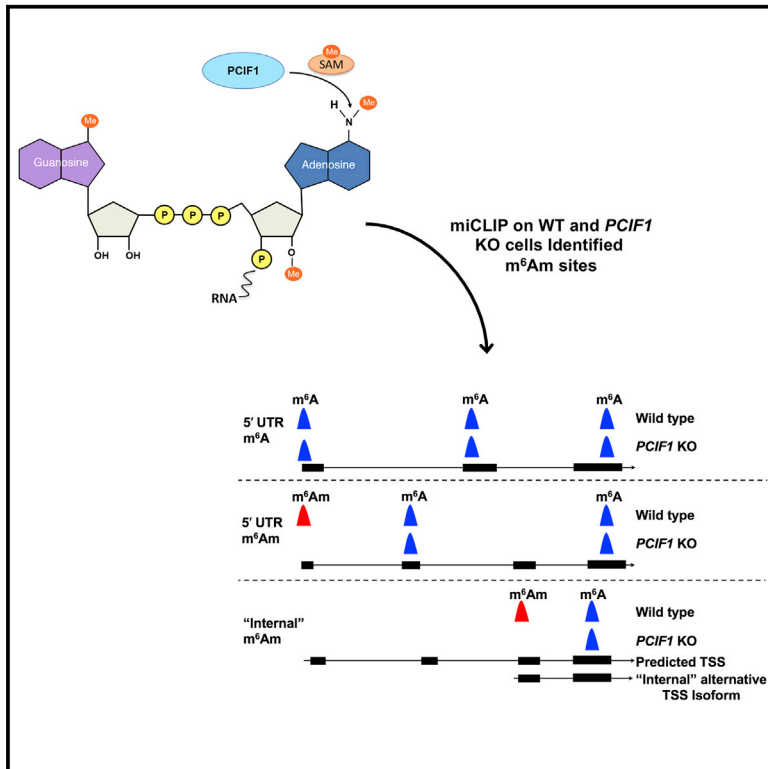


Molecular Cell

Identification of the m⁶Am Methyltransferase PCIF1 Reveals the Location and Functions of m⁶Am in the Transcriptome

Graphical Abstract



Authors

Konstantinos Boulias,
Diana Toczyłowska-Socha,
Ben R. Hawley, ..., L. Aravind,
Samie R. Jaffrey, Eric Lieberman Greer

Correspondence

srj2003@med.cornell.edu (S.R.J.),
eric.greer@childrens.harvard.edu (E.L.G.)

In Brief

m⁶Am is a prevalent mRNA modification occurring adjacent to the m⁷G cap. Boulias, Toczyłowska-Socha, Hawley et al. identify PCIF1 as the m⁶Am methyltransferase and perform transcriptome-wide mapping to distinguish m⁶Am from m⁶A and identify "internal" TSSs. m⁶Am increases stability of a subset of mRNAs but has minimal effects on translation.

Highlights

- PCIF1 is the N6-adenosine methylase that produces m⁶Am in an m⁷G cap-dependent manner
- PCIF1 depletion allows transcriptome-wide mapping of m⁶A and m⁶Am
- m⁶Am mapping identifies alternative "internal" transcription start sites
- m⁶Am increases stability of a subset of mRNAs and has no effect on translation

Identification of the m⁶Am Methyltransferase PCIF1 Reveals the Location and Functions of m⁶Am in the Transcriptome

Konstantinos Boulias,^{1,2,7} Diana Toczyłowska-Socha,^{3,4,7} Ben R. Hawley,^{3,7} Noa Liberman,^{1,2} Ken Takashima,^{1,2} Sara Zaccara,³ Théo Guez,⁵ Jean-Jacques Vasseur,⁵ Françoise Debart,⁵ L. Aravind,⁶ Samie R. Jaffrey,^{3,*} and Eric Lieberman Greer^{1,2,8,*}

¹Division of Newborn Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

²Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

³Department of Pharmacology, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA

⁴Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland

⁵IBMM, Université de Montpellier, CNRS, ENSCM, Montpellier, France

⁶NCBI, National Library of Medicine, NIH, Bethesda, MD 20894, USA

⁷These authors contributed equally

⁸Lead Contact

*Correspondence: srj2003@med.cornell.edu (S.R.J.), eric.greer@childrens.harvard.edu (E.L.G.)

<https://doi.org/10.1016/j.molcel.2019.06.006>

SUMMARY

mRNAs are regulated by nucleotide modifications that influence their cellular fate. Two of the most abundant modified nucleotides are N⁶-methyladenosine (m⁶A), found within mRNAs, and N⁶,2'-O-dimethyladenosine (m⁶Am), which is found at the first transcribed nucleotide. Distinguishing these modifications in mapping studies has been difficult. Here, we identify and biochemically characterize PCIF1, the methyltransferase that generates m⁶Am. We find that PCIF1 binds and is dependent on the m⁷G cap. By depleting PCIF1, we generated transcriptome-wide maps that distinguish m⁶Am and m⁶A. We find that m⁶A and m⁶Am misannotations arise from mRNA isoforms with alternative transcription start sites (TSSs). These isoforms contain m⁶Am that maps to "internal" sites, increasing the likelihood of misannotation. We find that depleting PCIF1 does not substantially affect mRNA translation but is associated with reduced stability of a subset of m⁶Am-annotated mRNAs. The discovery of PCIF1 and our accurate mapping technique will facilitate future studies to characterize m⁶Am's function.

INTRODUCTION

The most prevalent regulated methyl modifications in mRNA occur on two similar nucleotides: adenosine (A) and 2'-O-methyladenosine (Am) (Perry et al., 1975; Wei et al., 1975). METTL3 catalyzes the methylation on the N6 position of the adenine ring to form N⁶-methyladenosine (m⁶A) at internal sites in

mRNA (Bokar et al., 1997). At least 25% of mRNAs contain at least one m⁶A (Dominissini et al., 2012; Meyer et al., 2012).

N6 methylation also occurs on Am to form a dimethylated adenosine: N⁶,2'-O-dimethyladenosine (m⁶Am) (Keith et al., 1978; Wei et al., 1975). Am is primarily located at the first transcribed nucleotide position in mRNAs, adjacent to the m⁷G cap. Nucleotides located at the first transcribed nucleotide position in an mRNA are typically methylated on the ribose at the 2'-hydroxyl position. However, if this nucleotide is Am, it can undergo further N6 methylation to m⁶Am. Because m⁶Am is present at the first transcribed nucleotide in ~30% of all cellular mRNAs, m⁶Am can affect the fate of a large subset of the transcriptome (Wei et al., 1975).

Recent studies have started to reveal the functions of m⁶Am. m⁶Am is enriched in mRNAs with high stability and translation efficiency (Mauer et al., 2017). Mechanistically, m⁶Am was shown to impair mRNA decapping mediated by DCP2, leading to increased stability of at least some m⁶Am-modified mRNAs (Mauer et al., 2017). However, DCP2 does not regulate the stability of most mRNAs in the cell (Li et al., 2011, 2012). Instead, DCP2 is important for specific mRNA degradation pathways, such as nonsense-mediated decay, microRNA-mediated mRNA degradation, and mRNA degradation in response to interferon (Li et al., 2011, 2012). Therefore, it is not clear whether m⁶Am has a general role in promoting the high stability of m⁶Am-containing transcripts or whether m⁶Am confers mRNA stability to a subset of mRNAs that are degraded through selective decapping pathways.

Predicting the function of m⁶Am is complicated by the difficulty in determining whether an mRNA contains m⁶A or m⁶Am. Transcriptome-wide mapping of m⁶A and m⁶Am uses antibodies that bind the 6-methyladenine (6mA) nucleobase portion found in both of these methylated adenosine nucleotides. The two mapping methods, i.e., MeRIP-seq (methyl RNA immunoprecipitation followed by sequencing) (Dominissini et al., 2012; Meyer et al., 2012) and miCLIP (m⁶A individual-nucleotide-resolution

crosslinking and immunoprecipitation) (Linder et al., 2015), both map sites of 6mA rather than m⁶A or m⁶Am. The 6mA “peaks” are then interpreted to be either m⁶A or m⁶Am using a variety of criteria. For example, if the 6mA peak is in the 5' UTR, this suggests that the 6mA peak is caused by m⁶Am because this nucleotide is exclusively found as the transcription start nucleotides. Nevertheless, it can be difficult to distinguish m⁶Am from m⁶A located within the 5' UTR of mRNAs. As a result, previous maps of m⁶Am may have inaccuracies, which may make it difficult for predicting its function.

To definitively distinguish m⁶A and m⁶Am in transcriptome-wide maps, depletion of either m⁶A or m⁶Am would be required. m⁶A depletion cannot be readily achieved, as *Mettl3* is essential for survival in nearly all 341 cell lines that were screened (Tsherniak et al., 2017). The methyltransferase that generates m⁶Am is not known, but its depletion could enable the identification of the sites that are m⁶Am, because the remaining sites would be m⁶A.

Here, we describe the identification of PCIF1 as the methyltransferase that is responsible for generating m⁶Am in mRNA. We show that PCIF1 methylates Am in the context of the m⁷G cap and has negligible ability to methylate adenosine in mRNA outside this context. By mapping m⁶A in the transcriptome of PCIF1-deleted cells, we distinguish between m⁶Am and 5' UTR m⁶A. We find numerous examples where previously annotated m⁶Am sites reflect m⁶A and vice versa. We show that transcript isoforms with alternative transcription start sites (TSSs) account for many of these discrepancies facilitating the identification of these “internal” TSSs. Using this new high-confidence map of m⁶Am sites, we characterize the fate of m⁶Am-modified mRNAs in *PCIF1* knockout cells and show that m⁶Am has negligible effects on translation under basal conditions but is associated with increased stability of a subset of m⁶Am-initiated transcripts. Overall, our studies identify PCIF1 as the methyltransferase that generates m⁶Am in the transcriptome and provides revised transcriptome-wide maps that distinguish between m⁶A and m⁶Am.

RESULTS

Identification of PCIF1 as a Candidate m⁶Am-Forming Methyltransferase

Studies in the 1970s provided initial characterization of an enzymatic activity in HeLa cells that synthesizes m⁶Am (Keith et al., 1978). This enzyme selectively methylates Am adjacent to an m⁷G cap in synthetic RNA substrates (Keith et al., 1978).

In order to identify the m⁶Am-forming enzyme, we performed a comparative bioinformatic analysis of orphan adenosine methyltransferases. These enzymes contain the [DNSH]PP[YFW] motif, which is present in all adenine N6-methyltransferases (Iyer et al., 2016). Among these putative adenine methyltransferases, PCIF1 is notable because it evolved at the same time that the 5' cap emerged in mRNA (Iyer et al., 2016). It has been hypothesized that the 5' cap emerged with eukaryotic evolution to replace the Shine-Dalgarno sequence and direct ribosomes to mRNAs and to protect from 5' exoribonucleases to distinguish self-versus-foreign mRNAs (Furuichi et al., 1977; Shimotohno et al., 1977; Shuman, 2002). The PCIF1 methyltransferase family is derived from the prokaryotic M.EcoKI/M.TaqI methyltransferases of the bacterial restriction-modification systems (Iyer

et al., 2016). All of these methyltransferases contain helices before and after the conserved core strand-3, which display partial or complete degeneration into coil elements.

Another feature of these methyltransferases is the addition of a conserved residue from a helix N-terminal to the core methyltransferase catalytic domain. A PCIF1 crystal structure revealed that its putative methyltransferase domain indeed adopts the classical Rossmann fold of many RNA methyltransferases (Aki-chika et al., 2019). PCIF1 also contains a WW domain that interacts with the C-terminal domain of RNA polymerase II (Fan et al., 2003; Figure 1A), suggesting that its function is linked to transcription. Based on this, we asked whether PCIF1 is an adenine N6-methyltransferase in mRNA.

PCIF1 N6-Methylates 2'-O-Methyladenosine in an m⁷G Cap-Dependent Manner *In Vitro*

To identify potential PCIF1-dependent nucleotide methyltransferase activity, we bacterially expressed and purified glutathione S-transferase (GST)-tagged PCIF1. To test whether PCIF1 can methylate the cap-adjacent adenosine of mRNAs, we performed *in vitro* methyltransferase assays with an RNA oligonucleotide containing a 5' m⁷G cap followed by 2'-O-methyladenosine (m⁷G-ppp-Am-N₁₆) (Figure 1B). We found that PCIF1 methylates Am in this RNA to produce m⁶Am, as assessed by ultra-high-performance liquid chromatography coupled with triple-quadrupole tandem mass spectrometry (UHPLC-MS/MS; Figure 1C).

Interestingly, we did not detect m⁶A formation in these methylation reactions despite the presence of 5 internal adenosines in the RNA sequence (Figure 1C). Although PCIF1 may methylate an internal adenosine in a currently unknown sequence context, these findings suggest that PCIF1 preferentially N6-methylates 2'-O-methyladenosine rather than internal adenosines.

As a control, we generated predicted catalytically inactive PCIF1 by mutating both asparagine 553 and phenylalanine 556 to alanines (NPPF → APPA) or to a serine and a glycine (NPPF → SPPG). The corresponding mutations inactivate the EcoKI and Dam N6 methyltransferases (Guyot et al., 1993; Willcock et al., 1994). Neither the APPA nor SPPG mutant was able to methylate RNA (Figure 1C), suggesting that the PCIF1 catalytic domain is required for Am methylation *in vitro*.

We next asked whether PCIF1-mediated N6 methylation of the m⁷G-adjacent A requires 2'-O-methyl modification on the A. To test this, we used an RNA substrate with a 5' m⁷G cap followed by adenosine (m⁷G-ppp-A-N₁₆) rather than Am. Using *in vitro* methylation assays, we found that wild-type PCIF1, but not the SPPG or APPA PCIF1 mutant, was able to N6-methylate adenosine to m⁶A (Figure 1D). We next examined the rate and substrate preference of PCIF1 using a serial dilution of the m⁷G-ppp-Am and the m⁷G-ppp-A-capped oligonucleotides and tritiated S-adenosyl methionine [³H]-SAM as the methyl donor (Figure 1E). Michaelis-Menten analysis yielded a K_M = 82 ± 18.2 nM for the capped 2'-O-methylated RNA and a K_M = 630 ± 84.2 nM for the capped unmethylated A RNA (Figure 1F), suggesting that PCIF1 has an ~7.6-fold higher preference for binding the 2'-O-methylated adenosine substrate.

