Title

PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage

Permalink
https://escholarship.org/uc/item/4735b8qv

Journal
Cell Reports, 15(2)

ISSN
2639-1856

Authors
Hwang, Hun-Way
Park, Christopher Y
Goodarzi, Hani
et al.

Publication Date
2016-04-01

DOI
10.1016/j.celrep.2016.03.023

Peer reviewed

Published in final edited form as:

Cell Rep. 2016 April 12; 15(2): 423–435. doi:10.1016/j.celrep.2016.03.023.

PAPERCLIP identifies microRNA targets and a role of CstF64/64tau in promoting non-canonical poly(A) site usage

Hun-Way Hwang[1,5], **Christopher Y. Park**[1,3], **Hani Goodarzi**[2], **John J. Fak**[1], **Aldo Mele**[1,4], **Michael J. Moore**[1], **Yuhki Saito**[1], and **Robert B. Darnell**[1,3,5]

[1]Laboratory of Molecular Neuro-Oncology and Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA.

[2]Laboratory of Systems Cancer Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.

[3]New York Genome Center, 101 Avenue of the Americas, New York, NY 10013, USA.

Abstract

Accurate and precise annotation of the 3′ untranslated regions (3′ UTRs) is critical in understanding how mRNAs are regulated by microRNAs (miRNAs) and RNA-binding proteins (RBPs). Here we describe a method, PAPERCLIP (*P*oly(*A*) binding *P*rotein-mediated mRNA 3′ *E*nd *R*etrieval by *C*ross*L*inking *I*mmuno*P*recipitation), which shows high specificity for the mRNA 3′ ends and compares favorably to existing 3′ end mapping methods. PAPERCLIP uncovers a previously unrecognized role of CstF64/64tau in promoting the usage of a selected group of non-canonical poly(A) sites, the majority of them containing a downstream GUKKU motif. Furthermore, in mouse brain, PAPERCLIP discovers extended 3′ UTR sequences harboring functional miRNA binding sites and reveals developmentally regulated APA shifts including one in *Atp2b2* that is evolutionarily conserved in human and results in a gain of a functional binding site of miR-137. PAPERCLIP provides a powerful tool to decipher post-transcriptional regulation of mRNAs through APA *in vivo*.

[5]Corresponding authors CONTACT INFORMATION: Address: Box 226, 1230 York Avenue, New York, New York 10065, USA, Phone: (212) 327-7460, Fax: (212) 327-7109, RDarnell@nygenome.org (R.B.D.), hhwang@rockefeller.edu (H.W.H).
4Present address: Horizon Discovery, 7100 Cambridge Research Park, Waterbeach, Cambridge, CB25 9TL, UK.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Contributions: H-W.H. and R.B.D. conceived the project. H-W.H. designed the experiments. R.B.D. supervised the project. H-W.H. and A.M. developed the PAPERCLIP protocol. J.J.F. performed the northern blot experiments. Y.S. performed the mouse embryo dissection. H-W.H. performed all other experiments. H-W.H., C.Y.P., and H.G. performed the bioinformatics analyses. M.J.M. provided the Ago HITS-CLIP data. H-W.H. and R.B.D wrote the manuscript.

ACCESSION NUMBERS
The GEO accession number for PAPERCLIP and RNA-seq data originated from this study is GSE66092.

Conflict of interest statement: None declared.

INTRODUCTION

The 3′ untranslated regions (3′ UTRs) of mRNAs often contain sequence elements to interact with regulators such as microRNAs (miRNAs) and RNA-binding proteins (RBPs) (Elkon et al., 2013; Licatalosi and Darnell, 2010). Through alternative polyadenylation (APA), cells and tissues are able to fine-tune gene expression by altering inclusion of 3′ UTR regulatory elements to allow or deny access to regulators (Di Giammartino et al., 2011). Therefore, accurate and precise annotation of the 3′ UTRs is critical in understanding how mRNA stability, localization and translational efficiency are regulated by miRNAs and RBPs.

mRNA 3′ end mapping is commonly performed by reverse transcribing mRNAs from the poly(A) tails, utilizing oligo(dT) primers to construct the sequencing library (hereafter referred to as “3′ end sequencing”). Different varieties of this approach such as PAS-Seq, PolyA-Seq, 3′Seq and 3′-seq (different from the former) have been described recently in the literature (Derti et al., 2012; Jenal et al., 2012; Lianoglou et al., 2013; Shepard et al., 2011). (For a more complete list, see a recent review (Shi, 2012).) However, the oligo(dT) primers also anneal to internal adenine-rich sequences in mRNA and results in mispriming during reverse transcription (Di Giammartino et al., 2011; Nam et al., 2002). The presence of internal priming events, which may represent 30% of the entire library (Derti et al., 2012), diminishes the effective sequencing depth and requires strict *in silico* filtering to improve accuracy.

Different filtering strategies have been developed to remove potential internal priming events (Derti et al., 2012; Shepard et al., 2011). One commonly used filtering strategy is to remove reads mapping immediately upstream of genomic adenine-rich regions, as they are likely arising from mispriming events. However, these manipulations remove many authentic poly(A) sites in addition to misprimed sites (Hoque et al., 2013), and different algorithms exhibit distinct biases (Derti et al., 2012). To minimize false-positives, studies utilizing 3′ end sequencing are inclined to be conservative about poly(A) site annotations, which diminishes the possibility of identifying novel poly(A) sites and limits the scope of study.

To improve mRNA 3′ end mapping, we applied the CLIP (Crosslinking Immunoprecipitation) technique to poly(A)-binding protein, a factor with high specificity for mRNA poly(A) tails (Kahvejian et al., 2001). CLIP captures direct RNA-protein interaction *in situ*, allowing stringent immunopurification that effectively eliminates non-specific interactions to generate precise genome-wide interaction maps (Licatalosi et al., 2008). Our method, PAPERCLIP, minimizes the internal priming problem without utilizing *in silico* filtering and allows unbiased discovery of mRNA 3′ ends. We performed extensive validation to show that PAPERCLIP has high specificity for the mRNA 3′ ends and generates comprehensive maps of poly(A) sites at nucleotide resolution. Functional analysis using PAPERCLIP discovered a previously unrecognized role of CstF64/64tau in promoting the usage of non-canonical poly(A) sites. We further combined PAPERCLIP and Argonaute (Ago) HITS-CLIP (Chi et al., 2009) in both adult and developing mouse brain to discover regulatory relationships between mRNAs and miRNAs, demonstrating the capacity of

PAPERCLIP to provide critical knowledge in understanding post-transcriptional regulation of mRNAs *in vivo*.

RESULTS

PAPERCLIP ameliorates the internal priming problem

To develop an alternative strategy to globally map mRNA 3′ ends with high specificity, we took advantage of the well-documented property of PABP, which is its high selectivity for the poly(A) tails of mRNAs (Kahvejian et al., 2001), to develop PAPERCLIP (Figure 1A and 1B). In PAPERCLIP, only mRNA fragments containing the poly(A) tail are retained by crosslinking to PABP, while the internal adenine-rich fragments are excluded from the final sequencing library (Figure 1C). As a result, stringent sequence-based or genomic location-based filtering is not necessary, allowing the poly(A) sites to be mapped in an unbiased manner.

To directly compare PAPERCLIP to 3′ end sequencing, we used a 3′ end sequencing protocol adapted from the 3′Seq protocol (Jenal et al., 2012) and performed both methods to map poly(A) sites in HeLa cells (Library statistics are provided in Table S1). Consistent with previous reports (Derti et al., 2012; Wang et al., 2013), the 3′ end sequencing library contained a large number of internal priming events prior to filtering, as evidenced by the strong enrichment of adenines immediately downstream of putative poly(A) sites in the human genome (Figure 1D, right panel). In contrast, the composition of nucleotide sequence surrounding poly(A) sites identified by PAPERCLIP is devoid of downstream genomic adenines and is highly similar to the nucleotide composition at mRNA 3′ ends mapped by direct RNA sequencing (Ozsolak et al., 2010) (Figure 1D, left panel). These results demonstrate that PAPERCLIP is highly specific for the mRNA 3′ ends and ameliorates the internal priming problem seen in the 3′ end sequencing.

We next compared PAPERCLIP to other 3′ end mapping methods using publicly available data (Library statistics are provided in Table S1). To evaluate the performance without *in silico* filtering, all uniquely mapped reads from each method were used for comparison. We first counted the number of reads that were aligned immediately upstream to genomic adenine-rich regions, a method commonly used to estimate the occurrence of internal priming events (Derti et al., 2012; Shepard et al., 2011). Among all the methods that were compared, PAPERCLIP and direct RNA sequencing both had the lowest percentage of downstream adenine-rich reads (Figure 1E). We next compared the genomic localization of the mapped reads. Overall, PAPERCLIP had the strongest enrichment of reads in annotated 3′ UTRs and 10kb downstream of annotated genes, in addition to the lowest percentage of intergenic reads (Figure 1F). Taken together, these results provide strong evidence that PAPERCLIP not only minimizes the internal priming problem but also compares favorably to many other 3′ end mapping methods including direct RNA sequencing and newer methods such as 3P-Seq.

