Site-Specific Inhibition of Translation Initiation via 2'-O-methylation

Adam Suh[1,†], Stephanie Mou[1,2,†], Emmely A. Patrasso[1,†], Hannah Serio[1], Kevin Vasquez[1,3], Rui Wang[4] , Smriti Sangwan[1], 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.

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

[†] Adam Suh, Stephanie Mou, and Emmely A. Patrasso contributed equally to this work

ABSTRACT

Translation initiation involves a concerted set of intermolecular interactions that efficiently recognize optimal AUG start codons. However, translation initiation complexes often start at upstream non-optimal AUGs or near-cognate codons, leading to the expression of upstream Open Reading Frames (uORFs). Using retrospective analyses of translation preinitiation cryo-EM structures, we identified putative hydrogen bonds between the 2'-OH groups of mRNA start codons with 18S rRNA. Disruption of these interactions using a chemical modification of mRNA, 2'-O-methylation (N_m), repressed translation initiation by preventing the preinitiation complex from recognizing start codons. Notably, 2'-O-methylation in upstream AUG and near-cognate codons inhibits upstream translation initiation while enhancing the expression of canonical ORFs. These findings revealed a transcript- and site-specific inhibitory role for 2'-O-methylation in translation initiation, providing novel insights into the mechanisms of start codon selection in human transcriptomes.

INTRODUCTION

During translation initiation, a 43S ribonucleoprotein particle containing the small ribosome subunit (40S), the initiator tRNA ($tRNA_i^{met}$), and translation initiation factors, binds the 5' end of mRNA to form a 48S preinitiation complex (1). This complex scans the 5' UTR until it encounters an optimal start codon (1). The optimality of a start codon is typically determined by the Kozak sequence (<u>R</u>CCAUG<u>G</u>, where R is either adenosine or guanosine) (2). Once the start codon is recognized, GTP hydrolysis and rearrangement of the 48S complex enable the large 60S subunit to bind and form the 80S initiation ribosome (1). Subsequently, the 80S ribosome enters elongation and continues decoding through an Open Reading Frame (ORF) until a stop codon is encountered (1).

While the Kozak sequence defines the optimal translation initiation site, translation can also initiate at suboptimal or cryptic start sites characterized by the presence of AUGs in non-Kozak sequences or near-cognate start codons, which differ from AUG by one nucleotide (CUG, GUG, UUG, AAG, ACG, AGG, AUA, AUC, and AUU) (3). While these cryptic start codons are usually bypassed by the 43S scanning complex through a process known as leaky scanning (4), translation initiation at upstream sites results in the expression of upstream Open Reading Frames (uORFs) (3). Initially thought to be mere cis-regulatory sequences that inhibit the

expression of downstream canonical ORFs, recent studies demonstrated that uORFs also yield stable micropeptides with diverse biological functions, many of which are involved in various conditions such as cancer, neurodegeneration, and metabolic diseases (5-7). However, the mechanisms underlying the selection of non-canonical start codons and subsequent uORF expression in disease states remain unclear.

One poorly characterized layer of translation regulation is the epitranscriptome, defined as the set of more than 150 chemical modifications of ribonucleotides that alter RNA function. Several RNA modifications have been reported to regulate translation initiation. For instance, the hypermodified nucleotide 6-threonyl-carbamoyl-adenosine (t^6A) in position 37 of $tRNA_i^{met}$ directly interacts with Kozak sequences, where it stabilizes the translation preinitiation complex (8,9). Moreover, the mRNA cap is characterized by a non-templated 7-methylguanosine (m^7G), followed by one or two templated nucleotides with methylation in the 2'-OH of the ribose, a modification called 2'-O-methylation (also referred to as A_m, C_m, G_m, U_m, or simply N_m) (10). When the first nucleotide is an adenosine, it is also modified into N^6,2′-O-dimethyladenosine (m^6A_m) (11). The m^7G cap is 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 internal 2′O-methylation sites within 5'UTRs is unknown.

Using synthetic mRNAs with 2′-O-methylation at every position of the Kozak sequence, we found that N_m at the first nucleotide of the AUG, CUG, ACG, and AUC start codons inhibits translation initiation. Mechanistically, 2'-O-methylation prevents the preinitiation complex from recognizing start codons. Notably, N_m at upstream start codons inhibits upstream initiation while enhancing translation from downstream canonical AUGs. Thus, this study revealed an inhibitory effect of 2'-O-methylation in start codon recognition, providing novel insights into the mechanisms of start codon selection in human cells. This study also provides the molecular basis for the potential use of 2'-O-methylation to inhibit aberrant uORF expression or enhance downstream ORF expression during RNA therapeutics.

MATERIALS AND METHODS

Cell culture

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco, Cat. #: 11965092) supplemented with 10% Fetal bovine serum (Corning, Cat. #: 30-015-CV) and 1% Penicillin/Streptomycin (Dot scientific, Cat. #: DSP93560-50). HeLa cells were cultured in DMEM (Gibco, Cat. #: 11965092) supplemented with 10% bovine calf serum (Cytiva, Cat. #: SH30073.04) and 1% Penicillin/Streptomycin. Cells were grown at 37 °C in a humidified environment with 5% CO_2.

Synthesis and capping of 5' RNA fragments

To produce the 170 nt 5'UTR, an antisense DNA template containing a T7 promoter sequence was purchased from Integrated DNA Technologies (IDT) and annealed to a sense T7 primer (Supplementary Table S1). To produce the 504 nt 5'UTR, we generated the DNA template by PCR using primers that annealed to a plasmid containing the coding sequence of the micro red fluorescence protein, ensuring that the amplicon excluded any start codon. The forward primer contained the T7 sequence as an overhang. PCR was performed using the Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Cat. #: M0494).

The 170 nt and 504 nt 5'UTRs were *in vitro* transcribed using the HiScribe® T7 High Yield RNA Synthesis Kit (New England Biolabs, Cat. #: E2040), following the manufacturer's suggestions. RNA was purified using 1.8X SPRI beads (CleanNGS Cat. #: CNGS-0001). Purified RNA was capped *in vitro* using the Vaccinia Capping System (New England Biolabs, Cat. #: M2080) and re-purified using 1.8X SPRI beads. The capped 27 nt 5'UTR was custom-synthesized by TriLink BioTechnologies.

The integrity of RNA fragments was verified using polyacrylamide gels prepared in a tris-borate-urea buffer (2 mL of 40% PAA [19:1], 5 g of Urea, 1mL of 10X TBE, complete to 10 mL ddH2O). Electrophoresis was run in a 0.5X TBE running buffer and stained with a 1X SYBR gold solution (ThermoFisher, Cat. #: S11494). Gels were visualized in a Sapphire Biomolecular Imager (Azure Biosystems). Primers, RNA oligonucleotides, and resulting DNA sequences are described in Supplementary Table S1.

Dot blots

Dot blots were performed to verify the capping of *in vitro* transcribed RNAs. Briefly, 500 ng RNA was denatured at 65°C for 5 min and loaded onto Hybond-N+ membranes. Membranes were

123 crosslinked with 120 mJ^2/cm^2 in the UV Crosslinker. Membranes were first stained in a 0.2%
124 methylene blue solution (0.2% methylene blue in 0.4M sodium acetate and 0.4M acetic acid),
125 then washed three times in DI water, and imaged in a Sapphire Biomolecular Imager.
126
127 Membranes were then blocked with 1% bovine serum albumin (BSA) in PBS containing 0.05%
128 Tween-20 (PBST) for 15 min at room temperature and probed overnight with mouse anti-m^7G
129 antibodies (1:1000, MBL International Corporation, Cat. #: RN016M) in 1% BSA at 4°C.
130 Membranes were next washed three times with 0.1% PBST, incubated with HRP-conjugated
131 secondary anti-mouse IgG (1:10000 dilution, Cell Signaling Technologies, Cat. # 7076S) in 1%
132 BSA at room temperature for 1 hr, washed three times with 0.1% PBST and developed with
133 enhanced chemiluminescent (ECL) western blotting substrate (Promega, Cat #: W1001) on a
134 Sapphire Biomolecular Imager.
135
136 **Synthesis and polyadenylation of 3' RNA fragments**
137 To produce the ORF and 3'UTR of NanoLuc, we generated a DNA template by PCR using primers
138 that annealed to a plasmid containing the NanoLuc coding sequence, ensuring the amplicon
139 excluded the canonical start codon. The forward primer contained the T7 sequence as an
140 overhang. To produce the ORF and 3'UTR of GFP, we generated a DNA template by PCR using
141 primers that annealed to a plasmid containing the GFP coding sequence, ensuring the amplicon
142 excluded the canonical start codon. The forward primer contained the T7 sequence as an
143 overhang. PCR was performed using the Q5 Hot Start High-Fidelity 2X Master Mix (New England
144 Biolabs, Cat. #: M0494). To produce the mRNA reporter containing a uORF, the canonical
145 NanoLuc ORF, and the 3'UTR, we generated the DNA template by PCR using primers that
146 annealed to a plasmid that incorporated an uORF or an overlapping uORF upstream of the coding
147 sequence of NanoLuc, making sure to exclude the upstream canonical start codon from the
148 amplicon product. Primers and resulting DNA sequences are described in Supplementary Table
149 S1.
150
151 DNA duplexes were *in vitro* transcribed using the HiScribe® T7 High Yield RNA Synthesis Kit
152 (New England Biolabs, Cat. #: E2040) according to the manufacturer's instructions. RNA was
153 purified using 1X SPRI beads (CleanNGS Cat. #: CNGS-0001). RNA fragments were
154 dephosphorylated with 100 U of calf intestinal alkaline phosphatase (CIAP, Promega, Cat. #:
155 M2825), followed by mono-phosphorylation of the 5' end using 15 U of T4 Polynucleotide Kinase
156 (PNK, New England Biolabs, Cat. #: M0201L) in the presence of ATP. Alternatively, *in vitro*

157 transcribed RNAs were treated with RppH (New England Biolabs, Cat. #: M0356S), which
158 converts 5'-triphosphates into 5'-monophosphates. 5'-monophosphorylated RNA products were
159 purified using the RNA Clean & Concentrator-25 Kit (Zymo Research, Cat. #: R1018). RNA
160 products (10 μg) were polyadenylated using 5 U of *E. coli* Poly(A) Polymerase (New England
161 Biolabs, Cat. #: M0276). The integrity of polyadenylated RNA fragments was verified using TBU-
162 PAGE. Electrophoresis was run in 0.5X TBE buffer and stained with 1X SYBR Gold.
163
164 **Splint ligations**
165 5'-monophosphorylated RNA oligonucleotides (12 nt) containing the Kozak sequence with an
166 AUG, CUG, GUG, ACG, or AUC start codons were purchased from IDT. Likewise, the 2'-O-
167 methylated versions were purchased from IDT (Supplementary Table S1). For three-way splint
168 ligations, these RNA oligos were combined with 5' and 3' *in vitro* transcribed RNA fragments at
169 equimolar concentrations (15 pmol). An antisense DNA oligonucleotide (15 pmol) was added to
170 splint the different RNA pieces. Reactions were denatured at 65 °C for 2 minutes and ramped
171 down to room temperature for annealing. For two-way splint ligation, unmethylated or 2'-O-
172 methylated capped RNA oligonucleotides were purchased from TriLink BioTechnologies and
173 combined with a 3' *in vitro* transcribed RNA fragment at equimolar concentrations (15 pmol). An
174 antisense DNA oligonucleotide (15 pmol) was added to splint the different RNA pieces.
175
176 Ligations were performed using 10 U of T4 RNA ligase 2 (dsRNA ligase, New England Biolabs,
177 Cat. #: M0239) for 4 hr at 37 °C, according to the manufacturer's recommendations. Reactions
178 were stopped with 2 U of Turbo DNase I (ThermoFisher Scientific, Cat. #: AM2238) for 30 min at
179 37°C and purified using the RNA Clean & Concentrator-5 kit (Zymo Research, Cat. #: R1014). To
180 verify the integrity of splint-ligated NanoLuc RNA, 500 ng of ligated RNAs were separated by 4%
181 TBU-PAGE and stained with 1X SYBR gold (Invitrogen, Cat. #: S11494). To verify the integrity of
182 splint-ligated GFP RNA, 500 ng of ligated RNAs were separated by 1% denaturing agarose gels
183 and stained with 1X SYBR gold (Invitrogen, Cat. #: S11494). Gels were visualized in a Sapphire
184 Biomolecular Imager (Azure Biosystems).
185
186 To purify ligated RNA fragments, gels were stained with SYBR Gold, imaged, and aligned to a
187 printed scan to guide precise excision of RNA bands. Target gel slices were excised, crushed,
188 and RNA was eluted overnight at room temperature in elution buffer (300 mM NaCl, 10 mM Tris
189 pH 7.0, 1 mM EDTA, 0.25% SDS). Eluates were clarified by centrifugation, and RNA was
190 precipitated with glycogen and ethanol at −80 °C, followed by centrifugation. Pellets were washed

191 twice with 70% ethanol, air-dried, and resuspended in nuclease-free water. The integrity of
192 purified mRNA reporters was verified by TBU-PAGE and SYBR gold staining.
193
194 **NanoLuc and Firefly luciferase assays**
195 HEK293T or HeLa cells were seeded at a density of 30,000 cells in 100 μL complete DMEM
196 medium in 96-well plates. Cells were cultured for 24 hr before transfection with 300 ng of capped
197 and polyadenylated splint-ligated RNA or an unligated RNA control. Transfection reactions were
198 prepared using the TransIT-mRNA Transfection Kit (Mirus, Cat. #: MIR2225) following the
199 manufacturer's instructions. Following the addition of the transfection mixtures, cells were
200 incubated for 6 hrs at 37°C. For luminescence detection, cells were assayed with Nano-Glo®
201 Luciferase Assay System (Promega, Cat. #: N1110) according to the manufacturer's instructions
202 for 96-well plates. For the dual transfection experiments, cells were assayed using the Nano-Glo
203 Dual-Luciferase Reporter Assay System (Promega, Cat. #: N1610) according to the
204 manufacturer's instructions for 96-well plates. Luminescence was read in a BioTek Synergy LX
205 Multimode Reader (Agilent Technologies).
206
207 **Flow cytometry**
208 HEK293T or HeLa cells were seeded at a density of 150,000 cells in 1 mL medium in 12-well
209 plates. Cells were transfected with 600 ng of capped and polyadenylated splint-ligated mRNA or
210 unligated RNA control. Transfection reactions were prepared using the TransIT-mRNA
211 Transfection Kit (Mirus) following the manufacturer's instructions for 12-well plates. Following the
212 addition of the transfection mixtures, cells were incubated for 18 hrs at 37°C. Cells were
213 trypsinized and washed with PBS three times before they were resuspended in 500 μL PBS and
214 immediately analyzed via Flow cytometry in an LSR Fortessa 1 Analyzer or a BD FACSymphony
215 A5-Laser Analyzer. The percentage of GFP-positive cells was determined using FlowJo v10.10.0.
216
217 **RNA isolation and RT-qPCR**
218 RT-qPCR was performed to evaluate the efficiency of ligation before transfection or the stability
219 of ligated mRNA at different time points (0, 3, and 6 hr) post-transfection. Briefly, total RNA was
220 isolated from cells at 0, 3, and 6 hr post-transfection using the Trizol method. The extracted RNA
221 was reverse transcribed using SuperScript™ IV Reverse Transcriptase (ThermoFisher Scientific,
222 Cat. #: 18090010) using gene-specific RT primers for NanoLuc, GFP, and 18S rRNA
223 (Supplementary Table S1). qPCR was performed using the 2X PowerUp SYBR Master Mix
224 (ThermoFisher Scientific, Cat. #: A25742) and primers that amplify the ligated mRNAs, total

