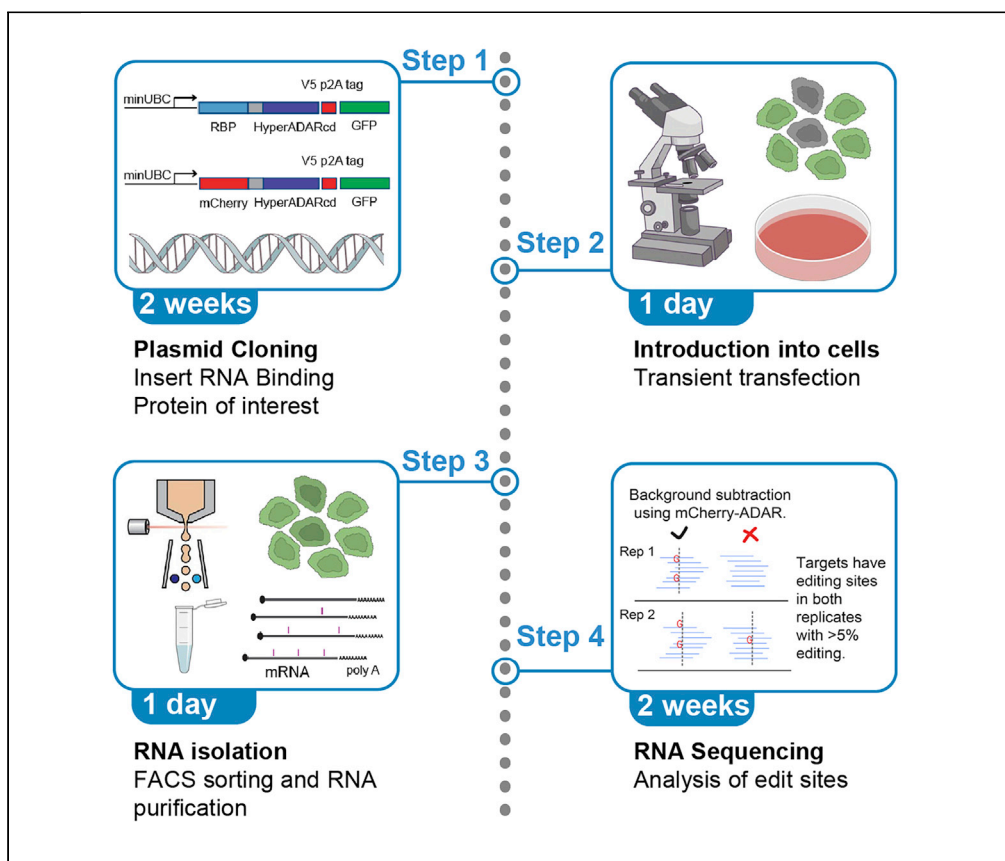


## Protocol

# Protocol for using TRIBES to study RNA-protein interactions and nuclear organization in mammalian cells



Targets of RNA-binding proteins discovered by editing (TRIBES) determines RNA-protein interactions and nuclear organization with minimal false positives. We detail necessary steps for performing mammalian cell RBP-TRIBES to determine the targets of RNA-binding proteins and MS2-TRIBES to determine RNA-RNA interactions within the nucleus. Necessary steps for performing a TRIBES experiment are detailed, starting with plasmid/cell line generation, cellular transfection, and RNA sequencing library preparation and concluding with bioinformatics analysis of RNA editing sites and identification of target RNAs.

Jeetayu Biswas,  
Michael Rosbash,  
Robert H. Singer,  
Reazur Rahman

jeetayu.biswas@alumni.  
einsteinmed.org (J.B.)  
reazur@brandeis.edu  
(R.R.)

**Highlights**  
TRIBES and  
HyperTRIBES can be  
readily used in  
mammalian systems

HyperTRIBES has  
minimal false  
positives and false  
negatives

The analysis pipeline  
has been made more  
accessible using a  
virtual machine

## Protocol

## Protocol for using TRIBE to study RNA-protein interactions and nuclear organization in mammalian cells

Jeetayu Biswas,<sup>1,5,6,\*</sup> Michael Rosbash,<sup>2</sup> Robert H. Singer,<sup>1,3,4</sup> and Reazur Rahman<sup>2,\*</sup><sup>1</sup>Department of Anatomy and Structural Biology, Bronx, NY 10461, USA<sup>2</sup>Department of Biology, Howard Hughes Medical Institute and National Center for Behavioral Genomics, Brandeis University, Waltham, MA 02451, USA<sup>3</sup>Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA<sup>4</sup>Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA 20147, USA<sup>5</sup>Technical contact<sup>6</sup>Lead contact\*Correspondence: [jeetayu.biswas@alumni.einsteinmed.org](mailto:jeetayu.biswas@alumni.einsteinmed.org) (J.B.), [reazur@brandeis.edu](mailto:reazur@brandeis.edu) (R.R.)  
<https://doi.org/10.1016/j.xpro.2021.100634>

## SUMMARY

**Targets of RNA-binding proteins discovered by editing (TRIBE) determines RNA-proteins interactions and nuclear organization with minimal false positives. We detail necessary steps for performing mammalian cell RBP-TRIBE to determine the targets of RNA-binding proteins and MS2-TRIBE to determine RNA-RNA interactions within the nucleus. Necessary steps for performing a TRIBE experiment are detailed, starting with plasmid/cell line generation, cellular transfection, and RNA sequencing library preparation and concluding with bioinformatics analysis of RNA editing sites and identification of target RNAs.**

**For complete details on the use and execution of this protocol, please refer to Biswas et al. (2020).**

## BEFORE YOU BEGIN

1. Clone RBP of interest into TRIBE backbone plasmid and then confirm expression.
2. Identify cell lines of interest for RBP-ADAR expression or prepare MS2 integrated cell line for MCP-ADAR expression.
3. Install and compile dependencies for computational analysis or utilized prepared virtual machine (Zenodo: <https://doi.org/10.5281/zenodo.4567690>).

## Clone RNA-binding protein of interest into TRIBE plasmid backbone

⌚ Timing: 1–2 weeks

4. Obtain the necessary, hyperactive ADAR containing, TRIBE plasmids.
  - a. A control, mammalian expression plasmid, containing mCherry ADAR control and p2A GFP reporter can be obtained from Addgene (Plasmid #154786).
  - b. For RBP-TRIBE, PCR amplify the RBP of interest from cDNA library, synthesize sequence of interest, or purchase plasmid containing RBP from Addgene.
  - c. For MS2-TRIBE, the MCP-ADAR plasmid can be obtained from Addgene (Plasmid # 154787).



After choosing a cell line of interest, the following samples should be generated:

**mCherry ADAR Control**

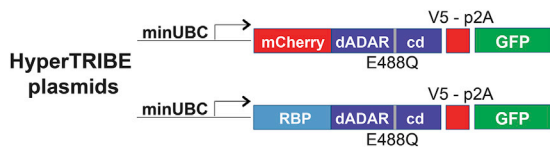
- Generate duplicate libraries
- Used to background subtract editing from ADAR catalytic domain alone

**RBP-ADAR Experiment**

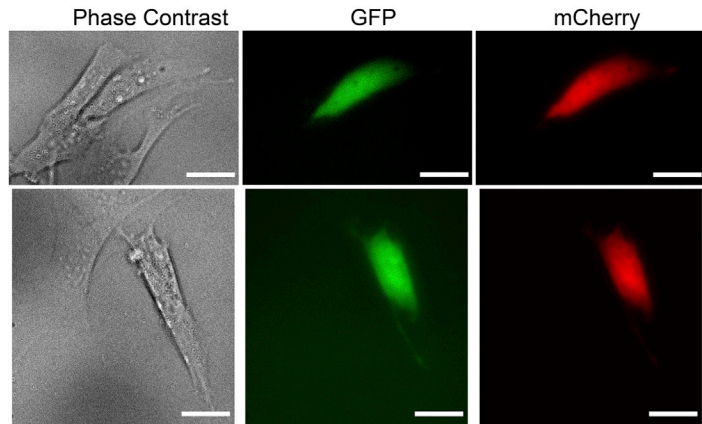
- Generate duplicate libraries
- Used to determine targets of RBP

**Untransfected cells or no ADAR control**

- Generate single library
- Used to confirm endogenous editing and mCherry ADAR editing levels



**Validate Transfection of mCherry and RBP-ADAR plasmids**



**Figure 1. Experimental design and validation of TRIBE constructs**

Top Left - Samples required and example of transient transfection to validate expression of TRIBE constructs. Bottom Left - Control and experimental plasmid design, highlighting features of the mammalian HyperTRIBE plasmids. Right - microscope images of cells transfected with mCherry-ADAR control plasmid to validate expression and calculate transfection efficiency. Scale bar is 200 pixels wide (35µm).

**Note:** Hyperactive ADAR is highly recommended (Kuttan and Bass, 2012; Xu et al., 2018; Herzog et al., 2020; Nguyen et al., 2020) to minimize the false negative rate and detect editing events with higher efficiency. This will also reduce the minimum sequencing depth required per sample.

- Clone RBP of interest into TRIBE plasmid using standard restriction enzyme-based cloning or Gibson assembly.
  - mCherry ADAR has NotI and BamHI sites flanking mCherry
  - MCP ADAR has NotI and KpnI sites flanking MCP
- Confirm proper insertion and frame by Sanger sequencing.
- Prepare high quality, transfection grade plasmid using an appropriate midiprep or maxiprep kit.
- After transient transfection into target cells of interest, confirm expression of RBP-ADAR fusion construct by GFP fluorescence (Figure 1). Perform western blot for the ADAR fusion protein using antibodies against the RBP or against V5 tag on C terminus of ADAR.

**△ CRITICAL:** The RBP must be in frame with the downstream catalytic domain of ADAR, this means that the endogenous stop codon for the RBP of interest must be excluded. Any issues with the cloning frame will disrupt the downstream p2A GFP and GFP fluorescence will not be observed.

**Install and compile software dependencies**

© Timing: 1 day

- A Unix based system is necessary for analysis. A dedicated RNA sequencing analysis server running Red Hat Enterprise Linux or Centos is recommended, however any Linux distribution can be used. For convenience, all scripts necessary for processing can be found in the supplement (Data S1) and we have also created a virtual Centos machine with all dependencies pre-installed (Zenodo: <https://doi.org/10.5281/zenodo.4567690>).

