# 热图上两个基因都是深红色，能说它们的表达量一样高吗？

用同一组原始计数绘制两张热图，逐格核对群平均表达、基因内标准化和检出比例，读懂两块同色方格到底代表什么。




## 运行教学示例

Python 3.10 或以上，需 numpy、scipy 与 matplotlib。安装 requirements.txt 后在解压目录运行。

```
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/、cells.tsv 九颗模拟细胞，带原始计数、群名和条码。归一化分母来自全部基因。

settings.json 展示的五个教学基因、三个细胞群及统一的每万分子系数。

## 核对输出，再换自己的数据

heatmap_values.tsv、per_cell.tsv 每格的线性平均、两种取对数方式、行标准分、检出率和细胞分母，以及逐细胞值。

paired_heatmaps.png 一份数据使用共同的表达量色标和逐基因标准分色标，用灰色标出无方差的行。

## 用示例验算

1 C 组的 GENE_HIGH 平均为 4,000，GENE_LOW 为 100，二者行标准分均约为 1.336。

2 FLAT 在三个群的均值同为 150，行内标准差为零，所以三格均标 NA。

3 SPARSE 在 B 组三个细胞中一个检出，先平均后取对数约 4.900，先取对数再平均约 1.998。

## 换用真实数据

先注明表达来自哪个数据层、计数分母、每群有哪些细胞，再决定是否按基因缩放。保留实际数值和每群检出比例；需要比较条件时按供者分别汇总。

本例用三个群均值等权，标准差分母为群数。不同软件在是否再缩放、方差分母和显示截断上的设置会改变数字。

## 方法资料

Seurat 对分组平均表达的说明

R 对中心化与缩放的定义