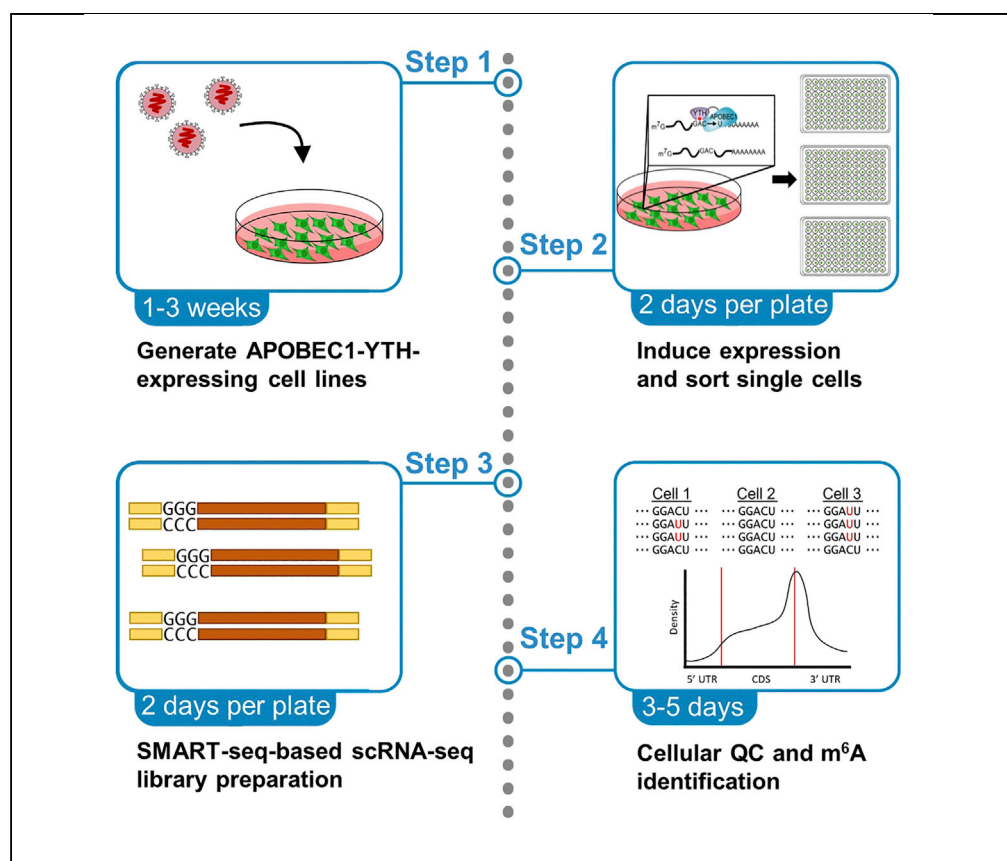


Protocol

Detection of m⁶A in single cultured cells using scDART-seq



Matthew Tegowski,
Kate D. Meyer

matthew.tegowski@duke.edu (M.T.)
kate.meyer@duke.edu (K.D.M.)

Highlights

Generation of transgenic cell lines expressing inducible APOBEC1-YTH

Transgene induction and sorting of individual cells

Preparation of sequencing libraries from single cells

Processing data to identify m⁶A sites from C-to-U mutations in the sequencing data

Most techniques for mapping m⁶A-methylated RNAs transcriptome-wide require large amounts of RNA and have been limited to bulk cells and tissues. Here, we provide a detailed protocol for the identification of m⁶A sites in single HEK293T cells using single-cell DART-seq (scDART-seq). The protocol details how to generate cell lines with inducible expression of the APOBEC1-YTH transgene and the use of important controls for minimizing false positives. We also describe the bioinformatic analysis to identify m⁶A sites.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Tegowski & Meyer, STAR
Protocols 3, 101646
September 16, 2022 © 2022
The Author(s).
<https://doi.org/10.1016/j.xpro.2022.101646>



Protocol

Detection of m⁶A in single cultured cells using scDART-seqMatthew Tegowski^{1,3,*} and Kate D. Meyer^{1,2,4,*}¹Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710, USA²Department of Neurobiology, Duke University School of Medicine, Durham, NC 27710, USA³Technical contact⁴Lead contact*Correspondence: matthew.tegowski@duke.edu (M.T.), kate.meyer@duke.edu (K.D.M.)<https://doi.org/10.1016/j.xpro.2022.101646>

SUMMARY

Most techniques for mapping m⁶A-methylated RNAs transcriptome-wide require large amounts of RNA and have been limited to bulk cells and tissues. Here, we provide a detailed protocol for the identification of m⁶A sites in single HEK293T cells using single-cell DART-seq (scDART-seq). The protocol details how to generate cell lines with inducible expression of the APOBEC1-YTH transgene and the use of important controls for minimizing false positives. We also describe the bioinformatic analysis to identify m⁶A sites.

For complete details on the use and execution of this protocol, please refer to Tegowski et al. (2022).

BEFORE YOU BEGIN

This protocol describes the identification of m⁶A sites in HEK293T cells expressing a doxycycline-inducible APOBEC1-YTH protein stably introduced using lentivirus. However, the following steps can be adapted to detecting m⁶A sites from any cell line. There are 4 major steps: (1) Generating a stable cell line with inducible APOBEC1-YTH or APOBEC1-YTH^{mut} expression, (2) inducing transgene expression and sorting single cells, (3) library preparation and Illumina sequencing, and (4) data analysis and identification of m⁶A methylation. The file sizes for the data analysis section are very large. About 2 TB of disk space and 64 GB of RAM are required on a Linux-based system analyze this dataset.

Software setup

⌚ Timing: 4 h

Most of bioinformatic steps of this protocol are designed to be accomplished on a Linux-based system using the Slurm workload manager. This protocol can still be accomplished without the Slurm workload manager, but the scripts will need to be modified. The instructions for installation provided use a miniconda environment to install the required programs without root access. Detailed instructions for installation of all required programs can also be found at our GitHub: <https://github.com/tegowski/scDARTHEKcells>.

Additionally, the single-cell quality control and gene expression analysis uses R, which can be downloaded to your computer from <https://cran.r-project.org/>, or alternatively installed within a conda environment on a Linux system as described below. Specifically, the R package Seurat (Stuart et al., 2019) is used to perform quality control and gene expression analysis of single cells. Detailed instructions for its installation and use can be found at (<https://satijalab.org/seurat/>) as well as below.



1. Create conda environment with required programs.

Note: \$WORKDIR is the path to your primary working directory for this project. In bash scripts, it is defined as a variable and the path to your working directory should replace the "/your/path/here/" string. In R scripts, all instances of \$WORKDIR should be directly replaced with the path to your working directory.

```
$ mkdir $WORKDIR/software
$ cd $WORKDIR/software
#Create conda environment and install required Perl modules
$ conda env create -f bullseye.yml
$ conda activate Bullseye
$ wget http://www.cpan.org/authors/id/T/TO/TODDR/XML-Parser-2.46.tar.gz
$ tar -xf XML-Parser-2.46.tar.gz
$ cd XML-Parser-2.46
$ perl Makefile.PL EXPATLIBPATH=$CONDA_PREFIX/lib EXPATINCPATH=$CONDA_PREFIX/include
$ make
$ make install
$ cpanm Bio::DB::Fasta
$ cpanm Text::NSP
$ cpanm Array::IntSpan
$ cpanm MCE
#Install flexbar and STAR
$ conda install -c bioconda flexbar=3.0.3
$ conda install -c bioconda STAR=2.7.5c
#Install R
$ conda install -c r r=4.0.0
```

2. Install Seurat and dependencies.

Note: Installation of Seurat in this manner will also install required dependencies.

```
#To start R if using in Linux terminal
$ R
#Once in R, install required packages
> if (!require('BiocManager', quietly = TRUE))
  install.packages('BiocManager')
> BiocManager::install(version = '3.15')
> Install.packages('Seurat')
```

```
> library(Seurat)
> install.packages('tidyverse')
```

3. Download sequencing files from SRA.

Note: This is only required if only performing the bioinformatic analysis to replicate the sequencing of 300 APOBEC1-YTH and 300 APOBEC1-YTH^{mut} single cells. This is not required if generating your own data. These are very large files and will require nearly 2 TB of disk space for all of them.

```
#In the $WORKDIR/software directory, run the script to download SMART-seq single-cell fastq
files

$SBATCH fastqdump.sh

#Below are the contents of fastqdump.sh

#!/bin/bash

for ((i = 68425; i <= 68461; i++))
do

fastq-dump -outdir ../YTH -split-files SRR152$i

done

for ((i = 68889; i <= 70105; i++))
do

fastq-dump -outdir ../YTH -split-files SRR152$i

done

for ((i = 461; i <= 876; i++))
do

fastq-dump -outdir ./ -split-files SRR15268$i

done

$SBATCH renameYTH.sh

#Below is the contents of renameYTH.sh

#!/bin/bash

#SBATCH -a 1-301

file=$(ls *1.fastq | sed -n ${SLURM_ARRAY_TASK_ID}p)

STEM=$(basename "$file" *_1.fastq)

mv ${STEM}_1.fastq WtCell${SLURM_ARRAY_TASK_ID}_1.fastq

mv ${STEM}_2.fastq WtCell${SLURM_ARRAY_TASK_ID}_2.fastq

$SBATCH renameYTHmut.sh

#Below is the contents of renameYTHmut.sh

#!/bin/bash

#SBATCH -a 1-300

file=$(ls *1.fastq | sed -n ${SLURM_ARRAY_TASK_ID}p)
```

```
STEM=$(basename "$file" _1.fastq)

mv ${STEM}_1.fastq WTCellmut${SLURM_ARRAY_TASK_ID}_1.fastq
mv ${STEM}_2.fastq WTCellmut${SLURM_ARRAY_TASK_ID}_2.fastq

$rm -r $WORKDIR/YTH
$rm -r $WORKDIR/YTHmut
```