PAPERCLIP provides comprehensive and precise mRNA polyadenylation maps in cultured cells

We next used the PAPERCLIP data to establish an mRNA polyadenylation map in HeLa cells. From two independent replicates (Figure S1A), we identified 17,652 high-confidence poly(A) sites located in 11,785 genes (Figure 2A). Importantly, PAPERCLIP was highly reproducible, as evidenced by the high correlation between the two replicates (Figure S1B) and the fact that the vast majority (15,176; 86%) of 17,652 poly(A) sites were annotated at the identical nucleotide position in both replicates (Figure S1C). Consistent with previous reports (Derti et al., 2012; Nam et al., 2014), we have observed considerable difference between the RefSeq annotation (Jun. 2014) and the poly(A) sites identified by PAPERCLIP. Among all 11,785 annotated genes detected in HeLa cells, 6,666 (57%) genes had at least one poly(A) site that was not present in the RefSeq annotation and 3,756 (32%) genes exclusively use non-RefSeq poly(A) sites. To validate results from PAPERCLIP, we first performed northern blotting and detected mRNAs at the expected size based on PAPERCLIP annotation (Figure 2B and 2C). Although it is known that RNA-seq lacks sufficient resolution to identify exact mRNA 3′ ends (Miura et al., 2014), we reasoned that it might provide additional support for 3′ UTR extensions identified by PAPERCLIP and broaden the scope of our validation in a high-throughput fashion. Indeed, through RNA-seq, we identified 17 genes that had 100% base coverage in the extended 3′ UTR regions annotated by PAPERCLIP and additional 61 genes with more than 90% coverage (Figure S2A and Table S2). Because most of the predicted mRNA transcripts with shortened 3′ UTRs also have overlapping longer isoforms, it is difficult to unequivocally demonstrate 3′ UTR shortening by RNA-seq alone given its limited power to resolve transcript isoforms. Nevertheless, we were able to confirm several examples of previously unrecognized short 3′ UTR isoforms predicted by PAPERCLIP in RNA-seq profiles (Figure S2B). As independent confirmation of the single nucleotide resolution of PAPERCLIP sites, we directly mapped selected individual polyadenylation events by 3′ RACE (Figure S2C). In all cases, the exact location of PAPERCLIP sites was validated. Overall, our validation data from northern blot, RNA-seq and 3′ RACE provide strong support for the high-resolution mRNA polyadenylation map established by PAPERCLIP in HeLa cells.

To test whether the observations in HeLa cells are generalizable, we performed PAPERCLIP in another cell-line, HEK293. Indeed, in HEK293 cells, we were able to obtain 16,414 high-confidence poly(A) sites in 11,488 genes with similarly high reproducibility and specificity for the 3′ ends (Figure S1D-H). As in HeLa cells, more than half of the annotated genes (6312; 55%) in HEK293 cells have at least one poly(A) site that is not present in the RefSeq annotation. Altogether, PAPERCLIP has identified 21,040 non-redundant poly(A) sites from HeLa and HEK293 cells. About 32% (6,778 of 21,040) of these poly(A) sites are not present in existing annotations (RefSeq and GENCODE, release 19). Of these un-annotated sites, 4,374 overlap with 3P-Seq read clusters (Nam et al., 2014) and 2,404 (35%) do not. Overall, 89% of the identified poly(A) sites are supported by either the standard annotations or 3P-Seq. Our results suggest that the mRNA polyadenylation events in individual cell-lines are incompletely covered by standard annotations and PAPERCLIP provides a robust tool to establish individualized mRNA polyadenylation maps with high precision and great depth.

PAPERCLIP offers transcriptome-wide assessment on APA in a quantitative manner

An important application for mRNA 3′ end mapping methods is to identify changes in poly(A) site usage. The mammalian mRNA 3′ processing complex contains a large number of proteins including 15 core factors such as CstF64 and CFIm68 (Shi et al., 2009). Recently, Shi and colleagues showed that co-depletion of CstF64 and its paralog CstF64tau in HeLa cells caused 201 genes to shift poly(A) site preference with a bias for the distal poly(A) site (Yao et al., 2012). In contrast, the Zavolan lab demonstrated that knockdown of CFIm68 in general induced the use of proximal poly(A) sites in HEK293 cells, although the numbers of genes that exhibited shift in APA preference were not discussed (Martin et al., 2012).

To examine the capacity of PAPERCLIP to identify shifts in poly(A) site usage, we performed siRNA-mediated knockdown for CFIm68 and CstF64/64tau in HeLa cells. Western blots demonstrated strong depletion of CFIm68 and CstF64/64tau in cells transfected with the corresponding siRNAs compared with control siRNAs (Figure 3A). We next performed PAPERCLIP to measure poly(A) site usage for all three conditions (Figure 3B). Both siRNA treatments resulted in a large number of APA shifts (1,911 genes in si-CFIm68 and 280 genes in si-CstF64/64tau) compared with control siRNAs (Figure 3C and Table S3). Consistent with previous reports (Martin et al., 2012; Yao et al., 2012), CFIm68 depletion and CstF64/64tau depletion had opposite effects on poly(A) site selection: the former caused a strong shift toward the proximal sites whereas the latter resulted in a preference for distal sites (Figure 3D). We next performed qRT-PCR as an independent measure to evaluate changes in APA in both conditions. In all cases, APA shifts identified by PAPERCLIP were also confirmed by qRT-PCR (Figure 3E and 3F). Although depletion of CFIm68 and CstF64/64tau overall had opposite effects on APA, we observed diverse effects for individual genes. For example, *DNAJB6* preferred the distal site in both conditions while *FAM3C* consistently shifted proximally (Figure 3E and 3F). A recent study identified genes that underwent 3′ UTR shortening after CFIm25 knockdown in HeLa cells (Masamha et al., 2014), and we found a strong agreement between the study and our data from CFIm68 depletion (525/538, 98%). This result is consistent with current knowledge that both CFIm25 and CFIm68 are part of the CFIm complex (Shi et al., 2009). We conclude that PAPERCLIP is able to offer a transcriptome-wide assessment on APA in a quantitative manner.

PAPERCLIP discovers a role of CstF64/64tau in promoting non-canonical poly(A) site usage

The binding motifs of the CFIm complex and CstF64/64tau have been characterized: the CFIm complex binds the UGUA motif upstream of the poly(A) sites while CstF64/64tau binds the DSE (downstream element) consisting of U/GU-rich motifs within 30nt 3′ to the poly(A) sites (Martin et al., 2012; Tian and Manley, 2013). Binding of the CFIm complex and CstF64/64tau generally promotes the usage of the poly(A) site. Moreover, for certain poly(A) sites that lack A(A/U)UAAA, the canonical poly(A) signal, the UGUA motif or the DSE becomes necessary for their usage (Nunes et al., 2010; Venkataraman et al., 2005). To identify sequence motifs that correlate with regulated APA, we performed an unbiased motif search for all possible 4-mers and 6-mers in proximity to APA sites identified in CFIm68 and CstF64/64tau depletion experiments.

For the 1744 genes showing loss of distal poly(A) site usage upon CFIm68 knockdown, there was a strong enrichment of UGUA (the top enriched 4-mer) and AAUAAA (the top enriched 6-mer) upstream the distal poly(A) site and an enrichment of U/GU-rich motifs (all top 10 enriched 6-mers) downstream (Table S3). These results are consistent with current knowledge that the distal poly(A) sites usually contain more canonical sequence elements (Shi, 2012) and explain the loss of distal poly(A) site usage upon depletion of CFIm68.

Interestingly, a different picture emerged from transcripts affected by CstF64/64tau depletion, which shifted predominantly from proximal to distal site usage. For the 241 genes showing loss of proximal poly(A) site usage upon depletion of CstF64/64tau, the proximal poly(A) sites have very strong enrichment of adenine-rich elements (defined as a hexamer with ≥5 adenosines excluding AAUAAA (Nunes et al., 2010); top 4 enriched 6-mers) in the upstream region and also enrichment of U/GU-rich elements downstream (8 of top 10 enriched 6-mers)(Table S3). It has been shown that the presence of an adenine-rich sequence and a strong DSE is sufficient to be recognized by the mRNA 3′ end processing complex for cleavage and polyadenylation (Nunes et al., 2010). The enrichment of these motifs near the proximal poly(A) sites shifted upon CstF64/64tau knockdown indicates a strong dependence on CstF64/64tau for selection and usage of these non-canonical sites.

Overall, the motif analysis results suggest that, in HeLa cells, CFIm68 is important in promoting distal poly(A) site usage for a large number of genes while CstF64 is necessary for selected non-canonical poly(A) sites that have strong DSE. For example, SERPINE1, which depends on CstF64 for the usage of proximal poly(A) site and CFIm68 for the usage of distal poly(A) site (Figure 3E, 3F and S3A), has its proximal poly(A) site flanked by an adenine-rich sequence and a sequence motif that closely resembles the consensus of mammalian DSE (Nunes et al., 2010) (Figure 3G and Figure S3B). In contrast, the distal poly(A) site of SERPINE1 lacks a DSE but has an UGUA motif and multiple AAUAAA motifs upstream (Figure 3G and Figure S3B).

PAPERCLIP identifies a GUKKU motif that contributes to non-canonical poly(A) site selection

The results from HeLa cells suggest a previously unrecognized role of CstF64 and CstF64tau in the selection of non-canonical poly(A) sites. To expand our findings in HeLa cells, we performed PAPERCLIP to examine APA shift from CFIm68 or CstF64/64tau depletion in another human cell-line, LN229. Similar to the HeLa results, CFIm68 depletion in LN229 resulted in a widespread proximal shift whereas CstF64/64tau depletion promoted distal shift for a smaller set of genes (Figure S3C-D and Table S3). Many genes, including SERPINE1, exhibited similar APA shifts in both cell-lines (Figure S3E and S3F).

There were 482 genes that showed distal shift (loss of proximal poly(A) site usage) upon CstF64/64tau depletion in HeLa and/or LN229 (Figure S3G, lower right panel). We focused on the subset of 105 genes that have non-canonical proximal poly(A) sites for motif search. Interestingly, a motif, GUKKU, was strongly enriched in the 50 nt region downstream to these 105 non-canonical proximal poly(A) sites (Figure 4A). The GUKKU motif is present in 58 genes (Table S4), which include SERPINE1, the prototypical gene whose dual-peak APA pattern depends on both CFIm68 and CstF64/64tau (Figure 4B, upper panel). EN2 is another GUKKU-containing gene that also requires both CFIm68 and CstF64/64tau for usage of its two main poly(A) sites (Figure S4A and Figure 4B, upper panel).

The G/U-rich nature of the GUKKU motif suggests that it is part of the DSE and might be involved in poly(A) site selection. To investigate whether the GUKKU motif is simply a sequence element associated with CstF64/64tau-dependent non-canonical poly(A) sites or it actually contributes to the usage of these sites, we took advantage of the absence of endogenous EN2 expression in HeLa cells (Figure 4B, lower panel) and made a construct expressing GFP upstream of the entire EN2 3′ UTR (pEN2, Figure 4C). pEN2 faithfully recapitulates the APA pattern observed from endogenous EN2 in LN229 when transfected into HeLa cells (Figure S4B) and exhibits APA shifts upon CFIm68 or CstF64/64tau depletion (Figure S4C). It was previously shown that adenine-rich elements are important for non-canonical poly(A) site recognition (Nunes et al., 2010). Therefore, we generated pEN2 mutants with mutated GUKKU motifs or adenine-rich elements (Figure 4D) and examined their APA patterns by PAPERCLIP. Indeed, both mutants had decreased proximal poly(A) site usage (Figure 4E and 4F), suggesting that both elements participate in the proximal poly(A) site selection and are necessary for its full usage.

