Targeted Inhibition of Translation Initiation via 2'O-Methylation of Start Codons

Emmely A. Patrasso[1,6], Hannah Serio[1,6], Adam Suh[1], Stephanie Mou[2], Kevin Vasquez[1,3], Siqi Wu[4], Yang Yi[4,5], Daniel Arango[1,5,*]

[1]Department of Pharmacology. Feinberg School of Medicine. Northwestern University. Chicago, IL. USA

[2]Driskill Graduate Program in Life Sciences. Feinberg School of Medicine. Northwestern University. Chicago, IL. USA.

[3]Department of Chemical & Biological Engineering. McCormick School of Engineering. Northwestern University. Evanston, IL. USA.

[4]Department of Urology. Feinberg School of Medicine. Northwestern University. Chicago, IL. USA.

[5]Robert H. Lurie Comprehensive Cancer Center. Feinberg School of Medicine. Northwestern University. Chicago, IL. USA.

[6] These authors contributed equally to this work

[*] To whom correspondence may be addressed: daniel.arangotamayo@northwestern.edu

Abstract

While protein synthesis typically initiates at an optimal AUG start codon, the 5' untranslated region (5'UTR) of mRNAs harbors non-canonical start codons that result in the translation of upstream Open Reading Frames (uORFs). However, the mechanisms underlying the selection of non-canonical start codons remain poorly understood. Structural analysis of translation pre-initiation complexes showed that the 2'OH group of the first nucleotide within start codons is monitored by 18S rRNA, allowing optimal translation initiation. We identified a chemical modification of mRNA, 2'O-methylation (N_m), that represses translation initiation by interfering with the interaction of mRNA and 18S rRNA, thereby preventing the pre-initiation complex from recognizing start codons. Significantly, we uncovered the presence of 2'O-methylation in the first nucleotide of non-canonical initiation sites within 5'UTRs, where it repressed upstream translation initiation. These findings revealed a potent regulatory function of 2'O-methylation in translation initiation and uORF expression, providing new insights into the mechanisms of non-canonical start codon selection in human cells.

Keywords

2'O-methylation, N_m, non-AUG translation, uORF, epitranscriptome

38 **Main**

39 During translation initiation, a 43S ribonucleoprotein particle containing the small ribosome
40 subunit (40S), the initiator tRNA ($tRNA_i^{met}$), and translation initiation factors, binds the 5' end of
41 mRNAs to form a 48S pre-initiation complex.[1] This complex scans through the 5'UTR until it
42 encounters an optimal start codon.[1] The optimality of a start codon is typically determined by the
43 Kozak sequence (RccAUGG, where R is either adenosine or guanosine).[2] Once the start codon
44 is recognized, GTP hydrolysis and rearrangement of the 48S complex enable the large 60S
45 subunit to bind and form the 80S initiator ribosome.[1] Subsequently, the 80S ribosome moves into
46 elongation and continues decoding through an Open Reading Frame (ORF) until a stop codon is
47 reached.[1]

48

49 While the Kozak sequence defines the optimal translation initiation site, translation can also
50 initiate at suboptimal or cryptic start sites characterized by the presence of AUGs in non-Kozak
51 sequences or near-cognate start codons, which differ from AUG by one nucleotide (CUG, GUG,
52 UUG, AAG, ACG, AGG, AUA, AUC, and AUU).[3] While these cryptic start codons are usually
53 bypassed by the 48S pre-initiation complex through a process known as leaky scanning,[4]
54 translation initiation at upstream sites results in the expression of upstream Open Reading
55 Frames (uORFs).[3] Initially thought to be mere cis-regulatory sequences that inhibit the expression
56 of downstream canonical ORFs, recent studies demonstrated that uORFs also yield stable
57 micropeptides with diverse biological functions, many of which are involved in neurological or
58 metabolic diseases and cancers.[5-7] However, the mechanisms underlying the selection of non-
59 canonical start codons and uORF expression under normal homeostasis and in disease remain
60 unclear.

61

62 One poorly characterized layer of translation regulation is the epitranscriptome, defined as the set
63 of more than 150 chemical modifications of ribonucleotides that alter RNA function. Several RNA
64 modifications have been reported to regulate translation initiation. For instance, the hypermodified
65 nucleotide 6-threonyl-carbamoyl-adenosine (t^6A) in position 37 of $tRNA_i^{met}$ directly interacts with
66 Kozak sequences, where it stabilizes the translation pre-initiation complex.[8,9] Moreover, the
67 mRNA cap is characterized by a non-templated 7-methylguanosine (m^7G), followed by one or two
68 templated nucleosides with methylation in the 2'OH of the ribose, a modification called 2'O-
69 methylation (also referred to as A_m, C_m, G_m, U_m, or simply N_m).[10] When the first nucleotide is an
70 adenosine, it is usually further modified into N^6,2′O-dimethyladenosine (m^6A_m).[11] The m^7G cap is
71 essential for translation, whereas the subsequent N_m modifications in the first or second

nucleotide regulate mRNA stability.[10-12] Beyond the cap, several mRNA modifications are found internally in the 5'UTRs, including m^6A,[13-15] N4-acetylcytidine (ac^4C),[8,16] and 2′O-methylation.[17-20] While m^6A regulates cap-independent translation by recruiting 'reader' proteins,[10-12] and ac^4C weakens the interaction of $tRNA_i^{met}$ with the Kozak sequence,[8] the function of 2′O-methylation within 5'UTRs is unknown.

Using synthetic mRNAs with 2′O-methylation in every position of the Kozak sequence and re-analysis of cryo-EM structures of translation pre-initiation complexes, we found that 2'O-methylation in the first nucleotide of the start codon completely abolishes protein synthesis. Mechanistically, 2'O-methylation interferes with the interactions between mRNA and 18S rRNA, thereby preventing the pre-initiation complex from recognizing start codons. Notably, we uncovered the presence of 2′O-methylation sites in upstream non-canonical start codons within human 5'UTRs, where N_m inhibits upstream translation initiation in a position-dependent manner. Collectively, this study revealed a potent regulatory role of 2'O-methylation in translation initiation, providing novel insights into the mechanisms of non-canonical start codon selection in human transcriptomes.

Results

2'OH hydrogen bonds within translation initiation complexes

To understand what defines an optimal start codon, several groups solved cryo-EM structures of mammalian pre-initiation and initiation complexes.[8,9,21,22] These efforts provided a snapshot of the crucial intermolecular interactions surrounding Kozak sequences (Fig. 1a). Besides the recognition of start codons by the anticodon loop in $tRNA_i^{met}$, the interactions of positions A/G[-3] and G[+4] with translation initiation factors are required for translation initiation (Fig. 1a). In revisiting several mammalian translation initiation structures[8,9,21], we observed that 2'OH groups of mRNA nucleotides within the Kozak sequence appeared to be monitored by other components of the pre-initiation complex. For instance, t^6A in position 37 of $tRNA_i^{met}$ forms a hydrogen (H)-bond with the 2'OH group of the nucleotide preceding the start codon, C[-1] (Fig. 1b and Extended Data Fig. 1a). The 2'OH group of the first nucleotide, A[+1], in the start codon interacts with U1830 of 18S rRNA, either with the phosphate (PO_4^{3-}) in the backbone (Fig. 1c and Extended Data Figs. 1b-f), or the keto (C=O) group in position C2, especially in the late stages of the pre-initiation complex (Fig. 1d and Extended Data Figs. 1e-f). Two additional H-bonds were observed between the 2'OH group of G[+3] with lysine seven (K7) of the translation initiation factor 1A (eIF1A, Fig.

106 1e) and between the 2'OH group of nucleotide A[-3] in the mRNA with G961 of 18S rRNA (Fig.
107 1f). Together, these observations suggest that 2'OH groups in mRNA are monitored by 18S rRNA,
108 $tRNA_i^{met}$, and eIF1A, potentially playing a crucial role in start codon recognition.
109
110 **2'O-methylation of start codons inhibits translation**
111 To probe the significance of the H-bonds between 2'OH groups in mRNA with $tRNA_i^{met}$, 18S rRNA,
112 and eIF1A, we implemented a three-way splint ligation approach (Fig. 2a).[23] Briefly, a 5'UTR
113 lacking a start codon was transcribed and capped *in vitro* (Extended Data Figs. 2a,b). In parallel,
114 the ORF of NanoLuciferase (NanoLuc)—also lacking a start codon—was transcribed and
115 polyadenylated *in vitro* (Extended Data Fig. 2c). These pieces were ligated to a chemically
116 synthesized RNA fragment containing the Kozak sequence using a DNA splint that hybridizes to
117 the three independent RNA pieces (Fig. 2a and Extended Data Fig. 2d). The ligated NanoLuc
118 mRNAs with a Kozak sequence, the ligated mRNAs that lack a start codon, and the unligated
119 RNA pieces were transfected into HEK293T cells. Crucially, the ligated mRNAs containing the
120 Kozak sequence exhibited an ~100-fold increase in NanoLuc activity compared to the unligated
121 RNA pieces or the ligated mRNAs lacking a start codon (Fig. 2b).
122
123 To control for possible confounding artifacts, wherein a high proportion of unligated products
124 affect translation, we transfected increasing concentrations of full-length NanoLuc mRNA with
125 constant amounts of unligated RNA (Fig. 2c). Reciprocally, we transfected constant amounts of
126 full-length NanoLuc mRNA with increasing concentrations of unligated RNAs (Extended Data Fig.
127 2e). In either case, NanoLuc activity was unaffected by the unligated RNAs, exhibiting linear
128 behavior across a broad range of mRNA concentrations.
129
130 Leveraging the ability of 2'O-methylation to disrupt the donor capacity of 2'OH groups, we
131 systematically methylated the 2'OH groups of every nucleotide within the Kozak sequence using
132 the three-way splint ligation approach (Fig. 1d and Extended Data Fig. 2f). The ligation efficiency
133 was verified through gel electrophoresis and RT-qPCR, obtaining similar ligation efficiencies
134 across multiple replicates (Extended Data Fig. 2g). Equivalent amounts of ligated mRNAs were
135 transfected into HEK293T cells, and translation efficiency was measured six hours after
136 transfection through a luminescence assay. Remarkably, methylation of the 2'OH group in the
137 first position of the AUG codon completely abolished NanoLuc activity (Fig. 2e). Significant
138 luminescence reduction was also observed when methylation occurred on positions G[+3] and
139 G[+4] (Fig. 2e). In contrast, no significant effects were observed in positions A[-3], C[-2], C[-1],

140 and U[+2] (Fig. 2e). The complete inhibition of NanoLuc activity by A_m[+1] remained after
141 normalization of NanoLuc activity by transfected mRNA levels (Extended Data Fig. 2h).
142
143 To guarantee that the inhibitory effect of A_m[+1] on NanoLuc activity is transcript-specific and not
144 a global perturbation of protein synthesis, we co-transfected unmodified or A_m[+1]-modified
145 NanoLuc mRNAs with unmodified Firefly luciferase mRNAs. We observed that A_m[+1] abolished
146 NanoLuc activity without significantly altering Firefly luciferase activity (Fig. 2f). In addition,
147 evaluation of mRNA levels at different time intervals post-transfection showed no significant
148 differences in mRNA stability between unmodified and 2'O-methylated transcripts (Fig. 2g),
149 indicating that the effect of 2'O-methylation on protein output occurs via the inhibition of
150 translation.
151
152 Next, we evaluated the reproducibility of these results across different mRNAs and cell types.
153 First, we transfected the ligated NanoLuc mRNAs into HeLa cells. Like HEK293T cells, A_m[+1]
154 completely abolished NanoLuc activity in HeLa cells (Extended Data Fig. 2i). G_m[+3] and G_m[+4]
155 significantly reduced NanoLuc activity, albeit to a lesser extent than A_m[+1] (Extended Data Fig.
156 2i). Second, we generated green fluorescent protein (GFP) mRNA reporters, each one containing
157 2'O-methylation at a different position in the Kozak sequence. Ligation efficiency was generally
158 higher than 50% for GFP constructs (Extended Data Fig. 2j), which include a much larger 5'UTR
159 than the NanoLuc mRNA reporters (Extended Data Fig. 2k and Supplementary Table 1).
160 Following transfection of the GFP mRNA reporters into HeLa cells, protein expression was
161 determined by flow cytometry. Consistently, the GFP fluorescence results mirror those of
162 NanoLuc activity: A_m[+1] completely abolished GFP expression; G_m[+3] and G_m[+4] reduced
163 protein expression, whereas A_m[-3], C_m [-2], C_m [-1], and U_m [+2] had no effect on GFP expression
164 (Figs. 2h,i).
165
166 Collectively, these findings demonstrate that 2'O-methylation of Kozak sequences inhibits
167 translation in a position-specific manner. 2'O-methylation of position A[+1] has the most potent
168 inhibitory effect, suggesting that H-bonds between the 2'OH group of nucleotide A[+1] and U1830
169 in 18S rRNA are required for translation initiation.
170
171 **2'O-methylation inhibits translation from non-AUG start codons**
172 Over 99% of human mRNAs contain alternative AUGs or near-cognate codons in their 5'UTRs
173 (Fig. 3a and Extended Data Fig. 3a-c). We previously performed Ribo-seq in the presence of

174 harringtonine.[8] This translation inhibitor allows the mapping of translation initiation sites
175 transcriptome-wide (Extended Data Fig. 3d).[8] These experiments identified alternative translation
176 initiation sites in ~70% of cellular mRNAs (Extended Data Fig. 3e). Distribution of upstream start
177 codons shows a higher prevalence of CUG, AUG, and GUG, with evidence for upstream initiation
178 in all near-cognate start codons (Fig. 3b). Building on these observations, we sought to determine
179 whether 2'O-methylation affects translation initiation in the most prevalent near-cognate start
180 codons: CUG and GUG (Fig. 3c).
181
182 Like canonical AUG codons, CUG, GUG, and their methylated versions (C_mUG and G_mUG)
183 exhibited comparable ligation efficiencies (Extended Data Fig. 3f). Notably, the unmodified CUG
184 and GUG near-cognate codons induced a ~50-fold reduction in NanoLuc activity compared to the
185 unmodified AUG codon (Fig. 3d). Yet, 2'O-methylation at position N[+1] reduced the remaining
186 activity (Fig. 3d). To better assess the effect of 2'O-methylation in near-cognate initiation, we
187 increased the time of luminescence acquisition to enhance the signal-to-noise ratio. As expected,
188 2'O-methylation significantly reduced translation from near-cognate start codons (Fig. 3e and
189 Extended Data Fig. 3g). These findings confirm that AUG is the optimal start codon while near-
190 cognate codons such as CUG and GUG can also start translation, albeit inefficiently (Fig. 3f).
191 Significantly, 2'O-methylation at position N[+1] abolishes translation independently on whether
192 the start codon is a canonical AUG or a near-cognate start codon (Fig. 3f).
193
194 **2'O-methylation is present in near-cognate codons within 5'UTRs**
195 The finding that N_m[+1] in cognate and near-cognate start codons represses protein synthesis
196 raised the possibility that 2'O-methylation may be a regulatory mechanism of translation initiation
197 in cellular mRNAs. Using chemical biology methods and Nanopore direct RNA sequencing (DRS)
198 approaches, previous studies identified N_m as a prevalent mRNA modification.[17-20] While the
199 highest proportion of internal N_m sites was found in coding sequences (CDS) and 3'UTRs, 2'O-
200 methylation sites were also identified in 5'UTRs (Extended Data Fig. 4a). Notably, the Nanopore-
201 DRS approach yielded the highest number of N_m sites, including sites mapped to 5'UTRs
202 (Extended data Fig. 4a).[17]
203
204 To determine whether N_m is present in canonical start codons, we extracted the N_m genomic
205 coordinates from four different studies[17-20] and intersected them with the genomic coordinates of
206 canonical start codons. To increase confidence in the candidate N_m sites, we analyzed sites with
207 an estimated modification ratio of more than 10% per position. We found a near complete absence

