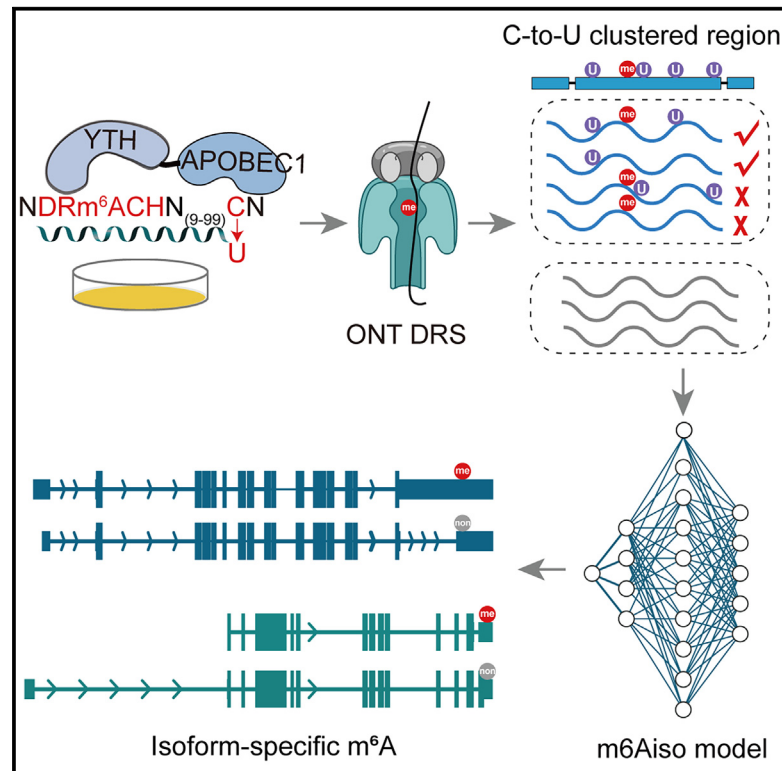# Single-molecule m^{6}A detection empowered by endogenous labeling unveils complexities across RNA isoforms

Graphical abstract

Authors

Wenbing Guo, Zhijun Ren, Xiang Huang, ..., Zhengming Feng, Jianbo He, Jinkai Wang

Correspondence

wangjk@mail.sysu.edu.cn

In brief

Guo et al. develop a deep-learning model, m6Aiso, which identifies m^{6}A on single ONT DRS reads through endogenously labeling m^{6}A on single RNAs. m6Aiso reveals isoform-specific m^{6}A on the same sites due to diverse mechanisms, elucidating the intricate dynamics and complexities of m^{6}A across RNA isoforms.

Highlights

- DART-seq C-to-U mutations 10–100 nt away can label m^{6}A on single RNAs

- m6Aiso is trained using 1,020,237 5-mer single-read m^{6}A signals of ONT DRS

- m6Aiso can accurately identify m^{6}A on single DRS reads and isoforms

- Isoform-specific m^{6}A on the same sites due to diverse mechanisms

CellPress

Molecular Cell

CellPress

Technology

Single-molecule m^{6}A detection empowered by endogenous labeling unveils complexities across RNA isoforms

Wenbing Guo,[1,2,3,7] Zhijun Ren,[1,2,3,7] Xiang Huang,[1,2,3,7] Jiayin Liu,[1,2] Jingwen Shao,[1,2] Xiaojun Ma,[1,2] Chuanchuan Wei,[1,2,3] Yixian Cun,[1,2] Jialiang He,[1,2] Jie Zhang,[1,2] Zehong Wu,[1,2] Yang Guo,[4] Zijun Zhang,[5] Zhengming Feng,[1,2] Jianbo He,[6] and Jinkai Wang[1,2,8,*]

[1]Department of Histoembryology and Cell Biology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China
[2]Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-sen University, Guangzhou 510080, China
[3]Medical College of Jiaying University, Meizhou 514031, China
[4]School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
[5]Division of Artificial Intelligence in Medicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[6]GeneMind Biosciences Company Limited, Shenzhen 518000, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: wangjk@mail.sysu.edu.cn
https://doi.org/10.1016/j.molcel.2025.01.014

SUMMARY

The landscape of N^6-methyadenosine (m^{6}A) on different RNA isoforms is still incompletely understood. Here, in HEK293T cells, we endogenously label the methylated m^{6}A sites on single Oxford Nanopore Technology (ONT) direct RNA sequencing (DRS) reads by APOBEC1-YTH-induced C-to-U mutations 10–100 nt away, obtaining 1,020,237 5-mer single-read m^{6}A signals. We then trained m6Aiso, a deep residual neural network model that accurately identifies and quantifies m^{6}A at single-read resolution. Analyzing m6Aiso-determined m^{6}A on single reads and isoforms uncovers distance-dependent linkages of m^{6}A sites along single molecules. It also uncovers specific methylation of identical m^{6}A sites on intron-retained isoforms, partly due to their differential distances to exon junctions and isoform-specific binding of TARBP2. Moreover, we find that transcription factor SMAD3 promotes m^{6}A deposition on its transcribed RNA isoforms during epithelial-mesenchymal transition, resulting in isoform-specific regulation of m^{6}A on isoforms with alternative promoters. Our study underscores the effectiveness of m6Aiso in elucidating the intricate dynamics and complexities of m^{6}A across RNA isoforms.

INTRODUCTION

N^6-methyadenosine (m^{6}A) is a prevalent and dynamic modification on mRNAs and diverse types of noncoding RNAs.[1] It is mainly catalyzed by the m^{6}A methyltransferase METTL3 at the DRACH (D = A, G or U; R = A or G; H = A, C, or U) motif on RNAs.[2] Recently developed methods, glyoxal and nitrite-mediated deamination of unmethylated adenosine (GLORI)[3] and evolved TadA-assisted N^6-methyladenosine sequencing (eTAM-seq),[4] have been proven to provide accurate identification and absolute quantification of m^{6}A at single-nucleotide resolution. Furthermore, more and more research emphasizes the intricate association between m^{6}A and the generation and metabolism of RNA isoforms.[3,5,6] Especially, recent studies have revealed that the exon-junction complex (EJC) plays an inhibitory role in the deposition of m^{6}A near splicing junctions.[7–10] These

findings strongly suggest the possibility of selective m^{6}A deposition on different mRNA isoforms, even at the same locations. However, due to the lack of methods to accurately identify m^{6}A on individual intact RNA molecules, the distribution of m^{6}A sites among various RNA isoforms remains unclear.

The direct RNA sequencing (DRS) technique developed by Oxford Nanopore Technology (ONT)[11] has been proven to be a powerful strategy for deciphering the complexities of RNA isoforms. It captures the electric current changes for each continuous five nucleotides on RNAs when they traverse the nanopore. Although numerous machine learning models, such as Epinano,[12] epitranscriptional landscape inferring from glitches of ONT signals (ELIGOS),[13] nanom6A,[14] NanoCompore,[15] TandemMod,[16] deeplearning explore nanopore m^{6}A (DENA),[17] and m6Anet,[18] have been developed based on this technology to identify m^{6}A sites at single-nucleotide resolution, training a

model to accurately identify m⁶A at single DRS reads is still of great challenge due to the lack of high-quality m⁶A-modified DRS signals on single molecules. Therefore, there is an urgent need for comprehensive and unbiased labeling of m⁶A on individual cellular RNA molecules to train a model with reliable single-read accuracy. Here, through endogenously labeling the m⁶A on single RNAs in live cells, we developed m⁶Aiso, which accurately identifies m⁶A on single reads, and observed isoform-specific m⁶A methylation driven by multiple potential mechanisms.

DESIGN

Deamination adjacent to RNA modification targets (DART-seq), which was reported to induce C-to-U mutations adjacent to m⁶A sites through the apolipoprotein B mRNA editing enzyme catalytic subunit 1 (APOBEC1)-YT521-B homology domain (YTH) fusion protein in live cells, was previously recognized as a method for identifying m⁶A at single-nucleotide resolution,[19] although its bias to certain 5-mers was also reported.[20] While the C-to-U mutations can endogenously label m⁶A on the same molecules, the mutations beside the m⁶A sites must distort the signals of m⁶A-modified 5-mers, which constitute the smallest unit for the ONT signal. In this study, we revealed that APOBEC1-YTH is also capable of inducing clustered C-to-U mutations away from the m⁶A sites without altering the 5-mers encompassing GLORI-identified m⁶A sites on the same molecules. Therefore, we reasoned that in the ONT DRS data of APOBEC1-YTH-expressed cells, the signals of annotated m⁶A sites on single reads without C-to-U mutations nearby but at reasonable distances away were undistorted m⁶A signals on single reads, based on which a deep-learning model could be trained to identify m⁶A on single DRS reads.

RESULTS

Determination of positive m⁶A signals on single DRS reads through endogenous labeling

Although the C-to-U mutations of DART-seq can label the m⁶A endogenously on the same molecules, the mutations beside m⁶A sites must distort the signals of m⁶A-modified 5-mers, which constitute the smallest unit for ONT signal. Upon reanalyzing the C-to-U mutations reported by the original DART-seq paper,[19] we realized that although 36,835 (91.4%) C-to-U mutations are immediately preceded by an A, only 16,223 (44.0%) of which were within the DRACH m⁶A motif and 1,273 (7.8%) were immediately preceded by GLORI-identified m⁶A sites in HEK293T (Figure S1A). Nevertheless, 1,708 (7.1%) and 6,815 (28.3%) of the C-to-U mutations outside DRACH motifs were located less than 10 and 10–100 nt away from the GLORI-identified m⁶A sites in the same cell line, respectively (Figure S1B), suggesting that APOBEC1-YTH induces a significant fraction of C-to-U mutations within 100 nt of m⁶A sites. Similar results were observed in HEK293T cells transfected with the copGFP-tagged APOBEC1-YTH fusion protein, which allows for the enrichment of transfected cells by flow cytometry (Figures S1C–S1F). In addition, the C-to-U mutations across the 3 replicates showed high reproducibility (Pearson's r of mu-

tation rate > 0.95) (Figures S1G–S1I). Thus, these C-to-U mutations away from m⁶A sites may be used to label the m⁶A without interfering with the electric current of m⁶A-methylated 5-mers.

To efficiently label the m⁶A endogenously on single RNAs, we performed ONT DRS deeply on the mRNAs of HEK293T cells with efficient translation of copGFP-tagged APOBEC1-YTH (5,503,279 reads) and copGFP-tagged empty vector (4,932,654 reads). After filtering out the known SNPs as well as the mutations that could be induced by APOBEC1 alone[19] or those mutations that occurred in the empty vector, we identified 384,096 C-to-U mutations from DRS long reads. These mutations were distributed in a manner resembling the distribution of m⁶A sites with an enrichment near-stop codons (Figure S2A), although exhibiting a slight 3′ bias, which is consistent with the general bias observed in ONT reads.[21] We found 166,507 (43.4%) of the C-to-U mutations, including those not within m⁶A motifs, were clustered within 100 nt. Furthermore, the clustered C-to-U mutations, rather than the non-clustered ones, were enriched near the stop codons of mRNAs (Figures S2B and S2C), suggesting that most of the non-clustered C-to-U mutations are not induced by the binding of the YTH domain but rather by the incidental contact of APOBEC1 with random RNAs.

Because the C-to-U mutations close to the m⁶A sites will distort the 5-mer signals of m⁶A, we did observe significantly different current signals between m⁶A sites with C-to-U mutations occurring within 10 nt and those 10–100 nt away (Figures S2D–S2F), while the 5-mer motif compositions of them are similar, suggesting the m⁶A sites with C-to-U mutations 10–100 nt away can well represent the m⁶A sites in all contexts (Figure S2G). Thereby, we proceeded to determine the proper single-read m⁶A signals at the GLORI-identified m⁶A sites by requiring the presence of at least one clustered C-to-U mutation 10–100 nt away, but no mutations occurring less than 10 nt away from the m⁶A sites on the same reads (Figure 1A). Assuming that the C-to-U mutations were caused by methylation of the nearest m⁶A sites on the same RNAs, we determined 2,275,985 m⁶A signals on 1,210,818 single DRS reads at 20,297 unique GLORI-identified m⁶A sites in HEK293T.[3] Meanwhile, from the same DRS dataset, we also identified 18,630,201 single-read-level unmodified signals at the DRACH motifs located at least 20 nt away from any m⁶A sites annotated by m⁶A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP),[22] GLORI,[3] m⁶A-selective allyl chemical labeling and sequencing (m6A-SAC-seq),[23] and m⁶A-crosslinking-exonuclease-sequencing (m6ACE-seq),[24] and outside m⁶A peaks of m6A-seq[25] in HEK293T cells.

Development of a deep-learning model to predict the m⁶A methylation states on single reads

Based on the m⁶A-modified and -unmodified signals at single reads, we aimed to develop a deep-learning model to predict the methylation states of DRACH motifs on single reads. To determine an optimized deep-learning model, we tried 1D-residual neural network (ResNet), 2D-ResNet, and self-attention to train 12 deep-learning models with the input of local RNA sequences and their local electric current signal features of 5-mer, 7-mer, 9-mer, and 11-mer, respectively.

