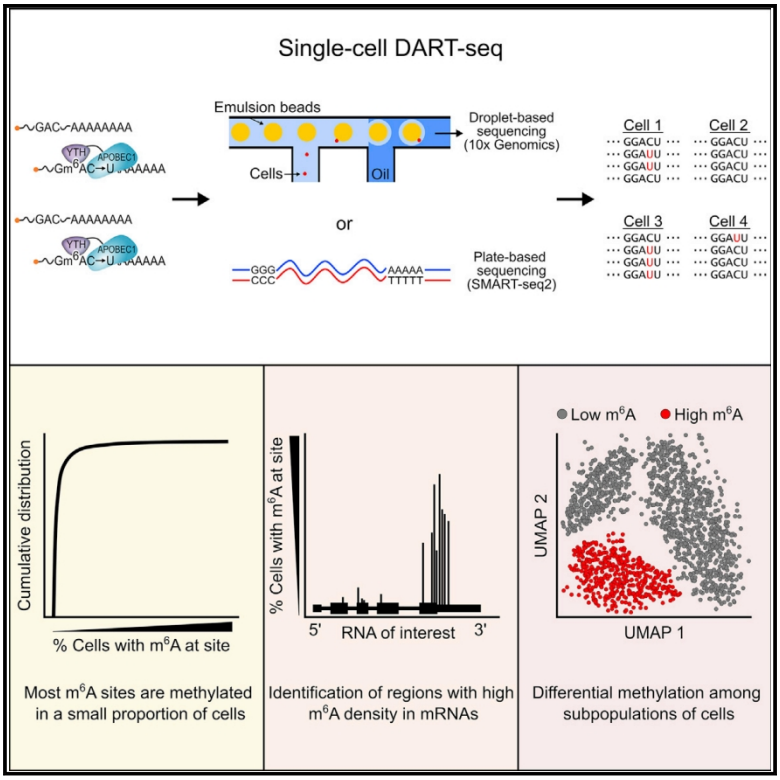
Molecular Cell

Resource

scDART-seq reveals distinct m^{6}A signatures and mRNA methylation heterogeneity in single cells

Graphical abstract



Highlights

- scDART-seq achieves global m^{6}A profiling in thousands of single cells
- Most mRNAs are methylated in a small fraction of cells
- RNAs have many more m^{6}A sites than previously reported, but most occur rarely
- m^{6}A stoichiometry is highly variable across individual cells of a population

Authors

Matthew Tegowski,
Mathieu N. Flamand, Kate D. Meyer

Correspondence

kate.meyer@duke.edu

In brief

Tegowski et al. use single-cell m^{6}A profiling to map the methylomes of thousands of individual cells. They find that m^{6}A is highly heterogeneous at the single-cell level, with most m^{6}A sites occurring in a small proportion of cells. They also uncover differential methylation within subpopulations of cells.

Tegowski et al., 2022, Molecular Cell *82*, 868–878
February 17, 2022
https://doi.org/10.1016/j.molcel.2021.12.038

CellPress

CellPress

Molecular Cell

Resource

scDART-seq reveals distinct m^{6}A signatures and mRNA methylation heterogeneity in single cells

Matthew Tegowski,[1] Mathieu N. Flamand,[1] and Kate D. Meyer[1,2,3,*]
[1]Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710, USA
[2]Department of Neurobiology, Duke University School of Medicine, Durham, NC 27710, USA
[3]Lead contact
*Correspondence: kate.meyer@duke.edu
https://doi.org/10.1016/j.molcel.2021.12.038

SUMMARY

N^6-methyladenosine (m^{6}A) is an abundant RNA modification that plays critical roles in RNA regulation and cellular function. Global m^{6}A profiling has revealed important aspects of m^{6}A distribution and function, but to date such studies have been restricted to large populations of cells. Here, we develop a method to identify m^{6}A sites transcriptome-wide in single cells. We uncover surprising heterogeneity in the presence and abundance of m^{6}A sites across individual cells and identify differentially methylated mRNAs across the cell cycle. Additionally, we show that cellular subpopulations can be distinguished based on their RNA methylation signatures, independent from gene expression. These studies reveal fundamental features of m^{6}A that have been missed by m^{6}A profiling of bulk cells and suggest the presence of cell-intrinsic mechanisms for m^{6}A deposition.

INTRODUCTION

N^6-methyladenosine (m^{6}A) is a widespread RNA modification that plays diverse roles in gene expression control and contributes to a variety of physiological processes and disease states (Shi et al., 2019; Zaccara et al., 2019). Thus, identifying the distribution of m^{6}A in distinct RNAs and within individual cells of a population is critical for understanding how m^{6}A residues contribute to normal cellular function and disease pathogenesis. However, current m^{6}A detection strategies require high amounts of input RNA, and as a result, all transcriptome-wide m^{6}A profiling studies to date have mapped m^{6}A in bulk cellular populations comprising thousands or millions of cells. Thus, there is a great need to develop m^{6}A profiling methods that can identify m^{6}A sites within single cells.

Recently, our group developed deamination adjacent to RNA modification targets (DART-seq), a method for transcriptome-wide m^{6}A mapping, which utilizes a fusion protein consisting of the m^{6}A-binding YTH domain tethered to the cytidine deaminase APOBEC1 to direct C-to-U editing at cytidine residues that invariably follow m^{6}A sites (Meyer, 2019) (Figure 1A). These editing sites are then identified from RNA-seq data, enabling nucleotide-resolution m^{6}A mapping. Unlike antibody-based approaches, DART-seq does not detect the cap-associated modification m^{6}A$_m$. Additionally, DART-seq can map m^{6}A from ultra-low-input amounts of RNA (Meyer, 2019). Therefore, we reasoned that this strategy could be used to identify m^{6}A sites in single cells.

Here, we describe single-cell DART-seq (scDART-seq), the first method for transcriptome-wide profiling of m^{6}A sites in single cells. Using both the droplet-based 10x Genomics and plate-based SMART-seq2 platforms, we identify the single-cell methylomes of thousands of individual cells. Although the general topology of m^{6}A identified from bulk cellular samples is recapitulated in single cells, we uncover substantial heterogeneity in the methylation of RNAs and individual m^{6}A sites across distinct cells within a population. Analysis of m^{6}A stoichiometry at individual sites within mRNAs also reveals broad variability across cells, with many m^{6}A sites that are highly methylated at the population level showing low or no methylation in a substantial number of individual cells. Finally, we uncover unique m^{6}A signatures associated with specific cell-cycle phases and show that individual cells can be distinguished based on their RNA methylation status, revealing a source of cellular heterogeneity independent from RNA abundance. Altogether, our studies uncover fundamental features of m^{6}A biology that have been missed by m^{6}A profiling in bulk cells and provide new insights into the regulation and function of m^{6}A within distinct cells of a population.

RESULTS

scDART-seq identifies m^{6}A sites in single cells

To determine whether DART-seq can map m^{6}A sites in single cells, we first created HEK293T stable cell lines expressing inducible APOBEC1-YTH or APOBEC1-YTH^{mut}, which lacks the m^{6}A-binding region of the YTH domain and facilitates the removal of off-target editing events in downstream analyses (Meyer, 2019) (Figures S1A–S1H). We then induced APOBEC1-YTH or APOBEC1-YTH^{mut} expression and performed DART-seq on bulk cells to identify m^{6}A sites transcriptome-wide. To identify m^{6}A sites, we developed Bullseye, a pipeline that identifies C-to-U editing events that are enriched in APOBEC1-YTH

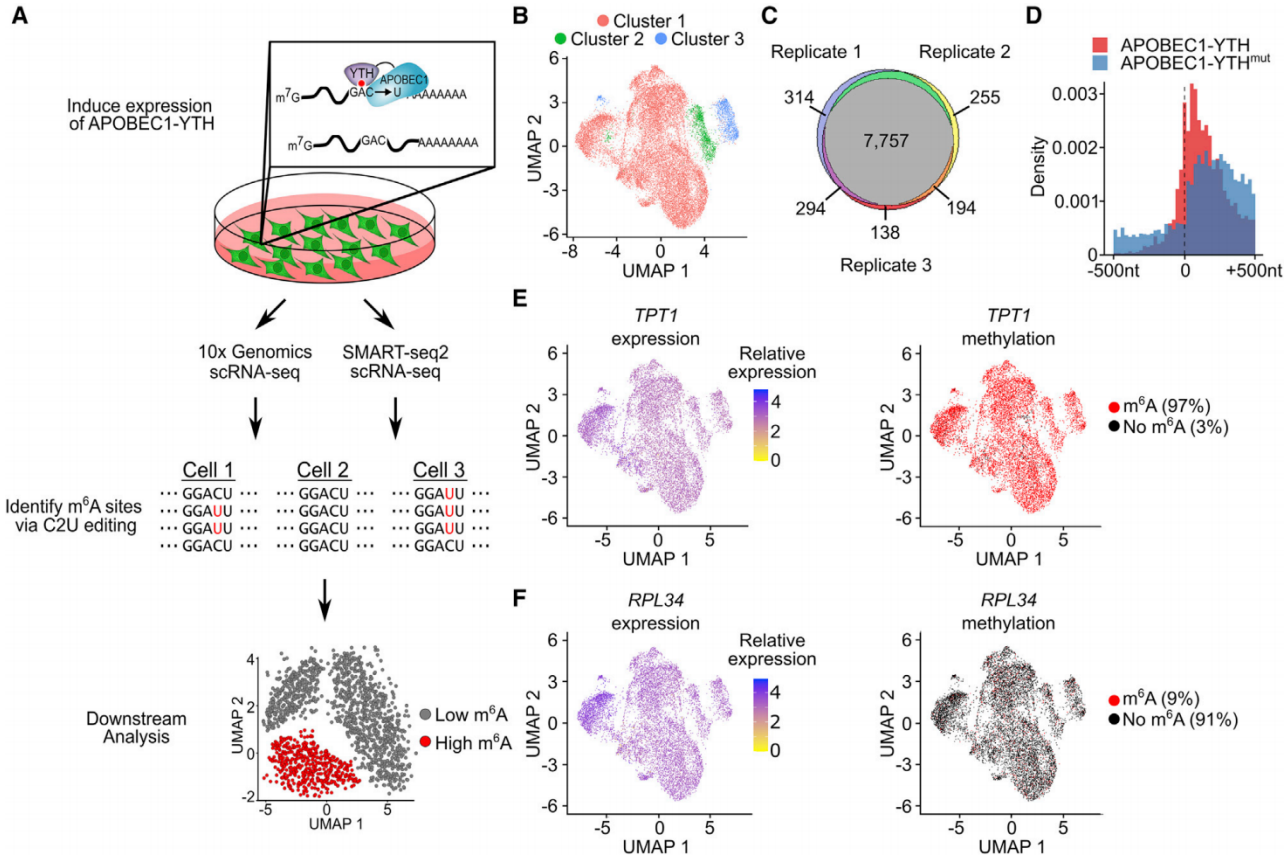


Figure 1. scDART-seq detects m^{6}A in single cells, see also Figures S1 and S2

(A) Experimental workflow of scDART-seq.

(B) UMAP visualization of all 10x Genomics-sequenced cells clustered by gene expression.

(C) Overlap of methylated RNAs identified in each 10x Genomics scDART-seq replicate, before filtering sites for those occurring in all 3 replicates and in at least 10 cells.

(D) Distribution of C-to-U editing events identified by scDART-seq. The stop codon is centered at position 0. Bins = 25 nt.

(E) Single-cell expression levels (left) and methylation status (right) for the *TPT1* mRNA in 10x Genomics-sequenced cells overlaid on the global gene expression UMAP plot.

(F) Single-cell expression levels (left) and methylation status (right) for the *RPL34* mRNA in 10x Genomics-sequenced cells overlaid on the global gene expression UMAP plot.

cells compared with APOBEC1-YTH^{mut} cells and that meet a minimum editing threshold (STAR Methods). Using this strategy, we identified 12,672 m^{6}A sites in 5,200 RNAs (Table S1). These sites showed characteristic features of m^{6}A, such as enrichment in proximal 3′UTRs and in the vicinity of the stop codon, and there was a high degree of overlap between methylated RNAs identified by DART-seq and those identified by MeRIP-Seq (Liu et al., 2020; Meyer et al., 2012) and miCLIP (Linder et al., 2015) (Figures S1I–S1M). These results are consistent with previous studies demonstrating the ability of DART-seq to accurately map m^{6}A transcriptome-wide in populations of cells.

To identify m^{6}A sites in single cells, we induced APOBEC1-YTH expression and performed droplet-based (10x Genomics) scRNA-seq on 10,352 total cells across three biological replicates (Figures 1A, 1B, and S2A–S2H). We then used Bullseye to identify m^{6}A sites in single cells and found 16,934 high-confidence sites in 3,844 RNAs in 10,352 cells (Table S2). Detection of m^{6}A sites and methylated RNAs was highly consistent across samples, indicating the reproducibility of the scDART-seq approach (Figure 1C). Single-cell m^{6}A sites exhibit similar features to those identified at the population level, including a strong enrichment in the vicinity of the stop codon and in the RAC (R=A/G) consensus sequence (Figures 1D and S2I). We also observed a high degree of overlap between the methylated RNAs identified using scDART-seq and those identified in bulk cells using antibody-based m^{6}A mapping (Linder et al., 2015; Liu et al., 2020; Meyer et al., 2012) (Figures S2J and S2K). Thus, these data demonstrate that scDART-seq can identify m^{6}A sites in single cells with high accuracy and reproducibility.

