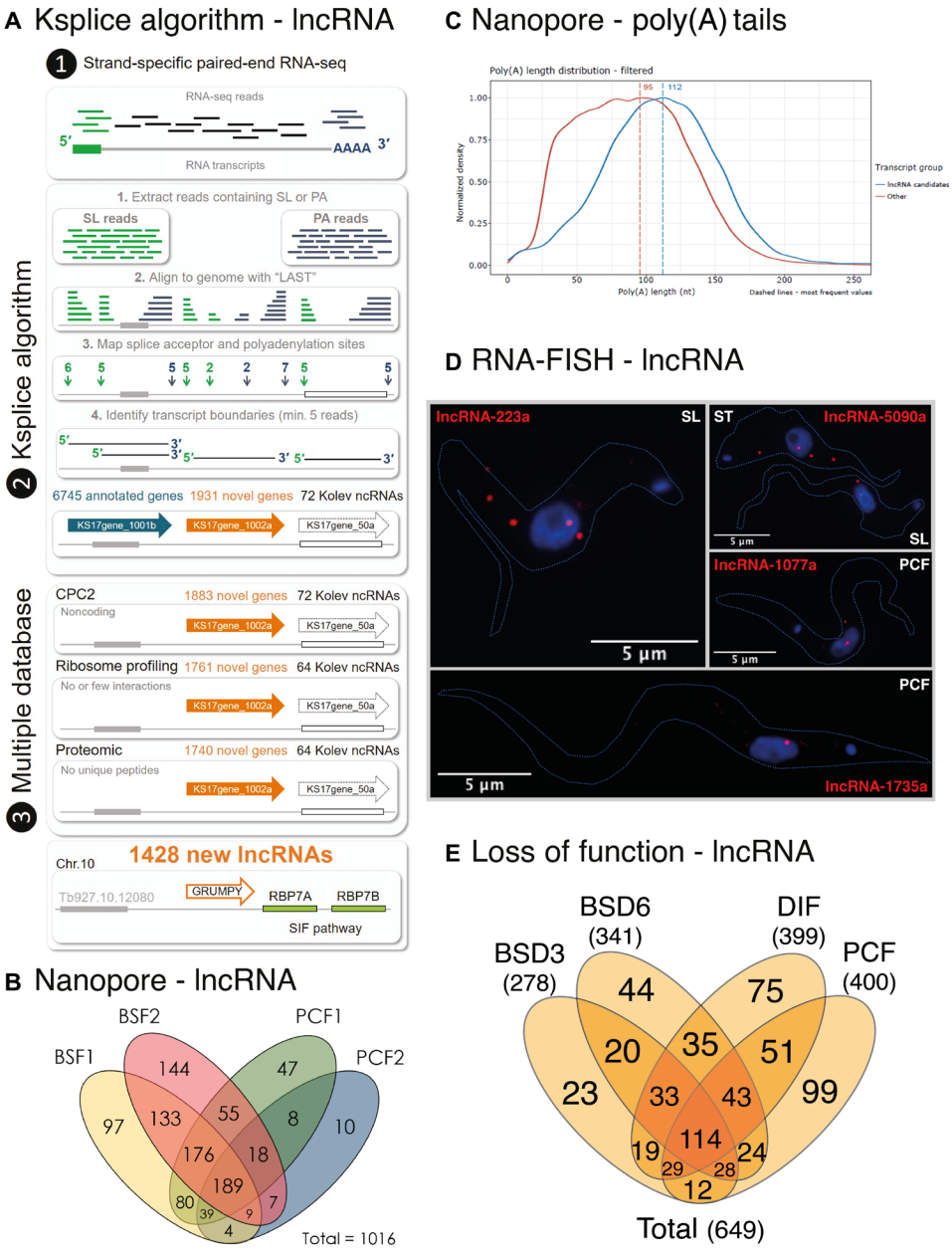


Fig. 1. Identification of 1428 lncRNAs in *T. brucei*. (**A**) Pipeline used for the identification of lncRNAs genes in *T. brucei*. (1) Strand-specific and paired-end RNA-seq. (2) Ksplice identified putative genes whose transcripts contained an SL sequence (SL) and a poly(A) tail (PA) at the extremities. Ksplice used LAST (*48*) to map RNA-seq reads to the *T. brucei* genome. (3) The noncoding nature of the putative lncRNAs was predicted from a low coding potential calculator (CPC) score, poor association with ribosomes, and no detectable peptides. *grumpy* lncRNA is intergenic and immediately upstream of RBP7 genes, previously shown to be involved in the SIF-dependent pathway. (**B**) The number of full-length lncRNAs [from the SL sequence to the poly(A) tail] sequenced with Nanopore in four RNA samples: two from BSF parasites (BSF) and two from procyclic forms (PCF). (**C**) Distribution of poly(A) tail lengths in lncRNA candidates versus other transcripts. (**D**) Subcellular localization of Ksplice lncRNA genes in slender forms (SL), stumpy forms (ST), and PCF of *T. brucei*, using RNA-FISH. (**E**) The number of Ksplice lncRNA genes that cause loss of parasite fitness upon down-regulation by RNAi [extracted from RIT-seq analysis (*23*)]. RNAi was induced in BSFs for 3 days (BSD3), 278 lncRNAs; in BSFs for 6 days (BSD6), 341 lncRNAs; during in vitro parasite differentiation from BSF to insect procyclic forms (DIF), 400 lncRNAs; in PCF, 402 lncRNAs. The total number of lncRNA genes essential for parasite fitness in this screen was 649.

snoGRUMPY (Fig. 4D). DRBD3 depletion did not affect *snoGRUMPY* levels (fig. S14A). *snoGRUMPY* localizes mainly in the nucleolus of the cells (Fig. 4E), but it can also be found associated with the Cajal body (fig. S14B). *snoGRUMPY*, similar to *grumpy*, is up-regulated during stumpy formation induced both by cell density (Fig. 4F) and by CPT (fig. S15). *snoGRUMPY* is up-regulated two times more than *grumpy* during stumpy formation, suggesting that *snoGRUMPY* is processed from *grumpy* (Fig. 4F).

To identify the function of *grumpy* and *snoGRUMPY*, we started by overexpressing the full-length *grumpy* lncRNA (fig. S12). Upon

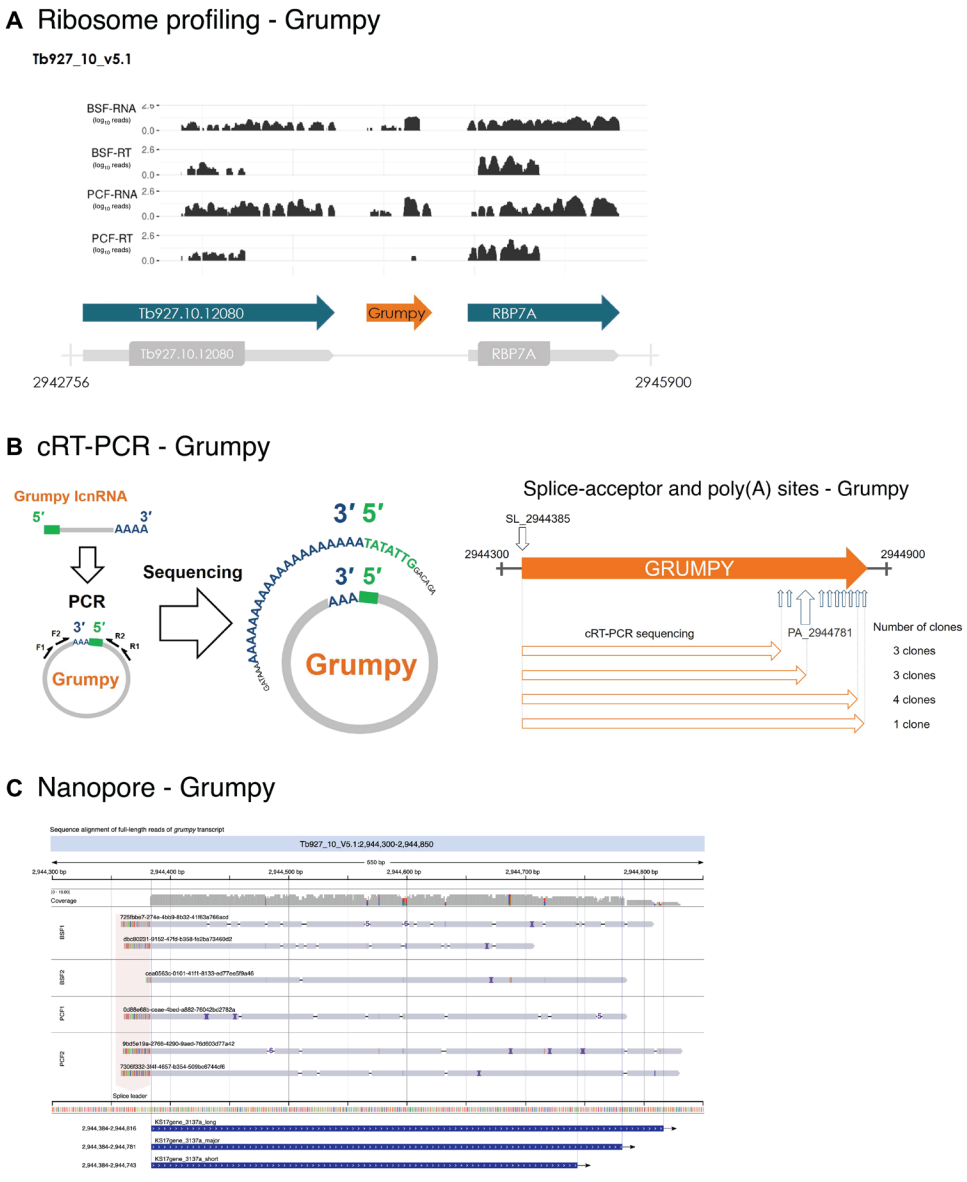


Fig. 2. *grumpy* is not associated with ribosomes and it has alternative 3′ ends. (**A**) Ribosome association of *grumpy* and its neighboring genes was assessed by analyzing previously published ribosome profiling datasets: Mapping of RNA-seq reads from BSFs (BSF-RNA) or PCF (PCF-RNA); mapping of ribosome profiling reads from BSFs (BSF-RT) or procyclic forms (PCF-RT). (**B**) Sequencing and accurate mapping of the 5′ and 3′ ends of the *grumpy* lncRNA using cRT-PCR. Black outlined arrows show the position of splice-acceptor site (SL) and polyadenylation sites (PA) identified with our Ksplice algorithm in the *grumpy* gene locus. Orange outlined arrows show the *grumpy* transcript isoforms that we sequenced using cRT-PCR and the number of clones sequenced for each isoform. (**C**) Identification of complete sequences of *grumpy* with Nanopore direct RNA-seq. Sequence alignment of identified grumpy reads [5′ splice-acceptor site were included, while 3′ poly(A) tail sequences were omitted for clarity]. The three main isoforms of grumpy (KS17gene_3137a) are annotated (in blue) below the reads alignment.

induction, the levels of *grumpy* transcripts increased three- to fourfold, and that of *snoGRUMPY* increased eight- to ninefold (fig. S16), while the transcript levels of RBP7A and RBP7B remained unchanged (Fig. 5A). FISH probes spanning the full-length *grumpy* (fig. S12) show that the overexpressed RNAs retained the original nucleolar localization (Fig. 5B). We observed that exogenous expression of *grumpy* repressed *T. brucei* growth and increased life span in vitro (Fig. 5, C and D). We asked whether this reduction in parasite growth could be explained by a higher proportion of the stumpy forms in culture. Stumpy formation occurs only at high parasite density via the SIF-dependent quorum-sensing mechanism and can be quantified using flow cytometry (*32*, *33*) by measuring the fraction of transgenic parasites expressing the fluorescent stumpy marker GFP::PAD1 (green fluorescent protein::protein associated with differentiation 1) 3′UTR. After 2 days in culture, 60% of the *grumpy*-overexpressing parasites were in the stumpy form, compared to 7% in the parental line cultured for the same time period, suggesting that *grumpy* accelerates stumpy formation (Fig. 5E). This phenotype is not observed upon overexpression of other lncRNAs (fig. S17). *grumpy* overexpression also led to a lower parasite density

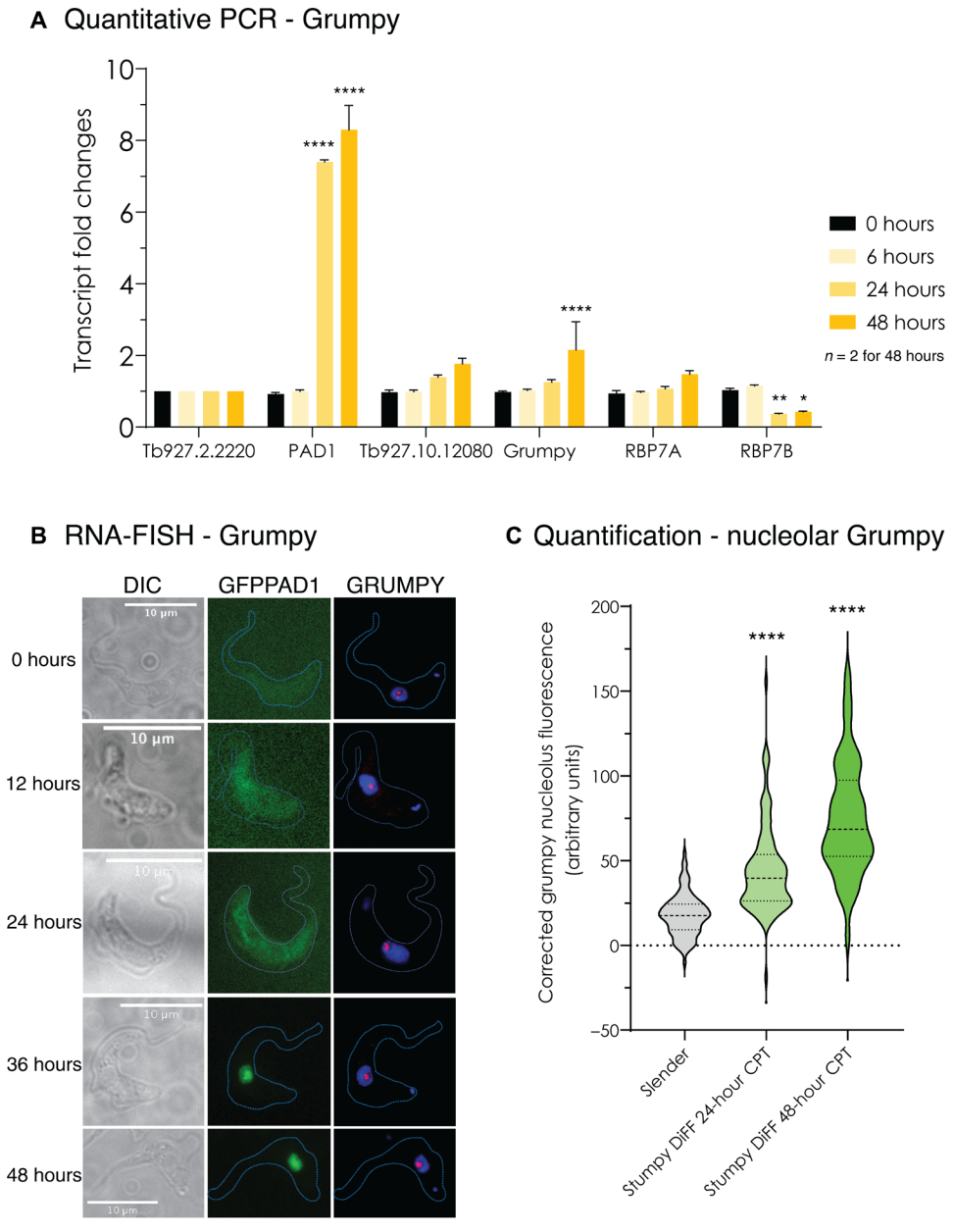


Fig. 3. Dynamic subnuclear localization of *grumpy* during parasite differentiation. (**A**) Transcript-level changes during the transition from slender to stumpy forms, measured by RT-qPCR. Stumpy formation was induced by pCPT-cAMP (C3912, Sigma-Aldrich). Tb927.2.2220 is used as a control to normalize transcript levels (*73*). PAD1 and GFP genes are used as controls to estimate parasite differentiation into stumpy forms. Tb927.10.12080 is the gene upstream of *grumpy*. Results are shown as means (SEM, $n = 3$ except for the 48-hour time point, $n = 2$). Dunnett's multiple comparisons test was used for statistical analysis using the 0-hour time point as the control for comparison (adjusted *P* values: $*P < 0.05$; $**P < 0.01$; $****P < 0.0001$). (**B**) Subcellular localization of *grumpy* during the transition from slender to stumpy forms using RNA-FISH. Time points (0, 12, 24, 36, and 48 hours) of parasite differentiation after addition of the pCPT-cAMP stimulus to the culture medium. DIC, differential interference contrast microscopy image of *T. brucei*; GFP, GFP::PAD1 signal expressed in the nucleus of stumpy forms; *GRUMPY*, *grumpy* signal using RNA-FISH (Stellaris probes). (**C**) Quantification of the intensity of the nucleolar *grumpy* RNA-FISH signal corrected with the cytoplasmic signal, during stumpy formation induced by pCPT-cAMP at time 0, 24, and 48 hours. Statistical test: one-way analysis of variance (ANOVA) (Dunnett's multiple comparisons test), adjusted *P* value: $****P < 0.0001$.

