

用Cytoscape绘制PPI网络图并 使用cytohubba识别hub基因

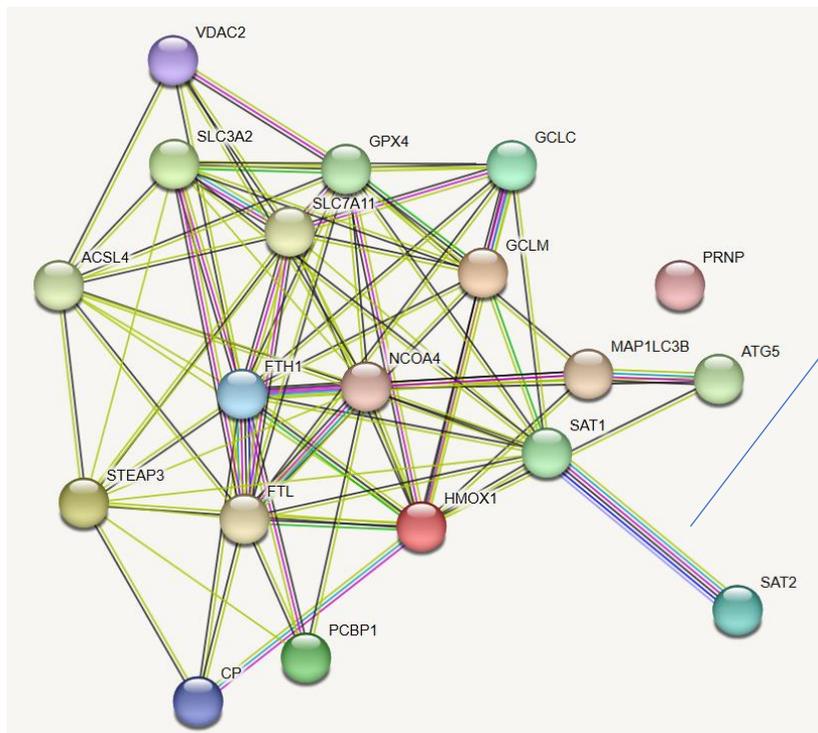
陈明杰 202411

Protein-Protein Interactions

- 蛋白质-蛋白质相互作用 (Protein-Protein Interaction, PPI) 是蛋白通过彼此之间的相互作用, 参与生物信号传递、基因表达调节、能量和物质代谢及细胞周期调控等生命过程的各个环节。
- 系统分析大量蛋白在生物系统中的相互作用关系, 对了解生物系统中蛋白质的工作原理, 了解疾病等特殊生理状态下生物信号和能量物质代谢的反应机制, 以及了解蛋白之间的功能联系都有重要意义。

Stringdb PPI数据库

- 蛋白质-蛋白质相互作用 (PPI) 数据库，它整合了已知和预测的蛋白质之间的功能性关联，包括直接的物理相互作用和间接的功能性关联。这些相互作用的数据来源于多种途径，包括计算预测、不同生物体间的信息转移、以及其他数据库汇总的相互作用。



不同颜色的线表示不同的证据

Neighborhood_on_chromosome: 在染色体上是邻居

Gene_fusion: 融合基因

Phylogenetic_cooccurrence: 系统发育上共现

Homology: 同源

Coexpression: 共表达

Experimentally determined interaction: 实验验证的

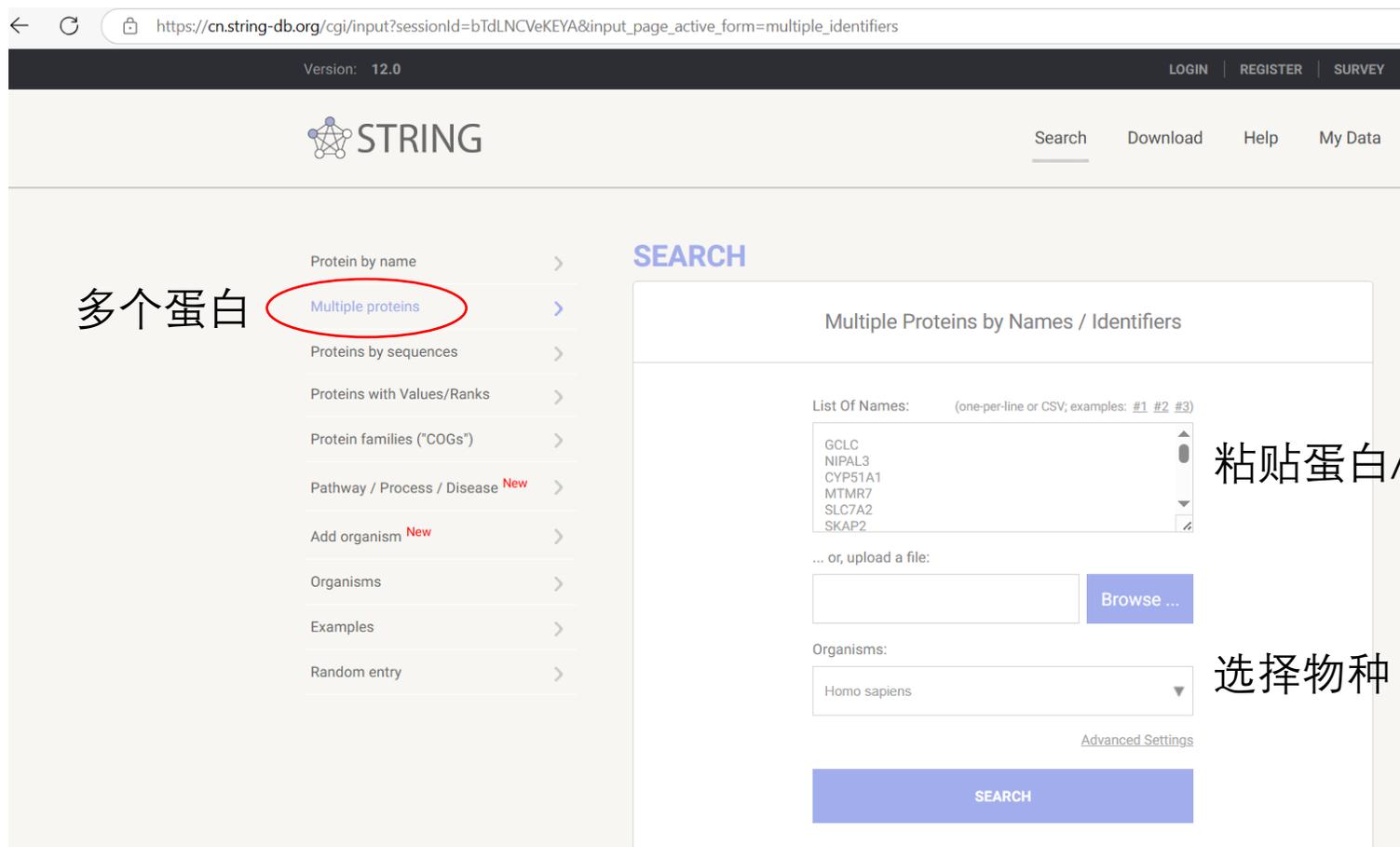
Database annotated: 其他数据库注释的

Automated text mining: 自动文本挖掘的

combined score
综合得分

1, 打开网站

https://string-db.org/



The screenshot shows the STRING database search page. On the left, a sidebar menu lists search options: 'Protein by name', 'Multiple proteins' (circled in red), 'Proteins by sequences', 'Proteins with Values/Ranks', 'Protein families ("COGs")', 'Pathway / Process / Disease New', 'Add organism New', 'Organisms', 'Examples', and 'Random entry'. The main content area is titled 'SEARCH' and 'Multiple Proteins by Names / Identifiers'. It features a text input field for 'List Of Names:' containing a list of gene symbols: GCLC, NIPAL3, CYP51A1, MTMR7, SLC7A2, and SKAP2. Below this is a file upload section with a 'Browse ...' button. An 'Organisms:' dropdown menu is set to 'Homo sapiens'. A 'SEARCH' button is at the bottom.

粘贴蛋白/基因, 一行一个

选择物种

2, 校验基因

Version: 12.0 LOGIN | REGISTER | SURVEY

 Search | Download | Help | My Data

The following proteins in *Homo sapiens* appear to match your input.
Please review the list, then click 'Continue' to proceed.

[<- BACK](#) [↓ MAPPING](#) [CONTINUE ->](#)

1001 query items showing page 1 of 51 • first • previous • next • last

1) 'GCLC':
 [GCLC](#) - Glutamate-cysteine ligase catalytic subunit.

2) 'NIPAL3':
 [NIPAL3](#) - NIPA-like protein 3; NIPA like domain containing 3.

3) 'CYP51A1':
 [CYP51A1](#) - Lanosterol 14-alpha demethylase; A cytochrome P450 monooxygenase involved in sterol biosynthesis. Catalyzes 14-alpha demethylation of lanosterol and 24,25- dihydrolanosterol likely through sequential oxidative conversion of 14- alpha methyl group to hydroxymethyl, then to carboxylaldehyde, followed by the formation of the delta 14,15 double bond in the sterol core and concomitant release of formic acid. Mechanistically, uses molecular oxygen inserting one oxygen atom into a substrate, and reducing the second into a water molecule, with two electrons provided by NADPH via cytochrome P45 [...]

4) 'MTMR7':
 [MTMR7](#) - Myotubularin-related protein 7; Phosphatase that specifically dephosphorylates phosphatidylinositol 3-phosphate (PtdIns(3)P) and inositol 1,3-bisphosphate (Ins(1,3)P2).
 [MTMR9](#) - Myotubularin-related protein 9; Acts as an adapter for myotubularin-related phosphatases. Increases lipid phosphatase MTMR6 catalytic activity, specifically towards phosphatidylinositol 3,5- bisphosphate and MTMR6 binding affinity for phosphorylated phosphatidylinositols. Positively regulates lipid phosphatase [MTMR7](#) catalytic activity (By similarity). Increases MTMR8 catalytic activity towards phosphatidylinositol 3- phosphate. The formation of the MTMR6-MTMR9 complex, stabilizes both MTMR6 and MTMR9 protein levels. Stabilizes MTMR8 protein levels. Plays a role in the late stages of ma [...] [a.k.a. ENSP00000221086, AJ297824, R-HSA-1483255]

一般不用管，基因名太旧的话，可以人工校验下

3, 网络图

太大，密密麻麻

Large Network Warning ...