Notably, the m⁷G moiety was required for methylation as PCIF1 efficiently methylated Am to m⁶Am in an m⁷G capped RNA (m⁷G-ppp-Am-N₁₆) but was unable to methylate Am in an

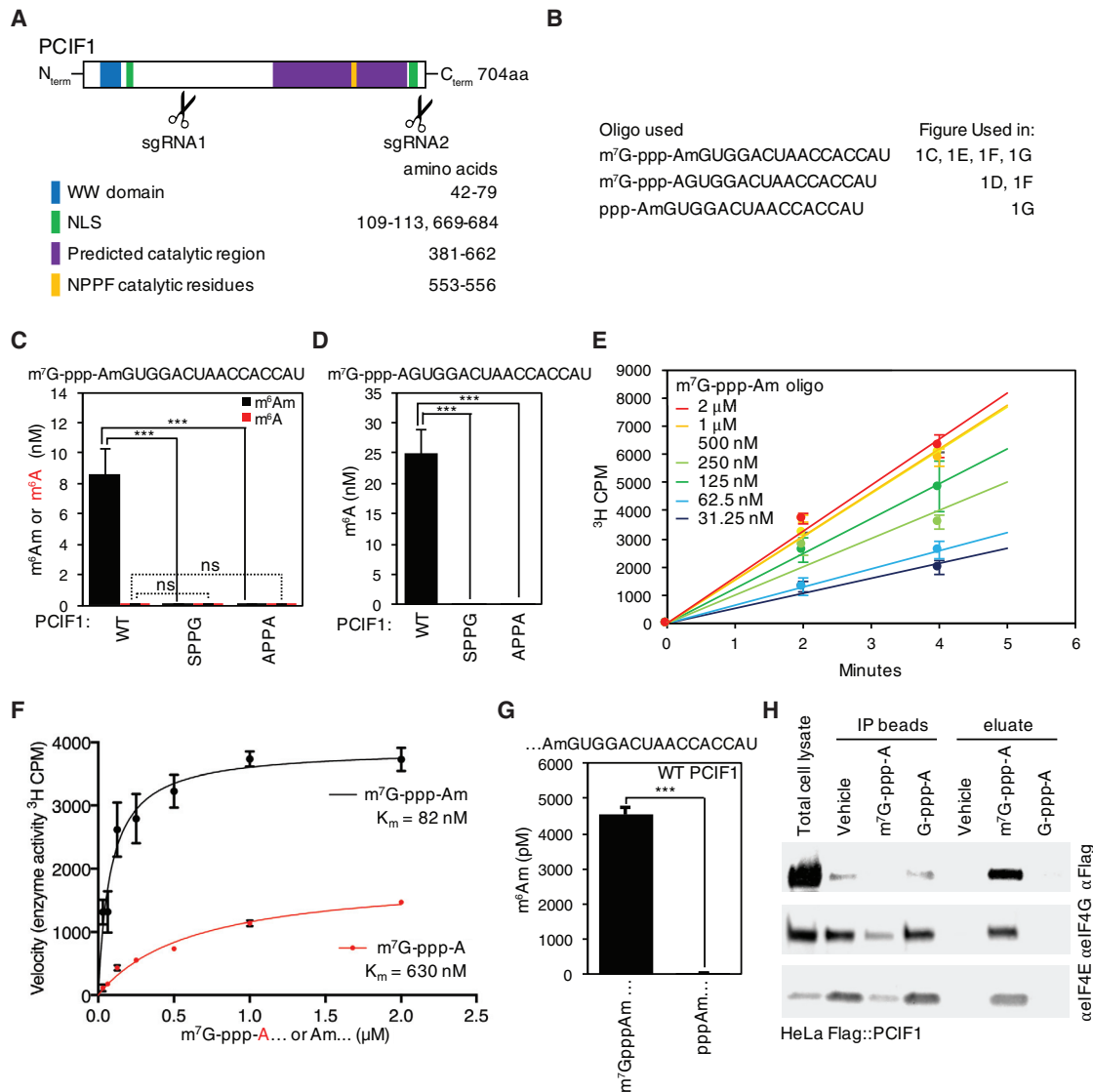


Figure 1. PCIF1 N6 Methylates 2'-O-Methyladenosine *In Vitro* in an m⁷G Cap-Dependent Manner

(A) Schematic of PCIF1 indicating the position of predicted functional domains. The location of the sites of mutations used in the study is shown. The catalytic domain includes a four-amino-acid motif, NPPF, which is predicted to be essential for mediating methylation (Iyer et al., 2016). The location of the site guide RNAs (gRNAs) (5'-CGGUUGAAAAGACUCCGUGG-3' and 5'-ACUUAACAUUAUCCUGCGGG-3') used in Figure 2 is indicated.

(B) Oligonucleotide sequences used in methyltransferase assays.

(C) PCIF1 methylates m⁷G-ppp-Am-N₁₆ RNA. GST-PCIF1 (50 nM), but not the catalytically inactive mutants APPA or SPPG, efficiently converts m⁷G-ppp-Am (4 μM) to m⁷G-ppp-m⁶Am as assessed by UHPLC-MS/MS. Under the same conditions (SAM, 160 μM, 10 min), PCIF1 does not convert any of the 5 internal adenosines to m⁶A. Each bar represents the mean ± SEM of 3 independent experiments. n.s., not significant; ***p < 0.001; as assessed by unpaired Student's t tests.

(D) PCIF1 methylates cap-adjacent adenosine regardless of 2'-O-ribose methylation. GST-PCIF1, but not the APPA or SPPG PCIF1 mutants, efficiently converts m⁷G-ppp-A-N₁₆ (4 μM) to m⁷G-ppp-m⁶A-N₁₆. Assays were performed as in (C). Each bar represents the mean ± SEM of 3 independent experiments. ***p < 0.001, as assessed by unpaired t tests.

(E) PCIF1 enzyme kinetics. m⁷G-ppp-Am-N₁₆ (at indicated concentration) was incubated with GST-PCIF1 (20 nM) for the indicated times in the presence of 1.33 μM ³H-SAM and 10 μM SAM. Methylation was determined by the presence of ³H in the RNA, as assessed by scintillation counting. Each point represents the mean ± SEM of 3 independent experiments.

(F) Michaelis-Menten kinetics of PCIF1 methyltransferase activity toward m⁷G-ppp-Am and m⁷G-ppp-A. Each point represents the mean ± SEM of 3 independent experiments.

(G) PCIF1 activity depends on the presence of the m⁷G cap. m⁷G-ppp-Am-N₁₆ or ppp-Am-N₁₆ (4 μM) was incubated with GST-PCIF1 as in (C). PCIF1 converted Am to m⁶Am specifically in the m⁷G capped RNA. Each bar represents the mean ± SEM of 3 independent experiments. ***p < 0.001, as assessed by unpaired t tests.

(H) PCIF1 directly binds the m⁷G cap. Anti-FLAG immunoblotting was used to detect binding of 3×FLAG-PCIF1 from HeLa cell extracts to m⁷GTP-conjugated beads. The beads were eluted with m⁷G-ppp-A or G-ppp-A. eIF4E and eIF4G were used to control for binding to m⁷G.

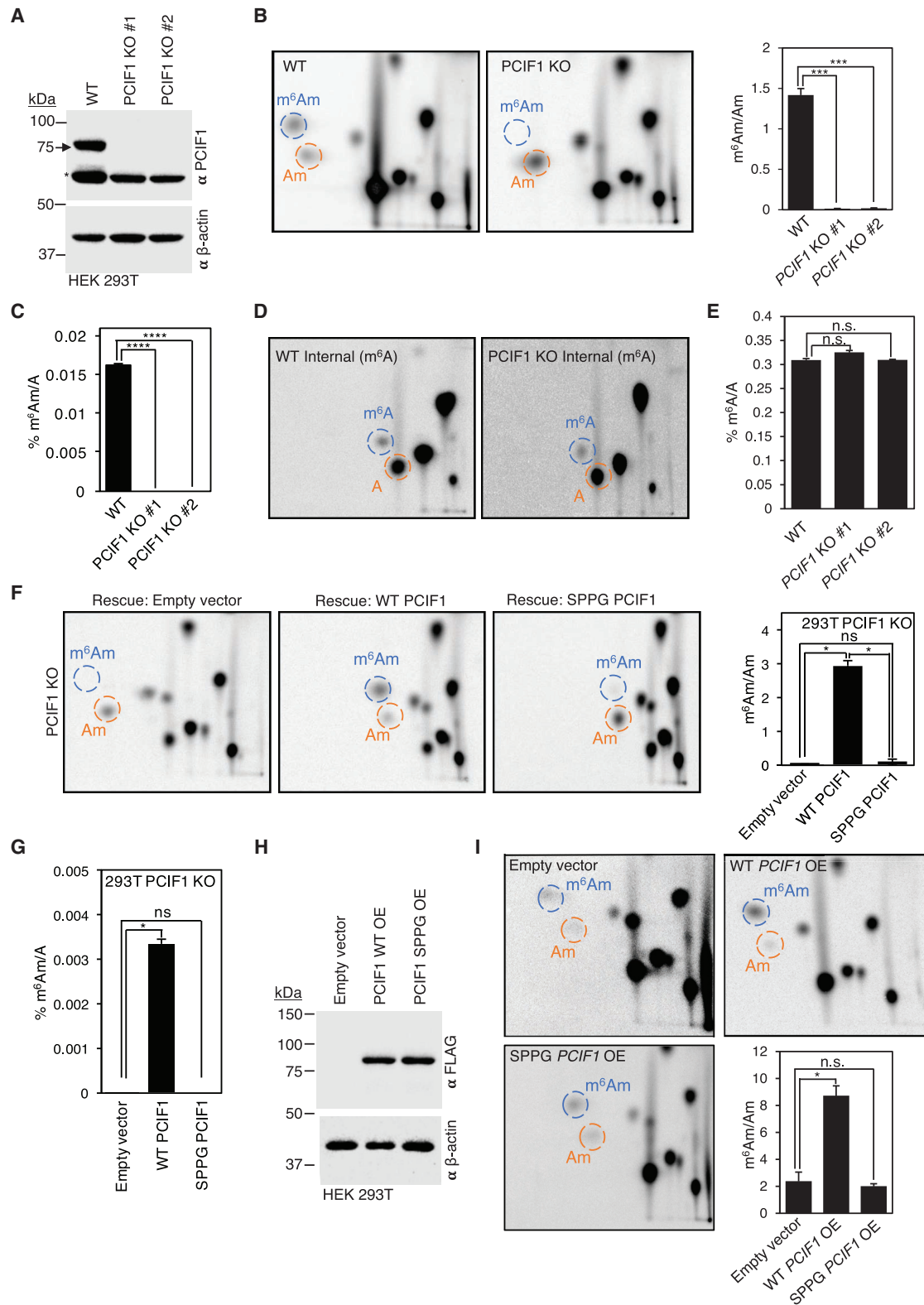


Figure 2. PCIF1 N6 Methylates 2'-O-Methyladenosine in Cells

(A) CRISPR-mediated *PCIF1* knockout (KO) in HEK293T cells was assessed by anti-PCIF1 immunoblotting. The upper band represents endogenous PCIF1, whereas the lower band is a non-specific band. β -actin, loading control.

(legend continued on next page)

RNA that lacked the m⁷G cap (ppp-Am-N₁₆; Figure 1G). Overall, these biochemical assays suggest that PCIF1 methyltransferase activity toward Am depends on the presence of the m⁷G cap but does not require 2'-O-methylation on the adenosine.

We next asked whether the PCIF1 preference for m⁷G-capped RNA was due to an ability to bind the m⁷G cap. We performed cap-binding assays with PCIF1 using 7-methylguanosine-5-triphosphate (m⁷G-ppp)-coupled Sepharose beads. In these experiments, we used lysates from HeLa cells expressing FLAG-tagged wild-type PCIF1. As expected, cap-binding proteins eIF4E and eIF4G were bound to m⁷G-ppp beads and were efficiently eluted using m⁷G-ppp-A, but not G-ppp-A (Figure 1H). Similarly, PCIF1 bound to the m⁷G-ppp beads and was eluted with m⁷G-ppp-A, but not G-ppp-A (Figure 1H). Together, these data suggest that PCIF1 binds directly to the m⁷G cap, which may account for its specificity toward adenosine adjacent to the m⁷G. These results are consistent with the crystal structure of PCIF1, which shows specific interactions with m⁷G (Akichika et al., 2019).

PCIF1 Knockout Abolishes m⁶Am Levels without Affecting m⁶A in RNA

To determine the ability of PCIF1 to generate m⁶Am in cells, we used CRISPR to delete *PCIF1* in various cell lines and examined levels of m⁶Am and m⁶A in RNA (Figures 2A and S1A). To measure m⁶Am, we used a two-dimensional thin-layer chromatography (2D-TLC)-based method that can measure both m⁶Am and Am, allowing the ratio of these modified forms of adenosine to be calculated in mRNA (Kruse et al., 2011). In this assay, mRNA is decapped, and the 5' nucleotide is selectively radiolabeled with [³²P]-ATP by polynucleotide kinase (PNK). Thus, the first transcribed nucleotide in RNA samples can be selectively detected and quantified. As expected, all the known nucleotides located at the first transcribed nucleotide in mRNA were detected, i.e., m⁶Am, Am, Gm, Cm, and Um. However, in *PCIF1* knockout cells, a selective and complete loss of m⁶Am was detected (Figures 2B and S1B). A similar effect was seen using UHPLC-MS/MS to quantify m⁶Am (Figure 2C). Thus, PCIF1 is required for the presence of m⁶Am at the first transcribed nucleotide in mRNA.

We next asked whether PCIF1 deletion affects m⁶A levels in mRNA. To test this, we used a 2D-TLC-based method that

selectively detects m⁶A in the G-A-C context (Zhong et al., 2008) and complemented this with UHPLC-MS/MS to quantify all m⁶A in mRNA. m⁶A was readily detected in mRNA in control cells, and no reduction was seen in *PCIF1* knockout cells (Figures 2D and 2E).

