

Jun 02, 2025

MAPIT-seq protocol V.1

DOI

<https://dx.doi.org/10.17504/protocols.io.q26g79ekkvwz/v1>

Qixuan Cheng¹, Gang Xie¹, Xiangyu Zhang¹, Jie Wang¹, Shuangjin Ding¹, Yi-Xia Wu¹, Ming Shi¹, Fei-Fei Duan¹,
Zi-Li Wan¹, Junyu Xiao¹, Yangming Wang¹

¹Peking University



Qixuan Cheng

Peking University

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.q26g79ekkvwz/v1>

Protocol Citation: Qixuan Cheng, Gang Xie, Xiangyu Zhang, Jie Wang, Shuangjin Ding, Yi-Xia Wu, Ming Shi, Fei-Fei Duan, Zi-Li Wan, Junyu Xiao, Yangming Wang 2025. MAPIT-seq protocol V.1. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.q26g79ekkvwz/v1>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: June 02, 2025

Last Modified: June 02, 2025

Protocol Integer ID: 219355

Keywords: RNA-binding protein; dual omics; tissue; single cell; RNA isoform, rna interactomes within the dynamic transcriptomic landscape, essential regulators of rna fate, dynamic transcriptomic landscape, rna editing technology, directed rna editing technology, rna interactions without genetic manipulation requirement, rna interactome, rna fate, rna interaction, utilizing fusion protein, rna, fusion protein, rna deamination, interacting transcript, rbp targets in situ, specific regulation of rbp, sequencing, binding protein, rbp target, rbp regulation, antibody incubation, studying rbp regulation, multiple rbp, antibody, protein, mapit, omics profiling

Abstract

RNA-binding proteins (RBPs) are essential regulators of RNA fate and function. A long-standing challenge in studying RBP regulation has been mapping their RNA interactomes within the dynamic transcriptomic landscape, especially in single-cell contexts and primary tissues. Here, we introduce MAPIT-seq (modification added to RBP interacting transcript-sequencing), an antibody-directed RNA editing technology utilizing fusion protein (pAG-hADAR2dd-rAPOBEC1) to label RBP targets in situ within fixed samples. This methodology integrates formaldehyde fixation, antibody incubation, and RNA deamination, enabling co-profiling of endogenous RBP-RNA interactions without genetic manipulation requirements. We optimize the protocol of MAPIT-seq and demonstrate its robustness and efficiency across multiple RBPs. MAPIT-seq achieves isoform resolution and is readily applied to frozen tissue sections, low-input and single cells, enabling the elucidation of cell stage-specific regulation of RBP. In summary, MAPIT-seq extends the dimensions of multi-omics profiling, offering an effective framework to dissect post-transcriptional regulation across dynamic biological processes and clinically relevant scenarios.

Materials

Instruments

Centrifuge (5810R) Eppendorf
Vortex Mixer (MX-S) Scilogex
Mini Centrifuge (D1008) Scilogex
1.5 mL Centrifuge Tubes (MCT-150-C) Axygen
Thermal Cycler (GE4852T) Biogener
0.2 mL PCR tubes (PCR-0208-C) Axygen
15/50 mL Centrifuge Tubes (352096, 352070) Falcon
PCR Magnet (QYM07A) Quayad

