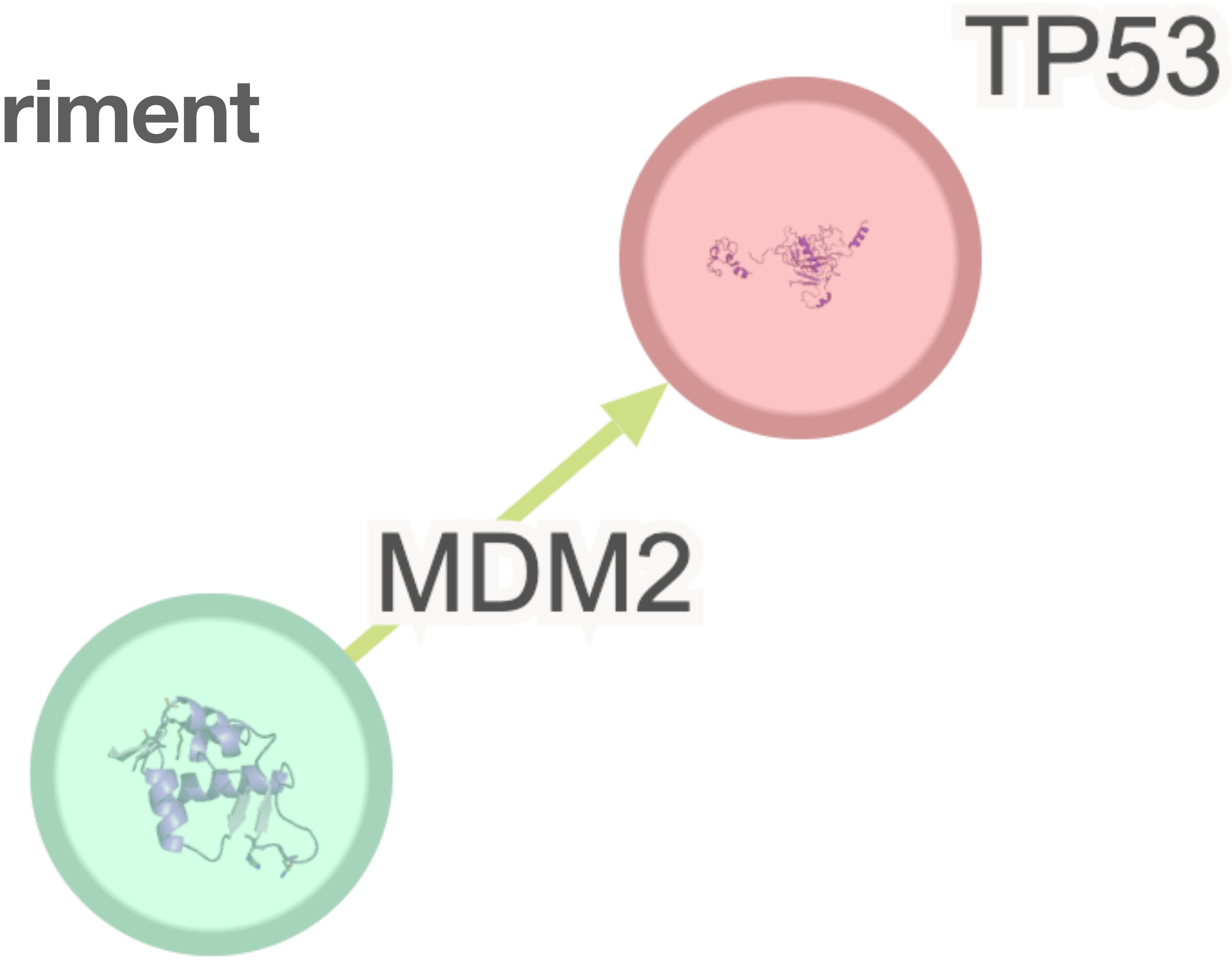
# The STRING database

## Data flow from the experiment to the discovery

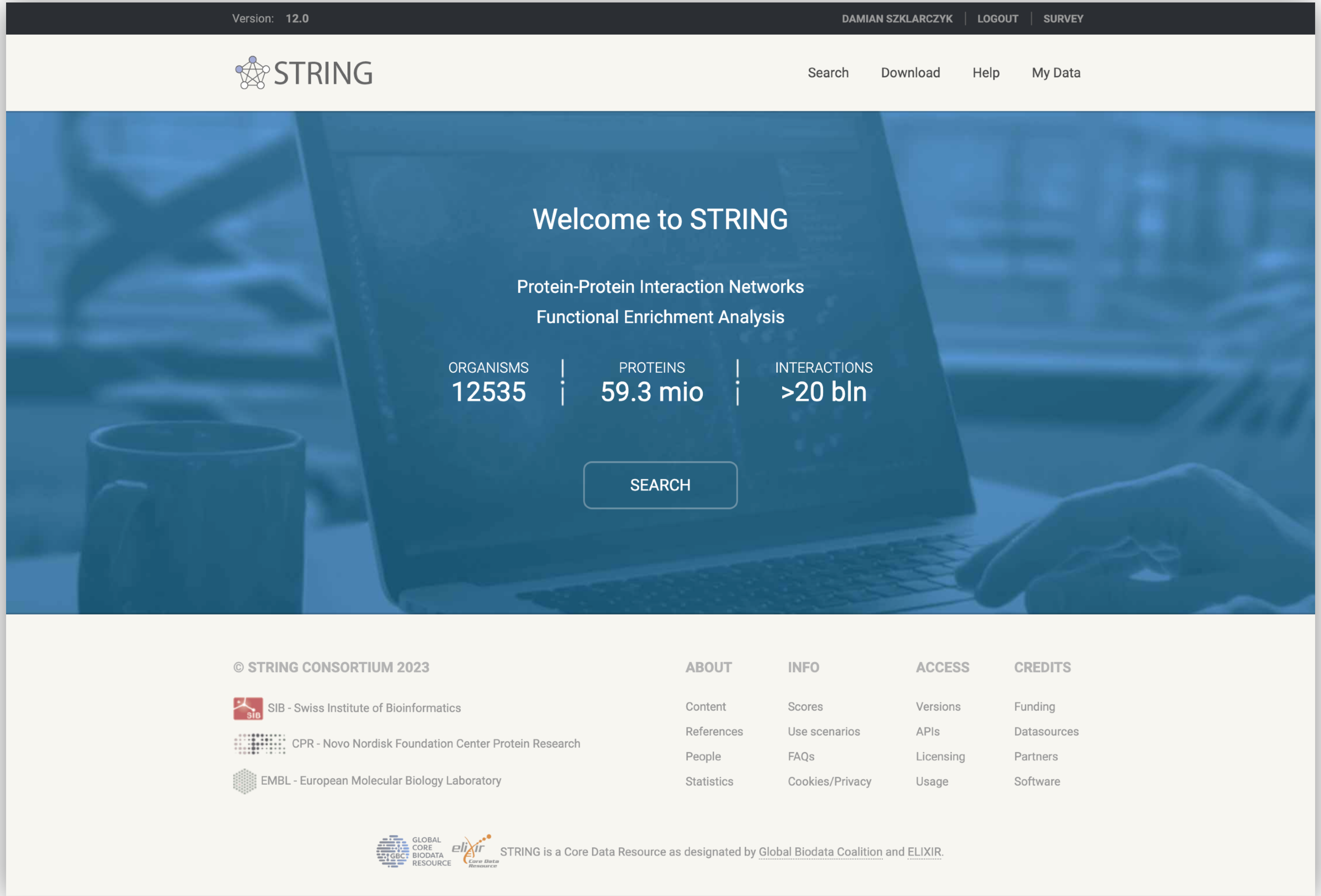
Damian Szklarczyk (UZH/SIB),




Swiss Institute of Bioinformatics

Universität Zürich[UZH]

2024, Life Science Industry Meets Data Science


Version: 12.0
DAMIAN SZKLARCZYK | LOGOUT | SURVEY
STRING
Search
Download
Help
My Data
Welcome to STRING
Protein-Protein Interaction Networks
Functional Enrichment Analysis
ORGANISMS 12535
PROTEINS 59.3 mio
INTERACTIONS >20 bln
SEARCH
© STRING CONSORTIUM 2023
SIB - Swiss Institute of Bioinformatics
CPR - Novo Nordisk Foundation Center Protein Research
EMBL - European Molecular Biology Laboratory
ABOUT
Content
References
People
Statistics
INFO
Scores
Use scenarios
FAQs
Cookies/Privacy
ACCESS
Versions
APIs
Licensing
Usage
CREDITS
Funding
Datasources
Partners
Software
GLOBAL CORE BIODATA RESOURCE
STRING is a Core Data Resource as designated by Global Biodata Coalition and ELIXIR.

# STRING

## What is it?

### What?

Protein-protein interaction network / DB

Functional Enrichment tool

### Access:

User Interface

Data dumps

API

Cytoscape / R packages

# STRING

## Functional Associations Networks

The linked proteins contribute to the same biological process.

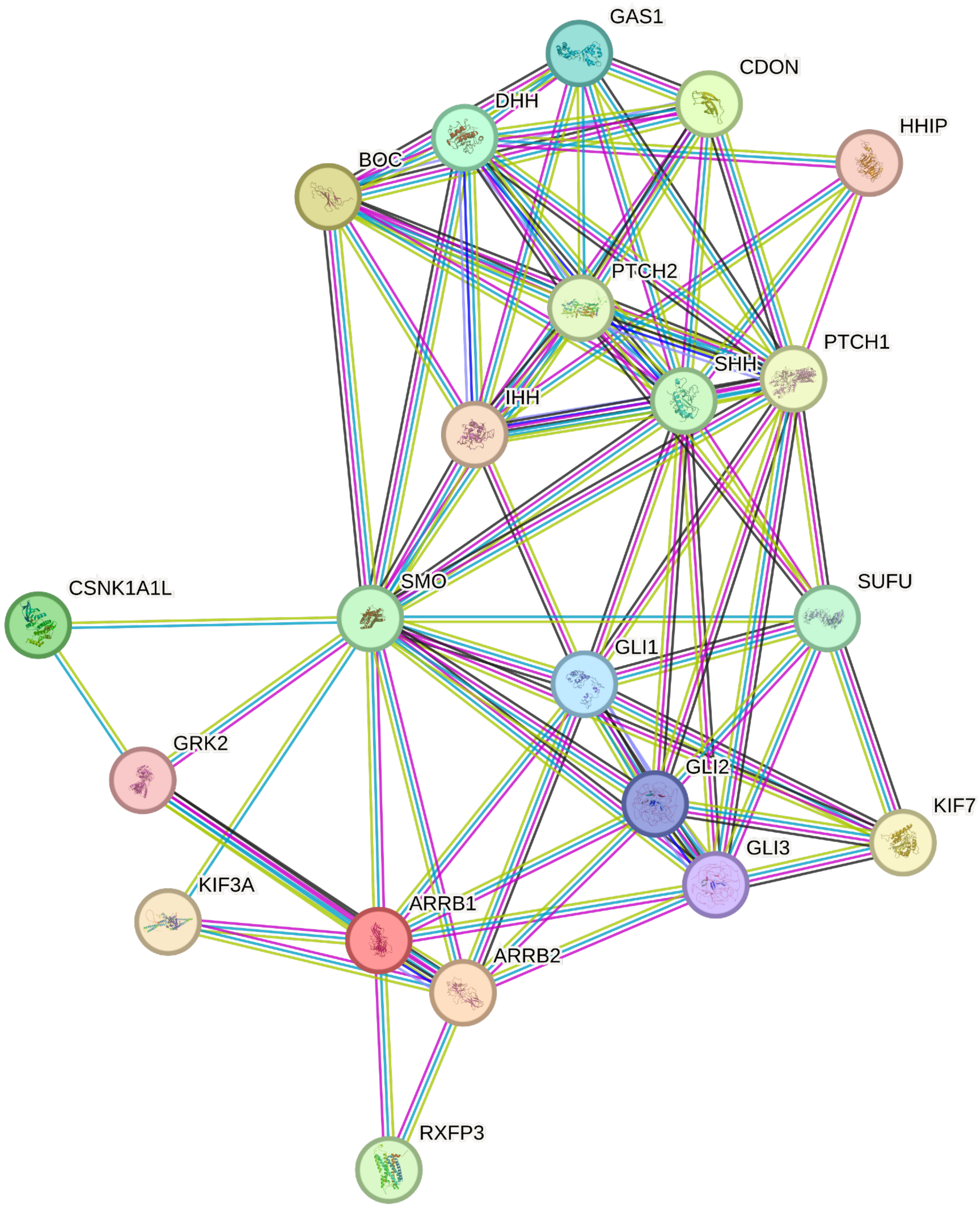
Useful abstraction / very predictive at various tasks