208 of N_m in A[+1] (Fig. 4a). For instance, in HeLa cells, a cell line used in all N_m mapping studies,
209 only one site in the A_m[+1] position of canonical start codons was obtained (Extended Data Fig.
210 4b), despite thousands of N_m sites reported in coding sequences (Extended Data Fig. 4a). A total
211 of six A_m[+1] sites in canonical start codons were obtained in all studies (Fig. 4a), with two sites
212 being reproduced across different cell lines (Extended Data Table 1). These findings suggest a
213 negative selection constraint for 2'O-methylation in canonical start codons.
214
215 To determine whether N_m is present in upstream AUGs or near-cognate codons, we extracted the
216 N_m sites that mapped to 5'UTRs of any annotated Ensembl transcript and examined whether
217 these sites were located in position N[+1] of an upstream AUG or an upstream near-cognate
218 codon (Extended Data Table 2). This analysis identified >600 N_m[+1] sites across all cell lines
219 (Fig. 4b and Extended Data Fig. 4c). While CUG had the highest prevalence of 2'O-methylation
220 in position N[+1] among all near-cognates, upstream AUGs are the least prevalent codons with
221 only nine A_m[+1] sites identified (Fig. 4b). Notably, the number of C_mUG, C_mUG, A_mUC, and A_mCG
222 sites were significantly enriched compared to the distribution of these codons in the 5'UTRs of
223 protein-coding genes (Fig. 4c), whereas AUGs were significantly de-enriched (Fig. 4c), reinforcing
224 the notion that A_mUG is under a strong negative selection constraint.
225
226 Next, we biochemically validated the presence of N_m[+1] in endogenous 5'UTRs. We employed
227 Reverse Transcription at Low dNTPs (RTL-P),[24] a sensitive method that induces reverse
228 transcriptase stops (RT-STOP) at low concentrations of dNTPs (Fig. 4d), resulting in a reduction
229 in amplicon density after PCR amplification of the RT product. While there are more than eleven
230 human 2'O-methyltransferase enzymes (Extended Data Fig. 5a),[25] Fibrillarin (FBL) is consistently
231 identified as the predominant enzyme responsible for introducing N_m at internal sites in mRNA.[17-
232 20,26] Thus, we performed RTL-P in total RNA isolated from wildtype HeLa cells transfected with
233 two different siRNAs against Fibrillarin (FBL) or a scramble siRNA control (Fig. 4e). Candidate
234 sites were selected based on the highest predicted modification ratio from the Nanopore-DRS
235 approach in HeLa cells (Extended Data Table 3). Ten of the twelve candidate sites tested exhibited
236 a substantial loss in amplicon density under limited dNTP conditions (Figs. 4f-h and Extended
237 Data Figs. 5b-h). Importantly, the RTL-P signature was lost upon knockdown of FBL (Figs. 4f-h
238 and Extended Data Figs. 5b-h), demonstrating the presence of FBL-dependent 2'O-methylation
239 sites in human 5'UTRs. Notably, these findings suggest that 2'O-methylation is preferentially
240 located in upstream near-cognate codons with significant enrichment and higher prevalence at
241 CUG and GUG codons.

242

243 **2'O-methylation inhibits upstream translation initiation**

244 The presence of N_m in upstream AUGs and near-cognate codons raises the possibility that human
245 cells use 2'O-methylation to inhibit upstream translation initiation. To evaluate this possibility, we
246 re-analyzed a previously published HR-Ribo-seq[8] in HeLa cells to compare the density of initiator
247 ribosomes over start codons. As expected, ribosome density around canonical AUGs is higher
248 than upstream AUGs or near cognate codons (Fig. 5a). To determine the effect of N_m in upstream
249 translation initiation, we computed the ribosome density at upstream AUGs or near-cognate
250 codons relative to the ribosome density at the canonical AUG within the same transcripts. Sites
251 with an RPKM value ≥ 1 were binned based on the methylation status (unmethylated versus
252 N_m[+1]) and their codon identity. Remarkably, A_mUG codons exhibited a significant reduction in
253 ribosome density compared to non-methylated AUGs (Fig. 5b and Extended Data Table 4) despite
254 the low number of A_mUG sites. We next focused on the near-cognate codons CUG, GUG, and
255 AUC, as those codons showed a significant N_m enrichment in human 5'UTRs (Fig. 4c). The
256 relative ribosome density at methylated C_mUG and G_mUG was significantly decreased compared
257 to the unmethylated codons, reaching similar levels found in methylated A_mUG (Fig. 5b and
258 Extended Data Table 4). Although methylated AUC showed a similar tendency, the differences
259 between A_mUC and the unmethylated codons were not significant (Fig. 5b).

260

261 Next, we compared the ribosome density over a window of 29 nucleotides, the ribosome-
262 protected mRNA length, on 2'O-methylated codons at position N_m[+1] compared to 2'O-
263 methylation in positions N_m[-3], N_m[-2], N_m[-1], and N_m[+2]. We reasoned that this comparison
264 would recapitulate the mRNA reporter studies, which established that A_m[+1] completely abolished
265 translation, whereas positions N_m[-3], N_m[-2], N_m[-1], and N_m[+2] had no effect in translation
266 initiation (Figs. 2e,i). Consistent with the mRNA reporters, A_mUG, C_mUG, and G_mUG exhibited
267 reduced ribosome density compared to methylated codons in positions N_m[-3], N_m[-2], N_m[-1], and
268 N_m[+2] (Figs. 5c-e and Extended Data Figs. 6a-c).

269

270 Though we observed only a handful of 2'O-methylation sites in canonical AUG start codons, we
271 evaluated whether these methylation events inhibited the translation of the canonical ORF. For
272 this purpose, we assessed differences in translation efficiency (TE) obtained from traditional Ribo-
273 seq in HeLa cells.[27] While TE is decreased in A_mUG compared to unmethylated canonical AUG
274 codons, the differences were not statistically significant, probably due to a lack of statistical power
275 in the A_mUG group (n=5, Extended Data Fig. 6d).

276

277 Overall, these findings demonstrate that 2'O-methylation in the upstream start codons AUG, CUG,
278 and GUG inhibits upstream translation initiation in cellular mRNAs. These upstream codons with
279 N_m[+1] are located in genes predominantly involved in cellular components, biological processes,
280 and molecular functions related to ribosome biogenesis and mRNA translation (Extended Data
281 Fig. 7).

282

283 **2'O-methylation inhibits start codon recognition**

284 Once the start codon is recognized, a 48S pre-initiation complex hydrolyzes GTP, resulting in the
285 binding of the large 60S subunit to form the 80S initiator ribosome (Fig. 6a). The inhibitory effect
286 of 2'O-methylation in position A_m[+1] suggests that the H-bond between the 2'OH group of
287 nucleotide A[+1] and U1830 in 18S rRNA is required for translation initiation. There are several
288 possible explanations for this phenomenon. First, 2'O-methylation in position A[+1] stalls the 48S
289 pre-initiation before or after recognizing the start codon. Second, A_m[+1] precludes the 80S
290 initiation complex from proceeding into elongation. Third, 2'O-methylation in position A[+1] makes
291 the start codon non-optimal or unrecognizable by the pre-initiation complex, thus inducing leaky
292 scanning. To test these possibilities, we performed *in vitro* translation with reticulocyte lysates
293 using a Cy5-labeled mRNA reporter in the presence of non-hydrolyzable GTP (GMPPNP) or
294 harringtonine (HR). Pre-treatment with GMPPNP stalls the 48S pre-initiation complex after codon
295 recognition (Fig. 6b). Relatedly, HR stalls the 80S initiation complex before proceeding into
296 elongation (Fig. 6c).

297

298 After *in vitro* translation, reactions were loaded on a 5-30% sucrose density gradient followed by
299 ultracentrifugation, fractionation, and fluorescence (Cy5) recording. As expected, treatment with
300 GMPPNP increases the abundance of the 48S peak compared to vehicle control (Figs. 6d,e, blue
301 lines), whereas harringtonine treatment increases the 80S peak in the Cy5-labeled unmodified
302 mRNA reporter (Fig. 6d,f, blue line). In contrast, A_m[+1] decreases the formation of both the 48S
303 peak and 80S peaks in the presence of GMPPNP or HR, respectively, as well as in the untreated
304 sample (Figs. 6d-f, orange arrows).

305

306 Collectively, these findings indicate that codon recognition by the 48S pre-initiation complex and
307 subsequent formation of an 80S initiator ribosome did not occur when the start codon was 2'O-
308 methylated in position N[+1]. Thus, N_m[+1] leads to leaky scanning, and the 48S subunit can

309 continue scanning until it encounters another downstream start codon or, alternatively,
310 dissociates from the mRNA (Fig. 6g).
311
312 **Discussion**
313 We uncovered a potent regulatory function of 2'O-methylation in translation initiation and uORF
314 expression, providing new insights into the mechanisms of start codon selection in human cells.
315 Three significant findings support this mechanism. First, H-bonds between the 2'OH group of
316 nucleotide A[+1] in mRNA with U1830 in 18S rRNA are required for translation initiation. Second,
317 A_mUG, C_mUG, and G_mUG inhibit translation initiation at canonical and upstream start codons.
318 Third, 2'O-methylation of position N[+1] of the start codon is preferentially deposited at upstream
319 near-cognate codons (C_mUG and G_mUG), exhibiting negative selection for canonical and
320 upstream AUGs in human transcriptomes.
321
322 The movement of the 48S subunit along the 5′ UTR is characterized by a conformational
323 intermediate between the closed state, in which the $tRNA_i^{met}$ is fully inserted into the P site trying
324 to make an interaction with the mRNA codon, and the open state, in which the anticodon in
325 $tRNA_i^{met}$ is distant from the mRNA.[22,28] This conformational intermediate enables the initiator
326 $tRNA_i^{met}$ to translocate along the mRNA while transiently sampling codons for base-pair
327 complementarity. The purpose of this transient sampling is to recognize an optimal start codon
328 while rejecting non-cognate codons and, in most cases, near-cognate codons or AUGs under
329 poor sequence contexts.[29-32] Once the start codon is entirely recognized, a fully closed and
330 arrested 48S subunit signals a conformational change that allows the 60S subunit to join.[9,21,28,32,33]
331 Since GMPPNP arrests the 48S after the start codon is recognized but before the 60S subunit
332 joins, the lack of GMPPNP-induced 48S arrest in mRNAs carrying A_mUG suggests that 2'O-
333 methylation inhibits start codon recognition (Fig. 6f), potentially preventing $tRNA_i^{met}$ from fully
334 inserting into the P site and thereby precluding the 48S subunits from forming a fully closed
335 conformation. Consequently, 2'O-methylation makes an otherwise optimal translation initiation
336 site look like a non-cognate codon, and the 48S subunit continues scanning until it encounters a
337 downstream start codon or, alternatively, dissociates from the mRNA. In either case, the A_m[+1]
338 modification does not impair mRNA stability or globally affect protein synthesis (Fig. 2f,g).
339
340 The intermolecular interactions of mRNA with $tRNA_i^{met}$, 18S rRNA, and translation initiation
341 factors, as well as the orientation and pairing of the codon:anticodon, are essential for start codon
342 selection (Fig. 1a).[9,22,28,32] Chemically, the 2'-OH group in RNA can act as both a donor and

343 acceptor in H-bonds. Methylation replaces the hydrogen atom of the 2'-OH with a methyl group (-
344 CH3), effectively eliminating the donor capability of the 2'-OH group. Since the 2'OH group of the
345 A[+1] position acts as a donor in the H-bond with U1830, either with the phosphate (PO_4^{3-})
346 backbone or the keto (C=O) group in position C2, our findings suggest that this interaction is
347 required for proper start codon recognition. Notably, while the 2'OH of position A[+1] is near PO_4^{3-}
348 in U1830 along all steps of codon recognition, rearrangement, and 60S joining (Figs. 1c,d and
349 Extended Data Figs. 1b-f), the 2'OH of position A[+1] is only near the keto (C=O) group in the late
350 stages of the pre-initiation complex (Extended Data Figs. 1e-f). We speculate that 18S rRNA
351 monitors for proper pairing and orientation of the codon:anticodon interaction. Thus, blocking the
352 $A[+1]^{mRNA}$:$U1830^{18S}$ interaction precludes the ribosome from properly proofreading start codon
353 recognition. Alternatively, since 2'O-methylation is known for altering RNA structure,[34,35] a
354 structural shift induced by 2'O-methylation may alter the orientation of A[+1] in the mRNA, thereby
355 preventing it from properly pairing with the anticodon base. Future structural analyses would be
356 necessary to determine the exact mechanism by which A_m[+1] inhibits start codon recognition.
357
358 Notably, the effect of 2'O-methylation in translation initiation is site-specific. 2'O-methylation of
359 position N[-3], N[-2], N[-1], and N[+2] had no effect in protein output (Figs. 2e,i). An unexpected
360 finding was the lack of effect in position N[-1], as previous studies reported a significant role of
361 this $C[-1]^{mRNA}$:t^6A^{tRNA} interaction in stabilizing the initiation complex.[8,9] For instance, we previously
362 reported that replacing cytosine with ac^4C at position N[-1] weakens the $C[-1]^{mRNA}$:t^6A^{tRNA}
363 interaction, resulting in reduced translation initiation.[8] Additionally, studies in yeast reported that
364 inhibition of t^6A biosynthesis increases upstream non-AUG translation.[36] While our findings using
365 2'O-methylation could suggest that the $C[-1]^{mRNA}$:t^6A^{tRNA} interaction may not be required for
366 translation initiation, it is important to note that 2'O-methylation does not alter the acceptor
367 capacity of the C-O-CH3 group. Thus, the $C[-1]^{mRNA}$:t^6A^{tRNA} hydrogen bonds may be the result of
368 the carboxyl group of t^6A acting as a donor and the C-O-H group of C[-1] acting as an acceptor.
369 Under these circumstances, the incorporation of a methyl group at position 2'OH would not disrupt
370 the $C[-1]^{mRNA}$:t^6A^{tRNA} hydrogen bond.
371
372 2'O-methylation in positions N[+3] and N[+4] significantly reduced protein synthesis, albeit to a
373 lesser extent than N[+1] (Figs. 2e,i). While positions N[+3] and N[+4] are also involved in crucial
374 intermolecular interactions within the translation initiation complex (Figs. 1a,e,f), the effect in
375 protein synthesis may be mediated through impairing elongation rather than translation initiation.
376 For instance, 2'O-methylation is known to be a potent repressor of translation elongation by

377 precisely inhibiting the interaction of mRNA with the monitoring bases of rRNA during the
378 proofreading steps of translation elongation.[26,37]
379
380 It is well established that 2′-O-methylation is highly abundant in noncoding RNAs, including rRNA
381 and tRNA, and occurs in the 5′ cap of all mRNAs in human transcriptomes.[25] The presence of N_m
382 at internal mRNA sites is more controversial,[38] with some studies suggesting that N_m does not
383 exist internally in mRNA, while others report thousands of N_m sites in mRNA.[17-20,39,40] 2′O-
384 methylation confers several chemical properties that can be exploited for the site-specific mapping
385 of N_m, including resistance to alkaline hydrolysis (RiboMeth-seq),[40] periodate oxidation (N_m-seq),[20]
386 enzymatic cleavage (NJU-seq),[19] specific sensitivity to engineered reverse transcriptase (N_m-mut-
387 seq),[18] steric hindrance during reverse transcription (2OMe-seq),[39] and unique electrical
388 signatures in Nanopore-DRS.[17] Despite the variety of approaches that can detect N_m sites in
389 mRNA, there is little to no reproducibility between the studies.[38] The large inconsistencies across
390 multiple studies could be due to technical factors, as some deep sequencing technologies are
391 more sensitive than others but are also prone to artifacts. For example, in the initially published
392 N_m-seq method,[20] a mispriming artifact led to many false positives—although this artifact was later
393 corrected in a corrigendum. Alternatively, inconsistencies across studies may have a biological
394 basis if N_m deposition in mRNA is context-dependent. Nevertheless, low throughput validations of
395 N_m sites,[26] including ours (Figs. 4d-h and Extended Data Figs. b-h), have undoubtedly identified
396 the presence of N_m at internal mRNA sites.
397
398 Our findings that 2'O-methylation is significantly enriched at near-cognate codons, especially
399 CUG, suggest that 2'O-methylation is leveraged by cells to control non-AUG translation. Such a
400 mechanism could be beneficial under stress conditions when canonical cap-dependent
401 translation is repressed, allowing non-AUG translation to serve as a stress response mechanism.
402 In this manner, cells could block unwanted non-canonical initiation sites through 2'O-methylation
403 while leaving desired CUG codons unmethylated to facilitate their initiation. If so, deposition of
404 2'O-methylation in near-cognate codons might be induced under stress conditions. Future
405 research outside the scope of this study is necessary to test these hypotheses.
406
407 Our findings also have a significant impact on RNA therapeutics. Aberrant translation initiation
408 has been reported in numerous diseases, with hundreds of pathogenic genetic variants in 5'UTRs
409 linked to the introduction or destruction of uORFs.[5,41-44] These pathological conditions would
410 benefit from strategies that inhibit the aberrant initiation site while enhancing the expression of

