


# 北京极客基因科技有限公司

## Smart-seq2 线上 DIY 生物信息分析平台使用说明

### 一、平台介绍

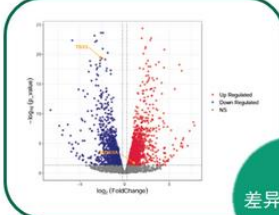
当前大多数的生物信息分析结果中只包含静态的分析报告。静态分析报告中包含的客户所需的大多数分析结果，通常是使用常见参数和阈值得到的固定表格、图片等结果。但很多客户仍会有许多个性化的需求，尤其是 Smart-seq2 转录组测序结果，往往需要反复校正和查看。如果需要在普通分析结果的基础上再进行参数调整、图片修改、查看其它结果该如何进行操作呢？北京极客基因生信团队经过不断设计与修改，开发了一套适用于 Smart-seq2 单细胞转录组数据“再分析”的在线分析平台（<http://bioinfo.geekgene.com.cn>）。将您的基础数据（计数矩阵）导入后台，就可以通过对各个功能参数的调整，来满足您的个性化分析需求，下图为在线分析平台主页面。



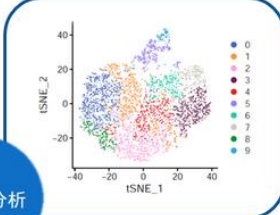
极客基因  
Smart-seq2单细胞转录组  
生信分析平台

主目录
数据准备
差异表达分析
差异分析可视化
聚类分析
单基因表达水平
热图
韦恩图
GO分析
Pathway分析
关于平台


**● 注**  
需先运行“数据准备”和“差异表达分析”步骤，才可运行其他功能。  
如果不是从本公司进行分析的数据，客户需提供Reads数、FPKM值和TPM值的表达矩阵。如果仅提供Reads数的表达矩阵，则FPKM值和TPM值的相关功能无法使用。



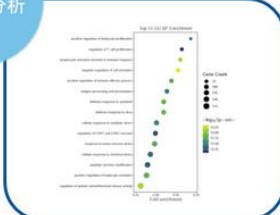
差异基因



聚类分析



基因表达



功能分析

**If using our platform in publication, please cite:**  
Geekgene Bioinformatics Service Portal/Platform (<http://bioinfo.geekgene.com.cn/>)

电话: 010-62897812 | 官网: [www.geekgene.com.cn](http://www.geekgene.com.cn) | E-mail: [service@geekgene.com.cn](mailto:service@geekgene.com.cn)  
 QQ: 3323103858 | 地址: 北京市海淀区天秀路10号中农大国际创业园1号楼603室  
 Copyright 2016 All Rights Reserved. 北京极客基因科技有限公司 版权所有 京ICP备19041323号

在我们给您发送的生信分析结果邮件中直接点击我公司生信分析平台网址，可以直接进入平台页面，输入您的专有项目编号进入您自己的分析主页（如下图所示）。

老师：

您好！附件为smartseq2\_example生信分析报告（打开html文件）、生信需求单。生信分析平台地址为<http://bioinfo.geekgene.com.cn/>，使用期限三个月，项目编号为smartseq2\_example。请您查收！  
祝好



## 平台入口

smartseq2\_example
开始分析

(若无法打开网页，请等待一分钟后刷新)

每份生信分析报告均提供对应平台编号。  
自报告向您发送之日起，平台可使用100天。如有延长需要，请与客服联系。

### 三、模块功能介绍

在线分析平台包括：数据准备、差异表达分析、差异分析可视化、聚类分析、单基因表达水平、热图、韦恩图、GO 分析和 Pathway 分析共九个模块。

#### 1、数据准备

对差异表达分析以及后续分析的数据进行过滤和处理。

①样本分组，您可以在这里选择需要的组数和每组包含的样本。需至少选择两个分组，每组至少两个重复。如果选择多个分组，差异分析会进行组间的两两比较。选择完样本后，在右边会出现一个线粒体含量表格，展示您所选样本对应的线粒体百分比，您可以按需求查看。

②自定义组名，根据您的需求自定义 group 名称，用于后续分析中的图像展示，默认是 treat1、control1、treat2、control2。

③数据过滤，通过不同参数对初始矩阵数据进行过滤，可以排除低丰度或者低表达基因，过滤后的矩阵用于差异表达分析。

The screenshot shows a web interface for data preparation, divided into three main sections:

- 1. 样本分组 (Sample Grouping):** Labeled with a red circled '1'. It contains four input fields for selecting groups: 'group1' (with options A1, A2, A3), 'group2' (with options M1, M2, M3), 'group3', and 'group4'.
- 2. 自定义组名 (Custom Group Names):** Labeled with a red circled '2'. It contains four input fields for naming the groups: 'group1' (with 'Treat1' entered), 'group2' (with 'Control1' entered), 'group3' (with 'Treat2' entered), and 'group4' (with 'Control2' entered).
- 3. 数据过滤 (Data Filtering):** Labeled with a red circled '3'. It includes a dropdown for '选择过滤参数' (Select filtering parameter) set to 'Readcounts'. Below are two input fields for filtering thresholds: '样本read counts总和 ≥' (Sample read counts total ≥) set to '1', and '样本read counts均值 ≥' (Sample read counts mean ≥) set to '0'. A note states: '仅显示read counts值，下载可查看包含FPKM、TPM的完整表格。' (Only display read counts values, download to view complete table including FPKM, TPM). At the bottom, there are three buttons: '运行' (Run), '下载原始矩阵表格' (Download original matrix table), and '下载筛选后矩阵表格' (Download filtered matrix table). A red circled '4' points to the '运行' button. A red arrow points from the '运行' button to the '点击下载表格' (Click to download table) text, which is positioned above the download buttons.

设置完成后点击④运行，可以得到原始矩阵表格和筛选后的矩阵表格。展示的结果中只显示了 geneid、genename 和样本的 read counts 数，想要查看 TPM、FPKM 可以通过下载表格进行查看。

主目录 数据准备 差异表达分析 差异分析可视化 聚类分析 单基因表达水平 热图 韦恩图 GO分析 Pathway分析 关于平台

1. 样本分组 ?

选择 group1  
A1 A2 A3

选择 group2  
B1 B2 B3

选择 group3

选择 group4

2. 自定义组名 ?

请输入group1的名称  
Treat1

线粒体含量

Show 10 entries

	A1	A2	A3	B1	B2	B3
Percentage of mitochondria	3.54%	3.68%	4.08%	3.62%	4.02%	3.67%

Showing 1 to 1 of 1 entries

共 29141 个基因; 过滤掉 8571 个基因; 其余 20570 个基因进入差异分析

原始矩阵

Show 10 entries

genename	A1	A2	A3	B1	B2	B3
ENSG000000000003	TSPAN6	0	4	6	3	5
ENSG000000000005	TNMD	0	0	0	0	0
ENSG000000000419	DPM1	431	503	414	391	409
ENSG000000000457	SCYL3	267	319	273	329	308

## 2、差异表达分析

差异表达分析部分对上一步数据准备中过滤出的数据进行差异表达分析。需要首先运行此步骤得到差异表达分析结果后才可进行后续分析。平台使用 R 包 DESeq2 进行差异表达分析，与生信报告一致。点击①运行。

主目录 数据准备 差异表达分析 差异分析可视化 聚类分析 单基因表达水平 热图 韦恩图 GO分析 Pathway分析 关于平台

1. 差异表达分析(请先运行此步骤) ?

