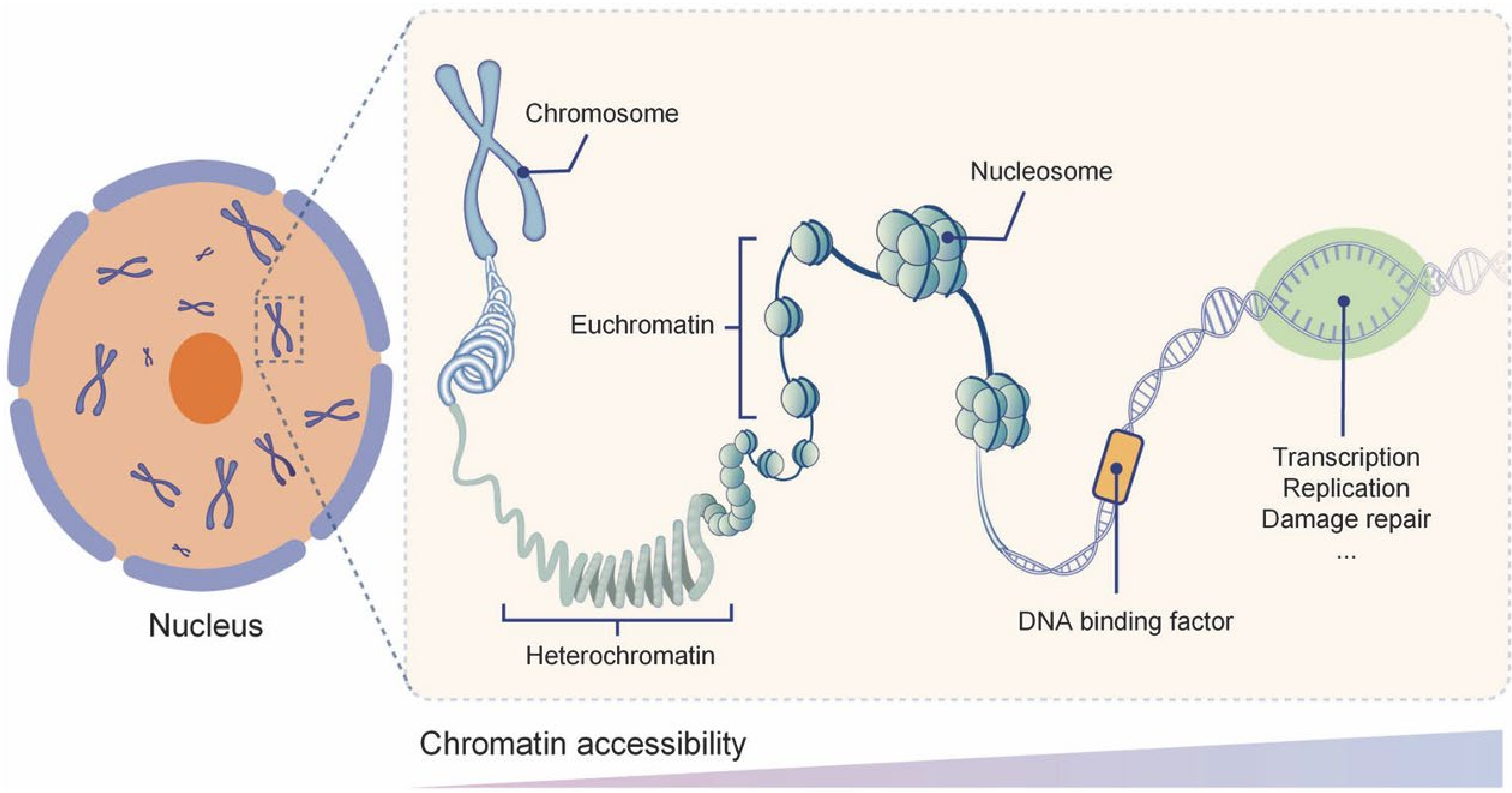
archR包的功能简介与示例

LEB G09

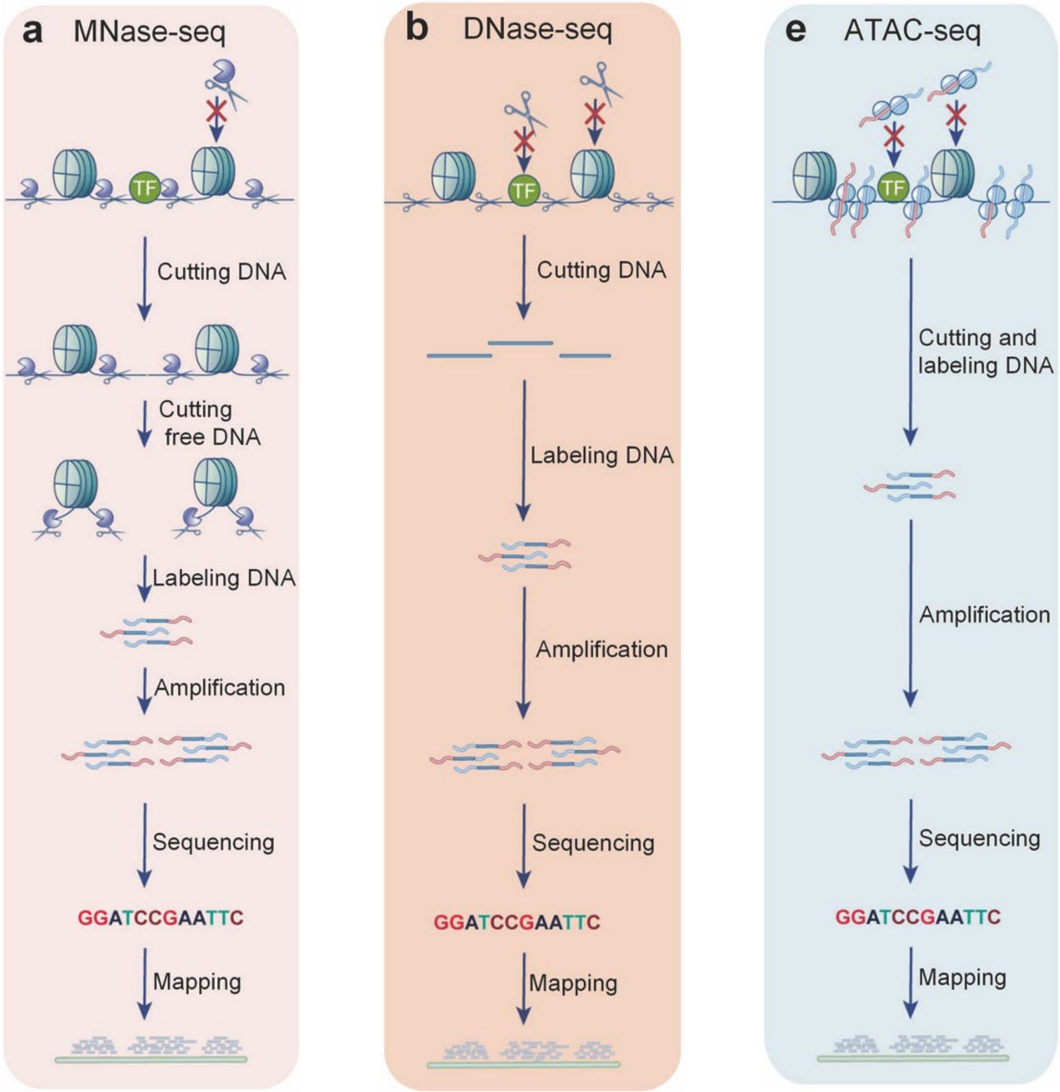
小组成员：刘晓涵 周逸飏 胡婧苑 魏欣怡

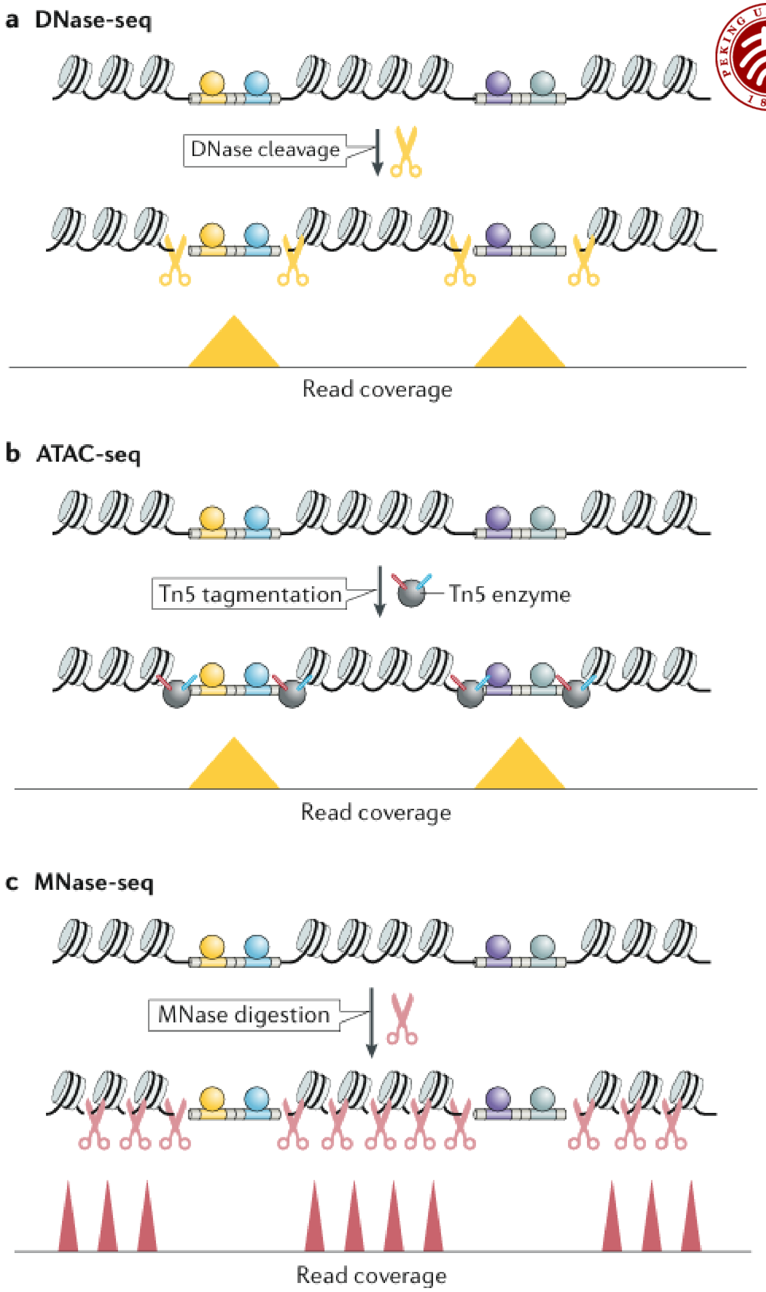
2025/6/11

染色质可及性


Chromosome
Nucleosome
Euchromatin
Transcription
Replication
Damage repair
...
Nucleus
Heterochromatin
DNA binding factor
Chromatin accessibility

