Probing N^6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA

NIAN LIU,[1] MARC PARISIEN,[2] QING DAI,[1] GUANQUN ZHENG,[1] CHUAN HE,[1,3] and TAO PAN[2,3,4]

[1]Department of Chemistry, [2]Department of Biochemistry and Molecular Biology, [3]Institute of Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, USA

ABSTRACT

N^6-methyladenosine (m^6A) is the most abundant modification in mammalian mRNA and long noncoding RNA (lncRNA). Recent discoveries of two m^6A demethylases and cell-type and cell-state-dependent m^6A patterns indicate that m^6A modifications are highly dynamic and likely play important biological roles for RNA akin to DNA methylation or histone modification. Proposed functions for m^6A modification include mRNA splicing, export, stability, and immune tolerance; but m^6A studies have been hindered by the lack of methods for its identification at single nucleotide resolution. Here, we develop a method that accurately determines m^6A status at any site in mRNA/lncRNA, termed site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET). The method determines the precise location of the m^6A residue and its modification fraction, which are crucial parameters in probing the cellular dynamics of m^6A modification. We applied the method to determine the m^6A status at several sites in two human lncRNAs and three human mRNAs and found that m^6A fraction varies between 6% and 80% among these sites. We also found that many m^6A candidate sites in these RNAs are however not modified. The precise determination of m^6A status in a long noncoding RNA also enables the identification of an m^6A-containing RNA structural motif.

Keywords: RNA modification; N^6-methyladenosine; single nucleotide resolution

INTRODUCTION

More than 100 types of post-transcriptional RNA modifications have been discovered in all three domains of life. RNA modifications generally fine-tune RNA structure and molecular interactions, and their detailed biological roles are under intense investigation (Grosjean 2005).

N^6-methyladenosine is the most abundant RNA modification in mammalian mRNA and long noncoding RNA (lncRNA), occurring on average in three to five sites per transcript (Bokar 2005). m^6A modification pattern is highly dependent on cell types and conditions (Dominissini et al. 2012; Meyer et al. 2012). Two recently discovered m^6A-demethylases are associated with obesity/diabetes and sperm development, indicating that the dynamics of m^6A modification is biologically important (Jia et al. 2011; Zheng et al. 2012). Proposed m^6A functions include mRNA splicing, export, stability, and immune tolerance (Bokar 2005; Karikó et al. 2005), but m^6A studies in mRNA/lncRNA were hindered by the lack of methods for its precise identification.

Although the m^6A modification in mammalian mRNA has been discovered for nearly 40 yr, robust methods were not available that determine its precise location and modification fraction, collectively defined as the m^6A status for each modification site. Challenges for the determination of m^6A status at single nucleotide resolution include the low abundance of mRNA/lncRNA, the inert reactivity of the methyl group, and interference of potential RNA structures around the modification site. m^6A cannot be identified by the traditional deep sequencing methods because reverse transcriptase is insensitive to its presence in cDNA synthesis. The recently developed m^6A/MeRIP-seq method combines m^6A antibody immune-precipitation and deep sequencing to locate m^6A residues in approximately 200 nucleotide RNA segments (Dominissini et al. 2012; Meyer et al. 2012). These results reveal that m^6A modifications are enriched in the 3′ UTR near the stop codons, and the m^6A modification patterns change in different cell types or when cells are stressed.

[4]**Corresponding author**
E-mail taopan@uchicago.edu

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.041178.113.

However, the m^{6}A/MeRIP-seq method cannot identify which adenosine residue under the deep sequencing peaks is actually modified, nor can it determine the modification fraction for any modification site. As the dynamic m^{6}A status could be crucial for its function, new methods that can unambiguously determine the m^{6}A status at single nucleotide resolution are needed to further understand the biological function of this highly abundant modification. To address this challenge, we developed a method that directly measures the precise location and modification fraction, i.e., the m^{6}A status in any candidate site in mRNA/lncRNA at single nucleotide resolution.

RESULTS AND DISCUSSION

A method to detect m^{6}A status in mRNA/lncRNA

Our method combines site-specific cleavage of each m^{6}A-containing candidate site followed by radiolabeling and site-specific ligation and complete nuclease digestion to enable isolation of the candidate m^{6}A containing residue in the mixture of total RNA or polyA$^+$ RNA (Fig. 1A). The candidate m^{6}A containing residue is then analyzed by thin-layer chromatography (TLC) to reveal the presence or absence of m^{6}A and its modification fraction. Using the site-specific RNase H digestion followed by radiolabeling and TLC analysis of an RNA modification residue was applied previously to investigate pseudouridine (Ψ) modification in U2 small nuclear RNA (Zhao and Yu 2004; Ma et al. 2005; Wu et al. 2011). These highly successful studies were made possible by first purifying the abundant U2 snRNA from total RNA. However, the low abundance of individual mRNA/lncRNA makes it very difficult and even impossible for efficient isolation; therefore, modifications in individual mRNA/lncRNA cannot be studied as those in the abundant U2 snRNA.

We aimed at developing a method that directly determines m^{6}A modification status at any mRNA/lncRNA site from the total RNA sample without the need of purifying a specific RNA. To achieve this, we combined RNase H site-specific cleavage, splinted ligation, ribonuclease digestion, and TLC to generate a new method, named site-specific cleavage and

