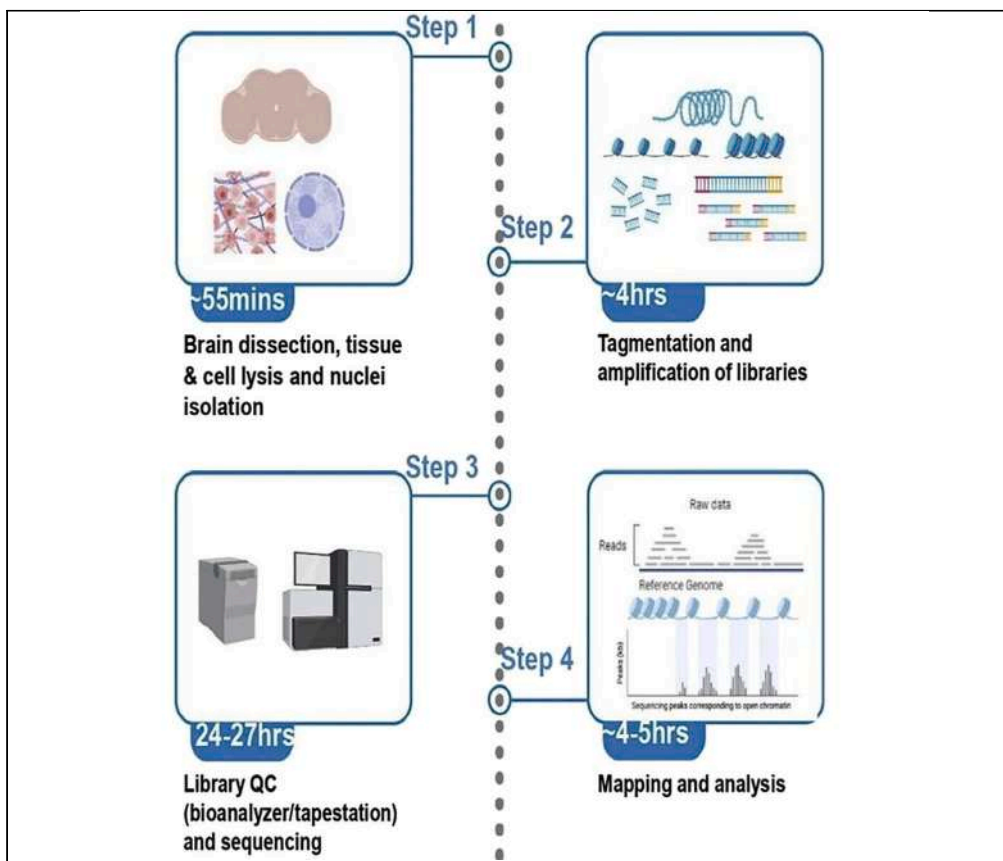


## Protocol

Optimized protocol for assay for transposase-accessible chromatin by sequencing (ATAC-seq) from *Drosophila melanogaster* brain tissue



Transposase-accessible chromatin by sequencing (ATAC-seq) has emerged as an advantageous technique to assess chromatin accessibility owing to the robustness of 'tagmentation' process and a relatively faster library preparation. A comprehensive ATAC-seq protocol from *Drosophila* brain tissue is currently unavailable. Here, we have provided a detailed protocol of ATAC-seq assay from *Drosophila* brain tissue. Starting from dissection and transposition to amplification of libraries has been elaborated. Furthermore, a robust ATAC-seq analysis pipeline has been presented. The protocol can be easily adapted for other soft tissues.

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

Jasmine Kaur Dhall,  
Nandashree  
Kasturacharya,  
Awadhesh Pandit,  
Lakshminarayanan  
CP

jasminek@ncbs.res.in  
(J.K.D.)  
awadhesh@ncbs.res.in  
(A.P.)

**Highlights**  
ATAC-seq protocol  
optimized for  
*Drosophila* brain  
tissue

Freshly dissected  
tissue samples were  
used for nuclei  
isolation

The protocol  
provides a detailed  
workflow from  
dissection to analysis  
pipeline

The protocol can be  
adapted for other soft  
tissue types

Dhall et al., STAR Protocols 4,  
102153  
June 16, 2023 © 2023  
[https://doi.org/10.1016/  
j.xpro.2023.102153](https://doi.org/10.1016/j.xpro.2023.102153)



## Protocol

Optimized protocol for assay for transposase-accessible chromatin by sequencing (ATAC-seq) from *Drosophila melanogaster* brain tissue

Jasmine Kaur Dhall,<sup>1,4,\*</sup> Nandashree Kasturacharya,<sup>1,2</sup> Awadhesh Pandit,<sup>1,3,\*</sup> and Lakshminarayanan CP<sup>1</sup>

<sup>1</sup>National Centre for Biological Sciences, TIFR, Bellary Road, Bengaluru 560065, India

<sup>2</sup>The University of Trans-Disciplinary Health Sciences and Technology (TDU), Bengaluru 560064, India

<sup>3</sup>Technical contact

<sup>4</sup>Lead contact

\*Correspondence: [jasminekd@ncbs.res.in](mailto:jasminekd@ncbs.res.in) (J.K.D.), [awadhesh@ncbs.res.in](mailto:awadhesh@ncbs.res.in) (A.P.)  
<https://doi.org/10.1016/j.xpro.2023.102153>

## SUMMARY

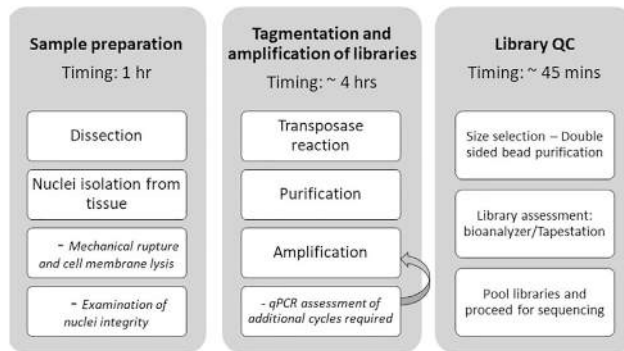
Transposase-accessible chromatin by sequencing (ATAC-seq) has emerged as an advantageous technique to assess chromatin accessibility owing to the robustness of "tagmentation" process and a relatively faster library preparation. A comprehensive ATAC-seq protocol from *Drosophila* brain tissue is currently unavailable. Here, we have provided a detailed protocol of ATAC-seq assay from *Drosophila* brain tissue. Starting from dissection and transposition to amplification of libraries has been elaborated. Furthermore, a robust ATAC-seq analysis pipeline has been presented. The protocol can be easily adapted for other soft tissues.

