

# An overview of the Chinese version [9]

## Fisher's exact test

**Fisher's exact test** 是基于一个  $2 \times 2$  的列联表 (contingency table) 来计算两个分类变量之间的关联性。该表中, 行代表第一个分类变量, 列代表第二个分类变量, 每个单元格中的数字表示同时具有这两种属性的个体数量。在计算 Fisher's exact test 之前, 需要明确一个假设, 即两个变量是独立的, 然后计算观察到的数据和期望数据之间的差异。期望数据是基于原始数据总和计算的, 假设两个变量是独立的情况下, 每个单元格中预期的数量。接着, 基于这些差异, 通过计算可能比当前结果更极端的结果的概率, 计算出  $p$  值。 $p$  值表示拒绝原假设 (即两个变量独立) 的程度,  $p$  值越小, 说明两个变量之间的关联性越显著。

$$P(X) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

Fisher's exact test 在生物学研究中广泛应用, 例如基因表达谱中的差异分析和基因功能富集分析等。这个测试可以帮助确定某些基因在特定条件下是否更可能与某些生物过程或疾病有关联。

## Transcript annotation & Epigenetic markers

**Transcript annotation** 是指识别和标记转录本或基因的不同区域和功能元素的过程。这通常包括确定编码区域、非编码区域 (UTRs)、内含子、外显子和其他功能元素的位置和边界。Transcript annotation 可以使用多种方法进行, 包括手动筛选、计算机算法和 RNA 测序等实验数据。

**Epigenetic markers** 是指对 DNA 或相关蛋白质的修饰, 可以在不改变基因序列的情况下影响基因表达。Epigenetic markers 的例子包括 DNA 甲基化、组蛋白修饰和染色质重塑。这些标记可以遗传, 并且可以影响整个发育过程中基因表达模式及其对环境因素的响应。Epigenetic markers 在许多生物学过程中都发挥重要作用, 包括细胞分化、发育和疾病。它们也可以作为疾病诊断和治疗的有用生物标志物。

## Correlational & causal gene networks

Correlational gene networks 和 causal gene networks 是基因表达分析中使用的两种基因网络类型。

**Correlational gene networks** 是基于样本间基因表达水平的相关性构建的网络。它们识别表达模式相似的基因并将它们分组为模块或聚类。Correlational gene networks 有助于识别潜在的基因调控关系, 以及了解在不同条件下基因共同调控的情况。然而, 这些网络无法提供基因之间调控关系的方向和性质等信息。

而 **Causal gene networks** 则是基于基因间的因果关系构建的网络。这些网络识别直接或间接调控其他基因表达的基因, 并提供这些调控关系的方向和强度等信息。Causal gene networks 比 Correlational gene networks 更加信息丰富, 因为它们允许识别控制特定生物过程的关键调控基因和途径。然而, 构建因果基因网络需要额外的实验数据和计算方法, 例如基因敲除或扰动以及可以从基因表达数据中推断因果关系的统计建模方法。

## Random Network & Scale-free Network

Random network 和 scale-free network 是两种常见的复杂网络拓扑结构。

**Random network** 是一种随机连接的网络, 其中每个节点具有相同的平均度数。在这种网络中, 节点之间的连接是完全随机的, 因此没有明显的聚类现象。然而, 随机网络在网络容错性、信息传播和同步等方面的研究具有重要的意义。

**Scale-free network** 是指节点度数呈幂律分布的网络, 其中少数节点具有非常高的度数, 而大多数节点的度数相对较低。这种网络结构在许多实际网络中都有发现, 例如社交网络、蛋白质相互作用网络和互联网拓扑结构等。Scale-free network 在信息传播、节点影响力、网络容错性等方面的研究具有重要意义。然而, 与 random network 相比, scale-free network 对节点攻击和失效的容忍度较低, 容易出现寡头垄断的情况。

总的来说，random network 和 scale-free network 具有不同的拓扑结构，这些结构反映了网络中节点之间的连接和调控方式，对于理解复杂系统和设计高效的网络算法都具有重要的意义。