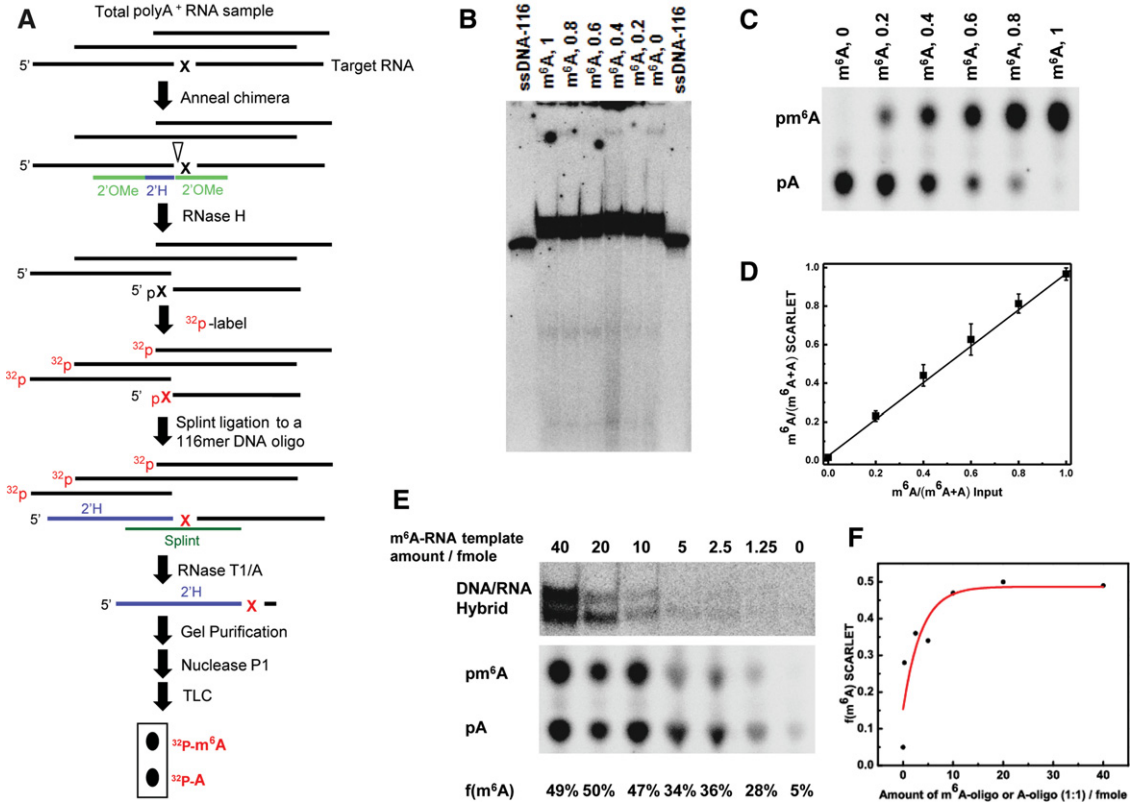

FIGURE 1. Feasibility of SCARLET to investigate m^{6}A status. (*A*) Schematic diagram of SCARLET. (*B*) Denaturing PAGE showing the ^{32}P-labeled bands after ribonucleases T1/A digestion. Lanes m^{6}A-1, 0.8, 0.6, 0.4, 0.2, 0 correspond to 200 fmol total of the synthetic RNA template mixture containing 100%, 80%, 60%, 40%, 20%, 0% m^{6}A in the presence of 1 µg polyA$^+$ RNA from HeLa. ^{32}P-labeled 116 nucleotides single stranded DNA (ssDNA-116) was loaded as a control. (*C*) TLC result of the purified gel bands from *B*. (*D*) Quantitative plot for the TLC result from *C* from three independent replicates. (*E*) SCARLET result for varying amount of oligonucleotide mixtures containing equal moles of m^{6}A-modified and unmodified RNA templates in 1 µg HeLa polyA+ RNA. The bands in the *upper* denaturing PAGE figure represent the ^{32}P-labeled DNA/RNA hybrid after RNase T1/A digestion. The *lower* TLC plate shows the m^{6}A modification fraction for each corresponding mixture. (*F*) The quantitative plot for *E* result. The sequence of the RNA template used is (A or m^{6}A at the underlined residue): 5′-UUAGACUCUGAACAACGCGCGCUAUAGAUAUAAAAAGUGG.

radioactive-labeling followed by ligation-assisted extraction and TLC (SCARLET) (Fig. 1A). The crucial new feature is the ligation-assisted extraction of the ^{32}P-labeled RNA of interest. SCARLET starts with the total RNA or total polyA$^+$ RNA sample. First, we choose a candidate site in a candidate RNA of interest. We apply RNase H cleavage guided by a complementary 2′-OMe/2′-H chimeric oligonucleotide to achieve site-specific cleavage 5′ to the candidate site (Yu et al. 1997; Zhao and Yu 2004). The cut site is radiolabeled with ^{32}P and the ^{32}P-labeled RNA fragment splint-ligated to a 116-nucleotide single-stranded DNA oligonucleotide using DNA ligase. The sample is then treated with RNases T1/A to completely digest all RNA, whereas the ^{32}P-labeled candidate site remains with the DNA oligonucleotide as DNA-^{32}P-(A/m^{6}A)p and DNA-^{32}P-(A/m^{6}A)Cp, which migrate as 117/118-mers on a denaturing gel. This labeled band is excised and eluted from the gel, digested with nuclease P1 into mononucleotides containing 5′ phosphate, and the m^{6}A modification status is determined by thin-layer chromatography. One crucial step of SCARLET is the splint-ligation, which attaches the candidate ribonucleotide to a DNA oligo and thus prevents its digestion by RNase T1/A.

We first tested this method using mixtures of two synthetic, 40-mer RNA oligonucleotides of the same sequence, one unmodified and the other containing one defined m^{6}A modification. To mimic the condition for biological site determination, this RNA oligo mixture at varying ratios was added to 1 μg HeLa total polyA$^+$ RNA before the start of SCARLET (Fig. 1B,C). The m^{6}A fraction determined by SCARLET accurately reflected the input m^{6}A fraction (Fig. 1D), indicating that SCARLET can quantitatively determine m^{6}A fraction from the total polyA$^+$ RNA pool. Furthermore, SCARLET was completely accurate when ~10 fmol of the RNA template was present and still sensitive when as little as ~1 fmol RNA template was present in 1 μg polyA$^+$ RNA (Fig. 1E,F).

To test the feasibility of SCARLET for biological sites, we first applied the method to determine the modification fraction of the previously reported m^{6}A sites in human rRNA in a total HeLa RNA sample (Piekna-Przybylska et al. 2008). The A1832 of human 18S rRNA is essentially fully modified at ~98% (Fig. 2A). We also determined the m^{6}A status at A4189 and A4190 of human 28S rRNA; previous methods could not make clear which one of these two sites is modified (Piekna-Przybylska et al. 2008). We found that A4190 is ~96% modified, whereas the A4189 site is minimally modified (Fig. 2B), indicating that SCARLET can easily resolve ambiguity of modification sites.

We also determined the m^{6}A fraction in an mRNA derived from a transfected plasmid in a human breast cancer cell line (Vilfan et al. 2013). This mRNA fuses the coding sequence of GFP with the 3′ UTR of bovine prolactin mRNA, which is known to have one strong and one cryptic m^{6}A modification site (Horowitz et al. 1984). We found that the modification fraction is 49% at the major site and 12% at the cryptic site in this breast cancer cell line (Fig. 2C). These results indicate that SCARLET can readily identify m^{6}A sites in a variety of cellular RNAs.

