Published in final edited form as:

Nat Methods. 2019 December ; 16(12): 1275–1280. doi:10.1038/s41592-019-0570-0.

DART-seq: an antibody-free method for global m^{6}A detection

Kate D. Meyer[*,1,2]

[1]Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710

[2]Department of Neurobiology, Duke University School of Medicine, Durham, NC 27710

Abstract

m^{6}A is a widespread RNA modification which influences nearly every aspect of the mRNA life cycle. Our understanding of m^{6}A has been facilitated by the development of global m^{6}A mapping methods, which use antibodies to immunoprecipitate methylated RNA. However, these methods have several limitations, including high input RNA requirements and cross-reactivity to other RNA modifications. Here, we present DART-Seq (deamination adjacent to RNA modification targets), an antibody-free method for detecting m^{6}A sites. In DART-Seq, the cytidine deaminase APOBEC1 is fused to the m^{6}A-binding YTH domain. APOBEC1-YTH expression in cells induces C to U deamination at sites adjacent to m^{6}A residues, which are detected using standard RNA-Seq. DART-Seq identifies thousands of m^{6}A sites in cells from as little as 10 nanograms of total RNA and can detect m^{6}A accumulation in cells over time. Additionally, we use long-read DART-Seq to gain new insights into m^{6}A distribution along the length of individual transcripts.

Introduction

m^{6}A is the most abundant internal mRNA modification and plays diverse roles in RNA regulation. Recently, m^{6}A has emerged as an important regulator of a variety of physiological processes[1, 2]; thus, detecting m^{6}A sites in cells is critical for understanding how this modification impacts gene expression to contribute to cellular function and disease states.

To date, most methods for global m^{6}A detection have relied on immunoprecipitation of methylated RNAs using m^{6}A-recognizing antibodies in a technique called MeRIP-Seq[3] or m^{6}A-Seq[4]. Subsequent improvements to this method have come with the addition of UV crosslinking steps to identify m^{6}A sites at single-nucleotide resolution[5, 6]. Although these methods have yielded unprecedented insights into the location and regulation of m^{6}A in cellular RNAs, they suffer from several limitations. First, they require large amounts of input RNA, which makes global m^{6}A detection prohibitive for limited-quantity samples. Second,

[*]correspondence to: kate.meyer@duke.edu.

Author Contributions

K.D.M conceived of the project, collected and analyzed the data, and wrote the manuscript.

Competing Interests

The author declares no competing financial interests.

m^{6}A antibodies also recognize the structurally similar cap modification, m^{6}Am, so immunoprecipitation of methylated RNAs does not exclusively enrich for m^{6}A-containing RNA. Finally, antibody-based approaches are costly and the associated library preparation steps are time-consuming, which can be major limiting factors for many experiments. Thus, there is a great need for a simple, sensitive, antibody-free method for global m^{6}A detection.

We reasoned that a strategy which alters the sequence near methylation sites would enable m^{6}A detection by standard RNA-seq and thus overcome the major limitations of current methods. APOBEC1 is a cytidine deaminase which targets DNA and RNA to induce cytidine to uridine (C to U) editing[7]. Although initially discovered for its ability to edit the *ApoB* mRNA, APOBEC1 has since been utilized in CRISPR/Cas9-based genome editing approaches to induce C to U conversion at targeted single-stranded DNA sites[8]. We speculated that a similar strategy could be used to edit m^{6}A-adjacent cytidines in RNAs by fusing APOBEC1 to the m^{6}A-binding YTH domain and detecting subsequent editing events with RNA-Seq. Here, we demonstrate the utility of this approach for detecting m^{6}A sites in cellular RNAs using transcriptome-wide mapping with as little as 10 nanograms of total RNA as input. Our strategy performs similarly to antibody-based approaches for methylated RNA detection and provides new insights into clustering of m^{6}A residues within individual transcript isoforms. This approach substantially improves the time and cost associated with global m^{6}A detection and will enable transcriptome-wide mapping in limited RNA samples.

Results

Development of an antibody-free method for m^{6}A detection

The preferred consensus sequence for m^{6}A contains an invariable cytidine residue immediately following the m^{6}A site (R$\underline{\text{m}^6\text{A}}$CH, where R = A or G; H=A, C, or U)[3–5, 9–12]. Thus, we speculated that recruitment of APOBEC1 to m^{6}A sites would enable deamination of the cytidine immediately following m^{6}A residues. To test this, we fused APOBEC1 to the m^{6}A-binding YTH domain of YTHDF2[13, 14] (Fig. 1). The APOBEC1-YTH fusion protein was then incubated with a synthetic RNA containing a single internal adenosine. Reverse transcription and Sanger sequencing indicated frequent editing of the cytidine immediately following m^{6}A in methylated RNA, but not in unmethylated RNA (Supplementary Fig. 1a).

To confirm that the observed editing was caused by targeting of APOBEC1 to the m^{6}A residue, we repeated the *in vitro* deamination assays using a mutant version of the APOBEC1-YTH fusion protein (APOBEC1-YTH^{mut}) in which the m^{6}A binding region of the YTH domain was deleted (Supplementary Fig. 1b). APOBEC1-YTH^{mut} was impaired in its ability to bind m^{6}A (Supplementary Fig. 1c) and failed to convert adjacent cytidines to uridines in m^{6}A-containing RNA (Supplementary Fig. 1a), indicating that the deaminase activity of APOBEC1-YTH is directed by the m^{6}A-binding activity of the YTH domain.

DART-Seq enables transcriptome-wide detection of m^{6}A

We next sought to determine whether APOBEC1-YTH could be used to detect endogenous m^{6}A sites in cells. To do this, we developed DART-Seq (deamination adjacent to RNA modification targets), in which we introduced APOBEC1-YTH into cells and then subjected

total RNA to next-generation sequencing followed by C to U mutation detection (Fig. 1, Methods). Comparison of three biological replicates indicated high reproducibility in C to U mutations in APOBEC1-YTH-expressing HEK293T cells (Supplementary Fig. 2), suggesting that APOBEC1-YTH targets specific RNAs for editing with high consistency across samples.

To determine whether DART-Seq can identify m^{6}A residues, we compared C to U editing sites from cells expressing APOBEC1-YTH and cells expressing APOBEC1 alone. DART-Seq editing events from APOBEC1-YTH-expressing cells occurred primarily in the 3'UTR and coding sequence (CDS) (Fig. 2a) and were enriched in the vicinity of the stop codon (Fig. 2b, Supplementary Fig. 3a), which mirrors the distribution of m^{6}A[3, 4]. In contrast, editing events from cells expressing APOBEC1 alone were located primarily in 3'UTRs and intergenic regions and failed to show an enrichment near the stop codon (Fig. 2b, Supplementary Fig. 3b). Furthermore, there was little overlap in C to U editing between the two datasets, as 96% of edited sites from APOBEC1-YTH-expressing cells were not detected in cells expressing APOBEC1 alone (56,603 out of 59,246 sites) (Supplementary Table 1). To further ensure that C to U editing was caused by recruitment of APOBEC1-YTH to m^{6}A, we performed RNA-seq on HEK293T cells expressing APOBEC1-YTH^{mut} and carried out the same C to U editing analysis (Supplementary Table 1). C to U editing events in cells expressing APOBEC1-YTH^{mut} showed a distinct distribution compared to those in cells expressing APOBEC1-YTH, characterized by an enrichment throughout the 3'UTR as opposed to in the vicinity of the stop codon (Fig. 2a, Supplementary Fig. 3a,c). Together, these results suggest that the specificity of APOBEC1-YTH editing throughout the transcriptome depends on its ability to bind m^{6}A.