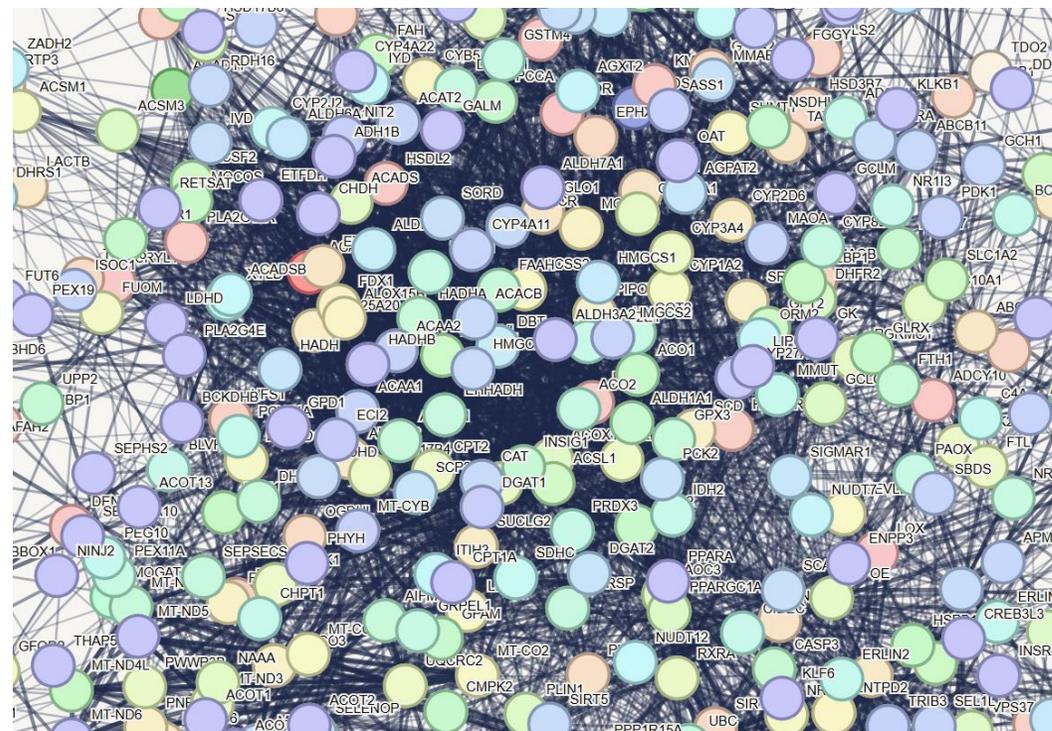
To keep your network manageable, STRING changed three settings:

- the nodes will have a simplified (not 3D) design
- previews of protein structures are not shown
- the network edges show interaction confidence only

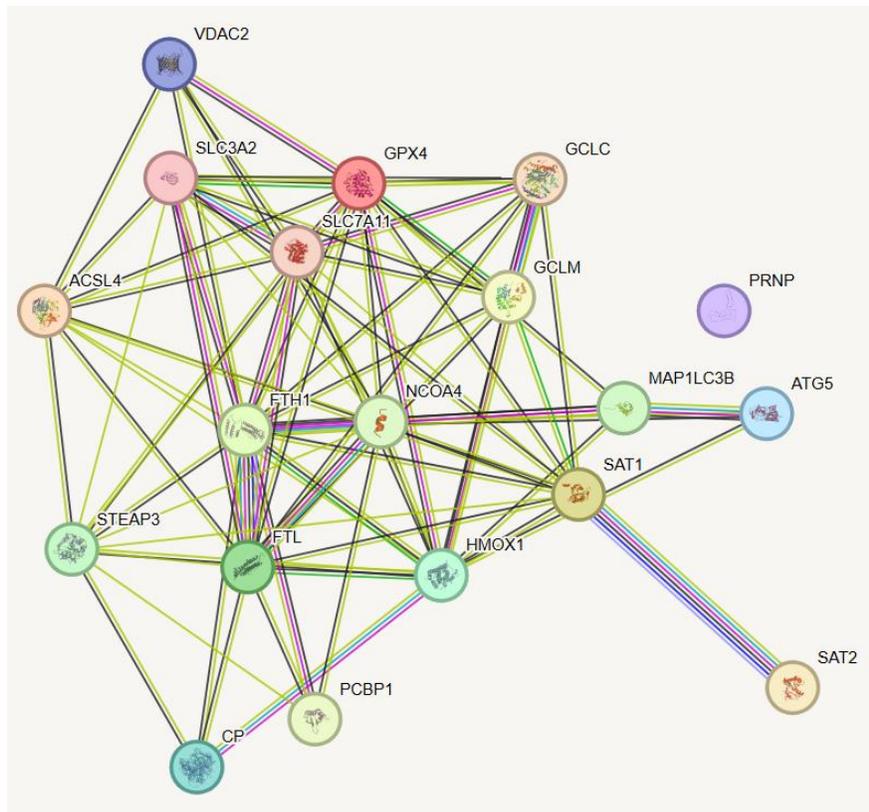
You can reverse these changes later, under the 'Settings' tab below the network.

<- BACK

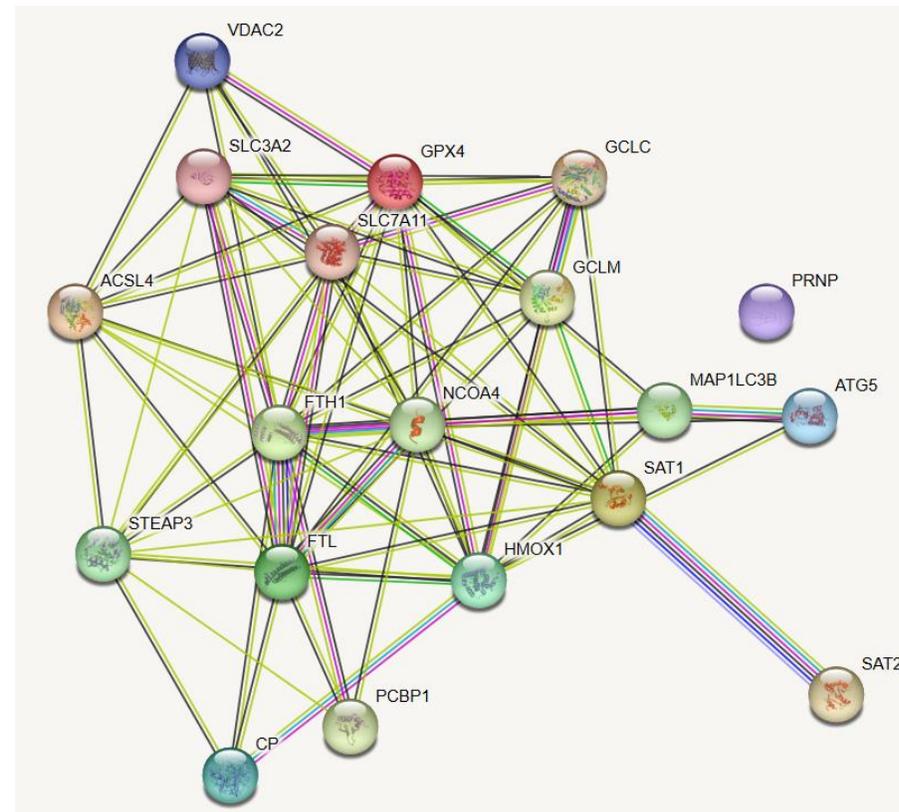
CONTINUE ->



v12扁平化



v12 3D效果



Basic Settings

Network type:

- full STRING network (the edges indicate both functional and physical protein associations)
- physical subnetwork (the edges indicate that the proteins are part of a physical complex)

UPDATE

meaning of network edges:

- evidence (line color indicates the type of interaction evidence)
- confidence (line thickness indicates the strength of data support)

active interaction sources:

来源

- Textmining
- Experiments
- Databases
- Co-expression
- Neighborhood
- Gene Fusion
- Co-occurrence

minimum required interaction score:

medium confidence (0.400)

置信度阈值

max number of interactors to show:

显示的最大

1st shell: - none / query proteins only -

相互作用节点数

2nd shell: - none -

Advanced Settings

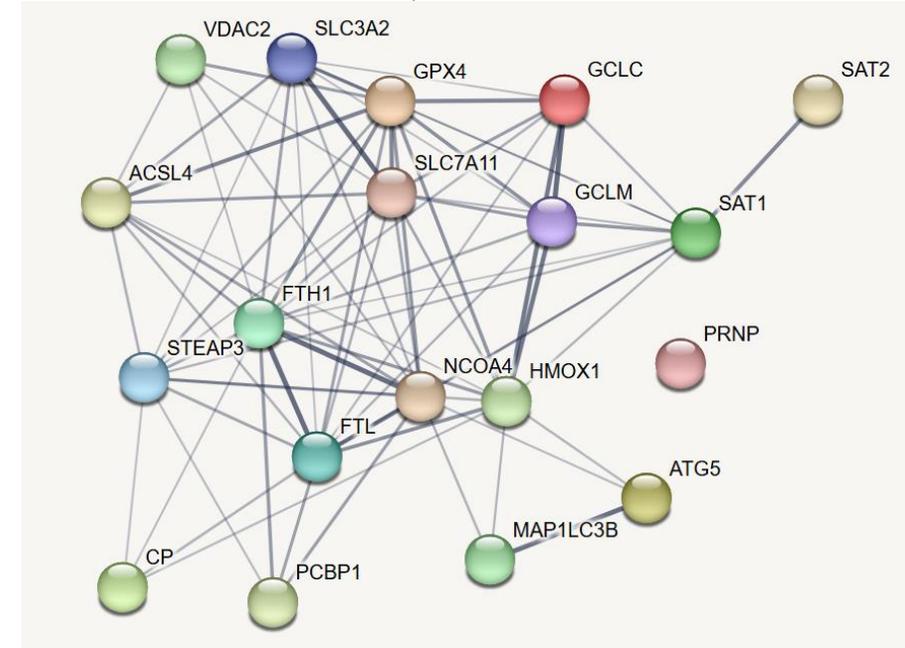
network display mode:

- static png (network is a simple bitmap image; not interactive)
- interactive svg (network is a scalable vector graphic [SVG]; interactive)

network display options:

