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

Konstantinos Boulias, Diana Toczydlowska-Socha, Ben Hawley, Noa Liberman, Ken Takashima, Sara Zaccara, Théo Guez, Jean-Jacques Vasseur, Françoise Debart, L. Aravind, et al.

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 Toczydłowska-Socha[3,4,7], Ben R. Hawley[3,7], Noa Liberman-Isakov[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,#]

[1] Division of Newborn Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA
[2] Department of Pediatrics, Harvard Medical School, Boston MA 02115, USA
[3] Department of Pharmacology, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA
[4] Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland
[5]IBMM, Université de Montpellier, CNRS, ENSCM, Montpellier, France.
[6] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
[7] Co-first authors

[#] Correspondence should be addressed to lead contact Eric L. Greer (eric.greer@childrens.harvard.edu) or Samie R. Jaffrey (srj2003@med.cornell.edu)

Summary
mRNAs are regulated by nucleotide modifications that influence their cellular fate. Two of the most abundant modified nucleotides are N^6-methyladenosine (m^6A), found within mRNAs, and N^6,2′-*O*-dimethyladenosine (m^6Am), which is found at the first-transcribed nucleotide. A long-standing challenge has been distinguishing these similar modifications in transcriptome-wide mapping studies. Here we identify and biochemically characterize, PCIF1, the methyltransferase that generates m^6Am. We find that PCIF1 binds and is dependent on the m^7G cap. By depleting PCIF1, we definitively identified m^6Am sites and generated transcriptome-wide maps that are selective for m^6Am and m^6A. We find that m^6A and m^6Am misannotations largely arise from mRNA isoforms with alternate transcription-start sites (TSSs). These isoforms contain m^6Am that appear to map to "internal" sites, increasing the likelihood of misannotation. Using the new m^6Am annotations, we find that depleting PCIF1 does not affect mRNA translation but is associated with reduced stability of a subset of m^6Am-annotated mRNAs. The discovery of PCIF1 and our accurate mapping technique will facilitate future studies to characterize m^6Am's function.

Keywords: m^6Am, m^6A, PCIF1, mRNA methylation, mRNA translation, mRNA stability

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^6A) at internal sites in mRNA (Bokar et al., 1997). At least 25% of mRNAs contain at least one m^6A (Dominissini et al., 2012; Meyer et al., 2012).

N6 methylation also occurs on Am to form a dimethylated adenosine: N^6, 2′-O-dimethyladenosine (m^6Am) (Keith et al., 1978; Wei et al., 1975). Am is primarily located at the first transcribed nucleotide position in mRNAs, adjacent to the m^7G 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 m^6Am. Since m^6Am is present at the first transcribed nucleotide in ~30% of all cellular mRNAs, m^6Am can affect the fate of a large subset of the transcriptome (Wei et al., 1975).

The functions of m^6Am are not clear. m^6Am is enriched in transcripts that exhibit enhanced stability and translation efficiency (Mauer et al., 2017a). However, it is not clear if the N6-methyl modification mediates these effects. Recent studies have shown that m^6Am impairs decapping of mRNAs by DCP2. However, DCP2 is not thought to mediate the majority of decapping activity in the cell (Li et al., 2012; Li et al., 2011). 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., 2012; Li et al., 2011). Therefore, it is not clear whether m^6Am mediates the generally high stability of m^6Am-containing transcripts.

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

To definitively distinguish m^6A and m^6Am in transcriptome-wide maps, depletion of either m^6A or m^6Am would be required. m^6A depletion cannot be readily achieved as *Mettl3* is essential for survival in nearly all of 341 cell lines that were screened (Tsherniak et al., 2017). The methyltransferase that generates m^6Am is not known, but its depletion could enable the identification of the sites that are m^6Am, since the remaining sites would be m^6A.

Here we describe the identification of PCIF1 as the methyltransferase that is responsible for generating m^6Am in mRNA. We show that PCIF1 methylates Am in the context of the m^7G cap, and has negligible ability to methylate adenosine in mRNA outside this context. By mapping m^6A in the transcriptome of PCIF1-deleted cells, we distinguish between m^6Am and 5′UTR m^6A. We find numerous examples where previously annotated m^6Am sites reflect m^6A and vice versa. We show that transcript isoforms with alternate TSSs account for many of these discrepancies facilitating the identification of these “internal” TSSs. Using this new high-confidence map of m^6Am sites, we characterize the fate of m^6Am-modified mRNAs in *PCIF1* knockout cells, and show that m^6Am has negligible effects on translation under basal conditions but is associated with increased stability of a subset of m^6Am-initiated transcripts. Overall, our studies identify PCIF1 as the methyltransferase that generates m^6Am in the transcriptome and provides revised transcriptome-wide maps that distinguish between m^6A and m^6Am.

RESULTS

Identification of PCIF1 as a candidate m^6Am-forming methyltransferase

Studies in the 1970’s provided initial characterization of an enzymatic activity in HeLa cells that synthesizes m^6Am (Keith et al., 1978). This enzyme selectively methylates Am adjacent to an m^7G cap in synthetic RNA substrates (Keith et al., 1978).

In order to identify the m^6Am-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-methylases (Iyer et al., 2016). Among these putative adenine methylases, PCIF1 is notable since 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 methylase family is derived from the prokaryotic M.EcoKI/M.TaqI methylases of the bacterial restriction-modification systems (Iyer et al., 2016). All of these methylases contain helices before and after the conserved core strand-3 which display partial or complete degeneration into coil elements.

Another feature of these methylases is the addition of a conserved residue from a helix N-terminal to the core methylase catalytic domain. A PCIF1 crystal structure revealed that its putative methyltransferase domain indeed adopts the classical Rossmann fold of many RNA methylases (Akichika et al., 2019). PCIF1 also contains a WW domain which 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 m^7G cap-dependent manner *in vitro*

To identify potential PCIF1-dependent nucleotide methylase 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* methylase assays with an RNA oligonucleotide containing a 5′ m^7G cap followed by 2′-*O*-methyladenosine (m^7G-ppp-Am-N_{20}) (**Figure 1B**). We found that PCIF1 methylates Am in this RNA to produce m^6Am, as assessed by UHPLC-MS/MS (**Figure 1C**).

Interestingly, we did not detect m^6A 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 suggests 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-methylases (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 if PCIF1-mediated N6 methylation of the m^7G-adjacent A requires 2′-*O*-methyl modification on the A. To test this, we used an RNA substrate with a 5′ m^7G cap followed by adenosine (m^7G-ppp-A-N_{20}) 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 m^6A (**Figure 1D**). We next examined the rate and substrate preference of PCIF1 using a serial dilution of the m^7G-ppp-Am- and the m^7G-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 unmethylated A RNA (**Figure 1F**), suggesting that PCIF1 has a ~7.6-fold higher preference for binding the 2′-O-methylated adenosine substrate.

Notably, the m^7G moiety was required for methylation as PCIF1 efficiently methylated Am to m^6Am in an m^7G capped RNA (m^7G-ppp-Am-N_{20}) but was unable to methylate Am in an RNA that lacked the m^7G cap (ppp-Am-N_{20}) (**Figure 1G**). Overall, these biochemical assays suggest that PCIF1 methyltransferase activity towards Am depends on the presence of the m^7G cap, but does not require 2′-O-methylation on the adenosine.

We next asked if the PCIF1 preference for m^7G-capped RNA was due to an ability to bind the m^7G cap. We performed cap-binding assays with PCIF1 using 7-methylguanosine-5-triphosphate (m^7G-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 m^7G-ppp beads and were efficiently eluted using m^7G-ppp-A but not G-ppp-A (**Figure 1H**). Similarly, PCIF1 bound to the m^7G-ppp beads and was eluted with m^7G-ppp-A but not G-ppp-A (**Figure 1H**). Together, these data suggest that PCIF1 binds directly to the m^7G cap, which may account for its specificity towards adenosine adjacent to the m^7G. These results are consistent with the crystal structure of PCIF1 which shows specific interactions with m^7G (Akichika et al., 2019).

PCIF1 knockout abolishes m^6Am levels without affecting m^6A in RNA

To determine the ability of PCIF1 to generate m^6Am in cells, we used CRISPR to delete *PCIF1* in various cell lines and examined levels of m^6Am and m^6A in RNA (**Figures 2A** and **S1A**). To measure m^6Am, we used a two-dimensional thin-layer chromatography (2D-TLC)-based method that can measure both m^6Am 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 [^{32}P]-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., m^6Am, Am, Gm, Cm, and Um. However, in *PCIF1* knockout cells, a selective and complete loss of m^6Am was detected (**Figures 2B** and **S1B**). A similar effect was seen using UHPLC-MS/MS to quantify m^6Am (**Figure 2C**). Thus, PCIF1 is required for the presence of m^6Am at the first transcribed nucleotide in mRNA.

We next asked if PCIF1 deletion affects m^6A levels in mRNA. To test this, we used a 2D-TLC-based method that selectively detects m^6A in the G-A-C context (Zhong et al., 2008) and complemented this with UHPLC-MS/MS to quantify all m^6A in mRNA. m^6A 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 m^6Am 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 m^6Am 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 if PCIF1 was sufficient to increase m^6Am levels in cells. We found that PCIF1 overexpression in HEK293T cells (**Figure 2H**) led to a ~3-fold increase in the m^6Am to Am ratio (**Figure 2I**). This increase in m^6Am levels was dependent on the catalytic activity of PCIF1, as overexpression of a catalytically inactive PCIF1 mutant had no effect on m^6Am levels (**Figure 2I**). Together this data suggests that PCIF1 is both necessary and sufficient to generate m^6Am in mRNA in cells.

miCLIP analysis of *PCIF1* knockout cells distinguishes m^6Am from 5′UTR m^6A residues

Next, we used the *PCIF1* knockout cells to distinguish m^6Am and m^6A in transcriptome-wide 6mA maps. We performed miCLIP, a method that produces narrow peaks, and nucleotide transitions at and adjacent to the m^6A (Linder et al., 2015). m^6A 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 m^6A (Linder et al., 2015). Because m^6Am can also be followed by cytosine, C to T transitions alone are not sufficient to distinguish m^6A from m^6Am. Peaks caused by m^6Am 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 m^6Am (Linder et al., 2015). However, because m^6A occurring near the TSS would also produce a similar shaped peak, this approach may result in false positive m^6Am 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^6Am and m^6A. 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**, ~57.5% 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 data set (**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 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 m^6Am.

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., 2017a), persist in the *PCIF1* knockout dataset (**Figure 3D** and **Figure 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., 2017a) 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^6Am identification.

A high-confidence transcriptome-wide map of m^6A and m^6Am based on PCIF1 depletion

To create a high confidence map of all m^6Am 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^6A (**Figure S2B**). However, 2360 peaks exhibited a significant reduction in both *PCIF1* knockout datasets (**Figure S2B**). In contrast, only 11 sites appeared to increase upon *PCIF1* knockout, suggesting a very low incidence of false positives.

We next identified the exact m^6Am 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^6Am 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^6Am within the PCIF1-dependent peaks. Previously we observed antibody-induced A to T transitions at the m^6A site in miCLIP (Linder et al., 2015). We confirmed that A to T transitions are readily detected at known m^6Am and m^6A throughout the transcriptome (**Figures S2C** and **S2D**). Therefore, we used a 10% A to T transition rate to identify the m^6Am 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^6Am sites that were called when using these methods separately (**Figure S2F**).

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

Next, we mapped m^6A in the 5′UTR. C to T transitions in a DRACH consensus were used to call m^6A sites based on the miCLIP protocol (Linder et al., 2015). This identified 399 5′UTR m^6A sites that were robustly called across all datasets (**Table S2**).

We next asked if mRNAs with 5′UTR m^6A and mRNAs that contain m^6Am are linked to different cellular processes, based on our updated m^6A and m^6Am sites. Functional annotation using DAVID shows that transcripts containing these distinct modified nucleotides are linked to different cellular processes, with 5′UTR m^6A associated with processes such as transcription and cell division, while m^6Am is primarily associated with splicing (**Figures S2H, I** and **Table S3**).

ATF4 contains a m^6Am rather than m^6A in its 5′UTR

We next wanted to understand if our revised map of m^6Am and m^6A can identify transcripts with misannotated modified nucleotides. A 5′UTR m^6A 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^6A 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^6A.

Based on the location of the peak, we asked if it instead reflects m^6Am rather than m^6A. 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^6Am (**Figure S3A**).

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

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

Identification of internal 6mA sites that reflect transcription-start m^{6}Am

We noticed two unusual features in our mapping results. First, not all m^{6}Am sites mapped to regions within annotated mRNA transcripts. Second, the m^{6}Am metagene showed that while 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 3E**).

