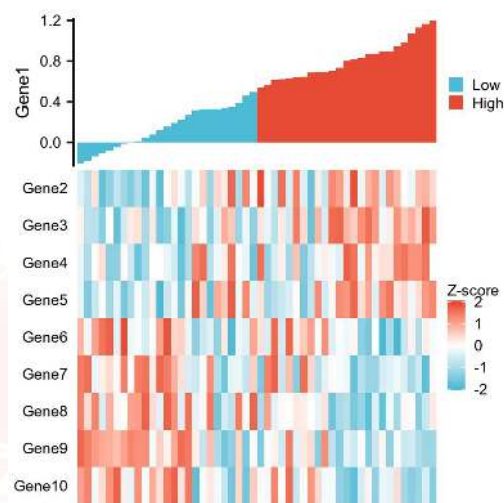


基础绘图 - 共表达热图



网址: <https://www.xiantao.love>



更新时间: 2023.02.17

目录	
基本概念	3
应用场景	3
分析过程	3
结果解读	6
数据格式	7
参数说明	8
数据处理	8
统计	9
上部分	10
热图	12
标注	13
标题文本	14
图注	15
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

- 热图：热图是一个以颜色变化来显示数据情况的矩阵
- 共表达热图：上部分用柱/点/线的形式来展示主要变量的值，下部分用热图的形式来展示其他变量的值，观察其他变量随主要变量的变化趋势

应用场景

- (1) 通过柱状图/点图/线图的形式来展示主要变量各样本的表达值
- (2) 通过热图的形式来展示其他变量经过 zscore 转换后的表达值相对变化趋势；
热图常用于展示多个基因在不同样本中的表达水平

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
1	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9
2	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557	0.518298418
3	-0.01368477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841	0.981862233
4	0.01182029	0.141322045	0.633965878	-0.263885834	0.542861073	0.490195153	1.183083571	0.987716302	1.067240227
5	-0.004488363	-0.058815726	0.586873137	0.222179483	0.179198701	0.44559659	0.548951482	0.965296724	1.13184169
6	-0.210829963	0.369931219	0.547144016	0.36844285	0.472425386	0.854233325	1.18447251	1.123654746	1.262049733
7	0.642966409	0.355974738	0.671350067	-0.173697559	-0.032966	0.767416042	1.249694698	0.595296227	0.60783876
8	-0.179486354	0.196057306	0.168928365	0.065455873	-0.014343253	0.448853135	1.244653087	0.678902879	1.150490547
9	0.119288873	-0.192532717	-0.155474752	0.501615558	0.246108244	0.940513758	1.272317943	0.347395517	0.745167982
10	0.15636136	0.52661018	-0.077797179	0.380145601	0.315129994	1.260823014	1.181884192	1.258837779	0.556531639
11	0.625642134	0.319952123	0.246807153	0.013511406	0.392323231	1.198206731	0.861498273	0.492483526	1.168313728
12	0.223237827	0.46813969	-0.270452353	-0.271382299	-0.042795272	0.734466762	0.696293477	0.3255452	1.235287427
13	0.327989421	0.053296945	-0.054571989	0.113712744	-0.114236178	0.36837606	1.103604986	1.17269313	0.833974877
14	-0.138863375	0.651505674	0.145345091	0.532905728	0.306025015	0.945230436	0.440612899	1.112037256	1.013758465
15	0.049065953	-0.14631829	0.063868247	0.010649082	0.333018047	0.33428674	1.225104532	1.255972546	1.076337558
16	0.391118731	0.140520319	-0.000645322	0.505075229	0.51432391	0.440272447	0.4341059	1.147909249	0.70803667

➤ 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)，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	0.000254573
Gene1	Gene3	48	4.98716	0.584217	8.42e-06	7818	0.624586	2.24115E-06
Gene1	Gene4	48	3.68768	0.469858	0.0006	1.131e+04	0.456711	0.000966584
Gene1	Gene5	48	3.71781	0.472841	0.0005	1.076e+04	0.483409	0.000441103
Gene1	Gene6	48	-2.65895	-0.358304	0.0106	2.778e+04	-0.333974	0.018195201
Gene1	Gene7	48	-4.23575	-0.521615	0.0001	3.202e+04	-0.537479	7.41699E-05
Gene1	Gene8	48	-4.39615	-0.535773	6.08e-05	3.186e+04	-0.529988	9.6553E-05
Gene1	Gene9	48	-5.74211	-0.638124	6.2e-07	3.373e+04	-0.619592	2.79649E-06
Gene1	Gene10	48	-4.34108	-0.530962	7.28e-05	3.196e+04	-0.534886	8.13118E-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；

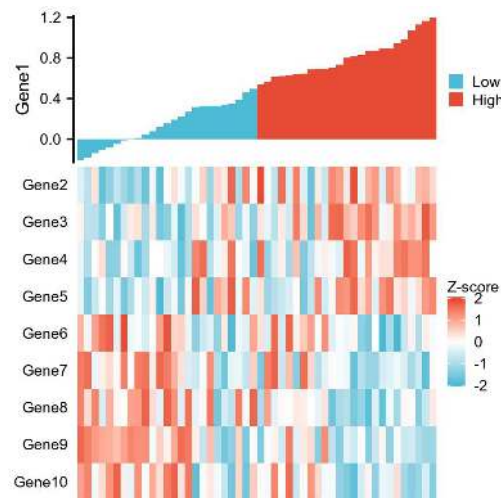
相关系数绝对值代表相关程度，0-0.3代表弱或者不相关；0.3-0.5代表弱相关；0.5-0.8代表中等程度相关；0.8-1代表强相关

相关是否有统计学意义还需要结合p值来查看

	A	B	C	D	E	F	G	H	I
1	var1	var2	df	statistic_pearson	cor_pearson	pvalue_pearson	statistic_spearman	cor_spearman	pvalue_spearman
2	Gene1	Gene2	48	3.997052511	0.499723597	0.000220055	10392	0.500984394	0.000254573
3	Gene1	Gene3	48	4.987164944	0.584216665	8.4218E-06	7818	0.624585834	2.24115E-06
4	Gene1	Gene4	48	3.687683713	0.469858183	0.000576033	11314	0.456710684	0.000966584
5	Gene1	Gene5	48	3.717805282	0.472840695	0.000525295	10758	0.483409364	0.000441103
6	Gene1	Gene6	48	-2.658946604	-0.358304357	0.010619887	27780	-0.333973589	0.018195201
7	Gene1	Gene7	48	-4.235745731	-0.521615313	0.000102532	32018	-0.537478992	7.41699E-05
8	Gene1	Gene8	48	-4.396150101	-0.535772808	6.08131E-05	31862	-0.529987995	9.6553E-05
9	Gene1	Gene9	48	-5.74211491	-0.638123529	6.1952E-07	33728	-0.619591837	2.79649E-06
10	Gene1	Gene10	48	-4.34108391	-0.5309623	7.28142E-05	31964	-0.534885954	8.13118E-05

- 将分析后得到的结果（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



对于上图（柱状图）：

- 横坐标表示各个样本
- 纵坐标表示主要变量在各样本之间的表达值
- 图形是**按照主要变量在各个样本间的表达值从低到高进行排序**，再以**表达值的中位数**为分界线划分为高低两个表达组

对于下图（热图）：

- 列表示各个样本
- 行表示除了主要变量之外的其他变量
- 每一个方块表示其他变量进行 z-score 转换之后在各样本间的表达值,颜色的深浅表示值的绝对值大小

补充：

- zscore 转换是绘制热图中常用的一种对数据进行转换的方法（每个变量在单个样本中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）
- zscore 转换可减少不同变量表达值差异过大而影响整个热图的可视化效果并且保留了单个变量在样本间的差异情况

数据格式

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

数据要求：

- 数据至少 2 列以上，每列至少 5 个观测（至少 5 行数据），最多支持 35 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
- 每 1 列代表变量，每 1 行代表样本。其中第 1 列为主要变量，其他列为其他变量
- 除第 1 列外，每一列列名即为热图的行名。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。
- 变量名（主要变量、其他变量名不能重复），即列名不能重复

参数说明

(说明：标注了颜色的为常用参数。)

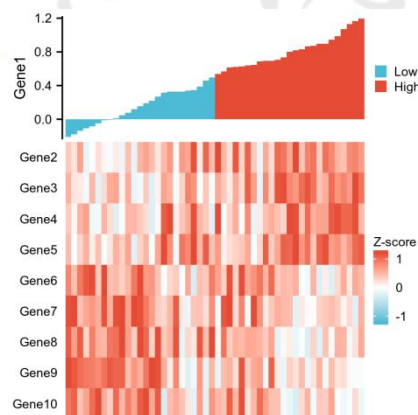
数据处理



➤ **数据处理**：可以选择是否对其他变量进行归一化处理（为了更好展示热图中不同变量之间的差异，一般都需要对热图中的变量进行归一化处理）

■ 默认选择 归一化(处理极值：归一化后大于3的值固定为3,小于-3固定为-3)

■ 还可以选择 归一化 或 不归一化，如下：为不归一化



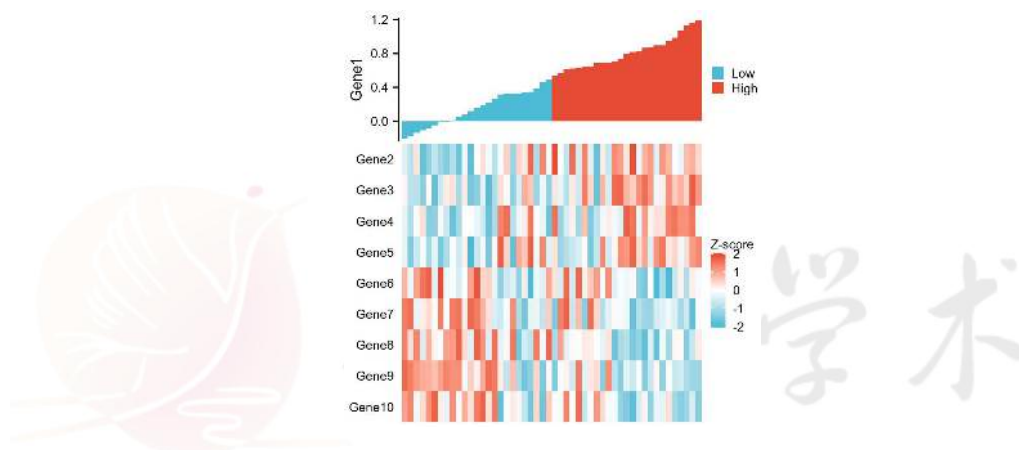
统计

统计

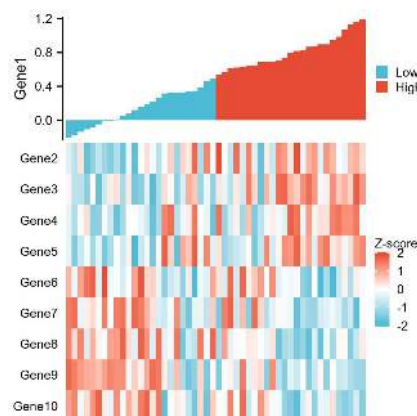
统计方法 Spearman

➤ 统计方法：可以选择对主要变量和其他变量进行相关性分析处理时的方法

- 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



上部分

上部分

样式

柱状图

分组

中位数分组

颜色

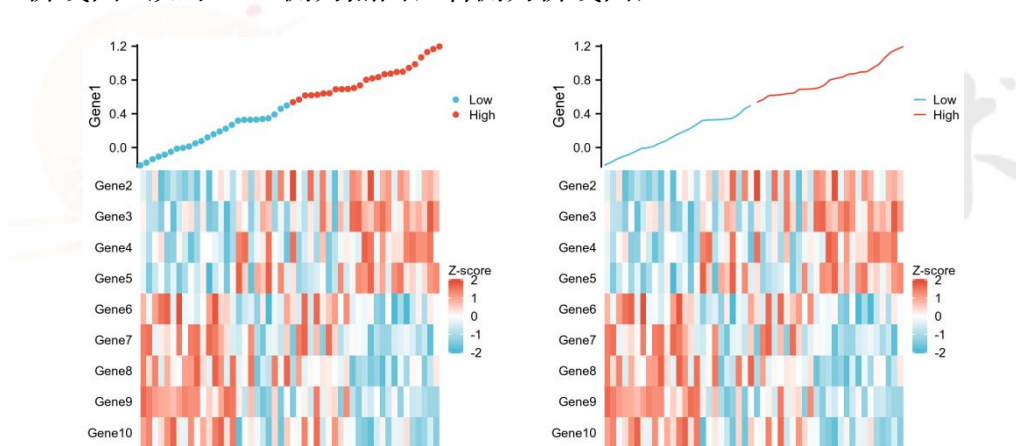
大小

0.75pt

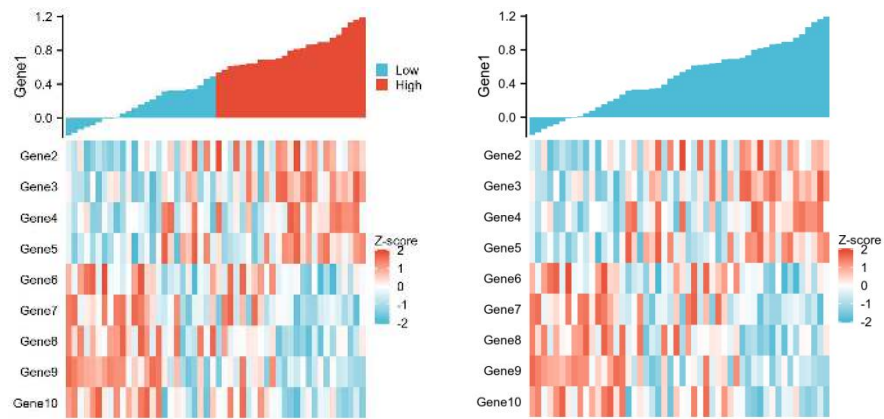
不透明度

1

- 样式：可以选择上部分图形的样式，默认为柱状图，还可以选择 点图 或者折线图（如下：左侧为点图，右侧为折线图）



- 分组：可以选择是否对主要变量进行分组，默认为以 中位数分组，分为高低两个表达组，还可以选择不分组，如下：（左侧为中位数分组（默认），右侧为不分组）



- 颜色：可以修改主要变量样本对应柱子/点/折线的颜色
- 大小：可以修改主要变量对应柱子的宽度/点的大小/折线的粗细
- 不透明度：可以修改主要变量样本对应高低表达组柱子/点/折线的不透明度，1 表示完全不透明，0 表示完全透明

热图



- 填充色：可以修改对应热图的颜色
- 描边色：可以修改对应热图的描边色
- 描边粗细：可以修改对应热图每个小矩形的描边粗细
- 不透明度：可以修改对应热图的不透明度，1 表示完全不透明，0 表示完全透明

标注

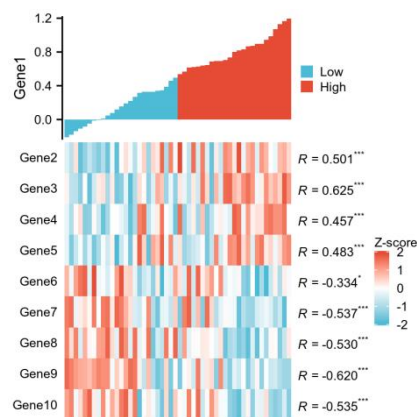
标注 ▼

内容 不标注 ▼

标注大小 6pt ▼

➤ 内容：可以选择是否对热图右侧进行标注操作，默认为不进行标注，还可以选择进行标注，如下：（相关系数-星号标注）

- 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,



➤ 标注大小：可以选择并修改标注的大小，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



- 是否展示：可以选择是否展示图注信息
- 图注位置：可以展示图注的位置，默认表示默认为右侧，还可以选择上侧
- 图注标题：可以修改图注标题内容

风格



- 边框：可以选择是否展示边框，默认不展示
- 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

图片

▼

宽度 (cm)

7

