PNAS RESEARCH ARTICLE | BIOCHEMISTRY OPEN ACCESS

Computational inference of eIF4F complex function and structure in human cancers

Su Wu[a,1] and Gerhard Wagner[a,1]

Edited by Natalie Ahn, University of Colorado Boulder, Boulder, CO; received August 10, 2023; accepted December 18, 2023

The canonical eukaryotic initiation factor 4F (eIF4F) complex, composed of eIF4G1, eIF4A1, and the cap-binding protein eIF4E, plays a crucial role in cap-dependent translation initiation in eukaryotic cells. An alternative cap-independent initiation can occur, involving only eIF4G1 and eIF4A1 through internal ribosome entry sites (IRESs). This mechanism is considered complementary to cap-dependent initiation, particularly in tumors under stress conditions. However, the selection and molecular mechanism of specific translation initiation remains poorly understood in human cancers. Thus, we analyzed gene copy number variations (CNVs) in TCGA tumor samples and found frequent amplification of genes involved in translation initiation. Copy number gains in *EIF4G1* and *EIF3E* frequently co-occur across human cancers. Additionally, *EIF4G1* expression strongly correlates with genes from cancer cell survival pathways including cell cycle and lipogenesis, in tumors with *EIF4G1* amplification or duplication. Furthermore, we revealed that eIF4G1 and eIF4A1 protein levels strongly co-regulate with ribosomal subunits, eIF2, and eIF3 complexes, while eIF4E co-regulates with 4E-BP1, ubiquitination, and ESCRT proteins. Utilizing Alphafold predictions, we modeled the eIF4F structure with and without eIF4E binding. For cap-dependent initiation, our modeling reveals extensive interactions between the N-terminal eIF4E-binding domain of eIF4G1 and eIF4E. Furthermore, the eIF4G1 HEAT-2 domain positions eIF4E near the eIF4A1 N-terminal domain (NTD), resulting in the collaborative enclosure of the RNA binding cavity within eIF4A1. In contrast, during cap-independent initiation, the HEAT-2 domain directly binds the eIF4A1-NTD, leading to a stronger interaction between eIF4G1 and eIF4A1, thus closing the mRNA binding cavity without the involvement of eIF4E.

eukaryotic initiation factor 4F (eIF4F) | cap-independent translation initiation | eIF4F dysregulation | eIF3e | computational analysis

Significance

Translation initiation is primarily governed by eIF4F, employing a "cap-dependent" mechanism, but eIF4F dysregulation may lead to a cap-independent mechanism in stressed cancer cells. We found frequent amplification of translation initiation genes and co-occurring copy number gains of *EIF4G1* and *EIF3E* genes in human cancers. *EIF4G1* amplification or duplication may be positively selected for its beneficial impact on the overexpression of cancer survival genes. The co-regulation of eIF4G1 and eIF4A1 protein levels, distinctly from eIF4E, reveals eIF4F dysregulation favoring cap-independent initiation. Alphafold predicts changes in the eIF4F complex assembly to accommodate both initiation mechanisms. These findings have significant implications for evaluating cancer vulnerability to eIF4F inhibition and developing treatments that target cancer cells with dependency on the translation initiation mechanism.

Translation initiation is the most crucial regulatory step in protein synthesis, where the eukaryotic initiation factor 4F (eIF4F) complex binds to activated mRNA and recruits ribosomes to translate it into a functional protein (1). Cancer cells rely heavily on eIF4F to drive aberrant protein synthesis that supports their survival, proliferation, and metastasis (2–4). The canonical eIF4F complex contains three core factors: the scaffold protein eIF4G1, the cap-binding protein eIF4E, and the RNA helicase eIF4A1. eIF4G1 has multiple binding domains, which variously can interact with the 5′UTR of mRNA, eIF4E, eIF4A1, the eIF3 complex, PABP1, and Mnk1/2 (5). When eIF4G1 binds to mRNA, its interaction with eIF4E can force proximity between eIF4E and the 7-methylguanosine cap at the 5′ end of mRNAs, stabilizing the cap and eIF4E interaction (6, 7). eIF4E in turn modulates eIF4G1's ability to stimulate eIF4A1, which unwinds the mRNA secondary structure for the ribosome attachment (8). The eIF3e subunit of the eIF3 complex directly interacts with eIF4G1, as a connection between eIF4G1 and the 40S ribosomal subunit (9, 10).

Overexpression or activation of eIF4F subunits by oncogenic signaling pathways can increase protein synthesis, promoting tumor growth (11). Cancer development involves genetic alterations enabling cell transformation, survival, drug resistance, and metastasis (12, 13). Selective genetic alterations in translation initiation genes are likely critical for tumor adaptation. However, quantification of positive selection in translation initiation across human cancers is lacking, impeding efforts to target eIF4F inhibition effectively.

eIF4F can initiate translation through a cap-dependent mechanism (with eIF4E) or a cap-independent mechanism (without eIF4E) (1, 14). eIF4E in its own right plays a dual role, in translation initiation, and the export of cell cycle gene mRNAs from the nucleus to cytoplasm through the nuclear pore complex (NPC) (15). Nutrient deprivation can cause eIF4E to accumulate in the nucleus in favor of mRNA transport, reducing its availability for eIF4F (16). Imbalanced expression of *EIF4G1* and *EIF4E* genes have been suggested to dysregulate cap-dependent initiation and favor cap-independent initiation

Author contributions: S.W. and G.W. designed research; S.W. performed research; S.W. contributed new analytic tools; S.W. analyzed data; G.W. conceptualized and critically reviewed the work for important intellectual content; and S.W. wrote the paper.

Competing interest statement: G.W. is a co-founder of and has equity in Enanta Pharmaceuticals, PIC Therapeutics, Eutropics, Olaris Therapeutics, Skinap Therapeutics, Cellmig Biolabs, NOW Scientific, Virtual Discovery, and QuantumTx.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: su_wu@hms.harvard.edu or gerhard_wagner@hms.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2313589121/-/DCSupplemental.

Published January 24, 2024.

in human cancers (17). Factors such as eIF4E binding protein 1 (4E-BP1) can facilitate cap-independent mechanisms under hypoxia (18) by preventing eIF4E from participating in eIF4F. On a molecular level, translation initiation typically requires eIF4A1 to interact with eIF4G1's HEAT-1 domain. For cap-independent initiation, eIF4A1 must furthermore interact with eIF4G1's HEAT-2 domain (19, 20). However, cellular pathways contributing to cap-independent initiation remain incompletely understood, and structural insight into eIF4F complex assembly for both mechanisms is limited. Addressing these questions is crucial for the development of cancer treatment drugs that target translation initiation.

To understand the importance of eIF4F components, we employed computational methods on large public datasets to investigate the impact of positive selection on eIF4F dysregulation in cancer. By analyzing copy number variation (CNV) and RNA-Seq data from over 10,000 tumors in The Cancer Genome Atlas (TCGA), we found that translation initiation genes often exhibit co-occurring amplification in tumors. We observed a strong correlation between the expression of *EIF4G1* and genes associated with cancer survival, in tumors with *EIF4G1* amplification or duplication (gain). Our analysis of proteomics data from the Cancer Cell Line Encyclopedia (CCLE) and CRISPR loss-of-function screen data from the Cancer Dependency Map (DepMap) revealed that eIF4E and eIF4G1 are essential for cancer cell survival but co-regulate with different protein complexes. Employing AlphaFold, we generated structural predictions for eIF4F complexes with and without eIF4G1–eIF4E binding. Our analysis revealed that the eIF4G1 HEAT-2 domain can adopt distinct conformations, facilitating both cap-dependent and cap-independent initiation mechanisms.

Results

Co-Occurring Copy Number Gain of Translation Initiation Genes in Tumors. Although the *EIF4G1* gene is frequently amplified in human cancers and is considered to be a driver gene (17), its frequency of amplification has not been assessed in the context of copy number variations (CNVs) across all genes in human cancers. To address this, we identified 703 genes amplified in a minimum of 5% of TCGA tumor samples across 33 cancer types (Fig. 1*A* and *SI Appendix*, Table S1). Gene amplification was less frequent than duplication (*SI Appendix*, Fig. S1*A*), as it often occurs through multiple steps following duplication (21). We focused on amplifications over duplications, because amplifications represent more persistent genomic changes under selection pressure (22).