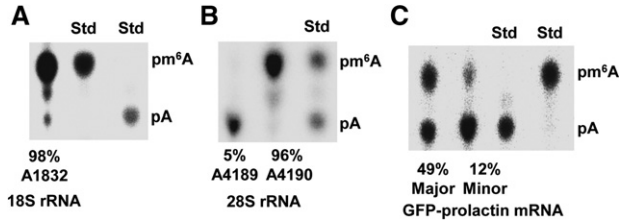

FIGURE 2. Determination of m^{6}A status of rRNA and an mRNA from a transfected plasmid. (*A*) SCARLET result determining the m^{6}A modification fraction at the A1832 site of human 18S rRNA. Standard ^{32}P-labeled m^{6}A and A nucleoside were run as controls. (*B*) SCARLET result determining the m^{6}A status at the A4189 and A4190 site of human 28S rRNA. (*C*) SCARLET result determining the m^{6}A status at two sites of the bovine prolactin mRNA segment derived from a transfected plasmid in the MDA-MB-231 cell line.

m^{6}A status in lncRNA and mRNA

We next applied SCARLET to determine the m^{6}A status in the nuclear-localized metastasis associated lung adenocarcinoma transcript 1 (MALAT1), a relatively abundant lncRNA of >6.5 kb and is known to regulate alternative splicing (Wilusz et al. 2008; Bernard et al. 2010; Tripathi et al. 2010). m^{6}A/MeRIP-seq of the polyA$^+$ RNA from three human cell lines identified three peaks of varying intensities, and seven m^{6}A-consensus sequences consisting of RRACH (Csepany et al. 1990; Harper et al. 1990) (R = A, G; H = A, C, U) are present in the largest peak (Fig. 3A). We found that m^{6}A is present at four of these seven sites (Fig. 3B; Table 1). Among the four modified sites, the modification fraction varies between 11% and 77% in HeLa and between 7% and 51% in HEK293T; the two upstream sites have higher m^{6}A fractions than the two downstream sites, although the upstream sites generally showed lower m^{6}A/MeRIP-seq signal. This result may be explained by the structural context of the two upstream m^{6}A sites (see below). We also determined the m^{6}A status at these four MALAT1 sites in a breast cancer line MDA-MB231 and human foreskin fibroblast HFF1. Among these lines, the modification fraction varied by up to approximately threefold, although the rank order of m^{6}A fraction at these four sites was similar (Table 1).

To validate that SCARLET accurately reports m^{6}A status of endogenous RNAs in the complex mixture of polyA$^+$ RNA, we affinity enriched MALAT1 RNA from the total polyA$^+$ RNA using MALAT1 RNA complementary biotinylated oligonucleotides and redetermined the m^{6}A fraction at three sites. Our purification enriched MALAT1 by >40-fold (Fig. 3C), and the m^{6}A fraction of the enriched MALAT1 was within 1.3-fold of those determined using total polyA$^+$ RNA (Fig. 3D), validating that SCARLET works well in the complex mixture of total polyA$^+$ RNA.

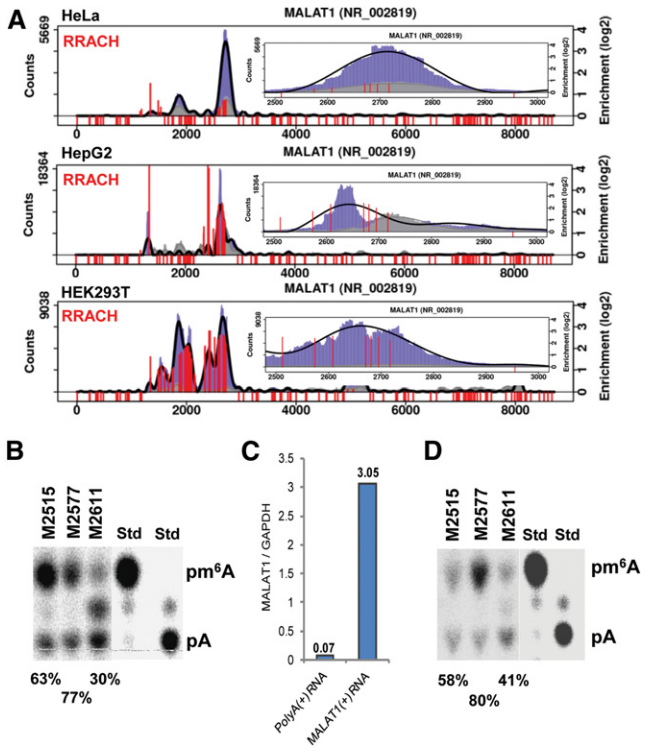

FIGURE 3. SCARLET determination of m^{6}A status in the MALAT1 lncRNA. (*A*) m^{6}A/MeRIP-seq data for the lncRNA MALAT1 in three cell lines. HeLa data are from our own experiment, HepG2 from Dominissini et al. (2012), and HEK293T from Meyer et al. (2012). The red lines correspond to all RRACH consensus motifs. The m^{6}A-antibody-IP data are in blue and the input sample is in gray. The RRACH consensus sites under the enriched peaks are shown as extended red lines. (*B*) SCARLET result for the three MALAT1 m^{6}A sites displaying the highest modification fraction, determined from total HeLa polyA$^+$ RNA. (*C*) Affinity enrichment of MALAT1 RNA using 3′-biotinylated DNA oligos complementary to MALAT1 residues 2449–2498. MALAT1 is ~43-fold enriched compared with GAPDH-mRNA, as demonstrated by real-time quantitative RT-PCR. (*D*) SCARLET result for the same three MALAT1 m^{6}A sites as in *B*, determined from MALAT1 enriched RNA.

We also determined the m^{6}A status in many RRACH sites in another lncRNA (TUG1) and three mRNAs (ACTB, TPT1, and BSG), chosen on the basis of the m^{6}A/MeRIP-seq results (Fig. 4). Among the sixteen sites tested, only four showed m^{6}A modification >5% (Table 2). This result shows that the previously reported 20% m^{6}A modification in bovine prolactin mRNA (Narayan and Rottman 1988) is not abnormal, suggests that many sites in mRNA carry incomplete m^{6}A modification, and highlights that not every RRACH consensus sequence sites under the m^{6}A/MeRIP-seq peaks are modified.

