Molecular Microbiology (2016) 100(3), 457–471

doi:10.1111/mmi.13328
First published online 10 March 2016

Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome

Smiths Lueong,[1] Clementine Merce,[2†] Bernd Fischer,[3] Jörg D. Hoheisel[1] and Esteban D. Erben[2*]

[1]*Functional Genome Analysis, Deutsche Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.*

[2]*Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 282.*

[3]*Computational Genome Biology, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg.*

Summary

Control of gene expression at the post-transcriptional level is essential in all organisms, and RNA-binding proteins play critical roles from mRNA synthesis to decay. To fully understand this process, it is necessary to identify the complete set of RNA-binding proteins and the functional consequences of the protein-mRNA interactions. Here, we provide an overview of the proteins that bind to mRNAs and their functions in the pathogenic bloodstream form of *Trypanosoma brucei*. We describe the production of a small collection of open-reading frames encoding proteins potentially involved in mRNA metabolism. With this ORFeome collection, we used tethering to screen for proteins that play a role in post-transcriptional control. A yeast two-hybrid screen showed that several of the discovered repressors interact with components of the CAF1/NOT1 deadenylation complex. To identify the RNA-binding proteins, we obtained the mRNA-bound proteome. We identified 155 high-confidence candidates, including many not previously annotated as RNA-binding proteins. Twenty seven of these proteins affected reporter expression in the tethering screen. Our study provides novel insights into the potential trypanosome mRNPs composition, architecture and function.

Accepted 14 January, 2016. *For correspondence. E-mail e.erben@zmbh.uni-heidelberg.de; Tel. 496221546861; Fax 496221545891. †Present address: Molecular Genome Analysis, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 460, D69120, Heidelberg, Germany.

© 2016 John Wiley & Sons Ltd

Introduction

RNA-binding proteins (RBPs) play an important role in controlling the life of mRNAs. They bind to nascent, mature and decaying mRNAs, packaging them into ribonucleoprotein particles (mRNPs) (Glisovic *et al.*, 2008). The coordinated temporal and spatial control of gene expression is in part determined by both the repertoire of RBPs and their activities, influencing transcript stability and/or translational efficiency. Since RBPs participate in these essential cellular processes, it is not surprising that mutations disrupt either the RNA or protein components of mRNPs can cause disease and be deleterious for life (Ramaswami *et al.*, 2013). Historically, annotation of RBPs was limited to proteins with known RNA-binding domains (RBDs). However, recent system-wide approaches allowed the identification of hundreds of new RBPs in both yeast (Scherrer *et al.*, 2010; Tsvetanova *et al.*, 2010; Klass *et al.*, 2013; Mitchell *et al.*, 2013) and mammalian cells (Baltz *et al.*, 2012; Castello *et al.*, 2012; Kwon *et al.*, 2013) confirming the existence of a plethora of non-classical RBDs. Although several genetic human diseases have been linked to mutations in genes coding for RBPs (Castello *et al.*, 2013a), the biological relevance for most of these newly discovered RBPs is largely unknown.

Kinetoplastid protists are exposed to environmental challenges in the host and vector, demanding fast and large changes in gene expression. Trypanosomes and related parasites are especially unusual in that the majority of its genes are regulated post-transcriptionally. Open reading frames are transcribed in long polycistronic arrays and primary transcripts are created by 5′ *trans* splicing and 3′ polyadenylation (Michaeli, 2011). However, the final output of mature mRNAs can greatly vary even between flanking genes. Additionally, mRNA and protein abundances of the same gene can widely differ between various

developmental stages (Queiroz *et al.*, 2009; Vasquez *et al.*, 2014). These developmental steps require coordinated modulation of gene expression, and RPB are the key determinants (Clayton, 2013). In *Trypanosoma brucei*, it has been shown that RBPs are involved in diverse molecular processes such as differentiation (Hendriks *et al.*, 2001; Hendriks and Matthews, 2005), development (Mani *et al.*, 2011; Subota *et al.*, 2011; Kolev *et al.*, 2012; Wurst *et al.*, 2012), the cell cycle (Archer *et al.*, 2009), rRNA processing (Droll *et al.*, 2010) and heat shock response (Droll *et al.*, 2013; Singh *et al.*, 2014). The trypanosome genome encodes over 150 proteins with conserved RBDs; however, it is highly likely that additional proteins also influence mRNA fate. In a genome-wide tethering screen, in which random protein fragments were artificially bound to a reporter mRNA, we recently identified over 300 proteins possibly regulating mRNA amounts or translation by post-transcriptional mechanisms (Erben *et al.*, 2014b). Although our findings expanded the repertoire of known proteins influencing the mRNA-fate, the results obtained with protein fragments may not be true for full-length proteins. For example, the use of fragments can result in defective protein folding or inactive proteins giving false negative hits. Several of these disadvantages could be abrogated by using a library of full-length open reading frames (ORFs).

Here, we describe the production of a small-scale ORF library encoded in a versatile Gateway vector system containing the most promising potential regulators. Using this 'mini-ORFeome' resource in a tethering screen, we established a functional catalogue of the *T. brucei* mRNA-fate regulators. This list of about 100 proteins contains not only canonical RBPs but also proteins without RNA-related ontology. Using the ORFeome resource, we also show the ability of multiple repressors to interact with the deadenylation CAF1/NOT complex. To determine which of these regulators are indeed bound to mRNAs *in vivo*, we captured poly(A) mRNA-bound proteins from bloodstream forms. Our mRNA-bound proteome contained 155 proteins, including many not previously annotated as RBPs. Twenty seven of these RPBs showed a reproducible effect on reporter expression in the tethering screen. Taken together, we provide a landscape of the trypanosome mRNPs by identifying not only the major RBPs of bloodstream parasites but also the potential effect on their bound mRNA targets.

Results and discussion

A small ORFeome collection for discovery of new mRNA-fate modulators

To screen for proteins that regulate gene expression, we created a small ORFeome. This collection contains 383 proteins (384 ORFs) that are of interest in post-transcriptional regulation, including all proteins with identified RBDs, some translation factors, enzymes involved in mRNA degradation and a subset of proteins without known RNA-binding features that were previously identified in our tethering screen (Erben *et al.*, 2014b) (Supporting Information Table S1). To allow for a systematic ORF cloning, we used the Gateway technology, which is based on cloning by site-specific recombination (Walhout *et al.*, 2000). For cloning, ORFs were amplified in 96-well format by two-step PCR using a high fidelity polymerase (Supporting Information Fig. S1A). After second PCR, the amplification products were grouped into three categories: small, medium and large sizes and each group were cloned separately into a Gateway donor vector. To create a tetracycline inducible over-expression library suitable for trypanosomes, a new library was created by pooling all plasmid preparations of entry clones and transferring them into a lambda-N vector. This plasmid has an N-terminal lambda peptide making it appropriate to perform tethering assay, in fusion with three copies of a myc tag facilitating detection (Supporting Information Fig. S1B). About 90% of randomly selected clones chosen for sequencing showed an in-frame insert (not shown). To determine the complexity of this new library, the ORFs were amplified by PCR only after transfection into trypanosomes (see below), using primers located within the lambda-N peptide and immediately 3' to the cloning site. Illumina sequencing of this PCR-amplified DNA showed that about 300 ORFs (~80%) were positively cloned (Supporting Information Table S2, Sheet 1). We considered successfully cloned to those genes displaying before selection, 10 or more reads counts. Failure in cloning can be explained by negative PCR, inefficient recombination or the presence of a NotI site within the ORFs since plasmid linearization with NotI before stable transfection is required. Transfection efficiency issues can be ruled out since more than fifty thousand clones were obtained in each experiment, vastly exceeding (>100-fold) the ORFeome complexity. Although our mini ORFeome collection does not contain all regulators, it provides a Gateway-compatible resource for proteins potentially involved in mRNA metabolism.

Screening for modulators of trypanosome mRNA fate

To screen for proteins that increase gene expression, the inducible library was transfected into bloodstream cells expressing the BLA-BoxB-ACT reporter. This mRNA reporter contains five copies of the boxB hairpin RNA sequence between the blasticidin (BLA) resistance cassette and the actin (ACT) 3'-UTR (Supporting Information Fig. 1B) (Erben *et al.*, 2014b). To perform the screen, we

induced N-peptide fusion protein expression for 24 h and then grew cells for four days under different blasticidin concentration as previously described (Erben *et al*., 2014b). A minor effect on overall population growth was seen under tetracycline induction, being slightly more pronounced with the 10-fold level of blasticidin (Supporting Information Fig. S2). As a control, we also ran a similar experiment with a cell line carrying the BLA reporter without boxB elements. After selection, the cloned ORFs were recovered by plasmid-specific PCR. Examination of small aliquots from the reactions by gel electrophoresis revealed similar patterns for all populations except the high blasticidin-treated cells. The PCR products from the control were not sequenced as growth was strongly inhibited ruling out unspecific effect on reporter expression (data not shown). We next transfected the cells with a reporter that selects for proteins that impair reporter expression upon tethering. In this cell line, the expression of the lethal phosphoglycerate kinase B (PGKB) is prevented only if the tethered protein represses its expression (Erben *et al*., 2014b) (Supporting Information Fig. 1B). Expression of both PGKB and the lambda-N fusions was induced with tetracycline and cells were grown for five days when the cloned ORFs were in each case recovered by plasmid-specific PCR. Examination by gel electrophoresis of small aliquots of these reactions revealed similar effects for the two replicates (Supporting Information Fig. S2). Then, all PCR mixtures were subjected to high-throughput sequencing.

