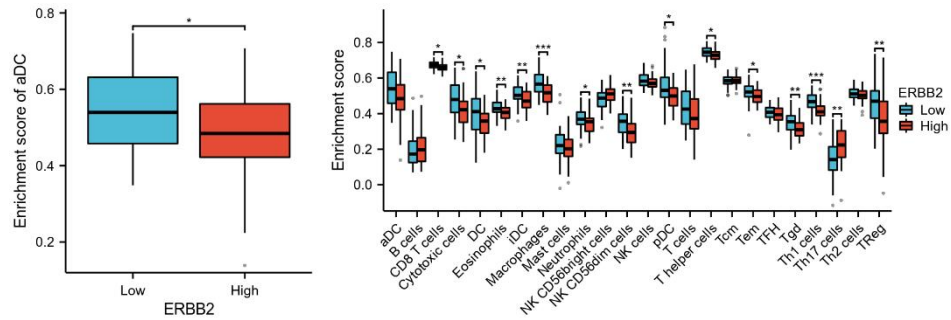


交互网络 - [免疫浸润-云] 单基因高低组差异



网址: <https://www.xiantao.love>



更新时间: 2023.03.09

目录

基本概念	4
应用场景	5
分析流程	5
主要结果	6
云端数据	10
参数说明	11
特殊参数	11
算法	12
统计分析	13
间距设置	14
点	15
箱/柱	16
小提琴	17
误差线	18
标题	19
图注(Legend)	19
坐标轴	20
风格	21
图片	22
结果说明	23
主要结果	23
补充结果	24
方法学	26
如何引用	27
常见问题	28



基本概念

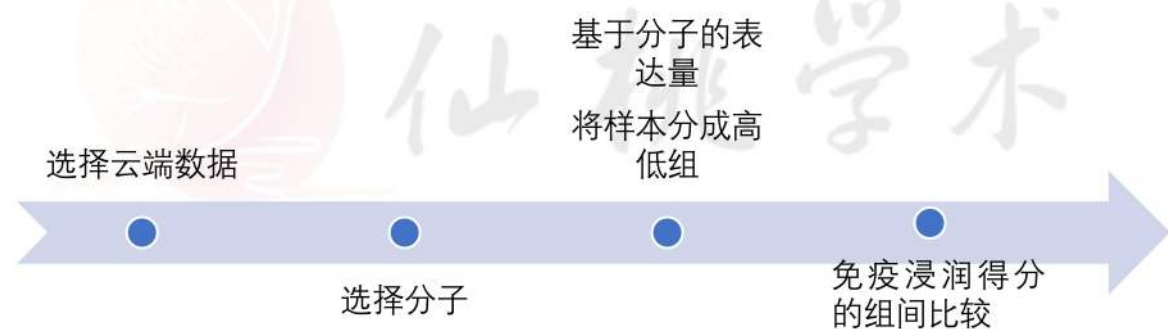
- 免疫浸润分析：利用转录组或者其他组学的数据，通过算法计算组织中免疫细胞的分数情况，推测组织中免疫细胞的构成情况。
- 免疫浸润算法
 - **ssGSEA (single sample GSEA)**：用给定的基因集对每一个样本计算对应基因集的富集得分。在免疫浸润中的 ssGSEA 则是利用每一类免疫细胞特定的 markers（来源于说明文本中的引文）作为基因集去计算每个样本中每一类免疫细胞的富集得分，推断单个样本的免疫细胞的浸润情况。
 - **estimate**：estimate 包提供的通过内置的 markers 计算样本的免疫浸润得分、基质得分和估计得分。（主要看基质得分和免疫浸润得分，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）
- 单基因-高度组比较：通过在同一个疾病状态中，对大量样本按照某个基因的表达值分成高低表达组（模拟类似过表达或者敲低这个基因的效果），进行两组间的比较分析。
- 统计方法：
 - **T test**，亦称 student t 检验 (Student's t test)，主要用于两组之间的比较，两组需要满足 正态性 和 方差齐性 的要求。
 - **Welch` t test**，又称不等方差检验，即当两组仅满足正态而不满足方差齐性的要求时，可以选择用该方法进行两组的比较。
 - **Wilcoxon rank sum test**，也叫 Mann-Whitney U test（曼-惠特尼 U 检验），或者 Wilcoxon-Mann-Whitney test。秩和检验是一个非参的假设检验方法，一般用于两组不满足正态性的情况。

