Unbiased screen of human transcriptome reveals an unexpected role of 3'UTRs in translation initiation

Yun Yang^{1,*,#}, Xiaojuan Fan^{1,*}, Yanwen Ye^{1,2}, Chuyun Chen^{1,2}, Sebastian E.J. Ludwig³, Sirui Zhang^{1,2}, Qianyun Lu^{1,2}, Cindy L. Will³, Henning Urlaub⁴, Jing Sun¹, Reinhard Lührmann³, Zefeng Wang^{1,2,5,#}

¹ Bio-med Big Data Center, CAS Key Laboratory of Computational Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, China

² University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

³ Max Planck Institute for Biophysical Chemistry, Cellular Biochemistry, Am Fassberg 11, D-37077, Göttingen, Germany

⁴ Bioanalytical Mass Spectrometry, MPI for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

⁵Lead Contact

* These authors contributed equally to this work

Corresponding to: wangzefeng@picb.an.cn; yangyun@picb.ac.cn

Running title: Translational enhancement from 3'UTR

Abstract

Although most eukaryotic mRNAs require a 5'-cap for translation initiation, some can also be translated through a poorly studied cap-independent pathway. Here we developed a circRNA-based system and unbiasedly identified more than 10,000 sequences in the human transcriptome that contain Cap-independent Translation Initiators (CiTIs). Surprisingly, most of the identified CiTIs are located in 3'UTRs, which mainly promote translation initiation in mRNAs bearing highly structured 5'UTR. Mechanistically, CiTI recruits several translation initiation factors including eIF3 and DHX29, which in turn unwind 5'UTR structures and facilitate ribosome scanning. Functionally, we showed that the translation of HIF1A mRNA, an endogenous DHX29 target, is antagonistically regulated by its 5'UTR structure and a new 3'-CiTI in response to hypoxia. Therefore, deletion of 3'-CiTI suppresses cell growth in hypoxia and tumor progression *in vivo*. Collectively, our study uncovers a new regulatory mode for translation where the 3'UTR actively participate in the translation initiation.

Introduction

The mRNA translation is an essential and energetically expensive process tightly regulated at the initiation stage. In eukaryotic cells, canonical translation initiation is mediated by the formation of eIF4F complex on the 5'-cap (7-methylguanosine), followed by the attachment of the ribosomal 43S preinitiation complex that scans along the mRNA in a 5'-to-3' direction for a start codon ^{1,2}. Multiple *cis*-elements and *trans*-factors are employed to accurately regulate mRNA translation initiation. For example, the RNA structures and upstream ORF (uORF) in 5'UTRs are well known *cis*-elements that regulate mRNA translation, whereas many protein factors or miRNAs can control translation *in trans* ¹⁻⁴.

In addition, many genes can also be translated in a cap-independent fashion, mostly through direct recruitment of ribosomes to the internal ribosome entry site (IRES), which often occurs during stress conditions (like virus infection, heat shock, or apoptosis) ^{5,6}. It has been estimated that up to 10-15% of cellular mRNAs undergo cap-independent translation under various conditions where cap-dependent translation is suppressed ⁷. A recent high-throughput screen targeting the viral RNAs and 5'UTRs of human mRNAs revealed that many human genes (583 out of 5058 genes tested) harbor the “cap-independent translation elements” that function as IRESs to drive mRNA translation ⁸. However, an unbiased screen of the translation initiation elements remains absent. Recently, thousands of circular RNAs (circRNAs) have been identified in eukaryotic cells ⁹⁻¹¹, many of which can be translated *in vitro* and *in vivo* ¹²⁻¹⁶. Since circRNAs lack free ends and can only be translated cap-

independently, they may serve as an improved system to study this noncanonical translation pathway.

Results

Identification of cap-independent translation initiation elements with a circRNA-based translation system

To unbiasedly screen the human transcriptome for endogenous sequences that drive cap-independent translation, we developed an efficient cell-based circRNA translation system (Fig. 1a) that has been extensively validated^{13,15-17}. We inserted a library of cDNA short fragments from normalized human transcriptome before start codon of the “split” EGFP (Fig. 1a)^{13,16}, and transfected the library to generate ~3 million stable clones. The green cells were selected by FACS (Extended data Fig. 1a), and the insertion libraries were sequenced at each stage (supplementary information).

The fragment abundance in this library correlated poorly with gene levels in HeLa cells ($r < 0.09$, Extended data Fig. 1b), with rRNAs representing <10% of the total fragments (Extended data Fig. 1c), indicative of an effective transcriptome normalization. After sorting, we recovered 82,367 unique fragments from the green cells with similar length distribution as the input (Extended data Fig. 1d and Extended data Table 1). The recovered fragments were not dominated by the pre-sorting library (Extended data Fig. 1e), suggesting the screen was unbiased. The majority of the reads from the pre- and post-sorting libraries were located in protein coding genes (Fig. 1b), with two examples shown in read density plots (Fig. 1c). Surprisingly, the

percentage of positive reads from 3'UTRs was substantially increased whereas the reads from 5'UTRs and coding sequences (CDS) was decreased after sorting (Fig. 1b), suggesting a prominent role of 3'UTRs in regulation of cap-independent translation.