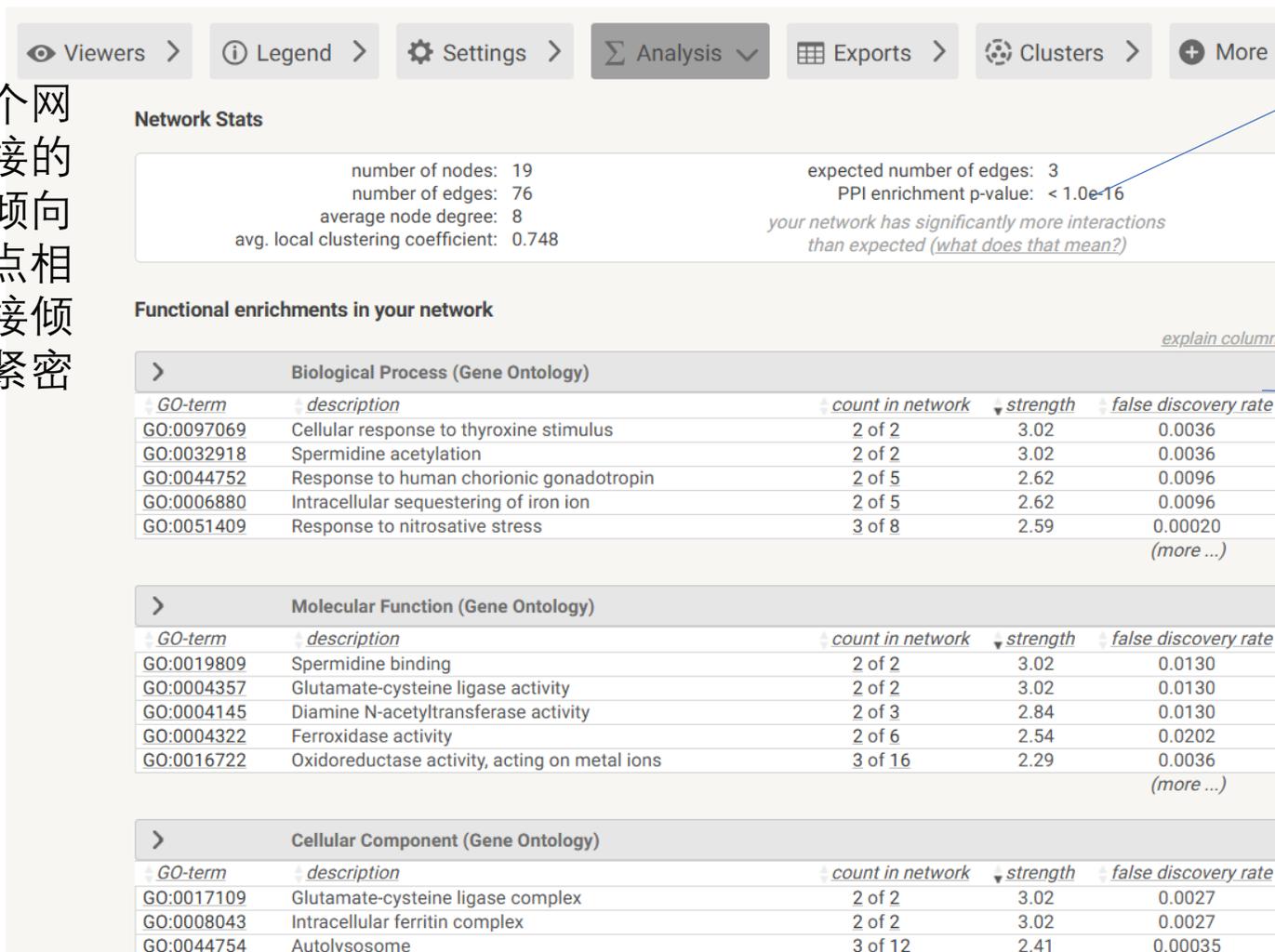
Enable node coloring mode

- enable 3D bubble design
- disable structure previews inside network bubbles
- center protein names on nodes
- show your query protein names
- hide disconnected nodes in the network
- hide protein names
- protein name font size



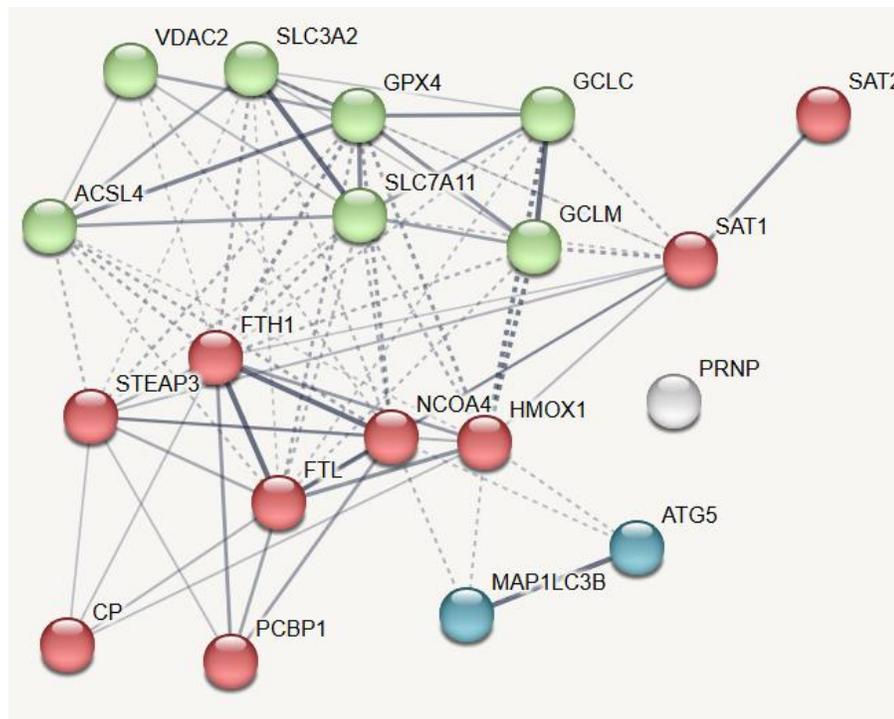
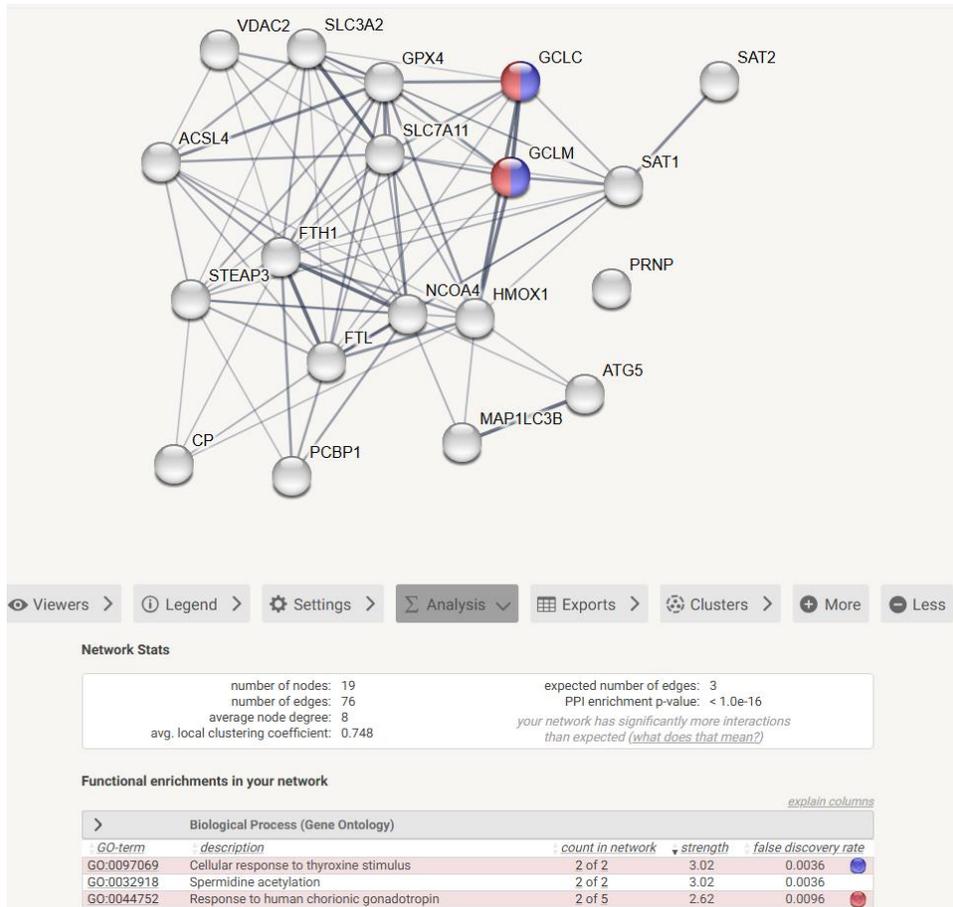
网络统计及功能富集分析

这些指标表明这个网络是一个高度连接的网络，其中节点倾向于与许多其他节点相连，并且这些连接倾向于在局部形成紧密的社区结构



这意味着你的蛋白质之间的相互作用比从基因组中随机抽取的具有相同大小和度分布的蛋白质集合预期的要多。这种富集表明，这些蛋白质至少在一定程度上是生物学上相连的，作为一个群体。

按照GO功能上色



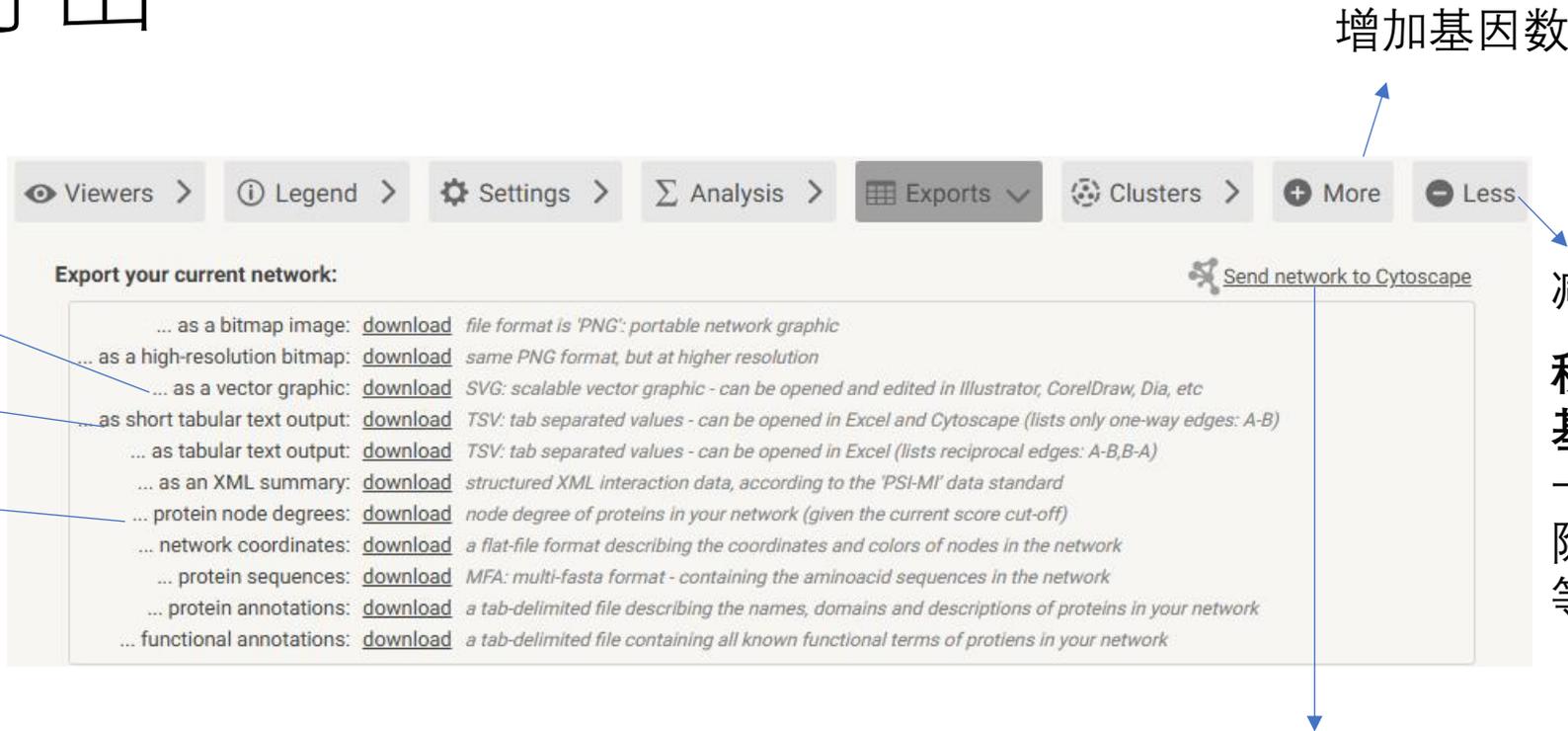
相同颜色表示
属于同一个
GO term

	A	B	C	D	E	F	G	H	I
1	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
2	hsa04216	Ferroptosis	19/19	42/8846	5.69368E-47	2.90378E-45	2.45727E-45	GCLC/GCLM/CP/	19
3	hsa00860	Porphyrim metabolism	4/19	46/8846	2.3431E-06	5.97491E-05	5.05617E-05	CP/FTL/HMOX1/	4
4	hsa00480	Glutathione metabolism	3/19	58/8846	0.000240595	0.003563442	0.003015503	GCLC/GCLM/GP/	3
5	hsa04978	Mineral absorption	3/19	61/8846	0.000279486	0.003563442	0.003015503	FTL/HMOX1/FTF	3
6	hsa04217	Necroptosis	3/19	159/8846	0.004468945	0.039610012	0.033519309	FTL/VDAC2/FTH	3
7	hsa00330	Arginine and proline metabolism	2/19	50/8846	0.005035412	0.039610012	0.033519309	SAT1/SAT2	2
8	hsa00270	Cysteine and methionine metabolism	2/19	52/8846	0.005436668	0.039610012	0.033519309	GCLC/GCLM	2
9	hsa04621	NOD-like receptor signaling pathway	3/19	189/8846	0.00722827	0.04608022	0.038994613	ATG5/MAP1LC3E	3
10	hsa00061	Fatty acid biosynthesis	1/19	18/8846	0.037999588	0.213831575	0.180951385	ACSL4	1
11	hsa01240	Biosynthesis of cofactors	2/19	153/8846	0.04192776	0.213831575	0.180951385	GCLC/GCLM	2

