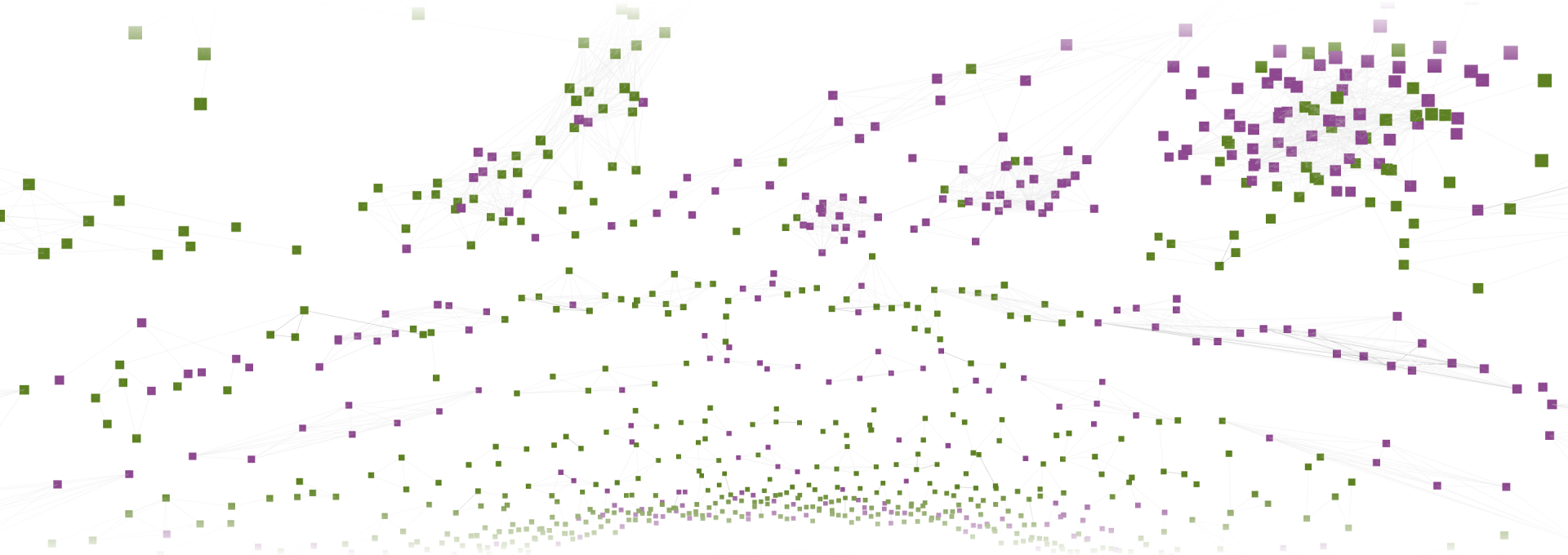
# Network biology




Lars Juhl Jensen

jensenlab.org

The Novo Nordisk Foundation
Center for Protein Research

KØBENHAVNS
UNIVERSITET

# what will you learn?

# STRING database

# network creation

# enrichment analysis

# FAVA

# co-expression networks

# Cytoscape tool

# STRING integration

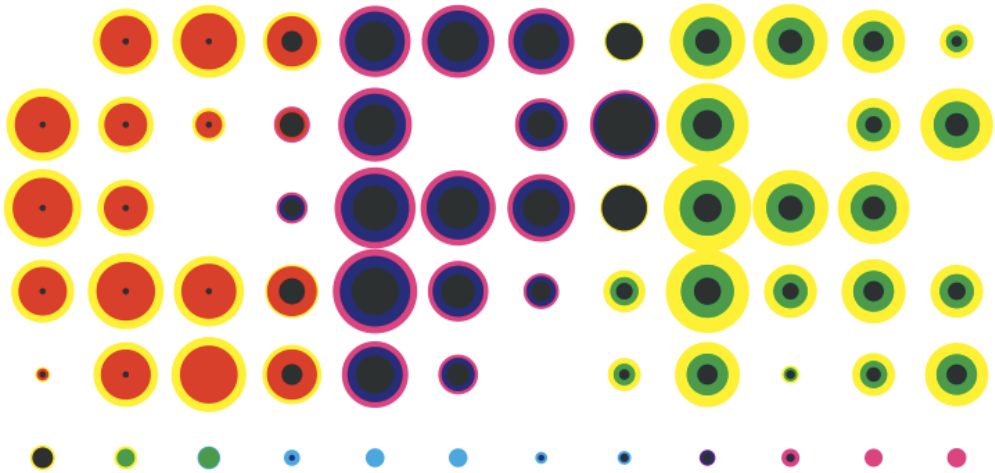
# network visualization

# who are we?

# Lars Juhl Jensen

# group leader

2009–



The Novo Nordisk Foundation

**Center for Protein Research**

# cofounder

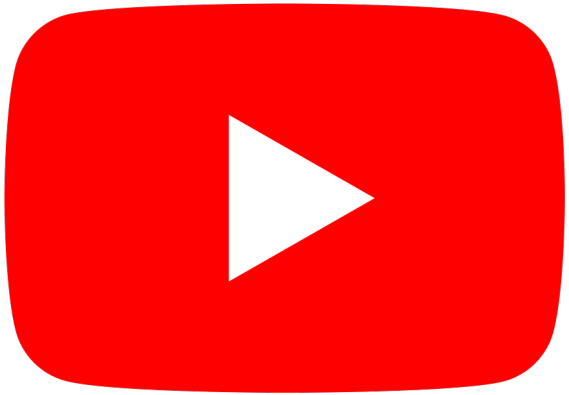
2009–2022


ZS


intomics

YouTuber

2020–



youtube.com/larsjuhljensen

# Mikaela Koutrouli

# PhD student

# single-cell transcriptomics

# Nadezhda Doncheva

assistant professor

# omics data analysis

# network analysis

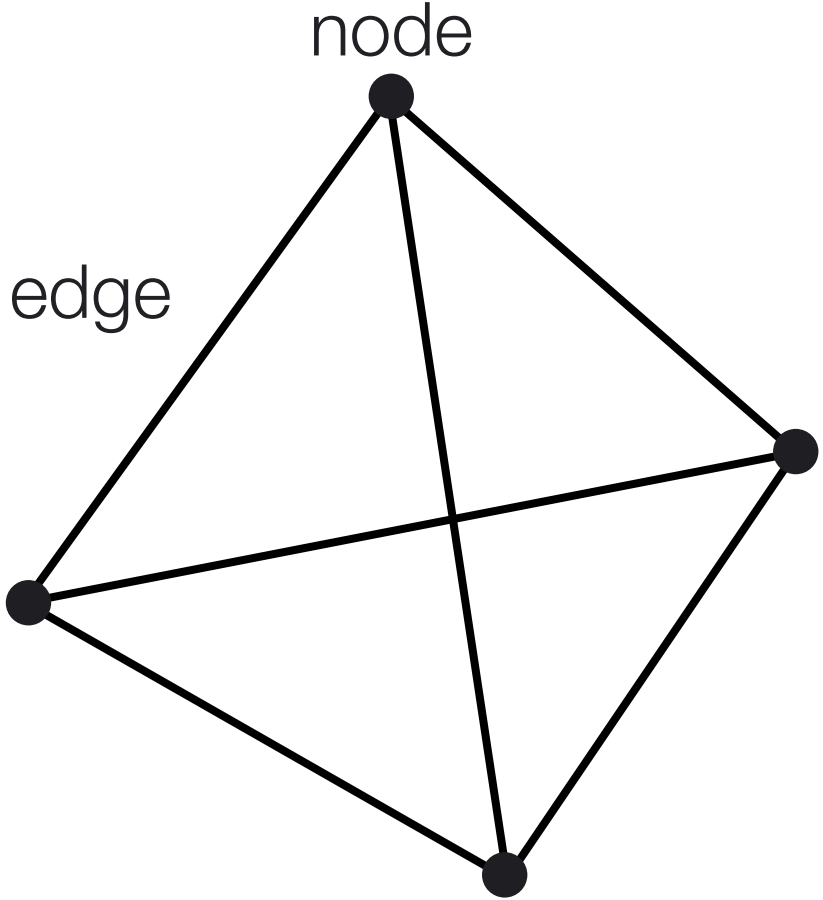
# why networks?