To further test the functionality of the GUKKU motif, we performed additional experiments using the pEN2 construct as a tool. We identified a non-canonical poly(A) site in NOTCH1 that did not contain downstream GU-rich elements nor exhibit shift upon CstF64/64tau depletion (NOTCH1-WT, Fig. S4D). Therefore, we generated a hybrid poly(A) site in which the upstream half sequence originates from the NOTCH1 poly(A) site and the downstream half sequence (containing the GUKKU motifs) comes from the EN2 proximal poly(A) site (NOTCH1-HYB, Fig. S4D). We also generated a mutant version of NOTCH1-HYB, in which mutations were introduced to the GUKKU motifs (NOTCH1-MUT, Fig. S4D). We then made 3 new constructs in which the pEN2 proximal poly(A) site is replaced by each of the aforementioned NOTCH1 sites and confirmed their usage by 3′ RACE (data not shown). qPCR experiments demonstrated a sensitivity to CstF64/64tau depletion for NOTCH1-HYB, which is completely abolished in NOTCH1-MUT (Fig. S4E). Altogether, these results provide further support for a functional role of the GUKKU motif in the selection of non-canonical poly(A) site by CstF64/64tau.

We next wished to determine whether other GUKKU-containing non-canonical poly(A) sites that are not part of the 58-gene list could substitute functionally for the EN2 proximal poly(A) site and exhibit CstF64/64tau dependence. We searched for candidate poly(A) sites in a group of genes that used only one poly(A) site throughout the siRNA experiments and therefore were not included in the APA shift analysis (Fig. S4F). We have identified two non-canonical poly(A) sites in SRSF9 and JUNB that contain an upstream adenine-rich element and downstream GUKKU motifs but otherwise do not resemble the EN2 proximal poly(A) site in sequence. Therefore, we made new pEN2 hybrid constructs in which the EN2 proximal poly(A) site is replaced by the SRSF9 or JUNB poly(A) site (Figure 4G). Usage of the inserted poly(A) sites was confirmed by 3′ RACE (data not shown). Indeed, upon

CstF64/64tau depletion, both constructs showed distal shift similar to EN2 (Figure 4H), suggesting that CstF64/64tau also promotes usage of the two poly(A) sites in pEN2. Moreover, these results also indicate that SRSF9 and JUNB might have non-permissive sequence context that prevents APA shift. To test the hypothesis that the lack of a second functional poly(A) site might contribute to the absence of APA shift in JUNB, we generate pAcGFP-JUNB by cloning the entire JUNB 3′ UTR plus ~100bp sequence 3′ to the JUNB poly(A) site into pAcGFP-C1, which contains a poly(A) site downstream of the cloning site (Fig. S4G). Interestingly, CstF64/64tau depletion resulted in a strong distal APA shift for pAcGFP-JUNB (Fig. S4H), providing evidence that the surrounding sequence context indeed contributes to poly(A) site usage. Taken together, these results demonstrate that PAPERCLIP is able to identify a functional sequence motif that participate in poly(A) site selection *de novo* and provide insights into APA regulation.

PAPERCLIP identifies unannotated 3′ UTR extensions that have regulatory functions in mouse brain

Neurons use multiple neuron-specific RBPs and miRNAs to regulate mRNA metabolism (Darnell, 2013). It is well documented that brain mRNAs as a group have the longest 3′ UTRs compared with mRNAs in other tissues (Ji et al., 2009; Lianoglou et al., 2013; Ulitsky et al., 2012). Many mRNAs have neural-specific 3′ UTR extensions, which may provide platforms to interact with RBPs and miRNAs (Hilgers et al., 2011; Smibert et al., 2012). To identify 3′ UTR regulatory elements in the brain, we first performed PAPERCLIP in adult mouse cortex and annotated 10,117 poly(A) sites from 8,354 genes (Figure S5A-C). Overall, PAPERCLIP identified 1,024 3′ UTR extensions that were not present in the RefSeq annotation (Figure S5D and Table S5). Lai and colleagues recently identified many 3′ UTR extensions in the mouse brain through RNA-seq analysis (Miura et al., 2013). Therefore, we further compared the identified 3′ UTR extensions with GENCODE annotation and the 3′ UTR extensions reported by Lai and colleagues. In more than half of cases (566 of 1,024, 55%), the distal poly(A) sites identified by PAPERCLIP that demarcate the 3′ UTR extensions do not overlap with either GENCODE M2 annotation or the 3′ ends of the 3′ UTR extensions reported by Lai and colleagues. To provide support for the presence of the 566 3′ UTR isoforms identified by PAPERCLIP, we performed RNA-seq and examined the extended 3′ UTR regions for mapped reads. 457 (81%) PAPERCLIP-identified 3′ UTR extensions have FPKM≥1 and 351 (62%) have RNA-seq coverage at above 90% (129 have 100% coverage), providing strong support for the annotation by PAPERCLIP. Taken together, these results further expand the list of 3′ UTR extensions in mouse brain and demonstrate the capacity of PAPERCLIP to provide comprehensive poly(A) site maps for complex living tissues.

To identify regulatory elements located in the 1,024 extended 3′ UTR regions identified by PAPERCLIP, we overlaid Argonaute (Ago) HITS-CLIP data from mouse cortex and searched for miRNA seed matches in Ago footprints located in the extended 3′ UTR regions. To make sure that the identified regulatory relationship occurs in the same cell type, we examined miR-128, a miRNA that is highly enriched in the brain and exhibits a neuron-specific expression pattern (Bruno et al., 2011). We have identified 48 miR-128 seed matches overlapping robust Ago binding sites in the extended 3′ UTR regions and selected 8 neuronal transcripts to test functionality of the predicted miR-128 binding sites by luciferase assay (Table S5). In all cases, the repression of luciferase reporter expression was dependent on the presence of a miR-128 mimic and the wild-type miR-128 binding site (Figure 5A), suggesting these sites are indeed functional. In a recent study, Tan et al. has identified 154 miR-128 targets in mouse D1-neurons by combining FLAG-Ago2 HITS-CLIP and ribosome-associated mRNA analysis (Tan et al., 2013). We hypothesized that the newly identified miR-128 sites in the 3′ UTR extension might allow us to find additional miR-128 targets using the mRNA expression dataset generated by Tan et al. Indeed, we have identified 5 genes that are not on the Tan et al. gene list and have increased ribosome-association in miR-128-deficient D1 neurons, including 3 genes that we have validated in luciferase assays (Figure 5B and 5C). The increase in ribosome-association of the 5 miR-128 targets identified in current study is similar to the previously identified 154 genes (Figure 5D, p-value = 0.37, Wilcoxon rank sum test). These results indicate that functional regulatory sequences might currently be overlooked because of incomplete coverage of mRNA 3′ end annotations and that PAPERCLIP can identify 3′ UTR extensions that mediate regulatory functions.

PAPERCLIP uncovers an evolutionarily conserved APA shift during brain development

Lastly, to investigate whether APA occurs during mouse brain development, we performed PAPERCLIP in embryonic mouse cortex (Figure S5E-G) and compared the relative usage of poly(A) sites between adult and embryonic mouse cortex. PAPERCLIP identified 444 genes that significantly changed poly(A) site usage (Figure 6A and Table S6). Follow-up qRT-PCR experiments validated PAPERCLIP results (Figure 6B and 6C).

A very recent RNA-seq study using human cortex samples identified many “differentially expressed regions” (DERs) in the human genome across 6 life stages from fetus to middle age (Jaffe et al., 2015). Although RNA-seq does not detect APA shifts with high sensitivity, we decided to explore the possibility that some of the APA shifts identified by PAPERCLIP could also be detected in human by comparing our data with the reported DERs. We first lifted the coordinates of both the proximal and distal mouse poly(A) sites from the 444 genes to the human genome and then compared the regions flanked by the lifted poly(A) sites to the DERs. This analysis identified 3 candidate genes (*DCX*, *XPR1*, and *ATP2B2*) in which the regions flanked by the lifted mouse poly(A) sites have >80% coverage by DERs. To exclude the possibility that the observed differential expression in Jaffe et al. was due to transcriptional changes, we calculated the overlap between DERs and all exons for each gene. The overlap between DERs and all exons in *ATP2B2* is 26%, as only the last exon contains DERs (Figure 6D). In contrast, *DCX* and *XPR1* have almost perfect overlap between DERs and all their exons (>98%), suggesting that the major source for differential expression is transcription instead of alternative usage of 3′ UTR sequence. Therefore, we excluded *DCX* and *XPR1* from further analysis.

In mouse *Atp2b2*, PAPERCLIP identified an increased usage of the distal 3′ UTR sequence in adult (Figure 6E). Interestingly, the mouse distal poly(A) site identified by PAPERCLIP actually corresponds to the annotated poly(A) site in human, suggesting that the mouse annotation is incomplete (Figure 6D). To examine whether usage of the distal 3′ UTR sequence is also increased in adult in human, we mapped the RNA-seq raw reads from Jaffe et al. and compared the distal 3′ UTR-to-proximal 3′ UTR read count ratio between fetus and adult. This analysis indeed revealed a significant increase in distal 3′ UTR usage in human adult (Fisher's exact test, $p<0.01$)(Figure 6E).

