Identification of the m^6Am Methyltransferase PCIF1 Reveals the Location and Functions of m^6Am in the Transcriptome

Graphical Abstract

Highlights

- PCIF1 is the N6-adenosine methylase that produces m^6Am in an m^7G cap-dependent manner
- PCIF1 depletion allows transcriptome-wide mapping of m^6A and m^6Am
- m^6Am mapping identifies alternative "internal" transcription start sites
- m^6Am increases stability of a subset of mRNAs and has no effect on translation

Authors

Konstantinos Boulias, Diana Toczydlowska-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

m^6Am is a prevalent mRNA modification occurring adjacent to the m^7G cap. Boulias, Toczydlowska-Socha, Hawley et al. identify PCIF1 as the m^6Am methyltransferase and perform transcriptome-wide mapping to distinguish m^6Am from m^6A and identify "internal" TSSs. m^6Am increases stability of a subset of mRNAs but has minimal effects on translation.

Boulias et al., 2019, Molecular Cell 75, 1–13
August 8, 2019 © 2019 Elsevier Inc. https://doi.org/10.1016/j.molcel.2019.06.006
Identification of the m^6^Am Methyltransferase PCIF1 Reveals the Location and Functions of m^6^Am in the Transcriptome

Konstantinos Boulias, 1,2,7 Diana Toczydlowska-Socha, 3,4,7 Ben R. Hawley, 3,7 Noa Liberman, 1,2 Ken Takashima, 1,2 Sara Zaccara, 3 Théo Guez, 6 Jean-Jacques Vasseur, 1 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 N^6^-methyladenosine (m^6^A), found within mRNAs, and N^6^,2^'-O-di-methyladenosine (m^6^Am), 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 m^6^Am. We find that PCIF1 binds and is dependent on the m^7^G cap. By depleting PCIF1, we generated transcriptome-wide maps that distinguish m^6^Am and m^6^A. We find that m^6^A and m^6^Am misannotations arise from mRNA isoforms with alternative transcription start sites (TSSs). These isoforms contain m^5^Am 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 m^6^Am-annotated mRNAs. The discovery of PCIF1 and our accurate mapping technique will facilitate future studies to characterize m^6^Am’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 N^6^-methyladenosine (m^6^A) at internal sites in mRNA (Bokar et al., 1997). At least 25% of mRNAs contain at least one m^6^A (Dominissini et al., 2012; Meyer et al., 2012).

N^6^-methylation also occurs on Am to form a dimethylated adenosine: N^6^,2^'-O-di-methyladenosine (m^6^Am) (Keith et al., 1978; Wei et al., 1975). Am is primarily located at the first transcribed nucleotide position in mRNAs, adjacent to the m^7^G 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 N^6^-methylation to m^6^Am. Because m^6^Am is present at the first transcribed nucleotide in ~30% of all cellular mRNAs, m^6^Am can affect the fate of a large subset of the transcriptome (Wei et al., 1975).

Recent studies have started to reveal the functions of m^6^Am. m^6^Am is enriched in mRNAs with high stability and translation efficiency (Mauer et al., 2017). Mechanistically, m^6^Am was shown to impair mRNA decapping mediated by DCP2, leading to increased stability of at least some m^6^Am-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 m^6^Am has a general role in promoting the high stability of m^6^Am-containing transcripts or whether m^6^Am confers mRNA stability to a subset of mRNAs that are degraded though selective decapping pathways.

Predicting the function of m^6^Am is complicated by the difficulty in determining whether an mRNA contains m^6^A or m^6^Am. Transcriptome-wide mapping of m^6^A and m^6^Am 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 (m^6^A individual-nucleotide-resolution
crosslinking and immunoprecipitation) (Linder et al., 2015), both map sites of 6mA rather than m6A or m5Am. The 6mA “peaks” are then interpreted to be either m6A or m5Am 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 m5Am because this nucleotide is exclusively found as the transcription start nucleotide. Nevertheless, it can be difficult to distinguish m5Am from m6A located within the 5′ UTR of mRNAs. As a result, previous maps of m5Am may have inaccuracies, which may make it difficult for predicting its function.