We further validated 14 out of 15 randomly selected fragments for their activity in driving circRNA translation (Fig. 1d, activities are independent on their read depth), indicating a low false positive rate. The IRES-like activities in 8 out of 9 positive fragments were also confirmed using a bicistronic translation reporter (Fig. 1e). Conversely, 3 out of the randomly selected 12 single clones from the input library promoted circRNA translation (Extended data Fig. 1f), suggesting that our screen had not reached saturation.

The overlapping positive fragments were merged into 10,455 peaks in human transcriptome, which were defined as cap-independent translation initiators (CiTIs). Most CiTIs were recovered only by a small number of reads in the post-sorting library while several CiTIs have exceedingly high read depth (Extended data Fig. 1g). The CiTIs were widely distributed across all human chromosomes (Extended data Fig. 1h), and their abundance correlated positively with the gene numbers ($r=0.96$) and the size ($r=0.76$) of each chromosome (Extended data Fig. 1i). The newly identified CiTIs included many previously reported IRESs^{8,18} (Extended data Fig. 1j), including ~10% of IRESs identified from a screen biased towards the 5'UTRs and viral sequences⁸.

The CiTI-containing genes were more conserved than an average human gene (Extended data Fig. 2a), and were actively expressed and translated in cultured cells

(Extended data Fig. 2b). Interestingly, CiTI containing genes were substantially enriched in functional categories of RNA translation, splicing, cell-cell adhesion and viral processes (Extended data Fig. 2c). Particularly, the CiTIs were enriched in genes encoding ribosomal proteins and translation initiation factors, suggesting a potential positive feedback regulation of translation.

Although the canonical IRESs are mostly located in the 5'UTR, the newly identified CiTIs were found in all gene regions including 5'UTRs (termed 5'-CiTI), CDS (termed C-CiTI), and 3'UTRs (termed 3'-CiTI). Consistent with our earlier finding, the CiTIs were enriched in 3'UTRs and peaked in the CDS near the start and stop codons (Fig. 1f). Compared to the cognate controls, we found a significant increase of sequence conservation in 3'-CiTIs, a modest increase in 5'-CiTIs, and no increase in C-CiTIs (Extended data Fig. 2d). The CiTIs generally had less structures as judged by higher minimal free energy (MFE) (Extended data Fig. 2e), whereas the canonical IRESs were significantly more structured¹⁸, suggesting that, unlike IRESs, most CiTIs do not function by forming a specific structure.

Several enriched motifs from the CiTIs, including the previously reported U-rich motifs^{8,19}, were identified and further experimentally validated (Extended data Fig. 3a-3d). Consistent with previous reports^{8,20}, we found that the 5'-CiTIs tend to interact with 18S rRNAs *via* base-pairing, and such interaction is sufficient to drive cap-independent translation (Extended data Fig. 3e-3g). Therefore, our circRNA-based translation system successfully identified several categories of CiTIs that

contain both previously confirmed elements/motifs and novel elements enriched within 3'UTRs.

Genes with 3'-CiTIs tend to contain downstream ORFs or structured 5'UTRs

The surprising enrichment of CiTIs in 3'UTRs rather than 5'UTRs implied an unexpected new role of 3'UTR in promoting translation initiation. Naturally, the 3'-CiTI may function as a *bona fide* IRES to drive cap-independent translation of downstream ORFs (dORF). By analyzing all potential dORFs, we found that the dORFs in the genes with 3'-CiTIs were more conserved than those without 3'-CiTIs (Extended data Fig. 4a), suggesting that 3'-CiTIs are probably under functional selection to drive dORF translation. Further analysis using published datasets of ribosome footprints^{21,22} showed that a significant fraction of these dORFs (177/564) also had an upstream 3'-CiTI (Extended data Fig. 4b, $p=2.9 \times 10^{-29}$). However, we only identified the potential translation products from 7 of these dORFs using public mass spectrometry datasets^{23,24}. The lack of strong proteomic evidences may be due to cell- or tissue-specific translation of these dORFs, and/or the rapid degradation of these dORF-encoded proteins just like 3'UTR-coded peptides from translation read-through^{25,26}. These observations suggest that driving cap-independent translation of dORFs might not be the function for most 3'-CiTIs.

The fact that >90% of 3'-CiTI-containing genes lack evidence for dORF translation (Extended data Fig. 4b) suggested unknown roles of 3'-CiTIs. Interestingly, the 3'-CiTI-containing genes were significantly more conserved and

more structured in their 5'UTRs than those without 3'-CiTIs (Fig. 2a-2b). They were also significantly overlapped with the genes containing upstream ORFs (uORF) (Extended data Fig. 4c)^{27,28}. Since most 5'UTR structures and uORFs play inhibitory roles in regulating translation²⁹, we speculated that 3'-CiTIs may counteract these 5'UTR structures or uORFs to enhance the main ORF (mORF) translation.

