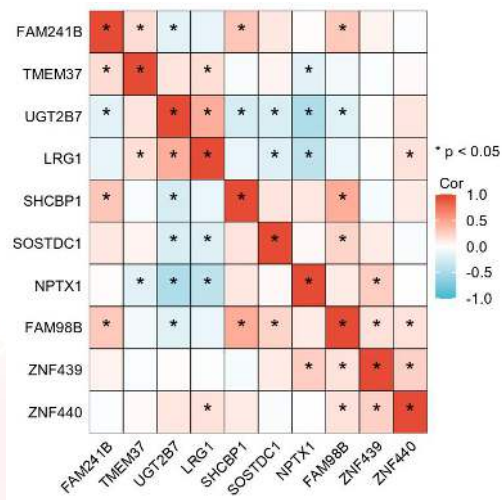


交互网络 - 相关性热图[云]



网址: <https://www.xiantao.love>



更新时间: 2023.02.24

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
ID 列表	8
统计	9
映射	10
布局	12
热图	13
标注	15
标题文本	16
图注	17
坐标轴	17
风格	18
图片	18
结果说明	19
主要结果	19
补充结果	20
方法学	22
如何引用	23
常见问题	24

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；相关性热图这是根据分子之间的相关性来展示的热图。
- 涉及的统计方法：
 - Pearson 相关：参数相关性检验，衡量两组之间是否存在线性关系
 - Spearman 相关：非参数相关性检验，通过秩次来判断两组是否存在相关性。如果不懂具体的选择条件，可以选择该方法
- 注意：相关不等于因果，也就是两者是可能不存在直接的关系

应用场景

基于云端数据 分析和可视化 各个基因之间（基因/分子两两间）表达的相关趋势情况

分析过程

云端数据 → 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
- 相关性分析
 - 将云端数据进行相关性分析

- ◆ 在参数[ID 列表]输入云端数据中需要进行分析的分子/基因名，输入多少个分子 ID 就对多少分子进行分析，如下：



- 如果输入分子 ID 少于 3 个或者输入分子 ID 与数据中匹配的个数少于 3 个，则不能进行分析：如下：



◆ 相关性分析表：将分子 ID 匹配到数据中的分子进行两两间相关性分析

● 相关性系数表

表1: 相关性系数表格

	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.183	0.326	0.2	0.299	-0.0651	-0.003	0.00586	0.0725	0.0798
TNMD	0.183		0.0379	0.177	0.0244	0.161	0.106	0.0636	-0.00236	0.119
DPM1	0.326	0.0379		0.132	0.452	-0.073	-0.00952	0.187	0.234	0.132
SCYL3	0.2	0.177	0.132		0.544	-0.04	0.0741	0.148	0.163	0.533
C1orf112	0.299	0.0244	0.452	0.544		-0.138	0.0032	0.181	0.329	0.41
FGR	-0.0651	0.161	-0.073	-0.04	-0.138		0.521	-0.00798	-0.0888	-0.0141
CFH	-0.003	0.106	-0.00952	0.0741	0.0032	0.521		-0.064	0.0836	-0.0934
FUCA2	0.00586	0.0636	0.187	0.148	0.181	-0.00798	-0.064		-0.144	0.171
GCLC	0.0725	-0.00236	0.234	0.163	0.329	-0.0888	0.0836	-0.144		0.147
NFYA	0.0798	0.119	0.132	0.533	0.41	-0.0141	-0.0934	0.171	0.147	

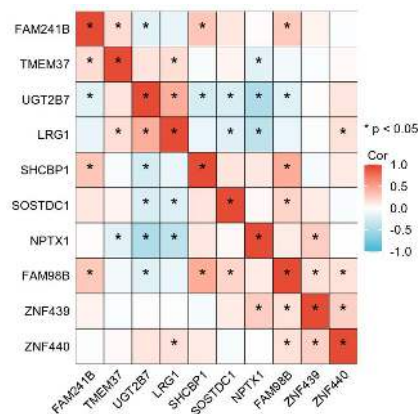
● 统计学检验 p 值

表2: 相关性检验表格

	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.0192	2.22e-05	0.0105	0.000108	0.409	0.97	0.941	0.358	0.311
TNMD	0.0192		0.631	0.024	0.757	0.0398	0.177	0.42	0.976	0.13
DPM1	2.22e-05	0.631		0.0941	1.35e-09	0.354	0.904	0.017	0.0026	0.0933
SCYL3	0.0105	0.024	0.0941		5.82e-14	0.613	0.347	0.0597	0.0374	2.5e-13
C1orf112	0.000108	0.757	1.35e-09	5.82e-14		0.0788	0.968	0.0211	1.83e-05	5.63e-08
FGR	0.409	0.0398	0.354	0.613	0.0788		9.73e-13	0.919	0.26	0.859
CFH	0.97	0.177	0.904	0.347	0.968	9.73e-13		0.417	0.289	0.236
FUCA2	0.941	0.42	0.017	0.0597	0.0211	0.919	0.417		0.0666	0.0289
GCLC	0.358	0.976	0.0026	0.0374	1.83e-05	0.26	0.289	0.0666		0.0614
NFYA	0.311	0.13	0.0933	2.5e-13	5.63e-08	0.859	0.236	0.0289	0.0614	