To identify regulatory elements located in mouse *Atp2b2* distal 3′ UTR, we again overlaid the Ago HITS-CLIP data to search for miRNA seed matches in Ago footprints located in the distal 3′ UTR region. MicroRNAs that have additional seed matches in the proximal 3′ UTR region or upstream exons were excluded from further analysis. We identified a conserved miR-137 binding site near the end of human *ATP2B2* distal 3′ UTR (Figure 6D and 6F). The finding is highly intriguing because both *Atp2b2* and miR-137 are mainly expressed in neurons (Smrt et al., 2010; Strehler and Zacharias, 2001), and miR-137 is known to play roles in neuronal development (Meza-Sosa et al., 2014) and human hearing (Schultz et al., 2005). Therefore, we performed luciferase assay in HEK293 cells to examine functionality of the identified miR-137 binding site, which resulted in repression similar in magnitude to that of a previously validated miR-137 binding site in *Mib1* (Smrt et al., 2010) (Figure 6G). Taken together, these results demonstrate that the PAPERCLIP is able to provide critical information to guide the discovery of evolutionarily conserved APA shifts and uncover regulatory relationships between mRNAs and miRNAs.

DISCUSSION

In this study, we describe PAPERCLIP, an mRNA 3′ end mapping method based on CLIP technique. In contrast to the original HITS-CLIP (Licatalosi et al., 2008), which aims to identify transcriptome-wide binding sites of the protein of interest, PAPERCLIP uses PABP as a tool to specifically retrieve mRNA fragments containing 3′ UTR sequences immediately upstream of the poly(A) sites. This use of a robust biological filter for poly(A) sites distinguishes PAPERCLIP and other mRNA 3′ end mapping methods, which require *in silico* filtering or sophisticated molecular biology techniques to achieve selection against internal adenine-rich sequence. We provide evidence that PAPERCLIP has excellent signal-to-noise ratio, ameliorates the internal priming problem commonly seen in 3′ end sequencing, and compares favorably to many other mRNA 3′ end mapping methods. We further demonstrate the capacity of PAPERCLIP in two major applications for mRNA 3′ end mapping methods, detecting APA shifts and discovering 3′ UTR-mediated mRNA regulation.

By combining PAPERCLIP and siRNA knockdown in two cell-lines, we provided a comprehensive list of genes that exhibited APA shifts upon CFIm68 or CstF64/64tau depletion. In line with recent studies (Martin et al., 2012; Yao et al., 2012; 2013), our results suggest that CstF64 and CstF64tau, originally considered as “general” mRNA 3′ end-processing factors, only contribute to a subset of poly(A) site selection. Additionally, we discovered a previously unrecognized role of CstF64/64tau in promoting the usage of non-canonical poly(A) sites. We further identified a GUKKU motif that is enriched in these CstF64/64tau-sensitive non-canonical poly(A) sites and established its direct role in promoting poly(A) site usage. It is well-established that human CstF64 and CstF64tau favors G/U-rich sequence for interaction (Takagaki and Manley, 1997; Yao et al., 2013). We believe that the GUKKU motif likely represents one of many naturally occurring motifs in DSEs or CstF64/64tau-interaction sequences, which is supported by its presence in some well-known canonical poly(A) sites (Figure S6A). We do not exclude the possibility that additional sequence element(s) in the 58 non-canonical poly(A) sites might also participate in their selection and future studies will be necessary to fully characterize their DSEs, which we note will be greatly facilitated by the pEN2 construct described in this study.

The advantages of PAPERCLIP are further illustrated in our studies using brain, which harbors many long 3′ UTR extensions compared with other tissues. In addition to providing nucleotide-resolution annotations to many 3′ UTR extensions identified in a previous study (Miura et al., 2013), PAPERCLIP further expanded the list of 3′ UTR extensions in mouse brain. Moreover, PAPERCLIP identified additional miR-128 targets in adult mouse cortex and discovered a regulatory relationship between miR-137 and *Atp2b2* during mouse cortex development. It is known that miRNA binding sites tend to locate near alternative and constitutive polyadenylation sites (Grimson et al., 2007). Indeed, we found a strong enrichment of miRNA binding sites upstream of PAPERCLIP identified poly(A) sites in adult mouse cortex (Figure S6B), which further supports the strength of PAPERCLIP.

Recent studies in *Drosophila* revealed the presence of a tissue-specific gene regulatory network consists of neuron-specific miRNAs and neural 3′ UTR extensions in the adult brain (Figure S6C) (Hilgers et al., 2011; Smibert et al., 2012). Interestingly, an examination of publicly available RNA-seq data (ENCODE Project Consortium, 2012) showed that the main APA isoform of *ATP2B2/Atp2b2* in both human and mouse adult liver is also the proximal form (Figure S6D and data not shown). Therefore, although further studies are necessary to fully elucidate the functional significance of the potential regulation between *Atp2b2* and miR-137 in adult brain, our results provide an example of mammalian neural 3′ UTR extension that is both tissue-specific and developmentally controlled.

The identification of APA shift during mouse brain development by PAPERCLIP also provided an opportunity to examine whether APA affects gene expression in this context. We found that single poly(A) site genes as a group was enriched in the differentially expressed genes between embryo and adult brain (Figure S6E). In addition, we also found that the overlap between multi-poly(A) site genes that changed abundance and multi-poly(A) site genes that exhibited APA shift was not higher than what would be expected by chance (Figure S6F). These two findings are consistent with a recent work from Mayr and colleagues in which they discovered that APA shift and changes in mRNA abundance are largely separate processes in two models of celluar transformation (Lianoglou et al., 2013). Nevertheless, we do not rule out the possibility that a link between APA and gene expression is present in certain cell type(s) that is not detectable at the whole cortex level.

Extending insights into post-transcriptional regulation *in vivo*, particularly brain, is vastly complicated by the coexistence of multiple cell-types that have distinct functions. It has been shown that a subset of neuronal genes exhibit 3′ UTR extension during development in *Drosophila* (Hilgers et al., 2011; Smibert et al., 2012). It remains unclear whether an increase in use of distal 3′ UTR during development is limited to neurons or is a general property for all major cell types in brain. Using data from a recent cell type-specific RNA-seq study (Zhang et al., 2014), we identified 84 cell type-enriched genes from our list of 444 genes that significantly changed poly(A) site usage between embryos and adults (Table S6). Interestingly, although neuron-enriched genes are the largest class (39/84, 46%, including *Atp2b2*), five major brain cell types (astrocytes, neurons, oligodendrocytes, microglia and endothelial cells) have at least one gene that exhibits APA shift, suggesting that APA shift occurs in more than one cell type during mouse brain development (Zhang et al., 2014). It is therefore increasingly clear that cell type-specific poly(A) site maps will be necessary to understand how individual cell-type regulates APA during brain development. A key feature that distinguishes PAPERCLIP from other mRNA 3′ end mapping methods is the potential to modify the target cell population by restricting PABP expression. As technologies to achieve cell type-specific control of transgene expression *in vivo* are readily available (Utomo et al., 1999), PAPERCLIP is uniquely suited for the development of cell type-specific mRNA 3′ end mapping compared with other methods, which require prior isolation of pure cell types to establish cell type-specificity, a process that is laborious and may introduce bias such as stress signals (Okaty et al., 2011).

EXPERIMENTAL PROCEDURES

Statistical methods

Wilcoxon rank sum test, Fisher's exact test, hypergeometric test, and Student's t-test were performed using R or Excel. For APA shift, EdgeR package (Robinson et al., 2010) was used to statistically test significant APA shifts between two experimental conditions, while accounting for biological and technical variability between experimental replicates.

PAPERCLIP

Sample preparation, immunoprecipitation, SDS-PAGE and RNA extraction were adapted from standard HITS-CLIP (Moore et al., 2014). Mouse monoclonal anti-PABP (Sigma, P6246) was used for immunoprecipitation. The sequencing library was constructed using BrdU-CLIP method (Weyn-Vanhentenryck et al., 2014) with modification to improve sensitivity. The library contains a 14-nt degenerate linker sequence at the 5′ end (6-nt random barcode followed by 8-nt sample multiplexing index) or a 11-nt degenerate linker sequence at the 5′ end (3-nt degenerate sequence, 4-nt sample multiplexing index and 7-nt random barcode). The complete protocol is detailed in Supplemental Protocol. To minimize batch effects, the entire process was performed independently for replicate experiments, sometimes using primers with different indices. Individual PAPERCLIP libraries were multiplexed and sequenced by HiSeq 2000 or MiSeq (Illumina) to obtain 100-nt (HiSeq) or 75-nt (MiSeq) single-end reads.

Comparison to other 3′ end mapping methods

For Figure 1E and 1F, raw reads were downloaded from the NCBI Sequence Read Archive (SRR568012, SRR1033820, SRR299108, SRR090236, SRR453410 and SRR317197) and processed before mapping to hg19 (Novoalign, same settings as for PAPERCLIP). When necessary, poly(A) sequence at the 3′ end was trimmed using CutAdapt. For 3P-Seq and 3′ Seq, two methods that often include untemplated adenines in the sequencing reads, only trimmed reads were used for mapping. For all libraries, only uniquely mapped reads were used for analysis. Downstream adenine-rich reads were defined as in the literature (Derti et al., 2012; Shepard et al., 2011): reads that have 6 consecutive adenines or at least 7 adenines total in the 10 nucleotides immediately downstream of the last aligned position in the human genome. Reference BED files were downloaded from the UCSC genome browser for annotating the genomic localization of mapped reads.

Northern blots

Total RNA from HeLa cells was prepared by Trizol (Invitrogen) extraction and column purification using High Pure RNA Isolation Kit (Roche). 10μg total RNA per lane was separated with 0.8% SeaKem Gold agarose (Lonza) along with Millennium RNA Markers (Ambion). The gel was treated with 0.05M NaOH; 1.5M NaCl for 20 min., then with 0.5M Tris, pH7.5; 1.5M NaCl for 10 min., before being equilibrated for blotting in 20×SSC for 20 min. The gel was blotted onto Hybond-N filter (Amersham). Probes were prepared using Prime-It II Random Primer Labeling Kit (Agilent) with [α-^{32}P] dCTP. All primer sequences are listed in Table S11. For hybridization, labeled probes were used at 1 × 10^6cpm/ml of hybridization solution (1% BSA Fraction V, 7% SDS, 0.5M NaH_2PO_4, pH 7, 1mM EDTA, pH 8). The filter was hybridized overnight at 68°C and washed with 2×SSC; 0.05%SDS at 60°C 2x 10min., then with 0.1×SSC; 0.1%SDS at 60°C 2x 30min.