染色质可及性检测方法



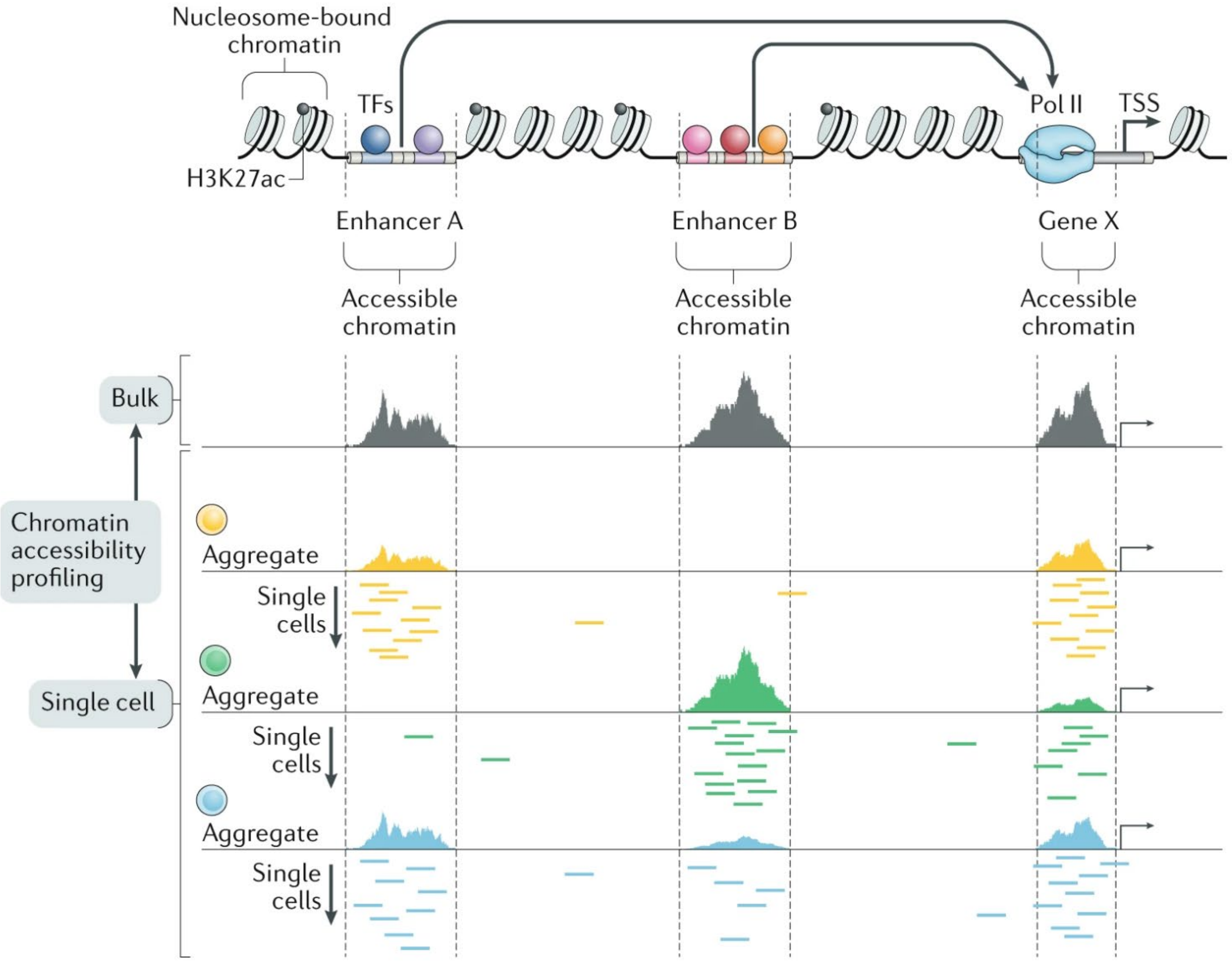


ATAC-seq

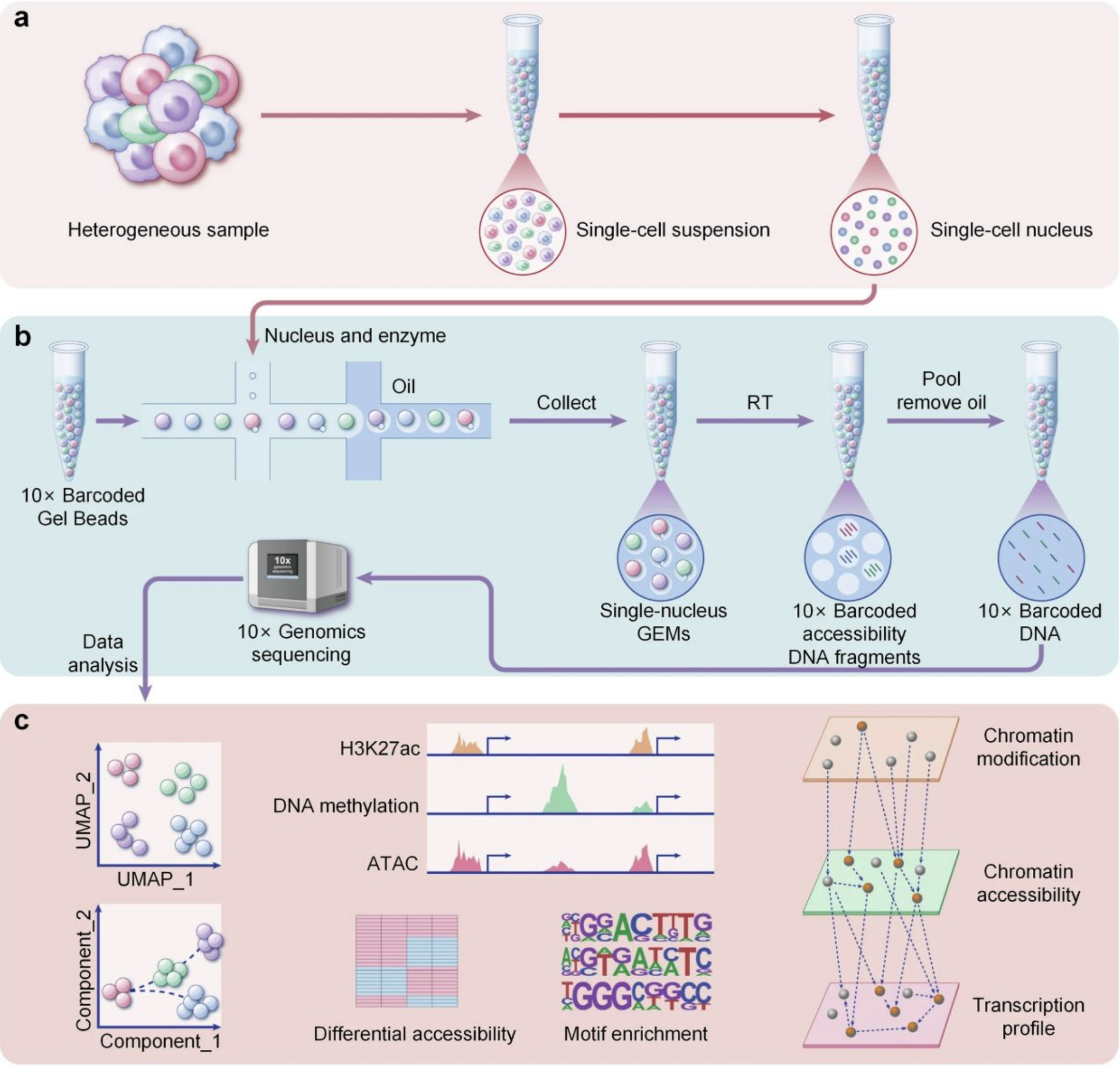
- ✓ 流程短
- ✓ 消耗细胞少
- ✓ 样品破坏小
- ✓ 灵敏且直接

→最常用方法

单细胞分析

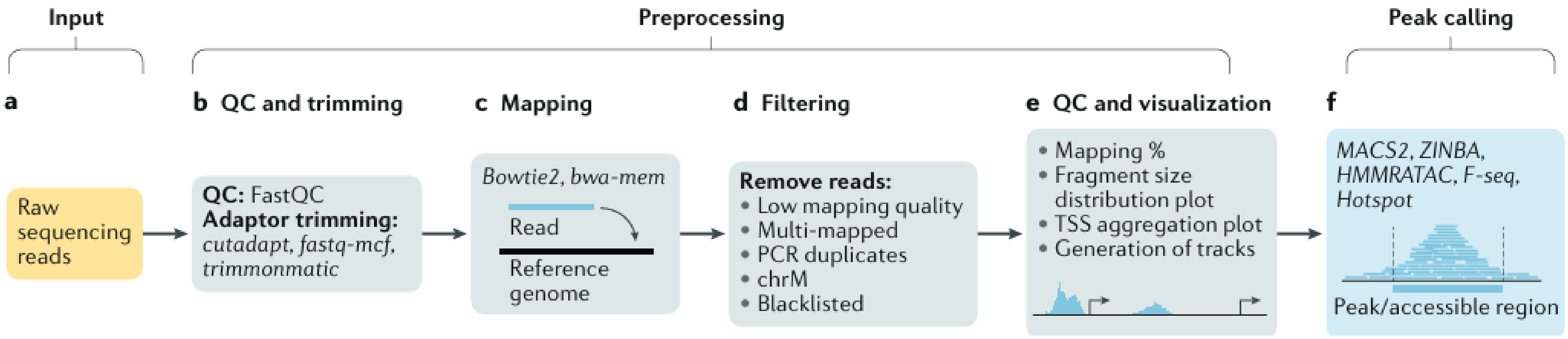


Nucleosome-bound chromatin
TFs
H3K27ac
Pol II
TSS
Enhancer A
Enhancer B
Gene X
Accessible chromatin
Accessible chromatin
Accessible chromatin
Bulk
Chromatin accessibility profiling
Single cell
Aggregate
Single cells
Aggregate
Single cells
Aggregate
Single cells



a
Heterogeneous sample
Single-cell suspension
Single-cell nucleus
b
Nucleus and enzyme
Oil
Collect
RT
Pool remove oil
10× Barcoded Gel Beads
Single-nucleus GEMs
10× Barcoded accessibility DNA fragments
10× Barcoded DNA
10× Genomics sequencing
Data analysis
c
UMAP_2
UMAP_1
Component_2
Component_1
H3K27ac
DNA methylation
ATAC
Differential accessibility
Motif enrichment
Chromatin modification
Chromatin accessibility
Transcription profile

数据处理



Input

Preprocessing

Peak calling

a

Raw sequencing reads

b QC and trimming

QC: FastQC
Adaptor trimming: *cutadapt, fastq-mcf, trimmonmatic*

c Mapping

Bowtie2, bwa-mem

d Filtering

Remove reads:

- Low mapping quality
- Multi-mapped
- PCR duplicates
- chrM
- Blacklisted

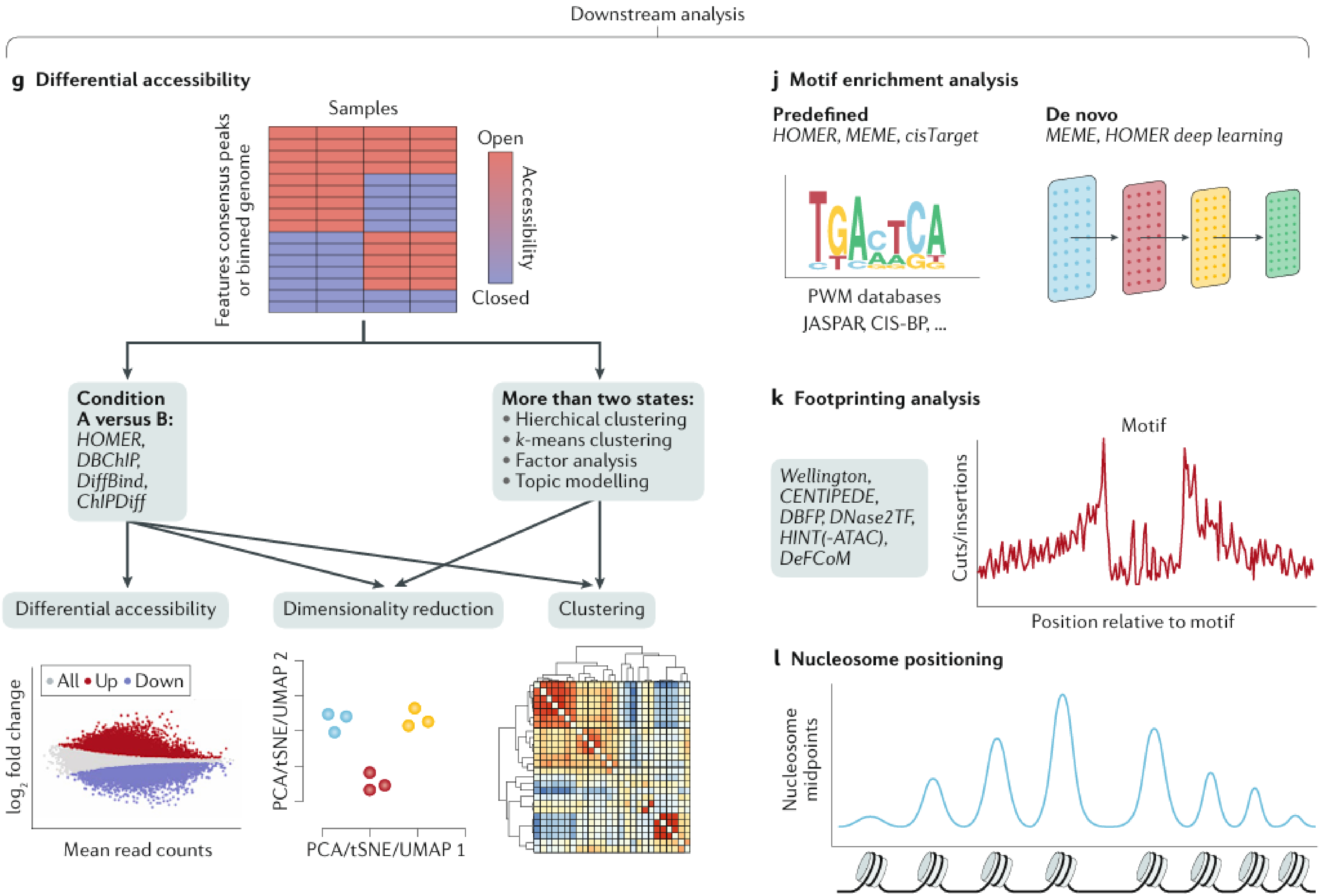
e QC and visualization

- Mapping %
- Fragment size distribution plot
- TSS aggregation plot
- Generation of tracks

f

MACS2, ZINBA, HMMRATAC, F-seq, Hotspot

下游分析


Downstream analysis
g Differential accessibility
Samples
Features consensus peaks or binned genome
Open
Accessibility
Closed
Condition A versus B:
HOMER,
DBChIP,
DiffBind,
ChIPDiff
More than two states:
• Hierchical clustering
• k-means clustering
• Factor analysis
• Topic modelling
Differential accessibility
Dimensionality reduction
Clustering
• All • Up • Down
$\log_2$ fold change
Mean read counts
PCA/tSNE/UMAP 2
PCA/tSNE/UMAP 1
j Motif enrichment analysis
Predefined
HOMER, MEME, cisTarget
De novo
MEME, HOMER deep learning
PWM databases
JASPAR, CIS-BP, ...
k Footprinting analysis
Motif
Wellington,
CENTIPEDE,
DBFP, DNase2TF,
HINT(-ATAC),
DeFCoM
Cuts/insertions
Position relative to motif
l Nucleosome positioning
Nucleosome midpoints

archR包简介

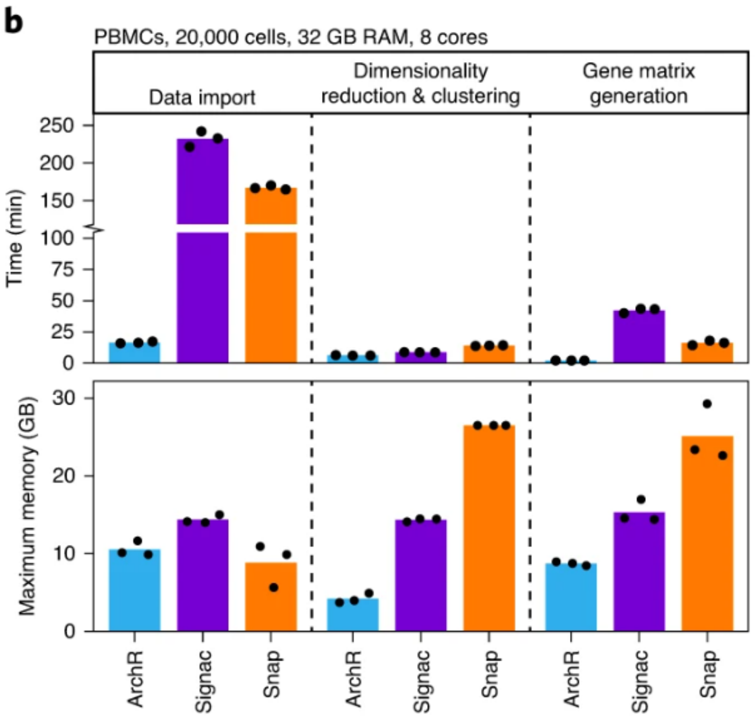
ArchR

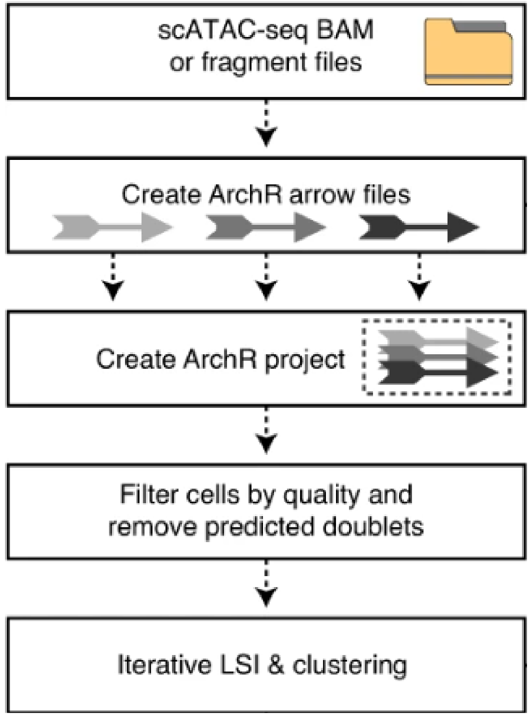
Analysis of regulatory chromatin in R

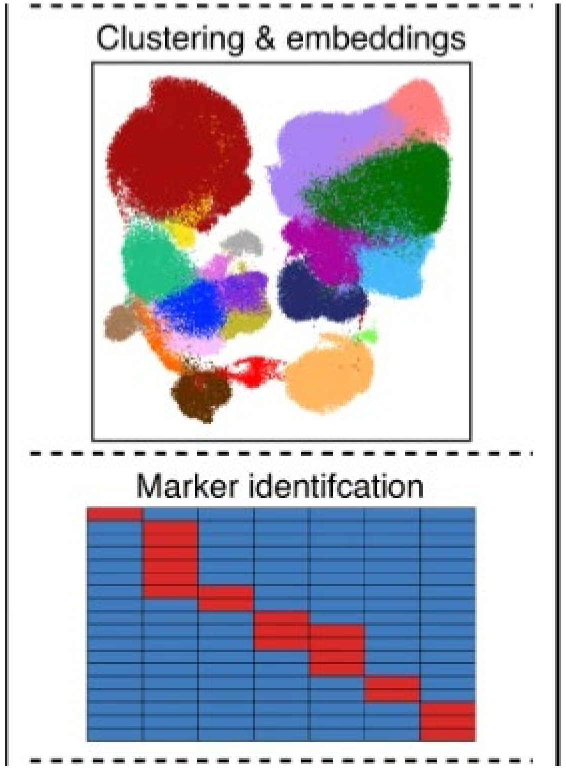
◆ 用来处理与分析scATAC-seq数据的程序

◆ 由 Howard Y. Chang 与 William J. Greenleaf 研发

◆ 功能齐全，使用广泛在速度与效率上表现出色







输入文件类型

- BAM 文件

二进制化的 tabix 排序文件，

包含每个 scATAC-seq 片段、原始序列、细胞条形码 ID 和其他信息

- Fragment 文件

tabix 排序的文本文件，

包含每个 scATAC-seq 片段和相应的单元 ID，每行一个片段

使用的输入格式取决于使用的具体程序

环境配置与安装

- Linux（不支持windows）

首先，如果尚未安装 devtools（用于安装 GitHub 包）：

```
if (!requireNamespace("devtools", quietly = TRUE)) install.packages("devtools")
```

