

STRING

Protein Association Network Database

Paul Henry Flynn

Sections

Description

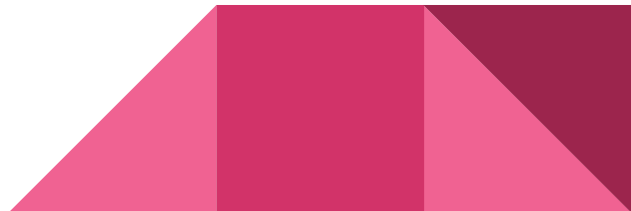
Features

Searching

Results

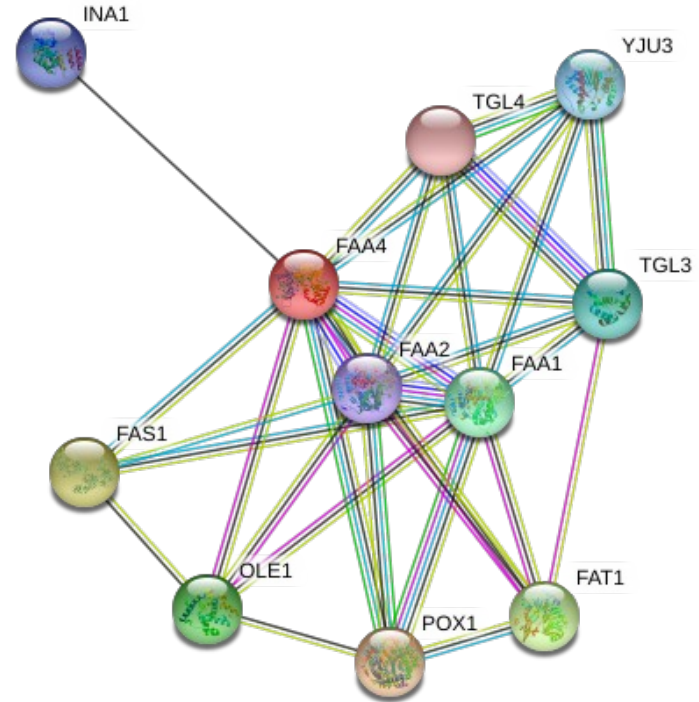
Algorithms

Critiques



STRING

- Visual network view of protein associations
 - Models known and predicted associations for a particular group
- Interactions include direct, indirect
- Developers
 - Swiss Institute of Bioinformatics, Center for Protein Research, University of Zurich



Features

Welcome to STRING

Protein-Protein Interaction Networks

ORGANISMS

5090

PROTEINS

24.6 mio

INTERACTIONS

>2000 mio

SEARCH

© STRING CONSORTIUM 2019



SIB - Swiss Institute of Bioinformatics



CPR - NNF Center for Protein Research



EMBL - European Molecular Biology Laboratory

ABOUT

[Content](#)

[References](#)

[Contributors](#)

[Statistics](#)

INFO

[Scores](#)

[Use scenarios](#)

[FAQs](#)

[Cookies/Privacy](#)

ACCESS

[Versions](#)

[APIs](#)

[Licensing](#)

[Usage](#)

CREDITS

[Funding](#)

[Datasources](#)

[Partners](#)

[Software](#)

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks ^{New}](#) >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Single Protein by Name / Identifier

Protein Name: (examples: [#1](#) [#2](#) [#3](#))

Organism:

 ▼**SEARCH**© **STRING CONSORTIUM 2019**

SIB - Swiss Institute of Bioinformatics



CPR - NNF Center for Protein Research



EMBL - European Molecular Biology Laboratory

ABOUT

[Content](#)[References](#)[Contributors](#)[Statistics](#)

INFO

[Scores](#)[Use scenarios](#)[FAQs](#)[Cookies/Privacy](#)

ACCESS

[Versions](#)[APIs](#)[Licensing](#)[Usage](#)

CREDITS

[Funding](#)[Datasources](#)[Partners](#)[Software](#)

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks](#) **New** >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Single Protein by Sequence

Amino Acid Sequence: (examples: [#1](#) [#2](#) [#3](#))

Organism:

SEARCH

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks](#) **New** >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Multiple Proteins by Names / Identifiers

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

... or, upload a file:

[Browse ...](#)

Organism:

[SEARCH](#)

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks](#) **New** >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Multiple Proteins by Sequences

Amino Acid Sequences: (examples: [#1](#) [#2](#))

... or, upload a file:

[Browse ...](#)

Organism:

[SEARCH](#)

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks ^{New}](#) >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Proteins with Values/Ranks - Functional Enrichment Analysis

Submit your entire experiment as a list of proteins - no cutoffs.

Behind each protein, put a meaningful value for ranking (fold-change, log-pvalue, abundance, ...).

Proteins with Values: (one per line; examples: #1 #2 #3)

... or, upload a file:

Browse ...