SDS-PAGE and western blots

20μg total protein per lane was separated on 10% Novex NuPAGE Bis-Tris gels (Invitrogen) and transferred to nitrocellulose membrane following standard procedures. The following antibodies are used for western blotting: rabbit polyclonal anti-CFIm68 (A301-356A, Bethyl Laboratories; 1:1000), rabbit polyclonal anti-CstF64 (A301-093A, Bethyl Laboratories; 1:1000), rabbit polyclonal anti-CstF64tau (A301-487A, Bethyl Laboratories; 1:1000), and mouse monoclonal anti-GAPDH (ab8245, Abcam; 1:2000).

Animal

Adult (8-14wk) and pregnant (E15.5) C57BL/6J mice were obtained from the Jackson Laboratory. Dissection of adult and embryonic brain cortex was performed following standard procedures. All procedures were conducted according to the Institutional Animal Care and Use Committee (IACUC) guidelines at the Rockefeller University.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Connie Zhao and Christine Lai at the Rockefeller University Genomics Resource Center for their support in high-throughput sequencing. We thank Nathalie Blachere, Jennifer Darnell, Shengdong Ke, Joseph Luna, Claudia Scheckel and Chaolin Zhang for comments and suggestions. H-W.H. is an HHMI Fellow of the Life Sciences Research Foundation. Y.S. is a JSPS Postdoctoral Fellow for Research Abroad. R.B.D. is an Investigator of the Howard Hughes Medical Institute. This work was also in part supported by grants from the National Institutes of Health (NS034389 and NS081706) and Simons Foundation (SFARI 240432) to R.B.D.

REFERENCES

Bruno IG, Karam R, Huang L, Bhardwaj A, Lou CH, Shum EY, Song H-W, Corbett MA, Gifford WD, Gecz J, et al. Identification of a MicroRNA that Activates Gene Expression by Repressing Nonsense-Mediated RNA Decay. Mol Cell. 2011; 42:500–510. [PubMed: 21596314]

Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 2009; 460:479–486. [PubMed: 19536157]

Darnell RB. RNA protein interaction in neurons. Annu. Rev. Neurosci. 2013; 36:243–270. [PubMed: 23701460]

Derti A, Garrett-Engele P, MacIsaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. A quantitative atlas of polyadenylation in five mammals. Genome Res. 2012; 22:1173–1183. [PubMed: 22454233]

Di Giammartino DC, Nishida K, Manley JL. Mechanisms and Consequences of Alternative Polyadenylation. Mol Cell. 2011; 43:853–866. [PubMed: 21925375]

Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet. 2013; 14:496–506. [PubMed: 23774734]

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007; 27:91–105. [PubMed: 17612493]

Hilgers V, Perry MW, Hendrix D, Stark A, Levine M, Haley B. Neural-specific elongation of 3′ UTRs during Drosophila development. Proc Natl Acad Sci USA. 2011; 108:15864–15869. [PubMed: 21896737]

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. Nat Methods. 2013; 10:133–139. [PubMed: 23241633]

Jaffe AE, Shin J, Collado-Torres L, Leek JT, Tao R, Li C, Gao Y, Jia Y, Maher BJ, Hyde TM, et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nature Neuroscience. 2015; 18:154–161. [PubMed: 25501035]

Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kühn U, Menzies FM, Oude Vrielink JAF, Bos AJ, Drost J, Rooijers K, et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell. 2012; 149:538–553. [PubMed: 22502866]

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci USA. 2009; 106:7028–7033. [PubMed: 19372383]

Kahvejian A, Roy G, Sonenberg N. The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. Cold Spring Harb Symp Quant Biol. 2001; 66:293–300. [PubMed: 12762031]

Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. 2013; 27:2380–2396. [PubMed: 24145798]

Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. Nat Rev Genet. 2010; 11:75–87. [PubMed: 20019688]

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008; 456:464–469. [PubMed: 18978773]

Martin G, Gruber AR, Keller W, Zavolan M. Genome-wide Analysis of Pre mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. Cell Reports. 2012; 1:753–763. [PubMed: 22813749]

Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature. 2014; 510:412–416. [PubMed: 24814343]

Meza-Sosa KF, Pedraza-Alva G, Pérez-Martínez L. microRNAs: key triggers of neuronal cell fate. Front Cell Neurosci 8. 2014

Miura P, Sanfilippo P, Shenker S, Lai EC. Alternative polyadenylation in the nervous system: to what lengths will 3′ UTR extensions take us? Bioessays. 2014; 36:766–777. [PubMed: 24903459]

Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. Widespread and extensive lengthening of 3′ UTRs in the mammalian brain. Genome Res. 2013; 23:812–825. [PubMed: 23520388]

Moore MJ, Zhang C, Gantman EC, Mele A, Darnell JC, Darnell RB. Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single- nucleotide resolution using HITS-CLIP and CIMS analysis. Nature Protocols. 2014; 9:263–293. [PubMed: 24407355]

Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, Chen J, Rowley JD, Wang SM. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. Proc Natl Acad Sci USA. 2002; 99:6152–6156. [PubMed: 11972056]

Nam J-W, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, Yildirim MA, Rodriguez A, Bartel DP. Global Analyses of the Effect of Different Cellular Contexts on MicroRNA Targeting. Mol Cell. 2014; 53:1031–1043. [PubMed: 24631284]

Nunes NM, Li W, Tian B, Furger A. A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. Embo J. 2010; 29:1523–1536. [PubMed: 20339349]

Okaty BW, Sugino K, Nelson SB. A Quantitative Comparison of Cell-Type- Specific Microarray Gene Expression Profiling Methods in the Mouse Brain. PLoS ONE. 2011; 6:e16493. [PubMed: 21304595]

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell. 2010; 143:1018–1029. [PubMed: 21145465]

Schultz JM, Yang Y, Caride AJ, Filoteo AG, Penheiter AR, Lagziel A, Morell RJ, Mohiddin SA, Fananapazir L, Madeo AC, et al. Modification of human hearing loss by plasma-membrane calcium pump PMCA2. N Engl J Med. 2005; 352:1557–1564. [PubMed: 15829536]

Shepard PJ, Choi E-A, Lu J, Flanagan LA, Hertel KJ, Shi Y. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. Rna. 2011; 17:761–772. [PubMed: 21343387]

Shi Y. Alternative polyadenylation: New insights from global analyses. Rna. 2012; 18:2105–2117. [PubMed: 23097429]

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL. Molecular architecture of the human pre-mRNA 3′ processing complex. Mol Cell. 2009; 33:365–376. [PubMed: 19217410]

Smibert P, Miura P, Westholm JO, Shenker S, May G, Duff MO, Zhang D, Eads BD, Carlson J, Brown JB, et al. Global Patterns of Tissue-Specific Alternative Polyadenylation in Drosophila. Cell Reports. 2012; 1:277–289. [PubMed: 22685694]

Smrt RD, Szulwach KE, Pfeiffer RL, Li X, Guo W, Pathania M, Teng Z-Q, Luo Y, Peng J, Bordey A, et al. MicroRNA miR-137 Regulates Neuronal Maturation by Targeting Ubiquitin Ligase Mind Bomb-1. Stem Cells. 2010; 28:1060–1070. [PubMed: 20506192]

Strehler EE, Zacharias DA. Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. Physiol. Rev. 2001; 81:21–50. [PubMed: 11152753]

Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. Mol Cell Biol. 1997; 17:3907–3914. [PubMed: 9199325]

Tan CL, Plotkin JL, Venø MT, Schimmelmann, von M, Feinberg P, Mann S, Handler A, Kjems J, Surmeier DJ, O'Carroll D, et al. MicroRNA-128 governs neuronal excitability and motor behavior in mice. Science. 2013; 342:1254–1258. [PubMed: 24311694]

Tian B, Manley JL. Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci. 2013; 38:312–320. [PubMed: 23632313]

Ulitsky I, Shkumatava A, Jan CH, Subtelny AO, Koppstein D, Bell GW, Sive H, Bartel DP. Extensive alternative polyadenylation during zebrafish development. Genome Res. 2012; 22:2054–2066. [PubMed: 22722342]

Utomo AR, Nikitin AY, Lee WH. Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice. Nat Biotechnol. 1999; 17:1091–1096. [PubMed: 10545915]

Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. 2005; 19:1315–1327. [PubMed: 15937220]

Wang L, Dowell RD, Yi R. Genome-wide maps of polyadenylation reveal dynamic mRNA 3′-end formation in mammalian cell lineages. Rna. 2013; 19:413–425. [PubMed: 23325109]

Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, et al. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Reports. 2014; 6:1139–1152. [PubMed: 24613350]

Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proc Natl Acad Sci USA. 2012; 109:18773–18778. [PubMed: 23112178]

Yao C, Choi E-A, Weng L, Xie X, Wan J, Xing Y, Moresco JJ, Tu PG, Yates JR, Shi Y. Overlapping and distinct functions of CstF64 and CstF64τ in mammalian mRNA 3′ processing. Rna. 2013; 19:1781–1790. [PubMed: 24149845]

Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, et al. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. J Neurosci. 2014; 34:11929–11947. [PubMed: 25186741]

