

Article

Transcription Start Site Choice Regulates m⁶A Stoichiometry in Cap-Proximal Regions

Jianheng Fox Liu and Samie R. Jaffrey * 

Department of Pharmacology, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA; jil4026@med.cornell.edu

* Correspondence: srj2003@med.cornell.edu

Abstract

Background/Objectives: N⁶-methyladenosine (m⁶A) is the most prevalent and functionally significant internal modification within eukaryotic mRNA. While m⁶A is known to be regulated at internal sites by factors such as splice junctions, the mechanisms governing deposition within the cap-proximal region remain poorly understood. This study aims to determine the patterns of m⁶A stoichiometry in cap-proximal regions and to investigate whether the choice of the specific transcription start site (TSS) can affect m⁶A stoichiometry. **Methods:** We re-analyzed our published single-nucleotide-resolution CROWN-seq data to quantify m⁶A stoichiometry across transcript isoforms with different TSSs, and assessed the relationship between specific TSSs and specific m⁶A sites (“TSS-m⁶A-site pairs”). **Results:** We established the first single-nucleotide-resolution dataset of m⁶A stoichiometry across the transcriptome in cap proximal regions, including stoichiometry of m⁶A across 5′ isoforms for each gene. We found that m⁶A deposition is markedly inhibited within a narrow cap-proximal region in a distance-sensitive manner. m⁶A sites located close to both 5′ and 3′ exon ends exhibit low methylation due to the overlap between the cap-proximal and 3′ exon-end exclusion zones. **Conclusions:** We find that the first exon contains a narrow m⁶A exclusion zone at its 5′ end. As a result, cap-proximal m⁶A sites can have different stoichiometries depending on the TSS choice. As the m⁶A site is positioned farther from the TSS, m⁶A stoichiometry increases. These results reveal that TSS switching is a regulatory mechanism for m⁶A stoichiometry in cap-proximal regions and provide a mechanism for fine-tuning gene expression and mRNA fate through isoform-specific m⁶A modification stoichiometry.

Keywords: m⁶A; m⁶Am; transcription start site; RNA modification; GLORI; CROWN-seq; 5′ UTR; cap-proximal region



Academic Editor: Laurent Metzinger

Received: 13 March 2026

Revised: 24 April 2026

Accepted: 30 April 2026

Published: 31 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

N⁶-methyladenosine (m⁶A) is the most prevalent modified nucleotide in mRNA in human cells [1,2]. Previous studies have shown that m⁶A is critical for multiple molecular processes, including RNA stability, splicing, and targeting to stress granules [2–4]. Dysfunction of m⁶A deposition is strongly linked to significant pathologies, including various cancers, neurological disorders, and immunological diseases [3,4]. Therefore, there is considerable interest in characterizing the molecular mechanisms that govern how m⁶A is deposited across the transcriptome.

One important question is why m⁶A is deposited unevenly throughout the mRNA transcript. It is known that m⁶A is deposited by the m⁶A writer complex, which pref-

entially modifies adenosines within the canonical DRACH (D = A, G, or U; R = A or G; H = A, C, or U) motif [5]. However, early studies demonstrated that across mRNA transcripts, DRACH motifs around stop codons are selectively methylated [6]. Recently, an “exon-end exclusion zone” model was proposed to solve this spatial puzzle. This model was suggested by the discovery that m⁶A deposition is repressed around the edges of exons due to competition between the m⁶A writing complex and the splicing machinery, including the exon junction complex (EJC) [2,7–10]. The splicing machinery comprises a large multiprotein complex that can span over ~100 nt on either side of the exon–exon junction. Since most exons are ~88 to 165 nt (25th to 75th quantile), they are completely covered by the splicing machinery, explaining the low stoichiometry of m⁶A in these exons [8–10]. However, for exons >400 nt, the central part of the exon is not covered by the splicing machinery, thus explaining why DRACH sites in the middle of long exons tend to be highly methylated [8–10]. Overall, these studies establish that m⁶A stoichiometry is regulated by the exon architecture [11]. Although much is known about m⁶A levels in large internal exons and the role of splicing machinery in shaping their stoichiometry, much less is known about m⁶A stoichiometry in 5′ regions of RNA. This region is derived from the first exon, which lacks an EJC binding on its 5′ side. Thus, it is unclear whether an exclusion zone exists in this region and if m⁶A stoichiometry is related to its position in the 5′ exon.

One major challenge to studying m⁶A sites in the 5′ exon is the lack of m⁶A stoichiometry data. Part of the reason why m⁶A stoichiometry is poorly annotated in mRNA 5′ regions is that current methods for determining m⁶A stoichiometry typically exhibit poor coverage in 5′ regions. Conventional methods, including antibody-based techniques like miCLIP and MeRIP-seq, as well as chemical methods like GLORI, often yield disproportionately low sequencing depth around 5′ ends [5,6,12]. Other methods that also provide quantitative m⁶A stoichiometry measurements, like Oxford Nanopore direct RNA sequencing, have low accuracy in 5′ regions [13]. Lastly, 5′ ends show considerable isoform diversity, and it is possible that m⁶A sites in 5′ regions could be affected by these 5′ isoforms. Indeed, we recently showed that 5′ isoform diversity can be substantial for many genes [14]. Therefore, while multiple studies have reported the importance of the 5′ region m⁶A [15–18], it remains unclear whether the stoichiometry of these sites is correctly determined or if they possess specific biological functions.

In this study, we describe m⁶A stoichiometries in mRNA 5′ UTRs, and we show that 5′ m⁶A stoichiometry can be highly affected by 5′ UTR isoform diversity. To assess 5′ m⁶A stoichiometry throughout the transcriptome, we used CROWN-seq, a method that uses the chemical conversion method of GLORI to quantify stoichiometries of cap-associated m⁶A in a 5′ isoform-specific manner [14]. By re-analyzing CROWN-seq data, we established a dataset for m⁶A sites across different 5′ isoforms, termed “TSS-m⁶A-site pairs”. By analyzing these pairs, we found that m⁶A deposition is markedly repressed in the cap-proximal region, specifically within the first 30–40 nucleotides. We also identified joint repression of m⁶A deposition driven by both the proximity to the cap and the 3′ end of the first exon. Our data indicate that an m⁶A site can have highly variable stoichiometry depending on the specific transcription start site (TSS) used in the transcript isoform, revealing an important role for alternative 5′ isoforms in establishing the stoichiometry of nearby m⁶A sites. Overall, this study reveals a novel cap-proximal mechanism for the spatial regulation of m⁶A deposition.

2. Materials and Methods

2.1. Genomic Assembly and Gene Annotations

Gencode v45 (GRCh38) was used in this study.

2.2. Reanalysis of Published Sequencing Data

The GLORI data of HEK293T samples were obtained from the Supplementary Table of the original study [12].

The CROWN-seq data were obtained from our previous study (GSE233655) [14]. Note that the HEK293T samples designated as replicate #1 and #2 are the corresponding samples designated as technical replicates #3 and #4 in our previous study. Replicates #1 and #2 were excluded because their low sequencing depths were insufficient for the analyses.

2.3. Estimating CROWN-Seq Non-Conversion Background

We used the RNA spike-in derived from ERCC-00057 to estimate the internal non-conversion background in CROWN-seq. These ERCC spike-ins were made via T7 in vitro transcription with Am/ m^6 Am cap-analogs (details see our previous study [14]). The sequences of the RNA spike-ins are: TAATACGACTCACTATAAGTGCATCCTCGTGGCATCATGCGTCTCCTCAGTAGGTCTGCNNNNNNGACTGATCCTAGTGCAATGCGTCTGAGCC TGAGCTACAGCGATATAGCCTGGATTGTGAGCGTATTTGCTGTCAGAACCTCAGCTCATCATGTATGATGCTGTACCATCCTGCGATTTA, where TAATACGACTCACTATA is the T7 promoter; the A nucleotide immediately adjacent to the promoter sequence is the TSS; NNNNNN is a barcode sequence without A's to distinguish different spike-ins in the same sequencing round (five different spike-ins in total). These spike-ins were generated by in vitro transcription with different ratios of m^6 Am/Am cap analog to achieve a m^6 Am stoichiometry ranging from 0% to 100%. No internal m^6 A modification was made. These RNA spike-ins were added to the RNA samples before CROWN-seq library preparation. In total, we generated three libraries with spike-ins. Therefore, we obtained 15 data points for each A nucleotide position (5 different barcodes $\times$ 3 replicates). All A positions analyzed in this study have a coverage >100 reads. Detailed data can be found in Table S3.

