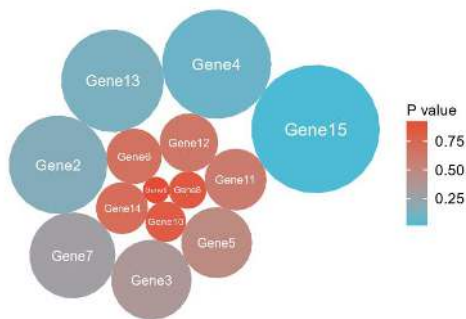


## 交互网络 - 相关性气泡图



网址: <https://www.xiantao.love>



更新时间: 2023.06.16

## 目录

|             |    |
|-------------|----|
| 基本概念 .....  | 3  |
| 应用场景 .....  | 3  |
| 分析过程 .....  | 3  |
| 结果解读 .....  | 5  |
| 数据格式 .....  | 6  |
| 参数说明 .....  | 7  |
| 统计 .....    | 7  |
| 映射 .....    | 8  |
| 图形样式 .....  | 10 |
| 气泡（点） ..... | 11 |
| 标注 .....    | 12 |
| 标题文本 .....  | 13 |
| 图注 .....    | 13 |
| 风格 .....    | 14 |
| 图片 .....    | 14 |
| 结果说明 .....  | 15 |
| 主要结果 .....  | 15 |
| 方法学 .....   | 16 |
| 如何引用 .....  | 17 |
| 常见问题 .....  | 18 |

## 基本概念

- 紧凑气泡图：散点图的一种，通过气泡/点的形式来展示数据的情况
- 相关性气泡图：分析 1 个主要变量和其他所有变量之间的相关性结果并用紧凑气泡图形式展示结果

## 应用场景

通过气泡/点的形式来展示主要变量与其他变量之间的相关性

## 分析过程

上传数据 → 数据处理(清洗) → 相关性分析 → 可视化

- 数据格式：（具体数据格式要求可以看后面过程的“数据格式”部分）
  - 数据每一列都代表一个变量/样本，都需要是数值类型的数据
  - ◆ 数据第 1 列表示主要变量
  - ◆ 数据第 2 列直至以后都是其他变量，对应紧凑气泡图各个点（气泡）

|    | A           | B           | C           | D           | E           | F           | G           | H           |
|----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1  | Gene1       | Gene2       | Gene3       | Gene4       | Gene5       | Gene6       | Gene7       | Gene8       |
| 2  | 0.454560567 | 0.961821065 | 0.386559976 | 0.372762927 | 0.811831074 | 0.295180814 | 0.784694036 | 0.228529026 |
| 3  | 0.95379953  | 0.704024889 | 0.160149079 | 0.850969005 | 0.318387697 | 0.800720281 | 0.041289769 | 0.527950544 |
| 4  | 0.08349129  | 0.55383833  | 0.445131116 | 0.791807708 | 0.646856103 | 0.104350423 | 0.577682909 | 0.861875697 |
| 5  | 0.926924902 | 0.493519535 | 0.279659963 | 0.494638234 | 0.609092163 | 0.753749829 | 0.307407179 | 0.328881482 |
| 6  | 0.295565529 | 0.835540217 | 0.952745714 | 0.928768046 | 0.465357961 | 0.768919581 | 0.582334603 | 0.961628989 |
| 7  | 0.515936849 | 0.998347631 | 0.005278934 | 0.014454146 | 0.089728537 | 0.868722979 | 0.33986496  | 0.673254072 |
| 8  | 0.234907919 | 0.084901414 | 0.785495181 | 0.702400689 | 0.117291042 | 0.768241607 | 0.538662346 | 0.757992964 |
| 9  | 0.558606646 | 0.69384991  | 0.438819759 | 0.360955941 | 0.096197276 | 0.012697684 | 0.524381363 | 0.792684477 |
| 10 | 0.754155403 | 0.613915667 | 0.881552035 | 0.903247918 | 0.332448788 | 0.492185116 | 0.785448442 | 0.239907264 |
| 11 | 0.388997389 | 0.414784534 | 0.875784324 | 0.611702966 | 0.769772245 | 0.583381805 | 0.012425477 | 0.575839543 |
| 12 | 0.286723127 | 0.425758032 | 0.001392136 | 0.744544761 | 0.890906309 | 0.091761208 | 0.963916105 | 0.412877348 |

