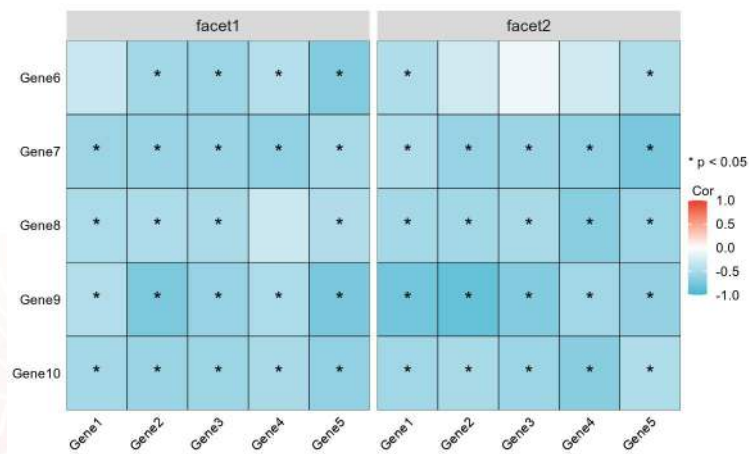


## 交互网络 - 相关性热图-分组 XY



网址: <https://www.xiantao.love>



更新时间: 2023.04.24

## 目录

基本概念 .....	3
应用场景 .....	3
分析流程 .....	3
结果解读 .....	6
数据格式 .....	8
参数说明 .....	9
统计 .....	9
映射 .....	10
样式 .....	11
热图 .....	12
标注 .....	14
分面 .....	15
标题文本 .....	15
图注 .....	16
坐标轴 .....	17
风格 .....	17
图片 .....	18
结果说明 .....	19
主要结果 .....	19
补充结果 .....	20
方法学 .....	22
如何引用 .....	23
常见问题 .....	24

## 基本概念

- 热图：热图是一个以颜色变化来显示数据情况的矩阵
- 相关性热图：通过热图的方式来展示变量之间的相关性

## 应用场景

相关性热图-分组 xy: 通过[将数据进行分组](#)的方式, 分别对各个组进行相应处理:  
[将各分组中的对应的数据分成两个不同方向](#)上变量(即将所有变量分成两个不同的组别)进行两两间(两分组间)相关性分析, 再以热图的形式展示其结果

## 分析流程

上传数据 ➡ 数据清洗 ➡ 数据处理 ➡ 将各分组进行相关性分析得到的结果进行可视化

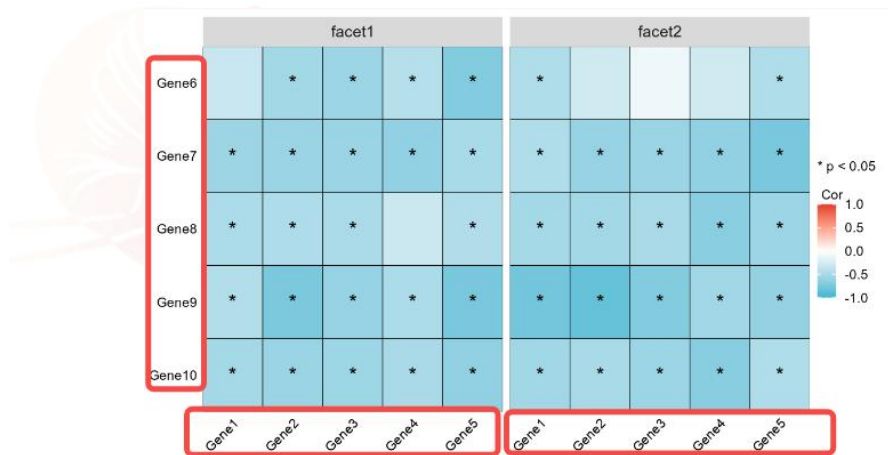
- 数据格式: csv / txt 文件格式: (具体格式可参考后续数据格式部分)