Because the correct m⁶A sites responsible for the C-to-U mutations were sometimes ambiguous, especially when multiple
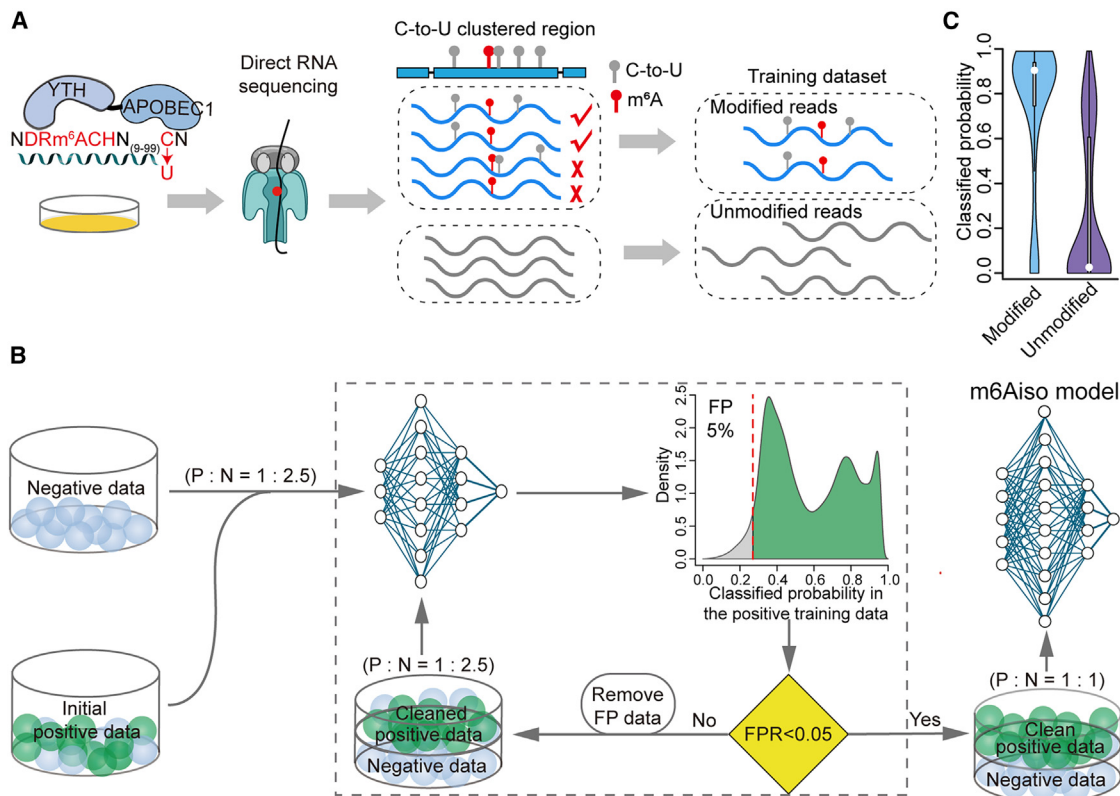
Molecular Cell
Technology

CellPress

Figure 1. Endogenously labeling of m^{6}A on single reads and assessment of m6Aiso at single-molecule level
(A) Flowcharts illustrating the classification of modified and unmodified native reads.
(B) Schematic illustrating a semi-supervised learning strategy for filtering out false positive signals. P:N, positive data: negative data; FP, false positive.
(C) The distribution of predicted modification probabilities using the m6Aiso model in the independent test dataset.
See also Figures S1–S4.

GLORI-identified m^{6}A sites were located near the mutations, we recognized that there should be some false positive signals in the positive training data. To address this issue, we employed a semi-supervised learning strategy[26] to train the deep-learning models, respectively (Figure 1B). During this training process, the initial models were trained using all the training data. We then discarded the 5% samples with the lowest predicted m^{6}A probabilities from the positive dataset and updated the training data to retrain the models, respectively. The updating and re-training processes were iteratively repeated and terminated till the false positive rate (FPR) < 0.05, which is about to converge in the positive training data (Figure S2H). We then trained 12 final models using the corresponding clean single-read m^{6}A signals. To evaluate the performance of these models, we used an independent flow cell of ONT DRS reads obtained from HEK293T cells infected with an empty vector. This test dataset comprises all the 171,791 single-read signals at 3,265 GLORI[3]-determined, extremely highly methylated (level > 0.95) m^{6}A sites in HEK293T cells as positive data as well as 858,955 unmodified single-read signals as the negative data. Through comparing the performances of the 12 final models, we found all models achieved similar area under curves (AUCs) of receiver operating characteristic (ROC) curves, ranging from 0.81 to 0.92 (Figures S3A, S3C, and S3E); however, only the 2D-ResNet model using signals of a

7-mer sequence, which comprised three 5-mer sliding windows from 1 nt upstream to 1 nt downstream of DRACH, exhibited an outstanding AUC of precision-recall (PR) curves at 0.69, which was significantly higher than the second-best at 0.36 (Figures S3B, S3D, and S3F). Moreover, the 7-mer 2D-ResNet model achieved an even higher ROC AUC of 0.9 and a PR AUC of 0.87 (Figures S3G and S3H) based on the synthetic in vitro transcribed data.[12] We then selected the 7-mer 2D-ResNet model as our final model and named it as m6Aiso (Figure S3I). As shown in Figure 1C, the modification probabilities determined by m6Aiso were concentrated at the expected extremes in both unmodified and modified reads in the test dataset (Figure 1C).

Based on m6Aiso, we ultimately preserved 1,020,237 (44.8%) clean, single-read m^{6}A-modified signals at 19,493 unique m^{6}A sites for final training. Although not a gold standard, we still found that 63.0% of these clean signals were also supported by m6Anet.[18] To further evaluate the accuracy of single-read m^{6}A signals, we used a 7-mer motif, AGGACUU, as an example due to its distinct 5-mer (GGACU) mean current signal between m^{6}A and A in this motif context (Figure S4A). As shown in Figure S4A, the distribution of the mean current signal of initial training data (epoch = 0) resembles a merged distribution of m^{6}A and A; however, after removing the 5% of training data with the lowest m^{6}A probabilities in each epoch over 27 times,

the distribution of mean current signal in the final training data (epoch = 26) closely resembles that of pure m⁶A signal, suggesting the data cleaning process effectively eliminated false positives in each epoch and resulted in clean training data. Additionally, the process continually reduced the proportion of single-read signals derived from lowly methylated m⁶A sites and those away from the C-to-U mutations (Figures S4B and S4C).

The 19,493 unique m⁶A sites that derived from the clean positive single-read m⁶A signals showed similar compositions of m⁶A motifs and distributions of m⁶A level with GLORI-identified m⁶A sites in HEK293T cells[3] (Figures S4D and S4E), suggesting it has the capacity of predicting all types of m⁶A sites. Notably, 256,080 (25.1%) of these single-read signals originated from lowly methylated m⁶A sites (level < 0.25), which were largely omitted in the training data of previous methods (Figure S4F).

m6Aiso accurately detects and quantifies both highly and lowly methylated m⁶A sites

Using m6Aiso, we then identified 23,146 m⁶A sites with strong enrichment near-stop codons in HEK293T cells infected with an empty vector (Figure 2A). These sites resembled the 5-mers profile of GLORI-identified m⁶A sites (Figure 2B). Moreover, in contrast to m6Anet,[18] which primarily captures the m⁶A sites with levels greater than 0.4, m6Aiso-identified m⁶A sites in HEK293T cells displayed a "saddle-shaped" distribution of modified levels, with 8,073 (34.9%) of them below 0.25, which is highly consistent with GLORI (Figure 2C). Similar results were observed using the published DRS data[27,28] from six human cell lines (A549, K562, HCT116, HEK293T, HeLa, and MCF7) (Figures S5A–S5C).

To further evaluate the accuracy of m6Aiso in identifying m⁶A sites, we took advantage of the previously known m⁶A sites in HEK293T cells identified by GLORI,[3] miCLIP,[22] m6A-SAC-seq,[23] and m6ACE-seq.[24] We found that 17,941 (77.5%) of the m6Aiso-identified m⁶A sites could be validated by at least one of the four experimental methods, while 5,205 (22.5%) sites were m6Aiso-specific (Figure 2D). In contrast to the shared m⁶A sites, these m6Aiso-specific sites exhibited very low m⁶A stoichiometry, with 73.0% of them having levels below 0.25 (Figure S5D). Furthermore, when considering only the 12,313 sites with modified levels greater than 0.4 to mimic the level distribution of m6Anet-identified m⁶A sites, 86.2% of m6Aiso-determined m⁶A sites were also detected by GLORI in HEK293T, which was comparable with m6Anet but substantially higher than DENA and nanom6A (Figure S5E). On the other hand, although methylated lowly, these 5,205 m6Aiso-specific sites still displayed a strong enrichment near the stop codons (Figure S5F), and 2,866 (55.1%) were annotated within the m6A-Atlas (v2.0) database.[29] Even for the remaining 2,339 m⁶A sites, which had not been previously identified, their motif distribution resembled the GLORI-determined lowly methylated m⁶A sites (Figure S5G), and 6 of 10 representative selected sites were still sensitive to depletion or overexpression of m⁶A methyltransferase or to in vitro fat mass and obesity-associated protein (FTO) treatment, as assessed using single-base elongation-and ligation-based qPCR amplification (SELECT) method[30] (Figures S5H–S5K).

We then tested the accuracy of m6Aiso in quantifying the m⁶A levels by comparing its outputs with those measured by GLORI in diverse human cell lines, including HEK293T, HeLa, and A549, as well as mouse embryonic stem cells (mESCs). We found a strong correlation between the m⁶A levels determined by m6Aiso and those determined by GLORI in all of these cell lines (Pearson's r = 0.86–0.90, Figures 2E, 2F, S6A, and S6B), with m6Aiso outperforming m6Anet, nanom6A, and DENA (Figures S6C–S6E). Additionally, a strong correlation of the measured m⁶A levels was also observed between m6Aiso and eTAM-seq[4] in HeLa (Pearson's r = 0.85, Figure S6F) and mESC cells (Pearson's r = 0.86, Figure S6G), respectively. Notably, m⁶A sites with levels below 0.25 were also highly consistent with GLORI or eTAM-seq, suggesting m6Aiso has sufficient power and accuracy to quantify the lowly methylated m⁶A sites. When we classified the m⁶A sites by the 5-mers, we found the correlation of m⁶A levels between m6Aiso and GLORI in HEK293T cells was particularly robust for the common motifs (Figure S7) and outperformed both m6Anet[18] and mAFiA.[31] We further found that m6Aiso measurements of m⁶A levels in HeLa cells were generally consistent with the six m⁶A sites measured by locus-specific extension of annealed DNA probes targeting m⁶A and sequencing (LEAD-m6A-seq)[32] (Figure S6H). Moreover, we observed the m⁶A levels determined by m6Aiso were correlated with the modeled m⁶A probability predicted by iM6A[33] in HEK293T, HeLa, A549, and mESCs, respectively (Figures 2G and S6I–S6K). Notably, the m6Aiso-specific m⁶A sites also displayed significantly lower iM6A scores than the shared sites (Figure 2H).

We next evaluated the accuracy of m6Aiso by using the previously published ONT DRS data from METTL3-knockout (KO) and wild-type (WT) in HEK293T cells[27] and mESCs,[34] respectively. Firstly, we found a high consistency of m6Aiso-determined m⁶A levels (Pearson's r = 0.98) between the two replicates of WT HEK293T cells (Figure S8A). We then observed robust downregulation of m6Aiso-determined m⁶A levels in METTL3-KO for the vast majority of sites (Figures 2I and S8B). Specifically, only 2.2% of the m⁶A sites with methylation levels > 0.25 in WT were above the diagonal. Notably, we found the m⁶A downregulation was more pronounced when determined by m6Aiso than by m6Anet (Figures S8C and S8D). Moreover, the majority of the m6Aiso-specific m⁶A sites and lowly methylated m⁶A sites (level < 0.25) were also downregulated, including those with levels < 0.1 (Figure S8E). Similar results were also observed for Mettl3-KO in mESCs (Figures 2J, S8F, and S8G). On the other hand, only 0.4% of the m6Aiso-determined downregulated (difference > 0.1) m⁶A sites upon METTL3-KO in HEK293T were previously unknown m⁶A sites (Figure S8H), while 88.2% of the sites sensitive to METTL3-knockdown (KD) in GLORI[3] (difference > 0.1) were also downregulated (difference > 0.1) in METTL3-KO based on m6Aiso (Figure S8I). Similar results were also observed in mESCs (Figures S8J and S8K).

m6Aiso performs well in Arabidopsis thaliana
To further assess the generalizability of m6Aiso, we used it to identify the m⁶A sites in three Arabidopsis thaliana samples: Col-0 (WT), VIRILIZER mutant (vir-1), and VIRILIZER
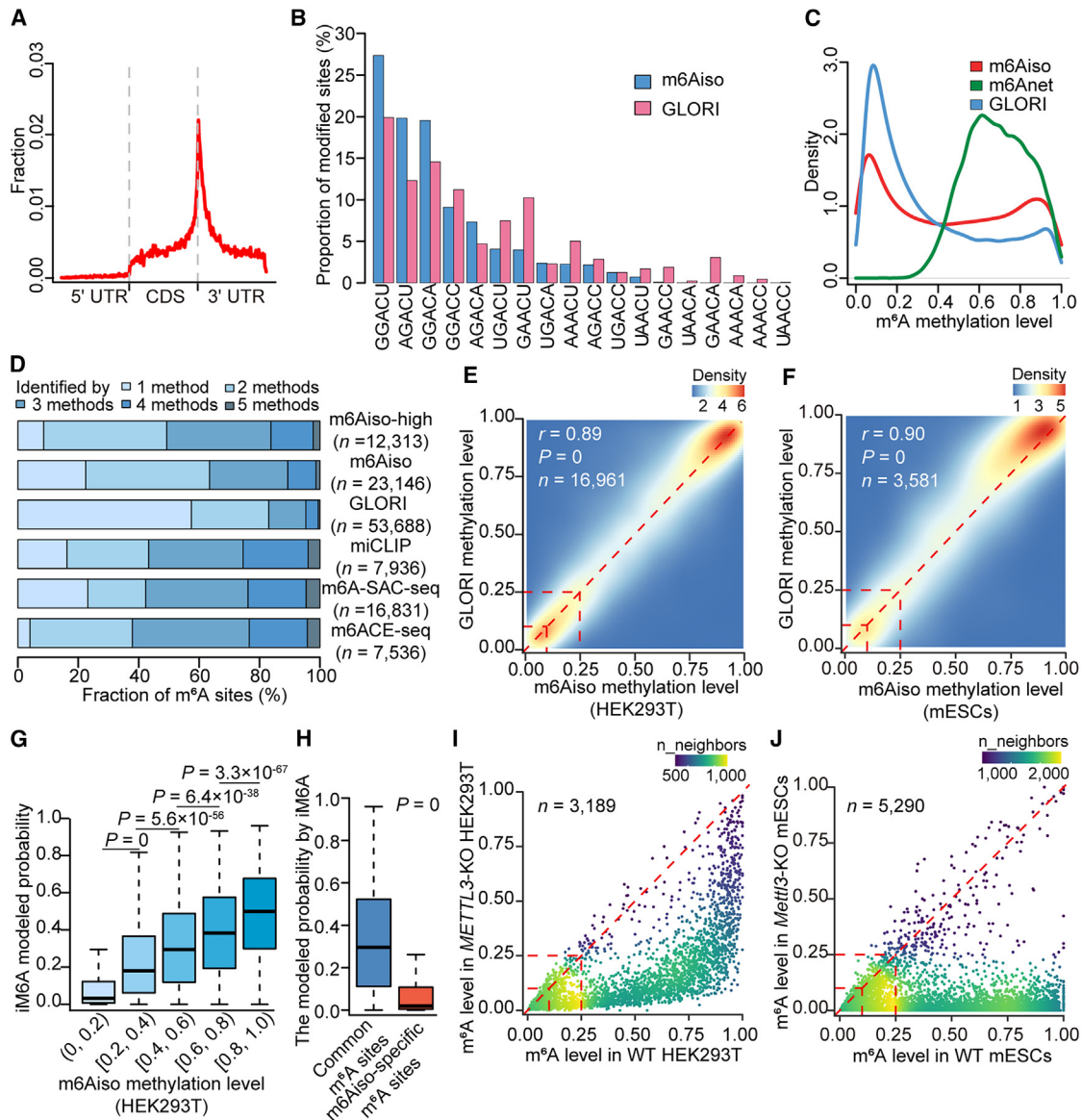
Figure 2. m6Aiso accurately detects and quantifies both highly and lowly methylated m⁶A sites

(A) Metagene profile illustrating the distribution of the modified sites captured by m6Aiso.

