

Translation initiation factor eIF4G1 preferentially binds yeast transcript leaders containing conserved oligo-uridine motifs

BORIS ZINSHTEYN,¹ MARIA F. ROJAS-DURAN,² and WENDY V. GILBERT²

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

ABSTRACT

Translational control of gene expression plays essential roles in cellular stress responses and organismal development by enabling rapid, selective, and localized control of protein production. Translational regulation depends on context-dependent differences in the protein output of mRNAs, but the key mRNA features that distinguish efficiently translated mRNAs are largely unknown. Here, we comprehensively determined the RNA-binding preferences of the eukaryotic initiation factor 4G (eIF4G) to assess whether this core translation initiation factor has intrinsic sequence preferences that may contribute to preferential translation of specific mRNAs. We identified a simple RNA sequence motif—oligo-uridine—that mediates high-affinity binding to eIF4G *in vitro*. Oligo(U) motifs occur naturally in the transcript leader (TL) of hundreds of yeast genes, and mRNAs with unstructured oligo(U) motifs were enriched in immunoprecipitations against eIF4G. Ribosome profiling following depletion of eIF4G *in vivo* showed preferentially reduced translation of mRNAs with long TLs, including those that contain oligo(U). Finally, TL oligo(U) elements are enriched in genes with regulatory roles and are conserved between yeast species, consistent with an important cellular function. Taken together, our results demonstrate RNA sequence preferences for a general initiation factor, which cells potentially exploit for translational control of specific mRNAs.

Keywords: RNA binding; eIF4G; ribosome footprint profiling; transcript leaders; translation initiation

INTRODUCTION

Eukaryotic cells in different tissues and developmental stages require different protein complements to achieve their form and function, despite having identical genomes. The contribution of translational control to differences in protein production is well established in specific cases, yet the general mechanisms governing the rate at which protein is produced from a particular species of mRNA are poorly understood. Transcript leaders (TLs; also known as 5'-untranslated regions) directly contact the translation initiation machinery and can strongly influence the rate of translation (Hinnebusch et al. 2016). Different TL sequences from yeast are sufficient to confer a 1000-fold range of translational efficiencies (TEs) both *in vivo* and in lysates (Rojas-Duran and Gilbert 2012), but the important functional elements within these TLs are mostly unknown. Several well-characterized TL features can repress translation to varying degrees. Inhibitory features include upstream open reading frames, which divert ribosomes from the main open reading frame of an mRNA

(Hinnebusch 2005; Arribere and Gilbert 2013; Pelechano et al. 2013); stable RNA secondary structures, which block or impede ribosome recruitment or scanning (Kozak 1990; Ding et al. 2014; Weinberg et al. 2016); and target sites for certain RNA-binding proteins, which lead to translational repression and mRNA decay of specific sets of messages (Hentze et al. 2004; Beckmann et al. 2005). In contrast, relatively little is known about TL features that can act as general translational enhancers, although such elements could contribute substantially to the wide range of TEs observed in eukaryotic cells (Gilbert 2010; Shatsky et al. 2014; Lee et al. 2015).

Eukaryotic translation initiation factor 4G (eIF4G) is a prime candidate to mediate the activity of translational enhancer sequences. eIF4G is the scaffold subunit of eIF4F, which initially recognizes translation-competent mRNAs (Jackson et al. 2010; Hinnebusch and Lorsch 2012). eIF4G bridges the cap-binding activity of eIF4E, the helicase activity of eIF4A, and the poly-A binding activity of PABP/Pab1, thereby forming the translational “closed loop” complex (for review, see Thompson and Gilbert 2017). eIF4G contains

¹Present address: Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

²Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06510, USA

Corresponding author: wendy.gilbert@yale.edu

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.062059.117>.

© 2017 Zinshteyn et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://majournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

three RNA-binding domains that directly interact with mRNA and are essential for yeast growth (Berset et al. 2003; Park et al. 2011a), although specific functional interactions between eIF4G and cellular mRNAs have not been characterized. In the context of some viral mRNAs, direct RNA binding by eIF4G is both necessary and sufficient to confer efficient translation, even in the absence of a cap or a poly(A) tail (Pestova et al. 1996a,b). High-affinity binding of eIF4G could facilitate the translation of those cellular mRNAs most dependent on the activity of the eIF4F complex, such as those with long or structured TLs (Sen et al. 2015, 2016), or allow coordinated translational control under conditions that limit eIF4F activity including viral infection (Castelló et al. 2006, 2011), nutrient depletion (Thoreen et al. 2012), and heat shock (Cuesta et al. 2000).

Here, we have comprehensively determined the RNA-binding preferences of eIF4G1 from *Saccharomyces cerevisiae* and tested the effects of these preferences on translation in vitro and in vivo. Using RNA Bind-n-Seq (RBNS), a quantitative high-throughput technique to measure RNA-binding affinities in vitro (Lambert et al. 2014), we tested approximately 87,380 distinct RNA motifs and showed that recombinant eIF4G1 preferentially binds to unstructured RNA sequences containing oligo-uridine (U). Consistent with this result, inserting U₁₀ into an unstructured RNA increased binding to eIF4G1 by 20-fold in electrophoretic mobility shift assays (EMSAs). Hundreds of yeast TLs contain oligo(U) sequences, which are evolutionarily conserved among budding yeast species and enriched in genes with regulatory roles. TL oligo(U) elements enhanced translation in vitro for four of eight TLs tested, although this effect could be mediated by a factor other than eIF4G. In vivo, we found that eIF4G preferentially bound mRNAs with unstructured oligo(U) motifs. However, TL length, and not the presence or absence of oligo(U) motifs, was the best predictor of translational sensitivity to eIF4G depletion in vivo. Taken together, our results show that short TL RNA sequences can act as translational enhancers and suggest that the intrinsic RNA-binding preferences of core translation factors may contribute to gene-specific differences in translation efficiency or regulation.

RESULTS

eIF4G1 binds oligo(U) sequences with high affinity

We used RBNS (Lambert et al. 2014) to determine the RNA-binding specificity of eIF4G. This competitive in vitro binding assay consists of mixing randomized RNA libraries with different concentrations of protein, and sequencing the bound RNA (Fig. 1A). RBNS reports directly and quantitatively on the innate RNA-binding preferences of a protein of interest. It is neither biased by the sequences present in a transcriptome, nor affected by processes that could indirectly alter protein–RNA association in vivo. We performed RBNS with recombinant eIF4G1 (Supplemental Fig. S1A,B), the

more abundant and better characterized of two *S. cerevisiae* paralogs of eIF4G (Goyer et al. 1993; Ghaemmaghami et al. 2003), and a library of random 20mer RNA.

