

Transcriptome-wide identification of 5-methylcytosine by deaminase and reader protein-assisted sequencing

Reviewed Preprint

v2 • October 11, 2024

Revised by authors

Reviewed Preprint

v1 • June 11, 2024

Jiale Zhou, Ding Zhao, Jinze Li, Deqiang Kong, Xiangrui Li, Renquan Zhang, Yuru Liang, Xun Gao, Yuqiang Qian, Di Wang, Jiahui Chen, Liangxue Lai , Yang Han , Zhanjun Li 

Key Laboratory of Zoonosis Research, Ministry of Education, College of Veterinary Medicine, Jilin University, Changchun, China • Laboratory Animal Center, College of Animal Science, Jilin University, Changchun, China • Guangzhou Regenerative Medicine and Health Guang Dong Laboratory (GRMH-GDL), Guangzhou, China • Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

5-Methylcytosine (m^5C) is one of the major post-transcriptional modifications in mRNA and is highly involved in the pathogenesis of various diseases. However, the capacity of existing assays for accurately and comprehensively transcriptome-wide m^5C mapping still needs improvement. Here, we develop a detection method named DRAM (deaminase and reader protein assisted RNA methylation analysis), in which deaminases (APOBEC1 and TadA-8e) are fused with m^5C reader proteins (ALYREF and YBX1) to identify the m^5C sites through deamination events neighboring the methylation sites. This antibody-free and bisulfite-free approach provides transcriptome-wide editing regions which are highly overlapped with the publicly available BS-seq datasets and allows for a more stable and comprehensive identification of the m^5C loci. In addition, DRAM system even supports ultra-low input RNA (10ng) and monitors the dynamic accumulation of cellular m^5C . We anticipate that the DRAM system could pave the way for uncovering further biological functions of m^5C modifications.

eLife Assessment

This potentially **useful** study introduces an orthogonal approach for detecting RNA modification, without chemical modification of RNA, which often results in RNA degradation and therefore loss of RNA molecules. While the authors have improved the work compared to a previous version, uncertainty regarding false positive and false negative rates leave the evidence for the broad applicability of the method **incomplete**. If properly validated, the approach might be of particular interest for sites where modifications are rare.

<https://doi.org/10.7554/eLife.98166.2.sa2>

Introduction

Epigenetics refers to stable inheritance without changing the basic sequence of DNA, involving various forms such as DNA methylation, histone modification and RNA modification. In recent years, RNA sequencing technology has boosted research on RNA epigenetics. More than 170 RNA modifications have been identified, mainly including m⁶A, m⁵C, m¹A, m⁷G and others ^{1,2}. Notably, RNA m⁵C methylation represents a crucial post-transcriptional modification observed across different RNA types, such as tRNA, mRNA, rRNA, vault RNA, microRNA, long non-coding RNA and enhancer RNA ³⁻⁸. Numerous studies have revealed multiple molecular functions of m⁵C in numerous key stages of RNA metabolism, such as mRNA stability, translation, and nuclear export ^{5,9-13}. The dynamic alterations of m⁵C play integral roles in many physiological and pathological processes, such as early embryonic development ¹⁴, neurodevelopmental disorder ^{15,16} and multifarious tumorigenesis and migration ¹⁷⁻²⁰. Moreover, this modification significantly contributes to the regulation of gene expression ^{5,9-13,17}. Therefore, the detection of m⁵C sites appears to be essential for understanding their underlying effects on cellular function and disease states.

With the recent advances in sequencing techniques, several high-throughput assays have been developed for qualitative or quantitative analysis of m⁵C. To date, bisulfite-sequencing (BS-seq) has been proven to be the gold standard method for RNA m⁵C methylation analysis ^{5,21,22}. This approach chemically deaminates unmethylated cytosine to uracil, while keeping methylated cytosine unchanged. The m⁵C methylation sites can be identified by subsequent library construction and sequencing. However, bisulfite treatment of BS-seq is extremely detrimental to RNA, thus resulting in unstable detection of m⁵C in low abundance RNA or highly structured RNA, which directly affects the confidence of results ^{23,24}. Another major type of global m⁵C analysis depends on antibody-assisted immunoprecipitation of m⁵C methylated RNAs, such as m⁵C-RIP-seq ²⁵⁻²⁷, AZA-IP-seq ²⁸ or miCLIP-seq ⁷. These methods are unable to recognize methylation on mRNAs with low abundance and secondary structure. Moreover, these methods are highly dependent on antibody specificity, which usually leads to unspecific binding of RNA and a low amount of m⁵C-modified regions. Moreover, TAWO-seq, originally developed for the identification of hm⁵C, is also capable of m⁵C analysis, but it highly depends on the oxidation efficiency of perovskite, which usually causes false positives and unstable conversion ^{29,30}. Furthermore, the emerging third-generation sequencing, such as Nanopore-seq, can directly map m⁵C by tracking the characteristic changes of bases, but it still faces challenges of a high error rate ³¹⁻³³. These together largely hamper its wide application on transcriptome profiling of m⁵C (Supplementary Table 1). Hence, there is an urgent need for a simple, efficient, sensitive, and antibody-independent method for global m⁵C detection.

The RNA-binding protein ALYREF is the initially recognized nuclear m⁵C reader that binds directly to m⁵C sites in mRNA and plays key roles in promoting mRNA nuclear export or tumor progression ⁵. Another well-known m⁵C reader, YBX1, can also recognize m⁵C-modified mRNA through its cold-shock domain and participates in a variety of RNA-dependent events such as mRNA packaging, mRNA stabilization and translational regulation ^{9,18}. RNA affinity chromatography and mass spectrometry analyses using biotin-labelled oligonucleotides with or without m⁵C were performed in previous reports, which indicated that ALYREF and YBX1 had a more prominent binding ability to m⁵C-modified oligonucleotides ^{5,18}. YBX1 can preferentially recognize mRNAs with m⁵C modifications via key amino acids W65-N70 (WFNVRN) ¹⁸, while K171 is essential for the specific binding of ALYREF to m⁵C sites ⁵. Previous studies have shown that mutations in key amino acids responsible for recognising m⁵C binding in ALYREF and YBX1 lead to a significant reduction in their binding levels to m⁵C-containing oligonucleotides ^{5,18}. Nucleic acid deaminases, primarily categorized as cytosine deaminases and adenine deaminases, are zinc-dependent enzymes which facilitate the deamination of cytosine or adenine within DNA

or RNA substrates³⁴, APOBEC1, an evolutionarily conserved family member of APOBEC proteins, can specifically catalyze the deamination of cytosine in single-stranded RNA (ssRNA) or DNA (ssDNA) to uracil^{35–37}. TadA8e is an adenine deaminase optimized through re-engineering of TadA and it induces conversion of adenine to inosine (eventually read as guanine by transcriptases) in ssRNA or ssDNA^{38,39}. APOBEC1 and TadA8e, with their prominent deamination efficiency, have been employed for the development of precise and efficient base editors such as CBE and ABE8e, which find widespread application in studies related to genome editing^{37,38}.

Here we aim to establish a deaminase and m⁵C reader-assisted RNA methylation sequencing approach (DRAM-seq), which identifies the m⁵C sites through reader-mediated recognitions and deaminase-mediated point mutations neighboring the m⁵C methylation sites. This bisulfite-free and antibody-free method is anticipated to provide more comprehensive and cost-effective transcriptome-wide detection of m⁵C methylation, which may better assist on exploring its further regulatory mechanisms.

Results

Development of DRAM system for m⁵C detection

Our sequencing platform is inspired by the concept of the m⁶A DART-seq assay, in which C near the m⁶A site is converted into U without affecting sequences near non-m⁶A sites⁴⁰. Therefore, we hypothesized that, by utilizing the targeted binding of m⁵C readers, deaminase can be recruited to achieve deamination of cytosine or adenine in the vicinity of the m⁵C sites on single-stranded RNA, thereby facilitating the detection of the m⁵C site. This approach was named DRAM (deaminase and m⁵C reader-assisted RNA methylation sequencing). As RNA-binding proteins, ALYREF and YBX1 also could bind to RNAs without m⁵C modification^{5,18}. To exclude the false-positive detection of DRAM due to the non-m⁵C specific binding of ALYREF and YBX1, knockout of W65-N70 (WFNVRN) amino acids in YBX1 and K171A mutation in ALYREF were introduced separately, resulting in the DRAM^{mut} system (Fig. S1A-S1D). DRAM^{mut} exhibits substantially reduced binding to m⁵C in RNA pulldown assays (Fig. S1D-S1F). To confirm the recognition of m⁵C site by DRAM system, DRAM, DRAM^{mut} and Deaminase system were transfected into the human HEK293T cells, respectively. Finally, we considered the presence of m⁵C modification in the vicinity only if the deamination changes produced under DRAM induction were significantly different from those produced under DRAM^{mut} or Deaminase induction (Fig. 1A).

Previous studies have indicated that there is no uniform intrinsic signature motif sequence that can characterize all m⁵C sites^{5,26,41,42}. To comprehensively detect the m⁵C loci, the readers of m⁵C (ALYREF and YBX1) were separately fused to the C-terminus of the deaminases (APOBEC1 and TadA-8e), namely DRAM-ABE and DRAM-CBE system (Fig. 1B).

DRAM detection system is assayed in an m⁵C-dependent form

To confirm the recognition of m⁵C site by DRAM system, DRAM, DRAM^{mut} and Deaminase were transfected into the human HEK293T cells, respectively. To evaluate candidate DRAM constructs within a cellular environment, we performed fluorescence microscopy to analyze the expression of DRAM. The results showed that DRAM-ABE and DRAM-CBE were properly expressed in HEK293T cells (Fig. S2A-S2B). In addition, flow cytometry displayed ~60% of cells were GFP-positive (Fig. S2C). Two previously reported m⁵C sites in RPSA and AP5Z1 were selected for the analysis^{5,21}, and their methylation status was verified by bisulfite sequencing PCR. The deep sequencing results showed that the m[C fraction of RPSA and SZRD1 was 75.5% and 27.25%, respectively (Fig. 2A and B). Sanger sequencing following RT-PCR was then performed to determine the editing of neighbouring m⁵C sites by DRAM system in these two mRNA. Notably,

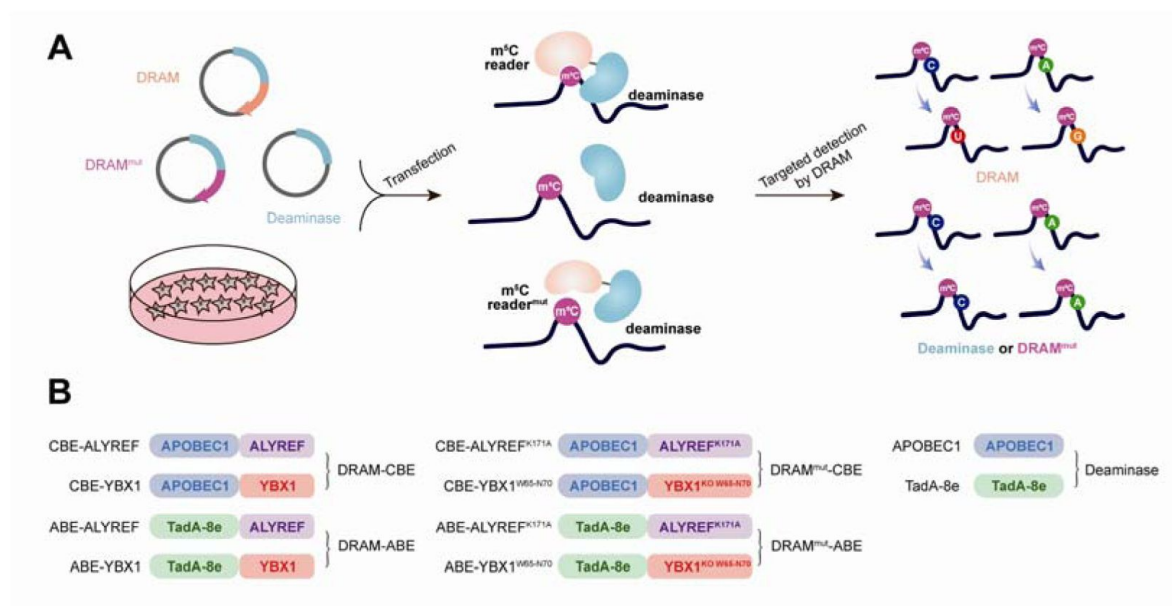


Fig. 1

Development of DRAM system for m⁵C detection.

