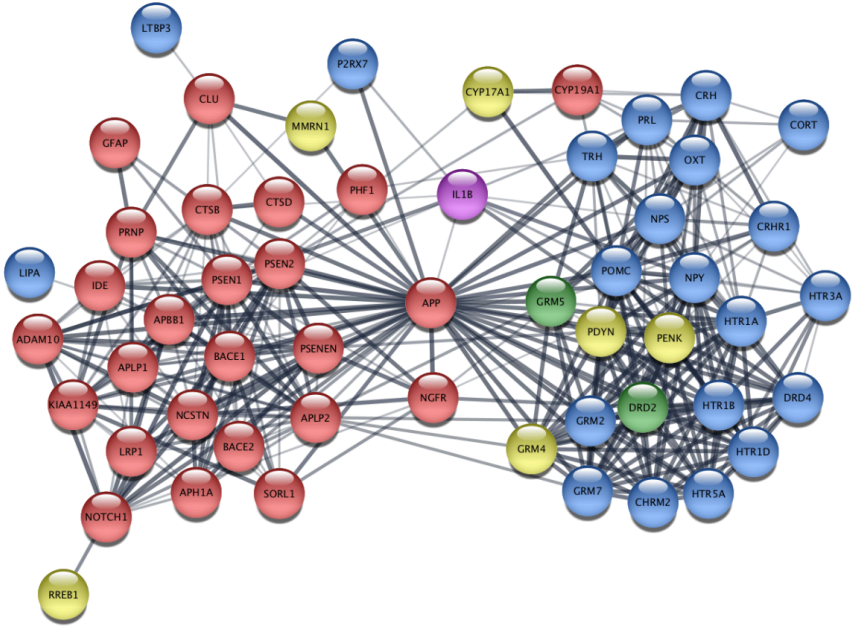
Network biology

A crash course on STRING and Cytoscape



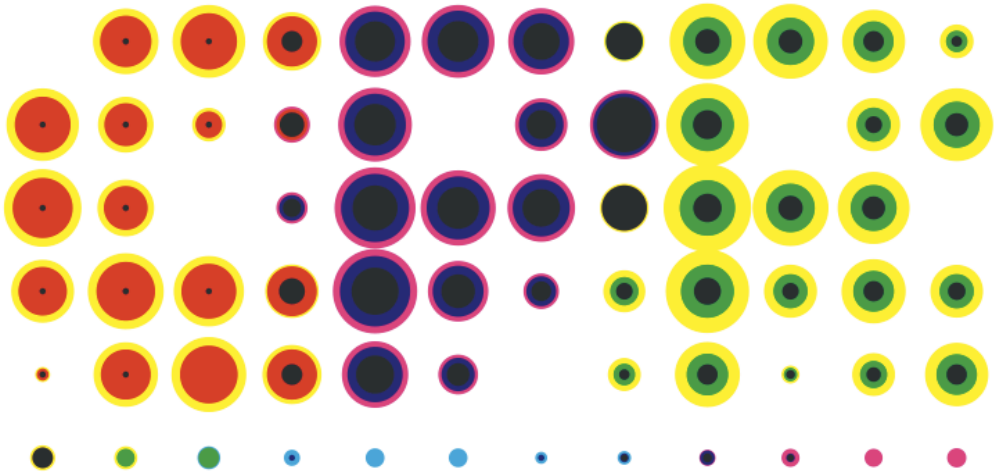
Lars Juhl Jensen

The Novo Nordisk Foundation
Center for Protein Research

FACULTY OF HEALTH SCIENCES
UNIVERSITY OF COPENHAGEN

me

group leader



The Novo Nordisk Foundation

Center for Protein Research

cofounder

intomics
from data to biology

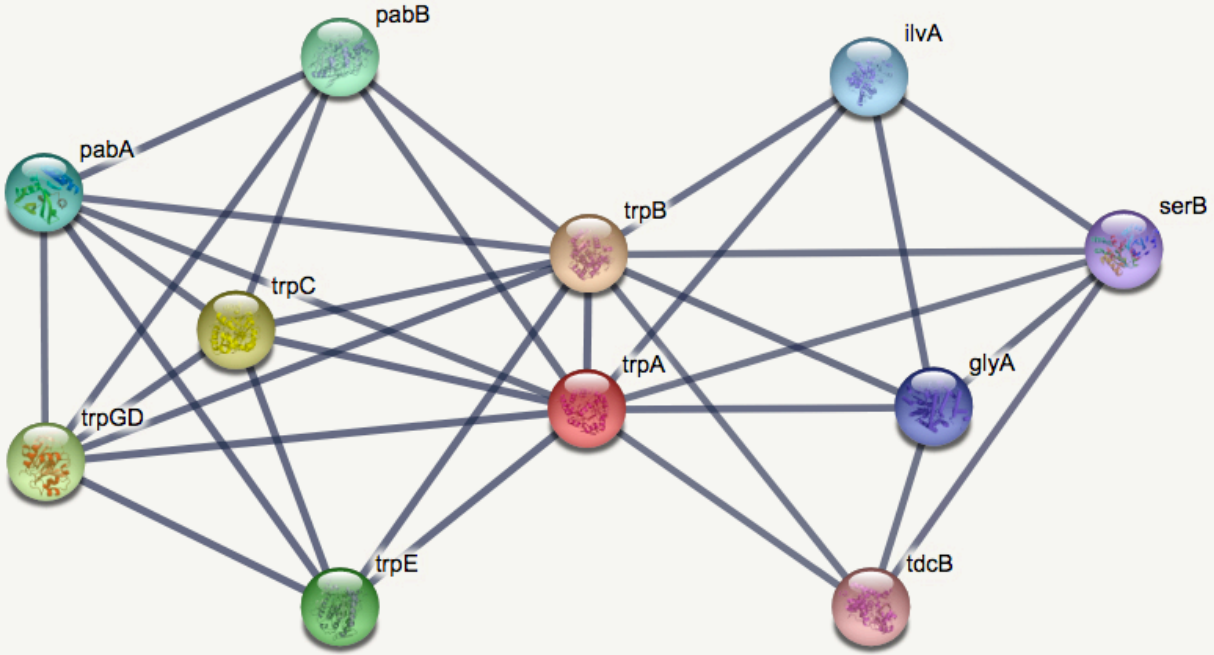
high-throughput data

proteins

understand their interplay

STRING

network database



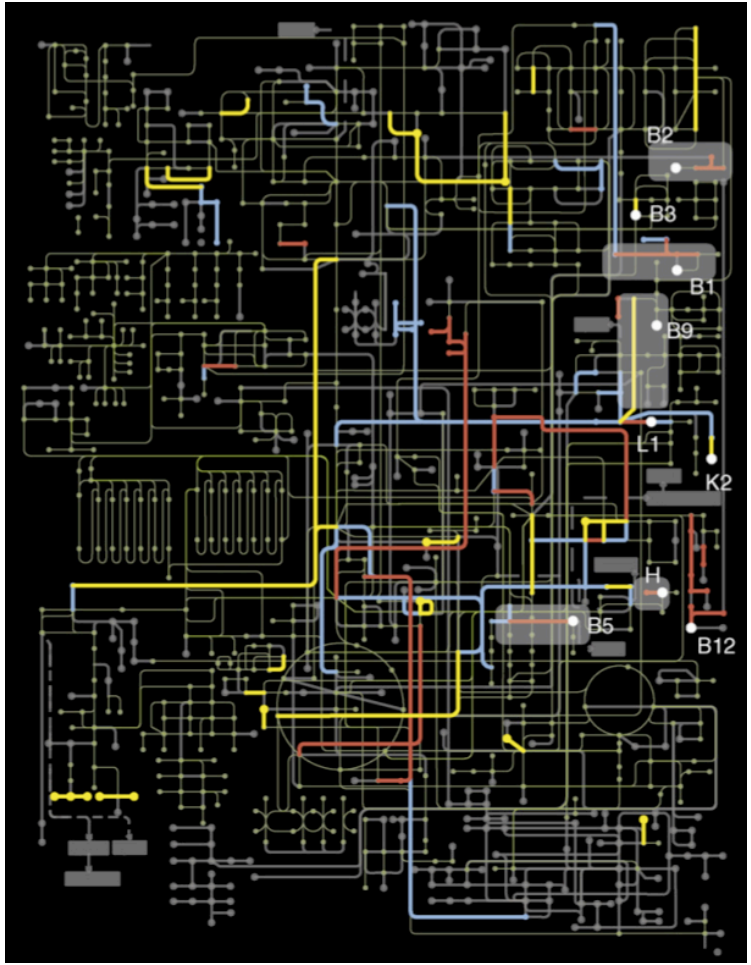