Analysis of the RNA sequences bound by eIF4G1 revealed a strong enrichment for sequences containing oligo(U), and a weaker enrichment for sequences containing five or more consecutive Gs. As expected, these enrichments were concentration-dependent; they peaked at an intermediate concentration (320 nM) and then decreased at higher protein concentrations where binding ceases to be competitive and all sequences are bound to a similar extent (Fig. 1B). The computation of enrichment values used here could lead to an artifactual enrichment of short homopolymer sequences if a longer homopolymer is enriched (Supplemental Fig. S1C). To account for this, the enrichment was recomputed, comparing only complete homopolymers of the same length. The enrichment for oligo(U) persisted, and it increased with the length of the homopolymer, indicating that eIF4G preferentially binds to longer oligo(U) stretches (Supplemental Fig. S1D). Strikingly, the enrichment for oligo(U) was similar in magnitude to that observed previously for known binding sites of highly sequence-specific mammalian splicing factors (Supplemental Fig. S1E).

EMSAs on homopolymeric U and A sequences confirmed the preference of eIF4G1 for binding oligo(U). eIF4G1 bound to U₄₀ with a 17-fold tighter affinity than A₄₀ (Fig. 1D), confirming the RNA sequence preference observed by RBNS. To determine if short stretches of the homopolymers were sufficient for tight binding in the context of a longer sequence, U₁₀ or G₁₀ were embedded in a sequence of 25 CA repeats and tested by EMSA with eIF4G1. The addition of G₁₀ increased binding moderately compared with the poly (CA) control, and the addition of U₁₀ resulted in tight binding comparable to the U₄₀ homopolymer (Fig. 1E), recapitulating the results from RBNS and confirming oligo(U) as the preferred binding site for eIF4G1 in vitro.

Oligo(U) sequences are conserved and enriched in genes with regulatory functions

Searching the *S. cerevisiae* transcriptome for TLs with continuous stretches of U's identified 380 genes containing a total of 463 oligo(U) motifs ≥ 7 -nt-long within their annotated TL (Supplemental Table S1; Pelechano et al. 2013). We assessed the conservation of oligo(U) sequences for each U₇ motif in yeast TLs by computing branch length scores (BLSs), a measure of the evolutionary distance over which a sequence is exactly conserved (Fig. 2A; Friedman et al. 2008). BLSs for oligo(U) motifs were compared with a set of control 7mers, which differ from U₇ by up to three U-to-A substitutions. Only U-to-A substitutions were used to maintain the A+U content in the motif, as G and C nucleotides have substantially lower representation in the yeast transcriptome. For each BLS, a significantly higher fraction of oligo(U) elements was conserved compared with the control set (Fig. 2B), and

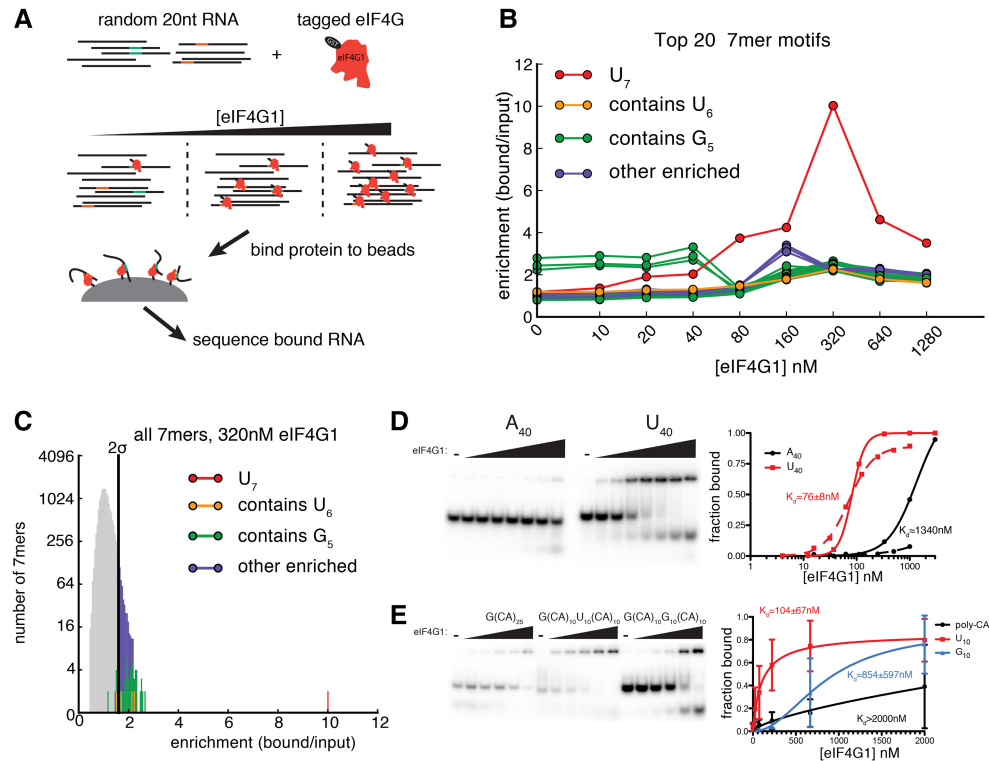


FIGURE 1. eIF4G1 preferentially binds to oligo(U) sequences. (A) RBNS method summary. GST-tagged eIF4G, at several concentrations, was incubated with a pool of random 20mer RNA. Protein and bound RNA were isolated with glutathione-conjugated magnetic beads, and the bound RNA was deep-sequenced to identify binding motifs. (B) RBNS enrichment scores at all concentrations of eIF4G1, for the twenty 7-nt sequences that were most enriched at 320 nM eIF4G1. Enrichment values are computed as the frequency of a motif in the bound library divided by its frequency in the input library. (C) Log-scaled histogram of RBNS enrichment values at 320 nM eIF4G1 for all 7-nt sequences. Vertical black line indicates 2 SDs above the mean of the distribution. (D) EMSA to measure binding affinity of eIF4G1 to homopolymer sequences. (Left) Representative gel. Binding reactions were performed with twofold dilutions of eIF4G1 ranging from 15 to 1000 nM. The fast-migrating band seen in some lanes is consistent with an RNA decay product, likely a mononucleotide. (Right) Quantification of homopolymer-binding experiments, in which solid and dotted lines indicate the two separate replicates. K_d s are mean \pm SD from two independent replicates. A_{40} binding was too weak to yield a reliable fit in one replicate. (E) EMSA on model TL sequences. (Left) Representative gel. Binding reactions were performed with threefold dilutions of eIF4G1 ranging from 25 to 2000 nM. (Right) Quantification of binding experiments. Each data point is mean \pm SD from two (G_{10}) or three (CA , U_{10}) independent replicates.

