用 CellBender 去完背景，怎样检查真实表达有没有一起被减掉?

用同一批细胞分别核对应该保留与应该减弱的标记，读出 CellBender 前后的检出率、分子数及条码覆盖，并筛出值得回看模型报告的细胞群。



运行教学示例

Python 3.10 或以上，需 h5py、numpy、scipy、matplotlib。用 pip 安装 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

raw.h5、corrected.h5 同一捕获的原始与校正矩阵，采用带 matrix 和 features 的 10x 风格 HDF5 结构。

labels.tsv、panel.tsv 原先的条码、独立谱系标签及按谱系定义的正向标记、应低表达标记。

核对输出，再换自己的数据

coverage.tsv、paired_counts.tsv 核对原注释条码是否仍在校正矩阵，并记录共同条码的逐基因原始与校正计数。

markers.tsv、marker-detection.png 按谱系列出同分母检出数、分子总数、失去检出的条码数，以及前后检出比例图。

用示例验算

1 11 个有标签条码中有 10 个在两边出现；缺少的 T5 单列。

2 T 群共同条码上 HBB 检出 3→0，CD3D 检出 4→4，CD3D 总量 25→24。

3 B 群的 MS4A1 总量从 12 降到 5，另一个 B 标记 CD79A 总量维持 16。

换用真实数据

从同一次试验取原始 Cell Ranger RNA H5 与 CellBender 输出的校正 H5，填入该批细胞已有的谱系注释以及控制标记 ID。保留模型的 HTML 报告、软件版本和 FPR 设置，按文库独立检查。

教学数据是按实际 H5 结构构造的前后计数，不含模型训练。真实标记明显下降意味着需要复查；单凭计数变化无法确定每个被减分子的真实来源。

方法资料

CellBender 使用与输出解释

10x HDF5 基因表达矩阵结构