Among the 703 most frequently amplified genes in human cancers, we found a significant enrichment of biological functions related to translation initiation, glycosylphosphatidylinositol-anchored protein synthesis, and FGFR2 signaling pathways (Fig. 1 *B* and *C*). Initiation factors such as *EIF4G1*, *EIF3E*, *EIF3H*, *EIF4A2*, *EIF2B5*, and *PABPC1*, and several genes encoding cellular and mitochondrial ribosomal large subunits, were among the most frequently amplified in TCGA cancers (*SI Appendix*, Fig. S1*B*). In addition, we analyzed the correlation of their copy number values across tumor samples. *EIF4G1, EIF4A2*, and *EIF2B5* display strong statistical associations (*SI Appendix*, Fig. S1*C*), likely due to their chromosomal proximity at 3q27. Similarly, we found strong statistical associations between *EIF3E*, *EIF3H*, and *PABPC1*, likely due to their chromosomal proximity at 8q23.

EIF4G1, *EIF4A2*, *EIF3E*, and *EIF3H* are frequently duplicated or amplified in most TCGA cancer types, particularly in lung squamous cell carcinoma and head & neck squamous cell carcinoma (Fig. 1*D* and *SI Appendix*, Figs. S1*D* and S2 *A* and *B*). Furthermore, we observed that a significant portion of the tumors had copy number gain in both *EIF4G1* and *EIF3E* (Fig. 1*E* and *SI Appendix*, Fig. S2*C*), even though they are located on different chromosomes. Fisher's exact test showed that the co-occurrence of copy number gains in these genes was significantly higher than what would be expected by chance (Fig. 1*F* and *SI Appendix*, Fig. S2*D*). We observed a comparable co-occurrence between the copy number gains of *EIF4G1* and *EIF3H*.

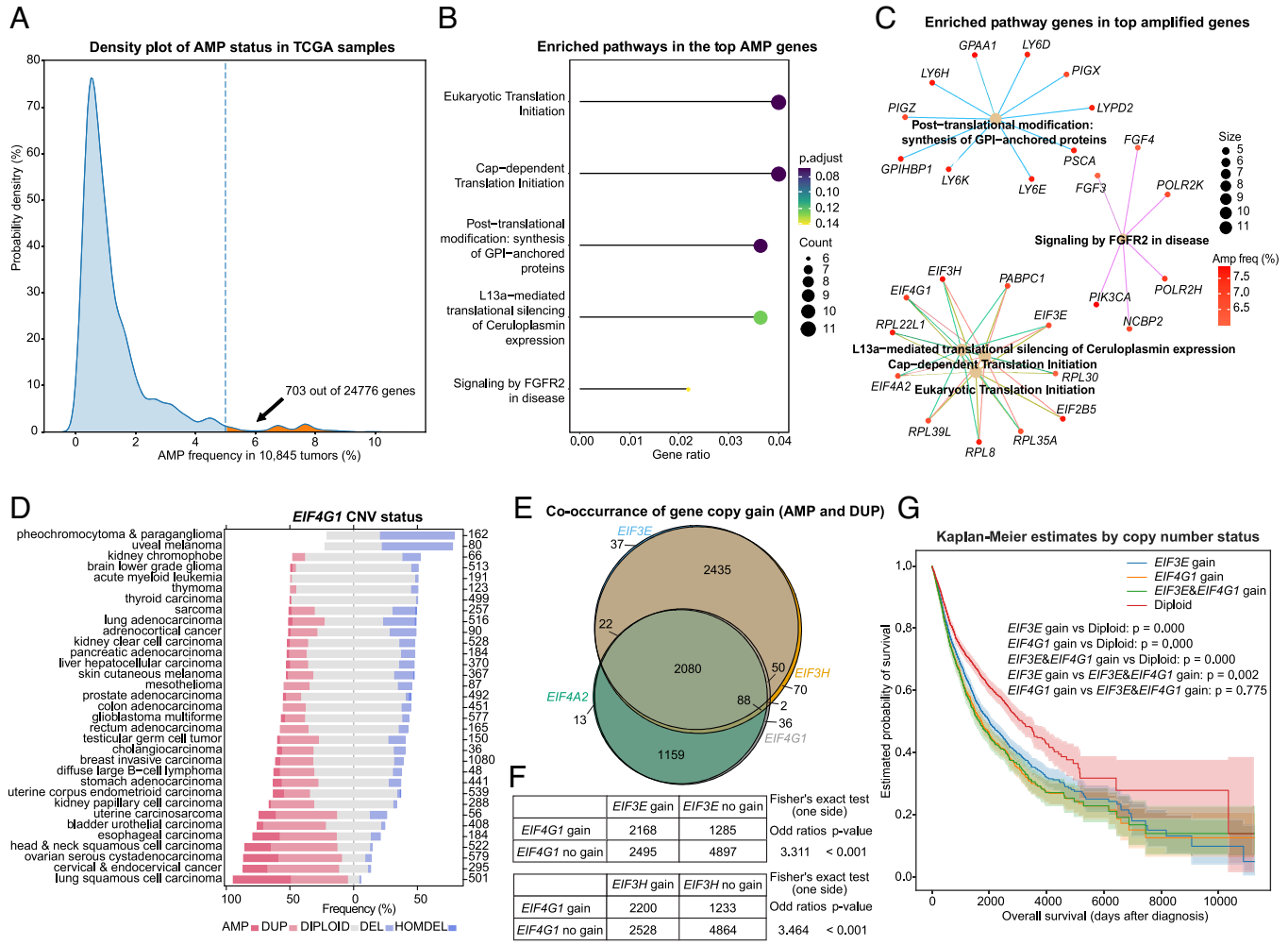
Kaplan–Meier analysis revealed that patients with copy number gains in either *EIF3E* or *EIF4G1* had significantly worse survival probabilities than those with diploid status for both genes in all TCGA cancer types (Fig. 1*G* and *SI Appendix*, Fig. S2*E*). Patients with copy number gains in both *EIF3E* and *EIF4G1* had even worse survival probabilities than those with gain in only *EIF3E*. Similar patterns were observed in many cancer types (*SI Appendix*, Fig. S3 *A* to *F*). These findings indicate that genes from the eIF4F and eIF3 complexes often exhibit co-occurring copy number gains, which is associated with cancer progression. Moreover, the benefits tumor cells derive from copy number gain in *EIF4G1* may depend on the co-occurring gain of other translation initiation genes.

A Strong Correlation of Expression between *EIF4G1* and Cancer Survival Genes in Tumors with *EIF4G1* Copy Number Gain. To determine whether the observed gene amplification is due to positive selection for the tumor-promoting effects of translation initiation genes or simply due to the susceptibility of these loci to amplification (23), we studied the cellular impact of initiation gene amplification on the cellular transcriptome. We identified genes that differentially correlate with *EIF4G1* mRNA expression in TCGA tumors categorized based on their *EIF4G1* CNV statuses and in GTEx healthy tissues.

We divided the correlating genes into five clusters (Fig. 2*A*) and performed pathway enrichment analysis (Fig. 2*B*). We found that the genes in "cluster 4" had strong positive correlations with *EIF4G1* expression in tumors that contained *EIF4G1* copy number gain. However, the strength of these correlations decreased in tumors with *EIF4G1* diploid or deletion, as well as in healthy tissues. In contrast, when we performed similar analyses on *EIF3E* and *EIF3H*, we did not identify any gene clusters that had stronger correlations in tumors with a gain of *EIF3E* or *EIF3H* (*SI Appendix*, Fig. S4 *A* to *D*).

Moreover, the cluster 4 genes from *EIF4G1* analysis (Fig. 2 *A* and *B*) are involved in the pathways crucial for cancer cell survival, such as regulation of cell cycle, cholesterol synthesis and lipogenesis, and cell adhesion pathways (24–26) (Fig. 2*C*). In contrast, the genes in clusters 3 and 5 had strong positive correlations with *EIF4G1* in healthy tissues, but not in tumors, and are involved in housekeeping pathways such as translation and ribosomal RNA processing (*SI Appendix*, Fig. S5 *A* and *B*). These findings suggest that *EIF4G1* positively influences pathways beneficial for cancer survival, especially in tumors with *EIF4G1* gain. However, in tumors, *EIF4G1*'s influence on housekeeping pathways is weaker compared to healthy tissues. Interestingly, in healthy tissues, cluster 1 and cluster 2 genes exhibit anti-correlations with *EIF4G1* expression (Fig. 2*A*), which include genes for neuronal proteins, G proteins, and ion channels.

