ADVANCED BIOLOGY
www.advanced-bio.com

Decoding the Heterogeneity and Specialized Function of Translation Machinery Through Ribosome Profiling in Yeast Mutants of Initiation Factors

Jia Wang, Geyu Zhang, Wenfeng Qian, and Ke Li**

The nuanced heterogeneity and specialized functions of translation machinery are increasingly recognized as crucial for precise translational regulation. Here, high-throughput ribosomal profiling (ribo-seq) is used to analyze the specialized roles of eukaryotic initiation factors (eIFs) in the budding yeast. By examining changes in ribosomal distribution across the genome resulting from knockouts of eIF4A, eIF4B, eIF4G1, CAF20, or EAP1, or knockdowns of eIF1, eIF1A, eIF4E, or PAB1, two distinct initiation-factor groups, the "looping" and "scanning" groups are discerned, based on similarities in the ribosomal landscapes their perturbation induced. The study delves into the cis-regulatory sequence features of genes influenced predominantly by each group, revealing that genes more dependent on the looping-group factors generally have shorter transcripts and poly(A) tails. In contrast, genes more dependent on the scanning-group factors often possess upstream open reading frames and exhibit a higher GC content in their 5′ untranslated regions. From the ribosomal RNA fragments identified in the ribo-seq data, ribosomal heterogeneity associated with perturbation of specific initiation factors is further identified, suggesting their potential roles in regulating ribosomal components. Collectively, the study illuminates the complexity of translational regulation driven by heterogeneity and specialized functions of translation machinery, presenting potential approaches for targeted gene translation manipulation.

1. Introduction

Translation initiation is complex and highly regulated process that ensure both the fidelity and efficiency of protein synthesis.[1] It has been widely accepted that translation machinery, encompassing both initiation factors (IFs) and ribosomes, maintains a consistent composition and translates all encountered mRNAs indiscriminately.[2] Generally, the eIF4F complex—eIF4E, eIF4G, and eIF4A—binds to the 5′-cap of the eukaryotic mRNA and recruits the 43S preinitiation complex (PIC) for attachment. The PIC, containing the 40S small ribosomal subunit along with eIF1, eIF1A, eIF3, and the eIF2-Met-tRNA$_i^{Met}$-GTP ternary complex, scans the mRNA to locate the initiation codon with the auxiliary of DEAD-box helicase eIF4A and its cofactor eIF4B.[3]

The prevailing notion that the translation machinery remains static is being re-evaluated.[4] For instance, while some initiation factors can potentially collaborate to form a closed-loop mRNA structure, specifically the 5′-cap-eIF4E-eIF4G-Pab1p-poly(A) tail complex,[5] the eIF4E-binding proteins Caf20p and Eap1p have been identified as competitors to eIF4G for eIF4E binding. This competition disrupts the complex,[6] thereby influencing both translation initiation and mRNA stability.[1b,7] Moreover, the poly(A)-binding (PAB) protein's affinity for the poly(A) tail has been found to vary based on the frequency of non-A nucleotides within the tail, which can then affect translational efficiency.[8]

While numerous biochemical studies have investigated on these initiation factors,[6b,9] their specific functions in selectively regulating various genes remain underexplored. Recent advances, such as RNA immunoprecipitation sequencing (RIP-seq), have begun to map how mRNAs from different genes are selectively bound by translation initiation factors.[1b] However, binding activity alone does not necessarily indicate regulatory influence on translation. Ribosome profiling (ribo-seq) offers a more direct measure of translational consequences upon perturbation of translation initiation factors.[10] Previous ribo-seq studies have assessed individual initiation factors (eIF1, eIF1A, eIF4A, eIF4B, and eIF4G1),[3c,9b,11] but have been hindered by inconsistences in experimental conditions and genetics

J. Wang, G. Zhang, W. Qian, K. Li
State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology
Innovation Academy for Seed Design
Chinese Academy of Sciences
Beijing 100101, China
E-mail: wfqian@genetics.ac.cn; like@genetics.ac.cn
G. Zhang, W. Qian
University of Chinese Academy of Sciences
Beijing 100049, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adbi.202300494

DOI: 10.1002/adbi.202300494

background,[12] making it challenging to form a cohesive understanding of their functional specialization.

In this study, we employed ribo-seq to investigate the heterogeneity and specialized function of translation machinery. Our findings highlight unique translational responses to perturbation of each of nine initiation factors. Data dimensionality reduction reveals two distinct clusters of initiation factors, each preferentially regulating a specific group of genes. We then pinpointed gene sequence features corresponding to this differential regulation, including poly(A) length, which we measured in this study. Analyzing ribosomal RNA fragments from the ribo-seq data, we also detected potential ribosomal heterogeneity linked to changes in initiation factors. These insights illuminate the sequence-specific dynamics of translation control.

2. Results

2.1. Exploring the Specialized Functions of Translation Initiation Factors in Yeast using Ribo-Seq Experiments

To investigate the differential regulation by initiation factors across the genome, we performed ribo-seq along with parallel RNA-seq experiments on yeast mutants of individual translation initiation factors (**Figure 1**A,B). Note that in our study, we also regarded non-canonical factors such as Caf20p, Eap1p, and Pab1p, which are involved in translation initiation, as initiation factors. Specifically, we either knocked out individual non-essential genes (*eIF4A*, *eIF4B*, *eIF4G1*, *CAF20*, or *EAP1*) in the haploid budding yeast strain BY4742, or reduced the mRNA levels of essential genes (*eIF1*, *eIF1A*, *eIF4E*, or *PAB1*) through Decreased Abundance by mRNA Perturbation (DAmP, Figure 1B).[13] For comparison, we also carried out ribo-seq and RNA-seq experiments on wild-type yeast cells. All DAmP-mediated knockdowns were highly efficient, reducing the mRNA levels of the target gene to no more than 25% of the levels in wild-type cells (Figure 1C; Figure S1A, Supporting Information).

To avoid the known issues of ribosome distribution bias toward the initiation codon, which can be induced by pre-treatment with the translation elongation inhibitor cycloheximide,[12a,14] we opted to omit cycloheximide pre-treatment in all our ribo-seq experiments. Furthermore, we performed computational quality assessment of our ribo-seq data and found consistent patterns. For example, across all strains tested, 28-nt ribosome-protected fragments (RPFs) showed a consistent 3-nt periodicity, in line with stepwise (i.e., codon-by-codon) ribosome movement along mRNA (Figure 1D,E; Figure S1B,C, Supporting Information). Moreover, the 5′ ends of the RPF were found to be approximately 12 nt upstream of the initiation codon (Figure 1E; Figure S1C, Supporting Information), consistent with prior findings that the nucleotides at positions 13 to 15 in the 28-nt RPFs are located at the P-site of a translating ribosome.[10]

To ensure the robustness and reproducibility of our findings, we performed two biological replicates for each yeast mutant. We quantified both RPF and mRNA levels for each gene in both mutants and wild-type strains (Tables S1 and S2, Supporting Information). We then calculated pairwise Spearman's correlation coefficients (ρ) to compare two biological replicates within the same mutants and between different mutants. The Spearman's correlation coefficients ranged from 0.88 to 0.98 for RPF levels (Figure 1F) and from 0.91 to 0.99 for mRNA levels (Figure 1G). Notably, the Spearman's correlation coefficients between biological replicates were significantly higher than those between different mutants (Figures 1F and 1G). These results collectively underscored the high reproducibility of our data.

2.2. Identification of Two Groups of Translation Initiation Factors based on Ribosomal Landscape Changes Following their Perturbation

To delve into the specific regulatory functions of translation initiation factors, we calculated the translational efficiency (TE) for each gene in both mutant and the wild-type yeast. TE was defined as the ratio of RPF abundance to mRNA abundance (**Figure 2**A; Table S3, Supporting Information). We then evaluated the influence of each perturbed initiation factor translation by calculating the fold change in TE (ΔTE), which was defined as the ratio of TE in the mutant to TE in the wild type (Figure 2A; Table S4, Supporting Information). To visualize these effects across the genome, we performed Uniform Manifold Approximation and Projection (UMAP) based on the ΔTE values. This analysis revealed that the initiation factors clearly clustered into two distinct groups (Figure 2A).

The first group consisted of Pab1p, eIF4G1, eIF4E, Eap1p, and Caf20p. These factors have been previously linked to forming a closed-loop structure in mRNA by connecting the cap to the poly(A) tail.[1a,6] This cluster was henceforth termed the "looping" group. The second group included eIF1, eIF1A, eIF4A, and eIF4B, factors previously known to be involved in ribosome scanning and initiation codon recognition.[1a,3a] We referred to this cluster as the "scanning" group. The distinct separation between these two groups indicated collaboration among initiation factors in translational regulation, and also suggested specialized functions of initiation factors.

2.3. Identification of *Cis*-Regulatory Elements Associated with Specialized Functions of the Two Groups of Initiation Factors

To discern the genetic basis for the preferential regulation of genes by either looping-group or scanning-group factors, we probed the genomic sequence for cis-regulatory elements that might interface with selective regulation by the two factor groups. To this end, we identified genes more influenced by one group over the other. For each gene, we calculated the average TE within each group and determined the TE ratio between the two groups (TE ratio), which was defined as the ratio of TE upon perturbation of looping-group factors to TE upon perturbation of scanning-group factors. A gene with a TE ratio significantly greater than 1 was classified as more dependent on scanning group, whereas a gene with a TE ratio significantly less than 1 was considered more dependent on looping group (Figure 2B).