We next analyzed the distribution of methylated RNAs within individual cells of the population. We observed a high degree of variability in the proportion of cells in which individual RNAs are methylated. Some transcripts, such as *TPT1*, are highly expressed in most cells and also methylated in the majority of cells

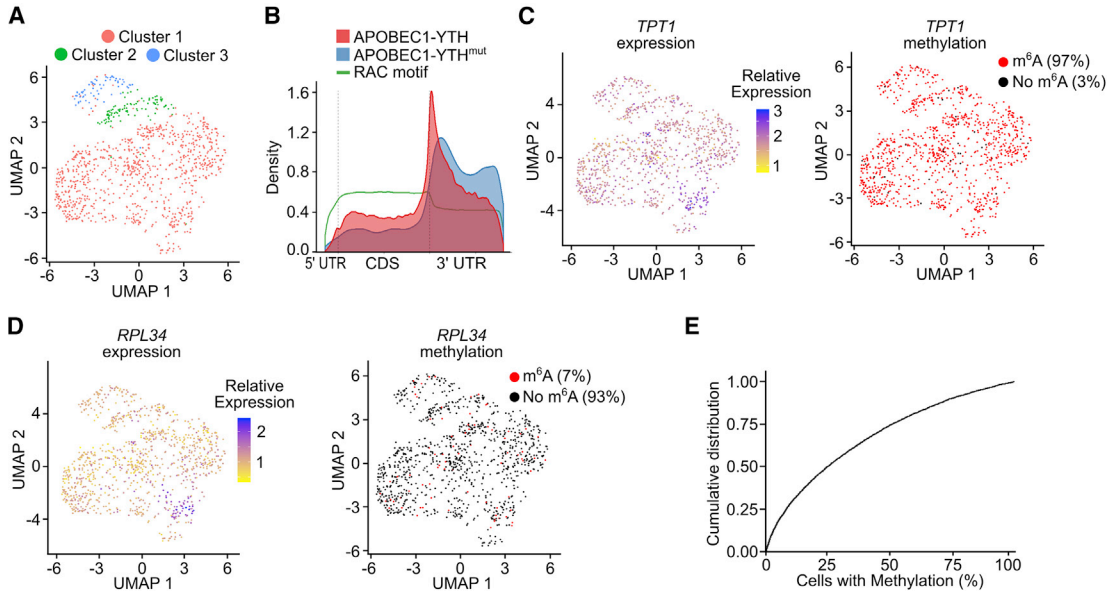


Figure 2. High-coverage scDART-seq detects m^{6}A transcriptome-wide in single cells, see also Figure S3

(A) UMAP visualization of SMART-seq2-sequenced cells clustered by gene expression.

(B) Distribution of C-to-U editing sites across mRNA length for cells expressing APOBEC1-YTH and APOBEC1-YTH^{mut}. Distribution of the RAC motif is also shown.

(C) Single-cell expression levels (left) and methylation status (right) for the *TPT1* mRNA in SMART-seq2-sequenced cells overlaid on the global gene expression UMAP plot.

(D) Single-cell expression levels (left) and methylation status (right) for the *RPL34* mRNA in SMART-seq2-sequenced cells overlaid on the global gene expression UMAP plot.

(E) Cumulative distribution plot showing the percentage of cells in which each methylated RNA contains at least one m^{6}A site.

in which they are expressed (Figure 1E). However, other RNAs, such as *RPL34*, also exhibit high expression levels but are rarely methylated (Figure 1F). This variability in methylation across the population is not limited to highly expressed RNAs, as we found many examples of weakly expressed transcripts that are methylated in a high or low proportion of cells (Figures S2L and S2M; Table S2).

To ensure that this variability in single-cell RNA methylation frequency is not due to the low coverage of the droplet-based approach, we also performed scDART-seq using the SMART-seq2 technology, which enables high single-cell read counts and improved coverage across transcripts (Picelli et al., 2013, 2014) (Figures 2A and S3A–S3J; Table S3). The improved coverage and sequencing depth enabled the use of more stringent criteria for identifying m^{6}A sites (STAR Methods), which we employed to obtain a list of 120,904 high-confidence m^{6}A sites in 7,820 RNAs in 1,068 cells (Figure S3K; Table S3). Metagene analysis of all high-confidence sites showed an m^{6}A-like distribution characterized by a strong enrichment in the vicinity of the stop codon (Figure 2B). Similar metagene profiles were observed in multiple randomly selected single cells, indicating that the distribution of m^{6}A within individual cells mirrors that of the larger population (Figure S3L). As expected, the improved read depth and coverage across transcripts of SMART-seq2 libraries enabled the identification of many more m^{6}A sites in the 5′UTR and CDS compared with 10x Genomics scDART-seq libraries (Figures S3M and S3N). Importantly, 91% of methylated mRNAs identified in the 10x Genomics dataset were also identified in the SMART-seq2 dataset, and the majority of methylated mRNAs identified in bulk HEK293T samples, using both DART-seq and antibody-based methods, were identified in single cells using SMART-seq2 scDART-seq (Figures S3O–S3Q). We also observed that mRNAs methylated in a small proportion of cells are less likely to be identified in bulk m^{6}A mapping datasets than mRNAs that are methylated in many cells (Figure S3R). Thus, scDART-seq enables robust m^{6}A detection in single cells and can identify infrequently methylated mRNAs that would otherwise evade detection by standard bulk m^{6}A mapping approaches.

Consistent with what we observed in the 10x Genomics scDART-seq dataset, we found widespread variability in the frequency with which individual RNAs are methylated. For instance, re-examination of the highly expressed mRNAs *TPT1* and *RPL34* again revealed extreme differences in methylation frequency (Figures 2C and 2D). Furthermore, when we analyzed all methylated RNAs, we found that 70% of m^{6}A-containing RNAs are methylated in fewer than half the cells in the population (Figures 2E and S3S). Thus, although some transcripts are methylated in a large proportion of cells, the majority of RNAs are rarely methylated.

Association of m^{6}A with RNA abundance in single cells

We next investigated the relationship between m^{6}A and RNA expression. We found a positive correlation between RNA

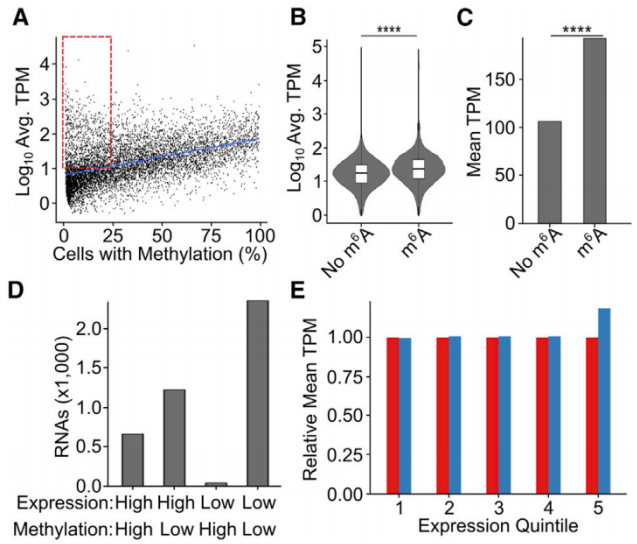


Figure 3. Relationship between RNA abundance and RNA methylation in single cells, see also Figure S4

(A) Relationship between RNA abundance and RNA methylation frequency at the population level. R = 0.5. Boxed region contains 1,231 RNAs that exhibit high expression in the population and rare methylation (< 25% of cells).

(B) Single-cell gene expression values of parent RNAs for each m^{6}A site in the SMART-seq2 scDART-seq dataset. Unmethylated sites: n = 5,947,374; methylated sites: n = 344,581. ****$p < 2.2 \times 10^{-16}$.

(C) Mean single-cell TPM values of RNAs in single cells containing unmethylated or methylated m^{6}A sites (unmethylated mean: 106.5, methylated mean: 193.0). Unmethylated sites: n = 5,947,374; methylated sites: n = 344,581. **** = $p < 2.2 \times 10^{-16}$.

(D) Number of RNAs grouped by expression and proportion of cells with methylation. High expression = TPM $\geq$ 10; low expression = TPM < 10; high methylation $\geq$ 75% of cells methylated; low methylation < 25% of cells methylated.

(E) Mean parent RNA TPM values for m^{6}A sites across all cells in which the site is methylated (blue) or unmethylated (red). Sites are divided into quintiles based on single-cell parent RNA abundance. n=1,258,391 sites for each quintile.

abundance at the population level and the percentage of cells in which the RNA is methylated (Figure 3A). However, we also identified a group of over 1,200 RNAs that are highly expressed but infrequently methylated, which includes genes involved in RNA regulation (box in Figure 3A; Table S4). We then examined the relationship between RNA methylation and RNA expression in single cells. For each m^{6}A site in the dataset, we compared its parent RNA expression in cells in which the site is methylated to those in which it is unmethylated. We found that parent RNA expression is 1.8-fold higher in cells in which a given site is methylated than in cells in which the site is unmethylated (Figures 3B–3D). To determine whether this association is consistent across a range of RNA expression levels, we divided the sites into quintiles based on parent RNA expression in each cell. Interestingly, only methylated sites in the highest expression quintile show increased parent RNA abundance compared with unmethylated sites (Figures 3E and S4; Table S4). However, for transcripts with low to moderate levels of expression, the presence of m^{6}A at any given site in an individual cell is not correlated with parent RNA expression in that cell (Figure 3E). Collectively, these data suggest that for a given m^{6}A site in a single cell, the likelihood of methylation is not associated with the expression level of the parent RNA within that cell. However, for sites occurring in a subset of highly expressed transcripts, there is a positive correlation between RNA abundance and the presence of m^{6}A. At the population level, this manifests as highly expressed RNAs exhibiting a greater likelihood of being methylated in more cells in the population (Figure 3A).

Most m^{6}A sites occur infrequently within individual cells of a population

Although the number of m^{6}A sites in a transcript is variable, most studies of m^{6}A in bulk cells estimate approximately 1 to 3 m^{6}A residues per mRNA (Dominissini et al., 2012; Linder et al., 2015; Meyer et al., 2012; Perry et al., 1975). Indeed, DART-seq profiling of bulk cells shows a median of 2 m^{6}A sites per mRNA, and only 2 mRNAs containing more than 20 sites (Figure 4A; Table S1). However, examination of scDART-seq data revealed a surprisingly high number of m^{6}A sites per transcript, with a median of 12 m^{6}A sites per mRNA and 237 mRNAs with at least 50 sites (Figure 4A; Table S4). To ensure that this discrepancy is not due to the detection of false-positive sites by scDART-seq, we used SELECT (Castellanos-Rubio et al., 2019) to validate a set of m^{6}A sites that were detected by scDART-seq but not by bulk antibody-based methods. Using this approach, we successfully confirmed the presence of m^{6}A at all tested sites, indicating that this effect is not due to false-positives (Figure S5A).

We therefore reasoned that the higher number of m^{6}A sites per mRNA detected in single cells could be due to the presence of rare m^{6}A sites that are found in subpopulations of cells that evade detection by bulk m^{6}A profiling. Indeed, we found that most m^{6}A sites identified in single cells are rare: 88% of sites occur in fewer than 20% of cells (Figure 4B). Additionally, while most mRNAs contain many total m^{6}A sites across all cells in the population, mRNAs are methylated at an average of only 1–3 sites in each cell, and individual sites are methylated in only 4.5% of cells on average (Figures 4C and 4D). Moreover, there is no correlation between the percentage of cells in which an individual m^{6}A site occurs and the total number of m^{6}A sites in the parent mRNA (Figure 4E). This suggests that most individual sites are rarely methylated, even those in mRNAs with many total m^{6}A sites. Indeed, mRNAs methylated in a high percentage of cells generally have more m^{6}A sites (Figure S5B), and only a small proportion of these sites are methylated per cell (Figures S5C–S5F). Collectively, these data reveal that mRNAs are susceptible to methylation at many more sites than has been suggested by m^{6}A profiling of bulk cells, and that most of these m^{6}A sites occur in a low proportion of cells within the population.

