多份样本一起跑 scDblFinder，应该按患者拆，还是按上机捕获批次拆?

把供者、冻存管和实际液滴捕获连接成表，按捕获运行 scDblFinder，再回到单个可疑条码和两套谱系标记，检查哪些组合在实验中真的有机会同滴。



运行教学示例

Python 3.10 以上，需 scipy、numpy。包内有 R 脚本和实跑分类结果；重跑模型需 Matrix、SingleCellExperiment、BiocParallel、scDblFinder。

```
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

capture_manifest.tsv、cells.tsv 文库装载记录和细胞注释，供者身份与捕获分两列保存。

counts/、markers.tsv、settings.tsv、calls.tsv RNA 稀疏原始计数、谱系标记、模型设置和 scDblFinder 所得条码分类。

核对输出，再换自己的数据

donor_pairs.tsv、capture_counts.tsv 每次捕获可能的跨供者组合、实际条码数和被标记的双细胞数。

flagged.tsv、marker_combinations.tsv 带评分和原始谱系标记分子数的候选条码及按捕获统计的标记组合。

用示例验算

1 480 颗模拟条码分两次独立捕获，各 240 颗。D1 分别与 D2、D3 共用捕获。

2 示例实际模型在每次捕获标出 25 颗，其中每次有 1 颗缺少两组达到教学阈值的谱系标记。

3 CAP_A_BC216 同时有 45 个 B 系和 48 个 T 系标记分子，两组均达到练习门槛。

换用真实数据

按同一物理捕获定义移交 capture 列，准备过滤后的真实细胞计数与唯一条码。在真实数据上按化学体系和回收规模核实预期双细胞率，修改设置后重跑 R 脚本，再在每一捕获内汇总。

示例含可核对的双谱系合成细胞，实际项目往往缺少完整真值。图表中的标记组合是审查线索，不是 scDblFinder 给出的亲本来源预测。

方法资料

scDblFinder 捕获分组与随机种子说明

scDblFinder 作者的双细胞检测教程