Redefining the functional catalogue: comparison of the high- and mid-throughput screens

Sequencing reads from all experiments were mapped to the *T. brucei* reference after trimming of the non-trypanosome boundary sequences. Results are tabulated in Supporting Information Tables S2. As mentioned, before selection we obtained 10 or more read counts for about 300 out of 384 ORFs; however, under blasticidin pressure or induction of the lethal PGKB expression, only particular genes were strongly enriched. We have previously found about 300 proteins potentially implicated in post-transcriptional mRNA regulation. Although our screen was able to identify known regulators, it relied on random shotgun clones. Since protein fragments may be improperly folded and consequently have atypical interactions, we validated the results by measuring CAT activity of full-length clones. For this, trypanosomes that constitutively expressed an mRNA encoding the CAT reporter with 5 *boxB* elements (*CAT-B-ACT)* were transfected with inducible lambda-Nmyc fusion proteins. When we compared the functional catalogue collected with the new ORF collection with the CAT activities we observed a good agreement (Fig. 1A). Since the measurement methods are different, quantitative agreement was not expected. However, when we compared the relative activity values determined in the previous high-throughput screening with the new results, several exceptions were detected. Since our shotgun library carries inserts only up to 3 kb in size, lack in correlation can be expected for large proteins. In addition, since expression shotgun libraries have a natural bias against N-terminal ends, proteins requiring this region for proper function might not be selected. For example, full-length MKT1 strongly increases expression of the reporter (Fig. 1A), but both the N- and C-terminal ends are required for such activity (Singh *et al*., 2014). Considering that MKT1 ORF is also too long to be included in our shotgun library, no selection for MKT1 was apparent at all (Erben *et al*., 2014b). Several other hits previously found to up-regulate reporter expression were analyzed. While the high-throughput survey suggested Tb927.8.4200, Tb927.11.12730, cytidine deaminase, Tb927.10.15760 and Tb927.3.1810 to be up-regulators, both CAT assay and blasticidin ORFs selection did not reveal significant changes (Fig. 1A). The same is true for ZC3H45. In contrast to the shotgun results, it displayed a slight increase in both CAT activity and ORF read counts. Thus, this mid-throughput screen, where full-length proteins were analysed, provides a cleaner snapshot of *T. brucei* genes able to regulate the mRNA-fate.

The screen detects expected and novel mRNA-fate up-regulators

When the number of reads per CDS from the two replicates of the blasticidin treated experiments was compared with the untreated condition, we obtained a list of candidate activators, defined here as proteins showing a reproducible RPKM enrichment of at least 1.5-fold. Applying this criterion, we identified 44 putative up-regulating proteins (Supporting Information Tables S2, Sheet 4 and Table 1). This subset includes 15 canonical RBPs (9 Zinc finger-, 4 RRM- and 2 PUF-containing proteins), 9 translation factors and interestingly, 10 uncharacterised hypothetical proteins. Supporting our results, several of the identified 'activators' have been previously shown to increase the abundance or translation of their target mRNAs. This includes ZC3H11 (Droll *et al*., 2013; Singh *et al*., 2014), LSM12 (Singh *et al*., 2014) and the master regulator MKT1 (Singh *et al*., 2014). The RRM protein RBP42 was also found in this group; it is associated with polysomes and binds to abundant mRNAs (Das *et al*., 2012), which would be consistent with a stabilizing function. The new up regulating candidates include the zinc finger proteins ZC3H18, ZC3H44 and ZC3H45 and the RRM-containing proteins RBP7A and DRBD6A.

This list contains also the hypothetical protein Tb927.10.14150, a putative homolog of the brefeldin A

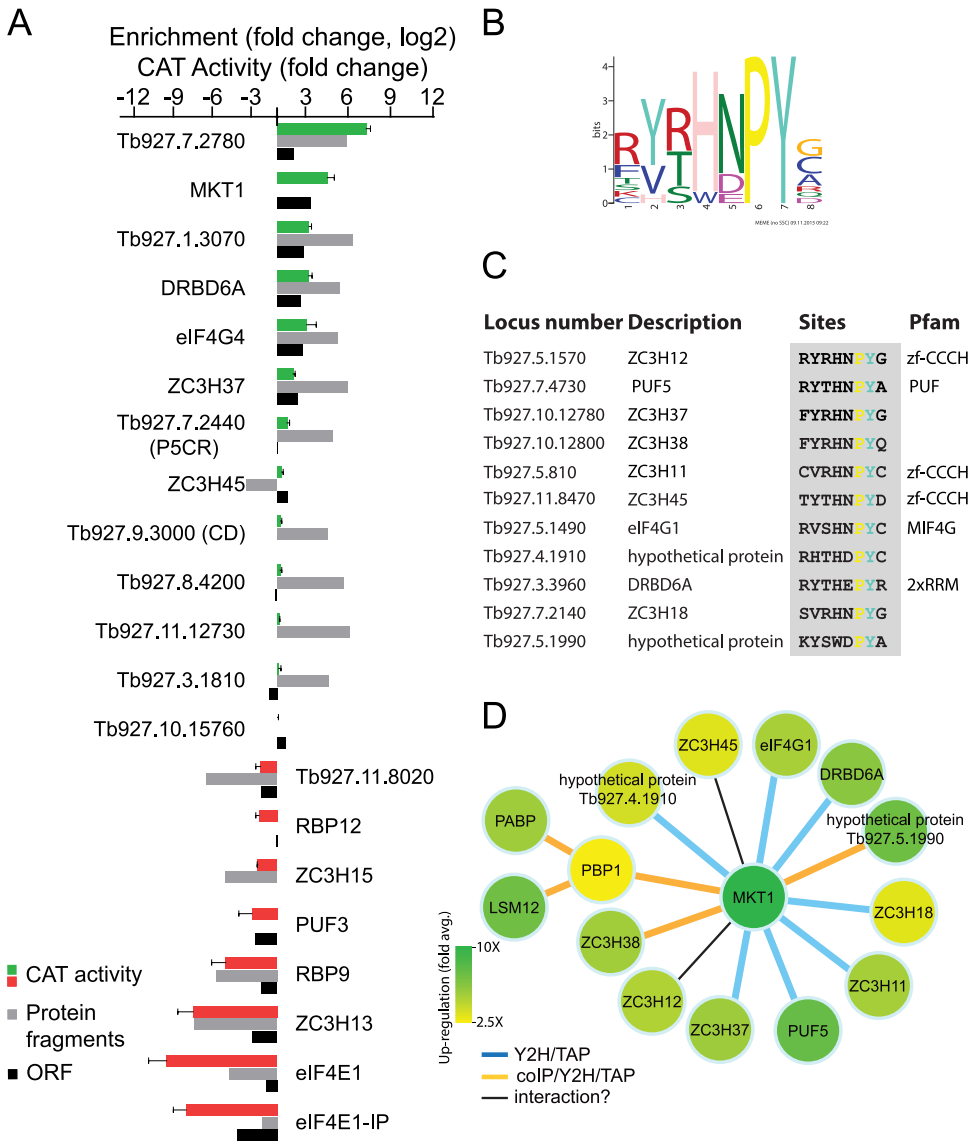

Fig. 1. A mid-throughput screening accelerating the discovery of new mRNA-fate modulators.
A. Activities comparison of selected mRNA-fate regulators: CAT activities (green and red for activators and repressors, respectively), protein fragment enrichment (*log2* values, grey) and full-length ORF enrichment (*log2* values, black) are shown. Data taken from (Erben *et al.*, 2014b) and from this work.
B. MEME analysis of the 44 up-regulating proteins identified an HNPY sequence motif to be the most highly enriched ($P = 4.0$ e -15).
C. This HNPY pattern is distributed in 11 putative RNA-binding proteins. Pfam detected domains are indicated.
D. Graphical representation of the MKT1 complex interactome. Most interaction partners of MKT1 carrying an HNPY sequence motif exhibit a positive effect on mRNA expression. Previously seen interactions are shown with blue or oranges edges (Singh *et al.*, 2014) while putative ones are denoted with black. Protein nodes are coloured according to the activation strength (average values) as judged by the tethering results (Supporting Information Table S2). Values for PABP1 and LSM12 were taken from (Singh *et al.*, 2014).

resistance protein (Bfr1p). In budding yeast, it associates with polysomes, preventing P-body formation under normal growth (Weidner *et al.*, 2014). Interestingly, several eIF4 initiation factors tethered to the 3'UTR can also increase gene expression. The trypanosome genome encodes multiple homologues for the eIF4A (two), eIF4E (six) and eIF4G (five) subunits (Dhalia *et al.*, 2006; Freire *et al.*, 2014; Moura *et al.*, 2015). However, only eIF4E3 and eIF4E4, which are in complex with eIF4G4 and eIF4G3, respectively, are thought to have a main role in the translation initiation complex (Zinoviev and Shapira, 2012). In agreement with these findings, we found that the four main players of this process (eIF4G3, eIF4G4, eIF4E3, eIF4E4) together with eIF4G1 are able to increase reporter expression as full-length proteins. The rest of the subunits were not up-regulating or possibly not

Table 1. Proteins that increase BLA resistance.