This pattern can be seen in the methylation landscape of individual mRNAs. For instance, the *PPP2CA* mRNA is methylated in 92% of cells in which it is expressed and contains a total of 33 m^{6}A sites. Most sites are methylated in fewer than 10% of cells, and the most frequently methylated site is present in only 39% of cells (Figure 4F). Furthermore, on average only 3 sites in *PPP2CA* are methylated per cell (Figure 4G). Notably, however, we found some transcripts with more distinct patterns of m^{6}A site frequency (Figures S5G and S5H). For instance, *STX4*

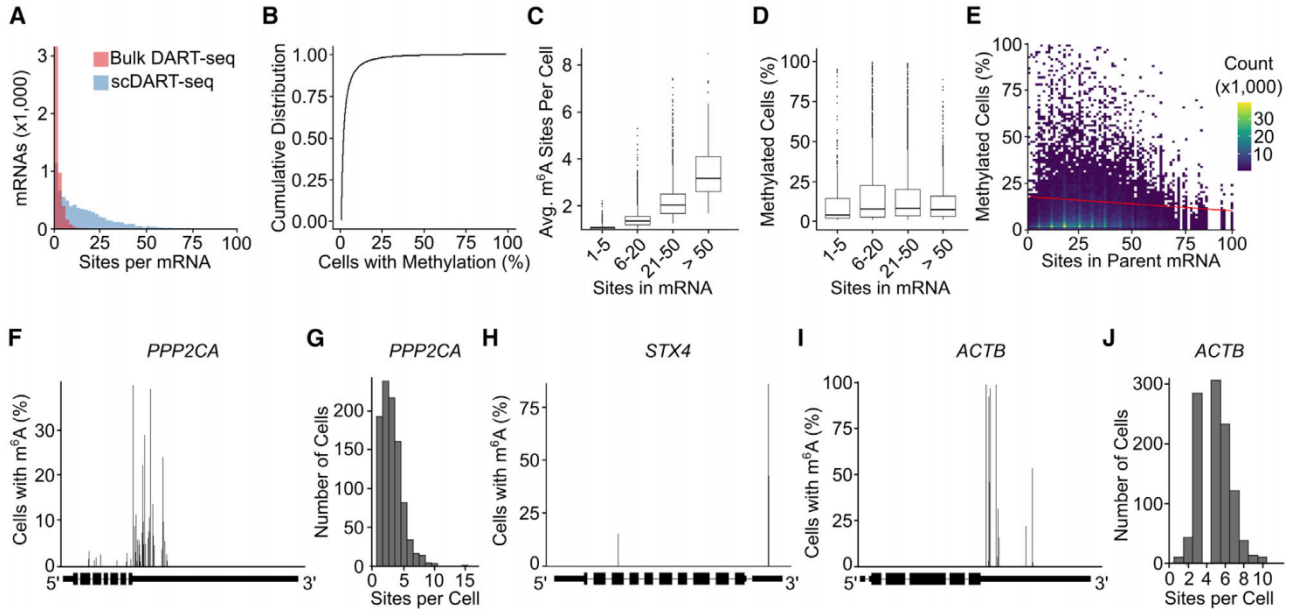


Figure 4. Distribution of m^{6}A sites in single cells, see also Figure S5

(A) Histogram showing the total number of m^{6}A sites identified in each methylated mRNA.

(B) Cumulative distribution plot showing the percentage of cells in which each methylated mRNA contains at least one site.

(C) Average number of m^{6}A sites per mRNA in single cells. The mRNAs are grouped by the total number of m^{6}A sites observed within the entire population. Only cells with detectable expression of the parent mRNA were considered.

(D) Percentage of cells in which each m^{6}A site is found, stratified by the total number of m^{6}A sites identified in the parent mRNA across all cells of the population. Only cells with detectable expression of the parent mRNA were considered.

(E) Relationship between the total number of m^{6}A sites identified within an mRNA and the percentage of cells in which each m^{6}A site occurs. Only mRNAs with 100 or fewer sites are shown for ease of visualization. n = 5,092,360 sites. $R = -0.07$. p = 0.23.

(F) Schematic of the *PPP2CA* mRNA with each vertical bar representing an individual m^{6}A site. Only cells in which *PPP2CA* is methylated were considered.

(G) Histogram showing the number of *PPP2CA* m^{6}A sites per cell.

(H) Schematic of the *STX4* mRNA with each vertical bar representing an individual m^{6}A site. Only cells in which *STX4* is methylated were considered.

(I) Schematic of the *ACTB* mRNA with each vertical bar representing an individual m^{6}A site. Only cells in which *ACTB* is methylated were considered.

(J) Histogram showing the number of *ACTB* m^{6}A sites per cell.

contains a total of just two m^{6}A sites, but only one of these sites is predominant and is found in 88% of the cells in which *STX4* is methylated (Figure 4H). Additionally, the *ACTB* mRNA is methylated in more than 99% of cells and contains only 11 total m^{6}A sites, although four of these sites are methylated in nearly all cells (Figures 4I and 4J). Thus, although the majority of m^{6}A sites in mRNAs are present in a low proportion of individual cells, some mRNAs contain one or more m^{6}A sites that are methylated in most cells of the population. Such "high-frequency" sites may serve important regulatory roles.

m^{6}A stoichiometry at individual sites is highly variable across single cells

We next sought to investigate m^{6}A stoichiometry in single cells, which is a critical determinant of the impact of m^{6}A on RNA function. Previous measurements have indicated that m^{6}A abundance can vary widely, from less than 5% to as high as 80% at individual sites (Garcia-Campos et al., 2019; Liu et al., 2013; Molinie et al., 2016). However, these studies used bulk cellular preparations, and the stoichiometry of m^{6}A within individual cells remains unknown. We previously demonstrated that %C2U values correlate with m^{6}A stoichiometry in bulk DART-seq data (Meyer, 2019). To confirm that %C2U values from scDART-seq data can also be used to estimate stoichiometry, we compared scDART-seq %C2U values with measurements from two independent methods for m^{6}A quantification—MAZTER-seq and SCARLET (Garcia-Campos et al., 2019; Liu et al., 2013). Both methods showed a positive correlation with %C2U values from scDART-seq data (Figures 5A and 5B). As further validation, we used a third independent method, SELECT (Castellanos-Rubio et al., 2019), to measure the stoichiometry of ten randomly selected m^{6}A sites from our scDART-seq dataset. Again, we observed a good correlation between %C2U and SELECT-based stoichiometry estimates (Figure 5C), confirming that C-to-U editing rates obtained with scDART-seq can be used to estimate m^{6}A stoichiometry.

We next examined m^{6}A abundance estimates from %C2U values transcriptome-wide in single cells. Surprisingly, we observed higher average %C2U values and a wider range of %C2U values for m^{6}A sites in single cells than in bulk cells (Figures 5D and S6A). These differences are not caused by variable levels of APOBEC1-YTH across individual cells, as there is no correlation between APOBEC1-YTH levels and average %C2U values in single cells (Figures S6B and S6C).

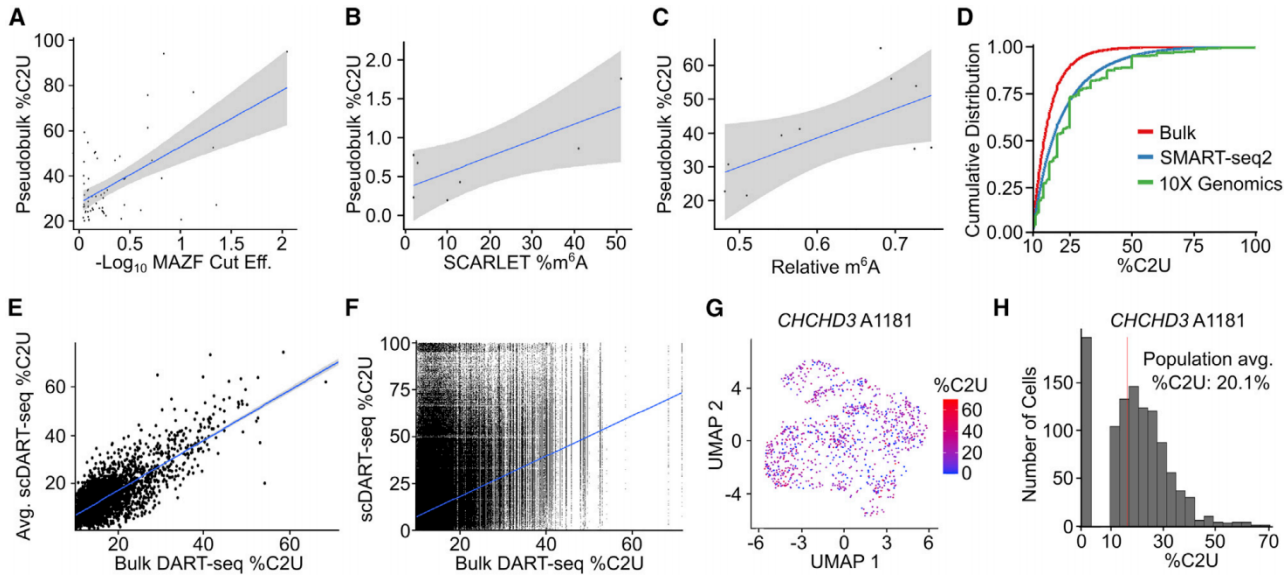


Figure 5. Heterogenous stoichiometry of m^{6}A in single cells, see also Figure S6

(A) Correlation between m^{6}A stoichiometry estimates using MAZTER-seq and scDART-seq (pseudobulk %C2U). $R = 0.57$; $p = 1.83 \times 10^{-6}$; n = 60 sites.

(B) Relationship between SCARLET and pseudobulk scDART-seq %C2U m^{6}A stoichiometry estimates for sites in the *MALAT1* RNA. $R = 0.79$; $p = 0.036$; n = 7 sites.

(C) Relationship between relative m^{6}A levels measured by SELECT and pseudobulk scDART-seq %C2U m^{6}A stoichiometry estimates at 10 random sites. n = 2. $R = 0.68$.

(D) Cumulative distribution plot showing the distribution of %C2U for all m^{6}A sites identified by bulk DART-seq and scDART-seq.

(E) %C2U values of m^{6}A sites observed in bulk DART-seq and average population-level %C2U values of the same sites identified with scDART-seq. $R = 0.78$; $p < 2.2 \times 10^{-16}$; n = 5,502 sites.

(F) Relationship between the %C2U values of m^{6}A sites identified with bulk DART-seq and the single-cell %C2U values identified with SMART-seq2 scDART-seq. $R = 0.32$; $p < 2.2 \times 10^{-16}$; n = 3,111,980 sites.

(G) Single-cell %C2U values for the m^{6}A site at A1181 in the *CHCHD3* 3′UTR. Only cells expressing *CHCHD3* were considered.

(H) Histogram showing the distribution of %C2U values for the m^{6}A site at A1181 in *CHCHD3*. Only cells expressing *CHCHD3* were considered.

We reasoned that the differences in m^{6}A stoichiometry estimates between single cells and bulk cells could reflect variability in m^{6}A abundance across individual cells. Since %C2U values in bulk DART-seq data include all cells in the population, cells with no methylation at a given site would reduce the overall %C2U value for that site. Indeed, when we compared the %C2U values of m^{6}A sites in bulk DART-seq with the population-averaged (pseudobulk) %C2U values in scDART-seq, we observed a strong correlation in m^{6}A abundance measurements. (Figures 5E, S6D, and S6E). However, single-cell %C2U values show substantial cell-to-cell variability (Figure 5F).

This single-cell variability is recapitulated in individual mRNAs. For instance, site A1181 in the *CHCHD3* mRNA has a pseudobulk %C2U value of 20.1%, similar to the 20.5% estimate of MAZTER-seq (Garcia-Campos et al., 2019). However, 18% of cells expressing *CHCHD3* have no methylation at A1181, whereas 56.7% of cells with A1181 methylation have a higher %C2U than the pseudobulk average (Figures 5G and 5H). A similar pattern is observed at A592 in the *ETFA* mRNA. This site has high methylation estimates (scDART-seq: 46.7%, MAZTER-seq: 55.9%), but most cells expressing *ETFA* are unmethylated at A592, and all cells with A592 methylation have higher %C2U values than the pseudobulk average (Figures S6F and S6G). These data suggest that m^{6}A stoichiometry is highly variable across individual cells, and that population-level estimates of m^{6}A stoichiometry fail to detect the extremes of m^{6}A abundance in single cells.

Differential methylation within subpopulations of cells

One advantage of scDART-seq is that it enables analysis of methylation patterns in different cellular states, such as distinct phases of the cell cycle. To determine whether mRNAs are differentially methylated throughout the cell cycle, we first identified cells in G1, S, and G2/M (Liu et al., 2017). Interestingly, we found no changes in global m^{6}A distribution or consensus site usage throughout the cell cycle (Figures 6A and 6B). However, we identified 122 mRNAs that are methylated in a significantly higher or lower proportion of cells within each cell cycle phase than would be expected based on the number of cells in which they are expressed (“CC-DMRNAs”; Table S5). The majority of CC-DMRNAs in G1 and G2/M phase cells are methylated more frequently than expected (53% and 83%, respectively), whereas most CC-DMRNAs in S phase cells (74%) are methylated less frequently (Figure 6C).

