

Video Article

# Mapping Genome-wide Accessible Chromatin in Primary Human T Lymphocytes by ATAC-Seq

Ivana Grbesa<sup>1</sup>, Miriam Tannenbaum<sup>1</sup>, Avital Sarusi-Portuguez<sup>1</sup>, Michal Schwartz<sup>1</sup>, Ofir Hakim<sup>1</sup>

<sup>1</sup>The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University

Correspondence to: Ofir Hakim at [Ofir.Hakim@biu.ac.il](mailto:Ofir.Hakim@biu.ac.il)

URL: <https://www.jove.com/video/56313>

DOI: [doi:10.3791/56313](https://doi.org/10.3791/56313)

Keywords: Genetics, Issue 129, Chromatin accessibility, ATAC-seq, human CD4+ lymphocytes, next-generation sequencing, regulatory elements, enhancer, promoter, Th1, Th2

Date Published: 11/13/2017

Citation: Grbesa, I., Tannenbaum, M., Sarusi-Portuguez, A., Schwartz, M., Hakim, O. Mapping Genome-wide Accessible Chromatin in Primary Human T Lymphocytes by ATAC-Seq. *J. Vis. Exp.* (129), e56313, doi:10.3791/56313 (2017).

## Abstract

Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) is a method used for the identification of open (accessible) regions of chromatin. These regions represent regulatory DNA elements (e.g., promoters, enhancers, locus control regions, insulators) to which transcription factors bind. Mapping the accessible chromatin landscape is a powerful approach for uncovering active regulatory elements across the genome. This information serves as an unbiased approach for discovering the network of relevant transcription factors and mechanisms of chromatin structure that govern gene expression programs. ATAC-seq is a robust and sensitive alternative to DNase I hypersensitivity analysis coupled with next-generation sequencing (DNase-seq) and formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) for genome-wide analysis of chromatin accessibility and to the sequencing of micrococcal nuclease-sensitive sites (MNase-seq) to determine nucleosome positioning. We present a detailed ATAC-seq protocol optimized for human primary immune cells *i.e.* CD4+ lymphocytes (T helper 1 (Th1) and Th2 cells). This comprehensive protocol begins with cell harvest, then describes the molecular procedure of chromatin tagmentation, sample preparation for next-generation sequencing, and also includes methods and considerations for the computational analyses used to interpret the results. Moreover, to save time and money, we introduced quality control measures to assess the ATAC-seq library prior to sequencing. Importantly, the principles presented in this protocol allow its adaptation to other human immune and non-immune primary cells and cell lines. These guidelines will also be useful for laboratories which are not proficient with next-generation sequencing methods.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/56313/>

## Introduction

ATAC-seq<sup>1,2</sup> is a robust method that enables identification of regulatory<sup>3</sup> open chromatin regions and nucleosome positioning. This information is applied for inferring the location, identity, and activity of transcription factors. The method's sensitivity for measuring quantitative variations in chromatin structure allows the study of the activity of chromatin factors, including chromatin remodelers and modifiers, as well as the transcriptional activity of RNA polymerase II<sup>1</sup>. Thus ATAC-seq provides a powerful and unbiased approach for deciphering mechanisms that govern transcriptional regulation in any cell type of interest. We describe the adaptation of ATAC-seq to primary human Th1 and Th2 cells.

In ATAC-seq, hyperactive Tn5 transposase loaded with adaptors for next-generation sequencing (NGS) couples DNA fragmentation with tagging of DNA with adaptors (*i.e.*, the "tagmentation" process)<sup>1</sup>. Following PCR amplification, the resulting DNA libraries are ready for next-generation sequencing (**Figure 1**). The preferential tagmentation of accessible chromatin is detected by the analysis of local enrichment of ATAC-seq sequencing reads.

The short experimental procedure and requirement for less starting material, relative to other methods for measuring chromatin accessibility and nucleosomal positioning such as DNase-seq<sup>4</sup>, FAIRE-seq<sup>5</sup>, and MNase-seq<sup>6</sup>, has promoted the use of ATAC-seq in multiple biological systems including human primary cells<sup>1,7</sup> and clinical samples<sup>8</sup>, as well as unicellular organisms<sup>9</sup>, plants<sup>10</sup>, fruit flies<sup>11</sup>, and various mammals<sup>12</sup>.

The identity of transcription factors which are bound to accessible loci can be uncovered by analyzing the enrichment of their binding sequence motifs or combining ATAC-seq with chromatin immunoprecipitation (ChIP) followed by high-throughput DNA sequencing (ChIP-seq). This approach enabled the identification of lineage-specific transcription factors important for hematopoiesis in mouse<sup>13</sup>. The unbiased and global nature of ATAC-seq allows studying gene regulation in organisms for which reagents such as antibodies for ChIP analysis are not available. For example, evolutionary variations in *cis*-regulatory regions have been identified by studying cranial neural crest cells from humans and chimpanzees<sup>14</sup>, developmental variations in regulatory elements during early mouse embryogenesis<sup>15</sup>, changes in the regulatory landscape during a life cycle of unicellular *C. owzarzaki*<sup>9</sup>, and the evolution of promoters and enhancers across 20 mammal species<sup>12</sup>.

ATAC-seq has also been instrumental for measuring chromatin accessibility in single cells, thus revealing variability within cell populations, which usually evades genome-wide studies<sup>7,16</sup>. In addition, ATAC-seq can be used to study changes that occur in DNA regulatory regions in disease

conditions, in which samples are rare. For example, ATAC-seq can be used to study changes in the regulatory landscape during the onset of acute myeloid leukemia (AML)<sup>17</sup> or Ras-driven oncogenesis<sup>11</sup>.

## Protocol

All the procedures were approved by institutional review board of Bar Ilan University and the protocol follows guidelines provided by the committee approving the experiments.