然后，安装 BiocManager（用于安装 bioconductor 包）（如果尚未安装）：

```
if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager")
```

然后，安装 ArchR：

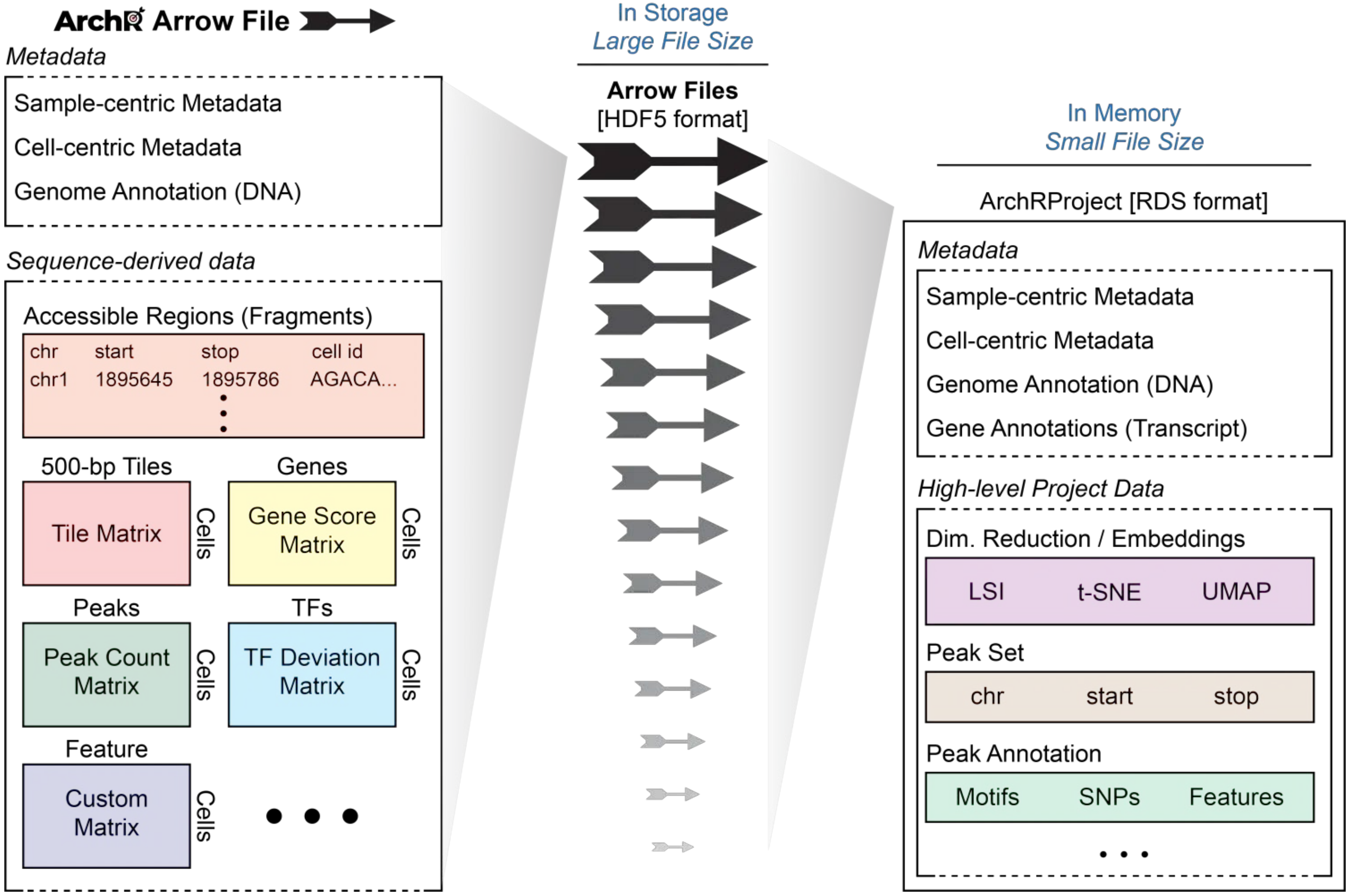
```
devtools::install_github("GreenleafLab/ArchR", ref="master", repos = BiocManager::repositories())
```

最后，安装默认情况下未安装的所有 ArchR 依赖项：

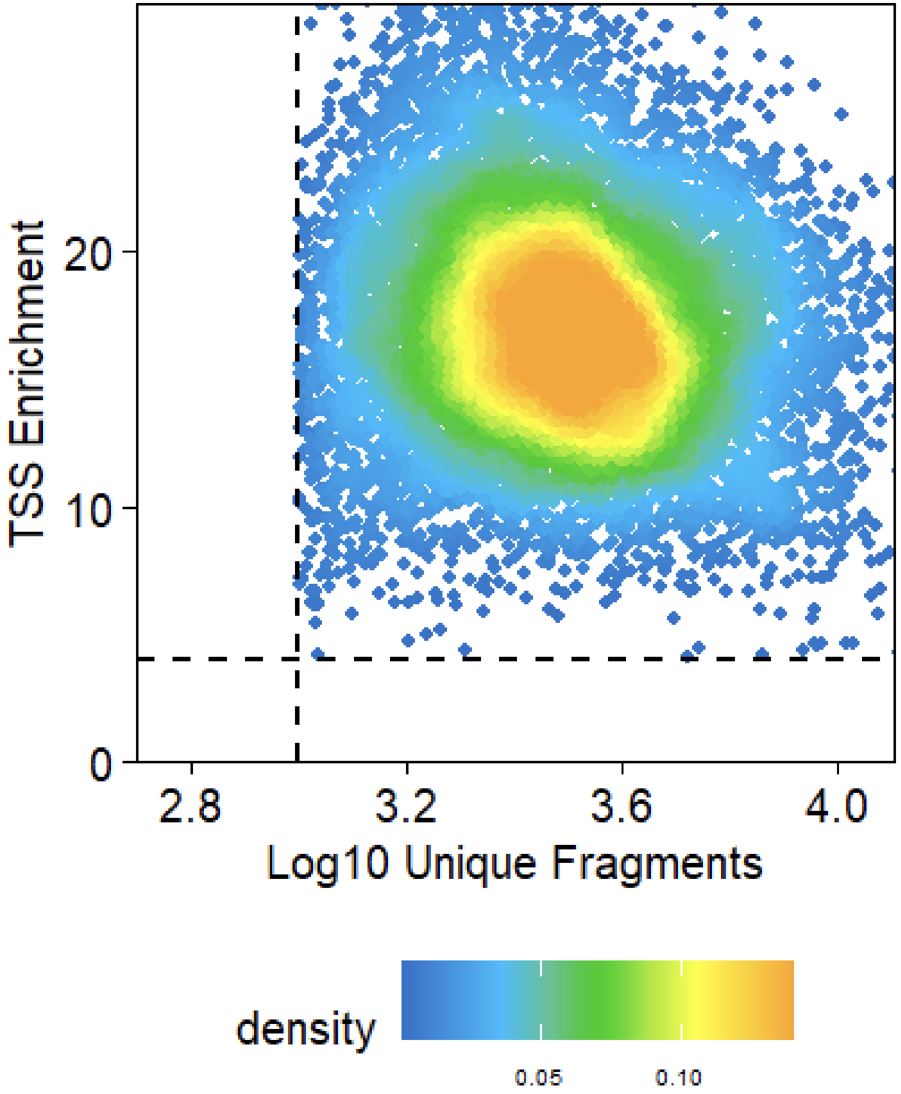
```
library(ArchR)
ArchR::installExtraPackages()
```