An m^{6}A-containing RNA structural motif

The precise mapping of m^{6}A sites enabled the determination of RNA secondary structural context in which m^{6}A occurs. According to the RNA secondary structure prediction, the two m^{6}A residues in MALAT1 lncRNA (M2515, M2577 site) displaying the highest modification fractions are located in two hairpin stems (Fig. 5A). We carried out phylogenic comparison for the MALAT1 RNA among primates (Fig. 5B,C). Phylogenetic covariations of two base pairs in the stem are present that support the hairpin structure of the M2515 site. M2577 site is only conserved among great apes and human and no covariation is present within this very limited data set.

Our structural mapping results in vitro of synthetic RNA oligonucleotides are consistent with the secondary structure prediction of these two hairpins (Fig. 6). The single-stranded selective nucleases T1 and S1 show strongest cuts in the predicted single-stranded regions, and the stacking/double-stranded selective nuclease V1 shows cuts almost exclusively in the predicted double-stranded regions. However, this hairpin structure seems to be quite dynamic as indicated by simultaneous signatures derived from T1/S1 and from V1 cuts. Interestingly, the m^{6}A residue in both sites is located adjacent to two consecutive noncanonical base pairs, and this arrangement likely results in the opening of the major groove to potentially facilitate the interaction of the m^{6}A residue with other cellular components such as m^{6}A binding proteins (Fig. 6; Dominissini et al. 2012). The presence of m^{6}A modification in the stem may increase the opening of the stem as indicated by the moderately increased cuts by nuclease S1 for M2515 and M2577 sites and by the moderately decreased cuts by nuclease V1 for the M2577 site (indicated by orange dots in Fig. 6). This result is consistent with a previous finding that m^{6}A modification in an RNA duplex moderately decreases duplex stability (Kierzek and Kierzek 2003).

The location of both m^{6}A sites in the stem of a hairpin loop, however, may hinder the access of the m^{6}A antibody. The structural context of these two m^{6}A sites is therefore consistent with the low-intensity m^{6}A/MeRIP-seq reads for these two sites.

TABLE 1. m^{6}A status in the seven RRACH consensus sites under the largest m^{6}A/MeRIP-seq peak of the MALAT1 lncRNA

		f(m^{6}A)[a]			
MALAT1 site	Consensus motif	HeLa[b]	HEK293T	MDA-MB231	HFF1
2515	**GGACU**	**0.61$^{\pm 0.03}$**	**0.41**	**0.39**	**0.67**
2577	**GGACU**	**0.80$^{\pm 0.03}$**	**0.51**	**0.58**	**0.88**
2611	**GGACU**	**0.38$^{\pm 0.07}$**	**0.13**	**0.20**	**0.49**
2674	AGACA	0.03	0.03	—[c]	—
2684	AGACA	0.02	0.02	—	—
2698	GAACC	0.02	0.03	—	—
2720	**GGACU**	**0.10$^{\pm 0.02}$**	**0.07**	**0.08**	**0.14**

[a]m^{6}A sites with >5% modification are shown in bold.
[b]f(m^{6}A) for HeLa have been tested three to six times, with the standard deviation of the average value shown.
[c]Not determined.

