



GCBI
Gene-Cloud of
Biotechnology Information

GCBI 平台快速数据分析 帮助文档

上海其明信息技术有限公司

目 录

1. 表格文件说明	1
1.1 文件意义说明	1
1.2 SAMPLES_GROUPS.TXT 列头说明	1
1.3 DIFF_RESULT.TXT_两分组列头说明.....	1
1.4 PATHWAY_RESULT.TXT 列头说明	1
1.5 PATH_NET_ATTRIBUTION_RESULT.TXT 列头说明	2
1.6 PATH_NET_RELATION_RESULT.TXT 列头说明	2
2. 图片文件说明	2
3. 其他文件说明	3
4. 方法学说明	3
4.1 差异方法学说明:	3
4.2 聚类图方法学说明:	3
4.3 功能和信号通路方法学说明:	3
4.4 网络图方法学说明:	3

1. 表格文件说明

Tips:建议所有 txt 文件使用 xls 打开。

1.1 文件意义说明

文件名称	内容
samples_groups.txt	样本分组信息表
diff_result.txt	差异结果表
pathway_result.txt	信号通路分析结果表
path_net_relation_result.txt	信号通路上下游分析信号通路关系表
path_net_attribution_result.txt	信号通路上下游分析信号通路属性表

1.2 samples_groups.txt 列头说明

列头	意义
sample_id	样本编号
sample_attr_id	样本对应临床信息

1.3 diff_result.txt_两分组列头说明

列头	意义
ENSG_id	TCGA 所使用的的 Ensembl 79 的 ID
Gene Symbol	某一探针编号对应的基因名称/lncRNA 名称
Fold change	差异倍数
Gene Feature	基因在实验中的变化属性，如“diff”表示差异，“up”表示上调，“down”表示下调
P_value	评估 ENSG 的表达值在组间的显著性水平，即探针集对应的基因在组间的差异
FDR	阳性误判率，差异基因中假阳性基因所占比例的期望，值越小假阳性率越低
Biotype	表示名称是属于 mRNA 或 ncRNA, coding 表示属于 mRNA, noncoding 表示属于 lncRNA 或 microRNA
sample1	样本对应的表达值
sample2	样本对应的表达值

1.4 pathway_result.txt 列头说明

列头	意义
Pathway ID	pathway 索引号，与 KEGG 数据库的 pathway 索引号一致

Pathway Name	pathway 名称, 与 KEGG 数据库中 pathway 的命名方式一致
Diff Gene Counts in Pathway	差异基因计数, 表示所有差异基因中属于某一 pathway 的差异基因数量
Gene Amount in Pathway	基因计数, 表示数据库中属于某一 pathway 的基因数量
Enrichment Score	富集度, 若 p 值相同, 富集度越大的 pathway, 表示该 pathway 受到实验的影响越大
p-value	p 值, 评估 pathway 的显著性水平, ($p < 0.05$ 表示 pathway 具有显著性差异, 用黄色标注)
FDR	误判率, 对 p 值准确率的判断, 对 pathway 显著性水平的再判断
Gene_symbol	基因的名称, 与 NCBI GenBank 的基因标识一致
Rank	排序

1.5 path_net_attribution_result.txt 列头说明

列头	意义
Pathway ID	pathway 索引号, 与 KEGG 数据库的 pathway 索引号一致
Pathway Name	pathway 名称, 与 KEGG 数据库中 pathway 的命名方式一致
Outdegree	出度, 某一 pathway 的下游 pathway 数量
Indegree	入度, 某一 pathway 的上游 pathway 数量
Degree	度, 某一 pathway 的上、下游 pathway 数量
Pathway Feature	基因在实验中的变化属性, 如“diff”表示差异, “up”表示上调, “down”表示下调

1.6 path_net_relation_result.txt 列头说明

列头	意义
Source Pathway ID	网络中上游 pathway 索引号, 与 KEGG 数据库的 pathway 索引号一致
Source Pathway Name	网络中上游 pathway 名称, 与 KEGG 数据库中 pathway 的命名方式一致
Target pathway ID	网络中下游 pathway 索引号, 与 KEGG 数据库的 pathway 索引号一致
Target pathway Name	网络中下游 pathway 名称, 与 KEGG 数据库中 pathway 的命名方式一致

2. 图片文件说明

文件名称	内容
diff_cluster.pdf	差异基因聚类图
diff_cluster_top20.pdf	差异基因聚类 TOP20 图
pathway_pvalue_top20.png	信号通路分析显著性最高 TOP20 图

3. 其他文件说明

文件名称	内容
path_net. cy s	图片结果的 cytoscape 文件，可以通过 3.5.1 以上版本的 cytoscape 软件调整图片。

Cytoscape 软件网址: <http://www.cytoscape.org/download.php>

4. 方法学说明

4.1 差异方法学说明:

使用 R 的 edgeR 的脚本包进行的差异分析。

EdgeR 的脚本包网址为:

<http://www.bioconductor.org/packages/release/bioc/html/edgeR.html>

4.2 聚类图方法学说明:

使用 R 的 pheatmap 的脚本包进行的聚类图分析制作。

Pheatmap 的脚本包网址为:

<https://cran.r-project.org/web/packages/pheatmap/index.html>

4.3 功能和信号通路方法学说明:

功能和信号通路方法学说明下载网址为:

<http://college.gcbi.com.cn/archives/534>

4.4 网络图方法学说明:

网络图方法学说明下载网址为:

<http://college.gcbi.com.cn/archives/534>