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



左图：（完整热图）

- 行、列都代表分子/基因
- 每一个小方格表示分子之间的相关系数，颜色越深，分子间越相关

右图：（完整热图-三角对角线）

- 行、列都代表基因/分子
- 每一个小方格分为两个部分（三角形），上部分表示分子间的相关系数，颜色越深，分子间越相关；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，分子间相关性越高，反之相关性越低
 - $\text{Correlation} < 0$ ，分子间呈负相关关系
 - $\text{Correlation} = 0$ ，分子间没有相关关系
 - $\text{Correlation} > 0$ ，分子间呈正相关关系
- Pvalue 代表分子间相关系数对应的 Pvalue 值，Pvalue 值越小，分子间相关系数越显著

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

（该样本数据：如下：）

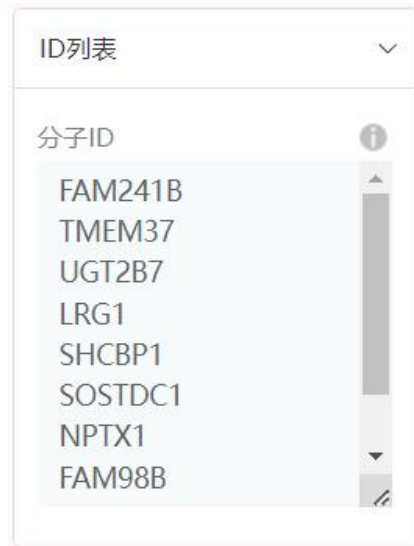
食管癌 / TCGA / TCGA-ESCA / RNAseq / STAR / TPM @过滤:去除正常 @处理:log2(value+1)



参数说明

(说明：标注了颜色的为常用参数。)

ID 列表



- 分子 ID：这部分输入的是热图部分的分子列表
 - 这部分分子可以来自「[单基因差异分析](#)」或者「[单基因相关性筛选](#)」两个模块筛选后再进行选择，建议是结合两者一起来看，如果想要热图结果好看一些，建议是从「单基因相关性筛选」模块中挑相关性高的分子进行可视化（因为相关趋势更加明显）
 - 分子 ID 不能少于 3 个或者分子 ID 与数据中匹配的个数少于 3 个
 - 一行一个 ID，可以是分子名，也可以是分子 ID，最多支持 20 个（最多支持 20 个分子进行相关性分析，也就是热图的行/列不能超过 20）

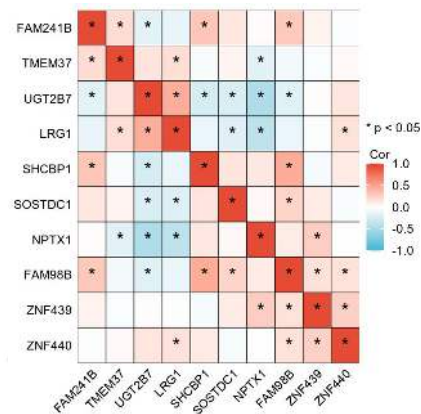
统计

统计

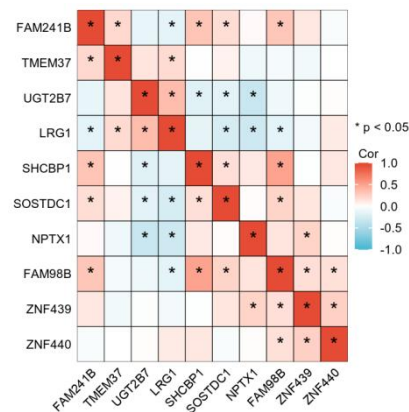
统计方法 Spearman

➤ 统计方法：可以选择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



映射

映射

上半颜色

映射

相关系数

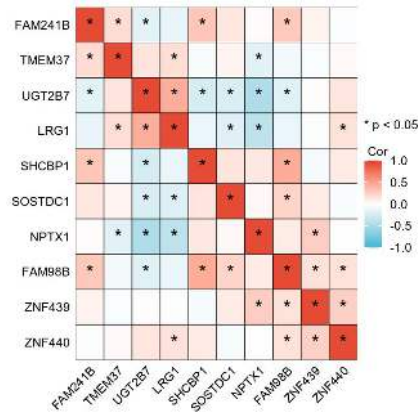
下半颜色

映射

p值

大小映射

不映射



- 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，热图-->类型-->三角对角线（第1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2个为上半颜色映射变量p值，下半颜色映射变量相关系数）

