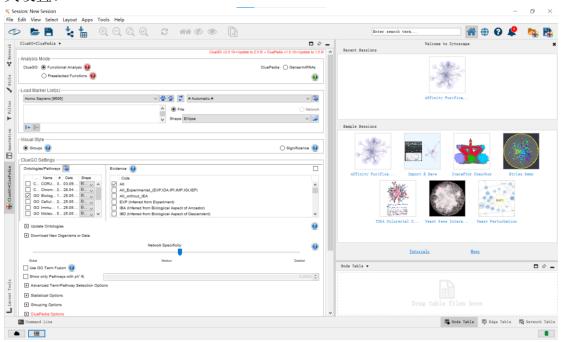
ClueGO 的功能探索和优势分析

ClueGO 是一个 Cytoscape 插件,它可以在一个功能分组的网络中展示 GO 富集结果,并可视化大簇基因的非冗余生物学术语。以下是对 ClueGO 的相关探索,所用 ClueGO 版本为 v2.5.10,Cytoscape 版本为 v3.9.1。

一、概览

下图为在 Cytoscape 中打开 ClueGO 时的界面, 左侧为 ClueGO 的控制界面, 可以对相关功能进行调控,包括选择分析模式,数据集上传,视图样式和一些相关设置。



二、功能介绍

下面以对 ClueGO 的示例文件"CluePediaExampleFile_Th1RelatedGenes.txt" 进行 GO 富集分析为例简单介绍 ClueGO 的功能。

1.分析模式(Analysis Mode)

根据不同目的选择不同的分析模式,如果既需要富集又需要可视化则选择 "Functional Analysis",如果完成富集分析只需要可视化则选择"Preselected Functions"。

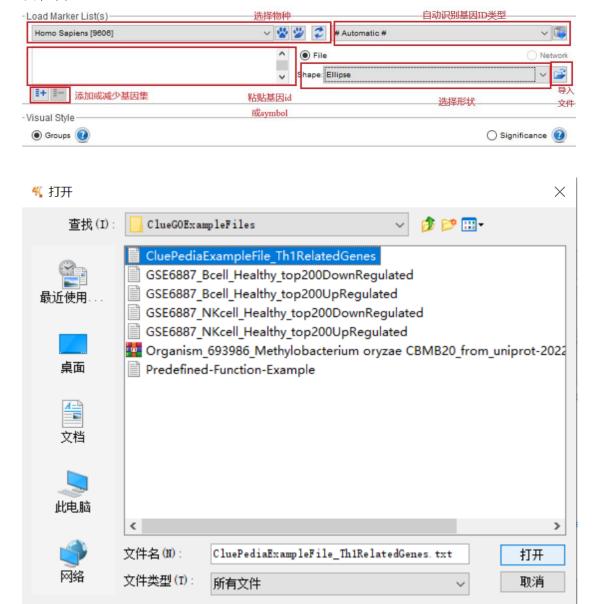


由于"CluePediaExampleFile Th1RelatedGenes.txt"文档中仅包含 17 个基因的

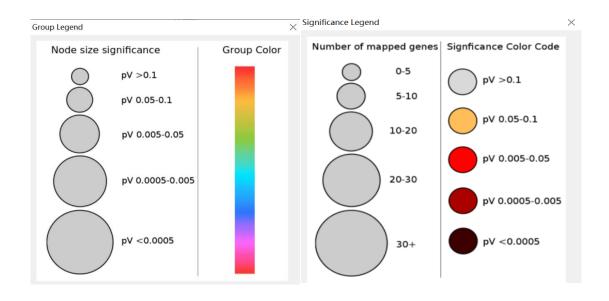
名字,因此勾选"Functional Analysis"(如上图)。

2.数据集上传和视图

物种选择"Homo Sapiens[9806]",点击导入文件按钮导入示例文件 "CluePediaExampleFile_Th1RelatedGenes.txt",17 个基因的名字被导入到左侧空 白框内。

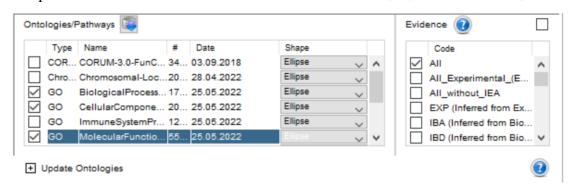


在"Visual Style"模块中勾选"Groups",则网络图中椭圆的大小由 pV 的值决定,pV 值越小,椭圆面积越大;而"Significance"则表示网络图中椭圆的大小由配对的基因数量决定(如下图)