运行 ①

注: normalized\_counts是DESeq2包标准化的结果

2. 分析结果过滤及下载: ?

选择用于比较的两组样本 ?

通过基因过滤结果

通过差异检验过滤结果

下载表格

Summary

Show 10 entries

geneID	genename	treat/control	log2FoldChange	p-value	q-value	nc
ENSG00000121410	A1BG	Treat1/Control1	-0.13	6.24e-01	8.00e-01	86.4
ENSG00000121410	A1BG	Control1/Treat1	0.13	6.24e-01	8.00e-01	86.4
ENSG00000268895	A1BG-AS1	Treat1/Control1	-0.09	7.82e-01	8.92e-01	36.1
ENSG00000268895	A1BG-AS1	Control1/Treat1	0.09	7.82e-01	8.92e-01	36.1
ENSG00000148584	A1CF	Control1/Treat1	4.49	1.20e-02	NA	0.00

表格包含了基因 ID、name，treat/control 信息（即实验组/对照组），log<sub>2</sub>FoldChange 值（其中 FoldChange 即实验组样本相对于对照组样本的表达量的差异倍数），p 值，q 值，nc 值（即 normalized\_counts，是 DESeq2 包在差异分析过程中进行 count 标准化处理的结果），FPKM 值及其总和和均值，TPM 值及其总和和均值，以及 Function，即基因的功能。

还可以对差异表达分析的结果进行筛选，可以①选择用于比较的两组样本，比如可以选择只查看 treat1/control1 的结果；也可以②通过基因过滤结果，在选择 geneID 下拉栏中选择想要查看的基因，或者输入基因名搜索，可以多选，就

可以得到想要查看的指定基因的差异表达分析结果；还可以③通过差异检验过滤结果，默认筛选出  $\log_2\text{FoldChange}$  值在-0.3 到 0.3 之间、p 值小于等于 0.05 的条目，这里的默认值与报告中是一致的。

