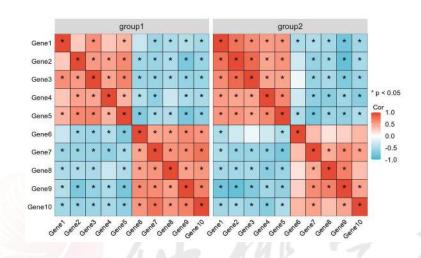


# 交互网络 - 相关性热图[分组统计]



网址: https://www.xiantao.love



更新时间: 2023.01.30



#### 目录

基本概念	3
应用场景	3
分析流程	3
结果解读	6
数据格式	8
参数说明	9
统计	9
映射	0
样式	2
热图 1	3
标注 1	5
分面	6
标 <mark>题文本</mark>	7
图注	
<del>坐标轴</del>	8
风格1	8
图片 1	9
结果说明 20	0
主要结果20	0
补充结果2	1
方法学 2	3
如何引用 2-	4
堂口问题 2	5



#### 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[分组统计]:通过将数据进行分组的方式,对各个分组中的变量进行两两间相关性分析,再以热图的形式展示其结果

## 分析流程

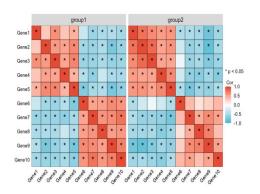
上传数据 → 数据清洗 → 数据处理 → 将各分组进行相关性分析得到的结果进行可视化

- ▶ 数据格式:
  - 第 1 列是<mark>分类类型</mark>,表示分组信息

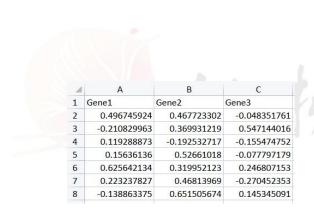


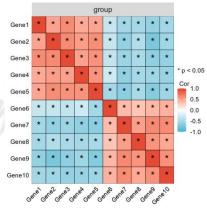
◆ 提供: 该列分类的个数不能少于 2 个,不能大于 4 个,每一个分类 对应的数量至少 5 个

1	Α	В	С	D	
1	group	Gene1	Gene2	Gene3	
2	group1	0.49674592	0.4677233	-0.0483518	
3	group2	-0.0136848	0.06421928	0.24973294	
4	group2	0.01182029	0.14132204	0.63396588	
5	group2	-0.0044884	-0.0588157	0.58687314	
6	group1	-0.21083	0.36993122	0.54714402	
7	group2	0.64296641	0.35597474	0.67135007	
8	group2	-0.1794864	0.19605731	0.16892836	



◆ <mark>不提供</mark>: 这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





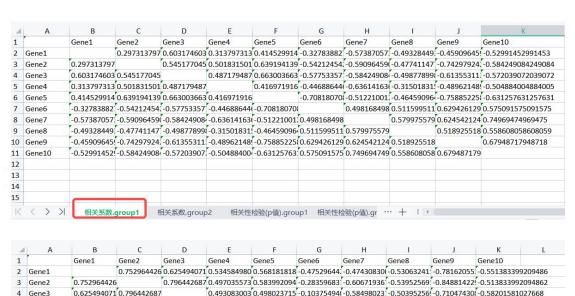
- 第2列及以后为数值类型,表示每个变量/样本值
- ▶ 数据清洗:对非字符和除第1列外非数值的数据进行清洗
- ▶ 数据处理:
  - ① 根据第 1 列分组信息对除第 1 列外的所有变量进行分组(即将数据分成几份 / 或者说一个变量的数据被分成了多少份),结果如下:



4	Α	В	С	D	E	F	G	Н	1
1	group	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8
2	group1	0.496745924	0.467723302	-0.048351761	-0.240196499	0.43226315	0.921879419	0.784514945	1.245615557
3	group1	-0.210829963	0.369931219	0.547144016	0.36844285	0.472425386	0.854233325	1.18447251	1.123654746
4	group1	0.119288873	-0.192532717	-0.155474752	0.501615558	-0.246108244	0.940513758	1.272317943	0.347395517
5	group1	0.15636136	0.52661018	-0.077797179	0.380145601	0.315129994	1.260823014	1.181884192	1.258837779
6	group1	0.625642134	0.319952123	0.246807153	0.013511406	0.392323231	1.198206731	0.861498273	0.492483526
7	group1	0.223237827	0.46813969	-0.270452353	-0.271382299	-0.042795272	0.734466762	0.696293427	0.3255452
8	group1	-0.138863375	0.651505674	0.145345091	0.532905728	0.306025015	0.945230436	0.440612899	1.112037256
9	group1	0.391187731	0.140520319	-0.000645322	0.505075229	0.51432391	0.440272447	0.4341059	1.147909249
10	group1	0.189545701	0.652805553	0.3729213	0.240424702	0.525400967	0.703610661	0.928725625	1.044582651

4	Α	В	C	D	E	F	G	Н	1
1	group	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8
2	group2	-0.01368477	0.06421928	0.249732944	0.537374657	-0.04316992	1.286957965	1.175662851	0.5182841
3	group2	0.01182029	0.141322045	0.633965878	-0.263885834	0.542861073	0.490195153	1.183083571	0.987716302
4	group2	-0.004488363	-0.058815726	0.586873137	0.222179483	0.179198701	0.44559659	0.548951482	0.965296724
5	group2	0.642966409	0.355974738	0.671350067	-0.173697559	-0.032966	0.767416042	1.249694698	0.595296227
6	group2	-0.179486354	0.196057306	0.168928365	0.065455873	-0.014343253	0.448853135	1.244653087	0.678902879
7	group2	0.327989421	0.053296945	-0.054571989	0.113712744	-0.114236178	0.36837606	1.103604986	1.17269313
8	group2	0.049065953	-0.14631829	0.063868247	0.010649082	0.333018047	0.33428674	1.225104532	1.255972546
9	group2	0.567831358	0.497207698	0.374835286	-0.068825581	0.103242341	0.334437394	1.163623489	0.971857322
10	group2	0.616972199	0.215477305	0.138560234	0.283830582	0.2192212	1.182833055	1.299929019	0.36170394

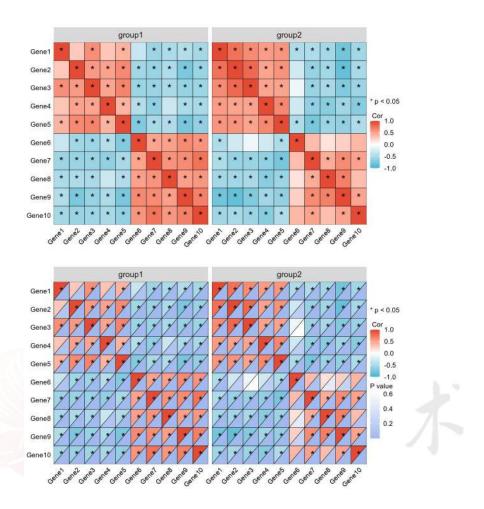
#### ■ ② 将个分组中的所有变量进行两两间相关性分析,结果如下:







#### 结果解读



第1个图: (完整热图)

- 通过分组的方式对所有变量进行相关性分析
- ▶ 每个分组热图的行、列都代表变量/样本
- ▶ 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

第2个图: (完整热图-三角对角线)

- 通过分组的方式对所有变量进行相关性分析
- ▶ 每个分组热图的行、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 Pvalue < 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