热图

类型

三角对角线

上半(全)颜色

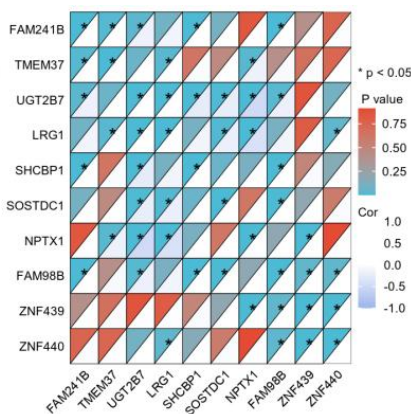
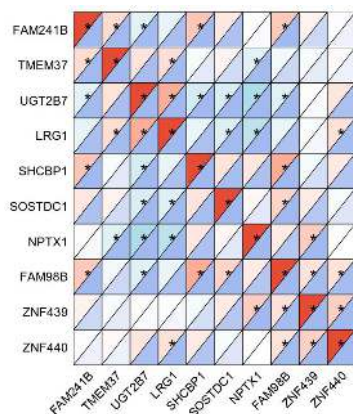
☒

☒

下半颜色

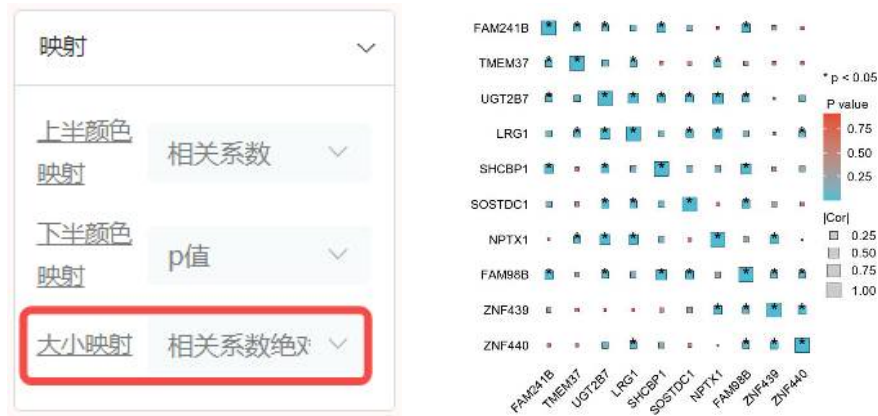
☒

☐



- 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：



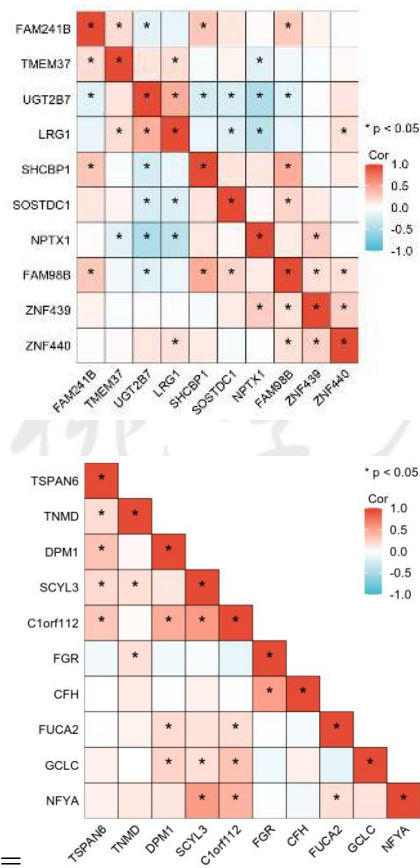
布局

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（只画上半部分热图）、下半热图（只画下半部分热图），如下：