($<0.7 \times 10^6$ cells/ml) compared to the parental culture (1.4×10^6 cells/ml) (Fig. 5C). Parasites overexpressing *grumpy* displayed all the hallmarks of being in stumpy form, including PAD1 protein expression at the cell surface (Fig. 5F), arrest at the cell cycle G_0-G_1 phase (Fig. 5G), and preadaptation to differentiate into the insect procyclic stage (fig. S18).

To confirm these results in vivo, we induced mouse infections with parasites overexpressing *grumpy* and measured parasitemia, mouse survival, and stumpy formation, which were compared to infection by the parasite parental line. Mice infected with the parental cell line showed a typical infection profile characterized by successive waves of parasitemia (Fig. 6A) and an average survival of 43 days

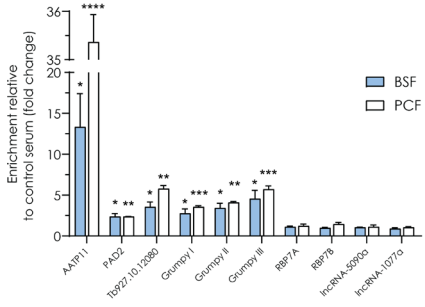
Grumpy-protein pulldown

A

Protein ID	Gene description	Mean difference Grumpy – control lncRNA	*P* value
Tb927.2.3880	Heterogeneous nuclear ribonucleoprotein H/F, putative (HNRNPH)	2.9	1.2×10^{-4}
Tb927.9.8740	RNA binding protein (DRBD3)	2.5	5.7×10^{-6}
Tb927.2.6070	Mitochondrial RNA binding complex 1 subunit (MRB6070)	2.3	1.5×10^{-4}
Tb927.3.3940	RNA binding protein, putative (DRBD11)	1.4	3.6×10^{-3}
Tb927.10.2370	RNA binding protein, putative (LA)	1.4	3.1×10^{-5}
Tb927.3.1590	Mitochondrial RNA binding complex 1 subunit (MRB1590)	1.3	7.6×10^{-4}
Tb927.4.1500	RNA editing–associated helicase 2 (REH2)	1.3	1.7×10^{-2}
Tb927.7.3810	Cold-shock' DNA binding domain containing protein, putative	1.3	9.8×10^{-4}
Tb927.10.16120	Inosine-5-monophosphate dehydrogenase, IMP dehydrogenase	1.2	4.6×10^{-3}

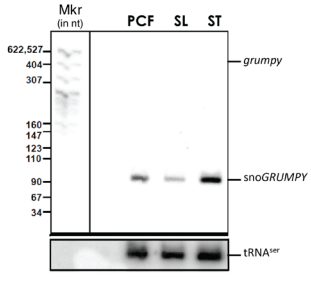
RNA immunoprecipitation (Ab DRBD3)

B

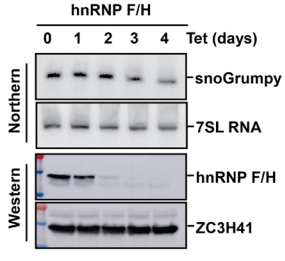


snoRNA grumpy (snoGRUMPY)

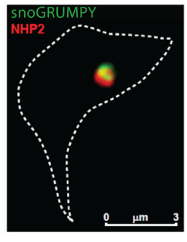
C



D



E



F

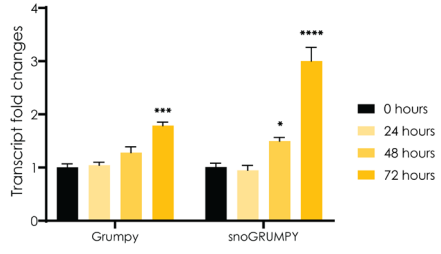


Fig. 4. *grumpy* lncRNA is processed into a snoRNA, called *snoGRUMPY*. (**A**) Protein partners of *grumpy* lncRNA identified using in vitro RNA-protein pulldown assay. Results are shown as the mean difference between pulldown assays performed with *grumpy* lncRNA and control lncRNA-5090a. Note that mitochondria proteins found in this in vitro RNA-protein pulldown assay could come from mitochondria contamination (see Materials and Methods). (**B**) RIP assay using anti-DRBD3 antiserum in BSFs and procyclic forms. Results are shown as the means (SEM, $n = 3$) and compared to RIP using control serum. Statistical test: two-sided *t* test, $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$, and $^{****}P < 0.00005$. (**C** to **F**) snoRNA encoded by *grumpy* lncRNA, named *snoGRUMPY*. (C) Northern blot analysis using probe against *snoGRUMPY* in procyclic forms, slender forms, and stumpy forms. $tRNA^{ser}$ serves as a loading control. The first lane shows the molecular marker (Mkr) and its different sizes annotated in nucleotides. (D) Northern blot analysis using probe against *snoGRUMPY* in RNAi hnRNP F/H cell line (in procyclic forms). 7SL RNA serves as a loading control for the Northern blot. Bottom: Western blot analysis showing the depletion of hnRNP F/H protein. ZC3H41 serves as a loading control for Western blot. (E) RNA-FISH analysis showing partial colocalization between *snoGRUMPY* and the nucleolar marker NHP2 (in procyclic forms). (F) *grumpy* and *snoGRUMPY* transcript level measured by RT-qPCR during the transition from slender to stumpy forms induced by the SIF signal (cell density). Tb927.2.2220 is used as a control to normalize transcript levels. Results are shown as means (SEM, $n = 3$), statistical test: two-way ANOVA (Dunnett's multiple comparisons test), adjusted *P* values: $^{*}P < 0.05$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

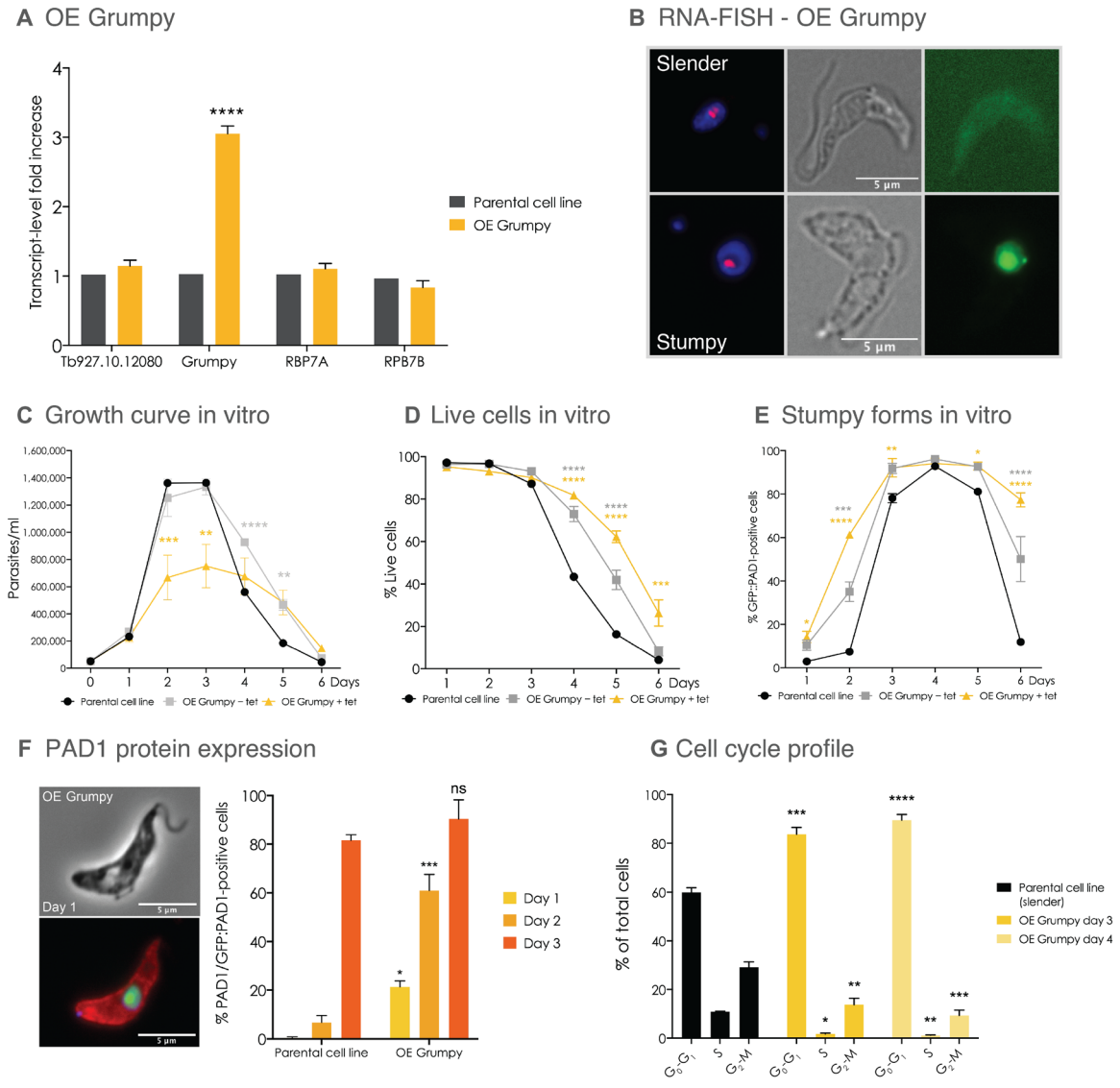


Fig. 5. *grumpy* overexpression promotes premature parasite differentiation. (**A**) Transcript levels measured by RT-qPCR of *grumpy* and its neighboring genes in parental cell line (black bars) and in *grumpy*-overexpressing (OE) cell line, 24 hours after tetracycline induction (yellow bars). Changes in transcript levels were measured by normalizing transcript level to a control gene (Tb927.10.12970) and to the parental cell line. (**B**) Subcellular localization of *grumpy* after overexpression using RNA-FISH. Left to right: *GRUMPY*, *grumpy* signal using RNA-FISH; phase-contrast signal of *T. brucei* parasite; GFP, GFP::PAD1 signal expressed in the nucleus of stumpy forms. (**C** to **F**) After inducing *grumpy* overexpression, we measured (C) parasite growth for 6 days (without passage), (D) percentage of live cells measured by fluorescence-activated cell sorting (FACS) of propidium iodide–stained cells, (E) percentage of GFP::PAD1-positive parasites (stumpy forms) measured by FACS, (F) percentage of parasites expressing both GFP::PAD1 and endogenous PAD1 protein that are measured by microscopy and image quantification. Microscopy picture (on the left) shows an example of a parasite expressing both GFP::PAD1 in the nucleus (in green) and the endogenous PAD1 protein at the cell surface (in red). Parasite DNA stained with 4′,6-diamidino-2-phenylindole (DAPI) (in blue). (**G**) Cell cycle profile of parental cell line (slender forms) and *grumpy*-overexpressing parasites at days 3 and 4 of in vitro culture without passage. All results are shown as means (SEM, $n = 3$). Multiple *t* tests (A) and Sidak's multiple comparisons test (C to G) are used for statistical analysis using the parental cell line as the control for comparison (adjusted *P* values: **P* <0.05; ***P* < 0.01; ****P* <0.001; *****P* < 0.0001).

(Fig. 6B) (*34*). By contrast, mice infected with parasites overexpressing *grumpy* showed no detectable parasitemia and did not die from the infection (>100 days) (Fig. 6, A and B). When *grumpy* overexpression was induced 4 days after infection, the parasites succeeded in establishing an infection (Fig. 6A), with three of four mice dying from the infection and mouse survival time increasing from approximately 43 to 72 days (Fig. 6B). Thus, *grumpy* overexpression substantially reduces parasite virulence in mice.

