The STRING database

Lars Juhl Jensen
jensenlab.org

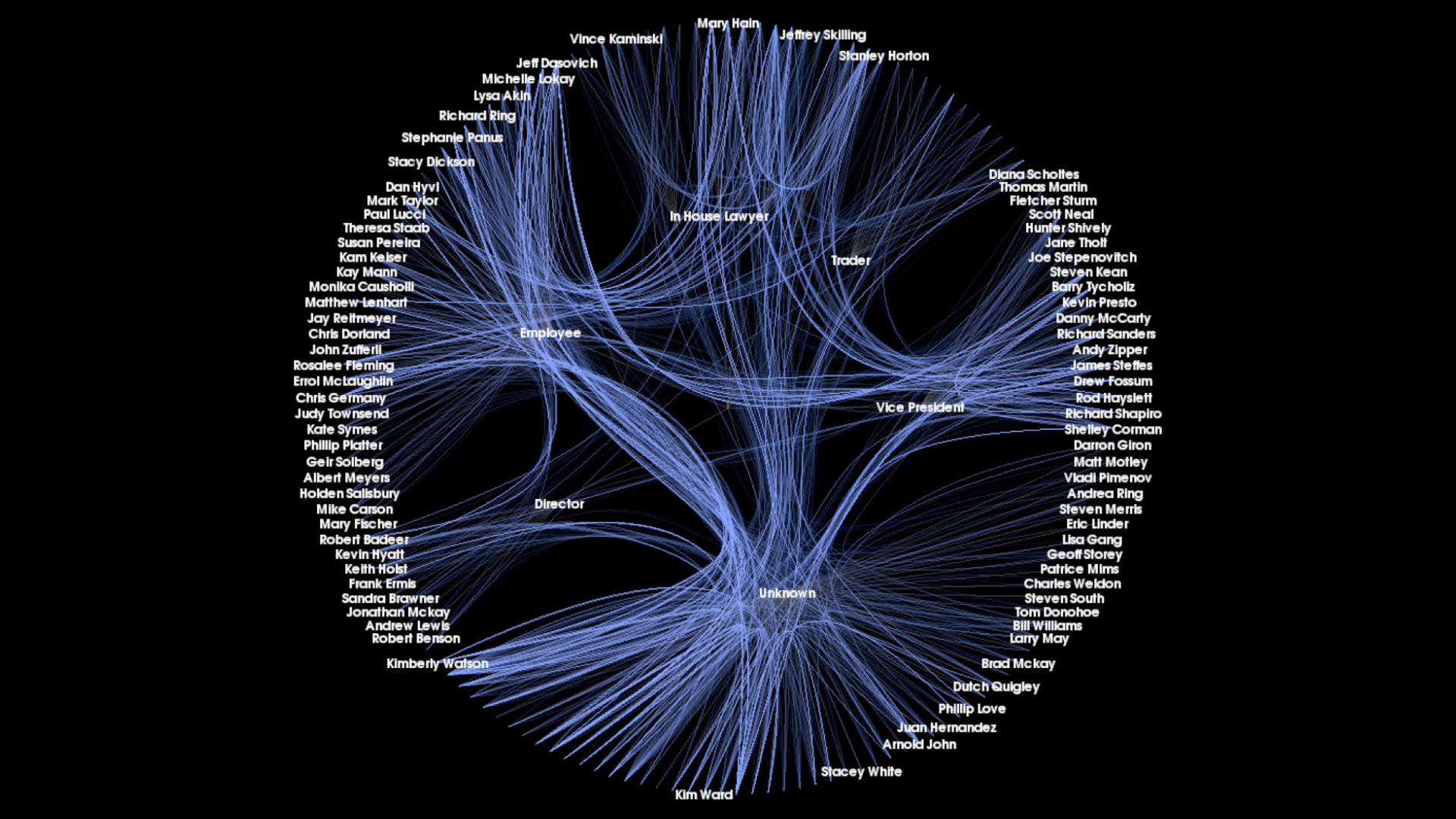
The Novo Nordisk Foundation
Center for Protein Research

Faculty of Health Sciences
University of Copenhagen

interaction networks

association networks

guilt by association


Mary Hain
Vince Kaminski
Jeffrey Skilling
Stanley Horton
Jeff Dasovich
Michelle Lokay
Lysa Akin
Richard Ring
Stephanie Panus
Stacy Dickson
Dan Hyvl
Mark Taylor
Paul Lucci
Theresa Staab
Susan Pereira
Kam Keiser
Kay Mann
Monika Causholli
Matthew Lenhart
Jay Reitmeyer
Chris Dorland
John Zufferli
Rosalee Fleming
Errol McLaughlin
Chris Germany
Judy Townsend
Kate Symes
Phillip Platter
Geir Solberg
Albert Meyers
Holden Salisbury
Mike Carson
Mary Fischer
Robert Badeer
Kevin Hyatt
Keith Holst
Frank Ermis
Sandra Brawner
Jonathan Mckay
Andrew Lewis
Robert Benson
Kimberly Watson
In House Lawyer
Trader
Employee
Vice President
Director
Unknown
Kim Ward
Stacey White
Arnold John
Juan Hernandez
Phillip Love
Dutch Quigley
Brad Mckay
Diana Scholtes
Thomas Martin
Fletcher Sturm
Scott Neal
Hunter Shively
Jane Tholt
Joe Stepenovitch
Steven Kean
Barry Tycholiz
Kevin Presto
Danny McCarty
Richard Sanders
Andy Zipper
James Steffes
Drew Fossum
Rod Hayslett
Richard Shapiro
Shelley Corman
Darron Giron
Matt Motley
Vladi Pimenov
Andrea Ring
Steven Merris
Eric Linder
Lisa Gang
Geoff Storey
Patrice Mims
Charles Weldon
Steven South
Tom Donohoe
Bill Williams
Larry May

