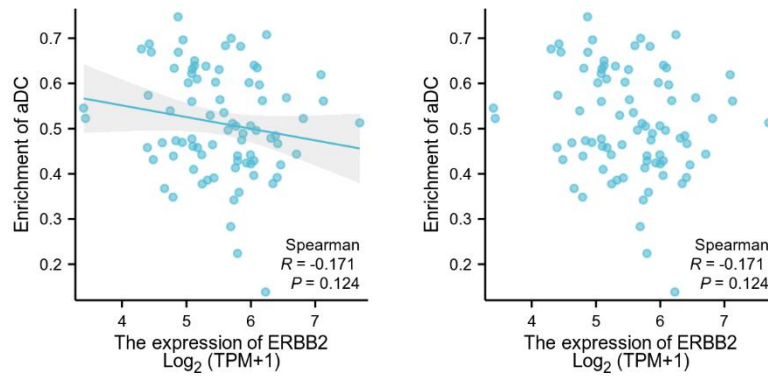


交互网络 - [免疫浸润-云] 相关性散点图



网址: <https://www.xiantao.love>



更新时间: 2023.03.10

目录

基本概念	3
应用场景	3
分析流程	4
主要结果	5
云端数据	8
参数说明	9
特殊参数	9
算法	9
统计	10
样式	11
点	12
拟合线	13
标题	14
风格	15
图片	15
结果说明	16
主要结果	16
补充结果	17
方法学	19
如何引用	20
常见问题	21

基本概念

- 免疫浸润分析：利用转录组或者其他组学的数据，通过算法计算组织中免疫细胞的分数情况，推测组织中免疫细胞的构成情况。
- 免疫浸润算法
 - **ssGSEA (single sample GSEA)**：用给定的基因集对每一个样本计算对应基因集的富集得分。在免疫浸润中的 ssGSEA 则是利用每一类免疫细胞特定的 markers（来源于说明文本中的引文）作为基因集去计算每个样本中每一类免疫细胞的富集得分，推断单个样本的免疫细胞的浸润情况。
 - **estimate**：estimate 包提供的通过内置的 markers 计算样本的免疫浸润得分、基质得分和估计得分。（主要看基质得分和免疫浸润得分，可以简单理解得分越高，内环境中基质成分和免疫浸润越高）
- 相关性分析：通过相关性分析的方法，获得两组数据之间的相关性程度。
- 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，数据需要满足双正态性。
 - 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。数据可以不需要满足正态性，如果不懂具体的选择条件，可以先选择该方法。

