分组后显著差异主要来自一个人，怎样把它从火山图里找出来?

对八位真实供者的单核细胞依次留出一位，重做八次配对 edgeR 模型，在火山图上标记效应最容易被某位患者拉动的基因。



运行教学示例

Python 3.10 或以上，安装 requirements.txt。若重跑全部九个计数模型，另需 R、edgeR 和 statmod。

```
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

pseudobulk.tsv.gz、samples.tsv 15,706 个基因，16 个样本的原始加和及每份的供者、细胞数和 UMI。

fit_influence.R、example/fit/ 真实 edgeR 九次拟合、全队列 TMM 对数 CPM 与版本。

核对输出，再换自己的数据

influence_all.tsv、volcano_influence.png 固定的 3,709 个基因，八次留一的最大效应变化、最敏感供者及火山图。

descriptive_intervals.tsv、donor_differences.tsv、donor_influence.png 逐人表达差、其均值的描述性 t 区间以及模型效应。

用示例验算

1 八位供者、三千七百零九个基因，每个基因有完整拟合及八次留出。

2 拿掉 1015，RWDD4 效应约 −0.67→−0.25，FDR 约 0.035→0.387。

3 ISG15 八次留出效应约 7.37 至 7.92；去掉 1015 后 KIAA0232 效应约 −1.38。

换用真实数据

按真正独立的供者留出其全部关联样本，重新拟合同一设计，固定候选基因范围，再对影响最大的供者回查取材、深度和原始样本。

这是探索性敏感性分析。逐人差的描述性区间与 edgeR 条件系数基于不同的量；九次拟合也不构成九个独立验证队列。

方法资料

Kang 等的原始刺激实验

edgeR 原始计数建模及手册