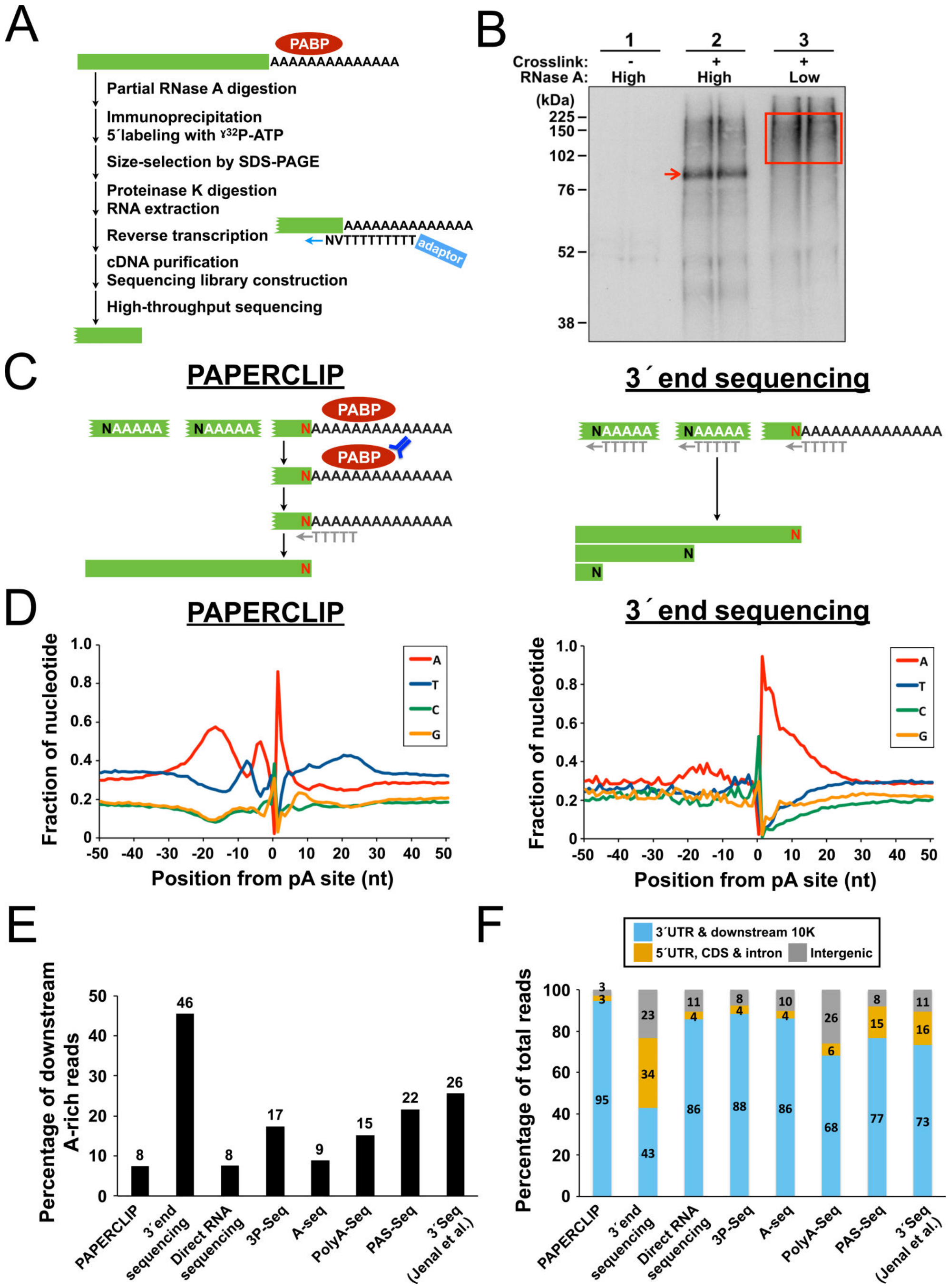

Figure 1.

PAPERCLIP ameliorates the internal priming problem commonly seen in 3′ end sequencing and compares favorably to other 3′ end mapping methods. (**A**) A diagram for the PAPERCLIP protocol. (**B**) An autoradiogram from PAPERCLIP experiment. Red arrow denotes the PABP-RNA complex. Red box shows the area of nitrocellulose membrane used for RNA extraction and subsequent sequencing library construction. (**C**) Schematics illustrating the difference between PAPERCLIP (left) and 3′ end sequencing (right). (**D**) Diagrams showing the genomic nucleotide sequence surrounding the putative poly(A) sites identified by PAPERCLIP (left) and 3′ end sequencing (right) in HeLa cells. (**E**) A bar graph showing the percentage of downstream adenine-rich reads for different 3′ end mapping methods. Data for PAPERCLIP and 3′ end sequencing are from the current study while data for other methods are obtained from NCBI. (**F**) A bar graph showing the percentage of reads mapped to different genomic regions. Data sources are the same as in (**E**).

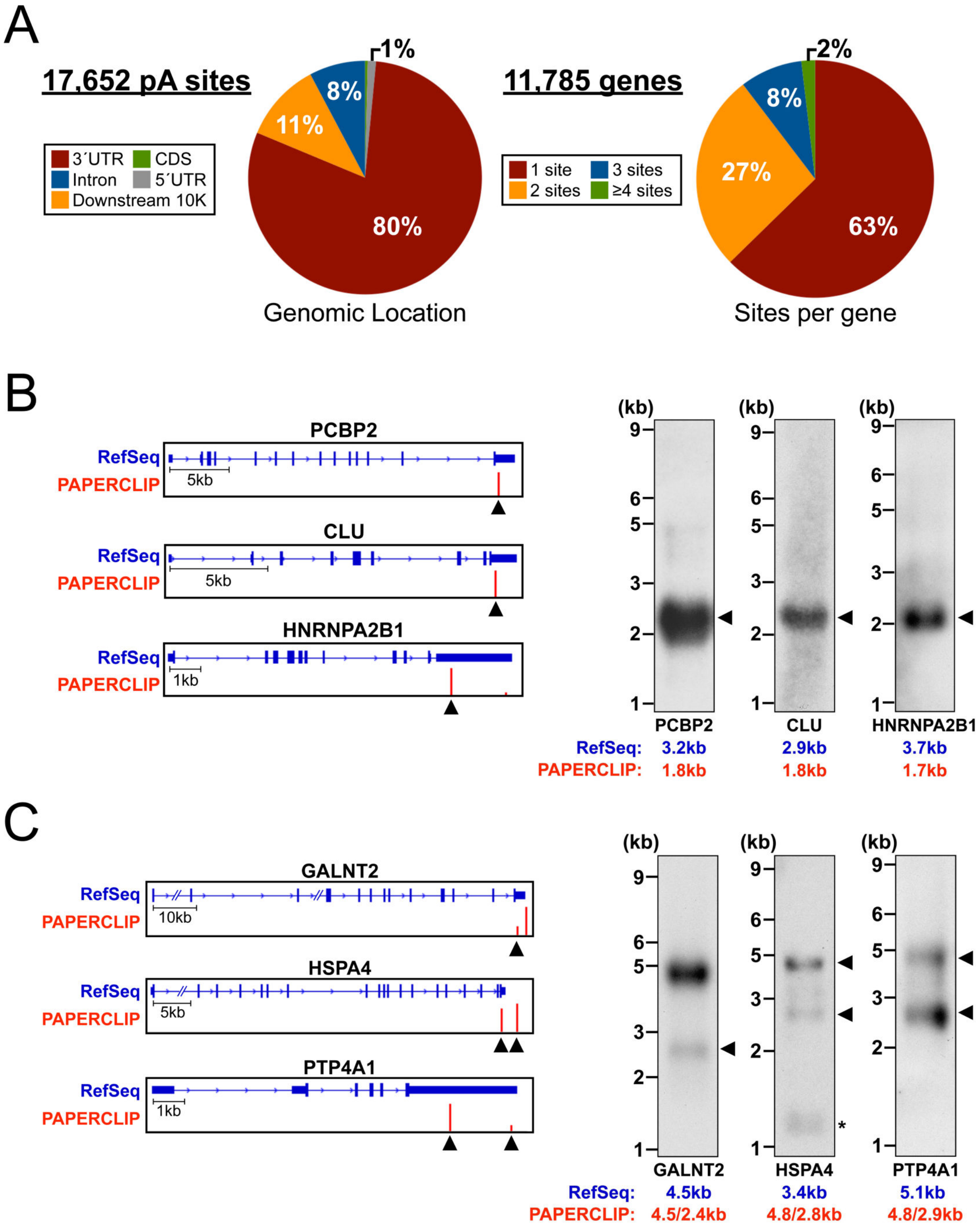

Figure 2.

PAPERCLIP depicts a high-resolution, comprehensive mRNA polyadenylation map of HeLa cells. (**A**) Pie-charts showing: the genomic location (left) and the per gene distribution (right) of poly(A) sites identified by PAPERCLIP in HeLa cells. (**B**) and (**C**) (left) Diagrams showing RefSeq annotation and PAPERCLIP results. The peak heights from PAPERCLIP are scaled for each gene. (right) Results from validating northern blot experiments. Numbers below denote the expected sizes of transcripts from RefSeq annotation and PAPERCLIP results. Arrowheads indicate poly(A) sites not present in the RefSeq annotation for the diagram and the corresponding mRNA isoforms for the northern blots.