To definitively distinguish m5Am and m6Am in transcriptome-wide maps, depletion of either m5A or m6A would be required. m5A depletion cannot be readily achieved, as MetM is essential for survival in nearly all 341 cell lines that were screened (Tsherniak et al., 2017). The methyltransferase that generates m5Am is not known, but its depletion could enable the identification of the sites that are m5Am, because the remaining sites would be m6A.

Here, we describe the identification of PCIF1 as the methyltransferase that is responsible for generating m5Am 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 m5A in the transcriptome of PCIF1-deleted cells, we distinguish between m5Am and 5′ UTR m5A. We find numerous examples where previously annotated m5Am sites reflect m5A 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 high-confidence map of m5Am sites, we characterize the fate of m5Am-modified mRNAs in PCIF1 knockout cells and show that m5A has negligible effects on translation under basal conditions but is associated with increased stability of a subset of m5Am-initiated transcripts. Overall, our studies identify PCIF1 as the methyltransferase that generates m5Am in the transcriptome and provides revised transcriptome-wide maps that distinguish between m5A and m5Am.

RESULTS

Identification of PCIF1 as a Candidate m5Am-Forming Methyltransferase

Studies in the 1970s provided initial characterization of an enzymatic activity in HeLa cells that synthesizes m5Am (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 m5Am-forming enzyme, we performed a comparative bioinformatic analysis of orphan adenosine methyltransferases. These enzymes contain the [DNSH][PPYPFW] motif, which is present in all adenine N6-methyltransferases (Iyer et al., 2016). Among these putative adenosine methyltransferases, PCIF1 is notable because it emerged 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 (Furuoichi et al., 1977; Shimotohno et al., 1977; Shuman, 2002). The PCIF1 methyltransferase family is derived from the prokaryotic M.EcoKl/M.Taq1 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 (Akiyama 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 (m5G-ppp-Am-N16) (Figure 1B). We found that PCIF1 methylates Am in this RNA to produce m5Am, 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 methyltransferase 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 EcoKl 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 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 (m5G-ppp-A-N16) rather than Am. Using in vitro methyltransferase assays, we found that wild-type PCIF1, but not the SPPG or APPA PCIF1 mutant, was able to N6-methylate adenosine to m5A (Figure 1D). We next examined the rate and substrate preference of PCIF1 using a serial dilution of the m5G-ppp-Am- and the m5G-ppp-A-capped oligonucleotides and titrated 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 m5G moiety was required for methylation as PCIF1 efficiently methylated Am to m5Am in an m5G capped RNA (m5G-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′-CGGUUGAAAGACUCCCGUGG-3′ and 5′-ACUUAACAUAUCCUGCGGGG-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-m7Am 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 3x-FLAG-PCIF1 from HeLa cell extracts to m7GTP-conjugated beads. The beads were eluted with m7G-ppp-A or G-ppp-A. elf4E and elf4G were used to control for binding to m7G.
Figure 2. PCIF1 N6 Methylation 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.

(B) WT and PCIF1 KO cells were analyzed for m6A levels by anti-m6A immunoblotting. The graph shows the quantification of m6A levels in WT and PCIF1 KO cells. Error bars represent the mean ± SEM of three independent experiments. **p < 0.01, ***p < 0.001.

(C) PCIF1 KO cells were rescued with empty vector, WT PCIF1, or SPPG PCIF1. The graph shows the quantification of m6A levels in rescued cells. Error bars represent the mean ± SEM of three independent experiments. *p < 0.05, ****p < 0.0001.

(D) Internal m6A analysis in WT and PCIF1 KO cells. The graph shows the quantification of m6A levels in internal positions. Error bars represent the mean ± SEM of three independent experiments. ns = not significant, n.s. = not significant.

(E) Rescue of PCIF1 KO cells with empty vector, WT PCIF1, or SPPG PCIF1. The graph shows the quantification of m6A levels in rescued cells. Error bars represent the mean ± SEM of three independent experiments. *p < 0.05.

(F) Rescue of PCIF1 KO cells with empty vector, WT PCIF1, or SPPG PCIF1. The graph shows the quantification of m6A levels in rescued cells. Error bars represent the mean ± SEM of three independent experiments. *p < 0.05, ****p < 0.0001.