protein networks

STRING

2031 genomes

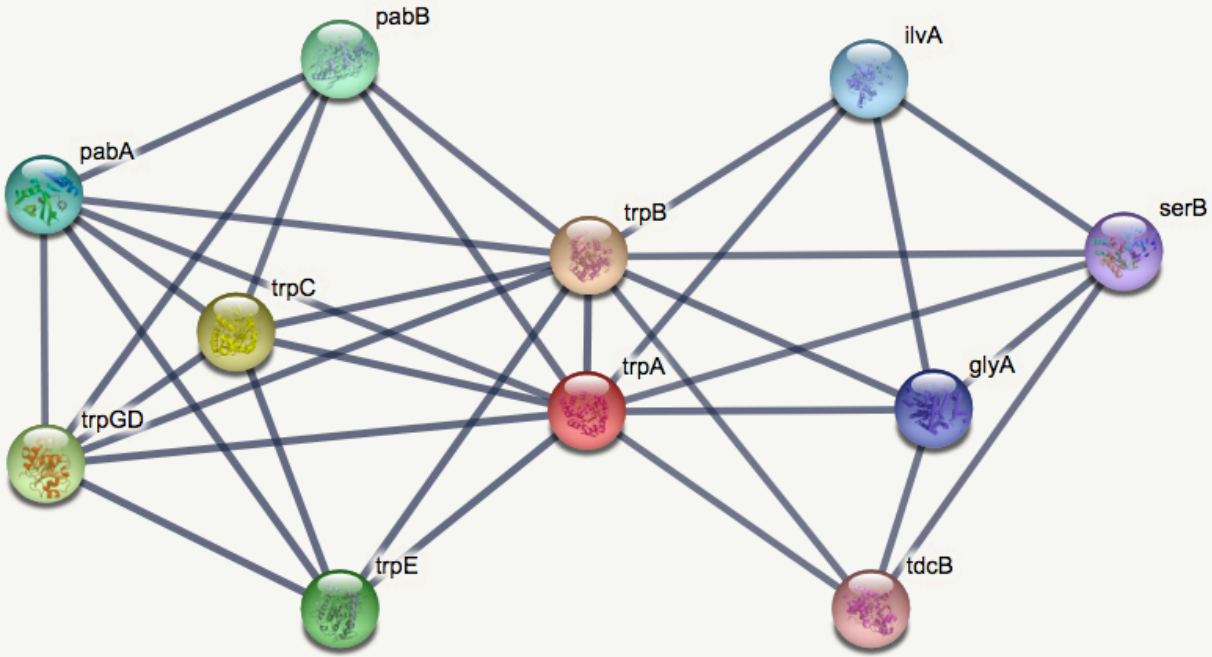
9.6 million proteins

functional associations

Version: 10.5 LOGIN | REGISTER

STRING

Search Download Help My Data

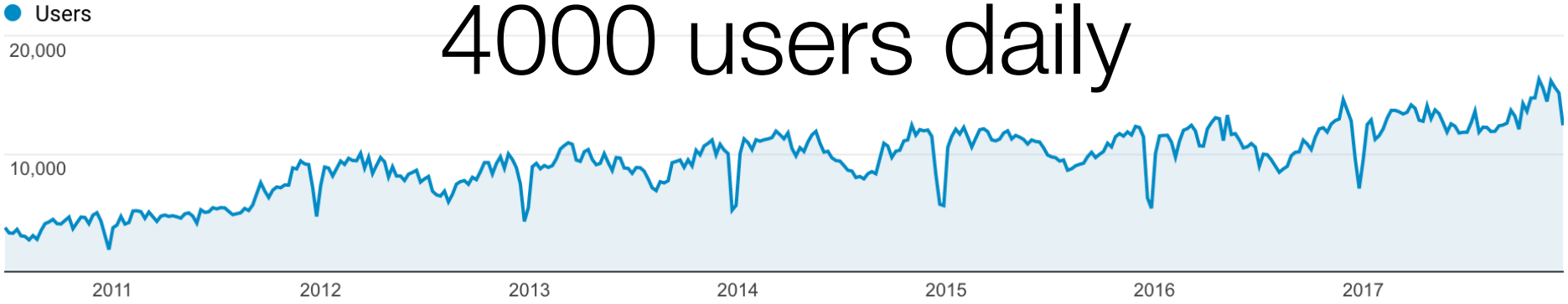


string-db.org

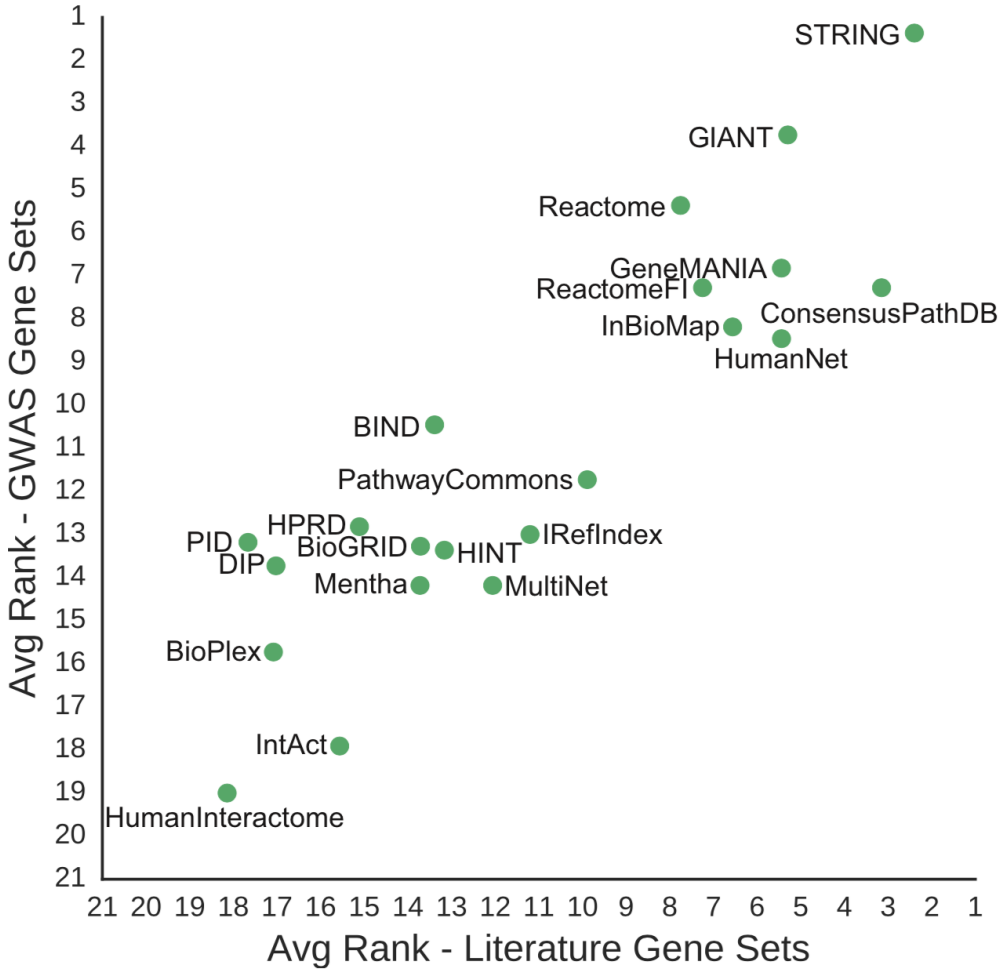
Szklarczyk et al., *Nucleic Acids Research*, 2017

