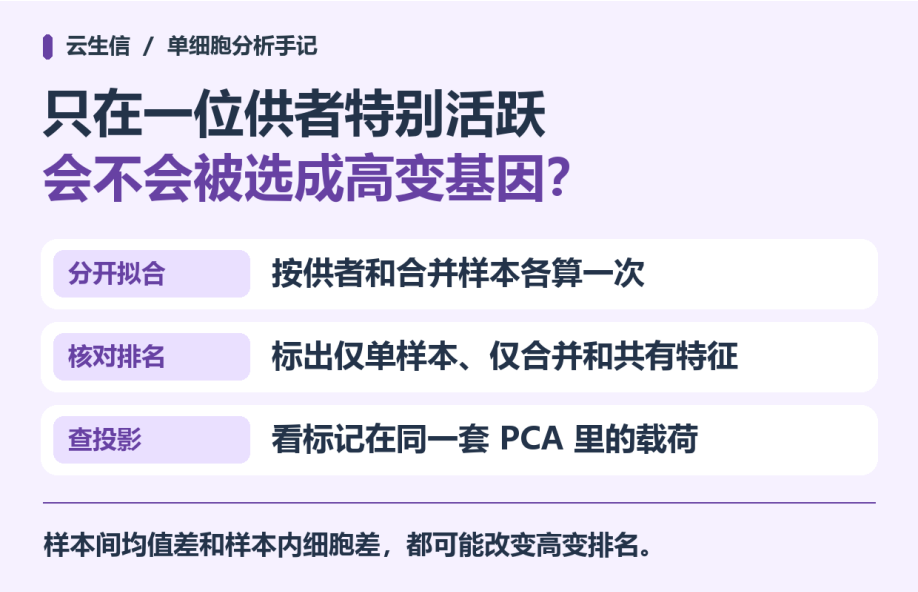
一个基因只在一位患者中特别活跃，会被选进高变基因吗？

用相同表达尺度分别拟合四位供者和合并样本的均值—方差趋势，找出由单一供者带入或被合并掩盖的高变基因，再核对它们实际参与哪些主成分。



运行教学示例

Python 3.10 或以上，需 numpy、scipy、scikit-learn。资料包含 scran 的实跑结果；重新拟合趋势需 R、Matrix 和 scran。

```
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/、cells.tsv、settings.json 原始基因表达矩阵、供者标签、所取高变数及主成分数。

variance.tsv、fit_hvg.R 四个供者和合并矩阵的 scran 方差分解结果，附可重建这张表的 R 代码和版本记录。

核对输出，再换自己的数据

feature_ranks.tsv、donor_expression.tsv 各作用范围的基因排名、入选供者数，以及入选基因逐供者的检出率和平均对数表达。

pca_loadings.tsv、cell_scores.tsv 在同一入选并集上的 PCA 平方载荷占比和各细胞的投影位置。

用示例验算

1 每个范围取 24 个基因，并集为 59 个。DONOR_SHIFT 在合并排名第 3，但在 4 位供者内部均未进入前 24。

2 DONOR_STATE 在 P1 内排名第 1，其余供者中未入选，合并时排到第 42。

3 DONOR_SHIFT 约占第二主成分平方载荷的 14.6%，计算对象为固定的 59 个特征。

换用真实数据

保持供者标签和归一化口径，按项目基因 ID 导出未被预先裁切的 RNA 计数。运行 fit_hvg.R 重估供者内及合并趋势，随后按研究目标检查入选覆盖与主成分载荷。

示例数值只对应给定计数、特征数和拟合方法。平方载荷描述标准化特征在该主成分轴中的权重，不等于该基因解释了整份数据方差的同等比例。

方法资料

OSCA 高变特征与生物学方差

OSCA 主成分与载荷