最后可以点击下载表格，下载过滤后的差异分析表格；如果没有选择过滤条件，则下载的是原始表格。

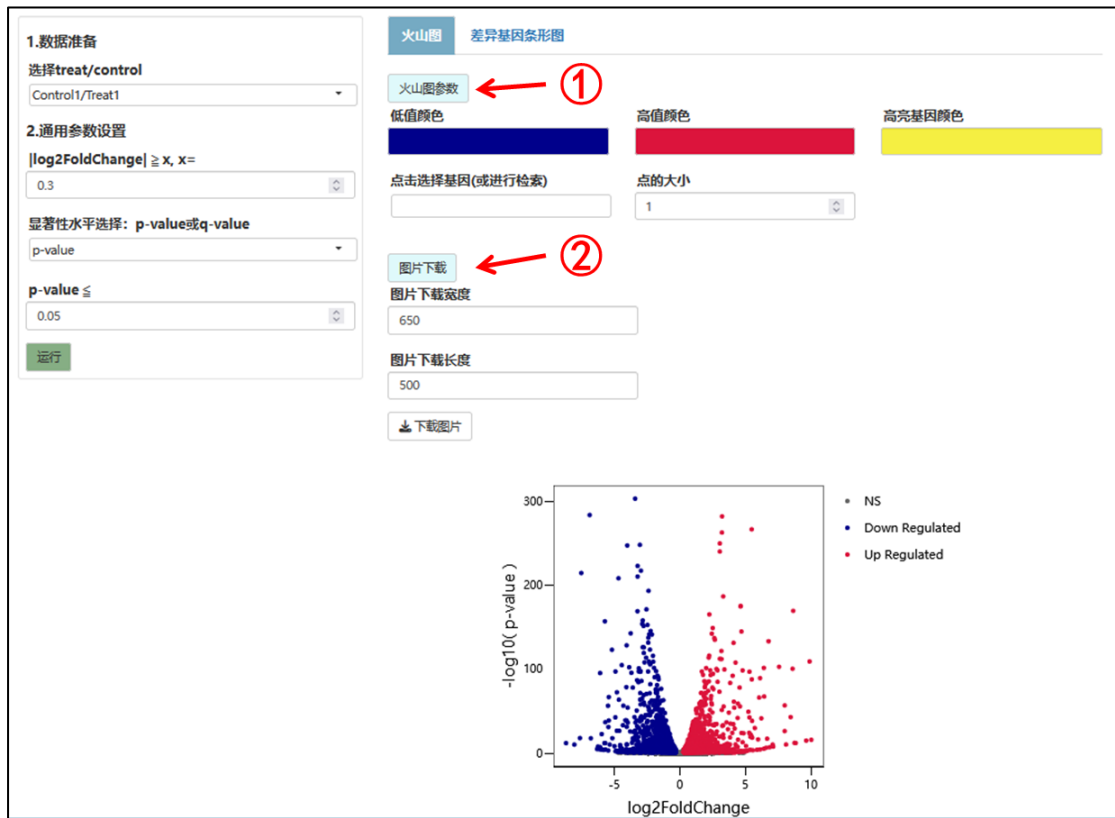
### 3、差异分析可视化

差异分析可视化部分主要将差异表达分析的结果进行可视化，包括火山图和差异基因条形图两部分。左侧栏是火山图和差异基因条形图通用的参数设置。

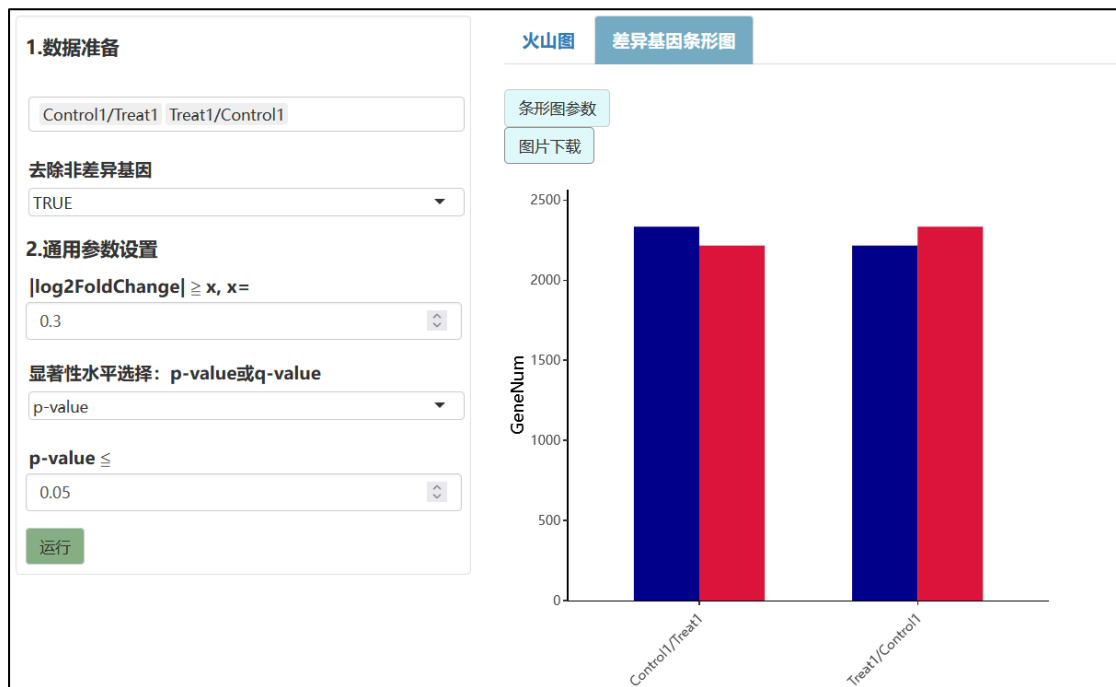
①选择数据类型，我们选择其中的一组 treat/control。

②选择差异倍数， $\log_2\text{FoldChange}$  值在-0.3 到 0.3 之间。选择使用 p 值，p 值小于等于 0.05。这里的默认值与报告中是一致的，您也可以自行调整通用参数，修改  $\log_2\text{FoldChange}$  值的范围，或选择使用 q 值过滤。

首先是火山图。点击③运行，得到 control1/treat1 的整体表达情况，其中蓝色代表表达量下调的基因，红色代表表达量上调的基因，灰色代表表达量无显著差异。点击①“火山图参数”，可以进行火山图专用参数的设置。您可以自行选择低值颜色、高值颜色，也可以对感兴趣的基因进行高亮，修改高亮基因的颜色，以及控制图上点的大小。选择高亮基因时，也可以通过输入基因名或 ID 进行检索。火山图是交互式的，可以将鼠标移动到点上查看具体的基因信息，也可以选择放大或缩小来查看某一区域。点击②“图片下载”，可以设置下载图片的宽度与长度，这里推荐您使用默认长宽；如果有高亮的基因，也会在下载的图片中显示其基因名。



我们再来看差异基因条形图。可以选择要展示的 treat/control 组别，也可以选择是否去除非差异基因，默认去除。点击条形图参数，可以进行差异基因条形图专用参数的设置，您可以自行选择低值颜色、高值颜色。条形图和火山图一样也是交互式的，您可以将鼠标移动到柱子上，查看具体信息。点击图片下载，可以设置下载图片的宽度与长度，下载下来的图片中将在柱子上方标注上调和下调的基因数的具体数字。



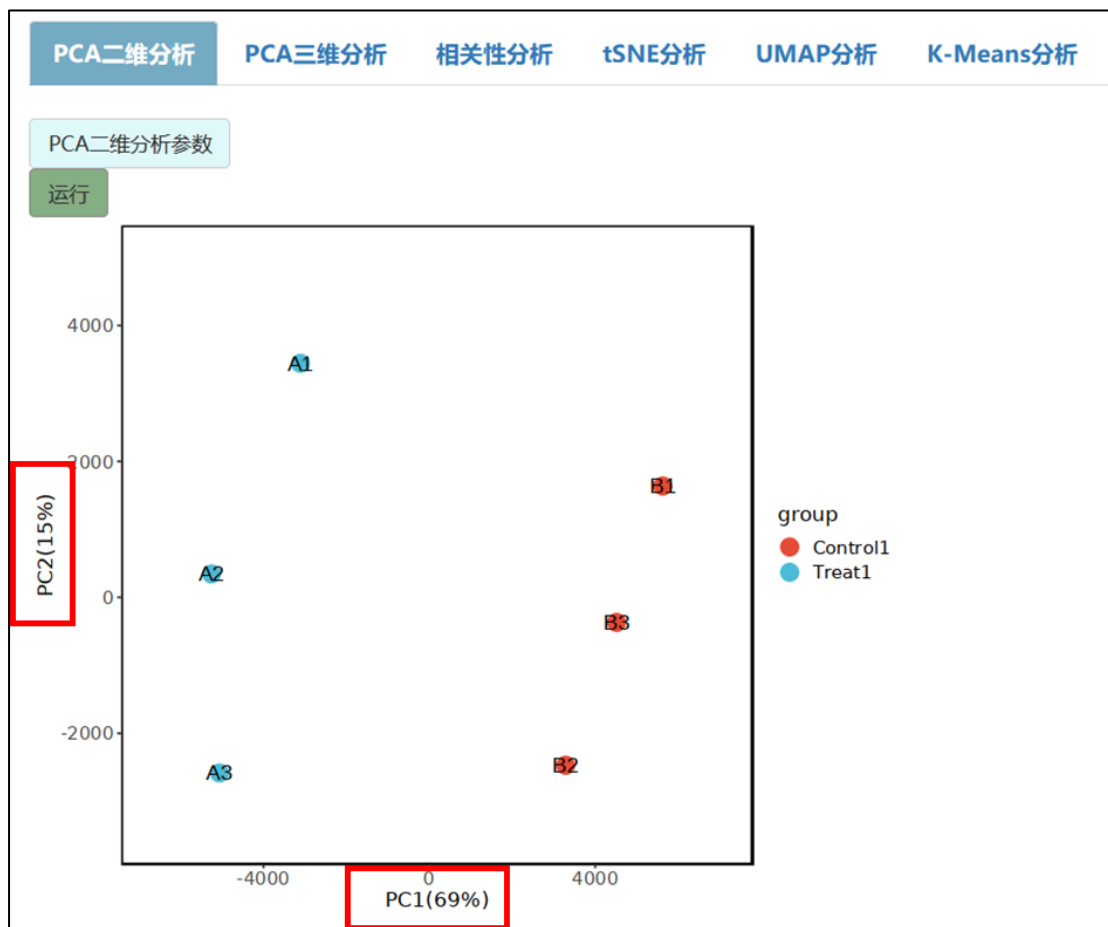
#### 4、聚类分析

聚类分析功能为您展示样本或组之间的异同关系。用于展示所有样本的总体情况，也能够帮助您快速找出异常样本。我们提供了六种方法：PCA 二维分析、PCA 三维分析、相关性分析、tSNE 分析、UMAP 分析和 K-Means 分析。鼠标悬停在数据准备右边的“？”上，可以查看每种方法的简介。在“数据准备”栏选择需要聚类的数据，包括 TPM 值、normalized\_counts 值、FPKM 值或 read counts 值。还可以选择要分析的组别和选择的样本。在“通用参数设置”栏对图片的配色方案进行调整，可以选择系统推荐的配色方案，也可以自定义；系统推荐的配色方案有五种，您可以自行选择。还可以修改图片的长宽。

This screenshot details the 'Data Preparation' (1.数据准备) and 'General Parameter Settings' (2.通用参数设置) sections. In 'Data Preparation', 'TPM' is selected for clustering data, and 'Treat1' and 'Control1' are chosen for grouping. In 'General Parameter Settings', 'System Recommendation' (系统推荐) is selected for color, 'palette1' is the recommended color scheme, and image dimensions are set to 600 width and 480 height. A note states: '注：修改图片大小后，需点击“运行”以更新。' (Note: After modifying the image size, click 'Run' to update). A 'Download Image' (下载图片) button is highlighted with a red arrow and the text '点击下载当前页面的图片' (Click to download the image of the current page).



