inTAC-Seq protocol

Stanford University CIMAC

Version: 1.0

Date: 1 Aug, 2018

Introduction

ATAC seq (Assay for Transposase-Accessible Chromatin using sequencing) is a highly validated epigenomic profiling assay that quantifies chromatin accessibility for regulation of transcription using preferential transposon insertion by Tn5 transposase enzyme (Buenrostro et. al., 2013). This bulk assay is traditionally carried out on live cells, but this limits its application to only assaying populations sorted on surface markers. We developed the ability to carry out ATAC seq on fixed and permeabilized cells that have been sorted on intracellular markers. The following protocol outlines the standard ATAC seq from the original paper and our novel inTAC seq (Intracellular fluorescent activated cell sorted transposase accessible chromatin sequencing) technique (Buenrostro et. al., 2013).

This protocol is carried out with live starting material with 50000 cells for the standard ATAC protocol and 3-5 million cells for inTAC protocol. There will be 3 technical replicates for each sample and if possible 2 biological replicates as well. The technical replicates will be processed together to ensure reproducible genome alignment and calling peaks of accessibility.

Reagents

1. FoxP3 fix/perm buffer kit --- eBiosciences
2. Qiagen MiniElute purification kit
3. Nextera library prep (Tn5) kit from Illumina [FC-121-1030]
4. NEBNext PCR kit (M0541S)
5. Nuclease-free Water
6. DNA LoBind Tubes (Eppendorf #022431021)
7. AMPure® XP Beads (Beckman Coulter, Inc. #A63881)
8. ATAC permeabilization buffer [prepared fresh]
    - 10mM Tris-Cl pH7.4
    - 10mM NaCl
    - 3mM MgCl2
    - 0.1% Igepal CA-630
9. 2X Reverse crosslinking buffer [prepared fresh]
    - 50mM Tris-Cl
    - 2mM EDTA
    - 2% SDS
    - 0.4M NaCl
    - [added last] 1:100 dilution NEB P8107S proteinase K
10. IDT forward primer and IDT backward barcode primers [primer seq info in Supplement]
    - IDT forward primer + each reverse barcode primer resuspended to 100mM stock concentration
    - Each barcode primer mix with forward and 1 reverse primer diluted to 25mM working concentration for application in PCR amplification
11. Sybr-Green in DMSO

Equipments

1. Magnetic rack
2. Cold room / 4C centrifuge
3. PCR tubes
4. PCR machine

5. qPCR machine [96 well plates, seals]
6. Shaking heating block
7. Bioanalyzer
8. Hi-Seq4000

Procedure

The protocol is divided into separate sections, but it is best to carry out the full protocol sequentially in one go. There are steps which can be delayed for up to a day or 2 as indicated but there will likely be DNA degradation. To carry out standard ATAC, skip directly to lysis step onwards and skip reverse crosslinking step.

Prepping samples

1. Harvest cells from culture as appropriate for sample type
2. Transfer 3-5 million cells per sample into FACS tubes

Cell surface staining

1. Block non-specific Fc-mediated interactions: Pre-incubate the cells with 20 µL of Human Fc Receptor Binding Inhibitor Purified per 100 µL for 10-20 minutes at 2-25°C before staining [optional]
2. Wash 1X with CSM at 250G
3. Resuspend the cells in predetermined titers of the cell surface staining master mix of flow cytometry antibodies in total volume of 100 ul in CSM.
4. Incubate for at least 30 minutes at room temperature. Protect from light.
5. Wash 2X with CSM at 250G for 5min.

Fix and Permeabilization buffer prep

1. Prepare fresh Foxp3 Fixation/Permeabilization working solution by mixing 1 part of Foxp3 Fixation/Permeabilization Concentrate with 3 parts of Foxp3 Fixation/Permeabilization Diluent. *One mL of the working solution is required for each sample*, if staining in tubes.
2. Prepare a 1X working solution of Permeabilization Buffer by mixing 1 part of 10X Permeabilization Buffer with 9 parts of distilled water. *8.5 mL of the working solution is required for each sample*, if staining in tubes.

