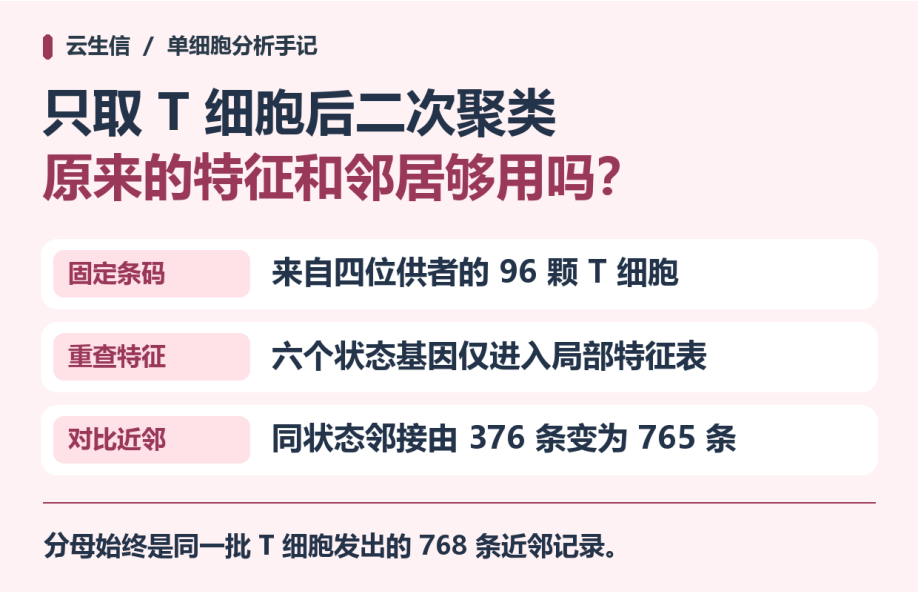
单独取出 T 细胞后二次聚类，还能沿用全图的高变基因和邻居吗?

固定同一批 T 细胞，分别继承原图的边、在原坐标重找邻居、在 T 细胞内重选特征并重算 PCA 和 Leiden，再从原表达核对状态。



运行教学示例

示例实测 Python 3.12，需 numpy、scipy、scikit-learn、matplotlib、igraph、leidenalg。安装包内 requirements.txt 后运行。

```
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

counts/、cells.tsv 原始表达、条码、供者、谱系和教学参考状态。参考状态只用于计算完成后的检查。

settings.json 目标谱系、变异特征数、均值分箱数、主成分和近邻数，外加 Leiden 分辨率与标记面板。

核对输出，再换自己的数据

features.tsv、cell_audit.tsv、edges.tsv 全局和局部入选特征、逐条码邻居组成和三种视图的完整边表。

clusters.tsv、cluster_composition.tsv、state_neighbors.tsv 实际 Leiden 标签、各簇的供者与参考状态组成，及按供者计算的同状态邻接比例。

raw_marker_check.tsv、subset_pca.png 以原始归一化表达复核每位供者的标记与检出率，同批 T 细胞的全局和局部投影。

用示例验算

1 同一批 96 颗 T 细胞，两种全局路线都得到 376/768 条同状态邻接，局部路线为 765/768。

2 CCR7、IL7R、TCF7、NKG7、GZMB、PRF1 均进入局部前 25 个特征，全局前 25 个未包含它们。

3 同一设置下局部 Leiden 得到两个各 48 颗的群；D1_T_01 在两套各八个的邻居中只保留一个。

换用真实数据

先核查谱系门槛和双细胞，再固定目标条码。在子集中按同一种有说明的特征选择方法建新图，保留原来的全局位置和供者信息。

教学的均值分箱与方差排序是透明的简化方法。模拟状态不代表真实患者亚型，给新群命名需独立表型和重复供者支持。

方法资料

Bioconductor 的子聚类与对照方法

Bioconductor 的高变特征选择