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

Jiale Zhou[1#], Ding Zhao[1#], Jinze Li[2#], Deqiang Kong[1], Xiangrui Li[2], Renquan Zhang[2], Yuru Liang[2], Xun Gao[2], Yuqiang Qian[2], Di Wang[2], Jiahui Chen[2], Liangxue Lai [1,3,4*], Yang Han[2*], Zhanjun Li [2*]

# These authors contributed equally: Jiale Zhou, Ding Zhao, Jinze Li.

[1]Key Laboratory of Zoonosis Research, Ministry of Education, College of Veterinary Medicine, Jilin University, Changchun, 130062, China

[2]Laboratory Animal Center, College of Animal Science, Jilin University, Changchun, 130062, China

[3]Guangzhou Regenerative Medicine and Health Guang Dong Laboratory (GRMH-GDL), Guangzhou, 510530, China

[4]Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, 510005, China

* To whom correspondence should be addressed:

Tel: (86)431-87836176; Fax: (86)431-87980131.

*e-mail:

hanyang8584@jlu.edu.cn (Yang Han); lizj_1998@jlu.edu.cn (Zhanjun Li); lai_liangxue@gibh.ac.cn (Liangxue lai).

Abstract

5-Methylcytosine (m^5C) is one of the post-transcriptional modifications in mRNA and is 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). We anticipate that the DRAM system could pave the way for uncovering further biological functions of m^5C modifications.

Keywords: RNA, 5-methylcytosine, deaminase, m^5C reader

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 m^6A, m^5C, m^1A, m^7G and others [1,2]. Notably, RNA m^5C 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 m^5C in numerous key stages of RNA metabolism, such as mRNA stability, translation, and nuclear export[5,9-13]. The dynamic alterations of m^5C 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 m^5C 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 m^5C. To date, bisulfite-sequencing (BS-seq) has been proven to be the gold standard method for RNA m^5C methylation analysis[5,21,22]. This approach chemically deaminates unmethylated cytosine to uracil, while keeping methylated cytosine unchanged. The m^5C 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 m^5C in low abundance RNA or highly structured RNA, which directly affects the confidence of results[23,24]. Another major type of global m^5C analysis depends on antibody-assisted immunoprecipitation of m^5C methylated RNAs, such as m^5C-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 m^5C-modified regions. Moreover, TAWO-seq, originally developed for the identification of hm^5C, is also capable of m^5C 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 m^5C 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 m^5C (Supplementary Table 1). Hence, there is an urgent need for a simple, efficient, sensitive, and antibody-independent method for global m^5C detection.

The RNA-binding protein ALYREF is the initially recognized nuclear m^5C reader that binds directly to m^5C sites in mRNA and plays key roles in promoting mRNA nuclear export or tumor progression[5]. Another well-known m^5C reader, YBX1, can also recognize m^5C-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 m^5C were performed in previous reports, which indicated that ALYREF and YBX1 had a more prominent binding ability to m^5C-modified oligonucleotides[5,18]. YBX1 can preferentially recognize mRNAs with m^5C modifications via key amino acids W65-N70 (WFNVRN)[18], while K171 is essential for the specific binding of ALYREF to m^5C sites [5]. Previous studies have shown

that mutations in key amino acids responsible for recognising m^5C binding in ALYREF and YBX1 lead to a significant reduction in their binding levels to m^5C-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 substrates[34]. 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 m^5C reader-assisted RNA methylation sequencing approach (DRAM-seq), which identifies the m^5C sites through reader-mediated recognitions and deaminase-mediated point mutations neighboring the m^5C methylation sites. This bisulfite-free and antibody-free method is anticipated to provide more comprehensive and cost-effective transcriptome-wide detection of m^5C methylation, which may better assist on exploring its further regulatory mechanisms.

Results

Development of DRAM system for m^5C detection

Our sequencing platform is inspired by the concept of the m^6A DART-seq assay, in which C near the m^6A site is converted into U without affecting sequences near non-m6A sites[40]. Therefore, we hypothesized that, by utilizing the targeted binding of m^5C readers, deaminase can be recruited to achieve deamination of cytosine or adenine in the vicinity of the m^5C sites on single-stranded RNA, thereby facilitating the detection of the m^5C site. This approach was named DRAM (deaminase and m^5C reader-assisted RNA methylation sequencing). As RNA-binding proteins, ALYREF and YBX1 also could bind to RNAs without m^5C modification[5,18]. To exclude the false-positive detection of DRAM due to the non-m^5C specific binding of ALYREF and YBX1, knockout of W65-N70 (WFNVRN) amino acids in YBX1and K171A mutation in ALYREF were introduced seperately, resulting in the $DRAM^{mut}$ system (Fig. S1A-S1D). Subsequently, we verified the affinity ability of YBX1 and ALYREF for m^5C-modified RNAs by RNA pull-down experiments. Consistent with previous reports[5,18], those two m^5C readers preferentially bound RNAs containing m^5C modifications. Furthermore, mutating key amino acids involved in their interaction with m^5C significantly reduced their binding ability, indicating that ALYREF and YBX1 exhibit specificity for m^5C-methylated mRNAs. (Fig. S1E-S1H). To confirm the recognition of m^5C site by DRAM system, DRAM, $DRAM^{mut}$ and Deaminase system were transfected into the human HEK293T cells, respectively. Finally, we considered the presence of m^5C 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 m^5C sites[5,26,41,42]. To comprehensively detect the m^5C loci, the readers of m^5C (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 m^5C-dependent form

To confirm the recognition of m^5C 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 m^5C 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^5C 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 m^5C sites by DRAM system in these two mRNA. Notably, adenine close to the m^5C 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 m^5C 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 m^5C sites 3782 and 4447 on 28S rRNA and several m^5C sites on tRNA, such as the m^5C48 and m^5C49 sites of $tRNA^{Val}$, the m^5C48 and m^5C49 sites of $tRNA^{Asp}$, and the m^5C48 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 m^5C binding proteins, and thus the DRAM system is more suitable for the mRNA m^5C detection (Fig. S3). Taken together, the fusion of m^5C reader and deaminase can effectively and selectively deaminate cytosine/adenine in the vicinity of the mRNA m^5C sites.

NSUN2[43] and NSUN6[44], two family members of NOL1/NSUN protein, were both identified as m^5C methyltransferase of mRNA[45]. To verify that the detection of DRAM occurs in the presence of m^5C, we performed knockdown experiments of NSUN2 and NSUN6 in HEK293T cells by base deletion, resulting in frameshift mutations that led to reduced expression of NSUN2 and NSUN6. These cells were then transfected with DRAM. The knockout efficiency has been confirmed by western blotting (Fig.2E, 2F and Fig. S4A,4B). It has been previously demonstrated that m^5C 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 m^5C 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 m^5C-dependent form.

DRAM enables transcriptome-wide analysis of m^5C 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 m^5C (Fig.3A). To serve as positive controls, two previously published BS-seq datasets were also integrated[5,21]. Mutations were detected near the m^5C 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 m^5C 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 m^5C 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 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 m^5C sites in other mRNAs (Fig. S6). These indicated that DRAM-seq analysis was detected in an m^5C-dependent manner. Unfortunately, motif analysis failed to identify any sequence preferences or consensus motifs associated with DRAM-edited sites mediated by loci associated with NSUN2 or 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 DRAM-edited mRNAs revealed a high degree of overlap across the three biological replicates (Fig. S4C). And a recent study by Wang et al. showed that ALYREF deletion affects the expression of 94 mRNAs[47], 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.

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 m^5C 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 m^5C (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 m^5C.

Comparative analysis of the DRAM-seq editing sites with the previously published BS-seq m^5C sites indicated that the likelihood of editing was notably higher in closer proximity to the m^5C sites (Fig.3I). Furthermore, the editing window of DRAM exhibited enrichment approximately 20bp before and after the m^5C 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 m^{5}C loci

Subsequently, we then evaluated the ability of DRAM-seq to detect m^{5}C 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. We first complied the overall distribution of mutant regions identified by DRAM-seq, presenting both the mutant sites detected by the DRAM system and those reported in previous studies[5,18] across each chromosome (Fig.4A). Our results indicated that DRAM-seq identified the presence of m^{5}C 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.4B and D). Remarkably, certain pivotal regulators with diverse biological functions, such as ATG16L1(coordinats autophagy pathway)[51] and ARHGEF25 (plays an important role in actin cytoskeleton reorganisation)[52], were identified by Zhang et al. and DRAM-seq, but not by Yang et al. (Fig.4C). 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.4E). Hence, DRAM-seq appears to offer a more stable and comprehensive identification of the m^{5}C loci.

To provide functional insights into m^{5}C 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.4F, 4G, Fig. S8C and S8D).

DRAM enables low-input m^{5}C profiling

A significant challenge in m^{5}C 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 assay[56]. Immunoprecipitation-based m^{5}C assays and LC-MS/MS also impose high demand for sample input[7,25,57]. Several experiments have highlighted the requirement of 100-500 ng of RNA for m^{5}C-RIP-seq, while BS-seq necessitates an even more demanding 750-1000 ng of RNA[21,25,58]. To assess the detection limits of DRAM-Sanger, we attempted to amplify two representative m^{5}C-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.5A and B). This finding underscores that the specificity of DRAM editing depended on its ability to bind m^{5}C, and DRAM is proficient in low-input m^{5}C analyses. Furthermore, cell viability was determined by CCK8 assay on HEK293T cells transfected with DRAM (Fig.5C). 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.5D).

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 m^{5}C region, we conducted a gradient transfection using plasmid concentrations of 1500 ng, 1000 ng and 500 ng. This approach allowed us to progressively reduce

the expression levels of the fusion proteins (Fig. 5E and 5F). Sanger sequencing revealed that the editing efficiency of A-to-G and C-to-U within the m^5C region significantly decreased as fusion protein expression diminished (Fig. 5G and 5H). 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 m^5C using the DRAM system.

Discussion

In recent years, m^5C methylation modifications have received increasing attention, with multiple reports detailing the distribution of RNA m^5C methylation modifications across various species and tissues, elucidating their characteristics. Despite the relatively low abundance of m^5C, 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 m^5C detection methods.

In this study, we developed a site-specific, depth-sequencing-free m^5C 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 m^5C detection.

DRAM-seq introduces a novel strategy for transcriptome-wide m^5C detection, overcoming inherent limitations in existing methods. Notably, DRAM-seq covered around 80% of the high-confidence m^5C-modified genes detected by BS-seq and identified more potential m^5C 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 m^5C 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 m^5C detection. Our analysis demonstrates that DRAM requires as low as 10ng of total RNA for m^5C detection. While DRAM is currently well-suited for detecting m^5C on a transcriptome-wide scale, the potential for future applications involving third-generation sequencing could extend its utility to individual mRNAs, particularly m^5C heterogeneity on mRNA splicing variants. In addition, the DRAM system depends on the specific recognition of m^5C 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, m^5C 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 m^5C detection. Our results confirm that YBX1 and ALYREF exhibit specificity as m^5C readers, binding preferentially to RNAs with m^5C modifications,

thereby validating the reliability of the DRAM detection system. However, mutations in the key amino acids responsible for m^5C binding reduced their affinity while retaining some binding capacity. DRAM-seq analysis identified a substantial number of m^5C sites. However, we cannot exclude the potential existence of false positive sites resulting from non-specific binding of the m^5C reader. Further elucidation of the key amino acids directing ALYREF and YBX1's binding to m^5C methylation sites should enable more accurate and sensitive m^5C detection by DRAM-seq. Due to the lack of a fixed base composition for characterizing all m^5C modification sites, DRAM has an apparent limitation in achieving single-base resolution for detecting m^5C. This technical constraint may explain the absence of identifiable sequence specificity in our analysis of m^5C sites catalyzed by NSUN2 and NSUN6, despite previous reports associating these methyltransferases with "G"-rich sequences and the "CUCCA" motif[59]. However, our present study proved that the measuring resolution of DRAM is around 40nt, which facilitates higher precision than that of m^5C-RIP-seq (~100nt). In the future, with more in-depth analyses of m^5C reader structures and the identification of new potential m^5C readers, we expect to achieve more precise m^5C localization and more comprehensive m^5C 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 [60-62].