Reagents

1x PBS Phosphate Buffered Saline (Cat. CC008) Macgene
DPBS (Cat. 14190144) Thermo Fisher Scientific
Formaldehyde (Cat. 47608) Sigma-Aldrich, Merck
Proteinase Inhibitor Cocktail (Cat. B14002) Selleck Chemicals
Dithiobis (succinimidyl propionate) (DSP) (Cat. 22586) Thermo Fisher Scientific
Dimethyl Sulfoxide (DMSO) (Cat. D10999) PSAITONG
Glycine (Cat. G0167) VWR Life Science
5% Digitonin (Cat. N253-YH01-01A) Novoprotein
1 M Dithiothreitol (DTT) (Cat. C6014) Bioss
100 mM PMSF (Cat. P0100) Solarbio
Recombinant RNase Inhibitor (Cat. 2313A) Takara
0.5 M EDTA, pH 8.0 (Cat. SL3069) Coolaber
30% (wt/vol) BSA (Cat. A8577) Sigma-Aldrich
Magzol RNA Extraction Reagent (Cat. R4801-02) Macgene
RiboLock RNase Inhibitor (Cat. EO0382) Thermo Fisher Scientific
BioMag® Plus Concanavalin A (Cat. BP531) Bangs Laboratories
Glycerol (Cat. M152) VWR Life Science
Proteinase K (Cat. 2546) Ambion
VAHTS Universal V6 RNA-seq Library Prep Kit (Cat. NR604) Vazyme
Normal Rabbit IgG (Cat. 2729S) Cell Signaling Technology
Guinea Pig anti-Rabbit IgG Antibody (Cat. ABIN101961) Antibodies-Online
Anti-G3BP1 (Cat. 13057-2-AP) Proteintech
rAPOBEC1-pAG-hADAR2dd (E488Q) fusion protein: The pFastBac-rAPOBEC1-pAG-hADAR2dd plasmid sequences are available in article. Fusion proteins were stored in a buffer containing 25 mM HEPES, pH 7.5 and 150 mM NaCl supplemented with 5% glycerol and snap-frozen in liquid nitrogen to store at -80°C

Biological materials

Cell suspension. We have used human HeLa (CCL-2, ATCC) and HEK293T (CRL-3216, ATCC) cell lines.

Software

miniconda (<https://www.anaconda.com/>)

trim-galore (<https://github.com/FelixKrueger/TrimGalore>): apply adapter and quality trimming to FastQ files

MAPIT (<https://github.com/wanglabpku/MAPIT-seq>): MAPIT-seq data analysis and processing

MAPIT dependencies

- bwa (v0.7.17)
- hisat2 (v2.2.1)
- python (at least v3.8)
- picard (at least v2.27)
- samtools (at least v1.14)
- gatk4 (at least v4.2)
- bedtools (at least v2.30.0)
- gffread
- reditools2 (<https://github.com/BioinfoUNIBA/REDIttools2>)
- pip (at least v21.2)
- python packages (pybedtools, pysam (at least v0.16.0.1), pandas (at least v1.4.3), numpy (at least v1.22.3), and scipy)

FLARE (<https://github.com/YeoLab/FLARE>): determine regions of enriched RNA editing

snakemake (v8.27.1): workflow management system

homer2 (v4.11): de novo motif finding

Required data

MAPIT-seq paired-end fastq files can be downloaded from the SRA archive under accession BioProject PRJNA1166181

Human GRCh38 reference genome

(https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/GRCh38.p13.genome.fa.gz)

Human GRCh38 transcript annotation GTF file

(https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/gencode.v40.chr_patch_hapl_scaff.annotation.gtf.gz)

Human GRCh38 transcript annotation GFF3 file

(https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/gencode.v40.chr_patch_hapl_scaff.annotation.gff3.gz)

Human GRCh38 RepeatMasker annotation for interspersed repeats and low complexity DNA sequences

(<http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz>)

Human GRCh38 dbSNP VCF file

(https://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh38p7/VCF/GATK/All_20180418.vcf.gz)



Human GRCh38 Exome Variant Server (EVS) VCF file (http://evs.gs.washington.edu/evs_bulk_data/ESP6500SI-V2-SSA137.GRCh38-liftover.snps_indels.vcf.tar.gz)

Human GRCh38 1000 Genomes Project VCF file

(http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/ALL.wgs.shapeit2_integrated_v1a.GRCh38.20181129.sites.vcf.gz)



Reagent setup

1 DSP Fixation Buffer

- 1.1 50× DSP Stock Solution (50 mg/mL): Balance DSP at room temperature for 30 min, dissolve 50 mg DSP in 1 mL 100% anhydrous DMSO, aliquot, and store at -80°C.
- 1.2 DSP Fixation Solution: Thaw the stock solution at room temperature for 10 min. Add 490 µL PBS dropwise to 10 µL DSP stock in a 1.5 mL tube using a P200 pipette while vortexing (important).
- 1.3 Note: Filter off any precipitates and use freshly within 5 min.



2 Glycine Solution (2.5 M Glycine)

- 2.1 Dissolve 9.38 g glycine in DEPC-treated water to a total volume of 50 mL. Filter with a 0.22 µm membrane.