2.4. 5' Isoform-Specific m^6 A Identification and Quantification of CROWN-Seq Data

We developed a customized computational pipeline to quantify m^6 A stoichiometry at the resolution of individual transcription start sites (TSSs). The CROWN-seq reads were first mapped as previously described as the mapping pipeline as previously described [14]. The resulting CROWN-seq alignments were then processed read-by-read to examine both the genomic coordinate of the transcription start site (i.e., the first nucleotide in a read) and the conversion state of every nucleotide mapped to an A nucleotide in the reference genome. In particular, when the A nucleotide is sequenced as G in the read, it indicates the adenosine is not methylated; while if the A nucleotide is sequenced as A, it indicates the adenosine might be m^6 A modified. Pysam (v0.19) [19] was used to perform BAM file processing. BAM file-related computation was performed on an Ubuntu 20.04.3 LTS environment. After iterating the BAM file, a table was generated, which contains the A or G reads of all the genomic A positions and their associated TSSs. This table was then processed for the identification and quantification of m^6 A sites as well as TSS- m^6 A-site pairs.

To call TSS- m^6 A-site pairs, we first identified the m^6 A sites. In this study, we focused on identifying high-confidence m^6 A sites. In previous modification studies, a modified site might be determined by ≥ 20 reads and ≥ 0.1 stoichiometry [14,20,21]. However, in this study, we used a stricter filter to avoid false positives and ensure quantification accuracy: a high-confidence m^6 A site must be (1) located in a DRACH motif; (2) have ≥ 50 reads and ≥ 0.2 stoichiometry in at least one 5' isoform.

This filter is related to the non-conversion background, which is ~1.5% to 4% in different positions in a read. As estimated by the binomial distribution, the non-conversion background might influence m^6 A identification when the background non-conversion rate is slightly higher (e.g., 5%) than our expectation, or when the sequencing depth is

low (e.g., 20 reads) (Figure S1). Therefore, we used a high coverage (≥ 50 reads) and high stoichiometry cutoff (≥ 0.2) to ensure both accuracy and precision.

After m⁶A site identification, we then quantify the stoichiometries in the corresponding TSS-m⁶A-site pairs. During this step, we only considered the TSS-m⁶A-site pairs that were covered by ≥ 50 reads. To increase sequencing depth for each cell type, we merged technical/biological replicates.

2.5. TSS and 3' Exon End Annotation of TSS-m⁶A-Site Pairs

We directly calculated the distance between a TSS and an m⁶A site based on their relative position in a read. We did not use genomic coordinates of the TSS and the m⁶A sites directly because splicing might occur at the 5' end.

To investigate the joint effects of transcript architecture, we annotated each m⁶A site with its distance from the 3' end of the first exon. To do so, we first examined the splicing events in each read. If splicing was observed, we used the distance between the m⁶A site and the splicing site. If no splicing event was found in the read, we then examined the closest annotated 3' exon end in Gencode v45 annotations. To do so, we used the “closest” subprogram of BEDTools [22] to annotate the distance between the TSS and all possible 3' exon ends. A customized script was used to rank all the distances and find the closest 3' exon end.

The related codes can be found on GitHub: https://github.com/jhfoxliu/TSS_m6A (accessed on 12 March 2026).

2.6. Linear Fixed-Effects Model

To quantify the impact of 5' proximity on m⁶A stoichiometry, while controlling for site-specific variability, we employed a linear fixed-effects regression model. Analysis was performed using the Python package “Statsmodels” [23]. TSS-m⁶A-site pairs were first categorized into 5-nucleotide (nt) bins based on their distance from the 5' end. We then modeled m⁶A levels using the formula: m⁶A level $\sim$ C(5' distance bin) + C(m⁶A site).

In this model, the m⁶A site was treated as a fixed effect to account for site-specific methylation efficiencies and local sequence contexts. The intercept represents the baseline methylation level of the reference bin (1–5 nt), while the coefficients for subsequent bins represent the mean deviation from this baseline. The model's adjusted R² was used to assess the proportion of variance explained by the combination of spatial positioning and site-specific factors. The detailed model output can be found in Table S1.

2.7. Motif Analysis

Motif analysis was performed with Weblogo 3.7.12 [24]. The 2 nt flanking region of the m⁶A site was used as input. Probability mode (--units probability) was used to generate the frequency logo.

2.8. Quantification and Statistical Analysis

Python (3.13.7) and its related packages were used to perform quantification and statistical analysis. Statistical analysis was performed with Numpy (v1.23.5), pandas (v1.5.2), Scipy (v1.9.3), and statsmodels (v0.14.2). Figures were generated with Matplotlib (v3.6.2) and Seaborn (v0.12.1). *p*-value significance levels: * (<0.05), ** (<0.01), *** (<0.001), **** (<0.0001).

3. Results

3.1. m^6A and m^6Am Are Easily Confused with Each Other in Non-5' Specific Analysis

We sought to quantify the stoichiometry of m^6A sites in mRNA 5' UTRs. A major problem with previous studies of m^6A in 5' UTRs is the presence of m^6Am , a modified nucleotide that is found exclusively at the first nucleotide position of RNA Pol II transcripts (Figure 1A). In previous antibody-based methods, like MeRIP-seq, the m^6A antibodies also bind m^6Am , causing peak enrichments in mRNA 5' regions. In many studies, 5' peaks are misattributed to m^6A , when they are instead much more likely to be m^6Am . We partially addressed this by our discovery of the m^6Am -forming enzyme, PCIF1 [25], and subsequent mapping of m^6A in *PCIF1* knockout cells [14]. Although this study showed that most 5'-associated peaks were due to m^6Am , it was clear that m^6A is located in 5' regions of RNA Pol II transcripts. However, MeRIP-seq does not provide quantitative measurements of m^6A .

