A Tale of Two Cities: eIF4G1 and eIF4G2's Unexpected Mechanisms in Translation Initiation

Meghna Pramoda, Siona Pramoda, & Sam Perli
Harvard College '29, Stanford '30, Massachusetts Institute of Technology Ph.D. '16

This study dissects the roles of two isoforms of eIF4G, eIF4G1 and eIF4G2, by combining in vitro luciferase assays on 48 putative transcripts with computational analysis. Unexpectedly, most putative eIF4G1-associated transcripts maintained or increased translation following eIF4G1 knockdown, suggesting that their translation can persist through alternative initiation routes when eIF4G1 levels are reduced. Similarly, a majority of putative eIF4G2-associated transcripts retained or increased translation following eIF4G2 knockdown, while a smaller subset exhibited reduced translation consistent with stronger eIF4G2 dependence. Notably, the addition of recombinant Nat1—the murine ortholog of human eIF4G2 (also known as DAP5/p97), a protein implicated in non-canonical translation initiation—partially altered translation output in a subset of eIF4G2-associated conditions, though no statistically significant global rescue relative to eIF4G2 knockdown was observed. Comprehensive motif discovery revealed a highly conserved GC-rich 16-nucleotide sequence whose composition is consistent with RNA G-quadruplex-forming sequences. Together, these results suggest that eIF4G isoforms act as context-dependent regulators of translational plasticity, with potential implications for stem cell biology and translational dysregulation in disease.

Introduction

Eukaryotic translation initiation is an intricate process. At least eleven eukaryotic initiation factors (eIFs), proteins that assist in assembling the ribosome on a messenger RNA (mRNA), work together to recruit the ribosome to mRNA and to position it at the start codon. First, they form a 43S pre-initiation complex recruited to the mRNA's 5' cap by the eIF4F complex. The eIF4F complex is made up of eIF4E (a cap-binding protein), eIF4G (a scaffolding protein that holds the complex together), and eIF4A (an RNA helicase). eIF4G1 is the predominant isoform of eIF4G and is known to be associated with internal ribosome entry site (IRES)-dependent translation, a cap-independent mechanism in which ribosomes are recruited directly to a structured RNA element within the mRNA, bypassing the need for a 5' cap.

As summarized in Figure 1, translation initiation can occur through cap-dependent mechanisms, relying on the eIF4F complex binding to the mRNA cap, or through cap-independent mechanisms, which utilize internal ribosome entry sites (IRES) or other structures to recruit ribosomes directly. Figure 1 highlights several key differences in the protein complexes involved in the initiation of translation, with Figure 1A showing canonical cap-dependent translation and Figure 1E showing IRES-mediated translation.

This study examines human embryonic stem cells as a defined model system and asks how candidate eIF4G1- and eIF4G2-associated transcripts behave when each factor is reduced. Additionally, this study examines whether adding recombinant Nat1, whose deletion impairs differentiation of mouse embryonic stem cells and whose protein complex lacks the eIF4E-binding capacity required for canonical cap-dependent translation, can modulate translational activity in eIF4G2 knockdowns (Sugiyama et al., 2017).

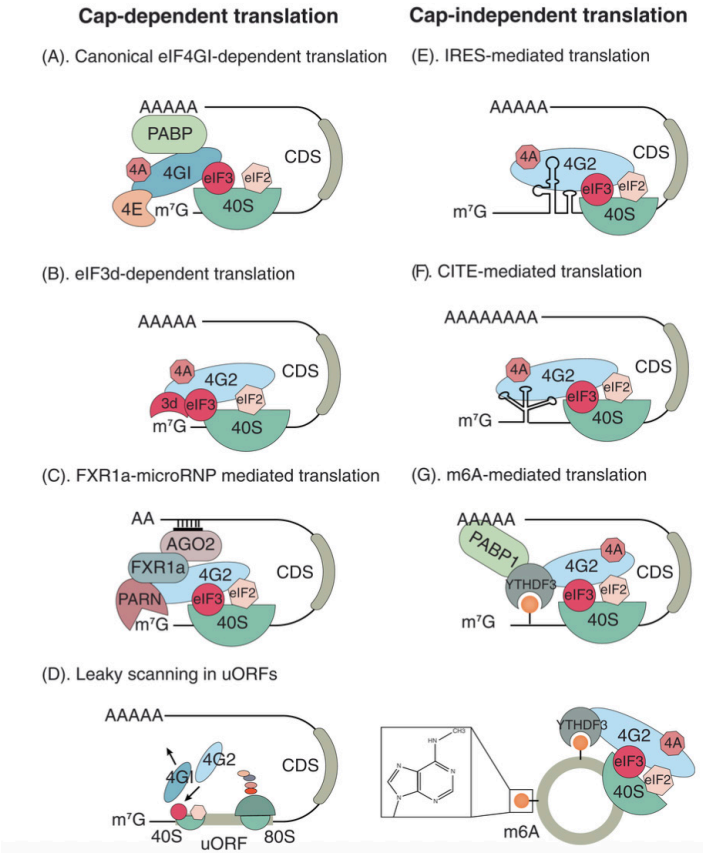

Figure 1. Known types of cap-dependent and cap-independent translation initiation (Liu et al., 2023).

Literature Review

Prior studies establish that eIF4G-family proteins are important organizers of translation initiation, but they also suggest that the two isoforms are not functionally identical. eIF4G1 is abundant and strongly associated with canonical cap-dependent initiation, whereas eIF4G2 has been implicated in specialized modes of initiation that become especially relevant for selected transcripts or under stress-related conditions.

Popular techniques to study the role of eIF4G have included in vitro translation (Smirnova, 2022) and ribosome profiling to examine how eIF4G proteins interact with ribosomes during translation (Mohammed, 2021) and polysome microarray analysis (Clarkson, 2010). Other techniques include proteomics and interactome analysis of eIF4G1 and eIF4G2 binding partners (Meril, 2023) and genetic manipulation of yeast and mammalian cell lines to clarify protein activities (Badura, 2012; Cunningham, 2023).

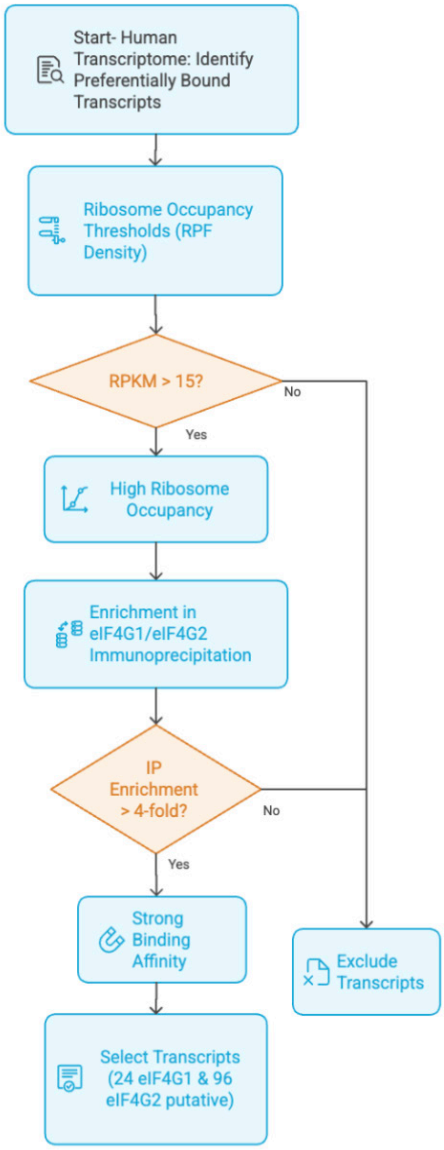

Figure 2. Schematic of the transcript-selection workflow for candidate eIF4G1- and eIF4G2-associated 5′ UTRs.

