Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome

Bastian Linder[1,4], Anya V Grozhik[1,4], Anthony O Olarerin-George[1], Cem Meydan[2,3], Christopher E Mason[2,3] & Samie R Jaffrey[1]

N^6-methyladenosine (m6A) is the most abundant modified base in eukaryotic mRNA and has been linked to diverse effects on mRNA fate. Current mapping approaches localize m6A residues to transcript regions 100–200 nt long but cannot identify precise m6A positions on a transcriptome-wide level. Here we developed m6A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP) and used it to demonstrate that antibodies to m6A can induce specific mutational signatures at m6A residues after ultraviolet light–induced antibody-RNA cross-linking and reverse transcription. We found that these antibodies similarly induced mutational signatures at N^6,2′-*O*-dimethyladenosine (m6Am), a modification found at the first nucleotide of certain mRNAs. Using these signatures, we mapped m6A and m6Am at single-nucleotide resolution in human and mouse mRNA and identified small nucleolar RNAs (snoRNAs) as a new class of m6A-containing non-coding RNAs (ncRNAs).

© 2015 Nature America, Inc. All rights reserved.

m6A is the most prevalent modified base in mRNA[1–4]. Although m6A was detected in poly(A)$^+$ RNA in the 1970s[3,4], it was not widely accepted as a component of mRNA until recent transcriptome-wide mapping studies showed that m6A is found in several thousand transcripts, typically near the stop codon, but also in the coding sequence, 3′ UTR and 5′ UTR of mRNAs[1,2].

The current m6A mapping approach, methyl-RNA immunoprecipitation and sequencing (MeRIP-Seq, also called m6A-Seq)[1,2], involves the immunoprecipitation of ~100-nt-long RNA fragments with m6A-specific antibodies. The approach generates m6A peaks, but it does not identify specific m6A residues.

Identifying m6A residues is challenging. Adenosine methylation appears to be restricted to adenosines in an R-A*-C context (where R denotes G or A, and A* denotes methylatable A), and recent transcriptome-wide m6A maps[1,2,5] suggest the broader consensus motif DRACH (where D denotes A, G or U, and H denotes A, C or U). Although DRACH motifs are prevalent, only a fraction are methylated *in vivo*[1,2]. One can predict exact m6A positions from MeRIP-Seq peaks by searching for DRACH motifs near the point of highest read coverage[5]. However, m6A often appears in clusters, which can result in large peaks spanning several m6A residues[1]. Additionally, multiple DRACH motifs can be present underneath a peak, making it difficult to predict the specific methylated adenosine(s).

Unlike other base modifications such as 5-methylcytosine, m6A and A have nearly identical chemical properties; this has prevented the development of a chemical method for the selective modification of m6A residues for detection at single-nucleotide resolution[6]. Moreover, m6A does not introduce errors during reverse transcription that would allow for direct mapping of its position[7]. Thus, a major goal is to develop a chemical signature that indicates the precise location of m6A residues in the transcriptome.

We reasoned that antibodies to m6A could be used to create such a highly selective mark: reverse transcription of RNA that was UV–cross-linked to the antibody would result in mutations or truncations in the cDNA, indicating the presence of the bound protein. This strategy resembles RNA cross-linking and immunoprecipitation (CLIP)-based techniques that cross-link proteins to RNA in living cells to enable mapping of RNA-binding sites throughout the transcriptome[8–11].

Here we show that reverse transcription of RNA cross-linked to certain m6A antibodies resulted in a highly specific pattern of mutations or truncations in the cDNA that enabled precise identification of m6A residues. These mutational signatures also enabled identification of the related modified nucleotide m6Am, which is the first nucleotide after the 7-methylguanosine cap of certain mRNAs[12]. Using these signatures, we mapped m6A and m6Am residues throughout the transcriptome at single-nucleotide resolution.

RESULTS

Determining the mutational profile of antibodies to m6A

Because UV-induced RNA-protein cross-links can be highly variable[13], we first established the mutational signatures of different commercially available antibodies to m6A. To do this, we cross-linked an *in vitro*–transcribed RNA containing a single m6A to different m6A antibodies. The RNA was reverse transcribed, and the cDNA was sequenced to identify mutations

[1]Department of Pharmacology, Weill Medical College, Cornell University, New York, New York, USA. [2]Department of Physiology and Biophysics, Weill Medical College, Cornell University, New York, New York, USA. [3]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College, Cornell University, New York, New York, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to S.R.J. (srj2003@med.cornell.edu).