与clusterProfiler
结果比较

红色的是一样的

4, 导出



增加基因数

减少基因数

矢量图

节点文件

属性文件

移除孤立节点；
基于阈值的简化：设置一个连接度的阈值，删除低于该阈值的节点等等

将网络图发送给cytoscape, 保持string的风格

下载节点文件, 导入cytoscape, 自定义风格

Cytoscape简介

- 专注于网络可视化和分析。允许用户将生物分子交互网络与高通量基因表达数据和其他分子状态信息整合在一起，特别擅长于大规模蛋白质-蛋白质相互作用、蛋白质-DNA和遗传相互作用的分析。
- 1. 网络可视化：** Cytoscape能够可视化包括蛋白质相互作用网络、代谢网络和基因调控网络等在内的生物网络。用户可以通过导入网络数据文件（如SIF、XGMML等格式）构建和展示网络图。
 - 2. 网络分析：** 提供多种网络分析工具，如计算节点的中心性指标、社区结构等，以及支持多种网络布局算法来优化网络图的展示。
 - 3. 数据整合：** 用户可以将基因表达数据、蛋白质功能注释等与网络数据集整合，在网络上展示附加信息。
 - 4. 模块和通路分析：** 通过插件进行更高级的分析，如寻找网络中的功能模块、通路分析等。
 - 5. 网络互动和分享：** 支持用户对网络图进行交互操作，并能保存网络图为图像或特定格式的文件，方便共享。
 - 6. 插件支持：** Cytoscape具有丰富的插件生态系统，用户可以根据需要选择和安装插件，以扩展Cytoscape的功能

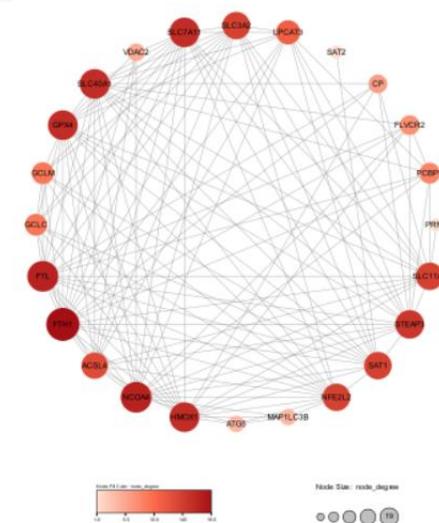
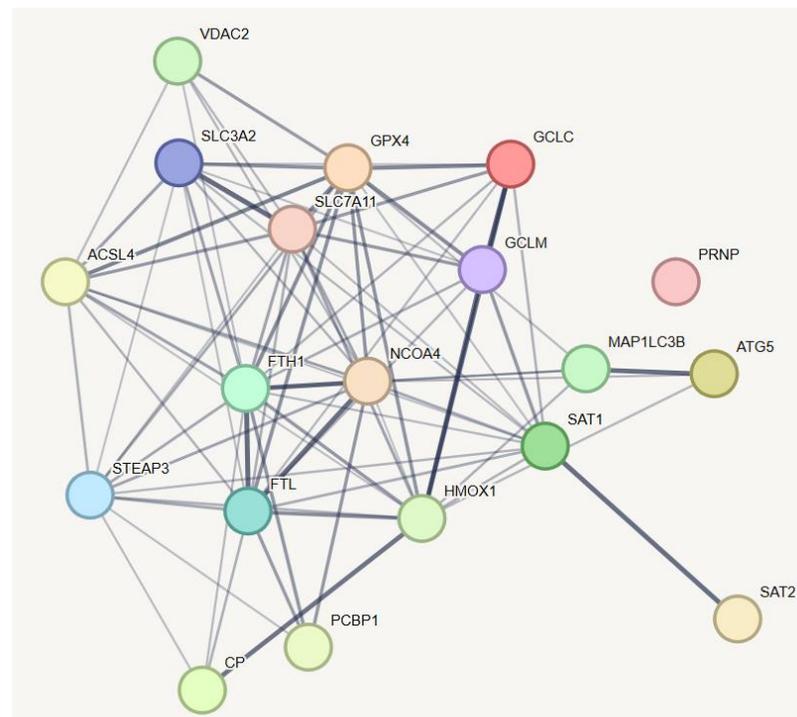
下载cytoscape并安装

<https://cytoscape.org/>



基本概念

- 节点 (Node, 例如基因、蛋白等)
 - 颜色 (单色, 渐变色)
 - 大小 (单一, 渐变大小)
 - 形状
 - 字体等
- 边 (edge, 例如靶向, 相关系数等)
 - 粗细 (单一, 渐变粗细)
 - 线型 (虚线, 实线)
 - 透明度 (美观度)
 - 箭头 (前, 后) 等
- 布局
 - preferred
 - 圆形
 - Grid
 - 节点排序
 - 缩放等



软件布局介绍

Session: New Session

菜单 File Edit View Select Layout Apps Tools Help

边文件

节点属性文件

可视化区

需选中

[全局][映射][个别]

缩放

节点, 边属性调整区

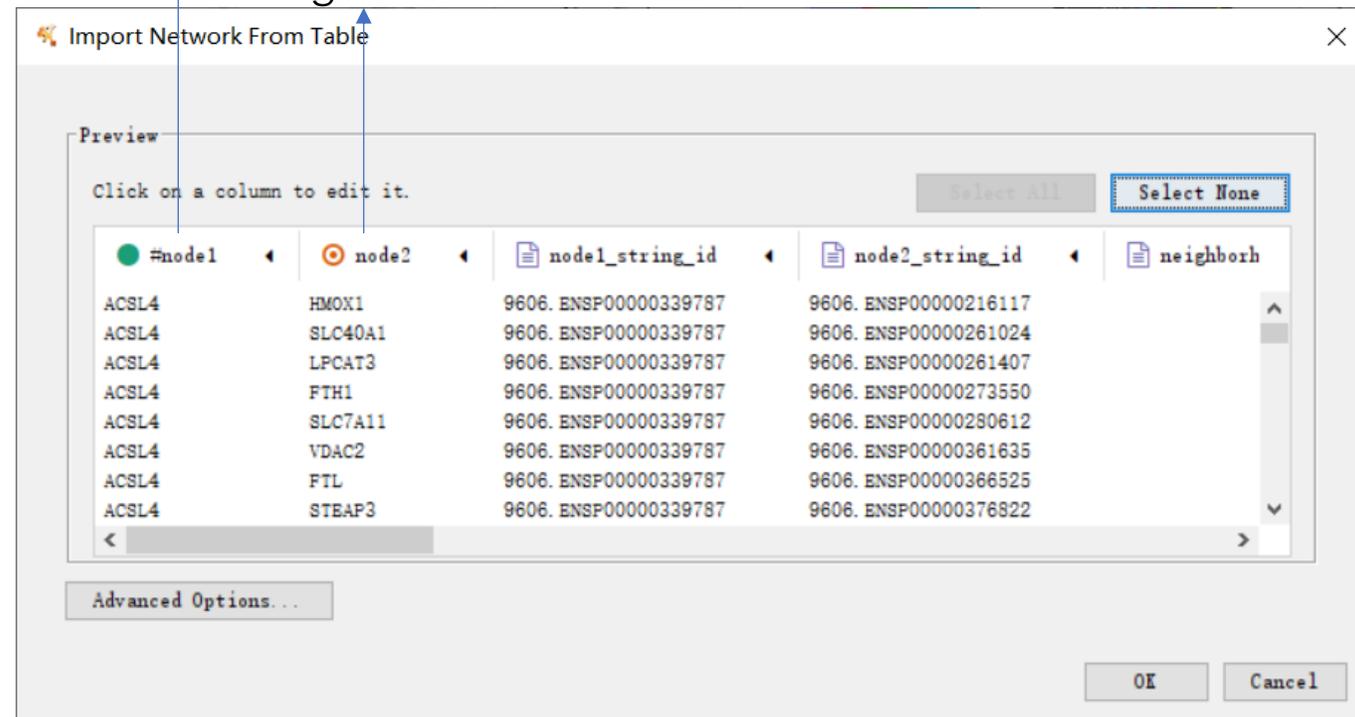
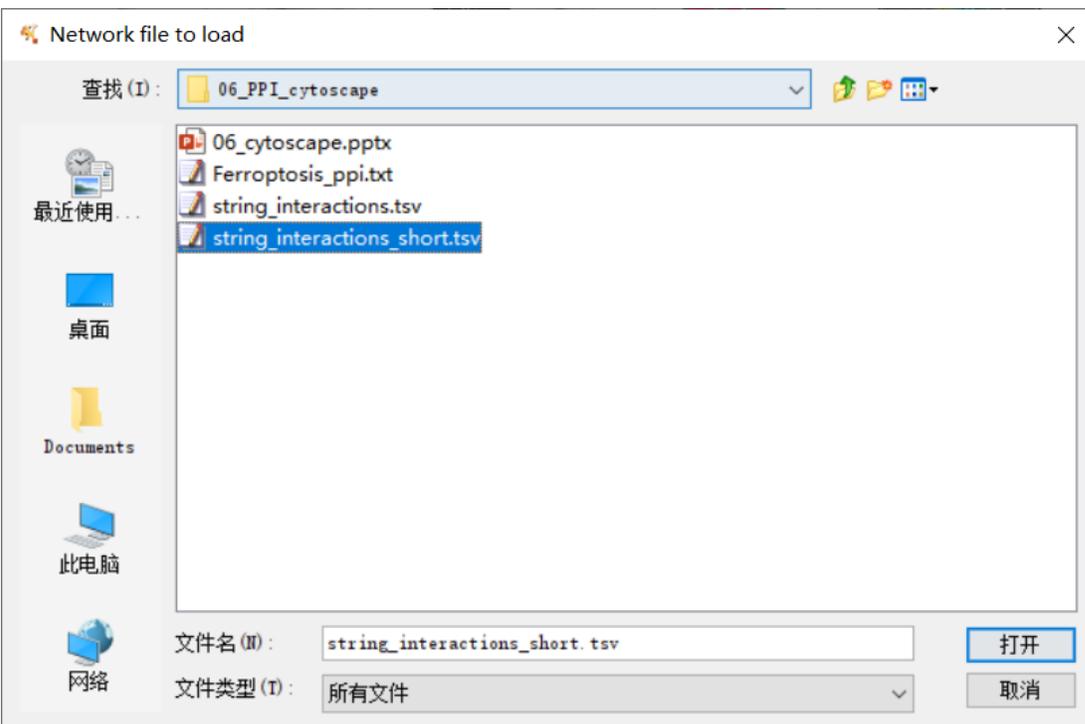
导出图片

shared name	name	identifier	node_degree	degree layout
ACSL4	ACSL4	9606. ENSP0000033...	13	13

导入边文件

Source node

Target node



带单

调整样式、插件等

Session: New Session

File Edit View Select Layout Apps Tools Help

Enter search term...