3 ConA Beads Binding buffer

3.1

	1 M HEPES-KOH, pH 7.5	1 mL
	1 M KCl	500 µL
	1 M CaCl ₂	50 µL
	1 M MnCl ₂	50 µL
	DEPC-treated water	to 50 mL

Store at 4°C.

4 Antibody Buffer 1



4.1

	100 mM PMSF	5 μ L
	Protease Inhibitor Cocktail	5 μ L
	RiboLock RNase Inhibitor	6.25 μ L
	Recombinant RNase Inhibitor	6.25 μ L
	5% Digitonin	1 μ L
	1 M DTT	0.5 μ L
	0.5 M EDTA	2 μ L
	30% BSA	1.67 μ L
	DPBS	to 500 μ L

Prepare fresh. Adjust the Digitonin concentration for different cell lines if necessary.

5 **Antibody Buffer 2**

5.1 Similar to Antibody Buffer 1, but exclude 0.5 M EDTA and 30% BSA. Make fresh.

6 **High-salt buffer**

7

	1 M HEPES, pH 7.5	1 mL
	5 M NaCl	3 mL
	DEPC-treated water	to 50 mL

Storage: Store at 4 °C.

8 **Deaminase-incubation buffer**



8.1

	100 mM PMSF	5 μ L
	Protease inhibitor cocktail	5 μ L
	RiboLock RNase Inhibitor	6.25 μ L
	Recombinant RNase Inhibitor	6.25 μ L
	5% Digitonin	1 μ L
	1 M DTT	0.5 μ L
	High-salt buffer	to 500 μ L

Make fresh.

9 **Deamination buffer**

9.1

	1 M HEPES, pH 7.9	150 μ L
	1 M KCl	600 μ L
	5 M NaCl	30 μ L
	80% Glycerol	625 μ L
	0.1 mM ZnCl ₂	10 μ L
	DEPC-treated water	to 10 mL

Storage: Store at 4 °C.

10 **Proteinase K digestion buffer**

10.1



	1 M Tris-HCl	500 µL
	5 M NaCl	1 mL
	0.5 M EDTA	100 µL
	10% SDS	2.5 mL
	DEPC-treated water	to 50 mL

Storage: Store at RT.

Equipment setup

11 Downloading and installing software and dependencies

- 11.1 Softwares in MAPIT pipeline are mainly available in conda. We recommend you to install these softwares by conda. Some softwares need to install by source code in Github repository.
- 11.2 Miniconda (see <https://www.anaconda.com/docs/getting-started/miniconda/install>)
- 11.3 MAPIT and dependencies `git clone https://github.com/WangLabPKU/MAPIT-seq` `cd MAPIT-seq` `conda env create -f env.yml -c bioconda -c conda-forge` `conda activate Mapit-seq` `chmod +x Mapit` `conda install trim_galore -c conda-forge`
- 11.4 Reditools2 and dependencies `cd ..` `git clone https://github.com/BioinfoUNIBA/REDIttools2` `cd REDIttools2` `pip install -r requirements.txt`
- 11.5 FLARE and dependencies `cd ..` `git clone https://github.com/YeoLab/FLARE` `conda install snakemake`

12 Downloading and preparing the required data

- 12.1 Downloading reference sequence and annotation `cd "your_ref_path"` `wget https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/GRCh38.p13.genome.fa.gz` `wget https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_40/gencode.v40.chr_patch_hapl_scaff.annotation.gtf.gz` `wget \`

```
https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/
release_40/gencode.v40.chr_patch_hapl_scaff.annotation.gff3.gz
```

