



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels

Citation for published version:

Fadda, A, Ryten, M, Droll, D, Rojas, F, Färber, V, Haanstra, JR, Merce, C, Bakker, BM, Matthews, K & Clayton, C 2014, 'Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels', *Molecular Microbiology*, vol. 94, no. 2, pp. 307-326. <https://doi.org/10.1111/mmi.12764>

Digital Object Identifier (DOI):

[10.1111/mmi.12764](https://doi.org/10.1111/mmi.12764)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Molecular Microbiology

Publisher Rights Statement:

© 2014 The Authors. Molecular Microbiology published by John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels

Abeer Fadda,^{1†} Mark Ryten,² Dorothea Droll,¹ Federico Rojas,³ Valentin Färber,^{1‡} Jurgen R. Haanstra,^{4,5} Clemetine Merce,¹ Barbara M. Bakker,^{4,5} Keith Matthews³ and Christine Clayton^{1*}

¹Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany.

²Flat 1, 172 Gloucester Terrace, London, UK.

³Centre for Immunity, Infection and Evolution, Institute for Immunology and Infection Research, School of Biological Sciences, Ashworth Laboratories, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.

⁴Department of Pediatrics and Systems Biology, Centre for Energy Metabolism and Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.

⁵Department of Molecular Cell Physiology, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands.

Summary

African trypanosomes are an excellent system for quantitative modelling of post-transcriptional mRNA control. Transcription is constitutive and polycistronic; individual mRNAs are excised by *trans* splicing and polyadenylation. We here measure mRNA decay kinetics in two life cycle stages, bloodstream and procyclic forms, by transcription inhibition and RNAseq. Messenger RNAs with short half-lives tend to show initial fast degradation, followed by a slower phase; they are often stabilized by depletion of the 5′–3′ exoribonuclease XRNA. Many longer-lived mRNAs show initial slow degradation followed by rapid destruction: we suggest that the slow phase reflects gradual deadenylation. Developmentally

regulated mRNAs often show regulated decay, and switch their decay pattern. Rates of mRNA decay are good predictors of steady state levels for short mRNAs, but mRNAs longer than 3 kb show unexpectedly low abundances. Modelling shows that variations in splicing and polyadenylation rates can contribute to steady-state mRNA levels, but this is completely dependent on competition between processing and co-transcriptional mRNA precursor destruction.

Introduction

The rates of transcription, RNA splicing, localization and decay all affect the levels of cellular mRNAs, while protein levels are additionally affected by the rates of translation and protein turnover. In mammalian cells, mRNA abundance makes a contribution of up to 40% to final protein levels (Schwanhäusser *et al.*, 2011). Results from high throughput techniques have revealed that decay rates are critical in overall mRNA homeostasis; genes encoding proteins with similar functions share similar transcript decay rates, and the decay rates respond to external stimuli (Shalem *et al.*, 2008; Munchel *et al.*, 2011; Rabani *et al.*, 2011; Schwanhäusser *et al.*, 2011).

Traditionally, decay rates have been measured by extracting mRNA at various times after transcription shut-off. More recently, a method involving pulse-chase with 4-thiouracil has enabled less invasive measurements, but this method is limited to cell types permissive to the label uptake and incorporation into newly synthesized RNA (e.g. Munchel *et al.*, 2011; Rabani *et al.*, 2011). Independently of the method used, transcriptome-wide decay studies can provide insight into the specificity (Grigull *et al.*, 2004; Mnaimneh *et al.*, 2004; Houalla *et al.*, 2006; Chekanova *et al.*, 2007; Kiss and Andrusis, 2010) and dynamics (Gudipati *et al.*, 2012; Harigaya and Parker, 2012; Schneider *et al.*, 2012; Deneke *et al.*, 2013) of the decay machinery. Some of these studies revealed the presence of feedback loops between synthesis and decay (Shalem *et al.*, 2008; Trcek *et al.*, 2011; Sun *et al.*, 2012; Haimovich *et al.*, 2013). In many cases, decay kinetics are not exponential; this can be explained by models in

Accepted 19 August, 2014. *For correspondence. E-mail cclayton@zmbh.uni-heidelberg.de; Tel. (+49) 6221 546876; Fax (+49) 6221 546891. Present affiliations: [†]Sidra Medical and Research Center, P.O. Box 2699, Doha, Qatar; [‡]Fresenius Kabi Deutschland GmbH, Bad Homburg, Germany.

which the 'age' of an mRNA influences its decay rate (Deneke *et al.*, 2013).

African trypanosomes are unicellular parasites with a mode of gene expression that is uniquely suited to the study of post-transcriptional mRNA control mechanisms. Transcription is polycistronic, with no apparent regulation of RNA polymerase II initiation (Palenchar and Bellofatto, 2006; Daniels *et al.*, 2010). Mature mRNAs are *trans* spliced, with addition of a capped spliced leader sequence at the 5' end (Michaeli, 2011); polyadenylation at the 3' end is determined by, and coupled to, the *trans* splicing of the next transcript downstream (Michaeli, 2011). Measurements that focused on *trans* splicing of the tubulin mRNA (in an *in vitro* system) (Ullu *et al.*, 1993) and the *PGKB* mRNA (*in vivo*) both found precursor half-lives of about 2 min (Haanstra *et al.*, 2008), which suggests that the mRNA may be spliced once the RNA polymerase is 1–2 kb further downstream (Haanstra *et al.*, 2008). Results for the *HSP70* tandem repeat were consistent with these estimates (Huang and van der Ploeg, 1991). However, it is not known whether this *trans* splicing rate is generally applicable.

Two trypanosome life cycle stages are routinely cultured in the laboratory: the bloodstream form, which is equivalent to the form in mammalian blood, and the procyclic form, equivalent to the form that grows in the midgut of the Tsetse vector. Most open reading frames have more than one possible spliced leader addition site, and often, several polyadenylation sites are used; moreover, the choice of both can vary between life cycle stages (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010; 2011; Zhang *et al.*, 2010). From RNASeq experiments, approximately 400 genes were judged to be developmentally regulated, with approximately equal numbers being preferentially expressed in each form (Siegel *et al.*, 2010). Numerous studies have shown that developmental regulation of individual mRNAs is associated with differences in their decay rates (Fernández-Moya and Estévez, 2010; Kramer and Carrington, 2011; Clayton, 2013). One example, that of *PGKB*, has been subjected to quantitative modelling. The *PGKB* mRNA is very unstable in bloodstream forms, and stable in the procyclic form. Measurement and modelling of its synthesis and degradation kinetics indicated that the level of the *PGKB* mRNA was predominantly controlled by the rates of transcription and degradation, and life cycle regulation was achieved by changes in the degradation rate (Haanstra *et al.*, 2008). The extent to which differences in mRNA degradation were responsible for developmental regulation on a transcriptome-wide scale was however unknown.

Degradation of trypanosome mRNAs follows the classic eukaryotic paradigm: the CAF1-NOT complex deadenylates the mRNA (Erben *et al.*, 2014), and this is followed by 5'–3' degradation by XRNA (Li *et al.*, 2006) or

3'–5' degradation by the exosome (Clayton and Estevez, 2010). The PAN2/PAN3 deadenylase also has a role in degradation of some mRNAs, and there is, in addition, a deadenylation-independent pathway that is initiated by 5' degradation (Schwede *et al.*, 2009; Fadda *et al.*, 2013). Studies of mRNA degradation currently rely on transcription inhibition. Unfortunately, pulse-labelling with 4-thiouracil is not feasible: not only is there hardly any uptake, but there is also a bypass pathway for uridine synthesis (Ali *et al.*, 2013) and a very strong flux of the nucleotide into glycosylation of surface proteins and other pathways (Ali *et al.*, 2012). We previously obtained approximate estimates of bloodstream-form trypanosome mRNA half-lives by inhibiting transcription for 30 min. We demonstrated that for most mRNAs, deadenylation precedes degradation (Manful *et al.*, 2011; Fadda *et al.*, 2013) and also found that the 5'–3' exoribonuclease XRNA was important mainly for degradation of very unstable mRNAs (Manful *et al.*, 2011). This confirmed previous observations made with individual mRNAs (Schwede *et al.*, 2008; 2009). The analyses were however limited by the use of a single time-point for transcription inhibition, meaning that half-life estimates were very approximate and complex kinetics could not be detected. Also, procyclic trypanosomes were not studied.

The work described here comprises part of a study that aims to model trypanosome metabolic pathways from genome to metabolome. To allow quantitative modelling, as many parameters as possible must be measured in trypanosomes cultured under tightly defined conditions (Bakker *et al.*, 2010). In this paper we describe a detailed analysis of mRNA decay in both bloodstream form and procyclic trypanosomes, including complex decay kinetics. We find extensive variations in mRNA decay, both between mRNAs and between life cycle stages, but these are usually insufficient to account for steady-state mRNA abundance. By modelling, we uncover a potentially important role for competition between mRNA processing and co-transcriptional RNA destruction in determining steady-state mRNA levels.

Results

Developmental regulation of steady-state mRNA levels

In this work we aimed to measure the half-lives of all trypanosome mRNAs in both bloodstream and procyclic forms, and to obtain an overview of *trans* splicing kinetics. To measure mRNA decay we inhibited *trans* splicing and transcription for various times, using Sinefungin to prevent splicing and Actinomycin D to stop transcription; to measure splicing we used Actinomycin D alone. We prepared RNA, depleted rRNA by subtraction and then measured relative RNA amounts with RNAseq. All

samples are listed in Supplementary Table S1, which includes the raw data. Supplementary Table S2 summarizes the main results. Details are in subsequent Supplementary Tables. Most analyses concentrate on a non-redundant list of genes in which repeated copies have been removed (Siegel *et al.*, 2010).

We first analysed steady state mRNA levels, comparing results for rRNA-depleted mRNA from four bloodstream form samples and four procyclic samples. Statistically significant differential expression was seen for 693 genes, of which 430 transcripts were more abundant in bloodstream forms (Supplementary Table S3, sheets 1 & 2). We compared these results with previously published data for poly(A) + mRNA (Siegel *et al.*, 2010; Singh *et al.*, 2014; Vasquez *et al.*, 2014) (Supplementary Figs S1 and S2, and Supplementary Table S3, sheet 3). There were clear differences between the datasets, which could have been caused by differences in the trypanosomes used, the culture conditions, the methods used for RNA preparation and sequencing, and the tools used for data extraction and analysis. This illustrates the fact that for quantitative work, it is vital to control all conditions and methods as tightly as possible.

The comparisons enabled us to compile a list of the 450 most reproducibly developmentally regulated mRNAs (at least twofold regulated in at least 3 of the 4 available RNA-Seq datasets, Supplementary Table S3, sheet 4). We grouped these according to manually annotated functional categories. Rather surprisingly, almost twice as many mRNAs were reproducibly upregulated in bloodstream forms than were upregulated in procyclics; and oddly, genes with no known function accounted for all of the difference. Figure 1 shows just the upregulated genes that could be functionally assigned. The regulation of energy metabolism was as expected – mitochondrial pathways present in procyclic forms, and reliance on glycolysis in bloodstream forms (Bringaud *et al.*, 2006). Bloodstream forms are also known to have a high requirement for vesicular transport and GPI anchor biosynthesis (Overath and Engstler, 2004). Bloodstream-form-specific expression of genes encoding cytoskeletal proteins, protein kinases and phosphatases, and RNA-binding proteins was less expected. Levels of RNA, however, do not always correlate with levels of protein (Gunasekera *et al.*, 2012; Urbaniak *et al.*, 2013). Translation (Vasquez *et al.*, 2014) and protein degradation are also important.

Our procyclic forms were grown with proline as the principal carbon source, with glucose only from the diluted serum. In contrast, the procyclic forms used by others (Siegel *et al.*, 2010; Vasquez *et al.*, 2014) were grown in medium with 5 mM glucose. Thirteen mRNAs reproducibly showed at least a threefold greater increase in our procyclics [ribo- and poly(A) + datasets] than was seen in the RNASeq data for glucose-grown cells (Supplementary

Table S3, sheet 3, columns O–R; names in bold in column C). The increased expression of proline dehydrogenase (Tb927.7.210) has been reported previously (Lamour *et al.*, 2005), and glutamine synthetase (Tb927.7.4970) is next on the same pathway (Bringaud *et al.*, 2006). The mRNAs encoding glycosomal malate dehydrogenase (Tb927.10.15410), ubiquinol-cytochrome *c* reductase (Tb927.4.4990) and an aquaporin (Tb927.6.1520) are also upregulated in proline-cultured cells. Specifically increased mRNAs encoding proteins of unknown function were Tb927.4.3500, Tb927.9.13200, Tb927.11.1690 and Tb927.10.2360. This shows again that carbon source availability affects expression levels in trypanosomes.