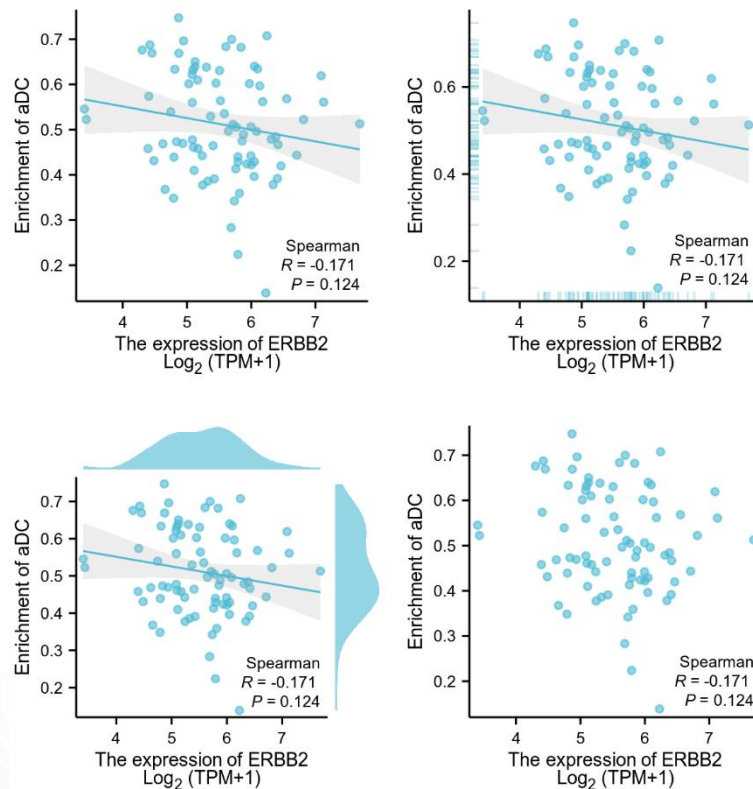
应用场景

基于公共数据（**云端数据**）直接分析所选分子与预先计算好的**单个**免疫细胞分数之间的相关性。一般绘制相关性散点图进行可视化。

分析流程



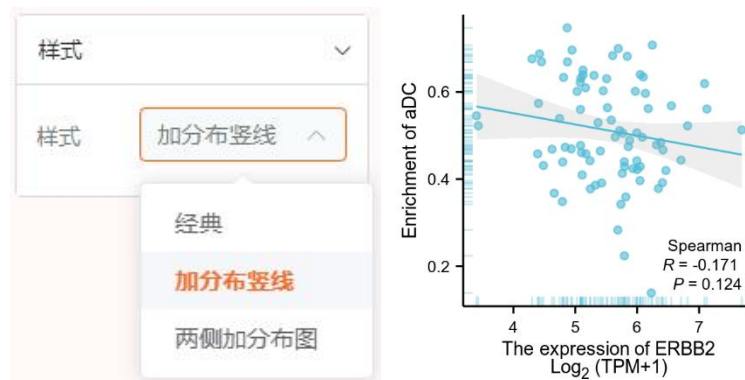
主要结果



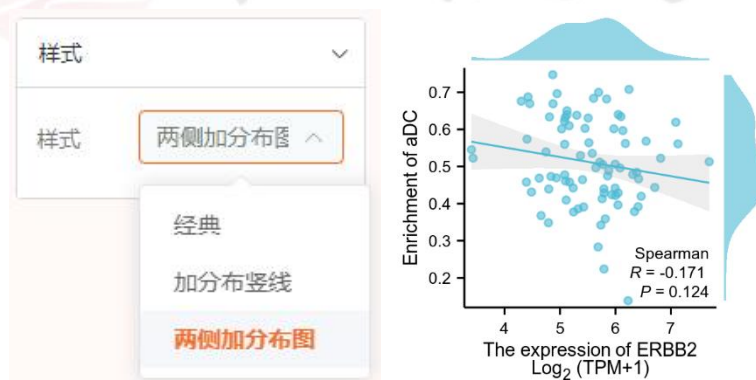
通过棒棒糖图展示相关性结果。

- 每个点为一个样本，对应有分子 ERBB2 和免疫细胞 aDC 的数值，横坐标为分子的 $\log_2(\text{value}+1)$ ，纵坐标为细胞的免疫浸润分数。
 - 图中的线代表拟合线，周围的阴影部分代表拟合直线的置信区间。
 - 图中的右下角有对应的统计方法和相关统计量。
 - 相关系数为正数，说明两个变量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；相关程度还需要看 相关系数的绝对值大小。
- 相关系数强弱，一般可以认为：
 - 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

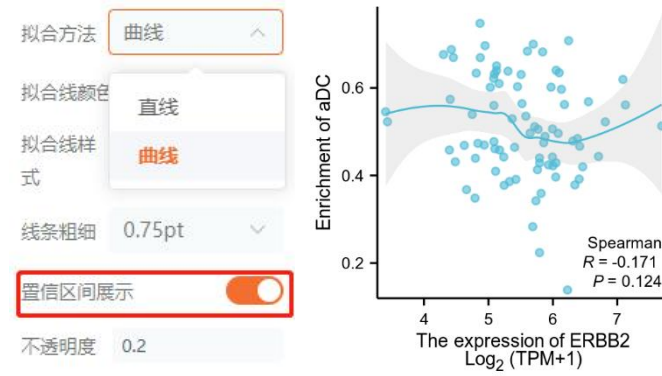
- 绝对值在 0.3-0.5: 相关程度一般
- 绝对值在 0.3 以下: 相对弱或者不相关
- 一般不是所有的分子都会有这么强的相关, 也可以只看 p 值, p 值越小越好。
- 可视化形式: 选择对应参数中【样式】, 默认-经典