Organism:

SEARCH

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks](#) **New** >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Organisms and Clades

Organism:

[SHOW INFO](#)© **STRING CONSORTIUM 2019**

SIB - Swiss Institute of Bioinformatics



CPR - NNF Center for Protein Research



EMBL - European Molecular Biology Laboratory

ABOUT

[Content](#)[References](#)[Contributors](#)[Statistics](#)

INFO

[Scores](#)[Use scenarios](#)[FAQs](#)[Cookies/Privacy](#)

ACCESS

[Versions](#)[APIs](#)[Licensing](#)[Usage](#)

CREDITS

[Funding](#)[Datasources](#)[Partners](#)[Software](#)

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks New](#) >[Organisms](#) >[Protein families \("COGs"\)](#) ▾[... by family name](#)[... by protein name](#)[... by protein sequence](#)[... multiple protein names](#)[... multiple sequences](#)[Examples](#) >[Random entry](#) >

SEARCH

Protein Family by Name

Protein Family Name: (examples: [#1](#) [#2](#) [#3](#))[SEARCH](#)

[Protein by name](#) >[Protein by sequence](#) >[Multiple proteins](#) >[Multiple sequences](#) >[Proteins with Values/Ranks](#) **New** >[Organisms](#) >[Protein families \("COGs"\)](#) >[Examples](#) >[Random entry](#) >

SEARCH

Examples and Highlights

Example 1

FAA4 and its ten most confident interactors.

FAA4 in yeast is a long chain fatty acyl-CoA synthetase; see it connected to other synthetases as well as regulators.

[SHOW](#)**Example 2**

The 20 most frequently mutated human cancer genes.

These have a higher-than-expected degree of interconnectivity, and are enriched in characteristic molecular functions.

[SHOW](#)**Example 3**

A poorly characterized protein from Methanosaeta.

It has no annotated domains or functions, but shows some intriguing connections to possible DNA-replication or -repair proteins.

[SHOW](#)

© **STRING CONSORTIUM 2019**



SIB - Swiss Institute of Bioinformatics



CPR - NNF Center for Protein Research



EMBL - European Molecular Biology Laboratory

ABOUT

[Content](#)[References](#)[Contributors](#)[Statistics](#)

INFO

[Scores](#)[Use scenarios](#)[FAQs](#)[Cookies/Privacy](#)

ACCESS

[Versions](#)[APIs](#)[Licensing](#)[Usage](#)

CREDITS

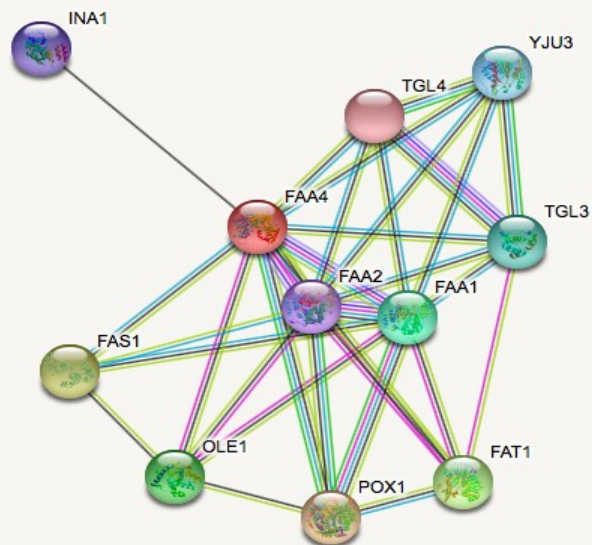
[Funding](#)[Datasources](#)[Partners](#)[Software](#)

Search Results

There are several matches for 'FAA4'.
Please select one from the list below and press Continue to proceed.

[<- BACK](#)
[CONTINUE ->](#)