4. Generate genome index for alignment. This will generate a genome index for the hg38 build of the human genome. This will need to be modified if using other organisms or genome builds.

```
#Download genome fasta files and gtf annotation file

$ mkdir $WORKDIR/genomefasta
$ cd $WORKDIR/genomefasta
$ for ((I = 1; I <= 22; i++))
do
wget http://ftp.ensembl.org/pub/release-98/fasta/homo\_sapiens/dna/Homo\_sapiens.GRCh38.dna\_sm.chromosome.1.fa.gz
done

wget http://ftp.ensembl.org/pub/release-98/fasta/homo\_sapiens/dna/Homo\_sapiens.GRCh38.dna\_sm.chromosome.X.fa.gz
wget http://ftp.ensembl.org/pub/release-98/fasta/homo\_sapiens/dna/Homo\_sapiens.GRCh38.dna\_sm.chromosome.Y.fa.gz
wget http://ftp.ensembl.org/pub/release-98/fasta/homo\_sapiens/dna/Homo\_sapiens.GRCh38.dna\_sm.chromosome.MT.fa.gz
wget http://ftp.ensembl.org/pub/release-98/fasta/homo\_sapiens/dna/Homo\_sapiens.GRCh38.dna\_sm.nonchromosomal.fa.gz

$ gunzip *
$ mkdir $WORKDIR/gtf
$ cd $WORKDIR/gtf
$ wget 'http://ftp.ensembl.org/pub/release-98/gtf/homo_sapiens/Homo_sapiens.GRCh38.98.gtf.gz'
$ gunzip *

#Create STAR index
$ cd $WORKDIR/software
$ sbatch STARindex.sh

#Contents of STARindex.sh

#!/bin/bash

#SBATCH --mem=60G

#SBATCH -c 4

$WORKDIR="/your/path/here"

STAR --runThreadN 4 \
--runMode genomeGenerate \
```

```
-genomeDir $WORKDIR/STARindex \
-genomeFastaFiles $WORKDIR/genomefasta/* \
-sjdbGTFfile $WORKDIR/gtf/Homo_sapiens.GRCh38.98.gtf \
-sjdbOverhang 49 \
-genomeSAindexNbases 3
```

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Phosphate Buffered Saline	Gibco	70011-044
OPTI-MEM	Gibco	11058021
Dulbecco's Modified Eagle Medium	Corning	10-013-CV
Fetal Bovine Serum	VWR	97068-085
Sucrose	VWR	BDH9308
Puromycin	Sigma-Aldrich	A1113803
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
1 M Tris HCl pH8.5	Rockland	MB-027-1000
Critical commercial assays		
Qubit 1x dsDNA HS Assay Kit	Invitrogen	Q33230
Qubit BR RNA Assay Kit	Invitrogen	Q10210
QIAquick Gel Extraction Kit	QIAGEN	28704
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 (Table S1 and Mendeley Data, https://doi.org/10.17632/96cpc2px6.1)
AMPure XP Beads	Beckman Coulter	A63880
Foil adhesive covers	Thermo Fisher Scientific	AB-0626
High Sensitivity DNA Kit	Agilent	5067-4626
Deposited data		
Raw and processed data	GEO	GEO: GSE180954
Bullseye code	GitHub	https://github.com/mflamand/Bullseye
List of identified sites	Mendeley Data	https://doi.org/10.17632/96cpc2px6.1
Experimental models: Cell lines		
HEK293T (Human embryonic kidney, female)	ATCC	ATCC CRL-3216
iDART-APOBEC1-YTH (Human embryonic kidney, female)	Tegowski et al. (2022)	N/A
iDART-APOBEC1-YTHmut (Human embryonic kidney, female)	Tegowski et al. (2022)	N/A
Oligonucleotides		
ActinFwd (Sanger sequencing)	Tegowski et al. (2022)	caacacagtgtctgtctggc
ActinRev (Sanger sequencing)	Tegowski et al. (2022)	caagatgagattggcatggc
Recombinant DNA		
TLCV2-APOBEC1-YTH	Addgene	Addgene: 178949

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
TLCV2-APOBEC1-YTHmut	Addgene	Addgene: 178950
pMD2.G	Unpublished	Addgene: 12259
psPAX2	Unpublished	Addgene: 12260
Other		
Bioanalyzer 2100	Agilent	G2939BA
Qubit 4 fluorometer	Thermo Fisher Scientific	Q33238
0.5–12 μ L 12-channel electronic pipette with adjustable spacers (Voyager II)	INTEGRA Biosciences	4731
BSLII biosafety cabinet	Labconco	332391001
Cell culture incubator supplying 5% CO ₂	Thermo Fisher Scientific	13-998-211
Ultracentrifuge compatible with sw28 rotor	Beckman Coulter	L8-60M
Sw28 rotor	Beckman Coulter	342207
50 mL ultracentrifuge tubes	Beckman Coulter	344058
Inverted fluorescence microscope		DMi8
PCR clean hood with HEPA filter	USA Scientific	4264-8588
96-well magnetic separator	Sergi Lab Supplies	B08134P9RT
Thermocycler with 384- and 96-well heat blocks	TEquipment	T5000-384
60 mm cell culture plates	Eppendorf	0030701119
10 cm cell culture plates	Eppendorf	0030702115
384-well plates	VWR	82006-678
40 μ m nylon cell strainers	Corning	431750
0.2 mL 96-well PCR plates	VWR	82006-636
Adhesive plate covers	Bio-Rad	MSB1001
Foil adhesive plate covers	Bio-Rad	MSF1001
p10 filter tips	Thomas Scientific	1159M41
P20 filter tips	Thomas Scientific	1159M43
P200 filter tips	Thomas Scientific	1159M40
P1000 filter tips	Thomas Scientific	P1126
Benchtop centrifuge for 1.7 mL tubes	Eppendorf	5404000014
Benchtop centrifuge for 15 mL conical tubes and PCR plates	Thermo Fisher Scientific	Sorvall Legend RT
Software and algorithms		
Single-cell Bullseye	This paper	https://github.com/tegowski/scDARTHEKcells https://doi.org/10.5281/zenodo.6869039
Bullseye	Tegowski et al. (2022)	https://github.com/mflamand/Bullseye
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
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

MATERIALS AND EQUIPMENT

Growth medium		
Reagent	Final concentration	Amount
Dulbecco's Modified Eagle Medium	90%	500 mL
Fetal Bovine Serum	10%	55.5 mL

Sucrose solution (20% w/v)		
Reagent	Final concentration	Amount
Phosphate-buffered saline (PBS)	N/A	200 mL
Sucrose	20% (w/v)	40 g

Selection medium		
Reagent	Final concentration	Amount
Dulbecco's Modified Eagle Medium	90%	500 mL
Fetal Bovine Serum	10%	55.5 mL
Puromycin	2 µg/mL	1 mg

SMART-seq HT lysis buffer (Included in SMART-seq HT Kit)		
Reagent	Final concentration	Amount
10× Lysis Buffer	1×	0.3 mL
RNase Inhibitor	1×	12 µL
3' SMART-Seq CDS Primer IIA	N/A	240 µL
ddH ₂ O	N/A	2.45 mL
Total	N/A	3 mL

Flow buffer		
Reagent	Final concentration	Amount
Dulbecco's Modified Eagle Medium	99%	495 mL
Fetal Bovine Serum	1%	5 mL

Elution buffer		
Reagent	Final concentration	Amount
1 M Tris HCl pH8.5	10 mM	0.1 mL
Nuclease-free water	N/A	9.9 mL

STEP-BY-STEP METHOD DETAILS

Production of lentivirus

⌚ Timing: 1 week

The main purpose of lentiviral transduction is to obtain clonal cell lines with transgene induction that is consistent from cell-to-cell. However, as an alternative, these plasmids can be used to perform transient transfection, followed by single-cell sorting of only GFP⁺ cells to perform scDART-seq. This may result in greater variation of transgene expression within each cell but would be much faster and easier than generating clonal cell lines. In addition, if using a heterogenous cell line, the process of selecting clones may eliminate most or all of the heterogeneity, which may be of interest in a

single-cell experiment. In this case, transient transfection is advised. Plasmid sequences and detailed plasmid maps are available at Addgene (see [key resources table](#)).

1. Culture HEK293T cells on a 15 cm plate in growth medium until they are at approximately 80% confluency.

Optional: HEK293T cells are easily disrupted from the plate. If this is an issue, the plates can be coated with gelatin or fibronectin so that they adhere more strongly.

2. Transiently transfect viral vectors.
 - a. In a sterile 15 mL tube, add 1 mL OPTI-MEM. Then add plasmids: 26.75 μ g of TLCV2-APOBEC1-YTH (or TLCV2-APOBEC1-YTH^{mut}), 20 μ g of psPAX2, and 6.25 μ g of pMD2.G. Then add 3 μ g of jet-PEI solution, gently mix. Incubate at 23°C for 30 min.
 - b. Apply to cells dropwise and culture cells at 37°C with 5% CO₂ for 24 h.
3. After 24 h, aspirate and add 20 mL of fresh growth medium.
4. Culture cells at 37°C with 5% CO₂ for a further 48 h. GFP fluorescence should be visible.
5. After 48 h, sterilize ultracentrifuge tubes with 70% ethanol and allow them to air dry in a biosafety cabinet. Aspirate to dry any residual ethanol if necessary.
6. Remove medium from cells and pipette into a 50 mL conical. Centrifuge at 5,000 \times g for 10 min.
7. Remove supernatant and filter with 0.45 μ m nylon filter.

Note: The following steps are to concentrate purified virus. These steps are optional, but recommended. If storing unpurified supernatant, 1 mL aliquots should be stored at –80°C.

8. Add filtered medium to ultracentrifuge tube.
9. Carefully add 4 mL of sucrose solution to the bottom of the ultracentrifuge tube to create distinct layers. Use a 5 mL serological pipette for this and remove slowly. There should be 2 distinct layers.

Note: To make 20% sucrose solution, dissolve sucrose in 150 mL of PBS. Once dissolved, add enough PBS to make 200 mL of solution, then filter-sterilize. Store solution at 4°C for up to 1 year.

10. Fill tubes to about $\frac{3}{4}$ of their height with PBS and use a balance to ensure tubes are of the same weight. Take care not to mix layers.
11. Centrifuge at 19,700 rpm for 2 h at 4°C in a Beckman sw28 swinging bucket rotor.
12. Decant supernatant into 10% bleach. Leave tube inverted on Kimwipes for ~5 min. Use vacuum carefully to remove beads of supernatant on the bottom of the tube without aspirating the viral pellet.
13. Add 100 μ L of PBS to the bottom of each tube.
14. Resuspend by incubating 16–18 h at 4°C in the dark, with rocking.
15. Centrifuge at 7,000 \times g for 1 min at 4°C.
16. Transfer to a new tube, and aliquot into 30 μ L aliquots on ice.
17. Freeze in dry ice/ethanol bath and store at –80°C.

Generation of stable cell lines

⌚ **Timing:** 3 weeks in generating single-cell clonal cell lines

Here, the purified lentivirus made in the previous section is applied to HEK293T cells to insert either the APOBEC1-YTH or APOBEC1-YTH^{mut} transgene. Then cells are diluted and plated at less than one cell per well in 96-well plates and allow to form clonal cell lines. The resulting clonal cell lines will have more consistent transgene induction than heterogenous mixture of virally transduced cell lines.