- Windows可用：本地下载

Arrow file / ArchRProject



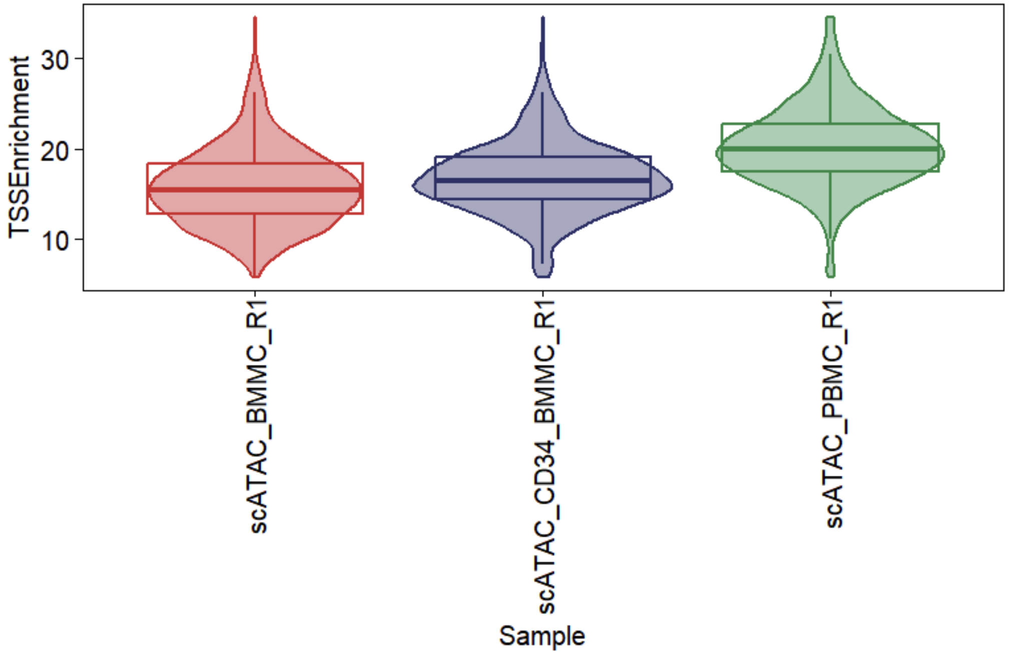
绘制QC指标



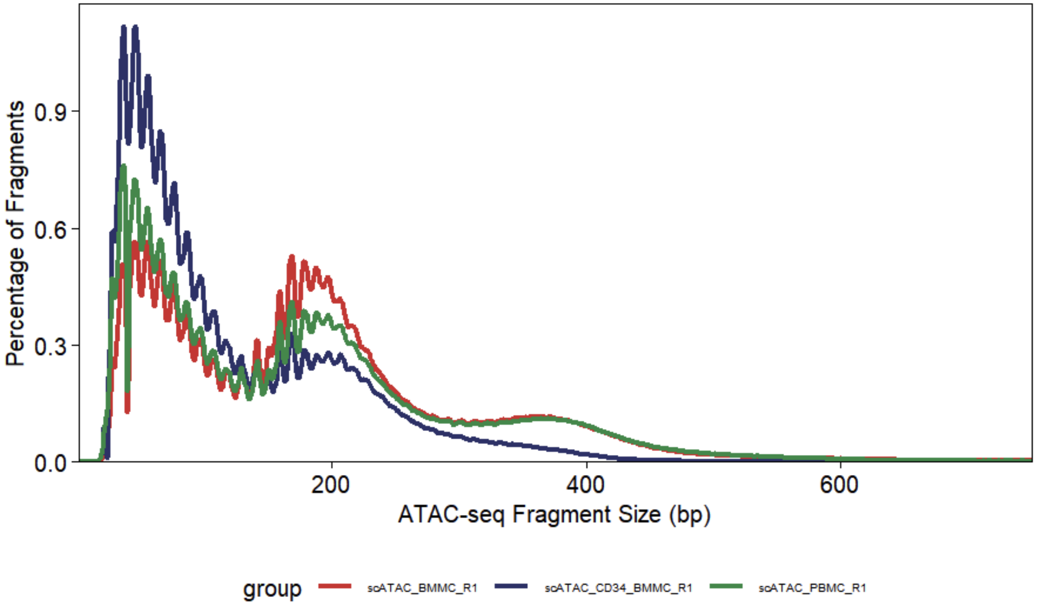
TSS Enrichment
20
10
0
2.8
3.2
3.6
4.0
Log10 Unique Fragments
density
0.05
0.10

