

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

Fadia Ibrahim^{1,2,3,4}, Manolis Maragkakis^{1,2,3,4}, Panagiotis Alexiou^{1,2,3} and Zissimos Mourelatos^{1,2,3*}

mRNAs 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 the native 3' and 5' ends of capped and polyadenylated RNAs, respectively, we show that canonical human mRNAs are subject to repeated cotranslational and ribosome-phased endonucleolytic cuts at the exit site of the mRNA ribosome channel, in a process that we term ribothrypsis. We uncovered RNA G quadruplexes among likely ribothrypsis triggers and show that ribothrypsis is a conserved process. Strikingly, we found that mRNA fragments are abundant in living cells and thus have 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.

Messenger RNAs, typically found in cells as ribonucleoproteins^{1,2}, engage ribosomes, tRNAs and other factors that translate the genetic information transcribed in mRNAs to instruct protein production³. Canonical mRNA decay⁴ involves deadenylation followed by 3'-to-5' exonucleolysis by the exosome^{5,6}, and decapping^{7,8} followed by 5'-to-3' exonucleolysis by XRN1 (ref. 4). Numerous trans-acting factors^{9,10} and cis RNA elements, including codons¹¹, 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 (XRN4 in plants) degrades decapped mRNAs following the last translating ribosome¹²⁻¹⁴. This view is conceptually appealing because the mRNA that undergoes 5'-to-3' decay will not be able to reinitiate translation, and 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, studies of mRNA-surveillance pathways, which initiate degradation of defective mRNAs during translation, have provided a better understanding of how cells handle aberrant mRNAs¹⁵⁻¹⁸. 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, aberrant mRNAs with strong stalls or lacking stop codons, respectively, trigger cleavages upstream of stalled ribosomes in a process mediated by an as-yet-unknown endonuclease^{16,18,20-24}. Stalled ribosomes are recognized by Dom34 (PELOTA) and Hbs1 proteins, homologs of the eukaryotic release factors eRF1 and eRF3, respectively, which along with Rli1 (ABCE1) split the stalled ribosome into 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 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 developed Akron-seq, a new approach capturing native 3' and 5' ends of capped and polyadenylated RNAs, respectively. We used this approach to demonstrate that canonical human mRNAs undergo repeated cotranslational and ribosome-phased endonucleolytic cuts at the exit site of the mRNA ribosome channel, in a process that we denote ribothrypsis. We further uncovered likely triggers and features of ribothrypsis and found that produced mRNA fragments are abundant in living cells, thus providing important implications for interpreting data from experiments, such as RNA-seq, that are predicated 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 through a method that we developed called Akron-SMRT (from *ακρον*, ultimate point, starting or ending; Supplementary Fig. 1a). In this method, capped RNAs are enriched through digestion with Terminator 5' phosphate (P)-dependent exonuclease. SMRT did not provide adequate sequencing depth for whole-transcriptome analysis or accurate quantification. Instead, we interrogated randomly selected 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 through high-resolution northern blotting (Supplementary Fig. 1d-f) and excluded exogenous RNA degradation as the source for fragmentation (Supplementary Fig. 1g).

Akron-seq uncovers ribothrypsis. Next, we sought to investigate the abundance of mRNA fragments in cells and the mechanism

¹Department of Pathology and Laboratory Medicine, Division of Neuropathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ²Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Penn Medicine Translational Neuroscience Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴These authors contributed equally: Fadia Ibrahim and Manolis Maragkakis. *e-mail: mourelaz@uphs.upenn.edu

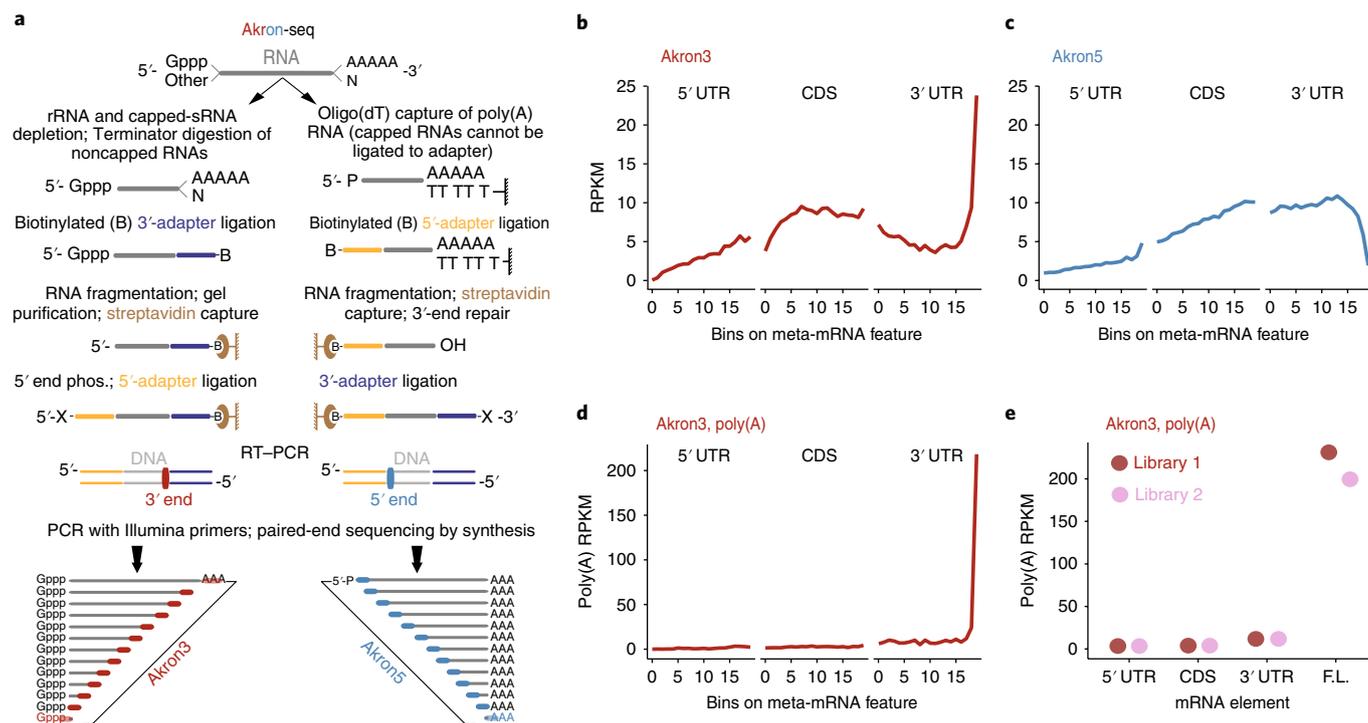


Fig. 1 | Akron-seq identifies native mRNA ends at single-nucleotide resolution and captures mRNA decay in vivo. **a**, Schematic of Akron-seq. Phos, phosphorylation. **b–d**, Distribution of Akron3 (**b**), Akron5 (**c**) and Akron3 poly(A) (**d**) ends from all libraries, plotted on meta-mRNA regions as reads or poly(A) reads per kilobase per million (RPKM), normalized by bin length. The drop in Akron5 reads close to the poly(A) is due to technical reasons, because of selection of RNAs longer than ~200 nt to avoid abundant noncoding sRNAs. **e**, Distribution of Akron3 poly(A) ends from the two Akron3 libraries corresponding to the 5' UTR, CDS, 3' UTR or full-length (F.L.) mRNAs.

through which they are generated. Methods that capture the 5' ends of polyadenylated mRNAs, such as parallel analysis of RNA ends (PARE)-seq^{33,34}, degradome-seq³⁵, genome-wide mapping of uncapped and cleaved transcripts (GMUCT)³⁶ and 5'-phosphorylated-end sequencing (5Pseq)¹³; or the 3' ends of transcripts, such as TAIL-seq³⁷, 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 and 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 methods, RNA native ends are marked by adapter ligation before fragmentation or cDNA generation and are identified by paired-end sequencing by synthesis (Fig. 1a). In Akron3, RNA is digested with Terminator to enrich for capped molecules. Because Terminator does not degrade RNAs bearing 5'-OH, we also set out to capture such molecules and compare their abundance with that of capped RNAs; however, we found that they were 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 because we did not obtain a specific PCR product (Supplementary Fig. 2b). This result was consistent with findings from the Steinmetz laboratory, which has also been unable to generate libraries from RNAs bearing 5'-OH in *Saccharomyces cerevisiae*¹³. Thus, Akron3 libraries are overwhelmingly derived from capped RNAs. For Akron5, we isolated polyadenylated RNAs by 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 were able to interrogate its positional relationships with respect to various mRNA features and processes, including translation, to uncover mechanistic insights into fragment generation. We prepared biological-triplicate

(Akron5) and biological-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 (all mRNAs binned across their length), 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 nontemplated adenines at the 3' end. The latter mapped primarily at the 3' ends of genes, as expected (Fig. 1d,e). Moreover, by interrogating the 50% most expressed transcripts, as quantified by total-RNA-seq, we found that ~63% of those transcripts had Akron3 ends that mapped within the CDS, thus indicating ubiquitous fragmentation.

