Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis

Fadia Ibrahim[1,2,3,*], **Manolis Maragkakis**[1,2,3,*], **Panagiotis Alexiou**[1,2,3], and **Zissimos Mourelatos**[1,2,3,‡]

[1]Department of Pathology and Laboratory Medicine, Division of Neuropathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[2]Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[3]Penn Medicine Translational Neuroscience Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Abstract

Messenger RNAs transmit the genetic information that dictates protein production, and are a nexus for numerous pathways that regulate gene expression. The prevailing view of canonical mRNA decay is that it is mediated by deadenylation and decapping followed by exonucleolysis from the 3′ and 5′ ends. By developing Akron-Seq, a novel approach that captures native 3′ and 5′ ends of capped and polyadenylated RNAs respectively, we show that canonical human mRNAs are subject to repeated, cotranslational, ribosome-phased, endonucleolytic cuts at the exit site of the mRNA ribosome channel, in a process that we term ribothrypsis. We uncover RNA G-quadruplexes among likely ribothrypsis triggers, and show that ribothrypsis is a conserved process. Strikingly, we find that mRNA fragments are prevalent in living cells with important implications for the interpretation of experiments, such as RNA-Seq, that rely on the assumption that mRNAs exist largely as full-length molecules *in vivo*.

‡Correspondence to: mourelaz@uphs.upenn.edu (Z.M.).
*These authors contributed equally to this work

INTRODUCTION

mRNAs, typically found in cells as ribonucleoproteins (mRNPs) [1,2], engage ribosomes, transfer RNAs and other factors that translate the genetic information transcribed in mRNAs to instruct protein production [3]. Canonical mRNA decay [4] involves deadenylation followed by 3′ to 5′ exonucleolysis carried by the exosome [5,6]; and decapping [7,8] followed by 5′ to 3′ exonucleolysis carried out by XRN1 [4]. Numerous *trans* acting factors [9,10] and *cis* RNA elements, including codons [11] regulate mRNA decay.

The textbook view that translation and decay are largely distinct processes has been challenged by the demonstration that in yeast and plants, Xrn1 (plant XRN4) degrades decapped mRNAs following the last translating ribosome [12–14]. This is conceptually appealing as the mRNA that undergoes 5′ to 3′ decay cannot re-initiate translation, while the last protein molecule that it produces will be full-length. Much less clear, however, is what happens to translated mRNAs when degradation occurs in the 3′ to 5′ direction. More generally, the prevalence and features of cotranslational decay of canonical mRNAs are still unclear. In contrast, we have a better understanding of how cells deal with aberrant mRNAs from studies of mRNA surveillance pathways, which initiate degradation of defective mRNAs during translation [15–18]. In Nonsense-mediated decay (NMD), premature termination codons are recognized by a multiprotein complex that accelerates decay from both ends and also recruits the SMG6 endonuclease [15,16,19]. In No-Go Decay (NGD) and Non-Stop Decay (NSD), aberrant mRNAs with strong stalls or lacking stop codons respectively, trigger cleavages upstream of stalled ribosomes mediated by an, as yet, unknown endonuclease [16,18,20–24]. Stalled ribosomes are recognized by Dom34 (PELOTA) and Hbs1 proteins, homologs of eukaryotic Release Factors eRF1 and eRF3 respectively, and along with Rli1 (ABCE1), split the stalled ribosome to a 40S subunit and a 60S subunit that still contains the peptidyl-tRNA-bound nascent polypeptide chain [16–18,20,25–28]. The interface of this 60S subunit is recognized by a multiprotein, ribosome quality-control complex (RQC) that ubiquitinates and extracts the nascent polypeptide for proteasomal degradation, and disassembles and recycles the 60S subunit and peptidyl-tRNA [17,29–32].

Here, we develop Akron-Seq, a novel approach that captures native 3′ and 5′ ends of capped and polyadenylated RNAs respectively, and show that canonical human mRNAs are subject to repeated, cotranslational, ribosome-phased, endonucleolytic cuts at the exit site of the mRNA ribosome channel, in a process that we term ribothrypsis. We uncover likely triggers and features of ribothrypsis and report that produced mRNA fragments are prevalent in living cells with important implications for the interpretation of experiments, such as RNA-Seq, that rely on the assumption that mRNAs exist largely as full-length molecules *in vivo.*

RESULTS

Cotranslational 3′ end trimming of capped mRNAs

To investigate human mRNA dynamics *in vivo,* we initially attempted whole transcriptome, single-molecule real-time (SMRT, PacBio), long-read sequencing of capped RNAs isolated from HeLa cells by a method we developed and termed Akron-SMRT (ακρον: ultimate

point, starting or ending; Supplementary Fig. 1a). Capped RNAs are enriched by digestion with the Terminator 5′ phosphate (P)-dependent exonuclease. SMRT did not provide adequate sequencing depth for whole transcriptome analysis, or accurate quantitation. Instead, we interrogated select, randomly chosen, transcripts and found that they were cotranslationally degraded from the 3′ end, while their fragments remained associated with translating ribosomes (Supplementary Fig. 1b, c). We verified these findings by high-resolution Northern blots (Supplementary Fig. 1d–f) and excluded exogenous RNA degradation as the source for fragmentation (Supplementary Fig. 1g).

Akron-Seq uncovers ribothrypsis

Next, we wished to investigate the prevalence of mRNA fragments in cells and how they are generated. Methods that capture the 5′ ends of polyadenylated mRNAs, such as parallel analysis of RNA ends (PARE)-Seq [33,34], degradome-Seq [35], GMUCT [36], and 5PSeq [13]; or the 3′ ends of transcripts, such as TAIL-Seq [37], have elucidated fundamental aspects of RNA processing and metabolism. However, existing 5′ end sequencing methods do not capture 3′ ends and TAIL-Seq does not discriminate between the 3′ ends of capped or uncapped RNAs. We developed Akron-Seq to simultaneously capture and sequence the 3′ ends of capped RNAs (Akron3) and the 5′ ends of polyadenylated RNAs (Akron5) (Fig. 1a). In both, RNA native ends are marked by adapter ligation before fragmentation or cDNA generation and are identified by paired-end, sequence-by-synthesis (Fig. 1a). In Akron3, RNA is digested with Terminator to enrich for capped molecules. Since Terminator does not degrade RNAs bearing 5′-OH, we also set out to capture such molecules and compare their abundance to capped RNAs, but found that they are present in negligible amounts (Supplementary Fig. 2a). We also attempted to generate Akron3 libraries of 5′-OH bearing RNAs but were unable to do so since we did not obtain a specific polymerase chain reaction (PCR) product (Supplementary Fig. 2b). This is consistent with findings from the Steinmetz lab that was also unable to generate libraries from RNAs bearing 5′-OH in *S. cerevisiae* [13]. Thus, Akron3 libraries are derived overwhelmingly from capped RNAs. For Akron5, we isolate polyadenylated RNAs using Oligo(dT) covalently bound to dynabeads (Fig. 1a).

By providing reference points for each mRNA end, relative to its beginning (cap) or its end (poly-A), we can interrogate their positional relationships to various mRNA features and processes, including translation, to uncover mechanistic insights of fragment generation. We prepared biological triplicate (Akron5) and duplicate (Akron3) libraries and found that they were highly reproducible (Supplementary Fig. 2c, d). We performed all analyses described below separately for each library and found them to be consistent across all libraries. By measuring the distribution of 3′ ends of capped mRNAs (Akron3) and 5′ ends of polyadenylated mRNAs (Akron5) on meta-mRNA, we found that most ends mapped within the mRNA body (Fig. 1b, c). Specifically, we found that ~63.5% of 3′ ends of capped mRNAs mapped within the coding sequence (CDS) of protein-coding genes and on average only 11% had non-templated adenines at the 3′ end. The later mapped primarily at the 3′ end of genes, as expected (Fig. 1d, e). Also, by interrogating the 50% most expressed transcripts, quantified by total RNA-Seq, we found that ~63% of those had Akron3 ends that mapped within the CDS indicating a ubiquitous process.

To examine whether the generation of these native 5′ and 3′ ends is related to translation, we compared Akron-Seq with ribosome profiling (Ribo-Seq) [38] data from HeLa cells [39]. We detected that ribosome-protected fragments (RPFs) mapped overwhelmingly in-frame with codons (Fig. 2a), and displayed three-nucleotide (3-nt) periodicity with stop codons (Fig. 2b), which we verified using Discrete Fourier Transform (DFT) (Fig. 2c), reflective of the stepwise ribosome movement during translation elongation. Next, we plotted the distribution of Akron5 and Akron3 ends upstream and downstream of RPFs' 5′ ends (Fig. 2d, e) either cumulatively for all transcripts, or after dividing them into ten quantiles according to their Akron5 or Akron3 abundance. For controls, we randomized the Akron end positions within the same transcripts, thus maintaining sequence and expression features and eliminating any potential confounding effects from the 3-nt ribosome movement alone. In contrast to random controls, we observed a striking 3-nt periodicity for both ends and for all transcripts (Fig. 2f), verified with DFT (Fig. 2g), indicating cotranslational generation of the 5′ and 3′ ends of mRNA fragments. We found a similar periodicity when we analyzed TAIL-Seq data [37] (Supplementary Fig. 3a) and when we interrogated long noncoding RNAs (lncRNAs) that we found loaded with ribosomes (Supplementary Fig. 3b), consistent with recent reports of polysomal association and degradation of lncRNAs [40]. To exclude that the observed pattern is created by the combined effect of codon-associated nt periodicity and any potential, though unlikely [41], nt preference by the T4 RNA ligase, we down-sampled our Akron libraries constraining the nt content at the Akron ends to 25% for each nt. We found that the plots for the down-sampled controls were almost identical to the original libraries verifying the independence of the observed results from any such bias (Supplementary Fig. 3c). Collectively, our results indicate a universal process of cotranslational RNA degradation that operates on most transcripts that engage with ribosomes.