We then examined 77 various gene features of gene sequence or expression to discern differences between the scanning- and looping-dependent genes. Our analyses revealed that looping-dependent genes generally exhibited significantly higher RPF abundance, mRNA abundance, and TE (Figure 2C). These genes also predominantly had shorter transcript lengths, primarily due

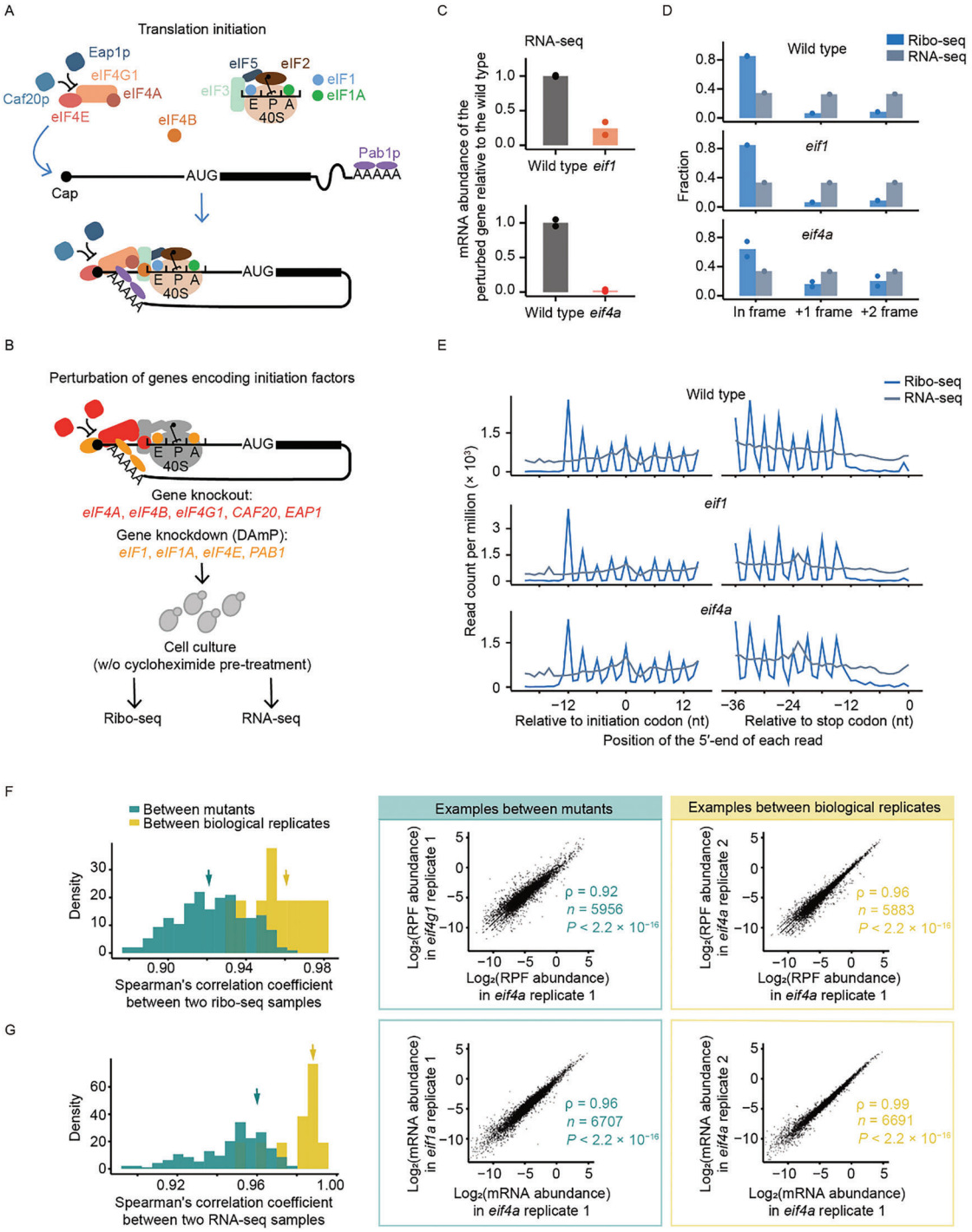

Figure 1. Ribosome profiling in yeast strains with perturbed initiation factors. A) Schematic diagram of the translation initiation process. B) Diagram illustrating the perturbation of genes encoding initiation factors and subsequent ribosome profiling. C) Perturbation efficiency in genes subjected to knockdown or knockout was measured as mRNA abundance in the respective mutants relative to the wild type using RNA-seq data. Biological replicates are shown in individual dots. Perturbation efficiencies of knockdown (*eIF1*) and knockout (*eIF4A*) are shown. For additional initiation factors, refer to Figure S1A (Supporting Information). D,E) Distribution of 28-nt RPFs and all RNA reads, relative to the reading frame D) or around the initiation and stop codons E). These counts represent the average of data obtained from two biological replicates and have been normalized by the total reads in each library. Shown are examples of *eif1, eif4a*, and the wild type. For additional initiation factors, refer to Figure S1B,C (Supporting Information). F,G) Density plots showing the reproducibility of ribo-seq F) and RNA-seq G) data between mutants or biological replicates. Scatter plots depict examples (indicated by arrows in the left images) showing correlations between mutants (cyan) or biological replicates (gold). Spearman's correlation coefficients (ρ) and the corresponding *P* values are shown.

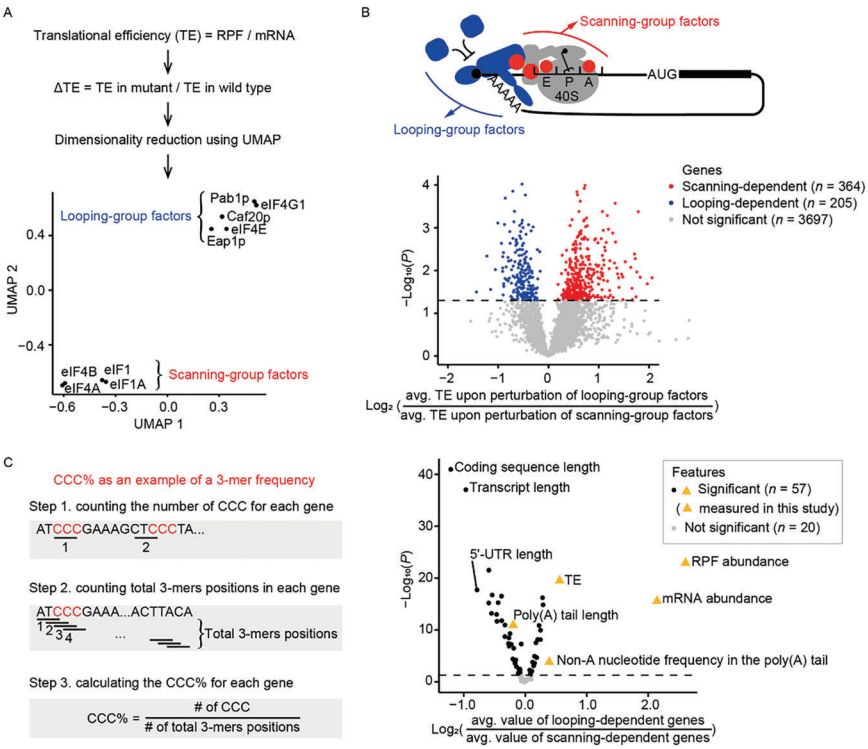

Figure 2. Identification of genes selectively regulated by initiation factors.
A) UMAP visualization of initiation-factor clusters based on ΔTE of mutants. A flowchart at the top illustrates the brief computational process. B) Identification of genes whose translation depends more on the looping- or scanning-group initiation factors. *P* values were given by the Student's *t* tests. C) Identification of sequence-based and expression-related gene features that distinguish between looping- and scanning-group initiation factors. Each dot represents a feature. *P* values were given by the Mann-Whitney rank sum *U* tests. The features experimentally estimated in this study are represented by gold triangles.

to shorter coding sequences (**Figure 3**A, $P < 2.2\times10^{-16}$, Mann-Whitney rank sum *U* test). The shorter transcript lengths likely lead to increased local concentrations of cap-binding protein eIF4E and the poly(A) binding protein Pab1p.[1,24] As a result, these genes become more sensitive to regulation by looping-group factors, which in turn enhances their translation efficiencies and/or mRNA stabilities.

To further validate this observation, we sorted genes into eight equal-size categories based on their coding sequence lengths. We found that genes with shorter coding sequence displayed a greater reduction in translational efficiencies in each looping-group mutant compared to the wild type (Figure 3B). This pattern was not evident in the case of the scanning-group factors (Figure S2A, Supporting Information). The confirmation of a previously hypothesized feature of looping-dependent genes encouraged further investigation into other sequence features associated with uniquely affecting translational regulation by either the looping- or scanning-group factors.

2.4. Genes with Shorter Poly(A) Tails are More Affected by Perturbation of Looping-Group Factors

Several studies have noted that poly(A) tail lengths vary among genes within the same species.[8,15] In principle, the poly(A) tail length could influence its binding affinity to PABP, which in turn may affect how different genes are regulated by looping-group factors. To test this hypothesis, we measured poly(A) tail lengths in our wild-type yeast strain BY4742 using the poly(A) inclusive RNA Isoform-sequencing (PAIso-seq, Figure 3C). The observed lengths ranged from 29–71 nt averaging 46 nt, which aligned with previously reported lengths in another yeast strain W303.[15] To investigate the role of poly(A) tail length in differential regulation by looping- and scanning-group factors, we compared poly(A) tail lengths between genes dependent on each group. Looping-dependent genes had significantly shorter poly(A) tails, averaging 41 nt (Figures 2C and 3D, $P = 1.2\times10^{-11}$, Mann-Whitney rank sum *U* test). This suggested that genes with shorter poly(A) tails were more prone to reduced binding with Pab1p, making these genes more susceptible to changes in looping-group factors, including Pab1p.

