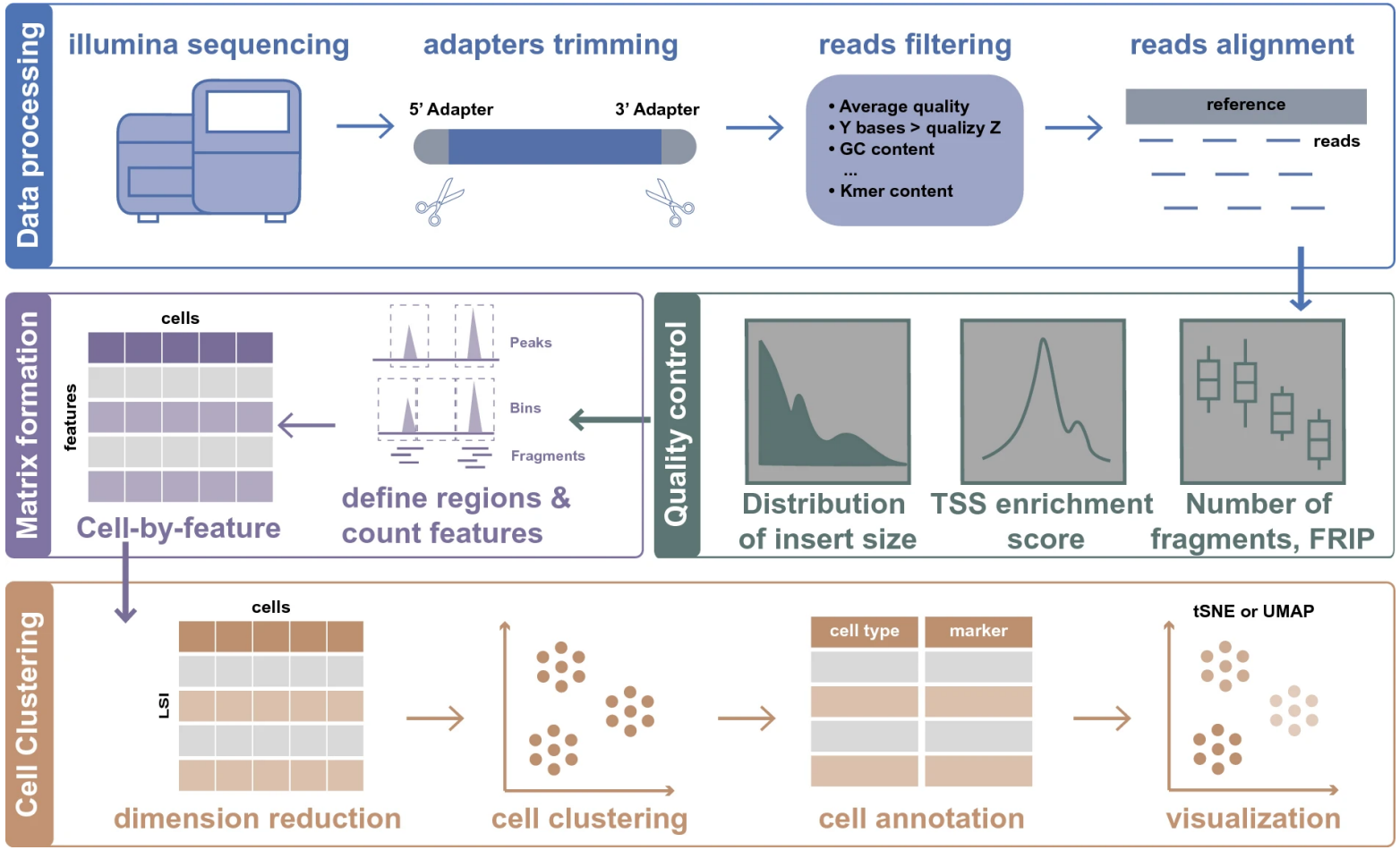
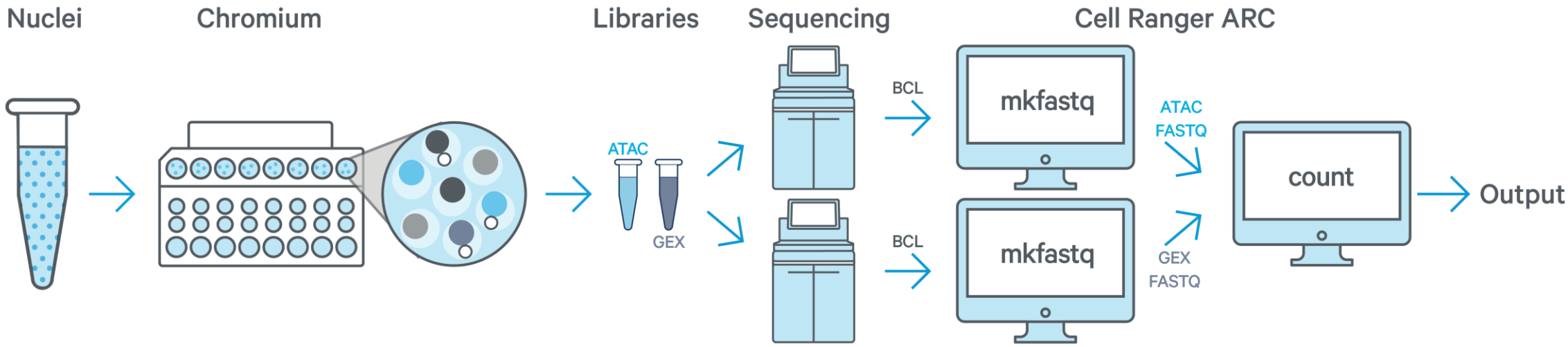
Basic scATAC-seq workflow

Main Steps for scATACseq Workflow



https://doi.org/10.1007/s42994-022-00082-5

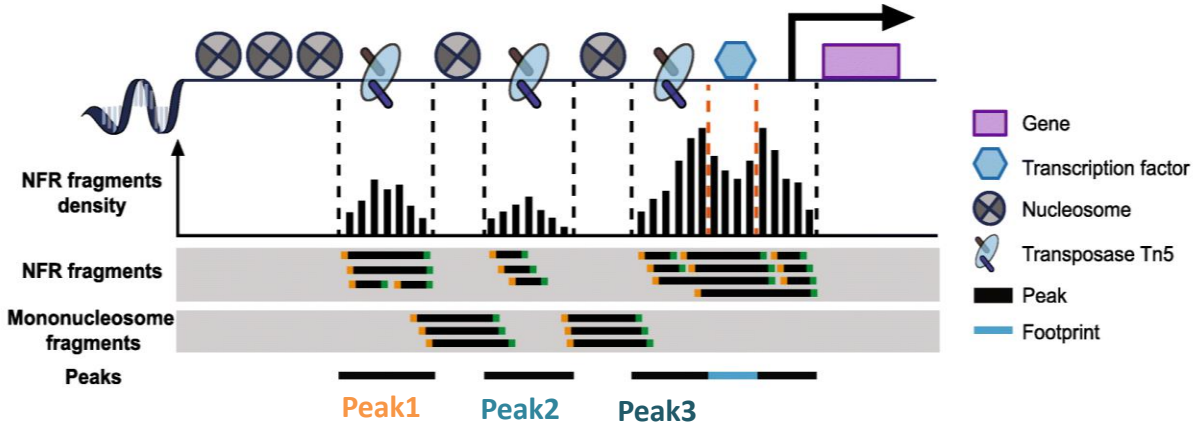
Step1. Data processing



- Sample preparation involves tissue dissociation and nuclei isolation.
- **Library construction**: Tn5 transposition and barcoding.
- **Alignment**: Mapping reads to the reference genome.