18. Culture HEK293T cells on 60 mm plate in DMEM + 10% FBS until the cells are at about 70% confluency in 3 mL of medium.
19. Apply 30 μ L of resuspended virus to medium and culture cells at 37°C with 5% CO₂ for 24 h.
20. Aspirate medium and add 2 mL of selection medium and culture cells at 37°C with 5% CO₂ for 7 days, replacing with fresh selection medium daily.
21. After 7 days of puromycin selection, add 500 μ L of trypsin and incubate at 37°C for 5 min to make a single cell suspension and resuspend in 10 mL of growth medium.
22. Count cells using a hemacytometer or automated cell counter.
23. Make serial dilutions using growth medium until cells are at a concentration of 1 cell per 200 μ L.
24. Plate 150 μ L of the cell suspension per well on 10 96-well plates.

Note: As an alternative to clonal selection, single cells can be directly sorted from this pool of transgenic cells. Direct sorting of GFP⁺ cells from this pool may be beneficial if using a more heterogenous cell line that contains multiple subpopulations. Clonal selection is recommended when consistent induction within all cells is important.

25. Using inverted microscope, identify and label wells containing only a single cell. Culture cells at 37°C with 5% CO₂.
26. Identify and culture colonies growing from wells containing single cells.
27. The cells should be expanded until at least a confluent 10 cm plate is obtained.

Testing of stable cell lines

⌚ Timing: 3 days

The major goal of generating clonal cell lines is to obtain consistent induction of APOBEC1-YTH in all cells within the population. This section describes testing of candidate cell lines to ensure consistent induction of APOBEC1-YTH and detectable C-to-U editing before preparing sequencing libraries. It may be necessary to screen 10–15 potential clonal cell lines to identify a high-quality candidate. After clonal cell line identification, we recommend proceeding with a single APOBEC1-YTH and APOBEC1-YTH^{mut} line, as this will ensure sequenced cells have as consistent transgene induction as possible.

28. Plate 200,000 cells of each inducible APOBEC1-YTH- and APOBEC1-YTH^{mut}-expressing clonal stable cell line to be tested on separate 10 cm plates and culture for 24 h.
29. Treat with 1 μ g/mL of doxycycline and culture for 24 h.
30. Assess percentage of cells expressing GFP using fluorescent microscopy. Only select cell lines with ~100% fluorescent cells (Figure 1A).

Note: Cell lines with highly variable induction of transgene or GFP signal suggest that they may not be truly clonal and deriving from a single cell. It is possible to sort GFP⁺ cells from such a cell line, however more variable transgene induction may be observed. We observed a high rate of cell lines with 100% of cells displaying GFP signal, and this should not be a rare event if a majority of clones are truly derived from a single cell.

31. Assess C-to-U mutations at a known site in ACTB mRNA (Figure 1B).
32. RNA isolation.
 - a. Aspirate medium, rinse plates with cold PBS, and pour off excess PBS.
 - b. Use a cell scraper to harvest cells and pipette into a 1.5 mL tube.
 - c. Centrifuge at 2,000 \times g for 1 min to pellet cells.
 - d. Remove PBS and add 1 mL of Trizol. Pipette to mix and homogenize.
 - e. Incubate at 23°C for 5 min.
 - f. Add 200 μ L of chloroform and vortex for 10–15 s.

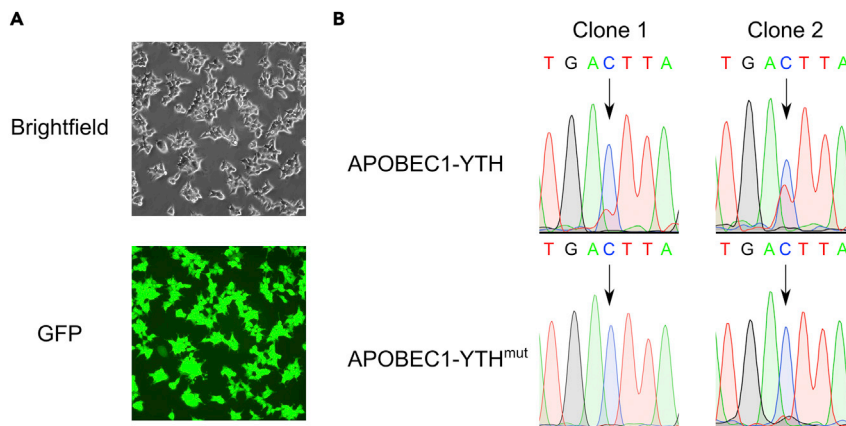


Figure 1. Validation of stable cell lines expressing inducible APOBEC1-YTH

(A) Fluorescence imaging showing transgene induction in all cells of a clonally selected cell line.

(B) Sanger sequencing traces showing expected C-to-U mutations at ACTB C1223 in cells expressing APOBEC1-YTH (top), but not those expressing APOBEC1-YTH^{mut} (bottom).

- g. Centrifuge at 13,000 × g for 15 min at 4°C.
- h. Transfer 400 µL of top aqueous layer to a new tube.
- i. Add 500 µL of isopropanol, mix by inverting, and incubate for 10 min at 23°C.
- j. Centrifuge at 13,000 × g for 10 min at 4°C.
- k. Remove supernatant and resuspend pellet in 70% ethanol. Vortex to dislodge pellet from bottom of tube.
- l. Centrifuge at 13,000 × g for 5 min at 4°C.
- m. Remove all of supernatant and air dry for ~5 min.
- n. Resuspend pellet in 90 µL of RNase-free water.
33. Removal of contaminating DNA.
 - a. Use RNase-free DNaseI to eliminate contaminating DNA fragments. To each tube with RNA solubilized in 90 µL of water.
 - i. Add 10 µL of DNaseI 10× buffer.
 - ii. Add 2 µL of DNaseI.
 - iii. Incubate at 37°C for 15 min.
 - b. Ethanol precipitation to recover RNA.
 - i. Add 10 µL of sodium acetate solution.
 - ii. Add 250 µL of 100% ethanol.
 - iii. Incubate at –80°C for 1 h.
 - c. Centrifuge at 13,000 × g for 15 min at 4°C to pellet RNA.
 - d. Completely remove supernatant and air dry for ~5 min.
 - e. Resuspend pellet in 90 µL of RNase-free water.

△ CRITICAL: After isolation, RNA should be kept on ice while thawed, and it should be stored at –80°C when not in use.

34. Quantify RNA using Qubit RNA BR Assay.
35. Synthesize cDNA using iScript Supermix in 0.2 mL PCR tubes. The thermocycler lid should be set to 105°C.

Note: Though 200 ng of total RNA are listed, as little as 10 ng can be used without affecting the ability to detect C-to-U mutations.

PCR reaction master mix	
Reagent	Amount
iScript Supermix	4 μ L
Total RNA	200 ng
ddH ₂ O	To 20 μ L

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial annealing	25°C	5 min	1
Extension	46°C	20 min	1
Inactivation	96°C	1 min	1
Hold	4°C	forever	

36. PCR amplify region surrounding *ACTB* C1223 in 0.2 mL PCR tubes, with the thermocycler lid set to 105°C.

PCR reaction master mix	
Reagent	Amount
cDNA template	1 μ L
CloneAmp Mastermix	5 μ L
ActinFwd primer	0.5 μ L
ActinRev primer	0.5 μ L
ddH ₂ O	3 μ L

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	10 s	30 cycles
Annealing	58°C	30 s	
Extension	72°C	5 s	
Final extension	72°C	5 min	1
Hold	4°C	forever	

37. Separate amplified fragment using 1% agarose gel (w/v) and use razor blade to extract the ~150 bp fragment.

Note: A suitable DNA ladder (ideally containing a 100 bp and 200 bp marker) should be run adjacent to the samples so the band size can be properly estimated, ensuring that the correct band is excised.

38. Purify DNA using QIAquick Gel Extraction Kit.
39. Send for Sanger sequencing using the ActinFwd primer.
40. Observe induction of C-to-T mutation at C1223 only in doxycycline-treated APOBEC1-YTH-expressing cell lines, but not in APOBEC1-YTH^{mut}-expressing cell lines (Figure 1B).

△ CRITICAL: We recommend collecting RNA less than 36 h after transgene induction, as prolonged APOBEC1-YTH expression (longer than 48 h) can negatively impact cell health.

Transgene induction and sorting of single cells

⌚ Timing: 2 days

In this step, expression of the APOBEC1-YTH or APOBEC1-YTH^{mut} transgenes is induced with doxycycline, resulting in C-to-U mutations in mRNAs adjacent to sites of m⁶A methylation. After doxycycline induction, single cells are sorted into 384-well plates prior to the preparation of sequencing libraries.

41. Plate 200,000 cells of inducible APOBEC1-YTH- and APOBEC1-YTH^{mut}-expressing stable HEK293T cells from clonal cell lines selected in the previous section on separate 10 cm plates for a total of 2 plates.
42. When cells are approximately 70% confluent, aspirate medium and add 10 mL of growth medium with 1 µg/mL doxycycline and incubate at 37°C for 24 h.

Note: Flow buffer should be prepared the day before and stored at 4°C until ready for use.

43. The next day, prepare 3 mL of SMART-seq HT lysis buffer and aliquot 1 µL into each well of 7 384-well PCR plates.

Note: The lysis buffer should be prepared on the same day as the cell sorting. After making the solution and aliquoting into 384 well plates, the plates should be covered with foil adhesive plate covers and kept on ice until ready for use. Use within the same day as aliquoting.

44. Generate a single cell suspension for sorting individual cells. The APOBEC1-YTH and APOBEC1-YTH^{mut} cells should be kept separate throughout this process and the tubes should be sorted sequentially.
 - a. Aspirate medium and add 10 mL of sterile PBS.
 - b. Repeat step 40a.
 - c. Aspirate PBS and add 2 mL of TrypLE and incubate at 37°C with 5% CO₂ for 5 min to dissociate the cells.
 - d. Add 12 mL of flow buffer and gently pipette cells with 10 mL pipette up and down 5–10 times to break up clumps. Transfer to 15 mL conical tubes.

△ **CRITICAL:** Cells should be pipetted to generate single cells and avoid clumps. However, vigorous pipetting will lead to reduced viability, potentially compromising the sample.

- e. Count cells using hemacytometer or automated cell counter.
- f. Add 2 million cells to a 1.5 mL tube and pellet by centrifuging at 1,000 × g for 3 min at 23°C.

△ **CRITICAL:** It is important not to centrifuge cells above 1,000 × g to maintain high cell viability. Centrifugation at lower speeds may improve viability but lead to greater cell loss. Use polypropylene tubes during these steps.

- g. Remove as much supernatant as possible without disturbing the pellet.
- h. Resuspend pellet in 200 µL of flow buffer. Gently dissociate the pellet using a wide-bore P1000 pipette tip with pipette set to 150 µL.

△ **CRITICAL:** Gently resuspending cell pellets using wide-bore pipette tips is recommended, as vigorous dissociation can damage cells, increase cellular debris, and decrease viability. Avoid creating large amounts of bubbles.