Allows us to derive the predictions from multiple sources at various modalities

Pretty dense

Direct / Indirect / stable / transient

Non-directional



# STRING

## Physical (Co-complex) Networks

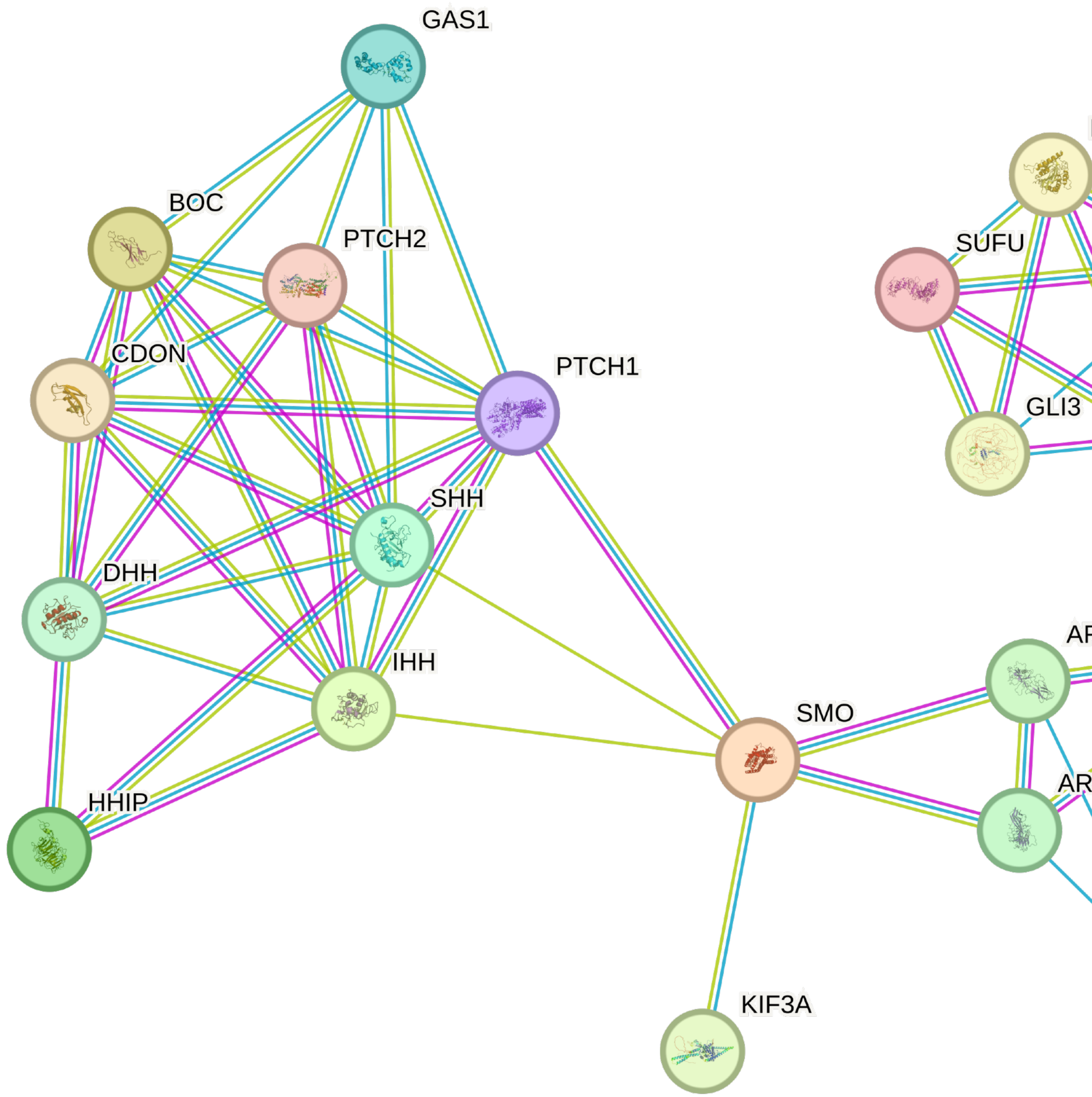
Proteins are part of the same physical complex.

Stable complexes / Transient interactions

The available data often does not define the complex structure

Curated knowledge / Text-mining / Experiments

Non-directional



# STRING

## Regulatory Networks

Protein A changes the activity / expression / stability of Protein B

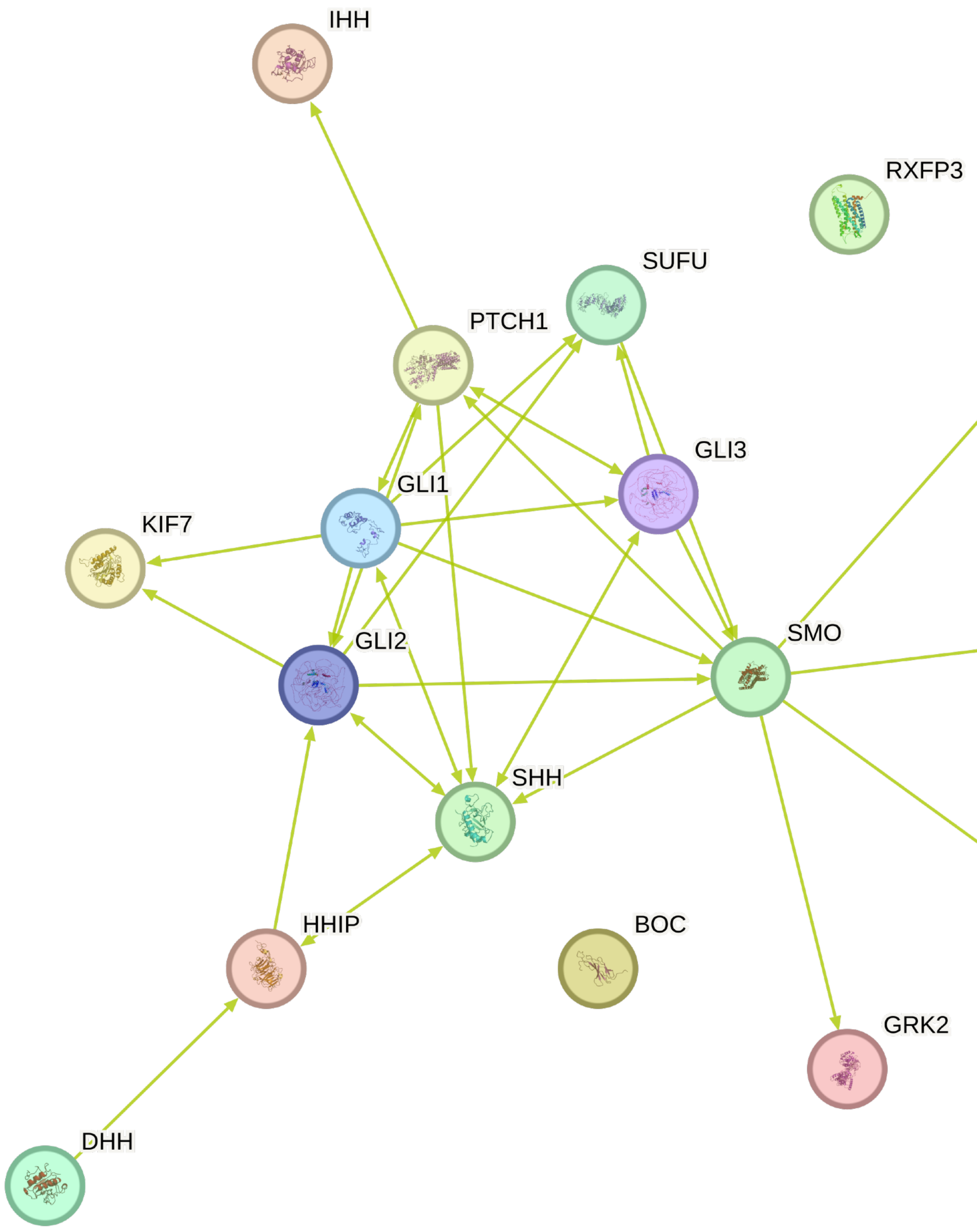
Focus on directionality of the interaction

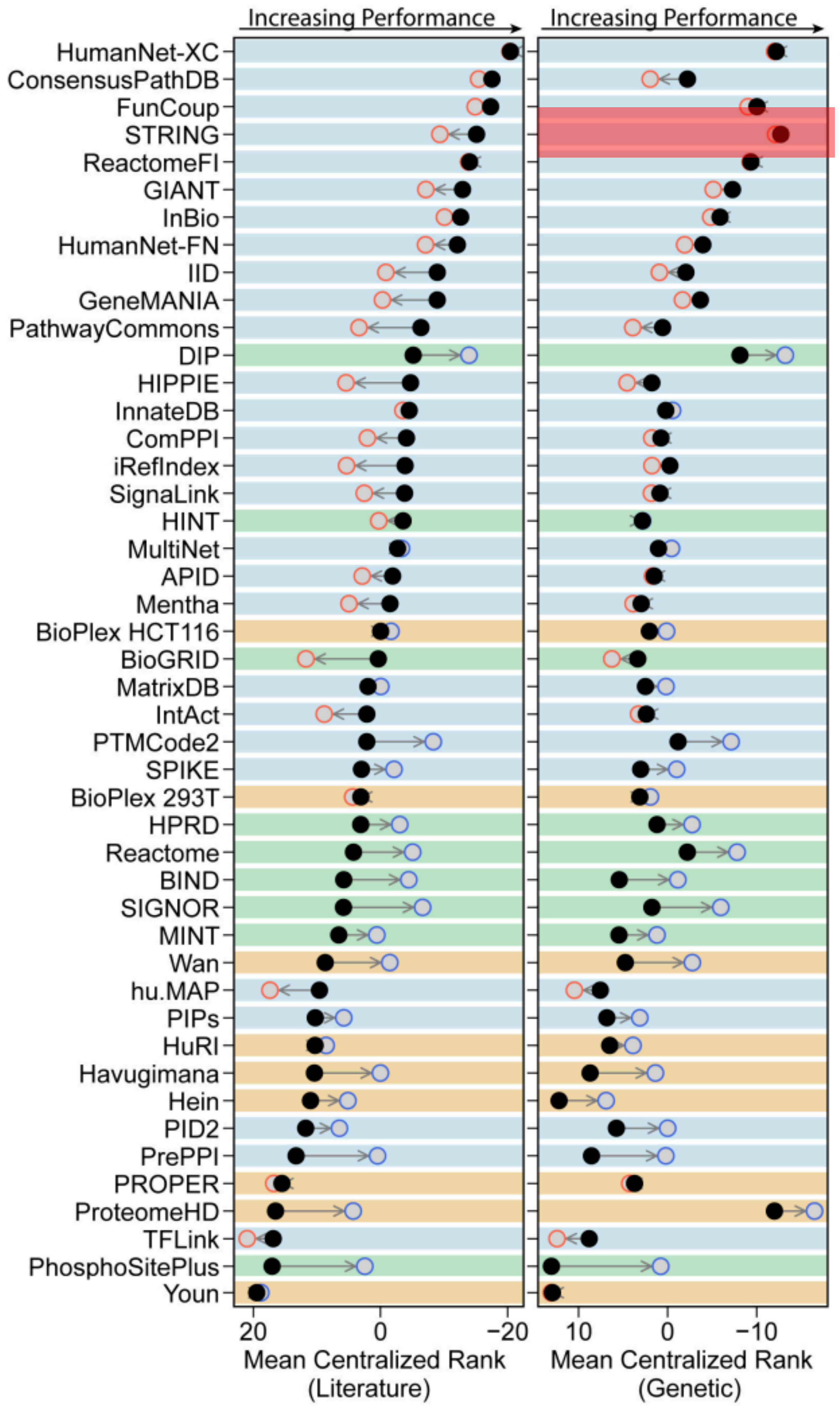
A regulates B

Postitive / Negative

Mode of regulation

Curated knowledge / Text-mining






# Best prediction performance for phenotype/disease associated genes on an independent benchmark

State of the Interactomes: an evaluation of molecular networks for generating biological insights.

