

# Probabilistic Significance of Enrichment by Using Combinatorics and the Hypergeometric Distribution

Estimate the probability for a coherent biological theme to be reflected in the annotations of a random unordered group of  $n \subseteq N$  genes with the largest or smallest expression, selected from the group of all  $N$  genes without replacements, where  $K \subseteq N$  of them are annotated to be of a particular molecular function or biological process, and where  $k \subseteq K$  of them are in the group of  $n$  genes, by using combinatorics

$$\begin{aligned} P - \text{value}(k; n, N, K) &= \sum_{i=k}^n P(i; n, N, K) \\ &= \binom{N}{n}^{-1} \sum_{i=k}^n \binom{K}{i} \binom{N-K}{n-i} \end{aligned}$$

Tavazoie et al., *Nat Genet* 22, 281 (1999);

Eden et al., *BMC Bioinformatics* 10, 48 (2009); <http://cbl-gorilla.cs.technion.ac.il/>