RECEIVED 3 DECEMBER 2014; ACCEPTED 28 MAY 2015; PUBLISHED ONLINE 29 JUNE 2015; DOI:10.1038/NMETH.3453

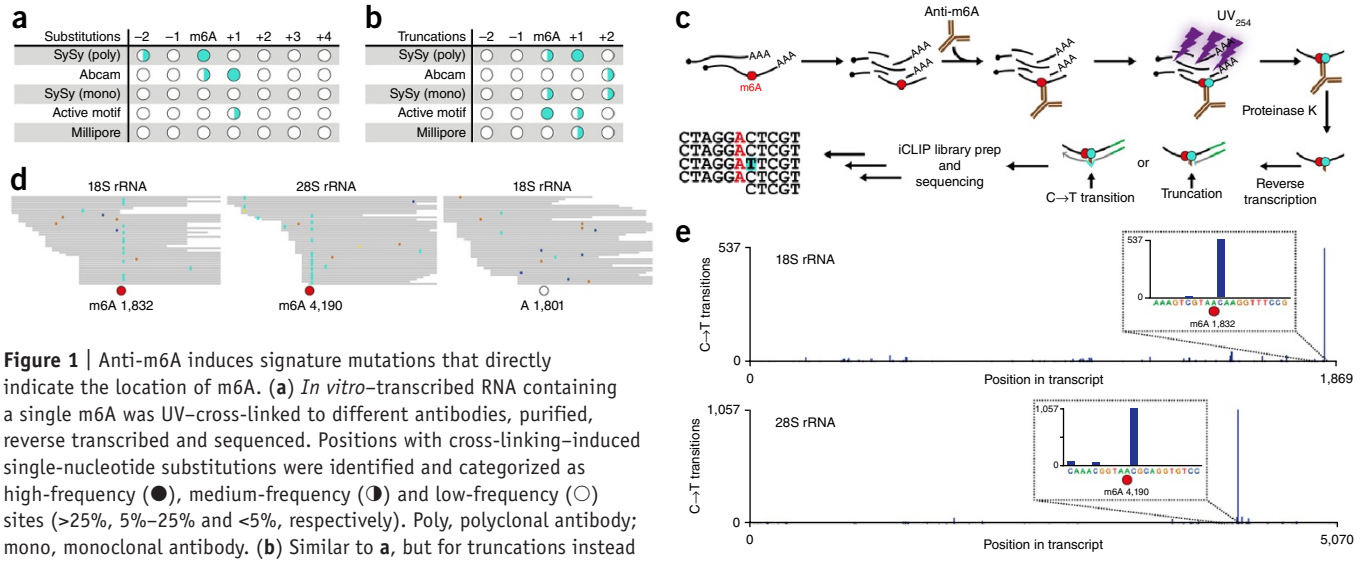

Figure 1 | Anti-m6A induces signature mutations that directly indicate the location of m6A. (**a**) *In vitro*–transcribed RNA containing a single m6A was UV–cross-linked to different antibodies, purified, reverse transcribed and sequenced. Positions with cross-linking–induced single-nucleotide substitutions were identified and categorized as high-frequency (●), medium-frequency (◑) and low-frequency (○) sites (>25%, 5%–25% and <5%, respectively). Poly, polyclonal antibody; mono, monoclonal antibody. (**b**) Similar to **a**, but for truncations instead of substitutions. (**c**) The miCLIP protocol. Purified cellular RNA is fragmented and incubated with anti-m6A. After cross-linking with UV light (254 nm), covalently bound antibody-RNA complexes are recovered by protein A/G-affinity purification, SDS-PAGE and nitrocellulose-membrane transfer. RNA is then released from the membrane by proteinase K and reverse transcribed. Peptide fragments that remain on the RNA lead to nucleotide-incorporation errors (indicated as C→T transition) and cDNA truncations. (**d**) miCLIP reads from HEK293 total RNA were aligned to rRNA and analyzed for mismatches at the two most abundant cellular m6A residues (filled red circles) at positions 1,832 and 4,190 of 18S and 28S rRNA, respectively. As a control, a non-methylated DRACH consensus site at position 1,801 of 18S rRNA was analyzed (open circle). C→T transitions are shown in turquoise; other single-nucleotide mismatches are indicated by dark blue, brown and yellow. (**e**) Quantitative representation of the C→T transitions in 18S (top) and 28S (bottom) rRNAs.

introduced during reverse transcription[9,14]. Some antibodies induced an unpredictable pattern of substitutions at different positions relative to the m6A residue, making it difficult to unambiguously determine the position of the m6A (**Fig. 1a**,**b**). However, other antibodies resulted in more consistent substitution patterns. For example, the Abcam antibody (Online Methods) induced a nucleotide substitution at the invariant cytosine residue adjacent to the m6A as well as at the m6A itself (**Fig. 1a**).

We also analyzed truncations and found that the Synaptic Systems (SySy) polyclonal antibody (Online Methods) efficiently induced truncations at the +1 position relative to the m6A (**Fig. 1b**). Although other antibodies also induced specific mutation patterns, the Abcam and SySy polyclonal antibodies were chosen for further investigation because of their high immunoprecipitation efficiency and predictable patterns of cross-link–induced substitutions and truncations.

Transcriptome-wide characterization of cross-linking sites

To exploit these mutational signatures for transcriptome-wide mapping of m6A, we developed m6A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP). In miCLIP, cellular RNA is sheared and cross-linked to anti-m6A using UV light (**Fig. 1c**). RNA fragments cross-linked to antibody are then purified and converted into a cDNA library according to the iCLIP protocol[9]. Cross-link–induced mutations and truncations introduced during reverse transcription are then analyzed to determine the precise positions of m6A throughout the transcriptome.

We generated two miCLIP libraries from human embryonic kidney (HEK293) cells using total cellular RNA cross-linked to either Abcam or SySy antibodies. Comparison of the two libraries revealed strong enrichment of C→T, but not of any other transitions, in the Abcam library (**Supplementary Fig. 1a**,**b**). Thus, C→T transitions may serve as a signature mutation for mapping m6A with miCLIP using the Abcam antibody.

To determine whether C→T transitions are markers of m6A in cellular RNA, we examined these mutations at known m6A residues in human 18S and 28S rRNA[15]. Mapped reads exhibited a high frequency of C→T transitions at +1 positions relative to these m6A residues, whereas reads covering a non-methylated site did not show such enrichment (**Fig. 1d**). Strikingly, quantitative analysis across the length of the 18S and 28S transcripts showed that C→T transitions were enriched by ~500-fold and ~1,000-fold at m6A positions, respectively (**Fig. 1e**). To test whether the Abcam antibody induced other mutations near m6A residues, we analyzed single-nucleotide substitutions around RAC triplets—the core m6A motif—throughout the transcriptome. The only mutation seen with high frequency at these triplets was the C→T transition at position +1 relative to the A (**Supplementary Fig. 1c**). Thus, the Abcam antibody induced C→T transitions at m6A in a highly position-specific manner.

Notably, the SySy antibody, which primarily induced truncations, also induced C→T transitions at the +1 position. Thus, C→T transitions may be a common feature with antibodies to m6A. However, in the SySy library, other substitutions were present, and the mutations occurred with lower positional accuracy (**Supplementary Fig. 1d**). Because *in vitro* analysis indicated that the SySy antibody efficiently induced cDNA truncations at the +1 position of m6A (**Fig. 1b**), truncations were used as this antibody's mutational signature.

Identification of m6A using antibody-induced mutations

To identify m6A residues throughout the transcriptome, we generated independent maps of human m6A residues in HEK293 mRNA using the Abcam and SySy antibodies.

© 2015 Nature America, Inc. All rights reserved.