entities or concepts

understand their interplay

useful abstraction

lends itself to visualization


node
edge

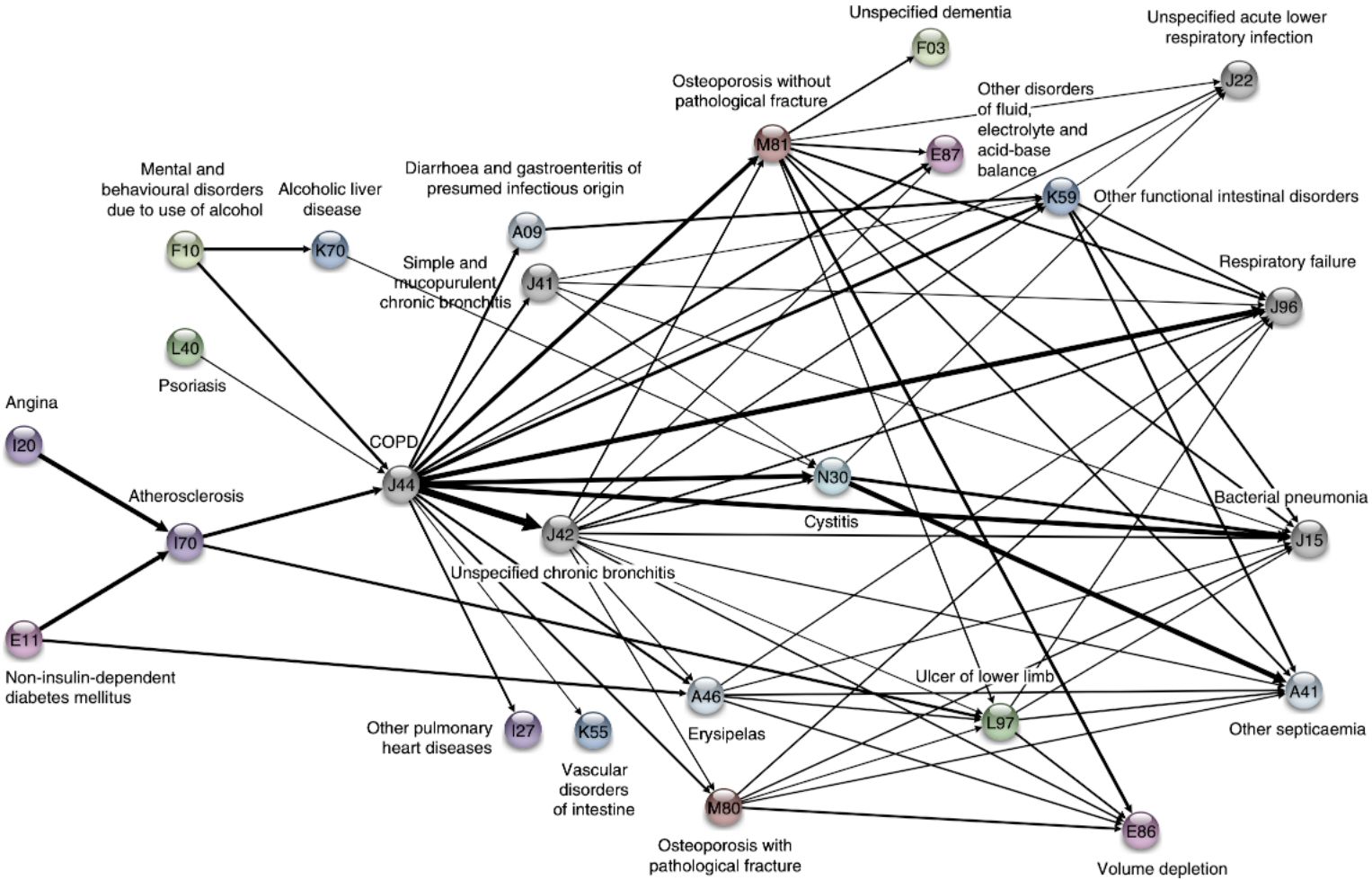
# nodes

# things to be linked

edges

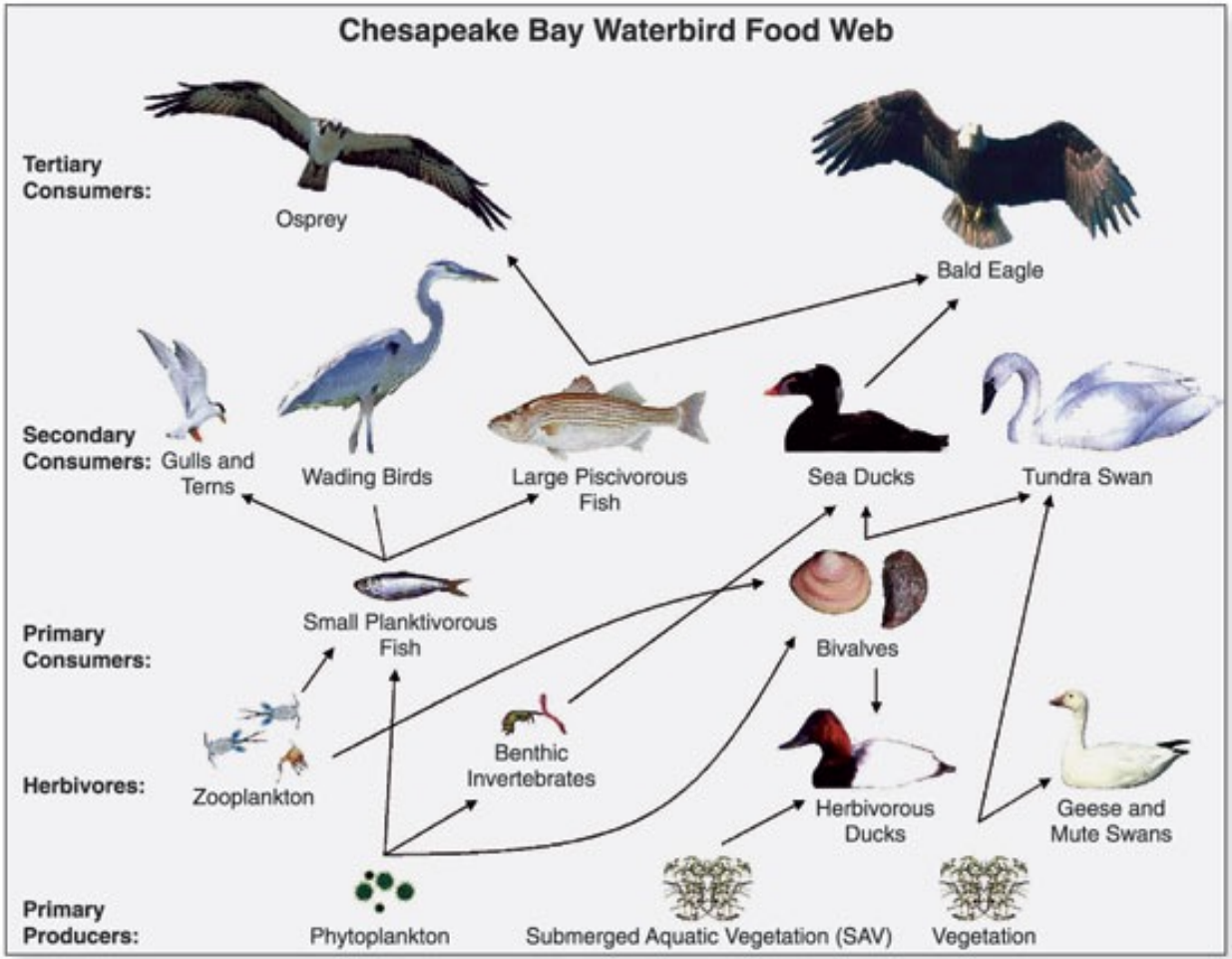
links between things

# disease networks



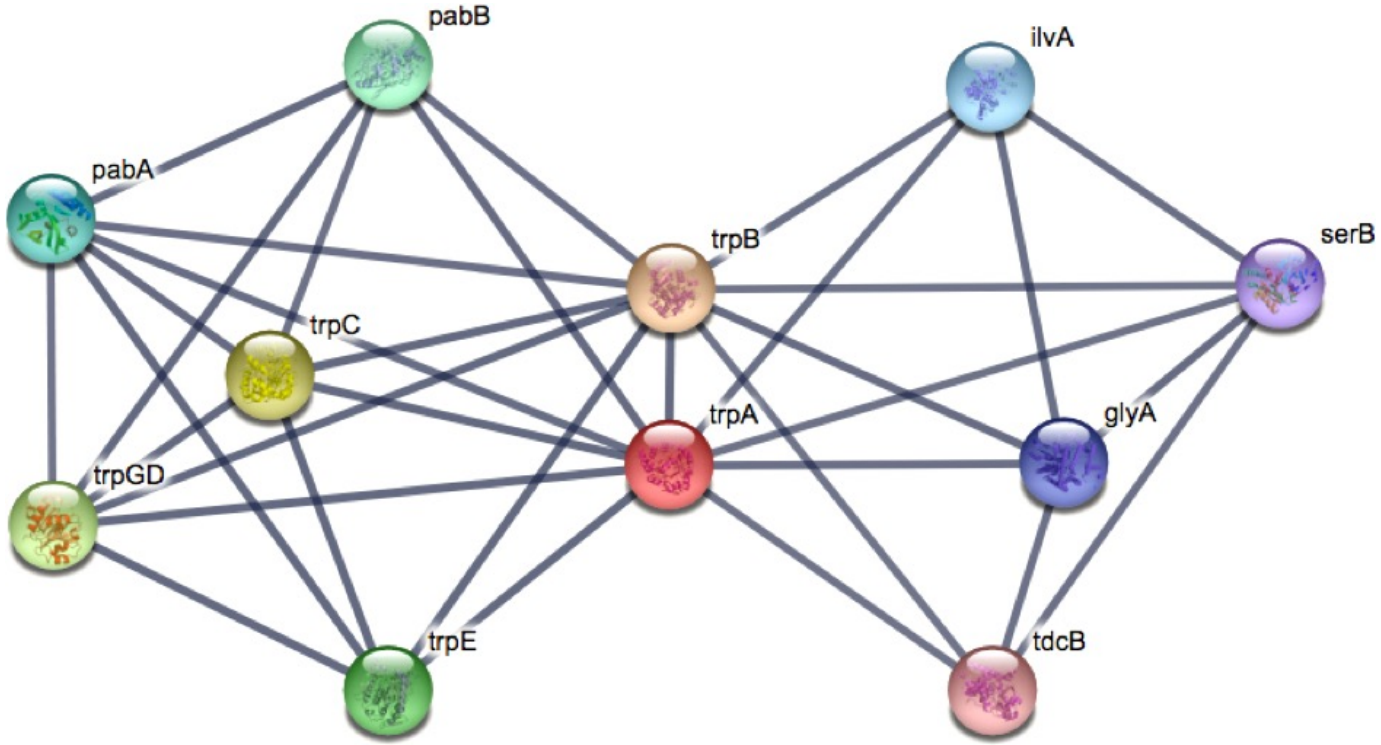

Jensen et al., *Nature Communications*, 2014

# food webs


Chesapeake Bay Waterbird Food Web
Tertiary Consumers:
Osprey
Bald Eagle
Secondary Consumers:
Gulls and Terns
Wading Birds
Large Piscivorous Fish
Sea Ducks
Tundra Swan
Primary Consumers:
Small Planktivorous Fish
Bivalves
Herbivores:
Zooplankton
Benthic Invertebrates
Herbivorous Ducks
Geese and Mute Swans
Primary Producers:
Phytoplankton
Submerged Aquatic Vegetation (SAV)
Vegetation

# protein networks

# The STRING database



14,094 genomes

67.6 million proteins

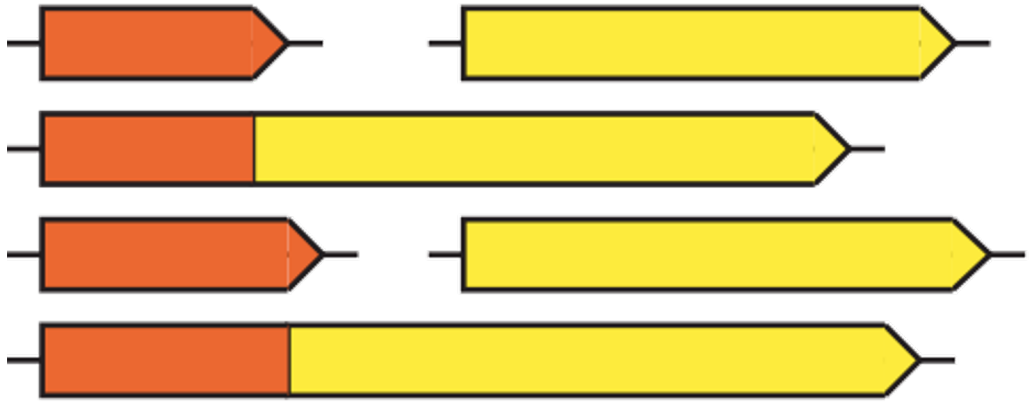
# physical interactions

# functional associations

# data integration

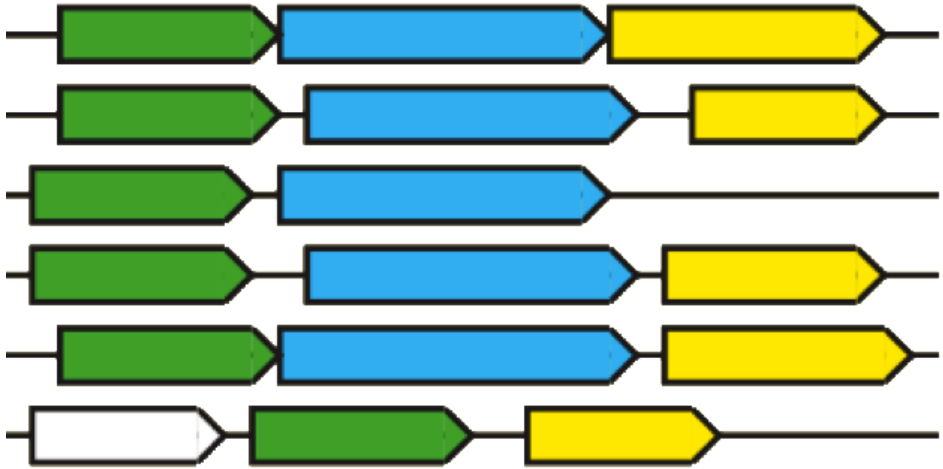
# genomic context

# gene fusion



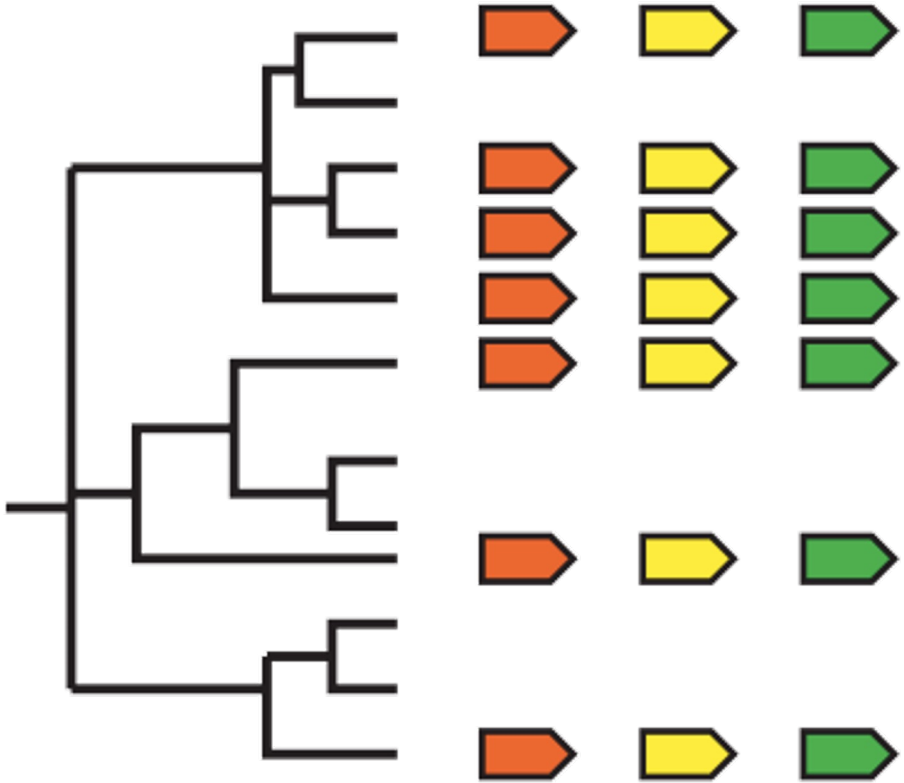
Korbel et al., *Nature Biotechnology*, 2004

