

=====文件=====

1. \*.all.xls 未经 log2fc 和 FDR 过滤的差异分析结果

注:第一列表示基因 ID,倒数第二列表示错误发现率阈值 FDR,倒数第一列表示 log2FC,其余各列表示基因在差异分组各样品中的 counts 值和表达量

2. \*.DEG\_final.xls 差异表达基因结果

注:第一列表示基因 ID,其后几列表示基因在各分组的 counts 值和表达量以及差异表达信息,包括错误发现率 FDR,差异倍数(Fold Change, FC),调控信息(up 表示上调表达基因, down 表示下调表达)。

注:log2FC 并不是用 FPKM 值直接计算的,而是DESeq 或者 EBSseq 软件对 read count 数进行转换,转换后的count值用于计算log2FC。这样的计算方法更精确,但会与用 FPKM 值计算有少量出入。

3. \*.cor.png 所有基因在分组内表达量相关性图

注:图中每一个点代表一个基因,横坐标和纵纵坐标分别对应基因在不同处理(如实验或对照)中的表达量以  $\log_2(\text{FPKM}+1)$  转换后的数值,同时在图中标识了对角线和相关性系数  $r$  的平方值。点在对角线附近越集中,则两个处理(实验和对照)中基因表达量相关性越强,偏离对角线的点代表差异表达基因。

4. \*.MA.png 差异表达 MA 图

注:差异表达基因 MA 图中每一个点代表一个基因。横坐标为 A 值: $\log_2(\text{FPKM})$ ,即两样品中表达量均值的对数值;纵坐标为 M 值: $\log_2(\text{FC})$ ,即两样品间基因表达量差异倍数的对数值,用于衡量表达量差异的大小。图中绿色的点代表下调差异表达基因,红色的点代表上调差异表达基因,黑色的点代表非差异表达基因。

5. \*.Volcano.png 差异表达火山图

注:差异表达火山图中的每一个点表示一个基因,横坐标表示某一个基因在两样品中表达量差异倍数的对数值;纵坐标表示基因表达量变化的统计学显著性的负对数值。

横坐标绝对值越大,说明表达量在两样品间的表达量倍数差异越大;纵坐标值越大,表明差异表达越显著,筛选得到的差异表达基因越可靠。

图中绿色的点代表下调差异表达基因,红色的点代表上调差异表达基因,黑色的点代表非差异表达基因。

6. \*.annotation.xls 差异基因的功能注释结果

注:展示差异基因在各数据库中的注释结果。\*\_count 为样本的 count 值;\*\_FPKM 为样本的 FPKM 值;FDR 差异表达 FDR 值;log2FC 为两样品间基因表达量差异倍数的对数值;regulated 为上下调信息;COG\_class,COG 分类结果;COG\_class\_annotation,COG 分类的功能描述;GO\_annotation,GO 功能注释结果;KEGG\_annotation,KEGG 功能注释结果;KEGG\_pathway\_annotation, 基因注释到的 pathway; KOG\_class, KOG 分类; KOG\_class\_annotation, KOG 分类的功能描述; Pfam\_annotation, Pfam 功能注释结果; Swiss-Prot\_annotation,Swiss-Prot 的功能注释结果; eggNOG\_class,eggNOG 数据库的分类;

eggNOG\_class\_annotation,eggNOG 的功能分类描述; NR\_annotation, NR 数据库的注释结果。