The *Trypanosoma brucei* RNA-binding protein DRBD18 ensures correct mRNA *trans* splicing and polyadenylation patterns

TANIA BISHOLA TSHITENGE and CHRISTINE CLAYTON

Heidelberg University Center for Molecular Biology (ZMBH), D69120 Heidelberg, Germany

ABSTRACT

The parasite *Trypanosoma brucei* grows as bloodstream forms in mammals, and as procyclic forms in tsetse flies. Transcription is polycistronic, all mRNAs are *trans* spliced, and polyadenylation sites are defined by downstream splicing signals. Expression regulation therefore depends heavily on post-transcriptional mechanisms. The RNA-binding protein DRBD18 was previously implicated in the export of some mRNAs from the nucleus in procyclic forms. It copurifies the outer ring of the nuclear pore, mRNA export factors and exon-junction-complex proteins. We show that for more than 200 mRNAs, DRBD18 depletion caused preferential accumulation of versions with shortened 3′-untranslated regions, arising from use of polyadenylation sites that were either undetectable or rarely seen in nondepleted cells. The shortened mRNAs were often, but not always, more abundant in depleted cells than the corresponding longer versions in normal cells. Their appearance was linked to the appearance of *trans*-spliced, polyadenylated RNAs containing only downstream 3′-untranslated region-derived sequences. Experiments with one mRNA suggested that nuclear retention alone, through depletion of MEX67, did not affect mRNA length, suggesting a specific effect of DRBD18 on processing. DRBD18-bound mRNAs were enriched in polypyrimidine tract motifs, and DRBD18 was found in both the nucleus and the cytoplasm. We therefore suggest that in the nucleus, DRBD18 might bind to polypyrimidine tracts in 3′-UTRs of mRNA precursors. Such binding might both prevent recognition of mRNA-internal polypyrimidine tracts by splicing factors, and promote export of the processed bound mRNAs to the cytosol.

Keywords: trypanosome; *trans* splicing; alternative splicing; polyadenylation; mRNA export

INTRODUCTION

Kinetoplastids are unicellular flagellated parasites that infect mammals and plants. The African trypanosome *Trypanosoma brucei* is a kinetoplastid that causes sleeping sickness in humans in Africa and infects livestock throughout the tropics, with a substantial economic impact (Shaw et al. 2014). *T. brucei* are transmitted by Tsetse flies, or during passive blood transfer by biting flies. In mammalian blood and tissue fluids, the parasites multiply extracellularly as long slender bloodstream forms, escaping the host immune response through antigenic variation of the variant surface glycoproteins (VSGs) (Gray 1965).

In kinetoplastids, nearly all protein-coding genes are arranged in polycistronic transcription units. Mature mRNAs are generated from the primary transcript by 5′-*trans*-splicing of a 39-nt capped leader sequence, and 3′-polyadenylation (Clayton and Michaeli 2011; Michaeli 2011). The parasite regulates mRNAs mainly by post-transcriptional mechanisms, supplemented, in the case of some constitutively abundant mRNAs, by the presence of multiple gene copies. Regulation of mRNA processing, degradation, and translation are therefore central to parasite homeostasis, and for changes in gene expression during differentiation (Fadda et al. 2014; Jensen et al. 2014; Vasquez et al. 2014; Antwi et al. 2016). The sequences required for regulation of mRNA stability and translation often lie in the 3′-untranslated regions (3′-UTRs) of the mRNAs, and most regulation so far has been found to depend on RNA-binding proteins (Clayton 2019). For example, RBP10 (Tb927.8.2780), which is exclusively expressed in mammal-infective forms (Wurst et al. 2012; Savage et al. 2016; Christiano et al. 2017; Shi et al. 2018; Vigneron et al. 2020) specifically associates with procyclic-specific mRNAs, targeting them for destruction (Mugo and Clayton 2017). The *RBP10* 3′-UTR is 7.3 kb

Corresponding author: cclayton@zmbh.uni-heidelberg.de

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.079258.122. Freely available online through the RNA Open Access option.

long and contains at least six different regions that are independently capable of causing bloodstream-form-specific expression of a reporter (Bishola Tshitenge et al. 2022).

In comparison with mRNA decay and translation, our knowledge of regulatory events during mRNA processing, and especially mRNA export, is rather limited. The mechanism of mRNA *trans*-splicing is basically similar to that for *cis* and *trans* splicing in other eukaryotes (Michaeli 2011; Preusser et al. 2012), as is the polyadenylation machinery (Koch et al. 2016). Unusually, however, the two processes are inextricably linked: the polyadenylation site of each mRNA is determined solely by the location of the next downstream splice site (Huang and van der Ploeg 1991; Ullu et al. 1993; Matthews et al. 1994; Vassella et al. 1994); and chemical or RNAi-mediated inhibition of either splicing or polyadenylation stops both processes (Ullu and Tschudi 1991; McNally and Agabian 1992; Hendriks et al. 2003; Begolo et al. 2018; Wall et al. 2018). Like other eukaryotes, trypanosomes have an exon junction complex, which is presumably deposited at the end of the spliced leader after or during processing: it includes Y14 (Tb927.7.1170), Mago-Nashi (Tb927.6.4950), and an NTF2-like protein (Tb927.10.2240) (Bercovich et al. 2009) and perhaps an eIF4AIII homolog, Tb927.11.8770 (Dhalia et al. 2006; Inoue et al. 2014). Bioinformatic analysis and reporter experiments indicate that—as in other eukaryotes—a polypyrimidine tract (PPT) serves as the main *trans*-splicing signal (Matthews et al. 1994; Schürch et al. 1994; Vassella et al. 1994; Nilsson et al. 2010; Siegel et al. 2010; Waithaka et al. 2022), but in contrast with at least some eukaryotes (Patzelt et al. 1989), there is no consensus branch-point sequence. The results of reporter experiments indicate that the nature of the sequence downstream from the splice site, as well as the length and composition of the PPT, influence splice site choice (Hartmann et al. 1998; Lopez-Estrano et al. 1998; Siegel et al. 2005; Waithaka et al. 2022). There are also indications that some mRNAs are more efficiently spliced than others (Kapotas and Bellofatto 1993; Fadda et al. 2014; Antwi et al. 2016); and when processing is inhibited, the precursors are substrates for the exosome (Kramer et al. 2016). Several RNA-binding proteins have been shown to influence splicing (Gupta et al. 2013a,b, 2014) but their mechanisms of action are not clear; often, they have double functions, also binding to the 3′-UTRs of mature mRNAs and affecting mRNA stability.

The mechanism of mRNA export from the nucleus in kinetoplastids has received relatively little attention (for review, see Kramer 2021). In opisthokonts, mRNA export requires a complex that is conserved across eukaryotes: it includes Mex67 (NXF1 or TAP in mammals) and the helicase Mtr2 (p15 or NXT1 in mammals) as well as various other less conserved proteins (for review, see Ashkenazy-Titelman et al. 2020). Mex67 interacts with both the mRNA and the nuclear pore. Gle2 (RAE1), a nuclear pore component, is also implicated in export. Recognition of mature mRNAs by export factors involves various proteins that can bind to the RNA, including cap-binding proteins, splicing factors and the exon junction complex (for review, see Ashkenazy-Titelman et al. 2020). Trypanosomes use a complex that contains the major export factor MEX67 (Tb927.11.2370) (Schwede et al. 2009; Dostalova et al. 2013), the helicase MTR2 (Tb927.7.5760), and a transportin-like protein, IMP1 (Tb927.9.13520) (Dostalova et al. 2013); trypanosome GLE2 (Tb927.9.3760) is also implicated. The inner ring of the trypanosome nuclear pore is very similar to those of opisthokonts, but outer components are more diverged and there are no cytoplasmic filaments (Obado et al. 2016). In trypanosomes, export of mRNAs can be initiated before they have been completely transcribed (Goos et al. 2018), so recognition must depend on the 5′-end: we do not know whether the export machinery recognizes the cap, cap-binding proteins, or the exon junction complex. The TREX and THO complexes, which are implicated in the integration of transcription, splicing and export in opisthokonts (Ashkenazy-Titelman et al. 2020), are absent in trypanosomes (Kramer 2021).

The double RNA-binding protein 18 (DRBD18) was initially identified as a substrate of arginine methylation (Lott et al. 2015). It is expressed in both bloodstream and procyclic forms. Various lines of evidence indicate that DRBD18 is involved in the export of a subset of mRNAs from the nucleus. DRBD18 interacts directly with MTR2, and coimmunoprecipitates MEX67 (Mishra et al. 2021). Moreover, in procyclic forms, depletion of DRBD18 by RNAi caused partial retention of MTR2 and MEX67, and some poly(A)$^+$ mRNAs, in the nucleus (Mishra et al. 2021). Levels of various mRNAs were affected, but the effects were different depending on whether cytoplasmic or whole-cell mRNAs were analyzed (Mishra et al. 2021).

One of the mRNAs that increased significantly in abundance after DRBD18 depletion in procyclic forms was that encoding RBP10 (Lott et al. 2015; Mishra et al. 2021). During studies of RBP10 regulation, we unexpectedly discovered that DRBD18 depletion led to accumulation of truncated versions of the *RBP10* mRNA. Analysis of transcriptomes from both bloodstream and procyclic forms revealed that DRBD18 depletion affects processing of over 200 mRNAs. These results suggest that DRBD18 may affect mRNA processing as well as mRNA export.

RESULTS

DRBD18 depletion affects *RBP10* mRNA processing

We initially studied DRBD18 because its depletion in procyclic forms had been shown to cause a modest increase in the abundance of *RBP10* mRNA (Lott et al. 2015; Mishra et al. 2021). We found that depletion of DRBD18 in bloodstream forms caused growth inhibition (Fig. 1A) and a

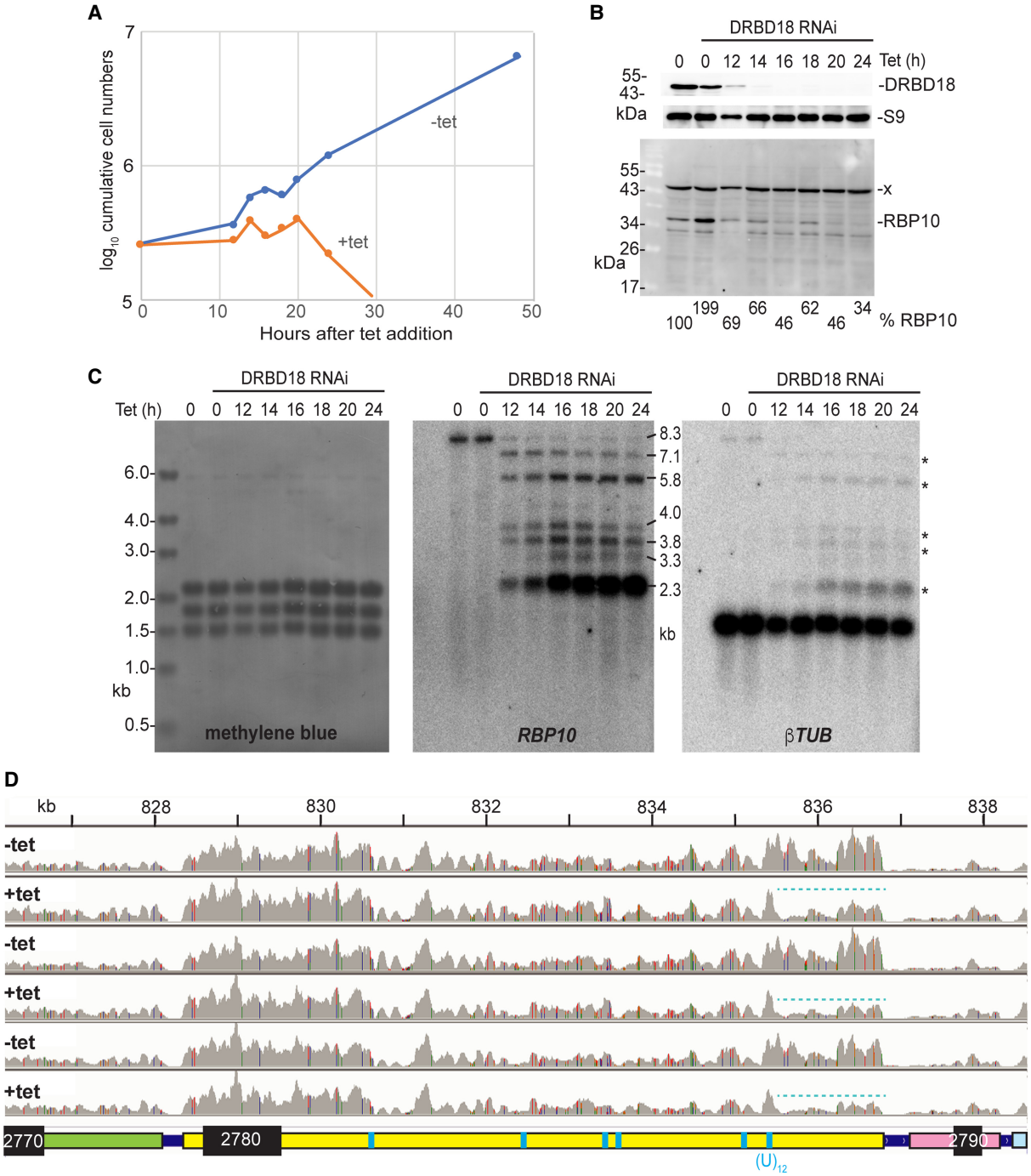

FIGURE 1. Effect of *DRBD18* RNAi on *RBP10* mRNA and expression levels. (*A*) Example of cumulative growth curves of bloodstream forms with and without induction of *DRBD18* RNAi. (*B*) DRBD18 and RBP10 protein levels after RNAi induction, corresponding to *A*. Ribosomal protein S9 is the loading control. Anti-RBP10 antibodies (Wurst et al. 2012) were used for detection of RBP10 and a nonspecific band (x). (*C*) RNA from the samples used in *A* was used to examine effects on *RBP10* mRNA. The methylene blue-stained membrane is shown on the *left*, with *RBP10* hybridization in the *center*. The membrane was then stripped and hybridized with a β-tubulin probe as a further control; **RBP10* signal that remained after stripping. (*D*) Visualization of RNA-seq results after 12 h tetracycline-mediated induction of *DRBD18* RNAi, using the Integrated Genomics Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2013). Three replicates are shown. Note that these results are shown on a linear scale whereas those in Figure 1A are on a log scale. Proposed full length mRNAs are shown *below* the tracks; positions of $(U)_{12}$ or $(U)_6C(U)_6$ sequences are indicated in cyan but there are numerous other polypyrimidine tracts present. The cyan dotted line highlights loss of reads at the end of the *RBP10* 3′-UTR after 12 h *DRBD18* RNAi. The read densities over the middle portion of the 3′-UTR are lower than the rest, probably because some of the reads will be aligned to the additional sequence copy in the TREU927 reference genome. Colored lines indicate read mismatches with the TREU927 reference genome.

modest (threefold) decrease in RBP10 protein (Fig. 1B). Remarkably, however, when we looked at *RBP10* mRNA by northern blotting, we discovered that DRBD18 depletion caused accumulation of progressively smaller *RBP10* mRNAs (Fig. 1C). Since the blot was hybridized to the open reading frame (coding sequence, CDS) probe, the mRNAs must have different 3′-UTRs. Eukaryotic mRNAs that lack poly(A) tails are generally extremely unstable, so we hypothesized that the various *RBP10* mRNAs were products of alternative polyadenylation. To find out whether the effect was specific for *RBP10*, we examined mRNA from the tandemly repeated tubulin genes, because processing inhibition causes accumulation of tubulin (*TUB*) RNA dimers and multimers. Processing of *TUB* mRNA was not affected, showing that the effect on *RBP10* mRNA was not due to a general RNA processing defect (Fig. 1C). A similar effect on *RBP10* mRNA was seen after DRBD18 depletion in a different strain, EATRO1125 (Supplemental Fig. S1A,B). Analysis of 3′-ends of the different *RBP10* mRNAs by RT-PCR (3′-RACE) using oligo $(dT)_{18}$ was not successful, perhaps because the sequence is generally of low complexity and the forward primers chosen were similar to sequences in other mRNAs.