Locus number	Description	Category	References
Tb927.9.7110	hypothetical protein	Unknown	
Tb927.7.2160	hypothetical protein	Unknown	
Tb927.7.2980	hypothetical protein	Unknown	
Tb927.5.1990	hypothetical protein	Unknown	
Tb927.11.6010	hypothetical protein	Unknown	
Tb927.1.3070	hypothetical protein	Unknown	
Tb927.10.15310	hypothetical protein	Unknown	
Tb927.4.1910	hypothetical protein	Unknown	
Tb927.7.2780	hypothetical protein	Unknown	
Tb927.11.3440	hypothetical protein	Unknown	
Tb927.5.4320	FIP1	DNA	
Tb927.2.4930	esterase	Enzyme	
Tb927.7.5680	deoxyribose-phosphate aldolase	Enzyme	
Tb927.7.2440	pyrroline-5-carboxylate reductase (P5CR)	Mito pathway	
Tb927.7.2140	ZC3H18	RNA binding	(Benz *et al.*, 2011)
Tb927.11.7890	ZC3H44	RNA binding	
Tb927.11.8470	ZC3H45	RNA binding	
Tb927.10.12090	RBP7A	RNA binding	(Mony *et al.*, 2014)
Tb927.5.810	ZC3H11	RNA binding	(Droll *et al.*, 2013)
Tb927.10.12780	ZC3H37	RNA binding	(Singh *et al.*, 2014)
Tb927.5.1570	ZC3H12	RNA binding	(Ouna *et al.*, 2012)
Tb927.3.3960	DRBD6A	RNA binding	
Tb927.10.12800	ZC3H38	RNA binding	(Singh *et al.*, 2014)
Tb927.10.11270	RBP23	RNA binding	(Wurst *et al.*, 2009)
Tb927.10.12330	ZC3H34	RNA binding	
Tb927.7.4730	PUF5	RNA binding	(Jha *et al.*, 2013)
Tb927.10.11760	PUF6	RNA binding	
Tb927.6.4440	RBP42	RNA binding	(Das *et al.*, 2012)
Tb927.3.790	ZC3H6	RNA binding	
Tb927.9.9060	LSM12	RNA degradation	(Singh *et al.*, 2014)
Tb927.6.4770	MKT1	RNA degradation	(Singh *et al.*, 2014)
Tb927.11.6870	14-3-3 protein	Signalling	(Inoue *et al.*, 2005)
Tb927.11.9530	14-3-3-I protein	Signalling	(Inoue *et al.*, 2005)
Tb927.6.1870	eIF4E4	Translation	(Freire *et al.*, 2011)
Tb927.11.10560	eIF4G4	Translation	(Moura *et al.*, 2015)
Tb927.11.5840	SUI1	Translation	
Tb927.5.1490	eIF4G1	Translation	
Tb927.11.2300	ERF1	Translation	
Tb927.11.6160	ERF3	Translation	
Tb927.8.4820	eIF4G3	Translation	(Moura *et al.*, 2015)
Tb927.9.10770	PABP2	Translation	(Kramer *et al.*, 2013)
Tb927.11.11770	eIF4E3	Translation	(Freire *et al.*, 2011)
Tb927.11.6130	Skp1 family protein	Ubiquitin proteasome	
Tb927.10.14150	BFR1	Vesicular transport	

We consider activators to be proteins causing at least a 1.5-fold increase in reads per million in the BLA10x condition in both replicates. Related references are shown.

cloned (Supporting Information Table S2). In different organisms, PUF proteins display either activator or repressor functions and trypanosomes are no exception (Quenault *et al.*, 2011). In our screen, both PUF6 and PUF5 were able to increase expression of the BLA reporter. Previously, a stabilizing function for PUF9 was also shown (Archer *et al.*, 2009; Erben *et al.*, 2014b). The component of the ubiquitin ligase complex SKP1 also constitutes this group. Interestingly, this protein is pulled down with the hub regulator MKT1, suggesting an active role in gene expression regulation (Singh *et al.*, 2014). In mammalian cells, 14-3-3 isoforms recruit phosphorylated tristetraprolin (TTP) and butyrate response factor-1 (BRF1), inhibiting target deadenylation (Schmidlin *et al.*, 2004; Stoecklin *et al.*, 2004). In trypanosomes there are two 14-3-3 isoforms (Inoue *et al.*, 2005) and we detected both of them increasing reporter expression. Interestingly, three metabolic enzymes were found to up-regulate reporter expression. This included the enzyme deoxyribose-phosphate aldolase (DERA, Tb927.7.5680). In HeLa cells, DERA is important for stress granule formation and interacts with the Y-box binding protein YB-1 (Salleron *et al.*, 2014). YB-1 participates in a wide variety of DNA/RNA-dependent events, including regulation of mRNA stability and translation. However, whether in trypanosomes DERA is related to RNA metabolism is yet

unknown. We also found a putative esterase, and the pyrroline-5-carboxylate reductase (P5CR), an enzyme that participates in arginine and proline metabolism. We have shown that P5CR increases reporter expression using a CAT reporter (Erben *et al.*, 2014b) (Fig. 1A). In HeLa cells, several enzymes involved in the synthesis of amino acids displayed RNA binding activities (Castello *et al.*, 2012). Using the MEME search algorithm, we identified three motifs specifically enriched in this activator group of proteins ($P<1$ e -5). The first motif was a histidine-rich 9mer motif at three sites ($P=3.6$ e -6). Remarkably, the other two were previously seen in a global survey on MKT1-interacting protein; a glutamine-rich 9mer was found at 11 sites ($P=1.3$ e -9) and the HNPY motif ($P=4.0$ e -15, 8mer) in 11 different proteins (Fig. 1B–D and Supporting Information Table S2). We have previously shown that MKT1 can interact with many proteins containing RNA-binding domains, and that such interaction occurs specifically through the HNPY motif (Singh *et al.*, 2014). For instance, MKT1 associate with ZC3H11 and PBP1 which in turn recruits poly(A)-binding protein (PABP1) promoting the stabilization of ZC3H11 targets. Our tethering results indicated that, as MKT1, RBPs carrying a HNPY motif have a positive regulatory effect on the fate of the target mRNA (Fig. 1D). As most of these proteins were shown to interact with MKT1, we can speculate that proteins carrying the HNPY feature will exert its function trough MKT1 increasing mRNA stability/translation. An *in silico* genome-wide search for additional proteins containing this feature identified three more hits ($P<10^{-4}$; $Q<0.01$) (Supporting Information Table S2). Interestingly, two of them were also pulled down with MKT1-TAP: the cyclin F box protein CFB2 and the PSP1 domain-containing protein Tb927.8.3850. In our ORFeome screen, overexpression of these proteins seems to affect negatively growth kinetics (see below), as it was shown previously for CFB1 (Benz and Clayton, 2007). Although overexpression of Tb927.8.3850 negatively affects cell growth, tethering Tb927.8.3850 for shorter time increases CAT reporter activity (Supporting Information Fig. S3). The other potential regulator is the hypothetical protein Tb927.7.3970 for which we have not data. All together, these data provide insights into the likely functions of the most relevant activating proteins involved in the trypanosome post-transcriptional control and indicate that the HNPY motif is a well-conserved feature unique to this group.

The screen also detects expected and novel repressors of gene expression

Forty six potential mRNA fate regulators were able to suppress PGKB expression (Table 2 and Supporting Information Table S2). The repressor list contains 25 canonical RBPs, and 10 proteins of unknown function among others. The 25 canonical RBPs include PUF, RRM, CCCH and PSP1 domain-containing proteins. Two canonical RBPs previously shown to negatively affect gene expression, RBP10 (Wurst *et al.*, 2012) and DRBD12 (Najafabadi *et al.*, 2013) are in this list. Not surprisingly, five subunits of the CAF1/NOT deadenylase complex reproducibly suppressed PGKB expression suggesting these subunits recruit the unique catalytic subunit CAF1 to the associated reporter mRNA (Farber *et al.*, 2013; Erben *et al.*, 2014a). As expected, the two components of the Pab1p-dependent poly(A) nuclease complex PAN2/PAN3, are also in this group. In Leishmania, it was shown that the eIF4E-interacting protein (eIF4E-IP) keeps eIF4E1 inaccessible for translation (Zinoviev *et al.*, 2011). In agreement with these observations, we observed that tethering of eIF4E1-IP reproducibly decreases PGKB expression. Unexpectedly, tethering of eIF4E1 also decrease expression. We don't know the reason but we can expect that the association with eIF4E1-IP must be involved. The MEME search algorithm identified only a glutamine-rich 10mer in 10 different proteins ($P=2.4$ e -11). PolyQ/N rich regions are typically enriched in RBPs where they are suggested to promote RNA granule assembly (Reijns *et al.*, 2008). How this subset of RBPs decreases mRNA abundance or translation is still not known but we can anticipate a simple model where RBPs recruit the mRNA degradation machinery to target mRNAs or to have inactivating interactions with the translation apparatus.

Over-expression of proteins involved in differentiation affects bloodstream-form viability

We observed that clones over-expressing some proteins exhibited a decline in read counts after tetracycline induction in our two complementary selections. Over-expression of the small ZFP1, ZFP2 and ZFP3 proteins was detrimental for bloodstream cell growth (Supporting Information Table S2). These proteins associate as a protein complex and are required for the normal differentiation process of the bloodstream form to the procyclic form (Paterou *et al.*, 2006; Walrad *et al.*, 2012). While the creation of a bloodstream cell line over-expressing ZFP1 was unsuccessful (Hendriks and Matthews, 2005), it was not shown if ZFP3 over-expression affects growth kinetics (Paterou *et al.*, 2006). In addition, it has been shown that over-expression of C-terminal tagged ZFP2 in bloodstream form had no effect on growth (Hendriks *et al.*, 2001). We speculate that the tagging location may influence the difference in the observed phenotype. We found that cells over-expressing the (ARE)-regulating protein DRBD13, is also disadvantageous for cell growth.

