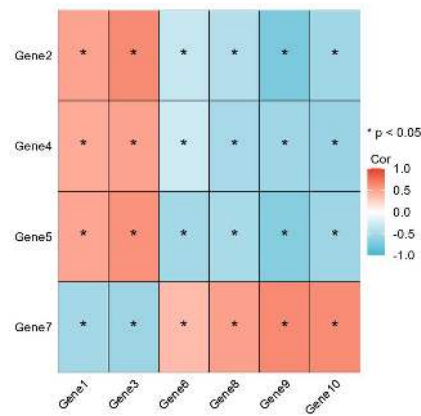


交互网络 - 相关性热图[指定 xy]



网址: <https://www.xiantao love>



更新时间: 2023.01.31

目录

基本概念	3
应用场景	3
分析流程	3
结果解读	7
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
标题文本	15
图注	15
坐标轴	16
风格	16
图片	17
结果说明	18
主要结果	18
补充结果	19
方法学	20
如何引用	21
常见问题	22

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[指定 xy]：将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

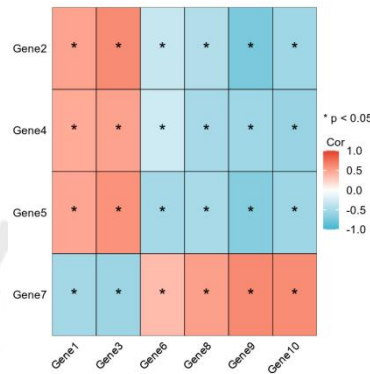
- 数据格式：csv / txt 文件格式：

- 第 1 行为 #注释的分组信息

- ◆ 提供：在数据处理过程中会根据注释信息将数据分成不同的组(不同的变量/列对应不同的分组)，再将两个分组进行相关性分析。#注释的信息中提供的分组个数不能超过 2 个，注释信息的长度与数据（除了第 1 列外）的列数一致，如下：

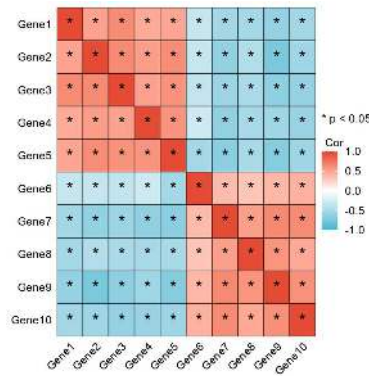
	A	B	C	D	E	F	G	H	I	J	K
1	#group	group2	group1	group2	group1	group1	group2	group1	group2	group2	group2
2	sample_id	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
3	sample1	0.4967459	0.4677233	-0.048352	-0.240196	0.4322632	0.9218794	0.7845149	1.2456156	0.5182984	0.7397056
4	sample2	-0.013685	0.0642193	0.2497329	0.5373747	-0.04317	1.286958	1.1756629	0.5182841	0.9818622	0.5688524
5	sample3	0.0118203	0.141322	0.6339659	-0.263886	0.5428611	0.4901952	1.1830836	0.9877163	1.0672402	1.033806
6	sample4	-0.004488	-0.058816	0.5868731	0.2221795	0.1791987	0.4455966	0.5489515	0.9652967	1.1318417	0.3547236
7	sample5	-0.21083	0.3699312	0.547144	0.3684429	0.4724254	0.8542333	1.1844725	1.1236547	1.2620497	0.9439753

- ① #group 表示分组的名字
- ② group2..表示分组的信息(最多两个分组: group1, group2),
group1 对应的变量为 Gene2、Gene4、Gene5、Gene7; group2 对应的变量为 Gene1、Gene3、Gene6、Gene8、Gene9、Gene10



- 此时的分组(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
- ◆ 不提供: 整个数据将会被作为一个大的分组进行两两间相关性分析及可视化

	A	B	C	D	E	F	G	H	I	J	K
1	sample_id	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
2	sample1	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557	0.518298418	0.739705603
3	sample2	-0.01368477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841	0.981862233	0.568852396
4	sample3	0.01182029	0.141322045	0.633965878	-0.263885834	0.542861073	0.490195153	1.183083571	0.987716302	1.067240227	1.033805987
5	sample4	-0.004488363	-0.058815726	0.586873137	0.222179483	0.179198701	0.44559659	0.548951482	0.965296724	1.13184169	0.354723611
6	sample5	-0.210829963	0.369931219	0.547144016	0.36844285	0.472425396	0.854233325	1.18447251	1.123654746	1.262049733	0.943975343
7	sample6	0.642966409	0.355974738	0.671350067	-0.173697559	-0.032966	0.767416042	1.249694698	0.595296227	0.60783876	0.513691745



