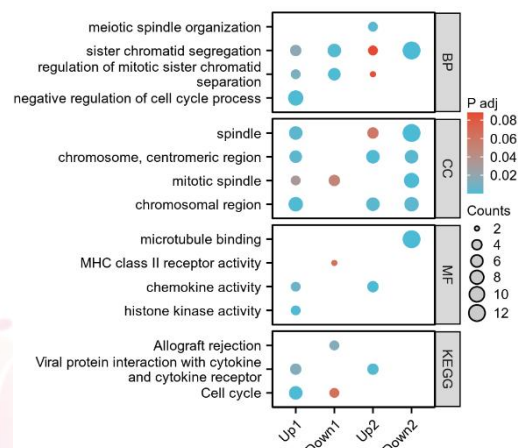


功能聚类 - 分组点图



网址: <https://www.xiantao love>



更新时间: 2023.11.16

目录

基本概念	3
应用场景	3
主要结果	4
云端数据	5
参数说明	6
ID 列表	6
样式	7
点	8
分面	9
标题	9
图注(Legend)	10
坐标轴	10
风格	11
图片	12
结果说明	13
主要结果	13
补充结果	14
方法学	15
如何引用	16
常见问题	17

基本概念

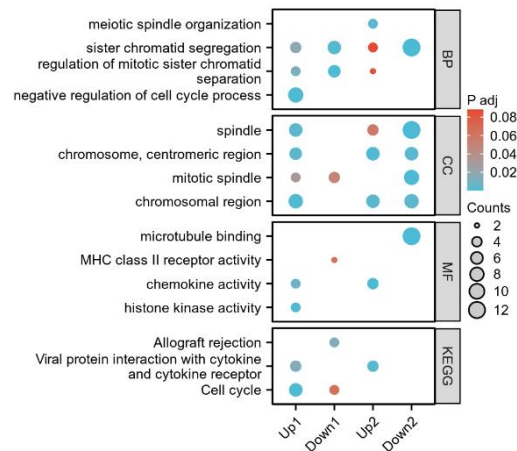
- 富集分析：简单而言，就是取一部分有功能注释的分子与所有有功能注释的分子去比较（超几何分布检验），确定这一部分分子中都涉及了哪些功能作用。注意：单独几个分子做富集分析意义并不大。
- GO (Gene Ontology, 基因本体) 数据库：把基因的功能分成了三类：生物过程 (biological process, BP)、细胞组分 (cellular component, CC)、分子功能 (molecular function, MF)。利用 GO 数据库，可以得到目标基因在 CC, MF 和 BP 三个层面上有什么关联。
- KEGG (Kyoto Encyclopedia of Genes and Genomes) 数据库：一种通路数据库，收集了很多通路相关的数据库。通路数据库还包括 wikipathway, reactome 等。
- 超几何分布检验：超几何分布 (hypergeometric) 是统计学上一种离散概率分布。它描述了在 N 个物件中指定 M 个种类的物件，不放回的抽取 n 个，成功抽中指定类型物件的个数 (k) 的事件。

应用场景

本模块为 GO|KEGG 分组 分析后结果的可视化展示。增加分组富集分析的功能，比较不同分组的富集结果可以揭示不同分组之间的功能共性和差异。

注意：模块需要先进行 GO|KEGG 分组 并 保存结果后，此处的云端数据才会有结果记录，然后才能进行可视化的操作。主要基本参数中有 ID 列表 输入框，可以将对应云端数据记录的富集分析表格中的感兴趣的 ID 列复制到此处，进行可视化。

主要结果



通过点图的形式展示 GOKEGG 分组 分析结果。

- 纵坐标为所选类目名称，横坐标为分组名称。
- 图中点代表了对应分组是否富集到了相应的功能类目，直观展示了各分组之间共有或特有的功能。
- 图中点的颜色为所选择的 颜色映射 内容。（如上图，对应类目的校正后 p 值(padj) ，注意有分面参数）
- 图中点的大小为所选择的 大小映射 内容。（如上图，对应类目包含 ID 的数量 ）
- 默认按照类目的类别进行分面展示，可以通过 分面 相关参数进行修改。

