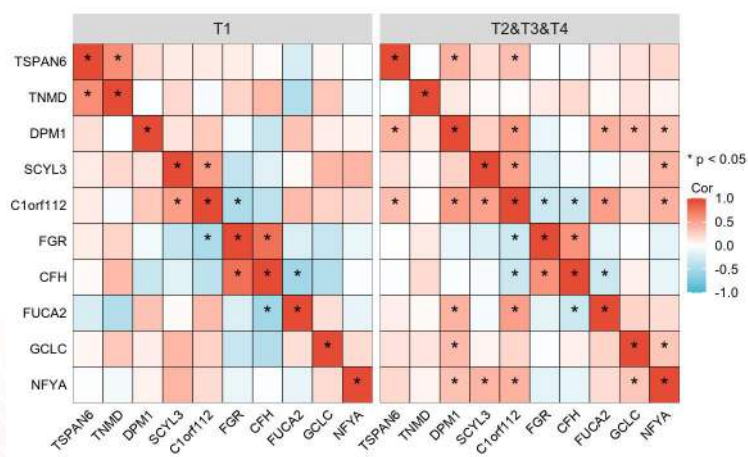


交互网络 - 相关性热图-亚组[云]



网址: <https://www.xiantao.love>



更新时间: 2023.03.08

目录

基本概念	3
应用场景	3
分析过程	4
结果解读	8
数据格式	9
参数说明	11
ID 列表	11
统计	12
映射	13
样式	15
热图	16
标注	18
分面	19
标题文本	19
图注	20
坐标轴	20
风格	21
图片	21
结果说明	22
主要结果	22
补充结果	23
方法学	25
如何引用	26
常见问题	27

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

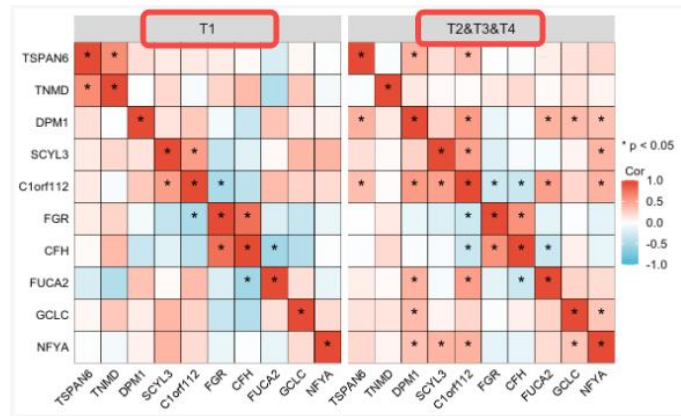
■ 将云端数据进行相关性分析

◆ 在特殊参数[临床变量]选择清洗好的临床变量以及对应进行分组，以将云端数据进行分组处理

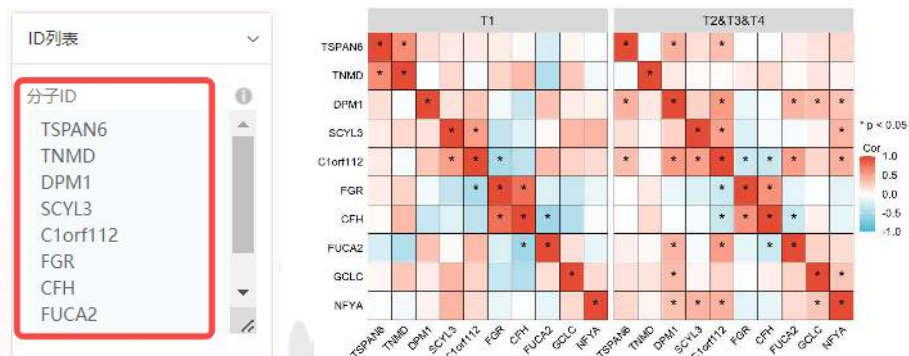


- 临床变量：选择需要用来将云端数据进行分组的变量，可以用输入关键词的形式进行搜索变量
- 分组：选择临床变量中包含的分类组成分组，对应热图中各个分面，多少个分组就代表多少个分面，一个框表示一个分组，可以通过“+”来添加分组，通过“-”来减少分组，一个分组里边可以点击“x”来减少分组中的分类个数，如下：第一个框为 T1 分类组成一个分组，第二个框为 T2、T3、T4 三个分类组成的分组





