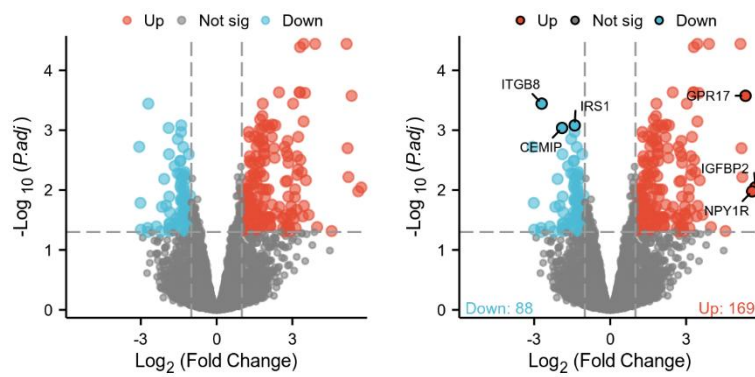


数据集工具 - [数据集]火山图



网址: <https://www.xiantao love>



更新时间: 2023.03.13

目录

基本概念	3
应用场景	3
分析流程	5
主要结果	6
云端数据	7
参数说明	8
数据设置	8
阈值	8
标注	10
点	11
标题	12
图注(Legend)	13
风格	13
图片	14
结果说明	15
主要结果	15
补充结果	16
方法学	17
如何引用	18
常见问题	19

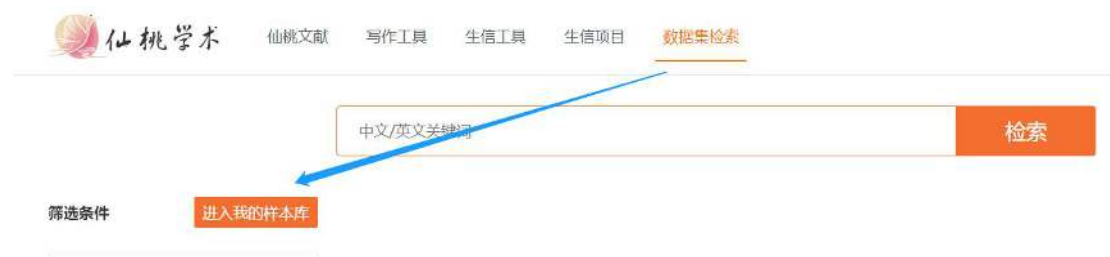
基本概念

- 数据集差异分析：从数据集检索模块中，针对特定 GEO 数据集的数据，进行芯片差异分析的过程，类似 GEO2R。
- 火山图：数据集检索差异分析后的可视化内容之一，火山图中每个点代表一个分子的差异倍数和显著性程度，整个图的形状形似火山喷发，所以称为火山图。

应用场景

本模块为 数据集检索 - 差异分析 后结果的可视化展示。用于可视化差异分析后的结果。

注意：模块需要**先进行 数据集检索 - 差异分析 后**，此处的云端数据才会有结果记录，然后才能进行可视化的操作。





数据集检索/样本库

刷新

<input type="checkbox"/>	分组	备注	数据集	平台	样本编号	Title
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214917	UET-13TR-EWS/FLI1 0hr
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214918	UET-13TR-EWS/FLI1 24hr
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214919	UET-13TR-EWS/FLI1 48hr
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214920	UET-13TR-EWS/FLI1 72hr
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214921	UET-13TR-EWS/FLI1 24hr tet+
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214922	UET-13TR-EWS/FLI1 48hr tet+
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214923	UET-13TR-EWS/FLI1 72hr tet+
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214924	UET-13TR-EWS/ERG 0hr
<input type="checkbox"/>		GSE8665	GSE8665	GPL570	GSM214926	UET-13TR-EWS/ERG 48hr

差异分析

缺失值处理

插补法

标准化处理

normalizeBctw

探针处理

处理

提交分析 (15/20)

