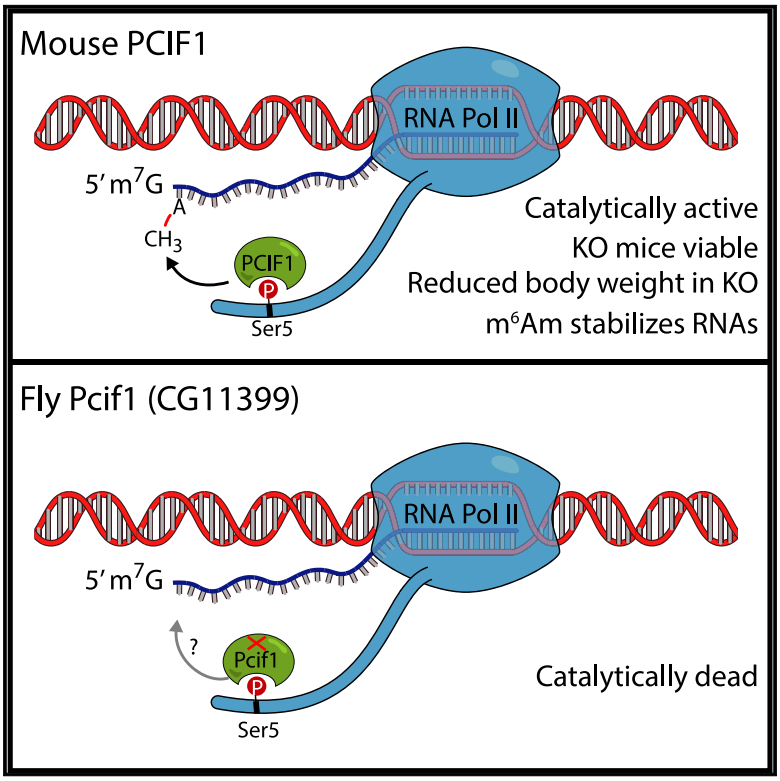
Cell Reports

The Mammalian Cap-Specific m^{6}Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues

Graphical Abstract



Highlights

- Mouse mutants of the m^{6}Am methylase *Pcif1* display reduced body weight
- Transcripts with a TSS adenosine are destabilized in the *Pcif1* mutant mice
- Catalytically dead *Drosophila* Pcif1 binds Ser5-phospho CTD
- *Trypanosoma* Pcif1 is an m^{6}Am methylase that creates the cap4 structure

Authors

Radha Raman Pandey, Elena Delfino, David Homolka, ..., Emmanuel Taillebourg, Marie-Odile Fauvarque, Ramesh S. Pillai

Correspondence

raman.pandey@unige.ch (R.R.P.), ramesh.pillai@unige.ch (R.S.P.)

In Brief

Pandey et al. demonstrate that a loss of the cap-specific m^{6}Am RNA methylation in mice destabilizes transcripts and that such *Pcif1* mutant mice have reduced body weight. In contrast, *Drosophila* Pcif1 is inactive as a methylase but, like its mammalian counterpart, still binds to Ser5-phosphorylated RNA Pol II CTD.

Pandey et al., 2020, Cell Reports *32*, 108038
August 18, 2020
https://doi.org/10.1016/j.celrep.2020.108038

Cell Reports

CellPress
OPEN ACCESS

Report

The Mammalian Cap-Specific m^{6}Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues

Radha Raman Pandey,[1,4,*] Elena Delfino,[1,4] David Homolka,[1,4] Adriana Roithova,[1] Kuan-Ming Chen,[1] Lingyun Li,[1] Giulia Franco,[2] Cathrine Broberg Vågbø,[3] Emmanuel Taillebourg,[2] Marie-Odile Fauvarque,[2] and Ramesh S. Pillai[1,5,*]

[1]Department of Molecular Biology, Science III, University of Geneva, 30 Quai Ernest-Ansermet, 1211 Geneva, Switzerland
[2]University Grenoble Alpes, CEA, INSERM, BGE, 38000 Grenoble, France
[3]Proteomics and Modomics Experimental Core (PROMEC), Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU) and St. Olavs Hospital Central Staff, Trondheim, Norway
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: raman.pandey@unige.ch (R.R.P.), ramesh.pillai@unige.ch (R.S.P.)
https://doi.org/10.1016/j.celrep.2020.108038

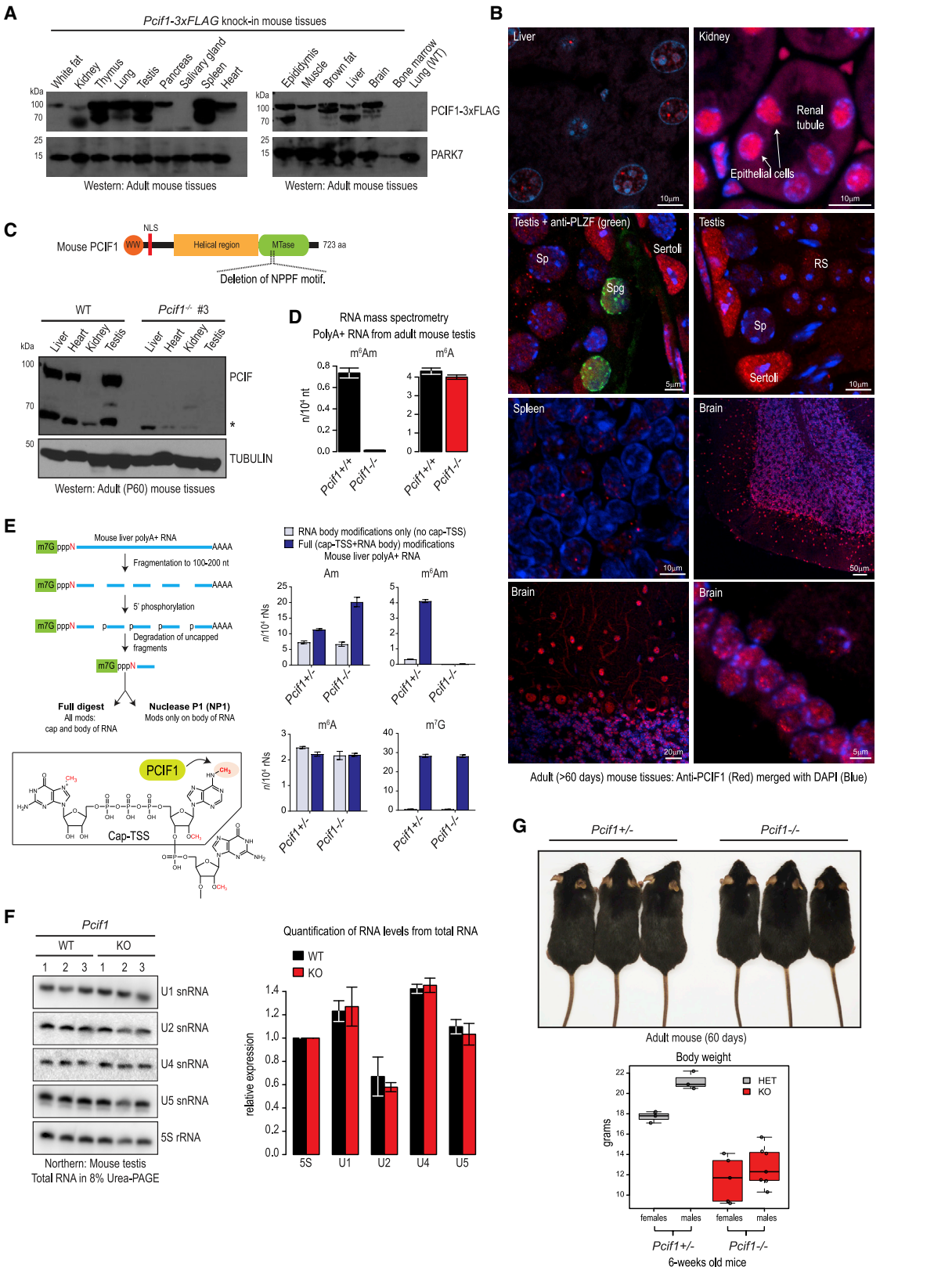
SUMMARY

The 5′ end of eukaryotic mRNAs is protected by the m^{7}G-cap structure. The transcription start site nucleotide is ribose methylated (Nm) in many eukaryotes, whereas an adenosine at this position is further methylated at the N^6 position (m^{6}A) by the mammalian Phosphorylated C-terminal domain (CTD)-interacting Factor 1 (PCIF1) to generate m^{6}Am. Here, we show that although the loss of cap-specific m^{6}Am in mice does not affect viability or fertility, the *Pcif1* mutants display reduced body weight. Transcriptome analyses of mutant mouse tissues support a role for the cap-specific m^{6}Am modification in stabilizing transcripts. In contrast, the *Drosophila* Pcif1 is catalytically dead, but like its mammalian counterpart, it retains the ability to associate with the Ser5-phosphorylated CTD of RNA polymerase II (RNA Pol II). Finally, we show that the *Trypanosoma* Pcif1 is an m^{6}Am methylase that contributes to the N^6,N^6,2′-*O*-trimethyladenosine (m$^6{}_2$Am) in the hypermethylated cap4 structure of trypanosomatids. Thus, PCIF1 has evolved to function in catalytic and non-catalytic roles.

INTRODUCTION

The 5′ end of eukaryotic RNA polymerase II (RNA Pol II) transcripts is protected with a N^7-methylguanosine (m^{7}G) cap structure (Furuichi et al., 1975a; Shatkin, 1976). This cap0 structure is added to the nascent RNA (~20–25 nucleotide [nt]) by an inverted 5′-5′ triphosphate linkage to the 5′ terminal transcription start site (TSS) nucleotide (Inesta-Vaquera and Cowling, 2017; Mao et al., 1995; Muthukrishnan et al., 1975; Shafer et al., 2005; Shatkin and Manley, 2000). The ribose of the TSS nucleotide in most eukaryotes is 2′-*O*-methylated (Am, Um, Cm, or Gm) by the action of CMTR1 to become the cap1 structure (Bélanger et al., 2010; Byszewska et al., 2014; Haline-Vaz et al., 2008; Wei et al., 1975b). CMTR1 uses its C-terminal WW domains to interact with the RNA Pol II C-terminal domain (CTD) (Haline-Vaz et al., 2008), enabling it to access the nascent RNAs. Some of the RNA Pol II transcripts also have a ribose methylation on the second transcribed nucleotide catalyzed by CMTR2, creating the cap2 structure (Furuichi et al., 1975b; Werner et al., 2011). It was recognized early on that when the TSS nucleotide is an adenosine, it can be further methylated at the N^6 position to create the dimethylated adenosine referred to as m^{6}Am (Wei et al., 1975a).

Apart from these cap-specific modifications, RNA Pol II transcripts contain abundant internal N^6-methyladenosine (m^{6}A) marks (Perry et al., 1975; Schibler et al., 1977; Wei et al., 1975b) that are catalyzed by the RNA methyltransferase heterodimer composed of METTL3-METTL14 (Liu et al., 2014; Śledź and Jinek, 2016; Wang et al., 2016). Mapping experiments with anti-m^{6}A antibodies reveal its presence at thousands of sites across the transcriptome, with multiple marks distributed along the length of the transcript body (Dominissini et al., 2012; Meyer et al., 2012; Schwartz et al., 2013). Additionally, the METTL16 writer catalyzes the internal m^{6}A mark on specific structured RNA contexts within a select set of RNAs that include the U6 small nuclear RNA (snRNA) and the S-Adenosyl methionine (SAM) synthetase mRNA (Doxtader et al., 2018; Mendel et al., 2018; Pendleton et al., 2017; Warda et al., 2017). These internal marks alter the fate of the RNA in terms of splicing, choice of polyadenylation site, RNA export, and RNA stability and translation. These consequences are mediated by reader proteins, especially those from the YTH family that can recognize the internal m^{6}A marks (Fu et al., 2014; Patil et al., 2018; Roignant and Soller, 2017). The essential role of the internal m^{6}A marks is highlighted by the embryonic lethality exhibited by mouse mutants of METTL3 (Batista et al., 2014; Geula et al., 2015), METTL16



(legend on next page)

(Mendel et al., 2018), and YTHDC1 (Kasowitz et al., 2018), whereas mutants of YTHDC2 (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017) and YTHDF2 (Ivanova et al., 2017) lead to infertility in mice.

Unlike the internal m^6A marks that have been shown to mediate RNA decay by the action of YTH proteins (Ke et al., 2017; Wang et al., 2014), the cap-specific m^6Am on the TSS adenosine is demonstrated to promote RNA stability by providing resistance to the action of the mRNA decapping enzyme DCP2 (Mauer et al., 2017). An activity responsible for cap-specific m^6Am methylation was partially purified as a ~65-kDa protein from human HeLa cell extracts (Keith et al., 1978), and this was recently revealed to be Phosphorylated CTD-Interacting Factor 1 (PCIF1) (Akichika et al., 2019; Boulias et al., 2019; Sendinc et al., 2019; Sun et al., 2019). PCIF1 was originally identified as a factor interacting with the phosphorylated CTD of RNA Pol II by its N-terminal WW domain (Fan et al., 2003; Hirose et al., 2008) and was later shown to have an affinity for the Ser5-phosphorylated CTD (Akichika et al., 2019). Methylation by PCIF1 seems to be efficient in human HEK293T cells, as up to 92% of m^7G-capped RNAs with a TSS adenosine carry the m^6Am mark, with the rest being Am (Akichika et al., 2019).

Early studies with *Pcif1* knockout (KO) chicken B cell DT40 lines did not identify any impact on cell cycle or viability (Yunokuchi et al., 2009). Removal of *PCIF1* in human cell lines also did not affect cell viability or growth (Akichika et al., 2019; Boulias et al., 2019; Sendinc et al., 2019). Investigations with *PCIF1* mutant human cells have supported the RNA stabilization role for m^6Am (Boulias et al., 2019) or implicated it in either enhancing (Akichika et al., 2019) or inhibiting (Sendinc et al., 2019) cap-dependent translation. Thus, although PCIF1-mediated m^6Am methylation is shown to impact diverse aspects of gene expression, its physiological role is currently not known. Here, we show that the cap-specific m^6Am modification catalyzed by PCIF1 is not essential for viability and fertility in mice. However, the *Pcif1* mutant mice display reduced body weight. Confirming an RNA-stabilizing role for m^6Am, transcripts starting with m^6Am are downregulated in the *Pcif1* mutant mouse tissues.

RESULTS

Loss of Mouse *Pcif1* Is Tolerated but Mutants Display Reduced Body Weight

To study the *in vivo* tissue distribution of mammalian PCIF1, we created a knockin mouse expressing a C-terminal FLAG-tagged version of the protein (Figure S1A and S1B; STAR Methods). The western blot analysis shows a near-ubiquitous expression of the protein at ~90 kDa, which is abundantly seen in the thymus, testis, and spleen. Expression in the kidney is very low, and the protein is not detectable in the salivary gland (Figure 1A). Similar results were obtained when detecting the endogenous protein in wild-type mouse tissues (Figure S1C). We consistently observed a second band at ~60 kDa in many of the tissues from both the wild-type and knockin mice (Figure 1A; Figure S1C), which we believe is an N-terminal truncated version of the protein (see STAR Methods). Immunofluorescence analysis using the anti-PCIF1 antibodies reveals a predominantly nuclear accumulation of PCIF1 in all tissues examined (Figure 1B).

To determine its physiological role, we created a *Pcif1* KO mouse model (Figures S1A and S1B; STAR Methods). Both heterozygous (HET) $Pcif1^{+/-}$ and homozygous $Pcif1^{-/-}$ KO animals of both sexes are viable, fertile, and born in the expected Mendelian ratio, indicating the absence of embryonic lethality (Figure S1D). Our western analyses of tissue lysates from the $Pcif1^{-/-}$ KO animals reveal a complete absence of the full-length protein, indicating that they are null mutants (Figures 1C and S1E). Consistent with the loss of the PCIF1 protein, the m^6Am modification is absent in polyA+ RNA from the $Pcif1^{-/-}$ KO testis (Figure 1D). We prepared ~200-nt m^7G-capped fragments of polyA+ RNAs from mouse liver (see STAR Methods) and analyzed separately the modifications on the full fragment (cap-TSS+ RNA body) versus that present only internally within the RNA body (no cap-TSS). The m^6Am modification is only detected on the TSS nucleotide and is lost in the $Pcif1^{-/-}$ mutant (Figure 1E). Expectedly, there is a slight increase of Am modification in the mutant, as it would normally be converted to m^6Am. Importantly, the overall levels of m^6A are not affected in the $Pcif1^{-/-}$ mutant (Figures 1D and 1E). Thus, mouse PCIF1 is the

Figure 1. *Pcif1* Mutant Mice Display Reduced Body Weight

(A) Multiple-tissue western analysis of *Pcif1-3xFLAG* knockin mouse showing presence of PCIF1-3xFLAG in most tissues. See also Figure S1C. Lung tissue from a wild-type (WT) animal is used to control signal from the anti-FLAG antibody. Signal from the PARK7 protein is used as a loading control.