Fix/Perm and stain:

1. Add 1 mL of Foxp3 Fixation/Permeabilization working solution to each tube and pulse vortex briefly.
2. Incubate for **5 minutes** at room temperature. Protect from light.
3. Add 2 mL of 1X Permeabilization Buffer to each tube and centrifuge samples at 600 x g for 5 minutes at room temperature. Discard the supernatant. Repeat 1X.
4. Block with CSM 1ml/sample for 15min at room temperature. [Optional]
5. Wash 1X with Permeabilization Buffer at 600 x g for 5 minutes.
6. Add intracellular antibody mix to cell pellet and re-suspend thoroughly in Permeabilization Buffer un to 100ul total volume.
7. Incubate for at least 30 minutes at room temperature. Protect from light.
8. Add 2 mL of 1X Permeabilization Buffer to each tube and centrifuge samples at 600 x g for 5 minutes at room temperature. Discard the supernatant. Repeat 1X.
9. Re-suspend stained cells in 200ul CSM.

FACS run

1. Run compensations and samples on FACS, gate out populations of interest [at least 200K cells per population]
2. Keep cells on ice at all times.
3. Wash 1X in cold PBS
4. Re-count cells and take 50K cells per iATAC sample (critical: Ensure counting is exact) in cold PBS.

inTAC Lysis

1. Spin down cells at 500G for 5min and aspirate (critical: very carefully, remove as much liquid as possible (Note: Pellet at an angle, may not be visible depending on cell type)
2. Cell lysis - with freshly prepared, cold lysis buffer - by adding 50ul onto cell pellet *slowly, dropwise without pipetting*
3. Centrifuge for 10min at 500G at 4C and aspirate carefully.

inTAC Transposition

1. Make up master mix (MM) to 50ul per sample . Preparation as below: 1 sample mix X number of samples + 1:

TD rxn buffer [2X]	25ul
Tn5 TDE	2.5ul
Nuclease-free water	22.5ul

2. Briefly vortex master mix and spin down
3. Add 50ul of MM to each sample and pipette 2-3 times only (do not pipette excessively).
4. Spin down quickly and place in heating block at 37C for 30min with shaking at 300RPM.

Reverse cross-linking of fixation

1. Prepare 2X reverse crosslinking buffer fresh and add 1:100 of proteinase K last and mix well
2. Keep samples on ice and add 50ul of 2X reverse cross-linking buffer, pipette 3X to mix
3. Place on 65C shaker for 6h at 300rpm [maximum overnight if needed]

DNA purification after reverse crosslinking

1. Carry out Qiagen Mini elute PCR purification with 100ul starting volume
2. Elute in 11.5ul of EB buffer 2X (23ul total)
3. Store at 4C for a few days if needed

Initial PCR amplification

This step is to ensure high enough library concentration to check quality in qPCR in the next step.
1 sample mix in PCR tubes -

NEBNext PCR mix [2X]	25ul
F+R primers [25mM]	2.5ul

Eluted DNA -each sample	22.5ul

Vortex briefly and spin down
Load on middle of PCR machine and run with the following settings.

PCR settings
Lid at 105°C and total vol 50ul
1 cycle -
5 min at 72°C
30 sec at 98°C
5 cycles -
10 sec at 98°C
30 sec at 63°C
1 min at 72°C
Keep tubes at 12°C

qPCR testing of DNA content

qPCR testing will check for optimal library concentration and amplification profile.
The following is a list of constituents for each qPCR sample.

F+R primer [25mM]	0.75ul
Amplified DNA	5ul
SyBr Green 100X	0.09ul
NEBNext MM	5ul
Nuclease-free water	4.16ul

1. Make a master mix of NEBNext, SyBr and water of sample number +1. (Note: Add Sybr quickly and protect from light)
2. Load samples in the middle of a 96-well plate (pipette up and down to mix).
   Note: Seal adhesive tape on plate well, smooth over with flat surface and do not touch with gloves

qPCR settings
Lid at 105°C and total vol 50ul
1 cycle -
30 sec at 98°C
20 cycles -
10 sec at 98°C
30 sec at 63°C (collect SYBR signal)
1 min at 72C
Keep tubes at 12C

Determination of additional cycles of amplification: Visualize data on linear scale and note number of cycles needed to achieve 1/3 of max intensity detected (signal should plateau). Final PCR amplification should be carried out additionally for calculated number of cycles in next step.