Although the m^5C assay can be performed using the DRAM system alone, comparing it with the $DRAM^{mut}$ and deaminase controls could enhance the accuracy of m^5C detection in specific regions. Given that the expression of DRAM fusion proteins significantly influences m^5C detection, it is advisable to transfect the same batch of cells during the assay to ensure consistent transfection efficiency across experimental groups and thus can better standardize the detection.

One future direction of endeavour is the purification of DRAM fusion proteins to facilitate *in vitro* detection of RNA m^5C methylation, which could extend the scope of DRAM-seq to diverse sample types. Another potential application for DRAM-seq could be the expression of drug-inducible DRAM systems *in vivo* using various animal models for m^5C analysis. These will together provide novel insights into m^5C modifications for biological and clinical research.

Conclusions

In summary, we developed a novel deaminase and reader protein-assisted RNA m^5C methylation approach that detects the m^5C region by deaminating As or Cs in close proximity to the m^5C sites, which does not rely on antibodies or bisulfite, thus leading to unprecedently comprehensive transcriptome-wide RNA m^5C methylation profiling. We anticipated that this system could pave the way for uncovering further biological functions of m^5C 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 fractions 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 (100μg/ml). The cells were seeded in 12-well plates and transfected using Hieff Trans™ Liposomal Transfection Reagent (Yeasen).

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 ± 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 m^5C 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 frequency[63]. 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 $2^{-\Delta\Delta Ct}$ 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 ± 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)[64]. 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 PyMOL[65].

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 m^5C methylation levels based on PCR amplification of cDNA combined with deep sequencing[23].

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 Buffe. 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 m^5C 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) [66]. 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

426 fragments of preferentially 370~420 bp in length, the library fragments were purified with AMPure
427 XP system (Beverly, USA). Then 3 μL USER Enzyme (NEB, USA) was used with size selected,
428 adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was
429 performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X)
430 Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed
431 on the Agilent 5400 system(Agilent, USA)and quantified by QPCR (library concentration ⩾ 1.5
432 nM). The qualified libraries were pooled and sequenced on Illumina platforms with PE150 strategy
433 in Novogene Bioinformatics Technology Co., Ltd (Beijing, China), according to effective library
434 concentration and data amount required.

435 **DRAM-seq analysis and calling of edited sites**

436 The raw fastq sequencing data were cleaned by trimming the adapter sequences using Fastp
437 (v0.23.1) and were aligned to the human genome (hg19) using STAR (v2.7.7) in paired-end mode.
438 The aligned BAM files were sorted and PCR duplicates were removed using Samtools (v1.12). The
439 cite calling pf DRAM-seq was performed using Bullseye, a previously customized pipeline to look
440 for C-to-U or A-to-G edited sites throughout the transcriptome[40]. Briefly, the sorted and
441 deduplicated BAM files were initially parsed by parseBAM.pl script.

442 Then, Find_edit_site.pl script was employed to find C-to-U or A-to-G editing events by
443 DRAM-seq with at least 10 reads of coverage, an edit ratio of 5%-95%, an edit ratio at least 1.5-
444 fold higher than NSUN2 or NSUN6-knockout samples, and at least 2 editing events at a given site.
445 Sites that were only found in one replicate of each DRAM protein variant were removed. Editing
446 events appeared in cells expressing merely APOBEC1 or TadA8e were also removed. For high
447 confidence filtering, we further adjusted the Find_edit_site.pl parameters to the edit ratio of 10%-
448 60%, an edit ratio of control samples at least 2-fold higher than NSUN2 or NSUN6-knockout
449 samples, and at least 4 editing events at a given site.

450 **Metagene and motif analyses**

451 Metagene analysis was performed using hg19 annotations according to previously reported
452 tool, MetaplotR[67]. For motif analysis, the 20bp flanking sequence of each DRAM-seq editing site
453 was extracted by Bedtools (v2.30.0)[68]. The motif logos were then plotted by WebLogo (v3.7.12)[69].

454 **Replicates analysis**

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

458 **GO and KEGG analysis**

459 GO and KEGG analysis of DRAM-seq edited mRNAs was performed using the DAVID
460 bioinformatic database [70]. GO terms with a P value of less than 0.05 were considered statistically
461 significant.

462 **RNA pulldown assay**

463 The biotin-labeled RNA oligonucleotides with (Oligo-m^5C) or without m^5C (Oligo-C) were
464 prepared in advance: 5′-biotin-GAGGUAUGAAXUGUAAGTT-3′ (X = C or m^5C, used in the
465 ALYREF and $ALYREF^{mut}$ group) and 5′-biotin-GAAAGGAGAUXGCCAUUAUCC-3′ (X = C or
466 m^5C, used in the YBX1 and $YBX1^{mut}$ group). Protein lysates were then isolated from HEK293T
467 cells transfected with DRAM-YBX1, DRAM-$YBX1^{mut}$, DRAM-ALYREF or DRAM-
468 $ALYREF^{mut}$ for 24 h using lysis buffer. RNA pull-down assays were performed with the Pierce™
469 Magnetic RNA-Protein Pull-Down Kit (Thermo) following the manufacturer's instructions, and the

results were finally analyzed by Western blotting.

Statistical analysis

All data are expressed as mean ± 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.

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.

ACKNOWLEDGEMENTS

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

FUNDING

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

REFERENCES

1 Wiener, D. & Schwartz, S. The epitranscriptome beyond m(6)A. *Nat Rev Genet* **22**, 119-131

(2021). https://doi.org:10.1038/s41576-020-00295-8

2 Li, S. & Mason, C. E. The pivotal regulatory landscape of RNA modifications. *Annu Rev Genomics Hum Genet* **15**, 127-150 (2014). https://doi.org:10.1146/annurev-genom-090413-025405

3 Van Haute, L. *et al.* NSUN2 introduces 5-methylcytosines in mammalian mitochondrial tRNAs. *Nucleic Acids Res* **47**, 8720-8733 (2019). https://doi.org:10.1093/nar/gkz559

4 Blaze, J. *et al.* Neuronal Nsun2 deficiency produces tRNA epitranscriptomic alterations and proteomic shifts impacting synaptic signaling and behavior. *Nat Commun* **12**, 4913 (2021). https://doi.org:10.1038/s41467-021-24969-x

5 Yang, X. *et al.* 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res* **27**, 606-625 (2017). https://doi.org:10.1038/cr.2017.55

6 Sharma, S., Yang, J., Watzinger, P., Kötter, P. & Entian, K. D. Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic Acids Res* **41**, 9062-9076 (2013). https://doi.org:10.1093/nar/gkt679

7 Hussain, S. *et al.* NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep* **4**, 255-261 (2013). https://doi.org:10.1016/j.celrep.2013.06.029

8 Aguilo, F. *et al.* Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1α. *Cell Rep* **14**, 479-492 (2016). https://doi.org:10.1016/j.celrep.2015.12.043

9 Yang, Y. *et al.* RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay. *Mol Cell* **75**, 1188-1202.e1111 (2019). https://doi.org:10.1016/j.molcel.2019.06.033

10 García-Vílchez, R., Sevilla, A. & Blanco, S. Post-transcriptional regulation by cytosine-5 methylation of RNA. *Biochim Biophys Acta Gene Regul Mech* **1862**, 240-252 (2019). https://doi.org:10.1016/j.bbagrm.2018.12.003

11 Trixl, L. & Lusser, A. The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdiscip Rev RNA* **10**, e1510 (2019). https://doi.org:10.1002/wrna.1510

12 Boo, S. H. & Kim, Y. K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med* **52**, 400-408 (2020). https://doi.org:10.1038/s12276-020-0407-z

13 Courtney, D. G. *et al.* Epitranscriptomic Addition of m(5)C to HIV-1 Transcripts Regulates Viral Gene Expression. *Cell Host Microbe* **26**, 217-227.e216 (2019). https://doi.org:10.1016/j.chom.2019.07.005

14 Liu, J. *et al.* Developmental mRNA m(5)C landscape and regulatory innovations of massive m(5)C modification of maternal mRNAs in animals. *Nat Commun* **13**, 2484 (2022). https://doi.org:10.1038/s41467-022-30210-0

15 Chen, P., Zhang, T., Yuan, Z., Shen, B. & Chen, L. Expression of the RNA methyltransferase Nsun5 is essential for developing cerebral cortex. *Mol Brain* **12**, 74 (2019). https://doi.org:10.1186/s13041-019-0496-6

16 Flores, J. V. *et al.* Cytosine-5 RNA Methylation Regulates Neural Stem Cell Differentiation and Motility. *Stem Cell Reports* **8**, 112-124 (2017). https://doi.org:10.1016/j.stemcr.2016.11.014

17 Chen, S. Y. *et al.* RNA bisulfite sequencing reveals NSUN2-mediated suppression of epithelial differentiation in pancreatic cancer. *Oncogene* **41**, 3162-3176 (2022).