The heterogeneity of trans splicing and polyadenylation sites precludes transcriptome-wide rate measurement of processing kinetics

In order to model gene expression, we needed an estimate of the variability in pre-mRNA splicing rates. In an attempt to obtain this, we treated cells with Actinomycin D alone for the shortest feasible time, 5 min, and subjected rRNA-depleted RNA to RNASeq. Since the three previously measured splice times were in the range of 1–2 min, time intervals in that range would have been preferable, but were not technically feasible. For the splicing measurements we used two replicates for bloodstream forms and one for procyclic forms, together with one untreated sample for each. Reads aligning to intergenic regions within a transcription unit come from un-spliced mRNA. *Trans* splicing involves *trans* esterification followed by destruction of the intergenic RNA by exoribonucleases. Most trypanosome mRNAs, however, have multiple possible splice sites (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010; 2011; Zhang *et al.*, 2010), which means that many RNA sequences can be either intergenic or within an mRNA. We collected all documented splice acceptor sites for both procyclic and bloodstream forms from TriTrypDB, and merged them with our own identified sites. For each site, the read counts in the 20mer upstream region were extracted, normalized to the read counts for the corresponding gene, and the ratio between untreated and Actinomycin D-treated samples was calculated. We then calculated the pre-mRNA half-life for each site, assuming exponential kinetics. If an open reading frame had more than one splice site, we chose the splice site with the shortest half-life to infer splicing kinetics for that open reading frame. 3029 measurements of 5' splicing for bloodstream forms and 1554 for procyclic forms are shown in Supplementary Table S2 and displayed in Supplementary Fig. S3. The measurements were limited by the numbers of reads, as well as by the use of only a single time interval. The median 5' pre-mRNA half-life measured in both life stages was a little

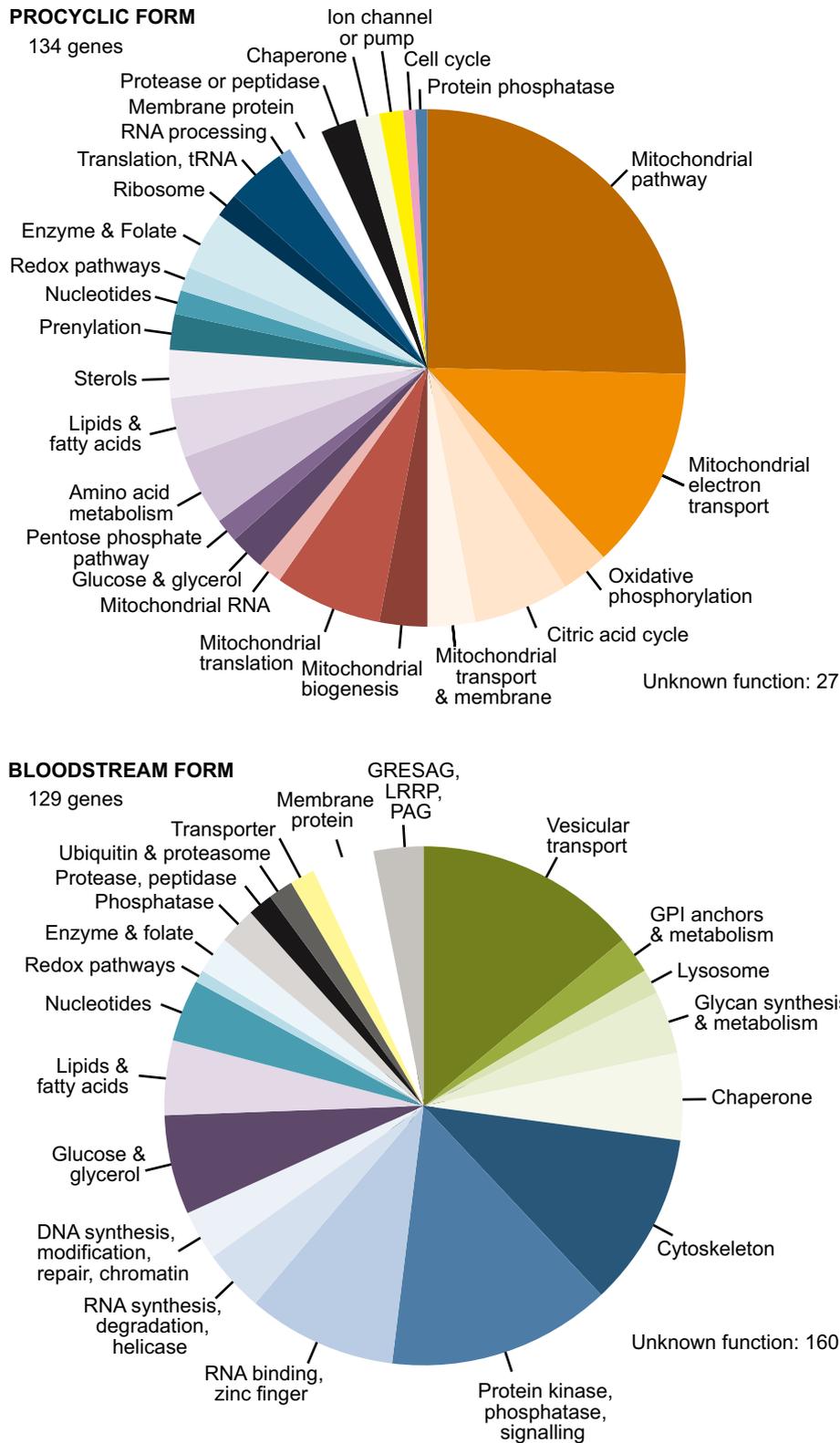


Fig. 1. Developmentally regulated mRNAs.

For each stage, genes that were upregulated in at least three of the four available RNASeq datasets were selected (Supplementary Table S3, sheet 4) and grouped according to manually assigned functional categories. Genes encoding proteins of no known function (27 for procyclic forms, 160 for bloodstream forms) are not shown. 'Upregulated' means at least twofold more abundant than in the other stage; for our current dataset it also means significantly regulated.

over 3 min, and many of the measured half-times were between 1–2 min. Since these times are similar to those measured for tubulin, HSP70 and PGK mRNAs, they may be correct. However, there was no correlation in 5' splicing times between the two life stages even for genes sharing the same major acceptor site in the two stages (not shown). Also, many precursor half-lives were over 4–5 min. Although this could reflect slow splicing, it could also indicate persistence of bicistronic RNAs, or of monocistronic mature mRNAs that had upstream *trans* splice sites. The un-spliced *PGK* precursor is degraded with a half-life of 8 min (Haanstra *et al.*, 2008) so half-lives above this are unlikely to reflect splicing, but may instead reflect precursor degradation. It was thus apparent that for most mRNAs, the complexity of trypanosome mRNA splicing patterns precluded reliable kinetic analysis using the high-throughput method that we had adopted. Our results therefore could not be used to assess the kinetics of splicing. Polyadenylation sites are even more heterogeneous so we made no attempt to analyse 3'-processing kinetics.

Despite the limitations in the results, they did suggest that many trypanosome mRNAs might be spliced within a few minutes of synthesis. Based on this, we adopted a conservative splicing half-time of 2 ± 1.5 min in the modelling (below).

Trypanosome mRNAs show complex decay kinetics

We next analysed the data for parasites treated with both Sinefungin and Actinomycin D, using multiple time intervals and at least two replicates for each time point. To normalize the data we hybridized Northern blots with a probe complementary to the spliced leader (see Methods and Supplementary Fig. S4). The spliced leader provides a good measure of the number of mRNAs present, because it is present at the 5'-end of every mRNA. Unlike the poly(A) tail, it does not vary in length; and it will be removed by XRNA very fast after the mRNA has been decapped.

To calculate half-lives, we considered three possible scenarios: simple exponential decay, or one of two non-exponential patterns (Deneke *et al.*, 2013). The non-exponential models represent a multi-step degradation process in which different mRNA molecules are in different stages of degradation; hence the 'age' of the mRNA population (i.e. the length of time after transcription has stopped) plays a role in the shape of overall decay curve (Deneke *et al.*, 2013). In the first scenario, the mRNA population initially seems relatively stable but becomes increasingly unstable over time: we call this the 'slow-fast' scenario. For example, the actin mRNA shows little degradation over the first 10 minutes, but then the rate increases (Fig. 2A). In the second scenario, part of the RNA is degraded very rapidly, but a fraction appears to be

Table 1. Summary of half-life results.

Category	Bloodstream	Procyclic
Slow-fast	1903	2632
Fast-slow	3074	1622
Exponential	1495	1704
Stable	92	149
Not measured	480	993

Results given are for 6951 unique open reading frames. mRNAs that had half-lives over 120 min (bloodstream forms) or 240 min (procyclic forms) were mostly fitted using the exponential model, but were classified as stable since the half-life could not be measured well using the time points that we had. 'Not measured' usually means that the numbers of counts were insufficient to calculate a half-life and/or that the data could not be fitted with any of the three models.

much more stable (fast-slow scenario). An example of this is shown in Fig. 2B. The model that fit the data best (smallest least-squares error) was used to calculate a single transcript lifetime and half-life for each mRNA and the resulting half-lives (Supplementary Table S4) were used in all further analyses. The results are summarized in Table 1. Sometimes the fit was always poor, especially when there was extensive scatter at early time points (Supplementary Fig. S5A and B), while in other cases, moving from an exponential model to one of the other models clearly gave a better fit (Supplementary Fig. S5C and D). As noted previously (Deneke *et al.*, 2013), transcripts with short half-lives tended to fit the fast-slow model, while those with long half-lives tended to fit a slow-fast model (Fig. 2C and D).

Some mRNAs showed no significant degradation over the incubation period, and for some, the abundance was too low for a half-life measurement. However, the use of multiple time points and complex kinetic models enabled us to estimate half-lives for most mRNAs. Bloodstream trypanosome mRNAs had a modal half-life of 9 min, median 12 min, and arithmetic mean 16 min (Fig. 3), similar to the numbers that we previously reported (Manful *et al.*, 2011). Procyclic trypanosomes are larger than bloodstream forms, grow at 27°C, have about twice as many mRNAs per cell and divide more slowly (Haanstra *et al.*, 2008); correspondingly, their mRNAs had a modal half-life of 12 min, median of 20 min, and arithmetic mean of 34 min (Fig. 3, Supplementary Table S3 sheet 5).