411 the canonical ORF. Since 2'O-methylation effect on translation initiation is transcript and position-
412 specific, our findings open the door to investigate approaches capable of guiding 2'O-methylation
413 to specific positions within aberrant mRNAs,[45-47] for the targeted inhibition of pathogenic
414 translation initiation in human disease.
415
416 In sum, we identified a chemical modification of mRNA, 2'O-methylation, that unambiguously
417 inhibits start codon recognition in a position-specific manner. Remarkably, we found that human
418 cells use 2'O-methylation to inhibit non-canonical translation. These findings uncovered a potent
419 regulatory mechanism of translation initiation with potential applications in RNA therapeutics
420 aimed at targeting aberrant translation initiation in human diseases.
421
422 **Author contributions**
423 Conceptualization, D.A; Methodology: D.A.; Investigation, E.A.P., S.M., A.S., K.V, S.W. and Y.Y;
424 Formal Analysis, D.A., H.S., and Y.Y.; Data Curation, D.A. and H.S.; Writing – Original Draft, D.A.;
425 Writing – Review & Editing, D.A. and H.S.; Visualization, D.A. and H.S.; Supervision, D.A., Project
426 Administration, D.A.; Funding Acquisition, D.A.
427
428 **Acknowledgments**
429 This study was supported with funds from the U.S. Department of Defense grant HT9425-24-1-
430 0352 and the Kinship Foundation (SSP-2023-102) to D.A. Y.Y. was supported by the U.S.
431 Department of Defense grant HT9425-23-1-0661 and NIH P50CA180995 SPORE in Prostate
432 Cancer Career Enhancement Award. K.V. was supported by the National Institutes of Health
433 Training Grant (T32GM008449) through Northwestern University's Biotechnology Training
434 Program.
435
436 **Competing Interest Statement**
437 Nothing to disclose
438
439

440 **Materials and Methods**

441

442 **Cell culture**

443 HEK293T cells were cultured in Dulbeccos' Modified Eagle Medium (DMEM, Gibco, Cat. #:
444 11965092) supplemented with 10% Fetal bovine serum (Corning, Cat. #: 30-015-CV) and 1%
445 Penicillin/Streptomycin (Dot scientific, Cat. #: DSP93560-50). HeLa cells were cultured in DMEM
446 (Gibco, Cat. #: 11965092) supplemented with 10% bovine calf serum (Cytiva, Cat. #:
447 SH30073.04) and 1% Penicillin/Streptomycin.

448

449 **Synthesis and capping of 5' RNA fragments**

450 To produce the NanoLuc 5'UTR, an antisense DNA template containing a T7 promoter sequence
451 was purchased from Integrated DNA Technologies (IDT) and annealed to a sense T7 primer
452 (Supplementary Table 1). To produce the GFP 5'UTR, we generated the DNA template by PCR
453 using primers that annealed to a plasmid containing the coding sequence of micro red
454 fluorescence protein, making sure to exclude the canonical start codon from the amplicon product.
455 The forward primer contained the T7 sequence as an overhang. PCR was performed using the
456 Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Cat. #: M0494). Primers and
457 resulting DNA sequences are described in Supplementary Table 1.

458

459 DNA duplexes were *in vitro* transcribed using the HiScribe® T7 High Yield RNA Synthesis Kit
460 (New England Biolabs, Cat. #: E2040), following the manufacturer's suggestions. RNA was
461 purified using 1.8X SPRI beads (CleanNGS Cat. #: CNGS-0001). Purified RNA was capped *in*
462 *vitro* using the Vaccinia Capping System (New England Biolabs, Cat. #: M2080) and re-purified
463 using 1.8X SPRI beads. The integrity of RNA fragments was verified using polyacrylamide gels
464 prepared in a tris-borate-urea buffer (2 mL of 40% PAA [19:1], 5 g of Urea, 1mL of 10X TBE,
465 complete to 10 mL ddH2O). Electrophoresis was run in a 0.5X TBE running buffer and stained
466 with a 1X SYBR gold solution (ThermoFisher, Cat. #: S11494). Gels were visualized in a Sapphire
467 Biomolecular Imager (Azure Biosystems).

468

469 **Dot blots**

470 Dot blots were performed to verify the capping of *in vitro* transcribed RNAs. Briefly, 500 ng RNA
471 was denatured at 65°C for 5 min and loaded onto Hybond-N+ membranes. Membranes were
472 crosslinked with 120 mJ^2/cm^2 in the UV Crosslinker. Membranes were first stained in a 0.2%

473 methylene blue solution (0.2% methylene blue in 0.4M sodium acetate and 0.4M acetic acid),
474 then washed three times in DI water, and imaged in a Sapphire Biomolecular Imager.
475
476 Membranes were then blocked with 1% bovine serum albumin (BSA) in PBS containing 0.05%
477 Tween-20 (PBST) for 15 min at room temperature and probed overnight with mouse anti-m^7G
478 antibodies (1:1000, MBL International Corporation, Cat. #: RN016M) in 1% BSA at 4°C.
479 Membranes were next washed three times with 0.1% PBST, incubated with HRP-conjugated
480 secondary anti-mouse IgG (1:10000 dilution, Cell Signaling Technologies, Cat. # 7076S) in 1%
481 BSA at room temperature for 1 hr, washed three times with 0.1% PBST and developed with
482 enhanced chemiluminescent (ECL) western blotting substrate (Promega, Cat #: W1001) on a
483 Sapphire Biomolecular Imager.
484
485 **Synthesis and polyadenylation of 3' RNA fragments**
486 To produce the ORF and 3'UTR of NanoLuc, we generated the DNA template by PCR using
487 primers that annealed to a plasmid containing the coding sequence of NanoLuc, making sure to
488 exclude the canonical start codon from the amplicon product. The forward primer contained the
489 T7 sequence as an overhang. To produce the ORF and 3'UTR of GFP, we generated the DNA
490 template by PCR using primers that annealed to a plasmid containing the coding sequence of
491 GFP, making sure to exclude the canonical start codon from the amplicon product. The forward
492 primer contained the T7 sequence as an overhang. PCR was performed using the Q5 Hot Start
493 High-Fidelity 2X Master Mix (New England Biolabs, Cat. #: M0494). Primers and resulting DNA
494 sequences are described in Supplementary Table 1.
495
496 DNA duplexes were *in vitro* transcribed using the HiScribe® T7 High Yield RNA Synthesis Kit
497 (New England Biolabs, Cat. #: E2040), following the manufacturer's suggestions. RNA was
498 purified using 1X SPRI beads (CleanNGS Cat. #: CNGS-0001). RNA fragments (80 μg) were
499 dephosphorylated with 100 U of calf intestinal alkaline phosphatase (CIAP, Promega, Cat. #:
500 M2825), followed by mono-phosphorylation of the 5' end using 15 U of T4 Polynucleotide Kinase
501 (PNK, New England Biolabs, Cat. #: M0201L) in the presence of ATP. Phosphorylated RNA
502 products were purified using 1X SPRI beads. RNA products (10 μg) were polyadenylated using 5
503 U of *E. coli* Poly(A) Polymerase (New England Biolabs, Cat. #: M0276). The integrity of
504 polyadenylated RNA fragments was verified using TBU-PAGE. Electrophoresis was run in a 0.5X
505 TBE running buffer and stained with a 1X SYBR gold solution.

506

507 **Three-way splint ligations**

508 Phosphorylated RNA oligonucleotides (12 nt) containing the Kozak sequence with an AUG, CUG,
509 or GUG start codons were purchased from IDT. Likewise, the 2'O-methylated versions were
510 purchased from IDT (Supplementary Table 1). These RNA oligos were combined with 5' and 3' *in*
511 *vitro* transcribed RNA fragments at equimolar concentrations (15 pmol). An antisense DNA
512 oligonucleotide (15 pmol) was added to splint the different RNA pieces. Reactions were denatured
513 at 65 °C for 2 minutes and ramped down to room temperature for annealing.

514

515 Ligations were performed using 10 U of the T4 RNA ligase 2 (dsRNA ligase, New England
516 Biolabs, Cat. #: M0239) for 4 hr at 37 °C according to the manufacturer's suggestions. Reactions
517 were stopped with 2 U of Turbo DNase I (ThermoFisher Scientific, Cat. #: AM2238) for 30 min at
518 37°C and purified using the RNA Clean & Concentrator-5 kit (Zymo Research, Cat. #: R1014). To
519 verify the integrity of splint ligated NanoLuc RNA, 500 ng of ligated RNAs were separated by 4%
520 TBU-PAGE and stained with 1X SYBR gold (Invitrogen, Cat. #: S11494). To verify the integrity of
521 splint ligated GFP RNA, 500 ng of ligated RNAs were separated by 1% denaturing agarose gels
522 and stained with 1X SYBR gold (Invitrogen, Cat. #: S11494). Gels were visualized in a Sapphire
523 Biomolecular Imager (Azure Biosystems).

524

525 **NanoLuc and Firefly luciferase assays**

526 HEK293T or HeLa cells were seeded at a density of 30,000 cells in 100 μL complete DMEM
527 medium in 96-well plates. Cells were cultured for 24 hr before transfecting with 300 ng of capped
528 and polyadenylated splint ligated RNA or unligated RNA control. Transfection reactions were
529 prepared using the TransIT-mRNA Transfection Kit (Mirus, Cat. #: MIR2225) following the
530 manufacturer's instructions. Following the addition of the transfection mixtures, cells were
531 incubated for 6 hrs at 37°C. For luminescence detection, cells were assayed with Nano-Glo®
532 Luciferase Assay System (Promega, Cat. #: N1110) according to the manufacturer's instructions
533 for 96-well plates. For the dual transfection experiments, cells were assayed using the Nano-Glo
534 Dual-Luciferase Reporter Assay System (Promega, Cat. #: N1610) according to the
535 manufacturer's instructions for 96-well plates. Luminescence was read in a BioTek Synergy LX
536 Multimode Reader (Agilent Technologies).

537

538 **Flow cytometry**

539 HEK293T or HeLa cells were seeded at a density of 150,000 cells in 1 mL medium in 12-well
540 plates. Cells were transfected with 600 ng of capped and polyadenylated splint ligated mRNA or
541 unligated RNA control. Transfection reactions were prepared using the TransIT-mRNA
542 Transfection Kit (Mirus) following the manufacturer's instructions for 12-well plates. Following the
543 addition of the transfection mixtures, cells were incubated for 18 hrs at 37°C. Cells were
544 trypsinized and washed with PBS three times before they were resuspended in 500 μL PBS and
545 immediately analyzed via Flow cytometry in an LSR Fortessa 1 Analyzer or a BD FACSymphony
546 A5-Laser Analyzer. The percentage of GFP-positive cells was obtained using FlowJo/v.10.10.0.
547
548 **RNA isolation and RT-qPCR**
549 Total RNA was isolated from the transfection reaction at time 0 or transfected cells at different
550 time points post-transfection using the Trizol method. RT-qPCR was performed to evaluate the
551 efficiency of ligation before transfection or the stability of ligated mRNA at different time points
552 post-transfection. Briefly, the extracted RNA was reverse transcribed using SuperScript™ IV
553 Reverse Transcriptase (ThermoFisher Scientific, Cat. #: 18090010) using gene-specific RT
554 primers for NanoLuc, GFP, and 18S rRNA (Supplementary Table 1). qPCR was performed using
555 the 2X PowerUp SYBR Master Mix (ThermoFisher Scientific, Cat. #: A25742) and primers that
556 amplify the ligated mRNAs, total mRNAs (ligated plus unligated), and 18S rRNA (Supplementary
557 Table 1). To obtain the efficiency of ligation, we applied the formula 100 * 2^[-($Ct^{ligated}$/$Ct^{total.mRNA}$).
558 To measure mRNA stability, we applied the formula final.time(2^[-($Ct^{ligated}$/$Ct^{18S.rRNA}$)/
559 initial.time(2^[-($Ct^{ligated}$/$Ct^{18S.rRNA}$).
560
561 ***In vitro* translation and sucrose density gradients**
562 Cy5-labeled synthetic mRNAs were purchased from Genescript (Supplementary Table 1).
563 Reticulocyte lysates (37.5 μL) were pre-incubated with vehicle DMSO, 1 mM GMPPNP
564 (Guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate, Millipore Sigma, Cat. #: G0635), or
565 100 μg/ml Harringtonine (Abcam, Cat. #: ab141941) prior to adding 150 nM of Cy5-labeled
566 synthetic mRNAs in a total volume of 50 μL. Reactions were incubated for 10 min at 30 °C and
567 immediately stopped on ice. Reactions were loaded on 5-30% sucrose density gradient (50 mM
568 Tris-acetate [pH 7.0], 50 mM NH_4Cl, 12 mM $MgCl_2$, 1 mM DTT) poured using a BioComp Gradient
569 Master. Gradients were spun at 42,000 rpm in a Beckmann Coulter SW41 Ti rotor for 2:26 hr at
570 4°C. Fractionation and recording of Cy5 fluorescence were performed using a BioComp Density
571 Gradient Fractionation system.
572

573 **siRNA transfection**

574 The FBL siRNAs used in this study were purchased from Thermo Fisher (siFBL-1: s4820, siFBL-
575 2: s4821). siRNAs were transfected into HeLa cells using the Lipofectamine RNAiMAX
576 Transfection Reagent (ThermoFisher Scientific, Cat. #: 13778150) by following the manufacturer's
577 manual. Total RNA and cell lysates were collected three days post-transfection. The efficiency of
578 FBL knockdown was verified via Western blots using specific anti-FBL antibodies (Abcam, Cat. #:
579 ab5821) and anti-β-actin antibodies (Cell Signaling Technologies, Cat. #: 3700) as a loading
580 control.

581

582 **RTL-P assay**

583 To confirm the N_m occurrence along with its alteration upon FBL depletion, the RTL-P assay was
584 conducted as previously described.[17] First, total RNAs from control and FBL-knockdown HeLa
585 cells were isolated using RNeasy Plus Mini Kit (Qiagen). Reverse transcription (RT) was then
586 performed using specific reverse primers targeting mRNA sequences downstream of the 2'O-
587 methylation site in the presence of either a low (1 μM) or a high (1 mM) concentration of dNTPs.
588 The obtained cDNA was subjected to two separate amplification reactions using one pair of
589 primers targeting upstream and downstream of the methylation site and another pair of primers
590 that amplified a downstream region that would not be interrupted by the 2'O-methylation site
591 (Supplementary Table 1). The PCR mixture consisted of 10 μL AmfiSure PCR master mix
592 (GenDEPOT), 2 μL cDNA, and 1 mM PCR primers. The PCR products were then separated using
593 1% agarose gels and photographed by a Bio-Rad imaging system. DNA band signal intensities
594 were calculated using the ImageJ software. The methylation ratio of each group was determined
595 from the densities of the PCR bands obtained using high and low dNTP concentrations. The RT
596 and PCR primers designed for the RTL-P assay are listed in Supplementary Table 1.