Our in vivo analysis also recapitulated in vitro observations with respect to stumpy forms and density. Wild-type parasites started differentiating into stumpy forms (>20% stumpy forms in the blood) only at high parasitemia (>1.5×10^7 parasites/ml), whereas *grumpy*-overexpressing parasites differentiated into stumpy forms when parasitemia was as low as 1.1×10^6 parasites/ml (Fig. 6C). These results support the notion that *grumpy* overexpression triggers premature *T. brucei* differentiation into

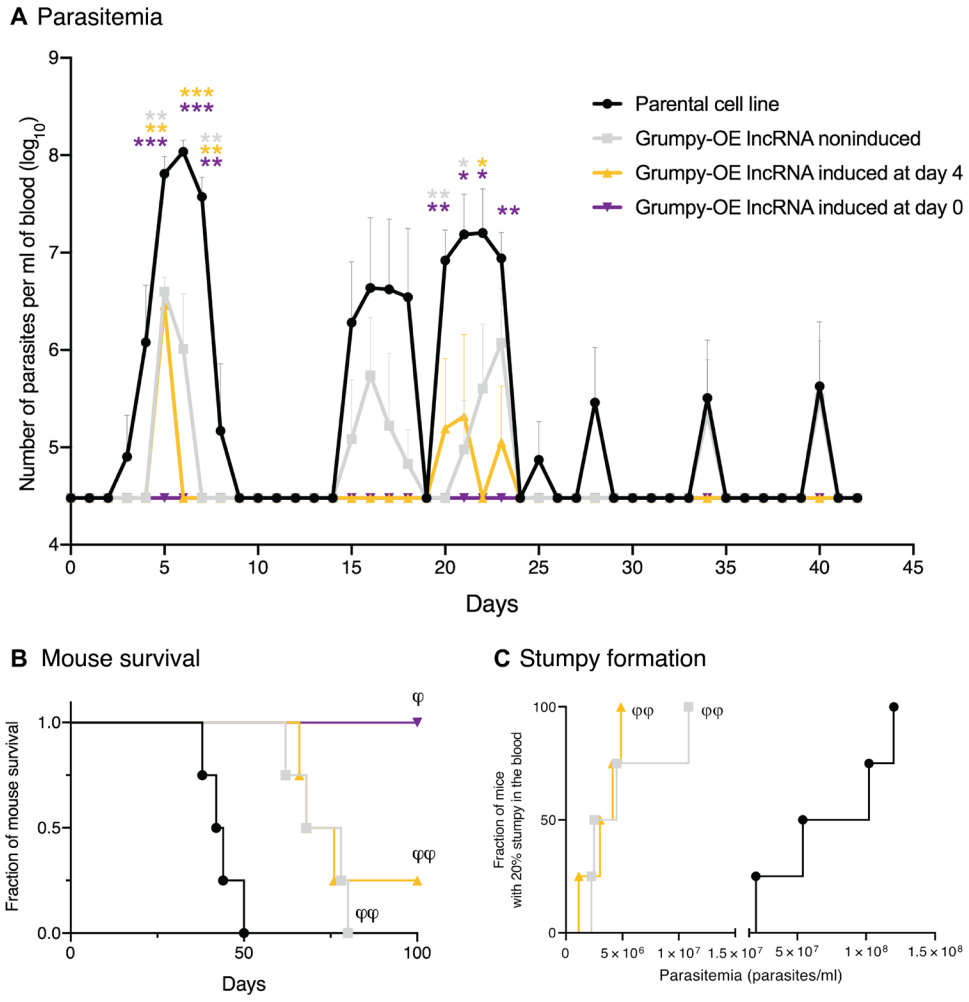


Fig. 6. Overexpression of *grumpy* promotes premature differentiation into stumpy forms in vivo and prolongs mouse survival. (**A**) Parasitemia in mice infected with the parental cell line (black line) or with a *grumpy*-overexpressing cell line. *grumpy* overexpression was either not induced or induced by adding doxycycline to drinking water either at day 0 (purple curve) or day 4 (yellow curve) of infection. Results are shown as the means (SEM, $n = 4$). Dunnett's multiple comparisons test was used for statistical analysis using the parental cell line as the control (adjusted P values: $^{**}P < 0.01$; $^{***}P < 0.001$). (**B**) Mouse survival rates according to the type of infection described in (A). Log-rank (Mantel-Cox) test for comparisons of Kaplan-Meier survival curves indicates a significant increase in the survival rates in mice infected with *grumpy*-overexpressing cell line parasites compared to mice infected with the parental cell line. φφ$P = 0.0067$ and φ$P = 0.0177$. (**C**) Fraction of mice with at least 20% of parasites GFP::PAD1 positive (arbitrary threshold that we defined to compare the initiation of stumpy formation between the different conditions) in the blood as a function of parasitemia. Log-rank (Mantel-Cox) test for comparisons of Kaplan-Meier curves indicates significant premature parasite differentiation into stumpy forms in mice infected with *grumpy*-overexpressing cell line parasites compared to mice infected with the parental cell line. φφ$P = 0.0067$ and φ$P = 0.0177$.

stumpy forms, which is associated with a reduction in parasite virulence.

Given that *snoGRUMPY* is processed from *grumpy*, we next studied the effect of overexpressing only *snoGRUMPY* in parasite differentiation (fig. S12). *snoGRUMPY* was overexpressed ~4 times from an exogenous genomic location (mini-chromosome) (Fig. 7A). Overexpressed *snoGRUMPY* localizes in the nucleolus (Fig. 7B), reduces parasite growth, and promotes differentiation into stumpy forms (Figs. 7C and 7D), recapitulating the *grumpy* lncRNA overexpression phenotype. Our results show that premature parasite differentiation into stumpy forms induced by *grumpy* lncRNA is likely mediated by the action of *snoGRUMPY*.

Next, we tried to study the function of *grumpy* or *snoGRUMPY* using loss-of-function strategies. All our attempts to knock out *grumpy* lncRNA were unsuccessful; therefore, we decided to use RNAi to strive for depletion of *grumpy* and/or *snoGRUMPY* (fig. S12). Although our RNAi targeting full-length *grumpy* successfully depleted ~80% of *grumpy* lncRNA at its 5′ end, the level of *snoGRUMPY* remained unchanged (fig. S19). Similarly, an RNAi construct that specifically targets the *snoGRUMPY* sequence did not succeed in reducing the levels of the snoRNA (figs. S12 and S20). Therefore, it appears that *snoGRUMPY* is protected from degradation, probably via interaction with snoRNPs (*35*, *36*).

Next, we tried gymnotic delivery of locked nucleic acid (LNA) gapmeRs (fig. S21). While this strategy failed to deplete *grumpy*, this gapmeR induced the cleavage of the *grumpy* lncRNA at the LNA target site and promoted the release of *snoGRUMPY* (figs. S12 and S22A), which accumulated inside the nucleolus (fig. S22B), resulting

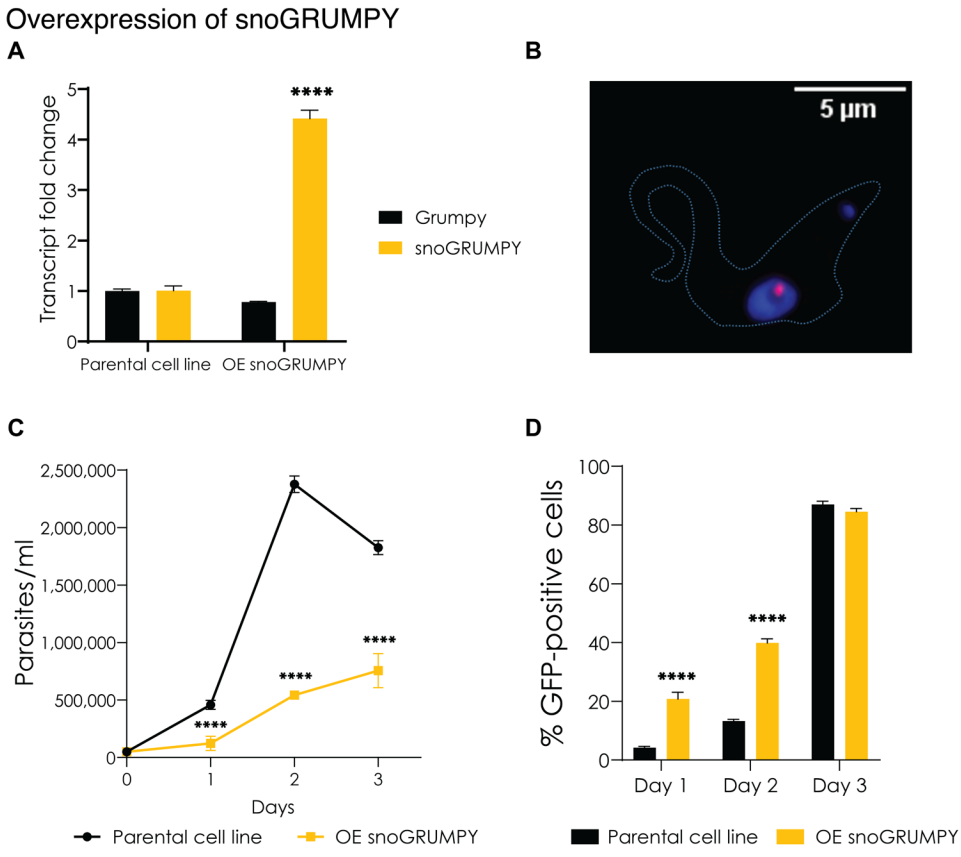


Fig. 7. Overexpression of *snoGRUMPY* promotes premature differentiation into stumpy forms. (**A**) Transcript levels measured by RT-qPCR of *grumpy* (black bars) and *snoGRUMPY* (yellow bars) in parasites overexpressing *snoGRUMPY*. Changes in transcript levels were measured by normalizing transcript level to a control gene (Tb927.2.2220) and to the parental cell line. Results are shown as the means (SEM, $n = 3$), statistical test: two-way ANOVA (Sidak's multiple comparisons test), adjusted *P* value: ****$P < 0.0001$. (**B**) Subcellular localization of *snoGRUMPY* after overexpression using RNA-FISH (red signal). Parasite DNA is stained using DAPI (blue signal). (**C**) Growth in parasites overexpressing *snoGRUMPY* (without passage) for 3 days. Results are shown as the means (SEM, $n = 3$), statistical test: two-way ANOVA (Sidak's multiple comparisons test), adjusted *P* value: ****$P < 0.0001$. (**D**) Percentage of GFP::PAD1-positive parasites (stumpy forms) measured by FACS after inducing *snoGRUMPY* overexpression. Results are shown as the means (SEM, $n = 3$), statistical test: two-way ANOVA (Sidak's multiple comparisons test), adjusted *P* value: ****$P < 0.0001$.

in a strong phenotype with premature parasite differentiation into stumpy forms (fig. S22, C and D). Thus, while our LNA gapmeR strategy failed as a loss-of-function strategy, it unexpectedly resulted in a gain of function of *snoGRUMPY*, confirming once again the phenotype of *snoGRUMPY* overexpression. This analysis confirmed that *snoGRUMPY* is a key player that causes premature parasite differentiation into stumpy forms upon *grumpy* lncRNA overexpression.

snoGRUMPY belongs to the C/D box snoRNA family, which guides 2′-*O*-methylation (Nm) in rRNA (*30*). To test whether *snoGRUMPY* plays a role in rRNA methylation, we used RiboMeth-seq (ribose methylations sequencing), which we previously showed that it can detect changes in Nm modification during the developmental cycle of the parasite (*8*, *37*). Unexpectedly, analysis of the methylation profile of rRNA upon overexpression of *grumpy* lncRNA using RiboMeth-seq revealed no changes in the methylation site guided by *snoGRUMPY* (Am622). No changes were observed during the slender to stumpy transition or during overexpression of *grumpy* lncRNA (table S8). We conclude that the premature parasite differentiation phenotype induced by *grumpy* is not caused by increased methylation of rRNA mediated by *snoGRUMPY*.

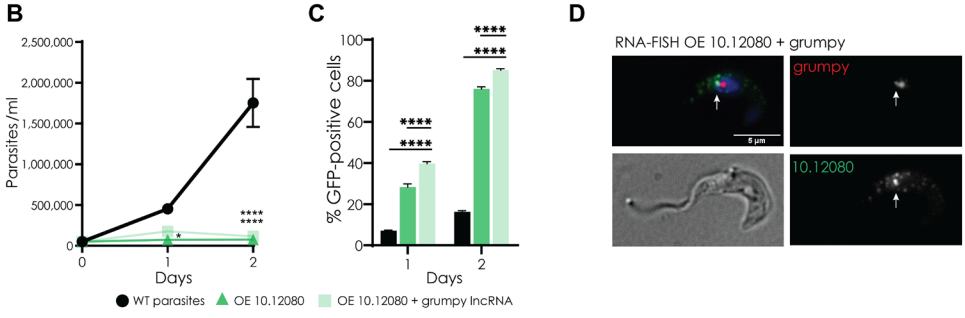
snoRNAs have also recently been shown to also act on mRNA and regulate its translation efficacy (TE) in other eukaryotes (*38*). More recently, small antisense RNAs have been shown to either inhibit or enhance the translation of target mRNA in *T. brucei* (*8*). Next, we performed an in vivo cross-linking experiment to map the interaction between *snoGRUMPY* and its mRNA targets (*8*). To obtain enough material for analysis, this experiment was performed in procyclic forms. We found that *snoGRUMPY* interacts with 36 mRNAs and 22 other ncRNAs (including rRNA, tRNA, and others snoRNAs) (table S9). The two most remarkable mRNA targets are two hypothetical proteins: Tb927.8.2860 (HYP5) and Tb927.10.12080 (Fig. 8A, fig. S23, and table S9). HYP5 has previously been identified to be involved in driving stumpy formation (*5*). HYP5 has probable functions in ubiquitination and could be involved in protein degradation in slender forms. Tb927.10.12080 is the gene located just immediately upstream of *grumpy* lncRNA locus, but its function is unknown. When we overexpressed the coding sequence of Tb927.10.12080 alone or together with *grumpy* lncRNA, parasites showed a strong premature differentiation to stumpy forms (Fig. 8B and fig. S24). Parasite growth almost completely stops after 24 hours, and almost all parasites expressed GFP::PAD1 after 48 hours of Tb927.10.12080

mRNA targets of snoGRUMPY

A

Gene ID	Gene description	Total number of reads
Tb927.10.12080	Hypothetical protein, conserved	9
Tb927.8.2860	Hypothetical protein, conserved (HYP5)	6
Tb927.11.18710	Hypothetical protein, conserved	4
Tb927.8.2861	SRP RNA, 7SL	4
Tb927.11.13120	PUB domain containing protein, putative	3
Tb927.9.11470	60S ribosomal protein L29	3
Tb927.7.7480	Trans-sialidase, putative	2
Tb927.11.16690	Voltage-gated chloride channel–like 3	2
Tb927.11.6890	DNA repair and recombination helicase protein PIF1	1
Tb927.6.4440	RNA binding protein 42 (RNA binding motif protein 42)	1
…	…	…

Function of snoGRUMPY's mRNA target



Regulation of snoGRUMPY's mRNA target proteins

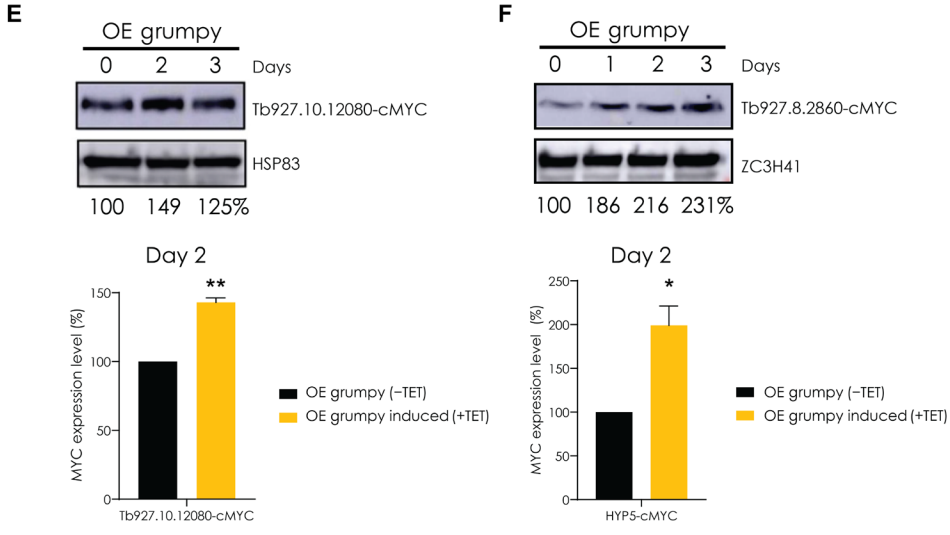


Fig. 8. *snoGRUMPY* interacts with stumpy-related transcripts and regulates their expression. (**A**) Top 10 mRNA targets of *snoGRUMPY* identified using in vivo psoralen ultraviolet (UV) cross-linking. Results are shown as the total number of reads sequenced that match the specific target gene. (**B** to **F**) Tb927.10.12080 is an mRNA target of *snoGRUMPY*. (B) Growth curve of parasite overexpressing Tb927.10.12080 gene alone (light blue) or together with *grumpy* (dark blue). Results are shown as means (SEM, $n = 3$), statistical test: two-way ANOVA (Dunnett's multiple comparisons test), adjusted P values: *$P < 0.05$; ****$P < 0.0001$. WT, wild-type. (C) Percentage of GFP::PAD1-positive parasites (stumpy forms) measured by FACS in parasite overexpressing Tb927.10.12080 gene alone (dark green) or together with *grumpy* (light green). Statistical test: two-way ANOVA (Tukey's multiple comparisons test), adjusted P value: ****$P < 0.0001$. (D) RNA-FISH showing the localization of Tb927.1012080 mRNA (green) and *grumpy* lncRNA (red), in parasite overexpressing Tb927.10.12080 together with *grumpy*. The white arrow shows the colocalization of signal between Tb927.10.12080 mRNA and grumpy lncRNA. (E) Change in Tb927.10.12080 protein expression in *grumpy*-overexpressing cell line is analyzed by Western blot and normalized by the expression of housekeeping protein HSP83 (in procyclic forms). (F) Change in HYP5 (Tb927.8.2860) protein expression in *grumpy*-overexpressing cell line is analyzed by Western blot and normalized by the expression of housekeeping protein HSP83 (in procyclic forms). Statistical test for (E) and (F): two-tailed paired t test (*$P < 0.05$; **$P < 0.01$).

overexpression (Fig. 8C). These data show that the two most remarkable targets of *snoGRUMPY*, HYP5 and Tb927.10.12080 mRNAs, are involved in stumpy formation in *T. brucei*.