(B) Subcellular localization of endogenous PCIF1 in mouse tissues. The protein is nuclear with some punctae observed (see liver and spleen). In testis sections, the protein is detected in the nucleus of PLZF-marked (green) undifferentiated spermatogonia (Spg), meiotic spermatocytes (Sp), post-meiotic round spermatids (RS), and somatic Sertoli cells.

(C) Domain architecture of mouse PCIF1. WW, tryptophan-rich domain; NLS, nuclear localization signal; helical region; MTase, methyltransferase domain. Creation of *Pcif1* knockout (KO) mouse by deletion of the region encoding the catalytic residues. See also Figures S1A and S1B. Western analysis with multiple tissues shows the absence of the protein in the KO ($Pcif1^{-/-}$) animals. See also Figure S1E for other biological replicates.

(D) RNA mass spectrometry confirms absence of m^6Am in polyA+ transcripts from the *Pcif1* KO mutant testis.

(E) Scheme showing preparation of m^7G-capped RNA fragments for RNA mass spectrometry to detect different modified nucleotides. Such fragments were treated to identify internal RNA modifications within the body of the transcript (excluding the cap and TSS nucleotide) and those from the entire fragment. RNA from mouse liver of indicated *Pcif1* genotypes was analyzed. Chemical structure of the eukaryotic cap structures indicating location of the m^6Am modification catalyzed by PCIF1 is shown.

(F) Northern analyses of total RNA from WT and *Pcif1* KO mutant mouse testis to detect U snRNAs. Signal from 5S rRNA is used as a loading control. The numbers indicate the biological triplicates tested. Quantification of signal intensities is shown on the right. Error bars show SD. See also Figure S1F for northern analyses with other tissues.

(G) Representative picture of adult $Pcif1^{-/-}$ KO mutant mice and control $Pcif1^{+/-}$ heterozygous (HET) littermates. The KO animals (6-weeks old) show reduced body weight. See also Figure S1G.

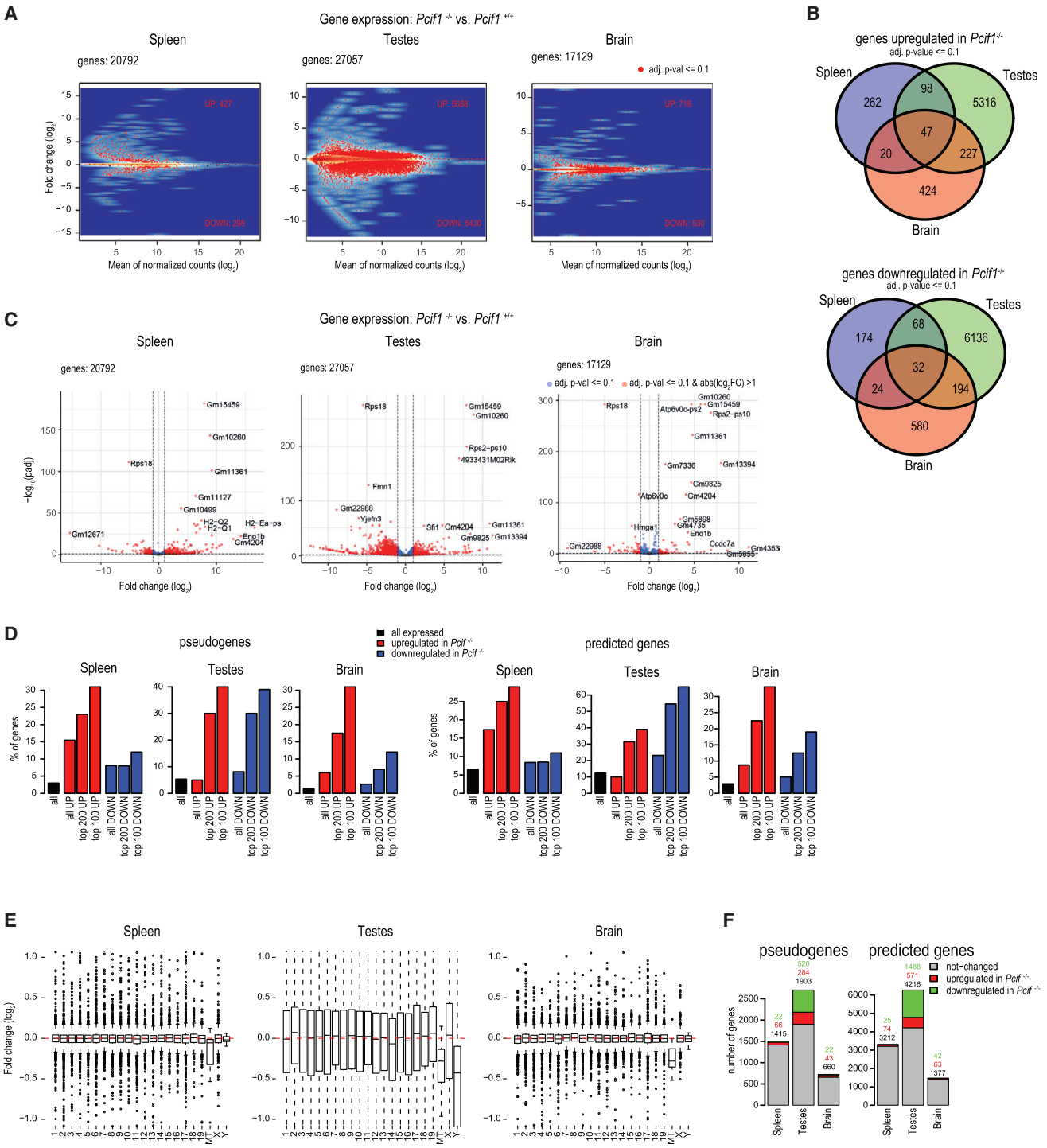


Figure 2. Pseudogene and Predicted Gene Transcript Levels Altered in the *Pcif1* Mutant Mouse Tissues

(A) MA plot showing transcript changes in the *Pcif1*$^{-/-}$ mutant mouse spleen, testes, and brain. Testicular transcriptome shows alteration of thousands of transcripts. See also Figure S2A.

(B) Venn diagram showing overlap in sets of upregulated and downregulated genes in different *Pcif1*$^{-/-}$ mutant mouse tissues.

(C) Volcano plot showing significant changes in levels for pseudogenes and predicted gene model (Gm) transcripts in the *Pcif1*$^{-/-}$ mutant tissues.

(D) Bar plots showing the enrichment of pseudogenes and predicted genes in the transcripts that are dysregulated in the *Pcif1*$^{-/-}$ mutant.

(legend continued on next page)

sole enzyme catalyzing the m^6Am modification on the TSS adenosine of m^7G-capped polyA+ transcripts.

Apart from messenger RNAs, RNA-Pol-II-transcribed splicing snRNAs like U1, U2, U4, and U5 also have the m^6Am modification on the TSS adenosine (Mauer et al., 2019; Wei et al., 2018). To examine the impact of the loss of PCIF1 on U snRNAs, we directly assessed their levels by northern blotting. In biological triplicates of total RNA from liver, spleen, and testis, we did not see any dramatic change in snRNA levels in the *Pcif1*$^{-/-}$ mutant (Figures 1F and S1F). Interestingly, despite the lack of any viability or fertility defects, we noticed that the *Pcif1*$^{-/-}$ mutant animals of both sexes consistently display reduced body weight compared to their HET littermates (Figure 1G). The overall body length of the *Pcif1*$^{-/-}$ mutants is not changed (Figure S1G), indicating that weight reduction is not due to a difference in growth during development. Taken together, our findings show that although a loss of mouse *Pcif1* is tolerated, the mutants have reduced body weight.

Dysregulation of Pseudogene and Predicted Gene Transcripts in *Pcif1* Mutant Mouse Tissues

To examine the global gene expression changes in the *Pcif1*$^{-/-}$ mutant, we carried out RNA sequencing (RNA-seq) analysis with total RNA from three different mouse tissues (Table S3). All the tissues show altered gene expression, with the testicular transcriptome being the most impacted (Figures 2A and S2A). Hundreds (in spleen and brain) to thousands (in testis) of transcripts are either up- or downregulated. There is no considerable overlap in terms of the altered transcripts across the tissues (Figure 2B), supporting the idea that PCIF1 regulates transcripts that are largely specific to each tissue. Surprisingly, among the top altered sequences in all the tissues are several pseudogenes and poorly characterized predicted protein-coding Gene model (Gm) transcripts (Figure 2C). Because some of the pseudogenes are annotated as predicted genes and vice versa, this is a largely overlapping set of transcription units.

Pseudogenes are inactive copies of functional genes that arose during evolution, either by tandem duplication or activity of retrotransposons (Pink et al., 2011). Compared to their parental genes, they have acquired mutations that render them non-functional in terms of protein-coding potential. Although duplicated pseudogenes still maintain their intron-exon organization, the processed pseudogenes arise from mature cDNAs due to retrotransposition activity of endogenous transposons. Pseudogenes are as numerous as protein-coding genes (up to 20,000 in the human genome), with the bulk of them being processed as pseudogenes (Zhang and Gerstein, 2004). Although many of the pseudogenes are not transcribed due to a loss of promoter sequences or due to an insertion of processed pseudogenes into regions lacking promoter activity, about 2%–20% of the human pseudogenes are transcribed (Harrison et al., 2005; Zheng et al., 2005). In our mouse tissue RNA-seq datasets, pseudogenes constitute only a small part (3% of spleen, 5% of testis, and 2% of brain) of the expressed genes in these tissues. Strikingly, despite being only a small part of the transcriptome, the pseudogenes constitute a significant chunk (30%–40%) of the top 100 upregulated genes (Figure 2D). Pseudogenes also form a substantial part (10%–40%) of the top 100 downregulated genes (Figure 2D). Similarly, predicted Gm transcripts constitute up to 25%–40% of the top 100 upregulated and 10%–60% of the top downregulated genes in the *Pcif1* mutant (Figure 2D).

The altered transcripts in the *Pcif1*$^{-/-}$ tissues do not show any particular genomic origins (Figure S2B). Interestingly, most of the testicular transcripts arising from the Y chromosome are largely downregulated (Figure 2E), with many being pseudogenes (Figure S2B). Likewise, transcripts from the mitochondrial genome are consistently downregulated in all the three tissues (Figure 2E). Because PCIF1 likely acts only in the nuclear compartment, we expect the change in mitochondrial transcriptome to be an indirect consequence of other gene expression changes. Although changes in the expression of pseudogene and predicted genes in the *Pcif1*$^{-/-}$ mutant are striking, they constitute only a small subset of such expressed sequences (Figure 2F). Taken together, we identify a set of pseudogenes and predicted Gm transcripts as being regulated by mouse PCIF1, with Y-chromosome-derived pseudogenes being heavily downregulated in the mouse testis.

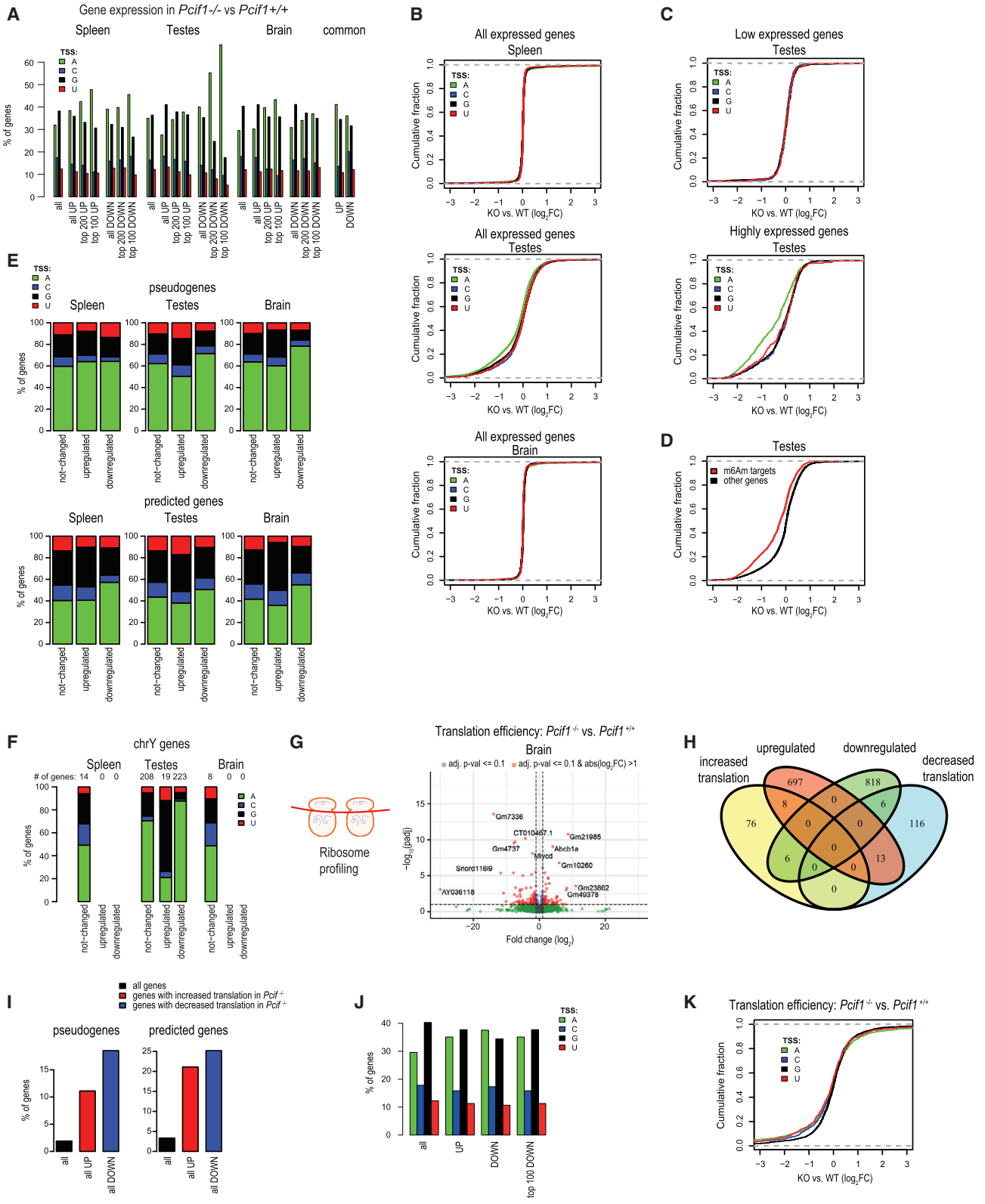
Transcripts with a TSS Adenosine Are Downregulated in *Pcif1* Mutant Mouse Tissues

To identify a link between the observed transcriptome changes and the activity of PCIF1, we first sorted the genes based on their annotated TSS nucleotide (STAR Methods). As expected (Carninci et al., 2006; Yamashita et al., 2006), most genes produce transcripts that start with either an adenosine or guanosine. By comparing the gene expression in the *Pcif1*$^{-/-}$ mutant with the control wild type (*Pcif1*$^{+/+}$), we calculated the proportion of dysregulated genes with a specific TSS nucleotide. Strikingly, in the testis datasets, we see an enrichment of transcripts starting with an adenosine among the top 100 and top 200 downregulated genes in the *Pcif1* mutant (Figure 3A). Likewise, when all the expressed genes are considered, those with an adenosine as the TSS nucleotide generally show lower expression in the *Pcif1*$^{-/-}$ mutant testes (Figure 3B). When the genes in the respective tissues were further sorted as high or low expressed (STAR Methods), we found that mainly highly expressed genes with a TSS adenosine are downregulated in the *Pcif1*$^{-/-}$ mutant testes (Figure 3C). Such a global relationship was not observed in the spleen and brain samples (Figures 3B and S3A) but was seen for specific transcript sets (see below).

To further link the change in testicular RNA levels to a loss of the m^6Am methylation, we experimentally defined a high-confidence set of m^6Am-target genes by using a previously described protocol (Sendinc et al., 2019). Briefly, ~200-nt capped RNA fragments from mouse testes were prepared and used for m^6A-IPseq, and enrichments were calculated (STAR Methods).

(E) Boxplots showing gene expression changes from individual chromosomes, as observed in the different *Pcif1*$^{-/-}$ mutant mouse tissues. The view is limited to only changes below 2-fold to better see the median differences. See also Figure S2B.

(F) Bar plots showing the total number of pseudogenes and predicted genes detected in the indicated mouse tissues and the fraction of those that are up- and downregulated in the *Pcif1*$^{-/-}$ mutant.