应用场景

单个分子高低表达组比较，是基于公共数据（**云端数据**）根据所选单个分子在所有样本中的**中位数**，将样本分为高低表达组，然后直接在预先计算好的每个样本的免疫细胞分数中，进行高低分组间的比较分析，分析其是否有统计学差异。

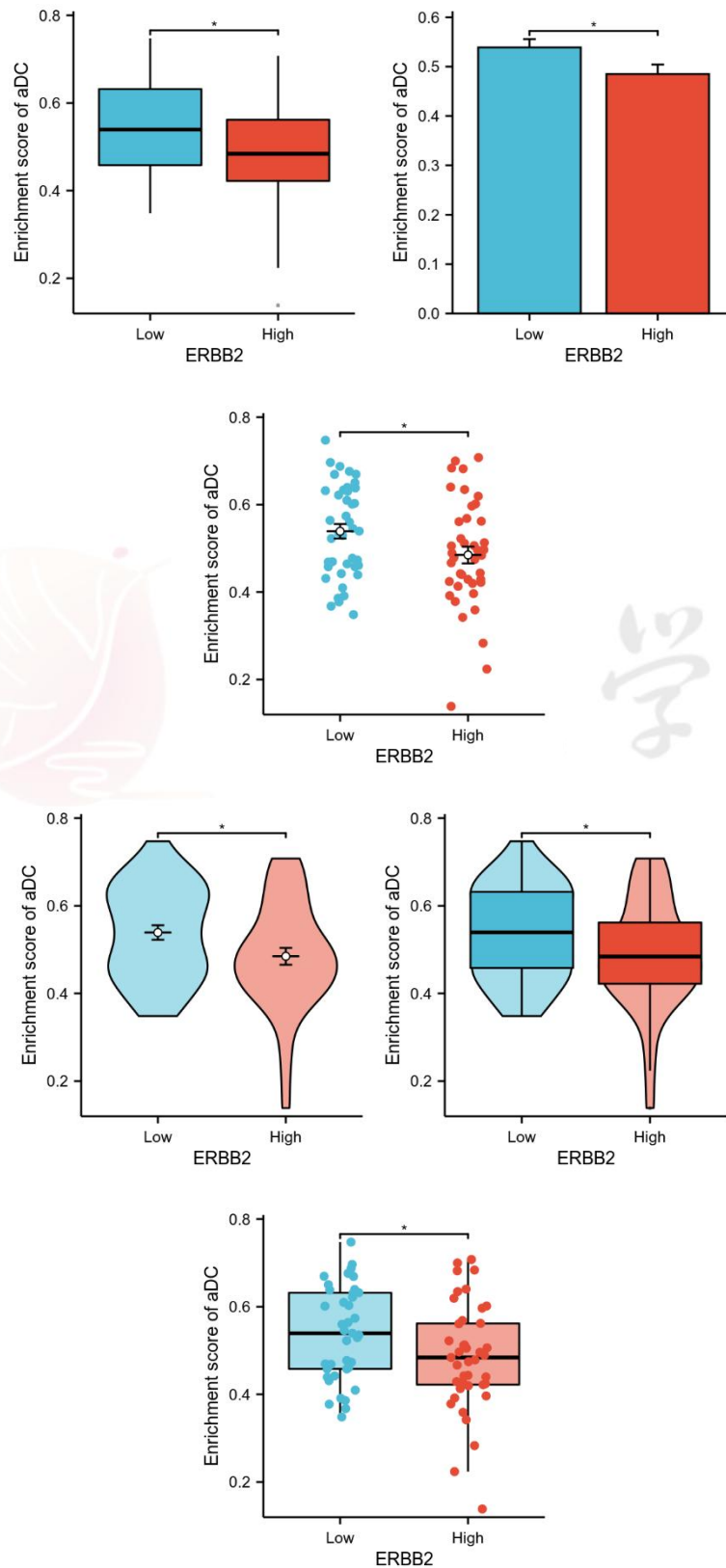
一般绘制箱式图进行直观比较，本模块支持点图、箱式图、柱状图、小提琴图及各自组合的可视化形式。