To examine whether the generation of these native 5' and 3' ends was related to translation, we compared Akron-seq with ribosome profiling (ribo-seq)³⁸ data from HeLa cells³⁹. We detected that ribosome-protected fragments (RPFs) mapped overwhelmingly in frame with codons (Fig. 2a) and displayed 3-nt periodicity with stop codons (Fig. 2b), as verified through discrete Fourier transform (DFT) (Fig. 2c), which is reflective of the stepwise ribosome movement during translation elongation. Next, we plotted the distribution of Akron5 and Akron3 ends upstream and downstream of RPF 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. We observed a striking 3-nt periodicity for both ends for all transcripts, in contrast to random controls (Fig. 2f).

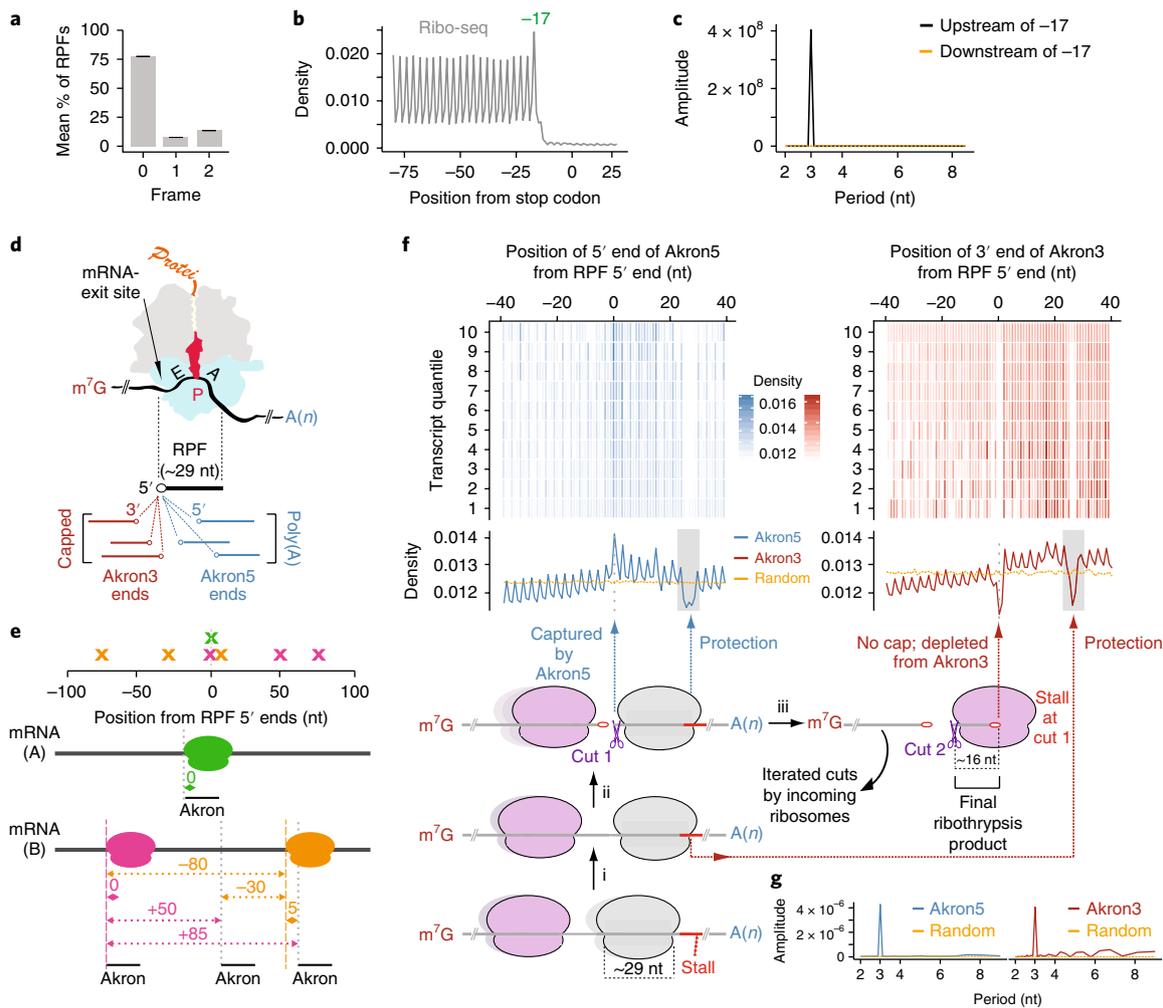


Fig. 2 | Ribothrypsis mediates ribosome-phased mRNA endonucleolytic cuts during translation elongation. **a**, Distribution of 5' ends of RPFs (derived from ribo-seq) in reading frames of HeLa protein-coding transcripts; mean \pm s.e.m., $n = 34,054$ protein-coding transcripts. **b,c**, Density of HeLa RPF 5' ends upstream and downstream of the stop codon (last nucleotide of the stop codon set as position 0; **b**) and 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; gray, 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 top is deconstructed in the schematic below. All pairwise relative positions of Akron5 ends with RPFs mapping on the same transcript are calculated. A density plot was created by measuring the number of times each relative position was found. Relative positions are color-coded depending on the corresponding RPF. The same approach was 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 numbers of ends. Bottom, schematic of the sequence of captured events. **g**, DFT of read density around RPFs for Akron5, Akron3 and random controls.

This result was verified with DFT (Fig. 2g) and indicated cotranslational generation of the 5' and 3' ends of mRNA fragments. We found a similar periodicity when we analyzed TAIL-seq data³⁷ (Supplementary Fig. 3a) and when we interrogated long non-coding RNAs loaded with ribosomes (Supplementary Fig. 3b), in agreement with recent reports of polysomal association and degradation of long noncoding RNAs⁴⁰. To exclude the possibility that the observed pattern was created by the combined effect of codon-associated nucleotide periodicity and any potential, though unlikely⁴¹, nucleotide preference by T4 RNA ligase, we downsampled our Akron libraries, constraining the nucleotide content at the Akron ends to 25% for each nucleotide. We found that the plots for the downsampled controls were almost identical to the original libraries, thereby verifying the independence of the observed results from any such bias (Supplementary Fig. 3c).

Collectively, our results indicated a universal process of cotranslational RNA degradation that operates on most transcripts that engage with ribosomes.