Final PCR amplification

Run original PCR tubes with the following program-
Lid at 105°C and total vol 45ul
1 cycle -
 30 sec at 98C
Added cycles -
 10 sec at 98°C
 30 sec at 63°C [collect data]
 1 min at 72°C
Keep tubes at 12°C
Store product at 4°C for a few days if necessary.

DNA purification

1. Make fresh 70% ethanol and keep on ice
2. Mix AMPure slurry well to re-suspend beads evenly and pipette slowly to keep volumes added consistent
3. *Off magnet* - Add 1.8 times PCR product volume [45ul] of bead slurry to DNA (reduce to 1:1 ratio to exclude more primer dimers).
4. Vortex briefly
5. Spin down and let it rest at RTP for 5min
6. *On magnet* - Pipette out the clear solution slowly, leaving 5ul and not taking up any beads
7. Add 200ul of 70% ethanol without pipetting, rest for 30s.
8. Pipette out as much as possible
9. Add 200ul of 70% ethanol without pipetting, rest for 30sec
10. Pipette out as much as possible
11. Leave beads to dry for 2min [Note: Do not over dry and crack beads]
12. *Off magnet* - Add 20ul of EB and pipette mix well [around 20 times] to get a homogenous mix without bubbles
13. Let it stand in RTP for 2min
14. *On magnet* - Take out ~18ul after full separation into DNA low bind eppendorfs slowly without taking up beads
15. Dilute DNA as appropriate (recommended dilution is 1:20).
16. Keep diluted DNA at 4C and store the rest at 20C (minimize freeze/thawing cycles of DNA).

DNA Library QC

1. Aliquot 2ul of each diluted sample to check the fragment length distribution on the bioanalyzer.
2. There should be the expected fragment length distribution at the nucleosome spacings (Figure 3).
3. Measure DNA concentration on Qubit.
4. Additional QC can be carried out by sequencing on a MiSeq to check the quality of the library before high depth HiSeq sequencing.

Next Gen Sequencing

1. Pool samples (up to 24) for a final concentration of around 10nM with 20ul volume.
2. Pooled sample is to be run on 1 lane on HiSeq4000 for paired end sequencing with 75bp read length (2x75 run) with index sequence matching to each sample for subsequent de-barcoding.
   Note: Index sequence is the reverse complement of first 8bp of barcode primer.
3. Stanford Functional Genomics Core [core services] carries out NGS on the HiSeq4000 [each barcoded sample is 1 lane] and they de-barcode the output before transferring the raw fastq files to Bendall lab for downstream processing.

Data processing and analytics

The HiSeq4000 outputs NGS data in fasta/fastq format. The Kundaje ENCODE processing pipeline is used to process the raw sequencing data to a genome aligned and filtered BAM file and a peak-called BED file (Figure 1, Left). The called peaks of accessibility are then annotated and compared between samples (Figure 1, Right). Transcription factor foot printing is then carried out to predict active gene expression pathways (Figure 1, Right).

