



Canadian Bioinformatics Workshops

www.bioinformatics.ca
bioinformaticsdotca.github.io



Canonical URL : <https://creativecommons.org/licenses/by-sa/4.0/> See the legal code

You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

No additional restrictions — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

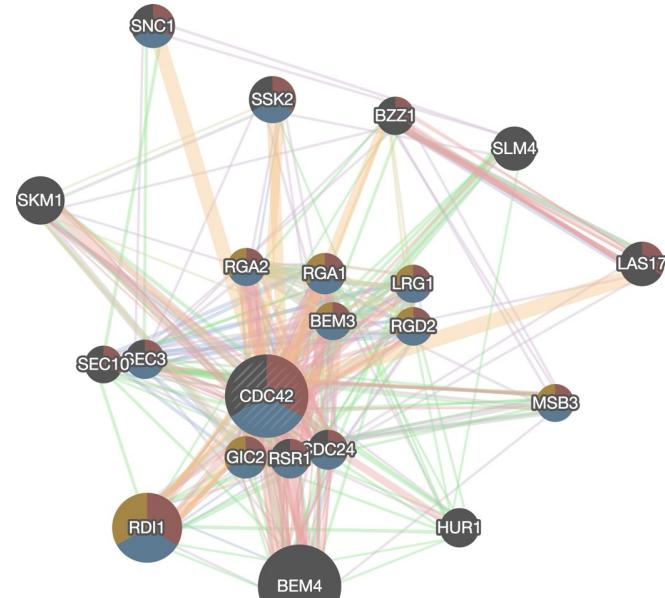
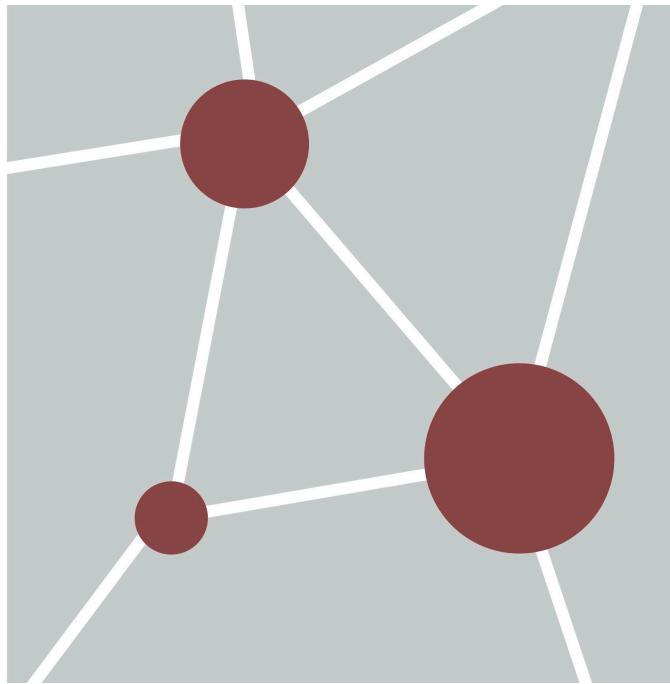
Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