Note: Instead of wide-bore tips, it is possible to cut the bottom of standard P1000 pipette tips using sterilized scissors or a razor blade to expand the bore size.

- i. Add 2 µL of near-IR LIVE/DEAD viability dye and incubate in the dark for 5 min at 23°C.
- j. Add 2 µL (4 units) of DNaseI. Gently mix by pipetting with a wide-bore P1000 pipette tip. Incubate for 15 min at 23°C in the dark.

- k. Pellet the cells by centrifugation at $1,000 \times g$ for 3 min.
 - l. Remove as much supernatant as possible without disturbing the pellet.
 - m. Add 1 mL of flow buffer and gently dissociate the pellet with a wide-bore P1000 pipette tip.
 - n. Filter cells using 40 μm nylon filter.
45. Use a FACS Aria II to sort 1 viable cell per well of each plate. Immediately after a plate is finished sorting, cover with a foil adhesive plate cover, spin plate at $1,000 \times g$ for 1 min in a swinging bucket centrifuge, and immediately place on dry ice.

Note: Cells should be gated on forward/side scatter, with doublet discrimination using forward scatter area vs. log settings. Single cells should be further gated on GFP expression. Only cells with a similar amount of GFP signal should be sorted to improve consistent transgene expression from cell-to-cell. Therefore, it is recommended to set a narrow gate on GFP expression, collecting cells with moderate levels of GFP (i.e., not a low or high extremes).

Note: Cells should not be highly concentrated, to reduce the rate of accidental double sorting and prevent doublets. Therefore, the suspension of single-cells to be sorted should be made to about 500,000 cells per 1 mL. Sorting can be stressful on the cells, so the largest nozzle size with the lowest pressure is recommended, we used a 100 μm nozzle with 20 psi of pressure.

Note: Although a FACS Aria II was used to sort, any cell sorter capable of sorting human cells can be used.

▮▮ Pause point: Cells may be stored at -80°C for at least 3 weeks.

SMART-seq2 cDNA synthesis

⌚ **Timing:** 1–2 days for a single 384-well plate

Here, the cDNA libraries that will be used to prepare sequencing libraries are created. In particular, this section outlines the protocol for reverse transcription followed by cDNA amplification and post-PCR clean-up.

Note: The reverse transcription and cDNA amplification steps are accomplished according to the manufacturer's protocol ([Takara SMART-seq HT Kit User Manual](#)). However, the reaction volumes have been decreased 12.5-fold.

Note: It is highly recommended to use a well-calibrated, electronic multichannel pipette with adjustable spacers for all PCR-related steps if performing these steps by hand. Adjustable spacers are optional, but useful for transferring between 384-well plates and 96-well plates. However, pipetting small volumes with manual pipettes will result in inaccuracies and increased variability in PCR efficiency from well-to-well and therefore electronic pipettors are highly recommended. Manual multichannel pipettes may be used for DNA purification steps using AmpPureXP beads. Additionally, if accessible, robotic liquid handlers may be used. It is also recommended that the library construction steps are performed in a clean hood or biosafety cabinet to reduce likelihood of contamination. Additionally, multichannel pipettes are recommended for bead purification steps.

Note: Do not re-use tips at any step, this could lead to sample cross-contamination. Additionally, filter tips should be used at all steps to reduce the risk of cross-contamination. The DNA purification steps with AmpPureXP beads require the use of many p20 and p200 tips, ensure that enough tips are readily available before starting.

46. Thaw one of the 384-well plates with sorted single cells on ice. Centrifuge at $2,000 \times g$ for 5 min at 4°C .
47. Incubate plate in a thermocycler set to 72°C for 3 min.
48. Immediately remove the plate from the thermocycler and incubate on ice for 2 min.
49. While on ice, add $1\ \mu\text{L}$ of One-step mastermix (provided in SMART-seq HT Kit) and briefly mix by pipetting.
50. Cover plate with adhesive plate cover and incubate in thermocycler with lid set to 105°C to perform reverse transcription immediately followed by cDNA amplification steps.

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Reverse transcription	42°C	90 min	1
Initial denaturation	95°C	1 min	1
Denaturation	98°C	10 s	18 cycles
Annealing	65°C	30 s	
Extension	68°C	3 min	
Final extension	72°C	10 min	1
Hold	4°C	forever	

▮▮ Pause point: Plate may be stored at -20°C for at least 4 weeks. Plates should be covered with foil adhesive plate covers.

51. Add $5.5\ \mu\text{L}$ nuclease-free water per well ($7.5\ \mu\text{L}$ total). Transfer samples from 384-well plate to four 96-well $0.2\ \text{mL}$ PCR plates.

Note: Perform steps 42–53 on each 96-well plate, keeping the other three plates on ice and covered.

52. Add $7.5\ \mu\text{L}$ AMPure XP beads to each well and mix by pipetting.
53. Incubate at 23°C for 8 min.
54. Place on a 96-well plate magnetic rack for 5 min or until solution is clear.
55. Remove supernatant and add $200\ \mu\text{L}$ of freshly prepared 80% ethanol.

Note: Perform this step using a multi-channel pipette on several rows (if using 12-channel pipette) or columns (if using 8-channel pipette) of wells at a time for efficiency but avoid letting the beads dry.

△ CRITICAL: Do not allow tips that have entered a sample well to enter any other sample well to avoid sample cross-contamination.

56. Incubate in ethanol for 30 s.
57. Remove supernatant and add 80% ethanol.
58. Using a multichannel P200 pipette, remove supernatant from one row. Then, use a P10 multi-channel pipette to remove as much ethanol as possible from the same row.

△ CRITICAL: Do not over-dry the beads, as it may affect yield. Change tip on aspirator between wells.

59. Using a multichannel pipette, add $17\ \mu\text{L}$ elution buffer and mix by pipetting.
60. Repeat steps 42–49 for all rows on plate. After elution buffer is added to the last row, incubate at 23°C for 5 min.

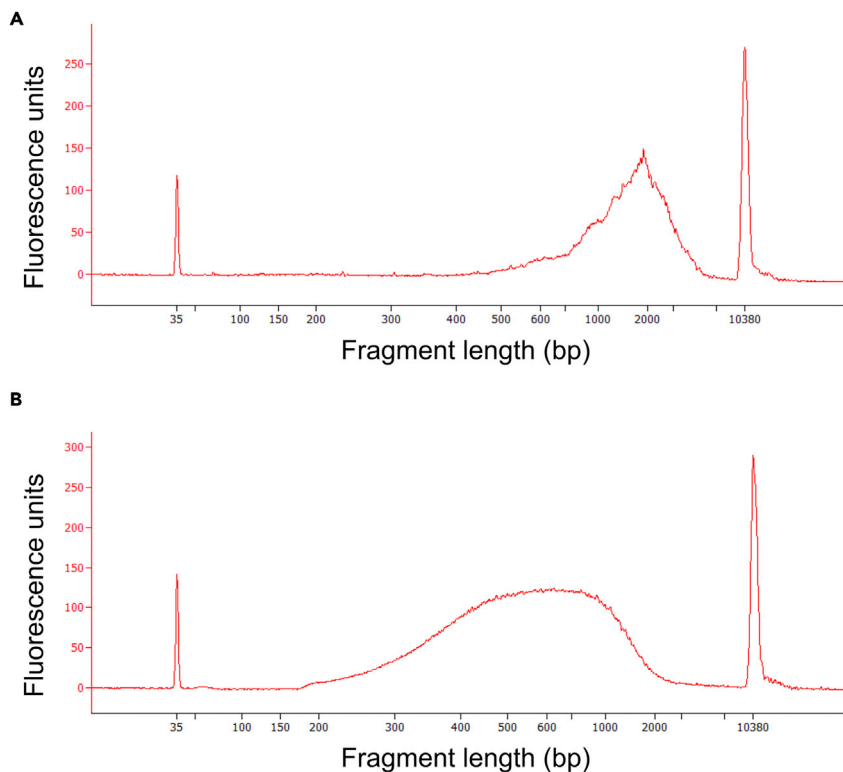


Figure 2. Assessment of fragment size distribution of cDNA and sequencing libraries

(A) cDNA libraries should have a broad distribution of fragment lengths with a peak around 1–2 kb.
(B) Final sequencing libraries should consist of an average fragment size of approximately 600 nt.

61. Place on a 96-well plate magnetic rack for 2 min or until supernatant is clear.
62. Transfer 15 μ L from each well to a new plate.

Pause point: Plates can be stored at -20°C . Plates should be covered with foil adhesive plate covers.

63. Use the Qubit 1 \times dsDNA HS Assay Kit to quantify the concentration of cDNA.

Note: A minimum concentration of 100 $\text{pg}/\mu\text{L}$ for each sample should be used. Concentrations below this should not be used as it may affect the efficiency of the following steps.

Note: There will be wells with insufficient material to continue with tagmentation. These wells should not be used. If using an adjustable-spacer pipette, cells can be easily transferred directly to a 384-well plate (step 61), skipping over these wells.

64. Verify the expected size distribution of cDNA fragments of several randomly selected cells using the Agilent High Sensitivity DNA Kit on the Agilent 2100 Bioanalyzer (Figure 2A).

Note: There should be an enrichment of fragments in the 1–2 kb range.

Tagmentation and final library preparation

Timing: 0.5 days for each 96-well plate

Here, final sequencing libraries are prepared from amplified cDNA. This protocol is designed to process each 96-well plate of purified cDNA from the previous section successively.

Note: The following steps are performed with the Illumina [Nextera XT Library Preparation Kit](#) using the protocol presented with the Takara SMART-seq HT Kit. The reaction volumes have been reduced 2-fold from those presented in the [Takara SMART-seq HT Kit instructions](#). 400 Custom Illumina-compatible CDI indexes were purchased from IDT. These dual indexes each consist of 2 indexing primers in the same tube at 10 μ M, and can be used in the protocol below as supplied. The sequences of each of these indexes can be found in [Table S1](#).

Note: Each 96-well plate of purified cDNA should be performed sequentially.

65. Dilute the cDNA in each well to a concentration of 200 pg/ μ L using elution buffer.

△ CRITICAL: The amount of cDNA in each reaction can influence the efficiency of the tagmentation reaction. Highly variable amounts of input cDNA can result in libraries with different average fragment sizes and result in some libraries outcompeting others in the sequencing reaction. If the concentration of cDNA is between 100–200 pg/ μ L do not dilute the sample.

66. As quickly as possible, use a multichannel pipette to add 1.875 μ L of the tagmentation premix (provided in Nextera XT Library Preparation Kit) to each well in top-left quadrant of a 384-well plate.