## BEFORE YOU BEGIN

The packaging of ~2 m DNA in less than 10 μm diameter space is attributed to the tightly regulated 3-D genome topology.<sup>1</sup> The spatial and temporal organization of the chromatin, comprising of chromatin remodelers, histone modifications, transcription factors and non-coding RNAs conform the precision of genome regulation. This organization mediates the enhancer-promoter interactions, envisage transcription factors and regulatory machinery to active regions and sequesters the inactive genomic regions.<sup>2–5</sup> The broad chromatin architecture and partitioning into the topologically associated domains orchestrate gene regulation and thus subsequently control all cellular processes and cell state. To facilitate such transcriptional regulation, plethora of protein complexes bind to chromatin in a spatiotemporal manner that leads to the opening of chromatin. Therefore, the open chromatin regions marks are associated with regulatory regions in the genome.

Further, examination of the chromatin accessibility profile (chromatin openness) and identification of active vs inactive regions mapped with gene expression provides a precise understanding of transcriptional landscape of the cell. Towards this, a plethora of techniques, such as, FAIRE-seq,<sup>6</sup> MNase-seq<sup>7</sup> and DNase-seq<sup>8</sup> have been employed to map the accessibility of regulatory elements. The advent of ATAC-seq assay of transposase accessible chromatin,<sup>9</sup> wherein, a hyperactive Tn5 transposase is utilized to conduct transposition of sequencing adapters at the regions of increased accessibility on chromatin has revolutionized the identification of regulatory elements. ATAC-seq superseded the traditional methods owing to the (1) faster and less complex library preparation based on the 'tagmentation' principle, (2).





**Figure 1. Overview of the ATAC-seq assay protocol**

Suitability for diverse input materials and low starting input material (50,000–60,000 nuclei) requirement. Despite the advantageous preparation, the protocol required systematic optimizations to yield better data quality. Omni-ATAC<sup>10</sup> is one such remarkable advancement of the method, which reduced mitochondrial reads, increased signal to noise ratio and was suitable for library prep from frozen tissues as well. The primary protocol steps of ATAC-seq involve: (i) input/sample preparation (ii) transposase mediated tagmentation (iii) library amplification (iv) sequencing and analysis.

While several detailed protocols are available to perform ATAC-seq in eukaryotic cell lines and mammalian tissue, a detailed protocol on specifically *Drosophila* brain is currently unavailable. Owing to the smaller size, several optimizations are required in the available protocols. Most ATAC-seq studies on *Drosophila* have been conducted from larvae and embryos.<sup>11,12</sup> A comprehensive ATAC-seq examination from *Drosophila* neurons has been carried out utilizing head dissections, followed by FACS sorting using specific GAL4 lines.<sup>13</sup> Here, we provide a detailed protocol of ATAC-seq assay from *Drosophila* brain tissue. First, sample preparation and transposition will be described. Thereafter, library amplification and size selection, pertaining to the assay, will be elaborated (Figure 1). Finally, we elucidate a robust ATAC-seq quality check and analysis pipeline (Figure 5).

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
DPBS	Gibco	D8537
Tris	Thermo Fisher Scientific	15-567-027
NaCl	HiMedia	GRM853-500G
KCl	Merck	104936
Digitonin	Promega	G9441
Tween20	Sigma	P9416-100ML
Nonidet™ P 40 substitute (NP40)	Sigma	74385
UltraPureDNase/RNase-free distilled water	Thermo Fisher Scientific	10977015
Ethanol	Sigma-Emsure	1009832511
Nextera DNA CD indexes	Illumina	20018708
NEBNext® Ultra™ II Q5 master mix	New England BioLabs (NEB)	M0544I
KAPA library quant kit (Illumina)	Roche	07960140001
Qubit™ dsDNA HS kit	Thermo Fisher Scientific	Q32851
Other		
Micro glass slides thickness 1.35 mm	BLUE STAR	NA
AMPure XP beads	Beckman Coulter	A63881
Microcentrifuge	Eppendorf	5430 R

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell Countess II FL	Invitrogen	NA
Confocal microscope -laser scanning	Olympus	FV3000
Thermomixer® C	Eppendorf	NA
Thermocycler	ABI Veriti™	NA
Real-Time PCR system	Applied Biosystems ViiA™ 7	NA
Qubit 4 Fluorometer	Invitrogen	NA
4200 TapeStation system	Agilent	NA
Hi-Seq 2500 sequencer	Illumina	NA
Cell strainer (40 µm)	Falcon	352340

**Experimental models: Organisms/strains**

<i>Drosophila melanogaster</i> : Canton S 2–5 days old flies comprising of both males and females. Grown in standard media comprising of commeal agar at 25°C and 12-h light/dark cycle	Bloomington Drosophila Stock Center (Indiana University, Bloomington, IN, USA)	BL64349
--	---	---------

**MATERIALS AND EQUIPMENT**

**Lysis buffer**

Reagent	Final concentration	Amount
Tris pH 7.5 (2 M)	10 mM	250 µL
NaCl (4 M)	10 mM	125 µL
KCl (3 M)	4 mM	66 µL
NP-40 (10%)	0.2%	1,000 µL
Tween20 (100%)	0.1%	50 µL
Digitonin (high purity) (1%)	0.01%	500 µL
<b>Total</b>	<b>N/A</b>	<b>50 mL</b>

**Note:** Make up the volume with nuclease free water. Buffer can be stored at 4°C for six months. Digitonin leads to permeabilization and disruption of plasma membrane owing to its steroidal saponin nature. It does not affect the membrane of the organelles (due to lower sterol composition).

