某位患者的样本质控后只剩一成，究竟是哪一步删掉的?

用过滤前的条码和最终交付名单重建质控损失，每个规则分别统计原始命中、与其他规则的重叠和按既定顺序首次剔除的人数，再按谱系检查稀有群是否集中丢失。



运行教学示例

Python 3.10 或以上，需 matplotlib。解压后在脚本目录运行 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

before.tsv、after.tsv 按样本和条码保存过滤前元数据与下游实际保留名单。

rules.json 检出基因下限、线粒体百分比上限和实际过滤顺序；边界以正文与脚本定义为准。

核对输出，再换自己的数据

rules.tsv、overlap.tsv、steps.tsv 各规则在同一起点的命中、重叠组合、每一步首次剔除和保留数。

barcodes.tsv、lineage.tsv、reconcile.tsv 逐条码的去向、各谱系的起止数，以及缺证据或与实际交付不符的条码。

用示例验算

1 P1 有 10 颗起始细胞，三项各命中 5、5、4 次，实际有 9 个不同条码未通过。

2 原顺序中各步首次剔除 5、3、1，最终保留 1 颗。

3 P2 的 BC10 缺少线粒体记录，进入差异清单，仍可看到它是否出现在交付名单。

换用真实数据

从每份文库导出过滤前的同一细胞全集、每个决策时点的原始证据以及实际最终名单，确认门槛单位、边界和顺序。对照表若还存在差异，应追查当时使用的数据版本和隐含过滤步骤。

示例重新应用已经固定的条码级判据。若真实流程会在删除细胞后重算指标、标记模型或聚类，更换步骤顺序还可能改变最终名单，需要复跑相应流程。

方法资料

Bioconductor 单细胞质控与多指标判断