(F) Rescue of PCIF1 KO cells with empty vector, WT PCIF1, or SPPG PCIF1. The graph shows the quantification of m6A levels in rescued cells. Error bars represent the mean ± SEM of three independent experiments. *p < 0.05, ****p < 0.0001.

(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 mG-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-cap 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, and 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 m6A (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 m6Am 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 m6A 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 m6A can also be followed by cytosine, C-to-T transitions alone are not sufficient to distinguish m6A from m6Am. Peaks caused by m6A 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 m6A (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 m6A identifications.
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 m^6Am peaks may have been discarded or thought to be m^6A based on their location away from a TSS.

We therefore performed miCLIP in control and PCIF1 knockout cells to distinguish m^6A and m^6Am. In control cells, reads were enriched in the vicinity of the stop codon as well as the TSS, which is generally assumed to reflect m^6A and m^6Am, 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 m^6Am residue.

A motif analysis of significant peaks showed the DRACH m^6A 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 m^6A is the most common modification mapped in both datasets, as expected.

To identify m^6Am 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 KDELR2 show called m^6A sites (false discovery rate [FDR] < 0.1) and m^6Am sites indicated by red circles and blue triangle, respectively. Zoomed insets show m^6A peaks can be distinguished from nearby m^6A sites.

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 m^6A peaks.

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 m^6Am 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 m^6A consensus motif, and C-to-T transitions are detected for RACK1 (Figure 3D), suggesting that these sites are actually m^6A.

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 m^6A peak.

Overall, only 60.2% of genes that had previously been annotated as m^6Am (Mauer et al., 2017) were validated as m^6Am 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 m\textsuperscript{6}Am identification.

### A High-Confidence Transcriptome-wide Map of m\textsuperscript{6}A and m\textsuperscript{6}Am Based on PCIF1 Depletion

To create a high-confidence map of all m\textsuperscript{6}Am 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 m\textsuperscript{6}A (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 m\textsuperscript{6}Am 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 m\textsuperscript{6}Am 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 m\textsuperscript{6}A within the PCIF1-dependent peaks. Previously, we observed antibody-induced A-to-T transitions at the m\textsuperscript{6}A site in miCLIP (Linder et al., 2015). We confirmed that A-to-T transitions are readily detected at known m\textsuperscript{6}A and m\textsuperscript{6}Am throughout the transcriptome (Figures S2C and S2D). Therefore, we used a 10% A-to-T transition rate to identify the m\textsuperscript{6}A 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 m\textsuperscript{6}A sites that were called when using these methods separately (Figure S2F).

Overall, the 2,350 m\textsuperscript{6}Am 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 m\textsuperscript{6}A nucleotide revealed the BCA motif, with A representing the m\textsuperscript{6}A and BC representing upstream genomic nucleotides (B = C, G, or T), as reported previously for m\textsuperscript{6}Am (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 m\textsuperscript{6}A occurs in specific sequence contexts within mRNA transcripts (Figure S2G).

Next, we mapped m\textsuperscript{6}A in the 5’ UTR. C-to-T transitions in a DRACH consensus were used to call on average 44,025 m\textsuperscript{6}A sites in 10,383 genes based on the miCLIP protocol (Linder et al., 2015). This identified 399 5’ UTR m\textsuperscript{6}A sites that were robustly called across all datasets (Table S2).

We next asked whether mRNAs with 5’ UTR m\textsuperscript{6}A and mRNAs that contain m\textsuperscript{6}Am are linked to different cellular processes, based on our updated m\textsuperscript{6}A and m\textsuperscript{6}Am 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 m\textsuperscript{6}A associated with processes such as transcription and cell division, and m\textsuperscript{6}Am is primarily associated with splicing (Figures S2H and S2I; Table S3).

### ATF4 Contains a m\textsuperscript{6}A Rather Than m\textsuperscript{6}A in Its 5’ UTR

We next wanted to understand whether our revised map of m\textsuperscript{6}Am and m\textsuperscript{6}A can identify transcripts with misannotated modified nucleotides. A 5’ UTR m\textsuperscript{6}A 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. m\textsuperscript{6}A 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 m\textsuperscript{6}A.

Based on the location of the peak, we asked whether it instead reflects m\textsuperscript{6}A rather than m\textsuperscript{6}A. 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 m\textsuperscript{6}A (Figure S3A).