**Wash buffer**

Reagent	Final concentration	Amount
Tris pH 7.5 (2 M)	10 mM	250 µL
NaCl (4 M)	10 mM	125 µL
KCl (3 M)	4 mM	66 µL
Tween20 (100%)	0.1%	50 µL
<b>Total</b>	<b>N/A</b>	<b>50 mL</b>

**Note:** Make up the volume with nuclease free water. Buffer can be stored at 4°C for six months.

**Transposition reaction mix**

Reagent	Final concentration	Amount
2× Tagment DNA buffer	50%	25 µL
1×PBS	33%	16.5 µL
10% Tween 20	1%	0.5 µL
1% Digitonin	1%	0.5 µL
Tn5 Transposase	5%	2.5 µL
Nuclease free water	10%	5 µL
<b>Total</b>	<b>N/A</b>	<b>50µL</b>

## STEP-BY-STEP METHOD DETAILS

### Sample preparation/nuclei isolation

⌚ Timing: 40 min

The foremost requirement of the assay is obtaining chromatin fragments using transposase while preserving the nuclei integrity during sample processing. Hence, ATAC-seq from tissue samples remains challenging since flash freezing the samples interferes with the library quality,<sup>14</sup> optimization of the assay will vary with respect to the type of tissue and often mechanical agitation is required for tissue rupture. Relatively few ATAC-seq examinations have been conducted in *Drosophila melanogaster*, and primarily focus to attain the regulatory information from embryos and larvae. Sorted neurons from *Drosophila* brains have been subjected to the transposase assay employing genetic labeling with *UAS-Gal4* system.<sup>13</sup> Some ATAC-Seq studies have used as low as 500 nuclei,<sup>9</sup> majorly 50000 nuclei have been posited in mammalian ATAC-seq assays. However, *Drosophila* genome (~180 megabase) is markedly smaller than the mammalian genome. Considering, a *Drosophila* brain comprises of ~100000 neurons,<sup>15</sup> bulk volume adding to the library complexity and some loss during sample preparation, 35 flies of each genotype were used for brain dissection. The current protocol follows the basis of Omni-ATAC method (addition of digitonin and Tween-20; inclusion of a wash step). With minor modifications in lysis buffer composition used previously,<sup>10</sup> the lysis was optimized for the brain tissue. The steps outlined below were used to obtain intact nuclei:

1. Dissect brains from 35 flies in pre-chilled (4°C) 1 × DPBS and add, separately for each genotype, in 1.5 mL DNA LoBind Eppendorf® tubes.
2. Centrifuge the tissue sample at 500 × *g*, 4°C for 5 min.
3. Remove the supernatant and resuspend the tissue in 250 μL pre-chilled lysis buffer.
4. Briefly vortex (15 s) the samples and incubate on ice for 10 min with intermittent pipetting with 200 μL tips to rupture the tissue and lyse the cells simultaneously.

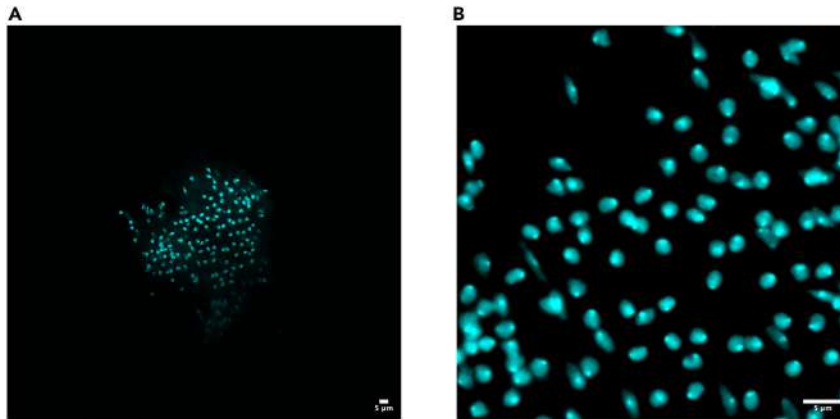
**Note:** Using Dounce homogenizer with the present buffer compositions resulted in nuclear lysis. To use Dounce homogenizer or automated homogenizers, it is recommended to slightly reduce surfactant concentration (0.1% NP-40 and no Tween20) and increase the total amount of sample and volume of lysis buffer (650 μL).

5. Add 250 μL of wash buffer and gently tap the tubes.
6. Examine that the cell membrane has been lysed for more than 90% of cells, as described earlier.<sup>16</sup>
  - a. Briefly, aliquot 5 μL of thoroughly mixed cell suspension in a fresh tube.
  - b. In equal parts (i.e., 1:1/5 μL), add 0.4% trypan blue.
  - c. Mix gently and incubate at RT for 3 min before examining for cell viability.
  - d. Confirm that the cell membrane has lysed for a minimum 90% cells.
7. Pass the suspension through cell strainer (40 μm) in a new 1.5 mL DNA LoBind Eppendorf® tube to remove large fragments of tissue debris.
8. Centrifuge at 500 × *g* for 10 min at 4°C in a fixed angle microcentrifuge.
9. Discard the supernatant (cytoplasm) and keep the pellet (nuclei).

### Examination of nuclei integrity

⌚ Timing: 15 min

All chromatin examination techniques, ranging from ChIP, conformation assays, ATAC-seq, Hi-ChIP etc require the preserved integrity of chromatin architecture during sample preparation. In this case, we have ascertained that the isolated nuclei are intact by examining that the nuclei are spherical and do not exhibit any leaking of DNA.



**Figure 2. Examination of nuclei integrity from an aliquot of isolated nuclei**  
(A and B) DAPI treated nuclei suspension under 20 $\times$  and 60 $\times$  respectively.

10. Gently but thoroughly mix each sample to attain a uniform suspension and then aliquot 10  $\mu$ L into respective fresh tubes.
11. Add 5  $\mu$ L of DAPI (0.2 ng/ $\mu$ L) in 1 $\times$ PBS to 10  $\mu$ L of nuclei suspension and incubate for 15 min at RT.
12. Mount 10  $\mu$ L of DAPI treated suspension on micro glass slide and determine the percentage of intact nuclei (Figures 2A and 2B).
13. Ascertain that more than 90%–95% of the nuclei are intact.

### Transposition

⌚ Timing: 35 min

The advantageous nature of ATAC-seq over other related traditional techniques emerges from the use of Tn5 transposase for the tagmentation based library preparation. The two Tn5 transposase monomers dimerize and form a synaptic complex. The complex then transfers the transposon onto the target DNA using oxygen molecules from the water for nucleophilic action, facilitated by divalent metal ions.<sup>17</sup> Herein, the process involves transposase-based fragmentation and simultaneous tagging of the DNA with adapter sequences. The two main parameters crucial before proceeding for the transposition reaction are (i) nuclei/transposase ratio (ii) genome size. Importantly, it is strongly recommended to keep the nuclei to transposase ratio same between samples, to compare their enrichment levels. We generated the libraries using 75000 nuclei, keeping transposase concentration similar to the mammalian ATAC-assays.