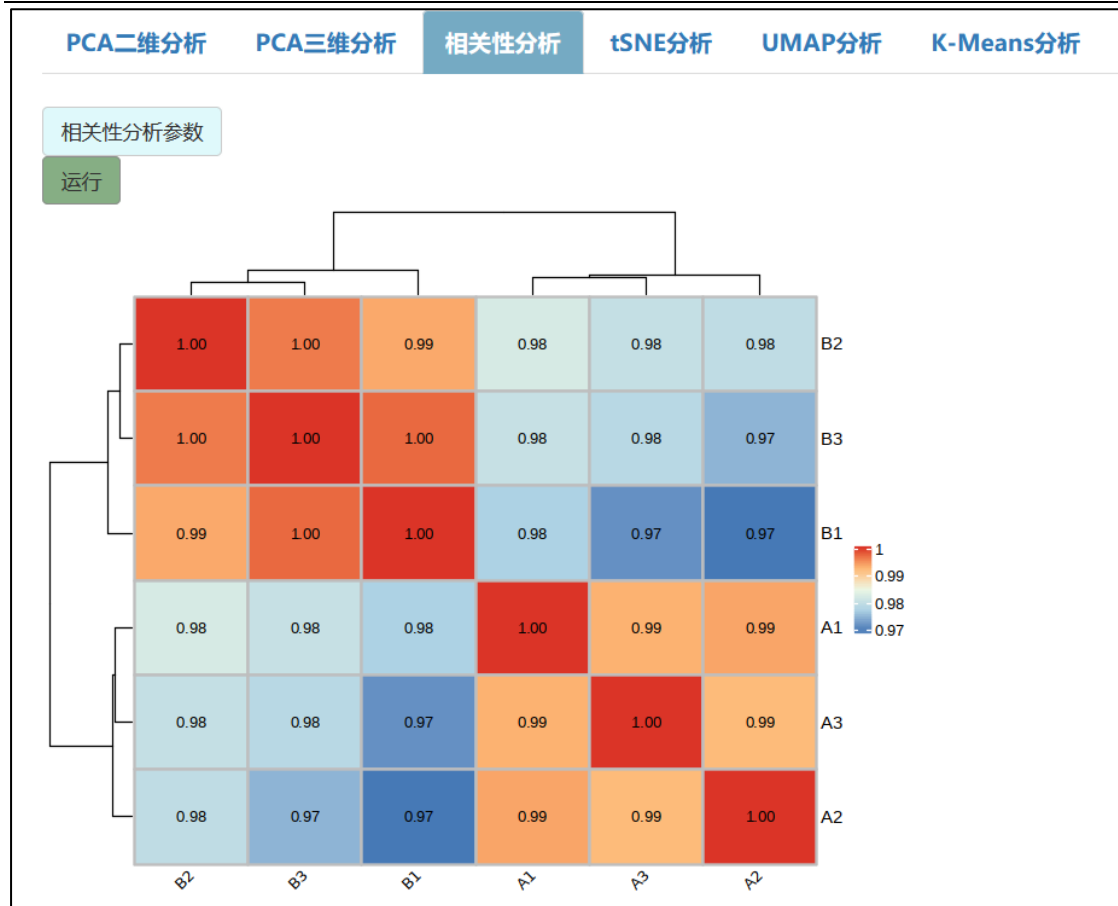
PCA 二维分析里，点击运行，得到图示。点击“PCA 二维分析参数”按钮，可以对图片进行修改。选择主成分（即 PC）中，第一个主成分是原始数据中方差最大的方向，第二个主成分是与第一个主成分正交的平面中使得方差最大的，第三个主成分是与第 1, 2 个主成分正交的平面中方差最大的。我们现在查看的是 PC1 和 PC2，我们也可以查看其他的 PC 值。在较好的情况下，只考虑排列靠前的几个 PC 就能较大程度地解释数据。在坐标轴标签里，我们提供了当前 PC 对数据的解释能力。“颜色分类”的 group 表示一组一个颜色，sample 表示一个样本一个颜色。也可以修改点的大小，或选择是否在点上添加标签。



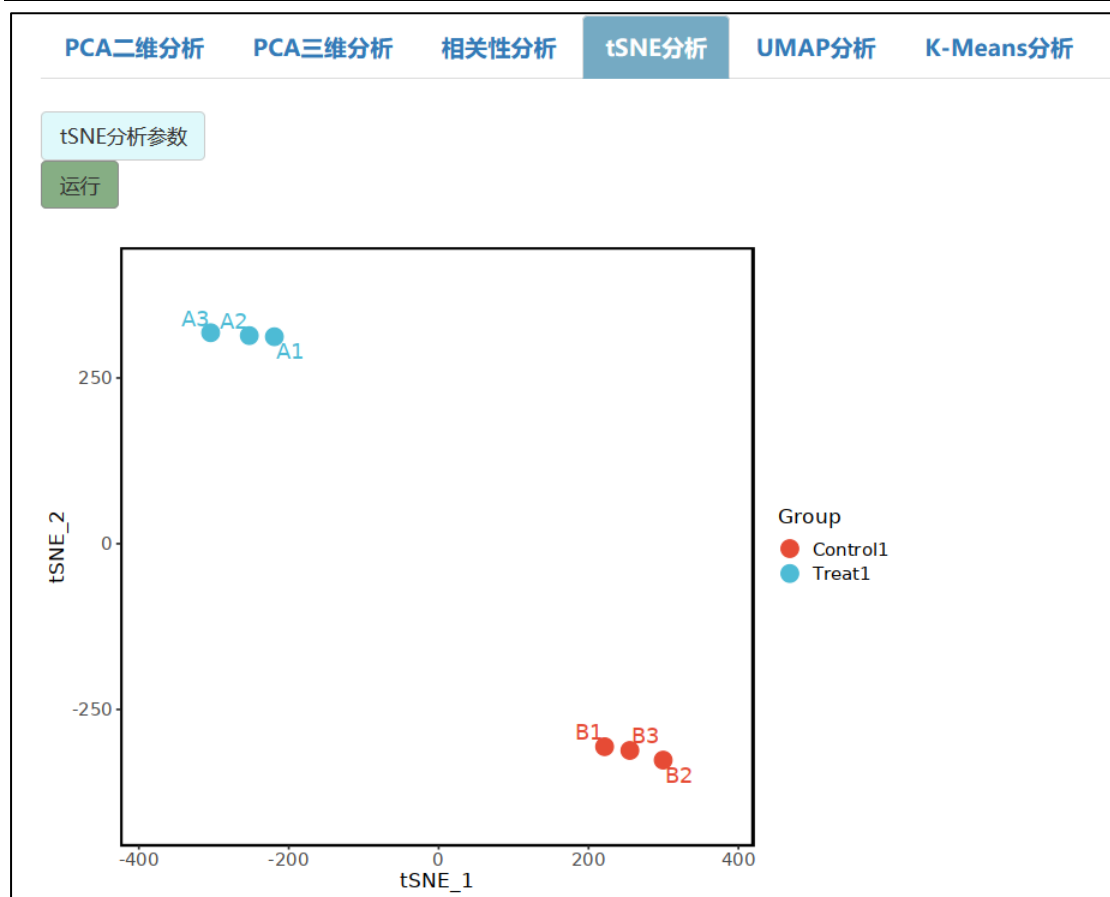
接下来看 PCA 三维分析，点击运行，得到图示。点击“PCA 三维分析参数”按钮，可以对图片进行修改。可以选择图示的主成分，也可以选择颜色分类。

