两个相邻簇的前十个 marker 几乎相同，怎样找出真正区分它们的基因?

让两个相邻 T 细胞簇先对照同一背景，再直接互比，结合检出率、表达效应和逐位供者的证据筛选更能区分二者的标记。



运行教学示例

Python 3.10 或以上，需 numpy、scipy 和 matplotlib。先在解压目录运行 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

counts/ 与 cells.tsv 有条码的模拟原始表达矩阵、供者和簇标签，A 与 B 为目标，M 为统一参照。

settings.json 比较方向、最小效应、最小检出率差与每位供者每侧最少细胞数。

核对输出，再换自己的数据

against_background.tsv、pair_candidates.tsv 对共同背景的两套排名、目标簇之间的效应和检出率、合并入选与同向通过的供者数。

within_donor.tsv、donor_effects.png 逐位供者的细胞分母、效应和可比情况，及合并与分供者的效应图。

用示例验算

1 两个簇各对 M 排名时前十的基因集合完全相同。直接比较后，CCR7、IL7R、NKG7、GZMB 在四位供者中均同向过线。

2 ISG15 合并效应约为 2.72，四位供者与合并同向通过的数量为零。第三位供者反方向达到阈值。

3 LINGER 仅在第一位供者同向过线，宜另列单样本观察。

换用真实数据

将原始计数、供者和聚类条码对齐，设定两个目标群及选择该背景的理由。按真实项目覆盖情况设置描述阈值，并保留每位供者的分母。

这些是描述性候选。教程的均值比有明确伪计数，不能直接等同其他软件默认的倍数变化，也不能把单细胞数作为独立患者重复。

方法资料

Seurat 的两组标记比较

Bioconductor 的成对标记检验