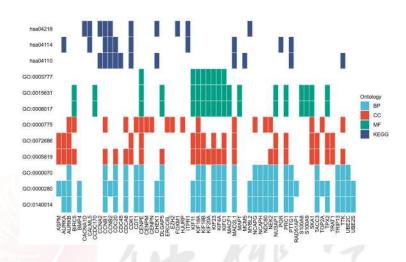


功能聚类 - GOKEGG 热图



网址: https://www.xiantao.love



更新时间: 2023.10.08



目录

基本概念
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
热图7
标题8
图注(Legend)9
坐标轴9
风格10
图片10
结果 <mark>说明</mark> 11
士要结里————————————————————————————————————
十 <u>五</u> 女
方法学13
如何引用14
常见问题15



基本概念

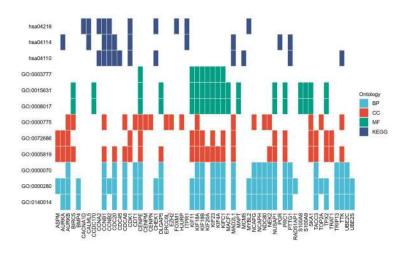
- ▶ 富集分析:简单而言,就是取一部分有功能注释的分子与所有有功能注释的分子去比较(超几何分布检验),确定这一部分分子中都涉及了哪些功能作用。注意:单独几个分子做富集分析意义并不大。
- ➤ GO (Gene Ontology,基因本体)数据库:把基因的功能分成了三类:生物过程(biological process, BP)、细胞组分(cellular component, CC)、分子功能(molecular function, MF)。利用 GO 数据库,可以得到目标基因在 CC, MF 和 BP 三个层面上有什么关联。
- ➤ KEGG (Kyoto Encyclopedia of Genes 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 富集分析后结果的可视化展示,可直观展示富集项中所包含的基因。

注意:模块需要先进行 GO|KEGG 富集分析 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

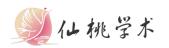


通过热图展示 GOKEGG 富集分析结果。

- 纵坐标为类目名称,横坐标为所选择类目所包含的基因信息。
- ➤ 图中方块显示与否,表示横坐标中的基因是否属于对应需要可视化的类目, 直观展示了类目与基因之间的关系。基因信息由【GOKEGG 富集分析】模 块的结果中获取,可下载【补充结果】中的 XLSX 文件查阅。



▶ 图中方块的颜色为所选择的 颜色映射 内容。(如上图,对应类别(Ontology)) 默认展示各类别的 top 几个结果(默认 满足校正后 p 值<0.05),分面是对应 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
GOKEGG	GOKEGG富集分析 @1.0	2023-02-06 16:14:19	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 富集分析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<mark>最近生成的分析</mark>记录。





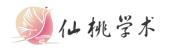
参数说明

(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 30 个类目,最少绘制 2 个类目。



样式

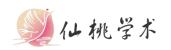


- ▶ ID 展示: ID 名称过长时,可以根据需要选择换行模式。可选择 全名(自动换行)、全名(一行 20 长度)、全名(一行 30 长度)、全名(一行 40 长度)、全名(一行 50 长度)、全名(一行 60 长度)、全名(一行 70 长度)、全名(一行 80 长度)、全名(不换行)、ID 号。
- ▶ 颜色映射:主要影响热图方块的颜色范围,默认无颜色映射,可选择 <u>类别</u> (Ontology)、无。

热图



- ▶ 填充色: 热图方块的填充色颜色选项,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: 热图方块的描边色颜色选项。默认白色, <mark>不受配色方案全局性修改</mark>。



▶ 描边粗细: 热图方块描边的粗细, 默认为 0.25pt。

▶ 不透明度: 热图方块的透明度。0 为完全透明, 1 为完全不透明。

标题

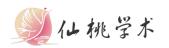


> 大标题: 大标题文本

➤ 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		~
是否展示		
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



