

ARTICLE

<https://doi.org/10.1038/s41467-019-13146-w>

OPEN

Antibody cross-reactivity accounts for widespread appearance of m¹A in 5'UTRs

Anya V. Grozhik^{1,2}, Anthony O. Olarerin-George ^{1,2}, Miriam Sindelar¹, Xing Li¹, Steven S. Gross¹ & Samie R. Jaffrey ^{1*}

N¹-methyladenosine (m¹A) was proposed to be a highly prevalent modification in mRNA 5'UTRs based on mapping studies using an m¹A-binding antibody. We developed a bioinformatic approach to discover m¹A and other modifications in mRNA throughout the transcriptome by analyzing preexisting ultra-deep RNA-Seq data for modification-induced misincorporations. Using this approach, we detected appreciable levels of m¹A only in one mRNA: the mitochondrial *MT-ND5* transcript. As an alternative approach, we also developed an antibody-based m¹A-mapping approach to detect m¹A at single-nucleotide resolution, and confirmed that the commonly used m¹A antibody maps sites to the transcription-start site in mRNA 5'UTRs. However, further analysis revealed that these were false-positives caused by binding of the antibody to the m⁷G-cap. A different m¹A antibody that lacks cap-binding cross-reactivity does not show enriched binding in 5'UTRs. These results demonstrate that high-stoichiometry m¹A sites are exceedingly rare in mRNAs and that previous mappings of m¹A to 5'UTRs were the result of antibody cross-reactivity to the 5' cap.

¹Department of Pharmacology, Weill Cornell Medicine, New York, NY 10065, USA. ²These authors contributed equally: Anya V. Grozhik, Anthony O. Olarerin-George. *email: srj2003@med.cornell.edu

The initial concept of the epitranscriptome was born with the transcriptome-wide mapping of thousands of internally located modified nucleotide N^6 -methyladenosine (m^6A) residues in the transcriptome^{1,2}. Two studies later identified N^1 -methyladenosine (m^1A) as another abundant epitranscriptomic modification^{3,4}. Both studies mapped m^1A in thousands of mRNAs by sequencing mRNA fragments immunoprecipitated with a monoclonal antibody (clone AMA-2) commercially distributed by MBL Bioscience, which was originally raised against KLH-conjugated 1-methyladenosine⁵. This antibody was previously shown to recognize m^1A -containing RNAs⁶. One study estimated the average stoichiometry of mapped m^1A sites at 20%³. Notably, most m^1A sites were located near start codons and proposed to provide a novel form of translational regulation³.

Subsequent work reported different distributions for m^1A , one arguing that m^1A was exceptionally rare in mRNA⁷. In that study, the antibody-bound RNA was reverse transcribed with an enzyme that efficiently introduces misincorporations at m^1A . Using this approach, m^1A was rarely observed in the RNA immunoprecipitated with m^1A antibodies⁷. Although mRNA fragments from 5'UTRs and start codon-proximal regions were immunoprecipitated, these fragments did not generate misincorporations. Thus it was concluded that mRNA fragments from the 5'UTR may be nonspecifically enriched during immunoprecipitation⁷. Ultimately it was concluded that only two mRNAs contained high-confidence m^1A sites: *C9orf100* and *MT-ND5*, a cytosolic and a mitochondrial mRNA, respectively⁷. Twelve other sites were detected at very low stoichiometry.

The second study mapped m^1A to 740 sites, 473 of which were in mRNA and lncRNA⁸. In mRNAs, the majority of sites were found in the 5'UTR; 22 of which the authors localized to the first nucleotide of the transcript. Based on this location, it was proposed that m^1A forms a novel cap structure in which m^1A immediately follows the 7-methylguanosine (m^7G) cap of mRNA (m^7G -ppp- m^1A). A re-analysis of these data showed that many of the sites that were mapped internally within the 5'UTR were actually transcription-start sites⁹.

It remained unclear why those studies produced divergent m^1A maps, and if m^1A exists at transcription-start sites or start codons, or neither, and why these particular sites are so prominent in m^1A -mapping studies. Additionally, whether m^1A sites are present with high stoichiometry as initially reported³, or low stoichiometry and rare⁷ also remained to be resolved.

Here, to address the question of the prevalence and location of m^1A in the transcriptome, we used both a high-resolution m^1A -mapping method as well as a bioinformatic approach, termed “misincorporation mapping”. Misincorporation mapping takes advantage of the ability of m^1A and numerous other modified nucleotides to induce misincorporations during the reverse transcription step common to most RNA-Seq protocols. By probing several ultra-deep RNA-Seq datasets for such misincorporations, we discovered that very few mRNAs contain misincorporations. Only the *MT-ND5* mitochondrial transcript and the *MALAT1* noncoding RNA generated statistically significant misincorporations, demonstrating the rarity of high stoichiometry m^1A sites. To understand why misincorporation mapping identified only a few m^1A sites while m^1A antibody-based mapping detects many, we mapped m^1A at high resolution using the same m^1A -directed antibody used in all previous studies. This mapping recapitulated the selective binding of the AMA-2 m^1A antibody to transcription-start nucleotides in mRNA. However, we also found that this m^1A antibody recognizes the m^7G cap structure, and that m^1A -independent binding explains why previous maps showed m^1A in mRNA 5'UTR regions. To further confirm this observation, we demonstrate that a different m^1A antibody, which we show does not bind the m^7G

cap, produces an m^1A map that no longer enriches for the 5' end of mRNAs. Overall, our data demonstrate that (1) m^1A and other hard stop nucleotides are rare in mRNA; (2) that—with the exception of *MT-ND5*— m^1A sites have very low stoichiometry; and (3) that cross-reactivity of the AMA-2 m^1A antibody with 5' caps leads to false-positive localization of m^1A to transcription-start nucleotides and start codons.

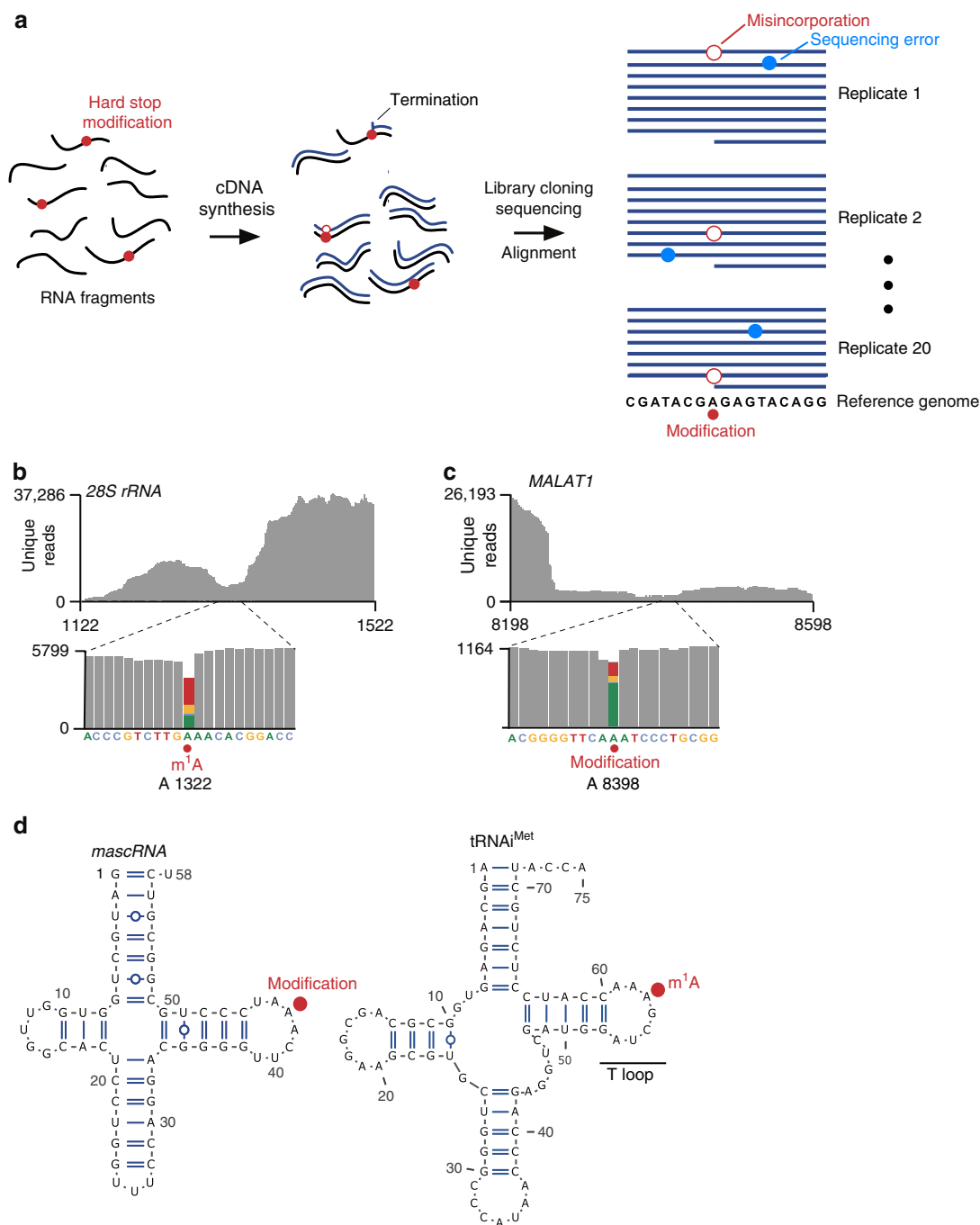
Results

Misincorporation mapping using ultra-deep RNA-seq. Given the inconsistency in the different antibody-dependent m^1A -mapping methods (Supplementary Fig. 1a–c), we sought to use an antibody-independent approach to detect m^1A at single-nucleotide resolution in mRNA. For this, we took advantage of existing ultra-deep RNA-seq datasets and the fact that m^1A is a “hard-stop” modification, meaning it typically arrests cDNA synthesized by standard reverse transcriptases^{10,11} (Supplementary Fig. 1d). However, SuperScript III will read through m^1A and other hard-stop nucleotides at low frequency, resulting in misincorporations that are variable and sequence dependent^{10,11}. Most m^1A -induced misincorporations are A→T transitions that can be detected by sequencing the cDNA^{10,11}. This approach can detect other hard-stop modifications, such as 3-methylcytidine (m^3C), 3-methyluridine, N^2,N^2 -dimethylguanosine, and N^6,N^6 -dimethyladenosine since these also produce misincorporations¹⁰. Therefore, misincorporations can directly localize m^1A and other hard-stop nucleotides in sequencing data^{10,11}.