The role of m\textsuperscript{6}A 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 m\textsuperscript{6}A site (Figure S3B). Conceivably, human cells exhibit stress-induced regulation of ATF4 translation through an m\textsuperscript{6}Am-independent pathway and mouse cells utilize an m\textsuperscript{6}A-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 m\textsuperscript{6}A residue. In comparison, there were low levels of 6mA reads throughout the transcript body, suggesting either background reads or low stoichiometry m\textsuperscript{6}A sites (Figure S3D).

Overall, these data suggest that ATF4 contains a m\textsuperscript{6}A at the TSS but no prominent m\textsuperscript{6}A site within the 5’ UTR as previously reported (Zhou et al., 2018). Thus, a role for a 5’ UTR m\textsuperscript{6}A in regulating ATF4 translation seems unlikely. Overall, these data demonstrate the ease with which m\textsuperscript{6}A and m\textsuperscript{6}Am sites can be confused for each other.

### Identification of Internal 6mA Sites that Reflect Transcription-Start m\textsuperscript{6}A

We noticed two unusual features in our mapping results. First, not all m\textsuperscript{6}A sites mapped to regions within annotated mRNA
transcripts. Second, the m^6^Am metagene showed that, although 94% of m^6^Am 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 2E).

We considered that these findings could be due to m^6^Am 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 m^6^Am metaplot relative to RefSeq-annotated TSSs (Figure 4A). We observed 16.7% of m^6^Am sites mapping within 250 nt upstream of annotated TSSs, suggesting that some m^6^Am occurs in isoforms with upstream TSSs. We similarly observed m^6^Am upstream of TSSs using GENCODE (Frankish et al., 2019) transcript annotations (Figure 4B). The FANTOM5 promoter-level expression atlas (Abegassia 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 m^6^Am sites, supporting the idea that these m^6^Am sites are indeed TSSs (Figure 4C).

TSS heterogeneity likely explains why some m^6^Am 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 m^6^Am (Figure 4D). This m^6^Am likely reflects an isoform with a TSS located at this m^6^Am site, based on its overlap with a CAGE peak (Figure 4D). Thus, the presence of m^6^Am within the 5’ UTR likely reflects TSS heterogeneity rather than “internal” m^6^Am nucleotides.

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

To test this idea further, we performed a metagene analysis on m^6^Am 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 m^6^Am sites in the coding sequence and 3’ UTR are located at or near CAGE sites. Additionally, these m^6^Am 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 m^6^Am mapped to canonical TSSs (Linder et al., 2015). 5’ RACE confirmed that our called m^6^Am sites are indeed transcription start nucleotides (Figure S3E). Overall, these data further suggest that m^6^Am is not internally located within transcripts but is instead found at the TSSs.

Approximately 8% of m^6^Am 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 m^6^A. These data highlight the value of using PCIF1 depletion to validate the transcriptome-wide m^6^Am and m^6^A maps.

### m^6^Am Correlates with Enhanced Translation, Expression, and Stability of mRNAs

In our previous studies, we found that m^6^Am 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 m^6^Am 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 m^6^Am for this analysis, and mRNAs previously annotated as m^6^Am were re-annotated as Am based on our revised mapping data. Analysis of mRNA expression and half-lives showed that transcripts that begin with m^6^Am are indeed more highly expressed and stable than mRNAs with other start nucleotides. Notably, m^6^Am 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 m^6^Am correlates with an overall increase in mRNA stability and that m^6^Am is the predominant starting nucleotide on “outlier” mRNAs with unusually high stability and expression. To determine whether the N6-methyl in m^6^Am 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 m^6^Am (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 m^6^Am 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 m^6^Am for transcript stability.

Previously, we found that m^6^Am-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 m^6^Am 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 m^6^Am 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 m^6^Am compared to transcripts annotated to begin with other nucleotides (Figure S4F). In agreement with the modest
Figure 4. Internally Mapped m^6^Am Sites Reflect m^6^Am in mRNA Isoforms with Alternative TSSs

(A) A metaplot centered on the closest RefSeq TSS for each called m^6^Am site shows most m^6^Am sites are found downstream of the annotated start site, but not at the annotated start site. The proportion of m^6^Am directly at the annotated TSS or up- or downstream is shown.