To obtain a set of high-confidence editing sites, we filtered our list of APOBEC1-YTH sites to include only those with at least a 1.5-fold enrichment over APOBEC1-YTH^{mut} samples. We also excluded all naturally occurring C to U mutations in HEK293T cells, as well as C to U editing sites detected in cells expressing APOBEC1 alone (**see** Methods). This resulted in a list of 100,636 C to U editing sites in 9,793 RNAs that occurred in at least 5% of all reads. Of these, a stringent list of 40,263 editing events in 7,707 RNAs was observed in at least 10% of all reads (Table 1).

Examination of sequences immediately surrounding DART-Seq sites revealed enrichment of a GGACU-containing motif, which matches the preferred consensus sequence for m^{6}A (Fig. 2c). In contrast, motifs detected in APOBEC1-YTH^{mut} and APOBEC1 samples did not match the m^{6}A consensus (Fig. 2c). Furthermore, DART-Seq sites were highly enriched within 3'UTRs and in the vicinity of the stop codon, as well as within long internal exons (Figs. 2b, Supplementary Fig. 3a,d), which matches the distribution of m^{6}A[3, 4]. Comparison of methylated RNAs detected by MeRIP-Seq[3] and those identified by DART-Seq showed a high degree of overlap, with 64% of m^{6}A-containing RNAs detected by DART-Seq (3,679 of 5,768 RNAs). Examination of individual RNAs showed that DART-Seq editing events occurred at sites of MeRIP-Seq enrichment (Fig. 2d, Supplementary Fig. 4). Furthermore, consistent with our *in vitro* deamination assays, we found that C to U editing events frequently occurred immediately downstream of known m^{6}A sites in cellular RNAs (Fig. 2e).

We next assessed the ability of DART-Seq to identify individual m^{6}A sites compared to antibody-based approaches. Comparison of global m^{6}A profiling datasets obtained by m^{6}A immunoprecipitation[3, 4, 15] showed that DART-Seq performs similarly in its ability to detect m^{6}A sites (Supplementary Fig. 5). We also observed an enrichment of DART-Seq editing adjacent to m^{6}A sites identified by single-nucleotide resolution m^{6}A profiling (miCLIP/m^{6}A-Seq)[5, 6] (Supplementary Fig. 6a,b). Additionally, the majority (91.4%) of C to U editing sites in APOBEC1-YTH-expressing cells are preceded by an A, compared to only 67.9% of C to U editing sites in cells expressing APOBEC1 alone (Supplementary Fig. 6c), suggesting that APOBEC1-YTH deamination is directed specifically toward cytidines adjacent to m^{6}A and that promiscuous editing of non-adjacent cytidines is rare. Further support for this comes from the finding that over 90% of DART-Seq sites are greater than 10 nucleotides away from the closest editing event, which is similar to the distribution seen in miCLIP (Supplementary Fig. 5c). Collectively, these data indicate that DART-Seq is capable of detecting m^{6}A sites in cellular RNAs transcriptome-wide.

Validation of the DART-Seq approach

To validate individual DART-Seq sites, we performed RT-PCR and Sanger sequencing to determine whether C to U editing occurs adjacent to m^{6}A sites previously quantified by miCLIP[5] or by SCARLET[16], another single-nucleotide resolution m^{6}A identification method. We confirmed the presence of editing adjacent to known m^{6}A sites in the *BSG* and *ACTB* mRNAs in cells expressing APOBEC1-YTH but did not observe robust editing in cells expressing APOBEC1 alone (Supplementary Fig. 7). To further validate that DART-Seq editing depends on the presence of m^{6}A, we performed DART-Seq using HEK293T cells depleted of the m^{6}A methyltransferase, METTL3 (Supplementary Fig. 8). METTL3-depleted cells exhibited fewer DART-Seq editing events in general and loss of the GGACU m^{6}A consensus sequence surrounding DART-Seq sites (Supplementary Fig. 9a, Supplementary Table 2). Furthermore, 97% of the DART-Seq sites detected in wild type cells were lost in METTL3-depleted cells (Supplementary Fig. 9b–e). These results further confirm that DART-Seq editing depends on m^{6}A.

DART-Seq enables low-input global m^{6}A profiling

One of the biggest challenges for global m^{6}A detection has been the large amount of input RNA required for effective immunoprecipitation and sequencing. Recent advances in library preparation have provided important improvements, with some studies reporting m^{6}A profiling using as little as 150 ng of mRNA or 500 ng of total RNA[17–19]. However, even with such improvements, the requirement for high nanogram amounts of poly(A) or rRNA-depleted RNA can be limiting for certain cell or tissue types.

We therefore sought to determine whether DART-Seq could be used to detect m^{6}A in low-input RNA samples. Using as little as 10 nanograms of total RNA as input, we detected over 79% of the DART-Seq edited mRNAs that were identified in our high-input DART-Seq library (Supplementary Fig. 10a,b, Supplementary Table 3). Low-input DART-Seq samples perform similarly to antibody-based approaches for m^{6}A detection, albeit with slightly reduced efficiency compared to high-input DART-Seq samples (Supplementary Fig. 10c). In addition, low-input DART-Seq sites are enriched for m^{6}A consensus motifs and near the 5’

end of the 3'UTR (Supplementary Fig. 10d,e). Thus, DART-Seq is capable of detecting m^{6}A sites from as little as 10 nanograms of total RNA.

m^{6}A detection using *in vitro* DART-Seq

APOBEC1-YTH-expressing cells exhibit normal levels of genes in the m^{6}A regulatory pathway and show no alterations in cell viability, suggesting that prolonged APOBEC1-YTH expression does not alter the m^{6}A landscape (Supplementary Fig. 11, Supplementary Table 4). Nevertheless, APOBEC1-YTH overexpression may not be possible or desirable in some cases, which would necessitate the use of *in vitro* deamination to perform DART-Seq. To test the ability of this approach to detect m^{6}A in cellular RNA, we performed *in vitro* DART-Seq using HEK293T cell RNA. We detected C to U editing at known m^{6}A sites, and global analyses revealed a distribution and motif enrichment similar to that of m^{6}A (Supplementary Figs. 12, 13). Although the majority (91%) of methylated mRNAs identified with *in vitro* DART-Seq were also identified using cellular DART-Seq, *in vitro* DART-Seq identified fewer methylated mRNAs than cellular DART-Seq, suggesting reduced efficiency (Supplementary Table 5). Thus, *in vitro* DART-Seq can reliably mark m^{6}A sites in cellular RNAs, although this approach will likely benefit from further optimization to increase identification of low-abundance m^{6}A sites.

DART-Seq distinguishes m^{6}A from m^{6}Am

A limitation of antibody-based m^{6}A detection strategies is cross-reactivity of m^{6}A antibodies with m^{6}Am[4, 5, 12, 20]. Hydrogen bonding between the YTH domain and the 2'-OH of m^{6}A suggests that the YTH domain used in DART-Seq may not recognize m^{6}Am[13, 21, 22]. Furthermore, unlike m^{6}A residues, m^{6}Am is not invariably following by a cytidine[5], which means that detection of m^{6}Am by APOBEC1-YTH would require deamination of cytidines further away from the modified base. We therefore wondered whether DART-Seq could be used to distinguish m^{6}A from m^{6}Am.