Network

Species: Homo sapiens X Giardia intestinalis

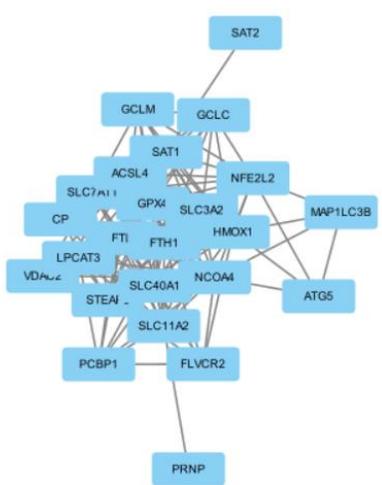
1 of 1 Network selected

string_interactions_short.tsv

string_interactions_short.tsv 24 133

网络导航栏

查看网络



string_interactions_short.tsv

Node Table

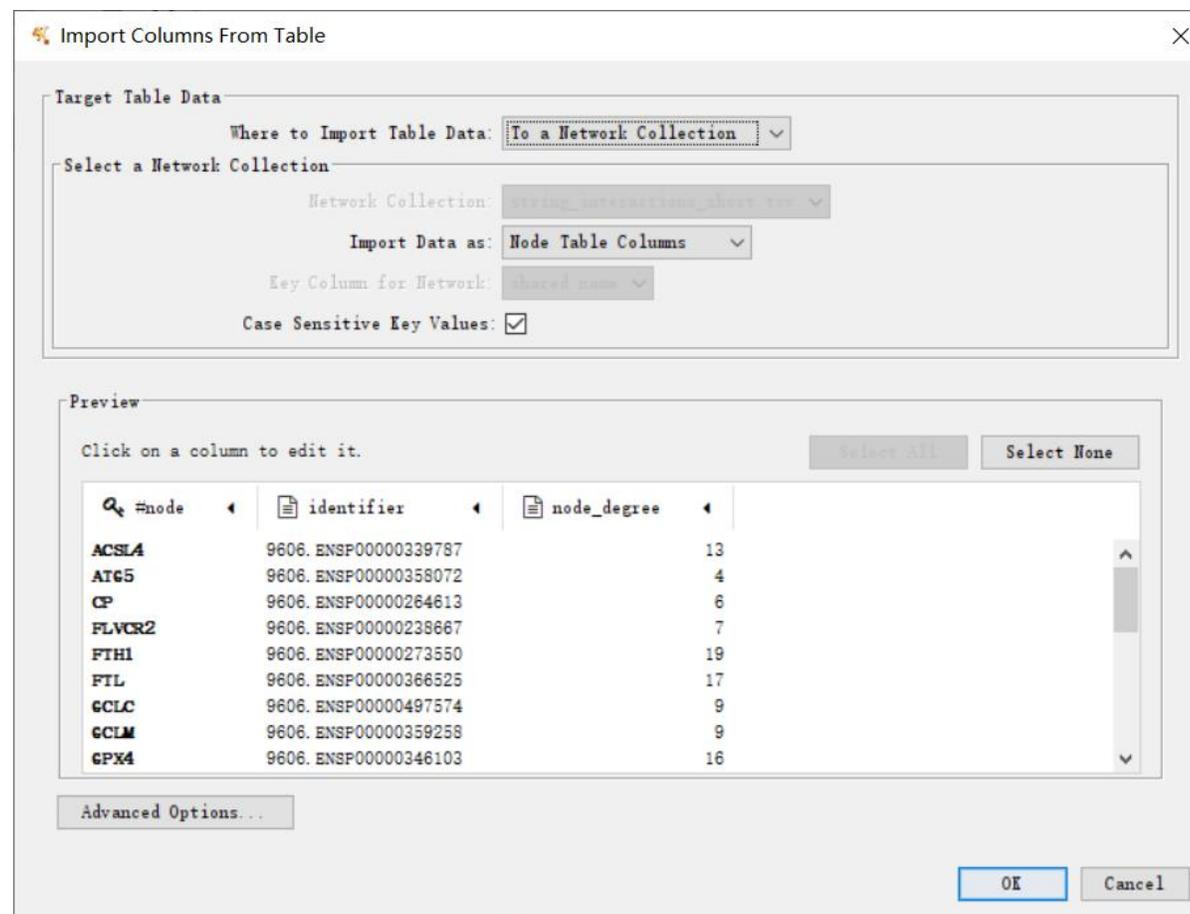
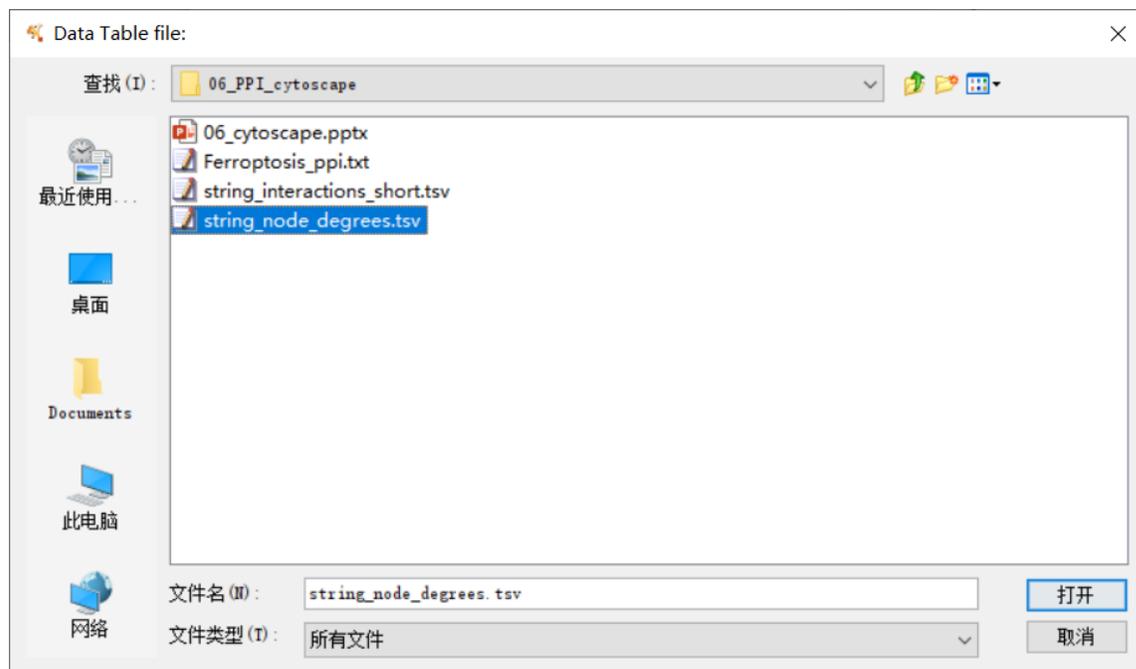
shared name	name
ACSL4	ACSL4
HMOX1	HMOX1
SLC40A1	SLC40A1
LPCAT3	LPCAT3
FTH1	FTH1

网络数据

Command Line

Node Table Edge Table Network Table Enrichment Table

导入属性文件



优化-颜色、形状、字体大小等

Session: New Session

File Edit View Select Layout Apps Tools Help

Enter search term...

Style: default

Properties:

Def.	Map.	By.	Property	Value
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Border Paint	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Border Width	0.0
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Fill Color	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Height	35.0
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Image/Chart 1	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Label	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Label Color	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Label Font Size	12
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Shape	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Size	<
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Transparency	255
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Width	75.0

Lock node width and height

Node Edge Network Column

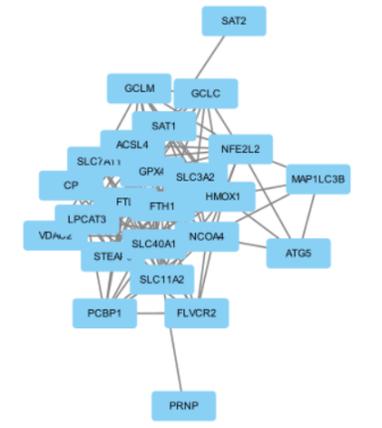
string_interactions_short.tsv

Node Table

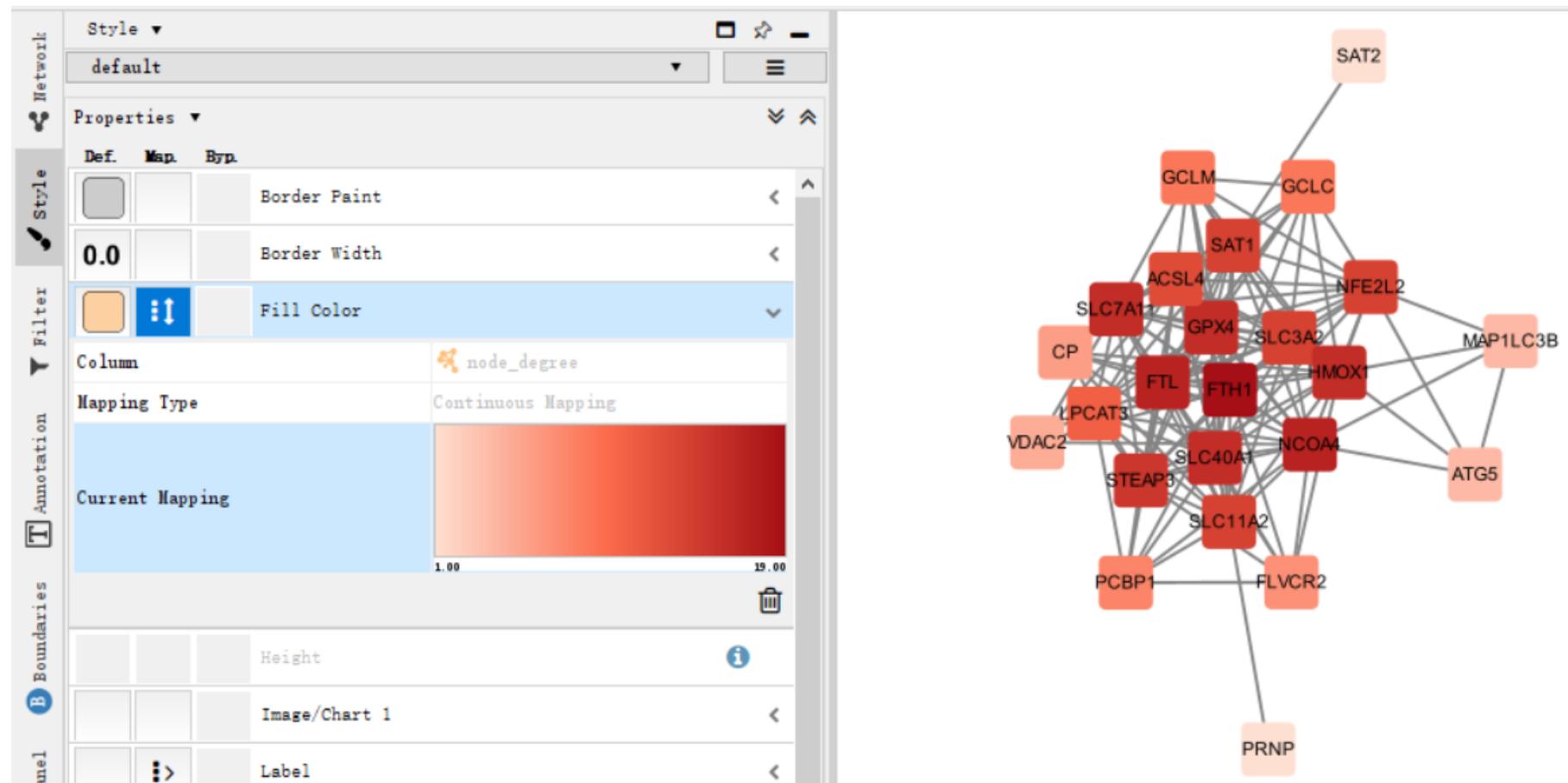
shared name	name	identifier	node_degree
ACSL4	ACSL4	9606.ENSP0000033...	13
HMOX1	HMOX1	9606.ENSP0000021...	16
SLC40A1	SLC40A1	9606.ENSP0000028...	16
LPCAT3	LPCAT3	9606.ENSP0000028...	11
FTH1	FTH1	9606.ENSP0000027...	19