(A) Schematic diagram of the DRAM assay. DRAM, DRAM^{mut} and Deaminase system were transfected into HEK293T cells separately. After DRAM transfection, the deaminase was directed by m⁵C reader to the vicinity of the m⁵C site and induce C-to-U/A-to-G mutations, whereas transfection of the DRAM^{mut} or Deaminase system failed to effectively induce similar mutations due to the absence of the m⁵C-recognition-binding domain.

(B) The overall design of DRAM, DRAM^{mut} and Deaminase system.

adenine close to the m⁵C site in RPSA mRNA was mutated into guanine, resulting in an A-to-G editing rate of 14.7% by DRAM-ABE, whereas this was rarely observed with TadA-8e or DRAM^{mut}-ABE (**Fig. 2C**). DRAM-CBE induced C to U editing in the vicinity of the m⁵C site in AP5Z1 mRNA, with 13.6% C-to-U editing, while this effect was significantly reduced with APOBEC1 or DRAM^{mut}-CBE (**Fig. 2D**). Subsequently, in order to investigate whether the DRAM system can detect other types of RNA, such as tRNA, 28S rRNA, or others, we performed PCR amplification of the flanking sequences of the m⁵C sites 3782 and 4447 on 28S rRNA and several m⁵C sites on tRNA, such as the m⁵C48 and m⁵C49 sites of tRNA^{Val}, the m⁵C48 and m⁵C49 sites of tRNA^{Asp}, and the m⁵C48 site of tRNA^{Lys}. But Sanger sequencing showed that there was no valid A-to-G/C-to-U mutation detected, which is most likely due to the fact that ALYTEF and YBX1 are mainly responsible for the mRNA m⁵C binding proteins, and thus the DRAM system is more suitable for the mRNA m⁵C detection (**Fig. S3**). Taken together, the fusion of m⁵C reader and deaminase can effectively and selectively deaminate cytosine/adenine in the vicinity of the mRNA m⁵C sites.

NSUN2⁴³ and NSUN6⁴⁴, two family members of NOL1/NSUN protein, were both identified as m⁵C methyltransferase of mRNA⁴⁵. To verify that the detection of DRAM occurs in the presence of m⁵C, we individually depleted NSUN2 and NSUN6 in HEK293T cells and performed DRAM transfection. The knockout efficiency has been confirmed by western blotting (**Fig. 2E, 2F** and **Fig. S4A, 4B**). It has been previously demonstrated that m⁵C methylation of AP5Z1 and RPSA is catalyzed by NSUN2 and NSUN6, respectively^{21, 46}. In line with this, sanger sequencing following RT-PCT showed a significant reduction in C-to-U or A-to-G mutations near the m⁵C sites in methyltransferase-deficient cells compared with WT cells (**Fig. 2G and H**). Overall, these findings suggest that the DRAM detection system is assayed in an m⁵C-dependent form.

DRAM enables transcriptome-wide analysis of m⁵C methylation

Subsequently, we performed RNA-seq analysis after DRAM transfection by detecting C-to-U/A-to-G editing events to accomplish transcriptome-wide detection of m⁵C (**Fig. 3A**). To serve as positive controls, two previously published BS-seq datasets were also integrated^{5, 21}. Mutations were detected near the m⁵C site in RPSA as A-to-G by DRAM-ABE (**Fig. 3B**), and DRAM-CBE detected the presence of C-to-U mutations near the AP5Z1 m⁵C site (**Fig. 3C**). However, the DRAM^{mut} and Deaminase systems induced few effective mutations close to these sites. Examination of multiple reported high-confidence RNA m⁵C sites showed that DRAM-seq editing events were also enriched in the vicinity of the BS-seq sites (**Fig. 3B, 3C** and **Fig. S5**).

DRAM-seq analysis further confirmed that DRAM was detected in an m⁵C-dependent manner and that mutations in AP5Z1 and RPSA mRNA were reduced in methyltransferase knockout cells compared to wild-type cells (**Fig. 3D, 3E**). Moreover, the knockout cells exhibited overall rare DRAM-seq editing events close to m⁵C sites in other mRNAs (**Fig. S6**). Unfortunately however due to the inability of the DRAM detection system to achieve m⁵C detection with single base resolution, we did not find sequence preferences for the m⁵C sites responsible for catalysis by each of the two methyltransferases NSUN2 and NSUN6 (**Fig. S4D**).

A comparison of three biological replicates from each experimental group revealed a strong reproducibility of A-to-G/C-to-U mutations in HEK293T cells expressing DRAM-ABE and DRAM-CBE (**Fig. S7**). Moreover, the integration and analysis of the experimental data revealed a high degree of overlap between the three biological replicates (**Fig. S4C**). And a recent study by Wang et al. showed that ALYREF deletion affects the expression of 94 mRNAs⁴⁷, and only 55.32% of these ALYREF-regulated mRNAs can be detected by the DRAM system (**Fig. S4E**). These findings suggest that DRAM selectively targets specific RNAs for editing, exhibiting a high degree of consistency across samples.

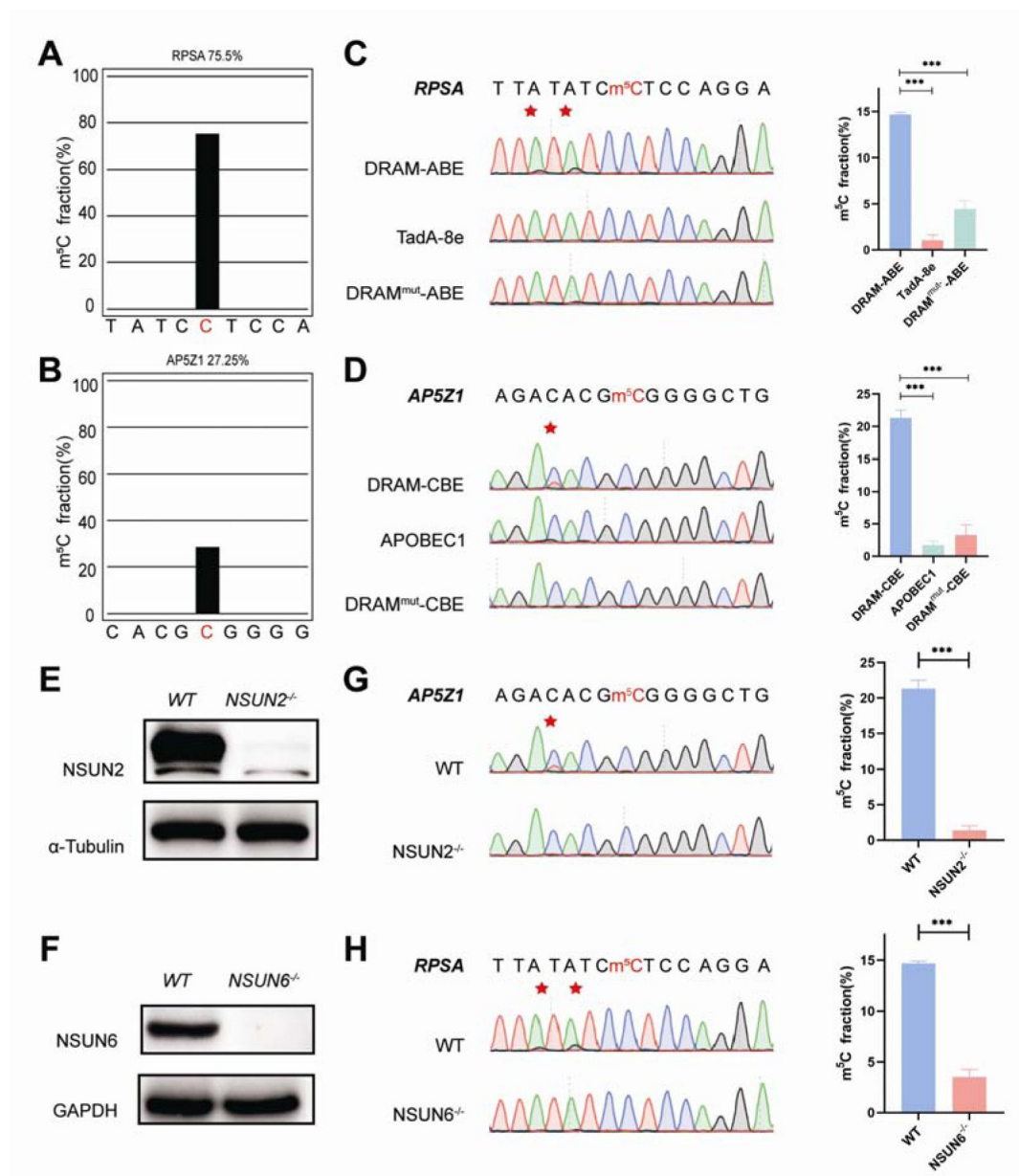


Fig. 2

DRAM detection system was assayed in an m⁵C -dependent form.

(A, B) Two m⁵C sites from RPSA (A) and AP5Z1 (B) mRNA detected by deep sequencing of bisulfite sequencing PCR in HEK293T cells. The m⁵C sites are highlighted by red color. The m⁵C fraction of RPSA and AP5Z1 were 75.5% and 27.25% (The number of reads is greater than 1000).

(C, D) Sanger sequencing following RT-PCR verified two m⁵C sites from RPSA (C) and AP5Z1 (D) mRNAs in DRAM-transfected HEK293T cells, respectively. HEK293T cells only expressing DRAM^{mut} or Deaminase were served as negative controls. The left panel illustrates the location of DRAM induced mutation sites, which is highlighted in red asterisk. The right panel shows the corresponding quantification of sanger sequencing.

(E, F) The knockout efficiency of NSUN2 (E) and NSUN6 (F) in HEK293T cell lines verified by Western blotting. The protein level of α-Tubulin and GAPDH were served as loading controls, separately.

(G, H) DRAM induced mutations close to m⁵C sites in AP5Z1 (G) and RPSA (H) mRNAs after NSUN2 and NSUN6 knockout in HEK293T cells. The left panel illustrates the location of DRAM induced mutation sites, which is highlighted in red asterisk. The right panel shows the corresponding quantification of sanger sequencing.

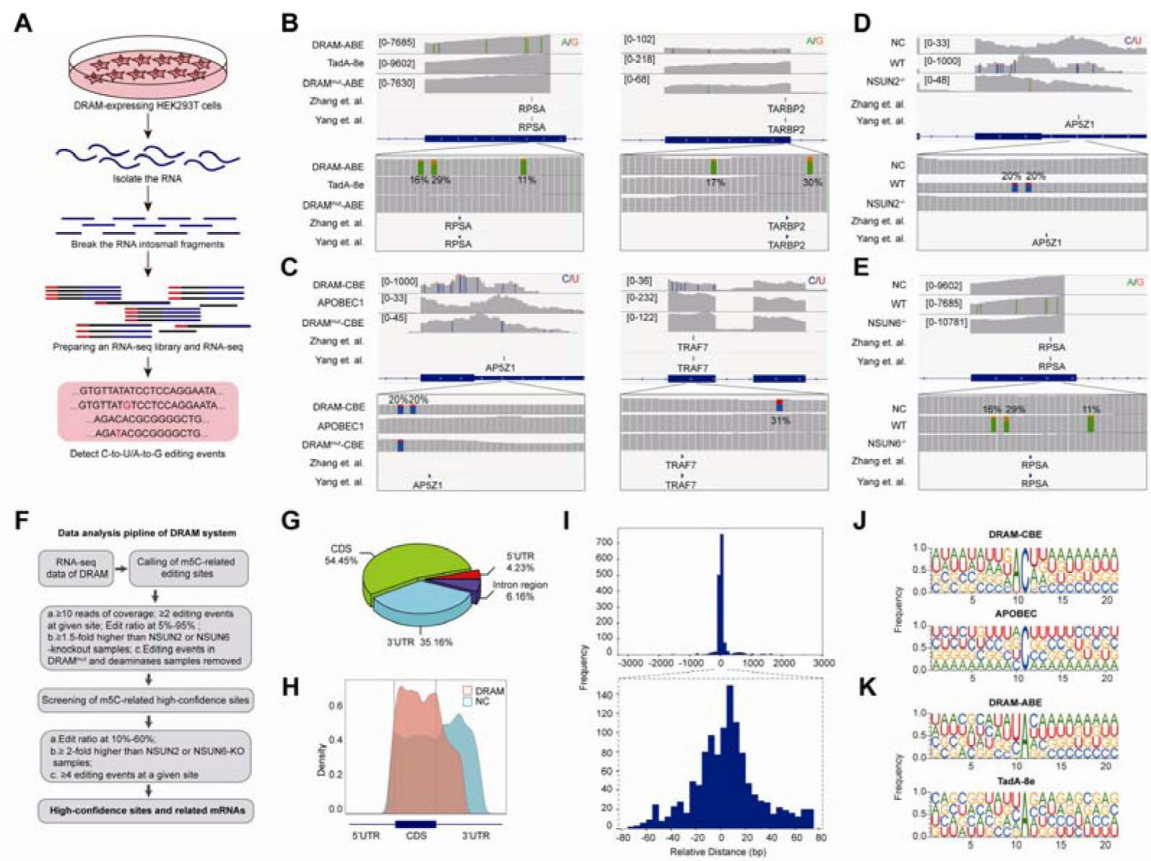


Fig. 3

DRAM enables transcriptome-wide analysis of m⁵C methylation.

