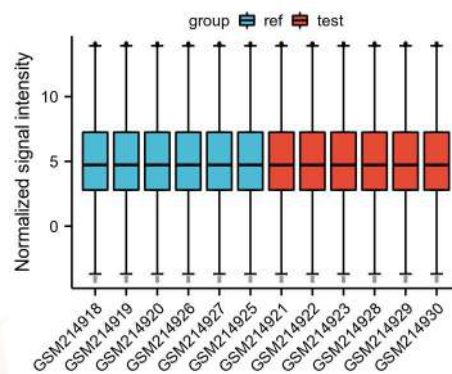


数据集工具 - [数据集] 箱式图



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
样式	8
箱	9
离群点	10
分面	错误! 未定义书签。
标题	11
图注(Legend)	11
坐标轴	12
风格	14
图片	14
结果说明	16
主要结果	16
补充结果	17
方法学	18
如何引用	19
常见问题	20

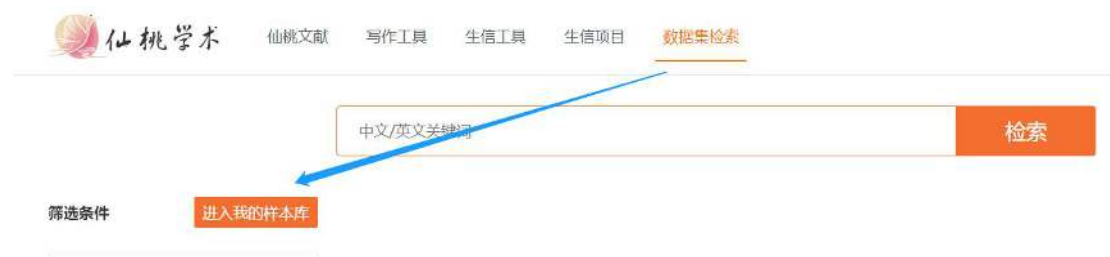
基本概念

- 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。
- 样本归一化箱式图：用箱式图来可视化数据集差异分析中各个样本中分子的芯片强度情况，比较常用于小样本的数据集中（当样本很多时，不适合用箱式图进行可视化）。

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。用于查看数据集中样本归一化情况。

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/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/高级版每日次数不同

选择参数进行分析

加入参考组
 加入实验组

未分组17个

取消分组

删除所选

分析记录

时间最新在最上方记录

① 历史记录中超过30天的记录会自动清理

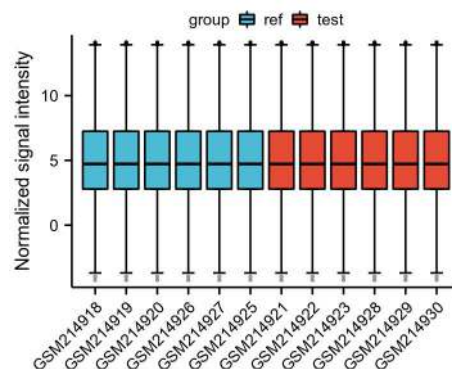
刷新

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



通过样本归一化箱式图展示 数据集检索 - 差异分析结果。

- 图中的横坐标为样本编号，纵坐标为芯片信号强度。
- 不同颜色代表在差异分析阶段，自定义的参考组（默认 ref）和实验组（默认 test）。图中箱子的颜色为所选择的【颜色映射】内容，选择 不映射 时将显示同样的颜色。



- 主要看各个样本的信号强度是否大致在一个中位数水平（代表样本已经校正好）。
- 一般单个数据集不是从最原始的 raw data 开始是不太需要做批间差去除或者校正的（因为一般数据集在上传 series_matrix 格式（GEOquery 下载就是这个格式）的时候作者都是做了的），但是也有作者强制会在文中进行校正，使得中位数、四分位数完全在一条线上。默认分析参数中是强制进行 normalizeBetweenArrays。

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。)

样式



- 颜色映射：主要影响箱子的颜色范围，**注意映射内容的类型，分类型数据为单个颜色**。可选择 分组 (group)、不映射。默认分组 ref、test。

箱

箱

填充色

描边色

描边粗细

0.75pt

不透明度

1

宽度

0.8

- **填充色**：箱子的填充色颜色选项，取决于 **颜色映射** 参数所选择的内容，展示分类型内容（如 分组 group）时，有多少个分组会提取多少个颜色，第一色卡控制参考组（默认 ref）分组，第二色卡控制实验组（默认 test）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**：箱子的描边色颜色选项，，取决于 **颜色映射** 参数所选择的内容，展示分类型内容（如 分组 group）时，有多少个分组会提取多少个颜色，第一色卡控制参考组（默认 ref）分组，第二色实验组（默认 test）分组，最多支持修改 2 个颜色。默认黑色，**不受配色方案全局性修改**。
- **描边粗细**：箱子描边的粗细，默认为 0.75pt。
- **不透明度**：箱子的透明度。0 为完全透明，1 为完全不透明。
- **宽度**：箱子的宽度。