▶ 是否显示 x 轴: 默认展示 x 轴

▶ 是否显示 y 轴: 默认展示 y 轴

➤ 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

风格		~
边框		
xy颠倒		
文字大小	6pt	~

▶ 外框:是否添加外框

> xy 颠倒:是否颠倒 xy 轴

> 文字大小: 针对图中所有文字整体的大小控制

图片



▶ 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



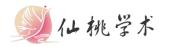
主要结果格式为图片格式,提供 PDF、TIFF 和 PPTX 格式下载,结果报告可以下载包括 pdf 以及说明文本的内容。



补充结果

前模块可视化所	选ID							
ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID
BP	GO:0140014	mitotic nuclear division	31/197	293/18800	2.61e-22	7.82e-19	6.71e-19	BMP4/CCNB1/CDC20/
BP	GO:0000280	nuclear division	35/197	446/18800	8.47e-21	1.27e-17	1.09e-17	BMP4/CCNB1/CDC20/.
BP	GO:0000070	mitotic sister chromatid se	24/197	171/18800	2.26e-20	1.83e-17	1.57e-17	CCNB1/CDC20/CENPE
CC	GO:0005819	spindle	26/203	402/19594	9.72e-14	2.88e-11	2.53e-11	BIRC5/CCNB1/CDK1/
CC	GO:0072686	mitotic spindle	17/203	160/19594	8.73e-13	1.29e-10	1.14e-10	CDK1/CENPE/KIF11/
CC	GO:0000775	chromosome, centromeric r	19/203	227/19594	2.81e-12	2.77e-10	2.44e-10	BIRC5/CCNB1/CENPE
MF	GO:0008017	microtubule binding	19/192	272/18410	7.14e-11	3.28e-08	3e-08	BIRC5/CENPE/KIF11
MF	GO:0015631	tubulin binding	19/192	376/18410	1.57e-08	3.6e-06	3.3e-06	BIRC5/CENPE/KIF11
MF	GO:0003777	microtubule motor activity	8/192	67/18410	4.66e-07	6.74e-05	6.18e-05	CENPE/KIF11/KIFC1
KEGG	hsa04110	Cell cycle	11/95	126/8164	1.99e-07	4.18e-05	3.98e-05	CCNA2/CCNB1/CDK1/.
KEGG	hsa04114	Oocyte meiosis	10/95	131/8164	2.55e-06	0.0003	0.0003	CCNB1/CDK1/CDC20/
KEGG	hsa04218	Cellular senescence	10/95	156/8164	1.22e-05	0.0009	0.0008	CACNA1D/CCNA2/CCN

此表格提供当前可视化的 GOKEGG 富集分析结果,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。



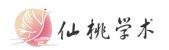


如何引用

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 3 个?

答:

在"ID 列表"选项卡中,有可视化 ID 的输入框:

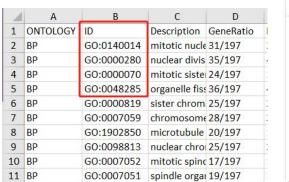


选项框内**默认**选择对应云端记录结果中前几个条目,可以在此处选择想要可视化的 ID。



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 30 个,至少展示 2 个。







2. 要选择哪些 ID 来进行可视化?

答:

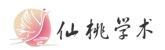
可以对应云端记录表格中的<u>各个分类的 TOP 几条目</u>,也可以是<u>自己感兴趣的想要展示的条目</u>。

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 可视化模块仅仅只是对已经完成 GOKEGG 富集分析的数据进行可视化,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 富集分析模块中,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 富集分析的结果,在历史记录中找到保存的记录:

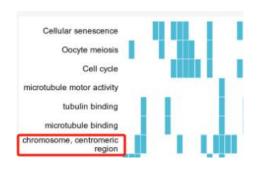




4. 名称太长了,如何修改?

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

5. 坐标轴有重叠,如何解决?

答:

- ① 增加图片高度或宽度
- ② 调整字体大小

6. 能否上传自己的富集数据进行可视化?



答:

由于涉及到类目相关的基因信息,暂时没有对应模块。