(A) Schematic of the DRAM-seq method.

(B, C) Integrative genomics viewer (IGV) browser traces of DRAM-seq data expressing the indicated constructs in RPSA (**B**, left panel), TARBP2 (**B**, right panel), AP5Z1 (**C**, left panel), and TRAF7 (**C**, left panel) mRNAs. C-to-U or A-to-G mutations found in at least 10% of reads are indicated by coloring. The previously published RNA BS-seq datasets from two individual studies were displayed as panel “Yang et al.” and “Zhang et al.”. (n(DRAM)=3 independent samples, n(Deaminase)=2 independent samples, and n(DRAM^{mut})=1 independent sample.)

(D, E) Integrative genomics viewer (IGV) browser traces of DRAM-seq data in wildtype and methyltransferases knockout cells in AP5Z1 (**D**) and RPSA (**E**) mRNAs. C-to-U or A-to-G mutations were found in at least 10% of reads are indicated by coloring. The previously published RNA BS-seq datasets from two individual studies were displayed as panel “Yang et al.” and “Zhang et al.”. n=3 independent samples.

(F) Screening process for DRAM-seq assays and principles for screening high-confidence genes.

(G) The pie chart shows the distribution of editing sites in different transcript region in cells expressing DRAM (n=3 independent samples).

(H) The density map showing the distribution of editing events across the mRNA transcripts detected by DRAM-seq.

(I) The frequency plot shows the distribution of the distances of edit events in DRAM-seq relative to the m⁵C sites from the published BS-seq datasets. The position of each m⁵C site of BS-seq is determined as 0, and the relative distance of each site to the nearest edit event in DRAM-seq is calculated and plotted. The plots are presented separately based on the cutoff of upstream and downstream 3000bp (above) and 80bp (below) windows.

(J, K) Motif analysis discovered within the ±20nt region around the C-to-U or A-to-G editing site in cells expressing DRAM-CBE (**J**), APOBEC1 (**J**), DRAM-ABE (**K**) and TadA-8e (**K**).

To obtain information on a set of high-confidence DRAM-seq data, we filtered the list of sites transfected with deaminase alone and screened the sequencing results with methyltransferase depleted, pooled editing events occurring in at least 10% of reads across multiple samples to obtain a set of high-confidence editing sites (**Fig. 3F** and Supplementary Table 2), and integrated genes with editing sites occurring in DRAM-ABE and DRAM-CBE (**Fig. 3F** and Supplementary Table 3).

Previous studies have indicated that m⁵C sites are predominantly distributed in the coding sequences (CDS) and notably enriched near the initiation codon^{5,25,26,48–50}. To further delineate the characteristics of the DRAM-seq data, we compared the distribution of DRAM-seq editing sites within the gene structure, specifically examining their occurrences in the 5′ untranslated region (5′UTR), 3′ untranslated region (3′UTR), CDS and Intergenic/Intron region. Our analysis revealed that DRAM-seq editing events in cells expressing DRAM-ABE and DRAM-CBE were primarily located in the CDS and 3′UTR, indicating a non-random distribution of m⁵C (**Fig. 3G**, Fig. S8A and 8B). Moreover, plotting the distribution of DRAM-seq editing sites in mRNA segments (5′UTR, CDS, and 3′UTR) highlighted a significant enrichment in the CDS (**Fig. 3H**). In contrast, cells expressing the deaminase exhibited a distinct distribution pattern of editing sites, characterized by a prevalence throughout the 3′UTR (**Fig. 3H**). This finding reaffirms that the specific editing pattern observed in DRAM-seq across the transcriptome depends on its capacity to bind m⁵C.

Comparative analysis of the DRAM-seq editing sites with the previously published BS-seq m⁵C sites indicated that the likelihood of editing was notably higher in closer proximity to the m⁵C sites (**Fig. 3I**). Furthermore, the editing window of DRAM exhibited enrichment approximately 20bp before and after the m⁵C site (**Fig. 3I**). Investigation into the sequences surrounding the editing window revealed that AC motifs were the most significantly enriched in DRAM-CBE, whereas (U/C) A motifs were most notably enriched in DRAM-ABE. In contrast, the APOBEC1 and Tada-8e samples displayed no significantly enriched motifs, with mutations being more randomly orientated (**Fig. 3J, 3K**).

DRAM-seq provides stable and comprehensive identification of m⁵C loci

Subsequently, we then evaluated the ability of DRAM-seq to detect m⁵C across the entire transcriptome and compared its performance to that of the previously reported BS-seq. Although both previous studies employed bisulfite treatment, the resulting data obtained significant discrepancies due to variations in their treatment and analysis methodologies. Our results indicated that DRAM-seq identified the presence of m⁵C modifications covering 79.6% of the genes detected by Yang et al.⁵ and 91.9% of the genes detected by Zhang et al.²¹ (**Fig. 4A** and C). Remarkably, certain pivotal regulators with diverse biological functions, such as ATG16L1 (coordinates autophagy pathway)⁵¹ and ARHGEF25 (plays an important role in actin cytoskeleton reorganization)⁵², were identified by Zhang et al. and DRAM-seq, but not by Yang et al. (**Fig. 4B**). Conversely, FANCD2 (Maintains chromosome stability)⁵³ and RPL15 (components of the large ribosomal subunit)^{54,55}, were discovered by Yang et al. and DRAM-seq, but not by Zhang et al. (**Fig. 4D**). Hence, DRAM-seq appears to offer a more stable and comprehensive identification of the m⁵C loci.

To provide functional insights into m⁵C RNA-modified genes in HEK293T cells, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. These results highlighted the involvement of these genes in the regulation of diverse key biological processes, such as cell division, cell cycle, mRNA splicing, protein processing in the endoplasmic reticulum, nucleocytoplasmic transport, translation, DNA repair and others (**Fig. 4E, 4F**, Fig. S8C and S8D).

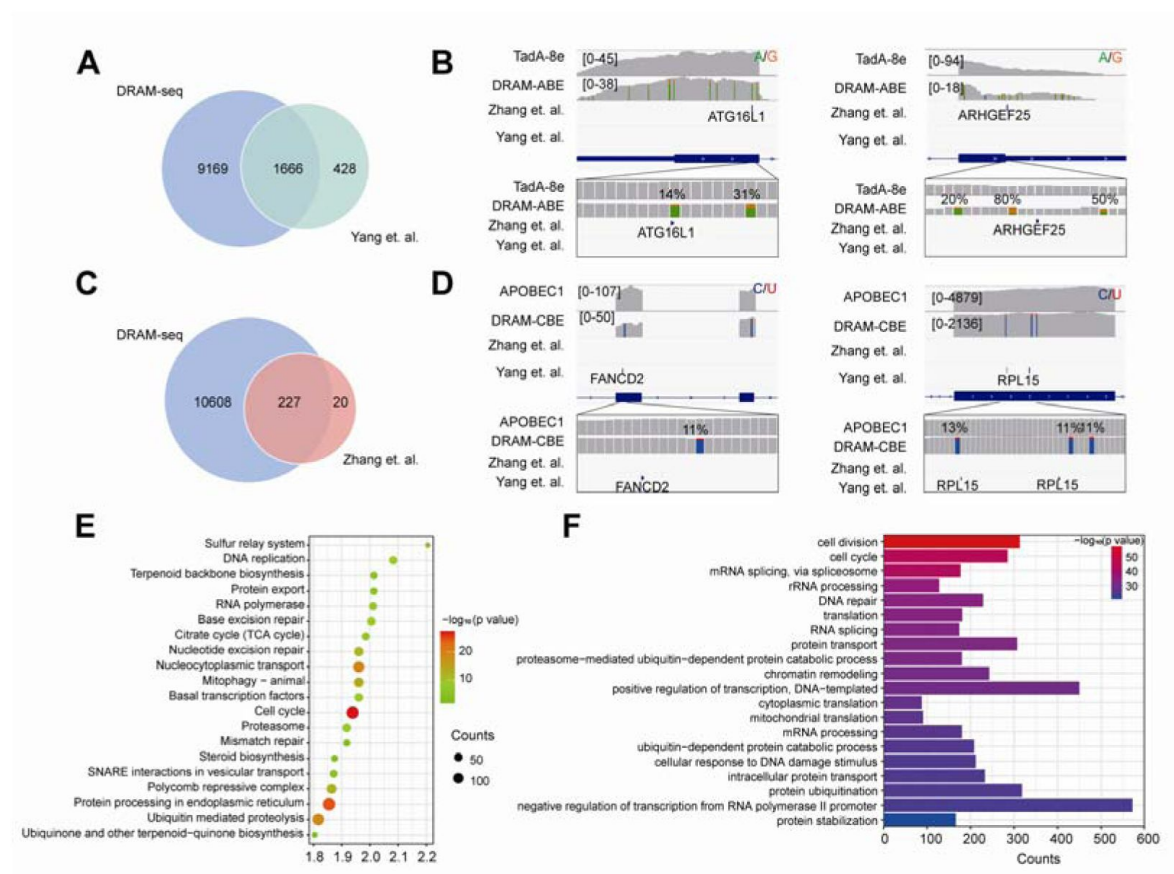


Fig. 4

DRAM monitors m⁵C dynamics in cellular RNA.

(A) Venn diagram showing the overlap between DRAM-seq and Yang et al.'s edited genes.

(C) Venn diagram showing the overlap between DRAM-seq and Zhang et al.'s edited genes.

(B, D) Integrative genomics viewer (IGV) browser traces of DRAM-seq data expressing the indicated constructs in the ATG16L1 (B), ARHGEF25 (B), FANCD2 (D), and RPL15 (D) mRNAs. C-to-U/A-to-G mutations found in at least 10% of reads are indicated by coloring, and the m⁵C site found by BS-seq is also labelled.

(E) Genes with DRAM-seq editing events were analyzed for KEGG bioprocess enrichment.

(F) GO biological processes enrichment analysis of genes with DRAM-seq editing events. Statistical analyses were performed using the DAVID tool.

DRAM monitors m⁵C dynamics in cellular RNA

Next, we aim to determine whether changes in m⁵C can be detected in real time by DRAM. Previous studies have established that oxidative stress induced by 200 μ M sodium arsenite (NaAsO₂) treatment inhibited the expression of NSUN2 and NSUN6, resulting in a gradual decrease in m⁵C methylation^{56,57}. As cells exposed to NaAsO₂ typically exhibited slowed cell cycle progression and NaAsO₂-induced apoptosis, we performed a low concentration screen^{58,59}. After gradient treatment of NaAsO₂ (50 μ M, 100 μ M, 150 μ M, and 200 μ M, separately) in cells, 200 μ M group was identified as the optimal concentration, leading to highly significant decreases of NSUN2 and NSUN6 expression (**Fig.5A**). Subsequently, HEK293T cells expressing DRAM with NaAsO₂ for 3 hours revealed a decreasing trend in the mRNA expression of NSUN2 and NSUN6 (**Fig.5B**). Bisulfite sequencing PCR also verified a decreasing trend in m⁵C levels in selected mRNAs after NaAsO₂ treatment (**Fig. 5C and D**).

Next, we performed DRAM system assays, and Sanger sequencing results showed that induction of the DRAM system after NaAsO₂ treatment resulted in a highly significant reduction of A-to-G/C-to-U mutations on both RPSA and AP5Z1 mRNAs compared to cells without NaAsO₂ treatment (**Fig.5E** and F). Taken together, with our previous evidence that the methyltransferase-deleted cells exhibited overall rare DRAM editing events close to m⁵C sites, we suggest that the DRAM could effectively monitor the accumulation of m⁵C in individual RNAs in response to the alteration of cellular conditions.