organism	protein
<input checked="" type="checkbox"/> Saccharomyces cerevisiae	FAA4 - Long-chain-fatty-acid-CoA ligase 4; Long chain fatty acyl-CoA synthetase; activates fatty acids with a preference for C12:0-C16:0 chain lengths; role in the competitive import of long-chain fatty acids and sphingoid long-chain bases; role in stationary phase survival; localizes to lipid particles and the plasma membrane; role in sphingolipid-to-glycerolipid metabolism; forms cytoplasmic foci upon replication stress; faa1 faa4 double null complemented by any of human ACSBG1, ACSL1, 3, 4, 5, 6, SLC27A2, or 4; Belongs to the ATP-dependent AMP-binding enzyme family
<input type="checkbox"/> Homo sapiens	MYH11 - Myosin-11; Muscle contraction; Belongs to the TRAFAC class myosin-kinesin ATPase superfamily. Myosin family [a.k.a. KIAA0866, NP_074035.1, ENSP00000484924, FAA4]
<input type="checkbox"/> Candida albicans	FAA4 - annotation not available
<input type="checkbox"/> Scheffersomyces stipitis	FAA4 - acyl-CoA synthetase (long-chain fatty acid CoA ligase) (fatty acid activator 2), activates imported fatty acids and provides substrates for N-myristoylation
<input type="checkbox"/> Wickerhamomyces ciferrii	FAA4 - annotation not available
<input type="checkbox"/> Saccharomyces cerevisiae	FAA1 - Long-chain-fatty-acid-CoA ligase 1; Long chain fatty acyl-CoA synthetase; activates fatty acids with a preference for C12:0-C16:0 chain lengths; role in the competitive import of long-chain fatty acids and sphingoid long-chain bases; accounts for most acyl-CoA synthetase activity; localizes to lipid particles and the plasma membrane; role in sphingolipid-to-glycerolipid metabolism; forms ER foci upon replication stress; faa1 faa4 double null complemented by any of human ACSBG1, ACSL1, 3, 4, 5, 6, SLC27A2, or 4 [a.k.a. YOR317W, O6136, Fatty acid activator 1]
<input type="checkbox"/> Schizosaccharomyces pombe	Icf1 - Long-chain-fatty-acid-CoA ligase 1; Esterification, concomitant with transport, of exogenous long-chain fatty acids into metabolically active CoA thioesters for subsequent degradation or incorporation into phospholipids. It may supplement intracellular myristoyl-CoA pools from exogenous myristate. Preferentially acts on C12:0-C16:0 fatty acids with myristic and pentadecanic acid (C15:0) having the highest activities (By similarity). Appears to play a role in the maintenance of cell viability during stationary phase [a.k.a. SPBC18H10.02, NM_001021624, ACSL3, FAA4]
<input type="checkbox"/> Xenopus tropicalis	myh11 - Myosin, heavy chain 11, smooth muscle; Belongs to the TRAFAC class myosin-kinesin ATPase superfamily. Myosin family [a.k.a. AAMC01009813, Myosin, heavy chain 11, smooth muscle, ENSXETG00000019650, faa4]
<input type="checkbox"/> Aspergillus flavus	CADAFLAP00005791 - Fatty acid activator Faa4 , putative [a.k.a. AFLA_044640, UPI00006784D1, XP_002377926]
<input type="checkbox"/> Aspergillus fumigatus	CADAFUBP00002517 - Fatty acid activator Faa4 , putative; AMP-binding enzyme (Fum16), putative [a.k.a. AFUA_2G09910, 6.2.1.3, CADAFUBT00002569]
<input type="checkbox"/> Aspergillus nidulans	AN6014.2 - Fatty acid activator Faa4 , putative (AFU_orthologue; AFUA_2G09910) [a.k.a. ANIA_06014, 6.2.1.3, Q5B0B6_EMENI]
<input type="checkbox"/> Blastomyces dermatitidis	BDCG_02451 - Long-chain acyl-CoA synthetase; Fatty acid activator Faa4 [a.k.a. XP_002620768.1, GeneID:8501372, XM_002620722.1]


[Viewers >](#)
[Legend >](#)
[Settings >](#)
[Analysis >](#)
[Exports >](#)
[Clusters >](#)
[+ More](#)
[- Less](#)

Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color



*colored nodes:
query proteins and first shell of interactors*



*white nodes:
second shell of interactors*

Node Content



*empty nodes:
proteins of unknown 3D structure*



*filled nodes:
some 3D structure is known or predicted*

Edges:


Edges represent protein-protein associations

Known Interactions


Predicted Interactions

Others


Result Options




Network currently showing
Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.




Experiments
Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.




Databases
Known metabolic pathways, protein complexes, signal transduction pathways, etc ... from curated databases.




Textmining
Automated, unsupervised textmining - searching for proteins that are frequently mentioned together.




Neighborhood
Groups of genes that are frequently observed in each other's genomic neighborhood.



Cooccurrence
Gene families whose occurrence patterns across genomes show similarities.



Coexpression
Proteins whose genes are observed to be correlated in expression, across a large number of experiments.



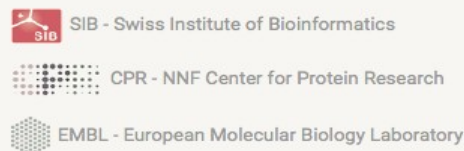
Fusion
Genes that are sometimes fused into single open reading frames.

STRING allows inspection of the interaction evidence for any given network. Choose any of the viewers above (disabled if not applicable in your network).

Server load: low (10%) [HD]

[GoPermalink](#)

© STRING CONSORTIUM 2019



ABOUT

[Content](#)
[References](#)
[Contributors](#)
[Statistics](#)

INFO

[Scores](#)
[Use scenarios](#)
[FAQs](#)
[Cookies/Privacy](#)

ACCESS

[Versions](#)
[APIs](#)
[Licensing](#)
[Usage](#)

CREDITS

[Funding](#)
[Datasources](#)
[Partners](#)
[Software](#)



STRING is part of the ELIXIR infrastructure: it is one of ELIXIR's Core Data Resources. [Learn more >](#)

Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color



*colored nodes:
query proteins and first shell of interactors*



*white nodes:
second shell of interactors*

Node Content



*empty nodes:
proteins of unknown 3D structure*



*filled nodes:
some 3D structure is known or predicted*

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.