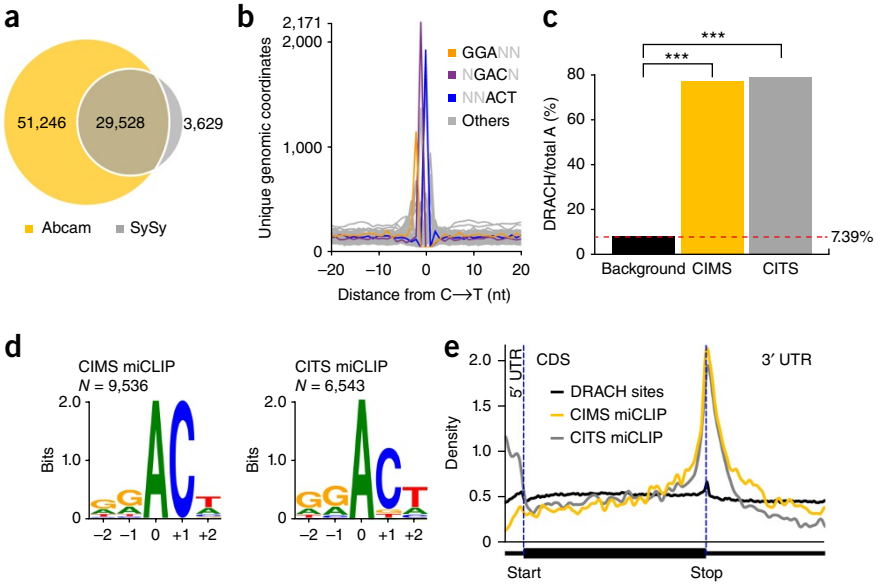

Figure 2 | C→T transitions and truncations map m6A throughout the transcriptome. (**a**) Overlap analysis of miCLIP peak clusters generated by the Abcam and SySy antibodies. (**b**) Frequency of nucleotide triplets at C→T transitions. The most abundant 3-mers at positions −1, 0 and +1 were GGA, GAC and ACT (shown in color). These reconstitute the most common m6A consensus sequence, GGACT. AAC, the other RAC motif known to be highly methylated, was the second most abundant 3-mer at the transition site (grouped here with "Others," in gray). (**c**) Adenosines called by miCLIP were significantly more frequent in the DRACH sequence context than expected on the basis of the background distribution of the motif (***$P < 1 \times 10^{-15}$, Fisher's exact test). (**d**) Transcriptome-wide sequence logos of C→T transitions and truncations were identical to the m6A consensus motif. Weblogo analyses of the sequence environment of C→T transitions (left) and truncations (right) (m6A is at position 0). (**e**) miCLIP-identified m6A residues showed a metagene distribution profile typical for m6A. Metagene distribution plots of DRACH consensus site background (black) and the m6A residues predicted by CIMS-based (yellow) and CITS-based (gray) miCLIP.

For the Abcam library, we prepared short antibody-bound RNA fragments (~35 nt) and used paired-end sequencing to reduce mutation noise (Online Methods). Mapping these reads to the reference genome resulted in ~40-nt-wide peaks, unlike the ~200-nt-wide peaks seen in MeRIP-Seq (**Supplementary Fig. 2**). Combining reads from four replicates resulted in the identification of 80,774 peaks on mRNA and 92,068 C→T transitions at 48,694 different genomic positions.

In the SySy library, peaks were similar in width and overall shape to those produced by the Abcam antibody (**Supplementary Fig. 2**). We detected 33,157 peaks in this library that mapped to mRNAs. These peaks had a high degree of positional overlap (89.06%) with the Abcam peaks (**Fig. 2a**).

We next sought to validate that C→T transitions induced by the Abcam antibody could be found at m6A residues throughout the transcriptome. Biochemical experiments have demonstrated that most m6A in mRNA is located in a GAC or AAC sequence[16], and indeed, we found that GAC and AAC were strongly enriched at transition sites (**Fig. 2b**). Furthermore, the GGA and ACT triplets were enriched at −1 and +1 positions, respectively, recapitulating the most prevalent m6A consensus sequence, GGACT (**Fig. 2b**). Thus, we concluded that C→T transitions predominantly occur at m6A consensus motifs.

To call C→T transitions, we used a computational pipeline designed for the identification of cross-linking–induced mutation sites (CIMSs) in high-throughput sequencing CLIP data[17] (Online Methods). This resulted in a set of 11,832 called sites. This set was enriched in adenosines at the −1 position of C→T transitions (80.59%), supporting the notion that these transitions largely reflect m6A. Furthermore, 77.29% of these adenosines occurred in a DRACH consensus motif, a value that is significantly higher than would be expected on the basis of the background distribution of this motif in mRNA (**Fig. 2c**; $P < 1 \times 10^{-15}$, Fisher's exact test). CIMS-based miCLIP (CIMS miCLIP) identified 9,536 putative m6A residues in the transcriptome (**Supplementary Table 1** and **Supplementary Fig. 3b**).

Identification of m6A using antibody-induced truncations

To analyze truncations in the SySy library, we used a computational pipeline for detecting cross-linking–induced truncation sites (CITSs) in CLIP data[18]. This resulted in 8,329 significant ($P < 0.05$) truncation sites that mapped to mRNAs. Most of these truncations occurred at adenosines (77.10%). CITS-based miCLIP (CITS miCLIP) identified 6,543 putative m6A sites (**Supplementary Table 2**). These were significantly enriched in DRACH consensus sites (**Fig. 2c**; 79.46%, $P < 1 \times 10^{-15}$, Fisher's exact test).

Validation of m6A residues identified by miCLIP

Both CIMS- and CITS-called sites were localized predominantly in the coding sequence and the 3′ UTR of mRNA (**Supplementary Fig. 3a**), consistent with the known distribution of m6A[1,2]. Sequence logo analysis of both data sets confirmed that called sites occurred in the m6A consensus motif DRACH (**Fig. 2d**). Additionally, both metagene profiles followed the typical distribution of m6A with strong enrichment at the stop codon (**Fig. 2e**). These data suggest that miCLIP identifies true m6A residues.

Next, we examined the accuracy of m6A identification by miCLIP. We compared miCLIP sites to a control set of adenosines that had been biochemically validated for their N^6-methylation status using the thin-layer chromatography–based method SCARLET[19]. This data set included a positive control set of eight m6A residues in five transcripts, as well as 15 adenosine residues in a DRACH sequence context that were not methylated.

To estimate the sensitivity and specificity of miCLIP, we determined the number of SCARLET-positive and SCARLET-negative sites that were called (**Supplementary Fig. 3c**). CIMS and CITS miCLIP identified six and five of the eight SCARLET-positive sites, respectively. Importantly, none of the 15 SCARLET-negative sites were called by CIMS miCLIP, and only one was called by CITS miCLIP. Despite the small size of the SCARLET-derived set, this suggests that miCLIP detects m6A with high specificity and sensitivity.