225 mRNAs (ligated plus unligated), and 18S rRNA (Supplementary Table S1). To obtain the
226 efficiency of ligation, we applied the formula 100 * 2^[-($Ct^{ligated}$/$Ct^{total.mRNA}$). To measure mRNA
227 stability, we applied the formula final.time(2^[-($Ct^{ligated}$/$Ct^{18S.rRNA}$)/initial.time(2^[-($Ct^{ligated}$/$Ct^{18S.rRNA}$).
228
229 ***S*ucrose density gradients**
230 Cy5-labeled synthetic mRNAs were purchased from Genescript (Supplementary Table S1).
231 Reticulocyte lysates (37.5 μL) were pre-incubated with vehicle DMSO, 1 mM GMPPNP
232 (Guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate, Millipore Sigma, Cat. #: G0635), or
233 100 μg/ml Harringtonine (Abcam, Cat. #: ab141941) prior to adding 150 nM of Cy5-labeled
234 synthetic mRNAs in a total volume of 50 μL. Reactions were incubated for 10 min at 30 °C and
235 immediately stopped on ice. Reactions were loaded on a 5-30% sucrose density gradient (50 mM
236 Tris-acetate [pH 7.0], 50 mM NH_4Cl, 12 mM $MgCl_2$, 1 mM DTT) poured using a BioComp Gradient
237 Master. Gradients were spun at 42,000 rpm in a Beckmann Coulter SW41 Ti rotor for 2:26 hr at
238 4°C. Recording of Cy5 fluorescence in different fractions was performed using a BioComp Density
239 Gradient Fractionation system.
240
241 **siRNA transfection**
242 The siRNAs used in this study were purchased from Thermo Fisher (siFBL-1 Cat. #: s4820; siFBL-
243 2 Cat. #: s4821; siCtrl Cat. #: 4390844) (Supplementary Table S1). The siRNAs were transfected
244 into HeLa cells using the Lipofectamine RNAiMAX Transfection Reagent (ThermoFisher
245 Scientific, Cat. #: 13778150) by following the manufacturer's instructions. Total RNA and cell
246 lysates were collected three days post-transfection. The efficiency of FBL knockdown was verified
247 by Western blot.
248
249 **Western blot**
250 Cells were washed with cold 1x phosphate-buffered saline (PBS, Corning, 21-040-CM) and lysed
251 in RIPA lysis buffer containing 150mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1%
252 SDS, and 50mM Tris (pH 8.0). Lysates were centrifuged at 13,200 rpm for 10 minutes at 4 °C,
253 and protein concentrations were determined using Micro BCA™ Protein Assay Kit (ThermoFisher
254 Scientific, 23235). Thirty micrograms of protein were separated by 12% SDS-PAGE Tris-Glycine
255 at 100 V for 1.5 hours and transferred overnight at 4 °C to a nitrocellulose membrane. Membranes
256 were blocked with 5% non-fat milk in Tris-buffered saline with 0.05% Tween-20 (TBST) for at least
257 15 minutes at room temperature. Membranes were then incubated overnight at 4 °C with one of
258 the following primary antibodies in 1% non-fat milk in TBST: rabbit polyclonal anti-FBL (1

259 microgram/mL, Cat. #: ab5821, abcam); rabbit polyclonal anti-β-actin (1:2000, Cat. #: ab8227,
260 abcam); mouse monoclonal anti-Nucleostemin (anti-GNL3) (1:1000, Cat. #: sc-166460, Santa
261 Cruz Biotechnology); mouse monoclonal anti-Nek2 (1:1000, Cat. #: sc-55601, Santa Cruz
262 Biotechnology); or mouse monoclonal anti-SF2/ASF (anti-SRSF1) (1:1000, Cat. #: sc-33652,
263 Santa Cruz Biotechnology). Membranes were washed three times (if monoclonal) or five times (if
264 polyclonal) with TBST and incubated with HRP-conjugated anti-rabbit IgG (1:5000, Cat. #: 7074,
265 Cell Signaling Technology) or HRP-conjugated anti-mouse IgG (1:5000, Cat. #: 7076, Cell
266 Signaling Technology) for 1 hour at room temperature. Signals were detected using ECL Western
267 Blotting Substrate (Promega, W1001) or SuperSignal™ West Femto Maximum Sensitivity
268 Substrate (ThermoFisher Scientific, 34095) and visualized with a Sapphire multimode imager
269 (Azure Biosystems). Band densitometry was quantified using ImageJ. The relative expression
270 levels were estimated by normalizing to the β-actin loading control.
271
272 **Identification of 2'-OH hydrogen bonds in translation preinitiation complexes**
273 The following structures were downloaded from the Protein Data Bank: 7UCJ (rabbit reticulocyte),
274 6YAL (rabbit reticulocyte), 8PJ1-6 (reconstituted human 48S). Structural visualization was
275 performed in PyMOL. Briefly, the background was set to white, and all elements were initially
276 hidden. Specific residues of the mRNA, centered around nucleotides N[-3], N[-2], N[-1], N[+1],
277 N[+2], N[+3], and N[+3] were selected and visualized as both ribbons (with increased width) and
278 sticks. Specific residues within $tRNA_i^{met}$, 18S rRNA, eIF1A, and eIF2α that interact with mRNA
279 were also selected. Hydrogen bond calculations were performed between mRNA, $tRNA_i^{met}$, 18S
280 rRNA, eIF1A, and eIF2α. Hydrogen-bond geometry involving the ribose O2′ group, distances
281 between O2′–P and O2′–O2 atoms, and corresponding heavy-atom proxy angles were measured
282 using PyMOL. Density maps and corresponding models were visualized in ChimeraX. Molecular
283 animation videos were generated using the movie-recording tools. Contour levels were set below
284 the EMDB-recommended values for all except 7UCJ and 6YAL, for which the levels were adjusted
285 to show the density clearly. Supplementary Video 1_7UCJ contour levels = 0.0183 emdb_26444,
286 Supplementary Video 2_6YAL contour levels = 0.0196 emdb_10760, Supplementary Video
287 3_7UCJ contour levels = 0.016 emdb_26444, Supplementary Video 4_8PJ3 contour levels = 5.82
288 emdb_17698, Supplementary Video 5_8PJ3 contour levels = 4.0 emdb_17698, Supplementary
289 Video 5_8PJ3 contour levels = 4.73 emdb_17698.
290
291 **Identification of N_m sites in AUG and near-cognate codons**

292 Maps of 2'-O-methylation in C4-2 and HeLa cells were extracted from Li et al. 2024 (17). N_m sites
293 with a 2'-O-methylation ratio above ten percent were selected for downstream analysis. To identify
294 N_m sites in canonical AUG codons, we extracted the genomic information for Ensembl canonical
295 start codons from a human genome annotation file (GRCh38). Then, bedtools/2.29.2 was used
296 to identify N_m sites around canonical start codons. N_m sites were then binned by their location
297 relative to the canonical AUG, selecting only sites at position N_m[+1]. To identify N_m sites in
298 upstream AUG and near-cognate codons, we selected N_m sites within protein-coding genes that
299 aligned to 5'UTR in any possible transcript. Then, we extracted the genomic sequence flanking
300 each N_m site. We filtered for sequences containing AUG and near-cognate codons. Each N_m site
301 location was derived relative to the AUG or near-cognate codons N_m[+1]. This analysis was
302 performed using custom scripts in Biopython 1.85.
303
304 **RTL-P assay**
305 To confirm the N_m occurrence along with its alteration upon FBL depletion, the RTL-P assay was
306 conducted as previously described (17). First, total RNAs from control and FBL-knockdown HeLa
307 cells were isolated using the RNeasy Plus Mini Kit (Qiagen). Reverse transcription (RT) was then
308 performed using specific reverse primers targeting mRNA sequences downstream of the 2'-O-
309 methylation site in the presence of either a low (1 μM) or a high (1 mM) concentration of dNTPs.
310 The obtained cDNA was subjected to two separate amplification reactions using one pair of
311 primers targeting upstream and downstream of the methylation site, and another pair targeting a
312 downstream region that would not be interrupted by the 2'-O-methylation site (Supplementary
313 Table S1). The PCR mixture consisted of 10 μL AmfiSure PCR master mix (GenDEPOT), 2 μL
314 cDNA, and 1 mM PCR primers. The PCR products were then separated using 1% agarose gels
315 and photographed using a Bio-Rad imaging system. DNA band signal intensities were calculated
316 using ImageJ. The methylation ratio of each group was determined from the densities of the PCR
317 bands obtained using high and low dNTP concentrations. The RT and PCR primers designed for
318 the RTL-P assay are listed in Supplementary Table S1.
319
320 **Codon bias analysis**
321 Using a list of genes that were expressed in HeLa HR-Ribo-seq (8), 5'UTR sequences of the
322 longest transcript per gene were extracted from Ensembl/113 with BioMart. This was also
323 performed for genes containing N_m[+1]. The frequencies of AUG and near-cognate codons were
324 calculated for modified codons in the 5'UTRs and for the presence of these codons in the 5'UTRs

of expressed protein-coding genes. Codon frequencies were analyzed using a chi-square test of proportions with a Benjamini-Hochberg correction.

Gene ontology analysis

Genes containing N_m[+1] in upstream AUG or near-cognate codons in their 5'UTR were used for gene ontology (GO) analysis for three ontologies: biological process, cellular component, and molecular function. An R script was used to perform and visualize this analysis using the clusterProfiler/4.14.6 and org.Hs.eg.db/3.20.0 packages. The top 5 categories for each ontology were displayed in descending order of *p-value*.

Analysis of translation efficiency in HeLa cells

We analyzed publicly available Ribo-seq performed in HeLa cells (GSE105248) (21). This dataset contained three Ribo-seq and Input RNA-seq from FBL knockdown (Doxycycline-inducible shRNA against FBL) and control groups. Ribosome-protected fragments (RPFs) were derived by trimming adapters from raw sequence files with TrimGalore/0.6.10, aligning to the transcriptome with STAR/2.7.9a, and featureCounts extracted RPFs with subread/2.0.3. mRNA and RPF counts were normalized with DESeq2/1.46.0, and translation efficiency was calculated as normalized RPF counts divided by normalized mRNA counts. We included only protein-coding genes with normalized mRNA counts> 0. The log2FoldChange of translation efficiency was calculated with Rstudio/v4.4.0. We compared genes with mapped N_m[+1] at an upstream AUG or near-cognate codon in the 5'UTR with those with no mapped N_m in the 5'UTR in HeLa cells. Differences in translation efficiency between groups were analyzed with a Wilcoxon rank-sum test.

Codon usage in upstream initiation sites

Codon identity of translation initiation sites from Harringtonine Ribo-seq in HeLa cells was downloaded from Arango et al., 2022 (8). Data were binned by codon identity and visualized as a pie chart in RStudio/v4.4.0 using the packages ggpubr and ggplot2.

Other statistical analysis

All other statistical tests and plotting were performed in Rstudio/v4.4.0. The specific statistical tests, graph type, variables, number of observations per variable, and *p-values* are indicated in the figures or their legends.

RESULTS

359

360 **2'-OH hydrogen bonds within translation initiation complexes**

361

362 To understand what defines an optimal start codon, several groups solved cryo-EM structures of
363 mammalian preinitiation and initiation complexes (8,9,22,23). These efforts provided a snapshot
364 of the crucial intermolecular interactions surrounding Kozak sequences (Fig. 1A). Besides the
365 recognition of start codons by the anticodon loop in $tRNA_i^{met}$, the interactions of positions A/G[-3]
366 and G[+4] with translation initiation factors are required for translation initiation (Fig. 1A). In
367 revisiting several mammalian translation initiation structures (8,9,22), we observed that 2'-OH
368 groups of mRNA nucleotides within the Kozak sequence appeared to be interacting with other
369 components of the preinitiation complex. For instance, t^6A in position 37 of $tRNA_i^{met}$ is in proximity
370 with the 2'-OH group of the nucleotide preceding the start codon, C[-1] (Fig. 1B, Supplementary
371 Fig. S1A, and Supplementary Table S2). The 2'-OH group of the first nucleotide, A[+1], in the start
372 codon appears to interact with U1830 of 18S rRNA, with the phosphate (PO_4^{3-}) in the backbone
373 (Fig. 1C), and possibly with the keto (C=O) group in position C2, especially in the late stages of
374 the preinitiation complex (Fig. 1D and Supplementary Fig. S1B-F). The H-bonds between the 2'-
375 OH groups in C[-1] and A[+1] with tRNA and 18S rRNA, respectively, are supported by structural
376 models, electron densities, distances between O2′–P and O2′–O2 atoms, and the corresponding
377 heavy-atom proxy angles (Fig. 1B and C, Supplementary Videos 1-4, and Supplementary Table
378 S2).

379

380 Additional putative H-bonds were observed between the 2'-OH group of nucleotide A[-3] in the
381 mRNA with G961 of 18S rRNA and between the 2'-OH group of G[+3] with lysine seven (K7) of
382 the translation initiation factor 1A (eIF1A), although the latter are not well supported by electron
383 densities or heavy-atom proxy angles (Fig. 1E-F, Supplementary Videos 5-6, and Supplementary
384 Table S2). Together, these observations suggest that 2'-OH groups in mRNA interact with 18S
385 rRNA and $tRNA_i^{met}$, potentially playing a crucial role in start codon recognition.

386

387 **2'-O-methylation in position A[+1] of AUG start codons inhibits translation**

388

389 To probe the significance of the putative hydrogen bonds formed by 2'-hydroxyl groups in mRNA,
390 we sought to systematically introduce a 2'-O-methylation modification in every nucleotide within
391 the Kozak sequence, as the 2'-OCH_3 group lacks the donor capacity of 2'-OH (Fig. 2A). While 2'-
392 O-methylation can be incorporated at specific positions within RNA molecules by chemical

393 synthesis, this approach is limited by the length of RNA fragments that can be chemically
394 synthesized. We therefore implemented a three-way splint ligation approach to introduce 2'-O-
395 methylation (Fig. 2B) (24). Briefly, a 5'UTR lacking a start codon was transcribed and capped *in*
396 *vitro* (Supplementary Fig. S2A and B). In parallel, the ORF of NanoLuciferase (NanoLuc)—also
397 lacking a start codon—was transcribed and polyadenylated *in vitro* (Supplementary Fig. S2C).
398 These pieces were ligated to a chemically synthesized RNA fragment containing the unmodified
399 or 2'-O-methylated Kozak sequence using a DNA splint that hybridizes to the three independent
400 RNA pieces (Fig. 2B).
401
402 To test the specificity of this system, we first transfected HEK293T cells with the ligated NanoLuc
403 mRNAs containing a start codon, ligated NanoLuc mRNAs lacking a start codon, or unligated
404 RNA pieces (Fig. 2C, Supplementary Fig. S2D). The ligated mRNAs containing the start codon
405 within a Kozak sequence exhibited an ~100-fold increase in NanoLuc activity compared to the
406 unligated RNA pieces or the mRNAs lacking a start codon (Fig. 2C), demonstrating that ligation
407 of a start codon is needed for translation of the mRNA reporters. To control for possible
408 confounding artifacts, wherein a high proportion of unligated products affects translation, we
409 transfected increasing concentrations of full-length NanoLuc mRNA with constant amounts of
410 unligated RNA (Fig. 2D). Reciprocally, we transfected constant amounts of full-length NanoLuc
411 mRNA with increasing concentrations of unligated RNAs (Supplementary Fig. S2E). In either
412 case, NanoLuc activity was unaffected by the unligated RNAs, exhibiting linear behavior across
413 a broad range of mRNA concentrations.
414
415 We next ligated unmodified or 2'-O-methylated Kozak sequences. Ligation efficiency was verified
416 by RT-qPCR, yielding similar values across multiple replicates (Supplementary Fig. S2F and G).
417 Equivalent amounts of ligated mRNAs were transfected into HEK293T cells, and translation
418 efficiency was measured six hours after transfection through a luminescence assay. Remarkably,
419 methylation of the 2'-OH group in the first position of the AUG codon significantly reduced
420 NanoLuc activity to background levels (Fig. 2E). Significant luminescence reduction was also
421 observed when 2'-O-methylation occurred on positions G[+3] and G[+4] (Fig. 2E). In contrast, no
422 effect was observed in positions A[-3], C[-2], C[-1], and U[+2] (Fig. 2E). The inhibition of NanoLuc
423 activity by A_m[+1] remained after normalization of NanoLuc activity by transfected mRNA levels
424 (Supplementary Fig. S2H).
425