By measuring the relative position of Akron5 and Akron3 ends to the RPFs' 5′ ends (Fig. 2f), we observed coincidence of an Akron5 peak and an Akron3 drop around position 0. By experimental design, Akron5 ends correspond to 5′ ends of polyadenylated mRNAs and Akron3 ends correspond to the 3′ ends of capped RNAs. Therefore, the coincidence of the two signals at position 0 cannot be the result of exonucleolytic trimming initiating from any direction. Along with the 3-nt periodicity and positioning of the Akron5 peak and Akron3 drop relative to the ribosome, these findings indicate that the ends must be created by an endonucleolytic activity that is associated with the ribosome and cleaves the translated mRNA as it exits the ribosome channel (Fig. 2d; mRNA exit is located between the head and the platform of the 40S subunit).

We hypothesized that the endonucleolytic cut is triggered during translation elongation just upstream of a transiently stalled ribosome (stall marked with a grey box in Fig. 2f, located ~25-nt downstream of RPFs' 5′ ends) and results in the creation of a 5′ and a 3′ mRNA fragments. The 5′ ends of the resulting 3′ fragments coincide with the 5′ of RPFs, which is evident by the increased density at position 0 (Fig. 2f). Our hypothesis predicts that the cut around position 0 will result in self-perpetuating mRNA cuts as incoming ribosomes (colored purple, Fig. 2f) will encounter a hard stall at the cleavage site, triggering upstream cuts on the 5′ fragment. The fragments that are generated by this iterative endonucleolysis are expected not to harbor a cap and therefore should not be captured by Akron3. This is verified by the characteristic depletion of Akron3 ends around position 0 (Fig. 2f).

According to our model, the incoming-stalled ribosomes on the 5′ fragment are expected to create accumulation of RPFs upstream of the initial cut sites that are captured by Akron5. This is reflected in our analysis as increased density in the area immediately to the right of position 0. This happens because RPFs are used as reference points and their 5′ ends are positioned at 0 on the x-axis (positive positions on the x-axis indicate Akron5 ends located downstream of RPFs). Interestingly, we observe that the highest density in this region is at position +15, corresponding to the distance of the 5′ end of an incoming ribosome that is stalled at the initial cut site, now located at its peptidyl (P) site, as previously described [42]. Finally, we hypothesize that the depletion of Akron5 and Akron3 ends at the stall site is the result of a protection of that region, likely attributable to the nature of the stall mechanism itself. We have simulated core aspects of this model such as the transient ribosome stall, the resulting upstream endonucleolytic cut, and propagating cuts by an incoming ribosome, as well as the putative protection at the stall site. The simulation recreates all characteristics of the experimentally derived Akron5 plot, supporting the theoretical foundation of our interpretation (Supplementary Fig. 3d).

The process that we describe is conceptually similar to NGD; however, it is not a surveillance pathway of aberrant mRNAs but rather a widespread and ubiquitous mechanism that degrades canonical mRNAs and we have termed it "ribothrypsis" ($\theta\rho\acute{\upsilon}\pi\tau\omega$: to fragment). Our model predicts that the ribothrypsis nuclease, which we will refer to as ribothrypsin, should act endonucleolytically so neither XRN1, nor the exosome, nor the SMG6 endonuclease, that can mediate NMD, is ribothrypsin. Also it is predicted that if the process is ubiquitous, then ribothrypsis products (RTPs) harboring 5′ and 3′ ends defined by this process, should be observable in RNA-Seq data where no RNA fragmentation is performed, such as small RNA-Seq. Furthermore, our model implies the generation of final RTPs (FRTPs) ~16 nt in length, corresponding to the distance from the first cut, that now resides at the P site of the incoming ribosome and stalls it (purple in Fig. 2f), and the subsequent cut upstream of the stalled incoming ribosome.

Ribothrypsis is not mediated by XRN1

Xrn1 has been shown to closely follow the last translating ribosome and define the 5′ ends of mRNAs, captured by 5PSeq, during 5′ to 3′ cotranslational degradation in yeast [13]. However, according to our model, XRN1 cannot be ribothrypsin although it could still polish the 5′ ends of RNAs after ribothrypsis cuts. To examine the role of XRN1, we analyzed along with Akron5, PARE-Seq (which is similar to Akron5) data obtained from HeLa cells after XRN1 knockdown (KD) [34]. If XRN1 mediates the process, we would expect abrogation of 3-nt periodicity and of the signal around the RPFs' 5′ end in XRN1-KD. We calculated the distribution of Akron5 and PARE-Seq 5′ ends around the stop codon and around the 5′ ends of RPFs. We detected 3-nt periodicity, confirmed by DFT, upstream, but not downstream, of the 5′ end of the terminating ribosome in Akron5 and PARE XRN1-KD (Fig. 3a, b) verifying that cotranslational ribothrypsis cuts are not mediated by XRN1. When plotting the distribution of 5′ ends around RPFs, the peak in PARE XRN1-KD was displaced ~2-nt upstream of position 0 (Fig. 3c) compared to Akron5 (Fig. 2f, middle panel) supporting a role for XRN1 in polishing the 5′ ends after ribothrypsis.

We wondered whether ribothrypsis was also operating in *S. cerevisiae* and we re-analyzed 5PSeq data from wild-type yeast (grown in complete media - YPD) and from an Xrn1-deletion (*Xrn1*) strain [13]. We calculated the distribution of 5PSeq 5′ ends around the stop codon and around the RPFs' 5′ ends, which we obtained from wild-type yeast grown in YPD (Supplementary Fig. 4a) [42]. As previously reported [13], we found 3-nt periodicity upstream of the 5′ end of the terminating ribosome in wild-type yeast, which we confirmed with DFT, and a peak at position −18 (as measured from the last nucleotide of the stop codon, which was set at 0), indicative of cotranslational degradation by Xrn1 that stopped at the terminating ribosome (Supplementary Fig. 4b). As reported [13], the −18 peak disappeared in *Xrn1* but we found, and verified by DFT, that the 3-nt periodicity upstream of the terminating ribosome persisted, arguably at a lower level (Supplementary Fig. 4c). We also observed marked 3-nt periodicity between RPFs and 5′ ends of mRNAs of wild-type (Supplementary Fig. 4d) and *Xrn1* yeast (Supplementary Fig. 4e). These findings indicate that ribothrypsis likely operates in *S. cerevisiae* and Xrn1 polishes the 5′ ends of transcripts after ribothrypsis cuts. Unlike 5PSeq in yeast, Akron5 did not show enrichment at position −17 from the stop codon (Fig. 3a) suggesting that Xrn1 may play a bigger role in RNA decay in yeast than in humans.

Ribothrypsis footprints in small RNA-Seq libraries

As hypothesized by our model, if ribothrypsis is prevalent, then its footprints should be observable in small RNA-Seq libraries, which capture endogenous 5′ and 3′ ends by adapter ligation and unlike classic RNA-Seq are not subjected to fragmentation, thus preserving the endogenously produced mRNA ends. We mined deeply sequenced HeLa small RNA libraries (sRNA; average read length of ~60-nt) [43] and examined the distribution of 5′ and 3′ ends of sRNAs mapping to mRNAs, relative to 5′ ends of RPFs, and found a striking 3-nt periodicity, indicative of cotranslational generation of these ends, and similar density plots to Akron5 and Akron3 respectively (Fig. 3d, e). As revealed by Akron5, polyadenylated RNAs are subject to ribothrypsis (Fig. 2f, g) eliminating the exosome as a candidate for ribothrypsin, as the exosome attacks mRNAs after deadenylation. As shown in Fig. 3f, g, and Supplementary Fig. 5a, b, the 3-nt periodicity and density plots were identical between sRNA libraries from wild-type HeLa cells and sRNA-Seq libraries after KD of RRP40, a core exosome component, or after triple KD of all exosome nucleases (DIS3, DIS3L, RRP6) [43]. These findings indicate that sRNAs may be used as proxies to investigate cotranslational mRNA decay and support our model's prediction that the exosome is not ribothrypsin. However, they do not preclude functions for the exosome and the Ski complex in clearing ribothrypsis products, even when bound to ribosomes [44]; or in initiating ribothrypsis during 3′-5′ exonucleolysis, as the exosome starts degrading CDS. In principle, exosome-mediated trimming [42] or cleaving by other ribonucleases (such as Agos programmed with miRNAs or siRNAs [45,46]) may be an important trigger for ribothrypsis.

Since our protocols probe canonical mRNAs, we reasoned that SMG6, an endonuclease that can initiate degradation of aberrant mRNAs subjected to NMD, is unlikely to be ribothrypsin. Indeed, when we analyzed the distribution of RPFs' 5′ end to 5′ ends of PARE-Seq reads from HeLa cells after double KD for SMG6 and XRN1, we found that the

3-nt periodicity and density plot profiles were retained (Fig. 3h), thus ruling out SMG6 as ribothrypsin.

To investigate whether we could detect final ribothrypsis products (FRTPs, Fig. 2f) in sRNA libraries, we interrogated RNAs smaller than 17-nt [43] and examined the distribution of their 5′ and 3′ ends relative to the 5′ ends of Akron5 reads. We found that the 5′ ends of sRNAs peaked ~16-nt upstream of Akron5 ends while the 3′ ends of sRNAs peaked immediately upstream of Akron5 ends, consistent with the designation of these ~16-nt sRNAs as FRTPs (Supplementary Fig. 5c). Collectively, all findings presented above support the ribothrypsis model.

Features of ribothrypsis

We initially asked whether ribothrypsis was related to transcript abundance (calculated by RNA-Seq from ribosomal RNA (rRNA)-depleted total RNA); length of CDS and untranslated regions (UTRs); average density of ribosomes per transcript (calculated by Ribo-Seq and RNA-Seq); or RNA stability (calculated by BRIC-Seq [47]) but we did not identify any correlations between Akron ends and these features (our unpublished data).