string-db.org

9.6 million genes

functional associations

curated knowledge

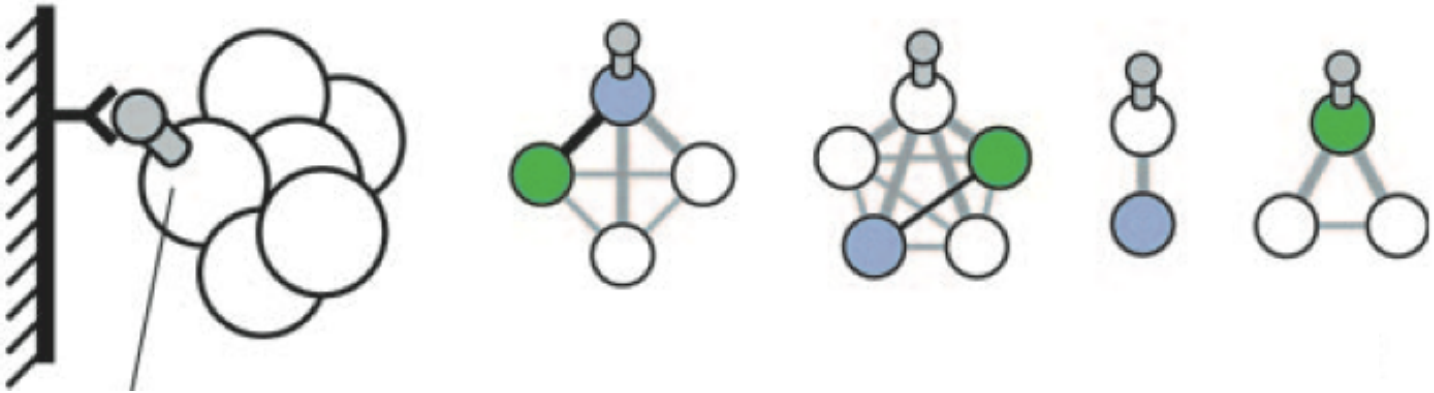
pathways



Letunic & Bork, *Trends in Biochemical Sciences*, 2008

experimental data

physical interactions

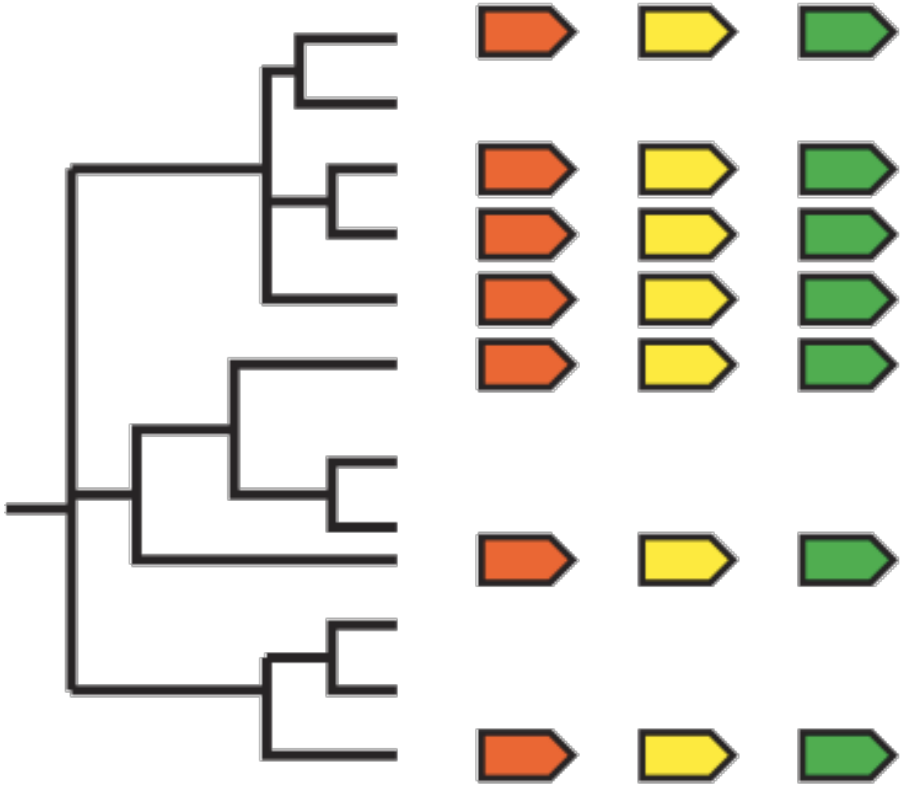


von Mering et al., *Nucleic Acids Research*, 2005

co-expression

genomic context

phylogenetic profiles



Korbel et al., *Nature Biotechnology*, 2004

many databases

different formats

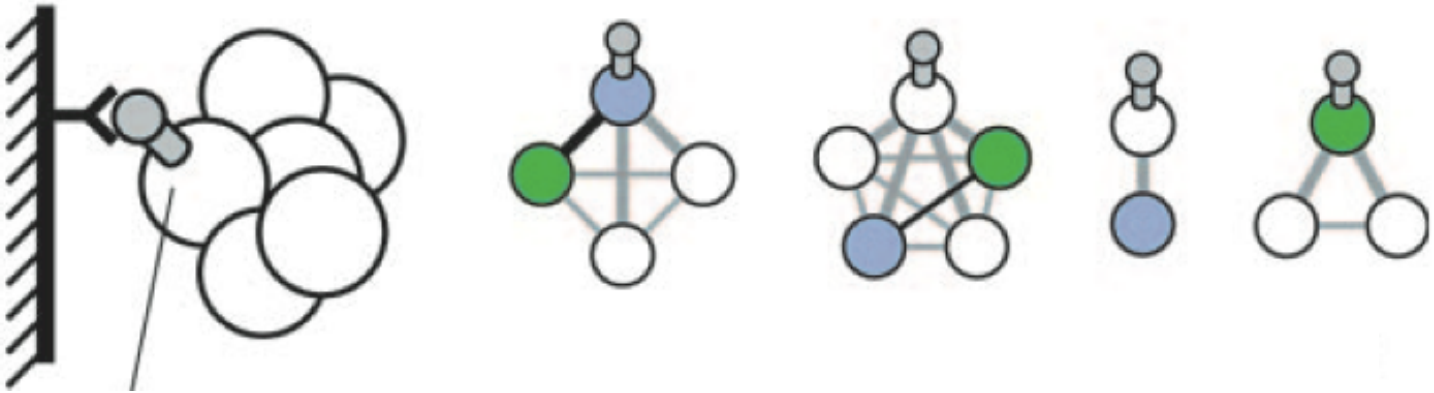
different identifiers

varying quality