S. Wright et al. Bioarxiv 2024

# STRING

## Users base ❤

1 mln unique users / >8500 citations (yearly)

Survey:

68% academia / hospitals

19% studnets

6% industry

5% gov- and non-gov orgs

# STRING

Licence

CC BY 4.0

# ATTRIBUTION 4.0 INTERNATIONAL

Deed

UNIVERSITY OF COPENHAGEN

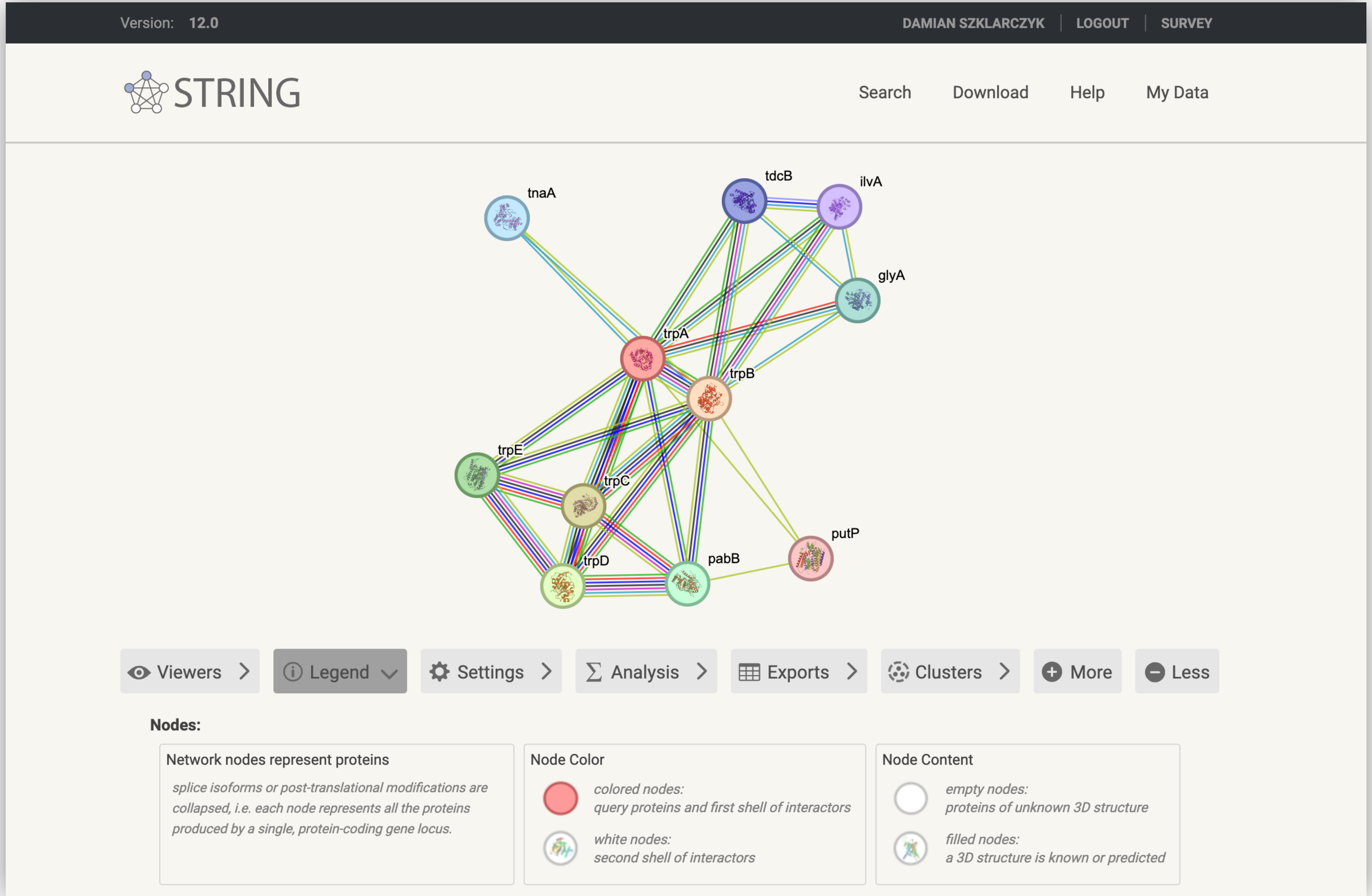
SIB Swiss Institute of Bioinformatics

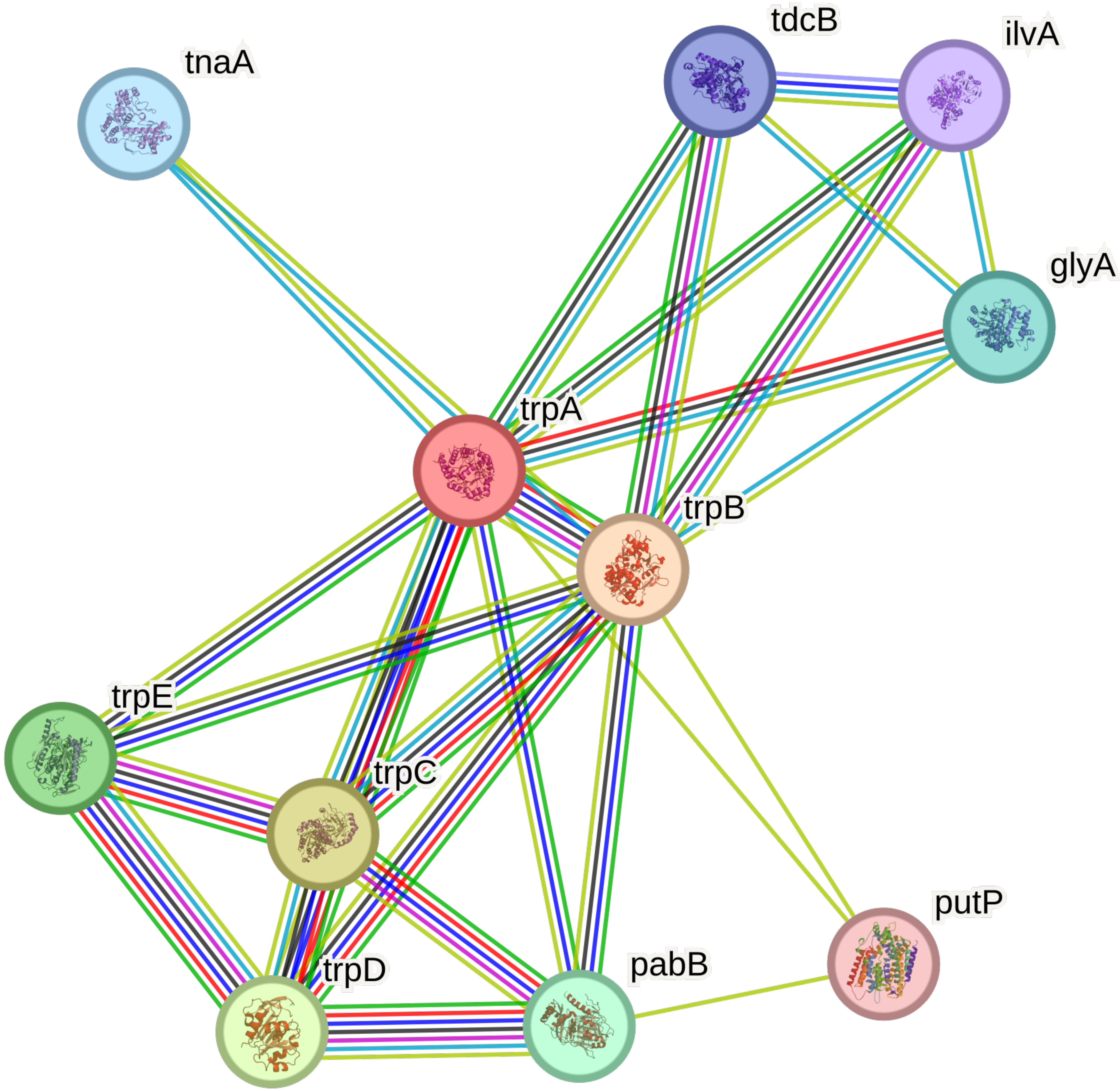
Novo Nordisk Foundation Center for Protein Research

EMBL

Universität Zürich[UZH]

UNIVERSITY OF TURKU


Version: 12.0
DAMIAN SZKLARCZYK
LOGOUT
SURVEY
STRING
Search
Download
Help
My Data
tnaA
tdcB
ilvA
glyA
trpA
trpB
trpE
trpC
trpD
pabB
putP
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:
a 3D structure is known or predicted


tnaA
tdcB
ilvA
glyA
trpA
trpB
trpE
trpC
trpD
pabB
putP

# STRING

## Evidence channels

- Gene neighborhood
- Gene fusions
- Phylogenetic profiles
- Experimentally validated
- Curated pathways
- Co-expression
- Text-mining