默认展示各分组中各类别的 top 几个结果，分面对应的数据库或者分类。可以挑选在满足阈值下的 top 的类目，或者一些感兴趣的类目。

云端数据

云端数据

	记录名称	来源模块	时间	补充说明
<input checked="" type="checkbox"/>	gokegg_group	GOKEGG分组 @1.0	2023-11-16 11:11:46	数据记录可以在历史记录中找到

这里的云端数据与历史记录汇总【GOKEGG 分组】模块的数据记录是保持一致的，可以在历史记录中找到相应的数据记录。

根据需要可视化的项目 选择好对应的云端数据记录。默认使用最近生成的分析记录。



参数说明

(说明：标注了颜色的为常用参数。)

ID 列表



- 可视化 ID：输入想要可视化的功能或者通路的条目 ID，默认为对应云端数据结果中各分组中每个类目的前几个条目，可以根据需要进行输入修改。注意：输入的 ID 来自所选云端数据记录的结果，需要先在历史记录中找到对应的记录，下载 excel 结果，复制想要展示的 ID 到这个输入框中，一行代表一个。最多支持 1 张图同时绘制 40 个类目。

样式

样式
▼

ID展示
全名(自动换行) ▼

颜色映射
校正后p值(p) ▼

大小映射
包含ID的数量 ▼

形状映射
无 ▼

- ID 展示: ID 名称过长时,可以根据需要选择换行模式。可选择 全名(自动换行)、全名(一行 20 长度)、全名(一行 30 长度)、全名(一行 40 长度)、全名(一行 50 长度)、全名(一行 60 长度)、全名(一行 70 长度)、全名(一行 80 长度)、全名(不换行)、ID 号。
- 颜色映射: 主要影响点的颜色范围, 注意映射内容的数值类型,数值型数据为渐变色,分类型数据为单个颜色。可选择 p 值(pvalue)、校正后 p 值(padj)、q 值(qvalue)(错误率)、类别(Ontology)、无。
- 大小映射: 主要影响点的大小,可选择 包含 ID 的数量、基因比例(GeneRatio)、无。
- 形状映射: 主要影响点的形状,可选择 类别(Ontology)、无。同时可以在 点-形状 中选择多种形状。

点



点

填充色

描边色

样式

大小比例 1

不透明度 1

- **填充色**：点的填充色颜色选项，取决于 **颜色映射** 参数所选择的内容，展示 数值型内容 时，修改第一和第二色卡作为数值从小到大的渐变色；展示 分类型内容（如 类别）时，有多少个功能类别会提取多少个颜色，最多支持修改 4 个颜色。受配色方案全局性修改。
- **描边色**：点的描边色颜色选项，取决于 **颜色映射** 参数所选择的内容，展示 数值型内容 时，修改第一和第二色卡作为数值从小到大的渐变色；展示 分类型内容（如 类别）时，有多少个功能类别会提取多少个颜色，最多支持修改 4 个颜色。受配色方案全局性修改。
- **样式**：点的样式类型，取决于 **形状映射** 参数所选择的内容，可选择 圆形、正方形、菱形、三角形、倒三角，多选后不同的分组中点的类型也会有不同。
- **大小比例**：点的相对大小，取决于 **大小映射** 参数所选择的内容。
- **不透明度**：点的透明度。0 为完全透明，1 为完全不透明。

分面

分面

分面映射

类别(Ontolo)