426 To ensure that the NanoLuc signal was not influenced by unligated RNA fragments or potential
427 artifacts from the three-way splint ligation, we used two complementary validation approaches
428 (Supplementary Fig. S2I). First, we gel-purified full-length unmodified or A_m[+1] mRNA reporters
429 (Supplementary Fig. S2J and K). For this experiment, a longer 5'UTR was used to facilitate clear
430 separation of the RNA fragments during gel purification (Supplementary Fig. S2I and J). Second,
431 we carried out a two-way splint ligation using a short (27 nt) 5' RNA fragment that was capped
432 and 2'-O-methylated through chemical synthesis (Fig. 2F). The two-way splint ligation achieved
433 nearly 100% ligation efficiency (Supplementary Fig. S2L and M).
434
435 Transfection of either gel-purified three-way ligated mRNA reporters or the two-way ligated
436 counterpart produced results consistent with those obtained using unpurified mRNA reporters:
437 the A_m[+1] modification significantly inhibited NanoLuc activity to background levels (Fig. 2G).
438 Although the two-way splint ligation approach provides high ligation and translation efficiency (Fig.
439 2G), chemically synthesized RNA has a 5'-OH end, which is unsuitable for enzymatic capping.
440 Consequently, chemical capping is required, which substantially increases costs and limits the
441 maximum length of RNAs that can be synthesized. Therefore, the two-way method is not cost-
442 effective for systematically screening the site-specific effect of multiple RNA modifications in
443 mRNA reporters. Taken together, our findings support the use of unpurified splint-ligated mRNA
444 reporters to screen site-specific effects of 2'-O-methylation in translation regulation.
445
446 We next evaluated the reproducibility of our finding across different mRNA reporters and cell
447 types. First, we transfected the ligated NanoLuc mRNAs into HeLa cells. Like in HEK293T cells,
448 A_m[+1] inhibited NanoLuc activity to background levels in HeLa cells (Supplementary Fig. S3A).
449 G_m[+3] and G_m[+4] significantly reduced NanoLuc activity, albeit to a lesser extent than A_m[+1]
450 (Supplementary Fig. S3A). Second, we generated green fluorescent protein (GFP) mRNA
451 reporters, each one containing 2'-O-methylation at a different position in the Kozak sequence
452 (Supplementary Fig. S3B and C). Following transfection of the GFP mRNA reporters into
453 HEK293T cells, protein expression was determined by flow cytometry. Consistently, the GFP
454 fluorescence results mirror those of NanoLuc activity: A_m[+1] inhibited GFP expression to baseline
455 levels; G_m[+3] and G_m[+4] reduced protein expression, whereas A_m[-3], C_m [-2], C_m [-1], and U_m
456 [+2] had no effect on GFP expression (Fig. 3A and B).
457
458 To guarantee that the inhibitory effect of A_m[+1] on NanoLuc activity is transcript-specific and not
459 a global perturbation of protein synthesis, we co-transfected unmodified or A_m[+1]-modified

460 NanoLuc mRNAs with unmodified Firefly luciferase mRNAs (Supplementary Fig. S3D). We
461 observed that A_m[+1] decreased NanoLuc activity to background levels without altering Firefly
462 luciferase activity (Fig. 3C). In addition, evaluation of mRNA levels at different time intervals post-
463 transfection showed no difference in mRNA stability between unmodified and 2'-O-methylated
464 transcripts (Fig. 3D), indicating that A_m[+1] inhibits translation only in 2'-O-methylated mRNAs.
465
466 In summary, consistent translation inhibition by A_m[+1] was observed across unpurified and gel-
467 purified three-way ligations, two-way ligations, three types of 5'UTRs (170 nt, 504 nt, and 27 nt),
468 two reporters (NanoLuc, GFP), two cell types (HEK293T, HeLa), and was not attributable to
469 mRNA instability. Since 2'-O-methylation eliminates the donor capacity of 2'-OH groups, these
470 findings suggest that H-bonds between the 2'-OH group of nucleotide A[+1] and U1830 in 18S
471 rRNA are relevant during translation initiation.
472
473 **2'-O-methylation reduces translation from non-AUG start codons**
474
475 We previously performed Ribo-seq in the presence of harringtonine (8), a translation inhibitor that
476 allows the mapping of translation initiation sites transcriptome-wide (Supplementary Fig. S4A) (8).
477 These experiments identified alternative translation initiation sites in ~70% of cellular mRNAs
478 (Supplementary Fig. S4B). Distribution of upstream start codons shows a higher prevalence of
479 CUG, AUG, and GUG, with evidence for upstream initiation in all near-cognate start codons (Fig.
480 4A). Building on these observations, we sought to determine whether 2'-O-methylation affects
481 translation initiation in the most prevalent near-cognate start codons: CUG and GUG. Constructs
482 were produced by three-way splint ligation followed by gel purification (Fig. 4B and Supplementary
483 Fig. S4C).
484
485 Expectedly, the unmodified CUG and GUG start codons exhibited reduced NanoLuc activity
486 compared to AUG (Fig. 4C). A_mUG reduced NanoLuc activity by ~11-fold, whereas C_mUG
487 significantly inhibited NanoLuc activity by only 1.8-fold (Fig. 4C). In contrast, a non-significant
488 effect was observed for G_mUG, although the NanoLuc levels for unmodified GUG were as low as
489 those of C_mUG (Fig. 4C). Given the more substantial effect of 2'-O-methylation in AUG codons,
490 we sought to determine whether this observation is an adenosine-specific effect or a cognate
491 codon-specific effect. Thus, we tested A_mCG and A_mUC, which retain adenosine at position [+1]
492 but differ at positions [+2] and [+3], respectively (Supplementary Fig. S4D). Though A_mCG and
493 A_mUC significantly inhibited NanoLuc activity, with a ~2-fold decrease for A_mCG and a ~3.5-fold

494 decrease for A_mUC (Fig. 4D), the effect of 2'-O-methylation in all near-cognate codons was
495 diminished compared to A_mUG (Fig. 4E). These findings indicate that 2'-O-methylation exerts the
496 most substantial effect in cognate AUG start codons.
497
498 Based on the modest inhibitory effect of 2'-O-methylation in near-cognate codons, we revisited
499 cryo-EM structures of eukaryotic translation initiation complexes on near-cognate initiation. The
500 only structure available was for the human 48S preinitiation complex on AUC, which shows an
501 absence of the H-bond between A[+1] in mRNA and U1830 in 18S rRNA (Supplementary Fig.
502 S4G).
503
504 Collectively, the presence of a putative hydrogen bond between the 2'OH group of A[+1] and 18S
505 rRNA, the absence of this H-bond in near-cognate initiation, the strong inhibitory effect of 2'-O-
506 methylation in A_mUG, which disrupts the hydrogen bond capacity of the 2'-OH group, and the less
507 significant effect of 2'-O-methylation in near-cognate codons, suggests that the interaction of 18S
508 with A[+1] strengthens the optimality of Kozak sequences.
509
510 **2'-O-methylation inhibits start codon recognition**
511
512 Possible explanations for the inhibitory effect of 2'-O-methylation in translation initiation include:
513 i) 2'-O-methylation prevents loading of the 40S subunit into mRNA; ii) 2'-O-methylation in position
514 A[+1] stalls or induces disassociation of the 48S preinitiation complex before or after recognizing
515 the start codon; iii) A_m[+1] stalls or induces disassociation of the 80S initiation complex; or iv) 2'-
516 O-methylation in position A[+1] makes the start codon non-optimal or unrecognizable by the
517 preinitiation complex.
518
519 To begin testing these possibilities, we generated uORF-reporter systems in which an upstream
520 uORFs is controlling the expression of NanoLuc (Supplementary Fig. S5A-C). We then
521 incorporated 2'-O-methylation at the upstream AUG, leaving the downstream AUG unmodified.
522 Expectedly, uORF inclusion decreased NanoLuc activity (Fig. 5A and B). Notably, we observed
523 that 2'-O-methylation in the upstream AUG codons exhibited significantly higher NanoLuc activity
524 compared to unmodified mRNA reporters, indicating that A_m[+1] inhibits recognition of the
525 upstream start codon and promotes downstream initiation (Fig. 5A and B). This observation
526 suggests that 2'-O-methylation does not block the 40S subunit from engaging mRNAs and does
527 not induce stalling/dissociation of the 48S or 80S complexes. Instead, the results of Fig. 5A and

528 B suggest that A_m[+1] hinders start codon recognition, allowing the 48S complex to initiate at a
529 downstream start codon.
530
531 To validate these findings, we incubated reticulocyte lysates with a Cy5-labeled mRNA reporter
532 in the presence of non-hydrolyzable GTP (GMPPNP) or harringtonine (HR) (Fig. 5C-E). Pre-
533 treatment with GMPPNP stalls the 48S preinitiation complex after codon recognition (Fig. 5D).
534 Relatedly, HR stalls the 80S initiation complex before proceeding into elongation (Fig. 5E).
535 Reactions were separated on a 5-30% sucrose density gradient followed by fluorescence (Cy5)
536 recording. As expected, treatment with GMPPNP increases the abundance of the 48S peak
537 compared to vehicle control (Fig. 5D, blue lines), whereas harringtonine treatment increases the
538 80S peak in the Cy5-labeled unmodified mRNA reporter (Fig. 5E, blue line). In contrast, A_m[+1]
539 decreases the formation of both the 48S peak and 80S peaks in the presence of GMPPNP or HR,
540 respectively (Fig. 5C-E, orange arrows).
541
542 Altogether, these findings support a model in which 2'-O-methylation at N[+1] reduces start codon
543 recognition and favors downstream initiation.
544
545 **2'-O-methylation is present in near-cognate codons within 5'UTRs**
546
547 Since N_m naturally occurs within mRNAs (17), our findings raise the possibility that 2'-O-
548 methylation is a regulatory mechanism of translation initiation in cellular mRNAs. Using Nanopore
549 direct RNA sequencing (DRS), members of this team previously identified N_m sites in HeLa and
550 C4-2 cell lines (Supplementary Fig. S6A) (17). We re-analyzed these Nanopore data sets to
551 determine whether 2'-O-methylation is present in canonical AUG start codons. To increase
552 confidence in the candidate N_m sites, only sites with an estimated modification ratio of more than
553 10% per position were included. Notably, no A_mUG sites in canonical start codons were identified
554 in either HeLa or C4-2 cell lines (Supplementary Fig. S6B).
555
556 We next extracted the N_m sites that mapped to 5'UTRs of any annotated Ensembl transcript and
557 examined whether these sites were located in position N[+1] of an upstream AUG or an upstream
558 near-cognate codon. This analysis identified 151 N_m[+1] sites in HeLa cells and 181 sites in C4-2
559 cells (Supplementary Fig. S6B and Supplementary Tables S3 and S4). These sites were located
560 in genes involved in biological pathways and molecular functions related to ribosome function and
561 translation (Supplementary Fig. S6C). Notably, 65 sites (~40%) were commonly observed in the

562 5'UTRs of HeLa and C4-2 cells, with only two A_mUG sites (Supplementary Fig. S6D and
563 Supplementary Table S5). The first A_mUG site mapped to the ribosome biogenesis factor GNL3,
564 although this site is polymorphic in the human population (rs1108842, A>C: ~50%) (6). For
565 instance, HeLa cells carry a CUG at this position (Supplementary Fig. S6E). The second AUG
566 site mapped to the ribosomal protein RPS3A. However, this site is located within a region that
567 could be part of the coding sequence or the 5'UTR, depending on the transcript isoform
568 (Supplementary Fig. S6F). These observations indicate that A_mUG is either absent or present at
569 very low stoichiometry in HeLa and C4-2 cells. Nevertheless, N_m sites were identified in near-
570 cognate codons within 5'UTRs (Supplementary Fig. S6B). CUG had the highest prevalence,
571 followed by AUC (Supplementary Fig. S6B), with C_mUG and A_mUC sites significantly enriched
572 relative to the distribution of these codons in 5'UTRs (Supplementary Fig. S6G).
573
574 To validate the presence of N_m[+1] in near-cognate codons, we employed Reverse Transcription
575 at Low dNTPs (RTL-P) (25). This method leverages the property of 2'-O-methylation to induce
576 reverse transcriptase stops (RT-STOP) at low concentrations of dNTPs (Fig. 6A), resulting in a
577 reduction in amplicon density after PCR amplification of the RT product (Supplementary Fig. S7A,
578 red arrowhead). While there are more than eleven human 2'O-methyltransferase enzymes (26),
579 fibrillarin (FBL) is consistently identified as the predominant enzyme responsible for introducing
580 N_m at internal sites in mRNA (17-20,27). Thus, we performed RTL-P in total RNA isolated from
581 HeLa cells transfected with two different siRNAs against Fibrillarin (FBL) or a siRNA control (Fig.
582 6B). Candidate sites were selected based on the highest predicted modification ratio from the
583 Nanopore-DRS approach in HeLa cells (Supplementary Table S6). Of the 12 candidate sites
584 tested, 10 exhibited a substantial loss in amplicon density under limited dNTP conditions, with
585 relative methylation values ranging from 30% to 99% in the validated sites (Fig. 6C-F and
586 Supplementary Fig. S7A-B). The RTL-P signature was sensitive to FBL knockdown (Fig. 6G-I and
587 Supplementary Fig. S7B), validating the presence of 2'-O-methylation sites in near-cognate
588 codons within human 5'UTRs.
589
590 To evaluate whether 2'-O-methylation in upstream near-cognate codons affects translation, we
591 analyzed publicly available Ribo-seq data from HeLa cells transduced with an inducible shRNA
592 against FBL. Translation efficiency (TE) was calculated for genes with N_m[+1] in upstream near-
593 cognate codons compared to genes with no 2'-O-methylation in the 5'UTR. The results suggest
594 that loss of N_m[+1] in upstream near-cognate codons, due to FBL knockdown, significantly
595 decreases translation efficiency compared to genes with no 2'-O-methylation in the 5'UTR

596 (Supplementary Fig. S7C and Supplementary Table S7). These findings are consistent with the
597 uORF reporter assays, where reporters with unmodified uORFs exhibited significantly less
598 NanoLuc activity compared to N_m[+1] in upstream start codons (Fig. 5A and B).
599
600 Lastly, we validated the effect of upstream 2'-O-methylation on canonical protein expression using
601 Western blots. We chose NEK2, GNL3, and SRSF1, which had N_m[+1] levels of ~99%, ~75%,
602 ~45%, and ~30%, respectively, in upstream CUG or AUC codons (Fig. 6C). These sites could
603 generate uORFs with inhibitory potential (Fig. 6G-I and Supplementary Fig. 7B, schematic), in
604 which loss of 2'-O-methylation, due to FBL knockdown, would increase upstream CUG or AUC
605 usage, leading to decreased canonical protein expression. NEK2 and SMARCE1 showed a
606 consistent decrease in canonical protein expression among the two different siRNAs, with a more
607 substantial effect in the siRNA-FBL-1 condition (Fig. 6J-K and Supplementary Fig. S7D). Notably,
608 the siRNA-FBL-1 condition exhibited a more pronounced FBL knockdown (Fig. 6B). The
609 consequence of FBL knockdown on GNL3 expression was weaker, with significant effects
610 observed only in the siRNA-FBL-1 condition (Fig. 6L and Supplementary Fig. S7D). In contrast,
611 SRSF1 protein levels were unchanged upon FBL knockdown (Supplementary Fig. S7D-E). These
612 findings indicate that endogenous 2'-O-methylation at upstream CUG or AUC codons significantly
613 affects downstream protein expression at high 2'-O-methylation levels, with a modest or
614 imperceptible effect at lower levels. Notably, the predicted stoichiometries of 2'-O-methylation in
615 upstream near-cognate codons from Nanopore DRS show a median of ~20% (Supplementary
616 Fig. S7F), explaining the modest, yet significant, effect of N_m[+1] loss on translation efficiency
617 transcriptome-wide in HeLa cells (Supplementary Fig. S7C).
618
619 Collectively, these findings confirmed N_m[+1] in upstream near-cognates in HeLa cells, where it
620 regulates the expression of downstream canonical proteins.
621
622 **DISCUSSION**
623
624 The results presented in this study suggest that 2'-O-methylation inhibits translation initiation by
625 interfering with start codon recognition, thereby allowing the scanning ribosome to initiate at
626 downstream start codons (graphical abstract). This model is supported by the presence of a
627 hydrogen bond between the 2'OH bond of AUG in mRNA and 18S rRNA (Fig. 1C), the absence
628 of this H-bond in near-cognate initiation (Supplementary Fig. S4E), the strong inhibitory effect of
629 2'-O-methylation in A_mUG (Fig. 2 and 3), a modification that disrupts the hydrogen bond capacity