Next, we examined the distribution of codons around Akron5 and Akron3 ends. As controls, we randomized end positions within the same transcripts maintaining the expression levels, nucleotide content and open reading frame at the cut site. Compared to control, we observed an enrichment of the lysine K(AAG) codon downstream of Akron5 ends (Supplementary Fig. 6) and an enrichment of K(AAG; AAA) at Akron3 ends (Supplementary Fig. 7) suggesting potential ribosome stalling at these codons. Prior reports have documented stalling in lysine codons or during translation of poly(A) tails that drive mRNA cleavage [24]. We note, however, that the ribothrypsis ends mapped by Akron5 and Akron3 are all within CDS of canonical mRNAs, and not in poly(A) sequences (generated from normal or premature polyadenylation), nor in consecutive stretches of lysines (our unpublished data), suggesting a role for the codon themselves as associating with ribothrypsis. Furthermore, we found enrichment of aspartate D(GAC) and glutamate E(GAA) codons upstream of Akron3 ends (Supplementary Fig. 7). Remarkably, aspartate and glutamate codons, and in particular the same E(GAA) codon that we find associating with Akron3 ends, associate with ribosome pausing in mouse ES cells [48] likely due to a more open RNA structure at these codons [49]. Our findings support the role of specific codons in ribothrypsis, although mechanisms are lacking and will be an important pursuit for future studies. More generally, the relationship between ribothrypsis, codons and translation will need to be addressed as better understanding of the codon impact on mRNA structure, codon optimality, codon usage and translation efficiency is developed for humans.

We then reasoned that mRNA structure might trigger or associate with ribothrypsis. We first used RNAfold to predict secondary structure potential around Akron5 ends that map in CDS and found a marked reduction of secondary structure around ribothrypsis cuts and an increase immediately downstream (Fig. 4a). Modification of RNA bases by chemicals, such as dimethyl sulfate (DMS) is widely used to probe RNA structure or protein footprints on RNA which are identified by next-generation sequencing [50] that captures the stalling of commonly used viral-derived reverse transcriptases (RT-stop) as they encounter modified

bases [51]. We compared Akron5 data with *in vivo* DMS-Seq data from HeLa cells [51] and found a sharp increase of DMS values around position 0, indicative of high DMS reactivity, consistent with high RNA accessibility around ribothrypsis cuts (Fig. 4b). Although some of the DMS value may be due to RT-fall at ribothrypsis cuts, rather than RT-stop at DMS-modified bases, when combined with structure prediction, these results are indicative of RNA accessibility to the ribothrypsis endonuclease at these sites and increased structure immediately downstream of the cuts.

RNA G-quadruplexes as potential ribothrypsis triggers

RNA G-quadruplexes (rG4) are secondary structures formed by self-association of guanine tetrads into square planar arrangements that stack on top of each other, stabilized by potassium or other cations [52]. rG4s are very stable *in vitro* [52,53], although largely unfolded by proteins and other factors *in vivo* [54]. We employed an algorithm that we had previously used and validated for rG4 detection in piRNA precursors [55], to predict genome-wide rG4 locations and compare them to ribothrypsis cut sites. We identified a striking rG4 enrichment downstream of Akron5 ends (green line in Fig. 5a), which corresponded closely to the predicted ribosome stall site of our model (Fig. 2f). Within CDS, rG4s are known to repress protein output by impeding ribosome progression and exhibit the largest suppression when found at position +2 relative to the open reading frame (ORF2) [56]. Remarkably, our data also show that rG4s in ORF2 correspond to the highest enrichment downstream of Akron5 ends (Fig. 5a, blue line). Overall, we found that ~43% of Akron5 ends had rG4s within a 100-nt region downstream. We found similar rG4 enrichment immediately downstream of PARE-Seq ends after XRN1, or XRN1 and SMG6 KDs (Fig. 5b, c). We did not observe any rG4 enrichment in random controls (Fig. 5a–c), which we generated by randomizing Akron end positions within the same transcripts (maintaining sequence and expression features). To verify that the signal was specific to rG4s and not due to increased G-content, we shuffled the sequences around Akron5 ends, maintaining nt, expression and codon composition and found that there was no rG4 enrichment in these shuffled controls (Fig. 5a–c). Next, we examined the distribution of experimentally detected rG4s (after rG4-Seq in HeLa cells [53,54]) and also detected enrichment immediately downstream of Akron5 ends (Fig. 5d, e). Moreover, we identified that the area around ribothrypsis cuts, and particularly those that have an rG4 downstream, were preferentially conserved in vertebrates (Fig. 5f). rG4 enrichment downstream of putative ribothrypsis cleavages was also conserved in yeast, in both wild-type and *Xrn1* strains (Supplementary Fig. 8). Taken together our findings support a role for conserved mRNA areas and increased rG4s, which likely cause transient ribosome stalling, among ribothrypsis triggers.

DISCUSSION

The prevailing view of canonical mRNA decay is that it is an exonucleolytic process mediated by XRN1 and by the exosome and Ski complex. Our findings uncover a novel pathway of mRNA decay, which we term ribothrypsis (Fig. 6), to convey its important features: first, that the decay of canonical mRNAs is also endonucleolytic; second, that it is mediated by translating ribosomes; third, that mRNA decay fragments persist in cells; and fourth, that just like any process that affects all mRNAs (such as mRNA splicing or mRNA

translation), ribothrypsis is almost certainly regulated. Why do cells employ ribothrypsis for mRNA decay? We hypothesize that the ribosome and its associated ribothrypsin can integrate numerous inputs in real time on each mRNA molecule. This intimate connection between mRNA translation and mRNA decay, orchestrated by the ribosome, allows for exquisite integration and compartmentalization of post-transcriptional gene expression regulation on each transcript.

While our manuscript was under review, two papers were published describing the contribution and features of NGD in the decay of Ire-1 cleaved mRNAs during activation of the Unfolded Protein Response in fission yeast [57]; and ribosome collision as a critical trigger of NGD in budding yeast [58]. In these cases, the endonuclease "NGDase" associates with the ribosome and cleaves the mRNA as it exits the ribosome [57,58]. Our findings are consistent with these reports and suggest that a ribosome-mediated mechanism of mRNA decay may be a unifying feature that underlies all mRNA decay and is not limited just for mRNA surveillance and clearance of aberrant or select mRNAs. Even though we do not know the identity of the nuclease, molecular mechanisms that mediate mRNA cleavage in ribothrypsis and NGD are likely connected. However, we expect that there will also be differences on how the two processes are triggered and regulated. For example, mRNA truncation or base damage by oxidation [59] elicits NGD on defective mRNAs. But we expect that numerous factors will trigger ribothrypsis on canonical (non-damaged mRNAs), for physiological regulation and not for quality control. We identified rG4s as likely triggers of ribothrypsis but we believe that this is just the tip of the iceberg and that many other factors, transcript-specific or global, can activate and regulate ribothrypsis.

Characterization of the chemical groups that mark the ends of the cleaved products may provide clues to assist identification of ribothrypsin. Based on our findings and those from the Steinmetz lab [13], 5′-P bearing mRNA fragments are prevalent in HeLa cells and budding yeast, suggesting that ribothrypsin generates 3′-fragments bearing 5′-P. In this case, ribothrypsis products are immediately accessible to XRN1, which degrades RNAs bearing 5′-P, and ribothrypsin is likely a protein whose cleaved products bear 5′-P (3′-fragment) and 3′-OH (5′-fragment). However, we note that the Hesselberth lab was able to selectively capture and sequence, from budding yeast, the small amounts of RNAs bearing 5′-OH termini by employing RtcB RNA ligase and found that such fragments originated from CDS regions of numerous mRNAs [60]. Intriguingly, 5′-OH fragments seemed to accumulate upstream of codons for acidic and basic aminoacids [60]. It is possible that these 5′-OH termini are products of ribothrypsis (but further work is required to test this notion); in that case ribothrypsin would likely generate 3′-fragments bearing 5′-OH, and 5′-fragments bearing 2′, 3′ cyclic phosphate. Ribozymes, in addition to proteins, generate such ends when cleaving RNA and given the fact that rRNA is essential for the peptidyl transferase activity of the ribosome, it is not inconceivable to hypothesize that rRNA may also play a key role in ribothrypsis. If ribothrypsin generates 5′-OH, a kinase would need to phosphorylate the 5′-OH bearing fragments to make them substrates for XRN1. We expect that identification of ribothrypsin will provide critical biological insights into mRNA translation and decay. Furthermore, we think that our work sets the stage for many other future investigations, including identification of factors and pathways that trigger or regulate ribothrypsis in normal cellular functions and in disease; and more mechanistic understanding

of ribothrypsis' intersection with mRNA translation and turnover. We also envision that the new methods for capturing RNA ends and the computational approaches, scripts and analyses that we report here, will be useful for analyzing and simulating RNA processes in general.

Finally, our findings have implications for studies that investigate gene expression by employing RNA-Seq, microarrays, RT-PCR and related approaches, which are among the most commonly used methods to profile gene expression. An implicit, almost axiomatic, assumption made by users of these methods is that mRNAs exist largely as full-length molecules *in vivo*. Based on this assumption, fragmentation of the isolated RNA is used to prepare libraries suitable for sequence-by-synthesis. The sequencing reads from each mRNA are then used for downstream analyses, such as ranking them to calculate expression levels; deciphering expression signatures that are used to identify cell types; or performing other analyses, limited only by the imagination of the scientists involved. Similarly, microarrays use probes that hybridize to only a few areas on each transcript but then the signals are extrapolated to full-length molecules. RT-PCR amplifies a small amplicon from each mRNA, yet it is assumed that the small, amplified fragment was derived from a larger, full-length molecule. Polysome profiling to assess ribosome loading and translation of mRNAs also assumes that mRNAs undergoing translation are largely full-length. Our finding that mRNA fragments are generated cotranslationally and are naturally prevalent in living cells, needs to be considered and opens new possibilities for the interpretation of RNA-Seq and other experiments used to interrogate gene expression.