549 https://doi.org:10.1038/s41388-022-02325-7
550 18 Chen, X. *et al.* 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing
551 mRNAs. *Nat Cell Biol* **21**, 978-990 (2019). https://doi.org:10.1038/s41556-019-0361-y
552 19 Yang, R. *et al.* The RNA methyltransferase NSUN6 suppresses pancreatic cancer development
553 by regulating cell proliferation. *EBioMedicine* **63**, 103195 (2021).
554 https://doi.org:10.1016/j.ebiom.2020.103195
555 20 Li, Y. *et al.* Long noncoding RNA DIAPH2-AS1 promotes neural invasion of gastric cancer via
556 stabilizing NSUN2 to enhance the m5C modification of NTN1. *Cell Death Dis* **14**, 260 (2023).
557 https://doi.org:10.1038/s41419-023-05781-5
558 21 Huang, T., Chen, W., Liu, J., Gu, N. & Zhang, R. Genome-wide identification of mRNA 5-
559 methylcytosine in mammals. *Nat Struct Mol Biol* **26**, 380-388 (2019).
560 https://doi.org:10.1038/s41594-019-0218-x
561 22 Trixl, L., Rieder, D., Amort, T. & Lusser, A. Bisulfite Sequencing of RNA for Transcriptome-
562 Wide Detection of 5-Methylcytosine. *Methods Mol Biol* **1870**, 1-21 (2019).
563 https://doi.org:10.1007/978-1-4939-8808-2_1
564 23 Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite
565 sequencing. *Nucleic Acids Res* **37**, e12 (2009). https://doi.org:10.1093/nar/gkn954
566 24 Amort, T. & Lusser, A. Detection of 5-Methylcytosine in Specific Poly(A) RNAs by Bisulfite
567 Sequencing. *Methods Mol Biol* **1562**, 107-121 (2017). https://doi.org:10.1007/978-1-4939-
568 6807-7_8
569 25 Cui, X. *et al.* 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Mol Plant* **10**, 1387-
570 1399 (2017). https://doi.org:10.1016/j.molp.2017.09.013
571 26 Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-wide
572 mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C
573 within archaeal mRNAs. *PLoS Genet* **9**, e1003602 (2013).
574 https://doi.org:10.1371/journal.pgen.1003602
575 27 Saplaoura, E., Perrera, V., Colot, V. & Kragler, F. Methylated RNA Immunoprecipitation Assay
576 to Study m5C Modification in Arabidopsis. *J Vis Exp* (2020). https://doi.org:10.3791/61231
577 28 Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA
578 cytosine methyltransferases. *Nat Biotechnol* **31**, 458-464 (2013).
579 https://doi.org:10.1038/nbt.2566
580 29 Yuan, F. *et al.* Bisulfite-free and base-resolution analysis of 5-methylcytidine and 5-
581 hydroxymethylcytidine in RNA with peroxotungstate. *Chem Commun (Camb)* **55**, 2328-2331
582 (2019). https://doi.org:10.1039/c9cc00274j
583 30 Hayashi, G. *et al.* Base-Resolution Analysis of 5-Hydroxymethylcytosine by One-Pot Bisulfite-
584 Free Chemical Conversion with Peroxotungstate. *J Am Chem Soc* **138**, 14178-14181 (2016).
585 https://doi.org:10.1021/jacs.6b06428
586 31 Harel, N., Meir, M., Gophna, U. & Stern, A. Direct sequencing of RNA with MinION Nanopore:
587 detecting mutations based on associations. *Nucleic Acids Res* **47**, e148 (2019).
588 https://doi.org:10.1093/nar/gkz907
589 32 Chang, J. J. *et al.* Transcriptional and epi-transcriptional dynamics of SARS-CoV-2 during
590 cellular infection. *Cell Rep* **35**, 109108 (2021). https://doi.org:10.1016/j.celrep.2021.109108
591 33 Wan, Y. K., Hendra, C., Pratanwanich, P. N. & Göke, J. Beyond sequencing: machine learning
592 algorithms extract biology hidden in Nanopore signal data. *Trends Genet* **38**, 246-257 (2022).

https://doi.org:10.1016/j.tig.2021.09.001

34 Budzko, L., Hoffa-Sobiech, K., Jackowiak, P. & Figlerowicz, M. Engineered deaminases as a key component of DNA and RNA editing tools. *Mol Ther Nucleic Acids* **34**, 102062 (2023). https://doi.org:10.1016/j.omtn.2023.102062

35 Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S. & Papavasiliou, F. N. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat Struct Mol Biol* **18**, 230-236 (2011). https://doi.org:10.1038/nsmb.1975

36 Qian, X. *et al.* 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 (1998). https://doi.org:10.1161/01.atv.18.6.1013

37 Siriwardena, S. U., Chen, K. & Bhagwat, A. S. Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. *Chem Rev* **116**, 12688-12710 (2016). https://doi.org:10.1021/acs.chemrev.6b00296

38 Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38**, 883-891 (2020). https://doi.org:10.1038/s41587-020-0453-z

39 Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* **38**, 892-900 (2020). https://doi.org:10.1038/s41587-020-0491-6

40 Meyer, K. D. DART-seq: an antibody-free method for global m(6)A detection. *Nat Methods* **16**, 1275-1280 (2019). https://doi.org:10.1038/s41592-019-0570-0

41 Sun, Z. *et al.* Effects of NSUN2 deficiency on the mRNA 5-methylcytosine modification and gene expression profile in HEK293 cells. *Epigenomics* **11**, 439-453 (2019). https://doi.org:10.2217/epi-2018-0169

42 Zhang, Q., Zheng, Q., Yu, X., He, Y. & Guo, W. Overview of distinct 5-methylcytosine profiles of messenger RNA in human hepatocellular carcinoma and paired adjacent non-tumor tissues. *J Transl Med* **18**, 245 (2020). https://doi.org:10.1186/s12967-020-02417-6

43 Zhang, Q. *et al.* The role of RNA m(5)C modification in cancer metastasis. *Int J Biol Sci* **17**, 3369-3380 (2021). https://doi.org:10.7150/ijbs.61439

44 Liu, J. *et al.* Sequence- and structure-selective mRNA m(5)C methylation by NSUN6 in animals. *Natl Sci Rev* **8**, nwaa273 (2021). https://doi.org:10.1093/nsr/nwaa273

45 Schumann, U. *et al.* Multiple links between 5-methylcytosine content of mRNA and translation. *BMC Biol* **18**, 40 (2020). https://doi.org:10.1186/s12915-020-00769-5

46 Fang, L. *et al.* CIGAR-seq, a CRISPR/Cas-based method for unbiased screening of novel mRNA modification regulators. *Mol Syst Biol* **16**, e10025 (2020). https://doi.org:10.15252/msb.202010025

47 Wang, N. *et al.* 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**, 139 (2023). https://doi.org:10.1038/s41419-023-05661-y

48 Tang, Y. *et al.* OsNSUN2-Mediated 5-Methylcytosine mRNA Modification Enhances Rice Adaptation to High Temperature. *Dev Cell* **53**, 272-286.e277 (2020). https://doi.org:10.1016/j.devcel.2020.03.009

49 Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-

637 coding RNA. *Nucleic Acids Res* **40**, 5023-5033 (2012). https://doi.org:10.1093/nar/gks144
638 50 Zhang, P., Wang, Y. & Gu, X. RNA 5-Methylcytosine Controls Plant Development and
639 Environmental Adaptation. *Trends Plant Sci* **25**, 954-958 (2020).
640 https://doi.org:10.1016/j.tplants.2020.07.004
641 51 Fletcher, K. *et al.* The WD40 domain of ATG16L1 is required for its non-canonical role in
642 lipidation of LC3 at single membranes. *EMBO J* **37** (2018).
643 https://doi.org:10.15252/embj.201797840
644 52 Guo, X. *et al.* A Rac/Cdc42-specific exchange factor, GEFT, induces cell proliferation,
645 transformation, and migration. *J Biol Chem* **278**, 13207-13215 (2003).
646 https://doi.org:10.1074/jbc.M208896200
647 53 Naim, V. & Rosselli, F. The FANC pathway and BLM collaborate during mitosis to prevent
648 micro-nucleation and chromosome abnormalities. *Nat Cell Biol* **11**, 761-768 (2009).
649 https://doi.org:10.1038/ncb1883
650 54 Anger, A. M. *et al.* Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80-85
651 (2013). https://doi.org:10.1038/nature12104
652 55 Liang, X. *et al.* Structural snapshots of human pre-60S ribosomal particles before and after
653 nuclear export. *Nat Commun* **11**, 3542 (2020). https://doi.org:10.1038/s41467-020-17237-x
654 56 Schaefer, M. RNA 5-Methylcytosine Analysis by Bisulfite Sequencing. *Methods Enzymol* **560**,
655 297-329 (2015). https://doi.org:10.1016/bs.mie.2015.03.007
656 57 Bourgeois, G. *et al.* Eukaryotic rRNA Modification by Yeast 5-Methylcytosine-
657 Methyltransferases and Human Proliferation-Associated Antigen p120. *PLoS ONE* **10**,
658 e0133321 (2015). https://doi.org:10.1371/journal.pone.0133321
659 58 Gu, X. & Liang, Z. Transcriptome-Wide Mapping 5-Methylcytosine by m(5)C RNA
660 Immunoprecipitation Followed by Deep Sequencing in Plant. *Methods Mol Biol* **1933**, 389-394
661 (2019). https://doi.org:10.1007/978-1-4939-9045-0_24
662 59 Selmi, T. *et al.* Sequence- and structure-specific cytosine-5 mRNA methylation by NSUN6.
663 *Nucleic Acids Res* **49**, 1006-1022 (2021). https://doi.org:10.1093/nar/gkaa1193
664 60 Kim, K., Shi, A. B., Kelley, K. & Chen, X. S. Unraveling the Enzyme-Substrate Properties for
665 APOBEC3A-Mediated RNA Editing. *J Mol Biol* **435**, 168198 (2023).
666 https://doi.org:10.1016/j.jmb.2023.168198
667 61 Barka, A. *et al.* The Base-Editing Enzyme APOBEC3A Catalyzes Cytosine Deamination in
668 RNA with Low Proficiency and High Selectivity. *ACS Chem Biol* **17**, 629-636 (2022).
669 https://doi.org:10.1021/acschembio.1c00919
670 62 Pan, Y. *et al.* Nanoscale Characterization of Interaction of APOBEC3G with RNA. *Biochemistry*
671 **56**, 1473-1481 (2017). https://doi.org:10.1021/acs.biochem.6b01189
672 63 Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing.
673 *Crispr j* **1**, 239-250 (2018). https://doi.org:10.1089/crispr.2018.0014
674 64 Bienert, S. *et al.* The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids*
675 *Res* **45**, D313-d319 (2017). https://doi.org:10.1093/nar/gkw1132
676 65 Delano, W. L. The PyMol Molecular Graphics System. *Proteins Structure Function and*
677 *Bioinformatics* **30**, 442-454 (2002).
678 66 Liu, Q. *et al.* Hi-TOM: a platform for high-throughput tracking of mutations induced by
679 CRISPR/Cas systems. *Sci China Life Sci* **62**, 1-7 (2019). https://doi.org:10.1007/s11427-018-
680 9402-9

67 Olarerin-George, A. O. & Jaffrey, S. R. MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. *Bioinformatics* **33**, 1563-1564 (2017). https://doi.org:10.1093/bioinformatics/btx002

68 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org:10.1093/bioinformatics/btq033

69 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004). https://doi.org:10.1101/gr.849004

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

FIGURES LEGENDS

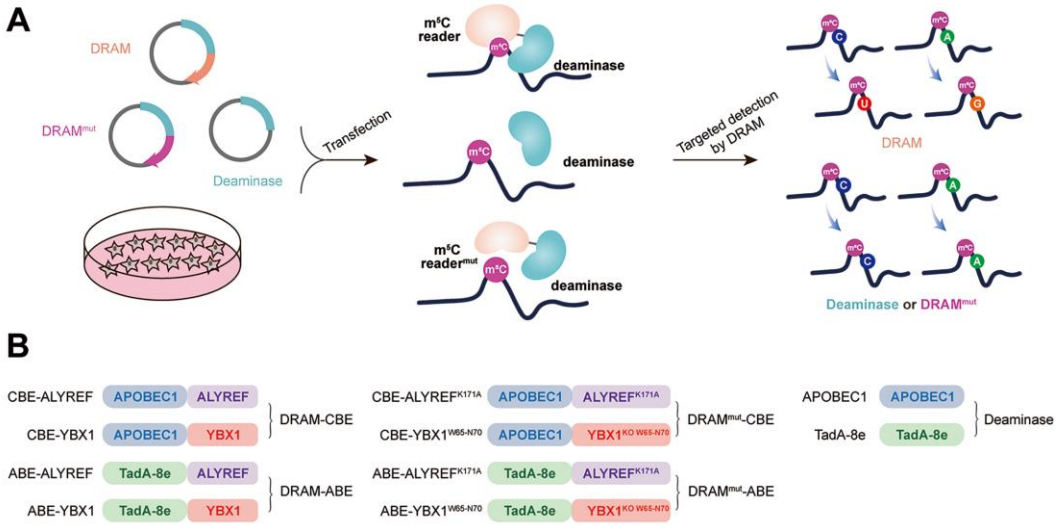


Fig.1: **Development of DRAM system for m^{5}C detection.**

(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 m^{5}C reader to the vicinity of the m^{5}C 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 m^{5}C-recognition-binding domain.