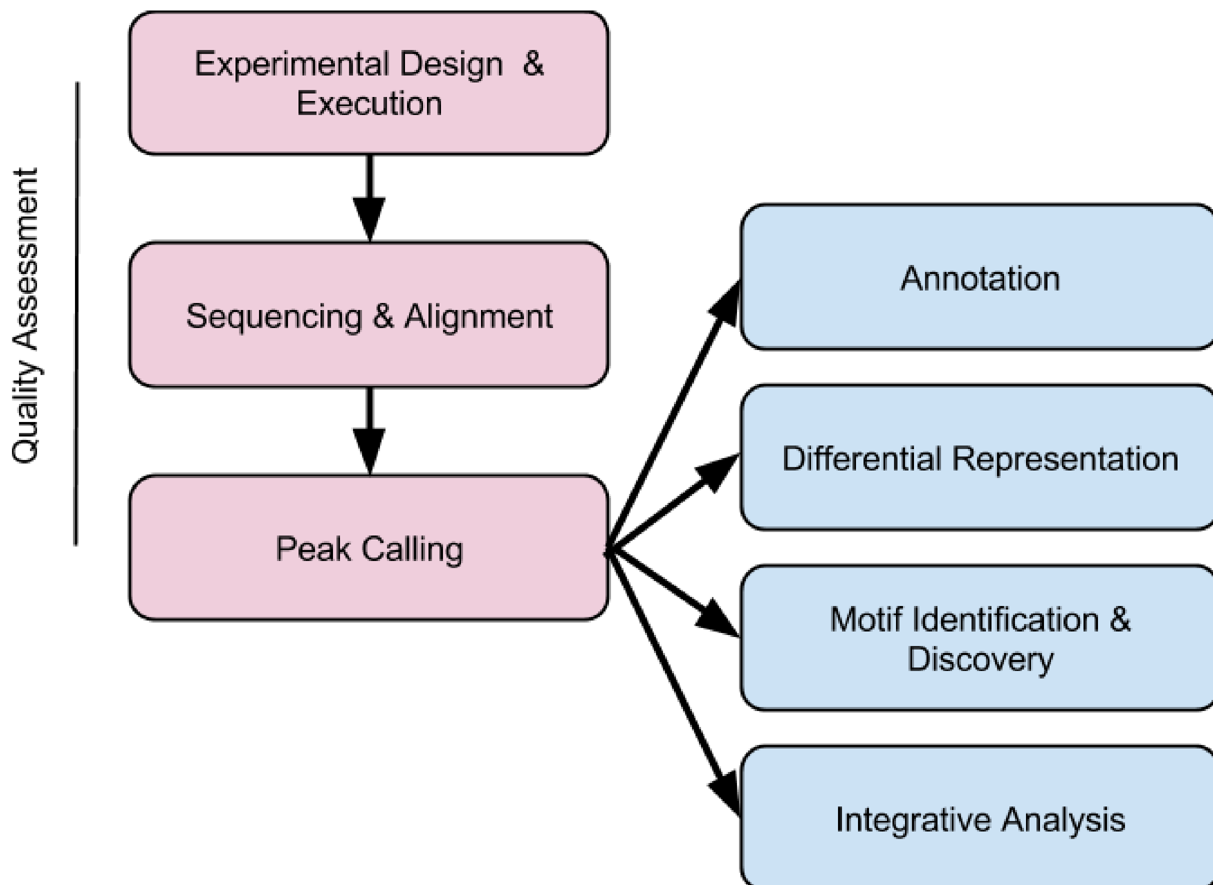

Figure 1. Overview of data processing (Left in red) and data analytics (Right in blue) pipeline.

Our data processing pipeline is the ENCODE standard Kundaje pipe with numerous quality assessment checks done on the genome aligned BAM files (Figure 1). The data processing pipeline starts with the paired end reads getting cleaned by trimming to remove the sequencing adapters using the cutadapt package (Martin, 2011). The reads are then aligned to the hg19 or hg38 human genome assembly using Bowtie2 algorithm (Langmead & Salzberg, 2012)(Figure 2). Unmapped reads, multi-mapped reads, PCR duplicates and reads mapping to mitochondrial DNA are removed using SAMtools (H. Li et al., 2009)(Figure 2). The genome aligned and filtered reads (BAM format) are fed into the MACS2 algorithm to call peaks of accessibility. The Irreproducible Discovery Rate (IDR) is used to filter significant hits across replicates (Q. Li, et. al., 2011).

