

Interactions of the *Trypanosoma brucei brucei* zinc-finger-domain protein ZC3H28Tania Bishola Tshitenge and Christine Clayton 

Heidelberg University Centre for Molecular Biology (ZMBH), Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany

Research Article

Cite this article: Bishola Tshitenge T, Clayton C (2022). Interactions of the *Trypanosoma brucei brucei* zinc-finger-domain protein ZC3H28. *Parasitology* **149**, 356–370. <https://doi.org/10.1017/S003118202100189X>

Received: 9 August 2021
Revised: 20 October 2021
Accepted: 20 October 2021
First published online: 2 November 2021

Keywords:

mRNA decay; RNA-binding protein; translation; *Trypanosoma brucei*

Author for correspondence:

Christine Clayton,
E-mail: cclayton@zmbh.uni-heidelberg.de

Abstract

In *Trypanosoma brucei* and related Kinetoplastids, regulation of gene expression occurs mostly post-transcriptionally, and RNA-binding proteins play a critical role in the regulation of mRNA and protein abundance. *Trypanosoma brucei* ZC3H28 is a 114 kDa cytoplasmic mRNA-binding protein with a single C(x)₇C(x)₅C(x)₅H zinc finger at the C-terminus and numerous proline-, histidine- or glutamine-rich regions. ZC3H28 is essential for normal bloodstream-form trypanosome growth, and when tethered to a reporter mRNA, ZC3H28 increased reporter mRNA and protein levels. Purification of N-terminally tagged ZC3H28 followed by mass spectrometry showed enrichment of ribosomal proteins, various RNA-binding proteins including both poly(A) binding proteins, the translation initiation complex EIF4E4/EIF4G3, and the activator MKT1. Tagged ZC3H28 was preferentially associated with long RNAs that have low complexity sequences in their 3'-untranslated regions; their coding regions also have low ribosome densities. In agreement with the tethering results, after ZC3H28 depletion, the levels of a significant proportion of its bound mRNAs decreased. We suggest that ZC3H28 is implicated in the stabilization of long mRNAs that are poorly translated.

Introduction

Trypanosoma brucei and related Kinetoplastid parasites rely heavily on post-transcriptional mechanisms for the control of gene expression. Control is required not just to determine steady-state expression, but also to respond to external stresses and to adapt to different environments in infected mammals and in the definitive host, the Tsetse fly (see e.g. Silvester *et al.*, 2018; Doleželová *et al.*, 2020; Vigneron *et al.*, 2020; Toh *et al.*, 2021). Most transcription in Kinetoplastids is polycistronic: mRNA levels are therefore determined mainly by gene copy numbers, the efficiency of mRNA processing (via *trans* splicing and polyadenylation) and the rate of mRNA decay, and rates of protein synthesis also vary considerably [reviewed in Clayton (2019)].

RNA-binding proteins are critical at all stages of Kinetoplastid gene expression control [reviewed in Clayton (2019)]. In this paper, we studied the role and interactions of ZC3H28 (Tb927.9.9450), a 114 kDa (1030-residue) protein with a single C(x)₇C(x)₅C(x)₅H (CCCH) (CCCH) zinc finger at the C-terminus. This motif is known to bind to mRNA (Hudson *et al.*, 2004). The *Trypanosoma brucei* genome encodes 48 CCCH-domain proteins (Kramer *et al.*, 2010), with various roles in mRNA metabolism (Hendriks and Matthews, 2005; Paterou *et al.*, 2006; Walrad *et al.*, 2009; Benz *et al.*, 2011; Ling *et al.*, 2011; Morking *et al.*, 2012; Ouna *et al.*, 2012; Klein *et al.*, 2017; Chakraborty and Clayton, 2018; Bajak *et al.*, 2020; Liu *et al.*, 2020; Erben *et al.*, 2021).

The ZC3H28 gene is unique, with no paralogues. The central part of the protein has multiple proline-, histidine- or glutamine-rich regions, including (H)₁₁, (H)₁₇, (Q)₁₁ and (H)₁₄ with a single interruption (Fig. 1A). Consequently, Phyre2 predicts that about 70% of the protein lacks an ordered secondary structure. Low-complexity regions in proteins are known to undergo liquid-liquid phase separation, and proteins containing them are enriched in mRNA-protein granules that form in response to stress (Guzikowski *et al.*, 2019). C-terminally GFP-tagged ZC3H28 was indeed found in cytoplasmic aggregates in slightly starvation-stressed procyclic forms (see <http://tryptag.org/?query=Tb927.9.9450>) (Dean *et al.*, 2016), but the N-terminally tagged version was uniformly distributed, and ZC3H28 was not enriched in granules that formed after more prolonged starvation (Fritz *et al.*, 2015). ZC3H28 has at least five phosphorylation sites (Benz and Urbaniak, 2019; Ooi *et al.*, 2020) and is about twice as abundant in the procyclic (Tsetse midgut) form as in the bloodstream form that grows in mammals (Siegel *et al.*, 2010; Dejung *et al.*, 2016).

Results from a high-throughput RNAi screen indicated that ZC3H28 is essential in bloodstream forms, and during differentiation to, and early survival as, the procyclic form (Alford *et al.*, 2011). It can also be cross-linked directly to mRNA (Lueong *et al.*, 2016). In the 'tethering' assay, we express in trypanosomes the protein of interest fused to the lambdaN peptide, together with a reporter mRNA containing 'BoxB' sequences. BoxB is bound with high affinity by the lambdaN peptide. In a high-throughput screen, lambdaN-ZC3H28 activated BoxB reporter expression (Erben *et al.*, 2014). This suggests that ZC3H28 can either increase the translation efficiency of the mRNA to which it is bound, or can stabilize it, or both. A

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

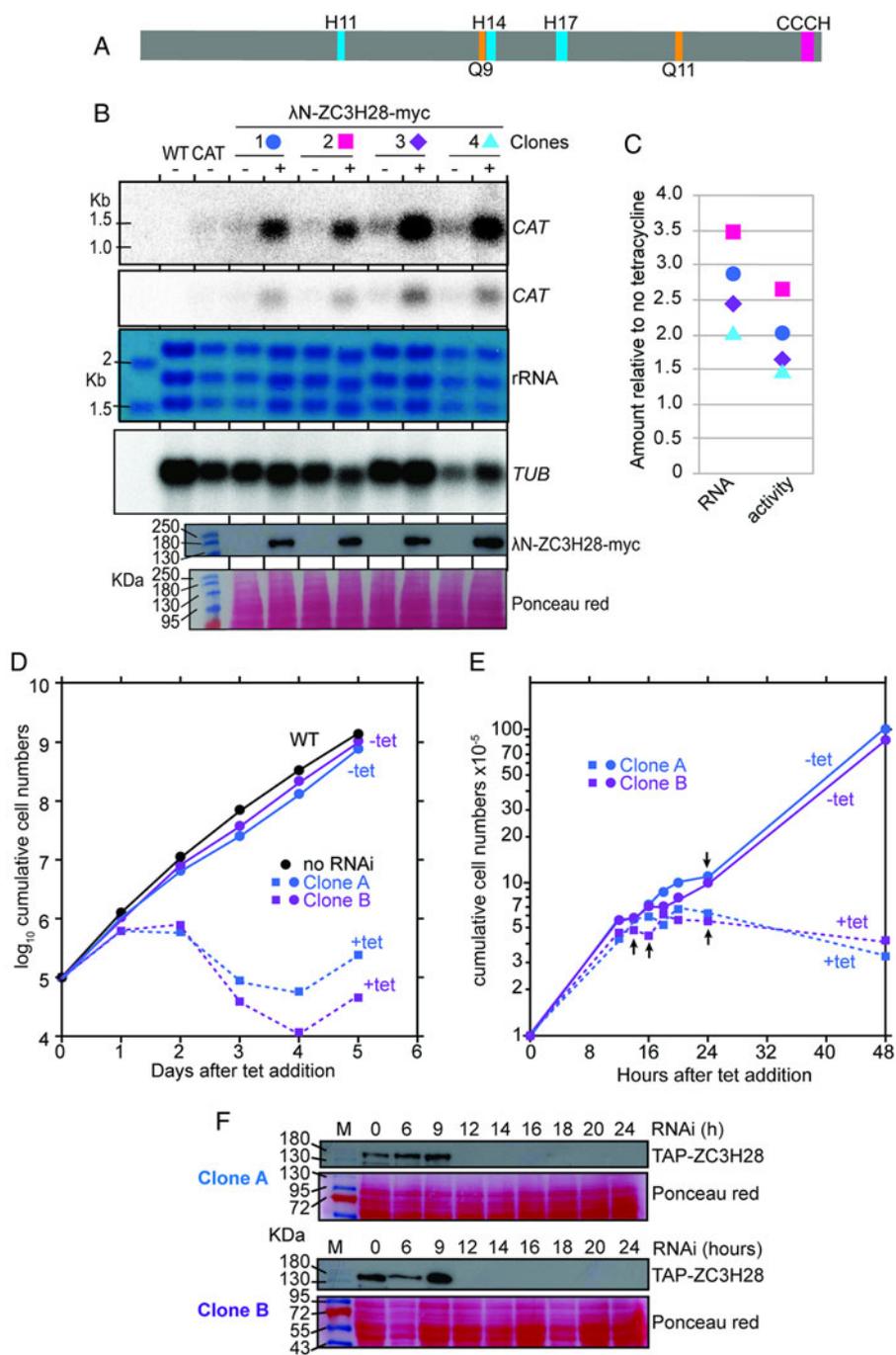


Fig. 1. ZC3H28 is essential for the normal growth of trypanosome bloodstream forms, and it increases the abundance of an attached mRNA. (A) Structure of ZC3H28, drawn to scale. The positions of the zinc finger domain, poly(His) and poly(Gln) sequences are indicated. (B) Effects of tethering lambdaN-tagged ZC3H28 on a boxB-bearing *CAT* reporter mRNA. This panel shows the expression of the *CAT* mRNA (2 different exposures), with a beta-tubulin control, in the upper four panels; the lower two panels show the expression of the lambdaN-tagged protein, which has a myc tag at the C-terminus. For the Northern blots, the first two lanes after the markers are cells with no *CAT* gene (WT) and cells with the reporter but no inducible lambdaN-tagged protein. These lanes are not present in the lowest two (Western blot) panels. (C) Effects of tethering on *CAT* activity and mRNA. For each cell line in (B), the amount of *CAT* mRNA or *CAT* activity in the presence of tetracycline was divided by the mRNA level or *CAT* activity in the absence of tetracycline. The symbols correspond to those in (B). (D) Effect of ZC3H28 depletion on cell numbers over 5 days. Expression of a dsRNA corresponding to part of the *ZC3H28* coding region was induced with tetracycline. Results for two different clones (A and B) containing the inducible construct are shown, either with or without tetracycline. Cells with no RNAi construct (WT) are also included for comparison. (E) Effect of ZC3H28 depletion on cell numbers over the first 48 h. The clones used were the same as in (D). The small arrows indicate time points at which RNA was harvested for sequencing. (F) Depletion of TAP-tagged ZC3H28 protein after RNAi for the experiment shown in E; the stained membrane serves as the loading control.

study of proteins associated with the two poly(A) binding proteins, PABP1 and PABP2, showed that ZC3H28 was preferentially associated with PABP2 (Zoltner *et al.*, 2018).