(B) A metaplot analysis of m^6^Am 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 m^6^Am site to the closest CAGE peak in the FANTOM5 database shows that m^6^Am sites are indeed TSSs. Here, the overlap of m^6^Am sites is highest, suggesting that m^6^Am sites are selectively localized to TSSs and not internal nucleotides within mRNA.

(D) The m^6^Am 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. m^6^Am peaks that appear within the 5^'^ UTR reflect m^6^Am in transcript isoforms with alternative TSSs. The exact m^6^Am site (blue triangle) was determined using the A-to-T transition within the PCIF1-dependent peak.

(E) Most m^6^Am sites are found in the annotated 5^'^ UTR of transcripts.

(F) The internally mapping m^6^Am in YOD1 derives from a TSS of a YOD1 transcript isoform. The m^6^Am 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 m^6^Am sites show overlap with CAGE data, indicating that m^6^Am 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.
effects of PCIF1 depletion on translation rates, we found that levels of proteins encoded by several m^6Am-modified mRNAs remained largely unchanged in PCIF1 KO cells (Figure S4 G).

Together, these experiments suggest that, under the conditions used in these experiments, N6 methylation does not mediate the increase in translation efficiency of m^6Am-initiated mRNAs in HEK293T cells.

**DISCUSSION**

A major challenge when mapping m^6A and m^6Am 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 m^6Am-forming methyltransferase and by depleting PCIF1 to definitively identify m^6Am sites, we present a revised annotation of m^6Am and m^6A 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 m^6A. The identification and characterization of PCIF1 coupled with precise m^6Am annotations generated by PCIF1 depletion will facilitate the identification of functions for m^6Am.

Our studies provide insights into the function of m^6Am. Using our new high-confidence m^6Am map, we find that m^6Am is found on unusually stable and highly abundant transcripts in cells. However, depletion of m^6Am by PCIF1 knockout does not markedly impair the stability of these unusual mRNAs under basal conditions. This suggests that m^6Am does not account for the stability of these unusual mRNAs. m^6Am therefore is likely to co-occur with other transcript features that confer these unusual properties to these mRNAs.
likely to have different sensitivities to decapping or bind to different m^6Am 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, m^6Am functions are more likely to be revealed by analysis of mRNAs binned based on m^6Am sequence contexts.

Another important factor that might affect whether m^6Am has a stabilizing effect is whether the mRNA utilizes DCP2 or potentially other m^6Am-sensitive decapping mechanisms. Previous studies showed that m^6Am 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, m^6Am may stabilize mRNAs when DCP2-dependent pathways are activated. Overall, our studies show that m^6Am acts in a transcript-selective manner rather than a general mRNA-stabilizing modification.

Although our study focused on mRNA stability, another study examined m^6Am mRNA abundance and found no effects upon PCIF1 depletion (Akichika et al., 2019). This might reflect compensatory upregulation of m^6Am mRNAs. Notably, Akichika et al. examined all mRNAs rather than specific subsets of m^6Am-annotated mRNAs.

Two recent studies also reported PCIF1 as the m^6Am cap-dependent methyltransferase (Akichika et al., 2019; Sun et al., 2019). Akichika et al. found that m^6Am 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 m^6Am on translation. Therefore, mRNAs modified with m^6Am 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 m^6Am-annotated mRNAs upon PCIF1 depletion. It should be noted that effects of m^6Am on both mRNA translation and mRNA stability are likely to depend on the sequence motif following m^6Am 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 overexpression cell lines
  - Protein expression and purification
  - In Vitro methyltransferase assays
  - UHPLC-MS/MS analysis
  - Cap-binding assay
  - Immunofluorescence
  - Determination of relative m^6Am, Am, and m^6A 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 R00AG043550 and DP2AG05947 (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 m^6Am 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**