To examine this in more detail and to look for effects on other mRNAs, we examined the transcriptomes of bloodstream forms 12 h after DRBD18 RNAi induction. This time was chosen as the first point at which DRBD18 was completely depleted without much effect on growth (Fig. 1A,B). We sequenced total rRNA-depleted RNA. The numbers of reads from the *RBP10* coding region were not much affected after 12 h (Supplemental Table S1), as expected from the northern blot results. Visualization in the Integrated Genome Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2013) also showed that with or without tetracycline induction of RNAi, the read density over the *RBP10* coding region was 2–3 times higher than the density over the upstream gene, Tb927.8.2770 (Fig. 1D). However, over the last 2 kb of the 3′-UTR, there were far fewer reads after DRBD18 depletion (cyan dotted line), indicating loss of the longest *RBP10* mRNAs and consistent with the northern blot result.

The results so far suggested that multiple alternative polyadenylation sites were being used after *DRBD18* depletion. The *RBP10* 3′-UTR contains several PPTs that might serve as splicing signals (Fig. 1D).

Depletion of MEX67 does not cause accumulation of shortened *RBP10* mRNAs

A previous paper implicated DRBD18 in export of mRNAs from the nucleus. We had previously shown that induction of RNAi targeting *MEX67* caused accumulation of poly(A)$^+$ RNA in the nucleus (Schwede et al. 2009). After 24 h RNAi induction in this cell line, growth of the parasites was clearly inhibited (Fig. 2A). After either 18 or 24 h, *MEX67* mRNA was reduced (Fig. 2B), but the migration of *RBP10* mRNA was unaffected (Fig. 2B). It might be argued that after MEX67 depletion, the retained *RBP10* mRNAs cannot be further processed because the splicing machinery is overwhelmed by extra substrates. Multimers of the tubulin mRNA accumulate rapidly after splicing inhibition (Ullu and Tschudi 1991; McNally and Agabian 1992; Hendriks et al. 2003; Begolo et al. 2018; Wall et al. 2018), but after MEX67 depletion a faint dimer band was barely visible (Fig. 2B), arguing against this hypothesis. These results therefore suggest that altered processing of *RBP10* mRNA after DRBD18 depletion was not caused by retention in the nucleus alone.

Composition and associations of proteins of DRBD18 mRNPs

There are two lines of evidence that DRBD18 is involved in mRNA export. The strongest argument is that its depletion in procyclic forms causes retention of mRNAs in the nucleus (Mishra et al. 2021). The second type of evidence is from protein–protein interactions. An RNA-independent interaction with MTR2 and MEX67 was confirmed by coimmunoprecipitation in the presence of RNase (Mishra et al. 2021) and a two-hybrid interaction with MTR2 was demonstrated. In addition, nonquantitative mass spectrometry results with DRBD18 preparations from procyclic forms suggested copurification not only with MEX67 and MTR2, but also some nuclear pore components and various RNA-binding proteins (Lott et al. 2015). Since label-free quantitative mass spectrometry is now available, we used it to reexamine DRBD18 protein associations in more detail. As some direct interactions of DRBD18 had already been demonstrated, we decided to focus

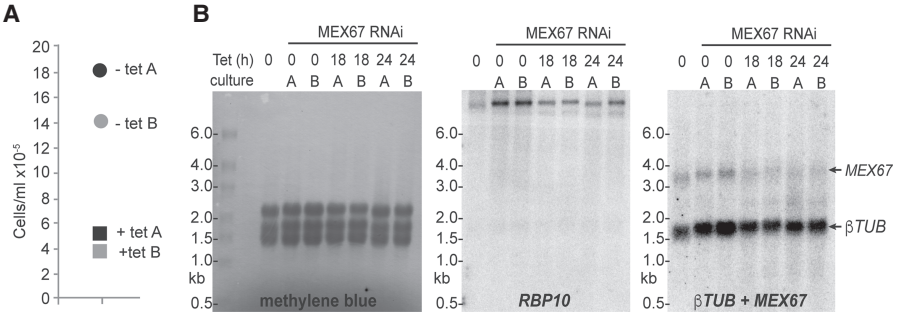

FIGURE 2. Effect of *MEX67* RNAi *on RBP10* mRNA and expression levels. (*A*) Cell densities in two cultures 24 h after addition of tetracycline to induce *MEX67* RNAi. (*B*) Northern blots using RNA from the two cultures, with detection of rRNA (methylene blue) and of *RBP10* and β-tubulin mRNAs.

on the components of DRBD18-containing "messenger ribonucleoprotein particles" (mRNPs) and their interactions, including other proteins that might be bound via RNA. DRBD18 with a carboxy-terminal TAP (tandem affinity purification) tag was expressed from one allele in bloodstream forms. Attempts to delete the other (untagged) copy failed. This might mean that the tagged protein has impaired function, but it is also possible that after the deletion, the amount of protein is insufficient for growth. It was not possible to obtain cells expressing DRBD18 with amino-terminal tags. We used just a single step of purification in order to avoid loss of material and interaction partners. DRBD18-TAP was pulled-down using IgG beads, then released by Tobacco Etch Virus (TEV) protease cleavage at a site within the tag. The resulting mixture was analyzed by quantitative mass spectrometry (Supplemental Table S2). GFP-TAP and TAP-ZC3H28 (Bishola Tshitenge and Clayton 2021) served as controls. ZC3H28 is an RNA-binding protein that is predominantly in the cytoplasm (Bishola Tshitenge and Clayton 2021). In addition, we compared the list of specifically associated proteins with that from a separate analysis of the putative splicing factor RBSR1 (Waithaka et al. 2022), because it is a nuclear protein (Supplemental Table S2, sheet 1).

Figure 3 shows analysis of the mass spectrometry data using the Perseus algorithm (Tyanova et al. 2016). Perseus yields probability values that are adjusted for multiple testing but handles absent values by substituting simulated intensities (Tyanova et al. 2016), which can give misleading results. In the following discussion, we therefore consider two overlapping sets of proteins as being significantly enriched. All of them were present in all three DRBD18 preparations. The first set are those that had adjusted *P*-values of less than 0.01, with greater than fourfold enrichment as calculated in Perseus. The second set were absent in the relevant control preparations but were not registered as significant by the algorithm because of the simulated values (Supplemental Table S2, sheet 1). The comparisons showed that 75 proteins were present specifically with DRBD18-TAP (Fig. 3; Supplemental Table S2, sheet 1). Of these, 19 (25%) were related to nuclear-cytoplasmic transport. These included eight components of the outer ring of the nuclear pore, which were (with the exception of single peptides from Sec13) absent from the other preparations. DRBD18 was also specifically associated with three nuclear pore components that comprise the Nup76 complex, which interacts with MEX67 (Degrasse et al. 2009; Obado et al. 2016, 2017). Other mRNA-export related associations were with MTR2, an importin-β subunit, a Ran binding protein (RANBP1), a transportin-2-like protein, and a possible Ran-GAP. MEX67 was not identified as being associated with DRBD18 because it was found in only two of the three DRBD18 preparations (three and four peptides); however, it was absent in all the others. GLE2 was not detected. EIF4AIII was present with DRBD18 but also in two of the three ZC3H28 preparations. One exon junction complex component, an NTF2-domain-protein (Bercovich et al. 2009), was detected with DRBD18, ZC3H28 and RBSR1, but not the GFP control. Just one or two peptides from the Mago-Nashi component of the exon junction complex were found with DRBD18 and ZC3H28 but not GFP.

Intriguingly, another quarter of the proteins associated with DRBD18 were components of the mitochondrial RNA editing machinery (Fig. 3). Although there is currently no evidence that DRBD18 is in the mitochondrion, and it lacks an amino-terminal targeting sequence, this probably warrants future investigation. Common contaminants that were found with both ZC3H28 and DRBD18 included histones, ribosomal proteins, and glycoytic enzymes; these are not highlighted in Figure 3.

The mass spectrometry results revealed other proteins related to mRNA metabolism that were enriched in DRBD18-containing mRNPs. The cytoplasmic RNA-binding proteins ZFP2 (Hendriks et al. 2001) and ZC3H30 (Chakraborty and Clayton 2018), and the nuclear RNA-binding protein RBP33 (Fernandez-Moya et al. 2014) were found specifically with DRBD18; and the RNA-binding proteins ALBA2, DRBD4, ZC3H39, ZC3H40, and ZC3H41 were found with both DRBD18 and ZC3H28. Proteins that were pulled down by RBSR1 as well included the poly(A) binding proteins, some splicing factors, the RNA interference Argonaute protein AGO1, the translation initiation complex EIF4E3/4G4, and the RNA-binding proteins ALBA3, DRBD2, DRBD3/PTB1, RBP42, ZC3H9, ZC3H34, UPF1, HNRNPH/F, and TRRM1 (Supplemental Table S2). Many of these proteins are likely to associate via RNA. Unlike ZC3H28 mRNPs, DRBD18 mRNPs did not contain MKT1, LSM12, XAC1, and PBP1 (Fig. 3B), which are the components of a complex that promotes translation and mRNA stability in the cytoplasm (Singh et al. 2014; Nascimento et al. 2020). Association with PUF1, PUF6, ZC3H35, and ZC3H18 was also specific to ZC3H28 (Fig. 3B). These results are consistent with at least partially different functions for ZC3H28 and DRBD18.

We next checked the location of DRBD18, since existing results were contradictory: Lott et al. (2015) reported that DRBD18 was in the cytosol whereas carboxy-terminally GFP-tagged DRBD18 was in the nucleus (Dean et al. 2016). After cell fractionation, the control cytoplasmic protein RBP10 and the nuclear exoribonuclease XRND (Li et al. 2006) showed the expected distributions. DRBD18 was associated with both compartments (Supplemental Fig. S1C,D), which is consistent with a role in export of mRNAs from the nucleus.

Long *RBP10* transcripts are mainly in the nucleus after DRBD18 depletion, while shorter ones are exported

Our mass spectrometry results (Fig. 3) supported previous observations implicating DRBD18 in mRNA export.

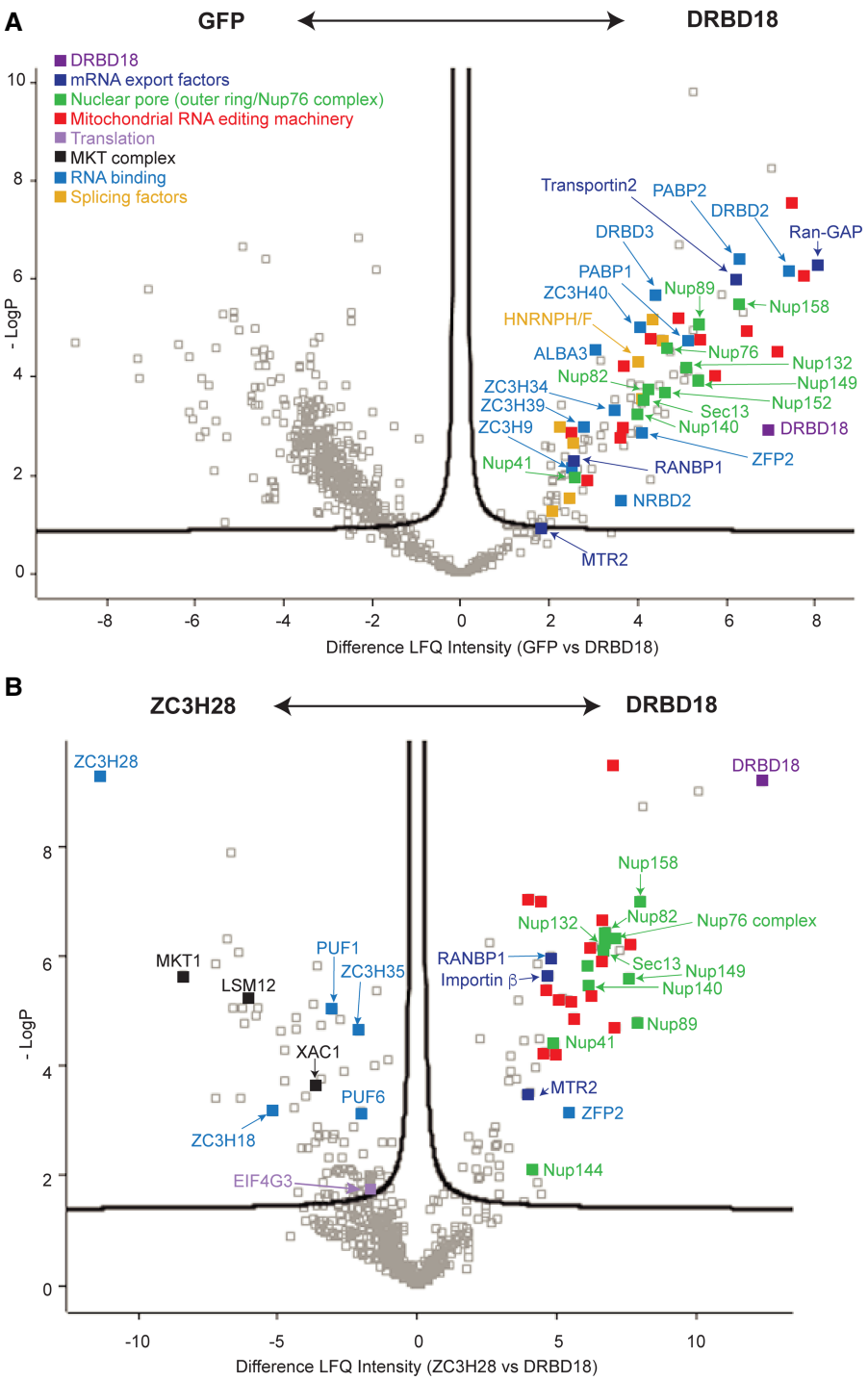

FIGURE 3. Proteins associated with DRBD18. (*A*) Volcano plot showing proteins that were significantly enriched with DRBD18-TAP relative to TAP-GFP. The *x*-axis shows $\log_2$ enrichment, while the *y*-axis is $-\log_{10}$ of the *P*-value generated by the Perseus algorithm (Tyanova et al. 2016). Proteins outside the curved lines are calculated to be significantly enriched according to the method described in Keilhauer et al. (2015), using a double-sided two sample *t*-test with permutation-based FDR of 0.01, and with parameter S_0 (for minimal fold change) set to 0.1. Spots for proteins that are mentioned in the text and showed specific enrichment with ZC3H18 are labeled; these were present in all three DRBD18 preparations, had *P*-values of less than 0.01, and were greater than fourfold enriched as calculated in Perseus. Spots that are not labeled include ribosomal proteins, abundant enzymes, and RNA processing factors. In addition, we have not labeled spots for proteins with anomalous enrichment values and probabilities due to the replacement, by the Perseus algorithm (Tyanova et al. 2016), of absent values (protein not detected) by random ones. For the DRBD18-GFP comparison, these include ALBA2, DRBD4, ZC3H41, and RBP42 and Nup109, all of which copurified with DRBD18 but not GFP. (*B*) Volcano plot showing proteins that were significantly enriched with DRBD18-TAP relative to TAP-ZC3H28. The data for ZC3H28 and GFP were described previously (Bishola Tshitenge and Clayton 2021). Details and color codes are as in panel *A*, except that some proteins that were selectively enriched with ZC3H28 are labeled.

Inhibition of export alone was however not sufficient to cause accumulation of shorter *RBP10* mRNA variants (Fig. 2), and the previous publication had already indicated that DRBD18 depletion affects the distributions of only some mRNAs (Mishra et al. 2021). We hypothesized, therefore, that DRBD18 binds to possible processing sites in the *RBP10* 3′-UTR, and that this not only promotes efficient export of the long mRNA, but also inhibits the use of the alternative sites. If this were true, DRBD18 RNAi might prevent export of longer *RBP10* mRNAs, while simultaneously allowing them to be processed to shorter versions. To test this, we examined the locations of the different *RBP10* mRNAs by single-molecule fluorescent in situ hybridization. To distinguish the different mRNAs, we used four probes (Fig. 4A). Details of the results are in Supplemental Table S3. The shortest species detected had signals from the coding region only (pink probe 1) and were mostly nuclear: these could represent precursors that were still undergoing transcription. The next shortest mRNAs detected hybridized only to probes 1 (coding region) and 2. These must be 2–7 kb long (Fig. 4A,B). The mRNAs that hybridize to probes 1 and 3 are at least of intermediate (or medium) length (5–7 kb long); and those that hybridize to probes 1, 2, and 4 or 1, 3, and 4 are more than 7.5 kb long (Fig. 4A,B). On average, 5% of the mRNAs hybridized only to probes 1 and 4: this is physically unlikely and may indicate the background failure rate for hybridization of probes 2 or 3 (Supplemental Table S3). The numbers of mRNAs hybridizing to probes 2 or 3 only were unexpectedly high (average 17% of signals); although these do not match any other sequences in the Lister427 genome, they do contain many low complexity sequences which might result in cross-hybridization. RNAs that hybridized to the green probe only, or to the red and green probes, could be artifacts, or represent products either from degradation, or from alternative splicing using 3′-UTR-internal signals (Kramer 2017). There were very few of either and the numbers were not reproducibly affected by *DRBD18* RNAi. If all mRNAs hybridizing to the coding region probe were included, wild-type cells contained 3–3.5 *RBP10* mRNAs per cell; this is consistent with previous calculations (Fadda et al. 2014), which suggested four *RBP10* mRNAs per cell.