One known mechanism for post-transcriptional activation in trypanosomes is the recruitment of a complex containing MKT1, PBP1 [Poly(A)-binding-protein-Binding-Protein 1], XAC1 (eXpression ACticator 1), LSM12 and PABP2 (Singh *et al.*, 2014; Nascimento *et al.*, 2020). The complex containing MKT1, PBP1 and LSM12 was originally discovered in yeast, where it is implicated in survival after various stresses (e.g. Tadauchi *et al.*, 2004; Sinha *et al.*, 2006; Demogines *et al.*, 2008). PBP1 was so named because it interacts with yeast poly(A) binding protein (Mangus *et al.*, 1998). The trypanosome MKT1-PBP1-LSM12-XAC1 complex was first found associated with ZC3H11 (Singh *et al.*, 2014), a zinc finger protein that is required to stabilize chaperone mRNAs after heat shock (Droll *et al.*, 2013). *Trypanosoma brucei* PBP1 can interact with both

PABP1 and PABP2, but the MKT1 complex preferentially contains PABP2 (Singh *et al.*, 2014; Nascimento *et al.*, 2020). Tethering of any of these proteins (MKT1, XAC1, LSM12, PABPs or PBP1) enhances reporter expression, although activation by PBP1 was weaker than the others (Singh *et al.*, 2014; Nascimento *et al.*, 2020). Within cells, most XAC1 and PBP1 is in complexes, but there may be some free MKT1 (Nascimento *et al.*, 2020). Although they are essential, the precise roles of LSM12 and XAC1 are unknown (Nascimento *et al.*, 2020). In addition to the PBP1 interaction with PABPs, MKT1 directs the recruitment of one of the *T. brucei* translation initiation complexes, EIF4E6/EIF4G5 (Melo do Nascimento *et al.*, 2021). We speculated that recruitment of the MKT1-XAC1-LSM12-PBP1 complex (hereafter shortened to 'MKT1 complex') to mRNAs increases mRNA stability and translation through the combination of PABPs and EIF4E6/EIF4G5 (Melo do Nascimento *et al.*, 2021). PBP1, XAC1 and LSM12 also form a complex which contains an MKT1-like

protein, MKT1L, instead of MKT1; the role of this complex is unknown (Nascimento *et al.*, 2020).

Some RNA-binding proteins that interact with MKT1 have a consensus motif, H(N/D)PY, which is essential both for the interaction of the RNA-binding proteins with the MKT1 complex and for their ability to increase the abundance or translation of their target mRNAs (Singh *et al.*, 2014; Liu *et al.*, 2020; Melo do Nascimento *et al.*, 2021). However, other copurifying RNA-binding proteins lack the motif. Among these is ZC3H28, which was significantly enriched after affinity purification of XAC1 from bloodstream forms, but was not enriched after purification of MKT1 from procyclic forms (Nascimento *et al.*, 2020).

The mRNA encoding variant surface glycoprotein (VSG), the major surface protein of bloodstream-form trypanosomes, is stable in bloodstream forms and unstable in procyclic forms. Analysis of proteins that copurified with the VSG mRNA, but not alpha-tubulin mRNA, revealed a cyclin-like F box protein, CFB2, and H(N/D)PY – containing RNA-binding protein that is the major regulator of VSG mRNA abundance, along with the components of the MKT1 complex and EIF4G5 (Melo do Nascimento *et al.*, 2021). Once again, ZC3H28 was also among the proteins that were specifically enriched with VSG mRNA (Melo do Nascimento *et al.*, 2021).

Since ZC3H28 had been identified in several different contexts, we decided here to find out more about its function, by investigating its interactions with mRNAs and proteins and the effects of ZC3H28 depletion on the transcriptome.

Materials and methods

Trypanosome culture and modification

The experiments in this study were carried out using monomorphic *T. brucei* Lister 427 bloodstream from parasites constitutively expressing the Tet-repressor (Alibu *et al.*, 2005). The parasites were cultured at 37°C as routinely in HMI-9 medium supplemented with 10% heat-inactivated fetal bovine serum (v/v), 1% (v/v) penicillin/streptomycin solution (Labochem international, Germany), 15 µM L-cysteine and 0.2 mM β-mercaptoethanol in the presence of 5% CO₂ and 95% humidity. During proliferation, the cells were diluted to 1 × 10⁵ cells mL⁻¹ and maintained in density between 0.2–2 × 10⁶ as described in Clayton (1999). Cell densities were determined using a Neubauer chamber. For generation of stable cell lines, ~1–2 × 10⁷ cells were transfected by electroporation with 10 µg of linearized plasmid at 1.5 kV on an AMAXA Nucleofector. Selection of newly transfectants was done after addition of appropriate antibiotic at the following concentrations: 1 µg mL⁻¹ puromycin, 2.5 µg mL⁻¹ phleomycin (InvivoGen), 5 µg mL⁻¹ hygromycin B (Calbio-chem) and 10 µg mL⁻¹ blasticidin (InvivoGen). Independent clones were obtained by serial dilution.

Genetic manipulation of trypanosomes

A cell line with *in-situ* TAP-ZC3H28 gene was generated by replacing one endogenous copy of ZC3H28 with a gene encoding N-terminally TAP tagged ZC3H28. For this purpose, a construct with puromycin resistance gene plus a TAP tag cassette was flanked on the 5'-end with a fragment of ZC3H28 5'-UTR. Also, downstream on the 3'-end, the N terminal region of ZC3H28 ORF was cloned in frame with the TAP tag. Prior to transfection, the plasmid (pHD3236) was cut with *Sac* I and *Apa* I enzymes to allow homologous recombination. Using the cell lines expressing the *in-situ* N-TAP ZC3H28, we were unable to knock-out the other copy of ZC3H28. Gene fragments for RNAi were selected based on default settings of the RNAi software (Redmond *et al.*, 2003) and cloned so as to give an inducible

expression of a stem-loop (Estévez *et al.*, 2001). For the tethering assays, cell lines constitutively expressing the CAT reporter with boxB and the actin 3'-UTR were co-transfected with plasmids encoding the ZC3H28 in fusion with the λN-peptide and a myc tag (Erben *et al.*, 2014). The primers and plasmids used are listed in Supplementary Table S5 and sequences (which were verified in critical areas) are available upon request.

DNA extraction

Genomic DNA from *T. brucei* was isolated using 1–2 × 10⁸ cells as follows. The cell pellet was collected by centrifugation (2300 rpm, 8 min, 4°C), washed once in cold 1x PBS, and lysed in 0.5 mL of EB buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 10 mM EDTA). RNA was digested with the addition of 12 µL RNase A (1 mg mL⁻¹ stock solution, Sigma-Aldrich) at 37°C for 30 min. Proteins were precipitated using 200 µL ice-cold 5 M ammonium acetate followed by centrifugation at maximum speed for 5 min. The supernatant containing the DNA was transferred to a new tube. The DNA was then precipitated with 0.7x isopropanol followed by centrifugation at maximum speed for 15 min. The pellet was then washed once with 75% ethanol to remove salts and then again with 100% ethanol followed by centrifugation for 5 min. The DNA pellet was then dried for approximately 5 min and dissolved in TE buffer (10 mM Tris pH 7.5, 1 mM EDTA pH 8.0) at 37°C. The concentration was measured using a Nanodrop. PCR was done using Taq or Phusion DNA Polymerase according to the manufacturer's instructions (New England Biolabs).

RNA manipulation

To identify RNAs bound to ZC3H28, approximately 1 × 10⁹ cells expressing *in-situ* N-TAP tagged ZC3H28 with a concentration of 1 × 10⁶ cells mL⁻¹ were pelleted by centrifugation at 3000 rpm for 13 min at 4°C. The pellet was washed twice in cold 1x PBS and collected by centrifugation at 2300 rpm for 8 min at 4°C and then snap frozen in liquid nitrogen. The RNA immunoprecipitation was done essentially as described in Mugo and Clayton (2017). The cell pellet was lysed in 1 mL of the lysis buffer (20 mM Tris pH 7.5, 5 mM MgCl₂, 0.1% IGEPAL, 1 mM DTT, 100 U RNasin, 10 µg mL⁻¹ leupeptin, 10 µg mL⁻¹ Aprotinin) by passing 20 times through a 21 G × ½ needle using a 1 mL syringe and 20 times through a 27 G × ¼ needle using a 1 mL syringe. The lysate was cleared by centrifugation at 15 000 g for 15 min at 4°C and the supernatant was transferred to a new tube. The salt concentration was then adjusted to 150 mM KCl. The cell extracts were afterwards incubated with 40 µL of IgG-coupled magnetic beads (Dynabeads™ M-280 Tosylactivated, Invitrogen) for 3 h at 4°C and the flow-through (unbound) fraction was collected by magnetic separation as the negative control. peqGOLD TriFast™ FL REAGENT was added to the unbound fractions and kept at –80°C for further RNA extraction. After three washing with IP buffer (20 mM Tris pH 7.5, 5 mM MgCl₂, 150 mM KCl, 0.1% IGEPAL, 1 mM DTT, 100 U RNasin, 10 µg mL⁻¹ leupeptin, 10 µg mL⁻¹ Aprotinin), the tagged protein was eluted from beads using 150 units of tobacco etch virus (TEV) protease at 4°C for overnight. The eluate was transferred to a fresh tube, two volumes of peqGOLD TriFast™ FL reagent were added, and samples were stored at –80°C until further processing. RNA was isolated from released and bound fractions according to the manufacturer's instructions. Total RNA from the unbound and the eluate fraction were depleted of ribosomal RNA (rRNA) using a cocktail of 131 DNA oligos (50 bases) complementary to the trypanosome rRNAs. The rRNAs hybridized to the oligonucleotides were digested with RNase H (NEB, M0297S) as previously described in Minia *et al.* (2016). Following rRNA depletion, the samples

were subjected to DNase I treatment in order to remove any trace of oligonucleotides using the Turbo DNase kit (Invitrogen, ThermoScientific). The RNA samples were afterward 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 were then analysed by RNA-Seq.

For Northern blotting, the purified RNA (5–10 µg) was mixed with 2x RNA loading dye (1.6x MOPS buffer, 7% formaldehyde, 65% formamide, 50 µg mL⁻¹ ethidium bromide, 0.025% bromophenol blue), denatured for 10 min at 65°C and then resolved on formaldehyde agarose gel. The RNA was afterwards blotted onto Nylon membranes (Amersham Hybond-N+, GE Healthcare, RPN203B) with 10x saline-sodium citrate buffer (SSC) by capillary transfer overnight. The RNA was then cross-linked to positively charged membranes using a UV-crosslinker (Stratagene UV Stratalinker 2400, 2 × 240 mJoules) and stained with methylene blue (SERVA) for 10 min. The northern blots were pre-hybridized in hybridization solution for 1 h at 65°C and then hybridized in the same solution with the appropriate (α -³²P) dCTP radioactively labelled DNA probes from CAT and tubulin genes for overnight at 65°C. Labelling the DNA probes was done with Prime-IT RmT Random Primer Labelling Kit, Stratagene. The following day, the blot was washed twice for 10 min at room temperature with wash solution 1 (2x SSC, 0.1% SDS), twice for 10 min with wash solution 2 (1x SSC buffer, 0.1% SDS) and twice for 10 min at 65°C with wash solution 3 (0.1x SSC, 0.1% SDS). For spliced leader detection, a 39-mer oligonucleotide complementary to the spliced leader was labelled with [³²P]-ATP using T4 polynucleotide kinase (NEB) and incubated with the membrane overnight at 42°C (Manful *et al.*, 2011; Fadda *et al.*, 2013). Afterwards, the blots were exposed to autoradiography films and the signals were detected with the phosphorimager (Fuji, FLA7000). The images were processed and quantified using ImageJ.

RNA sequencing and data analysis

RNA sequencing was done at the Cell Networks Deep Sequencing Core Facility at the University of Heidelberg. NEB Next Ultra RNA Library Prep Kit for Illumina (New England BioLabs Inc.) was used for library preparation. The libraries were multiplexed (6 samples per lane) and sequenced with a NextSeq 550 system, generating 75 bp single-end sequencing reads. This was done using a custom pipeline (Leiss *et al.*, 2016) that incorporated the following steps. 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, poly(A) tails and spliced leaders. After primer removal, the sequencing data were aligned to *T. brucei* 927 reference genome using Bowtie2 (Langmead and Salzberg, 2012), allowing 1 alignment per read, then sorted and indexed using SAM tools (Li *et al.*, 2009). Reads aligning to open reading frames, annotated 3'-untranslated regions and functional non-coding RNAs were counted. The alignment and counting were then repeated with Lister 427 genome (2018 assembly) (Müller *et al.*, 2018). For comparative and enrichment analyses, we used a list of unique genes modified from Siegel *et al.* (2010) in order to avoid giving excessive weight to repeated genes and multigene families. For the RIP-Seq, the reads per million were counted and the ratios of eluate vs unbound were calculated. An mRNA was considered as 'bound mRNA' if the lowest ratio was at least 3. The motif enrichment search was done using MEME with default settings (Bailey, 2011). Annotated 3'-UTRs were downloaded from TritrypDB. Analysis of differentially expressed genes after ZC3H28 RNAi was done in R using the DESeqUI (Leiss and Clayton, 2016), a customized

version of DESeq2 package (Love *et al.*, 2014) adapted for trypanosome transcriptomes. Statistical analyses were done using R and Microsoft Excel.