heavily used

4000 users daily



works well

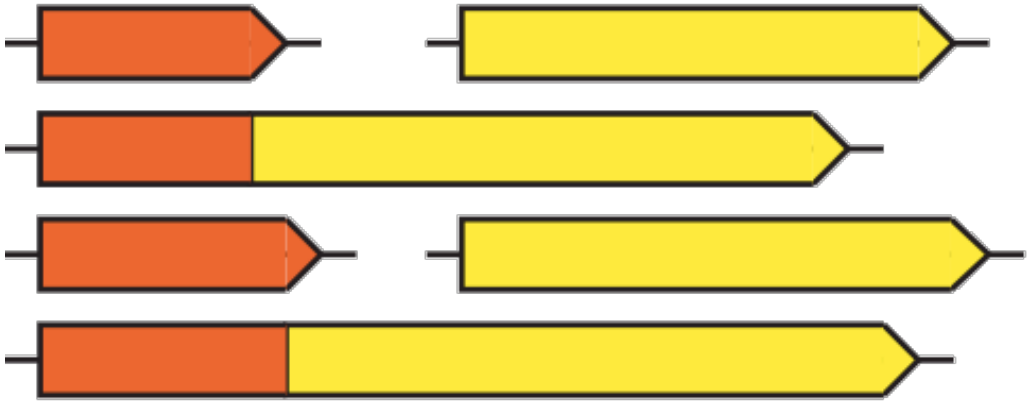


Huang et al., *Cell Syst.*, 2018

data integration

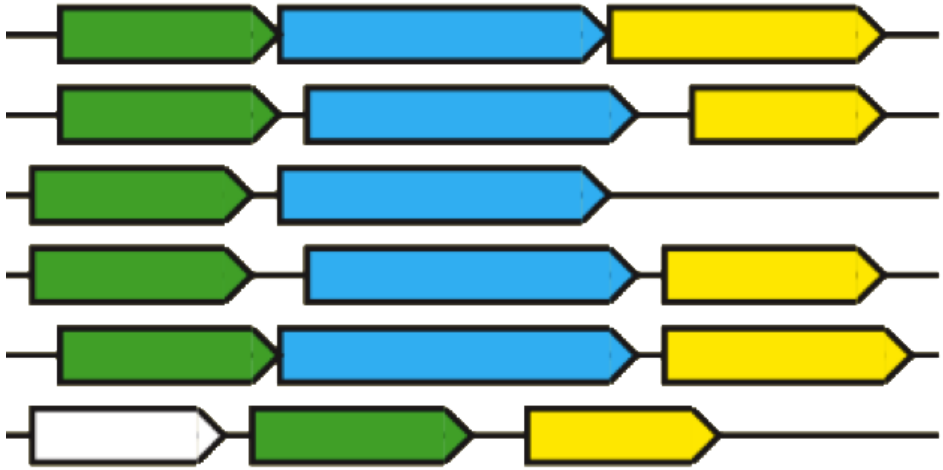
genomic context

gene fusion



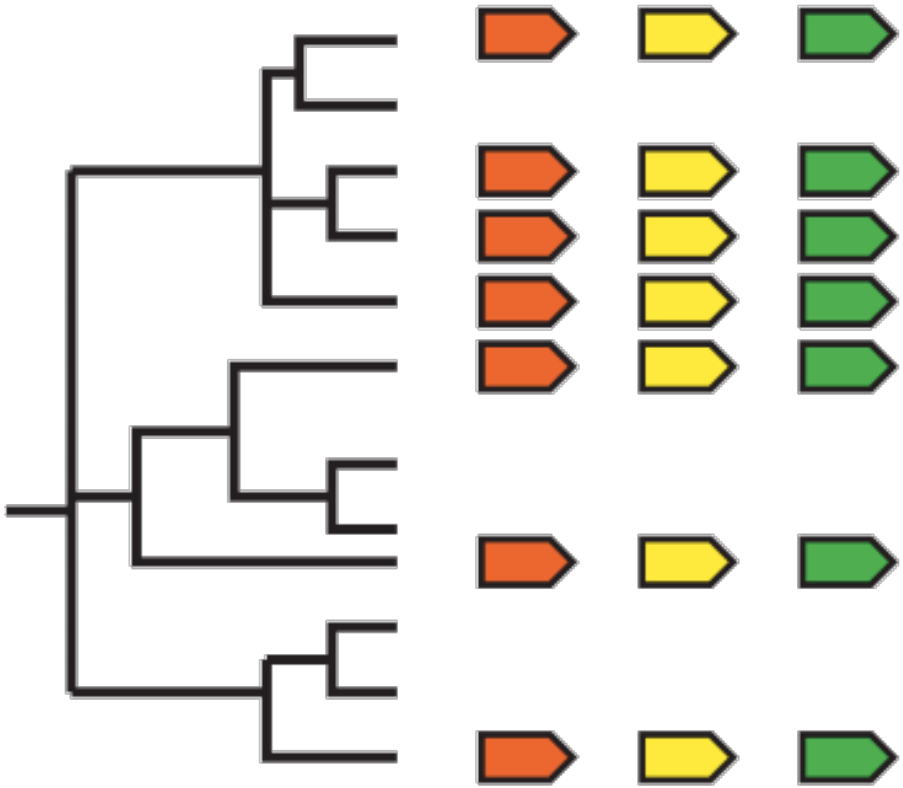
Korbel et al., *Nature Biotechnology*, 2004

