

Protocol

Chromatin Accessibility Profiling and Data Analysis using ATAC-seq in *Nothobranchius furzeri*

G. Adam Reeves,¹ Param Priya Singh,¹ and Anne Brunet^{1,2,3}

¹Department of Genetics, Stanford University, ²Glenn Laboratories for the Biology of Aging, Stanford University, Stanford, California 94305, USA

The state of genome-wide chromatin accessibility in cells, tissues, or organisms can be investigated with a technique called assay for transposase-accessible chromatin using sequencing (ATAC-seq). ATAC-seq is a powerful approach for profiling the epigenomic landscape of cells using very low input materials. Analysis of chromatin accessibility data allows for prediction of gene expression and identification of regulatory elements such as potential enhancers and specific transcription-factor binding sites. Here, we describe an optimized ATAC-seq protocol for the preparation of isolated nuclei and subsequent next-generation sequencing from whole embryos and tissues of the African turquoise killifish (*Nothobranchius furzeri*). Importantly, we provide an overview of a pipeline for processing and analyzing ATAC-seq data from killifish.

MATERIALS

It is essential to consult the appropriate Material Safety Data Sheets and the institution's Environmental Health and Safety Office for proper handling of equipment and hazardous materials used in this protocol.

Reagents

10,000× SYBR Green I (Biotium 40086)
AccuGreen High Sensitivity DNA Kit (Biotium 31066-T)
Agarose (optional; see Step 24)
Ampure Bead DNA size selection reagent (Beckman Coulter A63880) (optional; see Step 23)
EZ lysis buffer (Sigma-Aldrich 3408)
High Sensitivity DNA Detection Kit (Agilent 5067-4626)
High Yield Short Read NextSeq Kit (Illumina FC-131-1096)
Killifish of the desired age

Leading up to the desired time of dissection, embryos should be maintained as described in previous publications (Valdesalici and Cellerino 2003; Terzibasi et al. 2008) (see also Protocol: Husbandry of the African turquoise killifish Nothobranchius furzeri [Nath et al. 2023] and Protocol: Breeding and Reproduction of the African Turquoise Killifish Nothobranchius furzeri [Chen et al. 2023]). Adult fish or fry should be raised following standard killifish housing and husbandry protocols (Platzer and Englert 2016; Dodzian et al. 2018; Hu and Brunet 2018) (see also Protocol: Husbandry of the African turquoise killifish Nothobranchius furzeri [Nath et al. 2023]).

³Correspondence: abrunet1@stanford.edu

From the African Turquoise Killifish collection, edited by Anne Brunet.

© 2023 Cold Spring Harbor Laboratory Press

Advanced Online Article. Cite this protocol as *Cold Spring Harb Protoc*; doi:10.1101/pdb.prot107747

Liquid nitrogen (optional; see Steps 4, 8, 12)
MinElute Reaction Cleanup Kit (QIAGEN 28206)
NEBNext High-fidelity 2× PCR master mix (New England Biolabs M05415)
Nextera-Compatible Multiplex Primers (Active Motif 53155)
Nuclease-free H₂O (ThermoFisher AM9916)
Phosphate buffered saline pH 7.0 (PBS, 1×; Thermo Fisher AM9624)
Qubit Double-Stranded High-Sensitivity Assay Kit (Invitrogen Q32851) (optional; see Step 24)
TBE buffer (Tris-borate-EDTA pH 8.0; Thermo Fisher B52) (optional; see Step 24)
Tn5 Transposition DNA buffer and Enzyme kit (Illumina 20034197)
Tricaine methanesulfonate MS-222 (VWR IC10310625) (200–250 mg/L at pH 7.0–7.5)
Trypan blue stain (Thermo Fisher 15250061)
Trypsin-EDTA (Thermo Fisher R001100) (0.025% w/v) (optional; see Step 6)

Equipment

Annotation files

Annotation and reference genomes are publicly available on NCBI at the following link: https://www.ncbi.nlm.nih.gov/assembly/GCF_001465895.1

Genome Annotation File (GTF): GCF_001465895.1_annotation.gtf

Reference Genome: GCF_001465895.1_genome.fna

B2 Mini gel electrophoresis system (Thermo Fisher 09-528-110B) (optional; see Step 24)

Bright-Line cell counting chamber slide (Daigger Scientific EF16034F)

Computer equipped with internet connection and a web browser

Dissection dish, glass (PYREX 722085)

Heating block set to 37°C

Input files

Input files are available from the Figshare repository site at the following link: <https://figshare.com/account/home#/projects/140606>.

Lightcycler 480 II thermal cycler with 96-well plates (Roche)

Microcentrifuge (Eppendorf 2454)

Microcentrifuge tubes (1.5-mL, Eppendorf HEA4323)

Microscope for dissection with light source from both top and bottom

Mini-Douncers (DWK; Kimble 885300-0000)

R Packages

ChIPseeker/1.28.3

Diffbind/2.16.2

GenomicAlignments/1.24.0

GenomicFeatures/1.40.1

Tidyverse/1.3.1

Scripts

*All scripts are publicly available at the following github link and at the Figshare repository:
https://github.com/Dazam3/ATACseq_Pipeline
https://figshare.com/articles/software/ATACseq_Pipeline/19950143*

Python Scripts: Killifish_ATACseq_Pipeline.py

R Scripts: Killifish_ATACseq_Pipeline.R

Sequencing system (NextSeq; Illumina)

Standalone software

bedtools/2.27.1

bowtie2/2.4.1

deeptools/3.3.0

HOMER/4.4

IGV/2.15.4
macs2/2.2.71
picard/2.22.8
Samtools/1.5
trim_galore/0.4.5
Surgical scissors
Tweezers (two pairs, see Step 2) (biological-grade; Electron Microscopy Sciences 72700-D)
Qubit fluorometer (Invitrogen Q32857) (optional; see Step 24)

METHOD

Sample Preparation

As described in Figure 1A, each tissue, organ, or whole embryo of interest should be dissected and processed in lysis buffer to yield a nuclei suspension for quantification. The nuclei are then “tagmented” (DNA fragmented and tagged by the Tn5 transposase) and processed to generate a next-generation assay for transposase-accessible chromatin using sequencing (ATAC-seq) library. The following protocol provides specific instruction for these steps using African turquoise killifish embryos or adult tissue as the input material. It is adapted from previously published protocols for zebrafish (Doganli et al. 2017). Adjustments recommended for large tissues or whole organs are described below. Generally, embryo library preparations allow for comparisons between developmental stages, whereas tissue/organ library prep is ideal for comparison between tissues/organs or across organisms or genotypes.

Tissue Preparation

Embryos

1. On a glass dissection dish filled with chilled (4°C) PBS dissect away the chorion, yolk, and the embryonic membrane surrounding the yolk from a killifish embryo using biological-grade dissection tweezers. For optimal results, lightly puncture the chorion with tweezers and tease the chorion away without causing the yolk to spill out, leaving the embryo within the embryonic membrane (Fig. 1B).

Note that it is difficult to remove the chorion while keeping the yolk/embryonic membrane intact.

2. Using two sets of tweezers, pull the embryonic membrane surrounding the yolk over the embryo apart to remove the intact embryo (Fig. 1B).

It is imperative to remove all of the embryonic membrane from the embryo body. This is because the membrane is robust and adhesive. It will not break down during the physical or chemical lysis stages of the protocol and will act as coagulant surface for nuclei.

3. Transfer the dissected embryo bodies to a 1.5-mL microcentrifuge tube containing 20 μ L of chilled 1 \times PBS. Once collection is complete, pellet embryos by centrifugation at 500g for 5 min at 4°C. Remove as much liquid as possible from the tube.

To generate enough material to obtain a sufficient number of isolated nuclei, it is recommended to pool eight to 12 embryos in a single microcentrifuge tube. This represents ~1.12–1.68 mg of material and corresponds to 280,000–420,000 cells (~35,000 cells per embryo).

4. (Optional) Flash freeze the embryos in liquid nitrogen and store at –80°C for future processing.