To confirm that the loss of m⁶Am in the *PCIF1* knockout cells was due to a loss of PCIF1 itself, we performed rescue experiments. In these experiments, we used wild-type or the SPPG catalytically inactive PCIF1 mutant (Figures S1C and S1D). We found that re-expression of the wild-type, but not the catalytically inactive, PCIF1 restored m⁶Am levels in mRNA of HEK293T *PCIF1* knockout cells as assessed by 2D-TLC (Figure 2F) and by UHPLC-MS/MS (Figure 2G).

We next asked whether PCIF1 was sufficient to increase m⁶Am levels in cells. We found that PCIF1 overexpression in HEK293T cells (Figure 2H) led to a ~3-fold increase in the m⁶Am-to-Am ratio (Figure 2I). This increase in m⁶Am levels was dependent on the catalytic activity of PCIF1, as overexpression of a catalytically inactive PCIF1 mutant had no effect on m⁶Am levels (Figure 2I). Together, these data suggest that PCIF1 is both necessary and sufficient to generate m⁶Am in mRNA in cells.

miCLIP Analysis of PCIF1 Knockout Cells Distinguishes m⁶Am from 5' UTR m⁶A Residues

Next, we used the *PCIF1* knockout cells to distinguish m⁶Am and m⁶A in transcriptome-wide 6mA maps. We performed miCLIP, a method that produces narrow peaks, and nucleotide transitions at and adjacent to the m⁶A (Linder et al., 2015). m⁶A is nearly universally followed by cytosine in mRNA (Wei et al., 1976). This C is frequently observed to undergo a C-to-T transition as a result of antibody crosslinking in miCLIP, which can then be used to identify m⁶A (Linder et al., 2015). Because m⁶Am can also be followed by cytosine, C-to-T transitions alone are not sufficient to distinguish m⁶A from m⁶Am. Peaks caused by m⁶Am display a unique shape that exhibits a marked drop off of reads at an annotated A-starting TSS, and this feature can be used to identify m⁶Am (Linder et al., 2015). However, because m⁶A occurring near the TSS would also produce a similar shaped peak, this approach may result in false-positive m⁶Am identifications.

(B) PCIF1 is required for formation of m⁶Am in mRNA in cells. m⁶Am and Am levels in poly(A) RNA were detected by radiolabeling the 5' nucleotide after decapping. RNA hydrolysates were resolved by 2D-TLC. Representative images are shown from 3 biological replicates. The bar graph on the right represents the mean ± SEM of 3 independent experiments. ***p < 0.001; Student's t test.

(C) m⁶Am is depleted in *PCIF1* KO HEK293T cells as assessed by UHPLC-MS/MS. Each bar represents the mean ± SEM of 3 independent experiments. ****p < 0.0001 as assessed by paired t tests.

(D) PCIF1 does not affect the level of internal m⁶A. 2D-TLC analysis of poly(A) RNA from HEK293T (wild-type [WT]) and *PCIF1* KO HEK293T show no effect on the level of m⁶A.

(E) Internal m⁶A is not affected in *PCIF1* KO HEK293T cells as assessed by UHPLC-MS/MS. Each bar represents the mean ± SEM of 3 independent experiments. ns as assessed by paired t tests.

(F) Wild-type, but not a catalytically inactive PCIF1 mutant, restores m⁶Am levels in *PCIF1* knockout cells as assessed by 2D-TLC. Each bar represents the mean ± SEM of two independent experiments.

(G) Wild-type, but not catalytically inactive PCIF1 mutant, restores m⁶Am levels in *PCIF1* KO cells as assessed by UHPLC-MS/MS. Each bar represents the mean ± SEM of two independent experiments. *p < 0.05 as assessed by paired t tests.

(H) Western blot analysis demonstrates equivalent expression of wild-type and catalytically inactive FLAG-tagged PCIF1. β-actin, loading control.

(I) Overexpression of wild-type, but not catalytically inactive, PCIF1 increases m⁶Am levels in HEK293T cells as assessed by 2D-TLC. Upper and left panels show representative images of 3 independent experiments. The bar graph represents the mean ± SEM of 3 independent experiments. ****p ≤ 0.0001, unpaired t tests. See also Figure S1.

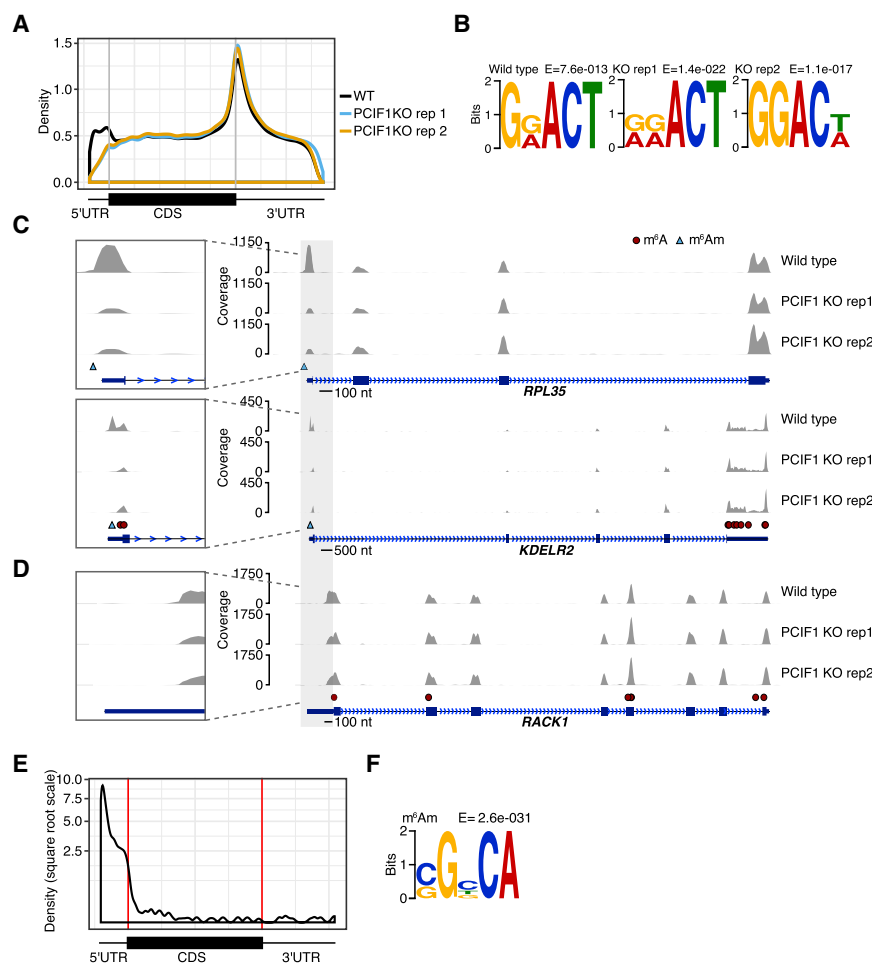


Figure 3. Depletion of PCIF1 Distinguishes m⁶A and m⁶Am in Transcriptome-wide 6mA Maps

(A) Metagene of miCLIP reads in wild-type and *PCIF1* KO HEK293T cells. Shown is a metagene analysis of reads from the wild-type or *PCIF1* KO miCLIP dataset. The first nucleotide of each read (with respect to the RNA strand) was extracted and plotted. Reads in the 5' UTR were lost in the *PCIF1* knockout, suggesting a complete loss of m⁶Am in the *PCIF1* knockout cells.

(B) DREME motif search within called peaks show the DRACH motif as the most enriched in all datasets, consistent with m⁶A as the most abundant 6mA-containing nucleotide mapped by miCLIP.

(C) miCLIP peaks can be identified as m⁶Am or m⁶A based on their decrease in *PCIF1* KO cells. Genome tracks were plotted for *RPL35* and *KDEL2* with called m⁶A sites (false discovery rate [FDR] < 0.1) and m⁶Am sites indicated by red circles and blue triangle, respectively. Zoomed insets show m⁶Am peaks can be distinguished from nearby m⁶A sites.

(D) The previously annotated m⁶Am site in *RACK1* is actually a 5' UTR m⁶A. The TSS-proximal peak in *RACK1* is not affected in the *PCIF1* knockout and overlaps with a DRACH motif.

(E) Metagene analysis of *PCIF1* KO-validated m⁶Am sites shows m⁶Am sites throughout the 5' UTR and in the transcript body. Shown is a metagene of the exact sites of m⁶Am within the *PCIF1*-dependent peaks as determined by A-to-T transitions and the read drop-off method. The metagene reveals an overall enrichment at the TSS, with some sites that appear to be within the coding sequence (CDS) and 3' UTR.

(F) DREME motif search of the nucleotides surrounding each m⁶Am was performed, confirms the previously reported BCA motif, and shows that the promoter sequence upstream of the m⁶Am is GC enriched.

See also [Figure S2](#).

Furthermore, these approaches are highly dependent on transcript annotations that may not have accurate TSS information for the cell type investigated. For example, annotated TSSs produced by RefSeq and ENSEMBL differ frequently for the same gene (Zhao and Zhang, 2015). Therefore, true m⁶Am peaks may have been discarded or thought to be m⁶A based on their location away from a TSS.

We therefore performed miCLIP in control and *PCIF1* knockout cells to distinguish m⁶Am and m⁶A. In control cells, reads were enriched in the vicinity of the stop codon as well as the TSS, which is generally assumed to reflect m⁶A and m⁶Am, respectively (Figure 3A). *PCIF1* knockout cells exhibited fewer reads mapping near the annotated TSS (Figure 3A; ~55% decrease in 5' UTR), suggesting these reads derive from an m⁶Am residue.

A motif analysis of significant peaks showed the DRACH m⁶A consensus (D = A, G, U; R = A, G; H = A, C, U) as the most common motif in each dataset (Figure 3B). This suggests that m⁶A is the most common modification mapped in both datasets, as expected.

To identify m⁶Am marked transcripts, we next examined the 6mA peaks that showed differences in the control and *PCIF1* knockout miCLIP datasets. As expected, we detected a loss of

peaks near the TSS of certain genes in the *PCIF1* knockout. For example, *RPL35* and *KDEL2* show peaks near the annotated TSS as well as at internal sites (Figure 3C). The TSS-proximal peaks were absent in the *PCIF1* knockout miCLIP dataset. These data are consistent with the idea that the TSS peaks predominantly reflect m⁶Am.

However, in some cases, the peaks near the TSS were not affected in the *PCIF1*-knockout dataset. For example, peaks near the TSSs of *RACK1* and *RPS5*, which were previously annotated as m⁶Am in HEK293T cells based on their location, peak shape, and lack of C-to-T transitions (Mauer et al., 2017), persist in the *PCIF1* knockout dataset (Figures 3D and S2A). These peaks contain a canonical DRACH m⁶A consensus motif, and C-to-T transitions are detected for *RACK1* (Figure 3D), suggesting that these sites are actually m⁶A.

The variability in C-to-T transitions reflects the low transition rate induced by the antibody adduct on this transcript. Altogether, these data indicate that *PCIF1* depletion can be used to determine the identity of an m⁶A peak.

Overall, only 60.2% of genes that had previously been annotated as m⁶Am (Mauer et al., 2017) were validated as m⁶Am based on their loss in *PCIF1* knockout cells. In some cases,

this could be explained by peaks being below the threshold for detection in one or both replicates. Nevertheless, this difference highlights the importance of depleting cells of PCIF1 to reduce false-positive m⁶Am identification.

A High-Confidence Transcriptome-wide Map of m⁶A and m⁶Am Based on PCIF1 Depletion

To create a high-confidence map of all m⁶Am sites in the transcriptome, we searched for all peaks that exhibit a marked reduction in miCLIP signal in the *PCIF1* knockout dataset. The majority of peaks showed no substantial difference between control and *PCIF1* knockout miCLIP datasets, suggesting that they are m⁶A (Figure S2B). However, 2,360 peaks overlapping 2,291 genes exhibited a significant reduction in both *PCIF1* knockout datasets (Figure S2B). In contrast, only 11 sites appeared to increase, suggesting a very low incidence of false positives.

We next identified the exact m⁶Am residue within each of these peaks. In our previous approach, we used a “pile up” of reads that drop off at the 5′ end of these read clusters in A-starting genes to predict the m⁶Am site (Linder et al., 2015). In some cases, the drop off is not easily detected or several of these were found in close proximity. This appears to occur when (1) the total reads are too few or (2) the reads terminate before the TSS, possibly due to impaired reverse transcription through the 2′-O-methyl modifications (Maden et al., 1995) in the cap-proximal nucleotides or due to non-templated nucleotide addition that occurs at the ends of cDNAs generated by reverse transcriptases (Chen and Patton, 2001).

Therefore, we wanted to develop an alternative approach to identify m⁶Am within the PCIF1-dependent peaks. Previously, we observed antibody-induced A-to-T transitions at the m⁶A site in miCLIP (Linder et al., 2015). We confirmed that A-to-T transitions are readily detected at known m⁶Am and m⁶A throughout the transcriptome (Figures S2C and S2D). Therefore, we used a 10% A-to-T transition rate to identify the m⁶Am within PCIF1-dependent peaks. The drop-off approach was used when the A-to-T transition rate did not meet these criteria (Figure S2E). There was high similarity in the m⁶Am sites that were called when using these methods separately (Figure S2F).

Overall, the 2,350 m⁶Am sites mapped based on their dependence on PCIF1 (Table S1) were primarily located throughout the 5′ UTR (~94%), with a prominent enrichment at the annotated TSS (Figure 3E). Motif analysis of the genomic context of the exact m⁶Am nucleotide revealed the BCA motif, with A representing the m⁶Am and BC representing upstream genomic nucleotides (B = C, G, or T), as reported previously for m⁶Am (Linder et al., 2015). Additionally, motif analysis shows the upstream promoter sequence is GC enriched (Figure 3F). Motifs downstream and including the transcription-start adenosine were also enriched, suggesting that m⁶Am occurs in specific sequence contexts within mRNA transcripts (Figure S2G).

Next, we mapped m⁶A in the 5′ UTR. C-to-T transitions in a DRACH consensus were used to call on average 44,025 m⁶A sites in 10,383 genes based on the miCLIP protocol (Linder et al., 2015). This identified 399 5′ UTR m⁶A sites that were robustly called across all datasets (Table S2).