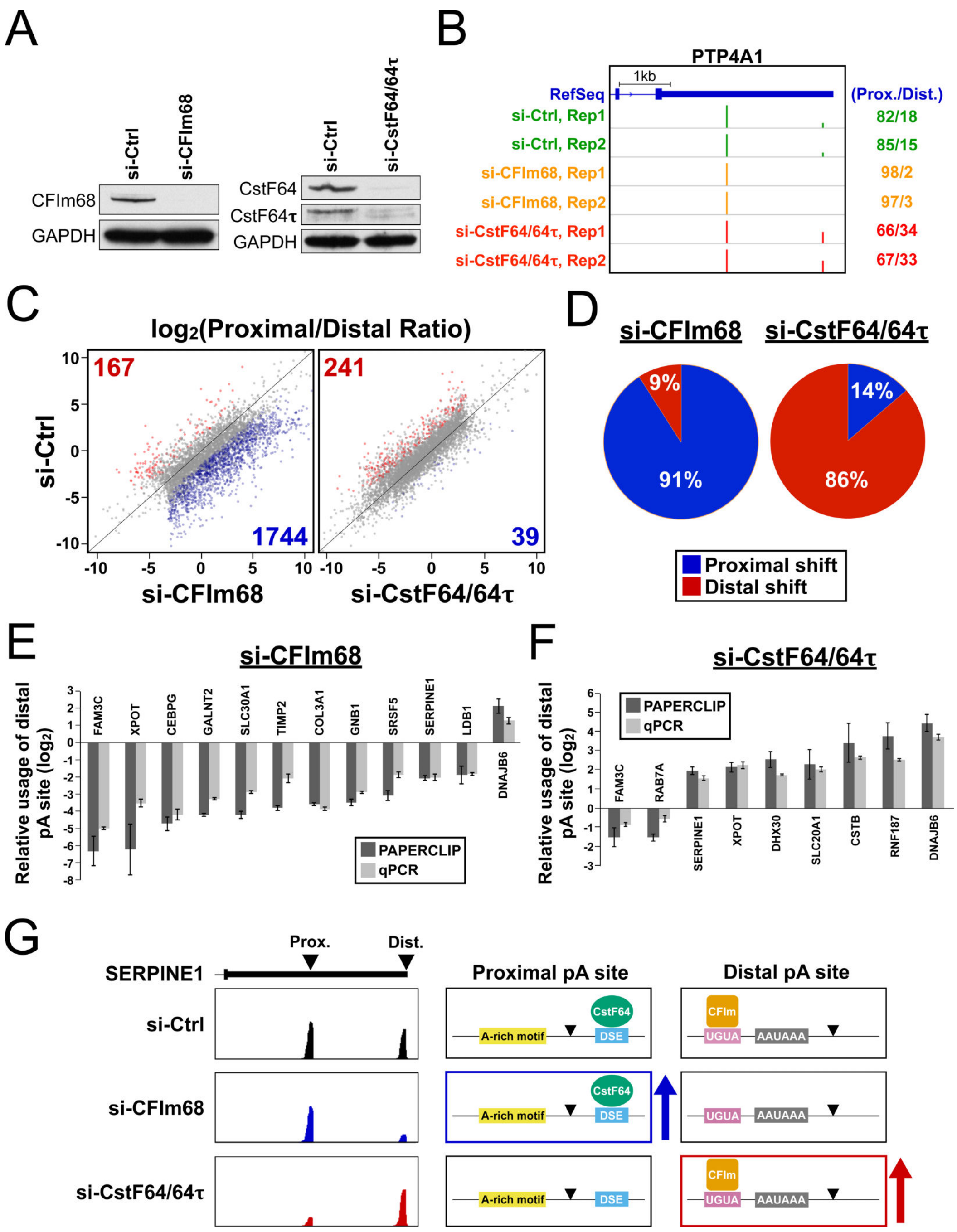

Figure 3.

PAPERCLIP discovers a role of CstF64/64tau in promoting non-canonical poly(A) site usage. (**A**) Western blots demonstrating siRNA knockdown of CFIm68 and CstF64/64tau in HeLa cells. (**B**) A diagram showing APA shifts in individual PAPERCLIP experiments following siRNA transfection. Relative peak height from PAPERCLIP at the proximal and distal poly(A) sites are shown as percentage of the sum of both peaks on the right. Rep, replicate. (**C**) Scatter plots comparing $\log_2$(proximal/distal ratio) by PAPERCLIP for 2-peak genes between control siRNAs (si-Ctrl) and treatment siRNAs (si-CFIm68 and si-

CstF64/64tau). Each dot represents a gene. Genes with FDR<0.05 and at least twofold change of P/D ratio are considered significantly shifted and are colored. Red, $\log_2$[(treatment siRNA P/D ratio)/(control siRNA P/D ratio)]≥1. Blue, $\log_2$[(treatment siRNA P/D ratio)/(control siRNA P/D ratio)]≤−1. Total numbers of significantly shifted genes for both directions are listed at the corners of plots. (**D**) Pie-charts summarizing the direction of APA shift in significantly shifted genes from (**C**). (**E**) and (**F**) Diagrams comparing PAPERCLIP results from (**C**) and validating qRT-PCR experiments for individual genes. Error bars represent standard errors. (**G**) Diagrams summarizing the sequence motifs correlated with APA in HeLa cells. SERPINE1 serves as an example. (left) Diagrams showing the last exon of SERPINE1, the location of proximal and distal poly(A) sites and PAPERCLIP read clusters in three experimental conditions. (middle and right) Diagrams showing the sequence motifs present in the flanking regions of SERPINE1 proximal and distal poly(A) sites in addition to the postulated interactions between them and CstF64/CFIm68. Black triangles denotes the poly(A) sites. Arrows denote the increase in poly(A) site usage. DSE, downstream element.

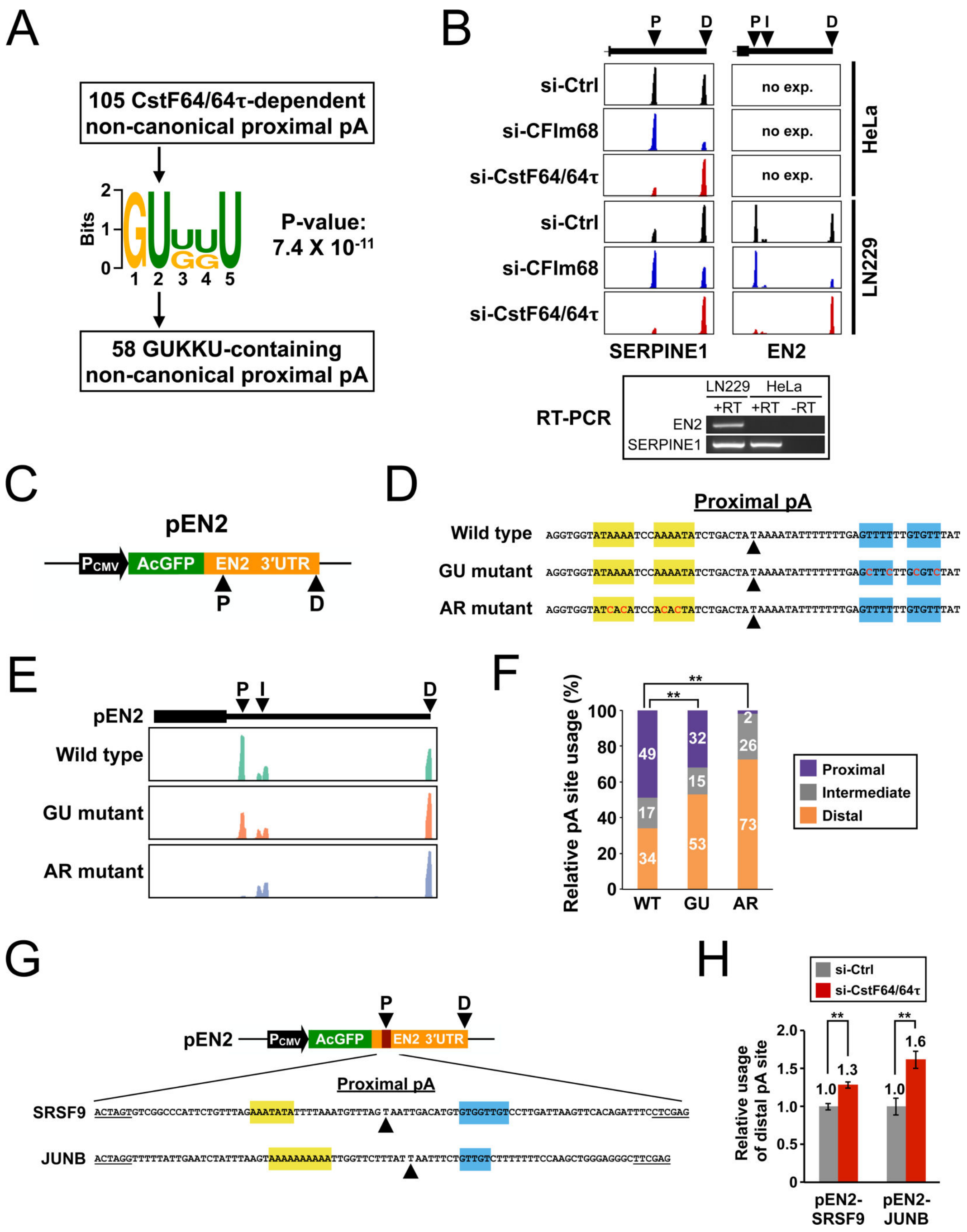

Figure 4.

PAPERCLIP identifies a GUKKU motif that contributes to non-canonical poly(A) site selection. (**A**) A diagram showing the identification of the GUKKU motif and GUKKU-containing non-canonical human poly(A) sites. (**B**) (upper) Diagrams showing the last exon of SERPINE1 and EN2, the location of proximal (P) and distal (D) poly(A) sites and PAPERCLIP read clusters in three experimental conditions in HeLa and LN229 cells. Black triangles denotes the poly(A) sites. I: two minor poly(A) sites in EN2. (lower) RT-PCR results showing the expression of SERPINE1 and EN2 in LN229 and HeLa cells. (**C**) A diagram showing the pEN2 construct. Black triangles denotes the poly(A) sites. (**D**) Diagrams showing the flanking nucleotide sequence of proximal poly(A) sites from pEN2 and two pEN2 mutants. Mutant nucleotide sequences are denoted in red. (**E**) Diagrams showing PAPERCLIP results from wild-type pEN2 and two mutants. Black triangles denotes the poly(A) sites. (**F**) Bar graphs summarizing poly(A) sites usage from two independent PAPERCLIP experiments. WT: wild-type; GU: GU mutant; AR: AR mutant. **: p<0.01. (**G**) Diagrams showing the nucleotide sequence of inserted poly(A) sites from SRSF9 and JUNB flanked by restriction sites for cloning (underline, destroyed in JUNB). Blue, GUKKU motif. Yellow, adenine-rich sequence. (**H**) Bar graphs showing qRT-PCR results from two independent siRNA experiments. Error bars represent standard errors. **: p<0.01.