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 nucleotides 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., 2018) 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 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 is likely to be a reflection of 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}A 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 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^6Am sites are indeed transcription start nucleotides (**Figure S3E**). Overall, these data further suggest that m^6Am is not internally located within transcripts but are instead found at the TSSs.

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

m^6Am correlates with enhanced translation, expression, and stability of mRNAs

In our previous studies, we found that m^6Am is correlated with transcripts that are highly expressed and have long half-lives in cells (Mauer et al., 2017a). We therefore wanted to reexamine this correlation based on the high-confidence m^6Am 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^6Am for this analysis, and mRNAs previously annotated as m^6Am 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^6Am are indeed more highly expressed and stable than mRNAs with other start nucleotides. Notably, m^6Am appears to be the predominant start nucleotide of the mRNAs that are the most abundant and have annotated half-lives greater than 24 hr (**Figure 5A-D**).

Overall, these data suggest that the presence of m^6Am correlates with an overall increase in mRNA stability, and that m^6Am is the predominant starting nucleotide on “outlier” mRNAs with unusually high stability and expression. To determine whether the N6-methyl in m^6Am 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^6Am (**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^6Am 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^6Am for transcript stability.

Previously we found that m^6Am-containing transcripts exhibit a subtle increase in translation relative to mRNAs with other start nucleotides (Mauer et al., 2017a). To more directly test the role of m^6Am 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^6Am 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^6Am compared to transcripts annotated to begin with other nucleotides (**Figure S4F**). In agreement with the modest effects of *PCIF1* depletion on translation rates, we found that protein levels of several m^6Am 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 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 a 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 transcripts and highly abundant transcripts in cells. However, depletion of m^6Am by *PCIF1* knockout does not markedly impair the stability of these unusual transcripts 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 confers these unusual properties to these mRNAs.

However, m^6Am 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 m^6Am, while others may not be affected? First, m^6Am is enriched in different sequence contexts. This contrasts with m^6A, which is nearly always found a single sequence context. m^6Am is therefore 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, this approach is likely to miss effects of m^6Am that depend on its various sequence contexts.

Another important factor that might affect whether m^6Am has a destabilizing 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., 2017b). DCP2 is not the major decapping enzyme in cells, but DCP2 targets mRNAs with specific 3′UTR features. Thus, m^6Am may be relevant only 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 context following m^6Am, and may be different during signaling or stress conditions that were not examined in our study.

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 DP2AG055947 (E.L.G.), and R01DA037755 (S.R.J.).

AUTHOR CONTRIBUTIONS

K.B. performed biochemical analysis of PCIF1 and generated PCIF1 knockout and overexpression cell lines, D.S. and K.B. performed assays of PCIF1 activity in cells, K.B., D.S., and S.Z. performed and analyzed ribosome profiling data, D.S. performed and analyzed SLAM-Seq experiments, B.H. performed and analyzed miCLIP experiments, N.L.I. 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, E.L.G and S.R.J. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors have no competing financial interests.

FIGURE LEGENDS

Figure legends

Figure 1. PCIF1 N6-methylates 2′-*O*-methyladenosine *in vitro* in an m^7G 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 are 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** are indicated.

B. Oligonucleotide sequences used in methyltransferase assays.

C. PCIF1 methylates m^7G-ppp-Am-N_{20} RNA. GST-PCIF1 (50 nM), but not the catalytically inactive mutants APPA or SPPG efficiently converts m^7G-ppp-Am (4 μM) to m^7G-ppp-m^6Am 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 m^6A. Each bar represents the mean ± s.e.m 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 m^7G-ppp-A-N_{20} (4 μM) to m^7G-ppp-m^6A-N_{20}. Assays were performed as in C. Each bar represents the mean ± s.e.m of 3 independent experiments. ***: $p < 0.001$, as assessed by unpaired t-tests.

E. PCIF1 enzyme kinetics. m^7G-ppp-Am-N_{20} (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 ± s.e.m of 3 independent experiments.

F. Michaelis-Menten kinetics of PCIF1 methylytransferase activity toward m^7G-ppp-Am and m^7G-ppp-A. Each point represents the mean ± s.e.m of 3 independent experiments.

G. PCIF1 activity depends on the presence of the m^7G cap. m^7G-ppp-Am-N_{20} or ppp-Am-N_{20} (4 μM) was incubated with GST-PCIF1 as in C. PCIF1 converted Am to m^6Am specifically in the m^7G capped RNA. Each bar represents the mean ± s.e.m of 3 independent experiments. ***: $p < 0.001$, as assessed by unpaired t-tests.

H. PCIF1 directly binds the m^7G cap. Anti-FLAG immunoblotting was used to detect binding of 3xFLAG-PCIF1 from HeLa cell extracts to m^7GTP-conjugated beads. The beads were eluted with m^7G-ppp-A or G-ppp-A. eIF4E and eIF4G were used to control for binding to m^7G.

Figure 2. PCIF1 N6-methylates 2′-*O*-methyladenosine in cells

A. CRISPR-mediated *PCIF1* knockout 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. PCIF1 is required for formation of m^6Am in mRNA in cells. m^6Am and Am levels in poly(A) RNA was 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 ± s.e.m of 3 independent experiments. *** $p < 0.001$, Student's t-test.

C. m^6Am is depleted in *PCIF1* KO HEK293T cells as assessed by UHPLC-MS/MS. Each bar represents the mean ± s.e.m of 3 independent experiments. ****: $p < 0.0001$ as assessed by paired t-tests.

D. PCIF1 does not affect the level of internal m^6A. 2D-TLC analysis of poly(A) RNA from HEK293T (WT) and *PCIF1* KO HEK293T show no effect on the level of m^6A.

E. Internal m^6A is not affected in *PCIF1* KO HEK293T cells as assessed by UHPLC-MS/MS. Each bar represents the mean ± s.e.m of 3 independent experiments. ns: not significant, as assessed by paired t-tests.

F. Wild-type but not a catalytically inactive PCIF1 mutant restores m^6Am levels in *PCIF1* knockout cells as assessed by 2D-TLC. Each bar represents the mean ± s.e.m of two independent experiments.

G. Wild-type but not catalytically inactive PCIF1 mutant restores m^6Am levels in *PCIF1* KO cells as assessed by UHPLC-MS/MS. Each bar represents the mean ± s.e.m of two independent experiments. Ns: not significant, * $p < 0.05$ as assessed by paired t-tests.

H. Western blot analysis demonstrate equivalent expression of wild-type and catalytically inactive FLAG-tagged PCIF1 in HEK293T cells. β-actin, loading control.

I. Overexpression of wild-type but not catalytically inactive PCIF1 increases m^6Am 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 ± s.e.m of 3 independent experiments. ns: not significant, ****$P \leq 0.0001$, unpaired t-tests.

Figure 3. **Depletion of PCIF1 distinguishes m^6A and m^6Am in transcriptome-wide 6mA maps.**

A. Metagene of miCLIP reads in wild-type and *PCIF1* knockout HEK293T cells. Shown is a metagene analysis of reads from the wild-type or PCIF1 knockout (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 m^6Am in the *PCIF1* knockout cells.

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

C. miCLIP peaks can be identified as m^6Am or m^6A based on their decrease in *PCIF1* knockout cells. Genome tracks were plotted for *RPL35* and *KDELR2* with called m^6A sites (FDR<0.1) and m^6Am sites indicated by red circles and blue triangle, respectively. Zoomed insets show m^6Am peaks can be distinguished from nearby m^6A sites.

D. The previously annotated m^6Am site in *RACK1* is actually a 5′UTR m^6A. The TSS-proximal peak in *RACK1* is not affected in the *PCIF1* knockout and overlaps with a DRACH motif.

E. Metagene analysis of *PCIF1* knockout-validated m^6Am sites shows m^6Am sites throughout the 5′UTR and in the transcript body. Shown is a metagene of the exact sites of m^6Am 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 CDS and 3′UTR.

F. DREME motif search of the nucleotides surrounding each m^6Am was performed confirms the previously reported BCA motif, and shows that the promoter sequence upstream of the m^6Am is GC-enriched.

Figure 4. Internally mapped m^6Am sites reflect m^6Am in mRNA isoforms with alternative TSSs.

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

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

D. The m^6Am 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^6Am peaks that appear within the 5′UTR reflect m^6Am in transcript isoforms with alternative TSSs. The exact m^6Am site (blue triangle) was determined using the A to T transition within the PCIF1-dependent peak.

E. Most m^6Am are found in the annotated 5′UTR of transcripts.

F. The internally mapping m^6Am in *YOD1* derives from a TSS of a *YOD1* transcript isoform. The m^6Am 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^6Am sites show overlap with CAGE data, indicating that m^6Am 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.

Figure 5. mRNA expression level and mRNA half-life depending on TSS

A. mRNAs with an annotated m^6Am 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 m^6Am map (Mauer et al., 2017a) . Transcripts that start with m^6Am are significantly upregulated. ****, $P < 2.2 \times 10{-}16$, Student's t-test. Cumulative distribution plot and boxplot represent the expression for mRNAs starting with m^6Am, Am, Cm, Gm and Um. Data shown are the average gene expression measured from two replicates for HEK293T cells.

B. m^6Am mRNAs annotated using the *PCIF1* knockout miCLIP dataset show increased expression compared to mRNAs with other start nucleotides. Cumulative distribution plots were prepared as in **A** using the high-confidence m^6Am dataset. The transcripts start with m^6Am are significantly upregulated as in **A**. ****, $P < 2.2 \times 10{-}16$, Student's t-test.

C. mRNAs with an annotated m^6Am start nucleotide show higher mRNA half-life than other mRNAs. Annotated mRNA half-lives were based on the first annotated nucleotide and an earlier m^6Am map (Mauer et al., 2017a). mRNAs with an annotated m^6Am exhibit a significantly elevated mRNA half-life than mRNAs with other annotated start nucleotides.
****, $P \leq 2.2 \times 10{-}16$, Student's t-test.

D. m^6Am mRNAs annotated using the *PCIF1* knockout miCLIP dataset show increased expression compared to mRNAs with other start nucleotides. Transcripts with m^6Am have significantly longer half-life with similar P-value. ****, $P \leq 2.2 \times 10{-}16$, Student's t-test.

E. Influence of PCIF1 depletion on mRNA half-life for transcripts in the lower half of gene expression. Transcripts with m^6Am have significantly shorter half-life in comparison to mRNAs annotated to begin with other nucleotides. *, $P = 0.0258$ by Student's t-test.

F. Influence of PCIF1 depletion on mRNA half-life for highly expressed transcripts. Transcripts with m^6Am show no significant decrease in mRNA half-life in comparison to mRNAs annotated to begin with other nucleotides. n.s., Student's t-test.

STAR METHODS

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-[illegible] mol 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 ppp(A_m)-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.

N^7-methylation of the purified G-ppp-Am-RNAs to give m^7G-ppp-Am-RNAs was carried out quantitatively using human mRNA guanine-N^7 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 m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP), rabbit anti-m^6A (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′-ACUUAACAUAUCCUGCGGGG-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 (1µg/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 seconds 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^7G 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^7G 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 $ZnCl_2$ 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 ddH20 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^7G-Am and m^7G-A RNA oligonucleotides were performed in methylation reaction buffer supplemented with 1.33 µM [^{3}H]-SAM (Perkin Elmer) and 10 µM SAM (NEB), using 20 nM GST-PCIF1 protein and a range of concentrations of m^7G-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 [^{3}H]-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^6A 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 ddH20 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^6Am 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 $ZnCl_2$ 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 ddH20 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 m). 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^6A), 296→150 (m^6Am), 244→112 (C). The amounts of A, C, Am, m^6A and m^6Am in the samples were quantified using corresponding calibration curves generated with pure standards. m^6Am and m^6A 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^7GTP conjugated Agarose beads (Jena Bioscience) for 1 hour at 4^{o}C degrees. 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^7G-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 resuspended 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 ddH_2O. 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 m^6Am, Am, and m^6A levels by thin layer chromatography
Levels of internal m^6A 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 m^6A. Thus, this method quantifies m^6A in a GA sequence context. 5′ ends were subsequently labeled with 10 units T4 PNK (NEB) and 0.4 mBq [γ-^{32}P] 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-H_2O 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 m^6Am:Am ratio was based on the protocol developed by Fray and colleagues (Kruse et al., 2011), with some modifications. Poly(A) RNA (1 µg) was used for the assay. 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 µl of DEPC-H_2O and 5′ ends were labeled using 30 units T4 PNK and 0.8 mBq [γ-^{32}P] 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-24/1.49m). For m^6Am experiments, the m^6Am: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 m^6A quantification, m^6A 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 m^6Am:Am ratios of

each experimental group (e.g. control vs. 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) capture. 10 µg poly(A) RNA was then used as input for single nucleotide-resolution m^6A 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 m^6A 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 m^6A residues with an FDR<0.1 and in a DRACH consensus were used as the final list of m^6A in this study.