免费版/基础版/高级版每日下载次数不同

选择参数进行分析

加入参考组 加入实验组

取消分组

删除所选

未分组17个

分析记录

时间最新在最上方记录

历史记录中超过30天的记录会自动清理

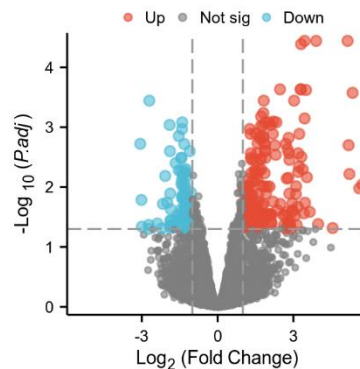
刷新

ID	名称	模块	状态	类型	时间	操作
1		芯片-差异分析	完成	表格	2023-03-12 20:35:56	更名 删除 下载 查看
2		芯片-差异分析	完成	表格	2023-03-12 20:14:16	更名 删除 下载 查看
3		芯片-差异分析	完成	表格	2023-03-12 20:13:12	更名 删除 下载 查看
4		芯片-差异分析	完成	表格	2023-03-12 20:10:10	更名 删除 下载 查看

分析流程



主要结果



通过火山图展示 数据集检索 - 差异分析结果。

- 图中横坐标为 \log_2FC 值，纵坐标为经过 $-\log_{10}$ 转换后的(校正后的) p 值。
- 每个点代表一个分子，对应有一个 \log_2FC 值和经过 $-\log_{10}$ 转换后的(校正后的) p 值。
- 不同颜色代表分子的调控情况，如上图，蓝色的点在左侧，代表 $\log_2FC < -1$ 且校正后 p 值 < 0.05 的下调分子；红色的点在右侧，代表 $\log_2FC > 1$ 且校正后 p 值 < 0.05 的上调分子。灰色的点代表 $|\log_2FC| < 1$ 或者校正后 p 值 > 0.05 的无差异分子。（阈值可以通过参数调整）
- （阈值线）图中还有两根竖线，分别代表 \log_2FC 为 -1 和 1 的竖线；还有一根横线，代表校正后 p 值在 0.05 的横线。（阈值可以通过参数调整）

云端数据

云端数据

	记录名称	来源模块	时间	补充说明
<input checked="" type="checkbox"/>		芯片-差异分析 @1.0	2023-03-12 20:35:56	数据记录可以在历史记录中找到

这里的云端数据与历史记录汇总 数据集检索工具样本库中【差异分析】的数据记录是保持一致的，可以在历史记录中找到相应的数据记录。

根据需要可视化的项目 选择好对应的云端数据记录。默认使用最近生成的分析记录。



参数说明

(说明：标注了颜色的为常用参数。)

数据设置



数据设置

p值类型 p.adj

p.adj

pvalue

- p 值类型：可选择 p.adj, pvalue。默认校正后 p 值。

阈值



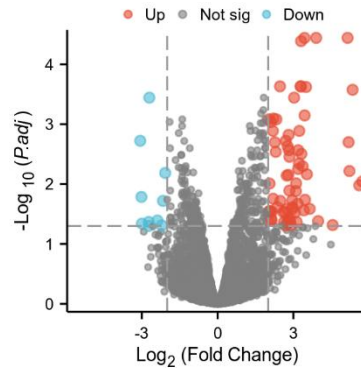
阈值

logFC阈值 1

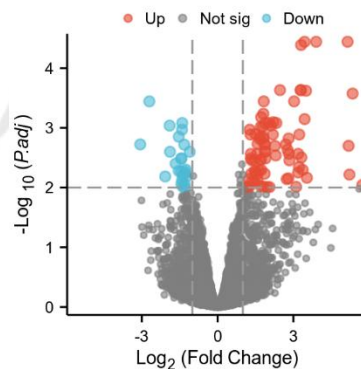
p值阈值 0.05

对称x轴长度 ☒

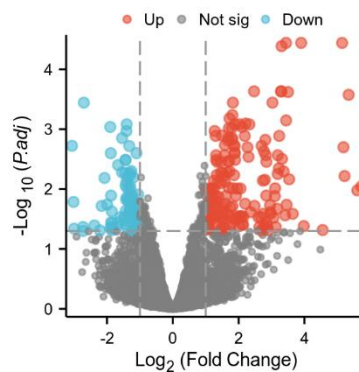
- \log_2FC 阈值：主要控制图中划分是否有显著差异表达对应的差异倍数 (FoldChange) 阈值，默认 1。设置 \log_2FC 为 2 时，即差异倍数为 4，注意与 x 轴垂直的两条竖线，如下：



- p 值阈值：主要控制图中划分是否有显著差异表达对应的显著性 (p_{adj} 活 p 值) 阈值，默认 0.05。设置 p 值阈值为 0.01 时，注意与 y 轴垂直的一条横线，如下：



