

project.

Salmon \rightarrow gene expression.

(分析基因表达)

量化大量不同的RNA转录。

优点：更快且更准。

expression levels ?

使用 maximum likelihood 来决定 基因表达水平。

(为什么) \rightarrow 这造成了 往往可以得到 转录数量 的精确估计，
但不给出 精确率。
(什么估计) \rightarrow abundances

这造成了 推测可能高度明确且准确，

或高度不准确， ?

这取决于 概率函数的形状以及优化怎么执行。

\Downarrow

因此，我们最好能 给出一个 精确率。

预测 准确性的方法很多;

一种是 bootstrapping

This approach treats the observed sample data as the population,
samples from the original data a number of
times, and reruns the maximum likelihood estimator
independently on all of these samples.

在原始数据中取样多次, 并对所有样本^{独立}返回最大
似然估计。

查看这些不同运行的分布, 可以形成一个经验的
准确率 间隔 (interval), 提供了一个关于最大
可能点估计的可能性记录。

然而, 他们倾向于低估这个不确定性。

目标是决定什么是 转录的什么具体属性倾向于脱离 预测的可能性间隔。

具体来说，如果真实值不在 95% 的间隔中，我们认为转录失败。

我们期望只有 5% 的转录落在间隔外。

step:

① filter the failing transcripts

② find some common properties

using sequence similarity - - - .

③ quality score. reflect the level of confidence.

关键点:

1. 关于 Salmon 的详细信息.

RNA-seq 测序的相关资料,

transcript abundances.