3.设置富集标准

默认是生物学进程 BP(Biological Process),需增加常用细胞组分 CC(cellular component) 和分子功能 MF (Molecular Function),右侧默认设置不作更改。



下图为一些基本设置的介绍。其中网络的特异性(Network Specificity)功能也可以在"Group Options"设置中调节;示例文件为人类基因故不作处理;由于示例文件中仅有 17 个基因,通路本身不会很多,因此不勾选"Use GO Term Fusion"。



为显示具有统计学意义的通路,应勾选以下选项。

4.开始富集和结果展示

选择合适的网络布局,点击"Start"开始富集分析。

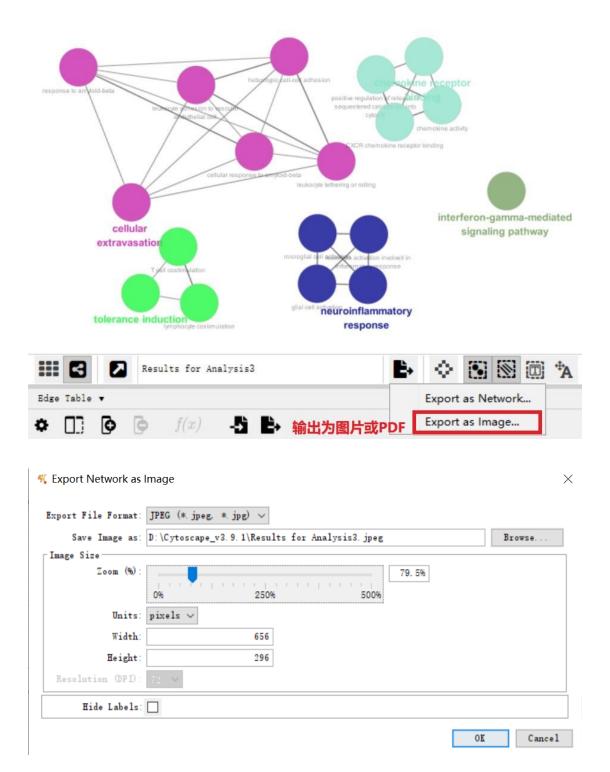


运行结束后,会跳出一个简述富集结果信息的提示框,包括物种,基因名字, 匹配的注释信息,重叠度和冗余度,节点数和节点的连接数等信息。



Cytoscape 的网络视图中展示的是富集结果的网络图。每个节点是代表性富集通路,节点的连线表示通路之间的共有的基因数,颜色表示该节点的富集情况分类(隶属于哪个功能组,颜色也可以和表格形式的的 ClueGO 结果对应,详见上文"数据集上传和视图"部分的介绍)。

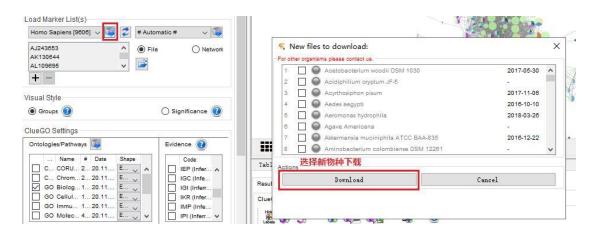
可将 GO 富集的结果导出为矢量图,并调整图片的宽度和长度。



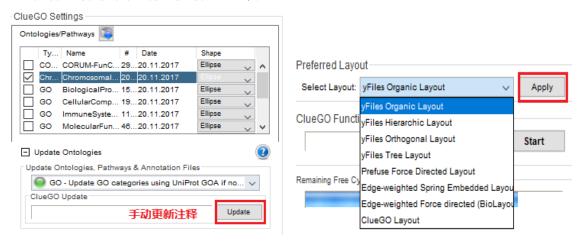
三、ClueGO 的优势分析

1. 物种丰富,并支持手动下载

本地有人类和小鼠可选,也可根据需要下载,支持近200个物种。

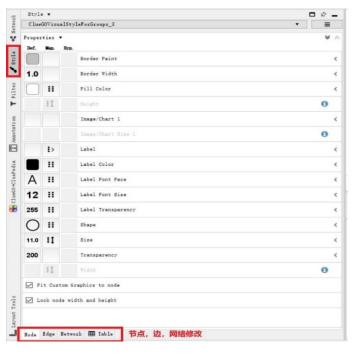


2. 物种注释手动更新和多种网络布局



由此也可以看出 ClueGO 的功能可以很好的与其他 Cytoscape 插件兼容(如 yFiles Layout Algorithm)