(legend on next page)

The anti-m^{6}A antibody which does not differentiate between m^{6}A and m^{6}Am, enriches the capped RNA fragments (Figure S3B). Supporting the fact that some of this enrichment is due to cap-specific m^{6}Am catalyzed by PCIF1, there is a clear decrease in the distribution of reads over the TSS in the *Pcif1*$^{-/-}$ mutant (Figure S3B). Furthermore, a clear and significant decrease (Figure S3C) in the peak is observed mainly at those TSSs with an annotated adenosine, indicating they are likely m^{6}Am targets *in vivo* (Table S4). Notably, the annotated TSS is the most dominant documented one from a collection of start sites at a transcription unit, so it is possible that the decrease seen in the *Pcif1*$^{-/-}$ mutant at non-adenosine TSS is just a reflection of this heterogeneity (Figure S3C). Importantly, when the expression levels of such high-confidence m^{6}Am-target genes were plotted, they show a clear decrease in the *Pcif1*$^{-/-}$ mutant testes (Figure 3D). When the TSS nucleotide was specifically examined for pseudogenes, we found those with an adenosine as the TSS to be downregulated in the testis and brain datasets (Figure 3E). Similarly, predicted genes from all three tissues that are downregulated tend to have an adenosine as the TSS nucleotide (Figure 3E). As noted previously, the Y chromosome genes tend to be downregulated in the *Pcif1*$^{-/-}$ mutant (Figure 2E), and ~90% of these downregulated genes use an adenosine as the TSS nucleotide (Figure 3F). These findings are consistent with the previous observation that the m^{6}Am modification has a transcript-stabilizing role (Mauer et al., 2017).

Finally, we examined if there is any impact on translation by carrying out ribosome profiling experiments with a *Pcif1* mutant brain (Table S5). An analysis of ribosome footprints indicates a significant alteration in the translation of several transcripts (Figure 3G), without any changes in transcript levels for most of them (Figure 3H). We find some predicted genes and pseudogenes to be among those that show changes in translation (Figure 3I). Note that although pseudogenes have mutations that prevent full-length protein production, this does not per se preclude their entry into ribosomes. However, the genes that show translation changes do not have any particular preference for the TSS nucleotide (Figures 3J and 3K). Considered together, these data indicate that cap-specific m^{6}Am methylation of RNA Pol II transcripts by mouse PCIF1 has an RNA-stabilizing role in mouse tissues, with a subset of pseudogenes and poorly characterized predicted genes being a common set of regulated transcripts.

Drosophila Pcif1 Is Catalytically Dead but Binds Ser5-Phosphorylated RNA Pol II CTD

The m^{6}Am modification is so far detected only in vertebrates (Sendinc et al., 2019). It is absent in worms and insects (Figures 4A and S4A). Although Pcif1 is absent in yeasts and worms, a clear ortholog with a high level of conservation is detected in several insect genomes (Figure S4B). It is represented in the fruit fly *Drosophila* by the uncharacterized gene CG11399 (dPcif1). Rabbit polyclonal antibodies raised against dPcif1 detects the endogenous protein in *Drosophila* ovaries and in the *Drosophila* S2 cell line (Figure 4B). Nevertheless, RNA mass spectrometry did not detect the m^{6}Am modification in adult flies, early fly embryos (2 h post-fertilization), and in the S2 cell line (Figure 4C). Transfecting S2 cells with a plasmid expressing human PCIF1 (hPCIF1) shows the ectopic installation of m^{6}Am on fly transcripts (Figure 4C), validating our detection strategy. The fly protein is nuclear in transfected *Drosophila* S2 cells (Figure S4C). Similarly, the endogenous dPcif1 protein is also nuclear in fly embryos (Figure S4D) and adult fly ovaries, in which it is excluded from the nucleoli that are positively stained by fibrillarin (Figure 4D). The protein is detected in both the somatic follicle cells and germline nurse cells of individual egg chambers in the fly ovaries (Figure 4D).

Examination of a protein sequence alignment shows that compared to mammalian PCIF1, the insect orthologs have mutations in the putative catalytic residues that can potentially render them inactive (Figures 4E and S4B). To directly verify this finding, we expressed a recombinant version of the full-length *Drosophila* Pcif1 protein (Figure S5A) and incubated it with m^{7}G-capped RNAs, together with radioactive ^{14}C-SAM as the methyl-donor. Recombinant hPCIF1 was used as a positive control (Figure S5A). As expected, the human protein shows methylation activity with an m^{7}G-capped RNA having an adenosine as the TSS nucleotide (m^{7}GpppA-RNA). However, consistent with mutation of the putative catalytic residues, the fly protein was

Figure 3. Transcripts Starting with an Adenosine Are Downregulated in the Mouse *Pcif1* Mutant

(A) Percentage of genes with a specific nucleotide as a transcription start site (TSS) is compared between all expressed genes, and the genes found to be up- and downregulated in the *Pcif1*$^{-/-}$ KO mutant mouse tissues compared to the WT, *Pcif1*$^{+/+}$) control.

(B) Plots showing the cumulative distributions of gene expression changes in the *Pcif1*$^{-/-}$ KO mutant mouse tissues compared to the WT control. The distributions of genes with different TSS nucleotides are shown separately.

(C) Plots showing the cumulative distributions of gene expression changes in the *Pcif1*$^{-/-}$ mutant mouse testes separately for low- and high-expressed genes. See also Figure S3A.

(D) Plot showing the cumulative distributions of gene expression changes in the *Pcif1*$^{-/-}$ mutant testes for those genes experimentally determined to express transcripts with m^{6}Am (m^{6}Am targets) and those without m^{6}Am (other genes). See also Figures S3B and S3C.

(E) Bar plots showing the percentage of pseudogenes and predicted genes with specific TSS nucleotides among the genes up- and downregulated in the *Pcif1*$^{-/-}$ mutant mouse tissues.

(F) Bar plots showing percentage of chromosome Y pseudogenes and predicted genes with specific TSS nucleotides among the genes with same or differential expression in the *Pcif1*$^{-/-}$ mutant mouse tissues compared to the WT control.

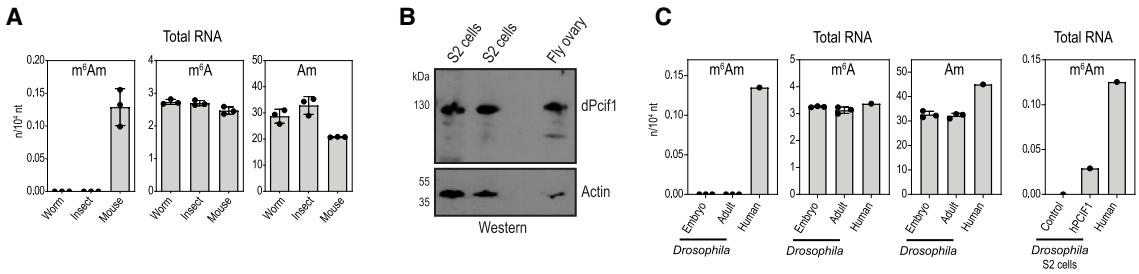
(G) Ribosome profiling of the *Pcif1*$^{-/-}$ mutant brain. Volcano plot shows changes in translation efficiency in the mutant compared to that of WT control. Several predicted genes are dramatically altered in their translation.

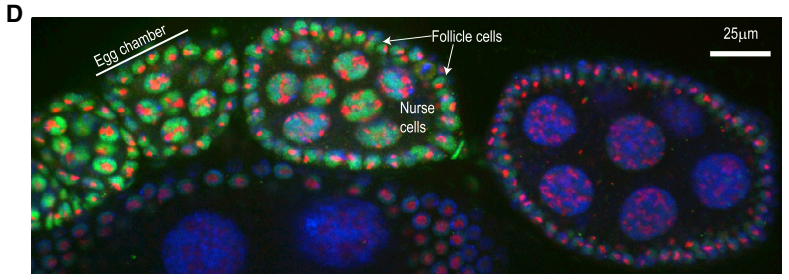
(H) Venn diagram comparing changes (upregulated and downregulated) in transcript levels and translation in the *Pcif1*$^{-/-}$ mutant brain.

(I) Bar plots showing the enrichment of predicted genes and pseudogenes among the genes with altered translation efficiency in *Pcif1*$^{-/-}$ mutant brain.

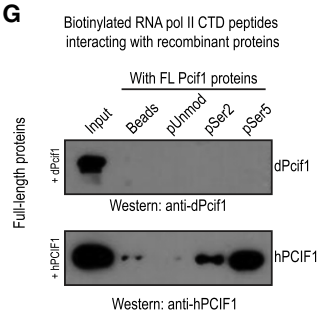
(J) Percentage of genes with specific TSS nucleotides is compared between genes with increased and decreased translation efficiency in the *Pcif1*$^{-/-}$ mutant brain.

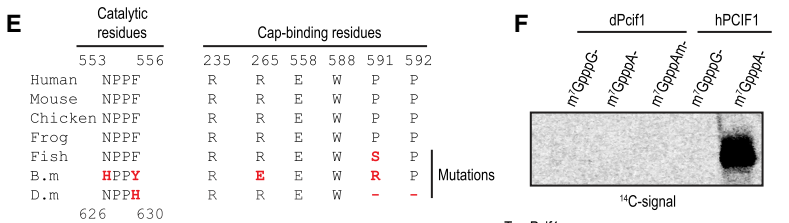
(K) Plot showing the cumulative distribution of changes in translation efficiency in the *Pcif1*$^{-/-}$ mutant mouse brain compared to the WT control.

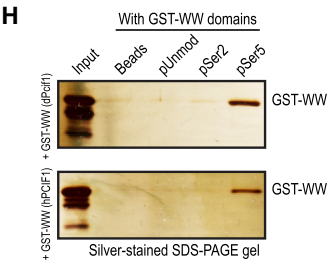


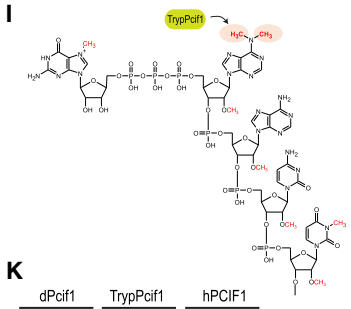


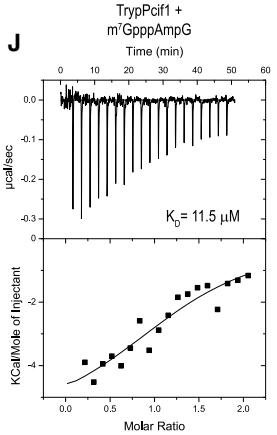
Immunofluorescence with adult fly ovary (5-days old): anti-dPcif1 (green), anti-Fibrillarin (red), DAPI (blue)

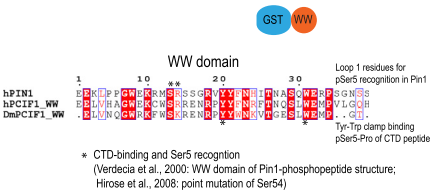


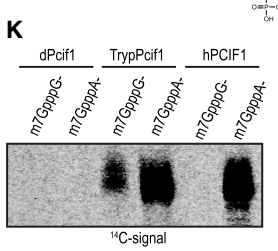


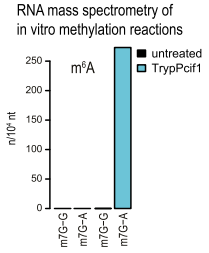


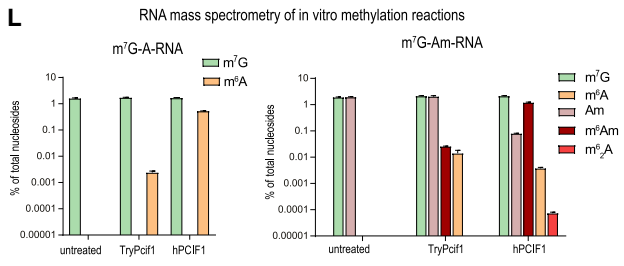












(legend on next page)

inactive as an RNA methylase on all tested RNAs (Figure 4F). Next, we examined whether the fly Pcif1 can bind the m^7G cap structure using isothermal calorimetry (ITC) experiments. Such direct binding experiments reveal that the recombinant full-length fly Pcif1 protein does not associate with the cap analogs m^7GpppA or m^7GpppG (Figure S5B). This is consistent with the fact that some of the cap-binding residues conserved in vertebrate PCIF1 proteins are absent in the fly protein (Figures 4E and S4B).

Next, we examined the CTD-interaction activity of fly Pcif1. The RNA Pol II CTD is composed of multiple repeats of a heptad sequence (YSPTSPS) that can be phosphorylated at different Serine (Ser) residues (Buratowski, 2009). We directly tested this interaction and confirm that recombinant full-length hPCIF1 preferentially associates with the Ser5-phosphorylated CTD peptide (Figures 4G and S5C). However, the full-length fly Pcif1 failed to bind any of the tested CTD peptides (Figures 4G and S5C). This finding was surprising, as the overall WW domain of PCIF1 is highly conserved from insects to human (Figure 4H) and is similar to that found in the prolyl isomerase hPIN1. The crystal structure of the hPIN1 WW domain in complex with the phospho CTD peptide is solved (Verdecia et al., 2000). This reveals key WW domain residues like the Tyr-Trp clamp (Y61-W72 in hPCIF1) that holds the Ser5-Pro motif of the CTD peptide [YSPT(pS)PS], and the Ser-Arg/Lys (S54-R55 in hPCIF1) of loop1 that recognizes the phosphorylated Ser5. Mutation of the S54 in the hPCIF1 WW domain abolishes the interaction with the CTD (Hirose et al., 2008). All these key residues are also conserved in the fly Pcif1 (Figure 4H). Importantly, when the binding assay was performed with the isolated WW domains from both proteins, fly Pcif1 also specifically binds the Ser5-phosphorylated CTD peptide (Figures 4H and S5D).

Taken together, our experiments reveal that the fly Pcif1 is biochemically inert as an RNA methylase and does not bind the m^7G cap structure. However, it retains an active WW domain that is able to bind Ser5-phosphorylated CTD of the RNA Pol II. We propose that fly Pcif1 might have evolved to function as a nuclear factor that transcriptionally regulates gene expression.

Trypanosoma Pcif1 Catalyzes m^6Am That Is Part of the cap4 Structure

The 5′ end of RNA Pol II transcripts in most eukaryotes is defined by transcription initiation. In contrast, nematodes receive a trimethylated guanosine cap ($m^{2,2,7}G$) by *trans*-splicing of a spliced leader (SL) sequence (Lasda et al., 2011), but they lack m^6Am due to the absence of a PCIF1 ortholog. The situation with trypanosomatid protozoans is partly similar to that of nematodes in that they also acquire an extensively modified cap4 structure (m^7Gpppm$^6{}_2$Am-p-Am-p-Cm-p-m^3Um) by *trans*-splicing. The structure consists of the first four nucleotides being ribose methylated (Nm), in addition to having a N^6,N^6-dimethyl modification on the first adenosine ($m^6{}_2$A), and N^3-methyl modification on the uridine (m^3U) at the fourth position (Figure 4I). The cap4 structure is first formed on the specialized SL RNA and then transferred to the pre-mRNAs by a *trans*-splicing reaction (Mair et al., 2000). Because the activity involved in the formation of the $m^6{}_2$A in the cap4 structure is currently not known, we wondered whether a Pcif1 ortholog could play a role.

We identified the Pcif1 ortholog in *Trypanosoma cruzi*, the trypanosomatid responsible for South American trypanosomiasis (Chagas disease). The *Trypanosoma* Pcif1 (TrypPcif1) lacks a WW domain but contains the NPPF catalytic residues as present in the vertebrate PCIF1 (Figure S5E). To test its biochemical properties, we expressed the TrypPcif1 as a recombinant protein (Figure S5F). TrypPcif1 has a strong affinity (dissociation constant, K_D = 11.5–29.9 μM) for the cap1 structure represented by the m^7GpppAmpG cap analog (Figures 4J and S5F), whereas binding to m^7GpppA and m^7GpppG are

Figure 4. *Drosophila* Pcif1 Is Catalytically Dead but Interacts with the Ser5-Phosphorylated RNA Polymerase II (RNA Pol II) C-Terminal Domain (CTD)

(A) RNA mass spectrometry analysis of total RNA from worm (*Caenorhabditis elegans*), insect (*Bombyx* BmN4 cell line), and adult mouse testis RNA. Worms and insects lack the m^6Am modification. See also Figure S4A.

(B) Western analysis of endogenous dPcif1 in fly ovaries and *Drosophila* S2 cell line.

(C) RNA mass spectrometry analysis of adult *Drosophila*, early (2-h) fly embryos, and human HAP1 cell line as a control. Also used was RNA from *Drosophila* S2 cells with or without (control) transfection of a plasmid expressing human PCIF1. See also Figure S4A.

(D) dPcif1 protein is detected in adult fly ovaries. Egg chambers show the protein being nuclear in the single layer of somatic follicle cells and the germline nurse cells contained within each chamber. See also Figures S4C and S4D.

(E) Insect Pcif1 proteins (B.m., *Bombyx mori*; D.m., *Drosophila melanogaster*) have mutations of the catalytic residues in the MTase and in the critical residues that are required for m^7G cap binding. See also Figure S4B.

(F) *In vitro* methylation assays with full-length recombinant *Drosophila* or human PCIF1 proteins. The reactions used m^7G-capped RNAs with either an adenosine or a guanosine or Am as the TSS nucleotide and radioactive ^{14}C-SAM as the methyl donor. See also Figure S5A.

(G) *In vitro* binding assay between recombinant full-length fly or human PCIF1 and different biotinylated peptides with heptad repeats from the CTD of RNA Pol II. Interaction was assayed by western blotting with specific antibodies. Phosphorylation at Ser2 or Ser5 of the CTD heptad sequence is indicated. Unmod, non-phosphorylated peptide; beads, beads alone without any bound peptide. See also Figure S5C.