https://www.10xgenomics.com/support/jp/software/cell-ranger-arc/latest/getting-started/what-is-cell-ranger-arc

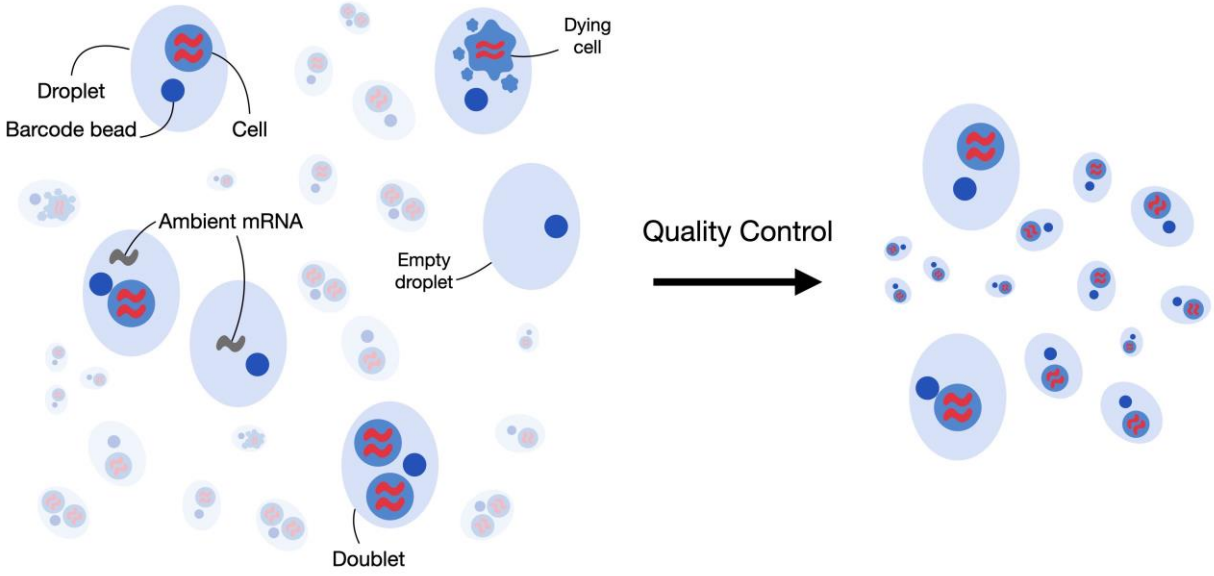
Step1. Data processing



	Cell1	Cell2	...	CellN
Peak1	3	1	.	2
Peak2	2	0	.	1
Peak3	1	1	.	0
...	.	.	.	.
PeakM	5	1	.	0

- **Peak Calling**: Identifying accessible chromatin regions
- **Fragment Counting**: Quantifying accessibility per region

Step2. Quality control



- Why QC is crucial in snATAC-seq: Ensures high-quality and biologically relevant data.
- Common QC Metrics: Read depth, TSS enrichment, fragment size distribution, and nucleosome positioning.
- Doublet detection

https://www.sc-best-practices.org/preprocessing_visualization/quality_control.html

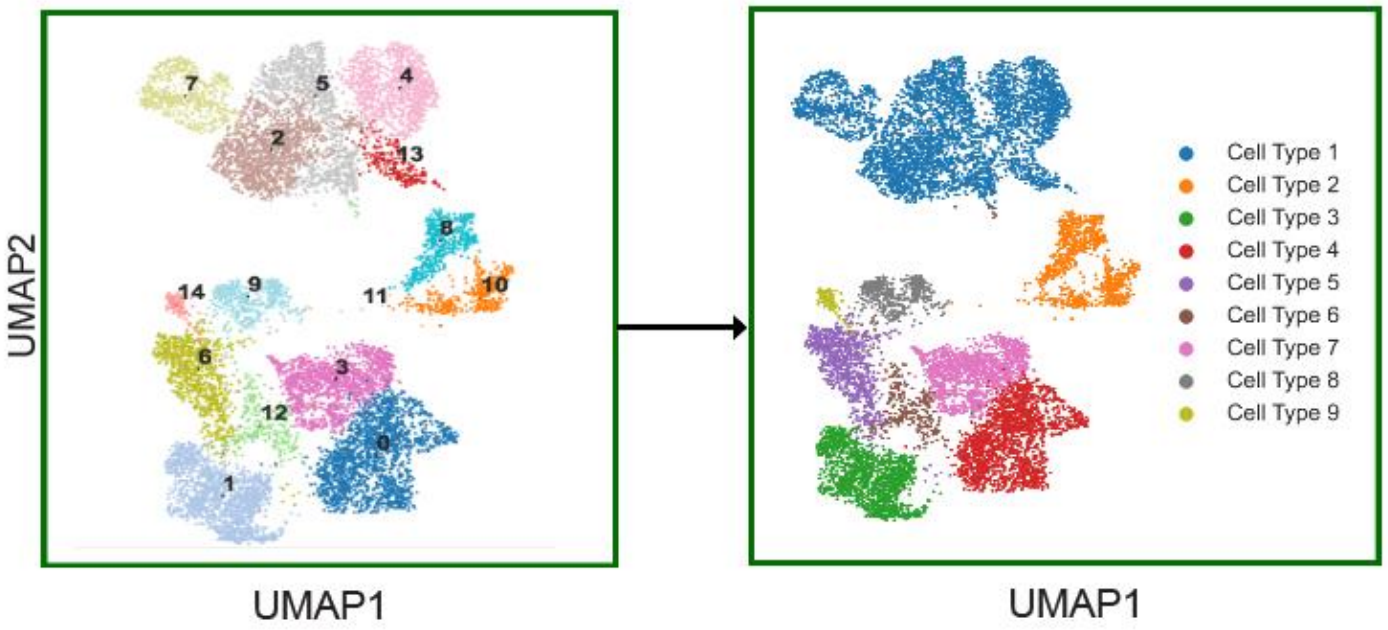
Step3. Matrix formation & Normalization

	Cell1	Cell2	...	CellN
Peak1	3	1	.	2
Peak2	2	0	.	1
Peak3	1	1	.	0
...	.	.	.	.
PeakM	5	1	.	0

	Cell1	Cell2	...	CellN
Peak1	0.1	0.2	.	0.3
Peak2	0.05	0	.	0.2
Peak3	0.02	0.2	.	0
...	.	.	.	.
PeakM	0.2	0.2	.	0