- a. After downloading the virtual machine files, the .vmdk or .ovf files can be opened with the following software that are free for noncommercial use:
    - i. VMware Workstation player: <https://www.vmware.com/products/workstation-player/workstation-player-evaluation.html>
      - Once installed, go to menu and select Player -> File -> Open
      - Select "Centos\_sequencing.ovf" file
      - Allocate resources to the virtual machine depending on what is available on the host computer (see below).
    - ii. Alternative software includes: VirtualBox: <https://www.virtualbox.org/>
  - b. For analysis with the virtual machine we recommend the following resources be allocated to it:
    - i. 8 vCPUs, 64gb of RAM and 4TB of space
    - ii. At least 32GB of RAM, 2TB of free disk space and a quad core processor is necessary, however more resources are highly recommended to process multiple samples efficiently.
10. To use the pipeline on a non-virtual machine, the following software must be installed in a Linux/Unix environment according to the software specific instructions:
- a. bedtools suite (v2.16.2, 2.26.0 or later): <http://bedtools.readthedocs.io/>
  - b. Bowtie2 (v2.1.0, v2.2.9): <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>
  - c. Cutadapt: <https://cutadapt.readthedocs.io/en/stable/installation.html>
  - d. HyperTRIBE software: <https://github.com/rosbashlab/HyperTRIBE>
  - e. MariDB v10.1 or later <https://downloads.mariadb.org/>:
  - f. Perl (v5.8.8, v5.12.5, v5.22.1): <https://www.perl.org/get.html>
  - g. Perl Module: DBI.pm (v1.631, v1.636)
  - h. Perl Module: DBD:mysql (v4.042)
  - i. Picard (v2.8.2): <https://broadinstitute.github.io/picard/>
  - j. Python (v2.7.2 or later): <https://www.python.org/downloads/>
  - k. SAMtools (v1.11): <http://samtools.sourceforge.net/>
  - l. SRA Toolkit (v2.10.8): <https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>
  - m. STAR (v2.7.3a): <https://github.com/alexdobin/STAR>
  - n. Stringtie: <https://ccb.jhu.edu/software/stringtie/>
  - o. Trimmomatic (v0.36): <http://www.usadellab.org/cms/?page=trimmomatic>

**△ CRITICAL: Software installation on linux machines should be performed by someone that is comfortable and knowledgeable, use of improper commands or copying / pasting can result in irreversible damage to the operating system. We highly recommend utilizing the generated virtual machine (Zenodo: <https://doi.org/10.5281/zenodo.4567690>) but have also included software installations below.**

11. Detailed software installation instructions:
- a. Open the terminal and create an "RNA" subfolder in your home directory where all the software and scripts will be installed, then navigate to that folder.
    - i. Download HyperTRIBE.tar.gz from the supplement (Data S1) or Zenodo archive.
    - ii. Make an RNA folder within your home directory.
      - /home/jbiswas/RNA
    - iii. Navigate to the newly created directory and unzip the script archive (in supplementary documents (Data S1) and on Zenodo archive).
      - cd RNA
      - tar -xvzf HyperTRIBE.tar.gz
    - iv. Update instructions for Ubuntu:
      - apt-get install build-essential
      - apt-get update

- apt-get upgrade
- v. Update instructions for Centos/Red Hat:
  - yum groupinstall 'Development Tools'
  - yum check-update
  - yum check-update
- b. Alternatively, scripts can be download via HyperTRIBE GitHub: <https://github.com/rosbashlab/HyperTRIBE>
  - i. git clone <https://github.com/rosbashlab/HyperTRIBE>
  - ii. May need the following command before installing HyperTRIBE:
    - For Ubuntu: apt-get install git
    - For Centos/Red Hat: yum install git
- c. Perl (v5.8.8, v5.12.5, v5.22.1): <https://www.perl.org/get.html>. All versions of Perl that have been tested are mentioned here, but newer versions of Perl should work as well.
  - i. Open the terminal and type the following:
    - curl -L <http://xrl.us/installperlnix> | bash
  - ii. May need the following commands before installing Perl:
    - For Ubuntu:
      - apt install curl
      - apt install make
    - For Centos/Red Hat:
      - yum install curl
      - yum install make
- d. Python (v2.7.2 or later): <https://www.python.org/downloads/>
  - i. For Ubuntu:
    - apt update
    - apt install software-properties-common
    - add-apt-repository ppa:deadsnakes/ppa
    - apt install python3.8
  - ii. For Centos/Red hat:
    - yum update -y
    - yum install -y python3
- e. Perl Module: DBI.pm (v1.631, v1.636) and DBD:mysql (v4.042)
  - i. cpan DBI
    - continue with default parameters
    - May need to install the following before proceeding in Centos/Red Hat:
      - yum install perl-CPAN
      - yum install gcc
      - cpan YAML
  - ii. cpan DBD::mysql
    - May need to do this after installing mariadb
- f. Bedtools suite (v2.16.2, 2.26.0 or later): <http://bedtools.readthedocs.io/>
  - i. For Ubuntu can use the following command:
    - apt install bedtools
  - ii. Alternatively can use the following:
    - wget <https://github.com/arq5x/bedtools2/releases/download/v2.29.1/bedtools-2.29.1.tar.gz>
    - tar -zxvf bedtools-2.29.1.tar.gz
    - cd bedtools2
    - make
    - May need the following:
      - For ubuntu: apt-get install g++

- For Centos/Red Hat:
    - yum groupinstall 'Development Tools'
    - yum install bzip2-devel
    - yum install xz-devel
    - yum install zlib-devel
- g. SAMtools (v1.11): <http://www.htslib.org/>
  - i. Download from htslib website (samtools 1.11).
  - ii. tar -vxjf samtools-1.11.tar.bz2
  - iii. cd samtools-1.x
    - In Ubuntu may need the following command before moving forward:
      - apt-get install libncurses-dev
      - apt-get install zlib1g-dev
      - apt-get install libbz2-dev
      - apt-get install liblzma-dev
    - In Centos/Red Hat may need the following command before moving forward:
      - yum install ncurses-devel ncurses
      - yum install bzip2-devel
      - yum install xz-devel
      - yum install curl-devel
  - iv. make
  - v. make install
- h. Cutadapt: <https://cutadapt.readthedocs.io/en/stable/installation.html>
  - i. On Ubuntu: apt install cutadapt
  - ii. On Centos/Red Hat (need python3 installed):
    - yum install epel-release
    - yum install pip
    - pip install upgrade-pip
    - python3 -m pip install --user --upgrade cutadapt
- i. Trimmomatic (v0.36): <http://www.usadellab.org/cms/?page=trimmomatic>
  - i. Compile by unzipping
- j. STAR (v2.7.3a): <https://github.com/alexdobin/STAR>
  - i. On Ubuntu may need the following before proceeding:
    - apt-get update
    - apt-get install g++
    - apt-get install make
    - apt-get install libz-dev
  - ii. wget <https://github.com/alexdobin/STAR/archive/2.7.3a.tar.gz>
  - iii. tar -xzf 2.7.3a.tar.gz
  - iv. cd STAR-2.7.3a/source
  - v. make STAR
- k. Bowtie2 (v2.1.0, v2.2.9): <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>
  - i. Download the source file.
  - ii. Unzip the folder into RNA directory.
  - iii. Move to the folder using cd command.
  - iv. Use command "make" to compile everything.
  - v. May need the following command before installing Bowtie2:
    - apt-get install libz-dev
- l. SRA Toolkit (v2.10.8): <https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>
  - i. Untar the archive (tar -xvf sratool...).
  - ii. cd into the directory.
    - cd bin

- on Ubuntu: `vdb-config --interactive`
    - on Centos/Red hat: `vdb-config -i`
  - iii. Other versions of this software should work as well.
  - m. Picard (v2.8.2): <https://broadinstitute.github.io/picard/>
    - i. Install java
      - `apt install default-jre`
      - `java -version`
    - ii. Download .jar file from website and place in RNA folder.
  - n. Stringtie: <https://ccb.jhu.edu/software/stringtie/>
    - i. `git clone https://github.com/gpertea/stringtie`
    - ii. `cd stringtie`
    - iii. `make release`
  - o. MariDB v10.1 or later <https://downloads.mariadb.org/>: <https://hypertribe.readthedocs.io/en/latest/mariadb.html>
    - i. For Centos/Red Hat:
      - `yum update`
      - `yum install mariadb-server`
      - `yum install mariadb-devel`
      - `systemctl enable mariadb`
      - `systemctl start mariadb`
      - `mysql_secure_installation`
    - ii. For ubuntu:
      - `apt-get install software-properties-common`
      - `apt-key adv --recv-keys --keyserver hkp://keyserver.ubuntu.com:80 0xF1656F24C74CD1D8`
      - `add-apt-repository 'deb [arch=amd64,i386,ppc64el] http://mirror.jmu.edu/pub/mariadb/repo/10.1/ubuntu xenial main'`
      - `apt update -y`
      - `apt install -y mariadb-server`
      - `mysql -V`
        - confirms installation
      - `apt-get install -y libmariadbclient-dev`
      - `systemctl start mariadb.service`
      - `systemctl enable mariadb.service`
      - `/usr/bin/mysql_secure_installation`
    - iii. Perform installation of mariadb with the following commands:
      - Enter current password for root (enter for none): <Enter>
      - Set root password? [Y/n]: Y
      - New password: !htribe01
      - Re-enter new password: !htribe01
      - Remove anonymous users? [Y/n]: Y
      - Disallow root login remotely? [Y/n]: Y
      - Remove test database and access to it? [Y/n]: Y
      - Reload privilege tables now? [Y/n]: Y
    - iv. Confirm that you can login to mysql with the following:
      - `mysql -h localhost -u root -p`
    - v. Create a database while in mysql:
      - `CREATE DATABASE hg38;`
      - `CREATE DATABASE mm10;`
12. Add software repositories to linux path, replacing the username and directories with what is present on your own system.
- a. Open the terminal
  - b. Edit the path file