Such possibility was directly tested using four translation reporters with different 5'UTRs (Extended data Fig. 5a): a canonical reporter with a regular 5'UTR, a hairpin reporter containing a 5'UTR hairpin immediately before the ORF, and two uORF reporters with the 5'UTRs containing two endogenous uORFs^{30,31}. We found that 6 out of 7 CiTIs inserted into the 3'UTR of the hairpin reporter significantly increased the translation (>10-fold increase) (Fig. 2c), and this translation enhancement was independent of the CiTI length or the cell lines used (Extended data Fig. 5b-5c). In contrast, no reliable translation increase or even inhibitory effect was detected when CiTIs were inserted into the canonical reporter (Extended data Fig. 5d). As another comparison, insertion of CiTIs into the 3'UTR of the uORF reporters showed a weaker effect (Extended data Fig. 5e).

We next tested several pairs of highly structured 5'UTR and 3'-CiTI from endogenous mRNAs for their activities to cooperatively regulate translation. As expected, the structured 5'UTRs from all endogenous genes suppressed translation, and such suppression could be “reversed” by the cognate 3'-CiTIs (Fig. 2d), again suggesting a general translation enhancer activity of 3'-CiTIs that counteracts the translation suppression mediated by secondary structures in the 5'UTR.

Most human mRNAs contain a 5'-cap and poly-A tail that are looped together by a protein complex to increase translation efficiency³²⁻³⁴. To determine if this looping is required for 3'-CiTI activity, we transfected cells with *in vitro* synthesized mRNAs containing optional 5'-cap or poly-A tail (Fig. 2e). In all mRNAs tested, insertion of a 3'-CiTI reversed the translation suppression by the 5'-hairpin, indicating the translation promotion by 3'-CiTI is independent of the 5'-cap or poly-A tail (Fig. 2e). This finding was further validated with a Poll reporter that can produce RNAs without a 5'-cap or poly-A tail, where a 3'-CiTI again enhanced translation of these “naked” mRNAs containing a 5'-hairpin (Extended data Fig. 6). Interestingly, the translation of mRNA that lacks a 5'-cap or poly-A tail could be enhanced by simply introducing a mRNA loop across the ORF using complementary sequences, and insertion of a 3'-CiTI further increased the translation of such mRNAs (Fig. 2f).

3'-CiTIs unwind the 5'UTR structures to facilitate ribosome scanning

During translation initiation, the preinitiation complex (PIC) usually needs to scan through the 5'UTR structures to initiate translation^{2,35-37}. Occasionally, the PIC can “jump over” the 5'UTR structures, a phenomenon called ribosome shunting³⁸ (Fig. 2g, left). To determine how 3'-CiTIs enhance translation of mRNAs containing 5'UTR structures, we generated several reporters with 5'-hairpins of different lengths and an upstream in-frame start codon (uAUG) inside the hairpin (Extended data Fig. 6a). The first start codon uAUG will be used by default to generate an extended Rluc (eRluc) *via* ribosome scanning. Alternatively, the PIC may “jump over” the structure

and use the second AUG (i.e., without structure unwinding) to produce the canonical Rluc.

We found that the uAUG was exclusively used during translation, suggesting a ribosome scanning in all reporter mRNAs (Fig. 2g). As expected, introduction of a 5'-hairpin reduced the translation efficiency in a length-dependent fashion, and insertion of a 3'-CiTI increased the translation from the uAUG as judged by both western blots and luciferase activities (Fig. 2g). This increase was achieved *via* promoting ribosome scanning rather than ribosome shunting, as only the proximal start codon was used in all cases. Additionally, we found that the reporter mRNA containing a 3'-CiTI had less structures (as judged by icSHAPE signals) than the mRNA without the 3'-CiTI, in both the entire 5'UTR and its hairpin region (Fig. 2h, and Extended data Fig. 7b-c), suggesting that the 3'-CiTI may unwind the upstream structures to help ribosomes scan through the 5'UTR.

***Trans*-acting factors are involved in 3'-CiTI-mediated translation enhancement**

To determine if the 3'-CiTIs recruit *trans*-acting factors to promote translation, we performed RNA affinity purification (Fig. 3a) using the MS2-fused RNAs containing three different 3'-CiTIs and controls (Fig. 2c, Extended data Fig. 8a). The purified RNP complexes were analyzed with mass spectrometry to identify 63 candidate proteins that specifically bind all three CiTIs (Extended data Fig. 8b and Table 2). These CiTI-binding proteins formed a densely-connected protein-protein interaction network with key functions in translation initiation and RNA processing

(Fig. 3b), including many components in the eIF2 or eIF3 complex that functions as a multi-tasking machine to promote translation³⁹. Interestingly, the newly identified CiTI-binding proteins also include DHX29 (Fig. 3b), a DExH-box RNA helicase that might help to unwind RNA structures.

In line with the possible role of DHX29 in CiTI-mediated translation regulation, knockdown of DHX29 specifically reduced the translation of several reporter mRNAs containing 3'-CiTIs (Fig. 3c, Extended data Fig. 8c). Furthermore, directly tethering DHX29 onto the 3'UTR significantly increased translation of a hairpin-containing mRNA (Fig. 3d, Extended data Fig. 8d), suggesting that DHX29 recruitment is sufficient for translation promotion. The recruitment of several eIF3 components also promoted the translation (Fig. 3d), supporting a previous finding that DHX29 cooperates with eIF3 during translation initiation^{29,40}. Since several eIF3 components were also pulled down by 3'-CiTIs (Fig. 3b), DHX29 may function together with the eIF3 complex to mediate the CiTI activity.