597

598 **Identification of 2'OH hydrogen bonds in translation pre-initiation complexes**

599 The following structures were downloaded from the Protein Data Bank: 7UCJ (rabbit reticulocyte),
600 6YAL (rabbit reticulocyte), 8PJ2, 8PJ3, 8PJ4, 8PJ5, and 8PJ6 (reconstituted human 48S).
601 Structural visualization was performed in Pymol. Briefly, the background was set to white, and all
602 elements were initially hidden. Specific residues of the mRNA, centered around nucleotides N[-
603 3], N[-2], N[-1], N[+1], N[+2], N[+3], and N[+3] were selected and visualized as both ribbons (with
604 increased width) and sticks. Specific residues within $tRNA_i^{met}$, 18S rRNA, eIF1A, and eIF2α that
605 interact with mRNA were also selected. Ribbon elements were colored blue-white for mRNA,
606 green for tRNA, yellow-orange for 18S rRNA, and black for eIF1A or eIF2α. Atomic elements were

607 colored red (oxygen), blue (nitrogen), and orange (phosphorus). Lighting parameters were
608 adjusted to enhance visualization, and hydrogen bond calculations between mRNA, $tRNA_i^{met}$, 18S
609 rRNA, eIF1A, and eIF2α were included.
610
611 **Identification of N_m sites in AUG and near-cognate codons**
612 Maps of 2'O-methylation were extracted from four publications using either biochemical or
613 Nanopore DRS coupled with machine learning approaches. These methods are named N_m-seq,[20]
614 N_m-mut-seq,[18] NJU-seq,[19] and Nanopore-DRS[17]. Publicly available BED files from these studies
615 were retrieved from the GEO and supplementary tables (GSE90164, GSE174518, Supplemental
616 Tables 4 and 5 from Li et al. 2024,[17] and Supplemental Table 2 from Tang et al. 2024).[19] Notably,
617 for the N_m-seq study,[20] we downloaded the N_m sites reported in the corrigendum (GSE90164),
618 which eliminated the mispriming artifacts of the original publication. Three datasets (NJU-seq,
619 Nanopore-DRS, and N_m-mut-seq) reported the stoichiometry of N_m sites; therefore, we chose to
620 analyze N_m sites with a 2'O-methylation ratio above ten percent.
621
622 To identify N_m sites present in canonical start codons, we extracted the genomic information for
623 Ensembl canonical start codons from a human genome annotation file (GRCh38). Then,
624 bedtools/2.29.2 was used to identify N_m sites around canonical start codons. N_m sites were then
625 binned by their location relative to the canonical AUG; N_m[-3], N_m[-2], N_m[-1], N_m[+1], N_m[+2],
626 N_m[+3], and N_m[+4]. The number of sites was plotted based on the position, cell line, and study.
627
628 To identify N_m sites in upstream AUG and near-cognate codons, we selected N_m sites within
629 protein-coding genes that aligned to 5'UTR in any possible transcript. Then, we extracted the
630 genomic sequence flanking every N_m site. We filtered for sequences containing AUG and near-
631 cognate codons. Each N_m site location was derived relative to the AUG or near-cognate codons
632 (N_m[-3], N_m[-2], N_m[-1], N_m[+1], N_m[+2], N_m[+3], and N_m[+4]). Note that N_m sites can be binned into
633 multiple positions when located near more than one AUG or near-cognate codon. This analysis
634 was performed with custom python/3.9.16 scripts using biopython/1.85 scripts.
635
636 **Analysis of HR-Ribo-seq**
637 Raw sequencing files from previously published HR-Ribo-seq in HeLa cells were downloaded
638 from GSE162043.[8] Adapter sequences were removed from sequence files using cutadapt/4.2.
639 Reads mapping to tRNA and rRNA sequences were removed with bowtie2/2.4.5. The remaining

640 reads were mapped to the human genome (GRCh38) using hisat2/2.1.0. Alignments were
641 normalized by reads per kilobase per million mapped reads (RPKM) using deeptools/3.5.1.
642
643 To compare ribosome density between the canonical AUG, upstream AUGs, and upstream near-
644 cognate codons, we obtained a BED file containing the genomic coordinates of canonical start
645 codons from Ensembl/113 using biomart. A custom python script extracted RPKMs in a 30-
646 nucleotide window, 13 nucleotides downstream and 16 nucleotides upstream of the +1 position
647 in the canonical start codon. Similarly, we used a custom python script to generate a BED file
648 containing the genomic coordinates of AUGs and near-cognate codons within 5'UTRs. RPKMs
649 were extracted in the 30-nucleotide window surrounding upstream AUGs and near-cognate
650 codons. RPKMs of two replicates were averaged together. To remove sites containing extremely
651 low RPKM values, sites with at least one position where the RPKM is less than 1 within the 30-
652 nucleotide range were removed from the analysis. An R script was used to visualize the
653 differences in average RPKMs as a line plot over the range of 30 nucleotides around start codons
654 or methylated positions.
655
656 To analyze the relative ribosome density at upstream codons compared to the canonical start
657 codon in the presence and absence of N_m[+1], we extracted the RPKM at position +1 for upstream
658 near-cognate codons and canonical AUGs. RPKMs were averaged for two replicates, and an R
659 script was used to remove N_m sites with an RPKM less than 1. The ratio of the upstream RPKM
660 to the canonical RPKM in the presence and absence of upstream N_m[+1] was visualized as box
661 plots using the ggpubr and ggplot2 packages in R. A Wilcoxon test was performed for each
662 upstream N_m[+1] codon and the corresponding non-methylated codon. For genes containing
663 multiple upstream near-cognate codons, each ratio was plotted independently.
664
665 **Codon bias analysis**
666 Using a list of genes that were expressed in HeLa HR-Ribo-seq, 5'UTR sequences of the longest
667 transcript per gene were extracted from Ensembl/113 with biomart. This was also performed for
668 genes containing N_m[+1]. The frequencies of AUG and near-cognate codons were calculated for
669 Modified codons in the 5'UTRs and for the presence of these codons in the 5'UTRs of expressed
670 protein-coding genes. Codon frequencies were analyzed by performing a chi-square test of
671 proportion with a Benjamini-Hochberg correction.
672
673 **Gene ontology analysis**

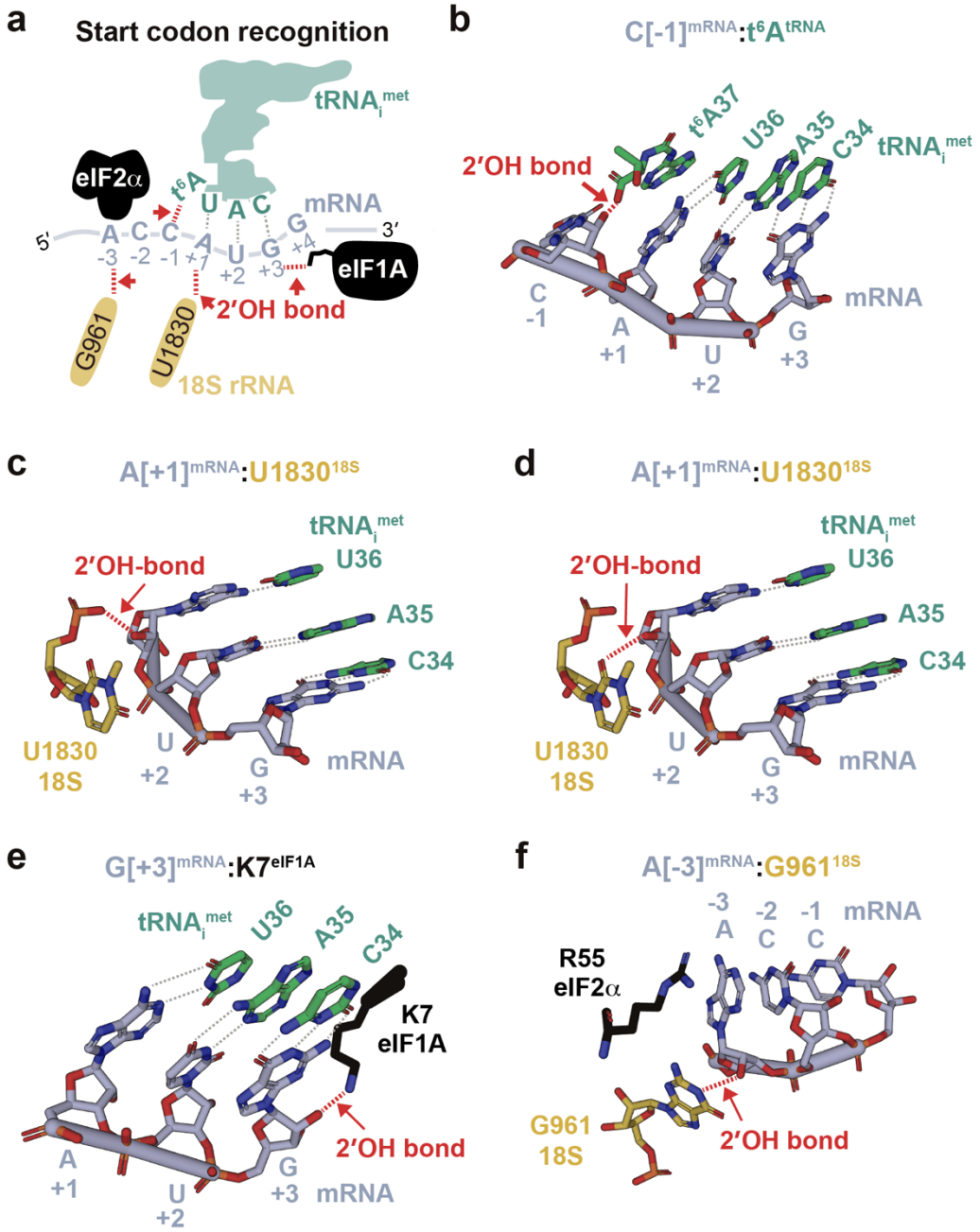
674 Genes containing A_mUG, C_mUG, and G_mUG in the 5'UTR were used for gene ontology (GO)
675 analysis for three ontologies: biological process, cellular component, and molecular function. An
676 R script was used to perform and visualize this analysis using the clusterProfiler/4.14.6 and
677 org.Hs.eg.db/3.20.0 packages. The top 5 categories for each ontology were displayed by
678 descending *p-value*.
679
680 **Analysis of translation efficiency**
681 To evaluate the impact of 2'O-methylation of cognate AUGs on the translation of the canonical
682 ORF, we downloaded the translation efficiency (TE) values of HeLa mRNAs reported in Arango
683 et al., 2018 (GSE102113).[27] mRNAs were binned based on the methylation status as AUG or
684 A_mUG and visualized as box plots in Rstudio/v4.4.0 using packages ggpubr and ggplot2.
685 Statistical significance between groups was assessed using a Wilcoxon rank-sum test.
686
687 **Other statistical analysis**
688 All other statistical tests and plotting were performed in Rstudio/v4.4.0. The specific statistical
689 tests, graph type, variables, number of observations per variable, and *p-values* are indicated in
690 the figures or their legends.
691
692 **Data availability**
693 This study did not generate original sequencing or structural data.
694
695 **References**
696 1. Brito Querido, J., Diaz-Lopez, I. & Ramakrishnan, V. The molecular basis of translation
697 initiation and its regulation in eukaryotes. *Nat Rev Mol Cell Biol* **25**, 168-186 (2024).
698 2. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that
699 modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-92 (1986).
700 3. Kearse, M.G. & Wilusz, J.E. Non-AUG translation: a new start for protein synthesis in
701 eukaryotes. *Genes Dev* **31**, 1717-1731 (2017).
702 4. Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation.
703 *Gene* **299**, 1-34 (2002).
704 5. Lee, D.S.M. et al. Disrupting upstream translation in mRNAs is associated with human
705 disease. *Nat Commun* **12**, 1515 (2021).

706 6. Calvo, S.E., Pagliarini, D.J. & Mootha, V.K. Upstream open reading frames cause
707 widespread reduction of protein expression and are polymorphic among humans. *Proc*
708 *Natl Acad Sci U S A* **106**, 7507-12 (2009).
709 7. Dasgupta, A. & Prensner, J.R. Upstream open reading frames: new players in the
710 landscape of cancer gene regulation. *NAR Cancer* **6**, zcae023 (2024).
711 8. Arango, D. et al. Direct epitranscriptomic regulation of mammalian translation initiation
712 through N4-acetylcytidine. *Mol Cell* **82**, 2797-2814 e11 (2022).
713 9. Simonetti, A., Guca, E., Bochler, A., Kuhn, L. & Hashem, Y. Structural Insights into the
714 Mammalian Late-Stage Initiation Complexes. *Cell Rep* **31**, 107497 (2020).
715 10. Galloway, A. & Cowling, V.H. mRNA cap regulation in mammalian cell function and fate.
716 *Biochim Biophys Acta Gene Regul Mech* **1862**, 270-279 (2019).
717 11. Mauer, J. et al. Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability.
718 *Nature* **541**, 371-375 (2017).
719 12. Akichika, S. et al. Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase
720 II-associated methyltransferase. *Science* **363**(2019).
721 13. Guca, E. et al. N(6)-methyladenosine in 5' UTR does not promote translation initiation.
722 *Mol Cell* **84**, 584-595 e6 (2024).
723 14. Zhou, J. et al. Dynamic m(6)A mRNA methylation directs translational control of heat
724 shock response. *Nature* **526**, 591-4 (2015).
725 15. Meyer, K.D. et al. 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999-
726 1010 (2015).
727 16. Beiki, H. et al. Detection of ac4C in human mRNA is preserved upon data reassessment.
728 *Mol Cell* **84**, 1611-1625 e3 (2024).
729 17. Li, Y. et al. 2'-O-methylation at internal sites on mRNA promotes mRNA stability. *Mol Cell*
730 **84**, 2320-2336 e6 (2024).
731 18. Chen, L. et al. Nm-Mut-seq: a base-resolution quantitative method for mapping
732 transcriptome-wide 2'-O-methylation. *Cell Res* **33**, 727-730 (2023).
733 19. Tang, Y. et al. An integrative platform for detection of RNA 2'-O-methylation reveals its
734 broad distribution on mRNA. *Cell Rep Methods* **4**, 100721 (2024).
735 20. Dai, Q. et al. Nm-seq maps 2'-O-methylation sites in human mRNA with base precision.
736 *Nat Methods* **14**, 695-698 (2017).
737 21. Petrychenko, V. et al. Structural basis for translational control by the human 48S
738 initiation complex. *Nat Struct Mol Biol* **32**, 62-72 (2025).