To be certain that only specific signals were counted, we decided to consider only the long, medium and short mRNAs (defined above) that hybridized to at least two probes. In wild-type cells, about 70% of *RBP10* mRNAs were longer than 7.5 kb; 17% of the long mRNAs and 9%–12% of the short or medium ones were in the nucleus (Fig. 4C). Cells containing the *DRBD18* RNAi plasmid, but with no induction, showed similar values to wild-type (Fig. 4C,D). After RNAi induction, the total number of full-length *RBP10* mRNAs was roughly halved (Student *t*-test *P*-value 2×10^{-6}) (Fig. 4D) and the number in the cytoplasm had decreased sevenfold (Student *t*-test *P*-value <0.0002). Thus, most of the full-length mRNAs that remained were now in the nucleus (Fig. 4C,D). Although the median number of full-length mRNAs in the nucleus had increased about 1.5-fold, the difference was not statistically significant. This mirrored the northern blot results. The amounts of medium-length and shorter *RBP10* mRNAs without RNAi were higher than expected from the northern blots, while the 1.4-fold increase after RNAi was unexpectedly small. However, the short and medium mRNAs were more than fourfold more abundant in the cytoplasm than in the nucleus at most time points, indicating that their export was unaffected by DRBD18 depletion. We know that reporter mRNAs with shortened *RBP10* 3′-UTRs can be translated (Bishola Tshitenge et al. 2022), so the reduced amount of RBP10 protein after DRBD18 RNAi (Fig. 1B) is probably caused by the decrease in total cytosolic *RBP10* mRNA (Fig. 4C,D).

Effects of DRBD18 depletion on mRNA abundance

DRBD18 depletion affected the length of *RBP10* mRNA, but not tubulin mRNAs. To find out whether *DRBD18* RNAi affects processing of additional mRNAs, we first examined the transcriptomes of *DRBD18*-depleted bloodstream forms (Supplemental Table S1) manually using the Integrative Genomics Viewer (Robinson et al. 2011), focusing on mRNAs with long 3′-UTRs. We found several that appeared to have altered processing. Northern blotting confirmed that one example, *DRBD12* mRNA—which is usually about 8.4 kb long—accumulated as a 2.6 kb mRNA from 12 h *DRBD18* RNAi onward, with an intermediate band of about 4 kb (Supplemental Fig. S2). We therefore undertook more detailed analysis of all available data sets. In addition to our bloodstream-form transcriptomes (12 h RNAi, whole-cell total RNA, processed with oligonucleotides and RNase H to deplete rRNA) (Supplemental Table S1), we examined the transcriptome data from procyclic forms after 19 h RNAi (Mishra et al. 2021). The procyclic-form RNA was poly(A)$^+$ purified, so we could be certain that it was processed. Moreover, results for procyclic forms are available both from whole cells and from the cytoplasmic fraction (Supplemental Table S4). Transcripts that are more abundant in the whole-cell mRNA than in the cytoplasmic mRNA are either poorly exported, associated with the nuclear membrane, or unstable in the cytoplasm.

We first looked at effects on mRNA abundance using the reads from coding regions, using DeSeq2 (Love et al. 2014; Leiss and Clayton 2016). There was only weak correlation between the effects in bloodstream and procyclic forms (Pearson correlation coefficient 0.32) (Supplemental Fig. S3A). However, considering only the significantly affected genes (adjusted *P*-value <0.05, change of at least 1.5-fold) there was considerable overlap. Using a set of "unique" genes that excludes paralogues (Supplemental Table S5, sheet 4), 397 mRNAs were decreased in

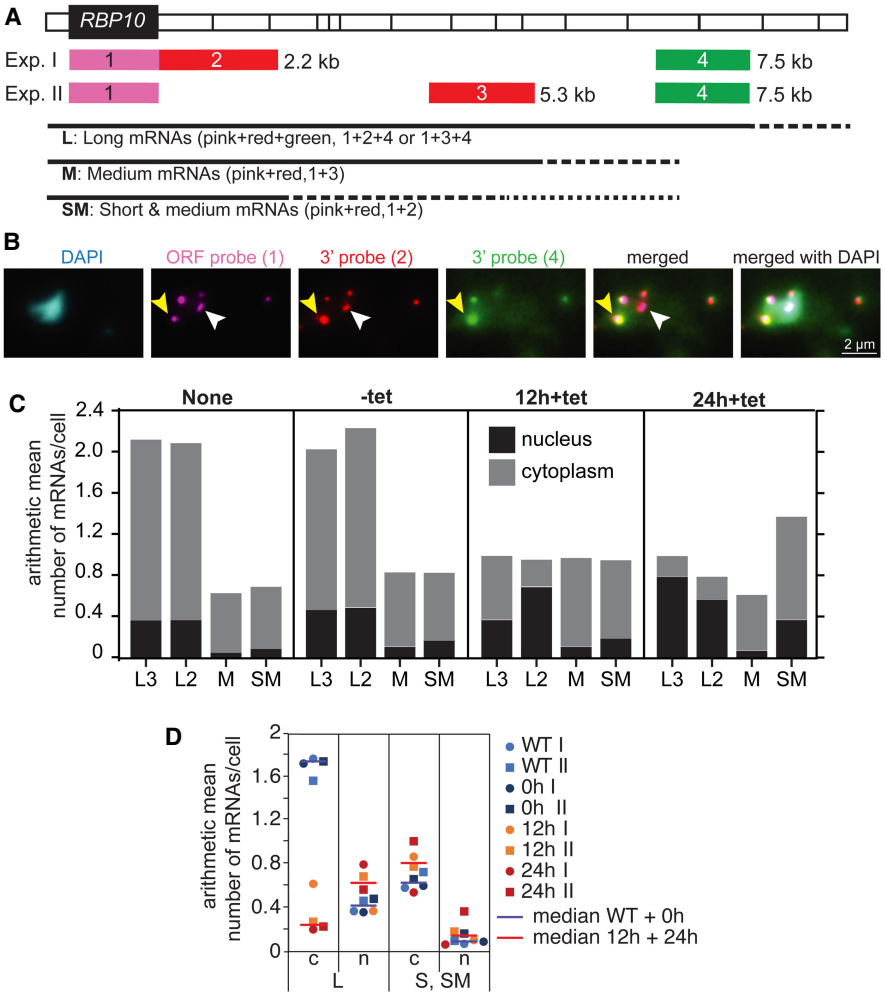

FIGURE 4. Long *RBP10* mRNAs remain in the nucleus after DRBD18 depletion. Individual *RBP10* mRNAs were detected by single-molecule fluorescent in situ hybridization with or without DRBD18 depletion for 12 or 24 h. In each case, 50 cells were analyzed. (*A*) Map showing the positions of the probes on the *RBP10* mRNA, approximately to scale. The 3′-UTR divisions on the map correspond to those in Figure 2. The bars indicate the extent of mRNAs detected by the probe combinations. The dotted lines indicate that RNAs hybridizing to the different probe sets could have various lengths. The distances next to the bars indicate the position of the probe end on the mRNA. "Exp.": Experiment. (*B*) Example image of *RBP10* mRNAs in a wild-type cell, detected by single-molecule fluorescent in situ hybridization using probes 1, 2, and 4 (experiment I in panel *A*). DNA was stained with DAPI. The white arrow indicates an mRNA that hybridizes only to the coding region (open reading frame, ORF) and 3′ probe 2; the yellow arrow points to one of the four mRNAs that hybridize to probes 1, 2, and 4. (*C*) Mean numbers of long, medium, and short mRNAs per cell under different conditions. The data were from 50 cells. "None": Cells with no RNAi plasmid; "-tet": RNAi cells without tetracycline; and cells after 12 or 24 h tetracycline treatment. All slides were hybridized with the red, pink and green probes, and results for hybridization with the two different red probes are shown separately. "L3" indicates long mRNAs that hybridized with probes 1, 3, and 4 (experiment II), and "L2" indicates long mRNAs that hybridized with probes 1, 2, and 4 (experiment I). The black bars indicate mRNAs in the nucleus and the gray bars, mRNAs in the cytoplasm. Numbers for mRNAs hybridizing to only one probe, or red and green only, are in Supplemental Table S3. (*D*) Compilation of data from panel *C*, to allow comparison of the numbers of full-length and shorter mRNAs in the nucleus and cytoplasm. The key to symbols is on the *right*.

bloodstream forms, and 81 in procyclic forms; of these, 61 were decreased in both (Supplemental Table S5, sheet 4) (Fisher test $P = 4 \times 10^{-60}$). There were no enriched functional categories among the mRNAs that were decreased in both forms.186 mRNAs were 1.5-fold significantly ($P_{adj} < 0.05$) increased in bloodstream forms and 304 in procyclic forms; of these, 72 were increased in both (Supplemental Table S5, sheet 4) (Fisher test $P = 4 \times 10^{-51}$).

In bloodstream forms, a minority of the RNA abundance changes after DRBD18 RNAi may have been due to the onset of growth arrest, or to stress. Examples include decreases in mRNAs encoding some cytoskeletal proteins, enzymes of glucose or glycerol metabolism, and increases in mRNAs encoding membrane proteins and mitochondrial proteins; such mRNAs also showed equivalent changes in the growth-arrested stumpy form (Naguleswaran et al. 2018; Silvester et al. 2018; Quintana et al. 2021). For the procyclic forms there was no significant overlap between the effects of DRBD18 depletion and the transcriptomes of growth-arrested metacyclic forms (Christiano et al. 2017).

DRBD18 is associated with mRNAs that have long 3′-UTRs containing polypyrimidine tracts

To find out whether mRNAs that were affected by *DRBD18* RNAi were also bound by DRBD18 at steady state, we purified DRBD18-TAP, released the protein with TEV protease, and sequenced both the eluted RNAs and those that did not bind to the beads (Supplemental Table S7). We found 226 mRNAs that were at least twofold enriched (normalized elute reads divided by corresponding unbound reads) in all three replicates, and defined these as "bound" RNAs (Supplemental Table S7, sheet 1). There was no overall correlation between binding of mature transcript by DRBD18 and effects on transcript abundance (Supplemental Fig. S3B). This argues against DRBD18 having either a stabilizing, or

destabilizing, effect on bound mRNAs. There was also no overall correlation with retention in the nucleus after DRBD18 depletion (Supplemental Table S4, sheet 5 and Supplemental Table S7, sheet 1; Mishra et al. 2021).

Some trypanosome RNA-binding proteins preferentially bind to mRNAs with long coding regions or untranslated regions (Erben et al. 2021). Although overall binding of DRBD18 did not correlate with either mRNA (Supplemental Fig. S3C) or 3′-UTR length (Fig. 5A), the 3′-UTRs of the "bound" mRNAs were indeed significantly longer than those of mRNAs that were not bound (Fig. 5A,B). ZC3H28 also preferentially binds mRNAs with long 3′-UTRs (Bishola Tshitenge and Clayton 2021), but there was less than 50% overlap in the reproducibly bound transcripts (Supplemental Table S6, sheet 1), consistent with the divergent mRNP compositions (Fig. 3B).

Lott et al. (2015) found a poly(A)-rich motif in mRNAs that were decreased after DRBD18 RNAi, and a CACCCAC motif in mRNAs that increased after DRBD18 RNAi. The functions of those motifs were however not tested for roles in either the RNAi response or DRBD18 binding. When the 3′-UTRs of our "bound" mRNAs were compared with length-matched controls (Supplemental Table S6, sheet 3) using MEME (Bailey et al. 2015), in the discriminative mode and looking for 6–20 nt motifs, three sequences were found (Fig. 5C). Of these, the one that was present in the most mRNAs (157/163) was a PPT. Only two motifs were identified by MEME in the differential enrichment mode, with lower significance, but they were again PPTs (Fig. 5C). The binding preference is not due to the DRBD3 that is present in the mRNP (Fig. 3A), since the consensus binding site for DRBD3 is UU(C/U)CCCUCU (Das et al. 2015). These results do not show that DRBD18 binds to PPTs—only that PPT motifs are enriched in its target mRNAs. PPTs by themselves are also clearly insufficient to distinguish between DRBD18-bound and -unbound mRNAs. For example, there are 528 copies of $(U)_8$ located in 138 of the 163 "bound" mRNA 3′-UTRs, but the 163 control "unbound" mRNAs also contain 157 copies of $(U)_8$, with up to seven copies per 3′UTR.

Effects of DRBD18 depletion on mRNA processing

To look for transcriptome-wide effects of DRBD18 depletion on mRNA processing, we compared the effects of RNAi on 3′-UTR reads with the effects on the coding region reads. The results for the 3′-UTRs are not as reliable as those for the coding-regions for two reasons. The 3′-UTRs contain many low-complexity sequences which may be repeated elsewhere in the genome, and more importantly, polyadenylation site mapping in the database is unreliable. Many genes have no annotated 3′-UTRs, and the annotated ones are quite often shorter than those suggested by visual examination of the read densities. This means that some effects of DRBD18 depletion will have been missed. Nevertheless, for both bloodstream forms and procyclic forms, some mRNAs showed clear discrepancies between coding region and 3′-UTR reads (Supplemental Fig. S3E; Supplemental Fig. S4A), and the differences were especially marked when procyclic-form cytoplasmic poly(A)$^+$ RNAs were examined (Supplemental Fig. S3B). To find mRNAs with truncated 3′-UTRs after RNAi, we chose those for which the change in 3′-UTR reads was at least 1.5-fold lower than the change in the coding region reads. This was true for 144 bloodstream-form whole-cell mRNAs, 142 procyclic-form whole-cell mRNAs, and 184 procyclic-form cytoplasmic mRNAs (Supplemental Table S5, sheet 2). Using the list of genes without paralogues ("unique gene list"), 190 coding regions showed loss of 3′-UTR reads. There were substantial overlaps in the lists (Fig. 5D; Supplemental Fig. S4E) with two-thirds of the mRNAs that were affected in bloodstream forms also showing 3′-UTR loss in procyclic forms. We also found that shortening of the mRNAs often, but not always, led to increases in mRNA abundance. Again, counting only genes from the "unique" list, for bloodstream forms, 48 of the affected mRNA coding regions increased significantly in read density after RNAi and 10 decreased, while for procyclic forms, 79 increased and none decreased.

Our working hypothesis at this stage was that binding of DRBD18 to PPTs might simultaneously prevent their recognition by the splicing machinery, and serve as a platform for recruitment of mRNA export factors. Subsequent to this, the DRBD18 might either remain bound to the 3′-UTR, or be replaced by other proteins that preferentially bind PPTs. We therefore checked whether DRBD18 was bound, at steady state, to the mRNAs that showed relatively reduced 3′-UTR reads after DRBD18 depletion. Results are plotted in Figure 5E and F. Overall, the mRNAs that had reduced 3′-UTR reads showed significantly more association with DRBD18 than the remaining mRNAs (Fig. 5F), although a minority showed no apparent association (Fig. 5; Supplemental Fig. S3B). *RBP10* mRNA was more than twofold enriched in two of the three replicates, and *DRBD12* mRNA was at least 2.7-fold enriched in all three. This suggests that in some, but not all cases, the DRBD18 that prevents additional processing may remain attached to 3′-UTRs after mRNA export.