We further tested this hypothesis by dividing genes into eight equal-size categories according to the poly(A) tail length. We found that genes with shorter poly(A) tails experienced a greater decrease in TE when looping-group factors were perturbed, compared to the longer-tailed counterparts (Figure 3E). This trend was not observed for scanning-group factors (Figure S2B, Supporting Information). Together, these results indicated that looping-group factors had a more pronounced impact on the translation of genes with shorter poly(A) tails.

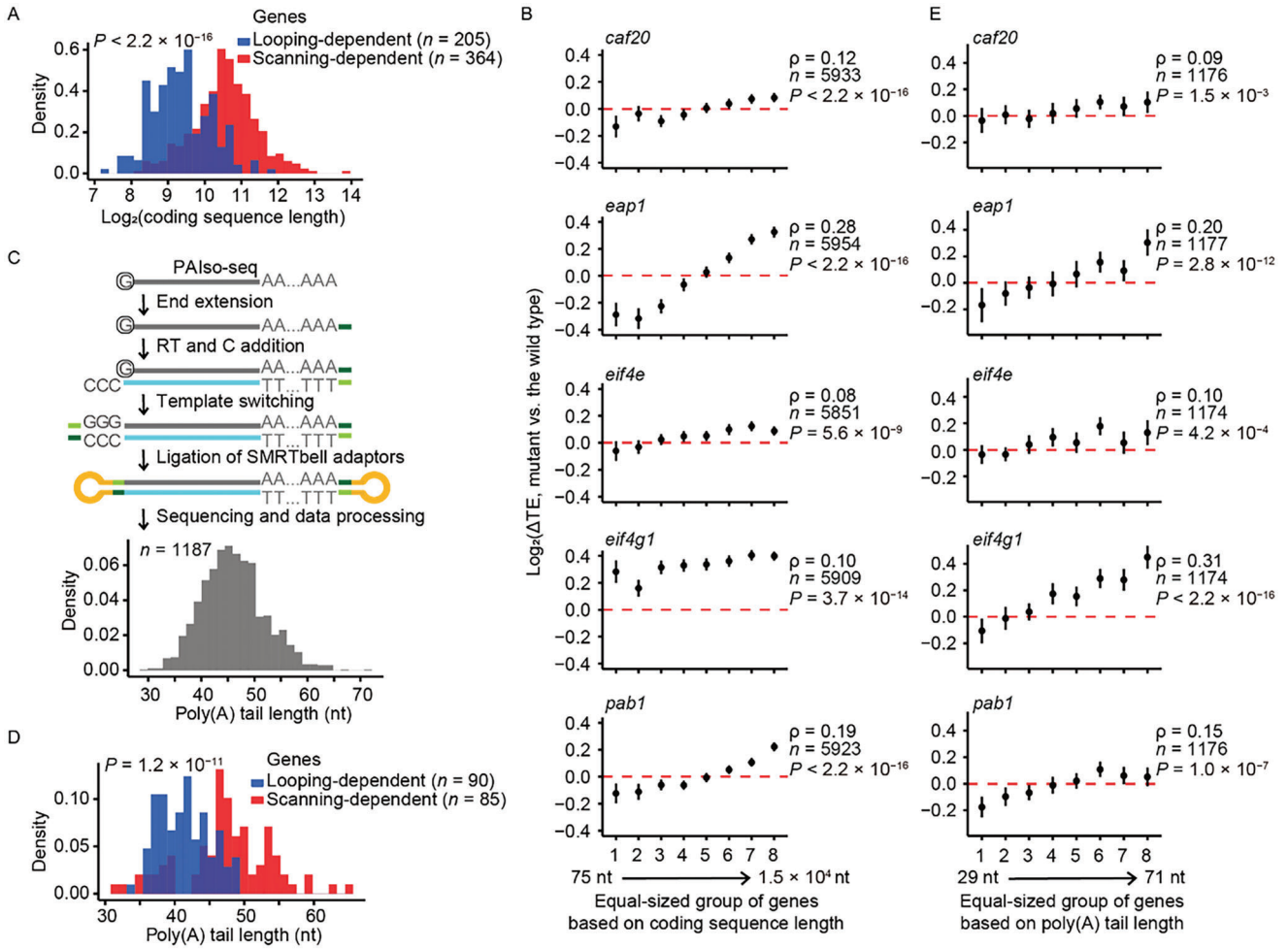

Figure 3. Translational regulation by looping-group initiation factors via coding sequence length and poly(A) tail length. A) Distribution of coding sequence length for both looping- and scanning-dependent genes. *P* value was given by the Mann-Whitney rank sum *U* test. B) Plots showing ΔTE for mutants of looping-group initiation factors. Genes were grouped by coding sequence length into eight equal-sized groups. Each dot represents the median ΔTE of a bin, and the error bars depict 95% confidence intervals. Spearman's correlation coefficients (ρ) and the corresponding *P* values are shown. C) Flowchart illustrating the PAIso-seq process for wild-type yeast BY4742. The distribution of poly(A) tail length is shown at the bottom. D) Distribution of poly(A) tail length for both looping- and scanning-dependent genes. *P* value was given by the Mann-Whitney rank sum *U* test. E) Plots showing ΔTE across eight bins of poly(A) tail length for mutants of looping-group initiation factors. Other aspects of the figure are consistent with (B).

2.5. Impact of Translation Initiation Factors on Translatome is Hard to Predict from mRNA Binding Affinity Alone

Previous research assessed the mRNA binding affinity of various translation initiation factors, including Caf20p, Eap1p, eIF4G1, eIF4E, and Pab1p, using RIP-seq.[1b] One might intuitively expect that stronger binding affinity between mRNA and a translation initiation factor would lead to greater effects on translation if the factor is perturbed. To explore this assumption, we calculated the correlation between mRNA binding affinity and fold change in TE (ΔTE) upon perturbation for each of these factors. Surprisingly, we found that the correlations between ΔTE and mRNA binding affinities were generally weak, ranging from −0.15 to 0.23 (Figure S3A, Supporting Information). Additionally, we observed an unexpected negative correlation between poly(A) tail length and mRNA binding affinity of Pab1p ($\rho = -0.22$, $P < 2.2 \times 10^{-16}$, Spearman's correlation, Figure S3B, Supporting Information). The result indicated that the relationship between a factor's binding affinity to mRNA and its regulatory role in translation is more complex than naturally assumed.

2.6. Scanning-Group Factors Contribute to Ribosome Accumulation at Upstream Open Reading Frames (uORFs)

The notably longer 5′-UTRs in scanning-dependent genes ($P < 2.2 \times 10^{-16}$, Mann-Whitney rank sum *U* test, Figures 2C and 4A) led us to investigate whether these genes contained more uORFs, which has been shown to interfere with the scanning process.[16] To explore this, we counted the number of genes, with or without uORFs, for both scanning-dependent and looping-dependent genes, referencing the uORF definitions from a previous study.[10] Our analysis revealed a significantly enrichment of genes with uORFs in scanning-dependent genes compared to

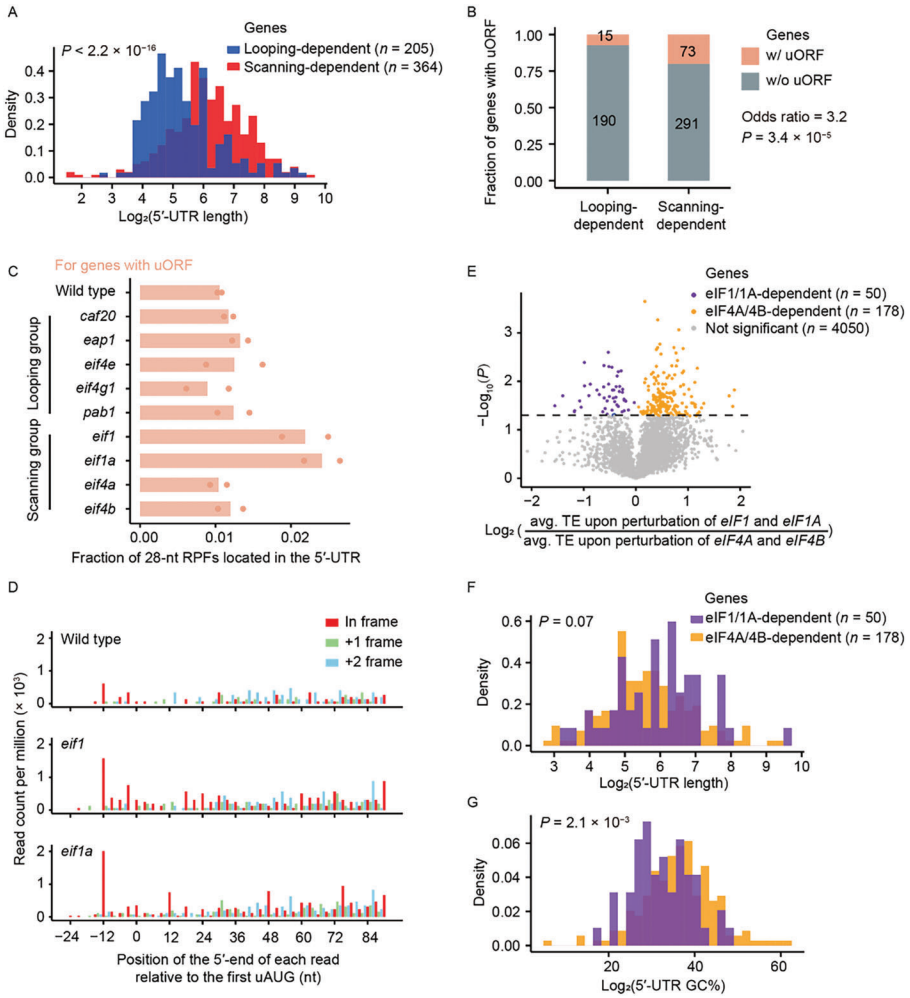

Figure 4. Translational regulation by scanning-group initiation factors via uORFs or GC% in the 5′-UTR. A) Distribution of the 5′-UTR length for both looping- and scanning-dependent genes. B) Fractions of genes containing uORFs in looping- or scanning-dependent genes. List of genes containing uORFs were retrieved from Ingolia et al. (2009), and the *P* value was determined using Fisher's exact test. C) Bar plots showing the average faction of 28-nt RPFs in the 5′-UTR compared to all 28-nt RPFs located in the entire transcript, for all genes with uORFs, among mutants of initiation factors. Biological replicates are shown in individual dots. D) Distribution of the 5′-end of 28-nt RPFs around the first uAUG, in *eif1*, *eif1a*, and the wild type. The read counts were normalized by the total read counts for all genes with at least one uORF. E) Identification of eIF1/1A- and eIF4A/4B-dependent genes. *P* values were given by the Student's *t* tests. F,G) Distribution of 5′-UTR length F) and 5′-UTR GC% G) for both eIF1/1A- and eIF4A/4B-dependent genes. *P* values were given by the Mann-Whitney rank sum *U* tests.