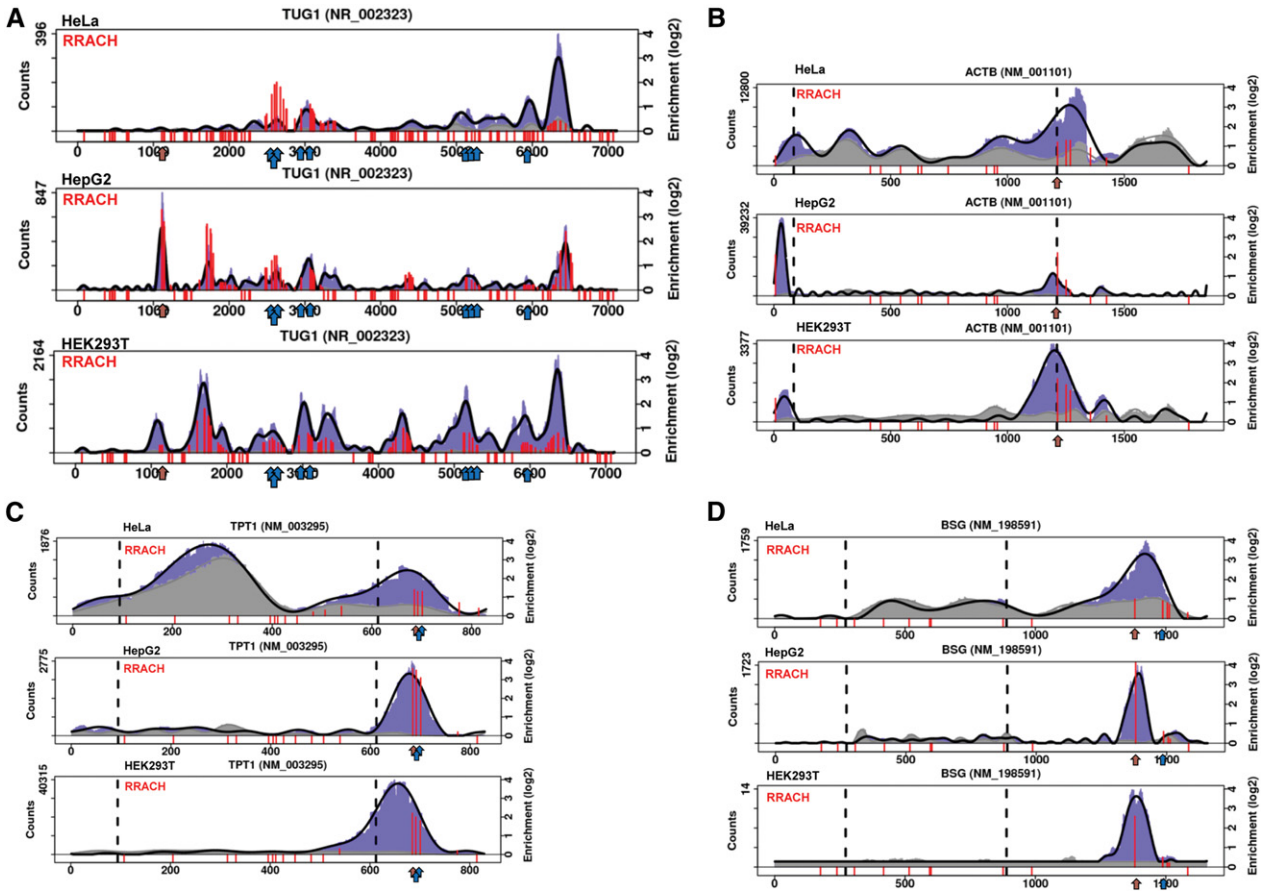

FIGURE 4. m^{6}A/MeRIP-seq analysis of four RNAs tested for m^{6}A status in this work. They include one lncRNA TUG1 (*A*), ACTB mRNA (*B*), TPT1 mRNA (*C*), and BSG mRNA (*D*) in three different human cell lines (HeLa, HepG2, and HEK293T). SCARLET tested sites are labeled with bold arrows. The sites tested positive for m^{6}A are indicated by purple arrows, whereas the sites tested negative for m^{6}A are indicated by blue arrows.

Concluding remarks

In conclusion, we report the first nucleotide-resolution method to determine m^{6}A location and modification fraction in mammalian mRNA/lncRNA. Starting from total polyA$^+$ RNA, SCARLET determined the m^{6}A status in several sites in two human lncRNAs and three human mRNAs, and we found that the m^{6}A fraction varies between 6% and 80% among these sites. We also found that many consensus RRACH sequences under the m^{6}A/MeRIP-seq peaks were not modified. One crucial advantage of SCARLET is its ability to directly determine the m^{6}A modification fraction at any site, an important consideration in elucidating the biological function of m^{6}A dynamics, which is currently not accessible using existing techniques.

In principle, SCARLET can also be used to investigate RNA modifications other than m^{6}A, such as 5-methyl-C (m^{5}C), pseudo-uridine (Ψ), 2′-*O*-methyl ribonucleosides (Nm), whose modification fractions or locations in mammalian mRNA/lncRNAs still remain obscure. SCARLET requires only common and readily available lab equipment and material and should be readily applicable to investigate the dynamics and biology of RNA modifications.

TABLE 2. m^{6}A status in the lncRNA TUG1 and three mRNAs (ACTB, TPT1, and BSG)

				f(m^{6}A)[a]	
ID	Name	Site	Motif	HeLa	MDA-MB-231
NR_002323	**TUG1**	**1114**	**GGACU**	**0.22**	**0.13**
NR_002323	TUG1	2497	GGACC	0.01	—[b]
NR_002323	TUG1	2564	GGACC	0.01	—
NR_002323	TUG1	2601	GAACA	0.01	—
NR_002323	TUG1	2953	AAACU	0	—
NR_002323	TUG1	3071	GGACU	0.04	—
NR_002323	TUG1	5125	AGACU	0.01	—
NR_002323	TUG1	5133	AAACA	0.01	—
NR_002323	TUG1	5138	GAACC	0.01	—
NR_002323	TUG1	5910	AGACU	0	—
NM_001101	**ACTB**	**1216**	**GGACU**	**0.21$^{\pm 0.09}$**	**0.28**
NM_003295	**TPT1**	**687**	**GGACU**	**0.15**	**0.19**
NM_003295	TPT1	694	AGACA	0.04	—
NM_003295	TPT1	703	GGACU	0.01	—
NM_198591	**BSG**	**1335**	**GGACU**	**0.06**	**0.08**
NM_198591	BSG	1442	GAACU	0.01	—

[a]m^{6}A sites with >5% modification are shown in bold.
[b]Not determined.