DRAM enables low-input m⁵C profiling

A significant challenge in m⁵C detection lies in the specificity of antibodies and the substantial amount of input RNA required for sequencing. RNA is susceptible to degradation during denaturation, sodium bisulfite treatment and desulfurization steps in the BS-seq assay⁶⁰. Immunoprecipitation-based m⁵C assays and LC-MS/MS also impose high demand for sample input^{7,25,61}. Several experiments have highlighted the requirement of 100-500 ng of RNA for m⁵C-RIP-seq, while BS-seq necessitates an even more demanding 750-1000 ng of RNA^{21,25,62}. To assess the detection limits of DRAM-Sanger, we attempted to amplify two representative m⁵C-containing sites in the RPSA and AP5Z1 transcripts from diluted RNA samples.

Remarkably, we successfully generated PCR products of these two mRNAs from cDNAs corresponding to 250 ng, 50 ng, and 10 ng of total RNA. Quantitative analysis by Sanger sequencing demonstrated nearly identical Sanger traces across these dilutions (**Fig.5G** and H). This finding underscores that the specificity of DRAM editing depended on its ability to bind m⁵C, and DRAM is proficient in low-input m⁵C analyses. Furthermore, cell viability was determined by CCK8 assay on HEK293T cells transfected with DRAM (**Fig.5I**). Importantly, there was no significant difference in the relative proliferative capacity of the cells compared to untransfected cells (NC), indicating that DRAM expression did not adversely affect cell viability (**Fig.5J**).

Transfection of the DRAM system in cells results in the transient overexpression of fusion proteins. To investigate how varying expression levels of these proteins influence A-to-G and C-to-U editing within the same m⁵C region, we conducted a gradient transfection using plasmid concentrations of 1500 ng, 750 ng and 300 ng. This approach allowed us to progressively reduce the expression levels of the fusion proteins (Fig. S9A and S9B). Sanger sequencing revealed that the editing efficiency of A-to-G and C-to-U within the m⁵C region significantly decreased as fusion protein expression diminished (Fig. S9C and S9D). These findings suggest that the transfection efficiency of the DRAM system is concentration-dependent and that the ratio of editing efficiency to transfection efficiency may assist in the quantitative analysis of m⁵C using the DRAM system.

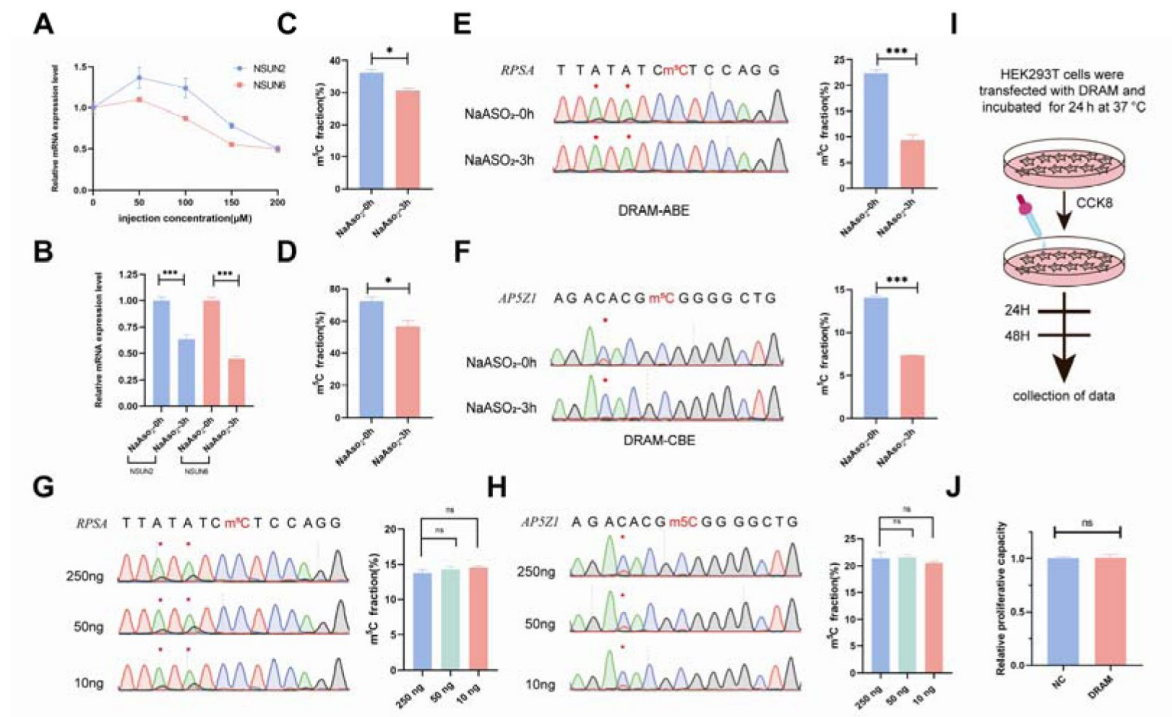


Fig. 5

DRAM monitors m⁵C dynamics in cellular RNA.

- (A) Relative expression levels of NSUN2 and NSUN6 mRNA in HEK293T cells treated with 50 μ M, 100 μ M, 150 μ M and 200 μ M NaAsO₂.
- (B) Relative expression levels of NSUN2 and NSUN6 mRNA in DRAM-expressing HEK293T cells treated with 200 μ M NaAsO₂ for 3 hours.
- (C, D) Bisulfite sequencing PCR showing the m⁵C levels in RPSA and AP5Z1 mRNAs after 3 h of control and NaAsO₂ treatment in cells.
- (E, F) The m⁵C adjacent mutations from RPSA (C) and AP5Z1 (D) mRNA in DRAM-expressing HEK293T cells treated with NaAsO₂. The left panel indicates the sanger sequencing results following RT-PCR, while the corresponding quantifications of DRAM-induced mutations are shown on the right panel.
- (G, H) DRAM analysis of RPSA (G) and AP5Z1 (H) mRNAs with 250 ng, 50 ng, and 10 ng of input RNA. Representative Sanger sequencing plots are shown on the left panel, with mutation sites marked with asterisks. The mutation rates are quantified on the right panel.
- (I) Flowchart illustrating Cell viability analysis by CCK8 reagent after DRAM transfection in HEK293T cells.
- (J) Quantitative comparison of the relative proliferative capacity of DRAM-expressing and untransfected cells.

Discussion

In recent years, m⁵C methylation modifications have received increasing attention, with multiple reports detailing the distribution of RNA m⁵C methylation modifications across various species and tissues, elucidating their characteristics. Despite the relatively low abundance of m⁵C, its highly dynamic changes hold significant implications for the regulation of physiological and pathological processes^{5,21,44}. However, due to the limitations of sequencing methods and the variability of data processing, there remains ample room for progress in the study of m⁵C detection methods.

In this study, we developed a site-specific, depth-sequencing-free m⁵C detection method using DRAM-Sanger. This workflow relies on conventional molecular biology assays such as RT-PCR and Sanger sequencing, eliminating the need for specialized techniques and thereby simplifying the process of m⁵C detection.

DRAM-seq introduces a novel strategy for transcriptome-wide m⁵C detection, overcoming inherent limitations in existing methods. Notably, DRAM-seq covered around 80% of the high-confidence m⁵C-modified genes detected by BS-seq and identified more potential m⁵C sites. This can be attributed to the avoidance of bisulfite treatment by DRAM-seq, preventing RNA damage and ensuring a more comprehensive representation of RNA samples. This feature also likely contributes to the observed stability of DRAM-seq in comparison to BS-seq. Additionally, DRAM-seq is not limited by antibody specificity and is resistant to chemical-induced damage.

A prominent challenge in existing m⁵C profiling methods is their reliance on substantial amounts of input RNA samples. In contrast, DRAM operates through the deamination activity of deamination activity of deaminase, preserving RNA integrity and preventing degradation. The notable advantage of DRAM lies in its capacity for low-input m⁵C detection. Our analysis demonstrates that DRAM requires as low as 10ng of total RNA for m⁵C detection. While DRAM is currently well-suited for detecting m⁵C on a transcriptome-wide scale, the potential for future applications involving third-generation sequencing could extend its utility to individual mRNAs, particularly m⁵C heterogeneity on mRNA splicing variants. In addition, the DRAM system depends on the specific recognition of m⁵C modifications on ssRNA by the reader protein, theoretically avoiding the false-positive effects of 5-hydroxymethylation modifications in other assays, such as BS-seq^{21,23}. This potential feature could enhance the accuracy of the DRAM assay, albeit it still requires careful validation.

In our study, m⁵C detection was performed following the transient transfection of the DRAM detection system into mammalian cells, which might result in a lower mutation rate at the corresponding site. Therefore, employing lentiviral-mediated transfection into cell lines of interest could potentially enhance the efficiency of m⁵C detection. In the present study, the editing efficiency of the DRAM system, which is believed to detect dynamic changes in m⁵C, was significantly reduced following sodium arsenite treatment. However, sodium arsenite can adversely affect all proteins containing thiol groups, impacting not only NSUN proteins but potentially also the function of reader proteins and deaminases. As a result, this validation cannot entirely rule out the possibility that the observed decrease in editing frequency is due to the inactivation of fusion proteins by sodium arsenite. The optimization of this validation may require the development of more specific m⁵C inhibitors in the future. Further elucidation of the key amino acids directing ALYREF and YBX1's binding to m⁵C methylation sites should enable more accurate and sensitive m⁵C detection by DRAM-seq. Due to the lack of a fixed base composition for characterizing all m⁵C modification sites, DRAM has an apparent limitation in achieving single-base resolution for detecting m⁵C. However, our present study proved that the measuring resolution of DRAM is around 40nt, which facilitates higher precision than that of m⁵C-RIP-seq

(~100nt). In the future, with more in-depth analyses of m⁵C reader structures and the identification of new potential m⁵C readers, we expect to achieve more precise m⁵C localization and more comprehensive m⁵C modification detection. Moreover, the substitution of deaminases, such as A3A and A3G (the family members of APOBEC), could also potentially enhance the efficiency of the DRAM detection [63](#)–[65](#).

Following the purification of DRAM fusion protein, achieving *in vitro* detection of RNA m⁵C methylation could broaden the applicability to a more diverse range of samples. Another potential application for DRAM-seq could be the expression of drug-inducible DRAM systems *in vivo* using various animal models for m⁵C analysis. These will together provide novel insights into m⁵C modifications for biological and clinical research.

Conclusions

In summary, we developed a novel deaminase and reader protein-assisted RNA m⁵C methylation approach that detects the m⁵C region by deaminating As or Cs in close proximity to the m⁵C sites, which does not rely on antibodies or bisulfite, thus leading to unprecedentedly comprehensive transcriptome-wide RNA m⁵C methylation profiling. We anticipated that this system could pave the way for uncovering further biological functions of m⁵C modifications and facilitate the development of therapeutic interventions for associated diseases.

Materials and methods

Plasmid construction

ALYREF and YBX1 expression plasmids were purchased from MIAOLING BIOLOGY (<http://www.miaolingbio.com/>), and the ALYREF and YBX1 fragments were then amplified using specific primer. The ALYREF and YBX1 portions were amplified using pCMV-APOBEC1-YTH (Addgene plasmid no. 131636; <https://www.addgene.org/131636/>) and ABE8e (Addgene plasmid no. 138489; <https://www.addgene.org/138489/>) to amplify the deaminase portion and the essential plasmid construct proxies, and finally the fragments were recombined by the ClonExpress Ultra One Step Cloning Kit to complete the plasmid vector construction. Both DRAM^{mut}-ABE and DRAM^{mut}-CBE related vectors were obtained by introducing the corresponding key amino acid mutations using Fast Site-Directed Mutagenesis Kit (TIANGEN Biotech). The primer sequences used are listed in Supplementary Table 4.

Cell culture and plasmid transfection

HEK293T cell line (ATCC) was cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (CLARK BIOSCIENCE) and 1% penicillin (100 U/ml)-streptomycin (100 µg/ml). The cells were seeded in 12-well plates and transfected using Hieff TransTM Liposomal Transfection Reagent (Yeasten).

NSUN2-depleted cell lines were generated by cloning NSUN2-targeting single guide RNA sequences into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene plasmid no. 62988; <http://n2t.net/addgene:62988>). Plasmids were then transfected into HEK293T cells and Puromycin (Meilunbio) was added at a final concentration of 3 µg/ml to enrich the positively transfected cells 24 h after transfection. After 72 h, the cells were collected and used for genotyping by Sanger sequencing. NSUN6-depleted cell lines were generated in the same way. The primers used for genotyping and single guide RNA sequences are listed in Supplemental Table 4.