Misincorporations are difficult to distinguish from sequencing errors using standard next-generation sequencing for two major reasons. First, substantial read depth is required to detect m^1A since m^1A typically induces a misincorporation in only approximately 20–30% of the cDNAs generated by SuperScript III (refs. ^{10,11}). The misincorporations would therefore be particularly difficult to detect for low stoichiometry m^1A residues. Second, misincorporations cannot be readily distinguished from stochastic errors originating during PCR amplification or during sequencing. Thus, m^1A cannot be definitively identified in standard RNA-seq experiments.

To overcome these problems, we developed a bioinformatic approach similar to high-throughput annotation of modified ribonucleotides (HAMR)¹⁰ to distinguish modified nucleotides from sequencing errors (Fig. 1a). We used an ultra-deep RNA-seq dataset from blood mononucleocytes comprising approximately three billion reads derived from 20 independent sequencing experiments (“replicates”)¹². These replicates were derived from a single human donor whose genome was sequenced, allowing any differences between the cDNA and genome to be readily detected. As with most RNA-seq datasets, the exact reverse transcriptase termination site is not detectable¹² (see Supplementary Fig. 1a). Instead, we localized hard-stop modifications by identifying all nucleotide positions in the transcriptome that showed misincorporations across multiple replicates (see Methods). Notably, for m^1A , we searched for A→T transitions alongside other less common transition types induced by m^1A ^{10,11} (see Methods). Importantly, this method only reveals misincorporations, not the identity of the modification; the identity would have to be determined by biochemical methods.

We first confirmed that we could detect known m^1A sites. After aligning reads to rRNA, we readily detected the known 28S rRNA m^1A at position 1322. (Fig. 1b, Supplementary Fig. 2a). As expected, the misincorporations were predominantly A→T transitions, which are characteristic of m^1A ^{10,11}. These site-specific misincorporations were detected in all 20 replicates, confirming that the A→T transitions were not stochastic sequencing errors.



Misincorporation mapping can detect other hard-stop modifications in rRNA, including 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine and m³U (Supplementary Fig. 2b). However, modifications that do not significantly affect reverse transcription, such as m⁶A, pseudouridine, N⁴-acetylcytidine, 2'-O-methylated nucleotides, and m⁷G, did not induce misincorporations (Supplementary Fig. 2c).

We considered the possibility that m¹A detection could be impaired because m¹A can convert to m⁶A through the Dimroth rearrangement, a heat and base-catalyzed reaction¹³ (Supplementary Fig. 2d). To estimate m¹A loss during the preparation of the ultra-deep sequencing libraries, we examined the m¹A at position 1322 in the 28S rRNA, which is methylated at near complete stoichiometry¹⁴. Since reverse transcription of m¹A results in read-through approximately 20–30% of the time^{10,11}, the fraction of read-through events can suggest the overall m¹A

stoichiometry. Notably, we found that m¹A at this position was associated with a ~15% read-through rate in this dataset (see Supplementary Fig. 2a). This suggests that the library preparation protocol did not cause substantial degradation of m¹A, and m¹A residues should be detectable throughout the transcriptome using this dataset.

m¹A is not readily detected in mRNA. In order to detect m¹A, the modified residue must be reverse transcribed a sufficient number of times during library preparation to generate misincorporations. Since m¹A sites were reported to have on average a 20% stoichiometry³, we set a threshold of 500 unique reads on any given nucleotide to detect m¹A sites. At this stoichiometry, 100 reverse transcription events would encounter m¹A. Of these 100 reverse transcription events, approximately 20% would read

Fig. 1 Misincorporation mapping identifies known and novel modifications. **a** Schematic of misincorporation mapping. An ultra-deep RNA-seq dataset was derived from 20 independent biological replicates. Conceivably, some of the RNA fragments (black lines) contain “hard-stop modifications” (red circles), and RNA fragments were reverse transcribed (cDNA, blue lines). For RNA fragments containing these modifications, the reverse transcriptase could terminate at the modification, produce a nucleotide misincorporation, or read through the modification. When the cDNA library is PCR-amplified and sequenced, nucleotide differences are detected between the aligned reads and genomic sequence matching the RNA-seq dataset. Misincorporations (open red circles) identified using this approach are indicative of putative modification sites. Examination of these misincorporations across multiple replicates is critical to distinguish modification sites and sequencing error, which occurs more dispersedly (blue circles). **b** m¹A in the 28S rRNA is detected by misincorporation mapping. To determine if known hard-stop nucleotide modifications are detected using misincorporation mapping in ultra-deep RNA-seq, we evaluated mapped sequence reads (gray) around a known m¹A site in the 28S rRNA (upper panel). Approximately 70% of the nucleotides that mapped to the m¹A position contained misincorporations (lower panel; colored bar, position of modified nucleotide and corresponding misincorporations). Additionally, most misincorporations were A→T transitions, typical of m¹A. **c** Misincorporation mapping detects a modified adenosine in *MALAT1*. To identify novel modified sites, we analyzed misincorporations in lncRNA and mRNA. A high-confidence site was identified in the lncRNA *MALAT1*. More than 25% of reads (gray) mapping to an adenosine in this RNA contained misincorporations (colored bar, position of modified nucleotide and corresponding misincorporations). **d** The modification in *MALAT1* described in **c** occurs in tRNA-like structure at a position that corresponds to the m¹A position in tRNAs. This region of *MALAT1* is known to be processed to yield a tRNA-like small RNA called *mascrRNA*¹⁶ (left), which is similar to human tRNAs (right, tRNA^{Met} is shown as an example). The location of the modified adenosine in *mascrRNA* is analogous to the position of m¹A in the T-loop structure of tRNA. This result suggests that the modified adenosine in *mascrRNA* is likely to contain m¹A

through, and most of these would be associated with a misincorporation^{10,11}. At this read depth, misincorporations should therefore be readily detected in multiple replicates. Thus, to detect m¹A in mRNA, we restricted our search to approximately eight million adenosine residues in the transcriptome that showed a read depth of >500 reads (Supplementary Fig. 3a).

Analysis of the three billion reads showed 14 high-confidence nucleotide positions across the transcriptome with misincorporations in more than one replicate (see Methods). Of these, 12 occurred at adenosine residues (Supplementary Data 1 and 2). Most of these modified adenines were found in mitochondrial tRNAs and occurred at known m¹A positions in mitochondrial tRNAs¹⁵ (Supplementary Data 2). We also detected a modified adenosine in *mascrRNA* (*MALAT1*-associated small cytoplasmic RNA), a short tRNA-like ncRNA that is derived from endonucleolytic processing of *MALAT1* (ref. 16) (Fig. 1c). Notably, this modified adenosine corresponds to position 58 within the T-loop of tRNAs (Fig. 1d), a conserved m¹A site in tRNAs¹⁷. This m¹A site in *MALAT1* may be similarly formed by T-loop-specific m¹A-synthesizing enzymes¹⁷.

Besides these noncoding RNAs, the previously reported⁷ m¹A-containing *MT-ND5* mitochondrial mRNA also contained a modified adenosine (Supplementary Data 2). This adenosine exhibited a misincorporation rate of 13.5%. Thus, misincorporation mapping resulted in the same major conclusion as previously reported⁷ that *MT-ND5* is the major m¹A-modified mRNA in the cell.

We next examined misincorporations at other mitochondrial mRNAs with previously annotated m¹A sites. Safra et al.⁷ and Li et al.⁸ identified 11 and 5 putative mitochondrial m¹A-containing protein-coding genes, respectively. Four mitochondrial mRNAs were common to both studies (*MT-ND5*, *MT-CO1*, *MT-CO2*, and *MT-CO3*). The misincorporation rates in poly(A) RNA-Seq for these mRNAs were very low (less than 0.7% and 2.1% in the Safra and Li studies^{7,8}, respectively). However, misincorporations could be detected when m¹A-containing mRNAs were enriched using the m¹A antibody, suggesting that m¹A-containing transcripts are indeed present in cells, but are so rare that they require enrichment to be detected. We therefore leveraged the exceptional mitochondrial read depth in our ultra-deep RNA-seq samples (average ~1 million reads/nucleotide). Here we found that with the exception of *MT-ND5* which had a misincorporation rate of 13.5%, the misincorporation rates for all other putative m¹A sites was less than 0.4%, which is close to the background rate (Supplementary Fig. 3b, Supplementary Data 3). This demonstrates that the baseline stoichiometry of m¹A is very

low in all mitochondrial mRNAs except for *MT-ND5*, and m¹A detection requires a pre-enrichment step due to its exceptionally low stoichiometry.

We also detected three additional sites of modifications in cytosolic mRNAs, with only one site being a modified adenosine (Supplementary Data 2). Therefore, it is unlikely that there are many high-stoichiometry hard-stop nucleotides in mRNA.

Notably, when we examined cytosolic mRNAs reported³ to have the highest stoichiometry of m¹A (i.e. >50%) such as *CCDC71*, *DLST*, and *STK16*, each lacked A→T transitions based on misincorporation mapping (Supplementary Fig. 3c, Supplementary Data 1).

m³C is also unlikely to be present at high stoichiometry in mRNA since we could not detect any high-confidence misincorporations at cytidine residues (Supplementary Data 2). The m³C previously detected in mRNA by mass spectrometry¹⁸ may therefore have originated from contaminating tRNA.