分析流程



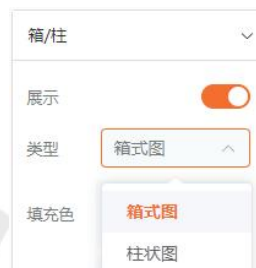
主要结果

单个免疫细胞类型的可视化



可视化形式：默认-箱式图

- 横坐标，分子名称及高低比较组信息，根据分子表达量的中位数所分成的高低表达组。
- 纵坐标，免疫浸润数据分布，算法为 `ssgsea` 时，代表各样本中不同免疫细胞类型的富集得分，算法为 `estimate` 时，代表各样本中整体的免疫得分（可选 免疫浸润得分、基质得分和估计得分）。
- 默认情况下，模块会根据数据的情况，如正态性和方差齐性自动选择合适的统计方法进行统计分析（具体方法见基本概念中的统计方法）。
- 可视化形式：需要选择对应参数中【展示】



■ 箱式图/柱状图

- ◆ 箱式图：常见分组比较图之一，箱子中间的横向代表中位数，箱子的上下边代表上四分位（75 百分位数）和下四分位（25 百分位数）。一般而言，箱子的上方和下方的线，如果分组内不存在离群值 ($Q1-1.5*IQR$ or $Q3+1.5*IQR$, 下四分位-1.5 倍四分位距), 那么线的最远位置就为最小值或者最大值。箱子的上方或者下方的点代表离群值的点。
- ◆ 柱状图：常见分组比较图之一，柱状图高度一般代表每组的均值情况，同时附带有误差线，表征组内变异的程度。



- 点图：将分组内所有的值用点的位置来进行表示，同时还会另外加上误差线以表征组内的变异情况。点图能够直接看到分组内各样本的分组情况。

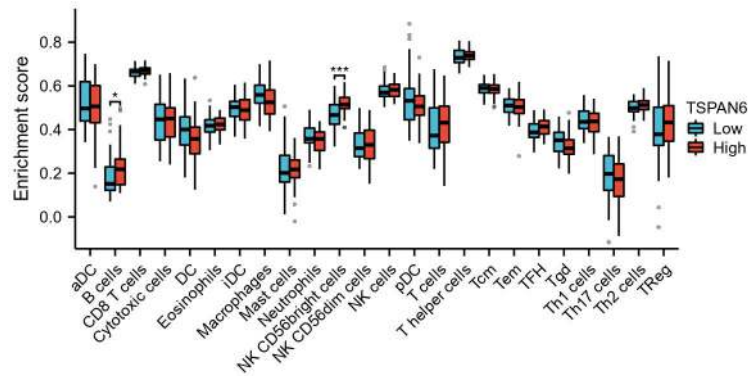


- 小提琴图：形状类似小提琴，同一水平线上分布的样本越多，则越宽，否则就越窄。小提琴图能有效展示分组内的样本情况的分布。



组合图：点图和箱式图

（多选）多个免疫细胞类型的可视化



上图为，免疫浸润算法-ssgsea 时，选择多个免疫细胞类型的可视化形式。

- 横坐标为免疫细胞类型，纵坐标为免疫浸润得分。
- 箱子（或其他形式）代表单基因高低表达组，如上图按照基因"TSPAN6"在云端数据集中的表达量将样本分为 Low、High 两组。

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
v

数据过滤:
去除正常
v

数据格式:
log2(value+1)
v

	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



- 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**：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。可多选。