- 第 1 行为 **#注释的 xy 方向信息**

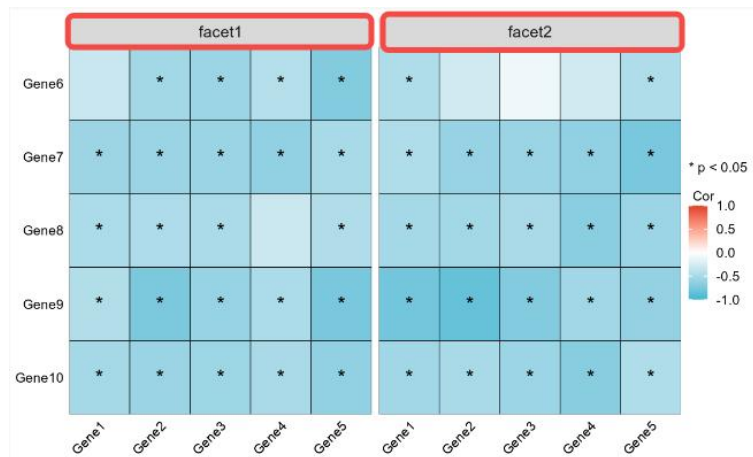
- ◆ **必须提供**: 在数据处理过程中会**根据注释信息将数据分成 xy 两个不同的方向(不同的方向对应不同的变量/列)**。#注释的信息中提供的**分类个数不能超过 2 个(二维坐标轴只支持 xy 两个方向)**, 注释信息的长度与数据(除了第 1 列外)的列数一致, 如下:

	A	B	C	D	E	F	G	H	I	J	K
1	#group	group1	group1	group1	group1	group1	group2	group2	group2	group2	group2
2	facet	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
3	facet1	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557	0.518298418	0.739705603
4	facet2	-0.01368477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841	0.981862233	0.568852396

- ① #group 表示不同方向的名字
- ② group1, ...group2 表示相关性热图对应 xy 两个方向(最多两个方向: group1, group2), group1 对应的变量为 Gene1、Gene2、Gene3、Gene4、Gene5; group2 对应的变量为 Gene6、Gene7、Gene8、Gene9、Gene10
- 此时的 group1、group2 在图形中代表 xy 两个不同的方向, 而各分组对应的变量代表 xy 两个方向上变量



- 第 1 列为分类类型, 表示分面/分组, 表示将数据分成多少个组 (对应热图的分面), 分组后的数据分别按照注释信息以 xy 两个方向分割成其对应的变量并进行相关性分析



■ 第2列及以后为数值类型数据，表示每个变量/样本值

➤ 数据清洗：对除了第1行注释信息外的非字符和除第1列外非数值的数据进行清洗

➤ 数据处理：

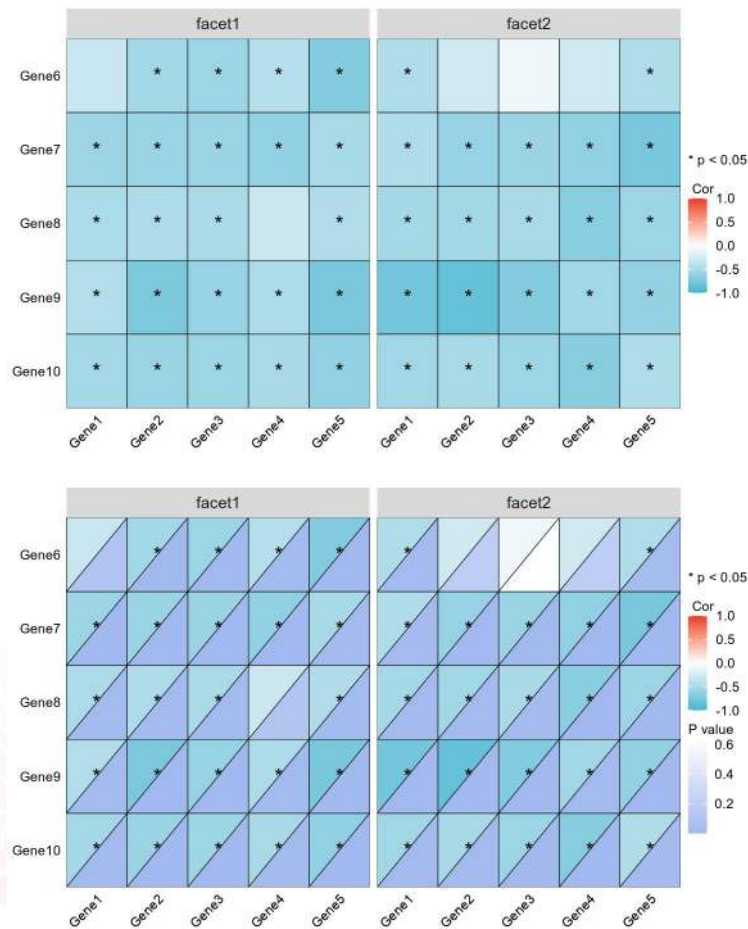
■ 根据上传数据第一行的 #注释信息将数据进行分组（除了第1列），此数据中将数据分成两个组(group1, group2)，其中 group1 对应的变量为 Gene2、Gene4、Gene5、Gene7；group2 对应的变量为 Gene1、Gene3、Gene6、Gene8、Gene9、Gene10