- i. `gedit ~/.bashrc`
  - c. Add the following commands to the end of the file, change the path to each file, will need to exit and re-open terminal before changes take effect, replace jbiswas with your username:
  - d. Bedtools
    - i. `export PATH=$PATH:/home/jbiswas/RNA/bedtools2/bin`
  - e. Bowtie2
    - i. For ubuntu: `export PATH=$PATH:/home/jbiswas/RNA/bowtie2-2.4.2-linux-x86_64`
    - ii. For Centos/Red Hat: `export PATH=$PATH:/home/jbiswas/RNA/bowtie2-2.2.9`
  - f. Picard
    - i. `export PICARD=/home/jbiswas/RNA/picard.jar`
    - ii. `alias picard="java -jar $PICARD"`
  - g. STAR
    - i. `export PATH=$PATH:/home/jbiswas/RNA/STAR-2.7.3a/bin/Linux_x86_64/`
  - h. SAMtools
    - i. `export PATH=$PATH:/home/jbiswas/RNA/samtools-1.11`
  - i. Stringtie
    - i. `export PATH=$PATH:/home/jbiswas/RNA/stringtie`
  - j. SRA toolkit
    - i. For ubuntu: `export PATH=$PATH:/home/jbiswas/RNA/sratoolkit.2.10.8-ubuntu64/bin`
    - ii. For Centos/Red Hat: `export PATH=$PATH:/home/jbiswas/RNA/sratoolkit.2.10.9-centos_linux64/bin`
13. The following annotations files will be of use for downstream analysis and can be downloaded from their corresponding databases. For convenience, Illumina has the latest annotation files on their iGenomes website (for *Homo Sapiens*, UCSC hg38 and for *Mus Musculus*, UCSC mm10 were used): [https://support.illumina.com/sequencing/sequencing\\_software/igenome.html](https://support.illumina.com/sequencing/sequencing_software/igenome.html)
- a. In your home directory create an index subfolder:
    - i. `mkdir index`
  - b. Once the iGenomes file has been downloaded, move it to the index subfolder of your RNA directory and use the following command to unpack:
    - i. `tar -xvf "nameoffilegoeshere.tar.gz"`
  - c. Note the following locations for use in the later scripts (Mouse, for human just replace *Mus\_musculus* with *Homo\_sapiens*).
    - i. Genome sequence file - `/home/jbiswas/RNA/index/Mus_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa`
    - ii. Genome index file - `/home/jbiswas/RNA/index/Mus_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa.fai`
    - iii. Gene features files - `/home/jbiswas/RNA/index/Mus_musculus/UCSC/mm10/Annotation/Archives/archive-current/Genes/genes.gtf`
  - d. Create a STAR index from the genome sequence file and gene features file above.
    - i. `mkdir STARIndex`
    - ii. copy the above genome sequence (genome.fa) and gtf file (genes.gtf) into this directory
    - iii. `STAR --runMode genomeGenerate --runThreadN 8 --genomeDir mm10_star_index --genomeFastaFiles /home/jbiswas/RNA/index/Mus_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa --sjdbGTFfile /home/jbiswas/RNA/index/Mus_musculus/UCSC/mm10/Annotation/Archives/archive-current/Genes/genes.gtf --outFileNamePrefix /home/jbiswas/RNA/index/mm10_star_index`
    - iv. `STAR --runMode genomeGenerate --runThreadN 8 --genomeDir hg38_star_index --genomeFastaFiles /home/jbiswas/RNA/index/Homo_sapiens/UCSC/hg38/Sequence/WholeGenomeFasta/genome.fa --sjdbGTFfile /home/jbiswas/RNA/index/Homo_sapiens/UCSC/hg38/Annotation/Archives/archive-current/Genes/genes.gtf --outFileNamePrefix /home/jbiswas/RNA/index/hg38_star_index`
      - May need to change the runThreadN depending on your machine
      - May need to change ulimit



- ulimit -n 30000
- 14. Useful path definitions to know for human samples.
  - a. star\_indices="/home/jbiswas/RNA/index/hg38\_star\_index"
  - b. TRIMMOMATIC\_JAR="/home/jbiswas/RNA/TRIMMOMATIC/0.39/trimmomatic.jar"
  - c. PICARD\_JAR="/home/jbiswas/RNA/picard.jar"
  - d. GENOME\_FAI\_FILE="/home/jbiswas/RNA/index/Homo\_sapiens/UCSC/hg38/Sequence/WholeGenomeFasta/genome.fa.fai"
  - e. gtf\_file="/home/jbiswas/RNA/index/Homo\_sapiens/UCSC/hg38/Annotation/Archives/archive-current/Genes/genes.gtf"
- 15. Useful path definitions to know for mouse samples.
  - a. star\_indices="/home/jbiswas/RNA/index/mm10\_star\_index"
  - b. TRIMMOMATIC\_JAR="/home/jbiswas/RNA/TRIMMOMATIC/0.39/trimmomatic.jar"
  - c. PICARD\_JAR="/home/jbiswas/RNA/picard.jar"
  - d. GENOME\_FAI\_FILE="/home/jbiswas/RNA/index/Mus\_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa.fai"
  - e. gtf\_file="/home/jbiswas/RNA/index/Mus\_musculus/UCSC/mm10/Annotation/Archives/archive-current/Genes/genes.gtf"

### Identify cell line of interest or generate MS2 tagged cell line

⌚ Timing: 1 day - 1 month

**Note:** The generation of stable cell lines constitutively expressing RBP-E488Q-ADAR fusions has been challenging (See [limitations](#) section). This can be circumvented by using transient expression either by lipid-based transfection, nucleofection or other equivalent methods. In the future, inducible systems may be employed to facilitate stable line generation.

- 16. If performing TRIBE on RBPs other than MCP, use of cell lines that are readily transfectable is recommended. For MS2-TRIBE, use of cell lines containing the MS2 stem loops is required and different available lines are detailed below.
- 17. Utilize previously published MS2 cell lines or mouse models that are available by request from the respective publications and laboratories:
  - a.  $\beta$ -Actin ([Lionnet et al., 2011](#)) and Arc ([Das et al., 2018](#)) mouse models have been generated.
  - b. Random integration libraries using retroviral central dogma (CD) tagging in human osteosarcoma (U2OS) cells ([Sheinberger et al., 2017](#)) and splice acceptor gene trap in human bronchial epithelial (HBEC) cells ([Wan et al., 2021](#)).
  - c. Other stable cell lines have been developed where the MS2 stem loops were targeted to a specific gene or reporter:
    - i. Immortalized mouse embryonic fibroblasts from the  $\beta$ -Actin MS2 mouse ([Lionnet et al., 2011](#)).
    - ii. Human osteosarcoma (U2OS) cells with MS2 tagged p21 ([Carvajal et al., 2018](#)) or a doxycycline inducible reporter RNA ([Janicki et al., 2004](#)).
    - iii. Mouse embryonic stem cells (mESCs) with MS2 tagged Esrrb ([Cho et al., 2018](#); [Spille et al., 2019](#)).
- 18. Generate an MS2 tagged cell line with CRISPR-Cas9:
  - a. Previously, mouse models and cell lines were generated with gene targeting and homologous recombination, however this process has been made easier with the widespread use of CRISPR-Cas9 homology directed repair ([Spille et al., 2019](#)).
- 19. For genome integrated MS2 lines, a modified genome index should be generated. This will allow the MS2 stem loops and surrounding gene regions to be available for mapping during RNA sequencing analysis.

**Table 1. Example Gene Transfer Format (GTF) file for custom genomes (tab separated text file without headers)**

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attribute
chrZ	unknown	start_codon	1	3	.	+	.	gene_id "Actb_ms2"; gene_name "Actb_ms2"; p_id "P111304811"; transcript_id "NM_111304811"; tss_id "TSS111304811";
chrZ	unknown	CDS	5	15587	.	+	0	gene_id "Actb_ms2"; gene_name "Actb_ms2"; p_id "P111304811"; transcript_id "NM_111304811"; tss_id "TSS111304811";
chrZ	unknown	exon	5	15587	.	+	.	gene_id "Actb_ms2"; gene_name "Actb_ms2"; p_id "P111304811"; transcript_id "NM_111304811"; tss_id "TSS111304811";
chrZ	unknown	stop_codon	15591	15593	.	+	.	gene_id "Actb_ms2"; gene_name "Actb_ms2"; p_id "P111304811"; transcript_id "NM_111304811"; tss_id "111304811";

After running the `maedit_wtRNA_RNA.sh` script, individual files containing editing sites from each replicate are generated:

**△ CRITICAL:** The nucleotide sequence of the MS2 integration site must be known for downstream analysis. The integration site should be sequenced for confirmation and both FASTA and gene transfer format (GTF) files containing the sequence of the integration site and the surrounding 2kb should be generated.

- a. Use the published sequence of the MS2 stem loops and the surrounding 2 kb to generate a FASTA file. See [expected outcomes](#) for example of FASTA file format.
  - i. Several MS2 sequences are available as plasmids on Addgene (Plasmid #s: 31865, 40651, 84561, 102718, 104391). MS2V5 (84561, 102718) or MS2V6 (Tutucci et al., 2018) are recommended for use in mammalian cells (Addgene Plasmid #104391).
  - ii. If using a previously published MS2 integrated cell line - the specific stem loop sequence and location of integration should be used.
  - iii. If generating a novel MS2 integrated cell line - MS2 sequence integration can be validated with either 3'/5' RACE or next generation DNA sequencing.
- b. Use the known gene annotations (start codon, exon, CDS, stop codon) to generate a GTF file for the MS2 integrated locus. See [expected outcomes Table 1](#) for example.
- c. Download the latest genome index and GTF files from Illumina iGenomes (mouse mm10, human hg38).
- d. Concatenate the FASTA file from iGenomes with the reference genome FASTA file (e.g., mm10.fa) to generate a modified genome FASTA (mm10\_chrZ.fa).
  - i. `cat mm10.fa chr_Z.fa > mm10_chr_Z.fa`
- e. Concatenate the MS2 GTF file (chrZ\_genes.gtf) to the reference genome GTF file (e.g., mm10\_genes.gtf) to generate a modified GTF (mm10\_chrZ\_genes.gtf):
  - i. `cat mm10_genes.gtf chrZ_genes.gtf > mm10_chrZ_genes.gtf`
- f. Create a bowtie index:
  - i. `nohup bowtie-build mm10_chrZ.fa mm10_chrZ &`
- g. Create a bowtie2 index:
  - i. `nohup bowtie2-build mm10_chrZ.fa mm10_chrZ &`
- h. Create an fai index using SAMtools:
  - i. `samtools faidx mm10_chrZ.fa`
- i. Create a STAR index:
  - i. `STAR --runMode genomeGenerate --runThreadN 8 --genomeDir star_indices --genomeFastaFiles mm10_chrZ.fa --sjdbGTFfile mm10_chrZ_genes.gtf`
- j. Preview the files on the Integrative Genomics Viewer (IGV) to confirm that the new MS2 chromosome appears and has associated references.