Gene Function Prediction

Veronique Voisin
Pathway and Network Analysis of -omics Data
June 26-28 2024



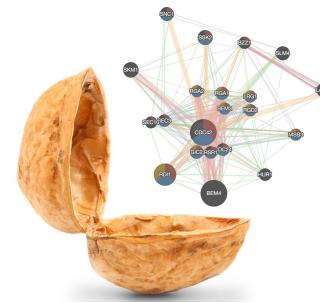
Learning Objectives of this module

- How to use GeneMANIA
- Practical lab: website and Cytoscape app

GeneMANIA: Two types of function prediction

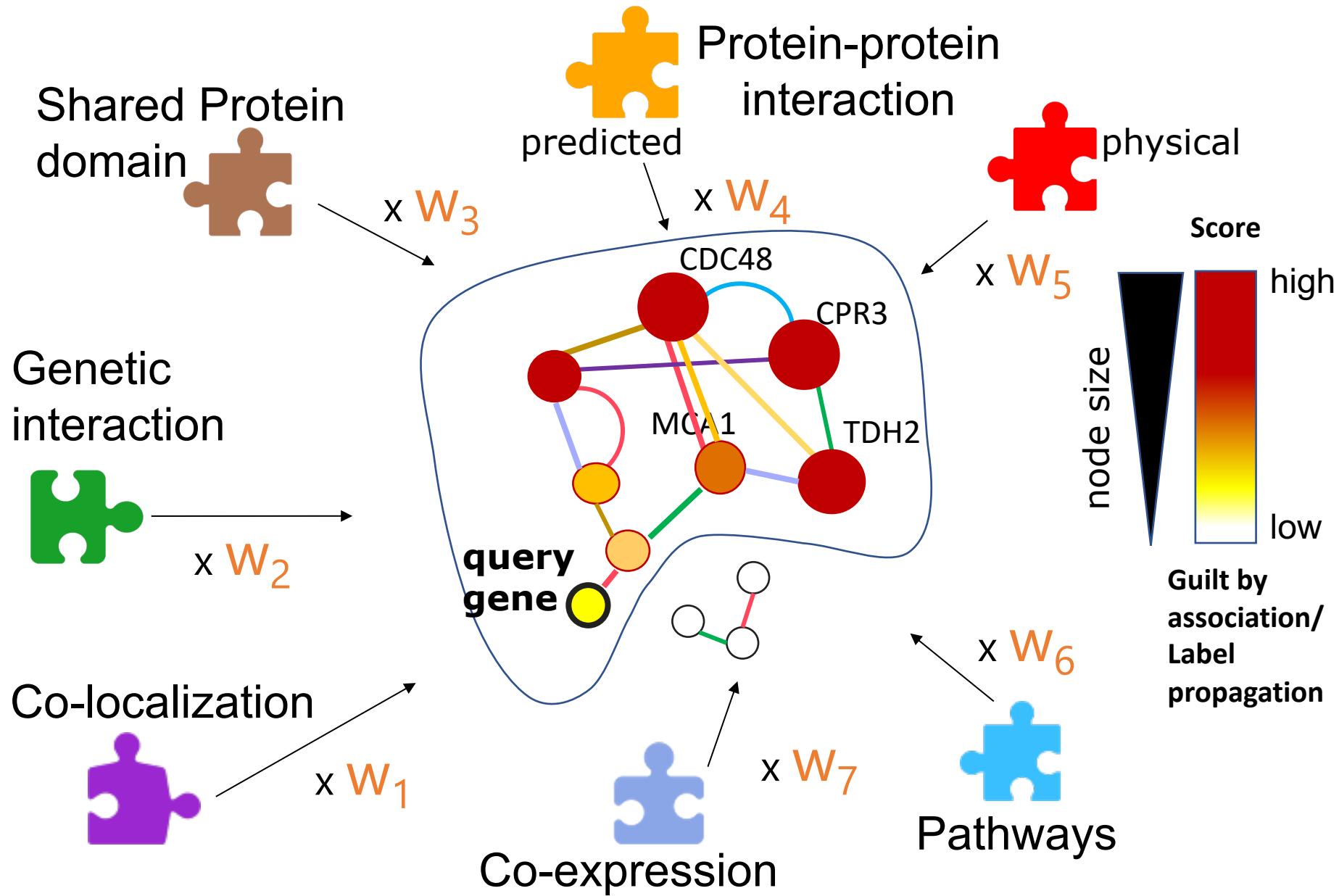
- “What does my gene do?” (1 gene as query)
- “Give me more genes like these” (gene list as query)

GeneMANIA in a



- Is a **functional interaction network**.
 - Is a **collection of interaction networks** that are summed up into one global network. To integrate all these networks into one, there are some **weights** associated with each network that are **context dependent**.
 - Is a **guilt by association/ label propagation** algorithm which finds genes and networks that are functionally associated with our **query gene list**.
 - Is an interactive **web app** and a **Cytoscape app**.

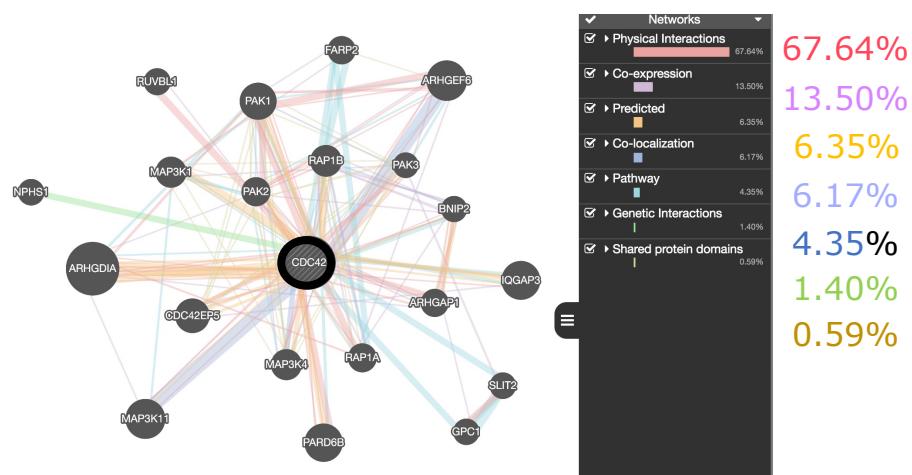
Functional interaction network



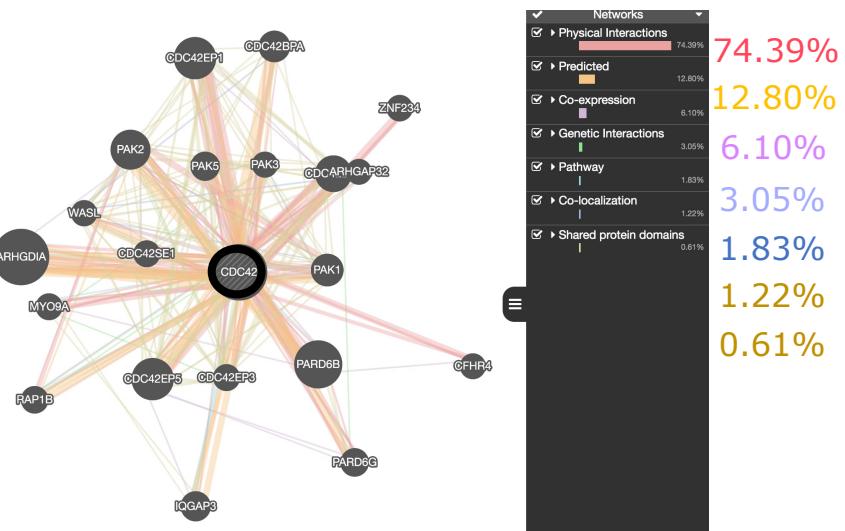
Context-dependent networks

= weight of different networks used to measure node connection

Network weighting=
Automatic (pathways)



