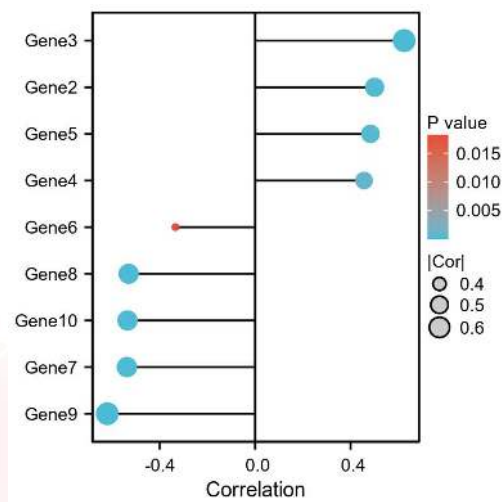


交互网络 - 相关性棒棒糖图



网址: <https://www.xiantao.love>



更新时间: 2023.02.16

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
数据处理	7
映射	8
点	10
线	11
标注	12
标题文本	13
图注	13
坐标轴	14
风格	15
图片	16
结果说明	17
主要结果	17
补充结果	18
方法学	19
如何引用	20
常见问题	21

基本概念

- 棒棒糖图：通过点图和线图组合成棒棒糖的形式
- 相关性棒棒糖图：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棒棒糖图形式展示结果

应用场景

通过点、线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 x 轴

	A	B	C	D	E	F	G	H	I	J
1	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
2	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557	0.518298418	0.739705603
3	-0.01360477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841	0.981862233	0.560852396
4	0.01182029	0.141322045	0.633965878	-0.263885834	0.542861073	0.490195153	1.183083571	0.987716302	1.067340227	1.033805987
5	0.004408363	-0.050815726	0.586873137	0.222179483	0.179198701	0.44559659	0.548951482	0.965296721	1.13184169	0.354723611
6	-0.210829963	0.369931219	0.547144016	0.36844285	0.472425386	0.854233325	1.18447251	1.123654746	1.262049733	0.943975343
7	0.642966409	0.355974738	0.671350067	0.173697559	0.032966	0.767416042	1.249694698	0.595296227	0.60783876	0.513691745
8	-0.179486354	0.196057306	0.168928365	0.065455873	-0.014343253	0.448853135	1.244653087	0.678902879	1.150490547	1.088849734
9	0.119288873	-0.192532717	-0.155474752	0.501615558	-0.246108244	0.940513758	1.272317943	0.347395517	0.745167982	0.645567114
10	0.15636136	0.52661018	-0.077797179	0.380145601	0.315129994	1.260823014	1.181884192	1.258837779	0.556531639	1.120612391
11	0.625642134	0.319952123	0.246807153	0.013511406	0.392332331	1.198206731	0.861498273	0.492483526	1.168313728	1.199737413
12	0.223237827	0.46813969	-0.270452353	-0.271382299	-0.042795272	0.734466762	0.696293427	0.3255452	1.235287427	0.623165723
13	0.327989421	0.053296945	-0.054571989	0.113712744	-0.114236178	0.36837606	1.103604986	1.17269313	0.833974877	0.594458569
14	-0.138863375	0.651505674	0.145345091	0.532905728	0.306025015	0.945230436	0.440612899	1.112037256	1.013758465	0.42780374
15	0.490465953	-0.14631829	0.063868247	0.010649082	0.333018047	0.33428674	1.225104532	1.255972546	1.076337558	0.507129556
16	0.391187731	0.140520319	-0.000645322	0.505075229	0.51432391	0.440272447	0.4341059	1.147909249	0.70803667	0.680611736

➤ 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	p值-Spearman
Gene1	Gene2	48	3.99705	0.499724	0.0002	1.039e+04	0.500984	0.0002
Gene1	Gene3	48	4.98716	0.584217	8.42e-06	7818	0.624586	2.24e-05
Gene1	Gene4	48	3.68768	0.469858	0.0006	1.131e+04	0.456711	0.0006
Gene1	Gene5	48	3.71781	0.472841	0.0005	1.076e+04	0.483409	0.0005
Gene1	Gene6	48	-2.65895	-0.358304	0.0106	2.778e+04	-0.333974	0.0006
Gene1	Gene7	48	-4.23575	-0.521615	0.0001	3.202e+04	-0.537479	7.42e-05
Gene1	Gene8	48	-4.39615	-0.535773	6.08e-05	3.186e+04	-0.529988	9.66e-05
Gene1	Gene9	48	-5.74211	-0.638124	6.2e-07	3.373e+04	-0.619592	2.8e-07
Gene1	Gene10	48	-4.34108	-0.530962	7.28e-05	3.196e+04	-0.534886	8.11e-05

相关性.xlsx

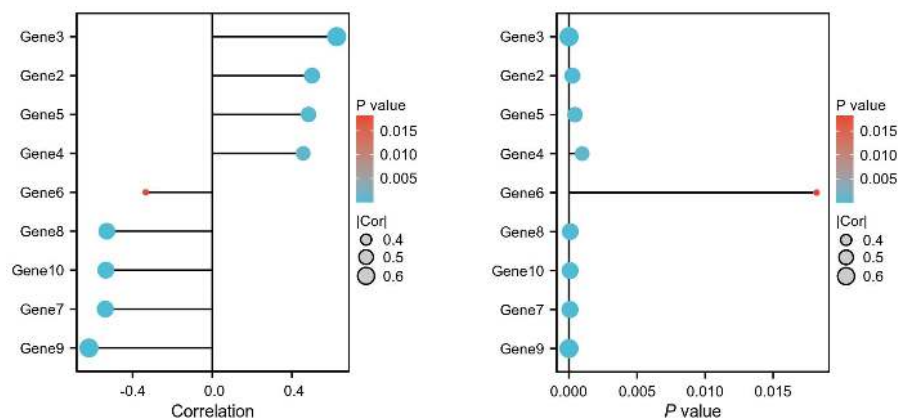
相关系数为正，说明两个变量之间存在正相关关系；相关系数为负，说明两个变量之间存在负相关关系；

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热图可视化

结果解读



- 横坐标表示其他变量（数据第 2 列开始每一列列名）（**图片为默认转置后**）
- 纵坐标表示主要变量（数据第 1 列）与其他变量（数据第二列开始）之间的（相关系数（左侧）/p 值（右侧）），取决于映射的内容（**图片为默认转置后**）
- 每个点表示主要变量与其他变量之间所映射的内容（相关系数/p 值/无）
- 连接点的横线表示相关性系数（相关系数/p 值），横线越长，表示其对应的值越大），取决于映射的内容
 - **如左图**：横线越长，变量之间相关系数的绝对值越大，表示主要变量与其他变量之间的相关性越强
 - **如右图**：横线越长，表示主要变量与其他变量之间的统计学 p 值越大，即相关性越没那么显著
- 横坐标等于 0 位置的竖线（虚线）表示相关性系数（相关系数/p 值）分割线

数据格式

	A	B	C	D	E	F	G	H	I	J
1	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
2	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557	0.518298418	0.739705603
3	-0.01368477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841	0.981862233	0.568852396
4	0.01182029	0.141322045	0.633965878	-0.263885834	0.542861073	0.490195153	1.183083571	0.987716302	1.067240227	1.033805987
5	-0.004488363	-0.058815726	0.586873137	0.222179483	0.179198701	0.44559659	0.548951482	0.965296724	1.13184169	0.354723611
6	-0.210829963	0.369931219	0.547144016	0.36844285	0.472425386	0.854233325	1.18447251	1.123654746	1.262049733	0.943975343
7	0.642966409	0.355974738	0.671350067	-0.173697559	-0.032966	0.767416042	1.249694698	0.595296227	0.60783876	0.513691745
8	-0.179486354	0.196057306	0.168928365	0.065455873	-0.014343253	0.448853135	1.244653087	0.678902879	1.150490547	1.088849734
9	0.119288873	-0.192532717	-0.155474752	0.501615558	-0.246108244	0.940513758	1.272317943	0.347395517	0.745167982	0.645567114
10	0.15636136	0.52661018	-0.077797179	0.380145601	0.315129994	1.260823014	1.181884192	1.258837779	0.556531639	1.120612391
11	0.625642134	0.319952123	0.246807153	0.013511406	0.392323231	1.198206731	0.861498273	0.492483526	1.168313728	1.199737413
12	0.223237827	0.46813969	-0.270452353	-0.271382299	-0.042795272	0.734466762	0.696293427	0.3255452	1.235287427	0.623165723
13	0.327989421	0.053296945	-0.054571989	0.113712744	-0.114236178	0.36837606	1.103604986	1.17269313	0.833974877	0.594458569
14	-0.138863375	0.651505674	0.145345091	0.532905728	0.306025015	0.945230436	0.440612899	1.112037256	1.013758465	0.42780374
15	0.049065953	-0.14631829	0.063868247	0.010649082	0.333018047	0.33428674	1.225104532	1.255972546	1.076337558	0.50712956
16	0.391187731	0.140520319	-0.000645322	0.505075229	0.51432391	0.440272447	0.4341059	1.147909249	0.70803667	0.680611736

数据要求：

- 数据至少 2 列以上，每列至少 5 个观测（即至少 5 行数据），数值类型，最多支持 30 列和 5000 行数据
 - 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
- 第一列为主要变量，用于计算自身与除自身外其他变量之间的相关性，图中所展示的变量（y 轴）是除第一列主要变量之外的其他变量，其顺序与上传数据中各分子的顺序保持一致，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。
- 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。)

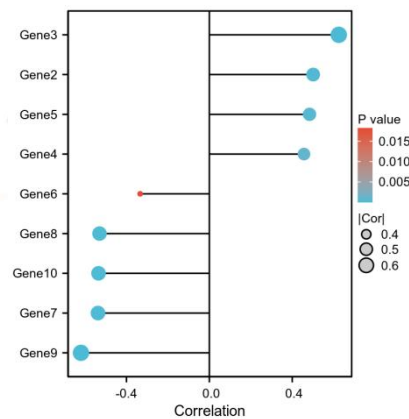
数据处理

统计 ▼

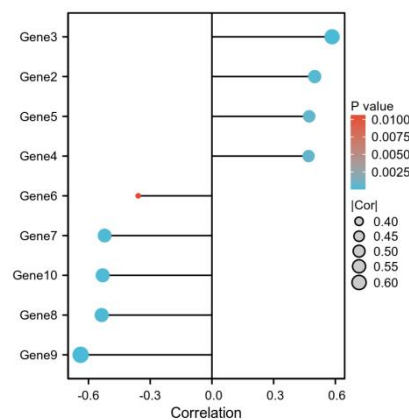
统计方法 Spearman ▼

➤ 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

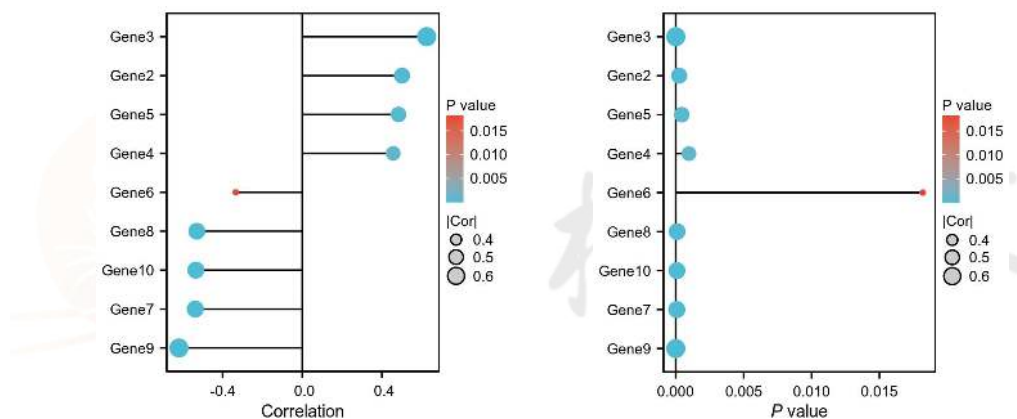
映射

y轴映射 相关系数

颜色映射 p值

大小映射 相关系数绝对值

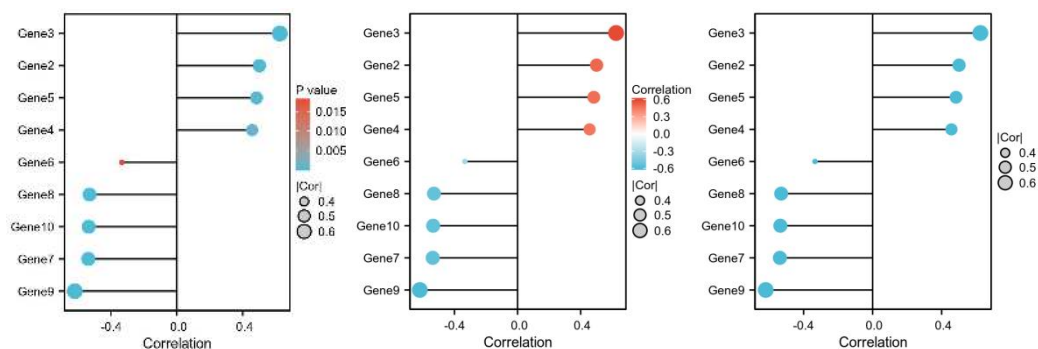
- y 轴映射：可以选择 y 轴映射的数据，即是 y 轴代表的信息，默认为相关系数，还可以选择 p 值，如下图：(图片为默认转置后的结果)



- 颜色映射：可以选择图中点映射的颜色信息，默认以 p 值作为映射，如图，颜色越偏向红色表示 p 值越大，颜色越偏向蓝色说明 p 值越小（如下左）

■ 还可以选择以相关系数进行映射（如下中）

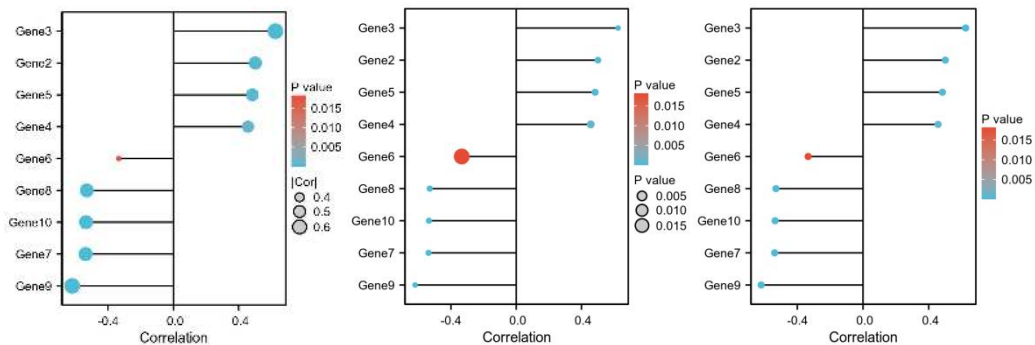
■ 还可以选择不映射（所有都是一个颜色）（如下右）



- 大小映射：可以选择图中点映射的大小信息，默认以相关系数绝对值作为映射，如图，点越大，说明主要变量与其他变量之间相关系数绝对值越大，即他们之间相关性越强（正相关/负相关）（如下左）

■ 还可以选择以 p 值进行映射（如下中）

■ 还可以选择不映射（所有点的大小都是一样的）（如下右）



点

点

填充色

描边色

样式

圆形

大小比例

1

不透明度

1

- 填充颜色：可以修改图中各点的填充颜色
- 描边颜色：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
- 大小比例：可以修改图中个点的大小比例，默认为 1
- 不透明度：可以修改图中各点不透明度，默认为 1，表示完全不透明,0 表示完全不透明

线



A configuration panel for lines, titled '线' (Line). It contains four settings: '颜色' (Color) with a black color swatch, '类型' (Type) with a dropdown menu showing '实线' (Solid line), '粗细' (Thickness) with a dropdown menu showing '0.75pt', and '不透明度' (Opacity) with a text input field showing '1'.

- 颜色：可以修改图中线的颜色
- 类型：可以选择连接点的线的类型，默认为实线，还可以选择虚线
- 粗细：可以选择修改线条的粗细
- 不透明度：可以修改线条的不透明度，默认为 1，表示完全不透明，0 表示完全透明

标注

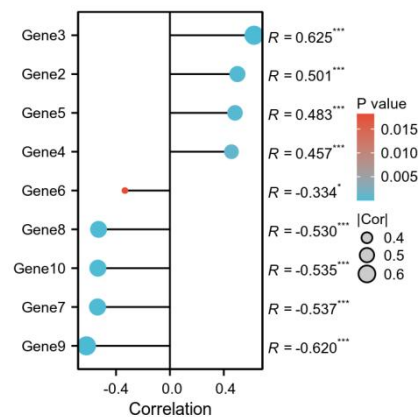
标注

内容
不标注

标注大小
6pt

- 内容：可以选择是否对图形进行标注映射，默认为不映射，还可以选择映射：
如下（为相关系数-星号）

- 相关系数
- 相关系数-星号
- 星号、p 值科学计数法
- p 值数值(小于 0.05 自动<)
- p 值数值(小于 0.001 自动<)



- 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

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

图注

是否展示

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
- 图注位置：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

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
- y 轴范围+刻度：可以控制 y 轴范围和刻度，可只提供 2 个值来控制范围。形如 0.1, 0.1, 0.2, 0.3 (最小值和最大值能不能可视化数据范围 20%，如果调整过大可能会无作用)



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
- 可以选择是否进行 x、y 轴颠倒，默认进行颠倒
- 文字大小：控制整体文字大小，默认为 7pt

图片

图片

▼

宽度 (cm)

6

高度 (cm)

7

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



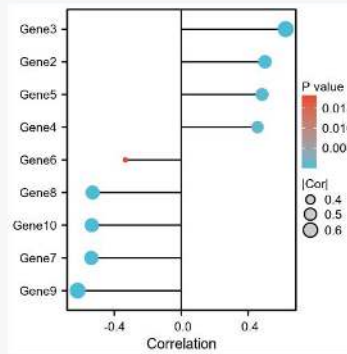
结果说明

主要结果

相关性棒棒糖图

相关性棒棒糖图: 分析1个主变量和其他所有变量之间的相关性结果并用棒棒糖图形式展示结果

统计方法: spearman



[相关性棒棒糖图.pdf](#)

[相关性棒棒糖图.tiff](#)

补充结果

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	p值-Spearman
Gene1	Gene2	48	3.99705	0.499724	0.0002	1.039e+04	0.500984	
Gene1	Gene3	48	4.98716	0.584217	8.42e-06	7818	0.624586	2
Gene1	Gene4	48	3.68768	0.469858	0.0006	1.131e+04	0.456711	
Gene1	Gene5	48	3.71781	0.472841	0.0005	1.076e+04	0.483409	
Gene1	Gene6	48	-2.65895	-0.358304	0.0106	2.778e+04	-0.333974	
Gene1	Gene7	48	-4.23575	-0.521615	0.0001	3.202e+04	-0.537479	7
Gene1	Gene8	48	-4.39615	-0.535773	6.08e-05	3.186e+04	-0.529988	9
Gene1	Gene9	48	-5.74211	-0.638124	6.2e-07	3.373e+04	-0.619592	4
Gene1	Gene10	48	-4.34108	-0.530962	7.28e-05	3.196e+04	-0.534886	8

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；

这里提供相关性分析表：可以查看主要变量与其他变量之间的相关系数与其对应的统计学 p 值

- 相关系数为正数，说明两个变量（主要变量与其他变量）之间可能存在正相关关系；相关系数为负数，说明两个变量可能存在负相关关系
 - 相关系数绝对值在 0.8-1.0 之间，说明两个变量之间强相关
 - 相关系数绝对值在 0.5-0.8 之间，说明两个变量之间中等程度相关
 - 相关系数绝对值在 0.3-0.5 之间，说明两个变量之间相关程度一般
 - 相关系数绝对值在 0.0-0.3 之间，说明两个变量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) 对数据中主变量和次要变量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)。

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

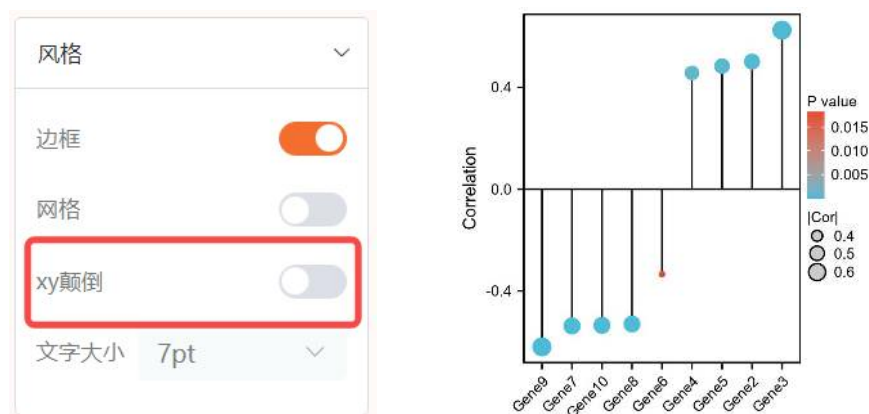
1. 为什么数据说明与主要结果说明跟图的坐标对应不上?



➤ 横坐标表示其他变量（数据第二列开始每一列列名）

纵坐标表示主要变量（数据第一列）与其他变量（数据第二列开始）之间的（相关系数（左侧）/p 值（右侧）），取决于映射的内容（**图片为默认转置**

答：主要结果显示的是默认将图形进行转置之后得到的，所以会跟数据说明与主要结果说明反过来了，如下（左侧为参数设置，右侧为不转置结果）



2. 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。

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