RNASeq half-lives are similar to those found by other methods

We previously had estimated half-lives for bloodstream forms by measuring the amount of mRNA remaining after 30 min of transcription inhibition (Manful *et al.*, 2011). The correlation between our new and old datasets was best for mRNAs with longer half-lives (Supplementary Fig. S6). One likely reason for this is that the reliability of RNASeq

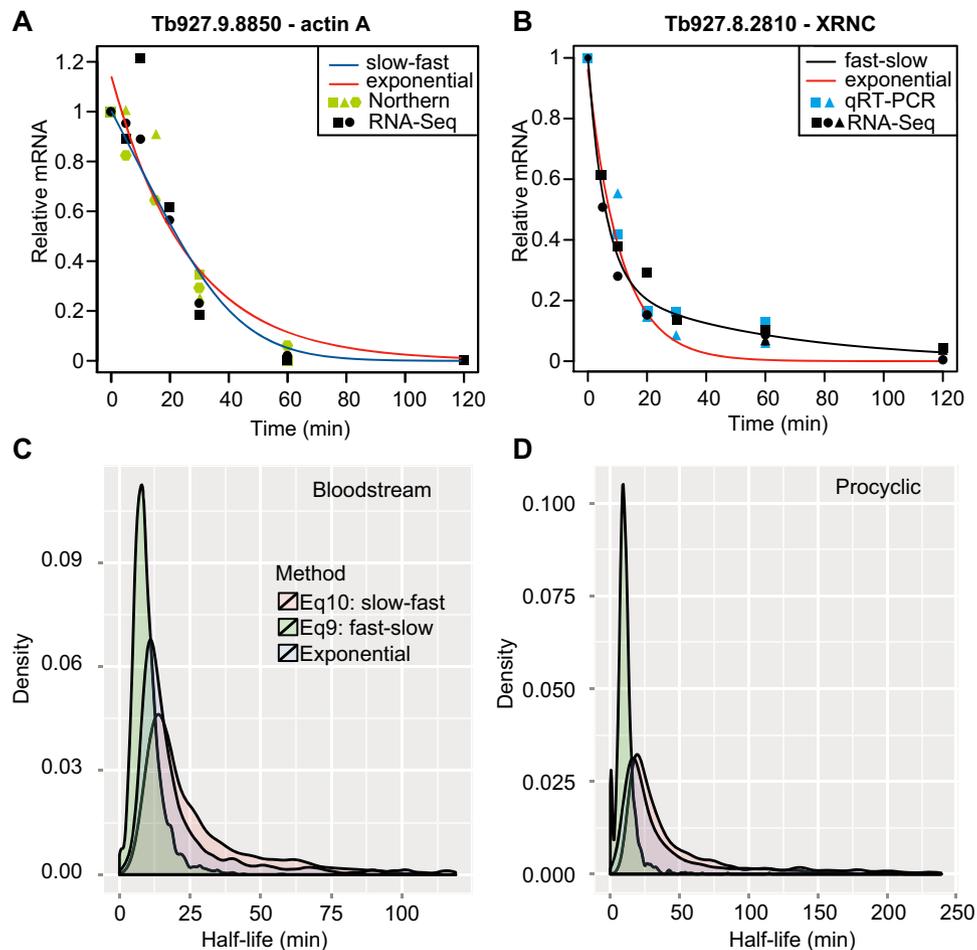


Fig. 2. Kinetics of mRNA decay.

A. Degradation of the mRNA encoding Actin A in bloodstream forms. RNASeq data are plotted in black, the exponential curve fit is in red, and the slow-fast curve fit is in blue. The slow-fast curve matches better, but is still not ideal. The green points are Northern blotting results (Schwede *et al.*, 2008; 2009).

B. Degradation of the *XRNC* mRNA in bloodstream forms fits the fast-slow model. Details as for (A), except that the cyan points are real-time qRT-PCR data (summarized in Manful *et al.*, 2011).

C. The numbers of mRNAs fitting each decay model in bloodstream forms were used to calculate and plot probability density functions for half-life values. The area under the curve represents the probability of having a range of half-life values. Each open reading frame is represented once, with the best-fitted model.

D. As for (C), but for procyclic forms.

decreases at low read densities, and unstable RNAs give very few reads after 30 min of degradation.

The half-lives that we measured by RNA-Seq were in quite good agreement with previously reported half-lives measured by real-time PCR or Northern blotting, with correlation coefficients (for log-transformed data) of over 0.8 (Fig. 2A and B; Fig. 4, Table 2). When we compared the RNASeq results with our previous measurements made using the same protocols for RNA synthesis inhibition, differences in the kinetics were rather minor (Supplementary Fig. S7). RNA-Seq might be somewhat more sensitive at the later time-points, when RNA levels dropped even below the levels that could be measured reliably by q-PCR. Disagreements with literature values

(Table 2) could be due to differences in the cell culture conditions as well as in methodology.

Changes in mRNA decay partially account for developmental regulation

We next asked whether mRNAs encoding proteins with particular functions showed similar half-lives. We grouped mRNAs into functional classes (Supplementary Table S2) and plotted half-lives for each class (Fig. 5). A one-way ANOVA showed significant life-time differences between the different classes (P -value $\ll 0.01$), suggesting coordinate regulation (Supplementary Table S5). Transcripts encoding ribosomal proteins are particularly stable in both

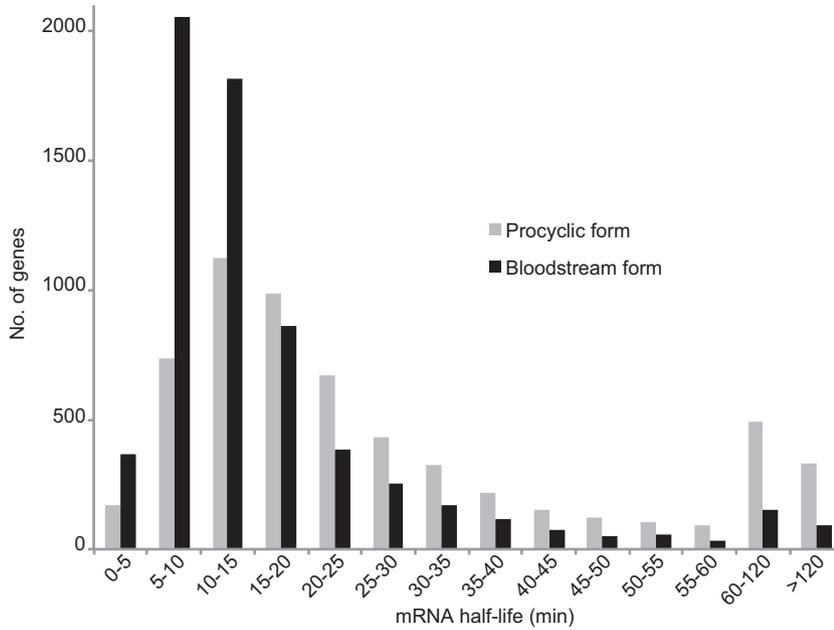


Fig. 3. Most trypanosome mRNAs have half-lives of less than 30 min. All mRNAs with measurable half-lives (stable ones are excluded) were placed into classes according to half-life range. The numbers of mRNAs in each class are plotted. All data are for ribo-minus RNA preparations.

stages: large amounts of mRNA are produced from only 1–3 open reading frames (Manful *et al.*, 2011). Although all RNA classes were overall more stable in the procyclic stage (T-test results in Supplementary Table S5), more pronounced shifts were apparent for functions related to procyclic energy metabolism, such as amino acids, citric acid cycle, mitochondrial biogenesis and oxidative phosphorylation. The small number of mRNAs involved in folate and sterol metabolism also showed greater increases in half-life than average. These results were consistent with the idea that developmental regulation is, at least in part, mediated through changes in half-life.

To analyse the relationship further we plotted half-life ratios (bloodstream half-life/procyclic half-life) against the abundance ratios (bloodstream abundance/procyclic abundance). There was almost no correlation, except for very strongly regulated mRNAs (Supplementary Fig. S8). Moreover, not all mRNAs with regulated turnover were regulated in abundance. Restriction of the dataset to the mRNAs that were rapidly spliced in both forms, or using only those mRNAs that had the same splice site in both forms, did not increase the correlation (not shown). However, this analysis almost certainly under-estimated the contribution of mRNA turnover to developmental

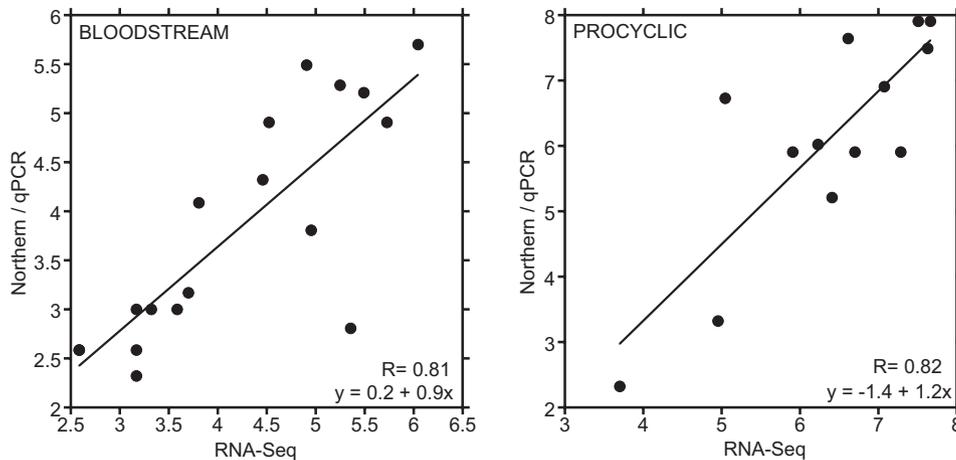


Fig. 4. Comparison between RNASeq and other methods.

All available RNA half-lives measured by qRT-PCR or Northern blotting were compared with the half-lives measured by RNA-Seq. Both axes show the log (base 2) of the half-life in min. References are in Table 2. If more than one Northern measurement was available the arithmetic mean was used. Only measurements with sufficient time points were included: for example, the numbers designated as 'more than' or 'less than' a certain half-life in Table 2 were excluded. The formulae and correlation coefficients were calculated with log₂-transformed numbers.

Table 2. Comparison of RNASeq half-lives with Northern blot results.

	Mantul BS	BS new	BS Northern/ qPCR	PC new	PC Northern	Ref
Not regulated						
Tb927.5.4170	68	45#	30–45	stable	> 60 min	PC (np); BS (Schwede et al., 2009) (Fåber et al., 2013)
Tb927.9.8850	28	22#	15–25	33#	~ 90, 52#, 70#	BS e.g. (np), (Holz et al., 1997); PC (np), (Estévez, 2008; Kramer et al., 2008)
Tb927.1.2330	70	66#	45–60#	135#	~ 120	(Manful et al., 2011) # (Kramer et al., 2008; Schwede et al., 2008)
Tb927.7.3780	9	9	8	15#	na	(Manful et al., 2011)
Tb927.8.2810	11	6	6	9	na	(Manful et al., 2011)
Tb927.4.410	26	23#	30	46#	na	(Manful et al., 2011)
Tb927.11.16600	9	12	8	5	na	(Manful et al., 2011)
Tb927.11.16310	16	31	14	11	na	(Manful et al., 2011)
Tb927.11.4710	9	10	8	13	na	(Manful et al., 2011)
Tb927.11.790	11	6	6	18	na	(Manful et al., 2011)
Tb927.11.8710	10	9	6	9	na	(Manful et al., 2011)
Tb927.6.2810	11	9	5	16	na	(Manful et al., 2011)
Tb927.7.6980	15	14	17	28	na	(Manful et al., 2011)
Tb927.6.4340	10	13#	9	15	na	(Manful et al., 2011)
Tb927.11.11330	21	38#	39	102	na	(Manful et al., 2011)
Tb927.6.3740	63	55	na	156	~ 60	(Kelly et al., 2012)
Tb927.11.5050	23	36	na	85#	37	(Estévez, 2008)
Tb927.11.10920	11	12	na	31	~ 10	(Stern et al., 2009)
Tb927.10.12700	53	36	na	stable	45	(Stern et al., 2009)
Bloodstream-specific						
Tb927.10.7090	stable	stable	stable	na	3.2 h	(Chaudhuri et al., 2002)
Tb927.10.5620	stable	stable	90#	182	~ 4 h#	(Mayho et al., 2006)
Tb927.2.6000	stable	53#	30#	13	5	(Webb et al., 2005)
Procytic-specific						
Tb927.10.10250	na	na	5–10	98	200	(Holz et al., 1997) and others
Tb927.1.4100	14	21#	< 30	216	> 4 h#	(Mayho et al., 2006)
Tb927.9.3170	15	9#	< 30	203	~ 4 h	(Mayho et al., 2006)
Tb927.10.280	15	11#	< 30	198	~ 3 h	(Mayho et al., 2006)
Tb927.3.1410	26	12	< 30	stable	~ 6 h#	(Mayho et al., 2006)
Tb927.4.4620	17	5	< 30	stable	~ 6 h#	(Mayho et al., 2006)
Tb927.10.8320	26	14#	< 30	stable	~ 8 h#	(Mayho et al., 2006)
Tb927.7.7420	17	22#	< 30	199	> 60#	(Brown et al., 2001)
Tb927.3.1380	32	27#	< 30	152	> 90#	(Brown et al., 2001)
Tb927.5.1710	30	22#	< 30	stable	> 240	(Brown et al., 2001)
Tb927.6.2170	38	30#	45	130#	na	(Manful et al., 2011)
Tb927.11.6280	36	41	7**	stable	na	(Haile et al., 2003)
Tb927.4.4730	18	28#	na	104#	82#, 40#	# (Estévez, 2008), (Stern et al., 2009)
Tb927.7.4970	12	10	na	75#	65#	(Estévez, 2008)
Tb927.10.10770	31	29	na	60#	60	(Estévez, 2008)
Tb927.3.4500	16	38	na	stable	116	(Estévez, 2008)
Tb927.6.1520	81	41#	na	stable	~ 80	(Stern et al., 2009)

Mantul BS – Bloodstream-form RNASeq results from (Manful et al., 2011). BS new, PC new: results from the current analysis; BS = bloodstream forms, PC = procytic forms. BS or PC Northern: Estimates from Northern blots, using data from the references cited on the right hand column. Note that for many of these, estimates of accuracy are not available and the methods for inhibiting transcription varied. All half-lives are in minutes. Non-standard gene abbreviations are: b-TUB – beta tubulin; Myosin HC – myosin heavy chain; Hyp – hypothetical open reading frame encoding protein of unknown function ABC – ABC transporter; PAP – poly(A) polymerase-like; ATP synth – ATP synthase (subunits a-alpha, b-beta, and 9). Values with asterisks are for reporter mRNAs with the 3'-UTR of the gene concerned, synthesized by RNA polymerase I (*) or T7 polymerase (**). Transcription by T7 polymerase might affect the half-life (Colasante et al., 2007). (np) – our data, not published. #Slow-fast kinetics.