We next asked whether mRNAs with 5′ UTR m⁶A and mRNAs that contain m⁶Am are linked to different cellular processes,

based on our updated m⁶A and m⁶Am sites. Functional annotation using DAVID (Database for Annotation, Visualization and Integrated Discovery) shows that transcripts containing these distinct modified nucleotides are linked to different cellular processes, with 5′ UTR m⁶A associated with processes such as transcription and cell division, and m⁶Am is primarily associated with splicing (Figures S2H and S2I; Table S3).

ATF4 Contains a m⁶Am Rather Than m⁶A in Its 5′ UTR

We next wanted to understand whether our revised map of m⁶Am and m⁶A can identify transcripts with misannotated modified nucleotides. A 5′ UTR m⁶A site has been described as mediating the unusual stress-regulated translation of *ATF4* (Zhou et al., 2018). *ATF4* has two upstream open reading frames (uORFs) in its 5′ UTR. In unstressed cells, the uORFs are translated, which prevents translation of the main open reading frame, which encodes the ATF4 protein (Vattem and Wek, 2004). However, during stress, the second uORF is skipped, and the ribosome scans to the main open read frame after translating the first uORF. This allows the ATF4 protein to be translated during stress. m⁶A was mapped to the second open reading frame and was described as disappearing in a stress-dependent manner, thus causing a stress-regulated switch in ATF4 translation (Zhou et al., 2018).

However, using miCLIP, it is apparent that the 6mA peak in the 5′ UTR of *ATF4* is not located within the second open reading frame (Figure S3A). Instead, the peak is located at the transcription-start nucleotide and does not overlap with the position of the putative m⁶A.

Based on the location of the peak, we asked whether it instead reflects m⁶Am rather than m⁶A. To test this, we examined *ATF4* in the *PCIF1* knockout miCLIP dataset. Here, we observed a complete loss of this peak, further confirming that this site is m⁶Am (Figure S3A).

The role of m⁶A in controlling stress-induced *Atf4* translation was described in mouse embryonic fibroblast cells (Zhou et al., 2018) rather than the HEK293T cells used here. Human cells appear to have lost the DRACH consensus sequence surrounding the putative m⁶A site (Figure S3B). Conceivably, human cells exhibit stress-induced regulation of ATF4 translation through an m⁶A-independent pathway and mouse cells utilize an m⁶A-dependent pathway. Therefore, we mapped 6mA in mouse embryonic fibroblasts using miCLIP (Figure S3C). Again, the 6mA peak was at the TSS, not at a position corresponding to the second uORF (Figure S3D). These data further show that this peak derives from a m⁶Am residue. In comparison, there were low levels of 6mA reads throughout the transcript body, suggesting either background reads or low stoichiometry m⁶A sites (Figure S3D).

Overall, these data suggest that *ATF4* contains a m⁶Am at the TSS but no prominent m⁶A site within the 5′ UTR as previously reported (Zhou et al., 2018). Thus, a role for a 5′ UTR m⁶A in regulating *ATF4* translation seems unlikely. Overall, these data demonstrate the ease with which m⁶A and m⁶Am sites can be confused for each other.

Identification of Internal 6mA Sites that Reflect Transcription-Start m⁶Am

We noticed two unusual features in our mapping results. First, not all m⁶Am sites mapped to regions within annotated mRNA

transcripts. Second, the m⁶Am metagene showed that, although 94% of m⁶Am sites were located in the 5' UTR, many were not directly at the annotated TSS and, in some cases, further downstream within the transcript body (Figure 3E).

We considered that these findings could be due to m⁶Am that occurs in mRNA isoforms with alternate TSSs upstream or downstream of the TSS in the RefSeq-annotated transcript. To test this, we created an m⁶Am metaplot relative to RefSeq-annotated TSSs (Figure 4A). We observed 16.7% of m⁶Am sites mapping within 250 nt upstream of annotated TSSs, suggesting that some m⁶Am occurs in isoforms with upstream TSSs. We similarly observed m⁶Am upstream of TSSs using GENCODE (Frankish et al., 2019) transcript annotations (Figure 4B). The FANTOM5 promoter-level expression atlas (Abugessaisa et al., 2017) uses a set of TSSs specifically mapped across multiple tissues using the cap analysis gene expression (CAGE) approach. Using FANTOM5, we observed a marked overlap with our m⁶Am sites, supporting the idea that these m⁶Am sites are indeed TSSs (Figure 4C).

TSS heterogeneity likely explains why some m⁶Am sites map within the 5' UTR rather than being solely located at the annotated TSS. In the case of *YBX1*, a 6mA peak is mapped to the 5' UTR and is lost in the *PCIF1* knockout miCLIP dataset, suggesting that this peak is due to m⁶Am (Figure 4D). This m⁶Am likely reflects an isoform with a TSS located at this m⁶Am site, based on its overlap with a CAGE peak (Figure 4D). Thus, the presence of m⁶Am within the 5' UTR likely reflects TSS heterogeneity rather than “internal” m⁶Am nucleotides.

We next wanted to understand why ~6% of m⁶Am sites (121 sites) appear to map to coding sequences or 3' UTR regions (Figure 4E). For example, *YOD1* shows an internal m⁶A peak in the first exon that is lost in the *PCIF1* knockout miCLIP dataset (Figure 4F). As with *YBX1*, we observed a TSS that overlapped with the m⁶Am site. Thus, this internal site, which would normally have been assumed to be m⁶A using MeRIP-seq and possibly miCLIP, derives from an isoform starting with m⁶Am.

To test this idea further, we performed a metagene analysis on m⁶Am sites mapping to coding sequences or the 3' UTR. Here, we plotted the distance to the nearest CAGE sites (Figure 4G). This analysis shows that many m⁶Am sites in the coding sequence and 3' UTR are located at or near CAGE sites. Additionally, these m⁶Am sites show the BCA motif, which resembles the transcription initiation site (initiator element [Inr]) motif of RNA polymerase II (Figure S2G; Yang et al., 2007). This motif was also previously found for m⁶Am mapped to canonical TSSs (Linder et al., 2015). 5' RACE confirmed that our called m⁶Am sites are indeed transcription start nucleotides (Figure S3E). Overall, these data further suggest that m⁶Am is not internally located within transcripts but is instead found at the TSSs.

Approximately 8% of m⁶Am sites that mapped to the coding sequence or 3' UTR also contained an adjacent C-to-T transition. As a result, these peaks would likely have been called as an m⁶A. These data highlight the value of using *PCIF1* depletion to validate the transcriptome-wide m⁶Am and m⁶A maps.

m⁶Am Correlates with Enhanced Translation, Expression, and Stability of mRNAs

In our previous studies, we found that m⁶Am is correlated with transcripts that are highly expressed and have long half-lives in

cells (Mauer et al., 2017). We therefore wanted to re-examine this correlation based on the high-confidence m⁶Am annotation based on peaks that were depleted in the *PCIF1* knockout miCLIP dataset. In some cases, mRNAs that had been previously annotated as beginning with Am, Cm, Gm, or Um were re-annotated as m⁶Am for this analysis, and mRNAs previously annotated as m⁶Am were re-annotated as Am based on our revised mapping data. Analysis of mRNA expression and half-lives showed that transcripts that begin with m⁶Am are indeed more highly expressed and stable than mRNAs with other start nucleotides. Notably, m⁶Am appears to be the predominant start nucleotide of the mRNAs that are the most abundant and have annotated half-lives greater than 24 h (Figures 5A–5D).

Overall, these data suggest that the presence of m⁶Am correlates with an overall increase in mRNA stability and that m⁶Am is the predominant starting nucleotide on “outlier” mRNAs with unusually high stability and expression. To determine whether the N6-methyl in m⁶Am was required for the unique properties of these outlier mRNAs, we examined mRNA stability in *PCIF1* knockout HEK293T cells. mRNA stability was measured using SLAM-seq (thiol(SH)-linked alkylation for the metabolic sequencing of RNA) (Herzog et al., 2017; Figure S4A).

To examine the outlier mRNAs, which are highly expressed, we separately examined mRNAs in the lower and upper half of gene expression. We only used transcripts that exhibited a minimum threshold of transitions required for mRNA half-life quantification. For mRNAs in the lower half of gene expression, we observed a marked decrease in mRNA half-life upon *PCIF1* depletion (Figure 5E). We confirmed this effect by examining the stability of individual mRNAs after treatment of control and *PCIF1* knockout HEK293T cells with actinomycin D. Both *NBR1* and *AKAP12* transcripts exhibited decreased expression after 8 h of actinomycin D treatment (Figure S4B). This effect was more prominent in *PCIF1* knockout cells, consistent with a stabilizing effect of m⁶Am (Figure S4B).

However, when we examined the more abundant mRNAs, which are enriched in the outlier transcripts, these transcripts did not show a substantial change in mRNA half-life (Figures 5F and S4C). We observed a slight reduction in stability relative to Am-annotated transcripts, but compared to all mRNAs (Am, Cm, Gm, and Um), these mRNAs appeared to show small but nonsignificant increase in mRNA stability in *PCIF1* knockout cells.

Thus, although m⁶Am is highly enriched in these outlier transcripts, the N6 methyl does not appear to account for their unusual stability. In contrast, mRNAs in the lower half of gene expression appear to utilize m⁶Am for transcript stability.

Previously, we found that m⁶Am-containing transcripts exhibit a subtle increase in translation relative to mRNAs with other start nucleotides (Mauer et al., 2017). To more directly test the role of m⁶Am on translation, we compared the translation efficiency of transcripts in control and *PCIF1* knockout cells by ribosome profiling (Figures S4D and S4E). Here, we found that transcripts that contained m⁶Am as the transcription-start nucleotide did not show a substantial change in translation efficiency upon *PCIF1* depletion (Figure S4F). Rather than showing a decrease in translation, we observed a slight increase in translation upon loss of m⁶Am compared to transcripts annotated to begin with other nucleotides (Figure S4F). In agreement with the modest

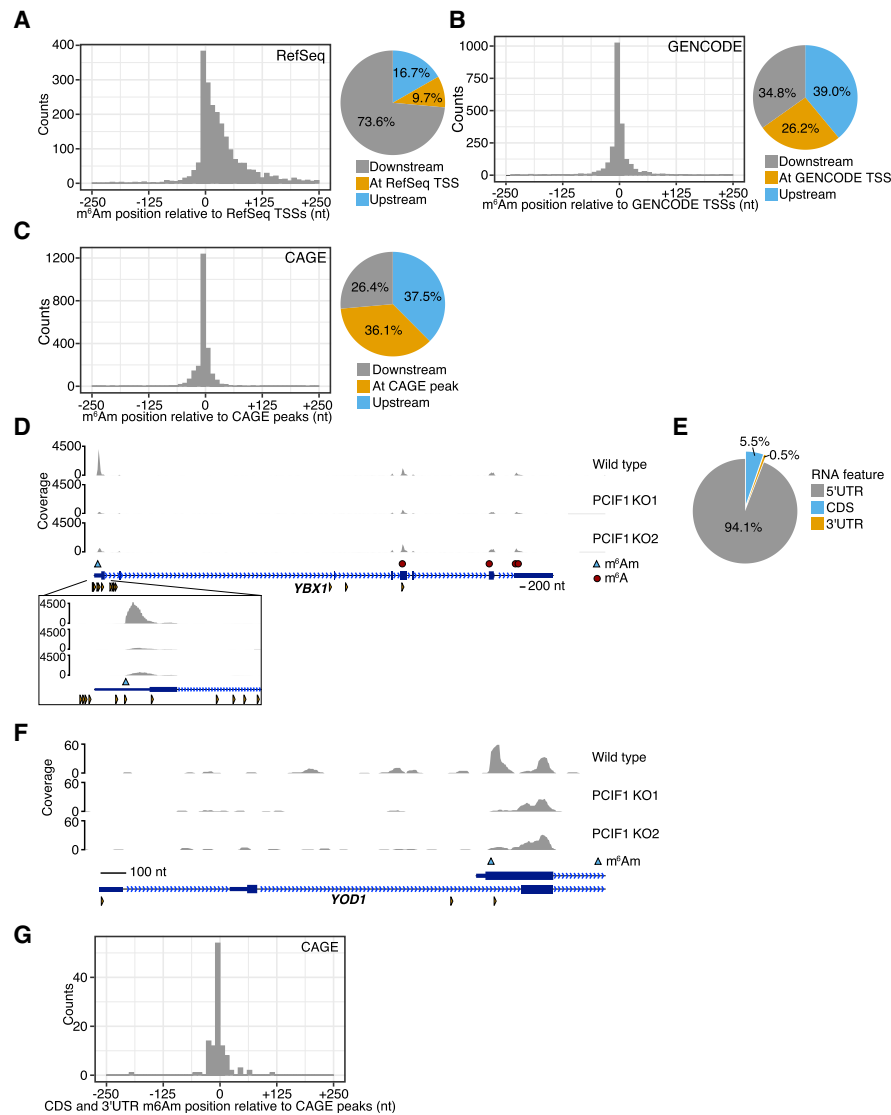


Figure 4. Internally Mapped m⁶Am Sites Reflect m⁶Am in mRNA Isoforms with Alternative TSSs

(A) A metaplot centered on the closest RefSeq TSS for each called m⁶Am site shows most m⁶Am sites are found downstream of the annotated start site, but not at the annotated start site. The proportion of m⁶Am directly at the annotated TSS or up- or downstream is shown.

(B) A metaplot analysis of m⁶Am locations using GENCODE TSS annotations shows higher overlap with TSSs. GENCODE annotations include more transcript isoforms and TSSs than RefSeq.

(C) A metaplot of the distance from each m⁶Am site to the closest CAGE peak in the FANTOM5 database shows that m⁶Am sites are indeed TSSs. Here, the overlap of m⁶Am was highest, suggesting that m⁶Am sites are selectively localized to TSSs and not internal nucleotides within mRNA.

(D) The m⁶Am mapping to the annotated 5' UTR of the *YBX1* transcript reflects a transcript isoform. The PCIF1-dependent 6mA peak in *YBX1* maps within the annotated 5' UTR of *YBX1*. However, this peak overlaps with a CAGE site (orange triangles), indicating the existence of a transcript isoform that initiates at this 6mA site. m⁶Am peaks that appear within the 5' UTR reflect m⁶Am in transcript isoforms with alternative TSSs. The exact m⁶Am site (blue triangle) was determined using the A-to-T transition within the PCIF1-dependent peak.