gene neighborhood



Korbel et al., *Nature Biotechnology*, 2004

phylogenetic profiles

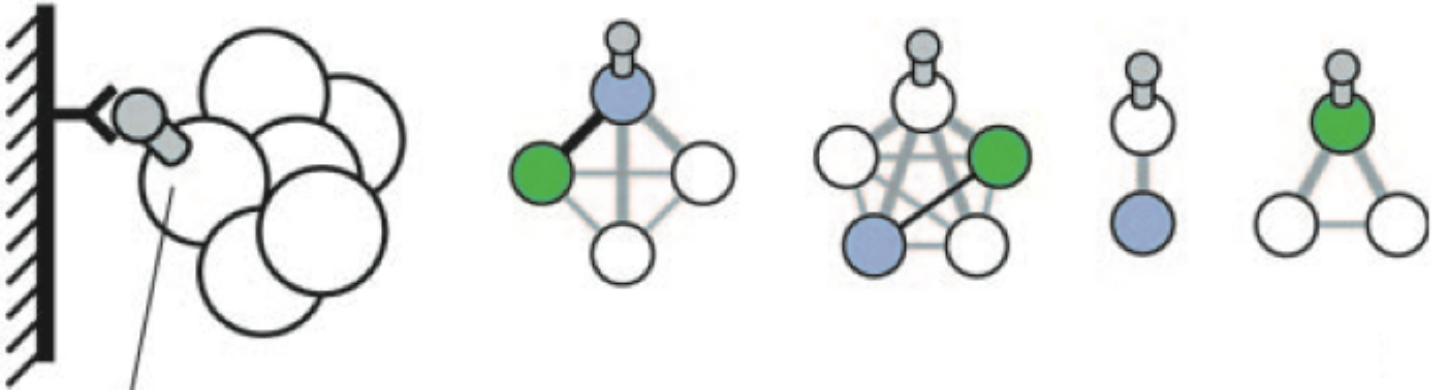


Korbel et al., *Nature Biotechnology*, 2004

experimental data

gene coexpression

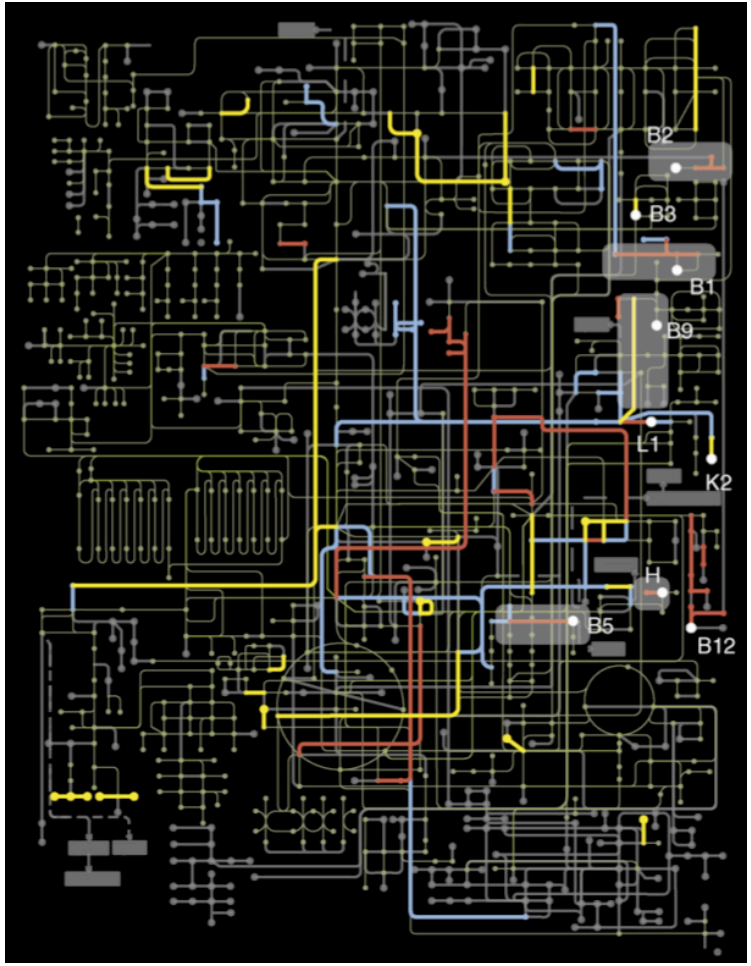
physical interactions



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

curated knowledge

pathways



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

many databases

different formats

different identifiers

varying quality

not comparable

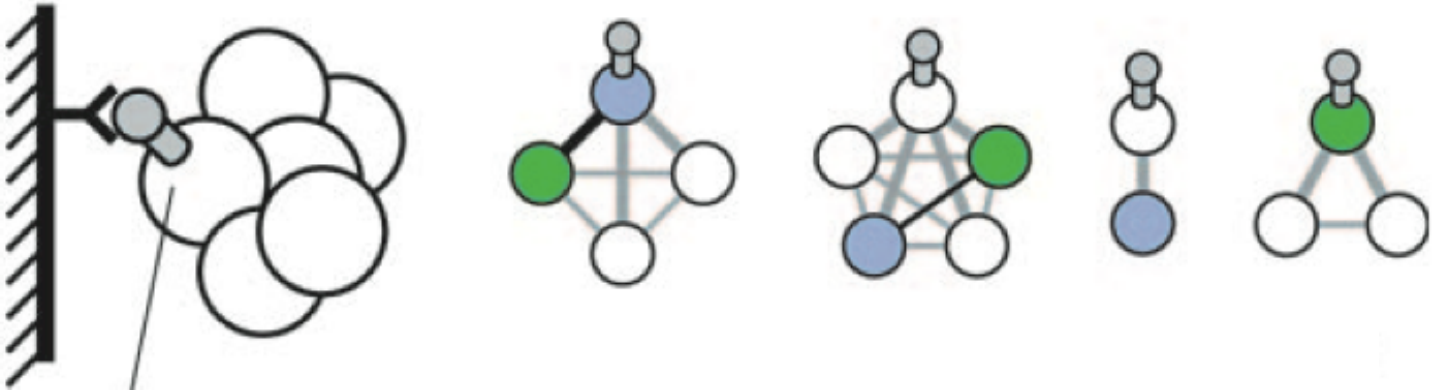
not same species

hard work

parsers

mapping files

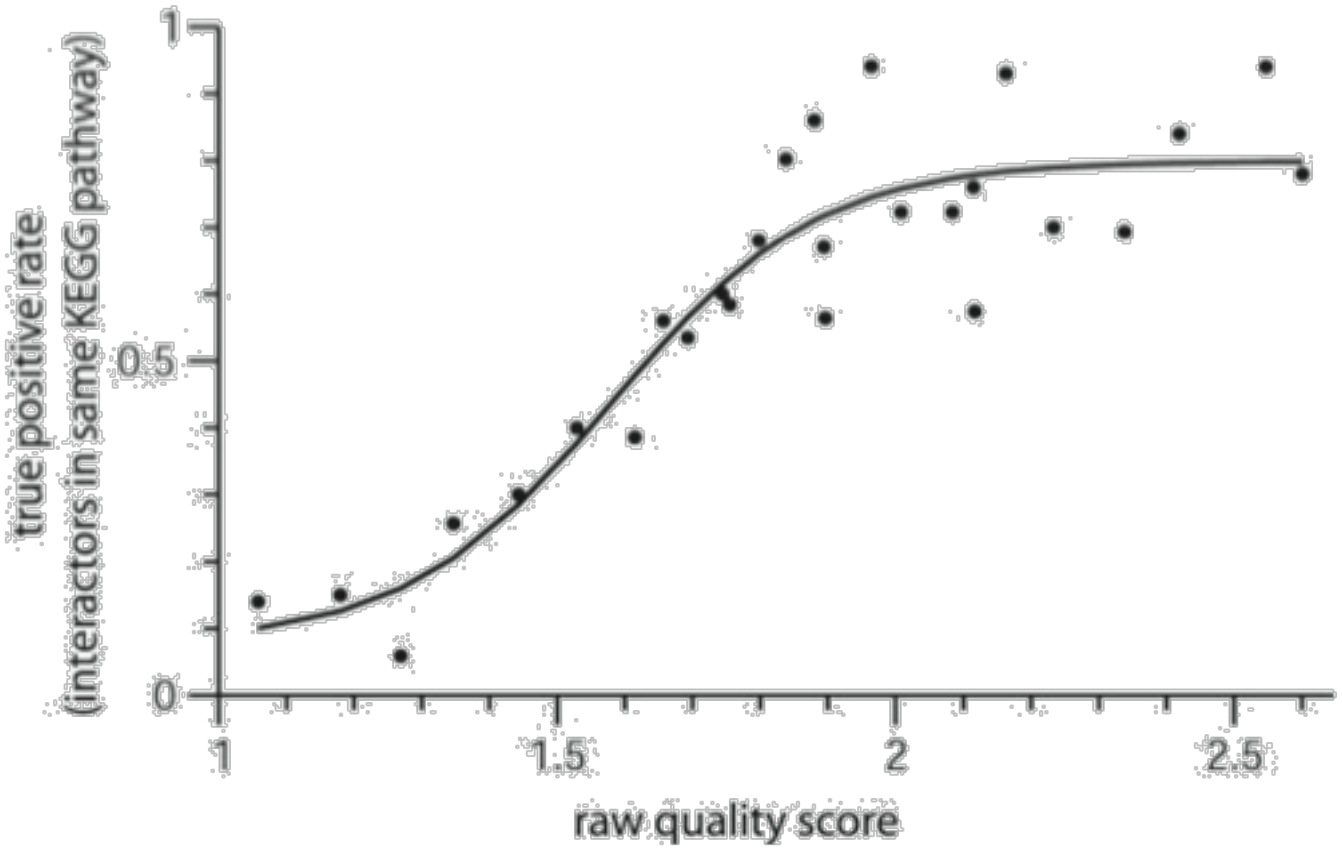
quality scores



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

score calibration

gold standard



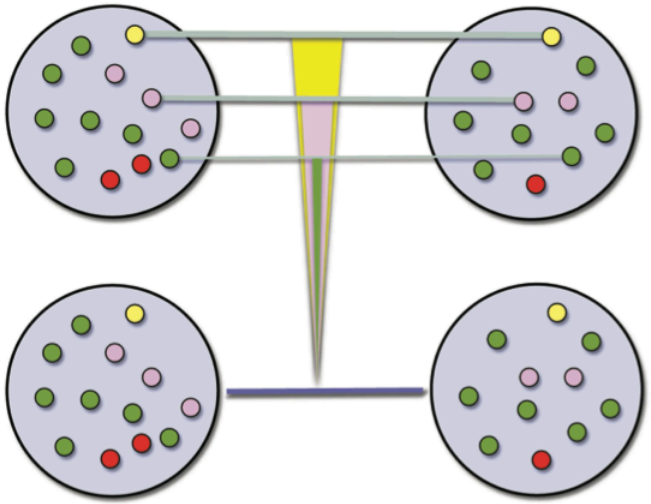
von Mering et al., *Nucleic Acids Research*, 2005

