肿瘤里增殖细胞特别多，细胞周期该不该回归掉？

在相同条码和特征上各做一次 PCA 与近邻计算，用谱系、周期和缺氧样上皮细胞三个读数决定一项肿瘤研究要保留哪种视图。



运行教学示例

Python 3.10 或以上，需 numpy、scipy、scikit-learn、matplotlib。解压后先运行 python -m pip install -r requirements.txt。

```
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/ 与 cells.tsv 原始基因计数和独立模拟标签，包括供者、谱系、相位和缺氧样状态。

settings.json 列出用于打分的基因、六个主成分、每个条码的八个邻居。

核对输出，再换自己的数据

raw_scores.tsv、pc_scores.tsv 原始对数表达上的标记均值，以及两套模型中各个细胞的主成分得分。

graph_summary.tsv、cell_audit.tsv 按供者和目标细胞检查邻居构成，逐个条码查八个近邻保留了几个。

neighbor_edges.tsv、cycle_comparison.png 每条近邻边及两行三列的原始与残差投影。

用示例验算

1 全体 144 个细胞中，平均同谱系比例从 1.000 到约 0.999，同相位由约 0.999 到 0.348。

2 18 个缺氧样细胞各有八个邻居。同为缺氧样细胞的邻居比例由 0.611 升至 0.903。

3 D1_01 的八个邻居保留两个；同相位邻居从八个变为两个，缺氧样邻居从五个变为八个。

换用真实数据

改用真实样本的原始计数与配套条码。随组织与物种调整周期面板，独立核对谱系与目标状态，记录供者内的变化。

本例的短标记集得分是教学用均值。回归只能消除所拟合的线性方向，不能从与相位完全重合的状态中识别独立效应。

方法资料

Seurat 细胞周期打分与回归实践

Bioconductor 单细胞降维