# gene neighborhood



Korbel et al., *Nature Biotechnology*, 2004

# phylogenetic profiles

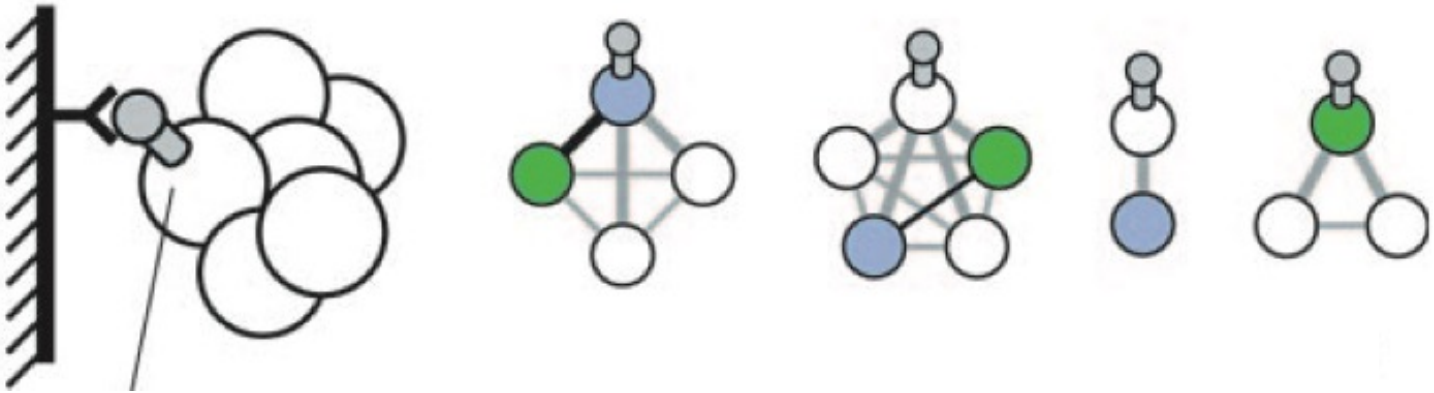


Korbel et al., *Nature Biotechnology*, 2004

experimental data

# gene coexpression

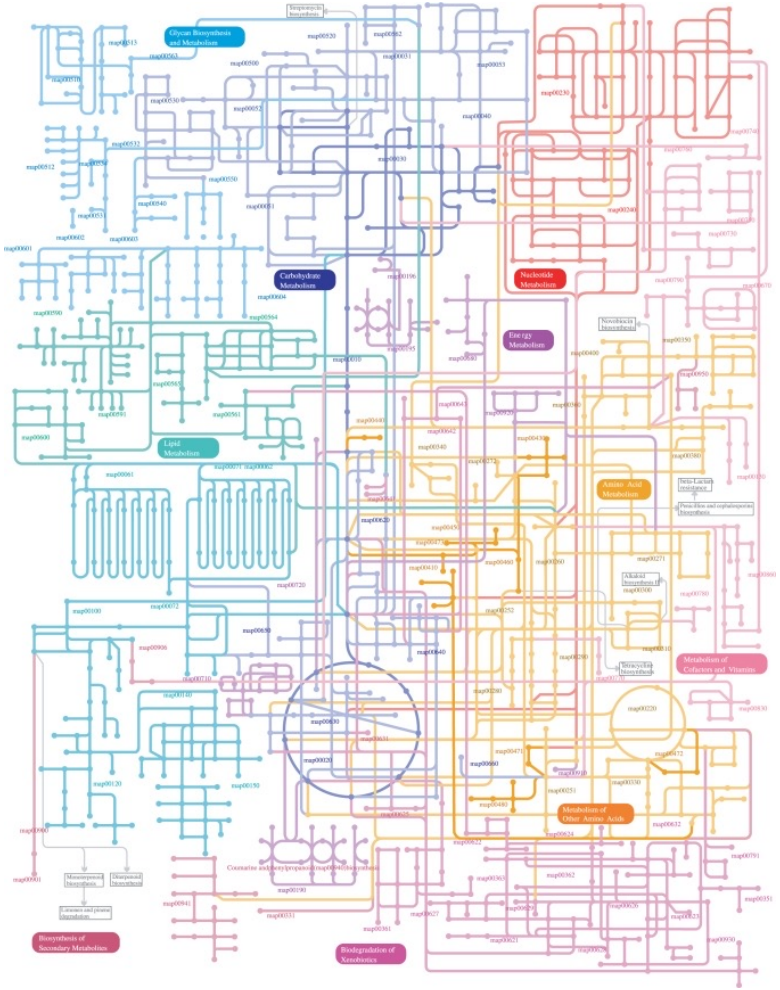
# interaction experiments



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

curated knowledge

# molecular pathways




Kanehisa et al., *Nucleic Acids Research*, 2008

# many databases

# different formats

different names

varying quality

not comparable

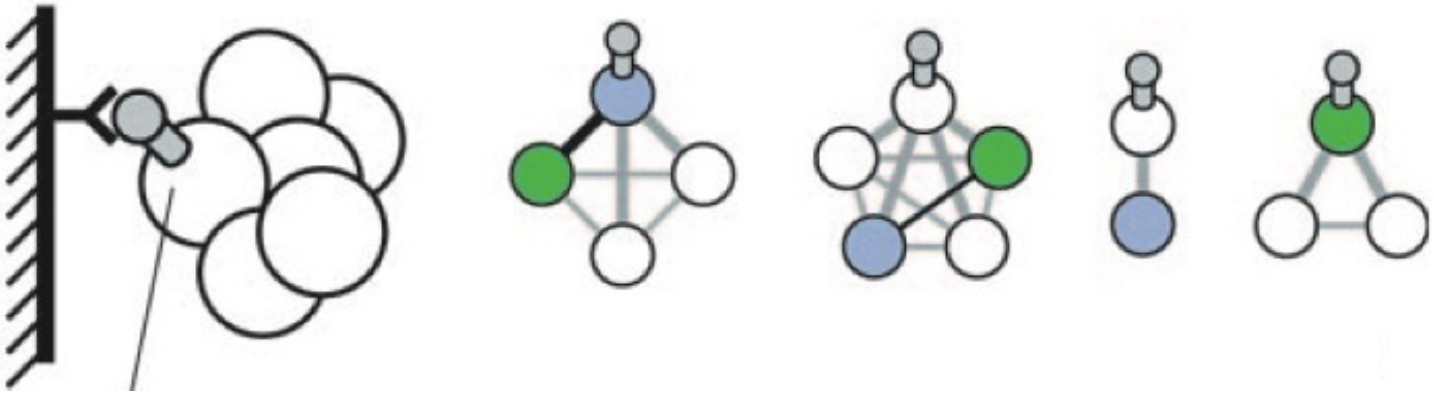
not same species

hard work

# parsers

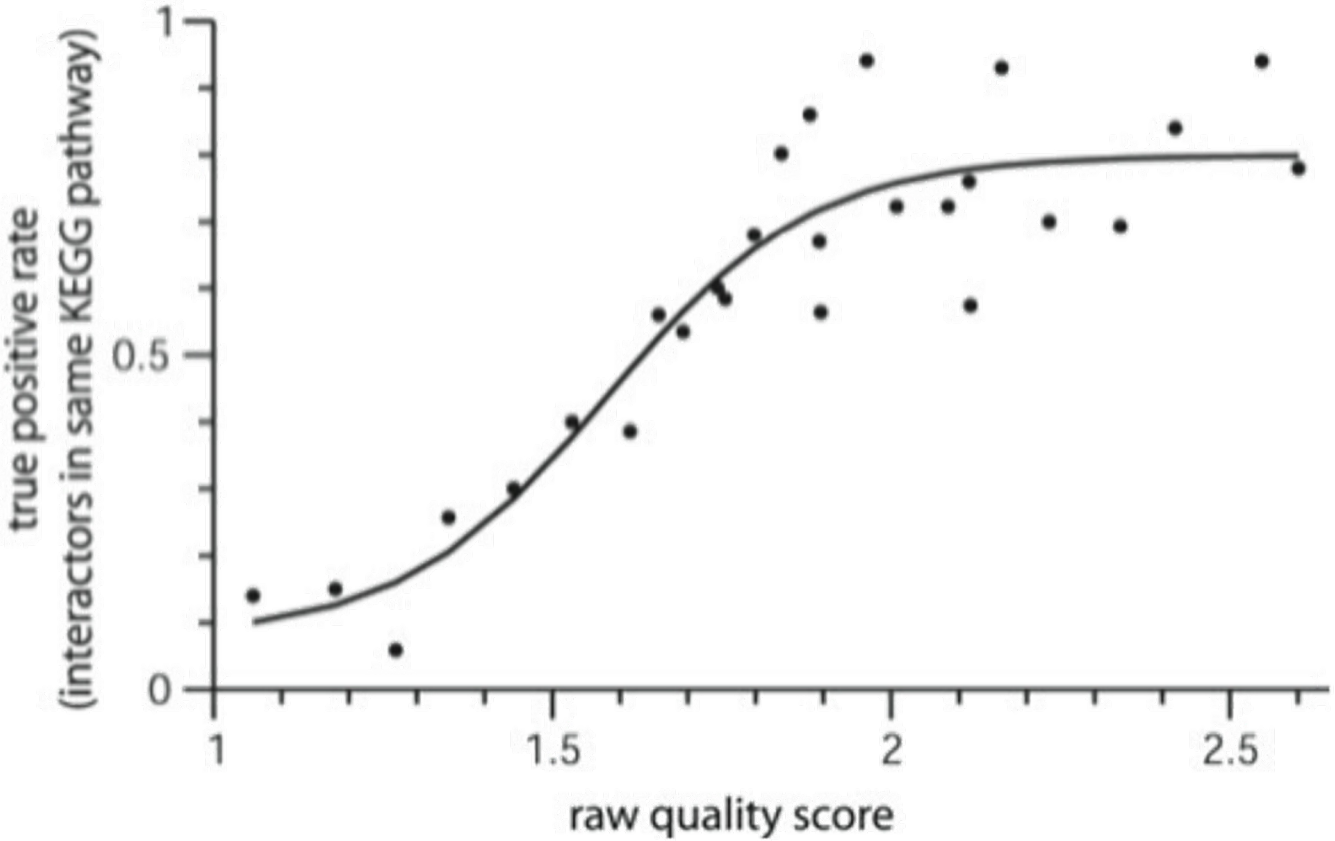
# mapping files

# scoring schemes



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

# score calibration




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

# transfer by orthology

missing most of the data

# text mining

>10 km

too much to read

