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. Ludwig [3], Sirui Zhang[1, 2], Qianyun Lu[1, 2], Cindy L. Will[3], Henning Urlaub[4], Jing Sun[1], Reinhard Lührmann[3], 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

[5]Lead 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 sequences [8].

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

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

116 Although the canonical IRESs are mostly located in the 5′UTR, the newly
117 identified CiTIs were found in all gene regions including 5′UTRs (termed 5′-CiTI),
118 CDS (termed C-CiTI), and 3′UTRs (termed 3′-CiTI). Consistent with our earlier
119 finding, the CiTIs were enriched in 3′UTRs and peaked in the CDS near the start and
120 stop codons (Fig. 1f). Compared to the cognate controls, we found a significant
121 increase of sequence conservation in 3′-CiTIs, a modest increase in 5′-CiTIs, and no
122 increase in C-CiTIs (Extended data Fig. 2d). The CiTIs generally had less structures
123 as judged by higher minimal free energy (MFE) (Extended data Fig. 2e), whereas the
124 canonical IRESs were significantly more structured [18], suggesting that, unlike IRESs,
125 most CiTIs do not function by forming a specific structure.

126 Several enriched motifs from the CiTIs, including the previously reported U-rich
127 motifs [8,19], were identified and further experimentally validated (Extended data Fig.
128 3a-3d). Consistent with previous reports [8,20], we found that the 5′-CiTIs tend to
129 interact with 18S rRNAs *via* base-paring, and such interaction is sufficient to drive
130 cap-independent translation (Extended data Fig. 3e-3g). Therefore, our circRNA-
131 based translation system successfully identified several categories of CiTIs that

132 contain both previously confirmed elements/motifs and novel elements enriched
133 within 3′UTRs.
134
135 **Genes with 3′-CiTIs tend to contain downstream ORFs or structured 5′UTRs**
136 The surprising enrichment of CiTIs in 3′UTRs rather than 5′UTRs implied an
137 unexpected new role of 3′UTR in promoting translation initiation. Naturally, the 3′-
138 CiTI may function as a *bona fide* IRES to drive cap-independent translation of
139 downstream ORFs (dORF). By analyzing all potential dORFs, we found that the
140 dORFs in the genes with 3′-CiTIs were more conserved than those without 3′-CiTIs
141 (Extended data Fig. 4a), suggesting that 3′-CiTIs are probably under functional
142 selection to drive dORF translation. Further analysis using published datasets of
143 ribosome footprints [21,22] showed that a significant fraction of these dORFs (177/564)
144 also had an upstream 3′-CiTI (Extended data Fig. 4b, $p=2.9x10^{-29}$). However, we only
145 identified the potential translation products from 7 of these dORFs using public mass
146 spectrometry datasets [23,24]. The lack of strong proteomic evidences may be due to cell-
147 or tissue-specific translation of these dORFs, and/or the rapid degradation of these
148 dORF-encoded proteins just like 3′UTR-coded peptides from translation read-through
149 [25,26]. These observations suggest that driving cap-independent translation of dORFs
150 might not be the function for most 3′-CiTIs.
151 The fact that >90% of 3′-CiTI-containing genes lack evidence for dORF
152 translation (Extended data Fig. 4b) suggested unknown roles of 3′-CiTIs.
153 Interestingly, the 3′-CiTI-containing genes were significantly more conserved and

154 more structured in their 5′UTRs than those without 3′-CiTIs (Fig. 2a-2b). They were
155 also significantly overlapped with the genes containing upstream ORFs (uORF)
156 (Extended data Fig. 4c) [27,28]. Since most 5′UTR structures and uORFs play inhibitory
157 roles in regulating translation [29], we speculated that 3′-CiTIs may counteract these
158 5′UTR structures or uORFs to enhance the main ORF (mORF) translation.

159 Such possibility was directly tested using four translation reporters with different
160 5′UTRs (Extended data Fig. 5a): a canonical reporter with a regular 5′UTR, a hairpin
161 reporter containing a 5′UTR hairpin immediately before the ORF, and two uORF
162 reporters with the 5′UTRs containing two endogenous uORFs [30,31]. We found that 6
163 out of 7 CiTIs inserted into the 3′UTR of the hairpin reporter significantly increased
164 the translation (>10-fold increase) (Fig. 2c), and this translation enhancement was
165 independent of the CiTI length or the cell lines used (Extended data Fig. 5b-5c). In
166 contrast, no reliable translation increasement or even inhibitory effect was detected
167 when CiTIs were inserted into the canonical reporter (Extended data Fig. 5d). As
168 another comparison, insertion of CiTIs into the 3′UTR of the uORF reporters showed
169 a weaker effect (Extended data Fig. 5e).