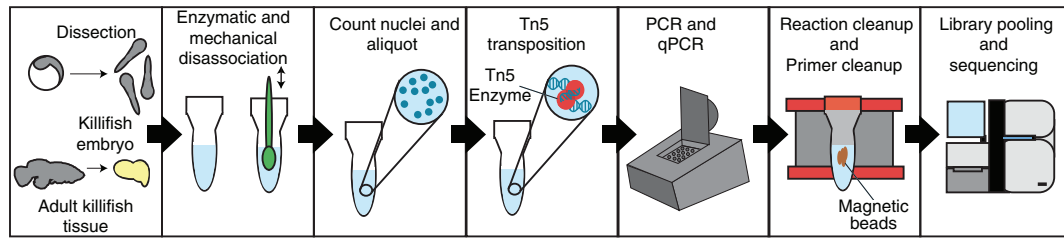
Upon flash freezing and storage at –80°C, samples should remain of viable quality for nuclear extraction for up to 1 yr.

Tissue and Organs

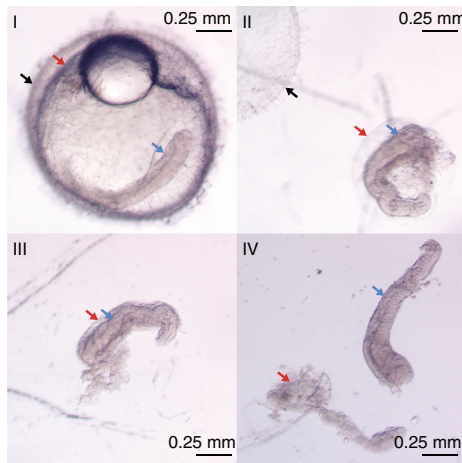
5. For dissection of tissues and organs from adult killifish, euthanize the fish by replacing holding tank water with a solution containing 200–250 mg/L of tricaine methanesulfonate MS-222 at pH 7.0–7.5.

This concentration of tricaine should lead to the animal being anesthetized in ~1–3 min and euthanasia within 15–17 min. Death can be confirmed at 10 min after the last opercular movement has occurred.

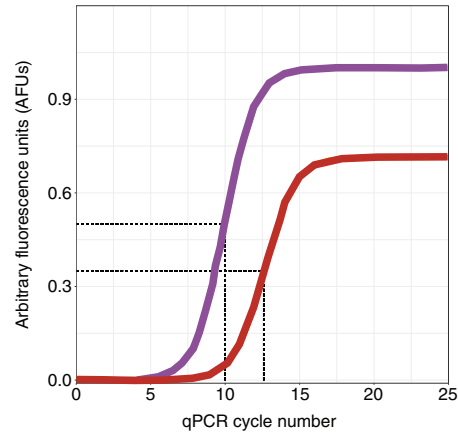
A Protocol for generating ATAC-seq libraries from embryos and dissected tissues



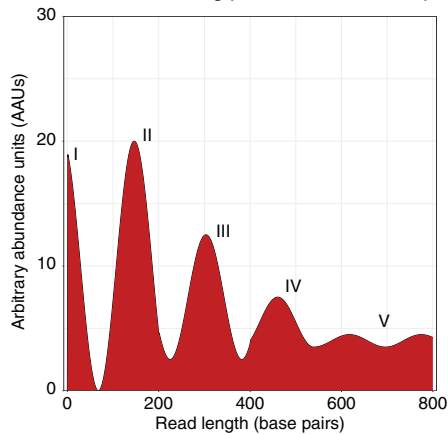
B Dissection of killifish embryos for disassociation



C qPCR curve for optimizing amplification cycle number



D Nucleosome banding pattern of ATAC-seq libraries



E Electrophoresis gel of ATAC-seq libraries

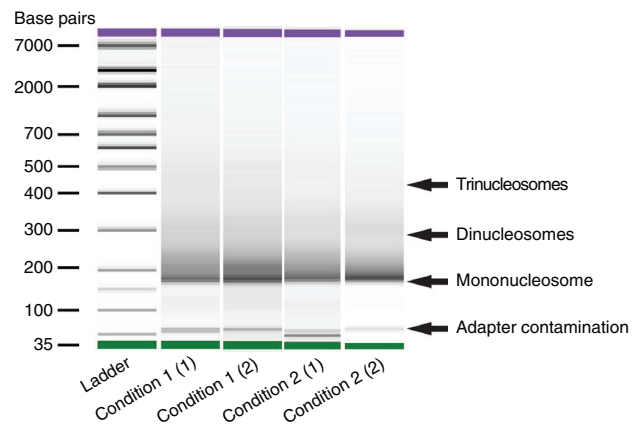


FIGURE 1. ATAC-seq library preparation and metrics. (A) Experimental steps and design for the generation of ATAC-seq libraries from either embryonic or adult killifish tissue. This process includes both physical and chemical lysis of material to isolate nuclei, fragmentation of chromatin using Tn5 transposase, and finally sample indexing and amplification. (B) Stepwise images of killifish embryo dissection. (I) The embryo initially has an intact chorion (black arrow) with the embryo body (blue arrow) resting atop the yolk bound by the embryonic membrane (red arrow). (II) The chorion (black arrow) has been dissected away exposing the embryo and yolk bound inside the embryonic membrane (red arrow), which also contains the embryo body (blue arrow). (III) The embryonic membrane (red arrow) has been punctured and the semiliquid yolk released into the surrounding medium, leaving only the deflated embryonic membrane (red arrow) around the embryo body (blue arrow). (IV) The embryonic membrane (red arrow) has been stretched, torn, and removed from the embryo leaving only the intact embryo body (blue arrow). (C) Example of ideal quantitative polymerase chain reaction (qPCR) curves generated from aliquots of partially amplified ATAC-seq libraries. The number of qPCR cycles required to reach half the amplitude of the qPCR amplification plateau is the recommended number of additional cycles to use to amplify the remaining ATAC-seq libraries. In this example plot, 10 additional cycles (purple library) and 13 additional cycles (red library) were performed. (Continued on next page.)

6. Dissect tissues and organs using surgical scissors and biological-grade dissection tweezers in an empty dissection dish.

For dissociation of tissue from adult fish, additional chemical dissociation may be required. We recommend using 500 μ L 0.025% trypsin for 5 min at 37°C before Step 7. We recommend using 1.12–1.68 mg of starting tissue (or 280,000–420,000 cells) per sample.

7. Transfer the dissected tissue or organs to a 1.5-mL microcentrifuge tube containing 20 μ L of chilled 1 \times PBS. Once collection is complete, pellet each tissue or organ(s) by centrifugation at 500g for 5 min at 4°C. Remove as much liquid as possible from the tube.

For large organs in which subsampling a small section is not possible, it is advisable to increase the initial volume of chilled PBS in the 1.5-mL microcentrifuge tube for initial containment. If the tissue selected is very sensitive and could break apart, omit centrifugation and directly remove the supernatant before freezing or further processing.

8. (Optional) Flash freeze the tissue or organs in liquid nitrogen and store at –80°C for future processing.

Upon flash freezing and storage at –80°C, tissues or organs should remain of viable quality for nuclear extraction for up to 1 yr.

Nuclei Isolation

If nuclei isolation is conducted on samples stored at –80°C, make sure the samples are given time to reach at least 4°C before Step 9, as this will give the best yield of intact nuclei.

9. Resuspend samples in 100 μ L of EZ lysis buffer at 4°C and transfer the mixture to a 250 μ L mini-Douncer. To ensure complete transfer of material to the Douncer, add an additional 100 μ L of EZ lysis buffer to the microcentrifuge tube and pipette vigorously before transferring. After the second transfer, homogenize the sample with glass pestles of increasing size (A and B respectively from the DWK Mini-Douncer set) for 25 strokes each. Use the pestle with force but slowly so as to not produce excess bubbles.

For large or pooled tissue samples, larger Douncers and scaled up reaction volumes are advised.

10. Transfer the homogenized samples back to a 1.5-mL microcentrifuge tube. To ensure complete transfer of material from the mini-Douncer, wash thoroughly with an additional 50 μ L of EZ lysis buffer and transfer to the same microcentrifuge tube for a final volume of 250 μ L. Allow the mixture to sit on ice (~4°C) for 2 min for chemical lysis to occur.