To investigate this, we compared a list of m^{6}Am sites[20] in HEK293 cells to DART-Seq datasets. Since m^{6}A sites in 5'UTRs may actually reflect m^{6}Am residues at misannotated start sites[5], we extended our DART-Seq sites to include regions 4 nt up- and downstream from the C to U editing site. We found only one RNA with overlap between extended DART-Seq sites and m^{6}Am sites. Upon closer examination, the DART-Seq editing site in this transcript is diminished in METTL3-depleted cells and is found internally within the 5'UTR, suggesting that it is not an m^{6}Am site (Supplementary Fig. 14). Thus, we conclude that DART-Seq does not recognize m^{6}Am and can be used to identify m^{6}A residues independently of m^{6}Am residues.

Estimation of m^{6}A abundance

Determining m^{6}A abundance within individual RNAs has been a major challenge to RNA methylation research. SCARLET enables quantitative measures of m^{6}A in individual RNAs[16], but this approach is not amenable to transcriptome-wide measurements. m^{6}A-LAIC-Seq[23] uses immunoprecipitation of full-length transcripts to estimate methylation levels of individual mRNAs, but it does not account for multiple m^{6}A sites or the presence

of m^{6}Am. Finally, peak over input (POI) can be used in MeRIP-Seq, but these measures provide only a rough estimate of m^{6}A abundance[3, 12].

We speculated that the degree of methylation may correlate with APOBEC1-YTH binding and C to U editing to enable global estimates of m^{6}A abundance in individual transcripts. To test this, we performed *in vitro* deamination assays using RNA with various amounts of m^{6}A. We found that C to U editing was positively correlated with m^{6}A levels at individual sites within an RNA (Supplementary Fig. 15a,b). Examination of DART-Seq editing of cellular RNAs also showed a positive relationship between m^{6}A abundance and editing efficiency (Supplementary Fig. 7a) Thus, DART-Seq can be used as an indicator of m^{6}A abundance in individual RNAs.

DART-Seq identifies m^{6}A accumulation in cellular RNAs

We next sought to determine whether changes in m^{6}A can be detected by DART-Seq. Previous studies have shown that treatment of cells with moderate concentrations of the topoisomerase inhibitor camptothecin (CPT) causes slowed transcription and an increase in m^{6}A abundance in the CDS[24]. We treated HEK293T cells expressing APOBEC1-YTH with CPT for 5 h and performed DART-Seq to identify m^{6}A sites. This led to 6,258 C to U sites that showed at least a 2-fold increase in editing compared to untreated cells (Supplementary Table 6). Metagene analysis of these sites indicated a slight enrichment in the CDS compared to untreated cells (Fig. 3a), and examination of individual mRNAs confirmed this analysis (Fig. 3b,c). We validated the increase in m^{6}A within the CDS of select mRNAs using m^{6}A immunoprecipitation followed by RT-qPCR (MeRIP-RTqPCR) (Fig. 3d). We also found that increased m^{6}A in these RNAs negatively correlated with their abundance (Fig. 3e), similar to what has been previously observed[24]. Thus, DART-Seq can be used to detect accumulation of m^{6}A in individual RNAs in response to changing cellular conditions.

Long-read DART-Seq reveals isoform-specific methylation patterns

Immunoprecipitation-based m^{6}A detection strategies have previously reported clustering of m^{6}A sites[3–5]. However, it remains unknown whether this reflects clustering of m^{6}A on the same or distinct RNA molecules. Since DART-Seq induces editing events in single transcripts, we reasoned that individual sequencing reads could be examined to determine whether m^{6}A sites are found in the same RNA molecule. To investigate this, we performed long-read DART-Seq using the PacBio platform. Examination of individual mRNAs showed that, although some transcripts exhibit isoform-specific regional editing, others contain DART-Seq sites in the 5’UTR, CDS, and 3’UTR (Fig. 4). In addition, 41% of reads spanning at least two editing sites contain two or more C to U editing events. These data suggest that the majority of individual RNA molecules have just one m^{6}A site, but that many RNAs harbor multiple sites, which is consistent with previous reports from isoform-specific m^{6}A immunoprecipitation (m^{6}A-LAIC-seq)[23]. Further studies will be needed to understand whether distinct m^{6}A residues on the same transcript work in a coordinated or competing manner. Additionally, although our data suggest that multiple C to U editing events caused by the same m^{6}A site are rare (Supplementary Fig. 5c, Supplementary Fig. 6), studies of clustered m^{6}A sites in individual transcripts may benefit from additional validation using miCLIP or SCARLET.

Discussion

A major challenge in global m^{6}A profiling has been the large amount of input RNA needed for m^{6}A detection. We find that DART-Seq can identify m^{6}A residues in cellular mRNAs using as little as 10 nanograms of total RNA. This is a substantial improvement over current m^{6}A detection methods which will undoubtedly facilitate global studies of m^{6}A in limited samples. Additionally, DART-Seq can be performed using standard RNA-seq library preparation methods, thus substantially reducing the time and cost associated with traditional antibody-based detection strategies. The versatility and sensitivity of DART-Seq also suggest the possibility of using this approach for m^{6}A detection in distinct cell types, and we envision that DART-Seq can potentially be coupled with single cell isolation and library preparation methods to achieve single-cell m^{6}A detection.

We find that DART-Seq marks more sites than antibody-based approaches, which could speak to the differences between the two techniques: while m^{6}A immunoprecipitation takes a snapshot of m^{6}A at any given time and is prone to accessibility of the antibody to m^{6}A during immunoprecipitation, DART-Seq irreversibly marks m^{6}A sites in cells over several hours. Thus, sites which would be otherwise buried in structure may be briefly accessible under physiological conditions and edited by APOBEC1-YTH. A potential future application of DART-Seq is to express APOBEC1-YTH in animals to enable m^{6}A profiling in cell types of interest during distinct physiological states.

Modifications to the YTH domain that improve its affinity for m^{6}A will likely enable more precise and sensitive m^{6}A detection using the DART-Seq approach. Similarly, DART-Seq can potentially be used to detect other RNA modifications by fusing APOBEC1 to small proteins engineered to bind them. Finally, fusion of localization elements to APOBEC1-YTH could potentially facilitate compartmentalized m^{6}A detection. For instance, forcing a nuclear, cytoplasmic, or mitochondrial localization could enable selective detection of methylated RNAs residing in distinct cellular compartments.

Online Methods

Antibodies

The following antibodies and concentrations were used: rabbit anti-HA (Cell Signaling; 3724S; 1:1000), rabbit anti-m^{6}A (Abcam; ab151230; 1:1000), HRP-conjugated goat anti-rabbit (Abcam; ab6721; 1:2500), HRP-conjugated sheep anti-mouse (GE Healthcare; 95017-554; 1:2500), mouse anti-βactin (Genscript; A00702; 1:5000), rabbit anti-METTL3 (Abcam; ab195352; 1:1000), rabbit anti-cleaved caspase 3 (Proteintech; 25546-1-AP, 1:1000), AlexaFluor 488-conjugated goat anti-rabbit (Thermo-Fisher; A-21206; 1:1000).

Constructs

YTH-HA was synthesized as a gene fragment (IDT), and YTH-HA or YTH^{mut}-HA were subsequently amplified using the YTH Fwd/YTH-HA Rev or YTH^{mut} Fwd/YTH-HA Rev primers (below). The YTH-HA sequence comprised amino acids 385-579 of human YTHDF2 fused at its C-terminal end to the HA tag (YPYDVPDYA). The YTH^{mut}-HA fusion lacked amino acids 385-409 comprising the m^{6}A-binding region. These YTH-HA

fusions were then inserted downstream of the rat APOBEC1 editing domain (APOBEC1) in the pCMV-APOBEC1 plasmid (a gift from David Liu; Addgene plasmid #73019; http://n2t.net/addgene:73019) using the XmaI and PmeI restriction sites. A 15 amino acid linker is present between the APOBEC1 domain and the YTH domain.