not comparable

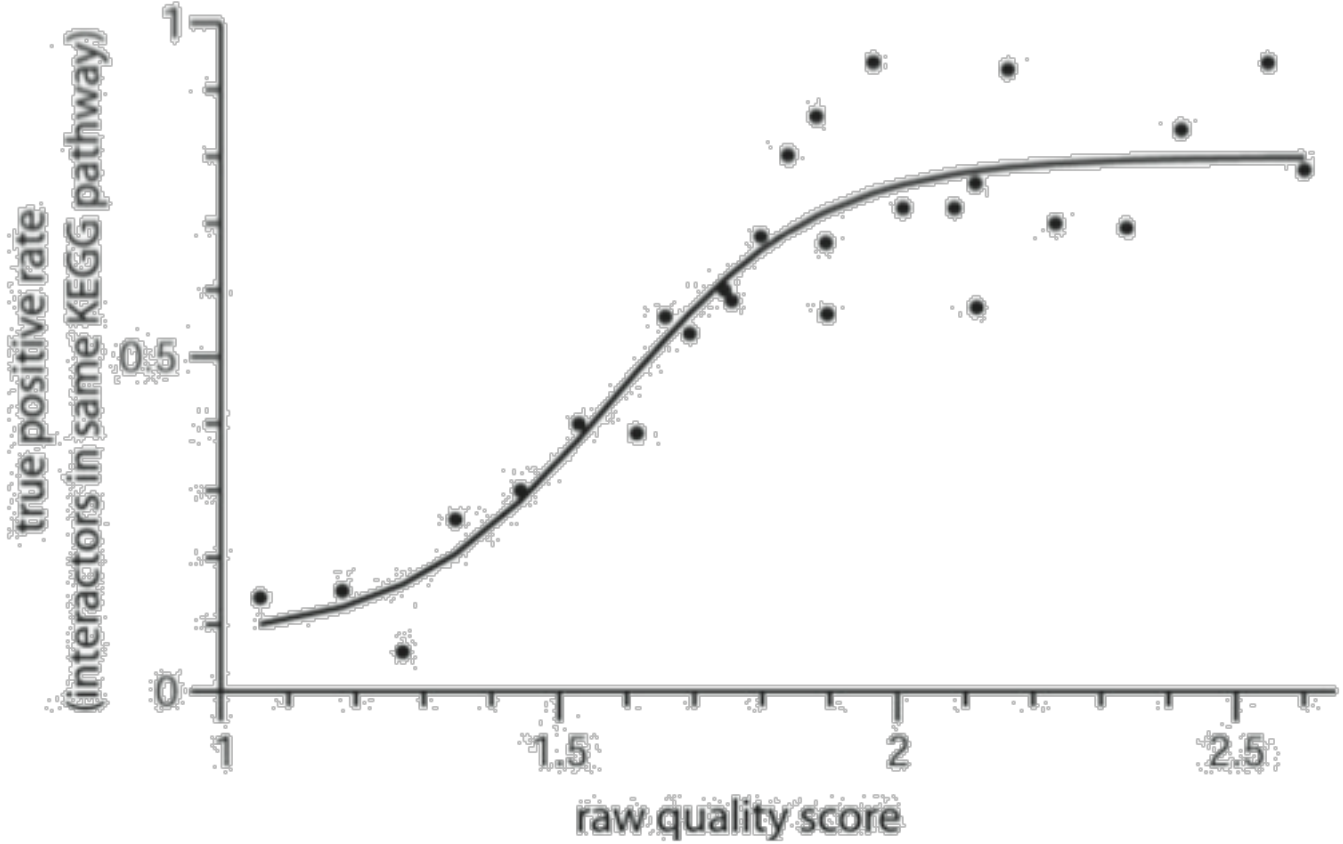
hard work

quality scores



von Mering et al., *Nucleic Acids Research*, 2005

score calibration



von Mering et al., *Nucleic Acids Research*, 2005

missing most of the data

>10 km

too much to read

computer

as smart as a dog

teach it specific tricks

What we say to dogs
Okay, Ginger! I've had it! You stay out of the garbage! Understand, Ginger? Stay out of the garbage, or else!

what they hear
blah blah GINGER blah
blah blah blah blah blah
blah blah GINGER blah
blah blah blah blah...