RNA-FISH in parasites overexpressing both Tb927.10.12080 and *grumpy* from a single construct shows that *grumpy* partially colocalizes with Tb927.10.12080 mRNA (Fig. 8D, white arrow). This result not only confirms that *grumpy* lncRNA is independent from Tb927.10.12080 mRNA but also shows that these two transcripts can be found associated inside the nucleus, validating our previous cross-linking experiment. Together, these results suggest that *snoGRUMPY* can regulate the metabolism of Tb927.10.12080 mRNA. We tested this hypothesis by overexpressing *grumpy* and following Tb927.10.12080 and HYP5 protein levels. Western blotting showed that both proteins are up-regulated (Fig. 8, E and F), indicating that *snoGRUMPY* interacts with differentiation-related transcripts to promote their stability or translation, ultimately leading to stumpy formation.

DISCUSSION

Of the 1428 lncRNA genes that we have identified in *T. brucei*, 649 have been predicted, via an RNAi screen, to play a role in parasite fitness, including 399 lncRNA genes that appear to be involved in cell differentiation (Fig. 1C). Among them, we identified *grumpy*, an lncRNA located just upstream of RBP7 genes, a family involved in driving stumpy formation in *T. brucei*. We found that *grumpy* is the host gene of *snoGRUMPY*, a C/D box snoRNA. Overexpression of *grumpy* lncRNA and *snoGRUMPY* leads to premature parasite differentiation into stumpy forms and a reduction in parasite virulence in vivo.

Overexpression of *snoGRUMPY* recapitulates the premature parasite differentiation phenotype observed with *grumpy* overexpression. These results indicate that splicing and maturation of *snoGRUMPY* from *grumpy* lncRNA are likely to be the key molecular mechanism that governs parasite differentiation into stumpy forms. The antisense LNA gapmeRs, which lead to the strongest phenotype in this work, also promoted the release of *snoGRUMPY*, thus promoting stumpy formation. We found that, besides the proteins hnRNP F/H and DRBD3, *grumpy* lncRNA also interacts with a protein involved in the SIF signaling pathway, the inosine-5′-monophosphate dehydrogenase (IMPDH; Tb927.10.16120) (Fig. 3A and table S7) (*3*, *39*). Therefore, it is possible that *grumpy* lncRNA also promotes stumpy formation via interaction with IMPDH protein. Reports in other eukaryotes indicate that lncRNA snoRNA host gene can perform different or dual functions depending on whether it is the lncRNA or the intronic snoRNA gene that is acting (*40*, *41*). Such a dual function may also exist for *grumpy*.

In the future, it will be interesting to study differentiation phenotypes associated with a version of *grumpy* carrying a splice site mutation that impairs the release of *snoGRUMPY*. This experiment will not only confirm the role of processed *snoGRUMPY* in stumpy formation but will also provide information about the subcellular localization of full-length *grumpy* and its function.

snoGRUMPY localizes mainly in the nucleolus of *T. brucei*, but it can also be found in the Cajal body associated with the methyltransferase-associated protein (MTAP), which was previously shown to interact with dual-function snoRNAs that guide modifications outside the nucleolus (*42*). C/D box snoRNAs, such as *snoGRUMPY*, usually direct the Nm of rRNA nucleotides. However, we found no evidence here that *snoGRUMPY* acts via this mechanism, since the effect on Nm modification under *grumpy* overexpression was only marginal. Instead, we found that *snoGRUMPY* directly interacts with a small subset of mRNAs involved in driving stumpy formation, including HYP5 (Tb927.8.2860) (*5*) and Tb927.10.12080. We propose that *snoGRUMPY* increases the translation efficiency of those genes either by direct antisense interaction, as described previously (*8*), or by promoting the methylation of its mRNA targets.

We also find that overexpression of Tb927.10.12080 leads to parasite growth arrest and premature transition to stumpy forms, indicating that this gene is important for differentiation. A recent report used single-cell sequencing to reveal that Tb927.10.12080 is highly expressed during gametogenesis and could therefore play an important role in gamete formation in *T. brucei*, another key differentiation process in the life cycle of *T. brucei* (*43*).

LNA treatment by gymnosis promoted the processing of *snoGRUMPY* from *grumpy* lncRNA. This easy-to-use methodology provides a convenient tool for scientists to study gene expression and RNA processing in trypanosomes. Alternatively, LNAs targeting essential parasite genes could be used to treat sleeping sickness.

In this work, we identified a novel molecular mechanism for parasite differentiation into stumpy forms: An lncRNA is processed into a snoRNA, which interacts with stumpy-inducing mRNAs to promote their stability or translation. *grumpy* is therefore the first functional lncRNA found in *T. brucei*. *grumpy*'s role in differentiation may be one example of a more general process used by the parasite to adapt to its environment. Understanding these regulatory processes may open up unexplored possibilities for developing therapeutic strategies to treat sleeping sickness.

MATERIALS AND METHODS

Ethics statement

Male C57BL/6J (6- to 8-week-old) mice were purchased from Charles River Laboratories (Lyon, France). All animal care and experimental procedures were performed according to EU regulations (Directive 2010/63/EU9) and approved by the Animal Ethics Committee of Instituto de Medicina Molecular João Lobo Antunes (AWB2016_19FG_RNA).

T. brucei cell culture

A stumpy reporter cell line with a GFP:PAD1UTR construct (*32*) integrated into the tubulin locus was generated in *T. brucei* Antat1.1e (90:13) strain (*44*). The stumpy reporter cell line was selected for the most intense GFP expression, which occurs in the nucleus in response to quorum-sensing signal. This reporter cell line was used as the genetic background to overexpress the *grumpy* lncRNA, *snoGRUMPY*, or *grumpy* RNAi construct. It was cultivated in HMI-11 at 37°C in 5% CO_2 with G418 (2.5 μg/ml), hygromycin B (5 μg/ml), blasticidin S (5 μg/ml), and phleomycin (2.5 μg/ml). Procyclic forms of *T. brucei* strain 29-13, which carries integrated genes for T7 polymerase and the tetracycline repressor, were grown in SDM-79 (*45*) supplemented with 10% fetal bovine serum in the presence of hygromycin B (50 μg/ml) and G418 (15 μg/ml). Transfected cells were cloned by serial dilution to obtain a clonal population. The cell lines used in this study to silence hnRNP F/H (*26*) and polypyrimidine tract binding protein 1 (PTB1) were as previously described (*27*).

RNA sequencing

T. brucei BSF and procyclic form parasites (strain Lister 427, antigenic type MiTat 1.2, clone 221a), from the PL1S cell line (*46*), were used to generate strand-specific libraries following the manufacturer's instructions (Encore Complete RNA-Seq Library Systems, NuGen) for Illumina next-generation paired-end sequencing. RNA-seq was performed by the Genomics Core Facility, EMBL Heidelberg. The RNA-seq data from this study have been submitted to the European Nucleotide Archive–PRJEB38238.

Reconstruction of *T. brucei* transcriptome

In *T. brucei*, all mature mRNAs are trans-spliced and polyadenylated, which means that all mRNA transcripts start with a conserved spliced-leader (SL) sequence and finish with poly(A) tail sequence (*47*). We hypothesized that any new *T. brucei* transcripts, including ncRNAs, will bear these features. RNA-seq reads were assessed for quality using FastQC. To improve genome mapping, RNA-seq read size was increased, if possible, by merging the paired-end reads using PEAR software (Paired-End reAd mergeR; https://cme.h-its.org/exelixis/web/software/pear/). Merged and forward unmerged reads containing a minimum of 8 bp matching the SL sequence on their 5′ ends were extracted for 5′ splice-acceptor site detection, and the SL sequence was removed from the read. Reads containing stretches of at least nine A's in the merged reads or nine T's in the unmerged reverse reads were extracted for poly(A) site identification, and poly(A) tails were removed from the read.

SL and poly(A) reads were aligned to *T. brucei* genome (https://tritrypdb.org/tritrypdb/; genome annotation: version v5.1) using LAST (version 959) alignment tools (http://last.cbrc.jp/) (*48*). 5′ splice-acceptor sites were determined by the first position of all SL-containing reads mapping uniquely to the genome. Poly(A) sites were determined by the last position of all uniquely mapped poly(A)-containing reads. SL acceptor or poly(A) sites were considered for further analysis if a splice-acceptor or poly(A) site is supported by at least five reads. Putative *T. brucei* genes were defined by all genomic regions separated by at least one 5′ acceptor site and one 3′ poly(A) site occurring before the next downstream 5′ site. For each gene region, the longest transcript isoform was defined by the association of the most upstream SL acceptor site and the most downstream poly(A) site. In contrary, the major isoform of the *T. brucei* gene transcript was defined by the gene region bordered by the major SL acceptor and poly(A) sites (i.e., the ones with most reads aligned). This analysis identified 8831 genes in the *T. brucei* genome.

Identification of Ksplice putative new noncoding genes

A stringent selection pipeline was developed to systematically identify *T. brucei* ncRNAs. This pipeline aims to discard housekeeping (tRNAs, snRNAs, and snoRNAs) *T. brucei* ncRNAs and transcripts with protein-coding potential. First, only transcripts that do not overlap that annotated protein-coding and ncRNA genes from TriTryp data (https://tritrypdb.org/tritrypdb/; genome annotation: version v5.1) were retained. Second, *T. brucei* transcripts with protein-coding potential were excluded. Protein-coding potential was determined using three different approaches. (i) The protein-coding potential for each transcript was calculated using coding potential calculator score (CPC2) (*49*). (ii) The association with *T. brucei* ribosomes and TE of each transcript was measured using the published ribosome profiling data from *T. brucei* (*50*) and reanalyzing it with our Ksplice gene annotation. (iii) The noncoding potential of each transcript was confirmed using proteomic data from three different life cycle stages of *T. brucei* (*51*). Each transcript with noncoding potential defined in parts (i) and (ii) and not encoding any peptides or encoding solely nonunique peptides in part (iii) was classified as a Ksplice ncRNA gene.

Coding potential calculator

The longest isoform of each Ksplice gene was used for CPC2 analysis. CPC2 (*49*) discriminates coding and noncoding DNA sequences based on four intrinsic features: Fickett TETSCODE score, open reading frame (ORF) length, ORF integrity, and isoelectric point (pI). The Fickett TESTCODE score was calculated from the weighted nucleotide frequency of the full-length transcript, whereas the ORF length, ORF integrity, and pI were calculated from the longest putative ORF identified in each gene. A CPC2 score below 0.5 defined a transcript as a noncoding gene, whereas a CPC2 score of $\geq$0.5 was used to classify a transcript as encoded by a protein-coding gene.

Ribosome profiling

T. brucei ribosome profiling data (*50*) were reanalyzed using our merged genome annotation that consisted of the annotated protein-coding genes from TriTrypDB and our newly annotated Ksplice noncoding genes (major isoforms). Quantification and statistical analysis were performed as described by Vasquez *et al.* (*50*). A *T. brucei* transcript was defined as engaged in productive interaction with ribosomes if its TE score was $\geq$1 [TE = RPKM (reads per kilobase million) of ribosome profiling/RPKM of RNA-seq]. Conversely, a *T. brucei* transcript was defined to be not interacting with ribosomes if its TE score was $\leq$0.2857, meaning its transcript levels (RNA-seq data in RPKM) were 3.5× higher than its level of association with *T. brucei* ribosomes. A *T. brucei* transcript with a TE score in between (0.2857 < TE < 1) was defined to have low or few interactions with *T. brucei* ribosome. In addition, as in the work of Vasquez *et al.* (*50*), we investigated the 5′ end periodicity of mapped reads of both coding and putative noncoding genes. For fig. S6, for each gene, the number of reads mapping to each frame of translation (represented as +0, +1, and +2) was calculated, and the frame with the highest number of mapped reads was determined. A *P* value indicating the likelihood of periodicity was calculated by a binomial test on the frame with the highest number of mapped reads under the null hypothesis that this number should be equal to one-third of all reads mapped to that gene.

Proteomics

The MS proteomic data from Dejung *et al.* (*51*) were analyzed following the author's methodology with some modifications using MaxQuant version 1.6.0.1 (*52*) and searching against our Ksplice protein database. Our Ksplice protein database is composed of three sets of proteins: protein-coding genes from TriTrypDB (version 33, 10,019 entries, excluding protein-coding genes with an internal codon stop), putative proteins that originate from the Ksplice new gene sequences (2003 Ksplice new genes + 72 Ksplice Kolev ncRNAs), and putative proteins originating from intergenic region sequences of *T. brucei* genome. Intergenic region sequences were selected to have, on average, the same size and number of sequences as Ksplice new genes. All putative protein sequences (enclosed by a start and stop codon) with a minimum of seven amino acids originating from Ksplice new genes or intergenic regions of *T. brucei* genome were extracted in order of the DNA sequence and from the

three possible translation frames (excluding sequences without a start codon or/and containing an ambiguous base). A total of 14,261 proteins were extracted from Ksplice new genes and 28,750 proteins from the selected intergenic region of *T. brucei* genome.

Full-length sequencing

To investigate the presence of full-length transcripts in our RNA-seq dataset, read pairs containing the SL sequence on the forward read and a poly(A) tail on the reverse read were extracted and mapped to the *T. brucei* genome as described above (in the "Identification of Ksplice putative new noncoding genes" section). For all concordant alignments (both paired reads aligned), the boundaries of the transcripts were determined by the mapping positions of the two reads. Paired-end reads provide the accurate boundaries of *T. brucei* transcripts as reads are sequenced from the same RNA molecule.