第三种图示是相关性图，相关性图采用的是皮尔森相关系数，每个方框中的数字都是皮尔森相关系数，代表两个样本的相关性，数值越大相关性越好。点击“相关性分析参数”按钮，出现“显示相关性数值”选项，可以选择是否在相关性图上显示数字。



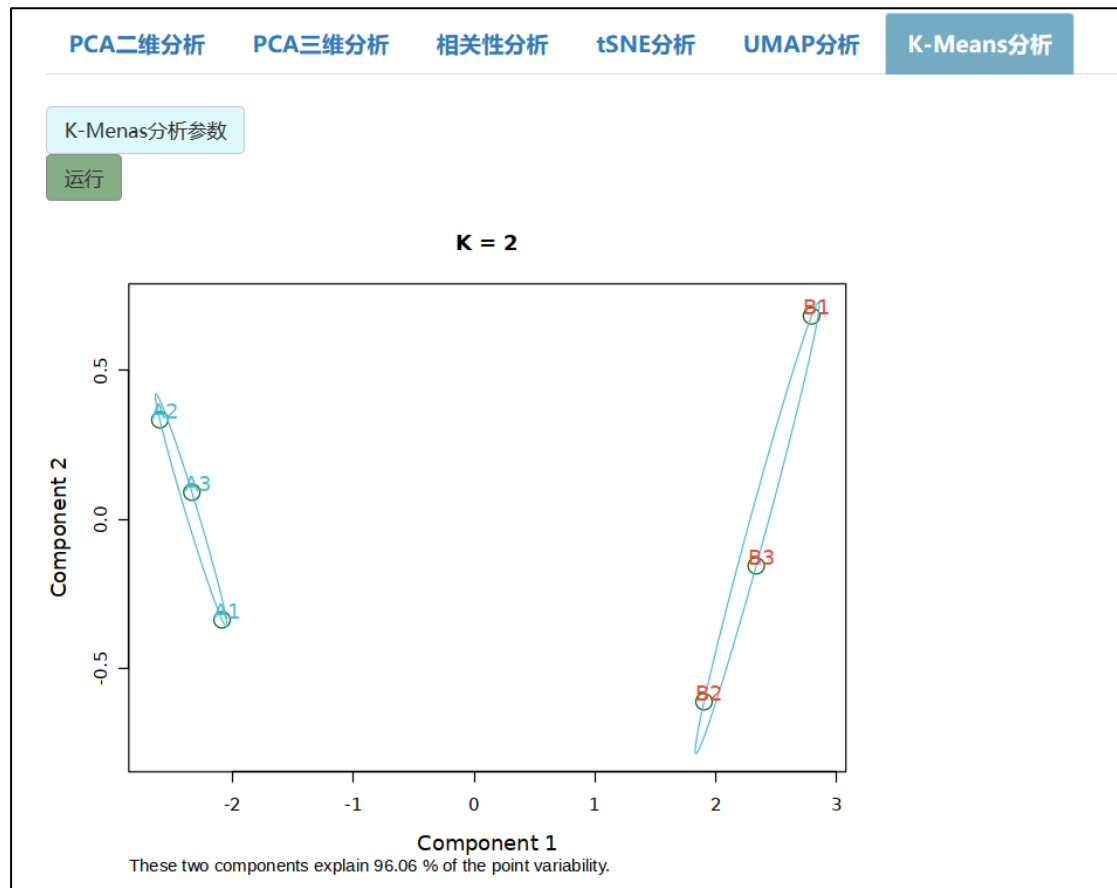


第四种图示是 tSNE 图，tSNE 是一种非线性降维技术，适合于将高维数据嵌入到二维或三维的低维空间中进行可视化。点击运行，得到图示。点击“tSNE 分析参数”按钮，可以对图片进行修改。颜色分类的 group 表示一组一个颜色，sample 表示一个样本一个颜色。也可以修改点的大小，或选择是否在点上添加标签。



第五种图示是 UMAP 图，UMAP 是一种基于流形学习技术和拓扑数据分析思想的降维算法。相比于 t-SNE，UMAP 提高了速度并更好地保存了数据的全局结构。点击“UMAP 分析参数”按钮，可以对图片进行修改。颜色分类的 group 表示一组一个颜色，sample 表示一个样本一个颜色。也可以修改点的大小，或选择是否在点上添加标签。

最后一种是 K-Means 分析，K-Means 是一种聚类算法，其中 K 表示类别数，Means 表示均值。K-means 算法通过预先设定的 K 值及每个类别的初始质心对相似的数据点进行划分，并通过划分后的均值迭代优化获得最优的聚类结果。点击运行，得到图示。点击“K-Means 分析参数”按钮，可以对图片进行修改。可以选择 k 值，即要得到的簇的个数，也可以修改点的大小。