Note: Due to the scaled-down volumes used in this protocol, 384-well plates are used. 96-well plates may be substituted while using double or triple the indicated volumes. If using fewer than all 96-wells in the top-left quadrant due to dropout samples with too little cDNA, those wells can be ignored to save reagent.

67. Add 0.625 μ L of 200 pg/ μ L cDNA to each well.

68. Quickly seal the plate with an adhesive plate cover and place in a thermocycler for 10 min at 55°C. Hold at 10°C.

69. As soon as the samples reach 10°C, remove the plate and quickly add 0.625 μ L of NT Buffer (provided in Nextera XT Library Preparation Kit).

70. Seal the plate with an adhesive plate cover and centrifuge at 2,000 \times g for 1 min at 23°C.

71. Incubate at 23°C for 5 min.

72. Add 1.88 μ L of NPM master mix (provided in Nextera XT Library Preparation Kit) to each well.

73. Add 0.625 μ L of nuclease-free water.

74. Add 0.625 μ L of duplexed, Nextera-compatible indexing primers.

75. Seal the plate with an adhesive plate cover and incubate in a thermocycler with lid set to 105°C.

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	72°C	3 min	1
Initial denaturation	95°C	30 s	1
Denaturation	95°C	10 s	13 cycles
Annealing	55°C	30 s	
Extension	72°C	1 min	
Final extension	72°C	5 min	1
Hold	4°C	forever	

▮▮ Pause point: Plate can be stored at –20°C. Plates should be covered with foil adhesive plate covers.

76. Add 9.375 μL of nuclease-free water to each well (12.5 μL total) and transfer samples to a 96-well plate.

Note: Transferring samples between 384-well plates and 96-well plates is aided by using an electronic pipette with adjustable spacers.

77. Add 7.5 μL AMPure XP beads to each well and mix by pipetting.
78. Incubate at 23°C for 8 min.
79. Place on a 96-well plate magnetic rack for 5 min or until solution is clear.
80. Remove supernatant and add 200 μL of freshly prepared 80% ethanol.

Note: Perform this step using a multi-channel pipette on several rows (if using 12-well multi-channel) or columns (if using 8-well multichannel) of wells at a time for efficiency, but avoid letting the beads dry. Additionally, always change tips between samples and do not let tips contact the wells of two different samples at any step to avoid cross-contamination.

81. Incubate in ethanol for 30 s.
82. Remove supernatant and add 80% ethanol.
83. Using a multichannel p200 pipette, remove supernatant from one row. Then, use a p10 multi-channel pipette to remove as much ethanol as possible from the same row.

⚠ CRITICAL: Do not over-dry the beads, as it may affect yield. Change tip on aspirator between wells.

84. Using a multichannel pipette, add 17 μL elution buffer and mix by pipetting.
85. Repeat steps 65–74 for all rows on plate. After elution buffer is added to the last row, incubate at 23°C for 5 min.
86. Place on a 96-well plate magnetic rack for 2 min or until supernatant is clear.
87. Transfer 15 μL from each well to a new plate.

⏸ Pause point: Plates can be stored at -20°C . Plates should be covered with foil adhesive plate covers.

88. For all wells, use the Qubit 1 \times dsDNA HS Assay Kit to quantify the concentration of cDNA.
89. Verify expected size distribution of cDNA fragments using the Agilent High Sensitivity DNA Kit on the Agilent 2100 Bioanalyzer (Figure 2B).

Note: There should be an average fragment size of about 500–600 bp. At least a random selection of 12 libraries should be tested.

90. Pool samples for sequencing. All samples should be pooled to an equal molar representation in the final library.
91. The pooled sample should be quantified again using the Qubit 1 \times dsDNA HS Assay Kit and fragment sizes should be assessed using the Agilent High Sensitivity DNA Kit.

Note: Fragment sizes should be 500–600 bp and DNA concentration should meet the requirements of the sequencer and flow cell being used.

92. Perform next-generation Illumina sequencing.

Note: We recommend using 150 bp paired-end sequencing (2 \times 150 bp PE) as more sequence information is obtained for the same number of total reads, but using 2 \times 50 bp PE or 2 \times 100 bp PE could also be used. Additionally, the depth of sequencing can have a

strong impact on the number of sites detected and the quality of the dataset. Although a standard SMART-seq scRNA-seq experiment is usually performed with at most 1 million reads per cell, it is recommended to exceed that for this dataset. 5 million reads per cell was targeted in (Tegowski et al., 2022).

Note: We used a Novaseq6000 with S4 flow cells to perform the sequencing for this experiment. However, NextSeq Illumina sequencers can also be used. Ensuring enough coverage on a cell-to-cell basis with high sequencing quality is the most critical consideration for successful identification of m⁶A sites (see above note).

Read preprocessing, genome alignment, and filtering of low-quality cells

⌚ Timing: 3–5 days

After Illumina sequencing and demultiplexing, the sequencing reads should all be separate fastq files. In this section, sample quality will be assessed, so that substandard cells will be eliminated from further analysis and m⁶A identification steps. Adaptor sequences will be trimmed from reads, followed by alignment, PCR duplicate removal, and quantification of RNA expression in each cell to generate a gene expression matrix. This matrix will then be used to perform quality control steps using Seurat (Stuart et al., 2019).

Note: Raw data from Tegowski et al. (2022) are available on SRA (GSE180954), and all instructions and code for reproducing the identification of m⁶A methylation sites is available at: <https://github.com/tegowski/scDARTHEKcells>.

Note: Other than R scripts (where indicated), the below are all bash scripts intended to be submitted as jobs on systems with a Slurm workload scheduler from the directory containing the files being processed. All instances of “WORKDIR” should be replaced with the working directory being used for this project. All scripts below and how to use them are also available at the indicated Github link.

93. Use Flexbar to remove adaptor sequences.

```
#First move to correct directory

$ cd $WORKDIR/software

#Submit flexbar script

$ sbatch flexbar.sh

##Contents of flexbar.sh

#!/bin/bash

#SBATCH -a 1-1667

WORKDIR=''/your/path/here''

FASTQ=$WORKDIR/rawfiles #input directory

TRIM=$WORKDIR/trim #output directory

mkdir -p $TRIM

#This step will loop through all fastq files in the directory (adjust the numbers in the -a option
at the top if using a different number of files)

file=$(ls $FASTQ/*_1.fastq.gz | sed -n ${SLURM_ARRAY_TASK_ID}p)
```

```
STEM=$(basename "$file" _1.fastq.gz)

R1="$STEM"_1.fastq.gz # reassign full name to Read1

R2="$STEM"_2.fastq.gz # assign name to Read 2

#This is the actual flexbar command.

flexbar -r $FASTQ/$R1 -p FASTQ/$R2 -aa Nextera -zip-output GZ -threads 16 -qf i1.8 -t $TRIM/
$STEM
```

94. Use STARsolo to align to the genome, remove exact duplicates, and generate a gene count matrix for all cells.

Note: The APOBEC1-YTH cells (first script) should be run separately from the APOBEC1-YTH^{mut} cells (second script).

```
# Copy scripts to the correct directory, then move there

$ cp $WORKDIR/software/StarsoloalignWT.sh $WORKDIR/rawfiles/

$ cp $WORKDIR/software/StarsoloalignMUT.sh $WORKDIR/rawfiles/

$ cd $WORKDIR/rawfiles

#Submit scripts to align the genome

$ sbatch StarsoloalignWT.sh

$ sbatch StarsoloalignMUT.sh

#Contents of StarsoloalignWT.sh

#!/bin/bash

#SBATCH -a 1-2

#SBATCH -c 8

#SBATCH -mem=60G

#This script should be run from the $WORKDIR/trim directory

#Replace $WORKDIR with the path to your directory

WORKDIR="/your/path/here"

mkdir -p $WORKDIR/aligned

#Defines list of fastq files for each batch to be aligned

filelist=$(cat $WORKDIR/software/WTfilelist${SLURM_ARRAY_TASK_ID}.txt)

#Run STARsolo

STAR --genomeDir $WORKDIR/STARindex/ \

-runThreadN 8 \

-readFilesIn $filelist \

-outFileNamePrefix $WORKDIR/aligned/WT${SLURM_ARRAY_TASK_ID} \

-outSAMtype BAM Unsorted \

-outFilterMismatchNoverReadLmax 0.06 \

-outSAMattrRGline ID: \
```

```
-soloType SmartSeq \

-readFilesManifest $WORKDIR/software/WTmanifest${SLURM_ARRAY_TASK_ID}.txt \

-soloUMIdedup Exact \

-soloStrand Unstranded

#Contents of StarsoloalignMUT.sh

#!/bin/bash

#SBATCH -a 1-2

#SBATCH -c 8

#SBATCH -mem=60G

#This script should be run from the $WORKDIR/trim directory

#Replace $WORKDIR with the path to your directory

WORKDIR="/your/path/here"

mkdir -p $WORKDIR/aligned

#Defines list of fastq files for each batch to be aligned

filelist=$(cat $WORKDIR/software/mutfilelist${SLURM_ARRAY_TASK_ID}.txt)

#Run STARSolo

STAR --genomeDir $WORKDIR/STARindex/ \

-runThreadN 8 \

-readFilesIn $filelist \

-outFileNamePrefix $WORKDIR/aligned/MUT${SLURM_ARRAY_TASK_ID} \

-outSAMtype BAM Unsorted \

-outFilterMismatchNoverReadLmax 0.06 \

-outSAMattrRGline ID: \

-soloType SmartSeq \

-readFilesManifest $WORKDIR/software/mutmanifest${SLURM_ARRAY_TASK_ID}.txt \

-soloUMIdedup Exact \

-soloStrand Unstranded
```

Note: This step can require significant amounts of time, especially if many cells (more than 100–200) are being aligned and counted. It is advisable to break this step into multiple groups to generate multiple gene expression matrices, which can be merged in the next steps using Seurat. For this example, there are two groups of APOBEC1-YTH (WT) cells and two groups of APOBEC1-YTH^{mut} (MUT) cells for a total of 600 cells being aligned. For more detailed information, please see <https://github.com/tegowski/scDARTHEKcells>.

Note: Several other files will be required for alignment. First, several lists of all the file names should be copied from the Github page into the path "\$WORKDIR/software/". Second, a file manifest needs to be generated and also contained within the "\$WORKDIR/software/" directory. These files are available from Github. They are tab-separated files that contain the read 1 file name, the read 2 file name, and a cell ID that will be added by STARSolo. Third, an indexed reference genome will need to be generated using STAR, and it should be located within the

“\$WORKDIR/STARindex” directory. Detailed instructions for generating this index are provided above, in the “before you begin” section. If not using this dataset, detailed instructions on generating these required files can be found at <https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md>.