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

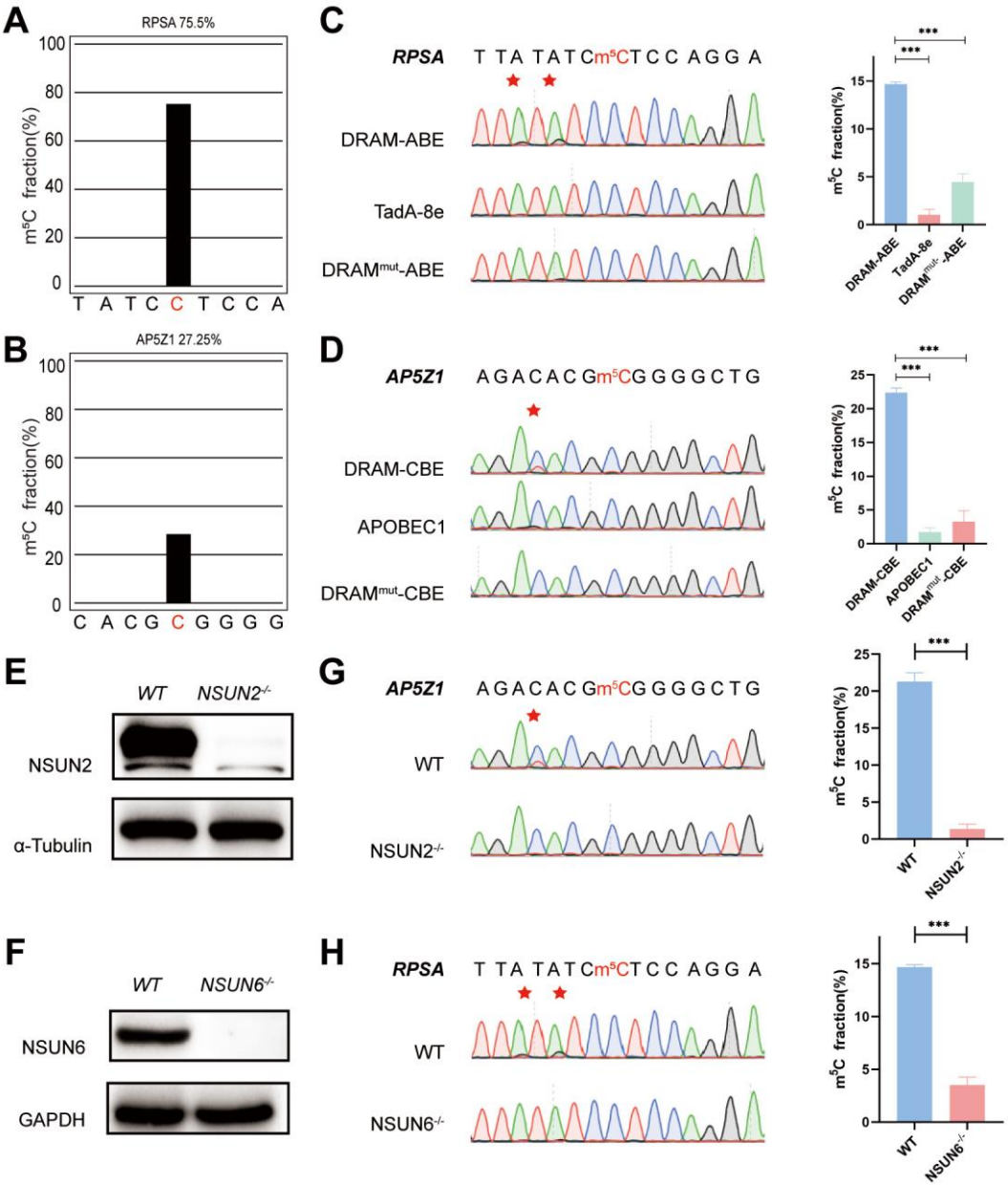


Fig.2: DRAM detection system was assayed in an m^{5}C -dependent form.

(A, B) Two m^{5}C sites from RPSA (**A**) and AP5Z1 (**B**) mRNA detected by deep sequencing of bisulfite sequencing PCR in HEK293T cells. The m^{5}C sites are highlighted by red color. The m^{5}C 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 m^{5}C 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.

715 **(G, H)** DRAM induced mutations close to m^5C sites in AP5Z1 (**G**) and RPSA (**H**) mRNAs after
716 NSUN2 and NSUN6 knockout in HEK293T cells. The left panel illustrates the location of DRAM
717 induced mutation sites, which is highlighted in red asterisk. The right panel shows the corresponding
718 quantification of sanger sequencing.
719

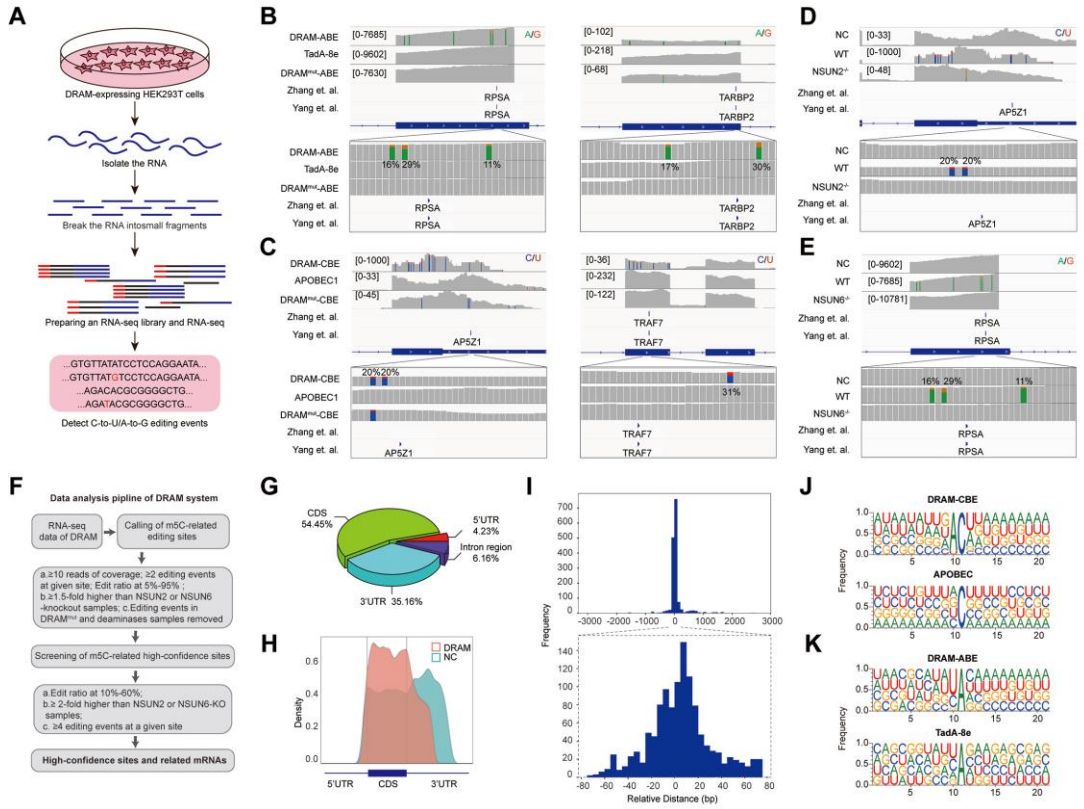


Fig.3: DRAM enables transcriptome-wide analysis of m^{5}C methylation.

(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 methyltranferases 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 m^{5}C sites from the published BS-seq datasets. The position of each m^{5}C 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.

744 **(J, K)** Motif analysis discovered within the ±20nt region around the C-to-U or A-to-G editing site
745 in cells expressing DRAM-CBE (**J**), APOBEC1(**J**), DRAM-ABE (**K**) and TadA-8e (**K**).
746

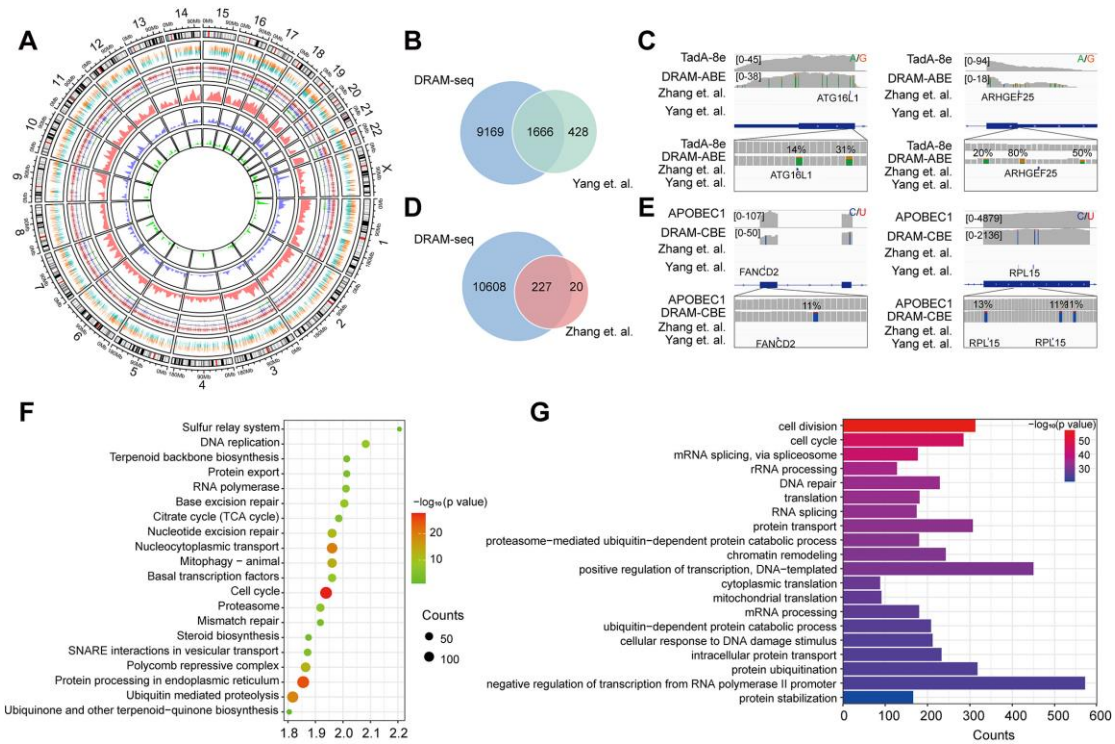


Fig.4: Stable and comprehensive cellular identification of m^{5}C loci by DRAM-seq.

(A) Comparison of the overall distribution of genes with m^{5}C modifications detected by DRAM-seq, Yang et al. and Zhang et al. on chromosomes. The mutation sites detected by DRAM-seq on each gene are categorized into dual-colored short lines, with positive strand mutations shown in orange and negative strand mutations in dark green. The line graph and kernel density plot in the inner ring represent the locations and distributions of overlapping genes detected by DRAM-seq (red), Yang et al. (blue) and Zhang et al. (light green).

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

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

(C, E) 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 m^{5}C site found by BS-seq is also labelled.