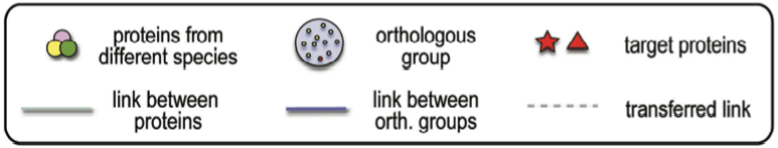
common identifiers

common scale

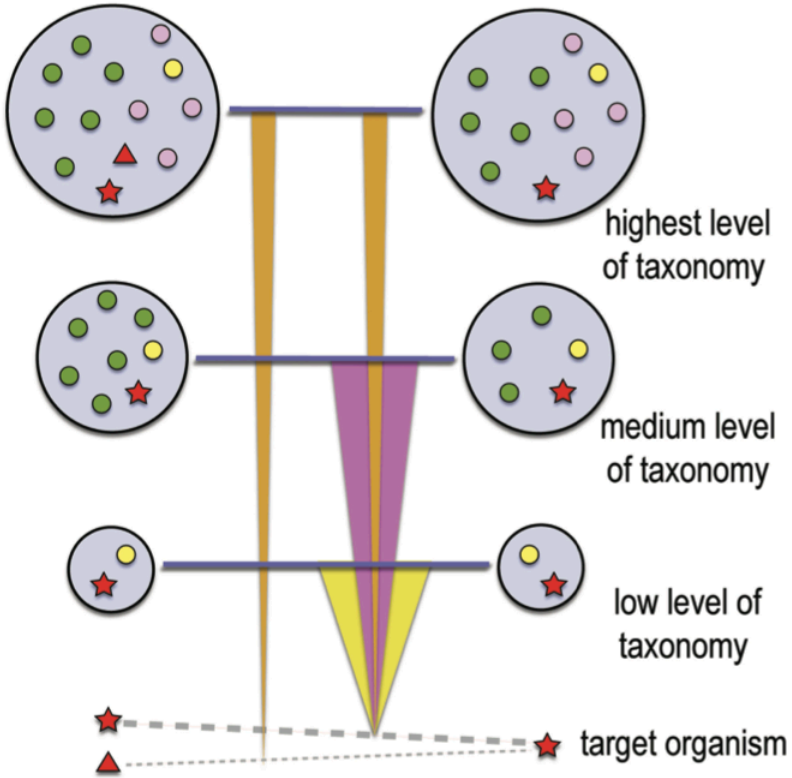
orthology transfer

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





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



Franceschini et al., *Nucleic Acids Research*, 2013

cross-species integration

missing most of the data

text mining

>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

synonyms

cyclin dependent kinase 1

CDC2

orthographic variation

cyclin dependent kinase 1

cyclin-dependent kinase 1

CDC2

hCDC2

black list

SDS

co-mentioning

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

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

score calibration

orthology transfer

STRING v11

5090 genomes

24.6 million proteins

more data

new enrichment modes

genome-wide screens

list of significant genes

arbitrary cutoffs

annotated pathways

arbitrary boxes

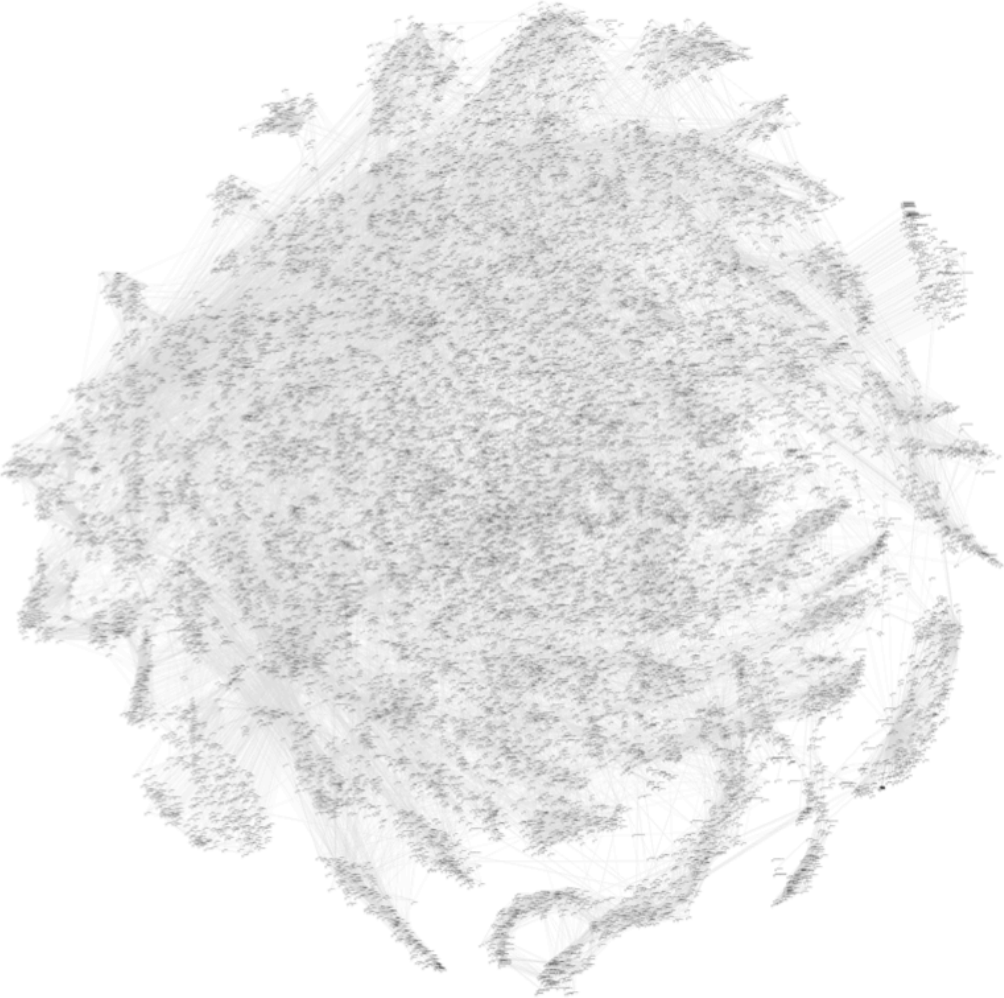
upload complete dataset