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

176 Most human mRNAs contain a 5′-cap and poly-A tail that are looped together by
177 a protein complex to increase translation efficiency [32-34]. To determine if this looping
178 is required for 3′-CiTI activity, we transfected cells with *in vitro* synthesized mRNAs
179 containing optional 5′-cap or poly-A tail (Fig. 2e). In all mRNAs tested, insertion of a
180 3′-CiTI reversed the translation suppression by the 5′- hairpin, indicating the
181 translation promotion by 3′-CiTI is independent of the 5′-cap or poly-A tail (Fig. 2e).
182 This finding was further validated with a PolI reporter that can produce RNAs without
183 a 5′-cap or poly-A tail, where a 3′-CiTI again enhanced translation of these "naked"
184 mRNAs containing a 5′-hairpin (Extended data Fig. 6). Interestingly, the translation of
185 mRNA that lacks a 5′-cap or poly-A tail could be enhanced by simply introducing a
186 mRNA loop across the ORF using complementary sequences, and insertion of a 3′-
187 CiTI further increased the translation of such mRNAs (Fig. 2f).
188

189 **3′-CiTIs unwind the 5′UTR structures to facilitate ribosome scanning**

190 During translation initiation, the preinitiation complex (PIC) usually needs to
191 scan through the 5′UTR structures to initiate translation [2,35-37]. Occasionally, the PIC
192 can "jump over" the 5′UTR structures, a phenomenon called ribosome shunting [38]
193 (Fig. 2g, left). To determine how 3′-CiTIs enhance translation of mRNAs containing
194 5′UTR structures, we generated several reporters with 5′-hairpins of different lengths
195 and an upstream in-frame start codon (uAUG) inside the hairpin (Extended data Fig.
196 6a). The first start codon uAUG will be used by default to generate an extended Rluc
197 (eRluc) *via* ribosome scanning. Alternatively, the PIC may "jump over" the structure

198 and use the second AUG (i.e., without structure unwinding) to produce the canonical
199 Rluc.
200 We found that the uAUG was exclusively used during translation, suggesting a
201 ribosome scanning in all reporter mRNAs (Fig. 2g). As expected, introduction of a 5′-
202 hairpin reduced the translation efficiency in a length-dependent fashion, and insertion
203 of a 3′-CiTI increased the translation from the uAUG as judged by both western blots
204 and luciferase activities (Fig. 2g). This increasement was achieved *via* promoting
205 ribosome scanning rather than ribosome shunting, as only the proximal start codon
206 was used in all cases. Additionally, we found that the reporter mRNA containing a 3′-
207 CiTI had less structures (as judged by icSHAPE signals) than the mRNA without the
208 3′-CiTI, in both the entire 5′UTR and its hairpin region (Fig. 2h, and Extended data
209 Fig. 7b-c), suggesting that the 3′-CiTI may unwind the upstream structures to help
210 ribosomes scan through the 5′UTR.
211
212 ***Trans*-acting factors are involved in 3′-CiTI-mediated translation enhancement**
213 To determine if the 3′-CiTIs recruit *trans*-acting factors to promote translation,
214 we performed RNA affinity purification (Fig. 3a) using the MS2-fused RNAs
215 containing three different 3′-CiTIs and controls (Fig. 2c, Extended data Fig. 8a). The
216 purified RNP complexes were analyzed with mass spectrometry to identify 63
217 candidate proteins that specifically bind all three CiTIs (Extended data Fig. 8b and
218 Table 2). These CiTI-binding proteins formed a densely-connected protein-protein
219 interaction network with key functions in translation initiation and RNA processing

220 (Fig. 3b), including many components in the eIF2 or eIF3 complex that functions as a
221 multi-tasking machine to promote translation [39]. Interestingly, the newly identified
222 CiTI-binding proteins also include DHX29 (Fig. 3b), a DExH-box RNA helicase that
223 might help to unwind RNA structures.