95. Next, the gene expression of the sequenced cells needs to be analyzed so that basic quality control steps can eliminate poor quality cells. To do this, start an R session and load required libraries for analysis using Seurat.

```
> library(Seurat)
> library(tidyverse)
> library(Matrix)
> library(scales)
> library(cowplot)
> library(RCurl)
```

96. Next, the output files from STARsolo need to be loaded and converted into Seurat objects.

Note: Each alignment will produce 3 files (barcodes.tsv, features.csv, and matrix.mtx), and all three of these files need to be in the same directory and zipped using gzip.

```
> for (file in c(`WT1`, `WT2`, `MUT1`, `MUT2`)) {
  seurat_data <- Read10X(data.dir = paste0(`$WORKDIR/aligned/`, file))
  seurat_obj <- CreateSeuratObject(counts = seurat_data, project = file)
  assign(file, seurat_obj)
}
> SeuratSMARTcombined <- merge(x = WT1, y = c(WT2, MUT1, MUT2))
```

97. Calculate key quality metrics: 1) Ratio of RNAs per cell to the reads per cell, 2) Proportion of mitochondrial reads.

Note: The pattern of “starts MT” (pattern = “^MT-”) in the code below is useful only for the human annotation of mitochondrial genes. This will need to be adjusted for other reference genomes (e.g., mouse should be “^mt-”).

```
> SeuratSMARTcombined$log10GenesPerRead <- log10(SeuratSMARTcombined$nFeature_RNA) /
log10(SeuratSMARTcombined$nCount_RNA)
> SeuratSMARTcombined$mitoRatio <- PercentageFeatureSet(object = SeuratSMARTcombined,
pattern = "^MT-")
> SeuratSMARTcombined$mitoRatio <- SeuratSMARTcombined@meta.data$mitoRatio / 100
> metadata <- SeuratSMARTcombined@meta.data
> metadata$cells <- rownames(metadata)
> SeuratSMARTcombined$log10GenesPerRead <- log10(SeuratSMARTcombined$nFeature_RNA) /
log10(SeuratSMARTcombined$nCount_RNA)
```

```
> SeuratSMARTcombined$mitoRatio <- PercentageFeatureSet(object = SeuratSMARTcombined,
pattern = "^MT-")

> SeuratSMARTcombined$mitoRatio <- SeuratSMARTcombined@meta.data$mitoRatio / 100

> metadata <- SeuratSMARTcombined@meta.data

> metadata$cells <- rownames(metadata)
```

98. Filter out low-quality cells.

Note: Cells are considered low-quality and removed if they meet any of the following criteria. 1) They have too few reads (cells with fewer than 1,000,000 total mapped reads to annotated genes are removed) 2) They have too few RNAs detected (cells with fewer than 9,000 RNAs (genes) identified are removed). 3) They have an altered ratio of RNAs identified to number of reads. A large increase in read numbers with no increase in RNAs detected can indicate doublets. (cells with a \log_{10} ratio of greater than 0.58 are removed). 4) They are not healthy or viable (cells with more than 10% of reads deriving from mitochondrial RNAs are removed).

```
> filtered_Seurat <- subset(x=SeuratSMARTcombined,
subset=(nCount_RNA >= 1000000) &
(nFeature_RNA >= 9000) &
(log10GenesPerRead > 0.58) &
(mitoRatio < 0.10))
```

△ CRITICAL: This is important to do before the identification of m⁶A sites, so that sites from dead cells or doublets can be ignored.

99. Generate quality control graphs showing the ratio of RNAs to reads per cell (Figure 3A) and proportion of mitochondrial reads (Figure 3B).

Note: A single peak is expected, and multiple peaks can indicate a high proportion of doublets that should be removed. The exact average ratio will depend on cell line used and sequencing saturation, but here only cells with a ratio of less than 0.58 are included in the following analyses.

```
> counts <- GetAssayData(object = filtered_Seurat, slot = "counts")
> nonzero <- counts > 0
> keep_genes <- Matrix::rowSums(nonzero) >= 10
> filtered_counts <- counts[keep_genes, ]
> filtered_Seurat <- CreateSeuratObject(filtered_counts, meta.data = filtered_Seurat
@meta.data)
> metadata_clean <- filtered_Seurat@meta.data
> metadata_clean %>%
  ggplot(aes(x=log10GenesPerUMI, color=sample, fill=sample)) +
  geom_density(alpha = 0.2) +
  theme_classic() +
```

```
geom_vline(xintercept = 0.8)
> metadata_clean %>%
  ggplot(aes(x=mitoRatio, color=sample, fill=sample)) +
  geom_density(alpha = 0.2) +
  scale_x_log10() +
  theme_classic() +
  geom_vline(xintercept = 0.2)
```

100. Obtain a list of cell IDs from the filtered dataset. This will be used after m⁶A site identification to remove sites found in low-quality cells after m⁶A identification in the following steps.

```
> filtered_Seurat@meta.data$cells <- row.names(filtered_Seurat@meta.data)
> FilteredCells <- filtered_Seurat@meta.data$cells
> write.table(FilteredCells, file="$WORKDIR/aligned/QCpasscells.txt", row.names=F, col.
names=T, sep="\t", quote=F)
```

Identification of m⁶A sites using Bullseye

⌚ Timing: 1 day

Here, adapter trimmed reads in fastq files (from step 82) will be aligned to the genome, but all APOBEC1-YTH-expressing cells will be kept as separate files. The aligned sequences will then be parsed to generate matrices containing all the basecall information at each mapped nucleotide position. These matrices will then be used to identify the C-to-U mutations that indicate methylation. Finally, poor quality sites will be filtered to generate a list of high-confidence sites.

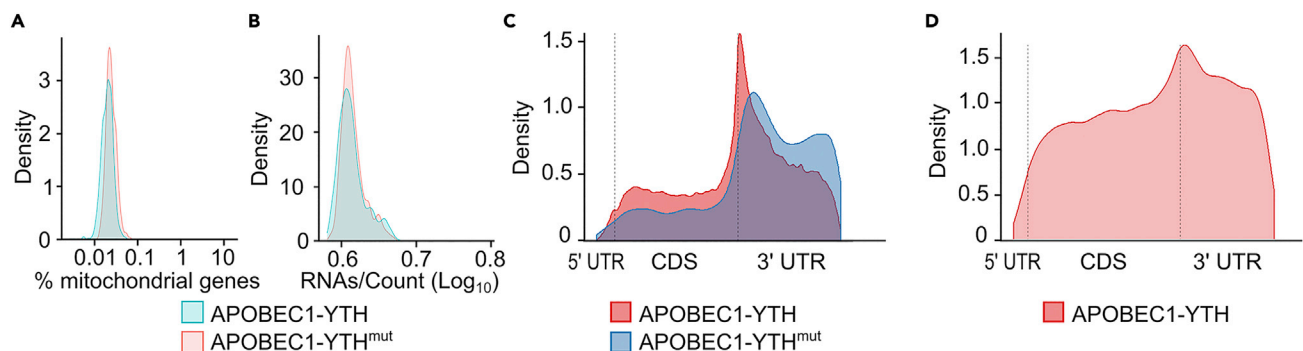


Figure 3. Quality control of sequenced cells and m⁶A sites identified

(A) Density plot showing the percentage of reads aligning to mitochondrial-encoded mRNAs from each cell. The average is less than 5%, indicating healthy and viable cells were sequenced.

(B) Density plot showing the ratio of the number of RNAs to the number of reads per cell (log₁₀ of each). The single distribution suggests the cells are primarily singletons.

(C) Metagene plot showing the distribution of identified sites along mRNA transcripts.

(D) Metagene plot showing improper distribution of identified sites. Poor stop codon enrichment would indicate that these sites do not accurately reflect m⁶A methylation. Alternative thresholds or quality control metrics should be tested. Figures reprinted with permission from (Tegowski et al., 2022).

Note: Raw data from Tegowski et al. (2022) are available on SRA (GSE180954), and all instructions and code for reproducing the identification of m⁶A methylation sites described here is available at <https://github.com/tegowski/scDARTHEKcells>.

Note: For detailed information on the installation and general use of the Bullseye pipeline, please also see <https://github.com/mflamand/Bullseye>.

101. Align the reads to the genome using STAR. The fastq files from each APOBEC1-YTH and APOBEC1-YTH^{mut} cell should be aligned separately to produce an independent bamfile.

```
# Copy scripts to the correct directory, then move there
$ cp $WORKDIR/software/StarsoloalignWT.sh $WORKDIR/rawfiles/
$ cp $WORKDIR/software/StarsoloalignMUT.sh $WORKDIR/rawfiles/
$ cd $WORKDIR/rawfiles

#Submit scripts to align to the genome
$ sbatch STARBullseyealignWT.sh
$ sbatch STARBullseyealignMUT.sh

#Contents of STARBullseyeWT.sh

#!/bin/bash

#SBATCH -a 1-601

#SBATCH -c 12

#SBATCH -mem=60G

#Define path to your working directory
WORKDIR="/work/mrt41/Ex96_SMARTseq/test/"

file=$(ls *_1.fastq | sed -n ${SLURM_ARRAY_TASK_ID}p) ## this is done for paired end libraries
ending with _1 and _2 for each pair.

STEM=$(basename "$file" *_1.fastq)

BAM=$WORKDIR/singlebams

mkdir -p $BAM

echo ``aligning $STEM"

STAR --runMode alignReads \
-genomeDir $WORKDIR/STARindex \
-runThreadN 12 \
-readFilesIn ${STEM}_1.fastq ${STEM}_2.fastq \
-outSAMtype BAM SortedByCoordinate \
-outFilterMismatchNoverReadLmax 0.06 \
-outFileNamePrefix $BAM/${STEM}

#Contents of STARBullseyeMUT.sh

#!/bin/bash

#SBATCH -a 1-601
```

```
#SBATCH -c 12

#SBATCH -mem=60G

#Define path to your working directory
WORKDIR="/work/mrt41/Ex96_SMARTseq/test/"

file=$(ls *_1.fastq | sed -n ${SLURM_ARRAY_TASK_ID}p) ## this is done for paired end libraries
ending with _1 and _2 for each pair.

STEM=$(basename "$file" *_1.fastq)

BAM=$WORKDIR/singlebams

mkdir -p $BAM

echo ``aligning $STEM"

STAR -runMode alignReads \
-genomeDir $WORKDIR/STARindex \
-runThreadN 12 \
-readFilesIn ${STEM}_1.fastq ${STEM}_2.fastq \
-outSAMtype BAM SortedByCoordinate \
-outFilterMismatchNoverReadLmax 0.06 \
-outFileNamePrefix $BAM/${STEM}
```

Note: When analyzing many cells, it is recommended to submit this script as an array so that many cells can be aligned at once.