Known Interactions



from curated databases



experimentally determined

Predicted Interactions



gene neighborhood



gene fusions



gene co-occurrence

Others



textmining



co-expression



protein homology

Your Input:

Long-chain-fatty-acid-CoA ligase 4; Long chain fatty acyl-CoA synthetase; activates fatty acids with a preference for C12-0-C16-0 chain lengths; role in the competitive import of long-chain fatty acids and sphingoid long-chain bases; role in stationary phase survival; localizes to lipid particles and the plasma membrane; role in sphingolipid-to-glycerolipid metabolism; forms cytoplasmic foci upon replication stress; *faa1 faa4* double null complemented by any of human ACSBG1, ACSL1, 3, 4, 5, 6, SLC27A2, or 4; Belongs to the ATP-dependent AMP-binding enzyme family (694 aa)

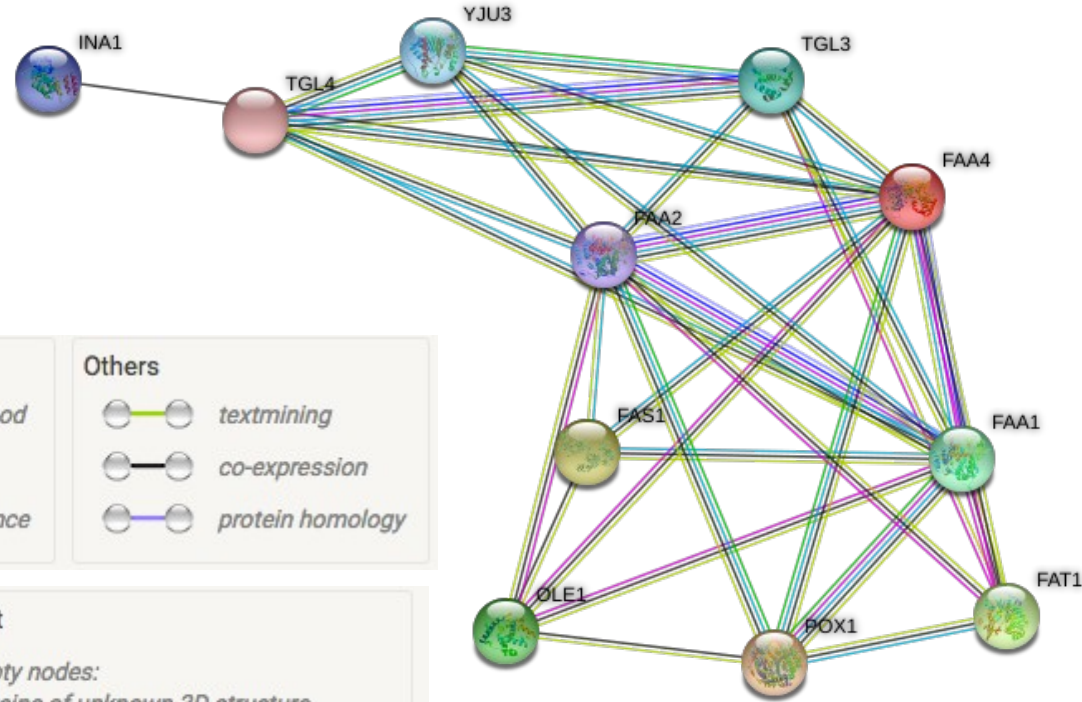
Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
POX1	Fatty-acyl coenzyme A oxidase; involved in the fatty acid beta-oxidation pathway; localized to the peroxisomal matrix (748 ...	●					●	●		0.980
FAS1	Beta subunit of fatty acid synthetase; complex catalyzes the synthesis of long-chain saturated fatty acids; contains acetyltr...							●	●	0.969
FAT1	Very long chain fatty acyl-CoA synthetase and fatty acid transporter; activates imported fatty acids with a preference for ve...						●	●	●	0.955
OLE1	Acyl-CoA desaturase 1; Delta(9) fatty acid desaturase; required for monounsaturated fatty acid synthesis and for normal di...						●	●	●	0.935
FAA1	Long-chain-fatty-acid-CoA ligase 1; Long chain fatty acyl-CoA synthetase; activates fatty acids with a preference for C12-0-...			●			●	●	●	0.923
TGL3	Lipase 3; Bifunctional triacylglycerol lipase and LPE acyltransferase; major lipid particle-localized triacylglycerol (TAG) lipas...						●	●	●	0.884
YJU3	Monoglyceride lipase (MGL); functional ortholog of mammalian MGL, localizes to lipid particles and membranes, also me...							●	●	0.859
INA1	Cell membrane protein YLR413W; Protein of unknown function; not an essential gene; YLR413W has a paralog, FAT3, that a...				●					0.834
FAA2	Long-chain-fatty-acid-CoA ligase 2; Medium chain fatty acyl-CoA synthetase; activates imported fatty acids; accepts a wid...			●		●	●	●	●	0.819
TGL4	Lipase 4; Multifunctional lipase/hydrolase/phospholipase; triacylglycerol lipase, steryl ester hydrolase, and Ca2+-independ...						●	●	●	0.819

Your Current Organism:

Saccharomyces cerevisiae

Network & Legend



Known Interactions

- from curated databases
- experimentally determined

Predicted Interactions

- gene neighborhood
- gene fusions
- gene co-occurrence

Others

- textmining
- co-expression
- protein homology

Node Color




- colored nodes:
query proteins and first shell of interactors
- white nodes:
second shell of interactors

Node Content

- empty nodes:
proteins of unknown 3D structure
- filled nodes:
some 3D structure is known or predicted

Basic Settings

meaning of network edges:

- ☒ evidence ( line color indicates the type of interaction evidence)
- ☐ confidence ( line thickness indicates the strength of data support)
- ☐ molecular action ( line shape indicates the predicted mode of action)

active interaction sources:

- ☒ Textmining ☒ Experiments ☒ Databases ☒ Co-expression
- ☒ Neighborhood ☒ Gene Fusion ☒ Co-occurrence

minimum required interaction score:

max number of interactors to show:

1st shell: 2nd shell:

UPDATE

Advanced Settings

network display mode:

- ☐ static png ( network is a simple bitmap image; not interactive)
- ☒ interactive svg ( network is a scalable vector graphic [SVG]; interactive)

display simplifications:

- ☐ disable structure previews inside network bubbles
- ☐ hide disconnected nodes in the network
- ☐ hide node labels

Network Stats

number of nodes: 11
 number of edges: 32
 average node degree: 5.82
 avg. local clustering coefficient: 0.843

expected number of edges: 10
 PPI enrichment p-value: 4.49e-08
*your network has significantly more interactions
 than expected (what does that mean?)*

Functional enrichments in your network

Note: some enrichments may be expected here (why?)

Biological Process (GO)

GO-term	description	count in gene set	false discovery rate
GO:0044255	cellular lipid metabolic process	10 of 280	2.85e-11
GO:0006631	fatty acid metabolic process	7 of 56	6.69e-11
GO:0001676	long-chain fatty acid metabolic process	4 of 7	4.38e-08
GO:0044539	long-chain fatty acid import into cell	3 of 3	1.49e-06
GO:0019433	triglyceride catabolic process	3 of 6	5.34e-06
(more ...)			

Molecular Function (GO)

GO-term	description	count in gene set	false discovery rate
GO:0015645	fatty acid ligase activity	4 of 5	3.70e-08
GO:0016405	CoA-ligase activity	4 of 8	7.24e-08
GO:0102391	decanoate-CoA ligase activity	3 of 4	1.72e-06
GO:0004467	long-chain fatty acid-CoA ligase activity	3 of 4	1.72e-06
GO:0031957	very long-chain fatty acid-CoA ligase activity	2 of 3	0.00022
(more ...)			

Cellular Component (GO)

GO-term	description	count in gene set	false discovery rate
GO:0005811	lipid droplet	7 of 55	9.84e-11
GO:0005777	peroxisome	3 of 83	0.0117
GO:0005886	plasma membrane	5 of 528	0.0181

Reference publications

publication	(year) title	count in gene set	false discovery rate
PMID:23275493	(2013) Lipid droplets and peroxisomes: key players in cellul...	10 of 148	3.93e-13
PMID:23139841	(2012) Neutral lipid metabolism influences phospholipid sy...	7 of 28	8.31e-12
PMID:25461829	(2015) Overproduction and secretion of free fatty acids thr...	6 of 11	1.15e-11
PMID:22245606	(2012) Metabolism and regulation of glycerolipids in the ve...	9 of 134	1.15e-11

Export your current network:

... as a bitmap image: [download](#) *file format is 'PNG': portable network graphic*
 ... as a high-resolution bitmap: [download](#) *same PNG format, but resolution at 400 dpi*
 ... as a vector graphic: [download](#) *SVG: scalable vector graphic - can be opened and edited in Illustrator, CorelDraw, Dia, etc*
 ... as simple tabular text output: [download](#) *TSV: tab separated values - can be opened in Excel*
 ... as an XML summary: [download](#) *structured XML interaction data, according to the 'PSI-MI' data standard*
 ... network coordinates: [download](#) *a flat-file format describing the coordinates and colors of nodes in the network*
 ... protein sequences: [download](#) *MFA: multi-fasta format - containing the aminoacid sequences in the network*
 ... protein annotations: [download](#) *a tab-delimited file describing the names, domains and annotated functions of the network proteins*

Browse interactions in tabular form:

node1	node2	node1 accession	node2 accession	node1 annotation	node2 annotation	score
FAA1	FAA2	YOR317W	YER015W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Long-chain-fatty-acid-CoA ligase 2; Medium...	0.822
FAA1	FAA4	YOR317W	YMR246W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Long-chain-fatty-acid-CoA ligase 4; Long ch...	0.923
FAA1	FAS1	YOR317W	YKL182W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Beta subunit of fatty acid synthetase; compl...	0.968
FAA1	FAT1	YOR317W	YBR041W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Very long chain fatty acyl-CoA synthetase an...	0.974
FAA1	OLE1	YOR317W	YGL055W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Acyl-CoA desaturase 1; Delta(9) fatty acid d...	0.915
FAA1	POX1	YOR317W	YGL205W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Fatty-acyl coenzyme A oxidase; involved in t...	0.993
FAA1	TGL3	YOR317W	YMR313C	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Lipase 3; Bifunctional triacylglycerol lipase a...	0.884
FAA1	TGL4	YOR317W	YKR089C	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Lipase 4; Multifunctional lipase/hydrolase/p...	0.887
FAA1	YJU3	YOR317W	YKL094W	Long-chain-fatty-acid-CoA ligase 1; Long ch...	Monoglyceride lipase (MGL); functional orth...	0.863
FAA2	FAA1	YER015W	YOR317W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Long-chain-fatty-acid-CoA ligase 1; Long ch...	0.822
FAA2	FAA4	YER015W	YMR246W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Long-chain-fatty-acid-CoA ligase 4; Long ch...	0.819
FAA2	FAS1	YER015W	YKL182W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Beta subunit of fatty acid synthetase; compl...	0.952
FAA2	FAT1	YER015W	YBR041W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Very long chain fatty acyl-CoA synthetase an...	0.893
FAA2	OLE1	YER015W	YGL055W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Acyl-CoA desaturase 1; Delta(9) fatty acid d...	0.817
FAA2	POX1	YER015W	YGL205W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Fatty-acyl coenzyme A oxidase; involved in t...	0.997
FAA2	TGL3	YER015W	YMR313C	Long-chain-fatty-acid-CoA ligase 2; Medium...	Lipase 3; Bifunctional triacylglycerol lipase a...	0.719
FAA2	TGL4	YER015W	YKR089C	Long-chain-fatty-acid-CoA ligase 2; Medium...	Lipase 4; Multifunctional lipase/hydrolase/p...	0.733
FAA2	YJU3	YER015W	YKL094W	Long-chain-fatty-acid-CoA ligase 2; Medium...	Monoglyceride lipase (MGL); functional orth...	0.861
FAA4	FAA1	YMR246W	YOR317W	Long-chain-fatty-acid-CoA ligase 4; Long ch...	Long-chain-fatty-acid-CoA ligase 1; Long ch...	0.923
FAA4	FAA2	YMR246W	YER015W	Long-chain-fatty-acid-CoA ligase 4; Long ch...	Long-chain-fatty-acid-CoA ligase 2; Medium...	0.819

[Viewers >](#)[Legend >](#)[Settings >](#)[Analysis >](#)[Exports >](#)[Clusters ▾](#)[+ More](#)[- Less](#)

Clustering method

network clustering:

- ☒ no clustering (network is shown as it is)
- ☐ kmeans clustering (network is clustered to a specified number of clusters)
- ☐ MCL clustering (network is clustered to a specified 'MCL inflation parameter')

[APPLY](#)

Server load: low (10%) [HD]

[GO Permalink](#)

© STRING CONSORTIUM 2019



SIB - Swiss Institute of Bioinformatics



CPR - NNF Center for Protein Research



EMBL - European Molecular Biology Laboratory

ABOUT

[Content](#)
[References](#)
[Contributors](#)
[Statistics](#)

INFO

[Scores](#)
[Use scenarios](#)
[FAQs](#)
[Cookies/Privacy](#)

ACCESS

[Versions](#)
[APIs](#)
[Licensing](#)
[Usage](#)

CREDITS

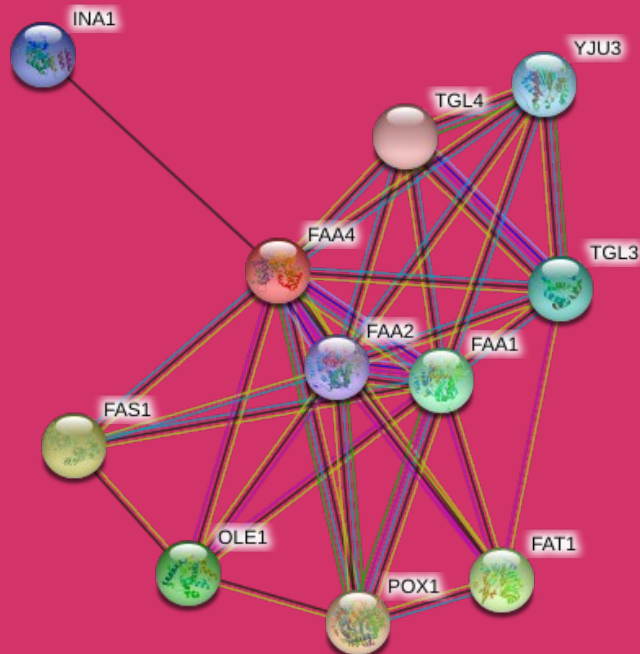
[Funding](#)
[Datasources](#)
[Partners](#)
[Software](#)




STRING is part of the ELIXIR infrastructure: it is one of ELIXIR's Core Data Resources. [Learn more >](#)