## 1. Purification of Naïve Human CD4<sup>+</sup> Cells and Polarization to T Helper 1 (Th1) and Th2 Cells

Note: Here we describe the procedure starting from frozen human peripheral blood mononuclear cells (PBMCs). The first step consists of isolating CD4<sup>+</sup> cells using microbeads and columns that usually give us more than 95% of CD4<sup>+</sup> cells. However, this step may vary according to the preferred protocol in each lab. The protocol for T cell activation and polarization was modified from Jenner *et al.* (2009)<sup>18</sup>. Isolation of CD4<sup>+</sup> cells from 10 million PBMCs gives rise to 4 - 6 million of CD4<sup>+</sup> cells. They are split into two flasks and grown under Th1 and Th2 polarizing conditions yielding 3-5 million Th1 and Th2 cells within just a week.

Note: Cool down the centrifuge to 4 °C before starting.

1. Thaw 1 mL of human PBMCs (10<sup>7</sup> cells) in a 50 mL tube containing 10 mL of RPMI medium supplemented with 1% Penicillin-Streptomycin, 2 mM L-glutamine and 10% heat-inactivated fetal bovine serum. Centrifuge at 500 x g for 5 min. Remove the supernatant and resuspend the cells with a sterile 25 mL pipette in 15 mL of supplemented RPMI medium. Transfer the cells (with sterile 25 mL pipette) to a T75 culture flask.
2. Leave the cells overnight in a humidified incubator (37 °C, 5% CO<sub>2</sub>).
3. Transfer the floating cells with a sterile 25 mL pipette to 50 mL tube. Determine cell number and viability by trypan blue exclusion.
4. Isolate CD4<sup>+</sup> cells from 10 million live non-adherent PBMCs by positive selection using CD4<sup>+</sup> microbeads and columns according to the manufacturer's recommendations (see **Table of Materials/Equipment**) with the following modifications: 10<sup>7</sup> PBMCs are labeled with 30 µL of CD4 microbeads in 120 µL of 0.5% BSA in PBS.
5. Activate the CD4<sup>+</sup> T cells for 72 h by rhIL-2 (10 ng/mL), plate-bound anti-CD3 (5 µg/mL) and soluble anti-CD28 (2 µg/mL). For Th1 polarization, add rhIL-12 (20 ng/mL) and anti-IL-4 (10 µg/mL). For Th2 polarization, add rhIL-4 (40 ng/mL) and anti-IFN-γ (10 µg/mL).
6. Culture the cells for additional 7 days in the presence of rhIL-2 (10 ng/mL) and the same polarizing cytokines (rhIL-12 for Th1 and rhIL-4 for Th2).

## 2. Nuclei Isolation

NOTE: ATAC-seq is performed with intact nuclei. Lysis buffer containing 0.05% nonylphenyl polyethylene glycol (see **Table of Materials/Equipment**) was calibrated for isolating nuclei from primary human Th1 and Th2 cells. We recommend calibrating this step with the laboratory reagents and cells. An excess of intact cells from insufficient detergent decreases the efficiency of the transposition reaction. Cell lysis efficiency is determined by the number of nuclei (trypan blue positive cells) relative to the total number of cells.

NOTE: Prepare the lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub>). Cool down a centrifuge with a swing-bucket rotor to 4 °C. Pelleting the cells in a swing-bucket centrifuge instead of a fixed angle centrifuge reduces cell/nuclei loss. To avoid nuclei or cell loss, pipet carefully when discarding the supernatant.

1. Add fresh nonylphenyl polyethylene glycol (to a final concentration of 0.05%) and 100x protease inhibitors (to a final concentration of 1x) to the cold lysis buffer immediately before use. Keep the buffer on ice.
2. Count T cells to determine their amount and viability using trypan blue method. Viability that is lower than 90% results in higher non-specific digestion.
3. Transfer 0.5 x 10<sup>6</sup> T cells (Th1 or Th2) to 1.5 mL microtubes. Spin down at 500 x g for 5 min at 4 °C.
4. Resuspend the cell pellet in 1 mL of cold phosphate-buffered saline (PBS) solution. Spin down at 500 x g for 5 min at 4 °C.
5. Resuspend the cell pellet in 1 mL of cold lysis buffer (containing nonylphenyl polyethylene glycol and protease inhibitors). Keep the tube on ice. Pipette gently to avoid disrupting the nuclei.
6. Quickly take 10 µL and count the cells with an automated cell counter while the microtube with lysed cells is on ice. This step should not exceed five minutes to avoid damaging the nuclei. At least 80% of the cells should be lysed.
7. Continue immediately with the transposition reaction. Keep the prepared nuclei on the ice.

## 3. Transposition reaction

NOTE: In this step, isolated nuclei are incubated with prokaryotic Tn5 transposase (TDE1) loaded with adapters for NGS sequencing. Hyperactive Tn5 simultaneously fragments DNA and ligates adapters into accessible regions of the genome (tagmentation process). The ratio between nuclei and Tn5 transposase is critical for preferential cleavage at accessible chromatin. This protocol is calibrated for 100,000 nuclei in a 100 µL reaction volume. However, the reaction can be scaled down.

1. Set temperature in a thermal shaker to 37 °C.
2. Transfer 100,000 nuclei to a 1.5 mL microtube.
3. Centrifuge at 500 x g for 10 min at 4 °C and gently remove the supernatant.
4. Add the transposition reaction components to the nuclei as specified in **Table 1**.
5. Resuspend by gentle pipetting.
6. Incubate the transposition reaction in a thermal shaker at 37 °C for 30 min with gentle shaking (500 rpm).

Note: DNA Cleanup is performed by solid-phase reversible immobilization beads<sup>19</sup> (see **Table of Materials/Equipment**) or PCR purification columns. At the end of the cleanup, elute the DNA fragments in 20  $\mu$ L of 10 mM Tris-HCl, pH 8. Avoid EDTA in the elution buffer.