To identify putative m^6Am sites, coverage of wild type and *PCIF1* knockout samples were compared genome-wide using the bamCompare 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 $\log_2$ ratio of BPM+1 for wild type to *PCIF1* knockout was calculated. A $\log_2$ 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 m^6Am regions across replicates was taken using bedtools intersect, resulting in 2360 high-confidence m^6Am peaks. To determine the precise m^6Am 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 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 (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 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′ RACE method was used. 10 µg poly(A) RNA was treated with Terminator exonuclease (Epicentre) 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 adapter (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 temperate 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 seconds on 1100 RPM, 30 seconds 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 adapter 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.

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 s^4U, 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 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 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 µM 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 $MgCl_2$, 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-H_2O 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 polysome 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 $MgCl_2$, 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 adapter 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:

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

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.

CONTACT FOR REAGENT AND RESOURCE SHARING

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.

DATA AND SOFTWARE AVAILABILITY

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

Reviewers may access these data deposited on GEO using the secure token qfmfaakyflkhvar.

Unprocessed and uncompressed imaging data is available at http://dx.doi.org/doi:10.17632/rnpfzjd7mj.1.

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 GRCm38 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 N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. Science (New York, NY *363*.

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.
Belanger, F., Stepinski, J., Darzynkiewicz, E., and Pelletier, J. (2010). Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. J Biol Chem *285*, 33037-33044.
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.
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.* (2018). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res.
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.
Hemeon, I., Gutierrez, J.A., Ho, M.C., and Schramm, V.L. (2011). Characterizing DNA methyltransferases with an ultrasensitive luciferase-linked continuous assay. Anal Chem *83*, 4996-5004.
Herzog, V.A., Reichholf, 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.
Hirose, Y., Iwamoto, Y., Sakuraba, K., Yunokuchi, I., Harada, F., and Ohkuma, Y. (2008). Human phosphorylated CTD-interacting protein, PCIF1, negatively modulates gene expression by RNA polymerase II. Biochem Biophys Res Commun *369*, 449-455.
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.*, et al.* (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, F., Kennedy, S., Hajian, T., Gibson, E., Seitova, A., Xu, C., Arrowsmith, C.H., and Vedadi, M. (2016). A Radioactivity-Based Assay for Screening Human m6A-RNA Methyltransferase, METTL3-METTL14 Complex, and Demethylase ALKBH5. J Biomol Screen *21*, 290-297.
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.
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, Y., Song, M., and Kiledjian, M. (2011). Differential utilization of decapping enzymes in mammalian mRNA decay pathways. RNA *17*, 419-428.
Linder, B., Grozhik, A.V., Olarerin-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.
Martin, S.A., and Moss, B. (1976). mRNA guanylyltransferase and mRNA (guanine-7-)-methyltransferase from vaccinia virions. Donor and acceptor substrate specificites. The Journal of biological chemistry *251*, 7313-7321.
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.* (2017a). Reversible methylation of m6Am in the 5' cap controls mRNA stability. Nature *541*, 371-375.
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.* (2017b). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. Nature *541*, 371-375.
McGlincy, 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.
Olarerin-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.
Robertson, A.K., Geiman, T.M., Sankpal, U.T., Hager, G.L., and Robertson, K.D. (2004). Effects of chromatin structure on the enzymatic and DNA binding functions of DNA methyltransferases DNMT1 and Dnmt3a in vitro. Biochem Biophys Res Commun *322*, 110-118.
Shimotohno, K., Kodama, Y., Hashimoto, J., and Miura, K.I. (1977). Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. Proceedings of the National Academy of Sciences of the United States of America *74*, 2734-2738.
Shuman, S. (2002). What messenger RNA capping tells us about eukaryotic evolution. Nature reviews *3*, 619-625.
Sonenberg, N., Morgan, M.A., Merrick, W.C., and Shatkin, A.J. (1978). A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. Proceedings of the National Academy of Sciences of the United States of America *75*, 4843-4847.
Sonenberg, N., and Shatkin, A.J. (1977). Reovirus mRNA can be covalently crosslinked via the 5' cap to proteins in initiation complexes. Proceedings of the National Academy of Sciences of the United States of America *74*, 4288-4292.
Sun, H., Zhang, M., Li, K., Bai, D., and Yi, C. (2019). Cap-specific, terminal N(6)-methylation by a mammalian m(6)Am 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-N(7))-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 e516.
Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A *101*, 11269-11274.
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. The EMBO journal *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., Bruning, J.C., and Qian, S.B. (2018). N(6)-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. Mol Cell *69*, 636-647 e637.
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.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. *PCIF1* deletion eliminates m^6Am in HeLa cells and both wild-type and catalytically dead PCIF1 express and localizes similarly

A. CRISPR-mediated knockout of *PCIF1* in HeLa cells is efficient as assessed by western blots of whole cell extracts blotted with an anti-PCIF1 antibody. The upper band represents endogenous PCIF1, whereas the lower band is a non-specific band. β-actin is shown as a loading control.

B. m^6Am is depleted in *PCIF1* KO HeLa cells as assessed by 2D-TLC. Left panels show representative images of three independent experiments with the blue circles representing m^6Am and the orange circles representing Am as assessed by standards run on separate blots in parallel. The bar graph on the right represents the mean ± s.e.m of three independent experiments. *: $p < 0.05$ as assessed by ratio paired t-test.

C. Western blot analysis demonstrates equivalent levels of wild-type and catalytically inactive PCIF1 rescue in HEK 293T *PCIF1* KO cells. Whole cell extracts from HEK293T *PCIF1* KO control cells (Empty Vector) and HEK 293T *PCIF1* KO cells stably overexpressing 3X-FLAG-PCIF1 WT or 3X-FLAG-PCIF1 SPPG catalytic mutant were blotted with anti-FLAG antibody (upper panel) and β-actin (lower panel).

D. Immunofluorescence staining of PCIF1 rescue demonstrates equivalent levels and localization of wild-type and catalytically dead PCIF1 in HEK293T *PCIF1* KO cells stably

expressing either 3X-FLAG-PCIF1 WT or 3X-FLAG-PCIF1 SPPG catalytic mutant. Both PCIF1 WT and SPPG mutant are expressed at similar levels and are primarily localized to the nucleus. DAPI was used to stain nuclei. Representative images are shown.

Figure S2. *PCIF1* knockout and miCLIP A to T transitions can be used to identify high-confidence m^{6}Am sites in the transcriptome

A. As in Figure 3D, *RPS5* is likely to be a 5′UTR m^{6}A-containing transcript instead of m^{6}Am, as was previously thought. The persistence of the TSS-proximal peak in the *PCIF1* knockouts suggests that this must be dependent on m^{6}A. No m^{6}A (as called by C to T transitions) was detectable at this peak, which may be due to the low and stochastic nature of C to T transitions. Inset shows the nucleotide sequence at this peak at the start site, with a potential DRACH motif (TGACA) indicated by the red line. Called m^{6}A sites are shown as red circles, and scale bar is shown underneath the gene structure.

B. Analysis of the fraction of miCLIP reads that are dependent on PCIF1. In this experiment, the genome was binned into 50 nt-wide windows and miCLIP read coverage within each bin was calculated with respect to strand. These values were normalized to the total number of reads mapping across all non-zero-coverage bins, per million (BPM). The $\log_2$ ratio of wild-type BPM+1 over knockout BPM+1 per was then plotted as a histogram with a square root y-axis. For each replicate, the skewness was ~1 with a kurtosis of ~9, showing that the very few deviations from no change were mainly found with a higher ratio. A cut off of 2 was chosen as being putative m^{6}Am regions per replicate. This was over 4 standard deviations from the mean. Overlap analysis showed a highly similar set of m^{6}Am regions. The intersection between the two replicates was then used as a list of high-confidence m^{6}Am regions as shown in the Venn diagram. These regions were then used for m^{6}Am site calling. p = 0 by hypergeometric probability.

C. A to T transitions are the most common transition at methylated adenosines and are markedly reduced in the *PCIF1* knockout miCLIP dataset. The transitions for each nucleotide at m^{6}A-called DRACH sites was plotted (left). As expected, the major m^{6}A signature mutation is the C to T transition as reported previously. However, at the m^{6}A site, A to T transitions, although less common than C to T transitions, are the most common mutation. These A to T transitions are markedly reduced at m^{6}Am sites in the *PCIF1* knockout (right). Thus, *PCIF1*-dependent A to T transitions can be used to confirm a m^{6}Am site.

D. Metagene analysis for all A to T transitions in miCLIP. A to T transitions are found primarily at the stop codon and TSS, likely reflecting antibody-induced signature mutations induced by m^{6}A and m^{6}Am, respectively.

E. Pie chart showing the methods used for m^{6}Am site calling. Of the 2360 high-confidence m^{6}Am regions, 1089 A to T transitions occurred at 10% or more reads. For those regions that did not meet these criteria, the most likely A-starting read pileup was chosen.

F. Overlap in m^{6}Am sites called by A to T transitions compared to the drop-off method. The A to T transitions provide high confidence that the chosen nucleotide is a 6mA; nevertheless, most sites were accurately called using the read-start method. p= 3.4e-04 by hypergeometric probability.

G. All m^{6}Am are found in a BCA context, with enrichment for specific, short sequence motifs following the m^{6}Am. Motifs within the genomic sequences surrounding each m^{6}Am were analyzed using MEME, further confirming the BCA context and revealing an enrichment for the motifs m^{6}Am-U-U-U and m^{6}Am-G-A-G. Subsets of m^{6}Am (those that appear to map internally (see **Figure 3**)) also show the BCA context and these short motifs, as would be expected for m^{6}Am at TSSs.

H. Functional annotation clustering shows 5'UTR m^6A is associated with transcription and the cell cycle. The non-redundant list of 5'UTR m^6A was annotated using DAVID v2.8. A dataset-specific background of all genes with at least 20 reads was used for this analysis. DAVID clusters enriched and similar Gene Ontologies and other pathways so that an overview of the general cellular processes of the gene sets investigated are involved in can be determined. Each cluster is similarly colored; the enrichment score is shown above for each cluster. Every other term title is shown due to space restraints; these are available in **Table S3**.

I. Function annotation clusters of m^6Am genes show these are strongly associated with mRNA splicing, mitochondrial processes, and translation. DAVID function annotation was performed as in (G) using the list of high-confidence m^6Am genes and the dataset-specific background. Clusters are shown in the same color and each clusters enrichment score is shown above. Every other term title is shown due to space restraints; these are available in **Table S3**.

Figure S3. *ATF4* shows a PCIF1-regulated m^6Am peak but no specific enrichment of m^6A in uORF2 or elsewhere in the mRNA

A. The largest 6mA peak in *ATF4* is due to m^6Am. The genome track for *ATF4* in wild-type HEK293T cells shows that the major peak is at the TSS. This site is lost in the *PCIF1* knockout cells. Shown is the location of the putative m^6A site in uORF2 as was reported in (Zhou et al., 2018). The schematic below indicates the complex uORF structure of the *ATF4* mRNA.

B. Pairwise alignment of the human and mouse *ATF4* sequences show that the putative m^6A DRACH consensus site is only seen in mouse *ATF4*, but not human *ATF4*. The mouse and human sequences from the TSS to the end of uORF2 (which extends beyond the start codon of the coding sequence of the main *ATF4* ORF) were aligned. The start codons for uORF1, uORF2, and *ATF4* coding sequence are indicated in bold and underlined. The uORF sequences share high (87%) sequence identity; however, at the putative mouse m^6A site, the DRACH consensus was lost (GGACA in mouse to CGACA in humans, shown in red).

C. miCLIP analysis of mouse embryonic fibroblast poly(A) RNA. The miCLIP analysis shows the expected stop codon enrichment of m^6A. The called m^6A sites were plotted as a metagene using MetaPlotR. Both MEF replicates show the typical m^6A metagene with a high degree of overlap.

D. A genome track of mouse *Atf4* (NM_009716.1; (Vattem and Wek, 2004)) using the MEF miCLIP data show a very similar profile to the human miCLIP profile on *ATF4* shown in (A). The major 5′UTR 6mA peak is similarly enriched at the TSS, with little to no reads at the putative m^6A site. Due to the highly similar profile, and the high identity of the mouse and human 5′UTR sequence (Supplementary Figure 4B), the major 6mA peak in MEFs is most likely m^6Am. m^6A is unlikely to be present within the body of this transcript except at trace levels.

E. An mRNA 5' end cloning method confirms that internally mapped m^6Am are indeed transcription start nucleotides. Polyadenylated RNA was either decapped by RppH or left capped as a control, ligated to a 5' adapter, reverse transcribed using gene-specific primers, and PCR amplified. A specific product at the expected size (indicated by the red asterisk) was detected for the transcripts tested and was not present in the capped (-RppH) control, confirming that the called m^6Am are bone fide transcription start nucleotides. Approximate m^6Am transcription start nucleotide product sizes: ACOT7 330 bp; BNIP2 330 bp; YOD1 263 bp.