(B) Comparison of the percentages of modified sites determined by m6Aiso and by GLORI on the DRACH motifs.

(C) Density plot comparing the distributions of m⁶A methylation levels determined by GLORI, m6Aiso, and m6Anet.

(D) Percentages for m⁶A sites detected by different number of methods. "m6Aiso-high" denotes modified level > 0.4.

(E and F) Correlation of the m⁶A levels estimated by m6Aiso and GLORI in HEK293T cells infected with an empty vector (E) and mESCs (F), respectively. The m⁶A levels of 0.1 and 0.25 are indicated by red lines.

(G) The m⁶A levels estimated by m6Aiso agreed with the modeled probability by iM6A in HEK293T cells.

(H) Comparison of iM6A scores between m6Aiso-specific sites and common sites.

(I) Scatterplot comparing the methylation levels in *METTL3*-KO HEK293T cells with WT cells.

(J) Scatterplot comparing the methylation levels of *Mettl3*-KO mESCs with WT mESCs.

For boxplots, *p* values were calculated by two-tailed Wilcoxon rank-sum test.

See also Figures S5–S8.

complemented (VIRc).[21] A total of 4,372, 1,877, and 5,152 m⁶A sites were identified in Col-0, *vir-1*, and VIRc, respectively, with enrichment in the 3′ UTR in all samples (Figures S8L and S8M). In contrast to mammals, the most frequently methylated

5-mers in *Arabidopsis thaliana* were AGACU, AGACA, and AAACU (Figure S8N). 94.1% of the m⁶A sites in the Col-0 sample were annotated by the RNA Epitranscriptome Collection (REPIC) database[35] (Figure S8O). As expected, compared with Col-0, the
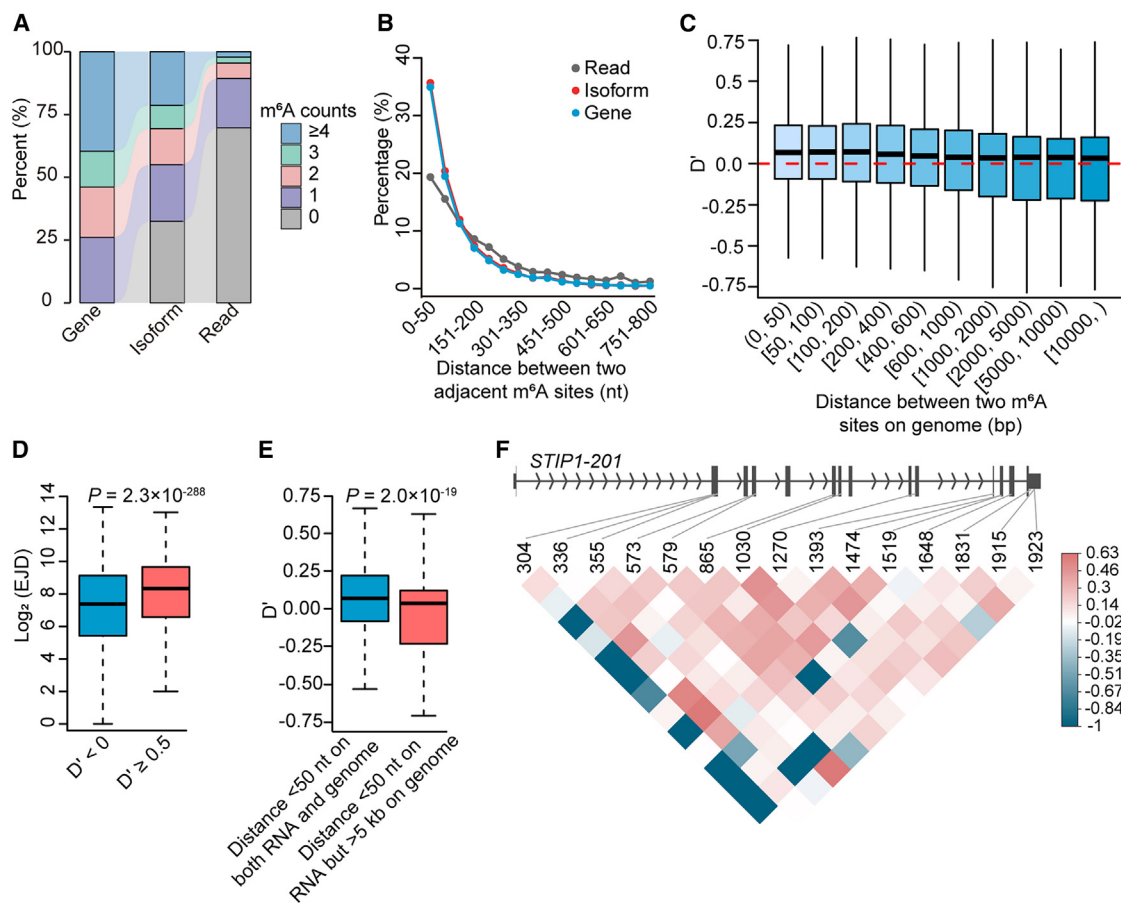
Figure 3. Linkage of m⁶A sites along the same RNA molecules

(A) Sankey diagram comparing the percentages of single reads, isoforms, and genes with different numbers of m⁶A sites.

(B) Line chart showing the percentages of m⁶A sites within various distances to adjacent m⁶A sites at the read, isoform, and gene levels.

(C) Boxplot showing the linkage (D′) of pairs of m⁶A sites with different distances on the genome.

(D) Boxplot comparing the distances of m⁶A sites with different D′ values to the exon junctions.

(E) Boxplot comparing the D′ for pairs of m⁶A sites with distance < 50 nt on RNAs but >5 kb on genes as compared with those with distance < 50 nt on both RNAs and genes.

(F) Representative example of the coordinated occurrence of m⁶A sites in *STIP1-201* transcript.

For boxplots, p values were calculated by two-tailed Wilcoxon rank-sum test.

See also Figure S9.

m6Aiso measured m⁶A levels of the vast majority of sites were remarkably decreased in both the *vir-1* and VIRc samples (Figures S8P and S8Q), indicating the reliability of m⁶Aiso in *Arabidopsis thaliana*.

Genomic distance-dependent linkage of m⁶A sites along single molecules

It has been reported that a significant portion of m⁶A sites tend to be clustered in short regions.[3,22,23] However, it remains unclear whether m⁶A sites are clustered on isoforms or single molecules. Using m6Aiso, we identified 23,146 m⁶A sites on 12,019 RNA isoforms of 6,010 genes in HEK293T cells transfected with an empty vector. On average, there were 3.9 m⁶A sites per modified gene, 2.3 m⁶A sites per expressed isoform of the modified gene, and 0.5 m⁶A sites per read of the modified gene (Figures S9A–S9C); of note, in contrast to the genes and isoforms, where

73.8% and 44.9% contained at least two m⁶A sites, respectively, only 10.6% of the DRS reads had at least two methylated m⁶A sites (Figure 3A). Moreover, consistent with previous reports,[3,22,23] approximately 35% of m⁶A modifications were clustered with adjacent m⁶A sites within 50 base pairs (bp), both on the same genes and the same isoforms (Figure 3B). However, only 19.3% of the m⁶A sites were clustered within 50 bp on the same ONT reads (Figure 3B), suggesting that the m⁶A sites are unlikely to be strongly linked on the same RNA molecules.

We then employed the parameter D′, a metric commonly used to assess the linkage disequilibrium (LD) between two genetic variants within a population,[36] to test whether different m⁶A sites on the same RNA isoforms tend to be linked on the same reads or co-occur by chance. As shown in Figure 3C, we observed a relatively weak but nonnegligible linkage between m⁶A sites with genomic distances of less than 200 bp, but the linkage

Molecular Cell
Technology

CellPress

decreased gradually as the distances increased and became less recognizable when the distances exceeded 1 kb, suggesting that a single binding of methyltransferase on an RNA can sometimes methylate multiple targets in a region. Moreover, pairs of m6A sites with higher linkage tended to occur on the same long exons of transcripts with fewer exon numbers (Figures S9D–S9F). We found the pairs of highly linked m6A sites had shorter genomic distances (Figure S9G) between each other and longer exon-junction distances (EJDs) (Figure 3D), suggesting that the inhibitory role of EJC on m6A sites near exon-exon junctions likely disrupts their linkage with other m6A sites. Gene Ontology (GO) enrichment analysis showed that genes with highly linked m6A sites were enriched in RNA processing and translation (Figure S9H). Consistent with the previous report that m6A can be deposited co-transcriptionally on chromatin RNAs,[37–39] we found that pairs of m6A sites with distances < 50 bp on RNAs but > 5 kb on the genome had significantly smaller D′ values than pairs of sites with distances < 50 bp on both RNAs and the genome (Figure 3E). As exemplified in Figure 3F, a subset of the 15 m6A sites on *STIP1-201* isoform are linked as a block with strong D′ for certain pairs of m6A sites. Therefore, the m6A sites are partially linked along single RNA molecules in a manner that depends on the distances on the genome rather than on RNAs.

m6Aiso reveals differential methylation of identical m6A sites on different isoforms due to EJC and TARBP2

To test whether m6Aiso can accurately quantify m6A on each RNA isoform, we calculated the m6A levels of each isoform as the mean level of all m6A sites. We found the average mutation rates of APOBEC-YTH-induced C-to-U mutations on individual isoforms had a significant positive correlation with the m6Aiso other than m6Anet-determined m6A levels of isoforms in HEK293T cells (Figures 4A and S10A). Moreover, the m6Aiso-determined isoform m6A levels had a significant negative correlation with the expression of isoforms (Figure S10B). In addition, consistent with previous reports that EJCs suppress the m6A deposition within 200 bp of exon boundaries,[7–10] we also observed lowly methylation of m6A within 200 bp of exon boundaries on the isoforms (Figure 4B).

We were then interested in whether the same m6A sites at identical genomic locations can be methylated differentially on different isoforms. By comparing the m6A levels of the same m6A sites among different RNA isoforms, we determined 287 (unique, $n = 147$) and 222 (unique, $n = 77$) sites that were methylated significantly higher and lower in specific isoforms than the combination of other isoforms of the same genes (Figure 4C). As compared with the isoforms with isoform-specific lowly methylated m6A sites, the isoforms with isoform-specific highly methylated m6A sites were significantly enriched in those isoforms with retained introns ($p = 0.0034$, Figure 4D). Furthermore, we observed the m6A sites on intron-retained isoforms tended to be located farther away from the exon junctions than in other transcripts ($p = 5.4 \times 10^{-30}$, Figure 4E). To test whether the inhibitory role of EJC on m6A results in differential methylation of m6A sites on the alternatively spliced isoforms, we determined the m6A sites with alternative EJDs on different isoforms. Compared with the levels of these sites on isoforms

with EJD < 100 bp, we found 33 and 7 unique m6A sites were methylated higher and lower on isoforms with EJD > 200 bp, respectively (Figure 4F). These results suggest the inhibitory role of EJC can account for at least a part of isoform-specific m6A methylation. As exemplified in Figures 4G and 4H, an m6A site in methyl-CpG binding domain protein 1 (*MBD1*) was methylated higher on the isoform with a retained intron (*MBD1–224*) than the other isoform with the intron spliced (*MBD1–201*) based on m6Aiso. Using the SELECT method,[30] we then found that specifically knockdown *MBD1–224* using the small interfering RNA (siRNA) targeting the retained intron could significantly downregulate the m6A level at this site in *MBD1*, suggesting *MBD1–224* is methylated higher than *MBD1–201* at this m6A site (Figures 4I and S10C). Similar results were also observed on the isoforms of the *SMUG1*, *C11orf24*, and *WRAP73* genes (Figures 4H, 4J, and S10D–S10G).

Besides EJC, TARBP2 subunit of RISC loading complex (TARBP2), which binds the intron regions of preRNAs, can recruit the methyltransferase complex (MTC) and promote the m6A deposition on the flanking exons to facilitate intron retention and RNA decay.[40] Interestingly, we found the levels of m6A sites on the flanking exons of TARBP2-bound introns were significantly higher than those not flanking TARBP2-bound introns on intron-retained isoforms (Figure 4K), whereas it was not significant on protein-coding isoforms (Figure S10H). Additionally, the 53 m6A sites flanking the TARBP2-bound introns were also significantly higher on intron-retained isoforms than the same sites on protein-coding isoforms (Figure S10I). These results suggest that isoform-specific m6A methylation may also result from the specific binding of RNA-binding proteins on RNA isoforms.