Network weighting=
Equal by network



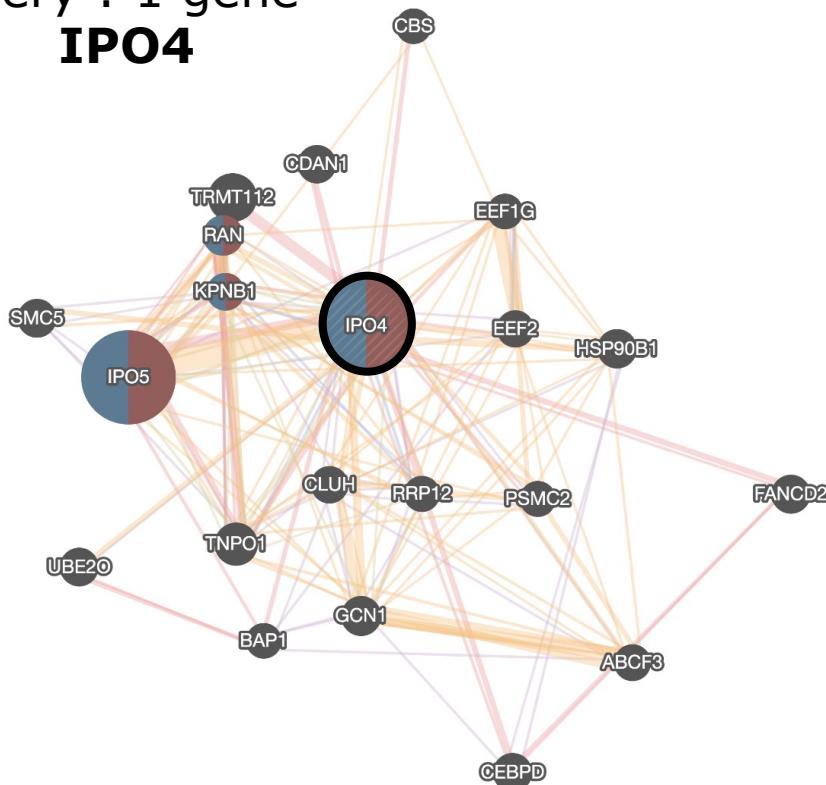
○ query gene

Automatic: GO-based weighting assigns network weights based on how well the networks reproduce patterns of GO co-annotations (“Are genes that interact in the network more likely to have the same annotation?”),

“Function prediction: What does my gene do?”

Query : 1 gene

IPO4



Default parameters

Max resultant genes: 20

Max resultant attributes: 10

Network weighting

Query-dependent weighting

Automatically selected weighting method

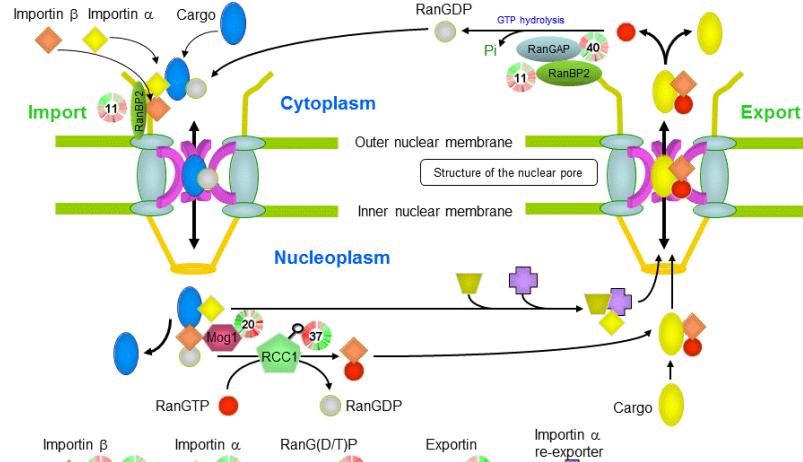
query gene

Genemania web app

Function	FDR	Coverage
pore complex	1.66e-3	4 / 47
nuclear pore	1.66e-3	4 / 41

IPO4: importin 4

Import and export through the nuclear pore



Directionality of transport. The RanGTP gradient determines the vectorial transport of cargo. Translocation through the NPC is an energy-independent, reversible reaction. To achieve directionality of substrate transport, translocation is coupled to a second reaction. During import, this reaction is the dissociation of the import complex by RanGTP, and the subsequent binding of the import substrate to nuclear partners. The export process gains directionality by export complex disassembly in the cytoplasm involving GTP hydrolysis on Ran.

