# 同一套干扰素基因拿去算 UCell，为什么一批样本的分数特别低?

从八名供者抽取 960 颗真实单核细胞，用真正的 UCell 分别检查基因名、原始深度、基因集删减与供者内评分。




## 运行教学示例

Python 3.10 或以上与 requirements.txt 可复核随包结果。重算排名和降深需 R、Matrix 和 UCell；此次实际拟合用 UCell 2.6.1。

```
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.mtx.gz、cells.tsv、genes.txt 960 颗真实细胞、一万五千七百零六个基因；逐条码附供者、条件与原库深度。

fit_ucell.R、example/fit/ 八个干扰素相关基因、受控降深、四基因缩写以及三套真实 UCell 分。

## 核对输出，再换自己的数据

identifier_audit.tsv、donor_scores.tsv 基因名匹配情况，各供者原始及降深后的中位深度、检出面板数和评分。

condition_summary.tsv、donor_ucell.png 两类条件的三个评分场景，逐供者折线。

## 用示例验算

1 八位供者各有对照和刺激各六十颗，共 960 颗。

2 原矩阵刺激侧中位分约 0.852；同样细胞降深后约 0.691，改用四基因约 0.730。

3 对照侧面板中位检出数从 1 到 0，刺激侧从 7 到 6。

## 换用真实数据

先检查物种与稳定标识，再确认同一表达层、基因排序背景和面板版本；将供者内分数与库深度、逐基因检出和独立条件一起解释。

本例降深是 UMI 计数的受控抽薄；无法重现重新建库或改变测序平台造成的所有影响。分数不能直接读成功能活性或响应概率。

## 方法资料

UCell 原始论文

作者软件及不同版本的归一化说明

Kang 等的原始刺激实验