11. Pellet the nuclei by centrifugation at 500g for 5 min at 4°C. Remove the lysis buffer supernatant without disturbing the nuclei pellet. Resuspend the nuclei in 50 μ L of ice-chilled 1 \times PBS.

It is important to remove all lysis buffer supernatant to ensure preservation of intact nuclei.

12. (Optional) Flash freeze samples in liquid nitrogen and store at –80°C.

It is possible to stop here and continue library preparation later; however, unless this is necessary because of a large sample number or other conditions, we recommend continuing with ATAC-seq library preparation because a fraction of nuclei will degrade with a freeze–thaw cycle. Our own studies have revealed a loss of ~10% of the total nuclei per sample for every 1–2 mo of storage, meaning that ATAC-seq libraries should be prepared as soon as possible and not later than 6 mo after the original isolation to avoid differences in numbers of viable nuclei.

FIGURE 1. (Continued from previous page.) (D) Theoretical nucleosome banding pattern displaying the presence/absence and intensity of the mono- (II), di- (III), and tri-nucleosome (IV) bands for each ATAC-seq library. The graphic also displays a peak that signifies adapter/barcode contamination that can be removed via bead cleanup (I). (E) Electrophoresis separation of four ATAC-seq libraries before adapter contamination cleanup on the Bioanalyzer platform. The ladder to the left displays the run position of DNA fragments of known length ranging from 35–7000 base pairs. The arrows to the right denote the expected location of contaminants, and mononucleosome (~146), dinucleosome (~292), and trinucleosome (~438) DNA fragments. The mononucleosome band will often be the most prominent.

13. Quantify nuclei. Transfer 10 μL of the sample to a new microcentrifuge tube and add 10 μL of 1 \times trypan blue. Pipette the mixture repeatedly. Transfer 10 μL of this mixture to a Bright-Line cell counting chamber slide and determine the concentration of nuclei using the manufacturer provided calculation.

During library preparation, nuclei yield from small amounts of tissue can be variable, ranging from $\sim 47,000$ to 378,000 nuclei (15%–90% maximal yield). It is recommended to only proceed with libraries that have at least 30% of the maximal yield to avoid cell type biases.

14. Aliquot $\sim 25,000$ nuclei to a new microcentrifuge tube and bring the total volume to 5 μL by either (a) adding ice-chilled 1 \times PBS or (b) centrifuging the sample at 500g for 5 min at 4 $^{\circ}\text{C}$, removing excess supernatant, and resuspending nuclei in an appropriate volume by repeated pipetting.

Library preparation requires an input of $\sim 25,000$ nuclei; therefore, intact nuclei should be quantified in each sample so that an exact aliquot and dilution can be prepared for each sample.

ATAC-seq Library Preparation

If libraries are prepared from samples stored at -80°C , make sure the samples are given time to completely thaw before beginning the transposition reaction. It is also advisable to quantify total intact nuclei (see Steps 13–14) after thawing because some nuclei degrade during long periods of cold storage and thawing.

15. To a 5 μL sample of $\sim 25,000$ nuclei in a microcentrifuge tube add 130 μL of transposition mix (comprising 64 μL of Transposition DNA buffer, 64 μL of nuclease-free H_2O , and 2.5 μL of Transposition DNA enzyme I [Tn5]). Mix by gentle pipetting and place tube in a heating block for 30 min at 37 $^{\circ}\text{C}$.

We recommend using a lower concentration of Tn5 enzyme for killifish embryo samples compared with those used for zebrafish (Buenrostro et al. 2013; Doganli et al. 2017) because killifish embryo nuclei are fragile, and chromatin can become overtransposed. This decreased concentration of Tn5 has been necessary to generate quality libraries from both embryonic and adult tissue killifish samples. See Troubleshooting.

16. To clean the sample after the Tn5-transposition reaction, run the sample through a MinElute reaction cleanup column following the manufacturer's instructions and with reference to previous ATAC-seq protocols (Buenrostro et al. 2013; Doganli et al. 2017).

For the final elution of cleaned, transposed DNA, it is important to use only 10 μL of elution buffer so the final volume will be appropriate for the subsequent PCR/qPCR steps.

17. For each sample, prepare a PCR reaction in a 96-well PCR plate (10 μL of purified DNA, 10 μL of nuclease-free H_2O , 2.5 μL custom Nextera primer I, 2.5 μL of custom Nextera primer II, and 25 μL of 2 \times high-fidelity PCR master mix) and run the following program (Doganli et al. 2017):

- i. 5 min at 72 $^{\circ}\text{C}$
- ii. 30 sec at 98 $^{\circ}\text{C}$
- iii. 5 cycles of 10 sec at 98 $^{\circ}\text{C}$, 30 sec at 63 $^{\circ}\text{C}$, 1 min at 72 $^{\circ}\text{C}$
- iv. Hold at 4 $^{\circ}\text{C}$.

It is important to tailor PCR amplification cycles to add next-generation sequencing barcodes (custom Nextera barcodes/primers are detailed in Supplemental Table 1 of Buenrostro et al. 2013 or custom kit from New England BioLabs) and to amplify DNA for sequencing. It is critical to adjust the PCR reaction cycle number for each individual sample to avoid each library being overamplified, therefore Step 18 must be completed before continuing the primary PCR step.

18. Prepare quantitative PCR (qPCR) reactions in a 96-well PCR plate to determine the optimal number of subsequent PCR amplification cycles. For each sample prepare qPCR mix as follows:

- i. 3.9 μL of nuclease-free H_2O
- ii. 0.25 μL of custom Nextera primer I
- iii. 0.25 μL of custom Nextera primer II
- iv. 0.6 μL of 100 \times SYBR Green I

- v. 5 μ L of high-fidelity PCR master mix
 - vi. Add 5 μ L of the sample from Step 17 (held at 4°C).
19. Run the qPCR on a Lightcycler 480 II using the following program:
 - i. 30 sec at 98°C
 - ii. 20 cycles of 10 sec at 98°C, 30 sec at 63°C, 1 min at 72°C.
 20. Determine the qPCR cycle number that generates half the total amplitude of the qPCR amplification plateau. This value is the correct number of continued PCR cycles for each original sample from Step 17 (Fig. 1C). Reallocate Step 17 samples to new PCR plates based on the remaining cycle number each sample requires (determined by qPCR in Step 19) and run the additional cycles using the same PCR program as in Step 17.
 21. Remove reaction contaminants from PCR using a MinElute reaction cleanup kit following the manufacturer's instructions. Use a final column elution volume of 25 μ L for a concentrated final library.
 22. Check for the presence of primer contamination by running a diluted aliquot of the sample on a high-sensitivity Bioanalyzer gel (from the Agilent High Sensitivity DNA Detection Kit) and looking for a DNA content peak below 100 bp in length.

Although the MinElute reaction cleanup kit will remove many contaminants, in many cases primer contamination will still exist and will need to be removed before sequencing. Regardless, the expected yield of each library should be ~4.5 μ g of total DNA at a concentration of 175 nM.

23. (Optional) If contamination is present, perform an additional cleanup step using Ampure beads following the manufacturer's protocol.

This selectively removes small (<100-bp) DNA fragments and limits primer contamination, which can interfere with Illumina next-generation sequencing.

24. Check library quality and concentration by one of the following methods:
 - Check the high-sensitivity Bioanalyzer gel from Step 22 for the presence or absence of a typical “nucleosome banding pattern” within the library (Fig. 1D,E). The Bioanalyzer platform will compute the concentration of DNA within each size bin of the gel automatically as well as an aggregate concentration value (both pg/ μ L and molarity).
 - Run a regular wide-range DNA electrophoresis gel to assess the nucleosome banding pattern in samples.

We suggest running the libraries on a 2% agarose gel using 1 \times TBE (pH 8.0) buffer at a voltage of ~100 mV for 45 min.

