一个转录因子表达不高，它的调控网络就一定不活跃吗？

用 GRNBoost2、pySCENIC 的基序筛选和 AUCell 依次追踪同一条网络，再找出转录因子计数为零、靶基因集得分却较高的具体条码。



运行教学示例

常规核验用 Python 3.10 或以上与 requirements.txt。重拟合用 model-environment.yml 创建独立环境，并运行 fit_scenic.py。

```
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.tsv、cells.tsv、make_example.py 256 个模拟条码，320 个基因，四个供者，生成过程与模拟真值。

toy_motif_ranks.tsv、toy_motif_annotations.tsv、fit_scenic.py 103 个模拟基序的排序、注释以及三步实际拟合脚本。

核对输出，再换自己的数据

target_history.tsv、barcode_evidence.tsv 逐候选的筛选去留、每个条码的因子 UMI 和 AUCell。

zero_umi_upper_auc.tsv、donors.tsv、tf_and_regulon.png 异常组合清单、分供者统计、同条码散点图。

用示例验算

1 STAT1 候选共 49 项，基序筛选后有 23 个靶基因。

2 256 个条码中 174 个 STAT1 为零，64 个处于所得 AUC 排名前四分之一。

3 两种条件均满足的 10 个条码分布于四位供者，分别为 2、1、3、4 个。

换用真实数据

真实数据须换成相符物种与基因组版本的 cisTarget 排序库和 motif2TF 注释，按供者、细胞类型与库深度复核候选，再用独立结合或扰动实验验证。

示例靶基因和基序排名人为设定。样本内分位数方便定位条码，不能视为具有通用意义的调控活性阈值。

方法资料

Aibar 等 SCENIC 原始论文

pySCENIC 作者教程