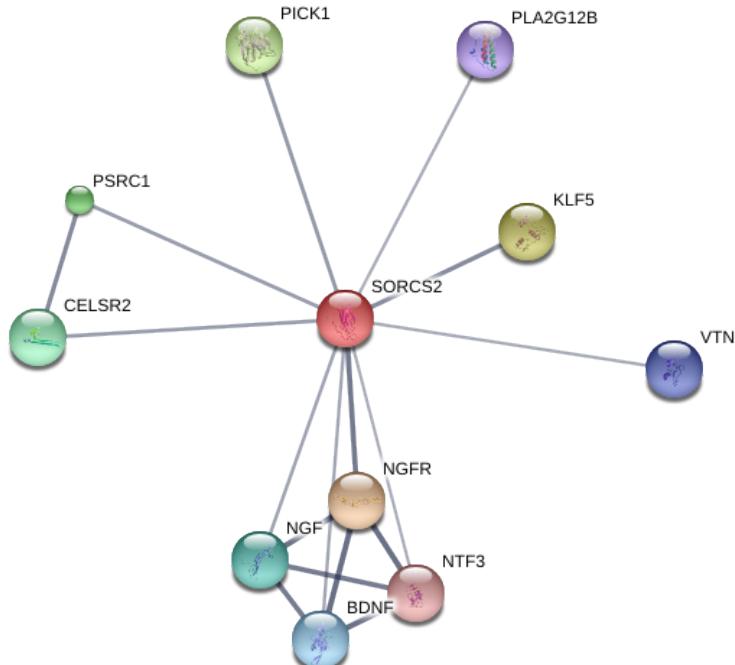


# The STRING database

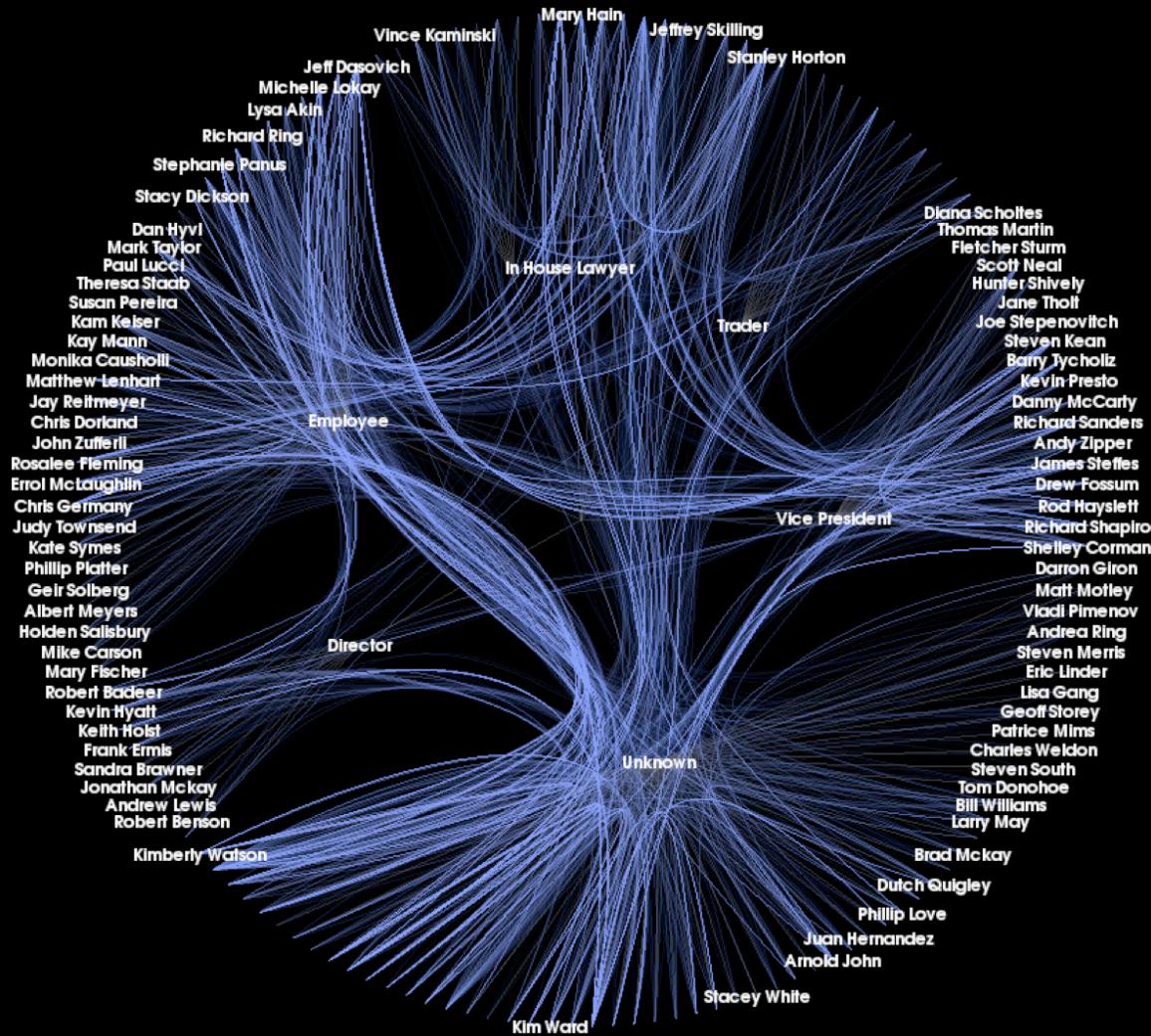


Lars Juhl Jensen  
[jensenlab.org](http://jensenlab.org)