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

234 Previously, DHX29 was shown to regulate the translation of a handful of genes,
235 including Cdc25C and XIAP [41,42]. However, ~2/3 of human mRNAs have structured
236 5′UTRs (i.e., with low MFE, $\Delta G < -40$ kcal/mol) [42], many of which may also be
237 regulated by DHX29. To systematically identify *bona fide* DHX29 targets, we
238 performed polysome profiling after DHX29 knockdown, and sequenced mRNAs from
239 different gradient fractions (Fig. 3e and Extended data Fig. 8c). We focused on the
240 genes with consistent profiles in two independent DHX29 knockdown lines (Pearson
241 correlation > 0.8), and calculated the translation efficiency by the percent of

242 polysome-associated mRNAs [43]. In total, 869 genes were identified with at least a
243 25% change in translation efficiency (Fig. 3f), most of which showed a translation
244 reduction (Fig. 3f and 3g, Extended data Table 3). Furthermore, the genes with
245 reduced translation efficiency in DHX29 knockdown cells had more RNA structures
246 in their 5′UTR (Fig. 3h), and 91 of these genes also contained 3′-CiTIs (p=0.001 by
247 hypergeometric test). Collectively, these results suggested that DHX29 plays a critical
248 role in unwinding mRNA structures and promotes translation upon specific
249 recruitment by 3′-CiTIs.
250

251 **3′-CiTI and DHX29 promote HIF1A translation in response to hypoxia *in vitro***
252 **and *in vivo***

253 An important new target of DHX29 is the hypoxia-induced transcription factor
254 HIF1A, which contains both a 3′-CiTI and strong 5′UTR structure [44,45]. As a key
255 factor for cellular response to hypoxia, HIF1A translation is suppressed in normoxia
256 but strongly induced by oxygen deprivation [46-48]. Using translation reporters
257 containing HIF1A 5′UTR and different fragments of its 3′UTR, we found that the
258 5′UTR significantly suppressed translation under hypoxia, however the HIF1A 3′-
259 CiTI fully restored translation in two different cell lines (Fig. 4a, top). Conversely,
260 deletion of the 3′-CiTI from the HIF1A 3′UTR reduced translation under hypoxia
261 (Fig. 4a, bottom). In contrast, under normoxia condition, translation enhancement by
262 the 3′-CiTI was observed only in HCT116 cells (Extended data Fig. 9a-b).

263 The translation promotion by the HIF1A 3′-CiTI was reduced by DHX29
264 knockdown under hypoxia (Fig. 4b). We further observed a direct interaction of
265 DHX29 with HIF1A mRNAs (Extended data Fig. 9c), consistent with its general
266 function as a subunit of the translation initiation complex [40]. Interestingly, the
267 interaction between DHX29 and HIF1A mRNA was nearly eliminated by the 3′-CiTI
268 deletion under hypoxia, whereas this interaction was only reduced by ~50% during
269 normoxia (Fig. 4c), suggesting that the 3′-CiTI is essential for DHX29 recruitment to
270 HIF1A mRNA during hypoxia.
271 To evaluate the functional impact of 3′-CiTI-mediated translation promotion, we
272 deleted the 3′-CiTI from endogenous HIF1A gene in SH-SY5Y neuroblastoma cells
273 with CRISPR-cas9 (Extended data Fig. 9d). Under normoxia, HIF1A mRNA was
274 mostly found in monosomes or light polysomes (Extended data Fig. 9e), consistent
275 with its low translation efficiency under this condition. Hypoxia induced the
276 association of most HIF1A mRNAs with polysomes (Fig. 4d). However, deletion of
277 the 3′-CiTI shifted HIF1A mRNAs (but not control mRNAs) from the polysome
278 factions to the monosome fractions (Fig. 4d), suggesting that 3′-CiTI is required to
279 specifically promote translation of HIF1A mRNA in response to hypoxia.
280 Furthermore, deletion of the 3′-CiTI significantly blunted the induction of HIF1A
281 protein by hypoxia (Fig. 4e), resulting in a small but significant reduction in cell
282 proliferation similar to that in HIF1A knockdown cells (Fig. 4f and Extended data Fig.
283 9f). Importantly, when inoculated in nude mice, 3′-CiTI deleted SH-SY5Y
284 neuroblastoma tumors exhibited a dramatic reduction of growth (Fig. 4g and 4h).

285 Taken together, these observations indicate that the 3′-CiTI in HIF1A mRNA
286 functions as a key *cis*-element that recruits DHX29 to mediate environmental
287 regulation of its translation, which plays a critical role in supporting *in vivo* tumor
288 growth.
289

290 **Discussion**

291 By unbiasedly screening human transcriptome, we uncovered a new class of
292 3′UTR elements that play an unexpected role in translation initiation. Specifically, the
293 3′-CiTIs actively recruit key initiation factors to unwind inhibitory RNA structures in
294 the 5′UTRs, thereby promoting the ribosomes scanning (Fig. 4i). Since actively
295 translated mRNAs generally form a loop through interactions of poly-A binding
296 proteins with eIF4G [49], the 3′UTR may generally serve as a repository of regulatory
297 elements that "catch and store" the translation initiation factors to increase their local
298 concentration for the next round of translation.