Nanopore direct RNA-seq

Direct RNA-seq was performed as described by Bilska *et al.* (*53*). Briefly, RNA libraries were prepared from 5 µg of total RNA isolated from the BSF and procyclic form, respectively, mixed with 50 to 200 ng of oligo(dT)25-enriched mRNA from *Saccharomyces cerevisiae* yeast with a Direct RNA Sequencing Kit (Oxford Nanopore Technologies, catalog no. SQK-RNA002) according to the manufacturer's instructions. Sequencing was performed with R9.4.1 RevD flow cells on a MinION device using MinKNOW software (Oxford Nanopore Technologies). The Nanopore direct RNA-seq data from this study have been submitted to the European Nucleotide Archive–PRJEB48655.

Direct RNA-seq bioinformatic analysis

Raw sequencing reads were base-called using Guppy v.4.4.1 (Oxford Nanopore Technologies). To find entirely sequenced reads (i.e., complete from 3′ to 5′ end), delimiting signals from both termini were assessed. The 3′ end was analyzed in the following manner: Obtained reads were mapped to custom references containing lncRNA candidates (provided by F.G.) using minimap2 v.2.17, with parameters -k 14 -ax map-ont –secondary=no. Supplemental alignments and reads mapping to the reverse strand were filtered out with SAMtools v.1.9 (SAMtools view -b -F 2320). The poly(A) tails of reads were assessed using the poly(A) estimation module of the program Nanopolish v.0.13.2 (*54*). Only reads fulfilling quality criteria (tagged as "PASS") were taken into consideration in further analyses. 5′ ends were examined using two approaches. First, obtained reads were mapped to the reference genome (TriTrypDB-54_TbruceiTREU927_Genome.fasta) obtained from TriTryp database using minimap2 v.2.17 (*55*) with the following parameters: -k 14 -ax splice -uf –secondary=no. Second, mapped reads were merged and converted to BLAST databases with makeblastdb command using default settings for nucleotide sequences (for each sample separately). Subsequently, the BLASTn v.2.11.0 (*56*) algorithm was run with SL-RNA sequence used as a query against built databases with the following parameters: -evalue 10000 -num_alignments 1000000 -dust no -max_hsps 1 -word_size 4 -num_threads 10 –outfmt 6. The output data were filtered on the basis of the following criteria: (i) reads containing a >10-bp region of homology with SL-RNA query, (ii) reads aligned with an *e* value of <100, (iii) alignment starts at the termini of read (subject) sequence (value of blast tabular output columns start/send == 1), and (iv) alignment starts at the termini of SL-RNA (query) sequence (value of blast tabular output columns qstart/qend == 1). Reads satisfying all of the above criteria were taken into account for further analysis. Second, reads mapping to the reference genome were filtered with cutadapt v.2.10 (*57*) with the following options: --discard-untrimmed -e 0.2 --action "lowercase," based on the presence of the potential SL RNA sequence. Last, unique read IDs obtained with both 5′ end analysis approaches and 3′ analysis were combined and used to filter the genomic alignment files. Filtering was performed in R 4.0.2 using Rsamtools 2.6.0 from Bioconductor. The results were saved in BAM and FASTA formats. Gene body coverages were plotted using RseQC genebodycoverage.py script (*58*). The obtained sequence alignments were saved in BAM format. Read sequences were extracted in FASTA format as two variants: entirely (encompassing total base-called length) and with soft clip masking [soft clipping 5′ SL and miscalled bases from 3′ poly(A) tail were removed on the basis of cigar scores] using biostar84452.jar from jvarkit toolset. Two corresponding FASTA datasets were coerced, and 5′ SLs were identified. The coordinates of 3′ termini of transcripts were pointed out on the basis of sequence alignment and Nanopolish poly(A) function output.

RNA-FISH

Between 2.5×10^5 and 1×10^6 cells were harvested by centrifugation (1800*g* for 10 min), washed with 1× phosphate-buffered saline (PBS) or trypanosome dilution buffer (TDB), and resuspended in between 500 µl and 1 ml of fixation buffer [3.7% formaldehyde diluted in ribonuclease (RNase)–free PBS] for 10 min at room temperature. Fixed cells were washed with between 500 µl and 1 ml of RNase-free PBS and resuspended with 150 µl of RNase-free PBS. Cells were then settled on precoated polylysine culture dishes (35-mm glass bottom, MatTek) for at least 20 min. PBS was removed, and cells were permeabilized with 1 ml of 70% ethanol (EtOH; in RNase-free water) for at least 1 hour at +2° to +8°C. EtOH (70%) is discarded, and cells were washed with 200 µl of wash buffer A [10% (v/v) formamide in 1× wash buffer A; Biosearch Technologies, catalog no. SMF-WA1-60]. Cells were incubated with 100 µl of hybridization buffer containing 1.25 µM RNA-FISH probes in the dark at 37°C overnight (~16 hours). Cells were washed with 200 µl of wash buffer A and incubated with 200 µl of wash buffer A in the dark at 37°C for 30 min. Cells were stained with a solution of 4′,6-diamidino-2-phenylindole (DAPI; 1 µg/ml) (in wash buffer A) in the dark at 37°C for 30 min. Cells were washed with 200 µl of wash buffer B (Biosearch Technologies, catalog no. SMF-WB1-20) and incubated with it at room temperature for 2 to 5 min. VECTASHIELD (100 µl) was added to the dishes before analysis with the Zeiss cell observer widefield microscope.

RNA-FISH probes were designed using the online tools provided by LGC Biosearch Technologies (Stellaris Probe Designer, www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer). A total of 17 probes were designed for *grumpy* lncRNA and 30 to 43 probes for Ksplice lncRNA223a, lncRNA1077a, lncRNA1735a, and lncRNA5090a.

Differential expression of Ksplice new genes between BSF and procyclic form

Differential expression analysis for Ksplice new genes between BSF and procyclic form was performed using our merged annotation of *T. brucei* genome (major isoform of Ksplice new genes + TriTryp

protein-coding genes) and the DEseq2 package. To that end, we used our previously published transcriptomic data (*59*) containing 13 RNA interference target sequencing (RNA-seq) samples replicated for both BSF and procyclic form.

RIT-seq analysis of Ksplice new genes

The RNA interference target sequencing (RIT-seq) data from Alsford *et al.* (*23*) were reanalyzed by aligning the sequence reads against our merged annotation of *T. brucei* genome (major isoform of Ksplice new genes + TriTryp protein-coding genes). Quantification and statistical analysis were performed as described by Alsford *et al.* (*23*).

Transcript quantification and cRT-PCR

Transcript quantification was performed by RT-qPCR, as described by Aresta-Branco *et al.* (*60*), except that random hexamer primers were used to generate cDNA. The cRT-PCR protocol was performed essentially as described in (*61*). Briefly, parasites were harvested by centrifugation at 677*g* for 10 min at 4°C and immediately resuspended in TRIzol (Life Technologies). Total RNA was isolated following the manufacturer's instructions, and RNA was quantified in a NanoDrop 2000 (Thermo Fisher Scientific). The ideal RNA concentration to perform the cRT-PCR protocol is 0.5 to 1 mg/μl. The RNA cap and poly(A) tail were removed by oligonucleotide-directed RNase H cleavage using SL and oligo(dT) primers. After RNase H treatment, RNA was extracted with phenol/chloroform approach and precipitated using EtOH precipitation protocol. RNase H–treated RNA (3 to 5 μg) was circularized using T4 RNA ligase 1 [ssRNA Ligase, New England Biolabs (NEB)], RNA was extracted using phenol/chloroform, and EtOH was precipitated. RNA was reverse-transcribed using the gene-specific primer R1 (100 nt from the 5′ end of the transcript or the RNase H cleavage site) and RT buffer and 5 mM magnesium from the SuperScript II kit (Life Technologies). The resulting cDNA molecules contained the juxtaposed 5′ and 3′ ends of circular RNA. PCR was performed on the produced cDNA using gene-specific primers R2 and forward F1. R2 primer is in a "nested" position relative to the R1 primer and contributes to the specificity of the PCR amplicon. PCR#1 product was purified using a MinElute PCR purification kit (QIAGEN), and a second round of PCR amplification was performed with gene-specific primers R2 and forward F2. The F2 primer is in a nested position compared to the F1 primer and contributes to PCR amplicon specificity. PCR#2 product was ligated to pGEM-T easy vector or TOPO vector following the manufacturer's instructions (Promega). After transformation in bacteria and plasmid amplification, the subcloned PCR#2 fragments were amplified and sequenced using T7 and SP6 primers.

Quantitative RT-PCR

RNA was prepared with TRIzol reagent (Invitrogen) according to manufacturer's instructions, and cDNA was synthesized with random primers and SuperScript II reverse transcriptase. qPCR was performed with AmpliTaq Gold DNA Polymerase (Power SYBR Green Master Mix, Applied Biosystems) and gene-specific primers as follows: control of differentiation (Tb927.10.12970), CCAGCCTTCTCAATCTCCAG (forward) and GGCCACAGTTGGATAGCTTG (reverse); Tb927.2.2220 (new control for the slender to stumpy transition), CCTGCCCCCAGTTTAGCG (forward) and CTGAAGTGTCACACAAGCGATG (reverse); Tb927.10.12080, CCTGCAGGCGTCACATTC (forward) and C A G T G A A G A A G A A A A G GCACG (reverse); *grumpy* I (*grumpy* or *grumpy* 5′ end), AACGGAAGGAAAGTTTGTGAATGC (forward) and GTGAATGAACTTTTTGTTTGGCGTC (reverse); *grumpy* II, GTCCATTGAAGTACACAGATAC (forward) and GGAAAAGACCACAACAGAAAG (reverse); *grumpy* III, GTCCATTGAAGTACACAGATAC (forward) and GAAAACCGGTGGAAATGC (reverse); LNA site, GTGCTATAAATTTAGGAGGCTGC (forward) and TGTGTACTTCAATGGACAAATCAG (reverse); *snoGRUMPY* (*grumpy* 3′ end), GCACGTGATGAGAGTATGTTC (forward) and CAGGGAAACGACAGTCTATG (reverse); RBP7A, GCTCGACTTTTTGTTGGGCAG (forward) and CATATTGTAGCGGTTGTGAAGCG (reverse); RBP7B, CTTTAACGCAACCGAAGATG (forward) and CAACGGTTGTGAAGTCCG (reverse); lncRNA-5090a, GTGAGAAACATGGAGCGGAG (forward) and GCGAGAGGGGAAATAAGGAG (reverse); lncRNA-1077a, TGACTTACTCCTGCGGTGATG (forward) and GAACGGCAACTTCTGTGGG (reverse); AATP11, CAAACAGCCCTTATTCAACACCTCACG (forward) and GGTTGAGTCCTCACTGCGTCACG (reverse); PAD2, GATAACCAAACA AATGCTGAAAATG (forward) and CTTGCTTGAATCAGCAGCTTTC (reverse). The quantitative PCR program was as follows: stage 1, 10 min at 95°C; stage 2, 15 s at 95°C, 15 s at 60°C, and 30 s at 72°C (40 cycles); melt curve, 15 s at 95°C, 1 min to 10 s at 60°C, and 15 s at 95°C.

Microscopy quantification (*SnoGRUMPY*)

The RNA-FISH signal of nucleolar *grumpy* (=*snoGRUMPY*) was quantified by correcting the mean fluorescence intensity of nucleolar *grumpy* (=*snoGRUMPY*) with the mean fluorescence intensity of equal area in the cytoplasm, defined here as the background signal of the cells. Note that although we cannot see a clear and defined signal in the cytoplasm, we cannot exclude that *grumpy* could occasionally be present in the cytoplasm.

Constructs

For the overexpression of *grumpy*, the noncoding sequence of *grumpy* lncRNA was amplified from *T. brucei* Antat 1.1E genomic DNA with forward (5′-CAAAAGGACAGAATTATAGGTTCA-3′) and reverse (5′-GATGCAGCTCAACAGCAAG-3′) primers and inserted into pDEX577 (phleo) between the Hind III and Bam HI sites of the plasmid. pDEX577 vectors are highly modular expression vectors for inducible expression of transgenes, integrating in the mini-chromosome repeats, which was designed and constructed by Kelly *et al.* (*62*). Moreover, two T7 terminator sequences were inserted between the Bam HI and Kpn I sites of the plasmid just downstream to the *grumpy* lncRNA construct. For the overexpression of *snoGRUMPY*, we used the same vector and cloning procedures as for overexpression of *grumpy*, except that *snoGRUMPY* was amplified from *T. brucei* Antat 1.1E genomic DNA with forward (5′-CCGTTCGCATCCTGTGTG-3′) and reverse (5′-TTATCGAATCAATCCATCCAAAG-3′) primers. For the RNAi construct of *grumpy*, lncRNA was designed as a long stem-loop double-stranded RNA and by following the experimental procedures described by Atayde *et al.* (*63*). Briefly, a fragment of *grumpy* lncRNA sequence was amplified from *T. brucei* Antat 1.1E genomic DNA with a forward primer (5′-CAAAGGCAAGATATGTGAAC-3′) containing either Hind III or Bam HI restriction site in its 5′ end and a reverse primer (5′-GCAGGGGTGTGGAAAAGAC-3′) containing Eco RI restriction site followed by a 50-nt randomized sequence at its 5′ end. Two PCR amplicons were therefore obtained: one containing

Hind III restriction site and the other containing Bam HI restriction site in their 5′ ends. These two PCR amplicons were then digested by Eco RI restriction enzyme and ligated together using 1000 U of T4 DNA ligase (NEB) for 16 hours at 16°C. The ligation product containing two *grumpy* fragments in opposite directions (sense and antisense) and separated by a 100-nt stuffer stem-loop was then gel-purified. Hind III and Bam HI restriction sites located at both ends of the ligated product were used to clone the stem-loop RNAi construct of *grumpy* into the pDEX577 vector (phleo). For the RNAi construct of *snoGRUMPY*, the sequence of *snoGRUMPY* was amplified from *T. brucei* Antat 1.1E genomic DNA with forward (5′-TATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCGCACGTGATGAGAGTATGTT-3′) and reverse (5′-AAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGGTCAGGGAAACGACAGTCT-3′) primers (note that the underlined sequences are extra and unrelated nucleotides added to the primers to increase the size of the RNAi construct). The *snoGRUMPY* RNAi construct was inserted between the Xho I and Hind III sites of the pZJM plasmid (head-to-head T7 promoters) (*64*). For the overexpression of Tb927.10.12080, the coding sequence and the 3′ untranslated region (3′UTR) were amplified from *T. brucei* Antat 1.1E genomic DNA with forward (5′-tccacttgacactagtATGTCTGTTTCCCCCAAGC-3′) and reverse (5′-tttagtgtttaatgggtaccAACAACGACAGCAAAAGACATC-3′) primers. For the overexpression of Tb927.10.12080 together with *grump*y, forward (5′-tccacttgacactagtATGTCTGTTTCCCCCAAGC-3′) and reverse (5′-tttagtgtttaatgggtaccGCAGAAGATACGGAGGAATGAG-3′) primers were used. Both PCR amplicons were inserted into pDEX577 (phleo) between the Spe I and Kpn I sites of the plasmid. All constructs were linearized with Not I before transfection. Stable transfectant clones were obtained by serial dilution of the transfected population and selected after 6 to 7 days after transfections. Expression was induced by adding tetracycline (in vitro) or doxycycline (in vivo) at the following concentrations: 1 μg/ml and 1 mg/ml. N-terminal MYC-tagged Tb927.10.12080 and Tb927.8.2860 constructs were prepared by cloning the N-terminal coding sequences of the respective genes into pNAT X-tag (a gift from S. Alsford, London School of Hygiene and Tropical Medicine, UK) between the Avr II and Bam HI restriction sites. Forward (5′-ATGCCTAGGTCTGTTTCCCCCAAGCGGAG-3′) and reverse (5′-CCCGGATCCCCATCCGGCCATCGATGGAT-3′) primers were used for the cloning of Tb927.10.12080. Forward (5′-ATGCCTAGGTCTGTTAGCGCCCCAACGAT-3′) and reverse (5′-CCCGGATCCGCTTACCCCCGTCTCCCTTG-3′) primers were used for the cloning of Tb927. 8.2860 (HYP5). The plasmids were linearized with Bsa AI and Bst API, respectively, for integration into specific chromosomal loci.