➤ 数据处理

■ 对数据中每一列非数值类型的数据进行处理

◆ 所有变量/列都需要纯数值类型的数据

◆ 不能有非数值，特殊值(特殊符号等)，并且每一个变量不能都是一个值

➤ 相关性分析：

■ 将处理(清洗)后的数据进行相关性分析

◆ 主要变量（数据第 1 列）与其它变量（数据第 2 列开始的列）之间

◆ 相关性分析表

- 包含不同方法（Pearson、Spearman）计算的相关性系数值与统计学 p 值等

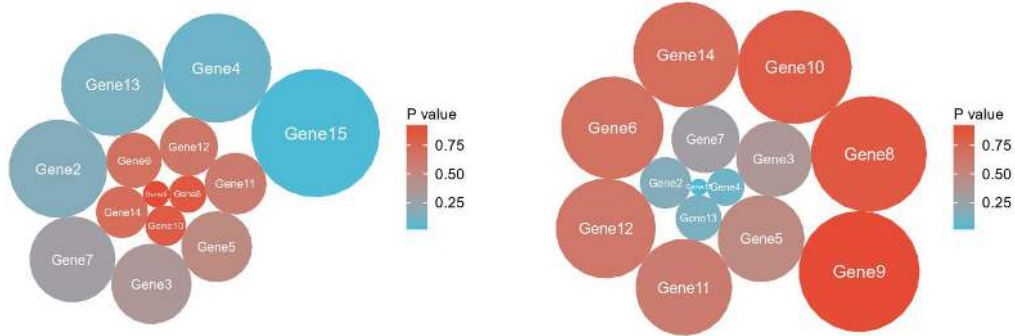
| 相关性分析                     |        |         |             |              |            |              |               |
|---------------------------|--------|---------|-------------|--------------|------------|--------------|---------------|
| 提供Pearson和Spearman统计方法的结果 |        |         |             |              |            |              |               |
| 主变量                       | 次变量    | 自由度(df) | 统计量-Pearson | 相关系数-Pearson | p值-Pearson | 统计量-Spearman | 相关系数-Spearman |
| Gene1                     | Gene2  | 98      | 1.46548     | 0.14644      | 0.1460     | 1.433e+05    | 0.139874      |
| Gene1                     | Gene3  | 98      | -1.0911     | -0.109554    | 0.2779     | 1.821e+05    | -0.0928053    |
| Gene1                     | Gene4  | 98      | -1.88289    | -0.186851    | 0.0627     | 1.954e+05    | -0.172517     |
| Gene1                     | Gene5  | 98      | 0.606654    | 0.0611665    | 0.5455     | 1.546e+05    | 0.0722832     |
| Gene1                     | Gene6  | 98      | -0.43411    | -0.0438097   | 0.6652     | 1.74e+05     | -0.0439364    |
| Gene1                     | Gene7  | 98      | -0.958806   | -0.0964029   | 0.3400     | 1.844e+05    | -0.106595     |
| Gene1                     | Gene8  | 98      | 0.222976    | 0.0225183    | 0.8240     | 1.633e+05    | 0.019922      |
| Gene1                     | Gene9  | 98      | -0.0309549  | -0.0031269   | 0.9754     | 1.683e+05    | -0.00976898   |
| Gene1                     | Gene10 | 98      | 0.0559933   | 0.00565608   | 0.9555     | 1.628e+05    | 0.0232943     |
| Gene1                     | Gene11 | 98      | 0.56414     | 0.0568944    | 0.5739     | 1.575e+05    | 0.0547855     |
| Gene1                     | Gene12 | 98      | -0.463974   | -0.0468171   | 0.6437     | 1.748e+05    | -0.0489889    |
| Gene1                     | Gene13 | 98      | 1.61035     | 0.16056      | 0.1105     | 1.413e+05    | 0.152031      |