computer

# named entity recognition

# dictionary

# cyclin dependent kinase 1

CDK1

# block list

# SDS

# co-mentions

# counting

# distance-weighted score

consensus view

no relation type

# deep learning

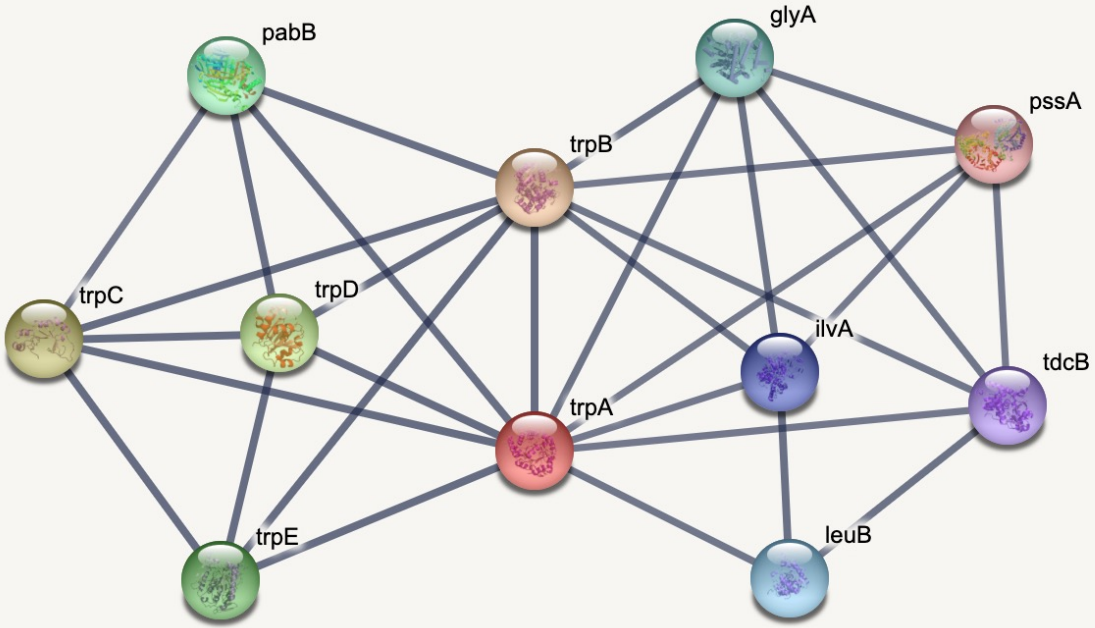
# pre-trained transformers

huge unlabeled corpora

finetune for specific task

# physical protein interactions

# practical session






Szklarczyk et al., *Nucleic Acids Research*, 2021

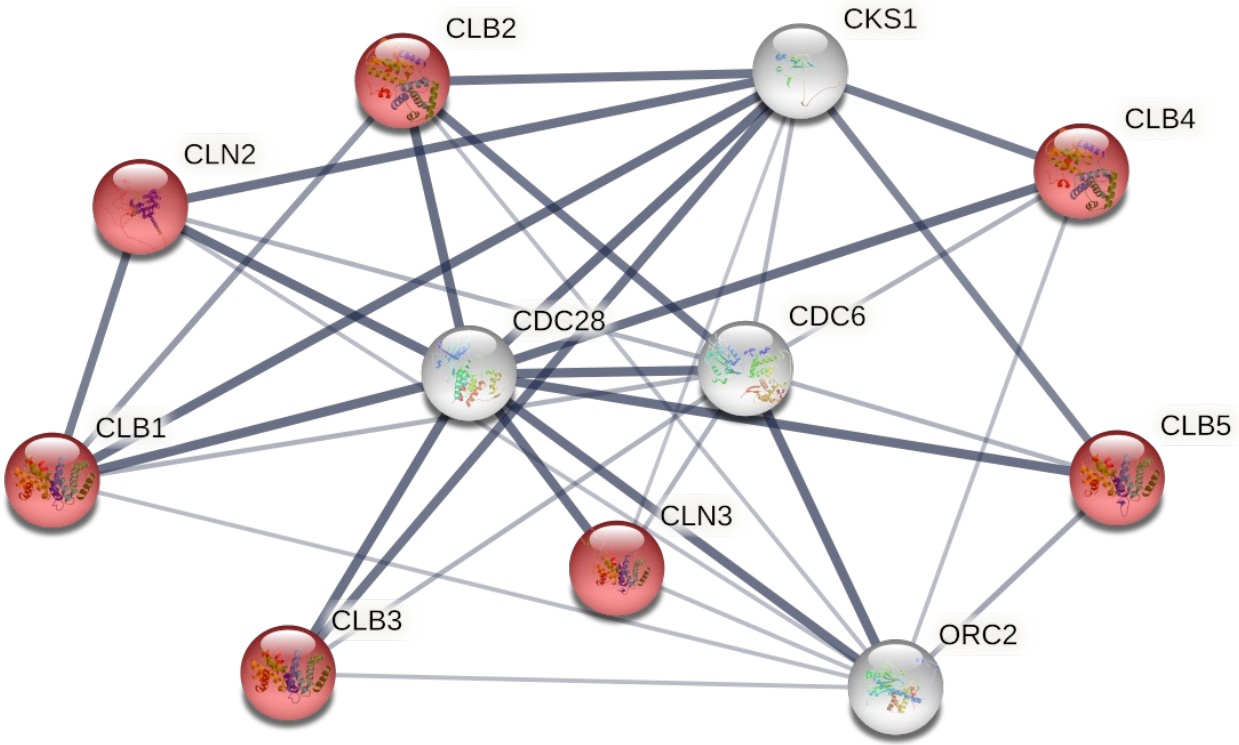
query the database

inspect the evidence

# exercise 1

jensenlab.org/training/string/

# Enrichment analysis



# characterize a gene list

# classes of genes

overrepresented

associated with the study

overrepresentation

# omics study

regulated proteins

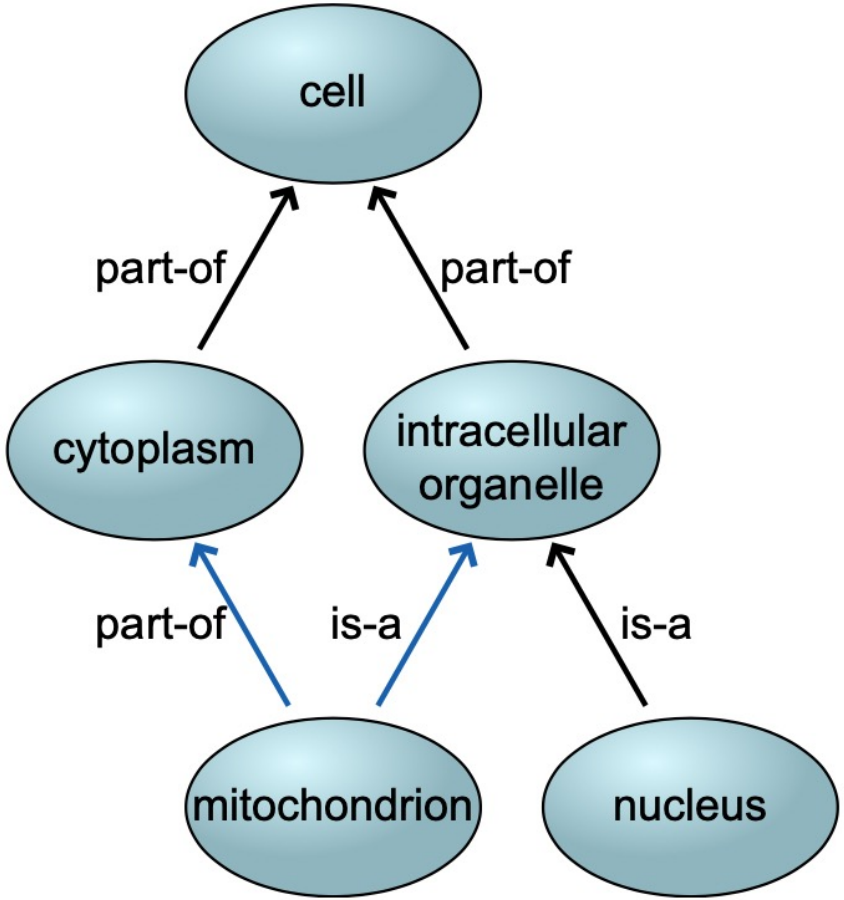
mitochondria enriched?

| | Mitochondrial | Not mitochondrial |
|---|---|---|
| Regulated | a | b |
| Not regulated | c | d |