- 此时 xy 两个方向上都是一样的，没有分组，也就是样本间的两两相关性分析

■ 第 1 列为分类类型，表示样本(样本标号/编号)

■ 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

■ 将个分组中的所有变量进行两两间相关性分析，结果如下：

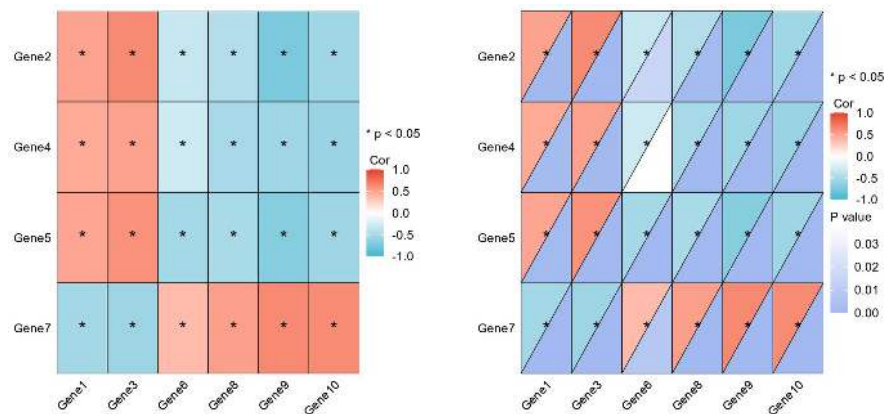
	A	B	C	D	E	F	G
1		Gene2	Gene4	Gene5	Gene7		
2	Gene1	0.50098439	0.45671068	0.48340936	-0.537479		
3	Gene3	0.62957983	0.50559424	0.59164466	-0.5768547		
4	Gene6	-0.3418487	-0.2952701	-0.5280672	0.37123649		
5	Gene8	-0.4590156	-0.5207683	-0.5155822	0.51711885		
6	Gene9	-0.7502041	-0.5568788	-0.6896999	0.63985594		
7	Gene10	-0.5556303	-0.592413	-0.5628331	0.62516206		
8							
9							
10							
11							
12							
13							
14							
15							

相关系数表格
相关性检验表格
+

➤ 将分析所得结果进行可视化



结果解读



左图：（完整热图）

- 行、列都代表变量/样本（分别代表不同方向/不同组别分别对应的变量）
- 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

右图：（完整热图-三角对角线）

- 行、列都代表变量/样本
- 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

补充：

- * 表示 $P\text{value} < 0.05$
- 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	group2	group1	group2	group1	group1	group2	group1	group2	group2	group2
2	sample_id	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
3	sample1	0.4967459	0.4677233	-0.048352	-0.240196	0.4322632	0.9218794	0.7845149	1.2456156	0.5182984	0.7397056
4	sample2	-0.013685	0.0642193	0.2497329	0.5373747	-0.04317	1.286958	1.1756629	0.5182841	0.9818622	0.5688524
5	sample3	0.0118203	0.141322	0.6339659	-0.263886	0.5428611	0.4901952	1.1830836	0.9877163	1.0672402	1.033806
6	sample4	-0.004488	-0.058816	0.5868731	0.2221795	0.1791987	0.4455966	0.5489515	0.9652967	1.1318417	0.3547236
7	sample5	-0.21083	0.3699312	0.547144	0.3684429	0.4724254	0.8542333	1.1844725	1.1236547	1.2620497	0.9439753
8	sample6	0.6429664	0.3559747	0.6713501	-0.173698	-0.032966	0.767416	1.2496947	0.5952962	0.6078388	0.5136917
9	sample7	-0.179486	0.1960573	0.1689284	0.0654559	-0.014343	0.4488531	1.2446531	0.6789029	1.1504905	1.0888497
10	sample8	0.1192889	-0.192533	-0.155475	0.5016156	-0.246108	0.9405138	1.2723179	0.3473955	0.745168	0.6455671
11	sample9	0.1563614	0.5266102	-0.077797	0.3801456	0.31513	1.260823	1.1818842	1.2588378	0.5565316	1.1206124
12	sample10	0.6256421	0.3199521	0.2468072	0.0135114	0.3923232	1.1982067	0.8614983	0.4924835	1.1683137	1.1997374
13	sample11	0.2232378	0.4681397	-0.270452	-0.271382	-0.042795	0.7344668	0.6962934	0.3255452	1.2352874	0.6231657
14	sample12	0.3279894	0.0532969	-0.054572	0.1137127	-0.114236	0.3683761	1.103605	1.1726931	0.8339749	0.5944586
15	sample13	-0.138863	0.6515057	0.1453451	0.5329057	0.306025	0.9452304	0.4406129	1.1120373	1.0137585	0.4278037
16	sample14	0.049066	-0.146318	0.0638682	0.0106491	0.333018	0.3342867	1.2251045	1.2559725	1.0763376	0.5071296