**Note:** Change in the number of nuclei would require optimization of transposition reaction. With 50,000 nuclei, we observed over-tagmentation. This can further result in increased noise. On the other hand, under-tagmentation can result in larger molecular weight fragments. However, if starting material is limited (eg a subset of neurons or segregation of neurons and glial cells is required), then it is recommended to reduce Tn5 Transposase concentration and optimize transposition reaction.

14. Add the freshly prepared transposition reaction mix to the nuclear pellet, followed by gentle pipetting (5–8 times, in 10 s).
15. Place the samples in thermomixer at 37°C for 30 min at 1,000 rpm.

### Purification of DNA fragments

⌚ Timing: 15 min

After transposition, the samples were immediately proceeded for DNA purification using Ampure XP beads (2.5×).

16. Keep AMPure XP beads at RT, for 30 min before use.
17. Vortex the the AMPure XP beads for 30 s to make sure that the beads are evenly distributed.
18. Resuspend and incubate the samples in 2.5× (125 μL) Ampure beads for 5 min at RT.
19. Place the tubes on magnetic stand, until it clears. Carefully remove and discard the clear supernatant without disturbing the beads.
20. Give two washes with 200–300 μL of 80% ethanol, while the beads are adhered towards the magnet. For proper washing, make sure that the beads are completely immersed in ethanol.
21. Carefully remove the leftover ethanol from each sample tube. With the sample tube on the magnetic stand, air-dry the beads for 1 min. Do not over-dry the beads.
22. Elute with 22.5 μL of 10 mM Tris pH 8.0 by leaving beads in Tris for approximately 2 min.
23. Place the tubes on the magnetic stand for 5 min or until the supernatant is clear and carefully transfer 21 μL of the supernatant to a new tube.
24. Using 1 μL of sample, take the qubit reading to ascertain the concentration and quality of the purified DNA.

⏸ **Pause point:** Purified transposed DNA can be stored at –20°C, until further processing, for ~ 15 days.

### PCR amplification and generation of libraries

⌚ Timing: 2 h 15 min

25. PCR reaction was followed as previously described<sup>9</sup> with minor modification, namely, increased volume of purified transposed DNA to increase the yield. PCR for each sample comprised of the following:

PCR reaction master mix	
Reagent	Amount
Purified transposed DNA	20 μL
Nextera™ DNA CD index	5 μL
Q5 High-Fidelity 2× PCR master mix	25 μL
Total	50 μL

**Note:** CD index is unique for each sample.

26. Samples were amplified using the following PCR program

Steps	Temperature	Time	Cycles
Extension	72°C	5 min	1
Initial denaturation	98°C	30 s	
Denaturation	98°C	10 s	5
Annealing	63°C	30 s	
Extension	72°C	1 min	

27. Conduct a q-PCR with 5  $\mu$ L of the library obtained from previous step.

**Note:** The q-PCR step allows monitoring of the amplification (of 5 cycles from PCR) in order to avoid over-amplification/saturation and variation among samples. In this step we used KAPA library quant kit for Illumina (Roche) with minor modifications, as listed below.

Reagent	Amount
2 $\times$ master mix	7.5 $\mu$ L
Template	5 $\mu$ L
Nextera™ DNA CD index	1 $\mu$ L
ROX	0.2 $\mu$ L
NFW	1.5 $\mu$ L
Total	15 $\mu$ L

**Note:** CD index is unique per sample and the same index should be used as used in the previous step (PCR amplification and generation of libraries step 25).

28. Plot linear Rn vs. Cycle.

29. Calculate the number of cycles corresponding to 1/3 of maximum intensity (Figure 3).

30. Continue PCR for each sample/library for respective number of cycles

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	10 s	Respective number of cycles calculated from previous step (step 29)
Annealing	63°C	30 s	
Extension	72°C	1 min	
Hold	4°C	forever	

### Library assessment and size selection

⌚ Timing: 45 min

It is recommended to assess the DNA concentrations in each library, both, pre and post size selection/cleanup.

31. In this step (pre size selection), we used Qubit™ dsDNA HS kit, following manufacturer's protocol for assessing DNA concentration using 1  $\mu$ L of the sample.

**Note:** Pre-size selection assessment of DNA concentration provides a better estimate of volume of sample to be used for further processing to obtain more concentrated samples.

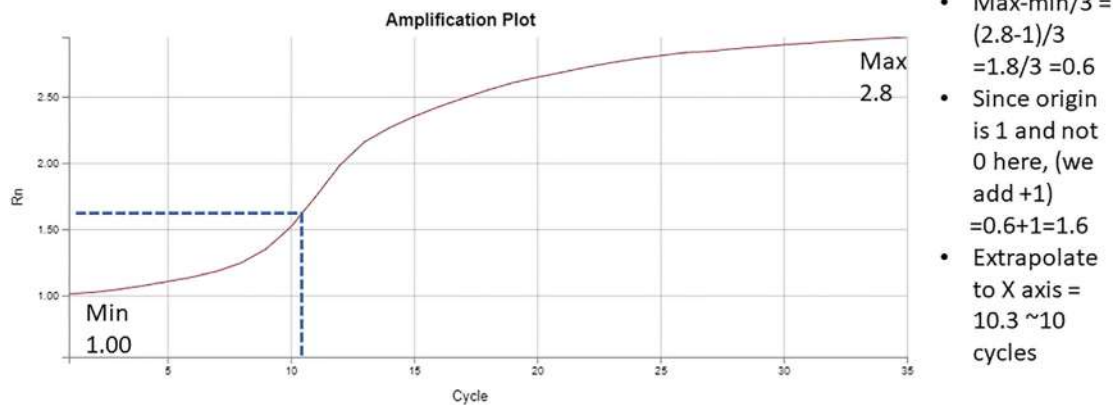
32. Carry out size selection using double-sided bead purification to eliminate adapter dimers (<150 bp) and larger molecular weight fragments (> 1,000 bp).

