

Identification of m⁶A residues at single-nucleotide resolution using eCLIP and an accessible custom analysis pipeline

JUSTIN T. ROBERTS,^{1,2} ALLISON M. PORMAN,¹ and AARON M. JOHNSON^{1,2,3}

¹Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045, USA

²Molecular Biology Graduate Program, University of Colorado-Denver Anschutz Medical Campus, Aurora, Colorado 80045, USA

³RNA Biosciences Initiative, University of Colorado School of Medicine, Aurora, Colorado 80045, USA

ABSTRACT

Methylation at the N⁶ position of adenosine (m⁶A) is one of the most abundant RNA modifications found in eukaryotes; however, accurate detection of specific m⁶A nucleotides within transcripts has been historically challenging due to m⁶A and unmodified adenosine having virtually indistinguishable chemical properties. While previous strategies such as methyl-RNA immunoprecipitation and sequencing (MeRIP-seq) have relied on m⁶A-specific antibodies to isolate RNA fragments containing the modification, these methods do not allow for precise identification of individual m⁶A residues. More recently, modified cross-linking and immunoprecipitation (CLIP)-based approaches that rely on inducing specific mutations during reverse transcription via UV cross-linking of the anti-m⁶A antibody to methylated RNA have been used to overcome this limitation. However, the most utilized version of this approach, miCLIP, can be technically challenging to use for achieving high-complexity libraries. Here we present an improved methodology that yields high library complexity and allows for the straightforward identification of individual m⁶A residues with reliable confidence metrics. Based on enhanced CLIP (eCLIP), our m⁶A-eCLIP (meCLIP) approach couples the improvements of eCLIP with the inclusion of an input sample and an easy-to-use computational pipeline to allow for precise calling of m⁶A sites at true single-nucleotide resolution. As the effort to accurately identify m⁶As in an efficient and straightforward way intensifies, this method is a valuable tool for investigators interested in unraveling the m⁶A epitranscriptome.

Keywords: RNA methylation; N⁶-methyladenosine; CLIP; CIMS; METTL3

INTRODUCTION

N⁶-methyladenosine (m⁶A) is a modification to RNA where a methyl group is added to the N⁶ position of adenosine. m⁶A is the most prevalent posttranscriptional modification of eukaryotic mRNA and has important roles in a variety of physiological processes including cell differentiation (Geula et al. 2015) and development (Wang et al. 2014b), alternative splicing (Xiao et al. 2016), and regulation of mRNA stability (Wang et al. 2014a). m⁶A residues are typically deposited cotranscriptionally (Ke et al. 2017) onto nascent pre-mRNA molecules in the nucleus via a “writer” consisting of a stable heterodimer enzyme complex composed of methyltransferase proteins METTL3/14 in association with the pre-mRNA regulator WTAP and additional accessory components such as KIAA1429 and RBM15 (Ke et al. 2017; Wu et al. 2017). This writer

complex targets RNAs containing a “DRACH” consensus sequence (where “D” is any nucleotide but cytosine, “R” is any purine, and “H” is any nucleotide but guanine), with the cytosine downstream from the substrate adenine being essential for methylation (Liu et al. 2014). The human methyltransferase METTL16 can also generate m⁶A modifications, though these residues do not occur within the same “DRACH” consensus motif and only a very few substrates are known (Doxtader et al. 2018; Ruskowska et al. 2018). While m⁶A modifications have been identified throughout the transcriptome, they are most-often enriched around 3'-UTRs and stop codons (Dominissini et al. 2012; Meyer et al. 2012; Ke et al. 2015). In contrast, a similar RNA modification, N⁶, 2'-O-dimethyladenosine (m⁶Am), is located on the 5' ends of mRNAs and is catalyzed by the methyltransferase PCIF1 (Akichika et al. 2019; Boulias et al. 2019; Sendinc et al. 2019; Sun et al.

Corresponding author: Aaron.m.johnson@CUAnschutz.edu

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.078543.120>. Freely available online through the RNA Open Access option.

© 2021 Roberts et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

2019). Following methylation, m⁶A containing transcripts are specifically recognized by “reader” proteins, the most well characterized being members of the YTH domain family (Wu et al. 2017). Depending on the subcellular localization of these readers, recognition of m⁶A can mediate critical cellular functions. For example, YTHDC1, the main nuclear YTH protein, regulates chromatin accessibility (Liu et al. 2020), pre-mRNA splicing (Xiao et al. 2016), nuclear export (Roundtree et al. 2017), and transcriptional repression (Patil et al. 2016), whereas binding of m⁶A via the cytoplasmic reader YTHDF2 leads to transcript decay (Wang et al. 2014a). m⁶A residues are also dynamically reversible via “erasure” by the demethylases ALKBH5 and FTO (Wu et al. 2017).

Current strategies to identify m⁶A residues are largely based on using m⁶A-specific antibodies to isolate transcripts containing methylated adenosine. The initial approaches, methyl-RNA immunoprecipitation, and sequencing (MeRIP-seq) (Meyer et al. 2012) and m⁶A-seq (Dominissini et al. 2012), involved the immunoprecipitation of ~100 nt long RNA fragments bound to anti-m⁶A antibodies whereby successive sequencing and mapping of the reads results in the selective enrichment for sequences that contain m⁶A. However, as the m⁶A residue could be anywhere within the precipitated fragment, single-nucleotide resolution can only be approximated using the “DRACH” motif as a guide to predict the specific m⁶A site. In contrast, more recently several techniques have demonstrated the ability to overcome this limitation by using cross-linking and immunoprecipitation (CLIP) (Licatalosi et al. 2008) based strategies where UV light is used to cross-link the m⁶A antibody to methylated transcripts. Two of these methods, m⁶A-CLIP (Ke et al. 2015) and “m⁶A individual-nucleotide-resolution CLIP” (miCLIP) (Linder et al. 2015), demonstrate that following cross-linking and IP of the antibody:RNA complex, removal of the antibody leaves a cross-linked amino acid “scar” near the m⁶A site, and reverse transcription over this adduct leads to distinct mutations that arise from the increased frequency of reverse transcriptase (RT) errors at the exact nucleotide where amino acids cross-link to RNA. Specifically, the miCLIP technique showed that there is a markedly high frequency for C-to-T transitions at the obligate C that occurs one nucleotide downstream from putative m⁶A sites within the resulting cDNA (an A is incorporated into the cDNA instead of a G). These mutations can then be used to identify individual m⁶A residues via computational screening of sequencing reads. Alternatively, miCLIP also demonstrated that cross-linking different anti-m⁶A antibodies can induce significant RT termination events which can be used to identify m⁶A. A similar method, photo-cross-linking-assisted m⁶A-sequencing (PA-m⁶A-seq) (Chen et al. 2015), uses PAR-CLIP (Hafner et al. 2010) to identify m⁶A residues based on the introduction of 4-thiouridine induced T-to-C transitions near the methylated adenosine. While these

strategies do resolve the lack of single-nucleotide resolution inherent in the previous non-CLIP-based methods, there are several limitations that remain to be addressed. Specifically, the miCLIP protocol itself uses several steps such as radiolabeling and circularization of the cDNA library that make it technically challenging. While all the methods described above outline a strategy to identify m⁶A sites from the resulting sequencing reads, they require a considerable amount of bioinformatic expertise in order to identify m⁶A positions.

We have developed an updated antibody-based approach to accurately call m⁶A residues at single-nucleotide resolution. m⁶A-eCLIP (meCLIP) is a modification of the existing enhanced CLIP (eCLIP) protocol (Van Nostrand et al. 2016) with changes specifically designed to identify m⁶As. Compared to existing strategies, the protocol is technically simplified and includes a comprehensive computational pipeline that runs to completion once it is executed. Using meCLIP yields higher numbers of identified sites compared to miCLIP while at the same time implementing further steps to increase confidence in site calls. We have successfully validated this strategy on several cell lines and confirmed its ability to accurately call individual m⁶A residues throughout the transcriptome in a high-throughput and more straightforward manner compared to previous approaches.

RESULTS

Overview of eCLIP library preparation

Our meCLIP approach adapts the eCLIP strategy, utilizing UV cross-linking to covalently link an anti-m⁶A antibody to fragmented poly(A)-selected transcripts containing m⁶A and then immunoprecipitating the antibody-bound RNA. This antibody:RNA complex is then run on a SDS-PAGE gel and transferred to a nitrocellulose membrane to remove any non-cross-linked RNA. Following treatment with proteinase K to remove nearly all of the antibody except the cross-linked amino acid, the RNA is isolated, one adapter is ligated, and the RNA is converted into cDNA. All first-strand cDNA products receive a second adapter required for sequencing, ensuring that efficiency of the reverse transcriptase crossing the amino acid adduct does not impede the library preparation. Reverse transcription over the anti-m⁶A cross-link site results in C-to-T mutations in the template strand read from the resulting sequencing, and a custom algorithm then identifies sites of elevated C-to-T conversion frequency that occur within the m⁶A consensus motif. An overview of the library preparation can be seen in Figure 1A.