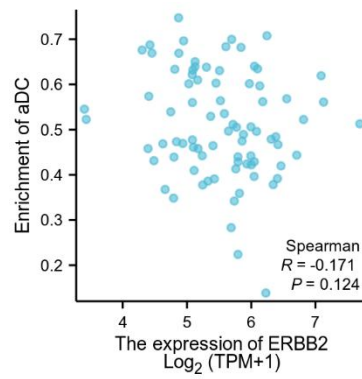
- **分布竖线**: 绘图的具体数据信息在 xy 轴上显示。



- **分布图**: 绘图数据的整体分布情况, x 轴对应图形的上方, y 轴对应图形的右侧。



■ 拟合曲线和置信区间




■ 不展示拟合线

云端数据


数据参数
重置参数

云端数据 ⓘ

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




疾病名/来源/数据集/平台/分析流程/数据格式




@数据处理方式


云端数据


选择疾病 

选择数据处理方式, 默认 去除正常 (过滤样本)、log2(value+1)

×

疾病 请选择 

数据过滤: 去除正常 

数据格式: log2(value+1) 

	疾病系统	疾病名	疾病英文	来源	获取时间	数据集	平台	Wo
<input checked="" type="checkbox"/>	食管	食管鳞癌	Esophagus squamous cell carcinoma	TCGA	202208	TCGA-ESCC	RNAseq	STA
<input type="checkbox"/>	食管	食管癌	Esophageal carcinoma	TCGA	202208	TCGA-ESCA	RNAseq	STA
<input type="checkbox"/>	食管	食管癌	Esophageal carcinoma	TCGA	202208	TCGA-ESCA	RNAseq	STA
<input type="checkbox"/>	食管	食管腺癌	Esophagus adenocarcinoma	TCGA	202208	TCGA-ESAD	RNAseq	STA

共 78 条 上一页 1 2 3 4 5 6 ... 8 下一页

ⓘ 只有合适这个模块的云端数据才会展示

确认

本模块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。注意查看当前数据参数选中的云端数据。

参数说明

(说明：标注了颜色的为常用参数。)

特殊参数



特殊参数

分子: [ERBB2[ENSG00000141736.14]]

主要参数

算法

标题

OAZ1[ENSG00000104904.12]
 TRMT1[ENSG00000104907.12]
 STX10[ENSG00000104915.15]
 RETN[ENSG00000104918.8]
 FCER2[ENSG00000104921.15]
 DMPK[ENSG00000104936.19]
 CLEC4M[ENSG00000104938.18]

- 分子：下拉框将列出对应所选数据集分子，可以输入关键字搜索分子，基因 symbol 或 Ensembl ID，[只能选单个分析](#)。

算法



算法

算法: ssgsea

细胞: aDC

- **算法**：选择预先计算好的每个样本的免疫细胞浸润分数的方法，可以选择 ssgsea、estimate。不同算法之间对应的内容会有一定差别，选择好算法后，

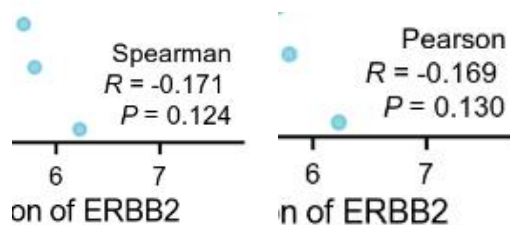
下面“细胞”参数需要点击“确认”后会进行更新，具体算法的参考文献可查看返回结果的<方法学>对应的内容。

- **细胞**：算法下提供的细胞，默认是选择第一个内容。当算法为 `ssgsea` 时是按照细胞类型排序后的第一个细胞类型，当算法为 `estimate` 时是 `StromalScore`（基质得分）。此参数需要点击“确认”后才会根据算法进行更新，只能选单个分析。