■ 将各分组中的所有变量分成 xy 两个方向，并将各分组中两个方向上的变量进行两两间相关性分析，结果如下：（[可查看模块补充结果部分](#)）

	A	B	C	D	E	F	G
1		Gene6	Gene7	Gene8	Gene9	Gene10	
2	Gene1	-0.327838828	-0.573870574	-0.493284493	-0.459096459	-0.52991453	
3	Gene2	-0.542124542	-0.590964591	-0.477411477	-0.742979243	-0.584249084	
4	Gene3	-0.577533578	0.584249084	0.498778999	0.613553114	0.572039072	
5	Gene4	-0.446886447	-0.636141636	-0.315018315	-0.48962149	-0.504884005	
6	Gene5	-0.708180708	-0.512210012	-0.464590965	-0.758852259	-0.631257631	
7							
8							
9							
10							
11							
12							
13							
14							
15							

➤ 将分析所得结果进行可视化

## 结果解读



上图：（完整热图）

- 行、列都代表变量/样本（分别代表不同方向对应的变量，[按照上传数据第 1 行#注释信息来划分](#)）
- 每一个小方格表示变量之间的相关系数，颜色越深，变量间越相关
- 两个热图表示两个不同的分面（表示不同的分组，[对应上传数据第 1 列的分组数据](#)，有多少个不同的分类就有多少个分面）

下图：（完整热图-三角对角线）

- 行、列都代表变量/样本，...[同上](#)....
- 每一个小方格分为两个部分（三角形），上部分表示变量间的相关系数，颜色越深，变量间越相关；下部分表示 p 值，颜色越深 p 值越小

- 两个热图表示两个不同的分面.....[同上](#)...

补充:

- \* 表示  $Pvalue < 0.05$
- Correlation 代表相关性系数(Correlation):  $|Correlation|$  越大, 变量间相关性越高, 反之相关性越低
  - $Correlation < 0$ , 变量间呈负相关关系
  - $Correlation = 0$ , 变量间没有相关关系
  - $Correlation > 0$ , 变量间呈正相关关系
- Pvalue 代表变量间相关系数对应的 Pvalue 值, Pvalue 值越小, 变量间相关系数越显著