© 2015 Nature America, Inc. All rights reserved.

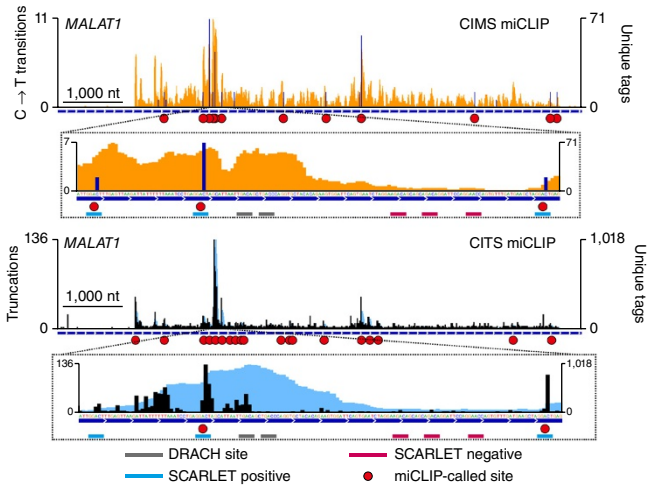

Figure 3 | miCLIP identifies m6A with single-nucleotide resolution. m6A residues detected by CIMS and CITS miCLIP in the ncRNA *MALAT1*. Orange and dark blue tracks denote CIMS miCLIP unique read coverage and C→T transitions, respectively. Light blue and black tracks denote CITS miCLIP unique read coverage and unique read starts, respectively. Horizontal blue bars denote transcript models. Filled red circles denote miCLIP-called m6A. Small horizontal bars in insets denote DRACH consensus sites with a methylation status that is undefined (gray), confirmed positive (turquoise) or confirmed negative (magenta) by SCARLET[19].

The accuracy of miCLIP could be seen when we examined individual transcripts. For example, in MALAT1, three SCARLET-negative sites were located in a 110-nt window between two SCARLET-positive m6A residues (**Fig. 3**). Although both SCARLET-positive sites were correctly identified by CIMS and CITS miCLIP, none of the interspersed negative sites were called. Thus, miCLIP showed excellent spatial resolution and a low false discovery rate.

m6A in clusters and infrequently methylated DRACH motifs

Clustered m6A residues have been predicted from the shape, size and distribution of MeRIP-Seq peaks[1]. Using CIMS miCLIP, we identified 958 clusters ranging in size from ~150 to ~500 nt (Online Methods and **Supplementary Fig. 4a**) that contained up to 15 m6A residues (**Supplementary Fig. 4b**). The average distance between m6As in these clusters was 64 nt.

Because MeRIP-Seq peaks are typically ~100–200 nt wide and the bioinformatic prediction of m6A residues is limited to one site per peak[5], this approach misses a substantial portion of clustered m6As. Indeed, although m6A clusters had 2.39 miCLIP-called m6As on average, they contained only 1.06 MeRIP-Seq–predicted m6As. This indicates that peak-based prediction algorithms may miss more than half of m6A residues occurring in clusters. In contrast, miCLIP identified individual m6A residues separated by as few as 3 nt (**Supplementary Fig. 4c**).

Bioinformatics methods use a predefined subset of consensus motifs to predict the m6A residue within a peak, according to the idea that some DRACH pentamers are preferred targets for methylation[5]. However, this approach misses m6As that occur in motifs outside of this predefined subset. As miCLIP does not require a priori assumptions about the sequence context of m6A (except for the invariant cytosine in CIMS miCLIP), it identifies m6A in all possible motifs. We determined the exact distribution of consensus sequences in which m6A occurs (**Supplementary Fig. 5**). Our findings confirmed that most m6A residues reside in a subset of DRACH motifs[5]. In fact, 41% and 50% of m6A residues detected by CIMS and CITS miCLIP, respectively, resided in just four subtypes of the DRACH motif. However, a considerable portion of m6As (23% and 31% as determined by CITS and CIMS miCLIP, respectively) occurred in DRACH motifs that would be missed by bioinformatic prediction.

CITS miCLIP identifies m6Am at the TSS

We next asked whether miCLIP can identify m6Am, a related base modification found in certain mRNAs in the first position after the 7-methylguanosine cap[12]. The function of m6Am is poorly understood; thus, mapping m6Am is important for elucidating its functional role *in vivo*.

Unlike m6A, which is not detected at the first position of mRNA, m6Am is limited to the first position in transcripts[12]. Recently, a MeRIP-Seq approach coupled with bioinformatic analysis that detects RNA 5′ ends was used to predict methylated adenines at transcription start sites (TSSs)[5]. We reasoned that miCLIP could also be used to map m6Am at TSSs throughout the transcriptome. Indeed, the metagene profile of sites identified by CITS miCLIP showed an enrichment of called sites in the 5′ UTR (**Fig. 2e**). This enrichment was absent in CIMS miCLIP (**Fig. 2e**), presumably because CIMS miCLIP requires a C at the +1 position of the modified A.

Because CITS miCLIP is not dependent on sequence context, we reasoned that it could identify m6Am at the TSS. We analyzed 797 truncations localized to 5′ UTRs and found that they occurred in the canonical m6A motif DRACH, as well as in a BCA* motif (where B denotes C, U or G, and A* denotes methylatable A) (**Fig. 4a**). 5′ UTRs contained nearly three times as many methylated BCA motifs as DRACH motifs (434 sites versus 151 sites). Interestingly, the extended form of the BCA motif (**Fig. 4a**) resembled the known pyrimidine-rich sequence at TSSs[20–23], further suggesting that these sites were m6Am rather than internal m6A.

To further address the position of putative m6Am sites, we performed metagene analysis of truncations occurring in DRACH and BCA sequence contexts (**Fig. 4b**). Whereas truncations in DRACH sites followed the canonical m6A distribution, truncations in BCA sequence contexts localized primarily to the 5′ UTR (**Fig. 4b**), with the greatest enrichment near the annotated TSS (**Fig. 4c**).

