Probing N\textsuperscript{6}-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA

NIAN LIU,\textsuperscript{1} MARC PARISIEN,\textsuperscript{2} QING DAI,\textsuperscript{1} GUANQUN ZHENG,\textsuperscript{1} CHUAN HE,\textsuperscript{1,3} and TAO PAN\textsuperscript{2,3,4}
\textsuperscript{1}Department of Chemistry, \textsuperscript{2}Department of Biochemistry and Molecular Biology, \textsuperscript{3}Institute of Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, USA

ABSTRACT

N\textsuperscript{6}-methyladenosine (m\textsuperscript{6}A) is the most abundant modification in mammalian mRNA and long noncoding RNA (lncRNA). Recent discoveries of two m\textsuperscript{6}A demethylases and cell-type and cell-state-dependent m\textsuperscript{6}A patterns indicate that m\textsuperscript{6}A modifications are highly dynamic and likely play important biological roles for RNA akin to DNA methylation or histone modification. Proposed functions for m\textsuperscript{6}A modification include mRNA splicing, export, stability, and immune tolerance; but m\textsuperscript{6}A 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\textsuperscript{6}A 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\textsuperscript{6}A residue and its modification fraction, which are crucial parameters in probing the cellular dynamics of m\textsuperscript{6}A modification. We applied the method to determine the m\textsuperscript{6}A status at several sites in two human lncRNAs and three human mRNAs and found that m\textsuperscript{6}A fraction varies between 6\% and 80\% among these sites. We also found that many m\textsuperscript{6}A candidate sites in these RNAs are however not modified. The precise determination of m\textsuperscript{6}A status in a long noncoding RNA also enables the identification of an m\textsuperscript{6}A-containing RNA structural motif.

Keywords: RNA modification; N\textsuperscript{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\textsuperscript{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\textsuperscript{6}A modification pattern is highly dependent on cell types and conditions (Dominissini et al. 2012; Meyer et al. 2012). Two recently discovered m\textsuperscript{6}A-demethylases are associated with obesity/diabetes and sperm development, indicating that the dynamics of m\textsuperscript{6}A modification is biologically important (Jia et al. 2011; Zheng et al. 2012). Proposed m\textsuperscript{6}A functions include mRNA splicing, export, stability, and immune tolerance (Bokar 2005; Karikó et al. 2005), but m\textsuperscript{6}A studies in mRNA/lncRNA were hindered by the lack of methods for its precise identification.

Although the m\textsuperscript{6}A 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\textsuperscript{6}A status for each modification site. Challenges for the determination of m\textsuperscript{6}A 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\textsuperscript{6}A cannot be identified by the traditional deep sequencing methods because reverse transcriptase is insensitive to its presence in cDNA synthesis. The recently developed m\textsuperscript{6}A/MeRIP-seq method combines m\textsuperscript{6}A antibody immune-precipitation and deep sequencing to locate m\textsuperscript{6}A residues in approximately 200 nucleotide RNA segments (Dominissini et al. 2012; Meyer et al. 2012). These results reveal that m\textsuperscript{6}A modifications are enriched in the 3′ UTR near the stop codons, and the m\textsuperscript{6}A modification patterns change in different cell types or when cells are stressed.
However, the m6A/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 m6A status could be crucial for its function, new methods that can unambiguously determine the m6A 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 m6A status in any candidate site in mRNA/lncRNA at single nucleotide resolution.

RESULTS AND DISCUSSION

A method to detect m6A status in mRNA/lncRNA

Our method combines site-specific cleavage of each m6A-containing candidate site followed by radiolabeling and site-specific ligation and complete nuclease digestion to enable isolation of the candidate m6A containing residue in the mixture of total RNA or polyA+ RNA (Fig. 1A). The candidate m6A containing residue is then analyzed by thin-layer chromatography (TLC) to reveal the presence or absence of m6A 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 m6A 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
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 $^{3'}$-OME/2-$^{3'}$-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 biological condition for site determination, this RNA oligo mixture at varying ratios was added to 1 $\mu$g HeLa total polyA\(^+\) RNA before the start of SCARLET (Fig. 1BC). 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 $\mu$g polyA\(^+\) RNA (Fig. 1EF).

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 $\sim$98% (Fig. 2A). We also determined the m$^{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 $\sim$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.

### 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 2. Determination of m$^{6}$A status of rRNA and an mRNA from a transfected plasmid.](image-url)
We also determined the m\textsuperscript{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\textsuperscript{6}A/MeRIP-seq results (Fig. 4). Among the sixteen sites tested, only four showed m\textsuperscript{6}A modification >5% (Table 2). This result shows that the previously reported 20% m\textsuperscript{6}A modification in bovine prolactin mRNA (Narayan and Rottman 1988) is not abnormal, suggests that many sites in mRNA carry incomplete m\textsuperscript{6}A modification, and highlights that not every RRACH consen-
sus sequence sites under the m\textsuperscript{6}A/MeRIP-seq peaks are modified.

An m\textsuperscript{6}A-containing RNA structural motif

The precise mapping of m\textsuperscript{6}A sites enabled the determination of RNA secondary structural context in which m\textsuperscript{6}A occurs. According to the RNA secondary structure prediction, the two m\textsuperscript{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\textsuperscript{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\textsuperscript{6}A residue with other cellular components such as m\textsuperscript{6}A binding proteins (Fig. 6; Dominissini et al. 2012). The presence of m\textsuperscript{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\textsuperscript{6}A modification in an RNA duplex moderately decreases duplex stability (Kierzek and Kierzek 2003).

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

| MALAT1 site | Consensus motif | HeLa\textsuperscript{a} | HEK293T | MDA-MB231 | FF1 |
|-------------|----------------|----------------------|----------|----------|-----|
| 2515        | GGACU          | 0.61 ± 0.03          | 0.41     | 0.39     | 0.67 |
| 2577        | GGACU          | 0.80 ± 0.03          | 0.51     | 0.58     | 0.88 |
| 2611        | GGACU          | 0.38 ± 0.07          | 0.13     | 0.20     | 0.49 |
| 2674        | AGACA          | 0.03                 | 0.03     | —        | —    |
| 2684        | AGACA          | 0.02                 | 0.02     | —        | —    |
| 2698        | GAACC          | 0.02                 | 0.03     | —        | —    |
| 2720        | GGACU          | 0.10 ± 0.02          | 0.07     | 0.08     | 0.14 |

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

In conclusion, we report the first nucleotide-resolution method to determine m6A location and modification fraction in mammalian mRNA/lncRNA. Starting from total polyA+ RNA, SCARLET determined the m6A status in several sites in two human lncRNAs and three human mRNAs, and we found that the m6A fraction varies between 6% and 80% among these sites. We also found that many consensus RRACH sequences under the m6A/MeRIP-seq peaks were not modified. One crucial advantage of SCARLET is its ability to directly determine the m6A modification fraction at any site, an important consideration in elucidating the biological function of m6A dynamics, which is currently not accessible using existing techniques.

In principle, SCARLET can also be used to investigate RNA modifications other than m6A, such as 5-methyl-C (m5C), 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. m6A status in the lncRNA TUG1 and three mRNAs (ACTB, TPT1, and BSG)

| ID        | Name   | Site   | Motif | f(m^6A)a | HeLa   | MDA-MB-231 |
|-----------|--------|--------|-------|----------|--------|------------|
| NR_002323 | TUG1   | 1114   | GGACU | 0.22     | 0.13   |
| NR_002323 | TUG1   | 2497   | GGACC | 0.01     | —      |
| NR_002323 | TUG1   | 2564   | GGACC | 0.01     | —      |
| NR_002323 | TUG1   | 2601   | GAACA | 0.01     | —      |
| NR_002323 | TUG1   | 2953   | AAACU | —        | —      |
| 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 | 0.15     | 0.28   |
| NM_001101 | ACTB   | 1216   | GGACU | 0.15     | 0.28   |
| NM_003295 | TPT1   | 687    | GGACU | 0.04     | —      |
| NM_003295 | TPT1   | 694    | AGACA | 0.01     | —      |
| NM_198591 | BSG    | 1335   | GGACU | 0.06     | 0.08   |
| NM_198591 | BSG    | 1442   | GGACU | 0.01     | —      |

m6A sites with >5% modification are shown in bold.

Not determined.
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’-mGmCmGmUmUmGmUCCCTCaAGGAGGCGATTGCTG-3’

PB-A/m^6^A Split: 5’-CCACCTTTTTTATATCTATAGCGCGTTGGTCTATTACTCACAGGACCGCGATGGCTG-3’

M2515 Chimera: 5’-mAmCmGmAmAmGmUCCCTCCmAmGmAmUmUmCmAmAmUmU-3’

M2515 Split: 5’-TGATCTTGAATTACTTCCGTTACGAAAGTCCTATTAACTCACAGGACCGGCGATGGCTG-3’

M2577 Chimera: 5’-mUmCmAmAmAmGmUCCCTCCmAmGmAmAmUmC-3’

M2577 Split: 5’-TGATCTTGAATTACTTCCGTTACGAAAGTCCTATTAACTCACAGGACCGGCGATGGCTG-3’

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 co-variations. 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.
FIGURE 6. Structural mapping of two MALAT1 RNA segments containing a structural motif for m^A modified sites. Structural mapping of the M2515-site hairpin (A) and of the M2577-site hairpin (B) with or without m^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^A site is indicated with a red dot, and the difference between A and m^A containing oligos is indicated with orange dots. The three-dimensional model was derived from MC-fold/MC-sym (Parisien and Major 2008).

**Synthesis of m^6-A-RNA oligo**
m^A-containing RNA oligonucleotides were synthesized according to a method reported previously (Dai et al. 2007).

**SCARLET (site-specific cleavage and 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 m^6-A-containing RNA oligonucleotides were synthesized according to a method reported previously (Dai et al. 2007).
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-<sup>32</sup>P] ATP by T4 PNK and gel purified. Mix 32P-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 MgCl₂ 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<sup>6</sup>A/MeRIP-seq and its data analysis

The m<sup>6</sup>A/MeRIP-seq for HeLa RNA was performed as previously described (Dominissini et al. 2013).

For m<sup>6</sup>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 (Ep) to the control (Ek), and to the height of the peak, Hp, to the average height across the whole gene, Hg. Ep = log((Hp/Hg)-/Ep/(Hp/Hg)-Ek). 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 *(d-p). 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.

REFERENCES

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedef F, Jour dend L, Coulipier F, et al. 2010. A long nuclear-retained non-coding RNA regulates syntagogenesis 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.

Csapany T, Lin A, Baldick CJ Jr, Beemon K. 1990. Sequence specificity of mRNA N<sup>6</sup>-adenosine methylation transferase. J Biol Chem 265: 20117–20122.

Dai Q, Fong R, Saikia M, Stephenson D, Yu YT, Pan T, Piccirilli JA. 2005. The biosynthesis and functional roles of methylated nucleosides by m6A-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 N<sub>6</sub>-adenosine methylase in vitro. Nucleic Acids Res 18: 5735–5741.

Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Fine-tuning of RNA functions by modification and quantitation of pseudouridine and N<sub>6</sub>-methyladenosine. Nucleic Acids Res 35: 6322–6329.

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Unger L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amargiolo N, Kupiec M, et al. 2012. Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. Nature 485: 201–206.

Dominissini D, Moshitch-Moshkovitz S, Salmon-Divon M, Amargiolo N, Rechavi G. 2013. Transcriptome-wide mapping of N<sup>6</sup>-methyladenosine by m<sup>6</sup>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.

Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Fine-tuning of RNA functions by modification and quantitation of pseudouridine and N<sub>6</sub>-methyladenosine. Nucleic Acids Res 18: 5735–5741.

Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Fine-tuning of RNA functions by modification and quantitation of pseudouridine and N<sub>6</sub>-methyladenosine. Nucleic Acids Res 18: 5735–5741.

Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Fine-tuning of RNA functions by modification and quantitation of pseudouridine and N<sub>6</sub>-methyladenosine. Nucleic Acids Res 18: 5735–5741.

Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Fine-tuning of RNA functions by modification and quantitation of pseudouridine and N<sub>6</sub>-methyladenosine. Nucleic Acids Res 18: 5735–5741.
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.