统计分析

- **统计方法**：统计方法默认为 auto（自动选择），当点击确认进行分析后，会自动替换成适合于对应公共数据的统计方法，之后可以自行选择和修改别的统计方法。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。
- **分组比较**：统计学差异标注的分组，默认为 all（全部都标注）。当点击确认进行分析后，会自动替换成对应数据的分组。之后可以自行选择想要保留和去掉的比较。（如果分组不满足>3 个观测以及标准差>0 的情况，则可能不会出现在此处。）。
- **显著性显示类型**：影响分组比较中显著性标注，默认为星号。可选择星号或者 p 值以及其他形式，可以选 星号、p 值科学计数法、p 值数值(小于 0.05

自动<)、p 值数值(小于 0.001 自动<)、p = 科学计数、p = 数值(小于 0.05 自动<)、p = 数值(小于 0.001 自动<)、无。



➤ 显著性大小：可以修改显著性标注的大小。

参数使用情况：

补充说明：

· 统计方法: T test

· 主分子: ERBB2[ENSG00000141736.14]

· 免疫浸润算法: 基于R包-GSVA[1.46.0](Hänzelmann et al., 2013)中提供的ssGSEA算法, 利用Immunity文章(Bindea, Gabriela, et al., 2013)提供的24种免疫细胞的markers来计算对应云端数据的免疫浸润情况, 具体24种免疫细胞可查看对应参考文献

· 免疫细胞: aDC[activated DC]

间距设置

间距设置 ✓

(二维)组内总宽度 0.8

➤ 组间距离：两组之间的宽度，只有在二维数据(含 legend)的时候才会有效果。主要控制单个分子两组之间的距离。

点



点

展示 ☐

填充色 ☐ ☐

描边色 ☐ ☐

样式 圆形 ×

大小 1

不透明度 1

分布宽度 0.6

- 展示：可选是否展示。可组合图形。
- 填充色：点的填充色颜色选项，有多少个分组会提取多少个颜色，第一色卡控制低表达分组，第二色控制高表达分组，最多支持修改 2 个颜色。受配色方案全局性修改。
- 描边色：点的描边色颜色选项，有多少个分组会提取多少个颜色，第一色卡控制低表达分组，第二色控制高表达分组，最多支持修改 2 个颜色。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。可以多选，多选后不同的分组中点的类型也会有不同。
- 大小：点的大小。
- 透明度：点的透明度。0 为完全透明，1 为完全不透明。
- 分布宽度：图中的点会在一个水平线上随机分布，此处影响点能随机水平移动的范围。

箱/柱



- 展示：可选是否展示。可组合图形。
- 填充色：箱子的填充色颜色选项，有多少个分组会提取多少个颜色，第一色卡控制低表达分组，第二色控制高表达分组，最多支持修改 2 个颜色。受配色方案全局性修改。
- 描边色：箱子的描边色颜色选项，有多少个分组会提取多少个颜色，默认黑色，最多支持修改 2 个颜色。不受配色方案全局性影响。
- 描边粗细：箱子描边的粗细，默认为 0.75pt。
- 不透明度：箱子的透明度。0 为完全透明，1 为完全不透明
- 宽度：箱子的宽度控制，默认 0.6。

小提琴

小提琴

展示

填充色

描边色

描边粗细

不透明度

宽度

宽度校正

0.75pt

0.5

0.8

1

- 展示：可选是否展示。可组合图形。
- **填充色**：小提琴的填充色颜色选项，有多少个分组会提取多少个颜色，第一色卡控制低表达分组，第二色控制高表达分组，最多支持修改 2 个颜色。受配色方案全局性修改。
- **描边色**：小提琴的描边色颜色选项，有多少个分组会提取多少个颜色，默认黑色，最多支持修改 2 个颜色。不受配色方案全局性影响。
- 描边粗细：小提琴描边的粗细，默认为 0.75pt
- 不透明度：小提琴的透明度。0 为完全透明，1 为完全不透明。
- 宽度：小提琴的宽度。
- 宽度校正：用于提高小提琴中较窄位置的宽度和整体宽度。