To investigate how the methylation of specific CC-DMRNAs changes during the cell cycle, we focused on two mRNAs, *TET1* and *TOPBP1*, which encode proteins that have been linked to cell cycle control through their roles in DNA regulation

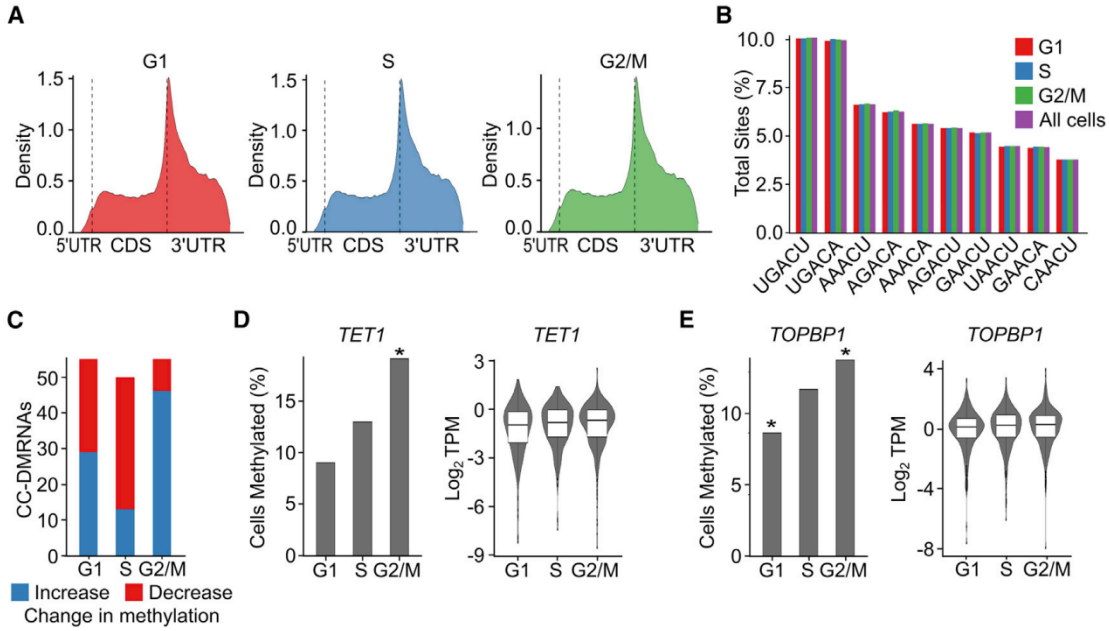


Figure 6. Differential RNA methylation throughout the cell cycle

(A) Metagene plots showing the distribution of m^{6}A sites identified in cells in the indicated cell cycle stage.

(B) The most common motifs surrounding m^{6}A sites identified in all cells, as well as in cells in different cell cycle stages. The y-axis indicates the percentage of all m^{6}A sites occurring in each motif.

(C) Number of differentially methylated RNAs (CC-DMRNAs) in cells within the indicated cell cycle phase.

(D) Percentage of cells in each cell cycle phase in which *TET1* is methylated. *$p < 0.05$ (left). Single-cell expression values for *TET1* in cells in each cell cycle phase (right).

(E) Percentage of cells in each cell cycle phase in which *TOPBP1* is methylated. *$p < 0.05$ (left). Single-cell expression values for *TOPBP1* in cells in each cell cycle phase (right).

(Bamezai et al., 2020; Wardlaw et al., 2014). Both *TET1* and *TOPBP1* are methylated in an increasing proportion of cells throughout cell cycle progression (Figures 6D and 6E). These differences are observed despite consistent expression of both mRNAs in each cell cycle phase (Figures 6D and 6E). Indeed, we found that CC-DMRNAs generally exhibit consistent expression levels throughout the cell cycle (Table S5). Thus, although global m^{6}A topology does not change across the cell cycle, we find that a subset of transcripts exhibit cell cycle-specific changes in methylation status independent of gene expression.

These findings prompted us to explore whether m^{6}A can be used more generally to distinguish subpopulations of cells. Identification of cellular subtypes is often done by using gene expression-based clustering to find heterogeneity in gene expression patterns. To determine whether mRNA methylation can be used to distinguish subpopulations of cells, we clustered cells using %C2U values identified by scDART-seq. This revealed two distinct clusters (me-cluster 1: 77.5% of cells; me-cluster 2: 22.5% of cells; Figures 7A and S7A). APOBEC1-YTH expression, cell cycle distribution, and global gene expression patterns are highly similar between the two me-clusters, indicating that these factors do not contribute to m^{6}A-based cell clustering (Figures 7B, 7C, and S7B–S7D). We searched for the sites that drive m^{6}A-based clustering and identified a total of 508 m^{6}A sites that are differentially methylated between the two me-clusters (Table S6). Most differentially methylated sites have lower stoichiometry estimates in me-cluster 2 relative to me-cluster 1, despite consistent expression of the parent RNAs (Figures 7D–7F and S7E; Table S6). When we examined the 393 mRNAs whose differentially methylated sites drive m^{6}A-based clustering, we found that many encode proteins involved in RNA regulation and RNA binding (Figures S7F and S7G), suggesting that the methylation of transcripts encoding RNA regulatory factors may contribute to cell subtype identity. Altogether, these data show that mRNA methylation can be used to distinguish subpopulations of cells and that the methylation status of a subset of mRNAs drives cell clustering independently of gene expression. It will be interesting in future studies to explore whether mRNA methylation can be used to distinguish distinct cell types that comprise complex cellular populations such as tumor samples or tissues.

DISCUSSION

In conclusion, we have developed a method for single-cell m^{6}A profiling and used it to identify the methylomes of 11,420 individual cells. Transcriptome-wide m^{6}A mapping studies have to date been hampered by high input RNA requirements and have therefore been limited to profiling m^{6}A in large populations of cells. scDART-seq circumvents these issues and, as reported here, uncovers fundamental features of RNA methylation that have been missed by population-level m^{6}A profiling studies. We find

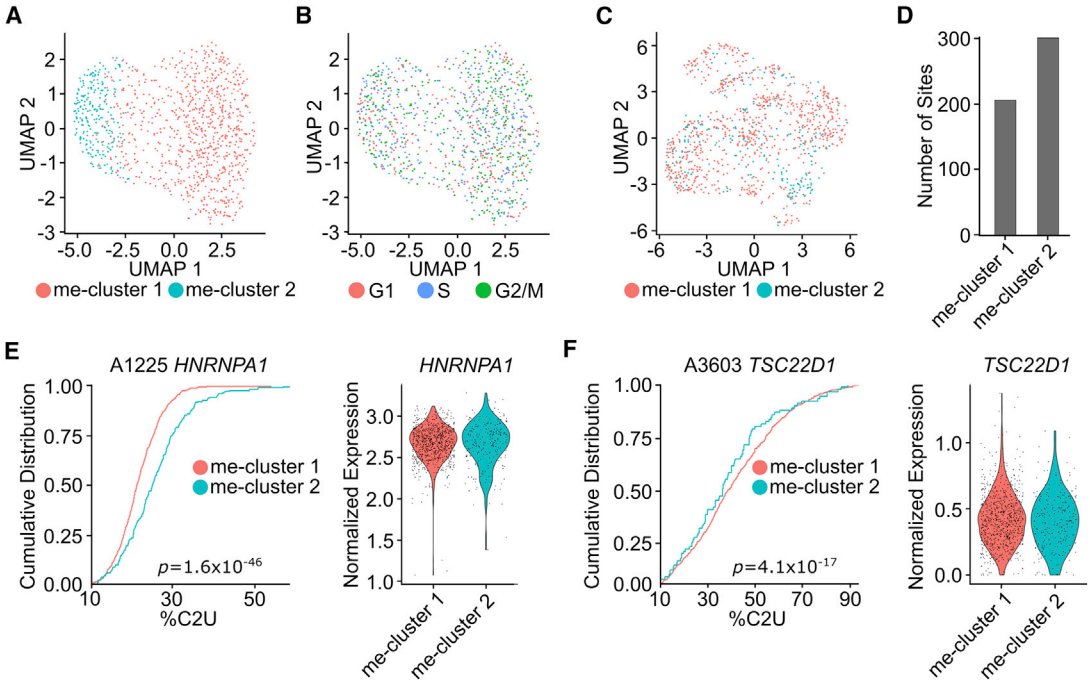


Figure 7. m^{6}A-based clustering to identify cellular subpopulations, see also Figure S7

(A) UMAP visualization of single cells clustered by m^{6}A.

(B) Distribution of cell cycle state in cells clustered by m^{6}A.

(C) me-cluster identity of single cells overlaid on the global gene expression UMAP plot.

(D) Differentially methylated m^{6}A sites with decreased %C2U in each me-cluster.

(E) Distribution of %C2U values for the m^{6}A site at A1225 in the *HNRNPA1* mRNA (left). Single-cell expression levels of the *HNRNPA1* mRNA within each me-cluster (right).

(F) Distribution of %C2U values for the m^{6}A site at A3603 in the *TSC22D1* mRNA (left). Single-cell expression levels of the *TSC22D1* mRNA within each me-cluster (right).

that single cells contain an average of 7,000 m^{6}A sites in 3,100 mRNAs, which is less than that predicted from bulk cellular m^{6}A profiling. At the same time, we find that individual mRNAs contain many more total m^{6}A sites than have been previously identified from population-level m^{6}A mapping. Our data show that the discrepancy between single-cell and population-level m^{6}A prevalence is largely driven by extreme heterogeneity in the frequency of individual m^{6}A sites in single cells. Although many mRNAs contain a surprisingly high number of total m^{6}A sites across the population, most of these sites occur in a low proportion of cells. This infrequent occurrence of individual m^{6}A sites contributes to the wide variability in transcript methylation frequency observed for many mRNAs across single cells. It further suggests that the density of m^{6}A sites within a transcript is likely to be an important factor for m^{6}A-mediated RNA regulation. Instead of a single site within an mRNA playing a dominant regulatory role, it may be that multiple, infrequently occurring sites in a particular region of an mRNA collectively contribute to transcript fate.

The fact that most mRNAs contain multiple m^{6}A sites that are methylated in a low proportion of cells suggests the possibility of cell-specific regulatory factors that direct the m^{6}A methyltransferase machinery to specific sites within an mRNA. Indeed, chromatin modifications, transcription factors, and transcription rate have been shown to influence m^{6}A methylation (Aguilo et al., 2015; Huang et al., 2019; Slobodin et al., 2020), so it will be interesting to apply scDART-seq in conjunction with the assessment of other gene regulatory signatures to determine whether cellular variation in site-specific m^{6}A modification is correlated with other factors. Notably, we also identified mRNAs containing "high-frequency" sites that are methylated in nearly all cells. The presence of such sites suggests that most cells have adapted the ability to methylate these sites and further underscores the importance of examining other factors at the single-cell level that contribute to their consistent methylation. These sites likely also serve important functional roles, and it will be important to investigate these and their impact on gene expression in future studies.

Our examination of single-cell m^{6}A stoichiometry reveals that population-level estimates of m^{6}A abundance at individual sites do not reflect the extreme variability that is observed among single cells. We found that many m^{6}A sites that show only modest stoichiometry estimates at the population level are in fact highly abundant within subpopulations of individual cells. To date, m^{6}A stoichiometry has been measured from bulk cellular samples, and it has been tempting to speculate that very low-stoichiometry m^{6}A sites may have a reduced functional impact compared with high-stoichiometry sites. Our studies challenge this notion

and suggest that m^{6}A sites with low-stoichiometry estimates at the population level may in fact be highly abundant within subpopulations of single cells and, therefore, have important functions in those cells.

Further supporting this idea, we find that we can cluster groups of cells based on their m^{6}A abundance patterns. This m^{6}A-based clustering was largely driven by the methylation of mRNAs encoding proteins involved in RNA binding and RNA processing, and it was independent of both gene expression and cell cycle phase. The ability to distinguish groups of cells by m^{6}A patterns alone reveals m^{6}A as a potential driving force for co-regulation of subpopulations of cells. It will be interesting to explore this phenomenon further in more complex biological systems, such as in the context of cancer development and cellular differentiation, as well as in tissues comprising heterogeneous cell types. Additionally, m^{6}A plays a critical role in stem cell maturation (Batista et al., 2014; Weng et al., 2018; Yoon et al., 2017) and has been implicated in numerous cancers (Jiang et al., 2021; Li et al., 2017; Liu et al., 2018; Zhang et al., 2017). Thus, developing scDART-seq tools in the form of cell lines and mouse models will enable m^{6}A mapping in unique contexts and improve our understanding of *in vivo* m^{6}A regulation in both normal and pathological states.

Limitations of the study

One limitation of scDART-seq in its current form is that it requires the expression of APOBEC1-YTH in cells or tissues of interest. *In vitro* DART-seq uses purified APOBEC1-YTH to globally profile m^{6}A from low-input samples (Meyer, 2019), so further development of this approach to make it compatible with single-cell sequencing may help overcome this limitation. Additionally, as with any m^{6}A mapping approach, there remains the possibility that scDART-seq can detect false-positive m^{6}A sites. Conversely, some sites may be missed due to low abundance or inaccessibility to the APOBEC1-YTH protein. Although we have taken several steps to minimize both outcomes, additional controls—such as the use of *METTL3* knockout cells or further optimization of the APOBEC1-YTH protein—may improve the accuracy of single-cell m^{6}A detection. Going forward, as more single-cell m^{6}A datasets are produced and compared with data from bulk cells, it will be critical to develop standards for determining the expected overlap between technologies and to eliminate potential false-positives. Lastly, although scDART-seq-based estimates of m^{6}A stoichiometry compare well with other independent methods for m^{6}A quantification, the accuracy of m^{6}A abundance estimates can likely be further improved with additional optimization.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Employed cell lines
  - Cell culture treatments
- METHOD DETAILS
  - Cloning and stable cell line creation
  - Western blots
  - Microscopy
  - DART-PCR
  - Cell cycle analysis using propidium iodide staining
  - Relative quantitation of m^{6}A using RT-qPCR
  - Bulk DART-seq
  - Single-cell DART-seq using 10X Genomics sample preparation
  - Single-cell DART-seq using SMART-seq2 sample preparation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Single-cell DART-seq (10X Genomics) gene expression analysis
  - SMART-seq2 scDART-seq data quality control, integration with 10x Genomics scDART-seq data, and gene expression analysis
  - Genome alignment and TPM calculation for SMART-seq2 scDART-seq data
  - reCAT for cell cycle analysis of SMART-seq2 scDART-seq data
  - Identification of m^{6}A sites from bulk DART-seq data
  - Identification of high-confidence m^{6}A sites in 10x Genomics scDART-seq data
  - Identification of high-confidence m^{6}A sites in SMART-seq scDART-seq data
  - m^{6}A metagene and motif enrichment analysis
  - Comparison of methylated RNAs with REPIC database
  - Comparison of RNA abundance and methylation likelihood
  - Comparison of RNA abundance and methylation likelihood among gene expression quintiles
  - Gene ontology
  - Distribution of individual m^{6}A sites
  - mRNA m^{6}A site distribution graphs
  - Comparison of %C2U values with MAZTER-seq and SCARLET
  - Analysis of %C2U in single cells
  - Differential methylation analysis across cell cycle phases
  - Clustering by m^{6}A using SMART-seq2 scDART-seq cells
  - Clustering by m^{6}A using 10x Genomics scDART-seq cells
  - Statistical analysis and plotting

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.12.038.