Previously, DHX29 was shown to regulate the translation of a handful of genes, including Cdc25C and XIAP^{41,42}. However, ~2/3 of human mRNAs have structured 5'UTRs (i.e., with low MFE, $\Delta G < -40$ kcal/mol)⁴², many of which may also be regulated by DHX29. To systematically identify *bona fide* DHX29 targets, we performed polysome profiling after DHX29 knockdown, and sequenced mRNAs from different gradient fractions (Fig. 3e and Extended data Fig. 8c). We focused on the genes with consistent profiles in two independent DHX29 knockdown lines (Pearson correlation > 0.8), and calculated the translation efficiency by the percent of

polysome-associated mRNAs⁴³. In total, 869 genes were identified with at least a 25% change in translation efficiency (Fig. 3f), most of which showed a translation reduction (Fig. 3f and 3g, Extended data Table 3). Furthermore, the genes with reduced translation efficiency in DHX29 knockdown cells had more RNA structures in their 5'UTR (Fig. 3h), and 91 of these genes also contained 3'-CiTIs (p=0.001 by hypergeometric test). Collectively, these results suggested that DHX29 plays a critical role in unwinding mRNA structures and promotes translation upon specific recruitment by 3'-CiTIs.

3'-CiTI and DHX29 promote HIF1A translation in response to hypoxia *in vitro* and *in vivo*

An important new target of DHX29 is the hypoxia-induced transcription factor HIF1A, which contains both a 3'-CiTI and strong 5'UTR structure^{44,45}. As a key factor for cellular response to hypoxia, HIF1A translation is suppressed in normoxia but strongly induced by oxygen deprivation⁴⁶⁻⁴⁸. Using translation reporters containing HIF1A 5'UTR and different fragments of its 3'UTR, we found that the 5'UTR significantly suppressed translation under hypoxia, however the HIF1A 3'-CiTI fully restored translation in two different cell lines (Fig. 4a, top). Conversely, deletion of the 3'-CiTI from the HIF1A 3'UTR reduced translation under hypoxia (Fig. 4a, bottom). In contrast, under normoxia condition, translation enhancement by the 3'-CiTI was observed only in HCT116 cells (Extended data Fig. 9a-b).

The translation promotion by the HIF1A 3'-CiTI was reduced by DHX29 knockdown under hypoxia (Fig. 4b). We further observed a direct interaction of DHX29 with HIF1A mRNAs (Extended data Fig. 9c), consistent with its general function as a subunit of the translation initiation complex⁴⁰. Interestingly, the interaction between DHX29 and HIF1A mRNA was nearly eliminated by the 3'-CiTI deletion under hypoxia, whereas this interaction was only reduced by ~50% during normoxia (Fig. 4c), suggesting that the 3'-CiTI is essential for DHX29 recruitment to HIF1A mRNA during hypoxia.

To evaluate the functional impact of 3'-CiTI-mediated translation promotion, we deleted the 3'-CiTI from endogenous HIF1A gene in SH-SY5Y neuroblastoma cells with CRISPR-cas9 (Extended data Fig. 9d). Under normoxia, HIF1A mRNA was mostly found in monosomes or light polysomes (Extended data Fig. 9e), consistent with its low translation efficiency under this condition. Hypoxia induced the association of most HIF1A mRNAs with polysomes (Fig. 4d). However, deletion of the 3'-CiTI shifted HIF1A mRNAs (but not control mRNAs) from the polysome fractions to the monosome fractions (Fig. 4d), suggesting that 3'-CiTI is required to specifically promote translation of HIF1A mRNA in response to hypoxia.

Furthermore, deletion of the 3'-CiTI significantly blunted the induction of HIF1A protein by hypoxia (Fig. 4e), resulting in a small but significant reduction in cell proliferation similar to that in HIF1A knockdown cells (Fig. 4f and Extended data Fig. 9f). Importantly, when inoculated in nude mice, 3'-CiTI deleted SH-SY5Y neuroblastoma tumors exhibited a dramatic reduction of growth (Fig. 4g and 4h).

Taken together, these observations indicate that the 3'-CiTI in HIF1A mRNA functions as a key *cis*-element that recruits DHX29 to mediate environmental regulation of its translation, which plays a critical role in supporting *in vivo* tumor growth.

Discussion

By unbiasedly screening human transcriptome, we uncovered a new class of 3'UTR elements that play an unexpected role in translation initiation. Specifically, the 3'-CiTIs actively recruit key initiation factors to unwind inhibitory RNA structures in the 5'UTRs, thereby promoting the ribosomes scanning (Fig. 4i). Since actively translated mRNAs generally form a loop through interactions of poly-A binding proteins with eIF4G⁴⁹, the 3'UTR may generally serve as a repository of regulatory elements that “catch and store” the translation initiation factors to increase their local concentration for the next round of translation.

This new model not only reveals a new role of 3'UTRs in translation initiation, but also explains why the 3'-CiTIs were initially identified as cap-independent translation initiator in our circRNA screening system. Unlike the canonical translation initiated with formation of eIF4F complex on 5'-cap, the 3'-CiTI in circRNAs may recruit initiation factors (e.g., DHX29 and eIF3 and eIF2 components), thus hijacking the initiation machine to bypass the requirement of eIF4F and drive cap-independent translation. Consequently, these 3'-CiTIs are conceptually equivalent to the “IRES”

when inserted into the suitable context, although they function as translation initiators/enhancers from 3'UTR in endogenous settings.