We further applied m6Aiso to A549 and MCF7, HCT116 cell lines to investigate whether changes in m6A site levels were associated with isoform expression. In A549 cells, we identified 283 and 70 m6A sites that exhibited significantly higher and lower methylation compared with MCF7 cells, respectively (Figure S10J), and only 6.1% and 6.0% of the significantly higher and lower modified isoforms happened to occur on isoform with changes of expression fractions, respectively (Figure S10K). Similar results were also observed in comparisons between A549 and HCT116 cells (Figures S10L and S10M), as well as between MCF7 and HCT116 (Figures S10N and S10O). These results suggest that the majority of cell-specific methylation of m6A sites represents a different layer of RNA processing from RNA splicing and expression.

m6Aiso reveals isoform-specific m6A changes during epithelial-mesenchymal transition

Recent studies have unveiled the pivotal roles of m6A in epithelial-mesenchymal transition (EMT),[41–43] a process that is also critically regulated by alternative RNA splicing.[44] To investigate whether m6A is dynamically regulated in an isoform-specific manner, we performed ONT DRS to examine the m6A changes on specific RNA isoforms in HeLa cells upon transforming growth factor β (TGF-β)-induced EMT. First of all, we confirmed the EMT was successfully induced through western blot and RNA sequencing (RNA-seq)-based expression analysis (Figures S11A–S11C). Through identification of the modified
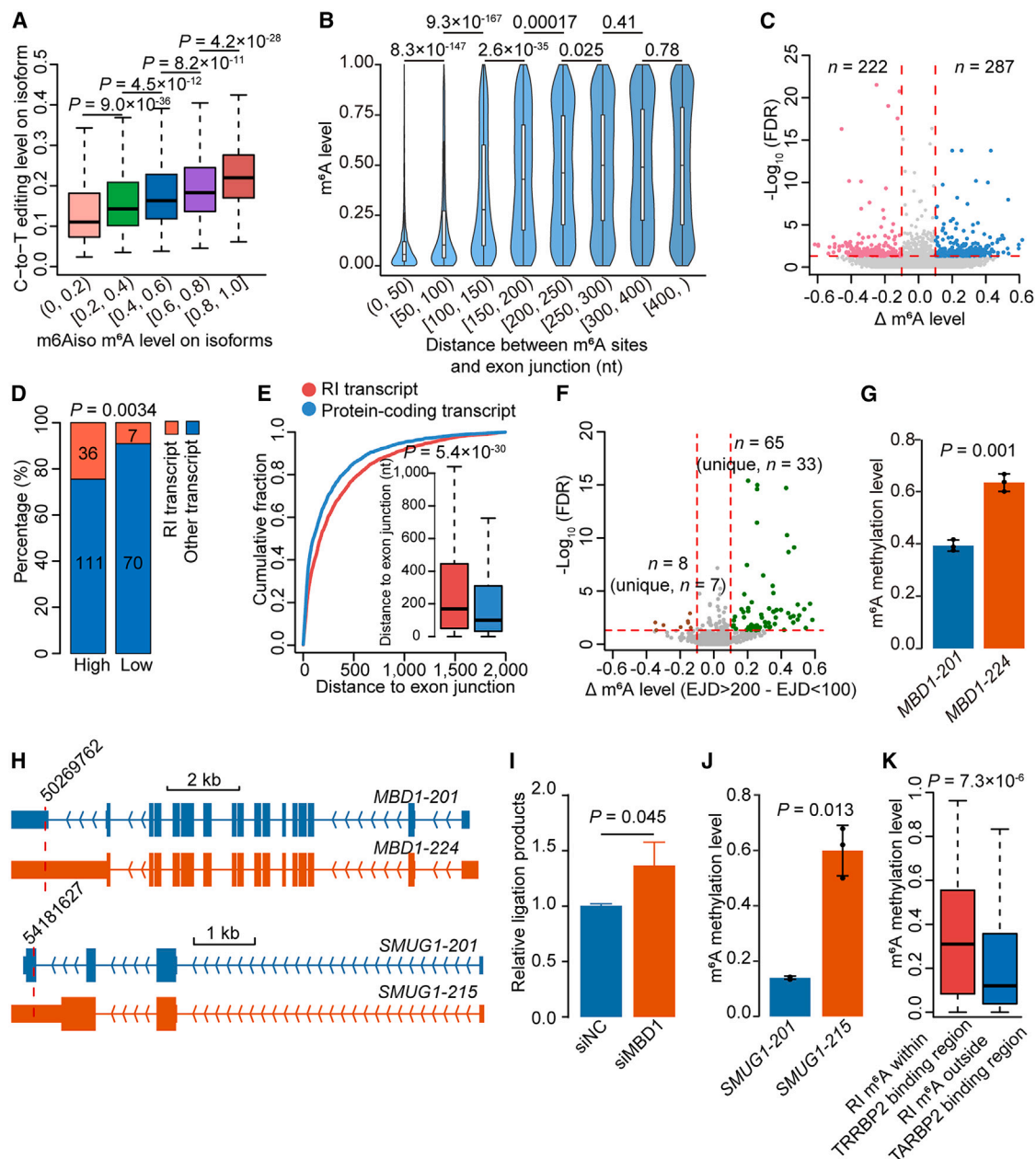
Figure 4. m6Aiso-identified m⁶A sites on different isoforms of the same genes

(A) Correlation of isoform m⁶A levels determined by m6Aiso and APOBEC1-YTH-induced C-to-U mutations.

(B) Violin plot depicting the correlation between methylation levels of m⁶A sites and their proximity to the exon junctions.

(C) Volcano plot depicting the differential methylation of single-nucleotide m⁶A sites on specific isoforms as compared with the m⁶A level of the identical sites on the combination of all other isoforms in the same genes.

(D) Bar plot depicting the number of retained-intron transcripts associated with isoform-specific, lowly methylated, and highly methylated m⁶A sites. p value was calculated by one-tailed Fisher's exact test.

(E) Plot comparing the cumulative fractions of EJDs for identical m⁶A sites on retained-intron isoforms and their corresponding protein-coding isoforms.

(F) Volcano plot showing the differential methylation at identical m⁶A sites on different isoforms with EJDs > 200 nt and <100 nt.

(G) Bar plot illustrating the differential m⁶A methylation levels at identical m⁶A sites on two isoforms of *MBD1*. p value was calculated by two-tailed t test. $n = 3$.

(H) Structures of two isoforms of *MBD1* and two isoforms of *SMUG1*. Red lines indicate the locations of m⁶A sites.

(I) SELECT assay showing m⁶A level changes at the site on *MBD1* with siRNA targeting the retained intron. p value was calculated by two-tailed t test. $n = 3$.

(J) Bar plot illustrating the differential m⁶A methylation levels at identical m⁶A sites on two isoforms of *SMUG1*. $n = 3$.

(K) Comparison of the levels of m⁶A sites between on the flanking exons of TARBP2-bound introns and those not flanking TARBP2-bound introns on intron-retained isoforms.

For boxplots and volcano plots, p values were calculated by a two-tailed Wilcoxon test.
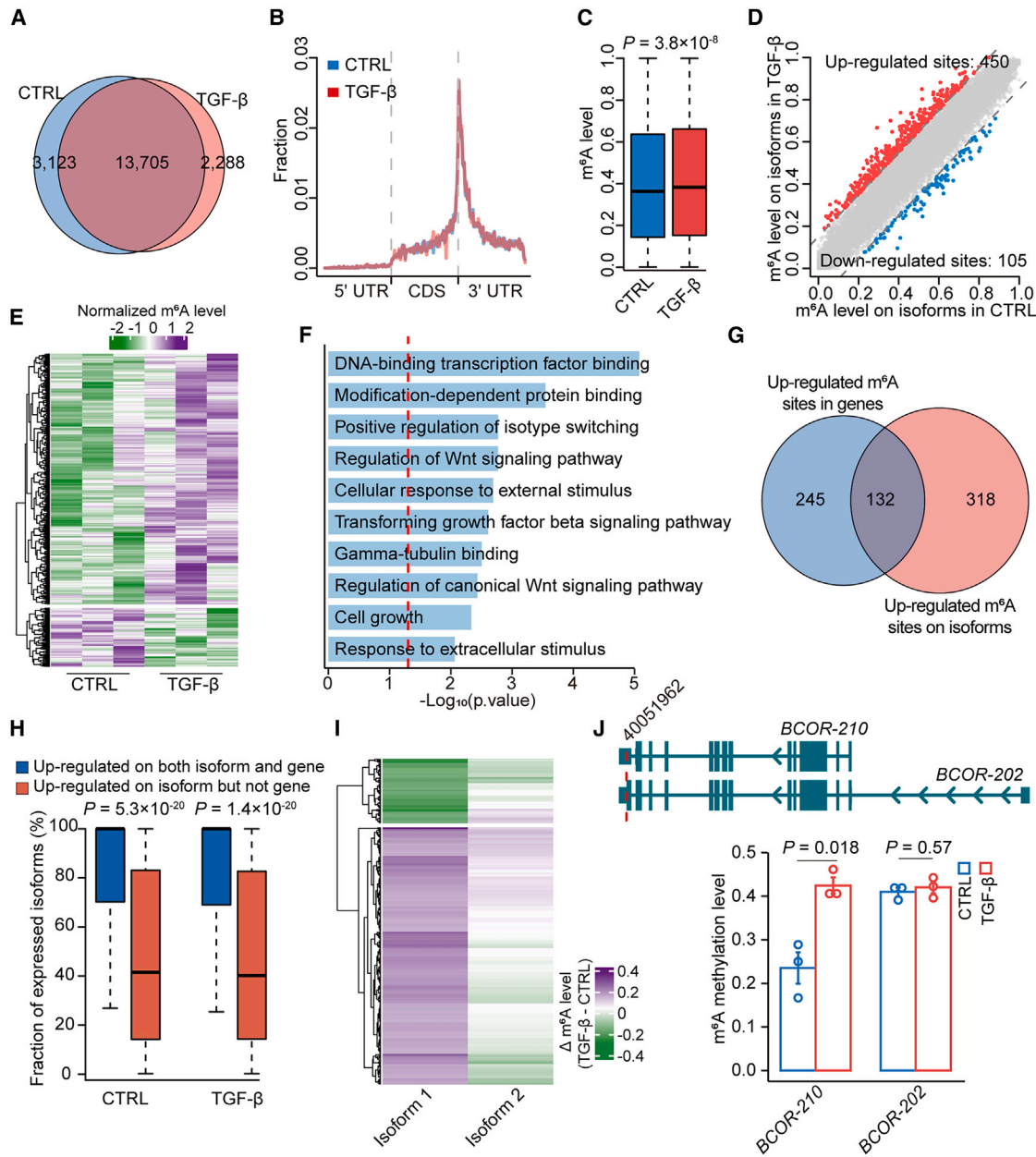
See also Figure S10 and Table S1.

Figure 5. Dynamical regulation of m⁶A in an isoform-specific manner during TGF-β-induced EMT

(A) Venn diagram comparing the number of m⁶A sites detected by m6Aiso in control and TGF-β-induced cells.

(B) Metagene profiles depicting the distributions of m⁶A sites in control and TGF-β-induced cells.

(C) Boxplot comparing the m⁶A levels between control and TGF-β-induced cells.

(D) Scatterplot depicting the differential m⁶A sites on isoforms between control and TGF-β-induced cells.

(E) Heatmap showing the methylation levels of differentially methylated m⁶A sites between control and TGF-β-induced cells.

(F) GO enrichment for the isoforms with upregulated m⁶A sites.

(G) Venn diagram comparing the upregulated m⁶A sites at the gene and isoform levels.

(H) Boxplot comparing the expression proportions of specific isoforms out of all isoforms of the same genes for the isoforms with upregulated m⁶A sites at both the isoform and gene levels with the isoforms with upregulated m⁶A sites at the isoform level but not the gene level.

(I) Heatmap comparing the changes of methylation levels of identical m⁶A sites on two different isoforms upon TGF-β induction. Isoform 1 denotes the isoforms with differentially methylated m⁶A sites, while isoform 2 denotes the randomly selected other isoforms originated from the same genes as isoform 1.

(J) Representative example of m⁶A sites in *BCOR* that exhibit changes to different directions on their two distinct isoforms during TGF-β induction. Red lines in the structures of two isoforms of *BCOR* indicate the locations of m⁶A sites. $n = 3$.

For boxplots, p values were calculated by two-tailed Wilcoxon rank-sum test.

See also Figure S11.

adenosines on individual single DRS reads using m6Aiso, we identified 16,828 and 15,993 unique m^{6}A sites in the control and TGF-β samples, respectively (Figure 5A). The m^{6}A levels were highly consistent across the three replicates in both conditions (Figures S11D–S11I). These m^{6}A sites were strongly enriched near-stop codons, with 83.4% of the sites overlapping between control and TGF-β samples (Figures 5A and 5B). Furthermore, we observed a significant global increase of m^{6}A levels in TGF-β samples ($p = 3.8 \times 10^{-8}$, Figure 5C), as confirmed by the results of m6A-seq (Figure S11J). Consistently, we determined 450 upregulated m^{6}A sites on 430 unique RNA isoforms and 105 downregulated m^{6}A sites on 103 RNA isoforms upon TGF-β induction of EMT (Figures 5D and 5E). The genes with upregulated m^{6}A on isoforms were significantly enriched in EMT-related pathways, such as the ''TGF-β signaling pathway'' and "regulation of Wnt signaling pathway" (Figure 5F).

To test whether the m^{6}A changes in EMT are isoform-specific, we compared the differential m^{6}A sites determined at gene and isoform levels. Of note, 318 (70.7%) of the 450 upregulated sites on isoforms were overlooked at the gene level comparison (Figure 5G). In contrast to the shared differential m^{6}A sites at both gene and isoform levels, the differentially methylated sites exclusively determined at isoform level were strongly enriched in those minor isoforms that expressed lower than the major isoforms of the same genes, suggesting the isoform-based m^{6}A analyses are critical in deciphering the dynamics of m^{6}A on minor isoforms (Figure 5H). Similarly, 89 (84.8%) of the 105 downregulated sites on isoforms were also overlooked at the gene level (Figure S11K). For the m^{6}A sites that were only differentially methylated on certain RNA isoforms other than genes, we found the randomly selected RNA isoforms of the same genes did not show significant m^{6}A changes at the same m^{6}A sites (Figure 5I), suggesting that these m^{6}A sites are differentially methylated upon TGF-β induction in an isoform-specific manner. For example, the m^{6}A sites in the BCL6 corepressor (*BCOR)*, which are critical for tumor suppression,[45] remained unchanged in the isoform *BCOR-202* but were upregulated in the isoform *BCOR-210* (Figure 5J). These findings suggest isoform-specific dynamic changes of m^{6}A during EMT.

SMAD3 selectively promotes m^{6}A deposition on isoforms with alternative promoters

A previous study has reported that upon stimulation with TGF-β signaling in hESCs, the transcription factors SMAD2/3 undergo phosphorylation and recruit m^{6}A MTC to facilitate m^{6}A methylation.[46] Indeed, we found the promoter regions of the isoforms with upregulated m^{6}A were significantly enriched in the motifs of SMAD family member 3 (SMAD3) (Figure 6A). We further found SMAD3 could interact with METTL3, METTL14, and WT1-associated protein (WTAP) in an RNA-independent manner (Figure 6B). Consistently, m6Aiso revealed that the isoforms with upregulated m^{6}A tend to have increased expression (Figure S11L). These results suggest that TGF-β treatment can promote m^{6}A on the RNA isoforms transcribed by SMAD3, which is activated upon TGF-β treatment.

Of note, we found both the isoforms with upregulated and downregulated m^{6}A sites were significantly enriched in isoforms with alternative promoters (Figures 5J, 6C, and 6D). By

comparing with the reference isoforms of the same genes, we found the promoters of the isoforms with upregulated m^{6}A sites were enriched in the motif of SMAD3 (Figure 6E). Interestingly, the upregulated m^{6}A sites on the isoforms with alternative promoters were mostly located in the 3′ UTRs and near-stop codons (Figures 6F and S11M), which is consistent with the previous report that SMAD3 promoted m^{6}A peaks upon stimulation with TGF-β signaling in hESCs that are near-stop codons.[46,47] These results suggest that transcription factor can selectively promote the m^{6}A on isoforms using specific promoters, resulting in isoform-specific regulation of m^{6}A upon stimulations and isoform-specific m^{6}A even at the same m^{6}A sites. On the other hand, since the short-read-based m^{6}A detection methods, such as GLORI, can only measure the differential m^{6}A sites located in the common 3′ UTRs without knowing the specific 5′ region of each isoform, this discovery further highlights the necessity of using long-read sequencing in studying the dynamics of m^{6}A.