http://mpmp.huji.ac.il/maps/import_export.html

“Give me more genes like these”

Query : 2 genes

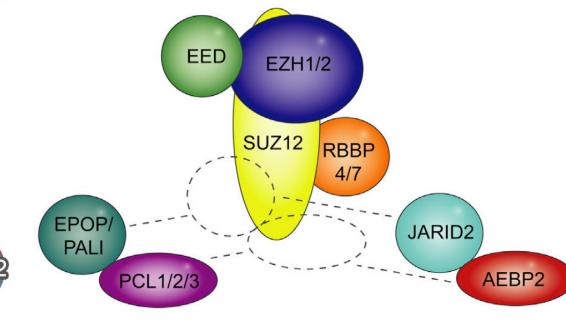
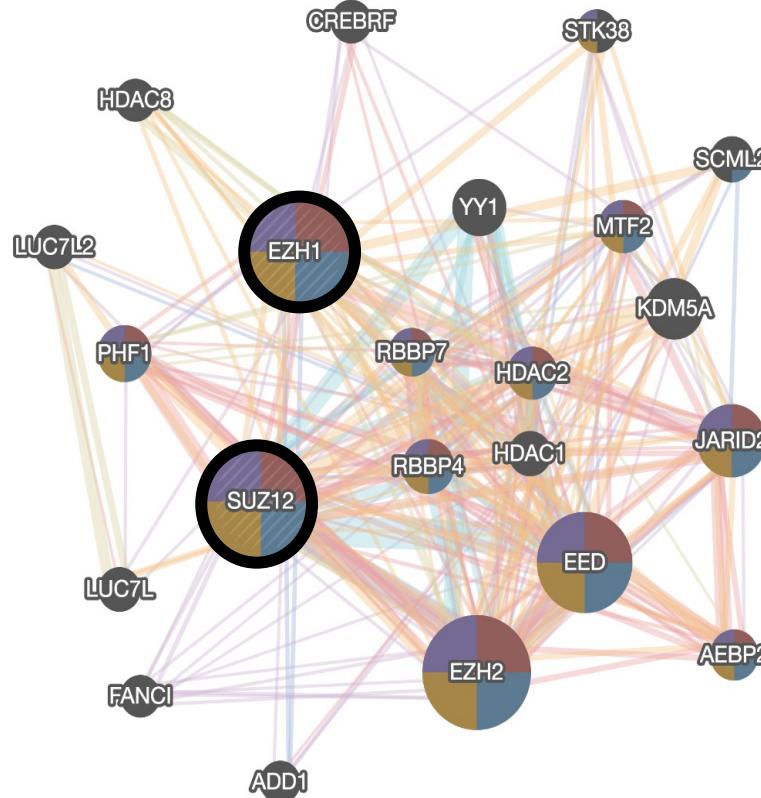
SUZ12 and **EZH1**

Genemania web app

Function	FDR	Coverage
ESC/E(Z) complex	1.71e-29	11 / 13
PcG protein complex	3.70e-24	12 / 39
histone methyltransferase complex	2.05e-21	12 / 64
methyltransferase complex	1.02e-20	12 / 74
NuRD complex	3.32e-6	4 / 14
CHD-type complex	3.32e-6	4 / 14
histone deacetylase complex	3.32e-6	5 / 42
SWI/SNF superfamily-type complex	1.81e-5	5 / 63
Sin3-type complex	5.27e-4	3 / 12
Sin3 complex	5.27e-4	3 / 12
transcriptional repressor complex	5.94e-4	4 / 54
positive regulation of histone methylation	5.94e-4	3 / 13
histone deacetylase activity	1.16e-3	3 / 17
regulation of histone methylation	1.16e-3	3 / 17
protein deacetylase activity	1.30e-3	3 / 18
negative regulation of histone modification	1.98e-3	3 / 21
deacetylase activity	2.83e-3	3 / 24
regulation of chromosome organization	3.12e-3	4 / 90
chromatin remodeling	3.23e-3	4 / 92
positive regulation of histone modification	3.88e-3	3 / 28
nuclear chromosome	5.50e-3	5 / 240
histone binding	5.68e-3	4 / 110

Default parameters

query gene



[https://www.cell.com/molecular-cell/pdf/S1097-2765\(19\)30185-6.pdf](https://www.cell.com/molecular-cell/pdf/S1097-2765(19)30185-6.pdf)

The polycomb repressive complex 2 (PRC2) is a chromatin-associated methyltransferase catalyzing mono-, di-, and trimethylation of lysine 27 on histone H3 (H3K27).

