Single-cell Epigenomics SinCellTE 2022

Andres Quintero
Carl Herrmann

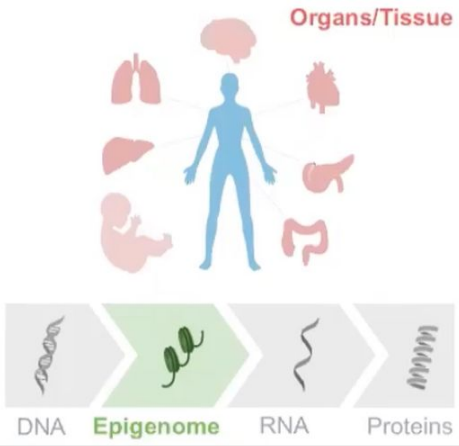
Health Data Science Unit
Heidelberg University
13/01/2022
Roscoff, France

Content

- Why do we study the epigenome?
- Epigenetic layers of gene regulation
- Single-cell epigenome profiling
- Single-cell ATAC-seq library preparation
- Single-cell ATAC-seq analysis
- Perspectives

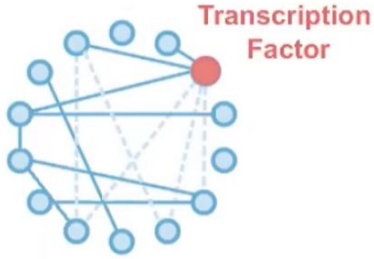
Why do we study the epigenome?

Elusive Cellular Variability



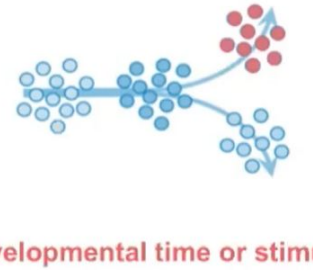
Understand **cellular diversity** despite nearly identical DNA sequences

Complex Gene Regulatory Networks



Identify **drivers** of differential gene expression across cells, tissues, disease

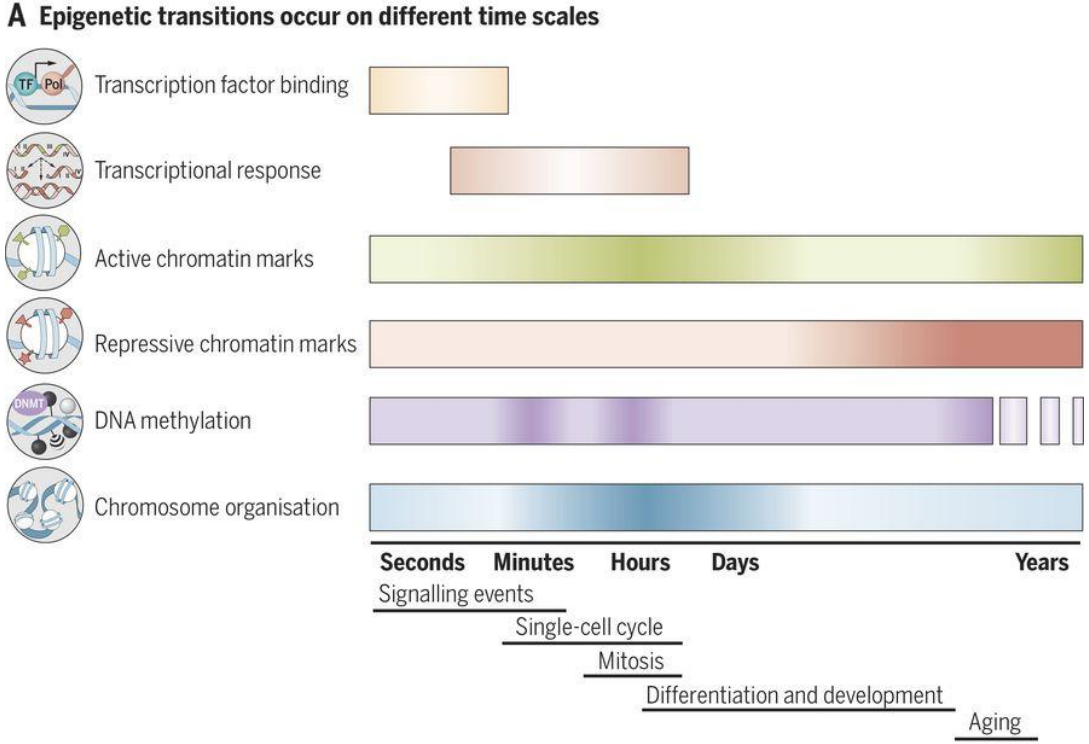
Multifaceted Cellular Memory & Potential



Characterize developmental **trajectory** or cellular **responses** to stimuli or drugs

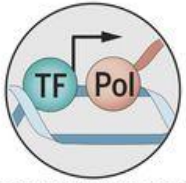
https://www.10xgenomics.com/products/single-cell-atac

Why do we study the epigenome?



Kelsey et al. 2017. Nature.
DOI: 10.1126/science.aan6826

Epigenetic layers of gene regulation



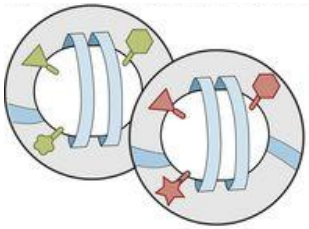
Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility

DNA modifications

C 5mC

5hmC / 5fC / 5caC



Histone modifications

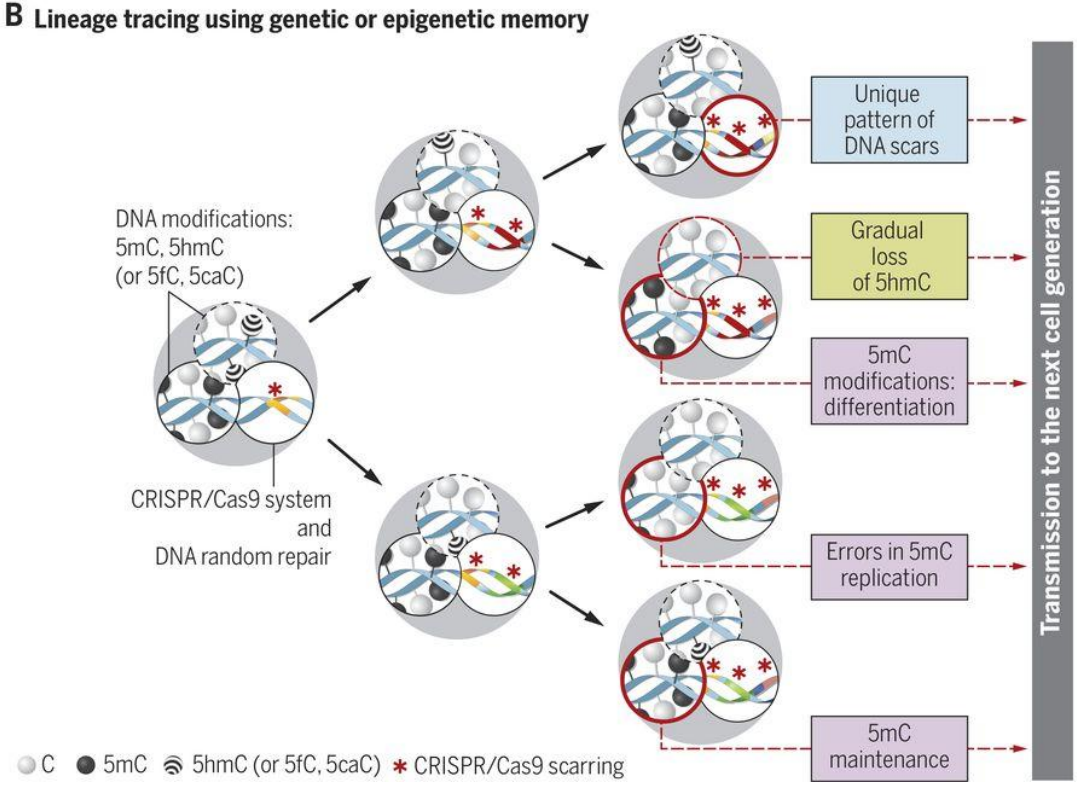
Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27m3 in red)

Chromosome organization

Higher-order chromatin organization into LADs and TADs

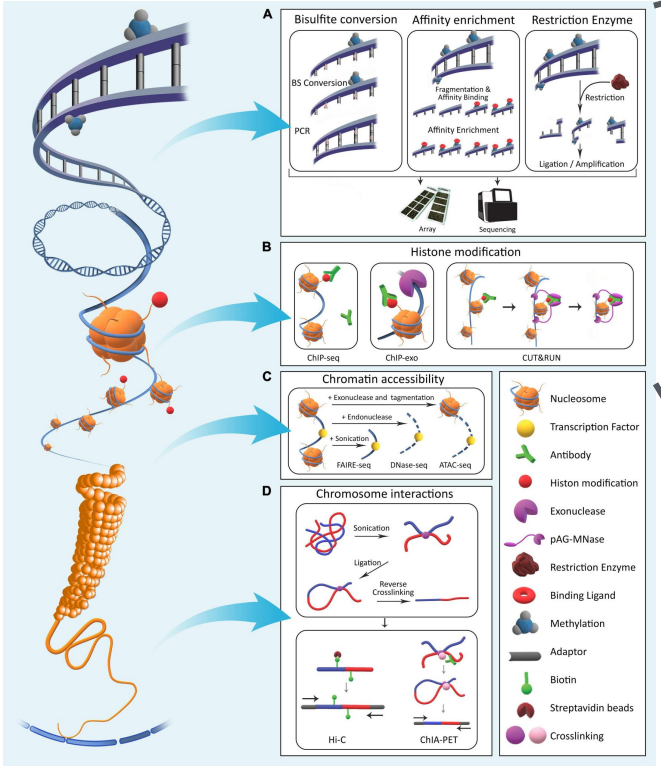
Kelsey et al. 2017. Nature.
DOI: 10.1126/science.aan6826