Although misincorporation mapping demonstrated that few mRNAs have m¹A or other hard-stop modifications, an important caveat is that the 5' end of mRNAs cannot be assessed. This is because RNA-seq typically provides less coverage at 5' ends of RNAs (see Supplementary Fig. 1a)¹⁹. Overall, the paucity of m¹A sites in mRNA using misincorporation mapping demonstrates that m¹A is not a prevalent high-stoichiometry modification in mRNA.

m¹A-miCLIP detects known m¹A sites at nucleotide resolution. To understand why m¹A antibody-based mapping approaches produce a prominent 5'UTR signal, we developed an approach to detect m¹A at single-nucleotide resolution: m¹A-miCLIP (m¹A-modification individual nucleotide resolution crosslinking and immunoprecipitation) (Fig. 2a). We initially used the AMA-2 m¹A antibody previously used to map m¹A sites in mRNAs^{3,4}. In m¹A-miCLIP, the m¹A antibody is crosslinked to sheared RNA (Fig. 2a). UV crosslinking with stringent washing reduces nonspecific RNA binding and increases peak resolution in mapping studies²⁰. RNA fragments crosslinked to the antibody are then purified and cloned as a cDNA library. Terminations introduced during reverse transcription are then analyzed to localize precise sites where the m¹A antibody binds throughout the transcriptome. Prior to performing transcriptome-wide m¹A mapping, we confirmed the AMA-2 m¹A antibody binds m¹A, and not other nucleotides (Supplementary Fig. 4a).

m¹A-miCLIP differs from earlier methods^{3,4} by preserving the cDNA 3' ends. Any m¹A antibody-binding site at the

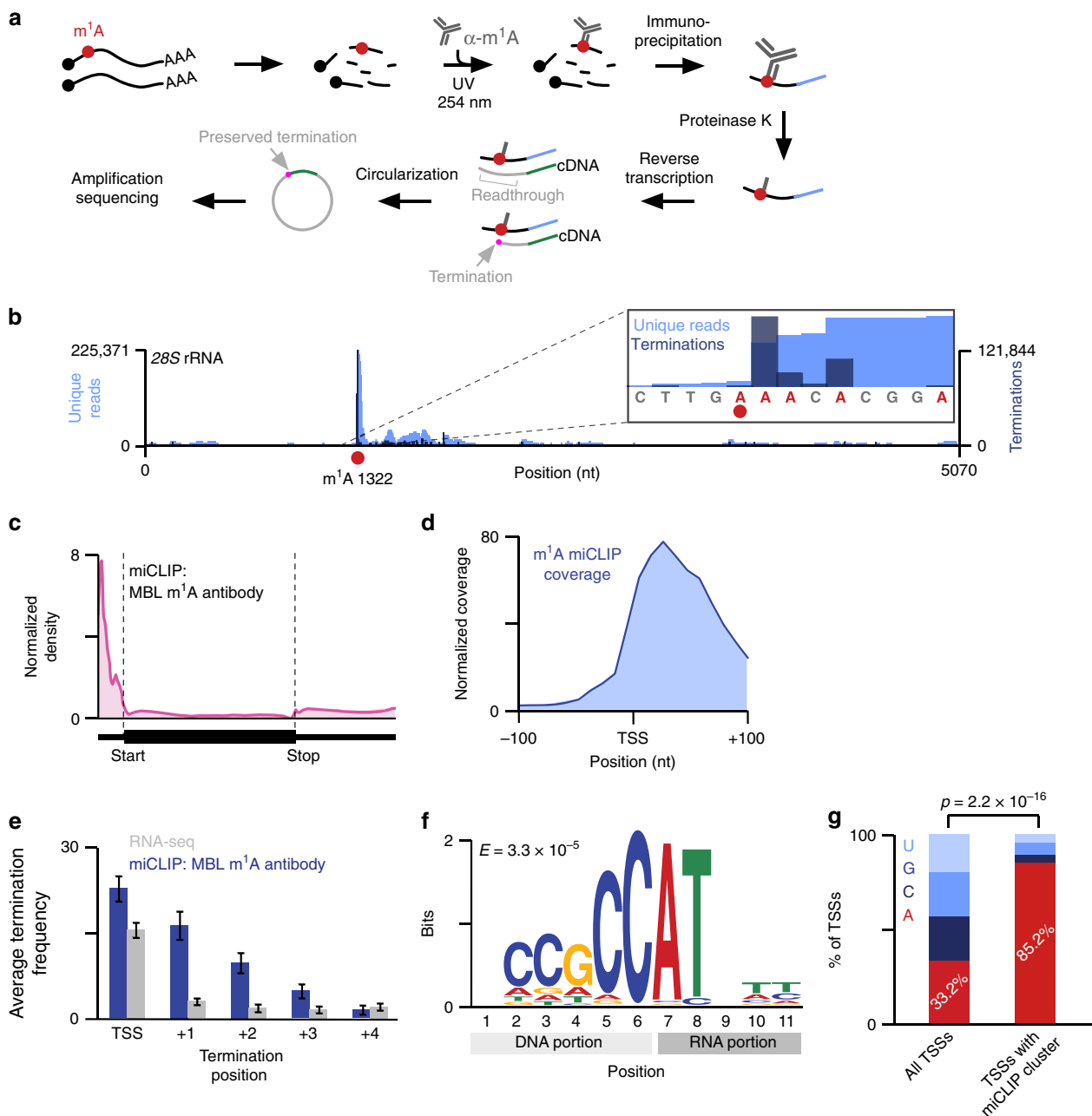


Fig. 2 Mapping m¹A in the transcriptome using m¹A-miCLIP. **a** m¹A-miCLIP workflow. Poly(A) RNA (black lines; black circles, RNA caps) was fragmented, incubated with m¹A antibody, and crosslinked to the antibody with UV. RNA-antibody complexes were immunoprecipitated, ligated to a 3' linker (blue line), and purified. RNA was released from crosslinked antibody using Proteinase K, generating RNA fragments with peptide adducts. After reverse transcription, first-strand cDNA (gray lines) was circularized to preserve m¹A-induced read terminations (magenta dot) at the cDNA 3' end. **b** m¹A-miCLIP detects the known m¹A site in the 28S rRNA. m¹A-miCLIP unique reads (light blue) were aligned to rRNA showing high enrichment at the m¹A site (red circle) at position 1322 of the 28S rRNA. The reads terminated predominantly at the +1 position relative to m¹A (inset, dark blue). **c** Relative distribution of m¹A-miCLIP clusters. The metagenome distribution of m¹A-miCLIP clusters that contained 20 or more stacked reads ($n = 3877$) was normalized to RNA-seq coverage and plotted. Clusters were predominantly enriched in the 5'UTR and next to the transcription-start site (start, start codon; stop, stop codon). The annotation of transcription-start sites is not exact, which results in slight coverage upstream of the transcription-start site. **d** m¹A-miCLIP coverage is enriched at and near transcription-start sites. m¹A-miCLIP unique reads were mapped relative to annotated transcription-start sites using DeepTools (see Methods). **e** While RNA-seq reads terminate predominantly at position 0 (i.e. due to the end of the transcript), m¹A-miCLIP reads terminate predominantly at positions 0 and +1, and to a lesser degree, at downstream nucleotides, suggesting that the m¹A antibody binds at or near the transcription-start site and blocks reverse transcription. Error bars represent SEM. Source data are provided as a Source Data file. **f** m¹A-miCLIP clusters found in 5'UTRs showed significant enrichment of the pyrimidine-rich Initiator sequence motif that facilitates transcription initiation of mRNAs at adenosine ($E = 3.3 \times 10^{-5}$). **g** m¹A-miCLIP clusters were overlapped with a collective set of transcription-start sites (TSS) that included RefSeq annotations and m⁶A_m mRNA extended caps⁴⁴. While 33.2% of these collective transcription-start sites were adenosines, the frequency of adenosine transcription-start sites overlapping m¹A-miCLIP clusters was significantly higher (85.2%, $n = 1,026$ TSSs, $p < 2.2 \times 10^{-16}$, Fisher's exact test)

transcription-start site would produce a peak that is displaced in the 3' direction (see Supplementary Fig. 1a). This can make a m¹A antibody-binding site at the transcription-start site appear to be located at the start codon. To avoid this problem, we chose to generate our libraries in a way that preserves the cDNA ends. Therefore, m¹A-miCLIP reveals exact m¹A antibody-binding sites within transcripts.

We performed m¹A-miCLIP with the AMA-2 antibody using poly(A) RNA from HEK293T cells and examined termination signatures at known m¹A sites in rRNA and tRNA, which typically co-purify to some extent with poly(A) RNA^{21–23}. The majority of reads truncated at the +1 position relative to the m¹A at position 1322 of the 28S rRNA (Fig. 2b, Supplementary Fig. 4b) as well as known m¹A sites in tRNA (Supplementary Fig. 4c). As expected, some read-through was also observed, reflecting the low read-through rate of SuperScript III when it encounters m¹A. These data demonstrate specific detection of m¹A by m¹A-miCLIP using the AMA-2 m¹A antibody. Notably, m¹A-miCLIP showed markedly improved peak resolution compared to the initial peak-based m¹A mapping studies^{3,4} (Supplementary Fig. 4b).

We next asked if m¹A detection in m¹A-miCLIP is adversely affected by m¹A conversion to m⁶A via the Dimroth reaction. m¹A-miCLIP does not use high temperatures and basic pH¹³ which are needed for this conversion (see Supplementary Fig. 2d). No m⁶A-miCLIP reads²⁴ could be detected at the 28S rRNA m¹A site (Supplementary Fig. 4d), demonstrating that m¹A does not appreciably convert to m⁶A during the m¹A-miCLIP protocol. Taken together, these data demonstrate that m¹A-miCLIP maps m¹A with high specificity and resolution.

The AMA-2 m¹A antibody binds near the first mRNA nucleotide. We next used m¹A-miCLIP to map m¹A in mRNA (Supplementary Fig. 5a). To do this, we aligned m¹A-miCLIP unique reads to the genome, generated m¹A-miCLIP clusters (see Methods, Supplementary Data 4), and analyzed their distribution. A metagene analysis of m¹A-miCLIP clusters obtained from HEK293T cells showed a marked enrichment in the 5'UTR (Fig. 2c). More precisely, the clusters were located at mRNA transcription-start sites (Fig. 2d). A similar enrichment was seen using mouse mRNA (Supplementary Data 5, Supplementary Fig. 5b).

Crosslinking of the m¹A antibody is expected to cause reverse transcription terminations within several nucleotides of the site of the antibody-RNA adduct²⁰. As expected, we found that in miCLIP, terminations were enriched not only at the transcription-start site, but also prominently at the +1 position relative to the transcription-start site (Fig. 2e), with additional terminations sometimes seen between position +2 and +3 (Fig. 2e). Thus, the AMA-2 m¹A antibody binds at or near mRNA transcription-start nucleotides.

We considered the possibility that terminations near the transcription-start nucleotide could simply reflect general behavior of the reverse transcriptase as it approaches the mRNA 5' end. To test this, we examined the input RNA fragments in the RNA-seq dataset prepared using the same library cloning strategy as m¹A-miCLIP. In general, RNA-seq reads terminated almost exclusively at the transcription-start nucleotide (Fig. 2e, Supplementary Fig. 5c). Therefore, read terminations seen in m¹A-miCLIP near the transcription-start nucleotide are likely induced by selective binding and crosslinking of the AMA-2 m¹A antibody rather than an artifact of reverse transcription near mRNA 5' ends.

m¹A-miCLIP motif analysis revealed a consensus sequence in the upstream genomic region that was pyrimidine-rich (Fig. 2f)

and highly similar to Initiator, a transcription-initiating sequence which produces transcripts that initiate with adenosine^{25,26}. Indeed, 85% of transcripts containing an m¹A-miCLIP cluster at their transcription-start nucleotide initiated with adenosine (Fig. 2g, Supplementary Data 6). Thus, we reasoned that adenosine at the transcription-start nucleotide was important for binding of the AMA-2 antibody at or near the transcription-start site.

m¹A and m¹A_m are not detected in extended cap structures.