looping-dependent genes (odds ratio = 3.2, $P = 3.4\times10^{-5}$, Fisher's exact test, **Figure 4**B). This suggested that scanning-group factors played a more critical role in ensuring accurate translation initiation for genes containing uORFs.

To examine whether these scanning-group factors help in the accurate recognition of the translation initiation sites, we looked at how perturbing these factors affected ribosome distribution along the transcript. Specifically, we compared the number of RPFs between the 5′-UTR and the entire transcript. We found that this ratio increased significantly in mutants of *eif1* and *eif1a* for genes containing uORFs (Figure 4C). Moreover, a more pronounced translation initiation signal at the upstream initiation codon (uAUG) was observed in *eif1* and *eif1a* mutants compared to the wild type (Figure 4D). Interestingly, this effect was most prominent at the upstream-most uAUG (Figure S4, Supporting Information). Taken together, these findings suggested that the perturbation of *eIF1* and *eIF1A* could lead to a higher proportion of RPFs in the 5′-UTR, likely due to increased translation initiation at the most upstream uAUGs.

2.7. eIF4A and eIF4B Specialize in Regulating Genes with High GC Content (GC%) in the 5′-UTR

In the UMAP analysis, scanning-group factors were further differentiated into two subgroups: eIF1/eIF1A and eIF4A/eIF4B (Figure 2A). This differentiation suggested that each subgroup had distinct preferences in translational regulation. To investigate these preferences, we identified genes that were differentially regulated by these subgroups. Specifically, genes with a TE fold-change ratio significantly greater than 1 in *eif1* and *eif1a* mutants compared to *eif4a* and *eif4b* mutants were classi-

fied as eIF4A/4B-dependent genes. Conversely, if the ratio was significantly less than 1, the genes were classified as eIF1/1A-dependent genes (Figure 4E).

In comparison to eIF1/1A-dependent genes, eIF4A/4B-dependent genes had similar 5′-UTR length ($P = 0.07$, Mann-Whitney rank sum U test, Figure 4F). However, these genes had a notably higher GC content in their 5′-UTRs ($P = 2.1 \times 10^{-3}$, Mann-Whitney rank sum U test, Figure 4G). Given that higher GC content is linked to increased mRNA secondary structure stability,[17] and considering eIF4A's known function as an RNA helicase,[18] these findings suggested that eIF4A and eIF4B collaboratively regulated the translation of genes with structurally complex 5′-UTRs.

2.8. The Two Initiation Factor Groups Preferentially Regulated Genes with Different Functions

What drives the specialized functions of translation initiation factors? We hypothesized that such specialization enabled cells to specifically regulate genes with different functions. To test this hypothesis, we performed an enrichment analysis, comparing genes primarily influenced by either initiation factor group (Figure S5, Supporting Information) based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The analysis unveiled pronounced difference in gene functions: genes encoding ribosomal proteins were predominantly regulated by looping-group factors (Figure S5A, Supporting Information), whereas genes central to metabolism relied heavily on scanning-group factors (Figure S5B, Supporting Information). Such patterns suggested a potential mechanism where initiation factors play a key role in tailoring the translational regulation of genes based on their functional attributes.

2.9. Potential Ribosomal Heterogeneity Caused by Perturbation of Initiation Factors

A previous study has demonstrated that variations in ribo-seq read distributions mapped to ribosomal RNA (rRNA) can be indicative of ribosome heterogeneity mediated by ribosomal proteins.[19] This stems from the premise that differences in ribosomal protein incorporation can lead to distinct rRNA fragments in the ribo-seq data, specifically near the regions adjacent to these proteins. Building on this idea, we probed whether perturbations in translation initiation factors could similarly influence ribosomal heterogeneity. Analyzing the read distribution across rRNAs revealed noticeable variability among yeast mutants of initiation factors (**Figure 5**A). To discern rRNA fragment patterns across mutants, we used UMAP analysis of changes in rRNA coverage at each nucleotide, which unveiled four prominent clusters of initiation factors (Figure 5B). These patterns diverged significant from the UMAP plot based on ΔTE of protein-coding genes (Figure 2A), suggesting an additional role of initiation factors in modulating ribosomal protein components engaged with specific transcripts.

Given that the production of ribosomal proteins can be influenced by perturbation of initiation factors (Figure S5A, Supporting Information), we further investigated whether the variation in rRNA coverage from the ribo-seq data could be partially explained by changes in ribosomal protein production rates. We found it indeed a possible mechanism. For instance, upon perturbation of *eIF1A*, *eIF4A*, *eIF4B*, *eIF4E*, *CAF20*, or *EAP1*, there was an increased production of ribosomal protein L19 (Table S5, Supporting Information), a protein component of the large ribosomal subunit. Notably, based on the yeast 80S ribosome structure (4V7R in the PDB database),[20] ribosomal protein L19 is located near the 1938–2116 nt region of the 25S rRNA, which corresponds with the region showing substantially increased coverage in our ribo-seq data (Figure 5A). This alignment suggested that initiation factors might modulate ribosomal heterogeneity through affecting ribosomal protein production rates.

2.10. Reduced mRNA Levels due to Perturbation of Translation Initiation Factors

Prior research has explored the influence of translation elongation rates on mRNA stability,[21] leading us to influence if translation initiation also has any impacts on mRNA stability. To this end, we analyzed ribo-seq and RNA-seq data from mutants with perturbed translation initiation factors to determine if changes in TE impact mRNA levels. To avoid the previously reported intrinsic negative correlation between changes in the mRNA level and changes in TE,[22] we used standardized major axis regression to find the slope between fold changes in mRNA abundance and RPF abundance for each initiation factor mutant. In every case, the observed slope exceeded one (**Figure 6**A), suggesting that decrease in translation efficiency also reduced mRNA abundance.

This synergistic regulation of translation initiation and mRNA levels might involve the nonsense-mediated decay (NMD) pathway, which triggers mRNA degradation in response to abnormal translation termination.[23] To further explore this possibility, we hypothesized that mRNAs with uORFs would be more susceptible to NMD and therefore show a stronger correlation between changes in translation initiation and mRNA levels upon perturbation of initiation factors. To test this hypothesis, we calculated the correlation between fold change in RPF and fold change in mRNA abundance, separately for genes with and without uORFs. Our analysis revealed that genes with uORFs had a steeper slope compared to those without, across all initiation factor mutants (Figure 6B; Figure S6A,B, Supporting Information). These observations indicated that the interaction between translation initiation and mRNA abundance was likely mediated through the NMD-related mRNA stability.

3. Discussion

3.1. Unraveling the Complex Nature of Translation Machinery

While many molecular biology textbooks portray the ribosome as a singular, uniform entity, our study delves into the intricate, specialized functions of translation initiation factors using ribosome profiling in yeast mutants. This approach highlights the heterogeneity and specialized functions within the translation machinery. Although prior research has reported the selective binding of transcripts by translation initiation factors based on RIP-seq