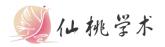


# 数据格式

4	Α	В	С	D	E	F	G	Н	I
1	group	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8
2	group1	0.49674592	0.4677233	-0.0483518	-0.2401965	0.43226315	0.92187942	0.78451495	1.24561556
3	group2	-0.0136848	0.06421928	0.24973294	0.53737466	-0.0431699	1.28695797	1.17566285	0.5182841
4	group2	0.01182029	0.14132204	0.63396588	-0.2638858	0.54286107	0.49019515	1.18308357	0.9877163
5	group2	-0.0044884	-0.0588157	0.58687314	0.22217948	0.1791987	0.44559659	0.54895148	0.96529672
6	group1	-0.21083	0.36993122	0.54714402	0.36844285	0.47242539	0.85423332	1.18447251	1.12365475
7	group2	0.64296641	0.35597474	0.67135007	-0.1736976	-0.032966	0.76741604	1.2496947	0.59529623
8	group2	-0.1794864	0.19605731	0.16892836	0.06545587	-0.0143433	0.44885313	1.24465309	0.67890288
9	group1	0.11928887	-0.1925327	-0.1554748	0.50161556	-0.2461082	0.94051376	1.27231794	0.34739552
10	group1	0.15636136	0.52661018	-0.0777972	0.3801456	0.31512999	1.26082301	1.18188419	1.25883778
11	group1	0.62564213	0.31995212	0.24680715	0.01351141	0.39232323	1.19820673	0.86149827	0.49248353
12	group1	0.22323783	0.46813969	-0.2704524	-0.2713823	-0.0427953	0.73446676	0.69629343	0.3255452
13	group2	0.32798942	0.05329695	-0.054572	0.11371274	-0.1142362	0.36837606	1.10360499	1.17269313
14	group1	-0.1388634	0.65150567	0.14534509	0.53290573	0.30602502	0.94523044	0.4406129	1.11203726
15	group2	0.04906595	-0.1463183	0.06386825	0.01064908	0.33301805	0.33428674	1.22510453	1.25597255
16	group1	0.39118773	0.14052032	-0.0006453	0.50507523	0.51432391	0.44027245	0.4341059	1.14790925

数据要求: (csv / txt 格式文件)

- ▶ 数据至少3列以上,每列至少10行,最多支持25列和5000行数据
  - 第1列为分类类型数据,对应分组信息
  - 除第1列外, 第2列开始每一列都为数值类型数据, 表示每一个变量
  - 每一行可以理解成1个样本
- ▶ 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

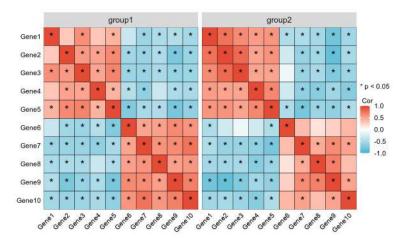


#### 映射

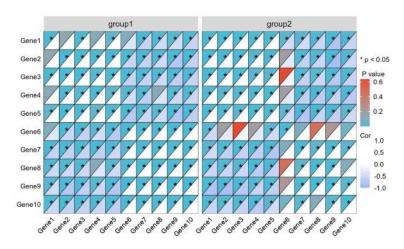


▶ 上半颜色映射:对应整体颜色的映射,当热图选择的是三角对角线类型时,则对应三角的上半颜色映射,如下:(第1个为热图类型-三角对角线;第2个为上半颜色映射变量相关系数,下半颜色映射变量 p值;第3个为上半颜色映射变量 p值,下半颜色映射变量相关系数



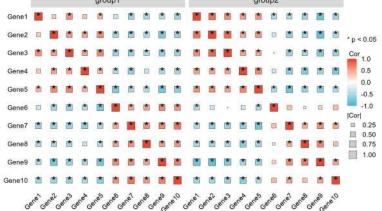






- ▶ 下半颜色映射: 当热图选择的是三角对角线类型时,则对应三角的下半颜色映射(如上: 上半颜色映射)
- 大小映射:可以对热图进行大小映射,只有在非三角对角线类型的时候会有效果,对应映射方块大小,默认为不映射,还可以选择相关系数绝对值,如下:



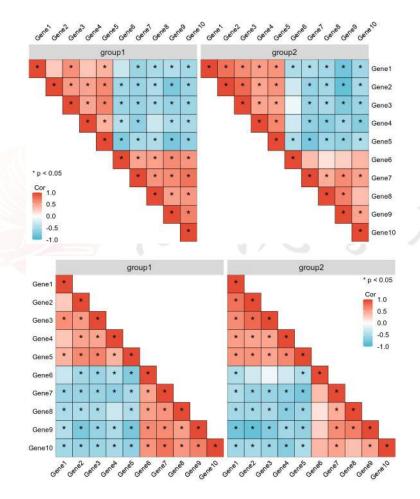




# 样式



▶ 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(第1个为上半,第2个为下半)

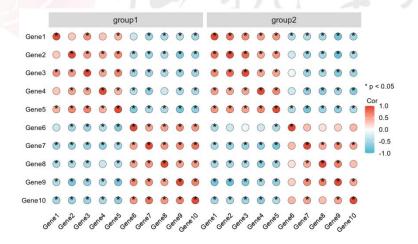


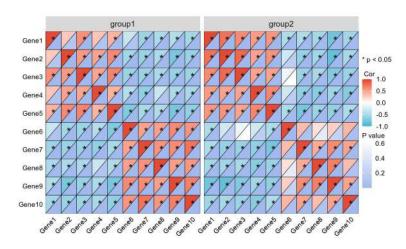


# 热图



▶ 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:





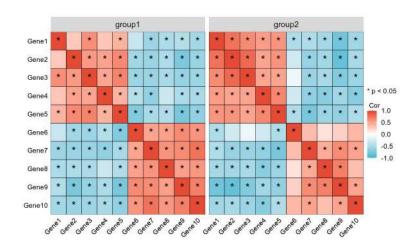
- ▶ 上半(全)颜色:选择三角类型时可以修改对应方块上半部分颜色,其他类型则修改对应整个方块颜色
- ▶ 下半颜色: 当选择三角类型时可以修改对应方块下半部分颜色, 其他类型无法修改
- ▶ 描边颜色: 可以修改热图对应方块的描边颜色
- ▶ 描边粗细:可以修改热图对应方块的描边粗细
- ▶ 大小比例:可以修改热图对应方块的大小比例,默认为1
- 不透明度:可以修改热图对应方块的不透明度,默认为1,表示完全不透明, 0表示完全透明

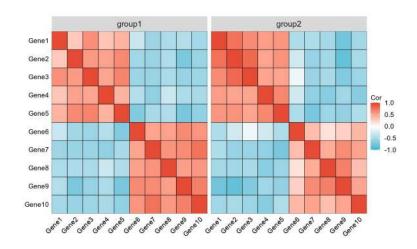


# 标注



- ▶ 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: (第1个为(\*p<0.05)映射,第2个为不映射)</p>
  - 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

分面



分面方向:可以选择并修改分组(分面)相关性热图进行分面的方向,默认为按列的方向,还可以选择按行进行分面(分组)

▶ 分面颜色: 可以修改各个分面(分组)的颜色

》 文字大小: 可以修改各个分面(分组)对应文本字体的大小, 默认为6pt



## 标题文本



#### ▶ 大标题: 大标题文本

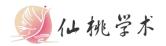
补充: 在要换行的中间插入\n。如果需要上标,可以用两个英文输入法下的大括号括住,比如 {{2}};如果需要下标,可以用两个英文输入法下的中括号括住,比如 [[2]]



▶ 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



▶ 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





# 补充结果

计方法: spea	arman									
A Jaren aper										
				gn	oup1组-相关系	数表格				
	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1		0.297	0.603	0.314	0.415	-0.328	-0.574	-0.493	-0.459	-0.53
Gene2	0.297		0.545	0.502	0.639	-0.542	-0.591	-0.477	-0.743	-0.584
Gene3	0.603	0.545		0.487	0.663	-0.578	-0.584	-0.499	-0.614	-0.572
Gene4	0.314	0.502	0.487		0.417	-0.447	-0.636	-0.315	-0.49	-0.505
Gene5	0.415	0.639	0.663	0.417		-0.708	-0.512	-0.465	-0.759	-0.631
Gene6	-0.328	-0.542	-0.578	-0.447	-0.708		0.498	0.512	0.629	0.575
Gene7	-0.574	-0.591	-0.584	-0.636	-0.512	0.498		0.58	0.625	0.75
Gene8	-0.493	-0.477	-0.499	-0.315	-0.465	0.512	0.58		0.519	0.559
Gene9	-0.459	-0.743	-0.614	-0.49	-0.759	0.629	0.625	0.519		0.679
Gene10	-0.53	-0.584	-0.572	-0.505	-0.631	0.575	0.75	0.559	0.679	

