

## 交互网络 – [免疫浸润] ssGSEA 算法

Sample	aDC	B cells	CD8 T cells	Cytotoxic cells	DC	Eosinophils	iDC	Macrophages
TCGA-23-1120	0.4847	-0.19717	0.48277	0.13524	0.0008679	-0.018543	0.12789	0.35284
TCGA-29-1695	0.12032	-0.23266	0.39599	0.027392	-0.064047	-0.017161	0.23579	0.49721
TCGA-61-2003	0.3162	-0.25387	0.42556	0.052807	0.075874	0.021128	0.32906	0.44635
TCGA-13-1404	0.040349	-0.33064	0.44538	-0.056773	-0.14368	-0.041679	0.13768	0.3119
TCGA-13-1512	0.3261	-0.31194	0.45166	-0.020136	-0.14931	-0.043032	0.12891	0.37954
TCGA-23-1022	-0.35101	-0.17447	0.41995	-0.23292	0.017819	-0.023762	0.09339	0.16733
TCGA-61-2088	-0.32491	-0.42369	0.41759	-0.31333	-0.214	-0.094925	0.050971	0.18297
TCGA-13-1507	0.39549	-0.074744	0.46522	0.28181	0.026806	0.0027336	0.24579	0.41642
TCGA-25-1628	0.45145	-0.25196	0.48708	0.16751	-0.13522	0.0048725	0.26012	0.44543
TCGA-31-1944	-0.025706	-0.25496	0.42979	-0.021788	-0.028971	0.02634	0.12452	0.34283

网址: <https://www.xiantao love>

更新时间: 2023.12.05

## 目录

基本概念 .....	3
应用场景 .....	4
主要结果 .....	5
数据格式 .....	6
参数说明 .....	7
分析参数 .....	7
结果说明 .....	8
主要结果 .....	8
方法学 .....	9
如何引用 .....	10
常见问题 .....	11



## 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。
- ssGSEA 算法
  - ssGSEA (single sample GSEA) 免疫浸润算法，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。
  - 24 种免疫细胞: aDC[activated DC]; B cells; CD8 T cells; Cytotoxic cells; DC; Eosinophils; iDC[immature DC]; Macrophages; Mast cells; Neutrophils; NK CD56bright cells; NK CD56dim cells; NK cells; pDC[Plasmacytoid DC]; T cells; T helper cells; Tcm[T central memory]; Tem[T effector memory]; TFH[T follicular helper]; Tgd[T gamma delta]; Th1 cells; Th17 cells; Th2 cells; Treg。

## 应用场景

- 免疫浸润分析主要是基于组织样本转录组测序数据或者微阵列芯片数据的分析。
- 在肿瘤研究中，所采集的组织样本并不止含有肿瘤细胞，还会有正常细胞、免疫细胞、基质细胞等，而不同的细胞具有一些标志性的 marker，免疫细胞也是一样。因此，可以根据这些 marker 基因在组织中的表达量，结合一些生物信息学的算法，评估和量化免疫细胞浸润、免疫微环境。
- 免疫浸润分析，正是为了弄清楚肿瘤组织当中免疫细胞的构成而产生的。
- 不同算法对应所使用的 marker 不同，如果手上的数据并未包含对应的 marker 基因，那么将由于匹配不全而无法进行免疫浸润分析。



## 主要结果

	A	B	C	D	E	F	G
1	Sample	aDC	B cells	CD8 T cells	Cytotoxic cells	DC	Eosinophils
2	TCGA-23-1120	0.484704092	-0.197165122	0.482774224	0.135244837	0.0008679	-0.01854293
3	TCGA-29-1695	0.120317708	-0.232663468	0.395989035	0.027392187	-0.0640465	-0.017160611
4	TCGA-61-2003	0.316197751	-0.253873671	0.425561676	0.052807412	0.07587361	0.021127642
5	TCGA-13-1404	0.040349106	-0.330636141	0.445376927	-0.056773214	-0.1436776	-0.041678526
6	TCGA-13-1512	0.326103674	-0.31193584	0.45165902	-0.020135667	-0.1493092	-0.043031697
7	TCGA-23-1022	-0.351007704	-0.174466367	0.419953251	-0.232917617	0.01781876	-0.02376191
8	TCGA-61-2088	-0.324912393	-0.423694889	0.417590524	-0.313327939	-0.2140033	-0.094924582
9	TCGA-13-1507	0.395492183	-0.074743979	0.465220621	0.281810328	0.02680585	0.002733587
10	TCGA-25-1628	0.4514537	-0.251962907	0.487077371	0.167508922	-0.1352232	0.004872521
11	TCGA-31-1944	-0.025705674	-0.254957339	0.429786134	-0.021787694	-0.028971	0.026340208

- 表格中的行名（第一列）代表样本，列名（从第二列开始）代表免疫细胞类型，数值代表每个样本在不同细胞的富集得分。

主要结果包含了 24 种细胞类型的富集得分数据，如果后续需要进行可视化，可以直接作为可视化输入数据进行下游分析。

## 数据格式

	A	B	C	D	E	F	G
1	gene_id	TCGA-23-1120	TCGA-29-1695	TCGA-61-2003	TCGA-13-1404	TCGA-13-1512	TCGA-23-1022
2	RAB4B	2.021775728	4.908403094	5.433856696	2.388734048	6.33655999	5.046220328
3	C12orf5	1.952965422	10.24395474	3.017626555	1.882028238	2.606324505	0.830262666
4	RNF44	28.59748336	21.00536905	28.13665217	28.41160017	14.89508077	16.6884242
5	DNAH3	0.084499389	0.016197327	0.168182336	0.167447976	0.511683969	0.059074999
6	RPL23A	391.8151425	254.0200414	223.8741998	562.7939575	277.7897572	742.1034512
7	ARL8B	27.91084812	24.8952554	23.64663025	16.26606641	15.37472808	11.97626001
8	CALB2	0.840168409	1.252111397	1.725005886	4.362976611	0.394821136	7.950225258
9	MFS3	18.51669087	16.81430036	9.568504292	21.97428333	33.25160649	36.52365769
10	PIGV	4.920701107	2.779924919	4.31818145	10.45757245	6.785078658	5.325567493
11	ZNF708	1.564035738	1.816362239	1.77258742	1.404986122	0.980998278	1.351677936
12	MYADM2	0.039560332	0.024645241	0.011436867	0.172733911	0.056112361	0.028763612
13	PHEX	0.093566277	0.267590803	0.317050584	0.179568589	0.507709635	1.655669383

数据要求：

- 数据至少有 2 列以上，至少需要 100 行数据。第一行为样本编号，第一列为分子 ID，不能含有缺失、重复及特殊字符。数值部分为不同分子在各样本中的表达量。
- ssGSEA 分析是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。输入表达谱数据支持 基因名 | Ensembl | Entrez ID 三类分子 ID，ID 会自动转换成 Entrez ID 后进行内置参数数据（marker）的富集分析。
- 最多支持 500 列，70000 行。若验证数据时返回报错，需要在上传数据内进行相应的调整，然后再上传数据。

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 左右，如果任务执行时间过长，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。)

## 分析参数



The image shows a dialog box titled "分析参数" (Analysis Parameters). It contains two settings: "是否矫正" (Whether to correct) with a red toggle switch turned on, and "物种" (Species) with a dropdown menu showing "人源(Homo)".

- 是否矫正：通过最小值和最大值之间的绝对差值来标准化分数(矫正)。
- 物种：物种选择，目前可以选人源。

## 结果说明

## 主要结果

[下载整份报告](#)

**免疫浸润-ssGSEA**

免疫浸润分析: 利用转录组或者表达谱芯片数据, 通过算法计算组织中免疫细胞的分值情况, 推测组织中免疫细胞的构成情况。  
过程: 表达谱 → 分子转成Entrez ID → 与内置参考数据进行ssgsea分析 → 获得分析结果

Sample	aDC	B cells	CD8 T cells	Cytotoxic cells	DC	Eosinophils	iDC	Macrophages	Mast
TCGA-23-1120	0.4847	-0.19717	0.48277	0.13524	0.0008679	-0.018543	0.12789	0.35284	-0.34
TCGA-29-1695	0.12032	-0.23266	0.39599	0.027392	-0.064047	-0.017161	0.23579	0.49721	-0.22
TCGA-61-2003	0.3162	-0.25387	0.42556	0.052807	0.075874	0.021128	0.32906	0.44635	-0.18
TCGA-13-1404	0.040349	-0.33064	0.44538	-0.056773	-0.14368	-0.041679	0.13768	0.3119	-0.28
TCGA-13-1512	0.3261	-0.31194	0.45166	-0.020136	-0.14931	-0.043032	0.12891	0.37954	-0.28
TCGA-23-1022	-0.35101	-0.17447	0.41995	-0.23292	0.017819	-0.023762	0.09339	0.16733	-0.32
TCGA-61-2088	-0.32491	-0.42369	0.41759	-0.31333	-0.214	-0.094925	0.050971	0.18297	-0.50
TCGA-13-1507	0.39549	-0.074744	0.46522	0.28181	0.026806	0.0027336	0.24579	0.41642	-0.22
TCGA-25-1628	0.45145	-0.25196	0.48708	0.16751	-0.13522	0.0048725	0.26012	0.44543	-0.28
TCGA-31-1944	-0.025706	-0.25496	0.42979	-0.021788	-0.028971	0.02634	0.12452	0.34283	-0.28

[免疫浸润ssGSEA.xlsx](#)

主要结果格式为表格格式, 提供 Excel 格式下载, 结果报告可以下载包括说明文本的内容。

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 左右, 如果任务执行时间过长, 刷新后任然在执行阶段, 建议删除后重新提交。) 任务完成后, 提供 Excel 格式下载。



## 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: GSVA 包

处理过程:

- 1) 基于 R 包-GSVA 中提供的 ssGSEA 算法
- 2) 利用 Immunity 文章(Bindea, Gabriela, et al., 2013)提供的 24 种免疫细胞的 markers
- 3) 计算各样本的免疫浸润情况。



## 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](http://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. 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。

### 2. 如何进行可视化的操作？

答：

提交分析任务完成后，历史记录中会有一条对应的结果记录，可以下载对应的结果表格（免疫浸润 ssGSEA.xlsx）。对数据进行相应绘图模块的整理后，即可进行对应的可视化。

### 3. 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。

### 4. 仙桃使用的算法（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。