(H) Alignment of the WW domains from human and *Drosophila* Pcif1 and human PIN1. CTD peptide binding assay with recombinant GST-WW domain fusions from human and fly Pcif1. Bound GST-WW proteins are detected by silver staining. See also Figure S5D.

(I) The cap4 structure of trypanosomatids showing the presence of a dimethyladenine ($m^6{}_2$A) at the TSS nucleotide.

(J) Isothermal calorimetry (ITC) experiment showing binding between recombinant full-length *Trypanosoma* Pcif1 (TrypPcif1) and the indicated cap1 analog. See also Figure S5F.

(K) *In vitro* methylation assays with recombinant full-length PCIF1 proteins from *Drosophila*, *T. cruzi*, and human. The m^7G-capped RNAs have either an adenosine or a guanosine as the TSS nucleotide, and radioactive ^{14}C-SAM was used as methyl donor. Reactions are resolved by urea-PAGE. The size of the capped RNA used is 40 nt. RNA mass spectrometry analysis of *in vitro* methylation reactions carried out with TrypPcif1, revealing the catalysis of m^6A.

(L) RNA mass spectrometry analysis of *in vitro* methylation reactions carried out with human PCIF1 or TrypPcif1 using m^7G-capped RNAs with either A or Am as the TSS nucleotide.

highly reduced (Figure S5F). To test its RNA methylation activity, TrypPcif1 was incubated with an m^7G-capped RNA and radioactive ^{14}C-SAM as a methyl donor. This revealed a robust methylation activity on the m^7GpppA-RNA (Figure 4K). We note that the TrypPcif1 shows an unexplained weak activity with the m^7GpppG-RNA, but this is not seen with the hPCIF1. RNA mass spectrometry analysis of reactions conducted in parallel with non-radioactive SAM shows the clear presence of m^6A in the reactions treated with TrypPcif1 (Figure 4K). When TrypPcif1 is incubated with the m^7GpppAmpG-RNA carrying the cap1 modification, we detect a signal for m^6Am, but the $m^6{}_2$A dimethyladenosine modification was never reliably detected (Figure 4L). Considered together, our findings show that the protozoan Pcif1 is a cap-dependent m^6Am RNA methylase that contributes to the cap4 structure of trypanosomatids.

DISCUSSION

Despite being a ubiquitous mark on RNA Pol II transcripts starting with an adenosine (Akichika et al., 2019), we were surprised to find that a loss of cap-specific m^6Am did not result in dramatic phenotypes in mutant mice (Figure 1). However, a clear contribution to optimal growth is evident, as the *Pcif1* KO mice display a reduced body weight phenotype (Figure 1G). RNA-seq analysis shows that PCIF1 has effects on the abundance of hundreds to thousands of transcripts in the mutant tissues (Figure 2A). One common set of dysregulated transcripts is a subset of pseudogenes and other poorly characterized predicted genes (Figure 2D). Given their sequence homology, pseudogenes that have lost their protein-coding capacity are proposed to regulate their parental genes by acting as non-coding RNAs (Yano et al., 2004). We explored this possibility by first identifying expressed pseudogene/parental gene pairs, and then we compared their levels in the *Pcif1* mutant. Only in the mutant brain we saw a negative correlation between expression levels of pseudogenes and that of their parental genes (Figure S2C). During spermatogenesis, mitotic spermatogonia enter meiosis to become spermatocytes, which later complete meiosis to form haploid spermatids, and further mature into sperm (Oakberg, 1956). We find that the top upregulated transcripts in the *Pcif1*$^{-/-}$ mutant testis are those normally present in mitotic spermatogonia, whereas the top downregulated transcripts are normally those highly expressed in the meiotic spermatocytes and haploid round spermatids (Figure S3D). This finding could suggest an apparent failure to proceed along spermatogenesis. However, we do not see any impact on fertility in the young *Pcif1*$^{-/-}$ mutants, although long-term impact in aged animals needs to be evaluated.

Several molecular roles are attributed to the m^6Am modification. The presence of m^6Am on the cap structure is shown to provide stability to transcripts in human HEK293T cell cultures (Boulias et al., 2019; Mauer et al., 2017). Our own study of the *Pcif1* KO mouse tissues confirms this impact on RNA levels (Figure 3). However, the impact on RNA stability is not an observation that is broadly shared by other studies that examined *Pcif1* KO human cell cultures (Akichika et al., 2019; Sendinc et al., 2019). Instead, they reveal a role for m^6Am in regulating translation, although opposing effects are proposed: promoting (Akichika et al., 2019) or repressing cap-dependent translation (Sendinc et al., 2019). Our own analysis of ribosome footprints in the *Pcif1* mutant mouse brain identified several transcripts that show either up- or downregulated translation (Figure 3G), but we do not find any correlation with the identity of the TSS nucleotide (Figures 3J and 3K). Given the results of earlier studies, a tissue-specific effect on translation should be left open as a possibility.

How the catalytically dead fly Pcif1 finds its target genes and what molecular consequences it has on them are not clear. One possibility is that it interacts with initiating RNA Pol II by its Ser5-phosphorylated CTD (Figure 4H) to transcriptionally regulate gene expression. Because only the isolated WW domain (Figures 4H and S5D), but not the full-length fly Pcif1 (Figures 4G and S5C), is able to interact with the CTD, we speculate that conformational changes might be needed to expose the WW domain. Perhaps interaction with some partner protein might allow this structural change. Interestingly, it was previously noted that due to its affinity for the CTD, the WW domain of hPCIF1 or full-length PCIF1 can interfere with dephosphorylation of the CTD *in vitro* (Hirose et al., 2008). Given that a change in CTD phosphorylation status from Ser5 to Ser2 is critical for transcription elongation (Buratowski, 2009), the observed CTD interaction of fly Pcif1 could have a regulatory role at the transcriptional level.

Finally, our analysis of the *Trypanosoma* Pcif1 identifies it as the activity that catalyzes N^6 methylation of the TSS adenosine within the cap4 structure of the protozoan mRNAs (Figure 4K). The TSS adenosine of the cap4 structure is dimethylated at the N^6 position ($m^6{}_2$A), but we did not find such an activity in the recombinant TrypPcif1 protein. Perhaps the sequential co-transcriptional installation of the other modifications in a hierarchical manner (Mair et al., 2000) is a prerequisite for dimethylation. Indeed, reduced ribose methylation on some of the positions can affect methylation at other sites of the cap4 structure in *Trypanosoma brucei* (Zamudio et al., 2006). Alternatively, TrypPcif1 might install m^6A, with yet another unidentified methylase completing the reaction. Collectively, our combined analyses of PCIF1 in three model systems shed light on its evolutionary functional plasticity in catalytic and non-catalytic roles.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal Work
  - Pcif1 knockout and Flag-Pcif1 knockin mice
  - Synthesis and annealing of gRNA
  - Preparation of injection mix
  - Genotyping
- METHOD DETAILS

- Clones and constructs
- Constructs for *Drosophila* S2 cell expression
- Constructs for recombinant protein production in prokaryotic expression systems
- Constructs for recombinant protein production in insect expression systems
- Recombinant protein production
- Purification of human, fly and Trypanosoma PCIF1
- Purification of GST-fused human and fly PCIF1 WW domain

- ANTIBODIES
  - Commercial antibodies
  - Antibodies generated for this study
  - Antibody purification
  - *Drosophila* S2 cell culture and transfections
  - Mouse multiple tissue western blot
  - Western Blot
  - Immunofluorescence analysis
  - *Drosophila* ovaries
  - *Drosophila* embryos
  - *Drosophila* S2 cells
  - Northern blot of U snRNAs
  - Quantification of RNA modifications using LC-MS/MS
  - Preparation of m^7G-cap enriched RNA fragment for RNA mass spectrometry
  - Preparation of m^7G-capped RNA by *in vitro* transcription
  - *In vitro* methylation assays with PCIF1
  - Phosphorylated CTD-binding assays with recombinant PCIF1
  - Isothermal calorimetry (ITC) experiments
  - Preparation of RNA libraries
  - Ribosome profiling
  - m^6Am-Exo-Seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of RNaseq data
  - Analysis of ribosome footprinting data
  - Analysis of m^6Am-Exo-Seq data

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108038.

ACKNOWLEDGMENTS

We thank François Karch for fly embryos and help with staining; Pascal Gos and Fabienne Fleury-Olela for assistance with experiments; Johanna Mattay and Michaela Dohnalkova for critical reading of the manuscript; and Nicolas Roggli for scientific illustration. We thank Andrew McCarthy, European Molecular Biology Laboratory (EMBL) Grenoble, France, for facilitating the ITC experiments. We thank the following University of Geneva core facilities: Institute of Genetics and Genomics of Geneva (iGE3) Genomics Platform and the Transgenic Mouse Facility. We also thank the EMBL Genomics core facility for deep sequencing. A.R. receives a Swiss Government Excellence Postdoctoral Scholarship, L.L. is supported by the iGE3 PhD Fellowship, and G.F. receives a PhD Fellowship from the Initiative d'excellence (IDEX) program of the University of Grenoble Alpes funded by the French l'Agence Nationale de la Recherche (ANR-15-IDEX-02). This work was supported by grants to R.S.P. from the Swiss National Science Foundation: European Research Council (ERC) transfer grant (CRETP3_166923), project grant (310030B_185386), and Sinergia grant (CRSII5_183524), and by funding from the Swiss National Center of Competence in Research (NCCR) RNA & Disease (51NF40_182880). Work in the Pillai lab is supported by the Republic and Canton of Geneva.

AUTHOR CONTRIBUTIONS

R.R.P. generated the Pcif1 mouse lines and prepared deep sequencing libraries and ribosome profiling experiments; E.D. carried out the biochemical analyses of Pcif1; D.H. performed computational analyses; K.-M.C. performed ITC experiments; A.R. and G.F. carried out immunofluorescence analyses and Northern analysis of snRNAs; L.L. produced recombinant fly Pcif1; G.F., E.T., and M.-O.F. carried out analysis of the fly *Pcif1* mutant not included in this study; C.B.V. analyzed RNA modifications; and manuscript preparation and writing were performed by R.S.P. and R.R.P. with input from everyone.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 23, 2020
Revised: July 10, 2020
Accepted: July 23, 2020
Published: August 18, 2020

REFERENCES

Akichika, S., Hirano, S., Shichino, Y., Suzuki, T., Nishimasu, H., Ishitani, R., Sugita, A., Hirose, Y., Iwasaki, S., Nureki, O., and Suzuki, T. (2019). Cap-specific terminal N^6-methylation of RNA by an RNA polymerase II-associated methyltransferase. Science *363*, eaav0080.

Bailey, A.S., Batista, P.J., Gold, R.S., Chen, Y.G., de Rooij, D.G., Chang, H.Y., and Fuller, M.T. (2017). The conserved RNA helicase YTHDC2 regulates the transition from proliferation to differentiation in the germline. eLife *6*, e26116.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707–719.

Bélanger, 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.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. *37*, 49–57.

Blighe, K. (2019). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling (Bioconductor). https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html.

Boulias, K., Toczydlowska-Socha, D., Hawley, B.R., Liberman, N., Takashima, K., Zaccara, S., Guez, T., Vasseur, J.J., Debart, F., Aravind, L., et al. (2019). Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. Mol. Cell *75*, 631–643.e638.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Byszewska, M., Śmietański, M., Purta, E., and Bujnicki, J.M. (2014). RNA methyltransferases involved in 5′ cap biosynthesis. RNA Biol. *11*, 1597–1607.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Coleman, T.M., Wang, G., and Huang, F. (2004). Superior 5′ homogeneity of RNA from ATP-initiated transcription under the T7 ϕ 2.5 promoter. Nucleic Acids Res. *32*, e14.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

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.

Doxtader, K.A., Wang, P., Scarborough, A.M., Seo, D., Conrad, N.K., and Nam, Y. (2018). Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. Mol. Cell *71*, 1001–1011.e1004.

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.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^6A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Furuichi, Y., Morgan, M., Muthukrishnan, S., and Shatkin, A.J. (1975a). Reovirus messenger RNA contains a methylated, blocked 5′-terminal structure: m-7G(5′)ppp(5′)G-MpCp-. Proc. Natl. Acad. Sci. USA *72*, 362–366.

Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-Georgieff, M., and Darnell, J.E. (1975b). Methylated, blocked 5 termini in HeLa cell mRNA. Proc. Natl. Acad. Sci. USA *72*, 1904–1908.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science *347*, 1002–1006.

Haline-Vaz, T., Silva, T.C., and Zanchin, N.I. (2008). The human interferon-regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity. Biochem. Biophys. Res. Commun. *372*, 719–724.

Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., and Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res. *33*, 2374–2383.

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.

Hsu, P.J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., et al. (2017). Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res. *27*, 1115–1127.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Inesta-Vaquera, F., and Cowling, V.H. (2017). Regulation and function of CMTR1-dependent mRNA cap methylation. Wiley Interdiscip. Rev. RNA *8*, e1450.

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P.N., Monahan, J., Carrieri, C., Enright, A.J., and O'Carroll, D. (2017). The RNA m(6)A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. Mol. Cell *67*, 1059–1067.e1054.

Jain, D., Puno, M.R., Meydan, C., Lailler, N., Mason, C.E., Lima, C.D., Anderson, K.V., and Keeney, S. (2018). *ketu* mutant mice uncover an essential meiotic function for the ancient RNA helicase YTHDC2. eLife *7*, e30919.

Jiao, X., Doamekpor, S.K., Bird, J.G., Nickels, B.E., Tong, L., Hart, R.P., and Kiledjian, M. (2017). 5′ End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. Cell *168*, 1015–1027.e1010.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet. *14*, e1007412.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbø, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., Jr., and Darnell, R.B. (2017). m^6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. *31*, 990–1006.

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.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Lasda, E.L., Kuersten, S., and Blumenthal, T. (2011). SL trans-splicing in a Caenorhabditis elegans in vitro extract. Cold Spring Harb Protoc. *2011*, pdb.prot5574.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Mair, G., Ullu, E., and Tschudi, C. (2000). Cotranscriptional cap 4 formation on the Trypanosoma brucei spliced leader RNA. J. Biol. Chem. *275*, 28994–28999.

Mao, X., Schwer, B., and Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. Mol. Cell. Biol. *15*, 4167–4174.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.J., Chen, Q., et al. (2017). Reversible methylation of m^6A_m in the 5′ cap controls mRNA stability. Nature *541*, 371–375.

Mauer, J., Sindelar, M., Despic, V., Guez, T., Hawley, B.R., Vasseur, J.J., Rentmeister, A., Gross, S.S., Pellizzoni, L., Debart, F., et al. (2019). FTO controls reversible m^6Am RNA methylation during snRNA biogenesis. Nat. Chem. Biol. *15*, 340–347.

Mendel, M., Chen, K.M., Homolka, D., Gos, P., Pandey, R.R., McCarthy, A.A., and Pillai, R.S. (2018). Methylation of Structured RNA by the m(6)A Writer METTL16 Is Essential for Mouse Embryonic Development. Mol. Cell *71*, 986–1000.e1011.

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.

Muthukrishnan, S., Filipowicz, W., Sierra, J.M., Both, G.W., Shatkin, A.J., and Ochoa, S. (1975). mRNA methylation and protein synthesis in extracts from embryos of brine shrimp, Artemia salina. J. Biol. Chem. *250*, 9336–9341.

Oakberg, E.F. (1956). Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. Am. J. Anat. *99*, 507–516.

Patil, D.P., Pickering, B.F., and Jaffrey, S.R. (2018). Reading m^6A in the Transcriptome: m^6A-Binding Proteins. Trends Cell Biol. *28*, 113–127.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Pendleton, K.E., Chen, B., Liu, K., Hunter, O.V., Xie, Y., Tu, B.P., and Conrad, N.K. (2017). The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell *169*, 824–835.e814.

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.

Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., and Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA *17*, 792–798.

Port, F., Chen, H.M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc. Natl. Acad. Sci. USA *111*, E2967–E2976.

R Core Team (2017). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Roignant, J.Y., and Soller, M. (2017). m^{6}A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. Trends Genet. *33*, 380–390.

Schibler, U., Kelley, D.E., and Perry, R.P. (1977). Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. J. Mol. Biol. *115*, 695–714.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409–1421.

Sendinc, E., Valle-Garcia, D., Dhall, A., Chen, H., Henriques, T., Navarrete-Perea, J., Sheng, W., Gygi, S.P., Adelman, K., and Shi, Y. (2019). PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. Mol. Cell *75*, 620–630.e629.

Shafer, B., Chu, C., and Shatkin, A.J. (2005). Human mRNA cap methyltransferase: alternative nuclear localization signal motifs ensure nuclear localization required for viability. Mol. Cell. Biol. *25*, 2644–2649.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645–653.

Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. Nat. Struct. Biol. *7*, 838–842.

Śledź, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. eLife *5*, e18434.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. *3*, 2179–2190.

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.

Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat. Struct. Biol. *7*, 639–643.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, P., Doxtader, K.A., and Nam, Y. (2016). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. Mol. Cell *63*, 306–317.