Command Line

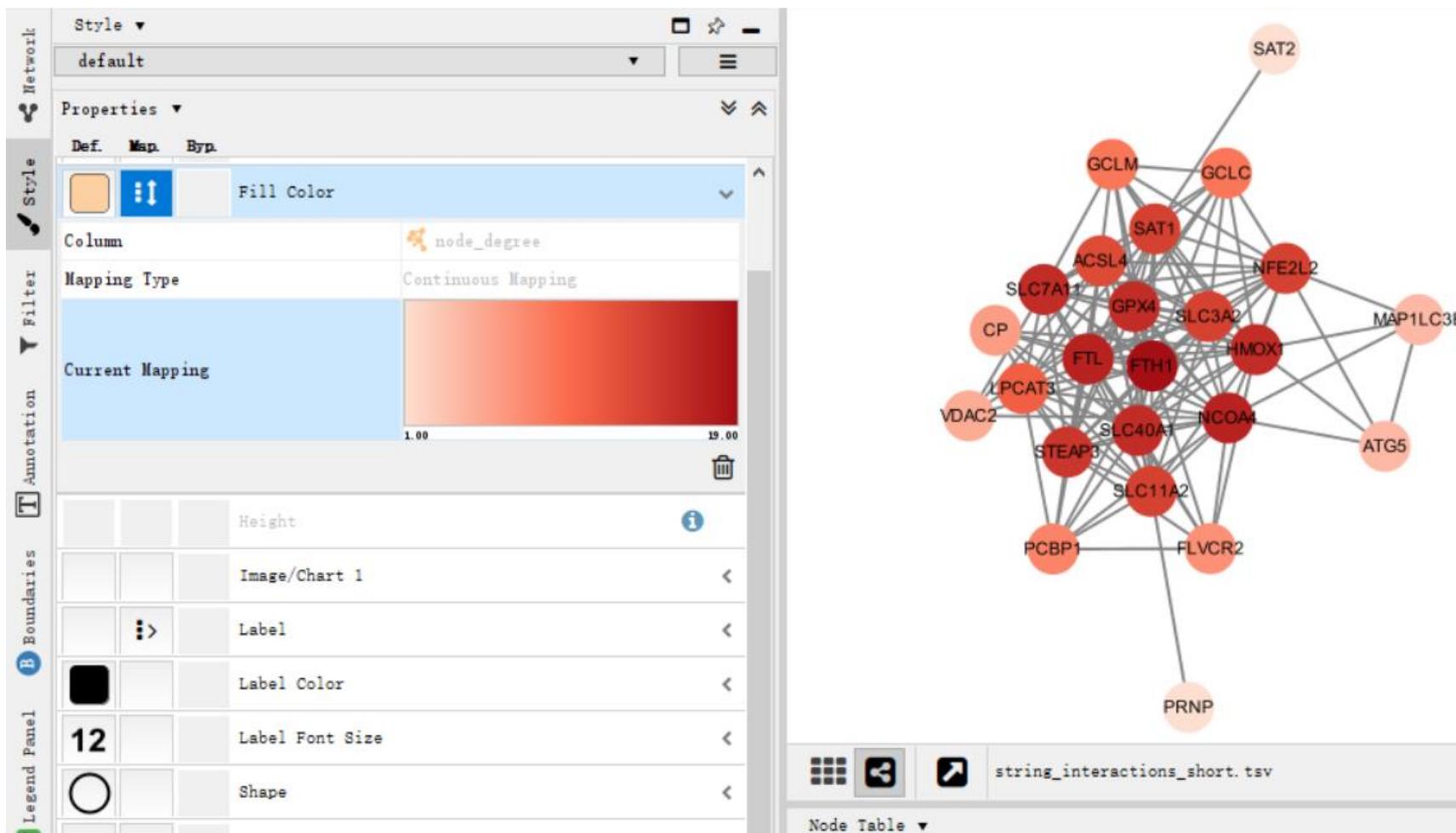
Node Table Edge Table Network Table Enrichment Table



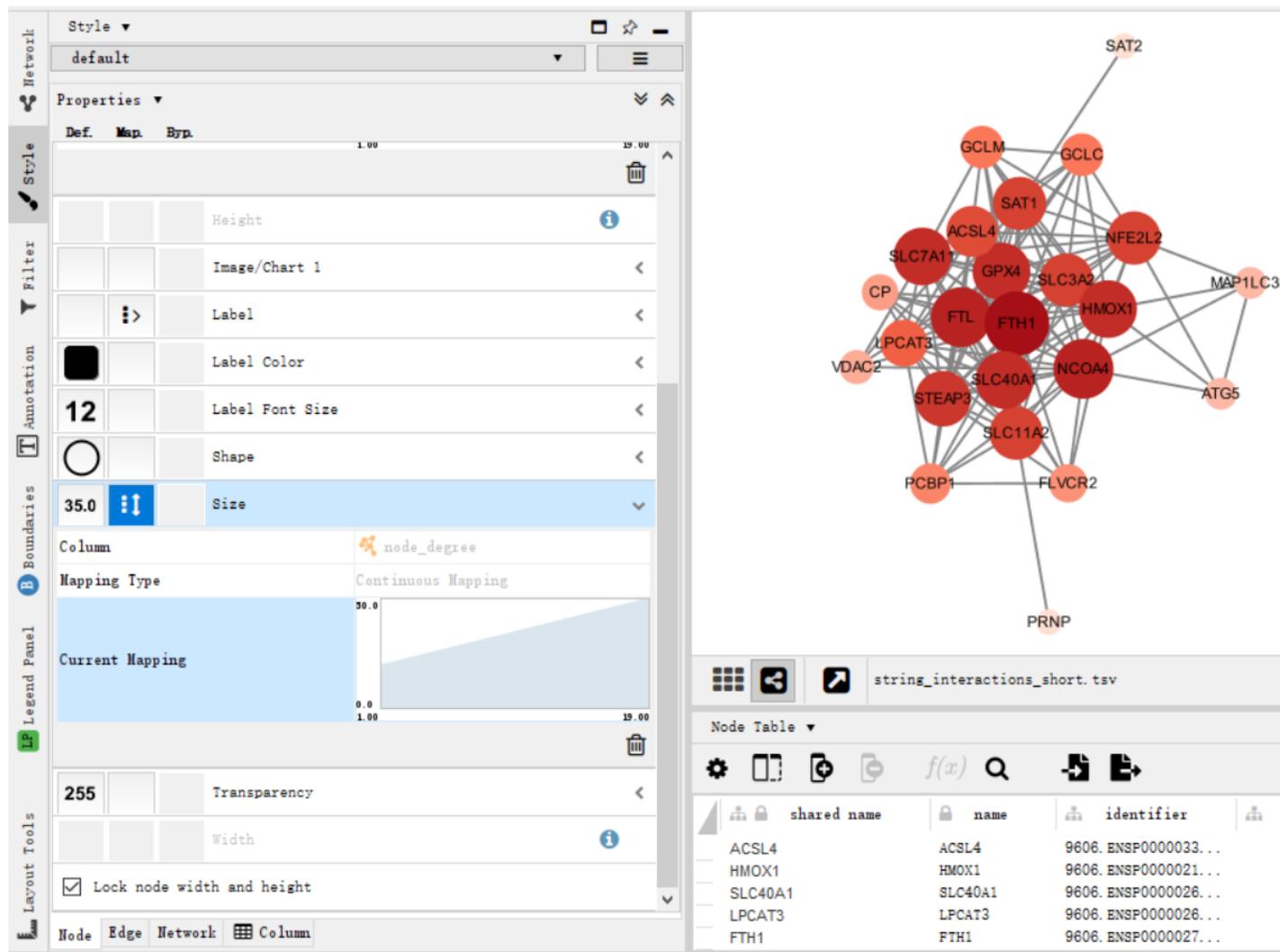
颜色-渐变



形状-圆形



节点大小 - 渐变



The screenshot displays a network visualization software interface. On the left, a 'Style' panel is open, showing the 'Size' property for nodes. The 'Size' is set to 35.0, and the 'Mapping Type' is 'Continuous Mapping'. A 'Current Mapping' graph shows a linear relationship between node degree (0.0 to 15.0) and size (0.0 to 50.0). Below the mapping graph, the 'Transparency' is set to 255. The 'Node Table' at the bottom lists several nodes with their identifiers.

Node Table

shared name	name	identifier	nc
ACSL4	ACSL4	9606. ENSP0000033...	
HMOX1	HMOX1	9606. ENSP0000021...	
SLC40A1	SLC40A1	9606. ENSP0000026...	
LPCAT3	LPCAT3	9606. ENSP0000026...	
FTH1	FTH1	9606. ENSP0000027...	