By measuring the positions of Akron5 and Akron3 ends relative to the RPF 5' ends (Fig. 2f), we observed the coincidence of an Akron5 peak and an Akron3 drop around position 0. By experimental design, Akron5 ends corresponded to the 5' ends of polyadenylated mRNAs, and Akron3 ends corresponded to the 3' ends of capped RNAs. Therefore, the coincidence of the two signals at position 0 could not have resulted from 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 indicated that the ends must have been created by an endonucleolytic activity that was associated with the ribosome and that cleaved the translated mRNA as it exited the

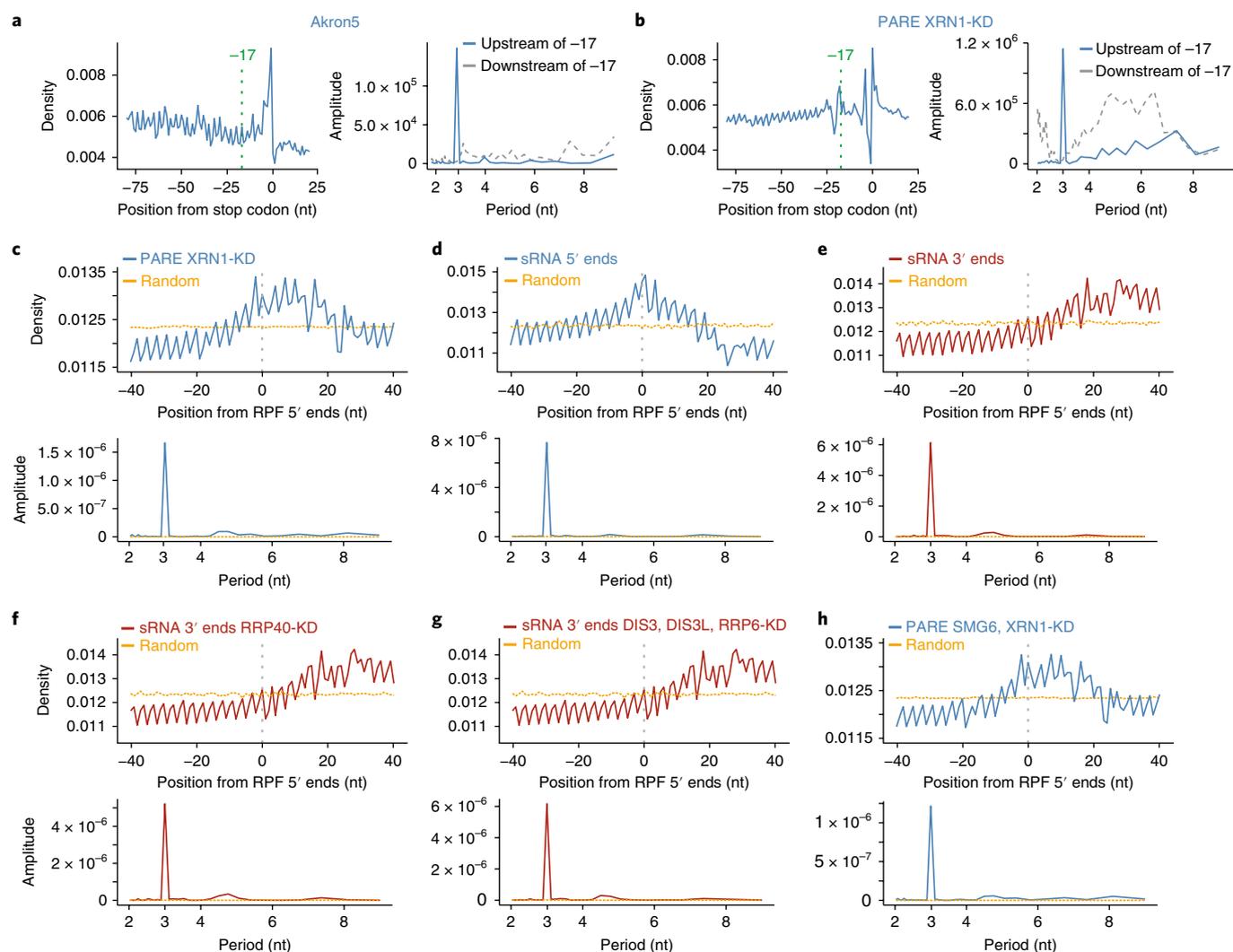


Fig. 3 | Ribothrypsis is not mediated by XRN1, the exosome or SMG6 endonuclease, and its end products are captured in sRNA-sequencing libraries. **a,b**, 5'-end density upstream and downstream of the stop codon (last nucleotide of stop codon set as position 0) and DFT upstream and downstream of position -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 RPF 5' end (position 0) of 5' or 3' ends from the indicated libraries in CDS. sRNA, sRNA libraries.

ribosome channel (Fig. 2d; mRNA exit is located between the head and the platform of the 40S subunit).

We hypothesized that the endonucleolytic cut was triggered during translation elongation just upstream of a transiently stalled ribosome (stall marked with a gray box in Fig. 2f, located ~25 nt downstream of RPF 5' ends) and resulted in the creation of a 5' and a 3' mRNA fragment. The 5' ends of the resulting 3' fragments coincided with the 5' ends of RPFs, as evidenced by the increased density at position 0 (Fig. 2f). Our hypothesis predicted that the cut around position 0 would result in self-perpetuating mRNA cuts as incoming ribosomes (purple in Fig. 2f) encountered a hard stall at the cleavage site, thus triggering upstream cuts on the 5' fragment. The fragments generated by this iterative endonucleolysis were expected not to contain a cap and therefore not to be captured by Akron3. This hypothesis was 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 were expected to result in accumulation of RPFs upstream of the initial cut sites captured by Akron5. This scenario was reflected in our analysis as increased density in the area immediately to the right of position 0. This result occurred because RPFs

were used as reference points, and their 5' ends were positioned at 0 on the x axis (positive positions on the x axis indicate Akron5 ends located downstream of RPFs). Interestingly, we observed that the highest density in this region was at position +15, corresponding to the distance of the 5' end of an incoming ribosome that is stalled at the initial cut site and the cut site itself, which is now located at the peptidyl (P) site, as previously described⁴². Finally, we hypothesized that the depletion of Akron5 and Akron3 ends at the stall site resulted from protection of that region, which was likely to be attributable to the nature of the stall mechanism itself. We 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, thus 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 instead is a widespread and ubiquitous mechanism that degrades canonical mRNAs. We have termed this process 'ribothrypsis' (from $\theta\rho\upsilon\pi\tau\alpha$, to fragment). Our model predicts that the ribothrypsis

nuclease, which we refer to as ribothrypsin, should act endonucleolytically; hence, ribothrypsin cannot be XRN1, the exosome or the SMG6 endonuclease, which can mediate NMD. Additionally, we predict that if the process is ubiquitous, then ribothrypsin products with 5' and 3' ends defined by this process should be observable in RNA-seq when no RNA fragmentation is performed, such as in small RNA (sRNA)-seq. Furthermore, our model implies the generation of final ribothrypsin products (FRTPs) ~16 nt in length, corresponding to the distance from the first cut, which now resides at the P site of the incoming ribosome and stalls the ribosome (purple in Fig. 2f), and the subsequent cut upstream of the stalled incoming ribosome.

Ribothrypsin is not mediated by XRN1. Xrn1 has been shown to closely follow the last translating ribosome and to define the 5' ends of mRNAs, captured by 5Pseq, during 5'-to-3' cotranslational degradation in yeast¹³. However, according to our model, XRN1 cannot be ribothrypsin, although it could still polish the 5' ends of RNAs after ribothrypsin cuts. To examine the role of XRN1, we analyzed data from Akron5 as well as PARE-seq (which is similar to Akron5), which were obtained from HeLa cells after XRN1 knockdown (KD)³⁴. If XRN1 mediates the process, we would expect abrogation of the 3-nt periodicity and of the signal around the RPF 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, as confirmed by DFT, upstream but not downstream of the 5' end of the terminating ribosome in Akron5 and PARE XRN1-KD (Fig. 3a,b), thus verifying that cotranslational ribothrypsin cuts are not mediated by XRN1. In a plot of the distribution of 5' ends around RPFs, the peak in PARE XRN1-KD was displaced ~2 nt upstream of position 0 (Fig. 3c) relative to Akron5 data (Fig. 2f), thus supporting a role of XRN1 in polishing the 5' ends after ribothrypsin.