**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anti-GFP from mouse IgG1κ (clones 7.1 and 13.1)	Roche/Sigma	Cat# 11814460001, RRID:AB_390913
<b>Bacterial and virus strains</b>		
Subcloning Efficiency DH5 Competent Cells	Thermo Fisher Scientific	Cat# 18265017
<b>Biological samples</b>		
Mouse embryonic fibroblasts, beta-actin MBS	This study. (Biswas et al., 2020; Lionnet et al., 2011)	N/A
Mouse embryonic fibroblasts, no beta actin MBS	This study. (Biswas et al., 2020; Lionnet et al., 2011)	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
EDTA	Sigma	Cat# EDS-100G
Ethanol 70% vol/vol made from 100% vol/vol ethanol and ultrapure water	Thermo Fisher Scientific	Cat# P288-500
jetPRIME transfection reagent	Polyplus	Cat# 114-15
TRizol LS Reagent	Thermo Fisher Scientific	Cat# 10296010
TURBO DNA-free Kit	Thermo Fisher Scientific	AM1907
<b>Critical commercial assays</b>		
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina with Sample Purification Beads	New England Biolabs	Cat# E7765
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	Cat# E7490
NEBNext Multiplex Oligos for Illumina (Index Primers Set 1)	New England Biolabs	Cat# E7335
PCR Purification Kit	Thermo Fisher Scientific	Cat# K310001
Qubit RNA HS Assay Kit	Thermo Fisher Scientific	Cat#Q32852
Plasmid DNA Midiprep/Maxiprep Kit	MACHERY-NAGEL	Cat#740410 or Cat#740414
<b>Deposited data</b>		
MS2-TRIBE data, raw and analyzed data	This study. (Biswas et al., 2020)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152855">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152855</a>
<b>Experimental models: cell lines</b>		
Human osteosarcoma, U2OS	(Janicki et al., 2004)	N/A
Immortalized mouse embryonic fibroblasts, MEFs	This study. (Biswas et al., 2020; Lionnet et al., 2011)	N/A
<b>Recombinant DNA</b>		
Plasmid: mCherry-ADAR	Addgene	<a href="https://www.addgene.org/154786/">https://www.addgene.org/154786/</a>
Plasmid: MCP-ADAR	Addgene	<a href="https://www.addgene.org/154787/">https://www.addgene.org/154787/</a>
<b>Software and algorithms</b>		
BEDTools	(Quinlan and Hall, 2010)	<a href="https://github.com/ark5x/bedtools2">https://github.com/ark5x/bedtools2</a>
Bowtie2	(Langmead and Salzberg, 2012)	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
Cutadapt	(Martin, 2011)	<a href="https://cutadapt.readthedocs.io/en/stable/installation.html">https://cutadapt.readthedocs.io/en/stable/installation.html</a>
FastQC	Babraham Bioinformatics	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
GraphPad Prism, version 8	GraphPad	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
HTSeq, version 0.6.1	(Anders et al., 2015)	<a href="https://github.com/htseq/htseq">https://github.com/htseq/htseq</a>
HyperTRIBE software	(Biswas et al., 2020)	<a href="https://github.com/rosbashlab/HyperTRIBE">https://github.com/rosbashlab/HyperTRIBE</a> All processing scripts can be found in Data S1 and are also present in the virtual machine (Zenodo: <a href="https://doi.org/10.5281/zenodo.4567690">https://doi.org/10.5281/zenodo.4567690</a> )
ImageJ v1.52p	(Schneider et al., 2012)	<a href="https://imagej.net">https://imagej.net</a>
Integrated Genome Viewer (IGV)	(Robinson et al., 2011)	<a href="https://software.broadinstitute.org/software/igv/">https://software.broadinstitute.org/software/igv/</a>
MariaDB (v10.1 or later)	MariaDB Foundation	<a href="https://downloads.mariadb.org/">https://downloads.mariadb.org/</a>
Perl (v5.8.8, v5.12.5, v5.22.1)	The Perl Programming Language	<a href="https://www.perl.org/get.html">https://www.perl.org/get.html</a>

(Continued on next page)

### Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Perl Module DBI.pm (v1.631, v1.636) and DBD mysql (v4.042)	MetaCPAN	<a href="https://www.perl.org/get.html">https://www.perl.org/get.html</a>
Picard (v2.8.2)	Broad Institute	<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
Python (v2.7.2 or later)	Python Software Foundation	<a href="https://www.python.org/downloads/">https://www.python.org/downloads/</a>
SAMtools	(Li et al., 2009)	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
SRA Toolkit	(Leinonen et al., 2011)	<a href="https://github.com/ncbi/sra-tools">https://github.com/ncbi/sra-tools</a>
STAR (v2.7.3a or later)	(Dobin et al., 2013)	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
StringTie	(Pertea et al., 2015)	<a href="https://ccb.jhu.edu/software/stringtie/">https://ccb.jhu.edu/software/stringtie/</a>
Unix based operating system	CentOS or RedHat Enterprise Linux	<a href="https://www.centos.org/">https://www.centos.org/</a>
<b>Other</b>		
Electrophoresis instrument	Agilent	Bioanalyzer 2100 or equivalent
Fluorescence-activated cell sorter (FACS)	BD Biosciences	FACSAria II or equivalent
Illumina sequencing system	Illumina	HiSeq 4000 or NovaSeq or equivalent
PCR Machine	Bio-Rad	T100 or equivalent
Fluorimeter	Thermo Fisher Scientific	Qubit 4 or equivalent
Spectrophotometer	Thermo Fisher Scientific	NanoDrop ND-1000 or equivalent

## MATERIALS AND EQUIPMENT

MEF FACS buffer	Final concentration (mM or $\mu$ M)	Amount
10 $\times$ PBS	1 $\times$	1 mL
10% (w/v) BSA	1% (w/v)	1 mL
EDTA (0.5 M)	5mM	100 $\mu$ L
DAPI (0.1 mg/mL)	0.1 $\mu$ g/mL	10 $\mu$ L
ddH <sub>2</sub> O (ultrapure)	n/a	7.89 mL
<b>Total</b>	<b>n/a</b>	<b>10 mL</b>

FACS buffer should be made fresh, in a sterile tissue culture hood and stored at 4°C until ready to harvest cells.

## STEP-BY-STEP METHOD DETAILS

### Cell culture and transfection

⌚ Timing: 1 day

ADAR fusion plasmids are introduced into cells by transient transfection.

**Note:** The following steps were optimized for adherent mouse embryonic fibroblasts and human osteosarcoma cells but may need to be adjusted depending on cell type. We recommend transient transfection of 0.5–1 million cells and FACS sorting of 10k GFP<sup>+</sup> DAPI<sup>-</sup> cells so that adequate amount of RNA (>100ng) can be acquired per sample before moving onto library preparation.

For each sample, prepare 10 cm dishes: 3 $\times$  10 cm dishes for mCherry-ADAR control, 3 $\times$  10 cm dishes for RBP-ADAR fusions, and one dish of un-transfected or mock transfected cells as a negative control. This final dish is useful for FACS gating and for downstream analysis of SNPs present in the parental cell line.

**Note:** RNA isolation is performed in triplicate for each sample, this serves as a backup in case one sample has low RNA integrity or if there is a failure during library preparation. While the extra sample can be omitted, we have found it to be helpful for those attempting the protocol for the first time.

**Note:** All dishes should be passed in the same manner and for equal number of passages relative to one another, this will help minimize the dish-to-dish variability. This consistency allows for RNA editing and gene expression comparisons to confidently be made between dishes (and eventually RNA sequencing libraries).

1. Culture the cells of interest using standard sterile tissue culture protocols, refer to American Type Culture Collection (ATCC) standards if needed. For our TRIBE experiments, mammalian cells are cultured and maintained in a humidified 37°C incubator with 5% CO<sub>2</sub>.
  - a. Mouse embryonic fibroblasts (MEFs) and human osteosarcoma (U2OS) cells are cultured in DMEM 4.5 g glucose, glutamine, supplemented with 1% penicillin streptomycin (P/S) and 10% heat-inactivated fetal bovine serum.
  - b. Once cell lines have been thawed, it is advisable to passage the cells at least two times before using for experiments. Cell passages should be tracked and cells with minimal passage number should be used.
2. For each RBP of interest, transfect the appropriate number of cells (see above Notes). For adherent MEFs or U2Os cells we use triplicate 10 cm dishes (70–80% confluence gives best results). Follow best practices for transfection reagent of choice and refer to [troubleshooting](#) section, [problem 1](#) if issues occur.

**Optional:** Before beginning next step, check transfection efficiency of each dish using GFP fluorescence. Additionally, for mCherry ADAR control plasmids, check RFP fluorescence. The ADAR p2A GFP design helps check that the reading frame of the RBP upstream of ADAR is correct.

**Optional:** If transfection efficiency is high (> 75%) it is possible to omit the following FACS sorting step and proceed directly to RNA isolation (Eg. HEK cell transfection). If the cells are not uniformly transfected, the WT unedited RNA will dampen the RNA editing levels below the threshold of detection.

**Note:** Prolonged expression of RNA editing constructs can lead to increasing levels of cellular death. Therefore, transfection should be performed the day before RNA isolation and cells should be sorted within 24 h of transfection as editing reaches steady state within that period ([Biswas et al., 2020](#)).

**Alternatives:** The following transfection reagents have been tested to be effective:

HEK293T cells (70–90% transfection efficiency) – Calcium Phosphate, Lipofectamine 2000, Lipofectamine 3000, JetPrime (Polyplus), JetOptimus (Polyplus)

SV40 Mouse embryonic fibroblasts (5–10% transfection efficiency), Human U2OS (25–50% transfection efficiency) – JetPrime (Polyplus), JetOptimus (Polyplus)

### FACS sorting

⌚ Timing: 4 h

For cell lines with lower transfection efficiency, FACS sorting allows isolation of live cells (DAPI negative) transfected with ADAR (GFP positive). Sorting is performed directly into TriZol or equivalent RNA isolation buffer.