739 22. Brito Querido, J. et al. Structure of a human 48S translational initiation complex. *Science*
740 **369**, 1220-1227 (2020).
741 23. Hertler, J. et al. Synthesis of point-modified mRNA. *Nucleic Acids Res* **50**, e115 (2022).
742 24. Dong, Z.W. et al. RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in
743 RNA molecules. *Nucleic Acids Res* **40**, e157 (2012).
744 25. Ayadi, L., Galvanin, A., Pichot, F., Marchand, V. & Motorin, Y. RNA ribose methylation (2'-
745 O-methylation): Occurrence, biosynthesis and biological functions. *Biochim Biophys Acta*
746 *Gene Regul Mech* **1862**, 253-269 (2019).
747 26. Elliott, B.A. et al. Modification of messenger RNA by 2'-O-methylation regulates gene
748 expression in vivo. *Nat Commun* **10**, 3401 (2019).
749 27. Arango, D. et al. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*
750 **175**, 1872-1886 e24 (2018).
751 28. Llacer, J.L. et al. Conformational Differences between Open and Closed States of the
752 Eukaryotic Translation Initiation Complex. *Mol Cell* **59**, 399-412 (2015).
753 29. Zhou, F., Zhang, H., Kulkarni, S.D., Lorsch, J.R. & Hinnebusch, A.G. eIF1 discriminates
754 against suboptimal initiation sites to prevent excessive uORF translation genome-wide.
755 *RNA* **26**, 419-438 (2020).
756 30. Thakur, A. & Hinnebusch, A.G. eIF1 Loop 2 interactions with Met-tRNA(i) control the
757 accuracy of start codon selection by the scanning preinitiation complex. *Proc Natl Acad*
758 *Sci U S A* **115**, E4159-E4168 (2018).
759 31. Thakur, A., Marler, L. & Hinnebusch, A.G. A network of eIF2beta interactions with eIF1
760 and Met-tRNAi promotes accurate start codon selection by the translation preinitiation
761 complex. *Nucleic Acids Res* **47**, 2574-2593 (2019).
762 32. Hussain, T. et al. Structural changes enable start codon recognition by the eukaryotic
763 translation initiation complex. *Cell* **159**, 597-607 (2014).
764 33. Yi, S.H. et al. Conformational rearrangements upon start codon recognition in human
765 48S translation initiation complex. *Nucleic Acids Res* **50**, 5282-5298 (2022).
766 34. Abou Assi, H. et al. 2'-O-Methylation can increase the abundance and lifetime of
767 alternative RNA conformational states. *Nucleic Acids Res* **48**, 12365-12379 (2020).
768 35. Kawai, G. et al. Conformational rigidity of specific pyrimidine residues in tRNA arises
769 from posttranscriptional modifications that enhance steric interaction between the base
770 and the 2'-hydroxyl group. *Biochemistry* **31**, 1040-6 (1992).
771 36. Thiaville, P.C. et al. Global translational impacts of the loss of the tRNA modification
772 t(6)A in yeast. *Microb Cell* **3**, 29-45 (2016).

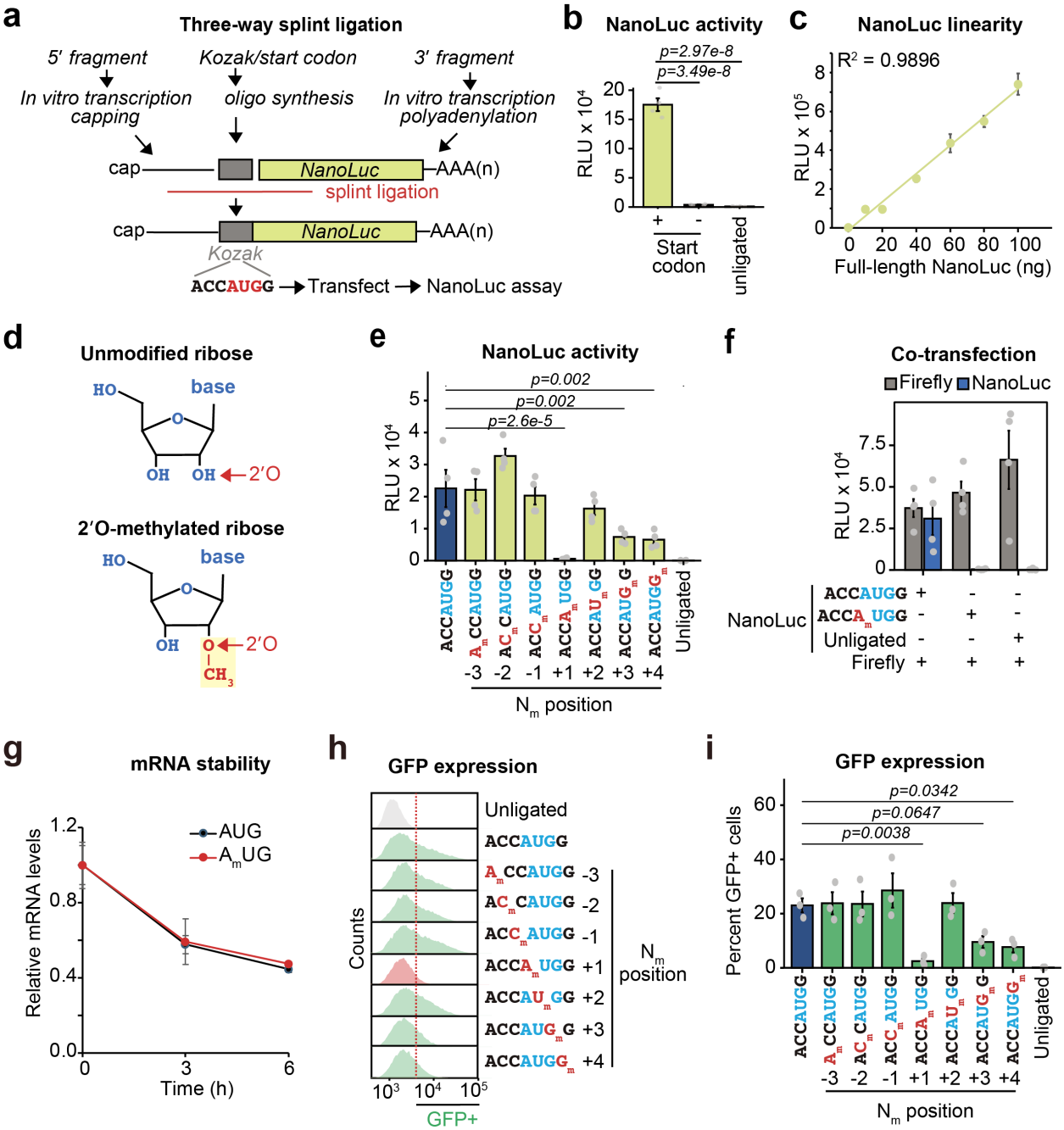
773 37. Choi, J. et al. 2'-O-methylation in mRNA disrupts tRNA decoding during translation
774 elongation. *Nat Struct Mol Biol* **25**, 208-216 (2018).
775 38. Zhou, K.I., Pecot, C.V. & Holley, C.L. 2'-O-methylation (Nm) in RNA: progress,
776 challenges, and future directions. *RNA* **30**, 570-582 (2024).
777 39. Incarnato, D. et al. High-throughput single-base resolution mapping of RNA 2´-O-
778 methylated residues. *Nucleic Acids Res* **45**, 1433-1441 (2017).
779 40. Marchand, V., Blanloeil-Oillo, F., Helm, M. & Motorin, Y. Illumina-based RiboMethSeq
780 approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res* **44**, e135 (2016).
781 41. Filatova, A. et al. Annotation of uORFs in the OMIM genes allows to reveal pathogenic
782 variants in 5'UTRs. *Nucleic Acids Res* **51**, 1229-1244 (2023).
783 42. Soukarieh, O. et al. Common and Rare 5'UTR Variants Altering Upstream Open Reading
784 Frames in Cardiovascular Genomics. *Front Cardiovasc Med* **9**, 841032 (2022).
785 43. Whiffin, N. et al. Characterising the loss-of-function impact of 5' untranslated region
786 variants in 15,708 individuals. *Nat Commun* **11**, 2523 (2020).
787 44. Martin-Geary, A.C. et al. Systematic identification of disease-causing promoter and
788 untranslated region variants in 8040 undiagnosed individuals with rare disease. *Genome*
789 *Med* **17**, 40 (2025).
790 45. Zhao, J. et al. Generating snoRNA-guided Programmable 2'- O -methylation. *bioRxiv*
791 (2024).
792 46. Huang, C. & Yu, Y.T. Targeted 2'-O methylation at a nucleotide within the pseudoknot of
793 telomerase RNA reduces telomerase activity in vivo. *Mol Cell Biol* **30**, 4368-78 (2010).
794 47. Zhao, X. & Yu, Y.T. Targeted pre-mRNA modification for gene silencing and regulation.
795 *Nat Methods* **5**, 95-100 (2008).
796
797

Figure 1. 2'OH hydrogen bonds within translation initiation complexes.



a, schematic representation of the intermolecular interactions within translation pre-initiation complexes. **b-d,** Cryo-EM visualization of rabbit 80S initiator ribosomes captured in complexed with harringtonine (PDB:7UCJ, Arango et al., 2022)[8] and centered around the interaction of the 2'OH group in nucleotide C[-1]^{mRNA} with t^6A in $tRNA_i^{met}$ (b); centered around the interaction of the 2'OH group in nucleotide A[+1]^{mRNA} with the PO_4^{3-} group of U1830 in 18S rRNA (c); or centered around the interaction of the 2'OH group in nucleotide A[+1]^{mRNA} with the C[2]=O group of U1830 in 18S rRNA. **e-f,** cryo-EM visualization of human 48S pre-initiation complexes (PDB:8PJ3 Petrychenko et al., 2025)[21] centered around the interaction of the 2'OH group in nucleotide G[+3]^{mRNA} with lysin 7 in eIF1A; or centered around the interaction of the 2'OH group in nucleotide A[-3]^{mRNA} with G961 in 18S rRNA. All structures were downloaded from PDB and visualized using Pymol.

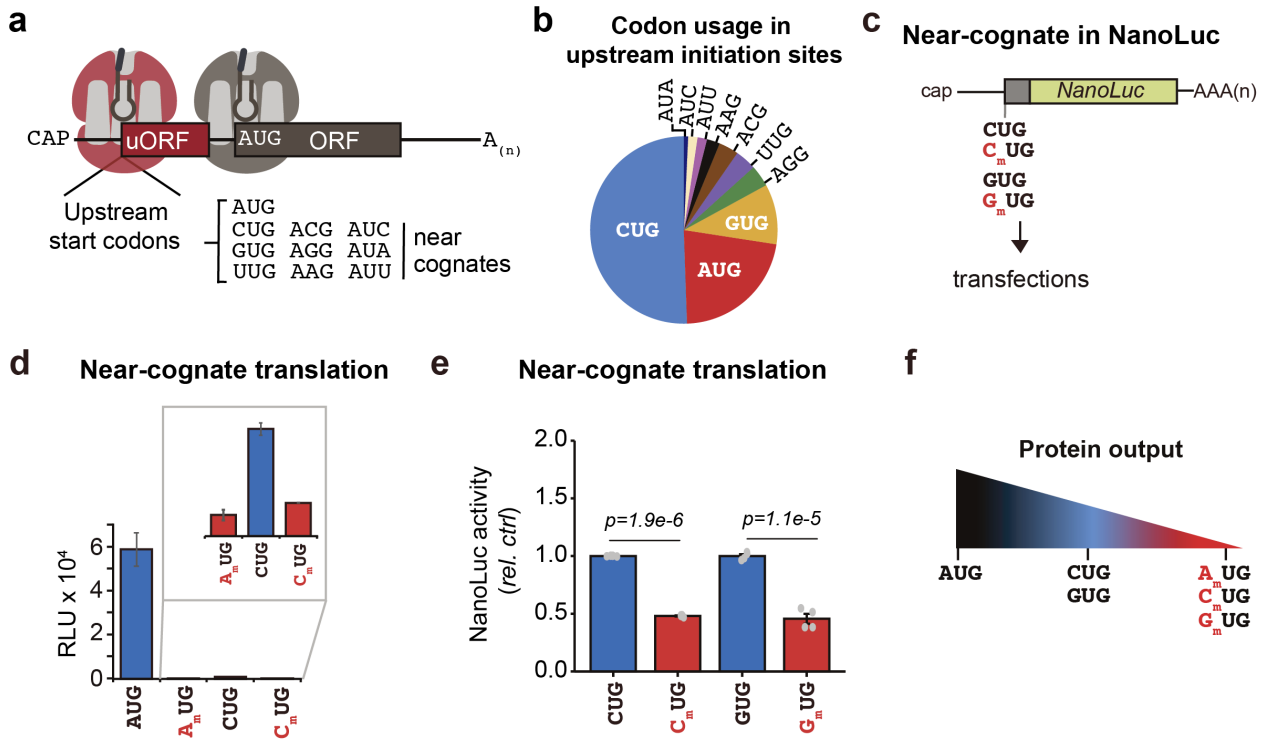
Figure 2. 2'O-methylation of start codons inhibits protein synthesis.



a, experimental design of a three-way splint ligation approach to insert 2'O-methylation in single positions within mRNA start codons. **b,** relative light units (RLU) were measured in cell lysates from HEK293T cells transfected with ligated NanoLuc mRNAs, with or without a start codon, or transfected with unligated mRNAs as control. Mean ± SEM, n = 3. p = One-way ANOVA with Tukey's post hoc analysis. **c,** relative light units were measured in lysates from HEK293T cells co-transfected with increasing amounts of full-length NanoLuc mRNA reporters and constant amounts (50 ng) of unligated mRNA. Mean ± SEM, n = 3. A linear regression demonstrates the linearity of the assay. **d,** chemical structures of unmethylated and 2'O-methylated nucleosides. **e,** relative light units were measured in lysates from HEK293T cells transfected with ligated NanoLuc

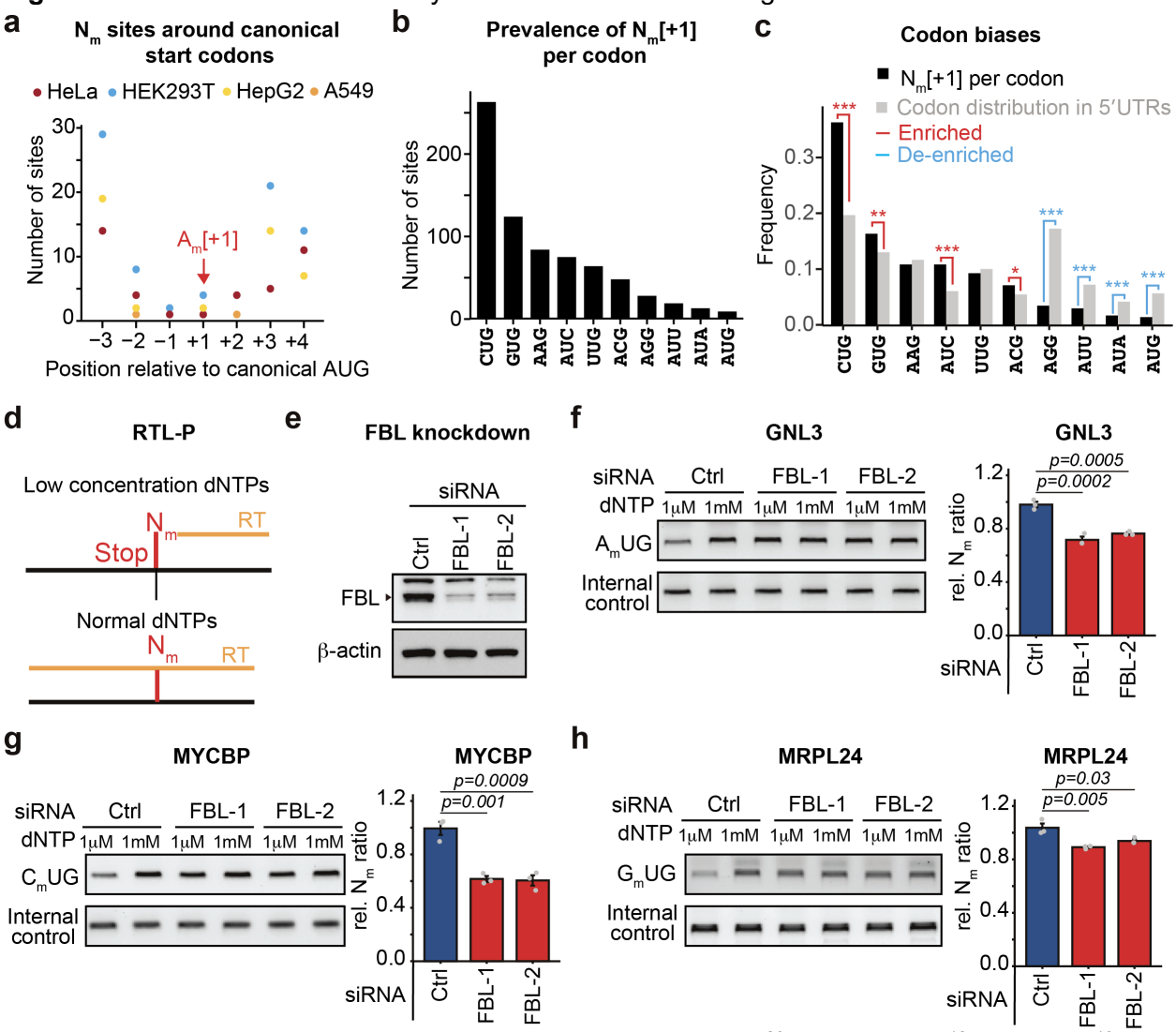
821 mRNAs carrying 2'O-methylation in different positions around the start codon. Unmethylated
822 (blue) and unligated mRNAs were used as controls. Mean ± SEM, n = 4. p = One-way ANOVA
823 with Tukey's post hoc analysis. **f,** relative light units were measured in lysates from HEK293T
824 cells co-transfected with unmodified mRNAs encoding Firefly luciferase and unmodified or A_m[+1]-
825 modified mRNAs encoding NanoLuc. Mean ± SEM, n = 4. **g,** NanoLuc mRNA levels were
826 measured using RT-qPCR at different time intervals post-transfection. Primers were designed to
827 ensure that only ligated mRNA is detected. 18S rRNA was used as a normalizing control. Mean
828 ± SEM, n = 3. **h,** representative histograms of fluorescence intensity measured by flow cytometry
829 in HeLa cells transfected with ligated GFP mRNA reporters carrying 2'O-methylation in different
830 positions around the start codon. Unmethylated (blue) and unligated mRNAs were used as
831 controls. **i,** quantification of data from (h). Mean ± SEM, n = 3. p = One-way ANOVA with Tukey's
832 post hoc analysis.