Warda, A.S., Kretschmer, J., Hackert, P., Lenz, C., Urlaub, H., Höbartner, C., Sloan, K.E., and Bohnsack, M.T. (2017). Human METTL16 is a N^6-methyladenosine (m^{6}A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Rep. *18*, 2004–2014.

Wei, C., Gershowitz, A., and Moss, B. (1975a). 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. (1975b). Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. Cell *4*, 379–386.

Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P.C., Shi, H., Cui, X., Su, R., Klungland, A., et al. (2018). Differential m(6)A, m(6)Am, and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. Mol. Cell *71*, 973–985.e975.

Werner, M., Purta, E., Kaminska, K.H., Cymerman, I.A., Campbell, D.A., Mittra, B., Zamudio, J.R., Sturm, N.R., Jaworski, J., and Bujnicki, J.M. (2011). 2′-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res. *39*, 4756–4768.

Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R., and Pillai, R.S. (2017). Regulation of m6A Transcripts by the 3→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. Mol. Cell *68*, 374–387.e312.

Wu, H., Li, L., Chen, K.M., Homolka, D., Gos, P., Fleury-Olela, F., McCarthy, A.A., and Pillai, R.S. (2019). Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs. Cell. Rep. *29*, 4422–4434.e4413.

Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K., and Sugano, S. (2006). DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. Nucleic Acids Res. *34*, D86–D89.

Yano, Y., Saito, R., Yoshida, N., Yoshiki, A., Wynshaw-Boris, A., Tomita, M., and Hirotsune, S. (2004). A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene. J. Mol. Med. (Berl.) *82*, 414–422.

Yunokuchi, I., Fan, H., Iwamoto, Y., Araki, C., Yuda, M., Umemura, H., Harada, F., Ohkuma, Y., and Hirose, Y. (2009). Prolyl isomerase Pin1 shares functional similarity with phosphorylated CTD interacting factor PCIF1 in vertebrate cells. Genes Cells *14*, 1105–1118.

Zamudio, J.R., Mittra, B., Zeiner, G.M., Feder, M., Bujnicki, J.M., Sturm, N.R., and Campbell, D.A. (2006). Complete cap 4 formation is not required for viability in Trypanosoma brucei. Eukaryot. Cell *5*, 905–915.

Zhang, Z., and Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev. *14*, 328–335.

Zheng, D., Zhang, Z., Harrison, P.M., Karro, J., Carriero, N., and Gerstein, M. (2005). Integrated pseudogene annotation for human chromosome 22: evidence for transcription. J. Mol. Biol. *349*, 27–45.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
For detecting mouse proteins		
Rabbit anti-PCIF1 (for mouse and human)	Bethyl Laboratories	Cat. No. A304-711A; RRID: AB_2620906
rabbit anti-PCIF (for mouse)	Sigma-Aldrich	Cat. No. HPA049517; RRID: AB_2680801
mouse anti-PLZF	Santa Cruz	Cat. No. SC-28319; RRID: AB_2218941
Rabbit anti-PARK7	ThermoFischer Scientitic	Cat. No. PA5-13404; RRID: AB_2160112
Rabbit anti-TUBULIN	Cell Signaling Technology	Cat. No. 2148; RRID: AB_2288042
Anti-m^{6}A antibody	Synaptic Systems	Cat. no. 202003; RRID: AB_2279214
For detecting fly proteins		
Rabbit anti-dPcif1 (CG11399)	This study	N/A
Rabbit anti-Actin	Santa Cruz Biotech	Cat. No. 1616-R
mouse anti-Fibrillarin	AbCam	Cat. No. ab4566; RRID: AB_304523
mouse anti-FLAG	Sigma	Cat. No. F3165; RRID: AB_259529
anti-rabbit IgG HRP-linked	GE Healthcare	Cat. No. NA934; RRID: AB_772206
anti-mouse IgG HRP-linked	GE Healthcare	Cat. No. NA931; RRID: AB_772210
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat. No. A11034; RRID: AB_2576217
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat. No. A11037; RRID: AB_2534095
Goat anti-Mouse IgG (H+L), Superclonal Recombinant Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat. No. A28175; RRID: AB_2536161
Goat anti-Mouse IgG (H+L) Cross-Adsorbed ReadyProbes Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat. No. R37121; RRID: AB_2556549
anti-mouse conjugated to Cyanine 3	Jackson ImmunoResearch	Cat. No. 111165144; RRID: AB_2338006
Bacterial and Virus Strains		
DH10EMBacY bacterial strain	Bieniossek et al., 2012	N/A
Top10	This study	
BL21(DE3)	This study	
Chemicals, Peptides, and Recombinant Proteins		
Sodium deoxycholate	Sigma	Cat. No. 30968
Complete EDTA-free protease inhibitor	Roche	Cat. No. 11 873 580 001
m7G(5′)ppp(5′)G RNA Cap Structure Analog	NEB	Cat. No. S1404S
m7G(5′)ppp(5′)A RNA Cap Structure Analog	NEB	Cat. No. S1405S
m7G(5′)ppp(5′)AmpG RNA Cap Structure Analog	Trilink	Cat. No. N-7413
Ponceau S	Sigma	P3504
Schneider's *Drosophila* Medium	GIBCO	Cat. No. 21720001
fetal calf serum	BioConcept	Cat. No. 2-01F30-I
30% acrylamide (37.5:1)	National Diagnostic	Cat. No. EC-890
N,N,N',N'-Tetramethylethylendiamin	Merck	Cat. No. 1107320100
Tween20	SIGMA	Cat. No. P7949
Amersham Prime Western Blotting Detection Reagent	GE Healthcare	Cat. No. RPN2232
SuperSignal West Femto Maximum Sensitivity Substrate	ThermoFisher	Cat. No. 34095
Pierce ECL 2 Substrate	ThermoFisher	Cat. No. 1896433A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays/Kits		
MinElute Gel Extraction Kit	QIAGEN	Cat. No. 28604
MEGAshortscript T7 Transcription Kit	Life technologies	Cat. No. AM1354
Pierce Detergent Compatible Bradford Assay Kit	ThermoFisher	Cat. No. 23246
SuperScript III Reverse Transcriptase kit	ThermoFisher	Cat. No. 18080085
The TruSeq Stranded Total RNA kit	Illumina	Cat. No. RS-122-2301
Deposited Data		
Deep sequencing datasets	This study	GEO: GSE151229
All raw gel data are deposited at Mendeley Data.	This study	https://doi.org/10.17632/jvfp8z5n8x.1
Experimental Models: Cell Lines		
Sf9 insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
High Five insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
Experimental Models: Organisms/Strains		
Mouse: *Pcif1* knockout	This study	Available from Lead Contact
Mouse: *Pcif1-3xFLAG* knockin	This study	Available from Lead Contact
Oligonucleotides		
DNA and RNA oligos		See Table S1
Recombinant DNA		
pACEBac2	Bieniossek et al., 2012	N/A
Human *Pcif1* cDNA	This study	GenBank: BAC45238
Drosophila Pcif1 cDNA	This study	Fly Base: CG11399-PB
Trypanosoma Pcif1	This study	GenBank: RNF19590.1
pIDK	Bieniossek et al., 2012	N/A
pU6-BbsI-chiRNA		
pAC5.1-EGFP		
pCFD4 vector	Port et al., 2014	
Software and Algorithms		
ENRICHR	Chen et al., 2013; Kuleshov et al., 2016	http://amp.pharm.mssm.edu/Enrichr/
R	R Core Team, 2017	https://www.r-project.org
DESeq2	Love et al., 2014	N/A
Bioconductor	Huber et al., 2015	https://www.bioconductor.org/
Salmon	Patro et al., 2017	N/A
tximport	Soneson et al., 2015	N/A
EnhancedVolcano	Blighe, 2019	https://github.com/kevinblighe/EnhancedVolcano
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
targetFinder		http://targetfinder.flycrispr.neuro.brown.edu/
Other		
Chelating Sepharose Fast Flow beads	GE Healthcare	Cat. No. 17-0575-01
StrepTrap HP	GE Healthcare	Cat. No. 28-9075-46
Superdex S75 10/300 GL	GE Healthcare	Cat. No. 17-5174-01
Superdex 200 10/300 GL	GE Healthcare	Cat. No. 17-5175-01
MethaPhor agarose	Lonza	Cat. No. 50180
Amersham Hyperfilm ECL	GE Healthcare	Cat. No. 28906837
Amersham Protran 0.45 μm nitrocellulose membrane	GE Healthcare	Cat. No. 10600002

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Ramesh S. Pillai (ramesh.pillai@unige.ch).

Materials Availability

All unique reagents (plasmids, antibodies, and animal mutants) generated in this study are available from the Lead Contact without any restriction. This includes the *Pcif1* mouse knockout and *Pcif1-3xFLAG* knockin lines.

Data and Code Availability

Deep sequencing data generated in this study are deposited with Gene Expression Omnibus (GEO) under accession number GEO: GSE151229. Code used in the current study is available from the Lead Contact upon reasonable request. Other raw data associated with this study are deposited with Mendeley Data (https://doi.org/10.17632/jvfp8z5n8x.1).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal Work

Mutant mice were generated at the Transgenic Mouse Facility of University of Geneva. The mice were bred in the Animal Facility of Sciences III, University of Geneva. The use of animals in research at the University of Geneva is regulated by the Animal Welfare Federal Law (LPA 2005), the Animal Welfare Ordinance (OPAn 2008) and the Animal Experimentation Ordinance (OEXA 2010). The Swiss legislation respects the Directive 2010/63/EU of the European Union. Any project involving animals has to be approved by the Direction Générale de la Santé and the official ethics committee of the Canton of Geneva, performing a harm-benefit analysis of the project. Animals are treated with respect based on the 3Rs principle in the animal care facility of the University of Geneva. We use the lowest number of animals needed to conduct our specific research project. Discomfort, distress, pain, and injury is limited to what is indispensable and anesthesia and analgesia is provided when necessary. Daily care and maintenance are ensured by fully trained and certified staff. This work was approved by the Canton of Geneva (GE/16/19).

Pcif1 knockout and Flag-Pcif1 knockin mice

This study concerns the *Pcif1* gene (MGI: 2443858; NCBI Gene: 228866) encoding for Phosphorylated CTD-interacting factor 1 (PCIF1) (Fan et al., 2003; Hirose et al., 2008), which was later shown to be a cap-dependent RNA methyltransferase (and termed as CAPAM) (Akichika et al., 2019). The *Pcif1* gene studied here is located on mouse chromosome 2 (Chr2: 164879368-164894454 bp, + strand) and consists of 17 exons (Figure S1A). It encodes for a protein with 723 aa (NCBI: NP_001361055).

We targeted the *Pcif1* locus in mouse embryos of the B6D2F1/J hybrid line (also called B6D2; The Jackson Laboratory, stock no. 100006). It is a cross between C57BL/6J (B6) and DBA/2J (D2), and heterozygous for all B6 and D2 alleles. Single-cell mouse embryos were injected with two different guide RNAs (gRNAs) that direct the DNA endonuclease Cas9 to delete a region between exon 15 and intron 16 (Figure S1A). Founder mice were identified by genotyping PCR (Figure S1B) and crossed with wild-type C57BL/6J (Janvier) partners to obtain germline transmission. We identified a mouse line with 89 bp deletion in the exon 15/intron 16 that removes the critical catalytic residues (NPPF) within the methyltransferase domain and also creates a premature termination codon. Indeed, Western analysis confirms the homozygous *Pcif1* mutant mice lack the full-length PCIF1 protein (Figures 1C and S1E). The antibody used for the western blot detects the C terminus of the PCIF1 protein. Heterozygous $Pcif1^{+/-}$ and homozygous $Pcif1^{-/-}$ mice of both sexes were viable and fertile. We did not find any obvious phenotypic alterations in the homozygous $Pcif1^{-/-}$ mice under the tested normal conditions. However, $Pcif1^{-/-}$ mice display reduced body-weight compared to heterozygous littermates (Figure 1G).

To create an allele expressing C-terminal FLAG-tagged PCIF1, we inserted a sequence corresponding to a four-amino acid flexible linker (SGGG) followed by the 3xFLAG tag (MDYKDHDGDYKDHDIDYKDDD) upstream of the translation stop codon (TAA) of the *Pcif1* open-reading frame (ORF). To this end, we used a single guide RNA to target the Cas9 endonuclease to the *Pcif1* gene locus in mouse embryos of the B6D2F1/J hybrid line (also called B6D2; The Jackson Laboratory, stock no. 100006) (Figures S1A and S1B). We used an ssDNA oligonucleotide with the required 3xFLAG sequence as the repair template (Table S1). Heterozygous and homozygous *Pcif1-3xFLAG* knockin mice were viable and fertile, without any impact on the PCIF1 protein. Based on the western blot analysis, PCIF1-3xFLAG expression (Figure 1A) was identical to that detected for endogenous PCIF1 (Figure S1C) in multiple mouse tissues.

Synthesis and annealing of gRNA

The gRNAs were commercially synthesized [Integrated DNA Technologies (IDT)] in two parts (Table S1). A tracrRNA (67 bases long; IDT, Cat. No. 1072533) which is the common part of the gRNA that binds to the Cas9 protein, and crRNA (IDT) that has a

complementarity of 20 bases with the genomic DNA sequence of interest plus common additional bases for pairing with tracrRNA at the 3′ end (GUUUUAGAGCUAUGCU). The complete gRNA was obtained by annealing 200 pmol each of tracrRNA and crRNA in 1xTE buffer (IDT 1X TE buffer, pH 7.5; Cat. No. 11-05-01-05) to a final volume of 10 uL. The gRNAs mix was heated at 95°C for 5 min and then allowed to cool at room temperature for 30 min to get annealed functional gRNA. The annealed gRNA was stored at −20°C until used for injection into mouse embryos.

Preparation of injection mix

Just before injection into mouse embryos, the annealed gRNAs (final concentration 0.6 pmol/μL) were mixed with the Cas9 protein at 30 ng/μL final concentration (Cat. No. 1081058, IDT) in a final volume of 9 μL. The mix was incubated at RT for 10 min for complex formation. For knockin mouse generation a ssDNA repair template (IDT; 20 ng/μL final concentration) was also added to the injection mix and volume was adjusted with 1x TE buffer (pH 7.5) to 100 μL. The injection mix was centrifuged at 13000 rpm for 5 min at 4°C, and 50 μL of supernatant was collected and stored on ice.

Injection of mouse embryos of the hybrid background B6D2F1/J (black coat color) was carried out at the Transgenic Mouse Core Facility, University Medical Centre (CMU), University of Geneva. The NMRI (Naval Medical Research Institute) mice, which have a white coat color were used as foster mothers.

crRNA: /Alt1/NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCU /Alt2/

N represent the gene-specific sequence. The sequence of crRNAs used are provided (Table S1).

The sequence of the ssDNA repair template for 3xFLAG sequence insertion is as follows:

CAAGGACCGGGACTCAGGCCGGGAACAGGGCCCTAGTAGAGAGCCTCACCCCACT*TCAGGCGGTGGA***ATG**GACTACAAAGA
CCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGAT**TAA**CATATCCTGCGGGGAGGAGGAGCTCCAGGGGTG
CTGTCACAGACTGCTAGGACTCAGCCTC**TAA** is the natural STOP sequence in the mouse PCIF1 coding sequence. The sequence (underlined) is the coding region for 3xFLAG sequence. The sequence in italics is the linker between the PCIF1 protein and the C-terminal 3xFLAG tag.

Genotyping

Ear-punches of the weaned animals (at 21 days after birth) were digested in 100 μL of buffer containing 10 mM NaOH, 0.1 mM EDTA for 120 min at 95°C. The digested ear-punch was centrifuged at 3000 rpm for 10 min and 50 μL of the supernatant was transferred to a new tube containing 50 μL of TE buffer (20 mM Tris-HCl, pH 8.0 and 0.1 mM EDTA). Primers (RRoligo1129 and RRoligo1130) flanking the targeted deletion region in the mouse *Pcif1* gene were used to genotype the knockout allele (Table S1). The expected sizes of PCR products are: $Pcif1^{+/+}$ (362 bp, WT) and $Pcif1^{-/-}$ (273 bp, knockout). For genotyping *Pcif1-3xFlag* knockin allele the primers RRoligo1132 and RRoligo1133 were used (Table S1). The expected sizes of PCR products are: 399 bp for the wild-type allele and 474 bp for the *Pcif1-3xFlag* allele. Identity of the bands was confirmed by Sanger sequencing.

Reaction mix for 20 μL PCR reactions: 10 μL of Phire Green Hot Start II PCR Master Mix (Thermo Scientific, Cat. No. F126L), 1.0 μL primer mix (stock 10 nM each), 2.0 μL ear-punch DNA (50-200 ng), and 7.0 μL water to make 20 μL final volume. Reactions were run using the following conditions: 98°C for 30 s, 35 cycles of [98°C for 5 s, 63°C for 5 s and 72°C for 10 s], 72°C for 1 min, and finally at 4°C to hold the reaction. PCR products were directly analyzed by electrophoresis on 2% agarose gel in 1xTBE buffer (Figure S1B). Primer sequences are given in Table S1.