- ◆ 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：将各分组中分子 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

T1组-相关系数表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.619	0.175	0.111	0.114	0.1	0.0333	-0.235	0.0456	-0.0175
TNMD	0.619		-0.01	0.205	-0.0501	0.231	0.376	-0.453	0.298	-0.0656
DPM1	0.175	-0.01		0.153	0.296	-0.0737	-0.326	0.321	0.0965	0.0702
SCYL3	0.111	0.205	0.153		0.526	-0.354	-0.181	0.0316	0.395	0.402
C1orf112	0.114	-0.0501	0.296	0.526		-0.489	-0.384	0.363	0.24	0.189
FGR	0.1	0.231	-0.0737	-0.354	-0.489		0.765	-0.218	-0.339	-0.125
CFH	0.0333	0.376	-0.326	-0.181	-0.384	0.765		-0.554	-0.449	-0.0211
FUCA2	-0.235	-0.453	0.321	0.0316	0.363	-0.218	-0.554		0.167	-0.137
GCLC	0.0456	0.298	0.0965	0.395	0.24	-0.339	-0.449	0.167		0.191
NFYA	-0.0175	-0.0656	0.0702	0.402	0.189	-0.125	-0.0211	-0.137	0.191	

T2&T3&T4组-相关系数表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		-0.0327	0.419	0.174	0.355	-0.00705	-0.0273	0.0942	0.156	0.207
TNMD	-0.0327		0.12	0.0261	0.0194	0.113	0.204	0.0328	0.142	0.0609
DPM1	0.419	0.12		0.24	0.538	-0.13	-0.0325	0.416	0.364	0.333
SCYL3	0.174	0.0261	0.24		0.516	-0.209	-0.0503	-0.0537	0.0594	0.405
C1orf112	0.355	0.0194	0.538	0.516		-0.301	-0.313	0.52	0.219	0.422
FGR	-0.00705	0.113	-0.13	-0.209	-0.301		0.59	-0.154	-0.0229	-0.157
CFH	-0.0273	0.204	-0.0325	-0.0503	-0.313	0.59		-0.308	0.0818	-0.141
FUCA2	0.0942	0.0328	0.416	-0.0537	0.52	-0.154	-0.308		0.23	0.188
GCLC	0.156	0.142	0.364	0.0594	0.219	-0.0229	0.0818	0.23		0.304
NFYA	0.207	0.0609	0.333	0.405	0.422	-0.157	-0.141	0.188	0.304	

● 统计学检验 p 值

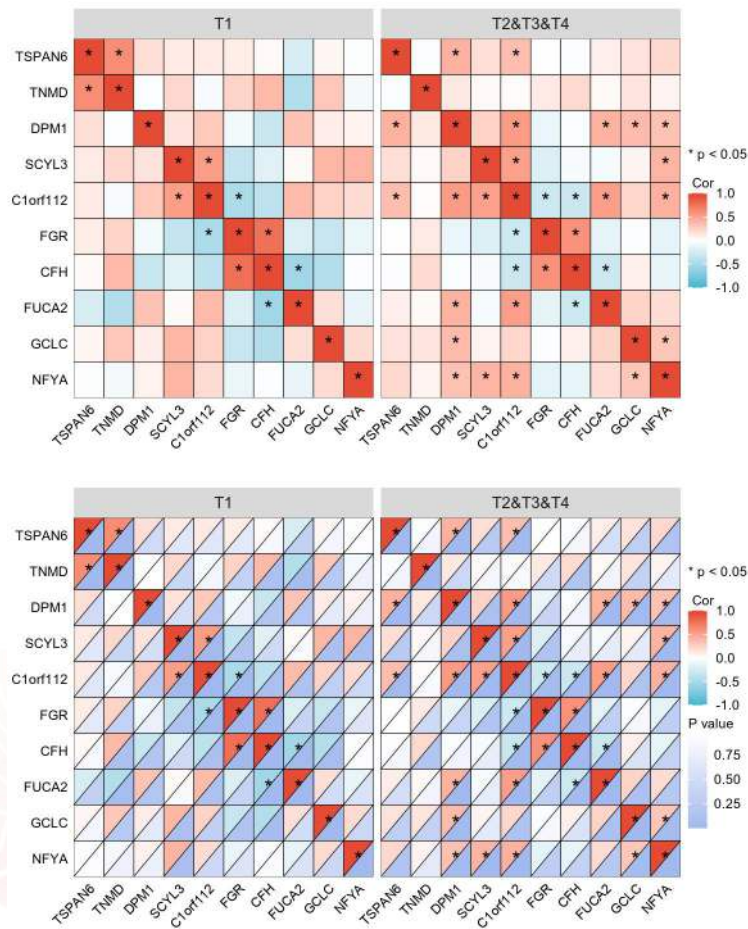
T1组-相关性检验(p值)表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.00473	0.473	0.652	0.642	0.684	0.892	0.333	0.853	0.943
TNMD	0.00473		0.968	0.4	0.839	0.34	0.112	0.0515	0.215	0.79
DPM1	0.473	0.968		0.533	0.218	0.764	0.173	0.18	0.694	0.775
SCYL3	0.652	0.4	0.533		0.0206	0.137	0.459	0.898	0.0944	0.0882
C1orf112	0.642	0.839	0.218	0.0206		0.0334	0.104	0.126	0.322	0.437
FGR	0.684	0.34	0.764	0.137	0.0334		0.000136	0.371	0.156	0.611
CFH	0.892	0.112	0.173	0.459	0.104	0.000136		0.0138	0.0537	0.932
FUCA2	0.333	0.0515	0.18	0.898	0.126	0.371	0.0138		0.495	0.576
GCLC	0.853	0.215	0.694	0.0944	0.322	0.156	0.0537	0.495		0.433
NFYA	0.943	0.79	0.775	0.0882	0.437	0.611	0.932	0.576	0.433	