# STRING

## Evidence channels

**Gene neighborhood**

Gene fusions

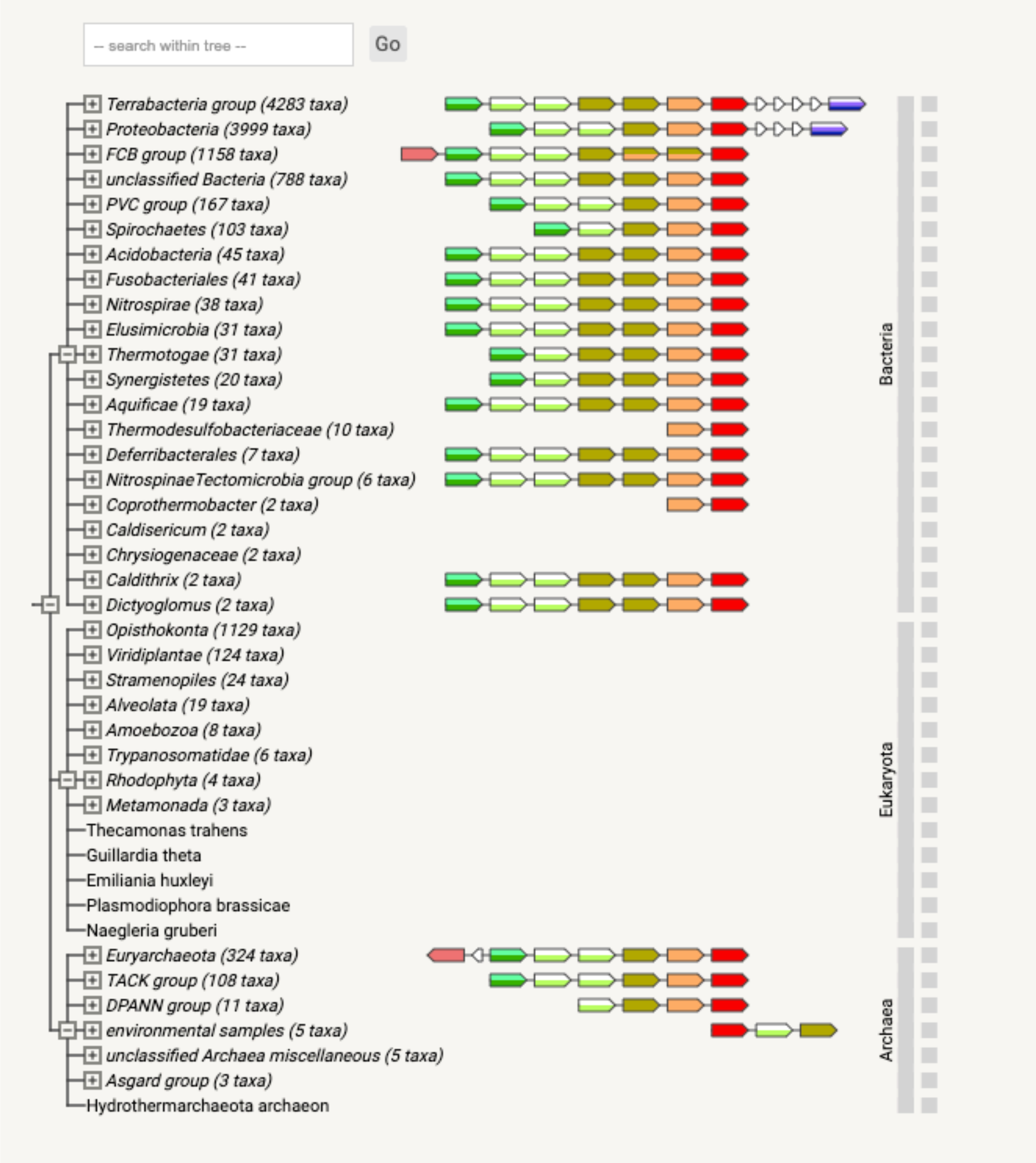
Phylogenetic profiles

Experimentally validated

Curated pathways

Co-expression

Text-mining



# STRING

## Evidence channels

Gene neighborhood

**Gene fusions**

Phylogenetic profiles

Experimentally validated

Curated pathways

Co-expression

Text-mining

# STRING
## Evidence channels

Gene neighborhood

Gene fusions

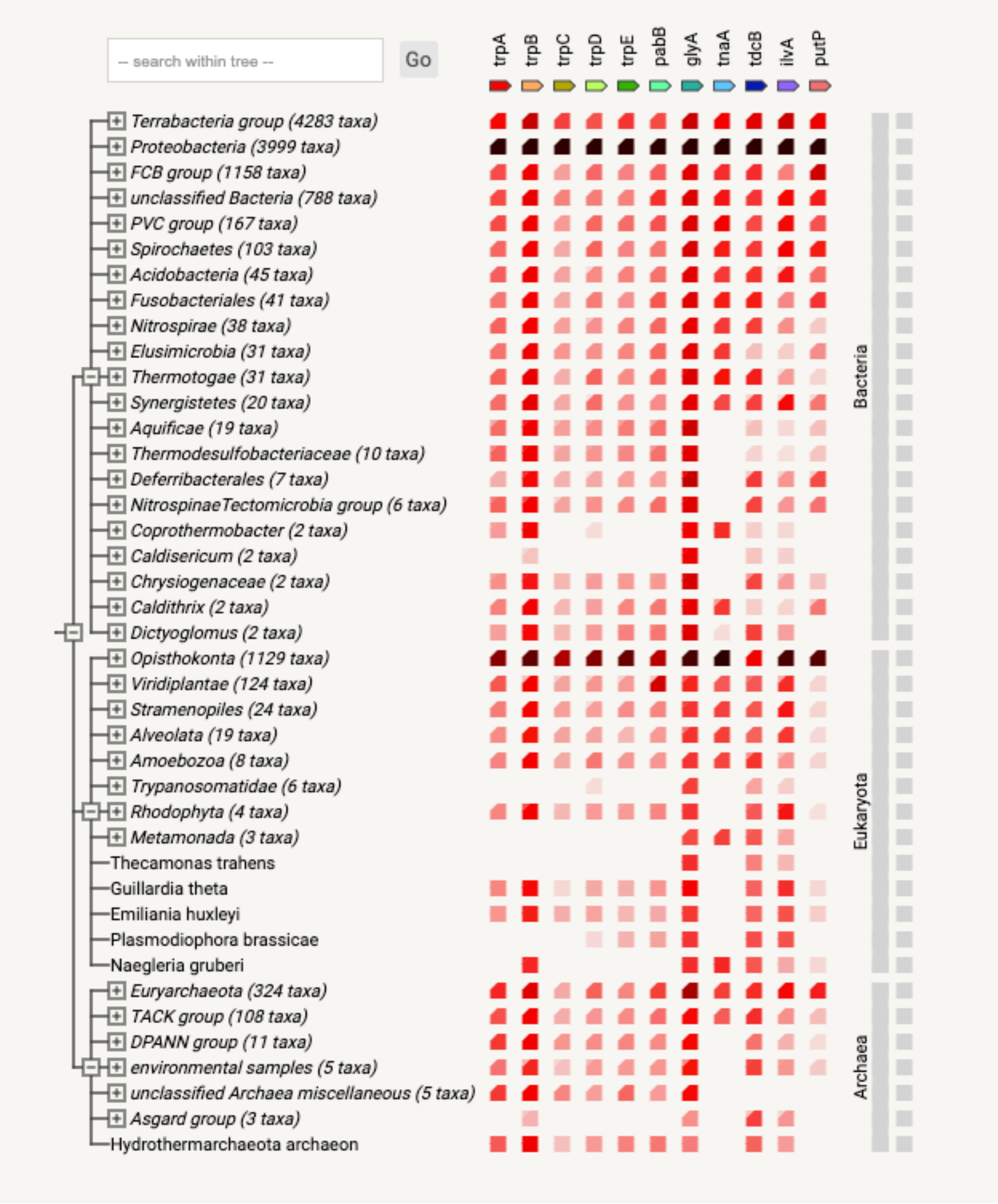
**Phylogenetic profiles**

Experimentally validated

Curated pathways

Co-expression

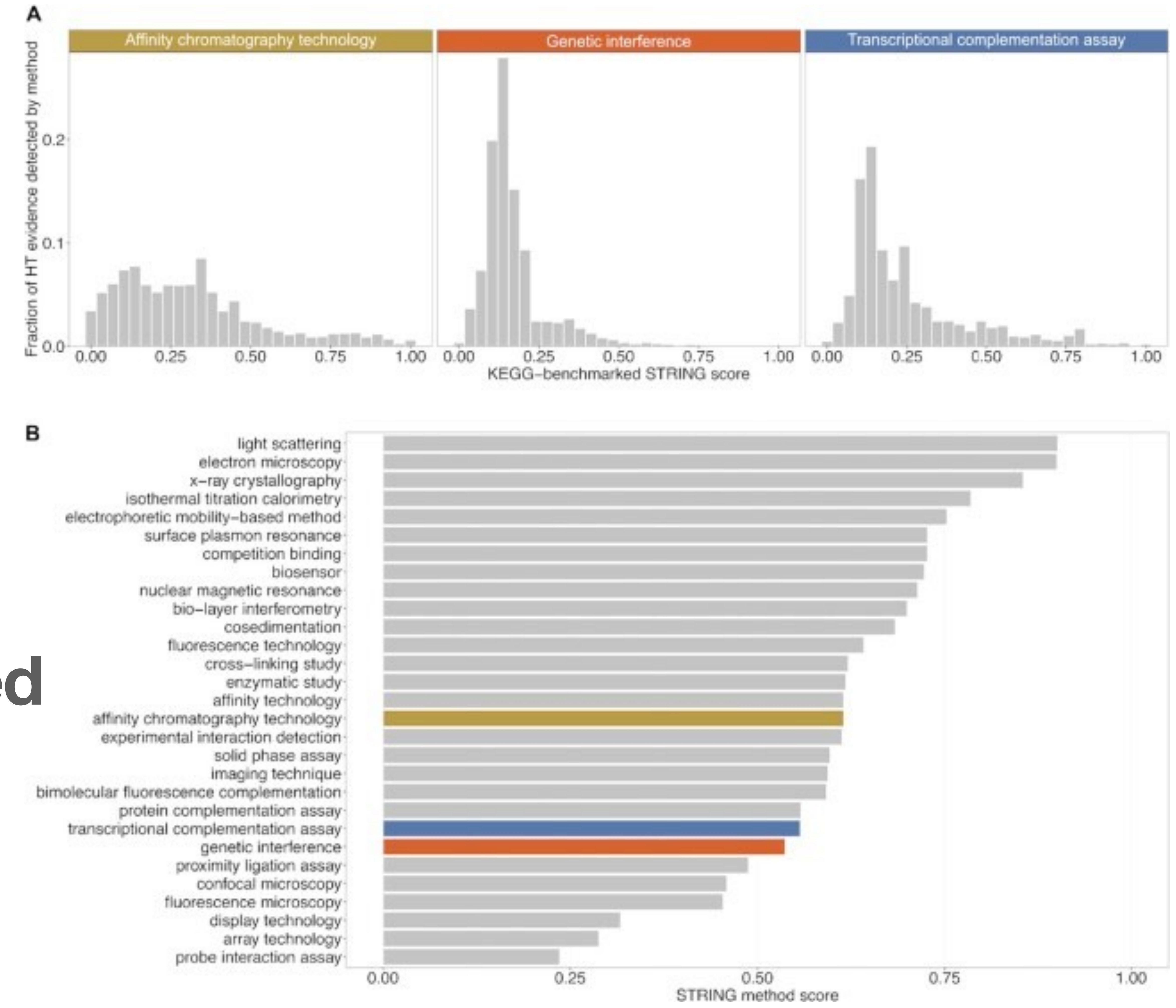
Text-mining



# STRING

## Evidence channels

- Gene neighborhood
- Gene fusions
- Phylogenetic profiles
- **Experimentally validated**
- Curated pathways
- Co-expression
- Text-mining

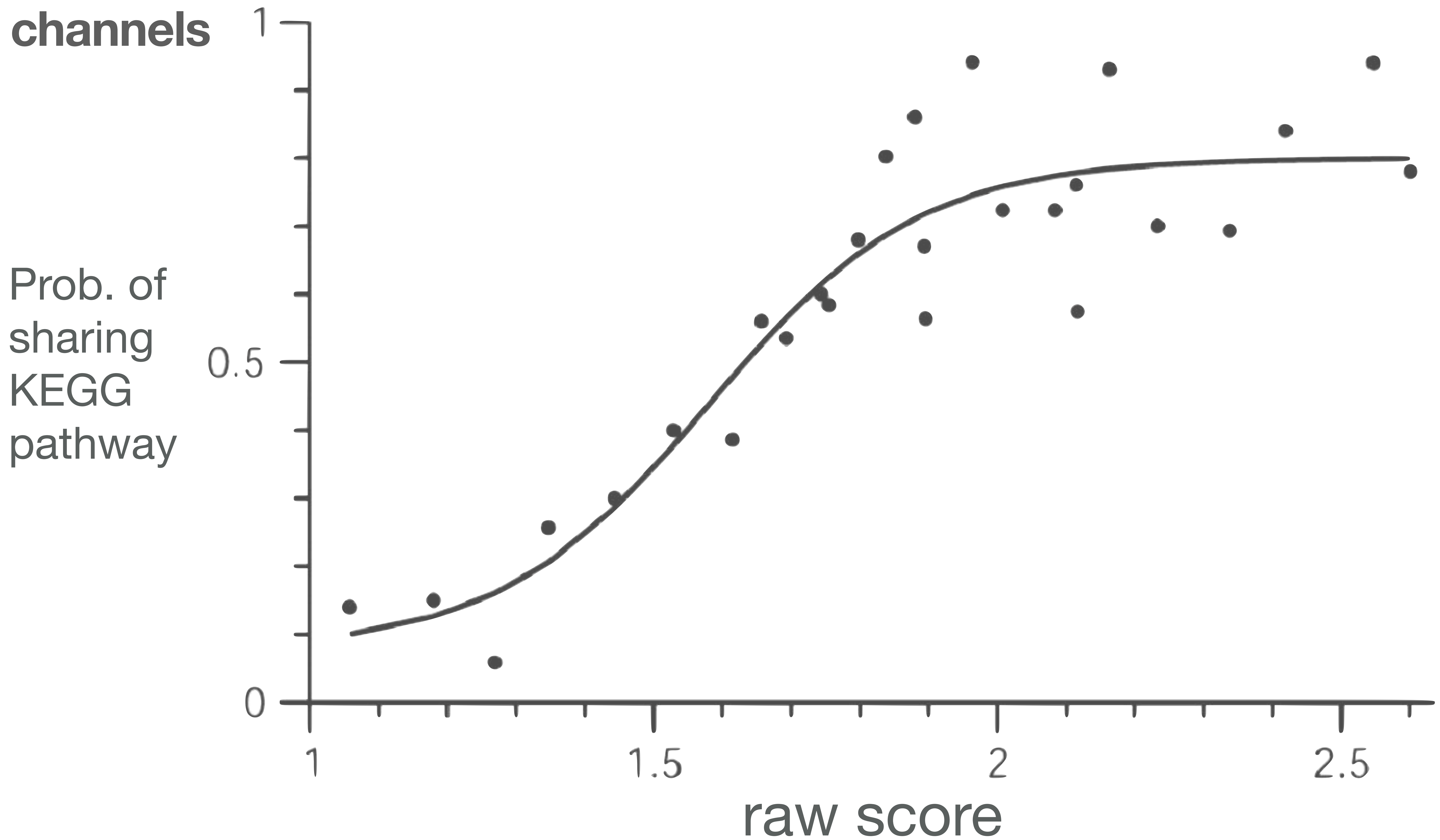



The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest.

D. Szklarczyk et al. NAR Db 2024

# STRING

## Evidence channels



# STRING

## Evidence channels

All channels are weighted equal.

# STRING

## Evidence channels

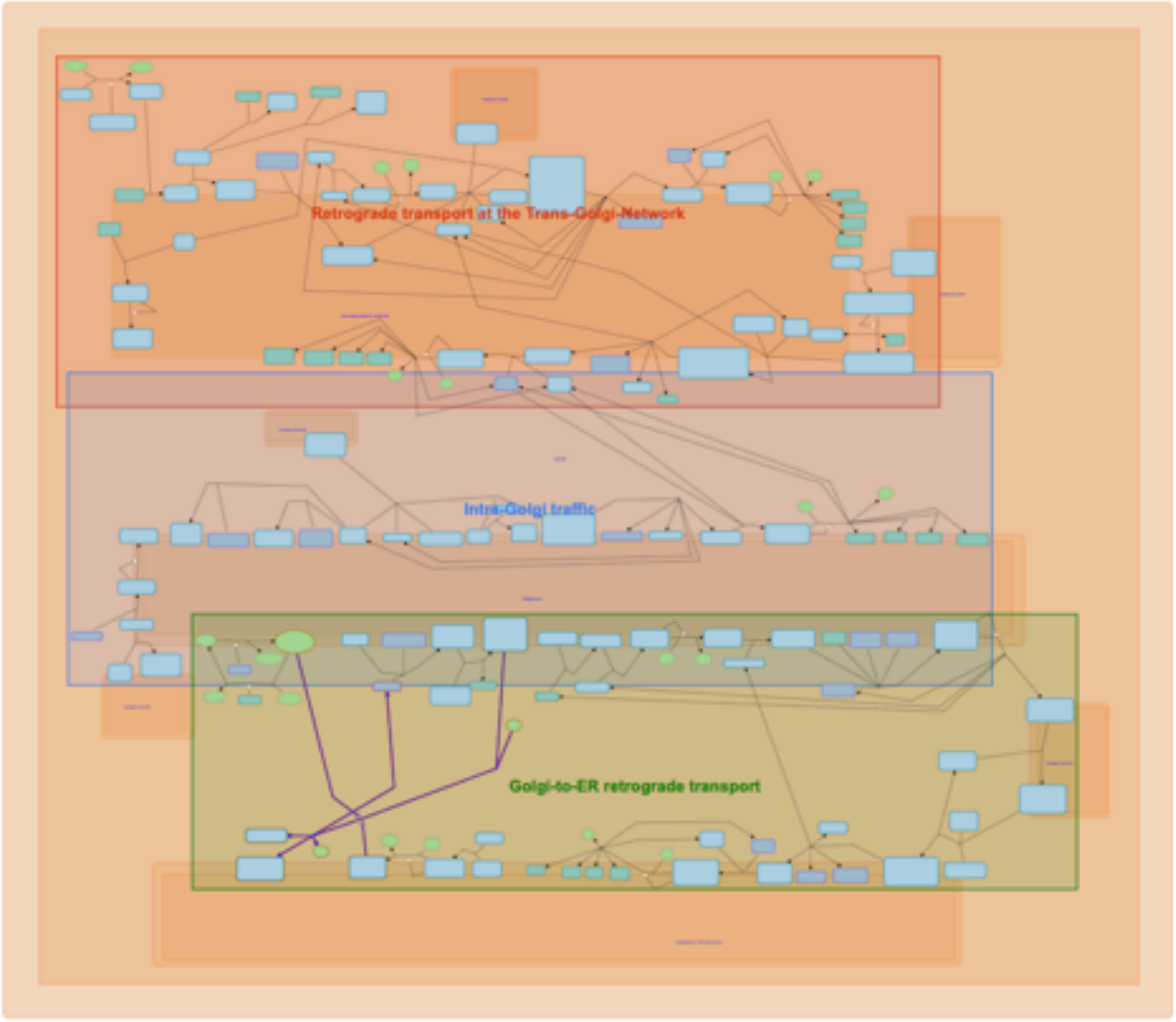
**0.5** Experimental = **0.5** Text-mining