We wondered whether ribothrypsin might also operate in *S. cerevisiae* and consequently reanalyzed 5Pseq data from wild-type yeast (grown in complete medium (YPD)) and from an Xrn1-deletion (*Xrn1Δ*) strain¹³. We calculated the distribution of 5Pseq 5' ends around the stop codon and around the RPF 5' ends, which we obtained from wild-type yeast grown in YPD⁴² (Supplementary Fig. 4a). As previously reported¹³, we found 3-nt periodicity upstream of the 5' end of the terminating ribosome in wild-type yeast, as confirmed by DFT, and a peak at position -18 (as measured from the last nucleotide of the stop codon, which was set at 0), a result indicative of cotranslational degradation by Xrn1 that stopped at the terminating ribosome (Supplementary Fig. 4b). As previously reported¹³, the -18 peak disappeared in *Xrn1Δ*, 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 the 5' ends of mRNAs of wild-type (Supplementary Fig. 4d) and *Xrn1Δ* yeast (Supplementary Fig. 4e). These findings indicated that ribothrypsin probably operates in *S. cerevisiae*, and Xrn1 polishes the 5' ends of transcripts after ribothrypsin cuts. Unlike 5Pseq in yeast, Akron5 did not show enrichment at position -17 from the stop codon (Fig. 3a), thus suggesting that Xrn1 may play a larger role in RNA decay in yeast than in humans.

Ribothrypsin footprints in sRNA-seq libraries. As hypothesized on the basis of our model, if ribothrypsin is prevalent, then its footprints should be observable in sRNA-seq libraries, which capture endogenous 5' and 3' ends through adapter ligation and, unlike classic RNA-seq libraries, are not subjected to fragmentation, thus preserving the endogenously produced mRNA ends. We mined deeply sequenced HeLa sRNA libraries (average read length of ~60 nt)⁴³ and examined the distributions of 5' and 3' ends of sRNAs mapping to mRNAs, relative to 5' ends of RPFs. We found a striking

3-nt periodicity indicative of cotranslational generation of these ends, as well as similar density plots to those of Akron5 and Akron3 (Fig. 3d,e). As revealed by Akron5, polyadenylated RNAs are subject to ribothrypsin (Fig. 2f,g), thus eliminating the exosome as a candidate for ribothrypsin, because the exosome attacks mRNAs after deadenylation. 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 and RRP6)⁴³ (Fig. 3f,g and Supplementary Fig. 5a,b). These findings indicated that sRNAs may be used as proxies to investigate cotranslational mRNA decay and supported our model's prediction that the exosome is not ribothrypsin. However, they do not preclude functions of the exosome and the Ski complex in clearing ribothrypsin products, even those bound to ribosomes⁴⁴; or in initiating ribothrypsin during 3'-to-5' exonucleolysis, as the exosome starts degrading CDS. In principle, exosome-mediated trimming⁴² or cleavage by other RNases (such as Ago proteins programmed with microRNAs or small interfering RNAs^{45,46}) may be an important trigger for ribothrypsin.

Because 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 RPF 5' ends compared with the 5' ends of PARE-seq reads from HeLa cells after double KD of 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 ribothrypsin products (FRTPs; Fig. 2f) in sRNA libraries, we interrogated RNAs smaller than 17 nt (ref. 43) and examined the distributions 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,

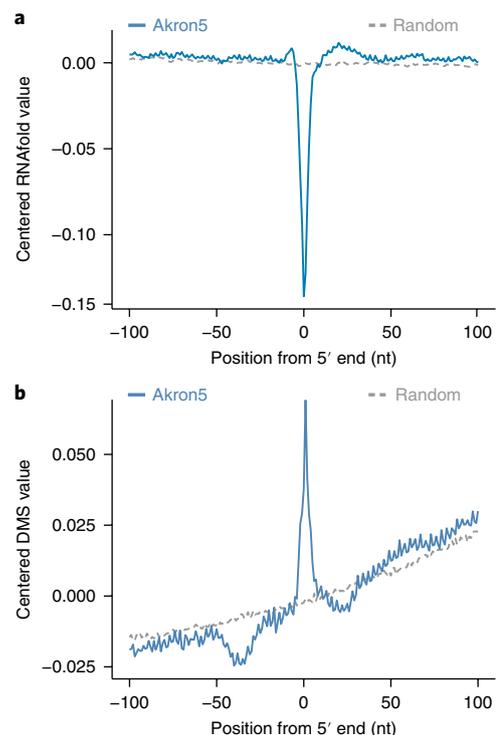


Fig. 4 | Ribothrypsin cuts occur in mRNA areas with open structure.

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.

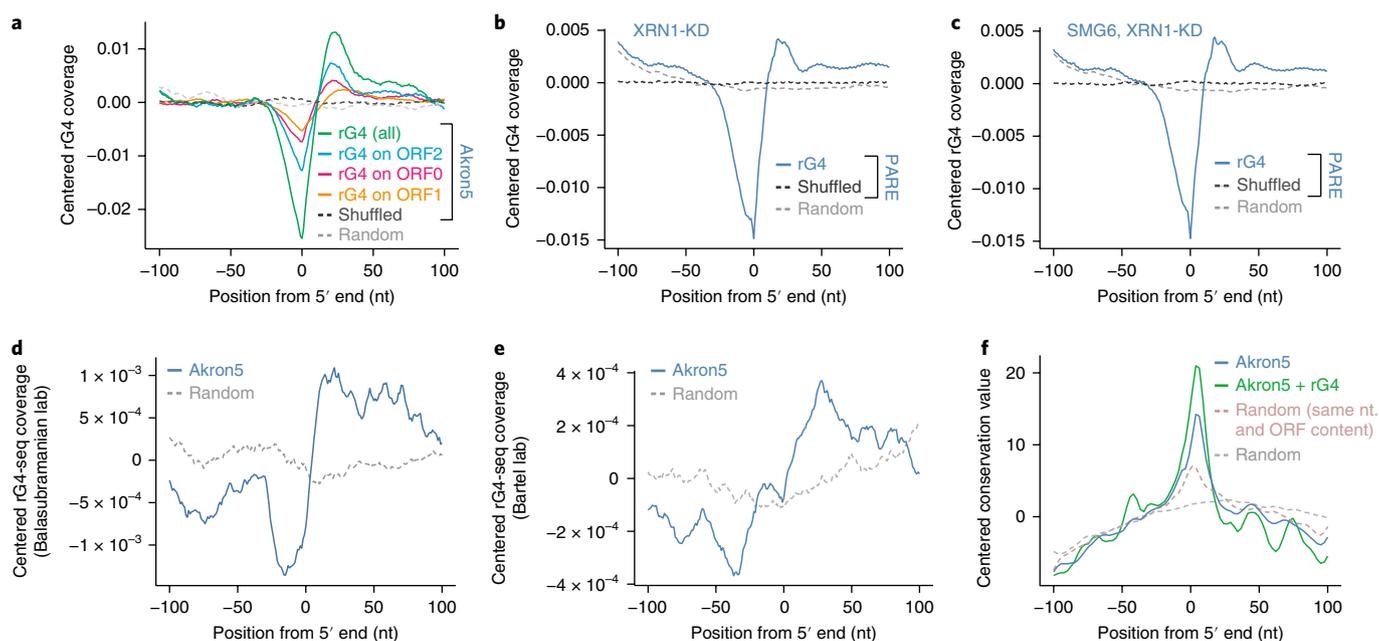


Fig. 5 | Ribothrypsis associates with RNA G quadruplexes and conserved RNA areas. **a**, Centered coverage of predicted rG4s located in the indicated positions relative to ORFs or cumulative (all), upstream and downstream of Akron5 ends that map in CDS, versus random and shuffled controls. **b,c**, Centered coverage of predicted rG4s upstream and downstream of PARE-seq 5' ends that map in CDS, versus random and shuffled controls, in the indicated knockdowns (KD). **d,e**, Centered coverage of rG4-seq from Balasubramanian and colleagues⁵³ (**d**) or Bartel and colleagues⁵⁴ (**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 maintaining the nucleotide and ORF context of Akron5 ends (dashed pink) and a completely random control (dashed gray) are shown.

whereas the 3' ends of sRNAs peaked immediately upstream of Akron5 ends, in agreement with the designation of these ~16-nt sRNAs as FRTPs (Supplementary Fig. 5c). Collectively, all findings presented above supported the ribothrypsis model.