Figure S4. m^6Am deficiency slightly influences translation efficiency

A. Correlation of Slam-Seq replicates derived from HEK293T wild-type and *PCIF1* knockout cells (0 h). The Spearman correlation coefficient (r) is shown.

B. The effect of *PCIF1* knockout on the stability of m^6Am transcripts can be rescued by overexpression of PCIF1. Wild-type or *PCIF1* knockout HEK293T cells were treated with DMSO or 5 µg/ml actinomycin D for 8 hours and RNA levels quantified by RT-PCR. Relative abundance compared to the control (normalized to a geometric mean of *RPS28* and *ACTB* control genes) of three biological replicates is shown as a boxplot. Wild-type PCIF1 expression restored mRNA levels, while the catalytically inactive mutant failed to restore expression to wild-type levels. *, $P < 0.1$ by Student's t-test, n.s., not significant.

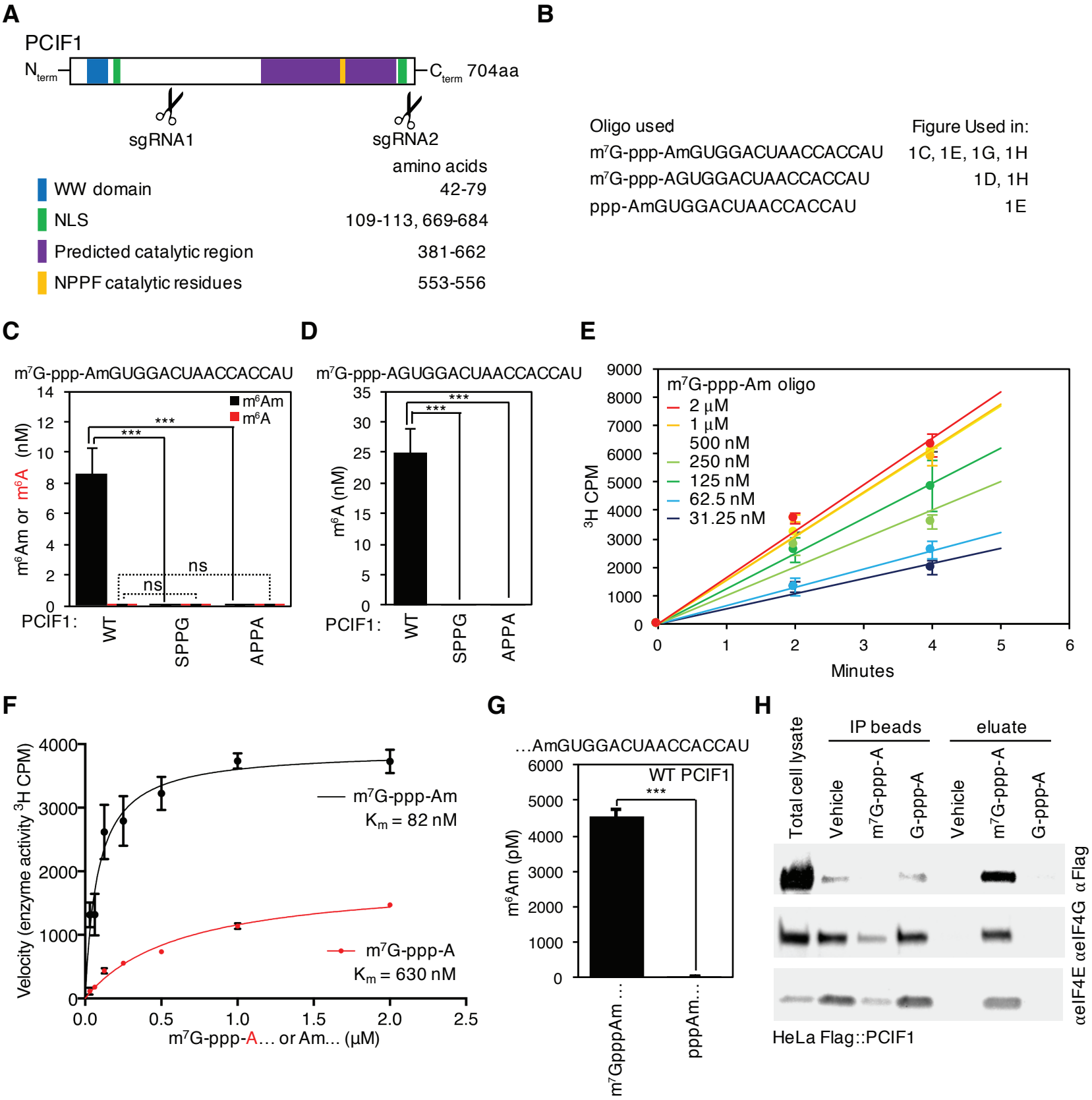
C. Venn diagram showing the intersection of transcripts showing altered expression in *PCIF1* HEK293T knockout cells and transcripts containing an annotated m^6Am sites. Upregulation or downregulation was defined as transcripts exhibiting a 0.2 $\log_2$-fold change in expression in *PCIF1* knockout cells relative to wild type as determined by RNA-seq. Only transcripts with a TPM ≥ 10 in wild-type cells were used for each category in this analysis.

D. Ribosome profiling dataset replicates show high correlation. Shown are the correlation plots of translation efficiency for two wild-type HEK293T replicates and three *PCIF1* knockout HEK293T replicates. The Spearman correlation coefficient (r) is shown.

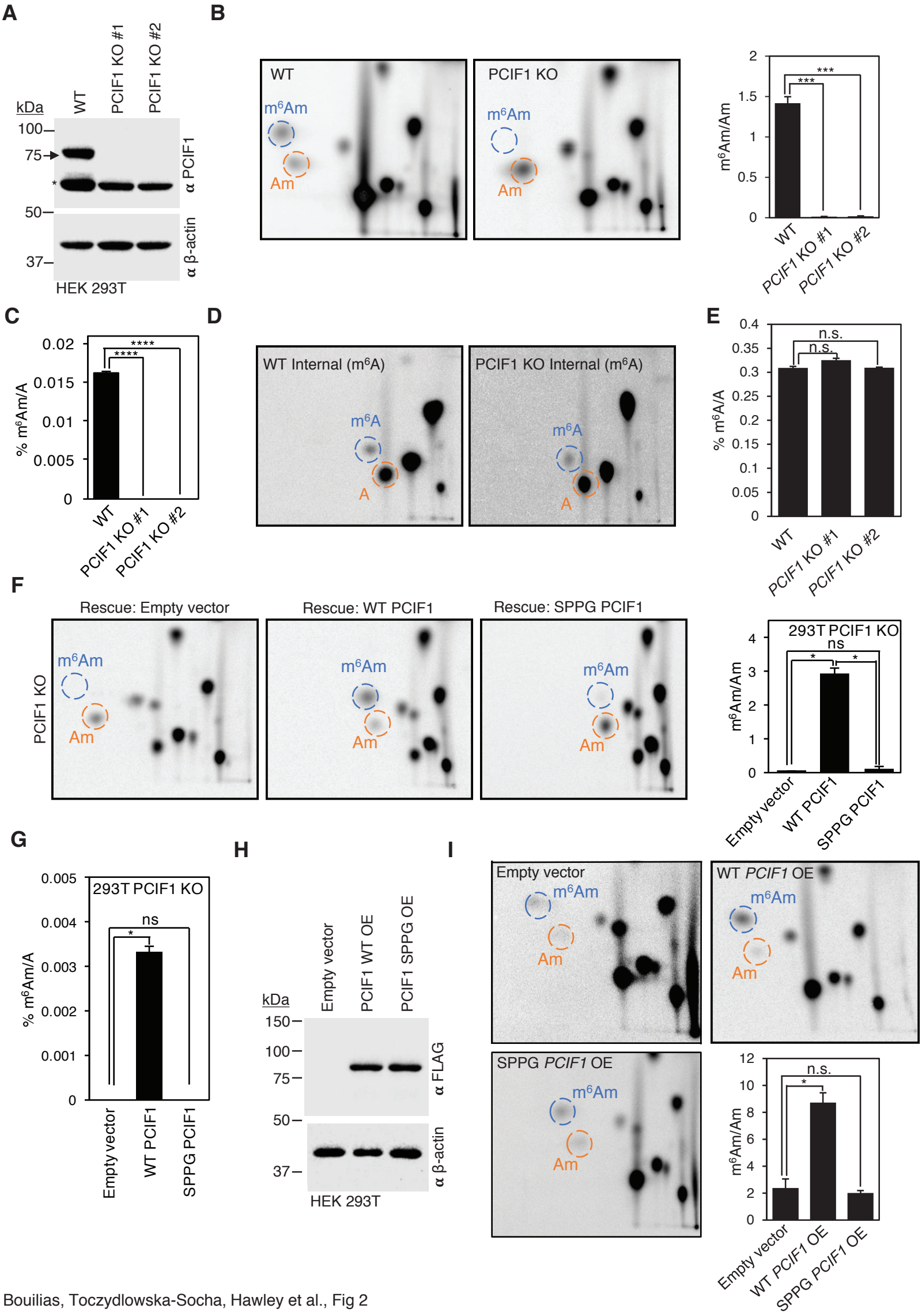
E. Ribosome-protected fragments are enriched in the coding sequence for all replicates. Shown is the distribution of ribosome profiling reads between the coding sequence (CDS) and UTRs. These data show the expected high coverage in the CDS compared to UTRs.

F. Effect of PCIF1 depletion of mRNA translation efficiency. Ribosome profiling was performed in wild-type and *PCIF1* knockout HEK293T cells. A matched RNA-Seq experiment was performed for each replicate in order to calculate translation efficiency. The boxplot represents the change in translation efficiency for mRNAs based on their annotated start nucleotide. The cumulative distribution plot indicates the change in translation efficiency in the *PCIF1* KO relative to wild-type HEK293T cells for mRNAs with an annotated m^6Am start nucleotide in comparison to mRNAs with an annotated Am start nucleotide. The translation efficiency of mRNAs starting with an m^6Am is slightly increased upon *PCIF1* KO. Data represent the average from two independent replicates of ribosome profiling data sets for wild-type HEK293T cells and three replicates of *PCIF1* KO HEK293T cells. *, $P = 0.02282$ by Student's t-test.

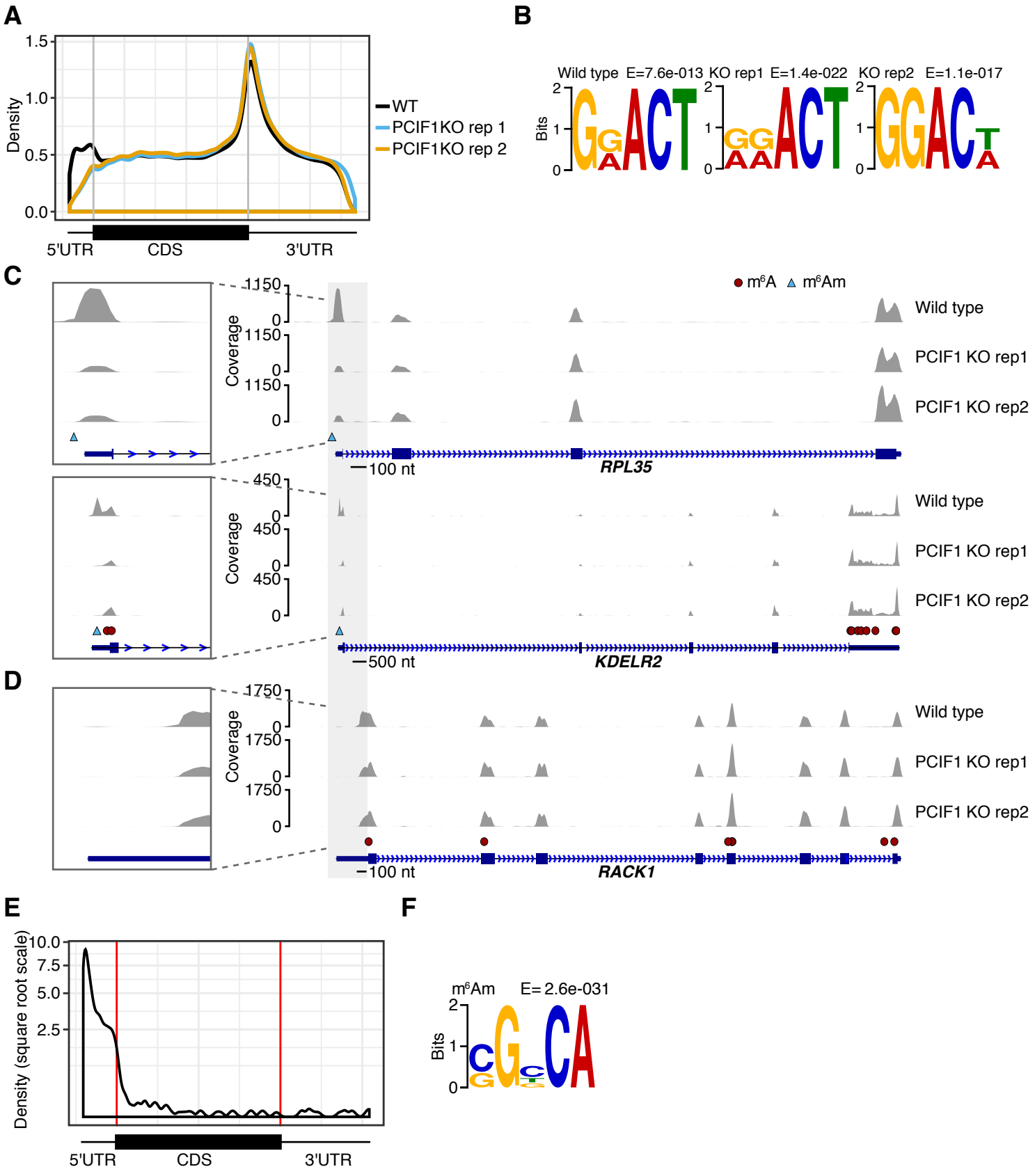
G. Western blot analysis shows that the protein levels of several m^6Am transcripts remain unchanged in *PCIF1* knockout cells. 10 µg of protein whole cell extract from three replicate samples of wild-type HEK293T and *PCIF1* KO cells were blotted with anti-TRIM28, anti-PARP1, anti-HSPA1B, anti-GAPDH, anti-HSPA8, anti-β-actin and anti-ATF5 antibodies (for m^6Am transcripts), and anti-EEF2 and anti-RACK1 antibodies (for non-m^6Am transcripts). Each bar represents the quantification of band intensity mean ± standard deviation for three replicate samples. Ns: not significant as assessed by unpaired t-tests.