Current findings have led to multiple areas of ongoing debate. For example, the degree of functional overlap between eIF4G1 and eIF4G2 continues to be debated; some studies propose specific responsibilities (Cunningham, 2023), whereas others imply possible redundancy (Clarkson, 2010). The significance of eIF4G2 in cap-independent translation is also contentious (Liu, 2023; Smirnova, 2022). Other sources of debate include the function of eIF4G isoforms in oxidative stress responses (Cunningham, 2023; Badura, 2012), the roles of eIF4G1 and eIF4G2 in cancer (Jaiswal, 2018; Meril, 2023), and the regulation of particular mRNA subsets (Smirnova, 2022).

Compelling arguments arise from research that integrates many methodologies, including genetic manipulation, in vitro experiments, and proteomics. Smirnova's (2022) research offers strong evidence for eIF4G2's involvement in leaky scanning and reinitiation, corroborated by in vitro and cellular experiments. The evolving perspective of eIF4G2 as a versatile factor engaged in several translation processes is gaining prominence (Liu, 2023). This viewpoint harmonizes apparently conflicting results and proposes a more sophisticated comprehension of translation regulation.

While the literature review provides valuable insights, there are still significant gaps in the understanding of eIF4G1 and eIF4G2 functional redundancy in human cells. Transcript-specific effects of eIF4G1 and eIF4G2 knockdowns or knockouts have not been thoroughly investigated using RNA-seq in human cells. Accordingly, this study asked whether candidate eIF4G1-associated transcripts show reduced translation when eIF4G1 is knocked down, whether candidate eIF4G2-associated transcripts would show reduced translation following eIF4G2 knockdown, and whether Nat1 can partially restore translation in the eIF4G2-knockdown condition.

Hypothesis

Three hypotheses were tested:

i. Candidate eIF4G1-associated transcripts would show reduced translation after eIF4G1 knockdown.
ii. Candidate eIF4G2-associated transcripts would show reduced translation following eIF4G2 knockdown.
iii. In the eIF4G2-knockdown condition, addition of Nat1 protein would partially restore translation of candidate eIF4G2-associated transcripts.

Methods

The study combined a wet-lab workflow and a computational analysis pipeline. The wet-lab component moved from transcript selection, to reporter-template preparation and lysate preparation (described in Supplementary Materials), to knockdown validation, and finally to luciferase-based in vitro translation measurements. The computational component was then used to analyze motifs, localization patterns, enrichment, and related sequence features among the candidate transcripts. Human embryonic stem cells from the H9 line (WAe009-A; RRID:CVCL_9773) were used throughout.

Gene Selection

Twenty-four putative eIF4G1-associated 5' UTRs and twenty-four putative eIF4G2-associated 5' UTRs were chosen from Sel-TCP-Seq data analysis, a technique that identifies mRNAs

associated with specific translation initiation factors. Ribosome density was calculated for each candidate transcript; only transcripts with RPKM (Reads per Kilobase per Million) ≥ 15 were retained, to exclude transcripts with low translational activity or background noise. Immunoprecipitation enrichment was then used as a second filter with a fold-enrichment threshold greater than 4x relative to input.

Using those criteria, 24 candidate eIF4G1-associated transcripts and 96 candidate eIF4G2-associated transcripts were initially identified, after which 24 transcripts from each group were advanced for downstream analysis. These transcript sets were used both for experimental reporter assays and for computational analyses aimed at identifying sequence motifs, structural features, and putative binding patterns that might help explain differences between canonical and non-canonical translation initiation.

Validate that knockdowns were successful: Western Blot Analysis

Protein levels of eIF4G1 and eIF4G2 in the mAID cell lines were analyzed using a Western blot - a standard technique in which proteins are separated by size, transferred to a membrane, and detected with antibodies - to confirm successful reduction of target protein levels in IAA-treated conditions relative to Dox controls. Protein detection was performed using rabbit primary antibodies against eIF4G1 and eIF4G2 followed by HRP-conjugated goat anti-rabbit secondary antibodies, with chemiluminescent detection on the Jess capillary-based Western blot platform. Knockdown was confirmed in the lysates as shown in Figure 3.

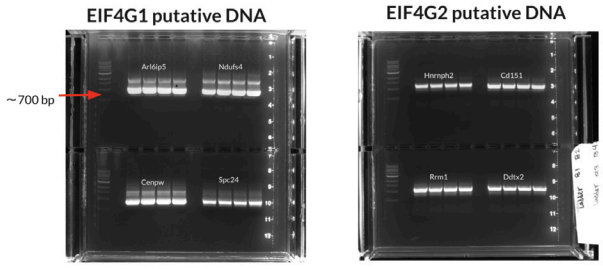

Figure 3. Western blot validation of eIF4G1 and eIF4G2 knockdown. Protein lysates from Dox (wild-type) and IAA-treated (knockdown) mAID cell lines were probed with antibodies against eIF4G1 and eIF4G2. Reduced band intensity in IAA-treated lanes relative to Dox controls confirms successful depletion of each target protein prior to use in luciferase assays (see Western Blot Analysis, Methods).

Experimental Variables

Independent variables were knockdown conditions (eIF4G1 mAID + 24-hour IAA treatment, eIF4G2 mAID + 24-hour IAA treatment, and eIF4G2 mAID + IAA + Nat1 rescue) and transcript groups (24 putative eIF4G1-dependent transcripts and 24 putative eIF4G2-dependent transcripts). The dependent variable was relative light units (RLUs), luminescence readout from the luciferase assay. Because luciferase signal is produced only after successful translation of the reporter, RLUs serve here as a proxy for translation-initiation output under each lysate condition. Baseline conditions were eIF4G1 mAID + Dox (wild type) and eIF4G2 mAID + Dox (wild type). Technical controls were identical lysate preparation protocols across conditions and uniform luciferase assay parameters (reagent concentrations, incubation times).

Western-blot confirmation of eIF4G1/eIF4G2 depletion was used to verify that the lysates reflected the intended experimental conditions. Candidate transcripts were chosen from prior Sel-TCP-Seq analysis and then tested in matched reporter-assay conditions. Western blot confirmation of eIF4G1/eIF4G2 depletion is shown in Figure 3 and described in the Western Blot Analysis subsection. For statistical analysis, outliers were identified using a 1.5×IQR rule, and a sensitivity analysis using a 2.0×IQR cutoff was performed to evaluate whether the main conclusions were robust to that choice. Because many fold change distributions did not satisfy normality assumptions, non-parametric tests were used where appropriate, and Benjamini-Hochberg correction was applied to control the false-discovery rate across multiple comparisons.