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

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

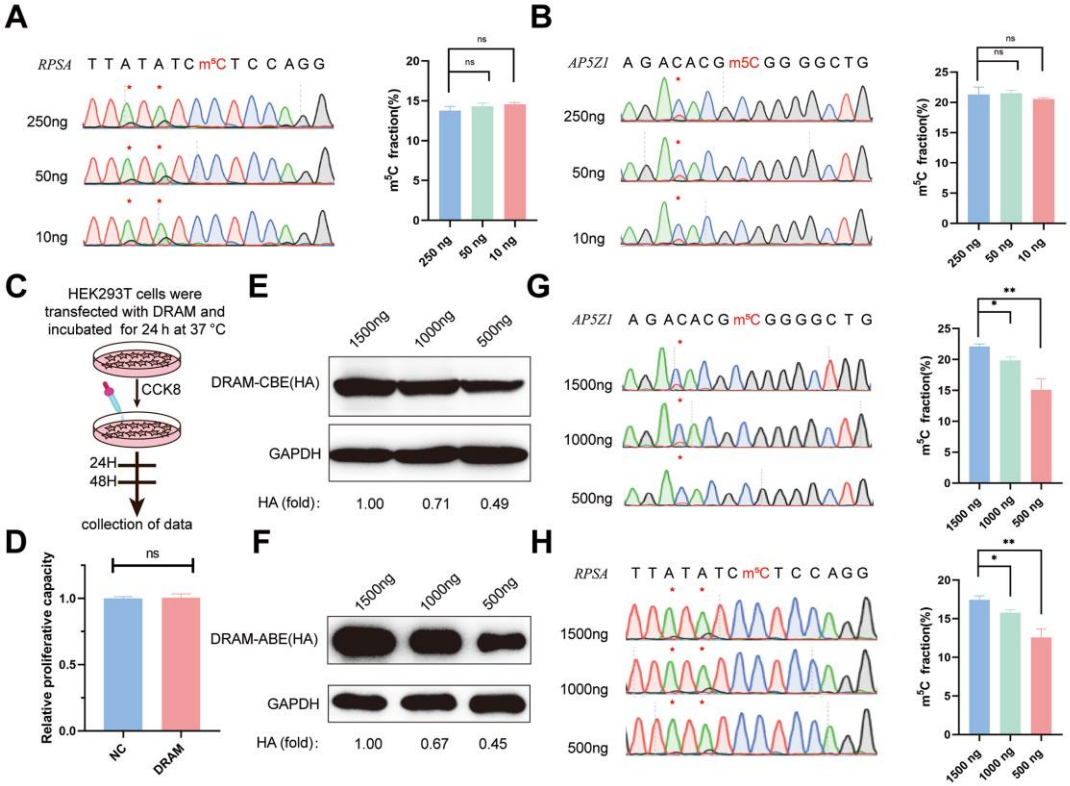


Fig.5: Low-input m^{5}C detection and transfection efficiency of DRAM system.

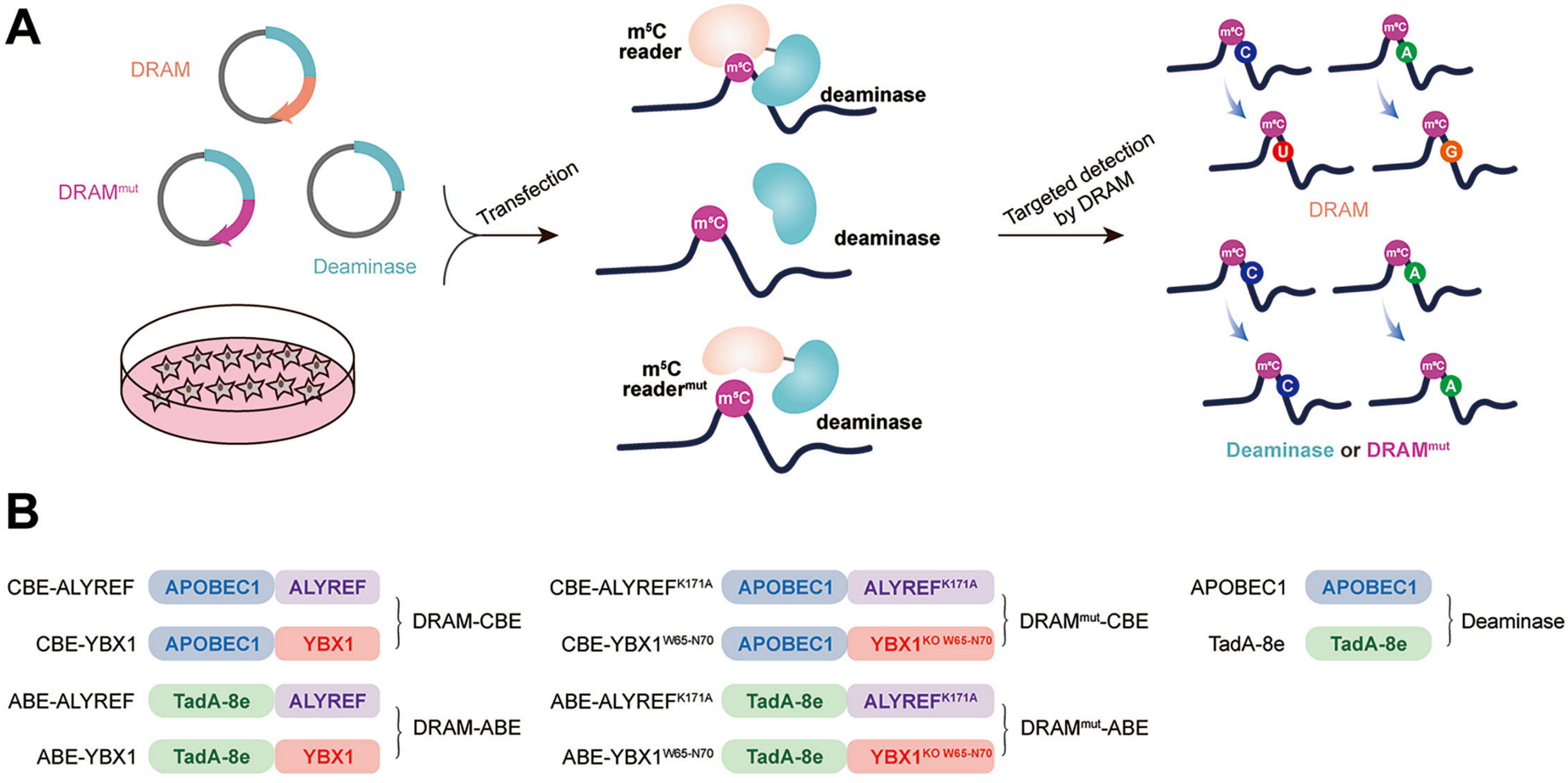
(**A**, **B**) DRAM analysis of RPSA (**A**) and AP5Z1 (**B**) mRNAs with 250 ng, 50 ng, and 10 ng of input RNA. Representative Sanger sequencing plots are shown on the left panel, with mutation sites marked with asterisks. The mutation rates are quantified on the right panel.

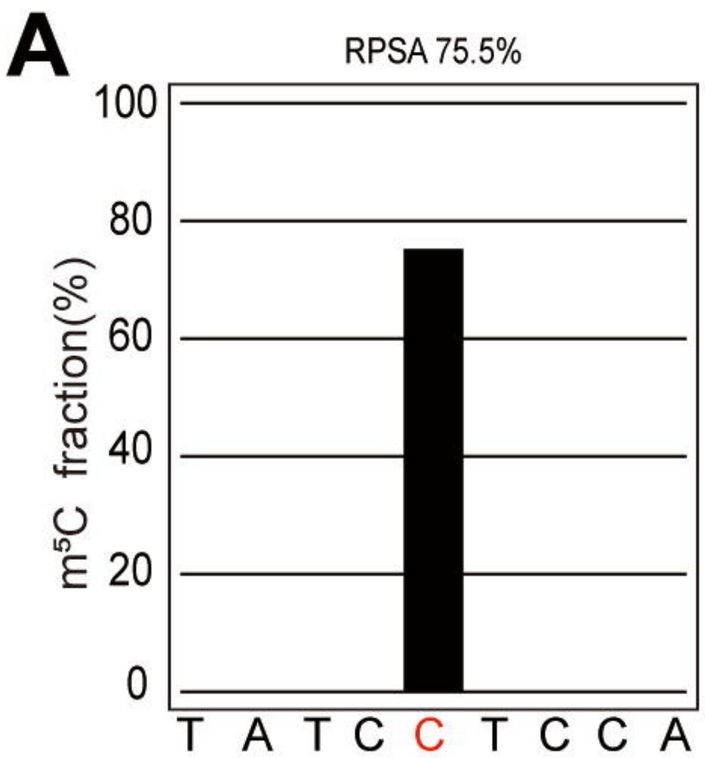
(**C**) Flowchart illustrating Cell viability analysis by CCK8 reagent after DRAM transfection in HEK293T cells.

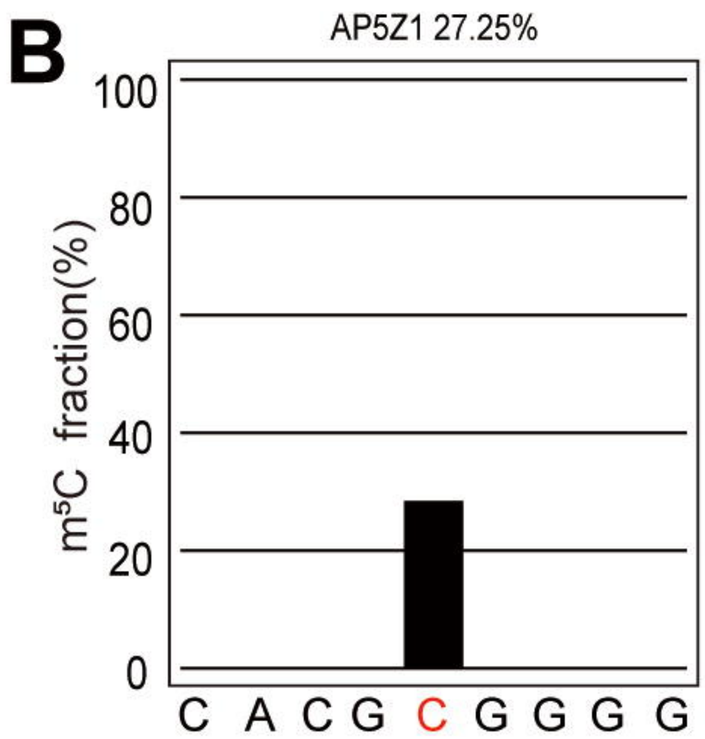
(**D**) Quantitative comparison of the relative proliferative capacity of DRAM-expressing and untransfected cells.