Cells

HEK293T cells (ATCC) were cultured at 37°C using DMEM supplemented with 10% FBS. METTL3-depleted cell lines were generated by cloning a *METTL3*-targeting sgRNA sequence (5'- GGAGTTGATTGAGGTAAAGCG -3') into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (a gift from Feng Zhang; Addgene plasmid # 62988; http://n2t.net/addgene:62988). Plasmids were then transfected into HEK293T cells and stable cells were selected with puromycin. Validation of METTL3 depletion was performed with western blot, RNA-Seq, and m^{6}A immunoblotting. Camptothecin treatment was carried out for 5 hours at 37°C using 6μM final concentration of camptothecin (Sigma) from a 3mM stock prepared in DMSO. Control cells were treated with the same volume of DMSO.

Cell viability measurements

HEK293T cells were transfected with APOBEC1-YTH and incubated for 24 hours at 37°C. Viability was assessed by trypan blue staining and manual counting of the proportion of viable cells compared to untransfected cells.

Differential gene expression analysis

Gene expression analysis was carried out using the deseq2 package[25]. mRNAs identified as being at least 2-fold increased/decreased in APOBEC1-YTH expressing cells relative to untransfected cells with a corrected p value < 0.05 were reported.

Immunofluorescence

HEK293T cells were transfected with APOBEC1-YTH and fixed 24 h later using 4% paraformaldehyde. Cells were then permeabilized in 0.1% Triton X-100 in PBS and blocked in 1% BSA/PBS for 15 min at 25°C. Rabbit anti-HA antibody was added overnight at 4°C. After 3 x 5 min washes in 1X PBS, secondary antibody (AlexaFluor488-conjugated goat anti-rabbit, 1:1000) was then added for 1 h at 25°C. Cells were washed again in 1 X PBS and incubated in DAPI solution (1:10000 in PBS) for 2 min. Images were acquired on a Leica DMi8 inverted fluorescence microscope.

m^{6}A immunoblotting

Equal amounts of total RNA were separated by agarose gel electrophoresis for 1 h at 70V. RNA was then transferred to a Hybond nylon membrane for 2h using downward transfer with Ambion NorthernMax/Gly transfer buffer. Membranes were then crosslinked using a handheld UV lamp for 1 min with 254 nm light. Membranes were blocked for ≥30 min with 5% nonfat dry milk in 0.1% PBST and probed with anti-m^{6}A antibody overnight at 4°C. Secondary antibody (HRP-goat anti-rabbit; 1:2500) was added for 1 h in blocking buffer and blots were developed with ECL (Amersham ECL Prime) and imaged using the BioRad Chemidoc imaging system.

m^{6}A immunoprecipitation (MeRIP)

30 μg of total RNA was fragmented using Ambion 10X fragmentation reagent at 70°C for 7 min. 1.5ul of the fragmentation reaction was saved as input. The remainder was subjected to m^{6}A immunoprecipitation by first coupling 12 μl of m^{6}A antibody to 50 μl of Protein A/G magnetic beads (Pierce) in 300 μl of IP buffer (10mM Sodium Phosphate, 0.05% Triton-X 100, 140mM NaCl) for 2 h at 4°C, rotating. Beads were then washed three times in IP buffer, and RNA was denatured for 5 min at 75°C followed by 2-3 minute incubation on ice. RNA was then coupled with antibody-bound beads in 300 μl IP buffer for 2 h at 4°C, rotating. Beads were washed five times in 500 μl IP buffer and eluted in 300 μl elution buffer (5mM Tris-HCl, pH 7.5, 1mM EDTA, pH 8.0, 0.05% SDS, 4.2 μl Proteinase K (20mg/ml; Thermo)) for 1.5 h at 50°C while mixing. RNA was then collected with phenol:chloroform extraction and ethanol precipitation.

RT-qPCR

RNA was reverse transcribed using Superscript III with random hexamers according to the manufacturer's instructions (Thermo). cDNA was then used for quantitative PCR using the indicated primers and iQ SYBR Green Super Mix (Bio-Rad) in an Eppendorf RealPlex thermocycler. RNA levels were determined using the ΔΔCt method and were normalized to *GAPDH* levels (MeRIP-RT-qPCR) or *ACTB* levels (mRNA abundance).

Western blotting

Protein was loaded in a NuPAGE 4-12% Bis-Tris precast gel (Thermo) and separated at 180V. Transfers were carried out at 105V for 90 min to a Hybond PVDF membrane. Blocking was carried out for ≥ 30 min in 5% nonfat dry milk/0.1% PBST and antibodies were added overnight in 0.1% PBST at 4°C. Secondary antibodies were incubated on membranes in blocking buffer for 1 h at room temperature. ECL reagent (Amersham ECL Prime) was mixed 1:1 and added to the membranes, which were imaged using the BioRad Chemidoc imaging system.

RNA pulldown assays

RNA pulldowns were performed as previously described[4]. Briefly, 5 μg of bait RNA which contained a single A or m^{6}A residue (5'- biotin-GUUCUUCUGUGGACUGUG -3') was bound to pre-washed streptavidin agarose beads (Sigma-Aldrich) in 200 μl binding buffer (10mM Tris-HCl, pH 7.5, 1.5mM $MgCl_2$, 150mM KCl, 0.5mM DTT, 0.05% (v/v) NP-40) at 4°C for 1 h on a rotator. Beads were then washed twice with 0.5 mL binding buffer. Protein lysates were isolated from HEK293T cells transfected with APOBEC1-YTH or APOBEC1-YTH^{mut} for 24 h by adding lysis buffer (10mM NaCl, 2mM EDTA, 0.5% Triton X-100, 0.5mM DTT, 10mM Tris-HCl, pH 7.5, complete mammalian protease inhibitor cocktail (Sigma) and phosphatase inhibitor cocktail 2 (Sigma)). Cells were lysed by 30 strokes of dounce homogenization and then centrifuged at 10,000 x g at 4°C for 15 min. Supernatants were collected and pre-cleared by mixing with streptavidin agarose beads at 4°C for 1 hour on a rotator. Beads were pelleted at 6,500 rpm for 1 min and supernatants were then mixed with binding buffer. Approximately 1 mg of lysate was mixed with 10 μl of RNasin (Promega) and then added to RNA-bound beads for 30 minutes at room temperature,

followed by a 2 h incubation at 4°C on a rotator. Beads were washed five times (6,500 rpm, 4°C) in binding buffer and proteins were eluted (60°C, 850 rpm for 30 minutes in a thermomixer) using elution buffer (50mM Tris-HCl, pH 8.0, 200mM NaCl, 2% SDS, 1mM biotin). Samples were spun down at 6,500 rpm for 1min at room temperature, and eluates were collected and mixed 1:1 with 2X NuPAGE sample buffer containing 2.5% β-mercaptoethanol and analyzed via western blot.

In vitro transcription

RNA was synthesized using the HiScribe T7 in vitro transcription kit (New England Biolabs). 1 μg of purified PCR product was used for each reaction. Transcription was carried out overnight at 42°C using either ATP or N^6-meATP (Trilink). For experiments using varying amounts of m^6A within an RNA, transcripts synthesized using all ATP or all N^6-meATP were mixed in the indicated proportions.