DISCUSSION

Endogenous labeling is critical for training models at single-read resolution

We found m6Aiso exhibited a reliable ability to detect m^{6}A at single-read resolution, even for the lowly methylated m^{6}A sites. This capability is largely attributed to the abundance of m^{6}A signals on single reads, which provides a comprehensive representation of endogenous m^{6}A sites and their stoichiometries. Although APOBEC1-YTH induces C-to-U mutations that are biased to certain 5-mers,[20] our clean training data have very similar 5-mer compositions and level distribution with GLORI (Figures S4D and S4E), and this is mostly due to the usage of mutations within ~100 nt regions to indicate the methylation status of GLORI-determined m^{6}A sites. On the other hand, a critical step in the development of m6Aiso is cleaning up the initial positive single-read m^{6}A signals using a semi-supervised learning strategy.[26] The success of data cleaning process suggests that the initial training using the initial training set with false positives is still able to capture the correct features, at least for the unmethylated signals. By mimicking the initial training data with different portions of false positives using all the single-read signals from the m^{6}A sites with different ranges of GLORI-determined m^{6}A levels in control HEK293T cells, we found the training with data cleaning process had sufficient power to generate well-performed models, provided that the positive training data have more than 40% true positive signals (Figures S4G–S4I), while the true positive rate of the initial positive training data was 50.2% as estimated by the final m6Aiso model. In this situation, since the FPRs of endogenous labeling generated initial training set were similar for the m^{6}A sites with different levels, the lowly methylated m^{6}A sites thus can also be well-trained.

Diverse mechanisms are responsible for the isoform-specific m^{6}A methylation

In this study, we revealed the existence of isoform-specific m^{6}A modifications and dynamic changes even at identical m^{6}A sites. This finding highlights the advantages of using ONT DRS to identify the locations and dynamics of m^{6}A. Previous studies have
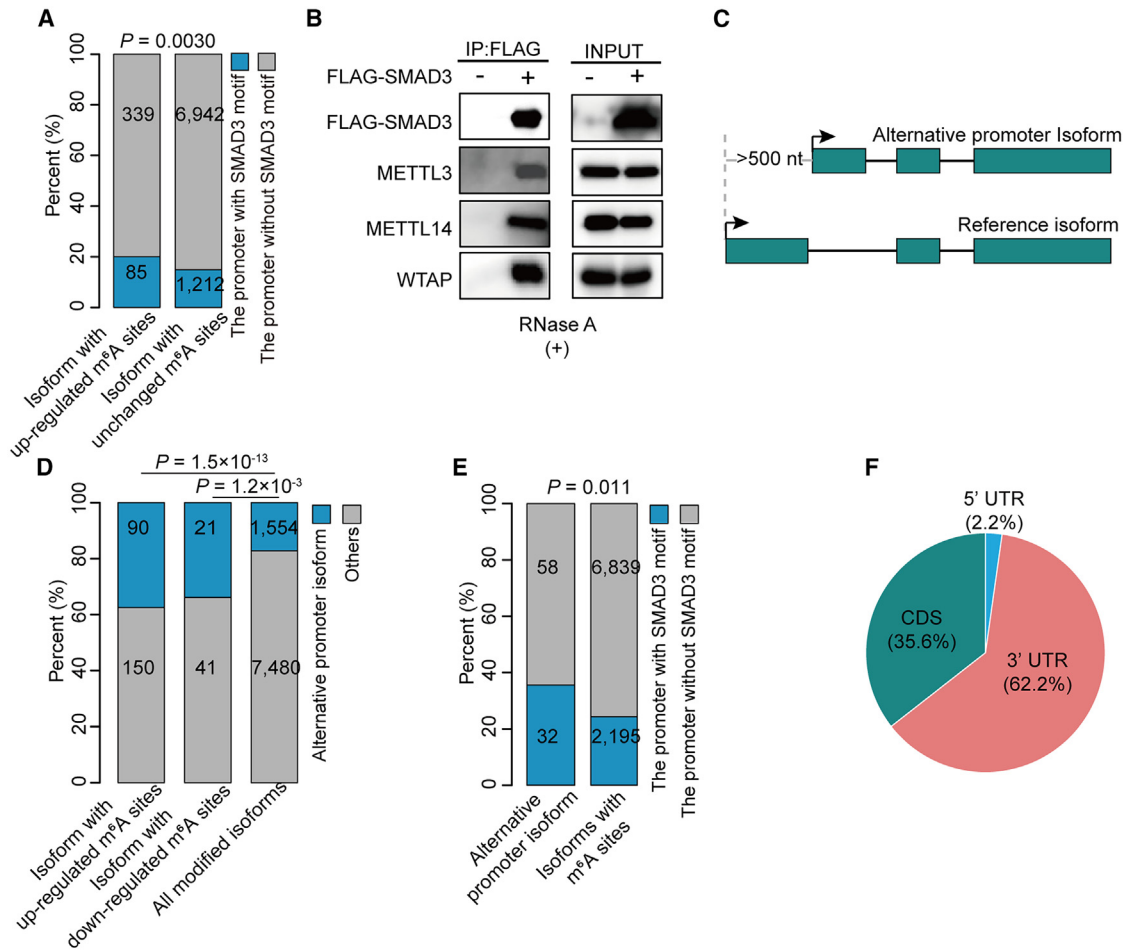
Figure 6. SMAD3 selectively promotes m[6]A deposition on isoforms with alternative promoters

(A) Stack bar plot showing the percentage of isoforms with upregulated m[6]A sites that with and without SMAD3 motif in promoters.

(B) Interactions between SMAD3 and the m[6]A methyltransferase components.

(C) Diagram illustrating the definition of alternative promoter isoforms.

(D) Bar plot displaying the number of alternative promoter isoforms containing differentially methylated m[6]A sites during EMT.

(E) Comparison of the percentages of alternative promoter isoforms with SMAD3 motif in promoters between the alternative promoter isoforms with upregulated m[6]A sites and all isoforms with m[6]A sites.

(F) Pie chart displaying the distribution of upregulated m[6]A sites within alternative promoter isoforms.

For bar plots, p values were calculated using one-tailed Fisher's exact test.

reported the inhibitory role of EJC on m[6]A deposition, as well as correlations of m[6]A and exon length or exon number across different genes.[7–10,48] These findings convergently suggest that alternative splicing can lead to alternative m[6]A deposition on different isoforms. Besides, many studies have reported the role of m[6]A in regulating alternative splicing,[5,48–50] it may also result in the preferred residence of m[6]A on specific isoforms. On the other hand, previous studies have reported the wide-spread existence of RNA-binding proteins that act as m[6]A *trans* factors to mediate the cell-specific m[6]A methylation through recruiting m[6]A MTC to the loci with the RBP binding.[51] These *trans* factors may bind to specific RNA isoforms and facilitate the m[6]A deposition on them. Additionally, these m[6]A *trans* factors include several canonical RNA splicing factors, such as SRSF7,[52] RBFOX2,[53] and TARBP2.[40] SRSF7 can interact and colocalize

with METTL3, METT14, and WTAP to facilitate the deposition of m[6]A near its binding sites.[52] It is possible that the splicing factors mediate the complex crosstalk between alternative splicing and m[6]A modification. Similarly, transcription factors can also promote the m[6]A deposition on their transcribed RNAs and result in selective regulation of m[6]A on isoforms using different promoters. Fortunately, we have provided a powerful tool m6Aiso to decipher the detailed mechanisms that mediate the isoform-specific m[6]A modification.

Limitations of the study

First, since ONT DRS is currently limited to sequencing the poly(A)+ RNAs, the non-poly(A) RNAs, such as some long non-coding RNAs (lncRNAs) and pre-mRNAs, cannot be measured using m6Aiso. Second, as the signal differences between m[6]A

and A are not uniformly robust for all 5-mers, the performance of m6Aiso for some 5-mers requires further improvement. Third, predicting the non-DRACH m⁶A sites, such as the targets of METTL16, remains challenging due to the limited availability of training data.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jinkai Wang (wangjk@mail.sysu.edu.cn).

Materials availability

All materials generated in this study are available from the lead contact upon request.

Data and code availability

The DRS sequencing and RNA sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) database under accession number SRA: PRJNA1044456. The original imaging data are available at Mendeley Data (https://doi.org/10.17632/gcypkw7jc4.1). These data are publicly available as of the date of publication. Published datasets utilized in this study are listed in the key resources table.

The code of the m6Aiso is available on GitHub (https://github.com/SYSU-Wang-LAB/m6Aiso) or Zenodo (https://doi.org/10.5281/zenodo.14650083).

Any additional information required to reanalyze the data reported in this study is available from the lead contact upon request.

AUTHOR CONTRIBUTIONS

J.W. conceived and supervised the research and revised the paper; W.G. analyzed DRS data and draft the manuscript; Z.R. analyzed NGS DART-seq data and trained the deep-learning model with the help of W.G., Y.G., and Z.Z.; X.H. performed experiments with the help of J.L., J.S., X.M., Y.C., J.Z., and Z.F.; C.W., Z.W., and J.H. provided critical technical supports.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell culture
  - Antibodies and reagents
- METHOD DETAILS
  - Validation of m⁶A changes using SELECT
  - GLORI-seq
  - Co-immunoprecipitation
  - RNA sequencing and data preprocessing
  - Endogenously labeling of m⁶A on single reads and generation of training data
  - m6Aiso model and its learning parameters
  - Model training and evaluation
  - Comparison between m6Aiso and NGS-based experimental methods
  - Comparison of m6Aiso with other machine learning models
  - Performance of m6Aiso on the synthetic Curlcakes dataset
  - Calculation of m⁶A methylation and expression levels for isoforms
  - Analysis of wild-type and *METTL3*-KO samples
  - Identification of m⁶A sites across cell lines
  - Co-occurrence analyses of m⁶A sites
  - Analysis of m⁶A sites on isoforms of single genes
  - Identification of dynamic m⁶A sites during the EMT process
  - Motif enrichment analyses for the promoter regions
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2025.01.014.

REFERENCES

1. Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187–1200. https://doi.org/10.1016/j.cell.2017.05.045.

2. Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95. https://doi.org/10.1038/nchembio.1432.

3. Liu, C., Sun, H., Yi, Y., Shen, W., Li, K., Xiao, Y., Li, F., Li, Y., Hou, Y., Lu, B., et al. (2023). Absolute quantification of single-base m⁶A methylation in the mammalian transcriptome using GLORI. Nat. Biotechnol. *41*, 355–366. https://doi.org/10.1038/s41587-022-01487-9.

4. Xiao, Y.L., Liu, S., Ge, R., Wu, Y., He, C., Chen, M., and Tang, W. (2023). Transcriptome-wide profiling and quantification of N⁶-methyladenosine by enzyme-assisted adenosine deamination. Nat. Biotechnol. *41*, 993–1003. https://doi.org/10.1038/s41587-022-01587-6.

5. Haussmann, I.U., Bodi, Z., Sanchez-Moran, E., Mongan, N.P., Archer, N., Fray, R.G., and Soller, M. (2016). m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. Nature *540*, 301–304. https://doi.org/10.1038/nature20577.

6. Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P., et al. (2016). m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. Nat. Methods *13*, 692–698. https://doi.org/10.1038/nmeth.3898.

7. He, P.C., Wei, J., Dou, X., Harada, B.T., Zhang, Z., Ge, R., Liu, C., Zhang, L.S., Yu, X., Wang, S., et al. (2023). Exon architecture controls mRNA m⁶A suppression and gene expression. Science *379*, 677–682. https://doi.org/10.1126/science.abj9090.

8. Uzonyi, A., Dierks, D., Nir, R., Kwon, O.S., Toth, U., Barbosa, I., Burel, C., Brandis, A., Rossmanith, W., Le Hir, H., et al. (2023). Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. Mol. Cell *83*, 237–251.e7. https://doi.org/10.1016/j.molcel.2022.12.026.

9. Luo, Z., Ma, Q., Sun, S., Li, N., Wang, H., Ying, Z., and Ke, S. (2023). Exon-intron boundary inhibits m⁶A deposition, enabling m⁶A distribution hallmark, longer mRNA half-life and flexible protein coding. Nat. Commun. *14*, 4172. https://doi.org/10.1038/s41467-023-39897-1.

Molecular Cell
Technology

CellPress

10. Yang, X., Triboulet, R., Liu, Q., Sendinc, E., and Gregory, R.I. (2022). Exon junction complex shapes the m6A epitranscriptome. Nat. Commun. *13*, 7904. https://doi.org/10.1038/s41467-022-35643-1.

11. Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods *15*, 201–206. https://doi.org/10.1038/nmeth.4577.

12. Liu, H., Begik, O., Lucas, M.C., Ramirez, J.M., Mason, C.E., Wiener, D., Schwartz, S., Mattick, J.S., Smith, M.A., and Novoa, E.M. (2019). Accurate detection of m6A RNA modifications in native RNA sequences. Nat. Commun. *10*, 4079. https://doi.org/10.1038/s41467-019-11713-9.

13. Jenjaroenpun, P., Wongsurawat, T., Wadley, T.D., Wassenaar, T.M., Liu, J., Dai, Q., Wanchai, V., Akel, N.S., Jamshidi-Parsian, A., Franco, A.T., et al. (2021). Decoding the epitranscriptional landscape from native RNA sequences. Nucleic Acids Res. *49*, e7. https://doi.org/10.1093/nar/gkaa620.

14. Gao, Y., Liu, X., Wu, B., Wang, H., Xi, F., Kohnen, M.V., Reddy, A.S.N., and Gu, L. (2021). Quantitative profiling of N6-methyladenosine at single-base resolution in stem-differentiating xylem of Populus trichocarpa using Nanopore direct RNA sequencing. Genome Biol. *22*, 22. https://doi.org/10.1186/s13059-020-02241-7.

15. Leger, A., Amaral, P.P., Pandolfini, L., Capitanchik, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A.J., et al. (2021). RNA modifications detection by comparative Nanopore direct RNA sequencing. Nat. Commun. *12*, 7198. https://doi.org/10.1038/s41467-021-27393-3.

16. Wu, Y., Shao, W., Yan, M., Wang, Y., Xu, P., Huang, G., Li, X., Gregory, B.D., Yang, J., Wang, H., et al. (2024). Transfer learning enables identification of multiple types of RNA modifications using nanopore direct RNA sequencing. Nat. Commun. *15*, 4049. https://doi.org/10.1038/s41467-024-48437-4.

17. Qin, H., Ou, L., Gao, J., Chen, L., Wang, J.W., Hao, P., and Li, X. (2022). DENA: training an authentic neural network model using Nanopore sequencing data of Arabidopsis transcripts for detection and quantification of N6-methyladenosine on RNA. Genome Biol. *23*, 25. https://doi.org/10.1186/s13059-021-02598-3.

18. Hendra, C., Pratanwanich, P.N., Wan, Y.K., Goh, W.S.S., Thiery, A., and Göke, J. (2022). Detection of m6A from direct RNA sequencing using a multiple instance learning framework. Nat. Methods *19*, 1590–1598. https://doi.org/10.1038/s41592-022-01666-1.