Yeast two-hybrid assays

The Matchmaker Yeast two-hybrid system (Clontech) was used to test direct protein-protein interactions according to the manufacturer's instructions. The coding sequence of ZC3H28 was PCR-amplified from genomic DNA and cloned into pGBKT7 and pGADT7 plasmids. The prey and the bait plasmids of ZC3H28 as well as those of the MKT1 complex proteins were co-transformed pairwise into AH109 yeast strains. Selection was done initially on double drop-out plates (i.e. SD medium lacking Tryptophan and Leucine) to check the expression of both bait and prey. The growth was then checked on quadruple drop-out plates (i.e. lacking Tryptophan, Leucine, Histidine and Adenine) that indicate positive interactions. The interaction between p53 and SV40 large T-antigen and the combination of LaminC and SV40 large T-antigen served as positive and negative controls, respectively.

Protein purification and mass spectrometry analysis

Approximately 1 × 10⁹ cells expressing either *in-situ* N-TAP ZC3H28 (pHD3236) or tetracycline-inducible GFP-TAP (pHD1743) with a concentration of 1 × 10⁶ cells mL⁻¹ were harvested by centrifugation. For each cell line, three technical replicates were done without RNase A treatment. The cell pellet was resuspended in 10 mL of cold 1x PBS and centrifuged at 2800 rpm for 8 min at 4°C and the pull-downs were done as described above, except that after TEV cleavage, the eluate was collected by magnetic separation. To remove His-tagged TEV, 10 µL of equalization buffer (200 mM sodium phosphate, 600 mM sodium chloride, 0.1% Tween-20, 60 mM imidazole, pH 8.5), as well as 30 µL of Ni-NTA-magnetic beads were added and incubated with the samples for 30 min at 20°C while rotating. Ni-NTA magnetic beads were separated using a magnetic rack and the supernatant was collected and stored in 6x Laemmli buffer at -80°C. Eluted proteins were separated on 12% SDS-polyacrylamide gel until the running front had migrated roughly 2 cm, thereafter the gel was stained with Coomassie blue and destained with destaining solution (10% acetic acid, 50% methanol in H₂O). Two areas per lane were cut and analysed in the ZMBH Mass Spectrometry facility *via* the Ultimate 3000 liquid chromatography system directly coupled to an Orbitrap Elite mass spectrometer (Thermo Fisher). MS spectra (m/z 400–1600) were acquired in the Orbitrap at 60 000 (m/z 400) resolution. Fragmentation in CID cell was performed for up to 10 precursors. MS2 spectra were acquired at rapid scan rate. Raw files were processed using MaxQuant (version 1.5.3.30; J. Cox, M. Mann, Nat Biotechnol 2008, 26, 1367) for peptide identification and quantification. MS2 spectra were searched against the TriTrypDB-8.1TREU927-Annotated Proteins-1 database (containing 11567 sequences). Data were analysed quantitatively and plotted using Perseus software (Version 1.6.15.0).

Protein detection by western blotting

For western blotting, 1–5 × 10⁶ cells were collected by centrifugation at 3000 rpm for 5 min, washed twice in ice-cold 1x PBS, lysed in 2 × Laemmli buffer and heated at 95°C for 10 min. The protein samples were separated using 12% SDS-PAGE gels, and processed as described previously in Minia and Clayton (2016). The following antibodies were used for specific protein detection: anti-PAP (1:15000, rabbit, Sigma), anti-myc (mouse, 1:1000); anti-rabbit IgG (for pull-downs). The proteins were detected by enhanced

chemiluminescence according to the manufacturer's instructions (Amersham Biosciences).

CAT Assay

To perform the CAT assay experiment, approximately 2×10^7 cells expressing the CAT reporter gene were harvested at 2300 rpm for 8 min and washed three times with 1x cold PBS. The pellet was re-suspended in 200 μ L of CAT buffer (100 mM Tris-HCl pH 7.8) and lysed by freeze-thawing three times using liquid nitrogen and a 37°C heating block. The supernatants were then collected by centrifugation at 15 000 $\times g$ for 5 min and kept on ice. The protein concentrations were determined by Bradford assay (BioRad) according to the manufacturer's protocol. For each setup, 0.5 μ g of protein in 50 μ L of CAT buffer, 10 μ L of radioactive butyryl CoA (^{14}C), 2 μ L of chloramphenicol (stock: 40 mg mL $^{-1}$), 200 μ L of CAT buffer and 4 mL of scintillation cocktail were mixed in a Wheaton scintillation tube HDPE (neoLab #9-0149). The incorporation of radioactive acetyl group on chloramphenicol was measured using program 7 of Beckman LS 6000IC scintillation counter.

Pulse labelling

For each time point, approximately 4×10^6 cells were collected at room temperature after the incubation times indicated in Supplementary Fig. S3B. The pellet was washed twice with ice-cold 1x PBS followed by centrifugation at 4000 g for 3 min. The cell pellet was then resuspended in 400 μ L labelling medium [Dulbecco's Modified Eagle Medium (Gibco) lacking L-methionine and cysteine] at 37°C for 1 h. Two microliters of L-[^{35}S] methionine (about 20 μ Ci) was then added. The cells were incubated for 1 h at 37°C and afterwards collected by centrifugation at 4000 g for 3 min at RT. The pellet was washed twice with 1x PBS and then resuspended in 15 μ L of Laemmli lysis buffer. The proteins were separated in a 12% SDS gel. The gel was then dried onto a Whatman paper, then exposed to autoradiography films and the signals were detected with the phosphorimager (Fuji, FLA7000).

The labelling medium was Dulbecco's modified Eagle's medium (Gibco, high-glucose, containing pyridoxine hydrochloride, lacking L-glutamine, sodium pyruvate, L-methionine and L-cysteine), supplemented with 25 mM HEPES, 2 mM glutamine, 0.1 mM hypoxanthine, 0.0028% b-mercaptoethanol, 0.05 mM bathocupron sulfate and 10% heat-inactivated fetal calf serum (previously dialysed against 30 mM HEPES, pH 7.3, 150 mM NaCl) (Leiss *et al.*, 2016).

Results

ZC3H28 is present in Kinetoplastids and a Bodonid

Examination of representative Kinetoplastid genomes revealed that ZC3H28 is present in all of them (Supplementary Fig. S1). The position in the genome upstream of the gene encoding peroxisomal protein PEX13 is also conserved although the syntenic region is not annotated. We also found a homologue in the Bodonid *Bodo saltans*, but not in *Euglena gracilis*. All of the sequences examined (except the *Leptomonas seymouri* sequence, which has a frame-shift near the C-terminus) have the C-terminal zinc finger and various histidine- and glutamine-rich regions (Fig. 1A, Supplementary Fig. S1).

ZC3H28 is essential for normal bloodstream-form trypanosome growth

Results of high-throughput screens suggested that ZC3H28 is essential for normal trypanosome growth, and that it activates expression when tethered to a reporter mRNA (see Introduction). Since results

obtained in high-throughput screens are not always reliable, we first confirmed that lambdaN-ZC3H28 indeed activates the expression of a boxB-containing chloramphenicol acetyltransferase (CAT) – encoding mRNA. Results for four independent clones confirmed 1.5–2.7-fold increases in CAT protein and 2–3.5-fold increases in the amount of mRNA (Fig. 1B, C). We do not know whether this small discrepancy is significant, since it could be caused by the different measurement methods. Tethering of MKT1 or LSM12 increased CAT protein 3 times more than it increased CAT mRNA (Singh *et al.*, 2014). However, when PBP1 was tethered, the effects on reporter mRNA and protein were lower, and similar to each other; and when ZC3H11 was tethered, the effects on mRNA exceeded those on protein (Droll *et al.*, 2013). The original tethering screen was done using a library of genomic fragments with an average size of 1.2 kb (Erben *et al.*, 2014). Since the open reading frame of ZC3H28 is over 3 kb long, it is unlikely that any of the fragments encoded the full-length protein. The raw results (Supplementary Fig. S2A) show that fragments that start in several positions within the open reading frame activated the expression of the boxB-bearing selectable marker. This suggests that ZC3H28 contains more than one sequence capable of gene expression activation.

Next, we confirmed the effects of ZC3H28 depletion on the proliferation of bloodstream-form trypanosomes. We initially integrated a plasmid for inducible RNAi in EATRO1125 cells, which are competent for differentiation into the procyclic (tsetse midgut) form. ZC3H28 RNAi only somewhat inhibited cell proliferation (data not shown). We therefore instead used a cell line that we had generated for affinity purification of ZC3H28. We integrated a sequence encoding a tandem affinity purification (TAP) tag in-frame with one ZC3H28 allele, which should result in the production of N-terminally tagged ZC3H28 (TAP-ZC3H28). This line was made with Lister 427 bloodstream forms, which tolerate higher densities than EATRO1125 but are unable to complete differentiation. Our initial test showed that after induction of RNAi, growth inhibition was detectable 24 h in two independent clones (Fig. 1D). Many of the cells subsequently died, but after 4 days, some survivors resumed growth. We did not investigate this recovery but it is most likely caused by re-expression of ZC3H28: loss of growth-inhibitory tetracycline-inducible RNAi is common in trypanosomes (e.g. Jha *et al.*, 2014; Klein *et al.*, 2017; Liu *et al.*, 2020). A more detailed investigation showed that ZC3H28 RNAi resulted in the loss of the protein within 12 h, and that the onset of growth inhibition occurred after about 18 h (Fig. 1D, E). These results confirmed that ZC3H28 is required for the normal growth of bloodstream forms.

To find out whether the tagged ZC3H28 protein was fully functional, we attempted to delete the unmodified ZC3H28 gene, which should result in cells expressing only TAP-ZC3H28. Unfortunately, we were unable to select such cells. This may mean that TAP-ZC3H28 is not fully functional. Alternatively, it is possible that trypanosomes require two copies of the ZC3H28 gene for growth, so that a deletion results in haploid insufficiency. We did not investigate this further but for interpretation of our results, it is necessary to bear in mind that TAP-ZC3H28 may not be fully functional.

Interactions of ZC3H28 with other proteins

Results from quantitative mass spectrometry using the MKT1-complex component XAC1 as bait suggested that ZC3H28 is associated with XAC1-containing complexes (Nascimento *et al.*, 2020). However, ZC3H28 lacks the H(N/D)PY consensus MKT1 interaction motif. We therefore examined the interactions of ZC3H28 with the different MKT1-complex components using the