named entity recognition

dictionary

cyclin dependent kinase 1

CDC2

orthographic variation

cyclin dependent kinase 1

cyclin-dependent kinase 1

CDC2

hCDC2

black list

SDS

co-mention score

Cytoscape

network visualization tool

not a database

stringApp

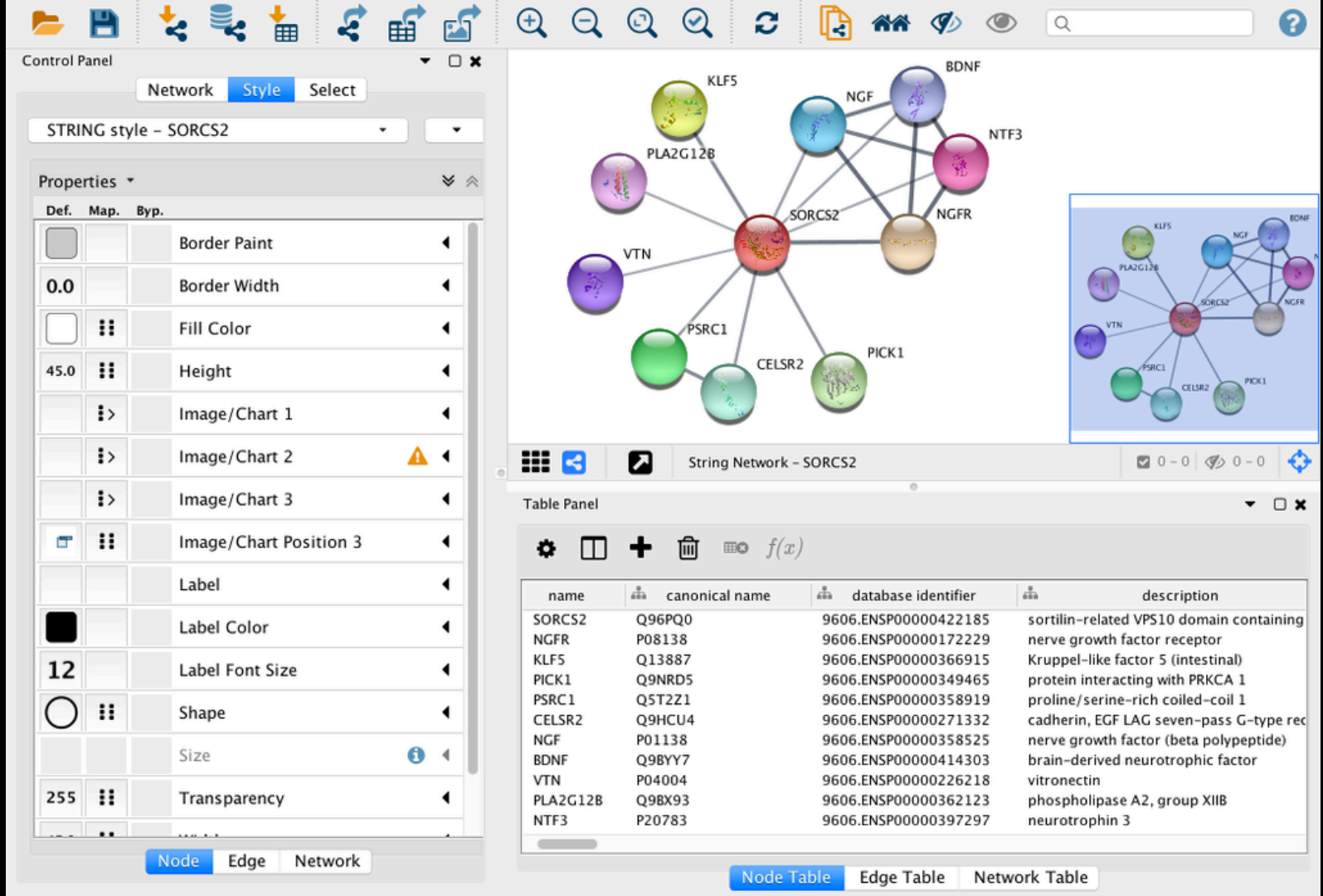
STRING → Cytoscape

user interface

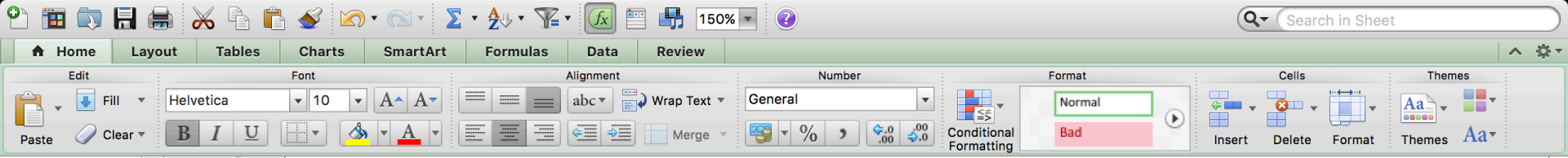
networks

tables

visual styles


Control Panel
Network
Style
Select
STRING style - SORCS2
Properties
Def.
Map.
Byp.
Border Paint
0.0
Border Width
Fill Color
45.0
Height
Image/Chart 1
Image/Chart 2
Image/Chart 3
Image/Chart Position 3
Label
Label Color
12
Label Font Size
Shape
Size
255
Transparency
Node
Edge
Network
KLF5
BDNF
NGF
NTF3
PLA2G12B
SORCS2
NGFR
VTN
PSRC1
CELSR2
PICK1
String Network - SORCS2
0 - 0
0 - 0
Table Panel
f(x)
name
canonical name
database identifier
description
SORCS2 Q96PQ0 9606.ENSP00000422185 sortilin-related VPS10 domain containing
NGFR P08138 9606.ENSP00000172229 nerve growth factor receptor
KLF5 Q13887 9606.ENSP00000366915 Kruppel-like factor 5 (intestinal)
PICK1 Q9NRD5 9606.ENSP00000349465 protein interacting with PRKCA 1
PSRC1 Q5T2Z1 9606.ENSP00000358919 proline/serine-rich coiled-coil 1
CELSR2 Q9HCU4 9606.ENSP00000271332 cadherin, EGF LAG seven-pass G-type rec
NGF P01138 9606.ENSP00000358525 nerve growth factor (beta polypeptide)
BDNF Q9BYY7 9606.ENSP00000414303 brain-derived neurotrophic factor
VTN P04004 9606.ENSP00000226218 vitronectin
PLA2G12B Q9BX93 9606.ENSP00000362123 phospholipase A2, group XIIB
NTF3 P20783 9606.ENSP00000397297 neurotrophin 3
Node Table
Edge Table
Network Table