833 **Figure 3.** Inhibition of non-AUG translation by 2'O-methylation of near-cognate start codons.



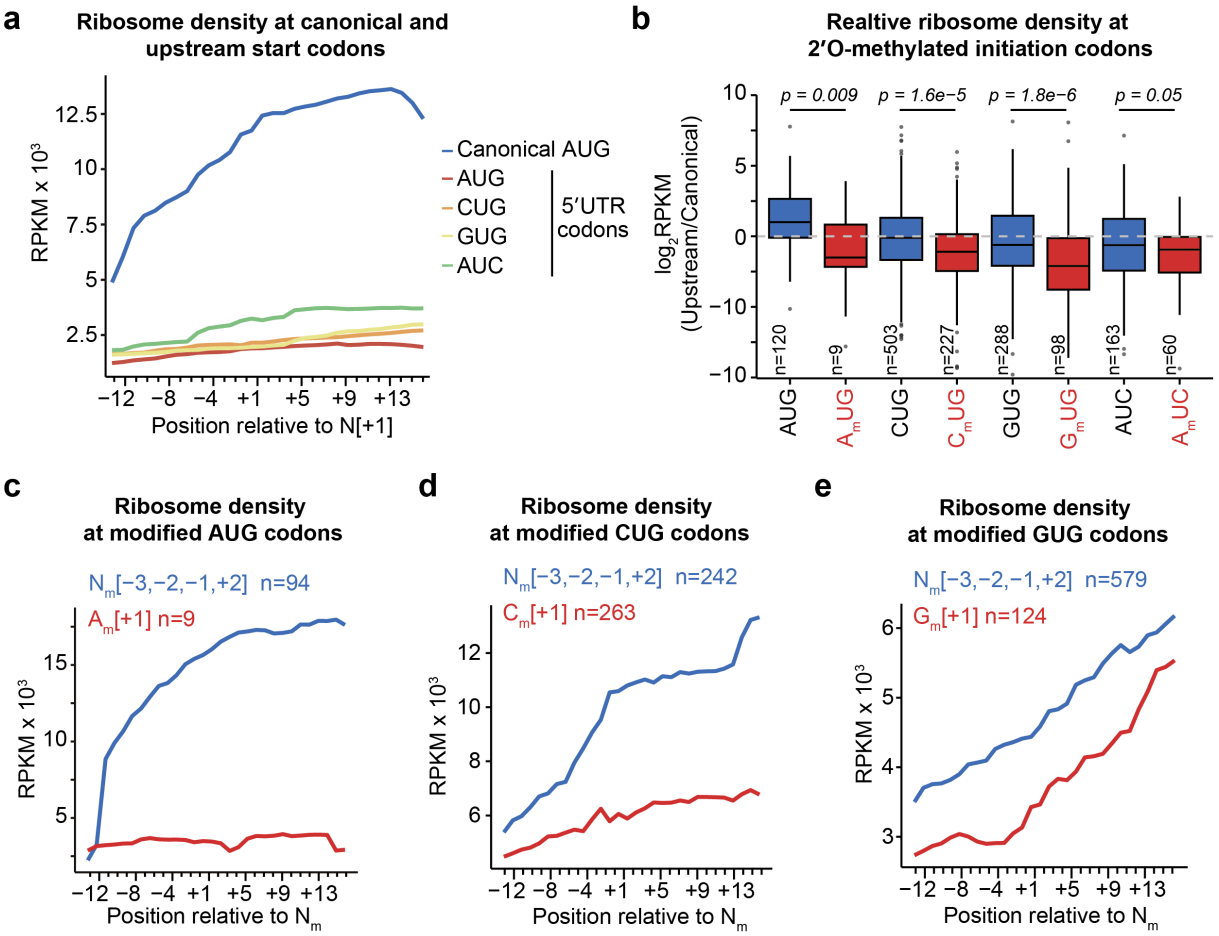
834
835 **a,** schematic representation of canonical and upstream translation initiation. **b,** distribution of
836 upstream initiation codons obtained from HR-Ribo-seq (Arango et al., 2022). 8 c, three-way splint
837 ligation approach to insert 2'O-methylation a position N[+1] of CUG and GUG start codons. **d,**
838 relative light units were measured in lysates from HeLa cells transfected with unmodified, C_mUG,
839 or G_mUG-modified NanoLuc mRNA reporters. Mean ± SEM, n = 2. **e,** normalized RLU relative to
840 unmodified mRNA. Mean ± SEM, n = 4. p = two-tailed *Student's t-test*. **f,** schematic representation
841 of translation initiation efficiency by AUG, CUG, GUG, and their respective 2'O-methylated
842 versions.
843
844
845

846 **Figure 4.** Identification of 2'O-methylation in AUG and near-cognate codons.



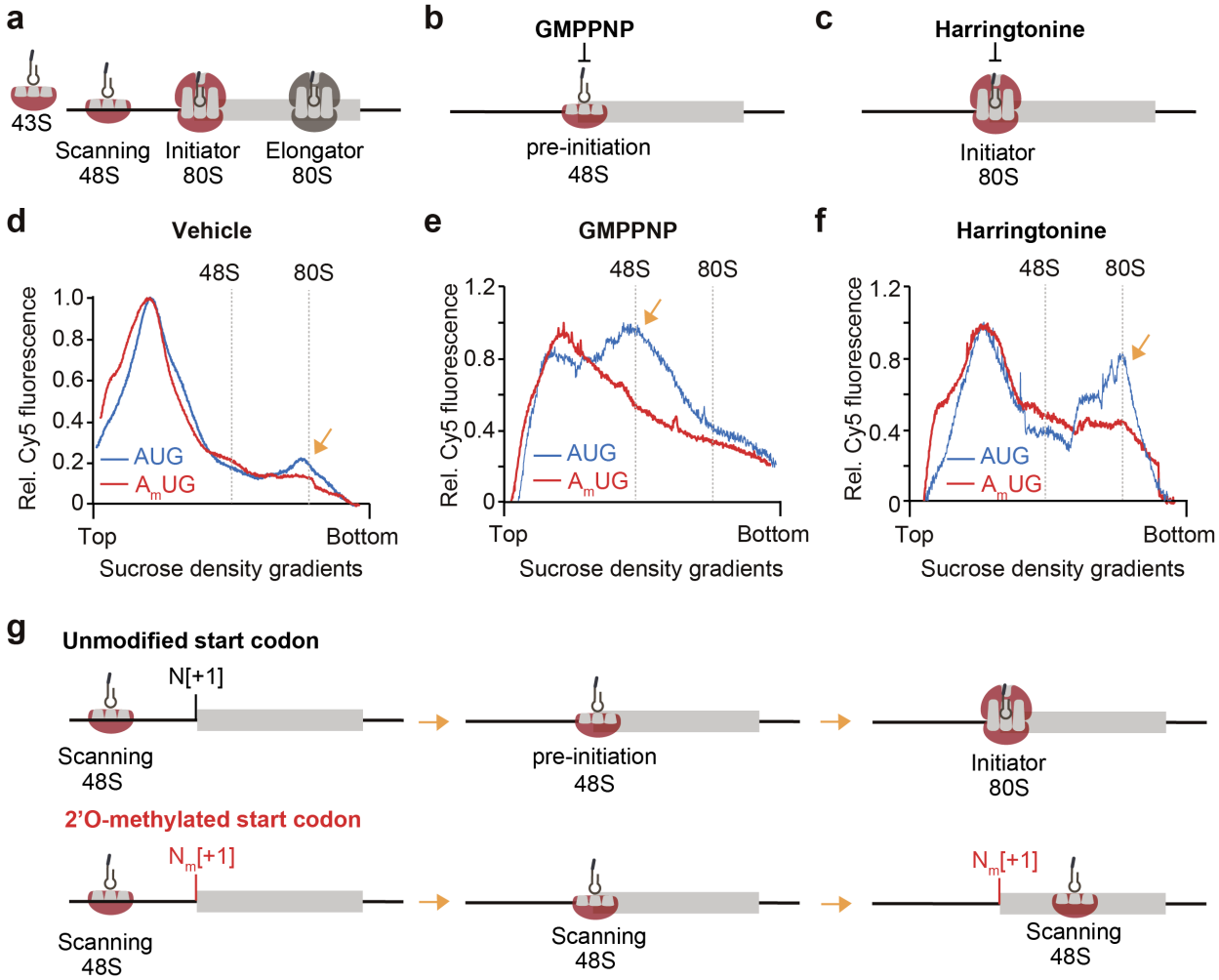
847 **a,** 2'O-methylation maps from four different studies, N_m-seq,[20] N_m-mut-seq,[18] NJU-seq,[19] and
848 Nanopore-DRS[17] were compiled and intersected with the coordinates for canonical AUG start
849 codons in human transcriptomes. Data from four different cell lines is shown in the dot plots. **b,**
850 sites with 2'O-methylation in position N[+1] of upstream AUG, or any near-cognate codon in
851 5'UTRs**.** Data from all different studies and cell lines described in (a) are included in the bar plots.
852 **c,** the fraction of each N_m[+1]-methylated codon was compared to the distribution of AUG and all
853 near cognate codons in the 5'UTR of human mRNAs. p = chi-squared test for proportions. * $p <$
854 0.01, ** $p < 0.001$, *** $p < 0.0001$. **d,** schematic of RTL-P. Reverse transcription was performed
855 in the presence of an optimal dNTP concentration (1 mM) or suboptimal dNTP concentration
856 (1 μM) that induces a premature RT-STOP. A pair of primers surrounding the 2'O-methylation is
857 used during PCR amplification. In this manner, the amplicon signal is an indicator of RT-STOP
858 efficiency and N_m stoichiometry. Another pair of primers, located downstream of the 2'O-
859 methylation site, serves as an internal control for mRNA levels. **e,** Western blot validation of FBL
860 knockdown using two different siRNAs in HeLa cells. **e-h,** *left:* representative agarose gels of
861 amplicon density for the 2'O-methylated site (top) and the loading control (bottom). *Right:*
862 quantification of the relative N_m ratio dependent on FBL. Mean ± SEM, n = 3. p = One-way ANOVA
863 with Tukey's post hoc analysis.

Figure 5. 2'O-methylation inhibits upstream translation initiation.



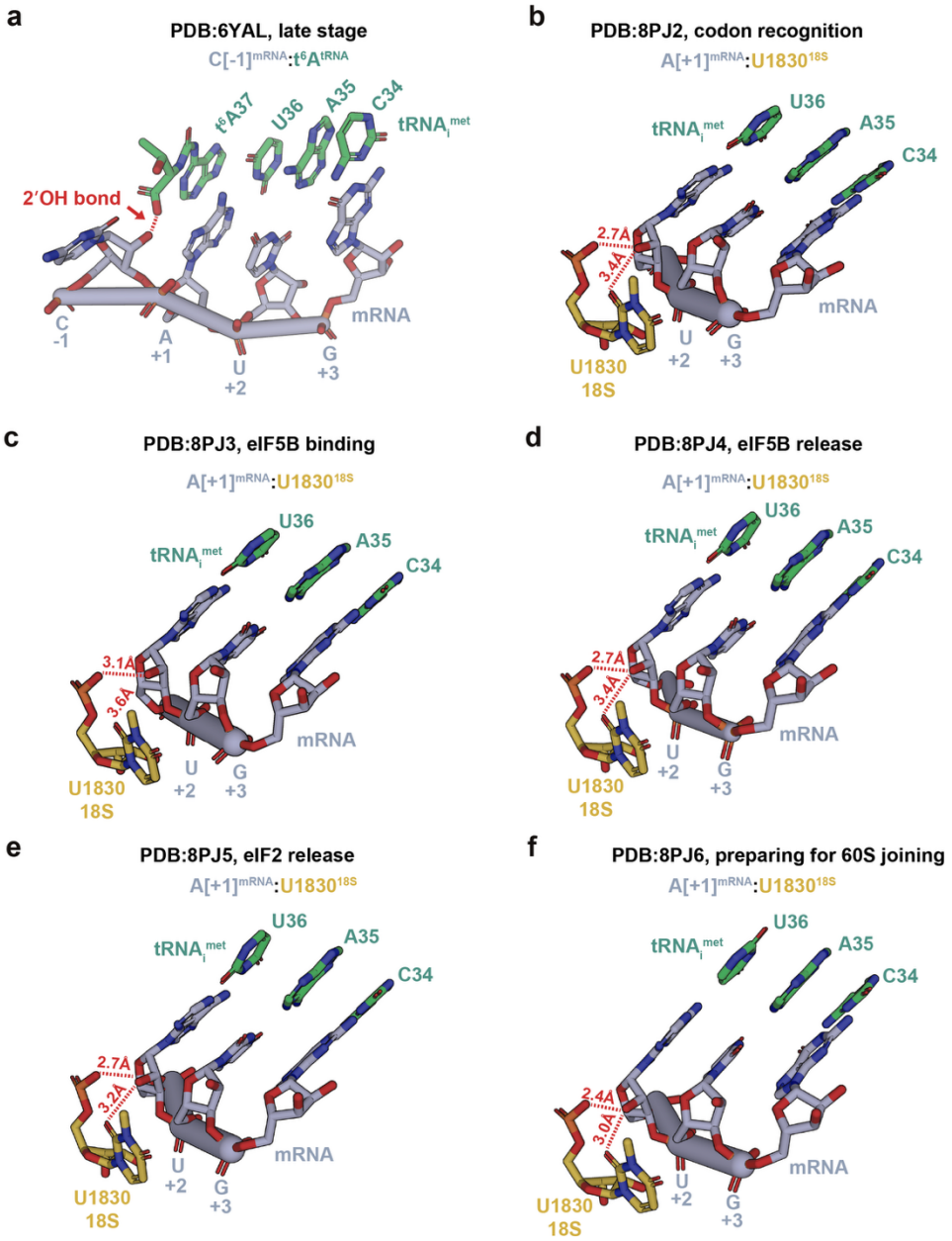
a, averaged HR-Ribo-seq density (RPKMs) over a window of -13 to +16 nucleotides around position N[+1] of canonical AUG or upstream AUG, CUG, GUG, and AUC. **b,** Relative HR-Ribo-seq density was obtained through the $\log_2$ of the fold change between the RPKMs at the unmodified (blue) or N_m[+1]-modified (red) upstream codon relative to the RPKMs at the downstream canonical AUG within the same transcript. Only sites with RPKM values ≥1 were included. Box plots summarize the median across all sites. p = two-sided Wilcoxon rank sum test. **c-d**, averaged HR-Ribo-seq density (RPKMs) over a window of -13 to +16 nucleotides around 2'O-methylated codons at position N_m[+1] (red) or 2'O-methylated codons at positions N_m[-3, -2, -1, and +2] (blue). Only sites with RPKM values ≥1 across all positions were included.