Exports & Views

BitMap Image




```
string_protein_sequences.fa-2.txt
>FAT1 4932.YBR041W Very long chain fatty acyl-CoA synthetase and fatty acid
transporter; activates imported fatty acids with a preference for very long lengths
(C20-C26); has a separate function in the transport of long chain fatty acids
MSPIQVVVFALSRIFFLLFRILIKLIITPIQKSLGSLFNGYFDELDRKYRYKEDWYIIPYFLKSVFCYIIDVRRHRFQNWYLFQ
VQNGDHLAISYTRPMAEKGEFQLETFYIETYNIVLRSLHILHFDYNNVQAGDYVAIDCTNKPLFVFLWLSLWNIGAIAPFLN
TKGTPLVHSLKISNITQVFIDPDASNPIRESEEEIKNALPDVKLNLYEEQDLMHLLNSQSPEFLQDQNVRTPLGLTDFKPSMLI
YTSGTTGLPKSAMSWSRKS SVGCQVFGHVLHMTNESTVFTAMPLFHSTAALLGACAILSHGGCLALSHKFSASTFWKQVYLTGAT
HIQYVGEVCRYLLHTPIISKYEKMHKVKVAYGNGLRPDIWQDFRKRNFIEVIGEFYAATEAPFATTTTQKGDFFGIGACRNYGTIIQ
WFLSFQQTLLVRMDPNDSDSVIYRNSKGFCEVAPVGPGEMLMRIFFPKKPKETSFGQYLGNAKETSKVVRDVFRRGDWYRCGDL
KADEYGLWYFLDRMGDTFRWKSENVSTTEVEDQLTASNKEQYQVLLVVGKVPKYEGRAGFAVILKTDNSLDITAKTKLLNDSLS
RLNLPYSAMPLFVKFVDEIKMTDNHKILKKVYREQKLPKGLDGNDFIWLKNYKRYEVLTAADWEAIDAQTIKL
>FAA2 4932.YER015W Long-chain-fatty-acid--CoA ligase 2; Medium chain fatty acyl-
CoA synthetase; activates imported fatty acids; accepts a wide range of fatty acid
chain lengths with a preference for medium chains, C9:0-C13:0; localized to the
peroxisome; comparative analysis suggests that a mitochondrially targeted form may
result from translation starting at a non-canonical codon upstream of the annotated
start codon
MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQPLTSPRYKTLFKKQAVAINSPDNEAGFSSIIYRSSLSSENLV
SCVDKNLRTAYDHMFSAARRWPQRDCLGSRPIDKATGTWEETFRFESYSTSVSKRCHNIGSGILSLVNTKRKRPLEANDFVVAILS
HNNPEWILTDLACQAYSLTNTALYETLGPNTESEYILNLTEAPILFAKSNMYHVLKMPVDMKFVNTLVCMDELTHDELRLMNL
LPVKNSNLNEKITFFSLQVEQVGCNFKIPAIPTPDLSLYTISFTSGTTGLPKGVEMSHRNIAISGIAFAFSTFRIPDKRNQLY
DMCFLPLAHIFERMVIAVDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVPRILTRFEAGIKNALDKSTVQRNVANTILDSKS
ARFTARGGPKDSIMNFLVYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYLTETFAVGVCLSEPFKDV
GSCGAIGISAECRLKSVPEMGYHADKDLKGLQIRGPQVFERFYFKNPNETSKAVDQDGFSTGDVAFIDGKGGRISVIDRVKNFFK
LAHGEYIAPEKIEINIYLSLSCPYITQIFVFGDPLKTLFVGIVGDVDDAAQPIAALKHPEVKTWKEVLVENLNRNKKLRKEFLNKI
NKCTDGLQGFEKLHNIVGLEPLTLEDDVVVPTFKIKRAKASKFFKDTLDQLYAEGLSVKTEKL
>OLE1 4932.YGL055W Acyl-CoA desaturase 1; Delta(9) fatty acid desaturase;
required for monounsaturated fatty acid synthesis and for normal distribution of
mitochondria
MPTSGTTIELIDDQFPKDDSSAGSIVDEVDLTEANILATGLNKKAPRIVNGFGSLMGSKEMVSVFEFKKGNEKKSNDRLLEKDN
```




Network currently showing

Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.




Experiments

Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.




Databases

Known metabolic pathways, protein complexes, signal transduction pathways, etc ... from curated databases.




Textmining

Automated, unsupervised textmining - searching for proteins that are frequently mentioned together.




Neighborhood

Groups of genes that are frequently observed in each other's genomic neighborhood.




Cooccurrence

Gene families whose occurrence patterns across genomes show similarities.



Coexpression

Proteins whose genes are observed to be correlated in expression, across a large number of experiments.



Fusion


Genes that are sometimes fused into single open reading frames.

STRING allows inspection of the interaction evidence for any given network. Choose any of the viewers above (disabled if not applicable in your network).