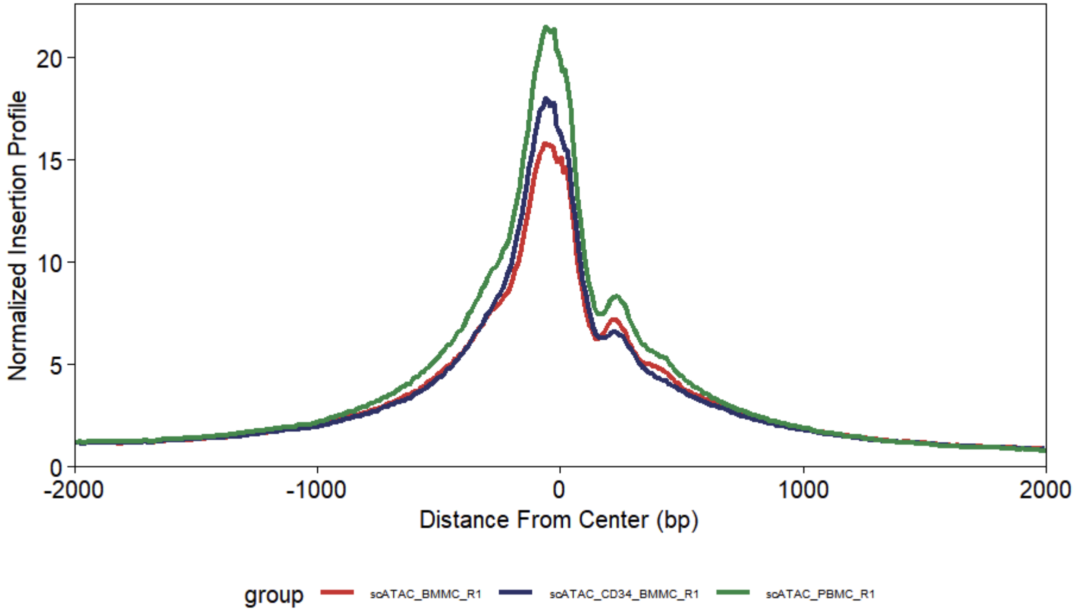
绘制样本统计量

- 小提琴图

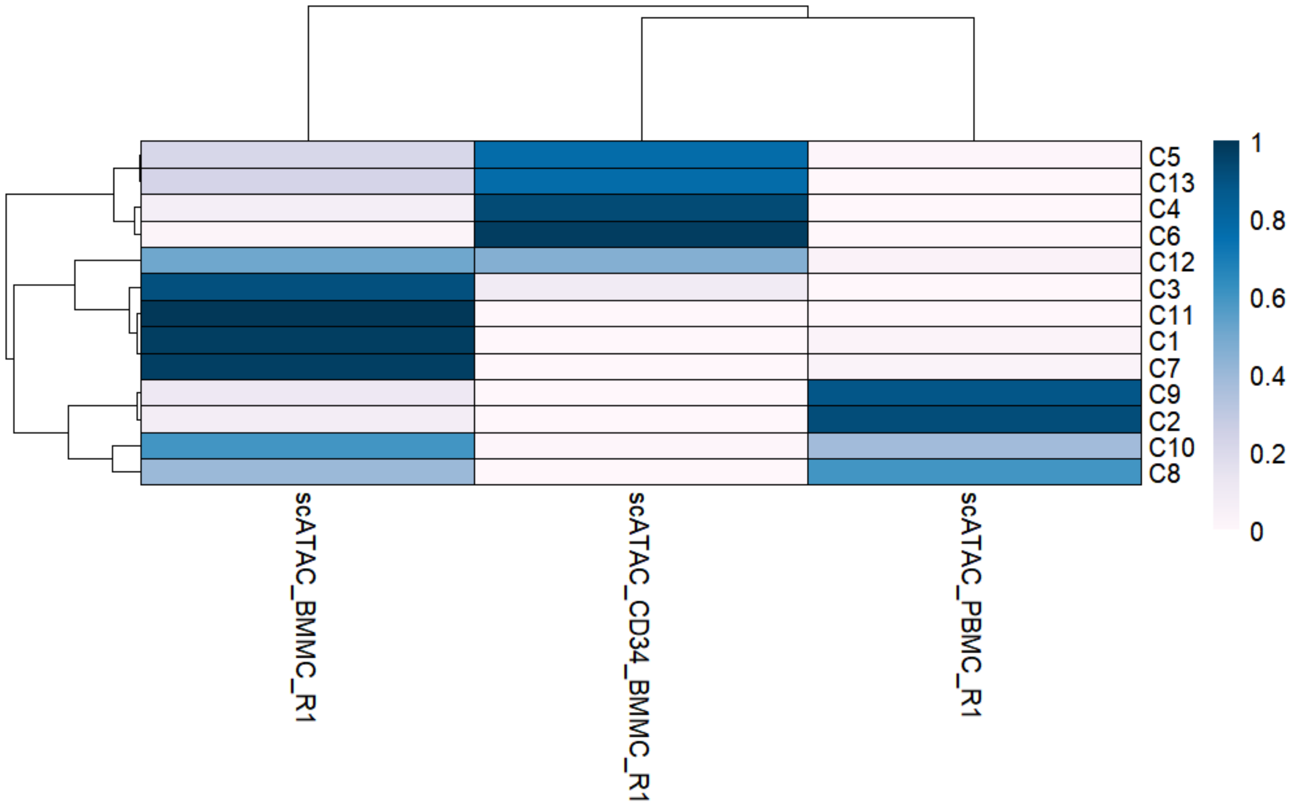


绘制样品片段大小分布和 TSS 富集曲线





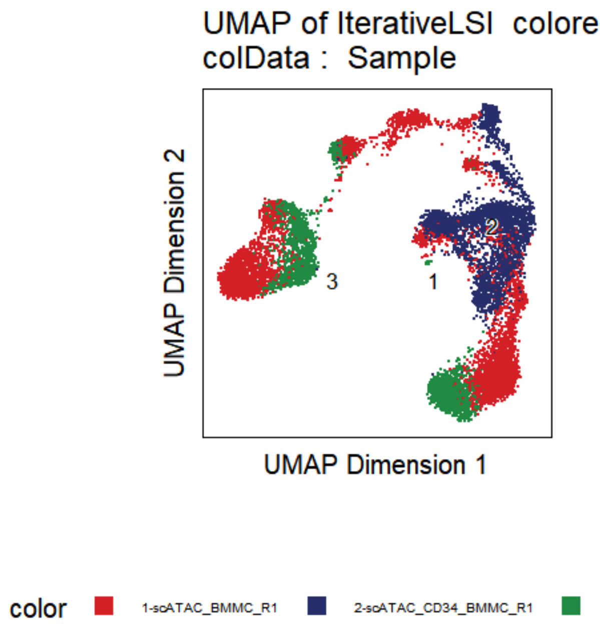
聚类

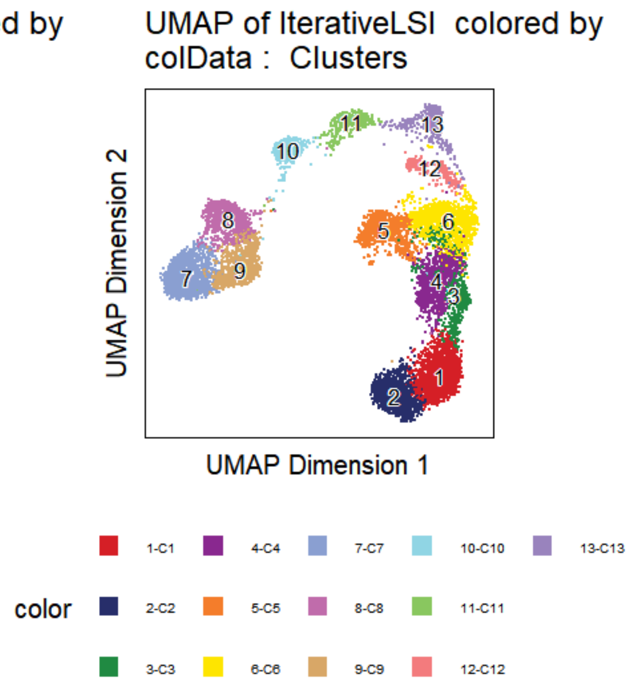


C5
C13
C4
C6
C12
C3
C11
C1
C7
C9
C2
C10
C8
scATAC_BMMC_R1
scATAC_CD34_BMMC_R1
scATAC_PBMC_R1
1
0.8
0.6
0.4
0.2
0

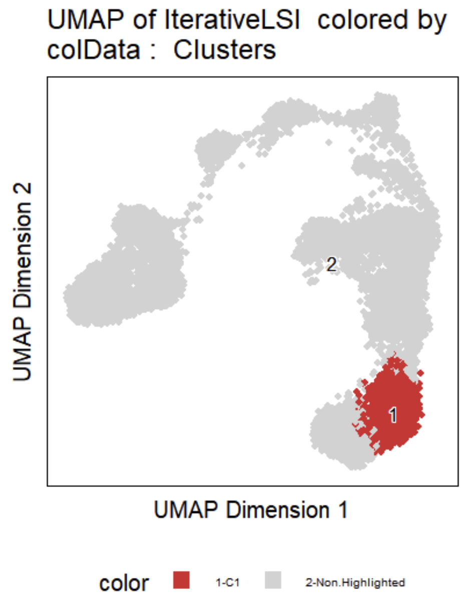
单细胞嵌入

- 运行UMAP
- 绘制UMAP结果（按不同方式着色





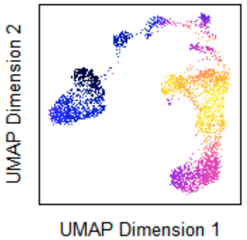
- 高亮显示特定细胞



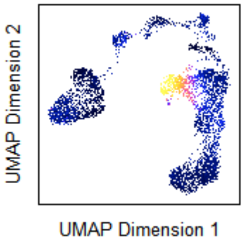
基因评分与标记基因

- 可视化标记基因

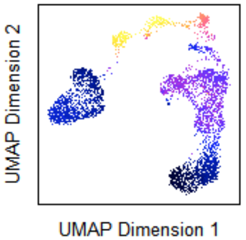
