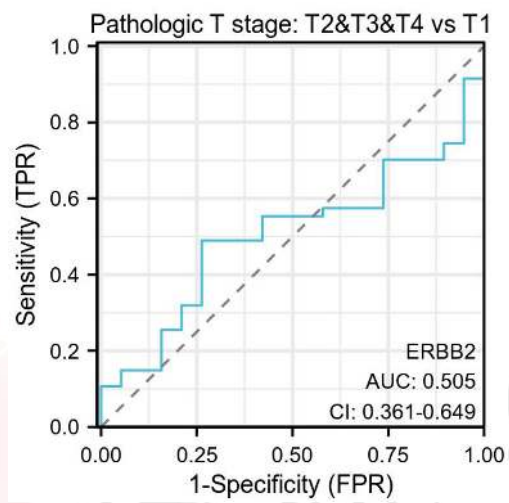


临床意义 - 诊断 ROC 曲线[云]



网址: <https://www.xiantao.love>



更新时间: 2023.03.09

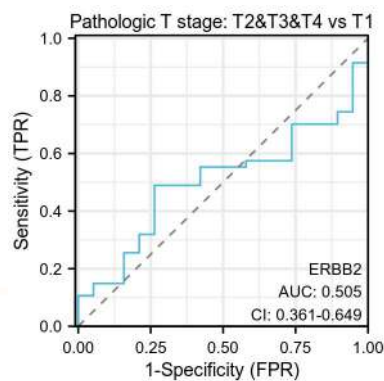
目录

基本概念	3
应用场景	3
分析流程	4
结果解读	5
数据格式	6
参数说明	7
特殊参数	7
分子	7
临床变量	8
主要参数	9
数据处理	9
统计	9
线	10
点	11
曲线下面积	12
标题	12
图注	13
风格	13
图片	14
结果说明	15
主要结果	15
补充结果	16
方法学	17
如何引用	18
常见问题	19

基本概念

诊断 ROC 曲线：受试者工作特征曲线（Receiver Operating Characteristic Curve, ROC 曲线）和 ROC 曲线下的面积（Area Under ROC Curve, AUC）常用于诊断试验的评估，评估预测准确率情况。例如一组数据的结局为 group1 和 group2，变量为 a、b 和 c，也就是评估 a、b 和 c 在预测 group1 和 group2 上的结局，哪个的准确性更高。ROC 曲线图是反映敏感性与特异性之间关系的曲线。AUC 取值范围一般在 0.5 和 1 之间，使用 AUC 值作为评价标准是因为很多时候 ROC 曲线并不能清晰的说明哪个分类器的效果更好，而作为一个数值，对应 AUC 更大的分类器效果更好。

➤ 图形构成



应用场景

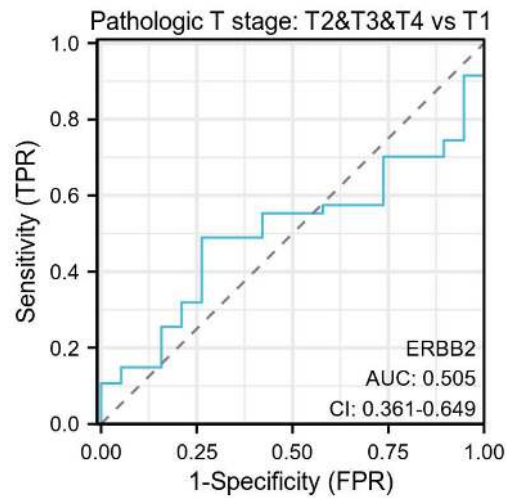
多应用在医学领域，判断某种因素对于某种疾病的诊断是否有诊断价值。

分析流程

云端数据 → ROC 相关分析 → ROC 曲线可视化

- 云端数据：不同平台的云端数据集的分子可能会有不同!
 - 通过特殊参数[分子]选择云端数据中需要进行分析以及可视化的预测变量/分子
 - 通过特殊参数[临床变量&二分类结局分组]选择预测结局，用于分析云端数据中所选择预测变量在改预测结局中的情况
- ROC 相关分析：
 - 统计描述：在不同的结局下，统计各预测变量/分子的生存情况
 - AUC 结果：不同的结局下，计算各预测变量/分子 ROC 曲线下面积(AUC)
 - ROC 信息：计算各预测变量/分子的最佳阈值（cut-off 值）、敏感度、特异度等
 -
- 将分析过程中得到的敏感度、特异度数据进行转换，得到可视化需要的数据，并进行可视化
 - 敏感度 = 真阳性率(TPR)
 - 特异度 = $1 - \text{假阳性率(FPR)} / \text{特异性}$