non-parametric tests

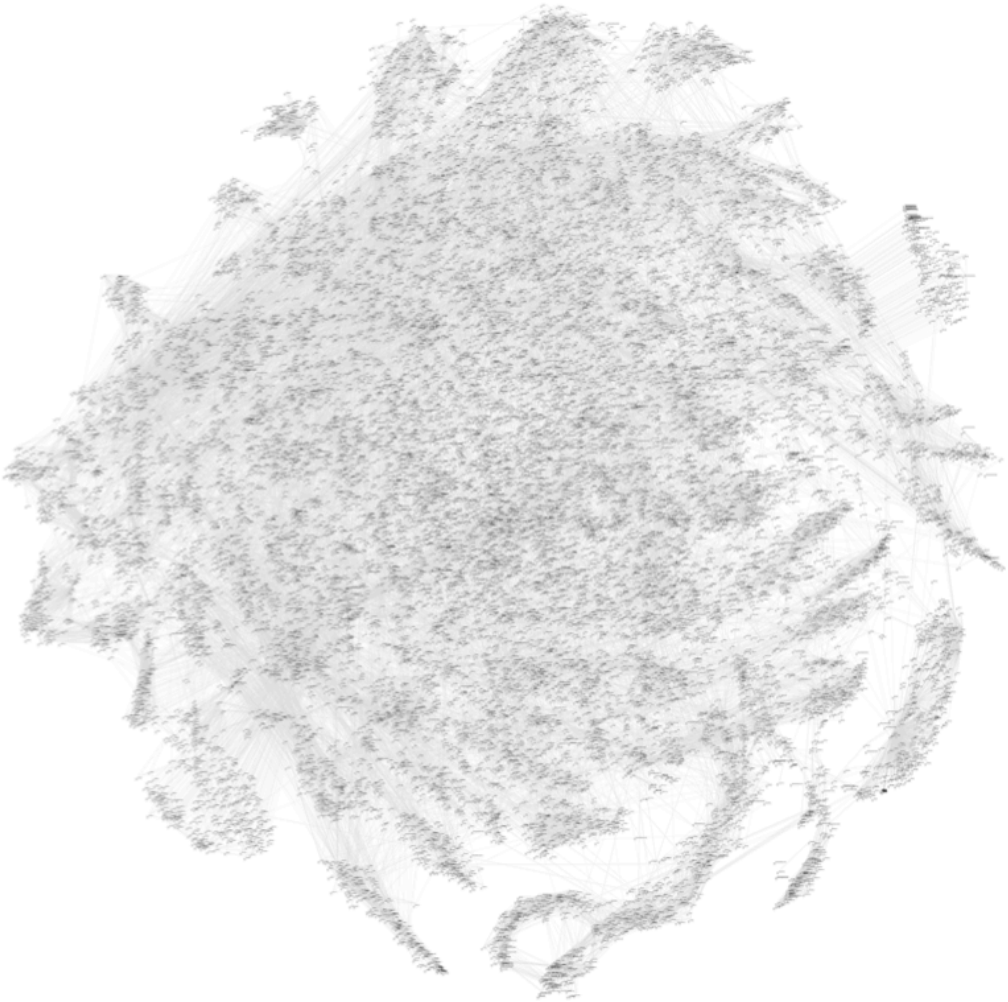
avoid arbitrary cutoff

hierarchical clustering

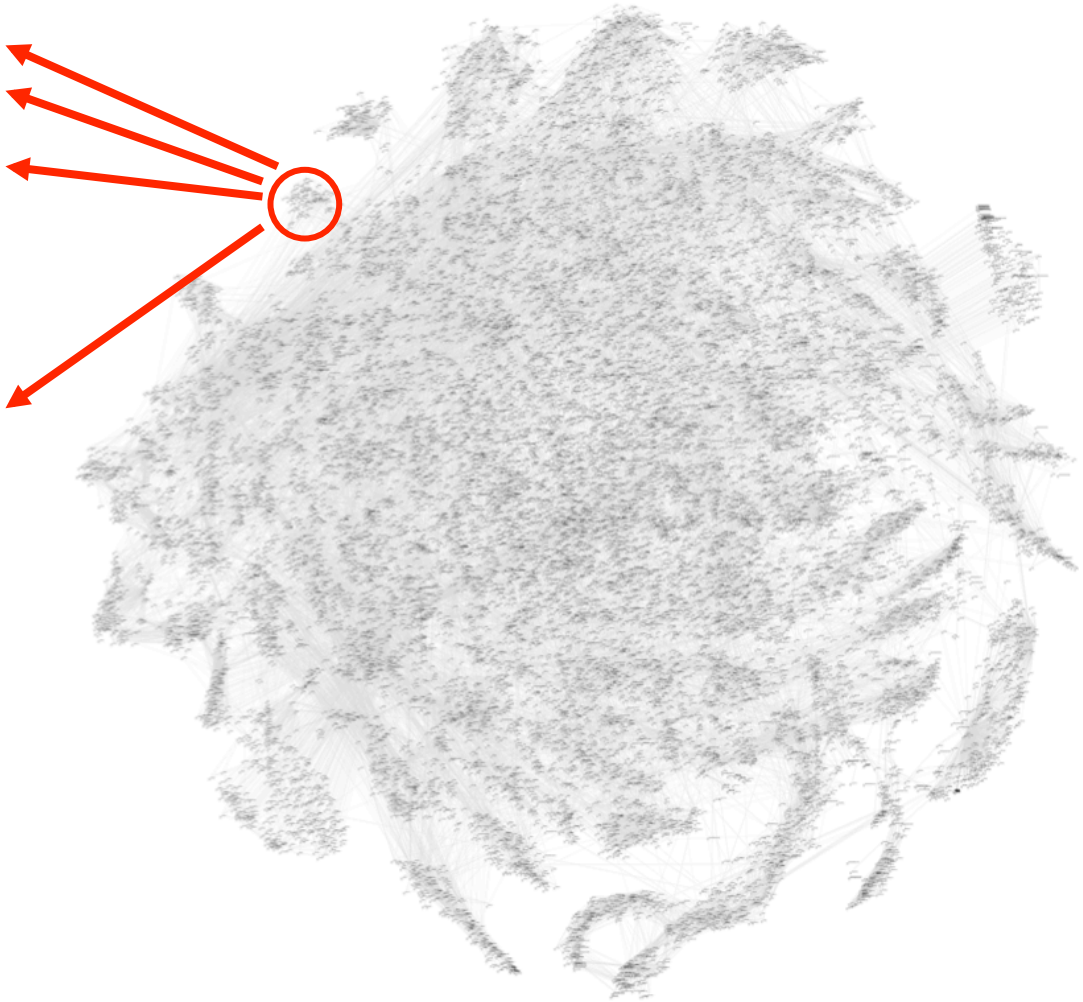
enriched network modules



1: PKP1	8.326
2: CDSN	8.130
3: SERPINB5	8.065
4: DSC1	7.917
5: DSG1	7.838
6: CALML5	7.706
7: ZNF750	7.527
8: SERPINB7	7.497
9: LCE2B	7.467
10: CHP2	7.423
11: GJB6	7.301
12: COL17A1	7.263
13: C19orf33	7.195
14: SBSN	7.140
15: LY6D	7.056
16: TRIM29	7.034
17: FLG	7.031
18: CRCT1	7.022
19: KRT15	6.867
20: SPRR1A	6.859
21: LOR	6.848
22: CLCA2	6.767
23: SLURP1	6.767
24: C1orf68	6.695
25: LGALS7	6.613



1: PKP1	8.326
2: CDSN	8.130
3: SERPINB5	8.065
4: DSC1	7.917
5: DSG1	7.838
6: CALML5	7.706
7: ZNF750	7.527
8: SERPINB7	7.497
9: LCE2B	7.467
10: CHP2	7.423
11: GJB6	7.301
12: COL17A1	7.263
13: C19orf33	7.195
14: SBSN	7.140
15: LY6D	7.056
16: TRIM29	7.034
17: FLG	7.031
18: CRCT1	7.022
19: KRT15	6.867
20: SPRR1A	6.859
21: LOR	6.848
22: CLCA2	6.767
23: SLURP1	6.767
24: C1orf68	6.695
25: LGALS7	6.613



