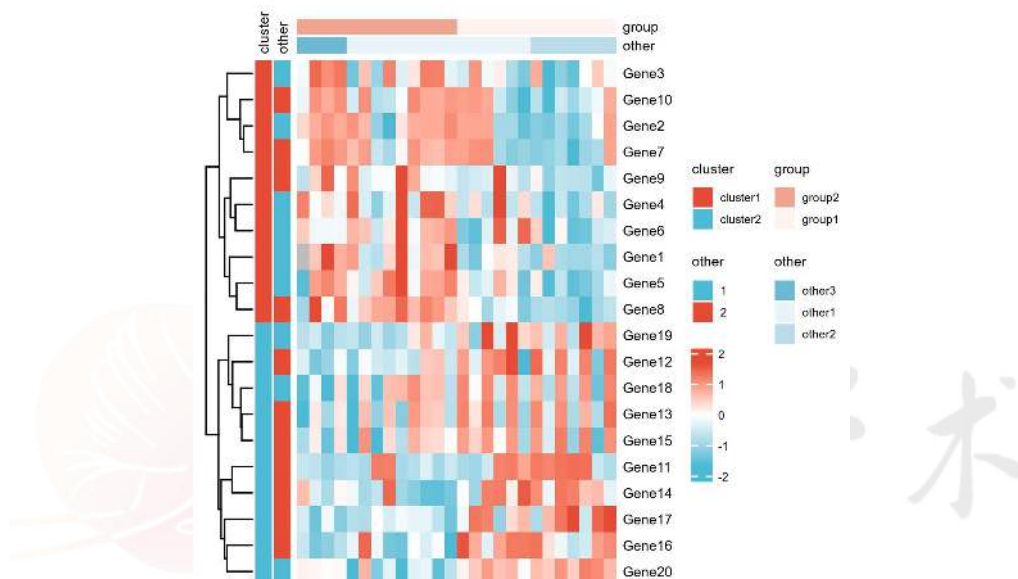


基础绘图 - 复杂数值热图



网址: <https://www.xiantao.love>



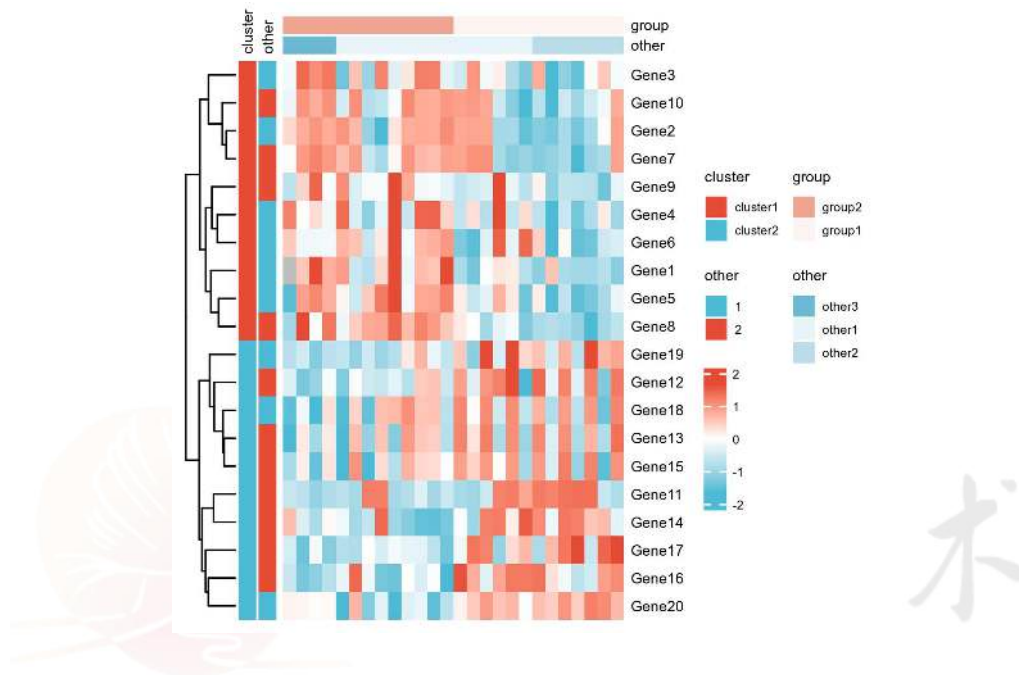
更新时间: 20230225

目录

基本概念	3
应用场景	3
结果解读	4
数据格式	6
参数说明	9
标注	9
数据处理	9
聚类(顺序)/分割	10
主图	11
文字	11
图注	12
图片	12
结果说明	13
主要结果	13
补充结果	13
方法学	14
如何引用	15
常见问题	16

基本概念

- 复杂数值热图：热图通过色彩变化来显示数据的数值，可以用于可视化表达矩阵情况。通过在图中添加注释信息，能够展示更多其他如分组等信息
- 图形构成：

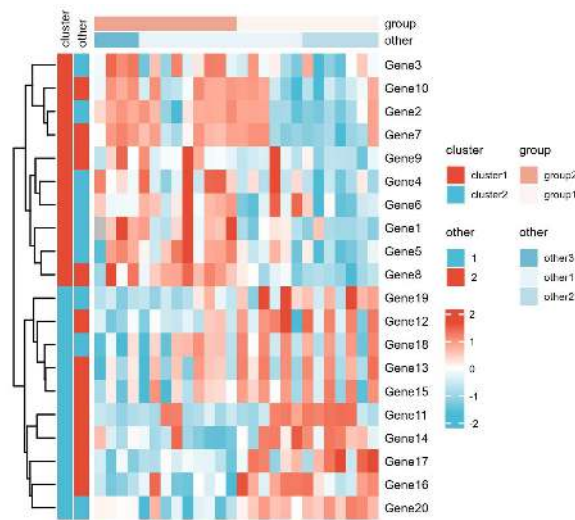


复杂数值热图

应用场景

热图经常用于展示多个基因在不同样本中的表达水平，然后可以通过聚类等方式查看不同组（如疾病组和对照组）特有的 pattern（如 CNV）。另外，热图还可以用于展示微生物的相对丰度、代谢组不同物质的含量等等。

结果解读