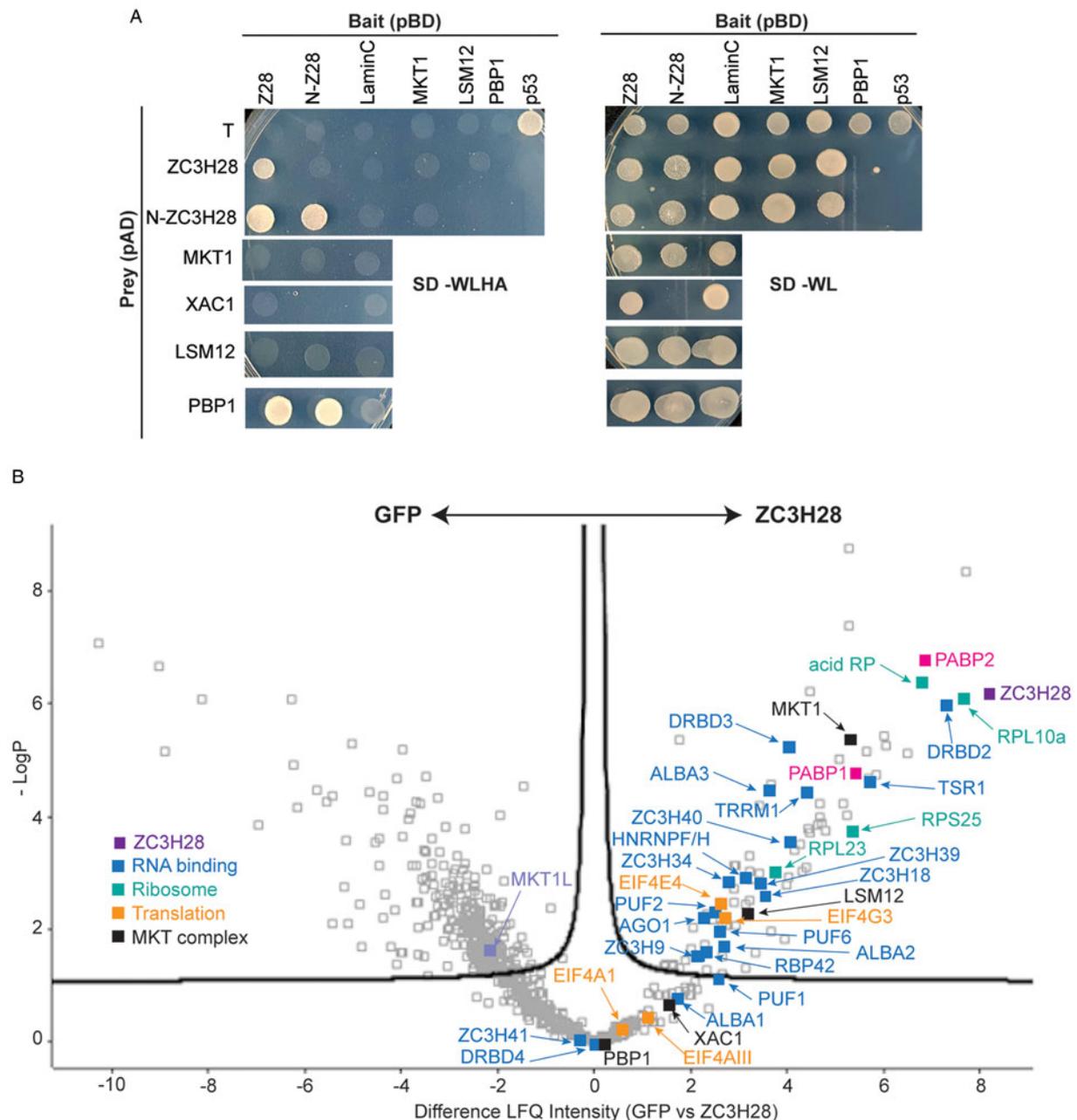


Fig. 2. Protein interactions of ZC3H28. (A) Interactions of ZC3H28 in the yeast two-hybrid assay. *Saccharomyces cerevisiae* AH109 was transformed with combinations of plasmids used as bait (pBD, DNA binding domain) and prey (pAD, transcription activating domain). Growth on media lacking adenine, histidine, leucine and tryptophan (SD-WLAH) indicates protein interactions while growth in media lacking leucine and tryptophan (SD-WL) controls for expression of the bait and prey. The interaction between SV40 large T-antigen (T) and p53 (first row, last column) serves as positive control and the combination of LaminC and SV40 large T-antigen (first row, third column) serves as negative control. PBP1 and XAC fused with the DNA-binding domain both activate when expressed alone (Nascimento *et al.*, 2020), therefore they are used here only as activation domain fusions, which give appropriate negative control results (Nascimento *et al.*, 2020). (B) Volcano plot of proteins interacting with TAP-ZC3H28, compared with GFP-TAP which served as the negative control. The figure was generated using Perseus software (Tyanova *et al.*, 2016). The y-axis is the negative log₁₀ of the *P* value for significant enrichment; higher values are more significant. The x-axis shows differential signal intensities on a log₂ scale. Selected proteins are indicated and all significantly enriched proteins are listed in Supplementary Table S2. The colour code is shown in the figure: purple is ZC3H28, blue indicates RNA binding proteins, cyan designates ribosomal proteins, orange is assigned to other proteins involved in translation, black denotes proteins of the MKT1 complex, and the polyA binding proteins are indicated in magenta.

yeast two-hybrid assay (Fig. 2A). In this assay, ZC3H28 interacted with itself, and with PBP1, but not with MKT1. A truncated version of ZC3H28 including the N-terminal 527 residues [containing two poly(His) and one poly(Gln) segments] also interacted with itself and with PBP1; this N-terminal version gave a detectable interaction with full-length ZC3H28 only when the latter was in the 'prey' configuration. These results suggest that ZC3H28 might interact with itself and with PBP1, but not with MKT1. It is important to bear in mind, however, that the DNA binding and activation domains that were fused to the tested proteins might affect their interactions;

and that post-translational modifications present in trypanosomes may well be absent in yeast.

To find out whether ZC3H28 is associated with the MKT1 complex *in vivo*, we purified TAP-ZC3H28 in triplicate and analysed the co-purifying proteins by quantitative mass spectrometry, with GFP-TAP as the control (Supplementary Table S1, Fig. 2B). We denoted proteins as being associated with ZC3H28 using two criteria: proteins that were at least 4-fold enriched, with *P* values of <0.01 [obtained using the Perseus algorithm (Tyanova *et al.*, 2016)]; and a partially overlapping set of proteins that were

present in all three ZC3H28 preparations and absent in all GFP preparations. (Some of the latter are not denoted as enriched by Perseus because it simulates the abundances of absent proteins.) MKT1, XAC1 and LSM12 consistently copurified with ZC3H28, MKT1 being one of the most significantly enriched proteins (Fig. 2B). In contrast, MKT1L was not associated. Comparing numbers of detected peptides, the ratios for ZC3H28 to the GFP control were as follows: XAC1: 1–2 peptides for ZC3H28, 0 for GFP; LSM12: 3–6 LSM12 peptides for ZC3H28, 0 for GFP; MKT1: 19–23 MKT1 peptides for ZC3H28, 1 in one GFP replicate. These numbers reflect the different protein lengths (117 residues for XAC1, 190 for LSM12 and 735 for MKT1). PBP1 (550 residues) was anomalous since it was present in only 2 of the 3 ZC3H28 replicates (but none of the controls). Loss of PBP1 due to the buffer conditions is unlikely, since the conditions were exactly the same as those used in previous analyses of the MKT1-PBP1-XAC1-LSM12 complex (Nascimento *et al.*, 2020). Since TAP-ZC3H28 may not be fully functional, it is possible that the N-terminal tag interfered with some interactions. However, the yeast two-hybrid constructs were also N-terminal fusions, so tags alone cannot explain the low number of PBP1 peptides that were detected. Comparison of these results with those obtained using DRBD18-TAP as bait (included in the same gel and mass spectrometry run) (Bishola Tshitenge *et al.*, 2021) confirmed selective co-purification of MKT1, XAC1, LSM12 and PBP1 with ZC3H28.

A large number of ribosomal proteins copurified with ZC3H28 (Supplementary Table S1, Fig. 2B), with a possible bias towards the small subunit. Fourteen of the 27 small subunit proteins were found, together with 14 of the 50 large subunit proteins, including the four acidic subunits and the Kinetoplastid-specific protein NRBD2 (Jaremko *et al.*, 2019). There was no obvious link between the detection of these proteins in the pull-down, and their protein sizes or their locations on the ribosome (Jenner *et al.*, 2012). Most were also found with DRBD18 (Bishola Tshitenge *et al.*, 2021). Interestingly, ZC3H28 was associated with just one of the five known EIF4E-EIF4G translation initiation complexes, EIF4E4-EIF4G3, but the only general translation factor found was the release factor subunit ERF1, which argues against the direct association of ZC3H28 with scanning small subunits (Bochler *et al.*, 2020).

Both of the poly(A) binding proteins, PABP1 and PABP2, co-purified with ZC3H28. These are thought to have somewhat different functions, although the details are unclear. A study of their interactions showed that ZC3H28 was preferentially associated with PABP2 (Zoltner *et al.*, 2018). EIF4E4 and EIF4G3 were found with both PABPs (Zoltner *et al.*, 2018). Of the many other RNA-binding proteins that copurified with ZC3H28, ALBA1, ALBA2 (Mani *et al.*, 2011), DRBD2, ZC3H34 and ZC3H41 were associated with both PABP1 and PABP2; HNRNPF/H, TSR1, TRRM1, ZC3H39 and ZC3H40 were preferentially associated with PABP2, and ALBA3 with PABP1 (Zoltner *et al.*, 2018). A pull-down of *T. cruzi* DRBD2, which is a likely homologue of the yeast translation factor Gbp2, also showed co-purification of ALBA proteins, UBP1, UBP2, DRBD3 and PABP2 (Wippel *et al.*, 2019). A Kinetoplastid-specific PSP1-domain protein (Tb927.10.9330) that was associated with ZC3H28 was also found with both PABPs (Zoltner *et al.*, 2018); the function of the domain is unknown. RNA-binding proteins that were specific to the ZC3H28 purification and not enriched with the PABPs were PUF2 (Jha *et al.*, 2014), RBP42 (Das *et al.*, 2012), ZC3H18 (Benz *et al.*, 2011) and ZC3H9. The comparison with DRBD18 showed that PUF1, PUF6, ZC3H18 and ZC3H35 were specific to ZC3H28 (Bishola Tshitenge *et al.*, 2021). ZC3H28 also pulled down the RNA interference effector AGO1 (Shi *et al.*, 2004a, 2004b), which does not copurify with

the PABPs. This was the only copurified protein that has any link with mRNA degradation. Some of the ZC3H28-linked RNA-binding proteins are implicated in splicing control: these include the putative regulators TSR1 (Gupta *et al.*, 2014) and HNRNPF/H (Gupta *et al.*, 2013), which are probably also associated with cytosolic mRNAs; and TRRM1, which has been implicated in transcription elongation (Levy *et al.*, 2015; Naguleswaran *et al.*, 2015; Banuelos *et al.*, 2019).

These mass spectrometry results were consistent with an association of ZC3H28 with ribonucleoprotein particles or polysomes including numerous RNA-binding proteins and the cap-binding initiation complex EIF4E4/EIF4G3.

mRNA interactions of ZC3H28

To find out which mRNAs are preferentially bound by ZC3H28, we purified the TAP-tagged protein, cleaved the tag with TEV protease and identified the co-purifying RNAs (Supplementary Table S2). In the three pull-downs, the mRNA encoding ZC3H28 was 2.3-, 2.8- and 4.5-fold enriched. A degree of 'self mRNA' purification is common when the affinity tag is at the N-terminus of the protein that is being purified (Erben *et al.*, 2021); it can probably be attributed to the purification of polyosomal RNA *via* the nascent polypeptide. Unexpectedly, VSG mRNA was not preferentially bound (Supplementary Table S2). The *T. brucei* genome includes many repeated genes, so in order to analyse the enrichment of mRNAs with particular characteristics, or functional classes of protein products, we used a list of 'unique' genes from which paralogues have been excluded. This list is the one published in Siegel *et al.* (2010), updated to include some additional genes including those with short open reading frames (Supplementary Table S2).