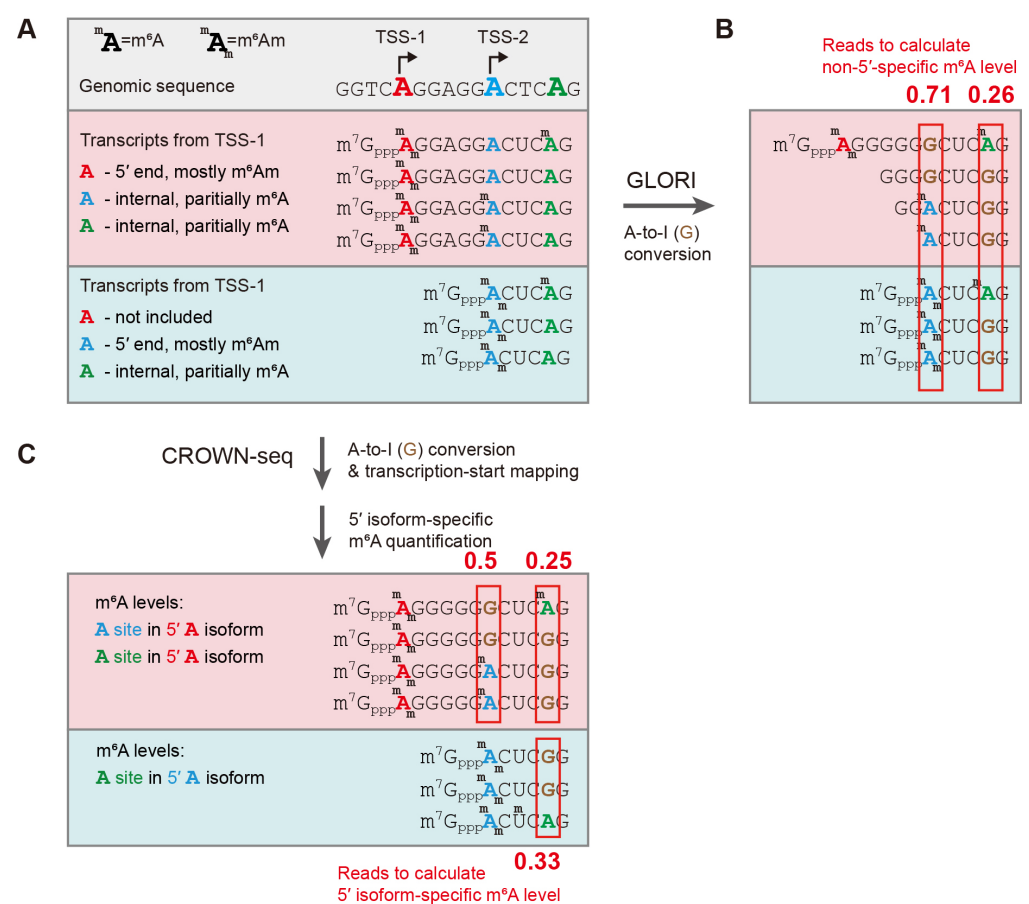


Figure 1. CROWN-seq provides accurate quantification of cap-proximal m^6A sites: (A) The complex sequence architecture at the 5' terminus, where a gene may utilize multiple genomic adenosines differently: a proximal site used exclusively as a transcription start site (TSS) (red), a middle site functioning as either a TSS or an internal site depending on isoform choice (blue), and a distal site used purely as an internal position (green). (B) Conventional methods like GLORI aggregate these signals, leading to inaccurate stoichiometry at the middle site due to the conflation of m^6A and m^6Am . Furthermore, for the distal site, non-5'-specific analysis fails to detect potential isoform-specific methylation dynamics. The way to calculate the methylation level and the results are indicated in red. (C) m^6A quantification at 5' isoform resolution using CROWN-seq. CROWN-seq resolves cap-proximal quantification issues by selecting and labeling RNA fragments by their m^7G caps following GLORI chemistry conversion. This approach physically separates m^6Am signals from internal sites, enabling the calculation of m^6A stoichiometry for each specific 5' isoform.

More recently, the GLORI method for mapping m⁶A was developed for quantitative measurements of m⁶A [12]. In this method, sodium nitrite is used to deaminate adenosines to inosine, while m⁶A is resistant to sodium nitrite-mediated deamination. During sequencing, inosines are read as G, while m⁶A sites are read as A. The fraction of reads that contain A rather than G at any individual site indicates the stoichiometry of m⁶A at any position (Figure 1B). However, certain limitations may affect the ability to detect m⁶A in 5' regions using GLORI. First, the original GLORI dataset showed low coverage at 5' ends, which was related to the fragmentation that occurs in the GLORI protocol. Second, the small fragments obtained using GLORI could not be definitively assigned to specific 5' isoforms. Thus, the stoichiometry of m⁶A sites near the 5' ends of mRNA reflects an average stoichiometry across all the 5' isoforms for any gene.

We first assessed the existing GLORI dataset to determine if 5' m⁶A sites could be detected. We examined the previously published GLORI dataset derived from HEK293T cells. Among the 170,240 high-confidence sites identified by GLORI [12], approximately 6016 are located within 100 nucleotides downstream of a TSS. Since m⁶A sites have previously been confused with m⁶Am sites, we compared the locations of these sites with transcription-start nucleotides identified with ReCappable-seq [26], a method for mapping the precise position of the first nucleotide in mRNAs. Notably, about half of the m⁶A sites in 5' regions (n = 3246) directly overlapped with a TSS, raising the possibility that a considerable fraction of m⁶A sites mapped to 5' ends in GLORI in fact are m⁶Am.

To address this problem, we decided to use our recent CROWN-seq dataset, which was originally developed to quantify m⁶Am stoichiometries in HEK293T mRNA [14]. CROWN-seq involves selective labeling of the 5' ends of mRNA using a decapping and recapping step, which ensures that true capped 5' ends of RNAs are recovered for analysis. These 5' fragments are subjected to the GLORI chemistry, which we showed can convert 2'-O-methyl adenosines at the first nucleotide position of mRNAs into 2'-O-methyl inosines, but leaves m⁶Am unmodified. Similar to GLORI measurements of m⁶A, the stoichiometry of m⁶Am can be assessed by measuring the fraction of reads that remain as A (Figure 1C).

We then further explored our CROWN-seq dataset, but instead focused on m⁶A sites that were internally located in each read. In this analysis, we examined 2231 of the 3246 m⁶A sites sequenced by both GLORI and CROWN-seq, which have high coverage in CROWN-seq (i.e., at least 50 reads) in two technical replicates. We observed that the majority of these sites are predominantly m⁶Am with high stoichiometry (typically >0.9), indicating that m⁶Am sites can easily be mischaracterized in GLORI datasets (Figure 2A). Remarkably, we identified 92 sites that exhibit both m⁶A and m⁶Am signatures, i.e., m⁶A in one transcript isoform and m⁶Am in another (Figure 2A; Table S2). Motif analysis showed the sites being modified as m⁶A and m⁶Am are in classic DRACH motifs, while the sites preferentially modified as m⁶Am are mostly non-DRACH motifs (Figure 2B). Therefore, in the traditional GLORI-based m⁶A analysis, m⁶A and m⁶Am would not be distinguished from each other.

3.2. CROWN-Seq Enables m⁶A Stoichiometry Quantification at 5' Isoform Resolution

A primary advantage of CROWN-seq is its ability to simultaneously capture the precise transcription-start nucleotide, which thus reveals the exact 5' isoform and internal m⁶A stoichiometry. This unique feature allows m⁶A modifications to be attributed to specific 5' transcript isoforms, providing a resolution that is often lost in non-5'-specific sequencing. Consequently, CROWN-seq uniquely reveals m⁶A stoichiometry, which may be different depending on the specific 5' isoform.

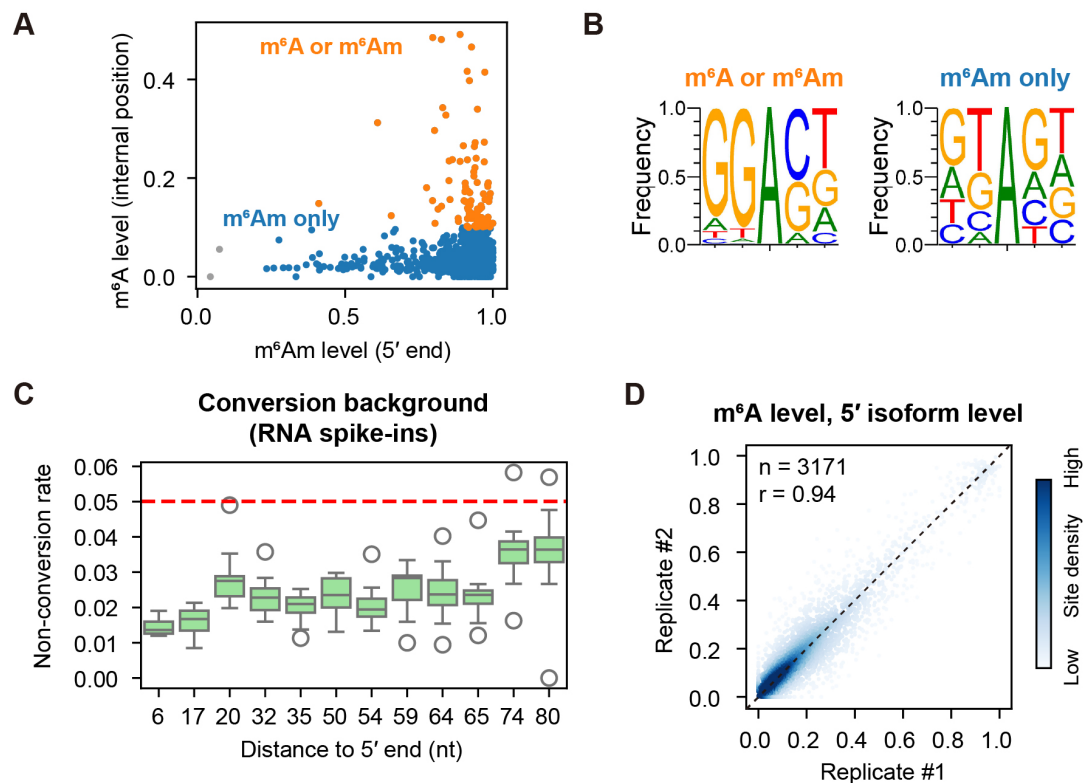


Figure 2. CROWN-seq provides accurate cap-proximal m⁶A quantification: (A) Global mischaracterization of m⁶Am as m⁶A in non-5'-specific datasets. Comparison of m⁶Am levels at 5' ends (X-axis) versus non-5'-specific m⁶A stoichiometry for the same genomic adenosine when transcribed as an internal site (Y-axis). The methylation levels are obtained from CROWN-seq. Data points highlighted in orange represent sites with internal m⁶A levels exceeding 0.1, identifying a subset of adenosines with dual modification identities. The CROWN-seq technical replicate #1 was used. (B) Corresponding to (A), shown are the motif logos of the sites that are modified as both m⁶Am and m⁶A, or the sites merely modified as m⁶Am. (C) The non-conversion background in CROWN-seq estimated by RNA spike-ins. In this analysis, we plotted the non-conversion rates of adenosines in the RNA spike-in sequence between position 6th and 80th. (D) High reproducibility of isoform-specific m⁶A quantification by CROWN-seq. Scatter plot demonstrating robust correlation of m⁶A stoichiometry in TSS-m⁶A-site pairs between technical replicates.