# interaction networks

# association networks

guilt by association



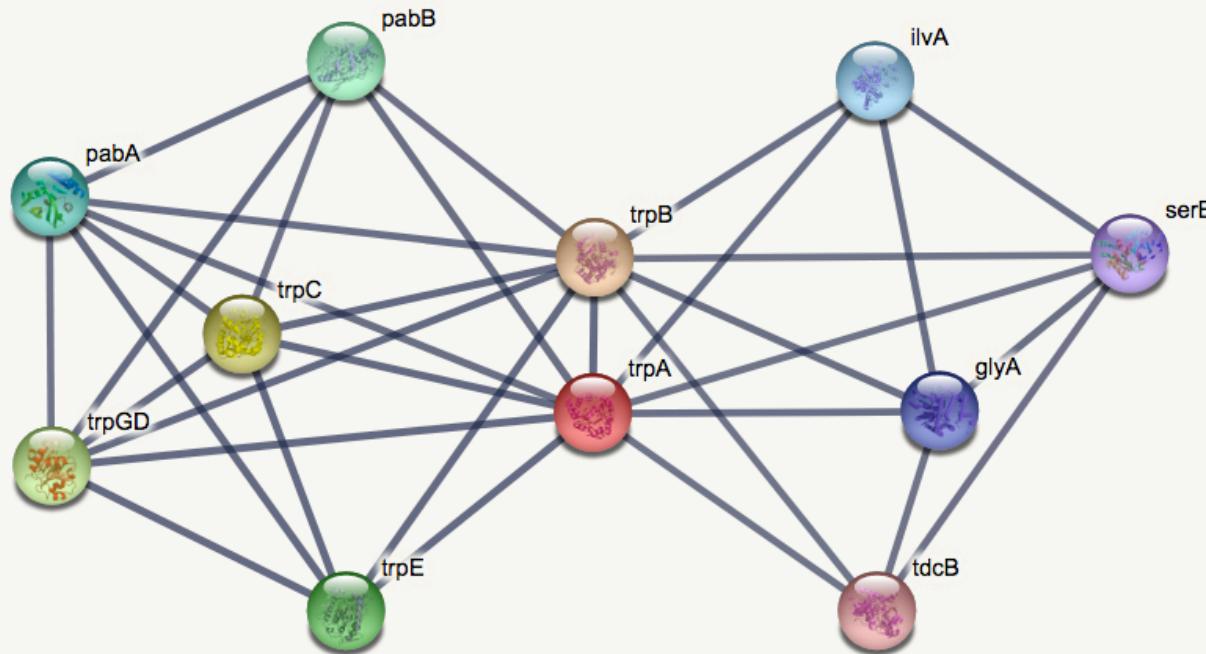
protein networks

STRING

2031 genomes

9.6 million proteins

# functional associations



heavily used

4000 users daily

● Users

20,000

10,000

2011

2012

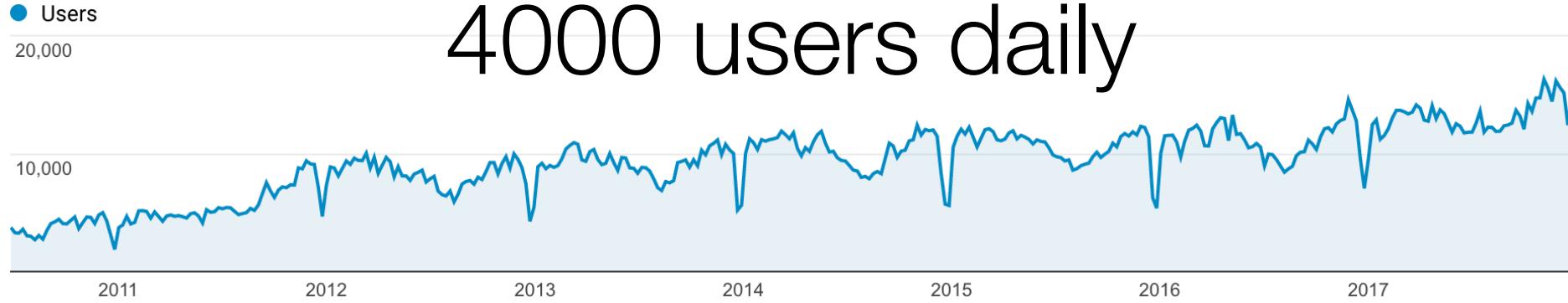
2013

2014

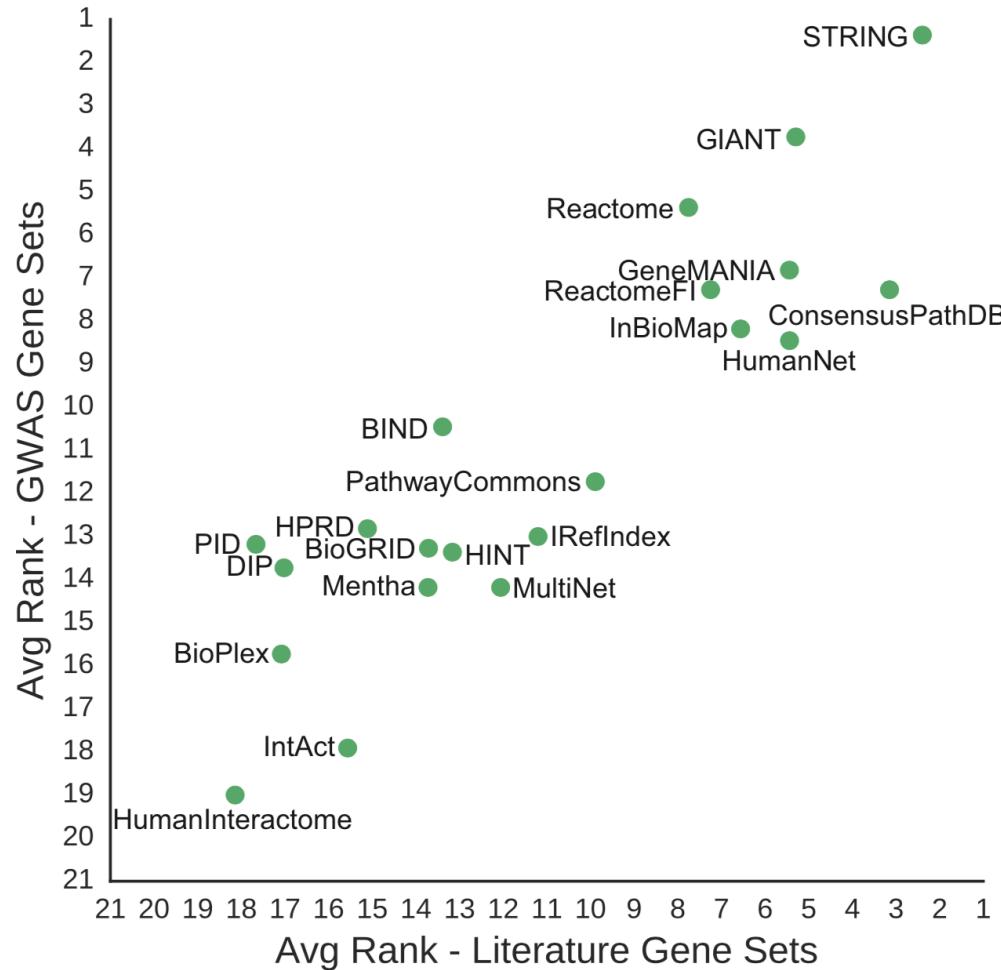
2015

2016

2017



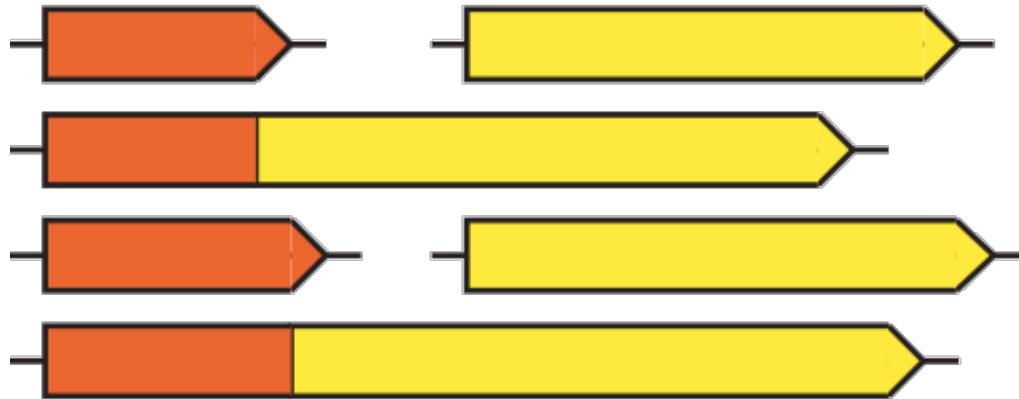
works well



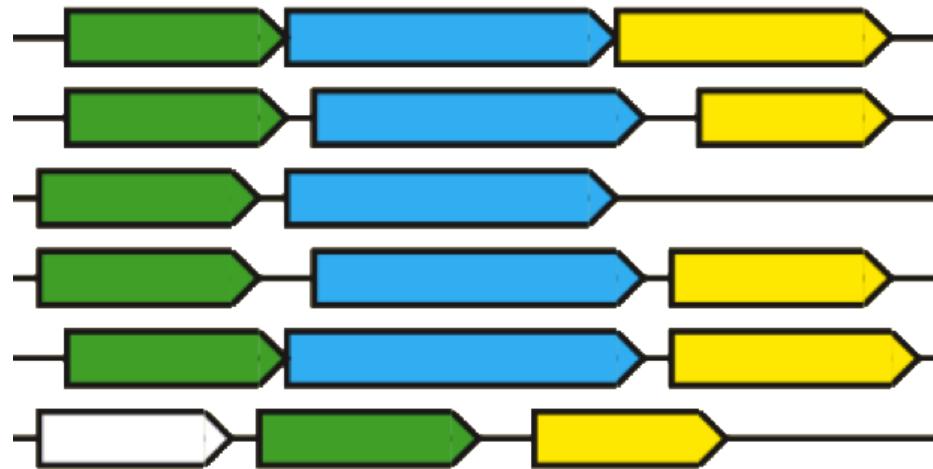