T2&T3&T4组-相关性检验(p值)表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.827	0.00336	0.241	0.0144	0.962	0.856	0.529	0.297	0.162
TNMD	0.827		0.423	0.862	0.897	0.451	0.168	0.827	0.342	0.684
DPM1	0.00336	0.423		0.104	9.75e-05	0.384	0.828	0.00364	0.0119	0.0223
SCYL3	0.241	0.862	0.104		0.000205	0.159	0.737	0.72	0.691	0.00476
C1orf112	0.0144	0.897	9.75e-05	0.000205		0.0397	0.0322	0.000178	0.139	0.00313
FGR	0.962	0.451	0.384	0.159	0.0397		1.3e-05	0.302	0.879	0.291
CFH	0.856	0.168	0.828	0.737	0.0322	1.3e-05		0.0351	0.585	0.344
FUCA2	0.529	0.827	0.00364	0.72	0.000178	0.302	0.0351		0.12	0.205
GCLC	0.297	0.342	0.0119	0.691	0.139	0.879	0.585	0.12		0.0378
NFYA	0.162	0.684	0.0223	0.00476	0.00313	0.291	0.344	0.205	0.0378	

- 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 Pvalue < 0.05

- Correlation 代表相关性系数(Correlation): $|\text{Correlation}|$ 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



数据参数

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：）

数据参数

云端数据 ⓘ

食管腺癌 / TCGA / TCGA-ESAD / RNAseq / STAR / TPM @过滤:去除正常+去除无临床信息 @处理:log2(...)