To verify whether the observed differential expression correlations reflect differential gene expression, we used the dimensionality reduction algorithm, uniform manifold approximation and projection (UMAP) (27), to assess cluster 4 gene expressions in healthy tissues and tumors. The UMAP analysis clearly distinguished healthy tissue samples from tumor samples based on the differential expression of cluster 4 genes. Multiple distinct clusters were observed



	EIF3E gain	*EIF3E* no gain	Fisher's exact test (one side)
EIF4G1 gain	2168	1285	Odd ratios p-value
EIF4G1 no gain	2495	4897	3.311 < 0.001

	EIF3H gain	*EIF3H* no gain	Fisher's exact test (one side)
EIF4G1 gain	2200	1233	Odd ratios p-value
EIF4G1 no gain	2528	4864	3.464 < 0.001

Fig. 1. Translation initiation genes frequently co-amplify in TCGA tumor samples. (*A*) The Kernel density estimate plot shows the frequency of cellular gene amplification in 10,845 tumors across 33 cancer types in the TCGA database. The highlighted region includes genes amplified in more than 5% of tumors (right of dashed line). (*B*) The dot plot displays pathway enrichment analysis results for the top 703 frequently amplified genes identified in (*A*). Gene ratio represents the ratio of pathway-associated genes to total genes of interest. Dot size indicates the number of genes in each pathway. "*P*-adjust" is the adjusted *P*-value using a hypergeometric distribution test for multiple comparisons. (*C*) The star network plot illustrates the genes in the enriched pathways from (*B*). Central node size represents the number of genes in each pathway. Peripheral node color indicates the amplification frequency of corresponding genes. (*D*) The stacked bar plot displays *EIF4G1* copy number variation (CNV) status in different cancer types. (*E*) The Venn diagrams show the co-occurrence of copy number gain (amplification or duplication) for indicated genes in TCGA tumors. (*F*) presents the co-occurrence analysis of copy number gain events for two genes using a contingency table. A positive odds ratio indicates a trend toward co-occurrence. A one-sided Fisher's exact test was performed to test the null hypothesis of no relationship between the copy number gain events for the two genes. A low *P*-value suggests that the observed pattern is unlikely due to chance and more likely driven by a specific biological mechanism. (*G*) The Kaplan-Meier plot displays the survival probabilities of 10,746 TCGA patients with cancer based on gene copy gain of *EIF4G1* and/or *EIF3E* statuses within their tumors (indicated inside each box). The statistical significance of differences was determined by *P* values from log-rank tests. The shaded areas represent a 95% confidence region for the curve.

among healthy tissues (green dots in Fig. 2*D*), likely indicating tissue-specific separation. Most tumors with *EIF4G1* amplification or duplication (red or orange dots in Fig. 2*D*) formed separate groupings from tumors with *EIF4G1* diploid or deletion, indicating differential expression of cluster 4 genes within tumor groups. Additionally, UMAP was employed for the expression of clusters 3 and 5 genes in healthy tissues and tumors, which confirmed the differential expression of clusters 3 and 5 genes between healthy tissue and tumors (*SI Appendix*, Fig. S5 *C* and *D*).

Finally, we confirmed that the expressions of *EIF4G1* and several cluster 4 genes were significantly elevated in tumors with *EIF4G1* gain (Fig. 2 *E* to *J* and *SI Appendix*, Fig. S5 *E* to *H*), suggesting increased activities of cancer survival genes co-regulated with *EIF4G1*, particularly in tumors with *EIF4G1* gain. Altogether, these findings suggest that *EIF4G1* gain is likely a result of positive selection for its tumor-promoting effects.

Dysregulated eIF4F Complex Subunits Are Essential for Cancer Cell Survival. Co-occurring amplification of specific initiation factors may indicate dysregulation of initiation complexes in cancers. To investigate the dysregulation of eIF4F, we first analyzed the correlations among eIF4G1, eIF4A1, eIF4E, 4E-BP1, eIF3E, and eIF3H protein levels in 375 cancer cell lines from CCLE (28). Strong positive correlations were found between eIF4A1 and eIF4G1 (r = 0.632), eIF4E and 4E-BP1 (r = 0.534), eIF3E and eIF3H (r = 0.885), eIF4G1 and eIF3E (r = 0.634), eIF4G1 and eIF3H (r = 0.563), eIF4A1 and eIF3E (r = 0.679), and eIF4A1 and eIF3H (r = 0.650) (Fig. 3*A*). However, eIF4E showed weak positive correlations with eIF4G1 (r = 0.358) and no correlations with eIF4A1, eIF3E, or eIF3H.

We then identified cellular proteins strongly correlated with these proteins, grouped them into four clusters (Fig. 3*B*), and conducted pathway enrichment analysis (Fig. 3*C*). Cluster 3