METHODS

Cell culture

HeLa cells, mycoplasma-free, were maintained in DMEM (Gibco) supplemented with 10% FBS (Sigma), at 37 °C, in 5% CO_2. Cells were harvested at ~80% confluence.

Polysome analysis

Cells were pretreated with 100 μg/ml cycloheximide (CHX, Sigma) at 37 °C for 10 min, and were washed twice with cold 1x PBS plus 100 μg/ml CHX, and collected in lysis buffer (20 mM Tris-HCL pH 7.5, 200 mM NaCl, 2.5 mM MgCl2) supplemented with 100 μg/ml CHX, complete EDTA-free protease inhibitor (Roche), and 500 U/ml RNasin ribonuclease inhibitor (RNasin, Promega). Cells were passed through a 25G syringe, incubated on ice for 10 min, and centrifuged to clear nuclei. Cytoplasmic extracts were layered on top of a 10–50% sucrose gradient, and were ultracentrifuged at 38,000 rpm for 2 hr (Beckman SW41-Ti rotor) at 4 °C. Polysome profile was recorded via absorbance at 254 nm (ISCO). Fourteen fractions of equal volume were collected and digested with 400 μg/ml proteinase K (Roche) and 1% SDS at 37 °C for 30 min; total RNA was then extracted with Trizol (Ambion).

Northern blotting

Total RNA was extracted from polysome fractions by Trizol. RNA was pooled from fractions corresponding to light (6–9) or heavy (10–14) polysomes or from all polysome fractions (6–14). Total RNA or RNA from input and sucrose gradient fractions were treated

with Terminator exonuclease (Epicentre) to remove non-capped, 5′-phosphate (P)-bearing RNAs by incubation at 30 °C for 1 hr according to the manufacturer's instructions. RNA was ethanol precipitated and resolved on 6% denaturing polyacrylamide gel electrophoresis (PAGE) and was transferred to a Hybond N+ nylon membrane (GE Healthcare) in 0.5x TBE in a semidry transfer system (Amersham Biosciences). RNA was UV-crosslinked (Stratagene) to the membrane and prehypridized in ULTRHyb-oligo buffer (Ambion). The probe was internally labeled with α^{32}P-UTP by T7 MEGAshortscript kit (Ambion) and purified using Sephadex G-50 columns (GE Healthcare). Hybridization was carried out in ULTRHyb-oligo buffer. Membranes were washed twice with 2x SSC buffer plus 1% SDS, and twice with 0.5x SSC buffer plus 0.1% SDS and signals detected by storage-phosphor autoradiography (Amersham). For quantification, the entire region that covers the full-length and smear was selected in a rectangular box using ImageJ (v. 1.46r) and the signal from the entire region was considered as a 100%. A rectangular box with the exact same width but covering only the area corresponding to the full-length signal was then selected. To get an estimate of the full-length percent, the signal of the full-length was divided over the total signal.

Ribosome immunoprecipitation from polysome fractions and Western blotting

Polysome fractions (6–14) were pooled from two sucrose gradients, and were diluted in IP binding buffer (20 mM HEPES pH 7.0, 150 mM NaCl, 5 mM MgCl2, 0.05% Triton X-100) supplemented with 500 U/ml RNasin, 40 µg/ml Digitonin, and complete EDTA-free protease inhibitor. Samples were incubated with pre-washed protein-A dynabeads (Invitrogen) that were conjugated to an antibody against RPL7A (#A300-749A, Bethyl Laboratories) or to non-immune serum as a negative control. Samples were nutated at 4 °C for 4 hr, followed by centrifugation at 3,000 rpm for 10 min to collect beads. Beads were washed five times with IP binding buffer. An aliquot was used for Western blot (WB), and the rest was incubated with IP elution buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10 mM EDTA, 1% SDS) at 95 °C for 3 min. RNA was recovered by Trizol extraction and processed for Northern. For WB, beads were heated at 95 °C for 10 min; proteins were resolved on 4–12% NuPAGE gels (Invitrogen), transferred to nitrocellulose membrane and probed with anti-RPL7A at 1:5,000 as described previously [61].

Preparation of Akron-SMRT sequencing libraries

Total RNA from HeLa cells and polysome fractions (light; 6–9, heavy; 10–14) was extracted by Trizol. RNA was treated with Terminator to remove non-capped, 5′-P bearing RNAs. RNA was precipitated by one volume of 5 M Lithium chloride and further treated with Calf Intestinal Phosphatase (CIP, NEB) at 37 °C for 30 min. RNA was recovered using AMPure-XP beads (Beckman) according to the manufacturer's protocol.

For library construction, RNA was ligated to a 3′-biotinylated RNA adapter (REL3) using T4 RNA ligase (Fermentas) plus 500 U/ml RNasin and incubated at 16 °C for 16 °C. Ligated RNA was treated with Terminator to remove excess adapter. RNA was recovered by AMPure-XP beads, and was captured by M-280 Streptavidin dynabeads (Invitrogen). Beads were washed twice with 2x bind/wash (BW) buffer (10 mM Tris pH 7.5, 1 mM EDTA, 2 M NaCl), and twice with 1x BW buffer (5 mM Tris pH 7.5, 0.5 mM EDTA, 1 M NaCl) and

resuspended in RNase-free water. cDNA was generated under conditions that favor full-length amplification. RNA was reverse transcribed by Superscript III (Invitrogen) primed with RT primer (REL3-RT; complementary to the 3′-adapter), plus saturated trehalose at (50 °C for 45 min, 55 °C for 15 min, heat inactivate at 70 °C for 10 min). RNA was hydrolyzed in 0.11 N NaOH at 98 °C for 20 min. cDNA was ethanol precipitated and PCR amplified by KAPA HiFiHotStartReadyMix (KAPA Biosystems), and 10 μM primers that contained SMRT-barcodes and corresponded to the 5′ UTR; forward (CDK4 F0) or TUBB (TUBB F0), and reverse (REL3-RT). Reactions were run at (98°C for 2 min, 20 cycles (98 °C for 30 s, 57 °C for 30 s, 72 °C for 2 min), and 72 °C for 5 min). Re-PCR was set up using 10 μM primers that corresponded to the start codon; forward (CDK4 F1 or TUBB F1) and reverse (REL3-1 or REL3-2), and run at (98 °C for 2 min, 6 cycles (98 °C for 30 s, 57 °C for 30 s, 72 °C for 2 min), and 72 °C for 5 min). PCR products were resolved on 2% Metaphor gel (Lonza), and fragments (350–850 bp and 900–4,000 bp) were excised per gene and purified using GeneJet kit (Thermo Scientific). SMRT-bell sequencing libraries were prepared by Pacific Biosciences using P6-C4 chemistry and sequencing protocol. Each size-selected library was sequenced on two SMRT cells. For all analyses, only Circular Consensus (CCS) reads were used.

As an internal control, polysome pool was spiked in with Renilla luciferase mRNA prior to total RNA extraction and Terminator digestion. Luciferase mRNA was generated *in vitro* by T7 MegaScriptTranscription kit (Ambion) in the presence of a synthetic RNA cap analog (NEB) and α^{32}P-UTP. RNA was resolved on 6% denaturing PAGE and visualized by autoradiography (GE Health).

Preparation of Akron sequencing libraries

DNA-free total RNA (~250 μg) was used as an input to generate both Akron3 and Akron5 libraries. The integrity of mRNA was assessed prior to library preparation using RNA Bioanalyzer or agarose gel. After the initial treatment, all protocols converge into the same RNA ligation and downstream library preparation steps with minor modifications.

Preparation of Akron3 libraries

Initially, we used ribosomal RNAs (rRNAs) depletion before proceeding to Akron3 library preparation but we noticed that our sequences contained significant amounts of small, capped, noncoding RNAs, including Uridine-rich, small nuclear RNAs (U snRNAs). To remove such abundant, capped RNAs, small RNAs <200 bp was first removed from total RNA with SPRIselect beads (Beckman) according to the manufacturer's protocol and was used an input for both Akron3 and Akron5.

To capture 3′ ends of capped (non- and poly(A) RNAs), snRNAs and rRNAs were depleted. We used our own sequencing data to design oligonucleotides (oligos) against abundant snRNAs and published oligos to remove rRNAs [62]. A cocktail mix of the sn/rRNAs oligos containing biotinylated-3′ ends was prepared based on the abundance of each targeted sn/rRNAs (that we calculated based on our initial sequencing runs). Total RNA (~5–6 μg per reaction) was incubated with the cocktail mix in 1x SSC buffer in a thermal cycler at (75 °C for 5 min, 0.1/s to 37 °C, 37 °C for 45 min). Samples were incubated with MyOne

Streptavidin C1 dynabeads (Invitrogen) that were resuspended in 2x BW buffer with mixing at 1,000 rpm for 15 min at 37 °C. Depleted-total RNA was precipitated using 100% ethanol, 3M NaOAc and 2 μg glycogen (Ambion). To remove non-capped, 5′-P bearing-RNAs, samples were Terminator digested, as described for Akron-SMRT. T4 PolyNucleotide Kinase (PNK, NEB) treatment is optional before Terminator digestion to remove residual 5′-OH RNAs.

