

临床意义 - 预后筛选[云]

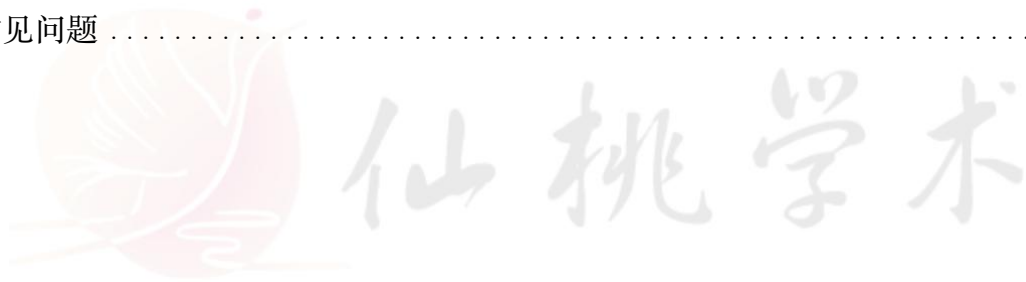
id	统计量	HR	pvalue	gene_name	gene_type
ENSG00000001036.14	3.0654	1.5594	0.0838	FUCA2	protein_coding
ENSG00000001084.13	0.092558	1.0799	0.7606	GCLC	protein_coding
ENSG00000001167.14	0.12821	0.91487	0.7203	NFYA	protein_coding
ENSG00000001460.18	0.00069118	0.99345	0.9790	STPG1	protein_coding
ENSG00000001461.17	0.582	0.82697	0.4463	NIPAL3	protein_coding
ENSG00000001497.18	1.8839	1.4113	0.1709	LAS1L	protein_coding
ENSG00000001561.7	1.3809	1.3441	0.2437	ENPP4	protein_coding
ENSG00000001617.12	0.72878	0.80835	0.3941	SEMA3F	protein_coding
ENSG00000001626.16	4.3623	1.7061	0.0407	CFTR	protein_coding
ENSG00000001629.10	3.2999	0.63465	0.0711	ANKIB1	protein_coding
ENSG00000001630.17	2.4748	0.67553	0.1178	CYP51A1	protein_coding
ENSG00000001631.16	0.014051	0.9708	0.9056	KRIT1	protein_coding
ENSG00000002016.18	0.10402	0.9224	0.7474	RAD52	protein_coding

网址: <https://www.xiantao love>

更新时间: 2023.03.07

目录

基本概念	3
应用场景	4
主要结果/结果解读	5
数据格式：云端数据	7
参数说明	8
预后参数	8
分组	9
统计	10
结果说明	12
主要结果	12
方法学	13
如何引用	14
常见问题	15



基本概念

- **风险比 (Hazard Ratio, HR)**：两个风险率的比值。当 $HR > 1$ 时，说明研究对象是一个危险因素；当 $HR > 1$ 时，说明研究对象是一个危险因素；当 $HR < 1$ 时，说明研究对象是一个保护因素；当 $HR = 1$ 时，说明研究对象对生存时间不起作用
- 预后分析方法：
 - 非参数法：寿命表、Kaplan-Meier(乘积极限法 Product limit method)、Cox 回归
 - 参数法
- Overall Survival (OS)，总体生存期，指结局指标是死亡时间，这个死亡是任何原因导致的死亡都算进去，只关心是否死亡，不关心因为何种原因死亡
- Disease Free Survival (DFS)，无病生存期，指经过治疗后未发现肿瘤，结局指标为疾病复发或死亡，同样不需要关心死亡原因。这一指标是临床获益的重要反映，随访时间可以缩短，因为增加了疾病复发这一节点。没有复发或没有死亡同样可以反映临床获益。这里也涉及到无疾病复发的一个定义，因此在临床资料纳入上比较困难
- PFI 无进展间隔：progression-free interval，从初次治疗的随机分组日期到疾病复发时间。（具体可以参考对应的引文）