# data integration

genomic context

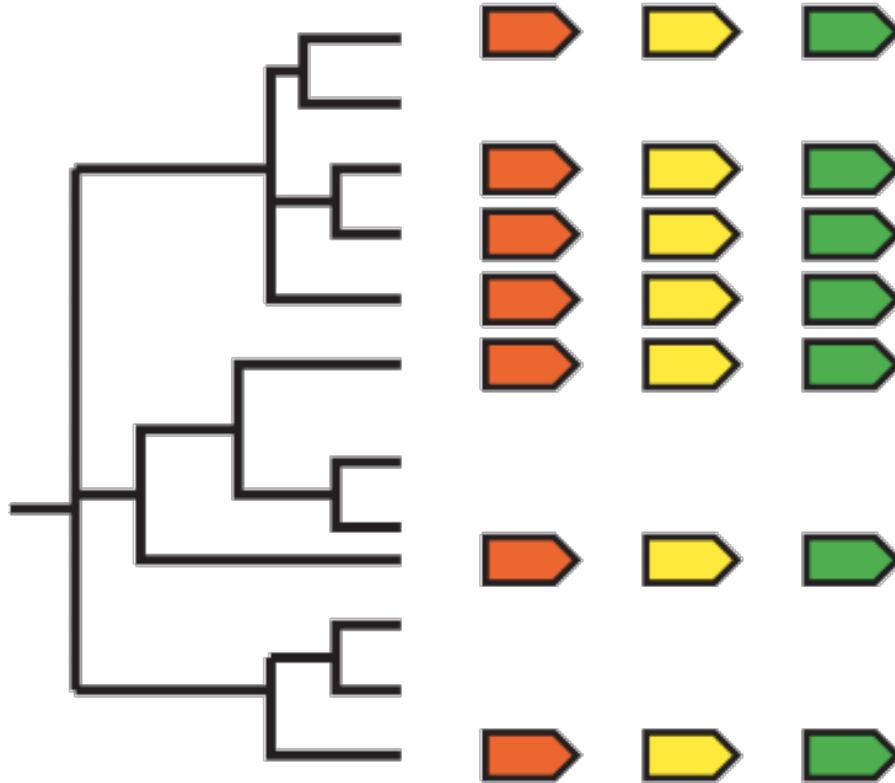
gene fusion



gene neighborhood



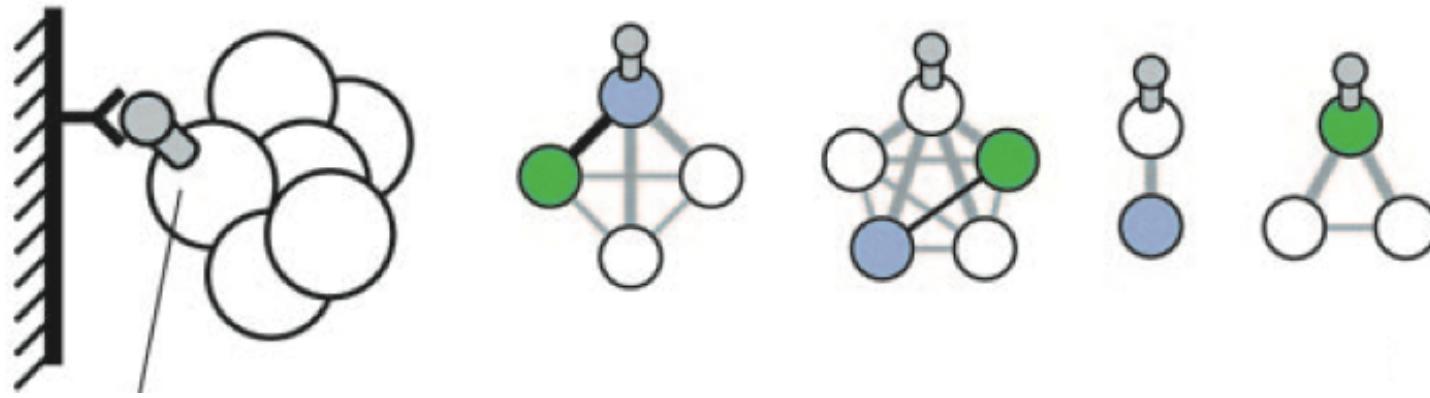
phylogenetic profiles



experimental data

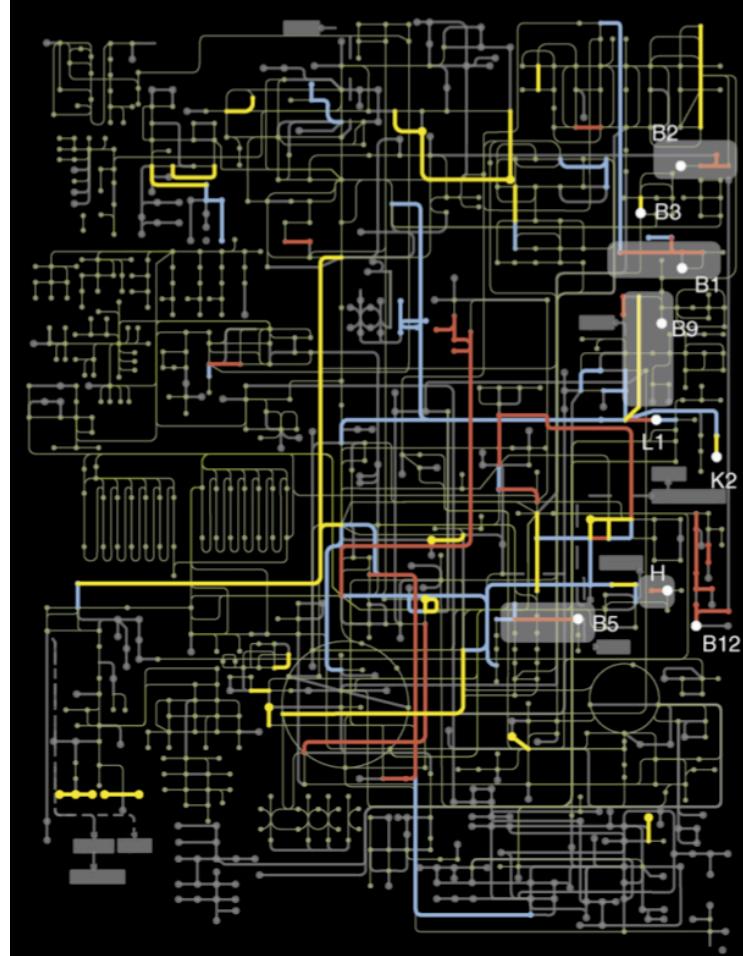
gene coexpression

physical interactions



curated knowledge

pathways



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

many databases

different formats

different identifiers

varying quality

not comparable