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

306 when inserted into the suitable context, although they function as translation
307 initiators/enhancers from 3′UTR in endogenous settings.
308 Our findings further indicate that the 3′-CiTI-mediated translational control is an
309 ancient and common mode of regulation. Firstly, the 3′-CiTIs are significantly more
310 conserved than the matched control regions (Extended data Fig. 1d) and the 5′UTR in
311 the 3′-CiTI-containing mRNAs are also more conserved and more structured (Fig. 2a-
312 2b), indicating that the pairs of 3′-CiTIs and 5′UTRs are under functional selection in
313 evolution. Moreover, ~44% of the 2681 newly identified 3′-CiTIs are located in
314 human mRNAs with structured 5′UTRs ($\Delta G < -40$ kcal/mol)[42], suggesting a common
315 role of 3′-CiTIs in counteracting the inhibitory 5′UTR structures in many genes.
316 This study also highlights the importance of DHX29 in 3′-CiTI-mediated
317 translation initiation in response to environmental stress. Several RNA helicases
318 (including DHX29, DHX9 and DDX3) were reported to promote translation [29,40]. In
319 particular, DHX29 has been shown to interact with eIF3 to form an active
320 43S/DHX29 complex [41]. However, it is unclear how these helicases are specifically
321 recruited to mRNAs. Our data indicate that many 3′-CiTIs can function as the *cis*-
322 acting elements to recruit DHX29 (and eIF3). The biological impact of such
323 regulation is exemplified by HIF1A: deletion of a 3′-CiTI on the endogenous *HIF1A*
324 locus directly impairs its translation and reduces cell survival under hypoxia, which
325 further blunts *in vivo* tumor progression.
326 Given the various stress conditions where the cap-dependent translation initiation
327 is inhibited (e.g., hypoxia, viral infection, heat/cold shock, etc.), the cooperative

328 regulation of translation by the structured 5′UTR and 3′-CiTI reported here adds an
329 additional layer of regulatory complexity during stress. Future studies are needed to
330 determine how CiTI-mediated translational initiation can interplay with other modes
331 of translation control from 3′UTRs, such as regulation by m6A and miRNAs [4,50].
332 In summary, our study uncovers an uncanonical mechanism of 3′UTR in
333 mediating translation initiation, and further reveals the importance of this regulation
334 in stress-induced cell survival and proliferation *in vitro* and *in vivo*. These findings
335 may pave the way for novel cancer therapies that target stress-induced translation.
336
337

Data availability

The raw sequencing data and filtered datasets are publicly available from NODE (assess number: OEP001285) and NCBI (assess 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 Plank 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).

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

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

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

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

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

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

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

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

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

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

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

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

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

31 Lee, Y. Y., Cevallos, R. C. & Jan, E. An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2alpha phosphorylation. *J Biol Chem* **284**, 6661-6673, doi:10.1074/jbc.M806735200 (2009).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

46 Belaiba, R. S. *et al.* Hypoxia up-regulates hypoxia-inducible factor-1alpha transcription by involving phosphatidylinositol 3-kinase and nuclear factor 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).