“Give me more genes like these”

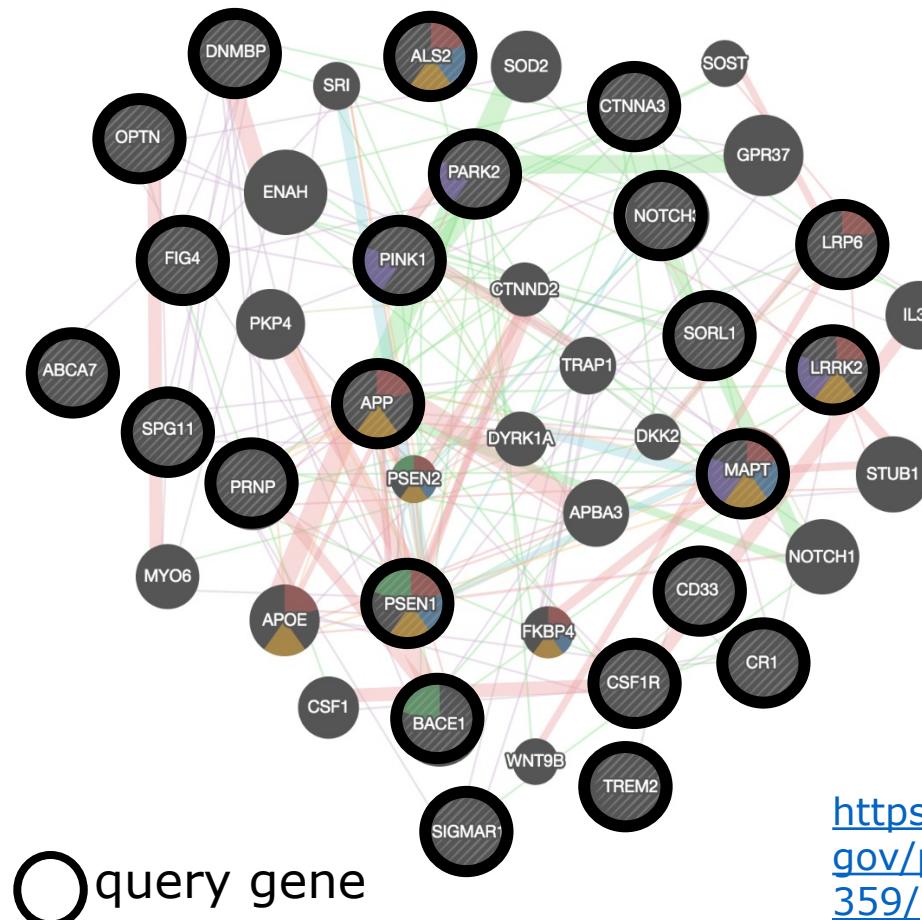
Query : 23 genes

genes with mutations and associated with Early-Onset Alzheimer's Disease



Cytoscape app

Function	FDR	Coverage
neuron part	2.73e-4	9 / 298
site of polarized growth	2.82e-4	5 / 47
neuron projection	2.82e-4	8 / 233
growth cone	2.82e-4	5 / 45
perinuclear region of cytoplasm	4.35e-4	8 / 276
Notch receptor processing	4.35e-4	4 / 21
dendrite	4.39e-4	6 / 112
developmental growth involved in morphogenesis	5.34e-4	5 / 61
neuronal cell body	6.06e-4	5 / 64
cell body	7.24e-4	5 / 69
receptor complex	7.24e-4	7 / 213
membrane protein ectodomain proteolysis	1.29e-3	4 / 33
beta-amyloid metabolic process	1.64e-3	3 / 10
dendritic shaft	1.64e-3	3 / 10
neuron death	2.49e-3	5 / 94
membrane protein proteolysis	2.69e-3	4 / 43
axon	2.69e-3	5 / 97
developmental growth	4.93e-3	5 / 112
negative regulation of multicellular organismal process	1.08e-2	6 / 230
regulation of anatomical structure size	1.90e-2	5 / 151
neuron apoptotic process	2.63e-2	4 / 80
autophagy	5.47e-2	4 / 98



Default parameters

- Max resultant genes: 20
- Max resultant attributes: 10
- Network weighting
 - Query-dependent weighting
 - Automatically selected weighting method

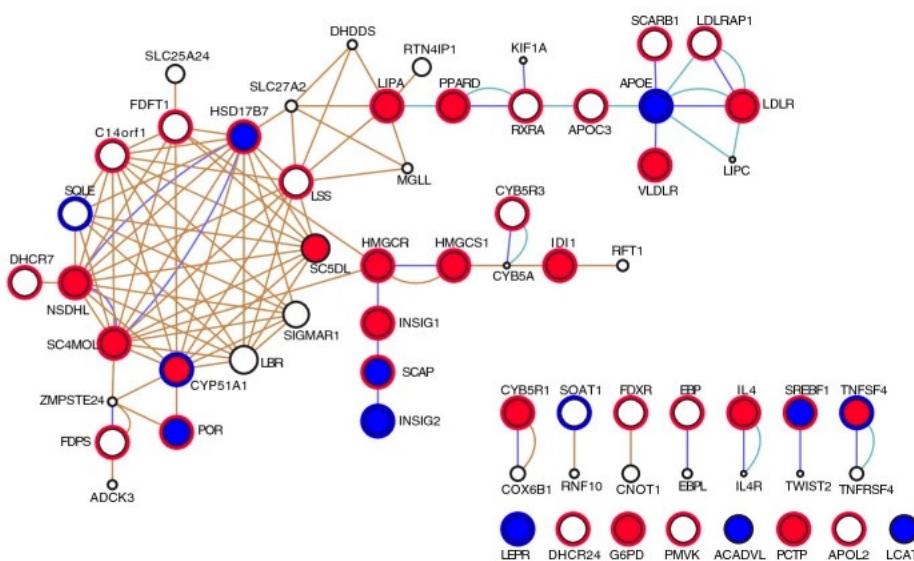
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6471359/>