Cell viability measurements

HEK293T cells were transfected with DRAM plasmid and cultured at 37°C for 24 h. Subsequently, 1000 cells were seeded in 96-well plates. After waiting for the cells to attach to the wall, the cell activity was detected by Cell Counting Kit-8 (Meilunbio). Cell Counting Kit-8 contains WST-8, which in the presence of the electronically coupled reagent 1-Methoxy PMS can be reduced by mitochondrial dehydrogenase to the orange-colored metazan product Formazan, the absorbance of which is measured at 450 nm to analyze cellular activity.

Western blotting

For protein blotting, samples were lysed in RIPA Lysis Buffer (Meilunbio) with Phenylmethanesulfonyl fluoride (PMSF) and the BCA protein assay kit (Beyotime Biotechnology) was used to Protein concentration was measured. Total protein extracts were separated by SDS-PAGE on a 10% gel and then transferred to 0.22 nm polyvinylidene fluoride membranes (Boster). Subsequently, the proteins were probed with specific antibodies after the blot was blocked with 5% non-fat milk (Boster). Images were quantified using ImageJ software and all data are expressed as mean $\pm$ SEM.

The following antibodies and concentrations were used: NSUN2 Polyclonal antibody (Proteintech; Cat No.20854-1-AP; 1:7500), NSUN6 Polyclonal antibody (Proteintech; Cat No. 17240-1-AP; 1:2000), RabbitAnti-GAPDH antibody (Bioss; bs-41373R; 1:2000), Alpha Tubulin Polyclonal antibody (Proteintech; Cat No. 11224-1-AP; 1:2000), HRP-labeled Goat Anti-Rabbit IgG(H+L) (Beyotime Biotechnology; A0208; 1:2000).

cDNA synthesis and Sanger sequencing

Total cellular RNA was extracted with TRIzol reagent (TIANGEN Biotech) and cDNA was synthesized using PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara Bio) according to the manufacturer's recommendations. PCR was then performed using 2×Taq PCR MasterMix II (TIANGEN Biotech) and primers flanking m⁵C target sites, and the purified PCR products were directly sequenced by Sanger sequencing. The Sanger sequencing results were analyzed using EditR 1.0.10 to calculate the mutation frequency⁶⁶. The primers used in this study are shown in Supplemental Table 4.

Real-time quantitative PCR

cDNA was synthesized using FastKing RT kit (with gDNase) (TIANGEN Biotech) according to the manufacturer's recommendations. RT-qPCR assay was performed using SuperReal PreMix Plus (SYBR Green) (TIANGEN Biotech). GAPDH was used as an endogenous control, and the expression levels were normalized to the control and calculated by the $2^{-\Delta\Delta C_t}$ formula. All samples were analyzed in triplicate and each mRNA quantification represents the average of at least three measurements. All data are expressed as mean $\pm$ SEM. The primers used in this study are shown in Supplemental Table 4.

Protein structure modelling

Protein structure simulations were performed using the SWISS-MODEL online website (<https://swissmodel.expasy.org/interactive>)⁶⁷. The SWISS-MODEL database is able to provide up-to-date annotated 3D protein models, which are generated from automated homology modelling of related model organisms and experimental structural information for all sequences in UniProtKB, with reliable structural information, and subsequently protein structure observations were performed using PyMOL⁶⁸.

Bisulfite sequencing PCR

We referenced bisulfite sequencing PCR, an assay established by Matthias Schaefer et al. We chemically deaminated cytosine in RNA using the EZ RNA methylation kit (50) (ZYMO RESEARCH) and then quantified m⁵C methylation levels based on PCR amplification of cDNA combined with deep sequencing²³ [↗](#).

RNA Conversion Reagent was premixed with prepared RNA samples, and the RNA was denatured at 70°C for 5 minutes, followed by a reaction period of 45 minutes at 54°C. Finally, the purified RNA samples were recovered after desulfurization by RNA Desulphonation Buffer. cDNA was synthesized using PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara Bio) according to the manufacturer's recommendations. PCR was then performed using 2× EpiArt HS Taq Master Mix (Dye Plus) (Vazyme) and m⁵C target site-specific Bisulfite Primer (primer sequences were designed at <https://zymoresearch.eu/pages/bisulfite-primer-seeker> [↗](#)), the products were purified by TIANGel Midi Purification Kit (TIANGEN Biotech), and the connectors for second-generation sequencing were attached at both ends of the products for sequencing. Finally, deep sequencing was performed by HiTOM analysis to detect the methylation level (The number of reads >1000 in deep sequencing)⁶⁹ [↗](#). The primers used in this study are shown in Supplemental Table 4.

Library construction and next-generation sequencing

1μg of total cellular RNA was used for sequencing library generation by NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA, Catalog #: E7530L) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEB Next First Strand Synthesis Reaction Buffer(5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs was converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEB Next Adaptor with hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of preferentially 370~420 bp in length, the library fragments were purified with AMPure XP system (Beverly, USA). Then 3 μL USER Enzyme (NEB, USA) was used with size selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent 5400 system(Agilent, USA)and quantified by QPCR (library concentration ≥ 1.5 nM). The qualified libraries were pooled and sequenced on Illumina platforms with PE150 strategy in Novogene Bioinformatics Technology Co., Ltd (Beijing, China), according to effective library concentration and data amount required.

DRAM-seq analysis and calling of edited sites

The raw fastq sequencing data were cleaned by trimming the adapter sequences using Fastp (v0.23.1) and were aligned to the human genome (hg19) using STAR (v2.7.7) in paired-end mode. The aligned BAM files were sorted and PCR duplicates were removed using Samtools (v1.12). The site calling of DRAM-seq was performed using Bullseye, a previously customized pipeline to look for C-to-U or A-to-G edited sites throughout the transcriptome⁴⁰ [↗](#). Briefly, the sorted and deduplicated BAM files were initially parsed by parseBAM.pl script.

Then, Find_edit_site.pl script was employed to find C-to-U or A-to-G editing events by DRAM-seq with at least 10 reads of coverage, an edit ratio of 5%-95%, an edit ratio at least 1.5-fold higher than NSUN2 or NSUN6-knockout samples, and at least 2 editing events at a given site. Sites that were only found in one replicate of each DRAM protein variant were removed. Editing events appeared

in cells expressing merely APOBEC1 or Tada8e were also removed. For high confidence filtering, we further adjusted the Find_edit_site.pl parameters to the edit ratio of 10%-60%, an edit ratio of control samples at least 2-fold higher than NSUN2 or NSUN6-knockout samples, and at least 4 editing events at a given site.

Metagene and motif analyses

Metagene analysis was performed using hg19 annotations according to previously reported tool, MetaplotR⁷⁰. For motif analysis, the 20bp flanking sequence of each DRAM-seq editing site was extracted by Bedtools (v2.30.0)⁷¹. The motif logos were then plotted by WebLogo (v3.7.12)⁷².

Replicates analysis

Independent biological replicates of DRAM-ABE or DRAM-CBE in DRAM-seq analysis were separately compared by computing the Pearson correlation coefficient between the number of C-to-U mutations per mRNA between any two replicate experiments.

GO and KEGG analysis

GO and KEGG analysis of DRAM-seq edited mRNAs was performed using the DAVID bioinformatic database⁷³. GO terms with a P value of less than 0.05 were considered statistically significant.

RNA pulldown assays

RNAs containing C or m⁵C residues and proteins expressing DRAM or DRAM^{mut} were prepared in advance, processed using the PierceTM Magnetic RNA-Protein Pull-Down Kit (Thermo) and finally analysed by Western blotting.

Statistical analysis

All data are expressed as mean $\pm$ S.E.M of three independent determinations. Data were analyzed through a two-tailed t-test. A probability of $P < 0.05$ was considered statistically significant; [, $P < 0.05$, [[, $P < 0.01$, *, $P < 0.05$, **, $P < 0.01$, ***, $P < 0.001$ and ****, $P < 0.0001$ denote the significance thresholds; ns denotes not significant.

Data and Materials Availability

The data supporting the findings of this study are available within the article and its Supplementary Information. Other data and reagents are available from the corresponding authors upon reasonable request.

Additional information

Author contributions

Conceptualization: JZ, YH, LL, ZL

Methodology: JZ, DZ, JL

Investigation: JZ, DZ, JL, DK, XL, RZ, YL

Visualization: XG, YQ, DW, JC

Supervision: DK, XL, RZ, YL, XG, YQ, DW, JC, YH

Funding acquisition: YH, LL, ZL

Data curation: JZ, YH

Writing—original draft: JZ, DZ, JL

Writing—review & editing: JZ, YH, LL, ZL

Competing Interests

All other authors declare they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (Nos.32200466).

Acknowledgements

We thank Yuning Song, Yuanyuan Xu and Tingting Sui for critical feedbacks on the work and manuscript.

References

- 1 Wiener D., Schwartz S (2021) **The epitranscriptome beyond m(6)A** *Nat Rev Genet* **22**:119–131 <https://doi.org/10.1038/s41576-020-00295-8>
- 2 Li S., Mason C. E (2014) **The pivotal regulatory landscape of RNA modifications** *Annu Rev Genomics Hum Genet* **15**:127–150 <https://doi.org/10.1146/annurev-genom-090413-025405>
- 3 Van Haute L., et al. (2019) **NSUN2 introduces 5-methylcytosines in mammalian mitochondrial tRNAs** *Nucleic Acids Res* **47**:8720–8733 <https://doi.org/10.1093/nar/gkz559>
- 4 Blaze J., et al. (2021) **Neuronal Nsun2 deficiency produces tRNA epitranscriptomic alterations and proteomic shifts impacting synaptic signaling and behavior** *Nat Commun* **12** <https://doi.org/10.1038/s41467-021-24969-x>
- 5 Yang X., et al. (2017) **5-methylcytosine promotes mRNA export -NSUN2 as the methyltransferase and ALYREF as an m(5)C reader** *Cell Res* **27**:606–625 <https://doi.org/10.1038/cr.2017.55>
- 6 Sharma S., Yang J., Watzinger P., Kötter P., Entian K. D (2013) **Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively** *Nucleic Acids Res* **41**:9062–9076 <https://doi.org/10.1093/nar/gkt679>
- 7 Hussain S., et al. (2013) **NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs** *Cell Rep* **4**:255–261 <https://doi.org/10.1016/j.celrep.2013.06.029>
- 8 Aguilo F., et al. (2016) **Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1 α** *Cell Rep* **14**:479–492 <https://doi.org/10.1016/j.celrep.2015.12.043>
- 9 Yang Y., et al. (2019) **RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay** *Mol Cell* **75**:1188–1202 <https://doi.org/10.1016/j.molcel.2019.06.033>
- 10 García-Vílchez R., Sevilla A., Blanco S (2019) **Post-transcriptional regulation by cytosine-5 methylation of RNA** *Biochim Biophys Acta Gene Regul Mech* **1862**:240–252 <https://doi.org/10.1016/j.bbagrm.2018.12.003>
- 11 Trixl L., Lusser A (2019) **The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark** *Wiley Interdiscip Rev RNA* **10** <https://doi.org/10.1002/wrna.1510>
- 12 Boo S. H., Kim Y. K (2020) **The emerging role of RNA modifications in the regulation of mRNA stability** *Exp Mol Med* **52**:400–408 <https://doi.org/10.1038/s12276-020-0407-z>
- 13 Courtney D. G., et al. (2019) **Epitranscriptomic Addition of m(5)C to HIV-1 Transcripts Regulates Viral Gene Expression** *Cell Host Microbe* **26**:217–227 <https://doi.org/10.1016/j.chom.2019.07.005>