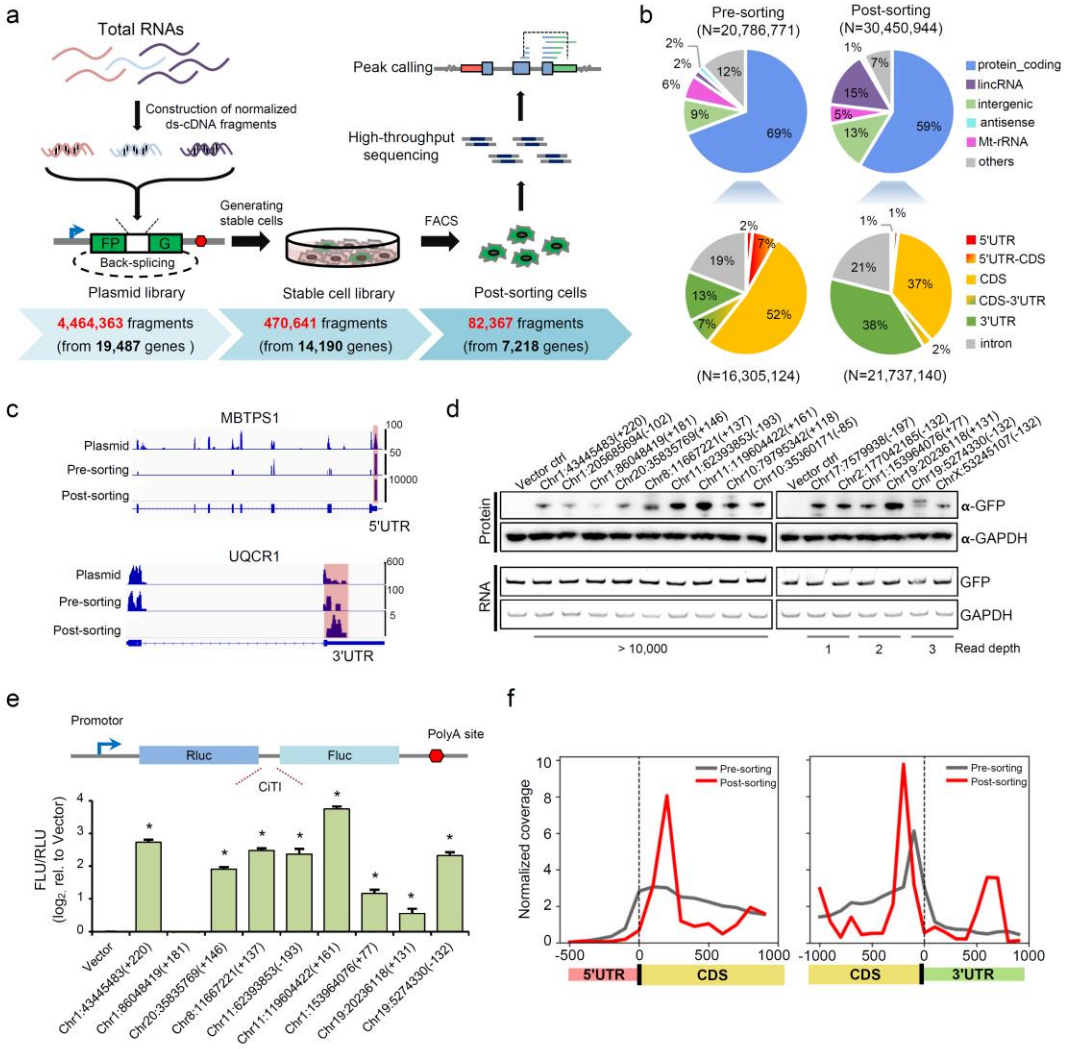
48 Hui, A. S., Bauer, A. L., Striet, J. B., Schnell, P. O. & Czyzyk-Krzeska, M. F. Calcium signaling stimulates translation of HIF-alpha during hypoxia. *FASEB J* **20**, 466-475, doi:10.1096/fj.05-5086com (2006).

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).

503 **Figures and figure legends**

504



505

506 **Figure 1. Systematic identification of cap-independent translation elements. a**,
507 Unbiased screen of cap-independent translation sequences using a circRNA reporter
508 containing a split EGFP. Total HeLa RNAs were fragmented into 200-300nt fragments,
509 reverse-transcribed, normalized and inserted into the circRNA reporter with FlpIn
510 backbone to generate a plasmid library. This library was co-transfected with pOG44
511 (ratio 1:9) into Flp-In™-293 cells and selected by hygromycin to generate a library of
512 stably transfected cells that were further sorted by FACS. The library was sequenced at
513 each stage to determine sequence complexity. **b**, Genomic distribution of inserted
514 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 ± 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.