Test hypotheses by measuring translation levels: Luciferase Assay

Translation activity was measured using a luciferase assay. In this assay, the candidate 5' UTR is fused to the coding sequence of luciferase, a light-producing enzyme. When the reporter mRNA is successfully translated, luciferase is produced and generates a measurable luminescence signal. The amount of light emitted therefore serves as a direct proxy for translation initiation activity under each lysate condition. Each reaction combined 100 µL Nano-Glo Buffer, 10 µL prepared IVT Buffer, 20 µL of diluted lysate, and the corresponding reporter mRNA. The assay was incubated for 1 hour at 37°C before luminescence was measured using a plate reader.

To summarize the experimental design: for each of the 48 candidate transcripts, one IVT reaction was run per lysate condition (five conditions total: eIF4G1 Dox, eIF4G1 IAA, eIF4G2 Dox, eIF4G2 IAA, eIF4G2 IAA + Nat1). Each plate was read three times (technical triplicates). Fold change was computed as the ratio of RLUs in the knockdown or rescue condition to RLUs in the corresponding wild-type (Dox) condition for each transcript. Average SEM values across triplicate plate readings are shown in Figure 4.

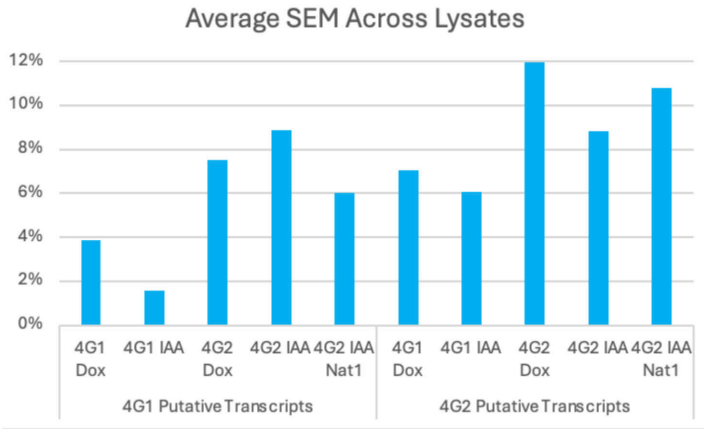

Figure 4. Technical replicate consistency across luciferase measurements. Average Standard Error of the Mean (SEM) across triplicate plate readings for each of the five lysate conditions, shown separately for eIF4G1-associated (n=24) and eIF4G2-associated (n=24) transcripts. SEM values below 12% indicate low within-experiment variability, supporting the reliability of the luminescence readout (see Luciferase Assay, Methods).

Hypothesis	Condition	% Upregulated	Mean Fold Change ± SEM (Pre-Outlier)	Mean Fold Change ± SEM (Post-Outlier)	Shapiro-Wilk P-value	Wilcoxon Statistic	Wilcoxon P-value
1: eIF4G1 Knockdown	IAA-4G1 / Dox-4G1	95.83%	6.50 ± 3.00	2.14 ± 0.16	0.000 (Normality NOT met)	0	0
2: eIF42 Knockdown	IAA-4G2 / Dox-4G2	58.33%	2.97 ± 0.89	1.87 ± 0.30	0.000 (Normality NOT met)	62.5	0.038
3A: Nat1 Restoration (vs. Dox-4G2)	Nat1 / Dox-4G2	66.67%	1.49 ± 0.17	1.38 ± 0.13	0.000 (Normality NOT met)	62	0.021
3B: Nat1 Restoration (vs. IAA-4G2)	Nat1 / IAA-4G2	37.50%	0.96 ± 0.13	0.96 ± 0.13	0.000 (Normality NOT met)	101	0.169

Table 1. Summary of translational changes and statistical analyses across the tested transcript groups. For each comparison, the table reports the number of transcripts analyzed before and after outlier removal, the mean fold change ± SEM, the percentage of transcripts with fold change greater than 1 (% upregulated), the Shapiro-Wilk normality result, and the Wilcoxon signed-rank p value with Benjamini-Hochberg-adjusted p value where applicable.

Computational Methods

All computational analyses were performed using a suite of tools that included custom Python code developed by the author or from the MEME Suite (version 5.5.7), along with complementary software packages. These tools were selected for their robust statistical frameworks and their ability to address various aspects of motif discovery, enrichment, localization, alignment, and annotation. The primary input sequences came from experimental data (48 transcripts), and corresponding control sequences were generated by shuffling the primary sequences while preserving k-mer frequencies, thereby maintaining intrinsic sequence characteristics. The integrated computation pipeline is described in the Supplementary Materials.

Selected Results

Translation initiation activity was quantitatively assessed by measuring the RLUs across 48 candidate transcripts under different lysate conditions (Figure 5). For the full fold-change summary across all transcripts and conditions, see Figure 5 and Table 1.

Hypothesis 1: Putative eIF4G1 transcripts would not be translated when eIF4G1 levels are significantly reduced.

Finding: The data do not support this hypothesis overall. Instead, most tested transcripts maintained or increased reporter output after eIF4G1 knockdown, as seen in Figure 6. Hypothesis 1 was therefore rejected.

A total of 24 transcripts were evaluated. Before outlier removal, the mean fold change was 6.50 ± 3.00 SEM, with 95.83% of transcripts exhibiting increased translation. After removing four outliers, twenty transcripts remained. These 20 transcripts had a mean fold change of 2.14 folds ± 0.16 SEM. The distribution of fold changes did not conform to normality (Shapiro-Wilk test, $W = 0.571$, $P < 0.001$), necessitating a Wilcoxon signed-rank test to test the hypothesis that the medians of the two samples differ.

The Wilcoxon signed-rank test indicated a statistically significant shift in translation after eIF4G1 knockdown ($P < 0.001$ see Table 1).

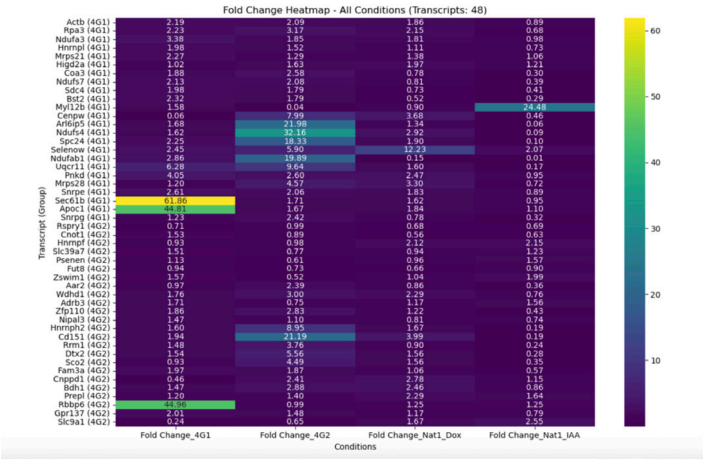

Figure 5. Summary of fold-change behavior across all 48 transcripts under four comparison conditions. (i) Fold change 4G1: eIF4G1 knockdown (IAA) vs. wild-type eIF4G1 (Dox); (ii) Fold change 4G2: eIF4G2 knockdown (IAA) vs. wild-type eIF4G2 (Dox); (iii-a) Fold change Nat1 Dox: Nat1-treated vs. wild-type eIF4G2 (Dox); (iii-b) Fold change Nat1 IAA: Nat1-treated vs. eIF4G2 knockdown (IAA). Values above 1 indicate increased translation in the first condition relative to the second.