## 4. PCR Enrichment of ATAC-seq Libraries

NOTE: This step is aimed to amplify the ATAC-seq library *i.e.*, DNA fragments with inserted adapters. To allow mixing of several ATAC-seq libraries in the same next-generation sequencing lane ("multiplexing") use unindexed Primer 1 (Ad1\_noMx)<sup>1</sup> for all samples, and a different indexed (barcoded) Primer 2 (Ad 2.1 - 2.24)<sup>1</sup> for each sample. The primer sequences are provided in the supplemented **Table of Materials/Equipment**.

### 1. Initial PCR amplification

NOTE: The working concentration of Primer 1 (Ad1\_noMx) and Primer 2 is 25  $\mu$ M. All the primers are diluted from original stock of 100  $\mu$ M to 25  $\mu$ M. In all of the PCR reactions, use Primer 1 (Ad1\_noMx) and only one of the indexed Primer 2.

1. Add the components of the PCR reaction as specified in **Table 2** to a sterile PCR tube.
2. Place PCR tube into a thermal cycler and perform PCR amplification using the cycling conditions detailed in **Table 3**.

### 2. Assessment of number of additional amplification cycles

NOTE: The number of additional PCR cycles should yield sufficient amount of library fragments for a successful next-generation sequencing run, while minimized to avoid GC and size bias<sup>20</sup>. The determination of the number of PCR cycles (*N*) required for optimal library fragment amplification is done by quantitative PCR (qPCR).

1. Dilute Primers 1 (Ad1\_noMx) and 2 (used for initial library amplification) from 25  $\mu$ M to 6.25  $\mu$ M.
2. Add the components to optical PCR tubes or a plate as stated in **Table 4**.
3. Place in a qPCR instrument and cycle as specified in **Table 5**.
4. To estimate the required number of additional amplification cycles (*N*), plot cycle number on the x-axis and relative fluorescence (RFU) on the y-axis.
5. The number of additional amplification cycles (*N*) is 1/3 of the number of cycles at which the qPCR reaction reached the plateau. **Figure 2** provides examples for three ATAC-seq libraries that reached plateau at ~2,350 relative fluorescence units, RFU (thick green line). The number of PCR cycles in which one third of the maximal amount (783 RFU, marked on y-axis) is amplified corresponds to 8 cycles for two of the libraries (red and blue amplification curves) and 9 PCR cycles for the third library (pink).

### 3. Final PCR amplification

1. Amplify the remaining 45  $\mu$ L of the PCR reaction. Place a PCR tube containing amplification reaction from step 4.1.2 in a thermal cycler. Run the PCR program described in **Table 6**. Use the previously determined (step 4.2.5) number of amplification cycles (*N*).

## 5. Size Selection of ATAC-Seq Libraries

NOTE: In our experience, size selection of amplified ATAC-seq libraries improves next-generation sequencing results because it eliminates high molecular weight library fragments from the final ATAC-seq library.

NOTE: Allow the magnetic beads to warm to room temperature 30 min before use.

Prepare fresh 70% ethanol in nuclease-free water.

1. Resuspend the magnetic beads by mixing.
2. Add nuclease-free water to the ATAC-seq libraries (obtained in step 4.3.1.) and bring up to 100  $\mu$ L.
3. Add 50  $\mu$ L (0.5x) of resuspended DNA-binding magnetic beads to 100  $\mu$ L of amplified libraries. Mix by pipetting up and down at least 10 times. Incubate samples for 5 min at room temperature. If necessary, quickly spin down the microtubes.
4. Place the tube on an appropriate magnetic stand for 2 min to separate the magnetic beads from the supernatant. After 2 min, transfer the supernatant to a new microtube.
5. Measure the volume of the supernatant by pipetting and add 0.7x of magnetic beads. Mix by pipetting up and down at least 10 times.
6. Incubate 5 min at room temperature. Place on a magnetic stand for 2 min.
7. After the 2 min incubation, discard the supernatant. Add 200  $\mu$ L freshly made 70% ethanol to wash the beads while the tubes are on the magnetic stand.
8. Keep the microtube on the magnet for 30 s and then discard the ethanol. Repeat step 5.7. for two final ethanol washes.
9. Completely remove the remaining ethanol and let the beads air-dry for 5 min while the tube is on the magnet. If necessary, briefly spin the microtube. Remove the traces of ethanol with a p10 pipette tip.
10. Remove the microtube from the magnet and add 22  $\mu$ L of 10 mM Tris-HCl, pH 8. Do not elute the ATAC-seq libraries in buffer containing EDTA.
11. Incubate the tube for 2 min at room temperature and then place on the magnetic stand.
12. When the solution is clear, transfer 20  $\mu$ L of eluted libraries to a new sterile microtube.
13. Store the size selected ATAC-seq libraries at -20 °C.

## 6. Quality Analysis of the ATAC-Seq Libraries

### 1. Validation of the quality of ATAC-seq libraries by Real-Time PCR

NOTE: It is important to assess the signal to noise ratio of the ATAC-seq libraries prior to next-generation sequencing. This is done by determining the relative amount of DNA fragments from accessible and inaccessible loci using quantitative PCR (qPCR). The inaccessible loci (negative control, chr1:48,137,860-48,137,934 and chr1:193,093,748-193,093,827) are amplified by primer pairs 1 and 2. The accessible loci (positive control, chr19:30,336,166-30,336,253 and chr19:11546154-11546237) are amplified by primer pairs 3 and 4. Positive and negative loci were defined from chromatin accessibility (DHS-seq) profiles of human CD4+ cells (ENCODE accessions ENCSR000EQE and

ENCSR000EQG). Negative primer 1 is located in a large heterochromatic intergenic region (230 kb from TRADB2 and 88 kb from FOXD2). Negative control region 2 is in within the first intron of CDC73 gene. Positive primer pair 3 is within an open chromatin region downstream of the Cyclin E (CCNE1) gene while positive primer pair 4 is centered within the promoter of protein kinase C substrate 80K-H (PRKCSH). Importantly, these control loci show a similar pattern of accessibility in other human cell types from the ENCODE project<sup>3</sup>, suggesting that they can be applied to monitor ATAC-seq libraries from a broad spectrum of human cell types. Primer efficiency and specificity of all primer pairs was verified by qPCR on a serial dilution of genomic DNA (from human Th cells) and a melting curve analysis of obtained amplified products.

