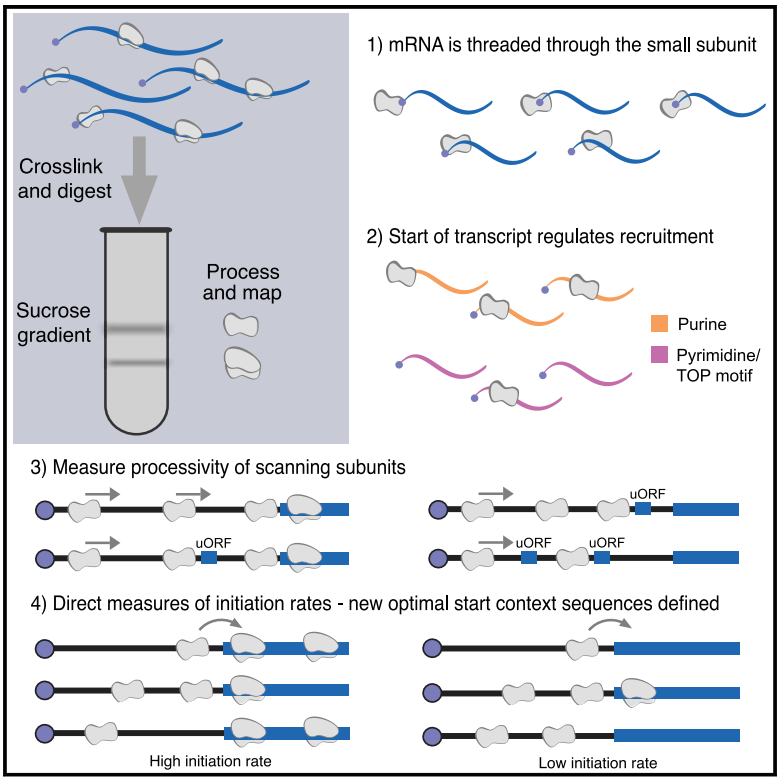
Report

Cell Reports

Profiling of Small Ribosomal Subunits Reveals Modes and Regulation of Translation Initiation

Graphical Abstract



Authors

Adam Giess, Yamila N. Torres Cleuren, Håkon Tjeldnes, ..., Aniekan Okon, Carston R. Wagner, Eivind Valen

Correspondence

yamilatorrescleuren@gmail.com (Y.N.T.C.), eivind.valen@gmail.com (E.V.)

In Brief

Giess et al. introduce RCP-seq, a method to study small ribosomal subunit (SSU) dynamics in zebrafish. It reveals threading as the main mode of recruitment to the mRNA and quantifies the impact of sequence features that affect SSU processivity. By contrasting scanning to translating ribosomes, the study calculates initiation rates and redefines the optimal translation initiation context for zebrafish.

Highlights

- RCP-seq allows the study of both scanning and translating ribosomal complexes
- Threading of mRNA through the small ribosomal subunit loads the subunits onto the 5′ UTR
- Processivity of small ribosomal subunits is high while scanning 5′ UTRs
- Direct measurements of initiation redefine optimal start codon context

Giess et al., 2020, Cell Reports *31*, 107534
April 21, 2020
https://doi.org/10.1016/j.celrep.2020.107534

CellPress

Cell Reports
Report

Profiling of Small Ribosomal Subunits Reveals Modes and Regulation of Translation Initiation

Adam Giess,[1,4] Yamila N. Torres Cleuren,[1,4,*] Håkon Tjeldnes,[1] Maximilian Krause,[1,2] Teshome Tilahun Bizuayehu,[1] Senna Hiensch,[2] Aniekan Okon,[3] Carston R. Wagner,[3] and Eivind Valen[1,2,5,*]
[1]Computational Biology Unit, Department of Informatics, University of Bergen, Bergen 5020, Norway
[2]Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen 5008, Norway
[3]Department Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: yamilatorrescleuren@gmail.com (Y.N.T.C.), eivind.valen@gmail.com (E.V.)
https://doi.org/10.1016/j.celrep.2020.107534

SUMMARY

Translation initiation is often attributed as the rate-determining step of eukaryotic protein synthesis and key to gene expression control. Despite this centrality, the series of steps involved in this process is poorly understood. Here, we capture the transcriptome-wide occupancy of ribosomes across all stages of translation initiation, enabling us to characterize the transcriptome-wide dynamics of ribosome recruitment to mRNAs, scanning across 5′ UTRs and stop codon recognition, in a higher eukaryote. We provide mechanistic evidence for ribosomes attaching to the mRNA by threading the mRNA through the small subunit. Moreover, we identify features that regulate the recruitment and processivity of scanning ribosomes and redefine optimal initiation contexts. Our approach enables deconvoluting translation initiation into separate stages and identifying regulators at each step.

INTRODUCTION

In eukaryotes, translation initiation is a highly orchestrated sequence of events where the ribosomal 43S pre-initiation complex (PIC) is first recruited to the beginning of the transcript through interactions with initiation factors and the 5′ m^7G cap (Sonenberg and Gingras, 1998). Previous studies have suggested two alternative models for the 43S PIC binding to mRNA (Kumar et al., 2016). In the first, mRNA is "threaded" through the mRNA channel of the complex, while in the second, mRNA "slots" directly into the channel, possibly leading to suboptimal scanning of the first nucleotides (Kumar et al., 2016). The 43S PIC then scans the transcript in a 5′-to-3′ direction until a suitable translation initiation site (TIS) is encountered. Upon recognition of the TIS, the large ribosomal subunit (60S) is recruited to form an elongation-capable 80S ribosome. These initiation steps are broadly acknowledged to be a rate-limiting factor in protein synthesis (Arava et al., 2003; Shah et al., 2013; Shirokikh and Preiss, 2018). Despite this, our knowledge of ribosome recruitment, scanning, and TIS recognition is limited.

Ribosome profiling (ribo-seq) has enabled global quantification and localization of translation through the capture of footprints from elongating 80S ribosomes (Ingolia et al., 2009). A limitation of ribo-seq, however, is that it is blind to ribosomes from other stages of translation. Recently, translation complex profiling (TCP-seq) was introduced in yeast, which circumvented this problem by crosslinking all stages of ribosomes to the mRNAs (Archer et al., 2016). However, because this technique relies on first purifying 80S ribosome-containing transcripts, it is limited to studying small ribosomal subunit (SSU) (40S) positioning to transcripts that have at least one 80S ribosome and are thus actively translated. Here, we have expanded this approach to capture footprints from all ribosome-associated mRNAs, including transcripts not bound by any 80S subunit. Our approach immobilizes all ribosomal subunits on the mRNA by paraformaldehyde crosslinking, followed by sucrose gradient separation of the small subunits from the 80S complexes (Archer et al., 2016; Figure 1A; STAR Methods). After extracting the RNA, sequencing libraries are made of each fraction using template switching, which enables the use of ultra-low input material (1 ng) (Hornstein et al., 2016). Because our method captures different populations of ribosomes than TCP-seq, we will refer to our modified protocol as "ribosome complex profiling" (RCP-seq).

Here, we use RCP-seq to capture footprints of both 80S ribosomes and SSUs across the transcriptome of a developing zebrafish embryo (Figure 1A; STAR Methods). Mapping scanning small subunits over 5′ UTRs allows us to distinguish three distinct phases during translational initiation: (1) recruitment of small subunits to the mRNAs, (2) progression along the 5′ UTR to the start codon, and (3) conversion of scanning to elongating ribosomes.

RESULTS

To investigate the regulation of translation initiation in a vertebrate, we performed RCP-seq during zebrafish embryo development (see STAR Methods). As expected under the scanning model of translation, the footprints from the small subunit fraction predominantly mapped to the 5′ UTR of the transcripts, while the elongating 80S footprints mainly mapped to the coding sequence (CDS). A sharp divide between the fractions occurred at the start codon consistent with the conversion of scanning

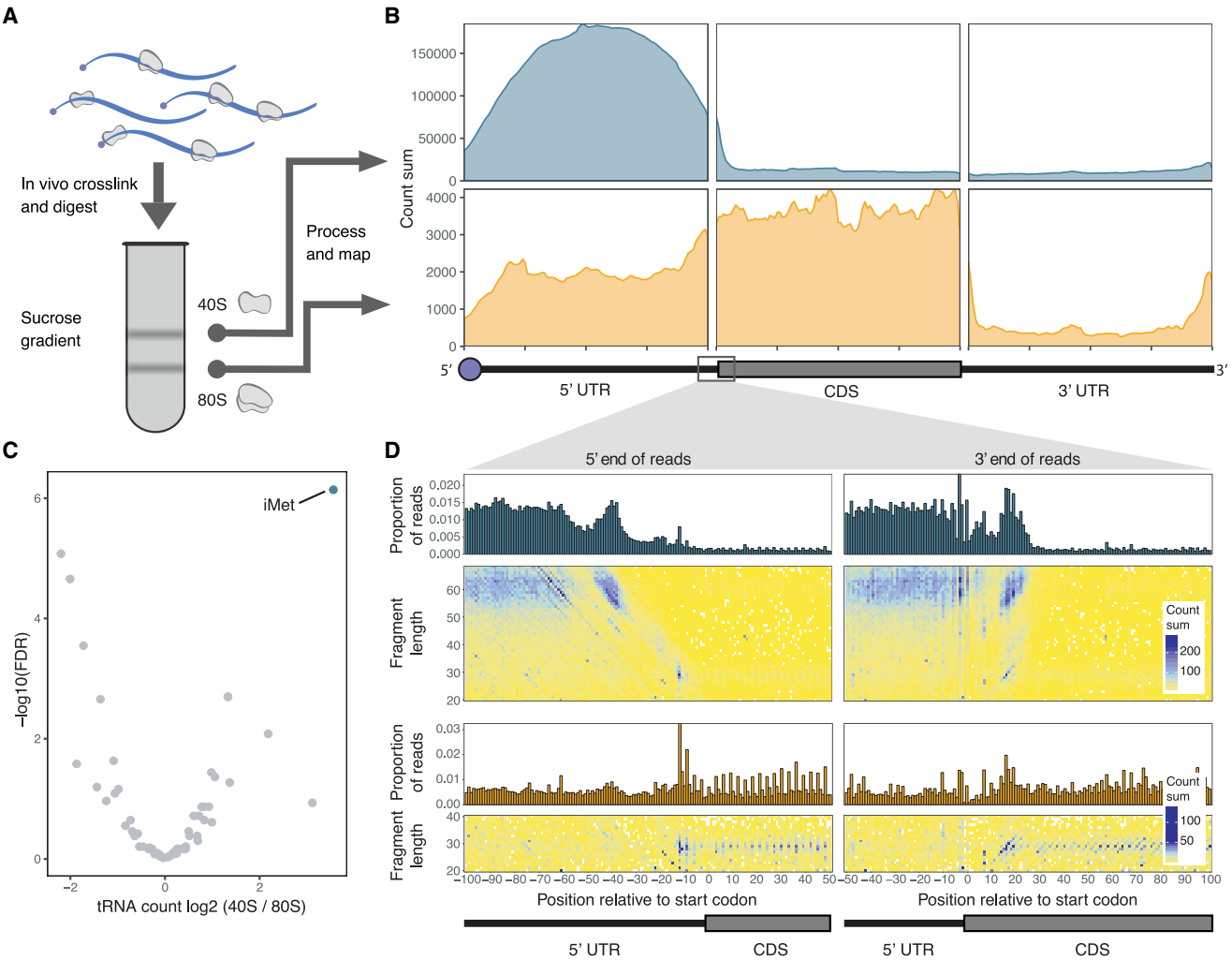


Figure 1. RCP-Seq Selectively Captures 80S Ribosomes and Small Subunits in Zebrafish

(A) Schematic representation of RCP-seq protocol.