By omitting the radiolabeling and autoradiographic visualization steps used in iCLIP, the eCLIP protocol is technically less challenging and can be completed in as few as 4 days. Additionally, optimizations to the library preparation itself, including replacing the circularization step used in

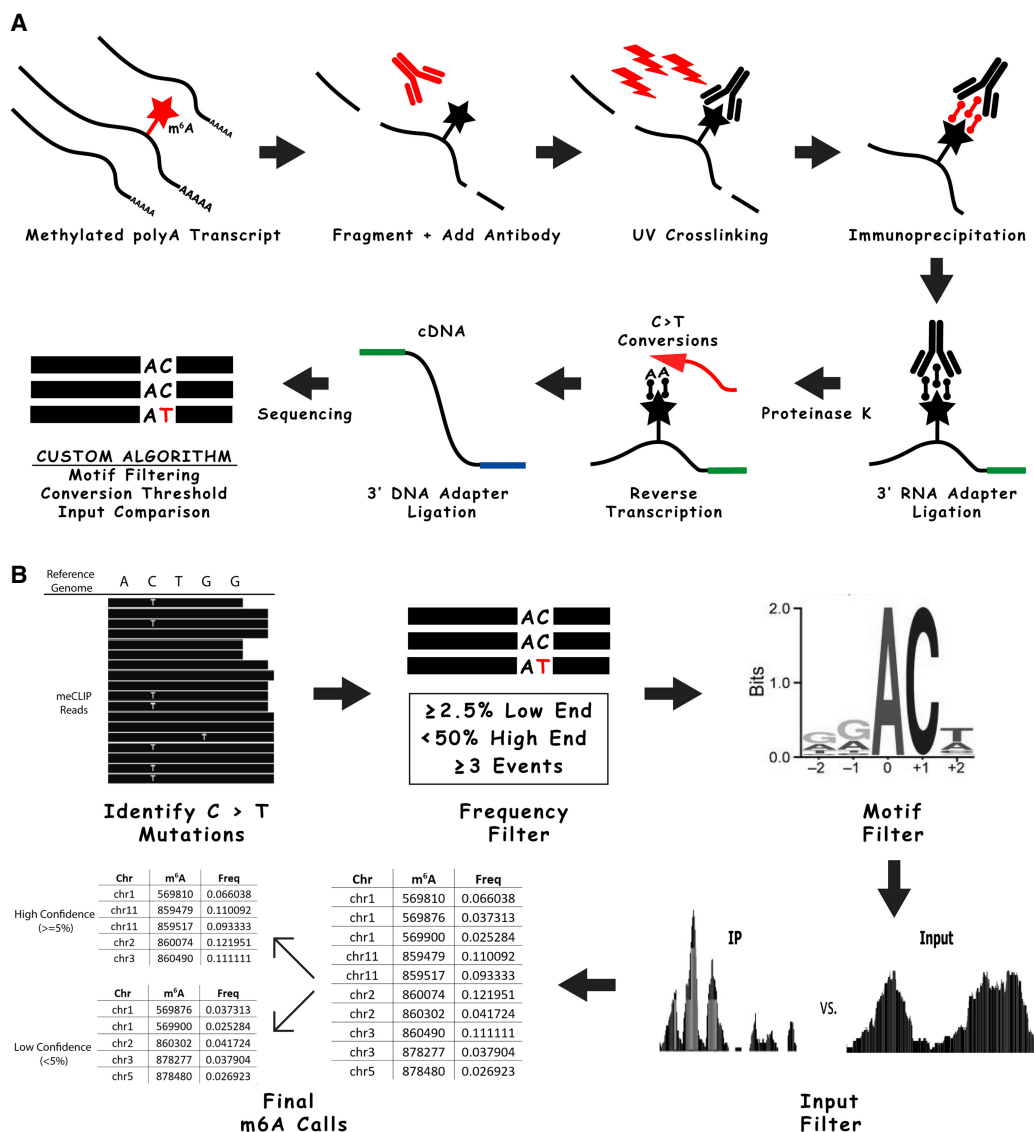


FIGURE 1. Overview of the meCLIP strategy, including summary of library preparation and the subsequent algorithm to identify m⁶A residues from the sequencing reads. (A) Following isolation of mRNA from total RNA samples, the transcripts are fragmented, and UV cross-linked to anti-m⁶A antibody (*top middle*). Following immunoprecipitation (*top right*), an indexed 3' RNA adapter is ligated to the cross-linked RNA fragment (*bottom right*). The antibody is then removed, the RNA is reverse transcribed, and a 3' single-stranded DNA adapter is ligated onto the resulting cDNA (*bottom middle*). Residual amino acid adducts resulting from the RNA:antibody cross-linking cause C-to-T mutations that are detectable in the resulting sequencing reads. These mutations are used as input for a custom algorithm that identifies sites of elevated C-to-T conversion frequency that occur within the m⁶A consensus motif (*bottom left*). (B) Following sequencing, the resulting reads are used in a custom algorithm that uses the "mpileup" command of SAMtools (Li et al. 2009) to identify sites of elevated C-to-T mutations. These positions are then filtered based on the frequency of the conversion ($\geq 2.5\%$ and $\leq 50\%$ with a minimum of three events) and their occurrence within the m⁶A consensus motif ("RAC," where "R" is any purine). Finally, the filtered positions are compared to the similarly analyzed input sample and any overlapping positions are removed. The resulting m⁶A calls are categorized into low and high confidence sets based on the mutational frequency ($< 5\%$ for low, $\geq 5\%$ for high).

iCLIP with two distinct adapter ligations, the second after cDNA synthesis, result in significant improvements to library complexity. While a direct comparison of library complexity between our method and miCLIP was not performed, Van Nostrand et al. found that the improvements implemented in eCLIP resulted in ~ 1000 -fold decreases in requisite library amplification compared to iCLIP and thereby decreased the number of discarded PCR reads by $\sim 60\%$.

Optimization of RNA fragmentation improves successful library generation

We have found that initial optimization of the cation-based RNA fragmentation step can increase the quality of the final library. While too little fragmentation results in amplicons that are outside the recommendations for Illumina sequencing (200–500 bp for NovaSEQ 6000), over-

fragmenting can result in a severe reduction in yield of appropriately sized immunoprecipitated RNA for input into the library preparation steps. While the actual conditions for fragmentation depend on the individual RNA sample, important factors to consider include: (i) the concentration of input RNA; (ii) the duration of time and temperature that the RNA is fragmented; and (iii) the length of reads used in sequencing. Based on these points, we recommend that a trial with total RNA be fragmented at 70°C with various durations ranging from 3 to 15 min and then analyzed on a TapeStation (Agilent) using High Sensitivity RNA ScreenTape (if unavailable, visualization by agarose gel electrophoresis can be performed instead). Further adjustments can then be made for the actual poly(A) RNA sample taking into consideration that poly(A)-selected RNA tends to fragment slightly faster. We recommend an optimal fragment size of 100 to 200 nt (for sequencing on NovaSeq with 2 × 150 run). For reference, we have included sample TapeStation results from appropriate and undesirable fragmentations (Supplemental Fig. 1). The TapeStation coupled with High Sensitivity RNA ScreenTape requires minimal RNA material for analysis, allowing for the same sample to be analyzed and then subsequently used in the experiment.

Improved adapter ligation to increase efficiency

Instead of addition of both sequencing adapters to the original RNA fragments as in m⁶A-CLIP (based on HITS-CLIP [Licatalosi et al. 2008]), or using circular ligation as implemented in miCLIP (based on iCLIP), our meCLIP protocol (based on eCLIP) adds adapters for sequencing in two separate steps. The first step uses an indexed 3' RNA adapter that is ligated to cross-linked RNA fragment while still on the immunoprecipitation beads, and the second is a 3' ssDNA adapter that is ligated to cDNA following reverse transcription. The first 3' RNA adapter is "in-line-barcoded" and may consist of a number of matched combinations (A01 + B06, X1A + X1B, etc.) that are detailed in the original eCLIP protocol and included here as well. The second ssDNA adapter (rand3Tr3) contains a unique molecular identifier (UMI) that allows for determination of whether two identical sequenced reads indicate two unique RNA fragments or PCR duplicates of the same RNA fragment. Therefore, the resulting reads generally have the following structure:

```
Read 1—NNNNNTGCTATT [Sequenced Fragment
(RC)] NNNNNNNNNNAGATCGGAAGAGCAC
Read 2—NNNNNNNNNN [Sequenced Fragment]
AATAGCANNNNN
```

where read 1 begins with the barcoded RNA adapter (X1B displayed here) and read 2 (corresponding to the sense strand) begins with the UMI (either N5 or N10) followed by a sequence corresponding to the 5' end of the original

RNA fragment or reverse transcription termination site. To account for errors in the UMI itself (which would impact the accuracy of quantifying unique molecules at a given genomic locus), our downstream analysis pipeline uses the software package UMI-tools (Smith et al. 2017) which uses network-based algorithms to correctly identify true PCR duplicates.

Standardized strategy to reduce PCR duplication

In addition to not requiring any radiolabeling, compared to iCLIP, the eCLIP protocol also decreases the required amplification by up to ~1000-fold resulting in far less reads being removed due to PCR duplication of the same molecule. To determine the optimal amount of amplification for the final sequencing library, miCLIP recommends that aliquots of the full cDNA library be amplified across a range of cycles (15, 20, or 25) and then visualized on a gel to determine the appropriate number of cycles for the remaining library. In contrast, the eCLIP protocol on which our strategy is based omits the gel visualization step and instead advises that a qPCR experiment using a 1:10 diluted sample of the cDNA library be performed to obtain a baseline cycle number that can then be adjusted for the final library amplification (the suggestion is to use three cycles less than the C_q obtained from the qPCR to account for the 1:10 dilution).

Our approach combines aspects of eCLIP and miCLIP to determine the optimal amount of amplification for the final sequencing library. Similar to eCLIP, we perform a qPCR experiment on diluted (1:10) cDNA to determine the final amplification based on the resulting cycle number (C_q – 3). We additionally perform an end-point PCR experiment on the same diluted cDNA using three different cycle numbers (C_q – 3, C_q, C_q + 3). Then, similar to miCLIP, the amplified samples are visualized on a polyacrylamide gel to further optimize for concentration and size distribution (Supplemental Fig. 2). By combining steps from multiple approaches, we ensure that the final library will not be saturated with PCR duplicates while still allowing for adequate sequencing depth to identify m⁶As.

Straightforward analysis pipeline using Snakemake

While other strategies have been developed to identify m⁶As from sequencing data (MeRIP-seq, m⁶A-seq, and m⁶A-CLIP), the protocol most directly comparable to our approach, miCLIP, relies on cross-linking-induced mutation site (CIMS) analysis, now a part of the CLIP Tool Kit (CTK) software package (Shah et al. 2017) which was originally designed to identify sites of RNA:protein cross-linking from CLIP data. While this method does ultimately provide a set of m⁶A sites deduced from the C-to-T mutations in the sequencing reads, the implementation requires considerable bioinformatic expertise in the form

of manually installing prerequisite software and executing the individual tools from the command line step-by-step. In comparison, our approach implements the workflow management system Snakemake (Köster and Rahmann 2012) to streamline the process, requiring only a single configuration file and no manual installation of software libraries. Specifically, as opposed to running multiple scripts one at a time, Snakemake workflows combine the execution of all the component commands into a human readable file that is easily modifiable.