Because the AMA-2 m¹A antibody binds at transcription-start sites, it has been proposed that mRNAs contain a novel mRNA cap structure comprising m⁷G followed by N¹-methylated adenine at the transcription-start nucleotide⁸. The methylated nucleotide would be m¹A or N¹,2'-O-dimethyladenosine (m¹A_m) since the first encoded nucleotide of mRNAs is typically subjected to 2'-O-methylation^{27,28}. To biochemically validate this, we used mass spectrometry to detect m¹A or m¹A_m in “cap dinucleotides,” i.e., m⁷G-ppp-m¹A_m. We treated cellular RNA with P1 nuclease, which digests internal nucleotides to mononucleotides, but leaves the cap dinucleotide intact (see Methods).

We readily detected diverse cap dinucleotides using high-resolution liquid chromatography and mass spectrometry using positive ion mode detection. We developed a multiple reaction monitoring protocol based on the fragment ion transitions from distinct dinucleotide precursor species (see Methods). To confirm that the N¹-methylated adenosine in a cap dinucleotide can be detected, we used synthetic RNA standards (see Methods). Using these standards, we readily detected m⁷G-ppp-m¹A as well as other cap dinucleotides, such as m⁷G-ppp-A_m, and m⁷G-ppp-m⁶A_m (Fig. 3a).

We next examined endogenous cap dinucleotides prepared by digesting HEK293T poly(A) RNA. We readily detected m⁷G-ppp-m⁶A_m ($m/z = 815.1$) and also m⁷G-ppp-A_m ($m/z = 801.1$), though to a lower degree. m⁷G-ppp-C_m ($m/z = 777.1$), m⁷G-ppp-G_m ($m/z = 817.1$), and m⁷G-ppp-U_m ($m/z = 778.1$) were also detected (Fig. 3b). The identity of each species was confirmed by detection of fragment masses corresponding to 7-methylguanine ($m/z = 166.1$) and the base comprising the first nucleotide ($m/z = m^6A_m 150.1$, A_m 136.1, C_m 112.1, G_m 152.1, U_m 112.1) within the extended cap.

Next, we asked whether either m⁷G-ppp-m¹A or m⁷G-ppp-m¹A_m is present in mRNA. Their masses ($m/z = 801.1$ and 815.1, respectively) are identical to cap dinucleotides containing A_m or m⁶A_m. Moreover, the mass of the fragment produced by the N¹-methylated adenine base ($m/z = 150.1$) would be identical to that produced by the m⁶A_m cap. However, m⁷G-ppp-m¹A and m⁷G-ppp-m¹A_m exhibit very distinct retention times from the A_m and m⁶A_m cap dinucleotides based on our synthetic RNA standards (see Fig. 3a). This allows us to differentiate N¹-methyl- and N⁶-methyl-containing adenine. Nevertheless, no N¹-methylated adenine-containing cap dinucleotide was detected in mRNA (Fig. 3b). Thus, m⁷G-ppp-A_m or m⁷G-ppp-m⁶A_m cap structures were readily detected, while m⁷G-ppp-m¹A_(m) was undetectable.

Since mass spectrometry analysis could not validate m¹A at the transcription-start nucleotide, we wanted to directly and sensitively determine the transcription-start nucleotide that is enriched by the m¹A antibody. We used two-dimensional thin layer chromatography (2D-TLC), which can identify and quantify the first encoded nucleotide²⁹. In this approach, mRNA is decapped, and the exposed 5' RNA end is radiolabeled, permitting sensitive detection of the first transcribed nucleotide. The radiolabeled nucleotide species are then resolved using 2D-TLC²⁹. Based in the mobility of each species, the transcription-start nucleotide can be determined.

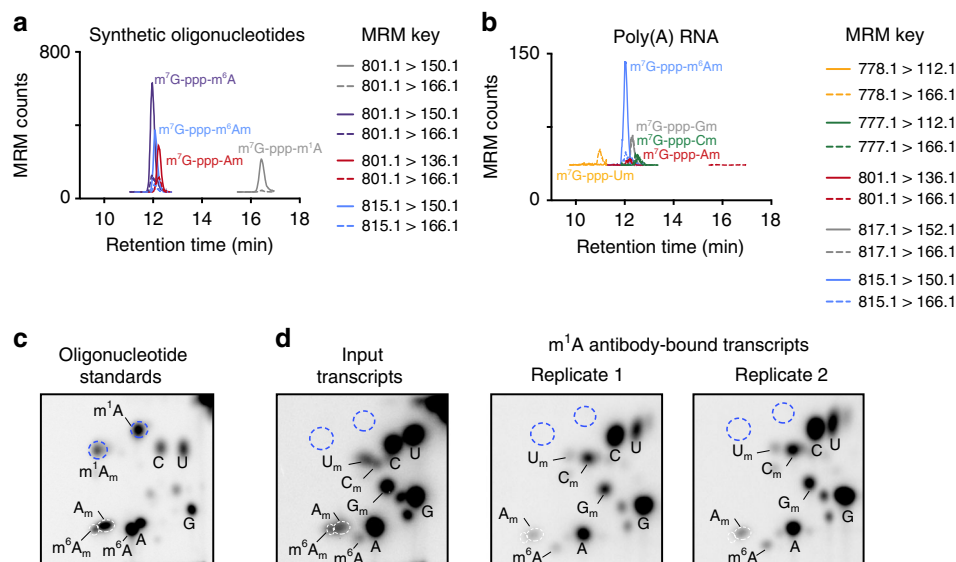


Fig. 3 The AMA-2 m^1A antibody does not enrich for a novel m^1A -containing cap structure. **a** Detection of synthetic 5' cap dinucleotides using mass spectrometry. Dynamic multiple reaction monitoring (dMRM) chromatograms were recorded in positive ion monitoring. Injected was a mixture of oligonucleotides containing m^7G -ppp- m^6A_m , m^7G -ppp- m^1A , m^7G -ppp- m^6A , or m^7G -ppp- A_m cap structures. Two MRM transitions were recorded for each standard (indicated in key in figure). Note the permanent positive charge on the adenine base of m^7G -ppp- m^1A resulted in an increased retention time on the ANP resin as expected. Thus, the m^7G -ppp- m^1A_m dinucleotide would be expected to elute at a retention time later than m^7G -ppp- m^6A due to the permanently positive charged adenine, and earlier than m^7G -ppp- m^1A due to the additional methyl group on the ribose. **b** Detection of 5' cap dinucleotides from cellular mRNA using mass spectrometry. Dynamic multiple reaction monitoring (dMRM) chromatograms were recorded in positive ion monitoring. Injected were mRNA samples from HEK293T cells. Two MRM transitions were recorded for each cap structure of interest (indicated in key in figure). We searched for peaks containing MRM transitions and retention times reflecting m^7G -ppp- m^1A or m^7G -ppp- m^1A_m (see note in **a**), but no such signal was detected. **c** 2D-TLC analysis of capped oligonucleotide standards containing various modified adenines as initiating nucleotides. The migration properties of these various initiating nucleotides are indicated (here, the initiating nucleotides include m^1A , m^1A_m , A_m , m^6A , and m^6A_m). Spots reflecting A, C, G, and U are also marked. Unmarked spots most likely reflect RNA dinucleotides that are not completely digested during RNA preparation⁴⁵. **d** 2D-TLC analysis of mRNA extended caps in antibody-unbound (left panel) or m^1A antibody-bound poly(A) RNA. Known identities of detected spots are marked. Expected positions of potential m^1A or m^1A_m -initiating nucleotides are marked (blue dashed circles) based on standard migrations in **a**

We analyzed the transcription-start nucleotide of poly(A) RNA and poly(A) RNA enriched with the m^1A antibody by 2D-TLC. We also used synthetic RNA containing m^7G -ppp- m^1A and m^7G -ppp- m^1A_m extended cap structures as standards (see Methods). We first optimized the solvent conditions so that m^1A and m^1A_m migrated to distinct positions by 2D-TLC (Fig. 3c). In poly(A) RNA, no m^1A or m^1A_m was detected at transcription-start nucleotides (Fig. 3d). Since these might be rare nucleotides at transcription-start nucleotides, we enriched for transcription start m^1A or m^1A_m -containing mRNAs by immunoprecipitating poly(A) mRNA with the AMA-2 m^1A antibody before 2D-TLC. Here, we again did not see m^1A or m^1A_m as the transcription-start nucleotide (Fig. 3d).

Taken together, the mass spectrometry and the TLC data suggest that m^1A and m^1A_m are not readily detectable at the transcription-start nucleotide, and m^7G -ppp- m^1A_m does not constitute a novel and prevalent mRNA cap structure as proposed⁸.

The AMA-2 m^1A antibody recognizes m^7G -ppp-A cap structures. At this juncture, we had contradictory results: m^1A -miCLIP suggested that m^1A is at the transcription-start nucleotide, but we did not observe m^1A or m^1A_m at this site by either mass spectrometry or TLC. Therefore, we wondered if the AMA-2 antibody binds the transcription-start region in an m^1A -independent manner. When we originally characterized the specificity of the antibody, we performed classic competition studies using nucleosides or nucleotides. However, based on the binding properties of the antibody revealed by mapping studies, we

considered the possibility that the AMA-2 m^1A antibody could recognize an epitope comprising the mRNA extended cap.

To test this, we used a dot blot assay to measure binding of the AMA-2 m^1A antibody to an m^1A -containing oligonucleotide in the presence of various competitors. We considered performing the dot blot assay using m^7G -ppp-A immobilized on the membrane. However, this approach could be misleading since we do not know if m^7G -ppp-A interacts with the membrane in a way that would prevent antibody binding. We therefore used the classic competition approach. In this approach, a m^1A -containing RNA is immobilized on the membrane and different competitors are added in solution. Competitors that bind the m^1A antibody will prevent the antibody from binding to m^1A on the membrane.