enriched publications

named entity recognition

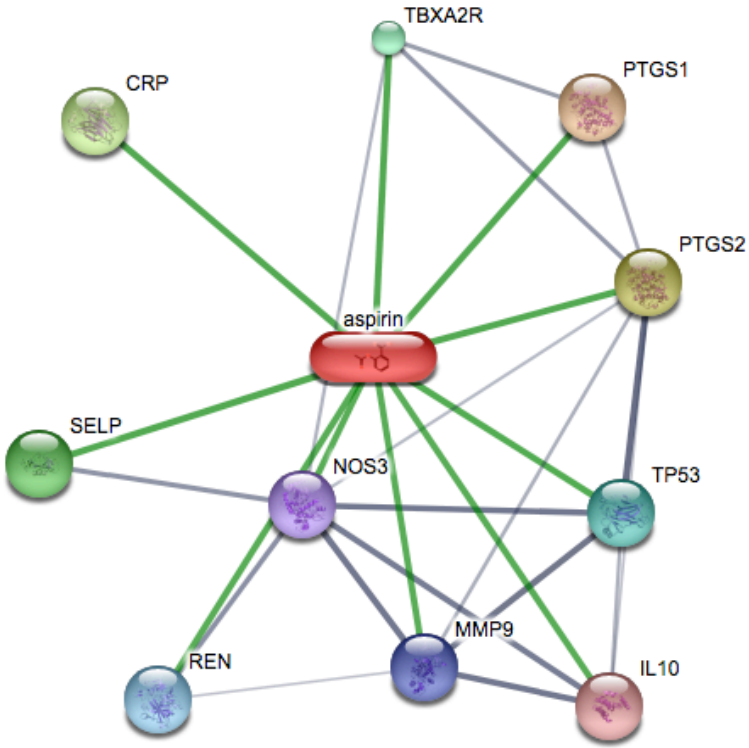
related resources

chemical compounds

Version: 5.0 LOGIN | REGISTER

STITCH

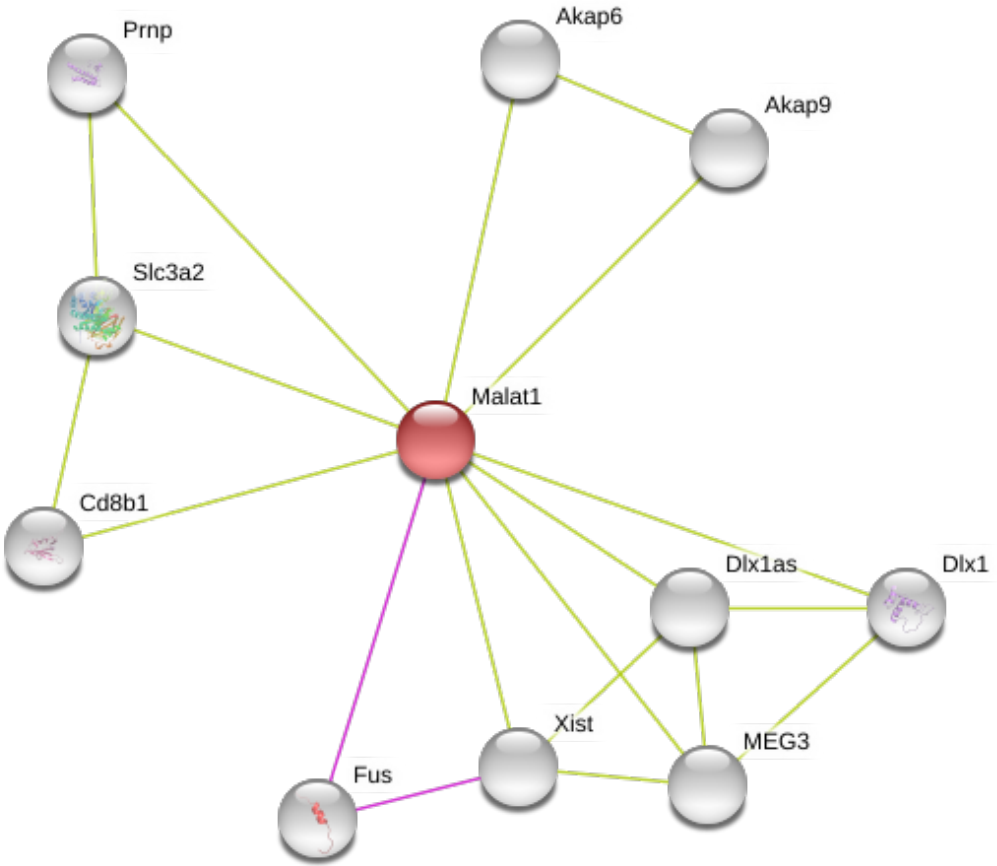
Search Download Help My Data



stitch-db.org

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

ncRNAs



rth.dk/resources/rain/

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

viruses

STRING Viruses Beta

Search Download Help

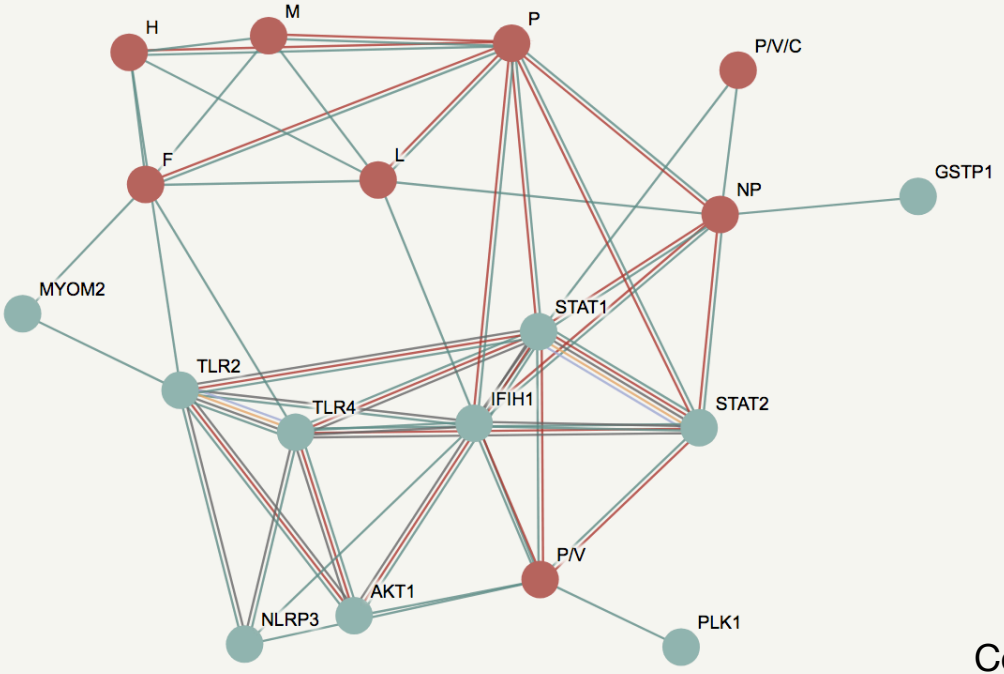
Measles virus (strain Ichinose-B95a)

Measles virus (strain Ichinose-B95a)