Our findings further indicate that the 3'-CiTI-mediated translational control is an ancient and common mode of regulation. Firstly, the 3'-CiTIs are significantly more conserved than the matched control regions (Extended data Fig. 1d) and the 5'UTR in the 3'-CiTI-containing mRNAs are also more conserved and more structured (Fig. 2a-2b), indicating that the pairs of 3'-CiTIs and 5'UTRs are under functional selection in evolution. Moreover, ~44% of the 2681 newly identified 3'-CiTIs are located in human mRNAs with structured 5'UTRs ($\Delta G < -40$ kcal/mol)⁴², suggesting a common role of 3'-CiTIs in counteracting the inhibitory 5'UTR structures in many genes.

This study also highlights the importance of DHX29 in 3'-CiTI-mediated translation initiation in response to environmental stress. Several RNA helicases (including DHX29, DHX9 and DDX3) were reported to promote translation^{29,40}. In particular, DHX29 has been shown to interact with eIF3 to form an active 43S/DHX29 complex⁴¹. However, it is unclear how these helicases are specifically recruited to mRNAs. Our data indicate that many 3'-CiTIs can function as the *cis*-acting elements to recruit DHX29 (and eIF3). The biological impact of such regulation is exemplified by HIF1A: deletion of a 3'-CiTI on the endogenous *HIF1A* locus directly impairs its translation and reduces cell survival under hypoxia, which further blunts *in vivo* tumor progression.

Given the various stress conditions where the cap-dependent translation initiation is inhibited (e.g., hypoxia, viral infection, heat/cold shock, etc.), the cooperative

regulation of translation by the structured 5'UTR and 3'-CiTI reported here adds an additional layer of regulatory complexity during stress. Future studies are needed to determine how CiTI-mediated translational initiation can interplay with other modes of translation control from 3'UTRs, such as regulation by m6A and miRNAs^{4,50}.

In summary, our study uncovers an uncanonical mechanism of 3'UTR in mediating translation initiation, and further reveals the importance of this regulation in stress-induced cell survival and proliferation *in vitro* and *in vivo*. These findings may pave the way for novel cancer therapies that target stress-induced translation.

Data availability

The raw sequencing data and filtered datasets are publicly available from NODE (access number: OEP001285) and NCBI (access number: PRJNA680213). The customized scripts for data analyses are also available from GitHub.

Acknowledgments

The authors thank Dr. Hao Chi for his support in identifying dORF peptides from proteomic data using open pFind(v3.1.3), Dr. Eran Segal and Dr. Shira Weingarten-Gabbay for kindly sharing the raw data of their previous screen, Dr. Qiangfeng Cliff Zhang for the suggestion in icSHAPE analysis, and Drs. Xiaoling Li, Dieter A. Wolf, Shu-Bing Qian and Nahum Sonenberg for useful discussions. This work is supported by the Natural Science Foundation of China to Z.W. (91940303, 31661143031, and 31730110) and Y.Y. (91753135, 31870814), the National Key Research and Development Program of China to Z.W. (2018YFA0107602), and funding from the Max Planck Society to H.U. and R.L. Z.W. is also supported by the type A CAS Pioneer 100-Talent program. Y.Y. is sponsored by the Youth Innovation Promotion Association CAS (2019267), SA-SIBS Scholarship Program and Shanghai Science and Technology Committee Rising-Star Program (19QA1410500).

References

- 1 Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745, doi:10.1016/j.cell.2009.01.042 (2009).
- 2 Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127, doi:10.1038/nrm2838 (2010).
- 3 Ma, X. M. & Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol* **10**, 307-318, doi:10.1038/nrm2672 (2009).
- 4 Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79**, 351-379, doi:10.1146/annurev-biochem-060308-103103 (2010).
- 5 Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J. & Embarek, A. M. Insights into structural and mechanistic features of viral IRES elements. *Front Microbiol* **8**, 2629, doi:10.3389/fmicb.2017.02629 (2017).
- 6 Yang, Y. & Wang, Z. IRES-mediated cap-independent translation, a path leading to hidden proteome. *J Mol Cell Biol*, doi:10.1093/jmcb/mjz091 (2019).
- 7 Spriggs, K. A., Stoneley, M., Bushell, M. & Willis, A. E. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biol Cell* **100**, 27-38, doi:10.1042/BC20070098 (2008).
- 8 Weingarten-Gabbay, S. *et al.* Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, doi:10.1126/science.aad4939 (2016).
- 9 Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, e30733, doi:10.1371/journal.pone.0030733.
- 10 Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141-157, doi:10.1261/rna.035667.112.
- 11 Zhang, X. O. *et al.* Complementary sequence-mediated exon circularization. *Cell* **159**, 134-147, doi:10.1016/j.cell.2014.09.001 (2014).
- 12 Chen, C. Y. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268**, 415-417 (1995).
- 13 Wang, Y. & Wang, Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**, 172-179, doi:10.1261/rna.048272.114.
- 14 Legnini, I. *et al.* Circ-ZNF609 Is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* **66**, 22-37 e29, doi:10.1016/j.molcel.2017.02.017 (2017).
- 15 Pamudurti, N. R. *et al.* Translation of circRNAs. *Mol Cell* **66**, 9-21 e27, doi:10.1016/j.molcel.2017.02.021 (2017).
- 16 Yang, Y. *et al.* Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res* **27**, 626-641, doi:10.1038/cr.2017.31 (2017).
- 17 Li, X. *et al.* Coordinated circRNA Biogenesis and function with NF90/NF110 in viral infection. *Mol Cell* **67**, 214-227 e217, doi:10.1016/j.molcel.2017.05.023 (2017).