1. Isolate genomic DNA using a commercially available kit (see **Table of Materials/Equipment**).
2. Dilute the amplified ATAC-seq library 1:10 (1  $\mu$ L library + 9  $\mu$ L of nuclease-free water) and genomic DNA to ~5 ng/ $\mu$ L.
3. Prepare reaction mixture (**Table 7**) for each positive and negative control primer pair, taking into account that the reactions are performed in triplicate.
4. Incubate in qPCR thermal cycler according to the protocol recommended by qPCR master mix supplier.
5. Analyze the results in qPCR instrument's software (see **Table of Materials/Equipment**). Choose genomic DNA as a control sample. The obtained values represent an enrichment of accessible regions (amplified by positive control primers). An example is shown in **Table 8**.

NOTE: Estimation of the average library size and concentration: the size distribution of DNA fragments from ATAC-seq libraries is determined by high-sensitivity automated electrophoresis systems according to the manufacturer's instructions. It is advisable to measure the sample concentration on a fluorometer using dsDNA high-sensitivity kit and at least 2  $\mu$ L of each DNA sample.

NOTE: Next-generation sequencing - prior to multiplexing the libraries, calculate the molarity of each ATAC-seq library by using the formula: (ng/ $\mu$ L  $\times 10^6$ )/(660  $\times$  average library fragment length). Aim for > 30 million reads of each ATAC-seq library to assess open chromatin regions of human samples. If you wish to determine if the library is good enough for NGS sequencing, initially aim for ~10 million reads (5% of the sequencing lane on DNA sequencing instrument in rapid run mode). Keep in mind that to infer nucleosome positioning, paired-end sequencing is needed<sup>1</sup>.

## 7. Analysis of the Obtained Next-Generation Sequencing Results

1. Infer the quality of the sequencing reads by inspecting the FastQC files, separately for every library.
2. Align the reads to human reference genome (hg19 assembly) using Bowtie<sup>21</sup> software in the Unix/Linux environment. The command is 'bowtie -m 1 -q -S genome\_directory.reads.fastq output\_aligned.sam'. Genome directory stands for the folder where the genome indexes of Bowtie are stored. The parameter -m 1 is for not allowing alignment of reads to more than one locus in the genome, -q is for the input file that should be in fastq format, -S is for the output that is in SAM format.
3. Remove duplicate reads using SAMtools<sup>22</sup> rmdup option in the Unix/Linux environment. The commands are 'samtools view -S output\_aligned.sam -b | samtools sort -o output\_aligned\_rmdup.bam', 'samtools rmdup -s output\_aligned.bam output\_aligned\_rmdup.bam'. The first command, view, changes the SAM format into BAM format which is then sorted. The option of rmdup is then applied on the sorted BAM file. Optionally one can adjust the reads for the transposon insertion site, as described in the original ATAC-seq paper<sup>1</sup>. This is done using BEDtools commands<sup>23</sup> in the Unix/Linux environment. The commands are 'bamToBed -i output\_aligned\_rmdup.bam > output\_aligned\_rmdup.bed', 'shiftBed -i output\_aligned\_rmdup.bed -p 4 -m -5 -g genome > output\_aligned\_rmdup\_adjusted.bed'. The first command, bamToBed, changes the BAM format into BED format which can then be used for the shiftBed command. The genome file is a tab delimited file that contains the length of each chromosome in the genome. The file is usually added in the BEDtools directory.
4. Perform peak calling using model-based analysis of ChIP-seq (MACS2)<sup>24</sup> software in the Unix/Linux environment on the shifted BED file with the following parameters: --nomodel --extsize 75 --shift -30. These parameters are used to adjust the reads so that the transposon insertion site is in the middle of each sequencing read.

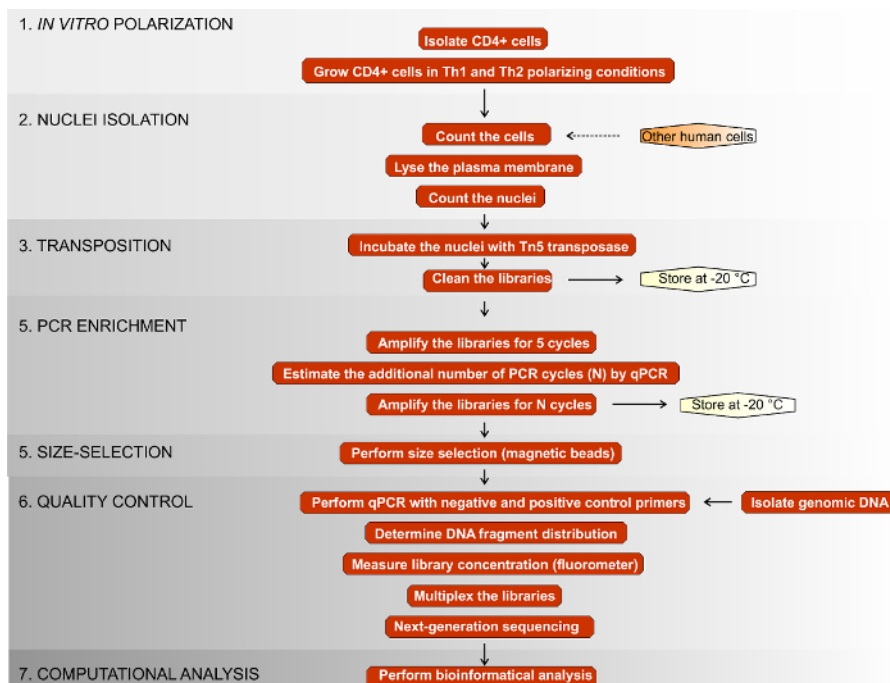
## Representative Results

The final outcome of this protocol is an ATAC-seq library of typically 3 - 20 ng/ $\mu$ L. When run on a system for DNA integrity analysis (see **Table of Materials/Equipment**), they show ladder-like appearance<sup>2</sup> (**Figure 3A**). The average size of DNA fragments is typically ~450 - 530 bp.