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

上半颜色

映射

相关系数

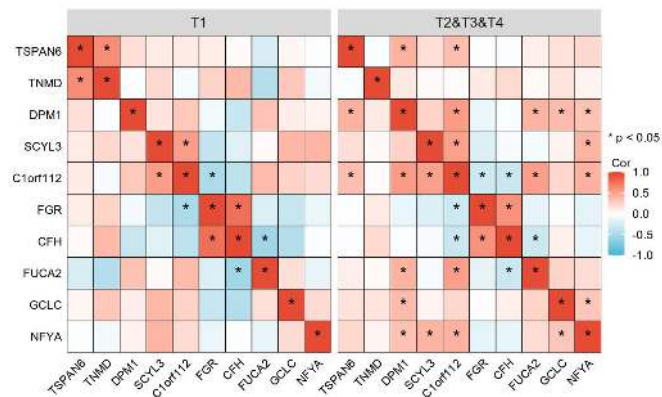
下半颜色

映射

p值

大小映射

不映射

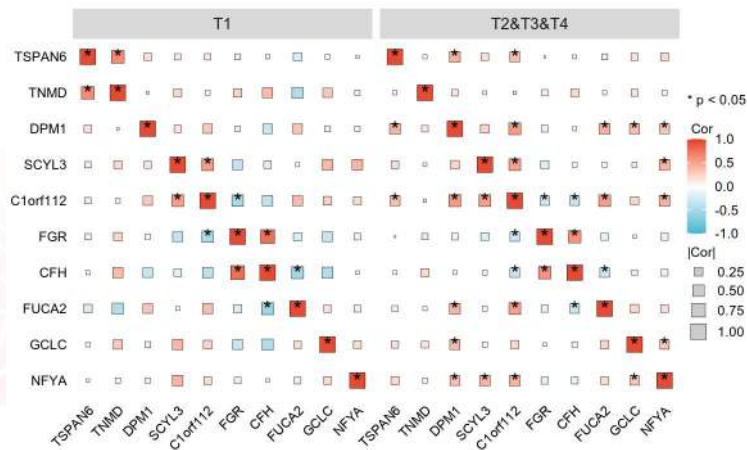


- 上半颜色映射：对应整体颜色的映射，当热图选择的是三角对角线类型时，则对应三角的上半颜色映射，热图-->类型-->三角对角线（第1个为上半颜色映射变量相关系数，下半颜色映射变量p值；第2个为上半颜色映射变量p值，下半颜色映射变量相关系数）