Finally, we sought to validate the idea that CITS miCLIP recognizes m6Am by comparing our data to a known set of m6Am-containing mRNAs[24]. We focused on histone mRNAs, which have been biochemically shown to start with m6Am[24]. Indeed, CITS miCLIP called truncation sites at the m6Am position of three histone mRNAs (**Fig. 4d**). Importantly, the truncation efficiency at these sites was not 100%, which indicated that these truncations were induced by antibody cross-linking rather than reverse transcriptase reaching the transcript end (**Fig. 4d** and **Supplementary Fig. 6**). These data show that CITS miCLIP was able to detect two distinct RNA modifications, m6A and m6Am, at single-nucleotide resolution throughout the transcriptome.

CIMS miCLIP identifies m6A residues in snoRNAs

An important criterion for calling peaks in MeRIP-Seq is that peaks must be higher than piled-up reads in adjacent areas in the same transcript, which define the background read level[1]. However, small RNAs can be entirely covered with reads, making

© 2015 Nature America, Inc. All rights reserved.

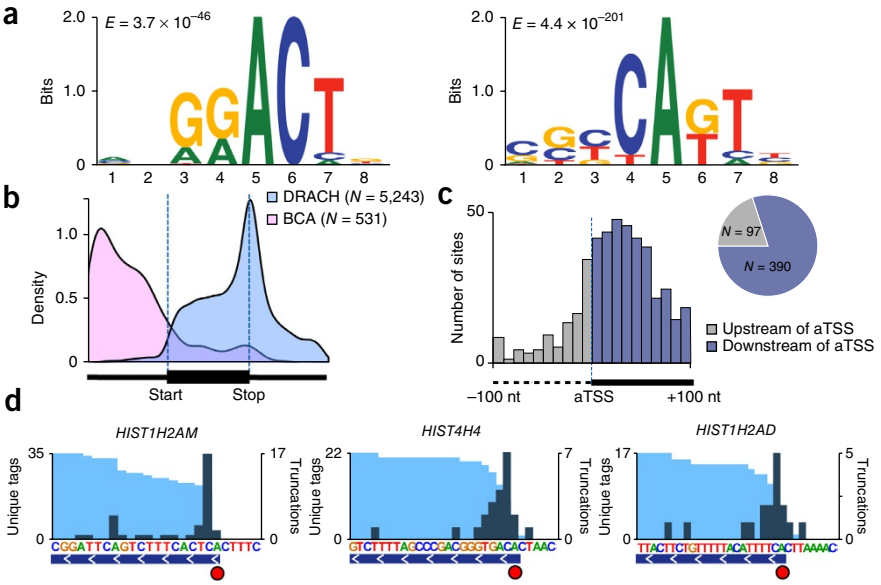

Figure 4 | Antibody-induced truncations map m6Am throughout the transcriptome. (**a**) MEME analysis of the sequence environment (−4 to +4) of 797 truncations localized to 5′ UTRs identified two predominant motifs: the canonical m6A motif DRACH (left) and a motif that is best described by the consensus sequence BCA (right). (**b**) Truncations in DRACH and BCA sequence contexts follow distinct metagene distribution profiles. (**c**) Truncations in CITS miCLIP clustered around the annotated transcription start site (aTSS). The plot shows the number of truncations in 10-nt bins between −100 nt and +100 nt relative to the aTSS. The pie chart shows the proportion of truncations found upstream and downstream of the aTSS. (**d**) Truncations identify m6Am with single-nucleotide resolution. Examples of known m6Am-containing 5′ UTRs are shown (light blue and teal tracks denote unique CITS miCLIP read coverage and unique read starts, respectively; horizontal blue bars denote transcript models; filled red circles denote called m6Am residues).

it difficult to establish the background. Thus m6A residues in small RNAs are particularly difficult to identify with MeRIP-Seq. However, because CIMS miCLIP identifies C→T transitions, m6A residues can be readily detected in small RNAs by that method.

To validate that CIMS miCLIP can detect m6A residues in small RNAs, we focused on snoRNAs. C→T transitions identified m6A residues in both H/ACA and C/D box snoRNA subclasses, with more than 25% of both classes having at least one m6A (**Fig. 5a**,**b**). For example, we detected a high rate of C→T transitions in snoRNAs *Snora64* and *Snord2* (**Fig. 5c**). These transitions were found in typical DRACH motifs (**Fig. 5d**). The discovery of m6A in snoRNAs complements the finding of pseudouridine in snoRNAs[25], which suggests that this class of ncRNAs is regulated by diverse RNA modifications.

DISCUSSION

m6A is the most widespread base modification in mRNA, but a method for identifying specific residues has been lacking. We showed that m6A residues can be mapped by generating signature mutations with m6A-specific antibodies and UV cross-linking. We mapped m6A residues using two antibodies, each of which had specific advantages: one antibody produced a C→T transition that was limited to detecting m6A but could readily detect clustered m6A residues, and the other antibody produced truncations that could be used to map m6A and m6Am residues simultaneously. Although our experiments used specific antibodies, our approach to screening cross-link–induced mutations should make it straightforward to identify other m6A antibodies suitable for miCLIP. This approach could additionally be used to map other RNA modifications for which antibodies exist.