While we previously demonstrated that CROWN-seq achieves high precision in quantifying m⁶Am stoichiometry at the +1 position [14], it was essential to verify the efficiency of sodium nitrite-mediated deamination within internal read regions. To evaluate the background non-conversion rate, we analyzed the synthetic RNA spike-ins utilized in our prior study [14], which contain unmodified internal adenosines and a modified m⁶Am at the first nucleotide. Our analysis revealed high conversion efficiency across internal adenosine residues: the background non-conversion rate at the 2nd and 17th nucleotide positions was approximately 1.5%, rising slightly to ~2.5% between the 20th and 65th nucleotides (Figure 2C, Table S3). At the 74th and 80th positions, the background rate reached approximately 3–4% (Figure 2C, Table S3). These results establish a low technical noise floor for internal m⁶A detection.

To further validate the technical robustness of CROWN-seq, we compared m⁶A site identification across two independent HEK293T technical replicates. We applied a high-stringency filter (50 reads and ≥ 0.2 stoichiometry in at least one 5' isoform, see Figure S1) to identify internal m⁶A sites across different 5' isoforms (Figure S3A). This led to the identification of 650 DRACH-context m⁶A sites across 2861 transcription start sites (TSSs), resulting in 2982 high-confidence TSS-m⁶A-site pairs. We observed a high degree of

correlation in m⁶A stoichiometry across the two replicates based on non-conversion rates (Pearson's $r = 0.94$; Figure 2D). Notably, 274 of the identified m⁶A sites had not been previously reported by GLORI [12], including sites within essential genes such as *MCM7*, *ATG5*, and *PTTG1* (Table S4). Collectively, these data demonstrate that CROWN-seq is a robust and reproducible method for the identification and quantification of m⁶A at 5' isoform resolution.

3.3. An Inhibitory Effect of m⁶A Deposition Around the Cap-Proximal Region

We next investigated whether m⁶A stoichiometry at a single site varies across different 5' isoforms. To address this, we compared the non-5'-specific (overall) m⁶A stoichiometry of a site with 5' isoform-specific stoichiometry using HEK293T CROWN-seq data. While many sites exhibited consistent methylation levels across isoforms, a significant proportion showed substantially higher stoichiometry (>0.1) in specific 5' isoforms compared to the overall estimation (Figure 3A). Conversely, we identified instances where isoform-specific methylation was markedly lower (<0.1) than overall stoichiometry (Figure 3A), indicating that m⁶A stoichiometry is highly specific to the 5' isoform context.

Our analysis revealed that this isoform-specific variation is strongly dictated by the spatial position of the m⁶A site relative to the transcription start site. For example, at the *CENPW* locus (chr6:126,340,145), a single m⁶A site is associated with over 25 distinct 5' isoforms. We observed a strong negative correlation between methylation and proximity to the 5' end: isoforms where the site was located within the first 10 nucleotides exhibited methylation levels below 0.1, significantly lower than the overall level of ~ 0.2 (Figures 3B and S2). However, in isoforms where the TSS was positioned >20 nt upstream of the same site, stoichiometry increased to 0.3–0.5 (Figures 3B and S2). The Kolmogorov–Smirnov test shows this methylation pattern is consistent between the two replicates. We observed similar distance-dependent inhibition at other loci, including *TIMMDC1* (chr3:119,498,582; Figure 3C) and the long non-coding RNA *XIST* (chrX:73,852,643; Figure 3D).

To determine if this “cap-proximal exclusion zone” is a transcriptome-wide phenomenon, we grouped TSS-m⁶A-site pairs by their distance from the TSS. To ensure high quantitative accuracy, we merged reads from two HEK293T technical replicates and selected high-confidence sites (at least 50 reads and ≥ 0.2 stoichiometry in at least one 5' isoform; see Section 2). This yielded 2982 TSS-m⁶A-site pairs originating from 650 DRACH sites. By binning these pairs by TSS distance, we observed that m⁶A stoichiometry is positively correlated with the distance from the 5' end (Figure 3E). This pattern is found in all subtypes of DRACH motifs (Figure S3B,C).

To account for the inherent variability in methylation efficiency between different DRACH sites, we employed a linear fixed-effects regression model. This model incorporated both the categorical distance bins and the specific m⁶A site as factors (see Section 2). The regression confirmed that distance to the 5' end remains a significant positive predictor of methylation levels, even when controlling for site-specific baseline efficiencies (Figure 3F; Table S1).

Importantly, sampling bias might contribute to an artifact in our statistics above, because different m⁶A sites are associated with varying numbers of TSS-m⁶A-site pairs. To ensure our observations were consistent across individual loci, we performed a site-specific correlation analysis. To do so, we calculated Spearman's correlation coefficients between the distance from the 5' end and m⁶A stoichiometry for 351 m⁶A sites associated with at least three alternative TSSs. Overall, we found that the majority of the m⁶A sites exhibit a higher methylation when it is located further away from the 5' end (Figure 3G). Notably, 192 of these sites (54%) exhibited significantly positive correlations ($r > 0$, FDR < 0.1 ; Figure 3H).

