High-throughput 5'P sequencing reveals environmental regulated ribosome stalls at termination level

Yujie Zhang[1] , Vicent Pelechano[1,*]

[1] SciLifeLab, Department of Microbiology, Tumor and Cell Biology, Karolinska Instituet, Solna, 171 65, Sweden

* Correspondence may be addressed to vicente.pelechano.garcia@ki.se

ABSTRACT

RNA degradation is critical for gene expression and mRNA quality control. mRNA degradation is intimately connected to the translation process up to the degree that 5'-3' co-translational mRNA degradation can be used as a proxy for ribosome dynamics. Here we present an improved high-throughput 5'P RNA sequencing method (HT-5Pseq). HT-5Pseq is easy, scalable and uses affordable duplex-specific nuclease based rRNA depletion. We investigate *in vivo* ribosome stalls in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* focusing on translation termination. We show that degradation intermediates of mRNAs with high termination pausing often use TGA as stop codon preceded by Lys and have relatively shorter poly(A) tails. We investigate the regulation of this process after glucose deprivation and reveal that, in addition to translation initiation, glucose starvation also leads to ribosome stall at the stop codon. Finally, we investigate gene-specific termination pausing across multiple conditions and show that it is a highly regulated process that crosstalks with codon optimality and the length of poly(A) remaining after deadenylation. In summary, HT-5Pseq is an improved and cost-effective approach that allows to investigate the crosstalk between RNA degradation and the translation process and that is particularly useful to examine its coupling at the level of translation termination.

Key points:

- HT-5Pseq facilitates high throughput investigation of 5'P mRNA degradation intermediates.
- Gene-specific ribosome pause at termination level is common and environmentally regulated.
- Ribosome stall at termination level are associated with TGA usage and shorter deadenylated poly(A) tails.

INTRODUCTION

RNA degradation is an integral player of gene expression, as its balance with RNA synthesis is what determines final mRNA abundance. Thus, to understand gene expression, it is important to explore not only mRNA synthesis but also the factors that shape the mRNA degradation. In addition to RNA binding proteins or RNA secondary structure, the translation process is also key to shape the mRNA degradome (1, 2). By investigating 5'phosphorylated (5'P) mRNA degradation intermediates, we have previously demonstrated that in vivo 5'-3' co-translational degradation provides information regarding ribosome dynamics and codon specific ribosome stalls. Specifically, in budding yeast the 5'-3' exonuclease Xrn1p follows the last translating ribosome generating an in vivo footprint of its position (1). A similar phenomenon has also been described in other organism such as *Arabidopsis thaliana* or rice (*Oryza sativa*) (3, 4). However, despite our ability to investigate the 5'P mRNA degradome, current sequencing approaches are still cumbersome. Methods originally developed to investigate miRNA cleavage such as PARE (parallel analysis of RNA ends) (4–6) and GMUCT (genome-wide mapping of uncapped and cleaved transcripts) (3) are labour intensive requiring multiple ligation steps, purification and, in some cases, labour-intensive gel size-selection. Even our original 5PSeq protocol requires multiple enzymatic steps and bead-based purifications that limits its application to investigate hundreds of samples (7).

In addition to investigate the presence of mRNA degradation intermediates, it is important to understand also their origin. Cytoplasmic mRNA decay in eukaryotes initiates through deadenylation (8, 9). After poly(A) tail shortening by deadenylase complexes PAN2/3 and CCR4-NOT, mRNA can be degraded exonucleolytically either in 5'-3' or 3´-5' orientation. In some cases, 5'-3' mRNA degradation by XRN1, is independent on deadenylation process (e.g. during NMD) or through the endoribonucleolytic decay (10). This has been confirmed in *Arabidopsis* showing that XRN4 substrates (XRN1 in yeast) can be both polyadenylated and non-polyadenylated. The existing diversity regarding 5'P mRNA degradation intermediates poly(A) length suggests that multiple subpopulations of degradation intermediates coexist. We have previously shown that both oligo-dT and random hexamer priming can be used to identify mRNA degradation intermediates. However, oligo-dT priming leads in general to a decrease in library complexity and introduces a mRNA length-dependent sequencing bias (1, 7). Additionally, up until now, we

lacked information regarding the environmental regulation of deadenylation activity of mRNA degradation intermediates and its potential relationship with the translation process and ribosome stalls. This lack of information is even more important when considering ribosome stalls at termination level, that have been generally less studied.

To facilitate the investigation of the 5'P mRNA degradome in relation with the translation process, we have developed a high throughput 5'P sequencing approach (HT-5Pseq). This approach is flexible, cheap and scalable. We investigate its applicability in *S. cerevisiae* and *S. pombe*. We investigate the 5'P degradation profiles associated to ribosome pausing, its regulation in stress conditions and the relative poly(A) length of mRNA degradation intermediates. Finally, we investigate the role of transient glucose deprivation on the regulation of ribosome stall at termination level.

MATERIAL AND METHODS

Growth Conditions and Sample preparations

Saccharomyces cerevisiae strain BY4741 (MAT a *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was grown to mid exponential phase (OD_{600}~0.8) at 30 °C using YPD (1% yeast extract, 2% peptone, 2% glucose). *Schizosaccharomyces pombe* (h-) was grown at 30ºC to mid-log phase (OD600~0.8) using YES media (0.5% yeast extract, 3% glucose, supplemented with 225 mg/l of adenine, histidine, leucine, uracil and lysine). For cycloheximide (CHX) treatment, CHX was added to final 0.1 mg/mL to the medium and incubated for 5 min at 30°C for *S. cerevisiae* and 10 min for *S. pombe*. For glucose deprivation, mid exponential *S. cerevisiae* cultures (OD_{600}~0.8) were spun down, washed twice and resuspended in pre-warmed YP (lacking glucose) and then grown at 30°C. For glucose re-addition, cultures with already 15 min glucose deprivation were washed twice and resuspended in YPD for 15 min at 30°C. For early exponential phase, strains were grown at 30°C from an initial OD_{600}~0.05 to a final OD_{600} of 0.3. To reach stationary stage, *S. cerevisiae* strains were grown during 60 h in YPD. All yeast samples were collected by centrifugation and pellets were frozen in liquid nitrogen. Total RNA was isolated by the standard phenol:chloroform method, and DNA was removed by DNase I treatment. RNA integrity was checked by agarose gel.