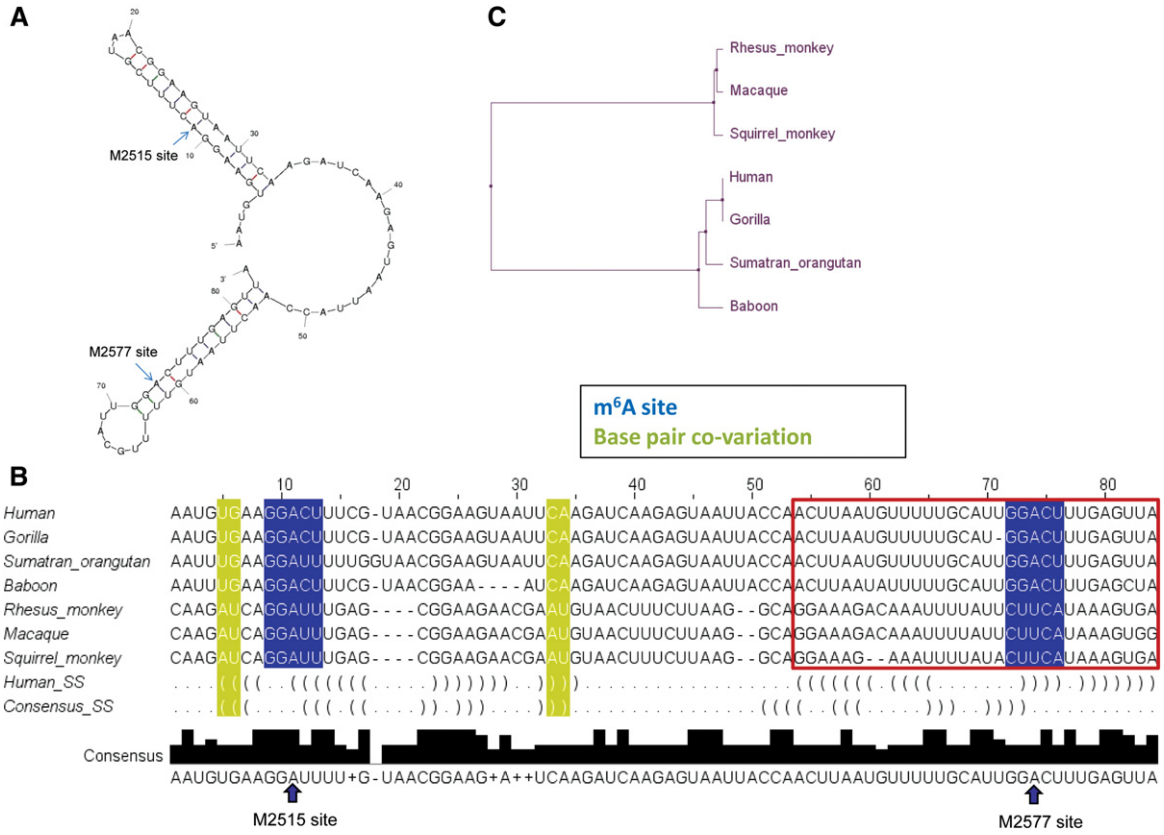

FIGURE 5. Prediction of secondary structure motifs around two MALAT1 m^{6}A residues displaying the highest modification fraction. (*A*) Mfold prediction of residues 2505–2587. The two m^{6}A modification sites are labeled with blue arrows and predicted to be located in two separate hairpin stems. (*B*) Sequence alignment of MALAT1 homologs among primates. The M2515 site is conserved and the hairpin structure supported by two base-pair covariations. The M2577 site is conserved only among human, gorilla, orangutan, and baboon; and no sequence change is present to provide information on structural conservation. (*C*) Phylogenetic tree of the MALAT1 homologs.

MATERIALS AND METHODS

Cell culture

Human cervical cancer cell line HeLa (CCL-2), embryonic kidney cell line HEK293T (CRL-11268), foreskin fibroblasts cell line HFF-1 (SCRC-1041), and breast cancer cell line MDA-MB-231 (HTB-26) were obtained from American Type Culture Collection (ATCC) and were cultured under standard conditions. Cells of ~70%–80% confluency were harvested for RNA extraction.

SCARLET oligonucleotide sequences

PB-A/m^{6}A Chimera: Nm = 2′-Ome-modified nucleotide. 5′-mGmCmGmUmUmGmUTCAGmAmGmUmCmUmAmA-3′

PB-A/m^{6}A Splint: 5′-CCACTTTTTATATCTATAGCGCGCGTTG TCTATTAACTCACAGGACCGGCGATGGCTG-3′

M2515 Chimera: 5′-mAmCmGmAmAmAmGmUCCTTmCmA mCmAmUmU-3′

M2515 Splint: 5′-TGATCTTGAATTACTTCCGTTACGAAAGTC TATTAACTCACAGGACCGGCGATGGCTG-3′

M2577 Chimera: 5′-mUmCmAmAmAmGmUCCAAmUmGmCm AmAmAmA-3′

M2577 Splint: 5′-GGATTTAAAAAATAATCTTAACTCAAAGTCT ATTAACTCACAGGACCGGCGATGGCTG-3′

M2611 Chimera: 5′-mUmGmCmUmAmGmUCCTCmAmGmGm AmUmUmU-3′

M2611 Splint: 5′-ACCTGGGTCAGCTGTCAATTAATGCTAGTCT ATTAACTCACAGGACCGGCGATGGCTG-3′

M2674 Chimera: 5′-mCmUmGmCmUmGmUCTTCmCmUmAm GmAmUmU-3′

M2674 Splint: 5′-CTGGTTCCTGGAATCCTGTCTGCTGCTGTC TATTAACTCACAGGACCGGCGATGGCTG-3′

M2684 Chimera: 5′-mAmUmCmCmUmGmUCTGCmUmGmCm UmGmUmC-3′

M2684 Splint: 5′-TCATCAAACACTGGTTCCTGGAATCCTGTC TATTAACTCACAGGACCGGCGATGGCTG-3′

M2698 Chimera: 5′-mCmAmCmUmGmGmUTCCTmGmGmAm AmUmCmC-3′

M2698 Splint: 5′-CCTCAGTCCTAGCTTCATCAAACACTGGTC TATTAACTCACAGGACCGGCGATGGCTG-3′

M2720 Chimera: 5′-mCmCmUmCmAmGmUCCTAmGmCmUm UmCmAmU-3′

M2720 Splint: 5′-AACTGCTGCTTGCTCGCTTGCTCCTCAGT CTATTAACTCACAGGACCGGCGATGGCTG-3′

ATCB1214 Chimera: 5′-mCmAmUmAmGmUCCGCmCmUmAm GmAmAmG-3′

ATCB1214 Splint: 5′-GAAAGGGTGTAACGCAACTAAGTCATA GTCTATTAACTCACAGGACCGGCGATGGCTG-3′

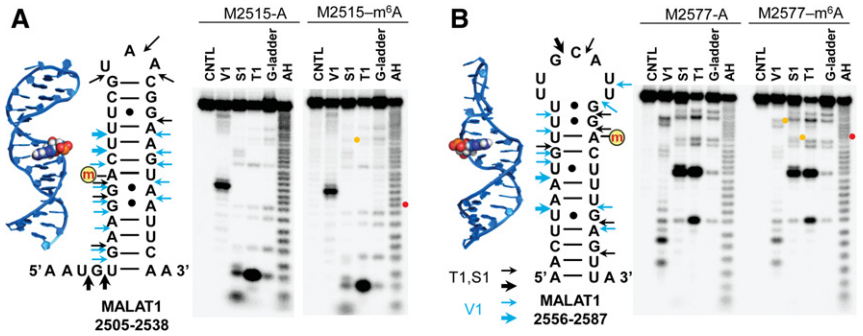

FIGURE 6. Structural mapping of two MALAT1 RNA segments containing a structural motif for m^{6}A modified sites. Structural mapping of the M2515-site hairpin (*A*) and of the M2577-site hairpin (*B*) with or without m^{6}A was performed using single-stranded selective nucleases T1 and S1 (black arrows) and stacking/double-stranded selective nuclear V1 (blue arrows). The position of RNA cuts were identified through the T1 ladder (G-ladder) and alkaline hydrolysis (AH). The m^{6}A site is indicated with a red dot, and the difference between *A* and m^{6}A containing oligos is indicated with orange dots. The three-dimensional model was derived from MC-fold/MC-sym (Parisien and Major 2008).