33. To each sample, added in a fresh 0.5 mL/1.5 mL tube, add 0.5  $\times$  AMPure XP beads. To comment precisely, here from second PCR we have 45  $\mu$ L. Hence, 22.5  $\mu$ L of the AMPure XP beads will be added.

34. Mix thoroughly with gentle pipetting.

35. Incubate at RT for 10 min.

36. Place against magnetic stand until the solution is completely transparent and all bead particles are adhered towards the magnet (usually 3–5 min).



**Figure 3. Examination of amplification plot to determine the additional number of cycles required**

**Note:** Here, fragments more than 1,000 bp in size are adhered to the beads and hence removed.

37. Transfer the supernatant to a new tube and add  $1.3 \times$  (of 45  $\mu\text{L}$  PCR product – i.e., 58.5  $\mu\text{L}$ ) AMPure XP beads.
38. Mix thoroughly with gentle pipetting and incubate for 5 min at RT.
39. Place the tubes on magnetic stand, until it clears. Carefully remove and discard the clear supernatant without disturbing the beads.
40. With the tubes on the magnetic stand, wash the beads by adding 200–300  $\mu\text{L}$  of freshly prepared 80% ethanol to each sample tube. Incubate for 30 s. Carefully remove and discard the supernatant.
41. Repeat the above step for a total of two washes and carefully remove the leftover ethanol from each sample tube.
42. With the sample tube on the magnetic stand, air-dry the beads for 1 min. Do not over-dry the beads.
43. Remove the tubes from magnetic stand and resuspend the beads in 22.5  $\mu\text{L}$  of 10 mM Tris-HCl pH 8.0 and incubate for 2 min at RT.
44. Place the tubes on the magnetic stand for 5 min or until the supernatant is clear and carefully transfer 20  $\mu\text{L}$  of the supernatant to a new tube.

**Pause point:** Libraries can be stored at  $-20^{\circ}\text{C}$  for up to 3 months or  $-80^{\circ}\text{C}$  for long-term storage (The prepared library can be stored for long but should not be subjected to freeze thawing).

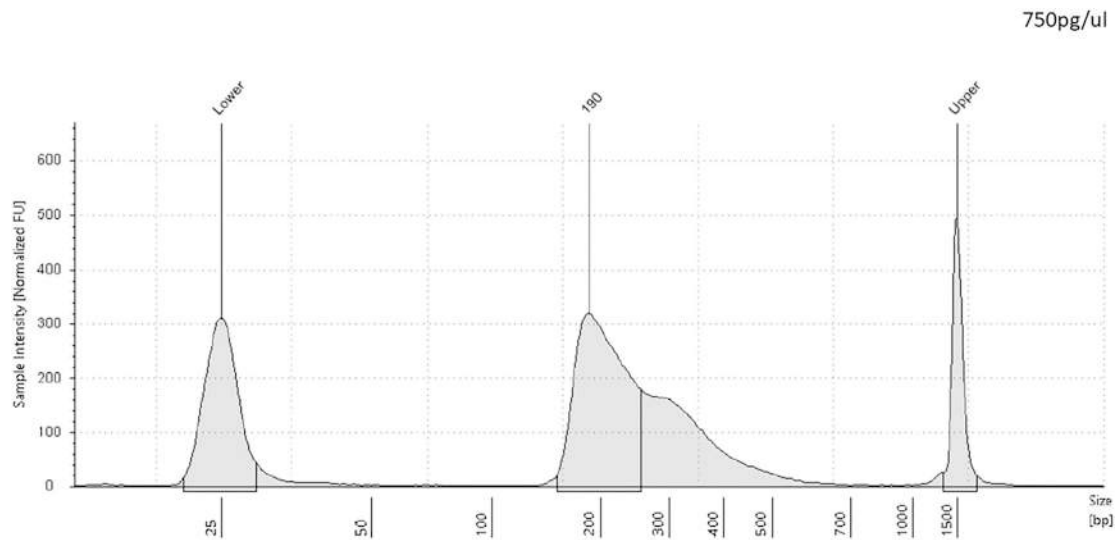
45. Post size selection library assessment:

**Note:** As mentioned earlier, we used Qubit™ dsDNA HS kit, following manufacturer's protocol for assessing DNA concentration. More importantly, to determine the quality, size and integrity of the library, we used Agilent 4200 TapeStation system. Furthermore, the pre-sequencing assessment of the library provides the elucidation of the nucleosomal periodicity.

**CRITICAL:** It is recommended to examine the fragment size distribution at this step (Figure 4).

### Sequencing of the libraries

© Timing: 24–26 h



**Figure 4. Bioanalyzer profile of a CS library depicting fragment size distribution (base pair) with upper and lower size markers**

46. Pool the libraries at equimolar quantities and subject to high depth sequencing. We sequenced 8.5 pM of the denatured pooled library.
47. Sequence the libraries, keeping the sequencing depth ~ 20–50 million reads.

**Note:** Suggested total number of read pairs per sample for examining chromatin accessibility is 20–50 million depending on the genome size. The depth of sequencing in the present study was 40 million. However, for examining nucleosome occupancy and footprint analysis, it is essential to further increase the depth. The scope of this protocol caters to only examination of chromatin accessibility.

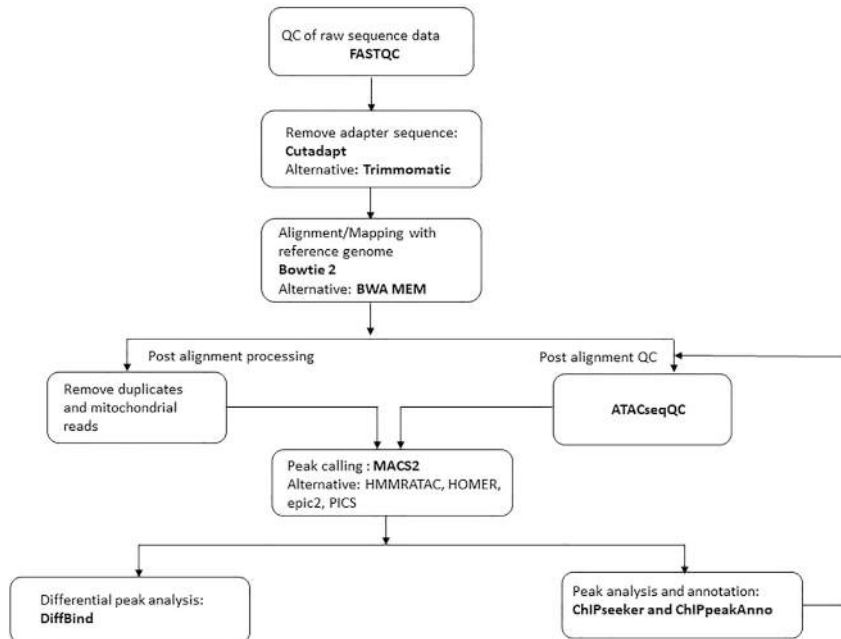
**Note:** In the current protocol, 2 × 100 base pair read length was used for sequencing with Illumina HiSeq 2500. Previously, even 36 base pair read length has been utilized successfully<sup>14</sup>.