METHOD DETAILS

Clones and constructs

The mouse PCIF1 protein depicted in the cartoons is based on the sequence from NCBI (NCBI: NP_001361055). Coding sequences for full-length (FL) human PCIF1 (704 aa; GenBank: BAC45238) was amplified from human HeLa cell total RNA by reverse transcription-PCR (RT-PCR). The *Drosophila melanogaster* Pcif1 (920 aa; Fly Base: CG11399-PB) was amplified from fly ovary total RNA by reverse transcription-PCR (RT-PCR). Coding sequence for full-length *Trypanosoma cruzi* Pcif1 (TrypPcif1; 532 aa; GenBank: RNF19590.1) was compared to protein sequences in the Kinetoplastid Genome Resource database TriTRypDB. Unlike the GenBank sequence (which has Proline at position 527), eleven strains of *T. cruzi* showed a serine at position 527. So for our study, we used the TrypPcif1 sequence from (GenBank: RNF19590.1), but carrying the P527S change in the coding sequence. The sequence was synthesized (ShineGene Bio-Technologies, Inc, Shanghai) after codon-optimization for insect cell expression.

Constructs for *Drosophila* S2 cell expression

For expression in *Drosophila* S2 cells, the required coding sequences were cloned into the expression vector pAC5.1-3xFLAG-HA vector at the XhoI-NotI sites. The pAC5.1-3xFLAG vector was prepared in-house by insertion of a multiple cloning site (MCS) cassette followed by a 3xFLAG tag at the site KpnI-XbaI of the pAC5.1vector, to allow expression of C-terminal 3xFLAG-HA fusions.

Sequence of the C-terminal 3xFLAG-HA tag is as follows:

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAG*GGCGGCAGCGGC*TA
CCCATATGATGTTCCAGATTACGCTGGT**TGA**

The linker between the 3xFLAG and the HA sequence is in italics. **TGA** is the stop codon.

Constructs for recombinant protein production in prokaryotic expression systems

For the production of full-length human PCIF1 protein in *E.coli*, the coding sequence (UniProtKB: Q9H4Z3) was cloned into pETM30 vector to express the N-terminal 6xHis-GST-PCIF1 fusion protein. For expression of isolated WW domains, sequences coding for human PCIF1 WW domain (42-80 aa of NCBI ID: NP_071387.1) and fly WW domain (138-172 aa of FlyBase ID: FBgn0037021) were cloned in-frame downstream of 6xHis-GST tag in the pETM-30 vector (EMBL Protein Expression Facility, Heidelberg), in order to express the 6xHis-GST-WW fusion protein.

Constructs for recombinant protein production in insect expression systems

For insect cell expression of fly Pcif1 (dPcif1), the coding sequence (UniProtKB: Q9VPB7) was cloned into pACEBac2S vector for the expression as N-terminal His-Strep-SUMO fusion protein. The constructs used for recombinant protein production were verified by restriction digest, as well as by Sanger sequencing. The coding sequence for *Trypanosoma cruzi* Pcif1 (GenBank: RNF19590.1) was codon-optimized for insect cell expression and synthesized (ShineGene Bio-Technologies, Inc, Shanghai) with a P527S mutation (see note above), and was cloned into pACEBac2S vector for protein expression in insect cell.

Recombinant protein production

Production of full-length fly and *Trypanosoma* Pcif1 was carried out in insect cell lines using the baculovirus expression system. The insect ovary-derived cell lines used are: High Five (Hi5) insect cell line originating from the cabbage looper (*Trichoplusia ni*) and the *Sf*9 cells derived from the fall army worm *Spodoptera frugiperda*. Briefly, the full-length fly Pcif1 coding sequence was cloned into pACEBac2S vector. Plasmids carrying target gene were transformed into DH10EMBacY competent cells for recombination with the baculovirus genomic DNA (bacmid). The bacmid DNA was extracted and transfected with FuGENE HD (Promega, cat. no. E231A) into the *Sf*9 insect cells for virus production. The supernatant (V_0) containing the recombinant baculovirus was collected after 48 to 72 h post-transfection. To expand the virus pool, 3.0 mL of the V_0 virus stock was added into 25 mL of *Sf*9 (0.5×10^6/mL) cells. The resulting cell culture supernatant (V_1) was collected 24 h post-proliferation arrest. For large-scale expression of the protein, Hi5 cells were infected with virus (V_1) and cells were harvested 72 h after infection.

For human PCIF1 expression pETM30-hPCIF1 plasmids were transformed into the *E. coli* BL21(DE3) strain and grown in LB media. Protein expression was induced by addition of 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when the culture density reached around 0.8 (OD_{600}). The proteins were then expressed overnight at 18°C following induction. Similarly, the expression of GST-fusions of WW domains (from human and fly PCIF1) was also performed in the prokaryotic system.

Purification of human, fly and Trypanosoma PCIF1

The cells (bacteria or insect cells) were collected by centrifugation and lysed by sonication in lysis buffer: 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 40 mM Imidazole and EDTA- free protease inhibitor (Thermo Scientific, Cat. No. PI88666). The lysate was centrifuged at 16000 rpm for 40 min at 4°C. Supernatant was collected in a new tube and incubated with with Ni^{2+} chelating Sepharose FF beads for 2 h in cold room. After 2 h of incubation, the beads were washed by imidazole gradient washing buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 40 mM Imidazole) and finally bound protein was eluted with 250 mM imidazole in lysis buffer. Subsequently, the N-terminal tag was cleaved by the TEV protease at 4°C overnight in the dialysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol). The cleaved tag was removed by a second purification on Nickel beads. The protein was further purified by gel-filtration chromatography with Superdex 200 columns (GE Healthcare, Cat. No. 28-9909-44), in the buffer (25 mM Tris-HCl pH 8.0, 150 mM KCl, 5% glycerol and 1 mM DTT). The pure fractions were verified by SDS-PAGE electrophoresis (Figures S5A and S5F) and used for biochemical assays.

Purification of GST-fused human and fly PCIF1 WW domain

The bacterial cells [*E. coli* BL21(DE3)] expressing GST-fused WW domain from human and fly PCIF1 were collected by centrifugation and lysed by sonication in buffer: 50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 0.1% Triton X-100, 5 mM $MgCl_2$, 5μM $ZnCl_2$ and EDTA-free protease inhibitor. Following the pre-wash (three times) of GST conjugated beads (Mecherey-Nagel, Cat. No. 745500.10) with lysis buffer, the lysate is incubated for 3 h at 4°C with shaking. Then the beads were washed five times: alternating with wash buffer (50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol) and high salt wash buffer (50 mM Tris-HCl, pH 8.0, 1000 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol, 0.1% Triton X-100). Finally, the bound protein was eluted with elution buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol, 15 mM Glutathione). The protein was further purified by gel-filtration chromatography with Superdex S75 columns (GE Healthcare, Cat. No. 29148721), in the buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol, 0.1% Triton X-100, 2mM MgCl2). The pure fractions were verified by SDS-PAGE electrophoresis (Figure S5D) and used for biochemical interaction assays.

ANTIBODIES

Commercial antibodies

Primary antibodies

rabbit anti-PCIF1 (Bethyl Laboratories, Cat. No. A304-711A) recognizing a C-terminal epitope (654-704 aa) was used for Western analysis; rabbit anti-PCIF (Sigma-Aldrich, Cat. No. HPA049517) recognizing an N-terminal epitope (98-188 aa) were used for immunofluorescence; mouse anti-PLZF (Santa Cruz, D-9, Cat. No. SC-28319); mouse anti-FLAG (Sigma, Cat. No. F3165), rabbit anti-PARK7 (ThermoFischer Scientific, Cat. No. PA5-13404) for detecting mouse protein, rabbit anti-TUBULIN (Cell Signaling Technology, Cat. No. 2148) for detecting mouse protein, and rabbit anti-Actin (Santa Cruz Biotech, Cat. No. 1616-R) to detect fly actin protein, were used for Western. For immunofluorescence analysis with *Drosophila* ovaries, mouse anti-Fibrillarin (AbCam, Cat. No. ab4566) were used. Anti-m^{6}A antibody (Synaptic Systems, Cat. no. 202003) were used for m^{6}A-IPseq experiments.

Secondary antibodies

For western blot analyses, the following secondary antibodies conjugated to Horse Radish Peroxidase were used: anti-rabbit IgG HRP-linked (GE Healthcare, Cat. No. NA934), anti-mouse IgG HRP-linked (GE Healthcare, Cat. No. NA931); for immunofluorescence studies, the following secondary antibodies coupled to fluorescent dyes were used: anti-rabbit (Alexa Fluor 488, Cat. No. A11034), anti-rabbit (Alexa Fluor 594, Cat. No. A11037), anti-mouse (Alexa Fluor 488, Cat. No. A28175), anti-mouse (Alexa Fluor 594, Cat. No. R37121), anti-mouse conjugated to Cyanine 3 (1:500; Jackson ImmunoResearch, Cat. No. 111165144).

Antibodies generated for this study

We generated rabbit polyclonal antibodies against fly Pcif1 protein. Two New Zealand White (NZW) rabbits were immunized with soluble antigens (Biotem, France). The antigen used was the purified untagged full-length fly Pcif1 (dPcif1) protein produced in insect cells. For each injection, 1 mg/mL protein was used. After six injections (at day 0, 14, 28, 56, 70 and 89) crude immune serum was collected (at day 39, 67, 82 and 96) and frozen. The crude sera and the affinity-purified antibodies detect a single band in fly ovary and *Drosophila* S2 cell lysates as detected by Western analysis (Figure 4B). The antibodies were also suitable for immunofluorescence studies of fly embryos and fly ovaries (Figures 4D and S4D). Antibodies were purified using recombinant full-length fly Pcif1 (dPcif1) protein.

Antibody purification

To purify antibodies against fly Pcif1 (dPcif1), large amount of recombinant untagged full-length fly Pcif1 (dPcif1) protein was resolved via SDS-PAGE and semi-dry transferred onto a nitrocellulose membrane. After reversible staining with Ponceau S (Sigma, Cat. No. P3504), part of the membrane containing the antigen was cut out and incubated overnight with 2 mL crude immune sera (diluted in 8 mL 1XPBS). After washes (1xPBS), bound antibodies were eluted with 1.0 mL (elute twice with 500 μL each) of low pH solution (0.1 M Glycine pH 2.5, 150 mM NaCl, three times) and immediately neutralized with 100 μL of neutralization buffer (0.5 M Tris-HCl pH 8.8, determine volume required by tests just before use). Antibodies were concentrated with 2.0 mL Amicon Ultra spin column (GE Health, 10 kDa cut-off) and stored in 1x PBS, 50% glycerol at −20°C.

Drosophila S2 cell culture and transfections

Drosophila melanogaster Schneider 2 (S2) cells were derived from a primary culture of late-stage (1d) *Drosophila* embryos. Based on molecular features, they likely have a macrophage-like lineage. Cells were cultured in Schneider's *Drosophila* Medium (GIBCO, Cat. No. 21720001) supplemented with 10% fetal calf serum (FCS) (BioConcept, Cat. No. 2-01F30-I) at 25°C.

For transfections, cells were plated into 60 mm culture dishes overnight and transfected with 1 μg of pAC5.1 vector expressing hPCIF1-3xFLAG-HA tagged proteins. Cells in a separate dish were transfected with pAC5.1-EGFP vector as control for transfection. To deliver the DNA into the cells 5 μL FuGENE HD (Promega, Cat. No. E231A) was used for each transfection, cells were collected after 3 days post-transfection.

For RNA mass spectrometry, S2 cells were transfected with plasmids expressing hPCIF1-3xFLAG-HA tagged proteins as described above, cells were collected after 3 days post-transfection. Total RNA from cells was extracted using TRIzol (Life technology, Cat. No. 15596-026), according to manufacturer instructions. RNA pellet was washed with 80% ethanol, briefly dried, and dissolved in 40 μL RNase-free water. To purify the polyA+ fraction from the total RNA, the samples were incubated with oligo dT Dynabeads (Thermo Fisher, Cat. No. 61006) and eluted in 20 μL of 10 mM Tris-HCl following the protocol suggested by the manufacturer. RNA mass spectrometry confirmed the absence of m^{6}Am RNA modification in total RNA from S2 cells (Figure 4C). When S2 cells were transfected with 3Xflag-HA-tagged human PCIF1 (Figure S4A), ectopic methylation of RNA with m^{6}Am was confirmed (Figure 4C).

Mouse multiple tissue western blot

Multiple tissues were isolated from an adult (> 60 d) mouse. After flash-freezing in liquid nitrogen, a piece of different tissues was homogenized in 1 mL lysis buffer [20 mM Tris pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 0.5% sodium deoxycholate, 1 mM DTT, Complete Protease Inhibitor Cocktail Tablet (Roche, Cat. No. 5056489001). The lysate was transferred to a 1.5 mL Eppendorf tube, centrifuged at 14000xg for 30 min, and the supernatant collected. An aliquot was taken to measure the concentration by Pierce

Detergent Compatible Bradford Assay Kit (Thermo Fisher, Cat. No. 23246). Lysate concentrations were normalized to 1 mg/μL. The SDS loading buffer was added and the samples were boiled at 95°C for 10 min. From each tissue sample, 20 mg of protein per lane was loaded and resolved by SDS-PAGE. Western blot analyses (Figures 1A and S1C) with anti-FLAG or anti-PCIF1 rabbit polyclonal antibodies (Bethyl Laboratories, Cat. No. A304-711A; recognizes a region between 654-704 aa) were performed. After stripping, the blot was re-probed with anti-PARK7 rabbit polyclonal antibodies, to serve as loading control.

The expected size of ~90 kDa is seen for mouse PCIF1 in all tissues, while in some tissues we consistently observed a second band at ~60 kDa (Figures 1A and S1C). Since the anti-PCIF1 antibody recognizes C-terminal epitopes, and since the tag in our PCIF1-FLAG knockin mouse is C-terminal, we conclude that the ~60 kDa band is likely to be an N-terminal truncated version.

Western Blot

Whole cell lysates and suspensions of boiled beads were separated on SDS-PAGE in order to detect proteins of interest. SDS-PAGE gels were prepared using Ultra-Pure ProtoGel 30% acrylamide (37.5:1) (National Diagnostic, Cat. No. EC-890), ultra-pure water, and resolving gel buffer (0.375 M Tris, 0.1% SDS, pH 8.8) to obtain 8% resolving gel, and with stacking gel buffer (0.125 M Tris, 0.1% SDS, pH 6.8) to obtain 12% stacking gel. N,N,N',N'-Tetramethylethylendiamin (TEMED; Merck, Cat. No. 1107320100) and 10% ammonium persulfate were added to catalyze the polymerization reaction. Gel electrophoresis was performed at 90 V for 30 min. and then at 120 V for 90 min. After separation, proteins from the gel were transferred to the Amersham Protran 0.45 μm nitrocellulose membrane (GE Healthcare, Cat. No. 10600002) overnight at 5 V at room temperature using Trans-Blot SD. Semi-Dry Transfer Cell system (Bio-Rad, Cat. No. 1703940). After transfer, membranes were washed with Tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.6) and blocked for 1 h at room temperature with 5% fat free milk in TBS with 0.05% Tween20 (TTBS) (Sigma, Cat. No. P7949). The blocked membranes were incubated with primary antibody for 3 h at room temperature in 5% milk with TTBS. Then, membranes were washed 3 times for 10 min with TTBS and incubated with HRP-conjugated secondary antibody at 1:10 000 dilution, either with anti-rabbit IgG HRP-linked (GE Healthcare, Cat. No. NA934) or anti-mouse IgG HRP-linked (Invitrogen, Cat. No. a27025) for 1 h at room temperature in 5% milk in TTBS. The membranes were washed 3 times for 10 min with TTBS and incubated with one of detection reagents: Amersham Prime Western Blotting Detection Reagent (GE Healthcare, Cat. No. RPN2232), SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher, Cat. No. 34095) or Pierce ECL 2 Substrate (ThermoFisher, Cat. No. 1896433A) for 5 min. at room temperature. Signal was detected using Amersham Hyperfilm ECL (GE Healthcare, Cat. No. 28906837). The processed films were scanned using Perfection 3200 Photo scanner (Epson) with XSane image scanning software (ver. 0999).

Immunofluorescence analysis

Mouse tissue sections

To prepare the paraffin sections, the mouse tissues were washed in PBS, and fixed in 4% paraformaldehyde overnight at 4°C. After washing in PBS, testes were dehydrated in 70% ethanol and stored in 70% ethanol at 4°C. Alternatively isolated tissues were fixed in Bouin's solution (Sigma, cat. No HT10132) overnight at room temperature and subsequently washed 3 times in 50% ethanol and in several changes of 70% ethanol, until no yellow dye could be extracted into solution. Samples were sent to Histology core service in University of Geneva, Switzerland where they were further dehydrated in 80%, 90%, 96% and 100% ethanol (90 min for each step), followed by incubation in xylene (3 times 30 min). Xylene was removed and replaced with paraffin and incubated at 56-58°C. Tissues were then transferred into plastic molds (Polysciences mold S-22; NC0397999) filled with melted paraffin, and paraffin was allowed to become solid at room temperature. The tissue sections (~5 μM thickness) were prepared using microtome. The sections were allowed to stretch at 42°C and then stored at room temperature. For histological analysis, the slides containing the paraffin sections were placed in a glass slide holder filled with xylene (3 × 5 min) to remove the paraffin. For rehydration, the slides were incubated in 100% ethanol, 96% ethanol, 70% ethanol, 50% ethanol (2 × 10 min for each step) and miliQ water (2x2 min for each step).