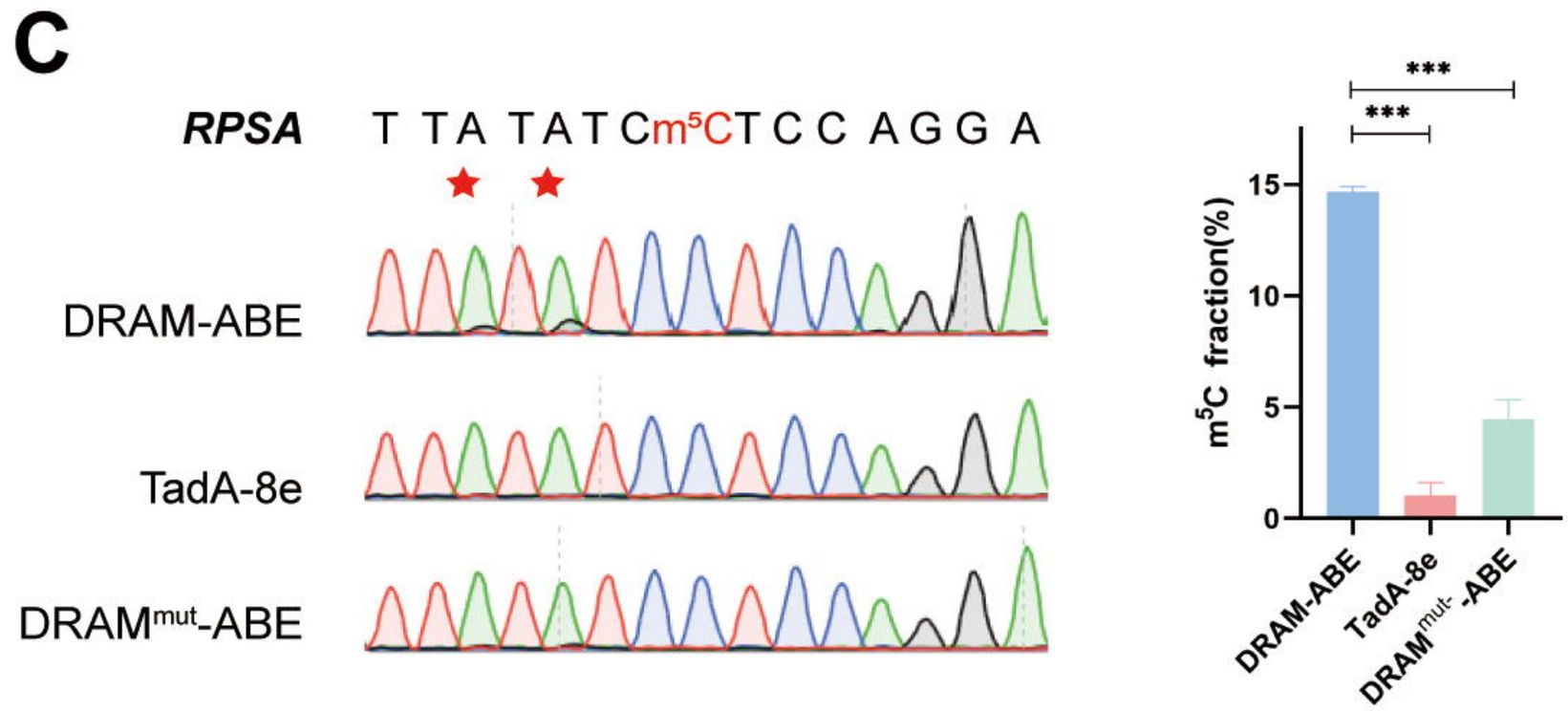
(**E**, **F**) The expression levels of DRAM-CBE **(E)** and DRAM-ABE **(F)** systems different plasmid transfection concentrations were verified by Western blotting.

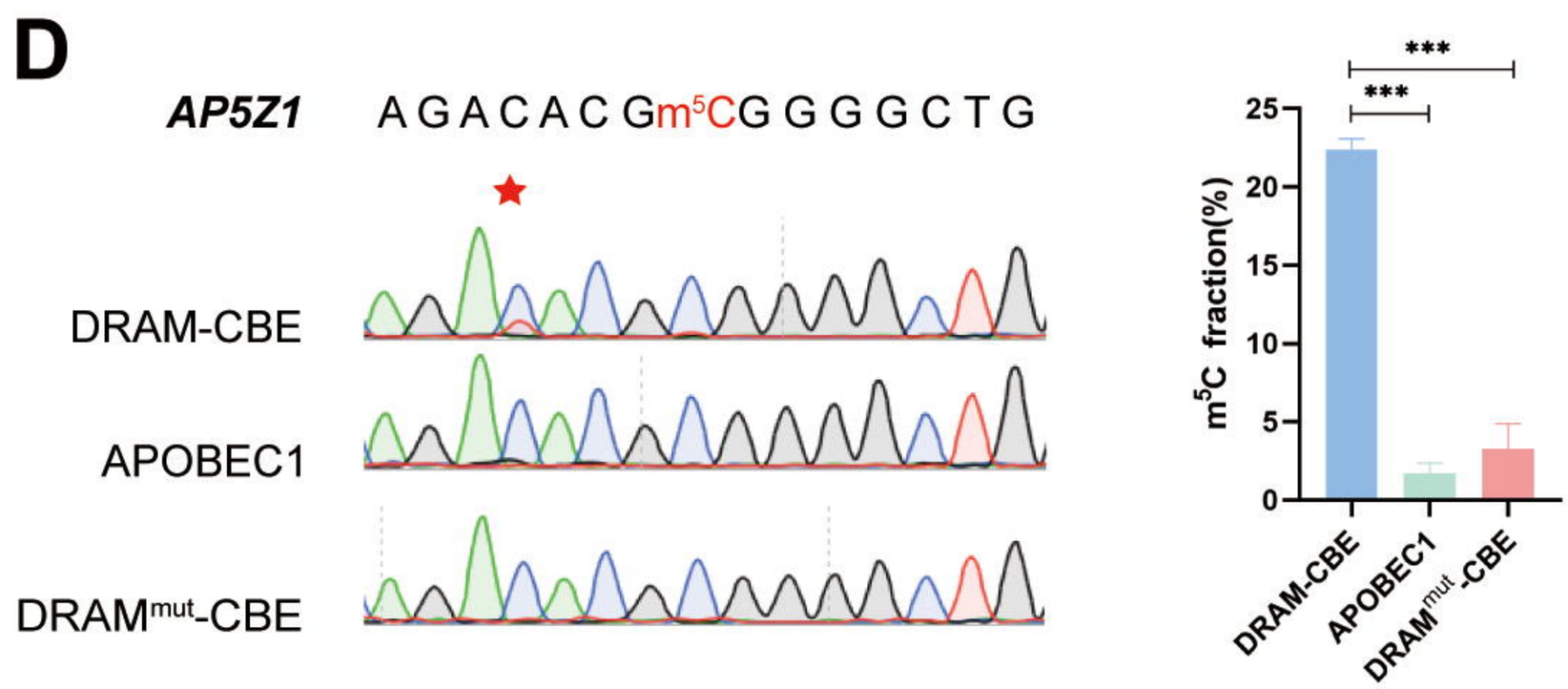
(**G**, **H**) Editing of RPSA **(G)** and AP5Z1 **(H)** mRNA at varying concentrations of DRAM protein expression. The left panels indicate Sanger sequencing results following RT-PCR, while the corresponding quantifications of DRAM-induced mutations are shown in the right panels.

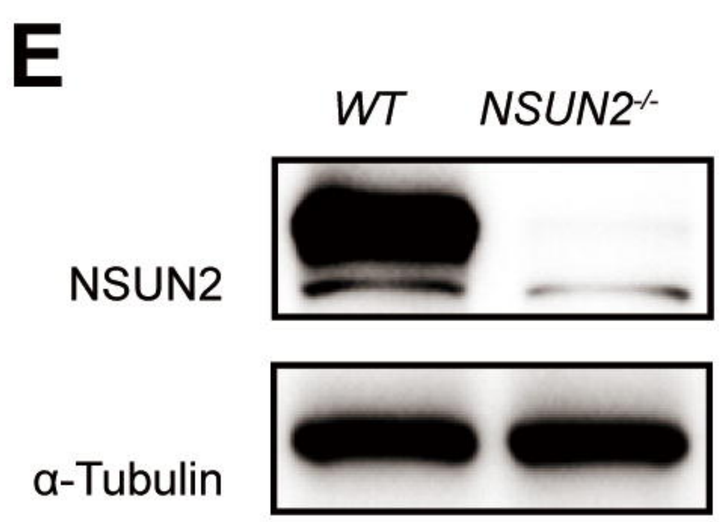

A
DRAM
DRAMmut
Deaminase
Transfection
m5C reader
m5C
deaminase
m5C
deaminase
m5C readermut
m5C
deaminase
Targeted detection by DRAM
m6C
C
A
U
G
DRAM
Deaminase or DRAMmut
B
CBE-ALYREF
APOBEC1
ALYREF
CBE-YBX1
APOBEC1
YBX1
DRAM-CBE
ABE-ALYREF
TadA-8e
ALYREF
ABE-YBX1
TadA-8e
YBX1
DRAM-ABE
CBE-ALYREFK171A
APOBEC1
ALYREFK171A
CBE-YBX1W65-N70
APOBEC1
YBX1KO W65-N70
DRAMmut-CBE
ABE-ALYREFK171A
TadA-8e
ALYREFK171A
ABE-YBX1W65-N70
TadA-8e
YBX1KO W65-N70
DRAMmut-ABE
APOBEC1
APOBEC1
TadA-8e
TadA-8e
Deaminase