In vitro deaminase assays

APOBEC1-YTH-HA and APOBEC1-YTH^{mut}-HA proteins were *in vitro* transcribed/translated using the Promega TNT T7 Quick Coupled In Vitro Transcription/Translation kit. Briefly, 1 μg of plasmid DNA was used in a 50 μl reaction and incubated for one hour at 30°C. 5 μl of each reaction was then mixed with 30 ng of a 1500 nt-long RNA with a single internal A or m^6A site, 0.5 μl RNasin (Promega), in 1X deaminase buffer (10mM Tris-HCl pH7.5, 50mM KCl, 0.1uM $ZnCl_2$). Reactions were incubated for 4 h at 37°C. RNA was isolated with the Qiagen RNeasy Plus Mini kit and treated with DNase I (New England Biolabs) for 15 min at 37°C. For assays using cellular RNA, *in vitro* deamination was carried out for 6 h at 37°C using 50 ng of total RNA from HEK293T cells. Sequencing libraries were then prepared using the NebNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs).

cDNA synthesis and Sanger sequencing

1 μg purified RNA from *in vitro* deamination assays or from cells expressing APOBEC1-YTH or APOBEC1 alone was used for cDNA synthesis using either a 1:1 mix of oligo(dT) and random hexamers or gene specific primers (below). cDNA synthesis was carried out using the SuperScript III reverse transcriptase kit according to manufacturer's instructions (Thermo). PCR was then performed using Phusion High-Fidelity PCR Mastermix (New England Biolabs) and primers flanking m^6A target regions. Purified PCR products were then either directly sequenced using Sanger sequencing (Genewiz) (Supplementary Fig. 15) or cloned into the pCR Blunt II TOPO vector (Thermo) (Supplementary Fig. 1a). For direct Sanger sequencing, measurements of C to U conversion were quantified from Sanger sequencing traces by calculating the height of T sequence peaks relative to C sequence peaks at individual mixed (C/T) sites. For TOPO cloning, a minimum of three individual clones were selected for each condition, and a representative Sanger sequencing trace for each is shown. Sequencing primers were used as indicated and comprised the sequences listed below.

Primers (5’-3’)

YTH Fwd: AGACTCCCGGGACCTCAGAG

YTH^{mut} Fwd:

ACTCCCGGGACCTCAGAGTCCGCCACACCAGAAGGCCGGGTTTTCATCATTAAG

YTH-HA Rev: CGGGTTTAAACTCAGGCGTAGTC

BSG RT primer: GTGGGGGCGATCTTTATTGTGGCGG

ACTB RT primer: TGTGCAATCAAAGTCCTCGGCCAC

BSG Fwd/Sanger: GCCAATGCTGTCTGGTTGCGCC

BSG Rev: GGAGGCTTCTGCGGTTCTGGAG

ACTB Fwd/Sanger: CAGCAAGCAGGAGTATGACGAGTC

ACTB Rev: CATGCCAATCTCATCTTG

ACTB MeRIP Fwd: CATGTACGTTGCTATCCAGGC

ACTB MeRIP Rev: CTCCTTAATGTCACGCACGAT

ATRX MeRIP Fwd: CGAAGATCCCCACGTGTAAAGACTAC

ATRX MeRIP Rev: CATCCTGCTCACCTCTTTGAGG

BPTF MeRIP Fwd: GTGTTAGATGATGTCTCCATTCGGAG

BPTF MeRIP Rev: CACTTTCCTCCTGTATGAGCGG

Single A 1500nt RNA RT primer: GCCAAGAGGCAACACACCAAC

Single A 1500nt RNA Fwd: CGGTTTCTCTCGGTCTGTTTTCC

Single A 1500nt RNA Rev: CAGAAGGCGACAACACAGCAACACC

Next-generation sequencing

All sequencing was performed by the Duke University Sequencing and Genomic Technologies Core facility. HEK293T cells were transfected with APOBEC1-YTH, APOBEC1-YTH^{mut}, or APOBEC1 alone using FuGENE HD according to the manufacturer’s instructions (Promega). After 24 h, total RNA was isolated with TRIzol (Thermo) and subjected to DNase I treatment using RNase-free DNaseI (Sigma) for 20 min at 37°C. 1 μg of total RNA was then used for sequencing library preparation using the NebNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). For low-input samples, the Single Cell/Low Input RNA Library Prep Kit (New England Biolabs) was used with either 10 ng or 100 ng of total RNA as input as indicated. We do not find it necessary to remove rRNA prior to sequencing library preparation, although doing so

may potentially be used to further increase the efficiency of C to U mutation detection. Prior to sequencing, samples were barcoded using NEBNext Multiplex Oligos for Illumina (New England Biolabs). Libraries were then sequenced on the Illumina HiSeq 4000. For PacBio sequencing, 1 μg of RNA was used for library preparation using the Iso-Seq system. Two samples were sequenced on one SMRT cell of the PacBio Sequel instrument and processed using the Iso-Seq analysis pipeline.

C to U editing site analysis

Sequencing reads were demultiplexed, adapters were removed, and strand-specific reads were reverse-complemented and aligned to the human genome (hg19) using Novoalign. PCR duplicates were collapsed, and individual C to U mutations were identified using CIMS[26]. C to U sites identified with the p<1 threshold were further filtered, and only those sites that had a minimum of 2 mutations, at least 10 reads per replicate, and a mutation/read (m/k) threshold of 10-60% (for high-stringency lists) were kept. We find that adjusting the number of mutations, reads per replicate, and m/k threshold is a good way to increase/decrease stringency of m^6A site calls to a desired level. If desired, sites can be further filtered to include only C to U editing events which are immediately preceded by an A; however, this could potentially exclude some m^6A sites for which editing occurs at a nearby C instead of the immediately adjacent C. In addition to these filtering steps, known mutations in the human genome (dbSNP 150), as well as endogenous C to U editing sites identified by sequencing of wild type HEK293T cells, were also removed. For APOBEC1-YTH or APOBEC1-YTH^{mut} expressing cells, the list of C to U editing sites was further processed by removing sites detected in cells expressing APOBEC1 alone. For determining enrichment of C to U editing between samples, a filter of m/k was used to find sites that were of the indicated fold-enrichment greater than the reference sample. PacBio datasets were aligned to the human genome (hg19) using GSNAP[27] and subjected to the same pipeline as above to identify C to U sites. *In vitro* DART-Seq datasets were also subjected to the same C to U mutation analysis using a m/k filtering threshold of 5-60%.

Exon length measurements

Exon length was determined using the RefSeq hg19 annotation. Sequencing reads spanning individual exons were processed by removing first and last exons, according to the consensus RefSeq annotation. In cases where reads overlapped with multiple isoforms and therefore different exons, the consensus Refseq sequence was used.

Metagene and motif analyses

Metagene analysis was performed using hg19 annotations according to previously published methods[28]. Discovery of enriched motifs was performed using HOMER[29] using sequences spanning a region 4 nucleotides up- and downstream of C to U editing sites as input.

Replicates analysis

Independent biological replicates of global DART-Seq experiments were compared by computing the Pearson correlation coefficient between the number of C to U mutations per gene between any two replicate experiments.

Calculating C to U editing events in individual reads

To determine the number of reads with one or more C to U editing events out of all reads that spanned at least two called C to U editing events, we first used Sam2Tsv[30] to identify individual reads containing C to U mutations. The first and last position of each read was then used in conjunction with bedtools intersect[31] to find reads that overlap more than one C to U editing site from our final list of high-confidence sites. The number of editing events within these reads was then counted and summed.