$$S = 1 - \prod_{i=1}^{n} \left( \frac{1 - p_i}{1 - p_x} \right)$$

# STRING

## Evidence channels

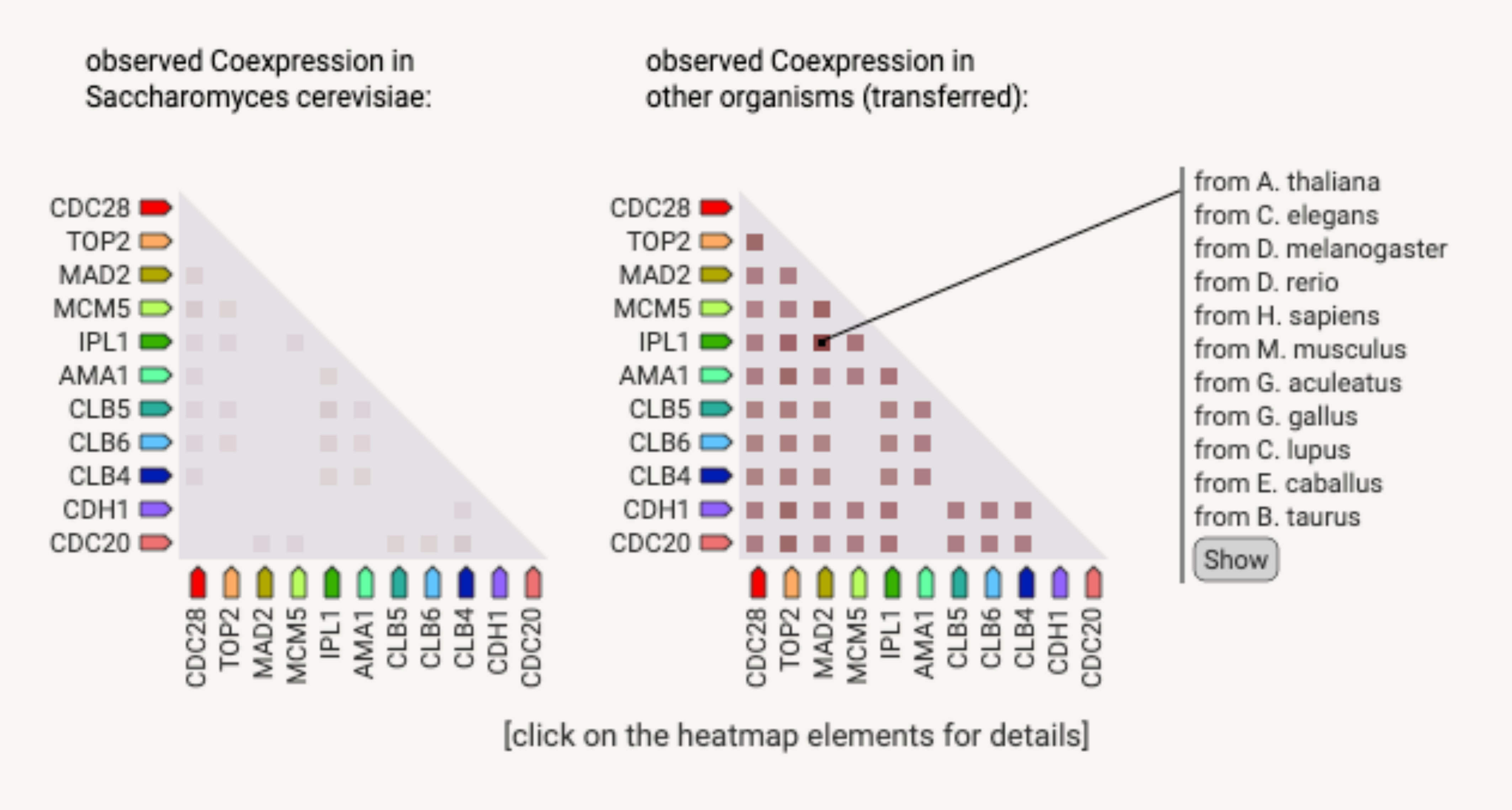
- Gene neighborhood
- Gene fusions
- Phylogenetic profiles
- Experimentally validated
- **Curated pathways**
- Co-expression
- Text-mining



# STRING

## Evidence channels

- Gene neighborhood
- Gene fusions
- Phylogenetic profiles
- Experimentally validated
- Curated pathways
- **Co-expression**
- Text-mining



# STRING

## Evidence channels

- Gene neighborhood
- Gene fusions
- Phylogenetic profiles
- Experimentally validated
- Curated pathways
- Co-expression
- **Text-mining**

**Relevant publications mentioning your query species (Escherichia coli K12):**

PMID:20534632 The synthetic integron: an in vivo genetic shuffling device.
*Bikard D, Julie-Galau S, Cambray G, Mazel D*
Nucleic Acids Res. 38(15):e153 2010.

**Abstract:**
As the field of synthetic biology expands, strategies and tools for the rapid construction of new biochemical pathways will become increasingly valuable. Purely rational design of complex biological pathways is inherently limited by the current state of our knowledge. Selection of optimal arrangements of genetic elements from randomized libraries may well be a useful approach for successful engineering. Here, we propose the construction and optimization of metabolic pathways using the inherent gene shuffling activity of a natural bacterial site-specific recombination system, the integron. As a proof of principle, we constructed and optimized a functional tryptophan biosynthetic operon in Escherichia coli. The **trpA** ()-E genes along with 'regulatory' elements were delivered as individual recombination cassettes in a synthetic integron platform. Integrase-mediated recombination generated thousands of genetic combinations overnight. We were able to isolate a large number of arrangements displaying varying fitness and tryptophan production capacities. Several assemblages required as many as six recombination events and produced as much as 11-fold more tryptophan than the natural gene order in the same context.

**Excerpts from full text:**
... ent integron cassettes and are abbreviations as follow. t, BioBrick terminator BBa_B0015; p, BioBrick promoter BBa_J23100 ('Materials and Methods' section); A, **trpA** (); B, **trpB** (); C, **trpC** (); D, **trpD** (); E, **trpE** (); Z, lacZalpha. The arrow means that the excised cassette was subsequently integrated at the attI site. The attC sites are numbered following their order in ...

PMID:33514556 Microbiota control of maternal behavior regulates early postnatal growth of offspring.
*Lee YM, Mu A, Wallace M, Gengatharan JM, Furst AJ, Bode L, Metallo CM, Ayres JS*
Sci Adv. 7(5) 2021.

**Abstract:**
Maternal behavior is necessary for optimal development and growth of offspring. The intestinal microbiota has emerged as a critical regulator of growth and development in the early postnatal period life. Here, we describe the identification of an intestinal Escherichia coli strain that is pathogenic to the maternal-offspring system during the early postnatal stage of life and results in growth stunting of the offspring. However, rather than having a direct pathogenic effect on the infant, we found that this particular E. coli strain was pathogenic to the dams by interfering with the maturation of maternal behavior. This resulted in malnourishment of the pups and impaired insulin-like growth factor 1 (IGF-1) signaling, leading to the consequential stunted growth. Our work provides a new understanding of how the microbiota regulates postnatal growth and an additional variable that must be considered when studying the regulation of maternal behavior.

**Excerpts from full text:**
... ecursor of serotonin, and gut microbes regulate tryptophan availability. We therefore directed our analyses toward genes associated with tryptophan metabolism (**tnaA** ()) and biosynthesis (trp operon). Analysis of the trp operon, by way of a multiple sequence alignment, revealed a concentrated number of amino acid substitutions in the region annotated as the bifunctional-**fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase** () (Fig. 5C and fig. S4). Furthermore, variant analysis identified a suite of misse [...] four strains, reported in parentheses as E. coli O6:H1 (ECN), E. coli O157:H7 (EHEC), and E. coli O21:H21 (ECO21) when compared with E. coli O16:H48 (MG1655): **tryptophan synthase subunit alpha** () (2, 3, and 1 variants), **tryptophan synthase subunit beta** () (2, 1, and 0 variants), **fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase** () (12, 10, and 4 variants), **anthranilate synthase subunit TrpD** () (1, 2, and 0 variants), and **anthranilate synthase subunit TrpE** () (5, 8, and 3 variants) (Fig. 5C, fig. S4, and data S1). Our analyses also included genes annotated as tryptophan transporters ...

PMID:26328893 XCluSim: a visual analytics tool for interactively comparing multiple clustering results of bioinformatics data.
*L'Yi S, Ko B, Shin D, Cho YJ, Lee J, Kim B, Seo J*
BMC Bioinformatics. 16 Suppl 11:S5 2015.

**Abstract:**
BACKGROUND: Though cluster analysis has become a routine analytic task for bioinformatics research, it is still arduous for researchers to assess the quality of a clustering result. To select the best clustering method and its parameters for a dataset, researchers have to run multiple clustering algorithms and compare them. However, such a comparison task with multiple clustering results is cognitively demanding and laborious. RESULTS: In this paper, we present XCluSim, a visual analytics tool that enables users to interactively compare multiple clustering results based on the Visual Information Seeking Mantra. We build a taxonomy for categorizing existing techniques of clustering results visualization in terms of the Gestalt principles of grouping. Using the taxonomy, we choose the most appropriate interactive visualizations for presenting individual clustering results from different types of clustering algorithms. The efficacy of XCluSim is shown through case studies with a bioinformatician. CONCLUSIONS: Compared to other relevant tools, XCluSim enables users to compare multiple clustering results in a more scalable manner. Moreover, XCluSim supports diverse clustering algorithms and dedicated visualizations and interactions for different types of clustering results, allowing more effective exploration of details on demand. Through case studies with a bioinformatics researcher, we received positive feedback on the functionalities of XCluSim, including its ability to help identify stably clustered items across multiple clustering results.

**Excerpts from full text:**
... elations in pathways. However, he recognized two problems in the result. First of all, a cluster that had both Arg and Art regulons also contained a gene named **tnaA** () that was considered to be noise. This was because **tnaA** () showed a different expression pattern and was not highly related to other cluster members in biologi [...] was the most satisfying result for the dataset. Additionally, our participant gained insight by seeing a stable group in XCluSim. Genes in the trp operon (i.e. **trpE** (), **trpD** (), **trpC** (), **trpB** (), and **trpA** ()) were stably clustered together with yciF through the four different results (see the highlighted stable group in Figure 7A). Since yciF was assigned to a ...

# STRING

## Textmining / Functional

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

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

# STRING

## Blocklisting

PMID:31212674

... elations in pathways. However, he recognized two problems in the result. First of all, a cluster that had both Arg and Art regulons also contained a gene named **tnaA** ( ) that was considered to be noise. This was because **tnaA** ( ) showed a different expression pattern and was not highly related to other cluster members in biologi [...] was the most satisfying result for the dataset. Additionally, our participant gained insight by seeing a stable group in XCluSim. Genes in the trp operon (i.e. **trpE** ( ), **trpD** ( ), **trpC** ( ), **trpB** ( ), and **trpA** ( )) were stably clustered together with yciF through the four different results (see the highlighted stable group in Figure 7A). Since yciF was assigned to a ...

PMID:32670211

... ecursor of serotonin, and gut microbes regulate tryptophan availability. We therefore directed our analyses toward genes associated with tryptophan metabolism (**tnaA** ( )) and biosynthesis (trp operon). Analysis of the trp operon, by way of a multiple sequence alignment, revealed a concentrated number of amino acid substitutions in the region annotated as the bifunctional-**fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase** ( ) (Fig. 5C and fig. S4). Furthermore, variant analysis identified a suite of misse [...] four strains, reported in parentheses as E. coli O6:H1 (ECN), E. coli O157:H7 (EHEC), and E. coli O21:H21 (ECO21) when compared with E. coli O16:H48 (MG1655): **tryptophan synthase subunit alpha** ( ) (2, 3, and 1 variants), **tryptophan synthase subunit beta** ( ) (2, 1, and 0 variants), **fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase** ( ) (12, 10, and 4 variants), **anthranilate synthase subunit TrpD** ( ) (1, 2, and 0 variants), and **anthranilate synthase subunit TrpE** ( ) (5, 8, and 3 variants) (Fig. 5C, fig. S4, and data S1). Our analyses also included genes annotated as tryptophan transporters ...