分面颜色

☐
☐
☐
☐

文字大小

6pt

- **分面映射**：主要影响图形的分面展示，默认按照类别（对应的数据库或者分类）按行排列分面。可选择 类别(Ontology)、不映射。
- **分面颜色**：分面的标题背景色颜色选项，当 **分面映射** 参数为类别时，有多少个功能类别会提取多少个颜色，最多支持修改 4 个颜色。默认灰色，**不受配色方案全局性修改**。
- **文字大小**：分面标题的文字大小。

标题

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

- **大标题**：大标题文本
- **x 轴标题**：x 轴标题文本
- **y 轴标题**：y 轴标题文本

- 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(Legend)



图注配置面板，包含以下选项：

- 图注：下拉菜单
- 是否展示：开关按钮（当前开启）
- 图注标题：图注标题内容
- 图注位置：默认

- 是否展示：是否展示图注
- 图注标题：可以添加图注标题
- 图注位置：可选择 默认、右、上、下。

坐标轴



坐标轴配置面板，包含以下选项：

- 坐标轴：下拉菜单
- x轴标注旋转：45

- x 轴标注旋转：支持对 x 轴文字进行旋转。适合于 x 轴文字过长的时候。（注意无论是否进行 xy 颠倒，均修改图形横坐标）

风格



- 外框：是否添加外框
- 网格：是否添加网格
- Xy 颠倒：可以颠倒 xy 轴
- 文字大小：针对图中所有文字整体的大小控制

图片

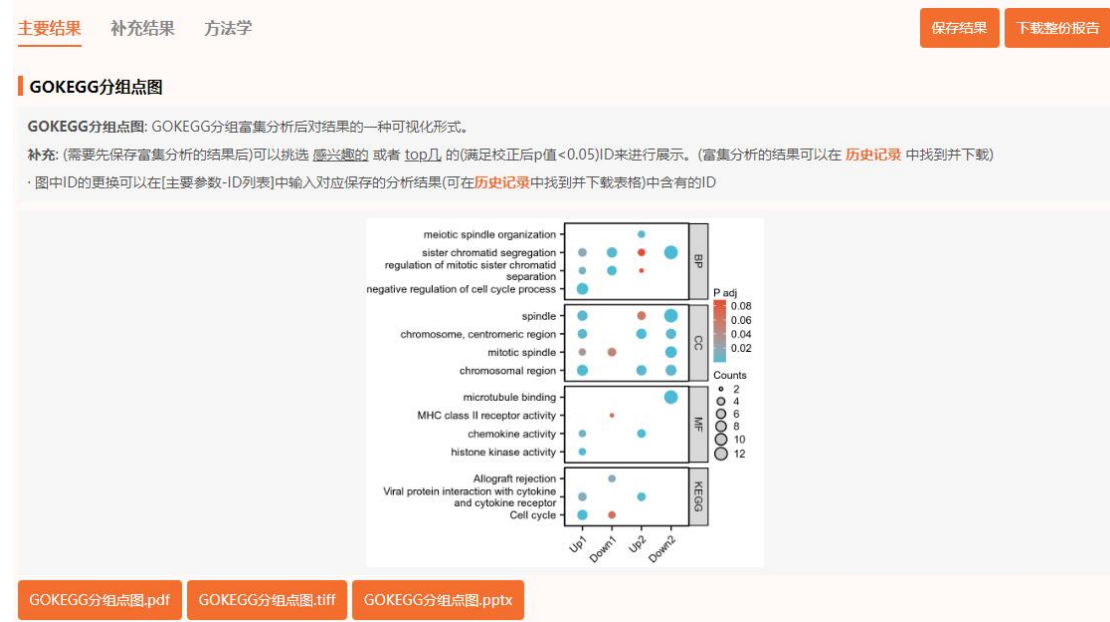
图片	▼
宽度 (cm)	7
高度 (cm)	8
字体	Arial ▼

- 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



结果说明

主要结果



主要结果格式为图片格式，提供 PDF、TIFF 和 PPTX 格式下载，结果报告可以下载包括 pdf 以及说明文本的内容。

补充结果

可视化ID

当前模块可视化所选ID

Group	ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	ger
Up1	BP	GO:0010948	negative regulation of cell...	8/40	301/18800	1.94e-07	0.0003	0.0002	AURKB/BMI
Up1	CC	GO:0098687	chromosomal region	7/40	366/19594	8.21e-06	0.0009	0.0006	AURKB/BIR
Up1	MF	GO:0035173	histone kinase activity	3/38	16/18410	4.46e-06	0.0006	0.0005	AURKB/CI
Up1	KEGG	hsa04110	Cell cycle	6/29	126/8164	4.26e-06	0.0004	0.0004	CCNA2/CCN
Up1	BP	GO:0010965	regulation of mitotic siste...	3/40	65/18800	0.0004	0.0105	0.0070	AURKB/CC
Up1	CC	GO:0072686	mitotic spindle	3/40	160/19594	0.0042	0.0319	0.0217	AURKB/CI