**Note:** The following steps were optimized for adherent mouse embryonic fibroblasts (MEFs) and human osteosarcoma cells (U2OS) but may need to be adjusted depending on cell type. We recommend FACS sorting of ~10k GFP<sup>+</sup> DAPI<sup>-</sup> cells so that >100ng of RNA can be acquired per sample.

△ **CRITICAL:** TriZol can cause serious chemical burns and is toxic upon inhalation, appropriate precautions should be taken, and reagent should be handled within chemical hood or class II biosafety cabinet.

3. Before removing cells (U2OS, MEFs) from incubator, prepare materials.
  - a. Place FACS buffer inside tissue culture hood.
  - b. Pre-warm tissue culture materials to 37°C (tissue culture media, DPBS, trypsin).
  - c. Pre-label 15 mL Falcon tubes, one corresponding to each FACS sample.
  - d. Prepare sterile 1.5 mL Eppendorf tubes corresponding to the same number of FACS samples, place 500µL of TriZol in each tube, keep tubes on ice until ready to use.
  - e. For each 10 cm dish, place one sterile 15 mL Falcon tube in the tissue culture hood.
4. Dissociate cultured cells from their 10 cm plates.
  - a. Wash each plate twice with 7 mL pre-warmed DPBS.
  - b. Use 1 mL pre-warmed trypsin to dissociate cells.
  - c. Place cells in incubator to complete dissociation, use gentle tapping of the dish to promote detachment.
  - d. Re-suspend the dissociated cells in complete media (with FBS) to neutralize trypsin activity and place each sample into pre-labeled a 15 mL conical Falcon tube.
5. Preparation of FACS samples: prepare three samples for each RBP-ADAR sample, three samples for the mCherry-ADAR control and one sample of un-transfected cells as a negative control (Figure 1).
  - a. Spin the re-suspended cells gently for 5 min at 500 g.
  - b. Remove the supernatant from the pelleted cells being careful not to disturb the cell pellets.
  - c. Re-suspend the cell pellets in 200–500µL of FACS buffer.
  - d. Pass the re-suspended cells through a mesh strainer to promote dissociation.

**Note:** The density of cells and size of the cell pellet is dependent on cell type. It may be necessary to optimize the amount of FACS buffer to re-suspend in so that sorting can be expedited efficiently.

▣ **Pause point:** Cells can remain on ice for one hour prior to sorting. However, prolonged incubation on ice may cause increased death, decreasing RNA yield and quality.

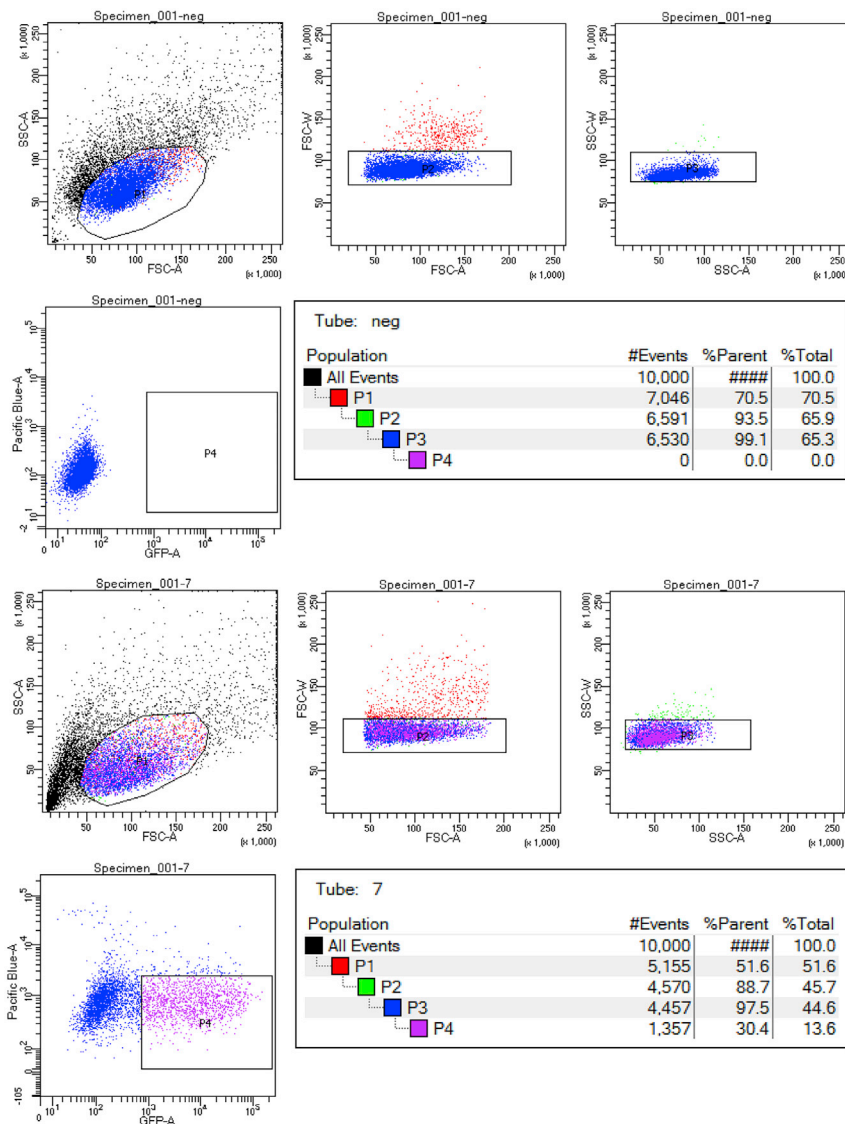
6. Keep FACS samples on ice and proceed immediately to FACS sorting, aim for 10,000 cells per sample with 1,000 cells as a minimum requirement.
  - a. Gate the cells using the negative sample to isolate un-transfected cells (Figure 2).
  - b. Select single cells using forward scatter and side scatter ranges. Use cells that are DAPI negative (alive) and GFP positive (transfected).
7. Ensure that the FACS solution is no more than 10% of total TriZol volume, start with at least 500µL of TriZol in each of the tubes and add more TriZol after sort if necessary.
8. Cap the tubes containing the samples in TriZol, mix well and store on ice until ready to proceed to RNA isolation.

**Optional:** After sorting, TriZol containing cell lysates can be stored at –80 until ready to proceed with RNA isolation.

### RNA isolation and quality control

⌚ **Timing:** 4 h

High quality RNA is crucial to the process of library preparation. We routinely isolate RNA from TriZol, however other methods such as RNA minipreps, spin columns and direct lysis can be used to isolate total RNA. The protocol is also compatible with nascent RNA isolation (McMahon et al., 2016).

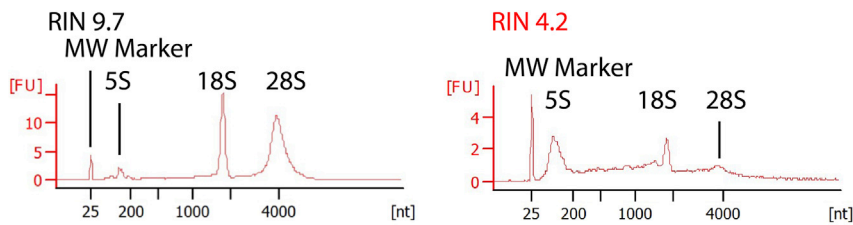


**Figure 2. Representative FACS plots**

Plots show stepwise selection of DAPI negative, GFP positive cell population for downstream RNA isolation and library preparation. Top plots show gate definition; it is important to identify gates that will prevent un-transfected cells from being included in the final population. Associated population counts can also be used to estimate transfection efficiency.

**△ CRITICAL:** TriZol and chloroform can cause serious chemical burns, appropriate precautions should be taken, and reagent should be handled within chemical hood.

9. RNA isolation using TRIZol is the preferred method when dealing with FACS sorted cells. If material is not limiting, RNA miniprep is an acceptable alternative.
  - a. Perform RNA isolation according to manufacturer instructions.
  - b. Isolated RNA should be placed into PCR tubes.
10. Perform DNase digestion of the RNA samples using Turbo DNA Free Kit (Ambion AM1907).
  - a. Add 0.1 volume of 10× DNase buffer to the isolated RNA.
  - b. Add 1uL of DNase to the RNA and mix gently.
  - c. Incubate at 37°C (in thermocycler) for 30 min.



**Figure 3. Examples of RNA integrity measured after RNA isolation**

High quality RNA is required (Left). RNA integrity is evaluated using ribosomal RNA peaks which constitute the majority of RNAs within the cell. Do not proceed with samples that have a low RIN score and are degraded (Right). X-axis, fragment size in nucleotides, y-axis, arbitrary fluorescence units (FU).

- d. While incubating – thaw out the DNase inactivation resin.
- e. Inactivate with 0.1 volumes of the inactivation resin (3.5uL resin for 30uL of RNA) and leave at RT for 2 min.
- f. Spin in PCR tubes for 30 s, careful not to disturb the resin.
- g. Transfer as much as possible to a fresh tube, use remainder for qubit (2uL into 500uL eppendorf tubes) and Bioanalyzer (3uL into PCR strip).

**Note:** Measuring the concentration of RNA with a Nanodrop will overestimate the amount in the sample. It is highly recommended that more specific assays such as Qubit be utilized so that libraries can be prepared with equal amounts of RNA input.

11. Measure concentration of RNA using QuBit RNA high sensitivity assay according to manufacturer instructions.
12. Perform RNA integrity analysis using RNA bio analyzer (Figure 3).
  - a. After sending out samples for quality control (QuBit and Bioanalyzer) we are left with approximately 30uL of RNA. 100 ng of total RNA is ideal for library preparation, however as little as 10 ng can be used. Proceed using samples with RIN scores > 8 (Figure 3).

**Optional:** Tape station can also be used for RNA integrity analysis.

▣ **Pause point:** Once RNA samples have been purified, 2uL and 3uL aliquot can be made for qubit and Bioanalyzer, respectively. All aliquots can be stored at -80 until the next step.

### RNA sequencing library preparation

⌚ **Timing:** 1–3 days

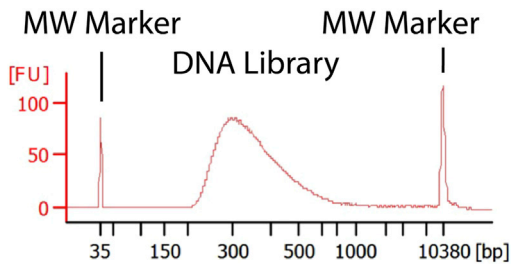
⚠ **CRITICAL:** Do not proceed with library preparation until sufficient, high quality RNA samples are available.