PMID:22380540

... of plant-associated bacteria amongst which many Pseudomonas spp.. Additionally, genes of the tryptophan biosynthesis gene cluster trp were present: trpD, trpG, **trpE** ( / ), **trpC** ( ), trpF, trpD, **trpB** ( ), **trpA** ( ). Strain VI4.1 has also the ability to produce IAA via an additional nitrilase-driven pathway converting indole-3-acetonitrile in IAA. The ability to produc ...

# STRING

## Blocklisting

### PMID:31212674

... elations in pathways. However, he recognized two problems in the result. First of all, a cluster that had both Arg and Art regulons also contained a gene named **tnaA** ( ) that was considered to be noise. This was because **tnaA** ( ) showed a different expression pattern and was not highly related to other cluster members in biologi [...] was the most satisfying result for the dataset. Additionally, our participant gained insight by seeing a stable group in XCluSim. Genes in the trp operon (i.e. **trpE** ( ), **trpD** ( ), **trpC** ( ), **trpB** ( ), and **trpA** ( )) were stably clustered together with yciF through the four different results (see the highlighted stable group in Figure 7A). Since yciF was assigned to a ...

### PMID:32670211

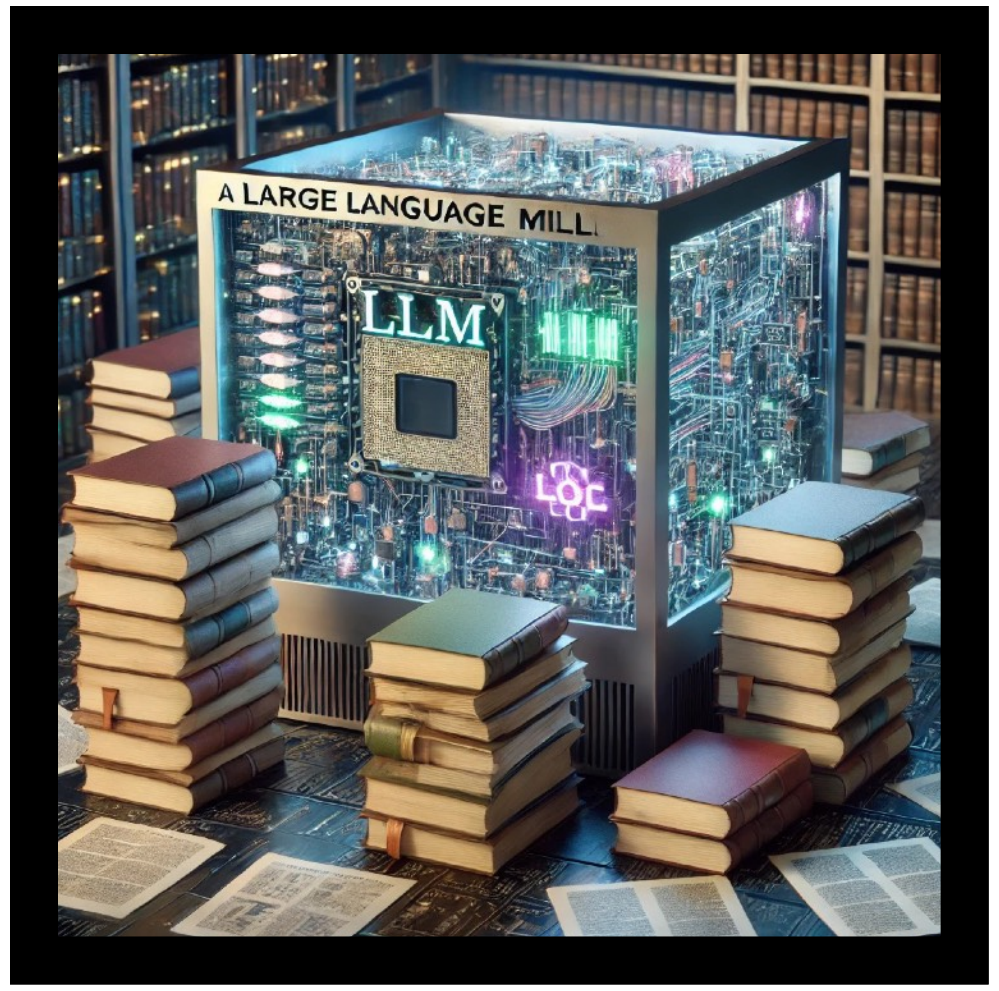
... ecursor of serotonin, and gut microbes regulate tryptophan availability. We therefore directed our analyses toward genes associated with tryptophan metabolism (**tnaA** ( )) and biosynthesis (trp operon). Analysis of the trp operon, by way of a multiple sequence alignment, revealed a concentrated number of amino acid substitutions in the region annotated as the bifunctional-**fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase** ( ) (Fig. 5C and fig. S4). Furthermore, variant analysis identified a suite of misse [...] four strains, reported in parentheses as E. coli O6:H1 (ECN), E. coli O157:H7 (EHEC), and E. coli O21:H21 (ECO21) when compared with E. coli O16:H48 (MG1655): **tryptophan synthase subunit alpha** ( ) (2, 3, and 1 variants), **tryptophan synthase subunit beta** ( ) (2, 1, and 0 variants), **fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase** ( ) (12, 10, and 4 variants), **anthranilate synthase subunit TrpD** ( ) (1, 2, and 0 variants), and **anthranilate synthase subunit TrpE** ( ) (5, 8, and 3 variants) (Fig. 5C, fig. S4, and data S1). Our analyses also included genes annotated as tryptophan transporters ...

### PMID:22380540

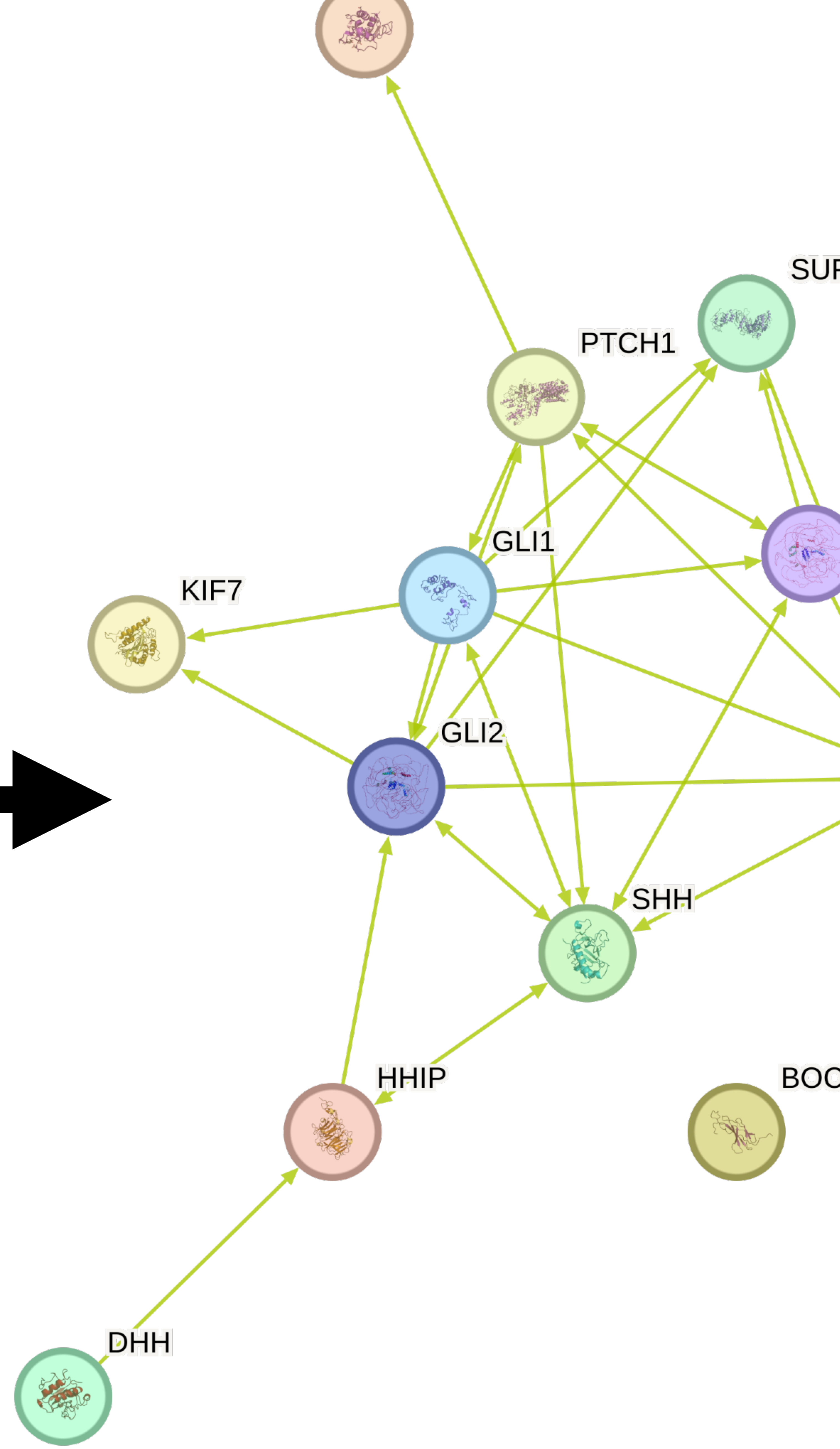
... of plant-associated bacteria amongst which many Pseudomonas spp.. Additionally, genes of the tryptophan biosynthesis gene cluster trp were present: trpD, trpG, **trpE** ( / ), **trpC** ( ), trpF, trpD, **trpB** ( ), **trpA** ( ). Strain VI4.1 has also the ability to produce IAA via an additional nitrilase-driven pathway converting indole-3-acetonitrile in IAA. The ability to produc ...

# STRING

## LLM - Regulatory / Physical Networks




| L | R | ACTION |
|---|---|---|
| - | - | - |
| - | - | - |
| - | - | - |
| - | - | - |



**Excerpts from full text:**

... s associated with a local regulation and the local regulators define a functional "elementary module" (see details in the Discussion). For example, the enzymes **TrpE** (), TrpG, **TrpD** (), **TrpF** () **TrpC** (), **TrpA** (), **TrpB** () involved in the tryptophan synthesis together with the regulators TRAP, RtpA, the tryptophan-specific T-box and the tryptophan-rich peptid [...] al regulation and the local regulators define a functional "elementary module" (see details in the Discussion). For example, the enzymes **TrpE** (), TrpG, **TrpD** (), **TrpF** () **TrpC** (), **TrpA** (), **TrpB** () involved in the tryptophan synthesis together with the regulators TRAP, RtpA, the tryptophan-specific T-box and the tryptophan-rich peptide leader Plead define an elementary module (see Additional file 5, Figure S26). Indeed the enzymatic complex **TrpE** ()+TrpG is regulated by TRAP at the genetic level, and by tryptophan at the enzymatic level. This elementary module definition is based on the mathematical de [...] amples of these ambiguities - glycine, proline and branched-chain amino acid syntheses - are presented below. Glycine is synthesized in one step from serine by **GlyA** () (see Additional file 5, Figure S13). The expression of **glyA** () is repressed by PurR, the global regulator involved in purine synthesis. Glycine is also requir [...] nscription involving the TRAP protein. TRAP binds to the nascent mRNA and prevents elongation of transcription. The translation of two genes of the trp operon (**trpE** () and **trpD** ()) is further modulated by TRAP (see below). tyrA and aroJ are also transcribed from an internal constitutive sigmaA-dependent promoter. In our model, the co [...] regulon. The state "on" of TRAP, which is defined as the conditions that enable the binding on mRNA, is described by TRAP also modulates the translation of the **trpE** () and **trpD** () genes by binding to the complete trp-mRNA. The **trpE** () ribosome-binding site (RBS) is then sequestrated in a RNA secondary structure preventing ribosome bindi [...] he **trpE** () ribosome-binding site (RBS) is then sequestrated in a RNA secondary structure preventing ribosome binding and **trpE** () translation. The coding sequences of **trpE** () and **trpD** () overlap by 29 nucleotides thereby coupling the translation of these two genes. The translation of **trpD** () is also affected by the binding of TRAP to the **trpE** () [...] g and **trpE** () translation. The coding sequences of **trpE** () and **trpD** () overlap by 29 nucleotides thereby coupling the translation of these two genes. The translation of **trpD** () is also affected by the binding of TRAP to the **trpE** () RBS. In our graphic representation, a mechanism of translational regulation is modelled by a hexagon, which is bound (state "on") or not (state "off") to t [...] TRAP-dependent modulation of the transcription of all the genes involved in the synthesis of tryptophan; (iii) TRAP-dependent modulation of the translation of **trpE** () and **trpD** (); (iv) TRAP sequestration by RtpA; and (v) conditions of transcription and translation of RtpA. The graphic representation of tyrosine synthesis displays th ...

# STRING

## LLM Model

### RoBERTa-large-PM

Fine-tuned:

**POSITIVE:** ~1000 curated abstracts containing complex formation information

**NEGATIVE:** ~300 curated abstracts containing information genetic interactions.

Max F1 score over 11 epochs

Softmax activation function on a decision layer.