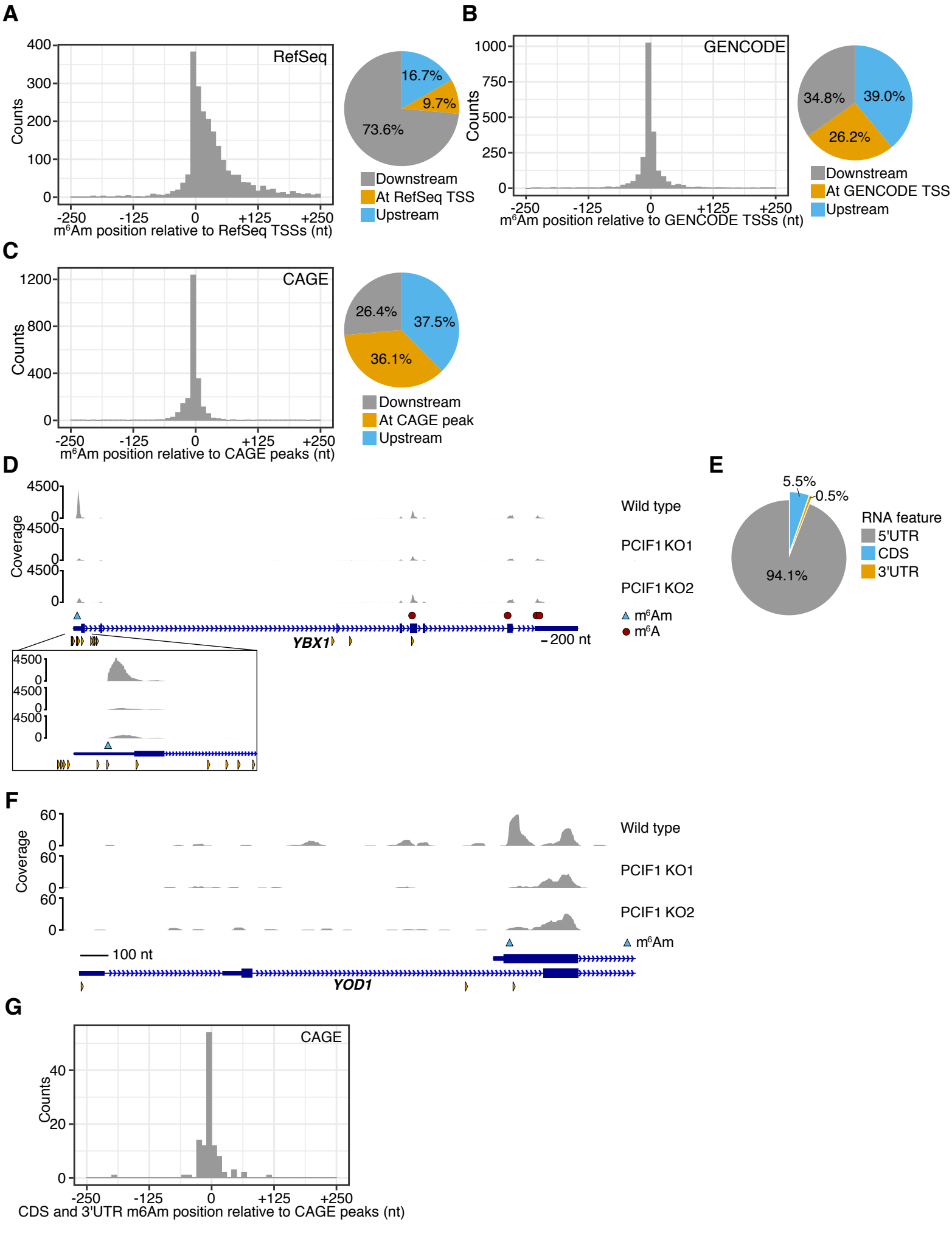
Boulias, Toczydlowska-Socha, Hawley et al Fig 1