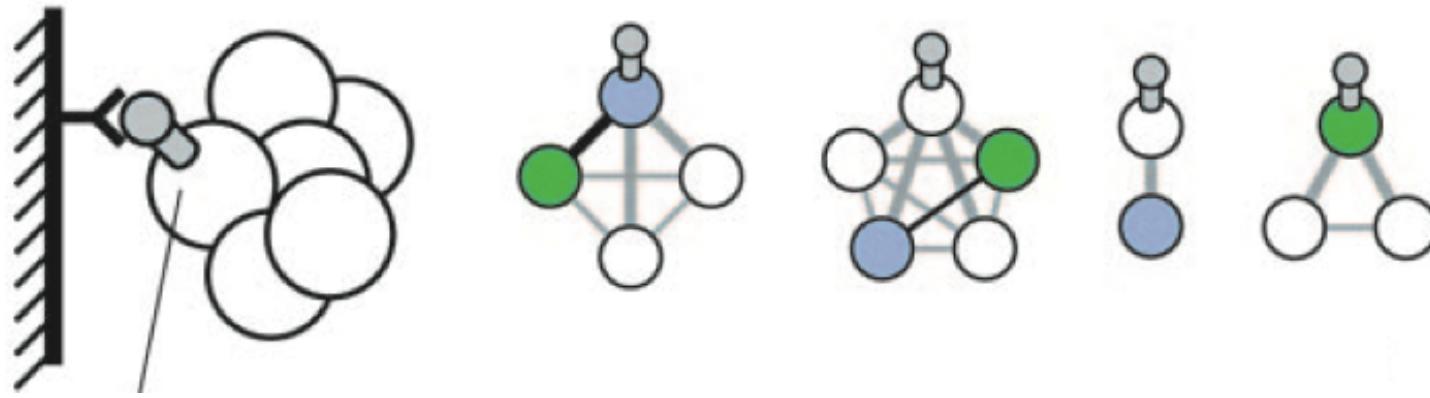
not same species

hard work

# parsers

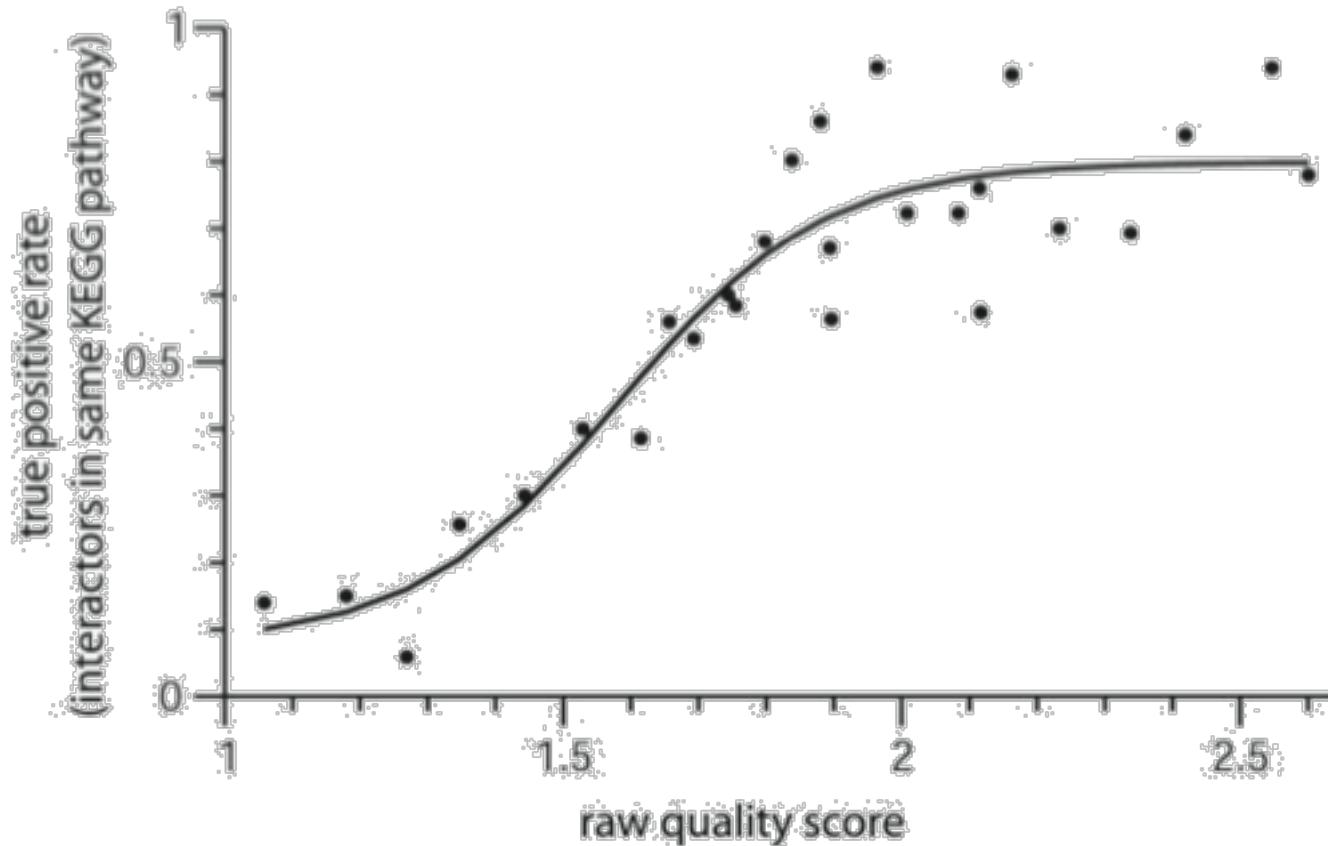
mapping files

quality scores



score calibration

gold standard

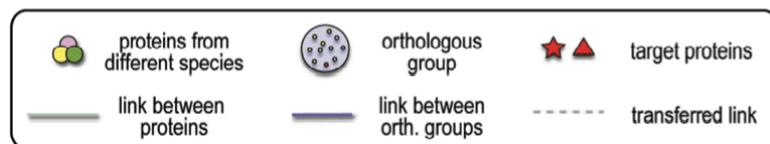
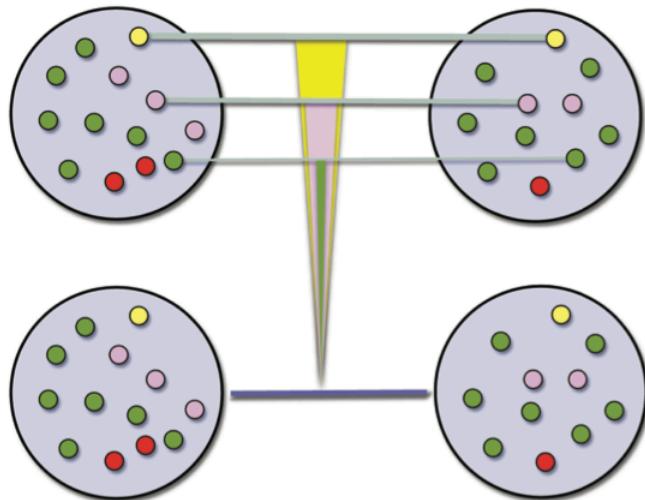


# common identifiers

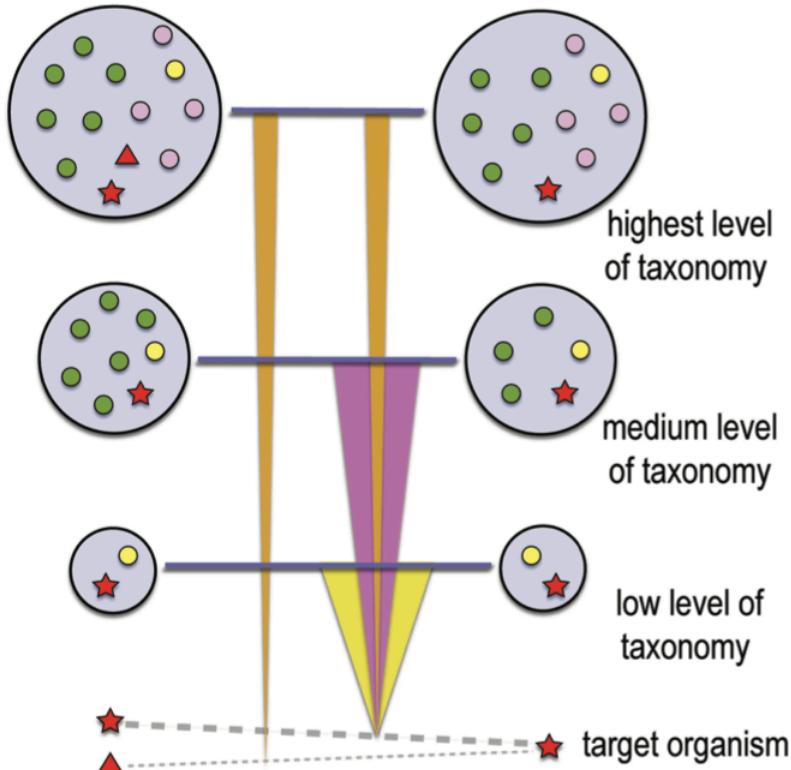
common scale

orthology transfer

### Step 1: Combining individual protein-protein links into links between orthologous groups.



### Step 2: Transferring orthologous group links back to the protein level



cross-species integration

missing most of the data

**text mining**

>10 km



too much to read

named entity recognition

dictionary

synonyms

cyclin dependent kinase 1

CDC2

orthographic variation

cyclin dependent kinase 1

cyclin-dependent kinase 1

CDC2

hCDC2

black list

SDS

co-mentioning

$$C_{ij} = \sum_{k=1}^n \delta_{dijk} w_d + \delta_{pijk} w_p + \delta_{sijk} w_s$$