Bowtie alignment log

```
31846218 reads; of these:
  31846218 (100.00%) were paired; of these:
    849939 (2.67%) aligned concordantly 0 times
    20973245 (65.86%) aligned concordantly exactly 1 time
    10023034 (31.47%) aligned concordantly >1 times
    ----
    849939 pairs aligned concordantly 0 times; of these:
      307838 (36.22%) aligned discordantly 1 time
    ----
    542101 pairs aligned 0 times concordantly or discordantly; of these:
      1084202 mates make up the pairs; of these:
        757501 (69.87%) aligned 0 times
        126049 (11.63%) aligned exactly 1 time
        200652 (18.51%) aligned >1 times
98.81% overall alignment rate
```

Mapping quality > q30 (out of total)	54,344,261	0.853
Duplicates (after filtering)	2,231,459.0	0.083
Mitochondrial reads (out of total)	4,767,840	0.076
Duplicates that are mitochondrial (out of all dups)	3,330,356	0.531
Final reads (after all filters)	49,102,386	0.771

Figure 2. Quality assessment 1: Alignment and read count log after alignment with bowtie2 and filtering with Kundaje ENCODE pipeline (ATAC data of CAR T therapy product).

The aligned reads are checked for their distribution around the transcription start site (a known site of high chromatin accessibility) (Figure 3). The read density also follows a known distribution of fragment lengths by nucleosome spacing (Figure 3). Further QC steps were carried out using the ATACseqQC R package to obtain the NFR score distribution and the read coverage of promoter vs transcript read coverage (Figure 4). We see an expectedly higher proportion of reads at the promoter. The final QC step is the read distribution at putative CTCF binding sites on the genome (Figure 4). CTCF is a well-known higher chromatin structural regulator by regulating topological association domain boundaries. We expect to see transposon insertion flanking CTCF sites and comparatively negligible insertion at the CTCF binding area. We see this

characteristic pattern from our CAR T therapy product ATAC data, thereby signifying accurate higher chromatin structure capture (Figure 4).

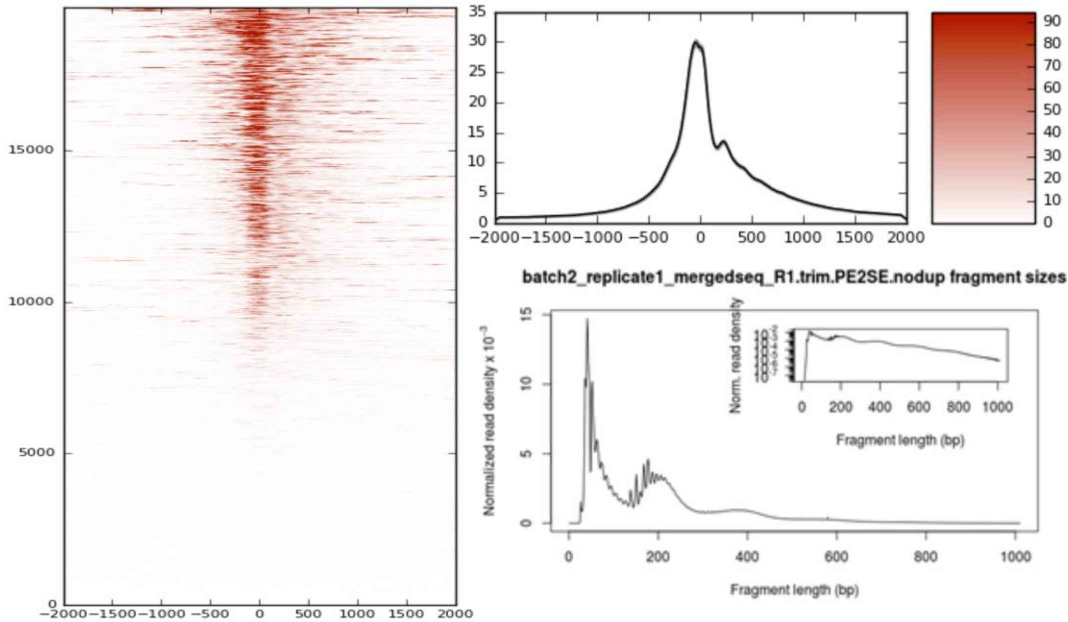

Figure 3. Quality assessment 2: Read density at transcription start sites have a characteristic open distribution in ATAC seq (Left and Right top). The fragment lengths are distributed by nucleosome spacing and the distribution is a secondary check for a successful transposition experiment.