点击左侧栏底部的“下载图片”按钮，能够下载当前页面的图示。

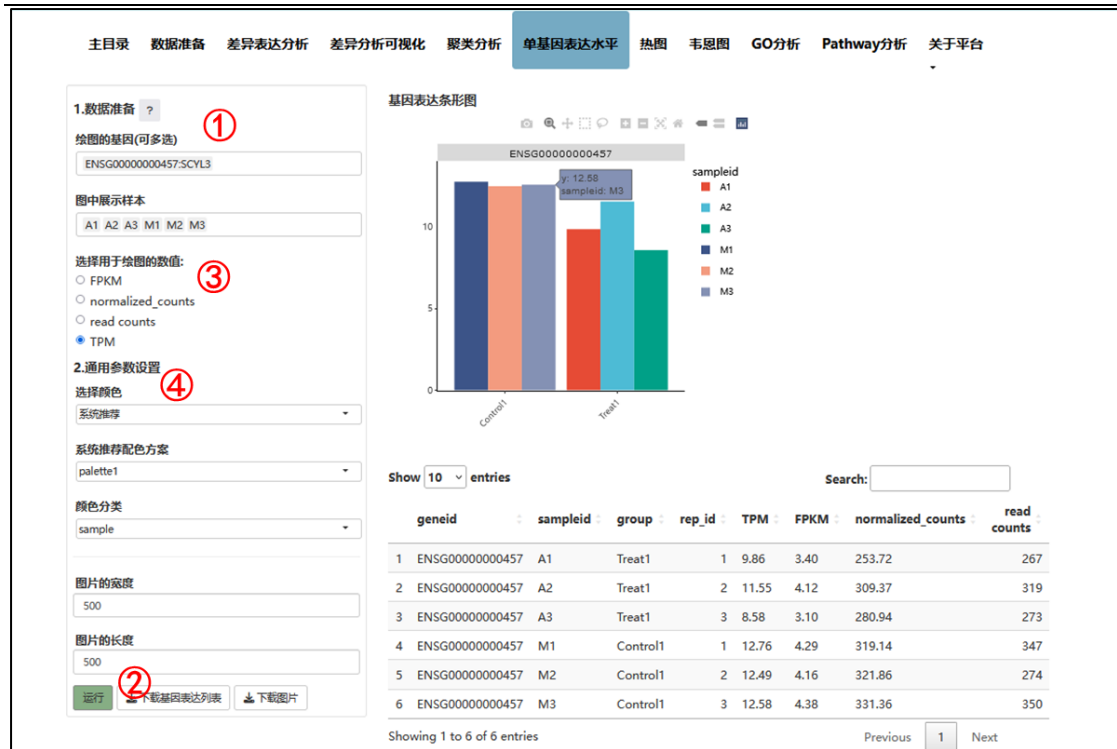
## 5、单基因表达水平

单基因表达水平是指将基因在所有样本中的表达情况用图示的方法展示出来。

①选择感兴趣的基因，点击②运行。还可以选择图中展示样本，即增删横轴代表的样本名。还可以③选择绘图数值，即 y 轴的值，可以是 count 值、FPKM 值、normalized\_counts 值或 TPM 值，默认为 TPM 值。还可以④在通用参数设置栏对图片的配色方案进行调整，可以选择系统推荐的配色方案，也可以自定义。颜色分类的 group 表示一组一个颜色，sample 表示一个样本一个颜色。

将指针放在图示上，可以查看每个柱代表的样本和它的表达情况。

下方的表格是画出图示的原始数据。



图片的宽度和长度可以进行调整。基因表达列表和图片可供下载，下载下来的图片中将在柱子上方标注 y 轴的值。

## 6、热图

热图功能将基因在所有样本中的表达情况用热图的方式展示出来。首先选择绘图数据，包括 TPM 值、read counts 值、normalized\_counts 值或 FPKM 值，这四种是不同的标准化方法。然后我们选择基因或上传基因绘图，如果选择“上传基因”，则下面的基因筛选和基因排序功能不运行，上传的基因形式为基因 ID，上传的文件包含一列，每行一个基因；可以点击下载“示例数据”进行参考。如果选择“选择基因”，因为所有的表达基因过多，我们只能选择部分感兴趣的基因进行画图。先通过差异表达结果对基因进行过滤，选择想查看的 treat/control 组别，然后筛选出  $\log_2\text{FoldChange}$  值在 -0.3 到 0.3 之间、p 值小于等于 0.05 的条目，这里的默认值与报告中是一致的，您也可以自行调整过滤条件，修改  $\log_2\text{FoldChange}$  值的范围，或选择使用 q 值过滤。然后将上一步骤过滤后的数据，按照某个数值排序，取前 n 个基因用于绘图。默认使用 p-value 升序，n 等于 100。

**1.数据准备 ?**

选择用于绘图的数据

TPM ①

read counts

normalized\_counts

FPKM

选择基因或上传基因绘图

选择基因 ②

上传基因

**2.基因筛选 ? ②-1 选择基因**

选择treat/control

Treat1/Control1

$|\log_2\text{FoldChange}| \geq x, x=$

0.3

显著性水平选择: p-value或q-value

p-value

p-value  $\leq$

0.05

④

**1.数据准备 ?**

选择用于绘图的数据

TPM

read counts

normalized\_counts

FPKM

选择基因或上传基因绘图

选择基因

上传基因 ②-2 上传基因

只能上传基因ID, 例如: ENSG000000000003.

示例数据

上传文件(每行一个基因名)

Browse... No file selected

**3.基因排序 ? ③**

基因排序方式

p-value升序

选择前n个基因(建议小于5000个)

100

运行 下载表格

Please select a file

点击运行，得到两张表格。基因筛选结果是包含参与绘图的基因的  $\log_2\text{FoldChange}$  值、p 值、q 值等信息的矩阵，用于在基因筛选和排序时进行参考。“此表格数据用于绘制热图”矩阵是绘图所用到的数据的矩阵。

作图数据准备
热图

**基因筛选结果**

Show 10 entries Search:

geneID	treat/control	log2FoldChange	p-value	q-value	genename	
1	ENSG00000006075	Treat1/Control1	-3.40	1.15e-100	2.28e-98	CCL3
2	ENSG00000010278	Treat1/Control1	2.83	5.75e-159	2.67e-156	CD9
3	ENSG00000010327	Treat1/Control1	4.36	0.00e+00	0.00e+00	STAB1
4	ENSG00000012223	Treat1/Control1	2.26	1.37e-106	3.09e-104	LTF
5	ENSG00000030582	Treat1/Control1	2.03	6.73e-110	1.61e-107	GRN
6	ENSG00000050730	Treat1/Control1	-4.65	4.35e-176	2.44e-173	TNIP3
7	ENSG00000085117	Treat1/Control1	-2.63	5.40e-138	1.83e-135	CD82
8	ENSG00000085265	Treat1/Control1	7.24	0.00e+00	0.00e+00	FCN1
9	ENSG00000090382	Treat1/Control1	2.55	3.70e-172	1.94e-169	LYZ
10	ENSG00000093072	Treat1/Control1	2.39	2.84e-194	1.71e-191	CECR1

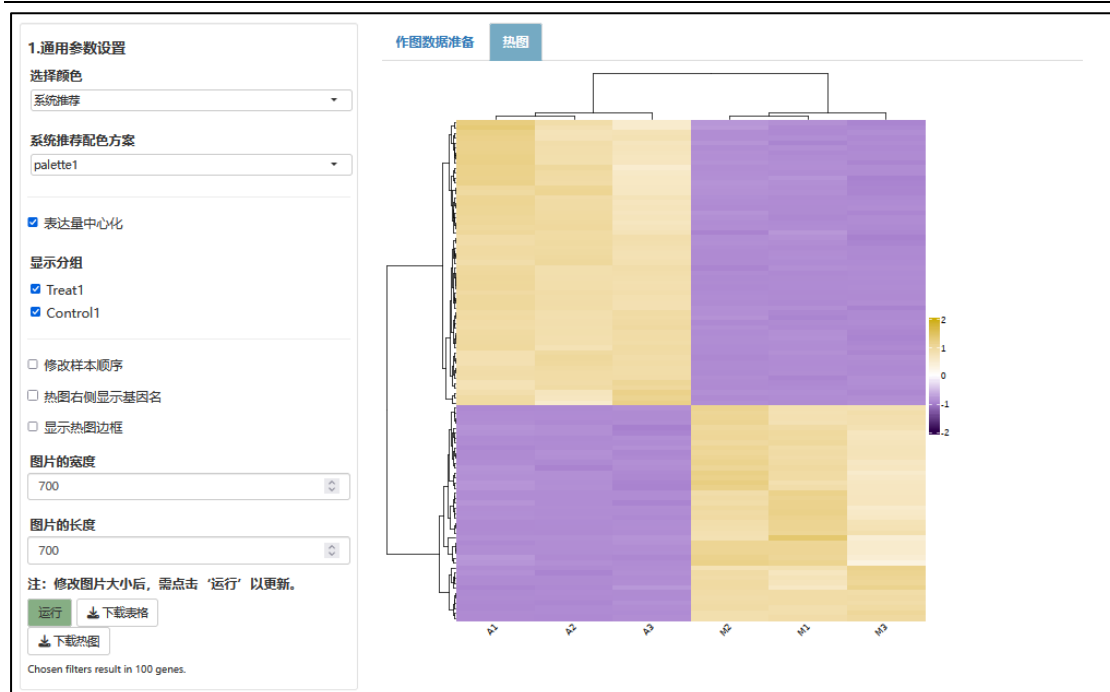
Showing 1 to 10 of 100 entries Previous      ...  Next