400 18 Mokrejs, M. *et al.* IRESite—a tool for the examination of viral and cellular
401 internal ribosome entry sites. *Nucleic Acids Res* **38**, D131–136,
402 doi:10.1093/nar/gkp981 (2010).

403 19 Gritsenko, A. A. *et al.* Sequence features of viral and human Internal Ribosome
404 Entry Sites predictive of their activity. *PLoS Comput Biol* **13**, e1005734,
405 doi:10.1371/journal.pcbi.1005734 (2017).

406 20 Pisarev, A. V., Kolupaeva, V. G., Yusupov, M. M., Hellen, C. U. & Pestova, T.
407 V. Ribosomal position and contacts of mRNA in eukaryotic translation
408 initiation complexes. *EMBO J* **27**, 1609–1621, doi:10.1038/emboj.2008.90 (2008).

409 21 Mackowiak, S. D. *et al.* Extensive identification and analysis of conserved
410 small ORFs in animals. *Genome Biol* **16**, 179, doi:10.1186/s13059-015-0742-x
411 (2015).

412 22 Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes
413 are translated and some are likely to express functional proteins. *Elife* **4**,
414 e08890, doi:10.7554/eLife.08890 (2015).

415 23 Chi, H. *et al.* Comprehensive identification of peptides in tandem mass spectra
416 using an efficient open search engine. *Nat Biotechnol*, doi:10.1038/nbt.4236
417 (2018).

418 24 Bekker-Jensen, D. B. *et al.* An Optimized shotgun strategy for the rapid
419 generation of comprehensive human proteomes. *Cell Syst* **4**, 587–599 e584,
420 doi:10.1016/j.cels.2017.05.009 (2017).

421 25 Arribere, J. A. *et al.* Translation readthrough mitigation. *Nature* **534**, 719–
422 723, doi:10.1038/nature18308 (2016).

423 26 Kramarski, L. & Arbely, E. Translational read-through promotes aggregation
424 and shapes stop codon identity. *Nucleic Acids Res* **48**, 3747–3760,
425 doi:10.1093/nar/gkaa136 (2020).

426 27 Chen, J. *et al.* Pervasive functional translation of noncanonical human open
427 reading frames. *Science* **367**, 1140–1146, doi:10.1126/science.aay0262 (2020).

428 28 Wan, J. & Qian, S. B. TISdb: a database for alternative translation initiation
429 in mammalian cells. *Nucleic Acids Res* **42**, D845–850, doi:10.1093/nar/gkt1085
430 (2014).

431 29 Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5'–
432 untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416,
433 doi:10.1126/science.aad9868 (2016).

434 30 Jousse, C. *et al.* Inhibition of CHOP translation by a peptide encoded by an
435 open reading frame localized in the chop 5'UTR. *Nucleic Acids Res* **29**, 4341–
436 4351, doi:10.1093/nar/29.21.4341 (2001).

437 31 Lee, Y. Y., Cevallos, R. C. & Jan, E. An upstream open reading frame regulates
438 translation of GADD34 during cellular stresses that induce eIF2 α
439 phosphorylation. *J Biol Chem* **284**, 6661–6673, doi:10.1074/jbc.M806735200
440 (2009).

441 32 Gallie, D. R. The cap and poly(A) tail function synergistically to regulate
442 mRNA translational efficiency. *Genes Dev* **5**, 2108–2116,
443 doi:10.1101/gad.5.11.2108 (1991).

444 33 Svitkin, Y. V. & Sonenberg, N. An efficient system for cap- and poly(A)-
445 dependent translation in vitro. *Methods Mol Biol* **257**, 155-170, doi:10.1385/1-
446 59259-750-5:155 (2004).

447 34 Kahvejian, A., Roy, G. & Sonenberg, N. The mRNA closed-loop model: the
448 function of PABP and PABP-interacting proteins in mRNA translation. *Cold*
449 *Spring Harb Symp Quant Biol* **66**, 293-300, doi:10.1101/sqb.2001.66.293 (2001).

450 35 Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary
451 structure within the 5' noncoding region of a eukaryotic mRNA reduces
452 translational efficiency. *Cell* **40**, 515-526, doi:10.1016/0092-8674(85)90200-4
453 (1985).

454 36 Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic
455 ribosomes. *Proc Natl Acad Sci U S A* **83**, 2850-2854, doi:10.1073/pnas.83.9.2850
456 (1986).

457 37 Kozak, M. Circumstances and mechanisms of inhibition of translation by
458 secondary structure in eucaryotic mRNAs. *Mol Cell Biol* **9**, 5134-5142,
459 doi:10.1128/mcb.9.11.5134 (1989).