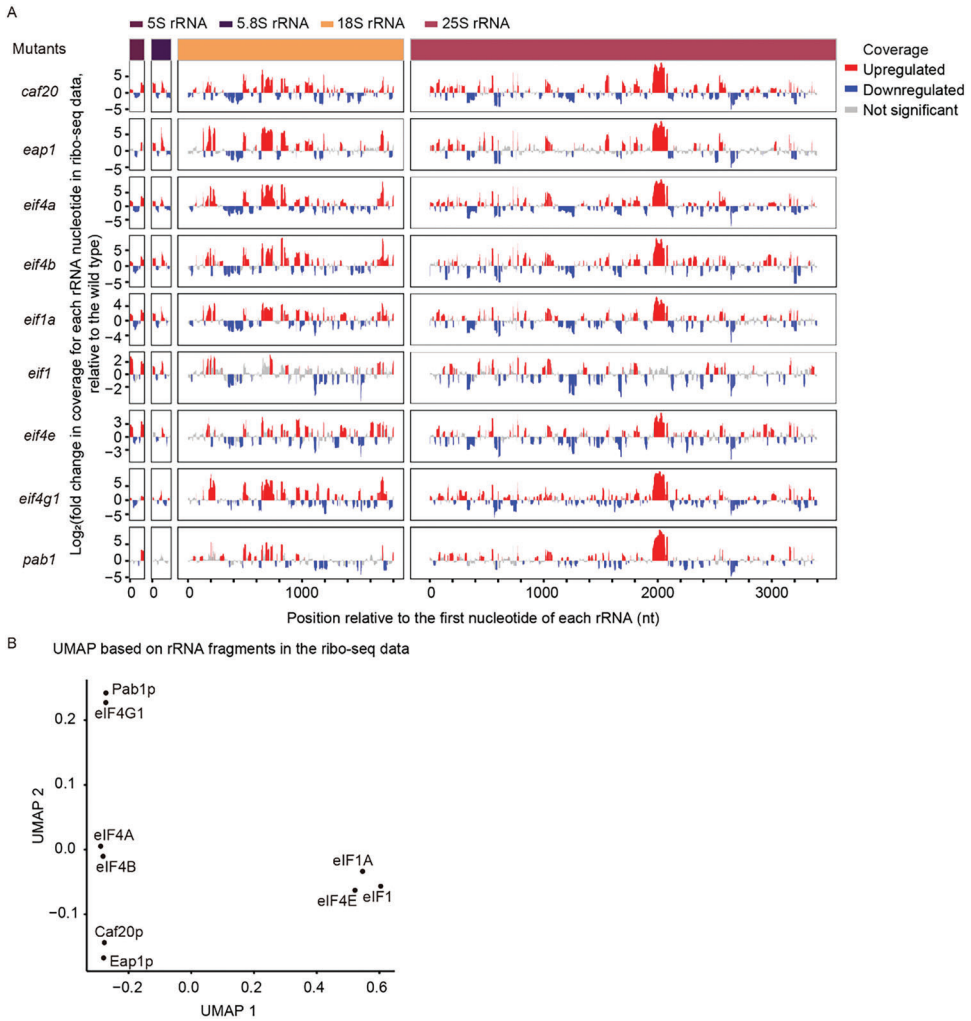

Figure 5. Ribosomal heterogeneity linked to perturbations of initiation factors. A) Fold change in coverage for each rRNA nucleotide along 5S, 5.8S, 18S, and 25S rRNAs. Two biological replicates were combined for each mutant. B) UMAP visualization of initiation-factor clusters based on fold changes in coverage for each rRNA nucleotide.

data,[1b] our findings challenge the idea that mRNA binding affinity to these factors directly correlates with translational outcomes upon factor perturbation. Rather, it seems that the inherent characteristics of the mRNA molecules themselves serve as more accurate predictors. These findings together offer new avenues for understanding the nuanced roles of translation initiation factors in translational regulation.

3.2. Looping- and Scanning-Group Initiation Factors Preferentially Regulated Different Biological Processes

Our KEGG pathway analysis underscores distinct regulatory approaches by looping- and scanning-group initiation factors (Figure S5, Supporting Information). Looping-group factors seem to favorably regulate the production of ribosomal proteins (Figure S5A, Supporting Information). This aligns with prior observations that genes encoding ribosomal proteins usually have shorter mRNA sequences and are highly expressed,[24] characteristics echoing the cis-regulatory features of looping-dependent genes we identified (Figure 2C). Conversely, scanning-group factors primarily influence metabolic pathways, such as starch and sucrose metabolism, and the biosynthesis of secondary metabolites (Figure S5B, Supporting Information). Collectively, these findings suggest a cellular approach for synchronizing protein production rates within particular functional pathways by selectively utilizing different initiation factors.

3.3. The Role of eIF4E Binding Proteins as Positive Regulator of Translation Initiation

Conventional view has held eIF4E binding proteins, Eap1p and Caf20p, could inhibit the formation of the 5′-cap-eIF4E-eIF4G1-Pab1p-poly(A) structure by competing with eIF4G for eIF4E binding at the same site.[6] This competition has been shown to lead to reduced translation initiation efficiency through decreased ribosome recycling[1,7] and disturbed 80S ribosome

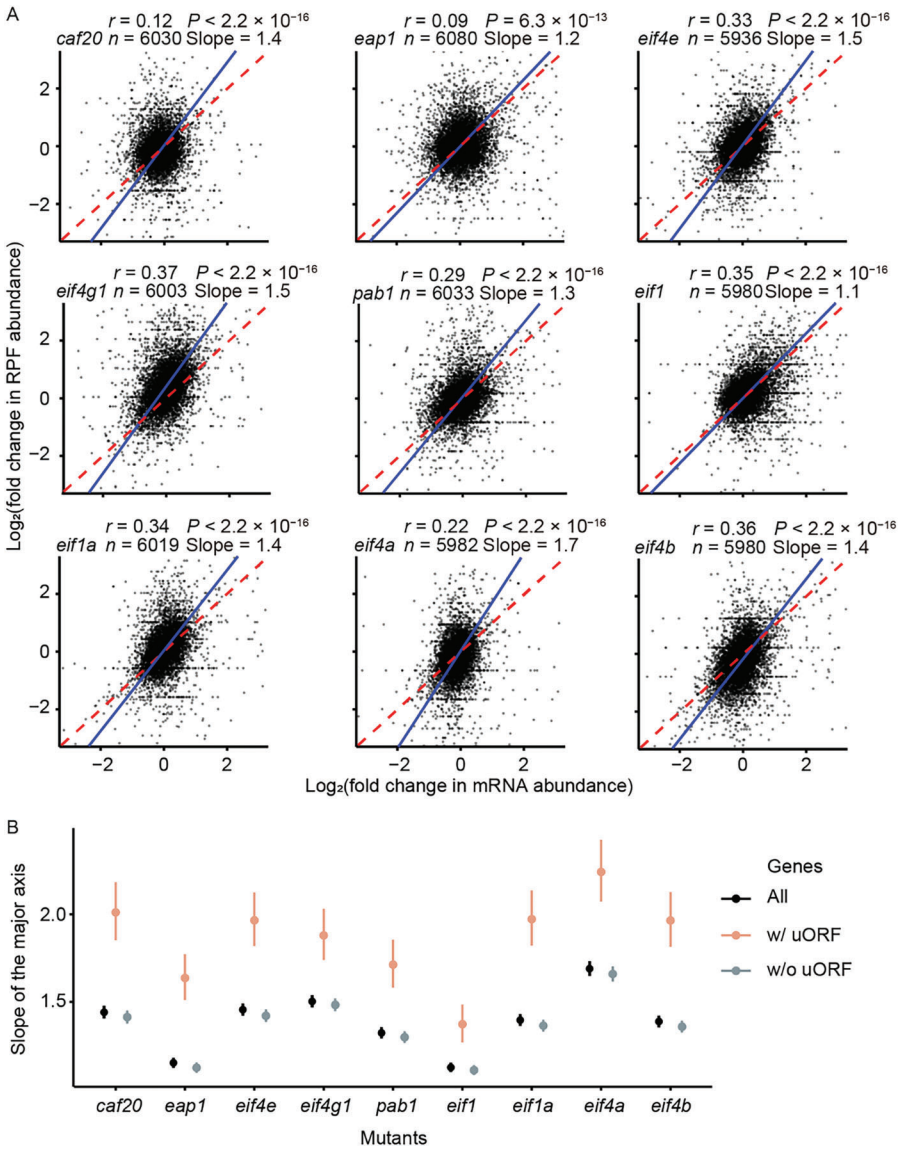

Figure 6. Translational efficiency and mRNA stability are coordinated upon purterbation of initiation factors. A) Scatter plots illustrating the correlation between changes in mRNA abundance and RPF abundance for mutants of initiation factors. Blue line represents the slope of standardized major axis (SMA) regression, while red dashed line represents the $y = x$ line. Pearson's correlation coefficients (r) and the corresponding P values are shown. B) SMA regression slopes across different gene categories (all genes, genes with uORF, and genes without uORF) for mutants of initiation factors. The error bars represent 95% confidence intervals.

loading.[25] Contrary to this, our UMAP analysis based on ΔTE profiles revealed that Caf20p and Eap1p clustered closely with eIF4E, eIF4G1, and Pab1p when translation initiation factors were perturbed (Figure 2A). Disrupting Eap1p and Caf20p resulted in ΔTE profiles that closely resembled that observed with eIF4E perturbation ($\rho = 0.57$, $P < 2.2 \times 10^{-16}$, Spearman's correlation, Figure S2C, Supporting Information). This is in line with a previous study that found increased eIF4E-cap binding and higher protein production levels in the presence of Caf20p.[26] Therefore, we proposed that eIF4E binding proteins should be reevaluated as positive, rather than negative, regulators of translation initiation, prompting a new direction of research in this area.

3.4. Link Between Translation Initiation and mRNA Stability

Previous research has shown that the use of unpreferred codons can lead to mRNA degradation and thereby reduced the mRNA concentration.[21,27] Specifically, slow translation elongation due to unpreferred codons recruits Not5p to the ribosome E-site, which then activates mRNA degradation mediated by the CCR4-NOT complex and a dead-box protein Dhh1p.[28] Our study extends this understanding by demonstrating a link between translation initiation and mRNA stability, particularly in mRNAs with uORFs (Figure 6B; Figure S6, Supporting Information). While NMD can partially explain this for uORF-containing mRNAs,[29] our finding suggested a boarder coupling between translation

initiation and mRNA degradation across all initiation factor mutants. This is also supported by studies showing reduced global mRNA half-life when translation initiation is inhibited by an inhibitor of eIF4A, hippuristanol.[30] Further research is needed to explore this relationship between translation initiation and mRNA stability.