复杂热图

- 主要部分：每一个小方块代表第 j 列样本对应第 i 行基因的表达值经过行 $zscore$ 转换后的值对应的颜色大小。其中红色代表高，蓝色代表低。
- $zscore$ 转换是绘制热图中常用的一种对数据进行转换的方法（每个基因的表达值减去其在所有样本中的表达均值后，再除以标准差），可以减少不同分子表达值差异过大而影响整个热图的可视化效果，并且 $zscore$ 转换保留了单个分子在样本间的差异情况（如果一个分子表达值在样本间很大，另外一个在样本间很小，可以想象到上一个的热图基本都是红色，另外一个基本都是蓝色，单个基因样本间的差异就很小，很难看出单个基因的效果）
- 主要部分的左侧：包含行的聚类树状图 和 左注释的内容。
 - 聚类树状图：根据每一行的具体情况进行聚类，不同的聚类方法会影响这部分的顺序。如果选择了行聚类，则最终行的顺序不同于上传数据中行的顺序。
 - 左注释：来自上传数据的左注释表的内容。

- 主要部分的上方：包含列的聚类树状图 和 上注释的内容
 - 聚类树状图：根据每一列样本的具体情况进行聚类，不同的聚类方法会影响这部分的顺序。如果选择了列聚类，则最终行的顺序不同于上传数据中行的顺序。
- 左注释：来自上传数据的上注释表的内容。