## 数据格式

	A	B	C	D	E	F	G	H	I	J	K
1	#group	group1	group1	group1	group1	group1	group2	group2	group2	group2	group2
2	facet	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
3	facet1	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557	0.518298418	0.739705603
4	facet2	-0.01368477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841	0.981862233	0.568852396
5	facet2	-0.01182029	0.141322045	0.633965878	-0.263885834	0.542861073	0.490195153	1.183083571	0.987716302	1.067240227	1.033805987
6	facet2	-0.004488363	-0.058815726	0.586873137	0.222179483	0.179198701	0.44559659	0.548951482	0.965296724	1.13184169	0.354723611
7	facet1	-0.210829963	0.369931219	0.547144016	0.36844285	0.472425386	0.854233325	1.18447251	1.123654746	1.262049733	0.943975343
8	facet2	0.642966409	0.355974738	0.671350067	-0.173697559	-0.032966	0.767416042	1.249694698	0.595296227	0.60783876	0.513691745
9	facet2	-0.179486354	0.196057306	0.168928365	0.065455873	-0.014343253	0.448853135	1.244653087	0.678902879	1.150490547	1.088849734
10	facet1	0.119288873	-0.192532717	-0.155474752	0.501615558	-0.246108244	0.940513758	1.272317943	0.347395517	0.745167982	0.645567114
11	facet1	0.15636136	0.52661018	-0.077797179	0.380145601	0.315129994	1.260823014	1.181884192	1.258837779	0.556531639	1.120612391
12	facet1	0.625642134	0.319952123	0.246807153	0.013511406	0.392323231	1.198206731	0.861498273	0.492483526	1.168313728	1.199737413
13	facet1	0.223237827	0.46813969	-0.270452353	-0.271382299	-0.042795272	0.734466762	0.696293427	0.3255452	1.235287427	0.623165723
14	facet2	0.327989421	0.053296945	-0.054571989	0.113712744	-0.114236178	0.36837606	1.103604986	1.17269313	0.833974877	0.594458569
15	facet1	-0.138863375	0.651505674	0.145345091	0.532905728	0.306025015	0.945230436	0.440612899	1.112037256	1.013758465	0.42780374
16	facet2	0.049065953	-0.14631829	0.063868247	0.010649082	0.333018047	0.33428674	1.225104532	1.255972546	1.076337558	0.50712956

数据要求：(csv / txt 格式文件)

- 数据至少 3 列以上，每列至少 5 行，最多支持 50 列和 5000 行数据
  - 第 1 行数据需要用 #注释的数据作为 xy 两个不同的方向信息，（将不同分组中对应的变量按照 xy 两个方向进行划分，以便进行不同方向上变量间相关性分析）
  - 除第 1 行外，第 1 列为分类类型数据，表示分组信息（将所有的数据分成不同的分组（相当于将同一个变量对应的数据分成了多少份）），然后按照 xy 两个不同的方向（相当于不同的分组中对应的变量分成了两个部分，两个分组对应着不一样的变量）
  - 除第 1 行、第 1 列外，从第 2 列开始，每一列都为数值类型数据，表示一个变量/基因/分子的值
    - ◆ 不能含有非数值之外的其它字符类型
- 变量名/列名不能重复



## 参数说明

(说明: 标注了颜色的为常用参数。)

## 统计



统计

统计方法 Spearman

- 统计方法: 可以选择变量间进行相关性分析的方法
  - spearman: Spearman(默认)为非参数检验方法, 数据可以不需要满足正态性
  - pearson: Pearson 为参数检验方法, 数据需要满足双正态

## 映射

映射

上半颜色映射

相关系数

下半颜色映射

p值

大小映射

相关性系数

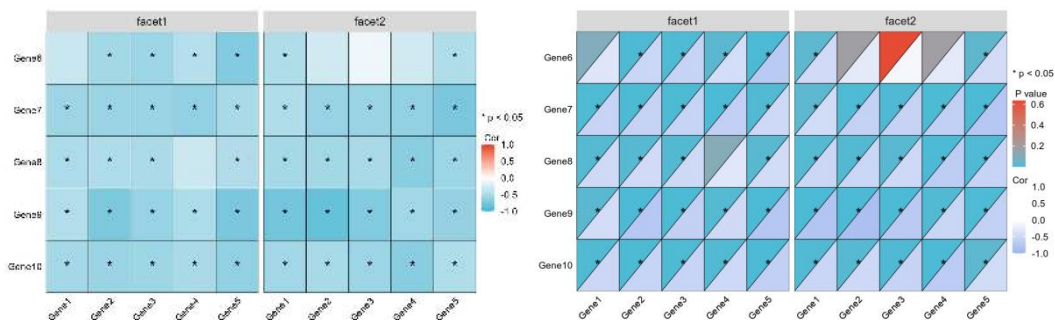
- 上半颜色映射：对应整体颜色的映射，当热图选择的是三角对角线类型时，则对应三角的上半颜色映射，如下：（上为热图类型-三角对角线；中为上半颜色映射变量相关系数，下半颜色映射变量p值；右为上半颜色映射变量p值，下半颜色映射变量相关系数