- 13

- 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：



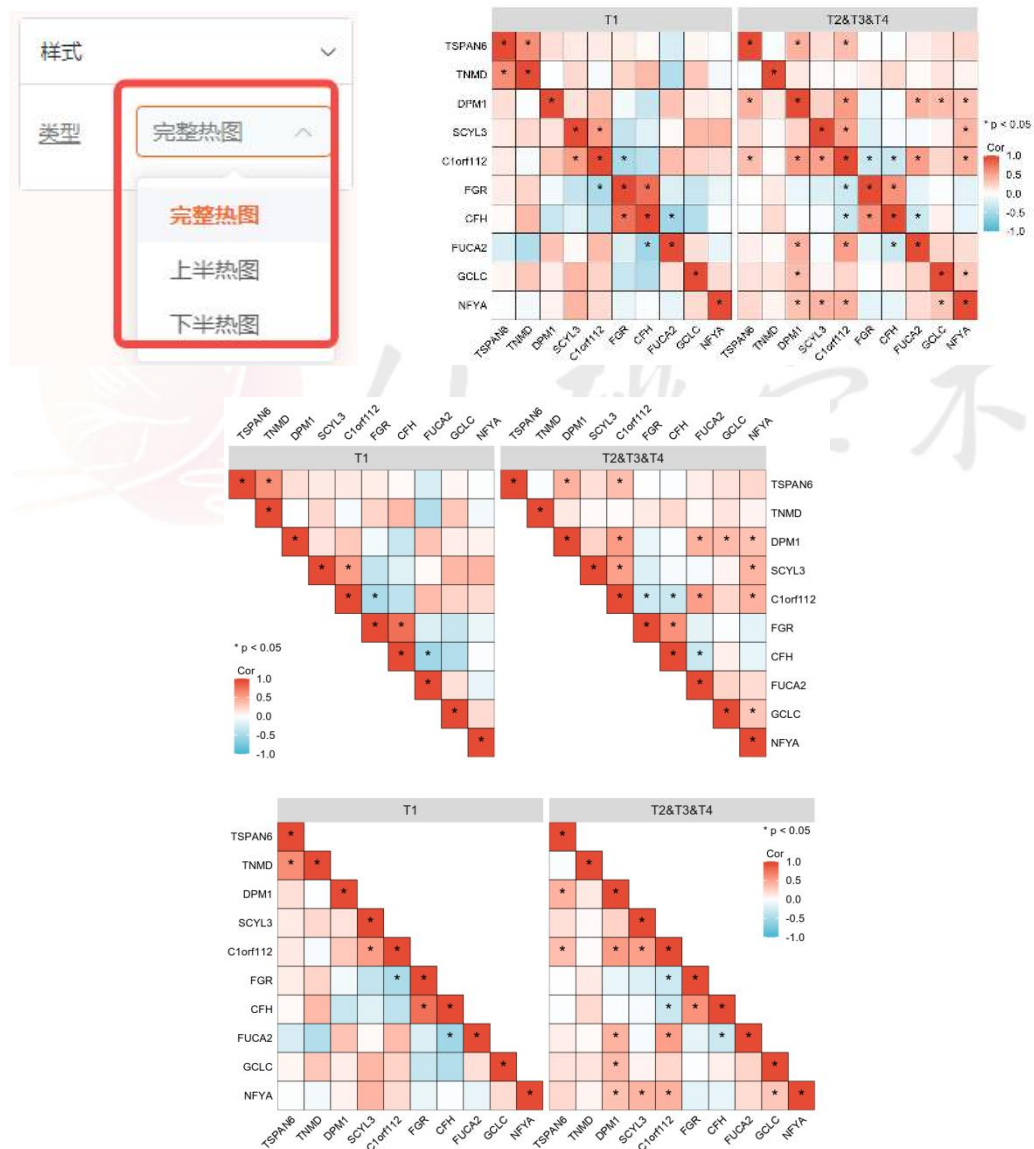
样式

样式

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 ☒ ☒

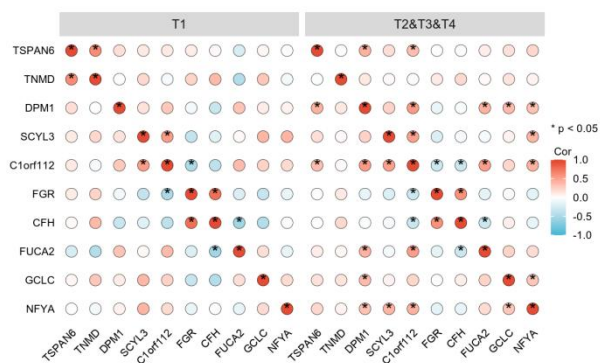
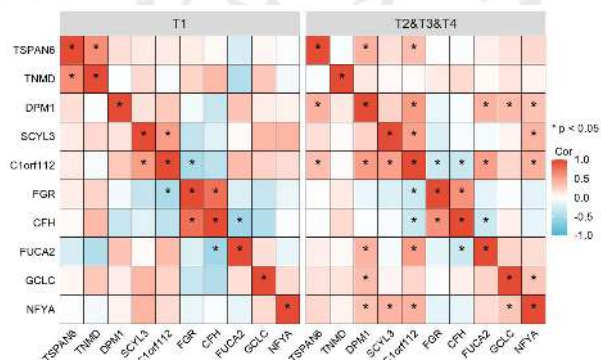
下半颜色 ☒ ☐

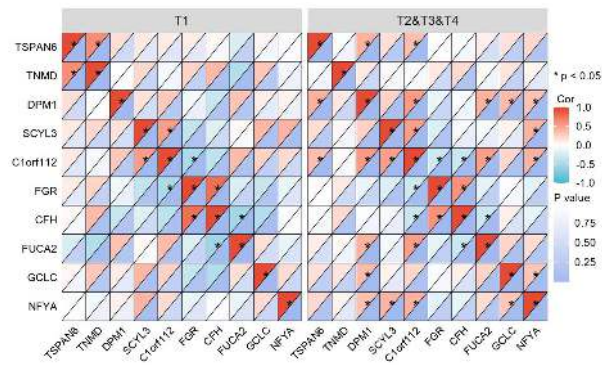
描边颜色 ☒

描边粗细 0.25pt

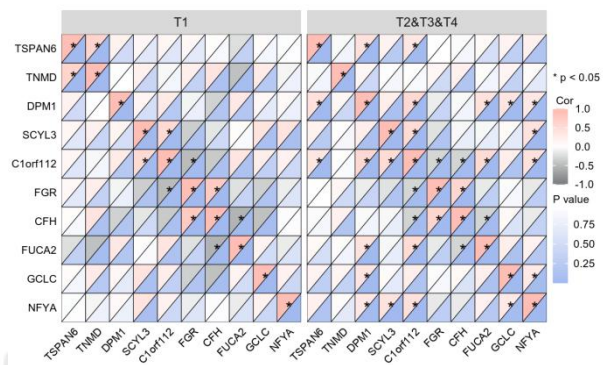
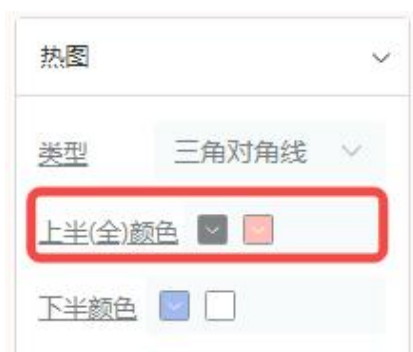
大小比例 1