- 12.2 RepeatMasker wget \ http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz gzip -d *
- 12.3 Downloading and processing known variants (SNP) annotation files: dbSNP, 1000Genome, and EVS cd "your_ref_path" mkdir "your_ref_path"/"genomeVersion_SNP" cd "your_ref_path"/"genomeVersion_SNP"
- 12.4 dbSNP wget \ https://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh38p7/VCF/GATK/All_20180418.vcf.gz
- 12.5 EVS mkdir EVS_split_chr wget \ http://evs.gs.washington.edu/evs_bulk_data/ESP6500SI-V2-SSA137.GRCh38-liftover.snps_indels.vcf.tar.gz tar -xvf ESP6500SI-V2-SSA137.GRCh38-liftover.snps_indels.vcf.tar.gz for i in {1..22} X Y; do awk '{if(substr(\$1, 1, 1) == "#"){print \$0}else if((length(\$4) == 1) && (length(\$5) == 1)) {gsub("MT","M");if(\$1 ~ "chr") print \$0; else print "chr"\$0 } }' ESP6500SI-V2-SSA137.GRCh38-liftover.chr\${i}.snps_indels.vcf | gzip > EVS_split_chr/chr\${i}.gz done
- 12.6 1000Genome wget \ http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/ALL.wgs.shapeit2_integrated_v1a.GRCh38.20181129.sites.vcf.gz mkdir 1000genomes_split_chr zcat ALL.wgs.shapeit2_integrated_v1a.GRCh38.20181129.sites.vcf.gz | awk -v dir_SNP=1000genomes_split_chr '{if(substr(\$1, 1, 1) == "#"){print \$0 > "1000genomes_header"}else if((length(\$4) == 1) && (length(\$5) == 1)) {gsub("MT","M");if(\$1 ~ "chr") print \$0 > dir_SNP/"\$1; else print "chr"\$0 > dir_SNP/chr"\$1 } }' gzip 1000genomes_split_chr/chr*

13 Creating configuration file

- 13.1 MAPIT configuration After downloading the genome sequence (FASTA), genome annotation (GTF and GFF3), RepeatMasker annotation, and known SNP datasets (split by chromosomes), the following command can be used to complete the configuration process. This step includes generating the genome sequence index, extracting gene element annotations, and creating chromosome-split dbSNP VCF files. Upon completion, a configuration file named "GenomeVersion.json" will be generated in the "conf" directory within the MAPIT-seq working directory, which will be used for downstream analyses. Mapit config \ --genomeVersion GRCh38 \ --genomeFasta "full_path"/GRCh38.p13.genome.fa \ --species human \ --outdir "full_path" \ --gff3 "full_path"/gencode.v40.chr_patch_hapl_scaff.annotation.gff3 \ --rmsk "full_path"/rmsk.txt \ --dbSNP "full_path"/GRCh38_SNP/All_20180418.vcf.gz \ -



```
-1000genomesDir "full_path"/GRCh38_SNP/1000genomes_split_chr \ --EVSEVADir  
"full_path"/GRCh38_SNP/EVS_split_chr \ --Reditools "full_path_to_Reditools" \ --FLARE  
"full_path_to_FLARE"
```

- 13.2 **FLARE configuration** To run the FLARE pipeline, you will first need a set of files specifying the genomic regions in which cluster identification will occur. Once the script finishes running, you will see a new folder called "genome_name"_regions, and within that folder a slew of files with increasing indices, i.e., GRCh38_regions_0, GRCh38_regions_1... "full_path_to_FLARE"/workflow_FLARE/scripts/generate_regions.py <full/path/to/your/genome/gtf/file> <genome_name>_regions
- 13.3 **Homer configuration** In an effort to make sure things are standardized for analysis, HOMER organizes promoters, genome sequences and annotation into packages. See <http://homer.ucsd.edu/homer/introduction/configure.html> in details. perl configureHomer.pl -install hg38

Procedure

14 Cell fixation (40 min–1 h)

- 14.1 Collect 5×10^2 – 5×10^5 cells or nuclei in centrifuge tubes.
- 14.2 Wash cells with PBS by centrifuging at 250g at RT for 5 min. 
- 14.3 For formaldehyde fixation, dilute 37% formaldehyde in room-temperature PBS to 0.5%.
- 14.4 Fix cells for 5 min at room temperature. Quench with 0.125 M glycine for 5 min.
- 14.5 (Optional) For DSP fixation, add 500 μ L DSP Fixation Solution to cells, rotate gently at room temperature for 15 min, mix with pipettes, and incubate for another 15 min. Quench with 1 M Tris-HCl, pH 7.4, for 5 min. 
- 14.6 Centrifuge cells at 4°C, 450g, for 5 min. Wash cells twice with pre-cooled DPBS. Then resuspend the fixed cells in 500 μ L DPBS. Fixed cells can be cryopreserved at -80°C with 10% DMSO or proceeded immediately 
- 14.7 Note: After fixation, transfer cells to centrifuge tubes pre-washed with 1% BSA. Usage of a swing-bucket rotor is highly recommended to reduce cell loss. 