As expected, competition with m^1A inhibited antibody binding, while related nucleotides, including adenosine, m^6A , ethenoadenosine, and N^1 -substituted nucleotides, like $N^1,6$ -dimethyladenosine ($m^1,6A$), did not compete with binding (Fig. 4a). Surprisingly, a commercially available cap analog, m^7G -ppp-A, was a relatively effective competitor, with an IC_{50} of 480 nM compared to 100 nM for m^1A (Fig. 4b). m^7G -ppp-G showed weaker inhibition ($IC_{50} \sim 4 \mu M$) (Fig. 4c). The higher binding to the m^7G -ppp-A cap analog compared to the m^7G -ppp-G cap analog may explain the preferential binding of the AMA-2 m^1A antibody to mRNAs that initiate with adenosine.

Notably, the antibody also showed binding to m^7G -ppp, but not m^7G or ATP (Fig. 4c, Supplementary Fig. 4a), suggesting that the antibody's binding specificity includes recognition of features all along the entire m^7G -ppp-A extended cap structure.

Since the AMA-2 m^1A antibody binds the cap structure in an m^1A -independent manner, the m^1A peaks seen in the 5'UTR

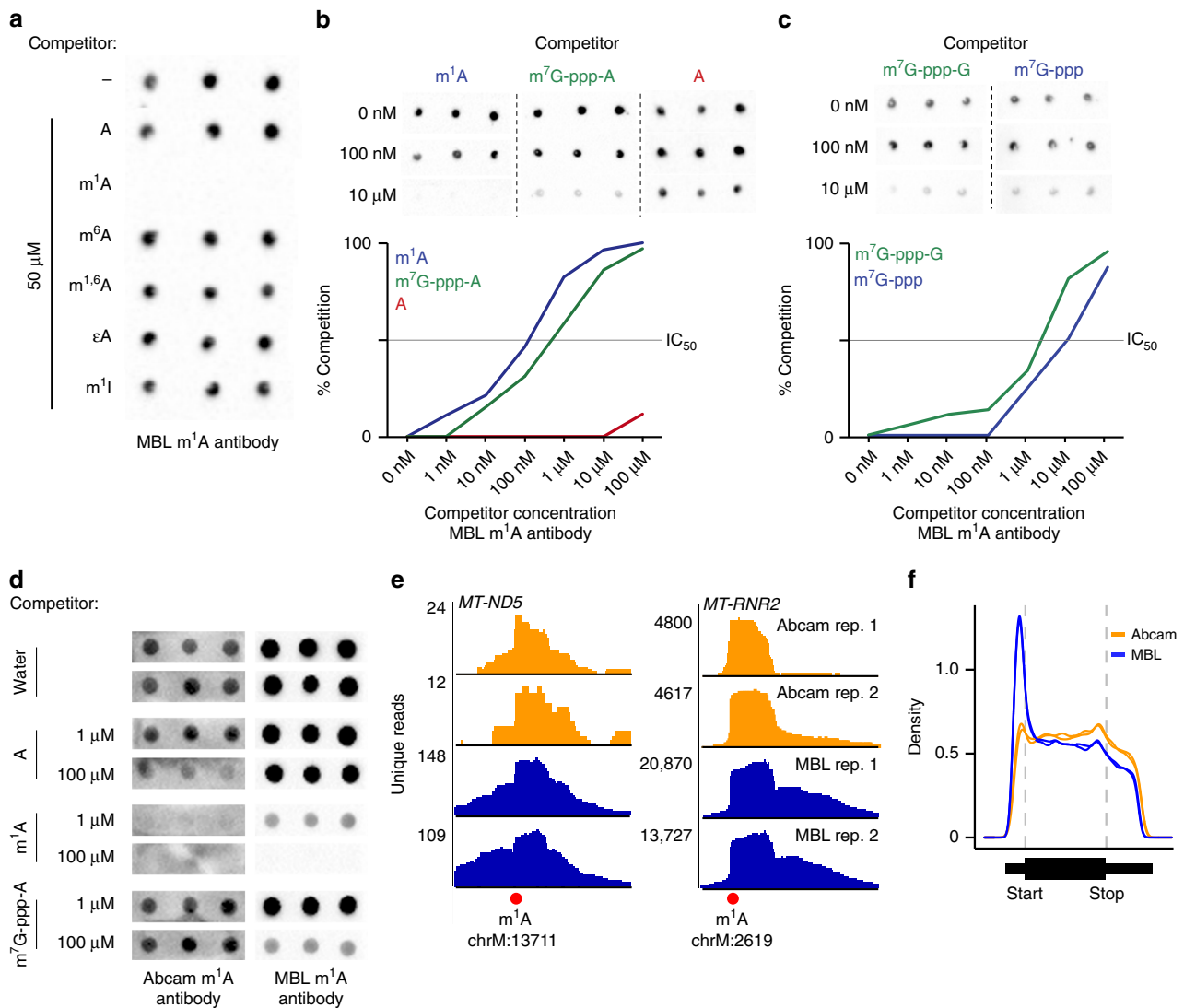


Fig. 4 The AMA-2 m¹A antibody binds to mRNA caps resulting in 5'UTR read enrichment. **a** Shown is a dot blot analysis of the AMA-2 m¹A antibody's specificity. An m¹A-containing oligonucleotide was spotted on a nylon membrane in triplicate and the antibody was incubated with the membrane along with m¹A-like nucleosides. Only m¹A prevents antibody binding to the spotted oligonucleotide. m^{1,6}A N1,6-methyladenosine, εA ethenoadenosine, m¹I 1-methylinosine. **b** Top, images of dot blot competition assays¹ where binding of the AMA-2 m¹A antibody to an m¹A-containing oligonucleotide was competed with m¹A (m¹ATP), m⁷G-ppp-A, or A (ATP) in triplicate. Bottom, quantification of dot blot competition experiments. m⁷G-ppp-A binds (IC₅₀ < 1 μM) to the antibody, although the binding is ~10-fold weaker than m¹A (IC₅₀ = ~100 nM). Source data are provided as a Source Data file. **c** The AMA-2 m¹A antibody recognizes the m⁷G-ppp portion of the cap structure. The membrane was probed with the antibody along with m⁷G-ppp or m⁷G-ppp-G. Both m⁷G-ppp and m⁷G-ppp-G competed with the m¹A oligonucleotide for binding to the antibody (IC₅₀ = ~4 μM for m⁷G-ppp-G, 10 μM for m⁷G-ppp). Source data are provided as a Source Data file. **d** The Abcam m¹A antibody (ab208196) does not bind the mRNA cap. Dot blots were performed with either of the two m¹A antibodies (AMA-2 or Ab208196) in the presence of A, m¹A, or m⁷G-ppp-A (A and m¹A were either tri- or mono-phosphorylated). While both antibodies bind m¹A, only the AMA-2 antibody binds m⁷G-ppp-A. **e** The m¹A antibodies from Abcam (ab208196) and MBL (AMA-2) both detect established m¹A sites. m¹A-miCLIP unique reads from two independent antibodies (in replicate) were aligned to the genome, and peaks at known m¹A sites (red circle) in *MT-ND5* and the mitochondrial 16S rRNA (*MT-RNR2*) exhibited a characteristic truncation at the m¹A site. **f** The m¹A antibody Ab208196 lacks the prominent 5'UTR metagenome peak seen with the AMA-2 antibody. As in **e**, unique m¹A-miCLIP reads were aligned to the genome and the relative location of the reads in each transcript is shown with a metagenome. Each trace is an independent biological replicate. The 5'UTR peak is essentially absent when the m¹A-specific Ab208196 antibody was used compared to the cap-cross-reacting AMA-2 antibody

likely reflect binding to the mRNA cap. This would explain why our m¹A-miCLIP shows read enrichment at the transcription-start nucleotide, as has also been seen using other m¹A mapping approaches⁸.

To test this hypothesis further, we used a second commercially available m¹A antibody from Abcam (catalog number ab208196). Unlike the AMA-2 m¹A monoclonal antibody available from MBL, the Abcam antibody did not bind the m⁷G-ppp-A cap

analog (Fig. 4d), indicating that it does not exhibit cross-reactivity with the mRNA cap.

We first confirmed by m¹A-miCLIP using HEK293T poly(A) RNA that both ab208196 and AMA-2 antibodies can detect authentic m¹A sites. In each case, we observed a robust peak at the m¹A sites in *MT-ND5* and *MT-RNR2*, the mitochondrially encoded 16S RNA (Fig. 4e), confirming the ability of both antibody to detect validated m¹A sites in mRNA.

We next asked if m¹A-miCLIP performed using the m¹A-specific Ab208196 would produce the same transcriptome-wide 5'UTR enrichment of m¹A-containing fragments observed with the AMA-2 m¹A antibody^{3,4}. As expected, the metagene of the miCLIP fragments using the AMA-2 antibody showed a prominent 5'UTR enrichment (Fig. 4f). However, a metagene analysis of all the immunoprecipitated reads using ab208196 lacked the 5'UTR enrichment (Fig. 4f). Together, these data demonstrate that only the AMA-2 antibody, which cross-reacts with the mRNA cap, results in a 5'UTR enrichment in read coverage. Overall, these data demonstrate that binding to the 5' UTR regions is not linked to the presence of m¹A at these sites, but rather attributable to cross-reactivity.

Comparison of m¹A-miCLIP with earlier m¹A maps. Although cross-reactivity with mRNA cap structures explains why m¹A was mapped to transcription-start nucleotides using the AMA-2 antibody, it does not explain the localization of m¹A to internal sites within the 5'UTR, such as the start codon-proximal region, which was proposed to mediate a novel form of translation initiation³. We therefore wanted to understand the exact location of these putative 5'UTR- and start codon-associated m¹A sites.

When we compared mRNAs that show AMA-2 m¹A antibody miCLIP coverage with the Dominissini et al.³ m¹A map that localized m¹A to start codons, we found considerable overlap (Supplementary Fig. 6a). However, the location of reads was different (Supplementary Fig. 6b). In particular, the 5' ends of the miCLIP reads approached the transcription-start site, while reads from Dominissini et al.³ were located downstream of the transcription-start site (Supplementary Fig. 6b, insets). This lateral displacement of peaks towards the start codon is consistent with the library cloning method used in this earlier method (see Supplementary Fig. 1a).

In earlier m¹A mapping studies, accumulations of reads, or in some cases, “troughs” of reduced read coverage due to a putative m¹A site, were used to predict m¹A residues to start codons in mRNA³. miCLIP provides more precise positioning by detecting exact sites of antibody-induced crosslinks, rather than using peaks and troughs, which are a common nonspecific feature in RNA-seq data (see Fig. 2b, e and Supplementary Figs. 5c and 6b).