TPT1 687 Chimera: 5′-mCmUmUmAmAmGmUCCTGmGmUmGmUmUmGmUmG-3′

TPT1 687 Splint: 5′-AGATGACATCAGTCCCATTTGTCTTAAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

TPT1 694 Chimera: 5′-mAmUmUmUmGmUCTTAmAmGmUmCmCmUmGmGmU-3′

TPT1 694 Splint: 5′-GAGCTCAAGATGACATCAGTCCCATTTGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

TPT1 703 Chimera: 5′-mAmCmAmUmCmAmGmUCCCAmUmUmUmGmUmCmU-3′

TPT1 703 Splint: 5′-ATAAATGAAGAGCTCAAGATGACATCAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

18S rRNA Chimera: 5′-mAmCmCmUmUmGmUTACGmAmCmUmUmUmUmA-3′

18S rRNA Splint: 5′-TTCCGCAGGTTcACCTACGGAAACCTTGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

GFP-Prolactin Major Chimera: 5′-mAmGmUmCmUmGmUTTTTmAmUmUmUmAmAmG-3′

GFP-Prolactin Major Splint: 5′-GCAGATTTTGACATCGCTACAGAGTCTGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

GFP-Prolactin Minor Chimera: 5′-mAmCmAmGmAmGmUCTGTmUmUmUmUmAmUmU-3′

GFP-Prolactin Minor Splint: 5′-GCGCGCAGATTTTGACATCGCTACAGAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-1114 Chimera: 5′-mUmAmAmAmAmGmUCCACmGmUmGmCmAmCmC-3′

T-1114 Splint: 5′-CTCTTCCAGTGAGCCCGCTTGCTAAAAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-2497 Chimera: 5′-mCmCmCmUmGmGmUCCAAmAmUmUmUmAmUmUmGmU-3′

T-2497 Splint: 5′-GAGCATTAATAACTAAAAAATCCCCTGGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-2564 Chimera: 5′-mGmUCCAAmUmGmCmUmGmCmUmCmU-3′

T-2564 Splint: 5′-TTCAATCATTTGAGATTGTGGGGTGTGGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-2601 Chimera: 5′-mUmCmCmUmAmGmAmUmGmUTCATmGmAmAmUmUmUmC-3′

T-2601 Splint: 5′-TCCAGTGACCTTCACGGGATCCTAGATGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-2953 Chimera: 5′-mGmCmAmCmAmGmUTTTCmAmCmUmCmAmAmA-3′

T-2953 Splint: 5′-TTCACTTTTACTCTGGGCTTCTGCACAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-3071 Chimera: 5′-mAmGmGmAmAmGmUCCCCmUmUmAmGmGmUmG-3′

T-3071 Splint: 5′-CGGAAGAGTTCAAGGTGTTGTAAGGAAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-5125 Chimera: 5′-mGmAmGmUCTATmGmCmUmAmCmAmUmCmU-3′

T-5125 Splint: 5′-CAATCAGACTTGAGGTTCTGTTTAGGAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-5133 Chimera: 5′-mGmGmUmUmCmUmGmUTTAGmGmAmGmUmCmUmA-3′

T-5133 Splint: 5′-CTTATCCTCAATCAGACTTGAGGTTCTGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-5138 Chimera: 5′-mUmUmGmAmGmGmUTCTGmUmUmUmAmGmGmA-3′

T-5138 Splint: 5′-AAGGCCTTATCCTCAATCAGACTTGAGGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

T-5910 Chimera: 5′-mGmUCTGAmGmAmGmCmCmCmAmA-3′

T-5910 Splint: 5′-CAACACAAGCAAACTCAGTGGTCATGAGTCTATTAACTCACAGGACCGGCGATGGCTG-3′

ssDNA-116: 5′-GGAGAGACAACTTAAAGAGACTTAAAAGATTAATTTAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGCCGGTCCTGTGAGTTAATAG-3′

Synthesis of m^{6}A-RNA oligo

m^{6}A-containing RNA oligonucleotides were synthesized according to a method reported previously (Dai et al. 2007).

SCARLET (site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and TLC)

Total RNA was isolated using PerfectPure RNA cultured cell kit (5 Prime). Polyadenylated RNA (polyA$^+$ RNA) was enriched via the GenElute mRNA miniprep kit (Sigma-Aldrich). One microgram polyA$^+$ RNA was annealed with 3 pmol corresponding Chimeric oligo in 3 µL 30 mM TrisHCl, pH 7.5 by heating for 1 min at 95°C. This sample was incubated with 1 µL 6× RNase H mix (2× T4 polynucleotide kinase buffer [T4 PNK, USB], 1 unit/µL RNase H [Epicentre #R0601K]) and 1 unit thermosensitive alkaline phosphatase (TAP, Thermo Scientific) for 1 h at 44°C. RNase H and TAP were inactivated by heating for 5 min at 75°C. The RNA fragment was radioactively labeled by adding 1 µL 6× T4 PNK mix (1× T4 PNK buffer, 6 units/µL T4 PNK, 28 µCi/µL [γ-^{32}P] ATP) and incubation for 1 h at 37°C. T4 PNK was inactivated for 5 min at 75°C. The preceding mix was annealed with 4 pmol corresponding Splint and 5 pmol 116-mer DNA oligo (ssDNA-116) by heating for 3 min at

75°C. Ligation reaction was performed by incubating with 2.5 µL 4× ligation mix (1.4× T4 PNK buffer, 0.27 mM ATP, 57% DMSO, 1.9 units/µL T4 DNA ligase) for 3.5 h at 37°C, and stopped by mixing with equal volume of 2× RNA loading buffer (9 M urea, 100 mM EDTA). RNA was digested by incubating with 1 µL RNase T1/A mixture (160 units/µL RNase T1, 0.16 mg/mL RNase A) at 37°C overnight. The ligation product was isolated by denaturing PAGE followed by crush and soak of corresponding gel slices. The RNA pellet was digested into individual nucleotides by incubation with 3 µL nuclease P1 mix (0.33 unit/µL nuclease P1 in 30 mM sodium acetate/acetic acid, pH 4.8) for 2 h at 37°C. The reaction mix was spotted on cellulose TLC plate (20 × 20 cm; Merck), and TLC was run with isopropanol:HCl:water (70:15:15, v/v/v). After that, the TLC plate was dried for 1 h at room temperature, wrapped in plastic film, and exposed to a blanked phosphorimager screen. The result can be visualized on a phosphorimager.

RNA structural mapping

The synthetic MALAT1 RNA oligos were 5′-labeled with [γ-^{32}P] ATP by T4 PNK and gel purified. Mix ^{32}P-labeled RNA, 0.25 µg/µL *E. coli* tRNA, 62.5 mM Tris-HCl and RNase free water in a total volume of 36.8 µL. Heat for 2 min at 90°C, then cool down at room temperature for 3 min. Add 4.6 µL of 24 mM $MgCl_2$ and 4.6 µL of 1 M KCl, then place at room temperature for 5 min. Split into 4 separate aliquots, and add 1 µL freshly made nuclease stock (nuclease V1, T1, S1, or RNase free water as control = CNTL) to each aliquot. Incubate for 15 min at room temperature. Add 5 µL RNA loading buffer, and immediately place on ice until loading on a 20% urea-denaturing gel. For the alkaline hydrolysis group, mix end-labeled RNA, 1 µL 1 µg/µL *E. coli* tRNA, 1 µL 5× BH buffer, and RNase free water in a total volume of 5 µL. Heat for 1 min at 95°C. Quick spin down to the bottom. Add 5 µL of urea loading buffer. Immediately place on ice until loading on the gel. For the G-ladder group, mix end-labeled RNA, 1 µL 1 µg/µL *E. coli* tRNA, 1 µL Nuclease T1 (1 U/µL), and RNase free water in 5 µL volume. Heat the sample for 1 min at 65°C and then spin down to bottom. Add 5 µL of urea loading buffer. Immediately put on ice until loading on the gel.

m^{6}A/MeRIP-seq and its data analysis

The m^{6}A/MeRIP-seq for HeLa RNA was performed as previously described (Dominissini et al. 2013).

For m^{6}A/MeRIP-seq data analysis, we used the deep sequencing data from GEO sets GSE37005 (HepG2) and GSE29714 (HEK293T). The deep sequencing data were mapped to a *Homo sapiens* genome provided in a GenBank flat file (retrieved April 1, 2013, ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/mRNA_Prot/human.rna.gbff.gz). The mapping program is an in-house mapper allowing for at most one mismatch for an alignment of at least 22 nucleotides. When a read aligns to many locations, the longest alignment is retained. For enrichment computations, both the control and the IPed data were mapped and compared for each gene (coding RNAs with IDs of the form NM_xx and noncoding RNAs of the form NR_xx). The raw data aligned on a gene make for a "mountain" plot. Because the plot is edgy, we smoothed out the plot using low-frequency fast Fourier transforms (FFT), yielding a smoothed mountain plot. From the smoothed mountain plots for both the control and IPed data, we compute the enrichment values for any given site on a gene (e.g., an RRACH position): We first find the peaks of the IPed data on the smoothed mountain plot. At these peak positions, their enrichment values, *Ep*, is the log ratio of the IPed data (*ip*) to the control (*ck*), and to the height of the peak, *Hp*, to the average height across the whole gene, *Hg*: $Ep = \log[(Hp/Hg)\_ip/(Hp/Hg)\_ck]$. The enrichment value at a given gene location, *El*, separate by a distance d to a peak is thus: *El* = sum over all peaks p of $Ep^* \exp(-d)$. On the mountain plots (Figs. 3A, 4), the raw IPed data are shown in blue, the control data in gray, and the FFT-smoothed curves in black. Red vertical bars indicate the RRACH sites with the height of the bar indicating the enrichment value at that genomic location.

ACKNOWLEDGMENTS

This work was supported by a NIH grant (GM088599 to T.P. and C.H.). We thank George Perdrizet for performing the phylogenetic analysis of the MALAT1 sites.

Received July 8, 2013; accepted August 8, 2013.

REFERENCES

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F, et al. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* **29:** 3082–3093.

Bokar JA. 2005. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. In *Fine-tuning of RNA functions by modification and editing* (ed. Grosjean H), pp. 141–178. Springer-Verlag, Berlin.

Csepany T, Lin A, Baldick CJ Jr, Beemon K. 1990. Sequence specificity of mRNA N^6-adenosine methyltransferase. *J Biol Chem* **265:** 20117–20122.

Dai Q, Fong R, Saikia M, Stephenson D, Yu YT, Pan T, Piccirilli JA. 2007. Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N^6-methyladenosine. *Nucleic Acids Res* **35:** 6322–6329.

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al. 2012. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485:** 201–206.

Dominissini D, Moshitch-Moshkovitz S, Salmon-Divon M, Amariglio N, Rechavi G. 2013. Transcriptome-wide mapping of N^6-methyladenosine by m^{6}A-seq based on immunocapturing and massively parallel sequencing. *Nat Protoc* **8:** 176–189.

Grosjean H. 2005. *Fine-tuning of RNA functions by modification and editing*. Springer-Verlag, Berlin.

Harper JE, Miceli SM, Roberts RJ, Manley JL. 1990. Sequence specificity of the human mRNA N6-adenosine methylase *in vitro*. *Nucleic Acids Res* **18:** 5735–5741.

Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Mapping of N^6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci* **81:** 5667–5671.

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, et al. 2011. *N*6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7:** 885–887.

Karikó K, Buckstein M, Ni H, Weissman D. 2005. Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23:** 165–175.

Kierzek E, Kierzek R. 2003. The thermodynamic stability of RNA duplexes and hairpins containing N^6-alkyladenosines and 2-methylthio-N^6-alkyladenosines. *Nucleic Acids Res* **31:** 4472–4480.

Ma X, Yang C, Alexandrov A, Grayhack EJ, Behm-Ansmant I, Yu YT. 2005. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J* **24:** 2403–2413.

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149:** 1635–1646.

Narayan P, Rottman FM. 1988. An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* **242:** 1159–1162.

Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452:** 51–55.

Piekna-Przybylska D, Decatur WA, Fournier MJ. 2008. The 3D rRNA modification maps database: With interactive tools for ribosome analysis. *Nucleic Acids Res* **36:** D178–D183.

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, et al. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39:** 925–938.

Vilfan ID, Tsai YC, Clark TA, Wegener J, Dai Q, Yi C, Pan T, Turner SW, Korlach J. 2013. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnology* **11:** 8.

Wilusz JE, Freier SM, Spector DL. 2008. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135:** 919–932.

Wu G, Xiao M, Yang C, Yu YT. 2011. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J* **30:** 79–89.

Yu YT, Shu MD, Steitz JA. 1997. A new method for detecting sites of 2′-O-methylation in RNA molecules. *RNA* **3:** 324–331.

Zhao X, Yu YT. 2004. Detection and quantitation of RNA base modifications. *RNA* **10:** 996–1002.

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, et al. 2012. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49:** 18–29.