Table 2. Proteins that decrease PGKB expression.

Locus number	Description	Category	References
Tb927.9.7480	hypothetical protein	Unknown	
Tb927.10.9330	hypothetical protein	Unknown	
Tb927.6.5010	hypothetical protein	Unknown	
Tb927.6.3950	hypothetical protein	Unknown	
Tb927.11.8020	hypothetical protein	Unknown	
Tb927.11.2900	hypothetical protein	Unknown	
Tb927.3.3060	hypothetical protein	Unknown	
Tb927.9.13970	hypothetical protein	Unknown	
Tb927.8.910	hypothetical protein	Unknown	
Tb927.1.670	hypothetical protein	Unknown	
Tb927.10.13540	RBP12	RNA binding	
Tb927.5.1580	ZC3H13	RNA binding	(Ouna *et al.*, 2012)
Tb927.10.310	PUF3	RNA binding	(Klein *et al.*, 2015)
Tb927.4.400	DRBD7	RNA binding	(Wurst *et al.*, 2009)
Tb927.3.740	ZC3H5	RNA binding	
Tb927.7.5380	DRBD12	RNA binding	(Najafabadi *et al.*, 2013)
Tb927.3.5250	ZC3H8	RNA binding	
Tb927.9.10280	ZC3H48	RNA binding	
Tb927.10.14950	ZC3H40	RNA binding	
Tb927.9.12360	RBP35	RNA binding	
Tb927.6.4050	ZC3H14	RNA binding	
Tb927.7.2680	ZC3H22	RNA binding	
Tb927.9.13990	DRBD2	RNA binding	
Tb927.11.12120	RBP9	RNA binding	(Wurst *et al.*, 2009)
Tb927.10.14930	ZC3H39	RNA binding	(Alves *et al.*, 2014)
Tb927.4.4230	RBP31	RNA binding	(Wurst *et al.*, 2009)
Tb927.11.16550	ZC3H46	RNA binding	
Tb927.11.3340	RBP34	RNA binding	
Tb927.11.7140	CSBPII	RNA binding	(Mahmood *et al.*, 2001)
Tb927.8.2780	RBP10	RNA binding	(Wurst *et al.*, 2012)
Tb927.6.3480	DRBD5	RNA binding	(Wurst *et al.*, 2009)
Tb927.3.3940	DRBD11	RNA binding	
Tb927.10.1540	ZC3H30	RNA binding	
Tb927.10.4430	PUF1	RNA binding	(Luu *et al.*, 2006)
Tb927.10.12660	PUF2	RNA binding	(Jha *et al.*, 2014)
Tb927.11.13970	PAN3	RNA degradation	(Schwede *et al.*, 2009)
Tb927.8.1960	C2ORF29	RNA degradation	(Farber *et al.*, 2013)
Tb927.6.1670	PAN2 exoribonuclease	RNA degradation	(Schwede *et al.*, 2009)
Tb927.6.600	CAF1 deadenylase	RNA degradation	(Schwede *et al.*, 2008)
Tb927.6.850	NOT2	RNA degradation	(Farber *et al.*, 2013)
Tb927.10.8720	CNOT10	RNA degradation	(Farber *et al.*, 2013)
Tb927.3.1920	NOT5	RNA degradation	(Farber *et al.*, 2013)
Tb927.11.15000	(SMN)-like protein (TbSMN)	RNA processing	(Jaé *et al.*, 2014)
Tb927.9.14120	NIF phosphatase	RNA processing	
Tb927.11.2260	eIF4E1	Translation	(Pereira *et al.*, 2013)
Tb927.9.11050	eIF4E-IP	Translation	(Zinoviev *et al.*, 2011)

We consider repressors to be proteins causing at least a 1.5-fold reduction in reads per million in the tetracycline induced condition in both replicates. Related references are shown.

DRBD13 down-regulates transcripts encoding cell surface coat proteins (Jha *et al.*, 2015). Although the effect of over-expressing DRBD13 in bloodstream forms has not been evaluated, DRBD13 levels are tightly regulated in the procyclic stage, as both DRBD13 over-expression and depletion were deleterious to the parasite's growth. Interestingly, we also identified a conserved hypothetical protein (Tb927.11.2250) negatively affecting cell growth. This protein is restricted to *T. brucei*, might be involved in driving non-dividing stumpy formation (Mony *et al.*, 2014) and is pulled-down with poly(A) RNA (see below). Other proteins with a detrimental effect on cell growth when over-expressed displaying mRNA-binding activities are ZC3H47, ZC3H29 and RBP38 (Supporting Information Tables S2 and S3) suggesting they may regulate essential target genes.

Several down-regulators interact with the deadenylation CAF1/NOT1 complex

In metazoans, several RNA-binding cofactors contribute to the recruitment of deadenylases to the mRNA target

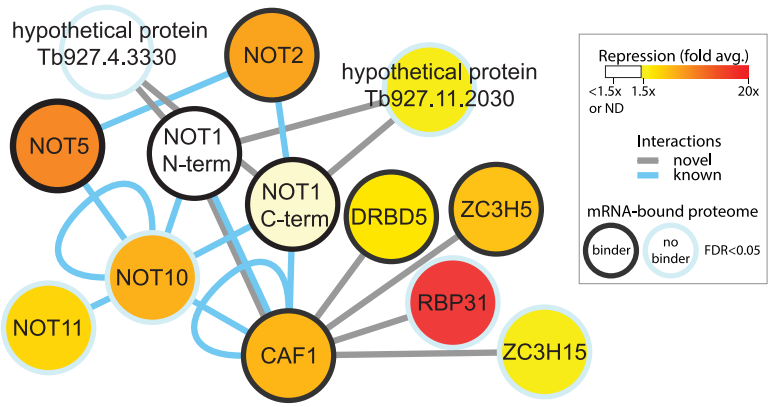

Fig. 2. Graphical representation of the CAF1/NOT complex interactome map found by yeast 2-hybrid. While known interactions are shown in blue (Farber *et al.*, 2013; Erben *et al.*, 2014a), novel interactions are shown with grey edges. Protein nodes are coloured according to the repression strength (average values) as judged by the tethering results (Supporting Information Table S2). Proteins with potential mRNA-binding activities (FDR<5%) are represented with bold black borders. ND: not determined.

by directly binding to them (Goldstrohm and Wickens, 2008). We speculated that some of our down-regulators might act in the same way. In trypanosomes, most mRNAs are deadenylated prior to degradation, and the major deadenylation activity resides in the CAF1/NOT complex (Schwede *et al.*, 2008; Fadda *et al.*, 2013; Erben *et al.*, 2014a). To identify co-repressors, we performed a yeast-two hybrid (Y2H) screen using as bait, the exoribonuclease CAF1 and both N- and C-terminal halves of the scaffold NOT1. For this, the ORFeome library was shuttled by recombination into a Gateway-compatible plasmid designed to make a fusion protein with GAL4 (Maier *et al.*, 2008). About 100 colonies which grew on highly stringent plates were analysed; plasmid DNA were isolated and the trypanosome DNA inserts were PCR-amplified and sequenced. Isolated prey plasmids were transformed back into the yeast strain used in the screen and checked for auto-activation. We have previously shown by Y2H assays the pairwise interaction between different subunits of the CAF1/NOT complex (Farber *et al.*, 2013; Erben *et al.*, 2014a). We showed that CAF1 interacts with the non-catalytic subunits NOT10 and both halves of NOT1. Here, we also detected positive interaction between CAF1 with both N- and C-terminal halves of NOT1 (Fig. 2). In addition, we found that both halves of NOT1 interact with two hypothetical proteins: an EF-hand-containing protein Tb927.4.3330 and Tb927.11.2030. Both proteins decreased reporter mRNA expression in our shotgun screening (Erben *et al.*, 2014b), as well the last one in this work (Fig. 2). Moreover, we found that the deadenylase interacts with the zinc finger proteins ZC3H15 and ZC3H5 and the RRM-containing proteins DRBD5 and RBP31. Remarkably, all of these proteins exhibited a down-regulating activity in our tethering assay suggesting a possible mode of action. Additionally, DRBD5 and ZC3H5 displayed RNA-binding activity *in vivo* (see below). This indicates that DRBD5 and ZC3H5 could directly repress mRNA expression by recruitment of CAF1.

The bloodstream form poly(A) mRNA-bound interactome

To find out whether the identified potential regulators bind directly to mRNA and to expand the composition of mRNPs, we analyzed the mRNA-bound proteome of bloodstream cells. For this, we performed *in vivo* cross-linking followed by poly(A) RNA enrichment (Castello *et al.*, 2012). To test whether the eluted proteins were enriched for *bona fide* RBPs, we first carried out western blotting to detect the well-known RBPs: PABP1 (Singh *et al.*, 2014), PUF2 (Jha *et al.*, 2014), DRBD3 (Estevez, 2008) and the potential binder RBP10 (Wurst *et al.*, 2012). As expected, all proteins were enriched in cross-linked (CL) samples and undetectable in controls (nonirradiated cells) (Fig. 3A and B). Conversely, both control and cross-linked samples were devoid of aldolase, one of the most abundant proteins in bloodstream form. We also checked the PSP1-containing protein Tb927.8.3850. Notably, it was barely detectable in whole lysates, but it was specifically enriched in CL samples. This suggests that our protocol can successfully select low-abundance proteins and also that Tb927.8.3850 may up-regulate target expression (Supporting Information Fig. S3) by direct contact.