边

Style ▾

default ▾

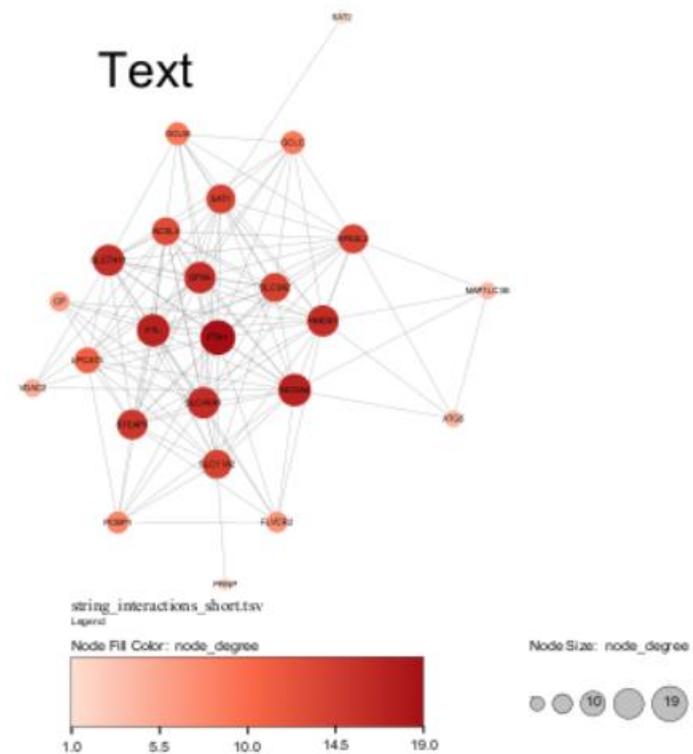
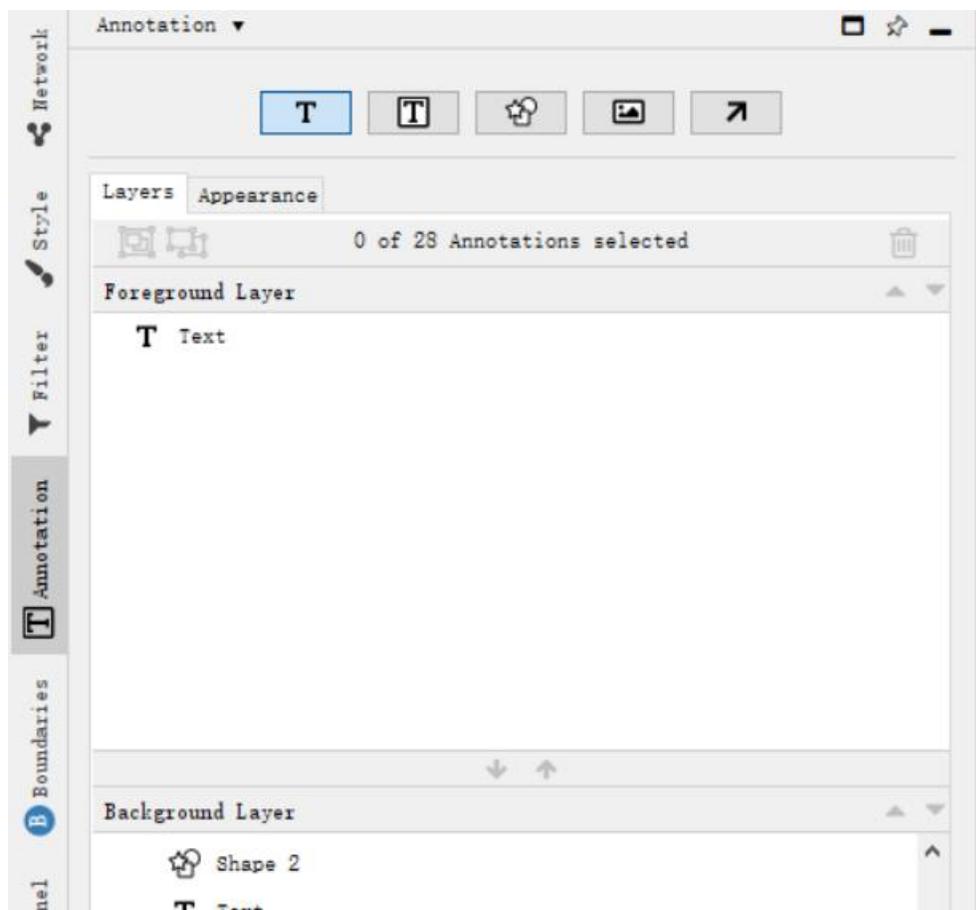
Properties ▾

Def.	Map.	By.	
			Color (Unselected) ⓘ
			Label <
			Label Color <
10			Label Font Size <
—			Line Type <
None			Source Arrow Shape <
			Source Arrow Unselected Paint <
			Stroke Color (Unselected) <
None			Target Arrow Shape <
			Target Arrow Unselected Paint <
150			Transparency <
1.0			Width <
<input type="checkbox"/> Edge color to arrows			

string_interactions_short.tsv

Node Table ▾

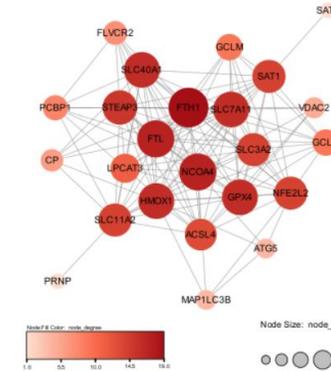
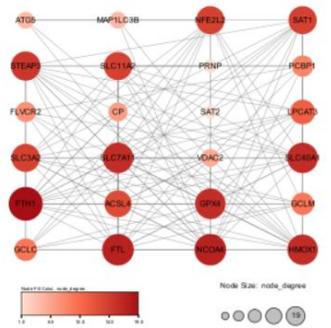
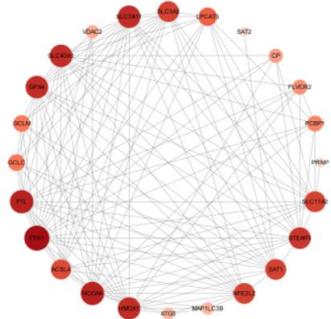
添加文字和Legend



布局及导出

The screenshot shows the Cytoscape interface with a network graph. A menu is open, showing layout options. The 'Attribute Circle Layout' is selected, and a sub-menu is open for 'Selected Nodes Only', listing attributes: shared name, name, selected, identifier, and node_degree. The 'Node Table' is visible at the bottom, showing a table with columns: shared name, name, identifier, and node_degree.

shared name	name	identifier	node_degree
FTH1	9606. ENSP0000027...		19
PRNP	PRNP	9606. ENSP0000039...	1
GCLC	GCLC	9606. ENSP0000049...	9
MAP1LC3B	MAP1LC3B	9606. ENSP0000049...	4
ACSL4	ACSL4	9606. ENSP0000033...	13



File → save session 保存网络文件

导出pdf图片

提问/实操：同心圆网络图布局

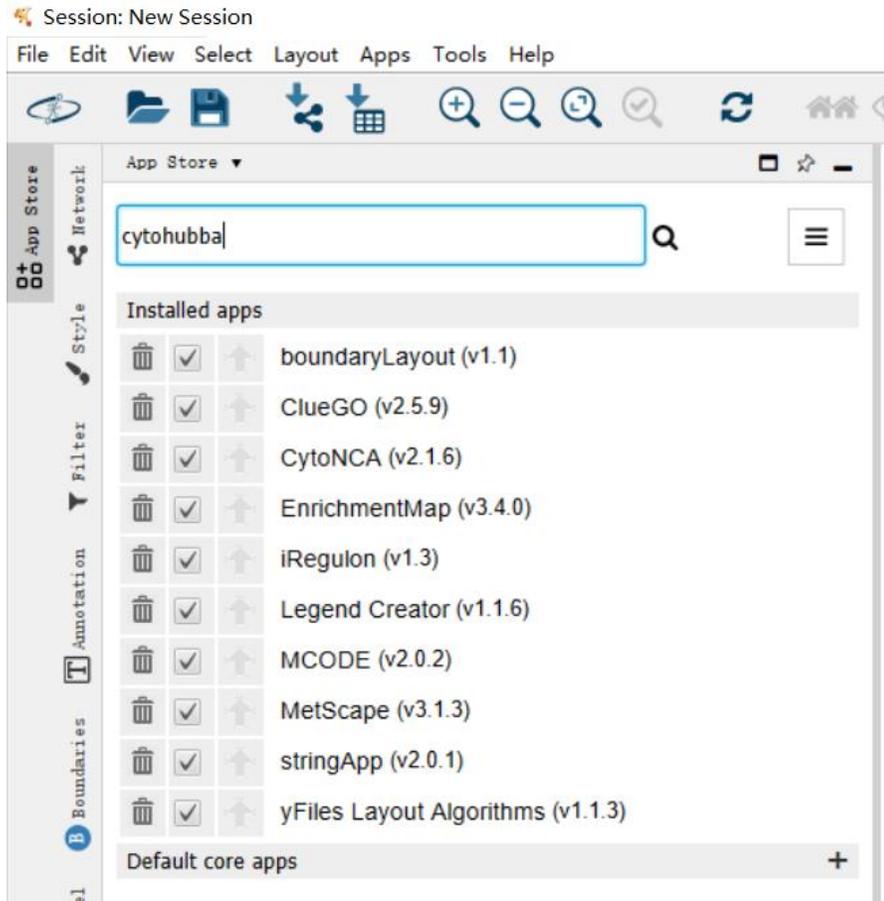
其他蛋白蛋白相互作用数据库

- 1. MINT (Molecular INTERaction Database):** MINT是一个公开的、经过专家策划的、实验验证的蛋白质-蛋白质相互作用数据库。它遵循国际PSI-MI标准进行注释，并包含多个元数据和其他资源的交叉引用。MINT是ELIXIR核心数据资源的一部分。
- 2. IntAct Molecular Interaction Database:** IntAct是一个提供分子间相互作用数据的数据库，这些数据通过PSICQUIC web服务可以以编程方式访问。
- 3. BioGRID (Biological General Repository for Interaction Datasets):** BioGRID是一个公共的蛋白质相互作用数据库，包含来自多种生物体的遗传、蛋白质和化学相互作用数据。
- 4. DIP (Database of Interacting Proteins):** DIP数据库提供了关于蛋白质相互作用的详细信息，包括文献引用和相互作用的证据。
- 5. HPRRD (Human Protein Reference Database):** HPRRD是一个详细的数据库，提供人类蛋白质的表达、PTMs（翻译后修饰）和蛋白质相互作用的信息。
- 6. MIPS (Munich Information Center for Protein Sequences):** MIPS提供了关于蛋白质相互作用的数据，以及蛋白质的功能分类和通路信息。
- 7. iRefIndex:** iRefIndex是一个参考数据库，整合了多个PPI数据源，并提供了一个非冗余的蛋白质相互作用数据集。
- 8. HitPredict:** HitPredict是一个综合资源网站，提供实验确定的蛋白质相互作用和可靠性评分。
- 9. HINT:** HINT数据库整合了来自多个PPI数据库的数据，支持单个或批量查询。
- 10. mPPI:** mPPI是一个数据库扩展，用于以一对多的方式可视化结构相互作用组。

Cytohubba插件

- CytoHubba是一个用于网络分析的插件，目的是识别生物网络中的关键节点（例如蛋白质-蛋白质相互作用网络中的枢纽蛋白质）和子网络。
- **多种分析方法：** CytoHubba提供11种不同的拓扑分析方法，包括度（Degree）、边渗透组件（Edge Percolated Component）、最大邻域组件（Maximum Neighborhood Component）、最大邻域组件的密度（Density of Maximum Neighborhood Component）、最大团中心性（Maximal Clique Centrality）以及基于最短路径的六个中心性指标（Bottleneck, EcCentricity, Closeness, Radiality, Betweenness, Stress）。
- **性能优越：** 新提出的方法MCC（Maximal Clique Centrality）在预测酵母蛋白质相互作用网络中的关键蛋白质方面表现更好，能够更精确地捕捉到高度和低度蛋白质中的关键蛋白质。
- **用户界面友好：** CytoHubba提供了一个简单的交互界面，用户可以通过这个界面分析网络，并选择特定的节点进行分析。
- **子网络提取：** CytoHubba还提供了一个功能强大的控制面板，帮助研究人员搜索和探索网络，并提取感兴趣的子网络。

安装cytohubba



Cytoscape App Store

Submit an App Search the App Store Sign In

cytoHubba
Predicts and explores important nodes and subnetworks in a given network by several topological algorithms.