(B) Coverage of RCP-seq reads across all transcripts. Footprints from small subunits (blue) map predominantly to 5′ UTRs, while 80S footprints (orange) map predominantly to coding sequence (CDS).

(C) Abundance of tRNA species (x axis) and false discovery rate (FDR) (y axis) between the RCP-seq small subunit (40S) and 80S fractions. Initiator Met-tRNA is highlighted (blue).

(D) Over-representation of RCP-seq small subunit (top) and 80S (bottom) fraction footprints around start codons. Counts of 5′ (left) or 3′ (right) ends of fragments are summed across highly expressed genes (≥10 fragments per kilobase of transcript per million mapped reads (FPKM). The barplots show the proportion of read counts per position (x axis), while the heatmaps show the same counts stratified by length (y axis) and colored by total count.

See also Figure S1.

43S PICs to elongating 80S ribosomes (Figure 1B). Here, the distribution of footprint lengths also revealed a range of ribosomal initiation conformations similar to those previously reported in yeast (Figures 1D and S1; Archer et al., 2016). As previously reported for TCP-seq, tRNA species contained within ribosomes are also selectively protected by RCP-seq, and consistent with capturing scanning ribosomes, we found initiator Met-tRNA strongly enriched in the small subunit fraction (Figure 1C). Taken together, these observations provide strong support for the selective capture of footprints from small subunits with RCP-seq.

We first sought to understand how the 43S PIC is recruited to the mRNA. The slotting and threading models are predicted to lead to substantially different profiles of protected fragments over the start of the transcript (Figure 2A). Zebrafish mRNAs have a strong enrichment of small subunit footprints coinciding with the transcription start site (Figure 2B). This enrichment is not present in non-coding RNAs, arguing that it is a feature only of translated RNA molecules and not an artifact of the method (Figure S3). The 5′ ends of these footprints all coincide with the transcription start sites (Figure 2B) and have a wide range of read lengths from the lower detection limit (~15 nt) up

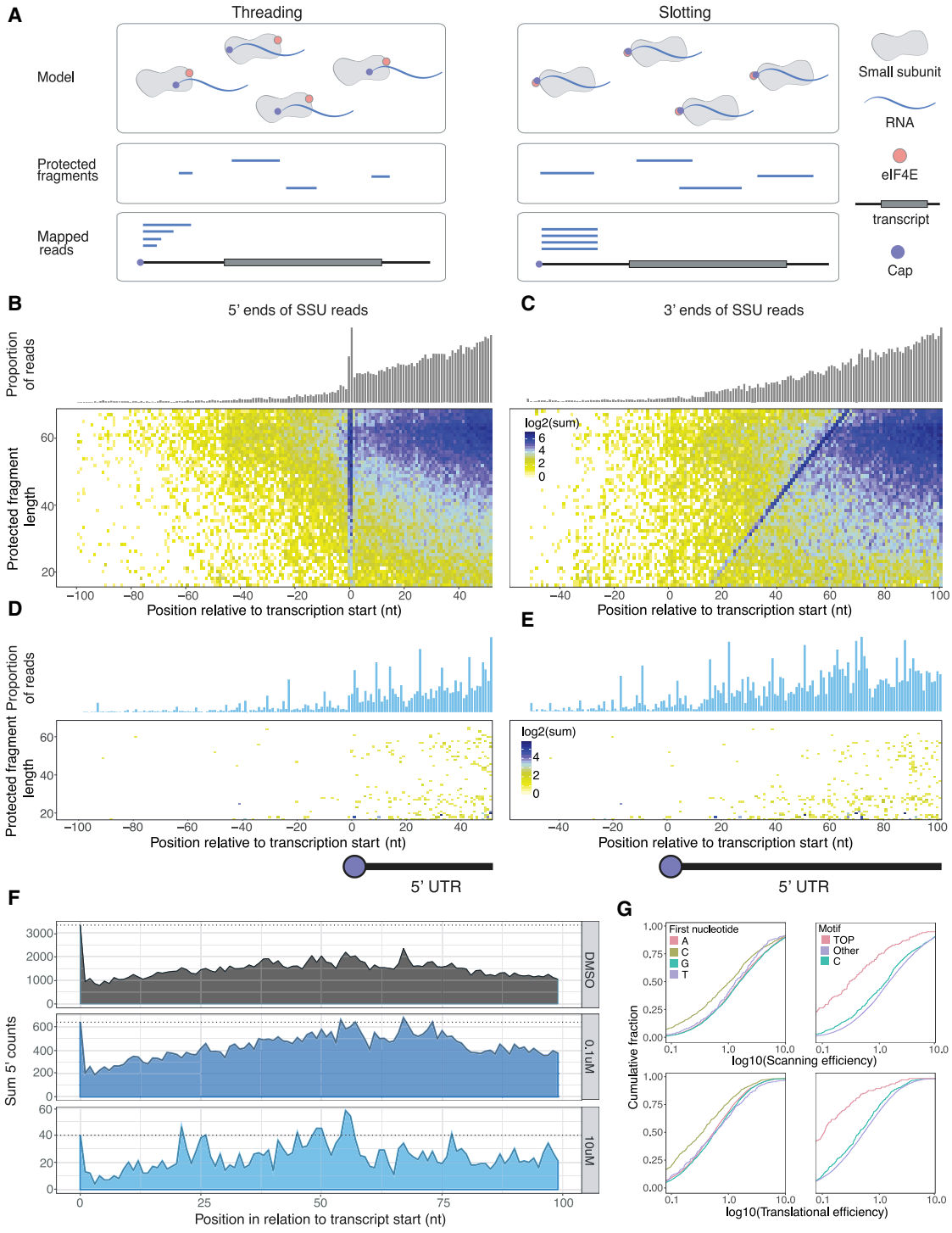


Figure 2. 43S PIC Recruitment and Impact of 5′ Transcript Features

(A) Schematic representation of two canonical recruitment models (top panel), "threading" (left) and "slotting" (right), the resulting protected fragments (middle panel), and the location of the mapped reads relative to the transcription start site (bottom panel).

(B) Heatmap of counts from 5′ ends of small subunit reads stratified by length (y axis) over each position (x axis) relative to transcription start site. Barplot above shows the proportion of reads at each position.

(C) Same as (B), but for 3′ ends of small subunit reads.

(D and E) Same as (B) and (C), but from 4Ei-10-inhibited samples (10 μM).

(legend continued on next page)

to ~80 nt, which is slightly longer than scanning 43S PICs (Figure 2C). The majority of footprints downstream of this peak correspond to the range commonly reported for 43S PICs (60–70 nt) (Kozak and Shatkin, 1978; Archer et al., 2016). A similar pattern was also observed when realigning data from TCP-seq in yeast to high-resolution mapping of transcription start sites (Wery et al., 2016; Figure S2; STAR Methods). These patterns of increasing lengths of small subunit footprints at the start of the transcript up to the size of the longest small subunit footprints are consistent with footprints from successive threading of the transcript through the mRNA channel of the 43S PIC complex.

Under the threading model, the cap-binding initiation factor eIF4E is placed at the leading edge of the 43S PIC and mRNA is threaded through the mRNA-binding channel (Elfakess et al., 2011; Kumar et al., 2016). To test the response of the transcription start site peaks to eIF4E inhibition, we sequestered eIF4E using the small molecule inhibitor 4Ei-10 (Okon et al., 2017), a cell-permeable prodrug improving upon 4Ei-1 (Smith et al., 2015), thereby specifically blocking eIF4E-cap binding, leading to a small but general inhibition of translation (Figure S4) followed by RCP-seq. This resulted in a global depletion of peaks at the transcription start sites consistent with these originating from eIF4E-dependent ribosome loading through threading (Figures 2D and 2E). The ratio of footprints starting at the transcription start site relative to footprints internal to the 5′ UTR (position 2–100) was reduced to ~63% and 1% of wild-type (WT) levels upon 0.1 and 10 μM 4Ei-10 treatment, respectively (Figures 2F and S4B; $p < 2.7 \times 10^{-9}$ and $p < 2.1 \times 10^{-15}$). In transcripts with very short 5′ UTRs, translation can be initiated through the translation initiator of short 5′ UTR (TISU) motif. In these transcripts, only threading is expected to be able to initiate translation, as slotting would deposit the small subunit too far downstream to scan the start codon (Elfakess et al., 2011; Kumar et al., 2016). Consistent with this and previous reports that TISU transcripts are eIF4E sensitive (Elfakess et al., 2011), upon eIF4E inhibition, we observed a strong reduction of transcription start site peaks compared to RNA levels across all transcripts initiated through the TISU motif (89%–92% reduction, $p < 2.2 \times 10^{-16}$; Figure S4C). The peak was also dose-dependently reduced relative to internal reads in the 5′ UTR (compare Figures S3E and S3F to Figure S3D). Collectively, this suggests threading is dependent on eIF4E and is a common recruitment pathway during early development.