**Note:** Library preparation times can vary widely depending on the number of samples and the individual pause points within the protocol.

**Alternatives:** Stranded, paired end RNA seq library preparation kits are ideal and allow for accurate mapping of transcripts. When dealing with limiting input amounts, we have had satisfactory performance from kits that have a lower required input (such as the NEB ultra II directional RNA seq library prep kit which can use as little as 10ng of total RNA input).

Either rRNA depletion or polyA selection can be used to purify total RNA before moving to library preparation. The decision to use rRNA depletion or poly-A selection should be guided by the model





**Figure 4. Proper cDNA library preparation**

DNA libraries should be correctly sized and absent of adapter dimer peaks X-axis, fragment size in nucleotides, y-axis, arbitrary fluorescence units (FU). See [troubleshooting, problem 2](#) for example of libraries requiring extra purification.

organism and biological question. Caveats of rRNA depletion include that the primers needed are often organism specific and finalized libraries may require increased sequencing depth to achieve an equivalent coverage.

Alternative mRNA seq library preparation kits that have been used in the literature for TRIBE include the TruSeq stranded mRNA from Illumina and the KAPA Stranded mRNA-Seq Kit from Roche.

Depending on the amount of RNA available, different library prep methods may be used. We have found that NEB provides robust kits and routinely use their Ultra II Directional RNA library prep kit. However, many other, equivalent kits are available.

13. Using the NEB Ultra 2 Kit with sample preparation beads – prepare stranded RNA seq samples according to manufacturer instructions. In addition to the kit, the following is required: NEB Index primers, polyA isolation module or rRNA depletion, a magnet appropriate for PCR tubes.
  - a. Isolate mRNA with poly T-beads (or rRNA depletion).
  - b. Fragment the RNA.

**△ CRITICAL:** Fragmentation time is critical, be sure to follow manufacturer’s instructions for proper library size. If using PE 150bp reads, it is recommended that fragmentation aims for ~300bp fragments. RNA fragment size can be checked using Bioanalyzer or Tape station.

- c. First and second strand synthesis (after this step the DNA is stable).

**▮▮ Pause point:** There are multiple pause points and the protocol can be split into two or three days. DNA can be stored after second strand synthesis, can also be stored after end prep and adapter ligation and finally is stable after library cleanup.

- d. End prep and adapter ligation, choose index primers for each sample.

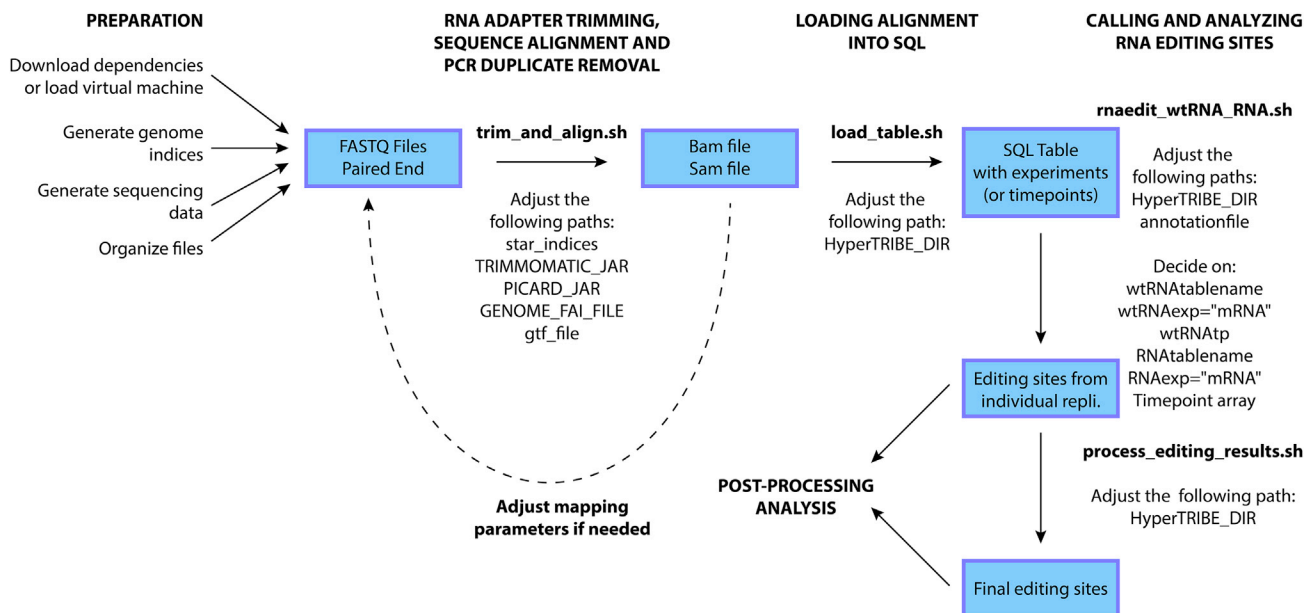
**Note:** NEB Index primers each contain a specific six nucleotide barcode sequence. Each index primer must be chosen as to not interfere with downstream de-multiplexing. If libraries are going to be sequenced in a single pool, each library within that pool must receive a unique index number for downstream demultiplexing of the sequence data. Library preparation kit manufacturers will sell index oligonucleotides in sets (of 12 or 24 indices).

- e. PCR amplification, double library cleanup to remove primer dimers.

**Note:** We recommend using the minimal amount of PCR cycles necessary to minimize the amount of PCR duplicates that remain in the final library.

14. Check proper fragmentation and DNA library size using Bioanalyzer ([Figure 4](#)).

**Optional:** Agilent TapeStation can also be used for DNA library size analysis.



**Figure 5. Overview of HyperTRIBE computational analysis**

Flow chart showing steps of data analysis and script paths that need to be changed throughout the protocol. Shell scripts are designated “.sh” and highlighted in bold.

## RNA sequencing and mapping

⌚ Timing: 1 day

The major data generation step for any TRIBE experiment is RNA sequencing. We currently use Illumina next generation sequencers; in the future this approach may be extended to other platforms. For mammalian cells, it is important to achieve the requisite sequencing depth for high confidence SNP calling, we have found that individual lanes of a HiSeq4000 or NovaSeq equivalents running in PE150 mode provide sufficient depth (40Gb per library) at a reasonable cost for a single experiment, where a single experiment contains duplicate RBP samples, duplicate mCherry controls and an untransfected library. Please refer to Figure 5 for a summary of the downstream bioinformatic analysis.

**Note:** The total number of bases output is important for TRIBE (with an ideal of 40Gb per mammalian library. Longer, paired-end reads and higher depth are both useful in performing mapping and SNP calling respectively. We recommend using no shorter than 100bp paired end reads for sequencing and analysis. Long paired end libraries that are sequenced to an appropriate depth will facilitate accurate RNA editing identification.

**Note:** Sequencing depth should be at least 50 million 150bp PE reads per sample when working with human and mouse samples. More reads will be necessary if using 100bp or shorter read lengths.

15. Operate the sequencer according to manufacturer’s instructions.
  - a. Demultiplex the sequencing reads using index primers that were added during library preparation as per manufacturer instructions.
16. Concatenate data from different lanes together, if applicable.
  - a. `cat data1_read1.fq.gz data data1_read2.fq.gz`

**Optional:** Multiple mCherry-ADAR control samples should be concatenated together to make a single, high coverage background file. When concatenated, control editing sites found in either sample are used to call background RNA editing events.

17. Quality control and mapping of sequenced reads using STAR:
  - a. Know the path to the following previously downloaded genome annotations (refer to before beginning section).
  - b. Prepare the shell scripts ([Data S1](#)) using the previously defined paths (refer to [install and compile software dependencies](#) section).
    - i. Star\_indices
    - ii. TRIMMOMATIC\_JAR
    - iii. PICARD\_JAR
    - iv. GENOME\_FAI\_FILE
  - b. Move the modified trim\_and\_align\_PE.sh file ([Data S1](#)) into the directory with your paired end RNA seq data.
  - c. Use the trim\_and\_align script to perform quality control, adapter trimming and mapping of the paired-end data using STAR. For each piece of data:
    - i. nohup trim\_and\_align\_PE.sh data1\_read1.fq.gz data1\_read2.fq.gz &
    - ii. Repeat the above command for each sample. Note that this process may be limited by the amount of system memory, for human or mouse samples ensure that the machine has 32GB of RAM per sample that is to be run.
    - iii. Please refer to [troubleshooting, problem 3](#) for common issues running shell scripts.
  - d. While the alignment is running, the output will be stored in the newly created "nohup.out" file, you can track the progress of the alignment using the following command:
    - i. tail nohup.out
  - e. The following files will be created from a successful alignment:
    - i. Log.out – log of alignment process.
    - ii. Log.final.out – contains important mapping quality metrics.
    - iii. Sort.bam – when used with the .bam.bai file can be loaded into IGV to visualize coverage.
    - iv. Sort.bam.bai
    - v. Sort.sam – used for downstream TRIBE analysis.
18. Loading alignments into MySQL database:
  - a. nohup /path\_from\_root/HyperTRIBE/CODE/load\_table.sh sam\_filename mysql\_tablename expt\_name replicate/timepoint &
  - b. example: nohup /home/jbiswas/RNA/HyperTRIBE/hg38/load\_table\_hg38.sh wtRNA.sort-sam testRNA rnalibs 2 &
  - c. Wait for nohup.out file to be updated, this may take several hours depending on the machine resources and sequencing depth. Refer to [troubleshooting, problem 3](#) if running into issues with script permissions.
19. Determination of RNA editing sites from alignments
  - a. Navigate to your working directory and copy over the script ([Data S1](#)) to call RNA editing sites.
    - i. cd /directory\_of\_choice/
    - ii. cp /home/jbiswas/RNA/HyperTRIBE/hg38/maedit\_wtRNA\_RNA\_hg38.sh .
  - b. Modify the heading of the copied script file ([Data S1](#)) to direct the script to the appropriate path to the CODE as well as the path to the genome annotation.
    - i. HyperTRIBE\_DIR="/home/jbiswas/RNA/HyperTRIBE/hg38"
    - ii. Annotationfile="/home/jbiswas/RNA/index/Homo\_sapiens/UCSC/hg38/Annotation/Archives/archive-current/Genes/refFlat.txt"
  - c. Modify the script ([Data S1](#)) to include your choice of table name, control sample (wtRNA), and different experimental samples or timepoints.
    - i. wtRNAtablename="testRNA"

**Table 2. Example of output from a single sample**

Chr	Edit coord	Name	Type	A count	T count	C count	G count	Total count	A_count gDNA/wtRNA	T_count gDNA/wtRNA
chr14	1.03E+08	Cln5	INTRON	18	0	0	2	20	34	0
chr14	14118481	Psm6	INTRON	0	20	3	0	23	0	38
chr14	25696325	Ppif	INTRON	18	0	0	2	20	43	0
chr14	1.02E+08	Lmo7	EXON	18	0	0	2	20	19	0
chr14	87417269	Tdrd3	INTRON	24	0	0	4	28	28	0

After running the process\_editing\_sites.sh script, an excel file containing the background subtracted intersection of replicates from both samples is generated:

- ii. wtRNAexp="rnalibs"
- iii. wtRNAtp="2"
- iv. RNAtablename="testRNA"
- v. RNAexp="rnalibs"
- vi. timepoint=(3 4 5)

**Optional:** Editing thresholds can be changed within the script. Recommended thresholds are 1% for the control sample, and 5% for each of the experimental samples. This provides stringent filtering of RNA editing sites.