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 #注释的数据作为分组信息，(将变量分为不同的组，以便进行不同组中变量间相关性分析)
 - 除第 1 行外，第 1 列为分类类型数据，表示样本信息/样本名
 - 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
- 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

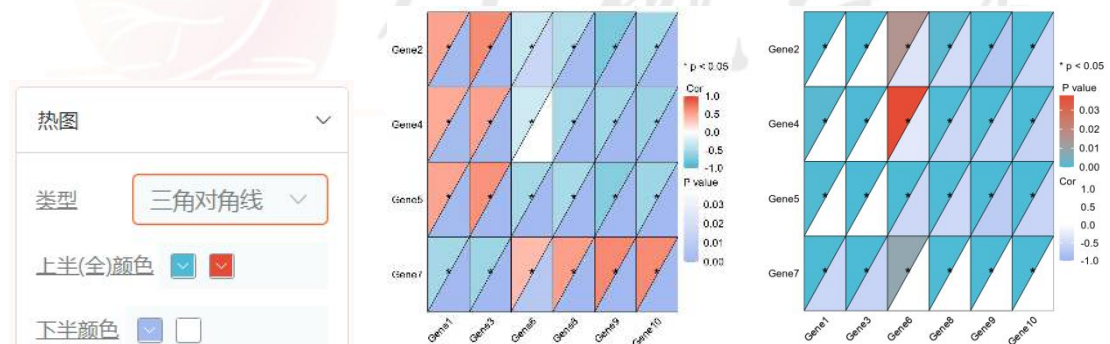
下半颜色映射

p值

大小映射

相关性系数绝对值

- 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：（左为热图类型-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