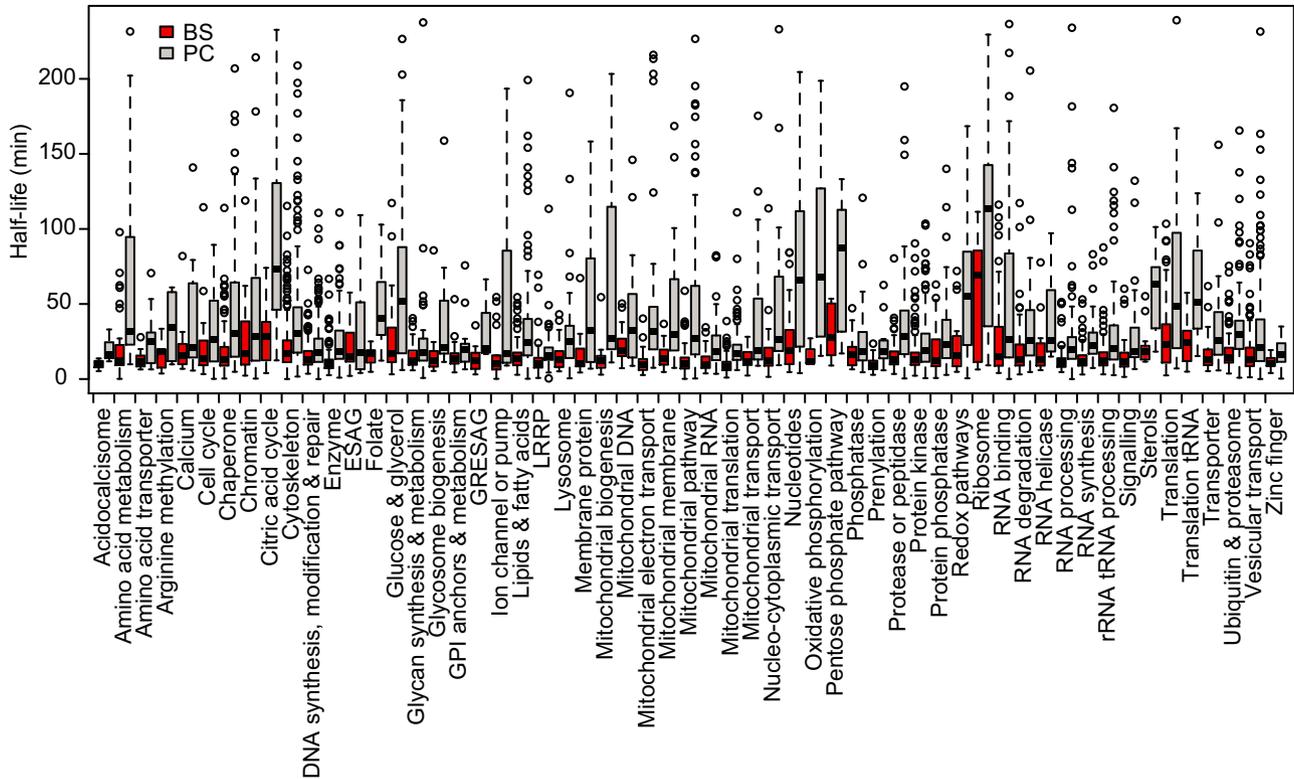


Fig. 5. Half-lives of mRNAs in particular functional classes.

Unique coding regions were placed into functional classes based on their genome annotation, user comments and publications. The classes used are indicated in Supplementary Table S5. In this box-and-whisker plot the bottom and top of the box are the first and third quartiles, the band inside the box is the median, the whiskers represent the lowest datum still within 1.5 of the lower quartile, and the highest datum still within 1.5 of the upper quartile and the circles represent outliers.

regulation. We had to exclude not only all stable mRNAs, but also those for which abundance in one stage was too low for the half-life to be measured. The latter included 29 procyclic-form-specific and 172 bloodstream-form-specific mRNAs (Supplementary Table S3, sheet 3). Excluding these, we found 179 mRNAs in which mRNA decay played a clear role in regulation (Supplementary Table S3, sheet 6). We saw regulated mRNA decay not only for components of the cytochrome oxidase (Mayho *et al.*, 2006) and ATP synthase (Brown *et al.*, 2001) complexes (Table 2), but also for 67 other mitochondrial proteins, the AATP11 amino acid transporter, and 9 glycosomal or cytosolic enzymes involved in procyclic energy metabolism. For bloodstream form-specific expression, we detected regulation for the GPI-phospholipase C (Webb *et al.*, 2005); the mRNAs encoding fructose biphosphate aldolase and the alternative oxidase were relatively stable in procyclic forms but showed no decay at all during our assay in bloodstream forms (Table 2).

Overall, more mRNAs fit a slow-fast model in procyclic than in bloodstream forms (Table 1), and no fewer than 1110 transcripts had different decay models in the two life stages. 976 mRNAs switched from fast-slow in blood-

stream forms to slow-fast in procyclics. These were particularly enriched for mitochondrial processes that are developmentally upregulated in procyclics: mitochondrial translation, RNA processing, electron transport, biogenesis and other pathways. Other classes that showed this switch were prenylation, RNA processing and the catch-all 'enzyme' group. The 134 that switched the other way were mostly upregulated in bloodstream forms and were enriched for GPI anchors and metabolism, vesicular transport, calcium regulation and DNA synthesis, modification and repair.

Many mRNAs are less abundant than expected from their half-lives

Since trypanosome mRNAs are constitutively transcribed, gene copy number and mRNA half-life are expected to make a significant contribution to steady-state mRNA abundance. We therefore normalized mRNA copy numbers per cell to the gene copy number, then compared these values with the half-life. Results suggested that half-life contributes to, but cannot completely account for, the mRNA level (Fig. 6). High abundance was

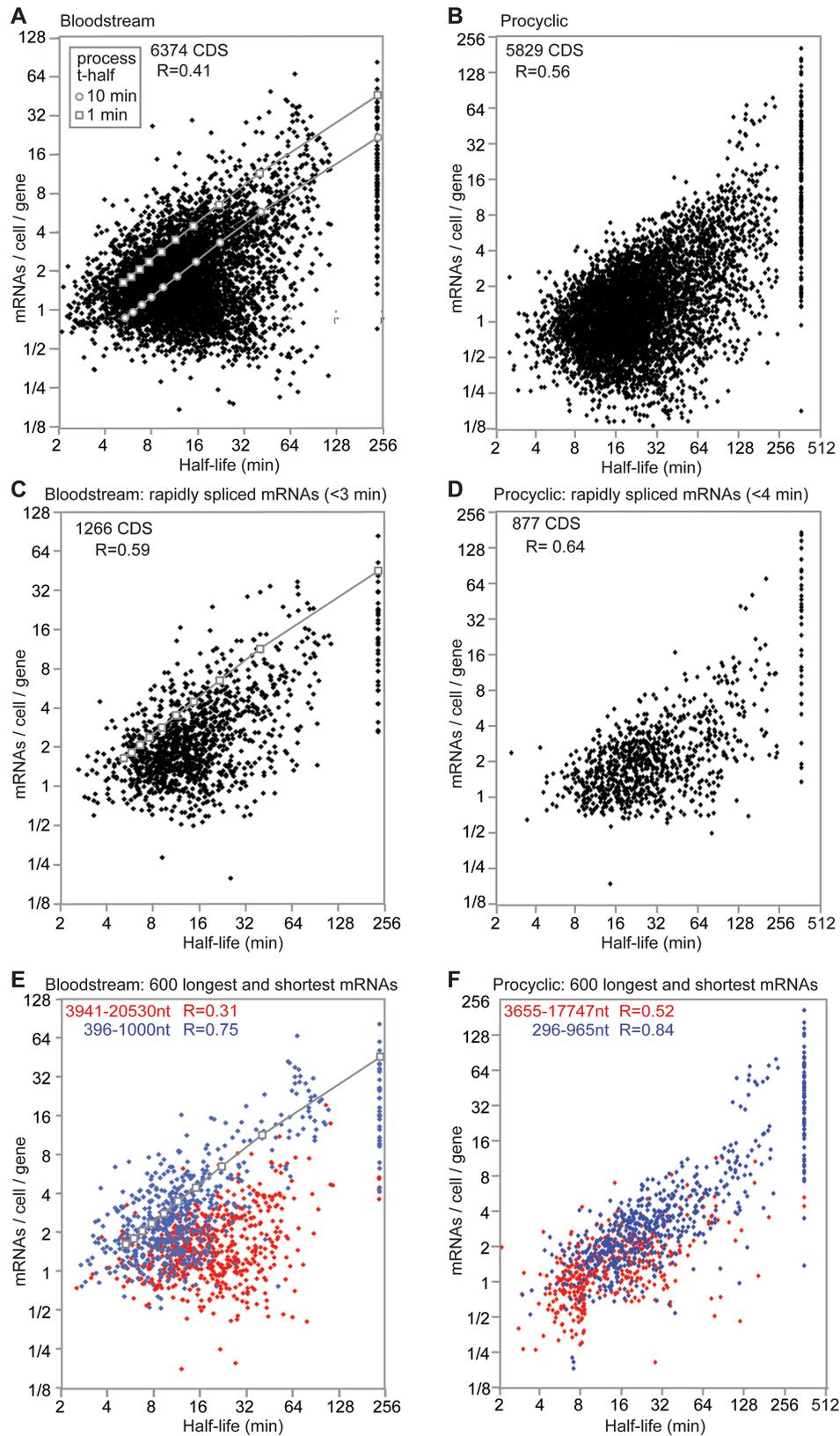


Fig. 6. Relationship between half-life and mRNA abundance.

- A. Relationship between the number of mRNAs per cell per gene and the half-life for 6367 non-redundant open reading frames in bloodstream forms. All mRNAs with apparent half-lives of less than 2 min were excluded since (given the 5 min Sinefungin treatment) this is outside the range of the assay. The mRNAs that appeared stable have been arbitrarily assigned half-lives of 240 min. The Pearson correlation coefficient (R) was calculated from log-transformed data. The grey squares indicate the results of modelling steady-state mRNA abundance in the published model (Haanstra *et al.*, 2008) assuming a splicing half-time of 1 min; the circles are for an assumed 5'-*trans* splicing half-time of 10 min. The gene copy numbers were estimated from random shotgun sequences of genomic DNA.
- B. As for (A), but for 5831 non-redundant open reading frames in procyclic forms. The mRNAs that appeared stable have been arbitrarily assigned half-lives of 360 min; their median number of transcripts per gene per cell is 9.
- C. As for (A) but including only mRNAs with precursor half-times of less than 3 min.
- D. As for (B) but including only mRNAs with precursor half-times of less than 4 min.
- E. As for (A), but including only the 600 longest (red) and shortest (blue) mRNAs. The length ranges and correlation coefficients are also shown for the two groups. The grey squares indicate the results of modelling steady-state mRNA abundance in the published model (Haanstra *et al.*, 2008) assuming a splicing half-time of 1 min. Lengths were taken from TritypDB; all mRNAs lacking annotated poly(A) sites were excluded.
- F. As for (B), but including only the 600 longest (red) and shortest (blue) mRNAs.

possible only if the mRNA was relatively stable: for example, nearly all mRNAs present at more than 16 copies/cell/gene in bloodstream forms or 32 copies per cell in procyclics had half-lives longer than 30 min.

