把小鼠基因名改成大写就能和人的细胞整合吗?

用实际查询的 Ensembl 同源关系对齐人鼠标记，亲手计算复制一对多计数带来的膨胀，再按供者核对可比较的表达。



运行教学示例

Python 3.10 或以上，安装 requirements.txt。随包冻结同源映射，可直接离线运行。联网刷新可另用 fetch_orthology.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

orthology.tsv、orthology_source.json 实际 Ensembl 同源边、版本、查询物种与检索时间。

features.tsv、cells.tsv、counts.tsv 两物种的已测基因、全库深度、供者和教学模拟稀疏计数。

核对输出，再换自己的数据

mapping_audit.tsv、mapping_cost.tsv 每个小鼠基因的目标、决定及原因，另列不当展开造成的计数变化。

donor_markers.tsv、ortholog_markers.png 共同标记的供者和谱系检出率、平均归一化计数与热图。

用示例验算

1 十五个小鼠基因中六个对齐，七个关系复杂，一个无返回，一个对侧面板未测。

2 Cd8b1 有两个目标，源计数 257 若逐边复制便变 514。Trbc1 能按大写匹配，却属于多对多。

3 两物种第一位供者的 CD79A 在示例 B 细胞里均为 12/12 检出。

换用真实数据

先固定每个物种的注释版本，用稳定编号映射并检查两边真实测量的特征。把一对多与未测量的候选另存，留在物种内核验。

同源基因是否具有类似细胞表达还需独立样本验证。演示表达由模拟产生；本例面板不代表全基因组映射率。

方法资料

Ensembl 同源关系分类

Ensembl REST 同源查询