Features of ribothrypsis. We initially asked whether ribothrypsis might be related to the transcript abundance (calculated from RNA-seq from rRNA-depleted total RNA); lengths of CDS and untranslated regions (UTRs); average density of ribosomes per transcript (calculated from ribo-seq and RNA-seq); or RNA stability (calculated from 5'-bromouridine immunoprecipitation chase (BRIC)-seq⁴⁷), but we did not identify any correlations between Akron ends and these features (data not shown).

Next, we examined the distribution of codons around Akron5 and Akron3 ends. For controls, we randomized end positions within the same transcripts, maintaining the expression levels, nucleotide content and open reading frame (ORF) at the cut site. We observed an enrichment in the lysine K(AAG) codon downstream of Akron5 ends (Supplementary Fig. 6) and an enrichment in K(AAG; AAA) at Akron3 ends (Supplementary Fig. 7), as compared with controls, thus 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 cleavage²⁴. We note, however, that the ribothrypsis ends mapped by Akron5 and Akron3 were within the CDS of canonical mRNAs, not in poly(A) sequences (generated from normal or premature polyadenylation) or in consecutive stretches of lysines (data not shown), thus suggesting a role for the codons themselves in ribothrypsis. Furthermore, we found enrichment in aspartate D(GAC) and glutamate E(GAA) codons upstream of Akron3 ends (Supplementary Fig. 7). Remarkably, aspartate and glutamate codons, particularly the same E(GAA) codon that we found to associate with Akron3 ends, have been reported to associate with ribosome pausing in

mouse embryonic stem cells⁴⁸, probably because of a more open RNA structure at these codons⁴⁹. Our findings support the roles of specific codons in ribothrypsis, although mechanisms are lacking and will be an important pursuit for future studies. More generally, the relationships among ribothrypsis, codons and translation will need to be addressed as understanding improves regarding codons' effects on mRNA structure, codon optimality, codon usage and translation efficiency in 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 mapped in CDS and found a marked decrease in 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, as identified by next-generation sequencing⁵⁰ that captures the stalling of commonly used viral-derived reverse transcriptases (RT-stop) as they encounter modified bases⁵¹. We compared Akron5 data against in vivo DMS-seq data from HeLa cells⁵¹ and found a sharp increase in DMS values around position 0, a result indicative of high DMS reactivity and consistent with high RNA accessibility around ribothrypsis cuts (Fig. 4b). Although some of the DMS value may have been due to RT-fall at ribothrypsis cuts, rather than to RT-stop at DMS-modified bases, these results, when combined with structure prediction, indicated 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 and are stabilized by potassium or other cations⁵². rG4s are very stable in vitro^{52,53}, although they are largely

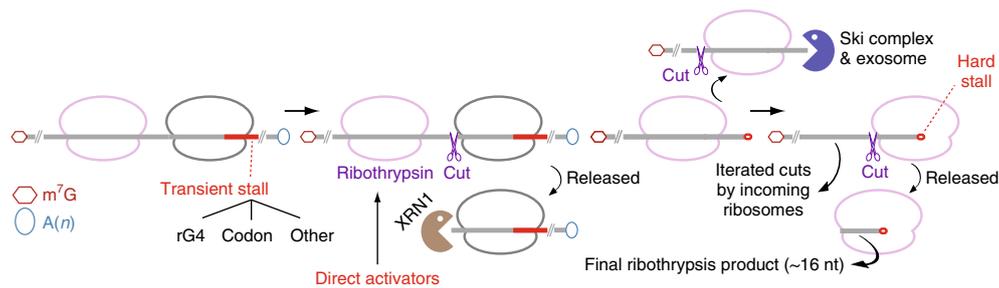


Fig. 6 | Ribothrypsis model. Transient ribosome stalling caused by various mechanisms can activate ribothrypsin to cleave the mRNA as it exits the channel of the ribosome that has encountered a transient stall. Ribothrypsin is propagated as incoming ribosomes encounter the initial ribothrypsin cut (hard stall). Transcript-specific or global direct activators of ribothrypsin may also initiate or propagate ribothrypsis. Intersection of ribothrypsis with other prominent RNA-decay pathways is shown.

unfolded by proteins and other factors *in vivo*⁵⁴. We used an algorithm that we had previously used and validated for rG4 detection in Piwi-interacting-RNA precursors⁵⁵, to predict genome-wide rG4 locations and compare them to ribothrypsin 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 in our model (Fig. 2f). Within CDS, rG4s are known to repress protein output by impeding ribosome progression and to exhibit the largest suppression when located at position +2 relative to the ORF (ORF2)⁵⁶. Remarkably, our data also showed that rG4s in ORF2 corresponded 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 KD of XRN1 or both XRN1 and SMG6 (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 was not due to increased G content, we shuffled the sequences around Akron5 ends, maintaining expression and nucleotide 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 areas around ribothrypsin cuts, particularly regions with an rG4 downstream, were preferentially conserved in vertebrates (Fig. 5f). rG4 enrichment downstream of putative ribothrypsin cleavages was also conserved in yeast, in both wild-type and *Xrn1Δ* strains (Supplementary Fig. 8). Together our findings support a role of conserved mRNA areas and increased rG4s, which probably cause transient ribosome stalling, among ribothrypsin 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 uncovered a previously undescribed pathway of canonical mRNA decay, which we term ribothrypsis (Fig. 6) and revealed its important features: first, that the decay of canonical mRNAs is also endonucleolytic; second, that this decay is mediated by translating ribosomes; third, that mRNA-decay fragments persist in cells; and fourth, that, like any process that affects all mRNAs (such as mRNA splicing or mRNA translation), ribothrypsin is almost certainly regulated. Why do cells use ribothrypsin 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.

Recently, two papers have described the contribution and features of NGD in the decay of Ire-1-cleaved mRNAs during activation of the unfolded-protein response in fission yeast³⁷ and the role of ribosome collision as a critical trigger of NGD in budding yeast⁵⁸. In these cases, the endonuclease ‘NGDase’ associates with the ribosome and cleaves the mRNA as it exits the ribosome^{57,58}. Our results are consistent with the findings from 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 to mRNA surveillance and clearance of aberrant or select mRNAs. Even though the identity of the nuclease remains unknown, molecular mechanisms that mediate mRNA cleavage in ribothrypsin and NGD are likely to be connected. However, we expect that there will also be differences in how the two processes are triggered and regulated. For example, mRNA truncation or base damage by oxidation⁵⁹ elicits NGD on defective mRNAs. However, we expect that numerous factors will trigger ribothrypsin on canonical nondamaged mRNAs, for physiological regulation and not for quality control. We identified rG4s as likely triggers of ribothrypsin, but we believe that this finding may represent the tip of the iceberg, in that many other factors, transcript specific or global, are likely to activate and regulate ribothrypsin.

Characterization of the chemical groups that mark the ends of the cleaved products may provide clues to assist in the identification of ribothrypsin. From our findings and those from the Steinmetz laboratory¹³, 5'-P-bearing mRNA fragments are abundant in HeLa cells and budding yeast, thus suggesting that ribothrypsin generates 3' fragments bearing 5'-P. In this case, ribothrypsin products are immediately accessible to XRN1, which degrades RNAs bearing 5'-P, and ribothrypsin is likely to be a protein whose cleaved products bear 5'-P (3' fragment) and 3'-OH (5' fragment). However, we note that the Hesselberth laboratory has been able to selectively capture and sequence the small amounts of RNAs bearing 5'-OH termini from budding yeast, by using RtcB RNA ligase, and has found that such fragments originate from CDS regions of numerous mRNAs⁶⁰. Intriguingly, 5'-OH fragments appear to accumulate upstream of codons for acidic and basic amino acids⁶⁰. These 5'-OH termini might be products of ribothrypsin (although further work is required to test this possibility); in that case, ribothrypsin would probably 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 that rRNA is essential for the peptidyltransferase activity of the ribosome, rRNA may conceivably also play a key role in ribothrypsin. 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 should provide critical

biological insights into mRNA translation and decay. Furthermore, we believe that our work may set the stage for many future investigations, including identification of factors and pathways that trigger or regulate ribothrypsis in normal cellular functions and in disease; and mechanistic investigation of the intersection of ribothrypsis with mRNA translation and turnover. The new methods for capturing RNA ends and the computational approaches, scripts and analyses described here may also be useful for analyzing and simulating RNA processes in general.