3.5. The Role of the Poly(A) Tail in Regulating Translation Initiation

Our study sheds light on the nuanced relationship between poly(A) tail length and translation initiation, particularly in the context of Pab1p knockdown in yeast. While previous study in *Xenopus* oocytes suggested that genes with shorter poly(A) tails are more sensitive to Pab1p levels,[31] our findings in yeast corroborate this observation (Figure 3E). Intriguingly, we noted that shorter poly(A) tails were more strongly bound with Pab1p (Figure S3B, Supporting Information), which contradicts earlier studies positing stronger Pab1p binding for longer poly(A) tails.[1] Previous studies also identified non-A bases (i.e., G, C, or T) in the poly(A) tails, especially guanine.[8,15,32] Although not discussed in the Results section, we also found a negative correlation between the non-A nucleotide frequency in the poly(A) tail and the poly(A) tail length ($\rho = -0.15$, $P = 4.8\times10^{-7}$, Spearman's correlation, Figure S3C, Supporting Information). Consistently, our findings also indicated that looping-dependent genes exhibited a significantly greater frequency of non-A nucleotide in the poly(A) tails (Figure 2C, $P = 1.4\times10^{-4}$). Taken together, these observations open up new avenues for understanding how the poly(A) tail and the closed-loop protein complex contribute to the regulation of translation initiation.

3.6. Limitations and Future Directions

While our study offers a comprehensive analysis of the heterogeneity and specialized function of translation machinery using yeast mutants of initiation factors, it has its limitations. For instance, we have yet to delineate the specific mechanisms through which certain translation initiation factors selectively regulate genes possessing particular attributes. Furthermore, the variation in rRNA coverage upon perturbations of translation initiation factors may not entirely accurately reflect the heterogeneity of ribosomal composition under natural conditions, and further research is warranted. For example, techniques such as translation complex profile sequencing (TCP-seq) or selective TCP-seq (sel-TCP-seq)[33] can be employed to pinpoint the locations of specific ribosomal 40S or 80S complexes on the mRNA, enabling a more in-depth investigation of ribosomal heterogeneity.

Additionally, our proposition that the NMD pathway connects initiation factors to mRNA stability also warrants further exploration. Our findings challenge the conventional view of eIF4E binding proteins as negative regulators of translation, but are based on correlations, necessitating further validation through mechanistic studies. Future research could bridge these gaps through complementary approaches such as in vitro assays and biochemical tests, or by extending the study to other organisms. Experiments involving genetic manipulation could also offer richer insights into the intricate genomic architecture governing translation.

4. Experimental Section

Construction of Yeast Mutants: *Knockout Strains*: Five initiation-factor encoding genes investigated in this study (*eIF4A*, *eIF4B*, *eIF4G1*, *CAF20*, or *EAP1*) are nonessential in *Saccharomyces cerevisiae*.[34] To delete these nonessential genes, we employed polymerase chain reaction (PCR)-mediated homologous recombination approach. Taking the *eif4g1*Δ strain as an example, we amplified a 54-bp segment of the 5′-UTR of eIF4G1, together with a *URA3MX* cassette as auxotrophic marker, and another 54-bp segment of the 3′-UTR of eIF4G1 (primers provided in Table S6, Supporting Information). Subsequently, we transformed the amplified PCR products into the BY4742 strain (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*). Successful transformants were selected on agar plates without uracil (comprising 8 g L^{-1} synthetic medium with uracil dropped-out, 2% glucose, and 2% agar) and incubated at 30 °C for two days. Positive colonies were confirmed by PCR, and two of these colonies were stored at −80 °C for use in future experiments. Moreover, to address the potential cellular effects arising from *URA3MX* expression when comparing deletions to the wild type, we generated a *ho*Δ strain. Specifically, we replaced the pseudogene *HO* in the BY4742 background with *URA3MX* cassette. Throughout the study, this *ho*Δ strain served as the wild-type strain, ensuring uniform conditions for comparison.

Knockdown Strains: Four initiation-factor encoding genes investigated in this study (*eIF1*, *eIF1A*, *eIF4E*, or *PAB1*) are considered essential.[34] To avoid the potentially fatal growth consequences arising from their deletion, we adopted the DAmP approach (Schuldiner et al., 2005) to downregulate their expression. In this method, we introduced a *natMX* selective marker right after the stop codon of each essential gene using PCR-mediated homologous recombination approach (primers provided in Table S6, Supporting Information). This insertion extended the native 3′-UTR, resulting in a decrease in the corresponding mRNA levels while ensuring cellular viability. We selected the transformants on the YPD solid medium with 100 μg mL^{-1} nourseothricin (Amresco, 6021–878).

Ribosome Profiling: Ribo-seq and corresponding RNA-seq were performed as described in a previous study.[10,14] For each strain, we selected two distinct colonies to serve as biological replicates. Initially, we cultured each colony overnight in YPD media (1% yeast extract, 2% peptone, and 2% glucose). Then, we transferred them into 500 mL of YPD at an optical density (OD) of ≈0.1 at 660 nm. The cultures were incubated at 30 °C with continuous agitation at 200 rpm until reaching an OD of ≈0.6 at 660 nm. After growth, we harvested the cells through filtration at room temperature. The collected cells were then ground and suspended in 1.5 mL of polysome lysis buffer (PLB) containing 200 mmol Tris-HCl (pH 8.0), 200 mmol KCl, 35 mmol MgCl2, 1% (v/v) Triton X-100, 5 mmol DTT, and 100 μg mL^{-1} cycloheximide (Sigma-Aldrich, C7698-1G). The resulting cell lysates were divided into two aliquots, with one aliquot utilized for ribo-seq and the other for RNA-seq experiment.

For the ribo-seq experiment, the cell lysates were digested with RNase I (Ambion, AM2295) for 1 h, and the digestion was subsequently halted using Superase-In (Invitrogen, AM2696). The ribosome-protected mRNA was directly extracted using hot acid phenol (pH < 5.0, Solarbio, P1012-100) following the termination of digestion.

For RNA-seq, the total RNA was similarly extracted using hot acid phenol (pH < 5.0) and the purified mRNA was isolated by the Dynabeads mRNA Purification Kit (Life Technologies, 61 006). Subsequently, the mRNA underwent fragmentation using Fragmentation Reagents (Thermo Fisher Scientific, AM8740).

The ribosome-protected mRNA and fragmented mRNA were then separated on 17% denaturing urea-PAGE gel, and the RNA fragments ≈28 nt in length were precisely excised, guided by 28-nt marker. Following the gel extraction, RNA was recovered using RNA extraction buffer (300 mmol NaOAc (pH 5.2), 10 mmol Tris-HCl (pH 8.0), 1 mmol EDTA). Antarctic phosphatase (NEB, M0289L) and T4 PNK (NEB, M0201S) were employed

to carry out RNA dephosphorylation and phosphorylation, respectively. Ultimately, the ribo-seq and RNA-seq libraries were constructed using the Balancer NGS Library Preparation Kit for small/microRNA (Gnomegen, k02420). The resulting libraries were subsequently sequenced using Illumina Hiseq2500 platform with single-end 50 or single-end 75 mode.

Construction of PAIso-Seq Library: We constructed the PAIso-seq library followed the protocols outlined in a prior study.[15] Initially, three separate cultures of the *ho*Δ strain (wild type) from different colonies were grown in YPD media until they reached an OD of ≈0.6 at 660 nm. Subsequently, total RNA was extracted from these cultured cells. mRNA was then isolated from 750 ng of total RNA using Dynabeads mRNA Purification Kit. A GI-anchor sequence (5′-GGGGGGGGG-3′) was added to the 3′-end of the mRNA, and the mRNA was further purified using magnetic beads. Next, the mRNA was converted into full-length cDNA through reverse transcription using the Smarter PCR cDNA Synthesis Kit (Clontech, 634 926). The PacBio transcriptome library was prepared using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, 100-991-900) and sequenced using the PacBio Sequel II platform with circular consensus sequencing (CCS) mode.

Statistical Analyses: *Quantification of RPF and mRNA Abundance in Ribosome Profiling—Mapping Reads to the Yeast Genome*: For both ribo-seq and RNA-seq data, the raw reads were processed for following procedures: (1) the low-quality reads were removed (using Cutadapt v4.1 with the parameter "-q 10""[35] (https://cutadapt.readthedocs.io/en/v4.1/installation.html); (2) reads lacking a 3′-adaptor (5′-TGGAATTCTCGGGTGCCAAGGAACTCC-3′) were removed, since they (> 46 nt) were much longer than the expectation (20-30 nt); (3) the 3′-adaptor sequences, as well as the 3 random bases at the 5′-end, were trimmed from the remaining sequences using Cutadapt; (4) trimmed sequences shorter than 10 nt were also exclude for future analysis.

To prevent contamination by ncRNA (mainly rRNA), we use Bowtie 2 v2.3.5.1[36] (https://bowtie-bio.sourceforge.net/bowtie2/index.shtml) to map and remove reads against the ncRNA in *S. cerevisiae* (http://ftp.ensembl.org/pub/release-107/fasta/saccharomyces_cerevisiae/ncrna/).

The remaining reads were aligned against the *S. cerevisiae* genome (http://ftp.ensembl.org/pub/release-107/fasta/saccharomyces_cerevisiae/dna/) using STAR v2.7.10a[37] (https://software.cqls.oregonstate.edu/updates/docs/STAR/). Reads were restricted to uniquely align to a single location in the genome, and a maximum of two mismatches were allowed.

Quantification of RPF and mRNA Abundance in Ribosome Profiling—Estimation of RPF and mRNA Abundance: The sequences filtered in the previous step, which varied in length from 12 to 56 nt, were subsequently used to compute the RPF and mRNA abundance for individual genes. Specifically, only those reads with their 5′- and 3′-ends precisely aligned within the ±50 nt boundaries of the coding sequence region were considered for further calculations. Both the RPF and RNA abundance were determined by the number of reads per kilobase of the coding sequence per million total reads (RPKM). The RPF and RNA abundance values were highly correlated (Figure 1F,G). To strengthen statistical robustness, we combined both replicates from the same strain in all subsequent data analyses.