Next, antigen retrieval was performed with Heat-Induced Epitope Retrieval (HIER). Briefly, slides were submerged in 1L of 10 mM Citrate Buffer pH 6.0 (10 mM Sodium citrate, 0.05% TWEEN) and heated in pressure cooker at 1 atmosphere at 120°C for 20 min. Tissues were allowed to cool down at room temperature for at least 1 h, washed twice in PBST (PBS, 0.5% Triton X-100) at RT for 10 min and 2x2 min in miliQ water. Slides were blocked in blocking soultion (5% normal goat serum, PBST) for 1 h at room temperature in a humidified chamber. Primary antibodies were diluted in blocking solution at different concentrations (see below) and incubated overnight at 4°C. Next day, slides were washed in PBST (3 × 5 min), incubated with secondary antibody (anti-mouse or anti-rabbit conjugated to Alexa 488, or 594 fluorophore) in a humidified chamber for 1 h (dilution 1:1000 in blocking solution). Slides were washed twice in PBST and incubated with DAPI (0.5 μg/mL, Bio-Rad, cat. No. 10043282) for 5 min to counterstain the nuclei. Sections were finally washed in PBS (5 min), in miliQ water (5 min) and mounted with ProLong Gold Antifade ReagentDiamond Antifade Mountant (Life technologies, cat. No. S36942P36970). Pictures were taken using Zeiss LSM800, point-scan confocal microscope (Bioimaging Center, University of Geneva). Primary antibodies concentrations: anti-PCIF1 1:100, anti-PLZF 1:100.

Drosophila ovaries

5-10 ovary pairs were dissected from wild-type or heterozygous *dPcif1*$^{F9/+}$ mutant *Drosophila* females in PBS. They were immediately fixed in 4% PFA for 30 min at room temperature (RT). Fixed ovaries were then washed (3 × 5 min) with 1xPBS, permeabilized for 5 min with PBT (1xPBS containing 0.1% Triton X-100) at RT and blocked in blocking buffer (1xPBS, 0.1% Triton X-100, 5% horse

serum) for 1 h at RT. They were incubated with rabbit anti-dPcif1 (1:100) and mouse anti-fibrillarin (1:100; abcam ab4566) antibodies in blocking buffer overnight at 4°C. Next day, they were washed (3 × 15 min) in PBT at RT and incubated for 4 h with anti-rabbit conjugated to Alexa 488 (1:500; Invitrogen, Cat. No. A11029), anti-mouse conjugated to Cyanine 3 (1:500; Jackson ImmunoResearch, Cat. No. 111165144) antibodies and Hoechst, 1:1000 (ThermoFisher, Cat. No. 33342) in blocking buffer at RT. Ovaries were then washed in PBT (3 × 10 min) at RT and stored in PBS before mounting in Dako (Cat. No. S3022). Images were captured with Nikon TI-E Eclipse confocal spinning-disk inverted microscope (CEA, Grenoble).

Drosophila embryos

Thousands of fly embryo (2 to 15 h) were collected in 0.4 mL 1xPBS. The embryos were dechorionated in 50% chlorex for 2 min followed by washing with tap water several times. Embryos were fixed in 2.0 mL of fixing solution (4% formaldehyde, 0.1M Pipes pH 6.9, 2 mM $MgSO_4$, 1mM EGTA). Equal volume of Heptane (2.0 mL) was added to the embryos, mixed by shaking and kept on rotating wheel at room temperature for 15-20 min. After fixing, aqueous phase (lower phase) was removed carefully and equal volume of methanol (equal volume to that of heptane) was added and embryo were shaken vigorously to devitilinize them. Once the embryos were settled in the bottom, heptane, and methanol was removed. Embryos were washed twice with a small volume of methanol (2.0 mL), and once with PBT (1xPBS, 0.1% Triton -X100, and 0.1% BSA). Care should be taken to minimize the exposure of embryos to methanol. Embryos were incubated in PBT buffer for 2 h at RT, followed by incubation with primary antibody (rabbit anti-dPcif1, 1:1000 dilution) in PBT overnight at 4°C on a wheel. Next day, embryos were washed 3x in PBT at RT and finally incubated with secondary anti-rabbit antibody coupled with Alexa Fluor 594 for 2 h at RT in dark. Embryos were washed 2x with PBT followed by incubation with 1:1000 diluted DAPI in PBT for 10 min at RT. After two more washes in PBT, embryos were mounted on glass slides in ProLong Gold Antifade Reagent mounting media (Thermo Scientific, Cat. No. P36931). Images were captured on Zeiss LSM 780 confocal microscope (Bioimaging Center, University of Geneva).

Drosophila S2 cells

Glass coverslips (Sigma, Cat. No. PEZGS0416) were previously coated with Concavalin A (0.5 mg/mL; Sigma, Cat. No. L7647), in order to allow the cells to attach on it. S2 cells were grown in dishes containing such coverslips and the cells were transfected with a plasmid expressing dPcif1-3xFLAG-HA, as previously described. After two days of transfection, the immunostaining protocol was started.

The cells were washed in 1xPBS, and then fixed in 4% paraformaldehyde at room-temperature (RT) for 15 min. The cells were washed 3 times and incubated for 5 min with 0.1M Glycine (MP Biomedicals, Cat. No. 808822). After three more washes, the cells were permeabilized with 0.1% Triton X-100 for 5 min. After the cells are incubated for 15 min in blocking buffer (0.2% BSA, 1% goat serum, 0.1% Triton X-100 in PBS). The cells were incubated for 90 min with the primary antibody (monoclonal mouse anti-FLAG; Sigma, Cat. No. F3165), diluted 1:1000 in blocking buffer. After 3 washes in 1xPBS, the cells were incubated for 60 min. in the dark with the secondary antibody conjugated to Alexa Fluor 488, diluted 1:250 in blocking buffer. Afterward, cell nuclei were stained for 30 min with DAPI 0.5 μg/mL (Bio-Rad, Cat. No. 10043282) diluted in 1xPBS. Finally, the cells were washed 3 times in 1xPBS and once in double-distilled H2O to remove the salts. The coverslip was mounted with Slowfade Gold Antifade Reagent (Life technologies, Cat. No. S36942). Images (Figure S4C) were taken using Zeiss LSM710 confocal microscope (Bioimaging Center, University of Geneva).

Northern blot of U snRNAs

Total RNA was extracted from mouse testes using TRIzol RNA extraction kit (Life technology, Cat. No. 15596-026), further purified with double phenol-chloroform treatment, precipitated in ethanol for at least 20 min at −20 °C and resuspended in milli-Q water. For Northern blotting, 2 μg of total RNA was resolved by 8% urea-PAGE. The RNA was transferred by capillarity to a Nylon membrane (Hybond N+, Amersham) for at least 16 h in in 20 × SSC solution (3 M NaCl, 300 mM sodium citrate). After the transfer, the RNA was UV cross-linked to the membrane using a Stratagene "cross linker" (120 mJ/cm^2 in auto-crosslinking mode). Pre-hybridization was performed for at least 2 h in Church buffer (0.25 M sodium phosphate buffer pH 7.2, 1 mM EDTA, 1% BSA, 7% SDS) at 65°C. Single-stranded DNA oligonucleotide probes (Table S1) were labeled with [γ-^{32}P]ATP (3000 Ci/mmol, 10 mCi/mL, Perkin Elmer) using the T4 polynucleotide kinase (Life technologies, No. EK 0031). The resulting probes were purified on Micro-Spin G25 columns (GE Healthcare, cat. No. 27-5325-01) to remove unincorporated nucleotides, denatured for 5 min at 95°C and incubated with the membrane in 10 mL Church buffer at 65°C overnight. Next day, washing was performed at 50°C as follows: twice 10 min each with Wash buffer-1 (2 × SSC, 0.1% SDS) and twice 10 min each with Wash buffer-2 (0.2 × SSC, 0.1% SDS). The membrane was wrapped in Saranfilm, exposed to Storage Phosphor Screens (GE Healthcare; BAS IP MS 2025 E, Cat no. 28-9564-75) and scanned (GE Healthcare; Typhoon FLA 9500 IP, Cat no. 29-1885-90). The signal of 5S rRNA was used as a loading control. Sequences of the probes are given (Table S1).

Quantification of RNA modifications using LC-MS/MS

RNA was hydrolyzed to ribonucleosides by 20 U benzonase (Santa Cruz Biotech) and 0.2 U nuclease P1 (Sigma) in 10 mM ammonium acetate pH 6.0 and 1 mM magnesium chloride at 40°C for 1 h. After that, ammonium bicarbonate to 50 mM, 0.002 U phosphodiesterase I and 0.1 U alkaline phosphatase (Sigma) were added, and incubated further at 37°C for 1 h. The hydrolysates were mixed with

3 volumes of acetonitrile and centrifuged (16,000xg, 30 min, 4°C). The supernatants were dried and dissolved in 50 μL water for LC-MS/MS analysis of modified and unmodified ribonucleosides. Chromatographic separation was performed using an Agilent 1290 Infinity II UHPLC system with an ZORBAX RRHD Eclipse Plus C18 150 × 2.1 mm ID (1.8 μm) column protected with an ZORBAX RRHD Eclipse Plus C18 5 × 2.1 mm ID (1.8 μm) guard column (Agilent). The mobile phase consisted of water and methanol (both added 0.1% formic acid) run at 0.23 mL/min. For modifications, starting with 5% methanol for 0.5 min followed by a 2.5 min gradient of 5%–15% methanol, a 3 min gradient of 15%–95% methanol and 4 min re-equilibration with 5% methanol. A portion of each sample was diluted for the analysis of unmodified ribonucleosides which was chromatographed isocratically with 20% methanol. Mass spectrometric detection was performed using an Agilent 6495 Triple Quadrupole system operating in positive electrospray ionization mode, monitoring the mass transitions 269.1-150.1 (m^6Am), 282.1-150.1 (m^6A), 282.1-136.1 (Am), 268.1 (A), 284.1-152.1 (G), 244.1-112.1 (C), 245.1-113.1 (U), 296.1-164.1 ($m^6{}_2A$), and 298.1-166.1 (m^7G).

Preparation of m^7G-cap enriched RNA fragment for RNA mass spectrometry

Liver Poly(A)$^+$ RNA (~8 μg) from three biological replicates of both *Pcif1*$^{+/-}$ and *Pcif1*$^{-/-}$ genotypes was purified with mRNA purification kit (Thermo Scientific, Catalog 61006). The poly(A)$^+$ RNA was further purified with RNA Clean & Concentrator-5 (ZYMO Research, R1014) and eluted in 17 μL water. The RNAs were fragmented with RNA fragmentation reagent (Thermo Scientific, catalog no. AM8740) in a final volume of 20 μL at 70°C for 7.5 min in a thermocycler. Fragmentation reaction was stopped by transferring the reaction mix quickly on ice and adding 2.2 μL of STOP solution (supplied with the fragmentation reagent) and gentle mixing. The fragmented RNAs were purified with RNA Clean & Concentrator-5 and eluted in 17 μL water. RNAs were 5′ phosphorylated by adding 2 μL of 10X T4 DNA ligase buffer (NEB) and 20 units of T4 PNK (ThermoScientific) in a final reaction volume of 20 μL. The phosphorylation reaction was carried out for 1.30 h at 37°C. The m^7G-capped RNA fragments were enriched by specifically degrading the 5′-phosphorylated RNA fragments with the Terminator 5′-Phosphate-Dependent Exonuclease (Epibio). To the phosphorylation reaction, 14 μL of water, 4 μL 10x Exonuclease buffer A and 2 μL Terminator exonuclease was added and the reaction mixture was incubated at 30°C for 3 h. RNAs were purified from above reaction with RNA Clean & Concentrator-5 and eluted in 15 μL water. At this point we get 300-400 ng m^7G-capped RNAs fragments from each sample.

The m^7G-cap enriched RNAs were divided into two parts. One part was treated with nuclease P1 (cleaves only phosphodiester bonds, leaving the cap structure linked to the TSS nucleotide) and alkaline phosphatase followed by LC-MS/MS analysis to identify internal (body) RNA modifications in the RNAs, whereas second part was completely digested with a mixture of nucleases [20 U benzonase (Santa Cruz Biotech) and 0.2 U nuclease P1 (Sigma) in 10 mM ammonium acetate pH 6.0 and 1 mM magnesium chloride at 40°C for 1 h] to obtain RNA modification present on the entire fragment in the RNAs. The nuclease P1-digest will have all internal modifications, but lack that of the m^7G cap structure and that on the transcription start site (TSS) nucleotide. The complete digest will have all modifications including that of the cap, TSS nucleotide and body of the RNA.

Preparation of m^7G-capped RNA by *in vitro* transcription

A 40 nt m^7G-capped RNA initiating with an adenosine was generated by *in vitro* transcription as previously described (Jiao et al., 2017). The commonly used T7 class III promoter ϕ6.5 initiates using GTP, while the class II promoter ϕ2.5 initiates with ATP (Coleman et al., 2004). We prepared a single-stranded DNA template containing the T7 class II promoter ϕ2.5 at the 5′ end, with the transcription start site adenosine, followed by residues composed of T, C and G (there is no adenosine except at the TSS position). The template with a G as the TSS nucleotide was prepared using the commonly used T7 class III promoter ϕ6.5. The templates (Wu et al., 2019) were amplified by PCR reaction to generate the dsDNA template for *in vitro* transcription (Table S1).

Sequences of DNA template used are indicated below. Only the top-strand (sense strand) is shown.

CAGTAATACGACTCACTATT**A**GGCCTCTCGCTCTGCTGGGTGTGCGCTTGCTTGGCTTGC.

The ϕ2.5 promoter (underlined) was used for the template with TSS as adenosine (A). The transcription-start site nucleotide is in bold.

CAGTAATACGACTCACTATA**G**GGCCTCTCGCTCTGCTGGGTGTGCGCTTGCTTGGCTTGC

The T7 class III promoter ϕ6.5 was used for template with TSS as guanosine (G).

In vitro transcription reaction was carried out at 37°C for 4 h in a 20 μL reaction containing 6 pmol DNA template, 2 μL 10x reaction buffer, 2 μL of 75 mM CTP and 75 mM UTP, 0.5 μL of 75 mM GTP and 2 μL 75 mM $m^7G(5)ppp(5)G$ (NEB, Cat. No. S1404S) or 75 mM $m^7G(5)ppp(5)A$ (NEB, Cat. No. S1405S) or $m^7G(5)ppp(5)(2'OMeA)pG$ (Trilink, Cat. No. N-7413) and 2 μL T7 enzyme mix (MEGAshortscript Kit; ThermoFisher Scientific). Generated RNAs were purified by phenol-chloroform extraction.

In vitro methylation assays with PCIF1

The m^7G-capped RNAs were prepared by *in vitro* transcription as described above. The *in vitro* methylation assay was performed overnight at 37°C in a reaction mix containing 50 mM HEPES-KOH pH 7.5, 1 mM DTT, 0.02 μCi C^{14}SAM (PerkinElmer, Cat. No. NEC363010UC), 500 pmol RNA substrate and 5 μg recombinant hPCIF1, dPcif1 or TrypPcif1. The next day, the RNAs in the samples were purified by phenol-chloroform extraction and precipitation. 50% of the samples were resolved via 15% UREA-PAGE. The gel was run 1 h at 12 W. After the run, the gel was stained with methylene blue staining to check for the presence of RNA. Afterward the

gel was dried in a dryer (Bio-Rad, model 583) with gradual heating and cooling, 80°C for 1.5 h. The dried gel was exposed to exposed to Storage Phosphor Screens (GE Healthcare; BAS IP MS 2025 E, Cat no. 28-9564-75) for 48 h and scanned (GE Healthcare; Typhoon FLA 9500 IP, Cat no. 29-1885-90).

For RNA mass spectrometry, the m^7G-capped RNAs were prepared by *in vitro* transcription as described above. The *in vitro* methylation assay was performed overnight at 37°C in a reaction mix containing 50 mM HEPES-KOH pH 7.5, 1 mM DTT 2.5 mM SAM (NEB, Cat. No. B9003S), 500 pmol RNA substrate and 5 μg recombinant hPCIF1, dPcif1 or TrypPcif1. The next day, the RNAs in the samples were purified by phenol-chloroform extraction and precipitation, RNA pellet was washed with 80% ethanol, briefly dried, and dissolved in 20 μL RNase-free water.