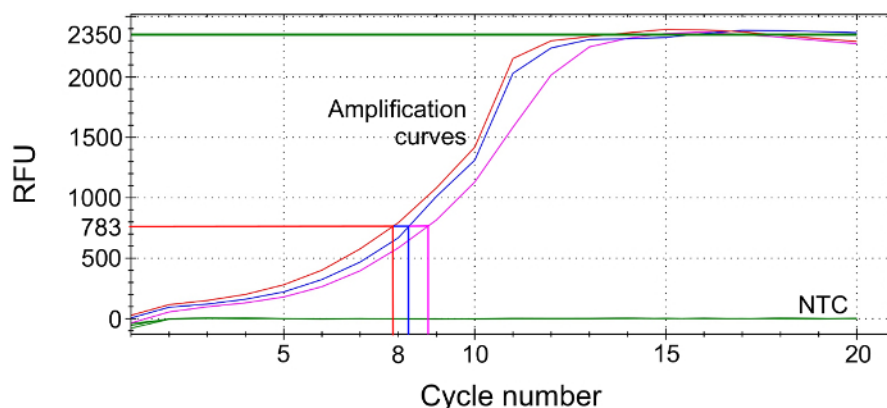
Proper quality control of the ATAC-seq libraries prior to performing next-generation sequencing is important to save time and money. We consider the libraries suitable for next-generation sequencing (NGS) when the enrichment of positive controls is more than 25- and 75-fold (for primer pairs 3 and 4, respectively) relative to negative controls (**Figure 3B**) and when they show the previously mentioned nucleosomal, ladder-like appearance. When analyzing qPCR data, we are also verifying that the Cq values for negative and positive control primers are similar in the reactions with the genomic DNA as a template and that the Cq values for the negative control regions (in the reactions with ATAC-seq DNA fragments) are constant between the experiments. For example, in the case of suddenly lower Cq values, we suspect that the nuclei were over-transposed and thus DNA fragments originating from heterochromatin regions are over-represented. Moreover, if the gel image (obtained by high-sensitivity tape-based electrophoresis) of ATAC-seq libraries indicates the presence of the adaptors (~120 bp), excessive DNA degradation (the majority of the fragments are ~200 bp), or large fragments (over 1 kb), we would not continue to the NGS step.

Furthermore, we always perform initial NGS in which we aim for 10 million reads per library. If the results of this sequencing are satisfactory (the sample passes all the parameters in FastQC report file and we can obtain 1000 - 2000 peaks after performing peak calling), the libraries are sequenced more deeply (more than 30 million reads/ATAC-seq library).

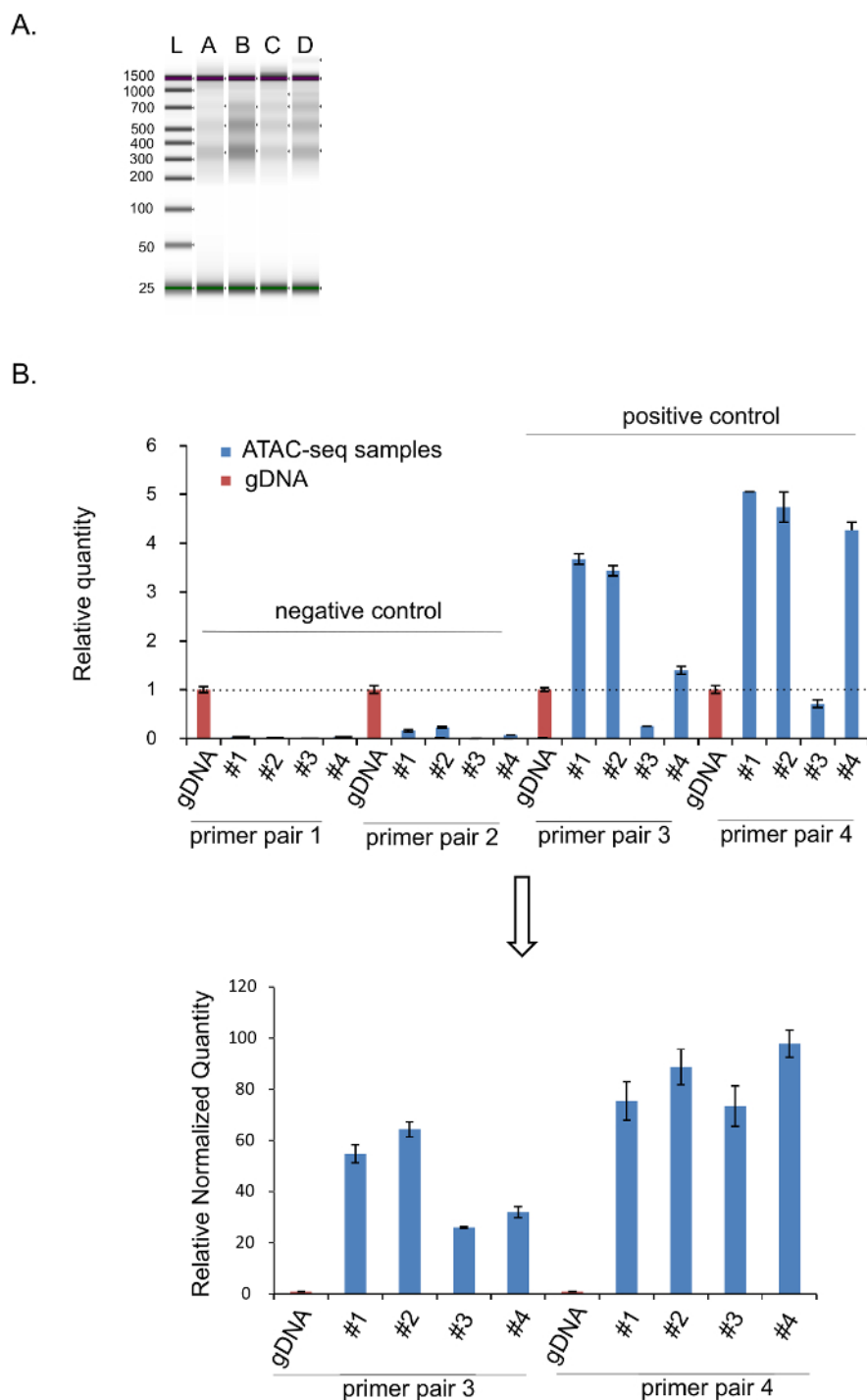
In ATAC-seq experiments on mammalian cells, anywhere from ~30 - 70% of the sequenced reads can come from mtDNA<sup>10</sup>. In contrast, our libraries contained 6-20% reads mapped to the mitochondrial genome. This is lower than the reported 46% in the ATAC-seq library from human CD4+ T lymphocytes<sup>1</sup>. After eliminating these contaminating reads, we were left with ~7 - 32 million unique reads mapping to the reference human genome (58%-91% of all reads). From these, 0.1 - 2 million reads (6.8% - 12% of all reads) were in ATAC-seq peaks (**Figure 4**).