- 对称 x 轴长度：可以强制调整图中 x 轴是否对称。当不强制时，如下：



标注

标注

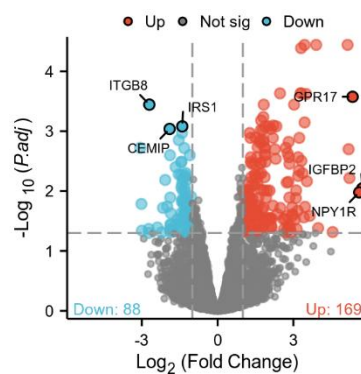
标注id(数据第1列)

可以输入想要标注的id, 1行1个

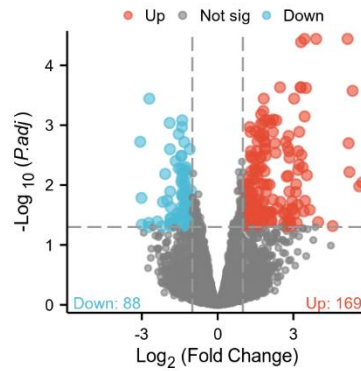
标注大小 5pt

标记差异数量

- 标注 id: 想要在图中标注的 id 信息，一般是基因 symbol。注意：输入的 ID 来自所选云端数据记录的结果，需要先在历史记录中找到对应的记录，下载差异表格结果，复制想要展示的 ID 到这个输入框中，一行代表一个。如果在对应的数据中没有找到对应的分子，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。



- 标注大小: 标注的字体大小，默认是 5pt。
- 标记差异数量: 可以在图中展示差异基因的数量，如下：



- 标注大小：控制图中需标注的文字大小，默认为 5pt。

点



- **填充色**：点的填充色颜色选项，主要控制显著上下调基因颜色，第一色卡控制**显著下调基因**分组，第二色卡控制**显著上调基因**分组，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，主要控制显著上下调基因颜色，第一色卡控制**显著下调基因**分组，第二色卡控制**显著上调基因**分组，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。多选，多选后不同的分组/分类中的点的类型也会有相应变化，循环取该参数值。
- 大小：点的大小。
- 不透明度：点的透明度。0 为完全透明，1 为完全不透明。