102. Remove duplicates from APOBEC1-YTH^{mut}.

```
cp $WORKDIR/software/RemoveduplicatesMUT.sh $WORKDIR/singlebams

cd $WORKDIR/singlebams

sbatch RemoveduplicatesMUT.sh

#Contents of RemoveduplicatesMUT.sh

#!/bin/bash

#SBATCH -p scavenger

#SBATCH -c 8

#SBATCH -mem=60G

#SBATCH -a 1-300

file=$(ls MUT*Aligned.sortedByCoord.out.bam | sed -n ${SLURM_ARRAY_TASK_ID}p)

STEM=$(basename "$file" .Aligned.sortedByCoord.out.bam)

echo "${SLURM_ARRAY_TASK_ID}"

echo ``Starting File: $file"

#First, use samtools to remove optical duplicates

samtools sort -n -o $STEM.name.bam $file

samtools fixmate -r -m $STEM.name.bam $STEM.fix.bam
```

```
samtools sort -o $STEM.possort.bam $STEM.fix.bam

samtools markdup -r $STEM.possort.bam $STEM.duprm.bam

rm $STEM.name.bam $STEM.fix.bam $STEM.possort.bam
```

103. Merge all APOBEC1-YTH^{mut} bamfiles and produce nucleotide representation matrix.

```
cp $WORKDIR/software/MakematrixYTHmut.sh $WORKDIR/singlebams

cd $WORKDIR/singlebams

sbatch Make_matrixYTHmut.sh

#Contents of Make_matrixYTHmut.sh

#!/bin/bash

#SBATCH --mem=60G

#SBATCH --c 12

#SBATCH --p scavenger

#Merge all MUT bamfiles

samtools merge MUTCellsmerged.bam MUT*duprm.bam

#Then use Bullseye (parseBAM.pl) to make nucleotide matrix

WORKDIR="/work/mrt41/Ex96_SMARTseq/test"

BAMS=$WORKDIR/singlebams

STEM=$(basename "$file" Aligned.sortedByCoord.out.bam.duprm.bam)

mkdir $WORKDIR/matrix

perl $WORKDIR/software/parseBAM.pl \

-i MUTCellsmerged.bam \

-o $WORKDIR/matrix/MUTCellsmerged.matrix \

-minCoverage 5 \

-verbose
```

Note: This step produces a matrix of all mapped residues with coverage. It lists the number of reads in which each genomic coordinate is found to be an A,U,G,C, or N. Additionally, it lists the total coverage. Matrixes will be made from APOBEC1-YTH cells (next step) as well and will be used to determine C-to-U mutation rate and find sites in APOBEC1-YTH cells with higher mutation rates than found in APOBEC1-YTH^{mut} control cells.

Note: This step generates a population average of all the APOBEC1-YTH^{mut}-expressing cells. The C-to-U mutation rate of this average in the population will be used as a control. Sites in which the C-to-U mutation rate (%C2U) is not higher than the APOBEC1-YTH^{mut}-expressing cells will be filtered out.

104. Parse the aligned sequences from each of the APOBEC1-YTH-expressing cells to generate a nucleotide representation matrix for each cell.

Note: The first step in this script removes duplicate reads from the bamfiles.

```
cp $WORKDIR/software/MakematrixYTH.sh $WORKDIR/singlebams
cd $WORKDIR/singlebams
sbatch MakematrixYTH.sh
#Contents of MakematrixYTH.sh
#!/bin/bash
#SBATCH -p scavenger
#SBATCH -c 8
#SBATCH -mem=60G
#SBATCH -a 1-301
file=$(ls WT*Aligned.sortedByCoord.out.bam | sed -n ${SLURM_ARRAY_TASK_ID}p)
STEM=$(basename "$file" .Aligned.sortedByCoord.out.bam)
echo "${SLURM_ARRAY_TASK_ID}"
echo "Starting File: $file"
#First, use samtools to remove optical duplicates
samtools sort -n -o $STEM.name.bam $file
samtools fixmate -r -m $STEM.name.bam $STEM.fix.bam
samtools sort -o $STEM.possort.bam $STEM.fix.bam
samtools markdup -r $STEM.possort.bam $STEM.duprm.bam
rm $STEM.name.bam $STEM.fix.bam $STEM.possort.bam
#Then use Bullseye (parseBAM.pl) to make nucleotide matrix
WORKDIR="/your/path/here"
BAMS=$WORKDIR/singlebams
file=$(ls WT*duprm.bam | sed -n ${SLURM_ARRAY_TASK_ID}p)
STEM=$(basename "$file" .duprm.bam)
mkdir $WORKDIR/matrix
perl $WORKDIR/software/parseBAM.pl \
-i $file \
-o $WORKDIR/matrix/$STEM.matrix \
-minCoverage 5 \
-verbose \
-removeDuplicates
```

Note: When analyzing many cells, it is recommended to perform this as an array so that many cells can be aligned at once. This step first removes optical duplicates from the bamfiles before making the nucleotide representation matrices.

105. Identify C-to-U mutations that indicate m⁶A methylation.

```
cp $WORKDIR/software/find_RNA_edit_sites.sh $WORKDIR/matrix
cd $WORKDIR/matrix
sbatch find_RNA_edit_sites.sh
#Contents of find_RNA_edit_sites.sh
#!/bin/bash
#SBATCH --mem=24G
#SBATCH --p scavenger
#SBATCH --c 8
#SBATCH --a 1-301
WORKDIR="/your/path/here/"
annotation_file=$WORKDIR/software/hg38refFlat_nochr.gtf
bed=$WORKDIR/bed
file=$(ls WT*matrix.gz | sed -n ${SLURM_ARRAY_TASK_ID}p)
STEM=$(basename "$file" .matrix.gz)
mkdir -p $bed
perl $WORKDIR/software/Find_edit_site.pl \
  -annotationFile $annotation_file \
  -EditedMatrix $file \
  -controlMatrix MUTCellsmerged.matrix.gz \
  -editType C2U \
  -minEdit 10 \
  -maxEdit 95 \
  -editFoldThreshold 1.5 \
  -EditedMinCoverage 20 \
  -ControlMinCoverage 10 \
  -MinEditSites 2 \
  -cpu 8 \
  -outfile $bed/$STEM.bed \
  -bed6 \
  -verbose
```

△ CRITICAL: Thresholds are used during this step to eliminate low quality editing events that are likely to be sequencing errors or APOBEC1-driven editing independent of YTH-mediated m⁶A binding. The criteria used here are: a minimum of 20 reads of coverage, a minimum of 10% C2U editing within the cell, and at least 1.5-fold higher %C2U than that observed in the APOBEC1-YTH^{mut} population average.

Note: A number of different thresholds have been tested, but these were shown to produce a list of sites that had important features. 1) An m⁶A-like metagene profile and 2) strong overlap of methylated RNAs with previously studies. However, if using a different dataset these

thresholds may need to be adjusted. Features like metagene analysis and overlap with existing datasets should serve as important factors for determining the validity of the sites identified.

Note: The annotation file should be in a refFlat format. This specific annotation file is available at the GitHub page <https://github.com/tegowski/scDARTHEKcells> and should be downloaded to the "\$WORKDIR/software/" directory. Detailed instructions for generating this annotation file for any organism are available at <https://github.com/mflamand/Bullseye>.

106. This step adds the cell ID (obtained from the file name) as the last column in each bedfile.

Note: All the potential sites for each cell are listed in separate bedfiles and they need to be combined into a single file. However, the cell ID (e.g., WTCell1) needs to be associated with each site before they are combined in order to keep track of which sites were identified within each cell.

```
cp $WORKDIR/software/addCellIDs.sh $WORKDIR/bed
cd $WORKDIR/bed
sbatch addCellIDs.sh
#Contents of addCellIDs.sh
#!/bin/bash
#SBATCH --mem=4G
for i in $(ls WTCell*.bed);
do
echo "`processing $i"
awk '{print $1"\t"$2"\t"$3"\t"FILENAME}' $i > $i.modbed
cut -f 4 $i.modbed > $i.Ids.txt
awk -F '[_]' 'OFS="\t" {print $2$3$4}' $i.Ids.txt > $i.Ids2.txt
paste $i $i.Ids2.txt > 2_$i.bed
rm $i.modbed $i.Ids.txt $i.Ids2.txt;
done
```

107. Then combine all the sites from each cell into one master bedfile.

```
$cat 2_WTCell*.bed > Sitesinallcells.bed
```

108. Eliminate any remaining sites that do not occur in the core R-A-C motif.

Note: This filtering step will enrich for site-specific editing events adjacent to the methylated adenosine at the expense of losing some non-adjacent sites in the vicinity of the methylated site.

```
cp $WORKDIR/software/RACfilter.sh $WORKDIR/bed
cd $WORKDIR/bed
sbatch RACfilter.sh
```

```
#Contents of RACfilter.sh

#!/bin/bash

#SBATCH -p scavenger

#SBATCH -mem=50G

#add 2nts upstream of C to T site if + strand, if - add downstream
awk '{ if ($6 == "+")
print $1"\t"$2-2"\t"$3"\t"$4"\t"$5"\t"$6;
else
print $1"\t"$2"\t"$3+2"\t"$4"\t"$5"\t"$6
}' $1.bed > $1.3nt.bed

echo -e '\n2nts added 5-prime of C:'
head -5 $1.3nt.bed

#get sequence at those sites
bedtools getfasta -bedOut -fo $1.3nt.sequence.bed -s -fi /datacommons/meyerlab/userdata/
matt_tego/genomes/hg38/Homo_sapiens.GRCh38.dna_sm.primary_assembly.fasta -bed $1.3nt.
bed

#Add sequence to tab
paste $1.bed $1.3nt.sequence.bed > $1.3ntwithseq.bed

echo -e '\n3mer context of edited C:'
head $1.3ntwithseq.bed

#retrieve lines that have RAC motif from pos file
awk 'BEGIN{IGNORECASE=1}{if($13!~"GAC|AAC") {print $0}}' $1.3ntwithseq.bed > $1.3nt.RAC.
bed

awk 'BEGIN{IGNORECASE=1}{if($13!~"GAC|AAC") {print $0}}' $1.3ntwithseq.bed > $1.3nt.
nonRAC.bed

rm $1.3nt.sequence.bed $1.nochr.bed $1.3nt.bed $1.3ntwithseq.bed
```

109. Eliminate sites that were previously found to be edited in HEK293T cells by overexpression of APOBEC1 alone ([Meyer, 2019](#)).