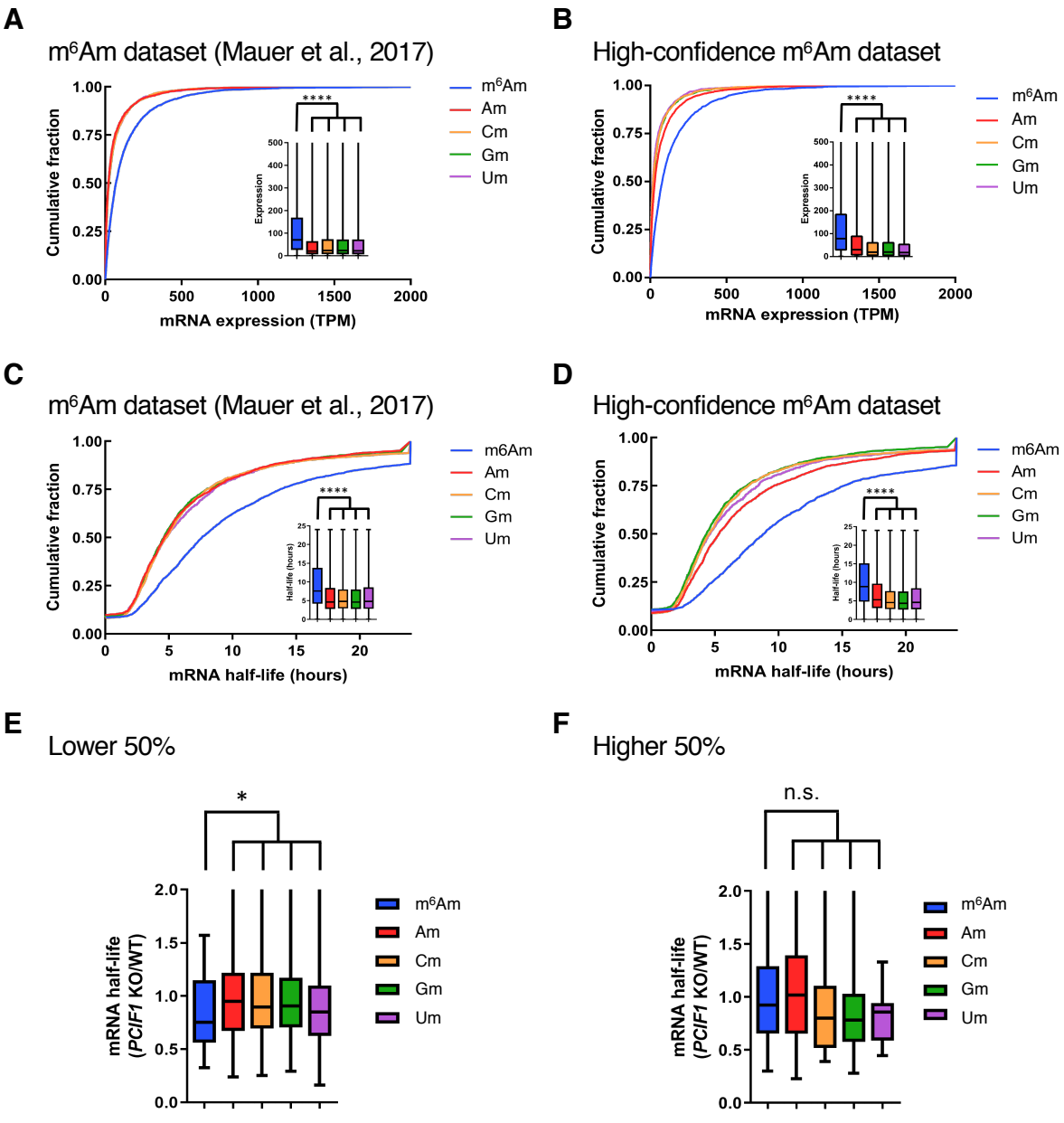


Bouilias, Toczydlowska-Socha, Hawley et al., Fig 2

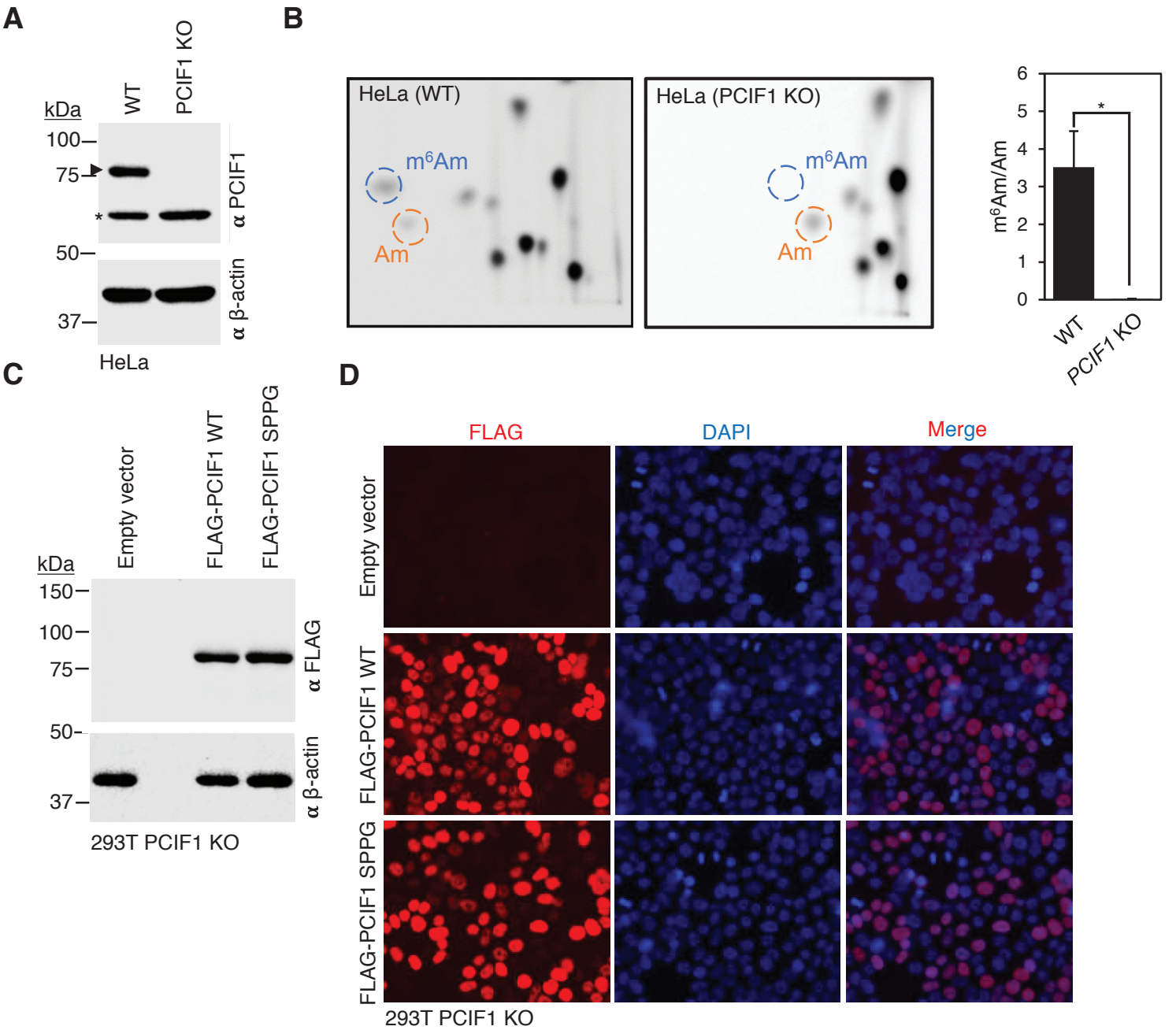


Boulias, Toczydlowska-Socha, Hawley et al Fig 3

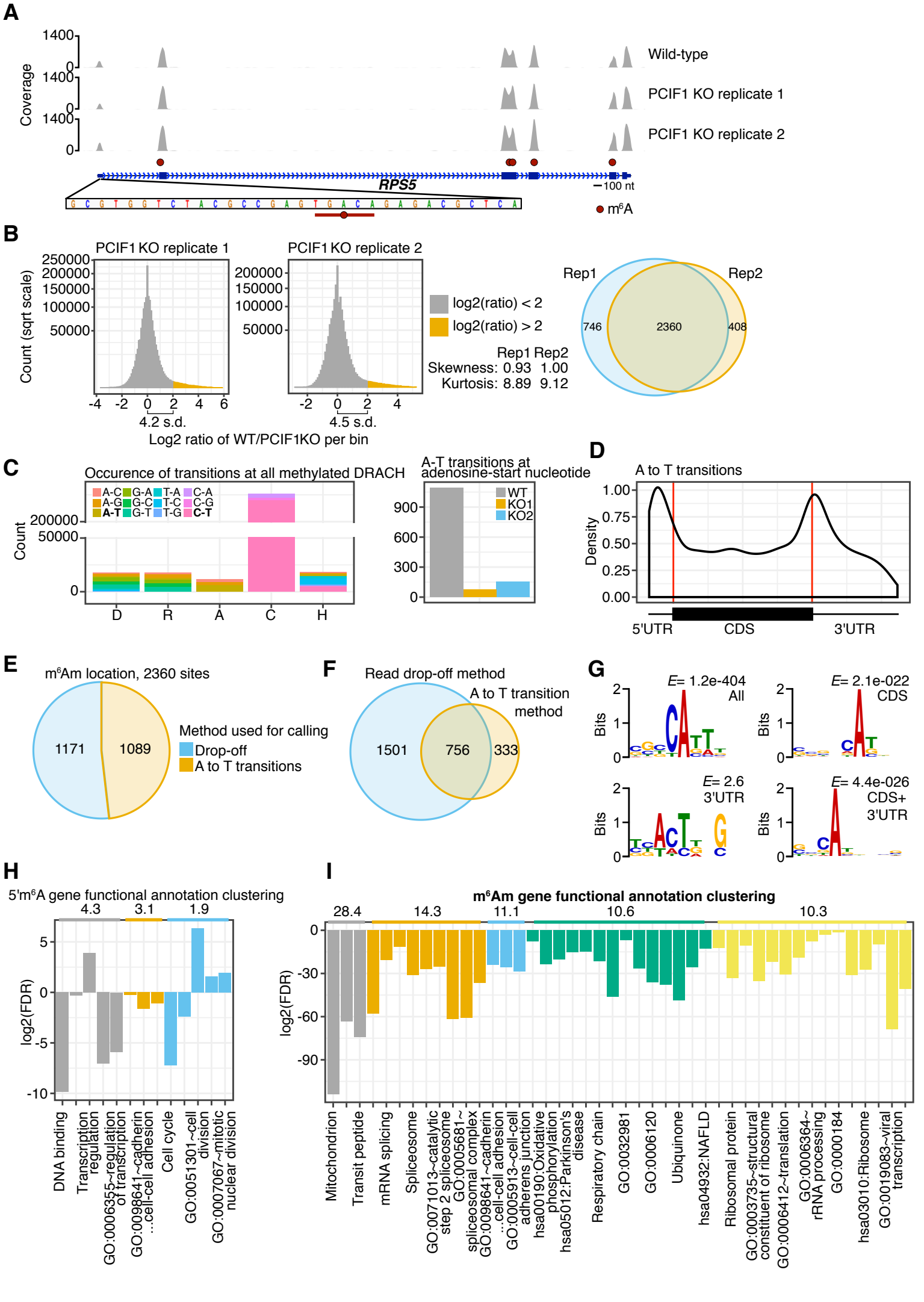

A
RefSeq
Counts
m6Am position relative to RefSeq TSSs (nt)
16.7%
9.7%
73.6%
Downstream
At RefSeq TSS
Upstream
B
GENCODE
Counts
m6Am position relative to GENCODE TSSs (nt)
34.8%
39.0%
26.2%
Downstream
At GENCODE TSS
Upstream
C
CAGE
Counts
m6Am position relative to CAGE peaks (nt)
26.4%
37.5%
36.1%
Downstream
At CAGE peak
Upstream
D
Coverage
Wild type
PCIF1 KO1
PCIF1 KO2
m6Am
m6A
YBX1
200 nt
E
5.5%
0.5%
94.1%
RNA feature
5'UTR
CDS
3'UTR
F
Coverage
Wild type
PCIF1 KO1
PCIF1 KO2
m6Am
100 nt
YOD1
G
CAGE
Counts
CDS and 3'UTR m6Am position relative to CAGE peaks (nt)



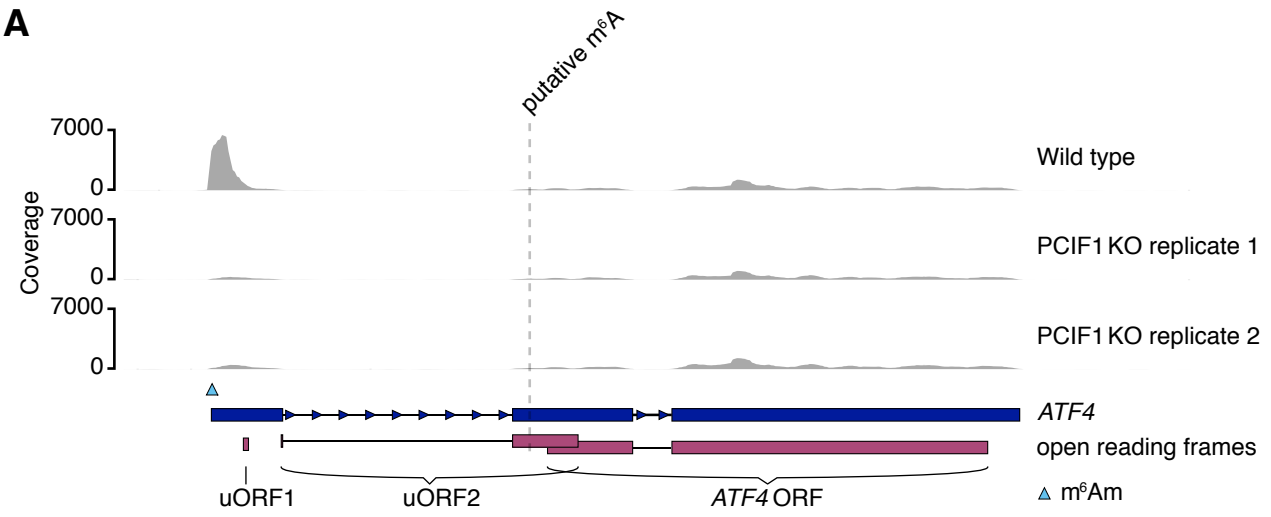
Bouilias, Toczydlowska-Socha, Hawley et al., Fig5

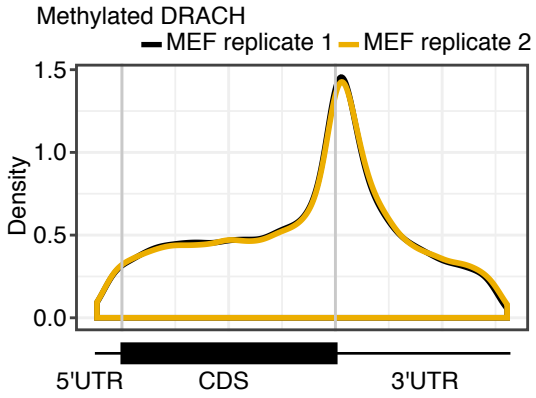


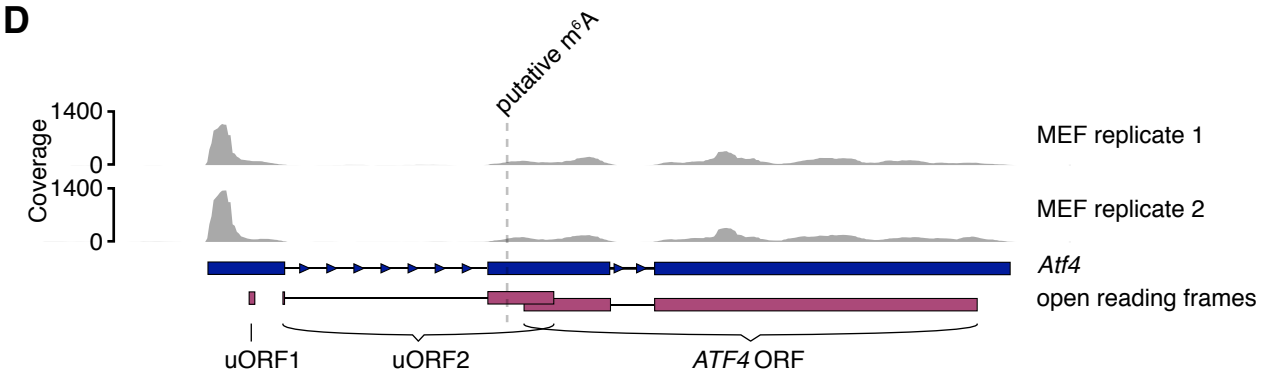
Bouilias, Toczydlowska-Socha, Hawley et al., Fig S1