Up1	CC	GO:0000775	chromosome, centromeric r...	5/40	227/19594	9.44e-05	0.0031	0.0021	AURKB/BIR
Up1	MF	GO:0008009	chemokine activity	3/38	49/18410	0.0001	0.0069	0.0049	CXCL10/CX
Up1	KEGG	hsa04061	Viral protein interaction w...	4/29	100/8164	0.0004	0.0137	0.0124	CX3CR1/C>
Up1	BP	GO:0000819	sister chromatid segregation	4/40	205/18800	0.0009	0.0171	0.0113	AURKB/CCN
Up1	CC	GO:0005819	spindle	6/40	402/19594	0.0002	0.0031	0.0021	AURKB/BIR
Down1	KEGG	hsa04110	Cell cycle	3/27	126/8164	0.0080	0.0652	0.0516	MAD2L1/M

GOKEGG可视化ID.xlsx

GOKEGG可视化ID.docx

此表格提供当前可视化的 GOKEGG 分组 分析结果，提供 Excel、Docx 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: ggplot2 包 (用于可视化)

处理过程: 使用 ggplot2 包对分组富集分析结果进行可视化。



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. 可视化结果能否更换别的 ID? 可视化结果的 ID 从哪里获得? 为什么某个条目 (BP、CC、MF、KEGG) 只有 1 个?

答:

在“ID 列表”选项卡中, 有可视化 ID 的输入框:

选项框内默认选择对应云端记录结果中前几个条目, 可以在此处选择想要可视化的 ID。

注意: 输入的 ID 来自所选云端数据记录的结果, 需要先在历史记录中找到对应的记录, 下载 excel 结果, 复制想要展示的 ID 到这个输入框中, 一行代表一个。最多同时支持 40 个。

默认展示前几个的结果, 所有的富集结果需要下载 excel 表格来进行查看。一般大于 4 个分组, 仅展示各分组各分类中的 1 个, 具体还是要下载所有结果来看。

2. 要选择哪些 ID 来进行可视化?

答：

可以对应云端记录表格中的各分组中各个分类的 TOP 几条目，也可以是自己感兴趣的想要展示的条目。

3. 为什么出来的图中少了 KEGG (或者 BP 或者 CC 或者 MF)，明明已经选了 GO+KEGG? (为什么出来的图里面某个分类只有 1 个或者没有?)

答：

GOKEGG 分组 可视化模块仅仅只是对已经完成 GOKEGG 分组 的数据进行可视化，如果对应保存的数据中就不存在某些类（没有富集出来某些类），可视化是不可能会有这些类的。而在 GOKEGG 分组 分析模块中，最终的结果表格只保留了满足较宽的阈值 ($p < 0.1$ 以及 $qvalue < 0.2$) 的结果，而不满足这一较宽阈值下的条目都会被过滤，如果整个类 (BP、CC、MF、KEGG) 都不满足这个阈值，那么最终的表格中就会缺少这个类。