Once the workflow is executed, the reads are first assessed for appropriate quality and presence of adapters which are removed accordingly. The reads are then mapped to RepBase (Bao et al. 2015) to remove repetitive elements and ribosomal RNA and then to the actual reference genome itself. Following mapping, any PCR duplicates are collapsed within the alignment file which is then sorted and indexed in preparation for calling mutations. The output files of each of these steps is automatically compiled and presented as a summary file to the user for reference. Finally, the custom algorithm (outlined in Fig. 1B) identifies variations from the reference genome at each position and then specifically identifies C-to-T conversions occurring between a frequency threshold of $\geq 2.5\%$ and $\leq 50\%$ (with a minimum of three reads supporting the mutation). Those positions meeting these thresholds are then analyzed for the presence of the m⁶A consensus motif "RAC" (where R is any purine). After comparing to the corresponding input sample (described below), a list of m⁶A sites is reported in the form of their individual coordinates within the genome along with gene and transcript annotations and supporting C-to-T mutation frequency. A metagene profile of the identified

residues summarizing how they localize within transcripts is also automatically generated (Fig. 2). All these steps are executed automatically with minimal input required from the user, taking ~6–8 h depending on the size of the sequencing library and available computational resources. The pipeline itself can also be scaled seamlessly to clusters and cloud servers depending on the available user environment without the need to modify the workflow.

Use of input sample to control for conversion calls not arising from m⁶A antibody

Similar to the eCLIP method on which it is based, our method includes a corresponding input sample that we use to identify C-to-T mutations that occur in the absence of anti-m⁶A antibody cross-links. A small aliquot of fragmented RNA is taken prior to introduction of the antibody and subsequent cross-linking (representing ~5% of the total sample) and prepared concurrently with the immunoprecipitated sample after the antibody is removed. Following sequencing, the same m⁶A identification workflow is performed on the input reads which effectively identifies C-to-T mutations that are not induced from anti-m⁶A antibody. The computational pipeline automatically takes these "false m⁶As" and compares them to the list of m⁶A sites obtained from the m⁶A immunoprecipitation. When multiple replicate experiments are performed on the same cell line, we create a combined list from all the input samples. By removing any positions that occur in both sets, our strategy specifically identifies adenosine residues that are recognized by the anti-m⁶A antibody.

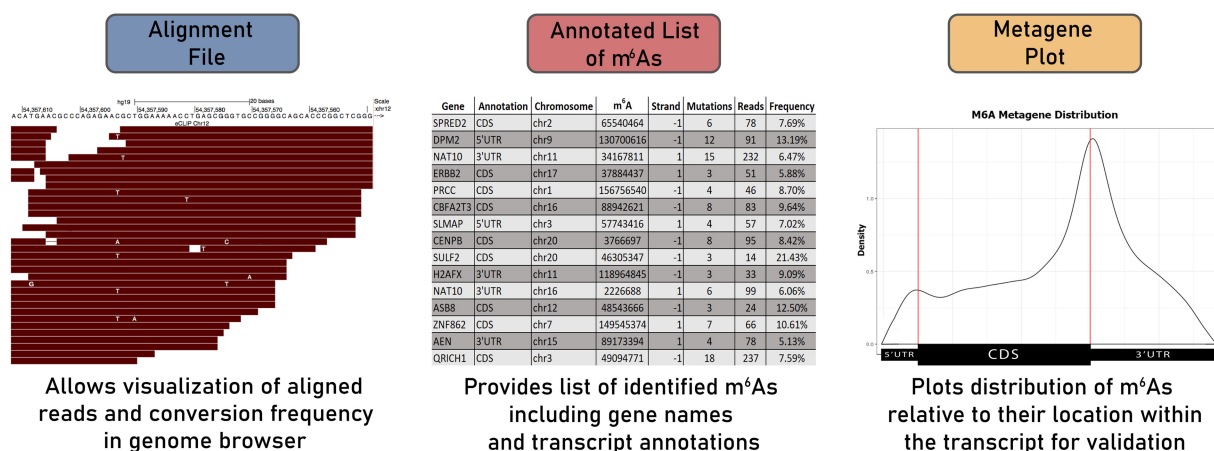


FIGURE 2. Overview of output files provided by the workflow. In addition to a summary file consisting of the relevant outputs and logs from the prerequisite software used in the workflow (not shown), an alignment file (BAM) consisting of reads that were successfully mapped to the genome is supplied for visualization of overall library quality and C-to-T conversion frequencies relative to identified m⁶A sites. A tab-delimited list of called m⁶As (sorted by confidence) with their genomic location, gene, and transcript annotation, number of supporting C-to-T mutations and relative conversion frequency are also provided (two separate BED files containing just the genomic coordinates of m⁶A residues in each confidence category are also generated for use in downstream analysis/visualization). A metagene profile summarizing where in the transcript the identified m⁶A residues occur is also automatically generated using MetaPlotR (Olarerin-George and Jaffrey 2017).

Our experiments to date have identified that such contaminating mutations account for up to 11% of the initial residues called from the immunoprecipitation (Supplemental Table 1). These input calls are likely a mixture of conversions due to some random errors across the library preparation/sequencing steps, while also including some nonrandom conversions due to SNPs and RT bias over specific sequences. This nonrandom fraction is highlighted by the use of as much as 20% of the input calls to filter out sites in the preliminary list of residues identified from the immunoprecipitated sample. Additionally, sites of high conversion rate (>50%) are much more likely to be filtered out by the input list (Fig. 3). As many of these m⁶A residues would otherwise likely remain included in the miCLIP approach, which does not use an input control, we consider this step to be a significant improvement and ultimately required for confidence in accurate identification of m⁶As. Further, as the input essentially represents an RNA-seq experiment of the same sample, it can be used to gauge overall gene expression within the cell line.

Successful identification of m⁶A sites categorized by confidence

Using the meCLIP strategy outlined above, we have successfully identified over 50,000 unique m⁶A residues in the two breast cancer cell lines that were analyzed (MCF-7 and MDA-MB-231) and over 8000 unique residues in HEK-293 cells. We find a considerable difference in the number of m⁶A sites called between experimental replicates of the same cell line (Table 1). These differences may be a result of variability in cross-linking efficiency, inconsistent frequency of reverse transcription errors, or evidence of the dynamic nature of m⁶A deposition itself.

Based on these results, we recommend calling m⁶As on replicates of the same sample and taking the consensus (“Multireplicate consensus” in Table 1) to increase confidence in generating a list of specific modifications. We also noted that the consensus among replicates increases when the C-to-T mutation frequency at a given residue is above 5%, and that the majority of raw C-to-T conversions are not located within the m⁶A “RAC” consensus motif when the mutation frequency is <5% (Fig. 3A). Grouping all the identified sites of C-to-T conversions between 1%–50% mutation frequency into bins of 2.5%, we observed that the 2.5%–4.9% bin has an “RAC” motif occurrence well below the ~50% which occurs in the next higher bin. Below 2.5%, the motif frequency trends close to the 12.5% random chance of having “RA” upstream of the C. This low motif frequency occurs above 50% conversion rate, and also occurs across all C-to-T conversion rate bins in the input sample (Supplemental Fig. 3). Further, the fraction of residues removed with our input filter was significantly higher at conversion rates >50% compared to the other conversion rate intervals (Fig. 3B). Based on these observations, we have chosen to automatically categorize our final m⁶A site calls into low and high confidence based on the frequency of the C-to-T mutation at that residue (where low confidence is ≥ 2.5 to <5% and high confidence is ≥ 5 to $\leq 50\%$, with a minimum of three reads that support the mutation). Sites we identify in the meCLIP with conversion rates above 50% also have nearly a random chance of having the “RAC” motif and the majority are filtered out with input control. For these reasons, our confidence is greatly reduced in calling these sites and we therefore do not include any sites >50% conversion frequency in the m⁶A call list. Even though the small number that remain after our multiple filters are not easily distinguished with confidence from

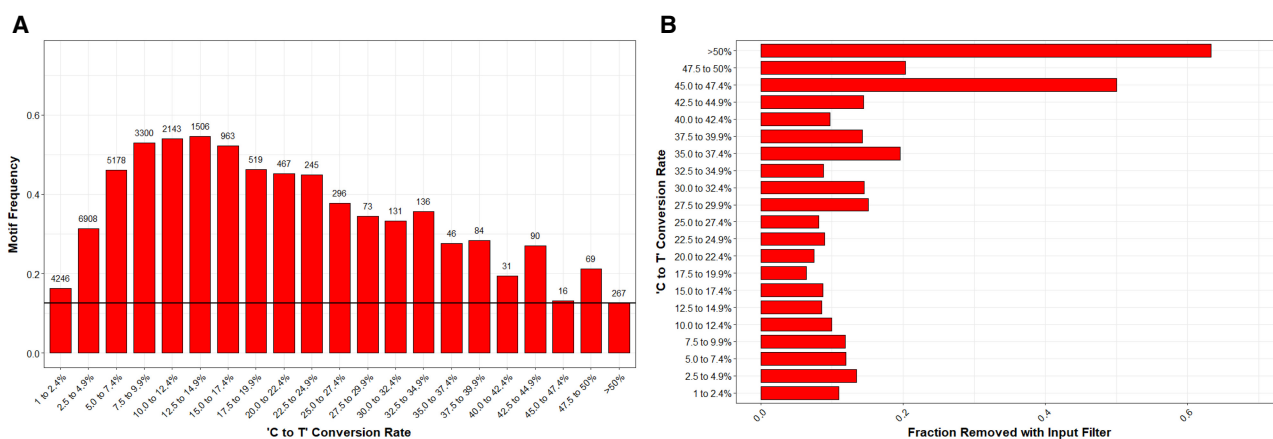


FIGURE 3. Evidence supporting use of motif and input filtering. (A) Analysis showing the occurrence of the “RAC” consensus motif relative to C-to-T mutations. Black line at 12.5% represents the random chance of having “RA” upstream of the C. The numbers above each bin indicate the total number of identified residues within each interval that occur within the consensus motif. At conversion rates of <2.5% and >50%, the occurrence of the “RAC” motif is close to what would be expected by random chance. (B) Analysis showing the fraction of residues removed with our input filter in each conversion rate interval. For residues at conversion rates >50%, the majority (63%) also occur in the input sample compared to the 14% average removal rate for conversions in the bins between 2.5% to 50%.