**△ CRITICAL:** (1) Only paired-end sequencing should be carried out for ATAC assay since it is essential to cover each nick (on either strand) introduced by the enzyme. (2) It is recommended to carry out a low depth sequencing first and conduct a quality check with respect to TSS enrichment, uniform coverage of the genome, mapping to the genome, percentage of mitochondrial reads and nucleosomal periodicity.

### ATAC-seq data analysis

⌚ Timing: 4–5 h

Even though ATAC-seq has become the ‘method of choice’, there has been a limited advancement in developing specific bioinformatic tools and pipelines for ATAC-seq. Owing to the similarity of the nature of data between ATAC-seq and ChIP-seq, a wide variety of bioinformatic tools available for the latter, have been utilized in ATAC-seq analysis. Recently ATACseqQC has been developed for specific quality checks for ATAC-seq data.<sup>18</sup> Here, in the second part of the protocol (ATAC-seq data analysis), we present the pipeline followed in the current examination for quality check, alignment, and peak calling utilizing available tools and guidelines.<sup>18–20</sup> The scope of the present protocol does not provide further downstream integrated analyses with RNA-seq, ChIP-seq etc. Figure 5 displays the detailed schematic followed along with some of the alternative tools available and Figure 6 presents the ATAC-seq quality check, annotation and ATAC-seq peaks profile [Figure 6: (a) Estimation



**Figure 5. Schematic for ATAC-seq analysis followed, along with some of the alternative tools available**

of library complexity using ATACseqQC in CS (b) Fragment size distribution in ATAC-seq library in CS generated using ATACseqQC (c) Genome coverage plot of reads using ChIPseeker (d) Enrichment plot using ChIPseeker showing enrichment at TSS (e) Tag plot showing reads mapping at TSS (+/- 3,000 bp), generated using ChIPseeker (f) Plot showing annotation of ATAC peaks, generated using ChIPpeakAnno (g) Heatmap showing ATAC peaks in two CS replicates, generated using plotHeatmap of deepTools 3.5].

Demultiplexing of raw sequence files was carried out using bcl2fastq(v1.8). Thereafter, the following pipeline was used and was found to perform robustly well.

48. Quality check: FASTQC.
49. Adapter removal: Cutadapt.
50. Alignment: Bowtie-2.
51. Remove duplicates and mitochondrial reads:
52. Post-alignment quality check: ATACseqQC.
53. Peak calling: MACS2.
54. Peak analysis and annotation: ChIPseeker and ChIPpeakAnno.

Commands/script used for the above-mentioned analysis tools and suites

```

>save.image("STAR Protocols.Rdata")

>fastqc "sample1.fastq" "sample2.fastq"

>cutadapt -a "adapter sequence found in fastqc.html file" -o "sample1trimmed.fastq" "sample1.fastq"

bowtie2 to build reference genome and mapping/alignment

>bowtie2-build "reference genome.fasta" "index genome"

>bowtie2 -very-sensitive -x "index genome" "sample1trimmed.fastq" > "sample1_mapped.sam"

Convert sam file to bam file
  
```

```

>samtools sort "sample1_mapped.sam" > "sample1.bam"

Remove PCR duplicates

>samtools rmdup -s "sample1.bam" "uniquesample1.bam"

Remove mitochondrial reads

>samtools view "uniquesample1.bam" egrep -v chrM samtools view -bT "dm6.fa" -o "uniquesample1_rmmt.bam"

Peakcalling

>macs2 callpeak -t "uniquesample1_rmmt.bam" -f BAM -g dm -outdir "MACS2_Output" -B -SPMR -q 0.05

Generating bigwig files for visualization of peaks and generating heatmaps using bamCoverage and deepTools respectively

>bamCoverage -b "uniquesample1_rmmt.bam" -o "sample1.bw" -bs 5 -smoothLength 20

>computeMatrix reference-point -S "sample1.bw" "sample2.bw" -R "NApeak.bed" -referencePoint center -a 2500 -b 2500 -skipZeros -o "sample1sample2matrix" -missingDataAsZero -skipZeros

>plotHeatmap -matrixFile "sample1sample2matrix" -outFileName "sample1sample2_heatmap"

```

**Note:** sample names have been italicized.

dm6.fa constitutes the reference fasta file of *Drosophila* genome. NApeak.bed file comprises of the bed file generated from MACS2 output files combined for both the samples. Alternatively, BED (browser extensible data) file for *Drosophila* genes can be downloaded from UCSC/Ensembl browser. Workflow for bioconductor packages, namely, ATACseqQC,<sup>18</sup> ChIPseeker<sup>21</sup> and ChIPpeakAnno<sup>22</sup> is followed as previously described.