875 **Figure 6.** 2'O-methylation inhibits start codon recognition.



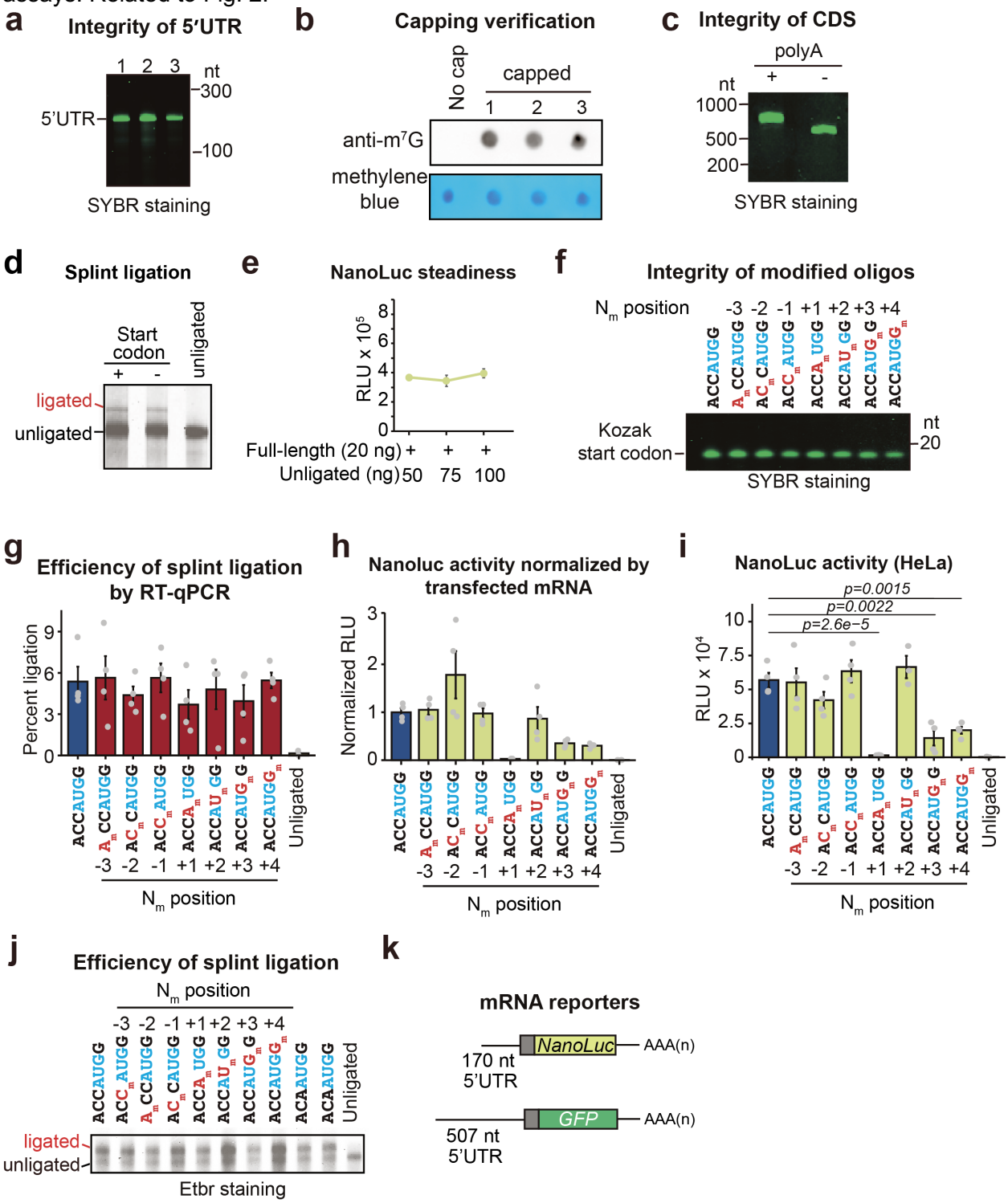
876
877 **a-c,** schematic representation of translation inhibition (a), mechanisms of 48S inhibition by
878 GMPPNP (b), and mechanisms of initiator 80S inhibition by HR (c). **d-f,** Cy5-labeled synthetic
879 mRNAs containing unmodified or A_mUG-modified start codons were incubated with untreated
880 reticulocyte lysates (c), treated with GMPPNP (d), or treated with HR (e). Reactions were
881 fractionated through a sucrose density gradient, and fluorescence (Cy5) intensity was recorded
882 concomitantly with fractionation. **g,** model: 2'O-methylation inhibits start codon recognition by
883 making an otherwise optimal translation initiation site look like a non-cognate codon.
884 Consequently, the 48S subunit continues scanning until encountering a downstream start codon
885 or, alternatively, dissociates from the mRNA.
886
887
888

Extended Data Figure 1. H-bonds between the 2'OH groups of mRNA nucleotides with $tRNA_i^{met}$ and 18S rRNA. Related to Fig. 1.



a, Cryo-EM visualization of late-stage rabbit 48S pre-initiation complexes (PDB:6YAL, Simonetti et al., 2022)[9] centered around the interaction of the 2'OH group in nucleotide $C[-1]^{mRNA}$ with t^6A in $tRNA_i^{met}$. This interaction was reproduced in rabbit 80S initiator ribosomes in complex with Harringtonine (Fig, 1B). **b-f,** Cryo-EM visualization of 48S pre-initiation complexes (Petrychenko et al., 2025)[21] centered around the interaction of the 2'OH group in nucleotide $A[+1]^{mRNA}$ with U1830 in 18S rRNA across different stages of 48S pre-initiation complexes, including the stage of codon recognition and GTP hydrolysis (b, PDB: 8PJ2), binding of eIF5B (c, PDB: 8PJ3), followed by sequential release of eIF5B (d, PDB: 8PJ4) and eIF2 (e, PDB: 8PJ5), and the late-stage readiness for 60S subunit joining (f, PDB: 8PJ6). All structures were downloaded from PDB and visualized using Pymol.

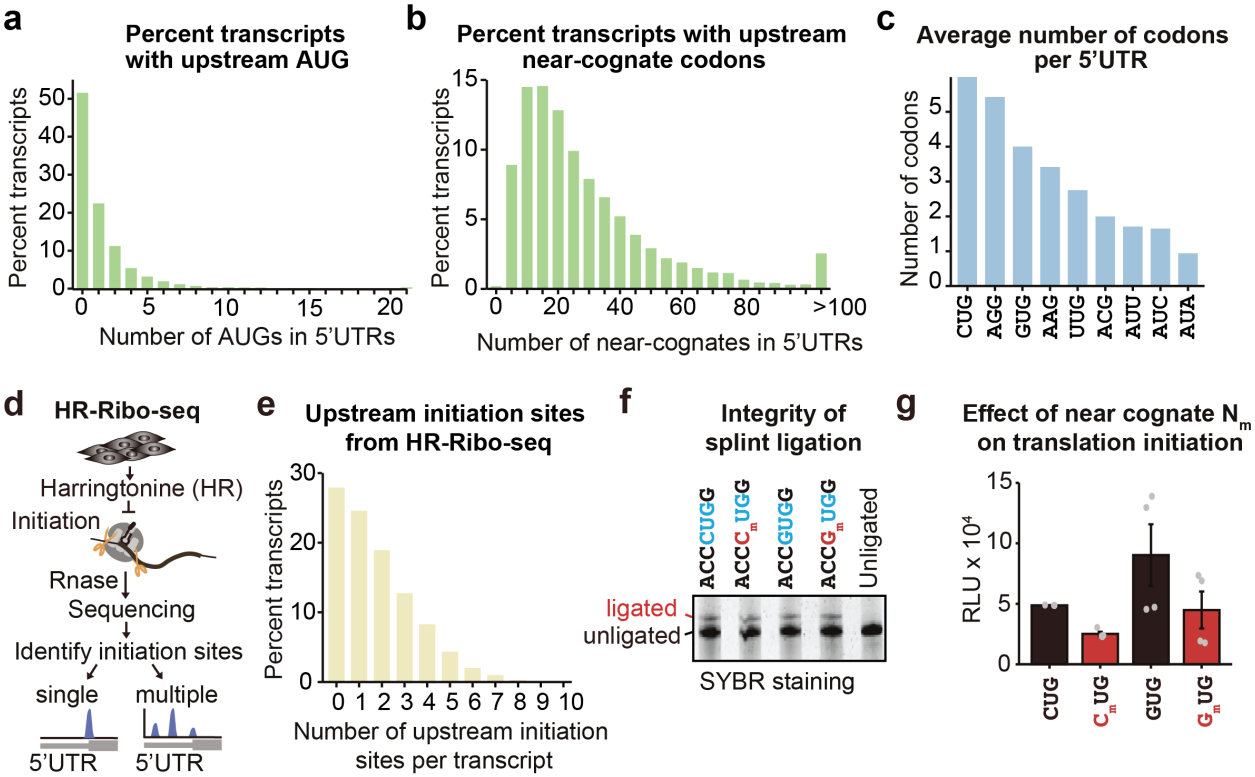
Extended Data Figure 2. Quality controls for the three-way splint ligation and translation assays. Related to Fig. 2.



a, the 5' end of the mRNA reporters, which lacked a start codon, was transcribed and capped *in vitro*. The integrity of the RNA was verified by polyacrylamide gel electrophoresis (PAGE) and SYBR gold staining. **b,** successful capping of the *in vitro* transcribed 5' ends was verified by dot blot using anti-m^{7}G antibodies. Methylene blue staining was used as a loading control. **c,** the 3'

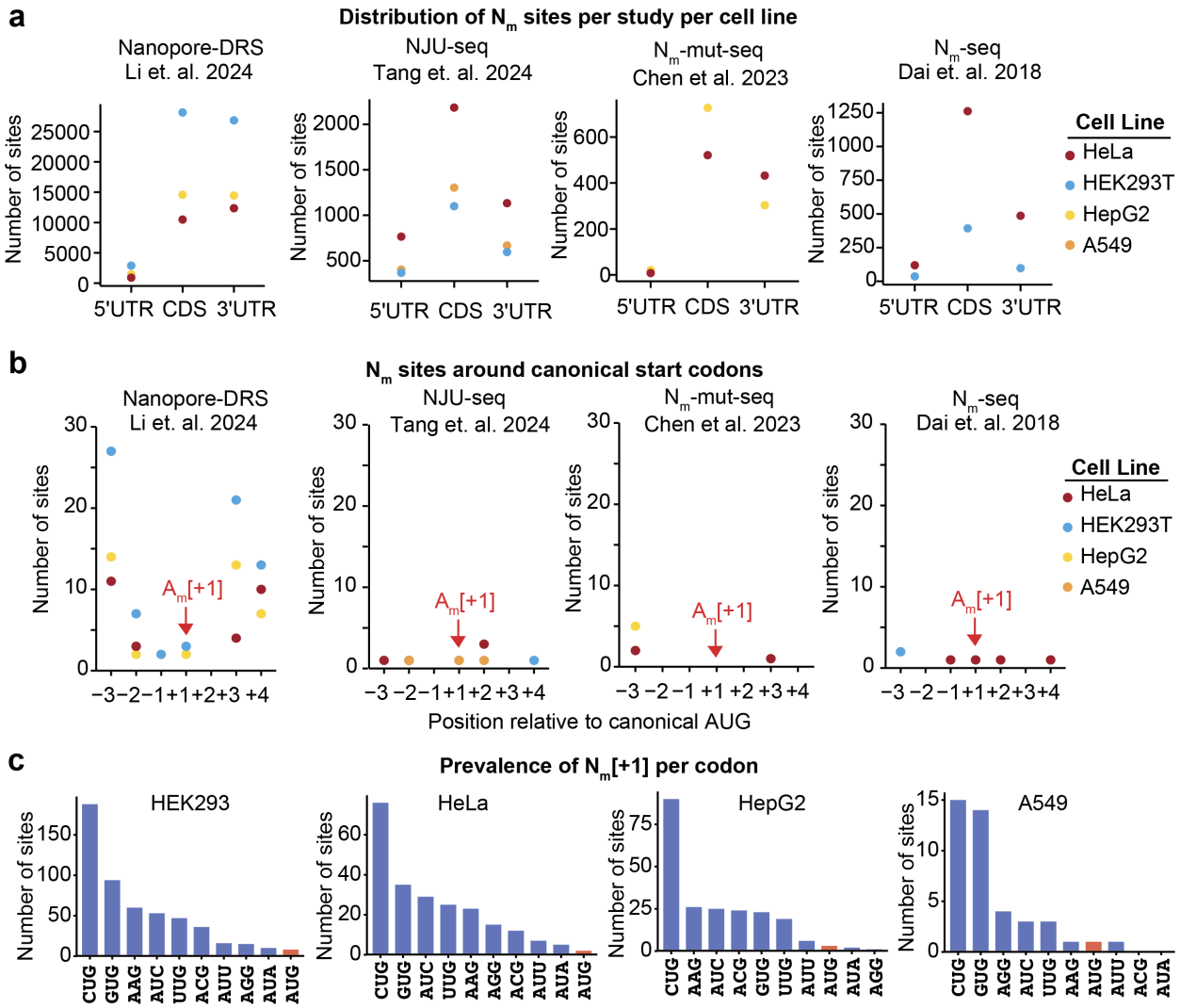
end of the mRNA reporters, containing the ORF and the 3'UTR but lacking a start codon, was transcribed and polyadenylated *in vitro*. The integrity of the RNA and the efficiency of polyadenylation were verified using PAGE and SYBR gold staining. **d,** the ligation efficiency of NanoLuc mRNA reporters was verified using PAGE and SYBR gold staining. Ligated mRNA shows a shift in molecular weight compared to unligated mRNA. **e,** relative light units were measured in lysates from HEK293T cells co-transfected with increasing amounts of unligated RNA pieces and constant amounts (20 ng) of full-length NanoLuc mRNA. Mean ± SEM, n = 3. The line plot shows stable NanoLuc activity across different amounts of unligated RNA. **f,** Integrity of the synthetic 5' phosphorylated RNA oligos containing unmodified or 2'O-methylation in different positions around the start codon. **g,** the efficiency of ligation was verified by RT-qPCR. Two sets of primers were used (Supplementary Table 1). The first set specifically recognizes the ligated mRNA, while the second set binds NanoLuc ORF to obtain total levels of mRNA. The efficiency of ligation was calculated as: (ligated mRNA)/(total RNA)*100. **h,** NanoLuc activity normalized by transfected mRNAs, which was obtained using RT-qPCR from the RNA transfected into HEK293T cells. Mean ± SEM, n = 3. **i,** relative light units were measured in lysates from HeLa cells transfected with ligated NanoLuc mRNAs carrying 2'O-methylation in different positions around the start codon. Unmethylated (blue) and unligated mRNAs were used as controls. Mean ± SEM, n = 4. p = One-way ANOVA with Tukey's post hoc analysis. **j,** denaturing agarose gel electrophoresis indicating the ligation efficiencies for the GFP mRNA reporters. These RNA pieces are larger than the ones used for NanoLuc reporters and do not resolve well in PAGE gels. Thus, agarose gels are better suited for their separation. **k,** schematic indicating the relative lengths of 5'UTRs in the NanoLuc and GFP mRNA reporters.