Inducible expression of RNAi and overexpression constructs

Cells were diluted at 5×10^4 parasites/ml and induced with tetracycline (1 μg/ml) for 3 days for *grumpy* RNAi and Tb927.10.12080 overexpression and for 6 days for *grumpy* overexpression cell lines. Cells were counted every day, live/dead cells were assessed by propidium iodide staining, GFP::PAD1-positive cells were scored, and all these parameters were quantified using an Accuri C6 flow cytometer. For *grumpy* RNAi/overexpression, at day 2 after tetracycline induction, RNA samples were collected by centrifugation of equivalent number of cells and addition of TRIzol reagent (Invitrogen) to the cell pellets. For the *grumpy* overexpression cell line, at day 3 after tetracycline induction, an equivalent number of cells were collected by centrifugation, and the cell cycle profiles were assessed using an Accuri C6 flow cytometer and propidium iodide staining in fixed cells.

Slender to stumpy differentiation assay

Cell cultures were started at 5×10^4 parasites/ml and induced or not with tetracycline (1 μg/ml). Every day of culture, a sufficient number of cells (>10,000 parasites) were collected, washed with TDB [5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20 mM Na_2HPO_4, 2 mM NaH_2PO_4, and 20 mM glucose (pH 7.4)], and resuspended in 200 μl of TDB with propidium iodide (1 μg/ml). A fixed volume of each cell culture was analyzed by flow cytometry (Accuri C6) to simultaneously measure parasite density, live and dead parasites, and the GFP::PAD1 expression.

Cell cycle profile assay

Cell cultures were started at 5×10^4 parasites/ml and induced or not with tetracycline (1 μg/ml). After 3 and 4 days of in vitro culture, 2×10^6 parasites were collected and spun down (1300*g* for 10 min at 4°C), washed once with ice-cold PBS, resuspended in 1 ml of PBS/2 mM EDTA, and fixed by dropwise addition of 2.5 ml of ice-cold 100% EtOH (stored at −20°C). Cells were fixed at 4°C for at least 1 hour, washed once with 1 ml of PBS/EDTA at room temperature, and resuspended in 1 ml of PBS/EDTA. RNA was digested by adding 1 μl of RNase A (10 μg/μl), and DNA was stained by adding 1 μl of propidium iodide (1 mg/μl) for 30 min at 37°C. Cell cycle profile was analyzed by flow cytometry using an Accuri C6 machine with FL3 channel.

Stumpy to procyclic differentiation assay

Cell cultures of BSFs were started at 5×10^4 parasites/ml, and *grumpy* lncRNA was induced or not with tetracycline (1 μg/ml). After 2 days of in vitro culture, the number of stumpy forms was assessed by measuring the GFP::PAD1 expression using flow cytometry (Accuri C6). BSFs were collected from culture, spun down, resuspended in differentiation trypanosome medium with 6 mM *cis*-aconitate at 1×10^6 parasites/ml, and incubated at 27°C. Parasite differentiation into procyclic forms was assessed at 12 hours after differentiation by flow cytometry using anti–*T. brucei* procyclin antibody (0.5 mg; Cedarlane, clone TBRP1/247, CLP001AP) conjugated with Alexa Fluor 647 (protein labeling kit, Molecular Probes) (1:500 dilution in TDB).

PAD1 staining

BSF parasites (5×10^5) were harvested by centrifugation (1800*g* for 10 min), washed with 1× PBS, and resuspended in 500 μl of fixation buffer (4% paraformaldehyde diluted in 1× PBS) for 10 min at room temperature. Fixed cells were washed with 500 μl of 1× PBS and resuspended with 100 μl of 1× PBS. Cells were then settled on precoated polylysine culture dishes (35-mm glass bottom, MatTek) for at least 20 min. PBS was removed, and cells were permeabilized with 100 μl of 0.1% Triton X-100 in PBS for 2 min at room temperature. Permeabilized cells were washed five times with 200 μl of PBS and blocked with 2% bovine serum albumin (BSA) in PBS for 45 min at 37°C in a humidity chamber. Cells were incubated with 100 μl of the primary antibody anti-PAD1 (1:1000 in 2% BSA in PBS; antibody provided by K. Matthews) overnight at 4°C in a humidity chamber. Cells were washed five times with 200 μl of PBS and incubated with 100 μl of the secondary antibody anti-rabbit

(1:1000 in 2% BSA in PBS; goat anti-rabbit Alexa Fluor 647, Invitrogen, #A21245) for 45 min at 4°C in a humidity chamber. Parasite DNA was stained using 100 µl of DAPI or Hoechst solution (1 µg/ml) for 20 min at room temperature. Cells were washed five times with 200 µl of PBS, and 100 µl of VECTASHIELD was added to the dishes before analysis with the Zeiss cell observer widefield microscope.

Infections and sample collection

Four-week-old male c57BL/6 mice (Charles River Laboratories, France) were inoculated intraperitoneally with 2000 parasites. Mice were infected with either Antat1.1 90:13 GFP::PAD1 cell line or Antat1.1 90:13 GFP::PAD1 *grumpy* overexpression cell line. Mice infected with Antat1.1 90:13 GFP::PAD1 *grumpy* overexpression parasites were separated in three different cages: One cage of four mice received only water, one cage of four mice received water with doxycycline hyclate (1 mg/ml; Sigma-Aldrich) at day 4 after infection, and one cage of three mice received water with doxycycline hyclate (1 mg/ml) at the day of infection. Parasitemia was monitored by tail-vein bleeds every other day and counted using a hemocytometer with 1:150 blood dilution in TDB. The percentage of stumpy forms in the mouse blood was assessed by measuring GFP::PAD1 expression in blood-diluted sample using an Accuri C6 flow cytometer. Mouse survival was monitored every other day until 100 days after infections. Mice were euthanized at the first signs of severe disease distress, with all efforts to minimize animal suffering. Humane end points were used.

grumpy LNA treatment

Antisense LNA gapmeRs targeting *grumpy* lncRNA (LNA sequence: AGGATGCGAACGGGA; from QIAGEN) were resuspended at 50 µM with tris-EDTA buffer. The stumpy reporter cell line cells (GFP::PAD1) were diluted at 5×10^4 parasites/ml and grown for 3 days with different concentrations of LNA gapmeRs (0 to 50 nM; diluted into culture medium). Cells were counted every day, live/dead cells were assessed by propidium iodide staining, GFP::PAD1-positive cells were scored, and all these parameters were quantified using an Accuri C6 flow cytometer. At day 1 after LNA exposure, RNA samples were collected by centrifugation of an equivalent number of cells and addition of TRIzol reagent (Invitrogen) to the cell pellets.

In vitro RNA-protein pulldown

Stumpy cells (350×10^6; GFP::PAD1-positive cells) were used for each condition of the RNA-protein pulldown. Cells were washed once with PBS and then kept on ice as a dry cell pellet (350×10^6 stumpy cells in ~30 µl of dry pellet). The nuclear proteins were then extracted from the dry cell pellets with the NE-PER kit according to the manufacturer's instructions (Thermo Fisher Scientific). The RNA baits used in the RNA-protein pulldown assays were synthetized in vitro with the MEGAscript T7 Transcription Kit according to the manufacturer's instructions (Thermo Fisher Scientific). After phenol:chloroform extraction, the RNA baits were biotinylated using the Pierce RNA 3′ End Desthiobiotinylation Kit and following the manufacturer's instructions (Thermo Fisher Scientific). RNA-protein pulldown assays were performed with the Pierce Magnetic RNA-Protein Pull-Down Kit according to the manufacturer's instructions (Thermo Fisher Scientific). Nuclear proteins (~75 µg) were used under each condition of RNA-protein pulldown. The primers used to synthetize the RNA baits were the following: *grumpy* lncRNA, taatacgactcactatagggagCAAAAGGACAGAATTATAGGTTCA (forward) and GATGCAGCTCAACAGCAAG (reverse); lncRNA-5090a (lncRNA control), taatacgactcactatagggagAACCCCCTCCACTCCTAC (forward) and TTGCCGTTGTTTTCGTGC (reverse).

Our *grumpy* RNAi vector was linearized using Bam HI restriction enzyme and used as a template for the RNA in vitro production of the *grumpy* RNAi bait. Note that we cannot exclude that the large proportion of mitochondria proteins found in our in vitro RNA-protein pulldown assay (Fig. 4A) could come from mitochondria contamination. The first lysis step of the NE-PER kit could be too mild to completely lyse mitochondria. Thus, mitochondria proteins will be extracted during the second lysis step together with the nuclear proteins. Under this condition, mitochondria and nuclear proteins will therefore be used together in our pulldown experiment.

MS sample preparation and measurement

Samples were loaded onto a Novex NuPage 4 to 12% precast gel (Thermo Fisher Scientific) and run in 1× Mops buffer for 10 min at 180 V. The Coomassie G250 (Roth)–stained gel lanes were minced and destained with 50% EtOH/25 mM ammonium bicarbonate (ABC). Subsequently, gel pieces were dehydrated with 100% acetonitrile (ACN) and dried for 5 min in a concentrator (Eppendorf). Samples were incubated with reduction buffer [10 mM dithiothreitol (DTT; Sigma-Aldrich)/50 mM ABC] for 30 min at 56°C and further alkylated for 30 min in the dark with iodoacetamide [50 mM indole-3-acetic acid (Sigma-Aldrich)/50 mM ABC]. Gel pieces were completely dehydrated with ACN and covered in MS-grade trypsin (Serva) solution (1 µg of trypsin per sample). Proteins were digested overnight at 37°C, and peptides were extracted twice by incubation with extraction buffer (3% trifluoroacetic acid and 30% ACN) for 15 min. The gel pieces were dehydrated with 100% ACN, and the extracted volume was reduced to approximately 150 µl in a concentrator (Eppendorf). Extracted peptides were desalted in StageTips using two layers of C_{18} material (Empore). Eluted peptides were injected via an autosampler into a uHPLC (EASY-nLC 1200, Thermo Fisher Scientific) and loaded on a 55-cm capillary (75-µm inner diameter; New Objective) packed in-house with Reprosil C18-AQ 1.9-µm resin (Dr. Maisch GmbH) for reversed-phase chromatography. The EASY-nLC 1200 HPLC system was directly mounted to an Exploris 480 mass spectrometer (Thermo Fisher Scientific). Peptides were eluted from the column with a 73-min optimized gradient from 2 to 40% ACN with 0.1% formic acid at a flow rate of 250 nl/min. Chromatography was performed with a column oven setup operating at 55°C (Sonation). The ion transfer tube temperature was set to 250°C. Spray voltage ranged from 2.2 to 2.4 kV. The mass spectrometer was operated in data-dependent acquisition mode with one MS full scan and up to 15 triggered tandem MS (MS/MS) scans using higher energy collisional dissociation (HCD) fragmentation. MS full scans were obtained in the Orbitrap at 60,000 resolution, while MS/MS scan resolution was set to 15,000 resolution. Charge states 2 to 6 were included from MS/MS selection, and peptide match was preferred.

MS data analysis

Raw files were processed with MaxQuant (version 1.5.2.8) (*52*) and searched against *T. brucei TREU927* database (version 8.1; 11,567 entries) downloaded from tritrypdb.org using the Andromeda search

engine. Carbamidomethylation was set as a fixed modification, while acetyl (protein N terminus) and oxidation (Met) were considered as variable modifications. Trypsin (specific) was selected as protease. Proteins were quantified with at least two quantification counts of which at least one of them was a unique peptide. Known contaminants and reverse hits were removed. For statistical analysis, the mean was calculated and Welch *t* test was performed between the conditions. Graphical data representation was done using the R environment (version 3.6.2).

RIP-DRBD3

T. brucei 449 procyclic cells (*65*) were cultured at 27°C in SDM-79 medium supplemented with 10% fetal bovine serum (*66*). For the RIP assays, 2×10^8 cells were washed in serum-free HMI-9 or SDM-79; resuspended in 1 ml of 10 mM tris-HCl (pH 7.4), 0.5% Igepal, 1 mM DTT, 2 mM ribonucleoside vanadyl complexes (Invitrogen), and protease inhibitors (Roche); and lysed by passing the cell suspension thrice on ice. Lysates were centrifuged at 16,000*g* for 10 min at 4°C, and NaCl was added to 0.9 ml of the supernatant to a final concentration of 150 mM. The supernatant was then split into two halves and rotated for 3 hours at 4°C in the presence of normal rabbit serum (NRS; Sigma-Aldrich) or anti-DRBD3 antiserum (*28*) coupled to 1.5 mg of Dynabeads Protein G {previously blocked overnight in IPP*-150 [10 mM tris-HCl (pH 7.4), 150 mM NaCl, and 0.5% Igepal CA-630] containing 0.1% BSA}. Beads were then washed four times in IPP*-150 and twice in IPP*-150 containing 1 M urea and incubated for 30 min at 50°C with 100 mg of proteinase K in 10 mM tris-HCl (pH 8), 100 mM NaCl, 0.5% SDS, and 1 mM EDTA. RNA was extracted with phenol:chloroform, precipitated with EtOH, and converted to cDNA using 0.5 μg of random hexamers (Invitrogen), 0.25 mM each dNTPs (deoxynucleotide triphosphates), 20 U of RiboLock (Thermo Fisher Scientific), and 200 U of Maxima reverse transcriptase (Thermo Fisher Scientific) in a final volume of 20 μl. Reactions were incubated for 5 min at room temperature, then incubated for 30 min at 50°C, and lastly heated at 85°C for 5 min. RT-qPCR reactions were carried out in a Bio-Rad CFX96 thermal cycler. Reactions were set up in a final volume of 10 μl containing 0.1 to 0.5 μl of cDNA, 1× SYBR Green Master Mix (Thermo Fisher Scientific) and 0.5 μM each oligodeoxynucleotide. Fold enrichments of target transcripts relative to control serum (NRS) were calculated using *actin* mRNA as a reference according to the formula $2^{-\Delta\Delta Ct}$, where $\Delta\Delta C_t = (C_t\text{NRS\_t.o.i.} - C_t\text{NRS\_actin}) - (C_t\text{DRBD3\_t.o.i.} - C_t\text{DRBD3\_actin})$, and t.o.i. is the transcript of interest. Experiments were carried out in three biological replicates, and the data are represented as the means ± SEM.

Western blot analysis

Whole-cell lysates (10^7 cells) were fractionated by 10% SDS–polyacrylamide gel electrophoresis, transferred to PROTRAN membranes (Whatman), and reacted with antibodies. The bound antibodies were detected with goat anti-rabbit immunoglobulin G coupled to horseradish peroxidase and were visualized by ECL (Amersham Biosciences). The dilutions used for the antibodies are PTB1 (1:10,000) (*27*), hnRNP F/H (1:10,000) (*26*), cMYC (1:10,000; 9E10, Santa Cruz), HSP83 (1:10,000) (*8*), and ZC3H41 (1:10,000) (*67*).