## EXPECTED OUTCOMES

ATAC-seq data analysis tools described in the protocol are used for both quality check as well as examining the differential accessibility. Prior to accessibility analysis, it is essential to determine the following:

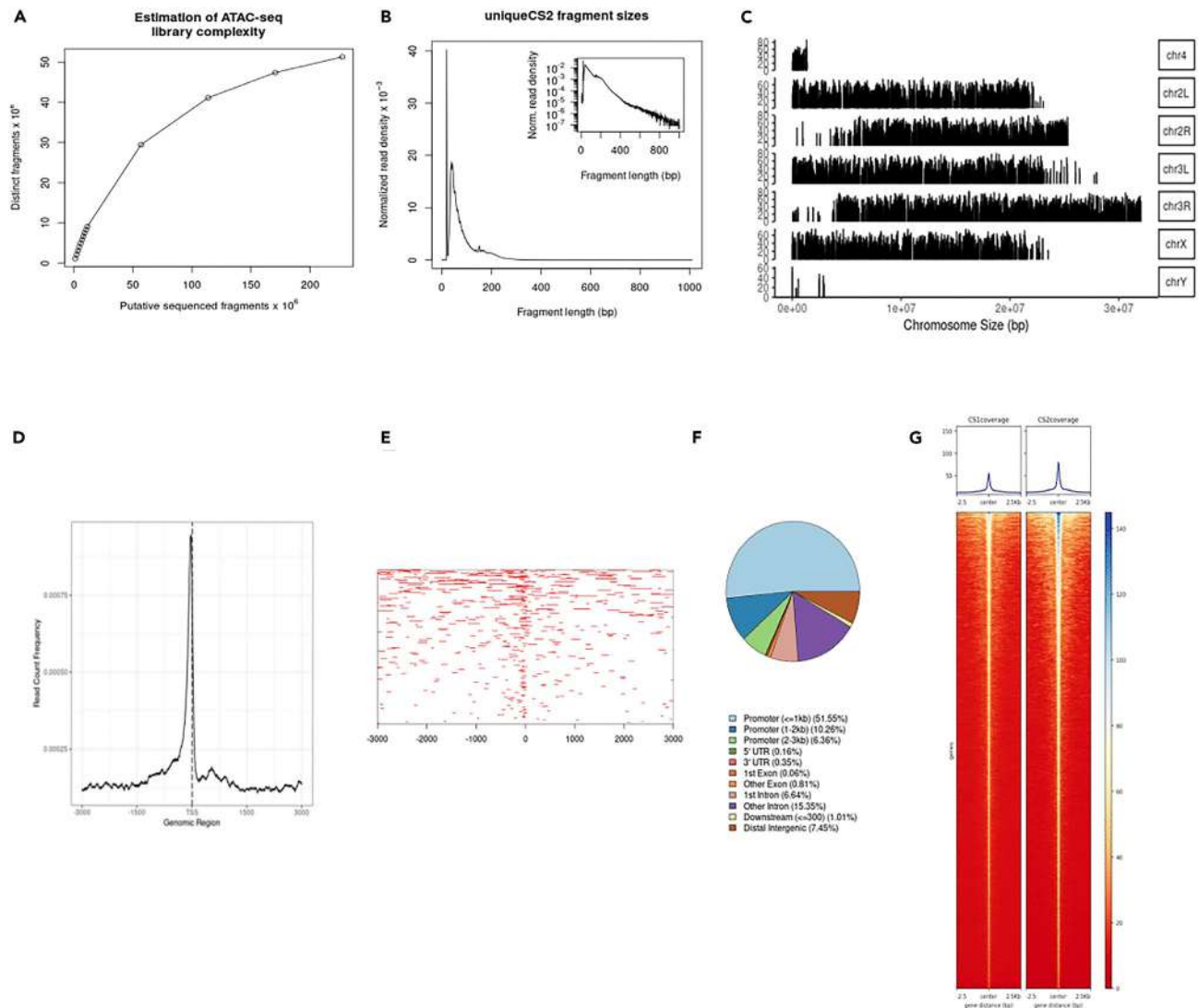
**TSS profile:** Measure of signal to background ratio as accessibility is increased at active promoters. It is one of the most important factor for library assessment and quality check before proceeding for downstream analysis.

**Mapping percentage:** More than 80% alignment rate is recommended after removing duplicates and mitochondrial reads.

**Fragment distribution:** Nucleosomal periodicity is a characteristic of ATAC peaks. Furthermore, ATACseqQC (post alignment Bioconductor package) analysis can be used to ascertain that the nucleosome free region presents reads  $\leq 150$  bp.

**Library complexity:** It constitutes the number of unique fragments (after removal of mitochondrial fragments/reads). This parameter is often not compared between different samples/conditions owing to the variation in the chromatin state.

Figure 6 displays some of the major expected outcome for ATAC-seq quality check and analysis. Library complexity is high depicting the low duplicate reads (Figure 6A). Characteristic nucleosomal periodicity is evident even though the transition is not started (Figure 6B). Coverage plot shows reads captured from all *Drosophila* chromosomes. The reads are depleted in chromosome 4, which



**Figure 6. ATAC-seq quality check, annotations and plot profiling**

- (A) Estimation of library complexity using ATACseqQC in CS.  
 (B) Fragment size distribution in ATAC-seq library in CS generated using ATACseqQC.  
 (C) Genome coverage plot of reads using ChIPseeker.  
 (D) Enrichment plot using ChIPseeker showing enrichment at TSS.  
 (E) Tag plot showing reads mapping at TSS ( $\pm$  3,000 bp), generated using ChIPseeker.  
 (F) Plot showing annotation of ATAC peaks, generated using ChIPpeakAnno.  
 (G) Heatmap showing ATAC peaks in two CS replicates, generated using plotHeatmap of deepTools 3.5.

is highly shortened in the model organism and chromosome Y, which majorly consists of silenced and repetitive DNA.<sup>23</sup> TSS enrichment plot (ATACseqQC) (Figure 6D) and tagHeatmap (ChIPseeker) (Figure 6E) exhibit the increased enrichment of signal around TSS region. Importantly, as expected, peak annotation plot presents relatively increased accessibility in the promoter and intergenic regions. Heatmap (Figure 6F) shows the differentially accessible regions in *Drosophila* genome in two CS replicates.

## LIMITATIONS

The protocol requires a sufficient input material  $\sim$  35 *Drosophila* brains for each library, mandating fast and efficient dissections.

### TROUBLESHOOTING

#### Problem 1

Biased fragment size distribution.

#### Potential solution

Biased fragment size distribution is attributed to undertagmentation (increase in larger fragments) or overtagmentation (increase in smaller fragments). With the given transposase reaction, for 75000 nuclei, reaction time of 20 min results in undertagmentation, while 40 min results in overtagmentation. This was resolved by optimizing reaction time to exactly 30 min.

#### Problem 2

Low TSS enrichment.

#### Potential solution

Low TSS enrichment and increased noise can result from poor amplification. Accurate determination of number of PCR cycles, using q-PCR (Figure 3), for each sample is essential to alleviate the problem.