相关性.xlsx

相关系数为正，说明两个变量之间存在正相关关系；相关系数为负，说明两个变量之间存在负相关关系；  
 相关系数绝对值代表相关程度，0-0.3代表弱或者不相关；0.3-0.5代表弱相关；0.5-0.8代表中等程度相关；0.8-1代表强相关  
 相关是否有统计学意义还需要结合p值来查看

➤ 将分析后得到的相关性系数与 p 值进行后续的相关性气泡图可视化

## 结果解读



- 每个点（气泡）表示各样本/基因（上传数据第 2 列开始每一列）对应的具体值（主要变量（数据第 1 列）与其他变量（数据第二列开始）之间的（相关系数绝对值（左侧）/p 值（右侧）））
- 点（气泡）的大小对应值的大小（相关系数（左侧）/p 值（右侧））的大小
- 不同的颜色对应不同的分组/值的大小（可选择映射的具体内容）

## 数据格式

|    | A           | B           | C           | D           | E           | F           | G           | H           |
|----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1  | Gene1       | Gene2       | Gene3       | Gene4       | Gene5       | Gene6       | Gene7       | Gene8       |
| 2  | 0.454560567 | 0.961821065 | 0.386559976 | 0.372762927 | 0.811831074 | 0.295180814 | 0.784694036 | 0.228529026 |
| 3  | 0.95379953  | 0.704024889 | 0.160149079 | 0.850969005 | 0.318387697 | 0.800720281 | 0.041289769 | 0.527950544 |
| 4  | 0.08349129  | 0.55383833  | 0.445131116 | 0.791807708 | 0.646856103 | 0.104350423 | 0.577682909 | 0.861875697 |
| 5  | 0.926924902 | 0.493519535 | 0.279659963 | 0.494638234 | 0.609092163 | 0.753749829 | 0.307407179 | 0.328881482 |
| 6  | 0.295565529 | 0.835540217 | 0.952745714 | 0.928768046 | 0.465357961 | 0.768919581 | 0.582334603 | 0.961628989 |
| 7  | 0.515936849 | 0.998347631 | 0.005278934 | 0.014454146 | 0.089728537 | 0.868722979 | 0.33986496  | 0.673254072 |
| 8  | 0.234907919 | 0.084901414 | 0.785495181 | 0.702400689 | 0.117291042 | 0.768241607 | 0.538662346 | 0.757992964 |
| 9  | 0.558606646 | 0.69384991  | 0.438819759 | 0.360955941 | 0.096197276 | 0.012697684 | 0.524381363 | 0.792684477 |
| 10 | 0.754155403 | 0.613915667 | 0.881552035 | 0.903247918 | 0.332448788 | 0.492185116 | 0.785448442 | 0.239907264 |
| 11 | 0.388997389 | 0.414784534 | 0.875784324 | 0.611702966 | 0.769772245 | 0.583381805 | 0.012425477 | 0.575839543 |
| 12 | 0.286723127 | 0.425758032 | 0.001392136 | 0.744544761 | 0.890906309 | 0.091761208 | 0.963916105 | 0.412877348 |

数据要求：

- 数据至少 2 列以上，每列至少 5 行，最多支持 30 列和 5000 行数据
  - 每个变量都需要是数值类型的数据
  - 不能有非数值，特殊值(特殊符号等)
  - 一个样本/变量不能全部为一样的值
  - 每一列列名对应热图的行、列名。图中各变量的顺序与上传数据中各变量的顺序保持一致，若需要调整图中各分类的顺序，需要在上传数据内进行调整，然后再上传数据
  - 每一行可以代表在不同样本对应的列变量的情况。例如表达谱数据
- 样本名不能重复(列名不能重复)

## 参数说明

(说明：标注了颜色的为常用参数。)