“Give me more genes like these”

Query : 43 genes
cholesterol metabolism related genes



Cytoscape app



Larger nodes: query genes
smaller white nodes: related genes
Red: up-regulation
Blue: down-regulation
2 time points: 3h and 21h (node center and node border)

Model of Atherosclerosis

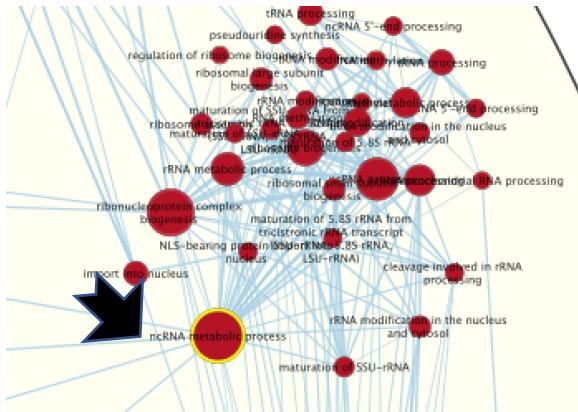
Quiescent human coronary artery SMCs were treated with minimally-oxidized LDL (moxLDL), for 3 hours and 21 hours, respectively.

“from EnrichmentMap to GeneMANIA”

Query : 502 genes
nc RNA metabolic process



Cytoscape app



rRNA metabolic process

tRNA metabolic process

DNA -templated Transcription, initiation

Post transcriptional gene silencing

Large ribosomal subunit

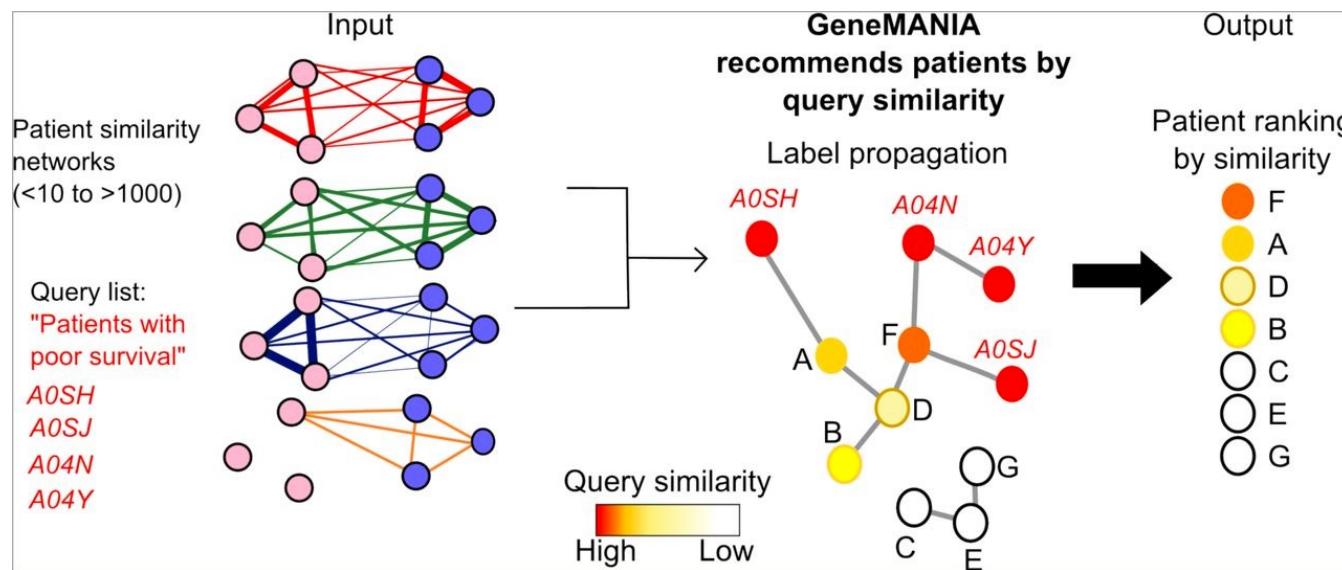
nc RNA metabolic process

Node size and color: relative to gene expression

“Advanced topics”

