Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2017 November 28.

Published in final edited form as:

Nat Methods. 2016 August ; 13(8): 692–698. doi:10.1038/nmeth.3898.

m^{6}A level and isoform characterization sequencing (m^{6}A-LAIC-seq) reveals the census and complexity of the m^{6}A epitranscriptome

Benoit Molinie[1,*], **Jinkai Wang**[4,*], **Kok-Seong Lim**[5], **Roman Hillebrand**[5], **Zhi-xiang Lu**[4], **Nicholas Van Wittenberghe**[1], **Benjamin D. Howard**[1], **Kaveh Daneshvar**[1], **Alan C. Mullen**[1,2], **Peter Dedon**[5], **Yi Xing**[4], and **Cosmas C. Giallourakis**[1,2,3]

[1]Gastrointestinal Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA

[2]Harvard Stem Cell Institute, Cambridge, MA

[3]Center for Study of Inflammatory Bowel Disease, MGH, Boston MA

[4]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles. CHS 33-260, 650 Charles E. Young Drive South, Los Angeles, CA

[5]Massachusetts Institute of Technology Department of Biological Engineering, 56-787b 77 Massachusetts Ave. Cambridge, MA

Abstract

N^6-Methyladenosine (m^{6}A) is a widespread, reversible chemical modification of RNA molecules, implicated in many aspects of RNA metabolism. Little quantitative information exists as to either how many transcript copies of particular genes are m^{6}A modified ('m^{6}A levels') or the relationship of m^{6}A modification(s) to alternative RNA isoforms. To deconvolute the m^{6}A epitranscriptome, we developed m^{6}A-level and isoform-characterization sequencing (m^{6}A-LAIC-seq). We found that cells exhibit a broad range of nonstoichiometric m^{6}A levels with cell-type specificity. At the level of isoform characterization, we discovered widespread differences in the use of tandem alternative polyadenylation (APA) sites by methylated and nonmethylated transcript isoforms of individual genes. Strikingly, there is a strong bias for methylated transcripts to be coupled with proximal APA sites, resulting in shortened 3′ untranslated regions, while

Correspondence to: Yi Xing; Cosmas C. Giallourakis.

*These authors contributed equally.

Yi Xing, PhD, Department of Microbiology, Immunology, and Molecular Genetics, UCLA, CHS 33-260, 650 Charles E. Young Drive South, Los Angeles, CA, 90095, 310-825-6806

Cosmas Giallourakis, MD, Harvard Medical School, Massachusetts General Hospital, Gastrointestinal Unit, Jackson Building Rm. J828, Boston, MA 02114, 617-726-2126

Contributions

C.C.G., Y.X., B.M., and J.W. conceived of the project, analyzed the data, and wrote the manuscript with input from all authors. B.M. did the experimental work related to m^{6}A-LAIC-seq along with Z.L., N.V.W., B.D.H., K.D., and A.C.M. Y.X. and J.W. did the computational analyses of data with input from C.C.G. and B.M. P.D., K.S.L., and R.H. synthesized m^{6}Am and performed RNA mass spectrometry.

Accession codes.

m^{6}A-LAIC-seq raw data of H1-ESC and GM12878 have been deposited in the Gene Expression Omnibus database under the accession number GSE66086.

Competing financial interests

C.C.G., Y.X., B.M., and J.W. are in the process of filing a patent application for m^{6}A-LAIC-seq.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

nonmethylated transcript isoforms tend to use distal APA sites. m^{6}A-LAIC-seq yields a new perspective on transcriptome complexity and links APA usage to m^{6}A modifications.

Introduction

m^{6}A is the most abundant known internal chemical modification of mRNAs and long noncoding RNAs (lncRNAs)[1, 2, 3, 4, 5, 6]. m^{6}A has been implicated in all forms of RNA metabolism[2, 3, 4, 7, 8, 9, 10, 11]. The 'writing' of m^{6}A RNA modification is accomplished via the m^{6}A methyltransferases METTL3 and METTL14 (refs. 7,12, 13,14,15). Furthermore, m^{6}A modification is reversible or 'erased' by the FTO and ALKBH5 m^{6}A demethylases and 'read' by the m^{6}A-binding family of YTH-domain RNA-binding proteins (RBPs), which promote degradation and translational regulation of m^{6}A-modified transcripts[8, 9, 10, 11, 16, 17]. Recent findings that METTL3 is required for efficient embryonic-stem-cell differentiation highlight the functional importance of m^{6}A modification[18, 19].

Although transcriptome-wide m^{6}A location analysis by m^{6}A-seq (or MeRIP-seq) has led to significant insights, methods to quantify the ratio of methylated to nonmethylated transcripts (defined here as m^{6}A level) on a transcriptome-wide scale have been lacking[3, 15, 20, 21]. The proportion of transcripts methylated per gene has been measured directly for only a handful of genes, and in those instances m^{6}A levels were found to be nonstoichiometric (i.e., fewer than 100% of transcript copies were methylated)[22, 23]. Genome-wide quantitative measurements of m^{6}A levels are critical to understanding the proportion of m^{6}A-modified versus unmodified transcripts for each gene to assess the magnitude of potential m^{6}A regulatory impact. Here, we report a new method, m^{6}A-LAIC-seq, which quantitatively deconvolutes methylated versus nonmethylated transcripts. Unlike m^{6}A-seq, our protocol does not fragment the RNA before anti-m^{6}A RNA immunoprecipitation (RIP); instead relying on sequencing intact full-length transcripts in both m^{6}A-positive and m^{6}A-negative fractions post-RIP. While allowing quantification of m^{6}A levels, the isolation and sequencing of intact full-length RNAs also allowed us to examine differential isoform usage in methylated and nonmethylated transcripts of each gene.

Results

In developing m^{6}A-LAIC-seq (Fig. 1a), we conducted initial experiments to establish the saturation curve, sensitivity, and specificity for anti-m^{6}A RIP of full-length transcripts. The anti-m^{6}A antibody became saturated at 500 ng of RNA per 1 μg of antibody (500 ng μg^{-1}) under our anti-m^{6}A RIP conditions, which used *in vitro*–synthesized *GAPDH* transcripts containing 100% m^{6}A-modified adenosines (Fig. 1b). We started m^{6}A-LAIC-seq with 2 μg of 2× polyadenylated (poly(A)$^+$) RNA, using an anti-m^{6}A antibody at 500 ng μg^{-1}. We employed a vast excess of antibody, as a second round of anti-m^{6}A RIP on the m^{6}A-negative fraction recovered no further RNA (Supplementary Fig. 1a–c). We also found that anti-m^{6}A RIP efficiency was independent of the number of modifications per transcript (Supplementary Fig. 1d,e). To examine the sensitivity of the anti-m^{6}A antibody *in vitro*, we performed anti-m^{6}A RIP experiments using *in vitro*–synthesized *GAPDH* with, on average, two random m^{6}A modifications per transcript as well as unmodified *GAPDH* transcripts,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

both of which were radiolabeled with CTP32. We then mixed the unmodified and m^{6}A-modified *GAPDH* transcripts at varying percentages. We detected as few as one m^{6}A-modified transcript mixed with 1,000 unmodified transcripts, which suggested a very high sensitivity of the anti-m^{6}A RIP (Fig. 1c).

Next, we added various spike-in controls to our starting input poly(A)$^+$ RNA to evaluate the specificity, sensitivity, and quantitation of the actual m^{6}A-LAIC-seq protocol (Fig. 1a). As intended, our protocol sequenced intact full-length RNAs with no local enrichment of m^{6}A sites (Fig. 1d). Furthermore, no reads were detected of the three nonmethylated RNA transcript spike ins in the eluate (m^{6}A-positive fraction), but they were readily detected in the supernatant (m^{6}A-negative fraction), confirming the high specificity. We quantified m^{6}A levels per gene by the ratio of RNA abundances, (eluate)/(eluate + supernatant). RNA abundances were represented by the RNA-seq fragment counts normalized across input, eluate, and supernatant using the 96 synthetic ERCC control RNAs equally spiked in just before library construction (see Online Methods). To assess m^{6}A-LAIC-seq's ability to quantify m^{6}A levels or stoichiometry, we mixed a series of nonmammalian m^{6}A-modified RNAs generated *in vitro*, each with one or two modifications, with an unmodified counterpart at ratios ranging from 0–80%. Quantitative agreement between the known stoichiometry and its m^{6}A-LAIC-seq quantification was excellent (Pearson $r = 0.995$; $P = 3.7 \times 10^{-5}$; Fig. 1e).

m^{6}A-LAIC-seq reveals the transcriptome-wide patterns of m^{6}A levels in H1-ESC

We recently characterized the m^{6}A-seq profile of the hESC line H1-ESC[18]. Here we applied m^{6}A-LAIC-seq to further characterize its m^{6}A epitranscriptome. On a transcriptome-wide scale, we observed a strong concordance of m^{6}A levels in two biological replicates ($r = 0.98$, $P < 2.2 \times 10^{-16}$; Fig. 2a), with m^{6}A levels following an almost bimodal distribution whereby most genes exhibited less than 50% methylation levels (Fig. 2b). The analysis of the relationship between the number of m^{6}A peaks and m^{6}A levels revealed that the more peaks present, the higher the m^{6}A levels (Supplementary Fig. 2a). Since m^{6}A RIP efficiency is independent of the number of m^{6}A sites and peaks (see above paragraph; Supplementary Fig. 1d,e), the increase in m^{6}A levels with number of peaks is consistent with at least some m^{6}A modifications occurring at different sites on separate transcripts, as opposed to all occurring *in cis* on the same transcripts. Notably, m^{6}A-seq-derived peak intensities correlated only modestly with m^{6}A levels ($r = 0.6$) (Supplementary Fig. 2b,c). Importantly, m^{6}A levels as measured by qRT-PCR correlated strongly with m^{6}A-LAIC-seq data ($r = 0.96$; $P < 2.2 \times 10^{-16}$, Supplementary Fig. 2d), crossvalidating the m^{6}A-LAIC-seq approach. Figure 2c and Supplementary Figure 2e show examples of m^{6}A-LAIC-seq profiles of genes with various m^{6}A levels. Our results offer a quantitative road map as to which genes are most (*SOX2* and *LINC-ROR*) or least (*POU5F1* and *TCF3*) likely to be influenced by m^{6}A-dependent regulatory networks (Supplementary Table 1).