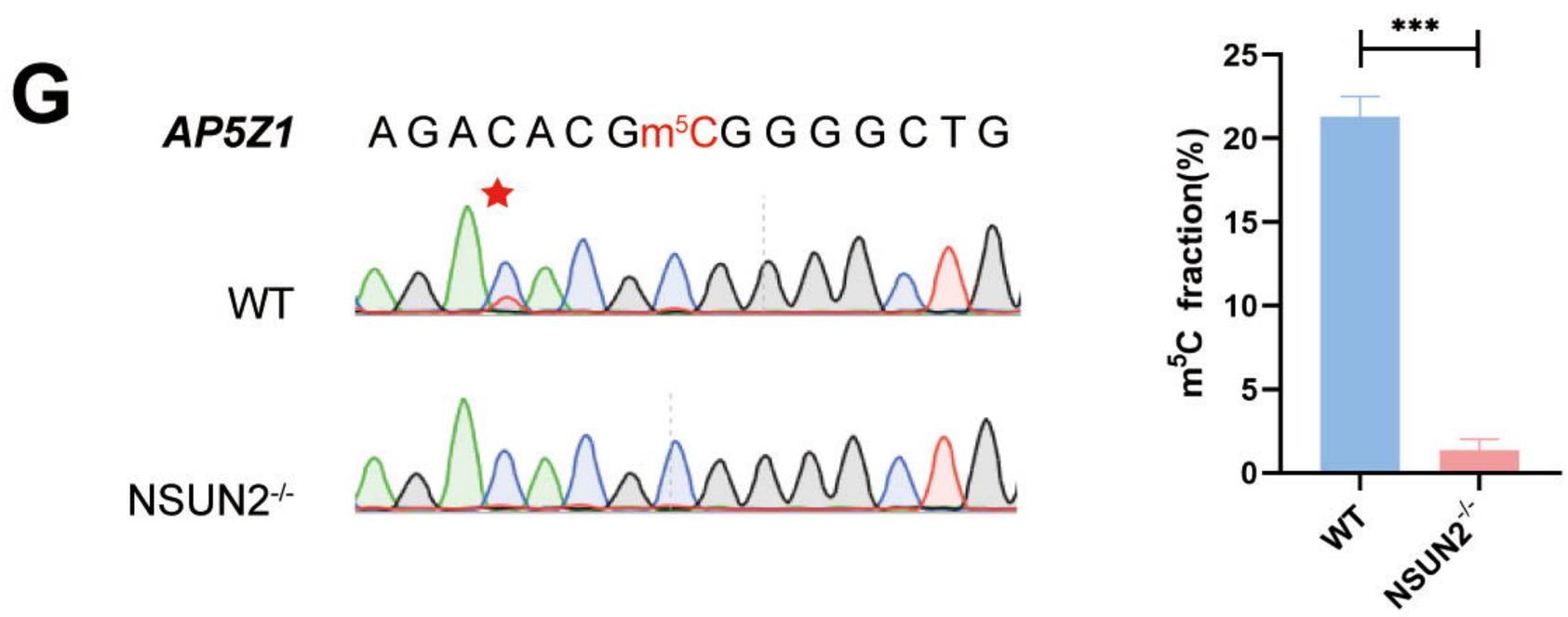

A
RPSA 75.5%
m5C fraction(%)
100
80
60
40
20
0
T A T C C T C C A

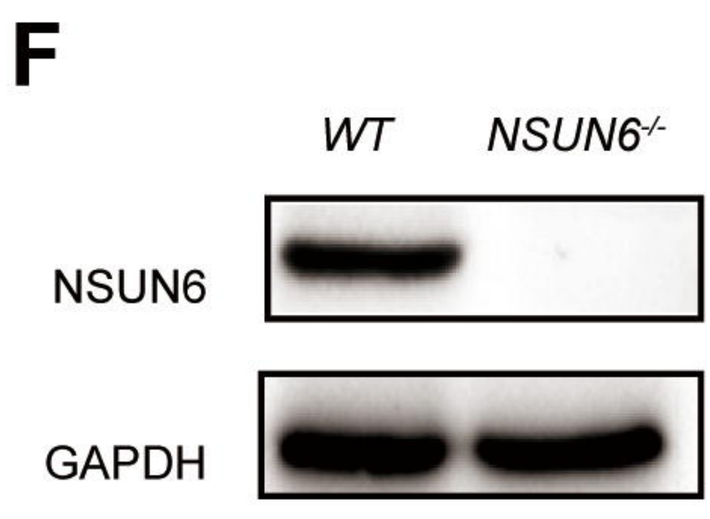

B
AP5Z1 27.25%
m5C fraction(%)
100
80
60
40
20
0
C A C G C G G G G

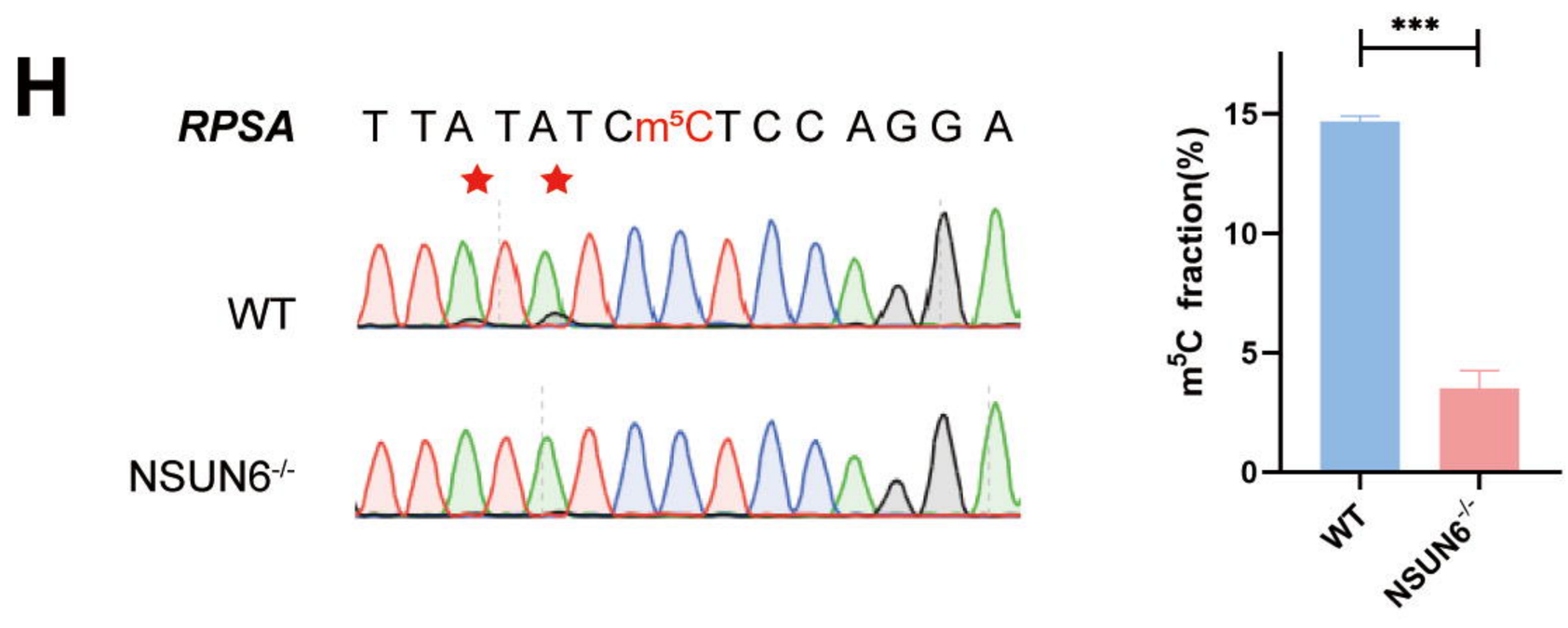

C
RPSA
T T A T A T C m5C T C C A G G A
DRAM-ABE
TadA-8e
DRAMmut-ABE
m5C fraction(%)
15
10
5
0
***
***
DRAM-ABE
TadA-8e
DRAMmut-ABE

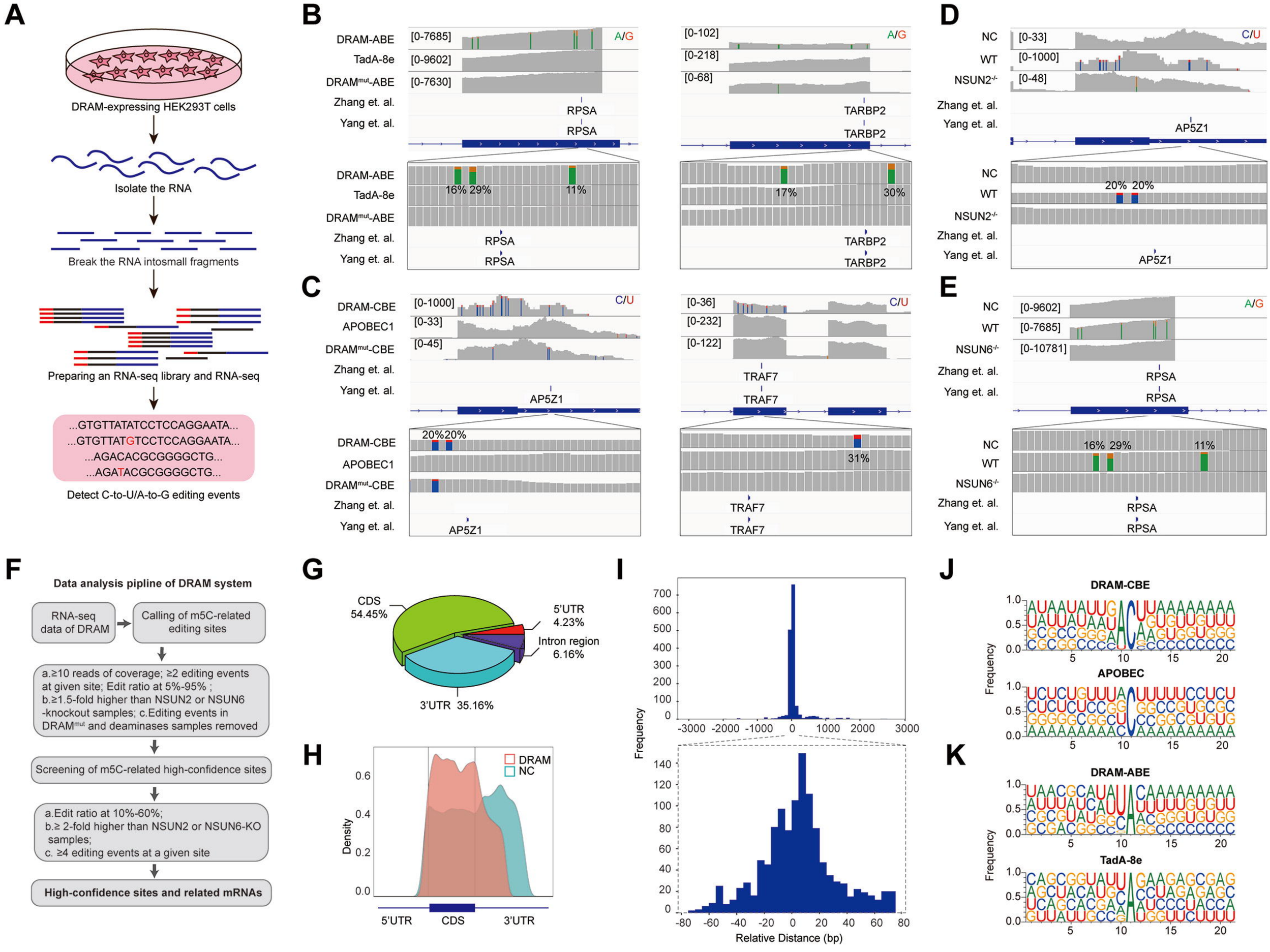

D
AP5Z1
A G A C A C G m5C G G G G C T G
DRAM-CBE
APOBEC1
DRAMmut-CBE
m5C fraction(%)
25
20
15
10
5
0
***
***
DRAM-CBE
APOBEC1
DRAMmut-CBE