This gel does not allow for quantitative concentration calculations but is ideal for qualitative evaluation of expected nucleosome banding pattern. If the majority of DNA fragment sizes are shifted smaller or larger than the expected banding pattern displayed in Figure 1D,E, the samples are over- or undertransposed, respectively, and the total time of the Tn5 reaction should be adjusted for subsequent samples.

- Evaluate the samples using the Qubit platform with the double-stranded DNA high-sensitivity kit for DNA quantification following the manufacturer's instructions.

The advantage of calculating concentration from a Bioanalyzer gel rather than Qubit is that the calculation can be restricted to a set size range (150–1000 bp) and the molarity of the DNA sample can be quantified. Trace amounts of primer contamination (<100 bp) or large DNA fragments that do not have strong binding affinity to the next-generation sequencing flow-cell substrate (>1000 bp) can be ignored.

Sequencing

The PCR amplification step above adds primers containing unique DNA barcodes to each sample, making it possible to pool samples together for a single sequencing run and to use the barcodes to deconvolute their origin during analysis.

25. Using the library concentration values acquired in Step 24, dilute libraries to be combined to the same concentration and combine in a 1:1 ratio to generate a single sample for sequencing.

26. Conduct paired-end sequencing of the pooled library with a read length of at least 75 bp.

A read length of 75 bp is commonly recommended in many ATAC-seq protocols (Buenrostro et al. 2013; Doganli et al. 2017). Longer read lengths can also be used, which can be later computationally truncated or “trimmed” to 75 bp. Generally, an ATAC-seq library of sufficient coverage and depth should contain at least 25 million uniquely mapping reads. To reach this suggested cutoff, it is advisable to pool and sequence libraries so that the estimated read yield per library is between 35–50 million reads. To acquire this depth of reads from a single sequencing run, it is advisable to run the pooled library on at least an Illumina NextSeq, HighSeq, or NovaSeq. These sequencing read requirements will change based on several factors including the number of libraries pooled for sequencing.

Sample Analysis

As described in Figure 2A and detailed below, the analysis of ATAC-seq data requires a set of scripts to generate a mapping index of the provided reference genome so that quality-assessed, trimmed, and filtered ATAC-seq fastq files can be aligned to produce Binary Alignment Maps (.BAM) files. Using these files, the pipeline defines accessible chromatin peaks along the reference genome followed by identification of differential peaks. Differential peaks are the genomic regions where chromatin accessibility significantly changes between provided conditions, time points, or tissue types. This provides a rich data set which can be used to evaluate promoter accessibility, identify proximal/distal enhancer sites, locate potential motif binding sites, and much more.

Trim Reads

Reads can be trimmed using a variety of programs (Corporation 1991; Kruger 2012; Bolger et al. 2014); however, we favor Trim Galore because of its modular ability to perform dynamic barcode quality trimming as well as static base-size trimming.

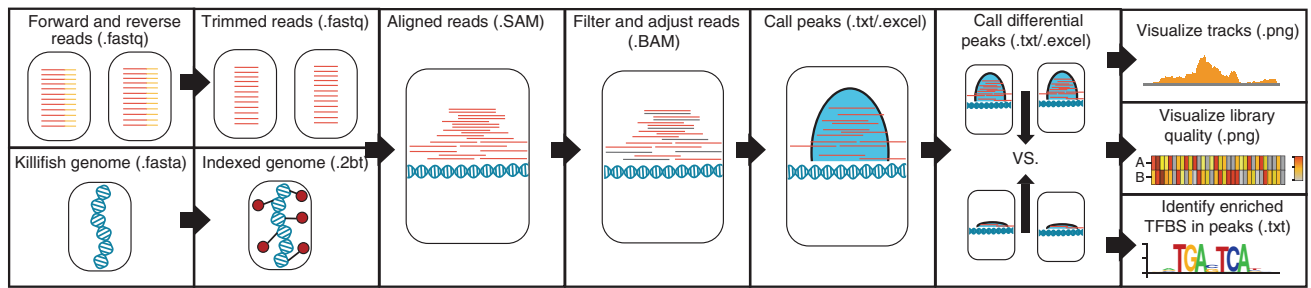
27. Use Trim Galore to trim reads contained in the fastq file generated from sequencing. Use the following command to remove barcodes and to trim low-quality bases from the ends of reads:

```
<trim_galore -fastqc -fastqc_args “-outdir out_directory/file.html” -gzip -output_dir out_directory/ -paired input_dir/forward_reads.fastq.gz input_dir/reverse_reads.fastq.gz>
```

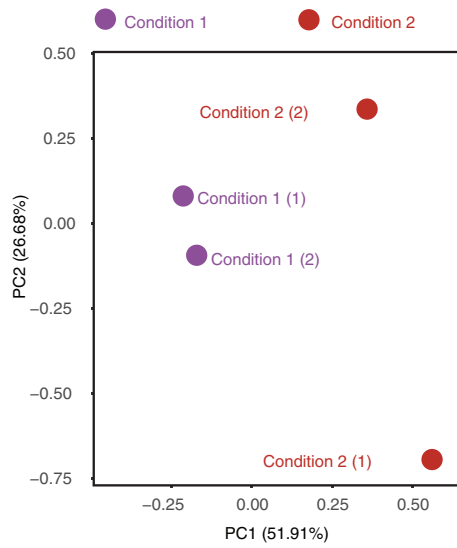
In the command, the argument `<-fastqc>` can be used to generate a library quality report during the trimming process with the output file for the report being directed to the desired file using `<-fastqc_args “-outdir Out_Directory/File.html”>` (Ewels et al. 2016). To conserve data storage of library fastq files, Trim galore can be used directly on zipped files by adding the argument `<-gzip>` to the issued command. It is important to include the `<-paired>` argument when working with paired-end reads, as this will ensure that the program retains the 1:1 forward/reverse read pairing when trimming or discarding reads. Once all these arguments are chosen, it is required to run the command for each forward and reverse read fastq file separately. Designate the directory location for both paired files and the desired output location for both the fastqc report and trimmed files. The command does not specify the type of barcode adapter (i.e., Nextera) attached to reads during PCR amplification of the library and makes use of Trim Galore’s auto barcode sequence detection. However, this can be supplied manually by the user, for example using the additional argument `<-nextera>` or `<-illumina>`.

FIGURE 2. ATAC-seq data processing and analysis. (A) Schematic detailing the computational design to process and analyze ATAC-seq data. This includes assessing the quality of sequencing files, indexing the genome to map sequencing reads to it, filtering reads to allow for detection of peaks generated by read pile-up, and finally contextualizing the biological relevance of the data through a variety of analyses, including principal component analysis (PCA), and motif binding site enrichment. (B) Illustration of a typical PCA of all chromatin accessibility between condition 1 (purple) and condition 2 (red) embryo samples of African turquoise killifish. Each point represents the consensus ATAC-seq peaks (chromatin accessibility) from an individual replicate of a given condition 1 or condition 2 state. Percentage of the variance explained by each principal component (PC) is shown in parentheses. (C) Example representation of transcription start site (TSS) read enrichment compared with neighboring 2-kb regions for each ATAC-seq library generated. Each line of the heatmap represents the enrichment of sequencing reads at a given TSS site and they are ordered by the strength of the enrichment. Enrichment of accessibility signal at a TSS indicates good quality. (D) Curated read coverage tracks of differential accessible chromatin peaks. The tracks were generated on the IGV platform by adding read pileup values across samples and then colored manually. The same genomic coordinates are inaccessible in condition 1 (purple) compared with condition 2 (red). (E) Heatmap depicting the enrichment of specific transcription factor binding motifs in accessible chromatin during condition 1 (upper) or condition 2 (lower). (Figure on next page.)