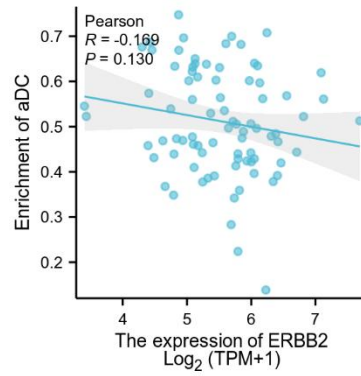
统计



- **统计方法**：可选择 `Spearman`、`Pearson`。统计方法的选择依据可以参考“基本概念”中统计方法的说明。
- **标注位置**：统计信息在图中的标注位置，一般展示 统计方法、相关系数值、p 值，可以选 `左下`、`左上`、`右下`、`右上`、`无`。



显示内容



位置 - 左上

- 标注颜色：标注文字的颜色，默认为纯黑，不受配色方案全局性影响。

样式



- 样式：绘图的样式。可选择 经典、加分布竖线、两侧加分布图。

点

点

填充色

描边色

样式 ×

大小

不透明度

- **填充色**：点的填充色颜色选项，只要修改第一色卡。受配色方案全局性修改。
- **描边色**：点的描边色颜色选项，只要修改第一色卡。受配色方案全局性修改。
- **样式**：点的样式类型，可选择 圆形、正方形、菱形、三角形、倒三角。单选，[选择后将全局变化](#)。
- **大小**：点的大小。
- **不透明度**：点的透明度。0 为完全透明，1 为完全不透明。

拟合线

拟合线

展示

☒

拟合方法

直线

拟合线颜色

☒
☒

拟合线样式

实线

线条粗细

0.75pt

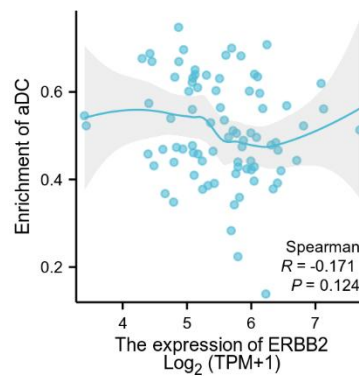
置信区间展示

☒

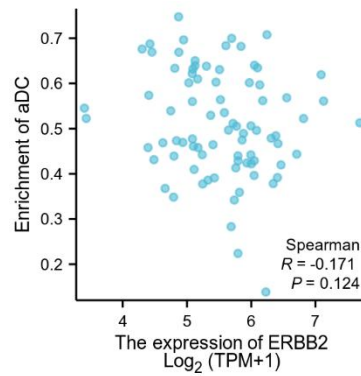
不透明度

0.2

- 展示：是否展示拟合线。
- 拟合方法：修改拟合方法，可以选 直线、曲线。曲线如下：



- 拟合线颜色：拟合线的颜色，只要修改第一色卡。受配色方案全局性修改。
- 拟合线样式：线条的类型，可选 实线、虚线。
- 线条粗细：线的粗细，默认为 0.75pt
- 置信区间展示：是否展示置信区间（阴影部分），默认为展示。不展示如下：