ACKNOWLEDGMENTS

We thank all members of the Meyer laboratory for helpful discussions. We are grateful to the Duke Center for Genomic and Computational Biology, the Duke Molecular Genomics Core, and the Duke Human Vaccine Institute Flow

Cytometry Core for providing the infrastructure and support for sequencing, 10X Genomics library prep, and sorting of single cells. This work was supported by the National Institutes of Health (R01MH118366, DP1DA046584, and RM1HG011563), the Rita Allen Foundation, and the Kinship Foundation Searle Scholars Program.

AUTHOR CONTRIBUTIONS

M.T. and K.D.M. designed experiments and interpreted data. M.T. performed the experiments and analyses. M.N.F. developed and tested Bullseye. M.T. and K.D.M. wrote the manuscript with input from M.N.F.

DECLARATION OF INTERESTS

K.D.M. has filed a patent application for the DART-seq technology through Duke University.

Received: September 7, 2021
Revised: December 22, 2021
Accepted: December 23, 2021
Published: January 25, 2022

REFERENCES

Aguilo, F., Zhang, F., Sancho, A., Fidalgo, M., Cecilia, S. Di, Vashisht, A., Lee, D.-F., Chen, C.-H., Rengasamy, M., Andino, B., et al. (2015). Coordination of m6A mRNA methylation and gene transcription by ZFP217 regulates pluripotency and reprogramming. Cell Stem Cell *17*, 689–704.

Bamezai, S., Demir, D., Pulikkottil, A.J., Ciccarone, F., Fischbein, E., Sinha, A., Borga, C., te Kronnie, G., Meyer, L.-H., Mohr, F., et al. (2020). TET1 promotes growth of T-cell acute lymphoblastic leukemia and can be antagonized via PARP inhibition. Leukemia *35*, 389–403.

Barger, C.J., Branick, C., Chee, L., and Karpf, A.R. (2019). Pan-cancer analyses reveal genomic features of FOXM1 overexpression in cancer. Cancers *11*, 251.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707–719.

Castellanos-Rubio, A., Santin, I., Olazagoitia-Garmendia, A., Romero-Garmendia, I., Jauregi-Miguel, A., Legarda, M., and Bilbao, J.R. (2019). A novel RT-QPCR-based assay for the relative quantification of residue specific m6A RNA methylation. Sci. Rep. *9*, 4220.

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

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. PLoS Comput. Biol. *3*, e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Garcia-Campos, M.A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A., et al. (2019). Deciphering the "m6A code" via antibody-independent quantitative profiling. Cell *178*, 731–747.e16.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors Prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

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

Jiang, X., Liu, B., Nie, Z., Duan, L., Xiong, Q., Jin, Z., Yang, C., and Chen, Y. (2021). The role of m6A modification in the biological functions and diseases. Signal Transduct. Target. Ther. *6*, 74.

Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: a method to quantify base editing from sanger sequencing. CRISPR J. *1*, 239–250.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., Huang, H., Nachtergaele, S., Dong, L., Hu, C., et al. (2017). FTO plays an oncogenic role in acute myeloid leukemia as a N6-methyladenosine RNA demethylase. Cancer Cell *31*, 127–141.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lichinchi, G., Zhao, B.S., Wu, Y., Lu, Z., Qin, Y., He, C., and Rana, T.M. (2016). Dynamics of human and viral RNA methylation during Zika virus infection. Cell Host Microbe *20*, 666–673.

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

Liu, J., Eckert, M.A., Harada, B.T., Liu, S.-M., Lu, Z., Yu, K., Tienda, S.M., Chryplewicz, A., Zhu, A.C., Yang, Y., et al. (2018). m6A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. Nat. Cell Biol. *20*, 1074–1083.

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA *19*, 1848–1856.

Liu, S., Zhu, A., He, C., and Chen, M. (2020). REPIC: a database for exploring the N6-methyladenosine methylome. Genome Biol. *21*, 100.

Liu, Z., Lou, H., Xie, K., Wang, H., Chen, N., Aparicio, O.M., Zhang, M.Q., Jiang, R., and Chen, T. (2017). Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat. Commun. *8*, 22.

Meyer, K.D. (2019). DART-seq: an antibody-free method for global m6A detection. Nat. Methods *16*, 1275–1280.

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

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P., et al. (2016). M6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. Nat. Methods *13*, 692–698.

Olarerin-George, A.O., and Jaffrey, S.R. (2017). MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. Bioinformatics *33*, 1563–1564.

Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5′ terminus. Cell *4*, 387–394.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rahman, R., Xu, W., Jin, H., and Rosbash, M. (2018). Identification of RNA-binding protein targets with HyperTRIBE. Nat. Protoc. *13*, 1829–1849.

Roehr, J.T., Dieterich, C., and Reinert, K. (2017). Flexbar 3.0 – SIMD and multicore parallelization. Bioinformatics *33*, 2941–2942.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. Cell Rep. *8*, 284–296.

Shi, H., Wei, J., and He, C. (2019). Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. Mol. Cell *74*, 640–650.

Slobodin, B., Bahat, A., Sehrawat, U., Becker-Herman, S., Zuckerman, B., Weiss, A.N., Han, R., Elkon, R., Agami, R., Ulitsky, I., et al. (2020). Transcription dynamics regulate poly (A) tails and expression of the RNA degradation machinery to balance mRNA levels. Mol. Cell *78*, 434–444.e5.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Wardlaw, C.P., Carr, A.M., and Oliver, A.W. (2014). TopBP1: a BRCT-scaffold protein functioning in multiple cellular pathways. DNA Repair (Amst) *22*, 165–174.

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B.S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., et al. (2018). METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m6A modification. Cell Stem Cell *22*, 191–205.e9.

Yoon, K.J., Ringeling, F.R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., Su, Y., Kim, N.S., Zhu, Y., Zheng, L., et al. (2017). Temporal control of mammalian cortical neurogenesis by m6A methylation. Cell *171*, 877–889.e17.

Zaccara, S., Ries, R.J., and Jaffrey, S.R. (2019). Reading, writing and erasing mRNA methylation. Nat. Rev. Mol. Cell Biol. *20*, 608–624.

Zhang, S., Zhao, B.S., Zhou, A., Lin, K., Zheng, S., Lu, Z., Chen, Y., Sulman, E.P., Xie, K., Bögler, O., et al. (2017). m6A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. Cancer Cell *31*, 591–606.e6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-βactin (mouse)	GenScript	A00702; RRID: AB_914102
Anti-HA (rabbit)	Cell Signaling Technologies	#3724; RRID: AB_1549585
Anti-rabbit-HRP	Abcam	ab6721; RRID: AB_955447
Anti-mouse-HRP	Fisher Scientific	62-6520; RRID: AB_2533947
Bacterial and virus strains		
APOBEC1-YTH lentivirus	This paper	N/A
APOBEC1-YTH^{mut} lentivirus	This paper	N/A
Chemicals, peptides, and recombinant proteins		
Doxycycline	Sigma-Aldrich	D3447
jetPEI	Polyplus	101-10N
TrypLE	Gibco	12604-013
DNase I	New England Biolabs	M0303L
Bst Polymerase	New England Biolabs	M0275S
Penicillin/Streptomycin	Gibco	15140-122
Critical commercial assays		
Qubit 1x dsDNA HS Assay Kit	Invitrogen	Q33230
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	New England Biolabs	E7760
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490
10x Genomics Next GEM Single Cell 3' Kit (v3)	10x Genomics	1000158
SMART-seq HT Kit	Takara	634438
Nextera XT DNA Library Prep Kit	Illumina	FC-131-1096
Nextera-Compatible Indexing Primers (CDI)	Integrated DNA Technologies	Custom
Ampure XP Beads	Beckman-Coulter	A63880
Deposited data		
Raw and processed data	This paper	GEO: GSE180954
Bullseye code	This paper	https://github.com/mflamand/Bullseye
Original Images (Mendeley Data)	This paper	https://doi.org/10.17632/99cjyffbyc.1
Experimental models: Cell lines		
HEK293T (Human embryonic kidney, female)	ATCC	ATCC CRL-3216
iDART-APOBEC1-YTH (Human embryonic kidney, female)	This paper	N/A
iDART-APOBEC1-YTH^{mut} (Human embryonic kidney, female)	This paper	N/A
Oligonucleotides		
ActinFwd (Sanger sequencing)	This paper	caacacagtgctgtctggc
ActinRev (Sanger sequencing)	This paper	caagatgagattggcatggc
ACTBqPCRFwd	Meyer, 2019	cagcaagcaggagtatgacgagtc
ACTB_nonAdjRev	Meyer, 2019	catgccaatctcatcttg
ActinA1222_AdjRev	This paper	ttgtcaagaaagggtgtaacgcaactaag
GPIqPCRFwd	This paper	gcagtggtgtgatctcggctcactg

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GPI_nonAdjRev	This paper	accatcctggctaacacggtgaaac
GPIA3874_AdjRev	This paper	gtgtggtggcgggcacctgtag
ETFAqPCRFwd	This paper	ggttgccccgatttctgacatc
ETFA_nonAdjRev	This paper	actggtgaagtacttgatgccttttc
ETFAA592_AdjRev	This paper	ggttcaccaaatgatggggcttg
G3BP1qPCRFwd	This paper	ctcactggtatgtttaaaaactgaataaaagg
G3BP1_nonAdjRev	This paper	acaacagttttgctctggttgttatc
G3BP1A2191_AdjRev	This paper	taaacatgaaaaaactacttcaaaaaaactaacag
EID1qPCRFwd	This paper	gtagttagatgctgttaaaagaggagg
EID1_nonAdjRev	This paper	tttctttaatgattctcttccttttcacaac
EID1A709_AdjRev	This paper	gtgttctattctgaaaaacaaatttttttgaagg
RANqPCRFwd	This paper	ctcattattatctagctaagcggaacatg
RAN_nonAdjRev	This paper	gaaaaggaatgggaatgacagtaacaaac
RANA939_AdjRev	This paper	tatgaaactcaaggaatcaactgcg
PWP1qPCRFwd	This paper	gatcatctaatttcctgcttaccttaactggg
PWP1_nonAdjRev	This paper	gcaacaagagccttacaaaccatccatc
PWP1A1726_AdjRev	This paper	ccaccgcagttgcagaaaggaatg
EXOSC2qPCRFwd	This paper	agaggcttttggaacaggaggg
EXOSC2_NonAdjRev	This paper	aaagggatgatctctcgagtctttgc
EXOSC2A1019_AdjRev	This paper	ggcctgtgggctggaggccg
DPM2qPCRFwd	This paper	ttcttgctcagggcctaaaagatg
DPM2_NonAdjRev	This paper	tgtcagggtaaggaaagctgcc
DPM2A771_AdjRev	This paper	agggctggtcctggggtctg
SRP9qPCRFwd	This paper	attggaattaactccacaccatagtatgc
SRP9_NonAdjRev	This paper	cacagtgatcatcattccattcatacactc
SRP9A1132_AdjRev	This paper	ctgcttctgatcactgaaaactagattcaaag
ACTN1qCPRFwd	This paper	tggatgccgaagacatcgttgg
ACTN1_NonAdjRev	This paper	ccttctgggctccagagaagg
ACTN1A1061_AdjRev	This paper	ctcgtagtcttccataagctgctcg
NDC1qPCRFwd	This paper	tttcagtaagcacccagaggcc
NDC1_NonAdjRev	This paper	gatgctgctactaagtgcgacag
NDC1A1791_AdjRev	This paper	caaagtattaaggatagctggtagtgtcg
TMEM167AqPCRFwd	This paper	ggattgttgggtatattttggaagtgtgc
TMEM167A_NonAdjRev	This paper	cattttccccagctactgtatgaagagg
TMEM167AA346_AdjRev	This paper	gagttaaacatcctttccattattttctgcag
TMEM192qPCRFwd	This paper	cctcggagactggattcagaactatttc
TMEM192_NonAdjRev	This paper	ggttcctctgcaattgctgtcatg
TMEM192A957_AdjRev	This paper	ggcagcttgtcaggtggtcag
ZDHHC16qPCRFwd	This paper	ctccaagggcagcttttctcagaatc
ZDHHC16_NonAdjRev	This paper	gcctgagtaaagggatgtcagtcc
ZDHHC16A1737_AdjRev	This paper	tagctagctgcctgacccctag
Recombinant DNA		
TLCV2	Barger et al., 2019	Addgene: 131637
TLCV2-APOBEC1-YTH	This paper	Addgene: 178949
TLCV2-APOBEC1-YTH^{mut}	This paper	Addgene: 178950
pCMV-APOBEC1-YTH	Meyer, 2019	Addgene: 131636
pCMV-APOBEC1-YTH^{mut}	Meyer, 2019	Addgene: 131637
pMD2.G	Unpublished	Addgene: 12259