### STAR METHODS

**KEY RESOURCES TABLE**

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibodies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mouse anti-FLAG M2</td>
<td>Sigma</td>
<td>F1804, RRID: AB_262044</td>
</tr>
<tr>
<td>rabbit anti-PCIF1</td>
<td>Abcam</td>
<td>ab205016, RRID: AB_2753142</td>
</tr>
<tr>
<td>mouse anti-β-actin</td>
<td>Sigma</td>
<td>A5441, RRID: AB_476744</td>
</tr>
<tr>
<td>anti-eIF4E</td>
<td>Cell Signaling</td>
<td>2067, RRID: AB_2097675</td>
</tr>
<tr>
<td>anti-eIF4G</td>
<td>Cell Signaling</td>
<td>2498, RRID: AB_2096025</td>
</tr>
<tr>
<td>rabbit anti-GAPDH</td>
<td>Abcam</td>
<td>ab181602, RRID: AB_2630358</td>
</tr>
<tr>
<td>mouse anti-KAP1</td>
<td>Abcam</td>
<td>ab225553, RRID: AB_447151</td>
</tr>
<tr>
<td>rabbit anti-EF2</td>
<td>Abcam</td>
<td>ab75748, RRID: AB_1310165</td>
</tr>
<tr>
<td>mouse anti-RACK1</td>
<td>Santa Cruz</td>
<td>B-3, RRID: AB_2247471</td>
</tr>
<tr>
<td>mouse anti-HSP70/72</td>
<td>Enzo Life Sciences</td>
<td>ADI-SPA-810-D, RRID: AB_2039260</td>
</tr>
<tr>
<td>rabbit anti-PARP1</td>
<td>Cell Signaling</td>
<td>9542, RRID: AB_10616513</td>
</tr>
<tr>
<td>rabbit anti-HSPA8</td>
<td>Cell Signaling</td>
<td>8444, RRID: AB_10831837</td>
</tr>
<tr>
<td>rabbit anti-m^6^A</td>
<td>Abcam</td>
<td>ab151230, RRID: AB_2753144</td>
</tr>
<tr>
<td>rabbit anti-ATF5</td>
<td>Abcam</td>
<td>ab60126, RRID: AB_940375</td>
</tr>
<tr>
<td><strong>Bacterial and Virus Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T7 Express lysY</td>
<td>New England Biolabs</td>
<td>C3010</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycloheximide</td>
<td>Sigma Aldrich</td>
<td>C4859</td>
</tr>
<tr>
<td>iodoacetamide</td>
<td>Sigma Aldrich</td>
<td>I1149</td>
</tr>
<tr>
<td>4-thiouridine (s^4U)</td>
<td>Sigma Aldrich</td>
<td>T4509</td>
</tr>
<tr>
<td>Actinomycin D</td>
<td>Sigma Aldrich</td>
<td>A1410</td>
</tr>
<tr>
<td>Nuclease P1</td>
<td>Sigma Aldrich</td>
<td>N8630</td>
</tr>
<tr>
<td>Nuclease P1</td>
<td>Wako USA</td>
<td>145-08221</td>
</tr>
<tr>
<td>RppH</td>
<td>New England Biolabs</td>
<td>M0356</td>
</tr>
<tr>
<td>Fast AP</td>
<td>Thermo</td>
<td>EF0654</td>
</tr>
<tr>
<td>T4 PNK</td>
<td>New England Biolabs</td>
<td>M0201</td>
</tr>
<tr>
<td>rSAP</td>
<td>New England Biolabs</td>
<td>M0371</td>
</tr>
<tr>
<td>Apyrase</td>
<td>New England Biolabs</td>
<td>M0398</td>
</tr>
<tr>
<td>RNase I</td>
<td>Epicenter</td>
<td>N6901K</td>
</tr>
<tr>
<td>RNase T1</td>
<td>Thermo</td>
<td>AM2283</td>
</tr>
<tr>
<td>T4 ligase 2, truncated K227Q</td>
<td>New England BioLabs</td>
<td>M0351</td>
</tr>
<tr>
<td>CircLigase II ssDNA Ligase</td>
<td>Lucigen</td>
<td>CL9021K</td>
</tr>
<tr>
<td>AccuPrime SuperMix I</td>
<td>Thermo</td>
<td>12342010</td>
</tr>
<tr>
<td>SuperScript III Reverse Transcriptase</td>
<td>Thermo</td>
<td>18080044</td>
</tr>
<tr>
<td>S1 Nuclease</td>
<td>Thermo</td>
<td>EN 0321</td>
</tr>
<tr>
<td>Terminator exonuclease</td>
<td>Epicenter</td>
<td>TERS020</td>
</tr>
<tr>
<td>CIP</td>
<td>New England Biolabs</td>
<td>M0290</td>
</tr>
<tr>
<td>T4 RNA ligase 1</td>
<td>New England Biolabs</td>
<td>M0437</td>
</tr>
<tr>
<td>RNase H</td>
<td>New England Biolabs</td>
<td>M0297</td>
</tr>
<tr>
<td>Phusion Master Mix with HF buffer</td>
<td>New England Biolabs</td>
<td>M0531</td>
</tr>
<tr>
<td>SuperScript IV Reverse Transcriptase</td>
<td>Thermo</td>
<td>18090010</td>
</tr>
<tr>
<td>iQ SYBR Green supermix</td>
<td>Bio-Rad</td>