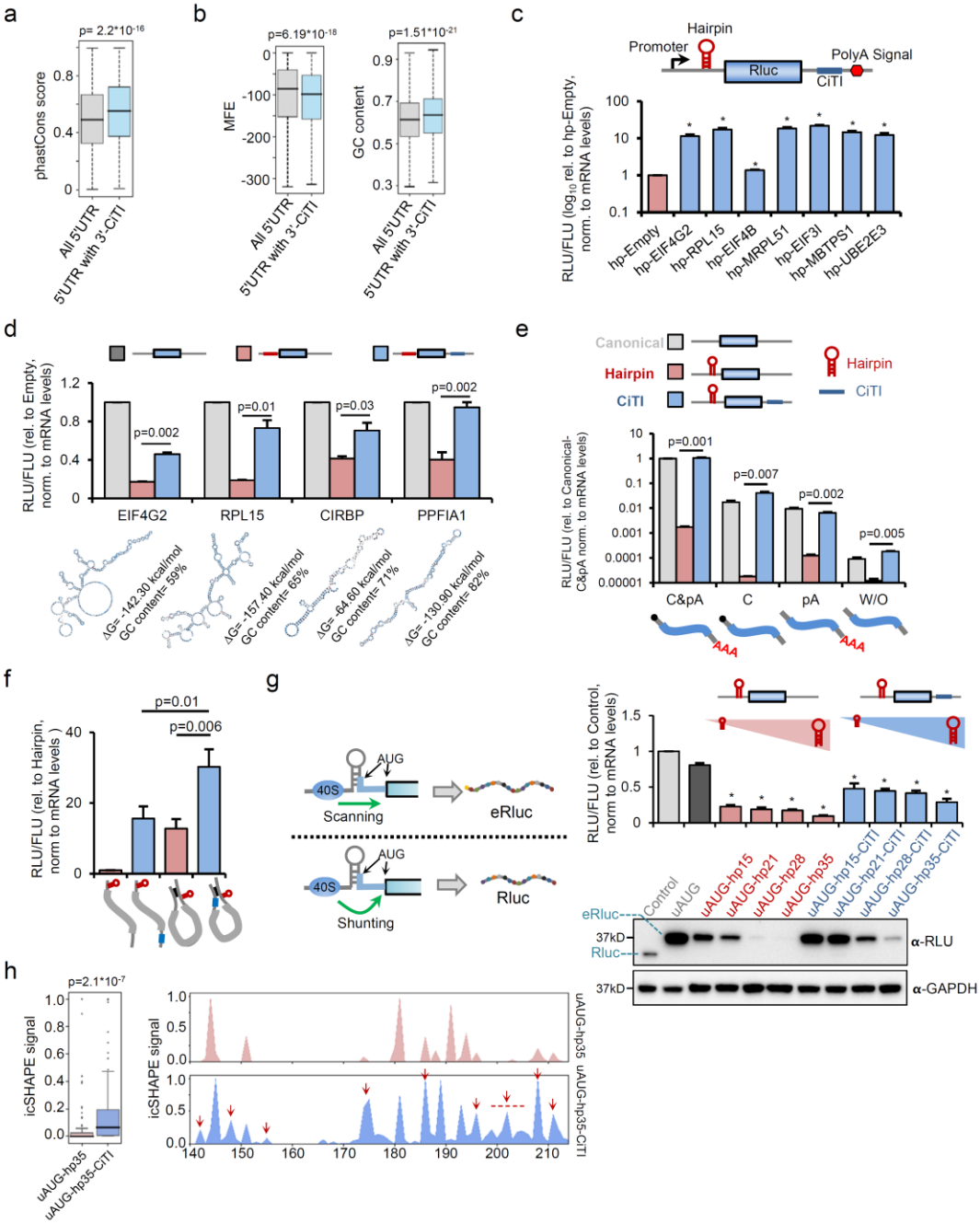


Figure 2. The 3′-CiTIs enhance translation of mRNAs with structured 5′UTRs. **a-b**, Sequence conservation the structure (predicted MFE, and GC contents) of the 5′UTRs from all human genes *vs.* the 3′-CiTI-containing genes. In **panels a,b and h**, p values were calculated by unpaired two-sided Mann–Whitney 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

533 of proteins (*via* luminescence assay) *vs* RNAs (*via* qPCR) normalized to the control
534 Fluc. In this and the following panels **(c-g)**, same procedures were used with Student′s
535 t-test to evaluate statistical significance (n = 3, mean ± SD, * indicated $p<0.05$). **d**,
536 Different combinations of endogenous structured 5′UTRs (shown as RNAfold
537 prediction) and their cognate 3′-CiTIs were inserted into the Rluc reporter to test their
538 effects on translation. **e**, Different Rluc mRNAs with an optional 5′-cap or poly-A
539 were transcribed *in vitro*, and co-transfected with control Fluc mRNA into HEK293
540 cells to test translation efficiency. **f**, Introduction of mRNA looping through
541 complementary sequences enhanced the activity of the 3′-CiTI. The experimental
542 procedures are same as **panel e**. **g**, Two potential mechanisms for translation
543 initiation through a 5′UTR structure (ribosome scanning *vs* shunting). The reporters
544 with different structures in their 5′UTR and optional 3′-CiTIs were transfected into
545 HEK293 cells, and the translation efficiency was determined by the relative
546 abundance of Rluc proteins (*via* luminescence assay) vs RNAs (*via* qPCR)
547 normalized to the control Fluc. **h**, *In vivo* icSHAPE signals of the hairpin regions in
548 uAUG-hp35 mRNA with optional 3′-CiTI mRNAs. The icSHAPE signals of the
549 hairpin region were plotted, and regions with increased opened structure by 3′-CiTI
550 are indicated with red arrows.
551