- Peak x Cell Matrix
- Cell-level Normalization: Adjusting for variations in sequencing depth across individual cells
- Peak-level Normalization: Scaling peak signals to assign higher values to rare peaks, enhancing the detection of less abundant regulatory regions

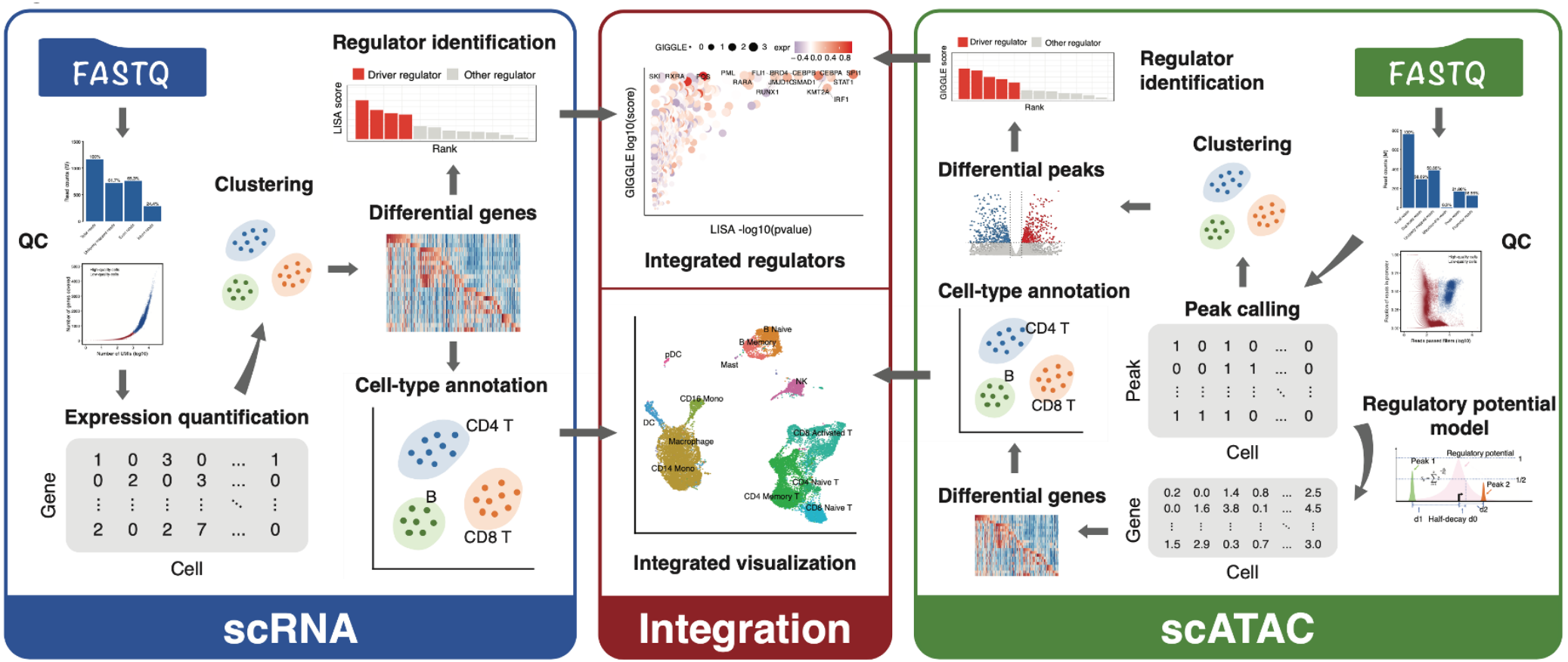
Step4. Cell clustering and annotation



- Cell Clustering: clustering after dimension reduction
- Cluster-level Annotation: using well-established cell type-specific markers
- Downstream Analysis: cluster analysis, differential accessibility analysis, and integration with multi-omics datasets

Jiang et al., scATAnno: Automated Cell Type Annotation for single-cell ATAC-seq Data

Step5. Downstream Analysis



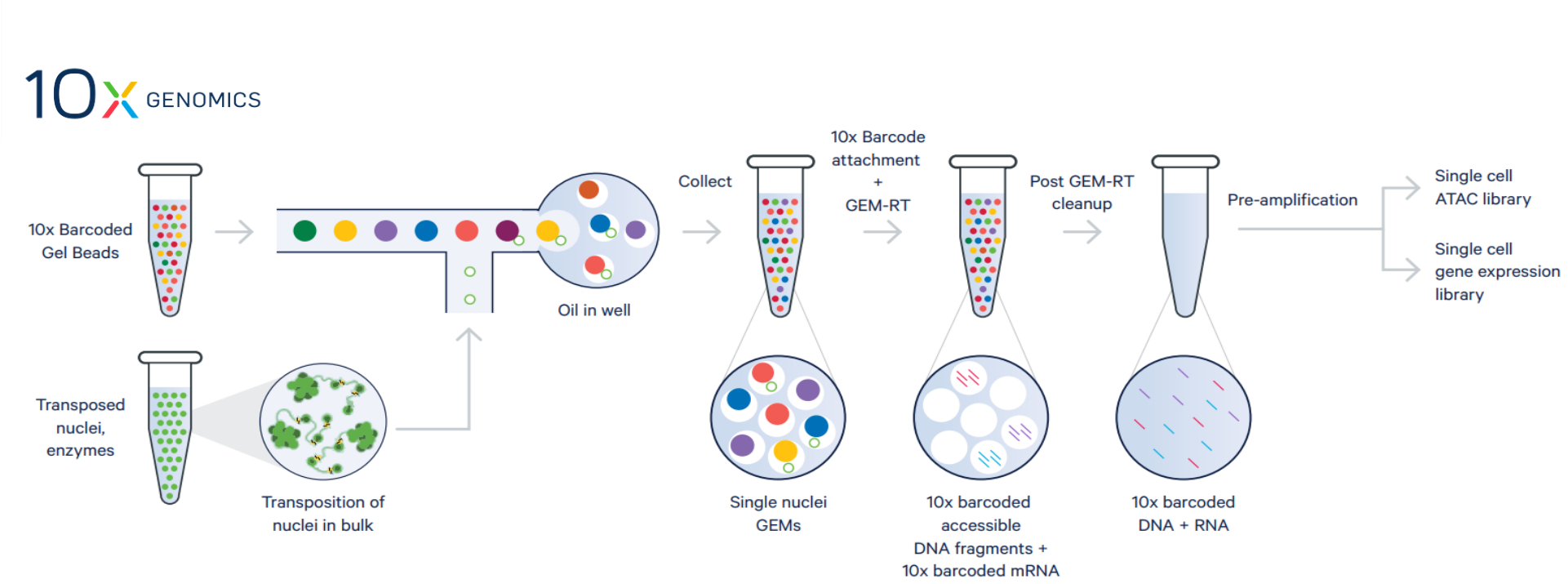
https://liulab-dfci.github.io/MAESTRO/

Hands-On Example Dataset

Assay: Single Cell Multiome (ATAC + Gene Expression)