We next asked which features could influence the recruitment of 43S PICs to the 5′ cap. To measure the amount of 43S PIC present on 5′ UTRs, we defined the scanning efficiency (SE) as the number of small subunit footprints over a 5′ UTR relative to its mRNA abundance (STAR Methods). This metric is conceptually identical to the widely used translational efficiency (TE), which measures elongating ribosomes relative to mRNA abundance (Ingolia, Lareau and Weissman, 2011). Using this metric, we observed that transcripts with a 5′ C, and to a lesser extent 5′ T, showed reduced SE and TE compared to transcripts beginning with an A or G (Figure 2G). This is consistent with *in vitro* biochemical studies which have shown that transcripts beginning with a pyrimidine (C/T) have a lower affinity for eIF4E binding than those starting with a purine (A/G) (Meyuhas and Kahan, 2015; Tamarkin-Ben-Harush et al., 2017). An initial C is a feature of transcripts containing a 5′ terminal oligo-pyrimidine (TOP) tract, a motif often present in mRNAs encoding the protein synthesis machinery and a target of mTOR-mediated translation control (Meyuhas and Kahan, 2015; Tamarkin-Ben-Harush et al., 2017; Danks et al., 2019). We found that while C has an effect, the overall effect on SE and TE reduction is dominated by mRNAs with the TOP motif (Figure 2G). This demonstrates that during early development, a reduced number of 43S PICs are recruited to TOP-motif-containing transcripts, resulting in reduced translation.

As the 43S PIC progresses through the 5′ UTR, it can encounter obstacles that can lead to termination of scanning. The RCP-seq data revealed this as a slight decline of scanning ribosomes throughout the 5′ UTR (Figure 3A). Under the assumption that averaged across all transcripts scanning proceeds at a uniform pace throughout the 5′ UTR, we compared the density of small subunit complexes of all transcripts at the 5′ end of the mRNA to the density proximal to the start codon (Figure 3A). Based on this analysis, we estimate that on average across all transcripts about 68% of all ribosomes recruited to the 5′ end reach the start codon. The loss of scanning ribosomes is largely contingent on whether the 5′ UTR contains one or more upstream open reading frames (uORF) (Calvo et al., 2009; Chew et al., 2013; Table S1) with only a weak correlation (Spearman's rho: −0.03) with 5′ UTR length if you control for the number of uORFs (Figure 3B). In transcripts that lack a uORF, we find that scanning overall maintains high processivity endogenously, consistent with previous results from reporter constructs (Berthelot et al., 2004; Andreev et al., 2009; Dmitriev et al., 2009), with a median of 95% of ribosomes retained. Collectively, this argues that scanning is highly stable and globally regulated through 5′ UTR elements promoting disassociation.

In transcripts containing uORFs, the CDS is translated either from ribosomes that fail to recognize the often suboptimal uORF TIS (Kozak, 2002; Ingolia et al., 2009; Fritsch et al., 2012; Lee et al., 2012) or by reinitiating ribosomes that continue scanning after translating the uORF (Kozak, 1987b; Grant and Hinnebusch, 1994). uORF regulation of protein synthesis can

(F) Counts of 5′ ends of reads from small subunits at transcription start site (TSS) and 100 nt downstream. Dotted lines show the number of reads starting at the TSS. Three conditions are shown: control (DMSO) and treatment with 0.1 μM and 10 μM 4Ei-10.

(G) Empirical cumulative density of scanning efficiency (SE) (top) or translation efficiency (bottom) for highly expressed transcripts (>10 FPKM). Left panels show transcripts colored by their first nucleotide. An initial pyrimidine (C/T) results in lower SE (C: $p < 2.2 \times 10^{-16}$, T: $p < 2.2 \times 10^{-16}$; number of transcripts per group, A = 10,388, C = 1,663, G = 8,335, T = 967) and translational efficiency (TE) (C: $p < 2.2 \times 10^{-16}$, T: $p < 2.2 \times 10^{-16}$) than a purine (A/G). Right panels show transcripts starting with a TOP motif. These have reduced SE ($p < 2.2 \times 10^{-16}$) and TE ($p < 2.2 \times 10^{-16}$), while non-TOP transcripts starting with a C show a minor reduction in SE ($p < 2.2 \times 10^{-16}$) and translation efficiency ($p < 2.2 \times 10^{-16}$) compared to non-TOP transcripts starting with the other 3 nt ("other" in figure). Number of transcripts per group: C = 1,257, other = 19,690, TOP = 406.

See also Figures S2–S4.

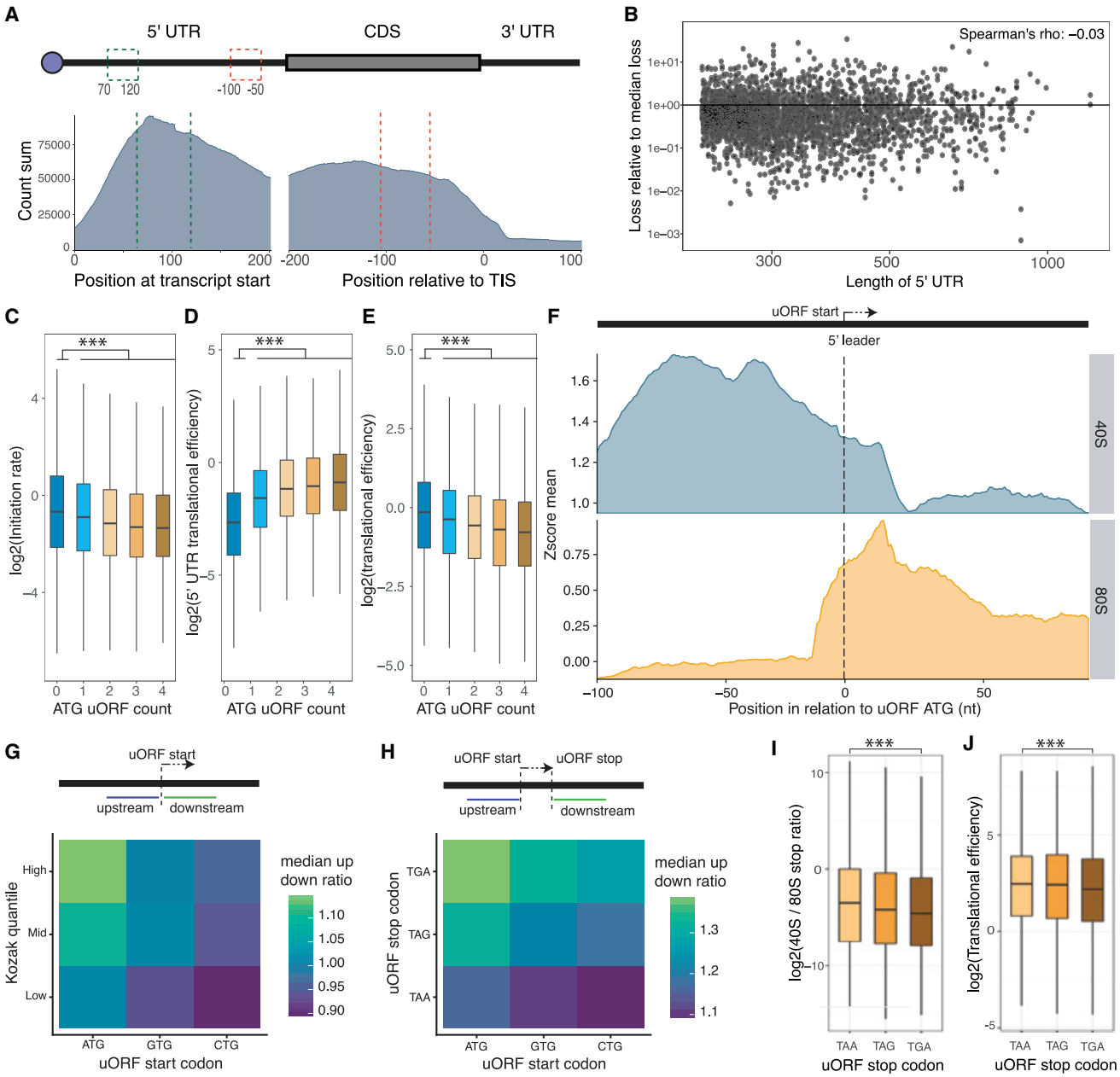


Figure 3. 43S PICs Processivity across 5′ UTRs Is Affected by uORFs

(A) Regions for estimations of small ribosomal subunit loss across 5′ UTRs, indicated by dashed lines. Transcripts were selected to be protein coding with 5′ UTRs regions ≥200 nt.

(B) Loss of small subunits scanning the 5′ UTR (y axis) as a function of 5′ UTR length (x axis). To control for the strong dependency between length of 5′ UTR and number of uORFs, the loss is calculated relative to the median loss of small subunits for all 5′ UTRs with the same number of ATG-initiated uORFs (horizontal line). Transcripts are selected to have 5′ UTR >220 nt and 5′ proximal SSU >10 FPKM and TIS proximal regions SSU >0 FPKM.

(C–E) The impact of the number of uORFs on (C) scanning subunits on 5′ UTR, (D) the TE of the 5′ UTR, and (E) the TE of the protein.

(F) Coverage of small subunit (40S) footprints (top, blue) and 80S complex footprints (bottom, orange) in fixed windows of 100 nt up- and downstream of the first ATG uORF.

(G) Heatmaps showing the rate of scanning subunit consumption as measured by the ratio of small subunit reads upstream versus downstream of all uORF start codons stratified by surrounding Kozak score and start codon.

(H) Similar to (G), but with ranking of start and stop codon, measuring upstream versus downstream of all uORFs.

(I and J) The impact of stop codon identity on the ratio of small subunit footprints to 80S complex footprints over uORF stop codons (I) and effects on TE of downstream CDS (J) (***p < 0.001).

therefore be assessed globally by measuring SSU consumption and relative translational levels in the 5′ UTR and CDS regions (Figures 3C–3E). Consistent with our global estimates, we find a local decline of 43S PIC footprints coinciding with an increase in 80S footprints at uORF TISs (Figure 3F). The ratio of the 43S PIC density upstream versus downstream of a uORF TIS can therefore quantify to what extent uORFs consume scanning 43S PICs (Figure 3F). As expected, uORFs starting with an ATG start codon (Figures 3G and 3H) and with a TIS context similar to the Kozak sequence (Figure 3G) have the highest 43S PIC consumption.