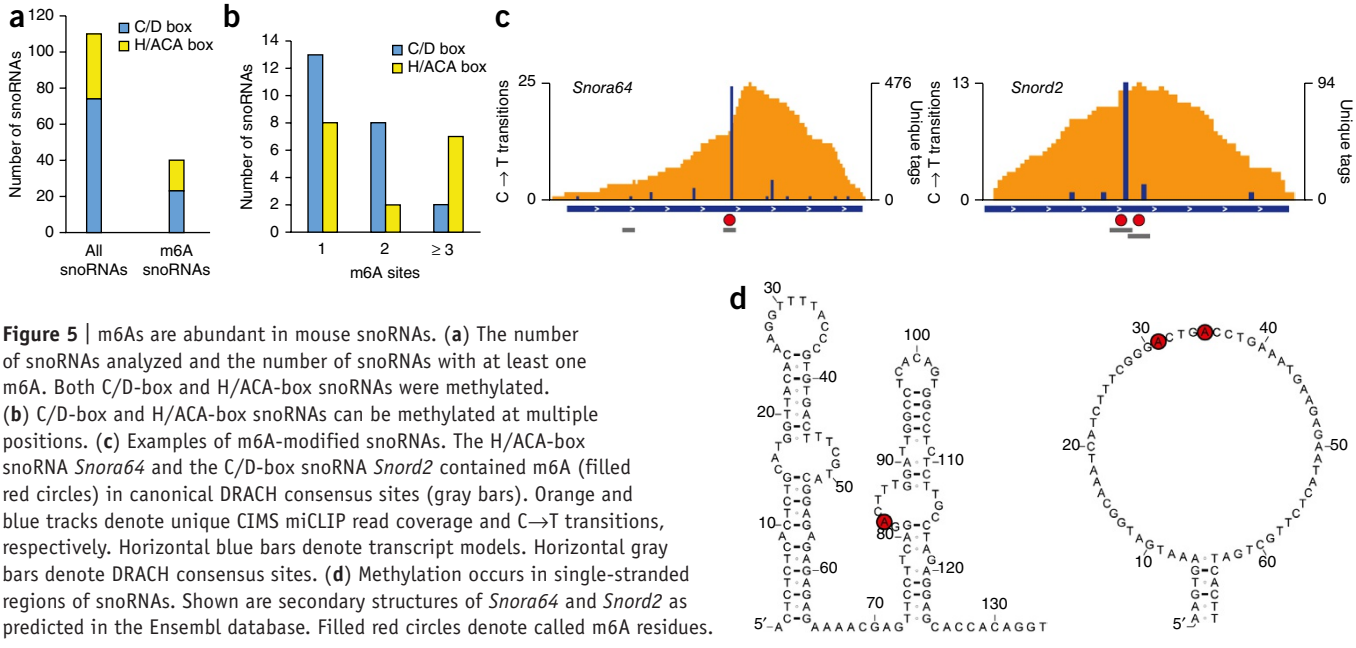

Figure 5 | m6As are abundant in mouse snoRNAs. (**a**) The number of snoRNAs analyzed and the number of snoRNAs with at least one m6A. Both C/D-box and H/ACA-box snoRNAs were methylated. (**b**) C/D-box and H/ACA-box snoRNAs can be methylated at multiple positions. (**c**) Examples of m6A-modified snoRNAs. The H/ACA-box snoRNA *Snora64* and the C/D-box snoRNA *Snord2* contained m6A (filled red circles) in canonical DRACH consensus sites (gray bars). Orange and blue tracks denote unique CIMS miCLIP read coverage and C→T transitions, respectively. Horizontal blue bars denote transcript models. Horizontal gray bars denote DRACH consensus sites. (**d**) Methylation occurs in single-stranded regions of snoRNAs. Shown are secondary structures of *Snora64* and *Snord2* as predicted in the Ensembl database. Filled red circles denote called m6A residues.

© 2015 Nature America, Inc. All rights reserved.

An important feature of miCLIP is that it can be used without pretreatment of cells with modified nucleotides such as 4-thiouracil (4SU), which is used in photoactivatable ribonucleoside–enhanced (PAR)-CLIP[10]. Indeed, in a recent study antibodies to m6A were cross-linked to 4SU-labeled RNA[26]. RNA fragments that cross-linked to the antibody contained characteristic T→C mutations arising from 4SU cross-links. This strategy is valuable, as it reduces noise by identifying and narrowing PAR-CLIP peak clusters. However, in that study the T→C mutations were not used to map specific m6A residues throughout the transcriptome. Unlike miCLIP, which is associated with signature mutations, PAR-CLIP experiments can generate multiple 4SU transitions at protein-binding sites in transcripts[10]. Because of the inconsistent number and position of transitions relative to each m6A residue, it is challenging to use T→C mutations for m6A identification. In contrast, miCLIP relies on characteristic mutational signatures that are highly predictive of m6A residues.

Direct identification of m6A provides advantages over bioinformatic prediction of m6A residues from MeRIP-Seq peaks. Bioinformatic prediction is reliable if the m6A peak has a single clear summit, is caused by a single m6A residue and has a single centrally located DRACH motif. However, m6A residues are often clustered in mRNAs[1]. As a result, MeRIP-Seq peaks are often broad with diverse shapes, and the summits of these peaks might not reflect the position of the m6A residues that account for the peak shape. In miCLIP, peak shape does not influence m6A identification. Furthermore, miCLIP does not constrain m6A identification to m6A residues that fall within a specified subset of DRACH motifs. Thus, miCLIP enables unbiased identification of m6A residues.

The inability to identify specific methylated adenosines in mRNA at single-nucleotide resolution has hampered functional studies of m6A. Indeed, recent studies proposing functions for m6A have not mutated specific adenosine residues to unequivocally determine the effect of m6A on a specific transcript. The single-nucleotide data set presented here will make it straightforward to test the role of any m6A in determining the fate or function of any transcript.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Data were deposited in NCBI's Gene Expression Omnibus (GEO) under accession number GSE63753.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank K. Meyer and D. Patil for useful comments and suggestions. This work was supported by NIH grants NIDA DA037150 (to S.R.J.), NS076465 (to C.E.M.), T32 HD060600 (to A.V.G.) and T32 CA062948 (to A.O.O.-G.); a German Research Foundation (DFG) fellowship (to B.L.); the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts; the STARR Consortium (I7-A765 to C.M.); the Vallee Foundation (C.M.); and the WorldQuant Foundation (C.E.M.).

AUTHOR CONTRIBUTIONS

B.L., A.V.G., A.O.O.-G. and S.R.J. conceived and designed the experiments and analyzed the data; C.M. and C.E.M. analyzed mutational profiles of initial miCLIP libraries; B.L., A.V.G., A.O.O.-G. and S.R.J. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Meyer, K.D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
2. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
3. 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).
4. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. USA* **71**, 3971–3975 (1974).
5. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. *Cell Rep.* **8**, 284–296 (2014).
6. Squires, J.E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **40**, 5023–5033 (2012).
7. Ryvkin, P. *et al.* HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684–1692 (2013).
8. Sugimoto, Y. *et al.* Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* **13**, R67 (2012).
9. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
10. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
11. Ule, J. *et al.* CLIP Identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
12. Schibler, U. & Perry, R.P. The 5′-termini of heterogeneous nuclear RNA: a comparison among molecules of different sizes and ages. *Nucleic Acids Res.* **4**, 4133–4149 (1977).
13. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070 (2014).
14. 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).
15. Piekna-Przybylska, D., Decatur, W.A. & Fournier, M.J. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.* **36**, D178–D183 (2008).
16. Schibler, U., Kelley, D.E. & Perry, R.P. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *J. Mol. Biol.* **115**, 695–714 (1977).
17. 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).
18. Weyn-Vanhentenryck, S.M. *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* **6**, 1139–1152 (2014).
19. Liu, N. *et al.* Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848–1856 (2013).
20. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
21. Ni, T. *et al.* A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods* **7**, 521–527 (2010).
22. Plessy, C. *et al.* Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nat. Methods* **7**, 528–534 (2010).
23. Frith, M.C. *et al.* A code for transcription initiation in mammalian genomes. *Genome Res.* **18**, 1–12 (2008).
24. Moss, B., Gershowitz, A., Weber, L.A. & Baglioni, C. Histone mRNAs contain blocked and methylated 5′ terminal sequences but lack methylated nucleosides at internal positions. *Cell* **10**, 113–120 (1977).
25. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
26. Chen, K. *et al.* High-resolution N6-methyladenosine (m6A) map using photo-crosslinking-assisted m6A sequencing. *Angew. Chem. Int. Ed. Engl.* **54**, 1587–1590 (2015).