In this dataset, the direction of that shift was predominantly upward rather than downward, indicating that the tested eIF4G1-associated transcripts did not behave as though they uniformly required eIF4G1 for reporter translation under these assay conditions. A secondary sensitivity analysis using a 2.0xIQR threshold reduced the number of excluded outliers from four transcripts to three. The qualitative direction of the result did not change, supporting the robustness of the main finding.

Hypothesis 2: Putative eIF4G2 transcripts would not be translated when eIF4G2 levels are significantly reduced.

Findings: Hypothesis 2 was partially supported.

A similar approach was applied to evaluate eIF4G2 knockdown effects. The initial dataset comprised 24 transcripts, with a mean fold change of 2.97 ± 0.89 SEM. Translation levels increased in

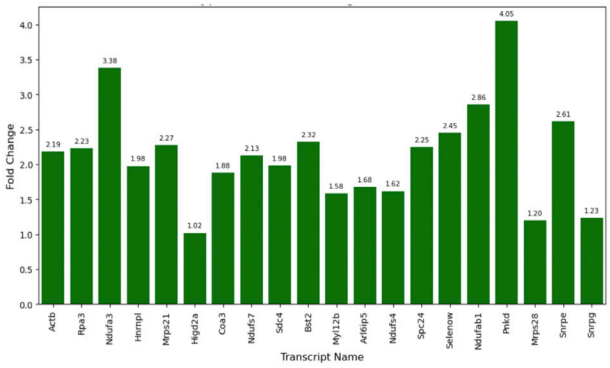

Figure 6. Fold changes in eIF4G1-associated transcripts upon eIF4G1 knockdown. Bar chart showing the ratio of RLUs in eIF4G1 knockdown (IAA) versus wild-type (Dox) conditions for each of the 20 remaining transcripts after outlier removal. Values above 1 indicate increased translation upon knockdown; values below 1 indicate decreased translation.

58.33% of transcripts and decreased in 41.67%. After removing two (2) outliers, a refined dataset of 22 transcripts remained with a mean fold change of 1.87 ± 0.30 SEM.

Similar to patterns seen with eIF4G1 putative transcripts, the distribution deviated from normality ($W = 0.781$, $p < 0.001$). The Wilcoxon signed-rank test confirmed a statistically significant alteration in translation ($p = 0.038$) and remained significant after Benjamini-Hochberg correction (adjusted $p = 0.038$) (see Table 1). It supports the observation that eIF4G2 knockdown significantly alters translation, though the effect is less pronounced compared to eIF4G1, as seen in Figure 7. The most upregulated transcripts included Cd151 (Fold Change (FC): 21.19), Hnrnph2 (FC: 8.95), and Dtx2 (FC: 5.56), while Zswim1 (FC = 0.52) exhibited a marked decrease (see Supplemental Figure 4).

A secondary sensitivity analysis using a 2.0xIQR threshold did not change the number of excluded outliers for Hypothesis 2. The qualitative direction of the result was unchanged, supporting the robustness of the main finding, although alternate-threshold summary statistics are not shown here.

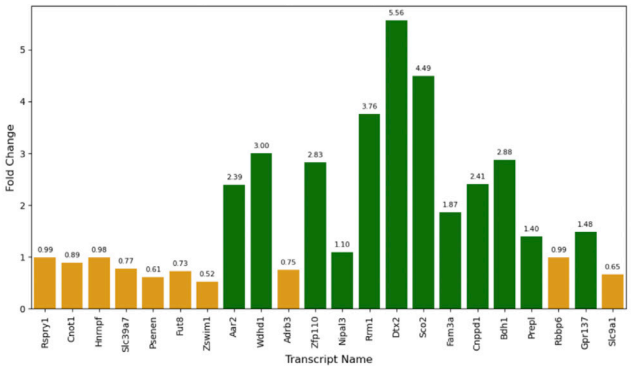

Figure 7. Fold changes in eIF4G2-associated transcripts upon eIF4G2 knockdown. Bar chart showing the ratio of RLUs in eIF4G2 knockdown (IAA) versus wild-type (Dox) conditions for each of the 22 remaining transcripts after outlier removal. Green bars indicate increased translation upon knockdown; orange bars indicate decreased translation.

Hypothesis 3: In eIF4G2 knockdown lysates, the addition of Nat1 protein will restore translational activity for putative eIF4G2 transcripts.

Findings: Our results partially support this hypothesis.

To assess whether Nat1 restores translation following eIF4G2 depletion, a comparative analysis of RLU values between Nat1 conditions and Dox (4G2) (Part A) and between Nat1 conditions and IAA (4G2) (Part B) was conducted.

In Part A, before outlier removal, the mean fold change was 1.49 ± 0.17 SEM, with 66.67% of transcripts showing increased translation. After removing one outlier, 23 transcripts remained, and the mean fold change was 1.38 ± 0.13 SEM. The data did not follow a normal distribution ($W = 0.623$, $p < 0.001$), and the Wilcoxon signed-rank test detected a statistically significant difference between the Nat1-treated condition and the Dox wild-type baseline ($p = 0.021$; adjusted $p = 0.042$ after Benjamini-Hochberg correction). This comparison shows that Nat1-treated lysates produce different translation output than untreated wild-type lysates, and indicates rescue relative to eIF4G2 knockdown - that comparison is addressed in Part B below. Notably, Cnppd1 (FC = 2.78) and Prepl (FC = 2.29) showed the largest upward shifts in this comparison (Figure 8).

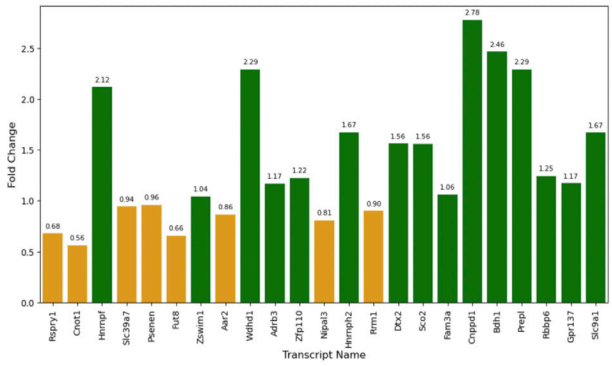

Figure 8. Fold changes comparing Nat1-treated to wild-type (Dox) eIF4G2 conditions (Part A). Bar chart showing the ratio of RLUs in the Nat1-treated condition versus the eIF4G2 Dox wild-type condition for each of the 23 remaining transcripts. Green bars indicate higher translation in the Nat1 condition; orange bars indicate lower. This comparison tests whether Nat1 shifts translation output relative to the untreated baseline, not relative to knockdown (see Part B, Figure 9).