离群点

离群点

展示

描边色

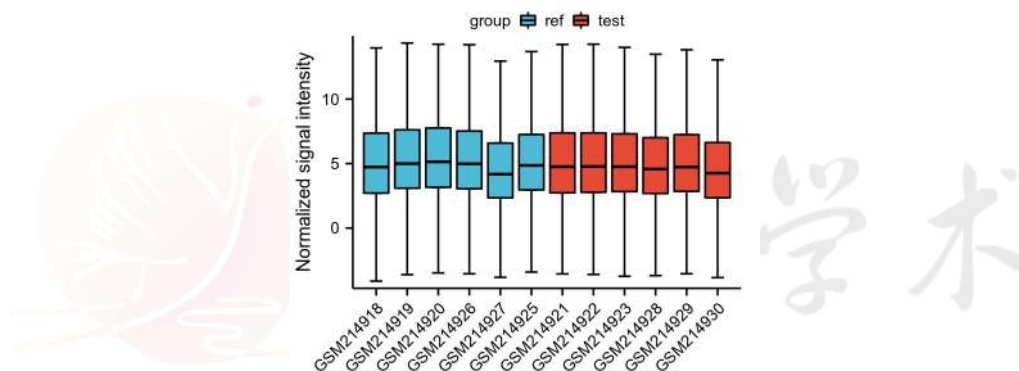
大小

0.4

不透明度

0.4

- 展示：可以选择是否展示离群点，默认展示，还可以选择不展示，如下：



- 描边色：在选择展示离群点时，可以修改离群点的颜色，默认黑色。
- 大小：在选择展示离群点时，可以修改离群点的大小。
- 不透明度：在选择展示离群点时，可以修改离群点的不透明度，默认为 0.4，1 表示完全不透明，0 表示完全透明。

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

图注标签

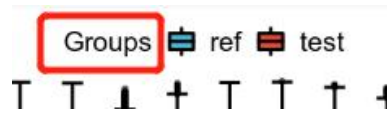
图注标签内容

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：



- 图注标签：可以添加图注标签信息，如：

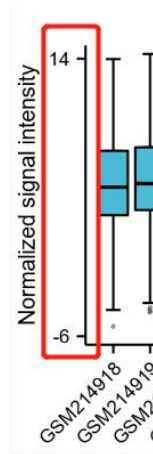


- 图注位置：可选择 默认、右、上。

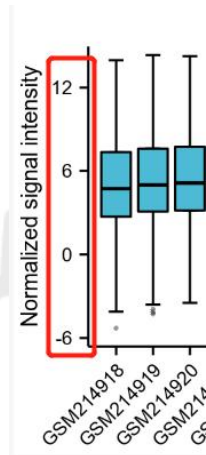
坐标轴



- x 轴标注旋转：支持对 x 轴文字进行旋转。适合于 x 轴文字过长的時候。
- y 轴范围+刻度：用于修改 y 轴范围以及刻度，如果需要分割，需要用小括号(英文输入法)隔开，数值间需要用逗号隔开，例如(1,1,2,5,5)。如果调整过大可能会无作用。
- 如果只是想要修改范围，可以只输入两个范围值，比如 -6,-6,14,14



- 如果同时想要修改范围+刻度，可以输入比如：-6,-6,0,6,12,12。注意，此时最大和最小值会被当做范围值，不会作为刻度，如果需要刻度，需要类似于 12 那样同时写两次。