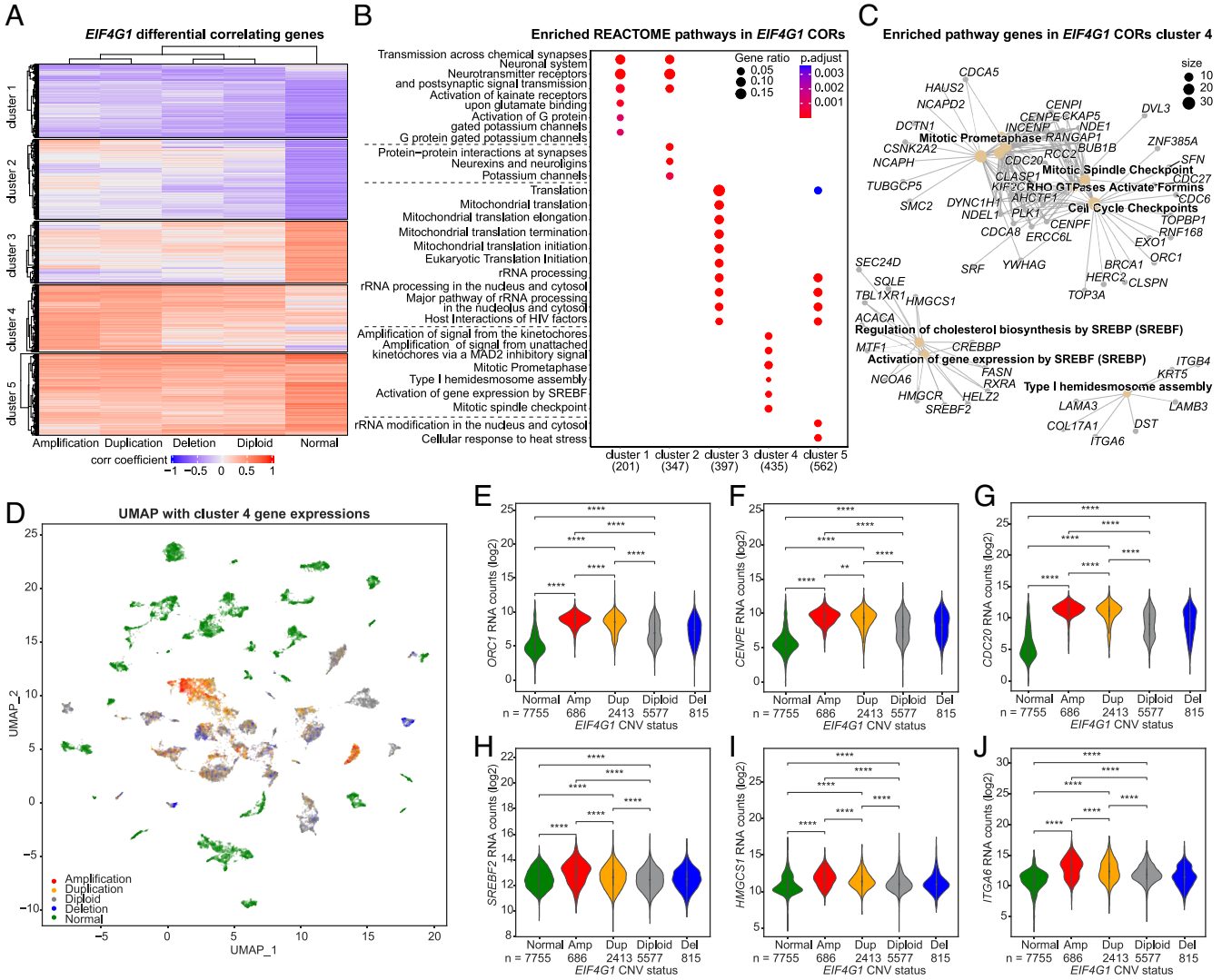


Fig. 2. A strong correlation of expression between *EIF4G1* and cancer survival genes in tumors with *EIF4G1* amplification or duplication. (*A*) The heatmap illustrates the differential expression correlation between cellular genes and *EIF4G1* in TCGA tumor samples with different *EIF4G1* CNV statuses. Pearson's correlation coefficients between *EIF4G1* and each of 58,582 other genes were calculated separately across 10,323 TCGA tumor samples with different *EIF4G1* CNV statuses [labeled as Amplification, Duplication, Diploid or Deletion (heterozygous and homozygous deletion)], or across 7,414 GTEx healthy samples from different tissue types (labeled as Normal). Genes with significant positive ($r > 0.3$) or negative ($r < -0.3$) correlations were selected for further analysis. Each row indicates the correlation of a gene with *EIF4G1* in the groups with the indicated *EIF4G1* CNV status. The heatmap cells' color and intensity correspond to Pearson's correlation coefficient values. The dendrogram at the top shows the hierarchical relationship between the columns. The rows were ordered and partitioned into five non-overlapping subgroups using a K-means clustering algorithm. (*B*) The dot plot displays the enriched pathways identified through REACTOME pathway analysis for the heatmap row clusters in (*A*). The six most significantly enriched pathways of each cluster are plotted, ranked by their adjusted *P*-values. (*C*) The star network plot illustrates the genes that belong to the significantly enriched pathways in cluster 4. (*D*) The UMAP plot illustrates the clustering of 10,323 TCGA tumor samples with varying *EIF4G1* CNV statuses, alongside 7,414 GTEx healthy samples from diverse tissue types. The clustering is based on the gene expression levels within cluster 4 genes identified from (*A*) within each sample. Normal (healthy) samples are shown in green, while other colors represent tumor samples with different *EIF4G1* copy number variation statuses. (*E–J*) The box plots compare the median expression of genes from cluster 4 in healthy samples and tumor samples with different *EIF4G1* copy number variation statuses. The two-tailed Student's *t* tests were performed. ns, not significant; $*P \le 0.05$; $**P \le 0.01$; $***P \le 0.001$; $****P \le 0.0001$.

proteins positively correlate with eIF4G1, eIF4A1, eIF3E, and eIF3H and are involved in translation initiation and ribosomal large and small subunits (*SI Appendix*, Fig. S6*A*). Cluster 4 proteins positively correlate with eIF4E and 4E-BP1, but not with eIF4G1, eIF4A1, eIF3E, or eIF3H. These proteins are involved in ubiquitination, nucleotide metabolism, and endosomal sorting complexes required for transport machinery (ESCRT) pathways (*SI Appendix*, Fig. S6*D*). Cluster 1 proteins negatively correlate with eIF4G1, eIF4A1, eIF3E, and eIF3H and are involved in mRNA splicing (*SI Appendix*, Fig. S7*A*), while cluster 2 proteins negatively correlate with eIF4E and 4E-BP1 and participate in extracellular matrix organization and viral infection pathways (*SI Appendix*, Fig. S8*A*). These findings indicate distinct co-regulation mechanisms for eIF4A1, eIF4G1, eIF3E, and eIF3H compared to eIF4E and 4E-BP1, suggesting dysregulation of cap-dependent initiation and potential cap-independent initiation mechanisms in cancer cells.

To identify the protein complexes involved in co-regulation pathways, we constructed a protein–protein interaction network using the STRING dataset (Fig. 3*D*). Cluster 3 contains numerous subunits from complexes such as the ribosome, eIF2, eIF3, and eIF4F. Our analysis of the cluster 3 network's degree centrality (Fig. 3*E* and *SI Appendix*, Fig. S6*B*) revealed extensive interactions among its proteins. Additionally, we investigated the relationship

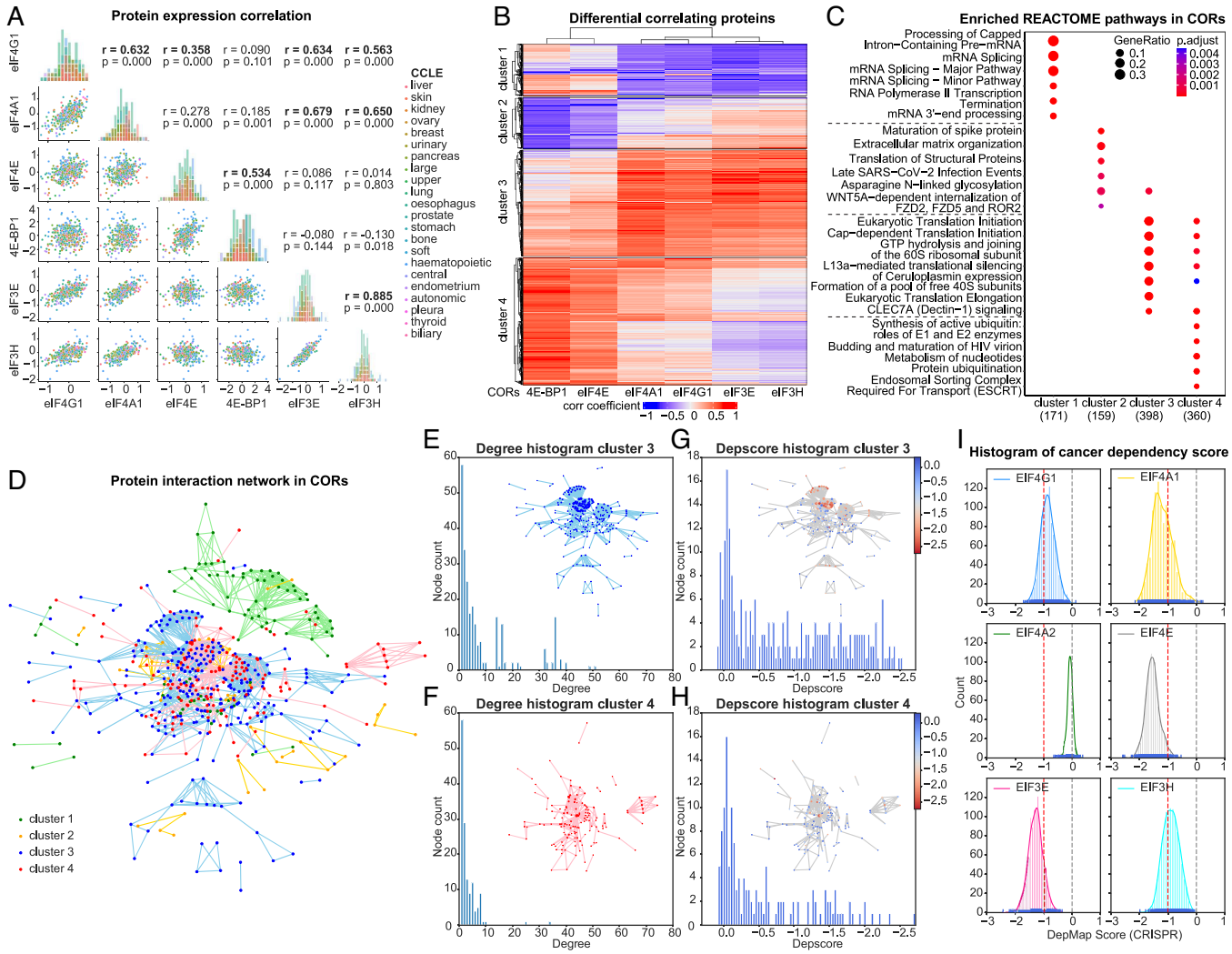


Fig. 3. The eIF4F complex components are vital for cancer cell survival and show dysregulation. (*A*) The scatter plots display the correlation between the protein expression levels of eIF4G1, eIF4A1, eIF4E, 4E-BP1, eIF3E, and eIF3H across CCLE cancer cell lines. Bold font denotes strong positive correlations. The colors in the plot represent the tissue origins of the cancer cell lines. Scatterplot axes show the protein expression levels (quantified as the relative abundance of detected peptides to reference, and log2 transformed) from CCLE proteomics data. The histograms show the binned expression distribution of each protein (by column) across cancer cell lines. For histograms, x-axes match scatterplot x-axes by column, y-axes (unlabeled) depict frequency counts, and coloring matches scatterplots. (*B*) The heatmap displays the differential expression correlation between cellular proteins and eIF4G1, eIF4A1, eIF4E, 4E-BP1, eIF3E, or eIF3H across CCLE cancer cell lines. Pearson's correlation coefficients between eIF4G1 and each of 12,755 other cellular proteins were calculated separately across 375 CCLE cancer cell lines. Proteins with strong positive ($r > 0.5$) or negative ($r < -0.5$) correlations were selected for analysis. (*C*) The dot plot shows the enriched pathways for the heatmap row clusters in (*B*), according to REACTOME pathway analysis. (*D*) The protein–protein interaction network shows the potential interactions between the proteins within clusters identified in the heatmap (*B*). The network was generated using evidence from the STRING database and includes proteins from all clusters. The colors in the network represent the cluster to which each protein (node) belongs and the potential interactions (edges) between them. (*E* and *F*) The histograms depict the distribution of node degree for the networks built from proteins in clusters 3 and 4 (as shown in the *Insets*). In the *Insets*, the node size reflects its degree—the number of connections it has to other nodes. (*G* and *H*) The histogram displays the distribution of the dependency score [a metric for gene essentiality measured through CRISPR knockout screens in cancer cell lines (29)] for each node in the network. In the *Insets*, the node size represents the degree, and the color reflects their median dependency score across CCLE cancer cell lines. (*I*) The kernel density plots show the distribution of dependency scores for a protein across all CCLE cancer cell lines. A lower score indicates that the gene is more likely to be essential for cancer cell survival in a particular cell line. A score of 0 indicates that the gene is not essential, while a score of −1 represents the median of all commonly essential genes.