UMAP of IterativeLSI colored by
GeneScoreMatrix : CD34



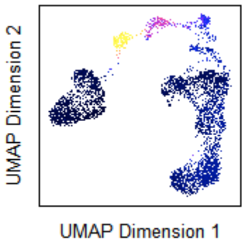
UMAP of IterativeLSI colored by
GeneScoreMatrix : GATA1



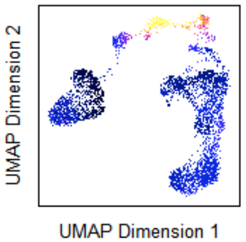
UMAP of IterativeLSI colored by
GeneScoreMatrix : PAX5



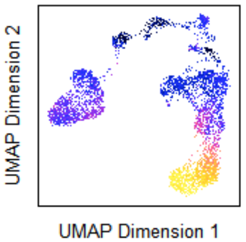
UMAP of IterativeLSI colored by
GeneScoreMatrix : MS4A1



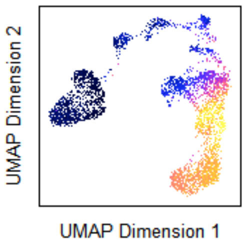
UMAP of IterativeLSI colored by
GeneScoreMatrix : MME



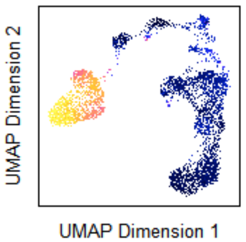
UMAP of IterativeLSI colored by
GeneScoreMatrix : CD14



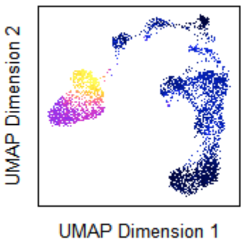
UMAP of IterativeLSI colored by
GeneScoreMatrix : MPO



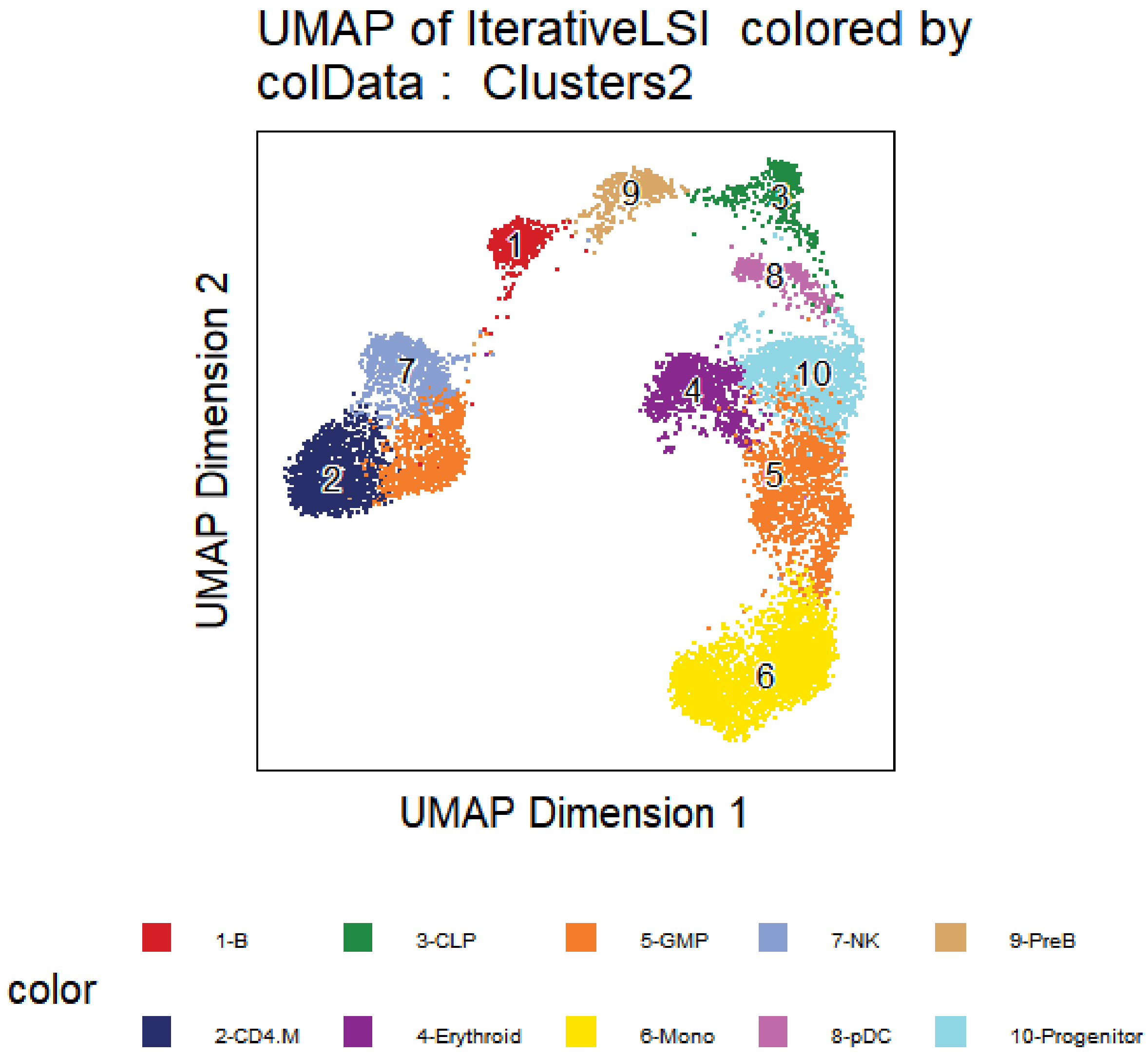
UMAP of IterativeLSI colored by
GeneScoreMatrix : CD3D



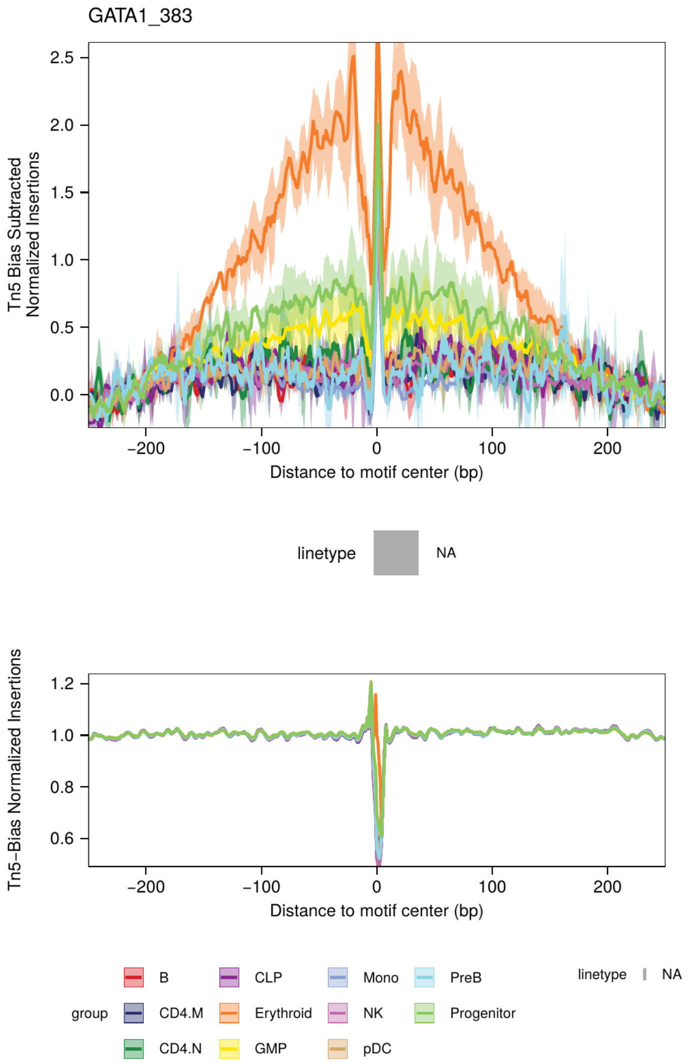
UMAP of IterativeLSI colored by
GeneScoreMatrix : CD8A



定义集群身份


UMAP of IterativeLSI colored by colData : Clusters2
UMAP Dimension 1
UMAP Dimension 2
1
2
3
4
5
6
7
8
9
10
color
1-B
2-CD4.M
3-CLP
4-Erythroid
5-GMP
6-Mono
7-NK
8-pDC
9-PreB
10-Progenitor

转录因子足迹


GATA1_383
Tn5 Bias Subtracted Normalized Insertions
2.5
2.0
1.5
1.0
0.5
0.0
−200
−100
0
100
200
Distance to motif center (bp)
linetype
NA
Tn5−Bias Normalized Insertions
1.2
1.0
0.8
0.6
−200
−100
0
100
200
Distance to motif center (bp)
group
B
CLP
Mono
PreB
CD4.M
Erythroid
NK
Progenitor
CD4.N
GMP
pDC
linetype
NA

参考资料

1. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

2. Chen, Y. *et al.* Chromatin accessibility: biological functions, molecular mechanisms and therapeutic application. *Sig Transduct Target Ther* **9**, 340 (2024).

3. Minnoye, L. *et al.* Chromatin accessibility profiling methods. *Nat Rev Methods Primers* **1**, 10 (2021).

4. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207–220 (2019).

5. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411 (2021).

6. Arch R官方网站操作手册 https://www.archrproject.com/bookdown/index.html

感谢观看

LEB G09

小组成员：刘晓涵 周逸飏 胡婧苑 魏欣怡

2025/6/11