After RNase treatment, proteins were cleaved into peptides with trypsin and resulting fractions were analyzed by high-resolution mass spectrometry. Eight hundred seventy three proteins were identified in all six proteomic analyses with one or more spectral counts, 544 of which were identified from at least two biological cross-linked replicates (Supporting Information Table S3). To apply statistical data analysis, protein enrichment in cross-linked samples over controls was assessed by considering the spectral counts (Liu *et al.*, 2004) as judged by the bioconductor package DESeq. The analysis resulted in a nonredundant list of 155 proteins (FDR 1%). This included nuclear, nucleolar, mitochondrial and cytoplasmic proteins, showing the diversity in the mRNPs pulled-down. While it is likely that the coverage of this apparent bloodstream form

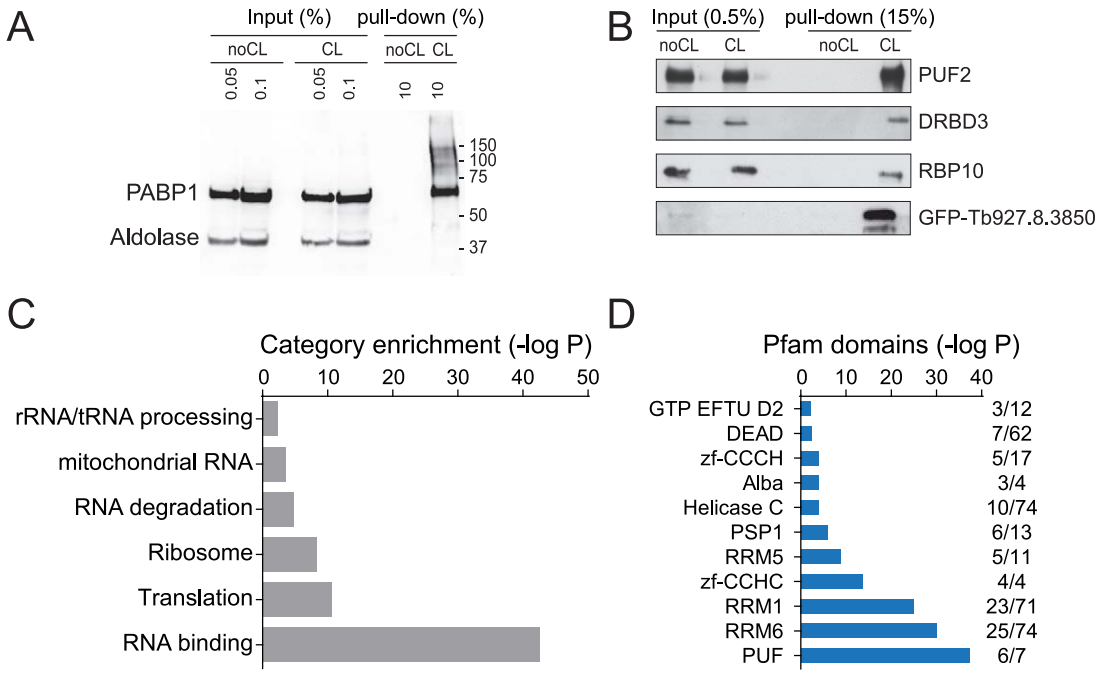

Fig. 3. *In Vivo* capture of bloodstream RBPs mRNA-protein interactions by UV cross-linking on cultured bloodstream cells.
A–B. After RNase treatment, released proteins were analyzed by western blotting against PABP1, aldolase, against the RBPs PUF2, DRBD3, RBP10 and the hypothetical protein Tb927.8.3850. CL: cross-linked, noCL: non-cross-linked.
C. The high-confidence RBPs (FDR< 0.01) are highly enriched in RNA-related processes. Shown are the most significantly enriched functional categories (Fisher's exact test, $P < 0.01$).
D. This group is also highly enriched in well-defined RNA-binding domains. Top protein domains with the smallest *P*-values from Pfam (Fisher's exact test, $P < 0.01$). Only domains with three hits or more are shown. The numbers on the right side of the graph indicate the numbers of proteins detected in the mRNA-bound proteome and genome-encoded proteins belonging to each Pfam domain category, respectively.

mRNA-bound proteome is not complete, it has a reasonable relative complexity compared with similar studies in metazoans (Baltz *et al.*, 2012; Castello *et al.*, 2012; Kwon *et al.*, 2013). For instance, PABP isoforms were also co-isolated in control poly(A) RNA without cross-linking and are excluded from the list. Anticipating this scenario, we also analyzed as a negative control, the mRNA-bound proteome of cross-linked cells treated with RNAses before elution. Although detected PABP isoforms increased, this treatment decreased the amount of proteins associated with poly(A) RNA (Supporting Information Table S3, compare column G with columns F and H). This implies that a fraction of proteins that were cut out from the high confidence list (FDR>0.01) are likely to be bona fide RBPs. For instance, the RNA helicase UPF1 (Delhi *et al.*, 2011) and the cell cycle sequence binding phosphoprotein homolog RBP33, (Mittra and Ray, 2004) among others are only removed from the beads after RNAse treatment. Although we considered for subsequent analysis only the high-confidence RBP list (FDR 1%), we recommend individual validation by other means.

We first classified the identified proteins into functional categories based on gene annotation. Analysis of the most significant over-represented categories revealed that the 155 proteins found in our RNA-bound proteome are highly enriched in RNA-related processes, making up close to 70% of the identified proteins. Proteins with 'RNA-binding' activities are the most enriched, followed by proteins involved in translation, ribosomal and mRNA degradative enzymes (Fig. 3C). In addition to the expected mRNA-interacting proteins, we identified many proteins which have not been previously annotated as RNA binding (Supporting Information Table S3). This group of potential binders includes 35 uncharacterized hypothetical proteins for which a putative function cannot be clearly assigned. Among others were also found the two universal minicircle sequence-binding proteins (UMSBPs), a kinetoplast conserved phosphatase and the CFB1/2 proteins.

To obtain an additional perspective of the mRNA-bound proteome, we also performed Pfam domain enrichment analysis using the identified proteins. Consistently, Pfam domain analysis showed that the most highly represented domains comprise canonical RBDs (RRM, PUF and Zinc finger; Fig. 3D). Proteins encoding PSP1 domains interact with mRNA in *Crithidia fasciculata* (Mittra and Ray, 2004) and with lncRNA in mammalian cell lines (Naganuma and Hirose, 2013). Our data

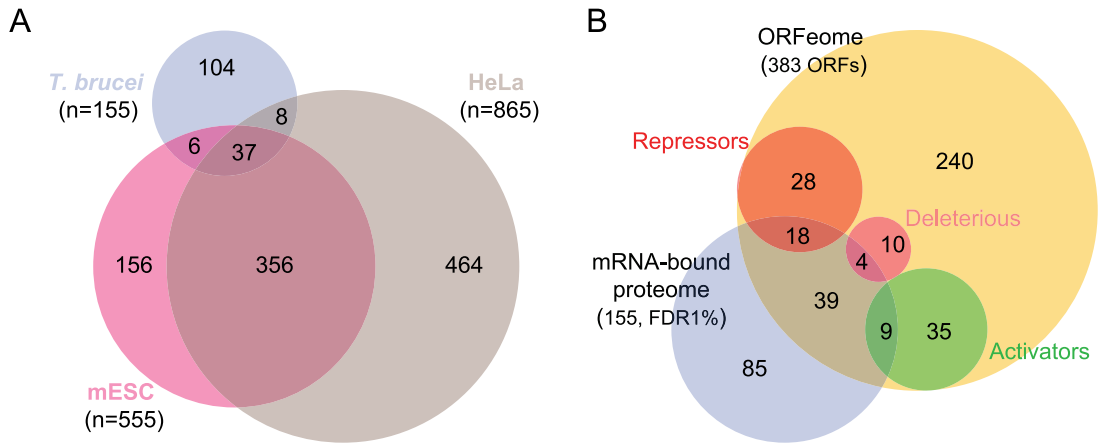

Fig. 4. A. Comparison of the *T. brucei* mRNA-bound proteome and two published mammalian data sets (Castello *et al.*, 2012; Kwon *et al.*, 2013). B. Venn diagram depicting the overlap between the detected mRNA-bound proteome (blue; FDR<0.01) with the intended mini-ORFeome collection (orange). Number of proteins found to up-regulate (green) or repress (red) reporter expression >1.5x upon tethering are indicated. Numbers of proteins whose overexpression affects negatively cell growth (deleterious) are also shown.

support a major function of PSP1 domains in RNA-binding with six out of 13 encoded proteins bound to poly(A) mRNA as it was shown for Tb927.8.3850 (Fig. 3B). Also several proteins encoding C/DEAD helicase, ALBA and translation elongation factors (GTP_EFTU_D2) were specifically pulled down.

Finally, we looked for evolutionarily conserved components of the mRNA-bound proteome. Recent studies identified hundreds of proteins from human and mouse embryonic stem cells (mESC) as candidate RBPs (Castello *et al.*, 2012; Kwon *et al.*, 2013). We compared the *T. brucei* data with these data sets and found that 37 out of 64 orthologs are common to all three data sets (Fig. 4A and Supporting Information Table 3). Six and eight extra proteins are found also in mESC and HeLa cells, respectively. These overlapping 37 proteins can be considered as evolutionarily conserved components and may have constitutive functions. Remarkable, the glycolytic enzyme pyruvate kinase binds to mRNA in *T. brucei*, mESC and Hela cells suggesting an evolutionary conserved function (Castello *et al.*, 2012; Kwon *et al.*, 2013). Other potential conserved RBPs include several helicases, ribosomal proteins, translation factors and the LA protein among others.