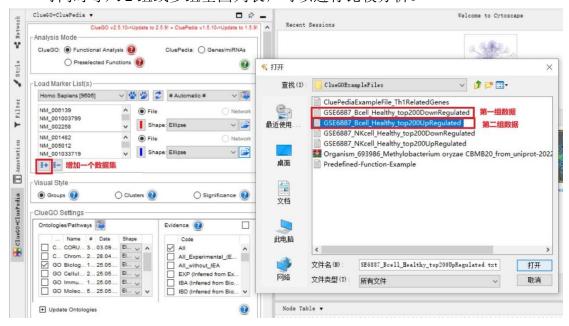
3. 自由网络设计



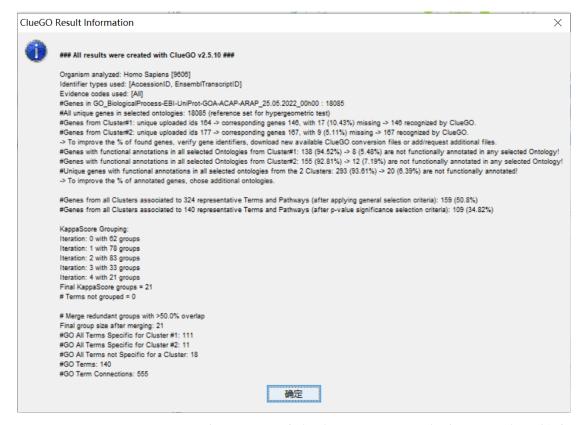
Style 中可以对网络的颜色, 节点, 网络等进行修改。

4. 多个数据集比较

可同时导入2组或多组基因列表,可以进行比较分析。

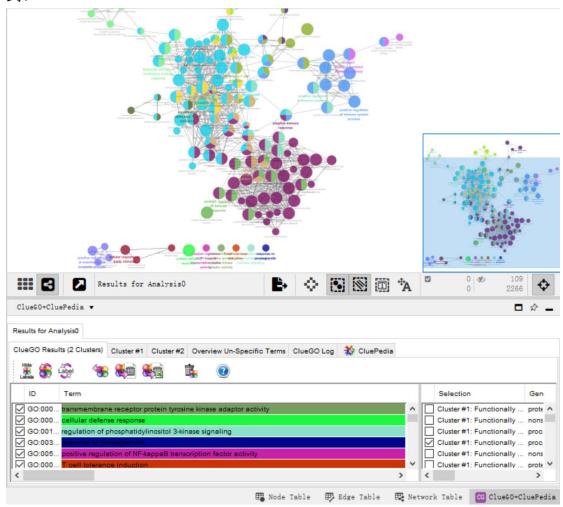


富集结果:首先跳出了一个简述富集结果信息的提示框,和单组分析类似,包括物种,基因名字,匹配的注释信息,重叠度和冗余度等信息,但是额外统计了了两组共有的基因的富集结果。



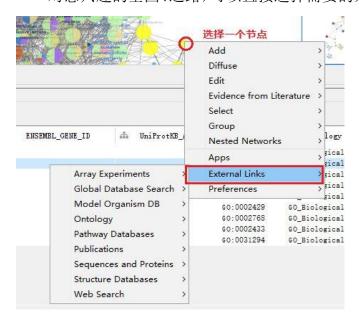
Cytoscape 的网络视图中展示的是富集结果的网络图。每个节点是代表性富集通路,节点的连线表示通路之间的共有的基因数,颜色表示该节点的富集情况

分类,而两组单独的,综合的,共有的富集通路等均可在 ClueGO results 详细看到。



5. 链接外源数据库

对感兴趣的基因、通路,可以直接选择需要的外源数据库查询获得更多信息。



四、ClueGO 合适的应用场景

- 1. 基因组学研究: 用于对基因组学数据进行功能注释和生物通路分析,以理解基因在不同生物过程中的功能和相互关系。
- 2. 蛋白质组学研究:适用于分析蛋白质相互作用网络,并寻找在特定生物学条件下显著富集的功能模块。
- 3. 药物开发研究: 用于解析药物与生物学通路之间的关联, 帮助理解药物的作用机制和潜在的副作用。
- 4. 疾病研究:可以帮助揭示基因或蛋白质与疾病之间的关联,促进对疾病机制的理解和潜在治疗靶点的发现。
- 5. 生物信息学分析:对于对大规模生物学数据进行功能注释和富集分析的研究人员,ClueGO可以提供可视化和结果解释的支持。