TABLE 1. Summary of m⁶A residues identified in three replicates of HEK-293 and breast cancer cell lines (MCF-7 and MDA-MB-231) using meCLIP

Experiment	Total m ⁶ A	Low confidence (<5%)	High confidence (≥5%)
HEK-293			
Replicate #1 ^a	4005	951	3054
Replicate #2	2358	583	1775
Replicate #3	4555	1644	2911
Multireplicate consensus	1748	344	1404
MCF-7			
Replicate #1	21,215	6644	14,571
Replicate #2	12,818	6766	6052
Replicate #3	2854	1055	1799
Multireplicate consensus	7391	1514	5877
MDA-MB-231			
Replicate #1	30,219	11,751	18,468
Replicate #2	14,885	5053	9832
Replicate #3	4837	1687	3150
Multireplicate consensus	10,939	2492	8447

^aHEK-293T.

“Multireplicate consensus” indicates the number of m⁶A residues that were called in at least two replicates. For consensus residues that were supported by different levels of confidence between replicates, if a residue had at least one supporting high confidence (≥5%) event it was deemed “High confidence,” otherwise it was categorized “Low confidence.” The consensus between replicates is significantly lower in the “Low confidence” category compared to “High confidence.”

random chance or error, these can be added to the site call list with slight changes to our pipeline, if desired.

Comparison to sites called using miCLIP

We performed meCLIP on HEK-293 cells and compared the identified m⁶As with those reported in the original miCLIP paper (Linder et al. 2015). We identified 8692 unique m⁶A sites, 1744 of which were called in at least two of our replicates. Comparing these m⁶As to those called by miCLIP, we found that 2252 (25.9%) were overlapping. Notably, almost half (784, 44.9%) of our multireplicate consensus m⁶As were overlapping with sites identified in Linder et al. (Supplemental Fig. 4A). We initially observed that the miCLIP approach identified significantly more residues compared to meCLIP. However, upon inspection of the output files for the CIMS analysis, we found that ~70% of the site calls were only supported by two conversion events (Supplemental Fig. 4B), whereas only ~40% of our site calls are supported by our minimum threshold of three supporting events and our average number of supporting events for a given m⁶A is 6–7.

To further compare our approach, we analyzed our raw meCLIP reads using the miCLIP analysis pipeline. We only compared the number of residues reported from the miCLIP CIMS approach, since it uses the same antibody and conversion-based strategy. The miCLIP CIMS pipeline identified roughly twice as many residues compared to our strategy, using the same raw reads. However, we noted that there are significant differences between the m⁶A calling methods. First, whereas our meCLIP strategy requires identified m⁶As to have a C-to-T conversion frequency of at least 2.5% and three reads supporting the mutation, the miCLIP approach only requires a 1% conversion frequency with two supporting reads (for example, only two conversions within 200 reads). Second, unlike our strategy, the miCLIP method does not require the identified residue to occur within the m⁶A consensus motif. Finally, the miCLIP strategy does not use an input filter to remove residues not induced by the anti-m⁶A antibody. We found that when we subjected the m⁶As called by miCLIP to our filtering thresholds the total number of m⁶A sites called between them was roughly equivalent, indicating that the majority of additional m⁶As reported by miCLIP were not what we consider high confidence calls (Fig. 4).

The remaining variability in identified residues is likely due to differences in the library analysis. For instance, whereas our mapping strategy utilizes both read pairs for alignment and then extracts read 2 for calling m⁶As, CTK only uses the read 2 throughout the pipeline. The aligner used is also different—our strategy uses the splicing-aware RNA aligner STAR (Dobin et al. 2013) whereas miCLIP uses NovoAlign/bwa (Li and Durbin 2009). In addition to the reduced ability of BWA to accurately map across splice-junctions, by default this aligner is more sensitive to mismatches within reads compared to STAR, and miCLIP adjusts this parameter even further to reduce the number of mismatches allowed in shorter reads. This more conservative approach often results in lower rate of uniquely mapped reads which ultimately leads to less reads being usable for calling m⁶As. The manner in which PCR duplicates are handled is also notably different. Our analysis pipeline uses UMI-tools to remove reads amplified from the same molecule in one step after alignment, while miCLIP uses custom scripts to first remove duplicate reads prealignment based on sequence alone and then collapses the UMI barcodes following mapping. Finally, the m⁶A calling algorithms themselves are distinct—our method is weighted heavily toward identifying raw mutations at a given position by comparing the alignment to the reference genome in a similar manner to how variants are typically called to identify SNPs, while the miCLIP strategy uses a custom alignment parser that extracts mismatch information from the alignment file without querying the reference. While these algorithmic differences are subtle, given

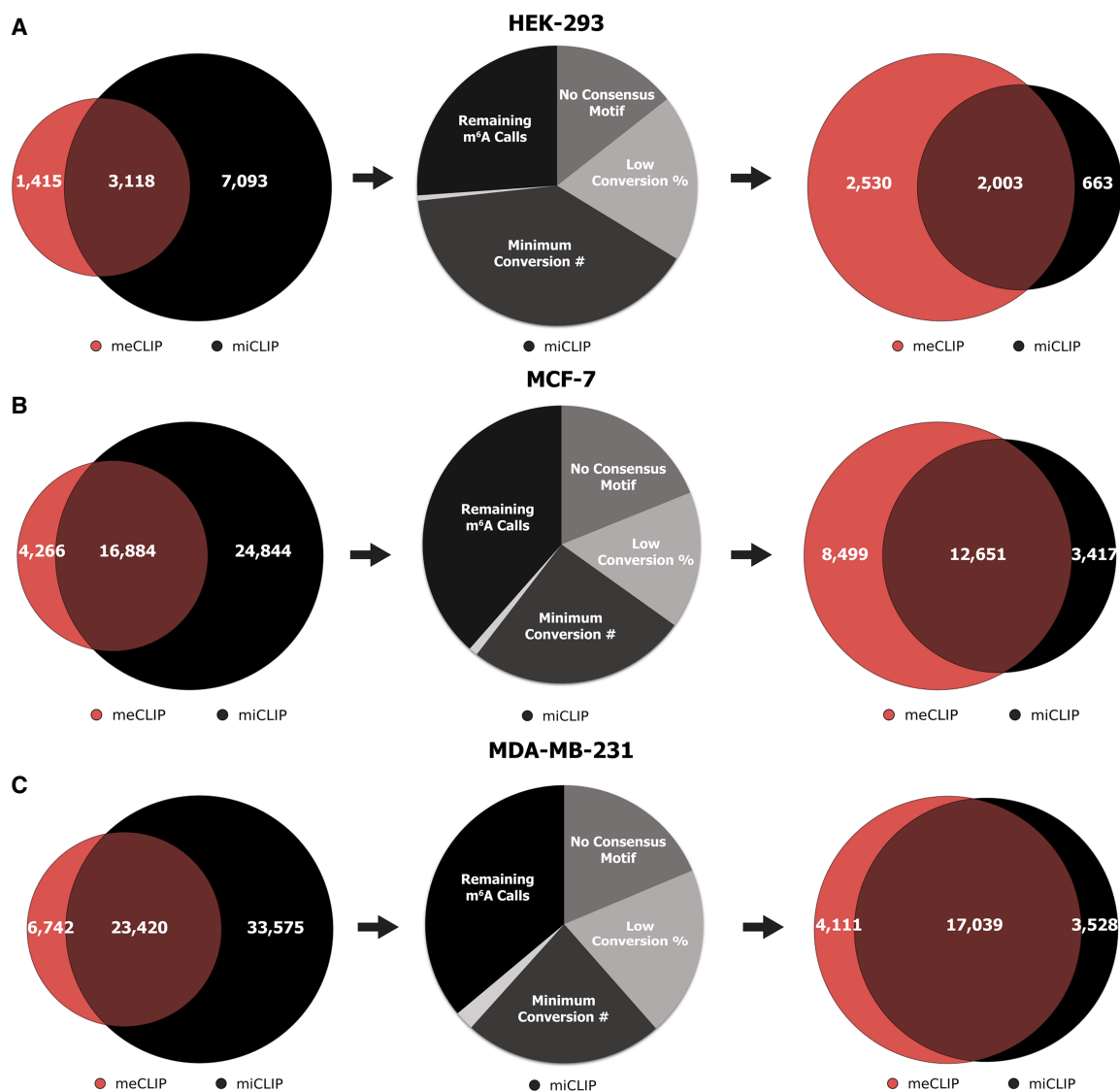


FIGURE 4. Comparison of m⁶A sites identified using the meCLIP analysis pipeline versus those identified from the miCLIP pipeline. (A) HEK-293 (Rep 3), (B) MCF-7 (Rep 1), (C) MDA-MB-231 (Rep 1). The *leftmost* Venn diagram depicts the number of m⁶As identified from both strategies using the same raw data. The *middle* shows the breakdown of m⁶As reported from miCLIP categorized by whether the residue would remain ("Remaining m⁶A Calls") if our filtering criteria were applied: If the conversion frequency is <2.5% ("Low Conversion %"), if the number of conversion events is <3 ("Minimum Conversion #"), if the conversion does not occur within the m⁶A "RAC" consensus motif ("No Consensus Motif"), and if the conversion was also identified in the combined input filter list generated from all the input samples for that cell line (unlabeled, small light gray bar). The Venn diagram on the *right* illustrates how the m⁶As correlate between the two strategies after miCLIP residues that did not meet the threshold of our identification strategy were removed. The replicate with the highest number of identified m⁶As for each cell line was chosen for this comparison.

that putative m⁶A residues are often only supported by a low number of conversion events (~10%–15% on average), slight changes in how the conversions are identified can often result in whether a residue is ultimately called. Overall, while the different approaches result in calling distinct sets of m⁶As, the increased thresholding in our meCLIP method produces a set of m⁶A site calls with confidence metrics that are useful for subsequent follow-up of specific modification events.