Finally, our findings have implications for studies investigating gene expression through RNA-seq, microarrays, RT-PCR and related approaches, which are among the most commonly used methods of profiling gene expression. An implicit, almost axiomatic, assumption made by users of these methods is that mRNAs exist largely as full-length molecules in vivo. On the basis of this assumption, fragmentation of the isolated RNA is used to prepare libraries suitable for sequencing 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 used to identify cell types; or performing other analyses limited only by researchers' imagination. 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 the small amplified fragment is assumed to derive 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 abundant 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

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41594-018-0042-8>.

Received: 4 September 2017; Accepted: 30 January 2018;
Published online: 5 March 2018

References

- Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195–205 (2002).
- Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.* **84**, 325–354 (2015).
- Green, R. & Noller, H. F. Ribosomes and translation. *Annu. Rev. Biochem.* **66**, 679–716 (1997).
- Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* **13**, 246–259 (2012).
- Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
- Lykke-Andersen, S., Tomecki, R., Jensen, T. H. & Dziembowski, A. The eukaryotic RNA exosome: same scaffold but variable catalytic subunits. *RNA Biol.* **8**, 61–66 (2011).
- Coller, J. & Parker, R. Eukaryotic mRNA decapping. *Annu. Rev. Biochem.* **73**, 861–890 (2004).
- Grudzien-Nogalska, E. & Kiledjian, M. New insights into decapping enzymes and selective mRNA decay. *Wiley Interdiscip. Rev. RNA* **8**, e1379 (2017).
- Jens, M. & Rajewsky, N. Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* **16**, 113–126 (2015).
- Vlachos, I. S. & Hatzigeorgiou, A. G. Functional analysis of miRNAs using the DIANA tools online suite. *Methods Mol. Biol.* **1517**, 25–50 (2017).
- Richter, J. D. & Coller, J. Pausing on polyribosomes: make way for elongation in translational control. *Cell* **163**, 292–300 (2015).
- Hu, W., Sweet, T. J., Chamnongpol, S., Baker, K. E. & Coller, J. Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* **461**, 225–229 (2009).
- Pelechano, V., Wei, W. & Steinmetz, L. M. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* **161**, 1400–1412 (2015).
- Yu, X., Willmann, M. R., Anderson, S. J. & Gregory, B. D. 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* **28**, 2385–2397 (2016).
- Isken, O. & Maquat, L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* **21**, 1833–1856 (2007).
- Shoemaker, C. J. & Green, R. Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* **19**, 594–601 (2012).
- Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nat. Struct. Mol. Biol.* **23**, 7–15 (2016).
- Inada, T. The ribosome as a platform for mRNA and nascent polypeptide quality control. *Trends Biochem. Sci.* **42**, 5–15 (2017).
- Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* **36**, 1073–1078 (2004).
- Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561–564 (2006).
- Frischmeyer, P. A. et al. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).
- 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* **46**, 518–529 (2012).
- 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.* **6**, 28234 (2016).
- Guydosh, N. R. & Green, R. Translation of poly(A) tails leads to precise mRNA cleavage. *RNA* **23**, 749–761 (2017).
- Shoemaker, C. J., Elyer, D. E. & Green, R. Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science* **330**, 369–372 (2010).
- Shoemaker, C. J. & Green, R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl. Acad. Sci. USA* **108**, E1392–E1398 (2011).
- Pisareva, V. P., Skabkin, M. A., Hellen, C. U., Pestova, T. V. & Pisarev, A. V. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J.* **30**, 1804–1817 (2011).
- Becker, T. et al. Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nat. Struct. Mol. Biol.* **18**, 715–720 (2011).
- Bengtson, M. H. & Joazeiro, C. A. Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature* **467**, 470–473 (2010).
- Brandman, O. et al. A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151**, 1042–1054 (2012).
- Defenouillère, Q. et al. Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proc. Natl. Acad. Sci. USA* **110**, 5046–5051 (2013).
- Verma, R., Oania, R. S., Kolawa, N. J. & Deshaies, R. J. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. *eLife* **2**, e00308 (2013).
- German, M. A. et al. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26**, 941–946 (2008).
- Schmidt, S. A. 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.* **43**, 309–323 (2015).
- Addo-Quaye, C., Eshoo, T. W., Bartel, D. P. & Axtell, M. J. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* **18**, 758–762 (2008).
- Gregory, B. D. et al. A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev. Cell* **14**, 854–866 (2008).
- Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* **53**, 1044–1052 (2014).
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
- Park, J. E., Yi, H., Kim, Y., Chang, H. & Kim, V. N. Regulation of poly(A) tail and translation during the somatic cell cycle. *Mol. Cell* **62**, 462–471 (2016).
- Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A. & Johnson, R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22**, 867–882 (2016).
- Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y. & Robb, G. B. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* **40**, e54 (2012).
- Guydosh, N. R. & Green, R. Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **156**, 950–962 (2014).
- Lubas, M. et al. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep.* **10**, 178–192 (2015).
- Schmidt, C. et al. The cryo-EM structure of a ribosome-Ski2-Ski3-Ski8 helicase complex. *Science* **354**, 1431–1433 (2016).

45. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
46. Orban, T. I. & Izaurralde, E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**, 459–469 (2005).
47. Tani, H. et al. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22**, 947–956 (2012).
48. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
49. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
50. Silverman, I. M., Berkowitz, N. D., Gosai, S. J. & Gregory, B. D. Genome-wide approaches for RNA structure probing. *Adv. Exp. Med. Biol.* **907**, 29–59 (2016).
51. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).
52. Song, J., Perreault, J. P., Topisirovic, I. & Richard, S. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation (Austin)* **4**, e1244031 (2016).
53. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13**, 841–844 (2016).
54. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371 (2016).
55. Vourekas, A. et al. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev.* **29**, 617–629 (2015).
56. Endoh, T. & Sugimoto, N. Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. *Sci. Rep.* **6**, 22719 (2016).
57. Guydosh, N. R., 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* **6**, e29216 (2017).
58. Simms, C. L., Yan, L. L. & Zaher, H. S. Ribosome collision is critical for quality control during no-go decay. *Mol. Cell* **68**, 361–373 e5 (2017).
59. Simms, C. L., Hudson, B. H., Mosior, J. W., Rangwala, A. S. & Zaher, H. S. An active role for the ribosome in determining the fate of oxidized mRNA. *Cell Rep.* **9**, 1256–1264 (2014).
60. Peach, S. E., York, K. & Hesselberth, J. R. Global analysis of RNA cleavage by 5'-hydroxyl RNA sequencing. *Nucleic Acids Res.* **43**, e108 (2015).

Acknowledgements

We thank the members of the laboratory of Z.M. for discussions; X. Ji and S. Liebhaber for assistance with ISCO polysome fractionation; and J. Schug (University of Pennsylvania) and D. Pouchnik (Washington State University) for technical help with Illumina and PacBio sequencing, respectively. We thank the University of Pennsylvania Diabetes Research Center (DRC) for the use of the Functional Genomics Core (P30-DK19525) for Illumina sequencing, and the Washington State University Molecular Biology and Genomics Core for the use of Pacific Biosciences sequencing. This work was supported by a Brody family fellowship to M.M. and by grants from the ALS Therapy Alliance and the NIH (GM072777) to Z.M.

Author contributions

F.I., M.M. and Z.M. conceived the study. F.I. designed, performed and interpreted all wet-lab experiments and developed Akron-SMRT and Akron-seq. M.M. designed, performed and interpreted all computational analyses and ribothrypsis modeling, to which P.A. substantially contributed. All authors analyzed the data. F.I., M.M. and Z.M. wrote the manuscript, and P.A. provided input and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-018-0042-8>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Z.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Cell culture. HeLa cells obtained from ATCC (ATCC0 CCL-2.1) were maintained in DMEM (Gibco) supplemented with 10% FBS (Sigma), at 37°C in 5% CO₂. Cells were harvested at ~80% confluence. The cells were free of mycoplasma.

