样本标签 HTO 显示 singlet，为什么表达谱还像两个细胞?

将 HTO 标签和 RNA 双细胞分类按实际条码配对，核对原始标签信号、同行谱系证据以及缺失分母，给相互矛盾的条码安排下一项复核。



运行教学示例

Python 3.10 或以上，需 numpy、scipy。示例演练用已给出的两类分类结果。另附已完成 HTO 分类的 Seurat v5 对象导出函数。

```
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

hto_calls.tsv、hto_counts.tsv 按条码给出 HTO 分类和各标签原始分子数，标签缺失的 RNA 条码不填入。

rna/、rna_calls.tsv、markers.tsv 10x RNA 稀疏矩阵、每个条码的实际捕获和 RNA 评分分类、谱系标记 ID。

核对输出，再换自己的数据

cross_table.tsv 捕获内的 HTO 与 RNA 分类交叉表，附每行的有效分母及 RNA 双细胞比例。

evidence.tsv、hto_singlet_rna_doublet.tsv 每颗条码的标签信号、RNA 评分与谱系分子数，同时单列单标签但 RNA 可疑的条码。

用示例验算

1 16 个 RNA 条码，14 个有 HTO 输入，2 个单列为 missing。

2 7 个 HTO Singlet 中有 2 个 RNA doublet；另有 2 个 HTO Doublet 被 RNA 分类为 singlet。

3 C06 的两个最高标签为 92 和 2 个分子，T 与 B 两组标记同时分别有 18 和 13 个分子。

换用真实数据

用同一物理捕获的 HTO 原始标签计数和分类结果、RNA 双细胞评分和原始基因表达矩阵替换示例。条码须能唯一标识捕获内的一颗液滴，跨捕获时对两个通道使用同样的前缀。

原始标签最高与次高差值帮助读图，不能代替 HTODemux 的逐标签背景校准。模拟例子的分类和分数是教学输入；脚本不将两种信号自动合成绝对真值。

方法资料

Seurat HTO 拆样与标签分类

Cell Hashing 原始研究