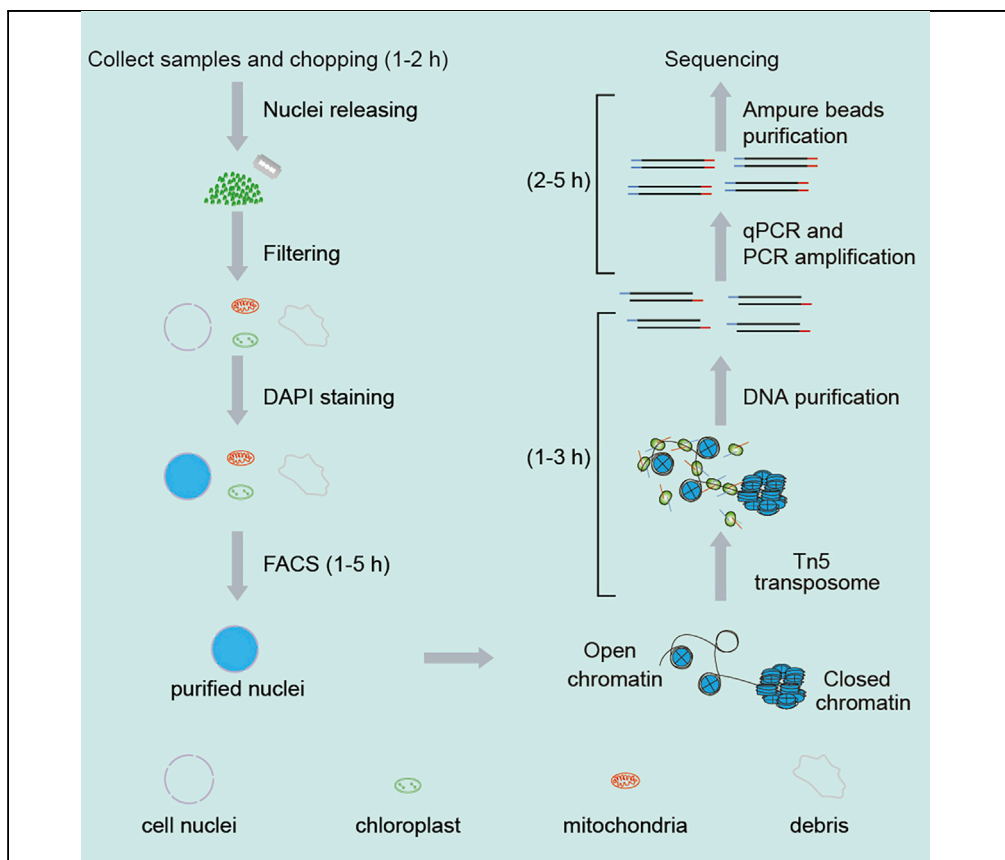


## Protocol

# Protocol for assaying chromatin accessibility using ATAC-seq in plants



Fu-Xiang Wang,  
Guan-Dong Shang,  
Lian-Yu Wu, Yan-Xia  
Mai, Jian Gao,  
Zhou-Geng Xu,  
Jia-Wei Wang

jwwang@sippe.ac.cn

### HIGHLIGHTS

A protocol to generate ATAC-seq libraries from fresh plant tissues

An easy-to-use bioinformatic analysis pipeline for ATAC-seq

Identification of differentially accessible peaks by ATAC-seq

Open or accessible regions of the genome are the primary positions of binding sites for transcription factors and chromatin regulators. Transposase-accessible chromatin sequencing (ATAC-seq) can probe chromatin accessibility in the intact nucleus. Here, we describe a protocol to generate ATAC-seq libraries from fresh *Arabidopsis thaliana* tissues and establish an easy-to-use bioinformatic analysis pipeline. Our method could be applied to other plants and other tissues and allows for the reliable detection of changes in chromatin accessibility throughout plant growth and development.

Wang et al., STAR Protocols 2, 100289

March 19, 2021 © 2020 The Author(s).

<https://doi.org/10.1016/j.xpro.2020.100289>



## Protocol

## Protocol for assaying chromatin accessibility using ATAC-seq in plants

Fu-Xiang Wang,<sup>1,2,4</sup> Guan-Dong Shang,<sup>1,2,4</sup> Lian-Yu Wu,<sup>1,3</sup> Yan-Xia Mai,<sup>1</sup> Jian Gao,<sup>1,2</sup> Zhou-Geng Xu,<sup>1,2</sup> and Jia-Wei Wang<sup>1,3,5,\*</sup>

<sup>1</sup>National Key Laboratory of Plant Molecular Genetics (NKLPMG), CAS center for Excellence in Molecular Plant Science, Institute of Plant Physiology and Ecology (SIPPE), Chinese Academy of Sciences (CAS), Shanghai 200032, China

<sup>2</sup>University of Chinese Academy of Sciences (UCAS), Shanghai 200032, China

<sup>3</sup>ShanghaiTech University, Shanghai 200031, China

<sup>4</sup>Technical contact

<sup>5</sup>Lead contact

\*Correspondence: [jwwang@sippe.ac.cn](mailto:jwwang@sippe.ac.cn)  
<https://doi.org/10.1016/j.xpro.2020.100289>

## SUMMARY

Open or accessible regions of the genome are the primary positions of binding sites for transcription factors and chromatin regulators. Transposase-accessible chromatin sequencing (ATAC-seq) can probe chromatin accessibility in the intact nucleus. Here, we describe a protocol to generate ATAC-seq libraries from fresh *Arabidopsis thaliana* tissues and establish an easy-to-use bioinformatic analysis pipeline. Our method could be applied to other plants and other tissues and allows for the reliable detection of changes in chromatin accessibility throughout plant growth and development.