★★★★★ (168) 137726 downloads | citations | discussions 3.0+

Details **Release History**

Categories: clustering, Connection sub-graph discovery, graph analysis, network analysis, Protein complex, social network analysis, Topology



Explore important nodes/hubs and fragile motifs in an interactive network by several topological algorithms including Degree, Edge Percolated Component (EPC), Maximum Neighborhood Component (MNC), Density of Maximum Neighborhood Component (DMNC), Maximal Clique Centrality (MCC) and centralities based on shortest paths, such as Bottleneck (BN), Eccentricity

CYTOSCAPE 3

Install

Or, for JAR file, click [here](#) to download

Version 0.1
Released 4 Jan 2017
Works with Cytoscape 3.0
Download Stats [Click here](#)

cytoscape v3.10.0

RESOURCES

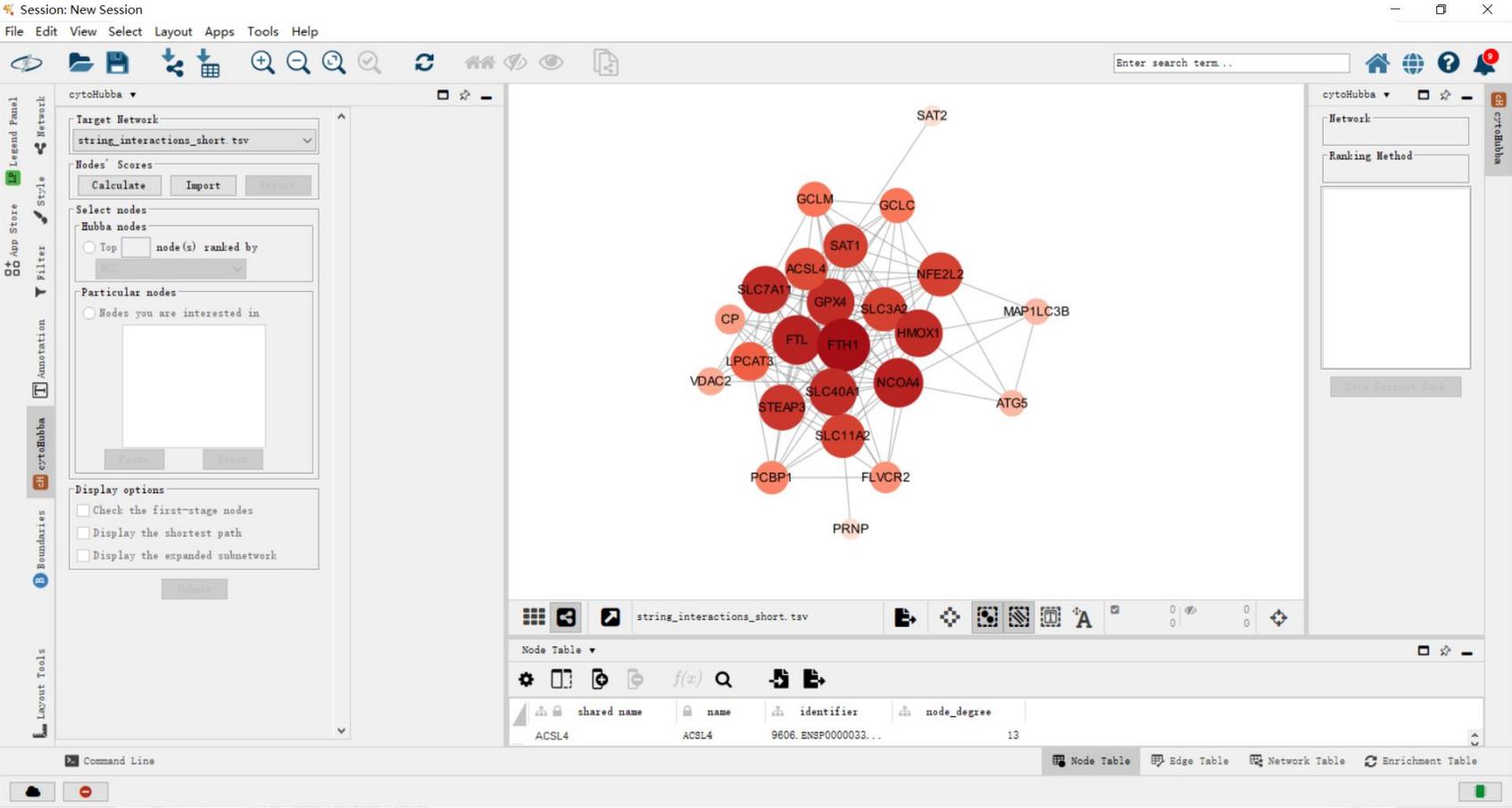
- Ask a question
- Search BioStars

CYTOSCAPE 3

Installed

Or, for JAR file, click [here](#) to download

使用



The screenshot displays the cytoHubba application interface. The central area shows a network graph with nodes of varying sizes and colors (red and orange) connected by edges. The nodes are labeled with gene symbols such as SAT2, GCLM, GCLC, SAT1, ACSL4, NFE2L2, SLC7A1, GPX4, SLC3A2, HMOX1, MAP1LC3B, CP, FTL, FTH1, SLC3A2, NCOA4, ATG5, VDAC2, LPCAT3, SLC40A1, STEAP3, SLC11A2, PCBP1, and FLVCR2. The interface includes a menu bar (File, Edit, View, Select, Layout, Apps, Tools, Help), a toolbar with various icons, and several panels:

- Legend Panel:** Contains 'Target Network' (string_interactions_short.tsv), 'Nodes' Scores' (Calculate, Import, Export), 'Select nodes' (Hubba nodes, Particular nodes), and 'Display options' (Check the first-stage nodes, Display the shortest path, Display the expanded subnetwork).
- Network Panel:** Shows 'Network' and 'Ranking Method' options.
- Node Table:** A table with columns: shared name, name, identifier, node_degree. The first row shows ACSL4, ACSL4, 9606. ENSP0000033..., and 13.
- Command Line:** Located at the bottom of the interface.

Session: New Session

File Edit View Select Layout Apps Tools Help

Legend Panel

App Store

Filter

Annotation

Boundaries

Layout Tools

cytoHubba ▾

Target Network: string_interactions_short.tsv

Nodes' Scores: Calculate Import Export

Select nodes

Hubba nodes

Top 10 node(s) ranked by

MCC

DKMC

MHC

Degree

EPC

BottleNeck

EcCentricity

Closeness

Display options

Check the first-stage nodes

Display the shortest path

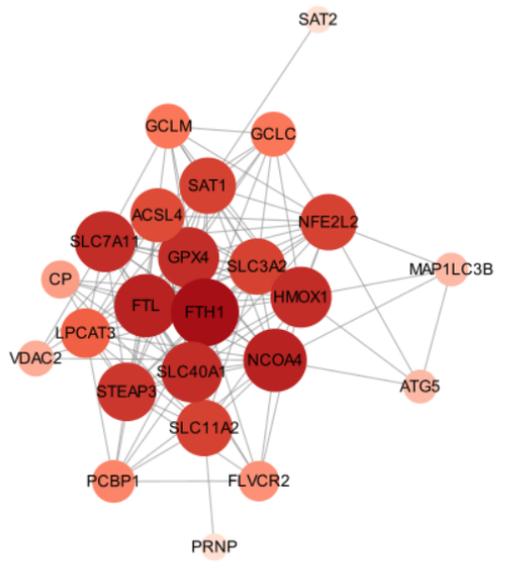
Display the expanded subnetwork

Submit

Network

Ranking Method

Save Current Rank

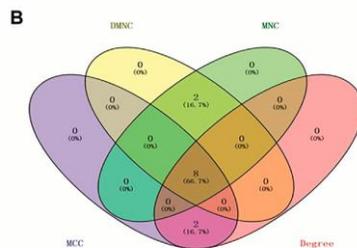


string_interactions_short.tsv

Node Table ▾

shared name	name	identifier	node_degree
ACSL4	ACSL4	9606.ENSP0000033...	13

Node Table | Edge Table | Network Table | Enrichment Table | Unassigned Tables



N种算法venn

Session: New Session

File Edit View Select Layout Apps Tools Help

cytoHubba

Target Network: string_interactions_short.tsv

Nodes' Scores: [Buttons]

Select nodes: Hubba nodes: Top 10 node(s) ranked by Degree

Particular nodes: Nodes you are interested in

Display options: Check the first-stage nodes, Display the shortest path, Display the expanded subnetwork

Network: string_interactions_short.tsv

Ranking Method: Degree

Rank	Node
1	FTH1
2	FTL
2	NCOA4
4	HMOX1
4	SLC7A11
4	GPX4
4	SLC40A1
8	STEAP3
9	SLC3A2
9	NFE2L2

Save Current Rank

string_interactions_short.tsv_Degree_t...

shared name	identifier	node_degree	name
SLC3A2	9606. ENSP0000036...	14	SLC3A2

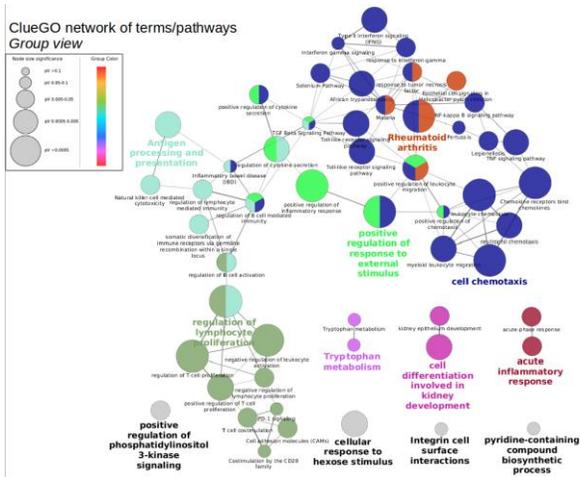
Node Table | Edge Table | Network Table | Enrichment Table | Unassigned Tables

保存Rank文件

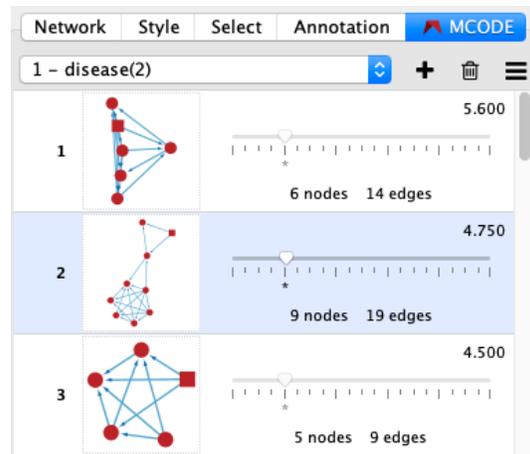
导出网络图

其他常用插件

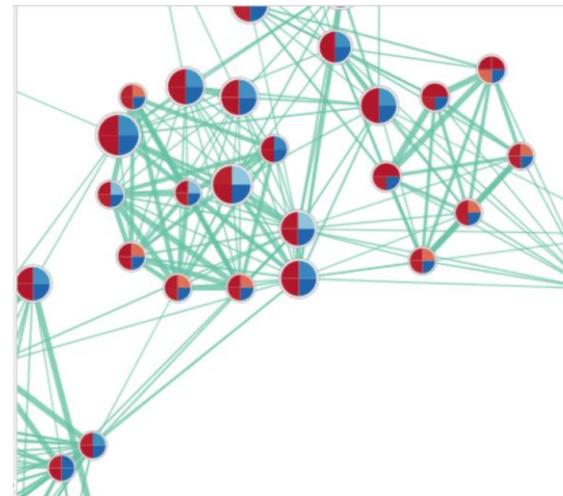
clueGO: 可视化基因集的富集分析结果



MCODE: 找相关模块



EnrichmentMap: GSEA结果网络化



Legend Creator: 创建图例