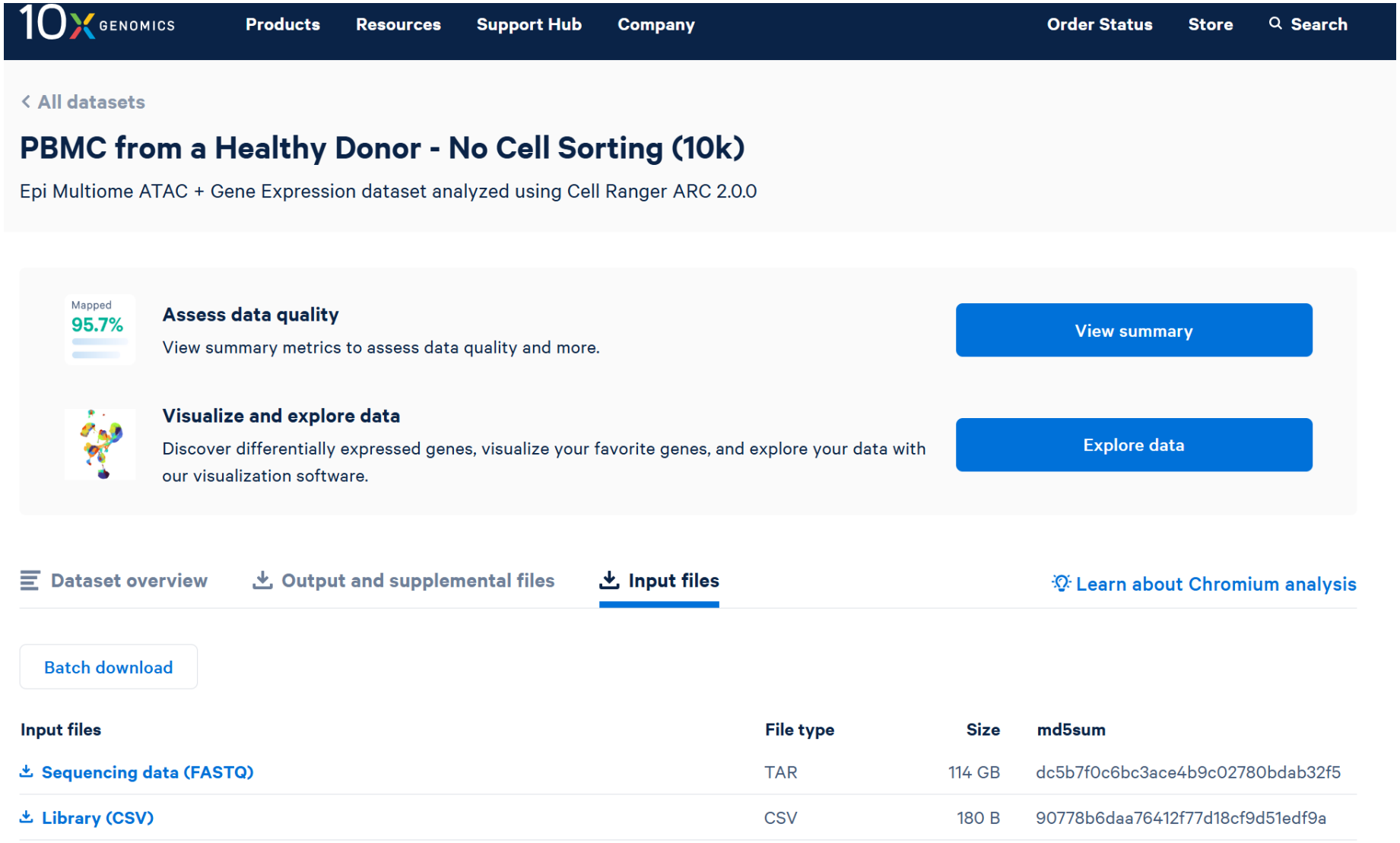
Sample Type: Peripheral Blood Mononuclear Cells (PBMCs)