标题



- 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

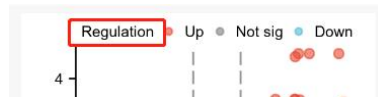
图注标题

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，如：



- 图注位置：可选择 默认、右、上。

风格

风格

边框

网格

xy颠倒

文字大小

7pt

- 外框：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

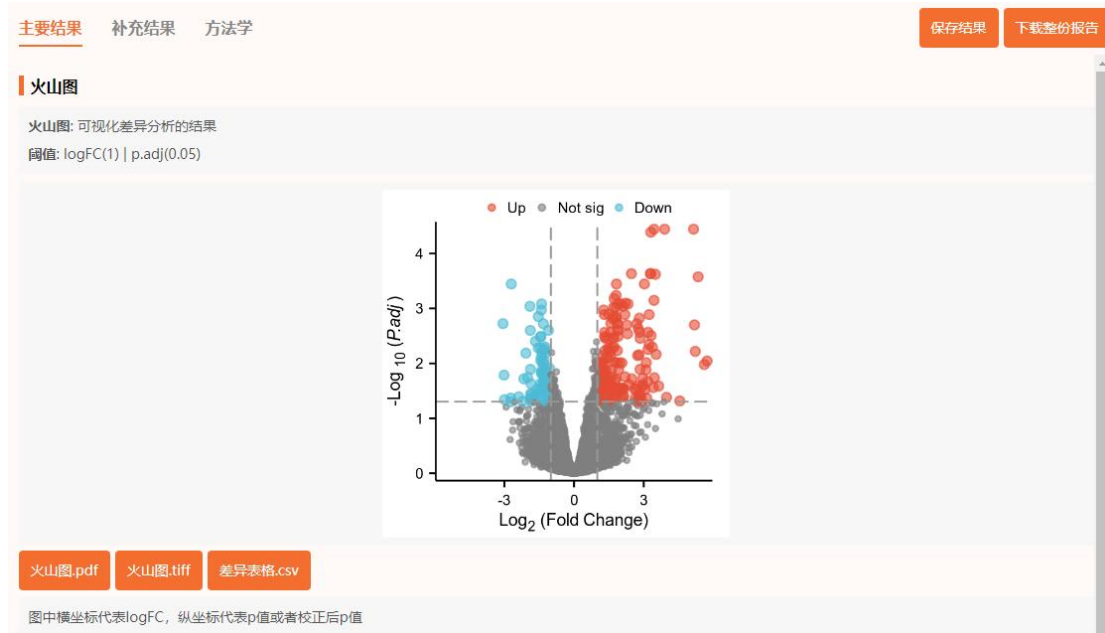


图片	
宽度 (cm)	5
高度 (cm)	5
字体	Arial

- 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。

➤ 另外, 提供对应云端数据的差异分析结果表格xlsx 下载。

	A	B	C	D	E	F
1	id	logFC	pvalue	p.adj	anno	sign
2	202718_at	5.728117201	2.56E-05	0.009053022	IGFBP2	Up
3	205440_s_at	5.608021617	3.52E-05	0.010506275	NPY1R	Up
4	206190_at	5.343444916	7.81E-08	0.000266787	GPR17	Up
5	205858_at	5.219314657	1.40E-05	0.006054541	NGFR	Up
6	236218_at	5.186295996	2.38E-06	0.002000403	PHOSPHO1	Up
7	238258_at	5.145286648	2.10E-09	3.62E-05	WBSCR28	Up
8	204818_at	4.553713216	0.000471459	0.048362152	HSD17B2	Up
9	1554012_at	4.482564097	0.001792781	0.102424537	RSPO2	Not sig
10	1556168_s_at	3.980265138	0.000349759	0.041481768	MROH2A	Up
11	222549_at	3.899954499	2.65E-09	3.62E-05	CLDN1	Up
12	205051_s_at	3.869707875	0.000510495	0.050381466	KIT	Not sig
13	213880_at	3.790425158	0.001242838	0.083071135	LGR5	Not sig

差异表格 (1)

补充结果

差异统计

统计一些常见阈值($|\log FC|$ 大于2或者1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
$ \log FC > 2 \ \& \ p.adj < 0.05$	81
$ \log FC > 1.5 \ \& \ p.adj < 0.05$	144
$ \log FC > 1 \ \& \ p.adj < 0.05$	257
$ \log FC > 0.58 \ \& \ p.adj < 0.05$	327

此表格提供在差异分析结果中，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)。

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

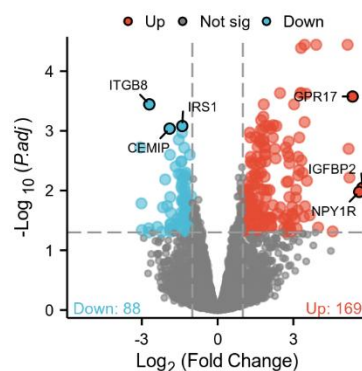
1. 如何标注分子?

答:

可以在 标注 参数中【标注 id】输入分子，[这里输入](#)的分子要和对应的云端记录的结果中的分子列对应，下载差异表格，复制 id 或者 anno 列中的分子，最终以基因 symbol 的形式标注。

- 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	D	E	F
1	id	logFC	pvalue	p.adj	anno	sign
2	202718_at	5.728117201	2.56E-05	0.009053022	IGFBP2	Up
3	205440_s_at	5.608021617	3.52E-05	0.010506275	NPY1R	Up
4	206190_at	5.343444916	7.81E-08	0.000268787	GPR17	Up
5	205858_at	5.219314657	1.40E-05	0.000054541	NGFR	Up
6	236218_at	5.186295996	2.38E-06	0.002000403	PHOSPHO1	Up
7	238258_at	5.145286648	2.10E-09	3.62E-05	WBSCR28	Up
8	204818_at	4.553713216	0.000471459	0.048362152	HSD17B2	Up
10	1556168_s_at	3.980265138	0.000349759	0.041481768	MROH2A	Up
11	222549_at	3.899954499	2.65E-09	3.62E-05	CLDN1	Up
14	210961_s_at	3.646783348	0.000154166	0.025935451	ADRA1D	Up
16	212670_at	3.53442207	1.69E-05	0.006860226	ELN	Up
18	205306_x_at	3.502286344	6.14E-08	0.000239656	KMO	Up



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云端记录的结果的分子列找到，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. 能否调整为可视化 p 值（校正后 p 值）？

答：

可以下载对应的云端记录的差异分析的结果，稍微整理数据后可到【表达差异】分类下的火山图上传数据进行可视化。