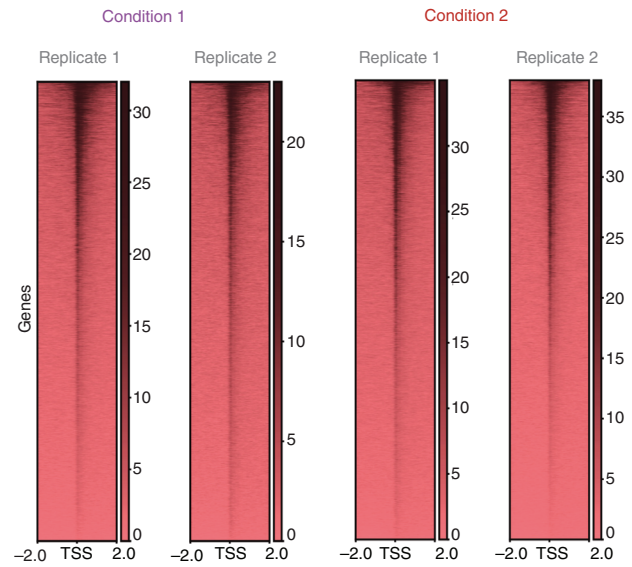
A Analysis pipeline for ATAC-seq libraries from dissected killifish embryos and tissues



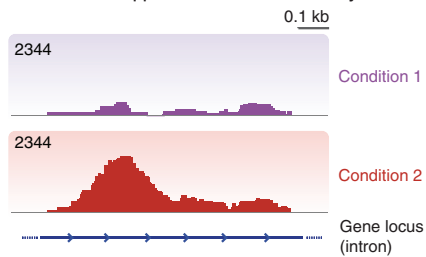
B Global library visualization via PCA



C Transcription start-site enrichment of libraries



D Visualization of ATAC-seq peaks with accessibility tracks



E Transcription factor enrichment within accessible chromatin

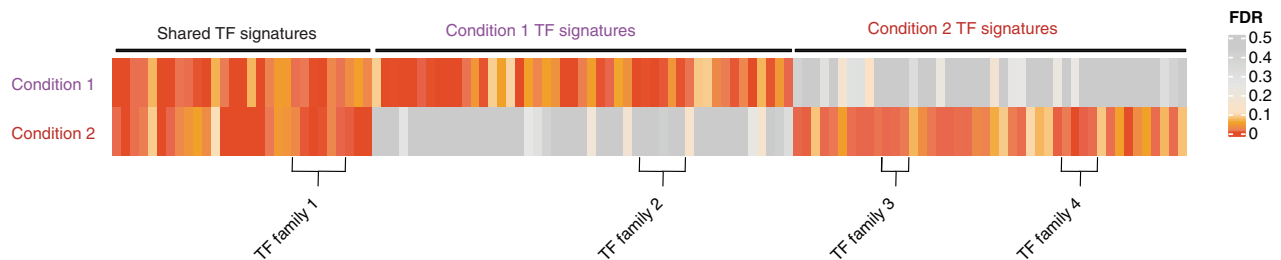


FIGURE 2. (Legend on previous page.)

Additionally, if sequencing of samples was performed using reads of differing lengths (i.e., 75-bp paired-end and 150-bp paired-end), the arguments `<-clip_R1 75>` and `<clip_R2 75>` can be used to standardize read lengths by reducing read size by the specified number of base pairs (e.g., 75 bp) from the forward and reverse read files, respectively.

Map Reads

Before reads can be mapped to the genome to define regions of chromatin accessibility, the genome requires indexing.

28. Index the genome using the mapping package of choice. The command for bowtie2 genome indexing using the provided wrapper script with the bowtie2 package is (Langmead and Salzberg 2012):

```
<bowtie2-build -f genome.fa base_name>
```

The above command uses the argument `<-f>` to identify the FASTA-formatted genome input file provided and to generate several output index files all beginning with the string provided in place of `<base_name>`.

29. Directly map reads using standard bowtie to input commands and arguments:

```
<bowtie2 -very-sensitive -x base_name -1 output_dir/  
forward_reads_trimmed.fastq.gz -2 output_dir/reverse_  
reads_trimmed.fastq.gz -S alignment_dir/library.sam & stdout>  
align_dir/align.log>
```

It is advisable to use the `<-very-sensitive>` argument to maximize the amount of mapped reads by allowing some mismatches (Cunningham et al. 2022). In addition to this designation, the user is asked to provide the base name string for genome index files, the path to forward and reverse trimmed read files, and a desired output destination for the alignment in SAM format (using arguments `<-x>`, `<-1>`, `<-2>`, and `<-S>`, respectively). Additionally, adding `<stdout> align_dir/align.log>` will write alignment rate raw numbers and percentages to the designated log file, which can be used as a metric to assess the quality of the library.

30. Compress large SAM format files to binary based-files (BAM) to reduce file size and make them more easily indexed. This step, as well as sorting the reads by genome position, can be accomplished using a series of commands relying on the Samtools suite of commands (Danecek et al. 2021):

```
<samtools view -S -b align_dir/library.sam> align_dir/library.  
bam>
```

```
<samtools sort -T temp_dir/library.bam -o align_dir/library_  
sorted.bam align_dir/library.bam>
```

```
<samtools index align_dir/library_sorted.bam>
```

Together, these three commands will convert the read alignment file to a BAM format, sort the reads in the file by genomic coordinates, and index the binary file for easy visualization in subsequent steps, respectively. In the first command, the input file and output file are included with the arguments for auto-detection of input file structure, `<-S>` and the designated output format of a binary alignment file `<-b>`. This alignment file is then used as the input for the second command in which reads are sorted by genomic location passing through a temporary file `<-T>` to conserve ram and is ultimately passed to a new sorted BAM format file `<-o>`. Finally, the mapped reads are sorted so that the file will later be compatible for various visualization platforms. After subsequent steps additional sorting will be required; however, visualization of files generated at this step offers the best look at the raw read data.

Filter Reads

31. Filter out low-quality mapped reads and exclude mitochondrial contamination. To accomplish this, evaluate reads and give them a proper SAM-format flag for later removal using two commands:

```
<100 * $(samtools view -c align_dir/library_sorted.bam Mito_Chr) /  
$(samtools view -c align_dir/library_sorted.bam) |bc - |> align_  
dir/mito_read_percent .txt>
```

```
<samtools view -h align_dir/library_sorted.bam|grep -v Mito_Chr  
| samtools sort -O bam -o filter_dir/library_sorted_mitoremoved.  
bam>
```

The first command reports the fraction of reads mapping to the killifish mitochondrial scaffold and logs this value for each library in a text file for the user to later reference. The second command passes the contents of the BAM file through several steps divided via the piping (the “|” character). These include accessing the reads in the BAM file, finding all reads that are flagged as mapping to the mitochondrial scaffold, and finally removing those reads from the library, generating a new BAM file in the process.

32. Remove optical and PCR duplicate reads from the library to improve quality. This is accomplished in several steps, the first of which is adding a flag to all reads that are duplicated in the file using the following command:

```
<picard MarkDuplicates QUIET=true INPUT=filter_dir/library_  
sorted_mitoremoved.bam OUTPUT=filter_dir/library_sorted_  
mitoremoved_dupmarked.bam MAX_FILE_HANDLES_FOR_READ_ENDS_  
MAP=1000 METRICS_FILE=filter_dir/dupMarked.sorted.metrics  
REMOVE_DUPLICATES=false CREATE_INDEX=true VALIDATION_  
STRINGENCY=LENIENT TMP_DIR=filter_dir/tmp/>
```

Specifically, this command takes the filtered BAM file produced in the previous step and generates a metrics file containing information about the duplication rate detected within the library. Additionally, an output BAM is generated in which duplicates have been flagged but not yet removed (designated by the <false> response to the <REMOVE_DUPLICATES> argument). This command has also been run with several arguments including the <LENIENT> setting for validating a read as being duplicated. This will generate the most stringent removal of reads as it increases the likelihood of false positives while limiting the false-negative rate. The run-time for this command is also drastically improved by increasing the number of temporary files the command can have open at once in the temporary directory (designated using the <TMP_DIR> argument) by adjusting the number provided to the <MAX_FILE_HANDLES_FOR_READ_ENDS_MAP> argument.

33. Remove duplicates using the following command:

```
<samtools view -h -b -F 1024 filter_dir/ library_sorted_  
mitoremoved_dupMarked.bam> filter_dir/library_dupRemoved.bam>
```

This command takes the flagged BAM file from previous steps as input and generates a new BAM output (designated by the argument) with SAM headers (<h>) and reads containing a SAM numeric flag that designates the read as duplicated <F 1024> removed.