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

## BEFORE YOU BEGIN

## Experimental considerations

1. Here, we describe a detailed ATAC-seq protocol as applied to embryos and seedlings in *Arabidopsis thaliana*. This protocol can also be used for other plants and other tissues. Before beginning this protocol, please ensure that your core facility has the instruments compatible for flow sorting as the sorting step is extremely important for acquiring high-quality data.
2. To ensure success, please perform pilot experiments to detect the position of nuclei with different ploidy levels in the peak diagram and to set appropriate gates for FACS sorting.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Gamborg B-5 basal medium	Phyto Technology	Cat#G398
Sucrose	ABCONE	Cat#57501
2-(N-morpholino) ethanesulfonic acid (MES)	BBI Life Sciences	Cat#145224948
Phytigel	Sigma-Aldrich	Cat#7101052-1
Agar Bacteriological Grade	Shanghai Jiafeng	Cat#H8145

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Triton X-100	Sigma-Aldrich	Cat#9002931
KCl	Sigma-Aldrich	Cat#7447407
NaCl	Sigma-Aldrich	Cat#7647145
MgCl <sub>2</sub> ·6H <sub>2</sub> O	Sigma-Aldrich	Cat#7791186
Tris base	Sigma-Aldrich	Cat#77861
DAPI	AAT Bioquest	Cat#28718903
2-Mercaptoethanol	Ruibio	Cat#60242
Spermine	Sigma-Aldrich	Cat#85590
Eva green dye	Biotium	Cat#31000
Phusion high-fidelity DNA polymerase	Thermo Fisher Scientific	Cat#022021
Complete protease inhibitor cocktail	Merck	Cat#04693132001

**Critical commercial assays**

MinElute Reaction Cleanup Kit	QIAGEN	Cat#28204
TruePrep DNA Library Prep Kit v2	Vazyme Biotech	Cat#TD50102
TruePrep Index Kit v2	Vazyme Biotech	Cat#TD202
2× NEBNext high-fidelity PCR mix	New England Biolabs	Cat#M0541L
AMPure beads	Beckman	Cat#A63880

**Deposited data**

ATAC-seq, ChIP-seq, and RNA-seq experiment data	This paper	BioProject <a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJCA002620">PRJCA002620</a> , Beijing Institute of Genomics Data Center ( <a href="http://bigd.big.ac.cn">http://bigd.big.ac.cn</a> )
---	------------	--

**Experimental models: organisms/strains**

<i>A. thaliana</i> : Col-0	N/A	N/A
----------------------------	-----	-----

**Software and algorithms**

R version 3.6	The R Foundation	RRID: SCR_001905 <a href="https://www.r-project.org/">https://www.r-project.org/</a>
Adobe Photoshop CC 2018	Adobe Acrobat	N/A
Adobe Illustrator CC 2018	Adobe Acrobat	N/A
Fastp	(Chen et al., 2018)	RRID: SCR_016962 <a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a>
FastQC v0.11.7	FastQC	RRID:SCR_014583 <a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
MultQC v1.6	(Ewels et al., 2016)	<a href="https://multiqc.info/">https://multiqc.info/</a>
Bowtie2 v2.3.4.3	(Langmead and Salzberg, 2012)	<a href="http://bowtie-bio.sourceforge.net/bowtie2/">http://bowtie-bio.sourceforge.net/bowtie2/</a>
Samtools v1.9	(Li et al., 2009)	RRID: SCR_002105 <a href="http://www.htslib.org/">http://www.htslib.org/</a>
sambamba v0.6.7	(Tarasov et al., 2015)	<a href="https://lomereiter.github.io/sambamba/">https://lomereiter.github.io/sambamba/</a>
bedtools v2.25.0	(Quinlan and Hall, 2010)	RRID: SCR_006646 <a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
deepTools v3.1.2	(Ramirez et al., 2014)	RRID: SCR_016366 <a href="https://deeptools.readthedocs.io/en/develop/">https://deeptools.readthedocs.io/en/develop/</a>
featureCounts v1.6.2	(Liao et al., 2014)	RRID: SCR_012919 <a href="http://bioinf.wehi.edu.au/featureCounts/">http://bioinf.wehi.edu.au/featureCounts/</a>
DESeq2 v1.26.0	(Love et al., 2014)	RRID: SCR_015687 <a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
MACS2 v2.1.2	(Zhang et al., 2008)	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
DiffBind v2.14.0	(Ross-Innes et al., 2012)	RRID: SCR_012918 <a href="https://bioconductor.org/packages/release/bioc/html/DiffBind.html">https://bioconductor.org/packages/release/bioc/html/DiffBind.html</a>
ChIPseeker v1.22.0	(Yu et al., 2015)	<a href="https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html">https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html</a>
HOMER v4.10	(Heinz et al., 2010)	RRID: SCR_010881 <a href="http://homer.ucsd.edu/homer/">http://homer.ucsd.edu/homer/</a>
Intervene v0.6.1	(Khan and Mathelier, 2017)	<a href="https://intervene.readthedocs.io/en/latest/">https://intervene.readthedocs.io/en/latest/</a>
Integrative Genomics Viewer	(Robinson et al., 2011)	RRID: SCR_011793 <a href="http://software.broadinstitute.org/software/igv/">http://software.broadinstitute.org/software/igv/</a>

**Other**

Eppendorf realplex2	Eppendorf	Cat#A248709R
The BD FACSAria II	BD	N/A
Percival chamber	Percival	N/A

### MATERIALS AND EQUIPMENT

The lysis buffer and wash buffer can be stored at 4°C for several months. Please add spermine and 2-mercaptoethanol before experiments.