#### Problem 3

Nucleosomal periodicity is not evident in bioanalyzer/tapestation analysis.

#### Potential solution

One of the major causes of lack of nucleosomal periodicity is undertagmentation. If there is no undertagmentation, yet the characteristic ATAC peaks pattern is not seen, give another round of 0.5 × AMPure bead purification to remove longer fragments.

#### Problem 4

Noise from mitochondrial reads.

#### Potential solution

Minor interference from mitochondrial fragments may be encountered even after washing (step 1 e). Even slight contamination from mitochondrial reads should be removed for better signal to noise ratio. Samtools (step 8) can be used to remove the mitochondrial reads.

#### Problem 5

Poor alignment percentage to the reference genome.

#### Potential solution

The most plausible reason for poor mapping is PCR duplicates and chimeras because of non-specific binding. PCR optimization using q-PCR should be conducted. Other reasons include- smaller fragments owing to over tagmentation and noise from mitochondrial reads. Transposition reaction should be optimized and mitochondrial reads should be removed, as explained above.

#### Problem 6

Uneven sequencing depth between samples.

#### Potential solution

To visualize or compare peaks between samples of different conditions/treatment, it is essential to normalize the sequencing depth between samples. This can be achieved using 'spmr' option in MACS2 peak calling or can be done manually using Samtools, as shown below:

```
>samtools view -b -s 0.x sample1.bam > subsampledsample1.bam
```

**Note:** x refers to the fractionation factor to be applied.

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jasmine Kaur Dhall ([jasminekd@ncbs.res.in](mailto:jasminekd@ncbs.res.in)).

### Materials availability

No unique reagent was generated in the study.

### Data and code availability

The study did not generate any novel codes.

The datasets supporting the current study have not been deposited in a public repository because the downstream integrated analyses with other OMICS data is ongoing but are available from the corresponding author on request.

## ACKNOWLEDGMENTS

We are grateful for Department of Biotechnology (DBT), Govt. of India (GOI) grant BT/PR28450/MED/122/166/2018, and intramural funds from NCBS-TIFR. We are grateful to Prof. Gaiti Hasan and Dr. Dimple Notani for their invaluable suggestions and inputs.

## AUTHOR CONTRIBUTIONS

The protocol is optimized and written by J.K.D. with feedback and suggestions from A.P. DAPI staining and confocal microscopy was carried out by N.K. Sequencing and demultiplexing were conducted by L.C.P. under A.P.'s supervision.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

- Annunziato, A. (2008). DNA packaging: nucleosomes and chromatin. *Nature E* 1, 26.
- Collas, P., Liyakat Ali, T.M., Brunet, A., and Germier, T. (2019). Finding friends in the crowd: three-dimensional cliques of topological genomic domains. *Front. Genet.* 10, 602.
- Saravanan, B., Soota, D., Islam, Z., Majumdar, S., Mann, R., Meel, S., Farooq, U., Walavalkar, K., Gayen, S., Singh, A.K., et al. (2020). Ligand dependent gene regulation by transient ER $\alpha$  clustered enhancers. *PLoS Genet.* 16, e1008516.
- Farooq, U., Saravanan, B., Islam, Z., Walavalkar, K., Singh, A.K., Jayani, R.S., Meel, S., Swaminathan, S., and Notani, D. (2021). An interdependent network of functional enhancers regulates transcription and EZH2 loading at the INK4a/ARF locus. *Cell. For. Rep.* 34, 12.
- Notani, D. (2022). First glimpse of enhancers in gene regulation. *Nat. Rev. Genet.* 23, 522–523.
- Giresi, P.G., Kim, J., McDaniel, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885.
- Chereji, R.V., Bryson, T.D., and Henikoff, S. (2019). Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.* 20, 198–218.
- Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* 2010. pdb.prot5384.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.
- Corces, M.R., Mumbach, M.R., Greenleaf, W.J., Montine, T.J., Khavari, P.A., Kundaje, A., Risca, V.I., Orloff, L.A., Kasowski, M., Carter, A.C., and Cho, S.W. Omni-ATAC-seq: improved ATAC-seq protocol. <https://doi.org/10.1038/protex.2017.096>.
- Haines, J.E., and Eisen, M.B. (2018). Patterns of chromatin accessibility along the anterior-posterior axis in the early *Drosophila* embryo. *PLoS Genet.* 14, e1007367.
- Bozek, M., Cortini, R., Storti, A.E., Unnerstall, U., Gaul, U., and Gompel, N. (2019). ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome Res.* 29, 771–783.
- Merrill, C.B., Pabon, M.A., Montgomery, A.B., Rodan, A.R., and Rothenfluh, A. (2022). Optimized assay for transposase-accessible chromatin by sequencing (ATAC-seq) library preparation from adult *Drosophila melanogaster* neurons. *Sci. Rep.* 12, 6043.
- Grandi, F.C., Modi, H., Kampman, L., and Corces, M.R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nat. Protoc.* 17, 1518–1552.
- Scheffer, L.K., Xu, C.S., Januszewski, M., Lu, Z., Takemura, S.Y., Hayworth, K.J., Huang, G.B., Shinomiya, K., Maitlin-Shepard, J., Berg, S., et al. (2020). A connectome and analysis of the adult *Drosophila* central brain. *Elife* 9, e57443.
- Strober, W. (2015). Trypan blue exclusion test of cell viability. *Curr. Protoc. Immunol.* 111, A3.B.1.
- Lovell, S., Goryshin, I.Y., Reznikoff, W.R., and Rayment, I. (2002). Two-metal active site binding of a Trn5 transposase synaptic complex. *Nat. Struct. Biol.* 9, 278–281.
- Ou, J., Liu, H., Yu, J., Kelliher, M.A., Castilla, L.H., Lawson, N.D., and Zhu, L.J. (2018). ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genom.* 19, 169–213.
- Yan, F., Powell, D.R., Curtis, D.J., and Wong, N.C. (2020). From reads to insight: a

- hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* 21, 22.
20. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, 187–191.
  21. 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.
  22. Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinf.* 11, 237–310.
  23. Brown, E.J., Nguyen, A.H., and Bachtrog, D. (2020). The *Drosophila* Y chromosome affects heterochromatin integrity genome-wide. *Mol. Biol. Evol.* 37, 2808–2824.