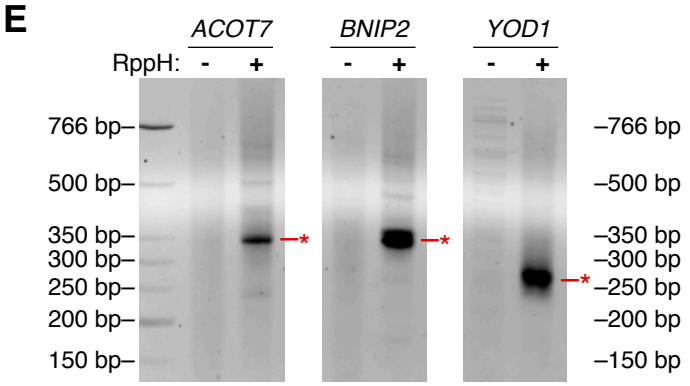


Boulias, Toczydlowska-Socha, Hawley et al Fig S2

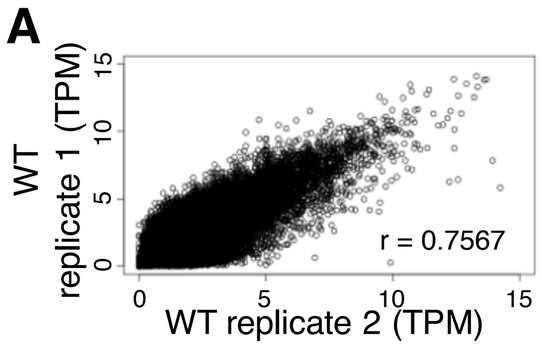


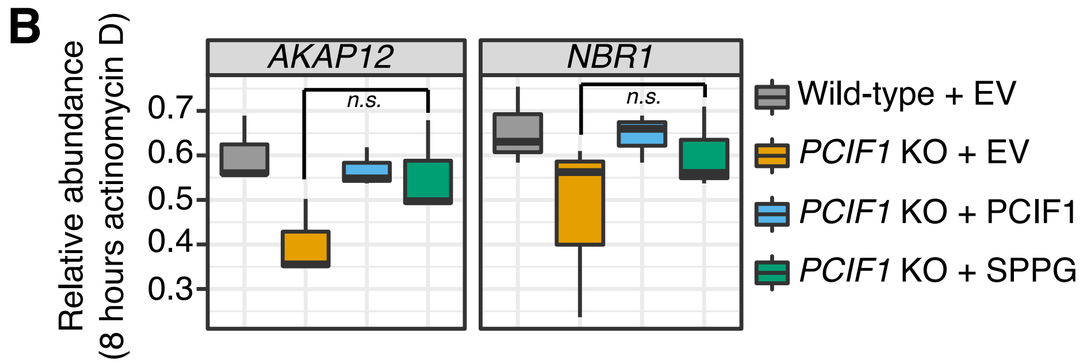


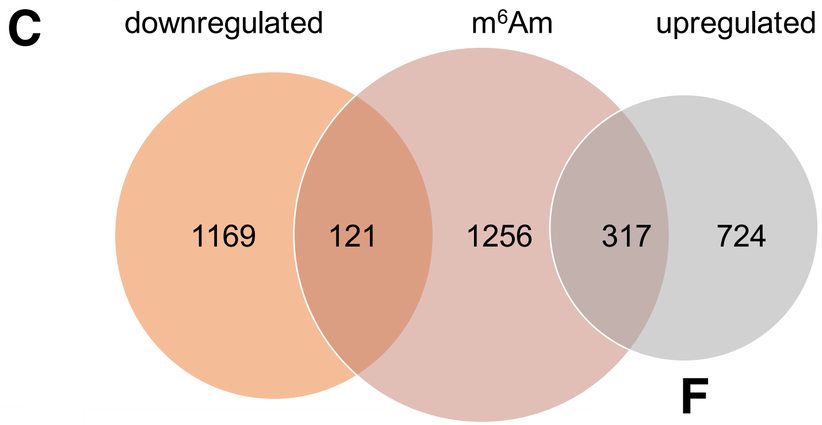


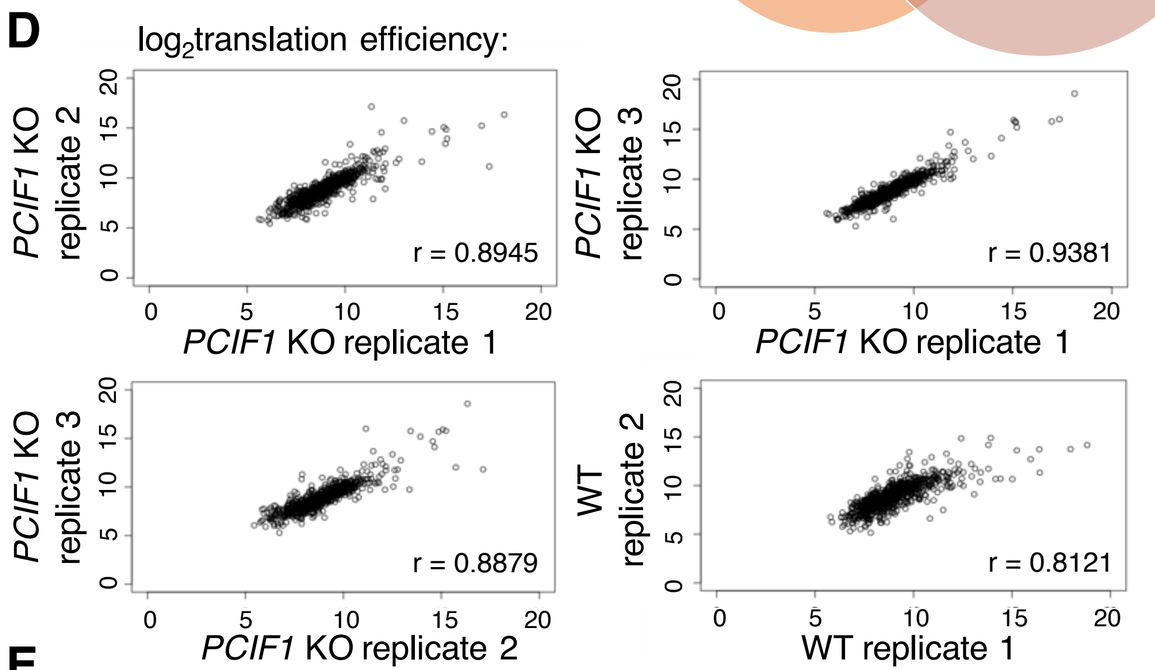


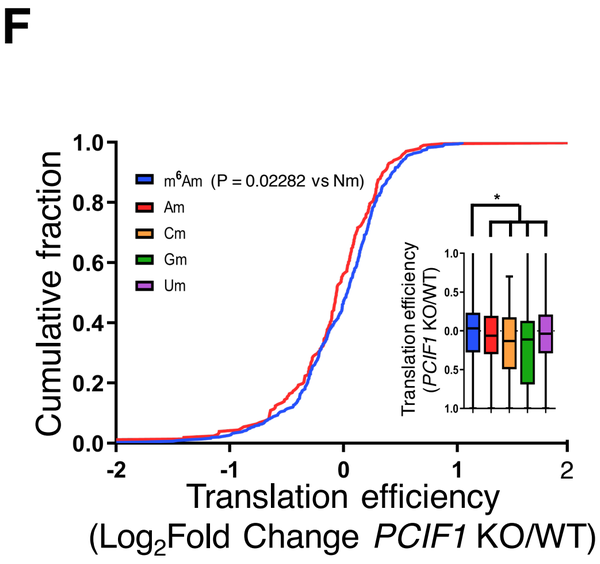
Boulias, Toczydlowska-Socha, Hawley et al Fig S3

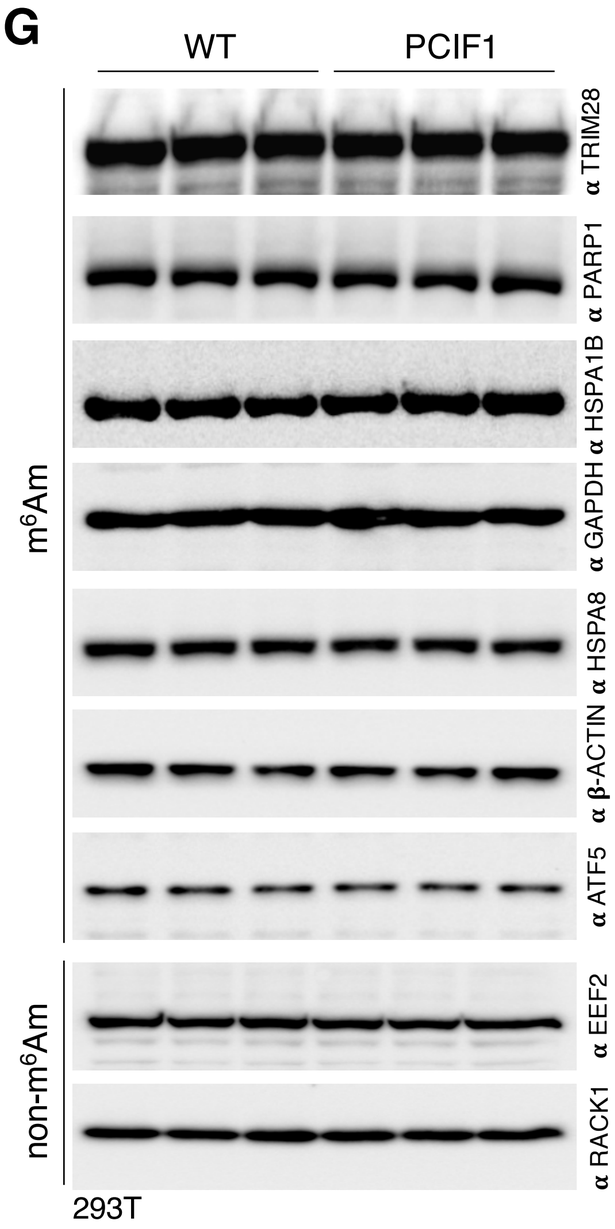


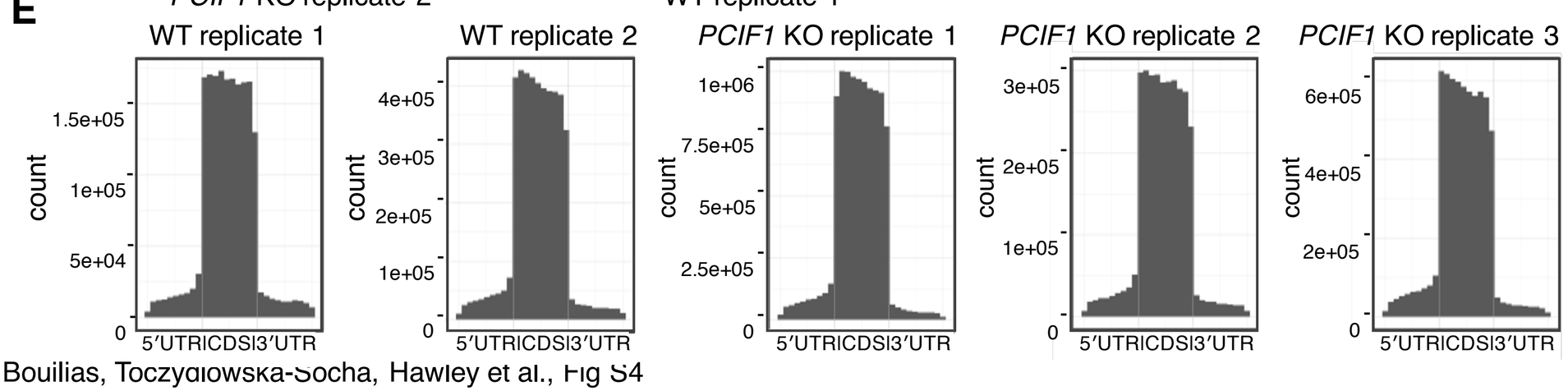


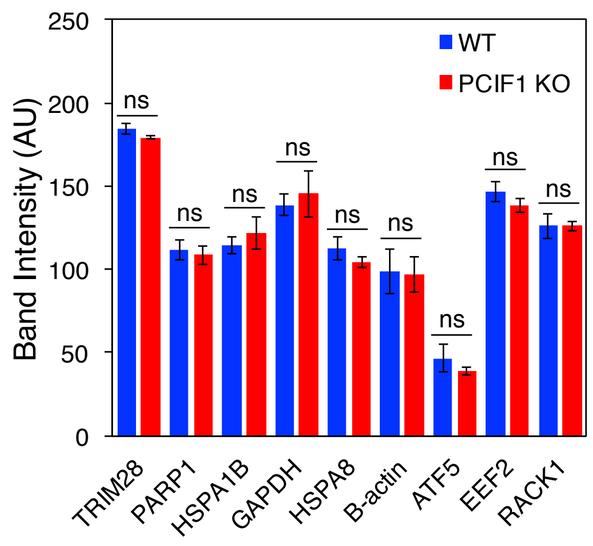












Bouilias, Toczydlowska-Socha, Hawley et al., Fig S4

KEY RESOURCES TABLE

The table highlights the genetically modified organisms and strains, cell lines, reagents, software, and source data **essential** to reproduce results presented in the manuscript. Depending on the nature of the study, this may include standard laboratory materials (i.e., food chow for metabolism studies), but the Table is **not** meant to be comprehensive list of all materials and resources used (e.g., essential chemicals such as SDS, sucrose, or standard culture media don't need to be listed in the Table). **Items in the Table must also be reported in the Method Details section within the context of their use.** The number of **primers and RNA sequences** that may be listed in the Table is restricted to no more than ten each. If there are more than ten primers or RNA sequences to report, please provide this information as a supplementary document and reference this file (e.g., See Table S1 for XX) in the Key Resources Table.

Please note that ALL references cited in the Key Resources Table must be included in the References list. Please report the information as follows:

- **REAGENT or RESOURCE:** Provide full descriptive name of the item so that it can be identified and linked with its description in the manuscript (e.g., provide version number for software, host source for antibody, strain name). In the Experimental Models section, please include all models used in the paper and describe each line/strain as: model organism: name used for strain/line in paper: genotype. (i.e., Mouse: OXTR$^{fl/fl}$: B6.129(SJL)-Oxtr$^{tm1.1Wsy/J}$). In the Biological Samples section, please list all samples obtained from commercial sources or biological repositories. Please note that software mentioned in the Methods Details or Data and Software Availability section needs to be also included in the table. See the sample Table at the end of this document for examples of how to report reagents.

- **SOURCE:** Report the company, manufacturer, or individual that provided the item or where the item can obtained (e.g., stock center or repository). For materials distributed by Addgene, please cite the article describing the plasmid and include "Addgene" as part of the identifier. If an item is from another lab, please include the name of the principal investigator and a citation if it has been previously published. If the material is being reported for the first time in the current paper, please indicate as "this paper." For software, please provide the company name if it is commercially available or cite the paper in which it has been initially described.