NCBI taxonomy Id: 11234

Homo sapiens

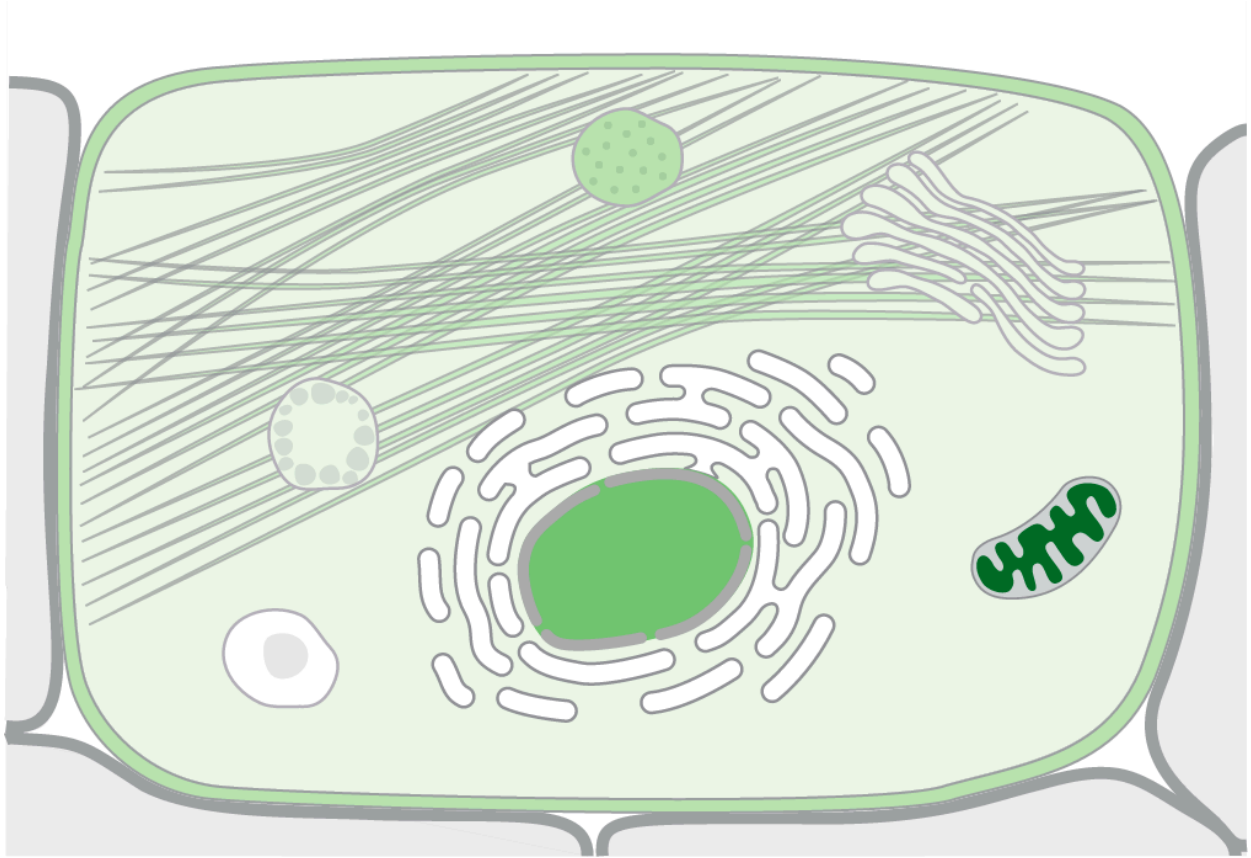
NCBI taxonomy Id: 9606



viruses.string-db.org

Cook et al., *Viruses*, 2018

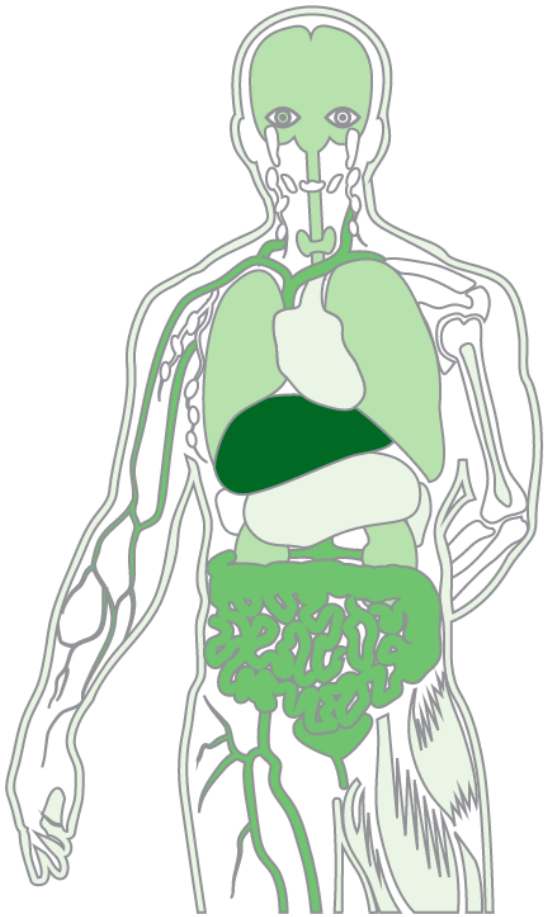
subcellular localization

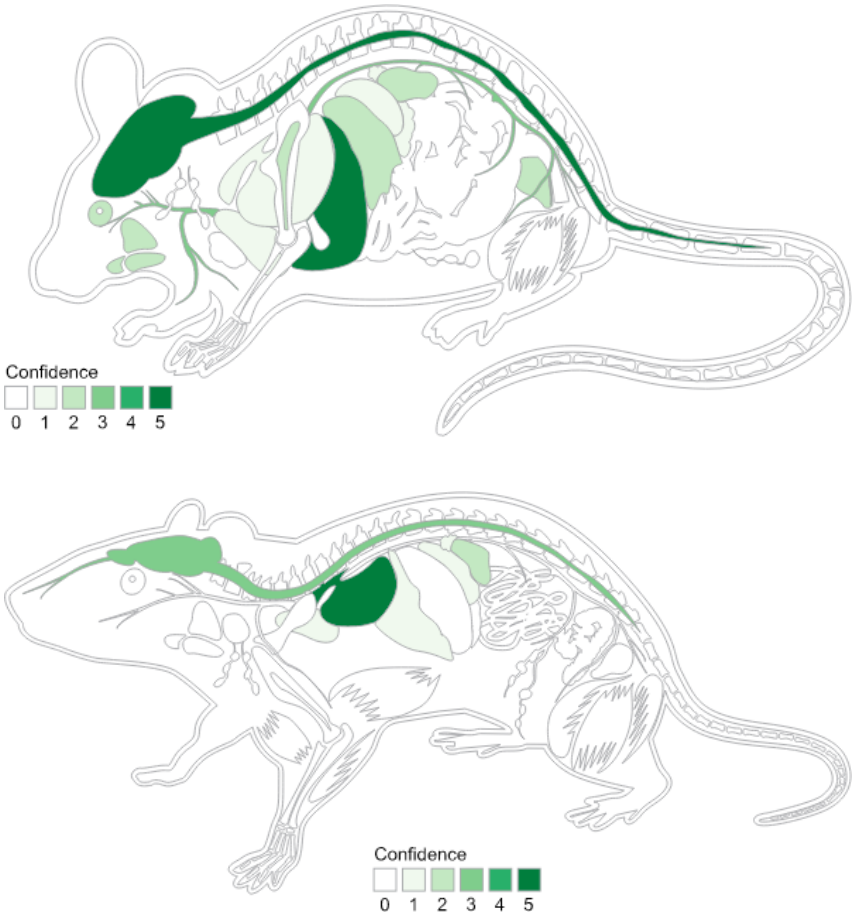


compartments.jensenlab.org

Binder et al., *Database*, 2014

tissue expression





tissues.jensenlab.org

Palasca et al., *Database*, 2018

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, Andrabi 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]

diseases.jensenlab.org

Pletscher-Frankild et al., *Methods*, 2015

same principles

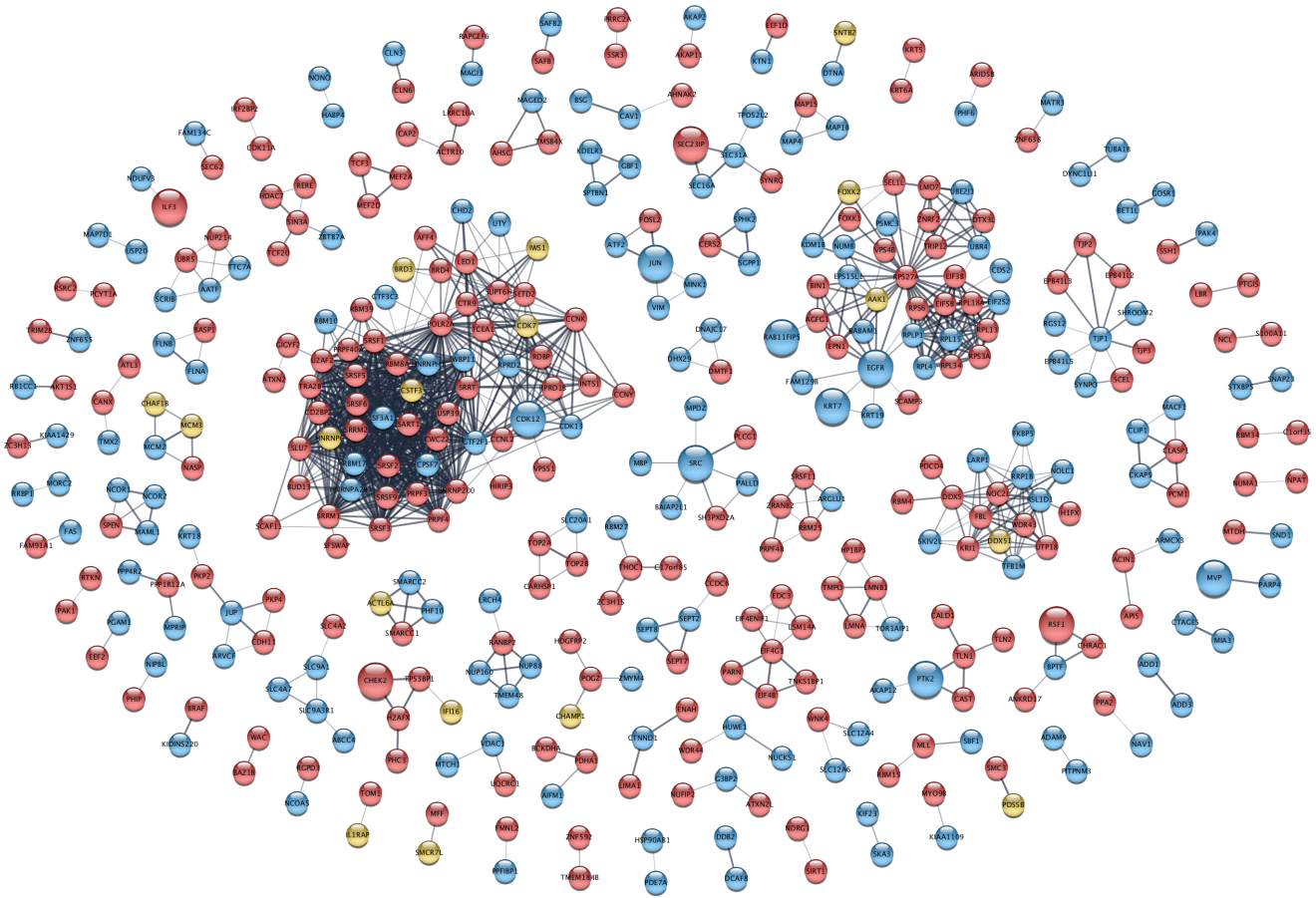
data integration

text mining

common identifiers

calibrated scores

Cytoscape stringApp



Doncheva et al., *BioRxiv*, 2018

Acknowledgments

STRING

Damian Szklarczyk
Annika Gable
David Lyon
Alexander Junge
Stefan Wyder
Jaime Huerta-Cepas
Milan Simonovic
Peer Bork
Christian von Mering

Related resources

Nadezhda Doncheva
Helen Cook
Alberto Santos
Sune Pletscher-Frankild
Oana Palasca
Janos Binder
Christian Stolte
Jan Gorodkin
John “Scooter” Morris

SIB

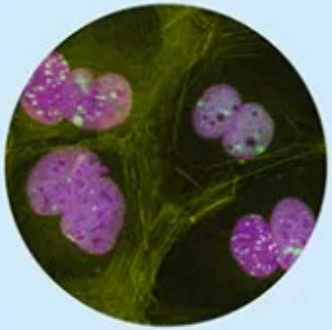
novo nordisk fonden

NIH
National Institutes of Health

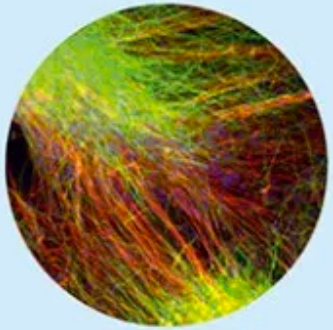
EMBL

Copenhagen Bioscience PhD

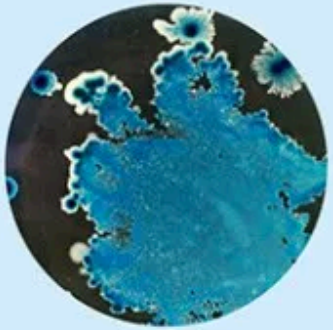
APPLICATION DEADLINE JANUARY 6, 2019:
www.novonordiskfonden.dk/da/cphbiosciencephd



Novo Nordisk Foundation
Center for Protein Research



Novo Nordisk Foundation
Danish Stem Cell Center



Novo Nordisk Foundation
Center for Biosustainability

Novo Nordisk Foundation
Center for Basic Metabolic Research

exercises

jensenlab.org/training/string/