For library construction, depleted- and Terminator-digested RNA (~5 μg) was ligated to a 5′-P and 3′-biotinylated RNA adapter (A3-RL3) by T4 RNA ligase in a mixture containing 5% polyethylene glycol (PEG, NEB) and 500 U/ml RNasin at 16 °C for 16 hr. Ligated RNAs were fragmented in 1x RNA fragmentation buffer (Ambion) at 70 °C for 7 min. The fragmented RNA was ethanol precipitated, and resolved on 6% denaturing PAGE. Fragments >400 bp were excised and RNA was eluted from gel pieces by crushing and shaking at room temperature in elution buffer (300 mM NaCl, 10 mM Tris-HCL pH 7.5, 1 mM EDTA) plus 500 U/ml RNasin. RNA was ethanol precipitated and incubated with M280 Streptavidin Dynabeads. RNA 5′ ends were phosphorylated on beads by T4 PNK treatment at 37 °C for 1 hr followed by beads washing with 1x BW buffer.

RNA was then ligated to a 5′-RNA adapter (A3-RL5) by T4 RNA ligase plus 15% PEG and 500 U/ml RNasin at 16 °C for 16 hr. RNA was reverse transcribed by Superscript III primed with 2.5 μM DP3-RT primer at (50 °C for 45 min, 55 °C for 15 min, heat inactivate at 90 °C for 3 min). Samples were PCR amplified by AccuPrimePfxSuperMix (Invitrogen), 10 μM primers (DP3, DP5) and run at (94 °C for 2 min, 16 cycles (94 °C for 20 s, 58 °C for 30 s, 68 °C for 45 s), and 68 °C for 5 min). PCR products were resolved on 3% Metaphor gel and fragments (~500–800 bp) were gel-excised and purified. Re-PCR was set up using 5 μM primers that were compatible with Illumina sequencer, and run at (94 °C for 2 min, 6–8 cycles (94 °C for 20 s, 60 °C for 30 s, 68 °C for 40 s), and 68 °C for 5 min). PCR products were resolved on 3% Metaphor gels and fragments (~500–800 bp) were gel-excised and purified.

To capture the fraction of 5′-OH RNA and compare it to that of capped RNAs, total RNA that had been depleted of rRNAs and small capped RNAs was Terminator treated to remove 5′-P bearing RNAs, and divided equally into two aliquots. To detect 5′-OH RNAs, the first aliquot was phosphorylated using T4 PNK in the presence of γ^{32}P-ATP and incubated at 37 °C for 1 hr. RNA was phenol:chloroform (Ambion) extracted and ethanol precipitated. To detect capped RNAs, the second aliquot was first treated with T4 PNK, to phosphorylate any RNAs bearing 5′-OH, followed by Terminator treatment to remove them. RNA was extracted and ethanol precipitated, and treated by Tobacco Acid Pyrophosphatase (TAP, Ambion) at 37 °C for 1 hr according to manufacturer's protocol to remove the 5′ cap, leaving a residual 5′-P. RNA was extracted and ethanol precipitated, and treated with CIP at 37° C for 30 min to remove 5′-P left by TAP and was then treated with T4 PNK in the presence of γ^{32}P-ATP. RNA was extracted and ethanol precipitated. Radiolabelled RNAs derived from 5′-OH RNA or from capped RNAs were resolved on 6% denaturing PAGE and visualized by autoradiography. We attempted to make Akron3 libraries of 5′-OH bearing RNAs but were unable to do so due to their negligible amount.

Preparation of Akron5 libraries

To capture 5′ ends of poly(A), non-capped RNAs, total RNA (~100–150 µg) was incubated with Oligo(dT)$_{25}$ Dynabeads (Ambion) according to the manufacturer's protocol. Beads were then incubated with T4 PNK at 37 °C for 1 hr to phosphorylate RNA 5′ ends. Beads were washed and poly (A)-enriched RNA was ligated to a 5′-biotinylated RNA adapter (A5-RL5) with T4 RNA ligase plus 15% PEG and 500 U/ml RNasin at 16 °C for 16 hr. Capped RNAs are carried over, but only non-capped RNAs ligate to the 5′-adapter. As described in Akron3, beads were washed and RNA was fragmented in 1x fragmentation buffer. Ligated RNA was captured by M280 Streptavidin Dynabeads and was treated with T4 PNK to repair RNA 3′ ends. After beads washes, RNA was ligated to a 3′-RNA adapter (A5-RL3) with T4 RNA ligase. cDNA was generated with Superscript III primed with DP3-RT primer and incubation at (50 °C for 45 min, 55 °C for 15 min, heat inactivate at 90 °C for 3 min). PCR and re-PCR reactions were set up as described in Akron3. PCR products were resolved on 3% Metaphor gel. Fragments (~500–800 bp) were gel-excised and purified. All Akron-Seq libraries were sequenced on Illumina NextSeq500 (75PE).

Primer and adapter sequences

See Supplementary Table 1

Oligonucleotides sequences

See Supplementary Table 2

Akron-SMRT data preprocessing

To avoid potential sequence fragments and make sure that we capture the whole insert RNA, we only used circular consensus (CCS) reads for Akron-SMRT analysis. Poly(A) sequences were trimmed from CCS reads using an in-house script that identifies A stretches longer than 6-nt and closer than 4-nt to the read 3′ end. The trimmed poly(A) tails were retained for further analysis. Trimmed reads were then aligned to the human genome (hg19) with STARlong (v2.4.1) using the parameters --outFilterMultimapScoreRange 0 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0.66 --outFilterMismatchNmax 1000 --winAnchorMultimapNmax 200 --seedSearchLmax 30 --seedSearchStartLmax 12 --seedPerReadNmax 100000 --seedPerWindowNmax 100 --alignTranscriptsPerReadNmax 100000 --alignTranscriptsPerWindowNmax 10000 --alignSJDBoverhangMin 1.

Akron5 data preprocessing

The adapter sequences were trimmed from the paired-end reads using the cutadapt software in paired-end mode. The adapter GTGTCAGTCACTTCCAGCGG was trimmed from the 3′ end of read1 whereas CCGCATCGTCCTCCCT was trimmed from the 3′ end of read2. A constant adapter sequence of 16-nt was removed from the 5′ end of read1. Only reads longer than 25-nt after adapter trimming were retained. The reads were aligned to the human genome (hg19) in a two-step approach using the STAR aligner (v2.5.2a). First, reads were aligned in paired-end mode. To exclude potential PCR artifacts, reads with the same 5′ end (defined from read1) and the same 3′ end (defined from read2) were collapsed. The retained

reads were then re-aligned in single end mode using read1 only and the following parameters --outFilterMultimapScoreRange 0 --alignIntronMax 50000 --outFilterMatchNmin 8 --outFilterMatchNminOverLread 0.7 --sjdbOverhang 80 --alignSJDBoverhangMin 1 --seedSearchStartLmax 15. Aligned reads were loaded into a SQLite3 database for further processing with CLIPSeqTools [63] and were annotated with information whether they overlap with elements from RepeatMasker (downloaded from UCSC), rRNAs (extracted from UCSC gene model annotation file) and genes (from UCSC gene model annotation file).

Akron3 data preprocessing

The adapter sequences were trimmed from the paired-end reads using the same process as for Akron5. The reads were aligned to the human genome in a two-step approach similar to Akron5. Unlike Akron5, in the second step only read2 were aligned after they were reverse complemented to facilitate downstream analysis. The STAR parameters were selected to permit the alignment of reads with poly(A) tails. Specifically, the following parameters were used --alignIntronMax 50000 --outFilterMultimapScoreRange 0 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --outFilterScoreMin 30 --outFilterScoreMinOverLread 0 --outFilterMatchNmin 30 --outFilterMatchNminOverLread 0 --sjdbOverhang 100 --alignSJDBoverhangMin 1. Similar to Akron3, aligned reads were loaded into a SQLite3 database for further processing and annotation.

Genomic distribution

All mapped reads were assigned to genomic categories based on whether they were contained in annotated elements of that category. Reads that were not contained in any element were marked as unannotated. For genes, 5′ UTRs, CDSs, and 3′ UTRs we measured uniquely mapped reads only.

Correlation of replicates

Read abundance was measured as Reads Per Kilobase of transcript per Million mapped reads (RPKM) for protein-coding transcripts. Pearson's correlation was calculated using R and plots were created with the corrplot package.

End distribution on meta-mRNA

The distribution of collapsed ends was measured in 20 bins across the whole mRNA, 5′ UTR, CDS and 3′ UTR (therein called elements) and for all protein coding transcripts. Only genic elements and mRNAs that are longer than 200-nt were used. For each element of size L, the number of ends was normalized by the bin length (L/20) and the library size to be comparable across libraries and different transcripts. For each bin, the normalized metric was summed across all elements and the resulting value was averaged for replicate libraries. The last bin of the 3′ UTR was defined as full-length for Akron3.

Identification of poly(A) tails

To identify reads with non-templated poly(A) tails for Akron3, specific parameters were used at the alignment step (see Akron3 data preprocessing). Poly(A) tails are reported by the

aligner as soft-clipped sequences. These soft-clipped sequences were extracted from the sam/bam alignment file using an in-house script and if the A content was higher than 80%, the read was annotated as having a poly(A) tail. For Akron-SMRT poly(A) identification see data preprocessing section.

Relative position to RPFs

The distribution of collapsed ends (ends on the same position as counted as one) from other libraries was measured in the area around RPF 5′ ends by measuring the pairwise distance of all ends to all RPFs (Fig. 1d, e). To reduce 5′ end uncertainty for RPFs, only RPFs of specific size were used. The size was selected so that it has the highest enrichment in any of the open reading frames. The selected RPF sizes were 29-nt for human and 28-nt for yeast. The analysis was performed for ends that map within the CDS only unless otherwise noted in the text with an offset of 100-nt from the start and stop codon to avoid distribution confounds from potential specific mechanisms acting close to the start and stop codons. All density plots were normalized so that the area under the plot is 1. Random controls were created by randomly assigning each end to a new random position within the same transcript, thus maintaining sequence and expression features.