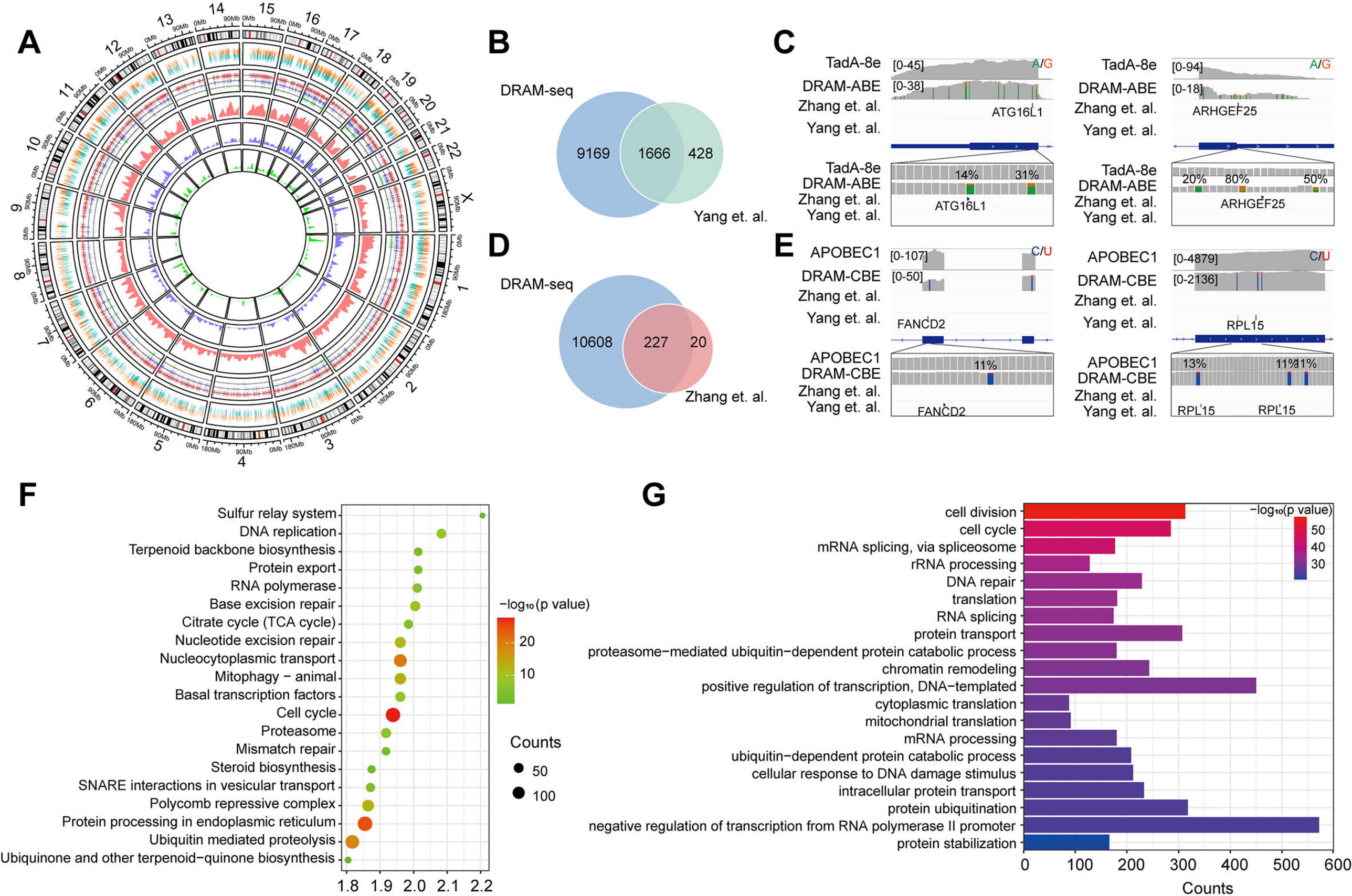

E
WT
NSUN2-/-
NSUN2
α-Tubulin

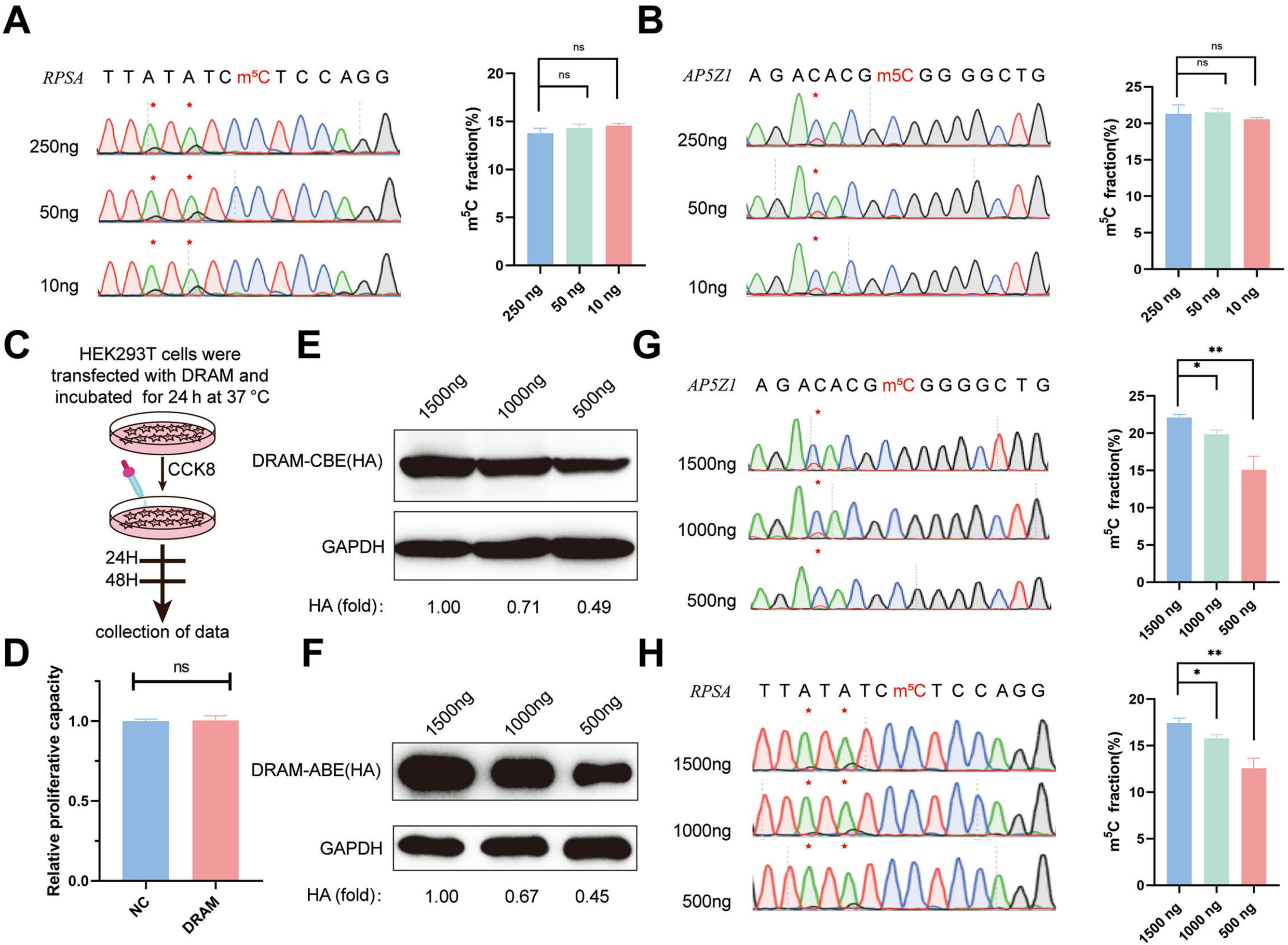

G
AP5Z1
A G A C A C G m5C G G G G C T G
WT
NSUN2-/-
m5C fraction(%)
25
20
15
10
5
0
***
WT
NSUN2-/-


F
WT
NSUN6-/-
NSUN6
GAPDH


H
RPSA
T T A T A T C m5C T C C A G G A
WT
NSUN6-/-
m5C fraction(%)
15
10
5
0
***
WT
NSUN6-/-


A
DRAM-expressing HEK293T cells
Isolate the RNA
Break the RNA intosmall fragments
Preparing an RNA-seq library and RNA-seq
...GTGTTATATCCTCCAGGAATA...
...GTGTTATGTCCTCCAGGAATA...
...AGACACGCGGGGCTG...
...AGATACGCGGGGCTG...
Detect C-to-U/A-to-G editing events
B
DRAM-ABE
TadA-8e
DRAMmut-ABE
Zhang et. al.
Yang et. al.
[0-7685]
[0-9602]
[0-7630]
A/G
RPSA
16% 29%
11%
[0-102]
[0-218]
[0-68]
TARBP2
17%
30%
C
DRAM-CBE
APOBEC1
DRAMmut-CBE
[0-1000]
[0-33]
[0-45]
C/U
AP5Z1
20%20%
[0-36]
[0-232]
[0-122]
TRAF7
31%
D
NC
WT
NSUN2-/-
[0-33]
[0-1000]
[0-48]
20% 20%
E
NSUN6-/-
[0-9602]
[0-7685]
[0-10781]
F
Data analysis pipline of DRAM system
RNA-seq data of DRAM
Calling of m5C-related editing sites
a.≥10 reads of coverage; ≥2 editing events at given site; Edit ratio at 5%-95% ; b.≥1.5-fold higher than NSUN2 or NSUN6 -knockout samples; c.Editing events in DRAMmut and deaminases samples removed
Screening of m5C-related high-confidence sites
a.Edit ratio at 10%-60%; b.≥ 2-fold higher than NSUN2 or NSUN6-KO samples; c. ≥4 editing events at a given site
High-confidence sites and related mRNAs
G
CDS 54.45%
5'UTR 4.23%
Intron region 6.16%
3'UTR 35.16%
H
DRAM
NC
Density
5'UTR
CDS
3'UTR
I
Frequency
Relative Distance (bp)
J
DRAM-CBE
APOBEC
K
DRAM-ABE
TadA-8e


A
B
DRAM-seq
9169
1666
428
Yang et. al.
C
TadA-8e
DRAM-ABE
Zhang et. al.
Yang et. al.
[0-45]
[0-38]
A/G
ATG16L1
14%
31%
[0-94]
[0-18]
ARHGEF25
20%
80%
50%
D
DRAM-seq
10608
227
20
Zhang et. al.
E
APOBEC1
DRAM-CBE
[0-107]
[0-50]
C/U
FANCD2
11%
[0-4879]
[0-2136]
RPL15
13%
11%
11%
F
Sulfur relay system
DNA replication
Terpenoid backbone biosynthesis
Protein export
RNA polymerase
Base excision repair
Citrate cycle (TCA cycle)
Nucleotide excision repair
Nucleocytoplasmic transport
Mitophagy − animal
Basal transcription factors
Cell cycle
Proteasome
Mismatch repair
Steroid biosynthesis
SNARE interactions in vesicular transport
Polycomb repressive complex
Protein processing in endoplasmic reticulum
Ubiquitin mediated proteolysis
Ubiquinone and other terpenoid−quinone biosynthesis
1.8 1.9 2.0 2.1 2.2
−log₁₀(p value)
20
10
Counts
50
100
G
cell division
cell cycle
mRNA splicing, via spliceosome
rRNA processing
DNA repair
translation
RNA splicing
protein transport
proteasome−mediated ubiquitin−dependent protein catabolic process
chromatin remodeling
positive regulation of transcription, DNA−templated
cytoplasmic translation
mitochondrial translation
mRNA processing
ubiquitin−dependent protein catabolic process
cellular response to DNA damage stimulus
intracellular protein transport
protein ubiquitination
negative regulation of transcription from RNA polymerase II promoter
protein stabilization
0 100 200 300 400 500 600
Counts
−log₁₀(p value)
50
40
30


A
RPSA T T A T A T C m5C T C C A G G
250ng
50ng
10ng
ns
ns
m5C fraction(%)
250 ng
50 ng
10 ng
B
AP5Z1 A G A C A C G m5C G G G G C T G
250ng
50ng
10ng
ns
ns
m5C fraction(%)
250 ng
50 ng
10 ng
C
HEK293T cells were transfected with DRAM and incubated for 24 h at 37 °C
CCK8
24H
48H
collection of data
D
ns
Relative proliferative capacity
NC
DRAM
E
1500ng
1000ng
500ng
DRAM-CBE(HA)
GAPDH
HA (fold): 1.00 0.71 0.49
F
1500ng
1000ng
500ng
DRAM-ABE(HA)
GAPDH
HA (fold): 1.00 0.67 0.45
G
AP5Z1 A G A C A C G m5C G G G G C T G
1500ng
1000ng
500ng
**
*
m5C fraction(%)
1500 ng
1000 ng
500 ng
H
RPSA T T A T A T C m5C T C C A G G
1500ng
1000ng
500ng
**
*
m5C fraction(%)
1500 ng
1000 ng
500 ng