Donors: Two healthy individuals



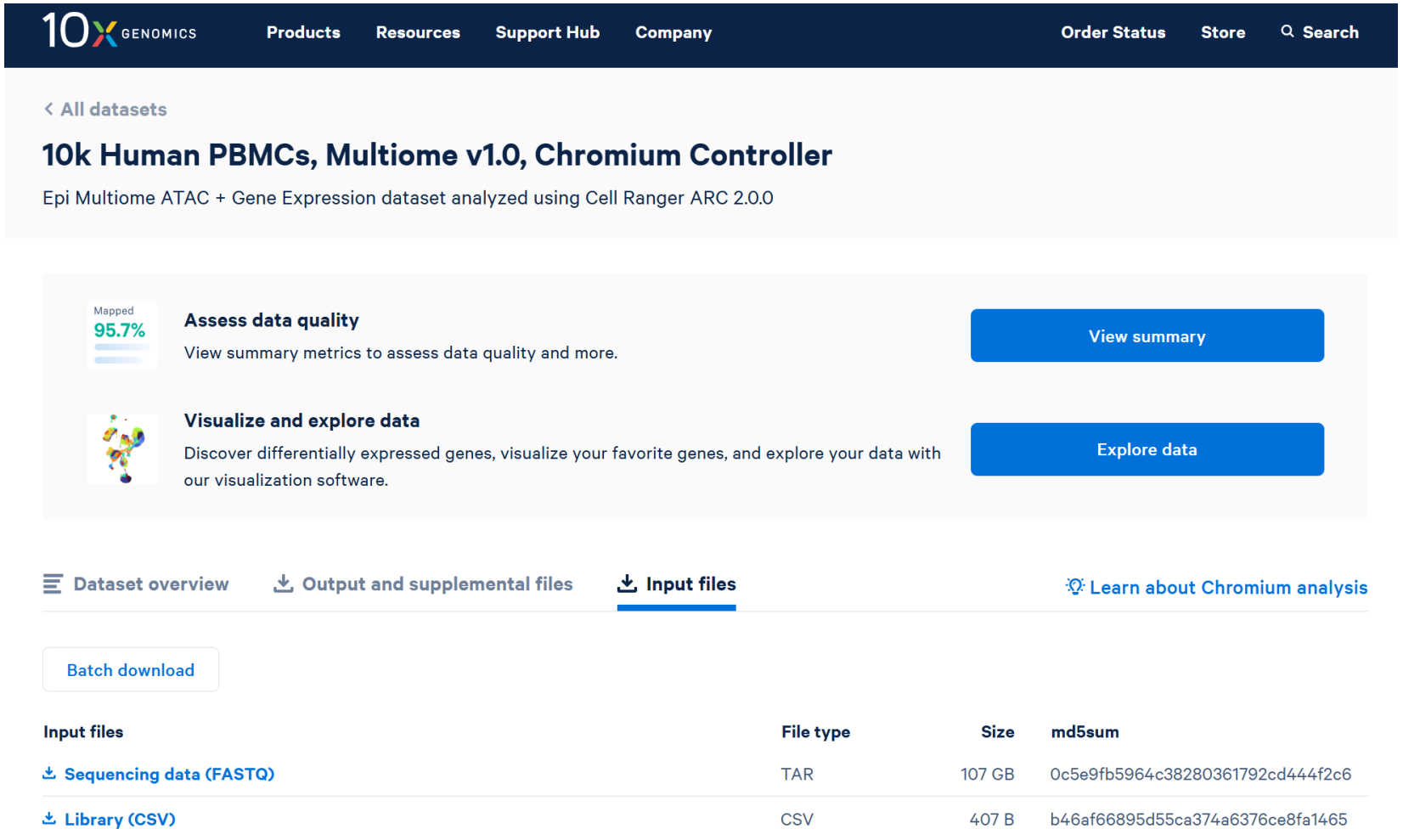
- a healthy female donor aged 25

https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-10-k-1-standard-2-0-0

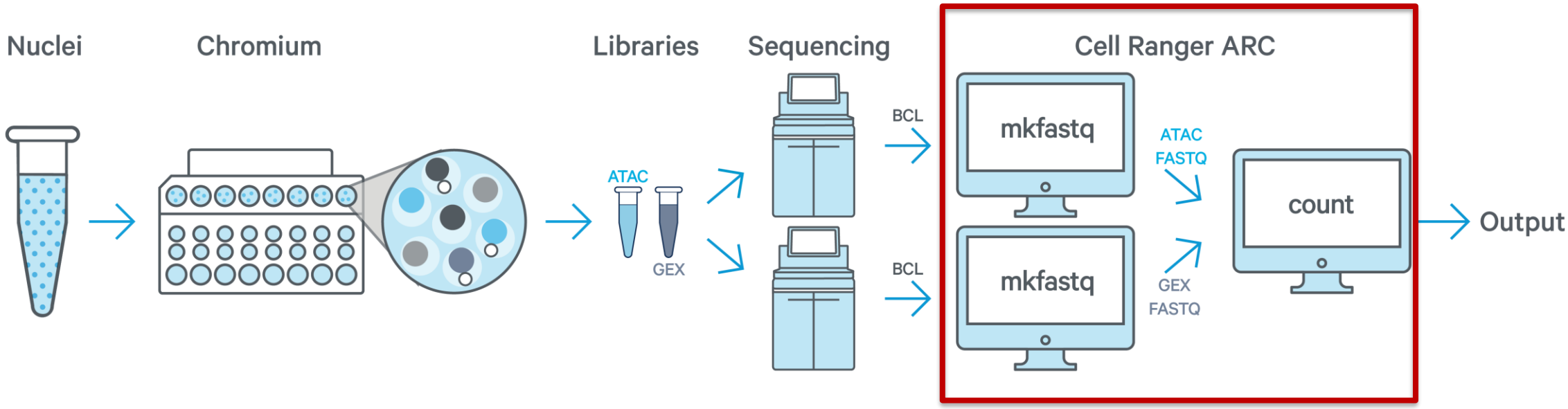


- a healthy male donor aged 30-35

  https://www.10xgenomics.com/datasets/10-k-human-pbm-cs-multiome-v-1-0-chromium-controller-1-standard-2-0-0



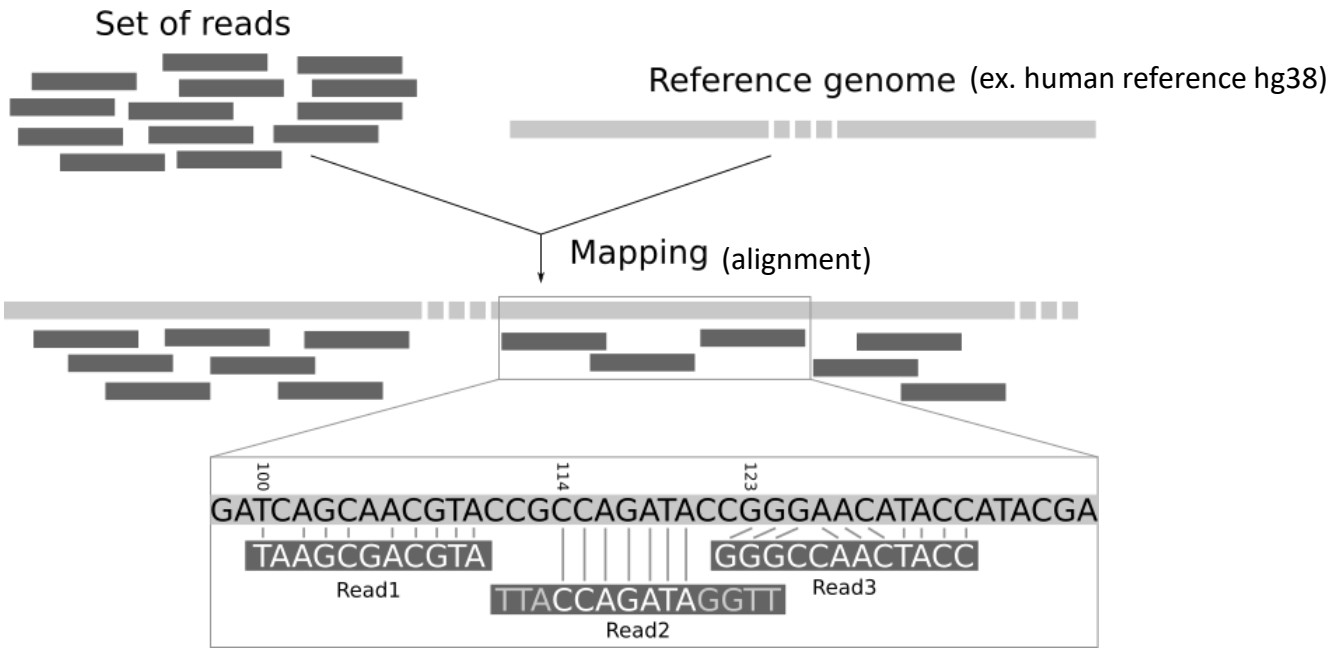
Step1. Generate single cell feature matrix



- **Alignment**: Mapping reads to the reference genome.
- **Peak Calling:** Identifying accessible chromatin regions
- **Fragment Counting:** Quantifying accessibility per region

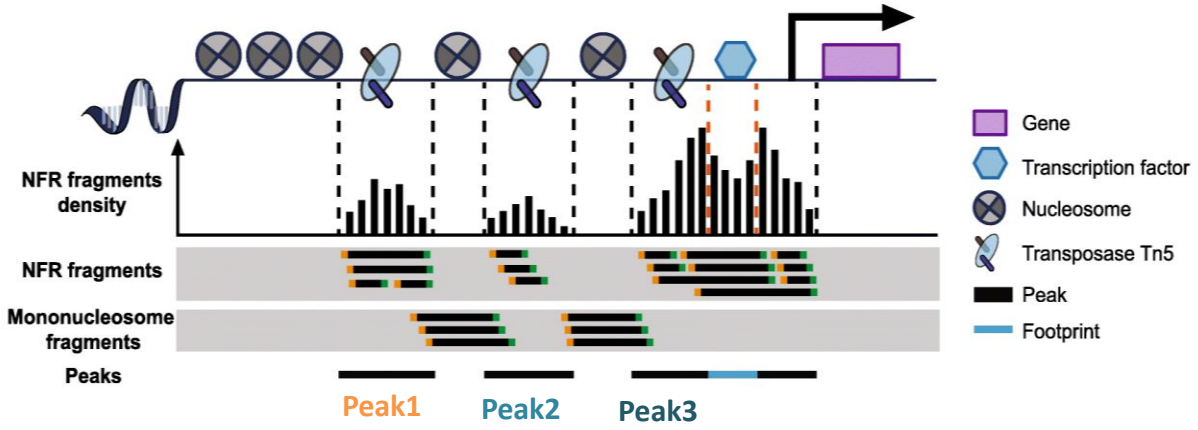
https://www.10xgenomics.com/support/jp/software/cell-ranger-arc/latest/getting-started/what-is-cell-ranger-arc