19. Meyer, K.D. (2019). DART-seq: an antibody-free method for global m6A detection. Nat. Methods *16*, 1275–1280. https://doi.org/10.1038/s41592-019-0570-0.

20. Liu, J., Huang, T., Yao, J., Zhao, T., Zhang, Y., and Zhang, R. (2023). Epitranscriptomic subtyping, visualization, and denoising by global motif visualization. Nat. Commun. *14*, 5944. https://doi.org/10.1038/s41467-023-41653-4.

21. Parker, M.T., Knop, K., Sherwood, A.V., Schurch, N.J., Mackinnon, K., Gould, P.D., Hall, A.J., Barton, G.J., and Simpson, G.G. (2020). Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. eLife *9*, e49658. https://doi.org/10.7554/eLife.49658.

22. Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods *12*, 767–772. https://doi.org/10.1038/nmeth.3453.

23. Hu, L., Liu, S., Peng, Y., Ge, R., Su, R., Senevirathne, C., Harada, B.T., Dai, Q., Wei, J., Zhang, L., et al. (2022). m6A RNA modifications are measured at single-base resolution across the mammalian transcriptome. Nat. Biotechnol. *40*, 1210–1219. https://doi.org/10.1038/s41587-022-01243-z.

24. Koh, C.W.Q., Goh, Y.T., and Goh, W.S.S. (2019). Atlas of quantitative single-base-resolution N6-methyl-adenine methylomes. Nat. Commun. *10*, 5636. https://doi.org/10.1038/s41467-019-13561-z.

25. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646. https://doi.org/10.1016/j.cell.2012.05.003.

26. Lee, D.-H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML, *2*, p. 896.

27. Ying, C., Nadia, M.D., Yuk Kei, W., Harshil, P., Fei, Y., Hwee Meng, L., Christopher, H., Laura, W., Andre, S., Chelsea, S., et al. (2021). A systematic benchmark of Nanopore long read RNA sequencing for transcript level analysis in human cell lines. Preprint at bioRxiv. https://doi.org/10.1101/2021.04.21.440736.

28. Acera Mateos, P., J Sethi, A., Ravindran, A., Srivastava, A., Woodward, K., Mahmud, S., Kanchi, M., Guarnacci, M., Xu, J., W S Yuen, Z., et al. (2024). Prediction of m6A and m5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. Nat. Commun. *15*, 3899. https://doi.org/10.1038/s41467-024-47953-7.

29. Liang, Z., Ye, H., Ma, J., Wei, Z., Wang, Y., Zhang, Y., Huang, D., Song, B., Meng, J., Rigden, D.J., et al. (2024). m6A-Atlas v2.0: updated resources for unraveling the N6-methyladenosine (m6A) epitranscriptome among multiple species. Nucleic Acids Res. *52*, D194–D202. https://doi.org/10.1093/nar/gkad691.

30. Xiao, Y., Wang, Y., Tang, Q., Wei, L., Zhang, X., and Jia, G. (2018). An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N6-Methyladenosine Modification. Angew. Chem. Int. Ed. Engl. *57*, 15995–16000. https://doi.org/10.1002/anie.201807942.

31. Chan, A., Naarmann-de Vries, I.S., Scheitl, C.P.M., Höbartner, C., and Dieterich, C. (2024). Detecting m6A at single-molecular resolution via direct RNA sequencing and realistic training data. Nat. Commun. *15*, 3323. https://doi.org/10.1038/s41467-024-47661-2.

32. Wang, Y., Zhang, Z., Sepich-Poore, C., Zhang, L., Xiao, Y., and He, C. (2021). LEAD-m6A-seq for Locus-Specific Detection of N6-Methyladenosine and Quantification of Differential Methylation. Angew. Chem. Int. Ed. Engl. *60*, 873–880. https://doi.org/10.1002/anie.202007266.

33. Luo, Z., Zhang, J., Fei, J., and Ke, S. (2022). Deep learning modeling m6A deposition reveals the importance of downstream cis-element sequences. Nat. Commun. *13*, 2720. https://doi.org/10.1038/s41467-022-30209-7.

34. Zhong, Z.D., Xie, Y.Y., Chen, H.X., Lan, Y.L., Liu, X.H., Ji, J.Y., Wu, F., Jin, L., Chen, J., Mak, D.W., et al. (2023). Systematic comparison of tools used for m6A mapping from nanopore direct RNA sequencing. Nat. Commun. *14*, 1906. https://doi.org/10.1038/s41467-023-37596-5.

35. Liu, S., Zhu, A., He, C., and Chen, M. (2020). REPIC: a database for exploring the N6-methyladenosine methylome. Genome Biol. *21*, 100. https://doi.org/10.1186/s13059-020-02012-4.

36. Lewontin, R.C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics *49*, 49–67. https://doi.org/10.1093/genetics/49.1.49.

37. Xu, W., He, C., Kaye, E.G., Li, J., Mu, M., Nelson, G.M., Dong, L., Wang, J., Wu, F., Shi, Y.G., et al. (2022). Dynamic control of chromatin-associated m6A methylation regulates nascent RNA synthesis. Mol. Cell *82*, 1156–1168.e7. https://doi.org/10.1016/j.molcel.2022.02.006.

38. Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B.S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F., et al. (2019). Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. Nature *567*, 414–419. https://doi.org/10.1038/s41586-019-1016-7.

39. Li, Y., Xia, L., Tan, K., Ye, X., Zuo, Z., Li, M., Xiao, R., Wang, Z., Liu, X., Deng, M., et al. (2020). N6-Methyladenosine co-transcriptionally directs the demethylation of histone H3K9me2. Nat. Genet. *52*, 870–877. https://doi.org/10.1038/s41588-020-0677-3.

40. Fish, L., Navickas, A., Culbertson, B., Xu, Y., Nguyen, H.C.B., Zhang, S., Hochman, M., Okimoto, R., Dill, B.D., Molina, H., et al. (2019). Nuclear

TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay. Mol. Cell 75, 967–981.e9. https://doi.org/10.1016/j.molcel.2019.06.001.

41. Jeschke, J., Collignon, E., Al Wardi, C., Krayem, M., Bizet, M., Jia, Y., Garaud, S., Wimana, Z., Calonne, E., Hassabi, B., et al. (2021). Downregulation of the FTO m⁶A RNA demethylase promotes EMT-mediated progression of epithelial tumors and sensitivity to Wnt inhibitors. Nat. Cancer 2, 611–628. https://doi.org/10.1038/s43018-021-00223-7.

42. Xu, X., Zhuang, X., Yu, H., Li, P., Li, X., Lin, H., Teoh, J.P., Chen, Y., Yang, Y., Cheng, Y., et al. (2024). FSH induces EMT in ovarian cancer via ALKBH5-regulated Snail m6A demethylation. Theranostics 14, 2151–2166. https://doi.org/10.7150/thno.94161.

43. Shu, F., Liu, H., Chen, X., Liu, Y., Zhou, J., Tang, L., Cao, W., Yang, S., Long, Y., Li, R., et al. (2024). m6A Modification Promotes EMT and Metastasis of Castration-Resistant Prostate Cancer by Upregulating NFIB. Cancer Res. 84, 1947–1962. https://doi.org/10.1158/0008-5472.CAN-23-1954.

44. Zhu, Z.M., Huo, F.C., Zhang, J., Shan, H.J., and Pei, D.S. (2023). Crosstalk between m6A modification and alternative splicing during cancer progression. Clin. Transl. Med. 13, e1460. https://doi.org/10.1002/ctm2.1460.

45. Sportoletti, P., Sorcini, D., and Falini, B. (2021). BCOR gene alterations in hematologic diseases. Blood 138, 2455–2468. https://doi.org/10.1182/blood.2021010958.

46. Bertero, A., Brown, S., Madrigal, P., Osnato, A., Ortmann, D., Yiangou, L., Kadiwala, J., Hubner, N.C., de Los Mozos, I.R., Sadée, C., et al. (2018). The SMAD2/3 interactome reveals that TGFbeta controls m⁶A mRNA methylation in pluripotency. Nature 555, 256–259. https://doi.org/10.1038/nature25784.

47. Alfonso-Gonzalez, C., Legnini, I., Holec, S., Arrigoni, L., Ozbulut, H.C., Mateos, F., Koppstein, D., Rybak-Wolf, A., Bönisch, U., Rajewsky, N., et al. (2023). Sites of transcription initiation drive mRNA isoform selection. Cell 186, 2438–2455.e22. https://doi.org/10.1016/j.cell.2023.04.012.

48. Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L., Lai, W.Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell 61, 507–519. https://doi.org/10.1016/j.molcel.2016.01.012.

49. Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Z., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019). Regulation of Co-transcriptional Pre-mRNA Splicing by m(6)A through the Low-Complexity Protein hnRNPG. Mol. Cell 76, 70–81.e9. https://doi.org/10.1016/j.molcel.2019.07.005.

50. Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet. 14, e1007412. https://doi.org/10.1371/journal.pgen.1007412.

51. An, S., Huang, W., Huang, X., Cun, Y., Cheng, W., Sun, X., Ren, Z., Chen, Y., Chen, W., and Wang, J. (2020). Integrative network analysis identifies cell-specific trans regulators of m6A. Nucleic Acids Res. 48, 1715–1729. https://doi.org/10.1093/nar/gkz1206.

52. Cun, Y., An, S., Zheng, H., Lan, J., Chen, W., Luo, W., Yao, C., Li, X., Huang, X., Sun, X., et al. (2023). Specific Regulation of m⁶A by SRSF7 Promotes the Progression of Glioblastoma. Genomics Proteomics Bioinformatics 21, 707–728. https://doi.org/10.1016/j.gpb.2021.11.001.

53. Dou, X., Xiao, Y., Shen, C., Wang, K., Wu, T., Liu, C., Li, Y., Yu, X., Liu, J., Dai, Q., et al. (2023). RBFOX2 recognizes N⁶-methyladenosine to suppress transcription and block myeloid leukaemia differentiation. Nat. Cell Biol. 25, 1359–1368. https://doi.org/10.1038/s41556-023-01213-w.

54. Pratanwanich, P.N., Yao, F., Chen, Y., Koh, C.W.Q., Wan, Y.K., Hendra, C., Poon, P., Goh, Y.T., Yap, P.M.L., Chooi, J.Y., et al. (2021). Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. Nat. Biotechnol. 39, 1394–1402. https://doi.org/10.1038/s41587-021-00949-w.

55. Goodarzi, H., Zhang, S., Buss, C.G., Fish, L., Tavazoie, S., and Tavazoie, S.F. (2014). Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. Nature 513, 256–260. https://doi.org/10.1038/nature13466.

56. Gao, Y., Wang, F., Wang, R., Kutschera, E., Xu, Y., Xie, S., Wang, Y., Kadash-Edmondson, K.E., Lin, L., and Xing, Y. (2023). ESPRESSO: Robust discovery and quantification of transcript isoforms from error-prone long-read RNA-seq data. Sci. Adv. 9, eabq5072. https://doi.org/10.1126/sciadv.abq5072.

57. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16, 284–287. https://doi.org/10.1089/omi.2011.0118.

58. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

59. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

60. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

61. Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410. https://doi.org/10.1038/nmeth.4184.

62. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 3. https://doi.org/10.14806/ej.17.1.200.

63. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

64. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

65. Liu, Y., Han, R., Zhou, L., Luo, M., Zeng, L., Zhao, X., Ma, Y., Zhou, Z., and Sun, L. (2021). Comparative performance of the GenoLab M and NovaSeq 6000 sequencing platforms for transcriptome and LncRNA analysis. BMC Genomics 22, 829. https://doi.org/10.1186/s12864-021-08150-8.

66. Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 15, 707–719. https://doi.org/10.1016/j.stem.2014.09.019.

67. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260.

68. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

69. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158, 1431–1443. https://doi.org/10.1016/j.cell.2014.08.009.

Molecular Cell
Technology

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Opti-MEM	GIBCO	Cat# 31985-062
Lipofectamine 3000	Thermo Fisher Scientific	L3000015
E-cadherin	Proteintech	Cat# 20874-1-AP; RRID:AB_10697811
N-cadherin	Proteintech	Cat# 22018-1-AP; RRID:AB_2813891
GAPDH	Proteintech	Cat# 10494-1-AP; RRID:AB_2263076
High-glucose (4.5.g/L) DMEM medium	Corning	10-013-CVRC
Fetal bovine serum	Sigma-Aldrich	F0193-500ML
Ham's F-12K (Kaighn's) medium	Gibco	21127030
Opti-MEM	Gibco	31985070
RIPA Buffer (10×)	Cell Signaling	9806S
PMSF	Beyotime	ST506
6 × SDS sample buffer	TransGen Biotech	DL101-02
Tris-HCl pH=7.5	ThermoFisher	15567027
IGEPAL CA-630	Sigma-Aldrich	I8896-50ml
NaCl	Sigma-Aldrich	S5150-1L
Glycerol	Sigma-Aldrich	G5516-100ML
0.5 M EDTA, PH 8.0	ThermoFisher	AM9261
Protease inhibitor cocktail	MedchemExpress	HY-K0010
Phosphatase Inhibitor Cocktail II	MedchemExpress	HY-K0022
Protein A/G Magnetic Beads	Vazyme	PB101-01
DMSO	Sigma-Aldrich	D2650-100ml
Fragment buffer	New England Biolabs	E6150S
Glycogen	Thermo Fisher Scientific	AM9510
H_3BO_3	Sigma-Aldrich	B0394
$NaNO_2$	Sigma-Aldrich	31443
MES	Sigma-Aldrich	M3671
Glyoxal	Sigma-Aldrich	50649
Triethylamine	Sigma-Aldrich	471283
$(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$	Sigma-Aldrich	09719
α-ketoglutarate	Sigma-Aldrich	K1128
L-ascorbic acid	Sigma-Aldrich	A5960
Sodium acetate solution	Thermo Fisher Scientific	R1181
Acetic acid	Sigma-Aldrich	695092–500 mL
Deionized formamide	Sigma-Aldrich	344206-M
Trypsin	GIBCO	25200072
Blotting-Grade Blocker, nonfat dry milk	Bio-RAD	1706404
RNase A	Thermo Fisher Scientific	EN0531
AMPure XP Beads	Beckman Coulter	A63881
$MgCl_2$	Invitrogen	AM9530G
RNasin Ribonuclease Inhibitors	Promega	N2615
Recombinant human FTO protein (Active)	Abcam	ab271525
Critical commercial assays		
Ligation Sequencing Kit	Oxford Nanopore Technologies	Cat# SQK-LSK110
TruSeq Stranded mRNA Library Prep Kit	Illumina	Cat# 20020594