between protein connectivity and essentiality by examining the depscores (30) of nodes. We found a significant number of high-connectivity protein nodes in cluster 3, and these proteins also displayed essential depscores (Fig. 3*G* and *SI Appendix*, Fig. S6*C*). Within the cluster 4 network, two ubiquitin ribosomal fusion proteins, Uba52 and Rps27A (31), were heavily connected to other proteins (Fig. 3*F* and *SI Appendix*, Fig. S6*E*) and are highly essential based on depscores (Fig. 3*H* and *SI Appendix*, Fig. S6*F*). These findings hint at a co-regulatory relationship between eIF4E activity and ubiquitination in cancer cells. Furthermore, the cluster 1 network exhibited modest connectivity with essential subunits, primarily consisting of the mRNA splicing complex (*SI Appendix*, Fig. S7 *B* to *E*). In contrast, the cluster 2 network showed low connectivity and lacked essential components (*SI Appendix*, Fig. S8 *B* to *E*).

Finally, to compare the essentiality of eIF4F and eIF3 subunits, we plotted the distribution of their depscores across CCLE cancer cell lines (Fig. 3*I*). eIF4G1, eIF4A1, eIF4E, eIF3E, and eIF3H are essential for viability in most cancer cell lines, whereas eIF4A2 is not. In summary, eIF4G1 and eIF4A1 co-regulate with essential ribosome, eIF2, and eIF3 complexes in cancer cell survival. Additionally, eIF4E strongly co-regulates with vital ubiquitin-synthesis complexes. These findings indicate dysregulations of eIF4E and eIF4G1 in cancer cells, suggesting the existence of

separate initiation mechanisms: cap-dependent and cap-independent, both crucial for cancer cell viability.

Predictive Structure Modeling Indicates Distinct Conformations of eIF4F Complex with and without eIF4E Binding. Using Alphafold multimer prediction, we modeled the two probable forms of the eIF4F complex crucial for translation initiation in cancer. We modeled the canonical eIF4F complex for cap-dependent initiation, consisting of eIF4G1$_{557–1437}$ [including eIF4E binding domain (4E-BD), HEAT-1, and HEAT-2], along with full-length eIF4A1 and eIF4E (Fig. 4*A*). We also modeled a dysregulated conformation of eIF4F for cap-independent initiation by including only eIF4G1$_{557–1437}$ and full-length eIF4A1.

In both eIF4G1•eIF4A1•eIF4E (Fig. 4 *B* and *C* and *SI Appendix*, Fig. S9 *A* and *B*) and eIF4G1•eIF4A1 complexes (Fig. 4 *D* and *E* and *SI Appendix*, Fig. S9 *C* and *D*), the two eIF4A1-binding domains in eIF4G1 (HEAT-1 and HEAT-2, colored yellow and orange, respectively) are modeled on opposite sides of eIF4A1. In both complexes, HEAT-1 interacts with the eIF4A1 C-terminal domain (CTD, light blue) and N-terminal domain (NTD, dark blue). However, in the eIF4G1•eIF4A1•eIF4E complex, HEAT-2 positions the cap-binding pocket of eIF4E adjacent to the RNA-binding cavity on eIF4A1-NTD (32). In contrast, in the eIF4G1•eIF4A1 complex, HEAT-2 directly interacts with eIF4A1-NTD. These findings indicate that in the absence of eIF4E, HEAT-2 takes a different conformation, resulting in a stronger interaction between eIF4G1 and eIF4A1.

Both the eIF4G1•eIF4A1•eIF4E and eIF4G1•eIF4A1 complexes (Fig. 5*A*) share three experimentally confirmed interaction interfaces between eIF4G1 and eIF4A1 (33). The first interface involves hydrogen bonds between residues R766 and N770 of HEAT-1-NTD and T269, L270, and I272 of eIF4A1-CTD (Fig. 5*B*). The second interface is characterized by Van der Waals interaction between F978 of HEAT-1-CTD and Y48 of eIF4A1-NTD (Fig. 5*C*). The

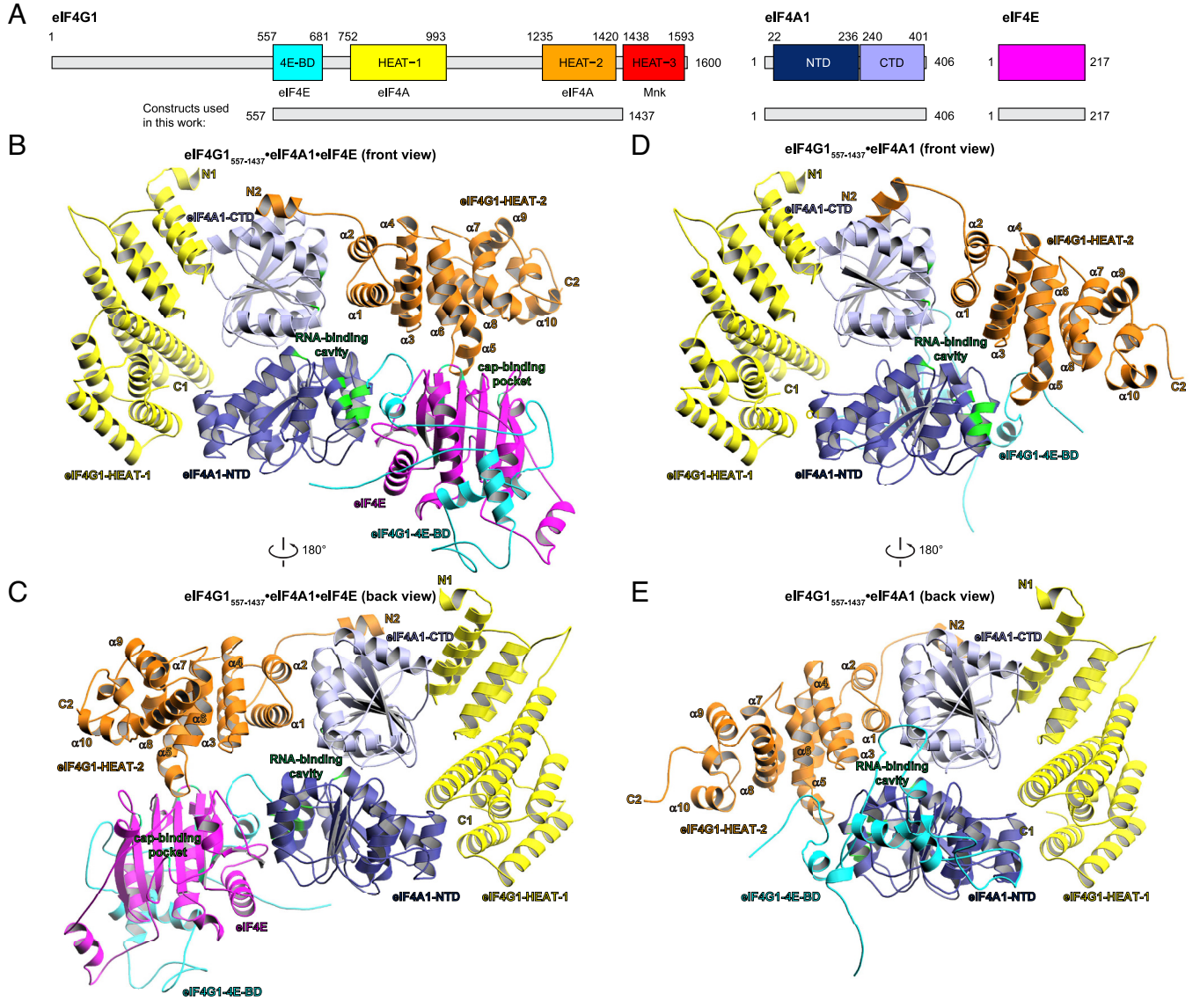


Fig. 4. The HEAT-2 domain of eIF4G1 adopts distinct conformations in eIF4F complexes with or without eIF4E. (*A*) The diagram illustrates the domain organization of eIF4G1, eIF4A1, and eIF4E constructs used in this work. (*B* and *C*) The cartoon representations of the structure of eIF4G1$_{557-1437}$•eIF4A1•eIF4E complex in two orientations, predicted by the AlphaFold2 multimer approach. Residues critical for RNA interaction and the formation of the RNA-binding cavity in eIF4A1 (R110, P159, G160, F163, D164, N167, Q195, D198, R282, D305, and G304) are highlighted in green. Additionally, residues crucial for the cap-binding pocket in eIF4E (W56, W102, E103, R112, R157, K159, K162, and W166) are highlighted in lime green. These residues are more visible from the back view compared to the front. (*D* and *E*) The predicted protein structures for the eIF4G1$_{557-1437}$•eIF4A1 complex in two orientations.

third interface involves salt bridges among R283 of eIF4A1-CTD, D1259, D1262, and Q1265 of the α-helix$_2$ of HEAT-2 (Fig. 5*D*).