In Part B, the mean fold change before and after outlier removal was 0.96 ± 0.13 SEM, with 37.50% of transcripts exhibiting increased translation and 62.50% showing decreased or unchanged translation. Following Benjamini-Hochberg correction (FDR < 0.05), the Wilcoxon signed-rank test remained non-significant (adjusted $p = 0.169$). The result indicates no statistically detectable difference in translation between Nat1-treated and eIF4G2-knockdown (IAA) conditions (adjusted $p = 0.169$; see Table 1, Figure 9). Accordingly, the data do not support a global rescue of eIF4G2-dependent translation by Nat1 under these conditions.

A secondary sensitivity analysis using a 2.0xIQR threshold did not change the number of excluded outliers for Hypothesis 3. The qualitative direction of the result was unchanged, supporting the robustness of the main finding, although alternate-threshold summary statistics are not shown here.

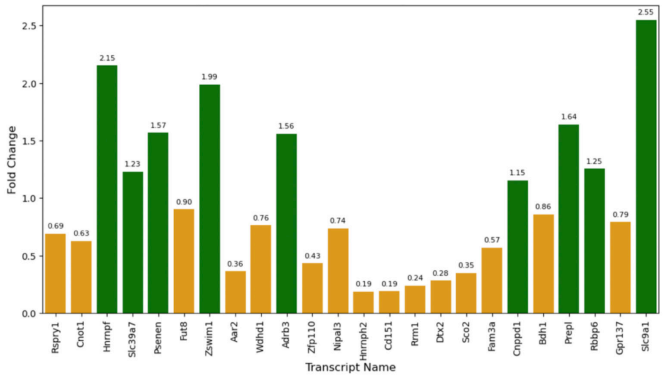

Figure 9. Fold changes comparing Nat1-treated to eIF4G2-knockdown (IAA) conditions (Part B). Bar chart showing the ratio of RLUs in the Nat1-treated condition versus the eIF4G2 IAA knockdown condition for each transcript. A ratio above 1 would indicate that Nat1 increased translation relative to knockdown. The Wilcoxon signed-rank test found no significant difference (adjusted p = 0.169), indicating no statistically detectable global rescue.

Comprehensive Computational Analysis Results

Out of the 48 transcripts tested, 40 showed increased translation when eIF4G1 or eIF4G2 was individually reduced. These are labeled Class 1. Seven transcripts selectively showed higher translation specifically under eIF4G1 knockdown conditions and displayed transcript-specific shifts in the Nat1-treated condition relative to eIF4G2 knockdown, but these shifts did not constitute a statistically significant global rescue (Part B adjusted p = 0.169); these are labeled Class 2. One additional transcript did not fit the defining pattern of either class and was treated as a Class 3 outlier, excluded from downstream computational analysis. Note that these classes are mutually exclusive - no transcript appears in more than one group. Because Class 2 contains only 7 transcripts, most computational results in that group did not reach statistical significance.

De Novo Motif Analysis

De novo motif analysis identified RNA structural features that enable transcripts to bypass canonical eIF4G1/2 dependency under stress. Using MEME, two key motifs were uncovered:

GC-Rich 16-nt Motif

The most statistically robust motif, a 16-nt sequence AGGAAGATGGAGGCCG (E-value = 3.7×10^{-6}), was present in 100% of transcripts (40/40). The motif's presence in 40 sites suggests a high level of functional significance. Many DNA-binding proteins recognize motifs in the 6-20 nucleotide range. Therefore, the width of 16 is within the expected range for regulatory function. The genes closest to the Class 1 motif include Gpr137, Cd151, Sdc4, Mrps21, Apoc1, Uqcr11, Bst2, Ndufab1, Snrpe, and Snrpg. Instances of the motif have consistent p-values ranging from 2.72e-4 to 1.70e-3, suggesting that these are likely bona fide binding sites (see Supplemental Figure 6).

Motif Enrichment Analysis Identifies YY1/YY2-Associated Motifs in Class 1 Transcripts

Motif enrichment analysis using AME (Analysis of Motif Enrichment) and SEA (Simple Enrichment Analysis) identified significant enrichment of transcription factor binding motifs in Class 1 transcripts. Among the enriched motifs were several associated with YY1 and YY2, which are best interpreted as candidate signals rather than as direct mechanistic proof of translation-initiation control. Enriched motifs identified by AME and SEA are shown in Supplemental Figures 7 and 8, respectively. RBPMap binding site analysis is also available in the Supplementary Materials.

In this dataset, enrichment of motifs annotated as YY1- or YY2-associated identifies candidate regulatory sequence features in Class 1 transcripts. The overlap between YY1/YY2-associated motif matches and transcripts with known stress-related functions (such as HIF1α and c-MYC) may be biologically interesting in the context of prior work on stress-responsive translation (Semenza, 2012), and these connections suggest context for future investigation.

Sequence Conservation and IRES Motif Analysis

To validate motif discovery and assess cap-independent translation potential, 40 Class 1 transcripts were analyzed using BLASTN/IRESite and Clustal Omega. Key findings include candidate IRES-like sequence similarities, evolutionary conservation patterns, and insights into structural limitations.

Two candidate IRES-like sequence similarities, c-MYC and HIF1α, were identified by BLASTN/IRESite. These matches provide directional guidance for studying IRES activity. The c-MYC mRNA contains an internal ribosome entry site (IRES) with a stem-loop III region that was identified by BLASTN analysis matching 12 plasmids with 100% identity (e.g., pSL200MLuc, E = 0.01). Clustal Omega alignment revealed conserved motifs in Zfp110 and Hnrnpl transcripts that resemble the c-MYC IRES region. Sequence similarities between the Class 1 transcripts and the c-MYC IRES region are suggestive of possible structural parallels and should be treated as candidate structural features for future investigation.

HIF1α mRNA contains an IRES-like site exhibiting 88% sequence identity to the hypoxia-responsive plasmid pRhifF as identified by BLASTN/IRESite (E = 0.5). Clustal Omega alignment revealed a conserved 16 of 18 base pair motif in Slc9a1 and Psenen transcripts. The absence of IRES-like matches in transcripts such as Actb and Snrpe is consistent with greater reliance on canonical initiation for those transcripts.

Deep evolutionary conservation of these motifs further underscores their functional importance. The c-MYC IRES motifs are conserved across diverse vertebrates, including humans, mice, cattle, sheep, chickens, and zebrafish, with structural elements like the 3' UTR Motif 17 preserved for over 350 million years. Similarly, HIF1α hypoxia-responsive motifs are conserved across species ranging from humans to lungfish, with origins predating the divergence of bony and cartilaginous fish approximately 450 million years ago. PhyloP analysis shows high conservation scores of 3.5–4.0 across Class 1 motifs, highlighting their evolutionary and functional significance.