$$S_{ij} = C_{ij}^\alpha \left( \frac{C_{ij} C_{..}}{C_{i..} C_{..j}} \right)^{1-\alpha}$$

score calibration

orthology transfer

# related resources

# chemical compounds

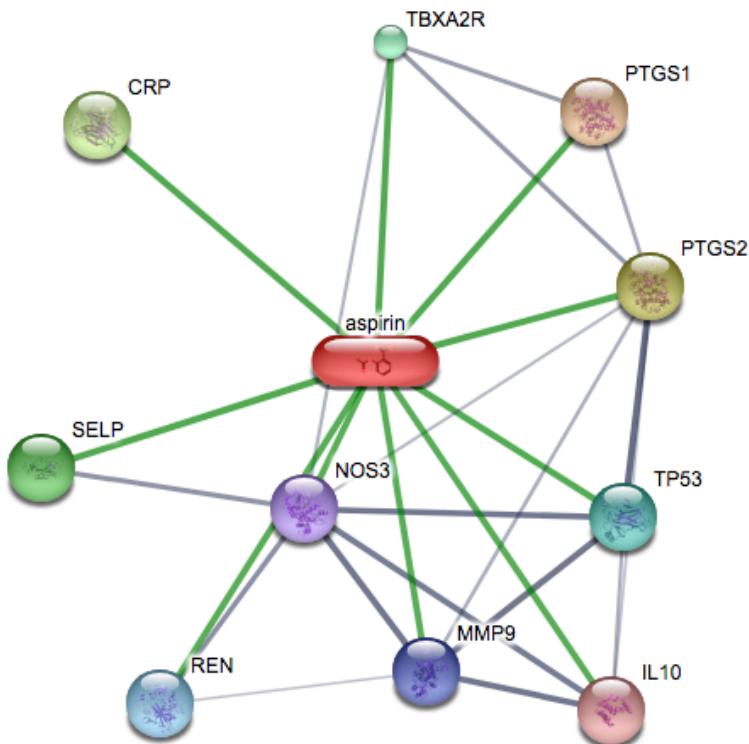


Search

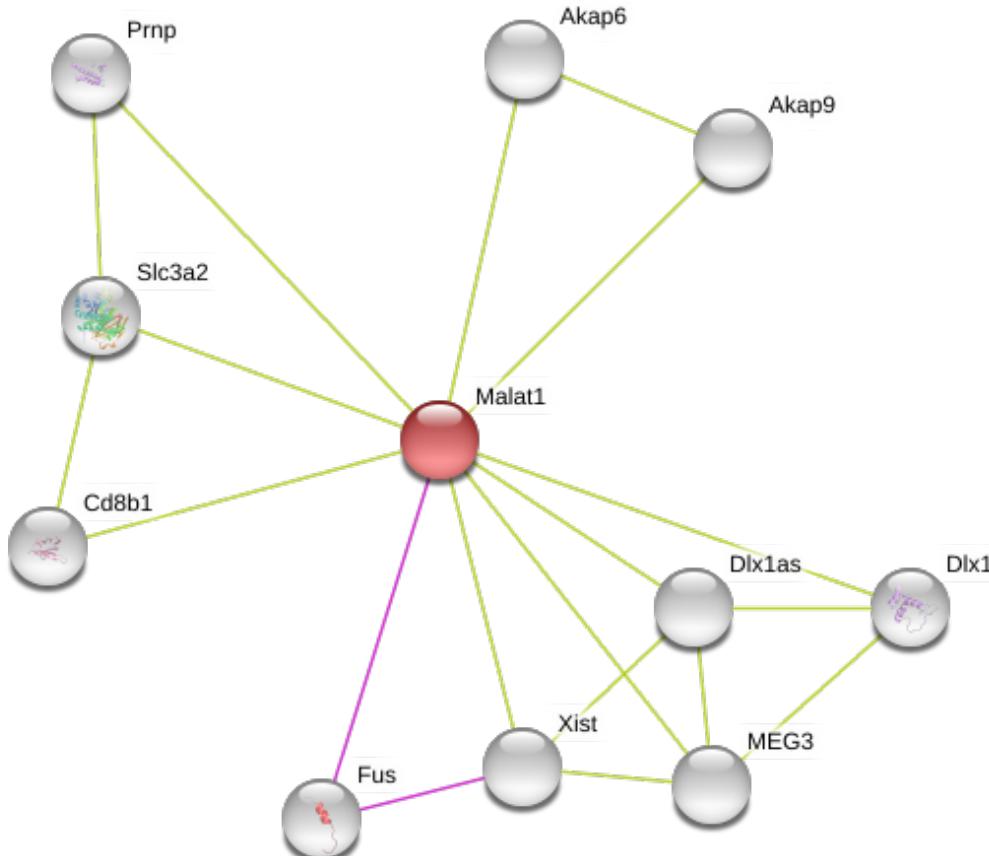
Download

Help

My Data

Szklarczyk et al., *Nucleic Acids Research*, 2016

# ncRNAs



viruses

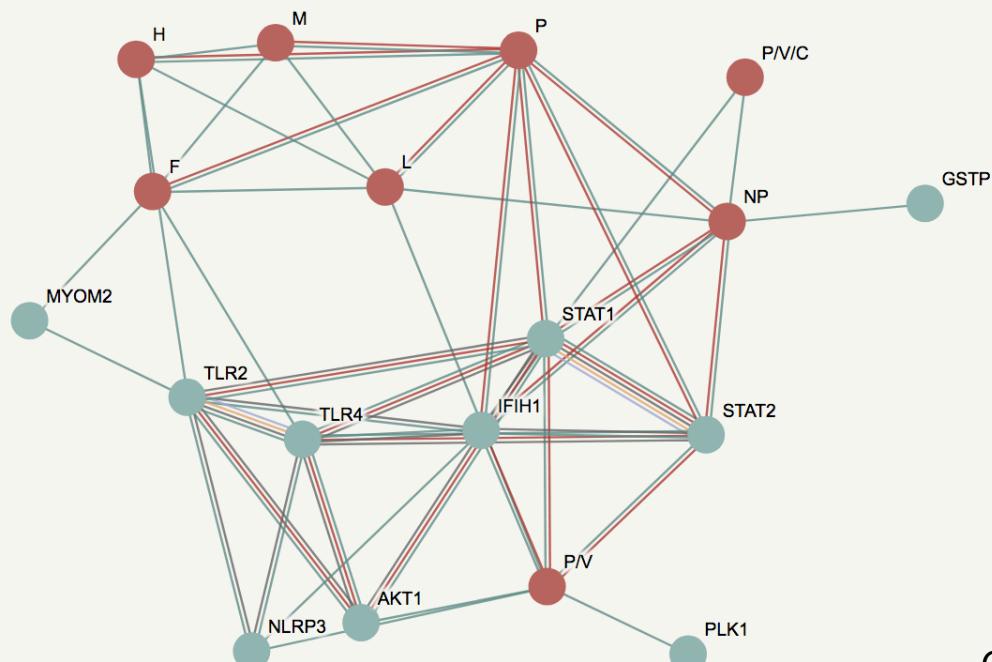
 Measles virus (strain Ichinose-B95a)