- 14 Liu J., et al. (2022) **Developmental mRNA m(5)C landscape and regulatory innovations of massive m(5)C modification of maternal mRNAs in animals** *Nat Commun* **13** <https://doi.org/10.1038/s41467-022-30210-0>
- 15 Chen P., Zhang T., Yuan Z., Shen B., Chen L (2019) **Expression of the RNA methyltransferase Nsun5 is essential for developing cerebral cortex** *Mol Brain* **12** <https://doi.org/10.1186/s13041-019-0496-6>
- 16 Flores J. V., et al. (2017) **Cytosine-5 RNA Methylation Regulates Neural Stem Cell Differentiation and Motility** *Stem Cell Reports* **8**:112–124 <https://doi.org/10.1016/j.stemcr.2016.11.014>
- 17 Chen S. Y., et al. (2022) **RNA bisulfite sequencing reveals NSUN2-mediated suppression of epithelial differentiation in pancreatic cancer** *Oncogene* **41**:3162–3176 <https://doi.org/10.1038/s41388-022-02325-7>
- 18 Chen X., et al. (2019) **5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs** *Nat Cell Biol* **21**:978–990 <https://doi.org/10.1038/s41556-019-0361-y>
- 19 Yang R., et al. (2021) **The RNA methyltransferase NSUN6 suppresses pancreatic cancer development by regulating cell proliferation** *EBioMedicine* **63** <https://doi.org/10.1016/j.ebiom.2020.103195>
- 20 Li Y., et al. (2023) **Long noncoding RNA DIAPH2-AS1 promotes neural invasion of gastric cancer via stabilizing NSUN2 to enhance the m5C modification of NTN1** *Cell Death Dis* **14** <https://doi.org/10.1038/s41419-023-05781-5>
- 21 Huang T., Chen W., Liu J., Gu N., Zhang R (2019) **Genome-wide identification of mRNA 5-methylcytosine in mammals** *Nat Struct Mol Biol* **26**:380–388 <https://doi.org/10.1038/s41594-019-0218-x>
- 22 Trixl L., Rieder D., Amort T., Lusser A (2019) **Bisulfite Sequencing of RNA for Transcriptome-Wide Detection of 5-Methylcytosine** *Methods Mol Biol* **1870**:1–21 https://doi.org/10.1007/978-1-4939-8808-2_1
- 23 Schaefer M., Pollex T., Hanna K., Lyko F (2009) **RNA cytosine methylation analysis by bisulfite sequencing** *Nucleic Acids Res* **37** <https://doi.org/10.1093/nar/gkn954>
- 24 Amort T., Lusser A (2017) **Detection of 5-Methylcytosine in Specific Poly(A) RNAs by Bisulfite Sequencing** *Methods Mol Biol* **1562**:107–121 https://doi.org/10.1007/978-1-4939-6807-7_8
- 25 Cui X., et al. (2017) **5-Methylcytosine RNA Methylation in Arabidopsis Thaliana** *Mol Plant* **10**:1387–1399 <https://doi.org/10.1016/j.molp.2017.09.013>
- 26 Edelheit S., Schwartz S., Mumbach M. R., Wurtzel O., Sorek R (2013) **Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs** *PLoS Genet* **9** <https://doi.org/10.1371/journal.pgen.1003602>
- 27 Saplaoura E., Perrera V., Colot V., Kragler F (2020) **Methylated RNA Immunoprecipitation Assay to Study m5C Modification in Arabidopsis** *J Vis Exp* <https://doi.org/10.3791/61231>
- 28 Khoddami V., Cairns B. R (2013) **Identification of direct targets and modified bases of RNA cytosine methyltransferases** *Nat Biotechnol* **31**:458–464 <https://doi.org/10.1038/nbt.2566>

- 29 Yuan F., et al. (2019) **Bisulfite-free and base-resolution analysis of 5-methylcytidine and 5-hydroxymethylcytidine in RNA with peroxotungstate** *Chem Commun (Camb)* **55**:2328–2331 <https://doi.org/10.1039/c9cc00274j>
- 30 Hayashi G., et al. (2016) **Base-Resolution Analysis of 5-Hydroxymethylcytosine by One-Pot Bisulfite-Free Chemical Conversion with Peroxotungstate** *J Am Chem Soc* **138**:14178–14181 <https://doi.org/10.1021/jacs.6b06428>
- 31 Harel N., Meir M., Gophna U., Stern A (2019) **Direct sequencing of RNA with MinION Nanopore: detecting mutations based on associations** *Nucleic Acids Res* **47** <https://doi.org/10.1093/nar/gkz907>
- 32 Chang J. J., et al. (2021) **Transcriptional and epi-transcriptional dynamics of SARS-CoV-2 during cellular infection** *Cell Rep* **35** <https://doi.org/10.1016/j.celrep.2021.109108>
- 33 Wan Y. K., Hendra C., Pratanwanich P. N., Göke J (2022) **Beyond sequencing: machine learning algorithms extract biology hidden in Nanopore signal data** *Trends Genet* **38**:246–257 <https://doi.org/10.1016/j.tig.2021.09.001>
- 34 Budzko L., Hoffa-Sobiech K., Jackowiak P., Figlerowicz M (2023) **Engineered deaminases as a key component of DNA and RNA editing tools** *Mol Ther Nucleic Acids* **34** <https://doi.org/10.1016/j.omtn.2023.102062>
- 35 Rosenberg B. R., Hamilton C. E., Mwangi M. M., Dewell S., Papavasiliou F. N (2011) **Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs** *Nat Struct Mol Biol* **18**:230–236 <https://doi.org/10.1038/nsmb.1975>
- 36 Qian X., et al. (1998) **Low expression of the apolipoprotein B mRNA-editing transgene in mice reduces LDL levels but does not cause liver dysplasia or tumors** *Arterioscler Thromb Vasc Biol* **18**:1013–1020 <https://doi.org/10.1161/01.atv.18.6.1013>
- 37 Siriwardena S. U., Chen K., Bhagwat A. S (2016) **Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases** *Chem Rev* **116**:12688–12710 <https://doi.org/10.1021/acs.chemrev.6b00296>
- 38 Richter M. F., et al. (2020) **Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity** *Nat Biotechnol* **38**:883–891 <https://doi.org/10.1038/s41587-020-0453-z>
- 39 Gaudelli N. M., et al. (2020) **Directed evolution of adenine base editors with increased activity and therapeutic application** *Nat Biotechnol* **38**:892–900 <https://doi.org/10.1038/s41587-020-0491-6>
- 40 Meyer K. D (2019) **DART-seq: an antibody-free method for global m(6)A detection** *Nat Methods* **16**:1275–1280 <https://doi.org/10.1038/s41592-019-0570-0>
- 41 Sun Z., et al. (2019) **Effects of NSUN2 deficiency on the mRNA 5-methylcytosine modification and gene expression profile in HEK293 cells** *Epigenomics* **11**:439–453 <https://doi.org/10.2217/epi-2018-0169>
- 42 Zhang Q., Zheng Q., Yu X., He Y., Guo W (2020) **Overview of distinct 5-methylcytosine profiles of messenger RNA in human hepatocellular carcinoma and paired adjacent non-tumor tissues** *J Transl Med* **18** <https://doi.org/10.1186/s12967-020-02417-6>

- 43 Zhang Q., et al. (2021) **The role of RNA m(5)C modification in cancer metastasis** *Int J Biol Sci* **17**:3369–3380 <https://doi.org/10.7150/ijbs.61439>
- 44 Liu J., et al. (2021) **Sequence-and structure-selective mRNA m(5)C methylation by NSUN6 in animals** *Natl Sci Rev* **8** <https://doi.org/10.1093/nsr/nwaa273>
- 45 Schumann U., et al. (2020) **Multiple links between 5-methylcytosine content of mRNA and translation** *BMC Biol* **18** <https://doi.org/10.1186/s12915-020-00769-5>
- 46 Fang L., et al. (2020) **CIGAR-seq, a CRISPR/Cas-based method for unbiased screening of novel mRNA modification regulators** *Mol Syst Biol* **16** <https://doi.org/10.15252/msb.202010025>
- 47 Wang N., et al. (2023) **m(5)C-dependent cross-regulation between nuclear reader ALYREF and writer NSUN2 promotes urothelial bladder cancer malignancy through facilitating RABL6/TK1 mRNAs splicing and stabilization** *Cell Death Dis* **14** <https://doi.org/10.1038/s41419-023-05661-y>
- 48 Tang Y., et al. (2020) **OsNSUN2-Mediated 5-Methylcytosine mRNA Modification Enhances Rice Adaptation to High Temperature** *Dev Cell* **53**:272–286 <https://doi.org/10.1016/j.devcel.2020.03.009>
- 49 Squires J. E., et al. (2012) **Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA** *Nucleic Acids Res* **40**:5023–5033 <https://doi.org/10.1093/nar/gks144>
- 50 Zhang P., Wang Y., Gu X (2020) **RNA 5-Methylcytosine Controls Plant Development and Environmental Adaptation** *Trends Plant Sci* **25**:954–958 <https://doi.org/10.1016/j.tplants.2020.07.004>
- 51 Fletcher K., et al. (2018) **The WD40 domain of ATG16L1 is required for its non-canonical role in lipidation of LC3 at single membranes** *EMBO J* **37** <https://doi.org/10.15252/embj.201797840>
- 52 Guo X., et al. (2003) **A Rac/Cdc42-specific exchange factor, GEFT, induces cell proliferation, transformation, and migration** *J Biol Chem* **278**:13207–13215 <https://doi.org/10.1074/jbc.M208896200>
- 53 Naim V., Rosselli F (2009) **The FANCD1 pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities** *Nat Cell Biol* **11**:761–768 <https://doi.org/10.1038/ncb1883>
- 54 Anger A. M., et al. (2013) **Structures of the human and Drosophila 80S ribosome** *Nature* **497**:80–85 <https://doi.org/10.1038/nature12104>
- 55 Liang X., et al. (2020) **Structural snapshots of human pre-60S ribosomal particles before and after nuclear export** *Nat Commun* **11** <https://doi.org/10.1038/s41467-020-17237-x>
- 56 Gkatza N. A., et al. (2019) **Cytosine-5 RNA methylation links protein synthesis to cell metabolism** *PLoS Biol* **17** <https://doi.org/10.1371/journal.pbio.3000297>
- 57 Ye F., et al. (2023) **SIRT1/PGC-1α is involved in arsenic-induced male reproductive damage through mitochondrial dysfunction, which is blocked by the antioxidative effect of zinc** *Environ Pollut* **320** <https://doi.org/10.1016/j.envpol.2023.121084>

- 58 McCollum G., Keng P. C., States J. C., McCabe M. J. (2005) **Arsenite delays progression through each cell cycle phase and induces apoptosis following G2/M arrest in U937 myeloid leukemia cells** *J Pharmacol Exp Ther* **313**:877–887 <https://doi.org/10.1124/jpet.104.080713>
- 59 Ran S., et al. (2022) **NaAsO(2) decreases GSH synthesis by inhibiting GCLC and induces apoptosis through Hela cell mitochondrial damage, mediating the activation of the NF- κ B/miR-21 signaling pathway** *Ecotoxicol Environ Saf* **234** <https://doi.org/10.1016/j.ecoenv.2022.113380>
- 60 Schaefer M (2015) **RNA 5-Methylcytosine Analysis by Bisulfite Sequencing** *Methods Enzymol* **560**:297–329 <https://doi.org/10.1016/bs.mie.2015.03.007>
- 61 Bourgeois G., et al. (2015) **Eukaryotic rRNA Modification by Yeast 5-Methylcytosine-Methyltransferases and Human Proliferation-Associated Antigen p120** *PLoS ONE* **10** <https://doi.org/10.1371/journal.pone.0133321>
- 62 Gu X., Liang Z (2019) **Transcriptome-Wide Mapping 5-Methylcytosine by m(5)C RNA Immunoprecipitation Followed by Deep Sequencing in Plant** *Methods Mol Biol* **1933**:389–394 https://doi.org/10.1007/978-1-4939-9045-0_24
- 63 Kim K., Shi A. B., Kelley K., Chen X. S (2023) **Unraveling the Enzyme-Substrate Properties for APOBEC3A-Mediated RNA Editing** *J Mol Biol* **435** <https://doi.org/10.1016/j.jmb.2023.168198>
- 64 Barka A., et al. (2022) **The Base-Editing Enzyme APOBEC3A Catalyzes Cytosine Deamination in RNA with Low Proficiency and High Selectivity** *ACS Chem Biol* **17**:629–636 <https://doi.org/10.1021/acscchembio.1c00919>
- 65 Pan Y., et al. (2017) **Nanoscale Characterization of Interaction of APOBEC3G with RNA** *Biochemistry* **56**:1473–1481 <https://doi.org/10.1021/acs.biochem.6b01189>
- 66 Kluesner M. G., et al. (2018) **EditR: A Method to Quantify Base Editing from Sanger Sequencing** *Crispr j* **1**:239–250 <https://doi.org/10.1089/crispr.2018.0014>
- 67 Bienert S., et al. (2017) **The SWISS-MODEL Repository-new features and functionality** *Nucleic Acids Res* **45**:D313–d319 <https://doi.org/10.1093/nar/gkw1132>
- 68 Delano W. L (2002) **The PyMol Molecular Graphics System** *Proteins Structure Function and Bioinformatics* **30**:442–454
- 69 Liu Q., et al. (2019) **Hi-TOM: a platform for high-throughput tracking of mutations induced by CRISPR/Cas systems** *Sci China Life Sci* **62**:1–7 <https://doi.org/10.1007/s11427-018-9402-9>
- 70 Olarerin-George A. O., Jaffrey S. R (2017) **MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites** *Bioinformatics* **33**:1563–1564 <https://doi.org/10.1093/bioinformatics/btx002>
- 71 Quinlan A. R., Hall I. M (2010) **BEDTools: a flexible suite of utilities for comparing genomic features** *Bioinformatics* **26**:841–842 <https://doi.org/10.1093/bioinformatics/btq033>
- 72 Crooks G. E., Hon G., Chandonia J. M., Brenner S. E (2004) **WebLogo: a sequence logo generator** *Genome Res* **14**:1188–1190 <https://doi.org/10.1101/gr.849004>