© 2015 Nature America, Inc. All rights reserved.

ONLINE METHODS

Antibodies. We used the following antibodies to m6A: rabbit polyclonal anti-m6A (202 003, Synaptic Systems; ab151230, Abcam; ABE572, Millipore; and 61495, Active Motif) and mouse monoclonal anti-m6A (202 011, Synaptic Systems).

***In vitro* mutagenesis assay.** A 1,502-nt-long RNA containing a single adenosine nucleotide at position 966 was transcribed in the presence of GTP, CTP, UTP and either ATP or m^6ATP with the Ampliscribe *in vitro* transcription kit (Epicentre). Then 6 µg of the fragmented transcript were incubated with 4 µg of each of the m6A antibodies tested. After UV cross-linking with 0.15 J cm^{-2} UV light (254 nm), antibody-RNA complexes were processed as described for cellular RNA, and a library was prepared for each antibody. Libraries were then sequenced on a MiSEQ instrument, and reads covering the m6A position with single mismatches at positions −2 to +4 were quantified. For each nucleotide at positions −2 to +4 of the m6A, we determined the frequency of truncation and transition events. For this, reads terminating at a given position (for truncations) and single-nucleotide mismatches at that position (for transitions) were counted and normalized to the total number of reads covering the m6A residue.

Cell lines and animals. For characterizing m6A residues in humans, we used HEK293 cells (CRL-1573, ATCC). Cells were tested for mycoplasma by Hoechst staining. Cell-line verification was not performed. For profiling m6A residues in mice, we used mouse liver (Charles River; CD-1/ICR, female, 8–10 weeks old). All experiments involving mice were approved by the Institutional Animal Care and Use Committee at Weill Cornell Medical College.

Preparation of mouse liver nuclei. For the identification of m6A residues in small nucleolar RNAs (snoRNAs), we collected intact nuclei from mouse liver using an iodixanol gradient as described previously[27]. All steps were performed at 4 °C. In brief, liver was homogenized in hypo-osmotic medium (250 mM sucrose, 25 mM KCl, 5 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.4) with a dounce homogenizer. The homogenate was then mixed with 50% iodixanol (D1556, Sigma-Aldrich, diluted with 250 mM sucrose, 150 mM KCl, 30 mM $MgCl_2$, 60 mM Tris-HCl, pH 7.4) to a final concentration of 25% iodixanol. This mixture was then underlayered with 30% and 35% iodixanol and centrifuged in a swinging-bucket rotor for 20 min at 10,000*g*. Nuclear bands were collected, and nuclear RNA was extracted with Trizol (Life Technologies).

Cross-linking of cellular RNA. For initial miCLIP libraries, total RNA was purified from HEK293 cells. Poly(A)$^+$ RNA was prepared from four biological replicates of HEK293 cells and processed in parallel for miCLIP. Mouse liver RNA was depleted of ribosomal RNA by Ribominus (Life Technologies).

We used fragmentation reagent (Life Technologies) to fragment RNA to a size between 30 and 130 nt. After the reaction had been stopped, 20 µg fragmented RNA was directly diluted in 450 µl immunoprecipitation buffer (50 mM Tris, pH 7.4, 100 mM NaCl, 0.05% NP-40) and incubated with 1–5 µg anti-m6A at 4 °C for 1–2 h, rotating head over tail. The solution was then transferred into a 3-cm cell culture dish and cross-linked twice with 0.15 J cm^{-2} UV light (254 nm) in a Stratalinker (Agilent).

After cross-linking, the solution was transferred into Eppendorf tubes and incubated with 30 µl Protein A/G beads (Thermo Scientific) for 1–2 h at 4 °C, rotating. Bead-bound antibody-RNA complexes were then recovered on a magnetic stand (Life Technologies) and washed twice with high-salt buffer (50 mM Tris, pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS), twice with immunoprecipitation buffer, and twice with polynucleotide kinase (PNK) wash buffer (20 mM Tris, 10 mM $MgCl_2$, 0.2% Tween 20).

3′ dephosphorylation, linker ligation and labeling. The protocol for these steps was similar to the protocol described for iCLIP[9]. In brief, RNA 3′ ends were dephosphorylated on beads with PNK (M0201S, New England BioLabs) for 30 min in dephosphorylation buffer (70 mM Tris, pH 6.5, 10 mM $MgCl_2$, 1 mM DTT). After another round of extensive washing (twice with PNK wash buffer, once with immunoprecipitation buffer, once with high-salt buffer, and twice with PNK wash buffer), the 3′ adaptor was ligated with T4 RNA ligase (New England BioLabs) overnight.

Antibody-RNA complexes were then eluted from beads with 1× NuPage sample buffer (Life Technologies) containing 50 mM DTT, subjected to NuPage gel electrophoresis and transferred onto 0.45-µm nitrocellulose membranes (Bio-Rad). After autoradiography (2 h to overnight), membrane regions containing RNA–cross-linked antibody heavy and light chains were excised and the RNA was released from the membrane by treatment with proteinase K (Life Technologies).

Library preparation. After phenol-chloroform extraction and precipitation, the RNA was reverse transcribed with Superscript III reverse transcriptase (Life Technologies) according to the manufacturer's protocol. First-strand cDNA was size-selected on a 6% TBE-Urea gel (Life Technologies), and regions corresponding to 80–120 nt (for the total RNA data set) or 70–100 nt (for the mRNA data set) were used for further analysis. Circularization and re-linearization of cDNA were performed as described in the iCLIP protocol[9]. Libraries were PCR amplified for 18–21 cycles and sequenced on an Illumina HiSeq 2500 instrument.