(E) Most m⁶Am sites are found in the annotated 5' UTR of transcripts.

(F) The internally mapping m⁶Am in *YOD1* derives from a TSS of a *YOD1* transcript isoform. The m⁶Am peak in *YOD1* begins beyond the start codon of both annotated isoforms. CAGE peaks (orange triangles) suggest this is indeed a TSS.

(G) A metaplot analysis of CDS and 3' UTR mapping m⁶Am sites show overlap with CAGE data, indicating that m⁶Am occurs at TSSs. The closest CAGE peak to each of the 6% of sites that appeared to not map to the 5' UTR (E) was calculated and plotted.

See also Figure S3.

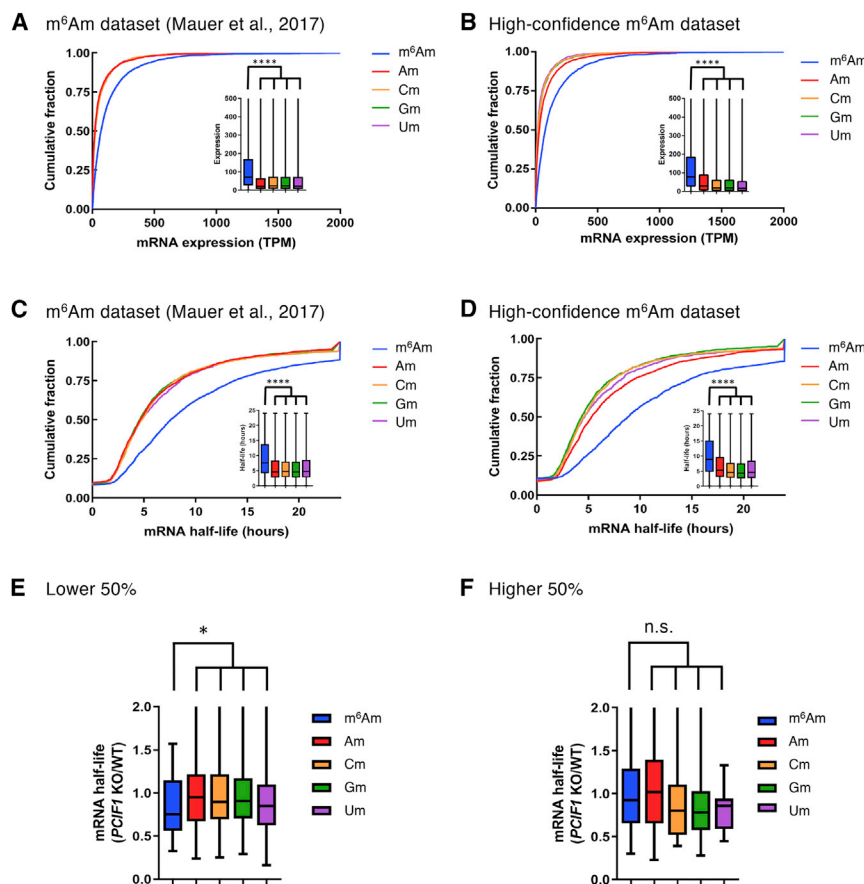


Figure 5. mRNA Expression Level and mRNA Half-Life Depending on TSS

(A) mRNAs with an annotated m⁶A start nucleotide show higher mRNA expression than other mRNAs. mRNA expression level in wild-type HEK293T cells was based on the first annotated nucleotide and an earlier m⁶A map (Mauer et al., 2017). Transcripts that start with m⁶A are significantly upregulated. ****p < 2.2 × 10⁻¹⁶; Student's t test. Cumulative distribution plot and boxplot represent the expression for mRNAs starting with m⁶A, Am, Cm, Gm, and Um. Data shown are the average gene expression measured from two replicates for HEK293T cells.

(B) m⁶A mRNAs annotated using the *PCIF1* KO miCLIP dataset show increased expression compared to mRNAs with other start nucleotides. Cumulative distribution plots were prepared as in (A) using the high-confidence m⁶A dataset. The transcripts starts with m⁶A are significantly upregulated as in (A). ****p < 2.2 × 10⁻¹⁶; Student's t test.

(C) mRNAs with an annotated m⁶A start nucleotide show higher mRNA half-life than other mRNAs. Annotated mRNA half-lives were based on the first annotated nucleotide and an earlier m⁶A map (Mauer et al., 2017). mRNAs with an annotated m⁶A exhibit a significantly elevated mRNA half-life than mRNAs with other annotated start nucleotides. ****p ≤ 2.2 × 10⁻¹⁶; Student's t test.

(D) m⁶A mRNAs annotated using the *PCIF1* KO miCLIP dataset show increased half-life compared to mRNAs with other start nucleotides. Transcripts with m⁶A have significantly longer half-life with similar p value. ****p ≤ 2.2 × 10⁻¹⁶; Student's t test.

(E) Influence of *PCIF1* depletion on mRNA half-life for transcripts in the lower half of gene expression. Transcripts with m⁶A have significantly shorter half-life in comparison to mRNAs with other annotated start nucleotides. *p = 0.0258 by Student's t test.

(F) Influence of *PCIF1* depletion on mRNA half-life for highly expressed transcripts. Transcripts with m⁶A show no significant decrease in mRNA half-life in comparison to mRNAs with other annotated start nucleotides. n.s., Student's t test.

See also Figure S4.

effects of *PCIF1* depletion on translation rates, we found that levels of proteins encoded by several m⁶A-modified mRNAs remained largely unchanged in *PCIF1* KO cells (Figure S4G).

Together, these experiments suggest that, under the conditions used in these experiments, N6 methylation does not mediate the increase in translation efficiency of m⁶A-initiated mRNAs in HEK293T cells.

DISCUSSION

A major challenge when mapping m⁶A and m⁶A is that both nucleotides are recognized by 6mA-specific antibodies and both can produce peaks in the 5' UTR of mRNA transcripts. Here, by identifying *PCIF1* as the m⁶A-forming methyltransferase and by depleting *PCIF1* to definitively identify m⁶A sites, we present a revised annotation of m⁶A and m⁶A in the transcriptome. We find that previous annotations contain errors that reflect the existence of mRNA isoforms that differ by TSSs. In some cases, the isoforms contain TSSs that map to internal sites within the annotated transcripts, resulting in the appearance of

peaks that would otherwise be attributed to m⁶A. The identification and characterization of *PCIF1* coupled with precise m⁶A annotations generated by *PCIF1* depletion will facilitate the identification of functions for m⁶A.

Our studies provide insights into the function of m⁶A. Using our new high-confidence m⁶A map, we find that m⁶A is found on unusually stable and highly abundant transcripts in cells. However, depletion of m⁶A by *PCIF1* knockout does not markedly impair the stability of these unusual transcripts under basal conditions. This suggests that m⁶A does not account for the stability of these unusual mRNAs. m⁶A therefore is likely to co-occur with other transcript features that confer these unusual properties to these mRNAs.

However, m⁶A does promote the stability of other mRNAs. When we examined mRNAs in the lower half of gene expression, we found a marked drop in mRNA stability in *PCIF1*-depleted cells. Why might some mRNAs be stabilized by m⁶A and others may not be affected? First, m⁶A is enriched in different sequence contexts. This contrasts with m⁶A, which is nearly always found in a single sequence context. m⁶A is therefore

likely to have different sensitivities to decapping or bind to different m⁶Am readers in a context-dependent manner. Thus, unlike traditional approaches, where mRNAs are binned and bioinformatically analyzed based on the presence or absence of a modification, m⁶Am functions are more likely to be revealed by analysis of mRNAs binned based on m⁶Am sequence contexts.

Another important factor that might affect whether m⁶Am has a destabilizing effect is whether the mRNA utilizes DCP2 or potentially other m⁶Am-sensitive decapping mechanisms. Previous studies showed that m⁶Am confers stability of mRNAs to DCP2-mediated decapping (Mauer et al., 2017). DCP2 is not the major decapping enzyme in cells, but DCP2 targets mRNAs with specific 3' UTR features. Thus, m⁶Am may stabilize mRNAs when DCP2-dependent pathways are activated. Overall, our studies show that m⁶Am acts in a transcript-selective manner rather than a general mRNA-stabilizing modification.

Although our study focused on mRNA stability, another study examined m⁶Am mRNA abundance and found no effects upon *PCIF1* depletion (Akichika et al., 2019). This might reflect compensatory upregulation of m⁶Am mRNAs. Notably, Akichika et al. examined all mRNAs rather than specific subsets of m⁶Am-annotated mRNAs.

Two recent studies also reported *PCIF1* as the m⁶Am cap-dependent methyltransferase (Akichika et al., 2019; Sun et al., 2019). Akichika et al. found that m⁶Am slightly enhances translation relative to the Am form of the transcript in *PCIF1* knockout cells, based on ribosome profiling (Akichika et al., 2019). Our ribosome profiling analysis of *PCIF1* knockout cells showed a slight repressive effect of m⁶Am on translation. Therefore, mRNAs modified with m⁶Am are efficiently translated, but Am-modified mRNAs are even slightly more efficiently translated. Regardless, both our study and the Akichika et al. study are consistent in finding a very minor effect on translation of m⁶Am-annotated mRNAs upon *PCIF1* depletion. It should be noted that effects of m⁶Am on both mRNA translation and mRNA stability are likely to depend on the sequence motif following m⁶Am and may be different during signaling or stress conditions that were not examined in our study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- METHOD DETAILS
 - Synthesis and characterization of synthetic oligonucleotides
 - Cell culture
 - Antibodies
 - Generation of *PCIF1* CRISPR knockout cells and over-expression cell lines
 - Protein expression and purification
 - In Vitro methyltransferase assays
 - UHPLC-MS/MS analysis
 - Cap-binding assay
 - Immunofluorescence

- Determination of relative m⁶Am, Am, and m⁶A levels by thin layer chromatography
- miCLIP
- miCLIP bioinformatic analyses
- Transcript 5' end cloning
- SLAM-seq
- Real-time PCR assay to determine transcript stability
- Ribosome profiling
- SLAM-seq bioinformatic analysis
- Statistics and software
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.molcel.2019.06.006>.

ACKNOWLEDGMENTS

We thank J. Lipton for reagents and A.O. Olarerin-George for assistance with data analysis. This work was supported by NCN 2017/24/T/NZ1/00170 (D.T.-S.) and NIH grants R01AG043550 and DP2AG055947 (E.L.G.) and R01DA037755 (S.R.J.).

AUTHOR CONTRIBUTIONS

K.B. performed biochemical analysis of *PCIF1* and generated *PCIF1* KO/OE cell lines; D.T.-S. and K.B. performed assays of *PCIF1* activity in cells; K.B., D.T.-S., and S.Z. performed and analyzed ribosome profiling data; D.T.-S. performed and analyzed SLAM-seq experiments; B.R.H. performed and analyzed miCLIP experiments; N.L. performed cap-binding experiments; K.T. performed experiments assessing the translational effect of *PCIF1* KO; T.G., J.-J.V., and F.D. synthesized capped and uncapped RNA; L.A. identified *PCIF1* as putative m⁶Am methyltransferase; and E.L.G. and S.R.J. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

S.R.J. is scientific founder of, advisor to, and owns equity in Gotham Therapeutics.

Received: November 29, 2018

Revised: April 8, 2019

Accepted: June 2, 2019

Published: July 3, 2019

REFERENCES

- Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H., and Kasukawa, T. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCh38 genome assemblies. *Sci. Data* 4, 170107.
- 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 N6-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* 363, eaav0080.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
- Barral, K., Sallamand, C., Petzold, C., Coutard, B., Collet, A., Thillier, Y., Zimmermann, J., Vasseur, J.J., Canard, B., Rohayem, J., et al. (2013). Development of specific dengue virus 2'-O- and N7-methyltransferase assays for antiviral drug screening. *Antiviral Res.* 99, 292–300.

- Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N⁶-adenosine)-methyltransferase. *RNA* 3, 1233–1247.
- Chen, D., and Patton, J.T. (2001). Reverse transcriptase adds nontemplated nucleotides to cDNAs during 5'-RACE and primer extension. *Biotechniques* 30, 574–580, 582.
- Dotz, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. *Biology (Basel)* 1, 895–905.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* 485, 201–206.
- Fan, H., Sakuraba, K., Komuro, A., Kato, S., Harada, F., and Hirose, Y. (2003). PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II. *Biochem. Biophys. Res. Commun.* 301, 378–385.
- Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 47, D766–D773.
- Furuichi, Y., LaFiandra, A., and Shatkin, A.J. (1977). 5'-terminal structure and mRNA stability. *Nature* 266, 235–239.
- Guyot, J.B., Grassi, J., Hahn, U., and Guschlbauer, W. (1993). The role of the preserved sequences of Dam methylase. *Nucleic Acids Res.* 21, 3183–3190.
- Herzog, V.A., Reichholz, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* 14, 1198–1204.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550.
- Iyer, L.M., Zhang, D., and Aravind, L. (2016). Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays* 38, 27–40.
- Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., and He, C. (2011). N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887.
- Keith, J.M., Ensinger, M.J., and Moss, B. (1978). HeLa cell RNA (2'-O-methyladenosine-N⁶-methyltransferase specific for the capped 5'-end of messenger RNA. *J. Biol. Chem.* 253, 5033–5039.
- Kruse, S., Zhong, S., Bodi, Z., Button, J., Alcocer, M.J., Hayes, C.J., and Fray, R. (2011). A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Sci. Rep.* 1, 126.
- Lavergne, T., Bertrand, J.R., Vasseur, J.J., and Debart, F. (2008). A base-labile group for 2'-OH protection of ribonucleosides: a major challenge for RNA synthesis. *Chemistry* 14, 9135–9138.
- Lavergne, T., Janin, M., Dupouy, C., Vasseur, J.J., and Debart, F. (2010). Chemical synthesis of RNA with base-labile 2'-o-(pivaloyloxymethyl)-protected ribonucleoside phosphoramidites. *Curr. Protoc. Nucleic Acid Chem. Chapter 3*, Unit3.19.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, Y., Song, M., and Kiledjian, M. (2011). Differential utilization of decapping enzymes in mammalian mRNA decay pathways. *RNA* 17, 419–428.
- Li, Y., Dai, J., Song, M., Fitzgerald-Bocarsly, P., and Kiledjian, M. (2012). Dcp2 decapping protein modulates mRNA stability of the critical interferon regulatory factor (IRF) IRF-7. *Mol. Cell. Biol.* 32, 1164–1172.
- Li, W., Wang, W., Uren, P.J., Penalva, L.O.F., and Smith, A.D. (2017). Riborex: fast and flexible identification of differential translation from Ribo-seq data. *Bioinformatics* 33, 1735–1737.
- Linder, B., Grozhik, A.V., Orlarier-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m⁶A and m⁶Am throughout the transcriptome. *Nat. Methods* 12, 767–772.
- Maden, B.E., Corbett, M.E., Heeney, P.A., Pugh, K., and Ajuh, P.M. (1995). Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* 77, 22–29.
- 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 m⁶A_m in the 5' cap controls mRNA stability. *Nature* 541, 371–375.
- McGlinchy, N.J., and Ingolia, N.T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods* 126, 112–129.
- Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646.
- Orlarier-George, A.O., and Jaffrey, S.R. (2017). MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. *Bioinformatics* 33, 1563–1564.
- Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. *Cell* 4, 387–394.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.
- Shah, A., Qian, Y., Weyn-Vanhenenryck, S.M., and Zhang, C. (2017). CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics* 33, 566–567.
- Shimotohno, K., Kodama, Y., Hashimoto, J., and Miura, K.I. (1977). Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl. Acad. Sci. USA* 74, 2734–2738.
- Shuman, S. (2002). What messenger RNA capping tells us about eukaryotic evolution. *Nat. Rev. Mol. Cell Biol.* 3, 619–625.
- Sun, H., Zhang, M., Li, K., Bai, D., and Yi, C. (2019). Cap-specific, terminal N⁶-methylation by a mammalian m⁶Am methyltransferase. *Cell Res.* 29, 80–82.
- Thillier, Y., Decroly, E., Morvan, F., Canard, B., Vasseur, J.J., and Debart, F. (2012). Synthesis of 5' cap-0 and cap-1 RNAs using solid-phase chemistry coupled with enzymatic methylation by human (guanine-N⁷)-methyl transferase. *RNA* 18, 856–868.
- Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. *Cell* 170, 564–576.e16.
- Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl. Acad. Sci. USA* 101, 11269–11274.
- Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* 15, R8.
- Wei, C., Gershowitz, A., and Moss, B. (1975). N⁶, O²-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. *Nature* 257, 251–253.
- Wei, C.M., Gershowitz, A., and Moss, B. (1976). 5'-terminal and internal methylated nucleotide sequences in HeLa cell mRNA. *Biochemistry* 15, 397–401.
- Willcock, D.F., Dryden, D.T., and Murray, N.E. (1994). A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* 13, 3902–3908.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389, 52–65.

Zhao, S., and Zhang, B. (2015). A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics* 16, 97.

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., and Fray, R.G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 20, 1278–1288.

Zhou, J., Wan, J., Shu, X.E., Mao, Y., Liu, X.M., Yuan, X., Zhang, X., Hess, M.E., Brüning, J.C., and Qian, S.B. (2018). N⁶-methyladenosine guides mRNA alternative translation during integrated stress response. *Mol. Cell* 69, 636–647.e7.

Zlatev, I., Lavergne, T., Debart, F., Vasseur, J.J., Manoharan, M., and Morvan, F. (2010). Efficient solid-phase chemical synthesis of 5'-triphosphates of DNA, RNA, and their analogues. *Org. Lett.* 12, 2190–2193.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mouse anti-FLAG M2	Sigma	F1804, RRID: AB_262044
rabbit anti-PCIF1	Abcam	ab205016, RRID: AB_2753142
mouse anti-β actin	Sigma	A5441, RRID: AB_476744
anti-eIF4E	Cell Signaling	2067, RRID: AB_2097675
anti-eIF4G	Cell Signaling	2498, RRID: AB_2096025
rabbit anti-GAPDH	Abcam	ab181602, RRID: AB_2630358
mouse anti-KAP1	Abcam	ab22553, RRID: AB_447151
rabbit anti-EEF2	Abcam	ab75748, RRID: AB_1310165
mouse anti-RACK1	Santa Cruz	B-3, RRID: AB_2247471
mouse anti-HSP70/72	Enzo Life Sciences	ADI-SPA-810-D, RRID: AB_2039260
rabbit anti-PARP1	Cell Signaling	9542, RRID: AB_10616513
rabbit anti-HSPA8	Cell Signaling	8444, RRID: AB_10831837
rabbit anti-m ⁶ A	Abcam	ab151230, RRID: AB_2753144
rabbit anti-ATF5	Abcam	ab60126, RRID: AB_940375
Bacterial and Virus Strains		
T7 Express lysY	New England Biolabs	C3010
Chemicals, Peptides, and Recombinant Proteins		
Cycloheximide	Sigma Aldrich	C4859
iodoacetamide	Sigma Aldrich	I1149
4-thiouridine (s ⁴ U)	Sigma Aldrich	T4509
Actinomycin D	Sigma Aldrich	A1410
Nuclease P1	Sigma Aldrich	N8630
Nuclease P1	Wako USA	145-08221
RppH	New England Biolabs	M0356
Fast AP	Thermo	EF0654
T4 PNK	New England Biolabs	M0201
rSAP	New England Biolabs	M0371
Apyrase	New England Biolabs	M0398
RNase I	Epicenter	N6901K
RNase T1	Thermo	AM2283
T4 ligase 2, truncated K227Q	New England BioLabs	M0351
CircLigase II ssDNA Ligase	Lucigen	CL9021K
AccuPrime SuperMix I	Thermo	12342010
SuperScript III Reverse Transcriptase	Thermo	18080044
S1 Nuclease	Thermo	EN 0321
Terminator exonuclease	Epicenter	TER51020
CIP	New England Biolabs	M0290
T4 RNA ligase 1	New England Biolabs	M0437
RNase H	New England Biolabs	M0297
Phusion Master Mix with HF buffer	New England Biolabs	M0531
SuperScript IV Reverse Transcriptase	Thermo	18090010
iQ SYBR Green supermix	Bio-Rad	1708880

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
oligo d(T) ₂₅ Magnetic mRNA isolation kit	New England Biolabs	S1550
Dynabeads oligo d(T) ₂₅	Thermo	61005
Quant-iT RiboGreen RNA Assay Kit	Thermo	R11490
Ribo-Zero Gold rRNA Removal Kit (Human, Mouse, Rat)	Illumina	MRZG12324
NEBNext Ultra Directional RNA Library Prep Kit for Illumina	New England Biolabs	E7420
Hydrophilic streptavidin magnetic beads	New England Biolabs	S1421
Deposited Data		
Raw and analyzed data	This Paper	GEO: GSE122948
Unprocessed and uncompressed imaging data	This Paper	https://doi.org/10.17632/mpfzjd7mj.1
Experimental Models: Cell Lines		
HEK293T	ATCC	CRL-3216
HeLa	ATCC	CCL-2
Oligonucleotides		
m7GpppAmGUGGACUAACCACCAU	This Paper and Trilink	N/A
m7GpppAGUGGACUAACCACCAU	This Paper and Trilink	N/A
pppAmGUGGACUAACCACCAU	This Paper	N/A
sgRNA1; 5'- CGGUUGAAAGACUC CCGUGG-3'	This Paper	N/A
sgRNA2; 5'- ACUUAACAUUCCU GCGGGG-3'	This Paper	N/A
Biotin 5' adaptor; biotin-GTTCAGAGT TCTACAGTCCGACGATC	This Paper	N/A
RT_ACOT7: cccgttctggctgttgcattgc	This Paper	N/A
RT_BNIP2: agcctggattcaatgtcaactac	This Paper	N/A
RT_YOD1: cggctggacagcccctgcaaaac	This Paper	N/A
PCR_5'adaptor: taatacactcactataggg tctcGGATCCcgacGTTTACAGATTCTA CAGTCCGACGATC	This Paper	N/A
PCR_ACOT7: GTAAAACGACGGCC AGTaaagaGAATTCggaacccgttctggct gttgcattgc	This Paper	N/A
PCR_BNIP2: GTAAAACGACGGC CAGTaaagaGAATTCggaagcctgg attcaatgtcaactac	This Paper	N/A
PCR_YOD1: GTAAAACGACGGCCAG TaagaGAATTCggaacgctggacagccc ctgcaaaac	This Paper	N/A
qPCR_ACTB_Fw: GAAGATCAAGA TCATTGCTCCTC	This Paper	N/A
qPCR_ACTB_Rv: ATCCACATCTG CTGGAAGG	This Paper	N/A
qPCR_RPS28_Fw: ATCAAGCTGGC TAGGGTAACC	This Paper	N/A
qPCR_RPS28_Rv: GGCCTTTGACAT TTCGGATGA	This Paper	N/A
qPCR_AKAP12_Fw: CATTGTCACAGAG GTTGGA	This Paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
qPCR_AKAP12_Rv: GCTGACTTAGTAG CCATCTC	This Paper	N/A
qPCR_NBR1_Fw: GTCTAATACCCTGAT GCTCCC	This Paper	N/A
qPCR_NBR1_Rv: GCAAATTCTCATCCA CAAATGC	This Paper	N/A
Recombinant DNA		
pSpCas9n(BB)-2A-Puro (PX459) V2.0 vector	Addgene	62988
pBABE-puro	Addgene	1764
pBABE-puro-PCIF1 WT	This Paper	N/A
pBABE-puro-PCIF1 SPPG	This Paper	N/A
pMD2.G	Addgene	12259
pUMVC	Addgene	8449
p3xFLAG-CMV-10	Sigma	E7658
p3xFLAG-CMV-10-PCIF1 WT	This Paper	N/A
p3xFLAG-CMV-10-PCIF1 SPPG	This Paper	N/A
Software and Algorithms		
Prism 5	Graphpad	N/A
MassHunter Suite	Agilent	N/A
NIS-Elements AR	Nikon	N/A
ImageQuantTL software	GE Healthcare	N/A
ImageJ (V2.0.0-rc-24/1.49 m)	NIH	N/A
Flexbar v2.5	Dodt et al., 2012	https://github.com/seqan/flexbar
pyCRAC	Webb et al., 2014	https://bitbucket.org/sgrann/pycrac
bwa v0.7.17	Li et al., 2009	http://bio-bwa.sourceforge.net/
CTK v1.1.3	Shah et al., 2017	https://github.com/chaolinzhanglab/ctk
deeptools v3.1.3	Ramirez et al., 2016	https://deeptools.readthedocs.io/en/develop/
bedtools v2.27.1	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
MetaPlotR	Olarerin-George and Jaffrey, 2017	https://github.com/olarerin/metaPlotR
DREME/MEME v5.0.2	Bailey et al., 2009	http://meme-suite.org/

LEAD CONTACT AND MATERIALS AVAILABILITY

Please contact E.L.G. (eric.greer@childrens.harvard.edu) or S.R.J. (srj2003@med.cornell.edu) for reagents and resources generated in this study.

METHOD DETAILS

Synthesis and characterization of synthetic oligonucleotides

The sequences of all the oligonucleotides used in this study are shown in [Figure 1B](#).

The synthetic RNA oligonucleotides, used in [Figure 1E](#), were chemically assembled on an ABI 394 DNA synthesizer (Applied Biosystems) from commercially available long chain alkylamine controlled-pore glass (LCAA-CPG) solid support with a pore size of 1000 Å derivatized through the succinyl linker with 5'-O-dimethoxytrityl-2'-O-Ac-uridine (Link Technologies). All RNA sequences were prepared using phosphoramidite chemistry at 1-μmol scale in Twist oligonucleotide synthesis columns (Glen Research) from commercially available 2'-O-pivaloyloxymethyl amidites (5'-O-DMTr-2'-O-PivOM-[U, C^{Ac}, A^{Pac} or G^{Pac}]-3'-O-(O-cyanoethyl-N,N-diisopropylphosphoramidite) ([Lavergne et al., 2010](#)) (Chemgenes). The 5'-terminal adenosine was methylated in 2'-OH (Am). The 5'-O-DMTr-2'-O-Me-A^{Pac}-3'-O-(O-cyanoethyl-N,N-diisopropylphosphoramidite) (Chemgenes) was used to introduce Am at

the 5' end of RNA. All oligoribonucleotides were synthesized using standard protocols for solid-phase RNA synthesis with the PivOM methodology (Lavergne et al., 2008).

After RNA assembly, the 5'-hydroxyl group of the 5'-terminal adenosine Am of RNA sequences, still anchored to solid support, was phosphorylated and the resulting *H*-phosphonate derivative was oxidized and activated into a phosphoroimidazolidate derivative to react with either pyrophosphate (for pppAm-RNA synthesis) (Zlatev et al., 2010) or guanosine diphosphate (for G-ppp-Am-RNA synthesis) (Thillier et al., 2012).

After deprotection and release from the solid support upon basic conditions (DBU then aqueous ammonia treatment for 4h at 37°C), all RNA sequences were purified by IEX-HPLC (Barral et al., 2013), they were obtained with high purity (> 95%) and they were unambiguously characterized by MALDI-TOF spectrometry.

N7 methylation of the purified G-ppp-Am-RNAs to give m⁷G-ppp-Am-RNAs was carried out quantitatively using human mRNA guanine-N7 methyltransferase and *S*-adenosylmethionine as previously described (Thillier et al., 2012). The oligonucleotides used in Figures 1C, 1D, 1G, and 1H were synthesized by TriLink.

Cell culture

HEK293T and HeLa cells were maintained in DMEM (11995-065, ThermoFisher Scientific) with 10% FBS and antibiotics (100 units/ml penicillin and 100 µg/ml of streptomycin) under standard tissue culture conditions. Cells were split using TrypLE Express (Life Technologies) according to the manufacturer's instructions. Mycoplasma contamination in cells were routinely tested by Hoechst staining.