热图

类型
三角对角线

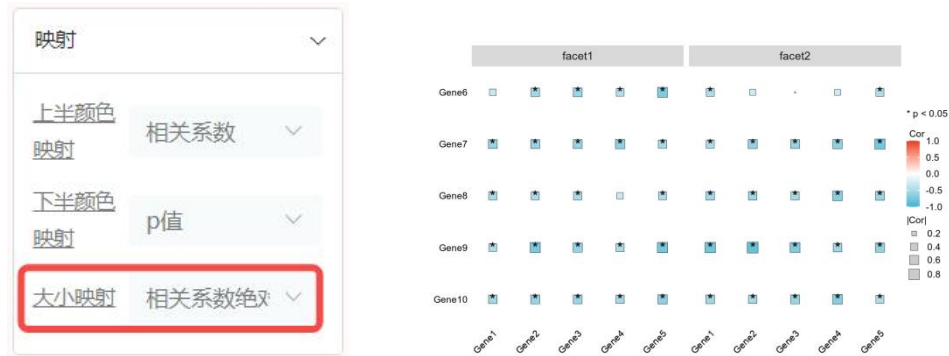
上半(全)颜色
☒ ☒

下半颜色
☒ ☐



- 下半颜色映射：当热图选择的是三角对角线类型时，则对应三角的下半颜色映射（如上：上半颜色映射）

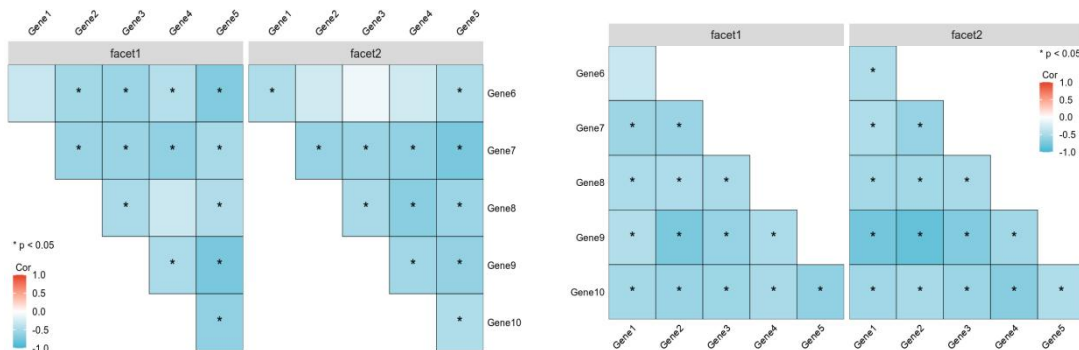
- 大小映射：可以对热图进行大小映射，只有在非三角对角线类型的时候会有效果，对应映射方块大小，默认为不映射，还可以选择相关系数绝对值，如下：



## 样式



- 类型：可以选择热图的类型，默认为完整热图，还可以选择上半热图、下半热图，如下（左侧为上半，右侧为下半）



## 热图

热图

类型

方块

上半(全)颜色

下半颜色

描边颜色

描边粗细

0.25pt

大小比例

1

不透明度

1

- 方块：可以选择热图中每个小块（一行一列）的类型，默认为方块，还可以选择圆形，三角对角线，如下：



- 上半（全）颜色：选择三角类型时可以修改对应方块上半部分颜色，其他类型则修改对应整个方块颜色

- 下半颜色：当选择三角类型时可以修改对应方块下半部分颜色，其他类型无法修改
- 描边颜色：可以修改热图对应方块的描边颜色
- 描边粗细：可以修改热图对应方块的描边粗细
- 大小比例：可以修改热图对应方块的大小比例，默认为 1
- 不透明度：可以修改热图对应方块的不透明度，默认为 1，表示完全不透明，0 表示完全透明