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
psPAX2	Unpublished	Addgene: 12260
Software and algorithms		
Bullseye	This paper	https://github.com/mflamand/Bullseye
Cell Ranger 3.1.0	10x Genomics	https://www.10xgenomics.com/ RRID:SCR_017344
Seurat 3.2.1	Stuart et al., 2019	https://satijalab.org/seurat/get_started.html RRID:SCR_016341
R/4.0.0	R Development Core Team	http://www.r-project.org/
Flexbar 3.0.3	Roehr et al., 2017	http://sourceforge.net/projects/flexbar/ RRID:SCR_013001
STAR 2.7.5c	Dobin et al., 2013	https://github.com/alexdobin/STAR/releases RRID:SCR_004463
Subread 1.6.3	Liao et al., 2014	http://subread.sourceforge.net/ RRID:SCR_009803
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/ RRID:SCR_010881
MetaPlotR	Olarerin-George and Jaffrey, 2017	https://github.com/olarerin/metaPlotR
BEDTools	Quinlan and Hall, 2010	https://github.com/arq5x/bedtools2 RRID:SCR_006646
SAMtools	Li et al., 2009	http://htslib.org/ RRID:SCR_002105
RSeQC	Wang et al., 2012	https://github.com/MonashBioinformaticsPlatform/RSeQC RRID:SCR_005275
Wiggleplotr	Unpublished	https://bioconductor.org/packages/release/bioc/html/wiggleplotr.html
FlowJo	FlowJo	https://www.flowjo.com/

RESOURCE AVAILABILITY

Materials availability

Plasmids generated in this study have been deposited to Addgene.

Data and code availability

- All raw fastq sequencing files and processed bed files have been deposited at GEO and are publicly available on the date of publication. Accession numbers are listed in the key resources table.
- All code for Bullseye is publicly available on Github (https://github.com/mflamand/Bullseye) on the date of publication.
- Other requests for data, scripts, or analyses are available upon reasonable request from the lead contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Employed cell lines

Human HEK293T cells (ATCC, Manassas, VA) were cultured in Dulbecco's Modified Eagle's Medium (Corning, Corning, NY) with 10% fetal bovine serum (Avantor, Radnor, PA) and 10 units/mL Penicillin, and 10 μg/mL Streptomycin (Gibco, Waltham, MA). Cells were cultured at 37°C with 5% CO_2. Cell lines have not been authenticated or tested for mycoplasma.

Cell culture treatments

All HEK293T transgenic cell lines were plated 2 days before the experiment. The following day, the cells were treated with 1 μg/mL doxycycline (Sigma-Aldrich, St. Louis, MO). The cells were harvested or prepared as described in Method Details for each experiment.

METHOD DETAILS

Cloning and stable cell line creation

The lentiviral vector TLCV2 (gift from Adam Karpf, Addgene #87360) was used as the backbone for the lentiviral expression construct used to generate TLCV2-APOBEC1-YTH and TLCV2-APOBEC1-YTH^{mut} plasmids, which were used to create stable

cell lines. To clone TLCV2-APOBEC1-YTH and TLCV2-APOBEC1-YTH^{mut}, the CAS9-FLAG sequence in TLCV2 was replaced with the APOBEC1-YTH (or APOBEC1-YTH^{mut} for control cell lines) cassette from pCMV-APOBEC1-YTH, Addgene #131636 (or pCMV-APOBEC1-YTH^{mut}, Addgene #131637) by Gibson Assembly. To make lentivirus, HEK293T cells at 80% confluency on a 15 cm plate were transfected with 26.75 μg of TLCV2-APOBEC1-YTH or TLCV2-APOBEC1-YTH^{mut}, 20 μg of psPAX2 (Gift from Didier Trono, Addgene plasmid #12260), and 6.25 μg of pMD2.G (Gift of Didier Trono, Addgene plasmid #12259) using 3 μg jet-PEI (Polyplus, Illkirch, France). The medium was changed to fresh DMEM/10% FBS after 8 hours. 72 hours after transfection, the virus was purified from the medium. To purify the virus, the medium was centrifuged at 5,000 x g for 10 minutes. The supernatant was then filtered through a 0.45 μm filter into an ultracentrifuge tube (Beckman-Coulter, Pasadena, CA). 4 mL of sterile 20% sucrose was added below the medium using a 5 mL serological pipette. The mixture was centrifuged at 19,700 rpm for 2 h at 4°C in a swinging bucket rotor (Beckman-Coulter). Then the supernatant was removed and 100 μL of PBS was added, and the virus was resuspended overnight at 4°C with rocking. The resuspended virus was stored at -80°C. To create stable cell lines, the virus was added to HEK293T cells at a 1:100 dilution. After 24 h, the medium was replaced with medium containing 2 μg/mL puromycin (Sigma-Aldrich). The medium was replaced daily to remove dead cells for 7 days. To select clonal cell lines, the transgenic cells were plated in 96-well plates at a dilution of 0.5 cells per well and only wells containing a single cell were used for expansion. The cell lines express a dox-inducible APOBEC1-YTH transgene with a T2A followed by EGFP, so induced cells have green fluorescence, but EGFP is not fused to the APOBEC1-YTH protein.

Western blots

Cells were quickly rinsed with cold 1x PBS and scraped from culture plates. Cells were then pelleted by centrifugation at 1,000 x g for 3 minutes at 4°C. Cell pellets were resuspended in lysis buffer (25 mM Tris-HCl, pH7.4; NaCl 150 mM; Triton X-100 1% (v/v); sodium dodecyl sulfate 0.1% (v/v); cOmplete proteinase inhibitor cocktail (Sigma-Aldrich)) and incubated on ice for 10 minutes. Lysates were then centrifuged at 13,000 x g for 15 minutes at 4°C. The supernatant was transferred to a new tube. Samples for SDS-PAGE were then prepared at a final concentration of 1μg/μL total protein in 1x NuPAGE LDS Sample Buffer (Invitrogen, Waltham, MA) and 0.1 M DTT (VWR, 97061-338). Samples were run on 4-12% SDS-PAGE gels (Invitrogen) and transferred for 60 minutes at 100V in Towbin transfer buffer (25 mM Tris Base, 192 mM Glycine, 20% methanol (v/v)) to a nitrocellulose membrane (GE Amersham, Amersham, UK). After transferring, the membrane was blocked in PBST (PBS with 0.1% Tween 20 (Sigma-Aldrich)) with 5% milk (w/v) (Quality Biological, Gaithersburg, MD) for 1 h at room temperature. Primary antibodies (anti-βactin (Genscript, Piscataway, NJ), anti-HA (Cell Signaling Technology, Danvers, MA)) were incubated with the blots overnight at 4°C. The membrane was washed 3 times with PBST before the secondary antibody was added for 1 h at room temperature in PBST. Anti-rabbit-HRP secondary (Fisher Scientific, Waltham, MA) was used at 1:10,000 dilution, while anti-mouse-HRP secondary (Fisher Scientific) was used at 1:2,500. The membrane was then washed 3 times with PBST for 5 minutes. The western blot was visualized using Amersham ECL Prime Reagent (Amersham) and imaged on a Chemidoc MP (BioRad, Hercules, CA).

Microscopy

Fluorescent images were captured using a Leica DMi8 inverted microscope (Excitation 470/40, Emission 525/50) using the LAS X software.

DART-PCR

Transgenic cells were treated with 1 μg/μL doxycycline for 0 h, 2 h, 4 h, 8 h, or 24 h. Total RNA was extracted using Trizol (Invitrogen, Paisley, UK) according to the manufacturer's instructions. Total RNA was treated with DNase I (NEB) for 15 min at 37°C to remove possible DNA contamination. RNA was isolated using ethanol precipitation. cDNA was made using iScript Reverse Transcription Supermix (Bio-Rad, Hercules, CA). PCR amplification of the region surrounding A1222 on the *ACTB* mRNA was carried out with CloneAmp HiFi PCR Mastermix (Takara, Mountain View, CA) with the following oligos: ActinFwd 5′-caacacagtgctgtctggc-3′; ActinRev 5′- caagatgagattggcatggc-3′. The resulting PCR product was gel-purified on a 1% agarose gel and gel extracted using the Qiaquick Gel Extraction Kit (Qiagen, Hilden, Germany). Samples were submitted for sanger sequencing (Genewiz, South Plainfield, NJ) and % C2U was quantified using EditR software (Kluesner et al., 2018).

Cell cycle analysis using propidium iodide staining

When APOBEC1-YTH- and APOBEC1-YTH^{mut}-expressing cells were approximately 50% confluent, they were treated with 1μg/mL doxycycline and cultured for 24 h. After 24 h, cells were removed from the plate by incubating with 1x TrypLE (Invitrogen) for 5 minutes at 37 °C. The cells were resuspended to 10 mL using 1x PBS with 1% FBS and counted using a hemacytometer. 1 million cells were then resuspended in 1 mL 1x PBS. While vortexing, 4 mL of ice-cold 100% was added dropwise to the cells. The cells were fixed for a minimum of 2 h at 4 °C. After ethanol fixation, cells were resuspended in 500 μL of labeling solution (1x PBS containing 0.05% TritonX-100, 0.2 mg/mL propidium iodide (Sigma-Aldrich), and 0.5 mg/mL RNase A (Promega)). The samples were incubated at 37 °C for 30 minutes, then filtered through a 40 μm filter (BD, Franklin Lakes, NJ). Samples were analyzed using a BD Canto flow analysis cytometer and the cell cycle distribution was determined using FlowJo software.

Relative quantitation of m^{6}A using RT-qPCR

Total RNA was extracted from HEK293T cells using Trizol (Invitrogen) according to the manufacturer's instructions, then treated with DNase I (NEB) for 15 minutes at 37°C to remove possible DNA contamination. RNA was isolated using ethanol precipitation. Two reverse transcription reactions using BstI polymerase (NEB) were assembled, one with a primer adjacent to a potential m^{6}A site (+), and one using a non-adjacent primer (-). Final reaction buffer consisted of 1x ThermoPol Buffer (NEB), 50 μM dNTPs, 500nM primer, 150 ng total RNA, and 10U BstI Polymerase. Samples were incubated in a thermocycler for the following protocol: 3 minutes 25 °C, X minutes 50 °C, and 3 minutes 85 °C. The extension times used for the following sites were as follows: *EID1* A709, *RAN* A939 and *G3BP1* A2191 – 10 minutes; *ACTB* A1222, *GPI* A3874, *ETFA* A592, *EXOSC2* A1019, *PWP1* A1726, *SRP9* A1132, *NDC1* A1791, *ACTN1* A1061, *TMEM167A* A346, *TMEM192* A957 – 15 minutes; *DPM2* A771, *ZDHHC16* A1737 – 40 minutes. Additionally, two reverse transcription reactions were performed using Super Script III (Invitrogen), one with the m^{6}A adjacent primer (+), and one using only the non-adjacent primer. Quantitative PCR was performed for all samples using the resulting cDNA, and relative m^{6}A levels were calculated from the Ct values obtained, using the following formula described in (Castellanos-Rubio et al., 2019): Rel. m^{6}A = $2^{-(Ct\ Bst(-) - Ct\ SSIII(-)/Ct\ Bst(+) - Ct\ SSIII(+))}$.

Bulk DART-seq

Three independent plating and RNA isolation experiments were performed for both APOBEC1-YTH and APOBEC1-YTH^{mut} cells. 24 h after doxycycline treatment, cells were briefly rinsed with cold 1X PBS and removed from the culture plate using a cell scraper (Corning). Cells were pelleted and supernatants were removed and stored at -80°C. Total RNA was isolated using Trizol (Invitrogen) according to the manufacturer's instructions. Total RNA was treated with DNase I (NEB) for 15 minutes at 37°C to remove possible DNA contamination. RNA was isolated using ethanol precipitation. Sequencing libraries were generated from 900 ng of total RNA for each replicate using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) with the Poly(A) Selection Module (NEB) according to the manufacturer's instructions. Sample QC was performed using a Bioanalyzer High Sensitivity DNA 1000 Chip (Agilent, Santa Clara, CA) and quantified using the Qubit 1X dsDNA HS Kit (Invitrogen). 50bp paired-end sequencing was performed on a NovaSeq 6000 using the S-Prime flow cell.

