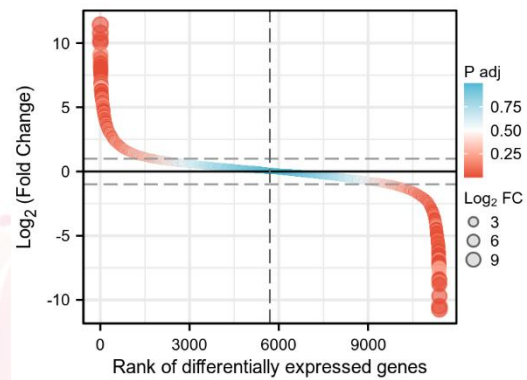


## 差异表达 - 差异排序图



网址: <https://www.xiantao love>



更新时间: 2023.03.14

## 目录

基本概念 .....	3
应用场景 .....	3
分析流程 .....	3
主要结果 .....	4
数据格式 .....	5
参数说明 .....	6
阈值 .....	6
标注 .....	7
点 .....	8
标题 .....	9
图注(Legend) .....	9
风格 .....	10
图片 .....	11
结果说明 .....	12
主要结果 .....	12
补充结果 .....	12
方法学 .....	14
如何引用 .....	15
常见问题 .....	16

## 基本概念

- 差异排序图：差异分析后得到差异倍数(FoldChange)和经过转换的校正后的 p 值，通过最终所得差异倍数数据的大小进行排序，做出差异排序图，展示差异分析的结果，也是火山图的另外一种可选的可视化形式。

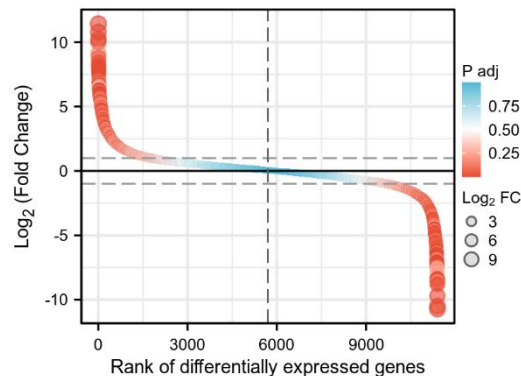
## 应用场景

主要用于可视化差异分析后的结果，需要提供 logFC 和校正后的 p 值（或者 p 值）

## 分析流程



## 主要结果



通过差异排序图展示差异分析结果。

- 图中横坐标为差异表达基因根据差异倍数排序后（从大到小）的**位置**，纵坐标为差异表达倍数（**log 化处理**）。图中的点越靠近左右两边，差异倍数的绝对值越大。
- 图中每个点代表一个分子（基因，即上传数据的 ID 列），**点的相对大小**对应其 logFC 值（即上传数据的第 2 列），**点的颜色**代表经过校正后的 p 值（或 p 值，即上传数据的第 3 列）。
- （**阈值线**）图中的两根横线，分别代表 logFC 为 -1 和 1 的横线，还有一根竖线，代表差异表达基因数量中位数的线。（**阈值可以通过参数调整**）
- **图注说明**，一个代表按照校正后的 p 值范围的颜色渐变，另一个图注代表按照 logFC 的值画出相对应差异基因代表的点的大小。
- **补充说明**：如果上传的数据中分子的信息含有缺失，这些分子信息会被删除，只会保留信息全的（都含有 logFC 和校正后 p 值或者 p 值的）的内容。所以如果上传的数据中含有缺失的，可能最终图中实际的分子数量会少。

## 数据格式

	A	B	C
1	gene	logFC	padj
2	TINAG	11.42176924	0.002045602
3	CPLX2	10.78346635	1.72926E-24
4	INS-IGF2	-10.72502805	3.40332E-08
5	CLEC4M	-10.4939252	1.13013E-21
6	HOXA13	10.24743564	6.05726E-09
7	PAGE4	10.02038069	0.001547695
8	CLEC1B	-9.694371799	2.12659E-27
9	CHP2	9.146296978	0.147290991
10	RTBDN	8.80808934	0.044341667
11	CFTR	-8.793154523	3.58571E-09
12	SPRR1B	-8.625761728	0.181574636
13	CXCL14	-8.488664898	7.93508E-33

