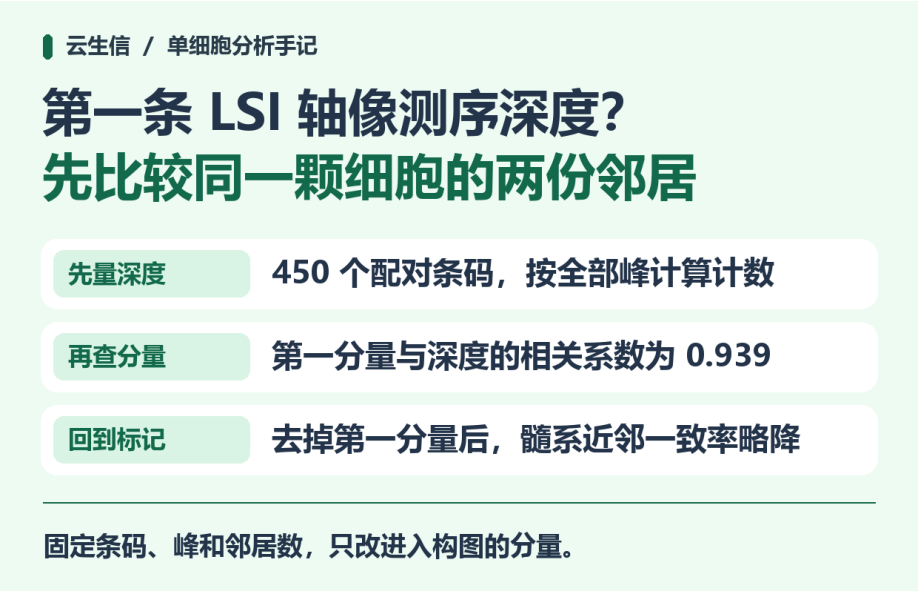
ATAC 降维后按测序深度排成一条带，第一个 LSI 分量该不该用?

把 450 颗真实配对多组学细胞的第一分量拿进、拿出近邻图，逐细胞核对深度和 RNA 标记。



运行教学示例

安装 Python 3.10 或以上及 requirements.txt。示例已含稀疏矩阵。从原始 HDF5 重做截取还需 h5py。

```
python run.py --input example --out results
```

在解压后的资料包根目录运行。example 是教学输入，results 保存本次输出，expected 保存可供核对的参考结果。

输入字段

peaks.npz、peak_index.tsv 450 个细胞与 2,500 个实测峰，矩阵行为细胞，列为峰。

cells.tsv、source.json 九个 RNA 标记、全峰深度、取样方式及来源文件校验值。

核对输出，再换自己的数据

component_depth.tsv、marker_groups.tsv 十五个分量的深度相关、三组标记一致率。

edges.tsv、cell_neighbors.tsv、components_depth_and_rna.png 两张逐条码近邻表及比较图。

用示例验算

1 第一分量对全峰计数的 Spearman 相关约 0.93937，第二分量约 0.08257。

2 两种模型各有 9,000 条定向邻居记录。

3 B 样标签的平均近邻一致率从 0.822 升至约 0.839，髓系样从约 0.867 降至约 0.859。

换用真实数据

保留完整峰深度、细胞顺序和独立标记，固定邻居数后替换分量。用多供者项目分别验证。

本例限定三类 RNA 标记清楚的细胞，频见峰面板和标签组成都会影响相关性与图结构。

方法资料

Signac 配对 PBMC 数据与流程

Signac 的深度相关诊断