Extended Data Figure 3. Upstream AUGs and near-cognate codons are prevalent in human transcriptomes. Related to Fig. 3.



a-b, the 5'UTR sequence of all canonical protein-coding transcripts was downloaded from Biomart and analyzed to estimate the percentage of transcripts with one or more upstream AUGs (a) and the percentage of transcripts with one or more upstream near-cognate codon (b). This analysis indicated that 99% of human transcripts contain at least one near-cognate codon in their 5'UTRs, whereas 50% of human transcripts contain at least one AUG in their 5'UTR sequences. **c,** distribution of upstream near-cognate or AUG codons per transcript. CUG is the most prevalent near-cognate codon in human transcripts. **d,** while virtually all human transcripts contain cryptic initiation sites in their 5'UTRs, not all are used at a given moment or physiological condition. To determine the upstream near-cognate or AUG codons used for initiation, we previously performed HR-Ribo-seq in HeLa cells.[8] In this method, the inhibitor Harringtonine (HR) stalls the initiator ribosome while letting the elongator ribosomes slide off of mRNAs. Subsequent RNAse footprinting and isolation of ribosome-protected fragments (RPFs) allow the identification of initiation sites transcriptome-wide. **e,** HR-Ribo-seq indicates that ~70% of transcripts initiate translation from at least one upstream site in wildtype HeLa cells. **f,** the integrity of ligated NanoLuc mRNAs carrying near-cognate codons was verified using PAGE and SYBR gold staining. **g,** relative light units were measured in lysates from HeLa cells transfected with unmodified, C_mUG, or G_mUG-modified NanoLuc mRNAs. Mean ± SEM, n = 4. p = two-tailed *Student's t-test*.

954 **Extended Data Figure 4.** Distribution of 2'O-methylation sites in human mRNAs. Related to Fig.
955 4.



956
957 **a,** 2'O-methylation maps from four different studies, N_m-seq, N_m-mut-seq, NJU-seq, and
958 Nanopore-DRS were compiled and binned based on the transcript location (5'UTR, CDS, or
959 3'UTR) within mRNA transcripts**.** Data from four different cell lines is shown in the dot plots. **b,** the
960 genomic coordinate of 2'O-methylation sites were intersected with the coordinates for canonical
961 AUG start codon in human transcriptomes. Sites within a window of -3 to +4 nucleotides are
962 reported in four different cell lines. **c,** the prevalence of 2'O-methylation sites in upstream AUGs
963 or near-cognate codons per cell line. Data from all different studies and cell lines described in (a)
964 are included in the bar plots.
965

966 **Extended Data Figure 5.** Validation of 2'O-methylation sites in near-cognate codons. Related to
967 Fig. 4.

a

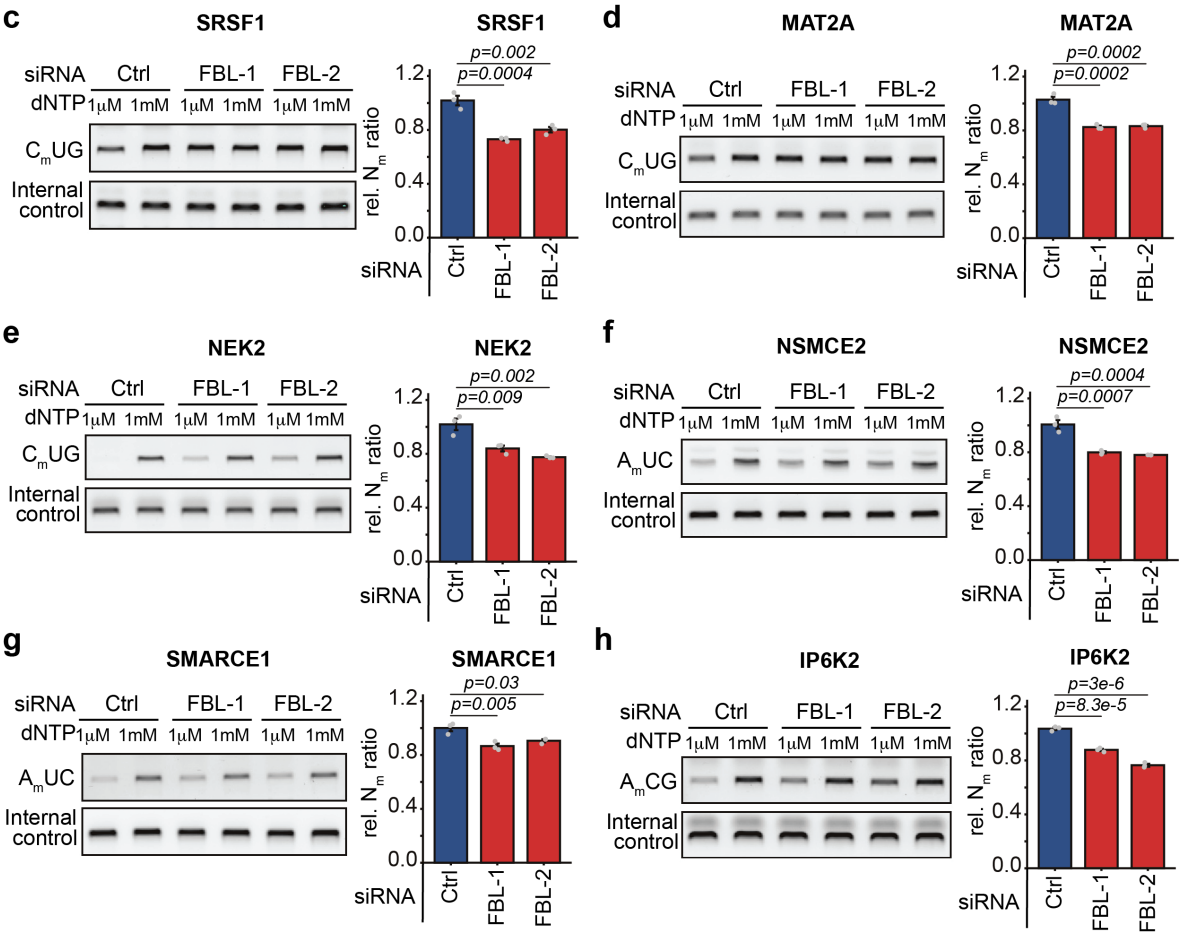
2′O-methyltransferases

Enzyme	Known substrates
FBL	rRNA, snRNA, and mRNA (internal sites)
FBLL1	mRNA (internal sites)
FTSJ3	Viral RNAs and mRNA (internal sites)
TRMT13	tRNA
FTSJ1	tRNA
FTSJ2	rRNA
HENMT1	3′end of piRNAs
CMTR1	mRNA cap
CMTR2	mRNA cap

b

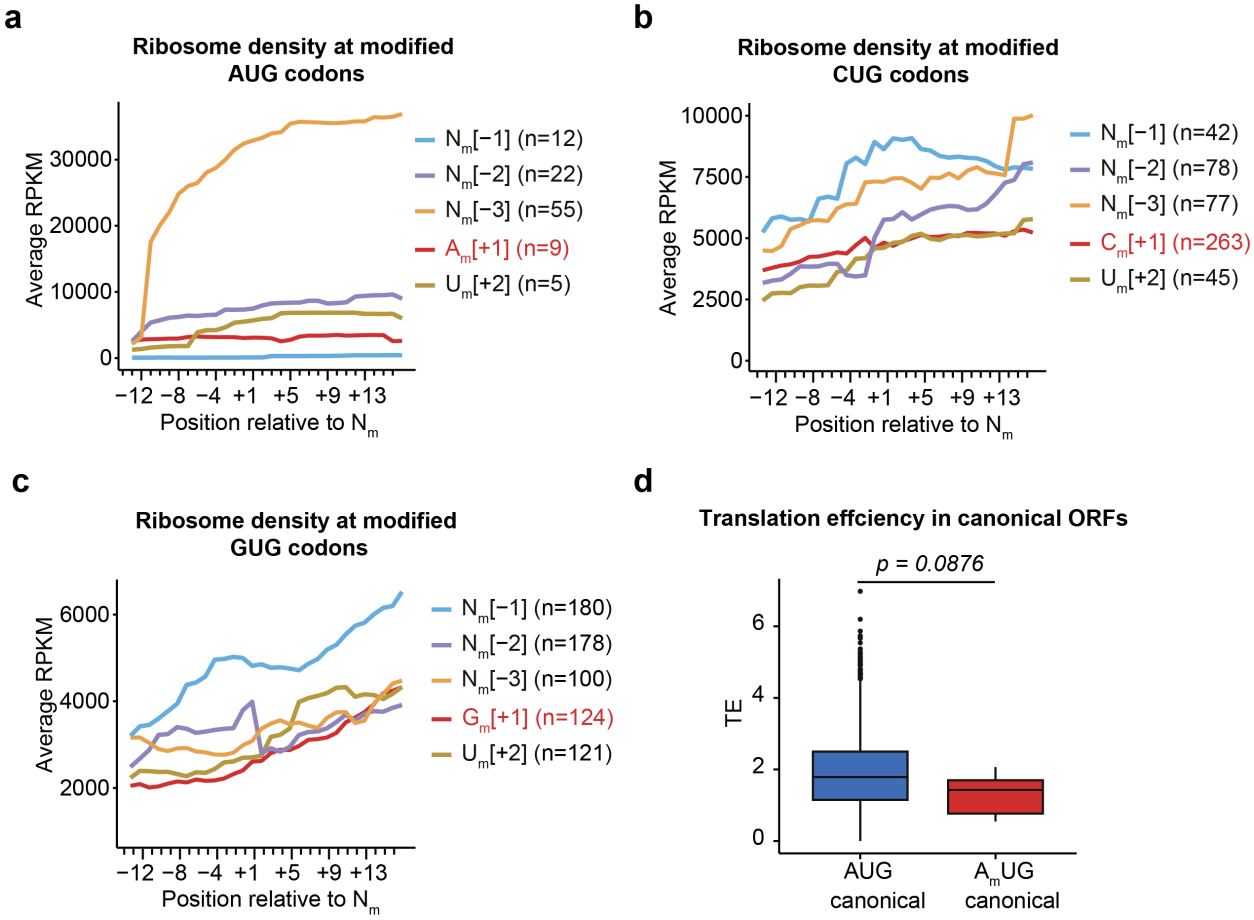
Validated N_m sites

Gene	Codon with N_m[+1]	Validation
GNL3	AUG	Validated
MYCBP	CUG	Validated
SRSF1	CUG	Validated
NEK2	CUG	Validated
MAT2A	CUG	Validated
MRPL24	GUG	Validated
SMARCE1	AUC	Validated
NSMCE2	AUC	Validated
IP6K2	ACG	Validated
RNPS1	Coding sequence	Validated
RPS9	GUG	Not validated
TAF9	UUG	Not validated



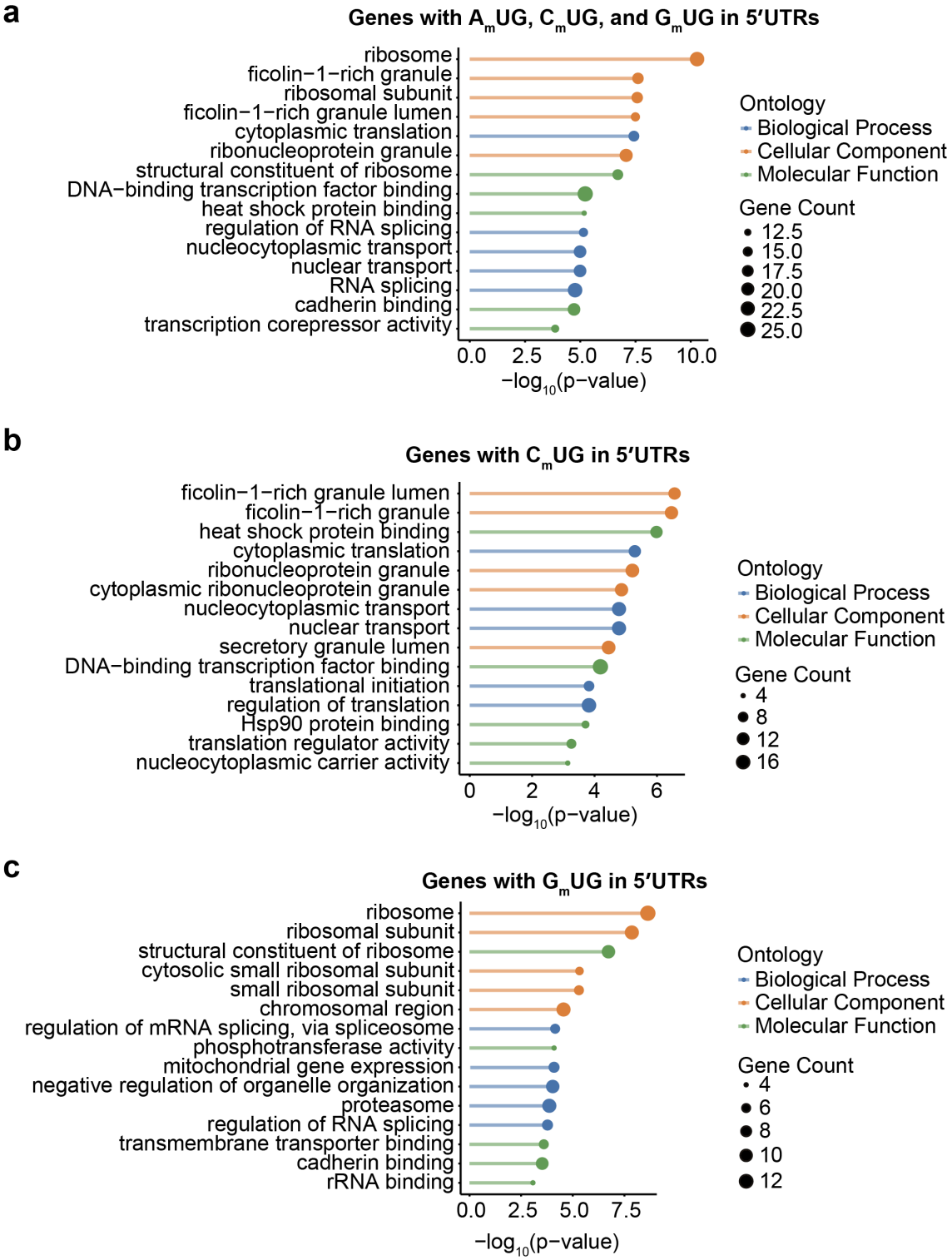
968
969 **a,** list of human 2'O-methyltransferase enzymes. **b,** list of validated sites. All tested sites were
970 upstream AUGs or near-cognate codons with 2'O-methylation in the first nucleotide of the codon.
971 Sites were selected based on the estimated modification ratio (the top sites) from the Nanopore-
972 DRS method[17]. **c-h,** *left:* representative agarose gels of amplicon density for the 2'O-methylated
973 site (top) and the loading control (bottom). *Right:* quantification of the relative N_m ratio dependent
974 on FBL. Mean ± SEM, n = 3. One-way ANOVA with Tukey's post hoc analysis.

975 **Extended Data Figure 6.** Ribosome density in upstream 2'O-methylated AUG, CUG, and GUG
976 codons. Related to Fig. 5.



977
978 **a-c,** averaged HR-Ribo-seq density (RPKMs) over a window of -13 to +16 nucleotides around
979 2'O-methylated codons at position N_m[+1] (red) or 2'O-methylated codons at positions N_m[-3], N_m[-
980 2], N_m[-1], and N_m[+2]. **d,** summarized translation efficiency (TE) in canonical A_mUG codons (n=5)
981 and canonical unmethylated AUG (n=7,091) start sites in HeLa cells. Translation efficiencies were
982 obtained from Arango et al., 2018.[27]
983

984 **Extended Data Figure 7.** Gene ontology analysis. Related to Fig. 5.
985



986 **a-c,** gene ontology analysis of all genes containing A_mUG, C_mUG, and G_mUG (a), containing only
987 C_mUG (b), or containing only G_mUG (c) codons in their 5'UTRs.