# Fisher’s exact test

# p-value

# Gene Ontology




Jensen and Bork, *PLOS Biology*, 2010

systematically test all

many gene sets

# Gene Ontology

pathways

# protein domains

# UniProt keywords

# subcellular localizations

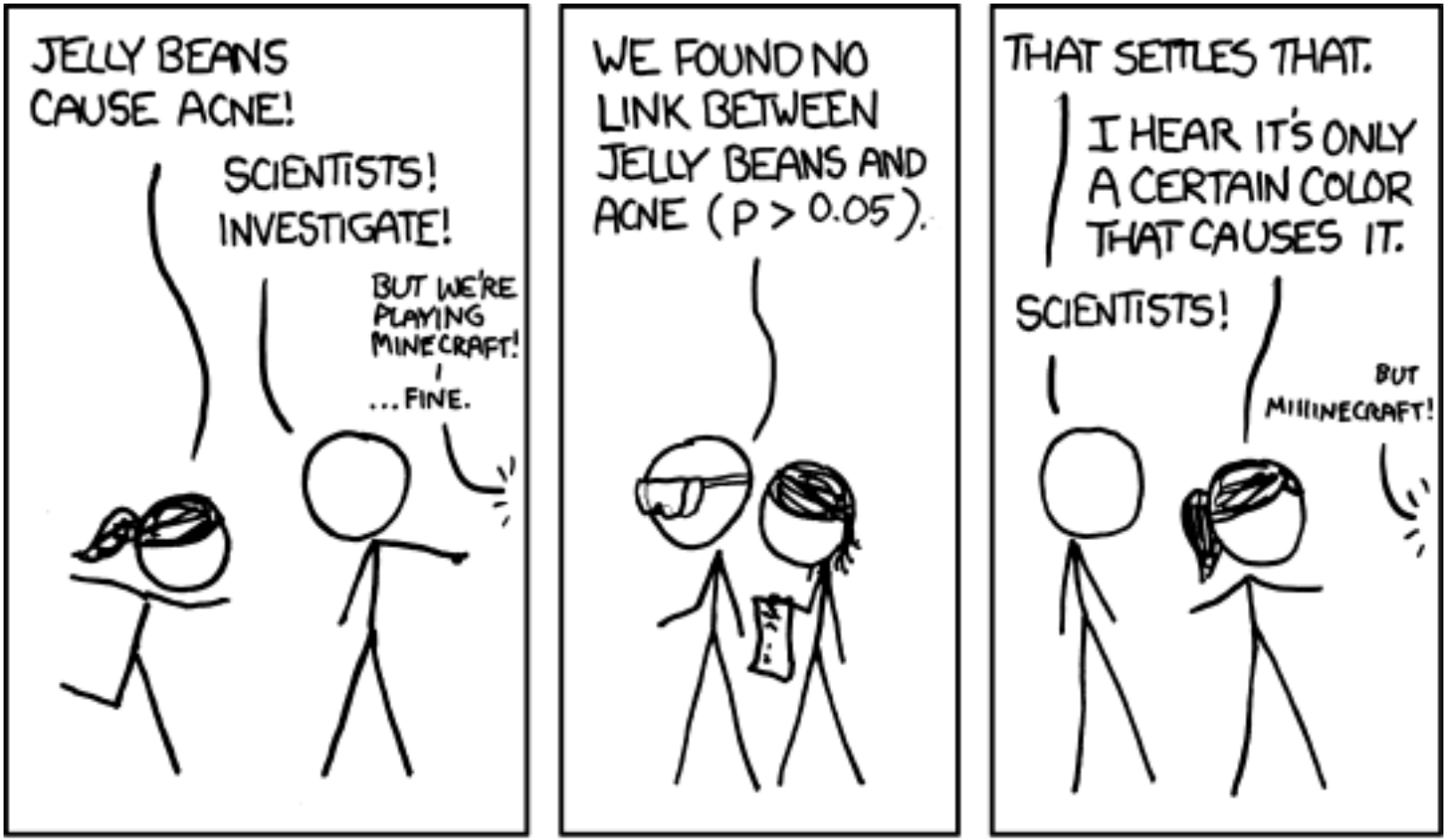
tissues

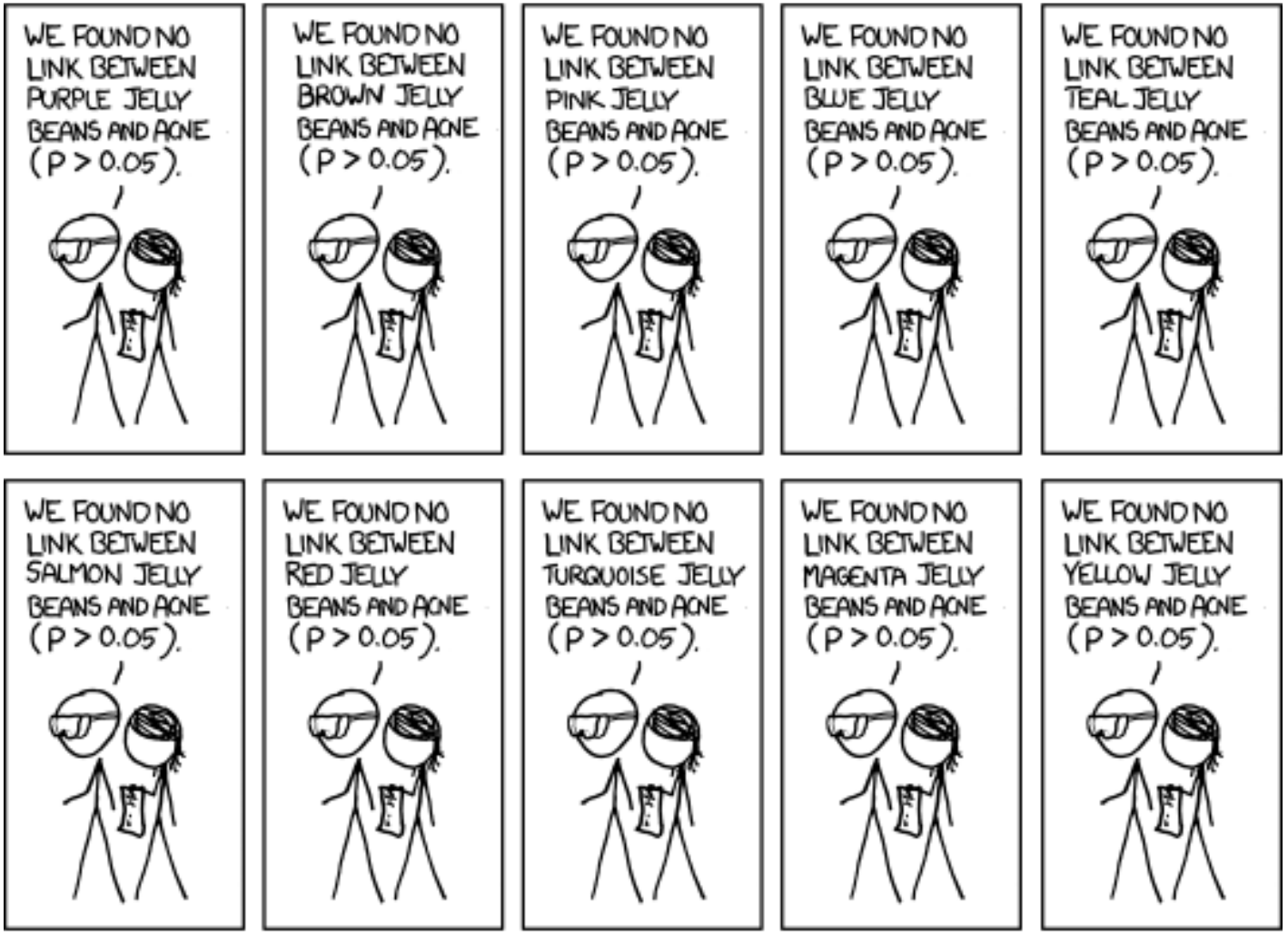
associate diseases

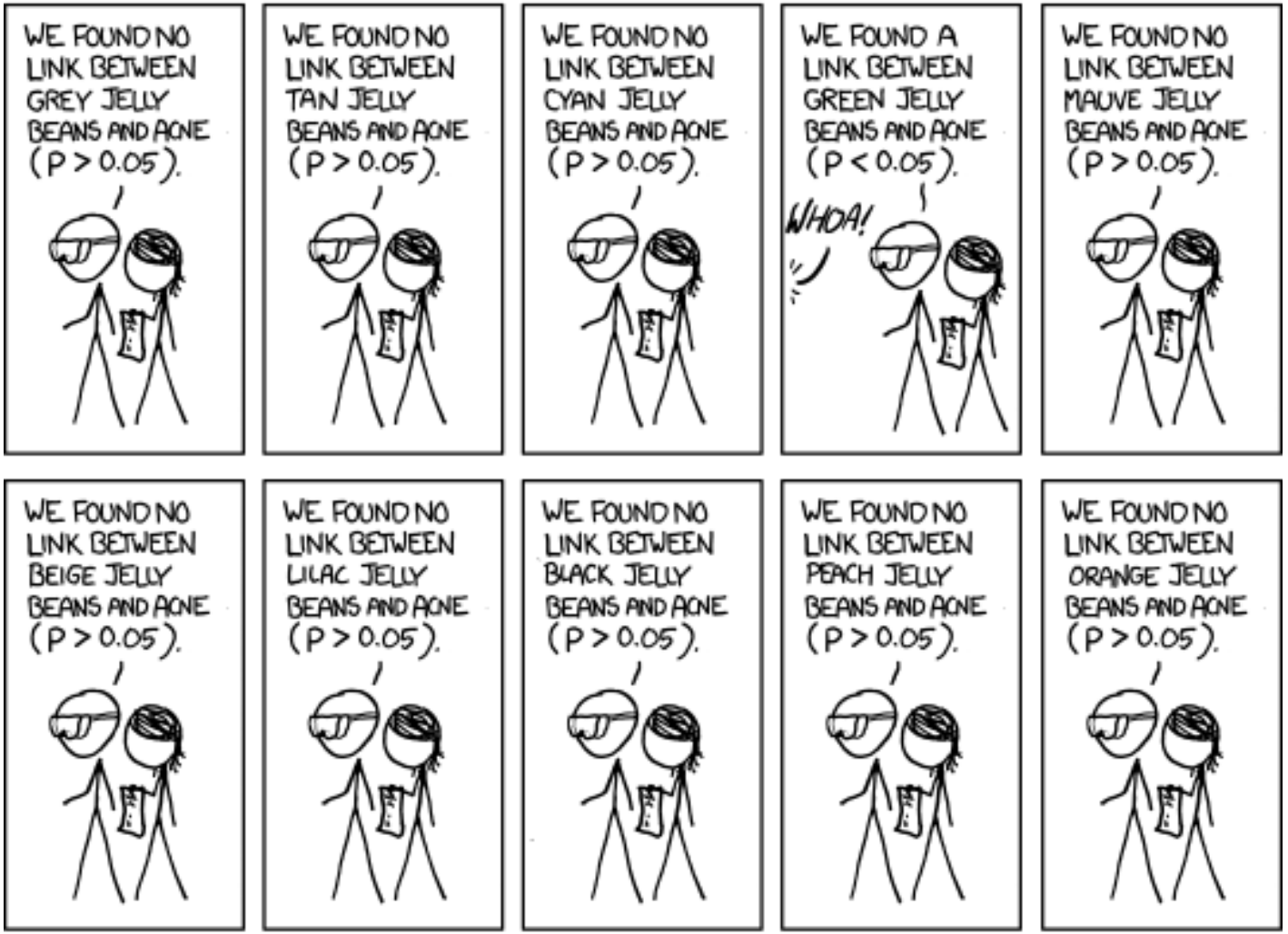
# local network clusters

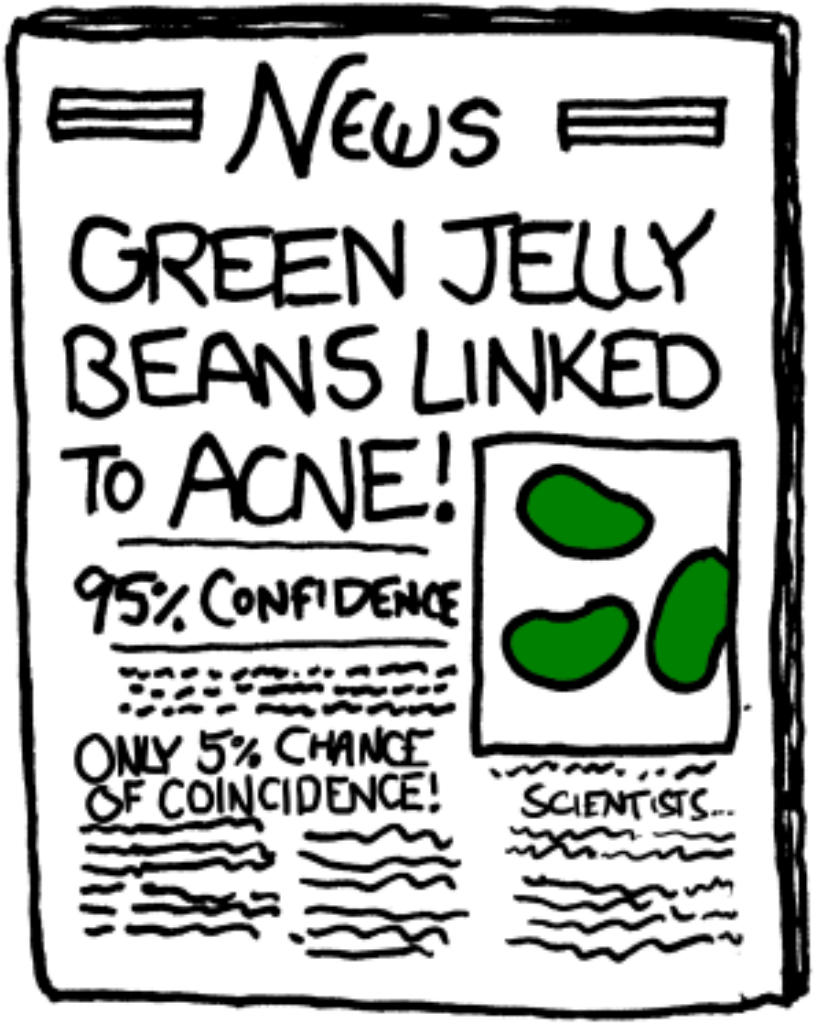
# publications

# multiple testing


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

# enrichment analysis

thousands of tests

crucial to correct

# Bonferroni correction

# false discovery rate

# practical session

# proteomics study

Protein by name
Protein by sequence
Multiple proteins
Multiple sequences
Proteins with Values/Ranks
Organisms
Protein families ("COGs")
Examples
Random entry

# SEARCH

## Multiple Proteins by Names / Identifiers

List Of Names: (one per line; examples: #1 #2 #3)

Q75VX8
P29353
O14492
P19174
Q07890
P62993

... or, upload a file:

Browse ...