Standard programs for statistical analysis of RNA-Seq data cannot be used for RNA pull-downs because their underlying assumption, which is that most mRNAs have similar abundances in both conditions, is not valid. For further study, we therefore simply selected only mRNAs that were at least 3-fold enriched in all of the purifications. This cut-off (which is arbitrary) resulted in a list of 180 different mRNAs (Supplementary Table S2 sheet 1). (This number is for a list of 'unique' genes which excludes paralogues; see Methods section and Supplementary Table S2.) Figure 3A shows that the median length of mRNAs that were associated with ZC3H28 was twice that for all mRNAs. The difference for 5'-UTRs was similar, while that for coding regions was smaller (1.5x). The biggest discrepancy was for 3'-UTRs, where the ZC3H28-bound median was 3.7 times longer than that for all mRNAs. However, the correlation between ZC3H28 binding and length was only partial, suggesting a degree of sequence specificity (Fig. 3B). (A caveat to Fig. 3A and B is that the database annotation of 3'-UTRs in the trypanosome database is unreliable.) We looked at functions of encoded proteins, using annotations and our own functional categories, which are derived from database annotation and the literature. Most *GRESAG4* mRNA representatives were enriched, but this result is difficult to interpret: this is a large gene family related to *ESAG4* genes, so read alignment can be problematic. A Fisher test for enrichment in the 3-fold bound list (corrected for multiple testing) gave a false discovery rate (FDR) of 2×10^{-8} for this category. Binding was on average highest for mRNAs encoding transporters and proteins involved in mitochondrial DNA synthesis (Fig. 3C), but examination of the list of 180 enriched mRNAs shows that the mRNAs in these categories that were selected were mostly unusually long. The Fisher test gave an FDR of 8×10^{-3} for the 'transporter' category, but non-significant for mitochondrial DNA synthesis. The bound fraction also included 13 mRNAs encoding protein kinases (FDR 1.3×10^{-3}) and 12 encoding RNA-binding proteins (FDR

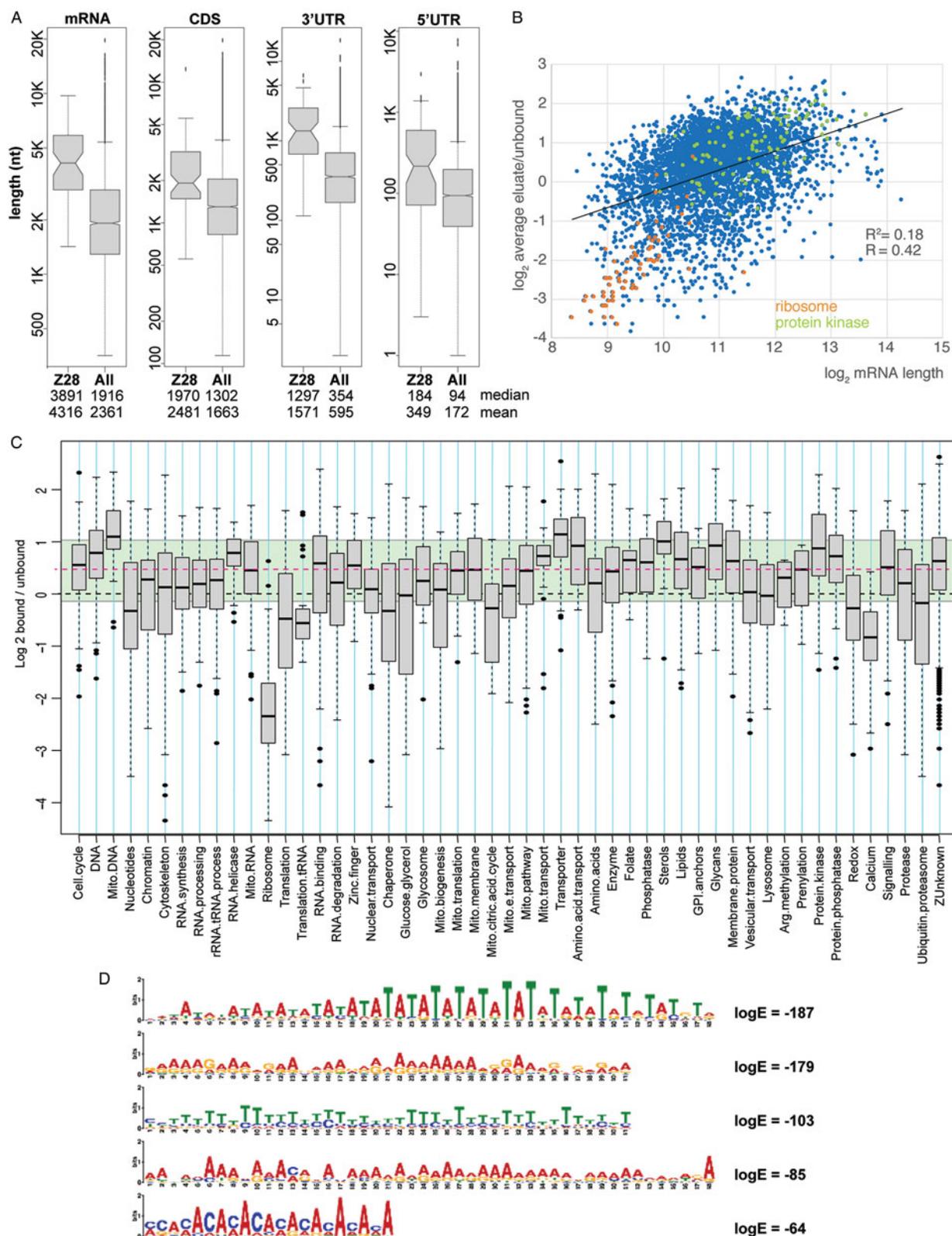


Fig. 3. Interactions of ZC3H28 with mRNA. (A) Characteristics of mRNAs that were at least 3-fold enriched with TAP-ZC3H28 relative to the unbound fraction. Results for the whole transcriptome are shown for comparison. All results in this figure are for the set of unique genes, to avoid over-counting repeated genes. In these box plots, the boxes extend from the 25th to 75th percentiles, and the whiskers to 1.5x the inter-quartile length; dots are outliers. The notches are an indicator of variability, $\pm 1.57x$ the inter-quartile range divided by the square root of the number of data points. Median and mean values are shown underneath the labels. (B) Plot for the whole dataset: enrichment with ZC3H28 (y -axis) relative to mRNA length (x -axis). (C) The set of unique genes was placed in functional categories (Supplementary Table S4) and binding to ZC3H28 was plotted for each category. The black dotted line indicates equal distribution in both fractions. The magenta dotted line is the median and the green background box indicates the range from the 25th to the 75th percentiles. (D) Motifs enriched in the 3'-UTRs of ZC3H28-bound mRNAs relative to size-matched controls showing less than 1-fold average enrichment in the bound fraction.

1.2×10^{-3}), but these again tend to be very long mRNAs (Clayton, 2019; Erben *et al.*, 2021) and many of the mRNAs in both categories showed only average enrichment (Fig. 3C). A GO term

analysis revealed enrichment only for 'intracellular signal transduction' (FDR 0.034) and related sub-categories, consistent with the enrichment of protein kinase and RNA binding protein

mRNAs (Supplementary Table S2 sheet 7). There was in contrast no association of ZC3H28 with ribosomal protein mRNAs, which are mostly very short (Clayton, 2019; Erben *et al.*, 2021) (Fig. 3B and C). A search for motifs using the 180 enriched full-length transcripts, compared with a set of length-matched controls with average binding ratios of <1, revealed weak enrichment of poly(U) and poly(AU) sequences, but these are common in trypanosome 3'-UTRs, and the median 3'-UTR lengths of the length-matched controls were half those for the 3xbound set. We therefore looked for motifs in the annotated 3'-UTRs only, again using appropriately length-matched controls. This revealed significant enrichment of poly(AU) and polypurine tracts in the bound mRNA 3'-UTRs (Fig. 3D). However, none of the motifs was present in all bound mRNAs, or exclusive to them. For example, the sequence 'AUAUAUAUA' was present at least once in 49 of the 106 'bound' 3'-UTRs but also in 36 of the 103 unbound ones.

To further study ZC3H28 specificity, we compared its binding with that of other proteins that preferentially bind to long mRNAs (Erben *et al.*, 2021 and see Discussion). Results showed clear differences. Among those for which comparable RNA-Seq data are available, the best correlation was for DRBD7 (Supplementary Fig. S2B, Supplementary Table S2). Nevertheless, even here the correlation was far from perfect ($R=0.57$), indicating that a degree of sequence specificity is superimposed on the length effects.

The effect of ZC3H28 depletion on the transcriptome

The results so far were consistent with partially selective association of ZC3H28 with mRNAs. They also demonstrated co-purification of many RNA-binding proteins, and the MKT1 complex. From the tethering result, we would expect the association of ZC3H28 to enhance the abundance of associated mRNAs. To test the role of ZC3H28 further, we therefore examined the effect of ZC3H28 depletion on the transcriptome. Total RNA from controls and tetracycline-treated cells (Fig. 1E and F) was purified, and subjected to rRNA depletion before sequencing. Neither tetracycline treatment, nor our rRNA depletion protocol has any significant effect on the transcriptome (Luu *et al.*, 2006; Mulindwa *et al.*, 2021). The results after alignment of the reads to all coding regions, non-coding RNAs (excluding rRNA) and 3'-untranslated regions are in Supplementary Table S3. The results for the 'unique' gene list, which was the basis for Figs 4C, D and E, are in Supplementary Table S4.

First, we looked at triplicate samples from an RNAi clone (clone A) grown without tetracycline, and after 10 h incubation with tetracycline. This time point was chosen because 10 h RNAi induction was insufficient to affect growth (Fig. 1E). Very few differences were found, so we also examined further samples from two clones (A and B) after 14 or 16 h with tetracycline, and 24 h either with or without tetracycline. At all these time-points after induction, the tagged protein was reproducibly undetectable (Fig. 1F). The cells with tetracycline initially grew only very slightly slower than the controls, but after 24 h the numbers started to decrease (Fig. 1E). Amounts of ZC3H28 mRNA were reduced 3–4-fold within 10 h and remained at that level throughout (Fig. 4A). Intriguingly, a principal component analysis (Fig. 4B) clearly separated the first set of controls (labelled A0-1, A0-2 and A0-3, grey squares) from the second set (labelled A0-4 and B0, black squares). The 10 h induction samples (A10-1, A10-2 and A10-3) clustered with their own control (Fig. 4B), while the 24 h induction (A24, B24) was similar to its controls (A0-4 and B0). In contrast, RNAi induction for 14 h (A14, B14) and 16 h (A16, B16) had clear effects on the transcriptome. For the 14 h time point, to check overall count reliability, we also looked at the reads aligned to the 3'-UTRs: there was an excellent

correlation with the coding region results (Supplementary Fig. S2C).

The cell densities for the triplicate untreated samples A0-1, A0-2 and A0-3 (Fig. 4B) were $5\text{--}6 \times 10^5 \text{ mL}^{-1}$, whereas the duplicate A0-4 and B0 (Fig. 4B) were harvested at approximately $1 \times 10^6 \text{ mL}^{-1}$ (Fig. 1E). The cells after 24 h RNAi had densities of $1.3\text{--}1.6 \times 10^6 \text{ mL}^{-1}$, whereas the densities for the 14 and 16 h time-points were lower (Fig. 1E, Supplementary Table S3). The maximum density attained by these cells is $2\text{--}3 \times 10^6 \text{ mL}^{-1}$, so we suspected that the density might have influenced the RNAi effect. We therefore first looked at the differences between the two sets of controls in order to assess the effects of cell density. The differences were modest, but the 180 mRNAs that were 1.5-fold significantly ($\text{Padj} < 0.05$) increased at the higher density included those encoding the procyclins. (These cut-offs are arbitrary but often used.) The category 'membrane proteins' was significantly enriched, $\text{FDR} 1.2 \times 10^{-5}$. Other enriched categories were 'cytoskeleton' ($\text{FDR} 0.024$, mostly flagellar proteins) and 'ESAG' ($\text{FDR} 0.017$). (No GO term enrichment was found.) A few mRNAs encoding proteins of mitochondrial metabolism were also increased. Eighty-two mRNAs were significantly decreased to a similar extent; there were no significantly enriched categories but the affected mRNAs included some encoding RNA polymerase I subunits, nucleotide transporters and translation elongation factor 2, all of which would be consistent with slowing growth. Some of these changes are consistent with very early differentiation (Silvester *et al.*, 2018) or a stress response (Quintana *et al.*, 2020) (Supplementary Table S4). A preliminary result suggested that the amounts of total mRNA or protein synthesis were not decreased by ZC3H28 RNAi alone (Supplementary Fig. S3A–C), but that trypanosomes at densities of over $10^6 \text{ cells mL}^{-1}$ showed 20–60% less translation compared with cells at $3\text{--}5 \times 10^5 \text{ mL}^{-1}$ (Supplementary Fig. S3B and C). A possible interpretation of these results is that all of the cells that had densities above $1 \times 10^6 \text{ mL}^{-1}$ were stressed, and that the effects of ZC3H28 RNAi on mRNA abundance can only be seen when the cells are growing optimally. Therefore, we decided to focus on the effects at the 14 and 16 h time-points.

The effects of ZC3H28 RNAi on the transcriptome peaked at about 14 h after tetracycline addition. In total, 561 mRNAs were significantly ($\text{Padj} < 0.05$) increased by at least 2-fold, and 673 were significantly decreased by at least 2-fold. (These cut-offs are again arbitrary.) Effects at 16 h were similar (Pearson correlation coefficient of 0.93) but slightly less pronounced. In order to examine the point with the greatest variation between the tested conditions, we concentrated on the effects after 14 h. The strongest effects were on mRNAs encoding ribosomal proteins (Fig. 4C and D): 48% of them were more than 2-fold increased ($\text{FDR} 7 \times 10^{-21}$). This must be a secondary effect since the ribosomal protein mRNAs were not associated with ZC3H28 at all.