Server load: low (10%) [HD]

[GoPermalink](#)

© STRING CONSORTIUM 2019

-  SIB - Swiss Institute of Bioinformatics
-  CPR - NNF Center for Protein Research
-  EMBL - European Molecular Biology Laboratory

ABOUT

- Content
- References
- Contributors
- Statistics

INFO

- Scores
- Use scenarios
- FAQs
- Cookies/Privacy

ACCESS

- Versions
- APIs
- Licensing
- Usage

CREDITS

- Funding
- Datasources
- Partners
- Software



STRING is part of the ELIXIR infrastructure: it is one of ELIXIR's Core Data Resources. [Learn more >](#)

Co-expression

trpB
trpA
trpC
trpD
trpE
pabA
pabB
glyA
ilvA
serB
sdaA

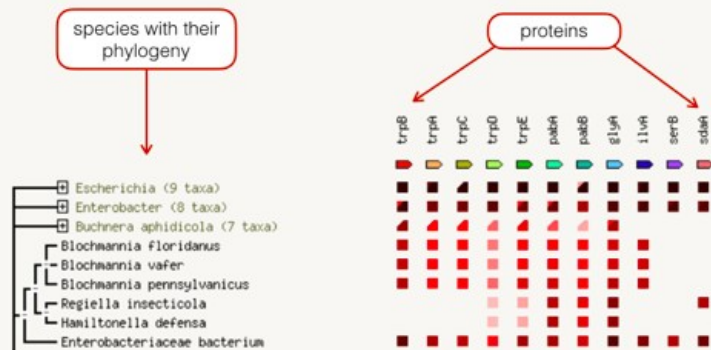
trpB
trpA
trpC
trpD
trpE
pabA
pabB
glyA
ilvA
serB
sdaA

species with their phylogeny

- Escherichia (9 taxa)
- Enterobacter (8 taxa)
- Buchnera aphidicola (7 taxa)
- Blochmannia floridanus
- Blochmannia vafer
- Blochmannia pennsylvanicus
- Regiella insecticola
- Hamiltonella defensa
- Enterobacteriaceae bacterium

proteins

trpB, trpA, trpC, trpD, trpE, padA, padB, glvA, ilvA, serB, sdhA



Experiments

protein-protein interaction (dip)
Detected by biochemical and biophysical assays

● trpB ● trpA

Databases

annotated pathway (KEGG)
Name: Biosynthesis of amino acids

● trpB ● trpA ● trpC ● trpD ● trpE ● glyA ● ilvA ● serB ● sdaA
[... and 109 other proteins]

Text mining

Investigation of the adaptation of *Lactococcus lactis* to isoleucine starvation integrating dynamic transcriptome and proteome information.
Dressaire C, Redon E, Giffon C, Loubière P, Monnet V, Cossign-Bousquet M
Microb Cell Fact. 10 Suppl 1:S18 (2011).

PubMed

● TrpB, trpB ● trpC ● trpD ● trpE ● pabB ● SdaA, sdaA ● GlyA, glyA ● pabA ● IlvA, ilvA ● SerB, serB ...



Methods & Algorithms

Sourcing

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:



Genomic Context
Predictions



High-throughput Lab
Experiments



(Conserved) Co-
Expression



Automated
Textmining



Previous Knowledge in
Databases

Coverage

The STRING database currently covers 9'643'763 proteins from 2'031 organisms.

Algorithm: Interaction Scoring

- Scores assigned to protein-protein interactions

- Calculated by integrating probabilities from various sources

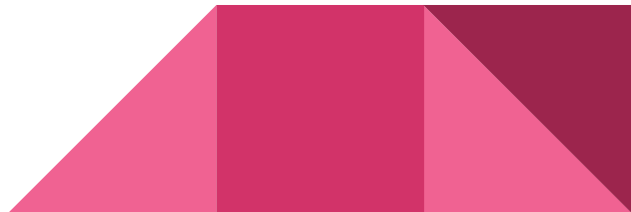
- Corrects for probability of random observation of interaction

- Indicate confidence, not strength or specificity
- 0.5 = Half of all interactions erroneous

- 'Transfer' Score

- Not originally observed in organism desired
- Normal score adapted based on homology

- low confidence - 0.15 (or better),
- medium confidence - 0.4,
- high confidence - 0.7,
- highest confidence - 0.9.




Experiments using STRING

- 1.) Researching protein-networks in the context of early immune system establishment
- 2.) Highly connected proteins have stable steady-state distribution of gene expression
- 3.) Searching for candidate genes involved in the immune response to gluten
- 4.) Identifying candidates for unknown enzyme in a pathway
- 5.) Using STRING to narrow the search space for two-locus epistasis
- 6.) Using STRING to show network connectivity
- 7.) STRING as a general purpose database
- 8.) STRING to guide experiments
- 9.) Prioritizing functional assignments in RNAi screens using interaction network data



Critiques

Critiques

- First time user - What is STRING?
 - About - Content
 - Powerful and Versatile tool
 - Text-mining
 - Several views of information
 - In-House view of gathered data
 - EMBL's IntAct
 - Provides links on Associations
 - Not visual
 - 6 Databases - Reduced Selection
 - High Quality User Manual - necessary?
- 



Thank You