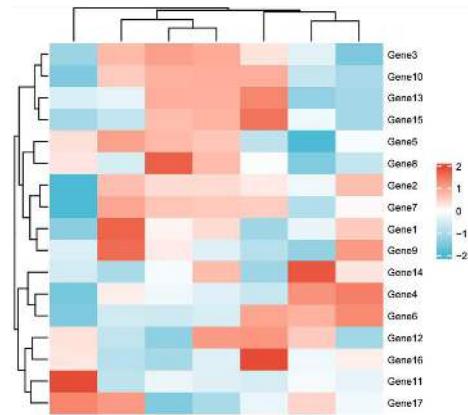


数据格式

- 主要部分：每一个小方块代表第 j 列样本对应第 i 行基因的表达值经过行 $zscore$ 转换后的值对应的颜色大小。其中红色代表高，蓝色代表低。
 - $zscore$ 转换是绘制热图中常用的一种对数据进行转换的方法（每个基因的表达值减去其在所有样本中的表达均值后，再除以标准差），可以减少不同分子表达值差异过大而影响整个热图的可视化效果，并且 $zscore$ 转换保留了单个分子在样本间的差异情况（如果有一个分子表达值在样本间很大，另外一个在样本间很小，可以想象到上一个的热图基本都是红色，另外一个基本都是蓝色，单个基因样本间的差异就很小，很难看出单个基因的效果）
- 主要部分的左侧：包含行的聚类树状图 和 左注释的内容。
 - 聚类树状图：根据每一行的具体情况进行聚类，不同的聚类方法会影响这部分的顺序。如果选择了行聚类，则最终行的顺序不同于上传数据中行的顺序。
 - 左注释：来自上传数据的左注释表的内容。
- 主要部分的上方：包含列的聚类树状图 和 上注释的内容
 - 聚类树状图：根据每一列样本的具体情况进行聚类，不同的聚类方法会影响这部分的顺序。如果选择了列聚类，则最终行的顺序不同于上传数据中行的顺序。
 - 左注释：来自上传数据的上注释表的内容。

数据格式

	A	B	C	D	E	F	G	H
1	id	sample1	sample2	sample3	sample4	sample5	sample6	sample7
2	Gene1	1.0857	2.2235	6.8363	3.0055	3.5615	-1.4764	-2.1033
3	Gene2	-0.6216	1.0857	2.033	1.1457	2.0632	0.6216	-6.7765
4	Gene3	-0.2429	5.5259	4.3451	5.0517	-4.9717	2.4139	-3.689
5	Gene4	4.2098	0.6785	1.6226	0.1889	4.7902	-0.552	-2.8266
6	Gene5	-3.6314	1.6453	2.1495	1.3583	-0.0633	-1.4107	0.7972
7	Gene6	1.6555	-0.1864	-0.1317	-0.0404	2.2153	1.8374	-1.2994
8	Gene7	-0.7368	1.4554	1.9224	1.3962	0.6254	1.3258	-2.1239
9	Gene8	-1.4519	3.6959	0.1097	2.1123	-0.2289	0.8314	1.3655
10	Gene9	-1.4127	2.0368	5.3161	0.6544	4.1518	-0.455	0.5713
11	Gene10	-0.1875	3.0719	2.424	3.1562	-0.9029	3.133	-1.7827
12	Gene11	-0.2351	-0.6641	-2.486	-1.2465	-0.8256	-1.6167	8.3875
13	Gene12	-0.3691	-3.8057	-2.7756	0.6089	-3.4537	0.6895	0.8223
14	Gene13	-8.3031	-0.1096	-4.5633	-0.1173	-7.5656	1.5439	-5.1931
15	Gene14	1.3715	-0.5592	-1.4696	0.2864	-0.1607	-1.6034	-1.0146
16	Gene15	-4.2134	-0.503	-6.3043	-0.1206	-7.7161	2.6279	-7.7402
17	Gene16	-0.7107	-2.0028	-1.697	-0.9882	-0.1341	2.6978	-0.0083
18	Gene17	-0.4036	-2.0148	0.1325	-1.6493	-0.9408	-1.0166	0.3084
19	Gene18	-8.0044	-2.0554	-8.6921	-0.2051	-8.2989	0.2051	-7.7849
20	Gene19	-0.5375	-0.0947	-1.4001	-0.5075	-0.3495	-1.1283	-0.1247
21	Gene20	0.0056	-0.021	-0.1307	-0.0713	-3.4778	0.9044	-2.0237



