样本标签可能贴反了，怎样用单细胞数据找到值得复查的那一管?

用装载清单、HTO 记录、独立供者基因型和逐条码等位基因分子数排查疑似错管，追踪哪些证据相互独立，并为覆盖不足的细胞保留未定状态。



运行教学示例

主脚本只需 Python 3.10 以上。可选的 cellSNP-lite 稀疏矩阵转换脚本另需 numpy 和 scipy。

```
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

manifest.tsv、barcodes.tsv 每管登记供者、实际条码和 HTO 供者，未分配 HTO 的留空。

donors.vcf、alleles.tsv、settings.tsv 配套供者的双等位 SNP 基因型与各条码 REF/ALT 的 UMI 数、深度和证据门槛。

核对输出，再换自己的数据

tube_summary.tsv、evidence_by_tube.tsv 分管记录支持各供者的人数、证据不足人数，以及登记与 HTO 的交叉核对。

cell_identity.tsv、unused_loci.tsv 每颗条码的可用位点、分子数、最佳候选、证据差和未成功匹配的位点。

用示例验算

1 三份管各有 7 个条码；每管 6 个达到覆盖门槛，余下 1 个为 insufficient。

2 TubeA 登记 D1，6 颗的基因型支持 D2；TubeB 的 6 颗反向支持 D1。TubeC 的支持与登记一致。

3 TubeA 的 BC01 有 24 个等位基因 UMI、6 个位点，第一和第二候选的自然对数似然差约 16.15。

换用真实数据

从独立登记且坐标版本一致的供者 VCF 取得基因型，从对应 RNA 条码的 SNP pileup 导出等位基因 UMI。可用 from_cellsnp.py 读取 cellSNP-lite 的 AD 与 DP 稀疏矩阵。对疑似错管，应回到原始装载和对照样本复核。

此例是教学用单供者模型，没有估计双细胞、环境 RNA、群体相关或等位特异表达。似然差是模型内部的相对证据，不是患者身份后验概率；证据不足时保留未定。

方法资料

demuxlet 的供者匹配研究

Vireo 已知和未知供者的输入说明