Single-cell DART-seq using 10X Genomics sample preparation

Samples were prepared according to the protocol from the 10x Genomics Cell Preparation Guide (Manual Part #CG00053). Briefly, three separate plates of APOBEC1-YTH and three separate plates of APOBEC1-YTH^{mut} cells were processed distinctly and were used as biological triplicates. 24 h after doxycycline treatment, cells were rinsed with 1X PBS and treated with 2 mL 1X TrpLE Express (Invitrogen, 12604039) for 5 minutes at 37°C. 6 mL of DMEM with 10% FBS was added, and the cells were gently disrupted with a wide bore p1000 pipette tip 10 times. The cells were counted using a hemacytometer and 2.5 million cells were transferred to a new tube and washed in 1 mL PBS with 0.04% BSA. 1 million cells were then resuspended in 1 mL DMEM with 10% FBS. Single-cell libraries were constructed as described for the 10x Genomics Next GEM Single Cell 3' Kit, v3.1 chemistry (10x Genomics, Pleasanton, CA). 50bp-paired end sequencing was performed on a NovaSeq 6000 using an S1 flow cell, obtaining approximately 80,000 reads per cell.

Single-cell DART-seq using SMART-seq2 sample preparation and sequencing

24 h after doxycycline treatment, APOBEC1-YTH and APOBEC1-YTH^{mut} cells were rinsed with 1X PBS and treated with 2 mL 1X TrpLE Express (Invitrogen) for 5 mins. at 37°C. Then 12 mL of flow buffer (DMEM with 1% FBS) was added. Cells were gently agitated and counted using a hemacytometer. 2 million cells were transferred to a sterile 1.5 mL tube and pelleted at 2,000 x g for 3 minutes. The supernatant was removed, and the cells were resuspended in 200 μL of flow buffer. 2 μL of near-IR LIVE/DEAD viability marker (Invitrogen) was added and the cells were incubated at room temperature for 5 minutes. Then 4 units of DNase I was added and incubated for 15 minutes at 37°C. The cells were pelleted at 2,000 x g for 3 minutes and resuspended in 2 mL of flow buffer. They were then filtered through a 40 μM filter. Single, viable, GFP+ cells were then sorted into individual wells of 384-well PCR plates containing 1 μL of the prepared lysis buffer supplied with the SMART-seq HT Kit (Takara) using a FACS Aria II (BD). Immediately after sorting, each plate was centrifuged at 1,000 x g for 1 min before flash-freezing on dry ice. Plates were stored at -80°C until library preparation. Manufacturer's instructions for the SMART-seq HT Kit were followed for reverse transcription/cDNA synthesis, but reactions were scaled down 12.5-fold. Briefly, each plate was thawed on ice and 1 μL of the One-Step mastermix was added to each well and incubated in a thermocycler (42°C for 90 minutes, 95°C for 1 minute, followed by 18 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 68°C for 3 minutes. A final extension at 72°C for 10 minutes was used followed by a hold at 4°C). Ampure XP beads (Beckman-Coulter) were used to purify the amplified cDNA, which was eluted into 17 μL of 10 mM Tris (pH 8.0). All cDNA samples were quantified using the Qubit 1X dsDNA HS Kit (Invitrogen). Samples above 200 pg/μL were diluted to 200 pg/μL, and samples with undetectable cDNA were discarded. The Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA) was used to prepare indexed libraries from the amplified cDNA. Manufacturer's instructions were used, but reactions were scaled down 8-fold. Briefly, 1.88 μL of tagmentation premix and 0.63 μL of cDNA was added to each well of a 384-well PCR plate and incubated in a thermocycler for 10 minutes at 55°C. Then 0.63 μL NT Buffer was added and the plate was vortexed and incubated at room temperature for 5 minutes. Then 1.88 μL of NPM buffer, 0.63 μL water, and 0.63 μL of custom duplexed CDI primers (IDT, 10 mM) were added

and library amplification was performed in a thermocycler (72°C for 3 minutes, 95°C for 30 seconds, followed by 13 cycles of 95°C for 10 seconds, 55°C for 30 seconds, and 72°C for 1 minute. A final extension at 72°C for 5 minutes was used followed by a hold at 4°C). After amplification, Ampure XP beads were used to purify the final libraries, which were eluted into 17 μL of 10 mM Tris, pH 8.0. Each library was quantified using the Qubit 1X dsDNA HS Kit (Invitrogen), and all samples above 200 pg/μL were pooled. Pooled samples were run on a Bioanalyzer High Sensitivity DNA 1000 Chip for QC (Agilent). Average fragment size was approximately 700 bp. 150bp paired-end sequencing was performed on a NovaSeq 6000 using an S4 flow cell to generate approximately 5 million reads per cell.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-cell DART-seq (10X Genomics) gene expression analysis

Raw reads were demultiplexed using Illumina bcl2fastq2 software. Fastq files from APOBEC1-YTH, APOBEC1-YTH^{mut} and wild-type HEK293T cells (10x Genomics) were aligned to the hg38 reference genome using Cell Ranger count. APOBEC1-YTH transgene expression was assessed by aligning to the rat APOBEC1 sequence. Filtered barcode matrices were loaded into Seurat (3.2.1). Cells were eliminated from analysis if they had high mitochondrial content (>20% of reads), low read or gene detection numbers (< 1,000 UMI reads or < 500 genes), or low transcriptional complexity (< 0.8 $\log_{10}$ genes/UMI). Each replicate was then log-normalized and scored for cell cycle using Seurat in the following way. Separately, each replicate was normalized using the SCTransform function, regressing out cell cycle genes to eliminate the effect of cell cycle on clustering. 3,000 features were used to find integration anchors, and the IntegrateData command was used to integrate all 6 replicates with wild-type HEK293T cells. PCA was performed on variable features, followed by the RunUMAP command. Clusters were identified by using the FindNeighbors command on dimensions 1:40, followed by FindClusters at different resolutions. The resolution used in this manuscript (0.05, followed by MergeClusters 0 and 1) was used as it was the highest resolution where all clusters had at least one significantly differentially expressed marker gene as determined by the FindConservedMarkers command.

SMART-seq2 scDART-seq data quality control, integration with 10x Genomics scDART-seq data, and gene expression analysis

Raw files were demultiplexed and converted to fastq format using bcl2fastq2 software from Illumina. Adapter sequences were removed using Flexbar (3.0.3) (Roehr et al., 2017). For gene expression analysis, reads were aligned to the hg38 reference genome using STAR (2.7.5c) (Dobin et al., 2013), with the –soloType option set to SmartSeq and –soloUMIdedup set to Exact. The filtered matrices were loaded into Seurat (3.2.1) (Stuart et al., 2019). Cells were eliminated from analysis if they had high mitochondrial content (> 10% of reads), low read or gene detection numbers (< 1,000,000 reads or < 9000 genes), or low transcriptional complexity (< 0.58 $\log_{10}$ genes/count). Using Seurat, the QC-filtered counts were normalized using the SCTransform function. Then, integration anchors were selected for all cells sequenced and passing QC filters using SMART-seq2 or 10x using 3,000 features. All SMART-seq2, 10x Genomics, and HEK293T samples were integrated using the IntegrateData function. Then PCA analysis was performed, followed by UMAP reduction using dimensions 1:40. Clusters were identified using the FindNeighbors and FindClusters commands; the resolution used in this analysis was 0.06. Marker genes were found using the FindConservedMarkers command.

Genome alignment and TPM calculation for SMART-seq2 scDART-seq data

After demultiplexing using bcl2fastq2 and trimming adapter sequences using Flexbar (3.0.3), each cell library was aligned to the hg38 reference genome using STAR (2.7.5c). Transgene expression was assessed by adding the rat APOBEC1 sequence to the reference genome. PCR duplicates were removed from the BAM files using SAMtools (1.11) (Li et al., 2009) fixmate and markdup with the -r option. featureCounts (Subread 1.6.3) (Liao et al., 2014) was used to generate a counts matrix for all features. TPM values were calculated in R using a custom script from the counts matrix and the resulting TPM-cell matrix was filtered to only include cells that passed QC filters in Seurat and were included in the integrated dataset analysis.

reCAT for cell cycle analysis of SMART-seq2 scDART-seq data

Cell cycle assignments for SMART-seq2 cells were determined using reCAT (Liu et al., 2017), available on GitHub (https://github.com/tinglab/reCAT). Briefly, $\log_2$TPM+1 was calculated from the TPM-cell matrix to generate the input data. Expression values for cell cycle genes only were prepared using the get_test_exp() function. Cells were then indexed by their order within the cell cycle using get_ordIndex(). G1, S, and G2/M phase scores were obtained using the get_score() function and plotted using plot_bayes() and plot_mean(). Finally, the indexed cells were then assigned a cell cycle stage using a hidden markov model with the get_hmm_order() function.

Identification of m^{6}A sites from bulk DART-seq data

Bullseye is a custom perl-based software designed to identify RNA editing sites and is modified from the HyperTRIBE pipeline (Rahman et al., 2018). Bullseye is available on GitHub (https://github.com/mflamand/Bullseye). To use Bullseye for bulk DART-seq analysis, raw sequencing data was converted to fastq files using bcl2fastq2 software from Illumina, and adapter sequences were trimmed using Flexbar (3.0.3). Sequences were aligned to the hg38 genome using STAR (2.7.5c). PCR duplicates were removed from the BAM files using Samtools (1.11) fixmate and markdup with the -r option. Then, using Bullseye, the parseBAM.pl script was used to parse

the BAM files and create a counts matrix of the number of reads for each nucleotide at all positions with coverage. The Find_edit_site.pl script was then used to find C-to-U mutations with at least 10 reads of coverage, an edit ratio of 10%-95% (T_{reads}/$Total_{reads}$), an edit ratio at least 1.5-fold higher than mutant control samples, and at least 2 C-to-U editing events at a given site. Those sites were further filtered to include only those occurring in an RAC (G/A-A-C) motif. Editing events observed when APOBEC1 alone was overexpressed in HEK293T cells (Meyer, 2019) were removed. Editing events fulfilling these criteria and occurring in at least 2 of 3 APOBEC1-YTH cell biological replicates were considered high-m^{6}A sites. For each C-to-U editing site, the m^{6}A is considered to be the A immediately preceding the edited C.

Identification of high-confidence m^{6}A sites in 10x Genomics scDART-seq data

The sequences in the fastq files were aligned to the hg38 reference genome using Cell Ranger count the create aligned, deduplicated BAM files. Then, Bullseye was used to identify m^{6}A sites. First, the parseBAM.pl script was used to parse the sequences from the position-sorted bam files so that the number of A,T,G,C, and N at all positions with coverage was obtained for each APOBEC1-YTH cell barcode. The APOBEC1-YTH^{mut} cell BAM files were parsed as an averaged population without tracking barcodes. The Find_edit_site.pl script was then used to find C-to-U mutations within each cell that has at least 3 reads of coverage, an edit ratio of at least 10% (T_{reads}/$Total_{reads}$), and an edit ratio at least 1.5-fold higher than the APOBEC1-YTH^{mut} cell population average. Editing sites observed when APOBEC1 alone was overexpressed in HEK293T cells (Meyer, 2019) were removed. This defines low-stringency sites. Figure 1C shows the overlap of low-stringency sites identified within each 10x Genomics biological replicate. Finally, high-confidence sites were found by filtering editing events to include only those within an RAC motif that additionally occur in all 3 biological replicates and in at least 10 cells in total to account for potential sequencing errors.

Identification of high-confidence m^{6}A sites in SMART-seq scDART-seq data

After demultiplexing using bcl2fastq2 and trimming adapter sequences using Flexbar (3.0.3), each cell library was aligned to the hg38 reference genome using STAR (2.7.5c). PCR duplicates were removed from the BAM files using Samtools (1.11) fixmate and markdup with the -r option. The resulting APOBEC1-YTH^{mut} control cell BAM files were concatenated and sorted by position using samtools sort. Bullseye was then used to identify m^{6}A sites. First, the parseBAM.pl script was used to generate the nucleotide matrix file. Each APOBEC1-YTH cell BAM file was parsed so that the number of A,T,G,C, and N reads at all positions with coverage was obtained for each APOBEC1-YTH cell barcode. The parseBAM.pl script was also used to parse the concatenated APOBEC1-YTH^{mut} cell BAM file so that the number of A,T,G,C, and N at all positions with coverage was obtained as a population average of all APOBEC1-YTH^{mut} cells. The Find_edit_site.pl script was then used to find C-to-U mutations within each APOBEC1-YTH cell with at least 20 reads of coverage at each site, an edit ratio of at least 10-95% (T_{reads}/$Total_{reads}$), an edit ratio at least 1.5-fold higher than the APOBEC1-YTH^{mut} population average, and at least 2 C-to-U editing events at that site within the cell. Editing sites observed when APOBEC1 alone was overexpressed in HEK293T cells (Meyer, 2019) were removed. Low-stringency sites were found by filtering all sites for those occurring in an RAC motif. Finally, high-confidence m^{6}A sites were defined as editing events occurring within RAC sites and identified in at least 10 APOBEC1-YTH cells.

m^{6}A metagene and motif enrichment analysis

metaPlotR was used to generate m^{6}A and RAC motif metagenes and stop codon enrichment profiles. (Olarerin-George and Jaffrey, 2017). HOMER (Heinz et al., 2010) was used to identify enriched motifs near sites identified before filtering for sites occurring in RAC (R=G/A) motifs only. The sequence around each site was expanded 10nt in each direction and 8bp enriched motifs were found within those 21nt windows. Metagene of RNA-seq relative RNA-seq read coverage was generated using RseQC (Wang et al., 2012).