15 Cell Binding to ConA Beads (20 min)

- 15.1 (Optional) For single-cell MAPIT-seq, the ConA beads-cells coupling step was omitted. Instead, formaldehyde or DSP fixed cells were collected by centrifuging at 650g, 4°C for 5 min in the following experiments. 
- 15.2 Activate ConA beads by introducing 5–10 µL ConA beads to a volume of binding buffer equivalent to 10 times the bead volume. Ensure thorough mixing and subsequently place the tubes on a magnetic stand for 30 s to 2 min to remove the supernatant. Repeat this activation step, and resuspend the beads in their initial volume of binding buffer.
- 15.3 Add the activated ConA beads to the cell suspension and incubate at room temperature under continuous rotation for 10–20 min.
- 15.4 Wash the cell beads mixture once with DPBS. Transfer the beads to PCR strip tubes.
- 15.5 Note: To mitigate the adhesion of beads to the walls of the centrifuge tubes, it is advisable to pretreat the tubes with 1% BSA. 

16 Cell permeabilization and antibody Incubation (4.5 h)

- 16.1 Resuspend the cells in 50 µL a pre-mixed solution of Antibody Buffer 1, incorporating both the RBP and IgG antibodies at a dilution of 100:1.
- 16.2 Note: The concentration of the antibodies can be adjusted following immunofluorescence assays. Mix gently and avoid repetitive pipetting to prevent material loss. 
- 16.3 Incubate the mixture at 4°C under continuous rotation for 3 h. 
- 16.4 Note: Avoid inversion of the tubes during rotation. It is recommended to mix the suspension intermittently to prevent bead aggregation. 
- 16.5 After a quick spin, place the tubes on a magnetic stand for 30 s to 2 min, then carefully remove the supernatant. Wash the beads once with pre-chilled DPBS on a roller at 4°C for 5 min.
- 16.6 Resuspend the beads in 50 µL Antibody Buffer 2, which contains secondary antibodies at a 100:1 ratio, and incubate at 4°C under continuous rotation for 1 h. 



17 Deaminase Incubation (1 h)

- 17.1 After a quick spin, place the tubes on a magnetic stand for 30 s to 2 min, then carefully remove the supernatant. Wash the beads once with pre-chilled DPBS on a roller at 4°C for 5 min.
- 17.2 Dilute 1 µg of rAPOBEC1-pAG-hADAR2dd in 50 µL Deaminase Incubation Buffer.
- 17.3 Resuspend the beads in 50 µL a pre-mixed Deaminase Incubation Buffer and incubate at 4°C under continuous rotation for 1 h.



18 Deamination Reaction (3.5–5 h)

- 18.1 After a quick spin, place the tubes on a magnetic stand for 30 s to 2 min, then carefully remove the supernatant. Wash the beads twice with 100 µL pre-chilled Washing Buffer, which includes High-Salt Buffer with 0.005% Digitonin, on a roller at 4°C for 5 min per wash.
- 18.2 Prepare 100 µL Deamination Buffer supplemented with the following components: 1 µL Proteinase Inhibitor Cocktail, 1.25 µL RiboLock RNase Inhibitor, 1.25 µL Recombinant RNase Inhibitor, and 0.1 µL 1 M DTT.
- 18.3 Resuspend the beads in 40 µL pre-mixed Deamination Buffer and incubate at 30°C for 3–4 h.
- 18.4 Note: To ensure specificity and minimize substantial RNA degradation, maintain a processing time of no more than 5 hours; optimal editing is typically achieved within 4 h.
- 18.5 Optional: For cells fixed with DSP, reverse crosslinking can be performed by incubating with 50 mM DTT at room temperature for 30 min.
- 18.6 Following the reaction, place the tubes on a magnetic stand for 30 s to 2 min, then carefully remove the supernatant.



19 RNA Extraction and Library Construction (10 h)

- 19.1 (Optional) For single-cell MAPIT-seq, cells were washed once, resuspended in DPBS with 0.05% BSA and 1 U/µl RiboLock RNase Inhibitor and subsequently loaded onto the





10x Genomics Chromium platform.