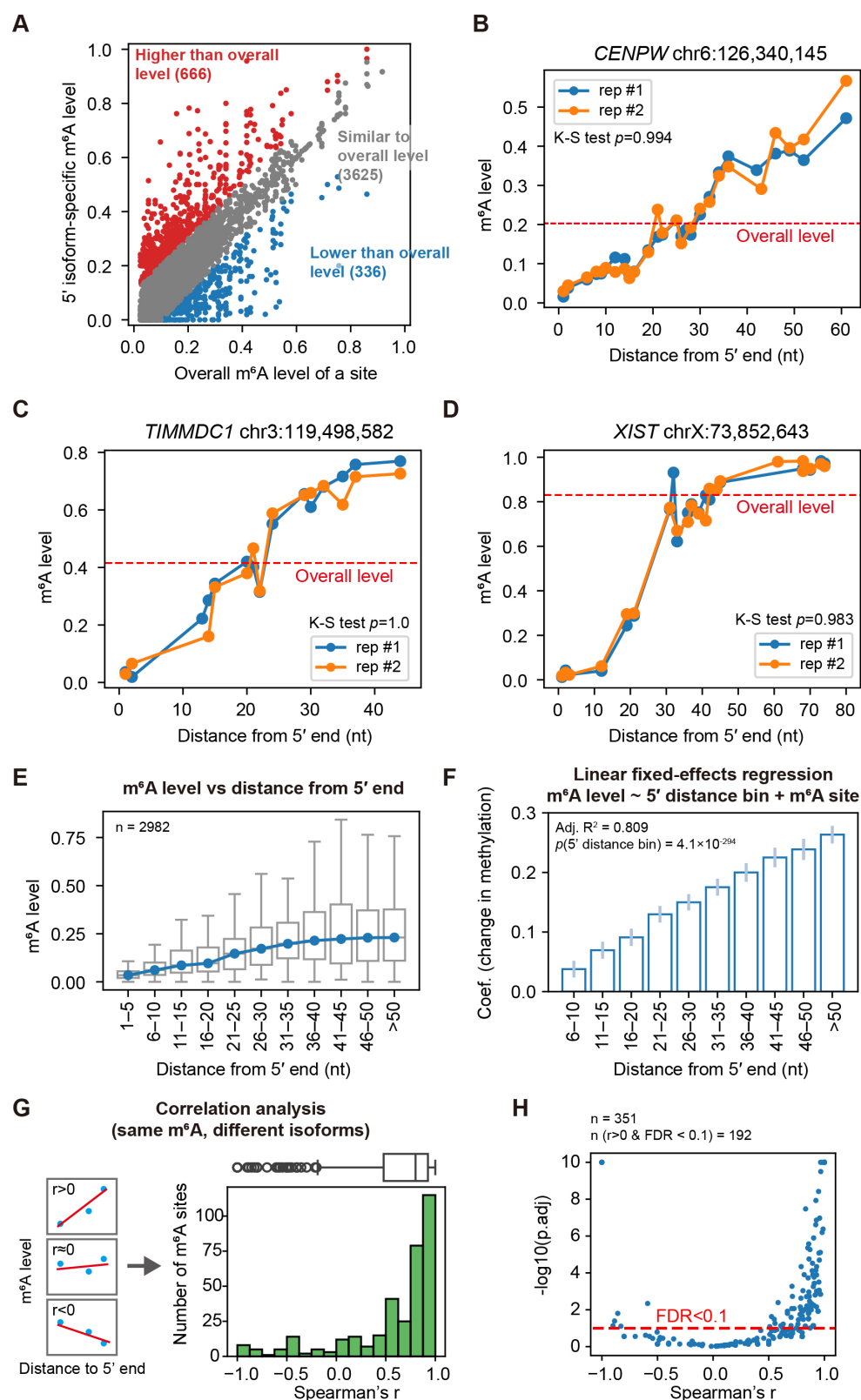


Figure 3. m⁶A deposition is inhibited in the cap-proximal region: (A) Heterogeneity of m⁶A stoichiometry across distinct 5' isoforms. Comparison of overall m⁶A levels versus 5' isoform-specific m⁶A stoichiometry derived from CROWN-seq. Each data point represents a unique TSS-m⁶A-site pair. Pairs exhibiting markedly lower methylation levels (stoichiometry < 0.1) relative to the overall stoichiometry of a site are indicated in blue, while those with elevated methylation levels are highlighted in red. (B–D) Representative loci demonstrating cap-proximal m⁶A inhibition. Examples of m⁶A

sites from the CENPW, TIMMDC1, and XIST transcripts show a marked reduction in stoichiometry when positioned near the 5' terminus. Overall methylation levels of the site are represented by red dashed lines, with isoform-specific stoichiometry from two technical replicates displayed for comparison. Kolmogorov–Smirnov (K-S) tests were performed to examine whether the distribution of m⁶A levels between replicates was different or not. (E) Transcriptome-wide positive correlation between m⁶A stoichiometry and cap distance. Grouped analysis shows that m⁶A levels increase as the distance from the m⁷G cap increases. Median stoichiometry values for each distance interval are plotted in blue and connected to highlight the distance-dependent inhibitory effect. (F) Corresponding to (E), distance from the 5' cap is positively associated with m⁶A stoichiometry. Plotted are the regression coefficients (Y-axis) for each 5' distance bin (X-axis) derived from a linear fixed-effects model ($m^6A \text{ level} \sim 5' \text{ distance bin} + m^6A \text{ site}$). This model considers both the effect of the 5' distance and the site-specific m⁶A deposition efficiency. The model achieves a high adjusted R² (0.802), indicating that the spatial position relative to the TSS, combined with site-specific factors, explains a substantial proportion of the variance in methylation levels. A higher coefficient indicates a more positive contribution to the m⁶A level. Error bars represent 95% confidence intervals (CIs). *p*-values of the coefficient are shown in asterisks. (G) Site-specific correlation analysis confirms widespread inhibition. Distribution of Spearman correlation coefficients calculated between m⁶A stoichiometry and distance from the 5' cap for 351 individual m⁶A sites. The predominant shift toward positive coefficients supports the model that m⁶A deposition is systematically repressed in the cap-proximal region across the transcriptome. (H) Corresponding to (G), among the 351 Spearman's correlation tests, 192 tests were positive and significant after multiple testing (FDR, Benjamini–Hochberg method). In this plot, FDR values smaller than 1×10^{-10} were shown as 1×10^{-10} .

Collectively, these transcriptome-wide analyses in HEK293T cells demonstrate a pervasive cap-proximal exclusion zone that governs m⁶A deposition in the first 30 to 40 nt region.

3.4. Cap-Proximal m⁶A Deposition Inhibition Is Seen in Multiple Human Cell Types

To determine whether the cap-proximal m⁶A inhibition effect is a general regulatory principle rather than a phenomenon unique to HEK293T cells, we extended our analysis to eight additional human cell lines from our previous CROWN-seq study [14]: K562, A549, CCD841 CoN, HCT116, HT29, HepG2, Huh7, and Jurkat E6.1.

Consistent with our initial HEK293T analysis, we applied a stringent filtering pipeline to identify high-confidence TSS-m⁶A-site pairs. Biological and technical replicates were merged, and sites were required to possess at least one 5' isoform with a minimum coverage of 50 reads and a methylation stoichiometry of ≥ 0.2 . Across these diverse cellular contexts, we identified a total of 1346 m⁶A sites within canonical DRACH motifs, representing 6296 high-confidence TSS-m⁶A-site pairs. This dataset is available via Zenodo (10.5281/zenodo.19546831).

We subsequently modeled the relationship between TSS distance and m⁶A levels across all evaluation pairs in these additional cell lines. Remarkably, we observed a robust and consistent correlation between distance from the 5' cap and methylation levels across every cell line tested (Figure S4). These results demonstrate that cap-proximal m⁶A inhibition is a pervasive feature of the human epitranscriptome, suggesting a fundamentally conserved mechanism governing m⁶A deposition at the 5' terminus of mRNAs.

3.5. TSS Usage Influences m⁶A Stoichiometry in the 5' Region

The observed heterogeneity in m⁶A stoichiometry at the isoform level suggests that the overall methylation level that is calculated at a specific site is a composite value, determined by the weighted average of the methylation stoichiometry in each 5' isoform. We hypothesized that differential TSS usage across cell types could therefore markedly impact the overall methylation stoichiometry of specific m⁶A sites, especially in the 5' UTR.

To test this idea, we examined genes with clear cell type-specific alternative 5' isoform usage and that contain m⁶A sites near the TSS. A representative example is the m⁶A site at chr11:83,157,157 within the *PCF11* gene. Based on CROWN-seq, we observed that the overall calculated stoichiometry at this site varies significantly across cell lines: in Huh7 and A549 cells, the overall stoichiometry is relatively low (~0.08), whereas in HEK293T and K562 cells, it increases more than two-fold to 0.21–0.23 (Figure 4A).

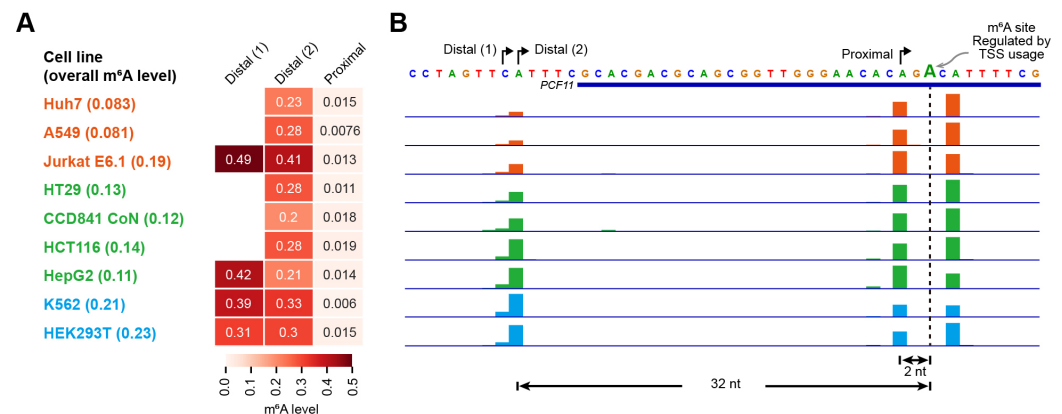


Figure 4. TSS usage determines overall m⁶A stoichiometry in *PCF11* transcripts: (A) Heatmap showing the methylation levels of the site chr11:83,157,157 in different isoforms across nine human cell lines. The overall methylation levels are also shown in brackets. The cell lines in orange preferentially use the proximal TSS; the cell lines in green tend to use both distal and proximal TSSs; the cell lines in blue preferentially use the distal TSS(s). Detailed coverages and methylation levels are available in Table S5. (B) Relative TSS usages of site chr11:83,157,157 across different cell lines.