630 of 2'-OH groups (28), and the less significant effect of 2'-O-methylation in near-cognate codons
631 (Fig. 4C-E). In addition, loss of N_m[+1] at upstream start codons decreased downstream
632 translation (Fig. 5A-B, Fig. 6J-K, and Supplementary Fig. S7C) and reduced GMPPNP- and HR-
633 trapped complexes (Fig. 5D-E), consistent with reduced start codon recognition at upstream
634 methylated sites.
635
636 Chemically, the 2'-OH group in RNA can act as both a donor and acceptor in H-bonds. Methylation
637 replaces the hydrogen atom of the 2'-OH with a $-CH_3$ group, effectively eliminating its donor
638 capability. Since the 2'-OH group of the A[+1] position can act as a donor in an H-bond with U1830,
639 either with the phosphate (PO_4^{3-}) backbone or the keto (C=O) group at C2, 2'-O-methylation is
640 likely to block such interaction. We speculate that the $A[+1]^{mRNA}$:$U1830^{18S}$ interaction helps 18S
641 rRNA monitoring for proper pairing and orientation of the codon:anticodon interaction. For
642 instance, the movement of the 40S subunit along the 5′ UTR is characterized by a conformational
643 intermediate between the closed state, in which the $tRNA_i^{met}$ is fully inserted into the P site, trying
644 to make an interaction with the mRNA codon, and the open state, in which the anticodon in
645 $tRNA_i^{met}$ is distant from the mRNA (23,29). This conformational intermediate enables the
646 preinitiation complex to proofread accommodation of optimal start codons (9,22,29-34). We
647 speculate that blocking the $A[+1]^{mRNA}$:$U1830^{18S}$ interaction by 2'-O-methylation precludes the
648 ribosome from proofreading start codon recognition. Alternatively, 2'-O-methylation can alter the
649 local structure of mRNA (35), thereby preventing it from properly pairing with the anticodon base.
650 Since 2'-O-methylation inhibits start codon recognition and promotes downstream initiation (Fig.
651 5), it likely promotes leaky scanning, allowing the 43S complex to continue scanning until it
652 encounters a downstream start codon. However, our findings do not rule out the possibility that
653 2'-O-methylation promotes reinitiation, rather than leaky scanning, or introduces internal ribosome
654 entry sites that enhance downstream initiation.
655
656 The inhibitory effect of 2'-O-methylation is less pronounced in near-cognate codons, likely
657 because they lack the proper pairing and orientation of the cognate codons. However, we
658 evaluated only the Kozak context most optimal for AUG initiation (ACCAUGG). We prioritized
659 CUG, GUG, AUC, and ACG based on HR-Ribo-seq prevalence (Fig. 4A), the A[+1] position, and
660 the prevalence of these codons in HeLa 5 "UTRs (Supplementary Fig. S6B), but focused on a
661 single Kozak context. Additional Kozak variants, near-cognate codons, and structured 5'UTRs
662 that favor non-AUG initiation remain to be tested to accurately rank the effects of 2'-O-methylation
663 in near-cognate initiation.

664

665 Notably, 2'-O-methylation of the start codon does not impair mRNA stability or globally affect
666 protein synthesis (Fig. 3C and D), indicating that its effect on translation initiation is transcript-
667 specific. The effect of 2'-O-methylation in translation initiation is also site-specific. 2'-O-
668 methylation of position N[-3], N[-2], N[-1], and N[+2] had no effect on protein output (Fig. 2E and
669 3B). An unexpected finding was the lack of effect at position N[-1], despite our previous report of
670 a significant role for the $C[-1]^{mRNA}$:t^6A^{tRNA} interaction in stabilizing the initiation complex (8).
671 However, 2'-O-methylation does not alter the acceptor capacity of the C-O-CH3 group, and thus,
672 the $C[-1]^{mRNA}$:t^6A^{tRNA} hydrogen bonds may still occur with N_m[-1]. 2'-O-methylation in positions
673 N[+3] and N[+4] reduced protein output, albeit to a much lesser extent than N[+1] (Fig. 2E and
674 3B). Alike N[-1], the 2'-OH group in N[+3] is an acceptor (Supplementary Table S2), rather than a
675 donor, and thus, 2'-O-methylation would not disrupt its hydrogen bond. The effect of N_m[+4] on
676 protein synthesis could be mediated through impairing elongation as the ribosome moves into the
677 second and third codons. In support of the latter, previous studies indicated that 2'-O-methylation
678 represses the translocation step during translation elongation by precisely inhibiting the
679 interaction of mRNA with the monitoring bases of rRNA during the proofreading step (27,28).
680

681 It is well established that 2′-O-methylation is highly abundant in noncoding RNAs, including rRNA
682 and tRNA, and occurs in the 5′ cap of all mRNAs in human transcriptomes (26). More recently,
683 the presence of N_m at internal mRNA has been reported in multiple studies (17-20,36-38). By re-
684 analyzing published Nanopore DRS datasets (17), we report the absence of 2'-O-methylation in
685 position A[+1] of canonical and upstream AUGs in HeLa and C4-2 cells under normal laboratory
686 growth conditions. However, our analysis does not conclusively rule out the possibility that AUG
687 codons are methylated in human transcriptomes. A more comprehensive study using additional
688 mapping methods such as RiboMeth-seq (39), N_m-seq (20), RibOxi-seq (37), NJU-seq (19), N_m-
689 mut-seq (18), or 2OMe-seq (36) in multiple cell types and tissues across different conditions would
690 help determine whether A_mUG codons exist in human transcriptomes.
691

692 We validated the presence of N_m sites in near-cognate codons within the 5'UTRs of HeLa mRNAs
693 (Fig. 6 and Supplementary Fig. S7) and observed that loss of upstream N_m, due to FBL
694 knockdown, reduces translation of downstream proteins, especially at high-stoichiometry sites
695 (Fig. 6J-K and Supplementary Fig. S7C). These observations suggest that cells can leverage 2'-
696 O-methylation to control non-AUG translation. However, FBL also catalyzes 2'-O-methylation of
697 ribosomal RNA, other noncoding RNAs, and the coding sequences of mRNAs, exerting pleiotropic

effects that could confound or mask the effect of FBL knockdown at low stoichiometric sites in 5'UTRs. Moreover, our findings do not preclude the possibility that near-cognate 2'-O-methylation plays a role under stress conditions when canonical cap-dependent translation is repressed, allowing non-AUG translation to serve as a stress response mechanism. In this manner, cells could block unwanted non-canonical initiation sites through 2'-O-methylation under specific cellular conditions.

The finding that 2'-O-methylation of upstream start codons inhibits uORF usage while enhancing canonical protein expression in transcript-specific manners has a significant impact on RNA therapeutics. Aberrant translation initiation has been reported in numerous diseases, with hundreds of pathogenic genetic variants in 5'UTRs linked to the introduction or destruction of uORFs (5,40-43). These pathological conditions would benefit from strategies that inhibit aberrant initiation sites while enhancing expression of the canonical ORF. Since 2'-O-methylation can be targeted to specific positions within transcripts (44-46), our findings open the door to investigating approaches that can guide 2'-O-methylation for the targeted inhibition of pathogenic translation initiation in human disease. Moreover, 2'-O-methylation could be incorporated into the 5'UTRs of synthetic mRNA medicines to enhance protein expression.

In sum, we identified a chemical modification of mRNA, 2'-O-Methylation, that unambiguously inhibits start codon recognition in a position-specific manner, a finding that could be leveraged to inhibit aberrant uORF expression or enhance downstream ORF expression for RNA therapeutics.

ACKNOWLEDGMENTS

We thank the Flow Cytometry Core Facility at the Robert H. Lurie Comprehensive Cancer Center for flow cytometry services, and the Quest High-Performance Computing Facility at Northwestern University for computational resources. We thank all members of the Arango lab for insightful discussions.

AUTHOR CONTRIBUTIONS

Adam Suh: Investigation, Methodology, Validation, Formal Analysis, Writing – review & editing. Stephanie Mou: Investigation, Methodology, Validation, Formal Analysis, Writing – review & editing. Emmely A. Patrasso: Investigation, Methodology, Validation, Formal Analysis, Writing –

review & editing. Hannah Serio: Formal analysis, Data curation, Visualization, Writing – review & editing. Kevin Vasquez: Investigation. Rui Wang: Investigation. Smriti Sangwan: Formal Analysis, Visualization. Yang Yi: Investigation, Validation, Formal Analysis. Daniel Arango: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

SUPPLEMENTARY DATA

Supplementary Figure S1. H-bonds between the 2'-OH groups of mRNA nucleotides with $tRNA_i^{met}$ and 18S rRNA

Supplementary Figure S2. Quality controls for the construction of NanoLuc mRNA reporters

Supplementary Figure S3. Quality controls for the construction of GFP mRNA reporters

Supplementary Figure S4. Quality controls for the construction of mRNA reporters with near-cognate codons

Supplementary Figure S5. Quality controls for the construction of uORF reporters

Supplementary Figure S6. Distribution of 2'-O-methylation sites in human mRNAs

Supplementary Figure S7. Validation of 2'-O-methylation sites in near-cognate codons

Supplementary Table S1: List of oligonucleotides used in this study

Supplementary Table S2. Hydrogen bond measurements

Supplementary Table S3. List of upstream N_m[+1] sites in HeLa cells

Supplementary Table S4. List of upstream N_m[+1] sites in C4-2 cells

Supplementary Table S5. List of upstream N_m[+1] sites common in both HeLa and C4-2 cells

Supplementary Table S6. List of genes used for biochemical validation of upstream N_m[+1] sites

Supplementary Table S7. Translation efficiency

Supplementary Video S1: Electron densities for PDB:7UCJ and centered around the interaction of the 2'-OH group in nucleotide $C[-1]^{mRNA}$ with t^6A in $tRNA_i^{met}$.

Supplementary Video S2: Electron densities for PDB: 6YAL and centered around the interaction of the 2'-OH group in nucleotide $C[-1]^{mRNA}$ with t^6A in $tRNA_i^{met}$.

Supplementary Video S3: Electron densities for PDB: 7UCJ and 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.

Supplementary Video S4: Electron densities for PDB: 8PJ5 and 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.

Supplementary Video S5: Electron densities for PDB: 8PJ3 and centered around the interaction of the 2'-OH group in nucleotide A[-3]^{mRNA} with G961 in 18S rRNA.

Supplementary Video S5: Electron densities for PDB: 8PJ3 and centered around the interaction of the 2'-OH group in nucleotide G[+3]^{mRNA} with lysin 7 in eIF1A.

COMPETING INTEREST STATEMENT

Nothing to disclose

FUNDING

This study was supported by the National Institute of General Medical Sciences from the National Institutes of Health [grant number R35GM159598 to D.A and R00GM143527 to S.S.]; the Department of Defense Congressionally Directed Medical Research Programs [grant numbers HT9425-24-1-0352 to D.A. and HT9425-23-1-0661 to Y.Y]; the Kinship Foundation through the Searle Scholars Program (Grant number SSP-2023-102 to D.A.]; the Robert H. Lurie Comprehensive Cancer Center through the Lefkofsky Innovation Research Award to D.A.; the Northwestern University's Biotechnology Training Program [grant number T32GM008449 to K.V.]; and the Elsa U. Pardee Foundation [grant number 21891 to Y.Y.].

DATA AVAILABILITY

This study did not generate original sequencing or structural data. Publicly available data used in this study include GEO: GSE162043 (8), GEO: GSE208837 (17), GEO: GSE105248 (21), PDB: 7UCJ (8), PDB: 6YAL (9), PDB: 8PJ1 (22), PDB: 8PJ2 (22), PDB: 8PJ3 (22), PDB: 8PJ4 (22), PDB: 8PJ5 (22), PDB: 8PJ6 (22).

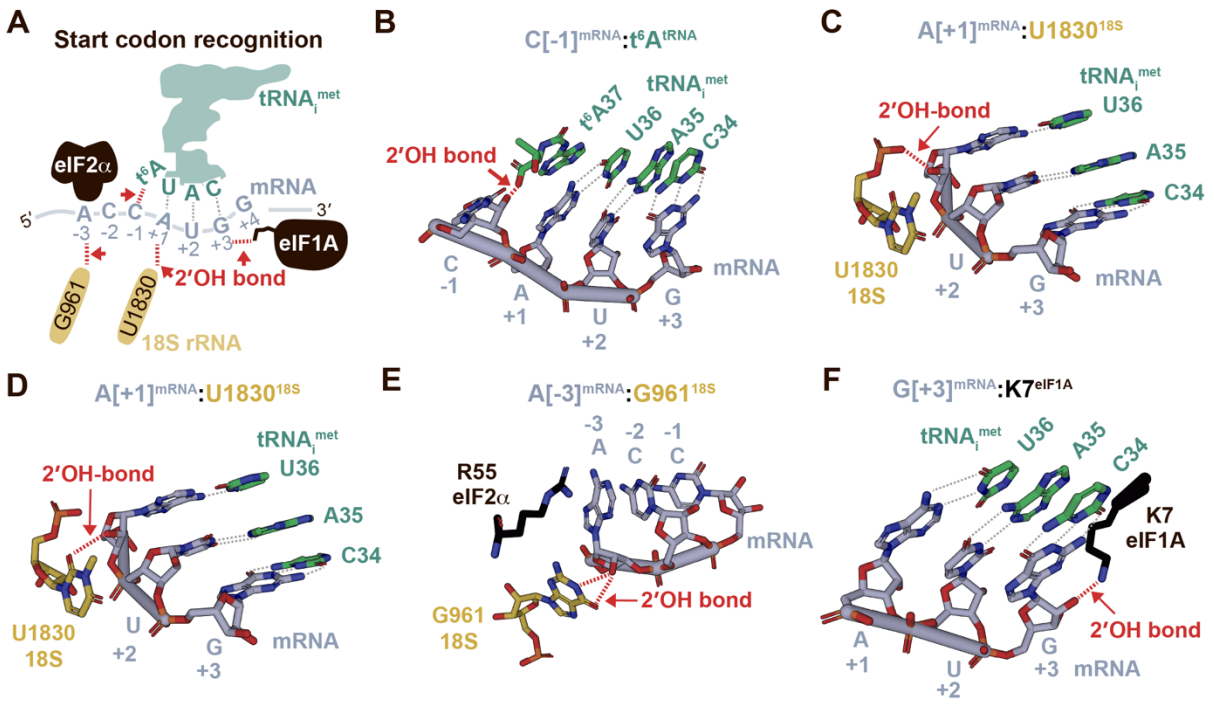
REFERENCES

1. Brito Querido, J., Diaz-Lopez, I. and Ramakrishnan, V. (2024) The molecular basis of translation initiation and its regulation in eukaryotes. *Nat Rev Mol Cell Biol*, **25**, 168-186.
2. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283-292.
3. Kearse, M.G. and Wilusz, J.E. (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev*, **31**, 1717-1731.
4. Kozak, M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1-34.

5. Lee, D.S.M., Park, J., Kromer, A., Baras, A., Rader, D.J., Ritchie, M.D., Ghanem, L.R. and Barash, Y. (2021) Disrupting upstream translation in mRNAs is associated with human disease. *Nat Commun*, **12**, 1515.
6. Calvo, S.E., Pagliarini, D.J. and Mootha, V.K. (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A*, **106**, 7507-7512.
7. Dasgupta, A. and Prensner, J.R. (2024) Upstream open reading frames: new players in the landscape of cancer gene regulation. *NAR Cancer*, **6**, zcae023.
8. Arango, D., Sturgill, D., Yang, R., Kanai, T., Bauer, P., Roy, J., Wang, Z., Hosogane, M., Schiffers, S. and Oberdoerffer, S. (2022) Direct epitranscriptomic regulation of mammalian translation initiation through N4-acetylcytidine. *Mol Cell*, **82**, 2797-2814 e2711.
9. Simonetti, A., Guca, E., Bochler, A., Kuhn, L. and Hashem, Y. (2020) Structural Insights into the Mammalian Late-Stage Initiation Complexes. *Cell Rep*, **31**, 107497.
10. Galloway, A. and Cowling, V.H. (2019) mRNA cap regulation in mammalian cell function and fate. *Biochim Biophys Acta Gene Regul Mech*, **1862**, 270-279.
11. Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.J., Chen, Q. *et al.* (2017) Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. *Nature*, **541**, 371-375.
12. Akichika, S., Hirano, S., Shichino, Y., Suzuki, T., Nishimasu, H., Ishitani, R., Sugita, A., Hirose, Y., Iwasaki, S., Nureki, O. *et al.* (2019) Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science*, **363**.
13. Guca, E., Alarcon, R., Palo, M.Z., Santos, L., Alonso-Gil, S., Davyt, M., de Lima, L.H.F., Boissier, F., Das, S., Zagrovic, B. *et al.* (2024) N(6)-methyladenosine in 5' UTR does not promote translation initiation. *Mol Cell*, **84**, 584-595 e586.
14. Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R. and Qian, S.B. (2015) Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature*, **526**, 591-594.
15. Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.B. and Jaffrey, S.R. (2015) 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell*, **163**, 999-1010.
16. Beiki, H., Sturgill, D., Arango, D., Relier, S., Schiffers, S. and Oberdoerffer, S. (2024) Detection of ac4C in human mRNA is preserved upon data reassessment. *Mol Cell*, **84**, 1611-1625 e1613.
17. Li, Y., Yi, Y., Gao, X., Wang, X., Zhao, D., Wang, R., Zhang, L.S., Gao, B., Zhang, Y., Zhang, L. *et al.* (2024) 2'-O-methylation at internal sites on mRNA promotes mRNA stability. *Mol Cell*, **84**, 2320-2336 e2326.
18. Chen, L., Zhang, L.S., Ye, C., Zhou, H., Liu, B., Gao, B., Deng, Z., Zhao, C., He, C. and Dickinson, B.C. (2023) Nm-Mut-seq: a base-resolution quantitative method for mapping transcriptome-wide 2'-O-methylation. *Cell Res*, **33**, 727-730.
19. Tang, Y., Wu, Y., Wang, S., Lu, X., Gu, X., Li, Y., Yang, F., Xu, R., Wang, T., Jiao, Z. *et al.* (2024) An integrative platform for detection of RNA 2'-O-methylation reveals its broad distribution on mRNA. *Cell Rep Methods*, **4**, 100721.

20. Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., Dominissini, D. and He, C. (2017) Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat Methods*, **14**, 695-698.
21. Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S.E., Garcia, M., Laforets, F., Marcel, V., Baudin-Baillieu, A. *et al.* (2017) Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc Natl Acad Sci U S A*, **114**, 12934-12939.
22. Petrychenko, V., Yi, S.H., Liedtke, D., Peng, B.Z., Rodnina, M.V. and Fischer, N. (2025) Structural basis for translational control by the human 48S initiation complex. *Nat Struct Mol Biol*, **32**, 62-72.
23. Brito Querido, J., Sokabe, M., Kraatz, S., Gordiyenko, Y., Skehel, J.M., Fraser, C.S. and Ramakrishnan, V. (2020) Structure of a human 48S translational initiation complex. *Science*, **369**, 1220-1227.
24. Hertler, J., Slama, K., Schober, B., Ozrendeci, Z., Marchand, V., Motorin, Y. and Helm, M. (2022) Synthesis of point-modified mRNA. *Nucleic Acids Res*, **50**, e115.
25. Dong, Z.W., Shao, P., Diao, L.T., Zhou, H., Yu, C.H. and Qu, L.H. (2012) RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res*, **40**, e157.
26. Ayadi, L., Galvanin, A., Pichot, F., Marchand, V. and Motorin, Y. (2019) RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions. *Biochim Biophys Acta Gene Regul Mech*, **1862**, 253-269.
27. Elliott, B.A., Ho, H.T., Ranganathan, S.V., Vangaveti, S., Ilkayeva, O., Abou Assi, H., Choi, A.K., Agris, P.F. and Holley, C.L. (2019) Modification of messenger RNA by 2'-O-methylation regulates gene expression in vivo. *Nat Commun*, **10**, 3401.
28. Choi, J., Indrisiunaite, G., DeMirci, H., Ieong, K.W., Wang, J., Petrov, A., Prabhakar, A., Rechavi, G., Dominissini, D., He, C. *et al.* (2018) 2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat Struct Mol Biol*, **25**, 208-216.
29. Llacer, J.L., Hussain, T., Marler, L., Aitken, C.E., Thakur, A., Lorsch, J.R., Hinnebusch, A.G. and Ramakrishnan, V. (2015) Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol Cell*, **59**, 399-412.
30. Yi, S.H., Petrychenko, V., Schliep, J.E., Goyal, A., Linden, A., Chari, A., Urlaub, H., Stark, H., Rodnina, M.V., Adio, S. *et al.* (2022) Conformational rearrangements upon start codon recognition in human 48S translation initiation complex. *Nucleic Acids Res*, **50**, 5282-5298.
31. Zhou, F., Zhang, H., Kulkarni, S.D., Lorsch, J.R. and Hinnebusch, A.G. (2020) eIF1 discriminates against suboptimal initiation sites to prevent excessive uORF translation genome-wide. *RNA*, **26**, 419-438.
32. Thakur, A. and Hinnebusch, A.G. (2018) eIF1 Loop 2 interactions with Met-tRNA(i) control the accuracy of start codon selection by the scanning preinitiation complex. *Proc Natl Acad Sci U S A*, **115**, E4159-E4168.
33. Thakur, A., Marler, L. and Hinnebusch, A.G. (2019) A network of eIF2beta interactions with eIF1 and Met-tRNAi promotes accurate start codon selection by the translation preinitiation complex. *Nucleic Acids Res*, **47**, 2574-2593.

887 34. Hussain, T., Llacer, J.L., Fernandez, I.S., Munoz, A., Martin-Marcos, P., Savva, C.G.,
888 Lorsch, J.R., Hinnebusch, A.G. and Ramakrishnan, V. (2014) Structural changes
889 enable start codon recognition by the eukaryotic translation initiation complex. *Cell*,
890 **159**, 597-607.
891 35. Abou Assi, H., Rangadurai, A.K., Shi, H., Liu, B., Clay, M.C., Erharter, K., Kreutz, C.,
892 Holley, C.L. and Al-Hashimi, H.M. (2020) 2'-O-Methylation can increase the
893 abundance and lifetime of alternative RNA conformational states. *Nucleic Acids*
894 *Res*, **48**, 12365-12379.
895 36. Incarnato, D., Anselmi, F., Morandi, E., Neri, F., Maldotti, M., Rapelli, S., Parlato, C.,
896 Basile, G. and Oliviero, S. (2017) High-throughput single-base resolution mapping of
897 RNA 2′-O-methylated residues. *Nucleic Acids Res*, **45**, 1433-1441.
898 37. Ye, X., Liu, Y., Zhu, Y., Alshawi, S.A., Elliott, B.A., Holley, C.L., Beaudoin, J.D.,
899 Graveley, B.R. and Carmichael, G.G. (2025) Genome-wide profiling of RNA 2'- O -
900 methylation in neurons and identification of orphan snoRNA targets. *bioRxiv*.
901 38. Zhu, Y., Pirnie, S.P. and Carmichael, G.G. (2017) High-throughput and site-specific
902 identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-
903 seq). *RNA*, **23**, 1303-1314.
904 39. Marchand, V., Blanloeil-Oillo, F., Helm, M. and Motorin, Y. (2016) Illumina-based
905 RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res*,
906 **44**, e135.
907 40. Filatova, A., Reveguk, I., Piatkova, M., Bessonova, D., Kuziakova, O., Demakova, V.,
908 Romanishin, A., Fishman, V., Imanmalik, Y., Chekanov, N. *et al.* (2023) Annotation of
909 uORFs in the OMIM genes allows to reveal pathogenic variants in 5'UTRs. *Nucleic*
910 *Acids Res*, **51**, 1229-1244.
911 41. Soukarieh, O., Meguerditchian, C., Proust, C., Aissi, D., Eyries, M., Goyenvalle, A.
912 and Tregouet, D.A. (2022) Common and Rare 5'UTR Variants Altering Upstream
913 Open Reading Frames in Cardiovascular Genomics. *Front Cardiovasc Med*, **9**,
914 841032.
915 42. Whiffin, N., Karczewski, K.J., Zhang, X., Chothani, S., Smith, M.J., Evans, D.G.,
916 Roberts, A.M., Quaife, N.M., Schafer, S., Rackham, O. *et al.* (2020) Characterising
917 the loss-of-function impact of 5' untranslated region variants in 15,708 individuals.
918 *Nat Commun*, **11**, 2523.
919 43. Martin-Geary, A.C., Blakes, A.J.M., Dawes, R., Findlay, S.D., Lord, J., Dong, S.,
920 Walker, S., Talbot-Martin, J., Wieder, N., D'Souza, E.N. *et al.* (2025) Systematic
921 identification of disease-causing promoter and untranslated region variants in 8040
922 undiagnosed individuals with rare disease. *Genome Med*, **17**, 40.
923 44. Zhao, J., Cockman, E., Yin, X., Chao, Y., Pecot, C. and Holley, C.L. (2024) Generating
924 snoRNA-guided Programmable 2'- O -methylation. *bioRxiv*.
925 45. Huang, C. and Yu, Y.T. (2010) Targeted 2'-O methylation at a nucleotide within the
926 pseudoknot of telomerase RNA reduces telomerase activity in vivo. *Mol Cell Biol*,
927 **30**, 4368-4378.
928 46. Zhao, X. and Yu, Y.T. (2008) Targeted pre-mRNA modification for gene silencing and
929 regulation. *Nat Methods*, **5**, 95-100.



930

931 **Figure 1.** 2'-OH hydrogen bonds within translation initiation complexes. **(A)** schematic
932 representation of the intermolecular interactions within translation preinitiation complexes. (**B-D)**
933 Cryo-EM visualization of rabbit 80S initiation ribosomes captured in complexed with harringtonine
934 (PDB:7UCJ, Arango et al., 2022) (8) and centered around the interaction of the 2'-OH group in
935 nucleotide C[-1]^{mRNA} with t^6A in $tRNA_i^{met}$ (B); centered around the interaction of the 2'-OH group
936 in nucleotide A[+1]^{mRNA} with the PO_4^{3-} group of U1830 in 18S rRNA (C); or centered around the
937 interaction of the 2'-OH group in nucleotide A[+1]^{mRNA} with the C[2]=O group of U1830 in 18S
938 rRNA (D). (**E-F)** cryo-EM visualization of human 48S preinitiation complexes (PDB:8PJ3
939 Petrychenko et al., 2025) (22) centered around the interaction of the 2'-OH group in nucleotide
940 A[-3]^{mRNA} with G961 in 18S rRNA (E); or centered around the interaction of the 2'-OH group in
941 nucleotide G[+3]^{mRNA} with lysin 7 in eIF1A (F). All structures were downloaded from PDB and
942 visualized using PyMOL.

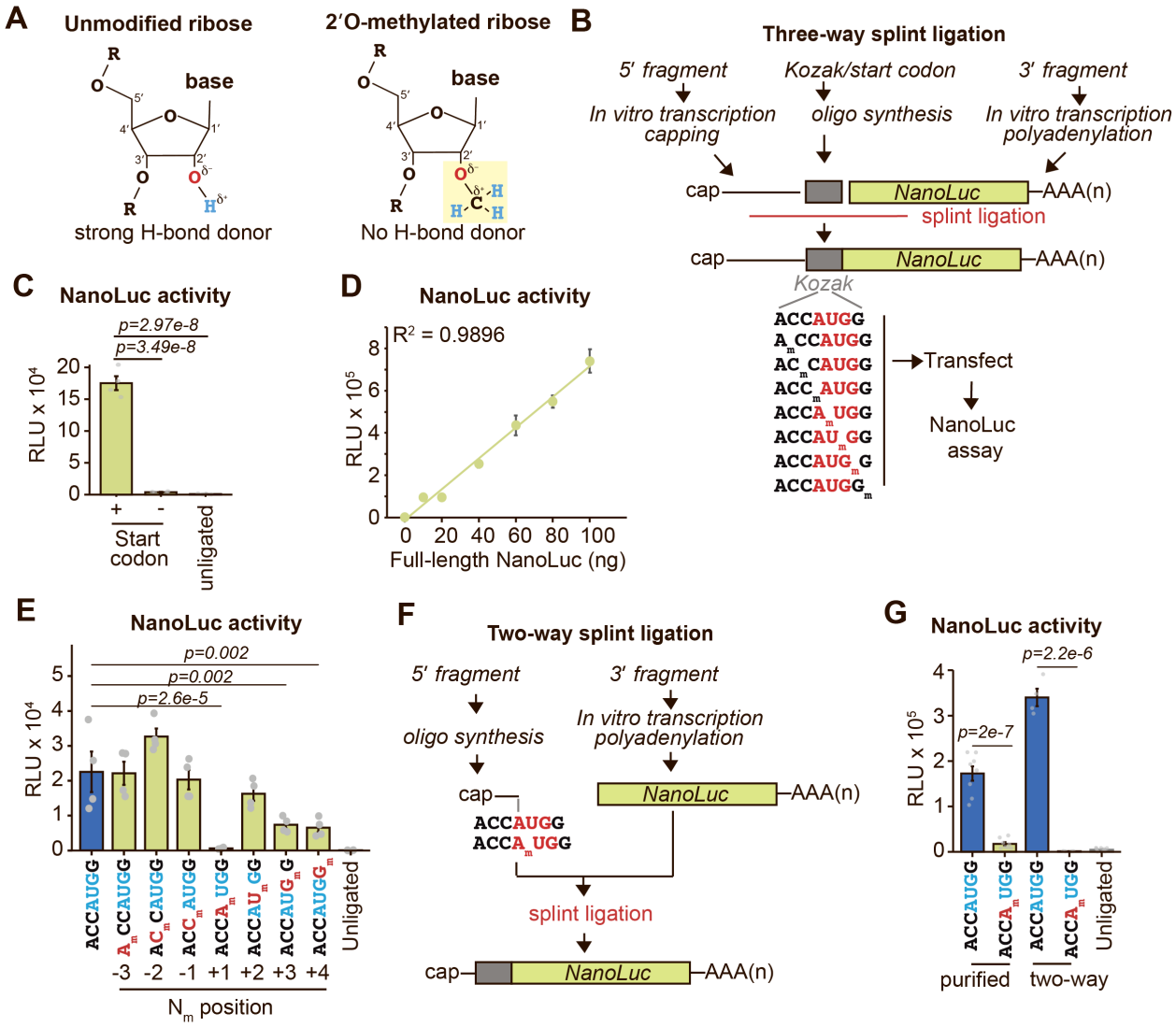


Figure 2. 2'-O-methylation of A[+1] within AUG start codons inhibits protein synthesis. **(A)** chemical structures of unmethylated and 2'-O-methylated nucleosides. (**B)** experimental design of a three-way splint ligation approach to insert 2'-O-methylation in single positions within mRNA start codons. **(C)** relative light units (RLUs) 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 a control. Mean ± SEM, n = 3. p = One-way ANOVA with Tukey's post hoc analysis. **(D)** RLUs 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. **(E)** RLUs were measured in lysates from HEK293T 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. **(F)** schematic of the two-way splint ligation approach. **(G)** RLUs were measured in lysates from HEK293T cells transfected with ligated and gel-purified NanoLuc mRNAs (n = 8) or the mRNA reporter produced by two-way-splint ligation (n = 4). Both constructs carried 2'-O-methylation at position A[+1] of the AUG start codon. Mean ± SEM. p = Two-tailed Student's t-test.