结果解读



诊断 ROC 曲线

- 横坐标 X 轴为 $1 - \text{特异性}$ ，也称为假阳性率，X 轴越接近零准确率越高；纵坐标 Y 轴称为敏感度，也称为真阳性率，Y 轴越大代表准确率越好。
- AUC (Area Under Curve, AUC)，ROC 曲线下的面积，常用于诊断试验的评估，AUC 取值范围一般在 0.5 和 1 之间，AUC 越接近于 1，说明该变量在预测结局上诊断效果越好。

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



参数说明

(说明：标注了颜色的为常用参数。)

特殊参数

分子

特殊参数

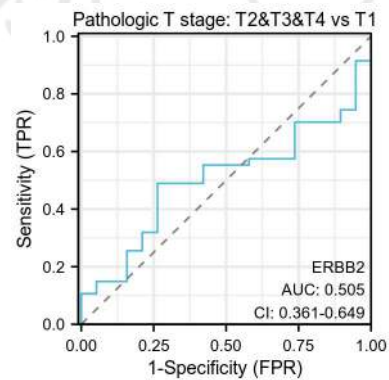
分子 ?

ERBB2[ENSG00000141736.14] ×

- 分子：可以输入需要进行诊断 ROC 相关分析的变量/分子/基因，可以输入关键词进行搜索，如下：

分子 ?

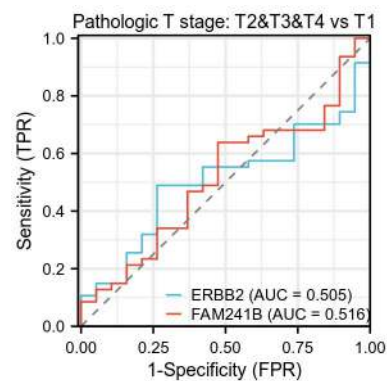
ERBB2[ENSG00000141736.14] ×



分子 ?

ERBB2[ENSG00000141736.14] ×
ERBB|

ERBB4[ENSG00000178568.15]
ERBB3[ENSG00000065361.16]
RBBP5[ENSG00000117222.14]



临床变量

临床变量 ⓘ

二分类结局分组(一个框内的分类组合成一个分类)

(临床)Pathologic_T_stage

T1 [21] ×

T2 [10] × T3 [38] ×
T4 [1] ×

− +

- 临床变量：可以选择清洗好的临床变量
- 二分类结局分组：通过选择清洗好的临床变量以及对应进行分组，用于分析云端数据中所选择预测变量在改预测结局中的情况



主要参数

数据处理

数据处理

缺失值处理

不处理变量缺失

缺失值处理：默认是不处理变量缺失值，可以选择去除任一变量缺失的样本

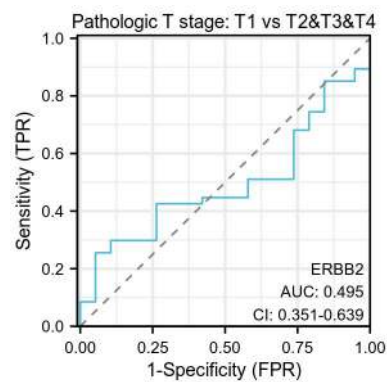
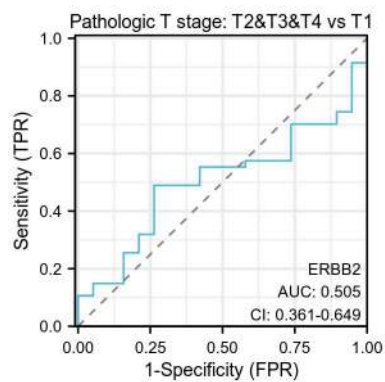
统计

统计

方向

自动

- 方向：可以设定二分类结局对应比较的方向(影响参考和实验组的设定)，默认自动（表示根据数据特点来），还可以自己选择正向或者反向，如下：左侧为正向，右侧为反向



线



- 颜色：可以修改每条曲线的颜色
- 样式：可以修改每条曲线的样式(线条类型)，默认是实线，也可以选择虚线
- 粗细：可以修改每条曲线的线条粗细，默认是 0.75pt
- 不透明度：可以修改每条曲线的不透明度，默认是 1，0 是完全透明，1 是完全不透明

点

点

展示

填充色

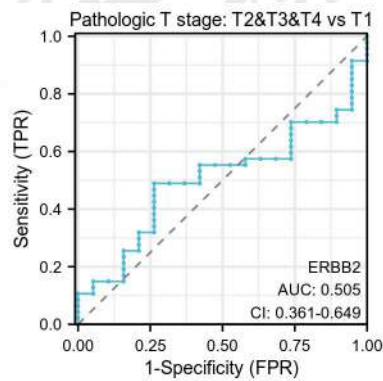
描边色

样式
圆形

大小
0.3

不透明度
1

- 展示：可以选择是否展示曲线上的点，默认不展示，还可以选择展示，如下：



- 填充色：可以修改点的填充色
- 描边色：可以修改点的描边色
- 样式：点的样式，可以选择圆形、三角形等形状选择
- 大小：点的大小，默认 0.3
- 不透明度：点的不透明度，默认是 1，0 是完全透明，1 是完全不透明

曲线下面积



- 是否展示：是否展示出每个变量曲线下的面积
- 不透明度：如果展示曲线下面积，可以设定面积的不透明度

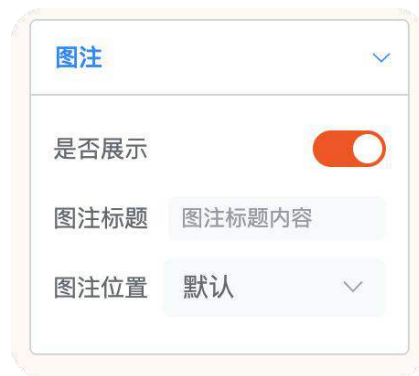
标题



- 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

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：图注内容是否展示
- 图注标题：可以填入图注标题
- 图注位置：默认是右下，还可以选右

风格



风格

边框 ☒

网格 ☒

文字大小 7pt

- 边框：是否在图中添加边框
- 网格：是否在图中添加网格线
- 文字大小：图中的文字部分的大小（包括标签文字和刻度数），默认是 7pt

图片



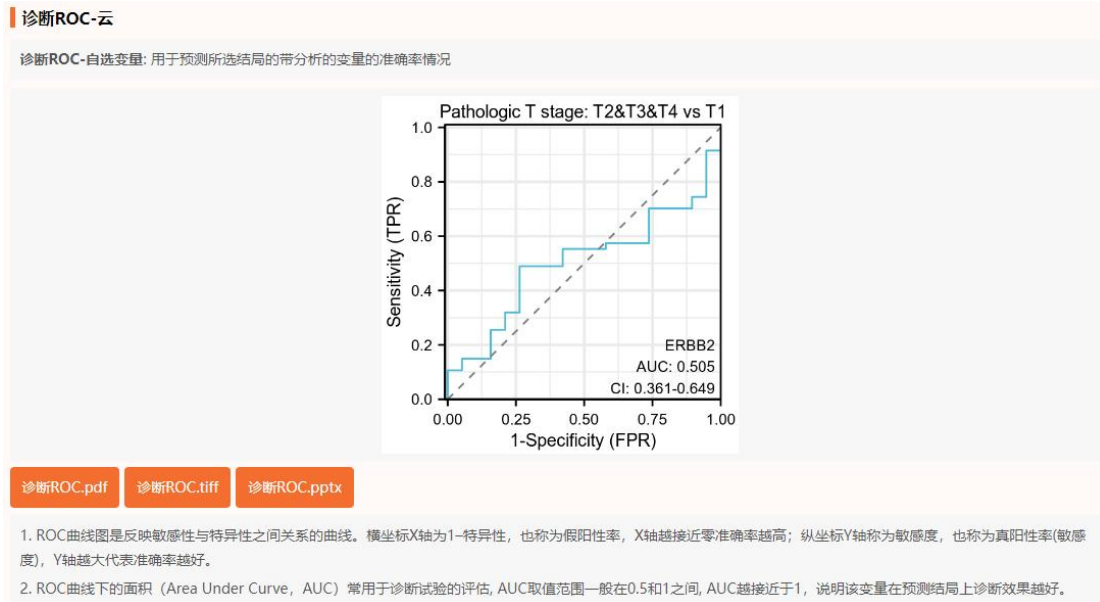
图片	
宽度 (cm)	5
高度 (cm)	5
字体	Arial

- 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：可以选择图中文本内容字体



结果说明

主要结果



主要结果格式为图片格式，提供 PDF、TIFF、PPT 格式下载。

补充结果

1. 统计描述表：上传数据的一些基本情况

统计描述

各个组常见「统计描述指标」

结局	变量	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)
T1	ERBB2	19	4.772	11.013	6.5099	0.69379	6.1748	6.8686	6.964	1.5472
T2&T3&T4	ERBB2	47	2.8556	12.878	6.6882	2.0376	5.7704	7.808	7.0713	2.2452

2. AUC 结果表

AUC结果表

预测变量	预测结局	曲线下面积(AUC)	置信区间(CI)
ERBB2	T2&T3&T4 vs T1	0.505	0.361 - 0.649

预测结局中, vs后面的结局事件为参考组(影响真/假阳性和真/假阴性的区分)(如果统计-方向参数选择的是“自动”,则会对结局的方向会进行调整保证曲线都是往上凸(pROC包提供))

在AUC > 0.5的情况下, AUC越接近于1, 说明该变量在预测结局上诊断效果越好。

AUC在0.5 ~ 0.7时有较低准确性, AUC在0.7 ~ 0.9时有一定准确性, AUC在0.9以上时有较高准确性。

AUC = 0.5时, 说明该变量不起作用, 无诊断价值。

3. ROC 信息表

ROC信息表

预测变量	cut-off值	灵敏度	特异度	准确率	真阳个数	真阴个数	假阳个数	假阴个数	阳性预测值	阴性预测值	约登
ERBB2	6.808	0.48936	0.73684	0.56061	23	14	5	24	0.82143	0.36842	0.18978

各预测变量在各自最佳cut-off值下部分ROC相关信息和数据。

4. 当特殊参数[分子]选择的预测变量大于 1 个时, 会提供预测变量间的 AUC 检验表：对变量 AUC 进行检验

AUC检验表

变量1	变量2	统计量	p值	检验方法	趋势方向
ERBB2	FAM241B	-0.10236	0.9185	DeLong's test	不一致

1. 变量在AUC > 0.5的情况下, AUC越接近于1, 该变量在预测结局上诊断效果越好,具体AUC值参考上面AUC结果表

2. DeLong's test检验结果中若P值小于0.05, 则表明两个变量在预测结局上诊断效果具有统计学意义, 若大于0.05,则表明两个变量在预测结局上诊断效果没有统计学意义。

3. 两个变量如果趋势方向(direction)不一致, 用DeLong方法可能结果会不准确

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: pROC[1.18.0] 用于 ROC 分析和 ROC 检验

分析过程:

- (1) 使用 pROC 包进行对云端数据进行 ROC 分析
- (2) 分析结果用 ggplot2 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. AUC 会出现 < 0.5 的情况吗?

答：一般情况下，pROC 分析结果中 AUC 面积是在 0.5-1 之间。

2. 1 个组的时候为什么没有给出统计学检验的 p 值?

答：

ROC 一般是看 AUC 的大小的,只有当存在有多个曲线的时候才会进行检验比较。如果只有 1 条曲线,是没办法进行统计检验的,除非是跟 0.05 的对角线比,这种比较其实是没有意义的,这种只要 AUC 的下限没有跨过 0.5,那么这个曲线肯定是有意义的,所以单个曲线是没有统计学比较的意义。

3. 这里能做联合指标的 ROC 分析吗?

答：

如果上传的数据是 logistic 模型的 predict 值,这个值就可以用来代表联合指标。在 logistic 回归模块或者是诊断列线图模块分析后均会有提供预测值.xlsx 文件下载,从这个里面就可以获取到联合指标的 predict 值,用这个值做 ROC 就是联合指标 ROC 分析。

4. 数据的结局是以哪个作为阴性（参考）？哪个作为阳性（实验）？

答：

默认上传数据的第一列（二分类）以第一个出现的分类组参考，后出现的分类作为实验。这个方向会影响最终的真阳、真阴、假阳、假阴个数。如果需要反过来，可以在<统计>-<方向>参数中进行修改。