To account for the numerous statistical tests conducted throughout the pipeline, all p-values were adjusted using the Benjamini–Hochberg (BH) correction. This procedure controlled the false discovery rate and minimized the risk of Type I errors, ensuring that the reported significant findings are statistically robust.

Discussion

Our wet lab experiments, alongside our integrated computational analysis pipeline, support a model in which eIF4G isoforms act as context-dependent regulators rather than universally essential scaffolds. While not conclusive, the data are consistent with how RNA structural motifs contribute to maintaining translation when canonical factors are reduced.

The computational analyses help generate preliminary hypotheses for how that flexibility might arise. Motif enrichment and structural predictions suggest that some transcripts may contain features compatible with non-canonical initiation, including G-rich elements and IRES-like signatures. Some of the GC-rich sequence features may also warrant future testing for possible G-quadruplex formation. At the same time, such standalone analyses do not establish mechanisms, and are candidates for future validation.

These findings extend the existing literature on eIF4G isoform function in important ways. Prior work in yeast established that eIF4G paralogs can target distinct mRNA subsets and differentially regulate translation under stress (Cunningham et al., 2023; Clarkson et al., 2010). Studies in mammalian cells have further suggested that eIF4G2 participates in non-canonical initiation modes including leaky scanning and reinitiation (Smirnova et al., 2022; Shestakova et al., 2023). The present results are consistent with those frameworks and extend them to human embryonic stem cells using a transcript-level reporter approach, though direct mechanistic comparisons with prior studies require care given differences in assay systems and model organisms.

Nat1 altered translation output in a subset of eIF4G2-associated conditions. However, direct comparison between Nat1-treated and eIF4G2-knockdown conditions was not statistically significant. The current dataset therefore supports a transcript-dependent modulatory effect in human embryonic stem cells. This is consistent with the established biology of Nat1 in stem cells: Sugiyama et al. (2017) showed that Nat1 promotes translation of specific differentiation-inducing proteins in mouse embryonic stem cells rather than acting as a general translation factor, which may explain why its effects here are transcript-selective rather than global.

More broadly, the study supports that eIF4G1 and eIF4G2 appear to participate in overlapping but non-identical initiation landscapes, and transcript structure likely helps determine which route remains available when one factor is reduced. That conclusion is consistent with the broader literature on translational plasticity under stress.

The computational framework complements the wet-lab findings by generating hypotheses about which sequence features might contribute to translational flexibility. The observation that 58.33% of eIF4G2-associated transcripts retained or increased translation after eIF4G2 knockdown is consistent with the presence of alternative initiation features in their 5' UTRs, and provides direction for future studies on the specific mechanism. In this context, it is important to note that CLIP-seq and similar approaches would be needed to directly map eIF4G-RNA interactions.

Implications

If future experiments confirm that the GC-rich motif identified here (AGGAAGATGGAGGCCG) forms a functional G-quadruplex structure—for instance via SHAPE-MaP or direct biochemical assay—such structures could represent candidate vulnerabilities in cancers reliant on stress-adapted translation. G-quadruplex-forming sequences in 5' UTRs have been independently shown to regulate translation of oncogenes such as c-MYC, and G-quadruplex structures have been explored as therapeutic targets in colorectal cancer and renal cell carcinoma (Neha et al., 2020; Bugaut & Balasubramanian, 2012). CX-5461 (Pidnarulex), a G-quadruplex-stabilizing small molecule currently in Phase I clinical trials for BRCA1/2-deficient solid tumors and with demonstrated preclinical activity in colorectal cancer models (Xu & Hurley, 2022; Otto et al., 2022), is one example of how G-quadruplex biology is being translated toward the clinic. Sequence features identified in the present study lend themselves to further validation in disease contexts.

In cancer biology, that possibility is preliminary. While current data are not sufficient to establish therapeutic targeting strategies, they do nominate transcript features and candidate genes that could be evaluated in more focused disease-specific studies. Finally, the identification of candidate eIF4G2-binding transcripts involved in processes such as zinc transport (Slc39a7) and protein ubiquitination (Rspry1) nominates specific pathways for further study in the context of stress-adapted translation, and could be therapeutically relevant in cancer.

Limitations

Potential G-quadruplex formation and other related RNA structures were inferred only indirectly from GC-rich sequence features and therefore require experimental confirmation. Methods such as SHAPE-MaP, mutational analysis, or orthogonal biochemical assays would be needed to determine whether any such elements are used during initiation in these transcripts.

The fixed 24-hour IAA treatment window may also obscure earlier or transient responses to factor depletion. Time-course experiments would be required to distinguish immediate initiation effects from later compensatory changes.

The study does not include patient survival data or disease-specific analyses; claims about biomarker utility or TCGA-derived hazard ratios are not made here, as no such analysis was conducted. Disease-oriented connections are exploratory and need a separate, fully described analysis pipeline.

Conclusion

This study provides evidence that eIF4G isoforms do not function as universally essential scaffolds for all transcripts, and instead participate in overlapping and context-dependent modes of translation initiation. Three key findings emerge: 1) translational flexibility, 2) candidate disease relevance, and 3) candidate eIF4G2 specialization.

Nearly all eIF4G1-associated transcripts maintained or increased translation after eIF4G1 knockdown, and a majority of eIF4G2-associated transcripts did likewise after eIF4G2 knockdown. This is consistent with the operation of alternative initiation features within specific 5' UTRs, though the precise mechanisms - which

may include G-quadruplex-forming motifs, IRES-like elements, or other structural features - require experimental confirmation.

Several transcripts elevated upon knockdown - including Apoc1 and Cd151 - have documented roles in cancer progression (Liu et al., 2021), and the GC-rich sequence motif identified across Class 1 transcripts is compositionally consistent with G-quadruplex-forming elements that have been studied as therapeutic targets in colorectal and renal cell carcinomas (Neha et al., 2020; Bugaut & Balasubramanian, 2012). G-quadruplex-targeting agents such as CX-5461 have shown preclinical activity in colorectal cancer (Otto et al., 2022) and are under clinical investigation. These connections are hypothesis-generating and need future validation.

RBPMap analysis (see Supplementary Materials) identified high-density predicted binding sites for eIF4G2 in transcripts involved in ion transport (Slc39a7) and protein ubiquitination (Rspry1), raising the hypothesis that eIF4G2 may preferentially associate with stress-relevant transcripts. This interpretation is preliminary and requires biochemical validation.

Future work could test individual candidate motifs by targeted mutagenesis in reporter constructs or by CRISPR-based editing in cellular models. Such experiments would help determine whether the motifs highlighted here are necessary for the observed translation patterns. Additional biochemical and structural approaches, including RNA-structure probing and direct assays of ribosome recruitment, could further clarify how Nat1, eIF4G1, and eIF4G2 contribute to transcript-specific initiation choices. Integrating those experiments with richer transcriptomic datasets may ultimately help distinguish which features are broadly generalizable and which are specific to particular cell states, stress conditions, or transcript classes.