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pierce™ Rapid Gold BCA Protein Assay Kit	Thermo Fisher Scientific	A53226
Endo-free Plasmid Mini Kit II	Omega	D6950-02
VAHTS Universal V10 RNA-seq Library Prep Kit for MGI	Vazyme	NRM606-01
SMARTer Stranded Total RNA-Seq Kit v2	TAKARA	634413
2× Universal SYBR Green Fast qPCR Mix	ABclonal	RK21203
ABScript III RT Master Mix for qPCR with gDNA Remover	ABclonal	RK20429
FastPure Cell/Tissue Total RNA Isolation Kit V2	Vazyme	RC112-01
Lipomaster 3000 Transfection Reagent	Vazyme	TL301-01
Mut Express MultiS Fast Mutagenesis Kit V2 (Vazyme)	Vazyme	C215-01
RNA Clean & Concentrator™-5 RNA	ZYMO RESEARCH	R1013
VAHTS mRNA Capture Beads 2.0	Vazyme	N403-01
Deposited data		
Mendeley data	This paper	DOI: https://doi.org/10.17632/gcypkw7jc4.1
ONT and RNA-seq data	This paper	SRA: PRJNA1044456
GLORI data	Liu et al.[3]	GEO: GSE210563
HEK293T METTL3 KO direct RNA sequencing data	Pratanwanich et al.[54]	SRA: PRJEB40872
mESCs METTL3 KO direct RNA sequencing data	Zhong et al.[34]	GEO: GSE195618
HeLa direct RAN sequencing data	Acera Mateos et al.[28]	GEO: GSE211759
Arabidopsis thaliana direct RNA sequencing data	Parker et al.[21]	SRA: PRJEB32782
Synthetic Curlcakes data	Liu et al.[12]	GEO: GSE124309
miCLIP data	Linder et al.[22]	GEO: GSE63753
m6A-SAC-seq data	Hu et al.[23]	GEO: GSE162357
m6ACE-seq data	Koh et al.[24]	GEO: GSE124509
m6A-seq data	Meyer et al.[25]	GEO: GSE29714
DART-seq data	Meyer[19]	GEO; GSE125780
eTAM-seq data	Xiao et al.[4]	GEO: GSE211303
TARBP2 HITS-CLIP data	Goodarzi et al.[55]	GEO: GSE49648
im6A data	Luo et al.[33]	Zenodo: https://doi.org/10.5281/zenodo.4734266
Experimental models: Cell lines		
Human HEK293T cells	ATCC	CRL-3216; RRID:CVCL_0063
Human HeLa cells	ATCC	CCL-2; RRID:CVCL_0030
Human A549 cells	Cellcook	CC0202; RRID:CVCL_0023
Software and algorithms		
ESPRESSO (v1.4.0)	Gao et al.[56]	https://github.com/Xinglab/espresso
clusterProfiler (v4.3.8)	Yu et al.[57]	https://github.com/YuLab-SMU/clusterProfiler; RRID:SCR_016884
FeatureCounts (v2.0.1)	Liao et al.[58]	https://github.com/byee4/featureCounts; RRID:SCR_012919
DESeq2	Love et al.[59]	https://github.com/thelovelab/DESeq2; RRID:SCR_015687
minimap2 (v2.17)	Li[60]	https://github.com/lh3/minimap2; RRID:SCR_018550
nanopolish (v0.13.2)	Simpson et al.[61]	https://github.com/jts/nanopolish/; RRID:SCR_016157
m6Anet (v2.0.0)	Hendra et al.[18]	https://github.com/GoekeLab/m6anet; RRID:SCR_025234
DENA (v1.0)	Qin et al.[17]	https://github.com/weir12/DENA
nanom6A (v2.0)	Gao et al.[14]	https://github.com/gaoyubang/nanom6A
Cutadapter (v3.5)	Martin[62]	https://github.com/marcelm/cutadapt/; RRID:SCR_011841

Molecular Cell
Technology

CellPress

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STAR (Version 2.7.3a)	Dobin et al.[63]	https://github.com/alexdobin/STAR; RRID:SCR_004463
Samtools (v1.9)	Li et al.[64]	http://samtools.sourceforge.net/; RRID:SCR_002105
Guppy (v4.2.2)	Oxford Nanopore Technologies	https://community.nanoporetech.com/downloads; RRID:SCR_022353
R (v4.3.0)	https://www.r-project.org/	https://www.r-project.org; RRID:SCR_001905
Python (v3.6.10)	https://www.python.org/	https://www.python.org/; RRID:SCR_008394
m6Aiso	This paper	https://github.com/SYSU-Wang-LAB/m6Aiso; Zenodo: https://doi.org/10.5281/zenodo.14650083
Other		
Detailed m6Aiso model	This paper	Method S1

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell culture

HeLa and HEK293T cells were cultured with Dulbecco's modified Eagle medium (DMEM, GIBCO, Carlsbad, CA, USA) with 10% fetal bovine serum (FBS). To establish the model of DART-seq, HEK293T cells were transfected with pcDNA3.1-APOBEC1-YTH-T2A-copGFP plasmid by using Lipofiter 3.0 (Hanbio) according to the manufacturer's instructions. The copGFP-positive cells were isolated through cytometry at 48 hours post-transfection. To establish the model of cancer cells undergoing EMT, HeLa cells were cultured with Dulbecco's modified Eagle medium (DMEM, GIBCO, Carlsbad, CA, USA) without FBS for 12 hours. Then, 10 ng/ml TGF-β was added to the DMEM medium for 72 hours to induce HeLa cells undergoing EMT.

Antibodies and reagents

Anti-FLAG (1:1000, FM1804), METTL14 (1:1000, HPA038002) antibodies were purchased From Sigma, USA. Anti-METTL3 (1:1000, ab195352) and WTAP (1:1000, ab195380) were purchased from Abcam, UK. Goat anti-rabbit (1:5000, SA00001) was purchased from Proteintech, China. Mouse IgG-HRP (1:5000, 7076S) antibodies were purchased from Cell Signaling Technology, USA. GAPDH Rabbit mAb (1:3000, A19056), E-Cadherin Rabbit mAb (1:500, A20798) and N-Cadherin Rabbit mAb (1:500, A19083) were purchased from ABclonal Technology, China. Protein A/G Magnetic Beads were purchased from Vazyme Biotech, China.

METHOD DETAILS

Validation of m⁶A changes using SELECT

500 ng total RNAs from the control group or expression level normalized amount of RNAs from the knockdown group were mixed with 40 nM upstream primer, 40 nM downstream primer, 5 μM dNTP, 1.7 μl 10× rCutSmart buffer (New England Biolabs, B6004V), DEPC H_2O to the final volume 17 μl. Then, the mixture of RNA and primers was incubated at the following temperatures: 1 min at 90 °C, 1 min at 80 °C, 1 min at 70 °C, 1 min at 60 °C, 1 min at 50 °C and 6 min at 40 °C. Subsequently, a 3 μl mixture containing 0.5 U SplintR ligase (New England Biolabs, M0375), 0.01 U Bst 2.0 DNA polymerase (New England Biolabs, M0537), and 10 nmol ATP was mixed with the annealing products. The final mixture was incubated at 40 °C for 20 min, and then at 80 °C for 20 min. qPCR was then performed in LightCycle 480 II (Roche) using 2× Universal SYBR Green Fast qPCR Mix (ABclonal, RK21203). Relative SELECT products between the experimental group and control group were calculated using the $2^{-\Delta\Delta Ct}$ method. Primers used in the SELECT assays were provided in Table S1.

The FTO demethylation reaction was executed in the reaction mixture that contained 1 μg of RNA, 0.25 μg of FTO demethylase, 283 μM $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, 300 μM α-ketoglutarate (α-KG), 2 mM L-ascorbic acid, 20 U of RNasin Ribonuclease Inhibitors (N2615, Promega, USA), and 50 mM tris-HCl buffer (pH 7.5). The reaction was stopped by 40 mM EDTA after 3 hours of incubation at 30 °C. Finally, RNA was purified by RNA Clean & Concentrator-5 (R1013, zymo research, USA) according to the manual.

GLORI-seq

First, mRNAs of A549 cells were captured by using VAHTS mRNA Capture Beads 2.0 (N403-01, Vazyme) from total RNA. Subsequently, GLORI was carried out according to previously published.[3] Briefly, mRNAs were fragmented at 94 °C for 30 s. Then, guanosine of fragmented mRNAs was protected by using a protection buffer (1.32 M glyoxal, 50% DMSO) at 50 °C for 30 min. Subsequently, 10 μl saturated H_3BO_3 solution was added and incubated at 50 °C for 30min. Next, deamination buffer (750 mM $NaNO_2$,

40 mM MES buffer, and 2.08 M glyoxal) was added and incubated at 16 °C for 8 h to deaminize adenine. Finally, deprotection buffer (0.25 M TEAA and 25% deionized formamide) was added and incubated at 95 °C for 10 minutes to deprotect guanosine.

For library construction, SMARTer Stranded Total RNA-Seq Kits v2 Pico Input Mammalian (Takara, 634411) was used according to the manufacturer's protocols. Briefly, about 50 ng RNA with GLORI treatment was used to generate first-strand cDNA. Then, Illumina adapters and indexes were mixed with first-stranded cDNA to generate PCR1 products by 5 cycles of PCR. Subsequently, PCR1 products were purified twice by using 0.8×AMPure beads, then skipped the removal of the Ribosomal cDNA step and eluting cDNA with 20 μL RNase-free H₂O. The purified cDNA was amplified by 9 cycles of PCR2, and followed by purification using 1×AM-Pure beads. Finally, the library size distribution was detected by Qsep100NGS. The library sequencing was performed at GeneMind Biosciences Company (Shenzhen, China) through the SURFSeq 5000 platform on PE100 mode.[65]

Co-immunoprecipitation

HEK293T cells were washed 2 times by using PBS buffer, and were lysed by lysis buffer (25 mM Tris-HCl pH = 7.5, 0.1% IGEPAL CA-630, 250 mM NaCl, 5% glycerol, 5 mM MgCl₂) containing 1% protease inhibitor cocktail (HY-K0010, MCE, USA), 1% Phosphatase Inhibitor Cocktail II (HY-K0022, MCE, USA), and 1 mM PMSF. Then, cell lysates were sonicated. Subsequently, RNase A (EN0531, Thermo Fisher Scientific, USA) was added to lysates with 30 μg/mL and incubated at 30 °C for 30 min. Next, lysates were centrifugated at 12,000 g for 15 min at 4 °C and added IP buffer (25 mM Tris-HCl pH = 7.5, 0.1% IGEPAL CA-630, 250 mM NaCl, 5% glycerol, 5 mM EDTA) with 1% protease inhibitor cocktail, 1% Phosphatase Inhibitor Cocktail II, and 1 mM PMSF to 1 mL. Lysates were incubated with appropriate antibodies overnight at 4 °C, followed by incubating Protein A/G Magnetic Beads for 2 hours at 4 °C (PB101-02, Vazyme Biotech, China). Immunoprecipitates were washed three times with cold IP buffer, and then eluted with 2 × SDS loading buffer by boiling for 10 min. Finally, SDS-PAGE was used to detect immunoprecipitates.

RNA sequencing and data preprocessing

RNA-seq was conducted on the HEK293T cells transfected with the copGFP-tagged APOBEC1-YTH fusion protein using Illumina HiSeq 2500, and C-to-U editing sites were identified following the original bioinformatics pipeline of DART-seq.[19] Nanopore direct RNA sequencing was performed in accordance with the guidelines provided by Oxford Nanopore Technologies (Oxford, UK) using DRS kits (SQK-RNA002) and R9.4.1 flow cells. The ionic current data from each FAST5 file were subjected to base calling using Guppy v4.2.2 with the high accuracy model. Only reads exceeding the quality threshold of 7 were selected for subsequent analyses. Sequencing reads were then aligned to the GRCh38 Ensembl annotated transcripts (v91) using minimap2 (v2.17)[60] with the following parameters: 'minimap2 -ax map-ont -k14 -uf -secondary=no --MD)'. Samtools (v1.9) was utilized to filter out secondary and supplementary alignments and convert the aligned reads to the BAM format.[64] ESPRESSO (v1.4.0)[56] was used to measure the expression of transcripts by counts. To align the read signals with the corresponding transcript references, the *eventalign* module of Nanopolish's (v0.13.2) was employed with the '--scale-events' and '--signal-index' options.[61] After re-squiggling, continuous ionic current measurements from each read were segmented into 5-mer events comprised of the mean, standard deviation, and dwelling time along the corresponding transcriptome coordinates.

Endogenously labeling of m⁶A on single reads and generation of training data

For the ONT DRS reads in HEK293T cells transfected with APOBEC1-YTH and an empty vector, we used Sam2Tsv pileup (v23c0a5c) to identify the C-to-U mutations and calculate the mutation rates. To identify the APOBEC1-YTH-induced C-to-U mutations, we first filtered out the mutations in dbSNP (v151) as well as the previously reported C-to-U mutations that could be induced by APOBEC1 alone.[19] We further removed the C-to-U mutations that could be observed in the control cells with an empty vector as determined by requiring the mutation rate exceeding 0.2 out of at least 10 reads. The remaining sites with C-to-U mutations on at least 10 reads and C-to-U mutation rate exceeding 0.05 but below 0.9 were considered as the preliminary APOBEC1-YTH induced C-to-U mutations for further analyses. To identify the clustered C-to-U mutations, we initially mapped the C-to-U mutations distributed across various isoforms of a single gene to the transcript with the longest coding regions. Then, we used 100 nt sliding windows with sliding step of 50 nt along the transcript with longest coding regions for each gene to search for C-to-U mutations. Windows containing at least one C-to-U mutation sites were merged if they have at least one nucleotide overlap. Subsequently, all the C-to-U mutation sites in the merged windows containing at least 3 C-to-U mutation sites were considered as clustered C-to-U mutations and preserved for the downstream analysis (Figure S2B). For each clustered C-to-U mutation on each single read, the nearest GLORI annotated m⁶A site in DRACH motif in HEK293T cells[3] within a 100 nt distance from the mutation was initially determined as the methylated m⁶A site on this single read. These methylated m⁶A sites on single reads were determined as the modified set of m⁶A on single ONT DRS reads only if there was an absence of C-to-U mutation within a 9 nt distance from the corresponding m⁶A sites on the same reads.

To determine the unmodified set of m⁶A on single ONT DRS reads of HEK293T cells transfected with APOBEC1-YTH, we first determined the unmodified 5-mers of DRACH by excluding the sites within a distance of 20 nt from the known SNPs in dbSNP (v151) and the known single-nucleotide m⁶A sites. Specifically, the previously determined single-nucleotide m⁶A sites in HEK293T cells by NGS-based experimental methods including GLORI,[3] miCLIP,[22] m6ACE-seq,[24] m6A-SAC-seq[23] were excluded; then the peak regions of m6A-seq[25] data in HEK293T cells determined as previously described[66] were also excluded.