HT-5Pseq library preparation and sequencing

For HT-5Pseq libraries construction, we used 6 µg of DNA-free total RNA. Samples were directly subjected to RNA ligation. The treated RNA samples were incubated with 100 µM RNA rP5_RND oligo (final 10 µM, Table S3) 2 h at 25°C with 10 Units of T4 RNA ligase 1 (NEB). Please note that we used an RNA oligo, and not the DNA/RNA oligonucleotide previously used (7). Ligated RNA was purified with RNA Clean XP (Beckman Coulter), according to the manufacturer's instructions. RNA was reverse transcribed with Superscript II (Life Technologies) and primed with Illumina PE2 compatible oligos containing random hexamers (20 µM, Table S3) and oligo-dT (0.05 µM, Table S3). Reverse transcription reaction was incubated for 10 min at 25°C, 50 min at 42°C and heat inactivated for 15 min at 70°C. To deplete RNA in RNA/cDNA hybrid after reverse transcription, we used sodium hydroxide (40 mM) for incubation 20 min at 65°C and then neutralized with Tris-HCl, pH =7.0 (40 mM). For DSN (Duplex-specific nuclease) based rRNA depletion, we used a mixture of probes (Table S1) targeting the 18S rDNA, 25S rDNA and 5.8S rDNA. The probes were designed to occupy the whole ribosomal RNA regions with consecutive 25-30 nt long unmodified DNA oligos. The hybridization of probes (2 µM each) with cDNAs were incubated at 68 °C for 2 minutes before adding pre-warmed DSN buffer mix with 1 Units of DSN enzyme (Evrogen). The reaction then performed at 68 °C for 20 minutes. To inactive DSN enzyme, we added 2X DSN stop solution and incubate 10 min at 68 °C. The final PCR amplification was performed using 2X Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB) and final 0.1 µM of PE1.0 and corresponding multiplex PE2.0_MTX (Table S3). The program followed this: 30s 98°C; 15 cycles (20s 98°C; 30s 65°C; 30s 72°C); 7min 72°C. Libraries were size selected using 0.7x-0.9x (v/v) AMpure XP beads (Beckman Coulter) to final length of 200-500 bp and sequenced by NextSeq 500 using 60 sequencing cycles for Read 1 and 15 cycles for Read 2.

Read preprocessing and analysis

3'-sequencing adaptor trimming was applied to 5'ends of reads using cutadapt V1.16 (http://gensoft.pasteur.fr/docs/cutadapt/1.6/index.html) . The 8-nt random barcodes on the 5' ends of reads were extracted and added to the header of fastq file as the UMI using UMI-tools. Reads were mapped to the reference genome (SGD R64-1-1 for *S. cerevisiae* genome, ASM294v2.20 for *S. pombe* genome) separately by star/2.7.0 (11) with the parameter --alignEndsType Extend5pOfRead1 to exclude soft-clipped bases on the 5' end. To calculate the fraction of rRNA, tRNA, snRNA snoRNA and mRNA in library compositions, the stepwise alignment was performed by corresponding index generated by star/2.7.0. Duplicated 5' ends of read introduced by PCR during library preparation were removed based on random barcodes sequences using UMI-tools. To compare the differences of 5'P read coverage in DSN based rRNA depletion at gene level, reads per gene were

counted using Subread package (featureCounts) (12). mRNA, tRNA, rRNA and snRNA and snoRNA transcripts were counted separately and combined for further analysis. Differential gene expression analysis were performed using the DESeq2 packages from R and Bioconductor (http://www.bioconductor.org/) (13). The threshold for differentially expressed genes were defined as p value < 0.005 and log2(fold-change) > 1. Analysis of 5' ends positions was performed using *Fivepseq* package (14) (http://pelechanolab.com/software/fivepseq), including relative to start, stop codon and codon specific pausing (Supplementary Data S1). Specifically, the unique 5'mRNA reads in biological samples were summed up and normalized to reads per million (rpm). Then the relative position of 5'mRNA reads to all codons of all ORF were summed at each position. Metagene plots were showed as the sum value versus the relative distance from respective codon. All clustering analyses was performed by k-means using Complexheatmap packages from R and Bioconductor (15). Gene Ontology enrichment analysis was performed with ClusterProfiler using Fisher's exact test (16). Significance for enrichment for particular nucleotides and amino acids at each position were performed using kpLogo (17). Dataset for *S. cerevisiae* tRNA adaptation index, mRNA codon stability index, translation efficiency were obtained from Carneiro *et al.* (18), gene expression level, NMD targeted genes, 3´UTR length and mRNA half-life were obtained from Xu *et al.* (19), Celik *et al.* (20), Pelechano *et al.* (21) and Presnyak *et al.* (22), respectively.

RESULTS

Development of HT-5Pseq for investigation of 5′-phosphorylated mRNA degradation intermediates

To enable the high-throughput investigation of co-translation mRNA degradation and ribosome dynamics, we set to improve our previous 5PSeq approach (7). We identified multiple bottlenecks for its systematic application: the high number of enzymatic reactions, the required DNA purification steps and the difficulty associated to rRNAs depletion (that

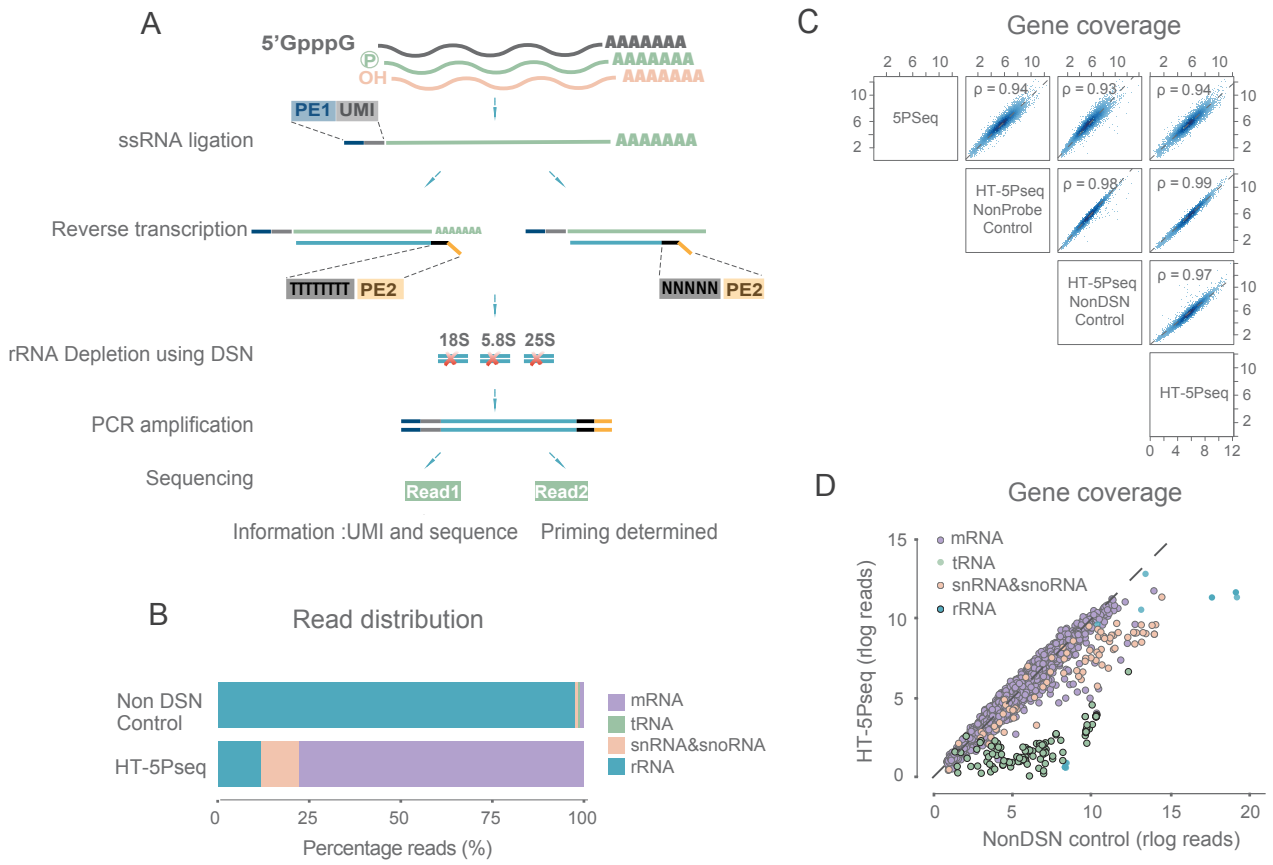

Figure 1. Development of HT-5Pseq. (A) HT-5Pseq method outline. A specific RNA oligo (PE1) is ligated to 5'P RNA molecules. RNA is reverse transcribed using as primer a mix of sequencing oligos (PE2) containing oligo-dT and random hexamers. The cDNA originating from rRNA is depleted using DNA oligos and double-strand specific nuclease (DSN). cDNA is PCR amplified and sequenced. (B) Improvement of mRNA mapability after rRNA depletion, comparing HT-5Pseq and same protocol omitting DSN treatment (%). (C) Spearman correlation between 5PSeq and HT-5Pseq. NonDSN refers to control libraries omitting DSN rRNA depletion. NonProbe refers to libraries treated with DSN but omitting the depletion oligos. 5'P read gene coverage shown in rlog. (D) Differential gene-specific 5'P read coverage.

compose the majority of the 5'P RNAs in the cell). By improving those steps, we decreased both the required hands-on time as well as the cost per library. In HT-5Pseq, we first ligate an RNA oligo containing unique molecular identifiers (UMI) to the 5'P ends of total RNA (Fig 1A). Total ligated RNA is subjected to reverse transcription using a mix of Illumina compatible oligos priming with a random hexamer and oligo-dT. This obviates the necessity of *in vitro* RNA or cDNA fragmentation and improves sequencing coverage around the stop codon region (see below). After the generation of cDNAs containing Illumina compatible adaptors at both sides, we degraded RNA with NaOH. To selectively deplete abundant cDNA complementary to rRNA, we designed a pool of affordable non-modified DNA oligos (Table S1). We annealed the rRNA depletion oligos with the single-stranded cDNA library at 68ºC and treated the mix with duplex-specific nuclease (DSN) (23, 24). After this, the rRNA depleted cDNA library is purified and PCR amplified to generate the final sequencing library. By eliminating the need for costly biotin-based rRNA depletion and the time-consuming mRNA fragmentation, we reduced the number of bead-based purifications (from 11 to 4) and eliminated the enzymatic steps required for end-repair, dA-tailing and DNA ligation. All these modifications reduced the total time from 35 h to 9 h and the library cost preparations 70% (Table S2).

DSN-based rRNA depletion increases the percentage of reads aligned to mRNAs from 1.1% to 77.6% (Fig 1B) and also the associated sequencing library complexity (Fig S1A). Accordingly, DSN treatment also decreases the rRNA reads percentage from 98.0% to 12.0% (Fig 1B). Although the main application of HT-5Pseq is the identification of regions with an increased 3-nt periodicity of 5'P fragments (e.g. as those caused by ribosome stalls) and not their overall abundance, we confirmed that DSN treatment is highly reproducible (Fig 1C and Fig S1C) and has limited effect on mRNA coverage (Fig 1D). DSN treatment decreases sequencing coverage of the depleted rRNA regions (in blue), and other features with significant secondary structure as tRNAs (in green). Only a few mRNAs (0.57%, 31 of the 5398 detected) present significant decreased

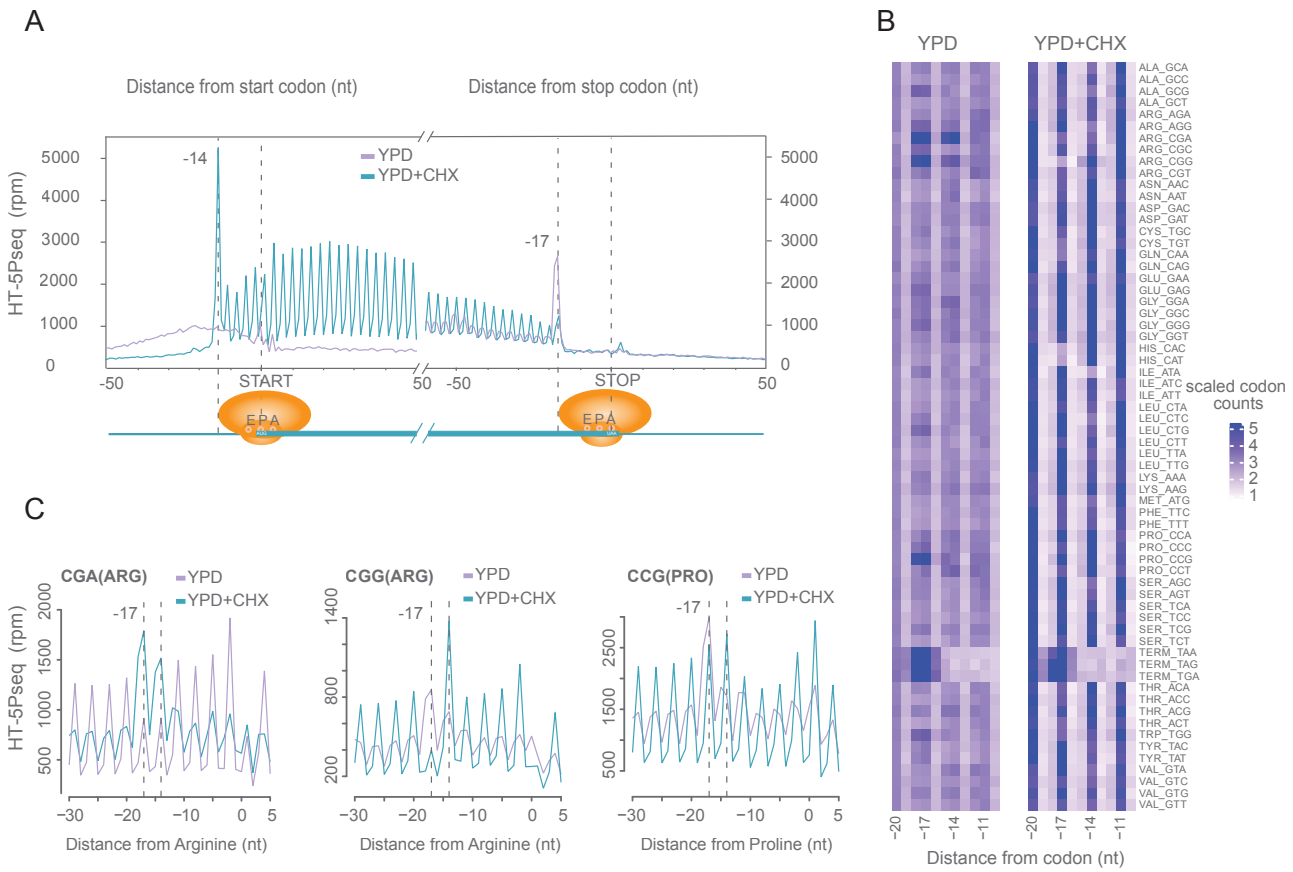

Figure 2. HT-5Pseq reveals ribosome dynamics at codon resolution. (A) Metagene analysis for 5'P read coverage relative to ORF start and stop codon. Cells grown in rich media (YPD) are shown in purple and cycloheximide treated cells (CHX) for 5 mins in blue. (B) Heat map for codon specific 5'P coverage. Positions -17, -14 and -11 nt represent ribosome protection at the A, P and E site respectively. For each codon, reads were normalized using the total 5'P reads between -30 to 5 nt. (C) 5'P reads coverage for rare arginine (CGA, CGG) and proline codons (CCG). Dotted lines at -17 and -14 corresponding to the expected 5' end of protected ribosome located at the A site or P site, respectively.

coverage, likely caused by their ability to form hairpins at 68°C. An advantage of HT-5Pseq is the simple and flexible design of rRNA depletion probes. Contrary to ribosome profiling, in HT-5Pseq samples are not subject to *in vitro* RNA degradation generating small rRNA fragments (25). In 5'P mRNA degradome sequencing, mature rRNA (and tRNAs) with well-defined 5'P boundaries are the main contaminants. To show the flexibility of our rRNA depletion approach, we used probes originally designed for *S. cerevisiae* (Table S1) to investigate the 5'P degradome in the distant yeast *S. pombe*. We confirmed that this treatment increased the percentage of usable mRNA molecules, from 5.0% unique mRNA degradation intermediates to 62.3% after DSN treatment (Fig S1C), even when using a suboptimal rRNA depletion panel. All this confirms that HT-5Pseq is a flexible protocol that can be easily adapted to the organism of interest.

HT-5Pseq provides single nucleotide resolution insights into 5'-3' co-translational decay

Once confirmed its general applicability, we verified HT-5Pseq ability to identify ribosome stalls at single-nucleotide resolution (1, 7). As expected, HT-5Pseq recovers the characteristic 3-nt periodicity associated to 5'-3' co-translational mRNA degradation (Fig 2A). To further confirm its sensitivity, we investigated codon specific ribosome stalling using our newly developed computational pipeline *Fivepseq* (14). We clearly identified pauses associated to slow codons like arginine (encoded by CGG, CGA) and proline (encoded by CCG) (Fig 2C) as we and other have previously shown (26, 27). To demonstrate the ability of HT-5Pseq to detect direct perturbations of the translation process, we investigated mRNA degradome changes after cycloheximide (CHX) treatment. As expected, CHX clearly increases the observed 3-nt periodicity, especially at the 5'regions of the genes (Fig 2A). In addition to its general effect inhibiting translation elongation, CHX can also mask *in vivo* ribosome pausing at codon level (28). As in ribosome profiling, in HT-5Pseq rare codons present higher 5'P protection indicating ribosome stalling, which was lost after CHX treatment (e.g. correlation between A-site 5'P footprint and tRNA adaptation index (tAI) goes from Pearson correlation r -0.385 to 0.061, Fig S2C). This is likely due to CHX allowing one round of translation elongation and thus displacing the protection pattern one codon downstream (29). Accordingly, we observe that the protection pattern for proline (CCG) and arginine (CGG) is displaced one codon after CHX treatment (Fig 2C). We also observed an increase protection for codons specifically pausing at A, P or E sites (Fig 2B). To investigate *in vivo* ribosome dynamics in a different species, we treated *S. pombe* with CHX and observed a clear increase of the 3-nt periodicity (Fig S2A), as expected by the general translation inhibition (Fig S2B). As in *S. cerevisiae*, rare codons such as CCG (proline) and CGG (arginine) showed higher enrichment without CHX treatment, but lost enrichment at A site in CHX samples (Fig S2D). We also found displaced CGG (arginine) codon from A site to P site in *S. pombe.* Our results thus confirm that HT-5Pseq offers codon-specific information of the co-translational mRNA decay process.

5'P degradome sequencing reveals clear ribosome pauses at the stop codon

An additional advantage of HT-5Pseq is that, by using a combination of random hexamers and oligo-dT primers during reverse transcription, it allows control of the sequencing depth around the stop codon (Fig 3A). In ribosome profiling, coverage around the stop codon can be modest, which can be explained by a combination of biological and technical reasons. For example, footprints associated to translation termination can be lost during the *in vitro* RNA digestion required in ribosome profiling (30, 31), especially when using high salt conditions (27) (Fig S3B), or because the associated footprints possess different length and thus can be excluded from the analysis. To increase the coverage around the stop codon, we focused on those HT-5Pseq reads primed by oligo-dT (Fig 3A). Oligo-dT primed HT-5Pseq reads allowed the observation of a clear protection pattern associated to single (-17nt) and two ribosomes (-47/-50nt, disomes), stalled at the stop codon. This could be observed both at metagene (Fig 3A) and single gene level (Fig 3C). This results also confirm that a fraction of mRNA degradation intermediates contain sizable poly(A) tails. Even if we used 10-dT to prime the libraries, 70% of the oligo-dT primed mRNA degradation intermediates had poly(A) tails of 15 or more As (Fig 3B). As deadenylation is tightly coupled with mRNA degradation, and degradation with the translation process, we decided to investigate up to what degree mRNA degradation, termination pausing, poly(A) length of the intermediate of degradation and other mRNA features were related.

First, we investigated genes according to their level of termination pause, defined as the relative 5'P protection, 17 nt upstream of the stop codon. Genes with high termination pauses have an increased usage of TGA as stop codon (29%, compared with the lower termination pausing of 19%) (Fig 3E, Fig S3D). While genes with lower termination pausing present an increased usage of TAA (62%, compared with the lower termination pausing of 48%). The termination paused was also influenced by the last amino acid preceding the stop codon (Fig 3E, Fig S3E). In particular, Lys was significantly enriched in genes with high termination pause, while Leu and Ala were enriched in genes with decreased pausing.

These amino acid-specific differences suggest that interaction with the ribosome release factors might play a role in the degree of termination pausing. We then examined the length of the remaining poly(A) tails present in the intermediates of degradation. In general, mRNAs with high termination pauses presented shorter remaining poly(A) tails (Fig 3D). Taking the data together, we hypothesized that high

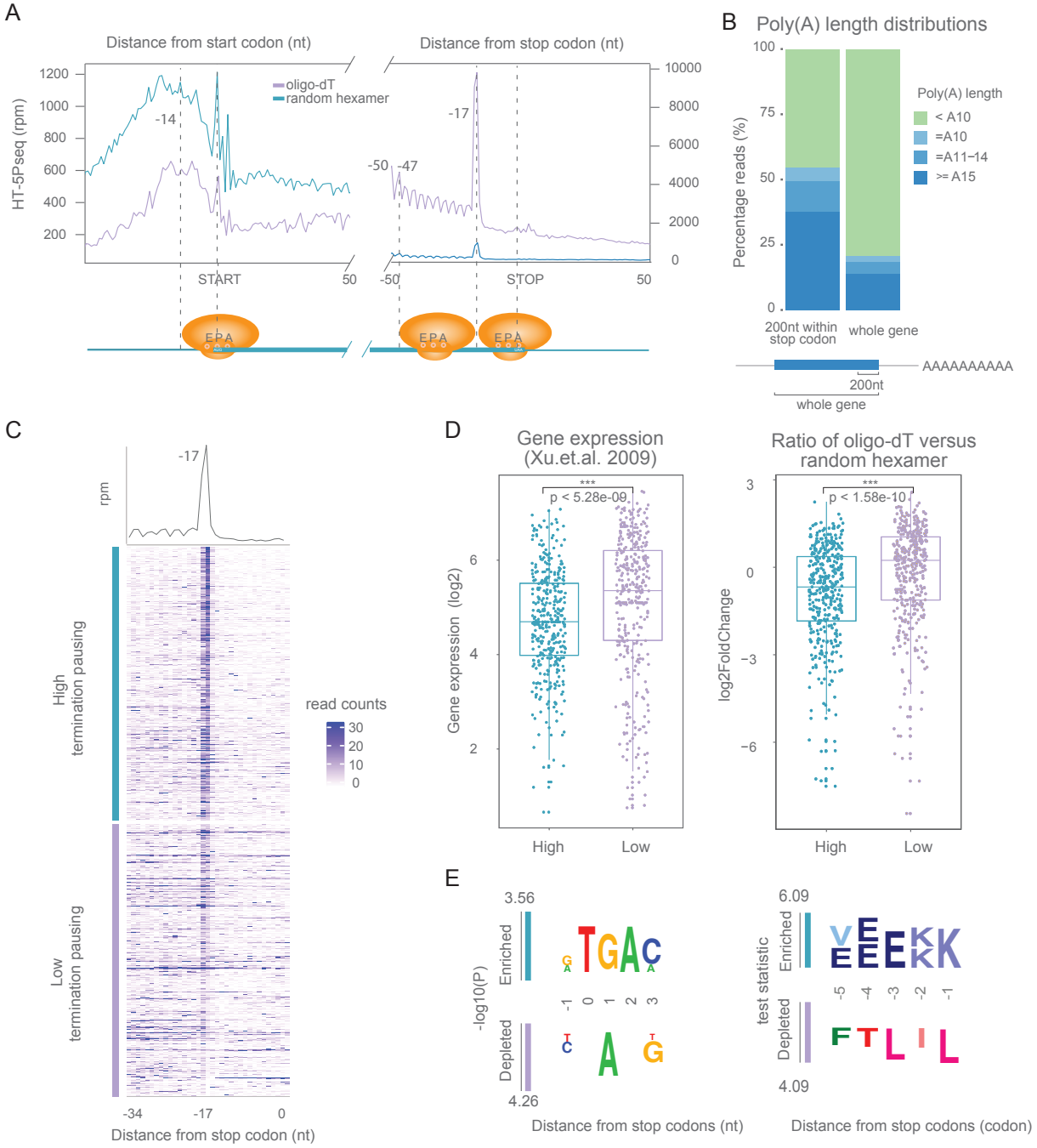

Figure 3. High and low ribosome termination pausing genes are associated with differential poly(A) length for 5'P intermediates of degradation. (A) Metagene analysis for 5'P read coverage relative to ORF start and stop codon for HT-5Pseq reads primed by oligo-dT(purple) or Random hexamer (blue). (B) inferred poly(A) length for mRNA intermediates of degradation for 5'P in the whole gene or the last 200 nt or the ORFs. (C) Genes were sorted by the relative ribosome protection at the stop codon (-17) in respect to the displayed region (-34 to 0 nt). Only genes with at least 50 reads in this region were considered. The top 50% genes were defined as the high termination pausing genes and the bottom half were defined as low termination pausing genes. (D) Boxplot for gene expression level (left)(19) and the ratio of oligo-dT versus random hexamer primed 5'P coverage in the last 200 nt. High termination pause are show in blue and low termination pause in purple. P-values were calculated by the Wilcoxon Rank Test (two-tailed test). (E) Significance for enrichment for particular nucleotides (left) and amino acids (right) at each position relative to stop codon by comparing high and low termination pausing groups using kpLogo (17).

expressed genes would present an overall decrease of termination pausing. In agreement, highly expressed genes associated with GOs related to translation, presented low termination pauses (e.g. cytoplasmic translation (GO:0002181) p-adj < 10e-09) (Table S4). On the contrary, those associated with the transcription process had in general increased termination pausing (e.g. DNA-templated transcription (GO:0006351) p-adj < 10e-04). Finally, we investigated if the observed differential termination pausing could reflect the intrinsic sensitivity of mRNAs to non-sense mediated decay (NMD) (20). However, our analysis showed that NMD sensitivity in exponential growth couldn't explain the observed differences (Fig S3C). Once described ribosome stalling at termination level in standard growth condition, we decided to take advantage of HT-5Pseq to investigate the environmental regulation of this phenomena.

Environmental regulation of ribosome stalling during termination

We have previously showed that during stationary phase, budding yeast displays massive increase in ribosome stalls at termination level (1). We hypothesized that this could be caused by nutrient exhaustion during stationary phase. To understand better this phenomenon, we used a more controlled experimental setting, glucose deprivation. Glucose deprivation is a well-characterized environmental challenge that leads to fast inhibition of translation acting mainly at the level of initiation in *S. cerevisiae* (32). However, much less is known regarding its potential role at translation termination. We investigated the 5'P degradome after 5 and 15 min of glucose deprivation and their recovery after 15 min of re-addition of glucose (Fig 4A). We observed a clear stop pause after 5 minutes of glucose deprivation, both at position -17 and at -47/-50 (corresponding to disomes). This pause is decreased after 15 min of glucose deprivation, likely as a downstream consequence of the inhibition of translation initiation. The normal termination pause observed in exponential growing cells is recovered after 15 min of glucose reintroduction. Although the observed pauses are an overall feature, some groups of genes respond differently. We classified genes according to their relatively high or low stop pausing at position -17nt in the different conditions. Genes classified with relatively low termination pausing are quite stable during the studied conditions while genes with high termination pausing are more variable (Fig 4B). The level of termination pause is gene-specific and differentially regulated across the studied conditions. For example, genes involved in sporulation display high termination pausing after glucose deprivation (5 and 15 min) (p-adj < 5e-02), while genes involved in macromolecule modification (GO:0043412) had high stop pausing after 15 min of glucose re-addition (p-adj < 10e-03) (Table S5). On the contrary, cytoplasmic translation genes were consistently observed with low termination pausing across all conditions (p-adj < 10e-05), while genes involved in oxidoreduction coenzyme metabolic process (GO: 0006733, p-adj < 5e-02) had low stop pausing after 15 min glucose re-addition. All these shows, that ribosome stalls at termination level are highly regulated during environmental change. In addition, we also observed environmental regulation of the mRNA degradation intermediates poly(A) length. Glucose deprivation lead to a drastic shortening of length of the partially deadenylated poly(A). This shortening is gene specific and could not be recovered after 15 min glucose re-addition (Fig S4A-B).

To investigate up to what degree the gene specific termination pauses associated to transient glucose deprivation are related to the ones observed during stationary phase, we used HT-5Pseq to investigate cells in stationary phase and exponentially growing cells at a lower cell density (OD ~ 0.3). We clustered genes according to their relative stalling at the 12 last codons as measured by HT-5Pseq. This includes the termination pauses associated to monosomes and disomes (from -50 to -17 nt). We observed clear differential regulation of termination pausing across conditions (Fig 4C). For example, genes involved in cellular respiration and oxidation-reduction process are specially paused in stationary condition (cluster 4, GO:0045333 (p-adj < 5.6e-06), GO:0009060 (p-adj < 5.6e-06)) (Table S6). While genes involved in transcription and cell cycle are paused both in low OD growth, stationary, as well as glucose deprivation conditions (cluster 1, GO:0006355 (p-adj < 3.9e-05), GO:0007049 (p-adj < 1.9e-04)) (Table S6). We observed previously that higher usage of TAA stop codon was associated to decrease termination pauses (Fig 3E). Consistently we observed that clusters 2, 5 and 6 with less stalling, have a higher usage of TAA stop codon (Fig 4C and Fig S4D). As ribosome stalling often leads to mRNA decay (33), we investigated factors involved with mRNA turnover. We found that those genes with less termination pausing (clusters 2, 5 and 6) have in general high codon stability index, codon optimality, translation efficiency and protein abundance (Fig S4E-I). Interestingly, this trend is clearer in those genes using TAA as stop codon (Fig S4E-I). Having classified genes according to their differential termination pauses, we investigated also the relative poly(A) length of the intermediates of degradation (Fig S4C). The previously described poly(A) shortening during glucose deprivation is particularly extreme in clusters 1, 4 and 5. In some cases, stationary phase and glucose deprivation present similar poly(A) shortening

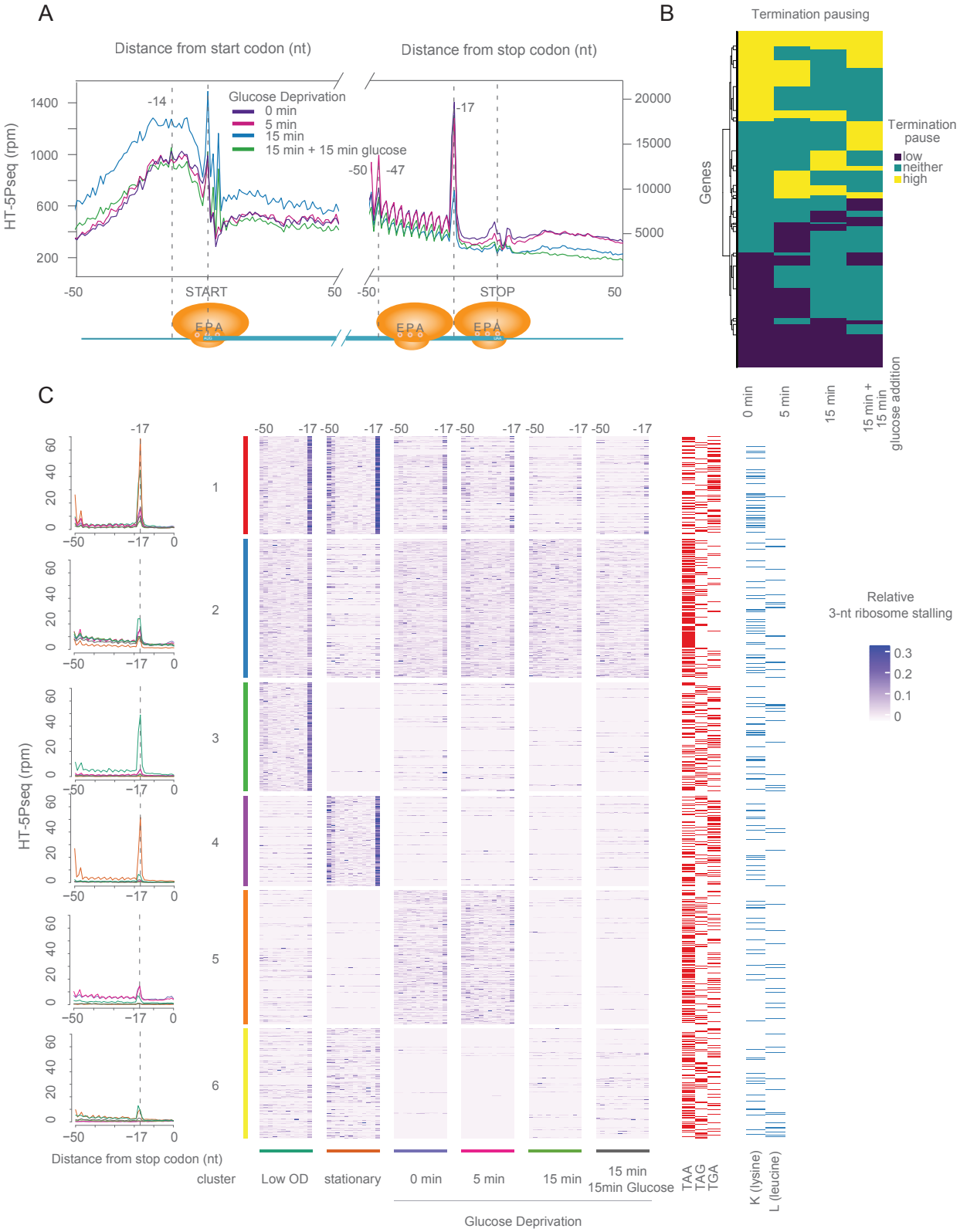

Figure 4. Ribosome termination pausing is regulated during glucose starvation. (A) Metagene analysis for 5'P read coverage relative to ORF start and stop codon. Cells grown at different points after glucose starvation. 0 mins (in dark purple), 5 mins (in pink), 15 mins (in blue) and 15 mins glucose deprivation with another 15 mins with glucose addition (in green). (B) Genes with high termination pausing (in yellow) and low termination pausing (in dark purple) during the analyzed glucose deprivation time course. High and low termination pausing defined as in Fig. 3. (C) Relative 3-nt periodicity for the last 12 codons clustered using K-means. Each point corresponds to the ratio at each codon (distance between peaks and valleys). -17nt indicates a ribosome paused at the stop codon and -50 nt corresponds to the protection by a disome. Line plots on the right side displays cluster specific metagene analysis for 5'P read coverage relative to ORF stop codons (low OD in green, stationary phase in orange, 0, 5 , 15 mins of glucose deprivation and another 15 mins of glucose addition were in purple, pink , green and grey respectively). The stop codon usage is represented in red and amino acids usage before stop codons (lysine and leucine) in blue.

(clusters 3 and 5). However, in other cases degradation intermediates have relatively long poly(A) tails in stationary phase, but short in glucose deprivations (cluster 1 and 4). In summary, our results show that translation termination pausing is often environmentally regulated, and it is coupled with differential level of deadenylation of the mRNA intermediates of degradation. Our results also demonstrate that glucose deprivation, in addition to its well know effect inhibiting translation initiation, leads to transient ribosome stalling at termination level.

DISCUSSION

We have presented here an optimized protocol for the easy measurement of 5′-phosphorylated mRNA degradation intermediates, HT-5Pseq. In this approach, 5′P RNA molecules are first ligated to an Illumina compatible RNA oligonucleotide with UMIs. Ligated RNA is reverse transcribed using a mix of random hexamers and oligo-dT. Contaminant cDNA originating from undesired rRNA molecules is selectively depleted using a pool of affordable DNA oligos and treatment with a duplex-specific nuclease (DSN) (24). This improved protocol decreases hands on time and allows the easy handling of ten to hundred samples, potentially even more if combined with early multiplexing and sample pooling. We have showed that HT-5Pseq is a simple and reproducible approach to detect codon-specific ribosome stalling in both *S. cerevisiae* and *S. pombe*.

We have shown that the use of an oligo-dT primer during reverse transcription increases the coverage at the stop codon. In addition to increase the utility of the approach, this also demonstrates that a fraction of mRNA degradation intermediates still possesses a significant poly(A) tail. Although HT-5Pseq with its current implementation can only measure subtle changes in the poly(A) tail of degradation intermediates (from 10 to 15+ As), we observed gene and environmental specific regulation. We observed that 38 % of the oligo-dT primed HT-5Pseq reads present tails of at least 15A, and longer tails are often associated with lower termination pause. Additionally we found that the proportion of poly(A) tails versus random hexamer primed transcripts could be regulated under glucose deprivation (for example in 15 min (Fig S4A-B)). As mRNA decay in eukaryotes is initiated by poly(A) tail shortening (34) and depends on the kinetics of the dissociation of Pab1 and the exonuclease activity of Ccr4-Not (35), our results suggest environmental regulation of the deadenylation process. However, more detailed molecular studies will be necessary to disentangle its potential crosstalk with ribosome stalling.

Finally, we have investigated the regulation of *in vivo* RNA footprints associated to ribosome stalling at termination level. We showed the genes presenting lower termination pausing use more frequently TAA as stop codons, and are often highly translated genes with high codon optimality. This could be attributed to secondary structure downstream of termination pausing (36) or due to the interaction of ribosomes with positive amino acid Lys before stop codon, which hinders binding of the release factors (37). Interestingly, termination stalling can be enhance by interfering with the translation termination process, for example, in strains depleted for eIF5A or Rli/ABCE (26, 27, 38). Termination pausing can also be regulated by environmental conditions, e.g. glucose deprivation or stationary phase. In particular, after 5 mins of glucose deprivation, the termination pausing is not limited to the peak at the stop codon of monosome, but also of disomes. Similar to the decrease of the termination pausing at 15 min that we observe after glucose deprivation, the queuing ribosome near the stop codon has been also shown to disappear upon oxygen and glucose deprivation (OGD) for 20min (39). The decrease of termination pausing may be explained by either slow translation elongation (limiting the ribosomes reaching termination) or by an increase of the termination/ recycling process. Although HT-5Pseq focuses exclusively on the subset of ribosomes undergoing co-translational degradation, we think it offers complementary information to general approaches like ribosome profiling. For example, using the *in vivo* toeprinting activity of Xrn1, HT-5Pseq can easily capture disomes or trisomes *in vivo*. Some of those can also be captured using specialized ribosome profiling approaches, such as Disome-seq (40, 41). However, depending on the stringency of the used *in vitro* RNase I treatment, ribosome footprints can be altered (42). This can lead to the loss of some disome complexes (e.g. those separated by an extra codon) (41). In summary, we think that HT-5Pseq will be a valuable tool facilitating the investigation of mRNA life from translation to decay. We expect that, by combining multiple genomic and structural approaches, in the future we will be able to understand better this fundamental process in biology.

ACCESSION NUMBERS

The raw and processed sequencing data are deposited at GEO with accession number GSE152375.

SUPPLEMENTARY DATA

ACKNOWLEDGEMENT

We thank all members of the Pelechano, Kutter and Friedländer laboratories for useful discussions. We kindly thank Lilit Nersisyan and Jingwen Wang for

bioinformatic support. Computational analysis was performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

FUNDING

This project was funded by the Swedish Foundation's Starting Grant (Ragnar Söderberg Foundation), a Wallenberg Academy Fellowship [2016.0123], the Swedish Research Council [VR 2016-01842], Karolinska Institutet (SciLifeLab Fellowship, SFO and KI funds) and a Joint China-Sweden mobility grant (STINT, CH2018-7750) to V.P. Y.Z. is funded by a fellowship from the China Scholarship Council.

CONFLICT OF INTEREST

None.