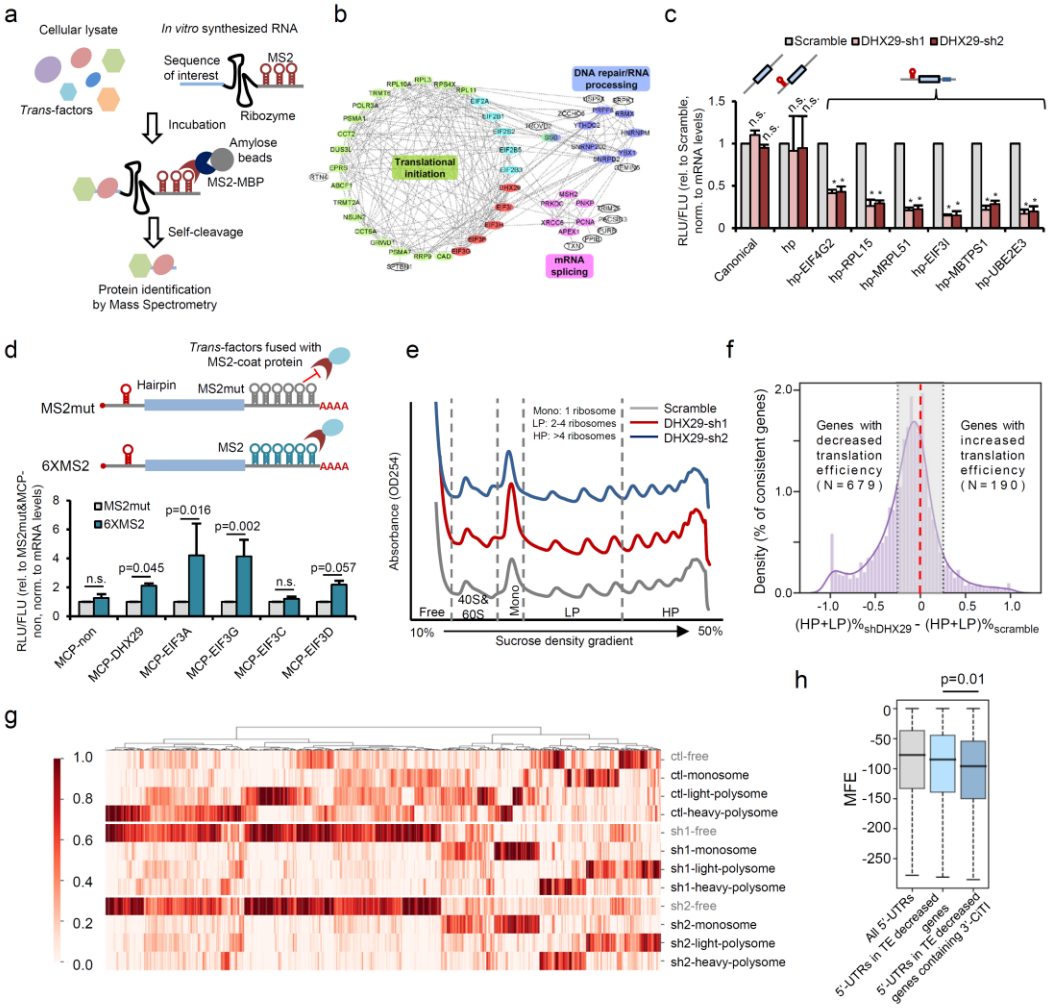
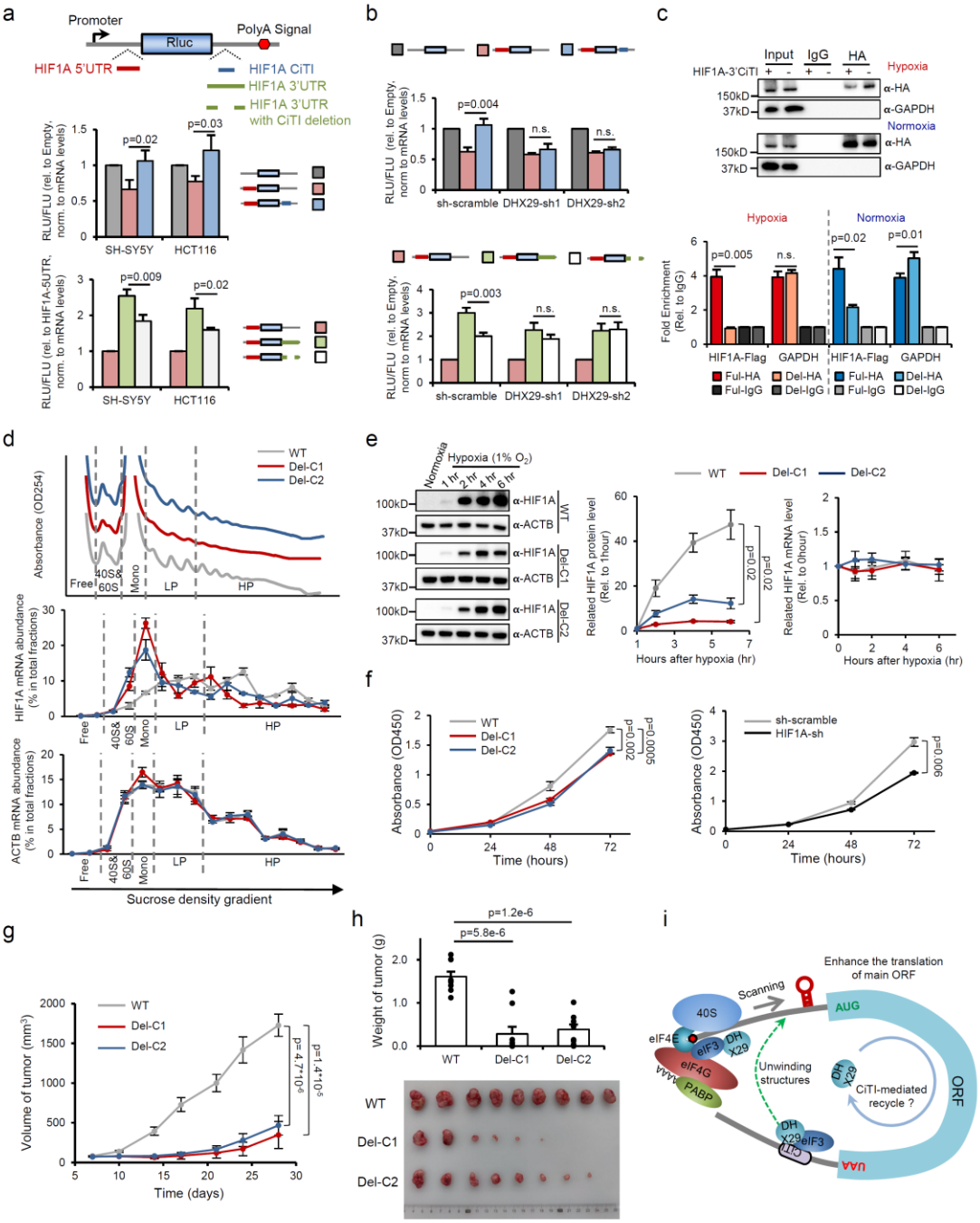