We additionally re-examined the Li et al.⁸ high-resolution m¹A mapping dataset in HEK293T cells. This study identified 474 m¹A sites in nuclear-encoded genes based on m¹A-induced reverse transcriptase misincorporations⁸. However, we eliminated 122 sites for the following reasons: three sites had gene identifiers missing or removed from Refseq, 37 sites did not map to adenosines, and 82 sites were duplicates resulting from mapping to transcript isoforms of the same gene. This left 352 unique putative m¹A sites in nuclear-encoded genes (see Methods and Supplementary Fig. 6c).

The Li et al.⁸ study located m¹A sites mostly within the 5'UTR. Only 19 sites were at annotated transcriptional-start sites. However, mRNAs can have alternative transcription-start sites, which differ from the RefSeq-annotated transcription-start site³⁰. To determine if the Li et al.⁸ m¹A sites mapped to alternative transcription-start sites, we compared the reported m¹A sites to a list of experimentally validated transcription-start sites in HEK293T cells. This list was derived from CAGE-seq and m⁶A_m mapping data^{31,32}. Of the putative 352 m¹A sites, 134 overlapped with CAGE and/or m⁶A_m-inferred transcription-start sites (Supplementary Fig. 6c). Hence, 140 m¹A sites occurred at transcription-start sites. The false-positive rates of m¹A mapping is not known, so it is possible that other 5'UTR m¹A sites are either false positives or map to currently unannotated transcription-start sites. Thus, most, if not all, of the putative

start codon/5'UTR m¹A sites mapped by that study are localized to alternative transcription-start sites, consistent with our mapping results and consistent with the cap-binding properties of the m¹A antibody.

Discussion

Considerable attention has revolved around m¹A based on its description as a high-stoichiometry, translation-promoting modification in thousands of mRNAs located near start codons³. Subsequent studies concluded that m¹A is less prevalent (~700 sites) with a fraction in mitochondrial mRNA⁸, while other studies suggest even fewer sites, with only two mRNAs having an m¹A at a stoichiometry above 5%⁷. To address these discrepancies, we developed two m¹A mapping approaches: (1) misincorporation mapping, a computational approach to discover m¹A-induced misincorporations in ultra-deep RNA-Seq datasets; and (2) m¹A-miCLIP, a high-resolution method for mapping m¹A antibody-binding sites in the transcriptome using two different m¹A-binding antibodies. Misincorporation mapping shows that m¹A is present at detectable stoichiometries only in the *MT-ND5* transcript, with no m¹A in other mitochondrial mRNAs or 5'UTRs of mRNAs as reported previously. Using m¹A-miCLIP, we find that the previously observed binding of the AMA-2 m¹A antibody to transcription-start nucleotides and the vicinity of start codons is due to a previously unrecognized cross-reactivity of the AMA-2 m¹A antibody to cap structures. We confirm this using a separate m¹A antibody that lacks this cap-binding cross-reactivity. We further show that m¹A is not detectable at transcription-start nucleotides, as previously proposed⁸, based on mass spectrometry and TLC. Overall, these data show that the divergent m¹A mapping data and large number of 5'UTR-mapped m¹A sites largely reflect cross-reactivity of the m¹A antibody with mRNA caps.

Both m¹A and m⁷G are positively charged purines. This common structural feature may be recognized by the AMA-2 antibody. Since the Ab208196 binds m¹A but not cap structures, this antibody does not generate 5'UTR false-positive signals. m¹A-miCLIP carried out using the Abcam antibody shows an m¹A signature in *MT-ND5* but no transcriptome-wide enrichment in 5'UTRs. These data, along with mass spectrometry, TLC, and misincorporation mapping data, support the idea that the m¹A localization to 5'UTR sites is specific to the AMA-2 antibody rather than a reflection of bona fide m¹A nucleotides at these sites.

Our results thus support the idea that m¹A is a rare and low stoichiometry modification except in the case of *MT-ND5* mRNA, as seen in another study⁷. We were not able to detect m¹A in any other cytosolic mRNAs or the mitochondrial mRNAs that were reported to contain m¹A⁸. This could reflect the inability of our method to detect very low stoichiometry m¹A modifications.

The initial m¹A mapping studies found that m¹A is highly prevalent based on mass spectrometry analysis of mRNA^{3,4}. However, more recent experiments showed that poly(A) preparations used for mass spectrometry are usually contaminated with tRNA and rRNA, and when these contaminants are meticulously removed, the m¹A signal in poly(A) preparations is absent³³. Thus, the newer mass spectrometry data also support the idea that m¹A is rare.

A pressing question is whether there are more yet-to-be-discovered modified nucleotides in mRNA. Misincorporation mapping suggests that this is not the case, at least for hard-stop nucleotides. The relative paucity of hard-stop nucleotides in mRNA may reflect the incompatibility of these nucleotides with translation. The ribosome mRNA surveillance pathway induces degradation of mRNAs when tRNAs cannot basepair with

codons³⁴. It is therefore notable that the original m¹A mapping studies localized many m¹A sites to coding sequences, which contributed to skepticism about these maps. If hard-stop nucleotides occur in mRNA, they would likely be transient and act to induce mRNA degradation through this surveillance pathway.

Methods

Cell lines and animals. For misincorporation mapping, an ultra-deep RNA-seq dataset that profiled RNA expression in blood mononucleocytes was used¹². For m¹A-miCLIP, HEK293T cells (passage 5–10, ATCC CRL-3216) or whole mouse brain (16 week age, pooled male and female brain, C57BL/6) was used. HEK293T cells were purchased directly from ATCC but not further validated for identity or tested for mycoplasma contamination. Experiments involving the use of animals were approved by the Institutional Animal Care and Use Committee at Weill Cornell Medicine.

Antibodies. The AMA-2 m¹A antibody is a mouse monoclonal antibody (MBL catalog number D345-3). This antibody is documented to react with both m¹A within RNA and the N¹-methylated adenine base, as documented in MBL's product specifications for AMA-2 (catalog no. D345-3) and prior studies^{3,4}. The specificity was validated previously^{3,4} and in the current study. The Abcam m¹A antibody (ab208196) is a rabbit monoclonal generated against m¹A. Its specificity for m¹A was validated in this study.

Alignment of reads for misincorporation mapping. Raw reads from the ultra-deep RNA-seq dataset used for this study¹² were downloaded from GEO (accession code: GSE33029). This RNA-seq dataset was prepared using standard reverse transcription with SuperScript III, an enzyme expected to produce misincorporations at m¹A positions¹¹. Variants identified in the genomic DNA corresponding to this dataset were acquired from <http://snyderome.stanford.edu>. Coordinates of other SNPs that may be present in the DNA sequence were downloaded from the SNP database dbSNP (February 2017 build; <https://www.ncbi.nlm.nih.gov/projects/SNP/>). Read alignment of forward and reverse read mates was performed using STAR (version 2.5.3a) and the hg19 genome build. Alignment incorporated removal of PCR duplicates, and clipping of 10 bases on either end of each read, since the ends of Illumina reads are prone to sequencing error³⁵. Only reads that mapped to a single location in the genome were used for downstream analysis. A maximum of one mismatch per read was permitted for alignment.

Misincorporation mapping. To identify misincorporations, aligned reads were analyzed using Rsamtools Pileup (version 1.27.16). This program was used to determine the frequency of each of the four nucleotides present in mapped reads at every genomic position with read coverage. We limited our analysis to nucleotide positions with a minimum combined read depth of 500 unique reads across the 20 biological replicates to maximize sensitivity of detecting modified nucleotides. To prevent calling genomic variants and SNPs as modification-induced misincorporations, we did not analyze nucleotide positions containing variants discovered in the genomic DNA corresponding to the RNA-seq dataset, or SNPs annotated in dbSNP. Importantly, our analysis could only be performed on transcripts longer than the library insert size of ~250 bases¹². For this reason, analysis of cytosolic tRNAs, which are ~75 nt-long RNAs that contain known conserved m¹A residues, could not be performed. However, short RNAs generated from polycistronic transcripts, like mitochondrial tRNAs³⁶, were represented in the analyzed library. To identify sites of modification throughout the transcriptome, we initially filtered for all nucleotide positions that were covered by at least 500 mapped reads and contained a 1% misincorporation rate, and that were present in at least half of biological replicates (Supplementary Data 1). To further obtain a high-confidence list of modification positions, we required that within the misincorporation profile at each initially identified position, a minimum of 5% of misincorporations were heterogeneous (i.e. transitions of the reference nucleotide to all three possible alternative nucleotides) in order to minimize detection of adenosine-to-inosine editing, and heterozygous alleles not reported as variants or SNPs. We chose this filter because hard-stop nucleotides have been shown to cause heterogeneous misincorporations, even when one type of misincorporation is predominant¹¹. This resulted in a high-confidence list of sites that were detected at known and novel modification positions (Supplementary Data 2).

m¹A-miCLIP. m¹A-miCLIP was performed as previously described^{20,37}, briefly described below along with any modifications: C57BL/6 mice (8 weeks) were sacrificed by CO₂ inhalation and cervical dislocation as approved by the Weill Cornell Medicine Institutional Animal Care and Use Committee (IACUC). Total RNA from HEK293T cells ($n = 2$ biological replicates) or whole mouse brain ($n = 6$ biological replicates) was extracted using TRIzol (ThermoFisher) and treated with RNase-free DNase I (Promega). Poly(A) RNA was isolated using one round of selection with oligo(dT)₂₅ magnetic beads (New England Biolabs). This resulted in approximately 10 µg of poly(A) RNA for each replicate used in this study. Poly(A)

RNA was subjected to fragmentation using RNA Fragmentation Reagents (ThermoFisher) for exactly 12 min at 75 °C. This fragmentation protocol is identical to the one in m¹A-seq, and has been reported not to facilitate substantial m¹A to m⁶A rearrangement³. Fragmented RNA was then incubated with 10–15 µg of m¹A antibody per replicate and the antibody-RNA complexes were processed for crosslinking, immunoprecipitation, RNA 3' linker ligation, purification, and reverse transcription^{20,37}. Following reverse transcription of purified peptide-RNA complexes, first-strand cDNA was circularized using CircLigase II ssDNA Ligase (EpiBio) to preserve the 3' end of the cDNA, and thus, sites of m¹A-induced terminations of reverse transcription. To generate priming sites for library amplification, the cDNA was cut in the middle of the cDNA primer sequence using a single-stranded DNA oligo complementary to this sequence and FastDigest BamHI (ThermoFisher)^{20,37}. This generated priming sites for the Illumina P5 and P3 primers on either side of the first-strand cDNA, eliminating the need for a second-strand synthesis step. For library amplification, Accuprime Supermix I (ThermoFisher) and Illumina P5 and P3 primers were used (see Supplementary Data 7). Amplified libraries were purified using AMPure XP magnetic beads (Beckman Coulter). Libraries were subjected to next-generation sequencing at the Epigenomics Core of Weill Cornell Medicine. Libraries were sequenced on an Illumina HiSeq 2500 and MiSeq instrument in single-end mode to generate 50-base reads.