43 U_7 motifs were completely conserved between five yeast species that shared their last common ancestor ~ 10 million years ago (Supplemental Table S1; Budovskaya et al. 2005). The conservation of oligo(U) sequences in TLs implies that these elements play a physiologically relevant role subject to selective pressure. Gene ontology analysis (Balakrishnan et al. 2012) on genes with TLs containing U_7 revealed an enrichment for genes with regulatory functions and for genes whose protein products localize to the cell periphery or cellular bud (Fig. 2C,D). Taken together, these results show that eIF4G1-binding motifs are evolutionarily conserved and raise the possibility that yeast cells could exploit mRNA affinity for eIF4G to tune the translation of genes necessary for specific adaptive responses.

Oligo(U) mRNAs preferentially copurify with eIF4G from yeast

In vitro binding preferences do not necessarily reflect in vivo association. To examine whether oligo(U) motifs recruit

eIF4G in vivo, we analyzed transcriptome-wide RNA immunoprecipitation and RNA sequencing (RIP-seq) data for several yeast translation initiation factors (Costello et al. 2015). Oligo(U) mRNAs were preferentially associated with both paralogs of eIF4G, when compared with all genes (median enrichment increased 1.13- and 1.11-fold for eIF4G1 and eIF4G2, respectively), and enrichment was increased for conserved oligo(U) sequences ($BLS > 1$) (Fig. 3). As mRNAs containing oligo(U) had longer TL lengths than most mRNAs (Supplemental Fig. S2B), we compared the enrichments of oligo(U) mRNAs with a set of mRNAs with a similar distribution of transcript-leader lengths. This group was also enriched in eIF4G IPs, indicating that association of oligo(U) mRNAs with eIF4G could alternatively be due to their longer than average TL lengths. Notably, this group displays an unexpected set of associations with translation initiation factors. Although eIF4E and eIF4G enrichment values were globally well correlated (Pearson $r = 0.74$ for eIF4G1 versus eIF4E, 0.72 for eIF4G2 versus eIF4E) (Supplemental Fig. S2A), the oligo(U) mRNAs were not enriched in IPs

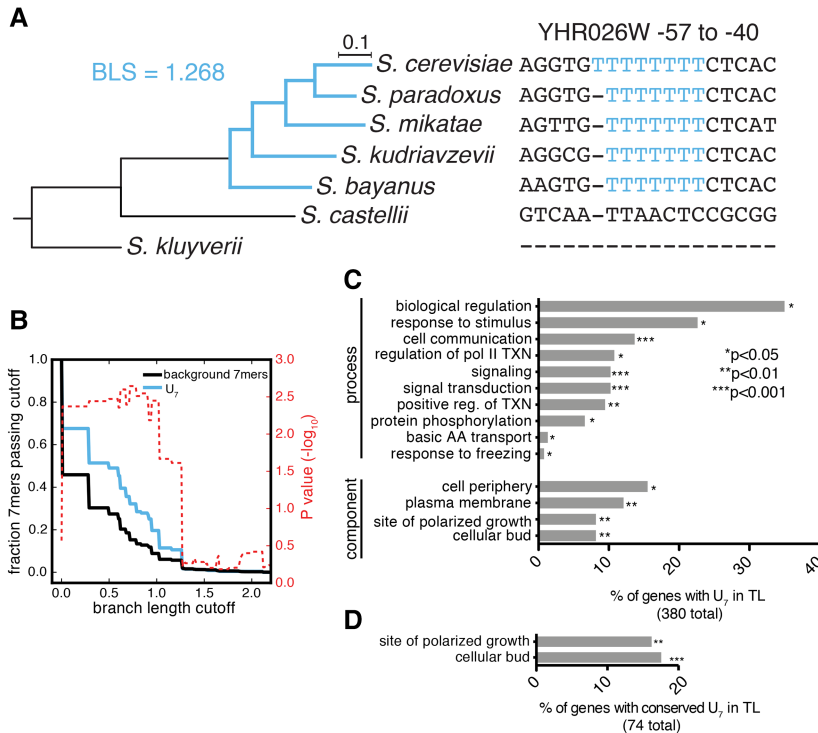


FIGURE 2. Oligo(U) sequences in TLs are conserved, and enriched in genes with regulatory functions. (A) Example of conservation scoring for one oligo(U) sequence. The branch length score (BLS) is the sum of the horizontal branch lengths connecting *S. cerevisiae* to all other species that contain a U_7 at the aligned position, as indicated in blue. BLSs increase with conservation over increasing evolutionary distance. Tree and branch lengths were downloaded from the UCSC genome browser. (B) Fraction of U_7 motifs or control 7mers in TLs passing each branch length cutoff. P values for the conservation of U_7 motifs are the probability of a background motif having an equal or greater conservation fraction than U_7 (Supplemental Methods). The P values decline sharply above a cutoff of 1.268, due to the lack of overall conservation in TLs between *Saccharomyces castellii* and *Saccharomyces kluyverii* and the other five yeast species. (C) Percent of genes with U_7 in their TL that belong to each significantly enriched gene ontology category. (TXN) transcription, (AA) amino acid. (D) Percent of genes with conserved U_7 (BLS > 1.0 in B) in their TL that belong to each significantly enriched gene ontology component category.

against eIF4E and were slightly depleted in IPs against cytoplasmic poly(A)-binding protein (Pab1), suggesting that some of these eIF4G–mRNA interactions happen outside the context of the canonical closed loop complex.

Consistent with the possibility that oligo(U) motifs promote stable binding to eIF4G in the absence of eIF4E, these oligo(U) mRNAs were also enriched in IPs against the eIF4E-binding protein (4E-BP) EAP1 (median enrichment increased 1.20-fold) (Supplemental Fig. S2C). 4E-BPs repress translation by competing with eIF4G for interaction with eIF4E, thereby excluding eIF4G from 4E-BP-bound mRNAs (Haghighat et al. 1995; Richter and Sonenberg 2005; Cridge et al. 2010). The observed association of oligo(U) mRNAs with both eIF4G and EAP1 could reflect distinct pools of these mRNAs bound to either factor. Alternatively, oligo(U) motifs or some function of TL length may allow eIF4E-independent recruitment of eIF4G to an EAP1-bound mRNA, thereby allowing these mRNAs to escape translational repression by EAP1.

eIF4G1 preferentially binds unstructured oligo(U) sequences

The known high-affinity binding sites for mammalian eIF4G in viral mRNAs are highly structured (Pestova et al. 1996a; Jackson et al. 2010), raising the possibility that yeast eIF4G1 has structural preferences as well. We determined the structural contexts of the sequence motifs bound by eIF4G1 by computing single-nucleotide pairing probabilities for each 20mer sequencing read containing a U_7 sequence (Lorenz et al. 2011). For each eIF4G concentration, the pairing probabilities in a window centered around bound U_7 motifs were averaged into pairing probability profiles and divided by the equivalent profile for the input sequences to determine the pairing preferences of the bound motifs (Fig. 4A). The sequences bound by eIF4G1 showed a sharp drop in pairing probability at the bound oligo(U) motif (Fig. 4B) compared with the input and no-protein controls. The preference for single-strandedness showed the same concentration dependence as the RBNS enrichment for oligo(U), with the strongest enrichment for oligo(U) coinciding with the strongest preference for unpaired sequences at 320 nM (compare Fig. 4B with Fig. 1B). Thus, eIF4G1 preferentially binds oligo(U) in unstructured contexts.

