一个样本的中位基因数很低，再测深一点有用吗？

把同一文库的中位检出基因和测序饱和度放到共同读深坐标，计算观测区间内的边际收益，找出哪份样本值得进一步评估补测。



运行教学示例

Python 3.10 以上，需 matplotlib。

```
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

depth.tsv 每行一个文库在一个降采样读深上的结果，含平均 reads/细胞、中位检出基因和 0 到 1 的饱和度。

settings.json 比较区间的每细胞平均读数上下限；只允许位于各文库观测范围内。

核对输出，再换自己的数据

intervals.tsv 每段读深增加带来的基因增量、每千 reads 增益和饱和度变化。

matched_depth.tsv / curves.png 在共同读深区间比较文库，双面板显示原始观测点。

用示例验算

1 在 1 万到 2 万 reads/细胞间，A 增加 80 个中位基因，即每千 reads 增加 8 个。

2 相同区间 B 增加 480 个，对应每千 reads 增加 48 个。

3 A 与 B 的终点位于已测曲线上；程序拒绝超出观测范围的区间。

换用真实数据

从每份同一实验类型的 Cell Ranger Web Summary 所报告的降采样点整理两个指标，记录软件版本、样本类型、细胞识别口径与读数单位。比较时选各文库都覆盖的区间，区间终点必须在已测深度以内。

每千 reads 增益基于曲线点之间的有限差分，非观测点使用分段线性插值。组织来源、细胞回收和基础 RNA 含量仍需独立解释，曲线没有测到的深度无法由本脚本预测。

方法资料

Cell Ranger 报告和两张降采样曲线