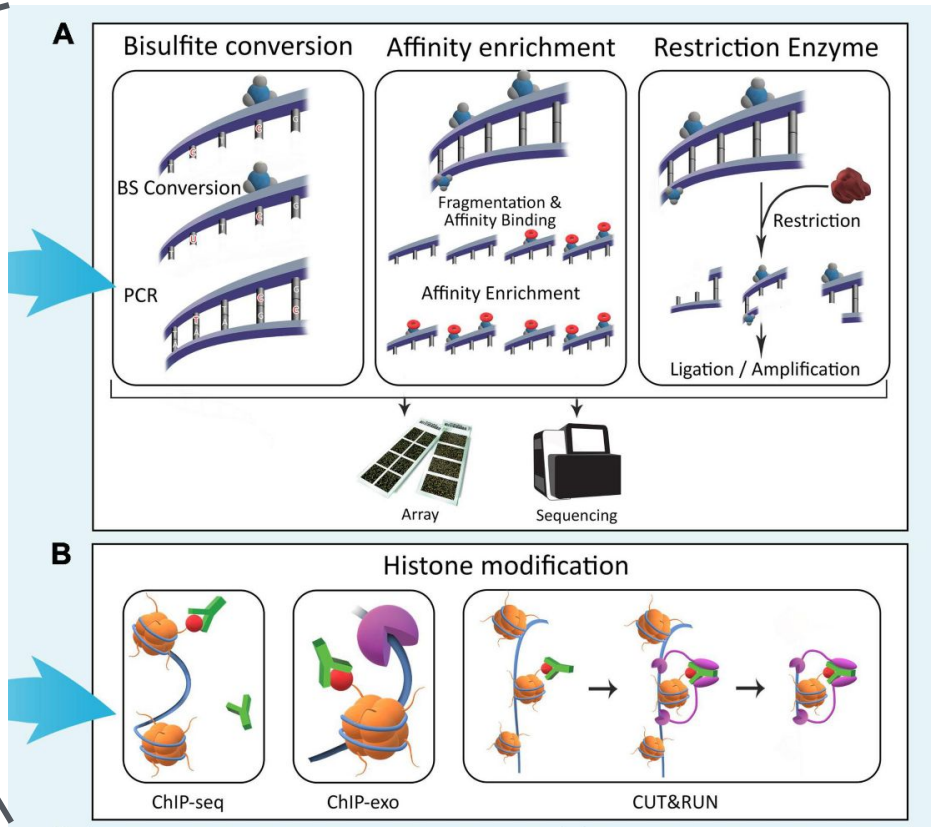
Why do we study the epigenome?



Kelsey et al. 2017. Nature.
DOI: 10.1126/science.aan6826

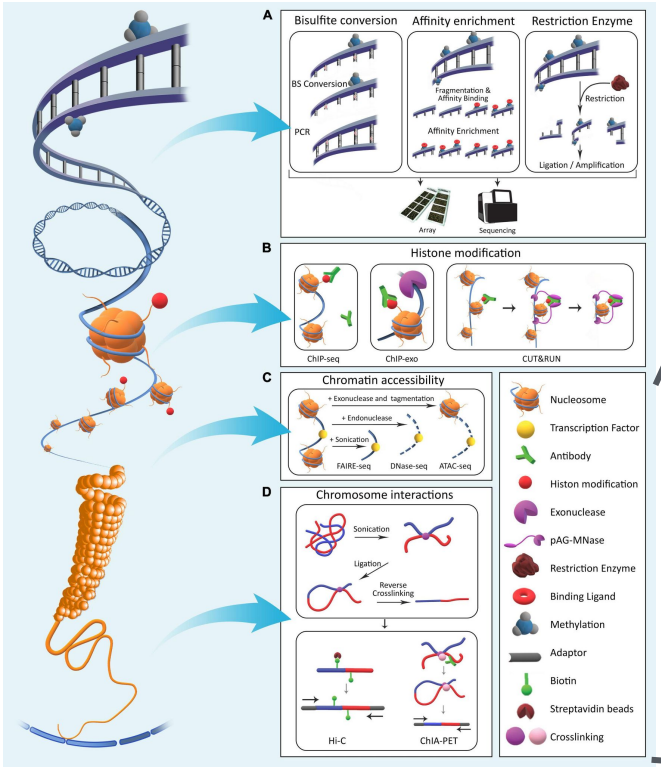
Epigenetic layers of gene regulation

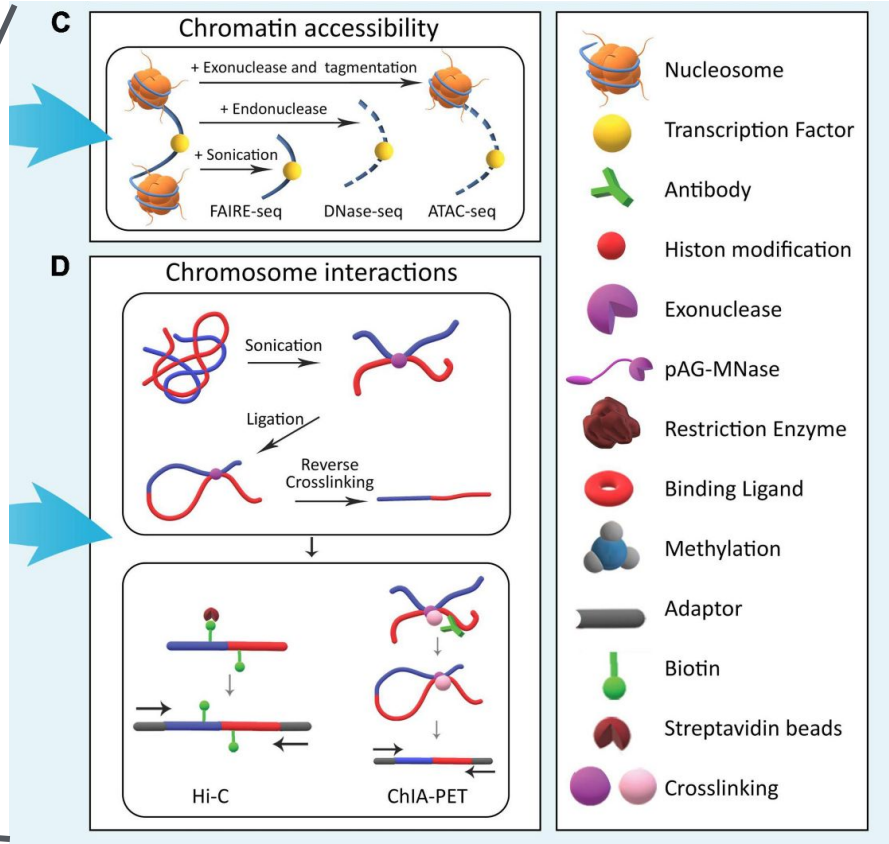




Mehrmohamadi et al. 2021. Front. Cell Dev. Biol.
DOI: 10.3389/fcell.2021.714687

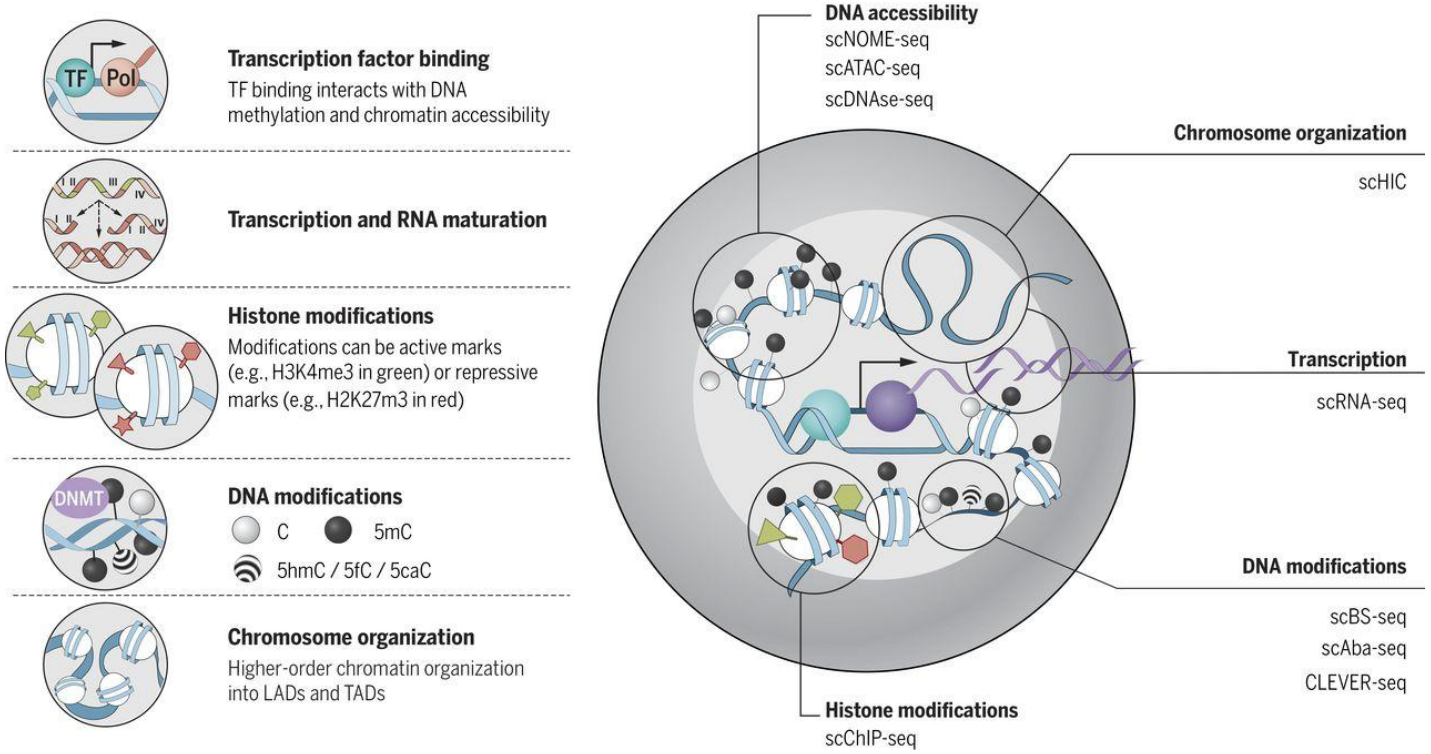
Epigenetic layers of gene regulation





Mehrmohamadi et al. 2021. Front. Cell Dev. Biol.
DOI: 10.3389/fcell.2021.714687

Single-cell epigenome profiling



Kelsey et al. 2017. Nature.
DOI: 10.1126/science.aan6826

Single cell DNA accessibility

Link to Publisher's site

eLife. 2017; 6: e23203.
Published online 2017 Jun 27.
doi: 10.7554/eLife.23203

PMCID: PMC5487215
PMID: 28653622

Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells

Sebastian Pott*

Published: 17 June 2015

Single-cell chromatin accessibility reveals principles of regulatory variation

Jason D. Buenrostro, Beijing Wu, Ulrike M. Litzenburger, Dave Ruff, Michael L. Gonzales, Michael P. Snyder, Howard Y. Chang & William J. Greenleaf