(默认是进行 行 Zscore 转换、行列聚类)

文件不能大于 6M，列数最多为 1200 列，行数最多 900。仅支持 xlsx 格式文件。

数据必须提供表达谱表（表一）：

- 第一列是列名，列名不能含有重复的内容
- 第一行是行名，行名不能含有重复的内容
- 中间部分是表达谱，也就是每个样本每个基因的表达值（一般是校正后的能够样本间比较的值，测序一般是校正后的值、FPKM、RPM、TMP、CPM 等等，一般不用 Counts），表达谱内不能含有一整行或者一整列都是非数值或者缺失数据的情况

如果需要增加热图上、左、右部分的注释信息，可在同一份 excel 数据中增加表（注意命名）：



- 第 1 列为样本名，与矩阵中第一行是对应的，
- 第 2 列以及之后的列为注释信息，对应的列名为注释名以及 legend 的标题名

如果需要修改相应注释内容的颜色，需要另外提供“颜色-上注释”表：



- 第一列的列名要跟 注释表中的列名是对应的（否则将找不到对应的信息）
- 分组信息只需要把单个注释中不同的分组列到一列，后面紧跟一列 colour 即可（这里的顺序会影响 legend 的上下的顺序）。每个注释的下一列的颜色都是与之一一对应的颜色。

参数说明

(说明：标注了颜色的为常用参数。)

标注



- 分子标注：可以填入数据中对应行名(分子名)，一行一个分子，如果为空，默认就标注数据中所有行名(分子)
- 标注的字体；默认是正体

数据处理



- 转换：可以对数据进行 log2 或者 log10 等转换
- 归一化：可以选择对行或列进行归一化

聚类(顺序)/分割

聚类(顺序)/分割

列聚类

不聚类

行聚类

欧式距离(euclidean)

列分割

不分割

行分割

不分割

左右颠倒

上下颠倒

- 列聚类：可以选择对列进行聚类分析，可以选择不同距离计算方法
- 行聚类：对行进行聚类分析，方法同列聚类
- 列分割：可以对列进行分割展示
- 行分割：同列分割
- 左右颠倒：可以对热图中的列的顺序进行颠倒
- 上下颠倒：可以对热图中的行的顺序进行颠倒

主图



- 色阶：热图块的颜色
- 色块描边：色块是否展示描边
- 外框：是否展示

文字



- 标题：填入热图的标题，换行可以在需要换行的位置插入\n
- 标题大小：标题的字体大小
- 列名大小：调节列名的大小或者不显示
- 行名大小：同列名大小

图注

图注

是否展示

☒

图注标题

图注标题内容

图注位置

默认

文字大小

6pt

- 是否展示：图注内容是否展示
- 图注标题：可以填入图注标题，换行可以在需要换行的位置插入\n
- 图注位置：默认是右，也可以选下
- 文字大小：图注文字的大小

图片

图片

宽度 (cm)

10

高度 (cm)