Conclusions

We have generated a small trypanosome ORFeome resource and established a functional tethering assay to identify the major regulators of gene expression. Our tethering screen identified 90 hits displaying not only features commonly found in the RNA biology but also novel players. The screen also showed that proteins carrying an HNPY sequence pattern exhibit a positive effect on mRNA expression and predicted novel members of this group. Whether all these proteins interact with MKT1 remains to be proven. Previous yeast two-hybrid experiments mapped the CAF1/NOT interactions (Farber *et al.*, 2013; Erben *et al.*, 2014a). Here we showed that the deadenylase CAF1 could interact with the repressors ZC3H15, ZC3H5, DRBD5 and RBP31. Presumably, mRNAs that undergo deadenylation associate with these sequence-specific factors which, in turn, recruit CAF1. If the central scaffold NOT1 has extra functions beside deadenylation is yet not known.

Next, we have captured the *in vivo* mRNA-bound proteome of bloodstream cells and uncovered numerous trypanosome proteins as novel RBPs enlarging the number of potential regulators. An overview of our main findings is shown in Fig. 4B. We found 155 high-confidence RBPs, including 35 hypothetical proteins with no RNA-related ontology. Strikingly, 13% (20 proteins) lacks any recognizable domain as judged by NCBI's conserved domain database suggesting the existence of unknown RNA-binding architectures. The mRNA-bound proteome also include 39 potential RBPs with no obvious reproducible effect on the tethering screen (Fig. 4B). In fact, only a fraction (~30%) of the analyzed regulators exhibited a clear effect on the expression of the mRNA reporter. Given proteins act as part of a complex, we can speculate that limiting amounts of other components will occur in false negatives. It is also plausible that some of the proteins are active only in a specific developmental stage or in the presence of certain post-translational modifications and/or specific substrates and cofactors. In addition, the possibility that the function is affected by the lambda fusion cannot be ruled out. Nevertheless, our previous random shotgun approach is

still a useful dataset as it has the power to delineate the involved functional domains.

Finally, we showed that 37 RBPs are potentially conserved from trypanosome to human (Fig. 4A). Interestingly, we found the enzyme pyruvate kinase able to bind mRNA adding further example of a possible regulatory link between gene expression and intermediary metabolism. During the revision of this article, two new studies also identify this enzyme as part of the RBP repertoire in budding yeast and human hepatocytic cells (Beckmann *et al*., 2015; Matia-Gonzalez *et al*., 2015) suggesting that RNA-binding activity may be an ancient and conserved function of this enzyme.

As many potential RBPs were clearly not included in our library, it will be important to expand our ORFeome collection to make it a more valuable resource for the study of RNA-protein networks. Since dynamic changes in RNA-binding are expected to occur, it will be interesting to perform similar experiments along the distinct developmental trypanosome stages helping to decipher the constitutive and stage-specific mRNA interactions.

Experimental procedures

Library generation

Gene-specific primers and ORFs amplification were previously described and can be found in Supporting Information Table S1 (Erben *et al*., 2014b). Briefly, the corresponding 384 ORFs were amplified directly from genomic DNA using Q5 high fidelity DNA polymerase (New England Biolabs); each primer included additional sequence suitable for amplification and directional Gateway® cloning. A second PCR amplification was performed to include full attB recombination sites. Then, the Gateway®-compatible PCR products were gel purified, grouped by size and recombined into pDONR^{TM}221 by using BP Clonase^{TM} II Enzyme Mix (Invitrogen) following manufacturer's instructions. After proteinase K treatment, the BP reactions were directly used to transform One Shot® TOP10 chemically competent *E. coli* (Invitrogen) to generate several entry Gateway® libraries. For the tethering screen, we shuttled all entry clones into a Gateway®-compatible tethering plasmid. This is a derivative of the tetracycline-regulated pHD678 (Biebinger *et al*., 1997): it contains a lambda-N peptide followed by three copies of c-myc epitope and a Gateway cassette in frame with the N-terminal tag. The quality of the destination library was verified by PCR-gel electrophoresis and DNA sequencing. In frame inserts were found in about 90% of plasmids. For yeast 2-hybrid, all entry clones were shuttled as before, into the Gateway®-compatible Y2H destination vector pAD-Gate2 (Maier *et al*., 2008).

T. brucei *growth, manipulation and Western blots*

Bloodstream form *T. brucei* 2T1 cells (Alsford and Horn, 2008) were transfected with the blasticidin (BLA) and PGKB reporters as described (Erben *et al*., 2014b) (Supporting Information Fig. S1). These cell lines were then transfected with the pRPaSce* plasmid which encodes the homing endonuclease I-*Sce*I gene and the cleavage site to the tagged rRNA spacer locus (Alsford *et al*., 2011). I-*Sce*I was induced using tetracycline at 1 mg/ml for 3 h and then the cells were electroporated with the ORFeome collection. For each series, an aliquot of the transfection was diluted to determine the transfection efficiency. Two biological replicates of each procedure were performed. To assay for stabilizing proteins, populations expressing the blasticidin resistance mRNA were preinduced for 24 h with 1 μg/ml tetracycline and then grown with 2x and 10x concentrations of blasticidin (1x = 5 mg/ml) for four days. For destabilizing proteins, populations expressing the *PGKB* mRNA were induced for 5 days with 1 μg/ml tetracycline. Samples were analyzed by western blot using polyclonal antibodies against aldolase (Clayton, 1987), PABP1 (gift from Dr. Laurie Read), PUF2 (Jha *et al*., 2014), DRBD3 (Estevez, 2008), RBP10 (Wurst *et al*., 2012), monoclonal antibodies against V5 (AbD seroTec) and GFP tags (Santa Cruz). For the tethering assays, cell lines constitutively expressing CAT reporter with boxB actin 3'-UTR (*CAT-B-ACT*) were co-transfected with plasmids encoding Tb927.10.15760, Tb927.11.12730, Tb927.8.4200 and Tb927.3.1810 in fusion with the lambda N-peptide. CAT activity was performed by triplicate as previously described (Erben *et al*., 2014b).

DNA sequencing and analysis

Genomic DNA was isolated from surviving populations and then, the cloned ORFs were in each case recovered by plasmid-specific PCR as described (Erben *et al*., 2014b). Adaptor-ligated sequencing libraries were prepared from PCR reactions and Illumina sequenced. The lambda-N sequence was removed, and the remainder was mapped to the *T. brucei* 927 reference genome (http://tritrypdb.org/tritrypdb) using Bowtie, allowing one base mismatch (Langmead *et al*., 2009). Read counts were extracted with custom script. To find proteins that either increased blasticidin resistance (up-regulators) or increased survival after PGKB expression (down-regulators) we analyzed all candidates with >25 read counts after blasticidin treatment (10-fold) and induction (tet+), respectively. We then calculated the RPKM (reads per kilobase per million) mapped reads for each experiment. For the blasticidin experiment, the normalized number of counts for 10-fold increased blasticidin (BLA 10X) was divided by the counts for minus tet (tet-), twofold blasticidin (BLA2x) and separately by the counts for cells with tetracycline but without blasticidin. To reduce the likelihood of identifying PCR artifacts, the lowest of these three values was taken to be the relative enrichment. We consider activators to be proteins showing a RPKM enrichment of at least 1.5-fold in both replicates. To find proteins that increased survival after PGKB expression, RPKM values from induced populations (tet +) were divided by the number of counts for the uninduced. Proteins showing at least 1.5-fold enrichment in both experiments were considered as repressors. We considered proteins detrimental for cell growth to those that showed a drop in the normalized count

reads (RPKM) upon tetracycline induction. This means that for all the conditions analyzed, we consider as 'deleterious' to those displaying a tet+/tet- count ratio lower than 1.

Bioinformatics

Motif identification was performed with the MEME program (Bailey *et al.*, 2015) using the discriminative mode on the set of activators and repressor sequences. The motif width was set up to 10 residues and specified the occurrences to one per sequence for the MEME search parameters. The motif position-specific frequency matrix derived from MEME was used as input to FIMO analysis (Bailey *et al.*, 2015). Statistical significance of the overrepresentation of Pfam domains and functional categories were tested by Fisher's exact test. To define orthologs of yeast and human, *T. brucei* gene IDs were converted to UniProt IDs and used for the InParanoid ortholog groups database (Version 8.0) (Sonnhammer and Ostlund, 2015). The orthologs between yeast and human RBPs were previously classified (Kwon *et al.*, 2013).

Yeast two-hybrid analysis

The yeast two-hybrid screen was performed using the bait vector pGBKT7 containing full-length CAF1 and both halves of the scaffold protein NOT1 (Farber *et al.*, 2013). The screening was done with the mini-ORFeome described in the library generation section according to protocols published previously (Mohr and Koegl, 2012). Before sequencing, prey and bait plasmids of candidate clones were co-transformed into AH109 yeast cells and grown on high stringency plates (Singh *et al.*, 2014). The interaction with Lamin served as negative bait control.