Comparison of methylated RNAs with REPIC database

A text file containing the genomic coordinates, gene annotation, and dataset information for MeRIP peaks reported in HEK293T cells from 3 separate studies (Lichinchi et al., 2016; Meyer et al., 2012; Schwartz et al., 2014) was downloaded from the REPIC database. (https://repicmod.uchicago.edu/repic/download.php) (Liu et al., 2020). Gene names were then retrieved from the Ensembl Gene ID annotations. RNAs with called peaks in at least 2 of the 3 studies were then compared to the list of RNAs containing high-confidence m^{6}A sites in the bulk DART-seq, 10x Genomics scDART-seq, or SMART-seq2 scDART-seq datasets.

Comparison of RNA abundance and methylation likelihood

The percentage of cells containing methylation in each RNA was obtained by calculating (# of cells with at least 1 m^{6}A site/# cells with expression of the parent RNA > 0.5) * 100. An RNA was considered expressed within a cell if the single-cell TPM was $\geq$ 0.5. The population-level gene expression was calculated by taking the $\log_{10}$ of the mean TPM value from all SMART-seq2 APOBEC1-YTH-expressing cells. The relationship between RNA abundance and likelihood of methylation was calculated in the following way. The nucleotide matrix generated by the parseBAM.pl script in the Bullseye pipeline was used to calculate the %C2U for all positions in all cells at which an m^{6}A site was found. Since detection of sites may potentially suffer from noise at locations of low coverage, and consequently bias the analysis, all positions with a coverage less than 250 reads were eliminated, and further, parent RNAs with a single-cell TPM value < 1 were also omitted. All sites with > 10% C2U were considered methylated and sites with < 10% C2U were considered unmethylated. This is the same cutoff used in identifying m^{6}A sites. The single-cell expression in TPM values for

the parent RNA each site is located on was also calculated. Then the mean TPM of the parent RNA for all methylated sites as well as all unmethylated sites was calculated, and significance determined using a Wilcoxon rank-sum test.

Comparison of RNA abundance and methylation likelihood among gene expression quintiles

For analysis of gene expression quintiles, the list of methylated and unmethylated sites in each cell was divided into 5 groups, with the sites occurring in the parent RNAs in 0^{th}-20^{th} percentiles of gene expression (by single-cell TPM values) binned in the lowest quintile, >20^{th}-40^{th} percentiles in the second quintile, >40^{th}-60^{th} percentiles in the third quintile, >60^{th}-80^{th} percentiles in the fourth quintile, and >80^{th}-100^{th} percentile in the fifth quintile. The mean TPM values of parent RNA expression for all methylated and unmethylated sites within each quintile was compared and normalized to the mean TPM of unmethylated sites within that quintile to give a relative expression level. Significance was determined using a Wilcoxon rank-sum test.

Gene ontology

Gorilla (http://cbl-gorilla.cs.technion.ac.il/) (Eden et al., 2007, 2009) was used for gene ontology analysis. For gene ontology analysis of differentially expressed RNAs, an unranked list of differentially expressed RNAs was compared to an unranked list of all RNAs with expression (population avg. TPM > 0.5). For gene ontology analysis of methylated RNAs or differentially methylated RNAs to a list all methylated RNAs. FDR values < $5.0x10^{-2}$ were considered significant.

Distribution of individual m^{6}A sites

First, the total number of high-confidence m^{6}A sites for each mRNA identified within the SMART-seq2 scDART-seq dataset was found (Figure 4A). To investigate the relationship between total sites in an mRNA and its expression within the entire population, the total number of m^{6}A sites in each mRNA was correlated with the $\log_{10}$ of the mean TPM value for the mRNA from all cells in the population. mRNAs with an average TPM value of at least 1 are shown (Figure S5F). The number of m^{6}A sites methylated on each mRNA per cell was assessed by finding the number of different, unique m^{6}A sites assigned to each mRNA that were associated with each SMART-seq2 cell ID. The percentage of m^{6}A sites methylated per mRNA per cell was calculated as (# of sites methylated on each mRNA within each cell ID/# of total sites on mRNA in population) * 100 (Figure S5C). The average percentage of cells each m^{6}A site is methylated in was calculated as (# of cells in which m^{6}A site is detected/# of cells with expression) * 100. For calculation of percent of cells with methylation, an mRNA was considered expressed when its single-cell TPM value is $\geq$ 0.5 (Figures 4B and 4D). Histograms of the number of sites per mRNA per cell show the number of single cells containing the indicated number of m^{6}A sites in the mRNA that are associated with the same cell ID (Figures 4G and 4J). Figure S5B shows the total number of m^{6}A sites on an mRNA across the entire population of SMART-seq2 cells, correlated with the percentage of cells in which that mRNA is methylated. The percentage of cells with methylation is calculated as (# of cells with at least 1 m^{6}A site in mRNA/# of cells with expression of mRNA) * 100. Single cell expression was defined as an mRNA having a TPM value of $\geq$ 0.5 in that cell.

mRNA m^{6}A site distribution graphs

To generate graphs showing the distribution of m^{6}A site frequency along transcripts, a BigWig file was first created with the 4^{th} column containing the percentage of cells in which that site is methylated (Of the total number of cells in which that RNA is methylated). The R package wiggleplotr (version 1.18) (https://bioconductor.org/packages/release/bioc/html/wiggleplotr.html) was then used to visualize the distribution of the sites along the transcript using the BigWig files to inform the height of each bar. Introns were shortened for ease of visualization.

Comparison of %C2U values with MAZTER-seq and SCARLET

First, the absolute %C2U value of the annotated cytidine following methylated adenosines identified by MAZTER-seq or SCARLET (Garcia-Campos et al., 2019; Liu et al., 2013) was calculated from all cells in the SMART-seq2 scDART-seq dataset. To do this, the percentage of total reads in which the annotated cytidine was called as a ''T'' (positive strand sites) or ''A'' (negative strand sites) was calculated. That number was correlated to the -$\log_{10}$ of the unnormalized MAZTER-seq cutting efficiency identified in HEK293T cells (Garcia-Campos et al., 2019). Due to noise of many sites containing very low %C2U in scDART-seq and low m^{6}A estimates in MAZTER-seq, sites were filtered to include only a high-stringency set. This set consists of sites with a cutting efficiency of less than 90% in the MAZTER-seq dataset (> 10% m^{6}A estimate), and 10%-95% population average %C2U estimates from scDART-seq data. The percentage of m^{6}A/A estimates from seven sites in the *MALAT1* RNA with m^{6}A estimates as measured by SCARLET (Liu et al., 2013) were compared to the %C2U calculated of the adjacent cytidine from the coverage in all SMART-seq2 scDART-seq cells. These sites in the *MALAT1* RNA were chosen as they were measured in HEK293T cells.

Analysis of %C2U in single cells

The %C2U values from the list of all high-confidence m^{6}A sites identified in bulk, 10x Genomics scDART-seq, and SMART-seq2 scDART-seq datasets was compared (Figures 5D and S6A). The %C2U was calculated only from sites in cells in which methylation was detected in scDART-seq data, omitting values from cells in which these sites were not identified as methylated. Similarly, to

compare the average %C2U in cells with APOBEC1-YTH expression, the single-cell TPM values were correlated to the mean %C2U of all m^{6}A sites identified within the same cell (Figure S6C). The correlation between APOBEC1-YTH^{mut} expression and %C2U in Figure S3J shows the %C2U at all called high-confidence scDART-seq sites compared to transgene expression. Figure S6D shows the distribution of all %C2U values for all bulk DART-seq sites that are also found in SMART-seq2 scDART-seq data. Shown is the population average pseudobulk %C2U for all SMART-seq2 sites in all SMART-seq 2 cells. Similarly, Figure S6E shows the population averaged %C2U in SMART-seq2 cells (y,-axis) for all sites identified in both bulk and SMART-seq2 datasets. Sites are stratified by the %C2U observed in bulk DART-seq (x-axis). Figure 5F shows all the same sites as Figure 5E, but instead of population average %C2U, the %C2U within each SMART-seq2 single cell (regardless of called methylation status) is shown as a data point. Histograms showing the distribution of %C2U values across cells (Figures 5H and S6G) show the distribution of the number of cells containing an identified m^{6}A site with the indicated %C2U value. Any cell without a called, high-confidence m^{6}A site (0%-10% C2U, or less than 1.5-fold higher than APOBEC1-YTH^{mut}-expressing cells), but with expression of the parent mRNA (TPM $\geq$ 0.5) is binned at 0 to indicate a lack of confident methylation in that cell.

Differential methylation analysis across cell cycle phases

To analyze differential methylation, cell cycle assignments obtained from reCAT were used for all SMART-seq2 APOBEC1-YTH cells. For each RNA, the number of cells with detectable expression within each cell cycle phase was found (single-cell TPM $\geq$ 0.5). Then the number of cells with expression in each phase was normalized to the total number of cells with expression. This proportion (always sums to 1 for all 3 phases) defines the expected distribution of methylation based on the proportion of cells with expression in each phase. Then, the number of cells in which the RNA contains at least 1 m^{6}A site was calculated. A chi-square test was performed to test whether the distribution of the number of cells with methylation within each phase of the cell cycle significantly differs from the expected distribution based on gene expression. A post-hoc exact test was also performed to identify the specific phases with differential methylation, and p-values less than 0.05 were considered significant. Significance = RNAs that were methylated in fewer than 25 total cells were eliminated from the analysis. Percentage of cells with methylation for all RNAs was then calculated within each phase of the cell cycle using (# cells with at least 1 m^{6}A site/total cells with expression) * 100. An RNA was considered expressed if its single-cell TPM value was $\geq$ 0.5.

Clustering by m^{6}A using SMART-seq2 scDART-seq cells

To cluster SMART-seq2 scDART-seq cells using m^{6}A methylation information, an input matrix was used that consists of the identified C2U editing ratio (values 0-1) at each m^{6}A site (Columns) identified in all cell IDs (Rows). For cells in which a site was not identified as methylated, a C2U editing ratio of 0 was used. This editing matrix was used as raw data input and loaded into Seurat using the CreateSeuratObject function. Then the following steps were performed using Seurat. The values were log-normalized using the NormalizeData function, and 2,000 variable features were found using the FindVariableFeatures function, using the "vst" selection method. The data was then scaled using ScaleData, followed by PCA. The RunUMAP function (first 20 PCA dimensions) was used for dimensionality reduction. The FindNeighbors and FindClusters functions were used to identify cells in different clusters. Different resolutions were tested and a final resolution of 0.29 was used as it identified distinct clusters that contained significantly differentially methylated sites. Differentially methylated sites were identified using the FindAllMarkers function, using the default Wilcoxon rank-sum test. Sites were considered significantly differentially methylated if the adjusted p-value < 0.05 and the magnitude of the $\log_2$-fold-change of the editing ratio between the me-clusters had an absolute value of 0.32 (1.25-fold) or greater. Differentially expressed RNAs were found by comparing the RNA expression values using FindAllMarkers between cells assigned to each me-cluster. RNAs were considered differentially expressed if they had an adjusted p-value of < 0.05 and the magnitude of the $\log_2$-fold-change of the expression between the me-clusters had an absolute value of 0.32 (1.25-fold) or greater.

Clustering by m^{6}A using 10x Genomics scDART-seq cells

To cluster 10x Genomics scDART-seq cells using m^{6}A methylation information, an input matrix was used that consists of the identified C2U editing ratio (values 0-1) at each m^{6}A site (Columns) identified in all cell IDs (Rows). For cells in which a site was not identified as methylated, a C2U editing ratio of 0 was used. This editing matrix was used as raw data input and loaded into Seurat using the CreateSeuratObject function. Then the following steps were performed using Seurat: The values were log-normalized using the NormalizeData function, and 2,000 variable features were found using the FindVariableFeatures function, using the "vst" selection method. The data were then scaled using ScaleData, followed by PCA. The RunUMAP function (first 20 PCA dimensions) was used for dimensionality reduction. The FindNeighbors and FindClusters functions were used to identify cells in different clusters. Different resolutions were tested and a final resolution of 0.05 was used as it identified distinct clusters that contained significantly differentially methylated sites. Differentially methylated sites were identified using the FindAllMarkers function, using the default Wilcoxon rank-sum test. Sites were considered significantly differentially methylated if the adjusted p-value < 0.05 and the magnitude of the $\log_2$-fold-change of the editing ratio between the me-clusters had an absolute value of 0.5849 (1.5-fold) or greater. Differentially expressed RNAs were found by comparing the RNA expression values using FindAllMarkers between cells assigned to each me-cluster. RNAs were considered differentially expressed if they had an adjusted p-value of < 0.05 and the magnitude of the $\log_2$-fold-change of the expression between the me-clusters had an absolute value of 0.32 (1.25-fold) or greater.

Statistical analysis and plotting

All statistical analysis and plotting was performed in RStudio. T-tests, Wilcoxon rank-sum tests (except for scRNA-seq gene expression analysis), chi-square tests (comparisons of cell cycle distribution), Pearson correlation coefficients, and p-values for correlations were calculated using the ''base'' package. Post-hoc exact tests after chi-square tests were performed with the chisq.theo.multcomp function in the ''RVAideMemoire'' package. All UMAP visualizations were generated using package ''Seurat''. The R package ''Eulerr'' was used to make Venn diagrams. All other charts were generated using the ''ggplot2'' package.