Step1-1. Alignment



- **Alignment**: Mapping reads to the reference genome.
- Peak Calling**:** Identifying accessible chromatin regions
- Fragment Counting: Quantifying accessibility per region

Step1-2. Peak calling & Fragment counting

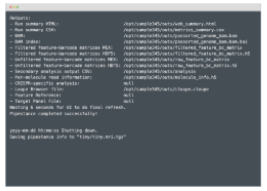


	Cell1	Cell2	…	CellN
Peak1	3	2	.	13
Peak2	2	3	.	1
Peak3	1	14	.	18
	.	.	.	.
	.	.	.	.
	.	.	.	.
PeakM	25	0	.	0

- Alignment: Mapping reads to the reference genome.
- **Peak Calling:** Identifying accessible chromatin regions
- **Fragment Counting:** Quantifying accessibility per region

Genome Biology volume 21, Article number: 22 (2020)

Step1. Generate single cell feature matrix

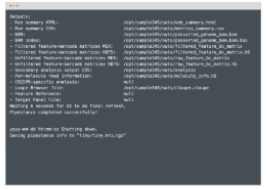


Cell Ranger

Analyze Single Cell Gene Expression and Single Cell Immune Profiling data with a set of free, easy-to-use analysis pipelines.

Learn more > Download >

single cell RNA,
single cell Immune

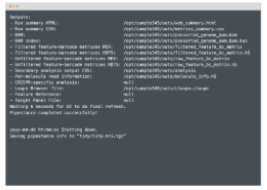


Cell Ranger ATAC

Analyze Single Cell ATAC data with a set of free, easy-to-use analysis pipelines for identification of open chromatin regions, motif annotation, and differential accessibility analysis.

Learn more > Download >

single cell ATAC



Cell Ranger ARC

Analyze Single Cell Multiome ATAC + Gene Expression data with a set of free, easy-to-use pipelines for primary and secondary analyses.

Learn more > Download >

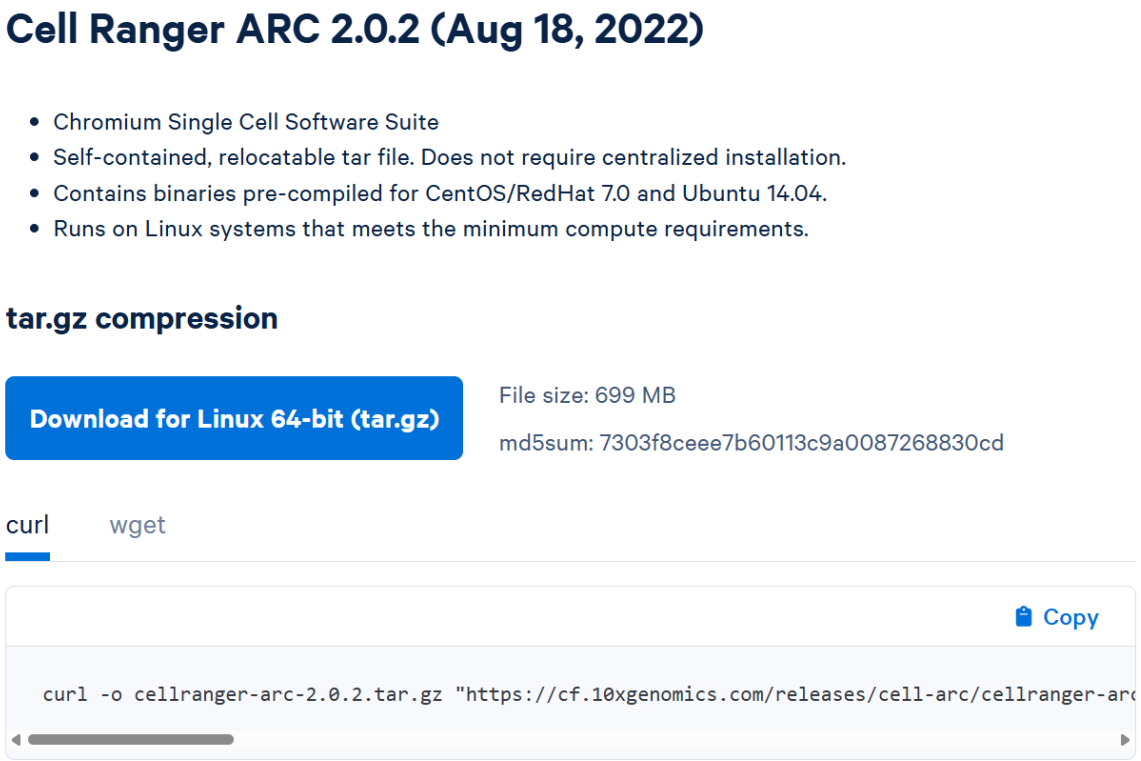
single cell Multiome
(ATAC +Gene expression)