proteomics experiment

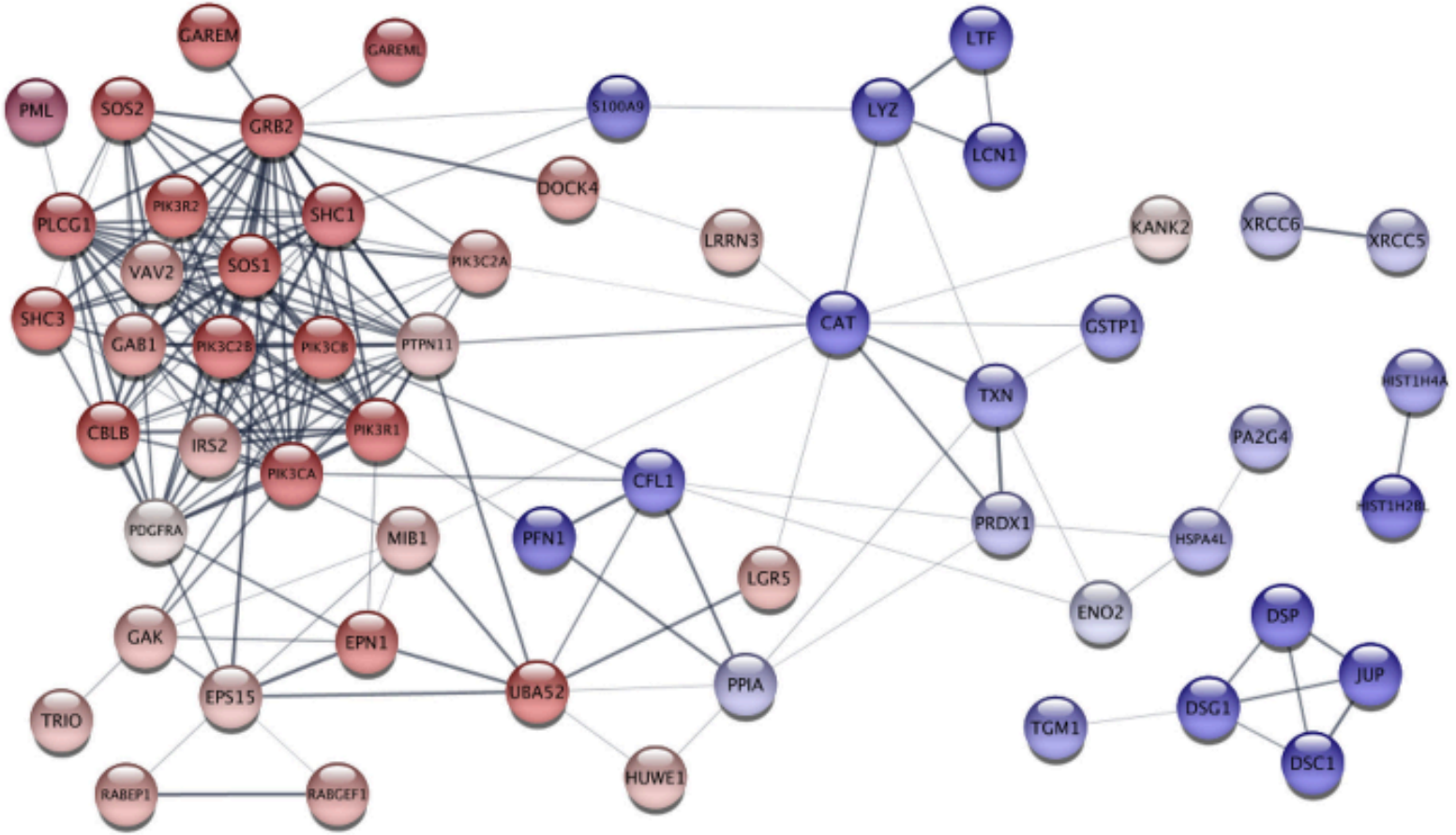


UniProt	Gene name	Peptides	Sequence coverage [%]	5 min ratio exp. 1	5 min ratio exp. 2	5 min log ratio	10 min ratio exp. 1	10 min ratio exp. 2	10 min log ratio
Q75VX8	GAREML	26	37.5	8.59	5.97	2.86	9.71	7.57	3.11
P29353	SHC1	21	44.1	6.13	7.50	2.77	6.03	9.01	2.91
O14492	SH2B2	5	9.5	6.55	6.69	2.73	7.05	10.85	3.16
P19174	PLCG1	69	50	5.63	7.23	2.69	6.06	8.07	2.82
Q07890	SOS2	30	22.5	4.65	6.35	2.46	4.48	8.55	2.70
P62993	GRB2	13	52.1	4.83	5.69	2.39	4.65	6.80	2.52
P42336	PIK3CA	21	22.8	4.43	4.95	2.23	6.09	5.92	2.59
Q9H706	GAREM	17	23.9	4.72	4.61	2.22	4.14	5.78	2.31
O00750	PIK3C2B	35	24.2	4.49	4.77	2.21	4.64	5.27	2.31
Q8TER5	ARHGEF40	34	28.3	3.65	4.09	1.95	2.64	3.24	1.56
P27986	PIK3R1	19	28.2	4.13	3.61	1.95	4.97	5.83	2.43
O00459	PIK3R2	21	40.2	3.88	3.77	1.94	4.59	4.89	2.24
Q07889	SOS1	35	27.8	3.76	3.72	1.90	3.97	5.35	2.22
Q92529	SHC3	5	13	2.85	3.86	1.75	3.89	5.55	2.24
P42338	PIK3CB	12	13.4	3.68	2.72	1.68	4.86	3.88	2.13
P29590	PML	4	4.8	0.59	16.61	3.10	0.67	114.56	5.85
P62987	UBA52	16	57.8	3.08	2.83	1.56	3.47	3.47	1.79
Q9UKV5	AMFR	8	19.1	3.13	2.57	1.51	2.60	2.42	1.33
Q13191	CBLB	10	13.1	2.45	3.24	1.51	3.77	5.63	2.23
Q13480	GAB1	5	9.2	2.38	2.69	1.34	2.08	3.13	1.38
Q9Y6I3	EPN1	14	27.8	2.32	2.64	1.31	2.83	3.49	1.66
O00443	PIK3C2A	29	17.8	2.32	2.07	1.13	2.39	2.41	1.26
Q8IZ07	ANKRD13A	12	19.2	2.22	1.99	1.07	1.99	2.25	1.08