To analyze alternative processing in detail, we focused on genes that showed a relative decrease in 3′-UTR reads in both procyclic and bloodstream forms (Supplemental Table S5, sheet 1), and for which the 3′-UTR had a single alternative processing site: the latter condition made quantitation possible. An example, Tb927.8.1580, is shown in Figure 6. In the absence of tetracycline, reads were fairly uniformly distributed across the mRNA, the 3′-UTR of which is 1 kb long (Fig. 6A; Supplemental Text 2). After RNAi induction, the read density over the coding region increased (Fig. 6A,B), but there was an abrupt drop about 330 nt into the 3′-UTR, resulting in a read density that was lower than in

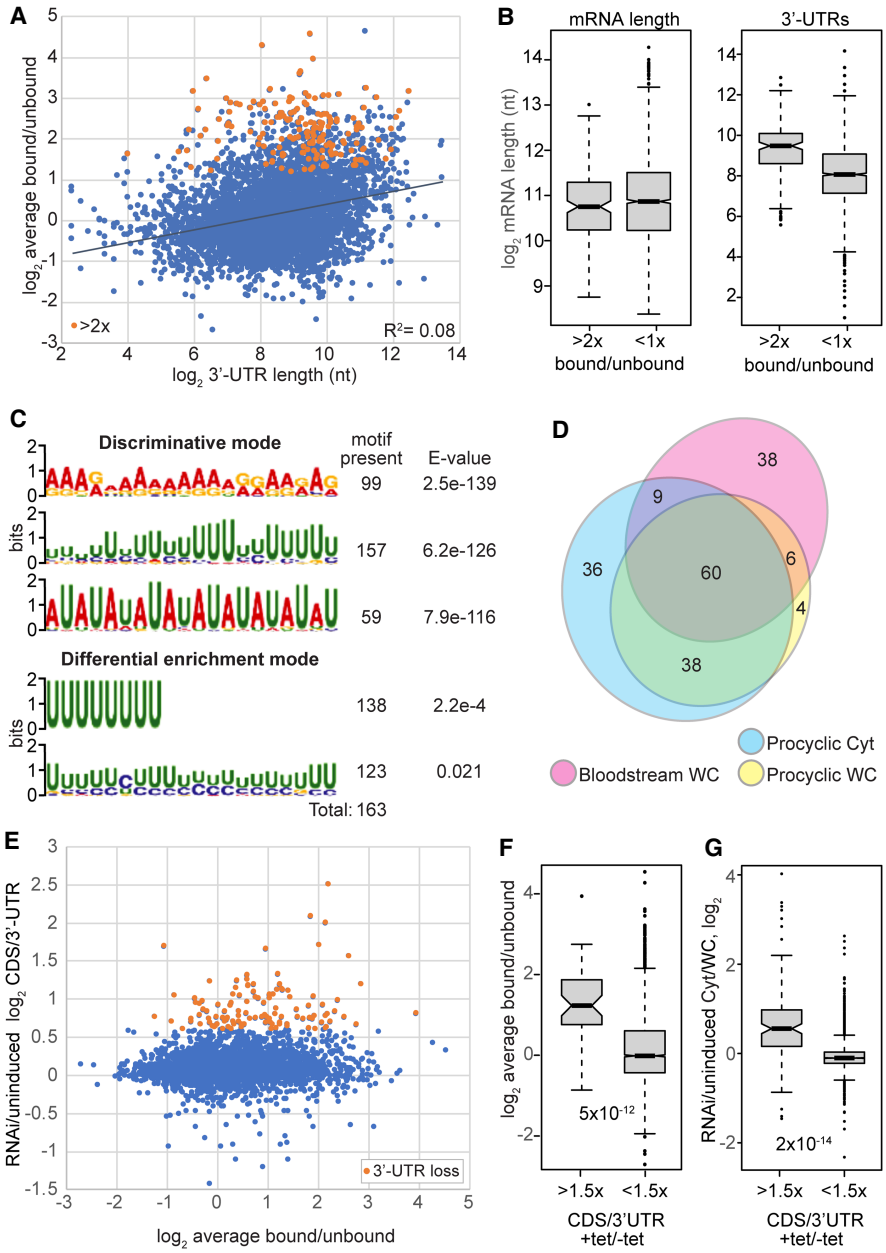

FIGURE 5. DRBD18 RNA binding and effects on the transcriptome. (*A*) Tagged DRBD18 was pulled down from bloodstream-form extracts, then associated (bound) RNAs were released using TEV protease. RNA that did not bind to the affinity beads served as the negative control ("unbound"). The average bound/unbound ratio for three biological replicates is on the *y*-axis and the annotated 3′-UTR length on the *x*-axis, both on a log scale. Each spot represents one open reading frame, with just one representative of repeated genes. The orange spots are genes with at least 10 reads in all samples and with at least twofold enrichment in all three comparisons. (*B*) Annotated mRNA lengths and 3′-UTR lengths for bound mRNAs (bloodstream forms, at least twofold enrichment in all three comparisons) and for mRNAs with, on average, less than onefold enrichment are plotted. The box plot shows the median and the box includes the interquartile range. Whiskers are up to 1.5-fold the inter-quartile ranges, and spots are outliers. (*C*) The 3′-UTRs of the 163 bound mRNAs (bloodstream forms) were compared with those of a length-matched set of mRNAs with less than onefold enrichment using MEME, using the discriminative mode and the more stringent differential enrichment modes. To the *right* are the numbers of bound mRNAs that contained the motif, and the probabilities from the program. (*D*) The $\log_2$ fold change for each coding region was subtracted from the $\log_2$ fold change of the corresponding 3′-UTR. If the result was less than −0.584 (1.5-fold excess of coding region reads) the mRNA was identified as having potential loss of 3′-UTR reads due to alternative polyadenylation. The numbers of genes for all three analyzed conditions are shown to scale: bloodstream-form whole-cell (WC) RNA, rRNA-depleted; procyclic-form whole-cell RNA, poly(A)$^+$; and procyclic-form cytoplasmic RNA (Cyt), poly(A)$^+$. Results were calculated for the "unique" gene set. (*E*) Effect of RNAi on the coding region/3′-UTR read ratio, plotted against the average DRBD18 RNA binding, in bloodstream forms. Results for unique coding regions and unique representatives of multigene families are shown. The orange dots are results for mRNAs that showed loss of 3′-UTR reads, as defined in *D*. (*F*) DRBD18 binding of mRNAs that showed loss of 3′-UTR reads relative to coding region reads, compared with binding of mRNAs that showed no (or less) 3′-UTR read loss. The Student *t*-test result is indicated. Results are for bloodstream forms. (*G*) Ratio of CDS reads in the cytoplasm (Cyt) relative to reads from the whole cell (WC), for procyclic forms. The Cyt/WC ratio with normal DRBD18 levels was subtracted from the ratio after RNAi. Positive results suggest an increase in the proportion of mRNA that was exported after RNAi. Results for the subset of mRNAs that showed loss of 3′-UTR reads relative to CDS reads are compared with the results for all other mRNAs. The Student *t*-test result is indicated.

the presence of DRBD18 (Fig. 6A). Scrutiny of the sequence reads (Fig. 6C; Materials and Methods section and Supplemental Fig. S5) revealed three new polyadenylation sites over a 40 nt region that is about 90 nt upstream of a 24 nt PPT, which is in turn 28-nt upstream of a new splice acceptor site (Fig. 6C; Supplemental Text S2, S3). The reads downstream from the new poly(A) sites represent either full-length mRNA, or a 3′-UTR-derived RNA (shown as [**] in Fig. 6A). We counted reads from the new processing site region for all replicates (Fig. 6C,D). It is important to note that we were analyzing steady-state mRNA levels and we do not know the relative half-lives of the different Tb927.8.1580 RNAs. The results therefore do not quantitate the absolute frequency of alternative processing, but they do allow qualitative assessment of the effects of DRBD18 depletion. The top two bar-graph panels in Figure 6D show the numbers of reads covering the novel poly(A) sites, normalized to the total number of the reads in each data set (Supplemental Table S5, sheet 6). The percentages of reads with alternative processing are shown as dot plots on the *right*. In the absence of tetracycline, all reads were wild-type (blue bars in Fig. 6D). After RNAi, considering whole-cell RNA, about 30% of the reads that were captured for the two upstream internal poly(A) sites, p(A)-1/2, were polyadenylated there (orange bars in Fig. 6D). 70% of the RNA would remain, and of this, about 30% terminated at poly(A) site 3. This analysis suggests that at least half of the mRNAs were alternatively processed at the 3′-end; the read density plots in Figure 6A suggest that the proportion was higher. Reads for the spliced leader acceptor site are shown in the lower part of Figure 6D. Only one read that showed this splicing event was observed in the presence of normal levels of DRBD18. After DRBD18 depletion, the number of full-length reads decreased, and 35% of the remaining reads were alternatively spliced. Interestingly, reads from alternatively spliced 3′-UTR-derived RNA were also present in the cytoplasmic fraction. The use of the alternative processing sites did not affect splicing of the downstream mRNA since there was no increase in intergenic reads (Fig. 6A) and no reproducible effect on the abundance of the downstream mRNA, Tb927.8.1590 (Fig. 6E).

We repeated this analysis for 11 more genes. Examples of screen shots are in Supplemental Figures S6 and S7, and quantification of alternative site use is in Figure 7, Supplemental Figure S8, and Supplemental Table S5, sheet 6. The gene sequences with identified processing sites are in Supplemental Text 2, and effects of RNAi on the abundances of the next downstream mRNAs are in Supplemental Figure S9. As expected from the read count values, the proportion of mRNAs that showed alternative polyadenylation was always higher after DRBD18 depletion. Usually, but not always, the shorter coding mRNA with alternative polyadenylation constituted a higher proportion of cytoplasmic RNA than the wild-type version; this could be because the shorter version was more readily exported, or had a longer cytoplasmic half-life, than the longer mRNA (e.g., Tb927.11.2500, Tb927.5.1130, and Tb927.7.18100 in Fig. 7). Reads from the distal parts of the 3′-UTR were lower in the presence of tetracycline since we had specifically selected for this; and the short noncoding spliced products that resulted from use of 3′-UTR-internal PPTs had varying relative abundances in the cytoplasm. There was no reproducible effect of DRBD18 depletion on the abundance or splicing pattern of the downstream mRNA (Fig. 7; Supplemental Figs. S8, S9; Supplemental Table S5, sheet 6). These observations suggest that at least for these mRNAs, loss of DRBD18 caused an increase in use of 3′-UTR-internal RNA processing signals, which was often accompanied by a decrease in the abundance of full-length mRNA. (We made a preliminary attempt to verify this for five mRNAs by northern blotting, but the mRNAs are present at only about one copy per cell and were not detected.) DRBD18 appears to have no role in defining the next exon, since the coding mRNAs immediately downstream from alternatively polyadenylated mRNAs were not reproducibly affected.

Finally, we looked at some of the mRNAs for which 3′-UTR reads increased, relative to coding-region reads, after DRBD18 RNAi (Supplemental Table S5, sheet 3). In some cases, the result was an artifact due to low read coverage, or the presence of sequences that were found several times in the genome. In other cases, the 3′-UTR was annotated so as to overlap the 5′-UTR of the downstream gene, making the source of reads unclear and with no evidence for novel processing events. However, we also found instances that were indeed caused by alternative splicing events. In these cases, the novel noncoding RNA that arose from within the 3′-UTR was more abundant than the new truncated mRNA containing the coding region. One example, Tb927.6.2830, is shown in Supplemental Figure S7.

Interestingly, after RNAi in procyclic forms the 215 mRNAs that showed possible nuclear retention (or decreased stability in the cytoplasm) were significantly longer than the remainder (Supplemental Fig. S4E, for definition see legend), and only five of them showed a possible change in processing. In contrast, the procyclic-form mRNAs with alternative processing mostly had increased ratios of cytoplasmic RNA to whole-cell RNA after RNAi (Fig. 5G; Supplemental Table S5, sheet 2). This suggested that the new, shorter mRNAs were either more readily exported, and/or had increased stability in the cytoplasm relative to the longer species.

Characteristics of alternatively processed mRNAs

We had eleven mRNAs for which we had mapped unique alternative processing sites and verified normal "wild-type" sites. This enabled us to investigate the difference

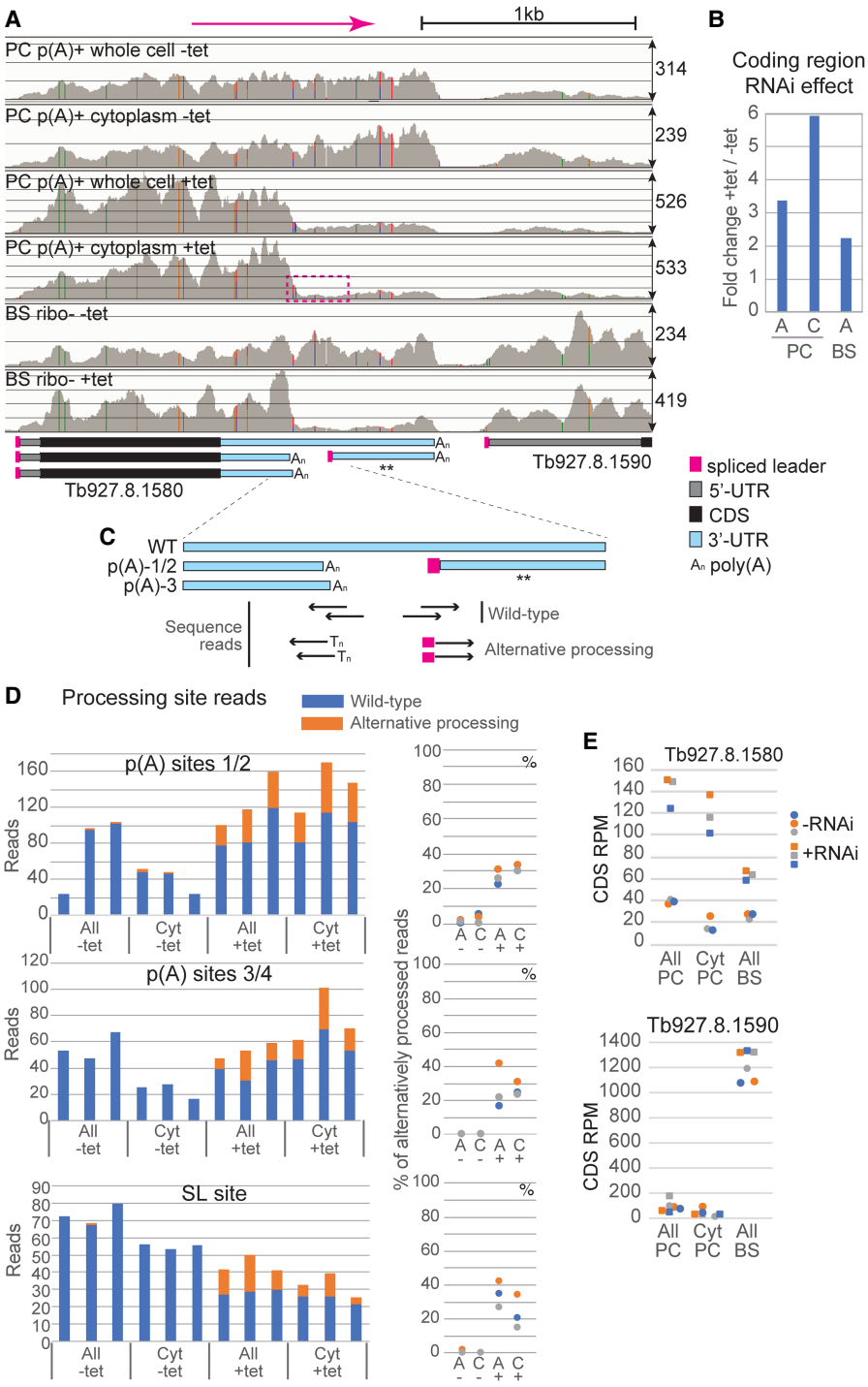

FIGURE 6. Effects of DRBD18 RNAi on the RNAs from Tb927.8.1580 and Tb927.8.1590. (*A*) Results from the Integrated Genome Viewer in the region containing Tb927.8.1580 and Tb927.8.1590. The direction of transcription and a scale bar are at the *top*. Representative views for all six types of samples are shown. The scales on the *y*-axis are different (see numbers on the *right*); the thin lines are drawn at intervals that show read depths of approximately 100. A map of the detected mRNAs is *below*, as for Figure 1. (*B*) Effects of RNAi on coding region reads ($\log_2$ fold change from DeSeq2). (*C*) Method to estimate the relative abundances of wild-type and alternatively processed reads. For more detail, see Supplemental Figure S5. Reads that spanned the sites were designated as wild-type, and those with polyadenylation or spliced leader sequences were classified as alternatively processed. "All"—whole-cell RNA; Cyt: cytoplasmic RNA. (*D*) The *upper* two bar graphs show numbers of reads (normalized to total data set size) that were either wild-type or alternatively processed at the different poly(A) sites; all of these events are directed by *trans* splicing at a single downstream site, results for which are shown in the *bottom* panel. The dot plots show the proportions of mRNA that were alternatively processed (A = All). (*E*) Reads per million for the coding regions of Tb927.8.1580 and the downstream gene, Tb927.8.1590, splicing of which directs wild-type Tb927.8.1580 polyadenylation. Results for all three replicates are shown for each data set.