Code Availability

The code for this project is available at https://github.com/chiwibeans/yamanaka_computational.

Supplementary Materials

Supplementary materials can be found online at thurj.com.

References

Adhikari, B., Narain, A., & Wolf, E. (2021). Generation of auxin-inducible degron (AID) knock-in cell lines for targeted protein degradation in mammalian cells. STAR Protocols, 2(4), 100949. https://doi.org/10.1016/j.xpro.2021.100949

Badura, M., Braunstein, S., Zavadil, J., & Schneider, R. J. (2012). DNA damage and eIF4G1 in breast cancer cells reprogram translation for survival and DNA repair mRNAs. Proceedings of the National Academy of Sciences, 109(46), 18767-18772. https://doi.org/10.1073/pnas.1203853109

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology (pp. 28-36). Menlo Park, CA: AAAI Press.

Bailey, T. L., & Grant, C. E. (2021). SEA: Simple enrichment analysis of motifs. bioRxiv. https://doi.org/10.1101/2021.08.23.457422

Bailey, T. L., & Machanick, P. (2012). Inferring direct DNA binding from ChIP-seq. Nucleic Acids Research, 40(17), e128. https://doi.org/10.1093/nar/gks433

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. Nucleic Acids Research, 43(W1), W39-W49. https://doi.org/10.1093/nar/gkv416

Bugaut, A., & Balasubramanian, S. (2012). 5′-UTR RNA G-quadruplexes: Translation regulation and targeting. Nucleic Acids Research, 40(11), 4727-4741. https://doi.org/10.1093/nar/gks068

Chen, L., Shioda, T., Coser, K. R., Sholl, L. M., Isselbacher, K. J., & Blumberg, R. S. (2010). Genome-wide analysis of YY2 versus YY1 target genes. Nucleic Acids Research, 38(12), 4011-4026. https://doi.org/10.1093/nar/gkq159

Clarkson, B. K., Gilbert, W. V., & Doudna, J. A. (2010). Functional overlap between eIF4G isoforms in Saccharomyces cerevisiae. PLOS ONE, 5(2), e9114. https://doi.org/10.1371/journal.pone.0009114

Costello, J. L., Kershaw, C. J., Castelli, L. M., Talavera, D., Rowe, W., Sims, P. F. G., Ashe, M. P., et al. (2017). Dynamic changes in eIF4F-mRNA interactions revealed by global analyses of environmental stress responses. Genome Biology, 18(1), 201. https://doi.org/10.1186/s13059-017-1338-4

Costello, J., Castelli, L. M., Rowe, W., Kershaw, C. J., Talavera, D., Mohammad-Qureshi, S. S., Ashe, M. P., et al. (2015). Global mRNA selection mechanisms for translation initiation. Genome Biology, 16(1), 10. https://doi.org/10.1186/s13059-014-0559-z

Cunningham, J., Sfakianos, A. P., Kritsiligkou, P., Kershaw, C. J., Whitmarsh, A. J., Hubbard, S. J., Grant, C. M., et al. (2023). Paralogous translation factors target distinct mRNAs to differentially regulate tolerance to oxidative stress in yeast. Nucleic Acids Research, 51(16), 8820-8835. https://doi.org/10.1093/nar/gkad568

De la Parra, C., Ernlund, A., Alard, A., Ruggles, K., Ueberheide, B., & Schneider, R. J. (2018). A widespread alternate form of cap-dependent mRNA translation initiation. Nature Communications, 9, 3068. https://doi.org/10.1038/s41467-018-05484-2

Liu, Y., Cui, J., Hoffman, A. R., & Hu, J.-F. (2023). Eukaryotic translation initiation factor eIF4G2 opens novel paths for protein synthesis in development, apoptosis and cell differentiation. Cell Proliferation, 56(3), e13367. https://doi.org/10.1111/cpr.13367

Endoh, T., & Sugimoto, N. (2016). Unusual RNA structures regulated by electrostatic interactions and their biological roles. Biophysical Reviews, 8(3), 211-218. https://doi.org/10.1007/s12551-016-0219-y

Feng, Y., Ni, Q., Wu, N., et al. (2024). Molecular mechanisms of MAZ targeting up-regulation of NDUFS3 expression to promote malignant progression in melanoma. Communications Biology, 7, Article 1491. https://doi.org/10.1038/s42003-024-07209-y

Feng, Y., Zhang, Y., Ying, C., Wang, D., & Du, C. (2014). Hypoxia promotes invasion and metastasis of gastric cancer through the HIF-1α pathway. Oncology Letters, 7(3), 879-883. https://doi.org/10.3892/ol.2014.1805

Frith, M. C., Saunders, N. F. W., Kobe, B., & Bailey, T. L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. PLoS Computational Biology, 4(5), e1000071. https://doi.org/10.1371/journal.pcbi.1000071

Ghosh, S., & Lasko, P. (2015). Loss-of-function analysis reveals distinct requirements of the translation initiation factors eIF4E, eIF4E-3, eIF4G and eIF4G2 in Drosophila spermatogenesis. PLOS ONE, 10(4), e0122519. https://doi.org/10.1371/journal.pone.0122519

Gordon, S., Akopyan, G., Garban, H., & Bonavida, B. (2006). Transcription factor YY1: Structure, function, and therapeutic implications in cancer biology. Oncogene, 25(8), 1125-1142. https://doi.org/10.1038/sj.onc.1209080

Grant, C. E., & Bailey, T. L. (2021). XSTREME: Comprehensive motif analysis of biological sequence datasets. bioRxiv. https://doi.org/10.1101/2021.09.02.458722

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. S. (2007). Quantifying similarity between motifs. Genome Biology, 8(2), R24. https://doi.org/10.1186/gb-2007-8-2-r24

Haimov, O., Sehrawat, U., Tamarkin-Ben Harush, A., Bahat, A., Uzonyi, A., Will, A., Hiraishi, H., Asano, K., & Dikstein, R. (2018). Dynamic interaction of eukaryotic initiation factor 4G1 (eIF4G1) with eIF4E and eIF1 underlies scanning-dependent and -independent translation. Molecular and Cellular Biology, 38(16), e00139-18. https://doi.org/10.1128/MCB.00139-18

Hinnebusch, A. G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiology and Molecular Biology Reviews, 75(3), 434-467. https://doi.org/10.1128/MMBR.00008-11

Jackson, R. J., Hellen, C. U. T., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nature Reviews Molecular Cell Biology, 11(2), 113-127. https://doi.org/10.1038/nrm2838

Jaiswal, P. K., Koul, S., Palanisamy, N., & Koul, H. K. (2019). Eukaryotic translation initiation factor 4 gamma 1 (EIF4G1): a target for cancer therapeutic intervention? Cancer Cell International, 19, 224. https://doi.org/10.1186/s12935-019-0947-2