<td>1708880</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oligo d(T)$_{25}$ Magnetic mRNA isolation kit</td>
<td>New England Biolabs</td>
<td>S1550</td>
</tr>
<tr>
<td>Dynabeads oligo d(T)$_{25}$</td>
<td>Thermo</td>
<td>61005</td>
</tr>
<tr>
<td>Quant-iT RiboGreen mRNA Assay Kit</td>
<td>Thermo</td>
<td>R11490</td>
</tr>
<tr>
<td>Ribo-Zero Gold rRNA Removal Kit (Human, Mouse, Rat)</td>
<td>Illumina</td>
<td>MRZG12324</td>
</tr>
<tr>
<td>NEBNext Ultra Directional RNA Library Prep Kit for Illumina</td>
<td>New England Biolabs</td>
<td>E7420</td>
</tr>
<tr>
<td>Hydrophilic streptavidin magnetic beads</td>
<td>New England Biolabs</td>
<td>S1421</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw and analyzed data</td>
<td>This Paper</td>
<td>GEO: GSE122948</td>
</tr>
<tr>
<td>Unprocessed and uncompressed imaging data</td>
<td>This Paper</td>
<td><a href="https://doi.org/10.17632/rnpfzjd7mj.1">https://doi.org/10.17632/rnpfzjd7mj.1</a></td>
</tr>
<tr>
<td><strong>Experimental Models: Cell Lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK293T</td>
<td>ATCC</td>
<td>CRL-3216</td>
</tr>
<tr>
<td>HeLa</td>
<td>ATCC</td>
<td>CCL-2</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m7GpppAmGUGGACUAACCACCAU</td>
<td>This Paper and Trilink</td>
<td>N/A</td>
</tr>
<tr>
<td>m7GpppAGUGGACUAACCACCAU</td>
<td>This Paper and Trilink</td>
<td>N/A</td>
</tr>
<tr>
<td>pppAmGUGGACUAACCACCAU</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>sgRNA1; 5'-CGGUUGAAAGACUCGCCUGG-3'</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>sgRNA2; 5'-ACUUAACAAUAUCCGCCGGG-3'</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Biotin 5' adaptor; biotin-GTTCAGAGT TCTACGTCCGACGATC</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>RT_ACOT7: cccgctctgctgtgcaatgc</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>RT_BNIP2: agcctggattcaatgtcaactac</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>RT_YOD1: cggcctgcagccctgcaaacc</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>PCR_5'adaptor: taatagctactaataggg tctcGTCTCCacGTGAGATGCTCTCAGTGCCAGATC</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>PCR_ACOT7: GTAAAACGACGGCCAGTGaagaGGTctgctgtgtgcaatgc</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>PCR_BNIP2: GTAAAACGACGGCCAGTAGTaagaGGTcagccctgcaatgc</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>PCR_YOD1: GTAAAACGACGGCCAGTAGTaagaGGTcagccctgcaatgc</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>qPCR_ACTB_Fw: GAAGATCAAGATCATTGCTCCTC</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>qPCR_ACTB_Rv: ATCCACATCTGCTGGAAGG</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>qPCR_RPS28_Fw: ATCAAGCTGGTACCTTGACGATC</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>qPCR_RPS28_Rv: GGCTTTTACATGATCGGAGG</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
<tr>
<td>qPCR_AKAP12_Fw: CATTGCTCACAGAGGTGGGA</td>
<td>This Paper</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
**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-mmol scale in Twist oligonucleotide synthesis columns (Glen Research) from commercially available 2'-O-pivaloyloxymethyl amidites (5'-O-DMTr-2'-O-PivOM-[U, C<sup>Ac</sup>, A<sup>Pac</sup> or G<sup>Pac</sup>]-3'-O-(O-cyanoethyl-N,N-diisopropylphosphoramidite) (Chemgenes). The 5'-terminal adenosine was methylated in 2'-OH (Am). The 5'-O-DMTr-2'-O-Me-A<sup>Pac</sup>-3'-O-(O-cyanoethyl-N,N-diisopropylphosphoramidite) (Chemgenes) was used to introduce Am at