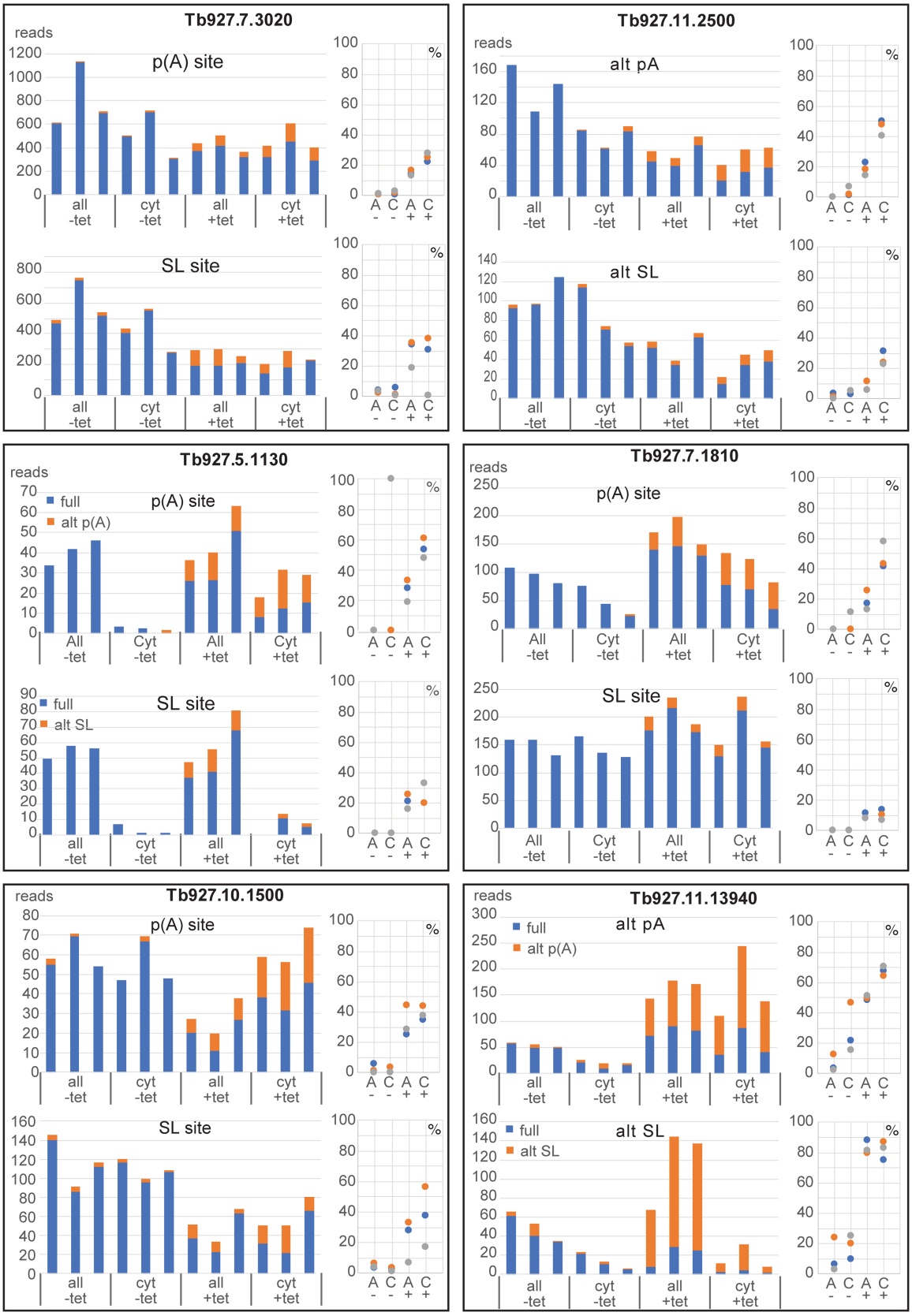

FIGURE 7. Effects of DRBD18 RNAi on additional mRNAs. For each gene, the *upper* two bar graphs show numbers of reads (normalized to total data set size) that were either wild-type or alternatively processed at the different poly(A) sites, and the *lower* panel shows mRNAs with or without the 3′-UTR-internal splice site. Details are as in Figure 6.

between normal and novel sites. We considered two main possibilities. One possibility is that DRBD18 inhibits alternative processing by binding specifically upstream or downstream from the alternative splice sites; the other is that DRBD18 can bind all splice acceptor signals, but that at the wild-type splice site, DRBD18 effects are inhibited by other sequence-specific proteins. We therefore compared sequences from −50-nt relative to the poly(A) sites to +50 downstream from the splice acceptor sites (Supplemental Text S3). In pair-wise comparisons, there were no reproducible differences between the wild-type and alternative sites in PPT length or composition, distance to the acceptor AG, poly(A)-polypyrimidine distance, or GC content downstream from the splice site. There were also no differences in the numbers of additional intron PPTs that did not direct splicing. Splicing and transcription elongation kinetics can be influenced by RNA folding (see, e.g., Turowski et al. 2020) but predicted folding energies of these RNA segments also showed no consistent differences. Finally, a comparison using MEME revealed no motifs that were specific to either the normal or the alternative sites. The sequences that differentiate wild-type from abnormal splice sites thus remain obscure. However, if wild-type sites are defined by various different proteins, with different specificities, or with binding sites that are more than 50 nt downstream, we would not have detected the relevant sequence motifs with this small sample.

DRBD18 depletion has polar effects on polycistronic transcription units

While examining the RNAi results, we noticed that some of the down-regulated genes appeared to be in clusters on the chromosomes. Most clusters were toward the ends of transcription units, but most transcription units were not affected (Fig. 8). Similar results were seen for procyclic forms (not shown). Manual examination of the affected regions did not reveal any common features: in particular, mRNAs with obviously altered processing were only sometimes at the start of a repressed region. RNAi did not cause significant decreases in any mRNAs encoding known proteins involved in transcription. These results might indicate links between mRNA processing or export and transcription elongation.

DISCUSSION

The mechanisms by which splice acceptor sites are chosen in animal cells have been investigated extensively. As in trypanosomes, the presence of a polypyrimidine tract is insufficient to define a splice acceptor site. Instead, wild-type exons are marked by the binding of specific proteins (Ule and Blencowe 2019). For example, proteins that contain serine-arginine rich regions (SR domain) as well as an RNA recognition motif (RRM), bind to exon sequences downstream from active splice sites (Howard and Sanford 2015). In trypanosomes, the only splicing signal that has been investigated in detail is the PPT. Results from one set of reporter experiments indicated that maximal splicing efficiency was obtained with PPTs of $(U)_{20}$ or with 17mers of poly(U) containing interpellated C residues (Siegel et al. 2005), but another series indicated that a 10mer PPT interrupted by a single "A" was as good as a 20mer (Matthews et al. 1994), and results of a recent screen suggest even more flexibility (Waithaka et al. 2022). In practice, most mRNAs are *trans*-spliced downstream from PPTs that are 12–24 nt long, with more U than C (Kolev et al. 2010); but for some acceptors the PPT is barely recognizable. (Tb927.11.5450 and Tb927.11.5460 in Supplemental Text S3 are examples.) Excellent PPTs are scattered throughout 3′-UTRs, but are used only if the normal splice site has been removed (see, e.g., Hug et al. 1994; Matthews et al. 1994; Schürch et al. 1994; Vassella et al. 1994). The most likely explanation is that splice sites that are upstream of coding regions are specifically chosen, as in animals; there is some extremely limited evidence for this from reporter experiments (Matthews et al. 1994; Hartmann et al. 1998).

We here show that DRBD18 preferentially binds mRNAs that contain several U-rich PPT motifs in their 3′-UTRs. Depletion of DRBD18 results in the accumulation of truncated mRNAs that arise from use of alternative splicing signals located in 3′-UTRs. In the examples examined, there was no evidence that DRBD18 loss had any effect on use of the normal splicing signal for the downstream mRNA. DRBD18 was already implicated in mRNA export (Mishra et al. 2021). We confirmed the association of DRBD18-containing mRNPs with proteins implicated in mRNA export and with the nuclear pore. Moreover, our results with bloodstream forms showed that the 8 kb wild-type *RBP10* mRNA was not exported from the nucleus in the absence of DRBD18. How can all of these results be interpreted? We considered several hypotheses.

The first possibility was that DRBD18 is required to define the position of the normal splice site. But if this were true, loss of DRBD18 should affect the abundance of the downstream mRNA or cause accumulation of RNAs that span the intergenic region. Neither effect was observed.

Another possibility was that alternative processing is constitutive, but the mRNAs with short 3′-UTRs, as well as the corresponding downstream spliced 3′-UTR fragments, are usually very unstable in the cytoplasm due to binding by DRBD18. If this were true, however, DRBD18 ought to destabilize the long mRNA as well. Moreover, there was no overall correlation between steady-state binding of mature transcript by DRBD18 and effects of RNAi on transcript abundance (Supplemental Fig. S3B), which argues against DRBD18 having either stabilizing, or destabilizing, effects on bound mature mRNAs. This hypothesis fails to explain why in some cases, DRBD18 depletion caused a decrease

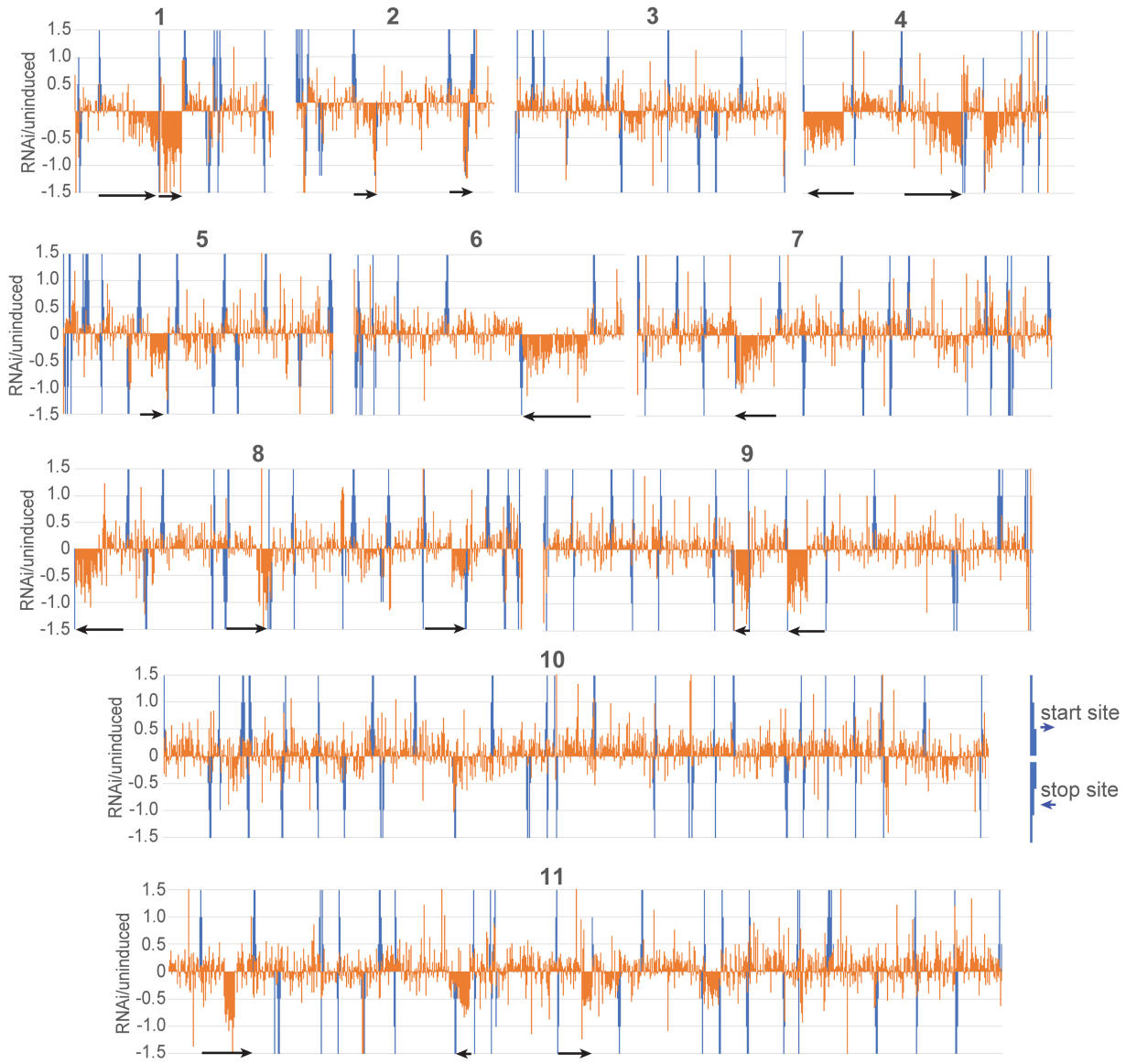

FIGURE 8. Effects of DRBD18 depletion show chromosomal clustering. The effects of DRBD18 depletion ($\log_2$ fold change) for unique genes (without paralogs) were plotted as orange bars according to the order of the genes on the chromosomes (positions not to scale). Start sites according to histone modifications (Siegel et al. 2009) and transcript orientation are upward blue bars, and stop sites are downward bars. Neighboring smaller bars indicate the orientation. The scale of the *y*-axis has been truncated so that all effects greater than 2.8-fold are shown as $\log_2(1.5)$. Black *horizontal* arrows indicate the directions of selected affected transcription units.

in the abundance of full-length mRNAs in favor of shorter variants. Conversely, the alternative—stabilization of the long mRNA by DRBD18 binding—fails to explain why the shorter mRNAs only appear after DRBD18 loss. Overall, hypotheses that depend on DRBD18 affecting mRNA half-life in the cytosol seem unlikely.

The next possibility depends on the known requirement for DRBD18 for the export of some mRNAs (Mishra et al. 2021). In the absence of DRBD18, its target mRNAs might get stuck in the nucleus and the 3′-UTR signals might become substrates for reprocessing. Once the processing had happened the shorter mRNAs would no longer require DRBD18 for export. However, at least for *RBP10* mRNA, the effects of DRBD18 depletion on splicing and polyadenylation appear not to be solely due to decreased export, because MEX67 depletion did not result in alternative processing (Fig. 2). This result indicates that effects on export from the nucleus cannot entirely explain the effects of DRBD18 action.

The hypothesis that we find to be most consistent with the data is shown in Figure 9. We speculate that in the nucleus, DRBD18 binds to PPTs in the 3′-UTRs of affected mRNAs. This prevents their recognition by the splicing machinery, while perhaps also simultaneously promoting

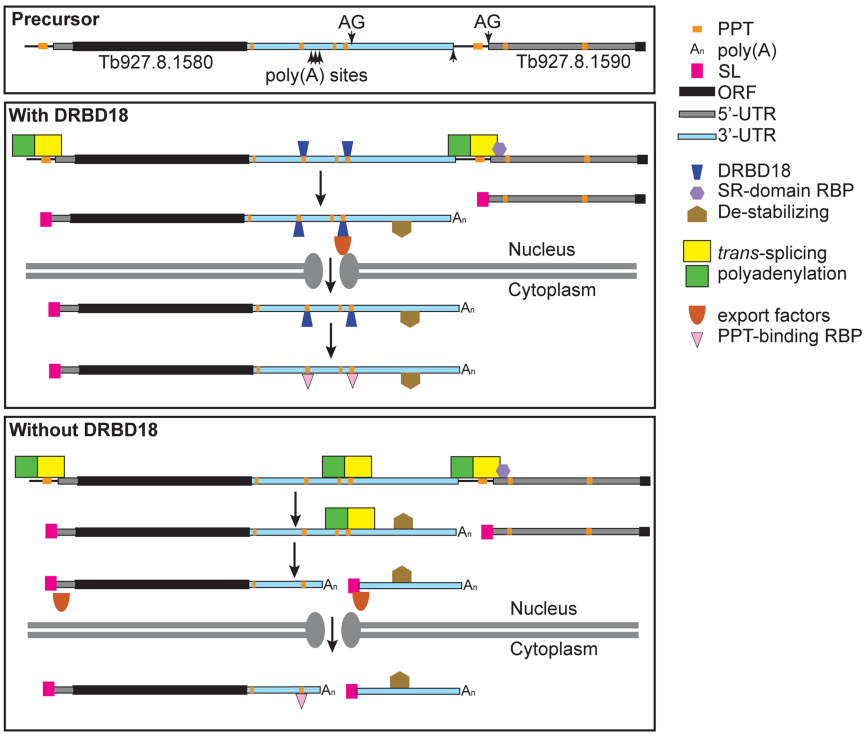

FIGURE 9. Model for the interaction between DRBD18 and *RBP10* mRNA. A key is on the Figure, which is explained in the text. RBP: RNA-binding protein.

mRNA export. (An alternative, for which there is less evidence, would be binding near the PPTs.) In most cases, the DRBD18 remains bound to the mRNA in the cytoplasm (Fig. 5E,F), but for a minority of affected mRNAs, the DRBD18 is replaced by other abundant PPT-binding proteins (e.g., D'Orso and Frasch 2001; Hartmann et al. 2007). When DRBD18 is depleted, the newly exposed alternative sites can be used in the nucleus, and shorter mRNAs accumulate. We sometimes found that the truncated mRNAs became more abundant in the cytoplasm than the longer version had been originally (Fig. 5G and Tb927.5.1130 and Tb927.11.13940 in Fig. 6); this might be because the longer version has destabilizing 3′-UTR motifs.