=

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：

热图

类型 方块

上半(全)颜色 ☒ ☒

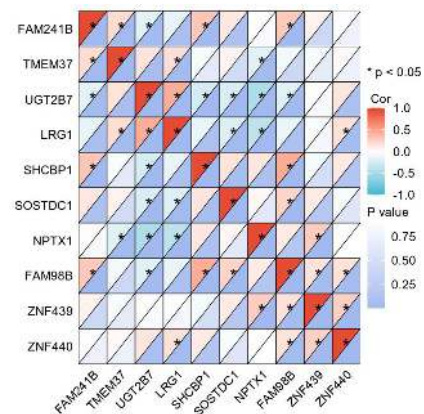
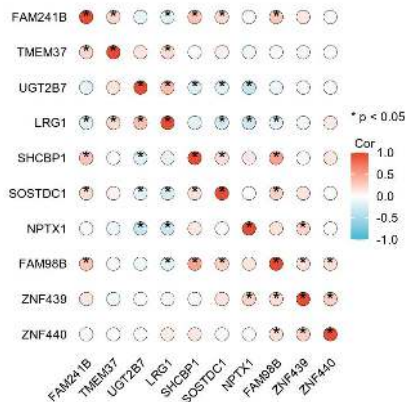
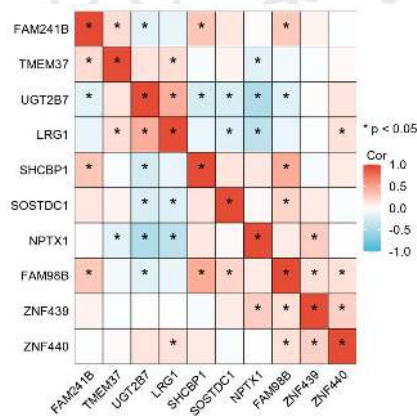
下半颜色 ☒ ☐

描边颜色 ☒

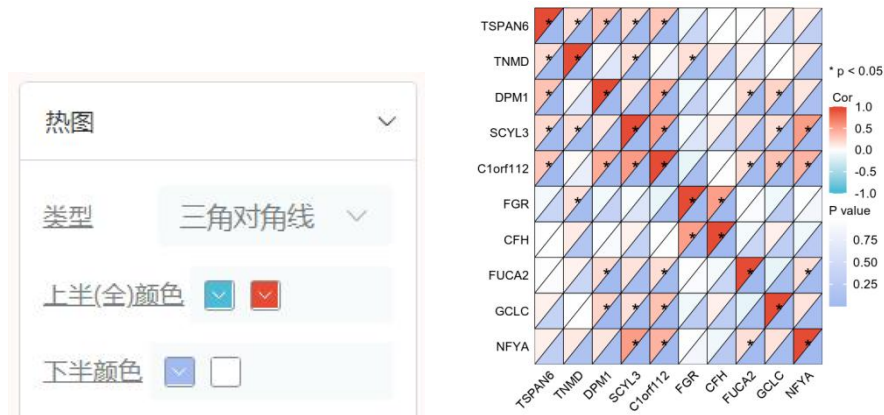
描边粗细 0.25pt

大小比例 1

不透明度 1



- 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: 可以修改热图对应块方的描边粗细
- 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

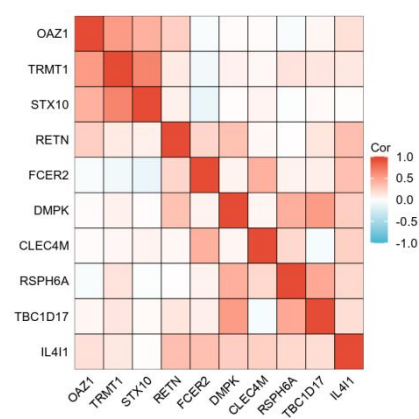
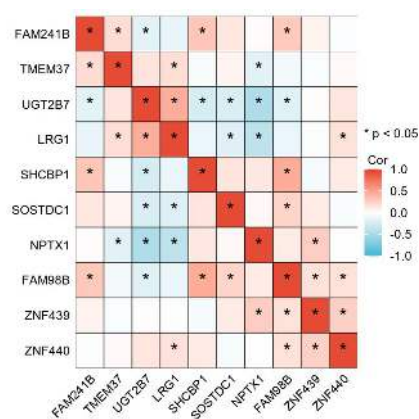
标注映射 星号(*p<0.05)

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



The screenshot shows a web form for entering title information. It consists of a main container with a light orange border. Inside, there is a top section with a label '标题' (Title) and a dropdown arrow. Below this is a section for the main title, containing a label '大标题' (Main Title) and a text input field with the placeholder '大标题内容' (Main Title Content).

