PNAS RESEARCH ARTICLE | MICROBIOLOGY

Life stage–specific poly(A) site selection regulated by *Trypanosoma brucei* DRBD18

Jonathan E. Bard[a,b,1], Brianna L. Tylec[c,1,2], Ashutosh P. Dubey[c], Natalie A. Lamb[a,3], Donald A. Yergeau[a], and Laurie K. Read[c,4]

Affiliations are included on p.11.

Edited by Margaret Phillips, The University of Texas Southwestern Medical Center, Dallas, TX; received February 21, 2024; accepted June 7, 2024

The kinetoplastid parasite, *Trypanosoma brucei*, undergoes a complex life cycle entailing slender and stumpy bloodstream forms in mammals and procyclic and metacyclic forms (MFs) in tsetse fly hosts. The numerous gene regulatory events that underlie *T. brucei* differentiation between hosts, as well as between active and quiescent stages within each host, take place in the near absence of transcriptional control. Rather, differentiation is controlled by RNA-binding proteins (RBPs) that associate with mRNA 3′ untranslated regions (3′UTRs) to impact RNA stability and translational efficiency. DRBD18 is a multifunctional *T. brucei* RBP, shown to impact mRNA stability, translation, export, and processing. Here, we use single-cell RNAseq to characterize transcriptomic changes in cell populations that arise upon DRBD18 depletion, as well as to visualize transcriptome-wide alterations to 3′UTR length. We show that in procyclic insect stages, DRBD18 represses expression of stumpy bloodstream form and MF transcripts. Additionally, DRBD18 regulates the 3′UTR lengths of over 1,500 transcripts, typically promoting the use of distal polyadenylation sites, and thus the inclusion of 3′UTR regulatory elements. Remarkably, comparison of polyadenylation patterns in DRBD18 knockdowns with polyadenylation patterns in stumpy bloodstream forms shows numerous similarities, revealing a role for poly(A) site selection in developmental gene regulation, and indicating that DRBD18 controls this process for a set of transcripts. RNA immunoprecipitation supports a direct role for DRBD18 in poly(A) site selection. This report highlights the importance of alternative polyadenylation in *T. brucei* developmental control and identifies a critical RBP in this process.

alternative polyadenylation | trypanosome | kinetoplastid | RNA-binding protein | life cycle

Significance

Trypanosoma brucei causes deadly diseases in humans and other mammals. RNA-binding proteins (RBPs) control differentiation during the *T. brucei* life cycle, which involves mammalian and insect hosts. RBPs act by incompletely understood mechanisms. Here, we show that in the procyclic insect stage, the RBP DRBD18 (Double RNA Binding Domain 18) represses expression of transcripts specific to stumpy bloodstream and metacyclic insect forms. Additionally, DRBD18 controls sites of polyadenylation on many of these transcripts, such that 3′ untranslated regions are shortened upon DRBD18 depletion. Sites of polyadenylation in DRBD18 depleted cells resemble those in stumpy bloodstream forms. This study provides evidence that poly(A) site selection is important in *T. brucei* differentiation and controlled by DRBD18 for some developmentally regulated transcripts.

Trypanosoma brucei is a single-celled, kinetoplastid parasite and the causative agent of Human African Trypanosomaisis and the livestock disease, nagana (1). As it progresses through its complex life cycle, *T. brucei* transitions between mammalian and tsetse fly hosts and between replicative and nonreplicative forms in both hosts. Remarkably, the dramatic changes in gene expression that drive these developmental cycles take place in the near absence of transcriptional control. Instead, RNA polymerase II transcription is polycistronic, and monocistrons are generated through 5′ *trans*-splicing and 3′ cleavage and polyadenylation. The lack of transcriptional control necessitates that gene regulation take place primarily posttranscriptionally, often at the levels of mRNA stability or translational efficiency (2). Thus, in *T. brucei*, as in other kinetoplastids, RNA-binding proteins (RBPs) constitute the critical gene regulatory *trans*-acting factors. RBPs act by binding key *cis*-acting elements that typically reside in mRNA 3′UTRs (2). Several RBPs that impact developmental progression in *T. brucei* have been described (3–7), although in most cases, their mechanisms of action are poorly understood.

Because 3′UTRs are the critical regulatory regions in *T. brucei* messenger (m)RNAs, it may not be surprising that they are often quite long, having a median length of around 400 nucleotides, and ranging to many thousands of nucleotides (8, 9). The handful of 3′UTRs that have been functionally characterized often contain numerous positive and negative regulatory elements, affecting both stability and translational efficiency (2, 10–18). Genome-wide studies showed that many transcripts exhibit heterogeneous poly(A) addition sites that differ by hundreds of nucleotides (8, 9), although these studies were not comprehensive with regard to 3′UTRs. Polyadenylation at alternative sites can lead to exclusion or inclusion of 3′UTR gene regulatory elements with dramatic functional consequences. In mammals, alternative polyadenylation is widespread and impacts metabolism, autophagy, cell differentiation, and development (19–22). Because kinetoplastid gene expression is primarily regulated by 3′UTR elements, alternative polyadenylation has an outsized potential to influence the transcriptome and proteome and promote responses to extracellular cues and developmental progression in these organisms. The *T. brucei* polyadenylation machinery is fairly conventional (23, 24), and polyadenylation of a given transcript is linked to 5′ *trans*-splicing of the downstream transcript (25). Both 5′ and 3′ precursor transcript processing requires a polypyrimidine tract, and many

Author contributions: J.E.B., D.A.Y., and L.K.R. designed research; J.E.B., B.L.T., A.P.D., N.A.L., and D.A.Y. performed research; J.E.B., B.L.T., A.P.D., N.A.L., and L.K.R. analyzed data; and J.E.B., B.L.T., and L.K.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]J.E.B. and B.L.T. contributed equally to this work.

[2]Present address: Foresight Diagnostics, Boulder, CO 80301.

[3]Present address: Bioenergy Science and Technology, National Renewable Energy Laboratory, Golden, CO 80401.

[4]To whom correspondence may be addressed. Email: lread@buffalo.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2403188121/-/DCSupplemental.

Published July 11, 2024.

intergenic regions harbor several such sequence elements that could signal upstream 3′ polyadenylation (2, 26). However, despite the likely importance of alternative polyadenylation in *T. brucei*, almost nothing is known regarding how poly(A) site selection is regulated or the impact of heterogeneous 3′UTR isoforms on trypanosome biology.

A few RBPs reportedly regulate the extent of *trans*-splicing of a subset of mRNAs (27, 28), and several SR proteins can alter the site of *trans*-splicing when tethered to a RNA (29) However, the only factor implicated to date in poly(A) site selection in *T. brucei*, and which was shown to do so in vivo, is DRBD18 (30). DRBD18 is abundant and essential in both mammalian bloodstream (BF) and insect procyclic (PF) forms of *T. brucei* (30, 31). This multifunctional protein is present in both nuclei and cytoplasm, dramatically impacts the transcriptomes of both BF and PF (30, 31), and plays likely direct roles in both translation (32) and nuclear export (33) of specific transcripts. It was recently shown that DRBD18 knockdown influences poly(A) site selection for a subset of mRNAs in both BF and PF *T. brucei* (30), although these authors note that many sites of alternative polyadenylation may have been missed in their study due to incomplete annotation of the genome. Nevertheless, by comparing the effects of DRBD18 depletion on 3′UTR reads to effects within the corresponding coding regions, they identified over 250 transcripts with apparent 3′UTR shortening, and these substantially overlapped between BF and PF. RNA immunoprecipitation and high throughput sequencing (RIP-seq) experiments in BF showed that transcripts whose 3′UTRs were altered were more highly enriched than those that were unchanged. Together, these results suggest a direct effect on DRBD18 on poly(A) site selection, at least in BF.

Here, we further examine the function of DRBD18 in *T. brucei*, using single-cell RNA sequencing (scRNAseq) of uninduced and induced DRBD18 knockdowns. In PF, we identified one major and one minor population of cells that dramatically increase upon DRBD18 knockdown. Both of these cell populations are significantly enriched for transcripts that are normally up-regulated in the quiescent stumpy BF and metacyclic form (MF) life cycle stages (34, 35), consistent with a role for DRBD18 in life cycle progression. Furthermore, the 3′ end bias of our sequencing technique allowed us to investigate the effect of DRBD18 on poly(A) site selection, and extension of gene annotations permitted a comprehensive analysis of these changes. We found that 3′UTR shortening upon DRBD18 depletion was prevalent, with over 1,500 transcripts exhibiting this phenotype in PF, and RIP-seq analyses showed significant overlap between bound transcripts and those with 3′UTR shortening. Supporting a connection between 3′UTR shortening and life cycle–specific gene regulation, those genes most highly expressed in stumpy BF and MF cells are significantly overrepresented in the lists of transcripts with shortened 3′UTRs upon DRBD18 RNAi. Comparison of polyadenylation patterns in uninduced and induced PF DRBD18 knockdowns and the same transcripts analyzed by similar methods in stumpy BF (36) revealed extensive life cycle–specific changes in polyadenylation sites and showed that many stumpy BF transcripts utilize the same polyadenylation sites that arise in DRBD18-depleted PF. Overall, these data support a model in which DRBD18 mediates life cycle stage-specific patterns of gene regulation by modulation of poly(A) site selection.