应用场景

- 此模块可以利用云端数据，自定义分子的分组，初步分析和筛选出与预后有一定关联的分子。
- 此模块可以作为筛选云端数据中的预后分子，能更加针对性筛选出有研究价值的分子

分析流程：

云端数据  预后筛选

- 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和临床变量可能会有不同
- 预后筛选

主要结果/结果解读

id	统计量	HR	pvalue	gene_name	gene_type
ENSG00000001036.14	3.0654	1.5594	0.0838	FUCA2	protein_coding
ENSG00000001084.13	0.092558	1.0799	0.7606	GCLC	protein_coding
ENSG00000001167.14	0.12821	0.91487	0.7203	NFYA	protein_coding
ENSG00000001460.18	0.00069118	0.99345	0.9790	STPG1	protein_coding
ENSG00000001461.17	0.582	0.82697	0.4463	NIPAL3	protein_coding
ENSG00000001497.18	1.8839	1.4113	0.1709	LAS1L	protein_coding
ENSG00000001561.7	1.3809	1.3441	0.2437	ENPP4	protein_coding
ENSG00000001617.12	0.72878	0.80835	0.3941	SEMA3F	protein_coding
ENSG00000001626.16	4.3623	1.7061	0.0407	CFTR	protein_coding
ENSG00000001629.10	3.2999	0.63465	0.0711	ANKIB1	protein_coding
ENSG00000001630.17	2.4748	0.67553	0.1178	CYP51A1	protein_coding
ENSG00000001631.16	0.014051	0.9708	0.9056	KRIT1	protein_coding
ENSG00000002016.18	0.10402	0.9224	0.7474	RAD52	protein_coding

9	ENSG00000001036.14	3.065426976	1.559401536	0.0837829	FUCA2	protein_coding
10	ENSG00000001084.13	0.092557944	1.079904061	0.760628506	GCLC	protein_coding
11	ENSG00000001167.14	0.12820672	0.914872947	0.72031514	NFYA	protein_coding
12	ENSG00000001460.18	0.000691179	0.993450087	0.979026712	STPG1	protein_coding
13	ENSG00000001461.17	0.582004349	0.826966474	0.446289566	NIPAL3	protein_coding
14	ENSG00000001497.18	1.883856242	1.411283584	0.170939314	LAS1L	protein_coding
15	ENSG00000001561.7	1.380874452	1.344070017	0.243742108	ENPP4	protein_coding
16	ENSG00000001617.12	0.728782314	0.808349565	0.394105236	SEMA3F	protein_coding
17	ENSG00000001626.16	4.362258854	1.706076146	0.040663128	CFTR	protein_coding
18	ENSG00000001629.10	3.299872093	0.634654416	0.071075306	ANKIB1	protein_coding
19	ENSG00000001630.17	2.474822869	0.675529352	0.117831935	CYP51A1	protein_coding
20	ENSG00000001631.16	0.01405123	0.970803464	0.905644003	KRIT1	protein_coding
21	ENSG00000002016.18	0.104022388	0.922395327	0.747364172	RAD52	protein_coding
22	ENSG00000002079.14	0.962077966	0.781281357	0.32933952	MYH16	ribosomal_protein
23	ENSG00000002330.14	1.78080351	0.715740902	0.185145164	BAD	protein_coding
24	ENSG00000002549.12	1.436053528	0.741426005	0.233128948	LAP3	protein_coding

- id: ensembl 库注释的分子 ID
- 统计量: 单因素 Cox 回归分析后得到的统计量
- HR: 单因素 Cox 回归后得到的风险比, 如果 $HR > 1$, 说明该分子可能是危险因素; 如果 $HR < 1$, 说明该分子可能是保护因素。如果为空, 则说明该分子在数据中存在有一半以上的样本表达值都为 0, 无法进行分组和分析
- pvalue: 单因素 Cox 回归分析后得到的 p 值
- gene_name: ensembl 库注释的分子名

- `gene_type`: ensembl 库注释的分子类型，其中包括了编码基因、lncRNA、miRNA 以及其他类型的分子



数据格式：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对应的分子可能会有不同。

数据参数

云端数据 ⓘ 食管癌 / TCGA / TCGA-ESCA / RNAseq / STAR / TPM @过滤:去除正常+去除无临床信息 @处理:log2(v...

这里为任务式模块，提交任务后需要到历史记录中刷新并等待任务完成，[\(分析时间大概在几分钟到十几分钟不等，具体要看对应的数据集的样本量。如果任务执行时间过长 \(>1h\)，刷新后任然在执行阶段，建议删除后重新提交\)](#)

参数说明

(说明：标注了颜色的为常用参数。)

预后参数



预后参数

预后类型 OS[Overall !]

- 预后类型：可选不同的预后类型。不同的数据集之间的预后类型可能不一样！
可以选择：

- OS[Overall Survival] (默认)：总体生存期
- DSS[Disease Specific Survival]：无病生存期
- PFI[Progress Free Interval]：无进展间隔

分组



分组

分组方法 0-50 vs 50-

- 分组方法：可以选择对于分子进行分组的方法，可以选择：
 - 0-50 vs 50-100 (默认, 表示中位数分组)
 - 0-33 vs 66-100
 - 0-25 vs 75-100
 - 0-20 vs 80-100
 - 数值纳入
- 当如果分子在一半的样本中的表达值都为 0，则无法进行分组和分析

统计

统计

统计方法 Cox回归

p值校正方法 none

- 统计方法：可以选择进行分子筛选的方法：默认为 Cox 回归，还可以选择 Logrank 检验，如下，上为 Cox 回归，下为 Logrank 检验

预后筛选-云

预后筛选: 针对数据批量进行预后分析

· 当前分组策略为: 0-50 vs 50-100; 当前分析方法: Cox回归

· 预后类型: OS[Overall Survival]

页面中仅仅展示20个结果, 更多的结果需要下载结果表格

id	统计量	HR	pvalue	gene_name	gene_type
ENSG00000000003.15	0.78722	1.2493	0.3743	TSPAN6	protein_coding
ENSG00000000005.6	0.010355	1.0258	0.9189	TNMD	protein_coding
ENSG000000000419.13	1.5431	1.3655	0.2145	DPM1	protein_coding
ENSG000000000457.14	2.9096	0.65355	0.0893	SCYL3	protein_coding
ENSG000000000460.17	0.25014	1.1325	0.6172	C1orf112	protein_coding
ENSG000000000938.13	0.84839	0.79278	0.3595	FGR	protein_coding
ENSG000000000971.16	1.8329	0.713	0.1781	CFH	protein_coding
ENSG000000001036.14	3.0654	1.5594	0.0838	FUCA2	protein_coding
ENSG000000001084.13	0.092558	1.0799	0.7606	GCLC	protein_coding
ENSG000000001167.14	0.12821	0.91487	0.7203	NFYA	protein_coding
ENSG000000001460.18	0.0069118	0.99345	0.9790	STPG1	protein_coding
ENSG000000001461.17	0.582	0.82697	0.4463	NIPAL3	protein_coding

预后筛选-云

预后筛选: 针对数据批量进行预后分析

· 当前分组策略为: 0-50 vs 50-100; 当前分析方法: Logrank检验

· 预后类型: OS[Overall Survival]

页面中仅仅展示20个结果, 更多的结果需要下载结果表格

id	统计量	HR	pvalue	gene_name	gene_type
ENSG00000000003.15	0.79257	1.2405	0.3733	TSPAN6	protein_coding
ENSG00000000005.6	0.010191	1.025	0.9196	TNMD	protein_coding
ENSG000000000419.13	1.555	1.3516	0.2124	DPM1	protein_coding
ENSG000000000457.14	2.9277	0.65962	0.0871	SCYL3	protein_coding
ENSG000000000460.17	0.24935	1.1294	0.6175	C1orf112	protein_coding
ENSG000000000938.13	0.84339	0.79879	0.3584	FGR	protein_coding
ENSG000000000971.16	1.8283	0.71862	0.1763	CFH	protein_coding
ENSG000000001036.14	3.0411	1.5349	0.0812	FUCA2	protein_coding
ENSG000000001084.13	0.092354	1.0769	0.7612	GCLC	protein_coding
ENSG000000001167.14	0.12891	0.91617	0.7196	NFYA	protein_coding
ENSG000000001460.18	0.00073733	0.99342	0.9783	STPG1	protein_coding
ENSG000000001461.17	0.58324	0.82982	0.4450	NIPAL3	protein_coding

- p 值矫正方法：可以选择进行筛选的时候对 p 值进行矫正的方法，默认为不进行矫正，还可以选择：BH、bonferroni、BY、fdr、holm、hochberg、hommel、none(不进行矫正)



结果说明

主要结果

预后筛选-云

预后筛选: 针对数据批量进行预后分析
 · 当前分组策略为: 0-50 vs 50-100; 当前分析方法: Cox回归
 · 预后类型: OS[Overall Survival]
 页面中仅仅展示20个结果, 更多的结果需要下载结果表格

id	统计量	HR	pvalue	gene_name	gene_type
ENSG00000000003.15	0.78722	1.2493	0.3743	TSPAN6	protein_coding
ENSG00000000005.6	0.010355	1.0258	0.9189	TNMD	protein_coding
ENSG000000000419.13	1.5431	1.3655	0.2145	DPM1	protein_coding
ENSG000000000457.14	2.9096	0.65355	0.0893	SCYL3	protein_coding
ENSG000000000460.17	0.25014	1.1325	0.6172	C1orf112	protein_coding
ENSG000000000938.13	0.84839	0.79278	0.3595	FGR	protein_coding
ENSG000000000971.16	1.8329	0.713	0.1781	CFH	protein_coding
ENSG00000001036.14	3.0654	1.5594	0.0838	FUCA2	protein_coding
ENSG00000001084.13	0.092558	1.0799	0.7606	GCLC	protein_coding
ENSG00000001167.14	0.12821	0.91487	0.7203	NFYA	protein_coding
ENSG00000001460.18	0.00069118	0.99345	0.9790	STPG1	protein_coding
ENSG00000001461.17	0.582	0.82697	0.4463	NIPAI3	protein_coding

确认

这里为任务式模块, 提交任务后需要到历史记录中刷新并等待任务完成, (分析时间大概在几分钟到十几分钟不等, 具体要看对应的数据集的样本量。如果任务执行时间过长 (>1h), 刷新后任然在执行阶段, 建议删除后重新提交)

- id: ensembl 库注释的分子 ID
- 统计量: 单因素 Cox 回归分析后得到的统计量
- HR: 单因素 Cox 回归后得到的风险比, 如果 $HR > 1$, 说明该分子可能是危险因素; 如果 $HR < 1$, 说明该分子可能是保护因素。如果为空, 则说明该分子在数据中存在有一半以上的样本表达值都为 0, 无法进行分组和分析
- pvalue: 单因素 Cox 回归分析后得到的 p 值
- gene_name: ensembl 库注释的分子名

- `gene_type`: ensembl 库注释的分子类型, 其中包括了编码基因、lncRNA、miRNA 以及其他类型的分子

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: survival、survminer 包

处理过程:

- (1) 数据部分跟所选的云端数据有关
- (2) 对云端数据进行预后筛选



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. 为什么结果表格中有一些分子是空值?

答： 筛选预后分子模块是逐个对分子参考设定的分组参数先进行分组，然后再进行 Cox 回归分析。如果该分子在一半的样本中的表达值为 0，则无法进行分组，也无法进行后续的 Cox 回归分析，所以在结果中对应的值都为空