Figure 3. *Trans*-factors involved in 3′-CiTI functions. a, Schematic diagram of RNP affinity purification. **b**, The protein-protein interaction network of identified CiTI-binding *trans*-factors. **c**, DHX29 knockdown reduces the 3′-CiTI activity. The reporters with a 5′UTR hairpin and different 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 ± 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

565 ribosomes); HP: heavy polysome (>4 ribosomes). The RNAs from each fraction were
566 sequenced using the ribo-minus protocol. **f**, Global translation changes after DHX29-
567 knockdown. The translation efficiency of each gene was defined by the percentage of
568 polysome-associated mRNAs (data of **panel e**). The genes with consistent profiles in
569 both knockdown lines were selected to plot percent difference of the polysome-
570 associated mRNAs between knockdown cells *vs.* control cells (± 25% cutoffs were
571 indicated). **g**, Relative distributions in polysome profiles were plotted as heatmap. The
572 genes harboring altered distribution in DHX29 knockdown SH-SY5Y cells were
573 shown. **h**, The predicted MFE of 5′UTRs from genes with reduced translation
574 efficiency by DHX29 knockdown (p values calculated with unpaired two-sided
575 Mann–Whitney U test).
576



578

579 **Figure 4. Biological function of HIF1A 3′-CiTI. a**, Different translation reporters
580 containing various combinations of the 5′UTR and 3′-CiTI from HIF1A were
581 transfected in SH-SY5Y or HCT116 cells, and the translation efficiencies were
582 examined as in Fig. 2c under hypoxia (1% O_2) (n = 3, mean ± SD, Student′s t-test in
583 **panel a-b**). **b**, The activity of the HIF1A 3′-CiTI was reduced in SH-SY5Y cells after
584 DHX29-knockdown under hypoxia (1% O_2). **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.