Read preprocessing. Fastq files were adaptor trimmed using flexbar[28] and demultiplexed on the basis of their experimental barcode using the pyBarcodeFilter.py script of the pyCRAC tool suite[29]. The second part of the iCLIP random barcode was then moved to the read headers with the Unix tool awk (awk -F "##" '{sub(/..../,"##"$2, $2); getline($3); $4 = substr($3,1,2); $5 = substr($3,3); print $1 $2 $4"\n"$5}'). For the HEK293 mRNA data set (CIMS miCLIP), reads from the four replicates were combined at this point. Sequence-based removal of PCR duplicates was then performed with the pyFastqDuplicateRemover.py script[29]. Awk was then used to generate read headers compatible with downstream processing by the CIMS pipeline (awk -F '[_/]' '/^>/{print $1"_"$2"_"$3"/"$4"#"$3"#"$2; getline($9); print $9}').

Processing of paired-end data. To increase base coverage and thus facilitate the discrimination of mutations (which should be at the same position in both reads) from random sequencing errors (which should be evenly distributed in the reads), we generated short insert libraries (~30–40 nt) and subjected them to paired-end sequencing. For these paired-end data, reverse

© 2015 Nature America, Inc. All rights reserved.

reads were reverse complemented and processed like their forward mates. However, to distinguish the forward and reverse mates as distinct reads and prevent their collapse during PCR duplicate removal, we assigned the reverse complement of the random barcode to the reverse reads using a custom Perl script (**Supplementary Software**).

Read mapping. Reads were mapped with Novoalign (Novocraft Technologies Sdn Bhd) to a custom build of rRNA (28S, acc. no. NR_003287.2; 18S, acc. no. NR_003286.2; and 5.8S, acc. no. NR_003285.2) or to the human (hg19) and mouse (mm10) genomes. A maximum alignment score of 85 was allowed, iterative trimming was used, and parameters were adjusted to allow mapping of short reads (Novoalign -t 85 -s 1 -l 16 -F FA -o Native -r None).

Mutation calling. Positions of C→T transitions and the coordinates of mapped reads were extracted from native Novoalign output files using the novoalign2bed.pl script of the CIMS software package[14]. Reads that mapped to the same start and end coordinates and that shared the same random barcode were then collapsed into unique tags using the CIMS script tag-2collapse.pl. For each mismatch position, the unique tag coverage (k) and the number of C→T transitions (m) were determined using the CIMS.pl program[14]. Because the cluster permutation approach of the CIMS algorithm[14] is not suited for assigning a false discovery rate to a specific transition subtype (i.e., the C→T transitions analyzed here), a filtering approach was used to minimize calling of false positive sites. In a first step, known SNPs (dbSNP 138) and regions marked as repetitive sequences in the respective genome were removed. Then we filtered sites by their number of C→T transitions (m) and the ratio of C→T transitions to unique tags covering that position (m/k). First, to reduce the number of mismatches caused by random errors introduced during reverse transcription, PCR amplification and library sequencing, each transition had to be called at least twice ($m2$). Second, the ratio of mutant reads to total reads covering a position was restricted to between 1% and 50% ($m/k1$–50). This further reduced noise and simultaneously removed sites with very high mismatch rates such as produced by SNPs and mismapping artifacts. Together, these filters ($m2$ and $m/k1$–50) reduced the signal-to-noise ratio more than twofold and the calling of non-A positions to less than 20%. This resulted in a final data set of 9,536 m6A residues in HEK293 cells and 780 m6A residues in mouse liver nuclei.

Truncation calling. Coordinates of mapped reads were extracted, and reads carrying the same random barcode that mapped to identical coordinates were collapsed as for mutation analysis (above). Then we used the CITS software package to call cross-link sites[18]. In brief, coverage of unique reads was calculated with the tag-2profile.pl script. Then, to detect potential cross-link–induced truncation sites, we used the bedExt.pl script of the CITS software package to map positions immediately upstream of the first nucleotide of each unique read. To differentiate reads that represented truncations from read-through reads, we used the CIMS algorithm to identify unique reads carrying cross-link–induced deletions. Then we used a Bash script and joinWrapper.py from the CITS package to remove unique reads carrying deletions at potential truncation sites. The remaining reads were then clustered into peaks with tag2cluster.pl. We then shuffled potential truncation positions (read-start "pileups") within their respective peak clusters using tag2peak.pl to identify truncation events of statistical significance. Events with a significance value of $P \leq 0.05$ that occurred at adenosines were retained to yield a list of 12,051 m6A residues in HEK293 cells. The sequence context of truncations was determined with the MEME motif discovery tool[30].

Peak calling. To determine the number of peaks generated by miCLIP, we clustered unique reads from each data set into peaks using the tag2cluster.pl script of the CIMS software package. The resulting list of peak clusters was filtered for those with at least four stacked reads to generate 80,774 HEK293 and 14,055 mouse liver nuclei peaks for CIMS miCLIP and 33,157 HEK293 peaks for CITS miCLIP.

Analysis of clustered m6A. To define sites of clustered methylation, we analyzed sliding windows (length, 100 nt; step size, 25 nt) that overlapped with annotated RefSeq transcripts for m6A site coverage. Windows with two or more m6As were then merged into a cluster. Finally, the number of miCLIP-called m6A residues in each cluster was determined. In addition, we counted the number of MeRIP-Seq coverage-predicted m6As[5] in each cluster.

Annotation of m6A residues. Called sites were annotated with the annotatePeaks.pl script from the Homer software suite[31]. To compare the genomic distribution of m6A residues with the distribution of meRIP-Seq peaks, we analyzed a previously published HEK293 MeRIP-Seq data set[1] in the same way.

27. Graham, J.M. Isolation of nuclei and nuclear membranes from animal tissues. *Curr. Protoc. Cell Biol.* Chapter 3, Unit 3.10 (2001).
28. Dodt, M., Roehr, J.T., Ahmed, R. & Dieterich, C. FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* (*Basel*) **1**, 895–905 (2012).
29. Webb, S., Hector, R.D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* **15**, R8 (2014).
30. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
31. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nmeth.3453