数据要求：

- 需要 3 列数据: ID 列 | logFC 列 | p 值或者校正后 p 值(p.adj)列, 至少 10 行以上。注意命名问题。
- 第 1 列为分子名 (也可以是其他名), 不能含有重复的名字, 列名可以随意。这部分数据将可能会用于标注分子功能。
- 第 2 列为 logFC 值, 对应分子经过差异分析后的 logFC 值, 不需要原始 Fold Change 值 (不需要转成 Fold Change 值), 这里的数值即为每个分子对应的 y 坐标, 不能有异常值。列名必须要为 logFC, 否则可能会存在识别错误。
- 第 3 列为校正后的 p 值, 列名必须为 pvalue 或者 p.adj。 (不需要手动转成 -log10 的形式)
- 补充说明: 如果上传的数据中分子的信息含有缺失, 这些分子信息会被删除, 只会保留信息全的 (都含有 logFC 和校正后 p 值或者 p 值的) 的内容。所以如果上传的数据中含有缺失的, 可能最终图中实际的分子数量会少。。
- 最多 70000 行, 若验证数据时返回报错, 需要在上传数据内进行相应的调整, 然后再上传数据。 (如果文件过大, 请适当减少 一些没用的列或者更换文件格式)

## 参数说明

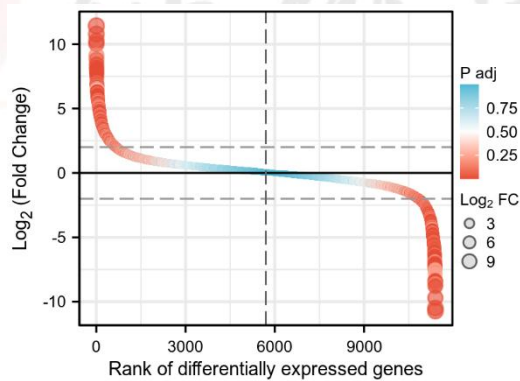
(说明：标注了颜色的为常用参数。)

## 阈值

阈值
 

logFC阈值 1

- $\log_2$ FC 阈值：主要控制图中划分是否有显著差异表达对应的差异倍数 (FoldChange) 阈值的阈值线，默认 1。设置  $\log_2$ FC 为 2 时，即差异倍数为 4，注意与 y 轴垂直的两条竖线，如下：



## 标注

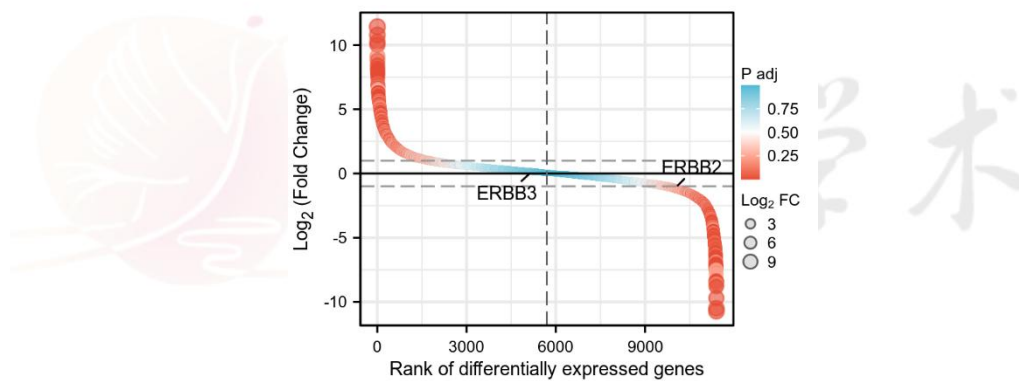
标注

标注id(数据第1列)

可以输入想要标注的id,  
1行1个

标注大小 5pt

- 标注 id: 可以输入想要标注的分子。这里输入的分子要和上传数据中的第一列对应，如果在上传数据的第一列中没有找到，则无法进行标注。另外，如果标注的分子比较多或者分子间距离比较近，可能也会无法标注。



- 标注大小: 控制图中需标注的文字大小，默认是 5pt。

## 点

点

填充色

描边色

样式

圆形

大小

1

不透明度

0.6

- **填充色**：点的填充色颜色选项，主要控制**显著性 p 值**的颜色范围，修改第一和第二色卡作为数值**从大到小**的渐变色，最多支持修改 2 个颜色。受配色方案全局性修改。
- **描边色**：点的描边色颜色选项，主要控制**显著性 p 值**的颜色范围，修改第一和第二色卡作为数值**从大到小**的渐变色，最多支持修改 2 个颜色。受配色方案全局性修改。
- **样式**：点的样式类型，可选择 圆形、正方形、菱形、三角形、倒三角，默认为圆形。**单选，选择后全局修改。**
- **大小**：点的相对大小，取决于 logFC 的数值范围。
- **不透明度**：点的透明度。0 为完全透明，1 为完全不透明。