- 73 Sherman B. T, et al. (2022) **DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)** *Nucleic Acids Res* **50**:W216–w221 <https://doi.org/10.1093/nar/gkac194>

Editors

Reviewing Editor

Detlef Weigel

Max Planck Institute for Biology Tübingen, Tübingen, Germany

Senior Editor

David Ron

University of Cambridge, Cambridge, United Kingdom

Reviewer #2 (Public review):

The fledgling field of epitranscriptomics has encountered various technical roadblocks with implications as to the validity of early epitranscriptomics mapping data. As a prime example, the low specificity of (supposedly) modification-specific antibodies for the enrichment of modified RNAs, has been ignored for quite some time and is only now recognized for its dismal reproducibility (between different labs), which necessitates the development of alternative methods for modification detection.

Furthermore, early attempts to map individual epitranscriptomes using sequencing-based techniques are largely characterized by the deliberate avoidance of orthogonal approaches aimed at confirming the existence of RNA modifications that have been originally identified.

Improved methodology, the inclusion of various controls, and better mapping algorithms as well as the application of robust statistics for the identification of false-positive RNA modification calls have allowed revisiting original (seminal) publications whose early mapping data allowed making hyperbolic claims about the number, localization and importance of RNA modifications, especially in mRNA. Besides the existence of m6A in mRNA, the detectable incidence of RNA modifications in mRNAs has drastically dropped.

As for m5C, the subject of the manuscript submitted by Zhou et al., its identification in mRNA goes back to Squires et al., 2012 reporting on >10.000 sites in mRNA of a human cancer cell line, followed by intermittent findings reporting on pretty much every number between 0 to > 100.000 m5C sites in different human cell-derived mRNA transcriptomes. The reason for such discrepancy is most likely of a technical nature. Importantly, all studies reporting on actual transcript numbers that were m5C-modified relied on RNA bisulfite sequencing, an NGS-based method, that can discriminate between methylated and non-methylated Cs after chemical deamination of C but not m5C. RNA bisulfite sequencing has a notoriously high background due to deamination artifacts, which occur largely due to incomplete denaturation of double-stranded regions (denaturing-resistant) of RNA molecules. Furthermore, m5C sites in mRNAs have now been mapped to regions that have not only sequence identity but also structural features of tRNAs. Various studies revealed that the highly conserved m5C RNA methyltransferases NSUN2 and NSUN6 do not only accept tRNAs but also other RNAs (including mRNAs) as methylation substrates, which in combination account for most of the RNA bisulfite-mapped m5C sites in human mRNA transcriptomes. Is m5C in mRNA only a result of the Star activity of tRNA or rRNA modification enzymes, or is their low stoichiometry biologically relevant?

In light of the short-comings of existing tools to robustly determine m5C in transcriptomes, other methods, like DRAM-seq, allowing to map m5C independently of ex situ RNA treatment

with chemicals, are needed to arrive at a more solid "ground state", from which it will be possible to state and test various hypotheses as to the biological function of m5C, especially in lowly abundant RNAs such as mRNA.

Importantly, the identification of >10,000 sites containing m5C increases through DRAM-Seq, increases the number of potential m5C marks in human cancer cells from a couple of 100 (after rigorous post-hoc analysis of RNA bisulfite sequencing data) by orders of magnitude. This begs the question, whether or not the application of these editing tools results in editing artefacts overstating the number of actual m5C sites in the human cancer transcriptome.

Remaining comments after resubmission:

(1) The use of two m5C reader proteins is likely a reason for the high number of edits introduced by the DRAM-Seq method. Both ALYREF and YBX1 are ubiquitous proteins with multiple roles in RNA metabolism including splicing and mRNA export. It is reasonable to assume that both ALYREF and YBX1 bind to many mRNAs that do not contain m5C. To substantiate the author's claim that ALYREF or YBX1 binds m5C-modified RNAs to an extent that would allow distinguishing its binding to non-modified RNAs from binding to m5C-modified RNAs, it would be recommendable to provide data on the affinity of these, supposedly proven, m5C readers to non-modified versus m5C-modified RNAs. To do so, this reviewer suggests performing experiments as described in Slama et al., 2020 (doi: 10.1016/j.ymeth.2018.10.020). Mind you that using dot blots like in so many published studies to show modification-specific antibody or protein binding, is insufficient as an argument because no antibody, nor protein encounters nanograms to micrograms of a specific RNA identity in a cell. This issue remains a major caveat in all studies using so-called RNA modification reader proteins as bait for detecting RNA modifications in epitranscriptomics research and becomes a pertinent problem, if used as a platform for base-editing similar to the work presented in this manuscript.

(2) Using sodium arsenite treatment of cells as a means to change the m5C status of transcripts through the downregulation of the two major m5C writer proteins NSUN2 and NSUN6 is problematic and the conclusions from these experiments are not warranted. Sodium arsenite is a chemical that poisons every protein containing thiol groups. Not only do NSUN proteins contain cysteines but also the base editor fusion proteins. Arsenite will inactivate these proteins, hence the editing frequency will drop, as observed in the experiments shown in Figure 5, which the authors explain with fewer m5C sites to be detected by the fusion proteins.

(3) The authors should move high-confidence editing site data contained in Supplementary Tables 2 and 3 into one of the main Figures to substantiate what is discussed in Figure 4A. However, the data needs to be visualized in another way than excel format. Furthermore, Supplementary Table 2 does not contain a description of the columns, while Supplementary Table 3 contains a single row with letters and numbers.

<https://doi.org/10.7554/eLife.98166.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Reviewer #1:

In this manuscript, Zhou et al describe a deaminase and reader protein-assisted RNA m5C sequencing method. The general strategy is similar to DART-seq for m6A sequencing, but the difference is that in DART-seq, m6A sites are always followed by C

which can be deaminated by fused APOBEC1 to provide a high resolution of m6A sites, while in the case of m5C, no such obvious conserved motifs for m5C sites exist, therefore, the detection resolution is much lower. In addition, the authors used two known m5C binding proteins ALYREF and YBX1 to guide the fused deaminases, but it is not clear whether these two binding proteins can bind most m5C sites and compete with other m5C binding proteins.

Thank you for your kind suggestion. RNA affinity chromatography and mass spectrometry analyses using biotin-labelled oligonucleotides with or without m5C were performed in previous reports (doi:10.1038/cr.2017.55 and doi: 10.1038/s41556-019-0361-y), and the results showed that ALYREF and YBX1 had a more prominent binding ability to m5C-modified oligonucleotides. Moreover, these two m5C-binding proteins are also responsible for mRNA m5C binding, so we chose to use their ability to bind targeted m5C to construct a DRAM detection system in anticipation of transcriptome-wide m5C detection. We hope to propose a suitable detection strategy for RNA m5C, and there will certainly be room for optimization of the DRAM system in the future with more in-depth studies of m5C binding proteins. We have discussed the above issue in lines 75-82 and 315-318.

It is well known that two highly modified m5C sites exist in 28S RNA and many m5C sites exist in tRNA, the authors should validate their methods first by detecting these known m5C sites and evaluate the possible false positives in rRNA and tRNA.

Thank you for your kind suggestion. We attempted PCR amplification of sequences flanking m5C sites 3782 and 4447 on 28S rRNA, as well as multiple m5C sites on tRNA, including m5C48 and m5C49 on tRNAVal, m5C48 and m5C49 on tRNAAsp, and m5C48 on tRNALys.

However, Sanger sequencing revealed no valid mutations, which was implemented in Figure S3. We believe this outcome indicates that the DRAM system is more suited for transcriptome-wide m5C detection of mRNAs. This is supported by current reports that ALYREF and YBX1 are responsible for the m5C-binding proteins of mRNAs (doi:10.1038/cr.2017.55 and doi: 10.1038/s41556-019-0361-y). The above results and descriptions were added to lines 136-143.

In mRNA, it is not clear what is the overlap between the technical replicates. In Figures 4A and 4C, they detected more than 10K m5C sites, and most of them did not overlap with sites uncovered by other methods. These numbers are much larger than expected and possibly most of them are false positives.

Thank you for your kind suggestion. We observed significant overlap between the technical repeats by comparing the data across biological repeats, as shown in Figure S4C and described in lines 174-175. We considered m5C modification in a region only when editing events were detected in at least two biological replicates, ensuring a high-stringency screening process (details seen in the revised method in lines 448-455 and Figure 3F). With more in-depth research into m5C readers, we aim to achieve more accurate detection in the future.

Besides, it is not clear what is the detection sensitivity and accuracy since the method is neither single base resolution nor quantitative.

Thank you for your suggestion. As shown in Figure 3G, we found that the editing window of the DRAM system exhibited enrichment of approximately 20 bp upstream and downstream of the m5C site. Previous reports identified Type I m5C sites, which tend to have a downstream "NGGG" motif, and Type II m5C sites, which often contain a downstream "UCCA" motif. However, these m5C motifs do not fully characterize all m5C sites, and their presence downstream of an m5C site is not guaranteed (doi: 10.1038/s41594-019-0218-x). This limitation

complicates single-base resolution analysis by the DRAM system. Nevertheless, we believe that with further exploration of m5C sequence features, precise single-base resolution detection can be achieved in the future. This point is also discussed in lines 314-322.

Regarding the quantitative level of the assay, we conducted additional experiments by progressively reducing the expression levels of the fusion proteins. Sanger sequencing revealed that the editing efficiency of A-to-G and C-to-U within the m5C region significantly decreased as fusion protein expression diminished (Figure S9). These findings suggest that the DRAM system's transfection efficiency is concentration-dependent and that the ratio of editing efficiency to transfection efficiency could aid in the quantitative analysis of m5C using the DRAM system. The relative results were supplemented in Figure S9 and discussed in lines 263-271.

There are no experiments to show that the detected m5C sites are responsive to the writer proteins such as NSUN2 and NSUN6, and the determination of the motifs of these writer proteins.

Thank you for your kind suggestion. We have performed a motif enrichment analysis based on the sequences spanning 10 nt upstream and downstream of DRAM-editing sites. The relative results of this analysis were supplemented in Figure S4D and lines 168-171. Unfortunately, we did not identify any clear sequence preferences for the m5C sites catalyzed by the methyltransferases NSUN2 and NSUN6, which have previously been associated with “G”-rich sequences and the “CUCCA” motif. This limitation is mainly due to the DRAM detection system's inability to achieve single-base resolution for m5C detection, which is also explained in the above response.

Reviewer #2:

(1) The use of two m5C reader proteins is likely a reason for the high number of edits introduced by the DRAM-Seq method. Both ALYREF and YBX1 are ubiquitous proteins with multiple roles in RNA metabolism including splicing and mRNA export. It is reasonable to assume that both ALYREF and YBX1 bind to many mRNAs that do not contain m5C.

To substantiate the author's claim that ALYREF or YBX1 binds m5C-modified RNAs to an extent that would allow distinguishing its binding to non-modified RNAs from binding to m5C-modified RNAs, it would be recommended to provide data on the affinity of these, supposedly proven, m5C readers to non-modified versus m5C-modified RNAs. To do so, this reviewer suggests performing experiments as described in Slama et al., 2020 (doi: 10.1016/j.ymeth.2018.10.020). However, using dot blots like in so many published studies to show modification of a specific antibody or protein binding, is insufficient as an argument because no antibody, nor protein, encounters nanograms to micrograms of a specific RNA identity in a cell. This issue remains a major caveat in all studies using so-called RNA modification reader proteins as bait for detecting RNA modifications in epitranscriptomics research. It becomes a pertinent problem if used as a platform for base editing similar to the work presented in this manuscript.