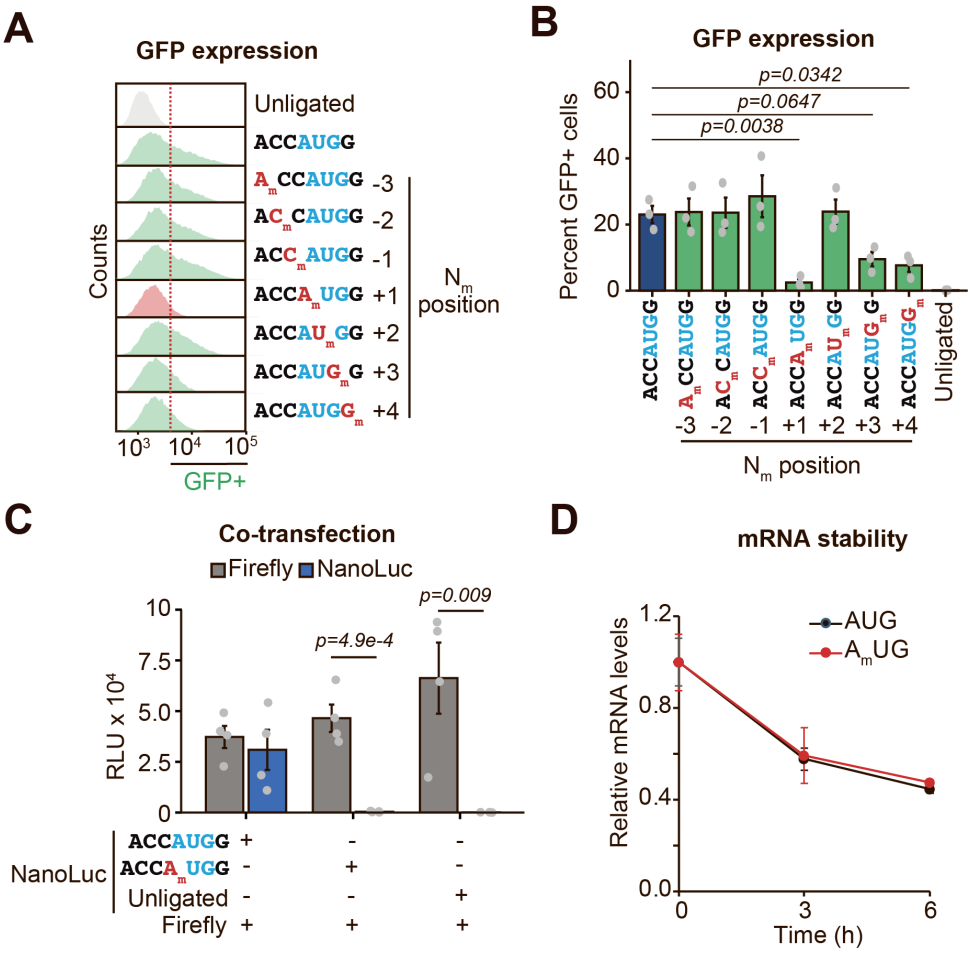


Figure 3. 2'-O-methylation of A[+1] within start codons inhibits translation in a transcript-specific manner. **(A)** representative histograms of fluorescence intensity measured by flow cytometry in HEK293T cells transfected with ligated GFP mRNA reporters carrying 2'-O-methylation in different positions around the start codon. **(B)** quantification of data from (A). 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 unmodified mRNAs encoding Firefly luciferase and unmodified or A_m[+1]-modified mRNAs encoding NanoLuc. Mean ± SEM, n = 4. p = Two-Tailed Student's *t-test*. **(D)** NanoLuc mRNA levels were measured using RT-qPCR at different time intervals post-transfection. Primers were designed to detect only ligated mRNA. 18S rRNA was used as a normalizing control. Mean ± SEM, n = 3.

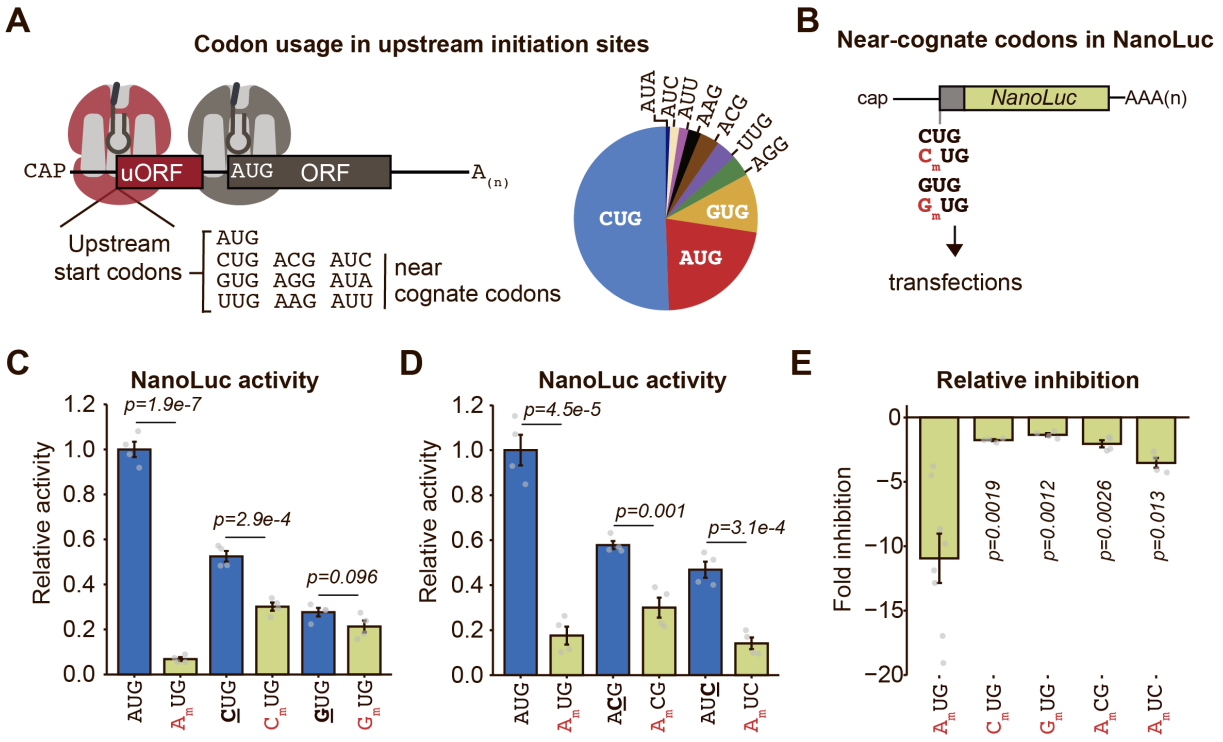


Figure 4. Inhibition of non-AUG translation by 2'-O-methylation. **(A)** distribution of upstream initiation codons obtained from HR-Ribo-seq (Arango et al., 2022) (8) **(B)** schematic of mRNA reporters with CUG and GUG initiation codons. **(C-D)** RLUs were measured in lysates from HEK293T cells transfected with gel-purified mRNA reporters carrying 2'-O-methylation in position N[+1] or left unmodified. RLUs were normalized relative to AUG. Mean ± SEM, n = 4. p = two-tailed *Student's t-test*. **(E)** the relative inhibition effect from (C-D) was calculated as the fold change between the methylated and unmethylated conditions. p = One-way ANOVA with Tukey's post hoc analysis.

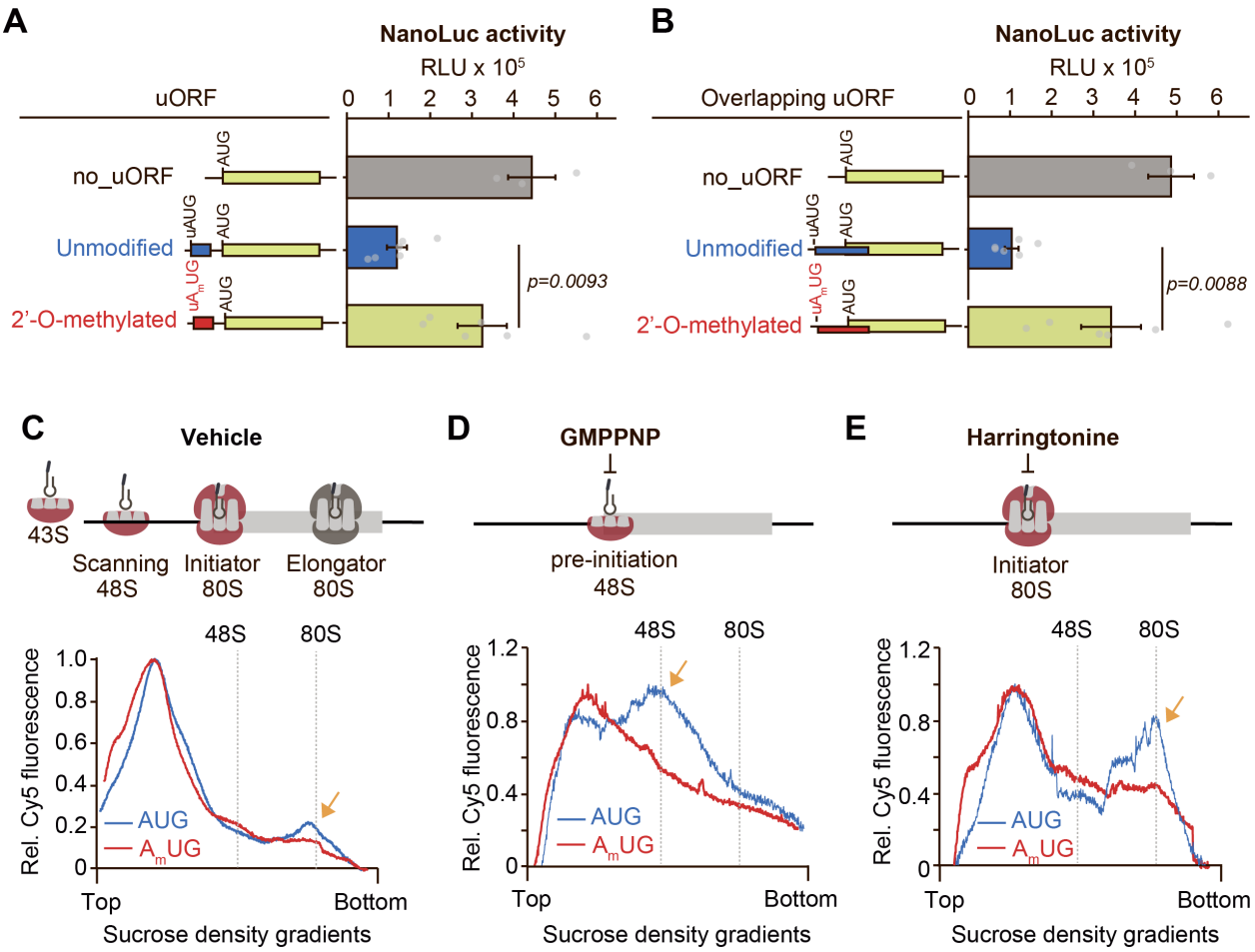


Figure 5. 2'-O-methylation inhibits start codon recognition. **(A-B)** mRNA reporters carrying 2'-O-methylated or unmethylated uORFs were generated by two-way splint ligation and transfected into HEK293T cells. RLUs were measured six hours after transfection. Mean ± SEM, n = 4. p = two-tailed *Student's t-test*. **(C-E),** *top:* schematic representation of translation initiation (C), 48S inhibition by GMPPNP (D), and 80S inhibition by harringtonine (E). *bottom:* Cy5-labeled synthetic mRNAs containing unmodified or A_mUG-modified start codons were incubated with untreated reticulocyte lysates (C), treated with GMPPNP (D), or treated with harringtonine (E). Reactions were separated on a sucrose density gradient followed by fluorescence (Cy5) recording.

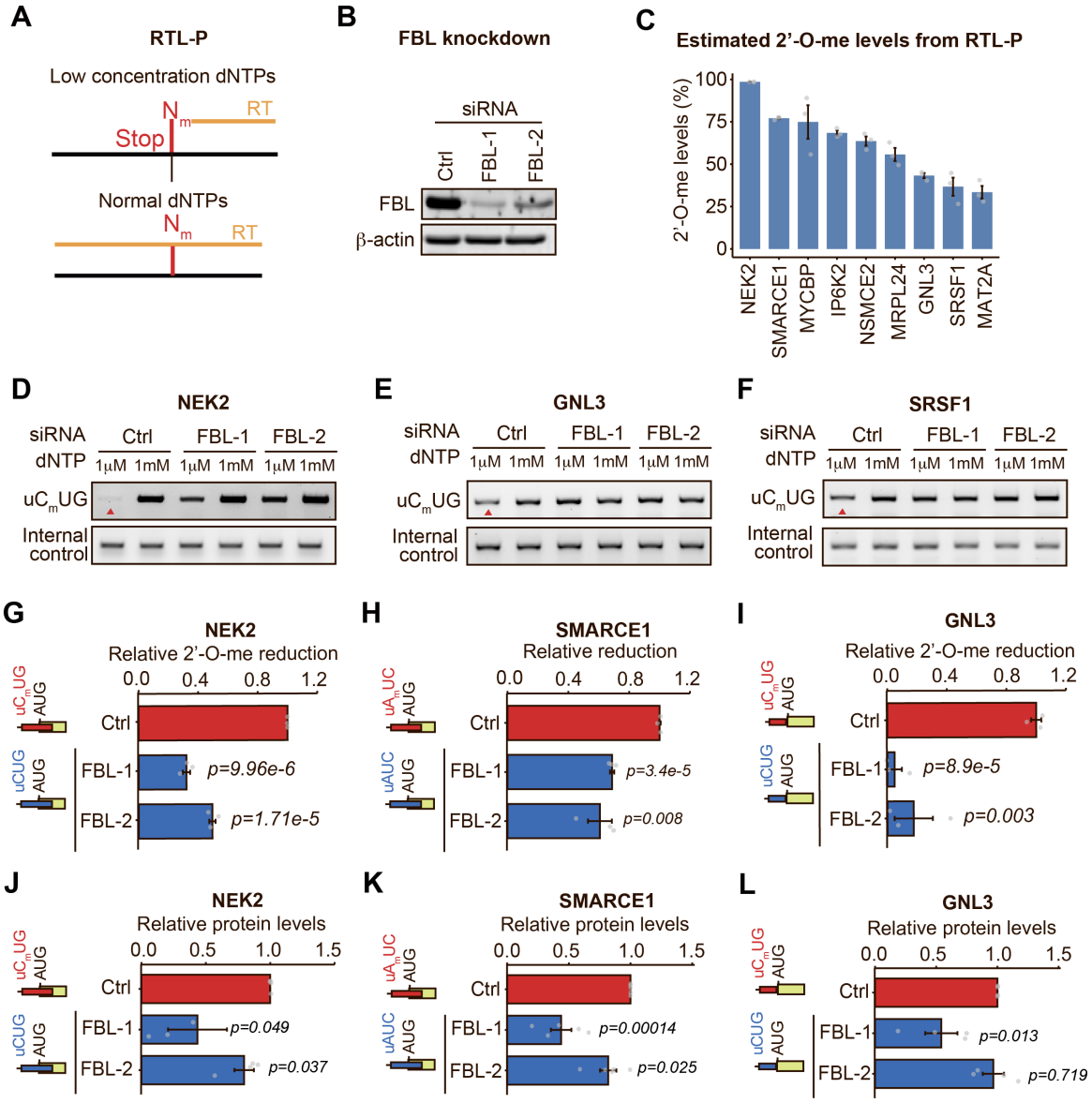
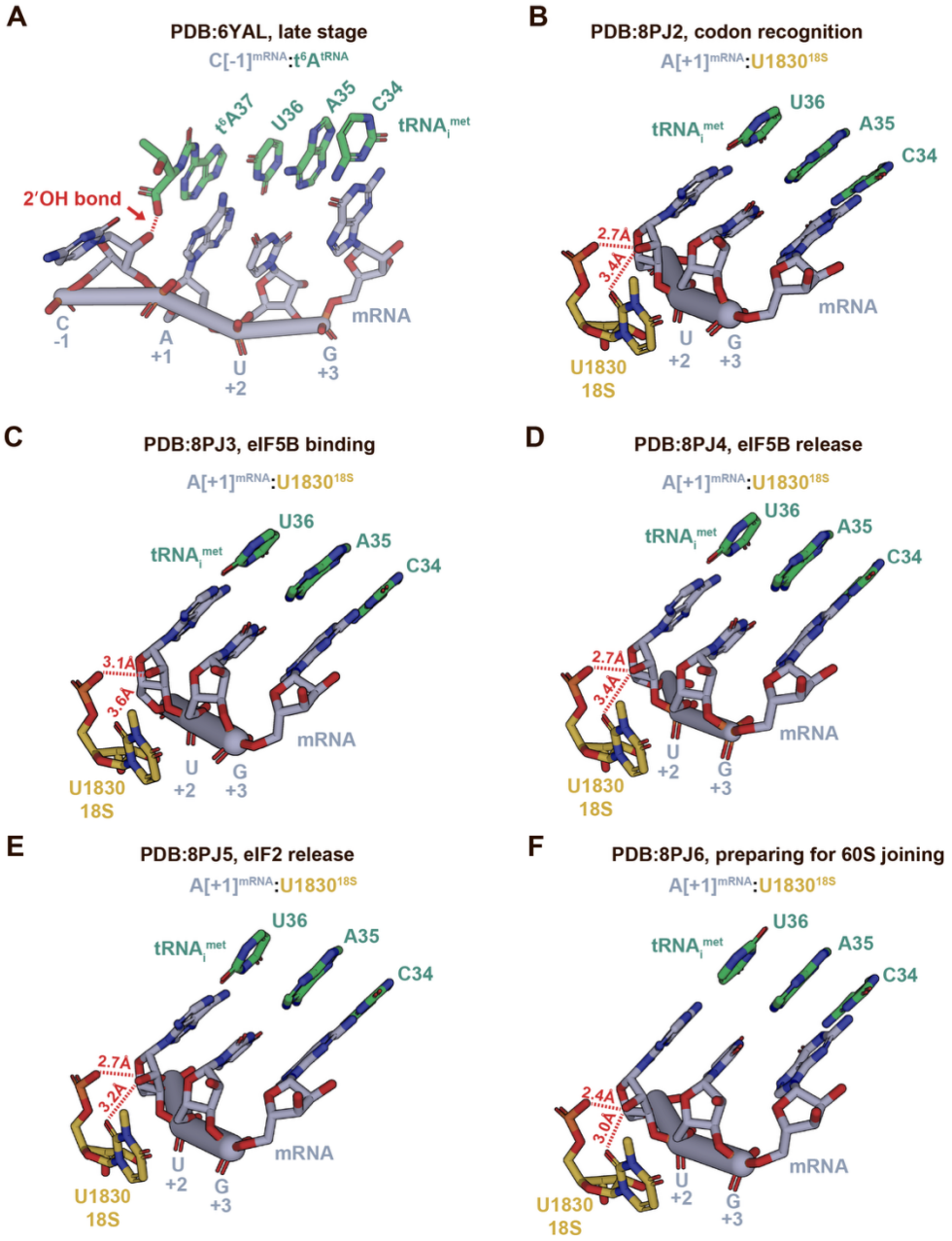