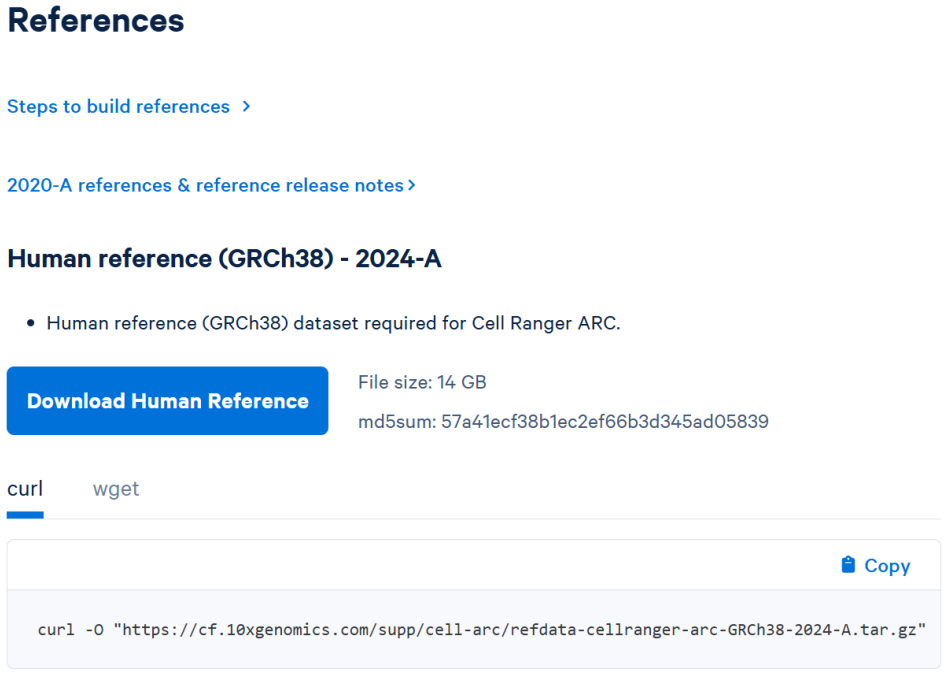
https://www.10xgenomics.com/software

Step1-1. Download Cell Ranger ARC

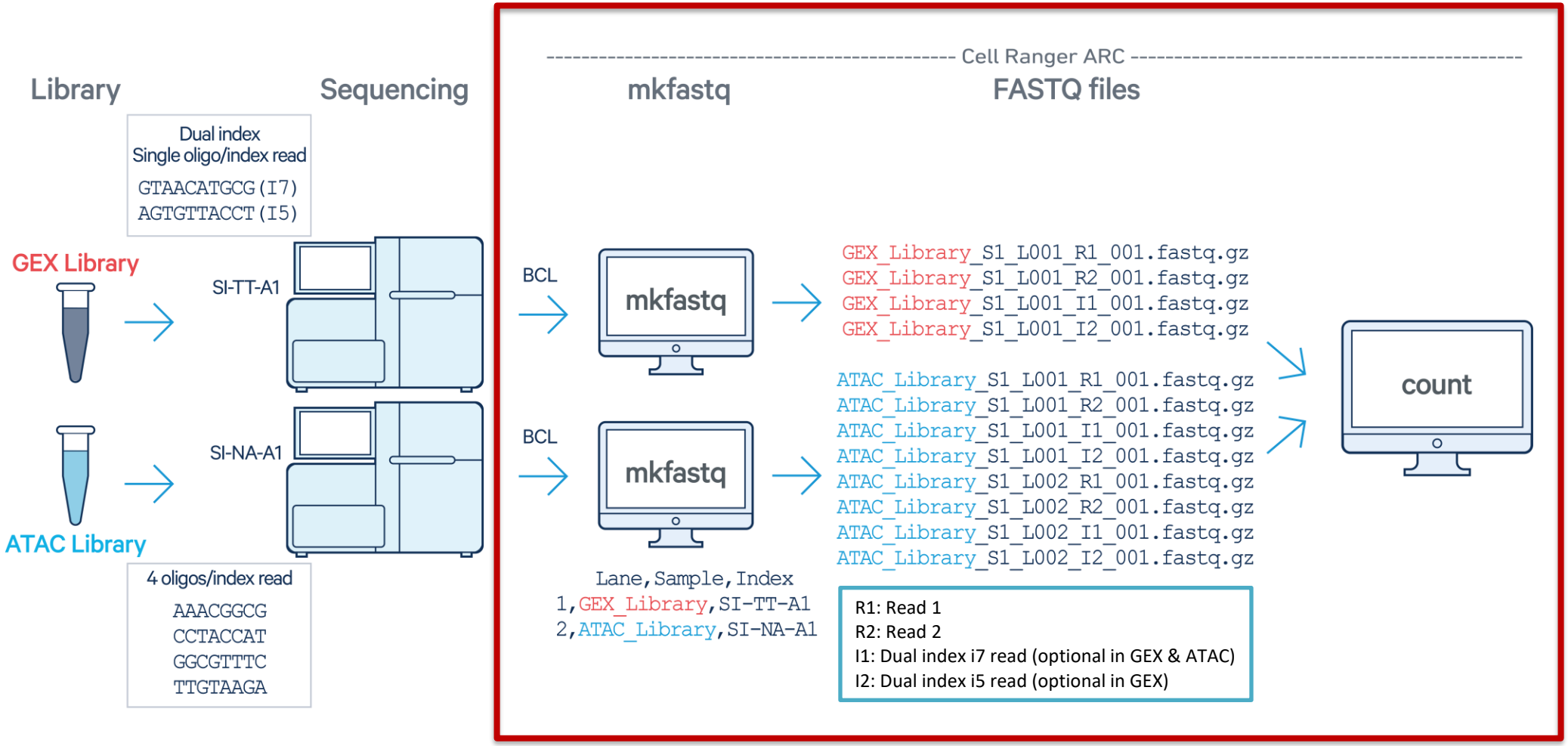
https://www.10xgenomics.com/support/software/cell-ranger-arc/downloads

Runs on Linux systems





Step1-2. Specifying input FASTQ files for cellranger-arc count



Step1-2. Specifying input FASTQ files for cellranger-arc count

ATAC FASTQs

`[Sample Name]`*S1_L00*`[Lane Number]``[Read Type]`_001.fastq.gz

Where `Read Type` is one of:

- `I1` : Dual index i7 read (optional)
- `R1` : Read 1
- `R2` : Dual index i5 read
- `R3` : Read 2

Cell Ranger ARC will also accept ATAC FASTQs in this format:

- `I1` : Dual index i7 read (optional)
- `R1` : Read 1
- `I2` : Dual index i5 read
- `R2` : Read 2

GEX FASTQs

`[Sample Name]`*S1_L00*`[Lane Number]``[Read Type]`_001.fastq.gz

Where `Read Type` is one of:

- `I1` : Dual index i7 read (optional)
- `I2` : Dual index i5 read (optional)
- `R1` : Read 1
- `R2` : Read 2

https://www.10xgenomics.com/support/software/cell-ranger-arc/latest/analysis/inputs/specifying-input-fastq-count

Step1-3. Create a libraries CSV file

- CSV file format

Column Name	Description
fastqs	A fully qualified path to the *directory* containing the demultiplexed FASTQ files for this sample. This field does not accept comma-delimited paths. If you have multiple sets of fastqs for this library, add an additional row, and use the use same library_type value.
sample	Sample name assigned as the Sample_ID in the demultiplexing sample sheet.
library_type	This field is case-sensitive and must exactly match Chromatin Accessibility for a Multiome ATAC library and Gene Expression for a Multiome GEX library.