We thank the reviewer for the valuable suggestion. Previous studies have shown that while ALYREF and YBX1 can bind mRNAs without the m5C modification, their binding affinity for m5C-modified oligonucleotides is significantly higher than for unmethylated controls. This has been demonstrated through experiments such as in vitro tractography, electrophoretic mobility shift assay (EMSA) (doi:10.1038/cr.2017.55), and UHPLC-MRM-MS/MS. Additionally, isothermal titration calorimetry measurements and PAR-CLIP experiments have shown that mutations in the key amino acids responsible for m5C binding in ALYREF and YBX1 result in a significant reduction in their ability to m5C (doi: 10.1038/s41556-019-0361-y).

Although Me-RIP analysis was unsuccessful in our laboratory, likely due to the poor specificity of the m5C antibody, we alternatively performed RNA pulldown experiments. These experiments verified that the ability of DRAMmut-expressing proteins to bind RNA with m5C modification was virtually absent compared to DRAM-expressing proteins, while their binding ability with non-modified RNA was not significantly affected. The relative RNA pulldown results were supplemented in Figure S1E, S1F and lines 110-111. Therefore, we believe that by integrating DRAMmut group, our DRAM system could effectively exclude the false-positive mutations caused by unspecific binding of DRAM's reader protein to non-m5C-modified mRNAs.

(2) Since the authors use a system that results in transient overexpression of base editor fusion proteins, they might introduce advantageous binding of these proteins to RNAs. It is unclear, which promoter is driving construct expression but it stands to reason that part of the data is based on artifacts caused by overexpression. Could the authors attempt testing whether manipulating expression levels of these fusion proteins results in different editing levels at the same RNA substrate?

Thank you for pointing this out. To investigate how different expression levels of these proteins influence A-to-G and C-to-U editing within the same m5C region, we conducted a gradient transfection using plasmid concentrations of 1500 ng, 750 ng and 300 ng. This approach allowed us to progressively reduce the expression levels of the fusion proteins. Sanger sequencing revealed that the editing efficiency of A-to-G and C-to-U within the m5C region significantly decreased as fusion protein expression diminished. These findings suggest that the transfection efficiency of the DRAM system is concentration-dependent and that the ratio of editing efficiency to transfection efficiency may assist in the quantitative analysis of m5C using the DRAM system. The relative results and hypotheses were added and discussed in Figure S9 and lines 263-271 of the revised manuscript.

(3) Using sodium arsenite treatment of cells as a means to change the m5C status of transcripts through the downregulation of the two major m5C writer proteins NSUN2 and NSUN6 is problematic and the conclusions from these experiments are not warranted. Sodium arsenite is a chemical that poisons every protein containing thiol groups. Not only do NSUN proteins contain cysteines but also the base editor fusion proteins. Arsenite will inactivate these proteins, hence the editing frequency will drop, as observed in the experiments shown in Figure 5, which the authors explain with fewer m5C sites to be detected by the fusion proteins.

Thank you for pointing this out. We used bisulfite sequencing PCR to determine that the m5C levels in RPSA and AP5Z1 were significantly reduced after sodium arsenite treatment. This was followed by a significant decrease in editing frequency detected by the DRAM system in sodium arsenite-treated samples compared to untreated samples. This reduction aligns with the decreased editing efficiency observed in methyltransferase-deficient cells (as shown in Figures 2G and 2H), which initially convinced us that these results reflected the DRAM system's ability to monitor dynamic changes in m5C levels.

However, as the reviewer pointed out, sodium arsenite treatment could potentially inactivate the fusion proteins, leading to the observed reduction in editing efficiency. This possibility has not been conclusively ruled out in our current experiments. Optimizing this validation may require the future development of more specific m5C inhibitors. In light of this, we have revised our previous results and conclusions in lines 235-244, and discussed these points in lines 308-315.

(4) The authors should move high-confidence editing site data contained in Supplementary Tables 2 and 3 into one of the main Figures to substantiate what is

discussed in Figure 4A. However, the data needs to be visualized in another way than an Excel format. Furthermore, Supplementary Table 2 does not contain a description of the columns, while Supplementary Table 3 contains a single row with letters and numbers.

Thank you for your kind suggestion. We have visualized the data from Supplementary Tables 2 and 3 into Figure 3F, presenting it as a screening flowchart for high-confidence editing sites. In Supplementary Table 3, we have displayed only the DRAM-mutated genes, which is why it contains a single row with letters and numbers. As requested, we have included descriptions of each column and reorganized the Supplementary table 2 and 3 accordingly.

(5) The authors state that "plotting the distribution of DRAM-seq editing sites in mRNA segments (5'UTR, CDS, and 3'UTR) highlighted a significant enrichment near the initiation codon (Figure 3F).", which is not true when this reviewer looks at Figure 3F.

Thank you for your kind suggestion, and we replaced the expression of " near the initiation codon" with "in the CDS" in lines 192-193.

(6) The authors state that "In contrast, cells expressing the deaminase exhibited a distinct distribution pattern of editing sites, characterized by a prevalence throughout the 5'UTR.", which is not true when this reviewer looks at Figure 3F.

Thank you for your kind suggestion. This distribution was actually characterized by a prevalence throughout the "3'UTR", but not "5'UTR". We have also made the necessary changes in lines 193-195.

(7) The authors claim in the final conclusion: "In summary, we developed a novel deaminase and reader protein assisted RNA m5C methylation approach...", which is not what the method entails. The authors deaminate As or Us close to 5mC sites based on the binding of a deaminase-containing protein.

Thank you for your kind suggestion, and we have made the necessary changes in lines 331-334.

(8) The authors claim that "The data supporting the findings of this study are available within the article and its Supplementary Information." However, no single accession number for the deposited sequencing data can be found in the text or the supplementary data. Without the primary data, none of the claims can be verified.

Thank you for pointing this out. The sequencing data from this study has already been deposited to the GEO database (GEO accession number: GSE254194, GEO token : ororioukdbdqtpcn), and we will ensure it is made publicly available in a timely manner.

(a) To underscore point (1), a recent publication (<https://doi.org/10.1038/s41419-023-05661-y>) reported: "To further identify the potential mRNAs regulated by ALYREF, we performed RNA-seq analysis in control or ALYREF knockdown T24 cells. After knockdown of ALYREF, 143 mRNAs differentially expressed, including 94 downregulated mRNAs (NC reads >100, |Fold change| >1.5, P-value <0.05). Functional enrichment analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG) indicated that regulated mRNAs by ALYREF are chiefly enriched in canonical cancer-related pathways (Fig. S4A), including TGF- β signaling, MAPK signaling, and NF- κ B signaling, strongly supporting the oncogenic function of ALYREF in tumor progression. Among these 94 downregulated genes, 11 mRNA showed a significant reduction in m5C methylation after NUSN2 silencing in T24 cells, combined with previously transcriptome-wide RNA-BisSeq data of T24 cells [21] (Fig. 4A)."

These results translate into 94 mRNAs are regulated by ALYREF in bladder cancer-derived cells. From those, very few (11) mRNA identities respond to NSUN2-dependent RNA methylation mediated by ALYREF binding. The question then arises, is that number sufficient to claim that ALYREF is a m5C-binding protein?

And if so, how does the identification of 10,000+ edits by DRAM-Seq compare with the 94 mRNAs that are regulated by ALYREF? Were these 94 mRNAs identified by DRAM-Seq.

Thank you for your kind suggestion. Previous reports by Yang et al. (doi: 10.1038/cr.2017.55), including the literature you refer to, have detailed the close relationship between ALYREF and m5C modification, and the ALY/REF export factor (ALYREF) was identified as the first nuclear m5C reader, and it was demonstrated that many mRNAs are regulated by ALYREF, and is therefore considered to be an m5C-binding protein.

As required, by comparing the DRAM-edited mRNAs with the reported 94 mRNAs, we found that only 55.32% of the 94 mRNAs regulated by ALYREF could be detected by the DRAM system. This indicates that the DRAM system specifically targets certain mRNAs, as illustrated in Figure S4E. The relevant results were described and discussed in lines 175-179.

(b) Line 123:

"The deep sequencing results showed that the deamination rates of RPSA and SZRD1 were 75.5% and 27.25%, respectively. (Fig. 2A, B)."

The Figure shows exactly the opposite of bisulfite-mediated deamination. These are the cytosines that were not deaminated by the chemical treatment and therefore can be sequenced as cytosines and not thymidines. Hence, the term deamination rate is wrong.

Thank you for your kind suggestion. We have made the necessary change in lines 129-130 to change the deamination rates to m⁵C fraction.

(c) Line 157:

"DRAM-seq analysis further confirmed that DRAM was detected in an m5C-dependent manner, with minimal mutations in AP5Z1 and RPSA mRNAs in methyltransferase knockout cells compared to wild-type cells (Fig. 3C, D)."

There is no indication of what the authors mean by minimal mutation in these Figures. The term "minimal mutation" should be reconsidered as well.

Thank you for your kind suggestion. We intended to express that "Mutations in AP5Z1 and RPSA mRNA are reduced in methyltransferase-deficient cells." There was an issue with the initial formulation, and we have made the necessary changes in lines 165-167.

(d) Line 167:

"To further delineate the characteristics of the DRAM-seq data, we compared the distribution of DRAM-seq editing sites within the gene structure, specifically examining their occurrences in the 5'untranslated region (5'UTR), 3' untranslated region (3'UTR), CDS and ncRNA."

Which part of a coding RNA is meant by "ncRNA"?

Thank you for pointing this out. This was actually the Intergenic or Intron region, but not ncRNA. We have also corrected this labelling in Figure 3G and lines 186-189 of the revised manuscript.

(e) Line 189:

"Subsequently, we assessed the capacity of DRAM-seq to detect m5C on a transcriptome-wide scale, comparing its performance to BS-seq that have been previously reported with great authority."

The term "great authority" is not a scientific term. Please, remove adulation to senior authors.

Thank you for your kind suggestion. We removed this unsuitable expression and made the necessary changes in lines 207-208.

(f) Line 233:

"Several experiments have highlighted the requirement of 100-500 ng of RNA for m5C-RIP-seq, while BS-seq necessitates an even more demanding 500-750 µg of RNA21,25,61."

This reviewer doubts that RNA bisulfite sequencing required half to one mg of RNA input. Please, check these references.

Thank you for your kind suggestion. According to the references, we corrected µg to ng and made the necessary changes in lines 251-252.

(g) Line 247:

"Several experiments have highlighted the requirement of 100-500 ng of RNA for m5C-RIP-seq, while BS-seq necessitates an even more demanding 500-750 µg of RNA21,25,61."

This reviewer doubts that RNA bisulfite sequencing requires half to one mg of RNA input. Please, check these references.

Thank you for your kind suggestion. According to the references, we corrected µg to ng and made the necessary changes in lines 251-252.

(h) Line 292:

"Since m5C lacks a fixed motif, DRAM has an apparent limitation in achieving single-base resolution for detecting m5C."

m5C deposition by NSUN2 and NSUN6 occurs in particular motifs that were coined Type I and II motifs. Hence, this statement is not correct.

Thank you for your kind suggestion. Previous reports identified Type I m5C sites, which tend to have a downstream "NGGG" motif, and Type II m5C sites, which often contain a downstream "UCCA" motif. However, these m5C motifs do not fully characterize all m5C sites, and their presence downstream of an m5C site is not guaranteed (doi: 10.1038/s41594-019-0218-x). Therefore, we have corrected the expression "fixed motif" to "fixed base composition for characterizing all m5C modification sites" in lines 317.

(i) Line 390:

"1 µl of total cellular RNA was used for sequencing library gene..."

1 uL does not allow us to deduce which RNA mass was used for cDNA synthesis.

Thank you for your kind suggestion. According to our cDNA synthesis protocol, we corrected “1 μ l” to “1 μ g” in lines 422-423.

(j) Line 405:

"...was assessed on the Agilent 5400 system (Agilent, USA) and quantified by QPCR (1.5 nM)"

What does the 1.5 nM refer to in this sentence?

Thank you for your kind suggestion. Here, "1.5nM" means that the concentration of the constructed library should be no less than 1.5nM. We have also revised this expression in the methods in lines 436-438.

<https://doi.org/10.7554/eLife.98166.2.sa0>