Random control maintaining ORF and nucleotide content around Akron ends

To select random positions for Akron5 and Akron3 and maintain the nucleotide content in the surrounding region x (x = [−1, 0] for Akron5 ends and x = [0, 1] for Akron3 ends) a generative model for the nucleotide content at each position within x is calculated. Three distinct nucleotide composition models for region x were calculated, corresponding to each of the three ORFs. For each end a model is selected based on the ORF that the end maps to. Using this model, one sequence of length equal to x was generated and an exact match within the same gene transcript was found. If no match was found the process was repeated until a match was found (max: 1000 times). The final random position was randomly set at one of the matching locations thus preserving the aggregate nucleotide distribution of the surrounding region x.

Relative position to stop codon

Reads with the same end were collapsed and their density was measured around the last nucleotide of the stop codon. The bottom 25th quantile of protein-coding transcripts for each library was excluded from the analysis. The analysis was performed independently for each library and replicates were only summed for visualization.

Discrete Fourier Transform

DFT converts a signal to a representation in the frequency domain. DFT was calculated using R and the GeneCycle package. To facilitate interpretation, we plot the power at each period instead of at each frequency.

G-quadruplex prediction

The G-quadruplex prediction was performed with an in-house script using the pattern G{2,4}.{1,7}G{2,4}.{1,7}G{2,4}.{1,7}G{2,4} as described in [55] within the CDS of

protein-coding genes. Prediction was performed in the sequence ± 150-nt of each read end as well as for random and shuffled controls. For random control, a random position within the same transcript was selected for each read end. For shuffled control, the sequence around the read ends was shuffled at the codon level. The later maintains the nucleotide and codon content and evaluates whether rG4 structures are preferentially selected for, beyond what would be expected by any increase in G content alone.

Ribothrypsis model simulation

For detailed description of ribothrypsis model simulation please see the legend of Supplementary Fig. 3.

External sequencing data

The following datasets were downloaded and used in this study:

Homo sapiens: PARE-Seq [34]; Ribo-Seq [39] (GSM2100602); Total RNA-Seq [39] (GSM2100594); TAIL-Seq [37]; Small RNA-Seq [43]; DMS-Seq [51]; rG4-Seq [53,54]; and BRIC-Seq [47].

Saccharomyces cerevisiae: 5PSeq [13] and Ribo-Seq [42].

CODE AVAILABILITY

Source code required for analysis has been deposited on GitHub (https://github.com/mnsmar/ribothrypsis).

DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available in the Gene Expression Omnibus repository, under accession number GSE107838.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol. 2002; 3:195–205. [PubMed: 11994740]

2. Singh G, Pratt G, Yeo GW, Moore MJ. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu Rev Biochem. 2015; 84:325–54. [PubMed: 25784054]

3. Green R, Noller HF. Ribosomes and translation. Annu Rev Biochem. 1997; 66:679–716. [PubMed: 9242921]

4. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. Nat Rev Genet. 2012; 13:246–59. [PubMed: 22392217]

5. Houseley J, Tollervey D. The many pathways of RNA degradation. Cell. 2009; 136:763–76. [PubMed: 19239894]

6. Lykke-Andersen S, Tomecki R, Jensen TH, Dziembowski A. The eukaryotic RNA exosome: same scaffold but variable catalytic subunits. RNA Biol. 2011; 8:61–6. [PubMed: 21289487]

7. Coller J, Parker R. Eukaryotic mRNA decapping. Annu Rev Biochem. 2004; 73:861–90. [PubMed: 15189161]

8. Grudzien-Nogalska E, Kiledjian M. New insights into decapping enzymes and selective mRNA decay. Wiley Interdiscip Rev RNA. 2017; 8

9. Jens M, Rajewsky N. Competition between target sites of regulators shapes post-transcriptional gene regulation. Nat Rev Genet. 2015; 16:113–26. [PubMed: 25488579]

10. Vlachos IS, Hatzigeorgiou AG. Functional Analysis of miRNAs Using the DIANA Tools Online Suite. Methods Mol Biol. 2017; 1517:25–50. [PubMed: 27924472]

11. Richter JD, Coller J. Pausing on Polyribosomes: Make Way for Elongation in Translational Control. Cell. 2015; 163:292–300. [PubMed: 26451481]

12. Hu W, Sweet TJ, Chamnongpol S, Baker KE, Coller J. Co-translational mRNA decay in Saccharomyces cerevisiae. Nature. 2009; 461:225–9. [PubMed: 19701183]

13. Pelechano V, Wei W, Steinmetz LM. Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell. 2015; 161:1400–12. [PubMed: 26046441]

14. Yu X, Willmann MR, Anderson SJ, Gregory BD. Genome-Wide Mapping of Uncapped and Cleaved Transcripts Reveals a Role for the Nuclear mRNA Cap-Binding Complex in Cotranslational RNA Decay in Arabidopsis. Plant Cell. 2016; 28:2385–2397. [PubMed: 27758893]

15. Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev. 2007; 21:1833–56. [PubMed: 17671086]

16. Shoemaker CJ, Green R. Translation drives mRNA quality control. Nat Struct Mol Biol. 2012; 19:594–601. [PubMed: 22664987]

17. Brandman O, Hegde RS. Ribosome-associated protein quality control. Nat Struct Mol Biol. 2016; 23:7–15. [PubMed: 26733220]

18. Inada T. The Ribosome as a Platform for mRNA and Nascent Polypeptide Quality Control. Trends Biochem Sci. 2017; 42:5–15. [PubMed: 27746049]

19. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet. 2004; 36:1073–8. [PubMed: 15448691]

20. Doma MK, Parker R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature. 2006; 440:561–4. [PubMed: 16554824]

21. Frischmeyer PA, et al. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science. 2002; 295:2258–61. [PubMed: 11910109]

22. Tsuboi T, et al. Dom34:hbs1 plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3′ end of aberrant mRNA. Mol Cell. 2012; 46:518–29. [PubMed: 22503425]

23. Ikeuchi K, Inada T. Ribosome-associated Asc1/RACK1 is required for endonucleolytic cleavage induced by stalled ribosome at the 3′ end of nonstop mRNA. Sci Rep. 2016; 6:28234. [PubMed: 27312062]

24. Guydosh NR, Green R. Translation of poly(A) tails leads to precise mRNA cleavage. RNA. 2017

25. Shoemaker CJ, Eyler DE, Green R. Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. Science. 2010; 330:369–72. [PubMed: 20947765]

26. Shoemaker CJ, Green R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. Proc Natl Acad Sci U S A. 2011; 108:E1392–8. [PubMed: 22143755]

27. Pisareva VP, Skabkin MA, Hellen CU, Pestova TV, Pisarev AV. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. EMBO J. 2011; 30:1804–17. [PubMed: 21448132]

28. Becker T, et al. Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. Nat Struct Mol Biol. 2011; 18:715–20. [PubMed: 21623367]

29. Bengtson MH, Joazeiro CA. Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. Nature. 2010; 467:470–3. [PubMed: 20835226]

30. Brandman O, et al. A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell. 2012; 151:1042–54. [PubMed: 23178123]

31. Defenouillere Q, et al. Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. Proc Natl Acad Sci U S A. 2013; 110:5046–51. [PubMed: 23479637]

32. Verma R, Oania RS, Kolawa NJ, Deshaies RJ. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. Elife. 2013; 2:e00308. [PubMed: 23358411]

33. German MA, et al. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol. 2008; 26:941–6. [PubMed: 18542052]

34. Schmidt SA, et al. Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. Nucleic Acids Res. 2015; 43:309–23. [PubMed: 25429978]

35. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol. 2008; 18:758–62. [PubMed: 18472421]

36. Gregory BD, et al. A link between RNA metabolism and silencing affecting Arabidopsis development. Dev Cell. 2008; 14:854–66. [PubMed: 18486559]

37. Chang H, Lim J, Ha M, Kim VN. TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. Mol Cell. 2014; 53:1044–52. [PubMed: 24582499]

38. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009; 324:218–23. [PubMed: 19213877]

39. Park JE, Yi H, Kim Y, Chang H, Kim VN. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. Mol Cell. 2016; 62:462–71. [PubMed: 27153541]

40. Carlevaro-Fita J, Rahim A, Guigo R, Vardy LA, Johnson R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA. 2016; 22:867–82. [PubMed: 27090285]

41. Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB. Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. Nucleic Acids Res. 2012; 40:e54. [PubMed: 22241775]

42. Guydosh NR, Green R. Dom34 rescues ribosomes in 3′ untranslated regions. Cell. 2014; 156:950–62. [PubMed: 24581494]

43. Lubas M, et al. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep. 2015; 10:178–92. [PubMed: 25578728]

44. Schmidt C, et al. The cryo-EM structure of a ribosome-Ski2-Ski3-Ski8 helicase complex. Science. 2016; 354:1431–1433. [PubMed: 27980209]

45. Martinez J, Tuschl T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. Genes Dev. 2004; 18:975–80. [PubMed: 15105377]

46. Orban TI, Izaurralde E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA. 2005; 11:459–69. [PubMed: 15703439]

47. Tani H, et al. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. 2012; 22:947–56. [PubMed: 22369889]

48. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011; 147:789–802. [PubMed: 22056041]

49. Spitale RC, et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature. 2015; 519:486–90. [PubMed: 25799993]

50. Silverman IM, Berkowitz ND, Gosai SJ, Gregory BD. Genome-Wide Approaches for RNA Structure Probing. Adv Exp Med Biol. 2016; 907:29–59. [PubMed: 27256381]

51. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature. 2014; 505:701–5. [PubMed: 24336214]

52. Song J, Perreault JP, Topisirovic I, Richard S. RNA G-quadruplexes and their potential regulatory roles in translation. Translation (Austin). 2016; 4:e1244031. [PubMed: 28090421]

53. Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods. 2016; 13:841–4. [PubMed: 27571552]

54. Guo JU, Bartel DP. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science. 2016; 353

55. Vourekas A, et al. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev. 2015; 29:617–29. [PubMed: 25762440]

56. Endoh T, Sugimoto N. Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. Sci Rep. 2016; 6:22719. [PubMed: 26948955]

57. Guydosh NR, Kimmig P, Walter P, Green R. Regulated Ire1-dependent mRNA decay requires no-go mRNA degradation to maintain endoplasmic reticulum homeostasis in S. pombe. Elife. 2017; 6

58. Simms CL, Yan LL, Zaher HS. Ribosome Collision Is Critical for Quality Control during No-Go Decay. Mol Cell. 2017; 68:361–373e5. [PubMed: 28943311]

59. Simms CL, Hudson BH, Mosior JW, Rangwala AS, Zaher HS. An active role for the ribosome in determining the fate of oxidized mRNA. Cell Rep. 2014; 9:1256–64. [PubMed: 25456128]

60. Peach SE, York K, Hesselberth JR. Global analysis of RNA cleavage by 5′-hydroxyl RNA sequencing. Nucleic Acids Res. 2015; 43:e108. [PubMed: 26001965]

61. Ibrahim F, et al. Identification of in vivo, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. Cell Rep. 2013; 3:301–8. [PubMed: 23416048]

62. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature. 2014; 508:66–71. [PubMed: 24476825]

63. Maragkakis M, Alexiou P, Nakaya T, Mourelatos Z. CLIPSeqTools--a novel bioinformatics CLIP-seq analysis suite. RNA. 2015; 22:1–9. [PubMed: 26577377]
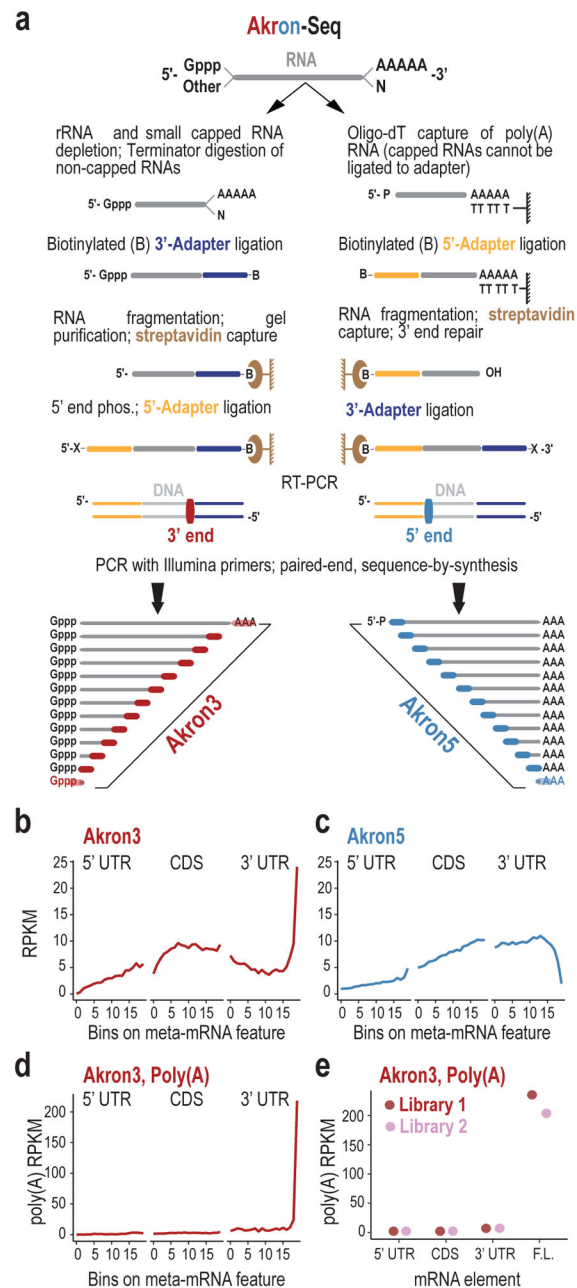
Figure 1. Akron-Seq identifies native mRNA ends at single nucleotide resolution and captures mRNA decay *in vivo*

(**a**) Schematic of Akron-Seq.

(**b** – **d**) Distribution of Akron3 (**b**), Akron5 (**c**) and Akron3 poly(A) (**d**) ends from all libraries, plotted on meta-mRNA regions as reads or poly(A) reads per kilobase per million (RPKM), normalized by bin length. The drop of Akron5 reads close to the poly(A) is technical, due to selection of RNAs longer than ~200 nt, to avoid abundant small noncoding RNAs.

(**e**) Distribution of Akron3 poly(A) ends from the two Akron3 libraries corresponding to the 5′ untranslated region (UTR), coding sequence (CDS), 3′ UTR or full-length (F.L.) mRNAs.
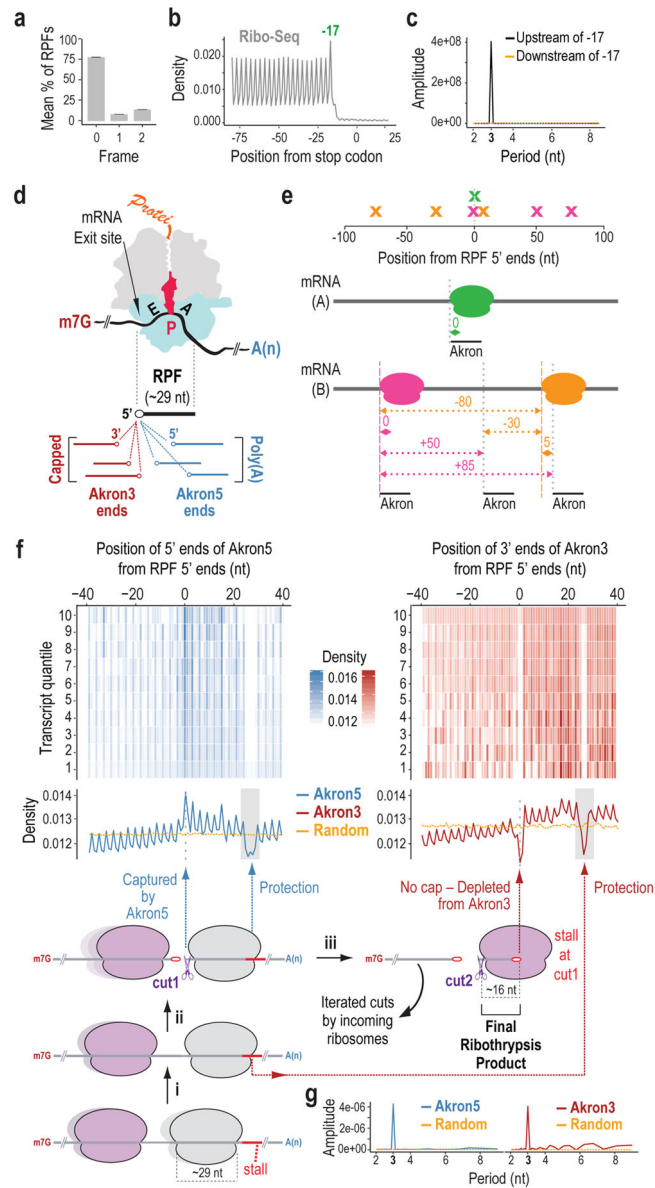
Figure 2. Ribothrypsis mediates ribosome-phased mRNA endonucleolytic cuts during translation elongation

(**a**) Distribution of 5′ ends of ribosome-protected fragments (RPFs; derived from Ribo-Seq) in reading frames of HeLa protein-coding transcripts; mean ± standard error of the mean (s.e.m.), n = 34,054 protein-coding transcripts.

(**b**, **c**) Density of HeLa RPFs' 5′ ends upstream and downstream of stop codon (last nucleotide of stop codon set as position 0) (**b**) and Discrete Fourier Transform (DFT) for the density upstream and downstream of position −17 (corresponding to the terminating ribosome 5′ end) (**c**).

(**d**) Schematic of elongating ribosome with RPF and relative positioning analyses for Akron3 and Akron5; E: tRNA exit site, P: peptidyl-tRNA site, A: aminoacyl-tRNA binding

site. Blue: 40S subunit with mRNA channel; Grey: 60S subunit with polypeptide channel; Red: peptidyl-tRNA attached to nascent protein.

(**e**) Schematic for calculation of Akron5 relative positioning plot to RPFs. The calculation for the density plot at the top is deconstructed in the schematic below. All pairwise relative positions of Akron5 ends with RPFs that map on the same transcript are calculated. A density plot is created by measuring the number of times each relative position is found. Relative positions are color-coded depending on the corresponding RPF. The same approach is used to calculate Akron3 relative positioning to RPFs.

(**f**) Density heat maps (protein-coding transcripts grouped in ten quantiles, top) and cumulative density plots (middle) of Akron5 and Akron3 ends, or random controls, relative to RPF 5′ ends (corresponding to position 0) in CDS. Quantiles are sorted (top to bottom) from high to low number of ends. Bottom: schematic of the sequence of captured events.