不透明度 1

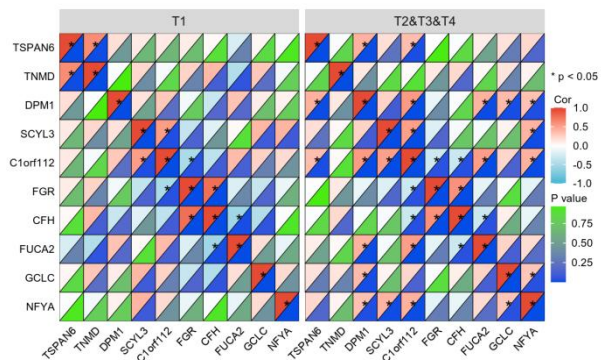




- 上半(全)颜色: 选择三角类型时可以修改对应方块上半部分颜色, 其他类型则修改对应整个方块颜色



- 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

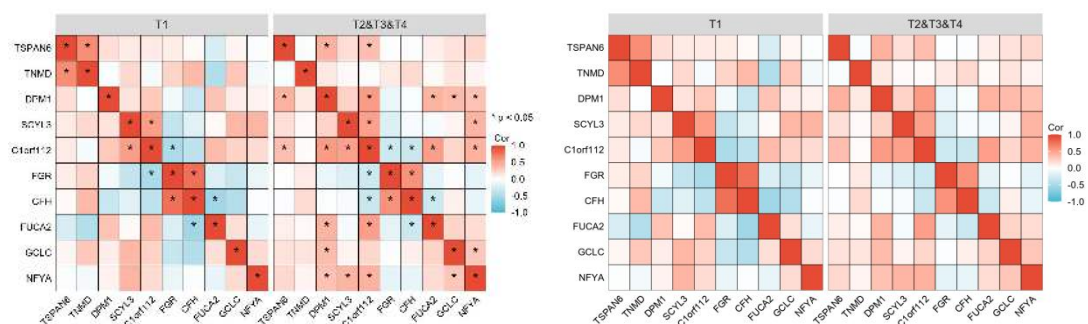
标注映射 星号(*p<0.0)

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

分面



- 分面方向：可以选择热图进行分面的方向，默认为按列进行分面，还可以选择按行进行分面
- 分面颜色：可以选择并修改进行分面时各分面的颜色
- 文字大小：可以选择并修改进行分面映射时各分面对应文本字体的大小

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]]

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

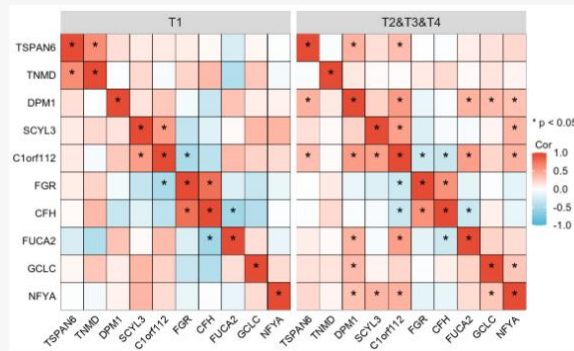
结果说明

主要结果

相关性热图-亚组

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



补充结果

T1组-相关系数表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.619	0.175	0.111	0.114	0.1	0.0333	-0.235	0.0456	-0.0175
TNMD	0.619		-0.01	0.205	-0.0501	0.231	0.376	-0.453	0.298	-0.0656
DPM1	0.175	-0.01		0.153	0.296	-0.0737	-0.326	0.321	0.0965	0.0702
SCYL3	0.111	0.205	0.153		0.526	-0.354	-0.181	0.0316	0.395	0.402
C1orf112	0.114	-0.0501	0.296	0.526		-0.489	-0.384	0.363	0.24	0.189
FGR	0.1	0.231	-0.0737	-0.354	-0.489		0.765	-0.218	-0.339	-0.125
CFH	0.0333	0.376	-0.326	-0.181	-0.384	0.765		-0.554	-0.449	-0.0211
FUCA2	-0.235	-0.453	0.321	0.0316	0.363	-0.218	-0.554		0.167	-0.137
GCLC	0.0456	0.298	0.0965	0.395	0.24	-0.339	-0.449	0.167		0.191
NFYA	-0.0175	-0.0656	0.0702	0.402	0.189	-0.125	-0.0211	-0.137	0.191	