Experimental validation of m⁶A sites via RNA immunoprecipitation

To experimentally confirm that the identified m⁶As are present within transcripts, we performed RNA immunoprecipitation followed by RT-qPCR on a select number of residues called in MCF-7 cells. These individual m⁶A sites were chosen based on whether they were called in multiple replicates and whether they were categorized as low

or high confidence according to our thresholding metrics in order to validate the ability of our method to identify m⁶A residues across a spectrum of observations. For all the profiled m⁶As, we saw increased RNA recovered using primers specific to the identified m⁶A site versus unmodified sites within the same transcript (Fig. 5).

DISCUSSION

Research into RNA modifications over the past decade has shown that m⁶A is involved in most aspects of RNA biology (Gilbert et al. 2016; Zhao et al. 2016). This ubiquitous regulation of cellular processes underscores the need to identify m⁶A residues accurately and reliably in a high-throughput transcriptome-wide manner so that context-specific m⁶A function can be better understood. While the use of antibody-based methods to identify specific m⁶A residues has become commonplace, these strategies are still challenging. We have outlined several improvement steps in our protocol which help generate m⁶A site lists that can guide subsequent research into specifically modified RNAs. These include optimization of the RNA fragmentation step, utilization of tools that account for errors in the UMI, and employment of strategies that further reduce PCR duplication. Notably, we have also streamlined the downstream analysis steps by implementing a workflow management system that automates the process of calling m⁶As.

Although our meCLIP approach to identify m⁶A residues does offer clear advantages compared to previous strate-

gies, a frequently cited challenge for using current m⁶A identification protocols is the large amount of input mRNA required for effective immunoprecipitation and sequencing. Our experience with eCLIP suggests that the input amounts we describe for our protocol could be reduced while still producing quality results, although we have not yet systematically tested the range of adequate RNA input. Consistent with recent reports (Zeng et al. 2018), however, we find that the number of unique m⁶A sites identified does increase with higher amounts of input RNA (see recommendations for sequencing in the “Materials and Methods” section). Therefore, in addition to ensuring accurate quantification of starting material via the methods outlined in our protocol, we also highly encourage the use of multiple replicates when starting RNA material is limited to gain confidence in the identified m⁶A residues.

While the majority of m⁶A deposition in humans occurs via the METTL3/METTL14 writer complex, a subset is generated via another methyltransferase, METTL16 (Doxtader et al. 2018; Ruzskowska et al. 2018). In contrast to METTL3/METTL14-dependent m⁶As which are typically found within a “DRACH” motif and often near stop codons, m⁶A modifications generated by METTL16 do not occur within a defined sequence motif and are more often found in introns and within noncoding RNAs (Ruzskowska et al. 2018). As our m⁶A identification protocol uses the consensus motif as a filtering mechanism and specifically isolates mRNA via poly(A) selection, identification of METTL16-dependent m⁶As will likely be limited. Similarly, although our

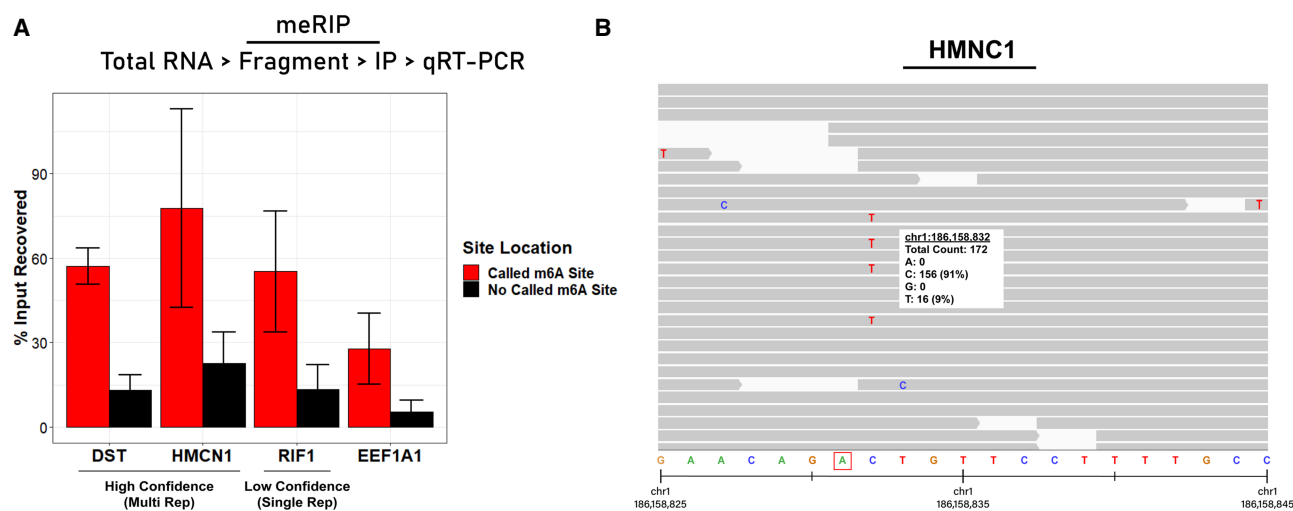


FIGURE 5. Experimental validation of select m⁶A residues. (A) RNA immunoprecipitation using anti-m⁶A antibody (meRIP) followed by RT-qPCR was used (protocol outlined in top flowchart) to confirm the presence of m⁶A residues identified in MCF-7 breast cancer cells. Residues were chosen based on whether they were called in multiple replicates and to assess low and high confidence, as indicated. Enrichment is measured as the percent of input recovered from the immunoprecipitation using primers specific to a called m⁶A site versus control primers to a nonspecific region of the same gene where no m⁶A was called. A known m⁶A site within the *EEF1A1* gene was used as a positive control. (B) Genome browser snapshot of the identified m⁶A residue in the *HMCN1* gene that was validated via meRIP. Red box surrounding the “A” at position chr1:186,158,831 indicates the m⁶A site, with the C-to-T mutations at the downstream “C” depicted in the reads (gray bars) aligning to that locus (white box depicts the number of each nucleotide at the position). See Supplemental Figure 6 for *DST* and *RIF1*.

anti-m⁶A antibody also recognizes the related RNA modification m⁶Am located on the 5' ends of mRNAs, m⁶Am is not invariably followed by a cytidine (Boulias et al. 2019) and thus could be filtered out in our pipeline as well. However, previous reports have noted that antibody-induced A-to-T transitions at the m⁶A site itself are also frequently observed. If specific modifications such as METTL16-dependent m⁶A and m⁶Am are of interest, the motif filtering step and conversion event of interest can be easily modified within the Snakemake workflow. The ease of making such ad hoc changes to the pipeline coupled with the automatic generation of data sets makes such focused identification approaches feasible and illustrates yet another benefit of our method.

In a further effort to overcome some of the limitations of previous m⁶A identification methods, recently several antibody-independent strategies have been developed to identify sites of m⁶A modifications. DART-seq (Meyer 2019) (deamination adjacent to RNA modification targets) uses a chimeric protein consisting of the YTH “m⁶A-reader” domain fused to the cytidine deaminase APOBEC1 in cells to induce C-to-U deamination events at sites adjacent to m⁶A residues and then detects the mutations using standard RNA-seq. Notably, the DART-seq method only calls for as low as 10 ng of total RNA. Another pair of methods, MAZTER-seq (Garcia-Campos et al. 2019) and m⁶A-REF-seq (Zhang et al. 2019), utilize the ability of the RNA endonuclease MazF to cleave single-stranded RNA immediately upstream of unmethylated sites occurring in “ACA” motifs, but not within their methylated “m⁶A-CA” counterparts (Imanishi et al. 2017). Finally, m⁶A identification approaches utilizing Oxford Nanopore’s direct RNA sequencing technology, including the MINES method (Lorenz et al. 2019; Price et al. 2019), have been developed to further facilitate accurate m⁶A detection. While all these methods offer unique advantages, they are not without limitations. For instance, the RNase MazF-based methods only allow for identification of the subset of m⁶As occurring within the defined motif “ACA” (estimated at ~16%–25% of methylation sites [Garcia-Campos et al. 2019]), making it more of a complementary strategy to quantify and validate select m⁶A residues rather than a standalone identification approach. The most significant barrier to such methods’ widespread utility however is the lack of a dedicated and straightforward analysis pipeline. We feel that coupling our m⁶A identification protocol with a package manager to easily install software dependencies and a workflow engine that automates the execution of each script is extremely valuable to those researchers with limited bioinformatic expertise and consider such an inclusion one of the most notable advantages to our calling method.

In summary, we have significantly improved the m⁶A CLIP library preparation to increase library complexity and introduced confidence metrics in identified m⁶A resi-

dues. We have also incorporated an easy-to-use analysis pipeline to facilitate the straightforward generation of lists and relevant figures detailing m⁶A deposition. Taken together, we believe our meCLIP approach to identify m⁶A modifications offers powerful benefits to investigators interested in deciphering the intricacies of m⁶A biology.

MATERIALS AND METHODS

RNA isolation and fragmentation assay

Cells were cultured in appropriate media and total RNA was isolated using the TRIzol (15596018, Invitrogen) method until ~1 mg of total RNA was obtained (for MCF-7 cells this was eight to ten 15 cm cell culture plates at ~80%–100% confluency). The MDA-MB-231 cells that were used for the described experiments contain a transgene that overexpresses the lncRNA HOTAIR. As noted, one replicate of HEK293s used was a HEK293-T derivative, and this line also overexpressed HOTAIR (Portoso et al. 2017). The total RNA samples were combined and diluted to make a 1 µg/µL stock solution (saving 2 µL to assess quality of RNA on TapeStation). To determine the optimal duration of fragmentation for the desired size (100–200 nt in length), 2 µg of total RNA was fragmented using RNA Fragmentation Reagents (AM8740, Ambion) for times ranging from 3 to 15 min. Following the manufacturer’s protocol, combining the appropriate amount of RNA (2 µL) with nuclease-free water to the recommended reaction volume of 9 µL, adding 1 µL 10× fragmentation reagent, incubating at 70°C for designated time, and then immediately quenching with 1 µL 1× stop reagent and placing on ice. The fragment size produced from each time point was then visualized by running a heavily diluted sample (~3 ng/µL) on an Agilent TapeStation 4200 with High Sensitivity RNA Screen Tape, and the appropriate fragmentation time for the actual poly(A) sample was approximated based on these results (see Supplemental Fig. 1).