**此表格数据用于绘制热图**

Show 10 entries Search:

geneid	A1	A2	A3	M1	M2	M3	
1769	ENSG00000090382	10860.58	11726.20	11505.15	2476.43	2292.30	2090.97
11417	ENSG00000169429	228.56	268.86	343.21	12357.71	10971.11	10182.04
14944	ENSG00000197746	3439.09	3362.42	3558.21	1057.26	1122.14	1059.35
3735	ENSG00000112096	308.95	294.29	339.72	2741.63	2647.63	2853.59
6663	ENSG00000136235	2241.67	2096.55	2016.89	52.44	57.83	46.27
1554	ENSG00000085265	1414.75	1381.24	1339.13	6.91	11.17	9.21
457	ENSG00000030582	1425.90	1348.83	1402.08	337.98	400.81	360.86
9972	ENSG00000163131	904.45	918.96	885.45	243.74	233.68	194.89
15130	ENSG00000198502	933.73	999.67	1068.50	189.16	194.24	205.70
2520	ENSG00000102575	1152.64	1116.79	1176.75	127.88	120.35	132.66

点击“热图”，我们可以看到上一步过滤得到的 100 个基因的表达热图。在通用参数设置栏对图片的配色方案进行调整，可以选择系统推荐的配色方案，也可以自定义。可以勾选是否进行表达量中心化，即将原始数值减去平均值后除以标准差，这样可以避免极端值影响图示美观。还可以选择热图显示的组别。勾选“修改样本顺序”，可以对热图中显示哪些样本及显示的顺序进行调整。勾选“热图右侧显示基因名”，可以在热图右侧显示基因的 ID，您也可以根据基因数的多少自行调节字号。勾选“显示热图边框”，则为每个小方格加框。还可以修改图片的长宽，修改图片大小后，需要点击“运行”以更新。



“运行”按钮旁边有下载表格和下载热图按钮，您可以下载您感兴趣的表格或热图。

## 7、韦恩图

韦恩图是用于显示元素集合重叠区域的图示，反映不同的 treat/control 组中上调、下调或差异的基因的大致关系，主要是交集。首先选择 treat/control 组别，最少选两组，最多选四组；&符前的文字代表组别，&后的 up、down、diffgene 分别代表该组中的上调、下调或差异基因。然后选择差异倍数， $\log_2\text{FoldChange}$  值在-0.3 到 0.3 之间。然后选择使用 p 值，p 值小于等于 0.05。这里的默认值与报告中是一致的，您也可以自行调整通用参数，修改  $\log_2\text{FoldChange}$  值的范围，或选择使用 q 值过滤。在通用参数设置栏对图片的配色方案进行调整，可以选择系统推荐的配色方案，也可以自定义。还可以修改图片的长宽，修改图片大小后，需要点击“运行”以更新。

点击“运行”，得到韦恩图和交集基因表格，其中交集基因即在所选的所有组别中都存在的基因。“运行”按钮旁边有下载表格和下载韦恩图按钮，您可以下载您感兴趣的表格或韦恩图。



主目录 数据准备 差异表达分析 差异分析可视化 聚类分析 单基因表达水平 热图 韦恩图 GO分析 Pathway分析 关于平台

1. 数据筛选 ?

选择treat/control组别(至少两个)

Control1/Treat1&up Treat1/Treat2&sup  
Control2/Control1&up Control1/Control2&up

$|\log_2\text{FoldChange}| \geq x, x =$

0.3

显著性水平选择: p-value或q-value

p-value

p-value  $\leq$

0.05

2. 通用参数设置

选择颜色

系统推荐

系统推荐配色方案

palette1

图片的宽度

500

图片的长度

500

注: 修改图片大小后, 需点击 '运行' 以更新。  
注: 下载查看完整交集基因表格。

运行 下载韦恩图 下载表格

韦恩图

交集 (Read counts值)

数据没有交集

## 8、GO 分析

Gene Ontology(简称 GO)是一个国际化的基因功能分类体系, GO 分析是将筛选出的所有差异表达基因进行功能注释, 将各个基因注释到不同的功能条目中, 并进行统计分析。

首先进行选择基因集或上传基因集。关于前景基因和背景基因, 富集就是比较某个 GOterm 里的基因在前景基因所占的比例, 是否显著高于这个 GOterm 里的基因在背景基因所占的比例, 然后根据  $p < 0.05$ , 判断是否有显著富集。背景基因使用全部表达基因。

然后选择物种。若选择上传基因集, 只能上传基因名, 上传的文件包含一列, 每行一个基因, 可以点击下载“示例数据”进行参考。若选择选择基因集, 首先选择 treat/control 组别。然后选择差异倍数,  $\log_2\text{FoldChange}$  值在-0.3 到 0.3 之间。然后选择使用 p 值, p 值小于等于 0.05。您也可以自行修改  $\log_2\text{FoldChange}$  值的范围, 或选择使用 q 值过滤。

主目录 数据准备 差异表达分析 差异分析可视化 聚类分析 单基因表达水平 热图 韦恩图 **GO分析** Pathway分析 关于平台

**1.数据准备 ?**

基因列表:

选择基因集

上传基因集

选择物种

human(Homo sapiens)

选择treat/control

Control1/Treat1

$|\log_2\text{FoldChange}| \geq x, x=$

0.3

显著性水平选择: p-value或q-value

p-value

p-value  $\leq$

0.05

数据准备 **GO分析结果**

用于富集的基因列表

Show 10 entries Search:

geneID	test	log2FoldChange	p-value	q-value	genename	nc_A1
ENSG00000000003	Control1/Treat1	1.88	2.73e-02	9.75e-02	TSPAN6	0
ENSG000000000938	Control1/Treat1	-0.53	2.17e-08	2.80e-07	FGR	2557.2
ENSG000000000971	Control1/Treat1	-0.61	2.53e-02	9.18e-02	CFH	88.38
ENSG000000001036	Control1/Treat1	-0.72	1.75e-05	1.51e-04	FUCA2	377.26
ENSG000000001084	Control1/Treat1	-1.23	2.69e-28	1.07e-26	GCLC	804.89
ENSG000000001460	Control1/Treat1	1.13	3.69e-04	2.47e-03	STPG1	34.21
ENSG000000001561	Control1/Treat1	0.43	9.76e-04	5.85e-03	ENPP4	405.77
ENSG000000001630	Control1/Treat1	-1.08	1.23e-03	7.17e-03	CYP51A1	71.27
ENSG000000002822	Control1/Treat1	-0.48	6.53e-04	4.10e-03	MAD1L1	806.79
ENSG000000002933	Control1/Treat1	-1.18	1.80e-18	4.67e-17	TMEM176A	1303.78