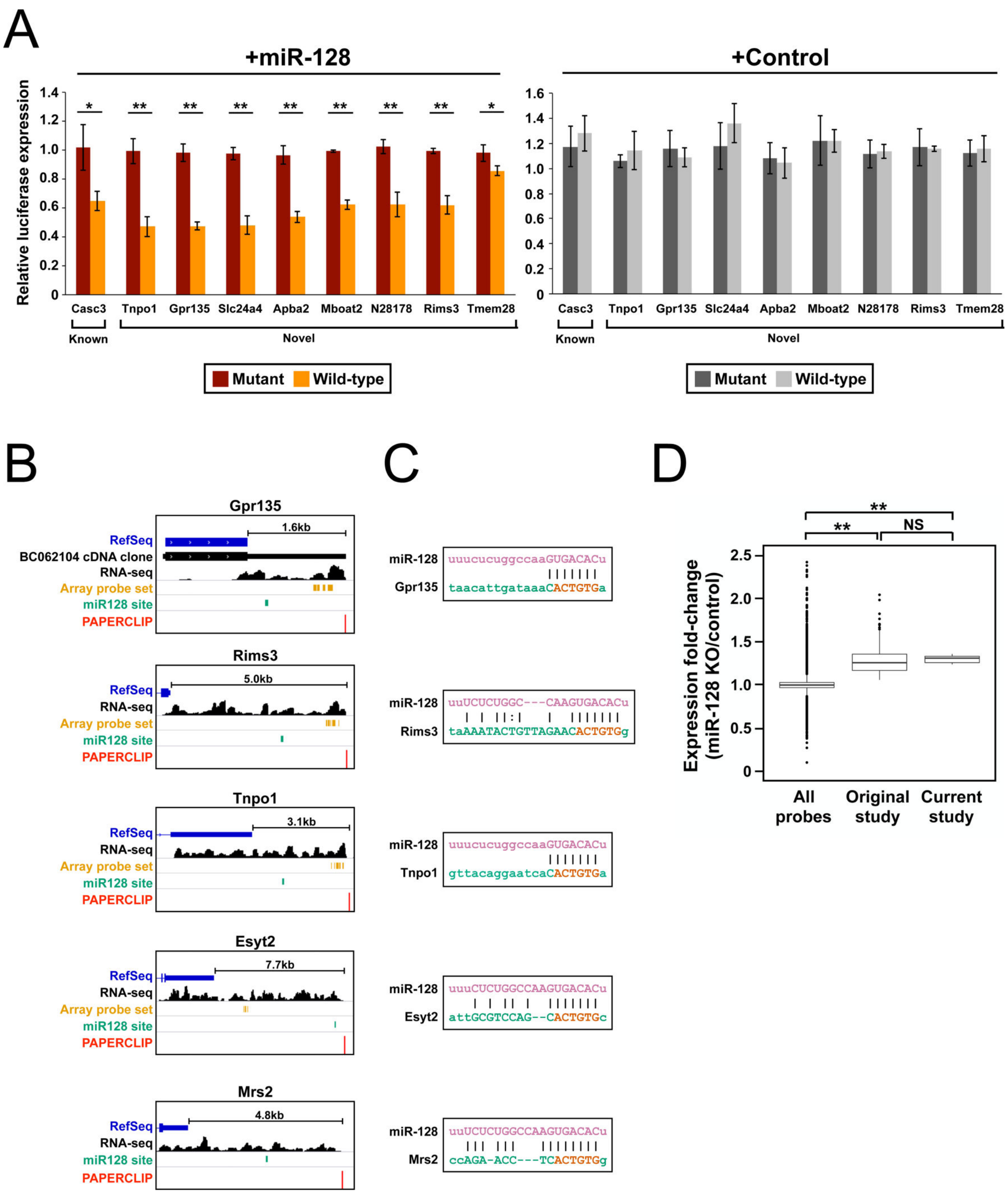

Figure 5.

PAPERCLIP identifies additional miR-128 targets. (**A**) Results of luciferase assay experiments in HEK293 cells co-transfected with miR-128 mimic (left) and control mimic (right) (n=3). Error bars represent standard deviation. *: p<0.05. **: p<0.01. Differences between mutant and wild-type for all constructs in the control mimic experiment are not statistically significant. (**B**) Diagrams showing RefSeq annotation, RNA-seq, PAPERCLIP results in addition to the locations of miR-128 binding site and the microarray probe set for the 5 miR-128 targets identified in current study. The peak heights from PAPERCLIP are scaled for each gene. The bar above RefSeq annotation indicates the length of 3′ UTR extension. For Gpr135, the poly(A) site identified by PAPERCLIP is at the 3′ end of a full-length cDNA clone BC062104, which is included in the diagram. The microarray probe set for Mrs2 is located 5′ to the last exon and therefore is not shown. For all five genes, no other miR-128 binding sites were found outside of the regions shown. (**C**) Diagrams showing the pairing between miR-128 and the 5 targets shown in (**B**) as determined by the miRanda algorithm. miR-128 sequence is shown from 3′ to 5′ and the target sequence is shown from 5′ to 3′. miR-128 seed sequence is shown in red. (**D**) Box plots showing the increased ribosome-association of 5 additional miR-128 targets (‘Current study’) in miR-128 deficient D1-neurons. All probes in the same dataset (‘All probes’) and the 154 miR-128 target genes identified in the original study (‘Original study’) (Tan et al. 2013) were shown for comparison. 6 outliers in the ‘All probes’ group were not shown. **: p<0.01. NS: not significant.

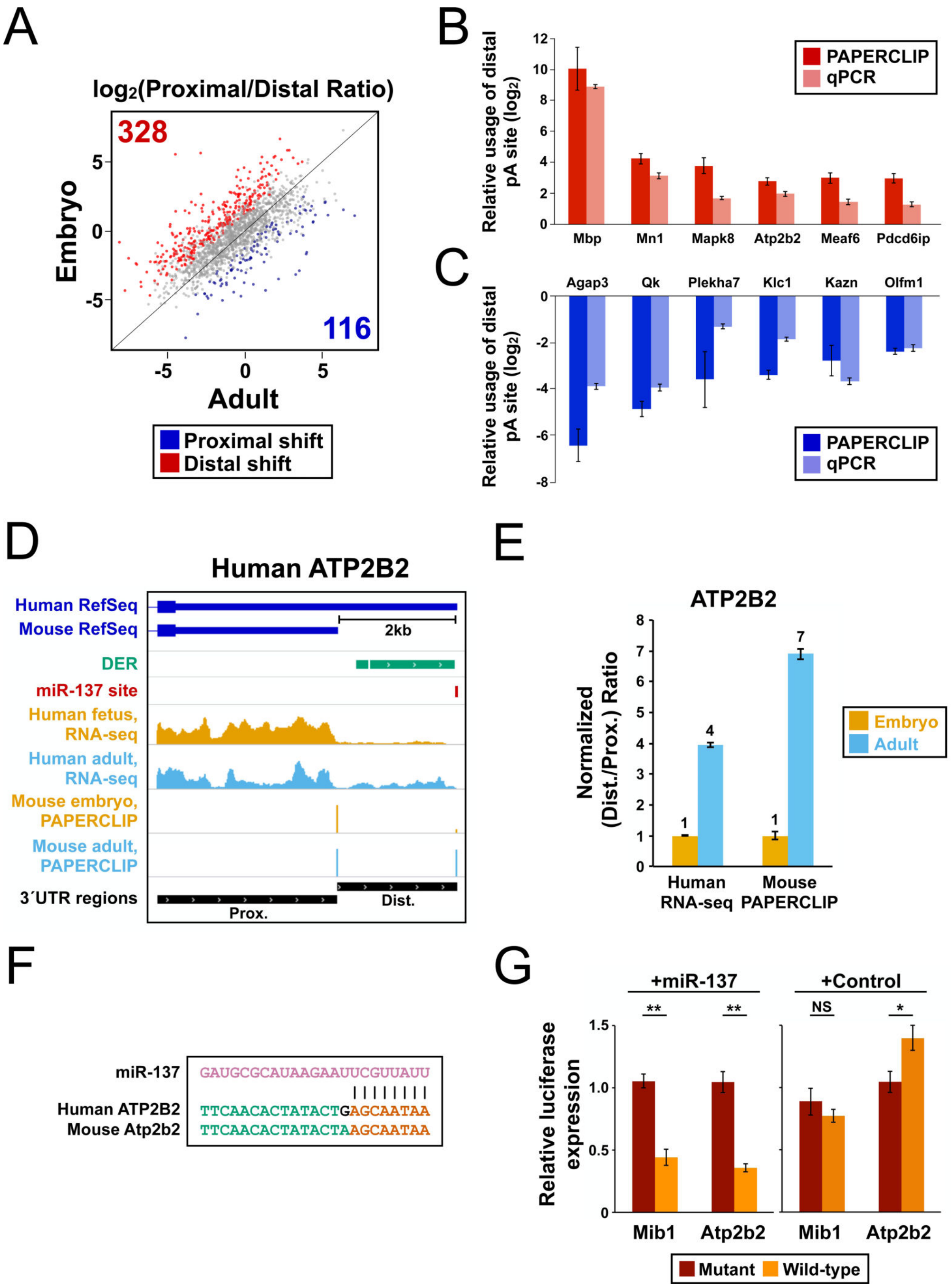

Figure 6.

PAPERCLIP identifies an evolutionarily conserved APA shift during brain development. (**A**) Scatter plots comparing $\log_2$(proximal/distal ratio) by PAPERCLIP for 2-peak genes between embryo and adult mouse cortex. Each dot represents a gene. Genes with FDR<0.05 and at least twofold change of P/D ratio are considered significantly shifted and are colored. Red, $\log_2$[(embryo P/D ratio)/(adult P/D ratio)]≥1. Blue, $\log_2$[(embryo P/D ratio)/(adult P/D ratio)]≤−1. Total numbers of significantly shifted genes for both directions are listed at the corners of plots. (**B**) and (**C**), Diagrams comparing PAPERCLIP results from (**A**) and validating qRT-PCR experiments for individual genes. Error bars represent standard errors. (**D**) Diagrams showing both human and mouse RefSeq annotation, the locations of DERs and the miR-137 site, human RNA-seq results from Jaffe et al., and mouse PAPERCLIP results for human *ATP2B2*. Mouse RefSeq annotation and PAPERCLIP results were lifted from mm10 to hg19. (**E**) A bar graph showing the relative usage of distal 3′ UTR in ATP2B2 gene for both human and mouse. The mouse PAPERCLIP results were re-plotted from (**B**) for comparison. Error bars represent standard errors. (**F**) Diagrams showing the pairing between miR-137 and the target site in ATP2B2 gene in both human and mouse. miR-128 sequence is shown from 3′ to 5′ and the target sequence is shown from 5′ to 3′. miR-137 seed sequence is shown in red. (**G**) Results of luciferase assay experiments in HEK293 cells co-transfected with miR-137 mimic (left) and control mimic (right) (n=3). Error bars represent standard deviation. *: p<0.05. **: p<0.01. NS: not significant.