In vivo *capture of RBPs*

For interactome capture, about 5 x 10^9 bloodstream cells in log phase (~1 x 10^6cells/ml) were pelleted and resuspended in HMI-9 medium lacking serum. Resuspended cells were exposed twice to 500 mJ/cm^2 of 254 nm UV in a Stratalinker 2400 (Stratagene) with a short break on ice and gentle mixing. After UV irradiation, cells were pelleted and washed twice with PBS and then resuspended in 30 ml of lysis buffer (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, 5 mM DTT and cOmplete EDTA-free protease inhibitors (Roche)). For lysis completion, cells were passed through a 27 gauge syringe needle 15 times to generate total cell extract. Pull down of poly(A) mRNA was performed exactly as previously described (Castello *et al.*, 2013b) using 1ml of magnetic oligo d(T)25 beads (NEB) per liter of starting culture. Briefly, after 1 h incubation at 4°C the beads were collected with a magnet and the supernatant was recovered for a second incubation. The beads were washed sequentially with two rounds of lysis buffer, Buffer 1 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.1% LiDS (wt/vol), 1 mM EDTA and 5 mM DTT), Buffer 2 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 1 mM EDTA and 5 mM DTT) and Buffer 3 (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 1 mM EDTA and 5 mM DTT) before elution with 20 mM Tris-HCl (pH 7.5) and 1 mM EDTA. The eluates were combined and concentrated with Amicon Ultra centrifugal filters (3 kDa cutoff). In total, we subjected six samples for MS analysis: three independent experiments of the RBP-bound proteome, two independent experiments without UV irradiation and one experiment with UV cross-linking but treated with RNAse A/T1 before elution (Supporting Information Tables S3, Sheet 1, column G). For this control, cells were UV cross-linked and the poly(A) mRNA pulled-down as before. Beads were sequentially washed with Buffers 1–3 and then treated with 50 U RNAse T1 and 20 μg RNAse A (Thermo Scientific) in RNAse buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% (vol/vol) NP-40 and 0.5 mM DTT) for 1h at 37°C to digest RNA. Beads were then washed again with Buffer 3, eluted and concentrated as before.

In-gel tryptic digestion and LC-MS/MS analysis

After SDS-PAGE, Coomassie stained bands were cut out with a scalpel. Gel slices were transferred to a 96 well plate and automatically reduced, alkylated and digested with trypsin. Peptides were extracted from the gel pieces with 50% acetonitrile/0.1% TFA, concentrated nearly to dryness in a SpeedVac vacuum centrifuge and diluted to a total volume of 30 μl with 0.1% TFA. 10 μl of the sample was analyzed by a Dionex UltiMate 3000RSLCnano HPLC system (Thermo Scientific) coupled to a LTQ Orbitrap Elite mass spectrometer (Thermo Scientific). Samples were loaded on a C18 trapping column (Acclaim PepMap100, C18, 5 μm, 300 μm i.d. x 5 mm, Thermo Scientific), using a loading buffer of 0.1% TFA at a flow rate of 30 μl/min. Peptides were eluted and separated on an C18 analytical column (Acclaim PepMap RSLC C18, 2 μm, 75 μm i.d. x 25 cm, Thermo Scientific) with a flow rate of 300 nL/min in a gradient of buffer A (1% acetonitrile, 0.1% formic acid, 5% DMSO, 93.9% water) and buffer B (90% acetonitrile, 0.1% formic acid, 5% DMSO, 4.9% water): 0–3 min: 4% B; 3-50 min: 4–40% B; 50-55 min: 40-95% B; 55–60 min: 95% B. One Orbitrap survey scan (scan range: 400–1600 m/z, resolution: 60,000, AGC target: 1 x 10^6, max fill time: 10 ms) was followed by 20 information-dependent product ion scans in the LTQ ion trap (normal scan rate, AGC target: 1 x 10^4, max fill time: 100 ms). 2+, 3+ and 4+ charged ions were selected for data dependent CID-fragmentation (norm. coll. energy: 35%, act. $q = 0.25$, act. time = 30 ms) with dynamic exclusion enabled (excl. duration: 60 s, parent mass accuracy: 5 ppm).

Protein identification and data analysis

For protein identification, spectra were correlated with a *T. brucei* protein database (Parsons *et al.*, 2015) containing 8429 protein entries using the software MaxQuant (v1.5.1.2) (Cox and Mann, 2008). All searches were performed with tryptic specificity allowing up to two missed cleavages. Carbamidomethylcysteine was set as fixed modification while oxidation of methionine and acetylation of protein N termini were considered as variable. The first search was performed with a mass accuracy of 20 ppm for precursor ions and 0.5 Da for fragment ions. A false discovery rate of 1% was applied to both the peptide and the

protein level. The minimum peptide length was 7 amino acids with the re-quantify option enabled. Results were cleaned for reverse, only identified by site and contaminants protein groups. For statistical analysis, we determined the protein enrichment in cross-linked samples over controls using the spectral counting method (Liu *et al*., 2004). To perform statistical data analysis taking into accounts reproducibility and natural variability, the bioconductor package DESeq2 was applied (Love *et al*., 2014).

Acknowledgement

EE thanks Christine Clayton for reading the manuscript. We appreciate the help of the ZMBH Mass Spectrometry Core Facility members, especially to Robert Hardt for support on Maxquant analysis. EE also thank Dr. J. Maldera and Dr. A. Leroux for fruitful discussions and for critical reading of the manuscript. The authors declare that they have no conflict of interest. This work was partly covered by the Deutsche Forschungsgemeinschaft (DFG) [Cl112/17-1].

References

Alsford, S., and Horn, D. (2008) Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei*. *Mol Biochem Parasitol* **161**: 76–79.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., *et al*. (2011) High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res* **21**: 915–924.

Alves, L.R., Oliveira, C., Morking, P.A., Kessler, R.L., Martins, S.T., Romagnoli, B.A., *et al*. (2014) The mRNAs associated to a zinc finger protein from *Trypanosoma cruzi* shift during stress conditions. *RNA Biol* **11**: 921–933.

Archer, S.K., Luu, V.D., de Queiroz, R.A., Brems, S., and Clayton, C. (2009) *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog* **5**: e1000565.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015) The MEME Suite. *Nucleic Acids Res* **43**: W39–W49.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., *et al*. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**: 674–690.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.M., *et al*. (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat Commun* **6**: 10127.

Benz, C., and. Clayton, C.E. (2007) The F-box protein CFB2 is required for cytokinesis of bloodstream-form *Trypanosoma brucei*. *Mol Biochem Parasitol* **156**: 217–224.

Benz, C., Mulindwa, J., Ouna, B., and Clayton, C. (2011) The *Trypanosoma brucei* zinc finger protein ZC3H18 is involved in differentiation. *Mol Biochem Parasitol* **177**: 148–151.

Biebinger, S., Wirtz, L.E., Lorenz, P., and Clayton, C. (1997) Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol Biochem Parasitol* **85**: 99–112.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., *et al*. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**: 1393–1406.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013a) RNA-binding proteins in Mendelian disease. *Trends Genet* **29**: 318–327.

Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., *et al*. (2013b) System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc* **8**: 491–500.

Clayton, C. (2013) The regulation of trypanosome gene expression by RNA-binding proteins. *PLoS Pathog* **9**: e1003680.

Clayton, C.E. (1987) Import of fructose bisphosphate aldolase into the glycosomes of *Trypanosoma brucei*. *J Cell Biol* **105**: 2649–2654.

Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.

Das, A., Morales, R., Banday, M., Garcia, S., Hao, L., Cross, G.A., *et al*. (2012) The essential polysome-associated RNA-binding protein RBP42 targets mRNAs involved in *Trypanosoma brucei* energy metabolism. *RNA* **18**: 1968–1983.

Delhi, P., Queiroz, R., Inchaustegui, D., Carrington, M., and Clayton, C. (2011) Is there a classical nonsense-mediated decay pathway in trypanosomes? *PLoS One* **6**: e25112.

Dhalia, R., Marinsek, N., Reis, C.R., Katz, R., Muniz, J.R., Standart, N., *et al*. (2006) The two eIF4A helicases in *Trypanosoma brucei* are functionally distinct. *Nucleic Acids Res* **34**: 2495–2507.

Droll, D., Archer, S., Fenn, K., Delhi, P., Matthews, K., and Clayton, C. (2010) The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. *FEBS Lett* **584**: 1156–1162.

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 Pathog* **9**: e1003286.

Erben, E., Chakraborty, C., and Clayton, C. (2014a) The CAF1-NOT complex of trypanosomes. *Front Genet* **4**: 299.

Erben, E.D., Fadda, A., Lueong, S., Hoheisel, J.D., and Clayton, C. (2014b) A genome-wide tethering screen reveals novel potential post-transcriptional regulators in *Trypanosoma brucei*. *PLoS Pathog* **10**: e1004178.

Estevez, A.M. (2008) The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res* **36**: 4573–4586.

Fadda, A., Farber, V., Droll, D., and Clayton, C. (2013) The roles of 3'-exoribonucleases and the exosome in trypanosome mRNA degradation. *RNA* **19**: 937–947.

Farber, V., Erben, E., Sharma, S., Stoecklin, G., and Clayton, C. (2013) Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. *Nucleic Acids Res* **41**: 1211–1222.

Freire, E.R., Dhalia, R., Moura, D.M., da Costa Lima, T.D., Lima, R.P., Reis, C.R., *et al*. (2011) The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. *Mol Biochem Parasitol* **176**: 25–36.

Freire, E.R., Malvezzi, A.M., Vashisht, A.A., Zuberek, J., Saada, E.A., Langousis, G., *et al*. (2014) *Trypanosoma brucei* translation initiation factor homolog EIF4E6 forms a tripartite cytosolic complex with EIF4G5 and a capping enzyme homolog. *Eukaryot Cell* **13**: 896–908.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **582**: 1977–1986.