- 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，还可以选择相关性系数绝对值，如下：

映射

上半颜色

映射

相关系数

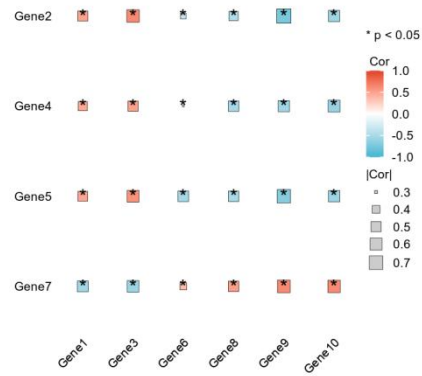
下半颜色

映射

p值

大小映射

相关系数绝对值



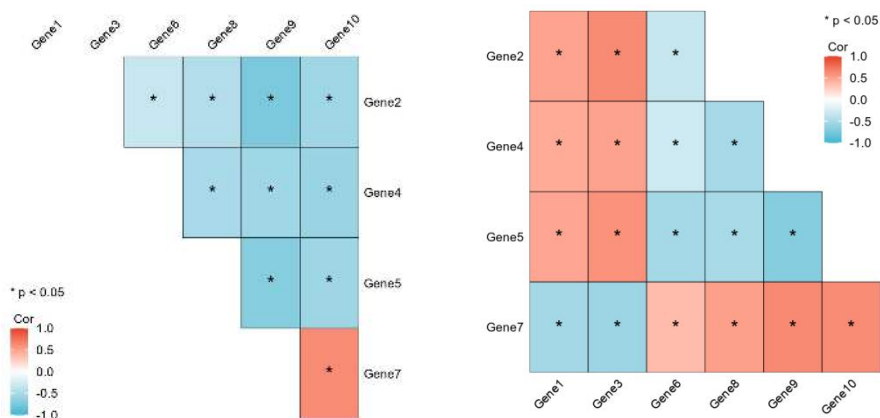
样式

布局

类型

完整热图

- 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

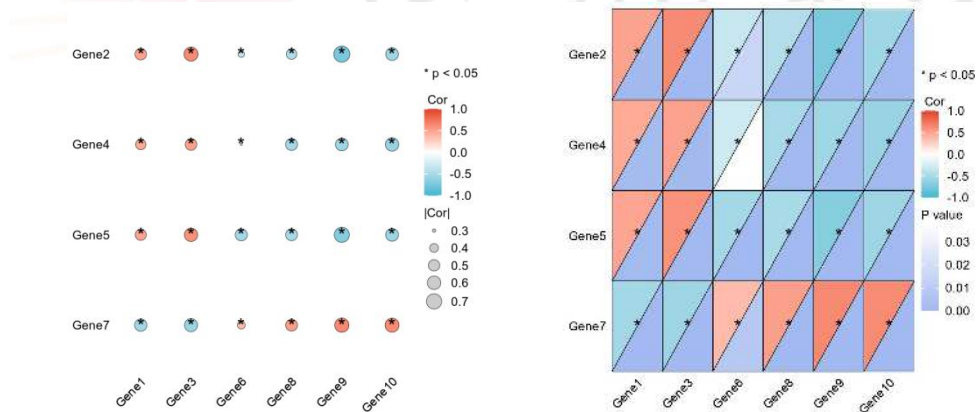
大小比例

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

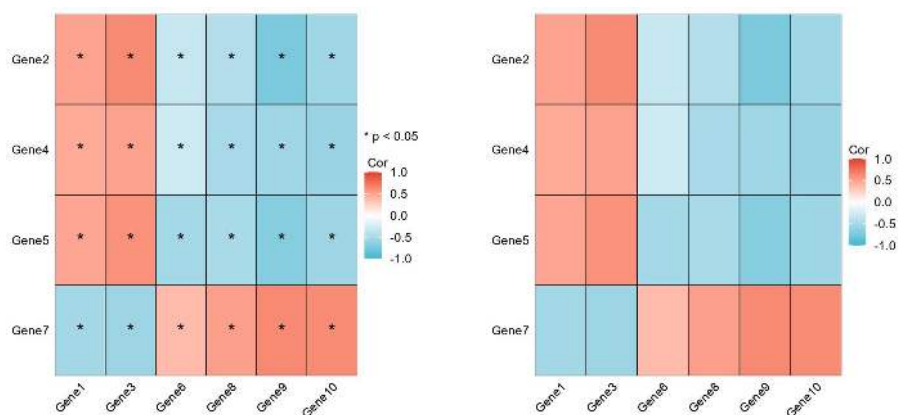
标注映射 星号(*p<0.05)

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

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

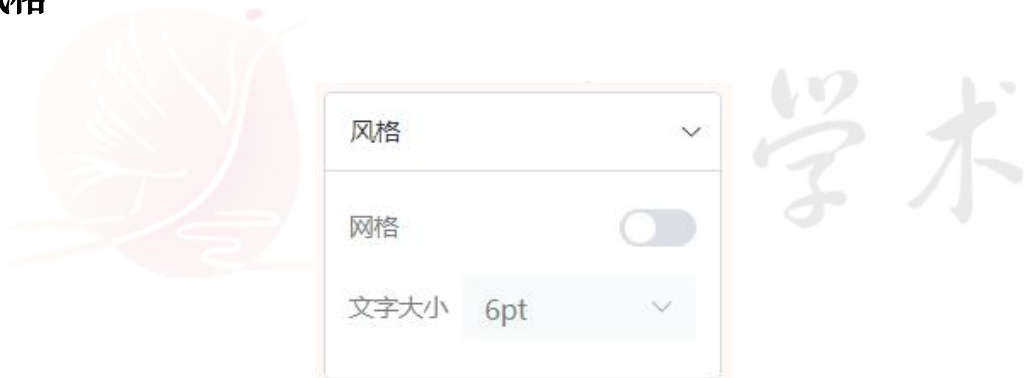
- 是否展示：可以选择是否展示图注信息，默认展示
- 图注标题：可以修改图注标题内容，默认没有
- 图注位置：默认为图片的右侧，还可以选择上、下

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)	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



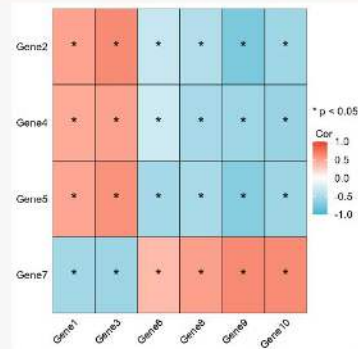
结果说明

主要结果

相关性热图-XY指定

相关性热图-XY指定: 用热图展示相关性情况

统计方法: spearman



相关性热图 XY指定.pdf

相关性热图 XY指定.tif



补充结果

相关性分析

统计方法: spearman

表1: 相关系数表格

	Gene2	Gene4	Gene5	Gene7
Gene1	0.501	0.457	0.483	-0.537
Gene3	0.63	0.506	0.592	-0.577
Gene6	-0.342	-0.295	-0.528	0.371
Gene8	-0.459	-0.521	-0.516	0.517
Gene9	-0.75	-0.557	-0.69	0.64
Gene10	-0.556	-0.592	-0.563	0.625

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

表2: 相关性检验表格

	Gene2	Gene4	Gene5	Gene7
Gene1	0.000255	0.000967	0.000441	7.42e-05
Gene3	1.79e-06	0.000219	9.24e-06	1.68e-05
Gene6	0.0155	0.0378	0.000103	0.00829
Gene8	0.000905	0.000133	0.000158	0.00015
Gene9	0	3.65e-05	1.1e-07	1.13e-06
Gene10	3.82e-05	8.95e-06	2.91e-05	2.18e-06

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)。

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