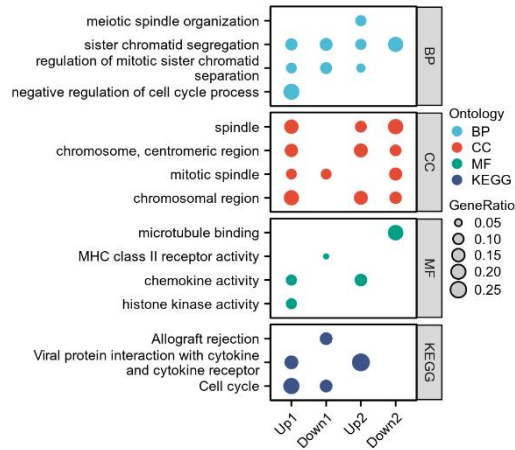
可以先检查 GOKEGG 分组 分析结果，在历史记录中找到保存的记录：



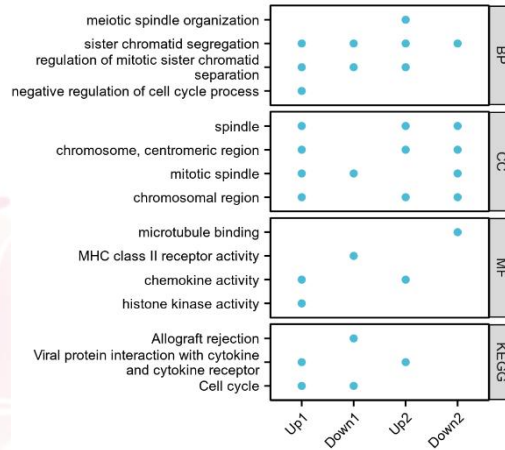
4. 如何修改展示的数据?

答：

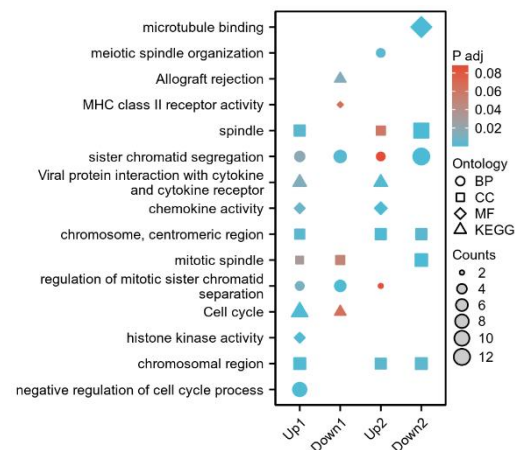
可以通过样式中的 颜色映射、大小映射、形状映射、分面 参数，下拉框选择 GOKEGG 分组 结果中不同的结果指标进行展示。



上图：颜色映射 - 类别(Ontology), 大小映射 - 基因比例(GeneRatio)



上图：颜色映射 - 无, 大小映射 - 无, 形状映射 - 无

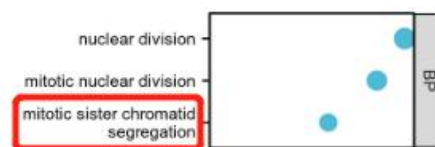


上图：颜色映射 - 校正后 p 值(padj), 大小映射 - 包含 ID 的数量, 形状映射 - 类别(Ontology), 分面映射 - 无

5. 名称太长了，如何修改？

答：

当类目名称太长时，可以在 [样式](#) 中的 [ID 展示](#) 参数中进行换行和修改。



上图：ID 展示 - 一行 20 长度



上图：ID 展示 - ID 号

6. 坐标轴有重叠，如何解决？

答：

- ✓ 增加 [图片宽度](#)
- ✓ 修改 轴标注旋转 角度

7. 能否上传自己的富集数据进行可视化？

答：

自己的富集分析的结果可以上传到基础绘图的点图或者一维柱状图模块进行可视化。