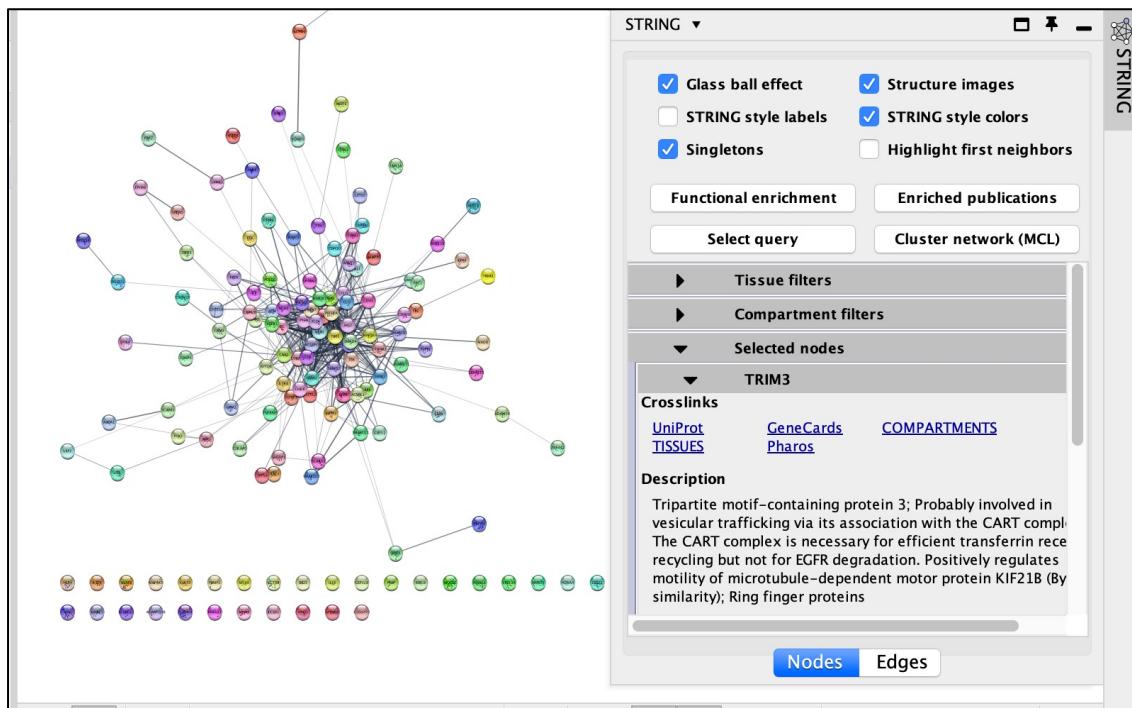
- Possibility to add your model organism if you can construct or have access to interaction networks for it.
- Possibility to use GeneMANIA as an example to assemble different interaction networks that you might have for your research question.

Example: netDx is a **patient classifier** algorithm based on the GeneMANIA algorithm



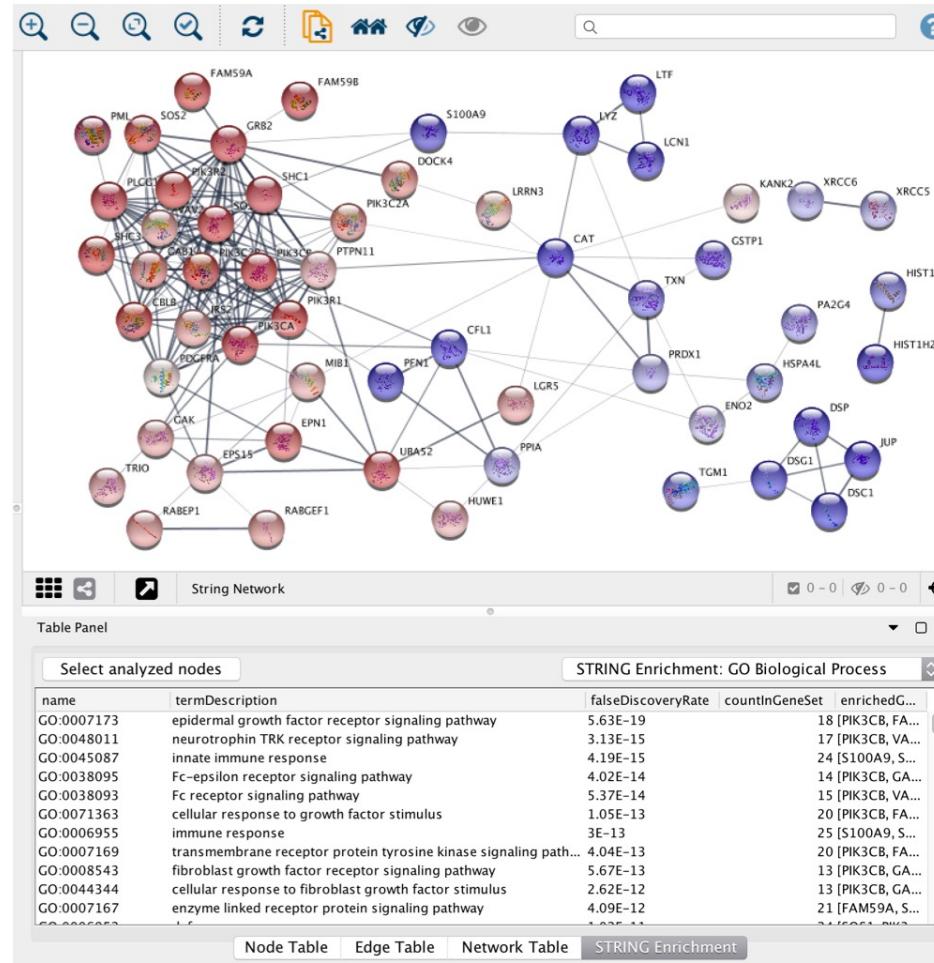
stringApp

stringApp imports **functional associations** or **physical interactions** between protein-protein and protein-chemical pairs from [STRING](https://string-db.org/) (<https://string-db.org/>) , [Viruses.STRING](#), [STITCH](#), [DISEASES](#) and from PubMed text mining into Cytoscape.