34. To complete the filtering of reads and improve library quality, remove reads that have a low mapping score (less than a mapQ of 20). This can be accomplished using two Samtools commands that are similar to the previous steps for read filtering:

```
<samtools view -h -q 20 filter_dir/library_dupRemoved.bam>  
filter_dir/library_mapQ20.bam>
```

```
<samtools view -h -b -F 1804 -f 2 filter_dir/library_mapQ20.bam>  
filter_dir/library_filtered.bam>
```

In this iteration, using the Samtool argument <q 20> will flag reads with a mapQ score less than 20, which can then be filtered by the second command using the adjusted SAM numeric flags <F 1804> and <f 2>.

Adjust Alignment

After filtering the aligned read BAM files, the next step is to shift read lengths to account for the 9-bp cutting interval between the two cut sites of the Tn5 enzyme. The shift can be introduced in several ways, such as extending reads by ± 9 bp causing increased read overlap or extending reads +5 and -4 to match them end-to-end over the 9-bp gap. Although this will allow for the evaluation of exact Tn5 cut sites, this step can be skipped if this information is not needed.

35. (Optional) Use a similar command to that in Step 31 to sort the filtered read BAM file:

```
<samtools sort -T filter_dir/library_filtered_temp.bam -o shift_  
dir/library_filtered_sorted.bam
```

```
filter_dir/library_filtered.bam>  
<samtools index shift_dir/library_filtered_sorted.bam>
```

The first command will sort the reads into a new BAM file, whereas the second command will generate an index for the BAM file. Although this is not necessary to perform the read shift, it will allow for visualization of nonshifted reads files later during the “Visualizing Peak Loci” section of the protocol (Step 50).

36. (Optional) The following command will directly shift the reads to account for the Tn5 9 bp cut spacing:

```
<alignmentSieve -numberOfProcessors 4 -ATACshift -bam shift_dir/  
library_filtered_sorted.bam -o shift_dir/library_shifted.bam>
```

The command can be lengthy to run as each read will have to be corrected. To improve the speed, use the argument <-numberOfProcessors> to adjust the reads in parallel. The -argument <-bam> designates the output will appear in BAM format and <-o> is used to name the output file.

37. (Optional) Resort the BAM file and index it for use in the “Visualizing Peak Loci” section of the protocol (Step 50). The commands are similar to those for sorting and indexing of BAM files, respectively:

```
<samtools sort -O bam -o shift_dir/library_FINAL.bam shift_dir/  
library_shifted.bam>  
<samtools index shift_dir/library_FINAL.bam>
```

38. (Optional) Although not necessary, it is advisable to remove excessive intermediate files as each takes a considerable amount of memory. These can be cleared using the basic shell command:

```
<rm folder/file.bam>
```

Intermediate files that should be kept include the unshifted BAM file and its associated index, the final BAM file and its associated index, and any and all log files that hold details about the library quality and run status of the pipeline (fastqc, mito.log, etc.).

Call Peaks and Differential Peaks

39. With reads successfully mapped, filtered, and shifted, use the MACS2 package to define the boundaries of accessible chromatin regions (peaks) (Zhang et al. 2008; Liu 2014). This is accomplished using the following command:

```
<macs2 callpeak -f BAMPE -g 8.568 108 -keep-dup all -cutoff-  
analysis -n killifish -B -t shift_dir/Library_FINAL.bam -outdir  
peaks_dir/> peaks_dir/mac2.log>
```

The command uses the argument <-f> to define the input format, which in this protocol is a paired-end BAM file (i.e., <BAMPE>). Macs2 also takes four inputs; (a) the name of the finalized BAM file <-t shift_dir/Library_FINAL.bam>, (b) the prefix to be used for the output files <-n killifish> designated using <-n>, (c) the directory to which the output should be written <-outdir. /peaks_dir/>, and (d) the user-specified genome size using the argument <-g> and scientific notation <8.568 × 10⁸>. Additionally, the arguments <-keep-dup all> and <-cutoff-analysis> have been added to allow for all read tags (because duplicates have been filtered in the step above) and to have Macs2 determine a dynamic cutoff range for peak calling using the data. A BedGraph-formatted output is also added using the argument <-B>, whereas all information output to standard-out is saved to a log for future reference <command> peaks_dir/mac2.log>.

40. With peak files generated, create an experimental template file to allow the R package, Diffbind, to compare across samples and to generate a list of peaks with differential accessibility between conditions. This file is in .csv format, as shown in the documentation for Diffbind (Stark and Brown 2011). This sheet includes a sample name, tissue, condition, status, peakcaller, and file column and is used to directly access BAM-formatted peak files. Once generated, provide a sample sheet and output directory for the current script run:

```
<sample_sheet=". /peaks_dir/sample_sheet.csv">  
<output_dir=". /DEseq2_dir/">
```

41. Read the sample sheet to generate a DBA object, which stores information about each sample. Use this DBA object to count and contrast read/peak contents between conditions and to call differential peaks. This is accomplished using the following five commands:

```
<samples<- read.csv(sample_sheet) >
<DBdata <- dba(sampleSheet=samples) >
<DBdata<- dba.count(DBdata) >
<DBdata <- dba.contrast(DBdata, categories=DBA_CONDITION,
minMembers=2) >
<DBdata <- dba.analyze(DBdata, method=DBA_DESEQ2) >
```

In the fourth line, the command <dba.contrast> is used to set the minimum number of required replicates <minMembers=2> as well as what conditions will be compared for differential peak accessibility <categories=DBA_CONDITION>. In the example above this contrast is set between condition 1 and condition 2 samples.

42. With the analysis complete, generate a consensus peak file for the entire sample set:

```
<consensusPeaks=dba.peakset(DBdata, peak.format="bed",
consensus=T, bRetrieve=T) >
```

These commands together generate a consensus peak file. The first command using <dba.peakset> extracts peak information in a bed format <peak.format="bed">, specifically including consensus peaks <consensus=T> and their rpk values <bRetrieve=T>.

43. (Optional) In addition to generating a consensus peak file, a differentially accessible peak file can also be generated using the additional data reorganization below:

```
<report=dba.report(DBdata, contrast=i, method=DBA_DESEQ2, )
<df1 <- data.frame(seqnames=seqnames(report), starts=start
(report)-1, ends=end(report), names=paste0("peak_", names
(report)), scores=c(rep("0", length(report))), strand=c(rep
(".", length(report))), fold=elementMetadata(report)$Fold) >
```

In summary, this pipeline evaluates each comparison made between samples (in the example above, one comparison is made – condition 1 vs. condition 2) and generates a sample-versus-sample comparison report, transforms it into a data frame with appropriate column names, and writes it as an output file in bed format, listing all DE-peaks in the respective condition. This differential pairwise comparison is done using DEseq2 (Love et al. 2014) in the script above but can also be run with EdgeR (Robinson et al. 2010) if designated <method=DBA_DESEQ2> or <method=DBA_EDGER>. Once each pairwise comparison has been generated and written, an aggregated CSV-formatted table is generated to include every differential peak between two conditions regardless of directionality.

Plotting Data Aggregates Using Principal Component Analysis

A cursory approach to evaluating how biological samples and replicates relate to one another is principal component analysis (PCA). PCA uses individual genes as features to collapse the multidimensional nature of the omics data into a two-dimensional representation that highlights the major contributors to similarity and difference between samples.

44. Perform PCA in R using the following commands to define the input data file and load the data set of RPKM values:

```
<Input <- read.table('./RPKM_Master_0.0.txt', header=T) >
```

The RPKM_Master_0.0.txt file is simply a truncation of the <consensusPeak> table generated in Step 42, containing one RPKM-value for each peak across each sample.