Nature **523**, 486–490 (2015) | Cite this article

Published: 25 November 2015

Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples

Wenfei Jin, Qingsong Tang, Mimi Wan, Kairong Cui, Yi Zhang, Gang Ren, Bing Ni, Jeffrey Sklar, Teresa M. Przytycka, Richard Childs, David Levens & Keji Zhao

Nature **528**, 142–146 (2015) | Cite this article

BioQuant MODEL base of LIFE

health data science unit

HEIDELBERG FACULTY OF MEDICINE

Single cell Histone modifications and Chromosome organization

Published: 06 July 2017

Cell-cycle dynamics of chromosomal organization at single-cell resolution

Takashi Nagano, Yaniv Lubling, Csilla Várnai ✉, Carmel Dudley, Wing Leung, Yael Baran, Netta Mendelson Cohen, Steven Wingett, Peter Fraser ✉ & Amos Tanay ✉

Nature **547**, 61–67 (2017) | Cite this article

Published: 12 October 2015

Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state

Assaf Rotem, Oren Ram, Noam Shoresh, Ralph A Sperling, Alon Goren, David A Weitz ✉ & Bradley E Bernstein ✉

Nature Biotechnology **33**, 1165–1172 (2015) | Cite this article

Single cell DNA accessibility

Published: 20 July 2014

Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity

Sébastien A Smallwood, Heather J Lee, Christof Angermueller, Felix Krueger, Heba Saadeh, Julian Peat, Simon R Andrews, Oliver Stegle, Wolf Reik & Gavin Kelsey

Nature Methods **11**, 817–820 (2014) | Cite this article

34k Accesses | **547** Citations | **139** Altmetric | Metrics

> Cell Stem Cell. 2017 May 4;20(5):720-731.e5. doi: 10.1016/j.stem.2017.02.013. Epub 2017 Mar 23.

Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution

Chenxu Zhu [1], Yun Gao [2], Hongshan Guo [2], Bo Xia [1], Jinghui Song [1], Xinglong Wu [3], Hu Zeng [1], Kehkooi Kee [4], Fuchou Tang [5], Chengqi Yi [6]

Affiliations + expand

PMID: 28343982 DOI: 10.1016/j.stem.2017.02.013

Free article

Published: 27 June 2016

Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction

Dylan Mooijman, Siddharth S Dey, Jean-Charles Boisset, Nicola Crosetto & Alexander van Oudenaarden

Nature Biotechnology **34**, 852–856 (2016) | Cite this article

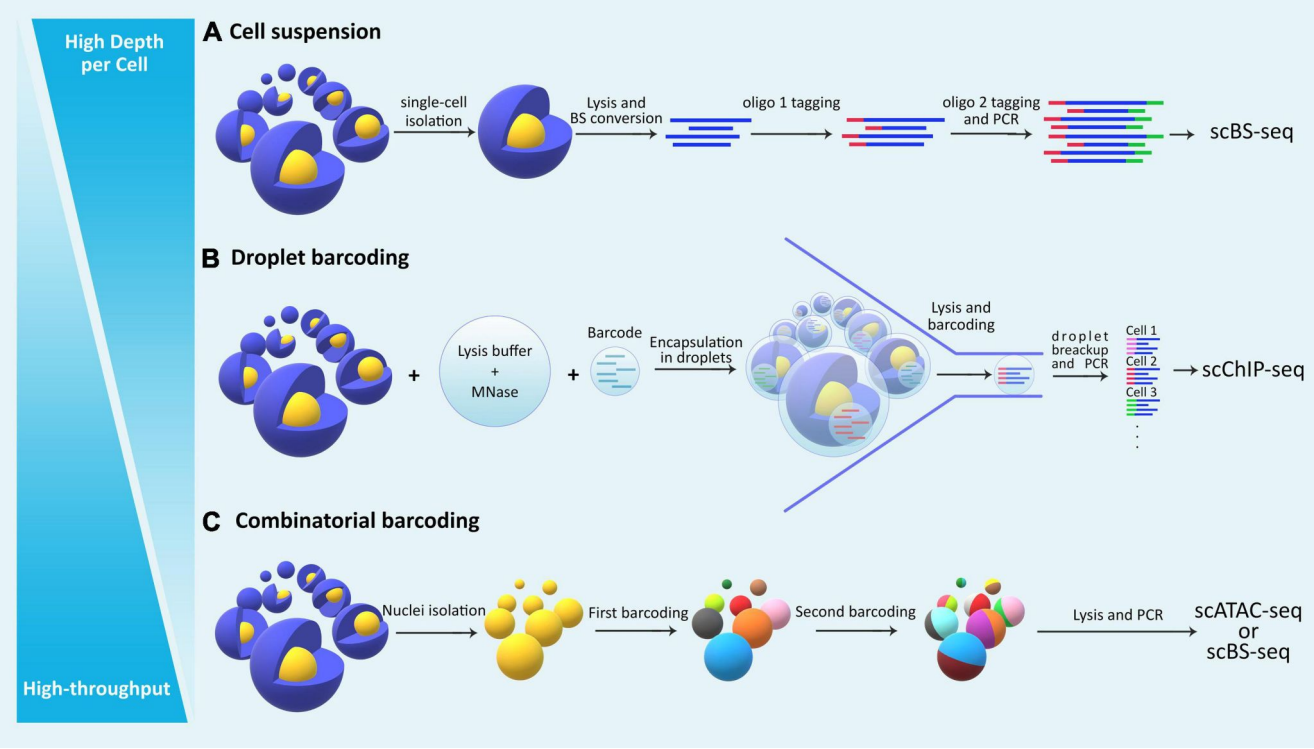
88 Citations | **33** Altmetric | Metrics

BioQuant MODEL base of LIFE

healthdata scienceunit

HEIDELBERG FACULTY OF MEDICINE

Single-cell epigenome profiling

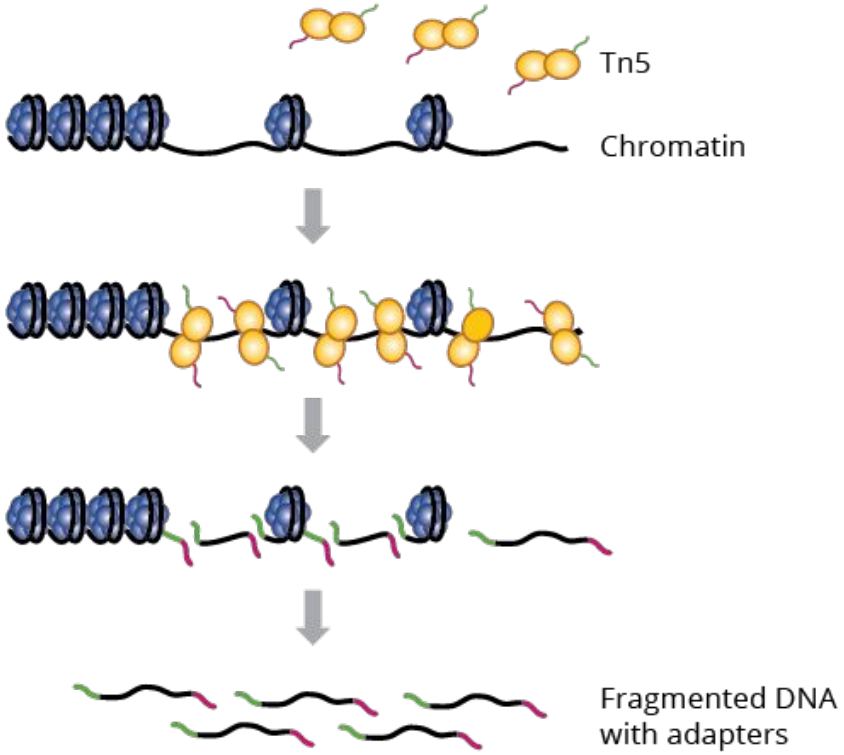


Mehrmohamadi et al. 2021. Front. Cell Dev. Biol.
DOI: 10.3389/fcell.2021.714687

Single-cell ATAC-seq library preparation

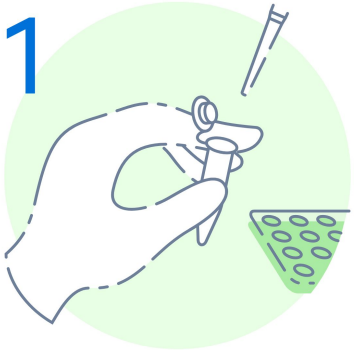
Assay for transposase-accessible chromatin sequencing (ATAC-Seq) employs a hyperactive form of Tn5 transposase to identify regions of open chromatin, which are important for global epigenetic control of gene expression.

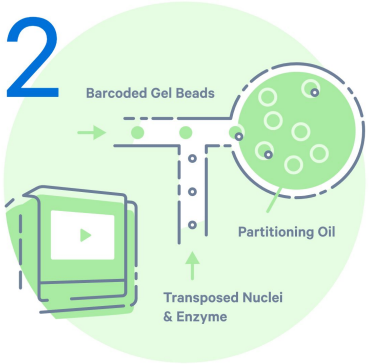
Tn5 simultaneously cleaves and adds adapters to nucleosome-free regions of DNA, priming them for sequencing.