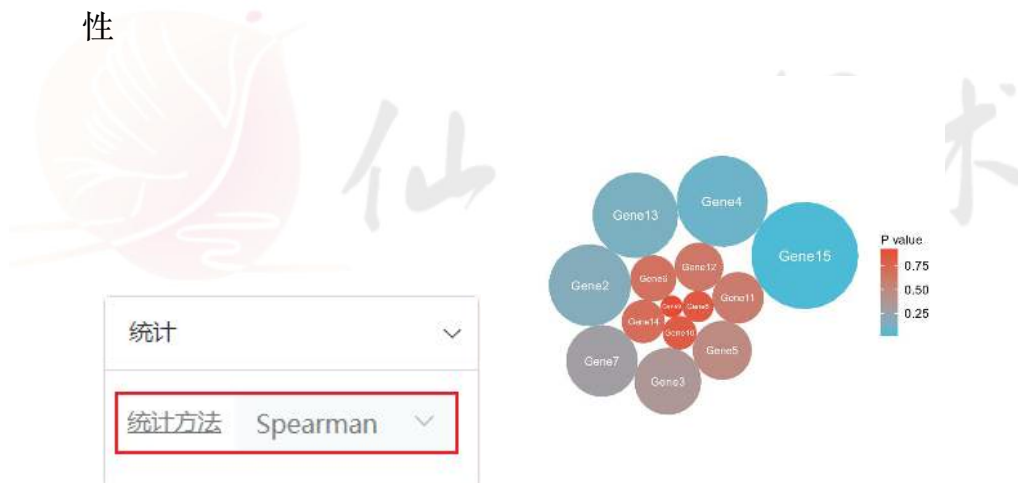
## 统计

统计

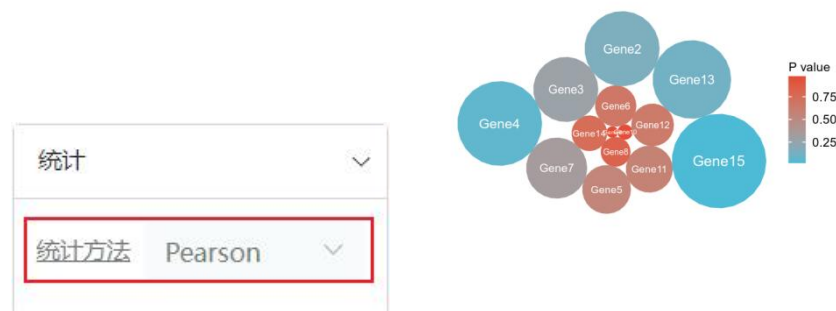
统计方法 Spearman

➤ 统计：可以选择主要变量与其他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 ▼

颜色映射 p值 ▼

大小映射 相关系数绝对值 ▼

- 颜色映射：根据上传数据特点，可以选择是否对紧凑气泡图进行颜色映射，如下：





映射

颜色映射 p值-分类

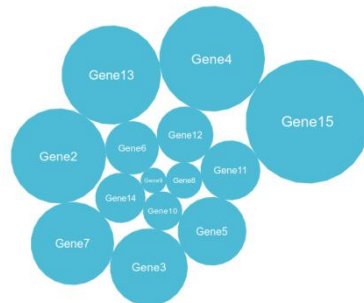
大小映射 相关系数绝对值



映射

颜色映射 不映射

大小映射 相关系数绝对值

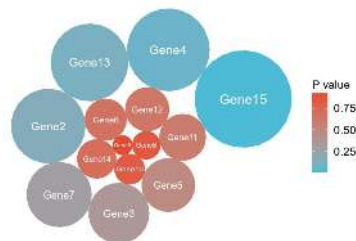


- 大小映射：根据上传数据特点，可以选择是否对紧凑气泡图进行大小映射，如下：

映射

颜色映射 p值

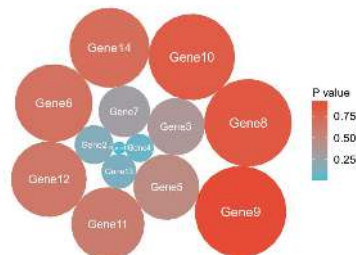
大小映射 相关系数绝对值



映射

颜色映射 p值

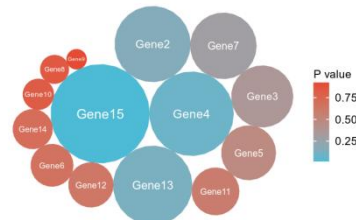
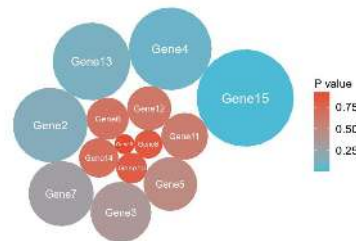
大小映射 p值



## 图形样式



- 样式：可以选择并修改紧凑气泡图螺旋的排序方式，默认表示根据上传数据的特点，不进行任何顺序操作，如下：



## 气泡（点）



- 填充颜色：可以修改图中各点（气泡）的填充颜色
- 描边颜色：可以修改图中各点（气泡）的描边颜色
- 描边粗细：可以修改图中各点（气泡）的描边粗细
- 不透明度：可以修改图中各点（气泡）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 ▼

颜色 ☐

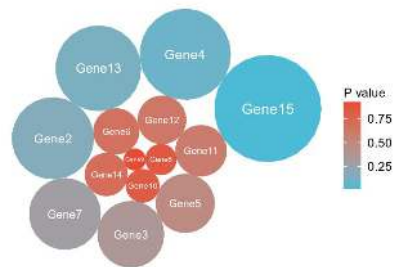
标注大小 6pt ▼

- 颜色：可以选择标注的颜色，如下：

标注 ▼

颜色 ☐

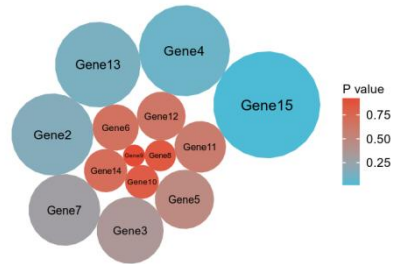
标注大小 6pt ▼



标注 ▼

颜色 ▼

标注大小 6pt ▼



- 标注大小：可以选择标注的大小，默认为 6pt

## 标题文本

标题