The >2x decreased mRNAs ($\text{Padj} < 0.05$) were enriched for the functions protein kinase ($\text{FDR} 5 \times 10^{-15}$), RNA binding ($\text{FDR} 5 \times 10^{-14}$), citric acid cycle ($\text{FDR} 4 \times 10^{-5}$) and amino acid transport ($\text{FDR} 1.7 \times 10^{-3}$) (Fig. 4C). The GO terms also reflected this (Supplementary Table S4). The effects were weakly ($R=0.37$) inversely correlated with mRNA length (Fig. 4D): in other words, there was a slight bias towards decreases in longer mRNAs. There was no overall correlation between the degree of ZC3H28 RNA binding and effects after RNAi (Fig. 4E), but 56 of the 180 mRNAs that were at least 3-fold enriched with TAP-ZC3H28 decreased after RNAi (Fisher test P value 6×10^{-22}) (Fig. 4E), whereas only two of the bound mRNAs were 2-fold significantly increased (Fisher test for under-representation: P value 3×10^{-6}). The median decrease for all the bound mRNAs was 1.65-fold. The median annotated length of the 56 enriched and decreased mRNAs was 4.8 kb and they

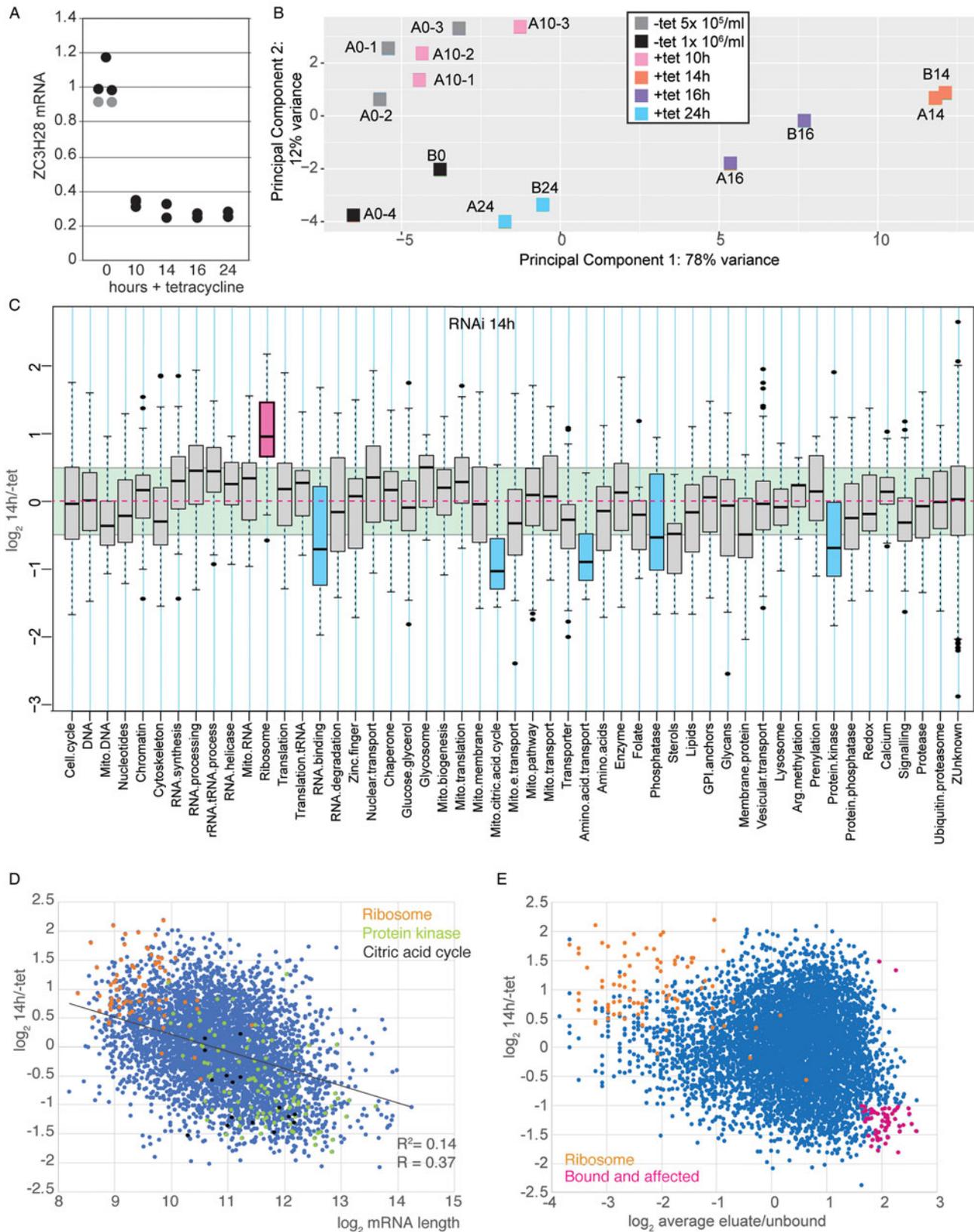


Fig. 4. Effects of ZC3H28 depletion on the transcriptome. (A) Amounts of ZC3H28 mRNA with and without tetracycline treatment. The amounts of RNA (reads per million reads from the coding region) were all compared with the arithmetic mean for untreated cells (time = 0). Results for the two control sets are colour-coded as in Fig. 4B. (B) Principal component analysis. Samples without induction are in black and grey; others are colour-coded according to the time with tetracycline. All results in this figure are for the set of unique genes, to avoid over-counting repeated genes. (C) The set of unique genes was placed in functional categories (Supplementary Table S4) and the effects of ZC3H28 depletion (14 h RNAi) were plotted for each category. The green background shows the range between 25th and 75th percentiles for the total dataset and the magenta dotted line is the median. The category with a median above the 75th percentile is coloured in pink while those with medians below the 25th percentile are in blue. (D) The effect of RNAi (y-axis) was plotted against the annotated mRNA length (x-axis). Only mRNAs with two annotated untranslated regions were included. Correlation coefficients were calculated in Microsoft Excel. Ribosomal protein mRNAs are in orange, protein kinase mRNAs in green, and those encoding enzymes of the citric acid cycle in black. (E) Relationship between the ZC3H28 RNAi effect after 14 h (y-axis) and the ZC3H28 mRNA binding (x-axis). Ribosomal protein mRNAs are in orange. The magenta spots are mRNAs that were at least 3-fold bound in all replicates, and 2-fold significantly increased or decreased after 14 h RNAi. (The two increased mRNAs do come from single-copy genes; this isn't an artefact due to paralogs having different behaviours.)

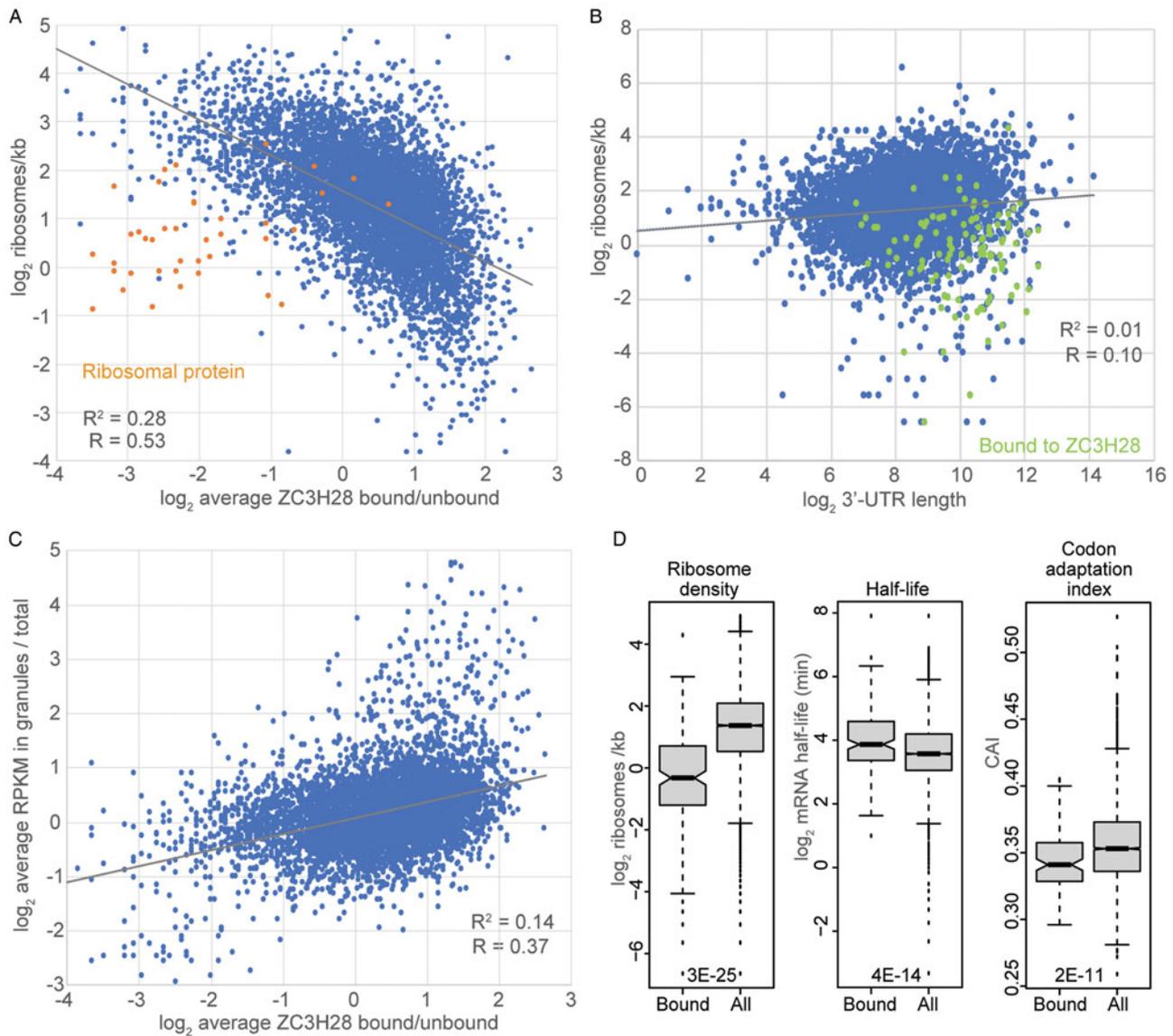


Fig. 5. ZC3H28 and translation. In A–C, each dot represents a single mRNA coding region. (A) Binding of the mRNA to ZC3H28 (data from Supplementary Table S2) is on the x-axis. The y-axis shows the number of ribosomes per kilobase of coding region, taken from Antwi *et al.* (2016). Ribosomal protein mRNAs are highlighted in orange. (B) The \log_2 of annotated 3'-UTR length is on the x-axis and the ribosome density is on the y-axis. mRNAs that were at least 3-fold enriched with ZC3H28 in all 3 replicate pull-downs are highlighted in green. (C) Binding of mRNAs to ZC3H28 is on the x-axis (Supplementary Table S2). (Note that these results, like all others in this paper, are from bloodstream forms.) The y-axis shows the proportion of the mRNA that is found in granules after starvation of procyclic forms (Fritz *et al.*, 2015). (D) Box plots showing properties of the mRNAs that were enriched with ZC3H28: ribosome density (as in A); half-lives (Fadda *et al.*, 2014); and the codon adaptation index (de Freitas Nascimento *et al.*, 2018). Student t -test results for the log-transformed data are shown on each plot.

included mRNAs encoding 7 protein kinases and 5 RNA-binding proteins. Nevertheless the majority of mRNAs with decreased abundances showed less, or no enrichment with ZC3H28, suggesting secondary effects.