**86.75%** average precision, **82.05%** average recall

# RoBERTa-large-PM-M3-Voc

**NLP-detected sentences suggesting physical associations:**
Genes encoding components of a strong complex (e.g., **trpA** ( )-**trpB** ( )) are more likely to be cotranscribed than genes encoding noncomplex-forming proteins acting in the same pathway (e.g., trpE-trpD). [...]

0.844

**NLP-detected sentence suggesting physical associations:**
Finally, tryptophan synthase (TS), also a tetrameric enzyme composed of two pairs of subunits (TS-alpha or **TrpA** ( ) and TS-beta or **TrpB** ( ) ( )), replaces the glycerol phosphate side-chain of indole-3-glycerol-phosphate by the alanyl moiety of an l-serine (fig. [...]

0.675

**NLP-detected sentences suggesting physical associations:**
Previous studies have indicated that in Chlamydia, **TrpA** ( ) has lost its catalytic activity yet remains associated with **TrpB** ( ) to support the beta reaction. [...]
Tryptophan synthase is a bifunctional protein-protein complex consisting of two subunits: alpha, encoded by **trpA** ( ), and beta, encoded by **trpB** ( ). [...]

0.490

**NLP-detected sentence suggesting physical associations:**
TrpS is comprised of two subunits, **TrpA** ( ) (alpha-subunit) and **TrpB** ( ) (beta-subunit) that function as a heterodimeric pair in an alphabetabetaalpha complex. [...]

0.905

**NLP-detected sentences suggesting physical associations:**
It is a heterotetramer, where two alpha-subunits (**TrpA** ( )) interact with the beta-subunit (**TrpB** ( )) dimer to form an alphabetabetaalpha complex. [...]
Bi-directional allosteric communication between **TrpA** ( ) and **TrpB** ( ) ensures that the generation of indole is coupled with its subsequent conversion to L-tryptophan. [...]

0.623

# RoBERTa-large-PM-M3-Voc

**NLP-detected sentences suggest** complex **s:**
Genes encoding components of a **pA** ( )-**trpB** ( )) are more likely to be cotranscribed than genes encoding noncomplex-forming proteins acting in the same pathway (e.g., trpE-trpD). [...]

0.844

**NLP-detected sentence suggesting physical associations:**
Finally, tryptophan synthase (TS), also a tetrameric enzyme composed of two pa subunits a or **TrpA** ( ) and TS-beta or **TrpB** ( ) ( )), replaces the glycerol phosphate side-chain of indole-3-glycerol-phosphate by the alanyl moiety

0.675

**NLP-detected sentences suggesting physical a**
Previous studies have indicated that in Chlamyd complex catalytic activity yet rer associated ) to support the beta reaction. [...]
Tryptophan synthase is a bifunctional protein-p g of two subunits: alpha, encoded by **trpA** ( ), and beta, encoded by **trpB** ( ). [...]

0.490

**NLP-detected sentence suggesting physical associations:**
TrpS is comprised of two subunits, **TrpA** ( ) (alpha-subunit) and **TrpB** ( ) (beta-subunit) that func heterodimeric phabetabetaalpha complex. [...]

0.905

**NLP-detected sentences suggesting physical associat** interact
It is a heterotetramer, where two alpha-subunits (**TrpA** ( beta-subunit (**TrpB** ( )) dimer to form an alphabetabetaalpha complex. [...]
Bi-directional allosteric communication between **TrpA** ( ) and **TrpB** ( ) ensures that the generation of indole is coupled with its subsequent conversion to L-tryptophan. [...]

0.623

# TRIGGER EXTRACTION

| | ĠThe | Ġformation | Ġof | Ġa | Ġstable | Ġtetrameric | Ġ | [unused1] | - | [unused1] | Ġcomplex | , | Ġwhich | Ġrequired | Ġthe | Ġdimerization | Ġdomain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 22 | 1.531 | 3.358 | 4.568 | 6.362 | 2.490 | 5.186 | 9.769 | 9.920 | 9.712 | 10.513 | 7.224 | 3.558 | 0.717 | 0.392 | 0.453 | 0.432 | 0.357 |
| 21 | 0.166 | 2.214 | 3.321 | 8.529 | 1.309 | 6.305 | 31.516 | 34.723 | 30.925 | 39.330 | 12.245 | 0.316 | 0.337 | 0.307 | 0.248 | 0.253 | 0.290 |
| 20 | -0.600 | 1.099 | 1.568 | 5.060 | 0.626 | 1.849 | 25.754 | 39.092 | 50.523 | 69.007 | 7.805 | -3.590 | 0.050 | 0.077 | -0.109 | -0.139 | 0.020 |
| 19 | -1.008 | 0.586 | 0.588 | 1.735 | 0.354 | 1.537 | 19.917 | 36.754 | 46.165 | 77.417 | 7.517 | -7.077 | 0.158 | 0.214 | 0.109 | 0.164 | 0.437 |
| 18 | -0.524 | -0.441 | -1.075 | -1.297 | -0.280 | -0.317 | 9.072 | 58.563 | 27.092 | 75.014 | -0.713 | -3.248 | -0.170 | 0.086 | 0.310 | 0.155 | 0.159 |
| 17 | -0.422 | -0.426 | -0.451 | -0.341 | -0.667 | -1.011 | 2.003 | 74.995 | 8.566 | 65.026 | -0.847 | -2.921 | -0.146 | 0.017 | 0.161 | -0.123 | -0.090 |
| 16 | -0.732 | 2.142 | 1.444 | 0.787 | 0.078 | -2.099 | 4.496 | 73.162 | 29.060 | 59.009 | 2.967 | -1.683 | 0.104 | -0.451 | -0.003 | 0.040 | -0.058 |
| 15 | -1.702 | 1.460 | 2.146 | 2.287 | 2.549 | 2.247 | 3.443 | 59.614 | 60.360 | 45.958 | 21.414 | -1.110 | -0.469 | -0.144 | -0.173 | 0.095 | -0.116 |
| 14 | -2.352 | 1.425 | 1.776 | 2.583 | 4.649 | 7.468 | 7.370 | 55.175 | 45.217 | 58.128 | 32.410 | -1.677 | -1.243 | -1.132 | -0.490 | 0.127 | -0.339 |
| 13 | -3.934 | -0.383 | 0.963 | 3.722 | 4.673 | 7.945 | 10.543 | 35.450 | 83.759 | 25.693 | 26.498 | -2.475 | -1.211 | 0.328 | -0.527 | 0.006 | -0.497 |
| 12 | -4.516 | 0.183 | 1.168 | 4.041 | 7.488 | 11.607 | 17.723 | 24.354 | 63.085 | -1.438 | 52.631 | -3.128 | -1.698 | -2.574 | -1.188 | 0.504 | -0.522 |
| 11 | -1.618 | 0.186 | -0.054 | 1.196 | 2.787 | 8.131 | 12.668 | 41.058 | 76.323 | 32.737 | 16.954 | -0.376 | -3.265 | 0.575 | 0.160 | 0.121 | 0.091 |
| 10 | 4.031 | 2.401 | 3.598 | 6.297 | 7.255 | 10.582 | 5.686 | -29.344 | 43.293 | -6.746 | 27.116 | -0.452 | 0.066 | 0.433 | 0.205 | -0.453 | -0.688 |
| 9 | 1.515 | 2.374 | 3.740 | 4.440 | 8.103 | 12.772 | 12.448 | 14.424 | 45.284 | 16.353 | 58.690 | 5.942 | 2.550 | -0.851 | 0.088 | -2.133 | -1.190 |
| 8 | 0.693 | 2.635 | 1.586 | 1.492 | 7.467 | 3.035 | 15.279 | -23.029 | 65.835 | -40.806 | 42.841 | 10.222 | 3.888 | 3.987 | 0.170 | -2.146 | -2.589 |
| 7 | 0.778 | 2.096 | 1.154 | 2.103 | 6.557 | 3.309 | 9.712 | -16.883 | 51.660 | -25.789 | 46.193 | 3.134 | 2.661 | 6.421 | 0.591 | -1.317 | -4.095 |
| 6 | 0.866 | -1.625 | 1.156 | 0.517 | 6.418 | 2.126 | 6.597 | -33.233 | 38.380 | 28.862 | 29.817 | 0.475 | -0.008 | 1.491 | -0.498 | -2.102 | -0.315 |
| 5 | 3.104 | 1.721 | 2.600 | 0.827 | 3.370 | -7.618 | 11.408 | -33.038 | 49.387 | -1.760 | 56.628 | 7.129 | -2.601 | 2.540 | 0.537 | -3.642 | -3.424 |
| 4 | 1.023 | -4.176 | -2.734 | -0.457 | 10.280 | -10.089 | -7.889 | -37.395 | 44.649 | 2.696 | 65.498 | 15.834 | 1.335 | 7.550 | 0.307 | -3.733 | -0.545 |
| 3 | -3.857 | -0.795 | -1.407 | 1.506 | 7.086 | -1.291 | -8.114 | -50.828 | 65.311 | -18.844 | 39.592 | -3.819 | 2.623 | -2.522 | -0.387 | -4.126 | -0.715 |
| 2 | -5.860 | -1.289 | -4.252 | 0.215 | 1.370 | 2.451 | -0.682 | -36.576 | 32.338 | -20.239 | 10.954 | -1.307 | 0.409 | 1.522 | -4.939 | -0.090 | -1.412 |
| 1 | -5.761 | 5.488 | -4.735 | 12.372 | 7.468 | 15.919 | -35.058 | -32.376 | 47.344 | -8.836 | 26.688 | -12.632 | -33.351 | 23.934 | 5.317 | -3.435 | -3.390 |
| 0 | 1.458 | -2.533 | -2.940 | -1.930 | 6.293 | 9.087 | -10.644 | 14.931 | 60.572 | 10.519 | 41.828 | -0.500 | 0.248 | -0.014 | 2.446 | -0.614 | 0.794 |
| embeddings | -2.905 | 0.707 | 4.800 | 5.773 | 5.332 | -1.487 | 0.904 | -0.386 | 7.242 | 6.091 | 11.238 | 16.226 | -1.650 | 1.032 | -2.435 | 0.429 | 0.244 |