Polysome analysis. Cells were pretreated with 100 µg/ml cycloheximide (CHX, Sigma) at 37°C for 10 min, washed twice with cold 1× PBS plus 100 µg/ml CHX, and collected in lysis buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl and 2.5 mM MgCl₂) supplemented with 100 µg/ml CHX, complete EDTA-free protease inhibitor (Roche) and 500 U/ml RNasin RNase inhibitor (RNasin, Promega). Cells were passed through a 25-G 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 r.p.m. for 2 h (Beckman SW41-Ti rotor) at 4°C. Polysome profiles were 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 with 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 noncapped, 5'-phosphate (P)-bearing RNAs by incubation at 30°C for 1 h according to the manufacturer's instructions. RNA was ethanol precipitated and resolved on 6% denaturing PAGE and was transferred to a Hybond N⁺ nylon membrane (GE Healthcare) in 0.5× TBE in a semidry transfer system (Amersham Biosciences). RNA was UV-cross-linked (Stratagene) to the membrane and prehybridized in ULTRHyb-oligo buffer (Ambion). The probe was internally labeled with [α -³²P]UTP with a T7 MEGAShortscript kit (Ambion) and purified with Sephadex G-50 columns (GE Healthcare). Hybridization was carried out in ULTRHyb-oligo buffer. Membranes were washed twice with 2× SSC buffer plus 1% SDS, and twice with 0.5× SSC buffer plus 0.1% SDS, and signals were detected through storage-phosphor autoradiography (Amersham). For quantification, the entire region covering the full length and smear was selected in a rectangular box in ImageJ (v 1.46r), and the signal from the entire region was considered to be 100%. A rectangular box with the same width but covering only the area corresponding to the full-length signal was then selected. To obtain an estimate of the full-length percentage, the signal of the full length was divided by 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 and 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 prewashed protein A Dynabeads (Invitrogen) that were conjugated to an antibody against RPL7A (A300-749A, Bethyl Laboratories) or to nonimmune serum as a negative control. Samples were nutated at 4°C for 4 h, then centrifuged at 3,000 r.p.m. for 10 min to collect beads. Beads were washed five times with IP binding buffer. An aliquot was used for western blotting (WB), and the rest was incubated with IP elution buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM EDTA and 1% SDS) at 95°C for 3 min. RNA was recovered through TRIzol extraction and processed for northern blotting. For WB, beads were heated at 95°C for 10 min; proteins were resolved on 4–12% NuPAGE gels (Invitrogen), transferred to nitrocellulose membranes and probed with anti-RPL7A at 1:5,000 as described previously⁶¹.

Preparation of Akron-SMRT sequencing libraries. Total RNA from HeLa cells and polysome fractions (light, 6–9; heavy, 10–14) was extracted with TRIzol. RNA was treated with Terminator to remove noncapped, 5'-P-bearing RNAs. RNA was precipitated with 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 with AMPure-XP beads (Beckman) according to the manufacturer's protocol.

For library construction, RNA was ligated to a 3'-biotinylated RNA adapter (REL3) with T4 RNA ligase (Fermentas) plus 500 U/ml RNasin and incubated at 16°C for 16 h. Ligated RNA was treated with Terminator to remove excess adapter. RNA was recovered with AMPure-XP beads and was captured with M-280 streptavidin Dynabeads (Invitrogen). Beads were washed twice with 2× bind/wash (BW) buffer (10 mM Tris, pH 7.5, 1 mM EDTA and 2 M NaCl) and twice with 1× BW buffer (5 mM Tris, pH 7.5, 0.5 mM EDTA and 1 M NaCl), then resuspended in RNase-free water. cDNA was generated under conditions favoring full-length amplification. RNA was reverse transcribed with 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 and heat inactivation 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 with KAPA HiFiHotStartReadyMix (KAPA Biosystems) and 10 µM primers that contained SMRT barcodes and corresponded to the 5' UTR (forward, CDK4 F0 or TUBB F0; reverse, REL3-RT). Reactions were run at 98°C for 2 min; 20 cycles of 98°C for 30 s, 57°C for 30 s

and 72°C for 2 min; and 72°C for 5 min. Re-PCR was set up with 10 µM primers that corresponded to the start codon (forward, CDK4 F1 or TUBB F1; reverse, REL3-1 or REL3-2) and was run at 98°C for 2 min; six cycles of 98°C for 30 s, 57°C for 30 s and 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 for each gene and purified with a GeneJet kit (Thermo Scientific). SMRT-bell sequencing libraries were prepared by Pacific Biosciences through a 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, a polysome pool was spiked in with *Renilla* luciferase mRNA before total RNA extraction and Terminator digestion. Luciferase mRNA was generated in vitro with a T7 MegaScriptTranscription kit (Ambion) in the presence of a synthetic RNA-cap analog (NEB) and [α -³²P]UTP. RNA was resolved on 6% denaturing PAGE and visualized by autoradiography (GE Healthcare).

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 before library preparation with an RNA Bioanalyzer or agarose gel. After the initial treatment, all protocols converged into the same RNA-ligation and downstream library-preparation steps with minor modifications.

Preparation of Akron3 libraries. Initially, we used rRNA depletion before proceeding to Akron3 library preparation, but we noticed that our sequences contained substantial amounts of capped noncoding sRNAs, including uridine-rich, small nuclear RNAs (U snRNAs). To remove these abundant capped RNAs, sRNAs <200 bp were first removed from total RNA with SPRIselect beads (Beckman) according to the manufacturer's protocol and then used as 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 rRNAs⁶². A cocktail of the sn/rRNA oligos containing biotinylated 3' ends was prepared on the basis of the abundance of each targeted sn/rRNA (that we calculated according to our initial sequencing runs). Total RNA (~5–6 µg per reaction) was incubated with the cocktail mix in 1× SSC buffer in a thermal cycler (at 75°C for 5 min, 0.1°C/s to 37°C and 37°C for 45 min). Samples were incubated with MyOne streptavidin C1 Dynabeads (Invitrogen) that were resuspended in 2× BW buffer through mixing at 1,000 r.p.m. for 15 min at 37°C. Depleted total RNA was precipitated with 100% ethanol, 3 M NaOAc, pH 5.5, and 2 µg glycogen (Ambion). To remove noncapped, 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) with T4 RNA ligase in a mixture containing 5% polyethylene glycol (PEG, NEB) and 500 U/ml RNasin at 16°C for 16 h. Ligated RNAs were fragmented in 1× 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, and 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 through T4 PNK treatment at 37°C for 1 h, and beads were washed with 1× BW buffer.

RNA was then ligated to a 5' RNA adapter (A3-RL5) with T4 RNA ligase plus 15% PEG and 500 U/ml RNasin at 16°C for 16 h. After beads were washed, RNA was reverse transcribed with Superscript III primed with 2.5 µM DP3-RT primer (at 50°C for 45 min, 55°C for 15 min and heat inactivation at 90°C for 3 min). Samples were amplified through PCR with AccuPrimePfxSuperMix (Invitrogen) and 10 µM primers (DP3 and DP5), run at 94°C for 2 min; 16 cycles of 94°C for 20 s, 58°C for 30 s and 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 with 5 µM primers that were compatible with the Illumina sequencer and were run at 94°C for 2 min; six to eight cycles of 94°C for 20 s, 60°C for 30 s and 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 against that of capped RNAs, total RNA that had been depleted of rRNAs and capped sRNAs was Terminator treated to remove 5'-P-bearing RNAs, then divided equally into two aliquots. To detect 5'-OH RNAs, the first aliquot was phosphorylated with T4 PNK in the presence of [γ -³²P]ATP and incubated at 37°C for 1 h. 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, then subjected to Terminator treatment to remove them. RNA was extracted and ethanol precipitated, then treated with tobacco acid pyrophosphatase (TAP, Ambion) at 37°C for 1 h according to the manufacturer's protocol to remove the 5' cap, thus leaving a residual 5'-P. RNA was extracted, 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 -³²P]ATP. RNA was extracted and ethanol

precipitated. Radiolabeled 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, owing to their negligible amounts.