We now looked for correlations with other published datasets. The bound mRNAs are mostly not cell-cycle regulated (Archer *et al.*, 2011). Binding and RNAi effects also did not correlate with developmental regulation of either mRNA level (Fadda *et al.*, 2014) or translation (Antwi *et al.*, 2016). ZC3H28-bound mRNAs were very slightly more stable than average (Fig. 5D). Although ZC3H28 protein was not enriched in starvation granules (Fritz *et al.*, 2015), there was a correlation between ZC3H28 binding to mRNAs and their granule association (Fig. 5C, Supplementary Table S4, sheet 4); here length could be a factor since enrichment of mRNAs in granules increases with mRNA length (Supplementary Table S4, sheet 4). The densities of ribosomes on the coding region are a surrogate for translation efficiency (though this ignores elongation kinetics). There

was no correlation between the effects of RNAi (Supplementary Table S4, sheet 4) and ribosome densities (Antwi *et al.*, 2016) but ZC3H28 clearly preferentially bound mRNAs with low densities (Supplementary Table S4, sheet 4 and Fig. 5A and D). The ribosome density correlation was not just caused by the lengths of the bound mRNAs, since there is no correlation between ribosome density on the coding region and either mRNA length or 3'-UTR length (Supplementary Table S4, sheet 4 and Fig. 5B).

Discussion

We have here confirmed that ZC3H28 is essential for normal growth of bloodstream-form trypanosomes, and that it increases mRNA abundance when tethered to a reporter mRNA. After purification of tagged ZC3H28, we saw co-purification of many ribosomal proteins and other proteins linked to cytoplasmic mRNAs. We also found that ZC3H28 binds preferentially to longer mRNAs. A

significant subset of the mRNAs that were bound by ZC3H28 decreased in abundance after ZC3H28 depletion.

We found that ZC3H28-bound mRNAs tended to be relatively long, especially in the 3'-UTR, with long low-complexity sequences (Fig. 3). However, there was no evidence for specific binding to a short sequence motif. This is probably because ZC3H28 has only a single zinc finger domain. A crystal structure of the two CCCH domains of the mammalian protein TIS11 showed that each CCCH domain could bind specifically to four nucleotides (Hudson *et al.*, 2004), and that the region 5' to the zinc finger contributed to specificity. With only a single zinc finger, not much specificity would be expected from ZC3H28 unless it acts as a multimer – though that might be the case, based on its self-interaction. The results for ZC3H28 mRNA binding are reminiscent of those from several other proteins: ZC3H30, DRBD7, ZC3H5, RBP9, UBP1 and RRM1 all select mRNAs that are significantly longer than average (Erben *et al.*, 2021). However, there is still specificity within this subset of proteins, since ZC3H30, DRBD7, ZC3H5 and RBP9 select mRNAs with long coding regions whereas UBP1 and RRM1 select mRNAs with long 3'-UTRs; and RRM1 selects mRNAs with long 5'-UTRs as well (Erben *et al.*, 2021). The preference for long RNAs is not an artefact of the method since other proteins analysed in exactly the same way showed preferences for short RNAs or for highly specific sequences (Erben *et al.*, 2021). We previously argued that the selection of long UTRs might be indicative that an RNA-binding protein preferentially binds AU-rich sequences (Erben *et al.*, 2021), which is exactly what was seen for ZC3H28. The length effects probably explain the preference of ZC3H28 for mRNAs encoding proteins of some functional classes: for example, mRNAs encoding protein kinases and RNA-binding proteins have longer-than-average 3'-UTRs (Clayton, 2019). Comparison of the RNA-binding data available for other proteins with those for ZC3H28 showed that the strongest correlation was with DRBD7 (Supplementary Table S2), which is somewhat surprising given the preference of DRBD7 for long coding regions (Erben *et al.*, 2021). None of the proteins that listed above co-purified with ZC3H28. Thus, the overlaps in RNA-binding specificity between these proteins did not suffice for the detection of co-purification at the protein level. This pattern of partial overlap between RNA-binding proteins in their RNA binding preference is not exceptional: results from a high-throughput survey of human binding proteins indicated that many bound similar – or even the same – motifs; probably, some specificity is conferred by additional flanking sequences or by secondary structures (Dominguez *et al.*, 2018).

On average, Opisthokont mRNAs are bound by 4–18 proteins per kilobase, with a rather higher density in untranslated regions than in coding regions (Khong and Parker, 2020). Some proteins actively influence mRNA turnover or translation, while others serve to shield the mRNA from binding of active regulators. The behaviours of most mRNAs are therefore the consequence of combinatorial effects. mRNAs that show a simple correlation between binding of a single protein, and either abundance or translation, are likely to be the exception. For example, RBP10 causes degradation of bound mRNAs, but of the 260 that were reproducibly bound, only 18 were at least 2-fold increased after RBP10 depletion (Mugo and Clayton, 2017). The longer the untranslated region is, the more redundancy is likely to be present (Bishola Tshitenge *et al.*, 2021). It is therefore quite surprising that of the 180 mRNAs that were at least 3-fold enriched with ZC3H28, 57 were also 2-fold significantly decreased after 14 h ZC3H28 RNAi. This is consistent with the stabilising function of ZC3H28 that was shown in the tethering assay.

EIF4E/EIF4G complexes bind to the mRNA cap and are critical for most translation initiation (Shirokikh and Preiss, 2018). Via

general translation factors, they recruit the small ribosomal subunit with charged initiator tRNA, initiating scanning of the 5'-UTR towards the start codon. Trypanosomes have six isoforms of EIF4E and five versions of EIF4G (Freire *et al.*, 2017). The EIF4E3/EIF4G4 complex is probably a major general initiation factor. EIF4E4/EIF4G3 is essential for normal bloodstream-form trypanosome growth and is also probably active in translation, but its precise role is unknown (Freire *et al.*, 2017). In contrast, the MKT1 complex recruits mainly EIF4E6/EIF4G5 (Nascimento *et al.*, 2020). Examination of the proteins associated with ZC3H28 revealed specific co-purification of EIF4E4/EIF4G3, despite the fact that MKT1 was also strongly enriched. There was no enrichment of any other general translation factors apart from a termination factor.

It is intriguing that despite its ability to increase mRNA abundance, ZC3H28 was preferentially associated with mRNAs that have low ribosome occupancy in their coding regions – a median of only 0.8 per Kb compared with the population median of 2.8 (Fig. 5D), which means that, given a mean coding region length of 1.6 kb (Berriman *et al.*, 2005), many of the bound mRNAs are not being translated at all. Since this is not a general characteristic of mRNAs with long lengths or 3'-UTRs, the link to ZC3H28 appears to be specific. Low ribosome densities can arise for two reasons: either a low rate of translation initiation, or an average rate of initiation combined with very rapid translation elongation. The mRNAs that are bound to ZC3H28 have somewhat more rare codons than normal (Fig. 5D) (de Freitas Nascimento *et al.*, 2018), which argues for poor initiation rather than rapid translation elongation. The picture that emerges, therefore, is that the role of ZC3H28 is to preserve long, poorly translated mRNAs.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S003118202100189X>

Acknowledgements. We thank Claudia Helbig and Ute Leibfried for technical assistance, David Ibberson of the Bioquant sequencing facility (University of Heidelberg) for cDNA library construction and RNA sequencing. Mass spectrometry was done in the ZMBH Core Facility by Thomas Ruppert and Sabine Merker. 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 including equipment and reagents after the flood in the ZMBH. We acknowledge the support of Andrea Zanotti (AG Lemberg) for the assistance in the pulse labelling experiment.

Author contribution. T.B. performed all the experiments. T.B. and C.C. were both involved in conceptualization, methodology, data curation, formal analysis, validation, investigation and visualization, writing-original draft, review and editing. C. C. was responsible for supervision, funding acquisition and project administration.

Financial support. This work was partially funded by Deutsche Forschungsgemeinschaft grant number Cl112/28-1 to CC, and by core support from the state of Baden-Württemberg.

Conflict of interest. None.

Ethical standards. Not applicable.

Data. The transcriptomes are available at Array express with accession numbers E-MTAB-10674 (ZC3H28-associated RNA) and E-MTAB-10751 (effect of RNAi). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol *et al.*, 2019) with the dataset identifier PXD027792.

References

Alibu VP, Storm L, Haile S, Clayton C and Horn D (2005) A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **139**, 75–82.

- Alsford S, Turner D, Obado S, Sanchez-Flores A, Glover L, Berriman M, Hertz-Fowler C and Horn D (2011) High throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Research* **21**, 915–924.
- Antwi E, Haanstra J, Ramasamy G, Jensen B, Droll D, Rojas F, Minia I, Terrao M, Mercé C, Matthews K, Myler P, Parsons M and Clayton C (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.
- Archer S, Inchaustegui D, de Queiroz R and Clayton C (2011) The cell-cycle regulated transcriptome of an early-branching eukaryote. *PLoS ONE* **6**, e18425.
- Bailey T (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics (Oxford, England)* **27**, 1653–1659.
- Bajak K, Leiss K, Clayton C and Erben E (2020) A potential role for a novel ZC3H5 complex in regulating mRNA translation in *Trypanosoma brucei*. *Journal of Biological Chemistry* **295**, 14291–14304.
- Banuelos CP, Levy GV, Nittolo AG, Roser LG, Tekiel V and Sanchez DO (2019) The *Trypanosoma brucei* RNA-binding protein TbRRM1 is involved in the transcription of a subset of RNA Pol II-dependent genes. *Journal of Eukaryotic Microbiology* **66**, 719–729.
- Benz C and Urbaniak M (2019). Organising the cell cycle in the absence of transcriptional control: dynamic phosphorylation co-ordinates the *Trypanosoma brucei* cell cycle post-transcriptionally. *PLOS Path.* doi: 10.1371/journal.ppat.1008129.
- Benz C, Mulindwa J, Ouna B and Clayton C (2011) The *Trypanosoma brucei* zinc finger protein ZC3H18 is involved in differentiation. *Molecular and Biochemical Parasitology* **177**, 148–151.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renaud H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mullard K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabinowitz E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE and El-Sayed NM (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science (New York, N.Y.)* **309**, 416–422.
- Bishola Tshitenge T, Liu B and Clayton C (2021) The RNA-binding protein DRBD18 regulates processing and export of the mRNA encoding *Trypanosoma brucei* RNA-binding protein 10. *bioRxiv*, BIORXIV/2021/460056.
- Bochler A, Querido JB, Prilepskaja T, Soufari H, Simonetti A, Del Cistia ML, Kuhn L, Ribeiro AR, Valášek LS and Hashem Y (2020) Structural differences in translation initiation between pathogenic trypanosomatids and their mammalian hosts. *Cell Reports* **33**, 108534.
- Chakraborty C and Clayton C (2018) Stress susceptibility in *Trypanosoma brucei* lacking the RNA-binding protein ZC3H30. *PLoS Neglected Tropical Diseases* **12**, e0006835.
- Clayton CE (1999) Genetic manipulation of Kinetoplastida. *Parasitology Today* **15**, 372–378.
- Clayton C (2019) Control of gene expression in trypanosomatids: living with polycistronic transcription. *Royal Society Open Biology* **9**, 190072.
- Das A, Morales R, Banday M, Garcia S, Hao L, Cross GA, Estevez AM and Bellofatto V (2012) The essential polysome-associated RNA-binding protein RBP42 targets mRNAs involved in *Trypanosoma brucei* energy metabolism. *Rna* **18**, 1968–1983.
- Dean S, Sunter J and Wheeler R (2016) TrypTag.org: a trypanosome genome-wide protein localisation resource. *Trends in Parasitology* **33**, 80–82.
- de Freitas Nascimento J, Kelly S, Sunter J and Carrington M (2018) Codon choice directs constitutive mRNA levels in trypanosomes. *Elife* **7**, e32467.
- Dejung M, Subota I, Bucarius F, Dindar G, Freiwald A, Engstler M, Boshart M, Butter F and Janzen CJ (2016) Quantitative proteomics uncovers novel factors involved in developmental differentiation of *Trypanosoma brucei*. *PLoS Pathogens* **12**, e1005439.
- Demogines A, Smith E, Kruglyak L and Alani E (2008) Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. *PLoS Genetics* **4**, e1000123.
- Doleželová E, Kunzová M, Dejung M, Levin M, Panicucci B, Regnault C, Janzen CJ, Barrett MP, Butter F and Zíková A (2020) Cell-based and multi-omics profiling reveals dynamic metabolic repurposing of mitochondria to drive developmental progression of *Trypanosoma brucei*. *PLoS Biology* **18**, e3000741.
- Dominguez D, Freese P, Alexis MS, Su A, Hochman M, Palden T, Bazile C, Lambert NJ, Van Nostrand EL, Pratt GA, Yeo GW, Graveley BR and Burge CB (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Molecular Cell* **70**, 854–867.e859.
- Droll D, Minia I, Fadda A, Singh A, Stewart M, Queiroz R and Clayton C (2013) Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Pathogens* **9**, e1003286.
- Erben E, Fadda A, Lueong S, Hoheisel J and Clayton C (2014) Genome-wide discovery of post-transcriptional regulators in *Trypanosoma brucei*. *PLoS Pathogens* **10**, e1004178.
- Erben E, Leiss K, Liu B, Inchaustegui Gil D, Helbig C and Clayton C (2021) Insights into the functions and RNA binding of *Trypanosoma brucei* ZC3H22, RBP9 and DRBD7. *Parasitology* **148**, 1186–1195.
- Estévez A, Kempf T and Clayton CE (2001) The exosome of *Trypanosoma brucei*. *EMBO Journal* **20**, 3831–3839.
- Fadda A, Färber V, Droll D and Clayton C (2013) The roles of 3'-exoribonucleases and the exosome in trypanosome mRNA degradation. *Rna* **19**, 937–947.
- Fadda A, Ryten M, Droll D, Rojas F, Färber V, Haanstra J, Bakker B, Matthews K and 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. *Molecular Microbiology* **94**, 307–326.
- Freire ER, Sturm NR, Campbell DA and de Melo Neto OP (2017) The role of cytoplasmic mRNA cap-binding protein complexes in *Trypanosoma brucei* and other trypanosomatids. *Pathogens (Basel, Switzerland)* **6**, 55.
- Fritz M, Vanselow J, Sauer N, Lamer S, Goos C, Siegel T, Subota I, Schlosser A, Carrington M and Kramer S (2015) Novel insights into RNP granules by employing the trypanosome's microtubule skeleton as a molecular sieve. *Nucleic Acids Research* **43**, 8013–8032.
- Gupta SK, Kosti I, Plaut G, Pivko A, Tkacz ID, Cohen-Chalamish S, Biswas DK, Wachtel C, Waldman Ben-Asher H, Carmi S, Glaser F, Mandel-Gutfreund Y and Michaeli S (2013) 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 Research* **41**, 6577–6594.
- Gupta SK, Chikne V, Eliaz D, Tkacz ID, Naboishchikov I, Carmi S, Waldman Ben-Asher H and Michaeli S (2014) Two splicing factors carrying serine-arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA stability, and rRNA processing in *Trypanosoma brucei*. *RNA Biology* **11**, 715–731.
- Guzikowski AR, Chen YS and Zid BM (2019) Stress-induced mRNP granules: form and function of processing bodies and stress granules. *Wiley Interdisciplinary Reviews. RNA* **10**, e1524.
- Hendriks EF and Matthews KR (2005) Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. *Molecular Microbiology* **57**, 706–716.
- Hudson B, Martinez-Yamout M, Dyson H and Wright P (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nature Structural & Molecular Biology* **11**, 257–264.
- Jaremko D, Ciganda M, Christen L and Williams N (2019) *Trypanosoma brucei* L11 is essential to ribosome biogenesis and interacts with the Kinetoplastid-specific proteins P34 and P37. *mSphere* **4**, e00475-19.
- Jenner L, Melnikov S, Garreau de Loubresse N, Ben-Shem A, Iskakova M, Urzhumtsev A, Meskauskas A, Dinman J, Yusupova G and Yusupov M (2012) Crystal structure of the 80S yeast ribosome. *Current Opinion in Structural Biology* **22**, 759–767.
- Jha B, Fadda A, Merce C, Mugo E, Droll D and Clayton C (2014) Depletion of the trypanosome pumilio domain protein PUF2 causes transcriptome changes related to coding region length. *Eukaryotic Cell* **13**, 664–674.
- Khong A and Parker R (2020) The landscape of eukaryotic mRNPs. *Rna* **26**, 229–239.