460 38 Sriram, A., Bohlen, J. & Teleman, A. A. Translation acrobatics: how cancer
461 cells exploit alternate modes of translational initiation. *EMBO Rep* **19**,
462 doi:10.15252/embr.201845947 (2018).

463 39 Wolf, D. A., Lin, Y., Duan, H. & Cheng, Y. eIF-Three to Tango: emerging
464 functions of translation initiation factor eIF3 in protein synthesis and
465 disease. *J Mol Cell Biol* **12**, 403-409, doi:10.1093/jmcb/mjaa018 (2020).

466 40 Shen, L. & Pelletier, J. General and target-specific DExD/H RNA helicases in
467 eukaryotic translation initiation. *Int J Mol Sci* **21**, doi:10.3390/ijms21124402
468 (2020).

469 41 Pisareva, V. P., Pisarev, A. V., Komar, A. A., Hellen, C. U. & Pestova, T. V.
470 Translation initiation on mammalian mRNAs with structured 5'UTRs requires
471 DExH-box protein DHX29. *Cell* **135**, 1237-1250, doi:10.1016/j.cell.2008.10.037
472 (2008).

473 42 Parsyan, A. *et al.* The helicase protein DHX29 promotes translation initiation,
474 cell proliferation, and tumorigenesis. *Proc Natl Acad Sci U S A* **106**, 22217-
475 22222, doi:10.1073/pnas.0909773106 (2009).

476 43 Floor, S. N. & Doudna, J. A. Tunable protein synthesis by transcript isoforms
477 in human cells. *Elife* **5**, doi:10.7554/eLife.10921 (2016).

478 44 Semenza, G. L. & Wang, G. L. A nuclear factor induced by hypoxia via de novo
479 protein synthesis binds to the human erythropoietin gene enhancer at a site
480 required for transcriptional activation. *Mol Cell Biol* **12**, 5447-5454,
481 doi:10.1128/mcb.12.12.5447 (1992).

482 45 Wang, G. L. & Semenza, G. L. Purification and characterization of hypoxia-
483 inducible factor 1. *J Biol Chem* **270**, 1230-1237, doi:10.1074/jbc.270.3.1230
484 (1995).

485 46 Belaiba, R. S. *et al.* Hypoxia up-regulates hypoxia-inducible factor-1alpha
486 transcription by involving phosphatidylinositol 3-kinase and nuclear factor
487 kappaB in pulmonary artery smooth muscle cells. *Mol Biol Cell* **18**, 4691-4697,

doi:10.1091/mbc.e07-04-0391 (2007).

47 Galban, S. *et al.* RNA-binding proteins HuR and PTB promote the translation of hypoxia-inducible factor 1alpha. *Mol Cell Biol* **28**, 93-107, doi:10.1128/MCB.00973-07 (2008).

48 Hui, A. S., Bauer, A. L., Striet, J. B., Schnell, P. O. & Czyzyk-Krzeska, M. F. Calcium signaling stimulates translation of HIF-alpha during hypoxia. *FASEB J* **20**, 466-475, doi:10.1096/fj.05-5086com (2006).

49 Vicens, Q., Kieft, J. S. & Rissland, O. S. Revisiting the closed-Loop model and the nature of mRNA 5'-3' communication. *Mol Cell* **72**, 805-812, doi:10.1016/j.molcel.2018.10.047 (2018).

50 Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **18**, 31-42, doi:10.1038/nrm.2016.132 (2017).

Figures and figure legends

Figure 1. Systematic identification of cap-independent translation elements. a, Unbiased screen of cap-independent translation sequences using a circRNA reporter containing a split EGFP. Total HeLa RNAs were fragmented into 200-300nt fragments, reverse-transcribed, normalized and inserted into the circRNA reporter with Fln backbone to generate a plasmid library. This library was co-transfected with pOG44 (ratio 1:9) into FLP-In™-293 cells and selected by hygromycin to generate a library of stably transfected cells that were further sorted by FACS. The library was sequenced at each stage to determine sequence complexity. **b,** Genomic distribution of inserted sequences in pre-sorting and post-sorting cells (N indicates the number of fragments).

c, The read coverage plots for MBTPS1 and UQCR10 as representative examples. **d**, Validation of the newly identified fragments (indicated by starting position in human genome and offset) using circRNA translation reporters. The fragments of both high (>10,000) and low (1 to 3) read depth were selected for testing. **e**, Validation of the newly identified fragments for cap-independent translation activation using bicistronic translation reporters inserted with different CiTIs. The reporters were transfected into HEK293 cells, and the products were determined by luminescence assay ($n = 3$, mean $\pm$ SD, * indicated $p < 0.05$ by Student's t-test). **f**, Positional density plots of the pre-sorting fragments and the CiTIs were in different regions of the gene body.