- d. Run the modified script to generate files containing output from a single replicate (Expected Outcomes, Table 2).
  - i. nohup./rnaedit\_wtRNA\_RNA.sh &

### TRIBE analysis – Postprocessing of data output

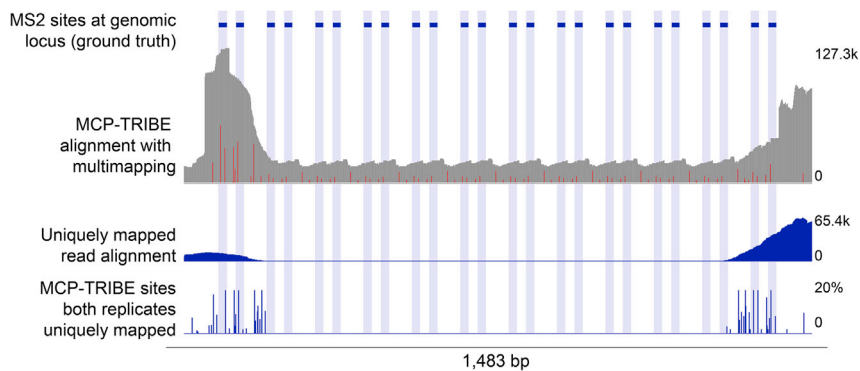
⌚ Timing: 1 day

Once mapped, the mapped files can be used for other standard quantification such as differential gene expression analysis. This section of the protocol will discuss further processing of the mapped data to identify SNPs specific to RNA editing events in both control and experimental samples.

20. Copy the process\_editing\_sites.sh file into your current working directory.
  - a. cp /home/jbiswas/RNA/HyperTRIBE/process\_editing\_results.sh.
  - b. Modify the location of the HyperTRIBE files and the file prefix.
    - i. HyperTRIBE\_DIR="/home/jbiswas/RNA/HyperTRIBE/
    - ii. file\_prefix="MCP\_TRIBE"
21. Use process\_editing\_results.sh script to find the intersection of both experimental replicates and remove the background sites (Expected Outcomes, Table 3).
  - a. Three inputs are required for the script:
    - i. Bedgraph file containing WT editing sites with 1% editing or more.
    - ii. Bedgraph file containing Replicate 1 editing sites with 5% editing or more.
    - iii. Bedgraph file containing Replicate 2 editing sites with 5% editing or more.
  - b. Alternatively, bedtools intersect can be used to perform these functions:
    - i. intersect replicates
    - ii. remove control editing sites

**Table 3. Example of output from intersection of two biological replicates**

Gene Name	Num Edit Sites	Average Editing	Features	Edit_percent_read_str	Identifier_str
Arid1a	4	12.1	EXON,EXON; ...	12.3%_154r,10.4%_134r; ...	chr4_133685443; ...
Alg13	3	10.7	EXON,EXON; ...	6.7%_89r,5.0%_100r; ...	chrX_144352673; ...
Dhx34	3	9.1	EXON,EXON; ...	8.9%_123r,10.2%_137r; ...	chr7_16198986; ...
Katna1	3	12.2	INTRON,INTRON; ...	7.1%_42r,6.8%_44r; ...	chr10_7744094; ...
Med22	3	24.1	EXON,EXON; ...	11.5%_192r,17.0%_165r; ...	chr2_26907730; ...



**Figure 6. RNA editing can be visualized with IGV and occurs both at and adjacent to the site of RNA binding**

Figure modified with permission from Biswas et al., *iScience* 2020.  $\beta$ -actin gene, focusing on the MBS array (x-axis) showing (from top to bottom). MS2 sites at genomic locus (ground truth): blue boxes represent the known location of the MS2 stem loops. MCP-TRIBE alignment with multimapping: reads are depicted in gray (scale bar for number of reads on right), editing sites are depicted as red bars. Uniquely mapped read alignment: mRNA coverage without multimapping depicted in blue (scale bar for number of reads on right). MCP-TRIBE sites both replicates uniquely mapped: editing events as indicated by dark blue bars where height corresponds to the average editing percentage across both replicates at that nucleotide (scale to right). Light blue shading indicates location of the stem loop nucleotides.

22. Visualize the editing sites by importing the editing bedgraph files into IGV (Figure 6). It is often useful to simultaneously visualize coverage using the .bam and .bam.bai files generated after alignment.
  - a. Please refer to [troubleshooting, problem 4](#) for issues with editing site discovery.
23. Determine intersections of gene symbols with other datasets (such as Hi-C, CLIP or RIP).
  - a. The gene symbols output from the TRIBE pipeline can be compared to gene symbols from other datasets using free online tools (<http://genevnn.sourceforge.net/>).
24. Determine relative positions of TRIBE sites to known binding motifs or published CLIP peaks.
  - a. The .bedgraph files from the TRIBE experiment can be intersected with bed files output from published CLIP datasets (<https://www.encodeproject.org/>) using bedtools function “intersect” <https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html>.
  - b. Distances can be measured using the bedtools function “closest” (<https://bedtools.readthedocs.io/en/latest/content/tools/closest.html>).

## EXPECTED OUTCOMES

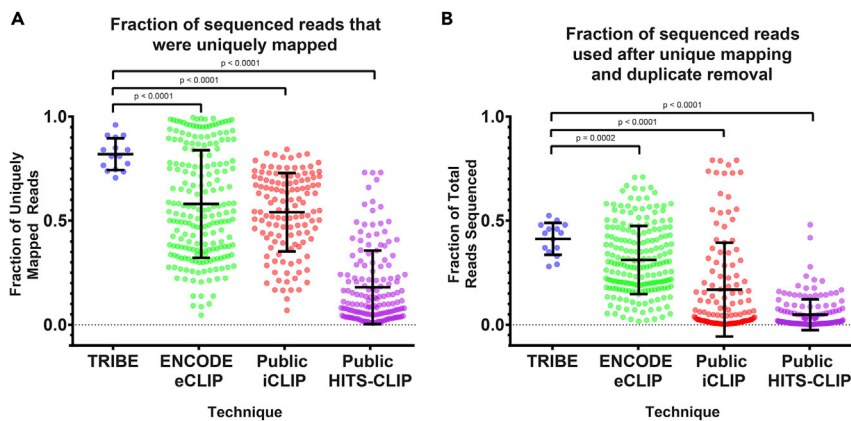
RNA yields can vary widely by cell type. To assist the reader, we have included estimated yields of RNA extraction from MEFs and U2OS cells:

- 1,000 FACS sorted cells = 10–20ng
- 10,000 FACS sorted cells = 100–200ng
- 1 × 10 cm dish of adherent cells = greater than 10ug of RNA

Input RNA should be uniform across samples with as much RNA as possible being used. After sequencing, reads should be quality controlled and libraries should have appropriate mapping statistics and minimal PCR duplicates (Figure 7).

If using the MS2-TRIBE protocol, it is required that one provide a FASTA file with the modified locus. An example FASTA file format for generating custom genome includes a header line beginning with the greater than symbol “>”, the second line contains the nucleotide sequence:

```
>chrZ
```



**Figure 7. Expected mapping parameters from mammalian HyperTRIBE**

Data from Biswas et al. *iScience* 2020 and Van Nostrand et al. *Nat. Methods* 2016.

(A) Percentage of reads in FASTQ file that were uniquely mapped to the genome. Individual points represent individual experiments or biological replicates and error bars represent standard deviation of the mean. P values calculated using two-tailed Welch's t-test. Processed CLIP data from (Van Nostrand et al., 2016).

(B) Percentage of reads in FASTQ file that are retained after complete processing. Major RNA seq processing steps include unique mapping and PCR duplicate removal. Individual points represent individual experiments or biological replicates and error bars represent standard deviation of the mean. P values calculated using two-tailed Welch's t-test. Figure reprinted with permission from Biswas et al. *iScience* 2020. Processed CLIP data from (Van Nostrand et al., 2016).

GCGGCCGCGCTCACAGCCATCTGTAATGGGATCTGATGCCTTCTTCTGTC...

If using the MS2-TRIBE protocol, it is required that one provide a Gene Transfer Format (GTF) file with annotations for the modified locus. Features such as start codon, CDS, exon and stop codon should be included.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Below we describe places in the protocol where editing thresholds are set. These variables are important to call editing sites with high confidence and were defined using the MCP/MBS system and hyperactive *Drosophila* ADAR. By comparing cells with and without the target RNA (MBS array), requiring a minimum of two editing sites per target, each at 5% or higher editing was found to minimize the background levels of RNA editing in the cells lacking MBS. If using a different ADAR catalytic domain (WT or human) or lower read depth the read coverage and percentage threshold levels in the RNA editing script may need to be adjusted.

Trim\_and\_align.sh - STAR alignment output filtering (Data S1):

- Unique alignment (outFilterMultimapNmax):
  - o Set at 1, only reads that map to a single place in the genome are allowed.
- Percentage of mismatches allowed per read (outFilterMismatchNoverLmax):
  - o Set at 0.06, or 6% of the total read length.
- More details can be found in the STAR manual.