mRNAs with unstructured oligo(U) motifs were also more enriched in pull-

downs of eIF4G from yeast cells. We used transcriptome-wide nuclease reactivity data (Kertesz et al. 2010) to group oligo(U) sequences in yeast TLs based on their average RNA secondary structure propensity (PARS score) and repeated the RIP-seq enrichment analysis (Fig. 4C; Supplemental Fig. S2D). Unstructured sequences were enriched significantly more than structured ones in RIP-seq for eIF4G1 (median change 1.08-fold, $P = 0.016$) and eIF4G2 (median change 1.19-fold, $P = 0.002$), whereas the group of structured oligo(U) TLs was not significantly enriched against the set of all TLs covered in the PARS data. This difference is not a result of TL length biases. The sets of structured and unstructured TLs have similar lengths, and the length-matched controls for these two gene sets had similar enrichments in the eIF4G RIP-seq (Fig. 4C). Furthermore, the difference in IP enrichment between structured and unstructured oligo(U) was not seen for other factors. This indicates that, in agreement with in vitro results, a base-paired oligo(U) motif is less likely to bind to eIF4G in vivo.

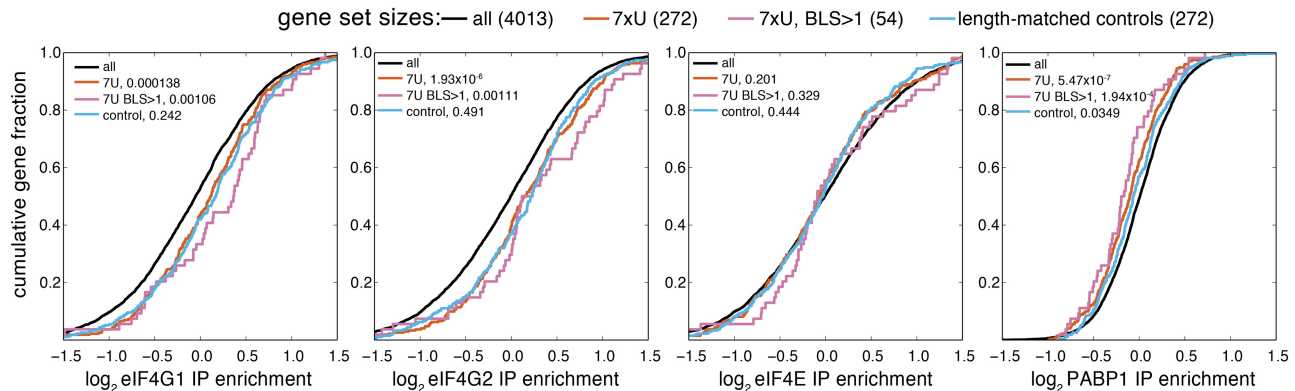


FIGURE 3. Oligo(U) mRNAs preferentially copurify with eIF4G from yeast. Cumulative distributions of RIP-seq enrichments from (Costello et al. 2015), for genes containing oligo(U) of length 7 or greater. *P* values for 7xU compare the indicated gene set with the set of all genes with RIP-seq data, as computed by the Mann–Whitney *U*-test. The *P* values for the length-matched controls compare the control gene set with the set of all 7xU genes in the RIP-seq data.

Oligo(U) motifs are correlated with eIF4G-dependent translation in vivo

We hypothesized that eIF4G-binding motifs would promote eIF4G-dependent ribosome recruitment. To determine the impact of eIF4G and oligo(U) sequences on translation in vivo, we performed ribosome footprint profiling (Ribo-seq) and total mRNA sequencing (Ingolia et al. 2009) on wild-type (WT) controls and mutant yeast in which eIF4G levels were reduced using a temperature-sensitive degron allele (4G-ts) (Fig. 5A; Park et al. 2011b). eIF4G levels were greatly reduced even in permissive conditions, as previously reported (Park et al. 2011b), and eIF4G was undetectable or barely detectable within 2 h of shifting the 4G-ts strain to the non-permissive condition (Fig. 5B; Supplemental Fig. S3A). Bulk polysome levels were moderately reduced (twofold) in the 4G-ts strain under permissive conditions, but collapsed (to 34-fold less than unshifted WT) after further eIF4G depletion (Fig. 5C; Supplemental Fig. S3B), indicating a severe global reduction in translation initiation.

We examined gene-specific differences in basal translation activity and sensitivity to eIF4G levels by computing translational efficiency (TE), defined as the ratio of ribosome footprint read density to RNA-seq read density for a particular mRNA (Ingolia et al. 2009). As oligo(U) mRNAs have longer than average TLs, which are correlated with reduced TE (Weinberg et al. 2016), we chose a set of mRNAs with a similar distribution of TL lengths as a control (Supplemental Fig. S3C). In WT cells, mRNAs with oligo(U) in their TLs had reduced TE when compared with the full set of genes, but only slightly less than the length-matched controls (Fig. 5D). This indicates that these genes are, on average, translationally disadvantaged, but primarily because of their longer TLs. In permissive conditions, oligo(U) mRNAs exhibited decreased TE in the 4G-ts strain compared with WT that could be entirely attributed to their longer TLs (Fig. 5E). Upon further depletion of eIF4G, the TE of oligo(U) mRNAs was

reduced more than the length-matched controls (Fig. 5E), and this effect was larger after longer depletion times (Supplemental Fig. S3D). This reduction in TE was consistent across four of the five experiments, although the effect for oligo(U) mRNAs was only statistically different from the length-matched controls in two of the five depletion experiments (Supplemental Fig. S3D). Although these results are consistent with the model that oligo(U) sequences help to recruit eIF4G to messages that are particularly dependent on its activity, the small effect sizes make it difficult to tease apart the contributions of TL length versus TL sequence in the translational response to eIF4G depletion.

We found that eIF4G depletion led to a marked decrease in TE for mRNAs with TLs longer than 135 nt (Supplemental Fig. S4A), consistent with the role of eIF4G in recruiting and activating eIF4A to unwind RNA structure and ensure processivity of the scanning pre-initiation complex on long and structured TLs (García-García et al. 2015; Sen et al. 2015; Gao et al. 2016). The same effect can be seen in published Ribo-seq data for strains depleted of eIF4A and its activator eIF4B (Supplemental Fig. S4B; Sen et al. 2015, 2016). We note that our findings contradict previous TE measurements conducted on the same eIF4G depletion strain using polysome microarrays (Supplemental Fig. S4C; Park et al. 2011b), which showed the opposite relationship between TE change upon eIF4G depletion and TL length. This discrepancy may be a result of the technical limitations of the polysome microarray assay, which assessed translation changes by monitoring the fraction of a message that sediments with an arbitrary heavy polysome fraction, a measure that is biased by coding sequence length (Arava et al. 2003) and thus skews translational changes for certain classes of genes.