- Klein C, Terrao M and Clayton C (2017) The role of the zinc finger protein ZC3H32 in bloodstream-form *Trypanosoma brucei*. *PLoS ONE* **12**, e0177901.
- Kramer S, Kimblin N and Carrington M (2010) Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*. *BMC Genomics* **11**, 283.
- Langmead B and Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359.
- Leiss K and Clayton C (2016) DESeqUI – trypanosome RNAseq analysis made easy. *Zenodo*. doi: 10.5281/zenodo.165132
- Leiss K, Merce C, Muchunga E and Clayton C (2016) TrypRNAseq – a easy to use pipeline for *Trypanosoma* RNAseq data. *Zenodo*. <http://doi.org/10.5281/zenodo.158920>.
- Levy GV, Banuelos CP, Nittolo AG, Ortiz GE, Mendiondo N, Moretti G, Tekiel VS and Sanchez DO (2015) Depletion of the SR-related protein TbRRM1 leads to cell cycle arrest and apoptosis-like death in *Trypanosoma brucei*. *PLoS ONE* **10**, e0136070.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R (2009) The sequence alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079.
- Ling A, Trotter J and Hendriks E (2011) A zinc finger protein, TbZC3H20, stabilises two developmentally regulated mRNAs in trypanosomes. *Journal of Biological Chemistry* **286**, 20152–20162.
- Liu B, Marucha K and Clayton C (2020) The zinc finger proteins ZC3H20 and ZC3H21 stabilise mRNAs encoding membrane proteins and mitochondrial proteins in insect-form *Trypanosoma brucei*. *Molecular Microbiology* **113**, 430–451.
- Love M, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biology* **15**, 550.
- Lueong S, Merce C, Fischer B, Hoheisel J and Erben E (2016) Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome. *Molecular Microbiology* **100**, 457–471.
- Luu VD, Brems S, Hoheisel J, Burchmore R, Guilbride D and Clayton C (2006) Functional analysis of *Trypanosoma brucei* PUF1. *Molecular and Biochemical Parasitology* **150**, 340–349.
- Manful T, Fadda A and Clayton C (2011) The role of the 5′-3′ exoribonuclease XRNA in transcriptome-wide mRNA degradation. *Rna* **17**, 2039–2047.
- Mangus D, Amrani N and Jacobson A (1998) Pbp1p, a factor interacting with *Saccharomyces cerevisiae* poly(A)-binding protein, regulates polyadenylation. *Molecular and Cellular Biology* **18**, 7383–7396.
- Mani J, Güttinger A, Schimanski B, Heller M, Acosta-Serrano A, Pescher P, Späth G and Sabel Roditi I (2011) Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS ONE* **6**, e22463.
- Martin M (2011) Next generation sequencing data analysis: Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**, 1.
- Melo do Nascimento L, Egler F, Arnold K, Papavasilou N, Clayton C and Erben E (2021) Functional insights from a surface antigen mRNA-bound proteome. *Elife* **10**, e68136.
- Minia I and Clayton C (2016) Regulating a post-transcriptional regulator: protein phosphorylation, degradation and translational blockage in control of the trypanosome stress-response RNA-binding protein ZC3H11. *PLoS Pathogens* **12**, e1005514.
- Minia I, Merce C, Terrao M and 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 Neglected Tropical Diseases* **10**, e0004982.
- Morking PA, Rampazzo Rde C, Walrad P, Probst CM, Soares MJ, Gradia DF, Pavoni DP, Krieger MA, Matthews K, Goldenberg S, Fragoso SP and Dallagiovanna B (2012) The zinc finger protein TcZFP2 binds target mRNAs enriched during *Trypanosoma cruzi* metacyclogenesis. *Memorias do Instituto Oswaldo Cruz* **107**, 790–799.
- Mugo E and Clayton C (2017) Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*. *PLoS Pathogens* **13**, e1006560.
- Mulindwa J, Ssentamu G, Matovu E, Kamanyi Marucha K, Aresta-Branco F, Helbig C and Clayton C (2021) *In vitro* culture of freshly isolated *Trypanosoma brucei* bloodstream forms results in gene copy-number changes. *PLoS Neglected Tropical Diseases* **15**, e0009738.
- Müller L, Cosentino R, Förstner K, Guizetti J, Wedel C, Kaplan N, Janzen C, Arampatzis P, Vogel J, Steinbiss S, Otto T, Saliba A-E, Sebra R and Siegel T (2018) Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**, 121–125.
- Naguleswaran A, Gunasekera K, Schimanski B, Heller M, Hemphill A, Ochsenreiter T and Roditi I (2015) *Trypanosoma brucei* RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure. *MBio* **6**, e00114.
- Nascimento L, Terrao M, Marucha K, Liu B, Egler F, Helbig C and Clayton C (2020) The RNA-associated proteins MKT1 and MKT1L form alternative PBP1-containing complexes in *Trypanosoma brucei*. *Journal of Biological Chemistry* **295**, 10940–10955.
- Ooi CP, Benz C and Urbaniak MD (2020) Phosphoproteomic analysis of mammalian infective *Trypanosoma brucei* subjected to heat shock suggests atypical mechanisms for thermotolerance. *Journal of Proteomics* **219**, 103735.
- Ouna B, Stewart M, Helbig C and Clayton C (2012) The *Trypanosoma brucei* CCCH zinc finger proteins ZC3H12 and ZC3H13. *Molecular and Biochemical Parasitology* **183**, 184–188.
- Paterou A, Walrad P, Craddy P, Fenn K and Matthews K (2006) Identification and stage-specific association with the translational apparatus of TbZFP3, a ccch protein that promotes trypanosome life cycle development. *Journal of Biological Chemistry* **281**, 39002–39013.
- Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu D, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak A, Ternent T, Brazma A and Vizcaino J (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* **47**, D442–D450.
- Quintana JF, Zoltner M and Field MC (2020) Evolving differentiation in African trypanosomes. *Trends in Parasitology* **37**, 296–303.
- Redmond S, Vadivelu J and Field MC (2003) RNAi: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **128**, 115–118.
- Shi H, Djikeng A, Tschudi C and Ullu E (2004a) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Molecular and Cellular Biology* **24**, 420–427.
- Shi H, Ullu E and Tschudi C (2004b) Function of the Trypanosome Argonaute 1 protein in RNA interference requires the N-terminal RGG domain and arginine 735 in the Piwi domain. *Journal of Biological Chemistry* **279**, 49889–49893.
- Shirokikh NE and Preiss T (2018) Translation initiation by cap-dependent ribosome recruitment: recent insights and open questions. *Wiley Interdisciplinary Reviews. RNA* **9**, e1473.
- Siegel T, Hekstra D, Wang X, Dewell S and 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 Research* **38**, 4946–4957.
- Silvester E, Ivens A and 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 Neglected Tropical Diseases* **12**, e0006863.
- Singh A, Minia I, Droll D, Fadda A, Clayton C and Erben E (2014) Trypanosome MKT1 and the RNA-binding protein ZC3H11: interactions and potential roles in post-transcriptional regulatory networks. *Nucleic Acids Research* **42**, 4652–4668.
- Sinha H, Nicholson B, Steinmetz L and McCusker J (2006) Complex genetic interactions in a quantitative trait locus. *PLoS Genetics* **2**, e13.
- Tadauchi T, Inada T, Matsumoto K and Irie K (2004) Posttranscriptional regulation of HO expression by the Mkt1-Pbp1 complex. *Molecular and Cellular Biology* **24**, 3670–3681.
- Toh JY, Nkouawa A, Sánchez SR, Shi H, Kolev NG and Tschudi C (2021) Identification of positive and negative regulators in the stepwise developmental progression towards infectivity in *Trypanosoma brucei*. *Scientific Reports* **11**, 5755.
- Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M and Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731–740.
- Vigneron A, O'Neill MB, Weiss BL, Savage AF, Campbell OC, Kamhawi S, Valenzuela JG and Aksoy S (2020) Single-cell RNA sequencing of *Trypanosoma brucei* from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens. *Proceedings of the National Academy of Sciences of the USA* **117**, 2613–2621.

- Walrad P, Paterou A, Acosta-Serrano A and Matthews K** (2009) Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. *PLoS Pathogens* **5**, e1000317.
- Wippel HH, Malgarin JS, Inoue AH, Leprevost FDV, Carvalho PC, Goldenberg S and Alves LR** (2019) Unveiling the partners of the DRBD2-mRNP complex, an RBP in *Trypanosoma cruzi* and ortholog to the yeast SR-protein Gbp2. *BMC Microbiology* **19**, 128.
- Zoltner M, Krienitz N, Field MC and Kramer S** (2018) Comparative proteomics of the two *T. brucei* PABPs suggests that PABP2 controls bulk mRNA. *PLoS Neglected Tropical Diseases* **12**, e0006679.