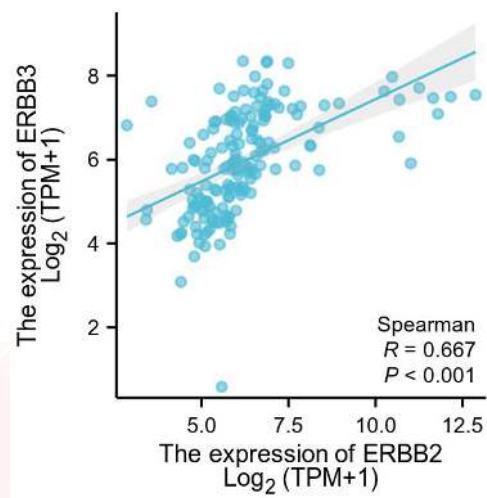


## 交互网络 - 相关性散点图[云]



网址: <https://www.xiantao.love>



更新时间: 2023.02.28

## 目录

基本概念 .....	3
应用场景 .....	3
分析过程 .....	3
结果解读 .....	7
数据格式 .....	8
参数说明 .....	9
数据处理 .....	9
点 .....	12
拟合线 .....	14
标题文本 .....	15
风格 .....	16
图片 .....	16
结果说明 .....	17
主要结果 .....	17
补充结果 - 统计描述 .....	18
补充结果 - 异常值分析 .....	18
补充结果 - 统计描述 .....	19
补充结果 - 统计描述 .....	19
方法学 .....	21
如何引用 .....	22
常见问题 .....	23

## 基本概念

- 散点图：通过点的形式来展示数据的分布情况
- 相关性散点图：分析 1 个变量和另外 1 个或者 2 个变量之间的相关性

## 应用场景

基于云端数据 通过点、线的形式来展示变量与变量之间的相关性

## 分析过程

云端数据 → 数据处理(清洗) → 相关性分析 → 可视化

- 云端数据：提供预清洗好的云端数据，不同平台的云端数据集的分子可能会有不同。 注意：选择了不同的平台，搜索出来的分子可能是不一样的数
- 分析
  - 数据准备
    - ◆ 通过特殊参数[主分子]、[补充分子]选择云端数据中需要进行分析的主要分子/变量和补充分子(其他分子/变量)，将主分子分别与补充分子进行相应分析

	A	B
1	ERBB2	ERBB3
2	12.16014207	7.499810816
3	2.855630972	6.819097174
4	6.509850996	7.597660034
5	8.122366776	6.345046802
6	6.953466634	7.300066968
7	5.585824854	6.280254833
8	3.562620987	7.385010765
9	5.816671379	6.457714118
10	5.96972841	7.140872706
11	5.270794284	6.159869319
12	6.454577317	6.676258944
13	5.939976415	6.886695239

- **主分子**：表示主要分子/变量 [ERBB2]

#### 特殊参数

主分子 ⓘ ERBB2[ENSG00000141736.14] 补充分子 ⓘ ERBB3[ENSG00000065361.16] ×

- **补充分子**：表示其他分子/变量 [ERBB3]

#### 特殊参数

主分子 ⓘ ERBB2[ENSG00000141736.14] 补充分子 ⓘ ERBB3[ENSG00000065361.16] ×

- 补充分子最多支持 2 个（也就是说在分析数据中，数据不能超过 3 列（主分子+补充分子）），如下：

#### 特殊参数

主分子 ⓘ ERBB2[ENSG00000141736.14] 补充分子 ⓘ ERBB3[ENSG00000065361.16] ×  
OAZ1[ENSG00000104904.12] ×  
TSPAN6[ENSG00000000003.15] ×

错误信息  
传入数据超过3列数据

确定

## 分析

### 统计描述

- 对主分子和补充分子进行常见统计描述指标统计分析

#### 统计描述

各个组对应常见「统计描述指标」

组别	数目	最小值	最大值	中位数(Median)	四分位距(IQR)	下四分位	上四分位	均值(Mean)	标准差(SD)	标准误(SE)
ERBB2	163	2.8556	12.878	5.94	1.4394	5.2574	6.6968	6.2346	1.6416	0.12858
ERBB3	163	0.57792	8.3489	5.9378	1.9775	5.0186	6.996	5.965	1.236	0.096812