Oligo(U) motifs promote translation in vitro

The results from ribosome profiling suggest that oligo(U) motifs promote eIF4G-dependent ribosome recruitment,

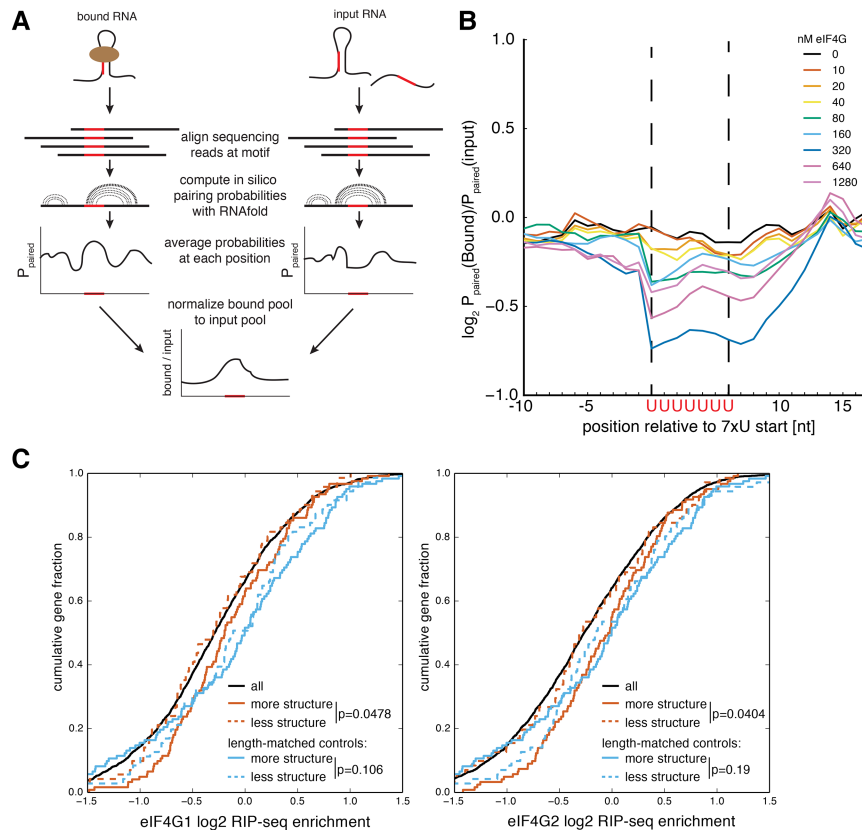


FIGURE 4. eIF4G1 binds oligo(U) in unstructured contexts. (A) Overview of RNA structure analysis. In this cartoon example, darker arcs represent larger pairing probabilities, and faint or absent arcs represent low probability pairings. See the main text for details. (B) Pairing probabilities for oligo(U) and surrounding sequence for each eIF4G1 concentration, normalized to the input sequences. The sequences are aligned at the 5' U of the homopolymer, but many extend more than 7 nt, which explains the sharp drop in pairing probability at the 5'-end of the sequence and the gradual rise at the 3'-end. (C) Cumulative distributions of RIP-seq enrichments from Costello et al. (2015), for genes containing oligo(U) of length 7 or greater, grouped by the average PARS structural propensity of the oligo(U) from Kertesz et al. (2010). “Less structured” oligo(U) have a mean PARS score from -2.0 to <0.1 , whereas “more structured” oligo(U) have a mean PARS score from 0.1 to 3.5 . Length-matched control sets have a similar distribution of TL lengths to the corresponding gene set. *P* values are for comparison of more structured and less structured gene sets (or their length-matched controls) to each other, as computed by the Mann–Whitney *U*-test.

but these results are complicated by the longer TL lengths of oligo(U) mRNAs. To directly test the impact of oligo(U) motifs on translation, we chose eight TLs of varying lengths containing oligo(U) motifs. These TLs and their oligo(U) mutant counterparts, plus partial coding sequences (12–87 nt), were cloned in-frame with the nanoLuc open reading frame, in vitro-transcribed, m⁷G-capped, and translated in yeast lysates. Of the eight TLs tested, four exhibited decreased luciferase production per mRNA when the oligo(U) was mutated (Fig. 6A; Supplemental Fig. S5), whereas one exhibited increased luciferase production per mRNA (Supplemental Fig. S5). We next compared the activity of these TL pairs in translation extracts genetically depleted of eIF4G. Although the overall level of translation was severely reduced in eIF4G-depleted extracts, the difference between the oligo(U)

and mutant mRNAs persisted (Fig. 6B). This result suggests that an mRNA feature or protein factor other than eIF4G is responsible for the difference in lysate translation between these mRNAs, or that the small amount of residual eIF4G in these depleted extracts is sufficient to confer an advantage on the oligo(U) mRNAs. Overall, these results demonstrate the ability of short oligo(U) sequences to stimulate protein production. However, the exact mechanism by which these sequences enhance ribosome recruitment remains unclear.

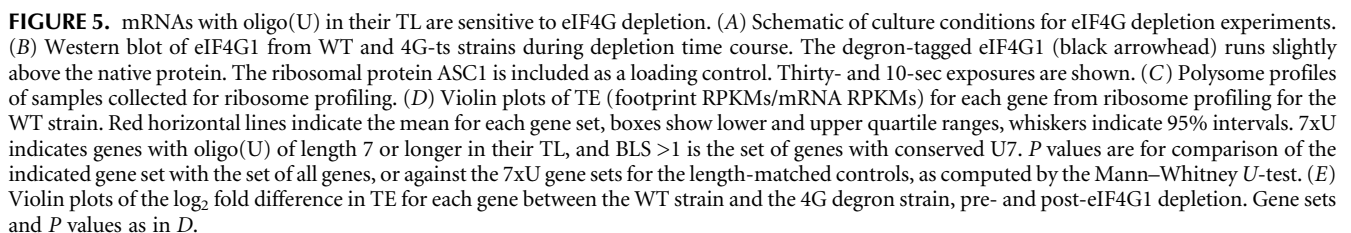
DISCUSSION

The RNA-binding activity of eIF4G1 is essential for yeast growth (Berset et al. 2003), yet the role of this activity in translation initiation is not currently understood. Although previous work suggested that eIF4G1 binds RNA non-specifically (Berset et al. 2003), certain viral mRNAs rely on high-affinity binding to eIF4G for their translation (Pestova et al. 1996a), raising the possibility that specific cellular mRNAs might also exploit the intrinsic RNA-binding preferences of eIF4G. Here, we have systematically determined the RNA-binding preferences of eIF4G1 in vitro and discovered a strong and specific affinity of eIF4G1 for oligo(U) sequences. We find that oligo(U) elements are present in the TLs of hundreds of yeast mRNAs and are frequently conserved. Under conditions of limiting eIF4G activity, mRNAs with oligo(U) in their TLs are poorly translated in vivo, as are other mRNAs

with long TLs. We further showed that oligo(U) motifs directly enhance translation initiation in lysates, but this effect seems likely to be mediated by a factor other than eIF4G.