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

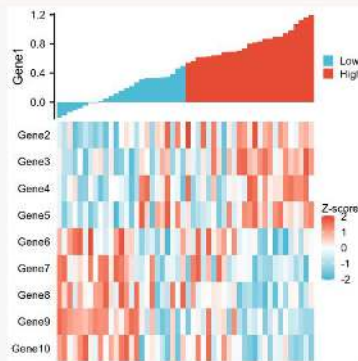


结果说明

主要结果

共表达热图

共表达热图: 上部分用柱/点/线的形式来展示主要变量的值, 下部分用热图的形式来展示其他变量的值, 观察其他变量随主要变量的变化趋势
统计方法: spearman



共表达热图.pdf

共表达热图.tiff

补充: zscore转换

- zscore转换是绘制热图中常用的一种对数据进行转换的方法 (每个变量在单个样本中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)
- zscore转换可减少不同变量表达值差异过大而影响整个热图的可视化效果并且保留了单个变量在样本间的差异情况

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	0.0002
Gene1	Gene3	48	4.98716	0.584217	8.42e-06	7818	0.624586	2.2e-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	0.0106
Gene1	Gene7	48	-4.23575	-0.521615	0.0001	3.202e+04	-0.537479	7.4e-05
Gene1	Gene8	48	-4.39615	-0.535773	6.08e-05	3.186e+04	-0.529988	9.6e-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	8.1e-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

这里提供（pearson，spearman 两种方法）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检验 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) 对数据中主变量和其它变量之间进行相关性分析
- (2) 分析结果用 ggplot 包进行共表达热图可视化



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. 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 减少数据中的列数（变量数量）。

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