风格



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

图片

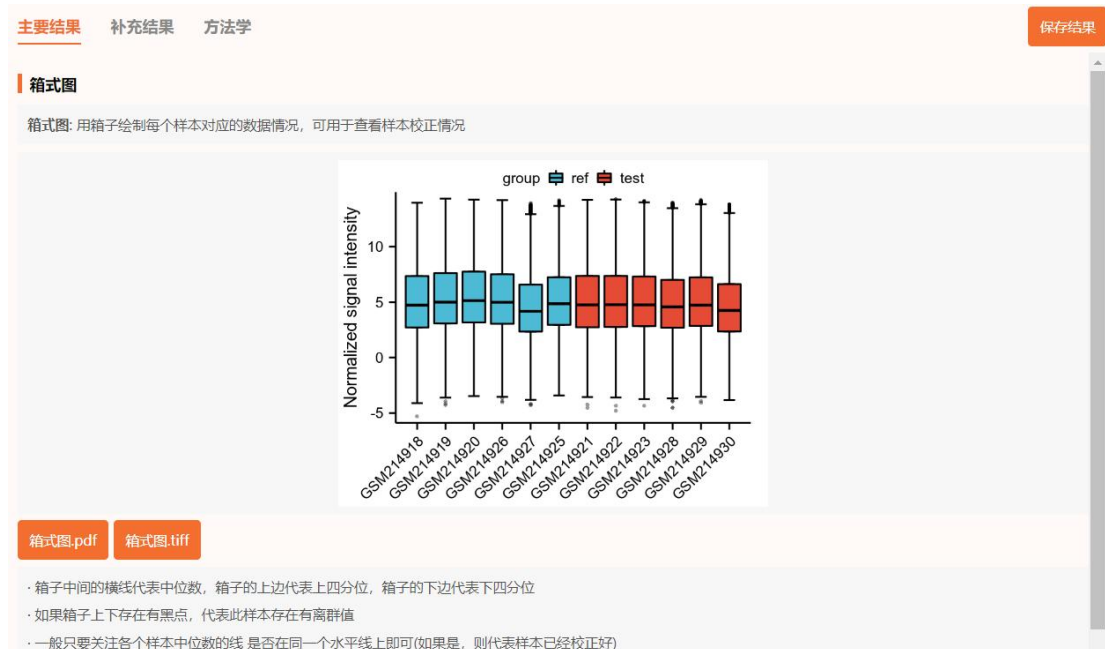


- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体



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

统计描述

各个样本对应下最小值、四分位、中位数、上四分位、最大值

样本	最小值	下四分位	中位数	上四分位	最大值
GSM214918	-5.3051	2.711	4.73	7.3561	13.964
GSM214919	-4.2817	3.0932	4.9975	7.6098	14.343
GSM214920	-3.476	3.158	5.1374	7.7507	14.25
GSM214926	-4.0323	3.0491	4.9876	7.5198	14.211
GSM214927	-4.3034	2.3445	4.1768	6.5866	13.951
GSM214925	-3.4114	2.9574	4.8556	7.2486	14.201
GSM214921	-4.5459	2.7346	4.7561	7.3729	14.236
GSM214922	-4.7923	2.7754	4.7698	7.3713	14.303
GSM214923	-4.3452	2.8321	4.7572	7.3026	14.141
GSM214928	-4.5353	2.6832	4.5737	7.0044	14
GSM214929	-4.0798	2.848	4.7295	7.2413	14.242
GSM214930	-3.8421	2.3488	4.2506	6.6251	13.845

[统计描述.xlsx](#)

此表格提供样本数据的统计描述结果，提供 Excel、Docx 格式下载。



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：用 ggplot2 包进行箱式图可视化。



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。



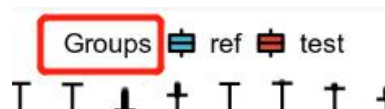
常见问题

1. 如何修改分组名?

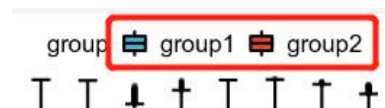
答:

可以在 图注 参数中修改【图注标题】、【图注标签】:

➤ 如修改标题为 Groups:



➤ 如修改标签为 group1,group2:



➤ 逗号为英文输入法下的逗号，其他的没办法识别。

2. 为什么只有一个箱式图? 如何绘制归一化前的箱式图?

答:

一般单个数据集不是从最原始的 raw data 开始是不太需要做批间差去除或者校正的（因为一般数据集在上传 series_matrix 格式（GEOquery 下载就是这个格式）的时候作者都是做了的），但是也有作者强制会再文中进行校正，使得中位数、四分位数完全在一条线上。

如果有强制归一化的，可以在数据集分析的时候标准化参数选择 normalizeBetweenArrays 参数

当纳入到对照组和实验组中含有 2 个以上数据集时，工具就会对数据按照不同的数据集作为批间差进行校正，对应的箱式图是批间差校正后的。

3. 能否上传自己的分析数据进行可视化?

答:

自己的差异分析的结果可以上传到基础绘图的箱式图等模块进行可视化。