In the eIF4G1•eIF4A1 complex, a fourth interface between eIF4A1-NTD and α-helix$_3$ and α-helix$_5$ of HEAT-2 was predicted, involving a hydrogen bond between E1289 of HEAT-2 and Q195 of eIF4A1-NTD (Fig. 5*E*). The proximity between eIF4A1-NTD and α-helix$_3$ and α-helix$_5$ of HEAT-2 has recently been experimentally confirmed in the HEAT-2•eIF4A1 complex (34), potentially contributing to the closure of the mRNA binding cavity in the cap-independent initiation mechanism. In contrast, two additional interactions were predicted in the eIF4G1•eIF4A1•eIF4E complex (Fig. 5*F*). K202 of eIF4A1-NTD interacts with V217 of eIF4E (Fig. 5*G*), while α-helix$_3$ and α-helix$_5$ of HEAT-2 interact with eIF4E (Fig. 5*H*), forming a hydrogen bond between E1289 of HEAT-2 and S209 of eIF4E. These two unverified interactions likely tether eIF4E close to the eIF4A1 mRNA binding cavity in the cap-dependent mechanism.

Moreover, in the eIF4G1•eIF4A1•eIF4E complex, three interaction interfaces are depicted in the eIF4G1-4E-BD•eIF4E complex (*SI Appendix*, Fig. S10 *A* and *D*). The first interface involves residues Q621, D626, D638, V639 and N645 on the non-canonical loop (NC-loop) of eIF4G1-4E-BD, interacting with E70, N77 and Q80 on the lateral side of eIF4E (*SI Appendix*, Fig. S10*B*). The second interface features R611, Y612, and Q621 on the canonical α-helix motif of eIF4G1-4E-BD, interacting with P38, G139, and D143 on the dorsal side of eIF4E (*SI Appendix*, Fig. S10*E*). These

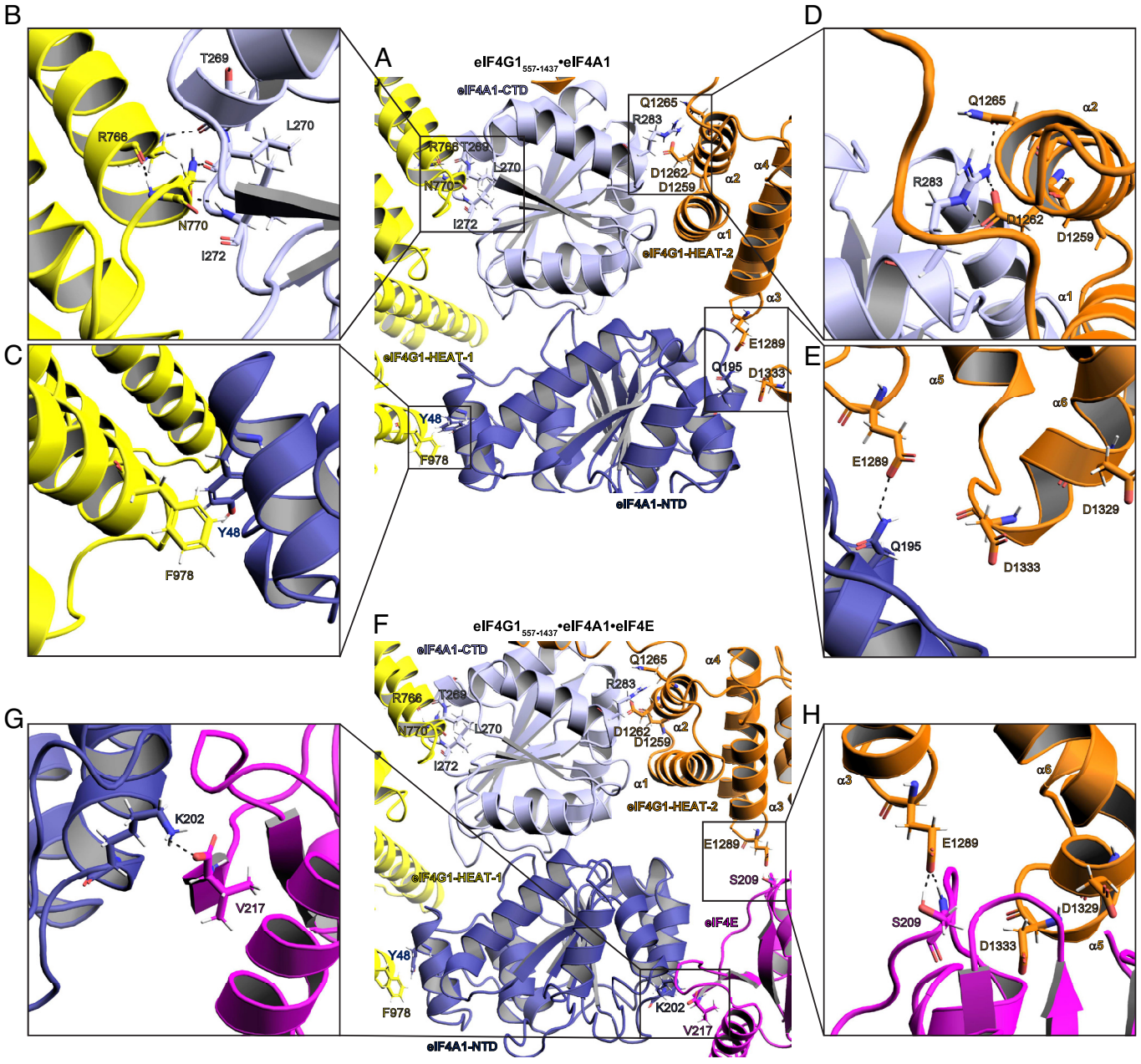


Fig. 5. Conformational stabilization of the eIF4F complex through multivalent interactions. (*A*) An overview of the four interaction interfaces between eIF4G1 and eIF4A1 in eIF4G1$_{557\text{-}1437}$•eIF4A1 complex. (*B*) Close-up view of hydrogen bonds (indicated by dashed black lines) between HEAT1-NTD and eIF4A1-CTD residues. (*C*) Close-up view of Van der Waals interaction between HEAT1-CTD and eIF4A1-NTD residues. (*D*) Close-up view of salt bridges formed between HEAT2 and eIF4A1-CTD residues. (*E*) Close-up view of hydrogen bonds between HEAT2 and eIF4A1-NTD residues. (*F*) An overview of the two interfaces between eIF4G1 and eIF4E residues, as well as eIF4A1 and eIF4E in the eIF4G1$_{557\text{-}1437}$•eIF4A1•eIF4E complex. The three interfaces between eIF4G1 and eIF4A1 remain identical to (*B*–*D*) (*G*) Close-up view of hydrogen bonds between eIF4A1-NTD and eIF4E residues. (*H*) Close-up view of hydrogen bonds between HEAT2 and eIF4E residues.

two interfaces have been experimentally confirmed (35–37). Additionally, a third unverified interaction was predicted between G663 and D665 of eIF4G1-4E-BD, and T116 of eIF4E (*SI Appendix*, Fig. S10*C*). These findings reveal extensive interactions between the N-terminal eIF4E-binding domain of eIF4G1 and eIF4E.

Discussion

The frequent amplification of translation initiation genes in TCGA tumor samples highlights their potential importance in tumor function. Our study identified translation initiation genes including *EIF4G1*, *EIF3E*, *EIF3H*, *EIF4A2*, *EIF2B5*, and *PABPC1*, as well as several ribosomal large-subunit genes, are co-located in two highly amplified chromosome loci, 3q27 and 8q23, in human cancers. The co-occurrence of copy number gain of *EIF4G1* at 3q27 and *EIF3E* at 8q23 from unlinked loci is noteworthy, as their encoded proteins play an essential role in ribosome recruitment on mRNA (9). Furthermore, we found that *PABPC1* proximal to *EIF3E* at 8q23 is also co-gained with *EIF4G1*. This co-occurrence aligns with the known biochemical interaction between their encoded proteins, which enhances initiation and ribosome recruitment (38). Functionally related metabolic genes often demonstrate genetic linkage by clustering closely in the genome to facilitate co-expression, making them susceptible to similar structural variations (39). Yet, it's also common for genetically unlinked driver genes, which activate collaborative oncogenic pathways, to display the co-occurrence of genetic alterations (40). Our findings suggest positive selections for interactions between eIF4G1 and subunits from other initiation complexes, implying their potential coordinated roles in human cancers.