19.2 Resuspend the beads in 100 μ L Proteinase K Digestion Buffer and add 1 μ L 20 mg/mL Proteinase K. Vortex the mixture briefly, after a quick spin, digest the beads at 56°C for 1 h.

19.3 Note: Resuspend samples in Proteinase K digestion buffer before transferred to a 1.5 ml tube and then add Proteinase K to prevent beads aggregation.



19.4 For large samples (> 100,000 cells): Extract total RNA following the standard TRIzol protocol. Use glycogen to facilitate RNA precipitation. For small samples (500–50,000 cells): Isolate mRNA using Oligo d(T)25 Magnetic Beads and proceed with the Smart-seq2 protocol.

19.5 Construct sequencing libraries from 50–500 ng of RNA using the VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina. Quantify the concentration of the libraries using Qubit and select 10–100 ng of cDNA for sequencing on an Illumina NovaSeq 6000 platform, employing 150 nt paired-end reads.

20 **MAPIT-seq data quality control (15 min)**



20.1 FASTQC: quality control mkdir ./fastqFileQC fastqc -t 2 -o ./fastqFileQC "fastq_file_path"/*.fq.gz

20.2 Apply adapter and quality trimming mkdir ./clean trim_galore -j 20 -q 20 --phred33 \ --stringency 3 --length 20 -e 0.1 \ --paired "fastq_file_path"/xxx_R1.fq.gz \ "fastq_file_path"/xxx_R2.fq.gz \ --gzip -o ./clean --basename \${samplename}_\${replicate} "fastq_file_path"/xxx_R1.fq.gz and "fastq_file_path"/xxx_R2.fq.gz denote the paired-end FASTQ files generated from MAPIT-seq experiments, corresponding to Read 1 and Read 2, respectively. The placeholders \${samplename} and \${replicate} indicate the abbreviated sample name and replicate identifier, respectively.

21 **Aligning and processing MAPIT-seq reads to reference genome sequence (3 h)**



21.1 Abundant RNA (rRNA, tRNA, and mt-tRNA) filtering and two-round uniquely mapping Abundant RNA species, including rRNA, tRNA, and mitochondrial RNA, were computationally removed prior to genome alignment to optimize sequencing read utilization, improve target gene detection sensitivity, and ensure high-quality clean data for downstream analyses. To further enhance the sensitivity and specificity of RNA editing site detection, a two-round uniquely mapping strategy was employed. In this approach, HISAT2 was used in the first round to capture transcriptomic complexity, such as splicing events and RNA modifications, while BWA-MEM was applied in the second

round to map genomic regions of editing sites using unmapped reads in the first round. The resulting alignments from both rounds were subsequently merged to generate a final BAM file for downstream analysis. Mapit mapping -v \$genomeVersion \ --fq clean/\${samplename}_\${replicate}_val_1.fq.gz \ --fq2 clean/\${samplename}_\${replicate}_val_2.fq.gz \ --rna-strandness {F,R,FR,RF} \ -n \${samplename} -r \${replicate} -o \${outpath} -t \${threads} "\$genomeVersion" represents the configuration file for specific genome assembly version created in "Equipment setup" session. "--fq" and "--fq2" specify the trimmed paired-end FASTQ files used as input for mapping. The "--rna-strandness" parameter defines the strand-specific information required by HISAT2 during alignment (refer to: <https://ccb.jhu.edu/software/hisat2/manual.shtml> for details). This step utilizes the specified number of computational threads (\${threads}) and automatically generates four sub-directories within the designated output directory (\${outpath}). Intermediate mapping and the final merged BAM files will be generated and stored within these sub-directories during the alignment process.

22 Fine-tuning alignments for RNA editing discovery

- 22.1 Following the GATK Best Practices workflow for RNA-seq short variant discovery (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels>), we customized and integrated key pre-processing steps — MarkDuplicates, SplitNCigarReads, and Base Quality Score Recalibration (BQSR) — into a single streamlined command. Mapit finetuning -v \$genomeVersion \ -n \${samplename} -r \${replicate} -o \${outpath} This integrated procedure enables efficient data cleanup and fine-tuning for RNA editing detection, directly using the designated output directory (\${outpath}) specified in Step 26 to process the final merged BAM file of "\${samplename}_\${replicate}".

23 RNA editing analysis (4–6 h)

- 23.1 Calling editing sites and filtering out known SNPs This step uses the fine-tuned BAM files to call single nucleotide variants (SNVs) with GATK HaplotypeCaller, followed by filtering out known SNPs based on previously curated SNP datasets. The remaining A-to-G or C-to-U editing sites are then annotated using genome annotation files. Mapit callediting -v \$genomeVersion \ --sampleList \${sample1},\${sample2},...,\$ \${sampleN} \ -o \${outpath} --prefix \${prefix} \ --enzyme {ADAR,APOBEC,Both} -t \${threads} This step allows for the simultaneous analysis of multiple samples. By providing a comma-separated list of sample names in the "--samplelist" parameter, replicate IDs from multiple samples in the designated output directory (\${outpath}) specified in Step 26–27 can be processed together. The parameter "--enzyme", limited to three words, represents the editing

enzyme introduced in MAPIT-seq samples. And after filtering, editing information will output in "\${outpath}/6-Edit_calling" directory with \${prefix} as the file prefix.

- 23.2 Differential editing analysis between RBP-MAPIT and IgG-MAPIT To identify RBP targets, the step uses a Wilcoxon signed-rank test to perform on the 'editing index' within 50-nt continuous and non-overlapping windows of transcripts between RBP-MAPIT and IgG-MAPIT samples. Mapit calltargets -v \$genomeVersion \ -i \${outpath}/6-Edit_calling/\${prefix}/\${prefix}_Edit_DATA.tsv \ --treatName \${treatSampleName} --controlName \${ctrlSampleName} \ -o \${outpath} -c \${coverage} -l \${level} "\${treatSampleName}" and "\${ctrlSampleName}" represent the abbr. sample name of RBP-MAPIT and IgG-MAPIT in the designated output directory (\${outpath}) specified in Step 26–28. And "\${coverage}" is the minimum coverage reads for an effective editing site (default 10), and "\${level}" represents the level performing in differential editing analysis. The step will output the "Differential editing analysis" results in the "\${outpath}/6-Edit_calling/\${prefix}_\${treatSampleName}/" directory.

24 Identifying high-confidence editing clusters and RBP binding motifs (6–8 h)



- 24.1 Identifying A-to-G and C-to-U editing using a modified SAILOR workflow for MAPIT Mapit prepare -v \$genomeVersion -n \${samplename} -r \${replicate} \ -o \${outpath} This step is for preparing files needed in SAILOR and FLARE workflow. Mapit SAILOR -v \$genomeVersion -n \${samplename} -r \${replicate} \ -c \${coverage} -o \${outpath} -t \${threads} This step integrated SAILOR workflow for MAPIT data and result directory. And "\${coverage}" is the minimum coverage reads for an effective editing site (default 10).
- 24.2 Identifying A-to-G or C-to-U editing clusters using FLARE Mapit FLARE -v \$genomeVersion -e {AG,CT} -n \${samplename} \ -r \${replicate} --regions \${regions} -o \${outpath} -t \${threads} This step integrated FLARE workflow for MAPIT data and result directory. And "\${regions}", created in configuration step, is the directory of genomic regions in which cluster identification will occur.
- 24.3 Identifying high-confidence editing clusters Mapit hc_cluster -v \$genomeVersion -n \${samplename} -o \${outpath} \ -s \${sloplength} This step identifies intersection C-to-U or A-to-G editing clusters between two replicates. And high-confidence MAPIT editing clusters are the intersection of two-type editing clusters and are extended or shrink into "\${sloplength}" nt windows.
- 24.4 De novo RBP binding motifs finding hc_cluster_path=\${outpath}/7-FLARE_for_MAPIT/confident_clusters
hc_cluster_bed=\${hc_cluster_path}/\${samplename}/\${samplename}_\${slop}.bed
output_motif=\${hc_cluster_path}/\${samplename}/\${samplename}.motif
findMotifsGenome.pl \$hc_cluster_bed hg38 \$output_motif \ -noknown -rna -size given -



len 5,6,7,8 -p \${threads} This step used the high-confidence MAPIT editing clusters to perform De novo RBP binding motifs finding.