Results

scRNAseq Defines Distinct Populations of DRBD18-Depleted Cells. Knockdown of DRBD18 in PF *T. brucei* leads to substantial dysregulation of the transcriptome, including increased abundance of numerous transcripts encoding RBPs, kinases, and phosphatases (*SI Appendix*, Fig. S1 and Dataset S1) (31, 33). Thus, perturbation of DRBD18 expression could result in different signaling cascades being activated in different cells leading to cell-specific transcriptome changes undetectable by bulk RNA sequencing. To investigate this possibility, we performed scRNAseq with the initial question of whether distinct cell states emerge upon DRBD18 knockdown. We utilized a published cell line containing a doxycycline (dox)-inducible RNAi construct targeting DRBD18 (Fig. 1*A*) (31). Biological replicate cultures were grown for 18 h with or without dox and subjected to scRNAseq capture and Illumina sequencing. We obtained data for 20,287 uninduced and 21,425 RNAi-induced cells, with a mean read depth per cell of 45,866 and 739 median genes per cell. Replicates were highly reproducible, with R values >0.99 (*SI Appendix*, Fig. S2). We noted that many genes had reads downstream of the 3′ ends annotated in TriTrypDB; thus, to capture all 3′ end data, we modified the TriTrypv46 gene annotation file to extend each transcript at its 3′ end to the annotated 5′ end of the downstream gene. Uniform Manifold Approximation and Projection (UMAP)-dimensional reduction and clustering (37) was performed, resulting in eight total clusters (denoted 0 to 7; Fig. 1*B*), with significant cell-to-cell heterogeneity observed even in the uninduced population. Upon DRBD18 depletion, the cell distribution shifts to populate the upper left region of the UMAP. We calculated the distribution of –dox and +dox cells by cluster (Fig. 1*C*), which confirmed a decrease in the sizes of clusters 1 through six upon DRBD18 knockdown and a corresponding increase in the size of clusters 0 and 7. Cluster 0, the largest group, is composed of almost entirely dox-treated (+dox) cells (94.78%), and the majority of +dox cells (55.84%) belong to cluster 0. Cluster 7, the smallest group, also contains a majority (71.57%) of +dox cells and is more populous upon DRBD18 knockdown. Thus, depletion of DRBD18 drives cells toward one major state and one minor state.

To establish cluster markers, we used Seurat (38) to identify transcripts that are differentially expressed (log2 FC > 0.25) in each cluster relative to the remaining combined population of cells (Dataset S2). Clusters 1 and 3 each contained only two enriched marker genes, indicating that their transcriptomes are less distinct from the rest of the population relative to other clusters. We analyzed the marker genes of the other clusters using a Gene Ontology enrichment tool (TriTrypDB) and GO-Figure! (39) (Fig. 1*D* and Dataset S3). Fig. 1*D* shows enriched GO terms with FDR < 0.01, with some exceptions for GO terms of particular interest that did not pass the 0.01 threshold. With regard to cells that arise upon DRBD18 knockdown, the enriched GO terms in cluster 0 marker genes corroborate previous bulk RNAseq findings: Protein phosphorylation, dephosphorylation, and mRNA binding were among the most enriched terms. Interestingly, among cluster 0 marker genes, we found RBPs that are drivers of the BF and MF states, including RBP10 and RBP6, respectively (Dataset S2). For cluster 7, only posttranscriptional regulation of gene expression passed our <0.01 FDR threshold. Upon closer inspection, however, we noticed genes involved with quorum sensing and cell–cell signaling were enriched ($P = 0.0435$), suggesting some derepression of BF functionality in these cells as well. Further analysis of cluster 0 and cluster 7 cells revealed that several putative VSGs were expressed two- to nine-fold higher in cluster 7 than in cluster 0 (*SI Appendix*, Fig. S3), suggesting a role for DRBD18 in repression of VSG expression.

Cells comprising clusters 2 and 4 are significantly reduced upon DRBD18 knockdown; thus, these appear to contribute the most cells to the major DRBD18 knockdown cluster, cluster 0 (Fig. 1*C*). Cluster 2 is enriched in transcripts involved with normal cell

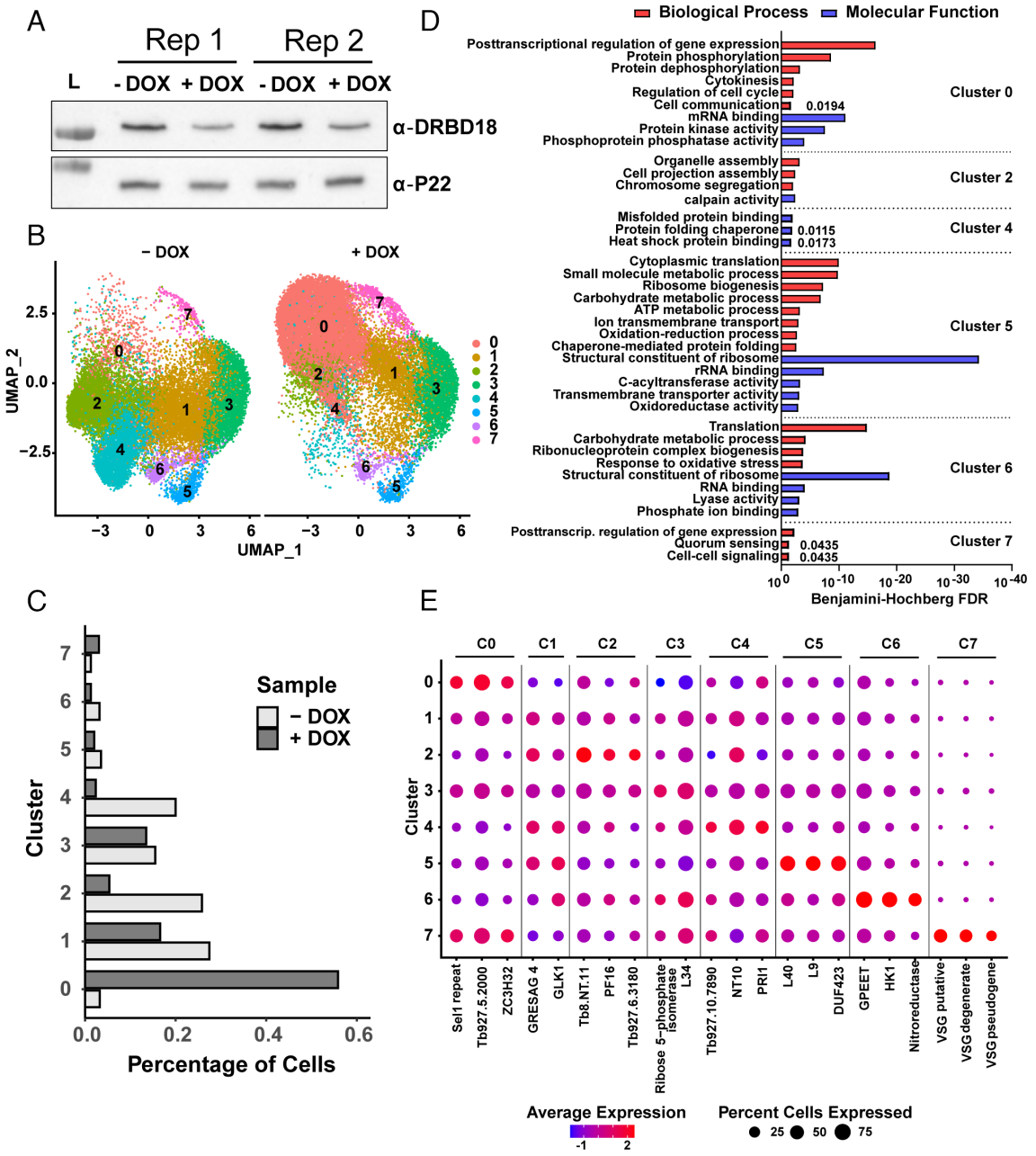


Fig. 1. scRNAseq analysis of DRBD18-depleted procyclic form *T. brucei*. (*A*) Anti-DRBD18 western blot of the duplicate samples used for scRNAseq. P22, loading control. L, ladder. (*B*) UMAP-dimensional reduction and clustering analysis of −DOX and +DOX scRNAseq datasets using Seurat. (*C*) Analysis of the total cell population, showing the proportions of −DOX and +DOX cells across clusters. (*D*) Enriched GO terms in cluster marker genes. (*E*) Relative expression of top cluster marker genes across all clusters.

growth and division, while cluster 4 is enriched in transcripts involved in misfolded protein binding, protein folding chaperone, heat shock protein binding, suggesting this cell population is experiencing stress (Fig. 1*D*). The cells of clusters 5 and 6 appear to be highly translationally and metabolically active, with slight differences in these parameters between the two clusters (Fig. 1*D*).

We next asked whether the most highly enriched transcripts in each cluster could reveal more about cluster characteristics and their relationships. We plotted the relative expression levels in each cluster of the top three marker genes from each cluster (with the exception of clusters 1 and 3, for which the sole two marker genes were plotted), and filtered to include those expressed in at least 25% of cells in that cluster (Fig. 1*E*). We note that the top two enriched marker genes for cluster 0 are 5.8S rRNA genes; however, upon closer inspection, we found that increased reads were only present in the intergenic regions (*SI Appendix*, Fig. S4), suggesting a modest impact of the knockdown on rRNA processing but not expression. The top protein coding genes in cluster 0 are essential SEL1-repeat containing and hypothetical proteins (Tb11.v5.0632 and Tb927.5.2000, respectively), and ZC3H32 (Tb927.10.5250), the latter being a repressive mRBP that is up-regulated and essential in BF *T. brucei* (40) (Fig. 1*E*). We also observed relatively high expression of these top cluster 0 marker genes in cluster 7. However, VSG-related cluster 7 marker genes were not enriched in cluster 0, again indicating that clusters 0 and 7 are similar, with the major exception of VSG-related genes (Fig. 1*E* and *SI Appendix*, Fig. S3). With regard to clusters that exhibit little change upon DRBD18 knockdown, GPEET is highly enriched in cluster 6, indicating that cluster 6 cells may be in the early PF state (34, 41) (Fig. 1*E*). This distinction may explain its separation from cluster 5 despite these

clusters having somewhat similar GO enrichment profiles. Top enriched genes in cluster 5 include two ribosomal proteins, in keeping with GO enrichment in cytoplasmic translation and ribosome biogenesis (Fig. 1*E*). Top markers for clusters 1 to 4 largely corresponded with their enriched GO terms; however, these clusters appeared somewhat less distinct than the other clusters, with top enriched marker genes exhibiting moderate expression in other clusters (Fig. 1*E*). Overall, we conclude that DRBD18 depletion results in the generation of two cell populations that are enriched for posttranscriptional gene regulatory protein mRNAs as well as several BF and MF up-regulated transcripts. As detailed below, separation of the cell populations that are the most affected by DRBD18 depletion from other cell types by scRNAseq now allows a more targeted examination of the effects of DRBD18 on gene regulation than previous bulk RNAseq studies.