RNA-seq. HEK293T cell total RNA was extracted with TRIzol (ThermoFisher), treated with RNase-free DNase I (Promega), and poly(A) RNA was isolated using one round of selection with oligo(dT)₂₅ magnetic beads (New England Biolabs). RNA was then subjected to fragmentation using RNA Fragmentation Reagents (ThermoFisher) for exactly 12 min at 75 °C. Fragmented RNA was then subjected to RNA 3' linker ligation using T4 RNA Ligase I (New England Biolabs) and reverse transcription using a primer complementary to the linker sequence and SuperScript III (ThermoFisher) (see Supplementary Data 7). First-strand cDNA was gel-purified using denaturing PAGE, and then circularized using CircLigase II ssDNA Ligase (EpiBio). Circularized cDNA was then cut and amplified exactly as described above for m¹A-miCLIP. Resulting RNA-seq libraries were subjected to next-generation sequencing at the Epigenomics Core of Weill Cornell Medicine. Libraries were sequenced on an Illumina HiSeq 2500 instrument in single-end mode to generate 50-base reads.

Read processing and alignment. After sequencing, reads from m¹A-miCLIP or RNA-seq libraries were trimmed of the 3' linker sequence and barcoded reverse transcription primer sequences using Flexbar (version 2.5) (see Supplementary Data 7). To demultiplex reads belonging to individual biological replicates, the pyBarcodeFilter.py script of the pyCRAC suite (version 1.2.2) was used. The random portion of the reverse transcription barcode was then moved into the sequence header using a custom awk script (available upon request). PCR duplicates were collapsed using pyFastqDuplicateRemover.py of the pyCRAC suite. Finally, reads were aligned to hg19 for HEK293T cells or mm10 for mouse brain using Bowtie (version 1.1.2).

Generation of m¹A-miCLIP clusters. m¹A-miCLIP clusters of unique reads were generated using the CIMS software package for analysis of HITS-CLIP data^{38,39}. To generate clusters and determine the cluster score (maximum of stacked reads), the tag2profile.pl, tag2cluster.pl and extractPeak.pl scripts of the CIMS software package were used. A custom awk script was then used to filter for clusters of a minimum score (at least 20 stacked reads; script is available upon request).

Motif analyses. To search for a possible common sequence motif present in our HEK293T cell m¹A-miCLIP dataset, we focused on potential motifs present in m¹A-miCLIP clusters in the 5'UTR, the region of predominant m¹A-miCLIP cluster enrichment. The genomic sequences of these clusters were retrieved using bedtools and subjected to motif discovery using the MEME suite (version 4.11.4).

Metagene distribution analyses. To analyze the metagene distribution of m¹A-miCLIP clusters on mRNAs, MetaPlotR was used⁴⁰, with in-house modifications. The density of m¹A-miCLIP coverage was normalized to that of RNA-seq coverage to reveal any enrichments using a custom R script (available upon request). For the HEK293T cell metagene, the in-house HEK293T cell RNA-seq dataset described above was used for normalization. For the mouse brain metagene, a published whole-brain RNA-seq dataset was used⁴¹ (accession code: GSE52564). To plot the coverage of transcription-start sites by m¹A-miCLIP at higher resolution, the plotProfile tool of the Deeptools suite was used.

Examination of antibody crosslink sites around the transcription-start site.

For analysis of antibody crosslinks at the transcription-start sites of mRNAs, we analyzed terminations of reverse transcription (i.e. 5' ends of reads) around these sites. To do so, the number of terminations was measured around RefSeq-annotated transcription-start sites that had coverage in both m¹A-miCLIP and RNA-seq. Terminations were counted at positions ranging from the transcription-start site to position +4 relative to the transcription-start site. Then, transcription-start sites

were filtered for those that contained a minimum coverage of five unique reads at the transcription-start site position in both m¹A-miCLIP and RNA-seq. This filtered set of transcription-start sites was then used to compare the distributions of read terminations in m¹A-miCLIP and RNA-seq. We focused on terminations rather than misincorporations in m¹A-miCLIP is because the misincorporation profile of m¹A is sequence dependent, with both upstream and downstream nucleotides contributing to misincorporation variability¹¹. Thus, we used the presence of terminations as a signature of antibody crosslinking events in our dataset. Additionally, while rare types of reverse transcriptases that read through m¹A have been described⁴², standard reverse transcriptases, like the SuperScript III used in m¹A-miCLIP, produce frequent terminations at m¹A residues^{11,43}.

Measurement of transcription-start sites enriched by m¹A-miCLIP. To determine the types of transcription-start sites overlapping m¹A-miCLIP clusters, we used a collection of transcription-start sites that included RefSeq transcription-start sites as well as recently-mapped transcription-start site regions containing the m⁶A_m mRNA extended cap²⁴. The frequencies of all transcription-start site types or those overlapping m¹A-miCLIP clusters were thus determined using this collective set.

Synthetic oligonucleotides used in this study. For biochemical analysis of various modifications present within the extended caps of mRNAs, synthetic oligonucleotides were generated as standards for mass spectrometry and/or thin layer chromatography (see below; see Supplementary Data 7). Oligonucleotides containing m⁷G-ppp-A_m, m⁷G-ppp-m⁶A, or m⁷G-ppp-m⁶A_m were synthesized chemically⁴⁴. Oligonucleotides containing m⁷G-ppp-m¹A or m⁷G-ppp-m¹A_m were synthesized enzymatically using an oligonucleotide initiating with ppp-m¹A (Trilink). This oligonucleotide was capped using ScriptCap Cap 1 Capping System (CellScript) to generate the m⁷G cap and, in the case of m⁷G-ppp-m¹A_m, 2'-O-methylation of m¹A.

Liquid chromatography and mass spectrometry (LC-MS). Poly(A) RNA was prepared for mass spectrometry as follows. Total RNA from HEK293T cells was treated with TURBO DNase (ThermoFisher) according to the manufacturer's instructions, followed by two rounds of poly(A) selection using oligo(dT) magnetic beads (NEB). Small RNAs shorter than 200 nt were then removed from the poly(A) RNA using the RNeasy kit (Qiagen). This size selection was performed to prevent detection of extended cap structures that are known to be present in certain small RNAs, like small nuclear RNAs (snRNAs). Approximately 5 µg of DNase-treated, poly(A)-selected, and size-selected RNA was thus generated for each sample for mass spectrometry analysis. To release extended cap structures from the nucleotides comprising the internal portion of the RNA, RNA was digested with 2–4 units of Nuclease P1 (Sigma Aldrich) in a final buffer concentration of 30 mM sodium acetate (pH 5.5) for 3 h at 37 °C. Following digestion, the nuclease was removed from the samples using molecular weight cutoff centrifugal filters (VWR). The digested and purified RNA was finally dried using an Eppendorf Vacufuge and reconstituted with 70% acetonitrile (LC-MS grade; Sigma Aldrich) to a final concentration of 0.5 µg/µl. Two microliters of the resulting solution were subjected to MS analysis.

Samples were injected into an LC-MS/MS system comprising an Agilent 1260 HPLC and an Agilent 6460 triple quadrupole mass spectrometer equipped with a JetStream electrospray ionization source. Positive ion monitoring and multiple reaction monitoring was used for detection of extended caps. The caps were resolved on an aqueous normal phase column (ANP, Cogent Diamond Hydride, 4 µm particle size, 150 mm × 2.1 mm; Microsolv). To achieve chromatographic separation of the cap structures from mononucleotides, the following gradient was used. The aqueous mobile phase (Buffer A) was 50% isopropanol with 0.025% acetic acid, and the organic mobile phase (Buffer B) was 90% acetonitrile containing 5 mM ammonium acetate. EDTA was added to the mobile phase in a final concentration of 6 µM. The final gradient applied was 0–10 min 99% B, 1.0–7.0 min to 80% B, 7.0–18.0 min to 50% B, 18.0–19.0 min to 0% B, and 19.1–29.0 min 99% B. The flow rate was 0.4 mL/min during data acquisition and 0.6 mL/min during column re-equilibration. Data were saved in centroid mode using MassHunter workstation acquisition software (Agilent). Data files were processed with MassHunter Qualitative Analysis Software (Agilent).

Exact operating source parameters for the LC-MS analysis are available upon request.

Biochemical examination of modifications at transcription-start sites. To identify the initiating nucleotide structure in mRNAs bound by the m¹A antibody, we utilized 2D-TLC analysis of mRNA extended caps²⁴. For analysis of cellular RNA, we used HEK293T cell poly(A) RNA or poly(A) RNA enriched using the m¹A antibody^{3,4}. Oligonucleotide standards, input (antibody-unbound) poly(A) RNA, and m¹A antibody-enriched poly(A) RNA were then subjected to 2D-TLC²⁴. To enhance resolution of m¹A and m¹A_m from other nucleotide species, the first dimension of 2D-TLC was resolved using 66% isobutyric acid and 1% NH₄OH in water, and the second dimension was resolved using 60% ammonium sulfate in 100 mM sodium phosphate buffer of pH 6.8 (w/v) with a final concentration of 2% *n*-propanol. Both dimensions were resolved overnight.

Synthesis of N^{1,6}-methyladenosine. To synthesize N^{1,6}-methyladenosine, N⁶-methyladenosine (Selleckchem) was dissolved in dry DMF and followed with addition of iodomethane (Acros Organics; 10:1 molar ratio iodomethane:N⁶-methyladenosine). The mixture was stirred overnight at room temperature. The product was purified by flash chromatography on silica gel (EMD), eluting with methanol and dichloromethane (1:10 to 1:5; ACS or HPLC grade solvents). This resulted in a product yield of 46.3% N^{1,6}-methyladenosine. Product identity was confirmed by nuclear magnetic resonance (NMR) and high-resolution mass spectrometry (HR-MS).