To generate an independent testing positive set of methylated m⁶A on single reads from the ONT DRS reads of HEK293T cells transfected with an empty vector, we selected the m⁶A sites on all the individual reads for the m⁶A sites with GLORI-determined methylation levels exceeding 0.95 in HEK293T cells.[3]

To train the models with input signals of 5-mer, 7-mer, 9-mer, and 11mer, respectively, the mean, standard deviation and dwell time of the ONT electric current signals at the 5-mers of DRACH motif (5-mer), the 5-mers located 1 nt (7-mer), 1–2 nt (9-mer), and 1–3 nt (11-mer) upstream and downstream of the DRACH motif, respectively, from the above generated modified set, unmodified set, and independent testing set were obtained for the downstream machine learning to develop m6Aiso.

m6Aiso model and its learning parameters
The overall framework of m6Aiso is depicted in Figure S3I. In brief, m6Aiso is a modified deep ResNet model with the input of RNA sequences and their local-signal features to predict the states of sequence modification. In m6Aiso (Figure S3I), the 1*N-dimension input sequences were first transformed to 4*N-dimension matrix data using one-hot encoding method; then, a convolutional neural network (CNN) layer was designed to learn the high-level sequence features by using two convolution filters (4*5*2) sliding on the whole sequences with stride 1. After that, we concatenated the output of the learned high-level features of local sequences in each filter with their corresponding signal features and then inputted the combined feature matrix (5*N-4) to a 6-layer ResNet unit, which is stacked by three ResNet blocks. The output of the first and last ResNet blocks were connected to a global maximum pooling layer and then connected to a two-layer fully-connected neural network with ReLU and sigmoid activation functions, respectively. In the first CNN layer, which accepts the input sequences, we used 2 convolution filters with size of 4*5; in the first ResNet block, we used 32 filters with size of (3×3) in all CNN layers; we used 128 filters with size of (3×3) in all CNN layers and 32 filters with size of (3×3) in all CNN layers in the second and third ResNet blocks, respectively. In each ResNet block, the dropout strategy was used in the forward and backward propagation between CNN layers and connected to a maximum pooling layer after each CNN layer. The first fully-connected layer was set to 16-way and the second fully-connected layer was set to 1-way to output the positive prediction. As a result, m6Aiso outputs the predicted modification probability at each DRACH site. The 1D-ResNet and self-attention models were constructed with similar block size and fully-connected layer as the 2D-ResNet m6Aiso model. A read was deemed modified if its predicted modification probability surpasses a threshold of $P = 0.9$. Conversely, a site on an individual read with predicted modification probability < 0.1 was classified as unmodified, while the remainder was not used in the m⁶A level calculation. The m⁶A methylation level for each site was then calculated by dividing the number of modified reads by the total number of predicted modified and unmodified reads.

Model training and evaluation
By considering the situation that there are some false positive sequence samples in our training positive samples, we used a semi-supervised learning strategy[26] to train a better m6Aiso model. In detail, there are pseudo-labels for some positive samples, while almost all labels of negative samples are real in our training sequence data. Meanwhile, the numbers of positive samples versus the number of negative samples are extremely unbalanced. It is hard to train a good prediction model if we use those data to train m6Aiso directly. To overcome the issues of pseudo-labels in positive samples and data unbalance in training data, we designed a semi-supervised learning strategy to train the m6Aiso model. Specifically, we first randomly selected a part of the negative samples and combined them with all positive samples to constitute a new training data set with relative balance samples. Because our data cleaning process relies on accurately identifying false positives from the positive training samples, we used relatively more negative samples (positive vs. negative at a ratio of 1:2.5) in the training to prioritize learning the negative signals. The negative samples were randomly sampled from all negative samples for each training; therefore, the majority of the large number of negative samples were used in the whole data cleaning process. Then we used the selected data set to train an initial prediction m6Aiso model and used it to predict the pseudo-label of each positive sample in the training data. After that, we then discarded 5% of samples with the lowest predicted m⁶A probabilities from the positive sample set and updated the training data to re-train the m6Aiso model. This update and re-training process will be repeated and terminated till the false positive rate (FPR) < 0.05 in the positive training data. The FPRs were calculated as the proportion of samples with predicted m⁶A probabilities < 0.5. Notably, although we used the trained model to re-predict the positive samples in the training data, this was different from the standard self-training pipeline in semi-supervised learning and may not be useless in most cases; while, as our training data is skewed to negative samples, the fitted model will tend to predict false positive samples with negative pseudo-labels. This will help us to obtain cleaner positive samples in training data gradually and fit a more conservative model. To train the m6Aiso model, we used the cross-entropy loss function and Adma optimizer to optimize the weights in neural networks using a learning rate of 0.001. In each round of model training, we set the batch size as 64, and the epoch as 50, and choose the model that performs the best on the remaining 10% of the training data.

Comparison between m6Aiso and NGS-based experimental methods
Based on the m6Aiso determined m⁶A sites on single reads, the genomic sites with at least 20 m⁶A modified reads were identified as m6Aiso determined m⁶A sites at the gene level. The m⁶A sites identified by miCLIP,[22] m6A-SAC-seq,[23] eTAM-seq,[4] and m6ACE-seq[24] were obtained as previously provided. In addition, to evaluate the lowly methylated m⁶A sites, we also included the GLORI identified m⁶A sites with levels < 0.1 in the m⁶A level comparisons by re-executing the original bioinformatics pipeline of GLORI[3] but without removing the sites with levels < 0.1. The GLORI data in A549 cells generated in this study was also processed in the same way.

Comparison of m6Aiso with other machine learning models

To evaluate the effectiveness of m6Aiso in mapping m^{6}A modifications, we conducted a comparative analysis with established computational tools that are capable of detecting m^{6}A at the read level and supporting methylation rate calculations. We ran m6Anet[18] (v2.0.0) with the default parameter, and only the DRACH sites with at least 20 reads were considered. For DENA[17] (v1.0), we required at least 20 reads for the m^{6}A sites and preserved the sites with modification levels below 0.1. For nanom6A[14] (v2.0), reads with a modification probability exceeding 0.8 were considered as modified in this study. The probability value of each nucleotide being an m^{6}A site modeled by iM6A was directly obtained from the work of Luo et al.[33]

Performance of m6Aiso on the synthetic Curlcakes dataset

To assess the capacity of m6Aiso to accurately distinguish between modified and unmodified reads, we performed inference using the synthetic Curlcakes dataset.[12] We combined all reads from two replicates of the modified and unmodified samples, respectively; and we excluded all 5-mer motifs that contained more than one adenosine. Thus, only reads from 43 sites within four DRACH motifs (GGACC, GGACU, UGACC, and UGACU) were preserved for subsequent analyses.

Calculation of m^{6}A methylation and expression levels for isoforms

For m6Aiso and m6Anet determined m^{6}A sites at the gene level, we calculated the m^{6}A level of these sites on each isoform according to the aligned isoforms of the long reads that carry the m^{6}A sites. At least 10 reads covering the m^{6}A sites on the same isoforms were required to determine the methylation levels of the m^{6}A sites on specific isoforms. Based on the m^{6}A sites on each isoform, the m^{6}A level of the isoform was calculated as the sum of all the modified reads at all m^{6}A sites on the isoform divided by the sum of the read coverages of all those sites. Meanwhile, the C-to-U mutation level of each isoform was determined as the sum of the mutated read counts of all C-to-U mutation sites on the isoform divided by the sum of the read coverages of all those mutation sites.

Analysis of wild-type and *METTL3*-KO samples

To assess the precision of m6Aiso, we obtained ONT DRS data for wild-type and corresponding *METTL3*-KO in HEK293T and mESCs from Pratanwanich et al.[54] and Zhong et al.[34], respectively. The datasets underwent identical preprocessing steps as previously described, with the exception that the mESCs samples were aligned to the GRCm38 Ensembl annotated transcripts (v91). The m^{6}A sites supported by at least 20 modified reads in WT or *METTL3*-KO and covered by more than 20 modified plus unmodified reads in both conditions were used in the comparisons of m^{6}A levels between WT and *METTL3*-KO samples. The GLORI dataset for WT and *METTL3*-KD in HEK293T was obtained from Liu et al.[3] The eTAM-seq dataset for WT and *Mettl3*-KO in mESCs was obtained from Xiao et al.[4]

Identification of m^{6}A sites across cell lines

The DRS datasets for five cell lines (A549, MCF7, K562, HCT116, and HEK293T), HeLa cells, and *Arabidopsis thaliana* were obtained from the Singapore Nanopore-Expression project (SG-NEx project),[27] Acera Mateos et al.,[28] and Parker et al.[21], respectively. The replicates of DRS data collected from different cell lines were merged, respectively. We then processed the data in the same way as we have above described, except that the *Arabidopsis thaliana* samples were aligned to the TAIR10 transcriptome.

Co-occurrence analyses of m^{6}A sites

Only the reads spanning at least 90% of the length of their corresponding mapped transcripts are used for the analyses. For each read and isoform, we calculated the transcriptomic distance between two adjacent m^{6}A sites. To calculate the number of m^{6}A modifications clustered with adjacent m^{6}A sites within 50 bp on the same genes for Figure 3B, we used the consolidated transcripts of all isoforms for each gene.

For each pair of m^{6}A sites in the same gene, we employed the parameter D',[36] a metric commonly used to assess the LD between two genetic variants within a population, to investigate the linkage of the pair of m^{6}A sites across all the reads that covered the two sites using the following formula:

$$D' = \frac{D}{D_{max}}, D_{max} = \begin{cases} \min(P_A * (1 - P_B), (1 - P_A) * P_B) & \text{if } D > 0 \\ \min(P_A * P_B, (1 - P_A) * (1 - P_B)) & \text{if } D < 0 \end{cases}$$

where $D = P_{AB} - P_A * P_B$, P_A and P_B represented the proportions of reads with modified A site and modified B site, respectively, while P_{AB} denoted the proportion of reads with both modifications. A positive D' value implies that the two modified sites tend to co-occur beyond random chance.

Analysis of m^{6}A sites on isoforms of single genes

Only the isoforms supported by at least 10 ONT DRS reads, were included in the downstream analysis. For a specific m^{6}A site on a particular isoform, we employed Fisher's exact test to assess whether its m^{6}A level significantly differed from the combination of other isoforms from the same gene followed by Benjamini-Hochberg (BH) based false discovery rate (FDR) correction for multiple testing. The m^{6}A sites exhibiting m^{6}A level changes exceeding 0.1 and FDR less than 0.05 were considered as isoform-specific highly or lowly methylated m^{6}A sites. To analyze the transcript type compositions of the isoform-specific highly and lowly methylated m^{6}A

sites, we excluded the genomic sites that exhibit highly methylated on one isoform while displaying lowly methylated on the other isoform.

To investigate the relationship between m⁶A and their proximity to exon junctions, we compared the m⁶A levels of identical sites on pairs of different isoforms. We exclusively used the sites that were more than 200 nt away from EJCs on one of the pairs of isoforms but less than 100 nt away from the EJC on the other isoforms. Fisher's exact test was used to evaluate the difference in modification level of m⁶A sites, with FDRs calculated using the BH method. The m⁶A sites were considered as differentially methylated m⁶A sites on isoform if the absolute changes of m⁶A levels between two isoforms exceeded 0.1 and FDRs less than 0.05. The TARBP2 HITS-CLIP data was obtained from Goodarzi et al.[55] Genome coordinates from hg19 were converted to hg38 using the LiftOver tool. Introns and flanking exons bound by TARBP2 were included in the downstream analysis.

To assess the relationship between m⁶A level changes and isoform expression, the isoform expression was quantified using ESPRESSO[56] and only the isoforms with CPM values greater than 10 were included in the analyses. A two-tailed Fisher's exact test was employed to compare modification levels among A549, MCF7, and HCT116 cell lines. Multiple testing correction was performed using the BH method. A transcriptomic m⁶A site was considered differentially methylated if the absolute change of m⁶A levels exceeded 0.2 and the FDR was less than 0.01. Isoform expression changes between two cell lines were defined as the difference in expression fractions exceeded 0.1.

Identification of dynamic m⁶A sites during the EMT process

Cutadapt[62] (v3.5) was used to remove the adapters from the Illumina sequencing reads. Then, the reads were aligned to the human reference genome (GRCh38) using STAR[63] (v2.7.4a). FeatureCounts[58] (v2.0.1) was utilized to quantify the number of reads mapped to each gene. Genes with counts exceeding 10 in both control and TGF-β samples were retained for subsequent analysis. Differential expression analysis was conducted using DESeq2.[59] The genes with fold changes > 2, and *P*-adjust values < 0.01 were identified as the differentially expressed during the EMT process. GO enrichment analysis was conducted using the R package clusterProfile[57] (v4.3.8). The m⁶A-seq data from control and TGF-β induced HeLa cells were obtained from the GEO database (GEO: GSE112795). The data were processed and m⁶A peaks were identified as described previously.[51,66] For the replicates of the ONT DRS control and TGF-β samples, we preprocessed them as described above previously. The m6Aiso determined common m⁶A sites of the two replicates with identical loci on the genome in control or TGF-β samples were used in the downstream isoform analyses. To investigate isoform-specific dynamic changes of identical m⁶A sites, we compared the same m⁶A sites on the same isoforms between control and TGF-β samples. Only the m⁶A sites on isoforms that covered at least 10 reads in all samples were used in the analyses. Differentially methylated m⁶A sites at isoform level were determined as the sites with absolute changes in their modification levels between control and EMT cells exceeding 0.15. Genes with up-regulated m⁶A sites on isoforms were selected for functional enrichment analysis using clusterProfiler[57] (v4.3.8) based on MSigDB C2 and C5 gene sets.[67,68] The fractions of expressed isoforms for each gene were calculated based on the ESPRESSO (v1.4.0)[56] determined isoform expression.

Motif enrichment analyses for the promoter regions

We defined an isoform as an alternative promoter isoform if its transcription start site (TSS) is located more than 500 nt away from the TSS of the reference isoform, which exhibits the highest expression level in the same gene in control cells. Promoter regions were from 1kb upstream to 500 bp downstream of the TSS. For motif enrichment analysis, we obtained the binding motifs of SMAD3 from the CIS-BP database.[69] The R package motifmatchr (https://github.com/GreenleafLab/motifmatchr.git) was then used to determine whether transcription factor motifs matched the promoter regions. The number of promoter regions matching a specific transcription factor motif was subsequently counted for different sets of isoforms. Finally, we performed a one-tailed Fisher's exact test to determine if alternative promoter isoforms were significantly enriched in the motif of SMAD3.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using R or Python software. Error bars represent the standard deviation (SD) of the mean of the replicates. Detailed information about statistical tests is provided in figure legends for respective figures.

ADDITIONAL RESOURCES

The detailed protocol for m6Aiso is available in Method S1.