m^{6}A-level-linked biological processes and protein domains

To understand potential links between biological pathways and m^{6}A levels, we analyzed Gene Ontology (GO) terms. The 1,000 most highly methylated genes based on m^{6}A levels were enriched in terms such as regulation of transcription (false discovery rate (FDR) = 2.9

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$\times 10^{-28}$) and regulation of RNA metabolic process (FDR = 7.5×10^{-24}) (Fig. 2d). This was consistent with the observation that high m^{6}A levels were enriched in transcription factor (TF) domains, the strongest being TFs containing the repressive KRAB domain (FDR = 8.3×10^{-35}; $n = 92$) and the ZNF-C2H2-type domain (FDR = 4.0×10^{-36}; $n = 143$). Conversely, the 1,000 genes with the lowest methylation levels were enriched in processes such as translational elongation (FDR = 5.7×10^{-46}) and generation of precursors of metabolites and energy (FDR = 4.2×10^{-12}). These results provided evidence that transcriptional regulators, especially transcriptional repressors, are likely the most susceptible to m^{6}A regulation.

m^{6}A levels are inversely correlated with steady-state RNA levels and half life

As m^{6}A is implicated in promoting RNA degradation[8, 18, 19], we asked how m^{6}A levels were related to RNA steady-state expression levels and half-life measurements[8, 18, 19]. We found an inverse correlation between m^{6}A levels and steady-state expression levels ($r = -0.39$; $P < 2.2 \times 10^{-16}$; Supplementary Fig. 2f). An inverse correlation was also identified between m^{6}A levels in human embryonic stem cells (hESCs) and half-life measurements in induced pluripotent stem cells (iPSCs) ($r = -0.14$; $P < 2.2 \times 10^{-16}$; Supplementary Fig. 2g)[24]. Furthermore, long intergenic noncoding RNAs (lincRNAs) had significantly higher m^{6}A levels than mRNAs or pseudogenes ($P = 3.0 \times 10^{-16}$; two-sided Wilcoxon test; Supplementary Fig. 2h and Supplementary Table 2).

m^{6}A-LAIC-seq reveals the extent of differential m^{6}A levels between cell types

To examine variation of m^{6}A levels among cell types, we performed m^{6}A-LAIC-seq on the B-cell lymphoblastoid cell line GM12878, which also exhibited highly correlated m^{6}A levels in replicates (Supplementary Fig. 2i). GM12878 cells also followed a non-Gaussian bimodal m^{6}A-level distribution, but exhibited higher methylation levels than H1-ESC ($P < 2.2 \times 10^{-16}$; two-sided Wilcoxon test; Fig. 3a and Supplementary Fig. 2j). We found that 1,509 genes exhibited a reproducible variation in m^{6}A levels of at least ±10% between cell types, but almost none showed variation larger than ±20%. Furthermore, such genes were found to largely exhibit a unidirectional increase in m^{6}A levels in GM12878 as compared to H1-ESC (ratio, 1,465:44) (Fig. 3b). qRT-PCR analyses of genes with differential m^{6}A levels between cell types showed highly concordant results compared to m^{6}A-LAIC-seq (Fig. 3c). These results imply that cell-type-specific m^{6}A levels of individual genes exhibit limited bidirectional variability, at least in the two cell types under the examined conditions. However, m^{6}A levels appear to follow more global shifts, perhaps representing different m^{6}A methyltransferase, demethylase, and/or reader expression and activity among conditions and cell types. Indeed, we found significantly higher mRNA levels of the YTH-domain-containing RBPs in H1-ESC (Fig. 3d), which would be expected to lead to higher rates of m^{6}A-targeted RNA degradation and thus to a global reduction of H1-ESC m^{6}A levels, compared to GM12878 as observed[8].

N^6,2′-*O*-Dimethyladenosine is a low-abundance modification

In transcripts starting with an adenosine, methylation of the 2′-*O* and the *N*-6 positions can lead to a modification structurally related to m^{6}A, *N*^{6},2′-*O*-dimethyladenosine (m^{6}Am), that appears to be recognized by anti-m^{6}A antibodies. m^{6}Am is thought to occur exclusively on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

first nucleotides of mRNAs, whereas m^6A has not been found at the first nucleotide position[25, 26, 27, 28]. To evaluate the impact of m^6Am on the quantification of m^6A level using m^6A-LAIC-seq, we analyzed mRNAs that were unlikely to carry m^6Am given that they lacked m^6A-seq peak signals within the first 200 nt of any annotated isoforms. We found very similar m^6A-level distributions between all mRNAs and the list of mRNAs lacking m^6Am in H1-ESC, suggesting that the presence of m^6Am was unlikely to cause significant distortions in m^6A-level quantification (Supplementary Fig. 3a). However, to our knowledge, there is no report of absolute quantitative measurement of m^6Am level in RNA. Thus, we synthesized m^6Am standards and used liquid chromatography–mass spectrometry spectrometry (LC-MS) (Supplementary Fig. 3b) to quantify m^6Am levels in poly(A)$^+$ RNA[29]. We found that H1-ESC poly(A)$^+$ RNAs contained ~3 m^6Am nucleotides per 10^5 nucleotides compared to ~100 m^6A nucleotides per 10^5 nucleotides, revealing 33 times more m^6A than m^6Am (Supplementary Fig. 3c). GM12878 also exhibited low levels of m^6Am compared to m^6A, suggesting that in both cell types m^6Am is unlikely to skew m^6A-LAIC-seq level results significantly (Supplementary Fig. 3c). Furthermore, multiple studies showed that fewer than 20% of methylated genes harbor m^6A peaks in their 5′ UTRs[18, 25]. Even with a liberal estimate that all of these m^6A peaks are m^6Am, the vast majority (>80%) of genes do not have any potential m^6Am signal, which further suggests that the impact of m^6Am on m^6A-LAIC-seq is limited.

Internal splicing differs in m^6A and non-m^6A RNA fractions

The m^6A methyltransferase METTL3 has been implicated in splicing of internal exons, although the true extent of involvement is unclear[19, 21, 30, 31, 32]. Unlike traditional m^6A-seq, m^6A-LAIC-seq can directly compare alternatively spliced isoforms between methylated and nonmethylated RNA. By analyzing differential splicing events between these fractions using our rMATS software[33, 34], at our current sequencing depth we detected only a limited number of differential splicing events in internal exons, including alternative cassette exons (n = 49 in H1-ESC; n = 77 in GM12878), and retained introns (n = 21 in H1-ESC; n = 78 in GM12878). There was a significant trend for alternative cassette exons included in the m^6A-positive fraction of H1-ESC to harbor m^6A sites as detected by m^6A-seq ($P = 2.7 \times 10^{-17}$; Fisher's exact test; Supplementary Fig. 4a,b and Supplementary Table 3). We independently validated an example of alternative-cassette-exon inclusion in the m^6A-positive fraction by RT-PCR (Supplementary Fig. 4c). We concluded that internal isoform-specific m^6A targeting may allow for differential regulation of individual isoforms by m^6A modification.

m^6A modification is associated with proximal APA usage

A prevalent form of mRNA isoform regulation outside of internal splicing is 3′-UTR length variation via APA. Visual inspection of m^6A-LAIC-seq data tracks revealed numerous examples of genes with significantly differential 3′-UTR signals in m^6A-positive versus m^6A-negative fractions. A systematic evaluation of APA based on m^6A-LAIC-seq (see Online Methods) revealed many genes (H1-ESC, n= 2,512; and GM12878, n = 2,260) in which methylated transcripts significantly used the proximal rather than the distal APA site and thus harbored short 3′ UTRs, compared to their nonmethylated transcripts (FDR < 0.01) (Fig. 4a, Supplementary Fig. 5a, and Supplementary Table 4). By contrast, we found a much smaller number of genes (H1-ESC, n = 410; and GM12878, n = 413) in which the reverse

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

was true (Fig. 4a, Supplementary Fig. 5a, and Supplementary Table 4). Furthermore, the m^{6}A-seq peak signals surrounding proximal poly(A) sites suggest that for genes with preferential m^{6}A modification of the shorter 3′-UTR isoform, the m^{6}A sites are located upstream of the proximal poly(A)$^+$ sites and are thus shared by both the shorter and longer 3′-UTR isoforms (Fig. 4b). Figure 4c and Supplementary Figure 5b show examples of m^{6}A-LAIC-seq results for genes with or without differential APA usage. We confirmed our findings in these examples, calculating the change in the percentage of distal polyadenylation (dPAS) usage index (DPDUI) by qRT-PCR (Fig. 4c and Supplementary Fig. 5b)[35]. Of note, genes in which the proximal APA usage in the m^{6}A-positive fraction is greater than that in the m^{6}A-negative fraction, designated P/D, m^{6}A$^+$ > m^{6}A$^-$, also exhibited higher m^{6}A levels than genes without differential usage ($P = 1.8 \times 10^{-152}$, two-sided Wilcoxon test; Supplementary Fig. 5c,d). A recent study reported that genes with longer last exons have a higher density of m^{6}A peaks[36]. Consistent with this observation, we found that m^{6}A levels of genes are positively correlated with maximum 3′-UTR lengths as measured in input steady-state RNA-seq data (Supplementary Fig. 5e). Detailed analyses showed that genes with higher proximal APA usage in the m^{6}A-positive fraction (P/D; m^{6}A$^+$ > m^{6}A$^-$) had significantly longer 3′ UTRs (Supplementary Fig. 5f) and lower ratios of proximal versus distal signals (Supplementary Fig. 5g,h) in the input RNA-seq data as compared to genes with higher distal APA usage in the m^{6}A-positive fraction (P/D; m^{6}A$^+$ < m^{6}A$^-$) or no change in APA usage between fractions (P/D; m^{6}A$^+$ = m^{6}A$^-$). Given that m^{6}A modification is strongly coupled with RNA degradation[3, 15, 20, 21], our data suggest one possible model in which preferential m^{6}A methylation and degradation of shorter 3′-UTR isoforms using proximal poly(A) sites simultaneously result in their increase in the m^{6}A-positive fraction and decrease in the total steady-state population of RNA.

