新增一位患者，怎样沿用旧图谱的坐标和注释，再验证映射靠不靠谱?

在三位参考供者上固定 PCA、UMAP 和标签，再将第四位已知标签供者作为盲测集，核对错分、拒判和参考图谱外的状态，最后映射第五位。



运行教学示例

示例实测 Python 3.12、umap-learn 0.6.0，推荐独立环境安装 requirements.txt。首次 UMAP 拟合和映射有编译开销。

```
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 UMI 矩阵、互不重叠的参考、留出、查询供者；只有留出集拥有用于盲测的 truth 列。

settings.json PCA 维数、邻居数、投票门槛、参考内部跨供者距离分位数和种子。

核对输出，再换自己的数据

assignments.tsv、validation_metrics.tsv 逐条码票数、邻居距离、拒判原因和标签；留出供者各类的分母、获标签数、正确和误判数。

validation_confusion.tsv、两张坐标表、reference_mapping.png 真实与预测标签的交叉表，固定参考和新条码坐标，以及留出与查询的分面图。

用示例验算

1 三位参考供者共 180 颗；留出 66 颗，第五位查询也有 66 颗。训练完毕后参考坐标逐元素未变。

2 留出供者 B 和髓系各 20 颗全部正确，T 的 20 颗里有 3 颗九票全投 B。

3 六颗参考未收录状态里五颗拒判，一颗错误接收为 T；不能用投票比例解释为真实正确率。

换用真实数据

换用相同物种、相同特征定义并按旧模型顺序对齐的原始 RNA 矩阵。整位供者留出，先在已知标签上检查误差和拒判，再处理没有真值的新患者。

练习采用标准化、PCA、近邻分类及 UMAP transform 构成的透明参考法。单一留出供者与六颗未知状态的数字不能代表临床准确率。

方法资料

UMAP 投影新数据的官方实例

Seurat 参考图谱映射教程