We found the ability of the small subunit to resume scanning after uORF translation to be dependent on the choice of stop codon. For proteins, TAA and TGA have been reported as the most and least efficient termination codons, respectively (Bonetti et al., 1995). Consistently, uORFs with TGA have the greatest reduction of downstream scanning small subunits (Figure 3H) and the lowest ratio of small subunit to 80S complexes over their stop codons (Figure 3I). Globally, this less efficient stop codon recognition leads to a small but significant effect on the TE of the downstream CDS (Figure 3J), suggesting that failure to recognize a stop codon can result in extended uORF translation and decreased rates of reinitiation after the translation of the extended uORF (Luukkonen et al., 1995; Kozak, 2001; Szamecz et al., 2008; Mohammad et al., 2017).

Whether a 43S PIC will recognize the TIS and trigger initiation of translation depends on the sequence surrounding the start codon. For many species, studies have defined an optimal consensus sequence for translation initiation (the Kozak sequence; Kozak, 1986, 1987a), often using indirect measures such as sequence conservation (Grzegorski et al., 2014) or reporter protein expression (Noderer et al., 2014). Uniquely, RCP-seq enables us to directly measure the average initiation rate (IR) on individual transcripts as the ratio of 80S ribosomes in the CDS to small subunit complexes in the 5′ UTR (Figure 4A; STAR Methods). By calculating the median IR of all transcripts containing a specific nucleotide at a specific position, this model revealed that the consensus of maximized IR is identical to the known zebrafish Kozak sequence (AAACATG) (Grzegorski et al., 2014; Figure 4B). This model, however, considers positions independently and therefore only reflects an average over sequences with high IR and not the efficiency of any particular sequence. To obtain this, we grouped all genes with identical sequence context and ranked these sequences by their median IR (Figures 4C and 4D). The resulting ranking was consistent with a previous assessment of a small number of sequences in zebrafish (Grzegorski et al., 2014) but surprisingly revealed that the Kozak sequence is not the optimal context. The highest scoring sequence was CATCATG, which differs by two bases from the consensus Kozak sequence (C at −4 and T at −2). More surprisingly, several sequences that differ strongly to the reported Kozak sequence rank above it. To test this new metric, we constructed GFP mRNA reporters with three different initiation sequences but otherwise identical: (1) AAGC, a sequence highly similar to the Kozak but with low IR; (2) AAAC, the Kozak sequence previously defined for zebrafish; and (3) TGGA, a sequence differing at all four bases from the Kozak but with greater IR. The TE (see STAR Methods) of these reporters when injected into zebrafish embryos confirmed our direct measurements from IR (Figure 4E), with statistically significant differences among all three contexts, with TGGA showing consistently higher TE than the other two. Taken together with previous reports of weaker than expected correlations between Kozak sequences and translation (Vogel et al., 2010; Pop et al., 2014), this demonstrates that a Kozak-similarity measure does not capture the complexity of start codon recognition but can be obtained by transcriptome-wide quantification through methods such as RCP-seq.

DISCUSSION

In this study, we expanded the TCP-seq protocol in two key aspects: (1) we capture all small subunits, not only those that co-occur on transcripts with 80S ribosomes; and (2) we use template switching in the library preparation to enable the use of less input material (Hornstein et al., 2016). This method, RCP-seq, therefore captures ribosomal complexes globally from all stages of the translation process and can be easily applied to other systems with limited input material, such as specific polysomal fractions or cell types. We used RCP-seq to study the dynamics of translation initiation during early stages of development in a vertebrate system, zebrafish. The longer 5′ UTRs of zebrafish allows for a detailed analysis of initiation by spatial separation of recruitment, scanning and start codon recognition.

Our data support the threading model of ribosome recruitment to mRNA. At the 5′ end of mRNAs, we observed a ‘‘ladder’’ of differentially sized fragments with 5′ ends coinciding with the transcription start site (Figures 2B and 2C). Fragment sizes shorter than the length of the 40S mRNA tunnel are consistent with the mRNA gradually entering the tunnel but conflicts with a slotting model where single-sized fragments would be expected (Figure 2A). However, two alternative explanations could also potentially account for these fragments. In the first, the SSU could be slotted adjacent to the 5′ cap but then proceed to backslide in the 5′ direction. However, previous studies have shown that mRNA binding by factors eIF4A, eIF4B/H, and eIF4F prevents the SSU from backsliding (Siridechadilok et al., 2005; Spirin, 2009), which makes it unlikely that the abundant 5′ reads (suggesting a frequent occurrence) are due to backsliding. The second possibility is that these reads are simply 3′-to-5′ degradation intermediates. However, two observations argue against this possibility. First, non-coding RNAs have very few 5′ reads, arguing for a translation-dependent origin (Figure S3G). Second, sequencing reads from degradation intermediates (and other possible artifacts) would be expected to increase when ribosome scanning is inhibited. Instead, upon eIF4E inhibition, we observe that these short 5′ fragments disappear together with fragments derived from SSU scanning (Figures 2D–2F). We therefore conclude that threading of mRNAs is the most likely explanation for the presence of these fragments. Moreover, TCP-seq libraries from yeast realigned to cap analysis gene expression (CAGE)-defined transcription start sites revealed a similar distribution of short 5′ fragments (Figure S2), supporting threading as a universal mechanism.

Our data allowed for the analysis of global processivity of scanning ribosomes, showing no correlation between SSU loss

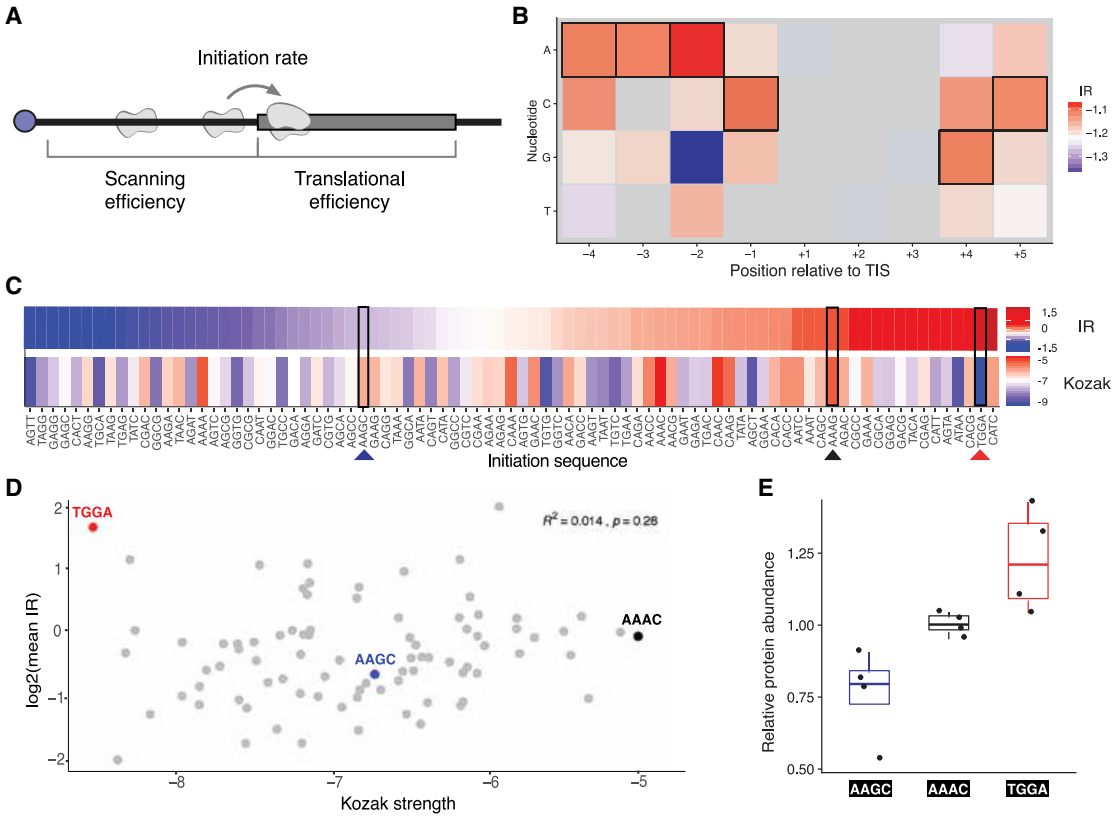


Figure 4. Direct Measurements of Initiation Rate

(A) Schematic representation of the three introduced metrics. SE is defined as the number of small subunit footprints over a 5′ UTR relative to its mRNA abundance. Initiation rate (IR) is defined as the ratio of 80S ribosomes in the CDS to small subunits in the 5′ UTR. TE measures elongating ribosomes relative to mRNA abundance.

(B) Median IRs for all transcripts containing nucleotide (y axis) at a specific position (x axis) relative to the protein TIS. The zebrafish Kozak sequence is highlighted with black borders (AAACATGGC).

(C) Mean IR for the entire sequence from −4 to −1 (top) with corresponding Kozak strength (bottom). Arrows indicate the sequences selected for reporter constructs.

(D) Correlation between Kozak strength and IR values shown in (C). The sites with very high similarity to the Kozak (more than −5.5) tend to have good IR, but overall, Kozak similarity is not a good predictor of IR.

(E) Relative protein abundance for GFP reporter constructs for three different initiation contexts, as measured by the protein/RNA abundance ratios in zebrafish embryos at 24 h post-fertilization.

and length of 5′ UTR, in agreement with reporter construct experiments in cell lines and *in vitro* (Andreev et al., 2009; Dmitriev et al., 2009). We found that the majority of scanning ribosomes reach the protein coding start codon and identified uORFs as a major cause of detachment. Consistent with previous studies on individual genes (Grant and Hinnebusch, 1994; McCaughan et al., 1995; Beznosková et al., 2016; Cridge et al., 2018), we find that the choice of stop codon affects uORF termination across the transcriptome and furthermore that poor stop codons can lead to an increase of readthrough 80S ribosomes, potentially decreasing the ability of SSUs to reinitiate at the CDS. This can result in a reduction of CDS translation, suggesting that the choice of uORF stop codon can globally tune protein expression. Since these estimates are calculated from global averages, they can represent a range of effects on individual genes. In fact, studies of individual cases have shown that many uORFs have little or no effect and that some may even lead to activation of downstream translation, where interplay between complex uORF arrangements can lead to dramatic change in TE. These effects are highly dependent on the uORF features and its location in the 5′ UTR (Gunišová and Valášek, 2014; Gunišová et al., 2016; Lin et al., 2019). Here, we have shown that, globally, uORFs are on average slightly inhibitory and that RCP-seq provides a novel view into this regulation by revealing the interplay of small and large ribosomal complexes.