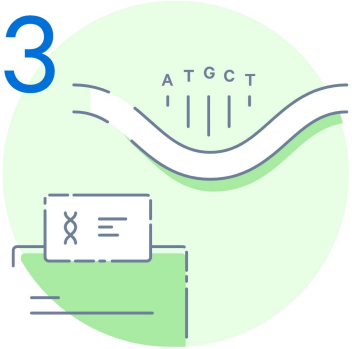


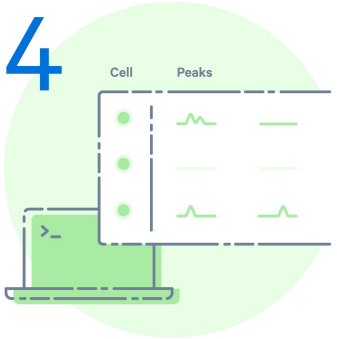
https://www.genewiz.com/en-GB/Public/Services/Next-Generation-Sequencing/Epigenomics/ATAC-Seq/

Single-cell ATAC-seq library preparation









Prepare sample:

Start with a nuclei suspension isolated from cell culture, primary cells, or fresh or frozen tissue.

Construct Library:

Construct a barcoded library. Each cell is encapsulated in a Gel Bead.

Sequence:

The resulting Barcoded single cell ATAC-seq library is compatible with standard NGS short-read sequencing.

Analysis:

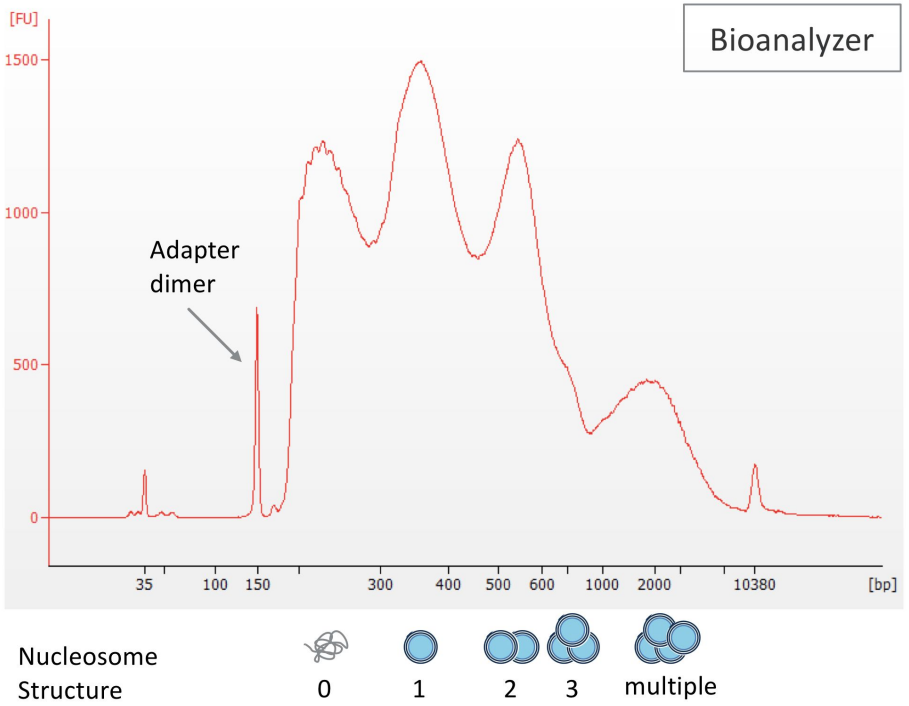
Identify clusters of cells with similar profiles and calls peaks.

https://www.10xgenomics.com/products/single-cell-atac#workflow

Single-cell ATAC-seq library preparation

ATAC trace:

The peaks of the final trace are indicative of the periodicity of the chromatin structure and show nucleosome-free, mononucleosome, dinucleosome and multinucleosome fragments.



https://kb.10xgenomics.com/hc/en-us/categories/360001072491-Single-Cell-ATAC

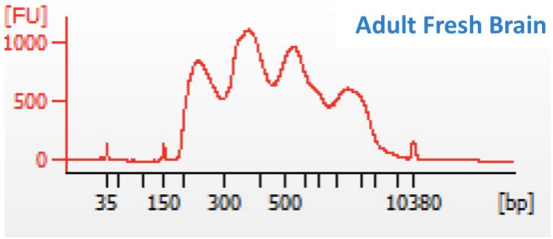
Single-cell ATAC-seq library preparation

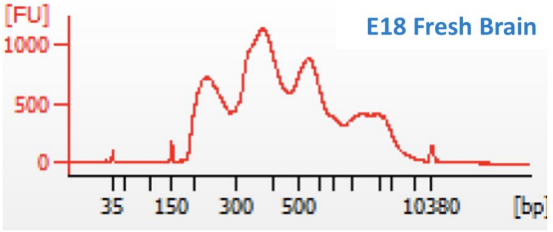
ATAC trace variations:

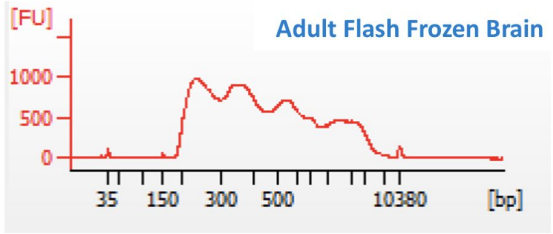
Some variation in the ATAC traces may be observed due to sample preservation or biological differences.

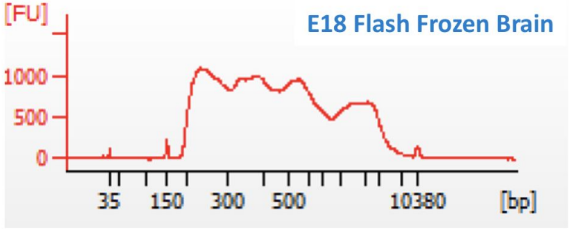
The periodicity of the chromatin is still observed, although the intensity of peaks changes.

Effects Due to Cryopreservation









https://kb.10xgenomics.com/hc/en-us/categories/360001072491-Single-Cell-ATAC

Single-cell ATAC-seq library preparation pitfalls

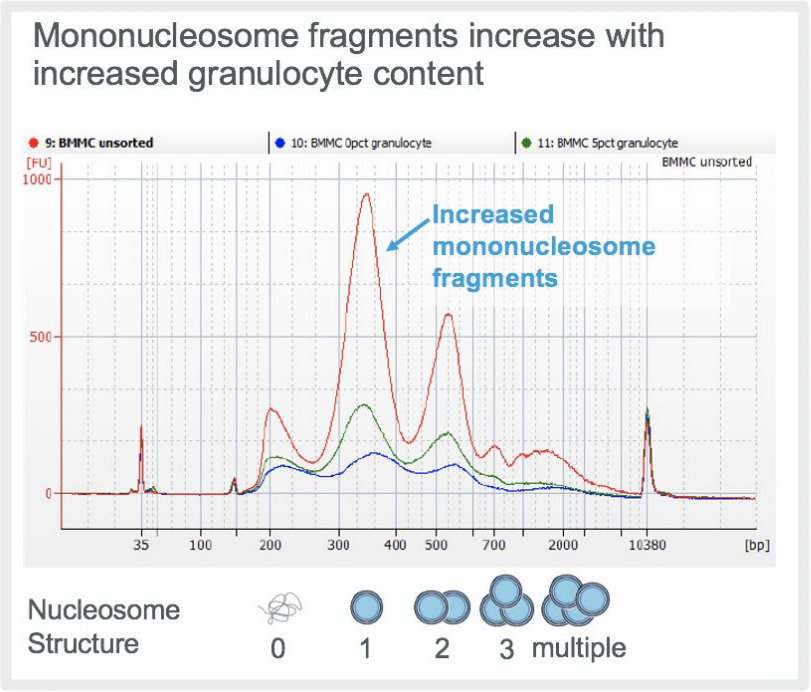
Presence of neutrophils:

Activated Neutrophils can participate in NETosis.

Forming complexes of proteins and chromatin that trap pathogens (neutrophil extracellular traps (NETs)).

During NETosis, neutrophils experience chromatin swelling.

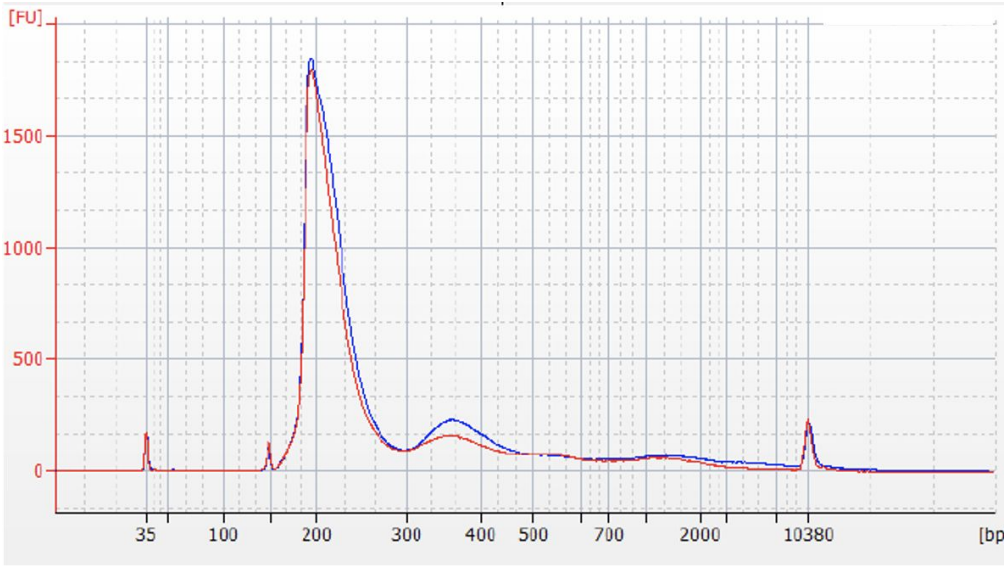
NETosis results in lysed neutrophil cells with lots of open chromatin.



Single-cell ATAC-seq library preparation pitfalls

Loss of chromatin structure:

When the sample has lost all chromatin structure and is completely open, nucleosome-free fragments will make up the majority of the recovered products.