REFERENCES

1. Pelechano,V., Wei,W., Steinmetz,L.M. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell*. 2015; 161: 1400–1412.
2. Hu,W., Sweet,T.J., Chamnongpol,S., Baker,K.E., Coller,J. Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature*. 2009; 461: 225–229.
3. 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*. 2016; 28: 2385–2397.
4. Hou,C.Y., Lee,W.C., Chou,H.C., Chen,A.P., Chou,S.J., Chen,H.M. Global analysis of truncated RNA ends reveals new insights into Ribosome Stalling in plants. *Plant Cell*. 2016; 28: 2398–2416.
5. Schon,M.A., Kellner,M.J., Plotnikova,A., Hofmann,F., Nodine,M.D. NanoPARE: Parallel analysis of RNA 5 ′ ends from low-input RNA. *Genome Res.* 2018; 28: 1919–1930.
6. German,M.A., Luo,S., Schroth,G., Meyers,B.C., Green,P.J. Construction of parallel analysis of rna ends (Pare) libraries for the study of cleaved mirna targets and the rna degradome. *Nat. Protoc.* 2009; 4: 356–362.
7. Pelechano,V., Wei,W., Steinmetz,L.M. Genome-wide quantification of 5′-phosphorylated mRNA degradation intermediates for analysis of ribosome dynamics. *Nat. Protoc.* 2016; 11: 359–376.
8. Brewer,G., Ross,J. Poly(A) shortening and degradation of the 3´A+U-rich sequences of human c-myc mRNA in a cell-free system. *Mol. Cell. Biol.* 1988; 8: 1697–1708.
9. Garneau,N.L., Wilusz,J., Wilusz,C.J. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 2007; 8: 113–126.
10. Nagarajan,V.K., Kukulich,P.M., von Hagel,B., Green,P.J. RNA degradomes reveal substrates and importance for dark and nitrogen stress responses of Arabidopsis XRN4. *Nucleic Acids Res.* 2019; 47: 9216–9230.
11. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M., Gingeras,T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29: 15–21.
12. Liao,Y., Smyth,G.K., Shi,W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30: 923–930.
13. Love,M.I., Huber,W., Anders,S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 1–21.
14. Nersisyan,L., Ropat,M., Pelechano,V. Fivepseq: analysis of 5′ count distribution in RNA-seq datasets for inference on mRNA degradation and translation. *bioRxiv*. 2020; 10.1101/2020.01.22.915421.
15. Gu,Z., Eils,R., Schlesner,M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016; 32: 2847–2849.
16. Yu,G., Wang,L.G., Han,Y., He,Q.Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* 2012; 16: 284–287.
17. Wu,X., Bartel,D.P. KpLogo: Positional k -mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Res.* 2017; 45: W534–W538.
18. Carneiro,R.L., Requião,R.D., Rossetto,S., Domitrovic,T., Palhano,F.L. Codon stabilization coefficient as a metric to gain insights into mRNA stability and codon bias and their relationships with translation. *Nucleic Acids Res.* 2019; 47: 2216–2228.
19. Xu,Z., Wei,W., Gagneur,J., Perocchi,F., Clauder-Münster,S., Camblong,J., Guffanti,E., Stutz,F., Huber,W., Steinmetz,L.M. Bidirectional promoters generate pervasive transcription in yeast. *Nature*. 2009; 457: 1033–1037.
20. Celik,A., Baker,R., He,F., Jacobson,A. High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *RNA*. 2017; 23: 735–748.
21. Pelechano,V., Wei,W., Steinmetz,L.M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*. 2013; 497: 127–131.
22. Presnyak,V., Alhusaini,N., Chen,Y.H., Martin,S., Morris,N., Kline,N., Olson,S., Weinberg,D., Baker,K.E., Graveley,B.R., *et al.* Codon optimality is a major determinant of mRNA stability. *Cell*. 2015; 160: 1111–1124.
23. Zhulidov,P.A. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res.* 2004; 32: e37.
24. Archer,S.K., Shirokikh,N.E., Preiss,T. Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics*. 2014; 15: 1–9.
25. Chung,B.Y., Hardcastle,T.J., Jones,J.D., Irigoyen,N., Firth,A.E., Baulcombe,D.C.,

Brierley,I. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA*. 2015; 21: 1731–1745.
26. Pelechano,V., Alepuz,P. EIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res.* 2017; 45: 7326–7338.
27. Schuller,A.P., Wu,C.C.C., Dever,T.E., Buskirk,A.R., Green,R. eIF5A Functions Globally in Translation Elongation and Termination. *Mol. Cell*. 2017; 66: 194-205.e5.
28. Hussmann,J.A., Patchett,S., Johnson,A., Sawyer,S., Press,W.H. Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. *PLoS Genet.* 2015; 11: 1–25.
29. Schneider-Poetsch,T., Ju,J., Eyler,D.E., Dang,Y., Bhat,S., Merrick,W.C., Green,R., Shen,B., Liu,J.O. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat. Chem. Biol.* 2010; 6: 209–217.
30. Guydosh,N.R., Green,R. Dom34 rescues ribosomes in 3′ untranslated regions. *Cell*. 2014; 156: 950–962.
31. Zhao,T., Chen,Y., Wang,J., Chen,S., Qian,W. Disome-seq reveals sequence-mediated coupling of translational pauses and protein structures. *bioRxiv*. 2019; 10.1101/746875.
32. Ashe,M.P., Long,S.K. De, Sachs,A.B. Initiation in Yeast. *Mol. Biol. Cell*. 2000; 11: 833–848.
33. Doma,M.K., Parker,R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*. 2006; 440: 561–564.
34. Parker,R., Song,H. The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.* 2004; 11: 121–127.
35. Webster,M.W., Chen,Y.H., Stowell,J.A.W., Alhusaini,N., Sweet,T., Graveley,B.R., Coller,J., Passmore,L.A. mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol. Cell*. 2018; 70: 1089-1100.e8.
36. Trotta,E. Selective forces and mutational biases drive stop codon usage in the human genome: A comparison with sense codon usage. *BMC Genomics*. 2016; 17: 1–12.
37. Kasari,V., Pochopien,A.A., Margus,T., Murina,V., Turnbull,K., Zhou,Y., Nissan,T., Graf,M., Nováček,J., Atkinson,G.C., *et al.* A role for the Saccharomyces cerevisiae ABCF protein New1 in translation termination/recycling. *Nucleic Acids Res.* 2019; 47: 8807–8820.
38. Young,D.J., Guydosh,N.R., Zhang,F., Hinnebusch,A.G., Green,R. Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3′UTRs In Vivo. *Cell*. 2015; 162: 872–884.
39. Andreev,D.E., O'Connor,P.B.F., Zhdanov,A. V., Dmitriev,R.I., Shatsky,I.N., Papkovsky,D.B., Baranov,P. V. Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20minutes. *Genome Biol.* 2015; 16: 1–14.
40. Han,P., Shichino,Y., Schneider-Poetsch,T., Mito,M., Hashimoto,S., Udagawa,T., Kohno,K., Yoshida,M., Mishima,Y., Inada,T., *et al.* Genome-wide Survey of Ribosome Collision. *Cell Rep.* 2020; 31: 107610.
41. Zhao Disome-seq reveals widespread ribosome collisions that recruit co-translational chaperones. *bioRxiv*. 2019; .
42. Mohammad,F., Green,R., Buskirk,A.R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife*. 2019; 8: 1–25.