We previously developed a model for gene expression in bloodstream trypanosomes, based on detailed measurements for the *PGK* and tubulin genes (Haanstra *et al.*, 2008). We expanded this model in order to account for our new results (Fig. 7A, Table 3). In the published model, the first step is transcription; we had set the overall rate (initiation plus elongation) so that it was sufficient to produce the measured steady state level of mRNA (step 1 in Fig. 7A). The next step (step 2 in Fig. 7A) is 5' splicing, with a fixed, half-time of 1 min. Splicing competes with precursor degradation rate (step 4 in Fig. 7A); for this, we used a half-life of ~8.7 min, as determined for the *PGK* precursor (Haanstra *et al.*, 2008). Mature mRNA degradation (step 6 in Fig. 7A) includes degradation in both the nucleus and the cytoplasm. The model also includes dilution of RNA species via parasite growth (specific growth rate, μ). To start the new analysis, we used this published model to calculate the expected amounts of mRNA at steady state for various half-lives (grey squares, Fig. 6A). We found that they were towards the upper edge of the cloud of data points. This indicates that the published model predicts almost the maximum amount of mRNA

that can be obtained per gene *in vivo*. Notably, however, the levels of many mRNAs fell well below the estimates from the published model. The first, and most obvious, explanation was that we had not allowed for the possibility that splicing rates might vary. We therefore went back to our measurements, and plotted the half-lives and abundances of mRNAs with relatively short measured 5' precursor half-lives (Fig. 6C and D): although the correlation was marginally better than for the bulk population, this could be an artefact of the lower sample size. Returning to the published model, we plotted the RNA abundances for a 5' *trans* splicing reaction half-time of 10 min (grey circles, Fig. 6A). This still failed to account for the unexpectedly low abundances of many mRNAs. Obviously, something was missing from the published model. Our mRNA decay rates (Fig. 7A, step 6) already included both intra-nuclear and cytoplasmic degradation of the mature mRNA, so differences in nuclear export of mRNA could not explain the discrepancies between the model and reality. We identified transcription, polyadenylation and precursor decay as possible variables.

Modelling predicts an important role for co-transcriptional mRNA decay

The first issue that we needed to address was the assumption that the transcription rate is constant within, and between, transcription units. It has previously been shown that genes that encode mRNAs that persist after heat shock are located towards the end of transcription units (Kelly *et al.*, 2012). This presumably enables synthesis of the mRNAs to continue for some time after stress-induced shut-off of transcription initiation (Kelly *et al.*, 2012). The distance of a gene from the transcription start site also bore a relationship with cell cycle control (Kelly *et al.*, 2012). We therefore considered the possibility that the position of the gene within the transcription unit was having an effect on the mRNA abundance at steady state. To assess this, we plotted the half-life : abundance ratios for each gene across the chromosomes. The plot for chromosome 10 is in

Table 3. The reactions in the model and their rate equations.

Rate	For reaction	Rate equation or value
v_1	Transcription into precursor	0.24 min^{-1}
v_2	5' <i>trans</i> -splicing of unspliced precursor	$k_2 \times [\text{unsplicedprecursor}]$
v_3	3' poly-adenylation	$k_3 \times [5'\text{splicedprecursor}]$
v_4	Degradation of unspliced precursor	$(k_4 + \mu) \times [\text{unsplicedprecursor}]$
v_5	Degradation of 5' <i>trans</i> -spliced precursor	$(k_5 + \mu) \times [5'\text{splicedprecursor}]$
v_6	Degradation of mature mRNA	$(k_6 + \mu) \times [\text{mRNA}]$

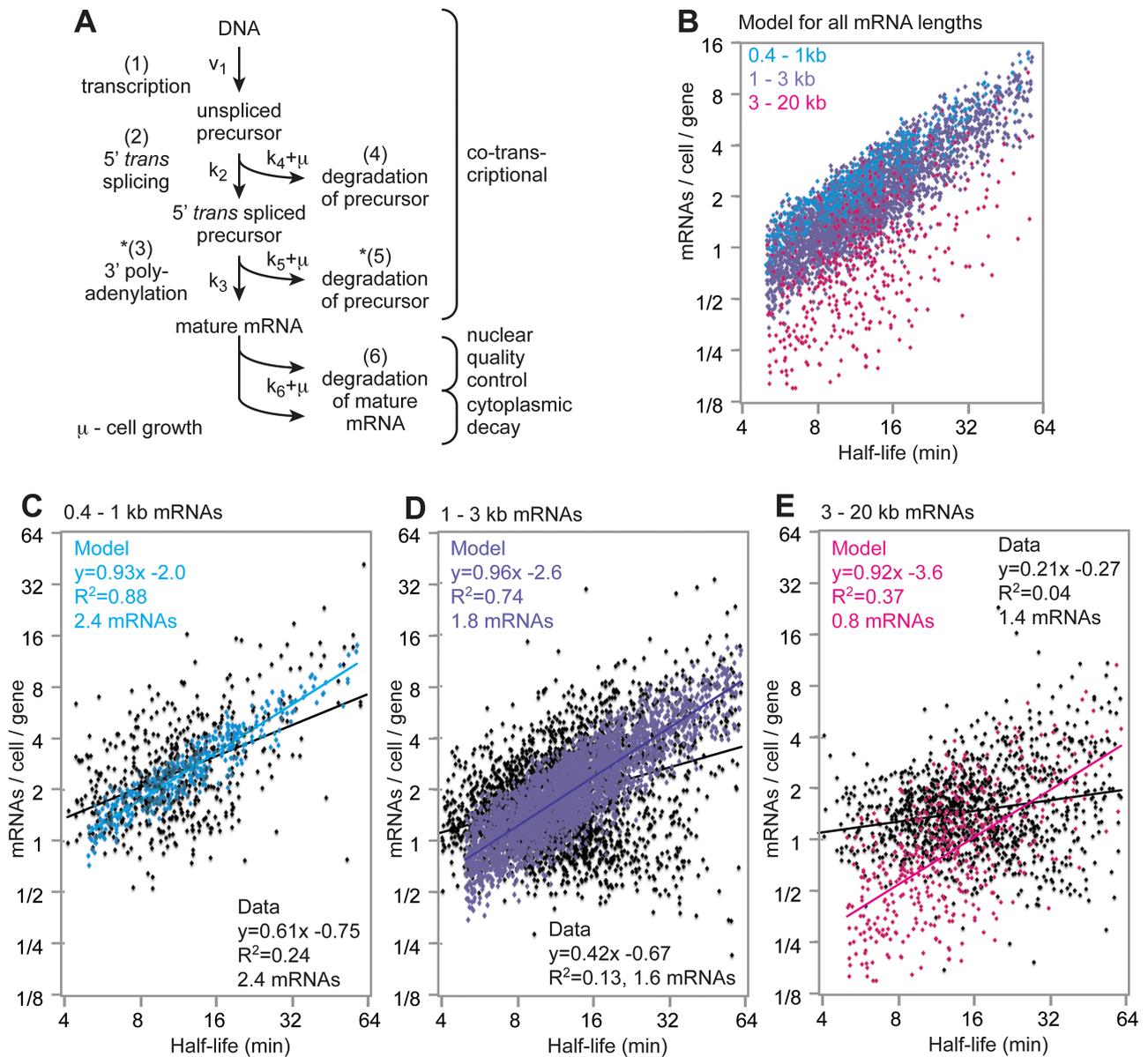


Fig. 7. The kinetic model for trypanosome gene expression.

A. Cartoon showing the main aspects of the model. *Steps 3 and 5 are extensions compared to (Haanstra *et al.*, 2008).

B. Trypanosome mRNA behaviour was modelled for three different groups: 550 mRNAs of 0.4–1, 4408 of 1–3 and 551 of 3–20. This distribution approximates to that seen in our dataset. For each set, half-lives were distributed as in Fig. 4 up to 120'. The amount of mRNA at steady state was then modelled using the following parameters: Transcription rate 0.24 molecules cell⁻¹ min⁻¹; Splicing and polyadenylation half-lives were sampled from a normal distribution around $2 \pm 1.5'$. The constant for degradation of the 5' spliced precursor k_5 (0.08 min⁻¹) was multiplied by the ratio (mRNA length in nt)/(600 nt). Shown is the mature mRNA half-life range of 5–60 min, since this includes most transcripts.

C. Measured abundances of mRNAs of 0.4–1 kb with half-lives of 5–60 min were plotted over the results from the model. The regression line formulae and correlation coefficients are for log₂-transformed data. Median mRNAs/cell/gene are also indicated.

D. As (C) but for 1–3 kb mRNAs

E. As (D) but for mRNAs longer than 3 kb.

Supplementary Fig. S9. Although more detailed analysis might reveal some trend for particular mRNA classes, we found no gradient within transcription units, and no differences between transcription units. We therefore did not add variations in transcription rate to the model.

Next, we added a step that was obviously missing from the published model: polyadenylation (step 3 in Fig. 7A). This was assigned the same kinetics as for *trans* splicing (Ullu *et al.*, 1993). We assumed that it, like *trans* splicing, was competing with precursor degradation

– in this case, degradation of a primary transcript that has been 5' *trans* spliced, but not yet polyadenylated (step 5 in Fig. 7A).

Meanwhile, we discovered that the correlation between half-life and abundance was much better for shorter mRNAs than for longer ones (Fig. 6E and F and Fig. S10A), and that the shortest mRNAs fitted the published model relatively well (Fig. 6E). Longer mRNAs were on average less abundant than shorter ones with the same half-life (Supplementary Fig. S10B). We concluded that at least one of the steps in the model must include a negative influence of mRNA length.

We now set up a simulation that incorporated variations in processing, splicing and polyadenylation rates as well as in mRNA decay, and that included a negative influence of mRNA length. Splicing and polyadenylation were varied somewhat conservatively such that the half-time for each step was 2 ± 1.5 min. To mimic length-dependent co-transcriptional destruction we multiplied kinetic constant k_5 (degradation of the *trans* spliced precursor) by a factor of (mRNA length in nt)/600. (The number 600 was chosen empirically.) Our rationale for this was that a longer transcript has more chance to be degraded before it is processed, since it takes longer for the polymerase to make a precursor that contains the downstream *trans* splicing signal. In addition, slow (inefficient) polyadenylation of an mRNA should also increase the likelihood of co-transcriptional destruction.

We now used this extended model to simulate steady-state abundances for transcripts with varying lengths, half-lives, 5' *trans* splicing and 3' polyadenylation kinetics, using the known length distribution for the mRNAs and a half-life distribution based on our measurements. The results are shown in Fig. 7B. We then compared the simulated data for each length group with the real data (Fig. 7C–E). There was a good qualitative agreement between simulations and the data. In the simulations the goodness-of-fit (R^2) for the relation between mRNA half-life and abundance decreased with increasing transcript length, and in both the simulations and in the experimental data, abundance was highest for the short transcripts and lower for the longer ones. The real data, though, had bigger variation (lower R or R^2) than the simulations. This may hint at additional regulation, or a need to adjust the parameters further. For instance, compared with the extended model, in the real dataset the intermediate length group (Fig. 7D) had more transcripts that were stable (half-life > 16') and showed very low abundance. These might be mRNAs that are polyadenylated slower than the maximum half-time of 3.5 min that we allowed for in the model. (Unfortunately we could not model using the real distribution of processing times because these could not be measured using the RNASeq approach.) The model also predicted a population of long, very unstable

transcripts with lower abundance than was seen in the data (less than one mRNA in every two parasites) (Fig. 7E). These would be very unstable, long transcripts that are processed slowly. It is possible that this particular combination of properties (long and unstable) is rare.