We do not know why the processing of some mRNAs is affected by DRBD18 loss, while others are unaffected, since PPTs are ubiquitously present in 3′-UTRs. We also do not yet know why wild-type splice sites are not inhibited by DRBD18, since in a (very small) sample no obvious sequence differences between them and the alternative sites were found. In humans, the sequences bound by different RNA-binding proteins often overlap, and it has been suggested that the context of a target sequence, or the secondary structure in the vicinity, is important in determining the extent of binding of particular proteins (Dominguez et al. 2018). The results of a recent study of a mouse RNA-binding protein indicated that its association and dissociation kinetics were affected by sequence context, and that association was enhanced if several binding sites were clustered (Sharma et al. 2021). The same may also apply to DRBD18: binding to mRNAs may be influenced by PPT frequency and context as well as competition with other proteins. The effects of DRBD18 could also be modulated by other proteins that bind the same mRNA precursor. By analogy with results from animal cells, DRBD18 activity at wild-type splice acceptor sites may be inhibited by proteins that show higher-affinity binding and specifically define exons. There is evidence for mRNA processing regulation by several nuclear trypanosome RNA-binding proteins (Stern et al. 2009; Gupta et al. 2013a,b, 2014; Waithaka et al. 2022). Further analysis of these and other nuclear RNA-binding proteins will be needed to understand how sites for trypanosome RNA processing are normally determined.

MATERIALS AND METHODS

Trypanosome culture

The experiments in this study were carried out using monomorphic *T. brucei* Lister 427 (Alibu et al. 2004) expressing the tetracycline repressor. The bloodstream-form parasites were cultured as routinely in HMI-9 medium supplemented with 10% heat inactivated fetal bovine serum at 37°C with 5% CO_2. During proliferation, the cells were diluted to 1×10^5 cells/mL and maintained in density between $0.2–1.5 \times 10^6$ (Clayton et al. 2005). For generation of stable cell lines, ~$1–2 \times 10^7$ cells were transfected by electroporation with 10 µg of linearized plasmid at 1.5 kV on an AMAXA Nucleofector. Selection of new transfectants was done after addition of appropriate antibiotic and serial dilution. The induction of RNAi was done using 100 ng/mL tetracycline, in the absence of other selective antibiotics.

Plasmid constructs

For endogenous tagging of DRBD18, a cell line with in situ TAP-DRBD18 was generated by replacing one endogenous copy of DRBD18 with a gene encoding a carboxy-terminally TAP-tagged DRBD18. For that, a construct with neomycin resistance gene plus TAP tag cassette was flanked on the 3′-end with a fragment of *DRBD18* 3′-UTR. Upstream on the 5′-end, the carboxy-terminal region of the *DRBD18* ORF without the stop codon was cloned in frame with the TAP tag. Prior to transfection in monomorphic bloodstream forms Lister 427, the plasmid (pHD3200) was cut

with Apa I and Xba I to allow homologous recombination. The tetracycline inducible construct for DRBD18 RNAi was done using a stem–loop vector targeting the coding region of DRBD18. The region used for the stem–loop vector was checked against off-targets using the RNAi target selection tool for trypanosome genomes (https://dag.compbio.dundee.ac.uk/RNAit/) as described in Redmond et al. (2003). Plasmids and oligonucleotides are listed in Supplemental Table S7.

RNA analysis and northern blotting

Total RNA was isolated from approximately 1×10^8 bloodstream-form trypanosome cells using either peqGold Trifast (PeqLab) or RNAzol RT following the manufacturer's instructions. To detect the mRNA by northern blot, 5 or 10 µg of the purified RNA was resolved on formaldehyde agarose gels and transferred onto nylon membranes (GE Healthcare) by capillary blotting and fixed by UV-crosslinking. The membranes were prehybridized in 5× SSC, 0.5% SDS with 200 mg/mL of salmon sperm DNA (200 mg/mL) and 1× Denhardt's solution, for an hour at 65°C. The probes were generated by PCR of the coding sequences of the targeted mRNAs, followed by incorporation of radiolabeled [α^{32}P]-dCTP and purification using the QIAGEN Nucleotide Removal Kit according to the manufacturer's instructions. The purified probes were then added to the prehybridization solution and the membranes were hybridized with the respective probes at 65°C for overnight (while rotating). After rinsing the membranes in 2× SSC buffer/0.5% SDS twice for 15 min, the probes were washed out once with 1× SSC buffer/0.5% SDS at 65°C for 15 min and twice in 0.1× SSC buffer/0.5% SDS at 65°C each for 10 min. The blots were then exposed onto autoradiography films for 24–48 h and the signals were detected with the phosphorimager (Fuji, FLA-7000, GE Healthcare). The signal intensities of the images were measured using ImageJ.

Total RNA used for sequencing analysis (RNA-seq) was prepared from cells collected 12 h after the induction of DRBD18 RNAi in monomorphic bloodstream forms cells (Lister 427). Induction of RNAi was done using tetracycline at a concentration of 100 ng/mL. The cells without induction of DRBD18 RNAi were used as controls. RNA was extracted using the phenol–chloroform separation method as described previously. The integrity of the total RNA was checked on a denaturing agarose gel. Afterwards, the ribosomal RNAs (rRNAs) were depleted from the total RNA samples using a cocktail of 131 50-base oligonucleotides complementary to rRNAs, combined with RNaseH (NEB, M0297S), as previously described (Antwi et al. 2016; Minia et al. 2016). Following rRNA depletion, the samples were subjected to DNase I treatment in order to remove oligonucleotides using the Turbo DNase Kit (Invitrogen, Thermo Scientific). The RNA samples were then purified using the RNA Clean & Concentrator -5 kit (ZYMO Research) following the manufacturer's instructions. The recovered purified RNA from both bound and unbound samples was then analyzed by RNA-seq.

RNA immunoprecipitation

A cell line expressing in situ carboxy-terminally TAP-tagged DRBD18 was used for the RNA immunoprecipitation. The TAP-tag consists of the protein A and calmodulin binding protein domains separated by a Tobacco Etch Virus (TEV) protease cleavage site. Approximately 1×10^9 cells expressing in situ C-TAP DRBD18 with a concentration of 1×10^6 cells/mL were pelleted by centrifugation at 3000 rpm for 13 min at 4°C. The pellet was washed twice in cold 1 × PBS and collected by centrifugation at 2300 rpm for 8 min at 4°C and then snap frozen in liquid nitrogen. The cell pellet was lysed in 1 mL of the lysis buffer (20 mM Tris pH 7.5, 5 mM $MgCL_2$, 0.1% IGEPAL, 1 mM DTT, 100 U RNAsin, 10 µg/mL leupeptin, 10 µg/mL Aprotinin) by passing 20 times through a 21G 1 × ½ needle using a 1 mL syringe and 20 times through a 27G 1 × ¾ needle using a 1 mL syringe. The lysate was cleared by centrifugation at 15,000*g* for 15 min at 4°C. Afterwards, the supernatant was transferred to a new microfuge tube and the salt concentration was adjusted to 150 mM KCl. 10% of lysate was collected as the input fraction for RNA extraction and 2% was used for western blotting. A total of 40 µL of magnetic beads (Dynabeads M-280 Tosyl activated, Invitrogen) coupled with Rabbit Gamma globulin antibodies (Jackson Immuno Research Laboratories) were washed three times with 500 µL of IP buffer (20 mM Tris pH 7.5, 5 mM $MgCl_2$, 150 mM KCl, 0.1% IGEPAL, 1 mM DTT, 100 U RNasin, 10 µg/mL leupeptin, 10 µg/mL aprotinin) then incubated with the cell lysate at 4°C for 3 h with gentle rocking. The unbound fraction was collected for RNA extraction (98%) and western blotting (2%). Subsequently, the beads were washed thrice with IP buffer. Two percent of each wash fraction was collected for protein detection. The TAP-tag was cleaved by incubating the beads with 500 µL of IP buffer containing 100 units of TEV protease (5 µL) with gentle rotation at 16°C for 2 h. The eluate was then collected afterward by magnetic separation for RNA isolation and protein detection. RNA was isolated from the input, the unbound and the eluate fraction using the peqGOLD TriFast FL (Peqlab, GMBH) according to the manufacturer's instructions. To assess the quality of the purified RNA, aliquots of the input, unbound and eluate fractions were resolved on formaldehyde agarose gels to check the integrity of the ribosomal RNAs.

Ribosomal RNAs were then depleted as described above and the fractions sent for RNA sequencing.

High-throughput RNA sequencing and bioinformatic analysis

RNA Sequencing was performed at the Cell Networks Deep Sequencing Core Facility of the University of Heidelberg. The library preparation was done using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760S). The libraries were multiplexed (six samples per lane) and sequenced with a NextSeq 550 system, generating single-end sequencing reads of about 75 bp. Raw data are available at Annotare as E-MTAB-9783 and E-MTAB-10735.

Sequence alignment and read counting were done using custom pipelines (Leiss and Clayton 2016; Leiss et al. 2016). Before analysis, the quality of the raw sequencing data was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Cutadapt (Martin 2011) was used to remove sequencing primers, then the sequencing data were aligned to *T. brucei* 927 reference genome using Bowtie (Langmead and Salzberg 2012) allowing 1 alignment per read, then sorted and indexed using SAMtools (Li et al. 2009). For RNAi, the reads that mapped to the open reading frames, 3′-UTRs and noncoding RNAs in the

TREU 927 genome were counted. Alignment and read-counting for the published results for procyclic-form whole-cell and cytoplasmic RNA (Mishra et al. 2021) were done using the same pipelines. Reads per million reads (RPM) were calculated after removal of rRNA reads, and the ratios of bound RPM to unbound RPM were calculated for each of the three purifications. To find reproducibly bound mRNAs, we considered only the lowest ratio for each mRNA.

In order to look for enrichment of particular functional characteristics, we used a list of unique genes modified from Siegel et al. (2010); this corrects for repeated genes and multigene families. To generate reads per million reads for the unique gene set, we first multiplied the reads for each gene in the list by the gene copy number (obtained in a separate analysis; see, e.g., Mulindwa et al. 2021). For the RIP-seq, the reads per million reads were counted and the ratios of eluate versus flow-through were calculated. An mRNA was considered as "bound mRNA" if the ratios from all the three pulldowns were higher than 1.5. The 3′-UTR motif enrichment search was done using MEME in the relative enrichment mode (Bailey 2011). Annotated 3′-UTRs were downloaded from TritrypDB. For some other 3′-UTRs, the manual annotation was performed using the RNA-seq reads (e.g., Jensen et al. 2014) and poly(A) site data in TritrypDB (Kolev et al. 2010). Analysis of differentially expressed genes after DRBD18 RNAi was done in R using DESeqUI (Leiss and Clayton 2016), a customized version of DESeq2 (Love et al. 2014) adapted for trypanosome transcriptomes. Statistical analyses were done using R and Microsoft excel.

To find alternative processing sites, aligned reads were visualized using the Integrated Genome Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2013). Terminal mismatches suggesting the presence of poly(A) tails or spliced leaders were identified. We then selected 17–20 nt sequences immediately upstream of the putative poly(A) sites or immediately downstream from the splice sites, to extract all overlapping reads from the raw data. This enabled us to find reads that had too many terminal mismatches to be present in the final aligned files. We used only the procyclic-form poly(A)$^+$ results for this; our bloodstream-form reads had almost no spliced leaders of poly(A) tails, perhaps due to the RNase H treatment. The lengths of sequences used for the extraction, and their distances from the processing sites, were not always the same, because highly repetitive sequences could not be used. Therefore, the proportion of overlapping reads extracted is expected to differ for different sites. The extracted reads were then edited manually to remove reads that come from the wrong gene, or had insufficient terminal bases to identify processing. Manual examination was particularly important for polyadenylation since often, several different alternative poly(A) sites were found within the region upstream of the new splice site. The numbers of sequences that spanned the processing site, or were processed, were then counted for all replicates.

RNA folding was predicted using the RNAFold Web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi).

Western blotting

Protein samples were collected from approximately 5×10^6 cells growing at logarithmic phase. Samples were run according to standard protein separation procedures using SDS-PAGE. The primary antibodies used in this study were: rabbit α-DRBD18 (1:2500) (Lott et al. 2015), rat α-ribosomal protein S9 (C Helbig, unpubl.) and rabbit anti-XRND (1:2500 [Li et al. 2006]). We used horseradish peroxidase coupled secondary antibodies (α-rat, 1:2000 and α-rabbit, 1:2500). The blots were developed using an enhanced chemiluminescence kit (Amersham) according to the manufacturer's instructions. The signal intensities of the images were quantified using ImageJ.

Protein–protein interactions

A cell line in which one allele of DRBD18 bears a sequence encoding a carboxy-terminal TAP tag was used to study the interactome of DRBD18. A cell line inducibly expressing GFP with a TAP tag at the amino terminus was used as a control. The expression of GFP-TAP was induced with 100 ng/mL tetracycline for 24 h, TAP-tagged protein was purified using IgG magnetic beads and the bound protein was eluted with TEV protease, which was then removed using HisPur Ni-NTA magnetic beads (Thermo Scientific) prior to mass spectrometry (Falk et al. 2021). Briefly, washed pellets from 2×10^9 cells were thawed on ice and resuspended in 0.5 mL of lysis buffer (20 mM Tris [pH 7.5], 5 mM $MgCl_2$, 1 mM DTT, 0.05% IGEPAL, 10 µg/mL aprotinin, and 10 µg/mL leupeptin). The cells were then lysed by passage 20× through a 21G × 1½″ needle and 20× through a 27G × ¾ needle. After centrifugation at 10,000*g* for 15 min, the salt concentration in the supernatant was adjusted to 150 mM KCl. We then added 100 µL of magnetic beads (Dynabeads M-280 Tosylactivated, Thermo Fisher Scientific) coupled to rabbit IgG, in wash buffer and incubated for 1–2 h at 4°C while rotating (20 rpm). After four washes, 20 µL of wash buffer and 1 µL of recombinant TEV protease (1 mg/mL) were incubated with the beads for 90 min at 20°C. IgG magnetic beads were then concentrated on one side, the supernatant containing eluted proteins was transferred to a fresh tube. A total of 10 µL of equalization buffer (200 mM sodium phosphate, 600 mM sodium chloride, 0.1% Tween-20, 60 mM imidazole, pH 8.5) were added, followed by 30 µL of Ni-NTA-magnetic beads. The mixture was incubated for 30 min at 20°C while rotating. Ni-NTA magnetic beads were then retained and the supernatant was collected and stored in Laemmli buffer at −80°C.

The proteins from three independent purifications were separated on a 1.5 mm NuPAGE Novex 4%–12% Bis-Tris protein gel (Thermo Fisher Scientific) until the running front had migrated roughly 2 cm, after which the gel was stained with Coomassie blue. Three areas per lane were cut and analyzed by Nanoflow LC-MS2 analysis with an Ultimate 3000 liquid chromatography system directly coupled to an Orbitrap Elite mass spectrometer (both Thermo Fisher). MS spectra (m/z 400–1600) were acquired in the Orbitrap at 60,000 (m/z 400) resolution. Fragmentation in the CID cell was performed for up to 10 precursors. MS2 spectra were acquired at a rapid scan rate. Raw files were processed using MaxQuant (version 1.5.3.30) (Tyanova et al. 2015) for peptide identification and quantification. MS2 spectra were searched against the TriTrypDB-8.1TREU927-AnnotatedProteins-1 database (containing 11567 sequences).

Data were analyzed quantitatively and plotted using Perseus (version 1.6.6.0) (Tyanova et al. 2016). The statistical analysis is described in more detail in Keilhauer et al. (2015). Briefly. to assess

significant enrichment, the program uses is a double sided two sample *t*-test with permutation-based FDR of 0.01. The parameter S_0 (minimal fold change) set to 0.1. The curved line is calculated from the two extreme values of S_0 at very high *P*-value, and *P* = (−10log 0.01) for large fold change. It takes into account that for small fold changes you should have a higher *P*-value whereas for large fold changes a FDR (false discovery rate) of 1%, *P* = 0.01, is sufficient for significance.

