

# KEGG 富集分析说明

KEGG(Kyoto Encyclopedia of Genes and Genomes)是系统分析基因功能、基因组信息数据库，它有助于研究者把基因及表达信息作为一个整体网络进行研究。基因组信息存储在 GENES 数据库里，包括完整和部分测序的基因组序列；更高级的功能信息存储在 PATHWAY 数据库里，包括图解的细胞生化过程如代谢、膜转运、信号传递、细胞周期，还包括同系保守的子通路等信息。KEGG 提供的整合代谢途径 (pathway)，包括碳水化合物、核苷、氨基酸等的代谢及有机物的生物降解，不仅提供了所有可能的代谢途径，而且对催化各步反应的酶进行了全面的注解，包含有氨基酸序列、PDB 库的链接等等。KEGG 是进行生物体内代谢分析、代谢网络研究的强有力工具。

将差异表达基因进行 KEGG 富集分析，可以把差异显著的 pathway 进行富集，有助于找到实验条件下显著性差异变化的生物学调控通路。

富集分析方法说明：

采取的方法是 fisher 精确检验，数据包是 clusterProfiler，来自 R/bioconductor；

挑选的标准是落在某个 term/pathway 上差异的基因数目  $\geq 2$ ,  $p\text{-value} < 0.05$ ,

画图中取得 term/pathway 是按照 enrich factor 的值从大小降序排列，取前 30 个结果。

$\text{enrich\_factor 定义} = (\text{某个 term 中的差异基因数目} / \text{总的差异基因数目}) / (\text{数据库 term 中总的基因数目} / \text{数据库中总的基因数目})$

字段说明：

ID	对应 PATHWAY 数据库中的 ID
Description	PATHWAY 的描述
GeneRatio	对应 PATHWAY 差异基因数 / 能够对应到 PATHWAY 数据库中的差异基因数
BgRatio	对应PATHWAY包含对应物种的基因数 / PATHWAY 数据库中包含对应物种的基因数
pvalue	富集分析得到的 p-value
p.adjust	校正后的 p-value

<b>qvalue</b>	富集分析得到的 qvalue
<b>geneID</b>	富集基因列表
<b>Count</b>	富集基因数目
<b>enrich_factor</b>	富集程度

上海伯豪生物技术有限公司

技术支持部

2016 年 1 月