- example of libraries CSV file

fastqs,sample,library_type
/home/jdoe/runs/HNGEXSQXXX/outs/fastq_path,example,Gene Expression
/home/jdoe/runs/HNATACSQXX/outs/fastq_path,example,Chromatin Accessibility

https://www.10xgenomics.com/support/software/cell-ranger-arc/latest/analysis/running-pipelines/single-library-analysis

Step1-4. Run cellranger-arc count

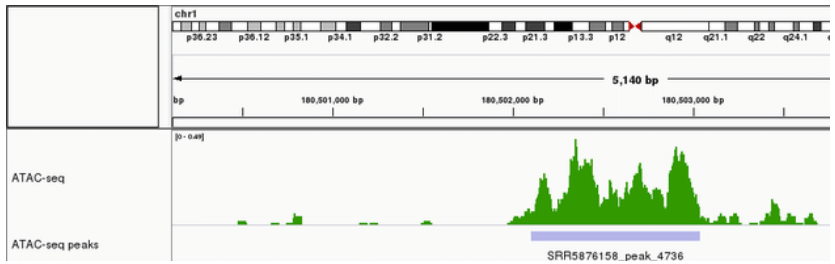
```
cellranger-arc count --id=[A unique run ID string] \
                    --reference=[Path to the cellranger-arc-compatible reference package] \
                    --libraries=[Path to a 30column CSV file]
```

Output files

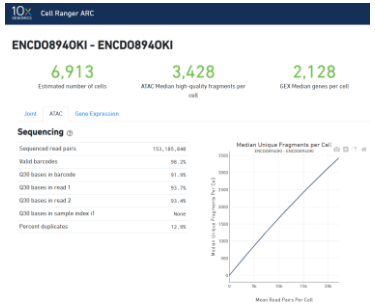
peak x count matrix

	Cell1	Cell2	...	CellN
Peak1	3	1	.	2
Peak2	2	0	.	1
Peak3	1	1	.	0
...	.	.	.	.
PeakM	5	1	.	0

fragment alignment file



QC info + QC summary in html



https://www.10xgenomics.com/support/software/cell-ranger-arc/latest/analysis/running-pipelines/single-library-analysis

web_summary.html generated by CellRanger

High-Quality Sample

10x GENOMICS Cell Ranger ARC

Sample A

6,913	3,428	2,128
Estimated number of cells	ATAC Median high-quality fragments per cell	GEX Median genes per cell

Sample requiring attention due to suboptimal QC metrics

Due to the high likelihood of doublet formation,
rigorous doublet filtering is recommended.

10x GENOMICS Cell Ranger ARC

Sample B

Alerts

The analysis detected 1 error.

	Alert	Value	Detail
!	Number of cells detected is high	15,469	Ideally < 10,000 cells. This can be caused by incorrect quantification of the nuclei suspension, improper handling of nuclei,excessive background RNA and DNA, or unexpected behavior in the cell calling algorithm.

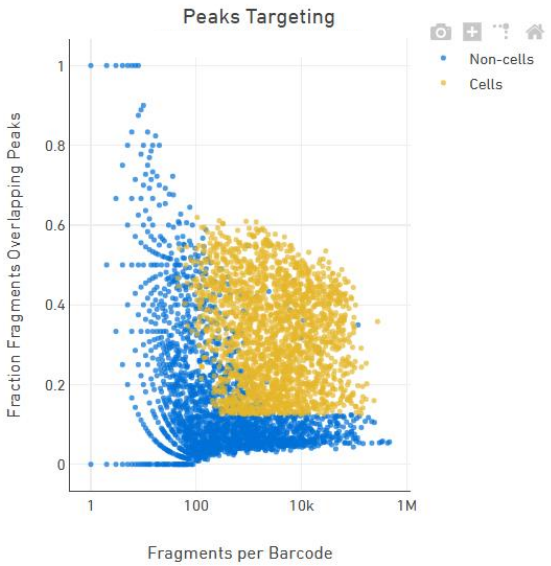
15,469	2,832	840
Estimated number of cells	ATAC Median high-quality fragments per cell	GEX Median genes per cell

web_summary.html generated by CellRanger

High-Quality Sample

Cells

Estimated number of cells	6,913
Mean raw read pairs per cell	22,159.10
Fraction of high-quality fragments in cells	68.0%
Fraction of transposition events in peaks in cells	33.3%
Median high-quality fragments per cell	3,428

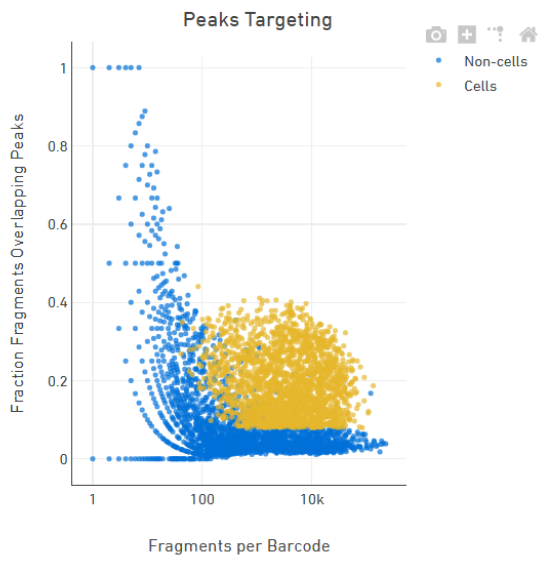


Sample requiring attention due to suboptimal QC metrics

	Alert	Value	Detail
⚠	ATAC Transposition events within peaks is low	18.8%	It is expected that more than 25% of the transposition events fall within peak regions. A lower value could suggest a problem during library preparation causing excessive background transposition, low sequencing depth, or unexpected behavior in the peak calling algorithm resulting in very few peaks detected.
⚠	ATAC Fragments in peaks is low	20.3%	Ideal > 25%. A low value can be caused by a problem during the transposition step, due to a population of cells with un-compacted DNA (e.g., activated granulocytes, dead or dying cells, unsupported organisms) or due to unexpected behavior in the peak calling algorithm resulting in very few peaks called.

Cells

Estimated number of cells	8,531
Mean raw read pairs per cell	26,011.97
Fraction of high-quality fragments in cells	43.3%
Fraction of transposition events in peaks in cells	18.8%
Median high-quality fragments per cell	4,559

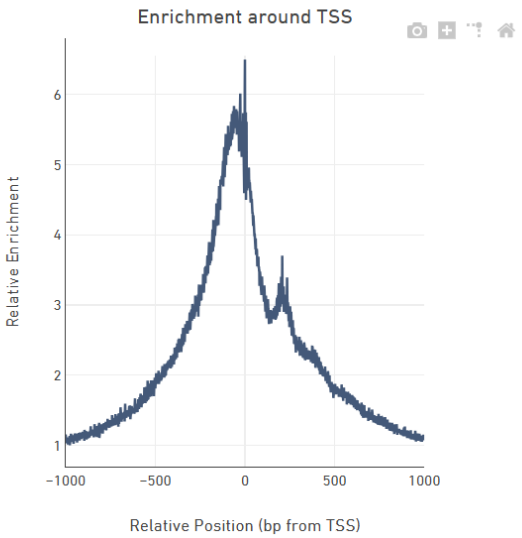


web_summary.html generated by CellRanger

High-Quality Sample

Targeting

Number of peaks	93,688
Fraction of genome in peaks	2.6%
TSS enrichment score	6.50
Fraction of high-quality fragments overlapping TSS	29.4%
Fraction of high-quality fragments overlapping peaks	36.0%



Sample requiring attention due to suboptimal QC metrics

⚠	ATAC TSS enrichment is low	4.67	Ideal > 5. A low TSS score can be caused by poor sample prep, poor sample quality, a population of cells with highly accessible DNA (e.g., activated granulocytes, dead or dying cells, unsupported organisms) or poor quality reference genome annotation.

Targeting

Number of peaks	55,882
Fraction of genome in peaks	1.6%
TSS enrichment score	4.67
Fraction of high-quality fragments overlapping TSS	21.8%
Fraction of high-quality fragments overlapping peaks	20.3%