When examining the 5' isoform usage, it was clear that this variation is directly attributable to a switching in the usage of specific 5' isoforms. In one 5' isoform (the “proximal” isoform), the m⁶A site is only two nucleotides away from the TSS, and exhibits almost no detectable methylation (Figure 4A). In contrast, other 5' isoforms (“distal” isoforms) of *PCF11* have their TSS ~30 nt from the m⁶A site, resulting in 0.2–0.49 m⁶A stoichiometry in these isoforms (Figure 4B). Importantly, Huh7 and A549 cells preferentially utilize the proximal 5' isoform, leading to an abundance of unmethylated transcripts and an overall low level of methylation at this site. In contrast, HEK293T and K562 cells preferentially express distal isoforms, thereby leading to an overall increase in m⁶A stoichiometry at this site (Figure 4A,B). Interestingly, in Jurkat E6.1 cells, while the proximal TSS is still favored, the exceptionally high methylation frequency of the few distal-TSS isoforms provides a compensatory effect, maintaining an intermediate overall methylation level (Figure 4A,B).

Taken together, the case of *PCF11* illustrates that m⁶A stoichiometry can be regulated by differential TSS usage, providing a mechanism for cells to tune the stoichiometry at specific m⁶A sites by switching the usage of specific 5' isoforms.

3.6. Methylation Pattern in the First Exon Is Primarily Regulated by the Cap-Proximal Exclusion Zone

We then asked whether m⁶A deposition is regulated by additional mechanisms beyond the cap-proximal exclusion zone. In the case of internal exons, suppression of m⁶A near exon–exon junctions occurs over ~150–200 nucleotides from the exon ends, and is thought to be driven by the exon junction complex [7–10]. In the last exon, only one end of the exon is bound by the splicing machinery, leading to an increase in m⁶A stoichiometry around the stop codons [7–10]. Notably, the first exon also only has one end bound by the splicing machinery. Since the median length of human 5' UTRs is approximately 218 nucleotides [27],

the first exon might be simultaneously regulated by both 5' cap-proximal and 3' exon-end exclusion zones (Figure 5A).

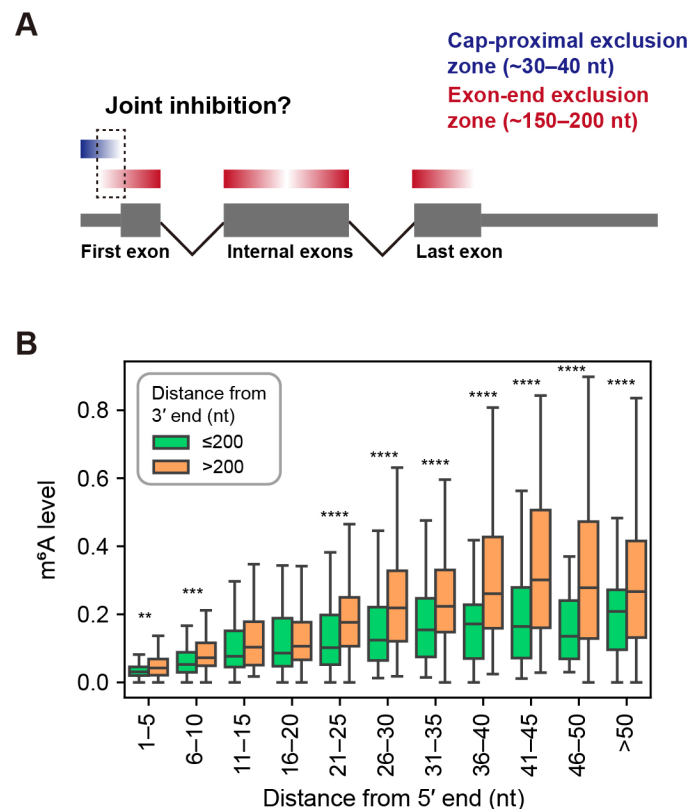


Figure 5. A joint inhibition effect of the cap-proximal exclusion zone and 3' exon-end exclusion zone: (A) Spatial intersection of m⁶A inhibitory mechanisms. This schematic illustrates the potential overlap between the cap-proximal exclusion zone (blue) and the exon-end exclusion zone (red). (B) The m⁶A sites located in the 3' exon-end exclusion zone undergo joint inhibition. Shown are stoichiometries of two groups of m⁶A sites, which are located ≤200 nt or >200 nt from the 3' exon end. 1296 and 1684 TSS-m⁶A-site pairs were analyzed in these two groups, respectively. *p*-values, non-parametric Mann–Whitney U test to compare the ≤200 nt and >200 nt populations in each bin. Detailed *p*-values can be found in Table S1. *p*-value significance levels: ** (<0.01), *** (<0.001), **** (<0.0001).

To test this idea, we first assessed the range of cap-proximal exclusion zone without the interference of exon-end suppression. We therefore established a control group of the m⁶A sites. These m⁶A sites are located >200 nt from the 3' end of the first exon and therefore are unlikely to be influenced by the exon-end exclusion zone. We found that the cap-proximal exclusion zone of these sites is around 30 to 40 nt long, consistent with our findings above (Figure 5B).

We then asked whether the m⁶A sites close to the 3' exon end exhibit joint inhibition by both the cap-proximal and the exon-end exclusion zones. We therefore selected the m⁶A sites ≤200 nt from the 3' end of the first exon, presumably located within the 3' exon-end exclusion zone. We found that these m⁶A sites have increased stoichiometry as the distance from the 5' end increases, indicating the cap-proximal inhibition. However, compared with the control group, the stoichiometries of these m⁶A sites increased significantly more slowly as the distance from the cap increased (Figure 5B). For instance, for sites located >200 nt from the 3' end, the median stoichiometry reached 0.3 when they are 41–45 nt away from the cap. In contrast, the sites ≤200 nt from the 3' end exhibit a median stoichiometry of 0.16 in this region, which is ~47% lower. Therefore, these m⁶A sites undergo a joint inhibition due to the overlapping of the cap-proximal and the 3' exon-end exclusion zones.

4. Discussion

Understanding the regulation of m⁶A deposition is necessary to determine how m⁶A might be regulated in different cellular contexts. Recently, it was proposed that m⁶A deposition is regulated primarily by an inhibitory effect of the splicing machinery, specifically the EJC, at exon–exon junctions [7–10]. Although this model effectively explains m⁶A levels within internal exons, this model does not address how m⁶A stoichiometry is determined in 5′ regions of RNA. Understanding how m⁶A levels in 5′ exons are regulated is further complicated due to the complex and often inaccurate mapping of mRNA 5′ ends [14]. In this study, we used a nucleotide-resolution map of transcription-start sites, along with a chemical method for exact quantification of m⁶A, and found that the stoichiometry of the cap-proximal m⁶A sites is repressed around the cap-proximal region within the first 30 to 40 nt of the TSS. We found that m⁶A stoichiometry is thus regulated by the choice of the transcription-start nucleotide in different transcript isoforms. We also found that m⁶A levels can be jointly suppressed by both the cap-proximal and 3′ exclusion zones of the first exon. This finding indicates that m⁶A stoichiometry is highly repressed in short first exons <200 nt, while high stoichiometry m⁶A sites are detected in large first exons. Overall, our results show that m⁶A stoichiometry in 5′ regions is determined by distinct exclusion zones at the 5′ and 3′ end of the first exon, and can be dynamically regulated by TSS usage.

Our study demonstrates that previous m⁶A studies related to the cap-proximal regions may have yielded inaccurate estimates of cap-proximal m⁶A stoichiometry. These previous studies relied on methods that did not distinguish between different 5′ isoforms. In this study, we re-analyzed our previous CROWN-seq datasets, which provide the stoichiometry of m⁶Am and internal m⁶A for each specific 5′ isoform. This allows, for the first time, accurate quantitative profiling of these cap-proximal m⁶A sites for each 5′ isoform. This analysis revealed that m⁶A stoichiometry can be highly different based on its distance from the TSS.