Life Cycle–Specific Effects of DRBD18. The increased expression in cluster 0 of several transcripts that are highly expressed in BF and MF led us to further examine a role for DRBD18 in developmental progression of the *T. brucei* life cycle. To determine whether DRBD18 has a broad effect across developmentally regulated transcripts, we compared the cluster marker gene lists to lists of transcripts up-regulated in specific life cycle forms. We analyzed published transcriptomes of EATRO 1125 cells transitioning from slender BF to PF after treatment with citrate, *cis*-aconitate, and decreased temperature, including slender BF, stumpy BF, early PF and late PF (additional file 5 in ref. 34, which lists stage-specific transcripts enriched ≥twofold compared to all other stages). We also analyzed Lister 427 cells transitioning from PF to MF through RBP6 overexpression (35). To determine the overlaps of cluster marker genes with life cycle–specific genes, we performed a hypergeometric test of enrichment for each group comparison (Fig. 2*A*). Remarkably, cluster 0 and cluster 7 markers exhibited highly significant overlap with the sets of transcripts that are enriched in stumpy BF and those up-regulated during metacyclogenesis. Conversely, the marker genes of the clusters containing majority untreated cells (clusters 2, 4, 5, and 6) are enriched for genes that are repressed during metacyclogenesis. Finally, we found that genes enriched in early PF cells were highly enriched among cluster 6 markers, confirming the early PF nature of this cell cluster.

To quantify the enrichment of life cycle stage-specific transcripts in our scRNAseq dataset, we performed module scoring (42). This approach calculates the average expression of a provided list of genes, subtracting the average expression of a list of randomly selected control genes. On average, both mRNAs that are expressed preferentially in stumpy BF and those that are up-regulated during metacyclogenesis are more highly enriched in dox-treated cells than in untreated cells, regardless of cluster (Fig. 2*B*). We observed little to no difference between untreated and dox-treated cells for slender BF or late PF genes, as defined in stage-specific gene lists from ref. 34, despite the upregulation of a few slender BF genes when DRBD18 is knocked down (Dataset S1). Finally, the early PF genes are greatly expressed above background for cluster 6 regardless of treatment. Together, these data demonstrate that DRBD18-depleted cells derepress expression of transcripts that are normally up-regulated in both stumpy BF and MF. Thus, we conclude that DRBD18 plays a role in maintaining the PF state by repressing expression of transcripts specific to other life cycle stages.

DRBD18 Knockdown Alters Poly(A) Site Selection. While analyzing gene tracks in our dataset, we observed that numerous transcripts had more reads mapping to a 5′ (open reading frame (ORF)-proximal) position in our +dox samples than in our −dox samples. Because our library generation and sequencing protocol selected 200 to 400 bp oligo(dT)-primed fragments of which the 5′ 96 bp was sequenced, a shift in peak position suggests a change in poly(A) site selection, although it does not precisely map the poly(A) addition site. Several examples of transcripts whose polyadenylation patterns changed in +dox cells are shown in Fig. 3*A*. In some cases, the position of the majority of poly(A) sites shifted to a proximal site upon DRBD18 depletion (Tb927.11.6390 and Tb927.10.14390; Fig. 3*A*). In others, a new, more proximal, site was apparent in DRBD18-depleted cells, while distal sites used in uninduced cells remained highly utilized (Tb927.9.2700; Fig. 3*A*). Finally, numerous transcripts had complicated polyadenylation patterns and changes, including both increases and decreases in the sites used in uninduced cells as well as utilization of new sites upon DRBD18 depletion (Tb927.10.15040; Fig. 3*A*).

To understand how poly(A) site selection is influenced by DRBD18 in a transcriptome-wide manner, we sought to identify all of the transcripts in our dataset with differential signal position upon DRBD18 knockdown. To this end, we used DeepTools (43) to extract the read counts for each transcript, normalize their lengths 5′ to 3′, and then plot read expression with respect to normalized position within each gene. For this analysis, we used combined (pseudobulk) reads from Rep1, as data from the two replicates were highly reproducible (*SI Appendix*, Fig. S2). To estimate the effect of DRBD18 knockdown on poly(A) site selection, we calculated the ratio of +dox to −dox signal across the full length of each gene and used DeepTools to cluster transcripts according to their signal ratios. DeepTools then classified each transcript as a member of one of five groups, which we named Patterns A to E (Fig. 3*B* and Dataset S4). Fig. 3 *B*, *Top* plots the log2 ratios of signal across the scaled regions, clearly highlighting increased proximal signal in Patterns D (orange) and E (red) upon DRBD18 depletion. The heat map in Fig. 3 *B*, *Bottom* shows the values for individual transcripts, plotted as the +dox/−dox ratio, with red indicating higher signal in +dox and blue indicating higher signal in −dox. Transcripts in Pattern A generally have higher signal in −dox samples, indicating higher expression in −dox cells. Transcripts in Patterns B and C generally demonstrate more modest differences in signal between −dox and +dox. Transcripts in Patterns D and E have a substantially higher +dox signal at more ORF-proximal positions along the gene. These characteristics are illustrated with representative gene tracks of transcripts from patterns A, C, and E to the right of the heat map (Fig. 3*B*). These data are also displayed as "unlogged" plots in which individual values for −dox and +dox are shown (Fig. 3*C*), again illustrating the dramatic 5′ shift in reads in Pattern E transcripts and a more modest shift for Pattern D transcripts. While these patterns hold true in all scRNAseq clusters (*SI Appendix*, Fig. S5*A*), markers for majority +dox clusters 0 and 7 were most highly enriched for Pattern D and E transcripts (*SI Appendix*, Fig. S5*B*). Altogether, these data demonstrate that numerous transcripts exhibit a 5′, or ORF-proximal, shift in poly(A) addition site in DRBD18-depleted cells. Thus, DRBD18 promotes distal poly(A) site selection in a subset of transcripts in PF cells.

Altered Poly(A) Site Selection Correlates with Altered Transcript Abundance and Translational Efficiency. Usage of a more proximal poly(A) site could result in the exclusion of 3′UTR elements that affect the stability of the mRNA. To assess this possibility, we performed module scoring of transcripts comprising Patterns A to E (Fig. 4*A*). Pattern A and B transcripts are more highly enriched in −dox cells relative to +dox cells, and expression of Pattern C transcripts is unchanged upon DRBD18 depletion. By contrast, Pattern D and E transcripts are more highly enriched

A

	Slender 57 genes	Stumpy 103 genes	Early PF 34 genes	Late PF 13 genes	MF Up 703 genes	MF Down 609 genes
Cluster 0 236 genes	0 overlap p = 1	**20 overlap p = 1.46e-14**	1 overlap p = 0.5	1 overlap p = 0.23	**60 overlap p = 2.46e-22**	10 overlap p = 0.80
Cluster 1 2 genes	0 overlap p = 1	0 overlap p = 1	0 overlap p = 1	**1 overlap p = 2.23e-3**	0 overlap p = 1	1 overlap p = 0.10
Cluster 2 84 genes	0 overlap p = 1	1 overlap p = 0.53	1 overlap p = 0.22	**2 overlap p = 3.82e-3**	2 overlap p = 0.97	**26 overlap p = 5.88e-14**
Cluster 3 2 genes	0 overlap p = 1	0 overlap p = 1	0 overlap p = 1	0 overlap p = 1	0 overlap p = 1	1 overlap p = 0.10
Cluster 4 74 genes	1 overlap p = 0.31	1 overlap p = 0.48	1 overlap p = 0.20	**2 overlap p = 2.97e-3**	3 overlap p = 0.83	**26 overlap p = 1.83e-15**
Cluster 5 402 genes	0 overlap p = 1	1 overlap p = 0.97	**6 overlap p = 9.7e-4**	2 overlap p = 0.07	9 overlap p = 1	**123 overlap p = 7.81e-63**
Cluster 6 175 genes	1 overlap p = 0.57	**5 overlap p = 0.02**	**14 overlap p = 1.94e-17**	0 overlap p = 1	9 overlap p = 0.74	**55 overlap p = 1.32e-28**
Cluster 7 167 genes	1 overlap p = 0.56	**10 overlap p = 2.13e-06**	**4 overlap p = 1.36e-3**	0 overlap p = 1	**52 overlap p = 6.42e-24**	9 overlap p = 0.51

B

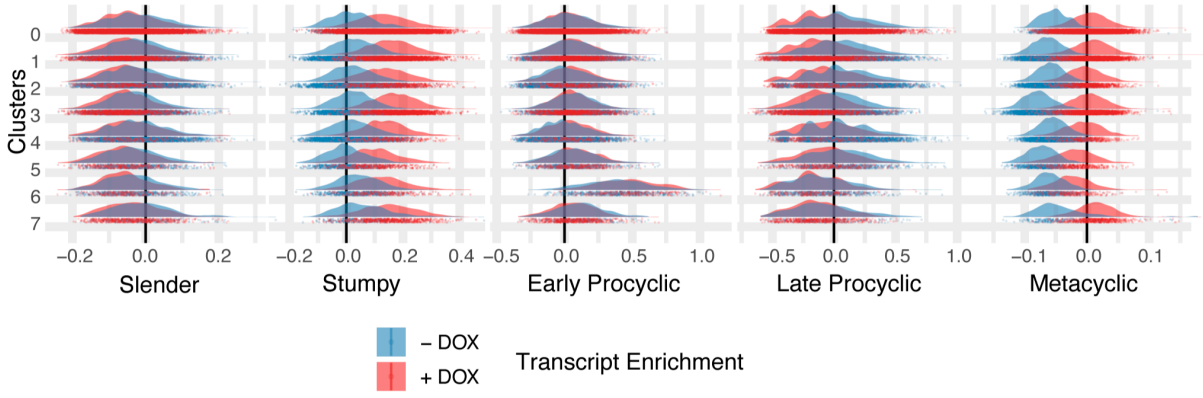


Fig. 2. Life stage–specific effects of DRBD18 depletion. (*A*) Stumpy BF-specific and MF-increased transcripts are highly overrepresented in clusters 0 and 7 as assessed using a hypergeometric test for overrepresentation. Light green: $P < 0.05$ and $>1 \times 10^{-5}$; dark green: $P < 1 \times 10^{-5}$. (*B*) Module scoring of life stage–associated genes. A module score was calculated for each cell and plotted as a violin plot above with single cells represented by dots below. The *x*-axis represents the average expression of the provided list of genes set against the average expression of a random selection of background genes. A module score of 0 means the provided list of genes is expressed at the same level as a selection of background control genes. A positive module score indicates higher expression than background genes; a negative module score indicates lower expression than background.

on average in +dox samples, indicating that the proximal shift in poly(A) site leads to increased stability of these mRNAs. We confirmed a general increase in transcript levels for Patterns D and E genes upon DRBD18 knockdown by comparing +dox/−dox transcript levels in the bulk RNAseq dataset from DRBD18 RNAi cells (*SI Appendix*, Fig. S1 and Dataset S1), calculating the average fold changes upon DRBD18 knockdown for genes in each 3′UTR Pattern, A to E (Fig. 4*B*). In keeping with Fig. 4*A*, Pattern A transcripts were expressed at a lower level in +dox cells, while Pattern D and E genes were expressed more highly on average in +dox cells (Fig. 4*B*). Moreover, Pattern D and E genes comprise the majority of the marker genes for clusters 0 (83.9%) and 7 (60.5%), suggesting their higher expression in these +dox clusters is related to 3′UTR shortening (*SI Appendix*, Fig. S5*B*). From these data, we conclude that the proximal poly(A) addition site shift upon DRBD18 RNAi in Pattern D and E transcripts generally results in increased mRNA expression, possibly due to exclusion of negative regulatory elements. To identify potential negative regulatory elements, we used STREME to analyze 3′UTRs from Pattern D and E transcripts for enriched motifs, with Pattern A transcripts serving as a negative control (44). The most significantly enriched motifs included AU-rich and AAGAW-containing elements reminiscent of those previously shown to mediate mRNA decay in *T. brucei* (10, 28) (Fig. 4 *C*, *Left*). Such motifs were often observed at positions between distal and proximal polyadenylation sites, consistent with their exclusion in some transcripts upon DRBD18 RNAi (Fig. 4 *C*, *Right* and *SI Appendix*, Fig. S6). From these data, we conclude that DRBD18 often acts as a repressor of mRNA stability by maintaining distal poly(A) addition sites and, likely, destabilization elements within long 3′UTRs.

Exclusion of 3′UTR elements could result in changes in other aspects of mRNA fate, including nuclear export or translational efficiency, both of which are affected by DRBD18 (32, 33). Therefore, we asked whether genes included in Patterns A to E are enriched in lists of transcripts with either altered translational efficiency or reduced nuclear export upon DRBD18 RNAi (32, 33) (Fig. 4*D*). Pattern D and E genes were significantly enriched in the set of 41 transcripts reported to have increased translational efficiency upon DRBD18 RNAi (TE up). This suggests that, for some transcripts, DRBD18-mediated 3′UTR shortening leads to increased translational efficiency through elimination of a negative

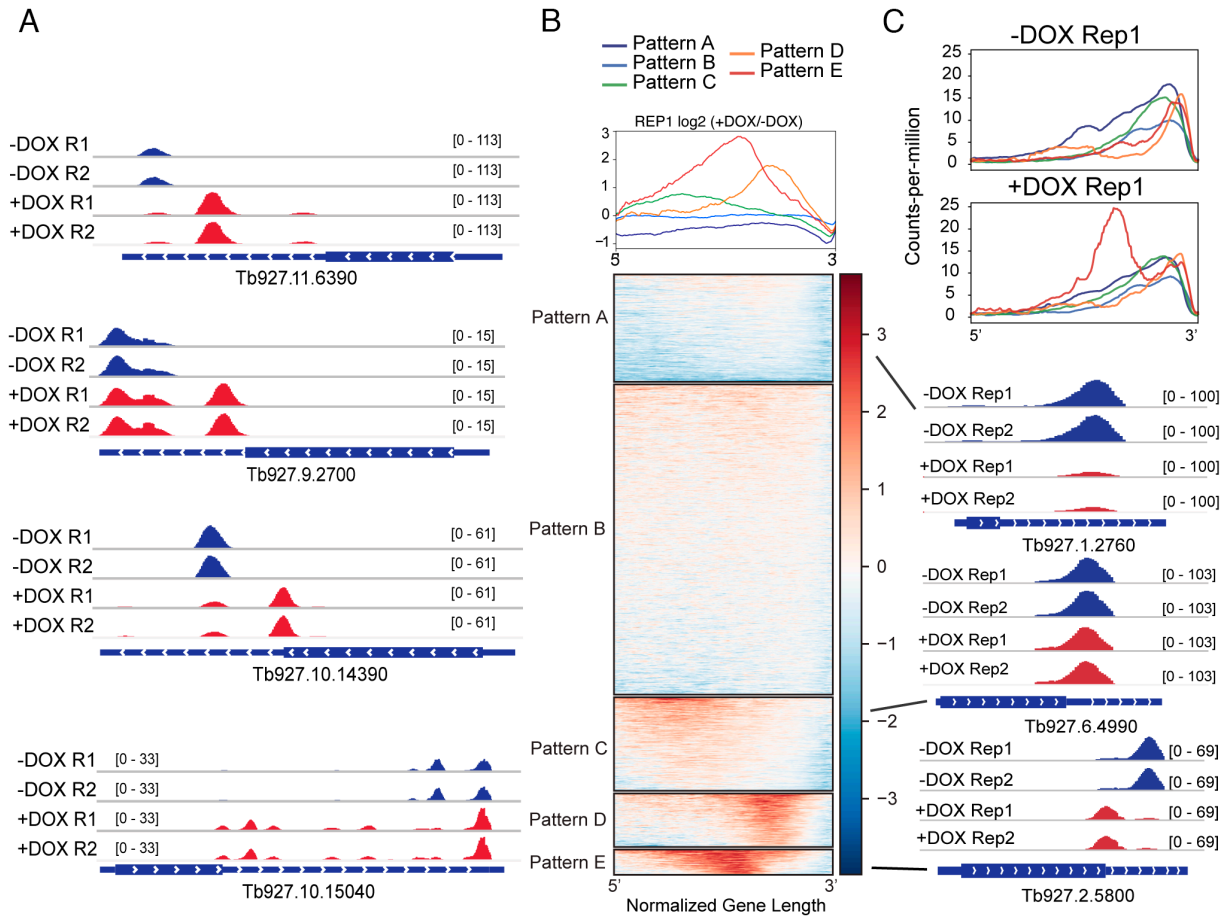
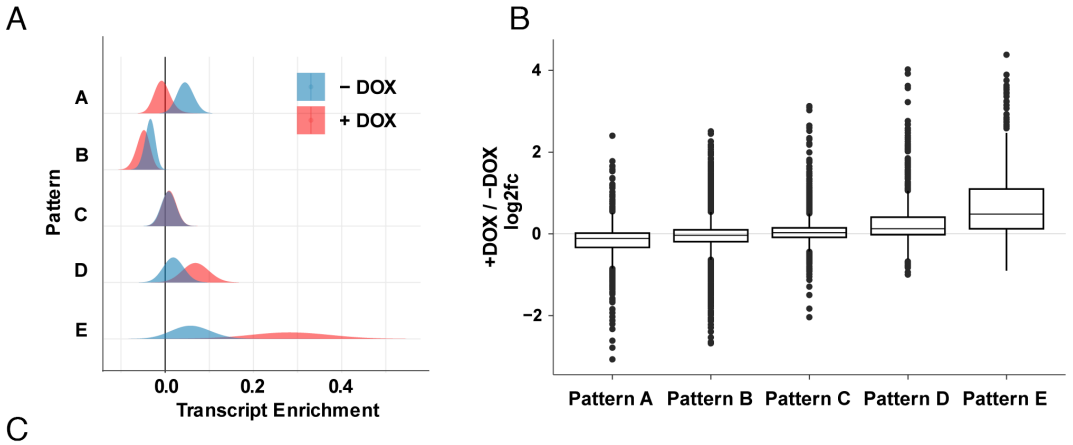


Fig. 3. Modulation of poly(A) addition sites upon DRBD18 depletion. (*A*) Examples of transcripts with altered poly(A) addition sites upon DRBD18 depletion in duplicate −DOX and +DOX samples, showing normalized read counts. Bigwig files were visualized using the Integrative Genomics Viewer (IGV). ORFs, thick blue line; intergenic regions, thin blue lines; direction of transcription, white arrowheads. (*B*) DeepTools analysis of poly(A) site changes upon DRBD18 depletion. Log2 read counts for each transcript with respect to position along the full length of the gene (defined as from the 5′ end of the 5′ UTR for a given gene to the 5′ end of the 5′ UTR of the next downstream gene in TriTrypDB) were extracted. DeepTools then classified the ratio of +DOX/−DOX signal for each transcript into one five distinct Patterns, A to E. (*Top*) Average signal of each Pattern shown as a ratio of +DOX to −DOX signals. (*Bottom*) Heatmap of transcriptome-wide analysis. Each row represents one gene classified as belonging to one of the five Patterns, A to E. Examples of Patterns A, C, and E transcripts are shown at the *Right*. (*C*) Unlogged average counts per million of transcripts comprising each Pattern, A to E, with respect to position along the full length of the gene.

translational regulatory element in the 3′UTR. Pattern A genes were enriched in the set of transcripts that decreased translational efficiency in DRBD18 knockdowns (TE down), implying a distinct mechanism of DRBD18-mediated translational control. With regard to nuclear mRNA export, Pattern A genes showed significant overlap with those transcripts exhibiting impaired nuclear export after DRBD18 knockdown, consistent with DRBD18-binding elements in long 3′UTRs to promote their nuclear export, as previously suggested (30). The Pattern A transcripts with impaired nuclear export and decreased translational efficiency were completely distinct. Conversely, Patterns B to E transcripts were not significantly enriched in export-impaired transcripts; thus, neither the increased abundance nor increased translation of transcripts with shortened 3′UTRs in DRBD18-depleted cells is correlated with modulation of their nuclear export by DRBD18. Together, these data implicate DRBD18 in a range of gene regulatory mechanisms, some of which are mediated through distinct 3′UTR elements.

Poly(A) Addition Sites in Many BF Transcripts Resemble Those in DRBD18-Depleted PF Cells. To this point, we have shown that DRBD18 depletion results increased expression of stumpy BF- and MF-specific transcripts (Fig. 2), as well as increased utilization of proximal poly(A) addition sites (Fig. 3). To begin to examine the relationship between 3′UTR length and life stage–specific gene regulation, we asked whether genes up-regulated in specific life cycle stages are enriched in Patterns D or E gene lists, that is, those exhibiting increased proximal poly(A) addition sites in DRBD18-depleted cells. Consistent with a role for altered poly(A) site selection in life stage–specific DRBD18-mediated gene regulation, we found significant overlap between Patterns D and E genes and those genes up-regulated in stumpy BF and MF (Fig. 5*A*). Indeed, 56/103 (54%) of stumpy BF-specific genes and 198/703 (28%) of MF up genes overlap with Pattern D or E datasets. Importantly, comparison of these 56 stumpy BF and 198 MF genes with reported protein levels in stumpy BF (45) and MF (35) stages showed that increased mRNA levels were reflected in increased protein levels (*SI Appendix*, Fig. S7). By contrast, pattern A and C transcripts were highly enriched in MF down transcripts.

The data in Fig. 5*A* suggest that DRBD18-regulated proximal poly(A) site utilization could be functionally connected to life stage–specific gene expression. If so, we expect that a subset of transcripts enriched in stumpy BF and MF cells have 3′UTR patterns resembling those in DRBD18-depleted PF cells. To address this hypothesis and identify specific stumpy BF transcripts with poly(A) addition sites that are proximal compared those in the same transcripts in PF, we compared Integrative Genomics Viewer read tracks from our scRNAseq dataset to



D

	Pattern A 2081 genes	Pattern B 6016 genes	Pattern C 1815 genes	Pattern D 1039 genes	Pattern E 488 genes
TE up 41 genes	4 overlap p = 0.95	4 overlap p = 0.99	9 overlap p = 0.18	8 overlap p = 0.03	13 overlap p = 6.42e-09
TE down 61 genes	17 overlap p = 0.04	28 overlap p = 0.85	9 overlap p = 0.63	0 overlap p = 1	2 overlap p = 0.73
Export Impaired 95 genes	41 overlap p = 8.61e-09	40 overlap p = 0.98	10 overlap p = 0.94	4 overlap p = 0.98	0 overlap p = 1

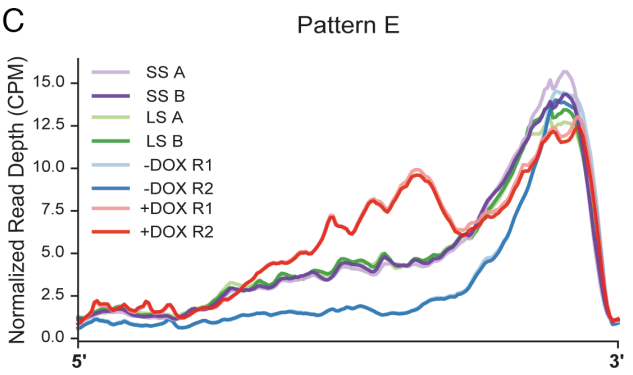
Fig. 4. Altered poly(A) site selection correlates with altered transcript abundance and translational efficiency. (*A*) Module scores, showing transcript enrichment, were calculated for each transcript set in Patterns A to E in −DOX and +DOX samples. (*B*) The average log2 fold changes of genes corresponding to Patterns A to E calculated from bulk RNAseq data (Dataset S1). The boxplot center line represents the median. (*C*) STREME was used to identify enriched motifs in 3′UTRs (from the stop codon to the 5′ end of the downstream gene) of Patterns D and E transcripts, with Pattern A transcripts serving as negative controls. Top four most highly enriched motifs (*Left*). Example of motifs that would be lost upon proximal polyadenylation site generation (*Right*). (*D*) The statistical significance of the overlap between each pair of gene lists was assessed using a hypergeometric test for overrepresentation. Light green: $P < 0.05$ and $>1 \times 10^{-5}$; dark green: $P < 1 \times 10^{-5}$.

those from published scRNAseq datasets that were generated at multiple points along the slender to stumpy transition by Briggs et al. (36). These authors generated scRNAseq data using a protocol similar to ours, allowing direct comparison of approximate poly(A) addition sites between datasets. In the previous study, the authors identified four different cell populations along the slender BF to stumpy BF transition, which they termed Long Slender A and B and Short Stumpy A and B. For this comparison, we normalized our datasets separately from those in ref. 36 due to the differing read depths in the two experiments. We combined the read counts from our replicates, and removed rRNA reads from all samples, as the BF datasets exhibited large rRNA peaks (36). Remarkably, we identified numerous examples of transcripts that, in BF, exhibited proximal poly(A) addition sites resembling those in our PF DRBD18 knockdown cells. Fig. 5*B* shows the comparison between our PF −dox and +dox samples for specific examples of such transcripts compared to those in all four stages of long slender BF to stumpy BF differentiation. This comparison highlights the utilization of life cycle–specific poly(A) site selection in *T. brucei*. Moreover, these data establish that for a subset of transcripts, DRBD18 maintains the PF-specific poly(A) site profile, which reverts to a BF-like pattern upon DRBD18 depletion.

To gain a transcriptome-wide view of the correspondence between poly(A) addition sites in PF cells replete for and depleted of DRBD18 and those in slender and stumpy BF, we turned to the transcript set with the most striking shift to proximal poly(A) addition site in DRBD18 knockdowns, those in our Pattern E dataset (Fig. 3 *B* and *C*). We analyzed read densities for the 488 Pattern E transcripts in the four published long slender and stumpy BF datasets and compared them to those in −dox and +dox DRBD18 RNAi cells. Comparison of −dox (Fig. 5*C*, light/dark blue) and +dox (Fig. 5*C*, pink/red) PF confirms the dramatic increase in proximal poly(A) addition sites upon DRBD18 knockdown in PF. The same transcripts in slender BF (Fig. 5*C*, light/dark green) and stumpy BF (Fig. 5*C*, light/dark purple) samples also exhibit a dramatic increase in 5′ shifted read densities compared to that in −dox PF, albeit not as dramatic as that in +dox PF. Overall, these data demonstrate that a substantial subset of BF transcripts is characterized by shortened 3′UTRs and increased abundance compared to those in the same transcripts in PF. The resemblance of this BF-specific 3′UTR pattern to that observed

A

	Pattern A 2081 genes	Pattern B 6016 genes	Pattern C 1815 genes	Pattern D 1039 genes	Pattern E 488 genes
Slender 57 genes	15 overlap p = 0.07	24 overlap p = 0.95	9 overlap p = 0.54	3 overlap p = 0.89	2 overlap p = 0.70
Stumpy 103 genes	7 overlap p = 1	26 overlap p = 1	11 overlap p = 0.94	**39 overlap p = 8.88e-16**	**17 overlap p = 1.20e-06**
Early PF 34 genes	5 overlap p = 0.75	11 overlap p = 0.99	**10 overlap p = 0.03**	4 overlap p = 0.36	2 overlap p = 0.42
Late PF 13 genes	1 overlap p = 0.92	7 overlap p = 0.55	1 overlap p = 0.89	2 overlap p = 0.33	2 overlap p = 0.10
MF Up 703 genes	127 overlap p = 0.47	248 overlap p = 1	124 overlap p = 0.07	**102 overlap p = 4.21e-07**	**96 overlap p = 5.32e-26**
MF Down 609 genes	**151 overlap p = 7.28e-06**	216 overlap p = 1	**135 overlap p = 7.28e-06**	64 overlap p = 0.09	**43 overlap p = 5.49e-04**



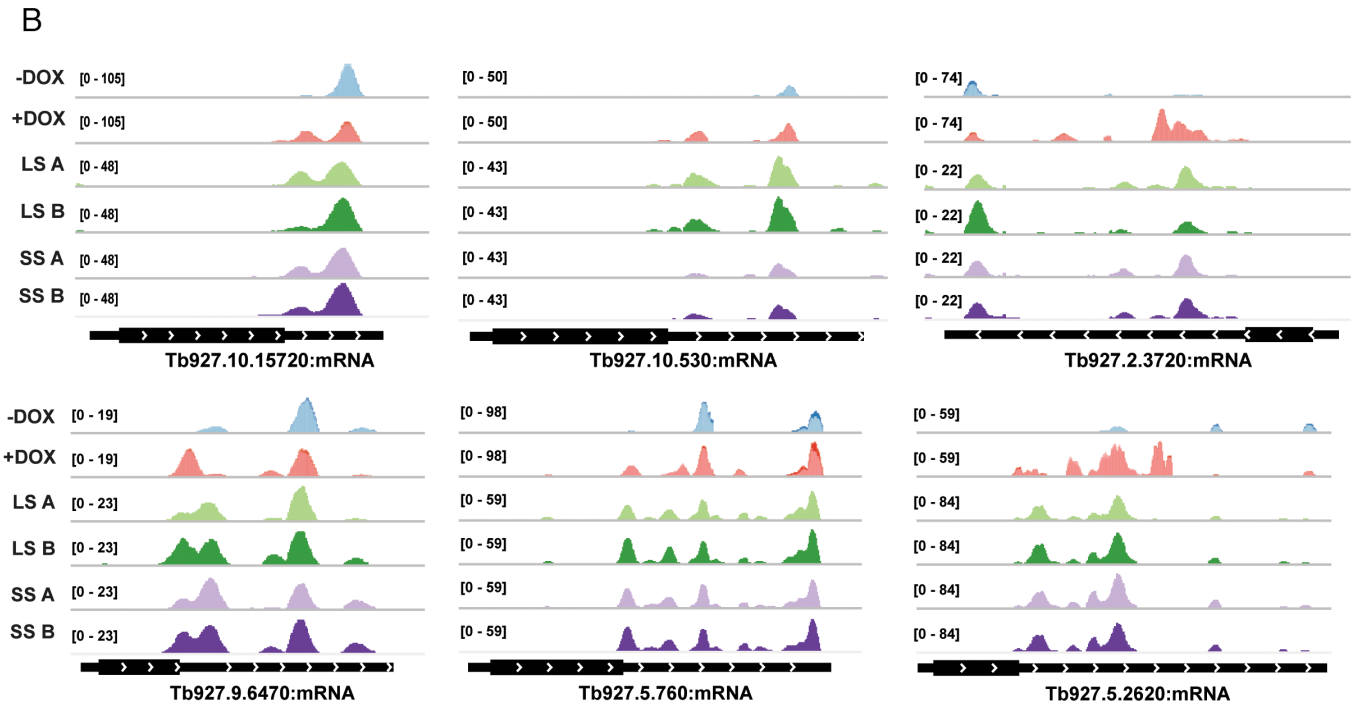


Fig. 5. Poly(A) site utilization is altered in different life cycle stages and is modulated by DRBD18. (*A*) Statistical significance of the overlaps between each pair of gene lists from each Pattern, A to E, compared to life cycle–specific genes was assessed using a hypergeometric test for overrepresentation. Light green: $P < 0.05$ and $>1 \times 10^{-5}$; dark green: $P < 1 \times 10^{-5}$. (*B*) Examples of transcripts with poly(A) site shifts in +dox PF DRBD18 RNAi cells that are similar to patterns in BF cells. Bigwig files were visualized using IGV. LS and SS clusters A and B from ref. 36 as compared to −/+DOX PF DRBD18 RNAi cells. PF replicates are depicted as lighter and darker versions of blue or red. ORFs, thick lines; intergenic regions, thin lines; direction of transcription, white arrowheads. (*C*) Average signal for binned and normalized read depths across all Pattern E genes in LS and SS clusters A and B from ref. 36 as compared to +/−DOX PF DRBD18 RNAi cells.

in the same transcripts in DRBD18-depleted PF indicates a role for DRBD18 in regulating life cycle–specific poly(A) site selection and transcript abundance.

DRBD18 Associates with Transcripts Whose Poly(A) Site It Controls. The observed effects of DRBD18 on poly(A) site selection could be directly caused by DRBD18 association with the affected transcripts. Alternatively, these effects could be mediated by other RBPs that change in abundance following DRBD18 depletion (*SI Appendix*, Fig. S1). To distinguish between these possibilities, we sought to identify those transcripts that are associated with DRBD18 in PF *T. brucei* using RNA Immunonprecipitation (RIP)-seq. To this end, strain 29-13 cells were UV cross-linked, lysed, and subjected to immunoprecipitation using anti-DRBD18 antibodies (31). RNA was isolated from both immunoprecipitate and input samples and sequenced. Transcript enrichment in DRBD18 immunoprecipitates compared to input was calculated by DEseq2 (46), resulting in identification of 579 transcripts that were enriched 1.5-fold (Padj < 0.05), which we term DRBD18-bound RNAs (Dataset S5). We note that DRBD18 interaction with these transcripts may be direct or mediated through other components of a ribonucleoprotein complex. We first asked whether DRBD18-bound RNAs are enriched for transcripts whose poly(A) sites shift 5′ upon DRBD18 depletion by sorting genes based on log2 fold-change RIP-seq intensities and examining the log2 ratio of the RNA densities for +dox/−dox. Genes with the highest RIP-seq binding signals correlated with shifts in 3′UTR usage to more proximal locations (Fig. 6*A*). This finding is supported by the relationship between enrichment of a given transcript in DRBD18 pull-downs and their poly(A) site addition pattern, Patterns A to E (Fig. 6*B*). Transcripts in Patterns D and E are enriched with DRBD18-bound transcripts, whereas Patterns A to C are enriched in transcripts that did not exhibit DRBD18 association. Hypergeometric tests showed that both Patterns D and E transcripts exhibit highly significant overlap with DRBD18-bound RNAs (Fig. 6*C*). These data support a model in which DRBD18-containing ribonucleoprotein complexes mediate poly(A) site selection by direct association with a cohort of transcripts.

We also established the relationships between DRBD18-bound transcripts and multiple additional parameters of gene regulation (*SI Appendix*, Fig. S8). Hypergeometric tests revealed significant

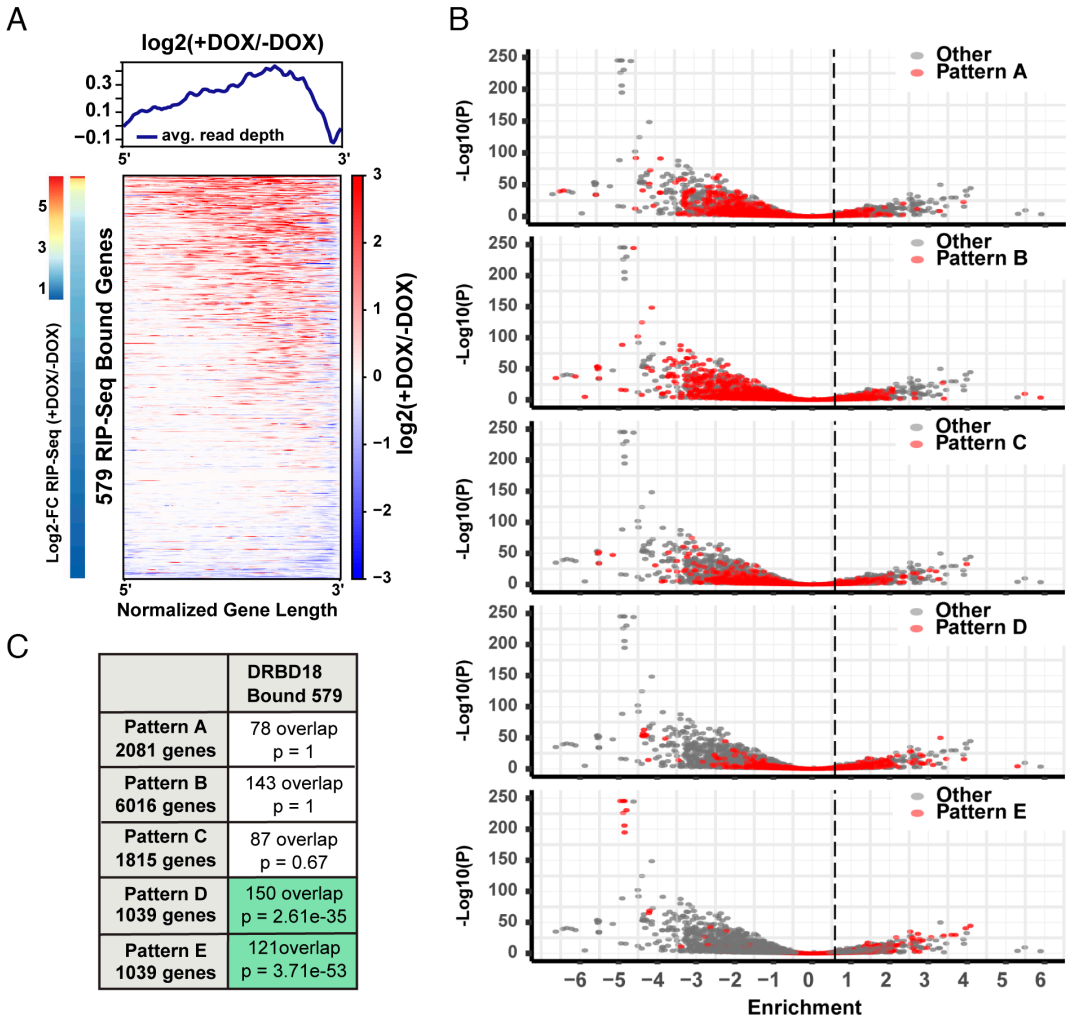


Fig. 6. DRBD18-bound transcripts are enriched for transcripts exhibiting poly(A) site shifts. (*A*) DeepTools analysis of poly(A) site changes upon DRBD18 depletion, sorted based on LogFC intensity of DRBD18-bound transcripts. (*Left*) Log2FC scaling for the 579 DRBD18-bound transcripts. Above: Average log2FC ratio of +DOX to −DOX signals for the 579 gene set. (*Below*) Heatmap of transcriptome-wide analysis. Increasing red intensity indicates higher expression in the +DOX samples; blue indicates higher −DOX expression. (*B*) Volcano plots showing the degree of transcript enrichment in DRBD18 RIP-seq. Individual plots illustrate 3′UTR Patterns A to E, with transcripts in a given pattern shown in red and the remainder of the transcriptome in gray. Transcripts to the *Right* of the vertical line (1.5 fold enriched) and above the horizontal dotted line ($P < 0.05$) are considered DRBD18-bound. (*C*) Statistical significance of the overlap between each pair of gene lists assessed using a hypergeometric test for overrepresentation. Light green: $P < 0.05$ and $>1 \times 10^{-5}$; dark green: $P < 1 \times 10^{-5}$.

overlap between those transcripts that are bound by DRBD18 and stumpy BF and MF markers, those with increased abundance or translational efficiency upon DRBD18 knockdown, and scRNAseq Cluster 0, Cluster 4, Cluster 7 markers (*SI Appendix*, Fig. S8*A*). Integration of these relationships revealed numerous transcripts that belong to multiple of these categories (*SI Appendix*, Fig. S8*B*). Overall, these data demonstrate that direct regulation of poly(A) site selection by DRBD18 often leads to downstream changes in gene expression.

Discussion

Here, we demonstrate that DRBD18 functions in the regulation of alternative polyadenylation and life cycle progression in *T. brucei*, and we establish a link between these processes. Enrichment of stumpy BF and MF transcripts in DRBD18 knockdown scRNAseq clusters indicates that DRBD18 promotes the PF state, in part, by repressing the abundance of transcripts specific to other life cycle stages. This finding adds to known impacts of DRBD18 on regulation of life cycle gene expression patterns. For example, in PF, DRBD18 inhibits the translation of a cohort of slender BF mRNAs, while promoting translation of PF mRNAs (32). Conversely, in slender BF, DRBD18 is essential for maintaining the abundance of RBP10, a critical factor in determining the BF state (30). Overall, a growing body of evidence marks DRBD18 as a key regulatory protein during both mammalian and insect stages of the *T. brucei* life cycle, with seemingly opposite effects in different life cycle stages. That is, DRBD18 is required for maintenance of both slender BF and PF life cycle stages.

DRBD18 depletion in PF dramatically altered 3′UTR lengths on over 1,500 transcripts (our combined Patterns D and E), in most cases leading to shortened 3′UTRs. A similar effect was previously reported for over 250 mRNAs in PF and BF *T. brucei* (30), although the previous study was somewhat limited by incomplete annotation of the genome, which we overcame by extending gene annotations to the 5′ end of the downstream transcript prior to analysis of sequencing results. The previous authors used a method different from ours to identify transcripts with altered 3′UTRs upon DRBD18 depletion and included BF transcripts in their analysis (30); nevertheless, we identified 50% of the transcripts that were identified in the previous study (Dataset S4). The increase in transcripts with shortened 3′UTRs when DRBD18 is depleted indicates that DRBD18 normally promotes the relative

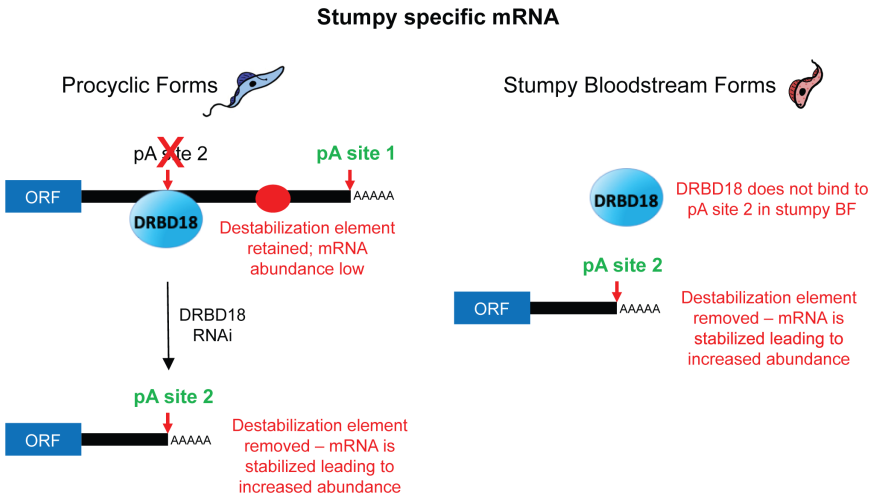


Fig. 7. Model of DRBD18 function in life cycle–specific poly(A) site selection. Shown is an example of a model stumpy BF-specific transcript whose polyadenylation site and abundance are regulated by DRBD18. See text for details.

utilization of distal poly(A) sites. Upon closer inspection of read tracks, it was clear that complex patterns of poly(A) site selection are at play. While proximal poly(A) addition sites were typically utilized in DRBD18 knockdown cells, the corresponding transcripts with distal sites could either remain or be diminished upon DRBD18 depletion. In many cases, numerous poly(A) addition sites became evident. Thus, mechanistically, DRBD18 appears to block the utilization of more proximal poly(A) addition sites in this cohort of transcripts (see model Fig. 7). Whether or not distal sites of polyadenylation continue to be utilized in DRBD18 knockdowns presumably depends on the strength of poorly understood sequence elements that surround poly(A) addition sites and to which the polyadenylation machinery binds.

A critical finding of this study is that the ability of DRBD18 to promote usage of distal polyadenylation sites, and thus likely inclusion of destabilization elements in mRNA 3′UTRs, plays an important role in its ability to alter the abundance of life cycle–specific transcripts. Stumpy BF and MF-specific transcripts increased in abundance upon DRBD18 knockdown, and this set of transcripts was also highly enriched for those with shortened 3′UTRs. To more specifically probe this correlation, we examined whether poly(A) site selection patterns in stumpy BF resembled those in PF DRBD18 knockdowns for a subset of transcripts. To this end, we compared published scRNAseq read tracks covering four stages in the slender to stumpy transition (36) to those from our uninduced and induced PF DRBD18 knockdown cells. Despite our experiments being done in a different *T. brucei* life cycle stage and strain, we observed a remarkable correspondence between the 3′ end peaks in all BF stages and PF DRBD18 knockdowns. Moreover, we identified a trend in which a cohort of transcripts exhibit both increased proximal poly(A) site utilization and abundance in slender and stumpy BF compared to PF, suggesting that numerous distal negative regulatory elements contribute to the maintenance of the PF state. Interestingly, while slender and stumpy BF generally exhibited similar 3′UTR patterns, transcripts with increased mRNA abundances in DRBD18 knockdown scRNAseq clusters overlapped only stumpy, but not slender BF-specific transcripts. This distinction between slender and stumpy BF suggests that additional controls are in place to modulate transcript abundances between these two stages, such as the presence of life cycle–specific RBPs (2).

Toward understanding the mechanism by which DRBD18 regulates poly(A) site selection in a life cycle–specific manner, we performed RIP-seq to identify mRNAs associated with DRBD18. DRBD18-bound transcripts were strongly enriched for both transcripts with shortened 3′UTRs in DRBD18 knockdowns, stumpy BF- and MF-specific transcripts, and markers for scRNAseq clusters 0 and 7. Together, these data strongly support the model of life cycle–specific gene regulation shown in Fig. 7. Fig. 7 depicts regulation of a general stumpy BF-specific transcript by DRBD18, shown here with just one proximal and one distal poly(A) selection site for simplicity. We envision that in PF (Fig. 7, *Left*) DRBD18 binds, either directly or a part of a complex, the proximal poly(A) site (pA site 2) of a subset of transcripts, thereby blocking access of the polyadenylation machinery. The polyadenylation machinery then utilizes the more distal site (pA site 1). Between pA sites 1 and 2 lies a sequence element that destabilizes the transcript, such as those identified in Fig. 4*C*, thereby keeping transcript abundance low in PF. When DRBD18 is depleted, pA site 2 is utilized in at least a portion of this transcript population. The resultant removal of the destabilization element leads to increased transcript abundance, as observed in our scRNAseq clusters 0 and 7. In stumpy BF (Fig. 7, *Right*), DRBD18 is unable to bind the more proximal poly(A) addition site (pA site 2), and this site is utilized, leading to increased transcript abundance. Because DRBD18 is similarly abundant in throughout the *T. brucei* life cycle, including in stumpy BF and MF stages (35, 45), additional factors must alter its functionality in distinct life cycle stages. One possible factor is differential arginine methylation, as this modification is known to regulate DRBD18's effects on transcript abundance (31). Identification of life cycle stage-specific DRBD18 posttranslational modifications and associated factors will inform how this protein modulates poly(A) site selection throughout the life cycle and will shed light on additional players in *T. brucei* developmental regulation.

Materials and Methods

Cell Culture. *T. brucei* PF 29-13 cells and the DRBD18 RNA interference (RNAi) derivative were grown as published (31). For scRNAseq, RNAi was induced using 4 μg/mL dox for 18 h. Cell densities at the time of harvest were as follows: Rep 1 −dox, 3.0×10^6, Rep 1 +dox 2.8×10^6, Rep 2 −dox, 3.4×10^6, Rep 2 +dox 3.0×10^6 cells/mL.

Combined Analysis of Published Bulk RNAseq Datasets. Bulk RNAseq data were retrieved from refs. 31 and 33 (total RNA) and aligned to the TriTrypDB-46 TREU927 genome using the HiSat2 algorithm. Per gene quantifications were calculated using the subread function FeatureCounts, using the strand-specific parameter enabled. Both paired-end and single-end data were then supplied

to the R package DESeq2, with a design controlling for experiment of origin (~ is_paired + treatment).

scRNAseq. Approximately 10,000 cells were loaded on a Chromium Single Cell Instrument (10× Genomics) to generate single-cell gel bead-in-emulsions (GEMs). ScRNAseq libraries were prepared using Chromium Single cell 3′ Reagent Kits v2 Library (10× Genomics). Libraries were run using paired-end sequencing on Illumina NovaSeq S1 Flowcell using the following cycles: Read 1 -28 cycles, i7-8 cycles, i5-0 cycles, Read 2-94 cycles. Cell Ranger version 6.1.2 software was used to perform sample demultiplexing and single-cell gene quantification. Reads were aligned to a modified TriTrypDB-46 *T. brucei* TREU927 version 46 reference genome where the 3′ ends of all gene annotations were extended until 1 bp adjected (nonoverlapping) of the downstream neighboring gene record.

Single-Cell Analysis Using the R Package Seurat. The scRNAseq dataset was demultiplexed using the 10× CellRanger pipeline, followed by downstream analysis in the R package Seurat v4. All four samples were introduced into the canonical correlation analysis integration workflow, followed by Louvain clustering, and UMAP projection. Marker genes were calculated using the Seurat function FindAllMarkers and FindConservedMarkers. To define clusters enriched for distinct cell types, we performed gene module scoring using the AddModuleScore function in Seurat (42). Slender, Stumpy, Early PF, and Late PF gene lists were obtained from ref. 34, Additional file 5. MF gene lists were from ref. 35.

Gene Ontology Analysis. Lists of enriched marker genes (avg_log2FC > 0.25) were submitted for Gene Ontology enrichment analysis on https://tritrypdb.org/ with a *P*-Value cutoff setting of 0.05. The resulting lists of enriched GO terms were then refined using GO-Figure! semantic redundancy reduction (version 1.0.1) under default settings.

Significance of Overlaps between Gene Lists. The statistical significance of overlaps between gene lists was calculated with a hypergeometric test using the phyper function in R with lower.tail set to FALSE.

DeepTools Pattern Analysis. BAM alignment files were used as input into the DeepTools analysis package to evaluate relative shifts in 3′UTR positioning. First, the alignment files were converted into bigWig format using the function bamCoverage, using the normalizing option of counts per million to scale for sequencing depth. Next, we calculated the log2ratio between the +dox and −dox sample using the bamCompare function (--operation log2). The normalized bigWig tracks were used as input into the computeMatrix function (scale-regions -a 0, -b 0, -skipZeros), followed by hierarchical clustering analysis (kmeans = 5) and visualization using the plotHeatmap tool, which established our patterns of 3′ enrichment.

RIP-seq. UV cross-linking of 5×10^9 *T. brucei* 29-13 cells was carried out with triplicate biological replicates as described (47). Cross-linked DRBD18-RNA complexes were immunopurified from the lysate using anti-DRBD18 antibodies (31) attached to protein A fast flow beads (GE Healthcare). The unbound fraction was also collected; 10% was used for RNA extraction and 2% for western blotting. Beads were washed three times with IP buffer (lysis buffer plus 150 mM NaCl) and treated with DNase 1 (Sigma) followed by proteinase K (Roche) after confirmation by western blot. RNA was extracted from beads with phenol/chloroform and ethanol precipitated. RNA was isolated from the input fractions using Trizol and ethanol precipitation.

Input RNA and DRBD18-pulldown RNA were paired-end sequenced as follows. RNA molecules were converted into libraries using the Takara SMART-Seq RNA kit for low input following (48) with modifications of no fragmentation and 16 cycles of PCR for final amplification. All final libraries were Ampure bead purified, pooled, and diluted to 10 nM for qPCR using QuantaBio qPCR Universal Library assay. Libraries were sequenced on an Illumina NextSeq500 platform at 1.5 pM (midoutput, PE75). Adapter sequences were trimmed using cutadapt. Fastq files were then mapped to the *T. brucei* Lister927 genome using Bowtie2. Gene count matrices were produced using featureCounts with the gtf file with extended 3′ ends. The matrices were then imported into RStudio for DESeq2 analysis. Differential expression between input and DRBD18-pulldown samples was calculated using the standard DESeq2 workflow followed by LFC shrinkage.

Data, Materials, and Software Availability. Raw and processed data are available on GEO/SRA at GSE255758 (49). The code used for this project is available at https://github.com/BriannaTylec/Publications/tree/main/2024_Bard_Tylec (50).

ACKNOWLEDGMENTS. This work was supported by NIH grants R01 AI141557 and R21 AI166820 to L.K.R. We thank the UB Genomics and Bioinformatics Core.

Author affiliations: [a]Genomics and Bioinformatics Core, University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY 14203; [b]Department of Biochemistry, University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY 14203; and [c]Department of Microbiology and Immunology, University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY 14203

1. E. Pays, M. Radwanska, S. Magez, The pathogenesis of African Trypanosomiasis. *Annu. Rev. Pathol.* **18**, 19–45 (2023).
2. C. Clayton, Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. *Open Biol.* **9**, 190072 (2019).
3. E. Mugo, C. Clayton, Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. *PLoS Pathog.* **13**, e1006560 (2017).
4. N. G. Kolev, K. Ramey-Butler, G. A. Cross, E. Ullu, C. Tschudi, Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. *Science* **338**, 1352–1353 (2012).
5. J. Y. Toh, A. Nkouawa, G. Dong, N. G. Kolev, C. Tschudi, Two cold shock domain containing proteins trigger the development of infectious Trypanosoma brucei. *PLoS Pathog.* **19**, e1011438 (2023).
6. E. Rico, A. Ivens, L. Glover, D. Horn, K. R. Matthews, Genome-wide RNAi selection identifies a regulator of transmission stage-enriched gene families and cell-type differentiation in Trypanosoma brucei. *PLoS Pathog.* **13**, e1006279 (2017).
7. L. McDonald *et al.*, Non-linear hierarchy of the quorum sensing signalling pathway in bloodstream form African trypanosomes. *PLoS Pathog.* **14**, e1007145 (2018).
8. N. G. Kolev *et al.*, The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS Pathog.* **6**, e1001090 (2010).
9. T. N. Siegel, D. R. Hekstra, X. Wang, S. Dewell, G. A. Cross, Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* **38**, 4946–4957 (2010).
10. L. Quijada *et al.*, Expression of the human RNA-binding protein HuR in Trypanosoma brucei increases the abundance of mRNAs containing AU-rich regulatory elements. *Nucleic Acids Res.* **30**, 4414–4424 (2002).
11. A. Furger, N. Schurch, U. Kurath, I. Roditi, Elements in the 3′ untranslated region of procyclin mRNA regulate expression in insect forms of Trypanosoma brucei by modulating RNA stability and translation. *Mol. Cell. Biol.* **17**, 4372–4380 (1997).
12. H. R. Hotz, C. Hartmann, K. Huober, M. Hug, C. Clayton, Mechanisms of developmental regulation in Trypanosoma brucei: A polypyrimidine tract in the 3′-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res.* **25**, 3017–3026 (1997).
13. S. K. Archer, V. D. Luu, R. A. de Queiroz, S. Brems, C. Clayton, Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog.* **5**, e1000565 (2009).
14. S. M. Fernandez-Moya, M. Carrington, A. M. Estevez, A short RNA stem-loop is necessary and sufficient for repression of gene expression during early logarithmic phase in trypanosomes. *Nucleic Acids Res.* **42**, 7201–7209 (2014).
15. M. Rico-Jimenez, G. Ceballos-Perez, C. Gomez-Linan, A. M. Estevez, An RNA-binding protein complex regulates the purine-dependent expression of a nucleobase transporter in trypanosomes. *Nucleic Acids Res.* **49**, 3814–3825 (2021).
16. S. Kramer, M. Carrington, An AU-rich instability element in the 3′UTR mediates an increase in mRNA stability in response to expression of a dhh1 ATPase mutant. *Translation (Austin)* **2**, e28587 (2014).
17. T. Bishola Tshitenge, L. Reichert, B. Liu, C. Clayton, Several different sequences are implicated in bloodstream-form-specific gene expression in Trypanosoma brucei. *PLoS Negl. Trop. Dis.* **16**, e0010030 (2022).
18. D. Droll *et al.*, Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Pathog.* **9**, e1003286 (2013).
19. Q. Zhang, B. Tian, The emerging theme of 3′UTR mRNA isoform regulation in reprogramming of cell metabolism. *Biochem. Soc. Trans.* **51**, 1111–1119 (2023).
20. Z. Ji, J. Y. Lee, Z. Pan, B. Jiang, B. Tian, Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7028–7033 (2009).
21. D. Heller-Trulli, H. Liu, S. Mukherjee, C. L. Moore, UBE3D regulates mRNA 3′-end processing and maintains adipogenic potential in 3T3-L1 cells. *Mol. Cell. Biol.* **42**, e0017422 (2022).
22. S. Mitschka, C. Mayr, Context-specific regulation and function of mRNA alternative polyadenylation. *Nat. Rev. Mol. Cell Biol.* **23**, 779–796 (2022).
23. E. F. Hendriks, A. Abdul-Razak, K. R. Matthews, tbCPSF30 depletion by RNA interference disrupts polycistronic RNA processing in Trypanosoma brucei. *J. Biol. Chem.* **278**, 26870–26878 (2003).
24. H. Koch, M. Raabe, H. Urlaub, A. Bindereif, C. Preusser, The polyadenylation complex of Trypanosoma brucei: Characterization of the functional poly(A) polymerase. *RNA Biol.* **13**, 221–231 (2016).
25. E. Ullu, K. R. Matthews, C. Tschudi, Temporal order of RNA-processing reactions in trypanosomes: Rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol. Cell. Biol.* **13**, 720–725 (1993).
26. K. R. Matthews, C. Tschudi, E. Ullu, A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* **8**, 491–501 (1994).
27. S. K. Gupta *et al.*, Two splicing factors carrying serine–arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA stability, and rRNA processing in Trypanosoma brucei. *RNA Biol.* **11**, 715–731 (2014).

28. S. K. Gupta *et al.*, 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).
29. A. Waithaka, O. Maiakovska, D. Grimm, L. M. do Nascimento, C. Clayton, Sequences and proteins that influence mRNA processing in Trypanosoma brucei: Evolutionary conservation of SR-domain and PTB protein functions. *PLoS Negl. Trop. Dis.* **16**, e0010876 (2022).
30. T. Bishola Tshitenge, C. Clayton, The Trypanosoma brucei RNA-binding protein DRBD18 ensures correct mRNA trans splicing and polyadenylation patterns. *Rna* **28**, 1239–1262 (2022).
31. K. Lott *et al.*, Arginine methylation of DRBD18 differentially impacts its opposing effects on the trypanosome transcriptome. *Nucleic Acids Res.* **43**, 5501–5523 (2015).
32. M. Ciganda *et al.*, Translational control by Trypanosoma brucei DRBD18 contributes to the maintenance of the procyclic state. *Rna* **29**, 1881–1895 (2023).
33. A. Mishra *et al.*, Selective nuclear export of mRNAs is promoted by DRBD18 in Trypanosoma brucei. *Mol. Microbiol.* **116**, 827–840 (2021).
34. A. Naguleswaran, N. Doiron, I. Roditi, RNA-Seq analysis validates the use of culture-derived Trypanosoma brucei and provides new markers for mammalian and insect life-cycle stages. *BMC Genomics* **19**, 227 (2018).
35. R. Christiano *et al.*, The proteome and transcriptome of the infectious metacyclic form of Trypanosoma brucei define quiescent cells primed for mammalian invasion. *Mol. Microbiol.* **106**, 74–92 (2017).
36. E. M. Briggs, F. Rojas, R. McCulloch, K. R. Matthews, T. D. Otto, Single-cell transcriptomic analysis of bloodstream Trypanosoma brucei reconstructs cell cycle progression and developmental quorum sensing. *Nat. Commun.* **12**, 5268 (2021).
37. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018), 10.1038/nbt.4314.
38. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
39. M. Reijnders, R. M. Waterhouse, Summary visualizations of gene ontology terms with GO-Figure! *Front. Bioinform.* **1**, 638255 (2021).
40. C. Klein, M. Terrao, C. Clayton, The role of the zinc finger protein ZC3H32 in bloodstream-form Trypanosoma brucei. *PLoS One* **12**, e0177901 (2017).
41. E. Vassella *et al.*, A major surface glycoprotein of trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes Dev.* **14**, 615–626 (2000).
42. I. Tirosh *et al.*, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
43. F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, T. Manke, deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
44. T. L. Bailey, STREME: Accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).
45. M. Dejung *et al.*, Quantitative proteomics uncovers novel factors involved in developmental differentiation of Trypanosoma brucei. *PLoS Pathog.* **12**, e1005439 (2016).
46. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. E. Mugo, E. D. Erben, Identifying trypanosome protein–RNA interactions using RIP-Seq. *Methods Mol. Biol.* **2116**, 285–294 (2020).
48. Y. Zeng *et al.*, Refined RIP-seq protocol for epitranscriptome analysis with low input materials. *PLoS Biol.* **16**, e2006092 (2018).
49. J. E. Bard *et al.*, Data from "Life stage specific poly(A) site selection regulated by Trypanosoma brucei DRBD18." Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE255758. Deposited 14 February 2024.
50. J. E. Bard *et al.*, R code for analysis of DRBD18 knockdown single cell RNAseq data. Github. https://github.com/BriannaTylec/Publications/. Deposited 15 January 2024.