---

**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.
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 phosphorimidazolodiazolate 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-EF2 (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’- CGGUUGAAGACUCCCGUGG-3’ and 5’- ACUUAACAUAUCCUGCGGG-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 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% confluence 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 m^7^G 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 m^7^G 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 m^7^G-Am and m^7^G-A RNA oligonucleotides were performed in methylation reaction buffer supplemented with 1.33 μM [³²H]-SAM (Perkin Elmer) and 10 μM SAM (NEB), using 20 nM GST-PCIF1 protein and a range of concentrations of m^7^G-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 [³²H]-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 m^5^A 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 m^6^Am 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 × 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→136 (A), 282→136 (Am), 282→150 (m^5^Am), 296→150 (m^6^Am), 244→112 (C). The amounts of A, C, Am, m^5^A and m^6^Am in the samples were quantified using corresponding calibration curves generated with pure standards. m^6^Am and m^5^A 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 m^7^GTP 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, m^7^G-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. 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 µl of DEPC-H₂O and digested to single nucleotides with 2 units of P1 nuclease (Sigma) for 1h at 60°C. 1 µl 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:A 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 SAP 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 µl 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 µl of labeled RNA were then digested with 2 units of P1 nuclease (Sigma) for 1 h at 60°C. 2 µl 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 µm 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:A 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:A 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)₉₉ capture. 10 µg 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 subsequently labeled with 0.4 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 µl of labeled RNA were then digested with 2 units of P1 nuclease (Sigma) for 1h at 60°C. 2 µl 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 µm 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:A 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 m⁶Am quantification, m⁶Am 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:A 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 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 m⁶Am within the DRACH consensus, C to T transitions were subsequently labeled with 0.4 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 µl of labeled RNA were then digested with 2 units of P1 nuclease (Sigma) for 1h at 60°C. 2 µl 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 µm 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:A 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 m⁶Am quantification, m⁶Am 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:A 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.
PCIF1-dependent peaks with an A to T transition occurring at a frequency of 10% or greater, this A was selected as the m^6Am. 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 m^6Am. 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 (Olaterin-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 m^6Am 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 m^6A, m^6Am, 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 m^6Am and 5'-UTR m^6A 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 m^6Am as transcription start nucleotides, an adapted 5'/C14 buffer was added, heated at 75°C molecular cell reagent (MRC) according to the manufacturer’s instructions, maintaining reducing conditions to prevent oxidation of s4U (0.1 mM with respect to strand) was selected.

0 with Terminator exonuclease (Epicenter) and then CIP and rSAP (NEB) to remove 5′ selected as m^6Am. 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 (Olaterin-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 m^6Am 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 m^6A, m^6Am, 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 m^6Am and 5'-UTR m^6A genes, DAVID v6.8 was used specifying a background of all genes covered with at least 20 reads.

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 (s^4U) for 24 h (pulse phase). s^4U 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 Rnaloz reagent (MRC) according to the manufacturer’s instructions, maintaining reducing conditions to prevent oxidation of s^4U (0.1 mM DTT final concentration). For thiol alkylation, a master mix (10 mM iodoacetamide, 50 mM NaPO_4 pH 8 and 50% DMSO) was prepared, centrifuged, and added to 20 μg of total RNA at 50°C for 15 minutes and room temperature for 30 minutes. 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 m^6Am starting transcripts.

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 (McGlincy 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 RNA 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:

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

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/rnpfzjd7mj.1.