We also tested the effect of a theoretical length-dependent 3' polyadenylation step – although it is difficult to imagine a mechanism for this (k_3 decreasing with increasing transcript length). This also resulted in the linear relationship between mRNA half-life and abundance being poorer for longer transcripts but the effect was not as pronounced as seen when we allowed for co-transcriptional degradation (Supplementary Fig. S11A). Notably, if we also removed degradation of the 5' *trans* spliced precursor (step 5), we obtained a virtually perfect relation between half-life and abundance for all lengths (Supplementary Fig. S11B). This is a consequence of the steady-state assumption. Simulations showed that without degradation of the 5' *trans* spliced precursor, if transcription were induced from an 'off' state, slower 3' polyadenylation lowered the mature mRNA abundance in the initial phase, but did not affect the abundance when the steady state was reached (not shown). If degradation of the un-spliced precursor was removed as well, a perfect correlation between abundance and half-life was obtained (Supplementary Fig. S11C). The modelling thus yielded the important conclusion that length-dependent variations in processing can only affect steady state mRNA abundance if processing is competing with degradation of the partly processed intermediates.

Discussion

At the start of this study it was already known that the rate of mRNA decay is an important determinant of trypanosome mRNA abundance, but detailed quantitative information, including time-courses for the degradation, was available only for a very small number of mRNAs (Table 2). We have now provided transcriptome-wide decay kinetics for both bloodstream and procyclic forms, and have developed a detailed model for gene expression. We could confirm the important role of mRNA decay in determining steady state mRNA levels. We also found that a previously unrecognized step in processing has to be present to explain the steady-state mRNA abundances we measured. This additional step is most likely co-transcriptional degradation. Differences in mRNA maturation rates can only affect mRNA abundance if co-transcriptional degradation is also occurring.

The kinetics of mRNA decay fell into at least three different patterns. The 'fast-slow' group of mRNAs showed initial fast degradation, followed by a slower phase. This is also seen, in bloodstream forms, for

mRNAs bearing the 3'-UTR of the mRNA encoding the major surface protein EP procyclin; in that case, the rapid degradation phase occurs without deadenylation and is dependent on the 5'-3' exoribonuclease XRNA (Li *et al.*, 2006; Schwede *et al.*, 2009). Those EP mRNAs that are not attacked in this way are deadenylated more slowly, then degraded by XRNA and the exosome (Irmer and Clayton, 2001; Haile *et al.*, 2003; Li *et al.*, 2006; Schwede *et al.*, 2009). Depletion of the exosome caused a slight delay in the initiation of EP mRNA degradation (Haile *et al.*, 2003), suggesting some cross-talk between the 5' and 3' pathways. To see if this pattern was general we examined previous RNASeq results for parasites defective in different steps of mRNA decay. The fast-to-slow method of decay was indeed the predominant one in the case of mRNAs whose half-lives were increased after XRNA depletion (Manful *et al.*, 2011) (Fisher's test P -value $\ll 0.01$) (Supplementary Table S2, sheet 2), suggesting that the fast-to-slow pattern is diagnostic for XRNA-dependent, deadenylation-independent decay. However, this mode was also predominant for mRNAs whose decay had been significantly affected by exosome (RRP45) depletion (Fadda *et al.*, 2013) (Fisher's test P -value $\ll 0.01$) (Supplementary Table S2, sheet 2).

The 'slow-fast' group of mRNAs showed degradation rates that increased with time. This pattern has previously been observed by Northern blotting for numerous long-lived kinetoplastid mRNAs (for some references see Table 2). We speculated that 'slow-fast' kinetics could be understood in terms of stochastic (non-processive) deadenylation. Indeed, alpha and beta tubulin and histone H4 have slow-fast kinetics in bloodstream forms, and show decreases in overall length prior to mRNA decay (Schwede *et al.*, 2008). Degradation of nearly all mRNAs is to some extent affected by depletion of the CAF1/NOT complex (Fadda *et al.*, 2013), so we could not use RNASeq data from such cells to look for differential effects. However, transcripts affected by depletion of the deadenylase PAN2 (Fadda *et al.*, 2013), were predominantly degraded in the slow-fast method (Fisher's test P -value $\ll 0.01$) (Supplementary Table S2, sheet 2). This supports the idea that the slow-fast pathway reflects deadenylation prior to decay. As expected, abundant mRNAs always have relatively long half-lives; we could now show that they also often decay with slow-fast or exponential kinetics.

We found 179 mRNAs in which mRNA decay played a dominant role in developmental regulation. Most likely, the number is higher, since 200 regulated mRNAs showed such a low abundance in one life cycle stage that the half-life could not be measured. The mRNAs can also be subdivided into classes according to whether they are preferentially affected by RNAi targeting the 5' or 3' degradation in bloodstream forms, or according to their deg-

radation pattern (fast/slow, slow/fast or exponential) in both forms. Combined with our previous analyses of the effects of inhibiting particular degradation pathways, we now have clear groups of mRNAs that are subject to similar degradation control, and which might therefore share sequences that govern that control. For example, half of the RNAs that showed lower abundances and half-lives in bloodstream forms than in procyclics (Supplementary Table S3, sheet 7) switched from the fast-slow to the slow-fast or exponential modes of degradation. We suggest that in bloodstream forms, some of these mRNAs are bound by proteins that recruit the decapping enzyme or XRNA; while in procyclic forms, the mRNAs are bound by different proteins and degraded by the slower, deadenylation-dependent route. Previous attempts to computationally predict regulatory sequences were limited by the resolution of the microarray data used and the fact that it showed only steady-state mRNA levels (Mao *et al.*, 2009; Queiroz *et al.*, 2009; Najafabadi *et al.*, 2013). The higher resolution of our new data may now facilitate identification of regulatory motifs.

Our experimental results showed that degradation of the mature mRNA is not the only factor that contributes to steady state abundance. Most notably, for a given half-life range, mRNA abundances decreased with length. There is clear evidence that mRNA processing is important in trypanosome gene expression regulation (Michaeli, 2011): experiments with reporter genes have shown strong variations in splicing efficiencies depending on the sequences preceding the AG dinucleotide (Kapotas and Bellofatto, 1993; Avliyakov *et al.*, 2003; Siegel *et al.*, 2005), and when alternative splice sites are available, different ones may predominate in procyclic and bloodstream forms (Nilsson *et al.*, 2010; Siegel *et al.*, 2010; Zhang *et al.*, 2010; Rettig *et al.*, 2012). Slow polyadenylation may also explain previous reports of kinetoplastid mRNA regulation by sequences in downstream intergenic regions (Ramamoorthy *et al.*, 1995; Brooks *et al.*, 2001; Fluck *et al.*, 2003). Our results now show that under steady-state conditions, the processing efficiency can only influence mRNA abundance if processing is competing with precursor degradation. Moreover, the length dependency of mRNA abundance can be explained by competition between length-dependent mRNA precursor degradation and polyadenylation.

Our kinetic model does not yet account for all of the features of the data. There is a variety of possible reasons for this. First, we restricted ourselves mainly to varying known kinetic parameters within the limits that are suggested by the available measurements. The simulation results might be improved if we allowed for more variation in these parameters, or added additional variables for which there is currently no experimental evidence. For example, polyadenylation was assigned the same rate as

trans splicing, but may in fact be slightly slower, since it is thought to occur after splicing has been completed (Ullu *et al.*, 1993). Second, in considering the effects of transcript length on degradation of the spliced precursor, we have opted for a simple length-dependence, which largely leaves transcripts < 600 nt unaffected by intermediate degradation. We could have changed the boundary chosen, or made this dependence more complex. Third, the model was evaluated at steady state, but we cannot rule out that pre-steady-state kinetics influenced the experimental results. If there are variations in the overall transcription initiation rate during the cell cycle, some slowly processed mRNAs may never reach steady state.

The detailed mechanism of trypanosome co-transcriptional degradation is currently unknown. Co-transcriptional degradation certainly occurs in other systems, although the details may not apply in trypanosomes. In Opisthokonts, the nuclear quality control machinery is triggered by failure of mRNA capping, and by failure to splice or polyadenylate a terminated mRNA (Schmid and Jensen, 2010; 2013; Kilchert and Vasiljeva, 2013). 5'→3' degradation by Xrn2/Rat1 requires a free 5' end, and has mainly been studied in relation to transcription termination beyond polyadenylation sites (Schmid and Jensen, 2010; 2013; Kilchert and Vasiljeva, 2013). With trypanosome polycistronic transcription, most mRNAs have no free 5' end, since the mature 5'-end is created by *trans*-esterification, and termination is relevant only for a tiny proportion of genes. Polyadenylation might, however, fail after downstream *trans* splicing has been completed, and the resulting free 3'-ends should be attacked by the nuclear exosome (Cristodero and Clayton, 2007; Etheridge *et al.*, 2009). There is, though, no reason to think that this would be more efficient for long mRNAs than for short ones and in fact, depletion of the exosome actually causes preferential loss of longer mRNAs (Fadda *et al.*, 2013; Jha *et al.*, 2014). The NOT deadenylation complex is associated with RNA polymerase II in mammals, but it is thought to promote (not inhibit) elongation (Reese, 2013). Previous experiments that aimed at characterization of RNA polymerase II subunits and transcription factors (e.g. Lee *et al.*, 2009) were not designed for detection of proteins present in sub-stoichiometric amounts, but after tandem affinity purification of the NOT complex, we detected two peptides from a putative polymerase II elongation factor (Tb927.2.3580) (Färber *et al.*, 2013).

Our data are best explained by the existence of an endonuclease that intermittently cleaves the associated mRNA precursor. This is highly speculative since there is, so far as we know, no real precedent for such a degradation process (Schmid and Jensen, 2010; 2013; Kilchert and Vasiljeva, 2013). To recognize the precursor specifically, we would expect the endonuclease to be associated either with chromatin, or with the elongating RNA poly-

merase II complex. One consequence of such a mechanism would be that a decrease in the polymerase elongation rate would increase the time that the precursor is accessible for digestion. Indeed, after depletion of trypanosome CTR9, a probable transcription elongation factor (Ouna *et al.*, 2012), the abundances of long mRNAs were drastically decreased (Fadda *et al.*, 2013; Jha *et al.*, 2014). In animal cells and yeast, elongating RNA polymerase II pauses, and even backtracks (Darzacq *et al.*, 2007; Churchman and Weissman, 2011). We speculate that such pauses might, in trypanosomes, be a trigger for cleavage without detachment of the polymerase, and could even be influenced by the DNA sequence. This would provide an additional means to regulate the formation of mRNA in the absence of any control of transcription initiation.

The simplicity of the trypanosome system has made it possible to construct a mathematical model of mRNA control which has revealed novel aspects and refined our view of the likely contribution of mRNA processing and degradation to mRNA abundance. When combined with ribosome profiling to determine translation efficiency, our measurements will enable us to mathematically model trypanosome gene expression from DNA to protein.

Experimental procedures

Sample preparation

Bloodstream and procyclic form trypanosomes of the Lister 427 strain, expressing the tet repressor, were used; the bloodstream forms were the 2T1 line described in (Alsford *et al.*, 2005), while procyclic cells expressed the tet repressor from plasmid pHD2060. Bloodstream forms were cultured in HMI-9 medium containing 20% FCS, to a density of $1.0\text{--}1.2 \times 10^6$ cells ml⁻¹; procyclic forms were cultured in MEM-Pros to a density of 2×10^6 (experiment 1) or 1.3×10^6 (experiment 2) cells ml⁻¹. For the pre-mRNA decay assay, transcription was inhibited briefly by incubating the cells with Actinomycin D (10 µg ml⁻¹) and samples were harvested immediately (equivalent to a 5 min incubation); the bloodstream form experiment was done twice; for procyclics a single experiment was done. For the mRNA decay assay, cap methylation was inhibited by adding 1 µg ml⁻¹ Sinefungin for 5 min to prevent splicing (t0 samples), then Actinomycin D was added for up to 120 min for bloodstream forms and 240 min for procyclics. All samples are listed in Supplementary Table S1. The parasites were harvested by centrifugation for 5 min (all times given include the centrifugation time); RNA was extracted using peqGOLD-Trifast (peqLab) and depleted for rRNA using Ribominus Eukaryotic kit for RNA-seq (Invitrogen, no. 10837-02 or -08). The RNA was then sequenced (Library preparation at the BioQuant facility, Heidelberg and HiSeq at EMBL). The first set of bloodstream-form decay samples was measured using Illumina GAIIX and all remaining samples with Illumina HiSeq (Supplementary Table S1). All data are available at the NCBI Sequence read archive, with accession number SRP042959, and a simplified version of the results

has been sent to TritypDB. We cannot rule out the possibility that the Sinefungin or Actinomycin D treatments influenced the mRNA decay rates, especially after long incubation times, but at present alternative methods are not available.