➤ 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

坐标轴

x轴标注旋
转

45

- 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

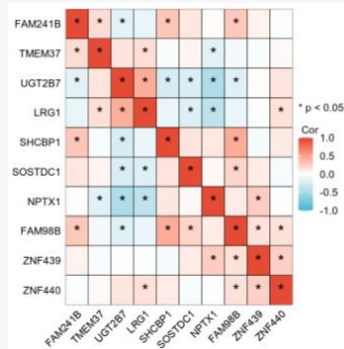
结果说明

主要结果

相关性热图-云

相关性热图: 用热图展示相关性情况

统计方法: spearman



相关性热图.pdf

相关性热图.tiff

相关性热图.pptx

分析数据.xlsx



补充结果

相关性分析

统计方法: spearman

表1: 相关系数表格

	FAM241B	TMEM37	UGT2B7	LRG1	SHCBP1	SOSTDC1	NPTX1	FAM98B	ZNF439	ZNF440
FAM241B		0.185	-0.169	-0.136	0.307	0.128	0.0152	0.293	0.0653	-0.0255
TMEM37	0.185		0.141	0.177	-0.0371	0.0581	-0.177	-0.0648	-0.034	0.0259
UGT2B7	-0.169	0.141		0.444	-0.247	-0.227	-0.476	-0.182	0.0138	0.133
LRG1	-0.136	0.177	0.444		-0.128	-0.197	-0.392	-0.126	-0.0191	0.16
SHCBP1	0.307	-0.0371	-0.247	-0.128		0.144	0.125	0.453	-0.0528	0.111
SOSTDC1	0.128	0.0581	-0.227	-0.197	0.144		0.0384	0.235	0.102	-0.0428
NPTX1	0.0152	-0.177	-0.476	-0.392	0.125	0.0384		0.104	0.273	-0.00871
FAM98B	0.293	-0.0648	-0.182	-0.126	0.453	0.235	0.104		0.159	0.16
ZNF439	0.0653	-0.034	0.0138	-0.0191	-0.0528	0.102	0.273	0.159		0.25
ZNF440	-0.0255	0.0259	0.133	0.16	0.111	-0.0428	-0.00871	0.16	0.25	

这里提供相关性分析表：可以查看分子之间的相关系数

- 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

	FAM241B	TMEM37	UGT2B7	LRG1	SHCBP1	SOSTDC1	NPTX1	FAM98B	ZNF439	ZNF440
FAM241B		0.0183	0.0312	0.0834	6.86e-05	0.104	0.847	0.000149	0.407	0.747
TMEM37	0.0183		0.0731	0.0242	0.639	0.461	0.0238	0.411	0.667	0.742
UGT2B7	0.0312	0.0731		2.91e-09	0.00148	0.0035	1.42e-10	0.02	0.861	0.09
LRG1	0.0834	0.0242	2.91e-09		0.103	0.0118	2.28e-07	0.108	0.809	0.041
SHCBP1	6.86e-05	0.639	0.00148	0.103		0.0671	0.111	1.26e-09	0.503	0.159
SOSTDC1	0.104	0.461	0.0035	0.0118	0.0671		0.626	0.00255	0.196	0.587
NPTX1	0.847	0.0238	1.42e-10	2.28e-07	0.111	0.626		0.188	0.000415	0.912
FAM98B	0.000149	0.411	0.02	0.108	1.26e-09	0.00255	0.188		0.0421	0.0407
ZNF439	0.407	0.667	0.861	0.809	0.503	0.196	0.000415	0.0421		0.00129
ZNF440	0.747	0.742	0.09	0.041	0.159	0.587	0.912	0.0407	0.00129	

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：可以查看各个分子间的相关性的检验 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) 对数据中主分子（变量）和次要分子（分子）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

数据：

- (1) 数据获取：从 TCGA 数据库（<https://portal.gdc.cancer.gov>）下载并整理 TCGA-ESCA(食管癌)项目 STAR 流程的 RNAseq 数据并提取 TPM 格式的数据 以及 临床数据
- (2) 数据过滤策略：去除正常
- (3) 数据处理方法： $\log_2(\text{value}+1)$

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. 选择哪些分子进行可视化？分子列表来自哪里？如何才能让热图更加好看？

答：分子可以来自「[单基因差异分析](#)」或者「[单基因相关性筛选](#)」两个模块筛选后再进行选择，建议是结合两者一起来看，可以分别从高和低各自挑选 10 个、15 个或者 20 个来进行可视化

如果想要[热图结果好看一些](#)，建议是从「单基因相关性筛选」模块中挑相关性高的分子进行可视化（因为相关趋势更加明显）

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. 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](#)。

4. 为什么在分子输入框内输入了很多的分子，但是出来的图只有几个分子，数目对不上？



答：输入的分子会进行匹配的，只有是正式分子名才会匹配上，而蛋白或者别名有可能会匹配不上，如果是要精准匹配，建议是输入 ENSG 编号（可以利用 ID 转换工具转换 ID）

5. 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