9

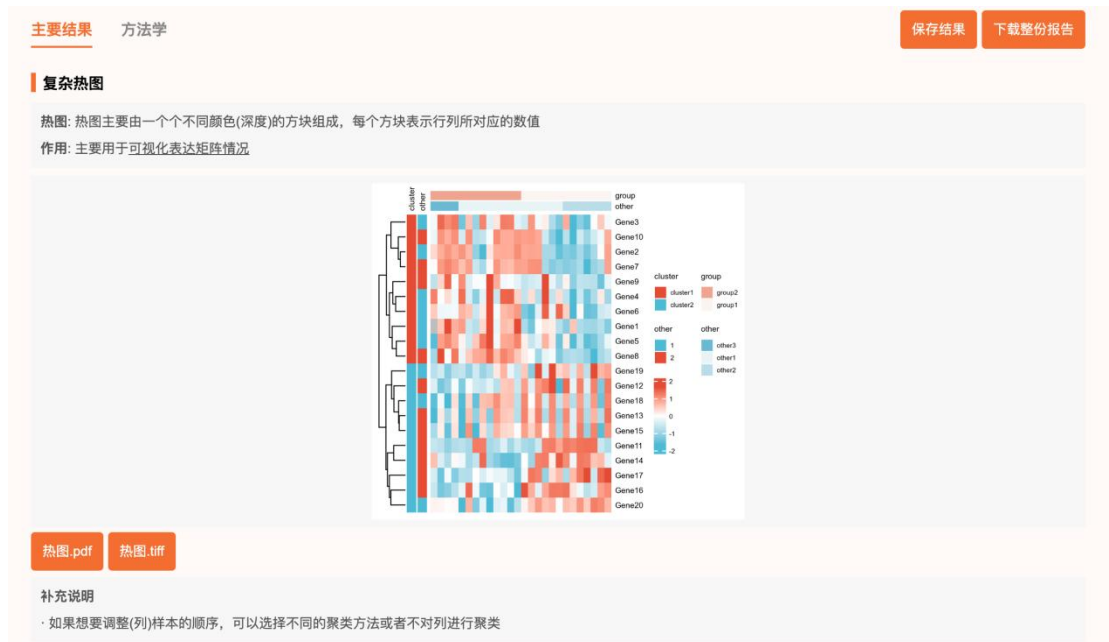
字体

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：可以选择图中文本内容字体

结果说明

主要结果



补充结果

无

方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包: ComplexHeatmap 包

处理过程:

- 利用 ComplexHeatmap 包进行热图的可视化

补充说明:

- 数据转换: 无
- 归一化处理: 对行归一化
- 行聚类: 欧式距离(euclidean)
- 列聚类: 不聚类

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。



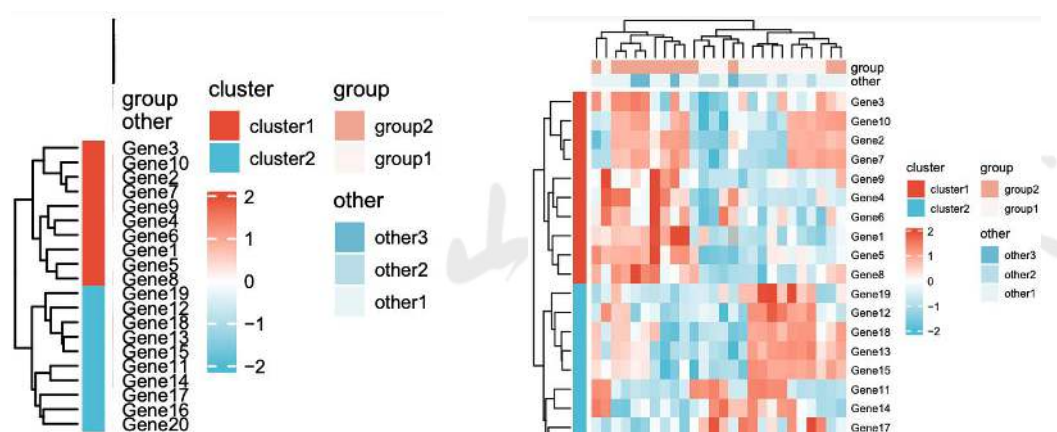
常见问题

1. 为什么要对数据进行转换?

答：实际应用中，异常值的出现可能会影响可视化结果，对数转换后的数据，看起来清晰很多，也会有更好的可视化效果。而且对数转换后，数据还保留着之前的变化趋势

2. 为什么图挤在了一起?

答：



由于文字不会被压缩，如果热图部分很长，就可能会导致热图部分重叠。解决方案可以是：

- ① 增加图片宽度和高度；
- ② 减少上传数据的行列。