Emdal et al., *Science Signaling*, 2015

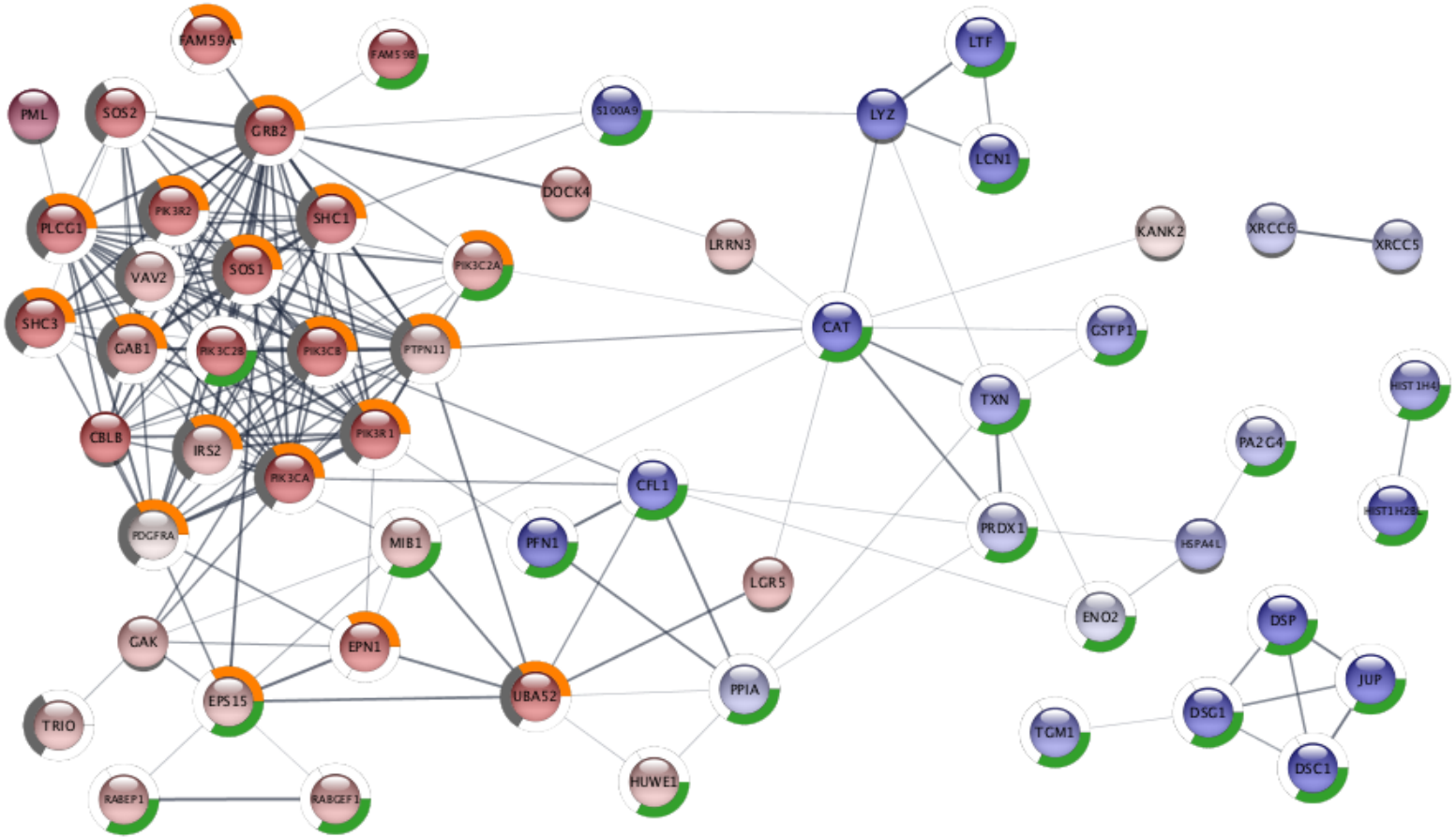
import network

map external data



Szklarczyk et al., *Nucleic Acids Research*, 2017

term enrichment analysis

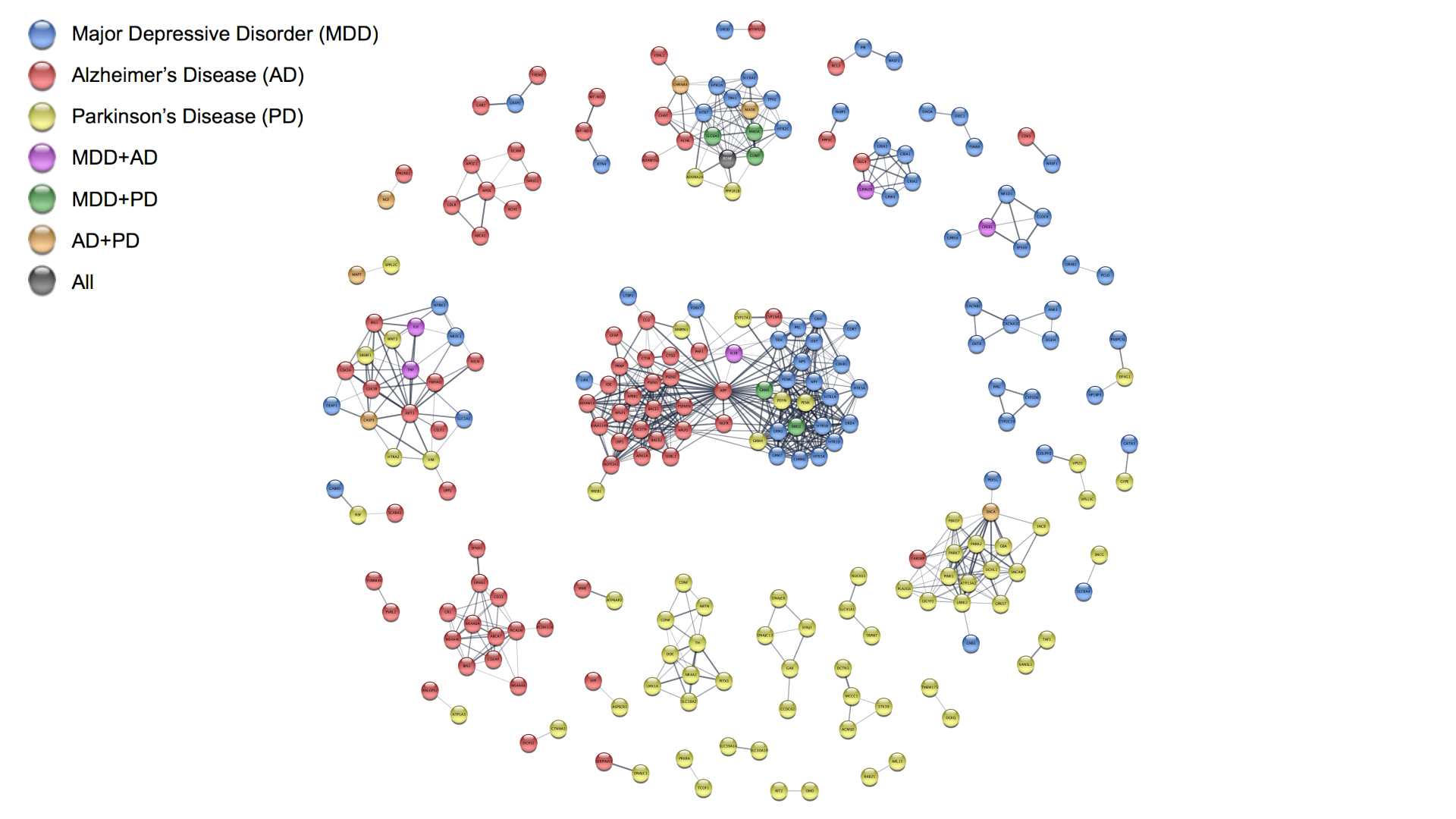

FAM59A
FAM59B
PML
SOS2
GRB2
S100A9
LTF
LYZ
LCN1
DOCK4
PLCG1
PIK3R2
SHC1
LRRN3
KANK2
XRCC6
XRCC5
VAV2
SOS1
PIK3C2A
SHC3
GAB1
PIK3C2B
PIK3CB
PTPN11
CAT
GSTP1
TXN
HIST1H4J
CBLB
IRS2
PIK3R1
PA2G4
PIK3CA
CFL1
HIST1H2BJ
PDGFRA
MIB1
PFN1
PRDX1
HSPA4L
LGR5
ENO2
GAK
EPN1
DSP
JUP
EPS15
UBA52
PPIA
TRIO
TGM1
DSG1
DSC1
HUWE1
RABEP1
RABGEF1

diseases of interest

associated genes

import network

clustering


Major Depressive Disorder (MDD)
Alzheimer's Disease (AD)
Parkinson's Disease (PD)
MDD+AD
MDD+PD
AD+PD
All

training material

LEARNING OBJECTIVES

PREREQUISITES

EXERCISE 1

Protein queries
Compound queries
Disease queries
PubMed queries
New search interface

EXERCISE 2

Protein network retreival
Discrete color mapping
Data import
Continuous color mapping
Functional enrichment