Previous *in vivo* analyses of start context optimality have focused mainly on common/conserved sequences (Hernández et al., 2019) and construct reporters (Kozak, 1986) and have been limited to measuring their optimality only indirectly through translational output. By directly contrasting the abundance of scanning complexes relative to translating ribosomes, RCP-seq offers a unique perspective into start codon recognition. This enabled us to confirm previous observations that the Kozak sequence provides a strong initiation context but furthermore

revealed that there are other endogenous sequences that give rise to equal or better rates of initiation. This is consistent with reports of weaker than expected correlations between Kozak sequences and protein abundance in humans (Vogel et al., 2010) and yeast (Pop et al., 2014). Together, this demonstrates that RCP-seq can give unique insight into start codon recognition and provide a valuable tool for assessing or optimizing translation.

Overall, our approach enables the deconvolution of translation initiation into distinct events, including recruitment, processivity of small subunits and IRs. The RCP-seq protocol can be further applied to study samples with limited input material, which will allow addressing heterogeneity and specialization of the translation machinery, compartmentalized translation or tissue-specific translation. This opens for the possibility to obtain novel insights into scanning and initiating mechanisms across organisms and disease models.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Embryo sample collection and crosslinking
  - Separation of ribosomal small subunit and ribosomal complexes
  - RNA isolation
  - Construction of RNA control libraries
  - Construction of RCP-seq sequencing libraries
  - eIF4E inhibitor assays
  - Polysome profiling
  - Reporter assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Transcript definitions
  - Read trimming and alignment
  - RCP-seq fractionation
  - Further read processing
  - Read counting
  - Fragment lengths heatmaps and fragment distribution
  - Scaled coverage meta plots
  - Estimates of 43S PIC loss across 5′ UTRs
  - 5′ feature plots (Figure 2)
  - uORF plots (Figure 3)
  - Initiation plots (Figure 4)
  - Statistical testing and plotting
  - Formulas used in this study

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107534.

ACKNOWLEDGMENTS

We wish to thank members of the Valen and Thompson labs (University of Bergen, Norway), Ignatova and Duncan labs (University of Hamburg, Germany), and Preiss lab (Australian National University, Australia) for helpful discussions of methodologies. We wish to thank Andrea Pauli (IMP, Austria), Sushma Nagaraja-Grellscheid, and Eric Thompson (University of Bergen, Norway) for providing valuable feedback on the manuscript.

The project was funded by the Trond Mohn Foundation, the Norwegian Research Council (#250049) and core funding from the Sars International Centre for Marine Molecular Biology.

AUTHOR CONTRIBUTIONS

A.G., Y.N.T.C., and E.V. designed the research. Y.N.T.C., M.K., T.T.B., and S.H. performed the experiments. A.G., Y.N.T.C., H.T., and E.V. analyzed the data. A.O. and C.R.W. designed and synthesized the 4Ei-10 compound. All authors discussed the results and contributed to writing the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 8, 2019
Revised: February 28, 2020
Accepted: March 27, 2020
Published: April 21, 2020

REFERENCES

Andreev, D.E., Dmitriev, S.E., Terenin, I.M., Prassolov, V.S., Merrick, W.C., and Shatsky, I.N. (2009). Differential contribution of the m7G-cap to the 5′ end-dependent translation initiation of mammalian mRNAs. Nucleic Acids Res. *37*, 6135–6147.

Arava, Y., Want, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci USA *100*. https://doi.org/10.1073/pnas.0635171100.

Archer, S.K., Shirokikh, N.E., Beilharz, T.H., and Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. Nature *535*, 570–574.

Avdesh, A., Chen, M., Martin-Iverson, M.T., Mondal, A., Ong, D., Rainey-Smith, S., Taddei, K., Lardelli, M., Groth, D.M., Verdile, G., and Martins, R.N. (2012). Regular care and maintenance of a zebrafish (Danio rerio) laboratory: an introduction. J. Vis. Exp. (*69*), e4196.

Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J., and McCarthy, J.E. (2004). Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. Mol. Microbiol. *51*, 987–1001.

Beznosková, P., Gunišová, S., and Valášek, L.S. (2016). Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. RNA *22*, 456–466.

Bonetti, B., Fu, L., Moon, J., and Bedwell, D.M. (1995). The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. J. Mol. Biol. *251*, 334–345.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc. Natl. Acad. Sci. USA *106*, 7507–7512.

Chan, P.P., and Lowe, T.M. (2019). tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol Biol *1962*. https://doi.org/10.1007/978-1-4939-9173-0_1.

Chew, G.-L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F., and Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. Development *140*, 2828–2834.

Cridge, A.G., Crowe-McAuliffe, C., Mathew, S.F., and Tate, W.P. (2018). Eukaryotic translational termination efficiency is influenced by the 3′ nucleotides within the ribosomal mRNA channel. Nucleic Acids Res. *46*, 1927–1944.

Danks, G.B., Galbiati, H., Raasholm, M., Torres Cleuren, Y.N., Valen, E., Navratilova, P., and Thompson, E.M. (2019). Trans-splicing of mRNAs links gene transcription to translational control regulated by mTOR. BMC Genomics *20*, 908.

Dmitriev, S.E., Andreev, D.E., Ad'ianova, Z.V., Terenin, I.M., and Shatskiĭ, I.N. (2009). Efficient cap-dependent in vitro and in vivo translation of mammalian mRNAs with long and highly structured 5′-untranslated regions. Mol. Biol. (Mosk) *43*, 119–125.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*. https://doi.org/10.1093/bioinformatics/bts635.

Elfakess, R., Sinvani, H., Haimov, O., Svitkin, Y., Sonenberg, N., and Dikstein, R. (2011). Unique translation initiation of mRNAs-containing TISU element. Nucleic Acids Res. *39*, 7598–7609.

Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., and Brosch, M. (2012). Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome Res. *22*, 2208–2218.

Grant, C.M., and Hinnebusch, A.G. (1994). Effect of sequence context at stop codons on efficiency of reinitiation in GCN4 translational control. Mol. Cell. Biol. *14*, 606–618.

Grzegorski, S.J., Chiari, E.F., Robbins, A., Kish, P.E., and Kahana, A. (2014). Natural variability of Kozak sequences correlates with function in a zebrafish model. PLoS ONE *9*, e108475.

Gunišová, S., and Valášek, L.S. (2014). Fail-safe mechanism of GCN4 translational control–uORF2 promotes reinitiation by analogous mechanism to uORF1 and thus secures its key role in GCN4 expression. Nucleic Acids Res. *42*, 5880–5893.

Gunišová, S., Beznosková, P., Mohammad, M.P., Vlčková, V., and Valášek, L.S. (2016). In-depth analysis of cis-determinants that either promote or inhibit reinitiation on GCN4 mRNA after translation of its four short uORFs. RNA *22*, 542–558.

Haberle, V., Forrest, A.R., Hayashizaki, Y., Carninci, P., and Lenhard, B. (2015). CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. Nucleic Acids Res. *43*, e51.

Hernández, G., Osnaya, V.G., and Pérez-Martínez, X. (2019). Conservation and variability of the AUG initiation codon context in eukaryotes. Trends Biochem. Sci. *44*, 1009–1021.

Hornstein, N., Torres, D., Das Sharma, S., Tang, G., Canoll, P., and Sims, P.A. (2016). Ligation-free ribosome profiling of cell type-specific translation in the brain. Genome Biol. *17*, 149.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283–292.

Kozak, M. (1987a). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. *15*, 8125–8148.

Kozak, M. (1987b). Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Mol. Cell. Biol. *7*, 3438–3445.

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. Nucleic Acids Res. *29*, 5226–5232.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. Gene *299*, 1–34.

Kozak, M., and Shatkin, A.J. (1978). Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. J. Biol. Chem. *253*, 6568–6577.

Kumar, P., Hellen, C.U.T., and Pestova, T.V. (2016). Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. Genes Dev. *30*, 1573–1588.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*. https://doi.org/10.1038/nmeth.1923.

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B., and Qian, S.B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc. Natl. Acad. Sci. USA *109*, E2424–E2432.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics *25*.

Lin, Y., May, G.E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., and McManus, C.J. (2019). Impacts of uORF codon identity and position on translation regulation. Nucleic Acids Res. *47*, 9358–9367.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

Luukkonen, B.G., Tan, W., and Schwartz, S. (1995). Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. J. Virol. *69*, 4086–4094.

McCaughan, K.K., Brown, C.M., Dalphin, M.E., Berry, M.J., and Tate, W.P. (1995). Translational termination efficiency in mammals is influenced by the base following the stop codon. Proc. Natl. Acad. Sci. USA *92*, 5431–5435.

Meyuhas, O., and Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. Biochim. Biophys. Acta *1849*, 801–811.

Mohammad, M.P., Munzarová Pondelíčková, V., Zeman, J., Gunišová, S., and Valášek, L.S. (2017). In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. Nucleic Acids Res. *45*, 2658–2674.

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S., et al. (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. Genome Res. *23*, 1938–1950.

Noderer, W.L., Flockhart, R.J., Bhaduri, A., Diaz de Arce, A.J., Zhang, J., Khavari, P.A., and Wang, C.L. (2014). Quantitative analysis of mammalian translation initiation sites by FACS-seq. Mol. Syst. Biol. *10*, 748.

Okon, A., Han, J., Dawadi, S., Demosthenous, C., Aldrich, C.C., Gupta, M., and Wagner, C.R. (2017). Anchimerically activated ProTides as inhibitors of cap-dependent translation and inducers of chemosensitization in mantle cell lymphoma. J. Med. Chem. *60*, 8131–8144.

Peterson, S.M., and Freeman, J.L. (2009). RNA isolation from embryonic zebrafish and cDNA synthesis for gene expression analysis. J. Vis. Exp. (*30*), 1470.

Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol. Syst. Biol. *10*, 770.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. Cell *153*. https://doi.org/10.1016/j.cell.2013.05.049.

Shirokikh, N.E., and Preiss, T. (2018). Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions. Wiley Interdiscip. Rev. RNA *9*, e1473.

Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., and Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science *310*, 1513–1515.

Smith, K.A., Zhou, B., Avdulov, S., Benyumov, A., Peterson, M., Liu, Y., Okon, A., Hergert, P., Braziunas, J., Wagner, C.R., et al. (2015). Transforming growth factor-β1 induced epithelial mesenchymal transition is blocked by a chemical antagonist of translation factor eIF4E. Sci. Rep. *5*, 18233.