### 正态性检验

- 对第主分子和补充分子列进行正态性检验

#### 正态性检验(Shapiro-Wilk normality test)

组别	自由度(df)	统计量	p值
ERBB2	163	0.83968	4.4e-12
ERBB3	163	0.96672	0.0006

正态性检验结果显示, 存在有不满足正态分布的变量( $P < 0.05$ ), 建议选择用 非参数检验方法(Spearman)

### 异常值分析

- 对变量进行异常值分析

#### 异常值分析

离群值 =  $Q1(\text{下四分位}) - 1.5 * IQR(\text{四分位间距})$  或者  $Q3(\text{上四分位}) + 1.5 * IQR(\text{四分位间距})$

异常值 =  $Q1(\text{下四分位}) - 3.0 * IQR(\text{四分位间距})$  或者  $Q3(\text{上四分位}) + 3.0 * IQR(\text{四分位间距})$

组别	离群值	异常值
ERBB2	12.160142068914, ...	12.160142068914, ...
ERBB3	0.577924244506702	

各组离群值和异常值如上所示, 如数据确认非人为记录错误, 可不进行处理

### 相关性分析: 将主分子和补充分子进行两两间相关性分析

- 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

同时提供Pearson和Spearman统计方法，可以根据需要选择标注在图中的方法

方法	组别I	组别J	自由度(df)	统计量	相关系数	置信区间(95%CI)	p值
Pearson	ERBB2	ERBB3	163	7.7469	0.5211	0.39937 - 0.62477	9.94e-13
Spearman	ERBB2	ERBB3	163	2.405e+05	0.66685		0

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；

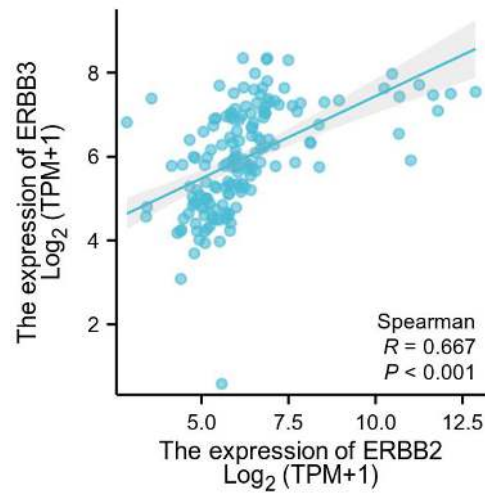
相关系数绝对值代表相关程度，0-0.3代表弱或者不相关；0.3-0.5代表弱相关；0.5-0.8代表中等程度相关；0.8-1代表强相关

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热图可视化



## 结果解读



- 横坐标表示主要分子的表达水平
- 纵坐标表示补充分子的表达水平
- 图中的线为拟合线，拟合线周围的阴影部分为置信区间
- 图中右下角
  - “Spearman”表示主分子与补充分子间进行相关性分析的方法
  - “R”表示主分子与补充分子间的相关性系数
  - “P”表示主分子与补充分子间的统计学 p 值

## 数据格式

提供预清洗好的云端数据，不同平台的云端数据集的分子可能会有不同。注意：选择了不同的平台，搜索出来的分子可能是不一样的





## 参数说明

(说明：标注了颜色的为常用参数。)

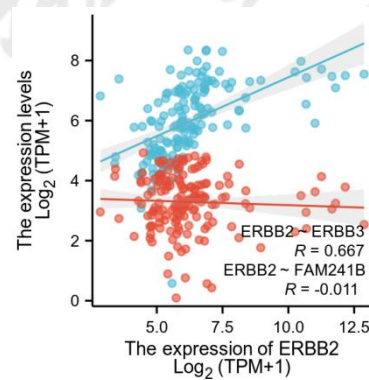
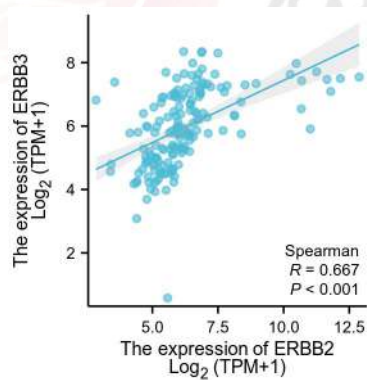
## 特殊参数

特殊参数
重置参数

主分子 ①

补充分子 ①

- 主分子：可以选择云端数据中进行分析时的主要分子/变量/基因
- 补充分子：可以选择云端数据中进行分析时的补充分子
  - 补充分子的个数不能超过 2 个，如下：左侧为一个补充分子，右侧为两个补充分子



## 主要参数

- 主要参数为分析以及可视化相关参数，如下[主要参数]

## 主要参数

### 统计

统计 ▼

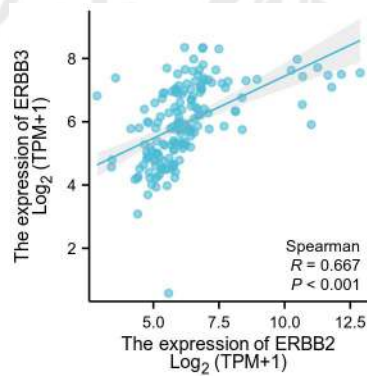
统计方法 Spearman ▼

标注位置 右下 ▼

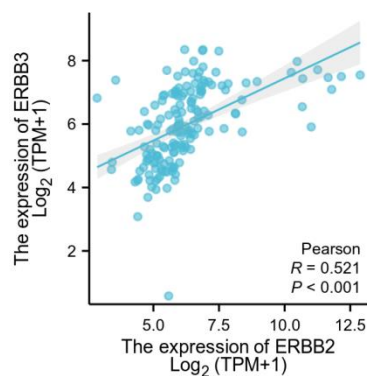
标注颜色 ■

➤ 统计方法：可以选择主要分子与其他分子间进行相关性分析的方法

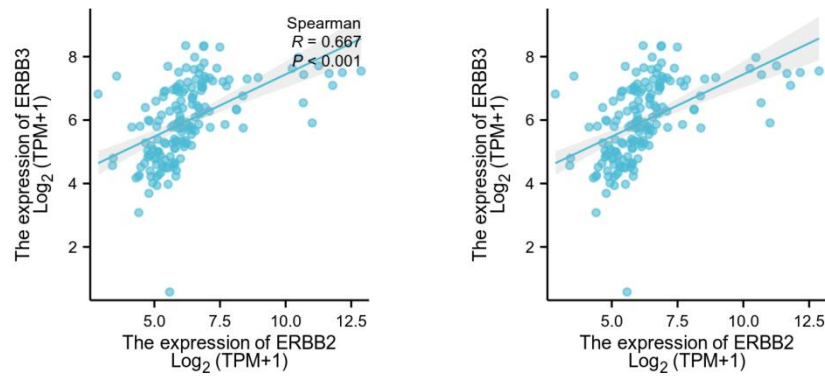
- spearman: Spearman(默认)为非参数检验方法，数据可以不需要满足正态性



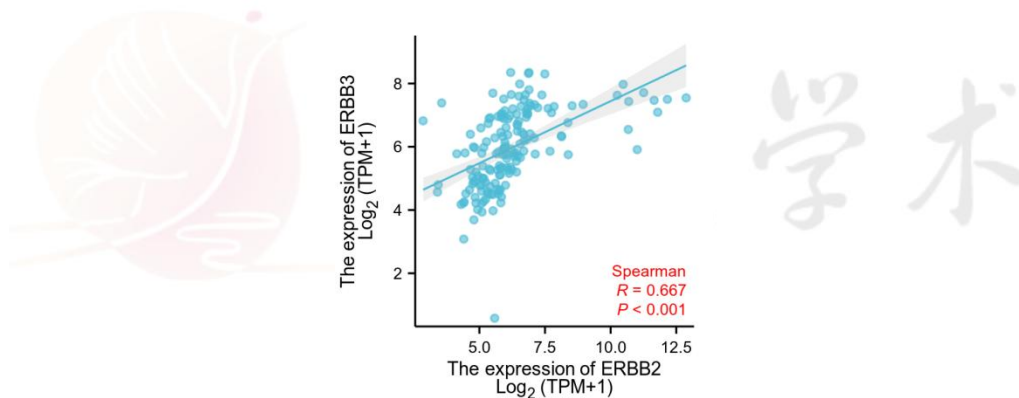
- pearson: Pearson 为参数检验方法，数据需要满足双正态



- 标注位置：可以修改图中相关性分析方法(Spearman)、相关性系数(R)，统计学 p 值的位置，默认在图形的右下，还可以选择左下、左上、右上、无(不进行标注)，如下：左侧为右上，右侧为无



- 标注颜色：当图形中进行标注的时候，可以修改标注的颜色，如下：

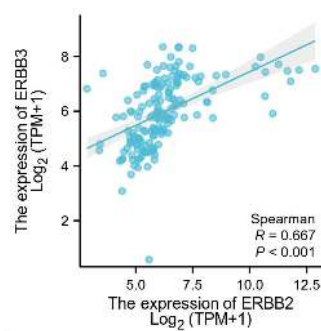


## 样式

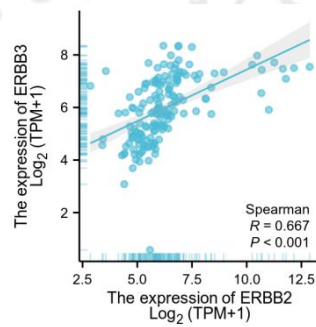


➤ 样式：可以选择图形的样式，默认为经典，还可以选择如下：

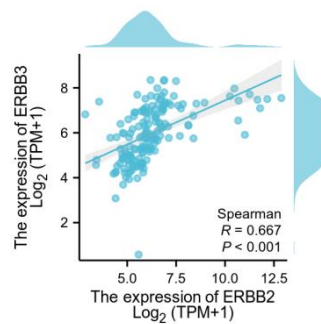
■ 经典：表示经典相关性散点图



■ 加分布竖线：表示以竖线的形式加在坐标轴两侧



■ 两侧加分布图：表示以数据分布图的形式加在坐标轴两侧



## 点

点

填充色

描边色

样式

圆形

×

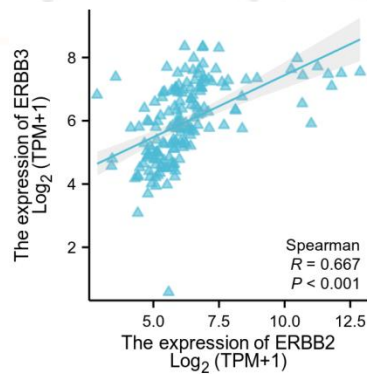
大小

1

不透明度

0.6

- 填充色：可以修改图中各点的填充颜色
- 描边色：可以修改图中各点的描边颜色
- 样式：可以修改图中各点的样式（形状），默认为圆形，还可以选择正方形、菱形、三角形、倒三角形，如下：



- 大小：可以修改图中个点的大小比例，默认为 1
- 不透明度：可以修改图中各点的不透明度，1 表示完全不透明,0 表示完全透明

## 拟合线

拟合线

展示

拟合方法

直线

拟合线颜色

拟合线样式

实线

线条粗细

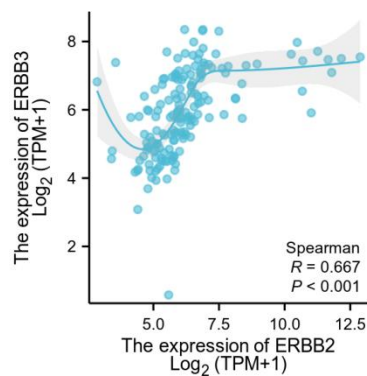
0.75pt

置信区间展示

不透明度

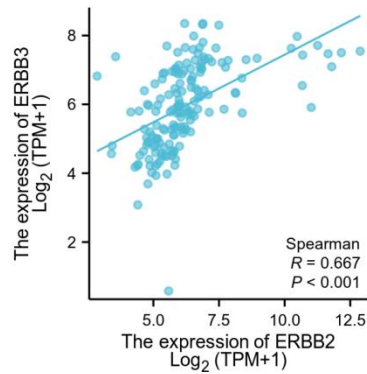
0.2

- 展示：可以选择是否进行展示拟合线的操作，默认为展示
- 拟合方法：可以修改图中拟合部分的拟合方法(类型)，默认为直线，还可以选择曲线的形式，如下：



- 拟合线颜色：可以修改图中拟合线的颜色
- 拟合线样式：可以修改图中拟合线的样式，默认为实线，还可以选择虚线
- 线条粗细：可以选择修改图中拟合线的线条粗细

- 置信区间展示：可以选择是否展示拟合线的置信区间（阴影部分），默认为展示，还可以选择不展示，如下：



- 不透明度：可以修改拟合线线条的不透明度，1 表示完全不透明，0 表示完全透明

## 标题文本

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]]

