某个基因在治疗组更高，是更多细胞表达，还是阳性细胞表达得更强?

从公开 RNA 原始计数重算 STAT1、B2M 和 IFIT1 的阳性比例及阳性者表达，把总体均值还原为两部分，再按供者匹配测序深度。



运行教学示例

Python 3.10 或以上及 requirements.txt。默认从三个基因的公开计数小文件直接运行，无需联网。

```
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

monocyte_genes.tsv 5,697 个公开条码的供者、条件、完整库深度，以及 STAT1、B2M、IFIT1 原始计数。

extract_public.py、source.json 从公开 H5AD 重建小表的代码与原始文件校验值。

核对输出，再换自己的数据

paired_metrics.tsv、pooled_descriptive.tsv 逐人及合并的检出数、检出比例、阳性均值、全体均值和深度。

depth_matched_ids.tsv、paired_two_part.png、positive_distributions.png 同供者按深度分箱匹配的条码，两部分逐人图，以及阳性细胞的累积分布。

用示例验算

1 STAT1 对照 485/2,932、刺激 1,277/2,765 检出。

2 B2M 两侧全检出，阳性均值约 273.93 与 345.63 CP10k。

3 深度匹配后每侧 2,311 颗；IFIT1 有一位供者的对照侧零阳性，阳性均值留空。

换用真实数据

在同一细胞类型内使用原始计数识别检出，用完整库深度归一化强度，保留供者、条件和无阳性样本，按实际问题检查深度及其他偏差。

检出率受捕获及库深度影响。合并细胞的均值是描述量，临床组间推断要在生物学重复层面建模。

方法资料

Kang 等真实干扰素刺激实验

MAST 两部分建模原始论文