Measles virus (strain Ichinose-B95a)

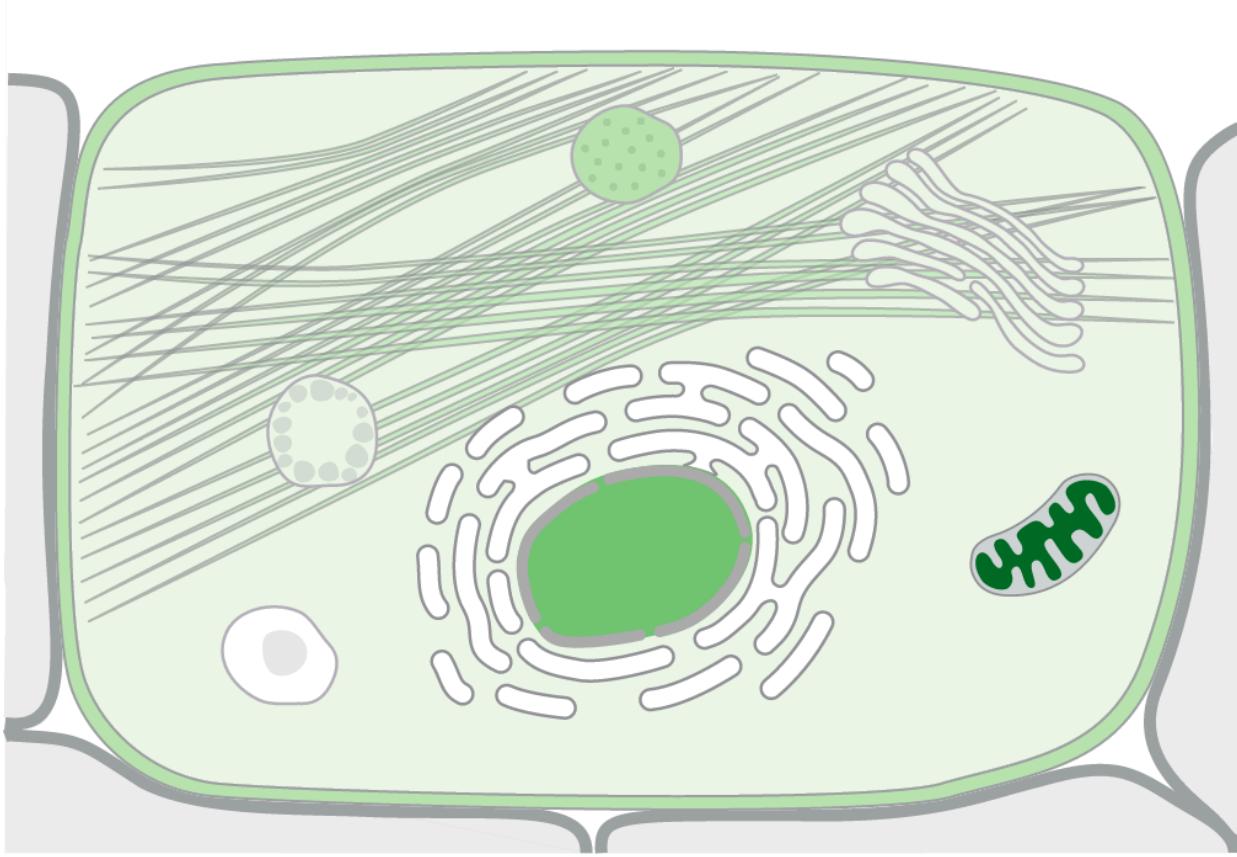
NCBI taxonomy Id: 11234

 Homo sapiens

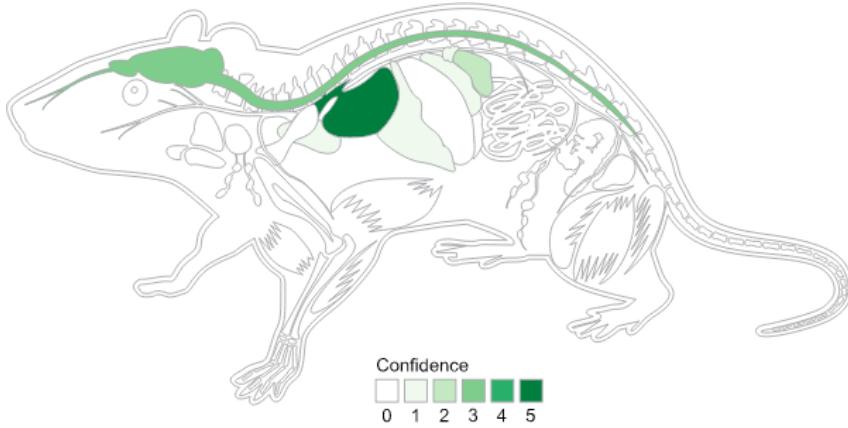
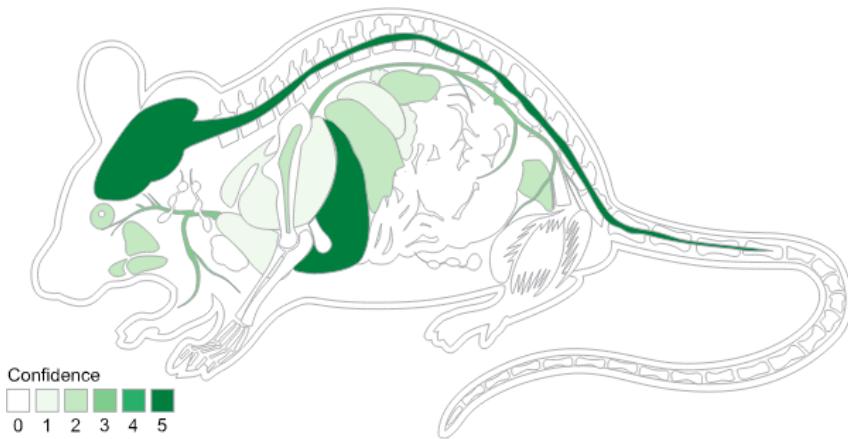
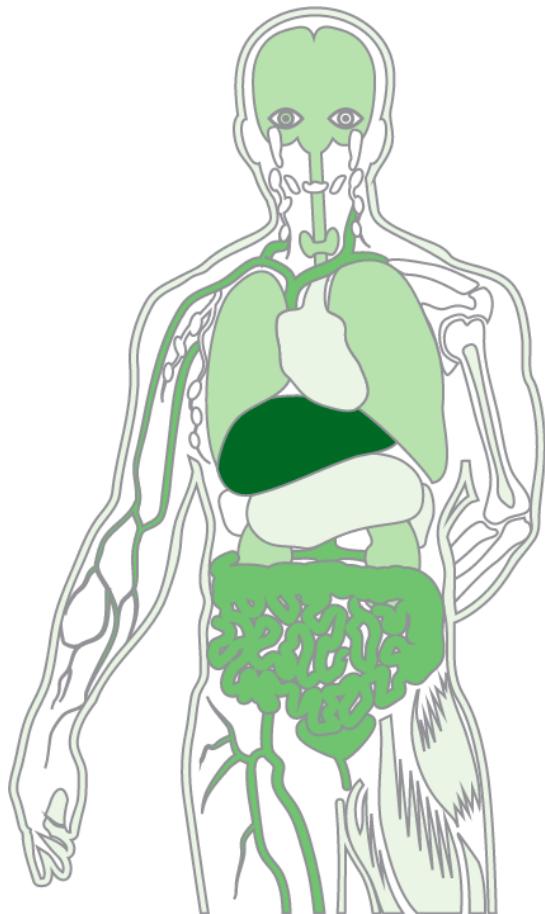
NCBI taxonomy Id: 9606



subcellular localization



tissue expression



disease associations

LRRK2 [ENSP00000298910]

Leucine-rich repeat kinase 2

Synonyms: LRRK2, AURA17, DKFZp686E15222, LRRK2p, hLRRK2 ...

## Text mining

Next >

Name	Z-score	Confidence
Parkinson's disease	6.4	★★★★★
Movement disease	4.0	★★★★★
Lewy body dementia	3.9	★★★★★
Multiple system atrophy	3.0	★★★★★
Leprosy	2.3	★★★★★
Alzheimer's disease	2.2	★★★★★
Frontotemporal dementia	2.2	★★★★★
Crohn's disease	2.0	★★★★★
Toxic encephalopathy	1.9	★★★★★
Amyotrophic lateral sclerosis	1.4	★★★★★

## Knowledge

Name	Source	Evidence	Confidence
Neurodegenerative disease	UniProtKB-KW	CURATED	★★★★★
Parkinson's disease	GHR	CURATED	★★★★★
Parkinson's disease	UniProtKB-KW	CURATED	★★★★★

## Experiments

Name	Source	Evidence	Confidence
Parkinson's disease	DistiLD	p-value = 2e-28	★★★★★
Crohn's disease	DistiLD	p-value = 3e-10	★★★★★
Carcinoma	COSMIC	108 samples	★★★★★
Lung cancer	COSMIC	34 samples	★★★★★
Kidney cancer	COSMIC	19 samples	★★★★★
Large intestine cancer	COSMIC	19 samples	★★★★★
Ovarian cancer	COSMIC	12 samples	★★★★★
Endometrial cancer	COSMIC	12 samples	★★★★★

## Parkinson's disease [DOID:14330]

[close]

A neurodegenerative disease that has\_material\_basis\_in degeneration of the central nervous system that often impairs the sufferer's motor skills, speech, and other functions.

Synonyms: Parkinson's disease, DOID:14330, Parkinsons disease, Parkinson's disorder, Parkinson's syndrome ...

< Prev | Next >

### Pharmacological rescue of mitochondrial deficits in iPSC-derived neural cells from patients with familial Parkinson's disease.

Cooper O, Seo H, Andрабi S, (and 36 more) ; *Sci Transl Med* (2012); PMID: 22764206