Lysis buffer		
Reagent	Final concentration	Amount
Tris-HCl pH 7.5 (1 M)	15 mM	0.75 mL
NaCl (5 M)	20 mM	0.2 mL
KCl (2.5 M)	80 mM	1.6 mL
2-Mercaptoethanol (14.3 M)	5 mM	17.5 $\mu$ L
Spermine (1 M)	0.5 mM	25 $\mu$ L
Cocktail (100 $\times$ )	1 $\times$	0.5 mL
Triton X-100 (100%)	0.2%	0.1 mL
ddH <sub>2</sub> O		47.3 mL
<b>Total</b>	<b>n/a</b>	<b>50 mL</b>

Wash buffer		
Reagent	Final concentration	Amount
Tris-HCl pH 8.0 (1 M)	10 mM	0.5 mL
MgCl <sub>2</sub> (1 M)	5 mM	0.25 mL
Cocktail (100 $\times$ )	1 $\times$	0.5 mL
ddH <sub>2</sub> O		49.25 mL
<b>Total</b>	<b>n/a</b>	<b>50 mL</b>

**△ CRITICAL:** Both 2-mercaptoethanol and spermine have pungent odors, please add them in a fume hood.

**Alternatives:** This protocol describes cell nuclear FACS using a BD FACSAria II machine, any other FACS machine can be used as well. The FACS parameters used here are: nozzle size (70  $\mu$ m), psi (70), sorting speed (1), sheath fluid (1  $\times$  PBS).

### STEP-BY-STEP METHOD DETAILS

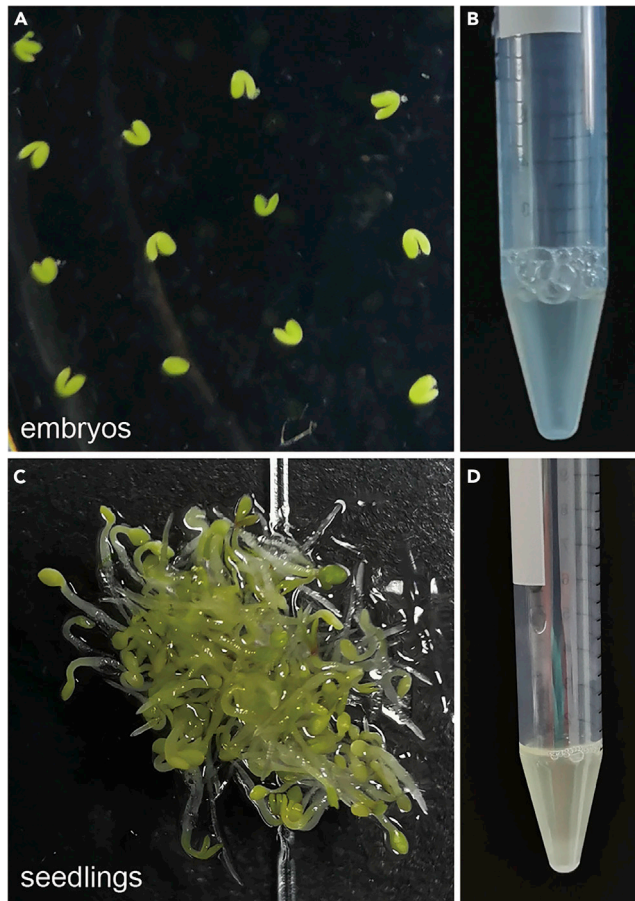
#### Extraction of cell nuclei

⌚ **Timing:** 1–2 h

This step describes the procedure for preparation of the nuclei suspension. Steps 1 and 2 are used for preparation of cell nuclei from embryos, while steps 3 and 4 are used for seedlings.

1. Approximately 200 embryos at the late-cotyledon stage were collected with a 1 mL injection syringe under a dissecting microscope at 25°C. The collected embryos were placed on a 3.5 cm diameter round dish supplemented with B5 agar media (Figure 1A). The B5 agar will keep the embryos alive. The dissected embryos were transferred into a 12 cm long square dish with 500  $\mu$ L pre-chilled lysis buffer on ice.
2. Finely chop embryos with a double-sided blade to release nuclei on ice in fume hood.

**Note:** The embryos naturally settle at the bottom of the lysis buffer, so we can collect the materials together and repeatedly cut them. Due to the small volume of embryos, please cut them repeatedly to release more nucleus until the color of the lysis buffer is yellow-green



**Figure 1. Generation of nuclei suspension**

- (A) The *Arabidopsis* embryos at the late-cotyledon stage were collected on the B5 agar.  
 (B) The nuclei lysis suspension of embryos after filtering.  
 (C) The 3-day-old *Arabidopsis* seedlings were collected on the B5 agar.  
 (D) The nuclei lysis suspension of seedlings after filtering.