Phosphorylated CTD-binding assays with recombinant PCIF1

The following N-term biotinylated peptides (Thermo Fisher Scientific) were purchased:

UnmodCTD: Heptad repeats of YSPTSPS, unmodified
Ser5-CTD: Heptad repeats of YSPT[(pS)]PS
Ser2-CTD: Heptad repeats of Y[(pS)]PTSPS

Dynabeads M-280 (Invitrogen, Cat. No. 11205D) were washed in binding buffer (25 mM Tris-HCl pH8.0, 50 mM NaCl, 1 mM DTT, 5% glycerol and 0.03% Triton X-100), and 1 nmol of the biotinylated CTD peptides were immobilized on 0.5 mg of beads in 50 μL of binding buffer. The beads were mixed with 3 μg of the recombinant full-length dPcif1, full-length hPCIF1 or the GST fused WW domains from the respective proteins, and incubated for 30 min on ice. After 3 washes, the beads were resuspended in 20 μL of binding buffer and boiled in 1x Laemmli buffer. The 20% of input protein, 40% of the eluate from the beads control or experimental samples were resolved via 15% SDS-PAGE. The bound full-length PCIF1 proteins were detected by Western analyses. The bound GST-fused WW domains were detected by Silver stain (Invitrogen, Cat. No. LC6070).

Isothermal calorimetry (ITC) experiments

ITC experiments were performed using a MicroCal iTC 200 (Malvern Panalytical) at 25°C. The dPcif1 protein was dialyzed in buffer (20 mM HEPES, pH 7.5, and 100 mM NaCl) overnight, and was titrated with two different RNA cap analogs, m^7G(5′)ppp(5′)G (NEB, Cat. No. S1404S) and m^7G(5′)ppp(5′)A (NEB, Cat. No. S1405S). The sample cell was filled with 13.5 μM dPcif1, and the syringe was filled with 135 μM RNA cap analog. Titrations were carried out by a preliminary 0.4 μL injection followed by constant volume injections (19 injections of 2 μL) with 150 s spacing. Data analysis was performed using Origin software. As for TrypPcif1, the protocol is same as dPcif1, except that the concentration of protein is 30 μM and the concentration of RNA cap analog is 300 μM [m^7G(5)ppp(5)G (NEB, Cat. No. S1404S) or m^7G(5)ppp(5)A (NEB, Cat. No. S1405S) or m^7G(5)ppp(5)(2′OMeA)pG (Trilink, Cat. No. N-7413)]. We thank Andrew McCarthy at EMBL Grenoble, France for access to the instrument.

Preparation of RNA libraries

Total RNA sequencing

Total RNA from biological triplicates of 2 months-old wild-type (WT) and *Pcif1* KO mice were used for the experiment with mouse brain, spleen, and testis samples. The tissues were snap-frozen in liquid nitrogen. Total RNA was isolated from tissues using TRIzol (Thermo Fisher Scientific) according to manufacturer instructions. The TruSeq Stranded Total RNA kit with Ribo-Zero Gold (Illumina, Cat. No. 20020612) was used for library preparation with 400 ng of total RNA as input. Ribodepleted RNAs were subjected to RNA-seq library preparation with NEB kit NEBNext® Ultra II Directional RNA Library Prep Kit (Illumina, Cat. No. E7760), following manufacturer's instruction. Library molarity and quality was assessed with the Qubit and Tapestation using a DNA High sensitivity chip (Agilent Technologies). Libraries were diluted at 2 nM and pooled before the clustering process on a HiSeq 4000 Single Read flow cell. Paired-end reads of 100 bases were generated using the TruSeq SBS reagents on the Illumina HiSeq 4000 sequencer (iGE3 Genomics Platform, University of Geneva).

Ribosome profiling

Brains from three adult wild-type (WT) controls and three *Pcif1* KO mice were isolated and snap-frozen in liquid nitrogen. Frozen brain (~400 mg) was homogenized in 1300 μL (3x weight of the tissue in miligrams) of ice cold polysome lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 100 μg/ml Cycloheximide, 40 U/ml RNasin plus, (Promega), Complete EDTA-free protease inhibitor (Thermo Scientific), 1% TritonX-100, 0.5% Sodium deoxycholate) using a glass homogenizer (6-8 strokes). The lysate was transferred to a 1.5 mL Eppendorf tube and incubated on ice for 10 min with intermittent mixing. The lysate was centrifuged at 3000 rpm for 3 min at 4°C. Supernatant was carefully collected in a new 1.5 mL Eppendorf tube. The OD_{260} of the lysate was measured with a Nanodrop instrument and approximately 15 OD_{260} (which correspond to 100 μL of lysate), was used for ribosome footprint preparation, and 50 μL was directly transferred to TriZol (Ambion) for total input RNA extraction. The remaining lysate was divided in aliquots of 150 μL and was snap frozen in liquid nitrogen and stored in −80°C freezer for later use. For ribosome footprint preparation, 100 μL of the supernatant was treated with 2.5 μL (5 units) of Turbo-DNase (Ambion) and 6.5 μL (650 units) of RNase I at 23°C for 45 min with rotation at 300rmp in a thermoblock. In meantime, the s-400 HR columns (GE healthcare) were washed three

times with 700 μL polysome buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 100 μg/mL Cycloheximide, 40U/mL RNasin plus, (Promega), Complete EDTA-free protease inhibitor (Thermo Scientific)) plus 20U/mL SUPERaseIn (Ambion). During each wash columns with polysome buffer were spun at 2000rpm for 1 min in a cold centrifuge. The columns were stored on ice with the polysome buffer in it (last wash). Once the RNaseI-treatment of lysate completed, 8.7 μL of SUPERaseIn RNase inhibitor was added to the reaction, gently mixed with the help of a pipette and stored on ice. The S-400 HR columns were spun at 2000 rpm for 1 min to remove the polysome buffer. RNaseI-treated lysate was applied on the top of the column and spun at 2400 rpm for 2 min. The flow through was transferred to a fresh Eppendorf tube and RNA was extracted using 1.0 mL TriZol reagent. The RNA contain ribosome protected fragments (RPF) of mRNA as well as ribosomal RNA.

Approximately 20-25 μg of RNaseI-treated RNAs were separated on 15% UREA-PAGE to purify RPF which are in the range of 26-34 nt. Radiolabelled RNA marker of length 28 and 30 nt were used to size select the ribosome protected RNA fragment (RPF). The urea polyacrylamide gel was run at 16 W for 1 h with the samples. Once the bromophenol blue reached the bottom of the gel, the power was switched off and the gel was taken out of the running apparatus. One glass plate was removed and the gel was wrapped in saran wrap and exposed to Storage Phosphor Screens (GE Healthcare; BAS IP MS 2025 E, Cat no. 28-9564-75) for 30 min, and scanned (GE Healthcare; Typhoon FLA 9500 IP, Cat no. 29-1885-90). Based on the position of markers, the gel piece corresponding to approximately 26-34 nt was excised and transfer to an RNase-free Eppendorf tube. The gel piece was ground to a fine powder and 600 μL of RNA extraction buffer (0.5 M Ammonium Acetate, 1 mM EDTA pH 8.0, and 0.1% SDS) was added and the tubes were incubated at 25°C overnight. Next day, the slurry was transferred to filter tubes (from Art-seq kit) and spun for 5 min at 5000 rpm. The flow-through was collected in a new RNase-free Eppendorf tube and precipitated with 1.0 μL of glycogen and 900 μL isopropanol overnight in −20°C freezer. The RNA solution was centrifuged at 13000 rpm for 20 min in a cold centrifuge to pellet RNA. RNA pellet was washed with 80% ethanol once, air-dried and resuspended in 20 μL of RNase-free water.

Ribosomal RNAs were depleted from both input RNAs (400 ng) and RPF RNAs with mouse ribosomal RNAs probe from siTOOLs (mouse riboPOOLs) following their protocol. The ribosomal RNA-depleted input and RPF RNAs were used to make next generation sequencing library with TruSeq Ribo Profile (mammalian) kit from illumina. The libraries were sequenced at Genomics platform (University of Geneva) for 50 cycles.

m^6Am-Exo-Seq

m^6Am-Exo-Seq (Sendinc et al., 2019) is a variation of m^6A-IP seq. In this protocol, first, the m^7G-capped RNA fragments (~100-200 nt) from poly(A)$^+$ RNAs are enriched by multiple enzymatic treatments followed by m^6A IP-seq to get m^6A modification near the 5′ end of the mRNAs.

Approximately 8.0 μg of poly(A)$^+$ RNAs from biological triplicates from *Pcif1*$^{+/-}$ and *Pcif1*$^{-/-}$ mouse testes were fragmented with RNA fragmentation buffer (Ambion) in a total volume of 20 μL for 7.30 min at 70°C in a thermcycler to obtain RNA fragments of ~100-200nt length. Fragmentation reaction was stopped by transferring the RNAs quickly to ice and adding 2.2 μL 10x stop solution (Ambion). The fragmented RNAs were purified with RNA Clean & Concentrator-5 (Zymo Research, R1014) and eluted in 17 μL water. Fragmented RNAs were 5′ phosphorylated by adding 10x, 2 μL T4 DNA ligase buffer (NEB) and 20 units of T4 PNK (ThermoScientific) in a final reaction volume of 20 μL. The phosphorylation reaction was carried out for 1.30 h at 37°C. The 5′ phosphorylated RNA fragments were degraded by Terminator 5′-Phosphate-Dependent Exonuclease (Epibio). To the phosphorylation reaction, 14 μL of water, 4 μL 10x Exonuclease buffer A and 2 μL Terminator exonuclease was added and the reaction mixture was incubated at 30°C for 3 h. RNAs were purified from above reaction with RNA Clean & Concentrator-5 and eluted in 15 μL water. At this point we get 300-400 ng of m^7G-cap-enriched RNA fragments from each sample.

The RNA obtained from above reaction were directly used for m^6A-IP (Wojtas et al., 2017) with some modifications. First, 100 μL of protein A dynabeads (Life Technologies; 10002D) were washed twice with PXL buffer (1X PBS, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% NP-40) and blocked with 1 μg/μL BSA in 400 μL PXL buffer for 45 min. Blocked beads were then incubated with 5 μg of anti-m^6A antibody (Synaptic Systems, Cat. no. 202003) in 200 μL PXL buffer supplemented with 4 μL of RNasin RNase inhibitor (Promega; N2611) for 1 h at room temperature. Antibody-conjugated beads were washed twice with PXL buffer and finally resuspended in 400 μL PXL buffer plus 4 μL RNAsin RNase inhibitor. Fragmented m^7G-cap enriched RNAs were added and beads were incubated at 4°C for 2 h with rotation. 10% of the RNA was kept separately and used to prepare input libraries. After 2 h incubation, the beads were washed twice by ice-cold Nelson low-salt buffer (15 mM Tris at pH 7.5, 5 mM EDTA), once by ice-cold Nelson high-salt buffer (15 mM Tris at pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1 M NaCl), once by ice-cold Nelson stringent wash buffer (15 mM Tris at pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 120 mM NaCl, 25 mM KCl), and last by ice-cold NT-2 buffer (50 mM Tris at pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40). Antibody-bound RNAs were eluted by adding 1mL TRIzol (Invitrogen) directly to the beads. RNAs were extracted and precipitated with Glycogen and ethanol overnight in −20°C, and used directly for the library preparation with NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina® (E7765) according to the protocol 3 of the kit (Protocol for use with FFPE RNA, step 3.5 onward). The final PCR was performed for 15 cycles (98°C – 30 s, 12 cycle of 98°C – 10 s, 65°C – 75 s and final extension at 65°C – 5 min). The libraries were sequenced on HiSeq4000 machine at genomics platform (University of Geneva), for single end 100 cycles.

All sequencing libraries prepared are listed in Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of RNaseq data

Analysis of gene expression data

Reads were sorted into individual libraries based on the barcodes and mapped to the mouse genome (Ensembl release 95) using salmon v0.12.0 (salmon quant with options -l A –validateMappings–gcBias) (Patro et al., 2017). Further analysis was performed using R version 3.6.2 (R Core Team, 2017) and Bioconductor (Huber et al., 2015). The DESeq function of DESeq2_1.25.10 bioconductor package (Love et al., 2014) was used to obtain log2 fold changes of gene expression between wild-type and mutant samples and the adjusted p values. Adjusted p value 0.1 was used as a threshold for statistical significance. The individual mouse tissues were analyzed separately. In MA plots the genes with significantly different expression between the mutant and wild-type samples were highlighted in red. The sets of genes upregulated and dowregulated in the mutant tissues were compared using Venn diagrams (VennDiagram 1.6.20). The Volcano plots were plotted using EnhancedVolcano function from EnhancedVolcano 1.3.5 package (https://github.com/kevinblighe/EnhancedVolcano). Significantly differentially expressed genes with at least two fold change in expression between mutant and wild-type were highlighted in red. We put a specific attention on pseudogenes and predicted genes which were among the highly dysregulated genes. In pseudogene category we included the genes which contain the "pseudogene" string in Ensembl biotype annotation. The genes containing "predicted gene" string in Ensembl description were considered as predicted genes. The frequency of the pseudogenes and predicted genes was compared among the expressed genes (genes with defined p value and log2 fold change), genes upregulated and downregulated in the mutant (adjusted p value < 0.1) and top 200 and 100 upregulated (adjusted p value < 0.1 and log2 fold change >0) and downregulated (adjusted p value < 0.1 and log2 fold change < 0) genes, ordered based on log2 fold change. In individual tissues we also compared the proportion of the pseudogenes which are upregulated, downregulated, and not changed in the mutant (adjusted p value > = 0.1). To check whether the expression of parental genes of dysregulated pseudogenes is affected, we obtained the information about the parental genes for the pseudogenes using C57BL/6NJ data from http://mouse.pseudogene.org/. Only for 20%–30% of dysregulated pseudogenes in individual tissues we were able to obtain the parental gene identifier. The gene expression of the dysregulated pseudogenes and their parental genes was visualized using boxplots with the individual genes plotted as gray dots. To see whether whole chromosome expression might be affected, we produced boxplots of log2 fold expression changes for genes lying on individual chromosomes with the y-axes limited to $< -1,1 >$ to better see whether the median fold change deviates from zero. To inspect the chromosomal localization of individual dysregulated genes, we plotted the ideograms where we highlighted their localization. To assess whether the transcripts of dysregulated genes preferentially start with specific nucleotide, for every expressed gene we computed the fractions of its transcript isoforms annotated in Ensembl starting with A,U,G and C. These fractions made for individual starting nucleotides (considered as TSS nucleotides) were summed for selected groups of genes and used to calculate the fraction of genes starting with individual nucleotides. Barplots were made to compare the genes which were dysregulated in the mutant with the genes without detected significant difference. To see whether the starting nucleotide affects the overall gene expression changes, we filtered only genes which produce transcripts with specific starting nucleotide and divided them into the groups based on that nucleotide. We then computed empirical cumulative distribution function on log2 fold changes between mutant and wild-type and plotted the distribution.

The group of genes found to be significantly up- or downregulated in the mutant were searched for enriched Gene Ontology terms in the Biological Process ontology using ENRICHR (Chen et al., 2013; Kuleshov et al., 2016) and the enriched categories were shown.

To find out whether the dysregulated genes in the *Pcif1* mutant testes, are expressed in the specific spermatogenic populations, we re-analyzed the published dataset containing gene expression data for Sertolli cells, spermatogonia, spermatocytes, spermatids, and spermatozoa (Soumillon et al., 2013). Using the boxplots we plotted the expression in these populations for the genes we found to be dysregulated in the mutant testes.

Analysis of ribosome footprinting data

To analyze the translational differences between the mutant and wild-type brain samples for individual genes, we compared the ratio of footprint versus the input reads (i.e., translation efficiency) for individual genes. We tested the differences in translation efficiency for individual genes using DESeq2 design: '~assay(input/footprint) + condition(KO/WT) + assay:condition' and applying the likelihood ratio test which removes the interaction term in the reduced model. For each gene we obtained log2 fold changes of gene translation efficiency difference between wild-type and mutant samples and the adjusted p value. Downstream analysis was analogous to the analysis of gene expression data.

Analysis of m^{6}Am-Exo-Seq data

Reads were sorted into individual libraries based on the barcodes and mapped to the mouse genome (Ensembl release 95) using STAR (Dobin et al., 2013) with parameters "–outFilterType BySJout–limitOutSJcollapsed 50000000–limitIObufferSize 1500000000." The genomic coverages for individual samples were created in R using GenomicAlignments::coverage method. Cumulative coverages of input and m^{6}A IP samples around the annotated TSS positions were calculated and plotted for −500:+500 distance, as well as an m^{6}A enrichment (log2 of m^{6}A/input ratio). As expected the sequenced fragments came predominantly from 5′ ends of the transcripts as demonstrated by increased coverage close to TSS. The TSSs were then divided according to the starting nucleotide and the m^{6}A enrichment around the TSSs was plotted. The m^{6}A enrichment at the starting nucleotide was

compared between in *Pcif1* knockout and wilt-type mice (t.test, n = 3). Significantly lower m^6A enrichment in *Pcif1* knockout mice was detected only for A starting transcripts (pval = 0.009).

For individual genes we tested the differences in m^6A enrichment of their TSSs using matrix of read counts covering TSS of individual genes and DESeq2 design: '~assay(input/m^6A_IP) + condition(KO/HET) + assay:condition' and applying the likelihood ratio test which removes the interaction term in the reduced model. For each gene we obtained log2 fold m^6A enrichment difference between wild-type and mutant samples and the adjusted p value. We identified 1370 genes which had significantly lower m^6A TSS enrichment in the mutants. Importantly these genes tend to have decreased expression in the mutant as shown by different empirical cumulative distribution curves of log2 fold changes in expression between mutant and wild-type mice.