# 富集分析换了背景基因，第一名通路为什么变了？

固定同一份四十基因上调名单，逐通路手算超几何检验，检查全基因组与真正进入比较的检测背景为何互换第一名。




## 运行教学示例

Python 3.10 或以上，安装 requirements.txt；SciPy 使用 1.11 或以上。

```
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

de_and_universe.tsv.gz、annotations.tsv.gz 模拟全基因组、进入检验标记、上调名单及固定的通路成员。

## 核对输出，再换自己的数据

enrichment.tsv 逐通路的四个整数、重叠基因、倍数和校正结果。

background_rank_change.png 同一候选名单在两种背景下的名次变化。

## 用示例验算

1 全基因组背景下 readily_detected 排第一，富集倍数为 35。

2 受检背景下 compact_response 排第一，富集倍数约 11.36。

3 更换背景前后，两条通路的命中数始终为 14 和 10。

## 换用真实数据

保存实际参与检验的基因、显著名单、物种和同版注释；对每条通路先与对应背景取交集。

是固定名单的过度代表检验；与使用全基因排序的富集方法所问的问题不同。

## 方法资料

Gene Ontology 富集背景说明

小基因名单功能富集的统计背景