## 标题

标题 ▼

大标题

大标题内容

x轴标题

x轴标题内容

y轴标题

y轴标题内容

- 大标题：大标题文本
- x 轴标题：x 轴标题文本
- y 轴标题：y 轴标题文本
- 补充：在要换行的中间插入\n。如果需要上标，可以用两个英文输入法下的大括号括住，比如 {{2}}；如果需要下标，可以用两个英文输入法下的中括号括住，比如 [[2]]。

## 图注(Legend)

图注 ▼

是否展示

☒

图注标题

图注标题内容

图注位置

默认 ▼

- 是否展示：是否展示图注
- 图注标题：可以添加图注标题，[修改颜色图注](#)

- 图注位置：可选择 默认、右、上。

## 风格



- 边框：是否添加外框
- 网格：是否添加网格
- xy 颠倒：可以颠倒 xy 轴
- 文字大小：针对图中所有文字整体的大小控制

## 图片

图片	▼
宽度 (cm)	7
高度 (cm)	5
字体	Arial ▼

- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体



## 结果说明

## 主要结果



主要结果格式为图片格式, 提供 PDF、TIFF 格式下载, 结果报告可以下载包括 pdf 以及说明文本的内容。

## 补充结果

### 差异统计

统计一些常见阈值(|logFC|大于2或者是0.58(0.58换算过来就是1.5倍))下的差异分子数量

筛选条件	筛选后的数量
LogFC >2 & p.adj<0.05	700
LogFC >1.5 & p.adj<0.05	881
LogFC >1 & p.adj<0.05	942
LogFC >0.58 & p.adj<0.05	942

此表格提供在上传数据中，一些常见阈值的差异分子数量统计。可以根据需要下载差异分析结果后用 excel 表进行过滤。



## 方法学

所有分析和可视化均在 R 4.2.1 中进行

涉及的 R 包：ggplot2 包（用于可视化）

处理过程：利用 ggplot2 包对上传的差异分析结果进行可视化。



## 如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 ([www.xiantao love](http://www.xiantao love))。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



## 常见问题

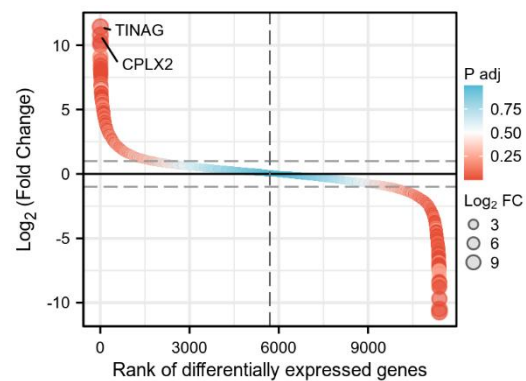
### 1. 如何标注分子?

答:

可以在 标注 参数中【标注 id】输入分子, [这里输入的分子要和上传数据中的分子列对应](#)。

- 如果输入了不存在的分子, 则无法进行标注。另外, 如果标注的分子比较多或者分子间距离比较近, 可能也会无法标注。

	A	B	C
1	gene	logFC	padj
2	TINAG	11.42176924	0.002045602
3	CPLX2	10.78346635	1.72926E-24
4	INS-IGF2	-10.72502805	3.40332E-08
5	CLEC4M	-10.4939252	1.13013E-21
6	HOXA13	10.24743564	6.05726E-09
7	PAGE4	10.02038069	0.001547695
8	CLEC1B	-9.694371799	2.12659E-27
9	CHP2	9.146296978	0.147290991
10	RTBDN	8.80808934	0.044341667



### 2. 为什么标注了分子没有在图中标注出来?

答:

首先, 需要保证输入的分子能在上传数据中的第一列找到, 如果在上传数据的分子列中没有找到, 则无法进行标注。另外, 如果标注的分子比较多或者分子间距离比较近, 可能也会无法标注。

### 3. 是否需要过滤掉不满足差异阈值的分子, 仅保留差异表达的分子?

答:

不建议只保留差异表达的分子, 建议是用差异分析后的全部的结果。



4. 上传数据中有一些分子没有  $\log FC$  或者  $p$  值，这种是否需要去掉？

答：

可以不用去掉，模块会自动去掉。分子中只要缺少了  $\log FC$  或者  $p$  值就无法在图中出现。