Showing 1 to 10 of 4,548 entries Previous 1 2 3 4 5 ... 455 Next

点击运行。GO 分析比较慢，一般需要一分钟左右。现在点击“GO 分析结果”查看，可以看到富集结果。在“富集结果筛选”栏可以对富集的结果进行筛选。富集结果中，ID 是 GOterm 的 ID，Description 为基因功能描述，GeneRatio 为差异表达基因集中有某种功能的基因数比上差异表达基因总数，BgRatio 表示具有某种功能的基因总数比上基因组中所有基因、所有分析的基因数，count 表示具有该条目功能的差异基因数，group 代表基因的分子功能(molecular function)、细胞组分(cellular component)和参与的生物过程(biological process)。另外还有该条目所包含的 geneID 列，因为比较长，在这里不予显示，如果您想要查看的话，可以点击下载 GO 结果查看完整表格。

主目录 数据准备 差异表达分析 差异分析可视化 聚类分析 单基因表达水平 热图 韦恩图 **GO分析** Pathway分析 关于平台

**1.富集结果筛选**

筛选的阈值选择: p-value或q-value  
p-value

GO p-value ≤  
0.1

**2.绘图结果展示**

选择图片显示的GO条目  
GO:0002283 GO:0043312 GO:0001819 GO:0071222  
GO:0032496 GO:0071219 GO:0071216 GO:0042581  
GO:002237 GO:0022407

选择图示  
 气泡图  
 条形图(geneNum)  
 条形图(pvalue)

**3.通用参数设置**

选择颜色  
系统推荐

系统推荐配色方案  
palette1

y轴字号  
11

气泡图中点的大小  
4

下载图片的宽度  
1000

下载图片的长度  
600

运行 下载GO结果 下载GO图示

**GO分析结果**

geneID过长, 不予显示, 查看可点击下载GO结果

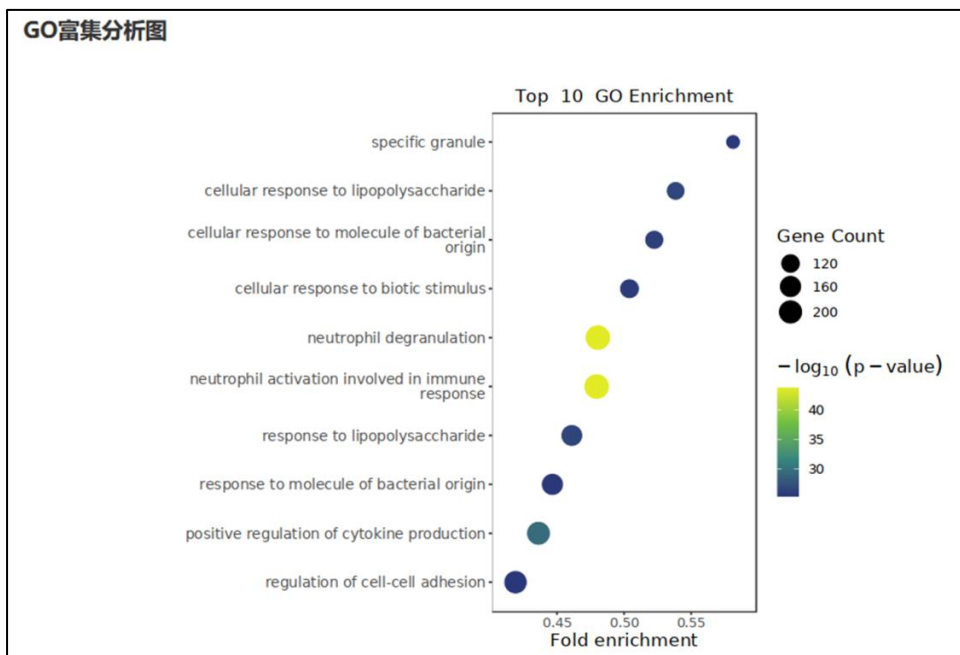
Show 10 entries Search:

ID	Description	GeneRatio	BgRatio	p-value	p.adjust	q-value	Count	group
GO:0002283	neutrophil activation involved in immune response	235/3797	490/18866	1.39e-44	4.61e-41	3.14e-41	235	biological_process
GO:0043312	neutrophil degranulation	234/3797	487/18866	1.44e-44	4.61e-41	3.14e-41	234	biological_process
GO:0001819	positive regulation of cytokine production	195/3797	447/18866	3.58e-30	7.67e-27	5.22e-27	195	biological_process
GO:0071222	cellular response to lipopolysaccharide	112/3797	208/18866	3.10e-27	4.64e-24	3.15e-24	112	biological_process
GO:0032496	response to lipopolysaccharide	154/3797	334/18866	3.61e-27	4.64e-24	3.15e-24	154	biological_process
GO:0071219	cellular response to molecule of bacterial origin	116/3797	222/18866	1.24e-26	1.33e-23	9.07e-24	116	biological_process
GO:0071216	cellular response to biotic stimulus	124/3797	246/18866	1.72e-26	1.58e-23	1.07e-23	124	biological_process
GO:0042581	specific granule	93/3924	160/19559	3.03e-26	2.27e-23	1.75e-23	93	cellular_component
GO:0002237	response to molecule of bacterial origin	159/3797	356/18866	3.86e-26	3.10e-23	2.11e-23	159	biological_process
GO:0022407	regulation of cell-cell adhesion	184/3797	439/18866	5.63e-26	4.02e-23	2.73e-23	184	biological_process

Showing 1 to 10 of 3,394 entries Previous 1 2 3 4 5 ... 340 Next

**GO富集分析图**

在“绘图结果展示”中选择图示的GO条目。在“选择图示”栏选择图示的类别，可以选择气泡图、基因数的条形图或p值的条形图。其中，气泡图的纵轴代表GO条目，横轴代表 $-\log_{10}(p\text{-value})$ ，点的大小代表基因数。在通用参数设置栏对图片的配色方案进行调整，可以选择系统推荐的配色方案，也可以自定义。还可以修改y轴字体的大小，以及设置下载图片的宽度和长度。



## 9、Pathway 分析

Pathway 分析中的 pathway 代表 KEGG 中的 pathway 数据库，是应用最广泛的代谢通路公共数据库。Pathway 分析的具体操作与 GO 分析非常类似，这里不展开介绍。

北京极客基因科技有限公司

官方网址：<http://www.geekgene.com.cn/>

生信分析平台：<http://bioinfo.geekgene.com.cn/>

Tel: 010-62897812 E-mail: [service@geekgenen.com.cn](mailto:service@geekgenen.com.cn)

地址：北京市海淀区天秀路 10 号中国农大国际创业园 1 号楼 603 室



微信公众号



公司官网



生信分析平台