复现公开肺癌图谱，怎样看出“转移灶特征”其实由取材部位驱动？

从 Kim 等肺腺癌图谱的原始样本工作簿与逐细胞注释重建 58 份标本的设计表，核查可比器官、病灶与配对患者。



运行教学示例

Python 3.10 或以上，先安装 requirements.txt。

```
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

samples.tsv、celltype_counts.tsv 原作者工作簿的患者、样本来源、报告细胞数与原注释按类型计数。

extract_samples.py 从 GEO 两个原始补充文件重建教学表的脚本。

核对输出，再换自己的数据

sample_design.tsv、organ_lesion.tsv、contrast_support.tsv 逐标本设计、器官病灶交叉表及真实配对患者名单。

paired_lung_b_fraction.tsv、design_and_paired_lung.png 限定同器官及患者的示范组成图。

用示例验算

1 58 份标本的汇总细胞数与作者逐细胞表完全相等，共 208,506 颗。

2 早期肺原发和脑转移在此数据中配对患者为零；同器官的正常与转移淋巴结也无同人配对。

3 早期肺肿瘤与远端正常肺有 10 位配对患者，其中 9 位肿瘤样本的 B 细胞注释比例较高。

换用真实数据

患者与器官必须来自原作者的标本表，不从条码末尾数字猜测。先检查交叉分层，再决定能估计哪个对比。

示范的是组成与可辨识性，没有重新拟合原论文基因签名，也不能据此断言任何一个作者结果由器官驱动。

方法资料

GEO 原数据及补充样本工作簿

Kim 等原始肺腺癌图谱