The translational effects of oligo(U) sequences in yeast TLs vary and can be modified by other sequence features. Our RBNS data indicate that sequestration of oligo(U) sequences in RNA secondary structures inhibits eIF4G binding. This may explain some of the spread in the data for IP enrichment of oligo(U) mRNAs with eIF4G and TE changes upon eIF4G depletion. In addition, upstream start codons can still divert ribosomes from translating the main open reading frame, thus reducing TE even if ribosome recruitment is enhanced by an oligo(U) motif.

Although the conditions used here to deplete eIF4G were genetic rather than environmental, many cellular stress



The portions of eIF4G1 responsible for specific binding to unstructured oligo(U) motifs are currently unknown. Although eIF4G was initially assigned an RNA recognition motif based on sequence (Goyer et al. 1993), this domain

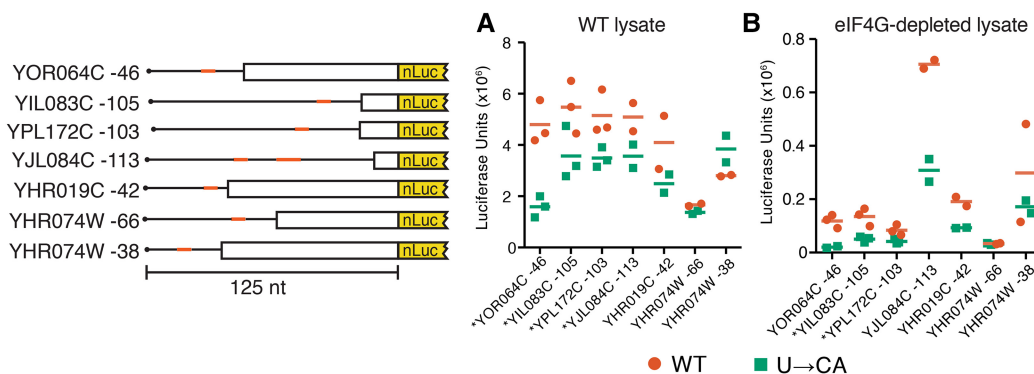


FIGURE 6. Oligo(U) motifs stimulate translation in vitro. NanoLuc activity from control (A) and eIF4G-depleted (B) in vitro translation extracts programmed with capped and polyadenylated NanoLuc mRNA bearing the indicated TL. Each point represents an independent experiment and is the mean of three technical replicates. Horizontal lines indicate the mean of the three independent replicates. For TL construct diagrams, TLs are drawn as lines, with $U \geq 7$ indicated in vermillion. Unfilled rectangles represent the N-terminal portion of the native coding sequence for the indicated gene. *Statistically significant ($P < 0.05$ by two-tailed Student's *t*-test) differences between WT and mutant NanoLuc activity.

was found not to bind RNA (Berset et al. 2003), and the empirically determined RNA-binding domains of eIF4G1 do not resemble any well-defined RNA-binding motifs at the sequence level (Burd and Dreyfuss 1994). However, two of these domains are arginine-serine rich (RS), resembling the unstructured RS domains commonly found in splicing factors. Although splicing factor RS domains are thought to mediate protein–protein interactions, not RNA binding (Graveley 2000), in eIF4G, these domains appear to mediate interactions with both RNA and other proteins (Berset et al. 2003; Singh et al. 2012). It will be interesting to determine which of the domains of eIF4G confer RNA-binding specificity and whether the many phosphorylation sites found within these domains (Albuquerque et al. 2008; Holt et al. 2009; Swaney et al. 2013) regulate translation of oligo(U) mRNAs.

The factors that influence the rate of translation initiation are just beginning to be described, and, until recently, the intrinsic RNA-binding preferences of the core initiation factors have largely been ignored. Several other translation initiation factors, including eIF4A, eIF4B, and multiple subunits of eIF3, have RNA-binding activity and may also have differential affinities for mRNAs. Indeed, human eIF3 was recently shown to preferentially associate with specific RNA structures in TLs in vivo, with enhancing or repressing effects on translation determined by contextual features that are not yet understood (Lee et al. 2015). Our results suggest that sequence-specific RNA binding by core initiation factors could have widespread effects on basal translation rates and enable coordinated translational control of specific subsets of mRNAs. The approach described here for comprehensive characterization of the RNA-binding preferences of purified factors is broadly applicable. With a complete affinity profile of the translation initiation machinery, we can begin to quantitatively predict the protein output of a transcriptome from the sequence of its mRNAs.

MATERIALS AND METHODS

Purification of yeast eIF4G1

Full-length eIF4G1 fused to N-terminal GST and C-terminal 6xHis tags was purified as described in (Mitchell et al. 2010), with the modifications described in [Supplemental Experimental Procedures](#).

RNA library synthesis

Randomized RNA libraries consisted of 20 nt of randomized RNA sequence, followed by an adaptor sequence used for priming reverse transcription and Illumina library preparation. All RNAs were produced by runoff T7 transcription from DNA oligos ([Supplemental Table S2](#)).

RNA Bind-n-Seq

The binding reactions were modified from Lambert et al. (2014) as follows. One micromolar RNA pool was incubated with concentrations of eIF4G ranging from 0 to 1280 nM for 30' at 22°C, then incubated with MagneGST glutathione beads (Promega) for 10' at 22°C. The supernatant was removed, the beads were washed once, and the bound RNA eluted at 70°C for 10 min. The bound RNA and input pool were reverse-transcribed and processed into Illumina sequencing libraries ([Supplemental Methods](#)).

Electrophoretic mobility shift assays

Approximately 60 femtomoles of 5'-end-labeled RNA was equilibrated at 22°C in 5 μ L of 2xTHEM buffer. eIF4G at twice the desired concentration in 5 μ L of storage buffer was added to the RNA and equilibrated for 30 min at 22°C. Samples were run on a 1xTHEM 7% acrylamide MIDI gel for 40–120 min at 140 volts. Gels were scanned on a Typhoon phosphorimager and bands were quantified using ImageQuant TL software (GE Healthcare). Fraction bound was defined as the ratio of the upper band intensity to the total intensity of upper and lower bands.