Cytoscape stringApp exercises

Learning objectives

In these exercises, we will use the stringApp for Cytoscape to retrieve molecular networks from the STRING and STITCH databases. The exercises will teach you how to:

- retrieve networks for proteins or small-molecule compounds of interest
- retrieve networks for a disease or an arbitrary topics in PubMed
- layout and visually style the resulting networks
- import external data and map them onto a network
- perform enrichment analyses and visualize the results

open licenses

outlook

viruses

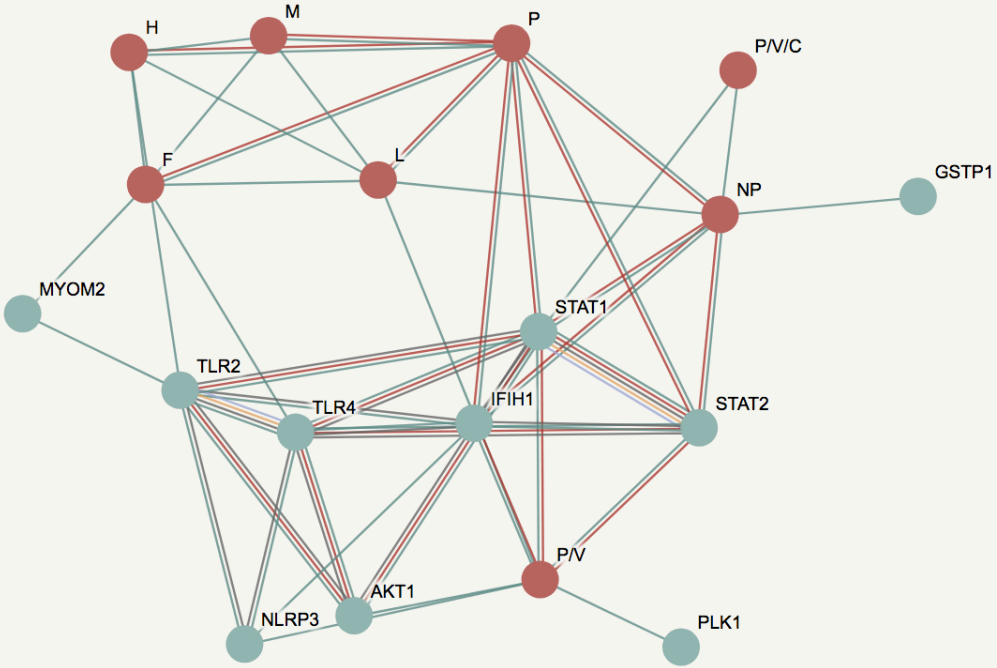
Measles virus (strain Ichinose-B95a)

Measles virus (strain Ichinose-B95a)

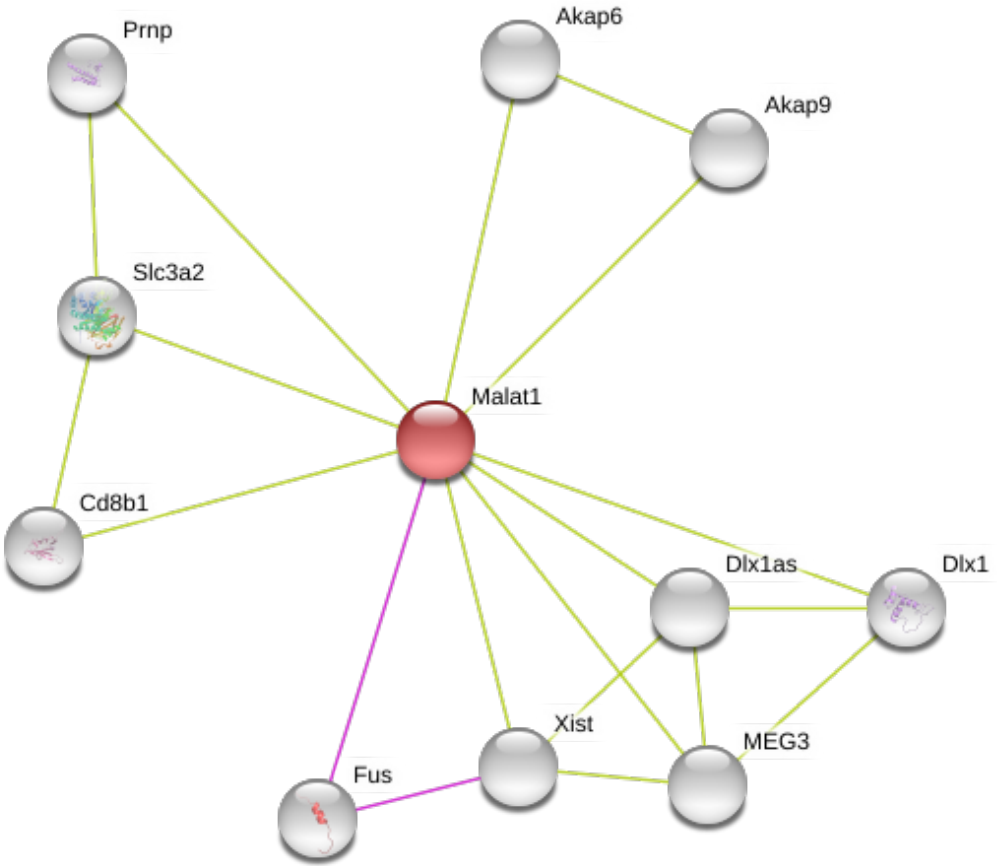
*NCBI taxonomy Id: **11234***

Homo sapiens

*NCBI taxonomy Id: **9606***



ncRNAs



rth.dk/resources/rain/

Junge, Refsgaards, Garde et al., *Database*, 2017

Acknowledgments

STRING

Damian Szklarczyk

Helen Cook

Michael Kuhn

Stefan Wyder

Milan Simonovic

Alberto Santos

Alexander Roth

Peer Bork

Christian von Mering

Cytoscape stringApp

Nadezhda Doncheva

John “Scooter” Morris

SIB

novo nordisk fonden

NIH National Institutes of Health

EMBL