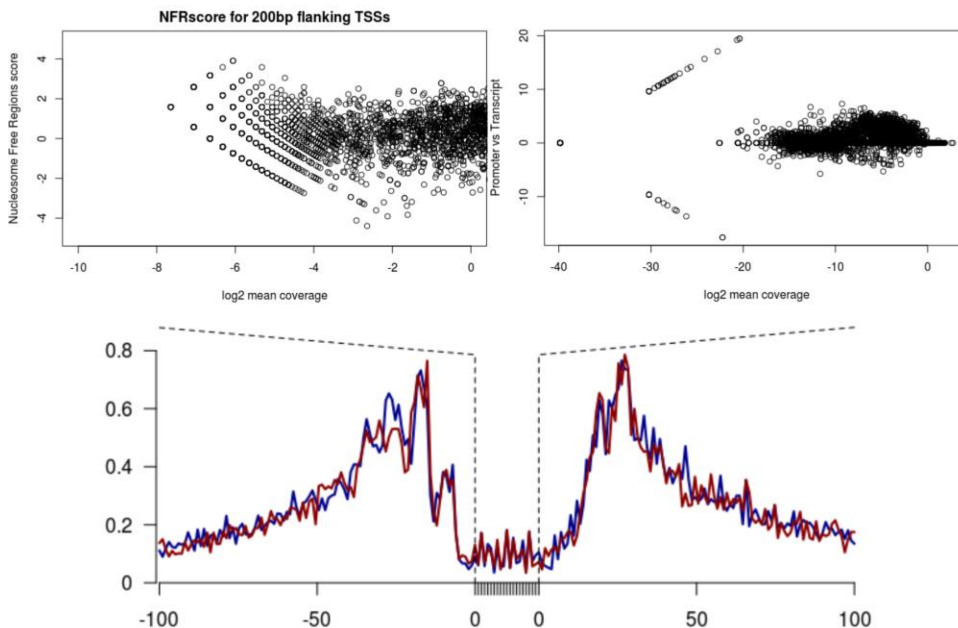

Figure 4. Quality assessment 3: NFR score distribution as well as promoter vs transcript read coverage is of CAR T therapy product is shown (top plot). CTCF motif footprint depicted by read density of positive strand (red) and negative strand (blue) at putative CTCF binding sites as curated by MotifDB (bottom plot).

Data analytics

Upon passing all the QC steps, all the samples from an experiment are combined into 1 matrix with peaks as features and the corresponding sequencing depth normalized read counts per sample for each peak. In order to compare between experiments, the data must be processed similarly and normalized together. The list of peaks is the union of significant peaks across samples and they are generated by MACS2 after combining all the sample BAM files and running it with a significance cutoff of q score < 0.01. These peaks are annotated using the GREAT algorithm (Mclean et al., 2010) (Figure 1). The peaks are annotated with significantly associated gene and genomic features such as regulatory regions and then proteins and features in similar pathways are associated using CHIPseeker package (Yu, Wang, & He, 2015). The matrix is then fed into DESeq2 to identify differentially accessible peaks between samples in an experiment (Love et. al., 2014). Further motif analysis is carried out using the Brockman algorithm with curated position weight matrices of transcription factors and chromatin-binding proteins from the JASPAR database (de Boer & Regev, 2018). The transcription factor occupancy variation across samples can shed light on differentially active gene expression programs. In conclusion, we have outlined our novel experimental technique (inTAC), as well as the processing and analytics pipeline for the data.

References

de Boer, C. G., & Regev, A. (2018). BROCKMAN: deciphering variance in epigenomic regulators by k-mer factorization. *BMC Bioinformatics*, *19*(1), 253. https://doi.org/10.1186/s12859-018-2255-6

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–9. https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–9. https://doi.org/10.1093/bioinformatics/btp352

Li, Q., Brown, J. B., Huang, H., & Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*, *5*(3), 1752–1779. https://doi.org/10.1214/11-AOAS466

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mclean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., … Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, *28*. https://doi.org/10.1038/nbt.1630

Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Supplement

ATAC forward and reverse primer seq [to order from IDT]

FORWARD primer:
Ad1_noMX AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG

REVERSE primers [24 unique barcodes]:
Ad2.1_TAAGGCGA CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT