



**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
2. 图片文件说明 .....	1
3. 方法学说明 .....	2
3.1 差异方法学说明: .....	2
3.2 聚类图方法学说明: .....	2

## 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

### 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

## 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

### 3. 方法学说明

#### 3.1 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>

#### 3.2 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>