45. Clean the data to transform it for PCA. This includes approximating all RPKM values using the “for loop” below and omitting genes that do not have an RPKM.


```
<for(j in 1:ncol(Input)) { Input[,j] <- as.integer(Input[,j]) }>  
<Nfur <- Input[!apply(Input, 1, is.na), ]>
```

Alternatively, the final line of code, which removes NA values, can be circumvented by converting NA values to 0, although this generates false negatives as it assumes a true no signal result at the site of the chromatin accessibility.

46. Once NA values have been handled, normalize and transform (i.e., row and column swap) the new integer table and use it to generate eigenvectors for each sample:

```
<Nfur <- vst(as.matrix(Nfur))>  
<Nfur <- t(Nfur)>  
<test1 <- prcomp(Nfur, scale.=T)>
```

In the example above, we have opted to use a variance stabilizing transform (VST) normalization; however, several other methods of normalization can be substituted depending on sample conditions, such as batch effects. The `<prcomp>` command generates not only eigenvectors using the integer table but also generates the loadings for the principal components to be graphed. However, the labels and sample information are decoupled from the integer table during this process, so several additional lines of code are used to generate labels for various groupings within the data:

```
<nNfur <- c(rep('Nfur', 4))>  
<sNfur <- c('Condition 1 (1)', 'Condition 1 (2)', 'Condition 2 (1)',  
'Condition 2 (2)')>  
<dNfur <- c('Con1', 'Con1', 'Con2', 'Con2')>  
<Nfur <- cbind(nNfur, sNfur, dNfur, Nfur)>
```

Each of the lines above generates a list to label samples according to collection time point and collection phenotype. Once each of these lists is generated, append them to the original integer converted RPKM table to be referred for graphing along with the `<prcomp>` object.

47. Generate a PCA plot from the data. Open a pdf, generate the plot for the file, and close and disconnect the pdf from R:

```
<autoplot(test1, x=1, y=2, data=Nfur, size=10, colour='dNfur',  
shape='nNfur') +  
scale_colour_manual(values=c('purple', 'red'))  
scale_shape_manual(values=c(19)) +  
labs(title="N. furzeri PCA")>
```

In the above plot example, we have defined the x- and y-axis of the PCA plot to be the first and second principal component (PC), respectively $x=1, y=2$; however, any principal component can be chosen and assigned to these axes. It is advisable to generate plots using several of the PCs to gather information about what biological relevance the underlying chromatin may represent in PCA space. For example, in Figure 2B, PC1 shows excellent separation between samples and replicates based on the phenotype of the given samples (condition 1 in purple and condition 2 in red), whereas PC2 is less clear, likely encapsulating the more variable nature of condition 2 replicates compared with those from condition 1.

Plotting Data Aggregates Using Transcriptional Start Site Enrichment

Assessing the enrichment of reads at transcription start sites is a gold standard for determining the quality of ATAC-seq libraries. Transcriptional start site (TSS) accessibility has good correlation to overall gene expression, allowing for comparison to paired RNA-seq samples or interpolation of relative gene expression levels.

48. Evaluate ATAC-seq libraries at a global level by determining chromatin accessibility enrichment levels at TSSs.

- i. To generate a plot for TSS enrichment, use a bigWig-formatted file, generated from the finalized bam file for each sample, as input to generate a compressed matrix file using the following command:

```
<computeMatrix reference-point --referencePoint TSS -b 2000 -a 2000 -R TSS_Dir/ -S shift_dir/Library_FINAL.bw --missingDataAsZero --skipZeros -o TSS_Dir/Library_FINAL_matrix.gz>
```

Chromatin accessibility enrichment levels at TSSs is another useful analysis for evaluating ATAC-seq libraries at a global level. The TSS of any gene being actively expressed needs to be accessible; therefore, an expected result for accessible chromatin enrichment is between five- and 15-fold genome-wide. Libraries with lower enrichment may denote problems during sample preparation and/or data processing but may still be usable.

The command above generates a matrix of read pileup based on a reference point (in this case the TSSs of genes <--referencePoint TSS>) using a genome loaded onto the deeptools platform (Ramírez et al. 2014). Additionally, the command allows for the window upstream and downstream to be specified, though it generally should remain equal and broad for this analysis (<-b 2000> and <-a 2000>, respectively). We have also opted to treat missing TSS site data <--missingDataAsZero> and skip all entries containing zeros <--skipZeros> to prevent errors when running a less established or custom input genome.

- ii. Use the following command with the output matrix generated, <-o TSS_Dir/Library_FINAL_matrix.gz>, to produce a heatmap, as shown in Figure 2C:

```
<plotHeatmap -m TSS_Dir/Library_FINAL_matrix.gz -out TSS_Dir/Library_Final_matrix.pdf --colorList='red,black' --plotFileFormat pdf>
```

This graph allows for a global look at TSS accessibility; however, by extracting the gene names associated with each entry the user can generate a rank list of genes based on TSS chromatin accessibility. It is often useful to identify highly expressed genes at a time point because TSS accessibility and expression correlate in many studies (Buenrostro et al. 2013; Lara-Astiaso et al. 2014; Pastor et al. 2014; Ampuja et al. 2017).

49. Use the HOMER platform to assess transcription factor binding site motifs within a set of given genomic intervals.

- i. Set up the killifish genome within the HOMER package using the following script and command:

```
<loadGenome.pl -name nfurzeri -org nfurzeri -fasta ./genome.fa -gff3 genome.gff3 -promoters nfurpromoter.txt -version ncbi -gid>
```

Many genomes are preloaded into the HOMER package and ready for use. However, because of killifish being a less widely used model organism, its genome will need to be manually added. This is accomplished by providing homer with a fasta-formatted genome <-fasta ./genome.fa> and its associated gff3-formatted features file <-gff3 ./genome.gff3>. In addition, providing names for the integrated genome and organism (<-name> and <-org> arguments respectively) will allow HOMER to access the output file containing promoter/TSS coordinates <-promoters nfurpromoter.txt> generated by the script.

- ii. Use the following script and command to annotate motifs within ATAC-seq peaks:

```
<findMotifsGenome.pl Nfurzeri ./shift_Dir/Peak_Set.bed ./Motif_Dir/ -mset vertebrates>
```

Although TSS enrichment is an accepted metric for library quality, many projects will focus on accessible chromatin sites that are not at a TSS site, such as those that are intergenic or intronic, a class that by far dominates the contents of most ATAC-seq data. These chromatin accessibility sites are often thought of as putative enhancer and gene regulatory regions; therefore, evaluating the contents of their sequences allows for a deeper understanding of the link between chromatin accessibility and biological function.

The script above requires the user to designate the name of the genome to use as background. The killifish genome must be added to the HOMER platform manually following the HOMER custom genome upload manual. Here it is designated as <Nfurzeri> (Bailey et al. 2009; Boeva 2016). As input, the command takes a bed-formatted list of any peak set of interest <./shift_Dir/Peak_Set.bed> to be compared against the background genome. For Figure 2D, the consensus peak sets for condition 1 and condition 2 were used. Finally, the output directory <./Motif_Dir/> as well as the motif sequence set to be used <mset vertebrates> are provided. Because there is currently no killifish-specific motif sequence list, the general list for vertebrates was used. Once this is achieved, a ranked list of motif binding sites enriched in the test set compared with the overall genome is generated.

- iii. Use the following command to generate a heatmap for easy comparison of overlap and unique motifs across samples and conditions. This is performed by loading results for differentially accessible chromatin peaks in condition 1 and condition 2 samples into R and merging them into a single data file:

```
<merged <- Reduce(function(x,y) merge(x,y,by="Motif",
all=TRUE), list(condition1,condition2))>
<merged=as.data.frame(merged)>
<rownames(merged)=make.unique(merged$Motif)>
<condition1.significant=merged[(merged$condition1.fdr<=0.1
&merged$condition2.fdr>0.1),]>
<condition2.significant=merged[(merged$condition1.fdr>0.1 &
merged$condition2.fdr<=0.1),]>
<both.significant=merged[(merged$condition1.fdr<=0.1 &
merged$condition2.fdr<=0.1),]>
<toPlot=rbind(condition1.significant, condition2.
significant, both.significant)>
```

Once each enrichment file is read in using the *Killifish_ATACseq_Pipeline.R* script, the data will be merged based on the motif entry existing in both files <by="Motif" all=TRUE>. This new merge motif set is then converted to a data frame and the row names are reassigned. Filtering of this set can be performed for a specified significance cutoff or to subset to any desired condition, such as <merged\$condition1.fdr <= 0.1> or <merged\$condition2.fdr <= 0.1>.