- 不透明度：阴影的透明度。0 为完全透明，1 为完全不透明。

标题



- 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 $[2]$

风格



- 边框：是否添加外框
- 网格：是否添加网格
- 文字大小：针对图中所有文字整体的大小控制



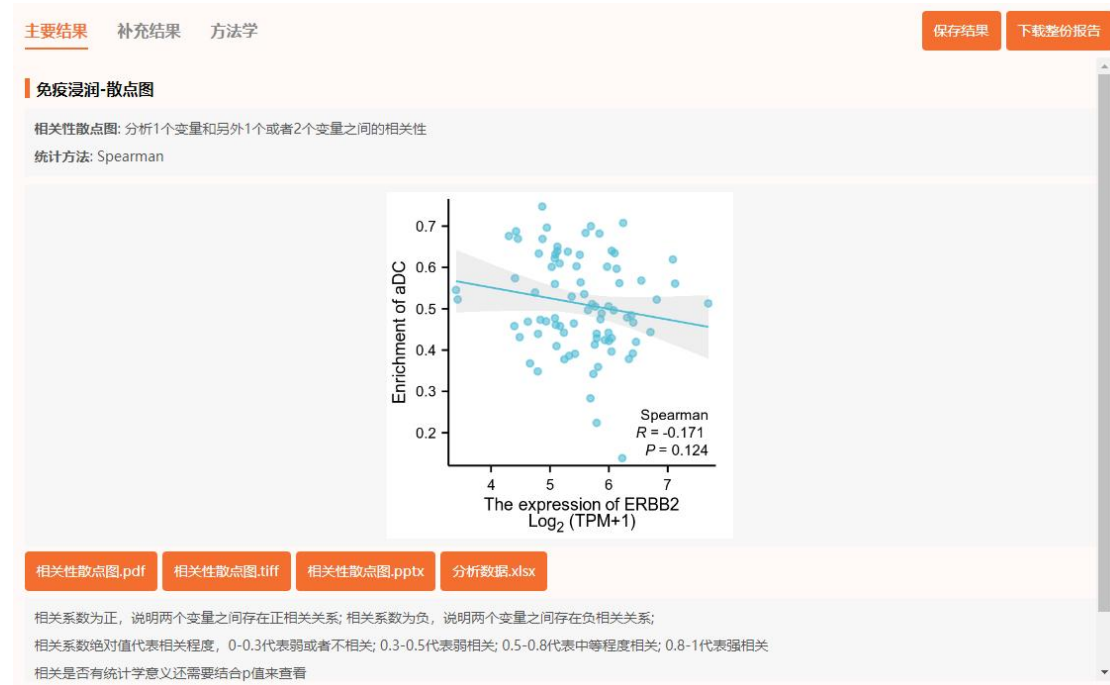
图片



- 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



主要结果格式为图片格式, 提供 PDF、TIFF、PPTX 格式下载, 结果报告可以下载包括 pdf 以及说明文本的内容。

- 此外, 还提供公共数据中分子在不同样本的表达量值及对应的免疫浸润数据, 提供 EXCEL 格式下载:

	A	B	C
1	sample_id	ERBB2	aDC
2	TCGA-IC-A6RF-01A-13R-A336-31	7.088179767	0.619236356
3	TCGA-IG-A3I8-01A-11R-A24K-31	5.326059103	0.38598375
4	TCGA-IG-A3QL-01A-11R-A24K-31	5.81655619	0.359050938
5	TCGA-IG-A3YA-01A-11R-A24K-31	5.080858024	0.621749382
6	TCGA-IG-A3YB-01A-11R-A36D-31	5.451385873	0.602773272
7	TCGA-IG-A3YC-01A-11R-A24K-31	5.302637096	0.638142372
8	TCGA-IG-A4P3-01A-11R-A260-31	6.244386671	0.707673335
9	TCGA-IG-A50L-01A-11R-A260-31	4.93491274	0.469485084
10	TCGA-IG-A51D-01A-11R-A36D-31	5.840435538	0.681957326
11	TCGA-IG-A5B8-01A-11R-A28J-31	3.431703181	0.522583507
12	TCGA-IG-A5S3-01A-11R-A28J-31	5.85764018	0.474614144
13	TCGA-IG-A625-01A-11R-A31P-31	5.405250406	0.464443361

补充结果

统计描述

各个组对应常见「统计描述指标」

组别	数目	最小值	最大值	中位数(Median)	四分位距(IQR)	下四分位	上四分位	均值(Mean)	标准差(SD)	标准误(SE)
ERBB2	82	3.4034	7.6902	5.5935	0.95265	5.0817	6.0343	5.5296	0.76868	0.084887
aDC	82	0.13856	0.74732	0.50106	0.17475	0.43326	0.60801	0.51188	0.11763	0.01299