Fractionation on sucrose gradient

Whole-cell extracts were prepared from 1010 *T. brucei* procyclic cells in a buffer containing 150 mM KCl, 20 mM tris (pH 7.6), 10 mM $MgCl_2$, 0.5 M DTT, and 0.1% NP-40. In addition, 1 μl of leupeptin (10 mg/ml) protease inhibitor and 1 μl of RNasin (Thermo Fisher Scientific) were added to the lysate. Lysates were fractionated on a 10 to 30% (w/v) sucrose gradient by centrifugation for 3 hours at 35,000 rpm in a Beckman SW41 rotor at 4°C. Fractions (500 μl) were collected, EtOH was precipitated, and RNA was extracted using phenol-chloroform. RNA extracted from each fraction was separated on a 6 or 10% (w/v) polyacrylamide gel containing 7 M urea gel and transferred onto nitrocellulose membranes. The RNA blots were hybridized to a ^{32}P-labeled antisense RNA probe and visualized by autoradiography (*8*).

Northern blot analysis

Total RNA was separated on a 6 or 10% (w/v) polyacrylamide gel containing 7 M urea or 1.2% agarose/formaldehyde gel and transferred onto nitrocellulose membranes. The RNA blots were hybridized to a ^{32}P-labeled antisense RNA probe. The results were analyzed by autoradiography (*8*, *37*). Primers used to amplify DNA templates for RNA probes synthesis (using T7 polymerase) were as follows: *snoGRUMPY*/TB10Cs2'C1_F, 5′-GCACGTGATGAGAGTATGTT-3′; *snoGRUMPY*/TB10Cs2'C1_T7_R, 5′-TTAATACGACTCACTATAGGGAGAGGTCAGGGAAACGACAGTCT-3′; 7SL_F, 5′-TTGCTCTGTAACCTTCGGGG-3′; 7SL_T7_R, 5′-TTAATACGACTCACTATAGGGAGACCGCCTCGCGACGACACTTG-3′; TB11Cs2C2_T7_R, 5′-TTAATACGACTCACTATAGGGAGAGATCAGTCAGGGCATAAAAT-3′; TB11Cs2C2_F, 5′-AAGTGATTGACACCTAGGCC-3′; TB11Cs2C1_T7_R, 5′-TTAATACGACTCACTATAGGGAGAAATCAGTCTGGGGCGACAAT-3′; TB11Cs2C1_F, 5′-TGAATGATGACTGACAAAAC-3′; TB11Cs3C2_F, 5′-ATGAAAAACCTTTCATGCTG-3′; TB11Cs3C2_T7_R, 5′-TTAATACGACTCACTATAGGGAGATTCAGAATGCCGTCTATAAT-3′.

In situ hybridization in procyclic cells

Procyclic *T. brucei* cells were fixed on circular coverslips using 1.6% formaldehyde and permeabilized using Triton X-100/Tween 20 (1:0.1%) in 1× PBS. In situ hybridization with specific CY3-labeled antisense RNA probes was performed as described (*8*). The nucleolus was localized using anti-NHP2 antibody (0.7:1000 dilution) (*68*), the MTAP protein was localized using an MTAP-YFP reporter (*69*), and the nucleus was stained with DAPI. Images were acquired using a Leica SP8 confocal microscope equipped with a white light laser and gating. Cells were oversampled both in the lateral and axial axis. The images were captured using an X100 HC Plan-Apo 1.4 numerical aperture objective at 512 by 512 pixels with Z slices taken every 200 nm. Excitation wavelengths used in this study were 405 nm for DAPI, 551 nm for detecting secondary anti-rabbit cy3-conjugated antibody, and 640 nm for detecting cy5-labeled antisense RNA (8 to 12 accumulations). The images were then deconvolved using Huygens Professional software with standard parameters (SVI, Laapersveld 6, 1213 VB Hilversum, The Netherlands).

RiboMeth-seq

Total RNA (5 μg) from *T. brucei* cells was initially denatured at 90°C for 2 min in a thermocycler, and an equal volume of buffer [$NaHCO_3/Na_2CO_3$ (pH 9.9)] was added. The RNA samples were subsequently incubated at 90°C for 20 min and later kept on ice (*37*, *70*). The hydrolyzed RNA was then used for library preparation. Briefly, 800 ng of the RNA was dephosphorylated with FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific) and cleaned by Agencourt RNA clean XP beads (Beckman Coulter).

The RNA was then ligated to 3′ linker using high-concentration T4 RNA Ligase 1 (NEB) in a buffer containing dimethyl sulfoxide (DMSO), adenosine 5′-triphosphate (ATP), polyethylene glycol (PEG) 8000, and RNase inhibitor (NEB) for 1.5 hours at 22°C. The ligated RNA was purified from excess linker using Dynabeads MyOne SILANE beads (Thermo Fisher Scientific), and first-strand cDNA was prepared using the AffinityScript Reverse Transcriptase enzyme (Agilent) at 55°C for 45 min. Next, the RNA was degraded using 2 μl of 1 M NaOH, and the cDNA was cleaned using Dynabeads MyOne SILANE beads (Thermo Fisher Scientific). The cDNA was further ligated to 3′ adapter using a high-concentration T4 RNA Ligase 1 (NEB) overnight at 22°C and cleaned of excess adapter using Dynabeads MyOne SILANE beads (Thermo Fisher Scientific). The adapter ligated cDNA was PCR-enriched using NEBNext High-Fidelity (NEB) polymerase (nine PCR cycles), separated on an E-Gel EX agarose gel (Invitrogen), and size-selected for the range of 150 to 300 bp (containing ~30 to 180 nt corresponding to RNA). The amplicons were gel-purified using a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and sequenced in a NextSeq system (Illumina) in paired-end mode (20 million reads for each sample). BioProject ID for the RiboMeth-seq is the PRJNA776556.

In vivo psoralen ultraviolet cross-linking and ligation to generate chimeric RNA

All samples described in this study were previously described (*8*, *42*). Briefly, *T. brucei* procyclic cells were washed twice with 1× PBS and concentrated (~5×10^9), and 4′-aminomethyltrioxsalen hydrochloride (AMT; Sigma-Aldrich) was added to the cells at a concentration of 0.2 mg/ml and incubated for ~40 min on ice. Cells were aliquoted for −UV (ultraviolet) and +UV treatments. For UV-treated samples, the cells treated with AMT were irradiated using a UV lamp at 365 nm, at a light intensity of 10 mW/cm^2 for 30 min on ice.

To enrich for small ncRNAs, postribosomal supernatant (PRS) was prepared from −UV and +UV samples. Whole-cell extracts were prepared from cells subjected to treatment with psoralen plus UV cross-linking and from samples not subjected to cross-linking in a buffer containing 24 mM KCl, 10 mM tris (pH 7.8), and 10 mM $MgCl_2$. RNPs were extracted with 300 mM KCl, and the ribosomes were removed by centrifugation for 3 hours at 35,000 rpm in a Beckman 70.1Ti rotor (150,000*g*). The supernatant was later deproteinized by digestion with proteinase K (Roche) (100 μg/ml in 1% SDS for 30 min), and RNA was prepared using TRIzol reagent (Sigma-Aldrich). RNA extracted from the PRS was used for library preparation, essentially as described below.

To prepare small RNA-seq libraries, the cross-linked RNA before and after UV treatment was fragmented by mild alkaline hydrolysis using FastAP buffer (Thermo Fisher Scientific), dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific), and cleaned by Agencourt RNA clean XP beads (Beckman Coulter), and the RNA-RNA hybrid was ligated using high-concentration T4 RNA Ligase 1 (NEB) in a buffer containing DMSO, ATP, PEG 8000, and RNase inhibitor (NEB). The cross-linking between the ligated RNAs was reversed by irradiation at 254 nm. The recovered RNA was then used for library preparation, as previously described (*31*, *37*, *42*). Briefly, 800 ng of PRS RNA was dephosphorylated with FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific), cleaned by Agencourt RNA clean XP beads (Beckman Coulter), and ligated to 3′ linker using high-concentration T4 RNA Ligase 1 (NEB) in a buffer containing DMSO, ATP, PEG 8000, and RNase inhibitor (NEB). The ligated RNA-RNA hybrid was cleaned from excess linker using Dynabeads MyOne SILANE beads (Thermo Fisher Scientific), and first-strand cDNA was prepared using AffinityScript Reverse Transcriptase (Agilent). The RNA was subsequently degraded using 2 μl of 1 M NaOH, and the cDNA was cleaned using Dynabeads MyOne SILANE beads (Thermo Fisher Scientific). The cDNA was further ligated to 3′ adapter using high-concentration T4 RNA Ligase 1 (NEB) and cleaned of excess adapter using Dynabeads MyOne SILANE beads (Thermo Fisher Scientific). The adapter ligated cDNA was PCR-enriched using NEBNext High-Fidelity (NEB) polymerase (nine PCR cycles), separated on an E-Gel EX agarose gel (Invitrogen), and size-selected in the range of 150 to 300 bp (containing ~30 to 180 nt corresponding to the input RNA). The amplicons were gel-purified using a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and sequenced in a NextSeq system (Illumina) in paired-end mode (20 million to 40 million reads for each sample).

Analysis of chimeric RNA molecules

The sequencing reads generated from in vivo psoralen UV cross-linking and ligation to generate chimeric RNA were retrieved from a previous study (*8*). The reads from the RNA-seq as described above were preprocessed using Trim Galore v0.5.0 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Overlapping read pairs were merged using FLASH v1.2.11 (*71*). BWA-MEM v0.7.17-r1188 (*72*) was used to align the reads to a custom-made *T. brucei* transcriptome database. The SAM output was converted to BAM using SAMtools. The scripts from the SPLASH computational pipeline (*73*) were used to identify the chimeric RNA from the BAM files. The "find_chimeras.py" script was used to detect chimeric RNA. In the case of "intramolecular" chimeric RNA, the ligated RNAs were required to be spaced at least 15 nt apart in the same transcript.

Statistical analysis

For all graphs in Figs. 3A, 3C, 4F, 6A and 8B and figs. S11, S15, S22 (C and D) and S24, the results are shown as means (SEM, $n = 3$), and all statistical analyses are done with two-way analysis of variance (ANOVA; Dunnett's multiple comparisons test). For all graphs in Figs. 5 (C to G) and 7 (A, C and D) and figs. S16A, S19A and S22A, the results are shown as means (SEM, $n = 3$), and all statistical analyses are done with two-way ANOVA (Sidak's multiple comparisons test). For the graph in Fig. 8C, the results are shown as means (SEM, $n = 3$), and all statistical analyses are done with two-way ANOVA (Tukey's multiple comparisons test). For the graph in Fig. 5A, the results are shown as means (SEM, $n = 3$), and all statistical analyses were done using multiple *t* test. For all graphs in Figs. 4B and 8 (E and F) and fig. S18, the results are shown as means (SEM, $n = 3$), and all statistical analyses were done with a two-tailed paired *t* test. For graphs in Fig. 6 (B and C), the results are shown as means (SEM, $n = 4$), and statistical analyses are done with log-rank (Mantel-Cox) test.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn2706

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. E. Vassella, B. Reuner, B. Yutzy, M. Boshart, Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J. Cell Sci.* **110**, 2661–2671 (1997).