Computational methods

RBNS data were analyzed with the tools developed in Lambert et al. (2014). The RBNS enrichment value for a kmer is the fraction of that kmer out of all kmers in the bound library, divided by the same fraction for that kmer in the input library. Sequence conservation analysis was done using the algorithm in Friedman et al. (2008). These tools were supplemented with custom Python scripts for all other analyses. For TL annotations, we used the median TL lengths from Pelechano et al. (2013), except for generating synthetic oligo pools, where multiple TL isoforms were included for some mRNAs, and for PARS structure analysis, where we used the matching annotations from Kertesz et al. (2010). For ribo-seq TE plots, all mRNAs with an AUG codon in their TL were excluded from analysis, to avoid complication from the known regulatory effects of upstream open reading frames. All plots from automated analyses were made with Matplotlib (Hunter 2007). Detailed methods for individual analyses are provided in [Supplemental Experimental Procedures](#). RBNS data were deposited into the NCBI short read archive with accession number SRR1909253. Ribo-seq data were deposited into the gene expression omnibus with accession number GSE87641.

In vitro translation

Yeast translation extracts and uncapped luciferase mRNA were made exactly as described in Rojas-Duran and Gilbert (2012). Strain YWG3, grown to mid-log phase, was used to make lysate for Supplemental Figure S5. For Figure 6, strains YAJ3 and YAJ41 (see eIF4G depletion) were grown in SC-URA and shifted to the nonpermissive condition for 2 h before harvesting. NanoLuc RNAs with various TLs were made by runoff transcription from plasmids (Supplemental Table S3) linearized with Ecl136 II (NEB) after the encoded poly(A)₆₂ tail. Transcripts were m⁷G-capped with recombinant vaccinia capping enzyme (Shuman and Moss 1990).

eIF4G depletion

For eIF4G depletion experiments, cultures of WT yeast YAJ3 (MATa trp1Δ leu2-3,112 ura3-52 gcn2Δ::hisG P_{GALI}-myc-UBR1::TRP1::ubr1, pRS316 <URA3>), or eIF4G depletion strain YAJ41 (MATa trp1Δ leu2-3,112 ura3-52 gcn2Δ::hisG P_{GALI}-myc-UBR1::TRP1::ubr1 tif4632Δ::kanMX6 P_{CUP1}-UBI-R-DHFR^{ts}-HA-tif4631-td::URA3::tif4631) (Park et al. 2011b), were grown at the permissive condition of 2% raffinose + 0.1 mM copper sulfate at 26°C (pre-shift). At $t = -30$ min, galactose was added to 2% to induce expression of the degron tag's cognate ubiquitin ligase. At $t = 0$, bathocuproinedisulfonic acid was added to 1 mM to chelate copper and shut off the CUP1 promoter driving eIF4G transcription, and the culture was shifted to 36°C to activate the degron tag. For YPA experiments, pre-shift, $t = 0$ h, and $t = 2$ h were collected for ribosome profiling. For SC experiments, $t = 2$ h and $t = 8$ h time points were collected.

Ribosome footprint profiling

Ribosome footprint profiling was performed essentially as described in Thompson et al. (2016), with cycloheximide added to 0.1 mg/mL

just prior to cell harvest and lysis. rRNA was subtracted from mRNA libraries using the Ribo-Zero Yeast rRNA Subtraction Kit (Illumina).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Sarah Walker and John Lorsch for providing protocols, plasmids, and advice for eIF4G purification; Nicole Lambert, Alex Robertson, and Chris Burge for providing RBNS methods and analysis algorithms prior to publication; Oksana Sergeeva and the King laboratory for sharing protein purification expertise and equipment; Vikram Agarwal for providing conservation analysis programs and advice; and David Bartel for discussions. Sequencing was performed by the MIT BioMicro Center. This work was supported by grants from the National Institutes of Health (NIH) (GM094303 and GM081399) to W.V.G. B.Z. was supported by the Henry and Francis Keady Rickard fund fellowship and by the NIH Pre-Doctoral Training Grant T32GM007287.

Received May 9, 2017; accepted May 23, 2017.

REFERENCES

- Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H. 2008. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol Cell Proteomics* 7: 1389–1396.
- Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. 2003. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* 100: 3889–3894.
- Arribere JA, Gilbert WV. 2013. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res* 23: 977–987.
- Ashe MP, De Long SK, Sachs AB. 2000. Glucose depletion rapidly inhibits translation initiation in yeast. *Mol Biol Cell* 11: 833–848.
- Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Cherry JM. 2012. YeastMine—an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database (Oxford)* 2012: bar062.
- Beckmann K, Grskovic M, Gebauer F, Hentze MW. 2005. A dual inhibitory mechanism restricts *msl-2* mRNA translation for dosage compensation in *Drosophila*. *Cell* 122: 529–540.
- Berset C, Zurbriggen A, Djafarzadeh S, Altmann M, Trachsel H. 2003. RNA-binding activity of translation initiation factor eIF4G1 from *Saccharomyces cerevisiae*. *RNA* 9: 871–880.
- Budovskaya YV, Stephan JS, Deminoff SJ, Herman PK. 2005. An evolutionary proteomics approach identifies substrates of the cAMP-dependent protein kinase. *Proc Natl Acad Sci* 102: 13933–13938.
- Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265: 615–621.
- Castelló A, Álvarez E, Carrasco L. 2006. Differential cleavage of eIF4GI and eIF4GII in mammalian cells. Effects on translation. *J Biol Chem* 281: 33206–33216.
- Castelló A, Álvarez E, Carrasco L. 2011. The multifaceted poliovirus 2A protease: regulation of gene expression by picornavirus proteases. *J Biomed Biotechnol* 2011: 369648.
- Costello J, Castelli LM, Rowe W, Kershaw CJ, Talavera D, Mohammad-Qureshi SS, Sims PFG, Grant CM, Pavitt GD, Hubbard SJ, et al.