Positive selection of *EIF4G1* amplification likely results from a feedback mechanism that regulates the expression of cell cycle and lipogenesis genes. *EIF4G1* expression strongly correlates with both cell cycle genes and SREBF-controlled lipogenesis genes in tumors exhibiting *EIF4G1* gain. The cell cycle and lipogenesis genes, as identified in Fig. 2, are not genetically linked to *EIF4G1*, yet they are overexpressed in tumors with *EIF4G1* gain. This suggests that cancer clones harboring *EIF4G1* gain may be positively selected because it facilitates the overexpression of cancer survival genes. Additionally, research has shown that *SREBF1* translation relies on cap-independent initiation under stress conditions that inhibit cap-dependent initiation (41). Further studies are required to evaluate whether inhibition of translation initiation could break the feedback loop and impede tumor growth, through the SREBF lipogenesis axis.

The strong co-regulation of eIF4G1, eIF4A1, eIF3E, and eIF3H, along with the independent co-regulation of eIF4E and 4E-BP1, suggests additional regulatory mechanisms for eIF4E beyond its role in the eIF4F complex. The co-regulation of eIF4G1, eIF4A1, eIF3E, and eIF3H with ribosomal subunits and various initiation complexes underscores their role in initiating translation, vital for cancer cell survival (Fig. 4). Interestingly, eIF4E is essential for cancer cell viability, and it also co-regulates with 4E-BP1, ubiquitination, and ESCRT proteins. Previous studies have shown that ubiquitination and 4E-BP1 can reduce eIF4E's ability to bind to eIF4G1, thus hindering cap-dependent initiation (42). Dysregulation of eIF4E and eIF4G1 in cancer cells could enable cap-independent mechanisms to meet critical translation initiation needs (2). Furthermore, ESCRT's role in binding to ubiquitinated proteins and its known regulation of the nuclear pore complex (NPC) (43), may facilitate eIF4E's mRNA export function. Dysregulation in cap-dependent translation initiation may result from an increased allocation of eIF4E to nuclear mRNA export, reducing its availability in the cytoplasmic eIF4F pool. Hence, eIF4E's essentiality may stem from both its mRNA transport function alongside its crucial role in cap-dependent initiation.

The Alphafold-predicted structures suggest potential mechanisms for regulating cap-dependent and cap-independent initiation through eIF4G1's HEAT-2 domain. α-helix$_3$ and α-helix$_5$ of HEAT-2 can interact with either eIF4E or eIF4A1-NTD. In the canonical eIF4F complex, eIF4G1 positions the cap-binding site of eIF4E close to the mRNA binding cavity of eIF4A1, facilitating the enclosure of the cavity by eIF4E and HEAT-2. In the absence of eIF4E, the interaction between HEAT-2 and eIF4A1-NTD (Fig. 5*D*) might enhance eIF4G1 and eIF4A1 interaction, resulting in a tighter enclosure of the mRNA binding cavity. The pivotal role of HEAT-2 in interacting with eIF4A1-NTD or eIF4E may help maintain the integrity of the eIF4F complex in both cap-dependent and -independent initiations. Further experimental investigations are needed to validate the impact of HEAT-2's interaction with either eIF4E or eIF4A1-NTD on cap-dependent or cap-independent initiation activity. This research will uncover promising drug-targetable sites within the eIF4F complexes and facilitate the design of inhibitors for specific initiation mechanisms in cancers.

Finally, our study proposes an intriguing hypothesis: eIF4E is crucial for cap-dependent initiation in human cancers; however, in situations of elevated mRNA transport demands, eIF4E may dissociate from the eIF4F complex. In this scenario, we postulate that the dysregulated eIF4F complex, mainly consisting of eIF4G1 and eIF4A1, could remain on the mRNA, potentially enabling cap-independent initiation via eIF3, ribosome, and mRNA circularization. Nonetheless, further experimental investigations are essential to validate this hypothesis.

Materials and Methods

Copy Number Variation Data and Co-Occurrence Analysis. Gene-level copy number data for 33 TCGA cancer types were obtained from the UCSC Xena data hub (44) (https://tcga.xenahubs.net and https://pancanatlas.xenahubs.net). To classify copy number variation statuses, the TCGA pan-cancer gene-level CNV threshold dataset we used, which combined GISTIC2-thresholded data from all TCGA cohorts, accessed through the Xena dataset ID: TCGA.PANCAN.sampleMap/Gistic2_CopyNumber_Gistic2_all_thresholded.by_genes. We grouped the estimated gene-level CNV values using thresholds 2, 1, 0, −1, −2, to represent high-level copy number gain (amplification), low-level copy number gain (duplication), diploid, shallow (possibly heterozygous) deletion, or deep (possibly homozygous) deletion.

To generate Likert plots of CNV statuses across different cancer types, we utilized clinically relevant phenotype information for TCGA samples, such as sample type and primary disease annotations, obtained from individual TCGA cohorts through the Xena dataset ID: TCGA_phenotype_denseDataOnlyDownload.tsv.

To conduct the co-occurrence analysis, we employed the VennCounts() function from the R package "limma" to calculate the number of overlapping genes between gene groups. We then utilized these counts to create proportional Venn diagrams using the euler() function from the R package "eulerr". Statistical analysis was performed using the fisher.test() function from the R package "stats".

Kaplan–Meier analysis was conducted using curated clinical data from the TCGA Pan-Cancer Clinical Data Resource (45), obtained through the Xena dataset ID: Survival_SupplementalTable_S1_20171025_xena_sp. Survival analysis utilized the fit() function from the class KaplanMeierFitter() in the Python package "lifelines". Differences in survival curves were evaluated using the log-rank test with the logrank_test() function from the statistics() class in the lifelines package.

RNA-Seq and Gene Expression Analysis. The original RNA-Seq data of tumor samples were obtained from TCGA and RNA-Seq data of healthy samples from GTEx (46). To ensure consistency and minimize computational

batch effects on read alignment and quantification, we used the reprocessed RNA-Seq read count data for both sources, available from the UC Santa Cruz computational genomics Lab, which was computed with the Toil-based RNA-Seq bioinformatic pipeline (47). We accessed the RNA-Seq datasets from the UCSC Xena data hub (https://toil.xenahubs.net) using the Xena dataset IDs: TcgaTargetGtex_RSEM_hugo_norm_count.

For the differential correlation analysis, we utilized the corrwith() function from the Python package "pandas" to calculate Pearson's correlation coefficient (r). To perform UMAP, we standardized the gene expression data by scaling it to unit variance using the fit_transform() function from the class StandardScaler() of the Python package "sklearn.preprocessing". Next, we used the fit_transform() function from the class UMAP() of the Python package "umap" to embed the standardized data into a Euclidean space.

Proteomics and Network Analysis. The CCLE proteomics data were obtained from the publication (28). We obtained the dependency score data of CRISPR knockout screens and sample information from the depmap portal, using the file name "CRISPR_gene_effect.csv" and "sample_info.csv". To construct the protein–protein interaction network, we used protein network data from the STRING database with the file name "9606.protein.physical.links.detailed.v11.5.txt". We constructed the protein–protein interaction networks using the from_pandas_edgelist() function and plotted them with the draw_networkx() and kamada_kawai_layout() functions from the Python package "networkx". We conducted the centrality analysis on the network using the degree_centrality() function from networkx.

Heatmap, Clustering, and Pathway Enrichment Analysis. To create heatmaps, we utilized the Heatmap() function from the R package "ComplexHeatmap". The heatmap rows were ordered and grouped into subgroups using the K-means clustering method while the heatmap columns were ordered by the hierarchical clustering method. To analyze the enriched biological pathways of the genes within each cluster, we used the enrichPathway() function from the R package "ReactomePA", which employed Reactome as a source of pathway data. We performed statistical analysis and visualization of these pathways using the compareCluster() function from the R package "clusterProfiler", by applying an over-representation analysis (ORA) method. The statistical significance (*P*-value) of the overlap between genes from a given pathway and the gene list was determined using the hypergeometric distribution test, and the *P*-values were adjusted for multiple comparison using Hochberg's and Hommel's method.