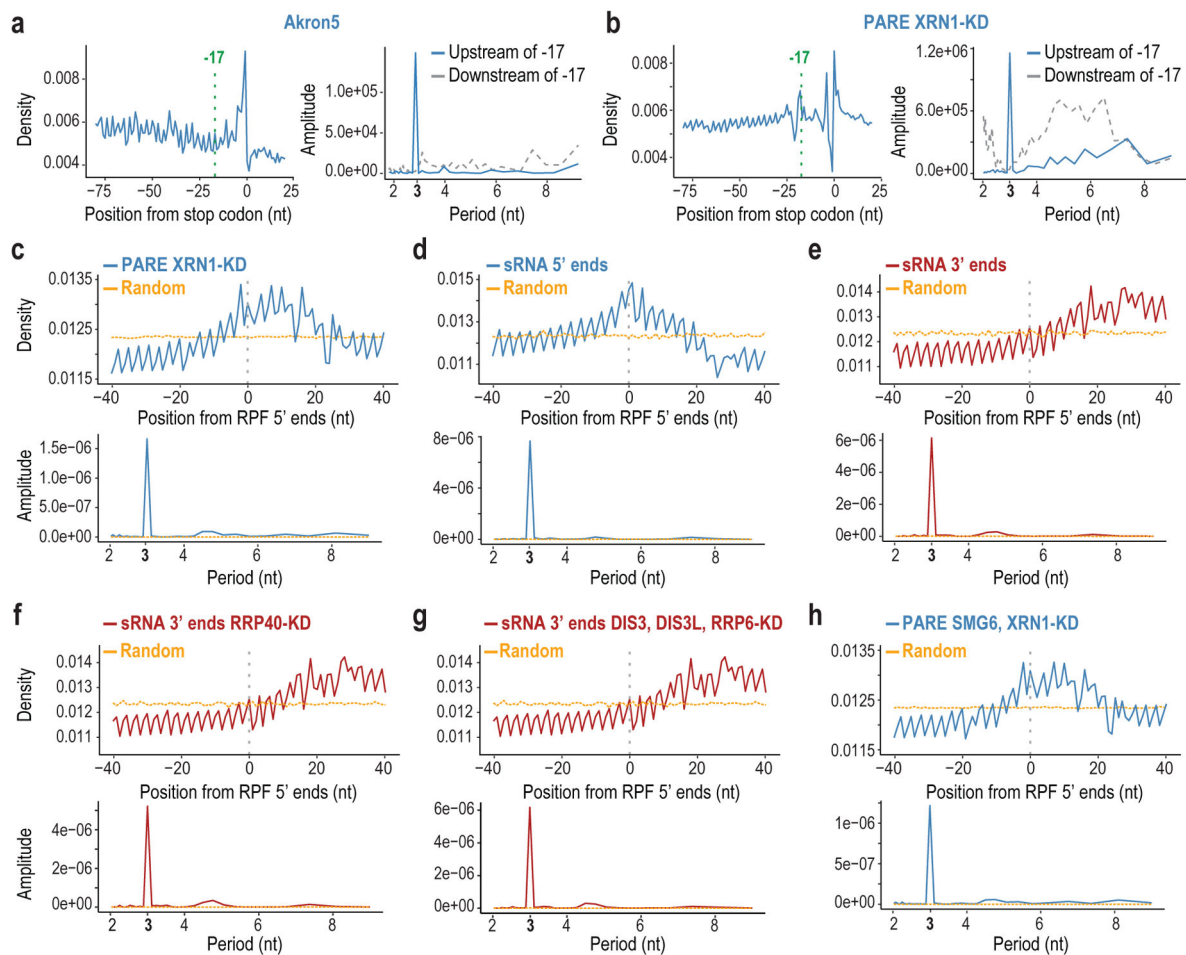(**g**) DFT of read density around RPFs for Akron5, Akron3 and random controls.

Figure 3. Ribothrypsis is not mediated by XRN1, the exosome or the SMG6 endonuclease and its end products are captured in small RNA sequencing libraries

(**a**, **b**) 5′ end density upstream and downstream of stop codon (last nucleotide of stop codon set as position 0) and DFT upstream and downstream of position −17 (corresponding to the terminating ribosome 5′ end, marked with a green dotted line) for Akron5 (**a**), and PARE-Seq after XRN1 knockdown (KD) (**b**).

(**c** – **h**) Density and DFT around the RPFs' 5′ end (position 0) of 5′ or 3′ ends from indicated libraries in CDS; sRNA = small RNA libraries.
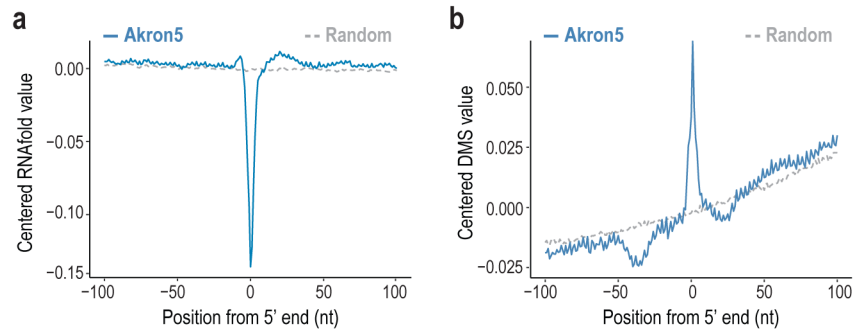
Figure 4. Ribothrypsis cuts occur in mRNA areas with open structure
(**a**, **b**) Centered RNAfold secondary structure values (**a**) and centered values from *in vivo* DMS-Seq
(**b**) upstream and downstream of Akron5 ends that map in CDS, versus random controls.
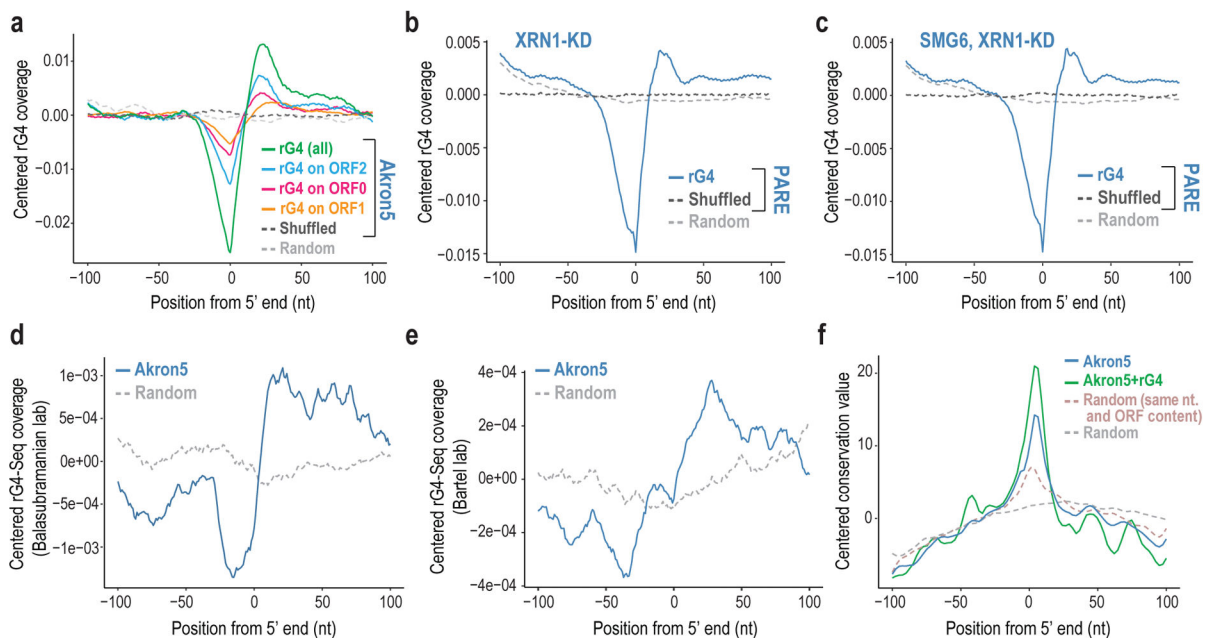
Figure 5. Ribothrypsis associates with RNA G-quadruplexes (rG4) and conserved RNA areas
(**a**) Centered coverage of predicted rG4s, located in indicated positions relative to reading frame (ORF) or cumulative (all), upstream and downstream of Akron5 ends that map in CDS, versus random and shuffled controls.

(**b, c**) Centered coverage of predicted rG4s upstream and downstream of PARE-Seq 5′ ends that map in CDS, versus random and shuffled controls, in indicated knockdowns (KD).

(**d, e**) Centered coverage of rG4-Seq from Balasubramanian lab [53] (**d**) or Bartel lab [54] (**e**), upstream and downstream of Akron5 5′ ends that map in CDS with corresponding random control.

(**f**) Evolutionary conservation for 100 vertebrates (PhastCons) upstream and downstream of Akron5 ends (blue) and Akron5 ends with rG4s (green) in CDS. A random control that maintains the nucleotide (nt) and ORF context of Akron5 ends (dashed pink) and a completely random control (dashed grey) are shown.
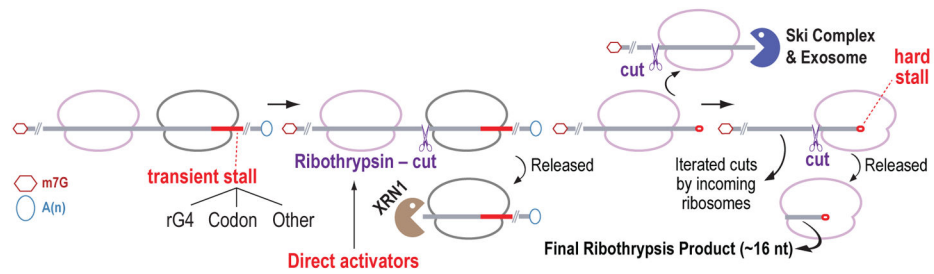
Figure 6. Ribothrypsis model

Transient ribosome stalling caused by various mechanisms can activate ribothrypsin to cleave the mRNA as it exits the channel of the ribosome that has encountered a transient stall. Ribothrypsis is propagated as incoming ribosomes encounter the initial ribothrypsis cut (hard stall). Direct activators, transcript-specific or global, of ribothrypsin may also initiate or propagate ribothrypsis. Intersection of ribothrypsis with other prominent RNA decay pathways is shown.