## 标注

标注

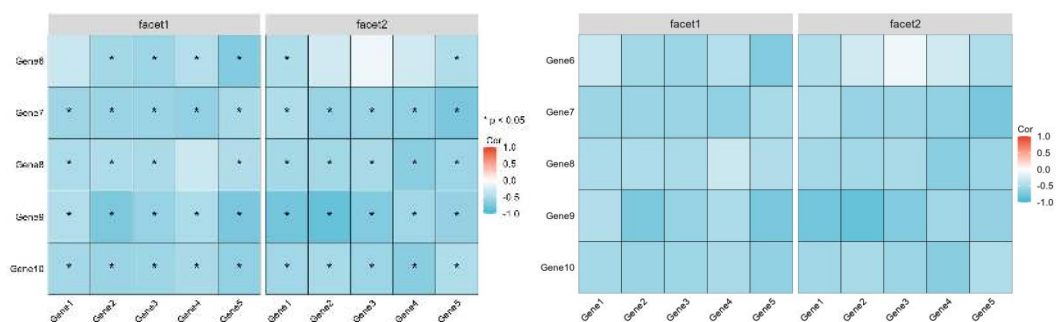
标注映射 星号(\*p<0.0

颜色

标注大小 6pt

- 标注映射：可以选择是否在热图矩阵对应的每一个小矩形上进行标注映射，默认为星号(\*p<0.05)进行标注，还可以选择：如下：（左侧为(\*p<0.05)映射，右侧为不映射）

- 星号(\*p<0.05|\*\*p<0.01)
- 星号(\*p<0.05|\*\*p<0.01|\*\*\*p<0.001)
- p 值(2 位小数)
- 相关系数(2 位小数)
- 无



- 颜色：可以选择当进行标注映射时，标注的颜色
- 标注大小：可以选择并修改标注的大小，默认为 6pt

## 分面

- 分面方向：可以选择并修改分组（分面）相关性热图进行分面的方向，默认为按列的方向，还可以选择按行进行分面（分组）
- 分面颜色：可以修改各个分面（分组）的颜色
- 文字大小：可以修改各个分面（分组）对应文本字体的大小，默认为 6pt

## 标题文本

- 大标题：大标题文本

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



The image shows a settings panel titled '图注' (Figure Annotation). It contains three main settings: '是否展示' (Whether to display) with a toggle switch turned on; '图注标题' (Figure Annotation Title) with a text input field containing '图注标题内容'; and '图注位置' (Figure Annotation Position) with a dropdown menu currently set to '默认' (Default).

- 是否展示：可以选择是否展示图注信息，默认展示
- 图注标题：可以修改图注标题内容，默认没有
- 图注位置：默认为图片的右侧，还可以选择上、下



## 坐标轴



- x 轴标注旋转：可选择并修改 x 轴对应刻度文本的旋转角度

## 风格



- 网格：可以选择是否展示网格，默认不展示
- 文字大小：控制整体文字大小，默认为 6pt

## 图片

图片	▼
宽度 (cm)	10
高度 (cm)	6
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



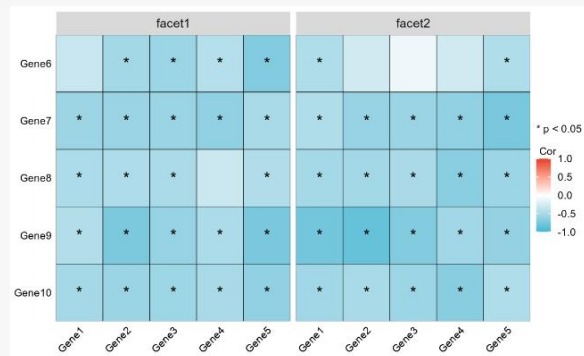
## 结果说明

## 主要结果

### 相关性热图-分组XY

相关性热图-分组XY: 用热图展示相关性情况

统计方法: spearman



[相关性热图-分组XY.pdf](#)