We determined the proportion of 28-nt RPFs in the 5′-UTR (Figure 4C) by dividing the count of 28-nt RPFs within the 5′-UTR by the total count of 28-nt RPFs spanning the entire transcript for each mutant.

Quantification of RPF and mRNA Abundance in Ribosome Profiling—Calculation of TE and ΔTE: The TE and ΔTE were calculated using the following formulas:

$$TE = \frac{RPF\ abundance}{RNA\ abundance} \quad (1)$$

$$\Delta TE = \frac{TE_{mutant}}{TE_{WT}} \quad (2)$$

KEGG Pathway Enrichment Analysis: To identify biological pathways specifically regulated by looping-group and scanning-group initiation factors, we performed KEGG pathway enrichment analysis. For each mutant strain, the analysis was performed separately for looping- and scanning-dependent genes using the "clusterProfiler" package in R (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html).[38]

Identification of Differentially Expressed Genes (DEGs) Encoding Ribosomal Proteins: To detect the changes of ribosomal protein production upon perturbation of initiation factors, we individually compared the mutant of each initiation factor with the wild type. The RPF read counts of each ribosomal protein gene from both replicates were used as input for the "DESeq2" package in R[39] (https://bioconductor.org/packages/release/bioc/html/DESeq2.html). In the output result, the *P* values were calculated using false discovery rate.

UMAP-Based Dimensionality Reduction: UMAP dimensionality reduction analysis was performed using the "umap" package in R v4.2.3[40] (https://cran.r-project.org/web/packages/umap/index.html). To elucidate the specialized function of initiation factors, we analyzed data comprising ΔTE (translation efficiency fold change) values for each gene in every strain, all of which were log2 transformed prior to the analysis. To examine changes in ribosomal heterogeneity, we assessed the fold change in rRNA abundance for each strain. The UMAP analysis was performed with the following parameters: metric = cosine, min_dist = 0.001, spread = 0.05, n_neighbors = 3.

Calculation of the Frequency of Sequence-Based Features: The sequence-based features (Figure 2C) include transcript length, coding sequence length, 5′-UTR length, 3′-UTR length, poly(A) tail length, and the GC content within the coding sequence, 5′-UTR, and 3′-UTR regions. Additionally, we considered non-A nucleotide frequency in the poly(A) tail, the Kozak score, and three expression-related features: RPF abundance, mRNA abundance, and TE. Using CCC frequency (CCC%) as an example, we calculated the occurrences of CCC within each gene spanning the entire transcript length, then divided this count by the total number of 3-mers (Figure 2C). The procedures for calculating poly(A) tail length and the non-A frequency in the poly(A) tail are outlined in the next Section. Kozak scores were computed following the methodology detailed in the referenced article.[41]

Calculation of Poly(A) Tail Length and Non-A Nucleotide Frequency in the Poly(A) Tail using PAIso-Seq Data—PAIso-Seq Data Processing: From subreads data generated by PacBio sequel II, we used ccs v6.4.0 (https://github.com/PacificBiosciences/ccs/releases) to call circular consensus sequence with specific parameters: –min-rq 0.99 –min-passes 5. All other parameters were kept at their default values. Next, the 5′-adapter (5′-AAGCAGTGGTATCAACGCAGAGTACGGGGG-3′) and 3′-adapter (5′-AAGCAGTGGTATCAACGCAGAGTAC-3′) were removed using lima v2.6.0 (https://lima.how/) with the –peek-guess parameter. The output file of the previous step was then converted to fastq format for further processing. A GI-anchor sequence (5′-GGGGGGGGG-3′) was introduced at the 3′-end of the reads as part of the library preparation process. For the removal of these sequences, Cutadapt was used, with excision being carried out only when the end of the read sequence had a minimum of 8 bases matching the GI-anchor.

Calculation of Poly(A) Tail Length and Non-A Nucleotide Frequency in the Poly(A) Tail using PAIso-Seq Data—Calculation of Poly(A) Tail Length and Non-A Nucleotide Frequency in the Poly(A) Tail: To calculating ploy(A) tail length and non-A nucleotide frequency, we firstly create an index and align the clean circular consensus sequencing reads to the *S. cerevisiae* genome using the pbmm2 v1.9.0 (https://github.com/PacificBiosciences/pbmm2/releases) with the following parameters: –sort –preset ISOSEQ –log-level INFO. After alignment, soft-clips (i.e., sequences that do not align with the reference sequence) at the terminal of the reads were considered as ploy(A) tails if they met the following criteria: (1) the length of the soft-clip sequence ranged from 15 to 500 nt; (2) the frequency of adenine bases in the soft-clip sequence exceeded 50%; (3) the number of non-A nucleotide in the soft-clip sequence was less than 20.

To assign circular consensus sequencing reads to genes, we use the "findOverlaps" function from "GenomicAlignments" (https://bioconductor.org/packages/release/bioc/html/GenomicAlignments.html) package in R.[42] To accurately assess the poly(A) tail length, genes with poly(A) tails detected more than 10 times were considered for further analysis. Non-A nucleotide frequency in the poly(A) tail was calculated as the frequency of non-A bases in a ploy(A) tail. Considering the various

transcript isoforms, the poly(A) tail length and non-A nucleotide frequency in the poly(A) tail were determined by calculating the arithmetic mean of these values across all transcripts from the same gene. This approach provides a representative estimation of the poly(A) tail length and non-A nucleotide frequency in the poly(A) tail for each gene.

Calculation of rRNA Abundance: In yeast, there are four distinct types of rRNAs: 18S, 25S, 5S and 5.8S. Each of these rRNAs is coded by multiple copies of rDNAs, and the sequence of these copies are identical. To simplify the analysis, we selected specific variants for each of these rRNAs. The identifiers for these selected variants are as follows: 18S (YNCL0016C), 25S (YNCL0012C), 5S (YNCL0018W) and 5.8S (YNCL0014C).

When constructing a ribo-seq library, the presence of a ribosomal protein (RP) can impede the digestion of rRNA at nearby sites by RNase I. As a result, by examining ribo-seq data, we can use the changes in the coverage of each nucleotide within rRNA (compared to the wild type) to deduce alterations in the composition of ribosomal proteins following the disruption of initiation factors. To determine the coverage of each nucleotide within rRNA, we used STAR v2.7.10a to align the RPF reads to the rDNA, following the same mapping approach used for mRNA. Then, the "depth" function in Samtools v1.15[43] (https://www.htslib.org/) was utilized to compute the coverage at each nucleotide position. We calculated the ratio of the coverage for each nucleotide relative to the total number of coverage across all positions and all rRNAs. To quantify changes in the coverage of each rRNA nucleotide, the fold change between mutant and wild type was calculated.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

J.W. and G.Z. contributed equally to this work. The authors thank Shuo Zhang from Institute of Genetics and Developmental Biology, CAS for technical support. This work was supported by grants from the National Key Research and Development Program of China (2019YFA0508700) and the National Natural Science Foundation of China (32100443 and 31922014).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study have been deposited at Genome Sequence Archive under the accession number CRA012379 and CRA012393. Codes to analyze the data are available at Zenodo, [44] under the terms of the Creative Commons Attribution 4.0 international license.

Keywords

eukaryotic translation initiation factors, gene length, poly(A) tail length, ribosome heterogeneity, ribosome profiling, translational regulation, upstream open reading frame

Received: September 13, 2023
Revised: September 24, 2023
Published online: November 23, 2023