2. E. Silvester, K. McWilliam, K. Matthews, The cytological events and molecular control of life cycle development of *Trypanosoma brucei* in the mammalian bloodstream. *Pathogens* **6**, 29 (2017).
3. B. M. Mony, P. MacGregor, A. Ivens, F. Rojas, A. Cowton, J. Young, D. Horn, K. Matthews, Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. *Nature* **505**, 681–685 (2014).
4. L. McDonald, M. Cayla, A. Ivens, B. Mony, P. MacGregor, E. Silvester, K. McWilliam, K. R. Matthews, Non-linear hierarchy of the quorum sensing signalling pathway in bloodstream form African trypanosomes. *PLOS Pathog.* **14**, e1007145 (2018).
5. B. M. Mony, K. R. Matthews, Assembling the components of the quorum sensing pathway in African trypanosomes. *Mol. Microbiol.* **96**, 220–232 (2015).
6. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
7. K. V. Morris, J. S. Mattick, The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437 (2014).
8. K. S. Rajan, T. Doniger, S. Cohen-Chalamish, P. Rengaraj, B. Galili, S. Aryal, R. Unger, C. Tschudi, S. Michaeli, Developmentally regulated novel non-coding anti-sense regulators of mRNA translation in *Trypanosoma brucei*. *iScience* **23**, 101780 (2020).
9. S. Hombach, M. Kretz, Non-coding RNAs: Classification, biology and functioning. *Adv. Exp. Med. Biol.* **937**, 3–17 (2016).
10. J. L. Rinn, H. Y. Chang, Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
11. J. Carlevaro-Fita, R. Johnson, Global positioning system: Understanding long noncoding RNAs through subcellular localization. *Mol. Cell* **73**, 869–883 (2019).
12. S. Geisler, J. Coller, RNA in unexpected places: Long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712 (2013).
13. R. W. Yao, Y. Wang, L. L. Chen, Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **21**, 542–551 (2019).
14. R. A. Flynn, H. Y. Chang, Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* **14**, 752–761 (2014).
15. J. D. Ransohoff, Y. Wei, P. A. Khavari, The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* **19**, 143–157 (2018).
16. M. Kretz, D. E. Webster, R. J. Flockhart, C. S. Lee, A. Zehnder, V. Lopez-Pajares, K. Qu, G. X. Y. Zheng, J. Chow, G. E. Kim, J. L. Rinn, H. Y. Chang, Z. Siprashvili, P. A. Khavari, Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev.* **26**, 338–343 (2012).
17. M. Kretz, Z. Siprashvili, C. Chu, D. E. Webster, A. Zehnder, K. Qu, C. S. Lee, R. J. Flockhart, A. F. Groff, J. Chow, D. Johnston, G. E. Kim, R. C. Spitale, R. A. Flynn, G. X. Y. Zheng, S. Aiyer, A. Raj, J. L. Rinn, H. Y. Chang, P. A. Khavari, Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**, 231–235 (2013).
18. I. Amit-Avraham, G. Pozner, S. Eshar, Y. Fastman, N. Kolevzon, E. Yavin, R. Dzikowski, Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E982–E991 (2015).
19. J. Guizetti, A. Barcons-Simon, A. Scherf, Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res.* **44**, 9710–9718 (2016).
20. K. L. Menard, B. E. Haskins, A. P. Colombo, E. Y. Denkers, *Toxoplasma gondii* manipulates expression of host long noncoding RNA during intracellular infection. *Sci. Rep.* **8**, 15017 (2018).
21. N. G. Kolev, J. B. Franklin, S. Carmi, H. Shi, S. Michaeli, C. Tschudi, The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLOS Pathog.* **6**, e1001090 (2010).
22. M. Aslett, C. Aurrecoechea, M. Berriman, J. Brestelli, B. P. Brunk, M. Carrington, D. P. Depledge, S. Fischer, B. Gajria, X. Gao, M. J. Gardner, A. Gingle, G. Grant, O. S. Harb, M. Heiges, C. Hertz-Fowler, R. Houston, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, F. J. Logan, J. A. Miller, S. Mitra, P. J. Myler, V. Nayak, C. Pennington, I. Phan, D. F. Pinney, G. Ramasamy, M. B. Rogers, D. S. Roos, C. Ross, D. Sivam, D. F. Smith, G. Srinivasamoorthy, C. J. Stoeckert, S. Subramanian, R. Thibodeau, A. Tivey, C. Treatman, G. Velarde, H. Wang, TriTrypDB: A functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38**, D457–D462 (2010).
23. S. Alsford, D. J. Turner, S. O. Obado, A. Sanchez-Flores, L. Glover, M. Berriman, C. Hertz-Fowler, D. Horn, High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* **21**, 915–924 (2011).
24. F. Constanty, A. Shkumatava, lncRNAs in development and differentiation: From sequence motifs to functional characterization. *Development* **148**, dev182741 (2021).
25. L. Statello, C. J. Guo, L. L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
26. S. K. Gupta, I. Kosti, G. Plaut, A. Pivko, I. D. Tkacz, S. Cohen-Chalamish, D. K. Biswas, C. Wachtel, H. Waldman Ben-Asher, S. Carmi, F. Glaser, Y. Mandel-Gutfreund, S. Michaeli, The hnRNP F/H homologue of *Trypanosoma brucei* is differentially expressed in the two life cycle stages of the parasite and regulates splicing and mRNA stability. *Nucleic Acids Res.* **41**, 6577–6594 (2013).
27. M. Z. Stern, S. K. Gupta, M. Salmon-Divon, T. Haham, O. Barda, S. Levi, C. Wachtel, T. W. Nilsen, S. Michaeli, Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA* **15**, 648–665 (2009).
28. A. M. Estévez, The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res.* **36**, 4573–4586 (2008).
29. S. M. Fernández-Moya, A. García-Pérez, S. Kramer, M. Carrington, A. M. Estévez, Alterations in DRBD3 ribonucleoprotein complexes in response to stress in *Trypanosoma brucei*. *PLOS ONE* **7**, e48870 (2012).
30. X. H. Liang, S. Uliel, A. Hury, S. Barth, T. Doniger, R. Unger, S. Michaeli, A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. *RNA* **11**, 619–645 (2005).
31. V. Chikne, T. Doniger, K. S. Rajan, O. Bartok, D. Eliaz, S. Cohen-Chalamish, C. Tschudi, R. Unger, Y. Hashem, S. Kadener, S. Michaeli, A pseudouridylation switch in rRNA is implicated in ribosome function during the life cycle of *Trypanosoma brucei*. *Sci. Rep.* **6**, 25296 (2016).
32. C. Batram, N. G. Jones, C. J. Janzen, S. M. Markert, M. Engstler, Expression site attenuation mechanistically links antigenic variation and development in *Trypanosoma brucei*. *eLife* **3**, e02324 (2014).
33. S. Dean, R. Marchetti, K. Kirk, K. R. Matthews, A surface transporter family conveys the trypanosome differentiation signal. *Biochemistry* **459**, 213–217 (2009).
34. S. Trindade, F. Rijo-Ferreira, T. Carvalho, D. Pinto-Neves, F. Guegan, F. Aresta-Branco, F. Bento, S. A. Young, A. Pinto, J. Van Den Abbeele, R. M. Ribeiro, S. Dias, T. K. Smith, L. M. Figueiredo, *Trypanosoma brucei* parasites occupy and functionally adapt to the adipose tissue in mice. *Cell Host Microbe* **19**, 837–848 (2016).
35. J. Kufel, P. Grzechnik, Small nucleolar RNAs tell a different tale. *Trends Genet.* **35**, 104–117 (2019).
36. S. Massenet, E. Bertrand, C. Verheggen, Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol.* **14**, 680–692 (2017).
37. K. S. Rajan, Y. Zhu, K. Adler, T. Doniger, S. Cohen-Chalamish, A. Srivastava, M. Shalev-Benami, D. Matzov, R. Unger, C. Tschudi, A. Günzl, G. G. Carmichael, S. Michaeli, The large repertoire of 2′-O-methylation guided by C/D snoRNAs on Trypanosoma brucei rRNA. *RNA Biol.* **17**, 1018–1039 (2020).
38. B. A. Elliott, H. T. Ho, S. V. Ranganathan, S. Vangaveti, O. Ilkayeva, H. Abou Assi, A. K. Choi, P. F. Agris, C. L. Holley, Modification of messenger RNA by 2′-O-methylation regulates gene expression in vivo. *Nat. Commun.* **10**, 3401 (2019).
39. E. Silvester, J. Young, A. Ivens, K. R. Matthews, Interspecies quorum sensing in co-infections can manipulate trypanosome transmission potential. *Nat. Microbiol.* **2**, 1471–1479 (2017).
40. A. A. Zimta, A. B. Tigu, C. Braicu, C. Stefan, C. Ionescu, I. Berindan-Neagoe, An emerging class of long non-coding RNA with oncogenic role arises from the snoRNA host genes. *Front. Oncol.* **10**, 389 (2020).
41. G. T. Williams, F. Farzaneh, Are snoRNAs and snoRNA host genes new players in cancer? *Nat. Rev. Cancer* **12**, 84–88 (2012).
42. K. S. Rajan, T. Doniger, S. Cohen-Chalamish, D. Chen, O. Semo, S. Aryal, E. Glick Saar, V. Chikne, D. Gerber, R. Unger, C. Tschudi, S. Michaeli, Pseudouridines on *Trypanosoma brucei* spliceosomal small nuclear RNAs and their implication for RNA and protein interactions. *Nucleic Acids Res.* **47**, 7633–7647 (2019).
43. V. M. Howick, L. Peacock, C. Kay, C. Collett, W. Gibson, K. N. Lawniczak, A. Health, M. Programme, Single-cell transcriptomics reveals expression profiles of, 1–30 (2021).
44. M. Engstler, M. Boshart, Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. *Genes Dev.* **18**, 2798–2811 (2004).
45. E. Wirtz, S. Leal, C. Ochatt, G. A. M. Cross, A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **99**, 89–101 (1999).
46. X. Yang, L. M. Figueiredo, A. Espinal, E. Okubo, B. Li, RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*. *Cell* **137**, 99–109 (2009).
47. T. N. Siegel, K. Gunasekera, G. A. Cross, T. Ochsenreiter, Gene expression in *Trypanosoma brucei*: Lessons from high-throughput RNA sequencing. *Trends Parasitol.* **27**, 434–441 (2011).
48. S. M. Kielbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
49. Y. J. Kang, D. C. Yang, L. Kong, M. Hou, Y. Q. Meng, L. Wei, G. Gao, CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
50. J. J. Vasquez, C. C. Hon, J. T. Vanselow, A. Schlosser, T. N. Siegel, Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res.* **42**, 3623–3637 (2014).
51. M. Dejung, I. Subota, F. Bucerius, G. Dindar, A. Freiwald, M. Engstler, M. Boshart, F. Butter, C. J. Janzen, Quantitative proteomics uncovers novel factors involved in developmental differentiation of *Trypanosoma brucei*. *PLOS Pathog.* **12**, e1005439 (2016).
52. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

53. A. Bilska, M. Kusio-Kobiałka, P. S. Krawczyk, O. Gewartowska, B. Tarkowski, K. Kobyłecki, D. Nowis, J. Golab, J. Gruchota, E. Borsuk, A. Dziembowski, S. Mroczek, Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. *Nat. Commun.* **11**, 2032 (2020).
54. R. E. Workman, A. D. Tang, P. S. Tang, M. Jain, J. R. Tyson, R. Razaghi, P. C. Zuzarte, T. Gilpatrick, A. Payne, J. Quick, N. Sadowski, N. Holmes, J. G. de Jesus, K. L. Jones, C. M. Soulette, T. P. Snutch, N. Loman, B. Paten, M. Loose, J. T. Simpson, H. E. Olsen, A. N. Brooks, M. Akeson, W. Timp, Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
55. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
56. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
57. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
58. L. Wang, S. Wang, W. Li, RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
59. F. Rijo-Ferreira, D. Pinto-Neves, N. L. Barbosa-Morais, J. S. Takahashi, L. M. Figueiredo, *Trypanosoma brucei* metabolism is under circadian control. *Nat. Microbiol.* **2**, 17032 (2017).
60. F. Aresta-Branco, S. Pimenta, L. M. Figueiredo, A transcription-independent epigenetic mechanism is associated with antigenic switching in *Trypanosoma brucei*. *Nucleic Acids Res.* **44**, 3131–3146 (2016).
61. S. Slomovic, G. Schuster, Chapter Thirteen-Circularized RT-PCR (cRT-PCR): Analysis of the 5′ ends, 3′ ends, and poly(A) tails of RNA, in *Laboratory Methods in Enzymology: RNA*, J. Lorsch, Ed. (Methods in Enzymology, Academic Press, 2013), vol. 530, pp. 227–251.
62. S. Kelly, J. Reed, S. Kramer, L. Ellis, H. Webb, J. Sunter, J. Salje, N. Marinsek, K. Gull, B. Wickstead, M. Carrington, Functional genomics in *Trypanosoma brucei*: A collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol. Biochem. Parasitol.* **154**, 103–109 (2007).
63. V. D. Atayde, E. Ullu, N. G. Kolev, A single-cloning-step procedure for the generation of RNAi plasmids producing long stem-loop RNA. *Mol. Biochem. Parasitol.* **184**, 55–58 (2012).
64. Z. Wang, J. C. Morris, M. E. Drew, P. T. Englund, Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *J. Biol. Chem.* **275**, 40174–40179 (2000).
65. S. Biebinger, L. Elizabeth Wirtz, P. Lorenz, C. Clayton, Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **85**, 99–112 (1997).
66. R. Brun, Schonenberger, Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication. *Acta Trop.* **36**, 289–292 (1979).
67. D. Eliaz, S. Kannan, H. Shaked, G. Arvatz, I. D. Tkacz, L. Binder, H. Waldman Ben-Asher, U. Okalang, V. Chikne, S. Cohen-Chalamish, S. Michaeli, Exosome secretion affects social motility in *Trypanosoma brucei*. *PLOS Pathog.* **13**, e1006245 (2017).
68. I. D. Tkacz, S. Cohen, M. Salmon-Divon, S. Michaeli, Identification of the heptameric Lsm complex that binds U6 snRNA in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **160**, 22–31 (2008).
69. N. G. Kolev, K. S. Rajan, K. T. Tycowski, J. Y. Toh, H. Shi, Y. Lei, S. Michaeli, C. Tschudi, The vault RNA of *Trypanosoma brucei* plays a role in the production of trans-spliced mRNA. *J. Biol. Chem.* **294**, 15559–15574 (2019).
70. U. Birkedal, M. Christensen-Dalsgaard, N. Krogh, R. Sabarinathan, J. Gorodkin, H. Nielsen, Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed.* **54**, 451–455 (2015).
71. T. Magoc, S. L. Salzberg, FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
72. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
73. J. G. A. Aw, Y. Shen, A. Wilm, M. Sun, X. N. Lim, K. L. Boon, S. Tapsin, Y. S. Chan, C. P. Tan, A. Y. L. Sim, T. Zhang, T. T. Susanto, Z. Fu, N. Nagarajan, Y. Wan, In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* **62**, 603–617 (2016).
74. P. Capewell, S. Monk, A. Ivens, P. MacGregor, K. Fenn, P. Walrad, F. Bringaud, T. K. Smith, K. R. Matthews, Regulation of *Trypanosoma brucei* total and polysomal mRNA during development within its mammalian host. *PLOS ONE* **8**, e67069 (2013).

Acknowledgments: We would like to thank M. Machado for valuable help with the in vivo mouse *T. brucei* infection experiments. We also thank H. Manso, A. R. Grosso, and N. Barbosa Morais for valuable help with computational analysis; L. Pinho for managing the orders, the production of culture medium, and the laboratory environment; and all members of the laboratory for stimulating discussion during this project. **Funding:** This work was supported in part by Fundação para a Ciência e Tecnologia (FCT) grant, awarded to F.G. and entitled "Long noncoding RNAs as new diagnostic biomarkers for African Sleeping sickness" (PTDC/DTPEPI/7099/2014, start date: 1 January 2016, end date: 31 December 2018); also by Howard Hughes Medical Institute International Early Career Scientist Program (project title: "How parasites use epigenetics to evade host defenses," project no. 55007419, start date: 1 February 2012, end date: 31 January 2017); and by the European Research Council (project title: "Exploring the hidden life of African trypanosomes: parasite fat tropism and implications for the disease," project no. 771714, start date: 1 August 2018, end date: 31 January 2024), both awarded to L.M.F. The project leading to these results have received funding from "la Caixa" Foundation under the agreement LCF/PR/HR20/52400019 [project title: "Mechanism and function of epitranscriptomic poly(A) tail modifications in African trypanosomes," project no. HR20-00361, start date: 1 March 2021, end date: 29 February 2024]. L.M.F. is supported by FCT (IF/01050/2014, project title: "Molecular basis for the efficient biology of trypanossome parasitism," start date: 1 January 2015, end date: 31 December 2019) and by CEEC institutional program (CEECINST/00110/2018, start date: 1 January 2020, end date: 14 December 2020). C.N. acknowledges the support of the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya. S. Michaeli acknowledges the support of the Israel Science Foundation (ref. 1959/20) from October 2020 to October 2025, entitled "Functional analysis of rRNA processing and the role of rRNA modification for specialized translation in the two life stages of trypanosomes" and U.S. Binational Science Foundation (ref. 2015/219) from October 2015 to October 2019, entitled "The role and mechanism of RNA pseudo-uridylation and sugar methylation (Nm) during the developmental cycle of trypanosomes." The work done in A.D.'s laboratory was supported by National Science Center SONATA BIS grant, entitled "Non-canonical RNA tailing and other post-transcriptional regulatory mechanisms in T cell-mediated adaptive immunity" (proposal ID: 492777, agreement no: UMO-2020/38/E/NZ2/00372, start date: 22 March 2021, end date: 21 March 2026); National Science Center OPUS grant, entitled "Analysis of the role of cytoplasmic polyadenylation in the regulation of the innate immune response" (proposal ID: 443521, agreement no.: UMO-2019/33/B/NZ2/01773, start date: 2 March 2020, end date: 1 March 2023); and European Union's Horizon 2020 (H2020-WIDESPREAD-03-2017)–ERAChair, entitled "MOlecular Signaling in Health and Disease - Interdisciplinary Centre of Excellence" (acronym: MOSaIC, agreement no.: 810425, implementation period: start date: 1 November 2018, end date: 31 October 2023). **Author contributions:** F.G., C.N., S. Michaeli, and L.M.F. designed the study. F.G., F.B., D.P.-N., M.S., K.S.R., N.G., S. Mroczek, A.D., S.C.-C., T.D., B.G., and A.M.E. designed and performed the experiments. F.G., D.P.-N., and L.M.F. analyzed the data and wrote the initial draft of the manuscript. F.G. and L.M.F. supervised the project. All authors edited and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The RNA-seq data from this study have been submitted to the European Nucleotide Archive–PRJEB38238. The Nanopore direct RNA-seq data from this study have been submitted to the European Nucleotide Archive–PRJEB48655.

Submitted 19 November 2021
Accepted 28 April 2022
Published 15 June 2022
10.1126/sciadv.abn2706

ScienceAdvances

A long noncoding RNA promotes parasite differentiation in African trypanosomes

Fabien GueganK. Shanmugha RajanFábio BentoDaniel Pinto-NevesMariana SequeiraNatalia Gumi#skaSeweryn MroczekAndrzej DziembowskiSmadar Cohen-ChalamishTirza DonigerBeathrice GaliliAntonio M. EstévezCedric NotredameShulamit MichaeliLuisa M. Figueiredo

Sci. Adv., 8 (24), eabn2706. • DOI: 10.1126/sciadv.abn2706

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.