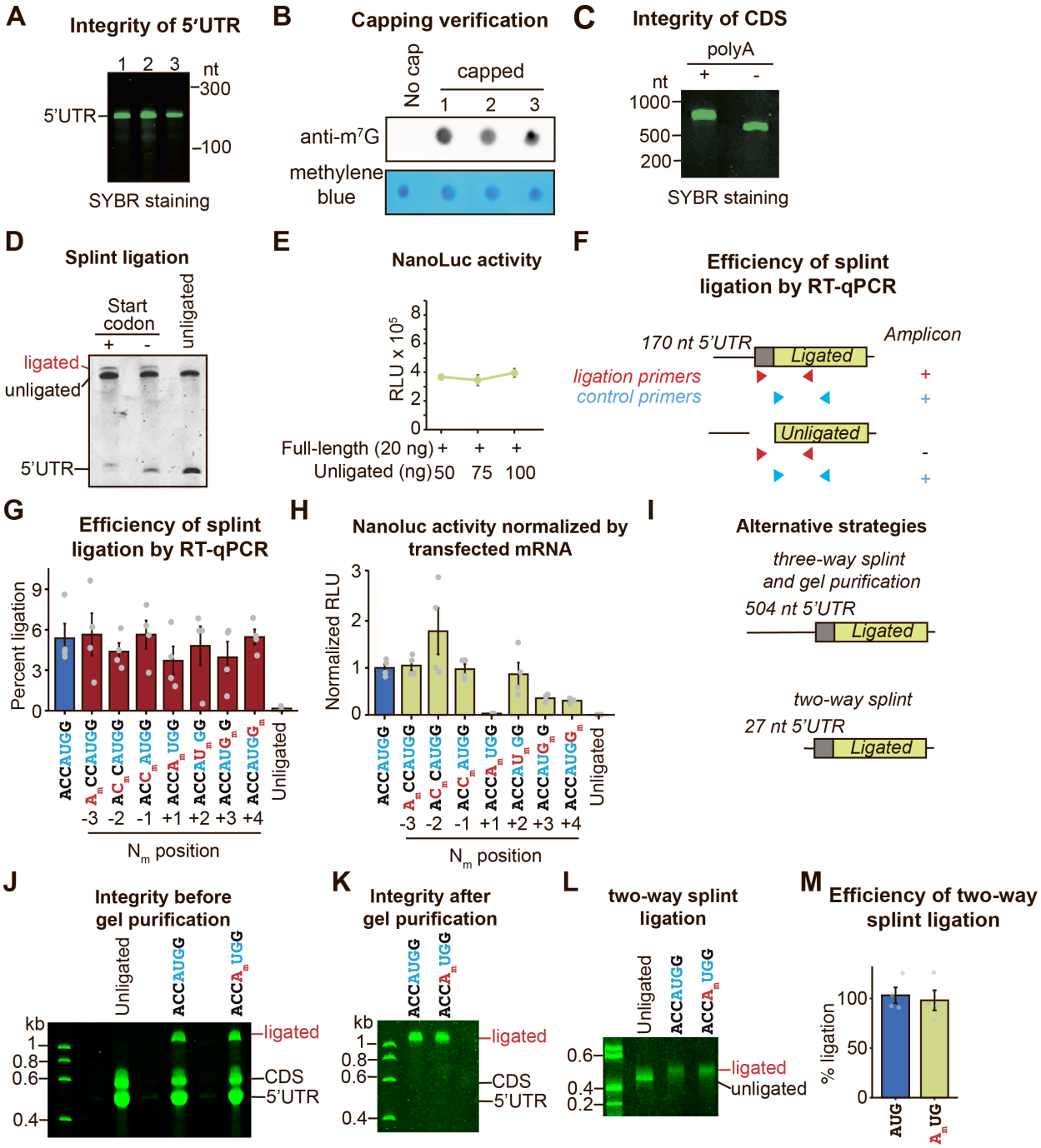
Figure 6. Identification of 2'-O-methylation in start codons of cellular mRNAs. **(A)** schematic of RTL-P. Reverse transcription was performed in the presence of an optimal dNTP concentration (1 mM) or a suboptimal dNTP concentration (1μM) that induces a premature RT-STOP at 2'-O-methylation sites. A pair of primers surrounding the 2'-O-methylation is used during PCR amplification. In this manner, the amplicon signal is an indicator of RT-STOP efficiency and N_m stoichiometry. Another pair of primers, located downstream of the 2'-O-methylation site, serves as an internal control for mRNA levels. **(B)** Western blot validation of FBL knockdown using two different siRNAs in HeLa cells. **(C)** quantification of the relative N_m ratio from the gels shown in Fig. 6D-F and Supplementary Fig. 7A. Mean ± SEM, n = 3. **(D-F)** representative agarose gels of amplicon density for the 2'-O-methylated site (top) and the loading control (bottom). Arrowheads indicate the reduction in amplicon signal when reverse transcription is performed at low dNTP concentration in the presence of FBL. **(G-I)** quantification of the relative N_m ratio dependent on FBL from the agarose gels shown in (D-F). Mean ± SEM, n = 3. p = two-tailed *Student's t-test*. (**J-L),** quantification of the relative protein levels dependent on FBL knockdown from the gels shown in Supplementary Fig. S7D. Mean ± SEM, n = 4-5. p = two-tailed *Student's t-test*.

1006 **Supplementary Figure S1.**



1007 **Supplementary Figure S1.** H-bonds between the 2'-OH groups of mRNA nucleotides with
1008 $tRNA_i^{met}$ and 18S rRNA. **(A)** Cryo-EM visualization of late-stage rabbit 48S preinitiation complexes
1009 (PDB:6YAL, Simonetti et al., 2022) (9) centered around the interaction of the 2'-OH group in
1010 nucleotide $C[-1]^{mRNA}$ with t^6A in $tRNA_i^{met}$. This interaction was reproduced in rabbit 80S initiator
1011 ribosomes in complex with Harringtonine (Fig. 1B). **(B-F),** Cryo-EM visualization of 48S
1012 preinitiation complexes (Petrychenko et al., 2025) (22) centered around the interaction of the 2'-
1013 OH group in nucleotide $A[+1]^{mRNA}$ with U1830 in 18S rRNA across different stages of 48S
1014 preinitiation complexes, including the stage of codon recognition and GTP hydrolysis (B, PDB:
1015 8PJ2), binding of eIF5B (C, PDB: 8PJ3), followed by sequential release of eIF5B (D, PDB: 8PJ4)
1016 and eIF2 (E, PDB: 8PJ5), and the late-stage readiness for 60S subunit joining (F, PDB: 8PJ6). All
1017 publicly available structures were downloaded from the PDB and visualized in PyMOL.

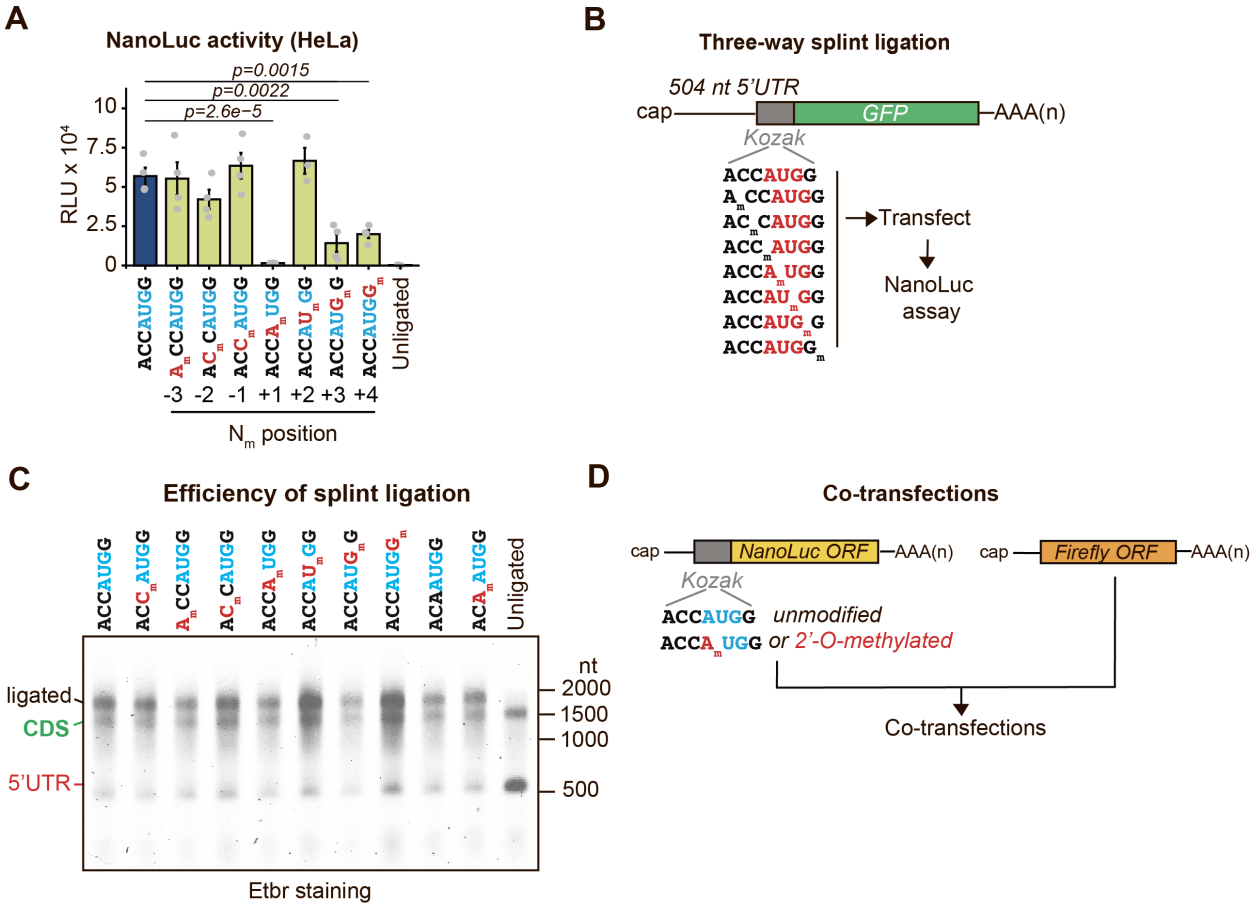
Supplementary Figure S2.



Supplementary Figure S2. Quality controls for the construction of NanoLuc mRNA reporters. **(A)** the 5' end of the mRNA reporters lacking 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. Three independent reactions are shown in the gel. **(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. Three independent capping reactions are shown in the membrane. **(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

1028 polyadenylation were verified using PAGE and SYBR gold staining. **(D)** the ligation of NanoLuc
1029 mRNA reporters was verified using PAGE and SYBR gold staining. Ligated mRNA shows a shift
1030 in molecular weight compared to unligated mRNA. **(E)** relative light units were measured in lysates
1031 from HEK293T cells co-transfected with increasing amounts of unligated RNA pieces and
1032 constant amounts (20 ng) of full-length NanoLuc mRNA. Mean ± SEM, n = 3. The line plot shows
1033 stable NanoLuc activity across different amounts of unligated RNA. **(F-G)** ligation efficiency was
1034 verified by RT-qPCR. Two primer sets were used (Supplementary Table S1). The first set
1035 specifically recognizes the ligated mRNA, while the second set binds the NanoLuc ORF to obtain
1036 total mRNA levels. The efficiency of ligation was calculated as: (ligated mRNA)/(total RNA)*100.
1037 **(H)** NanoLuc activity was normalized to the amount of transfected mRNA, which was quantified
1038 by RT-qPCR. Mean ± SEM, n = 3. **(I)** schematic of different constructs used in three-way and two-
1039 way splint ligations. **(J)** the ligation of NanoLuc mRNA reporters was verified using PAGE and
1040 SYBR gold staining. Ligated mRNA shows a shift in molecular weight compared to unligated
1041 mRNA. **(K)** the integrity of the gel-purified mRNA reporters was verified using PAGE and SYBR
1042 gold staining. **(L)** the ligation of NanoLuc mRNA reporters was verified using PAGE and SYBR
1043 gold staining. Ligated mRNA shows a slight shift corresponding to a 40 nt difference in molecular
1044 weight compared to unligated mRNA. **(M)** The efficiency of the two-way ligation was verified by
1045 RT-qPCR. Mean ± SEM, n = 4.