Trim\_and\_align.sh - PCR deduplication (Data S1):

- Recommended, protocol uses PICARD on the sorted sam file to remove duplicate reads. In the future UMI based deduplication methods may also prove useful.

RNAedit\_wtRNA\_RNA.sh - RNA Editing script (Data S1):

- Read coverage required for mapping:
  - o Set in `find_rnaeditsites.pl` at 9 reads.
  - o Set in `rnaedit_wtRNA_RNA.sh` at 20 read minimum coverage.
- Sets the percentage threshold for RNA editing events:
  - o Outputs editing events at both 1% and 5% editing thresholds.
    - o While the 5% editing threshold is used to find reproducible, high confidence editing sites in both replicates, the 1% file can be used to determine sub threshold editing events.

## LIMITATIONS

Currently, TRIBE does not provide the nucleotide resolution binding profile of techniques such as in CLIP (Ule et al., 2003), CIMS analysis (Zhang and Darnell, 2011) or RBP binding profiles of eCLIP (Van Nostrand et al., 2016). Prior work has shown that the ADAR fusion can reach up to 300 nt away from a putative or known RBP binding site (Biswas et al., 2020; McMahon et al., 2016). Additionally, genetic fusion of RBPs to ADAR currently prevents researchers from profiling the targets of endogenous proteins or accessing binding site information within patient samples.

While overexpression is minimized by a weak promoter, RBP fusions may be at higher levels than the endogenous promoter. This should be evaluated on a case by case basis and can be assisted by expressing the RBP of interest fused to a catalytically dead (E367A) adenosine deaminase control (Macbeth et al., 2005; Nguyen et al., 2020). While not limiting to TRIBE experiments, prolonged overexpression of hyperactive ADAR and other editing enzymes may cause editing within many coding sequence and lead to proteotoxic stress within the cell. This has made stable cell lines and animal models constitutively expressing the hyperactive ADAR difficult to generate. This does not appear to be the case with wtADAR (McMahon et al., 2016; Xu et al., 2018) and could be mitigated in the future with the use of inducible ADAR fusions.

In the case of MS2-TRIBE, MS2 stem loops are integrated within an RNA of interest. This has been made easier with the advent of CRISPR-Cas9 based methods to perform homology directed repair (Spille et al., 2019), however the cell lines must be validated to show that insertion of the MS2 stem loops does not cause unexpected gene expression changes at the locus of interest (Tutucci et al., 2018).

## TROUBLESHOOTING

### Problem 1

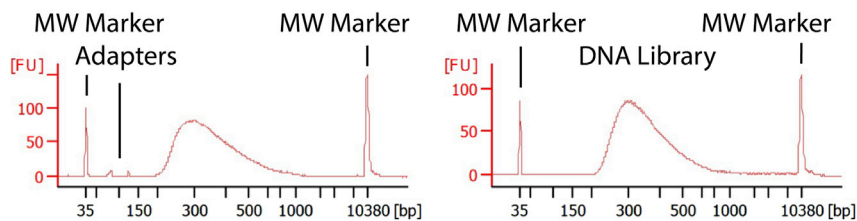
Low transfection efficiency (step 2).

#### Potential solution

There are several possible causes of low-transfection efficiency, the major categories to troubleshoot include plasmid preparation, cell line compatibility and transfection reagent choice. For plasmid preparation, ensure that cloning was completed properly and confirmed with Sanger sequencing. Follow manufacturer recommended protocols for preparing plasmid, ensuring that high purity DNA is obtained. As a control, optimize transfection efficiency with control mCherry-ADAR plasmid (Addgene Plasmid #154786). Certain cells, such as MEFs are difficult to transfect and may require trying different transfection reagents until a suitable reagent is found. Always ensure that the cells are low passage, have been recently plated and are appropriate confluency for transfection. For the transfection process, we recommend following manufacturer recommend protocols for optimizing the ratio of DNA to transfection reagent.

### Problem 2

Adapter dimers (127 nt) or primers (<85 nt) cause extra peaks in final library (Figure 8, left) that should be removed with an extra round of library purification (Figure 8, right) before proceeding to sequencing (step 14).



**Figure 8. cDNA library with adapter peaks requires further library purification**

Adapter dimers (127 nt) or primers (<85 nt) cause extra peaks in final library (Figure 8, left) that should be removed with an extra round of library purification (Figure 8, right) before proceeding to sequencing (step 14). X-axis, fragment size in nucleotides, y-axis, arbitrary fluorescence units (FU).

### Potential solution

Perform second round of library purification and re-analyze the fragment size until no adapter peak is present (right).

### Problem 3

Issues running the editing pipeline (steps 17–19) may include:

- 1) 'Permission denied' error when running shell scripts.
- 2) Error in database creation: DBD::mysql::st execute failed – unable to create SQL table.

### Potential solution

We highly recommend using the provided virtual machine, this has all dependencies pre-installed and will minimize versioning and permission issues.

- 1) Permission error can be due to user permissions on the script:
  - a. `chmod +x script.sh` can be used to make a script executable.
- 2) This error message occurs when tables are uploaded to the same database and can be ignored.

### Problem 4

An expected editing site is not being called by the pipeline (steps 19–22). This could be caused by several issues:

- 1) There is not enough read coverage to call the site with high confidence.
- 2) The site may be present in only one replicate and not the other.

### Potential solution

There are multiple ways to investigate this:

- 1) Low coverage sites can be improved biologically and computationally.
  - a. Biologically - Increase coverage at the site by increasing experimental sequencing depth. This is the preferred solution.
  - b. Computationally, the `process_editing_results.sh` script sets a threshold of 20 reads minimum coverage in the aligned file. This number can be made as low as 9 reads by changing the values in the script. Lowering the read threshold will increase the chance of false positive editing sites being called. The read threshold should be equal for both the control and experimental samples.
- 2) It is important to visualize both the levels of editing and the coverage simultaneously for both replicates as well as mCherry ADAR controls. This can be done by opening the `.bam` or `.bedgraph` files in IGV.



## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jeetayu Biswas ([jeetayu.biswas@alumni.einsteinmed.org](mailto:jeetayu.biswas@alumni.einsteinmed.org)).

### Materials availability

Plasmids encoding mCherry-ADAR (Addgene plasmid #154786) and MCP-ADAR (Addgene plasmid #154787) are available at Addgene.

### Data and code availability

The accession number for the raw sequencing data and identified RNA editing sites reported in [Biswas et al. \(2020\)](#) is NCBI's Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152855>) GEO: GSE152855. Code used to process and analyze the data is publicly available at (Zenodo: <https://doi.org/10.5281/zenodo.4567690> and Github: <https://github.com/rosbashlab/HyperTRIBE/>). All other relevant data are available from the authors upon request.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xpro.2021.100634>.

## ACKNOWLEDGMENTS

J.B. was supported with funding from an MSTP Training Grant T32GM007288 and predoctoral fellowship F30CA214009. R.R. and M.R. were supported by the Howard Hughes Medical Institute and R01DA037721. R.H.S. was supported by National Institutes of Health grant R01NS083085. FACS sorting was performed with core support from P30CA013330.

## AUTHOR CONTRIBUTIONS

Writing – original draft, J.B.; writing – review & editing, J.B., R.R., M.R., and R.H.S.; funding acquisition, J.B., R.H.S., and M.R.; resources, M.R. and R.H.S.; supervision, M.R. and R.H.S.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Biswas, J., Rahman, R., Gupta, V., Rosbash, M., and Singer, R.H. (2020). MS2-TRIBE evaluates both protein-RNA interactions and nuclear organization of transcription by RNA editing. *iScience* 23, 101318.
- Carvajal, L.A., Neriah, D.B., Senecal, A., Benard, L., Thiruthuvanathan, V., Yatsenko, T., Narayanagari, S.-R., Wheat, J.C., Todorova, T.I., Mitchell, K., et al. (2018). Dual inhibition of MDMX and MDM2 as a therapeutic strategy in leukemia. *Sci. Transl. Med.* 10.
- Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415.
- Das, S., Moon, H.C., Singer, R.H., and Park, H.Y. (2018). A transgenic mouse for imaging activity-dependent dynamics of endogenous Arc mRNA in live neurons. *Sci. Adv.* 4, eaar3448.
- 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.
- Herzog, J.J., Xu, W., Deshpande, M., Rahman, R., Suib, H., Rodal, A.A., Rosbash, M., and Paradis, S. (2020). TDP-43 dysfunction restricts dendritic complexity by inhibiting CREB activation and altering gene expression. *PNAS* 117, 11760–11769.
- Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., et al. (2004). From silencing to gene expression: real-time analysis in single cells. *Cell* 116, 683–698.
- Kuttan, A., and Bass, B.L. (2012). Mechanistic insights into editing-site specificity of ADARs. *PNAS* 109, E3295–E3304.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A.L., Chao, J.A., Park, H.Y., de Turris, V., Lopez-Jones, M., and Singer, R.H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* 8, 165–170.
- Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., and Bass, B.L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309, 1534–1539.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12.
- McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: hijacking an RNA-editing enzyme to identify cell-specific targets of RNA-binding proteins. *Cell* 165, 742–753.

- Nguyen, D.T.T., Lu, Y., Chu, K.L., Yang, X., Park, S.-M., Choo, Z.-N., Chin, C.R., Prieto, C., Schurer, A., Barin, E., et al. (2020). HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nat. Commun.* *11*, 2026.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* *33*, 290–295.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* *9*, 671–675.
- Sheinberger, J., Hochberg, H., Lavi, E., Kanter, I., Avivi, S., Reinitz, G., Schwed, A., Aizler, Y., Varon, E., Kinor, N., et al. (2017). CD-tagging-MS2: detecting allelic expression of endogenous mRNAs and their protein products in single cells. *Biol. Methods Protoc.* *2*, bpx004.
- Spille, J.-H., Hecht, M., Grube, V., Cho, W.-K., Lee, C., and Cissé, I.I. (2019). A CRISPR/Cas9 platform for MS2-labelling of single mRNA in live stem cells. *Methods* *153*, 35–45.
- Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2018). An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* *15*, 81–89.
- Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies nova-regulated RNA networks in the brain. *Science* *302*, 1212–1215.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhardt, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* *13*, 508–514.
- Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., et al. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* *184*, 2878–2895.e20.
- Xu, W., Rahman, R., and Rosbash, M. (2018). Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. *RNA* *24*, 173–182.
- Zhang, C., and Darnell, R.B. (2011). Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* *29*, 607–614.