[相关性热图-分组XY.tiff](#)

[相关性热图-分组XY.pptx](#)



## 补充结果

facet1-相关系数表格					
	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1	-0.328	-0.574	-0.493	-0.459	-0.53
Gene2	-0.542	-0.591	-0.477	-0.743	-0.584
Gene3	-0.578	-0.584	-0.499	-0.614	-0.572
Gene4	-0.447	-0.636	-0.315	-0.49	-0.505
Gene5	-0.708	-0.512	-0.465	-0.759	-0.631

facet2-相关系数表格					
	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1	-0.475	-0.474	-0.531	-0.782	-0.551
Gene2	-0.284	-0.607	-0.54	-0.849	-0.514
Gene3	-0.104	-0.585	-0.504	-0.71	-0.582
Gene4	-0.283	-0.635	-0.681	-0.555	-0.686
Gene5	-0.469	-0.756	-0.58	-0.621	-0.474

这里提供相关性分析表：可以查看变量之间的相关系数

- 相关系数为正数，说明两个分子（主要分子与其他分子）之间可能存在正相关关系；相关系数为负数，说明两个分子可能存在负相关关系
  - 相关系数绝对值在 0.8-1.0 之间，说明两个分子之间强相关
  - 相关系数绝对值在 0.5-0.8 之间，说明两个分子之间中等程度相关
  - 相关系数绝对值在 0.3-0.5 之间，说明两个分子之间相关程度一般
  - 相关系数绝对值在 0.0-0.3 之间，说明两个分子之间弱相关或者不相关

facet1-相关性检验(p值)表格					
	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1	0.0954	0.00209	0.0097	0.0169	0.00504
Gene2	0.00399	0.00144	0.0126	1.58e-05	0.00167
Gene3	0.00193	0.00167	0.00883	0.000852	0.00218
Gene4	0.0204	0.000484	0.11	0.0103	0.00794
Gene5	5.66e-05	0.00697	0.0155	8.54e-06	0.000549

facet2-相关性检验(p值)表格					
	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1	0.0231	0.0234	0.0102	1.59e-05	0.00722
Gene2	0.189	0.00262	0.00879	1.84e-06	0.0132
Gene3	0.636	0.00398	0.0153	0.000216	0.0042
Gene4	0.191	0.00144	0.00049	0.00675	0.000431
Gene5	0.0251	4.71e-05	0.00436	0.00198	0.0234

[相关性.xlsx](#)

相关性系数表格:

- 表中包含了各个变量间的相关系数(r)值, 相关系数一般是 -1到1 之间, 正负号表示正相关和负相关, 系数绝对值大小表示相关性大小
- 一般关系强度是:  $|r| > 0.95$ : 显著性相关;  $|r| \geq 0.8$ : 高度相关;  $0.5 \leq |r| < 0.8$ : 中度相关;  $0.3 \leq |r| < 0.5$ : 低度相关;  $|r| < 0.3$ : 弱相关

相关性p值表格:

- 表中包含了各个变量间的相关性的检验p值

这里提供相关性分析表: 可以查看各个变量间的相关性的检验 p 值

➤ p 值表示检验 p 值

## 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：ggplot2 包（用于可视化）

处理过程：

(1) 对数据各个组的变量进行两两相关性分析，分析结果用热图进行可视化





## 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. 方法里面的 Spearman 和 Pearson 方法，应该选择哪一个？

答：两种方法均可以选择。Pearson 会要求数据是满足正态性，Spearman 因为是非参数的方法，可以不需要满足。可以先选择非参数的 Spearman 相关进行尝试。

### 2. 图的内容被压缩了，如何处理？

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 增加图片高度；
- ② 减少分子列表中的分子。

### 3. 相关系数多少为好？

答：这个没有很统一的标准，可以参考以下：

- 相关系数强弱：
  - 绝对值在 0.8 以上：强相关
  - 绝对值在 0.5-0.8：中等程度相关
  - 绝对值在 0.3-0.5：相关程度一般
  - 绝对值在 0.3 以下：弱或者不相关