Preparation of Akron5 libraries. To capture 5' ends of polyadenylated, noncapped RNAs, total RNA (~100–150 µg) was incubated with oligo(dT)₂₅ Dynabeads (Ambion) according to the manufacturer's protocol to enrich for polyadenylated RNAs. Beads were then incubated with T4 PNK at 37°C for 1 h to phosphorylate RNA 5' ends. Beads were washed, and 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 h. Capped RNAs were carried over, but only noncapped RNAs ligated to the 5'-adapter. As described for Akron3, beads were washed, and RNA was fragmented in 1× fragmentation buffer. Ligated RNA was captured with M280 streptavidin Dynabeads and treated with T4 PNK to repair RNA 3' ends. After beads were washed, 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 incubated at 50°C for 45 min, then 55°C for 15 min, and heat inactivated at 90°C for 3 min. PCR and re-PCR reactions were set up as described for 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 the Illumina Nextseq500 platform (75PE).

Primer and adapter sequences. Primer and adapter sequences are shown in Supplementary Table 1.

Oligonucleotide sequences. Oligo sequences are shown in Supplementary Table 2.

Akron-SMRT data preprocessing. To avoid potential sequence fragments and to ensure that we captured the entire insert RNA, we used only CCS reads for Akron-SMRT analysis. Poly(A) sequences were trimmed from CCS reads with an in-house script identifying 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), with the following parameters: `--outFilterMultimapScoreRange 0 --outFilterScoreMinOverRead 0 --outFilterMatchNminOverRead 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 with cutadapt software in paired-end mode. The adapter GTGTCAGTCACTTCCAGCGG was trimmed from the 3' end of read 1, whereas CCGCATCGTCCCTCCCT was trimmed from the 3' end of read 2. A constant adapter sequence of 16 nt was removed from the 5' end of read 1. 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 with the STAR aligner (v 2.5.2a). First, reads were aligned in paired-end mode. To exclude potential PCR artifacts, reads with the same 5' end (defined from read 1) and the same 3' end (defined from read 2) were collapsed. The retained reads were then re-aligned in single-end mode, using read 1 only and the following parameters: `--outFilterMultimapScoreRange 0 --alignIntronMax 50000 --outFilterMatchNmin 8 --outFilterMatchNminOverRead 0.7 --sjdbOverhang 80 --alignSJDBoverhangMin 1 --seedSearchStartLmax 15`. Aligned reads were loaded into a SQLite3 database for further processing with CLIPseqTools⁴³ and were annotated with information on whether they overlapped with elements from RepeatMasker (downloaded from UCSC), rRNAs (extracted from the UCSC gene model annotation file) and genes (from the UCSC gene model annotation file).

Akron3 data preprocessing. The adapter sequences were trimmed from the paired-end reads through the same process as for Akron5. The reads were aligned to the human genome in a two-step approach similar to that of Akron5. Unlike Akron5, in the second step, only read 2 sequences 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 --outFilterScoreMinOverRead 0 --outFilterMatchNmin 30 --outFilterMatchNminOverRead 0 --sjdbOverhang 100 --alignSJDBoverhangMin 1`. Similarly to the Akron5 procedure, aligned reads were loaded into a SQLite3 database for further processing and annotation.

Genomic distribution. All mapped reads were assigned to genomic categories on the basis of 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 only uniquely mapped reads.

Correlation of replicates. Read abundance was measured in RPKM for protein-coding transcripts. Pearson's correlation was calculated in 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 entire mRNA, 5' UTR, CDS and 3' UTR (called elements), and for all protein-coding transcripts. Only genic elements and mRNAs 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 make it 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 nontemplated poly(A) tails for Akron3, specific parameters were used at the alignment step (described in Akron3 data preprocessing). Poly(A) tails were reported by the aligner as soft-clipped sequences. These soft-clipped sequences were extracted from the sam/bam alignment file with an in-house script, and if the A content was higher than 80%, the read was annotated as having a poly(A) tail. Akron-SMRT poly(A) identification is described in the data preprocessing section.

Relative position to RPFs. The distribution of collapsed ends (ends on the same position were 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 decrease 5'-end uncertainty for RPFs, only RPFs of specific sizes were used. The size was selected to ensure the highest enrichment in any of the ORFs. The selected RPF sizes were 29 nt for humans and 28 nt for yeast. The analysis was performed for ends that mapped within the CDS, unless otherwise noted in the text, with an offset of 100 nt from the start and stop codon to avoid distribution confounding from potential specific mechanisms acting close to the start and stop codons. All density plots were normalized so that the area under the plot was 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 to 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 was calculated. Three distinct nucleotide-composition models for region x were calculated, corresponding to each of the three ORFs. For each end, a model was selected on the basis of the ORF to which the end mapped. Using this model, we generated one sequence of length equal to x and identified an exact match within the same gene transcript. If no match was found, the process was repeated until a match was found (maximum of 1,000 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 twenty-fifth quantile of protein-coding transcripts for each library was excluded from the analysis. The analysis was performed independently for each library, and replicates were summed only for visualization.

Discrete Fourier transform. DFT converts a signal to a representation in the frequency domain. DFT was calculated in R with the GeneCycle package. To facilitate interpretation, we plotted the power at each period instead of at each frequency.

G-quadruplex prediction. The G-quadruplex prediction was performed with an in-house script by using the pattern $G\{2,4\},\{1,7\}G\{2,4\},\{1,7\}G\{2,4\},\{1,7\}G\{2,4\}$, as described in ref. ³⁵, 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 controls, a random position within the same transcript was selected for each read end. For shuffled controls, the sequence around the read ends was shuffled at the codon level. The latter maintains the nucleotide and codon content and can be used to evaluate whether rG4 structures are preferentially selected for, beyond what would be expected for any increase in G content alone.

Ribothrypsis-model simulation. A detailed description of the ribothrypsis-model simulation can be found in Supplementary Fig. 3.

External sequencing data. The following datasets were downloaded and used in this study:

Homo sapiens, PARE-seq³⁴, ribo-seq³⁹ (GSM2100602), total RNA-seq³⁹ (GSM2100594), TAIL-seq³⁷, sRNA-seq⁴⁵, DMS-seq⁴¹, rG4-seq^{33,54} and BRIC-seq⁴⁷. *S. cerevisiae*, 5Pseq¹³ and ribo-seq⁴².

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. Source code required for analysis has been deposited on GitHub at <https://github.com/mnsmar/ribothrypsis/>.

Data availability. Datasets generated during and/or analyzed during the current study are available in the Gene Expression Omnibus repository under accession number [GSE107838](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107838). All other data are available from the authors upon reasonable request.

References

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.* **3**, 301–308 (2013).
62. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66–71 (2014).
63. Maragkakis, M., Alexiou, P., Nakaya, T. & Mourelatos, Z. CLIPSeqTools: a novel bioinformatics CLIP-seq analysis suite. *RNA* **22**, 1–9 (2016).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

▶ Experimental design

1. Sample size

Describe how sample size was determined.

No sample size calculation was performed

2. Data exclusions

Describe any data exclusions.

No data exclusions

3. Replication

Describe whether the experimental findings were reliably reproduced.

There were three biological replicates for Akron5 and 2 biological replicates for Akron3

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Randomization was not relevant; we did not perform experimental manipulations that require randomization

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Randomization was not relevant; we did not perform experimental manipulations that require blinding

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Already published software has been described in the manuscript. Custom software tools have been described in the manuscript and deposited to GitHub.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Rabbit polyclonal antibody against RPL7A was purchased from Bethyl Laboratories, catalog #: A300-749A. Antibody was validated by Western blot and immunoprecipitation of human cells (HeLa)

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

HeLa cell line directly obtained from ATCC (ATCCO CCL-2.1)

b. Describe the method of cell line authentication used.

We did not authenticate the cell line since it was obtained from ATCC

c. Report whether the cell lines were tested for mycoplasma contamination.

We confirm that cell lines used were free of mycoplasma

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

Not applicable

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Not applicable

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Not applicable