Subcellular fractionation

Nuclear-cytosolic fractionation was performed as described in Biton et al. (2006). Briefly, 3×10^8 Lister 427 bloodstream-form wild-type cells, or those expressing the RNAi stem–loop vector targeting DRBD18 were harvested by centrifugation at 3000 rpm for 13 min at 4°C. The cell pellet was washed twice with cold 1× PBS and collected by centrifugation at 2300 rpm for 8 min at 4°C. Afterwards, the pellet was resuspended in 500 µL hypotonic buffer (10 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM dithiothreitol, 5 µg/mL leupeptin and 100 U RNasin) and lysed in the presence of 0.1% IGEPAL (Nonidet P-40) by passing 20 times through a 21G × ½ needle and 20 times through a 27G × ¾ needle. The cells were allowed to rest on ice for 20 min. Then, the nuclear fraction was pelleted by centrifugation at 15,000*g* for 15 min at 4°C. The supernatant containing the cytoplasmic fraction was transferred to a new tube. The pellet containing the nuclear fraction was resuspended in 500 µL of the same buffer containing 0.1% SDS. For western blotting, 20 µL of extract were used. The remainder was used for RNA preparation (not shown because only the shortest *RBP10* mRNAs were detectable).

Affymetrix smRNA FISH

The Affymetrix single molecule mRNA FISH was carried out as described in Kramer (2017). A total of 200 mL bloodstream-form trypanosomes at ~5–8 × 10^5 cells/mL were harvested by centrifugation (8 min, 1400*g*), resuspended in 1 mL PBS and pelleted again by centrifugation (5 min, 1400*g*). The cell pellet was resuspended in 1 mL PBS, followed by fixation with 1 mL of 8% formaldehyde (in PBS). The mixture was incubated at room temperature for 10 min with an orbital mixer. A total of 13 mL PBS were added and the cells were harvested by centrifugation (5 min, 1400*g*). The pellet of fixed cells was resuspended in 1 mL PBS and spread on glass microscopy slides (previously incubated at 180°C for 2 h for RNase removal) within circles of hydrophobic barriers (PAP pen, Sigma). The cells were allowed to settle at room temperature for 20 min. The slides were then washed twice in PBS. The protease solution was diluted 1:1000 in PBS and briefly vortexed to allow complete dissolution. Permeabilization of the fixed cells was done with addition of 50 µL of detergent solution QC in each circle on the slides. This was followed by a two-step washing in PBS. A total of 100 µL of the protease solution was added to each circle and incubated exactly for 10 min at 25°C. The slides were then washed twice in PBS and used for Affymetrix FISH experiments as described in the manual of the QuantiGene ViewRNA ISH Cell Assay (Affymetrix), protocol for glass slide format. The only modification from the kit protocol is that the protease digestion was done at 25°C rather than the normal room temperature, and we used a self-made washing buffer (0.1 × SSC, 0.1% SDS) instead of the washing buffer from the kit. All Affymetrix probe sets used in this work are described in Supplemental Table S7 (sheet 3). For visualization, the labeled cells were mounted with 4′,6-diamino-2-phenylindole dihydrochloride (DAPI) solution, diluted 1:1000 in PBS. Images were taken with a fluorescent inverted wide-field microscope Leica DMI6000B (Leica Microsystems GmbH) equipped with 100× oil immersion (NA 1.4) and a Leica DFC365 camera (6.45 m/pixel). Deconvolution was done using Huygens Essential software (SVI) and images are presented as *Z*-stack projection (sum slices). The image analysis was carried out using the available tools in Image J software, and 50 cells for each slide were selected for quantifying the number of mRNAs present in the cytosol and in the nucleus.

DATA DEPOSITION

Raw RNA-seq data are available at Annotare as E-MTAB-9783 and E-MTAB-10735. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al. 2019) with the data set identifier PXD027792.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We are very grateful to Professor Dr. Susanne Kramer, who supervised and provided facilities for the mRNA FISH experiment. We are thankful to Professor Laurie Read for anti-DRBD18 antibody donation and for extensive and friendly discussions, including communicating unpublished results. We thank Bernardo Gabiatti and Paula Andrea Castañeda Londoño (Biozentrum, Universität Würzburg) for discussion and assistance with the mRNA FISH experiment; Laura Armbruster (COS, University of Heidelberg) for help in mass spectrometry data analysis, and Kevin Leiss, Abeer Fadda, and Simon Anders for supplementing CC's rudimentary bioinformatics skills. All RNA-seq libraries and RNA sequencing were done by David Ibberson at the Bioquant in the University of Heidelberg. Mass spectrometry was done in the ZMBH Core Facility for Mass Spectrometry by Drs. Thomas Ruppert and Sabine Merker; we particularly acknowledge Thomas Ruppert for interrupting his holiday to provide a simple explanation of the statistical analysis. We thank Claudia Helbig and Ute Leibfried for technical assistance and for preparing media and buffers. We are indebted to Professor Dr. Nina Papavasiliou (DKFZ, University of Heidelberg) and Professor Dr. Luise Krauth-Siegel (BZH, University of Heidelberg) for allowing us to share their laboratories, equipment and reagents after a flood in ZMBH. This work was partially supported by the Deutsche Forschungsgemeinschaft, grant number CI112/28-1 to C.C., and partially by core funding.

Author contributions: T.B.T. was responsible for nearly all of the experimental work, provided figures and tables, wrote the first draft of the paper and was involved in subsequent editing. C.C.

devised and supervised the project, analyzed the procyclic-form RNA-seq data, edited the paper and provided funding.

NOTE ADDED IN PROOF

Since this paper was accepted, Obado et al. (2022) have demonstrated that the protein encoded by Tb927.11.2340, which was strongly and specifically pulled down with DRBD18, is a functional MEX67 paralog.

Received May 17, 2022; accepted June 23, 2022.

REFERENCES

Alibu VP, Storm L, Haile S, Clayton C, Horn D. 2004. A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. *Mol Biochem Parasitol* **139:** 75–82. doi:10.1016/j.molbiopara.2004.10.002

Antwi E, Haanstra J, Ramasamy G, Jensen B, Droll D, Rojas F, Minia I, Terrao M, Mercé C, Matthews K, et al. 2016. Integrative analysis of the *Trypanosoma brucei* gene expression cascade predicts differential regulation of mRNA processing and unusual control of ribosomal protein expression. *BMC Genomics* **17:** 306. doi:10.1186/s12864-016-2624-3

Ashkenazy-Titelman A, Shav-Tal Y, Kehlenbach RH. 2020. Into the basket and beyond: the journey of mRNA through the nuclear pore complex. *Biochem J* **477:** 23–44. doi:10.1042/BCJ20190132

Bailey T. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27:** 1653–1659. doi:10.1093/bioinformatics/btr261

Bailey T, Johnson J, Grant C, Noble W. 2015. The MEME suite. *Nucleic Acids Res* **43:** W39–W49. doi:10.1093/nar/gkv416

Begolo D, Vincent I, Giordani F, Pöhner I, Witty M, Rowan T, Bengaly Z, Gillingwater K, Freund Y, Wade R, et al. 2018. The trypanocidal benzoxaborole AN7973 inhibits trypanosome mRNA processing. *PLoS Pathog* **14:** e1007315. doi:10.1371/journal.ppat.1007315

Bercovich N, Clayton C, Levin M, Vazquez M. 2009. Identification of core components of the exon junction complex in trypanosomes. *Mol Biochem Parasitol* **166:** 190–193. doi:10.1016/j.molbiopara.2009.03.008

Bishola Tshitenge T, Clayton C. 2021. Interactions of the *Trypanosoma brucei brucei* zinc-finger-domain protein ZC3H28. *Parasitology* **149:** 356–370. doi:10.1017/S003118202100189X

Bishola Tshitenge T, Reichert L, Liu B, Clayton C. 2022. Several different sequences are implicated in bloodstream-form-specific gene expression in *Trypanosoma brucei*. *PLoS Negl Trop Dis* **16:** e0010030. doi:10.1371/journal.pntd.0010030

Biton M, Mandelboim M, Arvatz G, Michaeli S. 2006. RNAi interference of XPO1 and Sm genes and their effect on the spliced leader RNA in *Trypanosoma brucei*. *Mol Biochem Parasitol* **150:** 132–143. doi:10.1016/j.molbiopara.2006.07.004

Chakraborty C, Clayton C. 2018. Stress susceptibility in *Trypanosoma brucei* lacking the RNA-binding protein ZC3H30. *PLoS Negl Trop Dis* **12:** e0006835. doi:10.1371/journal.pntd.0006835

Christiano R, Kolev N, Shi H, Ullu E, Walther T, Tschudi C. 2017. The proteome and transcriptome of the infectious metacyclic form of *Trypanosoma brucei* define quiescent cells primed for mammalian invasion. *Mol Microbiol* **106:** 74–92. doi:10.1111/mmi.13754

Clayton C. 2019. Control of gene expression in trypanosomatids: living with polycistronic transcription. *R Soc Open Biol* **9:** 190072. doi:10.1098/rsob.190072

Clayton C, Michaeli S. 2011. 3′ processing in protists. *Wiley Interdiscip Rev RNA* **2:** 247–255. doi:10.1002/wrna.49

Clayton CE, Estévez AM, Hartmann C, Alibu VP, Field M, Horn D. 2005. Down-regulating gene expression by RNA interference in *Trypanosoma brucei*. In *RNA interference* (ed. Carmichael G). Humana Press, NJ.

Das A, Bellofatto V, Rosenfeld J, Carrington M, Romero-Zaliz R, del Val C, Estevez AM. 2015. High throughput sequencing analysis of *Trypanosoma brucei* DRBD3/PTB1-bound mRNAs. *Mol Biochem Parasitol* **199:** 1–4. doi:10.1016/j.molbiopara.2015.02.003

Dean S, Sunter J, Wheeler R. 2016. TrypTag.org: a trypanosome genome-wide protein localisation resource. *Trends Parasitol* **33:** 80–82. doi:10.1016/j.pt.2016.10.009

Degrasse JA, Dubois KN, Devos D, Siegel TN, Sali A, Field MC, Rout MP, Chait BT. 2009. Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol Cell Proteomics* **8:** 2119–2130. doi:10.1074/mcp.M900038-MCP200

Dhalia R, Marinsek N, Reis C, Katz R, Muniz J, Standart N, Carrington M, de Melo Neto O. 2006. The two eIF4A helicases in Trypanosoma brucei are functionally distinct. *Nucleic Acids Res* **34:** 2495–2507. doi:10.1093/nar/gkl290

Dominguez D, Freese P, Alexis MS, Su A, Hochman M, Palden T, Bazile C, Lambert NJ, Van Nostrand EL, Pratt GA, et al. 2018. Sequence, structure, and context preferences of human RNA binding proteins. *Mol Cell* **70:** 854–867.e859. doi:10.1016/j.molcel.2018.05.001

D'Orso I, Frasch ACC. 2001. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. *J Biol Chem* **276:** 34801–34809. doi:10.1074/jbc.M102120200

Dostalova A, Käser S, Cristodero M, Schimanski B. 2013. The nuclear mRNA export receptor Mex67-Mtr2 of *Trypanosoma brucei* contains a unique and essential zinc finger motif. *Mol Microbiol* **88:** 728–739. doi:10.1111/mmi.12217

Erben E, Leiss K, Liu B, Inchaustegui Gil D, Helbig C, Clayton C. 2021. Insights into the functions and RNA binding of *Trypanosoma brucei* ZC3H22, RBP9 and DRBD7. *Parasitology* **148:** 1186–1195. doi:10.1017/S0031182021000123

Fadda A, Ryten M, Droll D, Rojas F, Färber V, Haanstra J, Bakker B, Matthews K, Clayton C. 2014. Transcriptome-wide analysis of mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol Microbiol* **94:** 307–326. doi:10.1111/mmi.12764

Falk F, Kamanyi Marucha K, Clayton C. 2021. The EIF4E1-4EIP cap-binding complex of *Trypanosoma brucei* interacts with the terminal uridylyl transferase TUT3. *PLoS One* **16:** e0258903. doi:10.1371/journal.pone.0258903

Fernandez-Moya SM, Carrington M, Estevez AM. 2014. Depletion of the RNA-binding protein RBP33 results in increased expression of silenced RNA polymerase II transcripts in *Trypanosoma brucei*. *PLoS One* **9:** e107608. doi:10.1371/journal.pone.0107608

Goos C, Dejung M, Wehman A, M-Natus E, Schmidt J, Sunter J, Engstler M, Butter F, Kramer S. 2018. Trypanosomes can initiate nuclear export co-transcriptionally. *Nucleic Acids Res* **47:** 266–282. doi:10.1093/nar/gky1136

Gray AR. 1965. Antigenic variation in a strain of *Trypanosoma brucei* transmitted by *Glossina morsitans* and *G. palpalis*. *J Gen Microbiol* **41:** 195–214. doi:10.1099/00221287-41-2-195

Gupta SK, Carmi S, Waldman Ben-Asher H, Tkacz ID, Naboishchikov I, Michaeli S. 2013a. Basal splicing factors regulate the stability of mature mRNAs in trypanosomes. *J Biol Chem* **288:** 4991–5006. doi:10.1074/jbc.M112.416578

Gupta SK, Kosti I, Plaut G, Pivko A, Tkacz ID, Cohen-Chalamish S, Biswas DK, Wachtel C, Waldman Ben-Asher H, Carmi S, et al. 2013b. 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. doi:10.1093/nar/gkt369

Gupta SK, Chikne V, Eliaz D, Tkacz ID, Naboishchikov I, Carmi S, Waldman Ben-Asher H, Michaeli S. 2014. 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. doi:10.4161/rna.29143

Hartmann C, Hotz H-R, McAndrew M, Clayton C. 1998. Effect of multiple downstream splice sites on polyadenylation in *Trypanosoma brucei*. *Mol Biochem Parasit* **93:** 149–152. doi:10.1016/S0166-6851(98)00026-7

Hartmann C, Benz C, Brems S, Ellis L, Luu V-D, Stewart M, D'Orso I, Busold C, Fellenberg K, Frasch ACC, et al. 2007. The small trypanosome RNA-binding proteins *Tb*UBP1 and *Tb*UBP2 influence expression of F box protein mRNAs in bloodstream trypanosomes. *Eukaryot Cell* **6:** 1964–1978. doi:10.1128/EC.00279-07

Hendriks EF, Robinson DR, Hinkins M, Matthews KR. 2001. A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *EMBO J* **20:** 6700–6711. doi:10.1093/emboj/20.23.6700

Hendriks EF, Abdul-Razak A, Matthews KR. 2003. *Tb*CPSF30 depletion by RNA interference disrupts polycistronic RNA processing in *Trypanosoma brucei*. *J Biol Chem* **278:** 26870–26878. doi:10.1074/jbc.M302405200

Howard JM, Sanford JR. 2015. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip Rev RNA* **6:** 93–110. doi:10.1002/wrna.1260

Huang J, van der Ploeg LHT. 1991. Requirement of a poly-pyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3′ splice acceptor site. *EMBO J* **10:** 3877–3885. doi:10.1002/j.1460-2075.1991.tb04957.x

Hug M, Hotz HR, Hartmann C, Clayton CE. 1994. Hierarchies of RNA processing signals in a trypanosome surface antigen mRNA precursor. *Mol Cell Biol* **14:** 7428–7435. doi:10.1128/mcb.14.11.7428-7435.1994

Inoue AH, Serpeloni M, Hiraiwa PM, Yamada-Ogatta SF, Muniz JR, Motta MC, Vidal NM, Goldenberg S, Avila AR. 2014. Identification of a novel nucleocytoplasmic shuttling RNA helicase of trypanosomes. *PLoS One* **9:** e109521. doi:10.1371/journal.pone.0109521

Jensen BC, Ramasamy G, Vasconcelos EJ, Ingolia NT, Myler PJ, Parsons M. 2014. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics* **15:** 911. doi:10.1186/1471-2164-15-911

Kapotas N, Bellofatto V. 1993. Differential response to RNA *trans*-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*. *Nucleic Acids Res* **21:** 4067–4072. doi:10.1093/nar/21.17.4067

Keilhauer EC, Hein MY, Mann M. 2015. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol Cell Proteomics* **14:** 120–135. doi:10.1074/mcp.M114.041012

Koch H, Raabe M, Urlaub H, Bindereif A, Preusser C. 2016. The polyadenylation complex of *Trypanosoma brucei*: characterization of the functional poly(A) polymerase. *RNA Biol* **13:** 221–231. doi:10.1080/15476286.2015.1130208