Antibodies

Antibodies used for western blot analysis or immunostaining were as follows: mouse anti-FLAG M2 (F1804, Sigma), rabbit anti-PCIF1 (ab205016, Abcam), mouse anti-β actin (A5441, Sigma), anti-eIF4E (2067, Cell Signaling), anti-eIF4G (2498, Cell Signaling), rabbit anti-GAPDH (ab181602, Abcam), mouse anti-TRIM28 (ab22553, Abcam), rabbit anti-ATF5 (ab60126, Abcam), rabbit anti-EEF2 (ab33523), mouse anti-RACK1 (B-3, Santa Cruz), rabbit anti-PARP1 (9542, Cell Signaling), rabbit anti-HSPA8 (8444, Cell Signaling), mouse anti-HSP70/72 (C92F3A-5, Enzo Life Sciences). For m⁶A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP), rabbit anti-m⁶A (ab151230, Abcam) was used.

Generation of PCIF1 CRISPR knockout cells and overexpression cell lines

HEK293T and HeLa *PCIF1*-knockout cell lines were generated by CRISPR/Cas9 technology using two guide RNAs (gRNAs; 5'-CGGUUGAAAGACUCCCGUGG-3' and 5'-ACUUAACAUUCCUGCGGGG-3') designed to target the *PCIF1* genomic region between exon 8 and exon 17, that corresponds to the C-terminal catalytic domain. Double-stranded DNA oligonucleotides corresponding to the gRNAs were inserted into the pSpCas9n(BB)-2A-Puro (PX459) V2.0 vector (62988, Addgene). Equal amounts of the two gRNA plasmids were mixed and transfected into HEK293T and HeLa cells using FuGENE 6 (Promega). The transfected cells were then subjected to puromycin selection for three days and viable cells were used for serial dilution to generate single-cell clones. The genomic deletion was screened by PCR and was confirmed by Sanger sequencing. HEK293T and HeLa *PCIF1*-knockout lines used in this study contained a 4655 or 4656 nt homozygous deletion that removed the region between exon 8 and exon 17, including the stop codon, resulting in the disruption of *PCIF1* protein after P229 (aa 230-704). Loss of *PCIF1* protein expression was confirmed by western blot with anti-*PCIF1* antibody (Abcam).

Stable cell lines overexpressing *PCIF1* WT or catalytically inactive mutant proteins were generated through retroviral infection. The coding sequence of human *PCIF1* fused to a N-terminal 3X FLAG tag sequence that was cloned into the pBABE-puro retroviral vector (Addgene, 1764). Retroviral particles were generated in HEK293T cells through co-transfection of the packaging vectors pMD2.G (12259, Addgene) and pUMVC (8449, Addgene) with the appropriate pBABE-puro vectors. HEK293T and HeLa cells were infected with retroviral particles of pBABE-puro-3X-FLAG-*PCIF1* WT or pBABE-puro-3X-FLAG-*PCIF1* SPPG or control pBABE-puro empty vector, followed by puromycin selection (1 µg/ml).

Cells were maintained at 70%–80% confluency before harvesting for mRNA purification. Two rounds of poly(A) mRNA isolation from mammalian cells was performed using oligo d(T)₂₅ Magnetic mRNA isolation kit (NEB), according to the manufacturer's instructions.

Protein expression and purification

The coding sequence of human *PCIF1* was cloned as an in-frame fusion to the GST tagged vector pGEX-4T1. The catalytic site NPPF was mutated to APPA or SPPG thru site-directed mutagenesis using the Q5 mutagenesis kit (NEB), according to the manufacturer's instructions. Recombinant GST-*PCIF1* wild-type and catalytically inactive mutant proteins were expressed in *E. coli* T7 Express lysY. Overnight induction of protein expression was carried out with 0.5 mM IPTG at 18°C. Bacteria were harvested at 4000 rpm, 4°C and the cell pellet was resuspended in protein purification lysis buffer (50 mM Tris-HCl pH 7.5, 0.25 M NaCl, 0.1% Triton-X, 1 mM PMSF, 1 mM DTT, and protease inhibitors). The lysate was sonicated 6 times in 30 s on/off cycles and then centrifuged at 12,000 rpm for 20 minutes. Lysates were incubated with glutathione Sepharose 4B beads (Sigma). Proteins and beads were washed 3 times with protein purification lysis buffer before incubating the beads with elution buffer (12 mg/ml Glutathione in protein purification lysis buffer, pH 8.0) for 30 minutes. Eluates were dialyzed overnight at 4°C with enzyme storage buffer (40 mM Tris-HCl pH 8.0,

110 mM NaCl, 2.2 mM KCl, 1 mM DTT, 20% glycerol) and were subsequently stored at -80°C . Bradford assays and SDS-page gel electrophoresis followed by Coomassie staining was performed to determine integrity and quantity of purified proteins.

In Vitro methyltransferase assays

In vitro methylation reactions (50 μL) assaying PCIF1 activity against the m⁷G capped RNA oligonucleotides were performed in methylation reaction buffer (50 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT, 5% glycerol) supplemented with 160 μM SAM (NEB) using 50 nM GST-PCIF1 protein and 4 μM m⁷G capped oligonucleotide. Reactions were incubated for 10 minutes at 37°C , followed by heat inactivation for 20 minutes at 65°C and subsequent clean up and buffer exchange using Biospin P6 columns (Biorad). RNA oligonucleotides were decapped using 25 Units of RppH (NEB) in ThermoPol buffer for 3 hours at 37°C , followed by clean up and buffer exchange with Biospin P6 columns. Decapped RNA oligonucleotides were digested to nucleosides with 2 units of Nuclease P1 (Wako USA) at 37°C for 3 hours in a buffer containing 10 mM ammonium acetate pH 5.3, 2mM ZnCl₂ followed by treatment with 2 units of Fast Alkaline Phosphatase (FastAP, Thermo Scientific) in FastAP reaction buffer for 1 hour at 37°C . After digestion the sample volume was brought to 100 μL with ddH₂O followed by filtration using 0.22 μm Millex Syringe Filters (EMD Millipore). 5 μL of the filtered solution was analyzed by UHPLC-MS/MS.

Enzyme kinetics assaying PCIF1 activity against the m⁷G-Am and m⁷G-A RNA oligonucleotides were performed in methylation reaction buffer supplemented with 1.33 μM [³H]-SAM (Perkin Elmer) and 10 μM SAM (NEB), using 20 nM GST-PCIF1 protein and a range of concentrations of m⁷G-Am oligonucleotide for 2–4 min at 37°C in 50 μL reactions. The reactions were stopped with 0.1% TFA followed by removal of unincorporated [³H]-SAM with Biospin P30 columns (Biorad). The purified RNA oligonucleotide samples were then subjected to scintillation counting using a Perkin Elmer scintillation counter. The Michaelis-Menten curve and K_M value were determined using Graphpad Prism software.

UHPLC-MS/MS analysis

For the detection and quantification of internal m⁶A in mRNA, 500 ng of poly(A) mRNA was denatured at 70°C for 5 minutes followed by digestion to nucleotides using 20 units of S1 Nuclease (Thermo Scientific) in S1 Nuclease buffer for 2 hours at 37°C in 25 μL reactions. Nucleotides were then dephosphorylated to nucleosides by the addition of 2 units of Fast Alkaline Phosphatase (NEB) in FastAP reaction buffer for 1 hour at 37°C . After digestion the sample volume was brought to 100 μL with ddH₂O followed by filtration using 0.22 μm Millex Syringe Filters (EMD Millipore). 5 μL of the filtered solution was analyzed by LC-MS/MS.

For the detection and quantification of cap-adjacent m⁶Am in mRNA, 500 ng of poly(A) mRNA was decapped using 25 Units of RppH (NEB) in ThermoPol buffer for 3 hours at 37°C , followed by clean up and buffer exchange with Biospin P30 columns. Subsequently decapped RNA was denatured at 70°C for 5 minutes followed by digestion to nucleotides using 2 units of Nuclease P1 (Wako USA) in a buffer containing 10 mM ammonium acetate pH 5.3, 2mM ZnCl₂ for 3 hours at 37°C . Nucleotides were then dephosphorylated to nucleosides by the addition of 2 units of Fast Alkaline Phosphatase (NEB) in FastAP reaction buffer for 1 hour at 37°C . After digestion the sample volume was brought to 100 μL with ddH₂O followed by filtration using 0.22 μm Millex Syringe Filters. 5 μL of the filtered solution was analyzed by LC-MS/MS.

The separation of nucleosides was performed using an Agilent 1290 UHPLC system with a C18 reversed-phase column (2.1 \times 50 mm, 1.8 μm). The mobile phase A was water with 0.1% (v/v) formic acid and mobile phase B was methanol with 0.1% (v/v) formic acid. Online mass spectrometry detection was performed using an Agilent 6470 triple quadrupole mass spectrometer in positive electrospray ionization mode. Quantification of each nucleoside was accomplished in dynamic multiple reaction monitoring (dMRM) mode by monitoring the transitions of 268 \rightarrow 136 (A), 282 \rightarrow 136 (Am), 282 \rightarrow 150 (m⁶A), 296 \rightarrow 150 (m⁶Am), 244 \rightarrow 112 (C). The amounts of A, C, Am, m⁶A and m⁶Am in the samples were quantified using corresponding calibration curves generated with pure standards. m⁶Am and m⁶A levels in the RNA oligonucleotides after *in vitro* methylation reactions were normalized by cytidine concentration.

Cap-binding assay

Cells were lysed in buffer B (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 0.5 mM EDTA, 0.4% NP-40, 20% glycerol) supplemented with protease and phosphatase inhibitors (Roche), 1 mM dithiothreitol (DTT) and 80 units/ml RNasin (Promega). For pull down, 1–2.5 mg of total protein extract was first pre-cleared on Agarose beads (Jena Bioscience) followed by incubation with 25 μL m⁷GTP conjugated Agarose beads (Jena Bioscience) for 1 hour at 4°C . Following pull-down the beads were washed three times and the supernatant was removed and replaced by lysis buffer. Beads were incubated with 0.25 mM cap analog, m⁷G-ppp-A, or G-ppp-A, or water (mock) for 1 hour at 4°C . Supernatant (Eluate) was removed and diluted with Laemmli sample buffer. Beads were washed three times and re-suspended in Laemmli sample buffer. Samples were resolved on a 4%–15% Tris-HCl gradient gel (BioRad) and analyzed by western blotting using specific antibodies.

Immunofluorescence

Cells were grown on poly-L-lysine pre-coated coverslips that were sterilized under UV light for 30 minutes - 1 hour. Cells were rinsed in 1X phosphate-buffered saline (PBS) solution followed by fixation in ice-cold methanol at -20°C for 10 minutes. Coverslips were then washed 3 times with 1X PBS before being blocked for 30 minutes in 1% BSA in 1X PBS. Primary antibody was diluted 1/200 in 1% BSA 1X PBS and incubated for 1 hour at room temperature in a humidified chamber. Slides were subsequently rinsed 3 times and

washed 2 times for 15 minutes with 1% BSA in 1X PBS at room temperature before incubation with secondary antibody, diluted 1/200 in 1% BSA in 1X PBS, in a dark humidified chamber for 30 minutes at room temperature. Coverslips were then rinsed 3 times and washed 3 times for 15 minutes with 1% BSA in 1X PBS in the dark before being rinsed 3 times with ddH₂O. Coverslips were mounted using mounting medium containing DAPI. Image acquisition was carried out on a Nikon Eclipse Ti microscope (Nikon), using NIS-Elements AR software.

Determination of relative m⁶Am, Am, and m⁶A levels by thin layer chromatography

Levels of internal m⁶A in mRNA were determined by 2D-TLC essentially as previously described (Zhong et al., 2008). In brief, poly(A) RNA (100 ng) was digested with 2 units ribonuclease T1 (ThermoFisher Scientific) for 2h at 37°C in the presence of RNasin RNase Inhibitor (Promega). T1 cuts after every guanosine and exposes the 5'-hydroxyl of the following nucleotide, which can be A, C, U, or m⁶A. Thus, this method quantifies m⁶A in a GA sequence context. 5' ends were subsequently labeled with 10 units T4 PNK (NEB) and 0.4 mBq [γ -³²P] ATP at 37°C for 30 min followed by removal of the γ -phosphate of ATP by incubation with 10 units Apyrase (NEB) at 30°C for 30 min. After phenol-chloroform extraction and ethanol precipitation, RNA samples were resuspended in 10 μ l of DEPC-H₂O and digested to single nucleotides with 2 units of P1 nuclease (Sigma) for 1h at 60°C. 1 μ l of the released 5' monophosphates from this digest were then analyzed by 2D-TLC on glass-backed PEI-cellulose plates (MerckMillipore) as described previously (Kruse et al., 2011).

The protocol to detect the m⁶Am:Am ratio was based on the protocol developed by Fray and colleagues (Kruse et al., 2011), with some modifications. 300ng of poly(A) RNA was decapped with 15 units of RppH (NEB) for 3 h at 37°C. 5' monophosphates in the resulting RNA were removed by addition of 5 units of rSAP phosphatase (NEB) for 1 h at 37°C. Up to this point, all enzymatic reactions were performed in the presence of SUPERase In RNase Inhibitor (ThermoFisher Scientific). After phenol-chloroform extraction and ethanol precipitation, RNA samples were resuspended in 10 μ l of DEPC-H₂O and 5' ends were labeled using 30 units T4 PNK and 0.8 mBq [γ -³²P] ATP at 37°C for 30 min. PNK was heat inactivated at 65°C for 20 min and the reaction was passed through a P-30 spin column (Bio-Rad) to remove unincorporated isotope. 8 μ l of labeled RNA were then digested with 2 units of P1 nuclease (Sigma) for 1 h at 60°C. 2 μ l of the released 5' monophosphates from this digest were then analyzed by 2D-TLC on glass-backed PEI-cellulose plates (MerckMillipore) as described previously (Kruse et al., 2011).