Protein Complex Structure Prediction. We employed the AlphaFold multimer approach, which is an extension of the AlphaFold2 algorithm and is capable of predicting the structure of a protein complex as a single entity (48). All tasks were performed on the GPU cluster of Harvard medical school, with the default AlphaFold2 multimer settings. We used the amino acid sequences from the NCBI database with the following accession number: eIF4E, NP_001959.1; eIF4G1, NP_886553.3; eIF4A1, NP_001407.1; and 4E-BP1, NP_004086.1. We evaluated the consistency of the five prediction models generated by each task and selected the top-ranked model for illustration using PyMOL.

Data, Materials, and Software Availability. The detailed source code, which reproduces the results reported in this manuscript, is available online at https://github.com/a3609640/pyEIF (49) in the form of R and Python scripts. All the data utilized in our analyses are publicly accessible and the scripts for downloading them from the relevant data repositories are provided in our software repository. All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. The results published here are based on data generated by the TCGA Research Network: https://www.cancer.gov/tcga, the Genotype-Tissue Expression (GTEx) Project, and Cancer Cell Line Encyclopedia (CCLE). This work was supported by the National Cancer Institute (5R01CA200913-05 to G.W.) and the National Institute of Allergy and Infectious Diseases (5P01AI143565-03 to G.W.).

Author affiliations: [a]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

1. J. Pelletier, N. Sonenberg, The organizing principles of eukaryotic ribosome recruitment. *Annu. Rev. Biochem.* **88**, 307–335 (2019).
2. D. Silvera, S. C. Formenti, R. J. Schneider, Translational control in cancer. *Nat. Rev. Cancer* **10**, 254–266 (2010).
3. C. de la Parra, B. A. Walters, P. Geter, R. J. Schneider, Translation initiation factors and their relevance in cancer. *Curr. Opin. Genet Dev.* **48**, 82–88 (2018).
4. N. Robichaud, N. Sonenberg, D. Ruggero, R. J. Schneider, Translational control in cancer. *Cold Spring Harb Perspect Biol.* **11**, a032896 (2018), 10.1101/cshperspect.a032896.
5. R. J. Jackson, C. U. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
6. A. Yanagiya *et al.*, Requirement of RNA binding of mammalian eukaryotic translation initiation factor 4GI (eIF4GI) for efficient interaction of eIF4E with the mRNA cap. *Mol. Cell Biol.* **29**, 1661–1669 (2009).
7. A. Haghighat, N. Sonenberg, eIF4G dramatically enhances the binding of eIF4E to the mRNA 5′-cap structure. *J. Biol. Chem.* **272**, 21677–21680 (1997).
8. K. Feoktistova, E. Tuvshintogs, A. Do, C. S. Fraser, Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13339–13344 (2013).
9. N. Villa, A. Do, J. W. Hershey, C. S. Fraser, Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J. Biol. Chem.* **288**, 32932–32940 (2013).
10. A. K. LeFebvre *et al.*, Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J. Biol. Chem.* **281**, 22917–22932 (2006).
11. J. Pelletier, J. Graff, D. Ruggero, N. Sonenberg, Targeting the eIF4F translation initiation complex: A critical nexus for cancer development. *Cancer Res.* **75**, 250–263 (2015).
12. M. Greaves, C. C. Maley, Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
13. I. Martincorena *et al.*, Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e1021 (2017).
14. Y. V. Svitkin *et al.*, Eukaryotic translation initiation factor 4E availability controls the switch between cap-dependent and internal ribosomal entry site-mediated translation. *Mol. Cell Biol.* **25**, 10556–10565 (2005).
15. B. Culjkovic, I. Topisirovic, L. Skrabanek, M. Ruiz-Gutierrez, K. L. Borden, eIF4E is a central node of an RNA regulon that governs cellular proliferation. *J. Cell Biol.* **175**, 415–426 (2006).
16. L. Rong *et al.*, Control of eIF4E cellular localization by eIF4E-binding proteins, 4E-BPs. *RNA* **14**, 1318–1327 (2008).
17. S. Wu, G. Wagner, Deep computational analysis details dysregulation of eukaryotic translation initiation complex eIF4F in human cancers. *Cell Syst.* **12**, 907–923.e906 (2021).
18. S. Braunstein *et al.*, A hypoxia-controlled cap-dependent to cap-independent translation switch in breast cancer. *Mol. Cell* **28**, 501–512 (2007).
19. H. Imataka, N. Sonenberg, Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol. Cell Biol.* **17**, 6940–6947 (1997).
20. I. B. Lomakin, C. U. Hellen, T. V. Pestova, Physical association of eukaryotic initiation factor 4G (eIF4G) with eIF4A strongly enhances binding of eIF4G to the internal ribosomal entry site of encephalomyocarditis virus and is required for internal initiation of translation. *Mol. Cell Biol.* **20**, 6019–6029 (2000).
21. A. B. Reams, J. R. Roth, Mechanisms of gene duplication and amplification. *Cold Spring Harb. Perspect. Biol.* **7**, a016592 (2015).
22. F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* **279**, 5048–5057 (2012).
23. H. Tanaka, T. Watanabe, Mechanisms underlying recurrent genomic amplification in human cancers. *Trends Cancer* **6**, 462–477 (2020).
24. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
25. S. Wu, A. M. Naar, SREBP1-dependent de novo fatty acid synthesis gene expression is elevated in malignant melanoma and represents a cellular survival trait. *Sci. Rep.* **9**, 10369 (2019).
26. S. Wu, A. M. Naar, A lipid-free and insulin-supplemented medium supports De Novo fatty acid synthesis gene activation in melanoma cells. *PloS One* **14**, e0215022 (2019).
27. Y. Yang *et al.*, Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *Cell Rep.* **36**, 109442 (2021).
28. D. P. Nusinow *et al.*, Quantitative proteomics of the cancer cell line encyclopedia. *Cell* **180**, 387–402.e316 (2020).
29. A. Tsherniak *et al.*, Defining a cancer dependency map. *Cell* **170**, 564–576.e516 (2017).
30. R. M. Meyers *et al.*, Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
31. A. Martinez-Ferriz, A. Ferrando, A. Fathinajafabadi, R. Farras, Ubiquitin-mediated mechanisms of translational control. *Semin. Cell Dev. Biol.* **132**, 146–154 (2022).
32. S. Iwasaki *et al.*, The translation inhibitor rocaglamide targets a bimolecular cavity between eIF4A and Polypurine RNA. *Mol. Cell* **73**, 738–748.e739 (2019).
33. A. Marintchev *et al.*, Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* **136**, 447–460 (2009).
34. C. Wasmer, A. Marintchev, K. Edmonds, S. Wu, G. Wagner, Site-directed spin-labeling using an unnatural amino acid and application to the eIF4A–eIF4G translation initiation complex (Manuscript in Preparation).
35. C. Igreja, D. Peter, C. Weiler, E. Izaurralde, 4E-BPs require non-canonical 4E-binding motifs and a lateral surface of eIF4E to repress translation. *Nat. Commun.* **5**, 4790 (2014).
36. S. Gruner *et al.*, The structures of eIF4E-eIF4G complexes reveal an extended interface to regulate translation initiation. *Mol. Cell* **64**, 467–479 (2016).
37. N. Sekiyama *et al.*, Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4036–E4045 (2015).
38. A. Kahvejian, Y. V. Svitkin, R. Sukarieh, M. N. M'Boutchou, N. Sonenberg, Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev.* **19**, 104–113 (2005).

39. L. D. Hurst, C. Pal, M. J. Lercher, The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310 (2004).
40. G. El Tekle *et al.*, Co-occurrence and mutual exclusivity: What cross-cancer mutation patterns can tell us. *Trends Cancer* **7**, 823–836 (2021).
41. F. Damiano, S. Alemanno, G. V. Gnoni, L. Siculella, Translational control of the sterol-regulatory transcription factor SREBP-1 mRNA in response to serum starvation or ER stress is mediated by an internal ribosome entry site. *Biochem. J.* **429**, 603–612 (2010).
42. T. Murata, K. Shimotohno, Ubiquitination and proteasome-dependent degradation of human eukaryotic translation initiation factor 4E. *J. Biol. Chem.* **281**, 20788–20800 (2006).
43. M. Vietri, M. Radulovic, H. Stenmark, The many functions of ESCRTs. *Nat. Rev. Mol. Cell Biol.* **21**, 25–42 (2020).
44. M. J. Goldman *et al.*, Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).
45. J. Liu *et al.*, An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416.e411 (2018).
46. M. Mele *et al.*, Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
47. J. Vivian *et al.*, Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* **35**, 314–316 (2017).
48. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2021.10.04.463034 (Accessed 10 March 2022).
49. S. Wu, pyEIF. GitHub. https://github.com/a3609640/pyEIF. Deposited 28 February 2023.