## 风格



- 边框：可以选择是否展示图片边框，默认展示
- 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 7pt

## 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



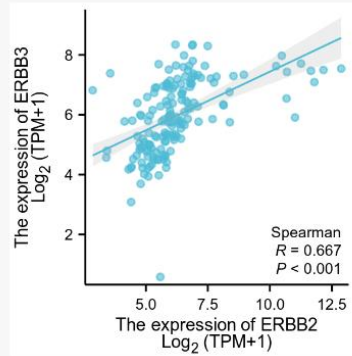
## 结果说明

## 主要结果

### 表达相关性散点图

相关性散点图: 分析1个变量和另外1个或者2个变量之间的相关性

统计方法: Spearman



[相关性散点图.pdf](#)

[相关性散点图.tiff](#)

[相关性散点图.pptx](#)

[分析数据.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

相关是否有统计意义还需要结合p值来查看

## 补充结果 - 统计描述

**统计描述**

各个组对应常见「统计描述指标」

组别	数目	最小值	最大值	中位数(Median)	四分位距(IQR)	下四分位	上四分位	均值(Mean)	标准差(SD)	标准误(SE)
ERBB2	163	2.8556	12.878	5.94	1.4394	5.2574	6.6968	6.2346	1.6416	0.12858
ERBB3	163	0.57792	8.3489	5.9378	1.9775	5.0186	6.996	5.965	1.236	0.096812

[统计描述.xlsx](#)

这里提供各个分子对应常见「统计描述指标」：最小值、最大值、中位数、标准差等

## 补充结果 - 异常值分析

**异常值分析**

离群值 =  $Q1(\text{下四分位}) - 1.5 * IQR(\text{四分位间距})$  或者  $Q3(\text{上四分位}) + 1.5 * IQR(\text{四分位间距})$   
 异常值 =  $Q1(\text{下四分位}) - 3.0 * IQR(\text{四分位间距})$  或者  $Q3(\text{上四分位}) + 3.0 * IQR(\text{四分位间距})$

组别	离群值	异常值
ERBB2	12.160142068914, ...	12.160142068914, ...
ERBB3	0.577924244506702	

各组离群值和异常值如上所示，如数据确认非人为记录错误，可不进行处理

这里统计各分子的离群值、异常值情况

- 离群值 =  $Q1(\text{下四分位}) - 1.5 * IQR(\text{四分位间距})$  或者  $Q3(\text{上四分位}) + 1.5 * IQR(\text{四分位间距})$
- 异常值 =  $Q1(\text{下四分位}) - 3.0 * IQR(\text{四分位间距})$  或者  $Q3(\text{上四分位}) + 3.0 * IQR(\text{四分位间距})$

## 补充结果 - 正态性检验

### 正态性检验(Shapiro-Wilk normality test)

组别	自由度(df)	统计量	p值
ERBB2	163	0.83968	4.4e-12
ERBB3	163	0.96672	0.0006

正态性检验结果显示, 存在有不满足正态分布的变量( $P < 0.05$ ), 建议选择用 非参数检验方法(Spearman)

这里提供各分子的正态性检验

- 分子接近正态分布( $P > 0.05$ ), 建议选择用参数检验方法(Pearson)

## 补充结果 - 相关性分析

### 相关性分析

同时提供Pearson和Spearman统计方法, 可以根据需要选择标注在图中的方法

方法	组别I	组别J	自由度(df)	统计量	相关系数	置信区间(95%CI)	p值
Pearson	ERBB2	ERBB3	163	7.7469	0.5211	0.39937 - 0.62477	9.94e-13
Spearman	ERBB2	ERBB3	163	2.405e+05	0.66685		0

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

这里提供相关性分析表: 可以查看第 1 列 (分子) 与第 2 列(第 2、3 列)之间的  
相关系数与其对应的统计学 p 值

- 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
- 相关是否有统计学意义还需要结合 p 值来查看



## 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) 对数据中主分子和补充分子之间进行相关性分析
- (2) 分析结果用 ggplot2 包进行棒棒糖图可视化



## 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