One possible interpretation of our findings is that m^{6}A and/or METTL3 may directly dictate APA choice. To test this possibility, we analyzed 3′-UTR read distribution of published RNA-seq or 3p-seq data in wild-type versus knockout or knockdown of m^{6}A writers (*METTL3*, *METTL14*, and *WTAP*) from multiple human and mouse cell lines[14, 15, 18, 19]. We observed no or limited numbers of significant APA changes in all data sets, with a trend for increased signals of proximal poly(A) sites upon depletion of m^{6}A writers (Supplementary Table 5). Thus, compared to thousands of differential APA usage events in methylated versus nonmethylated transcripts of the same genes, the direct effect of m^{6}A on APA was considerably smaller, suggesting that it is unlikely to explain the widespread association between m^{6}A and proximal APA usage revealed by m^{6}A-LAIC-seq.

N^6,2-*O*-Dimethyladenosine (m^{6}Am) does not correlate with APA usage

To remove the potential confounding effects of m^{6}Am, we repeated the APA analyses of m^{6}A-LAIC-seq data in genes without any peak within the first 200 nt of transcripts as detected by m^{6}A-seq and obtained similar results (Supplementary Fig. 6a). We also examined APA usage upon stratification of genes based on whether their transcription start sites (TSSs) are adenosine in H1-ESC and GM12878. The TSSs were annotated based on Ensembl annotation or cap analysis gene expression (CAGE) data. We found similarly strong association of m^{6}A with proximal APA usage in genes with or without A at TSSs based on TSS annotation (Supplementary Fig. 6b). These results suggest that the association

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

between m^{6}A and proximal APA usage revealed by m^{6}A-LAIC-seq is not dictated by the potential presence of m^{6}Am at the starting nucleotide position.

APA of m^{6}A versus non-m^{6}A fractions affects regulatory elements

The 3′ UTRs of transcripts serve as post-transcriptional nexuses of regulation by miRNAs as well as RBPs[37]. To assess the consequence of differential 3′-UTR usage patterns between modified and unmodified m^{6}A isoforms, we analyzed potential miRNA–mRNA and RBP–mRNA interactions. To do so, we analyzed predicted miRNA target sites as well as RBP-binding motifs in the extended segment of 3′ UTR between the proximal and distal poly(A)$^+$ sites in genes with differential APA usage, while using genes with m^{6}A modifications and tandem APA sites but no difference in APA usage between methylated and nonmethylated fractions as controls. Genes with higher proximal APA usage in the m^{6}A-positive fraction as compared to the m^{6}A-negative fraction (P/D; m^{6}A$^+$ > m^{6}A$^-$) show a higher degree of transcript regulation in the extended segment by miRNAs and RBPs, as compared to genes with higher distal APA usage in the m^{6}A-positive fraction (P/D; m^{6}A$^+$ < m^{6}A$^-$), or no change in APA usage between fractions (P/D; m^{6}A$^+$ = m^{6}A$^-$) (Fig. 5a,b). By contrast, genes with higher distal APA usage in the m^{6}A-positive fraction (P/D; m^{6}A$^+$ < m^{6}A$^-$) had a higher density of m^{6}A RRACU motif in the extended region (Fig. 5c).

Discussion

We envision that m^{6}A-LAIC-seq complements standard m^{6}A-seq identification of methylation sites and adds to our armamentarium for the study of m^{6}A biology. As demonstrated in this work, the combined use of these two technologies provides new insights into the dynamic range and isoform complexity of the m^{6}A epitranscriptome. Factors that determine m^{6}A levels remain to be elucidated, but they may be related to a number of interacting and stochastic processes ranging from the concentration of m^{6}A readers, erasers, and writers to the *in vivo* folding and accessibility of RNA. Future application of m^{6}A-LAIC-seq to a larger number of tissue types and cellular states will help further elucidate tissue or cell-state specificity of m^{6}A levels and identify the regulatory logic controlling m^{6}A levels. Theoretically, RNA secondary structures might affect m^{6}A-LAIC-seq results if the structures in full-length RNAs shield the m^{6}A sites from being recognized by the antibody. However, this scenario is unlikely, given that previous studies have shown that m^{6}A modifications occur in unstructured regions[31, 32].

A unique application of m^{6}A-LAIC-seq is to deconvolute and contrast the isoform complexity of m^{6}A-methylated versus nonmethylated RNA fractions in a given cell type. We found that thousands of genes with m^{6}A-modified transcripts exhibited a tendency to utilize proximal APA sites, while for hundreds of genes the converse was true. These data suggest that many genes generate at least two distinct pools of transcripts, encoding the same protein in the case of mRNAs, with differing metabolic fates based on m^{6}A coupled with differential binding of miRNAs and RBPs in their UTR regions. Indeed, our findings are consistent with a potential model that the m^{6}A fraction of genes that often harbor shorter 3′ UTRs turns over more rapidly, leading to the observation in the steady-state RNA population that such genes have longer 3′ UTRs. Collectively, these observations may reflect the need to fine-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

tune the amount of RNA or protein under different circumstances using distinct post-transcriptional mechanisms[38]. For genes with tandem APA sites in the 3′ UTR, the shorter transcript isoforms have less regulatory sites at the 3′ UTR than the longer isoforms. However, they preferentially acquire m^{6}A modifications and potentially turn on m^{6}A-mediated regulatory pathways. Furthermore, our results suggest that rather than m^{6}A METTL3 dictating APA choice, perhaps the opposite is true. APA factors may help determine m^{6}A site distribution along transcripts, consistent with the predominance of m^{6}A sites in 3′ UTR or last exon, along with data linking APA factors to internal exon splicing[39]. Further studies are needed to investigate the role of the APA machinery on the installment of m^{6}A along transcripts[40].

m^{6}A-LAIC-seq provides a new vantage point from which to dissect and classify the human transcriptome. For example, using m^{6}A-LAIC-seq as a road map, it should now be possible to study the effects of differentially depleting m^{6}A-modified versus m^{6}A-unmodified transcripts for the same gene by designing UTR-segment-specific shRNA or siRNAs. Furthermore, strategies employing RNA-targeting Cas9 systems will allow the rationale design of experiments based on LAIC-seq maps to track and manipulate methylated versus nonmethylated subfractions of a transcript based on their differential UTRs[41]. In addition, our work sets the foundation for subphenotyping methylated versus nonmethylated transcripts with respect to parameters such as their three-dimensional structures, half lives, and the binding specificity of RBPs on genome-wide scales.

In summary, m^{6}A-LAIC-seq has revealed a new design principle embedded in the human transcriptome based on m^{6}A modifications and 3′-UTR length combinatorics. Given the increasing recognition of other RNA modifications, such as 5mC, m1A, and pseudouridylation (ψψ)[42], our results may point to an even greater heterogeneity and isoform complexity of the transcriptome as part of an RNA epitranscriptome code.

Methods

Antibodies

Anti-m^{6}A antibody: the anti-m^{6}A antibody was obtained from Synaptic Systems (cat# 202 003) m^{6}A rabbit polyclonal, affinity purified. Normal rabbit serum R9133-5ML, Sigma (cat# R9133). Donkey secondary anti-rabbit antibody was obtained from Amersham (cat# NA934) (antibody dilution factor, 500 ng μg^{-1}).

Cell line

The GM12878 cell line was obtained by the Coriell Institute for Medical Research (cat# GM12878). The species of origin was confirmed by nucleoside phosphorylase, glucose-6-phosphate dehydrogenase, and lactate dehydrogenase. The cell lines were genotyped based on microsatellite polymorphism assay used by Coriell Institute. The GM12878 cell line tested negative for mycoplasma contamination using the polymerase chain reaction (PCR) method developed by Applied Biosystems and the MycoSEQ Mycoplasma Detection System (4460625, Life Technologies).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The H1 embryonic stem cell line was obtained from WiCell (name: WA1, alias: H1). The cell tested negative for mycoplasma infection based on PCR method.

Synthesis of m^{6}Am

The synthesis of N^6, 2′-*O*-dimethyladenosine was analogous to a published procedure for the synthesis of 2′-*O*-methyladenosine, except for the use of N^6-methyladenosine (Berry & Associates) as starting material[29]. Purification was performed on an Agilent 1200 series HPLC equipped with a diode array detector (DAD), fraction collector, and a Synergy Fusion RP column (4 μm particle size, 80 Å pore size, 250 mm length, and 4.6 mm inner diameter) from Phenomenex at 35 °C. The solvents consisted of 5 mM ammonium acetate buffer adjusted to pH 5.5 with acetic acid (solvent A) and acetonitrile (solvent B). The elution at a flow rate of 1 mL/min started with 90% solvent A followed by a linear gradient to 21% solvent B at 15 min. Solvent B was increased further to 50% over 5 min, followed by a decrease to 10% over 3 min. The desired fractions were dried under vacuum.