RNA decay analysis by Northern blotting

To determine the overall mRNA decay rate, aliquots of the sequenced samples were separated by agarose gel electrophoresis, blotted and probed with a spliced leader probe, using the rRNA signal for normalization, as previously described (Fadda *et al.*, 2013). The measured signal was used to fit an exponential decay curve with R (Supplementary Fig. S4) and the values corresponding to the curve were used to normalize the RNAseq decay data.

Data processing

Reads were aligned to the *T. brucei* 927 reference genome, and read counts were extracted using Bowtie, SamTools and custom-made PERL scripts as previously described (Fadda *et al.*, 2013). Read counts per ORF were first normalized for the total number of counts (RPM, reads per million reads). In order to build the libraries, the same amount of input RNA was used for each sample; and similar numbers of read counts were obtained for each library. After addition of transcription inhibitors, the proportions of the unstable mRNAs go down (giving a decrease in RPM) while the proportions of the stable RNAs go up (giving a decrease in RPM). The stable mRNAs seem to increase because the overall reduction in mRNA levels is not reflected in the RNAseq data. This necessitates the use of an external control to detect – and correct for – the decline in mRNA quantities in the cell with time. We could not normalize using endogenous stable RNAs, because they were excluded during the ribo-minus and size selection steps of library preparation. To measure global mRNA decay we therefore used Northern blots to assay the decline of the spliced leader (SL) signal, as previously described (Fadda *et al.*, 2013). Then, to account for the decreasing total amounts of RNA during the decay assay, the RPM for each time point were multiplied by the corresponding SL signal at that time. ORFs with an average of less than 50 reads at time 0 were excluded from the half-life analysis. Lastly, the read counts were calculated as a ratio to time point 0 (t_0). This format was used as input for a C++ code to calculate half-lives. Using the equations described by Deneke *et al.*, (2013) we allowed for three alternative decay patterns: simple exponential, initially fast, then slower (fast-slow, equation 9 of Deneke *et al.*, 2013) and initially slow, then fast (slow-fast, equation 10 of Deneke *et al.*, 2013). For the calculations each measurement was treated as an independent data point. Differential expression was analysed by DESeq (Anders and Huber, 2010).

To measure splicing kinetics, read counts were extracted for the 20mer upstream of all known splice acceptor sites for each gene (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010; Zhang *et al.*, 2010 and our data), as well as for each coding region. If the read counts for the 20mer were equal to, or more than, the length-normalized read counts in coding region, we assumed that the splice site was not being

used in our sample, and the data for the site were removed; we also removed sites if no reads existed in the 20mer region in the control (untreated) sample. Finally, zero values in the 20mer region in Actinomycin D samples were converted to 1 read. Read counts per 20mer were normalized to the read counts per coding region of the corresponding gene in each sample. The ratio of the normalized 20mer counts in the treated/untreated samples was used to calculate a pre-mRNA half-life based on an exponential model. While interpreting the results, we attribute this to splicing, although strictly speaking pre-mRNA degradation also contributes. Probability density plots were done using a kernel density function contained in the R package 'ggplot'.

To calculate gene copy numbers, we first did a control for uniformity of coverage. We mapped the reads from the reference genome against itself using reads available from Dr A. Ivens (University of Edinburgh). Allowing each read to map only once should yield a bell-shaped coverage plot. Next we allowed the reads to map up to 300 times and calculated RPKM for each gene. We assumed that the modal RPKM value (M) corresponded to a gene copy number of 1, since most genes should be unique. Next, we aligned the reads from sequenced, fragmented DNA from our lab strain, and mapped them to the reference genome, allowing each read to align only once. We again calculated the modal value (M'). If the number of reads from a gene was M' , then that gene has the same copy number as in the reference; if M' has a value $n \times M$, then the strain has n times the number of copies in the reference.

Calibrating decay models

The code for decay calibration is written in C++ and is provided as Supplementary file S1. We used the equations described by Deneke *et al.*, (2013) to model yeast mRNA decay; that paper explains the mathematical details. Complex, non-exponential mRNA degradation is modelled by a Markov chain with 5 states; each state represents a different biochemical process of mRNA degradation, where the molecule is likely to be degraded at any state. The rate of transition between states is assumed to be the same for all states and is denoted as λ . The degradation rate from any state is denoted as μ . In one variant of the complex model the initial rate of degradation μ_1 is assumed bigger than all subsequent rates, denoted ν . This is described by the parametric density Equation 9 (Deneke *et al.*, 2013) and referred to here as 'fast-slow' model. The other model variant assumes all degradation rates except the last to be equal, and the last rate $\mu_n \equiv \nu > 0$. This is described by the parametric density Equation 10 (Deneke *et al.*, 2013) and here referred to as 'slow-fast' model. We calibrated and compared three models: the exponential model, 'fast-slow' and 'slow-fast'. The calibration procedure worked in several stages. In the first stage we calibrated each model. The fitting procedure using non-linear least squares which we used (Dennis *et al.*, 1981) is extremely fast but suffers numerically from multiple local minima. These local minima are obtained from different 'initial parameter guesses'. For the three-parameter models we chose to calibrate the model using 125 different starting points, lying on a mesh between 0 and 3.0 with 5 points per axis. For each calibration we stored the residual model error and the calibrated parameters. In the next stage we ran a

rejection/acceptance procedure for each of the parametric forms, and for each of a form's 125 models. The rejection criteria were subjective, but were designed to reject models that were completely implausible, or had collapsed to simpler models. The rejection criteria for each model are found in the function `includeTentativeSolution()`. As an example, for a calibrated model of Equation 9's density we have parameters λ , μ , ν (Deneke *et al.*, 2013). The model is rejected if λ or ν are 0, or if any of the parameters have become infinite. The next stage of calibration chooses the best fitting accepted model for each density type. We call the three best-fit models the three 'candidate' models. The last stage of calibration selects the best-fit model from among the candidate models. The null hypothesis is that the distribution is Exponential E. For a candidate model $C \neq E$ we look at the relative

$$D(C, E) > \frac{R(C) - R(E)}{0.5 \cdot |R(C) + R(E)|}$$

where $R(M)$ is the residual sum of squares error from fitting with model M . If $D(C; E) > 5\%$ we rejected the null hypothesis and accepted C . The choice of threshold is essentially arbitrary. Deneke *et al.* (2013) used a boundary of 10%, we chose 5% after visual examination of individual examples. Using the lower threshold leads to a lower proportion of curves being designated as exponential. We did not favour either the Equation 10 or Equation 9 density over each other and simply took the candidate from the two with the minimal residual. The 'half lives' for the mRNAs with non-exponential decay in all the Tables are the time at which half of the mRNA is left (the value of time T where 50% of the distribution lies above T , and 50% lies below T).

Modelling mRNA processing and decay

The *PGK* mRNA model developed in Haanstra *et al.* (2008) was initially used. Then we made a new, modified version which applies to any transcript. The new model was extended to include polyadenylation and length-dependent degradation of the *trans* spliced intermediate. Model simulations were done, and extensions were made, in the open source software COPASI v.4.9.45 (Hoops *et al.*, 2006). Figure 7A shows a graphical representation of the reactions in the model, and their rate equations are listed in Table 1.

The constants k in the equations relate to half-lives (HL) of the corresponding species via

$$k = \frac{\ln(2)}{\text{HL}} \text{ min}^{-1}$$

For k_4 , k_5 and k_6 this relation has been explicitly included in the model, which enables us to affect the k 's by changing the half-life through random sampling (see below).

k_2 and k_3 each have a default value of 0.41 (half-life of 1.7') (as in Haanstra *et al.*, 2008). We assumed k_4 and k_5 to be equal to each other and took a value of 0.08 min^{-1} which is based on the previously measured half-life of the *PGK* precursor after inhibition of both transcription and splicing by Actinomycin D and Sinefungin (Haanstra *et al.*, 2008). μ is the specific growth rate and is 0.0019 min^{-1} (a population doubling time of 6 h).

The differential equations that describe the changes in species concentrations over time in the model are:

$$\frac{d[\text{unspliced precursor}]}{dt} = v_1 - k_2 \times [\text{unspliced precursor}] - (k_4 + \mu) \times [\text{unspliced precursor}]$$

$$\frac{d[5' \text{ spliced precursor}]}{dt} = k_2 \times [\text{unspliced precursor}] - k_3 \times [5' \text{ spliced precursor}] - (k_5 + \mu) \times [5' \text{ spliced precursor}]$$

$$\frac{d[\text{mRNA}]}{dt} = k_3 \times [5' \text{ spliced precursor}] - (k_6 + \mu) \times [\text{mRNA}]$$

Concentrations are in molecules cell^{-1} .

To simulate effects that depend on the length of the mature mRNA, the relevant constant was multiplied or divided by (length in nt)/(600 nt): For length-dependent degradation of the 5' spliced intermediate, k_5 was multiplied by length/600 so that longer transcripts would be affected faster. For length-dependent polyadenylation, k_3 was multiplied by 600/length so that polyadenylation would be slowed down for longer transcripts.

The model was used to simulate steady state abundances for individual transcripts (i.e. with their own values for the kinetic constants). To this end, 5510 transcripts were simulated by

- Sampling the half-life of the un-spliced precursor in the 5' *trans* splicing reaction from a normal distribution with a mean of 2' and a standard deviation of 1.5' (affects k_2),
- Sampling the half-life of the 5' spliced precursor in the 3' poly-adenylation reaction from a normal distribution with a mean of 2' and a standard deviation of 1.5' (affects k_3 ; only used when k_3 was not multiplied by a length-dependent factor), and
- Sampling the length of a transcript from a uniform distribution for short (551 transcripts in the range 400–1000 nt), intermediate (4408 transcripts in the range 1000–3000 nt) and long (551 transcripts in the range 3000–20000 nt) transcripts

For each length range, the half-life of the mature mRNA (affecting k_6) was sampled from a uniform distribution within the half-life range in each bin in Fig. 3, except that for the first bin, the half-lives ranged between 1–5'. The number of transcripts in the length range for each bin was taken proportionally to the distribution of lengths of the overall dataset (10% short, 80% intermediate and 10% long transcripts).

In some cases, the combination of parameters did not result in a steady state and the final number of solutions was therefore lower than the number of simulations.

Acknowledgements

We thank Janine Philipp for making the SL-hybridized Northern blots for bloodstream forms, and Ute Leibfried for technical assistance. We also thank Dr Carlus Deneke (Max Planck Institute of Colloids and Interfaces, Potsdam) for useful discussions, and Alasdair Ivens (University of Edinburgh) for the TREU927 shotgun reads.

Funding

This work was supported by the collaborative Sysmo grant 'The silicon trypanosome', with funds from the Bundesministerium für Bildung und Forschung, BBSRC (BB/1004602/1), and the Netherlands Organization for Scientific Research (NWO). Work in Edinburgh was also funded by the Wellcome Trust (088293; 095831). DD was supported by the Deutsche Forschungsgemeinschaft, Grant no CL112/17-1, and BMB is funded by a Rosalind Franklin Fellowship from the University of Groningen.

Author contributions

Abeer Fadda did most of the data analysis. Mark Ryten wrote the programme to fit the different decay models to the data and derive half-lives. Dorothea Droll, Federico Rojas and Valentin Färber were responsible for parasites, RNA samples and Northern blots. Clementine Merce did the gene copy number calculations. Keith Matthews supervised the bloodstream-form work. Jurgen Haanstra was responsible for the modelling, with supervision from Barbara Bakker. Christine Clayton was responsible for overall supervision and coordinating writing of the paper.

Conflict of interest

The authors declare no conflict of interest.