大标题 大标题内容

- 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

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：可以选择是否展示各指标映射的内容，即图注，默认展示、
- 图注标题：首先选择展示图注，可以修改图注的标题
- 图注位置：首先选择展示图注，可以选择图注的位置，默认表示默认展示在右侧，还可以选择上

## 风格

|      |       |
|------|-------|
| 风格   | ▼     |
| 文字大小 | 6pt ▼ |

- 文字大小：控制整体文字大小，默认为 6pt

## 图片

|         |         |
|---------|---------|
| 图片      | ▼       |
| 宽度 (cm) | 6       |
| 高度 (cm) | 6       |
| 字体      | Arial ▼ |

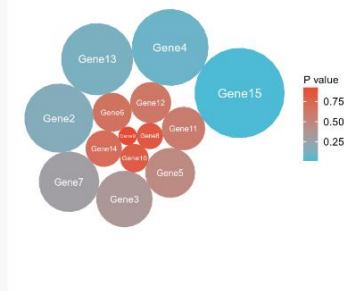
- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体

## 结果说明

## 主要结果

### 相关性气泡图

相关性气泡图: 散点图的一种, 分析1个主变量和其他所有变量之间的相关性结果并用气泡/点图形式展示结果



相关性气泡图.pdf

相关性气泡图.tiff

相关性气泡图.pptx



## 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：packcircles 包（用于计算数据），ggplot2 包（用于可视化）

处理过程：

- (1) 将清洗后的数据进行相关性分析
- (2) 用 packcircles 包对分析得到的数据进行处理得到绘制紧凑气泡图的数据
- (3) 用 ggplot2 包对处理后的数据进行紧凑气泡图可视化



## 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](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. 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