Dataset comparisons

DART-Seq and m^{6}A immunoprecipitation (MeRIP-Seq, miCLIP) datasets were analyzed using the *closest* and *intersect* features of the bedtools suite[31]. For comparison of methylated mRNAs, bed file coordinates were annotated using the annotation feature of the metagene analysis pipeline (above) to give individual mRNAs, which were then compared between datasets.

Statistics

Statistical analysis of cell viability and western blot data were performed using a two-tailed t-test. Analysis of C to U editing enrichment in various transcript regions following CPT treatment, as well as analysis of the proportion of C to U sites following each nucleotide, was performed using a chi-squared test.

Data availability

The data that support the findings of this study have been deposited in NCBI's Gene Expression Omnibus (GEO) under accession code GSE125780.

Reporting Summary

Further information on research design is available in the Life Sciences Reporting Summary linked to this article.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Seung H. Choi for generating METTL3-depleted cells. We also thank S. Horner and members of the Meyer laboratory for comments and suggestions. This study was supported by the National Institutes of Health (NIH) grants R00MH104712, R01MH118366 and DP1DA046584. K.D.M. is also supported by the Rita Allen Foundation, the Kinship Foundation Searle Scholars Program, and the Klingenstein-Simons Fellowship in Neurosciences.

References

1. Zhao BS, Roundtree IA & He C Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol 18, 31–42 (2017). [PubMed: 27808276]
2. Meyer KD & Jaffrey SR Rethinking m6A Readers, Writers, and Erasers. Annual Review of Cell and Developmental Biology 33, null (2017).

3. Meyer KD et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3’ UTRs and near stop codons. Cell 149, 1635–1646 (2012). [PubMed: 22608085]
4. Dominissini D et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485, 201–206 (2012). [PubMed: 22575960]
5. Linder B et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods (2015).
6. Ke S et al. A majority of m6A residues are in the last exons, allowing the potential for 3’ UTR regulation. Genes Dev 29, 2037–2053 (2015). [PubMed: 26404942]
7. Navaratnam N et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem 268, 20709–20712 (1993). [PubMed: 8407891]
8. Komor AC, Kim YB, Packer MS, Zuris JA & Liu DR Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424 (2016). [PubMed: 27096365]
9. Schibler U, Kelley DE & Perry RP Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. J Mol Biol 115, 695–714 (1977). [PubMed: 592376]
10. Wei CM & Moss B Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. Biochemistry 16, 1672–1676 (1977). [PubMed: 856255]
11. Wei CM, Gershowitz A & Moss B 5’-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA. Biochemistry 15, 397–401 (1976). [PubMed: 174715]
12. Li F, Zhao D, Wu J & Shi Y Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. Cell Res 24, 1490–1492 (2014). [PubMed: 25412658]
13. Wang X et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 505, 117–120 (2014). [PubMed: 24284625]
14. Schwartz S et al. Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5’ Sites. Cell Rep 8, 284–296 (2014). [PubMed: 24981863]
15. Chen K et al. High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. Angew Chem Int Ed Engl 54, 1587–1590 (2015). [PubMed: 25491922]
16. Liu N et al. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA 19, 1848–1856 (2013). [PubMed: 24141618]
17. Weng YL et al. Epitranscriptomic m(6)A Regulation of Axon Regeneration in the Adult Mammalian Nervous System. Neuron 97, 313–325 e316 (2018). [PubMed: 29346752]
18. Merkurjev D et al. Synaptic N6-methyladenosine (m6A) epitranscriptome reveals functional partitioning of localized transcripts. Nat Neurosci 21, 1004–1014 (2018). [PubMed: 29950670]
19. Zeng Y et al. Refined RIP-seq protocol for epitranscriptome analysis with low input materials. PLoS Biol 16, e2006092 (2018). [PubMed: 30212448]
20. Mauer J et al. Reversible methylation of m6Am in the 5’ cap controls mRNA stability. Nature 541, 371–375 (2017). [PubMed: 28002401]
21. Luo S & Tong L Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. Proc Natl Acad Sci U S A 111, 13834–13839 (2014). [PubMed: 25201973]
22. Xu C et al. Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat Chem Biol 10, 927–929 (2014). [PubMed: 25242552]
23. Molinie B et al. m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. Nat Methods 13, 692–698 (2016). [PubMed: 27376769]
24. Slobodin B et al. Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell 169, 326–337 e312 (2017). [PubMed: 28388414]

Methods-only References

25. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]
26. Shah A, Qian Y, Weyn-Vanhentenryck SM & Zhang C CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. Bioinformatics 33, 566–567 (2017). [PubMed: 27797762]
27. Wu TD & Watanabe CK GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859–1875 (2005). [PubMed: 15728110]
28. Olarerin-George AO & Jaffrey SR MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. Bioinformatics 33, 1563–1564 (2017). [PubMed: 28158328]
29. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589 (2010). [PubMed: 20513432]
30. Lindenbaum P JVarkit: java-based utilities for Bioinformatics. (2015).
31. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]

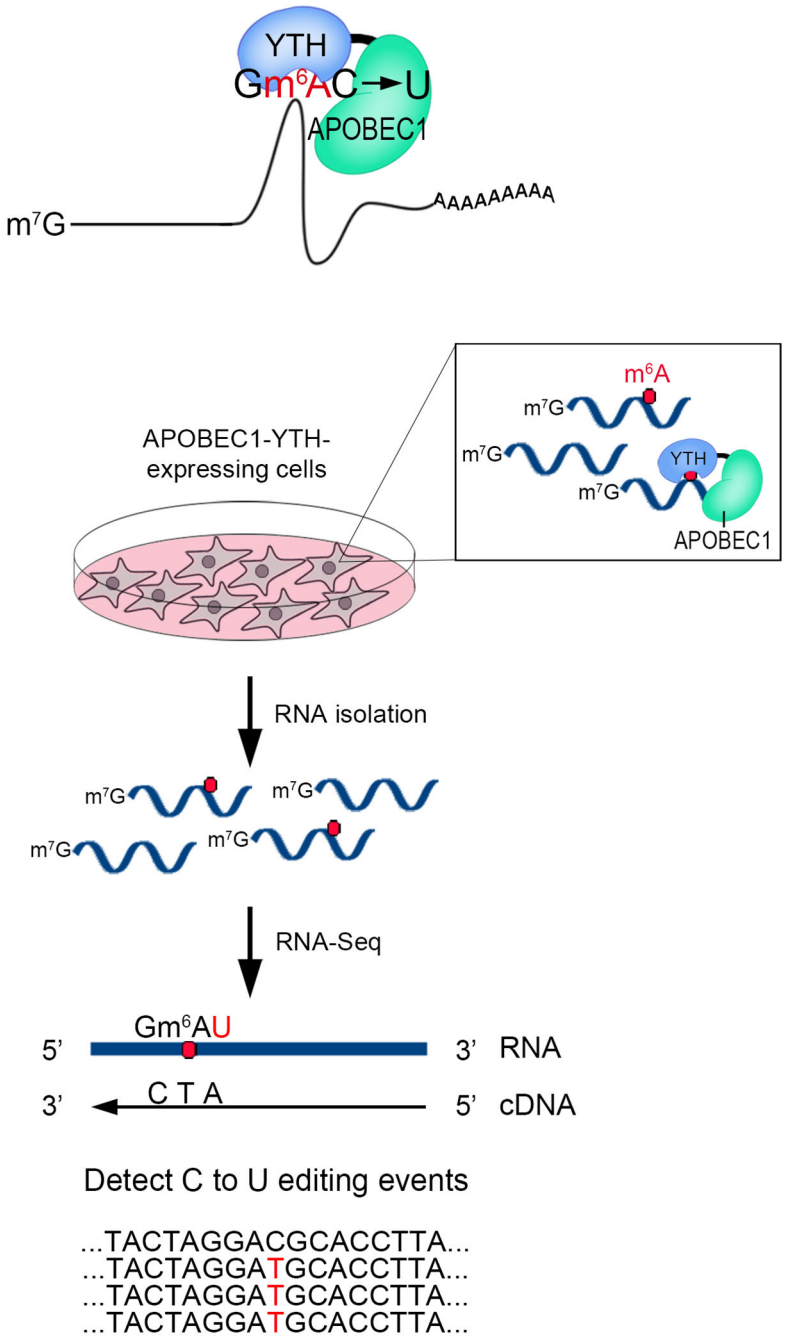

Figure 1. Development of a targeted deamination strategy to detect m^{6}A.