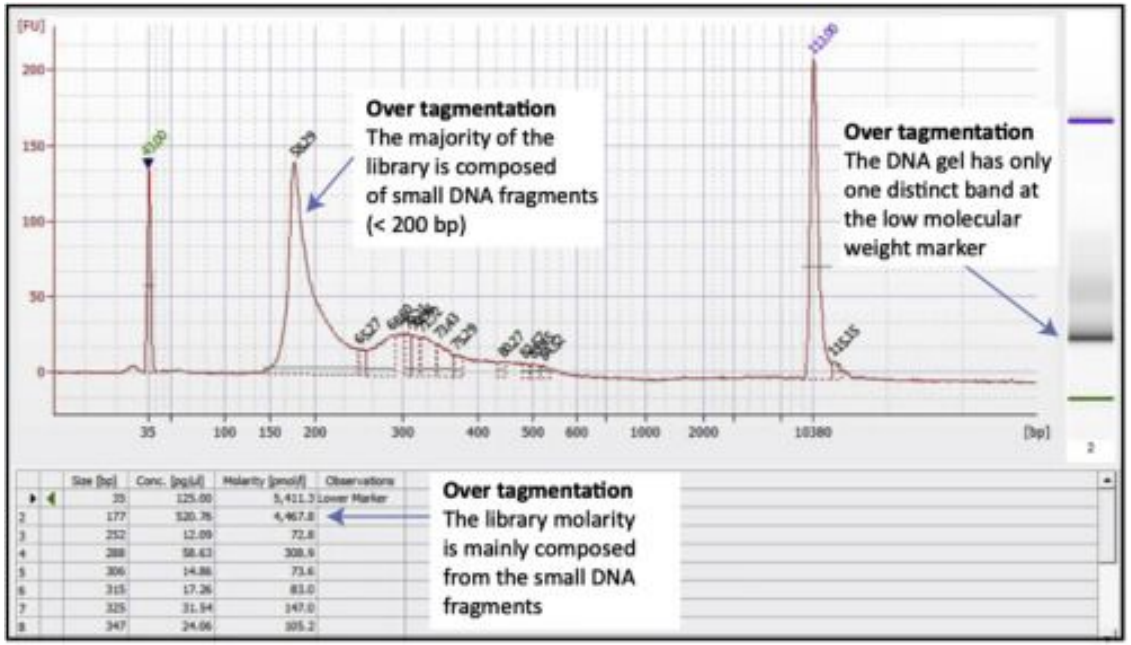


https://kb.10xgenomics.com/hc/en-us/categories/360001072491-Single-Cell-ATAC

Single-cell ATAC-seq library preparation pitfalls

Over tagmentation:

Most problems arise from not determining the appropriate number of cells to use in the tagmentation reaction. A low cell to transposase enzyme ratio can result in over-tagmentation, which produces a library mainly represented by small DNA fragments

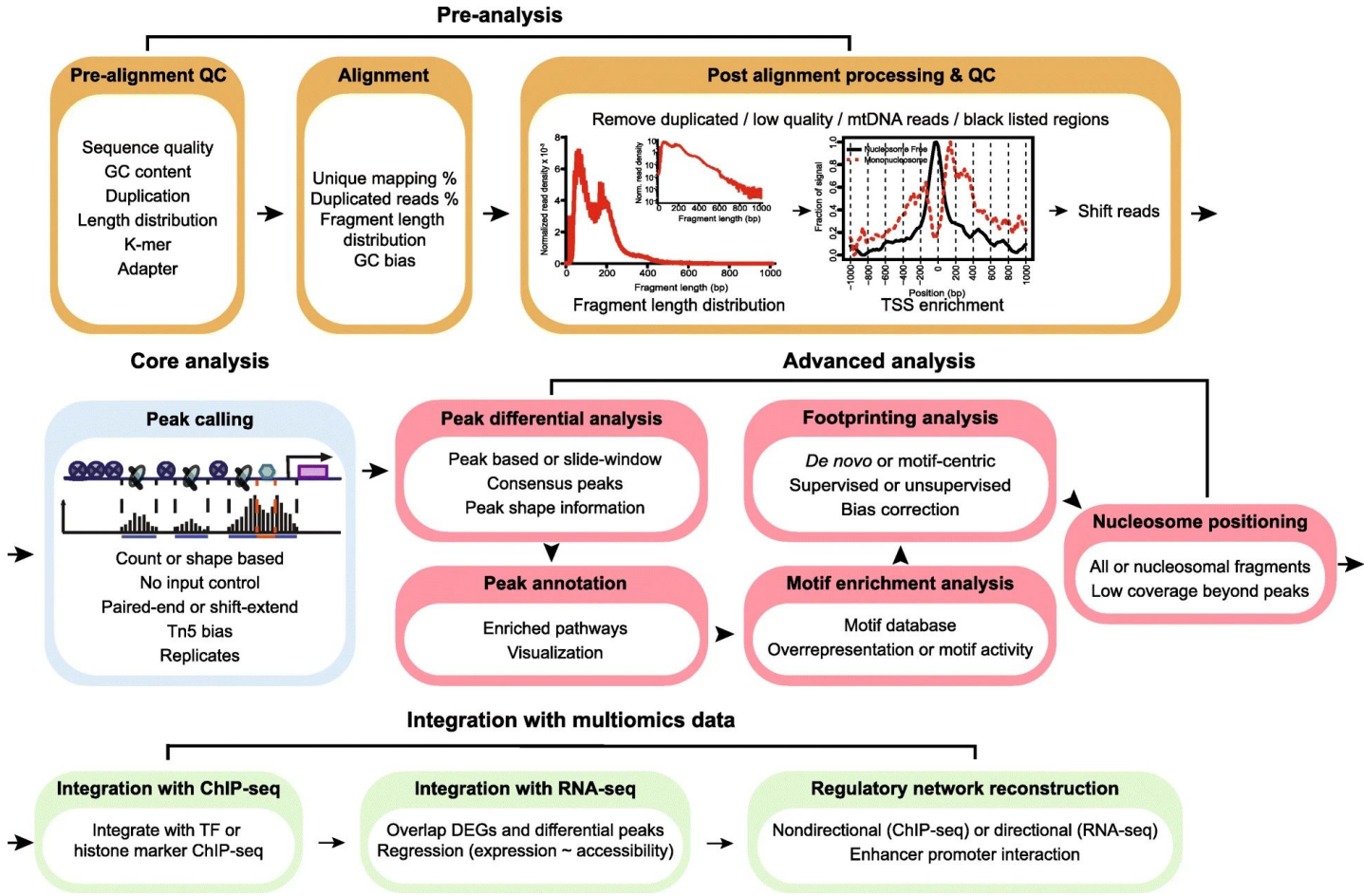


Brunton et al. 2020. STAR Protoc.
DOI: 10.1016/j.xpro.2020.100079

Single-cell ATAC-seq analysis

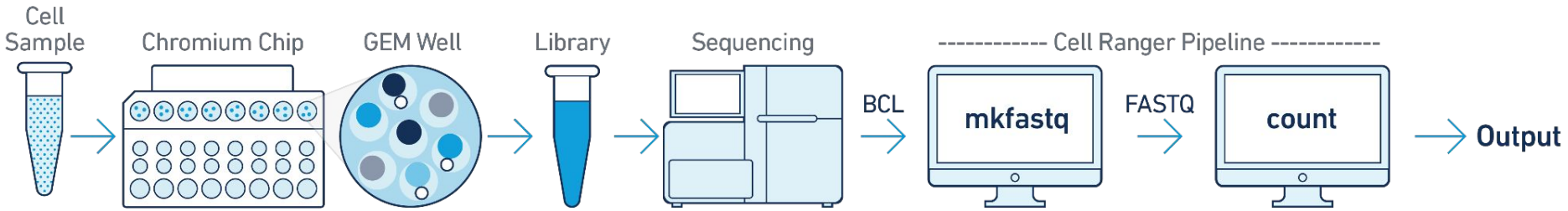
https://www.10xgenomics.com/products/single-cell-atac#workflow

ATAC-seq analysis steps



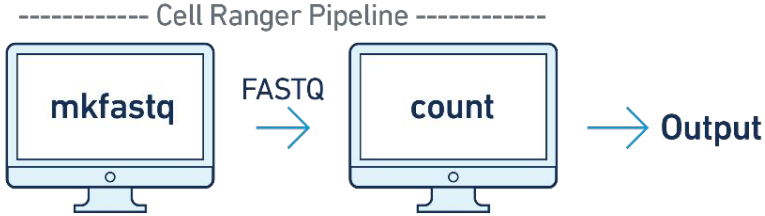
Yan et al 2020. Genome Biol.
DOI: 10.1186/s13059-020-1929-3

Cell Ranger ATAC pipeline



https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac

Cell Ranger ATAC pipeline



Demultiplexes raw base call (BCL) files generated by Illumina® sequencers into FASTQ files.