- iv. Generate an output file containing the heatmap:

```
<Heatmap(as.matrix(toPlot[, c(3, 5)]), col=colorRamp2(c(0,
0.001, 0.01, 0.05, 0.1, 0.25, 0.45), c("orangered", "tomato",
"coral", "orange", "bisque", "gray90", "gray80")), cluster_
rows=F, cluster_columns=T, show_row_names=T,
row_names_gp=gpar(fontsize=6), row_names_max_width=unit
(30, "cm"))>
```

Using the <Heatmap()> plotting function we designated the columns containing the enrichment scores for each condition <as.matrix(toPlot[, c(3, 5)])> and applied a scaling factor and color list for illustrating the heatmap. This can be adapted to accommodate any number of conditions or subsets (Fig. 2E).

Visualizing Peak Loci

50. Generate visual chromatin tracks using the Integrative Genome Viewer (IGV)

- i. Download and install the latest version of the IGV app and open a new visualization window.

Although global metrics of ATAC-seq libraries provide interesting insights into the regulatory landscape of the genome, often the most compelling data comes from specific examples of chromatin accessibly sites and their loci. The IGV is an excellent platform to explore and capture compelling examples from the data and to generate visual chromatin tracks (Robinson et al. 2011).
- ii. Provide the path of the fasta-formatted genome and its associated index (i.e., Nfur.fa and Nfur.fai, respectively). Gene location and structure information can also be loaded if the path to a gff/gtf genomic feature file is given (i.e., Nfur.gtf).

BAM-formatted or bigWig-formatted files are supported. However, it is recommended to use BAM files for libraries with less than 40 million unique mapping reads for the best quality and resolution, whereas bigWig-formatted files are recommended for all libraries above this quality threshold. It is also advisable to close the read pile-up windows (lower) for each sample and use only the aggregate windows for viewing (upper).

- iii. With all data loaded into IGV, right click on individual tracks to manipulate their aesthetics for figure production.

These parameters include: summation of replicate tracks, setting window size/height, making the y-axis of tracks automatically congruent, changing track color, and many more. For figure-quality tracks, setting a standard window size for comparable panels is recommended. This can be done manually by adding/subtracting a set value from the center of a peak of interest or automated for all peaks (i.e., the function of the optional ATACseq_IGV_Coordinates.py script provided).

- iv. View individual chromatin accessibility peaks by entering their coordinates, or by searching for by name if a BED-formatted peak file is also loaded into IGV.
- v. Save the individual chromatin accessibility tracks using the “capture screenshot” function in the tool’s dropdown menu.

TROUBLESHOOTING

Problem (Step 15): Poor-quality libraries result caused by over- or undertransposed chromatin.

Solution: Adjust the time the sample is placed at 37°C to increase or decrease the degree of transposition.

REFERENCES

- Ampuja M, Rantapero T, Rodriguez-Martinez A, Palmroth M, Alarmo EL, Nykter M, Kallioniemi A. 2017. Integrated RNA-seq and DNase-seq analyses identify phenotype-specific BMP4 signaling in breast cancer. *BMC Genomics* 18: 68. doi:10.1186/s12864-016-3428-1
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucl Acids Res* 37: W202–W208. doi:10.1093/nar/gkp335
- Boeva V. 2016. Analysis of genomic sequence motifs for deciphering transcription factor binding and transcriptional regulation in eukaryotic cells. *Front Genet* 7: 24. doi:10.3389/fgene.2016.00024
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. doi:10.1093/bioinformatics/btu170
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 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. doi:10.1038/nmeth.2688
- Chen J, Khondker RC, Brunet A. 2023. Breeding and reproduction of the African turquoise killifish *Nothobranchius furzeri*. *Cold Spring Harb Protoc* doi:10.1101/pdb.prot107816
- Corporation GC. 1991. Sequencher: DNA sequence analysis software (Gene Codes Corporation). <http://www.genecodes.com> [Accessed March 9, 2023].
- Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, et al. 2022. Ensembl 2022. *Nucl Acids Res* 50: D988–D995. doi:10.1093/nar/gkab1049
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10: giab008. doi:10.1093/giga-science/giab008
- Dodzian J, Kean S, Seidel J, Valenzano DR. 2018. A protocol for laboratory housing of turquoise killifish (*Nothobranchius furzeri*). *J Vis Exp* 134: 57073. doi:10.3791/57073
- Doganli C, Sandoval M, Thomas S, Hart D. 2017. Assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-Seq) protocol for zebrafish embryos. *Methods Mol Biol* 1507: 59–66. doi:10.1007/978-1-4939-6518-2_5
- Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048. doi:10.1093/bioinformatics/btw354
- Hu CK, Brunet A. 2018. The African turquoise killifish: a research organism to study vertebrate aging and diapause. *Aging Cell* 17: e12757. doi:10.1111/acer.12757
- Kruger F. 2012. Trim Galore. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Accessed March 9, 2023].
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359. doi:10.1038/nmeth.1923
- Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretzky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S, et al. 2014. Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345: 943–949. doi:10.1126/science.1256271
- Liu T. 2014. Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol Biol* 1150: 81–95. doi:10.1007/978-1-4939-0512-6_4
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. doi:10.1186/s13059-014-0550-8
- Nath RD, Bedbrook CN, Nagvekar R, Brunet A. 2023. Husbandry of the African turquoise killifish *Nothobranchius furzeri*. *Cold Spring Harb Protoc* doi:10.1101/pdb.prot107738
- Pastor WA, Stroud H, Nee K, Liu W, Pezic D, Manakov S, Lee SA, Moissiard G, Zamudio N, Bourc’his D, et al. 2014. MORC1 represses transposable



G.A. Reeves et al.

- elements in the mouse male germline. *Nat Commun* 5: 5795. doi:10.1038/ncomms6795
- Platzer M, Englert C. 2016. *Nothobranchius furzeri*: a model for aging research and more. *Trends Genet* 32: 543–552. doi:10.1016/j.tig.2016.06.006
- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42: W187–W191. doi:10.1093/nar/gku365
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140. doi:10.1093/bioinformatics/btp616
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29: 24–26. doi:10.1038/nbt.1754
- Stark R, Brown G. 2011. Diffbind: differential binding analysis of ChIP-seq data. <http://bioconductor.org/packages/release/bioc/html/DiffBind.html>. [Accessed March 9, 2023].
- Terzibas E, Valenzano DR, Benedetti M, Roncaglia P, Cattaneo A, Domenici L, Cellerino A. 2008. Large differences in aging phenotype between strains of the short-lived annual fish *Nothobranchius furzeri*. *PLoS ONE* 3: e3866. doi:10.1371/journal.pone.0003866
- Valdesalici S, Cellerino A. 2003. Extremely short lifespan in the annual fish *Nothobranchius furzeri*. *Proc Biol Sci* 270: S189–S191. doi:10.1098/rsbl.2003.0048
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137. doi:10.1186/gb-2008-9-9-r137





Cold Spring Harbor Protocols

Chromatin Accessibility Profiling and Data Analysis using ATAC-seq in *Nothobranchius furzeri*

G. Adam Reeves, Param Priya Singh and Anne Brunet

Cold Spring Harb Protoc; doi: 10.1101/pdb.prot107747; published online April 26, 2023

Email Alerting Service

Receive free email alerts when new articles cite this article - [click here](#).

Subject Categories

Browse articles on similar topics from *Cold Spring Harbor Protocols*.

[African Turquoise Killifish](#) (29 articles)

[African Turquoise Killifish: A Laboratory Manual](#) (30 articles)

[DNA Sequencing](#) (99 articles)

[Emerging Model Organisms](#) (325 articles)

To subscribe to *Cold Spring Harbor Protocols* go to:
<http://cshprotocols.cshlp.org/subscriptions>