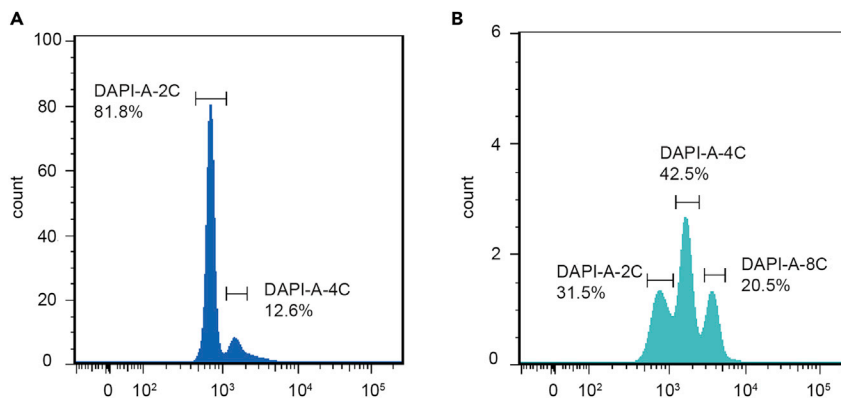
(Figure 1B), which should take approximately 10 min. 200 embryos should provide a sufficient number of nuclei.

3. Approximately 3-day-old 150 - 200 seedlings grown on B5 agar were collected with forceps at RT. The seedlings were transferred to a 12 cm long square dish with 500  $\mu$ L pre-chilled lysis buffer on ice.
4. Finely chop seedlings with a double-sided blade to release nuclei on ice in fume hood.

**Note:** For seedlings, 5 min of chopping is enough. The color of the lysis buffer will be bright green after chopping.

5. Filter the slurry through a 40  $\mu$ m filter into 15 mL collection tubes on ice. Rinse the petri dish with 1.5 mL lysis buffer and filter into the same 15 mL collection tube. The final nuclei suspension is about 2 mL in volume and the color should appear yellow-green.

**△ CRITICAL:** For acquiring more intact nuclei, we recommend extensively chopping plant materials in lysis buffer rather than grinding them in liquid nitrogen which leads to more tissue fragments that will disturb FACS.



**Figure 2. Gate set for FACS**

(A) FACS histogram of embryo nuclei suspension. The DAPI-A-2C and 4C represent cell ploidy with 2 times and 4 times, respectively. Only DAPI-A-2C cell nuclei was sorted.

(B) FACS histogram of seedling nuclei suspension. The DAPI-A-2C, -4C, and -8C represent cell ploidy with 2 times, 4 times, and 8 times, respectively. All DAPI-A-2C, -4C, and -8C cell nuclei were sorted.

### FACS of cell nuclei

© Timing: 1–5 h

This step describes the procedure for sorting nuclei by FACS.

- The nuclei suspensions are stained with DAPI for flow cytometric analysis. Set up an unstained control without DAPI.

**Note:** It is not necessary to estimate the density of nuclei suspension in this step because we will sort ~50,000 nuclei per replicate to ensure repeatability. The amount of nuclei suspension required for control samples is flexible. The purpose of control is to determine the appropriate voltage needed to exclude the interference of other particles.

- Add 2  $\mu\text{L}$  of 1 mg/mL DAPI to the nuclei suspension in a sample tube, flick gently to mix.
- Select parameters including FSC-A (forward scatter-A, particle size), SSC-A (side scatter-A, internal complexity), DAPI-A (intensity of DAPI staining), and DAPI-W (voltage pulse width).
- Use the unstained control and stained samples to set appropriate photomultiplier tube (PMT) voltages and adjust compensations on a BD FACSAria II cell analyzer.
- Create gates for the samples according to the position of 2C or 4C nuclei (Figure 2).
- Add 500  $\mu\text{L}$  of lysis buffer to a 10 mL BD collect tube and collect a total of 50,000 events per sample (each sample per tube). The nozzle size, psi, and sorting speed of FACS are 70  $\mu\text{m}$ , 70, and 1, respectively.

**Note:** 200 embryos and 150–200 3-day-old seedlings are enough material for getting 50,000 nuclei.

- Centrifuge at  $1,000 \times g$  for 10 min at  $4^\circ\text{C}$ .

**Note:** Since we cannot see the pellet at the bottom of the tube after centrifugation, please leave ~50  $\mu\text{L}$  of liquid to avoid nuclei loss when discarding the supernatant.

- Pipette 1 mL of pre-chilled wash buffer into the centrifuge tube.
- Centrifuge at  $1,000 \times g$  for 10 min at  $4^\circ\text{C}$  and discard the supernatant as much as possible.

**Note:** Immature embryo cells at late-cotyledon developmental stage rarely divide. DAPI-A detected nuclei almost with 2C from FACS (Figure 2A). The nuclei were sorted into lysis buffer as plant nuclei do not remain intact in PBS buffer that is used for FACS.

### ATAC-seq library constructing

⌚ Timing: 5–8 h

This step describes the procedure for generation of ATAC-seq library.

We use the TruePrep DNA Library Prep Kit V2 for Illumina including 5× TTBL and TTE Mix V50 (Vazyme, TD50102; see manual at [http://www.vazymebiotech.com/products\\_detail/productId=70.html](http://www.vazymebiotech.com/products_detail/productId=70.html)) to construct ATAC-seq libraries. We use the MinElute Reaction Cleanup Kit including Buffer PB, PE and EB (Qiagen) to purify transposed DNA.