**Figure 1. Workflow of experimental steps. Representation of the major steps in this ATAC-seq protocol.** Stopping points are indicated by yellow hexagons. An optional step is represented with an orange hexagon. [Please click here to view a larger version of this figure.](#)

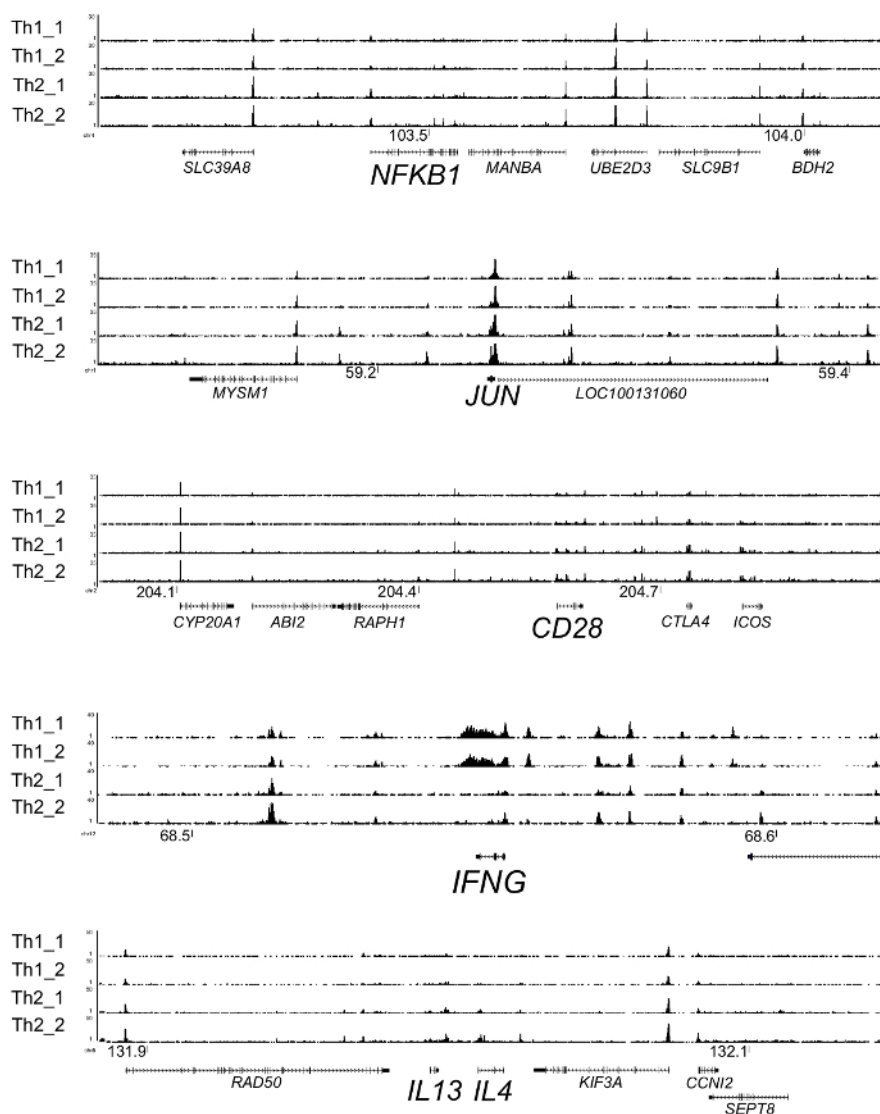


**Figure 2. Calculation of the additional number of PCR cycles required for the final enrichment of ATAC-seq libraries.** PCR amplification curves for three different ATAC-seq libraries are shown in red, blue and pink. The amplification reactions reached a plateau at 2,350 RFU (upper green horizontal line). The number of additional PCR cycles intersects with 783 RFU (2,350/3) on the amplification curve. Two libraries (red and blue) require 8 additional amplification cycles while the third library (pink) requires 9 cycles. Non-template control (NTC) is shown as the lower green line. X-axis, cycle number. Y-axis, RFU units. [Please click here to view a larger version of this figure.](#)



**Figure 3. Quality control analysis of ATAC-seq libraries.** (A) Results of tape-based automated electrophoresis of amplified ATAC-seq libraries. L, ladder; A-D, libraries of biological repeats from Th1 and Th2 cells. (B) qPCR quality control analysis of ATAC-seq libraries. Genomic DNA and four ATAC-seq libraries were amplified by negative control primer pairs (1 and 2) and positive control primers (3 and 4). The obtained values for ATAC-seq libraries (blue) were first normalized to the amplification levels of genomic DNA (arbitrarily set to a value of 1; red). Subsequently, values for the positive control regions were normalized to the negative control regions (lower chart). Libraries #1 and #2 were sent to NGS sequencing. The values represent mean  $\pm$  SEM. [Please click here to view a larger version of this figure.](#)





**Figure 4. Genome browser snapshots of accessible chromatin regions in Th1 and Th2 cells.** ATAC-seq tracks are shown for *NFKB1*, *JUN*, *CD28*, *IFNG* (expressed only in Th1) and *IL4* and *IL13* (expressed only in Th2) loci in Th1 and Th2 cells (two biological repeats for each). [Please click here to view a larger version of this figure.](#)

TD buffer	50	μL
TDE1 transposase	5	μL
Nuclease-free water	45	μL
Total volume	100	μL

**Table 1: Preparation of transposition reaction mixture.**

ATAC-seq library	15	μL
Primer 1 (Ad1_noMx) *	2.5	μL
Primer 2 *	2.5	μL
NEBNext High-Fidelity 2 x PCR Master Mix**	25	μL
Nuclease-free water	5	μL
Total volume	50	μL

\*Final concentration of each primer is 1.25 μM.

\*\*PCR reagent provided in the kit are not recommended

(Buenrostro et al., 2015. ATAC-seq: A method for assaying chromatin accessibility genome-wide) .

In addition, we have also performed successful amplifications with

NEBNext Q5 Hot Start HiFi PCR Master Mix (extension temperature is 65 °C).

**Table 2: Components of initial PCR reaction mix (step 5.1.1.).**

CYCLE STEP	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	3 min	
Hold	4 °C	∞	

\*This step is required to extend both

ends of the primers after transposition reaction.

**Table 3: PCR cycling conditions for initial library amplification (step 5.1.2.)**

Aliquot of PCR reaction (step 4.1.1) *	5	μL
Primer 1 (Ad1_noMx)**	1	μL
Primer 2	1	μL
2x SYBR master mix ***	7.5	μL
Nuclease-free water	0.5	μL
Total volume	15	μL

\*For non-template control instead of the DNA template add ultra-pure water.

\*\*The final concentration of each primer is 417 nM.

\*\*\*Use preferred master mix for qPCR.

It should contain polymerase, dNTPs, MgCl<sub>2</sub>, and fluorescent dye.

**Table 4: Preparation of qPCR reaction mixture (step 5.2.2).**

CYCLE STEP	TEMPERATURE	TIME	CYCLES
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	20
Annealing	63 °C	30 s	
Extension	72 °C	3 min	
Hold	4 °C	∞	

**Table 5: Cycling conditions for qPCR-based assesment of additional number of amplification cycles (N) (step 5.2.3.)**



CYCLE STEP	TEMPERATURE	TIME	CYCLES
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	N
Annealing	63 °C	30 s	
Extension	72 °C	3 min	
Hold	4 °C	∞	

**Table 6: Cycling conditions for the final PCR enrichment step (step 5.3.).**

For a single PCR reaction:		
Dilution of the library of genomic DNA	2.5	μL
10 μM Primer F (F1 or F2 or F3 or F4)*	0.3	μL
10 μM Primer R (R1 or R2 or R3 or R4)**	0.3	μL
2x SYBR master mix ***	5	μL
Nuclease-free water	1.9	μL
Total volume	10	μL
*The final concentration of each primer in the reaction mixture is 300 nM.		
**If using primer F1 than the R primer should be R1 etc.		
***Use preferred 2x master mix suitable for the qPCR instrument in your laboratory.		

**Table 7: Reaction conditions for QC analysis (step 7.1.3.).**

Target		Sample	Mean Cq	Cq SEM	Relative quantity	Normalized
1	Negative control	ATAC-seq	30.39	0.11	1	1.01
		gDNA	30.4	0.16	0.99	1
2		ATAC-seq	30.3	0.18	1	1
		gDNA	30.34	0.09	0.99	1
3	Positive control	ATAC-seq	23.27	0.03	1	<b>173.14</b>
		gDNA	30.7	0.06	0.01	1
4		ATAC-seq	21.55	0.24	1	<b>750.31</b>
		gDNA	31.1	0.03	0	1

**Table 8: Calculation of the enrichment of positive controls over negative controls (step 7.1.5.).**

## Discussion

The ATAC-seq protocol described here has been successfully employed for the analysis of accessible chromatin in primary cells (human Th1, Th2 cells, and B cells) as well as the cultured cell lines (MCF10A human breast cancer cells and U261 glioblastoma cells). Applying ATAC-seq to other cell types may require some protocol optimization, especially in the lysis step. If the concentration of non-ionic detergent is too high, there may be a higher percentage of mitochondrial DNA contamination. This can be reduced by decreasing the detergent concentration in the lysis buffer without reducing the nuclei yield. In our experience, lysis buffer with 0.05% of detergent gave the most satisfactory results. In addition, spinning the nuclei in a swing-bucket instead of fixed angle rotor allowed reducing the G force to 500, which reduced the mitochondria in the pellet. Researchers may also test other types of detergents, for example, digitonin<sup>17</sup>. However, even with optimized lysis conditions, mtDNA contamination is inevitable. To completely remove the mtDNA, the CRISPR/Cas9 system can be used<sup>15</sup>. In this case, ATAC-seq libraries are incubated with sgRNAs to target mtDNA and Cas9 nuclease to digest mtDNA<sup>15</sup>.