Analysis of m^{6}A and m^{6}Am by liquid chromatography-coupled mass spectrometric analysis (LC-MS)

Prior to mass spectrometric analysis, all RNA samples were hydrolyzed enzymatically to ribonucleosides as described previously[44]. Briefly, the digestion was carried out in 0.1 mM Tris buffer (pH 8), 5 mM MgCl2, 0.0375 U μL^{-1} benzonase, 0.17 U μL^{-1} alkaline phosphatase, 1 U mL^{-1} phosphodiesterase, 3 mM desferroxamine (antioxidant), 0.3 mM butylated hydroxytoluene (antioxidant), 0.05 μM [$^{15}N5$]-2′-deoxyadenosine (internal standard), and 5 μM 2′-deoxyinosine (internal standard) at 37 °C for 2 h. Enzymes were then removed using a YM-10 centrifugal spin column (Millipore).

Quantitative analyses of m^{6}A and m^{6}Am were achieved using an Agilent 1200 HPLC coupled to an Agilent 6430 triple quadrupole mass spectrometer in positive-ion mode using dynamic multiple reaction monitoring (MRM). The ribonucleosides in the hydrolyzed RNA samples were resolved on a Phenomenex C18 HPLC column (1.7 μm particle size, 100 Å pore size, 2.1 × 150 mm; 25 °C) at 330 μL/min using a solvent system consisting of 10 mM ammonium acetate in H2O (A) and acetonitrile (B). The elution profile was 0% B for 3 min, 0–7% B over 20 min, then to 7–40% B over 4 min, followed by a column washing at 80% B and column equilibration. The operating parameters for the mass spectrometer were as follows: gas temperature 350 °C; gas flow 10 l/min; nebulizer 50 psi and capillary voltage 3500 V, fragmentor voltage 100 V, and collision energy 15 V. The quantification of a ribonucleoside can be achieved using mass-to-charge ratio (*m*/*z*) of the parent ribonucleoside ion and *m*/*z* of the deglycosylated product ion. The nucleosides were quantified based on the transition of the parent ribonucleoside to the deglycosylated base ion: *m*/*z*, 282.1/150.1 for m^{6}A and *m*/*z*, 296.1/150.1 for m^{6}Am. Absolute quantities of each m^{6}A and m^{6}Am were determined respectively using an external calibration curve prepared with synthetic standards for each ribonucleoside.

m^{6}A-level dot blots

Amersham Hybond-XL (cat# RPN303s) membrane was rehydrated in H2O for 3 min. The membrane was then ‘sandwiched’ in Bio-Dot Microfiltration Apparatus (Bio-Rad, cat#

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

170-6545). Each well was then filled with H2O and flushed by gentle suction vacuum until it appeared dry. 5 µl of H2O alone was then applied to the membrane in each well followed by addition of indicated amount of RNA and this was allowed to bind to the membrane by gravity. The apparatus was disassembled and the membrane was crosslinked in a UV STRATALINKER 1800 using the automatic function, and then the membrane was placed back into the apparatus. The membrane was then blocked 10 min using sterile RNase DNase free TBST + 5% milk. The m^{6}A primary antibody (Anti-m^{6}A, Synaptic Systems, cat# 202 003) was then added at a concentration of 1:500 at room temperature for 1 h in TBST + 5% milk. The membrane was then washed four times in PBST. The membrane was then incubated with the secondary anti-rabbit antibody (1:5,000 dilution) for 30 min in TBST + 5% milk. The membrane was washed 4 times for 5 min in TBST and exposed on an autoradiographic film using Pierce ECL Western Blotting Substrate.

cDNA synthesis for qRT-PCR analysis

First, a mix was made of 100 ng of RNA in 5 µl volume, 2 µl of random hexamers (Roche), 1 µl of dNTP Mix (10 mM each) and 5 µl of ultrapure H2O was first generated, heated at 65 °C for 5 min and immediately put on ice. 4 µl of 5× first-strand buffer was added along with 1 µl of 0.1 M DTT, 1 µl RNase inhibitor and 1 µl of Superscript III reverse transcriptase (Invitrogen). The 20 µl reverse transcription reaction was then incubated 5 min at room temperature, then 60 min at 50 °C then 15 min at 70 °C. The freshly synthesized cDNA was treated with 1 µl of RNase H at 37 °C for 20 min.

SYBR Green qRT-PCR

To calculate m^{6}A levels by qPCR for Figures 2d and 3c and Supplementary Figures 2d and 3e, SYBR Green quantitative real-time PCR assays were performed utilizing a spike-in synthetic control called *Drosophila* E (see below) for normalization. Each PCR reaction was done in a 10 µl volume made of 5 µl of master mix (SYBR GreenER qPCR SuperMix for iCycler, Invitrogen), 2.5 µl of primer mix at 1.2 µM (each) and 2.5 µl of cDNA template at 20 ng µl^{-1}. The PCR was carried out using a standard protocol with melting curve. The amount of target was calculated using the formula: amount of target = $2^{-\Delta\Delta CT}$ (ref. 45).

Calculation of ΔPDUI

All ΔPDUI calculations were based on the method described in Masamha *et al.*[35], adopted to compare UTR length in positive and negative fractions of the m^{6}A IP of m^{6}A-LAIC-seq. The following was calculated using primer sets designed in the common part of the isoforms of a transcript (called ‘common primer set’) as well as in the distal part of long isoform of the same transcripts (called ‘distal primer set’): Δ*C*T (common and distal) was normalized to an artificial spike-in transcript DrosoE (described in spike-ins controls for m^{6}A-LAIC-seq). ΔΔ*C*T = Δ*C*T distal − Δ*C*T common. ΔΔΔ*C*T = ΔΔ*C*T negative fraction m^{6}A IP − ΔΔ*C*T positive-fraction m^{6}A IP. The increase or decrease in ΔΔΔ*C*T was given by $\pm 2^{\Delta - \Delta CT}$ (ref. 45).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA extraction, DNase I treatment, and poly(A)$^+$ selection

Total RNA was isolated from cells according to manufacturer's instructions using TRIzol LS reagent (Ambion). Total RNA was treated using DNase I (Promega) for 20 min at 37 °C. The treated RNA was then acid-phenol-chloroform extracted and chloroform extracted. The RNA was precipitated using 300 mM final concentration of NaCl2 spiked with 1 μl of 50 mg/ml of Ultra Pure Glycogen (Promega) and 2.5 volume of 100% ethanol at −20 °C either for 2 h or overnight. The precipitated RNA was then centrifuged using a refrigerated tabletop at maximum speed (>13,000*g*) at 4 °C for 20 min. The precipitated RNA was then washed with 70 °C ethanol and centrifuged at maximum speed for an additional 10 min. The final pellet was then resuspended in ultrapure H2O. Poly(A)$^+$ RNA selection was performed twice using Dynabeads mRNA Purification Kit (Invitrogen cat. # 610.06) according to the manufacturer's protocol. The second poly(A)$^+$ RNA selection was performed using the eluate of the first poly(A)$^+$ RNA selection as starting material according to the manufacture's instruction. The obtained poly(A)$^+$ RNA was evaluated both by NanoDrop and Bioanalyzer. For all RNA samples, the concentration, purity, and integrity of the RNA were verified using a NanoDrop and Bioanalyzer.

H1-ESC cell and GM12878 cell culture

H1 (WA01) cells were cultured in feeder-free condition using mTESR1 media (Stem Cell Technologies cat. # 05850) on 6-well plates coated with Matrigel (BD Biosciences, cat. # 354603), as described[46]. GM12878 cells were cultured on complete RPMI media (Invitrogen) supplemented with 15% FBS and 10% Pen–Strep antibiotic.

m^{6}A level and isoform-characterization sequencing (m^{6}A-LAIC-seq)

Of note for each biological replicate for m^{6}A-LAIC-seq, we started with 150 μg of total RNA yielding approximately 3.8 μg of double poly(A)-selected RNA which was resuspended in a final volume of 50 μl using UltraPure H2O (Life Technologies cat# 10977-015). The 2× poly(A)$^+$ RNA was then heated at 65 °C for 5 min and immediately put on ice. 50 μl of m^{6}A–DynaBeads (the m^{6}A antibody, Synaptic Systems cat# 202 003, was coupled to Dynabeads using the Life Technologies coupling kit cat# 14311D) were equilibrated by washing twice for 5 min in 500 μl of m^{6}A–Binding Buffer (50 mM Tris-HCl, 150 mM NaCl2, 1% NP-40, 0.05% EDTA). The RNA and the spike-ins controls were then added to the equilibrated m^{6}A–DynaBeads. The RNA was allowed to bind to the m^{6}A–Dynabeads (in 500 μl volume of m^{6}A–Dynabeads and m^{6}A–Binding Buffer at room temperature while rotating (tail over head) at 7 rotations per min for 1 h). The tubes containing the samples were placed on a magnet, allowing the beads complexes to cluster for 1 min or until the solution become clear. The liquid phase was carefully collected and placed on ice as this 500 μl fraction represented the 'supernatant' of the m^{6}A IP. Following the collection of the supernatant fraction, series of washes were performed using various buffers (see as follows). For all wash steps, with the exception of the elution step, the beads were washed for 3 min then placed on a magnet and the wash buffers were discarded. Following the supernatant collection, Wash step 1: the remaining fractions bound to the beads were washed twice in 500 μl of m^{6}A–Binding Buffer (50 mM Tris-HCl, 150 mM NaCl2, 1% NP-40, 0.05% EDTA). Wash step 2: the RNA–beads complexes were washed once in 500 μl

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of low-salt buffer (0.25× SSPE, 0.001 M EDTA, 0.05% Tween-20, 37.5 mM NaCl). Wash step 3: the RNA–beads complexes were washed once in 500 µl of high-salt buffer (0.25× SSPE, 0.001 M EDTA, 0.05% Tween-20, 137.5 mM NaCl). Wash step 4: the RNA–beads complexes were washed twice in 500 µl of TET (T.E. + 0.05% Tween-20). Elution step: The m^6A RNA was eluted from the beads by repeating the following four times: 125 µl of elution buffer (0.02 M DTT, 0.150 M NaCl, pH 7.5 0.05 M Tris-HCl, 0.001 M EDTA, 0.10% SDS) was added to the beads and incubated at 42 °C for 5 min. At the end of the 5 min the beads were gently placed on the magnet. The liquid phase was collected and transferred to a fresh tube as this will represent the eluate fraction containing the m^6A 'enriched RNA'. An additional 125 µl of elution buffer was then added to the beads and the process was repeated. The liquid phase obtained at each step was added to the 'fresh tube' containing the 125 µl of eluate from the previous step so the total final eluate volume was 500 µl.

