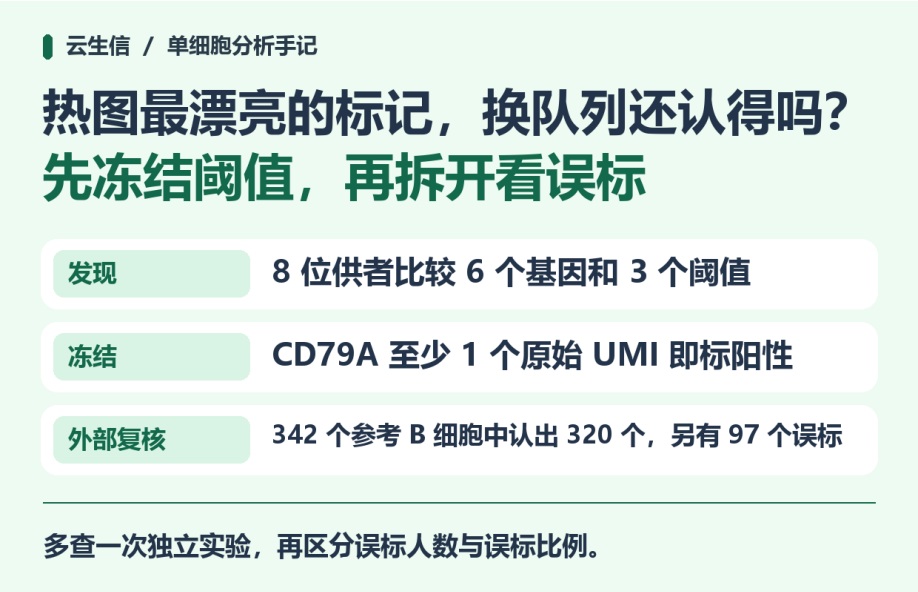
热图里“最漂亮”的 marker，换一批患者还能识别同一种细胞吗?

在八位 Kang 供者中选择 B 细胞标记与原始 UMI 阈值，先写出冻结规则，再打开独立 PBMC3K 数据核验误标。



运行教学示例

Python 3.10 或以上，先安装 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

discovery.tsv.gz、validation.tsv.gz 两个独立来源的实际标记原计数、条码、参考标签和深度。

核对输出，再换自己的数据

discovery_grid.tsv、frozen_rule.json 只用发现队列形成的供者等权排序与冻结规则。

by_donor.tsv、errors_by_type.tsv、frozen_marker.png 两批数据中的混淆计数、逐谱系误标和检出图。

用示例验算

1 冻结规则为 CD79A 原始 UMI 至少 1。

2 外部参考 B 细胞有 342 个，其中正确识别 320 个，漏掉 22 个。

3 外部非 B 细胞误标 97 个，55 个来自 CD4 T 参考组。

换用真实数据

明确发现与验证的供者边界、特征命名和原始计数口径。打开验证队列前冻结目标、阈值与报告指标。

外部队列仅一名供者；参考标签来自表达数据，无法代替独立蛋白或组织学真值。

方法资料

Kang 等的多供者 PBMC 研究

Scanpy PBMC3K 数据和参考流程