Jaiswal, P. K., Koul, S., Shanmugam, P. S. T., & Koul, H. K. (2018). Eukaryotic translation initiation factor 4G1 is upregulated during prostate cancer progression and modulates cell growth and metastasis. Scientific Reports, 8(1), 7459. https://doi.org/10.1038/s41598-018-25798-7

Jiang, F., et al. (2024). eIF4G2-dependent translation of pro-fibrotic mRNAs is essential for cardiac fibroblast activation (Polysome-Seq) [Data set]. NCBI Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE239823

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Raisanen, T., Satpathy, S., Jolma, I. W., Kulakovskiy, I. V., Lairson, L. L., Taipale, M., Dave, K., Huss, M., & Taipale, J. (2013). DNA-binding specificities of human transcription factors. Cell, 152(1-2), 327-339. https://doi.org/10.1016/j.cell.2012.12.009

Lin, Y., McMahon, A., Driscoll, G., Bullock, S., Zhao, J., & Yan, S. (2021). Function and molecular mechanisms of APE2 in genome and epigenome integrity. Mutation Research/Reviews in Mutation Research, 787, 108347. https://doi.org/10.1016/j.mrrev.2020.108347

Liu, L., Zhang, X., & Liu, J. (2021). CD151 promotes colorectal cancer progression by enhancing TGF-β signaling pathway activation via integrin α3β1 binding. Journal of Cellular Biochemistry, 122(5), 493-504. https://doi.org/10.1002/jcb.29967

McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. BMC Bioinformatics, 11, 165. https://doi.org/10.1186/1471-2105-11-165

Melanson, G., Timpano, S., & Uniacke, J. (2017). The eIF4E2-directed hypoxic cap-dependent translation machinery reveals novel therapeutic potential for cancer treatment. Oxidative Medicine and Cellular Longevity, 2017, Article 6098107. https://doi.org/10.1155/2017/6098107

Meril, S., Bahlsen, M., Eisenstein, M., Savidor, A., Levin, Y., Bialik, S., Kimchi, A., et al. (2023). Loss-of-function cancer-associated mutations in the EIF4G2 non-canonical translation initiation factor. bioRxiv. https://doi.org/10.1101/2023.08.22.554280

Mohammad, M., Smirnova, A., Gunišová, S., & Valášek, L. (2021). eIF4G is retained on ribosomes elongating and terminating on short upstream ORFs to control reinitiation in yeast. Nucleic Acids Research, 49(15), 8743-8756. https://doi.org/10.1093/nar/gkab652

Neha, Das, P., & Verma, S. P. (2020). Dual role of G-quadruplex in translocation renal cell carcinoma: Exploring plausible cancer therapeutic innovation. Biochimica et Biophysica Acta (BBA) - General Subjects, 1864(12), 129719. https://doi.org/10.1016/j.bbagen.2020.129719

Otto, C., Mallm, J.-P., Krug, S., et al. (2022). CX-5461 combined with two immune checkpoint inhibitors enhances in vivo therapeutic efficacy by increasing PD-L1 expression in colorectal cancer. Cancer Letters, 526, 182-194. https://doi.org/10.1016/j.canlet.2021.11.027

Park, E.-H., Zhang, F., Warringer, J., Sunnerhagen, P., & Hinnebusch, A. G. (2011). Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. BMC Genomics, 12, 68. https://doi.org/10.1186/1471-2164-12-68

Poulin, F., & Sonenberg, N. (2013). Mechanism of translation initiation in eukaryotes. In Madame Curie Bioscience Database [Internet]. Austin, TX: Landes Bioscience. https://pmc.ncbi.nlm.nih.gov/articles/PMC4461372/

Semenza, G. L. (2012). Hypoxia-inducible factors. Cell, 148(3), 399-408. https://doi.org/10.1016/j.cell.2012.01.021

Shestakova, E. D., Smirnova, V. V., Shatsky, I. N., & Terenin, I. M. (2023). Specific mechanisms of translation initiation in higher eukaryotes: The eIF4G2 story. RNA, 29(3), 282-299. https://doi.org/10.1261/rna.079462.122

Smirnova, V. V., Shestakova, E. D., Bikmetov, D. V., Chugunova, A. A., Osterman, I. A., Serebryakova, M. V., Terenin, I. M., et al. (2019). eIF4G2 balances its own mRNA translation via a PCBP2-based feedback loop. RNA, 25(7), 757-767. https://doi.org/10.1261/rna.065623.118

Smirnova, V. V., Shestakova, E. D., Nogina, D. S., Mishchenko, P. A., Prikazchikova, T. A., Zatsepin, T. S., Terenin, I. M., et al. (2022). Ribosomal leaky scanning through a translated uORF requires eIF4G2. Nucleic Acids Research, 50(2), 1111-1127. https://doi.org/10.1093/nar/gkab1286

Spriggs, K. A., Bushell, M., & Willis, A. E. (2010). Translational regulation of gene expression during conditions of cell stress. Molecular Cell, 40(2), 228-237. https://doi.org/10.1016/j.molcel.2010.09.028

Sugiyama, H., Takahashi, K., Yamamoto, T., Iwasaki, M., Narita, M., Nakamura, M., Rand, T. A., Nakagawa, M., Watanabe, A., & Yamanaka, S. (2017). Nat1 promotes translation of specific proteins that induce differentiation of mouse embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America, 114(2), 340-345. https://doi.org/10.1073/pnas.1617234114

Szostak, E., & Gebauer, F. (2013). Translational control by 3′-UTR-binding proteins. Briefings in Functional Genomics, 12(1), 58-65. https://doi.org/10.1093/bfgp/els056

Unbehaun, A., Borukhov, S. I., Hellen, C. U. T., & Pestova, T. V. (2004). Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. Genes & Development, 18(24), 3078-3093. https://doi.org/10.1101/gad.327804

WAe009-A cell line information. (n.d.). hPSCreg. https://hpscreg.eu/cell-line/WAe009-A

Weber, R., Kleemann, L., Hirschberg, I., Chung, M.-Y., Valkov, E., & Igreja, C. (n.d.). DAP5 enables translation re-initiation on structured messenger RNAs [Unpublished manuscript]. Department of Biochemistry, Max Planck Institute for Developmental Biology.

Xu, H., & Hurley, L. H. (2022). A first-in-class clinical G-quadruplex-targeting drug. The bench-to-bedside translation of the fluoroquinolone QQ58 to CX-5461 (Pidnarulex). Bioorganic & Medicinal Chemistry Letters, 77, 129016. https://doi.org/10.1016/j.bmcl.2022.129016

Young, S. K., & Wek, R. C. (2016). Upstream open reading frames differentially regulate gene-specific translation in the integrated stress response. Journal of Biological Chemistry, 291(33), 16927-16935. https://doi.org/10.1074/jbc.R116.733899

Zinshteyn, B., Rojas-Duran, M. F., & Gilbert, W. V. (2017). Translation initiation factor eIF4G1 preferentially binds yeast transcript leaders containing conserved oligo-uridine motifs. RNA, 23(9), 1365-1375. https://doi.org/10.1261/rna.062059.117