**Parkinson's disease** (PD) is a common neurodegenerative disorder caused by genetic and environmental factors that results in degeneration of the nigrostriatal dopaminergic pathway in the brain. We analyzed neural cells generated from induced pluripotent stem cells (iPSCs) derived from PD patients and presymptomatic individuals carrying mutations in the PINK1 (PTEN-induced putative kinase 1) and LRRK2 (leucine-rich repeat kinase 2) genes, and compared them to those of healthy control subjects. We measured several aspects of mitochondrial responses in the iPSC-derived neural cells including production of reactive oxygen species, mitochondrial respiration, proton leakage, and intraneuronal movement of mitochondria. Cellular vulnerability associated with mitochondrial dysfunction in iPSC-derived neural cells from familial PD patients and at-risk individuals could be rescued with coenzyme Q(10), rapamycin, or the LRRK2 kinase inhibitor GW5074. Analysis of mitochondrial responses in iPSC-derived neural cells from PD patients carrying different mutations provides insight into convergence of cellular disease mechanisms between different familial forms of PD and highlights the importance of oxidative stress and mitochondrial dysfunction in this neurodegenerative disease.

### The Gly2019Ser mutation in LRRK2 is not fully penetrant in familial Parkinson's disease: the GenePD study.

Latourelle JC, Sun M, Lew MF, (and 46 more) ; *BMC Med* (2008); PMID: 18986508

[View abstract]

### (G2019S) LRRK2 activates MKK4-JNK pathway and causes degeneration of SN dopaminergic neurons in a transgenic mouse model of PD.

Chen CY, Weng YH, Chien KY, (and 5 more) ; *Cell Death Differ* (2012); PMID: 22539006

[View abstract]

### Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies.

Nalls MA, Plagnol V, Hernandez DG, (and 15 more) ; *Lancet* (2011); PMID: 21292315

[View abstract]

### Genome-wide association study reveals genetic risk underlying Parkinson's disease.

Simón-Sánchez J, Schulte C, Bras JM, (and 44 more) ; *Nat Genet* (2009); PMID: 19915575

[View abstract]

same principles

data integration

text mining

common identifiers

calibrated scores

# exercises

jensenlab.org/training/string/