2015. Global mRNA selection mechanisms for translation initiation. *Genome Biol* **16**: 10.
- Cridge AG, Castelli LM, Smirnova JB, Selley JN, Rowe W, Hubbard SJ, McCarthy JEG, Ashe MP, Grant CM, Pavitt GD. 2010. Identifying eIF4E-binding protein translationally-controlled transcripts reveals links to mRNAs bound by specific PUF proteins. *Nucleic Acids Res* **38**: 8039–8050.
- Cuesta R, Laroia G, Schneider RJ. 2000. Chaperone hsp27 inhibits translation during heat shock by binding eIF4G and facilitating dissociation of cap-initiation complexes. *Genes Dev* **14**: 1460–1470.
- De Gregorio E, Preiss T, Hentze MW. 1999. Translation driven by an eIF4G core domain in vivo. *EMBO J* **18**: 4865–4874.
- Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Asmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**: 696–700.
- Friedman RC, Farh KKH, Burge CB, Bartel DP. 2008. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92–105.
- Gao Z, Putnam AA, Bowers HA, Guenther U-P, Ye X, Kindsfater A, Hilliker AK, Jankowsky E. 2016. Coupling between the DEAD-box RNA helicases Ded1p and eIF4A. *eLife* **5**: e16408.
- García-García C, Frieda KL, Feoktistova K, Fraser CS, Block SM. 2015. RNA biochemistry. Factor-dependent processivity in human eIF4A DEAD-box helicase. *Science* **348**: 1486–1488.
- Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425**: 737–741.
- Gilbert WV. 2010. Alternative ways to think about cellular internal ribosome entry. *J Biol Chem* **285**: 29033–29038.
- Gilbert WV, Zhou K, Butler TK, Doudna JA. 2007. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science* **317**: 1224–1227.
- Goyer C, Altmann M, Lee HS, Blanc A, Deshmukh M, Woolford JL Jr, Trachsel H, Sonenberg N. 1993. TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. *Mol Cell Biol* **13**: 4860–4874.
- Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA* **6**: 1197–1211.
- Haghighat A, Mader S, Pause A, Sonenberg N. 1995. Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. *EMBO J* **14**: 5701–5709.
- Hentze MW, Muckenthaler MU, Andrews NC. 2004. Balancing acts: molecular control of mammalian iron metabolism. *Cell* **117**: 285–297.
- Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59**: 407–450.
- Hinnebusch AG, Lorsch JR. 2012. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb Perspect Biol* **4**: a011544.
- Hinnebusch AG, Ivanov IP, Sonenberg N. 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**: 1413–1416.
- Holt LJ, Tuch BB, Villén J, Johnson AD, Gygi SP, Morgan DO. 2009. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**: 1682–1686.
- Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Comput Sci Eng* **9**: 90–95.
- Ibrahim S, Holmes LEA, Ashe MP. 2006. Regulation of translation initiation by the yeast eIF4E binding proteins is required for the pseudohyphal response. *Yeast* **23**: 1075–1088.
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223.
- Jackson RJ, Hellen CU, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**: 113–127.
- Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**: 103–107.
- Kozak M. 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci* **87**: 8301–8305.
- Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. 2014. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* **54**: 887–900.
- Lee ASY, Kranszuch PJ, Cate JHD. 2015. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**: 111–114.
- Liu B, Han Y, Qian SB. 2013. Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Mol Cell* **49**: 453–463.
- Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.
- Mitchell SF, Walker SE, Algire MA, Park EH, Hinnebusch AG, Lorsch JR. 2010. The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol Cell* **39**: 950–962.
- Park EH, Walker SE, Lee JM, Rothenburg S, Lorsch JR, Hinnebusch AG. 2011a. Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1-PABP mRNPs in vivo. *EMBO J* **30**: 302–316.
- Park EH, Zhang F, Warringer J, Sunnerhagen P, Hinnebusch AG. 2011b. Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. *BMC Genomics* **12**: 68.
- Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**: 127–131.
- Pestova TV, Hellen CU, Shatsky IN. 1996a. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol* **16**: 6859–6869.
- Pestova TV, Shatsky IN, Hellen CU. 1996b. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol Cell Biol* **16**: 6870–6878.
- Richter JD, Sonenberg N. 2005. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433**: 477–480.
- Rojas-Duran MF, Gilbert WV. 2012. Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA* **18**: 2299–2305.
- Sen ND, Zhou F, Ingolia NT, Hinnebusch AG. 2015. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res* **25**: 1196–1205.
- Sen ND, Zhou F, Harris MS, Ingolia NT, Hinnebusch AG. 2016. eIF4B stimulates translation of long mRNAs with structured 5' UTRs and low closed-loop potential but weak dependence on eIF4G. *Proc Natl Acad Sci* **113**: 10464–10472.
- Shalgi R, Hurt JA, Krykbaeva I, Taipale M, Lindquist S, Burge CB. 2013. Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell* **49**: 439–452.
- Shatsky IN, Dmitriev SE, Andreev DE, Terenin IM. 2014. Transcriptome-wide studies uncover the diversity of modes of mRNA recruitment to eukaryotic ribosomes. *Crit Rev Biochem Mol Biol* **49**: 164–177.
- Shuman S, Moss B. 1990. Purification and use of vaccinia virus messenger RNA capping enzyme. *Methods Enzymol* **181**: 170–180.
- Singh CR, Watanabe R, Chowdhury W, Hiraishi H, Murai MJ, Yamamoto Y, Miles D, Ikeda Y, Asano M, Asano K. 2012.

- Sequential eukaryotic translation initiation factor 5 (eIF5) binding to the charged disordered segments of eIF4G and eIF2 stabilizes the 48S preinitiation complex and promotes its shift to the initiation mode. *Mol Cell Biol* **32**: 3978–3989.
- Swaney DL, Beltrao P, Starita L, Guo A, Rush J, Fields S, Krogan NJ, Villén J. 2013. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* **10**: 676–682.
- Thompson MK, Gilbert WV. 2017. mRNA length-sensing in eukaryotic translation: reconsidering the “closed loop” and its implications for translational control. *Curr Genet* **63**: 613–620.
- Thompson MK, Rojas-Duran MF, Gangaramani P, Gilbert WV. 2016. The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs. *eLife* **5**: e11154.
- Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**: 109–113.
- Vaidyanathan PP, Zinshteyn B, Thompson MK, Gilbert WV. 2014. Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. *RNA* **20**: 912–922.
- Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP. 2016. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep* **14**: 1787–1799.
- Weingarten-Gabbay S, Elias-Kirma S, Nir R, Gritsenko AA, Stern-Ginossar N, Yakhini Z, Weinberger A, Segal E. 2016. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**: aad4939.



RNA

A PUBLICATION OF THE RNA SOCIETY

Translation initiation factor eIF4G1 preferentially binds yeast transcript leaders containing conserved oligo-uridine motifs

Boris Zinshteyn, Maria F. Rojas-Duran and Wendy V. Gilbert

RNA 2017 23: 1365-1375 originally published online May 25, 2017

Access the most recent version at doi:[10.1261/rna.062059.117](https://doi.org/10.1261/rna.062059.117)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2017/05/25/rna.062059.117.DC1>

References

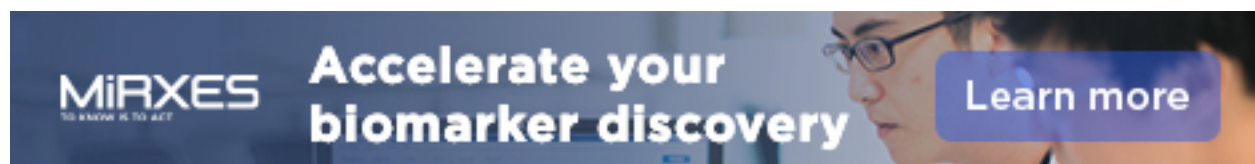
This article cites 60 articles, 28 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/23/9/1365.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