Poly(A) selection and fragmentation

Poly(A) selection was performed using the Magnosphere Ultrapure mRNA Purification Kit (9186, Takara) where 100 µL of the magnetic beads were combined with 250 µL of total RNA (1 µg/µL), and the beads were reused four times following the manufacturer’s protocol (the volume of binding buffer that the beads were resuspended in was scaled up to match the volume of RNA, i.e., 250 µL). The isolated poly(A) RNA was collected in four 50 µL aliquots, combined, and 1 µL was used to assess the quality of mRNA selection and determine percent recovery/concentration via TapeStation. The recommended amount of input mRNA into the antibody step is 5 to 20 µg. The RNA sample was then ethanol precipitated overnight at –20°C using standard methods with GlycoBlue Coprecipitant (AM9515, Invitrogen) added to the solution for easier recovery. The precipitated poly(A) RNA was resuspended in 15 µL of nuclease-free water and fragmented at 70°C for the amount of time determined previously. To further optimize the duration, initially 20% of the poly(A) sample was fragmented and visualized on the TapeStation as described above; adjustments were then made based on the observed size, or

just repeated with 20% aliquots, if the size of the first fragmentation is deemed appropriate. A small amount (5% of total mRNA) of fragmented mRNA was saved for use as the input sample (see "Input Sample Preparation" section).

Cross-linking and immunoprecipitation of m⁶A containing transcripts

The remaining fragmented RNA was resuspended in 500 μ L Binding/Low Salt Buffer (50 mM Tris-HCl pH 7.4, 150 mM sodium chloride, 0.5% NP-40), then 2 μ L RNase inhibitor (M0314, NEB) and 10 μ L (1 mg/mL) m⁶A antibody (Abcam Ab151230) were added, and the sample was incubated on a rotator for 2 h at 4°C. The RNA:antibody sample was transferred to one well of a prechilled 12-well plate and cross-linked twice at 150 mJ/cm² (254 nm wavelength) using a Stratilinker UV Crosslinker (or equivalent). A total of 50 μ L of protein A/G magnetic beads (88803, Pierce) were aliquoted into a fresh tube and washed twice with 500 μ L Binding/Low Salt Buffer. The beads were resuspended in 100 μ L Binding/Low Salt Buffer, added to the cross-linked RNA:antibody sample, and the bead mixture was incubated at 4°C overnight with rotation. The next day, the beads were washed twice with 900 μ L High Salt Wash Buffer (50 mM Tris-HCl pH 7.4, 1 M sodium chloride, 1 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate) and once with 500 μ L Wash Buffer (20 mM Tris-HCl pH 7.4, 10 mM magnesium chloride, 0.2% Tween-20). The beads were resuspended in 500 μ L Wash Buffer.

eCLIP-seq library preparation

FastAP treatment

The beads were magnetically separated, removed from the magnet, and the supernatant was combined with 500 μ L 1 \times Fast AP Buffer (10 mM Tris pH 7.5, 5 mM magnesium chloride, 100 mM potassium chloride, 0.02% Triton X-100). The beads were then placed on the magnet for 1 min and the combined supernatant was removed. The beads were washed once with 500 μ L 1 \times Fast AP Buffer (following this wash, the input sample can be prepared concurrently following the "Input Sample Preparation" instructions below). The beads were resuspended in Fast AP Master Mix (79 μ L nuclease-free water, 10 μ L 10 \times Fast AP Buffer, 2 μ L RNase Inhibitor, 1 μ L TURBO DNase [AM2238, Invitrogen], 8 μ L Fast AP enzyme [EF0654, Thermo Scientific]) was added, and the sample was incubated at 37°C for 15 min with shaking at 1200 rpm.

PNK treatment

PNK Master Mix (224 μ L nuclease-free water, 60 μ L 5 \times PNK Buffer [350 mM Tris-HCl pH 6.5, 50 mM magnesium chloride], 7 μ L T4 PNK [EK0031, Thermo Scientific], 5 μ L RNase inhibitor, 3 μ L dithiothreitol [0.1 M], 1 μ L TURBO DNase) was added to the sample and then incubated at 37°C for another 20 min with shaking. The beads were washed once with cold 500 μ L Wash Buffer, once with cold 500 μ L Wash Buffer and then cold 500 μ L High Salt Wash Buffer combined in equal volumes, once with cold 500 μ L High Salt Wash Buffer and then cold 500 μ L Wash Buffer combined in equal volumes, and once again with cold 500 μ L Wash Buffer. The beads were resuspended in 500 μ L Wash Buffer.

3' RNA adapter ligation

The beads were magnetically separated, removed from the magnet and then 300 μ L 1 \times RNA Ligase Buffer (50 mM Tris pH 7.5, 10 mM magnesium chloride) was added to the supernatant. The beads were placed on a magnet for 1 min and then the combined supernatant was removed. The beads were washed twice with 300 μ L 1 \times RNA Ligase Buffer and then resuspended in 3' RNA Ligase Master Mix (9 μ L nuclease-free water), 9 μ L 50% PEG 8000, 3 μ L 10 \times RNA Ligase Buffer (500 mM Tris pH 7.5, 100 mM magnesium chloride), 2.5 μ L T4 RNA Ligase I (M0437, NEB), 0.8 μ L 100% DMSO, 0.4 μ L RNase Inhibitor, 0.3 μ L ATP (1 mM), and 2.5 μ L (2 μ M) each of two matched barcoded RNA adaptors (X1A and X1B). The sample was incubated at room temperature for 75 min with flicking every 10 min.

Gel loading/membrane transfer

The beads were washed once with cold 500 μ L Wash Buffer, once with cold 500 μ L Wash Buffer and then cold 500 μ L High Salt Wash Buffer combined in equal volumes, once with cold 500 μ L High Salt Wash Buffer, once with equal volumes of cold 500 μ L High Salt Wash Buffer and then 500 μ L Wash Buffer, and once with cold 500 μ L Wash Buffer. The beads were resuspended in 20 μ L Wash Buffer and 7.5 μ L 4 \times NuPAGE LDS Sample Buffer (NP0007, Invitrogen) and 3 μ L DTT (0.1 M) was added. The sample was incubated at 70°C for 10 min with shaking at 1200 rpm and then cooled on ice for 1 min. The beads were magnetically separated, and the supernatant was transferred to a new tube. The sample was run on ice at 150 V in 1 \times MOPS Buffer for 75 min on a Novex NuPAGE 4%–12% Bis-Tris Gel (NP0321, Invitrogen) with 10 μ L dilute protein ladder (12 μ L Wash Buffer, 4 μ L M-XStable protein ladder [L2011, UBPBio], 4 μ L NuPAGE Buffer) on each side. The sample was then transferred to a nitrocellulose membrane overnight on ice at 30 V. After transfer, the area of the membrane containing the antibody:RNA sample (between 20 kDa and 175 kDa) was cut and sliced into small (~2 mm) pieces and placed into an Eppendorf tube.

Antibody removal/RNA cleanup

The membrane slices were incubated with 40 μ L proteinase K (3115828001, Roche) in 160 μ L PK Buffer (100 mM Tris-HCl pH 7.4, 50 mM sodium chloride, 10 mM EDTA) at 37°C for 20 min with shaking at 1200 rpm. An equal volume of PK Buffer containing 7 M urea was added to samples and incubated at 37°C for 20 min with shaking at 1200 rpm, then 540 μ L phenol:chloroform:isoamyl alcohol (25:24:1) was added and incubated at 37°C for another 5 min with shaking at 1200 rpm. The sample was centrifuged for 3 min at max speed, and the aqueous layer was transferred to a 15 mL conical. The RNA was isolated using the RNA Clean & Concentrator-5 Kit (R1013, Zymo) according to manufacturer's instructions.

Input sample preparation

The input sample (~1 μ L) was combined with FastAP Master Mix (19 μ L nuclease-free water, 2.5 μ L 10 \times Fast AP Buffer, 2.5 μ L Fast AP Enzyme, 0.5 μ L RNase Inhibitor) and incubated at 37°C for 15 min with shaking at 1200 rpm. PNK Master Mix (45 μ L nuclease-free water, 20 μ L 5 \times PNK Buffer, 7 μ L T4 PNK, 1 μ L RNase

Inhibitor, 1 μ L dithiothreitol (0.1 M), 1 μ L TURBO DNase) was added to the sample and then incubated at 37°C for another 20 min with shaking. The RNA sample was isolated using Dynabeads MyONE Silane (37002D, Thermo Fisher Scientific). Briefly, 20 μ L of beads were magnetically separated and washed once with 900 μ L RLT Buffer (79216, Qiagen), resuspended in 300 μ L RLT Buffer, and added to the sample. The bead mixture was combined with 615 μ L 100% ethanol and 10 μ L sodium chloride (5 M), pipette mixed, and incubated at room temperature for 15 min on a rotor. The sample was placed on a magnet, the supernatant was removed, and then resuspended in 1 mL 75% ethanol and transferred to a new tube. After 30 sec the bead mixture was placed on a magnet, the supernatant was removed, and the sample was washed twice with 1 mL 75% ethanol, waiting 30 sec between each magnetic separation. After the final wash, the beads were air dried for 5 min, resuspended in 10 μ L nuclease-free water and incubated at room temperature for 5 min. The sample was magnetically separated, and the elution was transferred to a new tube (an aliquot of this elution can be taken and stored at -80°C for backup if desired). The remaining eluted sample was combined with 1.5 μ L 100% DMSO, 0.5 μ L RIL19 RNA adapter, incubated at 65°C for 2 min, and placed on ice for 1 min. The sample was then combined with 3' RNA Ligase Master Mix (8 μ L 50% PEG 8000, 2 μ L 10 \times T4 RNA Ligase Buffer [B0216L, NEB], 1.5 μ L nuclease-free water, 1.3 μ L T4 RNA Ligase I [M0437, NEB], 0.3 μ L 100% DMSO, 0.2 μ L RNase Inhibitor, 0.2 μ L ATP [1 mM]) and incubated at room temperature for 75 min with mixing by flicking the tube every ~15 min. The sample was then reisolated using Dynabeads MyONE Silane following the same procedure described above (after the beads were initially washed with RLT Buffer, the sample was resuspended in 61.6 μ L RLT Buffer instead of 300 μ L and an equal volume of 100% ethanol was added). The eluted input sample was then prepared simultaneously with the immunoprecipitated (IP) sample following the same instructions.