Figure 2. The 3'-CiTIs enhance translation of mRNAs with structured 5'UTRs.

a-b, Sequence conservation the structure (predicted MFE, and GC contents) of the 5'UTRs from all human genes vs. the 3'-CiTI-containing genes. In **panels a,b and h**, p values were calculated by unpaired two-sided Mann–Whitney U test. **c**, The dual-luciferase reporter containing a hairpin at the 5'UTR of Rluc and different CiTIs in its 3'UTR. The translation efficiency of Rluc was determined by the relative abundance

of proteins (*via* luminescence assay) vs RNAs (*via* qPCR) normalized to the control Fluc. In this and the following panels (**c-g**), same procedures were used with Student's t-test to evaluate statistical significance ($n = 3$, mean $\pm$ SD, * indicated $p < 0.05$). **d**, Different combinations of endogenous structured 5'UTRs (shown as RNAfold prediction) and their cognate 3'-CiTIs were inserted into the Rluc reporter to test their effects on translation. **e**, Different Rluc mRNAs with an optional 5'-cap or poly-A were transcribed *in vitro*, and co-transfected with control Fluc mRNA into HEK293 cells to test translation efficiency. **f**, Introduction of mRNA looping through complementary sequences enhanced the activity of the 3'-CiTI. The experimental procedures are same as **panel e**. **g**, Two potential mechanisms for translation initiation through a 5'UTR structure (ribosome scanning vs shunting). The reporters with different structures in their 5'UTR and optional 3'-CiTIs were transfected into HEK293 cells, and the translation efficiency was determined by the relative abundance of Rluc proteins (*via* luminescence assay) vs RNAs (*via* qPCR) normalized to the control Fluc. **h**, *In vivo* icSHAPE signals of the hairpin regions in uAUG-hp35 mRNA with optional 3'-CiTI mRNAs. The icSHAPE signals of the hairpin region were plotted, and regions with increased opened structure by 3'-CiTI are indicated with red arrows.

Figure 3. Trans-factors involved in 3'-CiTI functions. **a**, Schematic diagram of RNP affinity purification. **b**, The protein-protein interaction network of identified CiTI-binding *trans*-factors. **c**, DHX29 knockdown reduces the 3'-CiTI activity. The reporters with a 5'UTR hairpin and different 3'-CiTIs were transfected into HEK293 cells after DHX29 knockdown, and the translation efficiency of Rluc was determined as in Fig. 2c ($n = 3$, mean $\pm$ SD in **panel c and d**, * indicates $p < 0.05$ by Student's t -test, n.s. = not significant). **d**, Tethering of different *trans*-factors promotes translation of mRNAs with structured 5'UTRs. The translation reporter containing 6xMS2 sites was co-transfected with expression vectors of different fusion proteins, and the translation efficiency was measured as in **panel c**. **e**, Polysome profiling of SH-SY5Y cells after DHX29 knockdown. Mono: monosome; LP: light polysome (2-4

ribosomes); HP: heavy polysome (>4 ribosomes). The RNAs from each fraction were sequenced using the ribo-minus protocol. **f**, Global translation changes after DHX29-knockdown. The translation efficiency of each gene was defined by the percentage of polysome-associated mRNAs (data of **panel e**). The genes with consistent profiles in both knockdown lines were selected to plot percent difference of the polysome-associated mRNAs between knockdown cells vs. control cells ($\pm 25\%$ cutoffs were indicated). **g**, Relative distributions in polysome profiles were plotted as heatmap. The genes harboring altered distribution in DHX29 knockdown SH-SY5Y cells were shown. **h**, The predicted MFE of 5'UTRs from genes with reduced translation efficiency by DHX29 knockdown (p values calculated with unpaired two-sided Mann–Whitney U test).

577

578

Figure 4. Biological function of HIF1A 3'-CiTI. **a**, Different translation reporters containing various combinations of the 5'UTR and 3'-CiTI from HIF1A were transfected in SH-SY5Y or HCT116 cells, and the translation efficiencies were examined as in Fig. 2c under hypoxia (1% O₂) (n = 3, mean ± SD, Student's t-test in **panel a-b**). **b**, The activity of the HIF1A 3'-CiTI was reduced in SH-SY5Y cells after DHX29-knockdown under hypoxia (1% O₂). **c**, HA-tagged DHX29 was co-

transfected with flag-tagged HIF1A (full length or 3'-CiTI deletion) into SH-SY5Y cells, and the interaction of HIF1A RNA and DHX29 was measured by RNA-IP and qPCR (n = 3, mean ± SD, Student's t-test). **d**, The SH-SY5Y cells with deletion of HIF1A 3'-CITI were analyzed with polysome profiling after 6 h under hypoxia. Top, polysome profiling of WT and two independent deletion clones. Middle and bottom, the mRNA abundance of HIF1A and ACTB control in each fraction. **e**, WT cells and two 3'-CITI deletion clones were collected at 1 h, 2 h, 4 h, and 6 h after hypoxia, and the protein or RNA level of HIF1A was determined by western blot (left) or RT-qPCR (right). Middle, quantification of western blot (n = 3, mean ± SD, Student's t-test). **f**, left, proliferation of WT SH-SY5Y cells and cells with HIF1A 3'-CiTI deletion under hypoxia as measured with CCK8 assay (n = 3, mean ± SD, Student's t-test). The HIF1A knockdown cells were used as a control (right). **g**, Growth of xenograft tumor generated with WT SH-SY5Y cells or two clones of HIF1A 3'-CiTI deletion (n = 9, mean ± SEM, Student's t-test). **h**, The weights and images of the xenograft tumors 4 weeks after injection. **i**, New model for 3'-CiTI-mediated translation regulation.