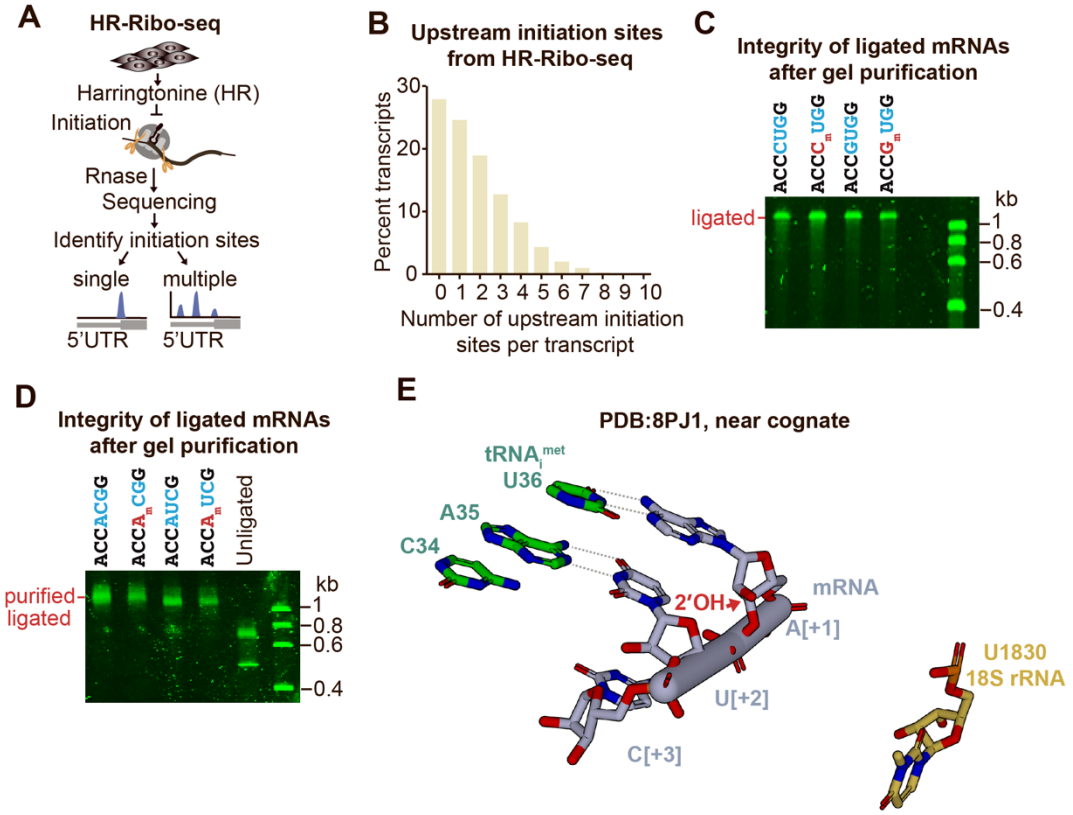
1046 **Supplementary Figure S3.**



1047

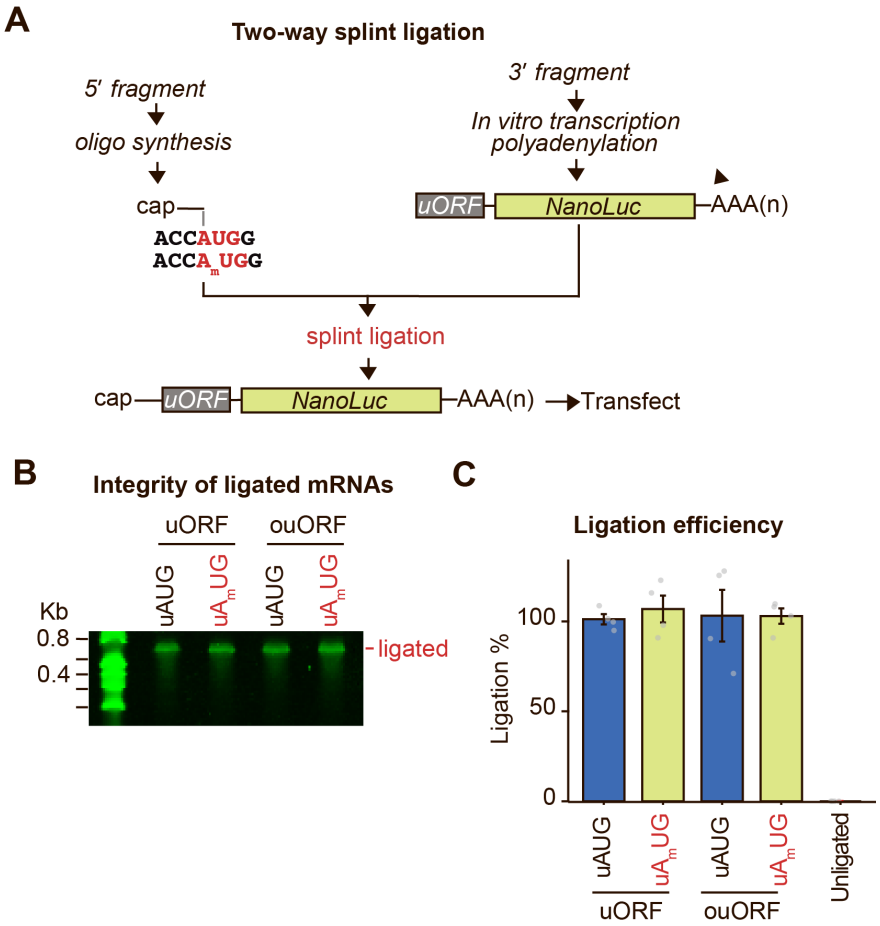
1048 **Supplementary Figure S3.** Quality controls for the construction of GFP mRNA reporters. **(A)**
1049 relative light units were measured in lysates from HeLa cells transfected with ligated NanoLuc
1050 mRNAs carrying 2'-O-methylation in different positions around the start codon. Unmethylated
1051 (blue) and unligated mRNAs were used as controls. Mean ± SEM, n = 4. p = One-way ANOVA
1052 with Tukey's post hoc analysis. **(B)** schematic of the GFP mRNA reporters. **(C)** denaturing agarose
1053 gel electrophoresis, indicating the ligation efficiencies for the GFP mRNA reporters. These RNA
1054 fragments are larger than those used for NanoLuc reporters and do not resolve well in PAGE
1055 gels. Thus, agarose gels are better suited for this separation. **(D)** schematic indicating the co-
1056 transfection strategy of the full-length firefly luciferase as a control compared to the splint ligated
1057 NanoLuc mRNA reporter.

1058 **Supplementary Figure S4.**



1059 **Supplementary Figure S4.** Quality controls for the construction of mRNA reporters with near-
1060 cognate initiation codons. **(A)** To determine the upstream near-cognate or AUG codons used for
1061 initiation, we previously performed HR-Ribo-seq in HeLa cells (8). In this method, the inhibitor
1062 Harringtonine (HR) stalls the initiator ribosome while letting the elongator ribosomes slide off of
1063 mRNA. Subsequent RNAse footprinting and isolation of ribosome-protected fragments (RPFs)
1064 allow the identification of initiation sites transcriptome-wide. **(B)** HR-Ribo-seq indicates that ~70%
1065 of transcripts initiate translation from at least one upstream site in wildtype HeLa cells. **(C-D)** the
1066 integrity of ligated NanoLuc mRNAs carrying near-cognate codons was verified using PAGE and
1067 SYBR gold staining. **(E)** Cryo-EM visualization of 48S preinitiation complexes (Petrychenko et al.,
1068 2025) (22) centered around nucleotide A[+1]^{mRNA} from an AUC start codon (PDB: 8PJ1). The
1069 distance between U1830 in rRNA and A[+1] in mRNA (10.5Å) does not support the formation of
1070 an H-bond. It is important to note that AUC is in different contexts in the cryo-EM structure
1071 (AGGAUCC) than our mRNA reporter system (ACCAUCG).

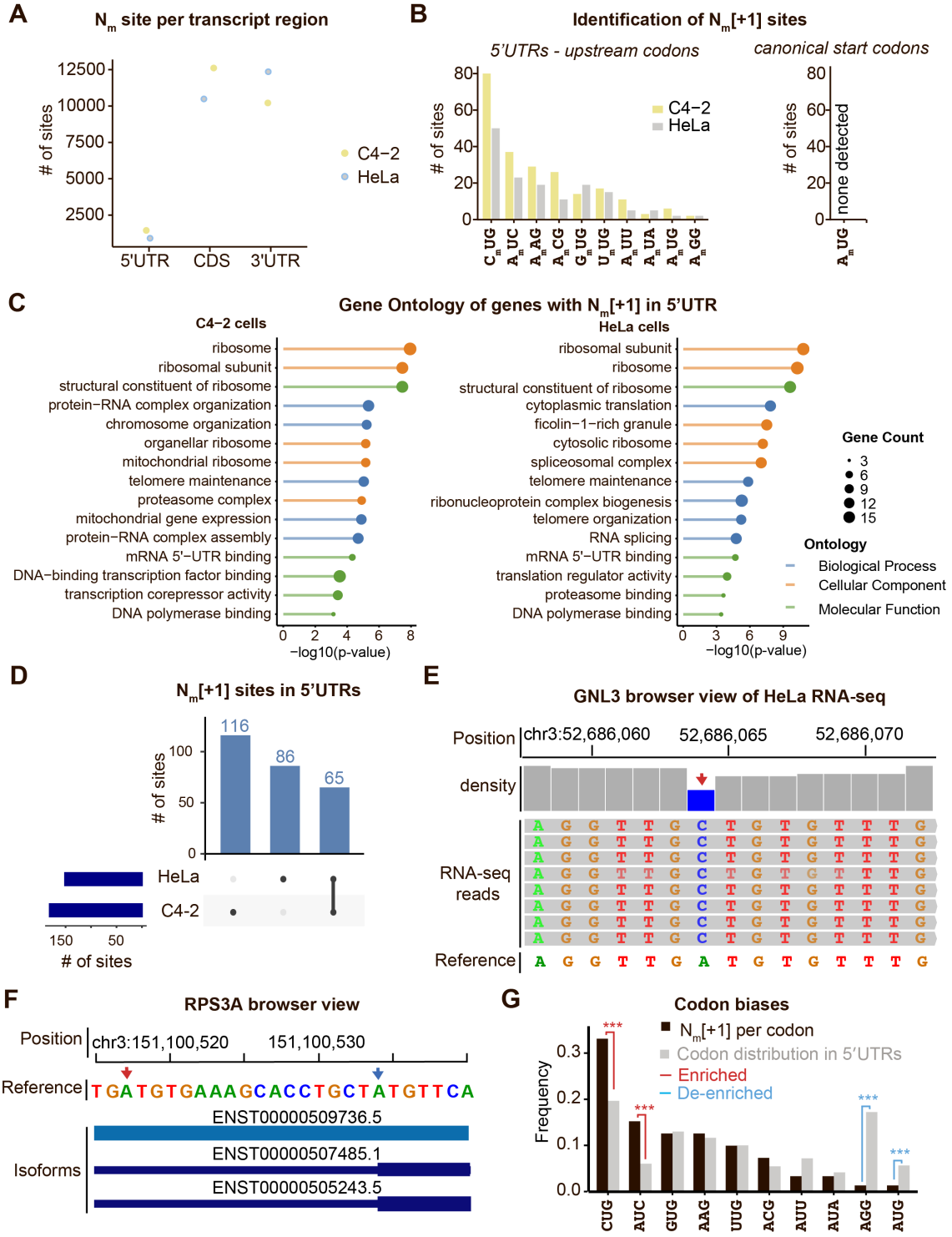
1072 **Supplementary Figure S5.**



1073

1074 **Supplementary Figure S5.** Quality controls for the construction of uORF reporters. **(A)** schematic
1075 of the two-way splint ligation strategy to generate mRNA reporters with 2'-O-methylated or
1076 unmethylated uORFs. **(B)** the integrity of ligated NanoLuc mRNAs was verified using PAGE and
1077 SYBR gold staining. **(C)** ligation efficiency was verified by RT-qPCR. Mean ± SEM, n = 4.

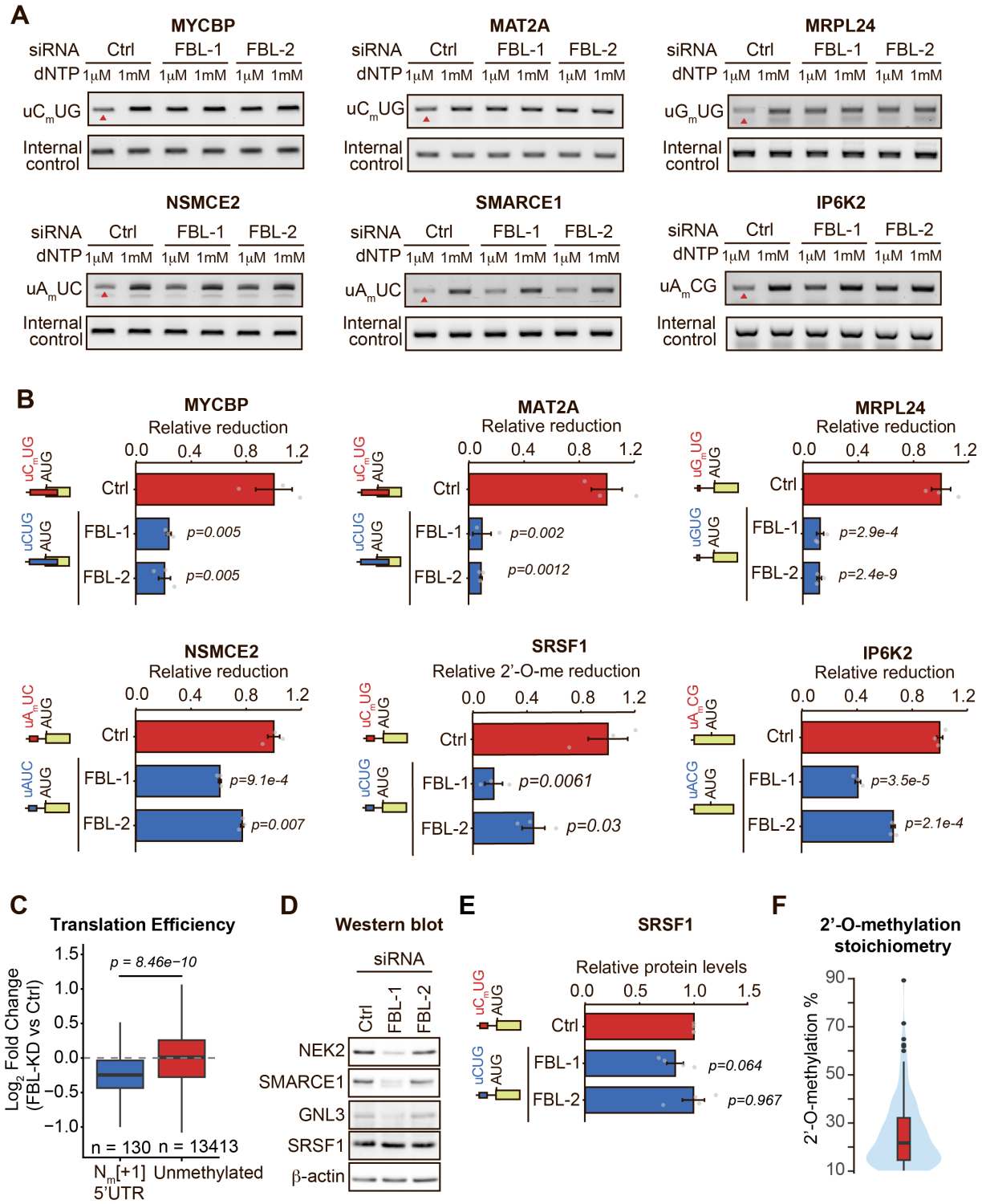
Supplementary Figure S6.



Supplementary Figure S6. Distribution of 2'-O-methylation sites in human mRNAs. **(A)** 2'-O-methylation maps in HeLa and C4-2 cells were compiled and binned by transcript location (5'UTR, CDS, or 3'UTR). **(B)** the genomic coordinate of 2'-O-methylation sites was intersected with the coordinates for canonical AUG, AUG within 5'UTRs, and near-cognate start codons within 5'UTRs of human transcriptomes. N_m[+1] sites were binned by codon identity. **(C)** gene ontology analysis

1085 of all genes from (B). **(D)** an upset plot indicating the overlap of the sites described in (B) between
1086 HeLa and C4-2 cells. **(E)** browser view of RNA-seq reads in HeLa cells indicates an A>C SNP at
1087 an annotated upstream AUG codon. **(F)** browser view of different transcript isoforms in RPS3A
1088 indicates that the annotated AUG codon is located in the 5'UTR of some isoforms but in the coding
1089 sequence of other isoforms. Red arrow indicates the 2'-O-methylation sites. Blue arrow indicates
1090 the canonical start codon. **(G)** the fraction of each N_m[+1]-methylated codon was compared to the
1091 distribution of AUG and all near-cognate codons in the 5'UTR of human mRNAs. p = chi-squared
1092 test for proportions. *** $p < 0.0001$.

Supplementary Figure S7.



Supplementary Figure S7. Validation of 2'-O-methylation sites in near-cognate codons. **(A)** representative agarose gels of amplicon density for the 2'-O-methylated site (top) and the loading control (bottom). Arrowheads indicate the reduction in amplicon signal when reverse transcription is performed at low dNTP concentration in the presence of FBL. **(B)** quantification of the relative

1099 N_m ratio dependent on FBL. Mean ± SEM, n = 3. p = two-tailed *Student's t-test*. **(C)** Analysis of
1100 publicly available Ribo-seq performed in HeLa cells (GSE105248) (21). This dataset contained
1101 three Ribo-seq and Input RNA-seq from FBL knockdown (Doxycycline-inducible shRNA against
1102 FBL) and control groups. Translation efficiency was calculated as normalized RPF counts divided
1103 by normalized mRNA counts. We compared genes with mapped N_m[+1] at an upstream AUG or
1104 near-cognate codon in the 5'UTR with those lacking any N_m in their 5'UTR in HeLa cells. p =
1105 Wilcoxon rank-sum test. **(D)** Western blot validation of NEK2, GNL3, and SRSF1 expression upon
1106 knockdown of FBL. Representative of n=4. **(E)** quantification of the relative protein levels
1107 dependent on FBL knockdown from the gels shown (D). Mean ± SEM, n = 4. p = two-tailed
1108 *Student's t-test*. **(F)** distribution of 2'-O-methylation stoichiometry in position N_m[-1] of near-
1109 cognate codons within human 5'UTRs.