Signal acquisition was carried out using a storage phosphor screen (GE Healthcare Life Sciences) at 200 μ m resolution and ImageQuantTL software (GE Healthcare Life Sciences). Quantification was carried out with ImageJ (V2.0.0-rc-24/1.49 m). For m⁶Am experiments, the m⁶Am:Am ratio was calculated. The use of this ratio has been described previously (Kruse et al., 2011). We confirmed that this assay is linear by spotting twice the sample material and confirming that the signal intensity doubles for the unmodified nucleotides (A, C, and U). Furthermore, exposure time of the TLC plates to the phosphor screen was chosen so the signal was not saturated. For m⁶A quantification, m⁶A was calculated as a percent of the total of the A, C, and U spots, as described previously (Jia et al., 2011). The use of relative ratios for each individual sample is important since it reduces the error derived from possible differences in loading. To minimize the effects of culturing conditions on the measured m⁶Am:Am ratios of each experimental group (e.g., control versus knockout), all replicates were processed in parallel to minimize any source of variability between samples being compared.

miCLIP

Total RNA from wild-type and *PCIF1* knockout HEK293T cells, and wild-type mouse embryonic fibroblasts, was extracted using TRIzol following the manufacturer's protocol. Any contaminating genomic DNA was degraded using DNase I and poly(A) RNA was isolated using two rounds of Dynabeads Oligo(dT)₂₅ capture. 10 μ g poly(A) RNA was then used as input for single nucleotide-resolution m⁶A mapping using the miCLIP protocol, as previously reported (Linder et al., 2015). Final libraries were amplified and subjected to 50-cycle paired-end sequencing on an Illumina HiSeq2500 at the Weill Cornell Medicine Epigenetic Core facility.

miCLIP bioinformatic analyses

The initial processing of raw FASTQ files was done as in the miCLIP protocol. Adapters and low quality nucleotides were first trimmed from paired reads using flexbar v2.5. The trimmed FASTQ file was then de-multiplexed using the pyBarcodeFilter.py script from the pyCRAC suite. The remainder of the random barcode was moved to the headers of the FASTQ reads using an awk script and PCR duplicates were removed using the pyCRAC pyDuplicateRemover.py script. Reads were aligned to hg38/mm10 using bwa v0.7.17 with the option "-n 0.06" as recommended in the CTK package. To identify m⁶A within the DRACH consensus, C to T transitions were extracted and the CIMS pipeline from the CTK package was used. Due to a high transition frequency in this dataset, putative m⁶A residues with an FDR < 0.1 and in a DRACH consensus were used as the final list of m⁶A in this study.

To identify putative m⁶Am sites, coverage of wild-type and *PCIF1* knockout samples were compared genome-wide using the bam-Compare tool from deeptools v3.1.3. In short, the genome was binned into 50 nt non-sliding windows and the coverage of reads in each was counted for each strand, discarding zero-coverage bins. This was normalized to the total number of reads in bins, per million (BPM) and the log₂ ratio of BPM+1 for wild-type to *PCIF1* knockout was calculated. A log₂ ratio threshold of 2 was chosen as the cutoff for each replicate. Adjacent bins passing threshold were merged using bedtools v2.27.1. The intersection of putative m⁶Am regions across replicates was taken using bedtools intersect, resulting in 2360 high-confidence m⁶Am peaks. To determine the precise m⁶Am nucleotide within these peaks, a combination of A to T transitions and a read pileup/drop-off method was used. In

PCIF1-dependent peaks with an A to T transition occurring at a frequency of 10% or greater, this A was selected as the m⁶A. For the remainder, a pileup/drop-off approach similar to the previous miCLIP criteria (Linder et al., 2015) was utilized. Here, the start nucleotide of each read (with respect to strand, i.e., the leftmost coordinate for + strand features and rightmost for – strand features) was extracted and piled up using the tag2cluster.pl script of the CTK package with the options “-s -v -maxgap -1.” Clusters of less than 5 reads were discarded, as were those that did not map to an A. When there was a single A-cluster in a PCIF1-dependent peak, this was selected as m⁶A. When more than one occurred, the most piled-up cluster of the two closest to the beginning of the peak (with respect to strand) was selected.

To generate metagenes, MetaPlotR (Olarerin-George and Jaffrey, 2017) was used. In all cases, the longest GENCODE transcript isoform for each gene was selected. For metaplots centered on reference annotations, the closest m⁶A to each feature was measured using bedtools closest and these distances were plotted as a histogram. Aligned reads in bigwig format and BED files with coordinates for m⁶A, m⁶Am, and CAGE peaks were used to generate genome tracks using pyGenomeTracks v1.0. Motif searches were performed using either DREME v5.0.2 or MEME v5.0.2. For functional annotation analyses of m⁶A and 5' UTR m⁶A genes, DAVID v6.8 was used specifying a background of all genes covered with at least 20 reads.

Transcript 5' end cloning

To validate called m⁶A as transcription start nucleotides, an adapted 5' RACE method was used. 10 μg poly(A) RNA was treated with Terminator exonuclease (Epicenter) and then CIP and rSAP (NEB) to remove 5' -monophosphate RNA following the manufacturer's recommendations. Half of this was then decapped using RppH (NEB) at a final concentration of 5 U per 100 ng RNA in 1X Thermopol buffer (NEB). The remaining half was incubated without RppH as a capped control, to control for any residual 5' -monophosphates due to RNA degradation. 75 pmol of a biotinylated adaptor (biotin-GTTCAGAGTTCTACAGTCCGACGATC) was then ligated onto the 5' end of the processed RNA using T4 RNA ligase 1 (NEB) in a 30 μl reaction containing 1X T4 RNA ligase buffer, 1 mM ATP, and 20% PEG-8000 for 3 hours at 23°C. This was then diluted to 200 μl in streptavidin bead wash buffer (SA wash; 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1 mM EDTA) and incubated with 0.5 mg hydrophilic streptavidin magnetic beads (NEB) for 15 minutes and room temperature. Beads were washed twice in SA wash buffer, then twice in annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA). 12.5 pmol of each gene-specific reverse transcription primers (STAR Methods) were annealed to RNA in 150 μl annealing buffer by heating to 90°C for 5 minutes and cooled to room temperature over 20 minutes. Beads were then resuspended in a 50 μl SuperScript III reverse transcription reaction (ThermoFisher Scientific) according to the manufacturer's protocol. Beads were washed and resuspended in a 50 μl RNase H (NEB) reaction and incubated for 30 minutes at 37°C. All reactions on beads were performed in a thermoshaker (15 s on 1100 RPM, 30 s off) to ensure beads remained in suspension. 100 μl wash buffer was added, heated at 75°C for 2 minutes, placed on beads and supernatant containing eluted cDNA immediately transferred to a fresh tube. The beads were resuspended in 50 μl and the elution repeated. Following an ethanol precipitation, 5% of the cDNA was used in 20 μl PCR reactions containing 1X Phusion master mix (NEB), 60% DMSO, 250 nM adaptor primer, and 250 nM gene-specific primer (see STAR Methods for primer sequences). 5 μl of this was then loaded on a 6% TBE-PAGE gel and visualized. Bands of the correct size that were absent in the capped control were then identified as m⁶A starting transcripts.

SLAM-seq

SLAM-seq was performed as described previously (Herzog et al., 2017) with minor modifications. HEK293T (WT and PCIF1 KO) cells (at 60% confluency) were incubated with cell culture growth medium supplemented with 25 μM 4-thiouridine (s⁴U) for 24 h (pulse phase). s⁴U incorporation was confirmed by HPLC analysis, as previously described (Herzog et al., 2017). The uridine chase was initiated by changing media containing 2.5 mM uridine (Sigma) and cells were collected for RNA extraction after 6 and 12 h. The 0 h sample were the cells that have completed the pulse with s⁴U, but without uridine-chase. Total RNA was extracted using RNAzol reagent (MRC) according to the manufacturer's instructions, maintaining reducing conditions to prevent oxidation of s⁴U (0.1 mM DTT final concentration). For thiol alkylation, a master mix (10 mM iodoacetamide, 50 mM NaPO₄ pH 8 and 50% DMSO) was prepared, centrifuged, and added to 20 μg of total RNA at 50°C for 15 min and then purified by ethanol precipitation. After that, two rounds of poly(A) mRNA enrichment was carried out with oligo d(T)₂₅ Magnetic Beads (NEB). Standard RNA-seq libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) following the instructions of the manufacturer. Sequencing was performed on a HiSeq2500 (Illumina) with 50 nucleotide reads.

Real-time PCR assay to determine transcript stability

Wild-type or PCIF1 knockout HEK293T cells were transfected with either empty vector or wild-type or SPPG mutant PCIF1 vectors for 48 hours and then treated with 5 μg/ml actinomycin D or vehicle (DMSO) for 8 hours. Total RNA was extracted using Trizol and 2 μg of this reverse transcribed using random hexamers and SuperScript IV (ThermoFisher Scientific) according to the manufacturer's protocol. RT-PCR was performed in 20 μl reactions containing 250 nM forward and reverse primers and iQ SYBR Green supermix (Bio-Rad) on an Eppendorf RealPlex2 RT-PCR machine. A delta cycle threshold (Ct) was calculated using the average Ct values across technical triplicates, by subtracting the geometric mean of two control genes (RPS28 and ACTB). A delta-delta Ct was then calculated by subtracting the vehicle control delta Ct value for each sample and untransformed to obtain relative abundances. Fold changes were tested for p < 0.1 by Student's t test.

Ribosome profiling

Ribosome profiling was performed as described previously (McGlinchy and Ingolia, 2017). In brief, wild-type and *PCIF1* knockout HEK293T cells were grown to ~70% confluence, washed twice with ice cold PBS supplemented with 50 μg/ml of cycloheximide (CHX) and collected by scraping. After pelleting, cells were resuspended in 400 μL lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT and 100 μg/ml CHX) After incubation on ice for 10 min, lysate was triturated 5 times through a 25-gauge needle and then lysate was centrifuged at 20,000 x g for 10 min. 5 μL of lysate was flash frozen and saved as input. To generate ribosome-protected fragments the lysates (30 μg) were first mixed with 200 μL DEPC-H₂O then incubated with 15 U RNase I for 45 min at room temperature. The reaction was stopped with 10 μL SUPERase[®]In RNase inhibitor. 0.9 mL of sucrose-supplemented lysis buffer was added to the digestion mixture and ultracentrifuged at 100,000 rpm, 4°C for 1 h. Pellets were resuspended in 300 μL of water and after phenol-chloroform extraction, precipitated with ethanol. The RNA was then run on a 15% 8 M urea TBE gel, stained with SYBR Gold, and a gel fragment between 17-34 nucleotides corresponding to ribosome-protected RNA was excised. RNA was eluted for 2 h at 37°C in 300 μL RNA extraction buffer (300 mM NaOAc pH 5.5, 1 mM EDTA, 0.25% v/v SDS) after crushing the gel fragment. RNA was ethanol precipitated and resuspended in 26 μL water and treated with RiboZero Gold kit. Libraries from RNA-protected fragments were generated as previously described in the protocol (Linder et al., 2015). In brief, the RNA fragments were dephosphorylated with T4 PNK for 1 h at 37°C in dephosphorylation buffer (70 mM Tris, pH 6.5, 10 mM MgCl₂, 1 mM DTT). The 3' adaptor was ligated using T4 RNA Ligase 2, truncated K227Q ligase (New England BioLabs) for 3h at 22°C. Ligated sRNAs were purified by ethanol precipitation, and reverse transcribed using the primers complementary to the 3' adaptor containing specific barcodes. After circularization with CircLigase II, cDNAs were relinearized by BamHI digestion and in the next step, PCR-amplified and subjected to Illumina HiSeq 2500 platform. Due to the similarity in size between ligated and unligated adapters, the libraries were gel purified.

RNA-Seq analysis was conducted using the ribosome profiling input material. Ribosomal RNAs were removed from the input RNA using the NEBNext rRNA Depletion Kit (NEB). Input RNA libraries were generated using the NEBNext Ultra Directional RNA library prep kit for Illumina (NEB). Libraries were sequenced using an Illumina HiSeq 2500 platform with 50 nt reads.

Ribosome footprint reads and corresponding RNA-Seq reads were processed essentially as described (Ingolia et al., 2012). Adaptors and short reads (< 17nt) were trimmed using FLEXBAR v2.5, demultiplexed using pyBarcodeFilter.py (pyCRAC software). PCR duplicates were collapsed by pyFastqDuplicateRemover.py script. Ribosomal RNA reads were removed by STAR aligner38. Remaining reads were then aligned to the hg38 genome with STAR v2.5.2a in a splicing-aware manner and using UCSC refSeq as a transcript model database (version from June 02/2014 downloaded from Illumina iGenomes). Two mismatches were allowed and only unique alignments were reported. Aligned reads were then counted on transcript regions using custom R scripts considering only transcripts with annotated 5' and 3' UTRs. Gene count tables generated from STAR were normalized using DESeq2 (R-Bioconductor). Translation efficiency was calculated using Riborex (Li et al., 2017), with pre-filtering for transcripts that had at least ten counted reads.

SLAM-seq bioinformatic analysis

Raw sequencing data were trimmed of adaptor sequences and filtered of reads with uncalled bases and reads < 17 nucleotides in length using Flexbar. Duplicate reads were further removed using pyFastqDuplicateRemover.py script and remaining reads were aligned to the human genome (GrCh38) using the STAR aligner.

To identify T → C conversions, aligned reads were analyzed using Rsamtools Pileup (version 1.27.16). This program was used to determine the frequency of each of the four nucleotides present in mapped reads at every genomic position with read coverage. After summation of all nucleotide mapped to each transcript, we selected only those with at least 100 T → C conversions at time point 0 h. Additionally, to select for those transcripts with a longer half-life, transcripts were filtered for those with at least 50 T → C conversions at time point 6h. The mRNA half-life for each transcript was calculated based on the equation:

$$t_{(1/2)} = - \frac{\ln 2}{\ln \left(\frac{N_t(t)}{N_0} \right)} t$$

Statistics and software

P-values were calculated with a two-tailed unpaired Student's *t* test or, for the comparison of more than two groups, with a one- or two-way ANOVA followed by Bonferroni's or Tukey's post-test. Reproducibility of half-life and translation efficiency measurements was assessed by calculating the Spearman correlation coefficient between replicates. Significance of list overlaps was calculated using hypergeometric probability.

DATA AND CODE AVAILABILITY

The accession number for the RNA-sequencing and ribosome profiling data reported in this paper is NCBI GEO: GSE122948.

Unprocessed and uncompressed imaging data is available at <https://doi.org/10.17632/rnptzjd7mj.1>.