References

- Ali, J., Tagoe, D., Munday, J., Donachie, A., Morrison, L., and de Koning, H. (2013) Pyrimidine biosynthesis is not an essential function for *Trypanosoma brucei* bloodstream forms. *PLoS ONE* **8**: e58034.
- Ali, J.A., Creek, D.J., Burgess, K., Allison, H.C., Field, M.C., Maser, P., and De Koning, H.P. (2012) Pyrimidine salvage in *Trypanosoma brucei* bloodstream forms and the trypanocidal action of halogenated pyrimidines. *Mol Pharmacol* **83**: 439–453.
- Alsford, S., Kawahara, T., Glover, L., and Horn, D. (2005) Tagging a *T. brucei* rRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol Biochem Parasitol* **144**: 142–148.
- Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.
- Avliyakov, N., Hines, J., and Ray, D. (2003) Sequence elements in both the intergenic space and the 3' untranslated region of the *Crithidia fasciculata* KAP3 gene are required for cell cycle regulation of KAP3 mRNA. *Eukaryot Cell* **2**: 671–677.
- Bakker, B.M., Krauth-Siegel, R.L., Clayton, C., Matthews, K., Girolami, M., Westerhoff, H.V., et al. (2010) The silicon trypanosome. *Parasitology* **137**: 1333–1341.
- Bringaud, F., Riviere, L., and Coustou, V. (2006) Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol Biochem Parasitol* **149**: 1–9.
- Brooks, D.R., Denise, H., Westrop, G.D., and Coombs, G.H. (2001) The stage-regulated expression of *Leishmania mexicana* CPB cysteine proteases is mediated by an intergenic sequence element. *J Biol Chem* **276**: 47061–47069.
- Brown, B.S.V., Chi, T.B., and Williams, N. (2001) The *Trypanosoma brucei* mitochondrial ATP synthase is developmentally regulated at the level of transcript stability. *Mol Biochem Parasitol* **115**: 177–187.
- Chaudhuri, M., Sharan, R., and Hill, G. (2002) Trypanosome alternative oxidase is regulated post-transcriptionally at the level of RNA stability. *J Eukaryot Microbiol* **49**: 263–269.
- Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., et al. (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell* **131**: 1340–1353.
- Churchman, L.S., and Weissman, J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**: 368–373.
- Clayton, C. (2013) The regulation of trypanosome gene expression by RNA-binding proteins. *PLoS Pathog* **9**: e1003680.
- Clayton, C., and Estevez, A. (2010) The exosomes of trypanosomes and other protists. In *RNA Exosome*. Jensen, T. (ed.). Austin, TX: Landes Bioscience, pp. 39–49.
- Colasante, C., Robles, A., Li, C.-H., Schwede, A., Benz, C., Voncken, F., et al. (2007) Regulated expression of glycosomal phosphoglycerate kinase in *Trypanosoma brucei*. *Mol Biochem Parasitol* **151**: 193–204.
- Cristodero, M., and Clayton, C. (2007) Trypanosome MTR4 involved in ribosomal RNA processing. *Nucleic Acids Res* **35**: 7023–7030.
- Daniels, J., Gull, K., and Wickstead, B. (2010) Cell biology of the trypanosome genome. *Microbiol Mol Biol Rev* **74**: 552–569.
- Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S., Phair, R., and Singer, R. (2007) In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14**: 796–806.
- Deneke, C., Lipowsky, R., and Valleriani, A. (2013) Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA. *PLoS ONE* **8**: e55442.
- Dennis, J., Gay, D., and Roy Welsch, R. (1981) Algorithm 573: an adaptive nonlinear least-squares algorithm. *ACM Trans Math Softw* **7**: 367–383.
- Erben, E., Chakraborty, C., and Clayton, C. (2014) The CAF1-NOT complex of trypanosomes. *Front Genet* **4**: 299.
- Estévez, A. (2008) The RNA-binding protein *TbDRBD3* regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res* **36**: 4573–4586.
- Etheridge, R.D., Clemens, D.M., Gershon, P.D., and Aphasizhev, R. (2009) Identification and characterization of nuclear non-canonical poly(A) polymerases from *Trypanosoma brucei*. *Mol Biochem Parasitol* **164**: 66–73.
- 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.
- Färber, 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.

- Fernández-Moya, S., and Estévez, A. (2010) Posttranscriptional control and the role of RNA-binding proteins in gene regulation in trypanosomatid protozoan parasites. *Wiley Interdiscip Rev RNA* **1**: 34–46.
- Fluck, C., Salomone, J., Kurath, U., and Roditi, I. (2003) Cycloheximide-mediated accumulation of transcripts from a procyclin expression site depends on the intergenic region. *Mol Biochem Parasitol* **127**: 93–97.
- Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M., and Hughes, T. (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol* **24**: 5534–5547.
- Gudipati, R.K., Xu, Z., Lebreton, A., Seraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012) Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol Cell* **48**: 409–421.
- Gunasekera, K., Wuthrich, D., Braga-Lagache, S., Heller, M., and Ochsenreiter, T. (2012) Proteome remodelling during development from blood to insect-form *Trypanosoma brucei* quantified by SILAC and mass spectrometry. *BMC Genomics* **13**: 556.
- Haanstra, J., Stewart, M., Luu, V.-D., van Tuijl, A., Westerhoff, H., Clayton, C., and Bakker, B. (2008) Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form *Trypanosoma brucei*. *J Biol Chem* **283**: 2495–2507.
- Haile, S., Estévez, A.M., and Clayton, C. (2003) A role for the exosome in the initiation of degradation of unstable mRNAs. *RNA* **9**: 1491–1501.
- Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., et al. (2013) Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* **153**: 1000–1011.
- Harigaya, Y., and Parker, R. (2012) Global analysis of mRNA decay intermediates in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **109**: 11764–11769.
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., et al. (2006) COPASI – a COMplex PATHway Simulator. *Bioinformatics* **22**: 3067–3074.
- Hotz, H.-R., Hartmann, C., Huober, K., Hug, M., and Clayton, C.E. (1997) Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a trypanosome surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res* **25**: 3017–3025.
- Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006) Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast* **23**: 439–454.
- Huang, J., and van der Ploeg, L. (1991) Maturation of polycistronic pre-mRNA in *Trypanosoma brucei*: analysis of trans splicing and poly(A) addition at nascent RNA transcripts from the hsp70 locus. *Mol Cell Biol* **11**: 3180–3190.
- Irmer, H., and Clayton, C.E. (2001) Degradation of the *EP1* mRNA in *Trypanosoma brucei* is initiated by destruction of the 3'-untranslated region. *Nucleic Acids Res* **29**: 4707–4715.
- 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. *Eukaryot Cell* **13**: 664–674.
- Kapotas, N., and Bellofatto, V. (1993) Differential response to RNA trans-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*. *Nucleic Acids Res* **21**: 4067–4072.
- Kelly, S., Kramer, S., Schwede, A., Maini, P., Gull, K., and Carrington, M. (2012) Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. *Open Biol* **2**: 120033.
- Kilchert, C., and Vasiljeva, L. (2013) mRNA quality control goes transcriptional. *Biochem Soc Trans* **41**: 1666–1672.
- Kiss, D.L., and Andrusis, E.D. (2010) Genome-wide analysis reveals distinct substrate specificities of Rrp6, Dis3, and core exosome subunits. *RNA* **16**: 781–791.
- Kolev, N., Franklin, J., Carmi, S., Shi, H., Michaeli, S., and Tschudi, C. (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog* **6**: e1001090.
- Kramer, S., and Carrington, M. (2011) Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol* **27**: 23–30.
- Kramer, S., Queiroz, R., Ellis, L., Webb, H., Hoheisel, J., Clayton, C., and Carrington, M. (2008) Stress granules and the heat shock response in *Trypanosoma brucei*. *J Cell Sci* **121**: 3002–3014.
- Lamour, N., Riviere, L., Coustou, V., Coombs, G.H., Barrett, M.P., and Bringaud, F. (2005) Proline metabolism in procyclic *Trypanosoma brucei* is down-regulated in the presence of glucose. *J Biol Chem* **280**: 11902–11910.
- Lee, J.H., Jung, H.S., and Gunzl, A. (2009) Transcriptionally active TFIIH of the early-diverged eukaryote *Trypanosoma brucei* harbors two novel core subunits but not a cyclin-activating kinase complex. *Nucleic Acids Res* **37**: 3811–3820.
- Li, C.-H., Irmer, H., Gudjonsdottir-Planck, D., Freese, S., Salm, H., Haile, S., et al. (2006) Roles of a *Trypanosoma brucei* 5'→3' exoribonuclease homologue in mRNA degradation. *RNA* **12**: 2171–2186.
- 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.
- Mao, Y., Shateri Najafabadi, H., and Salavati, R. (2009) Genome-wide computational identification of functional RNA elements in *Trypanosoma brucei*. *BMC Genomics* **10**: 355.
- Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S., and Matthews, K. (2006) Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res* **34**: 5312–5324.
- Michaeli, S. (2011) Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol* **6**: 459–474.
- Mnaimneh, S., Davierwala, A.P., Haynes, J., Moffat, J., Peng, W.T., Zhang, W., et al. (2004) Exploration of essential gene functions via titratable promoter alleles. *Cell* **118**: 31–44.
- Munchel, S., Shultzaberger, R., Takizawa, N., and Weis, K. (2011) Dynamic profiling of mRNA turnover reveals

- gene-specific and system-wide regulation of mRNA decay. *Mol Biol Cell* **22**: 2787–2795.
- 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.
- Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., et al. (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog* **6**: e1001037.
- Ouna, B., Nyambega, B., Manful, T., Helbig, C., Males, M., Fadda, A., and Clayton, C. (2012) Depletion of trypanosome CTR9 leads to gene expression defects. *PLoS ONE* **7**: e34256.
- Overath, P., and Engstler, M. (2004) Endocytosis, membrane recycling and sorting of GPI-anchored proteins: *Trypanosoma brucei* as a model system. *Mol Microbiol* **53**: 735–744.
- Palenchar, J.B., and Bellofatto, V. (2006) Gene transcription in trypanosomes. *Mol Biochem Parasitol* **146**: 135–141.
- Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J., and Clayton, C. (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics* **10**: 495.
- Rabani, M., Levin, J., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., et al. (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**: 436–442.
- Ramamoorthy, R., Swihart, K.G., McCoy, J.J., Wilson, M.E., and Donelson, J.E. (1995) Intergenic regions between tandem gp63 genes influence the differential expression of gp63 RNAs in *Leishmania chagasi*. *J Biol Chem* **270**: 12133–12139.
- Reese, J.C. (2013) The control of elongation by the yeast Ccr4-not complex. *Biochim Biophys Acta* **1829**: 127–133.
- Rettig, J., Wang, Y., Schneider, A., and Ochsenreiter, T. (2012) Dual targeting of isoleucyl-tRNA synthetase in *Trypanosoma brucei* is mediated through alternative splicing. *Nucleic Acids Res* **40**: 1299–1306.
- Schmid, M., and Jensen, T.H. (2010) Nuclear quality control of RNA polymerase II transcripts. *Wiley Interdiscip Rev RNA* **1**: 474–485.
- Schmid, M., and Jensen, T.H. (2013) Transcription-associated quality control of mRNP. *Biochim Biophys Acta* **1829**: 158–168.
- Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., and Tollervey, D. (2012) Transcriptome-wide analysis of exosome targets. *Mol Cell* **48**: 422–433.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011) Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.
- 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., 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.
- Shalem, O., Dahan, O., Levo, M., Martinez, M.F.R., Furman, I., Segal, E., and Pilpel, Y. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* **4**: 223.
- Siegel, T., Tan, K., and Cross, G. (2005) Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Mol Cell Biol* **25**: 9586–9594.
- 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 Res* **38**: 4946–4957.
- Siegel, T., Gunasekera, K., Cross, G., and Ochsenreiter, T. (2011) Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. *Trends Parasitol* **27**: 434–441.
- 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.
- Stern, M., Gupta, S., Salmon-Divon, M., Haham, T., Barda, O., Levi, S., et al. (2009) Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA* **15**: 648–665.
- Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., et al. (2012) Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* **22**: 1350–1359.
- Treck, T., Larson, D.R., Moldon, A., Query, C.C., and Singer, R.H. (2011) Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* **147**: 1484–1497.
- Ullu, E., Matthews, K.R., and Tschudi, C. (1993) Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol* **13**: 720–725.
- Urbaniak, M.D., Martin, D., and Ferguson, M.A. (2013) Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*. *J Proteome Res* **12**: 2233–2244.
- 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.
- Webb, H., Burns, R., Ellis, L., Kimblin, N., and Carrington, M. (2005) Developmentally regulated instability of the GPI-PLC mRNA is dependent on a short-lived protein factor. *Nucleic Acids Res* **33**: 1503–1512.
- Zhang, X., Cui, J., Nilsson, D., Gunasekera, K., Chanfon, A., Song, X., et al. (2010) The *Trypanosoma brucei* MitoCarta and its regulation and splicing pattern during development. *Nucleic Acids Res* **38**: 7378–7387.

Supporting information

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