T2&T3&T4组-相关系数表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		-0.0327	0.419	0.174	0.355	-0.00705	-0.0273	0.0942	0.156	0.207
TNMD	-0.0327		0.12	0.0261	0.0194	0.113	0.204	0.0328	0.142	0.0609
DPM1	0.419	0.12		0.24	0.538	-0.13	-0.0325	0.416	0.364	0.333
SCYL3	0.174	0.0261	0.24		0.516	-0.209	-0.0503	-0.0537	0.0594	0.405
C1orf112	0.355	0.0194	0.538	0.516		-0.301	-0.313	0.52	0.219	0.422
FGR	-0.00705	0.113	-0.13	-0.209	-0.301		0.59	-0.154	-0.0229	-0.157
CFH	-0.0273	0.204	-0.0325	-0.0503	-0.313	0.59		-0.308	0.0818	-0.141
FUCA2	0.0942	0.0328	0.416	-0.0537	0.52	-0.154	-0.308		0.23	0.188
GCLC	0.156	0.142	0.364	0.0594	0.219	-0.0229	0.0818	0.23		0.304
NFYA	0.207	0.0609	0.333	0.405	0.422	-0.157	-0.141	0.188	0.304	

这里提供各分组间的相关性分析表：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

T1组-相关性检验(p值)表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.00473	0.473	0.652	0.642	0.684	0.892	0.333	0.853	0.943
TNMD	0.00473		0.968	0.4	0.839	0.34	0.112	0.0515	0.215	0.79
DPM1	0.473	0.968		0.533	0.218	0.764	0.173	0.18	0.694	0.775
SCYL3	0.652	0.4	0.533		0.0206	0.137	0.459	0.898	0.0944	0.0882
C1orf112	0.642	0.839	0.218	0.0206		0.0334	0.104	0.126	0.322	0.437
FGR	0.684	0.34	0.764	0.137	0.0334		0.000136	0.371	0.156	0.611
CFH	0.892	0.112	0.173	0.459	0.104	0.000136		0.0138	0.0537	0.932
FUCA2	0.333	0.0515	0.18	0.898	0.126	0.371	0.0138		0.495	0.576
GCLC	0.853	0.215	0.694	0.0944	0.322	0.156	0.0537	0.495		0.433
NFYA	0.943	0.79	0.775	0.0882	0.437	0.611	0.932	0.576	0.433	

T2&T3&T4组-相关性检验(p值)表格										
	TSPAN6	TNMD	DPM1	SCYL3	C1orf112	FGR	CFH	FUCA2	GCLC	NFYA
TSPAN6		0.827	0.00336	0.241	0.0144	0.962	0.856	0.529	0.297	0.162
TNMD	0.827		0.423	0.862	0.897	0.451	0.168	0.827	0.342	0.684
DPM1	0.00336	0.423		0.104	9.75e-05	0.384	0.828	0.00364	0.0119	0.0223
SCYL3	0.241	0.862	0.104		0.000205	0.159	0.737	0.72	0.691	0.00476
C1orf112	0.0144	0.897	9.75e-05	0.000205		0.0397	0.0322	0.000178	0.139	0.00313
FGR	0.962	0.451	0.384	0.159	0.0397		1.3e-05	0.302	0.879	0.291
CFH	0.856	0.168	0.828	0.737	0.0322	1.3e-05		0.0351	0.585	0.344
FUCA2	0.529	0.827	0.00364	0.72	0.000178	0.302	0.0351		0.12	0.205
GCLC	0.297	0.342	0.0119	0.691	0.139	0.879	0.585	0.12		0.0378
NFYA	0.162	0.684	0.0223	0.00476	0.00313	0.291	0.344	0.205	0.0378	

这里提供各分组间的相关性分析表：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) 分别对各个组的数据的变量进行两两相关性分析
- (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