Note: This is an Rscript that compares sites found in the bedfile from the previous step with sites found when APOBEC1 was overexpressed in HEK293T cells ([Meyer, 2019](#)).

```
#First start R within the $WORKDIR/bed directory

$ cd $WORKDIR/bed

$ R

> #Then run R script that filters

> library(tidyverse)

> apoonly <- read.table('$WORKDIR/software/DARTseq_APOonly_hg38.bed', col.names=c
('`chr`, "start", "end"))
```

```
> head(apoonly)

> dartsites <- read.table(`Sitesinallcells.3nt.RAC.bed`)

> head(dartsites)

> #Add position column to compare with

> apoonly <- apoonly %>% mutate(pos = paste0(chr, "-", start))

> dartsites <- dartsites %>% mutate(pos = paste0(V1, "-", V2))

> #Filter dartsites for only those that don't appear in apoonly

> dartsitesfilt <- dartsites %>% filter(!pos %in% apoonly$pos)

> write.table(dartsitesfilt, file="Sitesinallcells_APOfilt.bed", col.names=F, row.names=F,
quote=F, sep="\t")
```

Note: If performing this experiment in cells other than HEK293T, a dataset of C-to-U editing events in cells expressing APOBEC1 alone may not be available. This is included for optimal high quality m⁶A identification, but it only eliminates a small fraction of sites (~1%). Therefore, this step is optional and high-quality sites can be determined without such a filtering step.

110. This step utilizes an R script to remove sites from cells that did not pass the quality control steps when analyzing the cells using Seurat (steps 91–96). Additionally, this script will also only keep sites found in at least 10 cells.

```
library(tidyverse)

> sitelist <- read.table(`bed/Sitesinallcells_APOfilt.bed`)

> FilteredCells <- read.table(`aligned/QCpasscells.txt`, header=T)

#Filter out sites in low quality cells

> sitelist2 <- sitelist %>% filter(V12 %in% FilteredCells$x)

#Now find the number of cells each site is in and filter out those in < 10 cells

> ncells <- sitelist2 %>% group_by(V14) %>% count()

> sitelist3 <- merge(sitelist2, ncells, by="V14")

> sitelist3 <- sitelist3 %>% filter(n > 9)

> write.table(sitelist2, file="/work/mrt41/Ex96_SMARTseq/test/bed/Finalsitelist.bed",
row.names=F, col.names=T, sep="\t", quote=F)
```

Note: This step is done to reduce the likelihood of calling sites that were found in very few cells as these may be likely to be noise or sequencing errors. These parameters can be optimized (Both cell quality parameters in steps 91–96 and the minimum number of cells a site must be found in). Although we found that these worked well in this case to identify m⁶A sites, optimal parameters may be affected by conditions like sequencing saturation, number of cells sequenced, and level of transgene expression. Therefore, if using another dataset, these thresholds may need to be adjusted.

At the end of these steps, a file containing 513,564 total sites (25,896 unique methylated site locations) should be obtained (Table S2). This file will contain the genomic coordinates, annotated gene name, % C-to-U editing, and cell ID, among other information that can be used in downstream analysis.

EXPECTED OUTCOMES

Individual cells from stable cell lines generated from clonal selection should exhibit APOBEC1-YTH expression as shown in [Figure 1A](#). Additionally, Sanger sequencing of regions surrounding known sites should reveal detectable C-to-U mutations ([Figure 1B](#)). After inducible cell lines are generated, single-cell suspensions with high viability and low debris are expected for optimal sorting of single cells into each well of 384-well PCR plates.

Most wells should have enough product after cDNA amplification to continue with library preparation (at least 100 pg/μL). The cDNA should show the expected size distribution ([Figure 2A](#)). Additionally, after final libraries are prepared, an average fragment size of about 500–600 bp is expected ([Figure 2B](#)).

A high number of cells are expected to pass the quality control thresholds used ([Figures 3A and 3B](#)). This includes a low percentage of reads deriving from mitochondrial mRNAs ([Figure 3A](#)). Further, the ratio of RNAs detected to number of reads per cell should produce a single distribution ([Figure 3B](#)). Bimodal or multimodal distributions may suggest high numbers of doublets.

The identified sites should show a characteristic m⁶A-like distribution across mRNA transcripts, with a strong enrichment in proximal 3'UTRs in the vicinity of the stop codon. [Figure 3C](#) demonstrates the expected m⁶A-like metagene. On the other hand, [Figure 3D](#) shows an improper metagene that would indicate poor quality site calling. Factors that may influence this are low sequencing coverage, low transgene expression, or improper selection of site-selection criteria. Many more m⁶A sites are expected to be identified using scDART-seq than reported using bulk m⁶A profiling. Sequencing 1,068 cells uncovered over 120,000 m⁶A sites in HEK293T cells ([Tegowski et al., 2022](#)). Reproducing this dataset with the scripts provided should result in the identification of over 513,000 sites in 300 APOBEC1-YTH cells. This number includes the same individual site that could be methylated in many cells, and over 25,000 unique methylated sites are identified. Although a much larger number of sites are identified using scDART-seq than observed with bulk m⁶A identification, most of the sites are present in a small percentage of cells within the population in HEK293T cells.

LIMITATIONS

Currently, scDART-seq requires the expression of the APOBEC1-YTH transgene within cells to detect sites of m⁶A methylation. Transient induction for 24 h or less in HEK293T cells does not impact cell health or viability, so this strategy can likely be adapted to many different cell types. One limitation of the present protocol is the generation of clonal cell lines. This was done so that the cells sequenced would have as little variability in transgene expression as possible. However, this process is time consuming. Additionally, if a cell line of interest consists of several subpopulations (an ideal application of scDART-seq) whose methylation needs to be compared, generating clonal lines would likely eliminate that desired variation. Instead, transient transfection of the plasmids used to generate viruses in this study can be used to express the transgene. While much easier and faster, transgene expression may vary widely, resulting in differing editing efficiencies between cells. Further, many cells may not express the transgene at all. In this case, the cells should be sorted for those expressing a very narrow range of GFP expression to help limit cell-to-cell transgene variability.

Although the protocol here outlines the use of SMART-seq2-based library preparation, scDART-seq is also compatible with the Chromium 3' sequencing kit (10× Genomics) ([Tegowski et al., 2022](#)). However, since a vast majority of reads map to the extreme 3' end of transcripts using that platform, most sites more than a few hundred nucleotides from the transcription stop site are undetectable in a majority of cells. While SMART-seq2 is more laborious and expensive, usually resulting in far fewer cells sequenced, a more comprehensive map of m⁶A methylation is possible since the entire transcript sequence is obtained. Furthermore, while using a SMART-seq2-based approach to sequence

many cells allows for the analysis of whole-transcriptome sequencing of single-cells, it lacks unique molecular identifiers (UMIs) and is prone to batch effects.

In order to avoid strong batch effects, care should be taken not to over amplify cDNA libraries and process samples as quickly and uniformly as possible. In addition, processing as many samples as possible at the same time will help limit batch effects. Importantly, sequencing on as large of a flow cell as possible, maximizes the number of samples that are sequenced in the same sequencing reaction, which can reduce batch effects. The result of strong batch effects can be observed when performing clustering analysis. Cells from independent replicates may form unique and separate clusters. However, these effects can be minimized when analyzing scRNA-seq data using Seurat by performing canonical correlation analysis (CCA)-based integration of all samples before clustering. This process identifies “anchor genes” within cells that are in the same biological state, allowing for the correction of gene expression values between samples (Stuart et al., 2019).

TROUBLESHOOTING

Problem 1

100% of cells do not express the APOBEC1-YTH transgene, as observed using GFP fluorescence (step 30).

Potential solution

The cell line selected is potentially not a true clonal cell line and a different clone should be selected. Alternatively, cells lacking transgene expression can be eliminated during sorting.

Problem 2

No editing is observed at expected locations (step 40).

Potential solution

Ensure that the APOBEC1-YTH transgene is expressed using western blot with an anti-HA antibody. If the transgene is expressed, make sure RNA has been treated with DNaseI before cDNA synthesis. Contaminating DNA can be efficiently amplified and sequenced, obscuring APOBEC1-YTH-mediated editing of RNA.

Problem 3

Low viability and/or high debris is observed during flow sorting of individual cells (step 45).

Potential solution

Use wide-bore pipette tips when resuspending or transferring cells, and always resuspend cells gently. Do not centrifuge cells at speeds greater than 1,000 × g.

Problem 4

Incorrect/unexpected fragment size distribution of amplified cDNA or final libraries.

Potential solution

Decrease PCR cycles to reduce likelihood of overamplification of products.

Problem 5

Low yields of amplified cDNA or final libraries (step 64 or step 89).

Potential solution

Amplified cDNA or final libraries with low yields should be discarded. If the concentration of the final pooled libraries for sequencing is too low, DNA can be bound to AMPure XP beads and eluted into a lower volume of water in order to concentrate the samples.

Problem 6

Error message when using Bullseye to parse bamfiles. (Can't locate MCE/Loop.pm in @INC) (step 103)

Potential solution

Ensure all required programs and Perl modules are installed and available in your path.

Problem 7

Many cells with very high numbers of reads, or a bimodal distribution of the number of RNAs to reads detected (step 99).

Potential solution

This would indicate the presence of many cell doublets. Ensure that cells are fully trypsinized and add DNaseI to the sorting buffer to prevent cell clumps from forming. If clumps are observed during cell counting, do not continue.

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kate D. Meyer (kate.meyer@duke.edu).

Materials availability

This study did not generate any new unique reagents.

Data and code availability

The published article (Tegowski et al., 2022) includes all datasets analyzed in this study, as well as the code for the Bullseye pipeline. The datasets are available on GEO Datasets using the accession code GEO Datasets: GSE180954. Bullseye code is available on GitHub: <https://github.com/mflamand/Bullseye>.

All other code used in this study for the analysis of scDART-seq data is available on GitHub: <https://doi.org/10.5281/zenodo.6869039>.

Data tables relating to the sequences of index primers used in the study and the m⁶A sites identified by following the indicated protocol are available at Mendeley Data: <https://doi.org/10.17632/96cpc2p6x6.1>.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xpro.2022.101646>.

ACKNOWLEDGMENTS

We thank the Duke Center for Genomic and Computational Biology and the Duke Human Vaccine Institute Flow Cytometry Core for providing the infrastructure for sequencing and the 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. conceived of the study, designed experiments, and interpreted data. M.T. performed the data analysis. M.T. and K.D.M. wrote the manuscript.

DECLARATION OF INTERESTS

K.D.M. has filed a pending patent application for the use of DART-seq technology with Duke University.

REFERENCES

- 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.
- 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.
- 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.
- Meyer, Kate D. (2019). DART-seq: an antibody-free method for global m⁶A detection. *Nature Methods* 16, 1275–1280.
- 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.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Roehr, J.T., Dieterich, C., and Reinert, K. (2017). Flexbar 3.0 – SIMD and multicore parallelization. *Bioinformatics* 33, 2941–2942.
- 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.
- Tegowski, M., Flamand, M.N., and Meyer, K.D. (2022). scDART-seq reveals distinct m⁶A signatures and mRNA methylation heterogeneity in single cells. *Mol. Cell* 82, 868–878.e10.