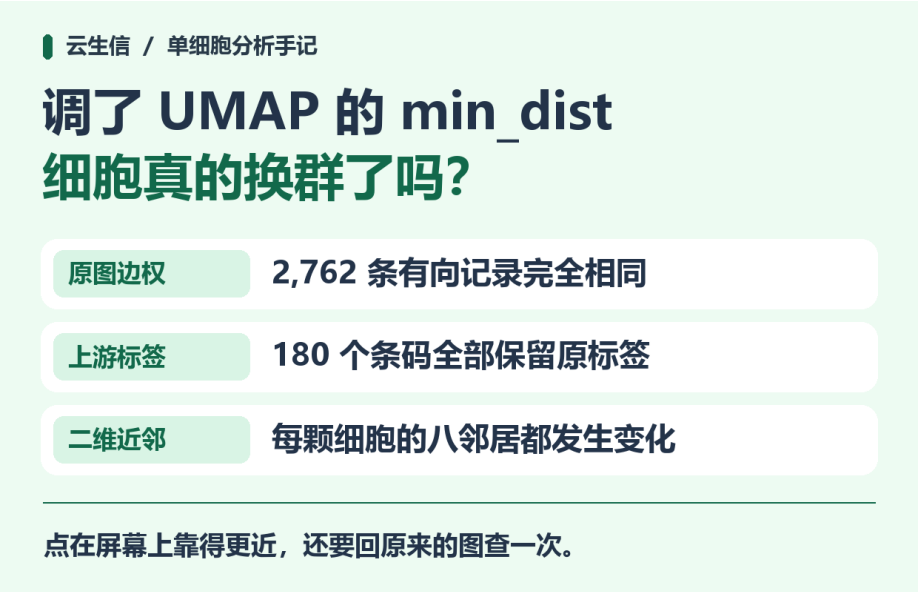
调了 UMAP 的 min.dist，两群离得更远，细胞的分群也变了吗?

固定八维主成分、近邻图、标签和随机种子，实际拟合两套 UMAP，沿着同一条码核对高维边权、二维近邻与标签。



运行教学示例

示例实测 Python 3.12、umap-learn 0.6.0。推荐新建 Python 3.12 环境后运行 python -m pip install -r requirements.txt；首次运行需编译。

```
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

pcs.tsv、labels.tsv 按条码连接的八维主成分和上游固定标签，两个输入的行序有意不同。

settings.json 两个 min_dist 值、固定随机种子、建图邻居参数和二维查找邻居数。

核对输出，再换自己的数据

graph_edges.tsv、barcode_audit.tsv 每一条原图边在两次拟合中的权重，及每个条码的标签、图上度数和二维邻居交集。

coordinates.tsv、map_neighbors.tsv、layouts.png 两套坐标、按二维距离排名的邻居名单、按同一图例绘制的对照图。

用示例验算

1 180 个条码，两次拟合的 2,762 条非零有向图记录完全一致，标签变化数为零。

2 cell_001 原图有 12 个非零近邻，两张二维图各取八个最近邻时重合五个。

3 配套版本下每颗细胞至少换了一个二维近邻，八个位置中重合数的中位数为四。

换用真实数据

导出项目最终使用的主成分和已有标签，用全局唯一条码连接。若目标是检查项目聚类的原图，还要分别保存该聚类实际用的图与权重。

模拟数据的二维名单与近邻数、实现版本和运行环境有关。本例构建的是 UMAP 自己的模糊邻接图，外部聚类标签保持固定。

方法资料

UMAP 参数和 min_dist

UMAP 使用预计算近邻