(a) Schematic of the DART-Seq method. APOBEC1 is fused to the YTH domain to guide C to U editing at cytidine residues adjacent to m^{6}A sites. APOBEC1-YTH is expressed in cells and total RNA is isolated and subjected to RNA-Seq. C to U mutations are then detected to identify sites of m^{6}A.

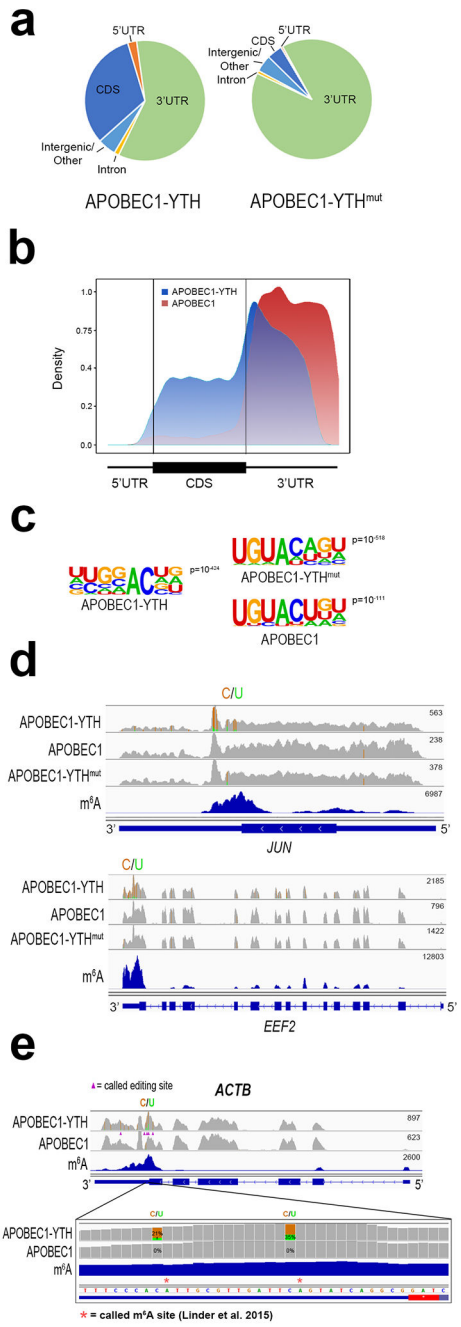

Figure 2. DART-Seq identifies m^{6}A-containing RNAs transcriptome-wide.

(a) Pie charts showing the distribution of C to U editing sites in distinct transcript regions. The majority of DART-Seq sites from APOBEC1-YTH-expressing cells are found in the CDS and 3'UTR, similar to the distribution of m^{6}A (left; n = 3 independent samples). In contrast, cells expressing APOBEC1-YTH^{mut} exhibit editing events mostly in 3'UTRs (right; n = 3 independent samples). (b) Metagene analysis showing a density plot of the distribution of C to U editing events detected by DART-Seq. There is an enrichment near the stop codon which resembles the global distribution of m^{6}A. In contrast, C to U editing

events in cells expressing APOBEC1 alone show a unique distribution characterized by editing throughout the 3'UTR. Results are representative of 3 independent experiments. (c) The most enriched motif found within a +/− 4 nt region surrounding C to U editing sites in APOBEC1-YTH-expressing cells contains a GGACU sequence, which matches the m^6A consensus (n = 39,069). In contrast, motifs from cells expressing APOBEC1-YTH^{mut} (n = 21,841) or APOBEC1 alone (n = 6,814) fail to show enrichment for m^6A motifs. P values for individual motif enrichment were calculated using the cumulative binomial distribution. (d) IGV browser tracks of DART-Seq data show C to U mutations in the *JUN* and *EEF2* mRNAs. C to U mutations found in at least 10% of reads are indicated by gold/green coloring (gold indicates the abundance of C sites, and green indicates the abundance of U sites at each position; both genes are transcribed from the negative strand). APOBEC1-YTH expression leads to robust C to U editing in regions of m^6A identified by MeRIP-Seq (blue track). In contrast, cells expressing APOBEC1 alone or APOBEC1-YTH^{mut} fail to show this editing. Results are representative of 3 independent experiments. (e) DART-Seq read coverage at the *ACTB* locus indicates C to U editing at cytidines adjacent to known m^6A sites in the *ACTB* 3'UTR. C to U mutations found in at least 10% of reads are indicated by gold/green coloring at individual sites (gold indicates the abundance of C sites, and green indicates the abundance of U sites at each position; *ACTB* is transcribed from the negative strand). Percentages of C to U mutations at two individual sites previously identified by miCLIP[5] are indicated in the expanded view. Results are representative of 3 independent experiments.