15. Prepare the transposition master mix in a 200  $\mu$ L Lobind tube (Eppendorf) on ice. Add reagents in the order shown below.

Reagent	Amount
5× TTBL	10 $\mu$ L
TTE Mix V50	5 $\mu$ L
Cocktail (100×)	0.5 $\mu$ L
Nuclei suspension	x $\mu$ L
ddH <sub>2</sub> O	34.5–x $\mu$ L
<b>Total</b>	<b>50 <math>\mu</math>L</b>

**Note:** The volume of nuclei suspension after step 14 is variable. Please sure that the remaining nuclear suspension is less than 34.5  $\mu$ L since the total reaction volume for transposition is 50  $\mu$ L. It is not necessary to optimize reaction conditions (e.g., transposition time) when working with different initial volumes or numbers of nuclei.

16. Dispense master mix into a PCR tube containing nuclear suspension. Pipette to mix.
17. Load the tube into a thermal cycler pre-warmed to 37°C and incubate for 30 min with occasional gentle mixing to keep the nuclei in suspension.
18. Add 3–5 volumes of Buffer PB to 1 volume of the PCR sample.
19. Transfer the mixture to a spin column that is situated in a provided 2 mL collection tube.
20. Centrifuge at 13,523  $\times$  g for 1 min at 4°C.
21. Discard the flow-through and add 750  $\mu$ L of Buffer PE to the spin column.
22. Centrifuge at 13,523  $\times$  g for 1 min at 4°C.
23. Discard the flow-through and centrifuge the column for an additional 2 min.
24. Place column in a clean 1.5 mL DNA Lobind microcentrifuge tube and add 11  $\mu$ L of buffer EB to the center of the column for 2 min.
25. Centrifuge at 13,523  $\times$  g for 1 min at 4°C.

⏸ **Pause point:** Samples can be safely stored overnight at 4°C or –20°C.

We use the TruePrep DNA Index Kit V2 for Illumina sequencing (Vazyme) and NEBNext high-fidelity PCR mix to amplify transposed DNA.

26. Prepare the PCR amplification mix in a 0.2 mL PCR tube on ice as follows:

Reagent	Amount
Transposed DNA	10 $\mu$ L
N5 Primer	2.5 $\mu$ L
N7 Primer	2.5 $\mu$ L
2 $\times$ NEBNext high-fidelity PCR mix	25 $\mu$ L
ddH <sub>2</sub> O	10 $\mu$ L
<b>Total</b>	<b>50 <math>\mu</math>L</b>

27. Incubate the reaction in a thermo cycler with the following PCR program.

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Pre-incubation	72°C	5 min	1
Initial denaturation	98°C	30 s	1
Denaturation	98°C	10 s	5 cycles
Annealing	63°C	30 s	
Extension	72°C	1 min	
Hold	4°C	Forever	

28. To determine the number of additional PCR cycles needed to adequately amplify the DNA library, prepare the quantitative PCR (qPCR) Library Amplification Mix as follows:

Reagent	Amount
Amplified library	5 $\mu$ L
N5 Primer	0.5 $\mu$ L
N7 Primer	0.5 $\mu$ L
2 $\times$ NEBNext high-fidelity PCR mix	7.5 $\mu$ L
20 $\times$ Eva green dye	0.75 $\mu$ L
ddH <sub>2</sub> O	0.75 $\mu$ L
<b>Total</b>	<b>15 <math>\mu</math>L</b>

29. Incubate the reaction in a qPCR thermocycler with the following PCR program.

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	10 s	20 cycles
Annealing	63°C	30 s	
Extension	72°C	1 min	

30. To determine the optimal number of cycles needed to amplify the remaining 45  $\mu$ L of each library, view the linear fluorescence versus cycle number plot on the qPCR machine once the reaction is finished. The cycle number at which the fluorescence for a given reaction is at 1/3 of its maximum is the number of additional cycles (N) that each library requires for adequate amplification (Figure 3).



31. Incubate the remaining reaction in a thermocycler with the following PCR program.

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	10 s	N cycles
Annealing	63°C	30 s	
Extension	72°C	1 min	
Hold	4°C	Forever	

32. Add 67.5  $\mu$ L (1.5 $\times$  of sample volume) resuspended Ampure XP beads to the PCR reaction. Mix well by pipetting up and down at least 10 times. Alternatively, samples can be mixed by vortexing for 3–5 s.

**Note:** If Ampure XP beads are stored at 4°C, please pre-warm to 25°C for at least 30 min and vortex before use.

33. Incubate samples on the bench for at least 5 min at 25°C.
34. Place the tube on an appropriate magnetic stand to separate the beads from the supernatant. Quickly centrifuge the sample to collect the liquid from the sides of the tube before placing it on the magnetic stand if necessary.
35. After 5 min, carefully remove and discard the supernatant. Be careful not to disturb the beads that contain bound DNA.
36. Add 200  $\mu$ L of freshly prepared 80% ethanol to the tube while in the magnetic stand.
37. Incubate at 25°C for 30 s, and carefully remove and discard the supernatant.
38. Repeat steps 34 and 35 once for a total of two washes.

**Note:** Be sure to remove all visible liquid after the second wash. Briefly centrifuge the tube on a benchtop centrifuge if necessary.