It takes FASTQ files from mkfastq and performs ATAC analysis, including:

- Read filtering and alignment
- Barcode counting
- Identification of transposase cut sites
- Detection of accessible chromatin peaks
- Cell calling
- Count matrix generation for peaks and transcription factors
- Dimensionality reduction
- Cell clustering
- Cluster differential accessibility

https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac

Cell Ranger ATAC fragments

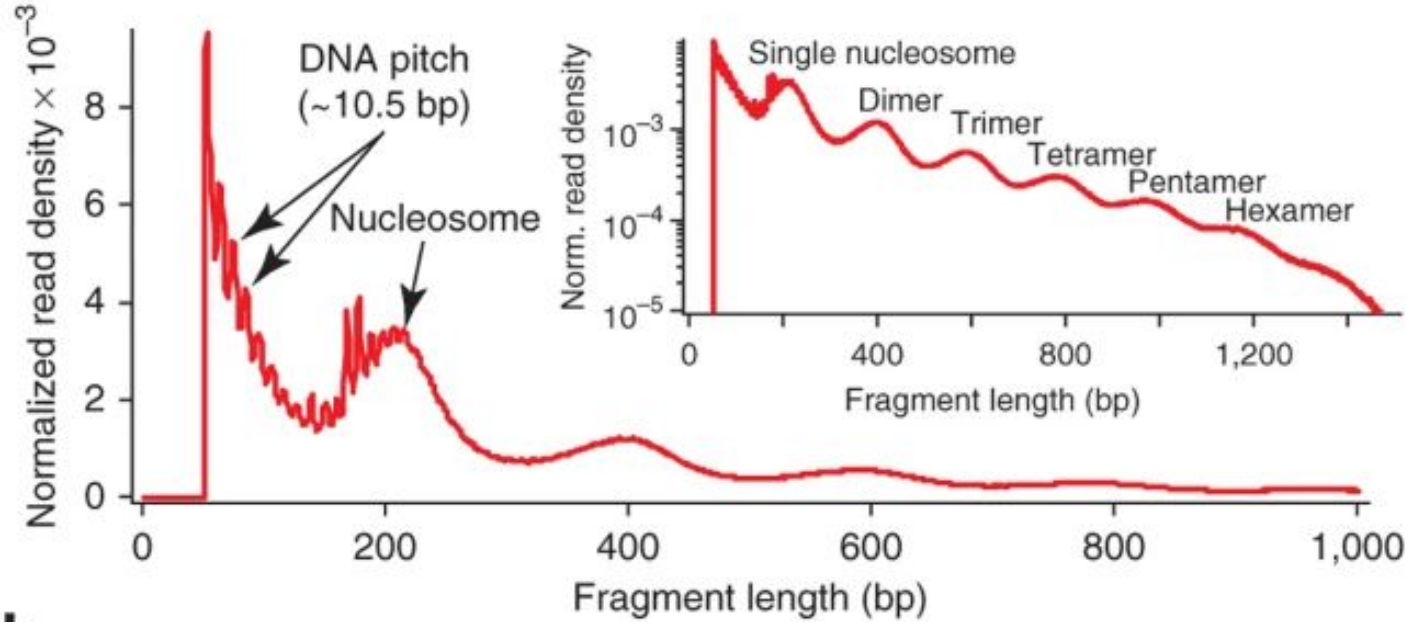
- BED-like tabular file
- Each line represents a unique ATAC-seq fragment.
- The transposase cuts the two DNA strands with a 9bp overhang, fragments positions are adjusted:
  - Moved forward by 4bp from a left-most alignment position
  - Moved backward 5bp from the right-most alignment position

Column Number	Name	Description
1	chrom	Reference genome chromosome of fragment
2	chromStart	Adjusted start position of fragment on chromosome.
3	chromEnd	Adjusted end position of fragment on chromosome. The end position is exclusive, so represents the position immediately following the fragment interval.
4	barcode	The 10x cell barcode of this fragment. This corresponds to the `CB` tag attached to the corresponding BAM file records for this fragment.
5	readSupport	The total number of read pairs associated with this fragment. This includes the read pair marked unique and all duplicate read pairs.

```
chr1    10091   10320   GACCTGATCAGCTAAC-1    1
chr1    10091   10340   ATAACGACACACCAAC-1    4
chr1    10095   10350   TTTAGCTTCCGCAACA-1    1
chr1    10096   10279   CATCATAAGGATCACT-1    3
chr1    10096   10308   AGTAGCTTCGGGATTT-1    2
chr1    10097   10334   TAGCTAGGTGTTGCTT-1    3
chr1    10097   10339   GCTGATCCAACCTAAT-1    3
chr1    10101   10346   CGCAATGTCGTTATCT-1    3
```

https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac

scATAC-seq QC - Fragment size



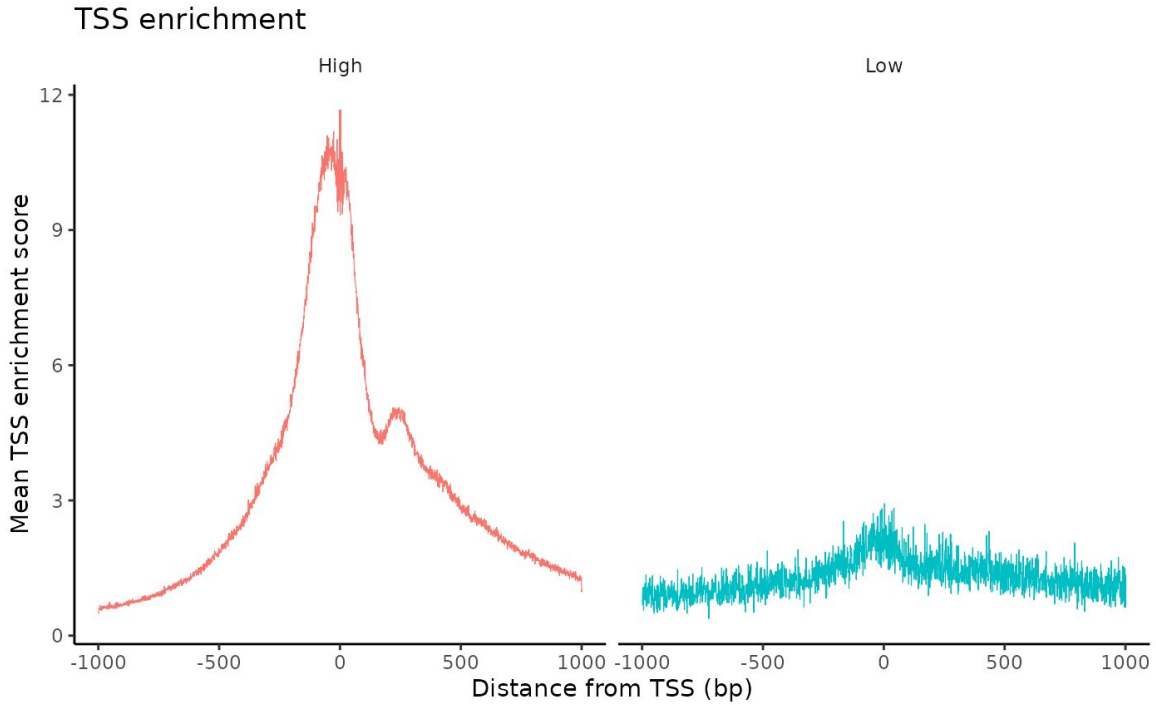
Buenrostro et al. 2013. Nature Methods.
DOI: 10.1038/nmeth.2688

scATAC-seq QC - TSS enrichment

Signal around a window of 2,000 bases around TSSs.

This profile is helpful to assess the signal-to-noise ratio of the library.

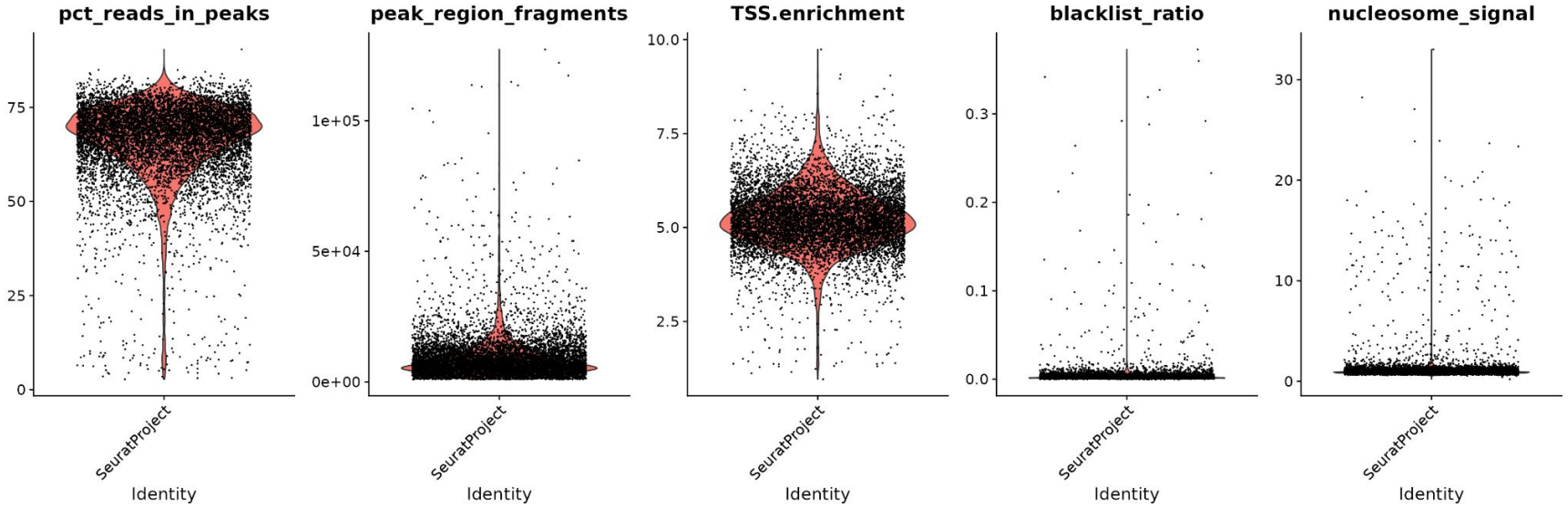
TSSs and the promoter regions around them have, on average, a high degree of chromatin accessibility compared to the intergenic and intronic regions of the genome.