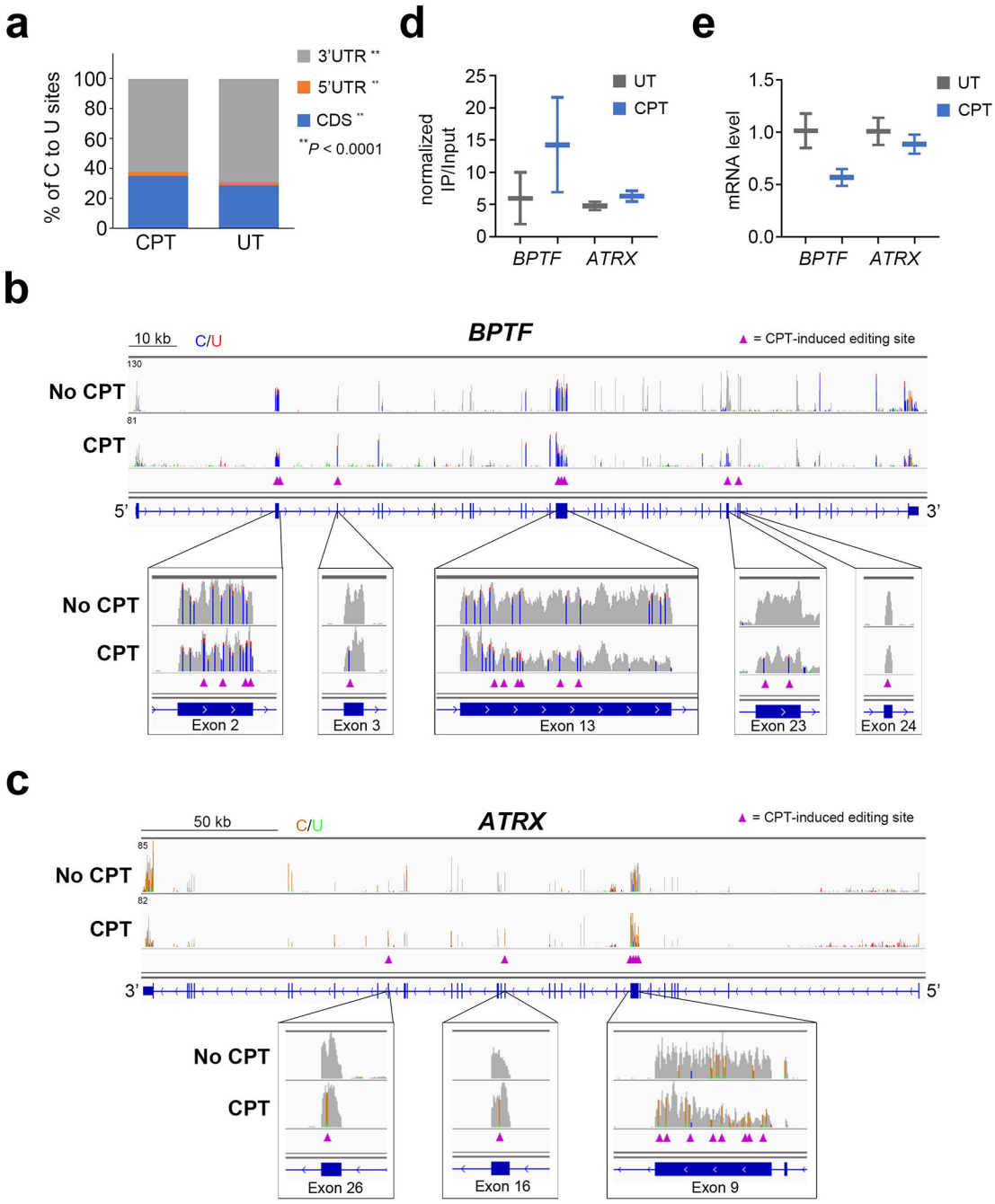

Figure 3. DART-Seq monitors changes in m^{6}A over time.

(a) Distribution of C to U mutations discovered by DART-Seq in HEK293T cells treated with camptothecin (CPT; n = 5,689) compared to untreated controls (UT; n = 40,594) indicates a slight enrichment within the CDS. (b,c) IGV browser images showing C to U mutations within the CDS of the *BPTF* and *ATRX* transcripts following CPT treatment. C to U mutations found in at least 10% of reads are indicated by blue/red (C/U) coloring in the *BPTF* transcript (positive strand) and by gold/green (C/U) coloring in the *ATRX* transcript (negative strand). C to U sites enriched after CPT treatment are indicated by purple triangles.

n = 2 independent samples. (d) MeRIP-RT-qPCR confirms enrichment of m^6A in the *BPTF* and *ATRX* mRNAs following CPT treatment. *n* = 2 biological replicates; box plot indicates mean and upper/lower limits. (e) RT-qPCR analysis using RNA from untreated and CPT-treated cells shows a decrease in abundance of the *BPTF* and *ATRX* transcripts following CPT treatment. *n* = 2 biological replicates; box plot indicates mean and upper/lower limits.

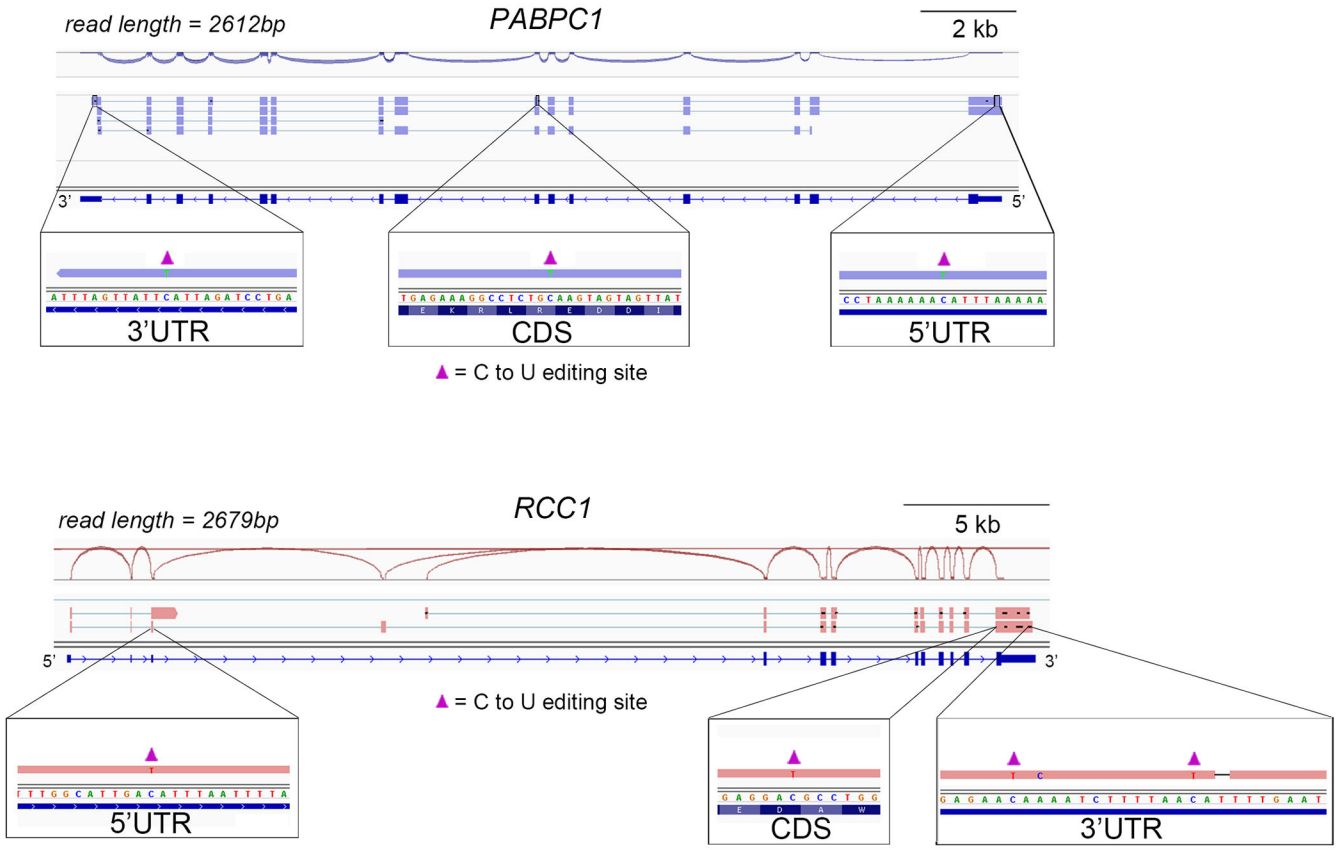

Figure 4. Long-read DART-Seq reveals m^{6}A distribution within individual RNA molecules.

IGV browser tracks showing PacBio DART-Seq data at two representative mRNAs (*PABPC1*; top and *RCC1*; bottom). C to U mutations found in at least 10% of reads are indicated by blue/red (C/U) coloring in the *PABPC1* transcript (positive strand) and by gold/green (C/U) coloring in the *RCC1* transcript (negative strand). Boxes show expanded views of C to U mutations within the same read spanning the 5'UTR, CDS, and 3'UTR. $n = 2$ independent samples.