- **IDENTIFIER:** Include catalog numbers (entered in the column as "Cat#" followed by the number, e.g., Cat#3879S). Where available, please include unique entities such as RRIDs, Model Organism Database numbers, accession numbers, and PDB or CAS IDs. For antibodies, if applicable and available, please also include the lot number or clone identity. For software or data resources, please include the URL where the resource can be downloaded. Please ensure accuracy of the identifiers, as they are essential for generation of hyperlinks to external sources when available. Please see the Elsevier list of Data Repositories with automated bidirectional linking for details. When listing more than one identifier for the same item, use semicolons to separate them (e.g. Cat#3879S; RRID: AB_2255011). If an identifier is not available, please enter "N/A" in the column.
    - ***A NOTE ABOUT RRIDs:*** We highly recommend using RRIDs as the identifier (in particular for antibodies and organisms, but also for software tools and databases). For more details on how to obtain or generate an RRID for existing or newly generated resources, please visit the RII or search for RRIDs.

Please use the empty table that follows to organize the information in the sections defined by the subheading, skipping sections not relevant to your study. Please do not add subheadings. To add a row, place the cursor at the end of the row above where you would like to add the row, just outside the right border of the table. Then press the ENTER key to add the row. Please delete empty rows. Each entry must be on a separate row; do not list multiple items in a single table cell. Please see the sample table at the end of this document for examples of how reagents should be cited.

TABLE FOR AUTHOR TO COMPLETE

Please upload the completed table as a separate document. ***Please do not add subheadings to the Key Resources Table.*** *If you wish to make an entry that does not fall into one of the subheadings below, please contact your handling editor. (**NOTE:** For authors publishing in Current Biology, please note that references within the KRT should be in numbered style, rather than Harvard.)*

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:
rabbit anti-PARP1	Cell Signaling	9542, RRID: AB_10616513
rabbit anti-HSPA8	Cell Signaling	8444, RRID: AB_10831837
rabbit anti-m^6A	Abcam	ab151230, RRID: AB_2753144
rabbit anti-ATF5	Abcam	ab60126, RRID: AB_940375
rabbit anti-m^6A	Abcam	ab151230. RRID:2753144
Bacterial and Virus Strains		
T7 Express lysY	New England Biolabs	C3010
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Cycloheximide	Sigma Aldrich	C4859
iodoacetamide	Sigma Aldrich	I1149
4-thiouridine (s^4U)	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	Epicentre	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	Epicentre	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
Actinomycin D	Sigma Aldrich	A1410
SuperScript IV Reverse Transcriptase	Thermo	18090010
iQ SYBR Green supermix	Bio-Rad	1708880
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	http://dx.doi.org/doi:10.17632/rnpfzjd7mj.1
Experimental Models: Cell Lines		
HEK293T	ATCC	CRL-3216
HeLa	ATCC	CCL-2
Experimental Models: Organisms/Strains		
Oligonucleotides		
m7GpppAmGUGGACUAACCACCAU	This Paper and Trilink	N/A
m7GpppAGUGGACUAACCACCAU	This Paper and Trilink	N/A
pppAmGUGGACUAACCACCAU	This Paper	N/A
sgRNA1; 5′- CGGUUGAAAGACUCCCGUGG-3′	This Paper	N/A
sgRNA2; 5′- ACUUAACAUAUCCUGCGGGG-3′	This Paper	N/A
Biotin 5’ adapter; biotin-GTTCAGAGTTCTACAGTCCGACGATC	This Paper	N/A

RT_ACOT7: cccgttctggctgttgcaatgc	This Paper	N/A
RT_BNIP2: agcctggattcaatgtcaactac	This Paper	N/A
RT_YOD1: cggctggacagcccctgcaaaac	This Paper	N/A
PCR_5'adapter: taatacgactcactatagggtctcGGATCCcgacGTTCAGAGTTCTACAGTCCGACGATC	This Paper	N/A
PCR_ACOT7: GTAAAACGACGGCCAGTaagaGAATTCggaacccgttctggctgttgcaatgc	This Paper	N/A
PCR_BNIP2: GTAAAACGACGGCCAGTaagaGAATTCggaaagcctggattcaatgtcaactac	This Paper	N/A
PCR_YOD1: GTAAAACGACGGCCAGTaagaGAATTCggaacggctggacagcccctgcaaaac	This Paper	N/A
qPCR_ACTB_Fw: GAAGATCAAGATCATTGCTCCTC	This Paper	N/A
qPCR_ACTB_Rv: ATCCACATCTGCTGGAAGG	This Paper	N/A
qPCR_RPS28_Fw: ATCAAGCTGGCTAGGGTAACC	This Paper	N/A
qPCR_RPS28_Rv: GGCCTTTGACATTTCGGATGA	This Paper	N/A
qPCR_AKAP12_Fw: CATTGTCACAGAGGTTGGA	This Paper	N/A
qPCR_AKAP12_Rv: GCTGACTTAGTAGCCATCTC	This Paper	N/A
qPCR_NBR1_Fw: GTCTAATACCCTGATGCTCCC	This Paper	N/A
qPCR_NBR1_Rv: GCAAATTCTCATCCACAAATGC	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-24/1.49m)	NIH	N/A
Flexbar v2.5	Dodt, Roehr, Ahmed, and Dieterich, 2012	https://github.com/seqan/flexbar
pyCRAC	Webb, Hector, Kudla, and Granneman, 2014	https://bitbucket.org/sgrann/pycrac
bwa v0.7.17	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/

CTK v1.1.3	Shah et al., 2017	https://github.com/chaolinzhanglab/ctk
deeptools v3.1.3	Ramírez 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/
Other		

TABLE WITH EXAMPLES FOR AUTHOR REFERENCE

REAGENT or RESOURCE	**SOURCE**	**IDENTIFIER**
Antibodies		
Rabbit monoclonal anti-Snail	Cell Signaling Technology	Cat#3879S; RRID: AB_2255011
Mouse monoclonal anti-Tubulin (clone DM1A)	Sigma-Aldrich	Cat#T9026; RRID: AB_477593
Rabbit polyclonal anti-BMAL1	This paper	N/A
Bacterial and Virus Strains		
pAAV-hSyn-DIO-hM3D(Gq)-mCherry	Krashes et al., 2011	Addgene AAV5; 44361-AAV5
AAV5-EF1a-DIO-hChR2(H134R)-EYFP	Hope Center Viral Vectors Core	N/A
Cowpox virus Brighton Red	BEI Resources	NR-88
Zika-SMGC-1, GENBANK: KX266255	Isolated from patient (Wang et al., 2016)	N/A
Staphylococcus aureus	ATCC	ATCC 29213
Streptococcus pyogenes: M1 serotype strain: strain SF370; M1 GAS	ATCC	ATCC 700294
Biological Samples		
Healthy adult BA9 brain tissue	University of Maryland Brain & Tissue Bank; http://medschool.umaryland.edu/btbank/	Cat#UMB1455

Human hippocampal brain blocks	New York Brain Bank	http://nybb.hs.columbia.edu/
Patient-derived xenografts (PDX)	Children's Oncology Group Cell Culture and Xenograft Repository	http://cogcell.org/
Chemicals, Peptides, and Recombinant Proteins		
MK-2206 AKT inhibitor	Selleck Chemicals	S1078; CAS: 1032350-13-2
SB-505124	Sigma-Aldrich	S4696; CAS: 694433-59-5 (free base)
Picrotoxin	Sigma-Aldrich	P1675; CAS: 124-87-8
Human TGF-β	R&D	240-B; GenPept: P01137
Activated S6K1	Millipore	Cat#14-486
GST-BMAL1	Novus	Cat#H00000406-P01
Critical Commercial Assays		
EasyTag EXPRESS 35S Protein Labeling Kit	Perkin-Elmer	NEG772014MC
CaspaseGlo 3/7	Promega	G8090
TruSeq ChIP Sample Prep Kit	Illumina	IP-202-1012
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE63473
B-RAF RBD (apo) structure	This paper	PDB: 5J17
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Nanog STILT inference	This paper; Mendeley Data	http://dx.doi.org/10.17632/wx6s4mj7s8.2
Affinity-based mass spectrometry performed with 57 genes	This paper; and Mendeley Data	Table S8; http://dx.doi.org/10.17632/5hvpvspw82.1
Experimental Models: Cell Lines		
Hamster: CHO cells	ATCC	CRL-11268
D. melanogaster: Cell line S2: S2-DRSC	Laboratory of Norbert Perrimon	FlyBase: FBtc0000181
Human: Passage 40 H9 ES cells	MSKCC stem cell core facility	N/A
Human: HUES 8 hESC line (NIH approval number NIHhESC-09-0021)	HSCI iPS Core	hES Cell Line: HUES-8
Experimental Models: Organisms/Strains		
C. elegans: Strain BC4011: srl-1(s2500) II; dpy-18(e364) III; unc-46(e177)rol-3(s1040) V.	Caenorhabditis Genetics Center	WB Strain: BC4011; WormBase: WBVar00241916
D. melanogaster: RNAi of Sxl: y[1] sc[*] v[1]; P{TRiP.HMS00609}attP2	Bloomington Drosophila Stock Center	BDSC:34393; FlyBase: FBtp0064874
S. cerevisiae: Strain background: W303	ATCC	ATTC: 208353
Mouse: R6/2: B6CBA-Tg(HDexon1)62Gpb/3J	The Jackson Laboratory	JAX: 006494

Mouse: OXTRfl/fl: B6.129(SJL)-Oxtr$^{tm1.1Wsy}$/J	The Jackson Laboratory	RRID: IMSR_JAX:008471
Zebrafish: Tg(Shha:GFP)t10: t10Tg	Neumann and Nuesslein-Volhard, 2000	ZFIN: ZDB-GENO-060207-1
Arabidopsis: 35S::PIF4-YFP, BZR1-CFP	Wang et al., 2012	N/A
Arabidopsis: JYB1021.2: pS24(AT5G58010)::cS24:GFP(-G):NOS #1	NASC	NASC ID: N70450
Oligonucleotides		
siRNA targeting sequence: PIP5K I alpha #1: ACACAGUACUCAGUUGAUA	This paper	N/A
Primers for XX, see Table SX	This paper	N/A
Primer: GFP/YFP/CFP Forward: GCACGACTTCTTCAAGTCCGCCATGCC	This paper	N/A
Morpholino: MO-pax2a GGTCTGCTTTGCAGTGAATATCCAT	Gene Tools	ZFIN: ZDB-MRPHLNO-061106-5
ACTB (hs01060665_g1)	Life Technologies	Cat#4331182
RNA sequence: hnRNPA1_ligand: UAGGGACUUAGGGUUCUCUCUAGGGACUUAG GGUUCUCUCUAGGGA	This paper	N/A
Recombinant DNA		
pLVX-Tight-Puro (TetOn)	Clonetech	Cat#632162
Plasmid: GFP-Nito	This paper	N/A
cDNA GH111110	Drosophila Genomics Resource Center	DGRC:5666; FlyBase:FBcl0130415
AAV2/1-hsyn-GCaMP6- WPRE	Chen et al., 2013	N/A
Mouse raptor: pLKO mouse shRNA 1 raptor	Thoreen et al., 2009	Addgene Plasmid #21339
Software and Algorithms		
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al., 2009	http://samtools.sourceforge.net/
Weighted Maximal Information Component Analysis v0.9	Rau et al., 2013	https://github.com/ChristophRau/wMICA
ICS algorithm	This paper; Mendeley Data	http://dx.doi.org/10.17632/5hvpvspw82.1
Other		
Sequence data, analyses, and resources related to the ultra-deep sequencing of the AML31 tumor, relapse, and matched normal.	This paper	http://aml31.genome.wustl.edu
Resource website for the AML31 publication	This paper	https://github.com/chrisamiller/aml31SuppSite

Supplemental Table 2

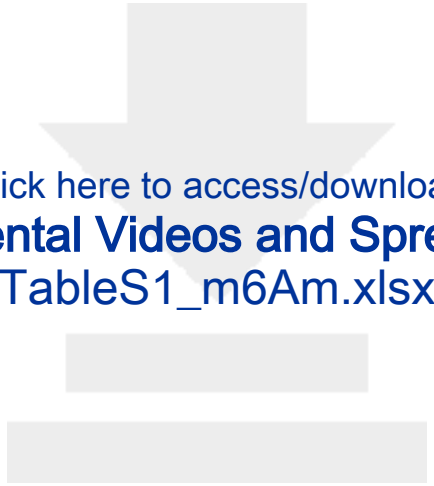
Supplemental Table 3

Click here to access/download

Supplemental Videos and Spreadsheets

TableS3_FunctionalAnnotations.xlsx

Supplemental Table 1



Click here to access/download

Supplemental Videos and Spreadsheets

TableS1_m6Am.xlsx