All RNA fractions were extracted as follow. 500 µl of acid phenol-chloroform (acid-phenol:chloroform), pH 4.5 (with IAA, 125:24:1; Ambion) was added to the 500 µl sample. The sample was centrifuged at 4 °C at 10,000*g* for 7.5 min. The upper phase was carefully collected, ensuring that it did not touch the interphase, and it was transferred to a clean 1.5 ml tube. 500 ml of chloroform was added to the fresh tube, mixed by gentle manual shaking, and centrifuged at 4 °C at 10,000*g* for 7.5 min. The upper phase was transferred to a fresh 1.5 ml tube and NaCl2 ethanol precipitated overnight at −20 °C in the presence of 1 µl of (20 mg/ml) Ultra Pure Glycogen. The following day the sample was centrifuged at 4 °C for 20 min at 16,000*g*. The pellet was then washed in 70% ethanol and centrifuged an additional 10 min at 4 °C at 16,000*g*. The pellet was then left to dry at room temperature for 10 min before it was resuspended in the desired volume of Ultra-Pure H20 (Invitrogen cat# 10977-015). Note that only one round of anti-m^6A RNA immunoprecipitation for m^6A-LAIC-seq was performed given our data that under the conditions we used there was essentially 100% efficiency in pulling down m^6A-positive transcripts (Fig. 1 and Supplementary Fig. 1).

Spike-in controls for m^6A-LAIC-seq

For each sample after double-poly(A)-RNA selection, but before anti-m^6A RIP, *in vitro*–transcribed transcripts with and without m^6A modification were mixed into the 2× poly(A)$^+$ RNA as spike-in controls at the indicated percentage of m^6A-modified to m^6A-unmodified transcript, the indicated number of copies, and the indicated percentage of m^6A in the modified transcript. For all samples these spike ins included: 80% mCHERRY (10^8 copies, 0.2% m^6A), 60% XEF (10^8 copies, 0.2% m^6A), 20% eGFP (10^8 copies, 0.2% m^6A), luciferase unmodified (10^8 copies, 0% m^6A), and Kanamycin unmodified (10^8 copies, 0% m^6A). An additional unmodified transcript originating from *Drosophila* called transcript DrosC (EST:RH10804; 10^8 copies, 0% m^6A) was also added before anti-m^6A RIP. Post-m^6A RIP 5 µl of 1:100 dilution of the universal ERCC spike-in control A (Invitrogen) was added to each fraction (input fraction, m^6A^- fraction and m^6A^+ fraction) as well as transcript originating from *Drosophila* called transcript-unmodified DrosE (EST:IP03042; 10^8 copies, 0% m^6A).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Library construction for RNA-seq

100 ng of RNA was used for library construction for m^{6}A-LAIC-seq input, supernatant (m^{6}A-negative fractions), and eluate (m^{6}A-positive fractions) using the TrueSeq Stranded mRNA Sample Preparation Guide, entering the protocol by adding the fragment, prime, finish mix, skipping the elution step and proceeding immediately to the synthesis of the first strand cDNA. From that point on, the exact steps of the Illumina TruSeq Stranded mRNA Sample Preparation Guide were followed to the end.

m^{6}A-antibody-binding saturation and sensitivity titration

We generated an m^{6}A antibody titration curve to identify the point of saturation of the anti-m^{6}A antibody in the context of performing m^{6}A RIPs (Fig. 1b). To do so, we used an *in vitro*–generated transcript from a plasmid containing cDNA of *GAPDH* transcript (1,008 bps; 233 ATPs). The plasmid was first linearized by restriction digest using SalI just downstream of the *GAPDH* cDNA cloning site. The linearized plasmid was gel purified and *in vitro* T7-mediated transcription was performed using the Ambion MEGAscript Kit (AM1334) as described in the user manual using 100% N^6-methyladenosine-5′-triphosphate (cat# N1013) nucleotide during synthesis adding α-32P-CTP as tracer (PerkinElmer cat# BLU008H250UC). The anti-m^{6}A RIP was performed as described in the m^{6}A-seq section, using anti-m^{6}A antibody (Synaptic system # 202 003) or normal rabbit serum on respectively 10 μg, 2 μg, and 400 ng of *in vitro*–transcribed transcript (antibody dilution factor, 500 ng μg^{-1}). m^{6}A dot blots were performed on the m^{6}A-positive (eluate) and m^{6}A-negative (supernatant) fractions resulting from the RIP as shown in Figure 1b. To test for sensitivity of anti-m^{6}A RIP *in vitro* on full-length transcripts, we synthesized α-32P-CTP *GAPDH* transcripts with on average 2 m6A modifications per transcript, as well as unmodified α-32P-CTP *GAPDH* transcripts. Subsequently, various ratios (shown as percentages) of unmodified versus m^{6}A-modified *GAPDH* transcripts were mixed (e.g., 0.1% indicates 1,000 unmodified to 1 modified transcripts). The various percentage ratios of mixed transcripts were then subjected to anti-m^{6}A RNA immunoprecipitation (m^{6}A-RIP) using anti-m^{6}A antibody or normal rabbit serum (NRS-RIP). The indicated fractions were subjected to gel electrophoresis and radiographically visualized (Fig. 1c).

RNA sequencing

All libraries were prepared as previously described for RNA-seq. Each individual library fragment size was verified on Agilent Bioanalyzer 2100 with high-sensitivity chip. Final quantification was done by qPCR on PerkinElmer 2500Fast with Kapa library quantification kit (# KK4824). Libraries were pooled at equimolar concentrations according to the manufacturer guidelines (TruSeq Stranded mRNA Sample Preparation Guide, September, 2012). After clustering on Illumina cBot, samples were run on Illumina HiSeq 2000.

Anti-m^{6}A RNA immunoprecipitation is efficient and independent of the number of m^{6}A modifications

Total RNA was first isolated using TRIzol reagent. Following RNA isolation the total RNA was further treated using RQ1 RNase free DNase (PROMEGA # M610A). The RNA was then poly(A)-selected twice using Ambion Dynabeads mRNA Purification Kit. The quality

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and the integrity of the RNA was checked by BioAnalyzer. The 2× poly(A)$^+$ RNA was split into 5 equal quantities. Each tube was spiked with *in vitro*–generated GFP transcripts containing one of five variant percentiles of incorporated m^{6}A nucleotides (100% m^{6}A, 20% m^{6}A, 2% m^{6}A, 0.2% m^{6}A, or 0% m^{6}A). Note that every percentage variant of m^{6}A spike in underwent one round of m^{6}A RIP and eluate was obtained to ensure the purity of each transcript after *in vitro* synthesis before being used as a spike in. Each of the five poly(A)$^+$ RNA pools with unique spike in were subject to m^{6}A RIP and fractions were subjected to a qRT-PCR for GFP to calculate the percentage of recovery of GFP post-m^{6}A-RIP. To check for overall efficiency of anti-m^{6}A RIP a similar experiment was performed with total RNA and a single spike in of 2% m^{6}A GFP and subjected to anti-m^{6}A RIP followed by qRT-PCR of GFP and anti-m^{6}A dot blot of input, m^{6}A positive (eluate), and m^{6}A negative (supernatant). The anti-m^{6}A RIP was performed a second time on the supernatant from the first round of anti-m^{6}A RIP and each subsequent fraction was also subject to qRT-PCR and m^{6}A dot blotting.

Read-mapping and gene-expression analyses

We mapped strand-specific paired-end RNA-seq reads first to Ensembl genes (release 74) and then to hg19 human genome incorporated with all spike-in sequences using tophat (v 1.4.1)[47]. Only the properly paired and uniquely mapped reads were used in future analyses. FPKMs of input fractions were calculated using Cufflinks (v 2.0.2)[48]. Differentially expressed genes (DEGs) between H1-ESC input and GM12878 input were determined by Cuffdiff (v 2.0.2)[48].

Read coverage distribution comparison

To compare the read coverage distributions around m^{6}A peaks during the three fractions of m^{6}A-LAIC-seq and m^{6}A-seq in H1-ESC, for each peak, we calculated the RNA-seq read coverage of the upstream and downstream 1,500 bp regions on the longest isoform of the Ensembl gene. The peaks with upstream or downstream transcript regions less than 1,500 bp were removed from the analyses. For each 3,000 bp region, the read coverage of each site was normalized by the average of the whole region. To plot Figure 1d, the normalized read coverage of all the 3,000 bp regions around m^{6}A peaks were averaged at each site.

m^{6}A-level calculation

The ratio of RNA abundance (copies) between eluate and supernatant was calculated first in order to calculate m^{6}A level for each gene or spike-in RNA before m^{6}A-RIP. Since equal amount of ERCC RNA Spike-In Mix were added to eluate and supernatant fractions right after m^{6}A RIP, the abundance of each gene or spike-in RNA in eluate or supernatant was represented by the RNA-seq fragment (read pair) counts normalized by the fragment counts of the ERCC RNAs. So, for each gene or spike-in RNA before m^{6}A RIP, the ratio of RNA abundance between eluate and supernatant could be calculated using the ratio of its RNA-seq fragment counts and the ratio of fragment counts of ERCC RNAs. Because there were 92 RNAs with a variety of concentrations in ERCC mixture, the ratio of ERCC RNA fragment counts between eluate and supernatant in each replicate could be accurately calculated using these RNAs together. As shown in Supplementary Figure 7, we used the log2-transformed counts of ERCC RNAs to fit a linear regression model (eluate counts as a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

function of supernatant counts) with coefficient of 1, and the log2 ratio between ERCC eluate counts and supernatant counts was indicated by the residual of the regression formula. To avoid using unreliable counts, only the ERCC RNAs with at least 100 counts were used to fit the regression model.

The m^{6}A levels of a specific Ensembl-annotated gene or spike-in RNA before m^{6}A-RIP were calculated using eluate counts (E), supernatant counts (S), and the residual of ERCC regression (R) according to the following formula:

$$\mathrm{m^6A\ level}=\frac{E}{E+S\times 2^R}$$

m^{6}A-level analyses

For all the analyses of m^{6}A levels, we required eluate counts ≥50 or supernatant counts ≥50 in both replicates to obtain reliable m^{6}A levels. To compare the m^{6}A levels with m^{6}A peaks, we used our previously published[18] m^{6}A peaks of H1 ESC cell line. The peak intensities of genes were represented by the maximum peak intensities of the peaks in the genes, and we also required the input RPKMs of the peaks ≥5 in both replicates to obtain reliable peak intensities. Gene ontology (GO) analyses for top 1,000 highly methylated genes and bottom 1,000 weakly methylated genes were conducted using DAVID[49]; all genes with the above required read counts for level calculation were used as background. Published half-life measurements of iPSC[24] were used to compare with m^{6}A levels, and the genes with average FPKMs of the two H1 replicates <5 were removed from the analyses.

Comparison of m^{6}A levels between cell lines

For all the comparisons of m^{6}A levels between H1-ESC and GM12878 cell lines, we required eluate counts ≥50 or supernatant counts ≥50 in both replicates of both H1-ESC and GM12878 cell lines to obtain reliable m^{6}A levels. Differentially methylated genes were determined as the genes with m^{6}A-level differences between cell types of > 10% or 20%.

Splicing analyses

We used rMATS (v 3.0.8)[34] to determine the differential alternative splicing events between different libraries. Four types of alternative splicing events (skipped exon (SE), alternative 5′ splice site (A5SS), alternative 3′ splice site (A3SS), and retained intron (RI)) based on annotated Ensembl genes were tested for inclusion-level difference ≥5%, and the events with FDR < 0.05 were determined as differential alternative splicing events.

Differential usage of alternative poly(A) sites

Poly(A) sites from the APADB database[50] were identified within the 3′-UTR regions of Ensembl protein-coding genes. For genes with multiple poly(A) sites, any pairs of poly(A) sites are defined as the end of the 3′ termini of common and extended 3′ UTR (the different region between longer 3′ UTR and shorter 3′ UTR) segments. If the lengths of the 3′-UTR segments are both greater than 100 nt (to rule out the regions determined by two close poly(A) sites), they are considered a valid pair of tandem UTRs. For each pair of tandem

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

UTRs, the number of RNA-seq reads mapped within the common and extended 3′ UTRs were calculated for each replicate of supernatant and eluate, respectively. The P:D (proximal:distal) ratio plus 1 (P:D + 1) was calculated as the ratio of region-length normalized read counts between common and extended 3′ UTR. We then used the statistic model of rMATS[34] to test the differences of the read counts ratio of common and extended 3′ UTRs between supernatant and eluate; the 3′-UTR pairs with read depths (total reads in common and extended 3′ UTR) <5 in any replicate of supernatant or eluate were removed from the calculation. The 3′-UTR pairs with FDR < 0.01 were then identified as differentially used alternative poly(A) sites between supernatant and eluate; we then took the most significant 3′-UTR pair of each gene for the downstream analyses. Based on the above analyses, we classified three categories of genes for further analyses: group 1, eluate has higher proximal poly(A) site usage; group 2, supernatant has higher proximal poly(A) site usage; group 3, eluate and supernatant have no significant (FDR > 0.5) difference of APA usage.

We then studied the m^{6}A peak coverage around proximal poly(A) sites for genes in group 1 and group 2, respectively, in H1-ESC. For the transcript with the extended 3′ UTR of each gene, the peak coverage of upstream and downstream 750 bp flanking the last nucleotide of proximal poly(A) site was calculated. The transcripts with either flanking regions <750 bp were removed from the analyses. To study the relationship between m^{6}A level and 3′-UTR length, we used the longest 3′ UTR of all isoforms for each gene.

m^{6}Am analyses

To create a list of protein-coding genes that were unlikely to harbor m^{6}Am, we removed the genes with m^{6}A peaks within the first 200 nt of at least one Ensembl-annotated transcript isoform in H1-ESC and GM12878, respectively. The m^{6}A peaks of GM12878 were obtained from previous publication[51]. To determine whether the genes had A at TSSs, we used the Ensembl-annotated TSSs of the longest protein-coding isoforms or TSSs annotated by ENCODE CAGE data of poly(A)$^+$ mRNAs of H1-ESC and GM12878. The processed data summarizing CAGE tags were downloaded from http://ccg.vital-it.ch/mga/hg19/encode/GSE34448/GSE34448.html. For the CAGE data, we only used the TSSs with CAGE tags ≥10 in both replicates and the TSSs must have been located in the 5′ UTR of first exon of the longest protein-coding isoforms. For each gene, the TSS with maximum average CAGE tags was used in the analyses.

miRNA and RBP analyses

We compare the densities of miRNA targets and RBP binding sites in the extended 3′ UTRs in the 3 categories of genes in terms of poly(A) usage as described above. miRNA targets predicted by TargetScanHuman 5.1 were downloaded from the UCSC genome browser[52]. RBP binding sites were predicted through matching RNAcompete-derived motifs of 110 RNA-binding proteins and other well-known RNA-binding motifs[53, 54, 55]. Only the genes with extended 3′-UTR lengths between 500 nt and 2,000 nt were used in RNA-binding-protein analyses.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data visualization

UCSC custom data tracks were generated using the uniquely mapped reads. For Figure 2c,e and Supplementary Figure 2e, to make the read coverage of supernatant and eluate reflect the real ratio of their RNA abundance, we used the raw reads coverage of eluate, but we multiplied the supernatant read coverage by 2^R (as described above, R was the residue of ERCC regression formula).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by an MGH startup and ECOR grants to C.C.G. This study was also supported by National Institutes of Health (NIH) grant R01GM088342 and an Eli & Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA and Rose Hills Foundation Research Award to Y.X. Y.X. is supported by an Alfred Sloan Research Fellowship. P.D. was supported by the National Research Foundation of Singapore through the Singapore–MIT Alliance for Research and Technology, National Institute of Environmental Health Science grants ES002109 and ES024615, and National Science Foundation grant CHE-1308839. A.C.M. was supported by NIH grant DK090122. We thank D. Mirsky for copy editing the manuscript. We thank J. Wan for technical support on data analyses. We thank H. Chang, P. Batista, and K. Jeffrey for reading the manuscript and providing helpful comments.

References

1. Fu Y, He C. Nucleic acid modifications with epigenetic significance. Curr Opin Chem Biol. 2012; 16:516–524. [PubMed: 23092881]
2. Sibbritt T, Patel HR, Preiss T. Mapping and significance of the mRNA methylome. Wiley Interdiscip Rev RNA. 2013; 4:397–422. [PubMed: 23681756]
3. Meyer KD, Jaffrey SR. The dynamic epitranscriptome: N^6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol. 2014; 15:313–326. [PubMed: 24713629]
4. Klungland A, Dahl JA. Dynamic RNA modifications in disease. Curr Opin Genet Dev. 2014; 26:47–52. [PubMed: 25005745]
5. Pan T. N^6-methyl-Adenosine modification in messenger and long non-coding RNA. Trends Biochem Sci. 2013; 38:204–209. [PubMed: 23337769]
6. Schwartz S, et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell. 2013; 155:1409–1421. [PubMed: 24269006]
7. Wang Y, et al. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat Cell Biol. 2014; 16:191–198. [PubMed: 24394384]
8. Wang X, et al. m^6A-dependent regulation of messenger RNA stability. Nature. 2014; 505:117–120. [PubMed: 24284625]
9. Meyer KD, et al. 5′ UTR m^6A promotes cap-independent translation. Cell. 2015; 163:999–1010. [PubMed: 26593424]
10. Zhou J, et al. Dynamic m^6A mRNA methylation directs translational control of heat shock response. Nature. 2015; 526:591–594. [PubMed: 26458103]
11. Wang X, et al. N^6-methyladenosine modulates messenger RNA translation efficiency. Cell. 2015; 161:1388–1399. [PubMed: 26046440]
12. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N^6-adenosine)-methyltransferase. RNA. 1997; 3:1233–1247. [PubMed: 9409616]
13. Liu J, et al. A METTL3–METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. Nat Chem Biol. 2014; 10:93–95. [PubMed: 24316715]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

14. Ping XL, et al. Mammalian WTAP is a regulatory subunit of the RNA N^6-methyladenosine methyltransferase. Cell Res. 2014; 24:177–189. [PubMed: 24407421]
15. Schwartz S, et al. Perturbation of m^{6}A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. Cell Rep. 2014; 8:284–296. [PubMed: 24981863]
16. Zheng G, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 2013; 49:18–29. [PubMed: 23177736]
17. Jia G. N^6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 2011; 7:885–887. [PubMed: 22002720]
18. Batista PJ, et al. m^{6}A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell. 2014; 15:707–719. [PubMed: 25456834]
19. Geula S, et al. Stem cells. m^{6}A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science. 2015; 347:1002–1006. [PubMed: 25569111]
20. Dominissini D, Moshitch-Moshkovitz S, Amariglio N, Rechavi G. Transcriptome-wide mapping of N^6-methyladenosine by m^{6}A-seq. Methods Enzymol. 2015; 560:131–147. [PubMed: 26253969]
21. Dominissini D, et al. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. Nature. 2012; 485:201–206. [PubMed: 22575960]
22. Liu N, et al. Probing N^6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA. 2013; 19:1848–1856. [PubMed: 24141618]
23. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. Mapping of N^6-methyladenosine residues in bovine prolactin mRNA. Proc Natl Acad Sci USA. 1984; 81:5667–5671. [PubMed: 6592581]
24. Neff AT, Lee JY, Wilusz J, Tian B, Wilusz CJ. Global analysis reveals multiple pathways for unique regulation of mRNA decay in induced pluripotent stem cells. Genome Res. 2012; 22:1457–1467. [PubMed: 22534399]
25. Linder B, et al. Single-nucleotide-resolution mapping of m^{6}A and m^{6}Am throughout the transcriptome. Nat Methods. 2015; 12:767–772. [PubMed: 26121403]
26. Wei C, Gershowitz A, Moss B. N^6, O2′-dimethyladenosine a novel methylated ribonucleoside next to the 5′ terminal of animal cell and virus mRNAs. Nature. 1975; 257:251–253. [PubMed: 1161029]
27. Kruse S, et al. A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. Sci Rep. 2011; 1:126. [PubMed: 22355643]
28. Keith JM, Ensinger MJ, Moss B. HeLa cell RNA (2′-O-methyladenosine-N^{6-})-methyltransferase specific for the capped 5′-end of messenger RNA. J Biol Chem. 1978; 253:5033–5039. [PubMed: 670176]
29. Robins MJ, Naik SR, Lee AS. Nucleic acid related compounds. 12. The facile and high-yield stannous chloride catalyzed monomethylation of the cis-glycol system of nucleosides by diazomethane. J Org Chem. 1974; 39:1891–1899. [PubMed: 4853388]
30. Zhao X, et al. FTO-dependent demethylation of N^6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. 2014; 24:1403–1419. [PubMed: 25412662]
31. Liu N, et al. N^6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature. 2015; 518:560–564. [PubMed: 25719671]
32. Spitale RC, et al. Structural imprints *in vivo* decode RNA regulatory mechanisms. Nature. 2015; 519:486–490. [PubMed: 25799993]
33. Shen S, et al. MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-seq data. Nucleic Acids Res. 2012; 40:e61. [PubMed: 22266656]
34. Shen S, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. Proc Natl Acad Sci USA. 2014; 111:E5593–E5601. [PubMed: 25480548]
35. Masamha CP, et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature. 2014; 510:412–416. [PubMed: 24814343]
36. Ke S, et al. A majority of m^{6}A residues are in the last exons, allowing the potential for 3′ UTR regulation. Genes Dev. 2015; 29:2037–2053. [PubMed: 26404942]
37. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. Mol Cell. 2011; 43:853–866. [PubMed: 21925375]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

38. Berkovits BD, Mayr C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature. 2015; 522:363–367. [PubMed: 25896326]
39. Misra A, Ou J, Zhu LJ, Green MR. Global promotion of alternative internal exon usage by mRNA 3′ end formation factors. Mol Cell. 2015; 58:819–831. [PubMed: 25921069]
40. Shi Y, Manley JL. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev. 2015; 29:889–897. [PubMed: 25934501]
41. Nelles DA, et al. Programmable RNA tracking in live cells with CRISPR/Cas9. Cell. 2016; 165:488–496. [PubMed: 26997482]
42. Chen K, Zhao BS, He C. Nucleic acid modifications in regulation of gene expression. Cell Chem Biol. 2016; 23:74–85. [PubMed: 26933737]
43. Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. 2013; 27:2380–2396. [PubMed: 24145798]
44. Chan CT, et al. Identification of N^6,N^6-dimethyladenosine in transfer RNA from *Mycobacterium bovis* Bacille Calmette-Guérin. Molecules. 2011; 16:5168–5181. [PubMed: 21694680]
45. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\Delta\Delta CT)}$ method. Methods. 2001; 25:402–408. [PubMed: 11846609]
46. Sigova AA, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc Natl Acad Sci USA. 2013; 110:2876–2881. [PubMed: 23382218]
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-seq. Bioinformatics. 2009; 25:1105–1111. [PubMed: 19289445]
48. Trapnell C, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. [PubMed: 20436464]
49. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37:1–13. [PubMed: 19033363]
50. Müller S, et al. APADB: a database for alternative polyadenylation and microRNA regulation events. Database (Oxford). 2014; 2014:bau076. [PubMed: 25052703]
51. Roost C, et al. Structure and thermodynamics of N^6-methyladenosine in RNA: a spring-loaded base modification. J Am Chem Soc. 2015; 137:2107–2115. [PubMed: 25611135]
52. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120:15–20. [PubMed: 15652477]
53. Ule J, et al. An RNA map predicting Nova-dependent splicing regulation. Nature. 2006; 444:580–586. [PubMed: 17065982]
54. Dittmar KA, et al. Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. Mol Cell Biol. 2012; 32:1468–1482. [PubMed: 22354987]
55. Anderson ES, et al. The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. RNA. 2012; 18:1041–1049. [PubMed: 22456266]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

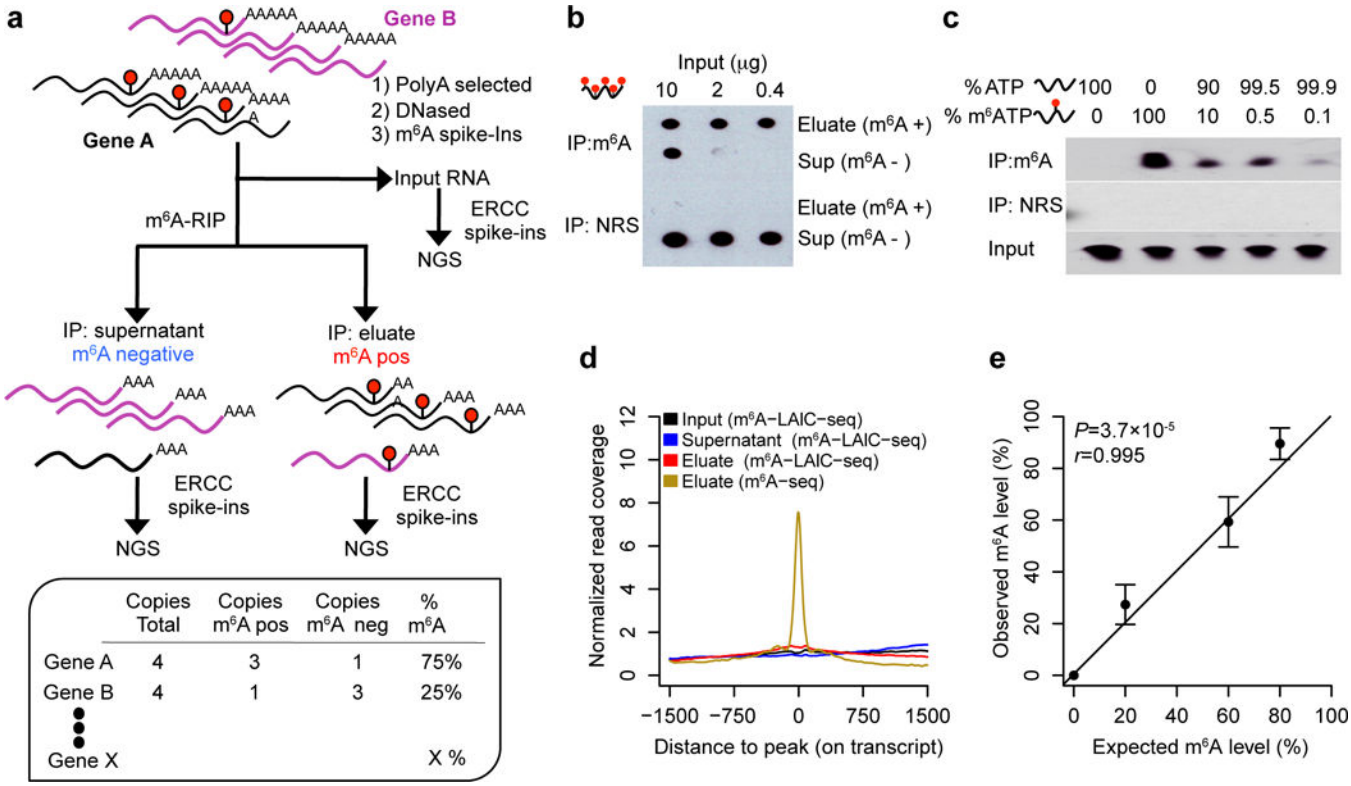

Figure 1.

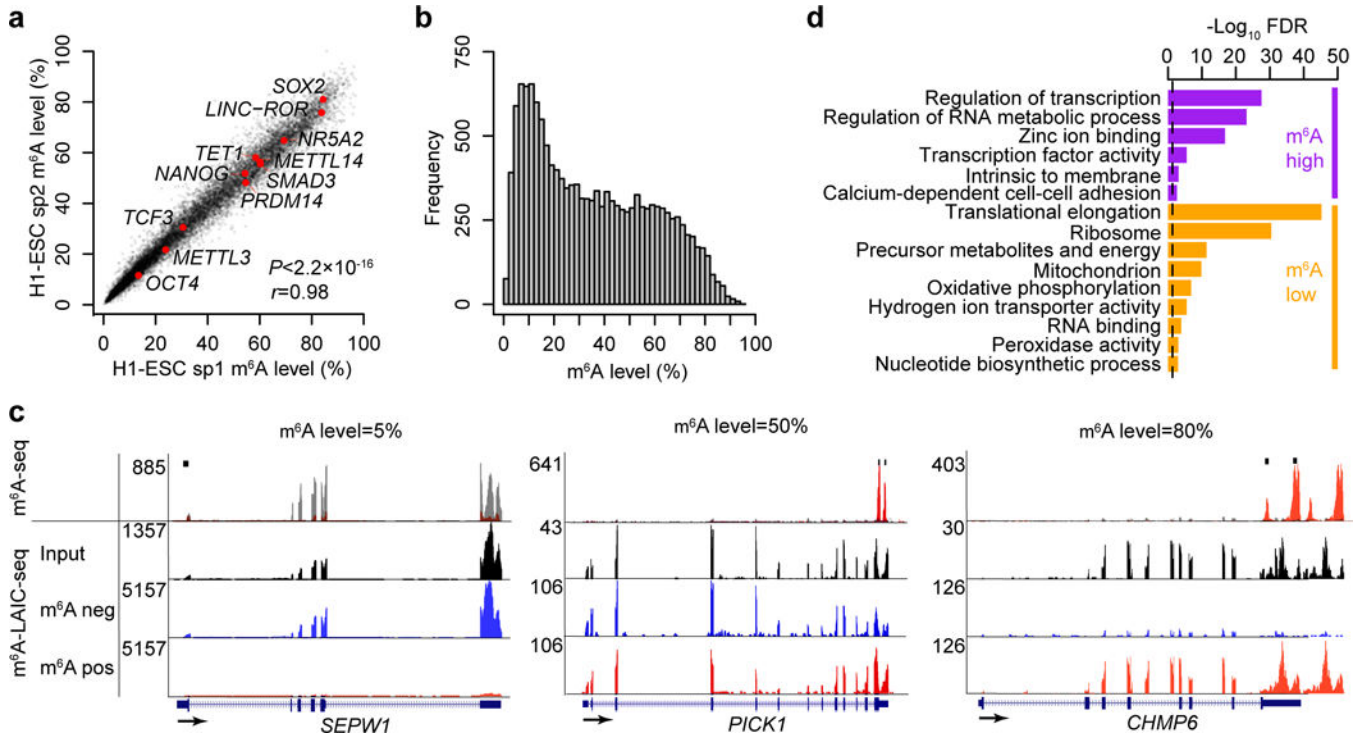

Figure 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

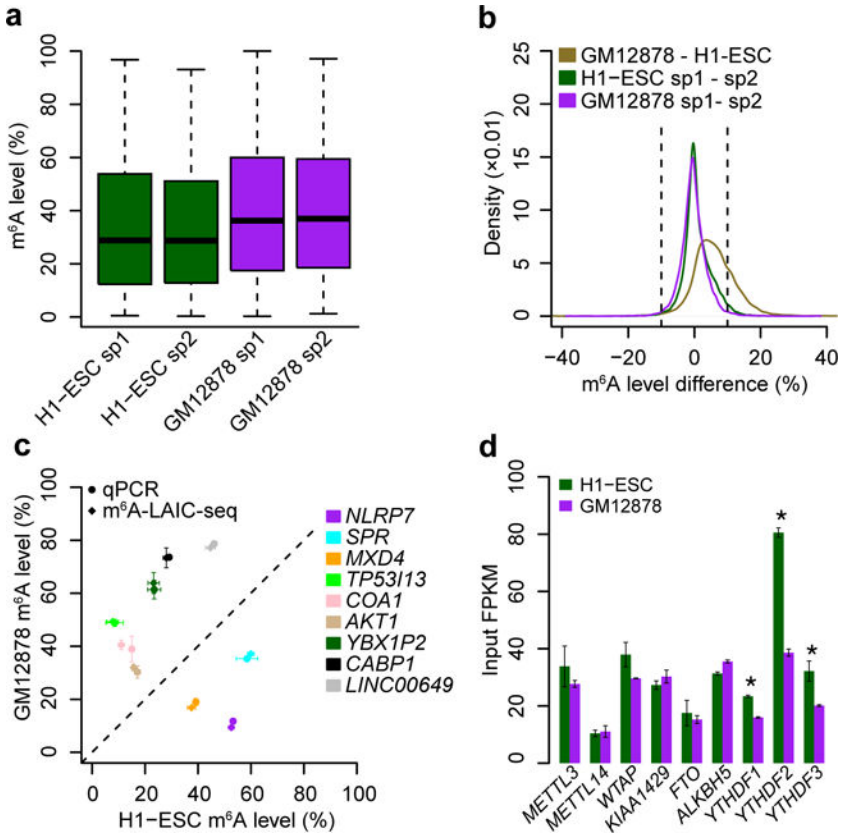

Figure 3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

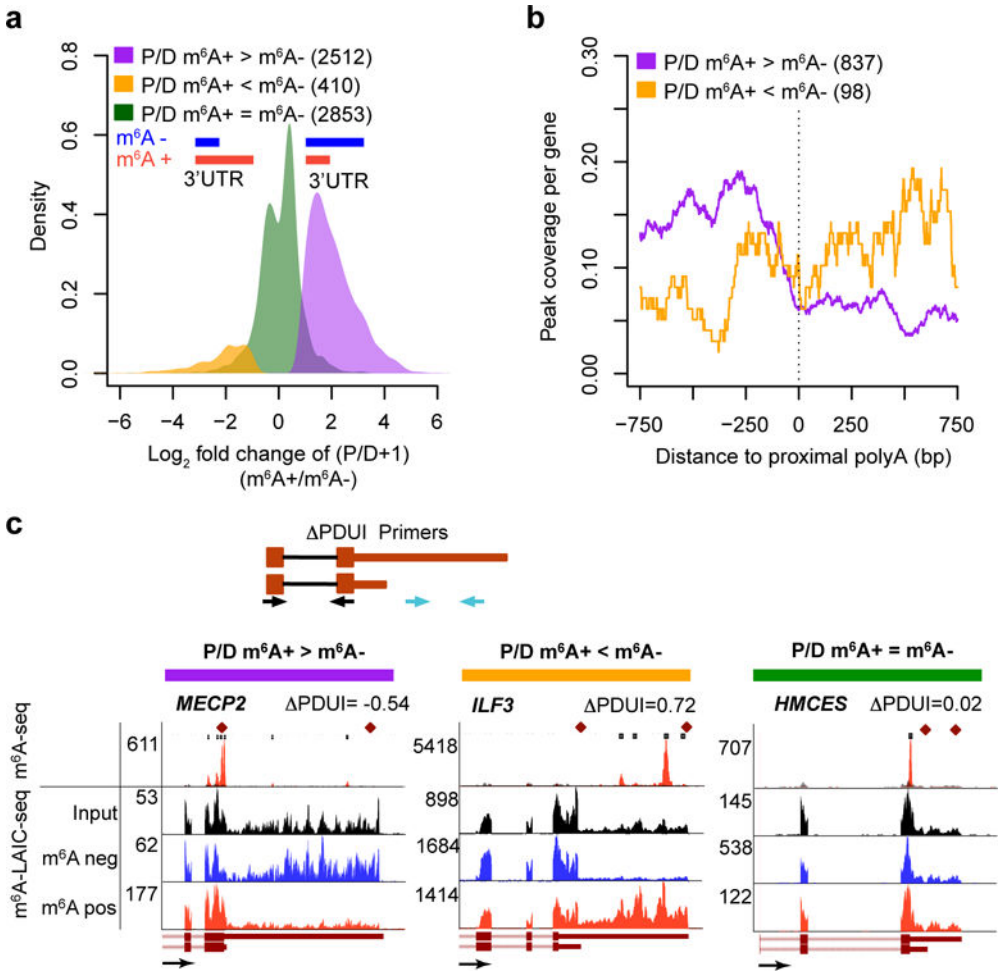

Figure 4.

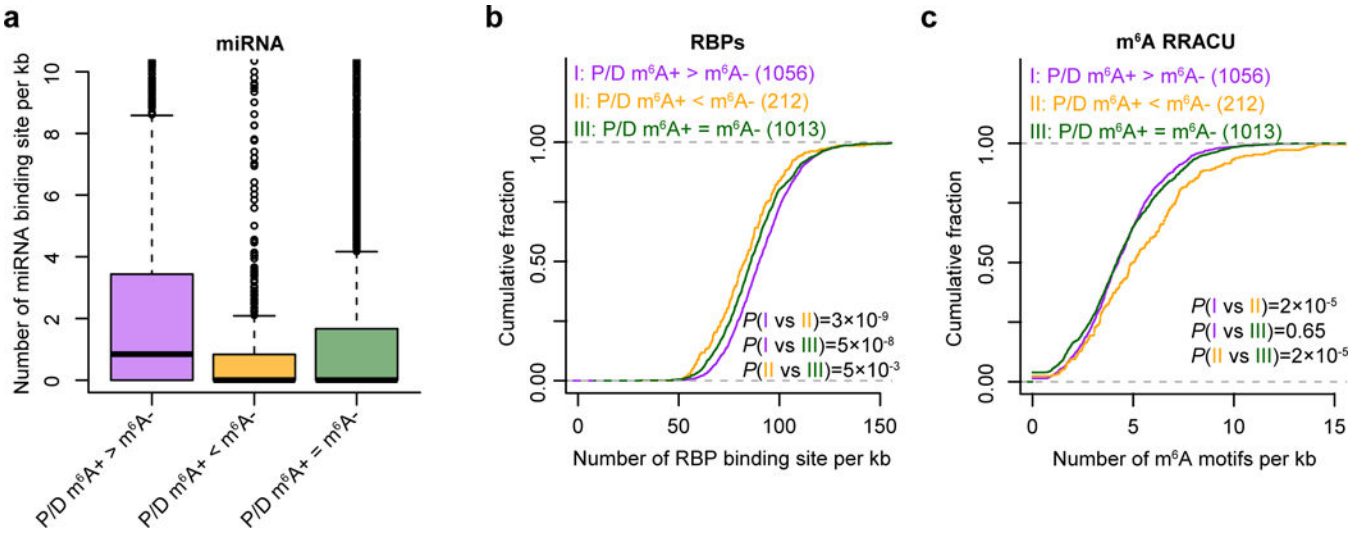

Figure 5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript