



1

目录

基本概念	3
应用场景	3
主要结果	4
数据格式	5
参数说明	6
映射	6
设置种子	7
文本参数	7
标题	8
风格	8
图片	9
结果说明	10
主要结果	10
方法学	11
如何引用	12
常见问题	13

基本概念

- 词云：又称文字云，是文本数据的视觉表示，由词汇组成类似云的彩色图形，用于展示大量文本数据。

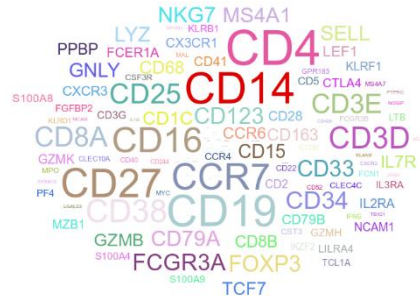


应用场景

词云通常用于描述大量关键字元数据（标签），或可视化自由格式文本。

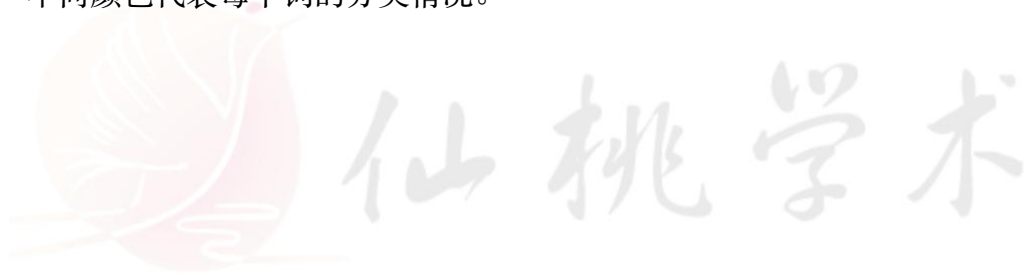
例如，可以根据 CellMarker 网站的检索数据绘制词云，从图中可以看出特定组织、特定细胞的 marker 基因所支持的文献数量多少。

主要结果



按照文本出现的频率对文本进行可视化。

- 图中以字体大小展示每个词的重要性程度/出现频率。
- 不同颜色代表每个词的分类情况。



数据格式

	A	B
1	Marker	Count
2	ALOX5AP	5
3	BCL6	5
4	CCR2	5
5	CCR4	11
6	CCR6	23
7	CCR7	111
8	CD123	44
9	CD14	160
10	CD15	31
11	CD16	107
12	CD160	5
13	CD163	25
14	CD19	159
15	CD1C	30
16	CD2	14

数据要求：

- 数据至少有 1 列以上，当仅有 1 列时，该列即为展示的文本（数据必须为分类型），并根据文本出现的频数绘制词云。当数据含 2 列及以上，根据所选列的数据类型对文本、颜色、大小进行映射。
 - 颜色映射：数据类型可以为数值型 或 分类型。本示例取 Marker 列信息作为文本颜色。
 - 大小映射：数据类型要求为数值型。本示例取 Count 列信息作为文字大小。
 - 文本映射：数据类型要求为分类型。本示例取 Marker 列信息作为文本展示。
- 最多支持 10 列，300 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。

参数说明

(说明：标注了颜色的为常用参数。)

映射



- 颜色映射：可选择 分类型、数值型 数据列信息，根据具体情况映射文本的颜色，默认 不映射。
- 大小映射：可选择 数值型 数据列信息，根据数值情况映射文本的大小，对应的数值越大，文本的大小相对越大，默认 不映射。
- 文本映射：可选择 分类型 数据列信息，将去重后的文本绘制词云，默认 选择数据中第 1 个出现的分类型数据列。

设置种子



- 种子号：设置种子数可以保证文本的排布可重复，默认为 2022，此参数请输入非零整数。调整该参数可改变文本的排布。

文本参数



- 文字颜色：文本的颜色选项，有多少个映射分组会提取多少个颜色，最多支持修改 10 个颜色，受配色方案全局性修改；超过 10 个分组时将由系统随机分配颜色，且不受配色方案全局性修改。
- 排布样式：文本整体的布局类型，可选择 圆形、心形、钻石、矩形、三角形(向上)、三角形(向右)、五边形、星形，默认为圆形。单选，选择类型后将改变整体形状。数据太少或文本字符太长时，排布的图形形状将不明显。
- 缩放大小：将整体图形按照一定比例自适应图片大小，取 1-6 之间，值越大图形相对图片越大。

标题

标题	▼
大标题	大标题内容

- 大标题：大标题文本

风格

风格	▼
文字大小	7pt ▼

- 文字大小：针对图中标题文字的大小控制，默认为 7pt

图片

图片	▼
宽度 (cm)	6
高度 (cm)	6
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



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、ggwordcloud 包（用于可视化）



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. 不适合的场景?

答:

- 1) 数据区分度不大的, 比如数据的频数都相同时, 使用词云则起不到突出的效果。
- 2) 数据特别少, 此时很难排布出好看的词云。
- 3) 文本字符太长, 词云的本质是点图, 是在相应坐标位置绘制具有特定样式的文字的结果, 字符过长时可能会超出布局样式的边界限制。

2. 字体都堆在一起, 怎么办?

答:

当文本特别多的时候, 可以通过同时调整 [图片宽高](#) 及 [缩放大小](#) 参数进行调整, 其中 [缩放大小](#) 参数, 可设置为 1-6 之间的, 值越大, 相对于画布的比例越大。