NMR spectra were recorded using a 500-MHz Bruker DMX-500 instrument at room temperature, and chemical shifts were referenced to the residual solvent peak. Shifts were as follows: ¹H NMR (500 MHz, DMSO-*d*₆) δ 8.11 (s, 1H), 8.04 (s, 1H), 5.76 (d, *J* = 5.8 Hz, 1H), 5.15 (s, 1H), 4.43 (t, *J* = 5.4 Hz, 1H), 4.15 (m, 2H), 3.92 (d, *J* = 3.7 Hz, 1H), 3.63 (dd, *J* = 12.0, 3.9 Hz, 1H), 3.56 (m, 2H), 3.50 (s, 3H), 3.45 (d, *J* = 6.5 Hz, 1H), 1.23 (s, 3H).

HR-MS data were recorded with Waters LCT-Premier XE at room temperature. For a predicted mass for N^{1,6}-methyladenosine, or C₁₂H₁₈N₅O₄⁺, of 296.1353, the mass found was 296.1361.

Analysis of Li et al. m¹A sites. The Li et al.⁸ m¹A sites for nuclear-encoded genes was obtained from supplementary Table 2 of the published manuscript. We annotated each of the 474 transcriptomic sites with their corresponding genomic coordinates and nucleotide sequences using an annotation file generated from Refseq with MetaPlotR. With a custom R script, we then filtered likely erroneous sites as specified in the Results section. Briefly, sites corresponding to gene IDs missing in Refseq, or that mapped to non-adenosine nucleotides, or with duplicate genomic coordinates were all removed.

Characterization of antibody affinity for various substrates. The specificity or affinity of the m¹A antibody for various nucleosides, nucleotides, or cap structures was determined as follows. To determine the specificity of the m¹A antibody for various nucleosides, two approaches were performed. For testing the specificity of the antibody for various nucleosides in the context of m¹A-miCLIP, the antibody was crosslinked to total cellular RNA in the presence of various competitor nucleotides. Antibody binding, crosslinking, and detection of crosslinked antibody-RNA complexes was performed exactly as in miCLIP, except with the inclusion of the competitor nucleotide during the antibody-binding reaction.

For testing the specificity of the antibody for modified adenines, especially those resembling N¹-methylated adenine, a dot blot assay was performed wherein the competing molecule is added during antibody binding¹. Competition assays are usually used to measure binding, rather than spotting the nucleotides to the membrane, since the manner of interaction of each nucleotide to the membrane is not known and can affect antibody binding. For measuring the affinity, the IC₅₀ of the various nucleotides and cap dinucleotides was measured. In these experiments, a series of the dot blot assays were performed, where serial dilutions of each competitor molecule (ranging from 10 µM to 1 nM) were used in parallel during antibody binding reactions. The dot blots were performed as follows: 250 ng of m¹A-containing synthetic oligonucleotide (see Supplementary Data 7) were spotted in triplicate on a BrightStar membrane (ThermoFisher), allowed to briefly air-dry, and auto-crosslinked twice in a Stratalinker 2400 (Stratagene). Each membrane was rinsed briefly in PBST, then blocked for 1 h at room temperature in 5% milk in PBST. Each membrane was then placed into a pouch containing a 1:1000 dilution of the m¹A antibody in 0.5% milk in PBST, and an appropriate concentration of competitor molecule. The antibody binding proceeded for 2 h at room temperature. Then, each membrane was washed three times in PBST (5 min per wash), and then incubated in a dilution of 1:2500 of secondary antibody (anti-mouse, GE # NA931; anti-rabbit, GE NA934) in 0.5% milk in PBST for 1 h at room temperature. Finally, the membrane was washed three times in PBST (5 min per wash), and developed using ECL Prime (GE). Membranes corresponding to a dilution series of a specific competitor molecule were imaged together using a ChemiDoc Imager (Bio-Rad).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

A reporting summary for this Article is available as a Supplementary Information file. Sequencing data have been deposited in GEO (GSE97909). The source data for Figs. 2e, 4b and 4c are provided in the Source Data File. All data are available from the corresponding author upon reasonable request.

Code availability

Custom scripts used in this study are available here: https://github.com/olarerin/misincorporation_mapping.

Received: 29 April 2019; Accepted: 10 October 2019;

Published online: 12 November 2019

References

- Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near Stop codons. *Cell* **149**, 1635–1646 (2012).
- Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
- Dominissini, D. et al. The dynamic N¹-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
- Li, X. et al. Transcriptome-wide mapping reveals reversible and dynamic N¹-methyladenosine methylome. *Nat. Chem. Biol.* **12**, 311–316 (2016).
- Itoh, K., Mizugaki, M. & Ishida, N. Preparation of a monoclonal antibody specific for 1-methyladenosine and its application for the detection of elevated levels of 1-methyladenosine in urines from cancer patients. *Jpn J. Cancer Res.* **79**, 1130–1138 (1988).
- Mishima, E. et al. Immuno-Northern blotting: detection of RNA modifications by using antibodies against modified nucleosides. *PLoS One* **10**, e0143756 (2015).
- Safra, M. The m1A landscape on cytosolic and mitochondrial mRNA at single base resolution. *Nature* **551**, 251–255 (2017).
- Li, X. et al. Base-resolution mapping reveals distinct m(1)A ethylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* **68**, 993–1005.e1009 (2017).
- Schwartz, S. m(1)A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map. *RNA* **24**, 1427–1436 (2018).
- Ryvkin, P. et al. HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684–1692 (2013).
- Hauenschild, R. et al. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res.* **43**, 9950–9964 (2015).
- Chen, R., Mias, G. I., Li-Pook-Than, J., Jiang, L. & Lam, H. Y. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307 (2012).
- Macon, J. B. & Wolfenden, R. 1-Methyladenosine. Dimroth rearrangement and reversible reduction. *Biochemistry* **7**, 3453–3458 (1968).
- Yang, J. et al. Mapping of complete set of ribose and base modifications of yeast rRNA by RP-HPLC and mung bean nuclease assay. *PLoS One* **11**, e0168873 (2016).
- Suzuki, T. & Suzuki, T. A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. *Nucleic Acids Res.* **42**, 7346–7357 (2014).
- Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919–932 (2008).
- Oerum, S., Degut, C., Barraud, P. & Tisne, C. m1A post-transcriptional modification in tRNAs. *Biomolecules* **7**, <https://doi.org/10.3390/biom7010020> (2017).
- Xu, L. et al. Three distinct 3-methylcytidine (m(3)C) methyltransferases modify tRNA and mRNA in mice and humans. *J. Biol. Chem.* **292**, 14695–14703 (2017).
- Heyer, E. E., Ozadam, H., Ricci, E. P., Cenik, C. & Moore, M. J. An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Res.* **43**, e2 (2015).
- Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
- Zhao, W. et al. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 419 (2014).
- Sultan, M. et al. Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics* **15**, 675 (2014).
- Cui, P. et al. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**, 259–265 (2010).
- Mauer, J., et al. Reversible methylation of m⁶Am in the 5' cap controls mRNA stability. *Nature* **541**, 371–375 (2016).
- Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell Biol.* **14**, 116–127 (1994).
- Furuichi, Y. et al. Methylated, blocked 5' termini in HeLa cell mRNA. *Proc. Natl. Acad. Sci. USA* **72**, 1904–1908 (1975).
- Perry, R. P., Kelley, D. E., Friderici, K. & Rottman, F. The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. *Cell* **4**, 387–394 (1975).
- Kruse, S. et al. A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Sci. Rep.* **1**, 126 (2011).
- Abugessaisa, I. et al. FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci. Data* **4**, 170107 (2017).
- Lykke-Andersen, S. et al. Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* **28**, 2498–2517 (2014).
- Mauer, J. et al. Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature* **541**, 371–375 (2017).
- Legrand, C. et al. Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* **27**, 1589–1596 (2017).
- Simms, C. L. & Zaher, H. S. Quality control of chemically damaged RNA. *Cell Mol. Life Sci.* **73**, 3639–3653 (2016).
- Schirmer, M. et al. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37 (2015).
- Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470–474 (1981).
- Grozhiik, A. V., Linder, B., Olarerin-George, A. O. & Jaffrey, S. R. Mapping m(6)A at individual-nucleotide resolution using crosslinking and immunoprecipitation (miCLIP). *Methods Mol. Biol.* **1562**, 55–78 (2017).
- Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* **29**, 607–614 (2011).
- Moore, M. J. et al. Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* **9**, 263–293 (2014).
- Olarerin-George, A. O. & Jaffrey, S. R. MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. *Bioinformatics* **33**, <https://doi.org/10.1093/bioinformatics/btx002> (2017).
- Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
- Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
- Motorin, Y., Muller, S., Behm-Ansmant, I. & Branlant, C. Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.* **425**, 21–53 (2007).
- Mauer, J. et al. FTO controls reversible m6Am RNA methylation during snRNA biogenesis. *Nat. Chem. Biol.* **15**, 340–347 (2019).
- Gott, J. M. (eds) *RNA Interference, Editing, and Modification: Methods and Protocols*. (Humana Press, Totowa, NJ, 2004).

Acknowledgements

We thank S. Schwartz for helpful discussions, and members of the Jaffrey laboratory for helpful comments and suggestions. We thank J. Mauer and S. Zaccara for experimental contributions that were not included in this manuscript. This work was supported by NIH grants R01DA037755 (to S.R.J.), T32 HD060600 and ULI TR000457 (to A.V.G.), T32 CA062948, KL2-TR-002385, NRSA 1F32GM120987 (to A.O.O.-G.), U01 HL121828 and P01 HD067244 (to M.S. and S.S.G.), and by a Postdoctoral Enrichment Program Award from the Burroughs Wellcome Fund (to A.O.O.-G.).

Author contributions

S.R.J., A.V.G., and A.O.O.-G. conceived the project and designed and carried out experiments. A.O.O.-G. designed and implemented the misincorporation mapping method. A.V.G., and partly A.O.O.-G. performed m¹A-miCLIP, characterized the specificity of the antibodies, and analyzed the miCLIP data. M.S. and S.S.G. performed mass spectrometry analysis of mRNA extended caps. X.L. performed chemical synthesis. S.R.J., A.V.G., and A.O.O.-G. wrote the manuscript with input from all the co-authors.

Competing interests

S.R.J. declares a competing interest; he is scientific founder, advisor to, and owns equity in Gotham Therapeutics. All other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13146-w>.

Correspondence and requests for materials should be addressed to S.R.J.

Peer review information *Nature Communications* thanks Yuri Motorin, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019