	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1		0.753	0.625	0.535	0.568	-0.475	-0.474	-0.531	-0.782	-0.551
Gene2	0.753		0.796	0.497	0.584	-0.284	-0.607	-0.54	-0.849	-0.514
Gene3	0.625	0.796		0.493	0.498	-0.104	-0.585	-0.504	-0.71	-0.582
Gene4	0.535	0.497	0.493		0.645	-0.283	-0.635	-0.681	-0.555	-0.686
Gene5	0.568	0.584	0.498	0.645		-0.469	-0.756	-0.58	-0.621	-0.474
Gene6	-0.475	-0.284	-0.104	-0.283	-0.469		0.361	0.159	0.253	0.394
Gene7	-0.474	-0.607	-0.585	-0.635	-0.756	0.361		0.452	0.625	0.503
Gene8	-0.531	-0.54	-0.504	-0.681	-0.58	0.159	0.452		0.689	0.349
Gene9	-0.782	-0.849	-0.71	-0.555	-0.621	0.253	0.625	0.689		0.497
Gene10	-0.551	-0.514	-0.582	-0.686	-0.474	0.394	0.503	0.349	0.497	

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



	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1		0.132	0.000867	0.111	0.0316	0.095	0.00175	0.00893	0.016	0.00447
Gene2	0.132		0.00327	0.00765	0.000332	0.00349	0.00117	0.0118	9.04e-06	0.00137
Gene3	0.000867	0.00327		0.00996	0.000164	0.00161	0.00137	0.00809	0.000665	0.00182
Gene4	0.111	0.00765	0.00996		0.0305	0.0194	0.000362	0.109	0.00954	0.00723
Gene5	0.0316	0.000332	0.000164	0.0305		3.58e-05	0.00631	0.0146	4.48e-06	0.000414
Gene6	0.095	0.00349	0.00161	0.0194	3.58e-05		0.00818	0.00638	0.000436	0.0017
Gene7	0.00175	0.00117	0.00137	0.000362	0.00631	0.00818		0.00152	0.000497	6.76e-06
Gene8	0.00893	0.0118	0.00809	0.109	0.0146	0.00638	0.00152		0.00555	0.00246
Gene9	0.016	9.04e-06	0.000665	0.00954	4.48e-06	0.000436	0.000497	0.00555		9.71e-05
Gene10	0.00447	0.00137	0.00182	0.00723	0.000414	0.0017	6.76e-06	0.00246	9.71e-05	

				group	2组-相关性检验	(p佰)表格				
	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10
Gene1		3.38e-05	0.00141	0.00859	0.00468	0.0219	0.0222	0.00919	1.06e-05	0.00639
Gene2	3.38e-05		5.45e-06	0.0158	0.00344	0.19	0.00214	0.00788	3.07e-07	0.0121
Gene3	0.00141	5.45e-06		0.0168	0.0156	0.638	0.00337	0.0142	0.000145	0.00357
Gene4	0.00859	0.0158	0.0168		0.000885	0.191	0.00112	0.000349	0.00594	0.000304
Gene5	0.00468	0.00344	0.0156	0.000885		0.0238	3.02e-05	0.00372	0.00158	0.0222
Gene6	0.0219	0.19	0.638	0.191	0.0238		0.0909	0.468	0.244	0.0627
Gene7	0.0222	0.00214	0.00337	0.00112	3.02e-05	0.0909		0.0305	0.00145	0.0144
Gene8	0.00919	0.00788	0.0142	0.000349	0.00372	0.468	0.0305		0.000279	0.103
Gene9	1.06e-05	3.07e-07	0.000145	0.00594	0.00158	0.244	0.00145	0.000279		0.0158
Gene10	0.00639	0.0121	0.00357	0.000304	0.0222	0.0627	0.0144	0.103	0.0158	

#### 相关性.xlsx

#### 相关性系数表格:

- 1. 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
- 2. 一般关系强度是: |r|>0.95: 显著性相关; |r|≥0.8: 高度相关; 0.5≤|r|<0.8: 中度相关; 0.3≤|r|<0.5: 低度相关; |r|<0.3: 剥相关

#### 相关性p值表格:

1. 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,<mark>可以直接写自己用 R 来进行分析和可视化即可</mark>,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