Reverse transcription/cDNA clean up

The samples (IP and input) were reverse transcribed using the oligonucleotide AR17 and SuperScript IV Reverse Transcriptase (18090010, Invitrogen). The resulting cDNA was treated with ExoSAP-IT Reagent (78201, Applied Biosystems) at 37°C for 15 min, followed by incubation with 20 mM EDTA and 0.1 M sodium hydroxide at 70°C for 12 min. Hydrochloric acid (0.1 M) was added to the sample to quench the reaction. The purified cDNA was isolated using Dynabeads MyONE Silane (37002D, ThermoFisher Scientific) according to manufacturer's instructions. Briefly, 10 μ L of beads were magnetically separated and washed once with 500 μ L RLT Buffer, resuspended in 93 μ L RLT Buffer, and added to the samples. The bead mixture was combined with 111.6 μ L 100% ethanol, pipette mixed, and incubated at room temperature for 5 min. The sample was placed on a magnet, the supernatant was removed, and then washed twice with 1 mL 80% ethanol, waiting 30 sec between each magnetic separation. After the final wash, the beads were air dried for 5 min, resuspended in 5 μ L Tris-HCl (5 mM, pH 7.5) and incubated at room temperature for 5 min.

5' cDNA adapter ligation

The sample was combined with 0.8 μ L rand3Tr3 oligonucleotide adaptor and 1 μ L 100% DMSO, incubated at 75°C for 2 min, and

then placed on ice for 1 min. Ligation Master Mix (9 μ L 50% PEG 8000, 2 μ L 10 \times NEB T4 RNA Ligase Buffer, 1.1 μ L nuclease-free water, 0.2 μ L 1 mM ATP, 1.5 μ L T4 RNA Ligase I) was added to the sample, mixed at 1200 rpm for 30 sec, and then incubated at room temperature overnight.

cDNA isolation and qPCR quantification

The cDNA was isolated using Dynabeads MyONE Silane following the instructions already described (5 μ L of beads per sample were used, washed with 500 μ L RLT Buffer, and resuspended in 60 μ L RLT Buffer and an equal volume of 100% ethanol). The samples were eluted in 25 μ L Tris-HCl (10 mM, pH 7.5). A 1:10 dilution of cDNA was used to quantify the sample by qPCR. Based on the resulting Cq values, a PCR reaction was run on the diluted sample using 25 μ L Q5 Hot Start PCR Master Mix (M0494S, NEB) and 2.5 μ L (20 μ M) each of two indexed primers (Illumina TruSeq Combinatorial Dual (CD) index adapters, formerly known as TruSeq HT). The sample was amplified using a range of cycles based on the Cq obtained from the qPCR (Cq - 3, Cq, Cq + 3) and then visualized on a 12% TBE polyacrylamide gel to determine the optimal amount of amplification for the final library (ideally a cycle number is chosen where the amplicon has just become visible) (Supplemental Fig. 2).

Library amplification and gel purification

The undiluted cDNA library was amplified by combining 12.5 μ L of the sample with 25 μ L Q5 Hot Start PCR Master Mix and 2.5 μ L (20 μ M) of the same indexed primers used previously (amplification for the full undiluted sample will be three cycles less than the cycle selected from the diluted sample). The PCR reaction was isolated using HighPrep PCR Clean-up System (AC-60050, MAGBIO) according to manufacturer's instructions. The final sequencing library was gel purified by combining the sample with 10 \times OrangeG DNA Loading Buffer and running on a 3% quick dissolve agarose gel containing SYBR Safe Dye (1:10,000). Following gel electrophoresis, a long wave UV lamp was used to extract DNA fragments from the gel ranging from 175 to 300 bp. The DNA was isolated using QiaQuick MinElute Gel Extraction Kit (28604, Qiagen). The purified sequencing library was analyzed via TapeStation using DNA ScreenTape (either D1000 or HS D1000) according to the manufacturer's instructions to assess for appropriate size and concentration (the final library should be between 175 and 300 bp with an ideal concentration of at least 10 nM).

Overview of Snakemake workflow

Sequencing of the cDNA libraries was primarily performed using an Illumina NovaSEQ 6000 to generate 2 \times 150 bp paired-end runs consisting of 40 million raw reads per sample (as frequency of conversions can be directly impacted by sequencing depth, we recommend a minimum of 40 million reads). The resulting reads are analyzed via a modified computational pipeline based on the original eCLIP strategy that has been converted into a Snakemake workflow (accessible via GitHub at <https://github.com/ajlabuc/meCLIP>). It can be executed according to Snakemake guidelines using a configuration file detailing the location of the respective sequencing files and relevant genomes.

Specific commands within the pipeline are as follows: the reads are initially inspected for appropriate quality using FastQC (v. 0.11.7) and the in-line unique molecular identifier (UMI) located within the ssDNA adapter (rand3Tr3) at the beginning of read 2 is extracted using UMI-tools (v. 1.0.0) to prepare the reads for downstream de-duplication. The remaining nonrandom ssDNA adapter and indexed RNA adapters are then removed using Cutadapt (v. 2.4), with any reads <18 bp being discarded. The trimmed reads are then briefly analyzed visually once more with FastQC to ensure all adapters are successfully removed. Two mapping steps are then performed using the splicing-aware RNA aligner STAR (v. 2.7.1a). First the reads are mapped to the species appropriate version of RepBase (v18.05) with any successfully mapped reads being removed from further analysis (this step ultimately leads to elimination of reads mapping to ribosomal RNA and other annotated repetitive sequences; however, if m⁶A identification within these loci are of interest then this filtering step can be turned off within the workflow and given sufficient read length many repeats should be able to be uniquely mapped). The remaining reads are then mapped to the full reference genome with only uniquely mapping reads being included in final alignment. Subsequent removal of PCR duplicates is performed with UMI-tools using the previously extracted UMIs, with the allowed error rate within the UMI itself determined by the default settings. The final alignment file is sorted and indexed and then used as input for a custom m⁶A identification algorithm (in keeping with the initial eCLIP pipeline, only read 2 is used).

Guidelines for assessing sequencing runs to ensure high-quality m⁶A identification

Uniquely mapped input reads should be ≥ 30 M for human m⁶A analysis, leading to 5%–7% filtering of m⁶As called in the immunoprecipitated (IP) sample. 20 M uniquely mapped IP reads is sufficient for m⁶A calling. Overamplification can result in poorer outcomes, although we have found that increased sequencing depth can help mitigate this effect in certain circumstances by increasing the number of usable reads. Ideally, $\geq 50\%$ of mapped reads should be unique (unique barcode, retained after UMI-based de-duplication).

m⁶A identification algorithm

Putative m⁶A residues are identified using a custom analysis pipeline that utilizes the “mpileup” command of SAMtools (v. 1.9) to identify variations from the reference genome at single-nucleotide resolution across the entire genome. An internally developed Java package is then used to identify C-to-T mutations occurring (i) within the m⁶A consensus motif “RAC”: “R” is any purine, A or G; A being the methylated adenosine; and C where the mutation occurs; and (ii) within a set frequency threshold of $\geq 2.5\%$ and $\leq 50\%$ of the total reads at a given position (with a minimum of three C-to-T mutations at a single site). The broader consensus motif “DRACH,” where “D” denotes A, G, or U, and “H” denotes A, C, or U can also be used for greater selectivity by modifying the configuration file. The resulting m⁶A sites are then automatically compared to those identified in the corresponding input sample and any sites occurring in both are removed from the final list of

m⁶As (this eliminates any mutations that are not directly induced from the anti-m⁶A antibody cross-linking).

Previous iCLIP-based m⁶A identification strategies (Linder et al. 2015) have used cross-linking-induced truncations (CITS) to further identify m⁶A sites based on the observation that reverse transcription often terminates at the RNA:antibody cross-link site. While eCLIP does maintain the ability to identify these events at single-nucleotide resolution via ligation of the ssDNA rand3Tr3 adapter to the cDNA fragments at their 3' ends, we do not often see this event in our meCLIP strategy (possibly due to increased fidelity of the reverse transcriptase used (SuperScript IV) compared to the older version (SuperScript III) or our use of a single antibody (Abcam) compared to others (Synaptic Systems) that are more prone to induce truncations). Therefore, our identification strategy does not include truncation-based identification of m⁶As.

m⁶A RNA immunoprecipitation (meRIP) and RT-qPCR

Total RNA from three biological replicates of MCF-7 cells was isolated with TRIzol (15596018, Invitrogen) according to the manufacturer's instructions. RNA was diluted to 1 $\mu\text{g}/\mu\text{L}$ and fragmented with RNA Fragmentation Reagents (AM8740, Invitrogen) at 70°C for 5 min. A small aliquot of fragmented RNA (500 ng) was reserved in 10 μL nuclease-free water for use as the input sample in qRT-PCR normalization. Protein A/G Magnetic Beads (88803, Pierce) were washed twice with IP Buffer (20 mM Tris pH 7.5, 140 mM sodium chloride, 1% NP-40, 2 mM EDTA) and coupled with anti-m⁶A antibody (ab151230, Abcam) or an IgG control (NB810-56910, Novus) for 1 h at room temperature. The beads were washed three times with IP Buffer, then 10 μg fragmented RNA and 400U RNase inhibitor was added to 1 mL IP Buffer. The antibody-coupled beads were resuspended in 500 μL RNA mixture and incubated 2 h to overnight at 4°C on a rotor. The beads were then washed three times with cold IP Buffer. Samples were eluted with 200 μL Elution Buffer (1 \times IP Buffer containing 10 U/ μL RNase inhibitor and 0.5 mg/mL N⁶-methyladenosine 5'-monophosphate [M2780, Sigma-Aldrich]) for 2 h at 4°C on a rotor. The supernatant was removed, and ethanol precipitated with 2.5 M Ammonium acetate, 70% Ethanol, and 50 $\mu\text{g}/\text{mL}$ GlycoBlue Coprecipitant (Invitrogen AM9515). The RNA was washed with 70% ethanol, dried for 10 min at room temperature, and then resuspended in 10 μL nuclease-free water. The RNA was reverse transcribed using SuperScript IV Reverse Transcriptase (18090010, Invitrogen) and quantified by qPCR using primers specific to an identified m⁶A site or region with no m⁶A in the same gene. Percent input calculation was performed based on the resulting Cq values.