https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac

https://satijalab.org/signac/articles/pbmc_vignette.html

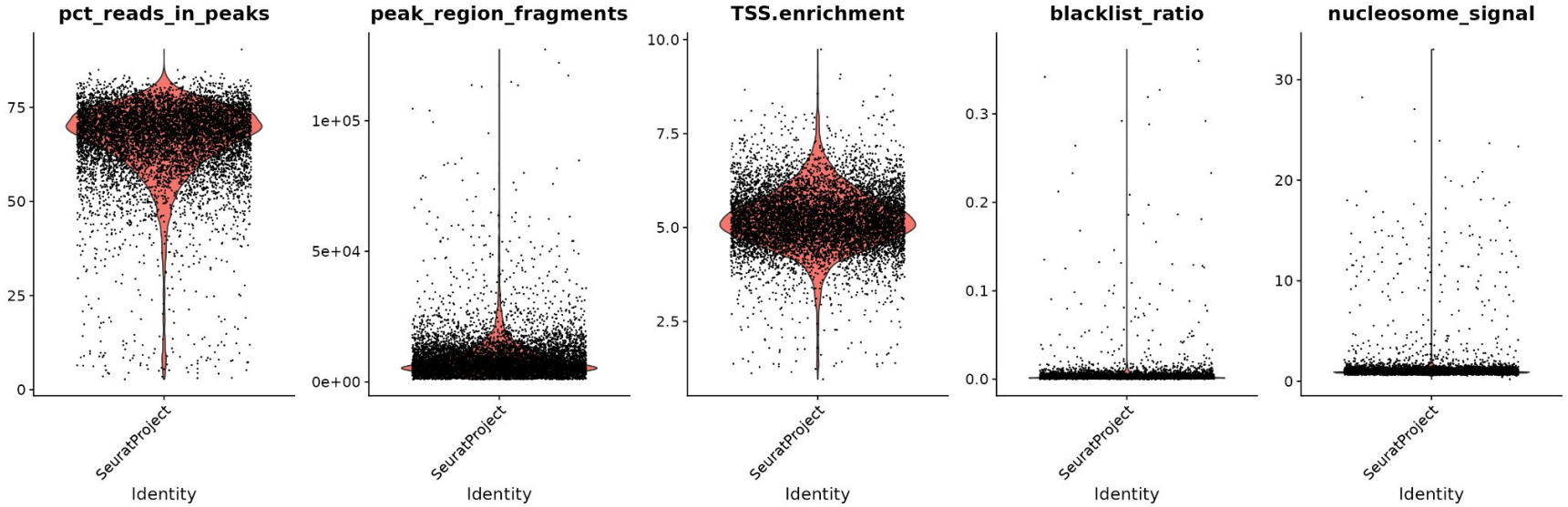
scATAC-seq QC - Metrics



Fraction of fragments in peaks: Represents the fraction of all fragments that fall within ATAC-seq peaks. Cells with low values (i.e. <15-20%) often represent low-quality cells or technical artifacts that should be removed. Note that this value can be sensitive to the set of peaks used.

https://satijalab.org/signac/articles/pbmc_vignette.html

scATAC-seq QC - Metrics



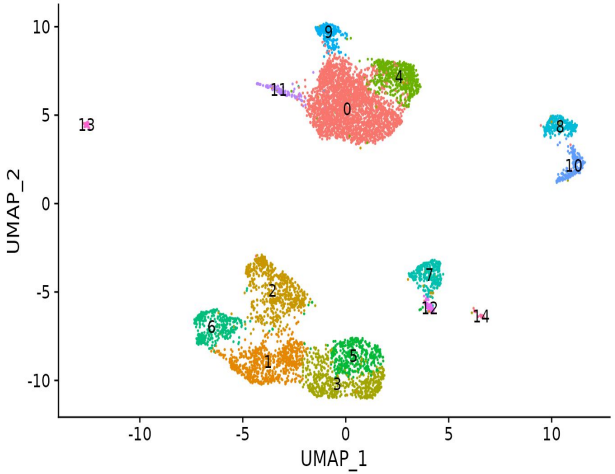
Ratio reads in genomic blacklist regions: The ENCODE project has provided a list of blacklist regions, representing reads which are often associated with artifactual signal. Cells with a high proportion of reads mapping to these areas (compared to reads mapping to peaks) often represent technical artifacts and should be removed.

https://satijalab.org/signac/articles/pbmc_vignette.html

Single-cell ATAC-seq dimension reduction

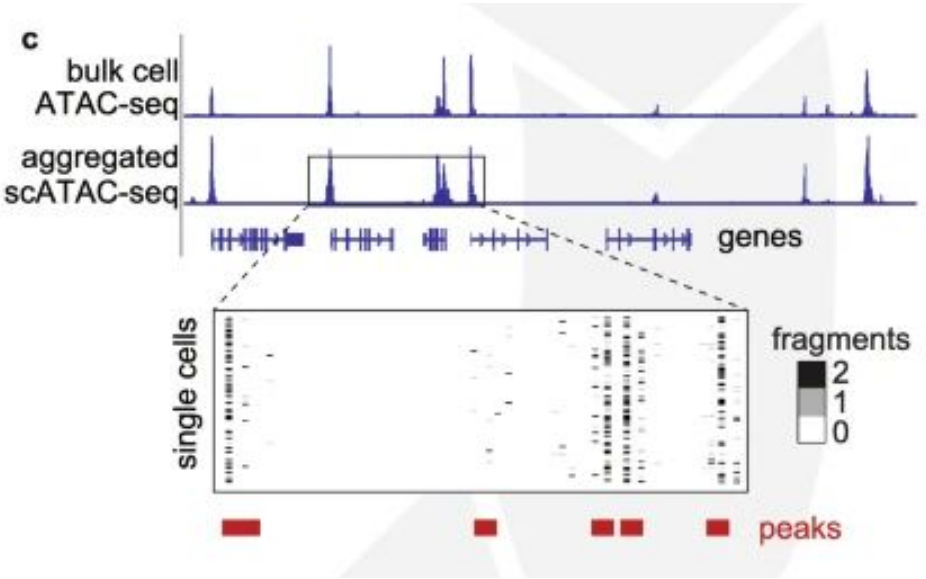
Latent Semantic Indexing (LSI) is an approach from natural language processing that was originally designed to assess document similarity based on word counts.

1. scATAC-seq: documents=samples, words=regions/peaks.
   scRNA-seq: documents=samples, words=genes.
2. Calculate word frequency by depth normalization per single cell.
3. Normalize word frequency by the inverse document frequency which weights features by how often they occur.
4. Results in a word frequency-inverse document frequency (TF-IDF) matrix, which reflects how important a word (aka region/peak) is to a document (aka sample).
5. Perform singular value decomposition (SVD) on the TF-IDF matrix.

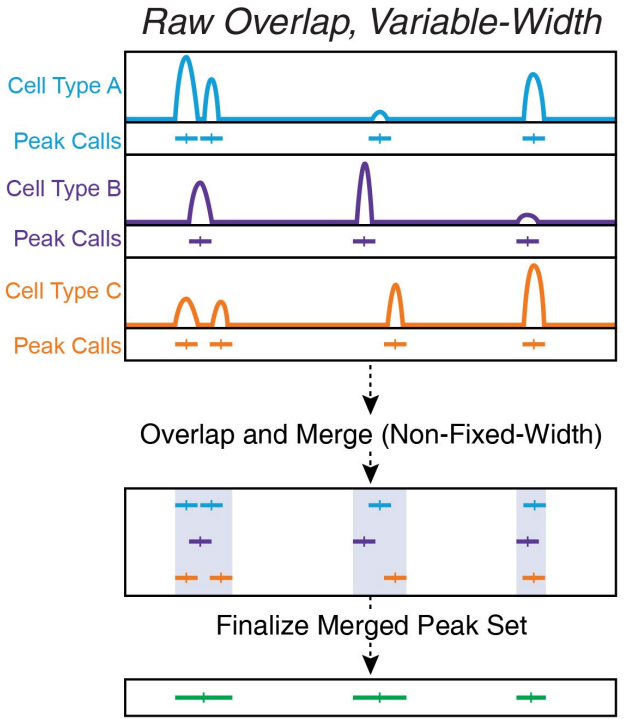


https://www.archrproject.com/bookdown/dimensionality-reduction-with-archr.html

Single-cell ATAC-seq peak calling

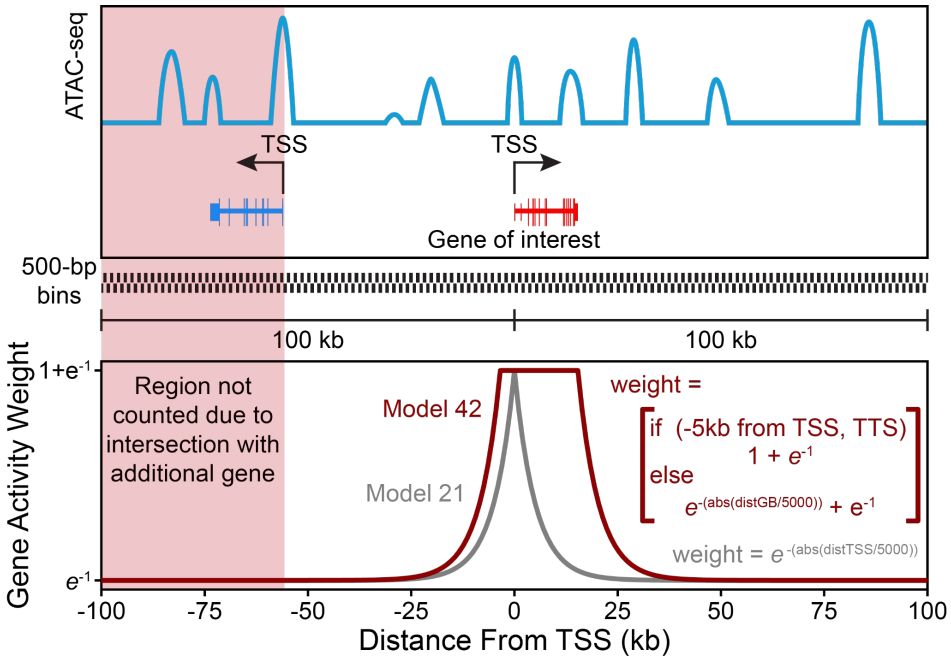


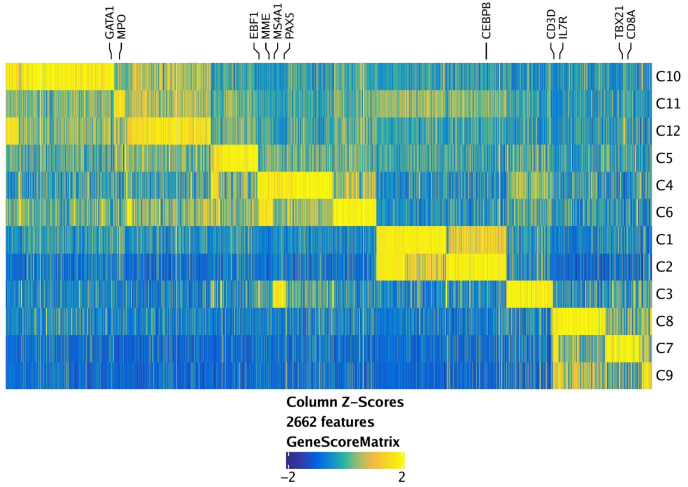
Chen et al. 2019. Genome Biology.
DOI: 10.1186/s13059-019-1854-5



https://www.archrproject.com/bookdown/the-iterative-overlap-peak-merging-procedure.html