39. Air-dry the beads for up to 5 min while the tube is on the magnetic stand with the lid open.

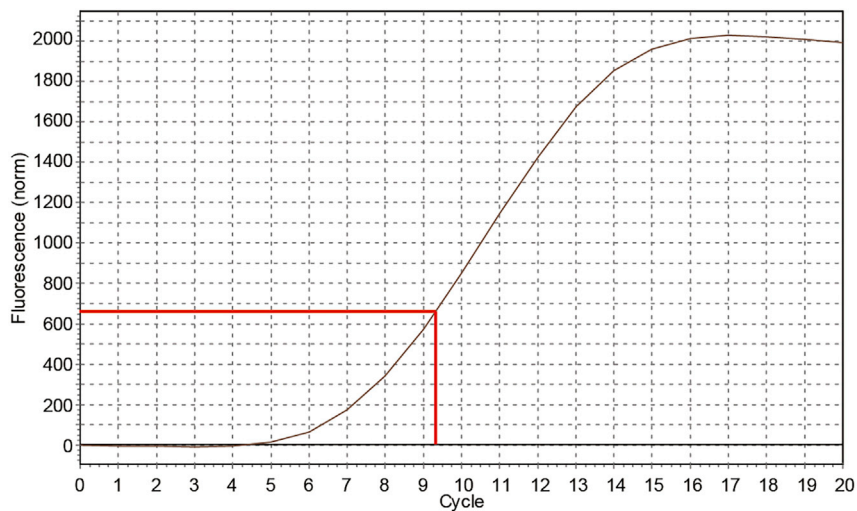
**Note:** Do not over-dry the beads. This may result in a lower recovery of DNA. Elute the samples when the beads are still dark-brown and glossy looking, but when all visible liquid has evaporated. When the beads turn lighter brown and start to crack, it shows beads are too dry.

40. Remove the tube from the magnetic stand. Elute the DNA from the beads by adding 30  $\mu$ L EB buffer provided by Qiagen.
41. Mix well by pipetting up and down 10 times, or on a vortex mixer. Incubate for at least 2 min at room temperature. Quickly spin the sample to collect the liquid from the sides of the tube if necessary.
42. Place the tube on the magnetic stand for 5 min. Transfer 28  $\mu$ L to a new 1.5 mL tube. Libraries can be stored at –20°C. The concentration of the library can be checked by NanoDrop.

### ATAC-seq library quality assessment

This step describes the procedure for assessment of ATAC-seq library.

We pipette 20  $\mu$ L of library samples and send them to Novogene for library quality determination and sequencing (Figure 4). The library quality platform: NGS3K; Sequencing platform: HiSeqPE150, paired-end.



**Figure 3. Schematic diagram of qPCR**

The maximum of fluorescence is 2,000. As such, one-third of the maximum is ~666, and the number of PCR cycles should be 9 (red line).

**Note:** If we started with 50,000 sorted nuclei, 20  $\mu$ L of library sample is enough for quality assessment and sequencing. Most sequencing companies perform library quality analysis before sequencing.

## EXPECTED OUTCOMES

We performed ATAC-seq on four samples following this protocol. The resultant datasets inform us of the dynamics of accessible chromatin regions during seed germination. First, our result shows the distribution of open chromatin regions along the genome (Figure 5A). Second, the analysis of differences between samples suggest the major variable regions and the principal biological processes encoded therein by gene ontology (GO) enrichment annotation (Figures 5C and 5D). Finally, we inferred putative transcription factor binding sites which are enriched in open regions by HOMER-motif analysis (Figures 5B and 5E).

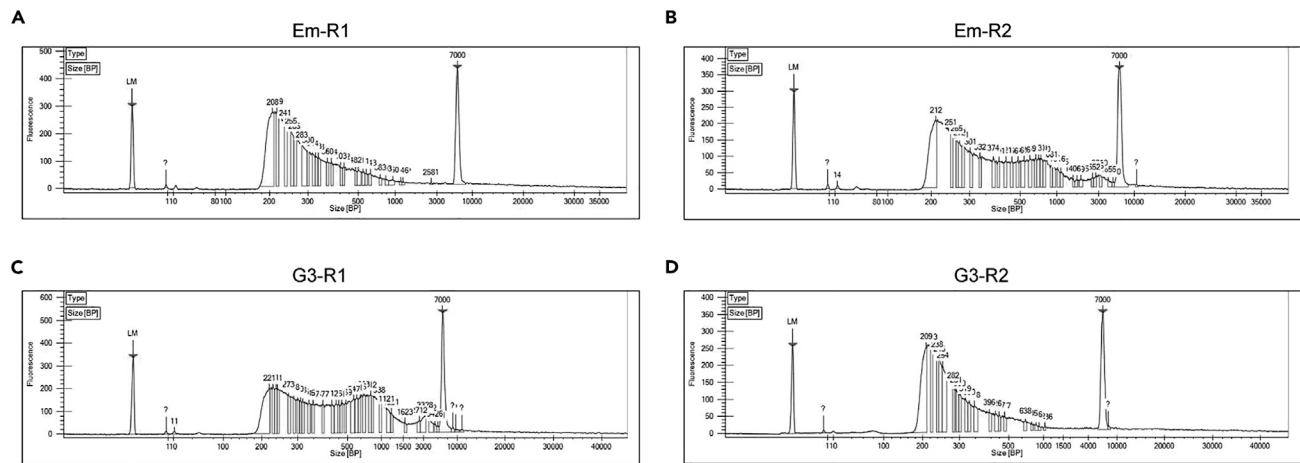
## QUANTIFICATION AND STATISTICAL ANALYSIS

### ATAC-seq analysis pipeline

This step describes the bioinformatic pipeline used for the analysis of ATAC-seq dataset and identification of differentially accessible peaks.

We will describe the basic analysis pipeline of processing and analyzing ATAC-seq including data filtering, quality control, matching, difference analysis, motif prediction, and GO set enrichment analysis as shown below. For a detailed analysis process and related code refer to the following link: [https://github.com/WangLab-SIPPE/2020STARProtocols\\_pipeline](https://github.com/WangLab-SIPPE/2020STARProtocols_pipeline).

1. For each library, raw.fastq was trimmed by fastp v0.20.0 with the parameters "-a CTGTCTCTTA-TACACATCT."
2. After trimming, FastQC v0.11.7 and MulitQC v1.6 were performed as quality control to obtain clean fq files.
3. Reads were aligned using Bowtie2 v2.3.4.3 to the *Arabidopsis* genome (TAIR10). Alternative align tools such as Bwa mean can be chosen.
4. The resulting SAM file containing mapped reads were converted to BAM format, sorted, and indexed using Samtools v1.9.



**Figure 4. Assessment of ATAC-seq library**

(A–D) The purified ATAC-seq DNA libraries of 4 samples on NGS3K platform. Em, immature embryo; G3, 3-day-old seedlings grown on the B5 agar plate.

- The biological replicates were merged by Samtools v1.9.
- The sorted BAM were processed to remove duplicated and organellar reads by sambamba v0.6.7 and bedtools v2.25.0. An alternative tool to mark and remove duplicated reads is Picard.
- To normalize and visualize the individual and merged replicate datasets, the BAM files were converted to bigwig using bamCoverage provided by deepTools v3.1.2 with a bin size of 10 bp and normalized by Bin Per Million mapped reads.
- MACS2 v2.1.2 was used to call peaks with default parameters "-t ATAC.bam -f BAMPE -g 1.1e8."
- For differentially accessible region analyses, we used "DiffBind" with the parameter "minOverlap = 1." "DiffBind" was also used to calculate merged peak locations, based on the outer boundaries of overlapping peaks from all the analyzed samples.
- The peaks called by MACS2 and differential peaks identified by DiffBind v2.14.0 were annotated by TxDb.Athaliana.BioMart.plantsmart28 and ChIPseeker v1.22.0 package with the "annotatePeak" function. The gene promoters are defined as  $\pm 3.0$  kb from the transcription start site (TSS).
- The differential regions were scanned for the enrichment of motifs using the "findMotifsGenome.pl" function provided by HOMER v4.10 with default parameters.
- The GO and Gene Set Enrichment Analysis of differentially accessible regions were performed using clusterProfiler v3.14.0 and org.At.tair.db v3.10.0.

## LIMITATIONS

This protocol is designed for fresh plant materials. If the materials are stored at  $-80^{\circ}\text{C}$  or in liquid nitrogen, it will lead to a high amount of tissue debris, which can interfere with the flow sorting of intact nuclei.

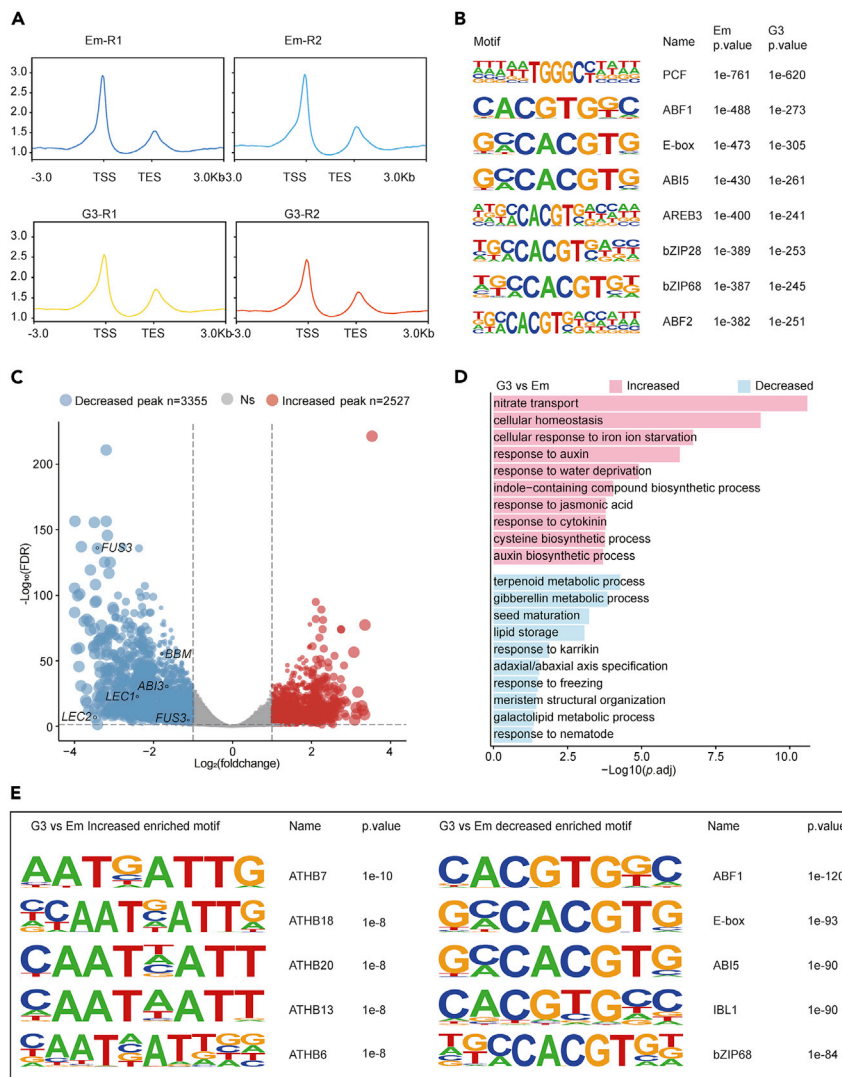
## TROUBLESHOOTING

### Problem 1

Sequenced reads have a low mapping rate.

### Potential solution

There are several reasons that can cause low mapping rates. Please ensure that all solutions, equipment, and work areas are free of DNA. Wear a mask and gloves when chopping samples and constructing libraries. Avoid excessive PCR cycles during amplification. In general, 12–17 cycles are sufficient.



**Figure 5. Results of ATAC-seq analysis**

(A) The genome-wide distribution of ATAC-seq peaks. Window size: gene body  $\pm$  3.0 kb.  
 (B) HOMER DNA-motif enrichment analyses of accessible peaks. The enrichment of binding motifs of 8 transcription factors per sample are shown.  
 (C) Volcano plot showing the genes associated with decreased (blue) or increased (red) accessible peaks between immature embryos (Em) and 3-day-old seedlings (G3). The known genes related to seed or embryo development are shown. Ns (gray in color), no difference between two samples.  
 (D) GO term analyses showing distinct gene ontologies of target genes linked to differentially accessible peaks (G3 versus Em). The selected 20 enriched GO biological processes are indicated. The  $-\log_{10}(p_{adj})$  is given.  
 (E) HOMER DNA-motif enrichment analyses of differentially accessible peaks (G3 versus Em). The enrichment of binding motifs of 5 transcription factors per sample are shown.

## Problem 2

Poor data quality.

## Potential solution

This might be caused by low quality or a low amount of cell nuclei. Please make sure that fresh samples are used, and an appropriate sorting speed and sheath pressure are selected.

**Problem 3**

Sequence reads do not map primarily to non-coding regions.

**Potential solution**

When you find that mapped sequences were mostly aligned to exonic regions in the reference genome, this might be caused by a low quality of nuclei or exogenous DNA pollution. Low-quality cell nuclei might cause nucleosome instability, which leads to an increase in the mapping rate at non-signal regions. To fix this problem, please make sure that only high-quality cell nuclei are used for the Tn5 tagmentation. As such, the examination of the sorted nuclei under a microscope is strongly recommended. Otherwise, 75% ethanol or DNA cleaner can be used to remove exogenous DNAs, and thereby improving your data quality.

**Problem 4**

The differentially accessible regions are far less as expected.

**Potential solution**

This problem can also be caused by poor data quality. Please check peak numbers, plot profiles, and the correlation between replicates. You can achieve these using the "plotCorrelation" function in deeptools, "plotPCA" function in DESeq2, or the "dba.plotPCA" function in DESeq2. Meanwhile, batch effects should also be considered. You can introduce the batch in the design effect with block parameter "dba.contrast" in DiffBind. Finally, you can try different ways to probe differentially accessible regions. According to a recent report (Reske et al., 2020), different bioinformatic tools may lead to different results.

**RESOURCE AVAILABILITY****Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jia-Wei Wang ([jwwang@sippe.ac.cn](mailto:jwwang@sippe.ac.cn)).

**Materials availability**

This study did not generate new unique materials.

**Data and code availability**

The data can be found online at <https://www.sciencedirect.com/science/article/pii/S1534580720305517>

The analysis code can be found online at [https://github.com/WangLab-SIPPE/2020STARProtocols\\_pipeline](https://github.com/WangLab-SIPPE/2020STARProtocols_pipeline).

**ACKNOWLEDGMENTS**

We thank Hong Zhu for technical support in FACS (Cell Biology Core Facility of Shanghai Center for Plant Stress Biology, Center for Excellence in Molecular Plant Sciences, CAS); members of J.-W.W.'s lab for discussion and comments on the manuscript. This work was supported by National Natural Science Foundation of China grant numbers 31788103, 31525004, and 31721001 to (J.-W.W.) and the Strategic Priority Research Program of the Chinese Academy of Sciences grant number XDB27030101 to (J.-W.W.).

**AUTHOR CONTRIBUTIONS**

F.-X.W., S.-G.D., and Y.-L.W. contributed equally. Y.-X.M. and J.G. helped with FACS. Z.-G.X. helped with bioinformatic analysis. F.-X.W. and S.-G.D. wrote the manuscript. J.-W.W. polished the paper and supervised the study.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### REFERENCES

- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Khan, A., and Mathelier, A. (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* 18, 287.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Methods* 9, 357–359.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–191.
- Reske, J.J., Wilson, M.R., and Chandler, R.L. (2020). ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin* 13, 22.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389–393.
- Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034.
- Wang, F.X., Shang, G.D., Wu, L.Y., Xu, Z.G., Zhao, X.Y., and Wang, J.W. (2020). Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. *Dev. Cell* 54, 742–757.e8.
- Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.