Sonenberg, N., and Gingras, A.-C. (1998). The mRNA 5′ cap-binding protein eIF4E and control of cell growth. Curr. Opin. Cell Biol. *10*, 268–275.

Spirin, A.S. (2009). How does a scanning ribosomal particle move along the 5′oes a scanning ribosomal particle mmRNA? Brownian Ratchet model. Biochemistry *48*, 10688–10692.

Szamecz, B., Rutkai, E., Cuchalová, L., Munzarová, V., Herrmannová, A., Nielsen, K.H., Burela, L., Hinnebusch, A.G., and Valásek, L. (2008). eIF3a cooperates with sequences 5′ of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. Genes Dev. *22*, 2414–2425.

Tamarkin-Ben-Harush, A., Vasseur, J.J., Debart, F., Ulitsky, I., and Dikstein, R. (2017). Cap-proximal nucleotides via differential eIF4E binding and alternative promoter usage mediate translational response to energy stress. eLife *6*, e21907.

Vogel, C., Abreu, Rde.S., Ko, D., Le, S.Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. *6*, 400.

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.S., Gautheret, D., and Morillon, A. (2016). Nonsense-mediated decay restricts LncRNA levels in yeast unless blocked by double-stranded RNA structure. Mol. Cell *61*, 379–392.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Danio rerio embryonic samples	This study	NA
Chemicals, Peptides, and Recombinant Proteins		
cOmplete, Mini, EDTA-free protease inhibitor cocktail	Roche	Cat#COEDTAF-RO
4Ei-10	Wagner lab	NA
Pronase	Roche	Cat#PRON-RO
Superase-In (RNase Inhibitor)	Invitrogen	Cat#AM2694
RNase I	Thermo Scientific	Cat#EN0601
T4 PNK	New England Biolabs	Cat#M0201L
High Fidelity Phusion MasterMix	Thermo Fisher	Cat#F-531L
Critical Commercial Assays		
Ribo-zero Gold rRNA Removal kit	Illumina	Cat#MRZG126
TaKaRa SMARTer smRNA-Seq Kit for Illumina	Clontech	Cat#635029
RNA Clean & concentrator-5	Zymo Research	Cat#R1013
Agilent Bioanalyzer DNA High Sensitivity kit	Agilent	Cat#5067-4626
Pierce BCA kit	Thermo Scientific	Cat#23225
mMessage mMachine	Thermo Fisher	Cat#AM1340
Deposited Data		
Raw and analyzed data	This study	PRJEB33323
CAGE zebrafish (2hpf, 4hpf, 6hpf)	Nepal et al., 2013	SRA055273
Ribo-seq zebrafish (2hpf, 4hpf, 6hpf)	Chew et al., 2013	GSE46512
CAGE yeast	Wery et al., 2016	GSE69384
Experimental Models: Organisms/Strains		
Danio rerio AB strain	This study	NA
Oligonucleotides		
SP6-AAAC-fw: CTTGATTTAGGTGACACTATAGTACGGATTCG TACACCAGTAAAGGCG**AAAC**ATGGTGAGCAAGGGCGAG	Sigma	NA
SP6-AAGC-fw: CTTGATTTAGGTGACACTATAGTACGGATTCG TACACCAGTAAAGGCG**AAGC**ATGGTGAGCAAGGGCGAG	Sigma	NA
SP6-TGGA-fw: CTTGATTTAGGTGACACTATAGTACGGATTCG TACACCAGTAAAGGCG**TGGA**ATGGTGAGCAAGGGCGAG	Sigma	NA
M13-rev: GGAAACAGCTATGACCATG	Sigma	NA
SP6-RFP-AAAC-fw: CTTGATTTAGGTGACACTATAGTACGGA TTCGTACACCAGTAAAGGCG**AAAC**ATGGTGTCTAAGGGCGA AGA	Sigma	NA
RFP-rev: CAGCCATACCACATTTGTAGAG	Sigma	NA
eGFP qPCR fw: TCAAGGAGGACGGCAACATC	Sigma	NA
eGFP qPCR rev: AACTCCAGCAGGACCATGTG	Sigma	NA
RFP qPCR fw: TGCAGAAGAAAACACTCGGC	Sigma	NA
RFP qPCR rev: TGTCGGCCTCCTTGATTCTT	Sigma	NA
Serp1 qPCR fw: GTGGATCAGCGATATTCCAG	Sigma	NA
Serp1 qPCR rev: AGAGAAGCGGAATGGTCGAG	Sigma	NA
Elf1a qPCR fw: CTCCTCTTGGTCGCTTTGCT	Sigma	NA
Elf1a qPCR rev: GCCTTCTGTGCAGACTTTGTGA	Sigma	NA

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pT2KXIGdeltaIn-MCS-birA-tagRFP	Addgene	Cat#58378; RRID: Addgene_58378
Software and Algorithms		
bowtie v2.2.4	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
STAR v.2.5.1a	Dobin et al., 2013	NA
tophap2 v2.0.14	Kim et al., 2013	NA
samtools v1.2	Li et al., 2009	http://www.htslib.org/
tRNAscan-SE v2.0	Chan and Lowe, 2019	NA
R v3.6.0	R foundation	NA
CageR v1.12.0	Haberle et al., 2015	NA
edgeR v3.12.1	Robinson et al., 2010	NA
All scripts and data analyses	This study	https://github.com/agiess/RCP_processing
Other		
Agencourt RNAClean XP beads	Beckman Coulter	Cat #A63987

RESOURCE AVAILABILITY

Lead Contact

Further information and request for resources and reagents should be directed to the Lead Contact, Eivind Valen (eivind.valen@gmail.com).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Custom scripts used to process the RCP-seq libraries are available at the following link: https://github.com/agiess/RCP_processing. The RCP-seq libraries have been uploaded to the ENA database (accession number PRJEB33323).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Danio rerio adult fish (AB type), aged 6 months to 2 years, were used to collect embryos using standard zebrafish husbandry (Avdesh et al., 2012). In short, AB males and females were separated the day before mating. Shortly after first light, fish were put together and allowed to mate for 10 min. Embryos were collected in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$), cleaned, and dechorionated using pronase (1 mg/ml, Sigma) for 5 min. Embryos were cleaned thoroughly after dechorionation and grown on 2% agarose plates containing E3 medium until the desired stage.

METHOD DETAILS

Embryo sample collection and crosslinking

For RCP-seq zebrafish embryonic samples and their respective RNA-seq controls, dechorionated embryos were stage-matched for sample collection (at stages 64-cell, Sphere and Shield). 200 embryos were transferred to 2 mL Eppendorf tubes per sample and washed twice in PBS with protease inhibitors (1:100 dilution, cOmplete, Mini, EDTA-free protease inhibitor cocktail) immediately prior to crosslinking, and left in 250 μl.

Embryos were snap-chilled by addition of 750 μL ice-cold PBS with 4% paraformaldehyde (PFA, freshly prepared) and immediately placed on ice. Samples were incubated for 15 min on ice, with gentle agitation. PFA medium was fully removed and 1 mL lysis buffer (20 mM HEPES-KOH pH 7.4, 100 mM KCl, 2 mM $MgCl_2$) added. Glycine was added to a final 0.25 M concentration for PFA quenching, and samples incubated for 5 min on ice. Embryos were then washed twice in lysis buffer and resuspended in lysis buffer supplemented with 0.5 mM DTT, 2 μL Superase-In (RNase Inhibitor) and 1x protease inhibitor. Samples were immediately flash frozen in liquid nitrogen and stored at −80°C until used.

Separation of ribosomal small subunit and ribosomal complexes

Samples were lysed in a cold room (4°C) by first shaking at 1,300 rpm for 10 min, followed by passing six times through a 27G needle. Samples were clarified by centrifuging for 15 min at 14,500 g, 4°C. The OD260 absorbance of the supernatant was measured by Nanodrop. 1/10th of each lysate was kept for RNA sequencing of that sample (RNA controls). Based on the absorbance readings, samples were digested using RNase I (0.0383 U x sample volume (μl) x OD 260 absorbance) for 45 min, at 23°C, 300 rpm shaking. Linear sucrose gradients were made from 5% to 30% sucrose solutions (containing 50 mM Tris-HCl pH 7.0, 50 mM NH_4Cl, 4 mM $MgCl_2$, 1 mM DTT) with Biocomp Gradient Station (long cap program, 5% to 30%). Gradients were cooled down for 45 min at 4°C. Samples were layered on top of each gradient and tubes were centrifuged in a SW-41 rotor (Beckman-Coulter) at 4°C, 38,000 rpm for 4 h. The gradients were fractionated using Biocomp Gradient Station and the small subunit and 80S fractions were identified and collected by monitoring the absorbance profile at 254 nm.

RNA isolation

The collected fractions and RNA controls were supplemented with 1% SDS, 10 mM EDTA, 10 mM Tris-HCl pH 7.4, and 10 mM glycine. One volume of phenol:chloroform:isoamyl alcohol (pH 4.5) was added to each sample, immediately placed on a shaker at 65°C, 1,300 rpm for 45 min. After a 5 min centrifugation at 15,000 g at room temperature, the aqueous phase was transferred to a new tube and precipitated by addition of 20 μg glycogen, 0.1 volume 3M sodium acetate (pH 4.5) and 2.5 volumes absolute ethanol. Samples were precipitated at −20°C for at least 3 h. RNA was pelleted by centrifugation at 21,000 g for 40 min, followed by two washes with 80% ethanol and 20 min centrifugation at 21,000 g, 4°C. After drying the pellet, it was resuspended in 17 μL water, and concentration assessed on Nanodrop.

Construction of RNA control libraries

After RNA extraction, each RNA control sample was DNase-treated with TURBO DNase (Invitrogen) for 30 min at 37°C. The reaction was stopped by addition of EDTA and incubation at 70°C for 10 min. DNase-treated RNA samples were cleaned up with RNA clean & concentrator-5 spin columns (Zymo Research). Next, rRNA was removed by using Ribo-Zero Magnetic Gold Kit (Illumina) following manufacturer's instructions and purifying RNA using clean & concentrator-5 spin columns (Zymo Research) in a final volume of 13 μl. RNA samples were then fragmented for 40 min at 90°C in an alkaline fragmentation solution (2x solution: 0.5 Vol 0.5 M EDTA, 15 Vol 100 mM Na_2CO_3, 110 Vol 100 mM $NaHCO_3$). Fragmentation was stopped by addition of precipitating reagents (20 μg glycogen, 0.1 volume 3M sodium acetate (pH 4.5) and 2.5 Vol absolute ethanol) and RNA was precipitated by incubation at −20°C for minimum 3 h. Fragmented RNA control samples were end-repaired together with the RCP-seq fractions as below. Followed by library construction, following manufacturers' instructions as described below. Sequencing was performed at 75bp single-end reads.