统计描述.xlsx

此表格提供分子和免疫细胞统计描述的结果，提供 EXCEL 格式下载。

异常值分析

离群值 = $Q1(\text{下四分位}) - 1.5 * IQR(\text{四分位间距})$ 或者 $Q3(\text{上四分位}) + 1.5 * IQR(\text{四分位间距})$

异常值 = $Q1(\text{下四分位}) - 3.0 * IQR(\text{四分位间距})$ 或者 $Q3(\text{上四分位}) + 3.0 * IQR(\text{四分位间距})$

组别	离群值	异常值
ERBB2	3.43170318148885,...	
aDC	0.138556252326031	

各组离群值和异常值如上所示，如数据确认非人为记录错误，可不进行处理

此表格为分子和免疫细胞异常值情况表，可以判断数据是否存在异常值。

正态性检验(Shapiro-Wilk normality test)

组别	自由度(df)	统计量	p值
ERBB2	82	0.98771	0.6286
aDC	82	0.9756	0.1199

正态性检验结果显示，提供的变量均接近正态分布($P > 0.05$)，建议选择用 参数检验方法(Pearson)

此表格为分子和免疫细胞正态性检验的结果。

相关性分析

同时提供Pearson和Spearman统计方法，可以根据需要选择标注在图中的方法

方法	组别I	组别J	自由度(df)	统计量	相关系数	置信区间(95%CI)	p值
Pearson	ERBB2	aDC	82	-1.5296	-0.16857	-0.37197 - 0.050275	0.1301
Spearman	ERBB2	aDC	82	1.076e+05	-0.17132		0.1237

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

此表格提供 Pearson 和 Spearman 统计方法的分析结果，即分子和免疫细胞的相关性分析结果。可以根据需要选择及修改标注在图中的方法。



方法学

所有分析和可视化均在 R 4.2.1 中进行

涉及的 R 包：ggplot2 包（用于可视化）

处理过程：分析数据变量之间相关性后，用 ggplot2 可视化结果。



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 可以不需要满足正态性，可以先选择非参数的 Spearman 相关进行尝试。

2. 怎么判断相关性好坏（结果好坏）？

答：

这个没有很统一的标准，可以是相关系数在 0.3、0.5、0.7 以上（不一定说要非常相关(>0.7)才能放文章里面），也可以是 p 值，具体根据研究情况进行判断。

➤ 相关系数强弱，一般可以认为：

- 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：相对弱或者不相关

3. ssGSEA 算法（其他一些算法）给到的结果跟 其他方法（或者别的数据库 TIMER 等）趋势不一样，这个是什么原因？

答：

不同算法之间可能是会存在有一定的差别，况且算法只是一种推测手段，实际是什么情况还是需要通过做实验来确定的。所以，如果只是单纯想要拿一些结果来充实自己的研究，那么可以只放满足自己想要的趋势的数据。

4. 免疫浸润可以做什么实验验证？

答：

可以通过免疫组化检测对应的免疫细胞的 markers，也可以对组织做流式分析分析细胞的情况等等。具体要根据研究情况进行安排。

5. 为什么“算法选项卡”中的细胞选项的内容对不上？

答：

如果更换了算法的参数，需要重新点击“确认”按钮后，选项卡中的【细胞】选项才会变成对应算法的选项。

6. 结果（ssGSEA）里面的有一些英文简写的细胞具体是哪些细胞？

答：

可以看方法学中对应的 Immunity 的引文：Bindea, Gabriela, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 39.4 (2013): 782-795. 本模块 ssGSEA 算法基于这篇引文（补充材料）提供的 24 种细胞 markers 计算这 24 种细胞的得分情况。如果想要知道具体某些细胞是什么，可以先查阅上面那篇引文。

7. 在云端数据框内看到的例数和分析时候的例数不同，这个是什么情况？

答：

云端数据的例数一般是对应组学所有的例数，分析时候可能会有剔除样本的情况，本模块使用 去除正常+去除无临床信息 的样本，具体需要看说明文本中对于数据的处理情况的说明。

8. 云端数据在哪可以查询?

答:

模块分析后，在方法学标签中，提供了公共数据（云端数据）的具体信息及下载链接。