Single-cell ATAC-seq gene activity



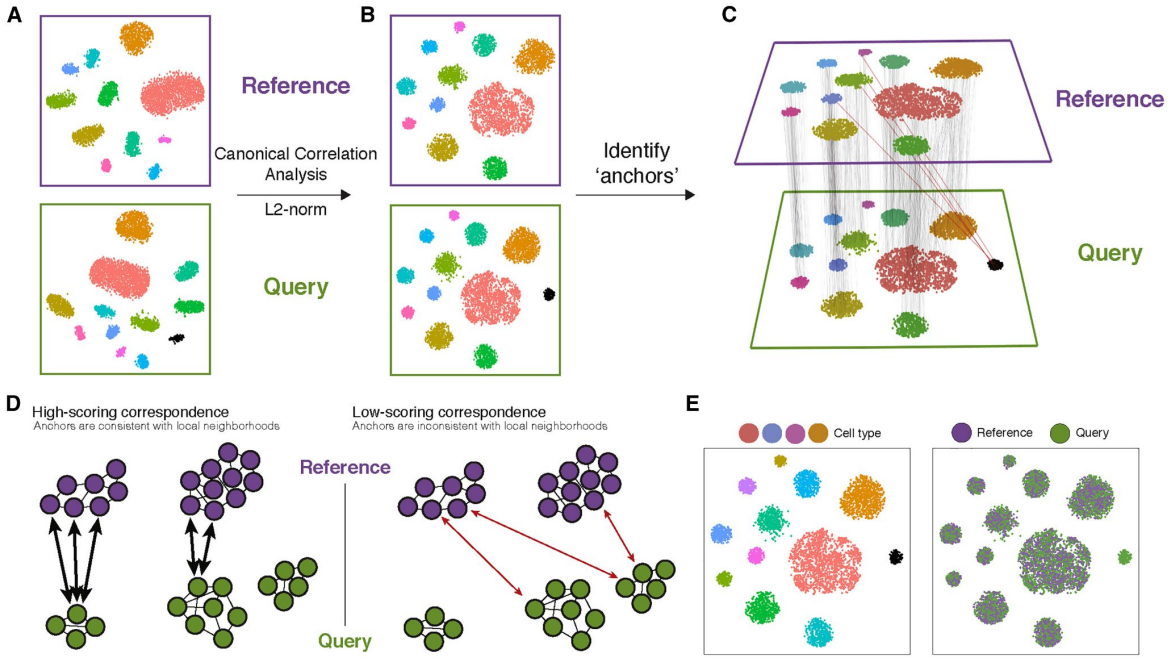


https://www.archrproject.com/bookdown/calculating-gene-scores-in-archr.html

scATAC-seq & scRNA-seq integration Label transfer

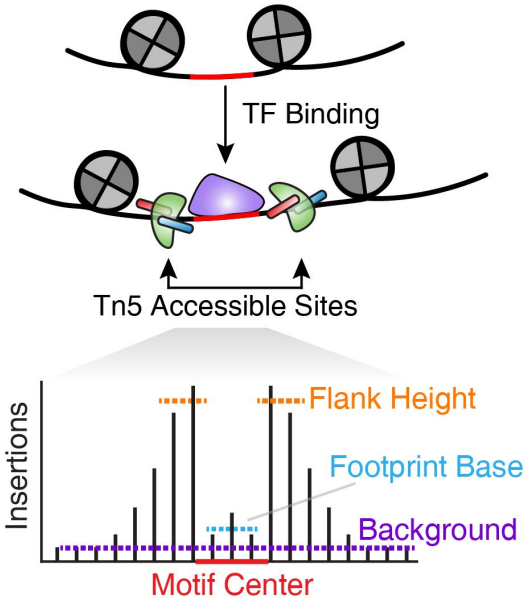
Identify shared correlation patterns in the gene activity matrix and scRNA-seq dataset to identify matched biological states across the two modalities.

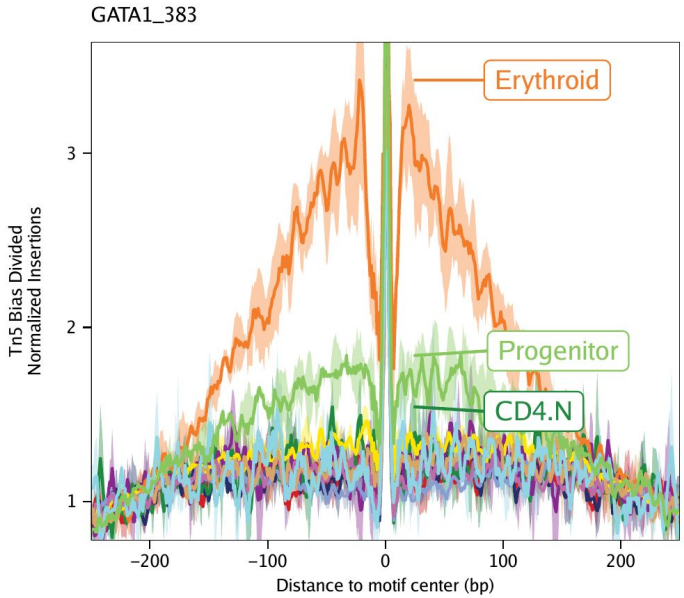
This procedure returns a classification score for each cell for each scRNA-seq-defined cluster label.



Stuart et al. 2019. Cell.
DOI: 10.1016/j.cell.2019.05.031

Single-cell ATAC-seq Footprinting

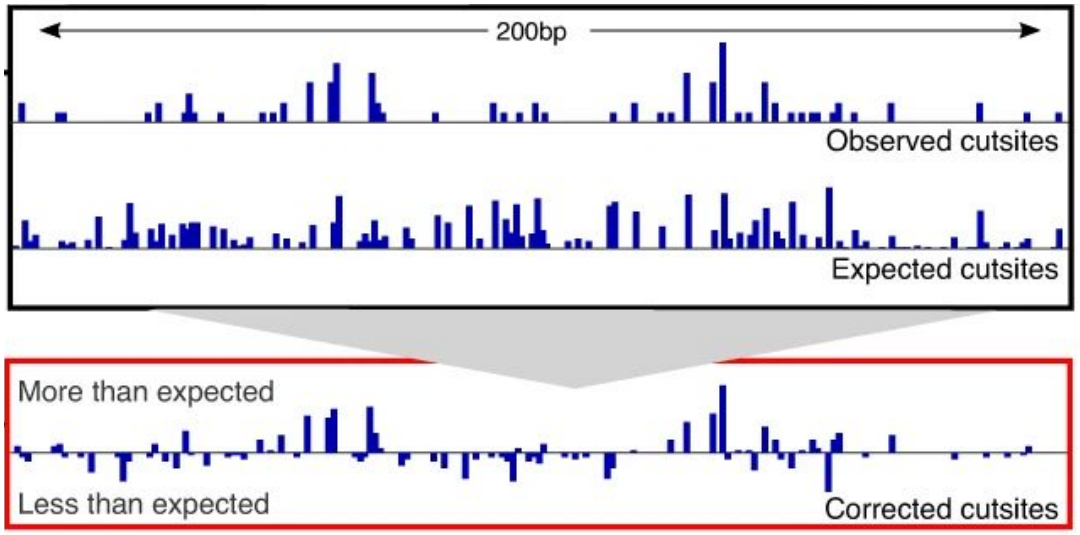


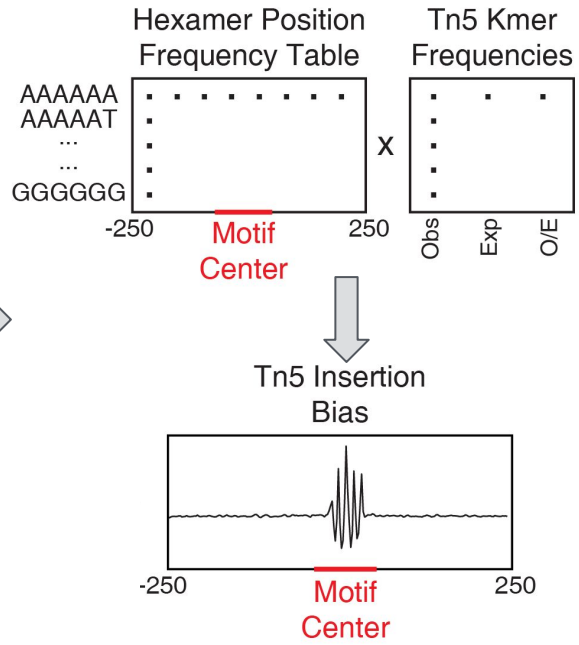


TF footprinting allows for the prediction of the precise binding location of a TF at a particular locus. The DNA bases that are bound by the TF are protected from the Tn5 while the DNA bases immediately adjacent to TF binding are accessible.

https://www.archrproject.com/bookdown/footprinting-with-archr.html

Single-cell ATAC-seq Footprinting





Bentsen et al. 2020. Nature Com.
DOI: 10.1038/s41467-020-18035-1

The Tn5 has an insertion bias which needs to be corrected for; if not, false positive/negative predictions!

https://www.archrproject.com/bookdown/footprinting-with-archr.html

Perspectives

https://www.archrproject.com/

https://satijalab.org/signac/index.html

https://github.com/colomemaria/epiScanpy

https://support.10xgenomics.com/single-cell-atac

Session 1 - scATAC-seq analysis

Hands-on sessions:

https://www.hdsu.org/sincellTE_2022/

RStudio server:

https://rstudio-singlecell.sb-roscoff.fr/

BioQuant
MODEL base of LIFE

healthdata
scienceunit

HEIDELBERG
FACULTY OF
MEDICINE

Dr. Carl Herrmann
Dr. Carlos Ramirez
Dr. Andres Quintero
Daria Doncevic
Ana Luísa Costa
Youcheng Zhang

healthdata scienceunit

HEIDELBERG FACULTY OF MEDICINE

Thank you!