Goldstrohm, A.C., and Wickens, M. (2008) Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* **9**: 337–344.

Hendriks, E.F., and Matthews, K.R. (2005) Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. *Mol Microbiol* **57**: 706–716.

Hendriks, E.F., Robinson, D.R., Hinkins, M., and Matthews, K.R. (2001) A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *EMBO J* **20**: 6700–6711.

Inoue, M., Nakamura, Y., Yasuda, K., Yasaka, N., Hara, T., Schnaufer, A., *et al*. (2005) The 14-3-3 proteins of *Trypanosoma brucei* function in motility, cytokinesis, and cell cycle. *J Biol Chem* **280**: 14085–14096.

Jaé, N., Preußer, C., Krüger, T., Tkacz, I.D., Engstler, M., Michaeli, S., and Bindereif, A. (2014) snRNA-specific role of SMN in trypanosome snRNP biogenesis in vivo. *RNA Biol* **8**: 90–100.

Jha, B.A., Archer, S.K., and Clayton, C.E. (2013) The trypanosome Pumilio domain protein PUF5. *PLoS One* **8**: e77371.

Jha, B.A., Fadda, A., Merce, C., Mugo, E., Droll, D., and Clayton, C. (2014) Depletion of the *Trypanosome Pumilio* domain protein PUF2 or of some other essential proteins causes transcriptome changes related to coding region length. *Eukaryot Cell* **13**: 664–674.

Jha, B.A., Gazestani, V.H., Yip, C.W., and Salavati, R. (2015) The DRBD13 RNA binding protein is involved in the insect-stage differentiation process of *Trypanosoma brucei*. *FEBS Lett* **589**: 1966–1974.

Klass, D.M., Scheibe, M., Butter, F., Hogan, G.J., Mann, M., and Brown, P.O. (2013) Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Res* **23**: 1028–1038.

Klein, C., Terrao, M., Inchaustegui Gil, D., and Clayton, C. (2015) Polysomes of *Trypanosoma brucei*: association with initiation factors and rna-binding proteins. *PLoS One* **10**: e0135973.

Kolev, N.G., Ramey-Butler, K., Cross, G.A., Ullu, E., and Tschudi, C. (2012) Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science* **338**: 1352–1353.

Kramer, S., Bannerman-Chukualim, B., Ellis, L., Boulden, E.A., Kelly, S., Field, M.C., and Carrington, M. (2013) Differential localization of the two *T. brucei* poly(A) binding proteins to the nucleus and RNP granules suggests binding to distinct mRNA pools. *PLoS One* **8**: e54004.

Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., *et al*. (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20**: 1122–1130.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Liu, H., Sadygov, R.G., and Yates, J.R., III. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**: 4193–4201.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Luu, V.D., Brems, S., Hoheisel, J.D., Burchmore, R., Guilbride, D.L., and Clayton, C. (2006) Functional analysis of *Trypanosoma brucei* PUF1. *Mol Biochem Parasitol* **150**: 340–349.

Mahmood, R., Mittra, B., Hines, J.C., and Ray, D.S. (2001) Characterization of the Crithidia fasciculata mRNA cycling sequence binding proteins. *Mol Cell Biol* **21**: 4453–4459.

Maier, R., Brandner, C., Hintner, H., Bauer, J., and Onder, K. (2008) Construction of a reading frame-independent yeast two-hybrid vector system for site-specific recombinational cloning and protein interaction screening. *Biotechniques* **45**: 235–244.

Mani, J., Guttinger, A., Schimanski, B., Heller, M., Acosta-Serrano, A., Pescher, P., *et al*. (2011) Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS One* **6**: e22463.

Matia-Gonzalez, A.M., Laing, E.E., and Gerber, A.P. (2015) Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat Struct Mol Biol* **22**, 1027–1033.

Michaeli, S. (2011) Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol* **6**: 459–474.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013) Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**: 127–133.

Mittra, B., and Ray, D.S. (2004) Presence of a poly(A) binding protein and two proteins with cell cycle-dependent phosphorylation in Crithidia fasciculata mRNA cycling sequence binding protein II. *Eukaryot Cell* **3**: 1185–1197.

Mohr, K., and Koegl, M. (2012) High-throughput yeast two-hybrid screening of complex cDNA libraries. *Methods Mol Biol* **812**: 89–102.

Mony, B.M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., *et al*. (2014) Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. *Nature* **505**: 681–685.

Moura, D.M., Reis, C.R., Xavier, C.C., da Costa Lima, T.D., Lima, R.P., Carrington, M., and de Melo Neto, O.P. (2015) Two related trypanosomatid eIF4G homologues have functional differences compatible with distinct roles during translation initiation. *RNA Biol* **12**: 305–319.

Naganuma, T., and Hirose, T. (2013) Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biol* **10**: 456–461.

Najafabadi, H.S., Lu, Z., MacPherson, C., Mehta, V., Adoue, V., Pastinen, T., and. Salavati, R., (2013) Global identification of conserved post-transcriptional regulatory programs in trypanosomatids. *Nucleic Acids Res* **41**: 8591–8600.

Ouna, B.A., Stewart, M., Helbig, C., and Clayton, C., (2012) The *Trypanosoma brucei* CCCH zinc finger proteins ZC3H12 and ZC3H13. *Mol Biochem Parasitol* **183**: 184–188.

Parsons, M., Ramasamy, G., Vasconcelos, E.J., Jensen, B.C., and Myler, P.J. (2015) Advancing *Trypanosoma brucei* genome annotation through ribosome profiling and spliced leader mapping. *Mol Biochem Parasitol* **202**: 1–10.

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. *J Biol Chem* **281**: 39002–39013.

Pereira, M.M., Malvezzi, A.M., Nascimento, L.M., Lima, T.D., Alves, V.S., Palma, M.L., *et al*. (2013) The eIF4E subunits of two distinct trypanosomatid eIF4F complexes are subjected to differential post-translational modifications associated to distinct growth phases in culture. *Mol Biochem Parasitol* **190**: 82–86.

Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J.D., and Clayton, C. (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics* **10**: 495.

Quenault, T., Lithgow, T., and Traven, A. (2011) PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol* **21**: 104–112.

Ramaswami, M., Taylor, J.P., and Parker, R. (2013) Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell* **154**: 727–736.

Reijns, M.A., Alexander, R.D., Spiller, M.P., and Beggs, J.D. (2008) A role for Q/N-rich aggregation-prone regions in P-body localization. *J Cell Sci* **121**: 2463–2472.

Salleron, L., Magistrelli, G., Mary, C., Fischer, N., Bairoch, A., and Lane, L. (2014) DERA is the human deoxyribose phosphate aldolase and is involved in stress response. *Biochim Biophys Acta* **1843**: 2913–2925.

Scherrer, T., Mittal, N., Janga, S.C., and Gerber, A.P. (2010) A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One* **5**: e15499.

Schmidlin, M., Lu, M., Leuenberger, S.A., Stoecklin, G., Mallaun, M., Gross, B., *et al*. (2004) The ARE-dependent mRNA-destabilizing activity of BRF1 is regulated by protein kinase B. *EMBO J* **23**: 4760–4769.

Schwede, A., Ellis, L., Luther, J., Carrington, M, Stoecklin, G., and Clayton, C. (2008) A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. *Nucleic Acids Res* **36**: 3374–3388.

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

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 Res* **42**: 4652–4668.

Sonnhammer, E.L., and Ostlund, G. (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res* **43**: D234–239.

Stoecklin, G., Stubbs, T., Kedersha, N., Wax, S., Rigby, W.F., Blackwell, T.K., and. Anderson, P. (2004) MK2-induced tristetraprolin:14-3-3 complexes prevent stress granule association and ARE-mRNA decay. *EMBO J* **23**: 1313–1324.

Subota, I., Rotureau, B., Blisnick, T., Ngwabyt, S., Durand-Dubief, M., Engstler, M., and Bastin, P. (2011) ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. *Mol Biol Cell* **22**: 4205–4219.

Tsvetanova, N.G., Klass, D.M., Salzman, J., and Brown, P.O. (2010) Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. *PLoS One* **5**: e12671.

Vasquez, J.J., Hon, C.C., Vanselow, J.T., Schlosser, A., and Siegel, T.N. (2014) Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res* **42**: 3623–3637.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. (2000) GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328**: 575–592.

Walrad, P.B., Capewell, P., Fenn, K., and Matthews, K.R. (2012) The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in *Trypanosoma brucei*. *Nucleic Acids Res* **40**: 2869–2883.

Weidner, J., Wang, C., Prescianotto-Baschong, C., Estrada, A.F., and Spang, A. (2014) The polysome-associated proteins Scp160 and Bfr1 prevent P body formation under normal growth conditions. *J Cell Sci* **127**: 1992–2004.

Wurst, M., Robles, A., Po, J., Luu, V.D., Brems, S., Marentije, M., *et al*. (2009) An RNAi screen of the RRM-domain proteins of *Trypanosoma brucei*. *Mol Biochem Parasitol* **163**: 61–65.

Wurst, M., Seliger, B., Jha, B.A., Klein, C., Queiroz, R., and. 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.

Zinoviev, A., and. Shapira, M. (2012) Evolutionary conservation and diversification of the translation initiation apparatus in trypanosomatids. *Comp Funct Genomics* **2012**: 813718.

Zinoviev, A., Leger, M., Wagner, G., and Shapira, M. (2011) A novel 4E-interacting protein in Leishmania is involved in stage-specific translation pathways. *Nucleic Acids Res* **39**: 8404–8415.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.