[1] a) R. J. Jackson, C. U. T. Hellen, T. V. Pestova, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 113; b) J. Costello, L. M. Castelli, W. Rowe, C. J. Kershaw, D. Talavera, S. S. Mohammad-Qureshi, P. F. G. Sims, C. M. Grant, G. D. Pavitt, S. J. Hubbard, M. P. Ashe, *Genome Biol.* **2015**, *16*, 10.
[2] a) F. Jacob, J. Monod, *J. Mol. Biol.* **1961**, *3*, 318; b) S. Brenner, F. Jacob, M. Meselson, *Nature* **1961**, *190*, 576.
[3] a) L. A. Passmore, T. M. Schmeing, D. Maag, D. J. Applefield, M. G. Acker, M. A. Algire, J. R. Lorsch, V. Ramakrishnan, *Mol. Cell* **2007**, *26*, 41; b) P. Yourik, C. E. Aitken, F. Zhou, N. Gupta, A. G. Hinnebusch, J. R. Lorsch, *eLife* **2017**, *6*, e31476; c) N. D. Sen, F. Zhou, M. S. Harris, N. T. Ingolia, A. G. Hinnebusch, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10464.
[4] a) N. R. Genuth, M. Barna, *Nat. Rev. Genet.* **2018**, *19*, 431; b) S. Xue, S. Tian, K. Fujii, W. Kladwang, R. Das, M. Barna, *Nature* **2015**, *517*, 33; c) K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S.-B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999; d) M. L. Truitt, C. S. Conn, Z. Shi, X. Pang, T. Tokuyasu, A. M. Coady, Y. Seo, M. Barna, D. Ruggero, *Cell* **2015**, *162*, 59; e) Z. Shi, K. Fujii, K. M. Kovary, N. R. Genuth, H. L. Röst, M. N. Teruel, M. Barna, *Mol. Cell* **2017**, *67*, 71.
[5] S. Z. Tarun, A. B. Sachs, *EMBO J.* **1996**, *15*, 7168.
[6] a) M. Altmann, *EMBO J.* **1997**, *16*, 1114; b) G. P. Cosentino, T. Schmelzle, A. Haghighat, S. B. Helliwell, M. N. Hall, N. Sonenberg, *Mol. Cell Biol.* **2000**, *20*, 4604.
[7] Q. Vicens, J. S. Kieft, O. S. Rissland, *Mol. Cell* **2018**, *72*, 805.
[8] T. Zhao, Q. Huan, J. Sun, C. Liu, X. Hou, X. Yu, I. M. Silverman, Y. Zhang, B. D. Gregory, C.-M. Liu, W. Qian, X. Cao, *Genome Biol.* **2019**, *20*, 189.
[9] a) P. Martin-Marcos, Y.-N. Cheung, A. G. Hinnebusch, *Mol. Cell Biol.* **2011**, *31*, 4814; b) P. Martin-Marcos, F. Zhou, C. Karunasiri, F. Zhang, J. Dong, J. Nanda, S. D. Kulkarni, N. D. Sen, M. Tamame, M. Zeschnigk, J. R. Lorsch, A. G. Hinnebusch, *eLife* **2017**, *6*, e31250; c) S. E. Wells, P. E. Hillner, R. D. Vale, A. B. Sachs, *Mol. Cell* **1998**, *2*, 135; d) L. M. Castelli, D. Talavera, C. J. Kershaw, S. S. Mohammad-Qureshi, J. L. Costello, W. Rowe, P. F. G. Sims, C. M. Grant, S. J. Hubbard, M. P. Ashe, G. D. Pavitt, *PLoS Genet.* **2015**, *11*, e1005233; e) A. Z. Andreou, U. Harms, D. Klostermeier, *RNA Biol.* **2017**, *14*, 113.
[10] N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, *Science* **2009**, *324*, 218.
[11] a) F. Zhou, H. Zhang, S. D. Kulkarni, J. R. Lorsch, A. G. Hinnebusch, *RNA* **2020**, *26*, 419; b) N. D. Sen, F. Zhou, N. T. Ingolia, A. G. Hinnebusch, *Genome Res.* **2015**, *25*, 1196; c) B. Zinshteyn, M. F. Rojas-Duran, W. V. Gilbert, *RNA* **2017**, *23*, 1365.
[12] a) D. E. Weinberg, P. Shah, S. W. Eichhorn, J. A. Hussmann, J. B. Plotkin, D. P. Bartel, *Cell Rep.* **2016**, *14*, 1787; b) M. V. Gerashchenko, V. N. Gladyshev, *Nucleic Acids Res.* **2014**, *42*, e134; c) N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **2011**, *147*, 789; d) J. Gardin, R. Yeasmin, A. Yurovsky, Y. Cai, S. Skiena, B. Futcher, *eLife* **2014**, *3*, e03735; e) N. R. Guydosh, R. Green, *Cell* **2014**, *156*, 950; f) D. D. Nedialkova, S. A. Leidel, *Cell* **2015**, *161*, 1606; g) C. Pop, S. Rouskin, N. T. Ingolia, L. Han, E. M. Phizicky, J. S. Weissman, D. Koller, *Mol. Syst. Biol.* **2014**, *10*, 770.
[13] M. Schuldiner, S. R. Collins, N. J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J. F. Greenblatt, J. S. Weissman, N. J. Krogan, *Cell* **2005**, *123*, 507.
[14] L. F. Lareau, D. H. Hite, G. J. Hogan, P. O. Brown, *eLife* **2014**, *3*, e01257.
[15] a) Y. Liu, H. Nie, H. Liu, F. Lu, *Nat. Commun.* **2019**, *10*, 5292; b) A. Tudek, P. S. Krawczyk, S. Mroczek, R. Tomecki, M. Turtola, K. Matylla-Kulinska, T. H. Jensen, A. Dziembowski, *Nat. Commun.* **2021**, *12*, 4951.
[16] K. Li, J. Kong, S. Zhang, T. Zhao, W. Qian, *Genome Biol.* **2022**, *23*, 254.

[17] C. Y. Chan, C. S. Carmack, D. D. Long, A. Maliyekkel, Y. Shao, I. B. Roninson, Y. Ding, *BMC Bioinformatics* **2009**, *10*, S33.
[18] S. Blum, S. R. Schmid, A. Pause, P. Buser, P. Linder, N. Sonenberg, H. Trachsel, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7664.
[19] F. Alkan, O. G. Wilkins, S. Hernández-Pérez, S. Ramalho, J. Silva, J. Ule, W. J. Faller, *Nucleic Acids Res.* **2022**, *50*, e95.
[20] A. Ben-Shem, L. Jenner, G. Yusupova, M. Yusupov, *Science* **2010**, *330*, 1203.
[21] a) S. Chen, K. Li, W. Cao, J. Wang, T. Zhao, Q. Huan, Y.-F. Yang, S. Wu, W. Qian, *Mol. Biol. Evol.* **2017**, *34*, 2944; b) V. Presnyak, N. Alhusaini, Y.-H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K. E. Baker, B. R. Graveley, J. Coller, *Cell* **2015**, *160*, 1111.
[22] a) G. Csárdi, A. Franks, D. S. Choi, E. M. Airoldi, D. A. Drummond, *PLoS Genet.* **2015**, *11*, e1005206; b) C. G. Artieri, H. B. Fraser, *Genome Res.* **2014**, *24*, 411; c) C. J. Mcmanus, G. E. May, P. Spealman, A. Shteyman, *Genome Res.* **2014**, *24*, 422.
[23] a) C. I. González, A. Bhattacharya, W. Wang, S. W. Peltz, *Gene* **2001**, *274*, 15; b) J. Kong, S. Zhang, W. Qian, K. Li, *J. Genet. Genomics* **2023**, *50*, 447.
[24] M. K. Thompson, M. F. Rojas-Duran, P. Gangaramani, W. V. Gilbert, *eLife* **2016**, *5*, e11154.
[25] J. D. Gross, N. J. Moerke, T. Von Der Haar, A. A. Lugovskoy, A. B. Sachs, J. E. G. Mccarthy, G. Wagner, *Cell* **2003**, *115*, 739.
[26] N. Arndt, D. Ross-Kaschitza, A. Kojukhov, A. A. Komar, M. Altmann, *Sci. Rep.* **2018**, *8*, 6707.
[27] a) A. A. Bazzini, F. Del Viso, M. A. Moreno-Mateos, T. G. Johnstone, C. E. Vejnar, Y. Qin, J. Yao, M. K. Khokha, A. J. Giraldez, *EMBO J.* **2016**, *35*, 2087; b) Y. Mishima, Y. Tomari, *Mol. Cell* **2016**, *61*, 874; c) G. Boël, R. Letso, H. Neely, W. N. Price, K.-H. Wong, M. Su, J. D. Luff, M. Valecha, J. K. Everett, T. B. Acton, R. Xiao, G. T. Montelione, D. P. Aalberts, J. F. Hunt, *Nature* **2016**, *529*, 358.
[28] a) A. Radhakrishnan, Y.-H. Chen, S. Martin, N. Alhusaini, R. Green, J. Coller, *Cell* **2016**, *167*, 122; b) R. Buschauer, Y. Matsuo, T. Sugiyama, Y.-H. Chen, N. Alhusaini, T. Sweet, K. Ikeuchi, J. Cheng, Y. Matsuki, R. Nobuta, A. Gilmozzi, O. Berninghausen, P. Tesina, T. Becker, J. Coller, T. Inada, R. Beckmann, *Science* **2020**, *368*, eaay6912; c) A. J. Veltri, K. N. D'orazio, L. N. Lessen, R. Loll-Krippleber, G. W. Brown, R. Green, *eLife* **2022**, *11*, e76038.
[29] a) F. He, X. Li, P. Spatrick, R. Casillo, S. Dong, A. Jacobson, *Mol. Cell* **2003**, *12*, 1439; b) H. Alalam, J. A. Zepeda-Martínez, P. Sunnerhagen, *RNA* **2022**, *28*, 905.
[30] L. Y. Chan, C. F. Mugler, S. Heinrich, P. Vallotton, K. Weis, *eLife* **2018**, *7*, e32536.
[31] K. Xiang, D. P. Bartel, *eLife* **2021**, *10*, e66493.
[32] H. Chang, J. Lim, M. Ha, V. N. Kim, *Mol. Cell* **2014**, *53*, 1044.
[33] a) S. K. Archer, N. E. Shirokikh, T. H. Beilharz, T. Preiss, *Nature* **2016**, *535*, 570; b) N. E. Shirokikh, S. K. Archer, T. H. Beilharz, D. Powell, T. Preiss, *Nat. Protoc.* **2017**, *12*, 697; c) J. Bohlen, K. Fenzl, G. Kramer, B. Bukau, A. A. Teleman, *Mol. Cell* **2020**, *79*, 561.
[34] G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.-D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, et al., *Nature* **2002**, *418*, 387.
[35] M. Martin, *EMBnet.journal* **2011**, *17*, 10.
[36] B. Langmead, S. L. Salzberg, *Nat. Methods* **2012**, *9*, 357.
[37] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, *Bioinformatics* **2013**, *29*, 15.
[38] a) G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, *OMICS* **2012**, *16*, 284; b) T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, *Innovation (Camb)* **2021**, *2*, 100141.
[39] M. I. Love, W. Huber, S. Anders, *Genome Biol.* **2014**, *15*, 550.
[40] J. H. Leland McInnes, J. Melville, (Preprint) arXiv:1802.03426, v3 **2020**.
[41] S. Lee, B. Liu, S. Lee, S. X. Huang, B. Shen, S. B. Qian, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2424.
[42] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, *PLoS Comput. Biol.* **2013**, *9*, e1003118.
[43] P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, *GigaScience* **2021**, *10*, giab008.
[44] J. Wang, G. Zhang, W. Qian, K. Li, *Zenodo* **2023**.