误差线

误差线只有在**在没有箱式图时才会显示**（箱式图本身自带类似误差线）。

- 展示：可选是否展示。
- 样式：可选 上、上下。
- 类型：可选均值±标准差、均值±标准误、中位数~上下四分位，建议选择均值±标准差。
- 颜色：误差线颜色，默认为纯黑，不受配色方案全局性影响。
- 描边粗细：误差线粗细，默认为 0.75pt
- 宽度：误差线的宽度。

标题

标题 ▼

大标题

大标题内容

x轴标题

x轴标题内容

y轴标题

y轴标题内容

- 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 $[2]$

图注(Legend)

图注 ▼

是否展示

☒

图注标题

图注标题内容

图注位置

默认 ▼

- 展示：是否展示图注
- 图注标题：可以添加图注标题

- 图注位置：可选 默认、右、上，默认为右。

坐标轴

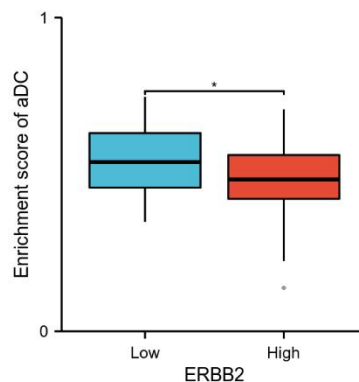
坐标轴

x轴分组名 ,+空格隔开

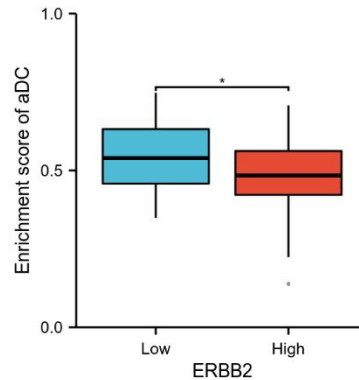
x轴标注旋
转 0

y轴范围+刻度 ()包裹,内容用','+

- **X 轴分组名**：支持直接修改 x 轴各个分组的名字，每个名字之间需要用英文输入法的逗号隔开，比如 group1,group2。这里支持换行，需要换行的位置可以插入\n
- **X 轴标注旋转**：支持对 x 轴文字进行旋转。适合于 x 轴文字过长的时候。
- **Y 轴范围+刻度**：（注意：范围的修改如果调整过大会失效）
 - 如果只是想要修改范围，可以只输入两个范围值，比如 0,0,1,1



- 如果同时想要修改范围+刻度，可以输入比如：0,0,0.5,1,1 。注意，此时最大和最小值会被当做范围值，不会作为刻度，如果需要刻度，需要类似于 0,1 那样同时写两次



风格

风格

边框

网格

xy颠倒

文字大小 7pt

- 外框：是否添加外框
- 网格：是否添加网格
- 是否颠倒 XY 轴：可以颠倒 xy 轴
- 文字大小：针对图中所有文字整体的大小控制

图片

图片	▼
宽度 (cm)	5
高度 (cm)	5
字体	Arial ▼

➤ 宽度：图片横向长度，单位为 cm

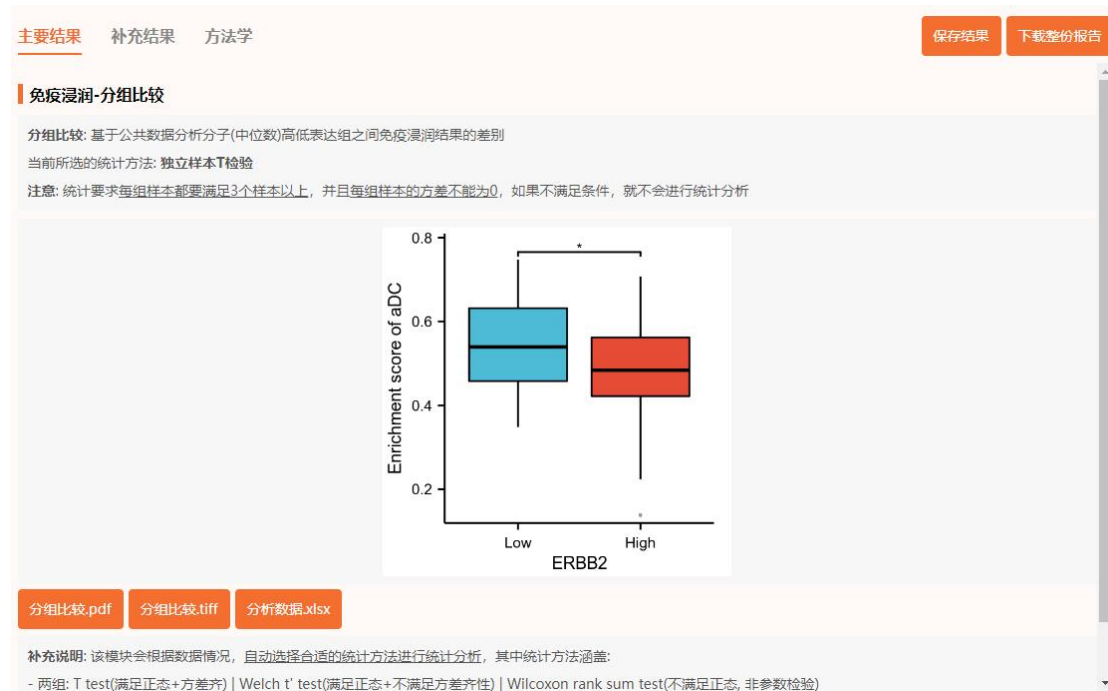
➤ 高度：图片纵向长度，单位为 cm

字体：可以选择图片中文字的字体



结果说明

主要结果



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

- 如果数据可以进行统计分析, 将会进行统计分析。统计分析默认是根据数据情况选择合适的统计方法。统计要求每组样本都要满足 3 个样本以上, 并且每组样本的方差不能为 0, 如果不满足条件, 就不会进行统计分析。
- 此外, 还提供公共数据中分子在不同样本及分组的免疫浸润数据, 提供 EXCEL 格式下载:

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

补充结果

统计描述

各个组常见「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)
Low	41	0.34841	0.74732	0.53936	0.17384	0.45803	0.63186	0.53897	0.10618	0.016582
High	41	0.13856	0.70767	0.48417	0.13985	0.42208	0.56193	0.48479	0.12346	0.019281

统计描述.xlsx

此表格提供统计描述的结果，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
High	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

此表格异常值情况表，可以判断数据是否存在异常值。

正态性检验

检验方法: Shapiro-Wilk normality test

组别	自由度(df)	统计量	p值
High	41	0.96925	0.3258
Low	41	0.96012	0.1588

正态性检验结果显示，观测变量在各组内接近正态分布($P > 0.05$)，建议选择用 参数检验的方法

此表格为正态性检验的结果。

方差齐性检验

检验方法: Levene's test

· Base on Mean

自由度1(df1)	自由度2(df2)	统计量	p值
1	80	0.010904	0.9171

方差齐性检验显示，各组观测变量的方差相等($P > 0.05$)

此表格为方差齐性检验的结果。

独立样本T检验

应用条件: 两组独立数据, 满足正态性检验和方差齐性检验

组别I	组别J	自由度(df)	统计量t	差值(J-I)	置信区间(95%CI)	p值
Low	High	80	-2.1305	-0.05418	-0.10479 - -0.0035719	0.0362

p值满足 <0.05 时，可认为两组存在统计学上差异

此表格为 2 组比较统计检验的结果。

(注意: 不同的统计方法会有不一样的统计检验的表格)



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)、stats、car (用于统计分析)

处理过程: 对主变量进行分组后, 根据数据格式特征情况选择合适的统计方法进行统计(stats 包以及 car 包)([如果不满足统计要求将不会进行统计分析](#)), 用 ggplot2 包对数据进行可视化。



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. 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。

2. 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。

3. 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。

4. 统计学标注可以用具体 p 值吗？

答：

在“统计分析”选项卡中，【显著性显示类型】参数，里面有显示具体 p 值的选项。另外，需要【分组比对】选择了分组才会显示。

5. 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。

6. 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。