DATA DEPOSITION

All sequencing reads and final m⁶A calls for each of the cell lines tested have been uploaded to NCBI and are available in the Gene Expression Omnibus (GEO) database as GSE147440. The m⁶A calls from the Linder et al. miCLIP paper that were used in our comparative analysis can be found at GSE63753.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We would like to thank K. Riemondy, R. Sheridan, N. Mukherjee, and M. Taliaferro for helpful comments on the manuscript. We acknowledge the University of Colorado Cancer Center Genomics Core (supported by National Institutes Health [NIH] grant P30-CA46934) for technical support. This work was supported by NIH grants T32GM008730 and F31CA247343 (J.T.), T32CA190216 (A.M.P.), and R35GM119575 (A.M.J.); a Department of Defense Breast Cancer Research Program Breakthrough Fellowship Award W81XWH-18-1-0023 (A.M.P.), and a seed grant from the University of Colorado School of Medicine RNA Bioscience Initiative.

Received December 4, 2020; accepted December 18, 2020.

REFERENCES

- Akichika S, Hirano S, Shichino Y, Suzuki T, Nishimasu H, Ishitani R, Sugita A, Hirose Y, Iwasaki S, Nureki O, et al. 2019. Cap-specific terminal N⁶-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* **363**: eaav0080. doi:10.1126/science.aav0080
- Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**: 11. doi:10.1186/s13100-015-0041-9
- Boulias K, Toczyłowska-Socha D, Hawley BR, Liberman N, Takashima K, Zaccara S, Guez T, Vasseur JJ, Debart F, Aravind L, et al. 2019. Identification of the m⁶A_m methyltransferase PCIF1 reveals the location and functions of m⁶A_m in the transcriptome. *Mol Cell* **75**: 631–643.e8. doi:10.1016/j.molcel.2019.06.006
- Chen K, Lu Z, Wang X, Fu Y, Luo GZ, Liu N, Han D, Dominissini D, Dai Q, Pan T, et al. 2015. High-resolution N⁶-methyladenosine (m⁶A) map using photo-cross-linking-assisted m⁶A sequencing. *Angew Chem Int Ed Engl* **54**: 1587–1590. doi:10.1002/anie.201410647
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635
- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al. 2012. Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* **485**: 201–206. doi:10.1038/nature11112
- Doxtader KA, Wang P, Scarborough AM, Seo D, Conrad NK, Nam Y. 2018. Structural basis for regulation of METTL16, an S-adenosylmethionine homeostasis factor. *Mol Cell* **71**: 1001–1011.e4. doi:10.1016/j.molcel.2018.07.025
- García-Campos MA, Edelheit S, Toth U, Safra M, Shachar R, Viukov S, Winkler R, Nir R, Lasman L, Brandis A, et al. 2019. Deciphering the 'm⁶A code' via antibody-independent quantitative profiling. *Cell* **178**: 731–747.e16. doi:10.1016/j.cell.2019.06.013
- Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, Hershkovitz V, Peer E, Mor N, Manor YS, et al. 2015. m⁶A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**: 1002–1006. doi:10.1126/science.1261417
- Gilbert WV, Bell TA, Schaening C. 2016. Messenger RNA modifications: form, distribution, and function. *Science* **352**: 1408–1412. doi:10.1126/science.aad8711
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Jr AM, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129–141. doi:10.1016/j.cell.2010.03.009
- Imanishi M, Tsuji S, Suda A, Futaki S. 2017. Detection of N⁶-methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. *Chem Commun* **53**: 12930–12933. doi:10.1039/c7cc07699a
- Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, Haripal B, Zucker-Scharff I, Moore MJ, Park CY, et al. 2015. A majority of m⁶A residues are in the last exons, allowing the potential for 3'UTR regulation. *Genes Dev* **29**: 2037–2053. doi:10.1101/gad.269415.115
- Ke S, Pandya-Jones A, Saito Y, Fak JJ, Vågbø CB, Geula S, Hanna JH, Black DL, Darnell JE, Darnell RB. 2017. m⁶A mRNA modifications are deposited in nascent Pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev* **31**: 990–1006. doi:10.1101/gad.301036.117
- Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**: 2520–2522. doi:10.1093/bioinformatics/bts480
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760. doi:10.1093/bioinformatics/btp324
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079. doi:10.1093/bioinformatics/btp352
- Licalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464–469. doi:10.1038/nature07488
- Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. 2015. Single-nucleotide-resolution mapping of m⁶A and m⁶A_m throughout the transcriptome. *Nat Methods* **12**: 767–772. doi:10.1038/nmeth.3453
- Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, et al. 2014. A METTL3–METTL14 complex mediates mammalian nuclear RNA N⁶-adenosine methylation. *Nat Chem Biol* **10**: 93–95. doi:10.1038/nchembio.1432
- Liu J, Dou X, Chen C, Chen C, Liu C, Xu MM, Zhao S, Shen B, Gao Y, Han D, et al. 2020. N⁶-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science* **367**: 580–586. doi:10.1126/science.aay6018
- Lorenz DA, Sathe S, Einstein JM, Yeo GW. 2019. Direct RNA sequencing enables m⁶A detection in endogenous transcript isoforms at base specific resolution. *RNA* **26**: 19–28. doi:10.1261/rna.072785.119
- Meyer KD. 2019. DART-Seq: an antibody-free method for global m⁶A detection. *Nat Methods* **16**: 1275–1280. doi:10.1038/s41592-019-0570-0
- Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3'-UTRs and near stop codons. *Cell* **149**: 1635–1646. doi:10.1016/j.cell.2012.05.003
- Olarerin-George AO, Jaffrey SR. 2017. MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. *Bioinformatics* **33**: 1563–1564. doi:10.1093/bioinformatics/btx002
- Patil DP, Chen C-KK, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR. 2016. m⁶A RNA methylation promotes XIST-mediated

- transcriptional repression. *Nature* **537**: 369–373. doi:10.1038/nature19342
- Portoso M, Ragazzini R, Brenčić Ž, Moiani A, Michaud A, Vassilev I, Wassef M, Servant N, Sargueil B, Margueron R. 2017. PRC2 is dispensable for HOTAIR mediated transcriptional repression. *EMBO J* **36**: 981–994. doi:10.15252/emboj.201695335
- Price AM, Hayer KE, McIntyre ABR, Gokhale NS, Abebe JS, Della Fera AN, Mason CE, Horner SM, Wilson AC, Depledge DP, et al. 2020. Direct RNA sequencing reveals m⁶A modifications on adenovirus RNA are necessary for efficient splicing. *Nat Commun* **11**: 6016. doi:10.1038/s41467-020-19787-6
- Roundtree IA, Luo GZ, Zhang Z, Wang X, Zhou T, Cui Y, Sha J, Huang X, Guerrero L, Xie P, et al. 2017. YTHDC1 mediates nuclear export of N⁶-methyladenosine methylated MRNAs. *Elife* **6**: e31311. doi:10.7554/eLife.31311
- Ruszkowska A, Ruszkowski M, Dauter Z, Brown JA. 2018. Structural insights into the RNA methyltransferase domain of METTL16. *Sci Rep* **8**: 1–13. doi:10.1038/s41598-018-23608-8
- Sendinc E, Valle-Garcia D, Dhall A, Chen H, Henriques T, Navarrete-Perea J, Sheng W, Gygi SP, Adelman K, Shi Y. 2019. PCIF1 catalyzes m⁶A_m mRNA methylation to regulate gene expression. *Mol Cell* **75**: 620–630.e9. doi:10.1016/j.molcel.2019.05.030
- Shah A, Qian Y, Weyn-Vanhentenryck SM, Zhang C. 2017. CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics* **33**: 566–567. doi:10.1093/bioinformatics/btw653
- Smith T, Heger A, Sudbery I. 2017. UMI-Tools: modelling sequencing error in unique molecular identifiers to improve quantification. *Genome Res* **27**: 491–499. doi:10.1101/gr.209601.116
- Sun H, Zhang M, Li K, Bai D, Yi C. 2019. Cap-specific, terminal N⁶-methylation by a mammalian m⁶A_m methyltransferase. *Cell Res* **29**: 80–82. doi:10.1038/s41422-018-0117-4
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, et al. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (ECLIP). *Nat Methods* **13**: 1–9. doi:10.1038/nmeth.3810
- Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al. 2014a. N⁶-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**: 117–120. doi:10.1038/nature12730
- Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC. 2014b. N⁶-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**: 191–198. doi:10.1038/ncb2902
- Wu B, Li L, Huang Y, Ma J, Min J. 2017. Readers, writers and erasers of N⁶-methylated adenosine modification. *Curr Opin Struct Biol* **47**: 67–76. doi:10.1016/j.sbi.2017.05.011
- Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, Sun HY, Li A, Ping XL, Lai WY, et al. 2016. Nuclear m⁶A reader YTHDC1 regulates mRNA splicing. *Mol Cell* **61**: 507–519. doi:10.1016/j.molcel.2016.01.012
- Zeng Y, Wang S, Gao S, Soares F, Ahmed M, Guo H, Wang M, Hua JT, Guan J, Moran MF, et al. 2018. Refined RIP-Seq protocol for epitranscriptome analysis with low input materials. *PLoS Biol* **16**: 1–20. doi:10.1371/journal.pbio.2006092
- Zhang Z, Chen LQ, Zhao YL, Yang CG, Roundtree IA, Zhang Z, Ren J, Xie W, He C, Luo GZ. 2019. Single-base mapping of m⁶A by an antibody-independent method. *Sci Adv* **5**: 6. doi:10.1016/j.jmb.2019.09.021
- Zhao BS, Roundtree IA, He C. 2016. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **18**: 31–42. doi:10.1038/nrm.2016.132