Our data suggest a new function for alternative TSSs: to regulate m⁶A stoichiometry. Alternative TSSs are known to regulate gene expression by changing the regulatory elements in the 5′ UTR region. Different RNA 5′ isoforms can exhibit variable 5′ regulatory elements, including upstream ORFs (uORFs), RNA-binding protein binding motifs, 5′ terminal oligopyrimidine motifs, and pyrimidine-rich translational elements [28]. In addition to these elements, TSS switching from adenosine to non-adenosine also leads to the loss of m⁶Am at the first nucleotide [14]. Our study shows that a ~10-nucleotide switching in TSS usage might lead to a significant change in m⁶A stoichiometry for an internal cap-proximal DRACH motif. Notably, TSS switching has been observed in diverse cellular contexts, such as different tissues and developmental stages, and is often attributed to different activities of different transcription factors that may use different spatially proximal promoters [29,30]. TSS switching can also be seen as a result of mutations in general transcription factors or RNA Pol II [28] or genetic mutations in the promoter region [31], or as a result of various environmental stimuli (e.g., hypoxia) or drug treatment [32].

The resulting change in m⁶A stoichiometry might significantly influence the binding of m⁶A readers to the cap-proximal region, for example, the binding of YTHDF1/2/3, YTHDC1, and other proteins [4,33]. These m⁶A readers could influence translation, RNA stability, RNA subcellular location, or other functions [4]. In addition, the gain or loss of m⁶A in the 5′ end could function as a regulator of RNA structure in the 5′ UTR since m⁶A tends to disrupt base pairing [34]. Therefore, our study reveals a new mechanism by which transcription-start site choice can influence gene expression, i.e., by altering m⁶A stoichiometry.

One potential model for reduced methylation near the TSS is that the RNA is too short to be a substrate for the METTL3/METTL14 writer complex. Indeed, a previous

study showed that RNA substrates should be at least an 8–11 nt RNA strand for efficient m⁶A deposition [35]. Methylation rates achieved highest efficiency when the substrate was 13–14 nt long [35]. However, this is unlikely to fully explain the results here. We observed a gradual increase in m⁶A deposition within the first 30–40 nt from the TSS. Therefore, the cap-proximal exclusion zone cannot be simply explained by the size of the RNA substrate adjacent to the cap.

Here, we propose two different but related models to explain the 5' cap proximal exclusion zone. First, similar to the inhibitory effect of the splicing machinery, the cap-binding complex (CBC) and its associated factors [36–38] might also prevent access of the METTL3/METTL14 writer complex to the 5' end of mRNAs. The second model is that the recruitment of the m⁶A writer complex requires a specific transcriptional state, for example, a specific phosphorylation state of the C-terminal domain (CTD) of RNA Pol II [39]. The lack of such a signal during the very beginning of transcription might lead to a lag of m⁶A deposition. Future studies will reveal the precise mechanism of the suppressive effect of the 5' cap proximal exclusion zone.

A limitation in our analysis is that CROWN-seq reads are generally short due to the highly fragmenting conditions of RNA deamination (i.e., 30–60 nt on average). In our dataset, many reads do not contain the entire first exon. As a result, the 3' end of the exon needs to be inferred based on genomic annotations. This read length limitation also prevents us from investigating the exact cap-proximal exclusion zone for every m⁶A site. Future studies might use newer RNA deamination methods with less fragmentation [40] or full-length direct RNA sequencing of nascent RNA to reveal the diversity of m⁶A deposition around the 5' ends.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/genes17060653/s1>. Figure S1. Evaluation of m⁶A call threshold by binomial distribution; Figure S2. Corresponding to Figure 3B; Figure S3. Corresponding to Figure 3E; Figure S4. Cap-proximal inhibition across other human cell lines. Table S1, summary of statistical analyses; Table S2, the A nucleotides transcribed as both m⁶A and m⁶Am; Table S3, the non-conversion background in RNA spike-in transcripts; Table S4, 5' isoform-specific m⁶A stoichiometry in HEK293T cells; Table S5, methylation level and coverage of 5' isoforms of *PCF11* loci chr11:83,157,157.

Author Contributions: J.F.L. analyzed experimental data. J.F.L. and S.R.J. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NIH grants R35 NS111631 (S.R.J.), RM1 HG011563 (S.R.J.), and R01 CA186702 (S.R.J.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The comprehensive dataset of TSS-m⁶A-site pairs is available on Zenodo (10.5281/zenodo.19546831).

Acknowledgments: We thank members of the Jaffrey lab for their comments and suggestions throughout this project. During the preparation of this work, the authors used Gemini 3.5 Flash to rephrase text to enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflicts of Interest: S.R.J. is a scientific founder of, advisor to, and/or owns equity in Chimerna Therapeutics, Lucerna Technologies, and 858 Therapeutics. The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

m ⁶ A	N6-methyladenosine
m ⁶ Am	N6,2'-O-dimethyladenosine
TSS	Transcription start site
5' UTR	5' Untranslated region
ERCC	External RNA Controls Consortium
RNA Pol II	RNA Polymerase II
EJC	Exon junction complex
GLORI	Glyoxal and nitrite-mediated deamination of unmethylated adenosines
CROWN-seq	Conversion Resistance detection On Whole-transcriptomic transcription-start N6,2'-O-dimethyladenosine by sequencing
CTD	C-terminal domain
CENPW	Centromere protein W
TIMMDC1	Translocase of Inner Mitochondrial Membrane Domain Containing 1
XIST	X Inactive Specific Transcript
CBC	Cap-binding complex
uORF	Upstream open reading frame
FDR	False discovery rate
BH	Benjamini–Hochberg
PMF	Probability mass function
BAM	Binary alignment map
CI	Confidence interval

References

1. He, B.; Chen, Y.; Yi, C. Quantitative mapping of the mammalian epitranscriptome. *Curr. Opin. Genet. Dev.* **2024**, *87*, 102212. [[CrossRef](#)]
2. Sendinc, E.; Shi, Y. RNA m6A methylation across the transcriptome. *Mol. Cell* **2023**, *83*, 428–441. [[CrossRef](#)]
3. Wang, Z.; Zhou, J.; Zhang, H.; Ge, L.; Li, J.; Wang, H. RNA m(6) A methylation in cancer. *Mol. Oncol.* **2023**, *17*, 195–229. [[CrossRef](#)]
4. Jiang, X.; Liu, B.; Nie, Z.; Duan, L.; Xiong, Q.; Jin, Z.; Yang, C.; Chen, Y. The role of m6A modification in the biological functions and diseases. *Signal Transduct. Target. Ther.* **2021**, *6*, 74. [[CrossRef](#)]
5. Linder, B.; Grozhik, A.V.; Olarerin-George, A.O.; Meydan, C.; Mason, C.E.; Jaffrey, S.R. Single-nucleotide-resolution mapping of m⁶A and m6Am throughout the transcriptome. *Nat. Methods* **2015**, *12*, 767–772. [[CrossRef](#)]
6. Meyer, K.D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C.E.; Jaffrey, S.R. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **2012**, *149*, 1635–1646. [[CrossRef](#)] [[PubMed](#)]
7. Yang, X.; Triboulet, R.; Liu, Q.; Sendinc, E.; Gregory, R.I. Exon junction complex shapes the m(6)A epitranscriptome. *Nat. Commun.* **2022**, *13*, 7904. [[CrossRef](#)] [[PubMed](#)]
8. Uzonyi, A.; Dierks, D.; Nir, R.; Kwon, O.S.; Toth, U.; Barbosa, I.; Burel, C.; Brandis, A.; Rossmannith, W.; Le Hir, H.; et al. Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. *Mol. Cell* **2023**, *83*, 237–251.e237. [[CrossRef](#)] [[PubMed](#)]
9. 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. Exon architecture controls mRNA m(6)A suppression and gene expression. *Science* **2023**, *379*, 677–682. [[CrossRef](#)]
10. Luo, Z.; Ma, Q.; Sun, S.; Li, N.; Wang, H.; Ying, Z.; Ke, S. Exon-intron boundary inhibits m6A deposition, enabling m6A distribution hallmark, longer mRNA half-life and flexible protein coding. *Nat. Commun.* **2023**, *14*, 4172. [[CrossRef](#)]
11. Murakami, S.; Jaffrey, S.R. Hidden codes in mRNA: Control of gene expression by m⁶A. *Mol. Cell* **2022**, *82*, 2236–2251. [[CrossRef](#)]
12. Liu, C.; Sun, H.; Yi, Y.; Shen, W.; Li, K.; Xiao, Y.; Li, F.; Li, Y.; Hou, Y.; Lu, B.; et al. Absolute quantification of single-base m(6)A methylation in the mammalian transcriptome using GLORI. *Nat. Biotechnol.* **2023**, *41*, 355–366. [[CrossRef](#)] [[PubMed](#)]
13. Lane, R.E.; Calcutt, E.; Srinivasan, A.; Gamble, V.; Oppermann, U.; Sun, J.; Cribbs, A.P. Distinct 5' and 3' Coverage Biases Shape Transcriptome Interpretation in Nanopore Direct RNA versus PCR-cDNA Sequencing. *bioRxiv* **2025**. [[CrossRef](#)]
14. Liu, J.F.; Hawley, B.R.; Nicholson, L.S.; Jaffrey, S.R. Decoding m6Am by simultaneous transcription-start mapping and methylation quantification. *eLife* **2025**, *13*, RP104139. [[CrossRef](#)]
15. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S.R.; Qian, S.B. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **2015**, *526*, 591–594. [[CrossRef](#)]

16. Cieřla, M.; Ngoc, P.C.T.; Muthukumar, S.; Todisco, G.; Madej, M.; Fritz, H.; Dimitriou, M.; Incarnato, D.; Hellström-Lindberg, E.; Bellodi, C. m(6)A-driven SF3B1 translation control steers splicing to direct genome integrity and leukemogenesis. *Mol. Cell* **2023**, *83*, 1165–1179.e1111. [[CrossRef](#)]
17. Chen, H.; Zhang, M.; Li, J.; Liu, M.; Cao, D.; Li, Y.Y.; Yamashita, T.; Nio, K.; Tang, H. BMP9-ID1 Pathway Attenuates N(6)-Methyladenosine Levels of CyclinD1 to Promote Cell Proliferation in Hepatocellular Carcinoma. *Int. J. Mol. Sci.* **2024**, *25*, 981. [[CrossRef](#)]
18. Wang, J.; Yin, H.; Li, G.; Wu, D.; Xu, Y.; Chen, Y.; Wang, X.; Xing, Y.; Zhang, T.; Fei, D.; et al. METTL14 promotes neuroblastoma formation by inhibiting YWHAH via an m6A-YTHDF1-dependent mechanism. *Cell Death Discov.* **2024**, *10*, 186. [[CrossRef](#)] [[PubMed](#)]
19. Bonfield, J.K.; Marshall, J.; Danecek, P.; Li, H.; Ohan, V.; Whitwham, A.; Keane, T.; Davies, R.M. HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* **2021**, *10*, giab007. [[CrossRef](#)]
20. Huang, T.; Chen, W.; Liu, J.; Gu, N.; Zhang, R. Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat. Struct. Mol. Biol.* **2019**, *26*, 380–388. [[CrossRef](#)] [[PubMed](#)]
21. Ramaswami, G.; Zhang, R.; Piskol, R.; Keegan, L.P.; Deng, P.; O'Connell, M.A.; Li, J.B. Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **2013**, *10*, 128–132. [[CrossRef](#)] [[PubMed](#)]
22. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)]
23. Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. *Scipy* **2010**, *7*, 92–96.
24. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)] [[PubMed](#)]
25. Akichika, S.; Hirano, S.; Shichino, Y.; Suzuki, T.; Nishimasu, H.; Ishitani, R.; Sugita, A.; Hirose, Y.; Iwasaki, S.; Nureki, O.; et al. Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* **2019**, *363*, eaav0080. [[CrossRef](#)] [[PubMed](#)]
26. Yan, B.; Tzertzinis, G.; Schildkraut, I.; Ettwiller, L. Comprehensive determination of transcription start sites derived from all RNA polymerases using ReCappable-seq. *Genome Res.* **2022**, *32*, 162–174. [[CrossRef](#)]
27. Leppke, K.; Das, R.; Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 158–174. [[CrossRef](#)]
28. Alfonso-Gonzalez, C.; Hilgers, V. (Alternative) transcription start sites as regulators of RNA processing. *Trends Cell Biol.* **2024**, *34*, 1018–1028. [[CrossRef](#)]
29. Zhang, P.; Dimont, E.; Ha, T.; Swanson, D.J.; Hide, W.; Goldowitz, D. Relatively frequent switching of transcription start sites during cerebellar development. *BMC Genom.* **2017**, *18*, 461. [[CrossRef](#)]
30. Duttke, S.H.; Guzman, C.; Chang, M.; Delos Santos, N.P.; McDonald, B.R.; Xie, J.; Carlin, A.F.; Heinz, S.; Benner, C. Position-dependent function of human sequence-specific transcription factors. *Nature* **2024**, *631*, 891–898. [[CrossRef](#)]
31. Schor, I.E.; Degner, J.F.; Harnett, D.; Cannavò, E.; Casale, F.P.; Shim, H.; Garfield, D.A.; Birney, E.; Stephens, M.; Stegle, O.; et al. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat. Genet.* **2017**, *49*, 550–558. [[CrossRef](#)]
32. Watt, K.; Dauber, B.; Szkop, K.J.; Lee, L.; Jovanovic, P.; Chen, S.; Palia, R.; Vassalakis, J.A.; Cooper, T.T.; Papadopoli, D.; et al. Epigenetic alterations facilitate transcriptional and translational programs in hypoxia. *Nat. Cell Biol.* **2025**, *27*, 1965–1981. [[CrossRef](#)]
33. Zaccara, S.; Jaffrey, S.R. A Unified Model for the Function of YTHDF Proteins in Regulating m(6)A-Modified mRNA. *Cell* **2020**, *181*, 1582–1595.e1518. [[CrossRef](#)]
34. Roost, C.; Lynch, S.R.; Batista, P.J.; Qu, K.; Chang, H.Y.; Kool, E.T. Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J. Am. Chem. Soc.* **2015**, *137*, 2107–2115. [[CrossRef](#)] [[PubMed](#)]
35. Qi, S.; Mota, J.; Chan, S.-H.; Villarreal, J.; Dai, N.; Arya, S.; Hromas, R.A.; Rao, M.K.; Corrêa, I.R., Jr.; Gupta, Y.K. RNA binding to human METTL3-METTL14 restricts N6-deoxyadenosine methylation of DNA in vitro. *eLife* **2022**, *11*, e67150. [[CrossRef](#)] [[PubMed](#)]
36. Schulze, W.M.; Cusack, S. Structural basis for mutually exclusive co-transcriptional nuclear cap-binding complexes with either NELF-E or ARS2. *Nat. Commun.* **2017**, *8*, 1302. [[CrossRef](#)]
37. Hallais, M.; Pontvianne, F.; Andersen, P.R.; Clerici, M.; Lener, D.; Benbahouche Nel, H.; Gostan, T.; Vandermoere, F.; Robert, M.C.; Cusack, S.; et al. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1358–1366. [[CrossRef](#)] [[PubMed](#)]
38. Giacometti, S.; Benbahouche, N.E.H.; Domanski, M.; Robert, M.-C.; Meola, N.; Lubas, M.; Bukenborg, J.; Andersen, J.S.; Schulze, W.M.; Verheggen, C.; et al. Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. *Cell Rep.* **2017**, *18*, 2635–2650. [[CrossRef](#)]

39. Appel, L.M.; Franke, V.; Bruno, M.; Grishkovskaya, I.; Kasiliauskaite, A.; Kaufmann, T.; Schoeberl, U.E.; Puchinger, M.G.; Kostrhon, S.; Ebenwaldner, C.; et al. PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat. Commun.* **2021**, *12*, 6078. [[CrossRef](#)]
40. Sun, H.; Lu, B.; Zhang, Z.; Xiao, Y.; Zhou, Z.; Xi, L.; Li, Z.; Jiang, Z.; Zhang, J.; Wang, M.; et al. Mild and ultrafast GLORI enables absolute quantification of m6A methylome from low-input samples. *Nat. Methods* **2025**, *22*, 1226–1236. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.