The number of nuclei required for this ATAC-seq protocol (100,000) may be limiting in cases where cell number from clinical samples is scarce<sup>7</sup>. ATAC-seq has been successfully scaled down to 5,000 and 500 cells<sup>1,13,17</sup> by reducing the reaction volume<sup>1,13</sup>, performing cleanup with magnetic beads instead of with columns<sup>13</sup>, or avoiding the cleanup step<sup>1</sup>. In addition, single-cell ATAC-seq (scATAC-seq) may be performed using a microfluidic device to separate individual nuclei<sup>16</sup>. Notably, other methods for measuring chromatin accessibility, such as DNase-seq and FAIRE-seq, require more starting material and a few more days to perform the experiment.

It is now widely appreciated that biases in various NGS methods are common phenomena<sup>25</sup>. One of the causes for variations in different chromatin accessibility measures is the enzymatic cleavage step<sup>25</sup>. It has been shown that DNase I shows cleavage bias<sup>26</sup> and now it is recognized that Tn5 transposase shows cleavage preference as well<sup>27</sup>. Fortunately, the potential bias can be identified computationally by using a quality control pipeline such as ChiLin<sup>28</sup>. Another common source of bias is the PCR amplification step<sup>20,25</sup>. This frequently manifests as a bias

for the amplification of GC-rich fragments<sup>25</sup>. Since the bias increases with every PCR amplification step, we do not exceed 9 additional cycles (N) in the final PCR amplification of ATAC-seq libraries.

The proper ratio between Tn5 transposase and nuclei is critical for successful ATAC-seq<sup>2</sup>. An excess of enzyme would lead to "over transposition" in inaccessible loci and high background. This is reflected in high amplification of the negative control regions in the qPCR quality control step. An excess of nuclei would lead to "under-transposition". In this case, too distant cleavage sites will result in fewer PCR-amplified fragments and low library complexity. To determine the number of nuclei accurately, we introduced an additional counting step after cell lysis and prior to the transposition reaction.

Chromatin accessibility is an important epigenetic regulatory layer as it permits the activity of transcription regulators at specific genomic locations. Thus, the DNA sequence within the accessible chromatin profile provides a wealth of information about the identity and target loci of tens of possible transcription factors. Combining this information (obtained by ATAC-seq) with an RNA expression profile allows a focus on the relevant transcription factors that can be further analyzed by ChIP-seq<sup>13,22</sup>. In addition, only 10% of disease-associated polymorphisms are in the coding genome<sup>30</sup>. Thus applying ATAC-seq to clinical samples can reveal which of the non-coding variants are at regulatory elements that may affect gene regulation in the disease state (for example, in systemic lupus erythematosus<sup>8</sup>). We hope that this ATAC-seq protocol will help and encourage interested investigators to use this powerful method to advance their research.

## Disclosures

The authors have nothing to disclose.

## Acknowledgements

This work is supported by the Israel Science Foundation (grant 748/14), Marie Curie Integration grant (CIG)- FP7-PEOPLE-20013-CIG-618763 and I-CORE Program of the Planning and Budgeting Committee and The Israel Science Foundation grant no. 41/11.

## References

1. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. **10**, (12), 1213-1218 (2013).
2. Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. **2015**, (January), 21.29.1-21.29.9 (2015).
3. Thurman, R. E., Rynes, E., et al. The accessible chromatin landscape of the human genome. *Nature*. **489** (7414), 75-82 (2012).
4. Song, L., & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*. **2010** (2) (2010).
5. Simon, J. M., Giresi, P. G., Davis, I. J., & Lieb, J. D. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc*. **7** (2), 256-267 (2012).
6. Cui, K., & Zhao, K. Genome-Wide Approaches to Determining Nucleosome Occupancy in Metazoans Using MNase-Seq. *Methods Mol Biol*. **833**, 413-419 (2012).
7. Qu, K. et al. Individuality and Variation of Personal Regulomes in Primary Human T Cells. *Cell Syst*. **1** (1), 51-61 (2015).
8. Scharer, C. D. et al. ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. *Sci Rep*. **6**, 27030 (2016).
9. Seb  Pedr  s, A. et al. The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell*. **165** (5), 1224-1237 (2016).
10. Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M., & Schmitz, R. J. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res*. **45** (6), e41 (2016).
11. Davie, K. et al. Discovery of Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by ATAC-seq and FAIRE-seq Open Chromatin Profiling. *PLOS Genet*. **11** (2), e1004994 (2015).
12. Villar, D. et al. Enhancer Evolution across 20 Mammalian Species. *Cell*. **160** (3), 554-566 (2015).
13. Lara-Astiaso, D. et al. Chromatin state dynamics during blood formation. *Science*. **345** (6199), 943-949 (2014).
14. Prescott, S. L. et al. Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimpanzee Neural Crest. *Cell*. **163** (1), 68-83 (2015).
15. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. **534** (7609), 652-657 (2016).
16. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. **523** (7561), 486-490 (2015).
17. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. **48** (10), 1193-1203 (2016).
18. Jenner, R. G. et al. The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes. *Proc Natl Acad Sci USA*. **106** (42), 17876-17881 (2009).
19. DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res*. **23** (22), 4742-4743, (1995).
20. Aird, D. et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*. **12** (2), R18 (2011).
21. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10** (3), R25 (2009).
22. Li, H., Handsaker, B., et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25** (16), 2078-2079 (2009).
23. Quinlan, A. R., & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26** (6), 841-842 (2010).
24. Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. **9** (9), R137 (2008).
25. Meyer, C. A., & Shirley Liu, X. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet*. **15** (11), 709-721 (2014).

26. He, H. H. *et al.* Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Methods*. **11** (1), 73-78 (2013).
27. Madrigal, P. On Accounting for Sequence-Specific Bias in Genome-Wide Chromatin Accessibility Experiments: Recent Advances and Contradictions. *Front Bioeng Biotechnol*. **3**, 1-4 (2015).
28. Qin, Q. *et al.* ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. *BMC Bioinformatics*. **17** (1) (2016).
29. Bao, X. *et al.* A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biol*. **16** (1), 284 (2015).
30. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science*. **337** (6099), 1190-1195 (2012).