Construction of RCP-seq sequencing libraries

RNA samples were end-repaired by first incubating for 2 min at 80°C and on ice for 5 min, followed by the addition of 2 μL 10x T4 PNK buffer, 1 μL SUPERase-In, and 1 μL T4 PNK (10U/μl), and incubated for 2 h at 37°C. rRNA fragments were removed by using Ribo-Zero Magnetic Gold Kit (Illumina) following manufacturer's instructions and purifying RNA using clean & concentrator-5 spin columns (Zymo Research) in a final volume of 13 μl. Libraries were constructed using the TaKaRa SMARTer smRNA-Seq Kit for Illumina, following manufacturers' instructions without intermediate freezing points. In general, ATP was used for polyadenylation depending on the amount of starting material (less than 25 ng, no ATP). The number of cycles used was optimized per sample in the PCR amplification step and samples were eluted in a final volume of 20 μl. RCP-seq library sizes were checked on Agilent Bioanalyzer DNA High Sensitivity chips. Depending on size distribution, small and/or large fragments were removed by using AMPure XP beads (in order to remove adaptor dimers and too large fragments not resulting from ribosomal protection). Sequencing was performed on a NextSeq 500 (Illumina) at the Norwegian Sequencing Centre in Oslo, high output mode with single reads of 75 bp or 150 bp.

eIF4E inhibitor assays

The eIF4E inhibitor 4Ei-10 (or 6a; Okon et al., 2017) was synthesized at the Wagner lab (University of Minnesota, USA). The inhibitor was diluted in DMSO to a concentration of 100 μM, and kept frozen as stock. Further dilutions were performed in water. One nl of two concentrations (10 μM and 100 nM) were injected into dechorionated zebrafish embryos between 1-4 cell stages, in parallel with DMSO injections as controls. Embryos were allowed to continue development and samples were collected for RCP-seq at 64-cell and Shield stages (as described above). Flash frozen samples at Shield stages were also collected for polysome profiling in order to quantify effects on global translation for both 4Ei-10 and control DMSO injected samples. All comparisons between treated and untreated samples are done with the DMSO as the untreated control.

Polysome profiling

Control (DMSO-injected) and inhibitor-injected samples were collected at Shield stage and flash frozen to halt ribosomes. Samples were lysed in polysome lysis buffer (10 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 100 mM KCl, 1% Triton X-100) by shaking for 10 min at 1,300 rpm and 4°C, followed by 6x lysing through a 27G needle. Samples were centrifuged at 14,500 g, for 15 min, 4°C. The supernatant was transferred to a new tube and its absorbance at 260 nm was measured with Nanodrop One to quantify RNA content. Linear sucrose gradients (15%–45%) were prepared in a buffer containing 20mM HEPES KOH (pH 7.4), 5 mM $MgCl_2$, 100 mM

KCl, 2 mM DTT and 10 μL Superase-In. Gradients were cooled down for 45 min at 4°C. Samples were layered atop the gradients and ultracentrifuged in a Beckman-Coulter SW-41 rotor at 36,000 rpm, 2 h, 4°C. Polysome profiles of the gradients were obtained by in-line 254 nm absorbance measuring with Biocomp Gradient Station. Translation was quantified as the area under the curve (AUC), by comparing the ratio of polysomes AUC to monosome AUC, and the overall AUC (general translation, monosome + polysomes).

Reporter assays

Translational efficiency of initiation contexts was tested using eGFP reporters. Three eGFP reporters with different initiation contexts were synthesized. The coding sequence of eGFP was amplified from pCS2+-eGFP vector using High Fidelity Phusion MasterMix (ThermoFisher, #F-531L). The forward primers for the PCR included the SP6 promoter sequence, followed by 26 bp of eIF3d leader sequence and the respective start codon context; the reverse primer (M13-rev) was located after the common SV40 termination signal included in the pCS2+ vector backbone (see Key Resources Table). Following synthesis by a two-step PCR, samples were gel-purified and RNA was synthesized with SP6 mMessage mMachine (Thermo Fisher, #AM1340), following manufacturer's recommendations, and cleaned-up using Zymo RNA Clean & Concentrator-25 columns (Zymo Research, #R1013).

To control for GFP expression changes, we further synthesized RFP mRNA with a fixed start codon context. The coding sequence of RFP was amplified from the pT2KXIGdeltaIn-MCS-birA-tagRFP vector (AddGene #58378) using High Fidelity Phusion MasterMix. The forward primers for the PCR included the SP6 promoter sequence and the reverse Primer was located after the common SV40 termination signal included in the pT2KXIG vector backbone (see Key Resources Table). Following synthesis by a two-step PCR, samples were processed as above to obtain mRNA.

Stock RNA solutions for injections at 120 ng/ul were made based on Nanodrop & Qubit RNA concentration measurements. RNA solutions containing RFP and eGFP reporters were co-injected at a final concentration of 50 ng/ul per reporter (with phenol red added for injection visualization) and injected at 1 nL per embryo at 2-4 cell stage (dechorionated embryos). Embryos were collected and dechorionated as described above.

Groups of 25 embryos for each of the samples were collected at 24 hpf and used for qRT-PCR and eGFP protein quantification. For qRT-PCR, RNA was extracted using Trizol, and processed following the protocol previously described (Peterson and Freeman, 2009). eGFP RNA expression was quantified against RFP for injection control and endogenous controls *serp1* and *Elf1a*.

For eGFP protein quantification, the 25 embryos collected were homogenized in 100 mM Tris (pH 7.5) with 1% Triton X-100, and processed as described in Grzegorski et al. (2014). After quantifying total protein concentration with the Pierce BCA kit (Thermo Scientific), the samples were diluted to equal total protein concentrations. A Qubit fluorometer (Life Technologies) was used to measure eGFP fluorescence, using 9 technical replicates for each sample, and with GFP dilutions as measurement controls.

Translational efficiency was calculated as described previously (Grzegorski et al., 2014), by quantifying RNA with qRT-PCR and quantifying eGFP protein by measuring eGFP fluorescence in Qubit. Relative eGFP protein (average of all 9 technical replicates for each sample) was calculated compared to AAAC (zebrafish Kozak), divided by the relative abundance of eGFP RNA in each corresponding sample as measured by normalized qPCR abundance. For the AAAC samples, we compared their eGFP values to the average of the group to get their individual TE values.

QUANTIFICATION AND STATISTICAL ANALYSIS

Transcript definitions

The analysis was performed on the most highly expressed transcript from each gene, calculated from total RNA-seq coverage. Cap analysis gene expression (CAGE) was used to update the 5′ UTR on a per sample basis as follows. The highest CAGE peak was selected in a search region from the 3′ most end of the transcript, to the greater of the 5′ most end of the 5′ UTR or 1000 nt upstream of the annotated start codon (If the highest CAGE peak was called downstream of the protein TIS the transcript was excluded for further analysis). Transcripts were excluded from further analysis if they overlapped with the most highly expressed transcript of another gene, or with an annotated non-coding transcript (defined as all Ensembl transcripts with biotypes other than protein_coding).

TISU motif containing transcripts were defined as those with 5′ UTRs of $\leq$ 30 nt in length and a PWM score against the consensus TISU sequence SAASATGGCGGC (where S is C or G) of $>= -12$. TOP motif containing transcripts were defined as those beginning with a C followed by at least 4 T nucleotides. Kozak sequence strength was determined through PWM scores against the zebrafish Kozak matrix taken from Grzegorski et al. (2014). Three genes (ENSDARG00000077330, ENSDARG00000102873, ENSDARG00000089382) contained strong coverage peaks across repetitive regions in their 3′ UTR and were excluded from plots showing densities over 3′ UTRs (Figure 1B).

Read trimming and alignment

RCP-seq reads were trimmed with cutadapt searching for the "AAAAAAAAAA" added to the 3′ of each fragment during library preparation, allowing 1 mismatch and at least 5 nt of overlap. The first 3 nt of each read were removed, reads shorter than 15 nt after trimming were discarded. The remaining reads were aligned to, in order, the PhiX genome, rRNA from the silva database (Quast et al., 2013) (version 119), organism-specific ncRNA as defined by Ensembl (zebrafish GRCz10), and organism-specific

tRNA produced with tRNA-scan SE (Lowe and Eddy, 1997) (default settings). Reads that did not match to any of the above were aligned to the *D. rerio* GRCz10 genome. Total RNA-seq reads were trimmed and aligned to the *D. rerio* GRCz10 genome. Ribo-seq reads were trimmed and aligned to rRNA and ncRNA as above, unaligned reads where then aligned to the *D. rerio* GRCz10 genome. CAGE reads were trimmed and aligned to the *D. rerio* GRCz10 genome. Alignments were performed using tophat2 (Kim et al., 2013) against the *D. rerio* GRCz10 genome and ensembl version 81 gene annotations, reporting up to 20 hits for reads mapping to multiple locations (later filtered with MAPQ, see below).

TCP-seq data from *Saccharomyces cerevisiae* (Archer et al., 2016) (SRA: SRP074093) was processed as above (with the exception of 1st 3nt removal and using STAR as aligner with default parameters) to the R64_1_1 genome with Ensembl version 79 gene annotations. *S. cerevisiae* 5′ UTR were defined with CAGE (Wery et al., 2016) (GEO: GSE69384).

RCP-seq fractionation

The RCP-seq sedimentation fractions corresponding to small subunits and 80S complexes were determined from sequencing all sedimentation fractions. Based on coverage profiles with 20 fractions, fractions 12-14 were determined to contain small subunit fragments. Fractions 18-19 were determined to contain 80S complex fragments. RCP-seq small subunit counts in this study are reported from a pooled set of all relevant fractions, unless otherwise stated. In order to determine the maximum length of small subunit RCP-seq reads, one sample was re-sequenced for 150 cycles (shown in Figures S2A–S2D), as opposed to the 75 cycles used for the rest of the samples.