- Using STRING protein, it creates a protein-protein interaction network from an input gene list.
- It is possible to use the interaction score to reduce the number of interactions and keep the interactions associated with the strongest scores.

- It is possible to perform functional enrichment using the genes on the network



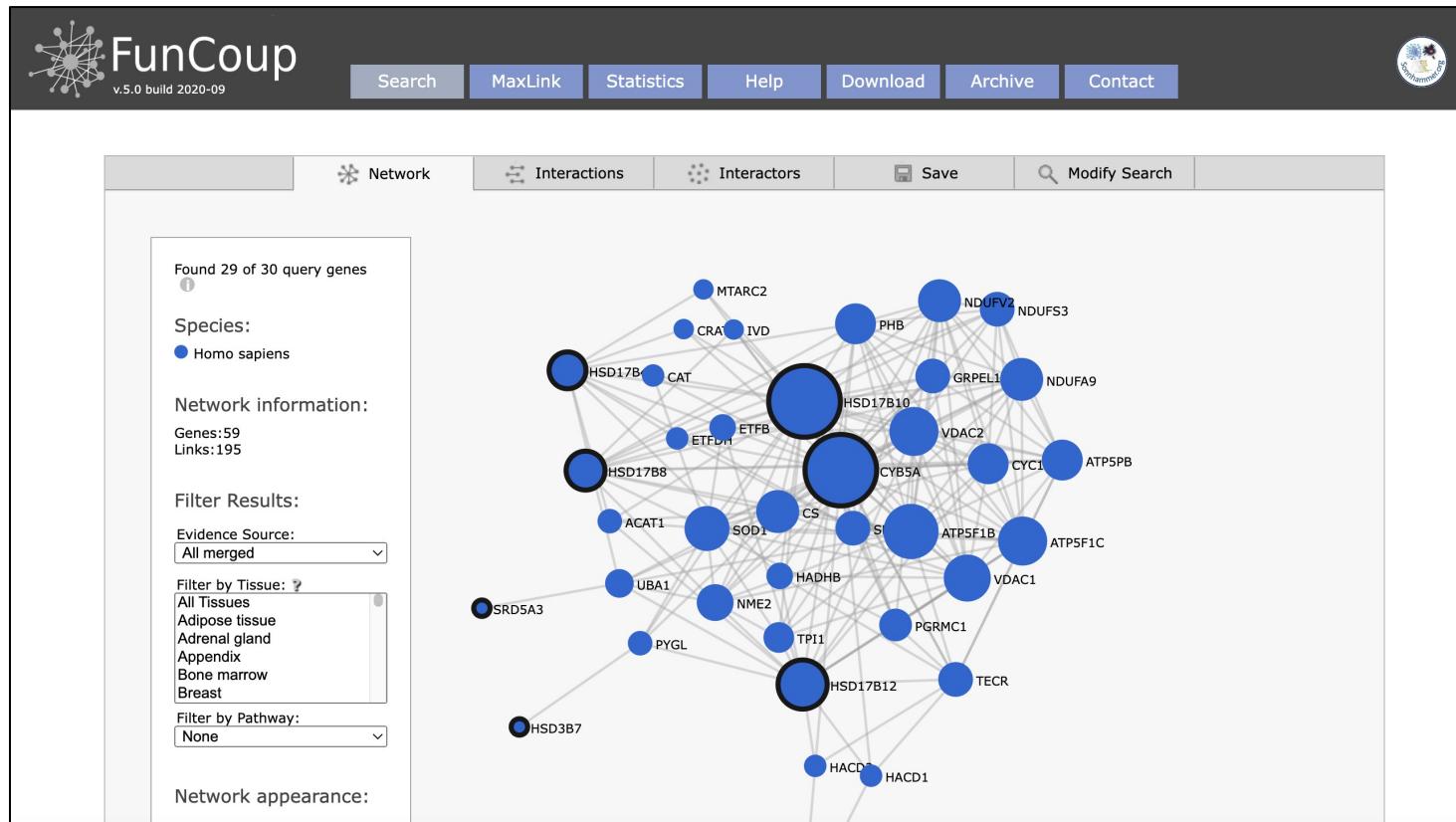
STRING network visualization within Cytoscape for a set of proteins identified using mass spectrometry-based quantitative proteomics (top) and the corresponding enriched GO Biological Process terms (bottom). Figure adapted from Szklarczyk et al., NAR, 2016.

FunCoup

<https://funcoup.org>

The name FunCoup [fən kəp] stands for functional coupling.

FunCoup is a framework to infer genome-wide functional couplings in 21 model organisms. Functional coupling, or functional association, is an unspecific form of association that encompasses direct physical interaction but also more general types of direct or indirect interaction like regulatory interaction or participation in the same process or pathway.



Acknowledgment

Quaid Morris for his slides and lecture:

<https://www.youtube.com/watch?v=2KrUq9ad2xc&list=PL3izGL6oi0S9amh6Czalbuo4ICcvZIWF0&index=6>

GeneMANIA developper team: Christian Lopez, Max Franz, Ruth Isserlin, Quaid Morris , Gary Bader

We are on a Coffee Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



HPC4Health