Kolev N, Franklin J, Carmi S, Shi H, Michaeli S, Tschudi C. 2010. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog* **6:** e1001090. doi:10.1371/journal.ppat.1001090

Kramer S. 2017. Simultaneous detection of mRNA transcription and decay intermediates by dual colour single mRNA FISH on subcellular resolution. *Nucleic Acids Res* **45:** e49. doi:10.1093/nar/gkw1245

Kramer S. 2021. Nuclear mRNA maturation and mRNA export control: from trypanosomes to opisthokonts. *Parasitology* **148:** 1196–1218. doi:10.1017/S0031182021000068

Kramer S, Piper S, Estevez A, Carrington M. 2016. Polycistronic trypanosome mRNAs are a target for the exosome. *Mol Biochem Parasitol* **205:** 1–5. doi:10.1016/j.molbiopara.2016.02.009

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Leiss K, Clayton C. 2016. DESeqUI—Trypanosome RNAseq analysis made easy. *Zenodo* doi:10.5281/zenodo.165132

Leiss K, Merce C, Muchunga E, Clayton C. 2016. TrypRNAseq: a easy to use pipeline for *Trypanosoma* RNAseq data. *Zenodo* doi:10.5281/zenodo.158920

Li C-H, Irmer H, Gudjonsdottir-Planck D, Freese S, Salm H, Haile S, Estévez AM, Clayton CE. 2006. Roles of a *Trypanosoma brucei* 5′→3′ exoribonuclease homologue in mRNA degradation. *RNA (New York, NY)* **12:** 2171–2186. doi:10.1261/rna.291506

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Lopez-Estrano C, Tschudi C, Ullu E. 1998. Exonic sequences in the 5′ untranslated region of α-tubulin mRNA modulate *trans* splicing in *Trypanosoma brucei*. *Mol Cell Biol* **18:** 4620–4628. doi:10.1128/MCB.18.8.4620

Lott K, Mukhopadhyay S, Li J, Wang J, Yao J, Sun Y, Qu J, Read LK. 2015. Arginine methylation of DRBD18 differentially impacts its opposing effects on the trypanosome transcriptome. *Nucleic Acids Res* **43:** 5501–5523. doi:10.1093/nar/gkv428

Love M, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Martin M. 2011. Next generation sequencing data analysis: cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17:** 1. doi:10.14806/ej.17.1.200

Matthews KR, Tschudi C, Ullu E. 1994. A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev* **8:** 491–501. doi:10.1101/gad.8.4.491

McNally KP, Agabian N. 1992. *Trypanosoma brucei* spliced-leader RNA methylations are required for *trans* splicing in vivo. *Mol Cell Biol* **12:** 4844–4851. doi:10.1128/mcb.12.11.4844-4851.1992

Michaeli S. 2011. *Trans*-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol* **6:** 459–474. doi:10.2217/fmb.11.20

Minia I, Merce C, Terrao M, Clayton C. 2016. Translation regulation and RNA granule formation after heat shock of procyclic form *Trypanosoma brucei*: many heat-induced mRNAs are increased during differentiation to mammalian-infective forms. *PLoS Negl Trop Dis* **10:** e0004982. doi:10.1371/journal.pntd.0004982

Mishra A, Kaur J, McSkimming D, Hegedűsová E, Dubey A, Ciganda M, Aris Z, Read L. 2021. Selective nuclear export of mRNAs is promoted by DRBD18 in *Trypanosoma brucei*. *Mol Microbiol* **116:** 827–840. doi:10.1111/mmi.14773

Mugo E, Clayton C. 2017. Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*. *PLoS Pathog* **13:** e1006560. doi:10.1371/journal.ppat.1006560

Mulindwa J, Ssentamu G, Matovu E, Kamanyi Marucha K, Aresta-Branco F, Helbig C, Clayton C. 2021. *In vitro* culture of freshly isolated *Trypanosoma brucei brucei* bloodstream forms results in gene copy-number changes. *PLoS Negl Trop Dis* **15:** e0009738. doi:10.1371/journal.pntd.0009738

Naguleswaran A, Doiron N, Roditi I. 2018. RNA-Seq analysis validates the use of culture-derived *Trypanosoma brucei* and provides new

markers for mammalian and insect life-cycle stages. *BMC Genomics* **19:** 227. doi:10.1186/s12864-018-4600-6

Nascimento L, Terrao M, Marucha K, Liu B, Egler F, Helbig C, Clayton C. 2020. The RNA-associated proteins MKT1 and MKT1L form alternative PBP1-containing complexes in *Trypanosoma brucei*. *J Biol Chem* **295:** 10940–10955. doi:10.1074/jbc.RA120.013306

Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T. 2010. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog* **6:** e1001037. doi:10.1371/journal.ppat.1001037

Obado SO, Brillantes M, Uryu K, Zhang W, Ketaren NE, Chait BT, Field MC, Rout MP. 2016. Interactome mapping reveals the evolutionary history of the nuclear pore complex. *PLoS Biol* **14:** e1002365. doi:10.1371/journal.pbio.1002365

Obado SO, Field MC, Rout MP. 2017. Comparative interactomics provides evidence for functional specialization of the nuclear pore complex. *Nucleus* **8:** 340–352. doi:10.1080/19491034.2017.1313936

Obado SO, Stein M, Hegedűsová E, Zhang W, Hutchinson S, Brillantes M, Glover L, Paris Z, Chait BT, Field MC, et al. 2022. Mex67 paralogs mediate division of labor in trypanosome RNA processing and export. bioRxiv doi:10.1101/2022.06.27.497849

Patzelt E, Perry KL, Agabian N. 1989. Mapping of branch sites in *trans*-spliced pre-mRNAs of *Trypanosoma brucei*. *Mol Cell Biol* **9:** 4291–4297. doi:10.1128/mcb.9.10.4291-4297.1989

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, et al. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47:** D442–d450. doi:10.1093/nar/gky1106

Preusser C, Jae N, Bindereif A. 2012. mRNA splicing in trypanosomes. *Int J Med Microbiol* **302:** 221–224. doi:10.1016/j.ijmm.2012.07.004

Quintana JF, Zoltner M, Field MC. 2021. Evolving differentiation in African trypanosomes. *Trends Parasitol* **37:** 296–303. doi:10.1016/j.pt.2020.11.003

Redmond S, Vadivelu J, Field MC. 2003. RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*. *Mol Biochem Parasitol* **128:** 115–118. doi:10.1016/S0166-6851(03)00045-8

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26. doi:10.1038/nbt.1754

Savage AF, Kolev NG, Franklin JB, Vigneron A, Aksoy S, Tschudi C. 2016. Transcriptome profiling of *Trypanosoma brucei* development in the Tsetse fly vector *Glossina morsitans*. *PLoS One* **11:** e0168877. doi:10.1371/journal.pone.0168877

Schürch N, Hehl A, Vassella E, Braun R, Roditi I. 1994. Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. *Mol Cell Biol* **14:** 3668–3675. doi:10.1128/mcb.14.6.3668-3675.1994

Schwede A, Manful T, Jha B, Helbig C, Bercovich N, Stewart M, Clayton C. 2009. The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. *Nucleic Acids Res* **37:** 5511–5528. doi:10.1093/nar/gkp571

Sharma D, Zagore LL, Brister MM, Ye X, Crespo-Hernández CE, Licatalosi DD, Jankowsky E. 2021. The kinetic landscape of an RNA-binding protein in cells. *Nature* **591:** 152–156. doi:10.1038/s41586-021-03222-x

Shaw AP, Cecchi G, Wint GR, Mattioli RC, Robinson TP. 2014. Mapping the economic benefits to livestock keepers from intervening against bovine trypanosomosis in Eastern Africa. *Prev Vet Med* **113:** 197–210. doi:10.1016/j.prevetmed.2013.10.024

Shi H, Butler K, Tschudi C. 2018. Differential expression analysis of transcriptome data of *Trypanosoma brucei* RBP6 induction in procyclics leading to infectious metacyclics and bloodstream forms *in vitro*. *Data Brief* **20:** 978–980. doi:10.1016/j.dib.2018.08.169

Siegel T, Tan K, Cross G. 2005. Systematic study of sequence motifs for RNA *trans* splicing in *Trypanosoma brucei*. *Mol Cell Biol* **25:** 9586–9594. doi:10.1128/MCB.25.21.9586-9594.2005

Siegel T, Hekstra D, Kemp L, Figueiredo L, Lowell J, Fenyo D, Wang X, Dewell S, Cross G. 2009. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* **23:** 1063–1076. doi:10.1101/gad.1790409

Siegel T, Hekstra D, Wang X, Dewell S, Cross G. 2010. 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. doi:10.1093/nar/gkq237

Silvester E, Ivens A, Matthews KR. 2018. A gene expression comparison of *Trypanosoma brucei* and *Trypanosoma congolense* in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. *PLoS Negl Trop Dis* **12:** e0006863. doi:10.1371/journal.pntd.0006863

Singh A, Minia I, Droll D, Fadda A, Clayton C, Erben E. 2014. Trypanosome MKT1 and the RNA-binding protein ZC3H11: interactions and potential roles in post-transcriptional regulatory networks. *Nucleic Acids Res* **42:** 4652–4668. doi:10.1093/nar/gkt1416

Stern M, Gupta S, Salmon-Divon M, Haham T, Barda O, Levi S, Wachtel C, Nilsen T, Michaeli S. 2009. Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA (New York, NY)* **15:** 648–665. doi:10.1261/rna.1230209

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192. doi:10.1093/bib/bbs017

Turowski TW, Petfalski E, Goddard BD, French SL, Helwak A, Tollervey D. 2020. Nascent transcript folding plays a major role in determining RNA polymerase elongation rates. *Mol Cell* **79:** 488–503.e411. doi:10.1016/j.molcel.2020.06.002

Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M, Cox J. 2015. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15:** 1453–1456. doi:10.1002/pmic.201400449

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13:** 731–740. doi:10.1038/nmeth.3901

Ule J, Blencowe BJ. 2019. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol Cell* **76:** 329–345. doi:10.1016/j.molcel.2019.09.017

Ullu E, Tschudi C. 1991. *Trans* splicing in trypanosomes requires methylation of the 5′ end of the spliced leader RNA. *Proc Natl Acad Sci* **88:** 10074–10078. doi:10.1073/pnas.88.22.10074

Ullu E, Matthews KR, Tschudi C. 1993. Temporal order of RNA-processing reactions in trypanosomes: rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol* **13:** 720–725. doi:10.1128/mcb.13.1.720-725.1993

Vasquez JJ, Hon CC, Vanselow JT, Schlosser A, Siegel TN. 2014. Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res* **42:** 3623–3637. doi:10.1093/nar/gkt1386

Vassella E, Braun R, Roditi I. 1994. Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site: a dual role for polypyrimidine tracts in trypanosomes? *Nucleic Acids Res* **22:** 1359–1364. doi:10.1093/nar/22.8.1359

Vigneron A, O'Neill MB, Weiss BL, Savage AF, Campbell OC, Kamhawi S, Valenzuela JG, Aksoy S. 2020. Single-cell RNA

sequencing of *Trypanosoma brucei* from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens. *Proc Natl Acad Sci* **117:** 2613–2621. doi:10.1073/pnas.1914423117

Waithaka A, Maiakovska O, Grimm D, Melo do Nascimento L, Clayton C. 2022. Sequences and proteins that influence mRNA processing in *Trypanosoma brucei*: evolutionary conservation of SR-domain and PTB protein functions. *bioRxiv* doi:10.1101/2022.04.25.489340

Wall RJ, Rico E, Lukac I, Zuccotto F, Elg S, Gilbert IH, Freund Y, Alley MRK, Field MC, Wyllie S, et al. 2018. Clinical and veterinary trypanocidal benzoxaboroles target CPSF3. *Proc Natl Acad Sci* **115:** 9616–9621. doi:10.1073/pnas.1807915115

Wurst M, Selinger B, Jha B, Klein C, Queiroz R, Clayton C. 2012. Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. *Mol Microbiol* **83:** 1048–1063. doi:10.1111/j.1365-2958.2012.07988.x

MEET THE FIRST AUTHOR

Tania Bishola Tshitenge

***Meet the First Author(s)* is a new editorial feature within *RNA*, in which the first author(s) of research-based papers in each issue have the opportunity to introduce themselves and their work to readers of *RNA* and the RNA research community. Tania Bishola Tshitenge is the first author of this paper, "The *Trypanosoma brucei* RNA-binding protein DRBD18 ensures correct mRNA *trans* splicing and polyadenylation patterns." Tania did her BSc in molecular biology at the University of Kinshasa, D.R. Congo, and an MSc in biotechnology in the Center of Biotechnology and Bioinformatic at the University of Nairobi (Kenya). Subsequently, Tania did her PhD and a brief post-doc in Christine Clayton's laboratory at the Heidelberg University Center for Molecular Biology, investigating the mechanisms of post-transcriptional regulation of gene expression using trypanosomes as a model system.**

What are the major results described in your paper and how do they impact this branch of the field?

In trypanosomes, Pol II transcription is polycistronic, and polyadenylation of each mRNA is determined by *trans*-splicing of the mRNA that is immediately downstream. *Trans*-splicing signals include polypyrimidine tracts. However, 3′-UTRs include many polypyrimidine tracts and it was previously unknown why these are not recognized by the splicing machinery. DRBD18 is a 2-RRM RNA-binding protein which had previously been shown to be needed for efficient export of some mRNAs. I showed that depletion of DRBD18 affects processing of more than 200 mRNAs, including the *RBP10* mRNA, which has an 8.5 kb 3′-UTR. DRBD18 depletion caused accumulation of versions with shortened 3′-untranslated regions, which were often, but not always, more abundant in depleted cells than the corresponding longer versions in normal cells. Our cumulative results suggested that in the nucleus, selective binding of DRBD18 to the 3′-UTRs of some mRNA precursors actively prevents recognition of 3′-UTR-internal polypyrimidine tracts by splicing factors, thus ensuring correct 3′-UTR lengths. At the same time, the DRBD18 that is bound to the mRNA promotes its export to the cytosol, perhaps particularly important for long 3′-UTRs.

This is the first time that a trypanosome RNA-binding protein has been implicated in determining *trans*-splicing and polyadenylation patterns of specific mRNAs.

What led you to study RNA or this aspect of RNA science?

As a child, I was always very curious to understand the principles of life. This led me to study biology at the University of Kinshasa. During my bachelor studies, I became very fascinated about the central dogma of molecular biology, and I wanted to understand the mechanisms that regulate gene expression. During my master studies, I did a three-month summer internship in 2016 in the Clayton laboratory, and I was captivated by the trypanosome polycistronic mRNAs and the unique way in which their gene expression regulation mainly relies upon RNA-binding proteins. Therefore, I decided to focus on the role of RBPs in regulating gene expression and understand the RNA–protein interactions in a deeper manner.

During the course of these experiments, were there any surprising results or particular difficulties that altered your thinking and subsequent focus?

There were many surprising results in the course of these experiments. I was not expecting that DRBD18 might affect the processing of *RBP10* mRNA; instead, I thought that it might be stabilizing *RBP10* mRNA. Funnily, my supervisor's old-fashioned insistence on northern blots was essential; I would never have seen the altered polyadenylation with a qPCR! This first surprising result changed my focus from mRNA stability toward mRNA processing. Secondly, as DRBD18 affects processing, I expected to find splicing factors or the polyadenylation complex associating with DRBD18, but again, this was not so; DRBD18 was only associated with proteins of the outer ring of the nuclear pore and nuclear export factors. That suggested that DRBD18 was preventing recognition by the processing machinery rather than actively promoting use of downstream sites.

Continued

If you were able to give one piece of advice to your younger self, what would that be?

I would tell my younger self that she could have been more confident in herself and not stress a lot about experiments that did not work, because that was part of the training and will help develop the resilient spirit she will need for her next career step.

Are there specific individuals or groups who have influenced your philosophy or approach to science?

Yes, there are. First is my PhD supervisor, Christine Clayton, who taught me to trust my instinct, to interact a lot with my results and consider all hypotheses available. Christine taught me that the hypothesis remains a hypothesis until we have evidence based on experiments and the results we get. She has influenced my philosophy a lot. The second person is my fiancé, Dieudonné, who taught me to consider all aspects of science when dealing with interpretation of the results, and to express my thoughts clearly so that I am easily understood.

What are your subsequent near- or long-term career plans?

I work now in Bayer Pharmaceuticals as a cell biologist and next generation sequencing expert, developing chemical transcriptomics to understand the drug mode of action. I also plan to be a visiting lecturer at my home university, the University of Kinshasa, teaching techniques of molecular biology as well as genetic engineering, because I would like to pass on the knowledge that I have gained to the younger generation.