Further read processing

Inappropriately truncated RCP-seq reads after polyA trimming were updated by extending alignments where trimmed regions exactly matched transcriptomic references. Unusually high RCP-seq coverage peaks were removed from transcripts by filtering out reads with the same 5′ and 3′ coordinates that were present at > = 200 times the average coverage of each transcript. A subset of small subunit reads (~25-35 nt in length), were observed to show 3nt periodicy over the CDS region (Figure 1D, lower left, CDS region). This periodicity is indicative of translation, but it was not clear if these reads represent the leaky scanning of 43S PICs, queued behind translating ribosomes, or footprints of translating complexes, where possibly the 60S subunit has become detached, before sedimentation. As such, reads corresponding to the length of typical translating fragments (length 25-35 nt) were considered ambiguous and removed from RCP-seq small subunit libraries. Conversely the RCP-seq 80S libraries used in Figure 1B, were filtered to only include the lengths of typical translating fragments (length 25-35 nt), akin to ribo-seq libraries. RCP-seq reads that mapped to positions overlapping the last 10 nt of each transcript were discarded, to remove 3′ peaks that likely result from polyA selection during the library preparation.

Read counting

The relative enrichment of tRNA species between small subunit and 80S complex fractions was calculated with edgeR (Robinson et al, 2010) using a binomial generalized log-linear model and likelihood ratio test. Read counts were summed per tRNA anticodon type. CAGE reads counts were normalized using the power law method from the cageR package (Haberle et al., 2015). The FPKMs (fragments per kilobase, per million mapped reads) of RCP-seq, Ribo-seq and Total RNA-seq reads with a MAPQ > = 10 were calculated for transcript 5′ UTR, CDS, and 3′ UTR regions. Reads that overlapped multiple regions were preferentially assigned to CDS > 5′ UTR > 3′ UTR.

Fragment lengths heatmaps and fragment distribution

Metaplots of RCP-seq footprint distributions were plotted in windows across at the 5′ most end of transcripts (TSS) or around protein TIS, for transcripts with at total RNA-seq FPKM > 10 and 5′ UTRs at least 100 nt long. Fragment counts are assigned to either the 5′ or the 3′ end of the fragment. The heatmaps of counts per fragment length are colored by sum of counts from all transcripts at a given position for a given fragment length. The proportion of coverage window counts are summed for all transcripts at given position. The transcripts of genes ENSDARG00000036180, ENSDARG0000001479 were observed to have strong artifactual peaks caused by premature read trimming in polyA regions upstream of the protein TIS and were removed from the TIS plots in Figure 1D.

For the yeast TCP-seq data, the footprints were similarly plotted as heatmaps for the area around the TSS in Figure S2, using a median based filter to remove transcripts that had extreme peaks in the +2 to +40 region relative to TSS (total of 21 genes). Filter removed transcript if: peak at any position in +2 to +40 > median footprints in transcript per position * 99 quantile rank of transcript's footprints per position + 99.9 quantile rank of the medians of all transcripts.

Scaled coverage meta plots

The coverage, or 5′ counts of RCP-seq and ribo-seq reads with a MAPQ > = 10 was calculated across transcript 5′ UTRs, CDS, and 3′ UTR. Transcripts with 5′ UTRs, CDS, and 3′ UTR greater than a length cutoff (typically 100 nt) were scaled to the length value. Values are displayed as the sum of all selected transcripts, mean normalized values or z-score for all selected transcripts. Counts from each transcript normalized by z-score across the whole transcript allow for comparisons of transcripts across wide expression ranges.

Estimates of 43S PIC loss across 5′ UTRs

The coverage of small subunit footprints in regions proximal to the beginning of the transcript and the protein TIS were used to infer the number of 43S PICs recruited to a transcript and those available for initiation at the protein TIS. These regions were defined based on coverage metaplots (Figure 3A) as +70 to +120 nt relative to the beginning of the transcript and −100 to −50 nt relative to the protein TIS. The ratio of coverage in these regions was used to estimate the loss of 43S PICs across 5′ UTRs for all protein coding transcripts, with RNA-seq FPKM > = 10 and 5′ UTR > = 220 nt in length.

5′ feature plots (Figure 2)

Empirical cumulative density for scanning efficiency and translational efficiency (Figure 2G) were plotted for all protein coding transcripts with > = 10 RNA FPKM. For groups of transcripts starting with; i) an A, C, G or T; or ii) transcripts starting with a TOP motif, transcripts starting without a TOP motif and also starting with a C, and transcripts starting without a TOP motif and also starting with a A, G or T.

uORF plots (Figure 3)

Upstream open reading frame coverage metaplots were produced for all protein coding transcripts with an ATG uORF starting > 100 nt from the 5′ most end of the transcript and the TIS, centering on the first (5′ most) ATG uORF within the transcript.

The proportion of small subunit reads mapping upstream (from the beginning of the transcript to the uORF start codon) or downstream of the uORF (uORF start codon to the protein TIS) were calculated for all ATG, CTG and GTG uORFs, starting > = 50 nt from the beginning of the transcript and the protein TIS, in transcripts with RNA FPKM > = 1, stratified by uORF start codon, and Kozak strength quantile. Similarly the proportion of small subunit reads mapping upstream (from the beginning of the transcript to the uORF start codon) or downstream of the uORF stop codon to the protein TIS were calculated for all ATG, CTG and GTG uORFs starting > = 50 nt from the beginning of the transcript and > = 50 nt from uORF stop codon to the protein TIS, in transcripts with > = 1 RNA FPKM, stratified by uORF start codon, and uORF stop codon.

Scanning efficiency, translational efficiency and 5′ UTR translational efficiency were calculated for all protein coding transcripts with 5′ UTRs > = 100 nt in length and RNA FPKM > = 1, that contained 0, 1, 2, 3 or 4 ATG uORFs. Statistical significance was reported between transcripts containing 0 ATG uORFs versus those containing 1-4 ATG uORFs as: W = 112420000, p value < $2.2x10^{-16}$ for 5′ UTR RCP-seq 40S FPKM, W = 138210000 p value < $2.2x10^{-16}$ for translational efficiency and W = 73953000, p value < $2.2x10^{-16}$ for 5′ UTR translational efficiency.

The ratio of RCP-seq small subunits to ribo-seq 80S complexes over the uORF stop codon, per uORF stop codon, and the translational efficiency (Figures 3I and 3J) were calculated for all translated ATG, CTG or GTG uORFs starting > = 50 nt from the beginning of the transcript and > = 50 nt from uORF stop codon to the protein TIS, in transcripts with > = 1 RNA FPKM. Translated uORFs were defined as those with a normalized uORF coverage of > = 10. Normalized uORF coverage is calculated as the coverage of ribo-seq 80S complexes directly over the uORF, divided by uORF length and ribo-seq 80S library size, normalized by the CDS RNA FPKM of the transcript containing the uORF. Statistical significance was reported as the ratio of small subunits to 80S over the stop codon of transcripts containing TAA uORFs versus TGA uORFs (W = 57466000, p value = $2.46x10^{-14}$), and the translational efficiency of transcripts containing TAA uORFs versus TGA uORFs (W = 73381000, p value = $1.703x10^{-14}$). Note that for Figures 3I and 3J, a pseudocount of 0.1 was added to each value in order to plot all values when log transforming (including those that are 0).

In all uORF plots, uORFs that could be considered extensions to the protein coding region, i.e., those inframe with the CDS but without an inframe stop codon before the CDS TIS, were excluded.

Initiation plots (Figure 4)

Initiation rates were calculated as the ratio of RCP-seq small subunits in the 5′ UTR to RCP-seq 80S complexes in the CDS, for all protein coding transcript sequences with RNA FPKM > = 10 and 5′ UTRs > = 100 nt in length. Initiation rates for transcripts were then grouped by each nucleotide, per position in a −4 to +5 window surrounding the protein coding start codon, excluding the start codon. Median initiation rate was calculated for nucleotides that were present more than 1000 times. Initiation rates were also calculated over continuous sequence contexts, using a smaller window of −4 to +3 nucleotides surrounding protein coding start sites, calculating the mean initiation rate for all sequences present > = 20 times in the selected transcripts. This smaller window was used in order to increase the number of transcripts present in each sequence bin.

Statistical testing and plotting

Significance testing was performed in R using the Wilcoxon rank sum test with continuity correction. The metrics used to investigate relationships between small subunits and 80S footprints are summarized below. Boxplot upper whiskers extend from the 1st quartile to the largest value no further than 1.5 times the distance between the first and third quartile, from the first quartile. Boxplot lower whiskers extend from the third quartile to the smallest value no further than 1.5 times the distance between the first and third quartile, from the third quartile.

Formulas used in this study

uORF small subunit (SSU) consumption rate for start codons:

$$\frac{\text{SSU coverage upstream of uORF start/length of upstream region}}{\textit{SSU coverage downstream of UORF start/ length of downstream region}}$$

uORF small subunit consumption rate for stop codons:

$$\frac{\text{SSU coverage upstream of uORF start/ length of upstream region}}{\textit{SSU coverage downstream of uORF start/ length of downstream region}}$$

uORF stop codon recognition rate:

$$\frac{\textit{SSU count at uORF stop codon}}{\textit{80S count at uORF stop codon}}$$

uORF readthrough rate:

$$\frac{\text{80S count downstream of uORF stop codon/length of downstream region}}{\textit{CDS RNA FPKM}}$$

5′ UTR translational efficiency:

$$\frac{5'\textit{UTR 80S FPKM}}{5'\textit{UTR RNA FPKM}}$$

Scanning efficiency (SE):

$$\frac{5'\textit{UTR SSU FPKM}}{\textit{CDS RNA FPKM}}$$

Initiation rate (IR):

$$\frac{\textit{CDS 80S FPKM}}{5'\textit{UTR SSU FPKM}}$$

Translational efficiency (TE):

$$\frac{\textit{CDS 80S FPKM}}{\textit{CDS RNA FPKM}}$$

Scanning efficiency and initiation rate are equivalent to translational efficiency:

$$\frac{5'\textit{UTR SSU FPKM}}{\textit{CDS RNA FPKM}} \times \frac{\textit{CDS 80S FPKM}}{5'\textit{UTR SSU FPKM}} = \frac{\textit{CDS 80S FPKM}}{\textit{CDS RNA FPKM}}$$

$$\textit{SE} \times \textit{IR} = \textit{TE}$$