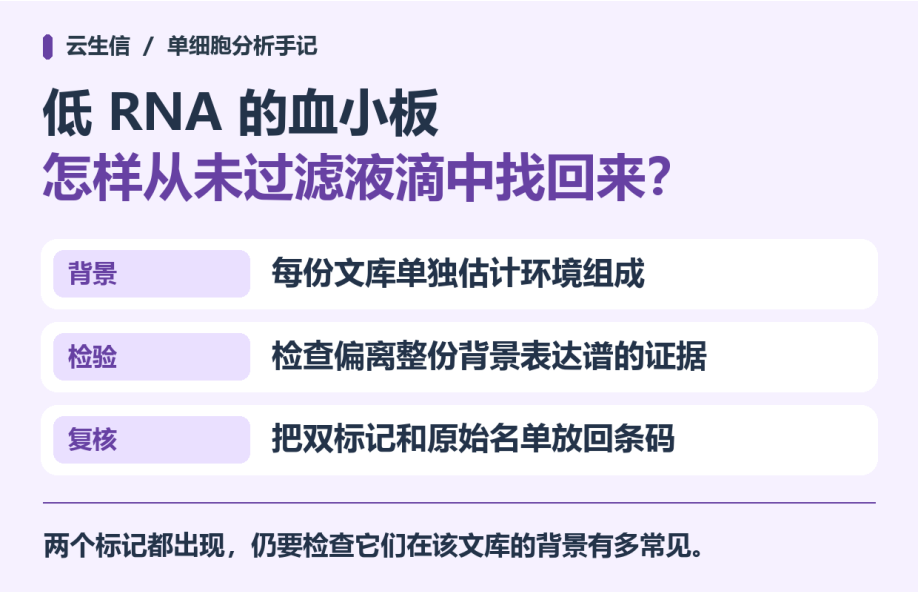
血小板本来就只有少量 RNA，怎样从空液滴里把候选细胞找回来?

从两份未过滤液滴矩阵分文库估计背景，重读 EmptyDrops 的整谱检验，再到低计数条码逐颗核查 PF4、PPBP 和原始细胞名单。



运行教学示例

复现主练习用 Python 3.10 以上、numpy 和 scipy。包内附实跑的 EmptyDrops 表。重算统计检验另需 R、DropletUtils、Matrix 和 BiocParallel。

```
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

L1/、L2/ 各含 raw/ 的 10x MEX、上游已保留条码 called.txt 和 EmptyDrops 的逐条码检验表。

runs.tsv 每文库的背景上限、自动保留下限、模拟次数、随机种子、复查范围、FDR 与两个标记 ID。

核对输出，再换自己的数据

background.tsv 列出候选背景条码数、分子总量、两个标记计数及合并比例。

low_barcodes.tsv、review_counts.tsv 按条码保存两个标记、FDR、Limited 及复核判断，并按文库汇总。

用示例验算

1 两份文库各有 15 个处于复查范围且未在原始名单的条码，其中各有 7 个同时满足标记与整谱检验。

2 两标记在候选背景中的合计为 L1 的 16/875 和 L2 的 100/889。

3 只满足两个标记条件的条码数在 L1 为 1 个，在 L2 为 6 个。

换用真实数据

替换成每一份文库原始 RNA 矩阵与该文库原先识别的条码，按实际参考写入标记 ID。选择合适的背景与检验参数，用 rerun_emptydrops.R 重新生成逐条码结果，最后运行主脚本汇总。

统计检验评估与环境 RNA 谱的偏离程度，不能认定细胞类型、排除聚集体或确认完整血小板。教学数据固定阈值用于手算；真实文库的细胞识别和背景边界需要分别诊断。

方法资料

EmptyDrops 原始研究与血小板结果

DropletUtils 统计检验教程