Organism:

Homo sapiens

Advanced Settings

SEARCH

# protein network

# change query parameters

# functional enrichment

## KEGG Pathways

| pathway | description | count in network | strength | false discovery rate |
|---|---|---|---|---|
| hsa01521 | EGFR tyrosine kinase inhibitor resistance | 12 of 78 | 1.59 | 6.17e-13 |
| hsa04012 | ErbB signaling pathway | 12 of 83 | 1.56 | 6.17e-13 |
| hsa04722 | Neurotrophin signaling pathway | 13 of 114 | 1.46 | 6.17e-13 |
| hsa04650 | Natural killer cell mediated cytotoxicity | 13 of 121 | 1.43 | 6.57e-13 |
| hsa05214 | Glioma | 11 of 72 | 1.58 | 2.17e-12 |

*(more ...)*

## Reactome Pathways

| pathway | description | count in network | strength | false discovery rate |
|---|---|---|---|---|
| HSA-8853659 | RET signaling | 12 of 40 | 1.88 | 2.00e-14 |
| HSA-168249 | Innate Immune System | 29 of 1025 | 0.85 | 1.44e-13 |
| HSA-177929 | Signaling by EGFR | 11 of 50 | 1.74 | 3.83e-12 |
| HSA-9006335 | Signaling by Erythropoietin | 9 of 25 | 1.96 | 3.20e-11 |
| HSA-2424491 | DAP12 signaling | 9 of 28 | 1.91 | 6.02e-11 |

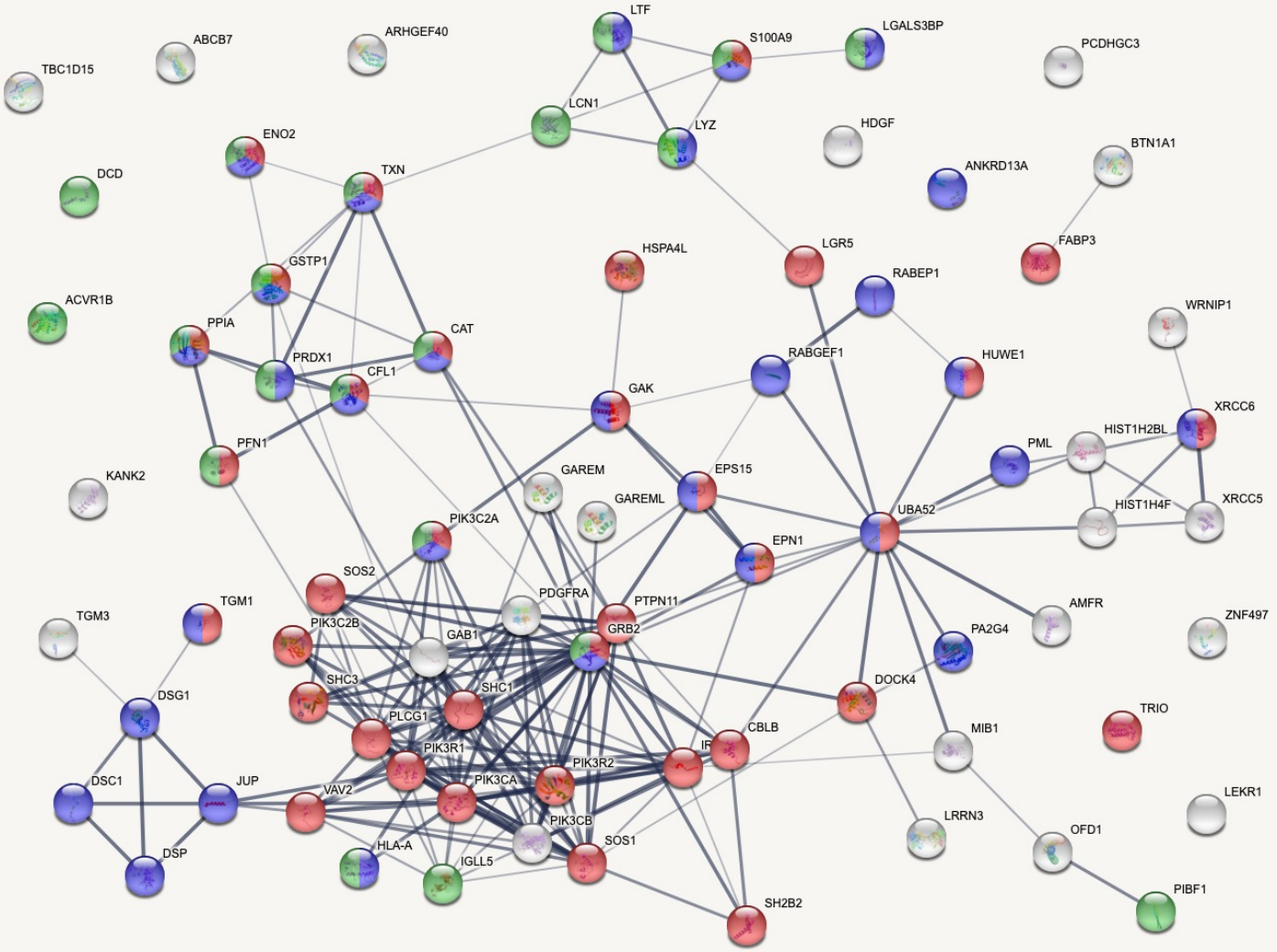
*(more ...)*

## WikiPathways

| pathway | description | count in network | strength | false discovery rate |
|---|---|---|---|---|
| WP437 | EGF/EGFR signaling pathway | 15 of 162 | 1.37 | 4.14e-13 |
| WP481 | Insulin signaling | 15 of 159 | 1.37 | 4.14e-13 |
| WP4806 | EGFR tyrosine kinase inhibitor resistance | 12 of 83 | 1.56 | 1.56e-12 |
| WP673 | ErbB signaling pathway | 12 of 90 | 1.52 | 2.82e-12 |
| WP2037 | Prolactin signaling pathway | 11 of 76 | 1.56 | 1.43e-11 |

*(more ...)*

show on network


LTF
ABCB7
ARHGEF40
S100A9
LGALS3BP
TBC1D15
PCDHGC3
LCN1
LYZ
HDGF
ENO2
BTN1A1
DCD
TXN
ANKRD13A
FABP3
HSPA4L
LGR5
GSTP1
RABEP1
ACVR1B
WRNIP1
PPIA
CAT
RABGEF1
HUWE1
PRDX1
CFL1
GAK
XRCC6
PML
HIST1H2BL
PFN1
KANK2
GAREM
EPS15
GAREML
HIST1H4F
XRCC5
UBA52
PIK3C2A
EPN1
SOS2
PDGFRA
PTPN11
AMFR
ZNF497
TGM3
TGM1
PIK3C2B
GAB1
GRB2
PA2G4
SHC3
SHC1
DOCK4
DSG1
TRIO
PLCG1
CBLB
MIB1
PIK3R1
IR
PIK3R2
DSC1
JUP
PIK3CA
VAV2
LEKR1
PIK3CB
LRRN3
HLA-A
SOS1
OFD1
DSP
IGLL5
PIBF1
SH2B2

# exercise 2

jensenlab.org/training/string/