# TRIGGER EXTRACTION

| 23 | ĠThe | Ġformation | Ġof | Ġa | Ġstable | Ġtetrameric | Ġ | [unused1] | - | [unused1] | Ġcomplex | , | Ġwhich | Ġrequired | Ġthe | Ġdimerization | Ġdomain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 21 | 1.531 | 3.358 | 4.568 | 6.362 | 2.490 | 5.186 | 9.769 | 9.920 | 9.712 | 10.513 | 7.224 | 3.558 | 0.717 | 0.392 | 0.453 | 0.432 | 0.357 |
| 20 | 0.166 | 2.214 | 3.321 | 8.529 | 1.309 | 6.305 | 31.516 | 34.723 | 30.925 | 39.330 | 12.245 | 0.316 | 0.337 | 0.307 | 0.248 | 0.253 | 0.290 |
| 19 | -0.600 | 1.099 | 1.568 | 5.060 | 0.626 | 1.849 | 25.754 | 39.092 | 50.523 | 69.007 | 7.805 | -3.590 | 0.050 | 0.077 | -0.109 | -0.139 | 0.020 |
| 18 | -1.008 | 0.586 | 0.588 | 1.735 | 0.354 | 1.537 | 19.917 | 36.754 | 46.165 | 77.417 | 7.517 | -7.077 | 0.158 | 0.214 | 0.109 | 0.164 | 0.437 |
| 17 | -0.524 | -0.441 | -1.075 | -1.297 | -0.280 | -0.317 | 9.072 | 58.563 | 27.092 | 75.014 | -0.713 | -3.248 | -0.170 | 0.086 | 0.310 | 0.155 | 0.159 |
| 16 | -0.422 | -0.426 | -0.451 | -0.341 | -0.667 | -1.011 | 2.003 | 74.995 | 8.566 | 65.026 | -0.847 | -2.921 | -0.146 | 0.017 | 0.161 | -0.123 | -0.090 |
| 15 | -0.732 | 2.142 | 1.444 | 0.787 | 0.078 | -2.099 | 4.496 | 73.162 | 29.060 | 59.009 | 2.967 | -1.683 | 0.104 | -0.451 | -0.003 | 0.040 | -0.058 |
| 14 | -1.702 | 1.460 | 2.146 | 2.287 | 2.549 | 2.247 | 3.443 | 59.614 | 60.360 | 45.958 | 21.414 | -1.110 | -0.469 | -0.144 | -0.173 | 0.095 | -0.116 |
| 13 | -2.352 | 1.425 | 1.776 | 2.583 | 4.649 | 7.468 | 7.370 | 55.175 | 45.217 | 58.128 | 32.410 | -1.677 | -1.243 | -1.132 | -0.490 | 0.127 | -0.339 |
| 12 | -3.934 | -0.383 | 0.963 | 3.722 | 4.673 | 7.945 | 10.543 | 35.450 | 83.759 | 25.693 | 26.498 | -2.475 | -1.211 | 0.328 | -0.527 | 0.006 | -0.497 |
| 11 | -4.516 | 0.183 | 1.168 | 4.041 | 7.488 | 11.607 | 17.723 | 24.354 | 63.085 | -1.438 | 52.631 | -3.128 | -1.698 | -2.574 | -1.188 | 0.504 | -0.522 |
| 10 | -1.618 | 0.186 | -0.054 | 1.196 | 2.787 | 8.131 | 12.668 | 41.058 | 76.323 | 32.737 | 16.954 | -0.376 | -3.265 | 0.575 | 0.160 | 0.121 | 0.091 |
| 9 | 4.031 | 2.401 | 3.598 | 6.297 | 7.255 | 10.582 | 5.686 | -29.344 | 43.293 | -6.746 | 27.116 | -0.452 | 0.066 | 0.433 | 0.205 | -0.453 | -0.688 |
| 8 | 1.515 | 2.374 | 3.740 | 4.440 | 8.103 | 12.772 | 12.448 | 14.424 | 45.284 | 16.353 | 58.690 | 5.942 | 2.550 | -0.851 | 0.088 | -2.133 | -1.190 |
| 7 | 0.693 | 2.635 | 1.586 | 1.492 | 7.467 | 3.035 | 15.279 | -23.029 | 65.835 | -40.806 | 42.841 | 10.222 | 3.888 | 3.987 | 0.170 | -2.146 | -2.589 |
| 6 | 0.778 | 2.096 | 1.154 | 2.103 | 6.557 | 3.309 | 9.712 | -16.883 | 51.660 | -25.789 | 46.193 | 3.134 | 2.661 | 6.421 | 0.591 | -1.317 | -4.095 |
| 5 | 0.866 | -1.625 | 1.156 | 0.517 | 6.418 | 2.126 | 6.597 | -33.233 | 38.380 | 28.862 | 29.817 | 0.475 | -0.008 | 1.491 | -0.498 | -2.102 | -0.315 |
| 4 | 3.104 | 1.721 | 2.600 | 0.827 | 3.370 | -7.618 | 11.408 | -33.038 | 49.387 | -1.760 | 56.628 | 7.129 | -2.601 | 2.540 | 0.537 | -3.642 | -3.424 |
| 3 | 1.023 | -4.176 | -2.734 | -0.457 | 10.280 | -10.089 | -7.889 | -37.395 | 44.649 | 2.696 | 65.498 | 15.834 | 1.335 | 7.550 | 0.307 | -3.733 | -0.545 |
| 2 | -3.857 | -0.795 | -1.407 | 1.506 | 7.086 | -1.291 | -8.114 | -50.828 | 65.311 | -18.844 | 39.592 | -3.819 | 2.623 | -2.522 | -0.387 | -4.126 | -0.715 |
| 1 | -5.860 | -1.289 | -4.252 | 0.215 | 1.370 | 2.451 | -0.682 | -36.576 | 32.338 | -20.239 | 10.954 | -1.307 | 0.409 | 1.522 | -4.939 | -0.090 | -1.412 |
| 0 | -5.761 | 5.488 | -4.735 | 12.372 | 7.468 | 15.919 | -35.058 | -32.376 | 47.344 | -8.836 | 26.688 | -12.632 | -33.351 | 23.934 | 5.317 | -3.435 | -3.390 |
| embeddings | 1.458 | -2.533 | -2.940 | -1.930 | 6.293 | 9.087 | -10.644 | 14.931 | 60.572 | 10.519 | 41.828 | -0.500 | 0.248 | -0.014 | 2.446 | -0.614 | 0.794 |
| | -2.905 | 0.707 | 4.800 | 5.773 | 5.332 | -1.487 | 0.904 | -0.386 | 7.242 | 6.091 | 11.238 | 16.226 | -1.650 | 1.032 | -2.435 | 0.429 | 0.244 |

# TRIGGER EXTRACTION

| | ĠThe | Ġformation | Ġof | Ġa | Ġstable | Ġtetrameric | Ġ | [unused1] | - | [unused1] | Ġcomplex | , | Ġwhich | Ġrequired | Ġthe | Ġdimerization | Ġdomain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 22 | 1.531 | 3.358 | 4.568 | 6.362 | 2.490 | 5.186 | 9.769 | 9.920 | 9.712 | 10.513 | 7.224 | 3.558 | 0.717 | 0.392 | 0.453 | 0.432 | 0.357 |
| 21 | 0.166 | 2.214 | 3.321 | 8.529 | 1.309 | 6.305 | 31.516 | 34.723 | 30.925 | 39.330 | 12.245 | 0.316 | 0.337 | 0.307 | 0.248 | 0.253 | 0.290 |
| 20 | -0.600 | 1.099 | 1.568 | 5.060 | 0.626 | 1.849 | 25.754 | 39.092 | 50.523 | 69.007 | 7.805 | -3.590 | 0.050 | 0.077 | -0.109 | -0.139 | 0.020 |
| 19 | -1.008 | 0.586 | 0.588 | 1.735 | 0.354 | 1.537 | 19.917 | 36.754 | 46.165 | 77.417 | 7.517 | -7.077 | 0.158 | 0.214 | 0.109 | 0.164 | 0.437 |
| 18 | -0.524 | -0.441 | -1.075 | -1.297 | -0.280 | -0.317 | 9.072 | 58.563 | 27.092 | 75.014 | -0.713 | -3.248 | -0.170 | 0.086 | 0.310 | 0.155 | 0.159 |
| 17 | -0.422 | -0.426 | -0.451 | -0.341 | -0.667 | -1.011 | 2.003 | 74.995 | 8.566 | 65.026 | -0.847 | -2.921 | -0.146 | 0.017 | 0.161 | -0.123 | -0.090 |
| 16 | -0.732 | 2.142 | 1.444 | 0.787 | 0.078 | -2.099 | 4.496 | 73.162 | 29.060 | 59.009 | 2.967 | -1.683 | 0.104 | -0.451 | -0.003 | 0.040 | -0.058 |
| 15 | -1.702 | 1.460 | 2.146 | 2.287 | 2.549 | 2.247 | 3.443 | 59.614 | 60.360 | 45.958 | 21.414 | -1.110 | -0.469 | -0.144 | -0.173 | 0.095 | -0.116 |
| 14 | -2.352 | 1.425 | 1.776 | 2.583 | 4.649 | 7.468 | 7.370 | 55.175 | 45.217 | 58.128 | 32.410 | -1.677 | -1.243 | -1.132 | -0.490 | 0.127 | -0.339 |
| 13 | -3.934 | -0.383 | 0.963 | 3.722 | 4.673 | 7.945 | 10.543 | 35.450 | 83.759 | 25.693 | 26.498 | -2.475 | -1.211 | 0.328 | -0.527 | 0.006 | -0.497 |
| 12 | -4.516 | 0.183 | 1.168 | 4.041 | 7.488 | 11.607 | 17.723 | 24.354 | 63.085 | -1.438 | 52.631 | -3.128 | -1.698 | -2.574 | -1.188 | 0.504 | -0.522 |
| 11 | -1.618 | 0.186 | -0.054 | 1.196 | 2.787 | 8.131 | 12.668 | 41.058 | 76.323 | 32.737 | 16.954 | -0.376 | -3.265 | 0.575 | 0.160 | 0.121 | 0.091 |
| 10 | 4.031 | 2.401 | 3.598 | 6.297 | 7.255 | 10.582 | 5.686 | -29.344 | 43.293 | -6.746 | 27.116 | -0.452 | 0.066 | 0.433 | 0.205 | -0.453 | -0.688 |
| 9 | 1.515 | 2.374 | 3.740 | 4.440 | 8.103 | 12.772 | 12.448 | 14.424 | 45.284 | 16.353 | 58.690 | 5.942 | 2.550 | -0.851 | 0.088 | -2.133 | -1.190 |
| 8 | 0.693 | 2.635 | 1.586 | 1.492 | 7.467 | 3.035 | 15.279 | -23.029 | 65.835 | -40.806 | 42.841 | 10.222 | 3.888 | 3.987 | 0.170 | -2.146 | -2.589 |
| 7 | 0.778 | 2.096 | 1.154 | 2.103 | 6.557 | 3.309 | 9.712 | -16.883 | 51.660 | -25.789 | 46.193 | 3.134 | 2.661 | 6.421 | 0.591 | -1.317 | -4.095 |
| 6 | 0.866 | -1.625 | 1.156 | 0.517 | 6.418 | 2.126 | 6.597 | -33.233 | 38.380 | 28.862 | 29.817 | 0.475 | -0.008 | 1.491 | -0.498 | -2.102 | -0.315 |
| 5 | 3.104 | 1.721 | 2.600 | 0.827 | 3.370 | -7.618 | 11.408 | -33.038 | 49.387 | -1.760 | 56.628 | 7.129 | -2.601 | 2.540 | 0.537 | -3.642 | -3.424 |
| 4 | 1.023 | -4.176 | -2.734 | -0.457 | 10.280 | -10.089 | -7.889 | -37.395 | 44.649 | 2.696 | 65.498 | 15.834 | 1.335 | 7.550 | 0.307 | -3.733 | -0.545 |
| 3 | -3.857 | -0.795 | -1.407 | 1.506 | 7.086 | -1.291 | -8.114 | -50.828 | 65.311 | -18.844 | 39.592 | -3.819 | 2.623 | -2.522 | -0.387 | -4.126 | -0.715 |
| 2 | -5.860 | -1.289 | -4.252 | 0.215 | 1.370 | 2.451 | -0.682 | -36.576 | 32.338 | -20.239 | 10.954 | -1.307 | 0.409 | 1.522 | -4.939 | -0.090 | -1.412 |
| 1 | -5.761 | 5.488 | -4.735 | 12.372 | 7.468 | 15.919 | -35.058 | -32.376 | 47.344 | -8.836 | 26.688 | -12.632 | -33.351 | 23.934 | 5.317 | -3.435 | -3.390 |
| 0 | 1.458 | -2.533 | -2.940 | -1.930 | 6.293 | 9.087 | -10.644 | 14.931 | 60.572 | 10.519 | 41.828 | -0.500 | 0.248 | -0.014 | 2.446 | -0.614 | 0.794 |
| embeddings | -2.905 | 0.707 | 4.800 | 5.773 | 5.332 | -1.487 | 0.904 | -0.386 | 7.242 | 6.091 | 11.238 | 16.226 | -1.650 | 1.032 | -2.435 | 0.429 | 0.244 |

# TRIGGER EXTRACTION

| | ĠThe | Ġformation | Ġof | Ġa | Ġstable | Ġtetrameric | Ġ | [unused1] | - | [unused1] | Ġcomplex | , | Ġwhich | Ġrequired | Ġthe | Ġdimerization | Ġdomain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 22 | 1.531 | 3.358 | 4.568 | 6.362 | 2.490 | 5.186 | 9.769 | 9.920 | 9.712 | 10.513 | 7.224 | 3.558 | 0.717 | 0.392 | 0.453 | 0.432 | 0.357 |
| 21 | 0.166 | 2.214 | 3.321 | 8.529 | 1.309 | 6.305 | 31.516 | 34.723 | 30.925 | 39.330 | 12.245 | 0.316 | 0.337 | 0.307 | 0.248 | 0.253 | 0.290 |
| 20 | -0.600 | 1.099 | 1.568 | 5.060 | 0.626 | 1.849 | 25.754 | 39.092 | 50.523 | 69.007 | 7.805 | -3.590 | 0.050 | 0.077 | -0.109 | -0.139 | 0.020 |
| 19 | -1.008 | 0.586 | 0.588 | 1.735 | 0.354 | 1.537 | 19.917 | 36.754 | 46.165 | 77.417 | 7.517 | -7.077 | 0.158 | 0.214 | 0.109 | 0.164 | 0.437 |
| 18 | -0.524 | -0.441 | -1.075 | -1.297 | -0.280 | -0.317 | 9.072 | 58.563 | 27.092 | 75.014 | -0.713 | -3.248 | -0.170 | 0.086 | 0.310 | 0.155 | 0.159 |
| 17 | -0.422 | -0.426 | -0.451 | -0.341 | -0.667 | -1.011 | 2.003 | 74.995 | 8.566 | 65.026 | -0.847 | -2.921 | -0.146 | 0.017 | 0.161 | -0.123 | -0.090 |
| 16 | -0.732 | 2.142 | 1.444 | 0.787 | 0.078 | -2.099 | 4.496 | 73.162 | 29.060 | 59.009 | 2.967 | -1.683 | 0.104 | -0.451 | -0.003 | 0.040 | -0.058 |
| 15 | -1.702 | 1.460 | 2.146 | 2.287 | 2.549 | 2.247 | 3.443 | 59.614 | 60.360 | 45.958 | 21.414 | -1.110 | -0.469 | -0.144 | -0.173 | 0.095 | -0.116 |
| 14 | -2.352 | 1.425 | 1.776 | 2.583 | 4.649 | 7.468 | 7.370 | 55.175 | 45.217 | 58.128 | 32.410 | -1.677 | -1.243 | -1.132 | -0.490 | 0.127 | -0.339 |
| 13 | -3.934 | -0.383 | 0.963 | 3.722 | 4.673 | 7.945 | 10.543 | 35.450 | 83.759 | 25.693 | 26.498 | -2.475 | -1.211 | 0.328 | -0.527 | 0.006 | -0.497 |
| 12 | -4.516 | 0.183 | 1.168 | 4.041 | 7.488 | 11.607 | 17.723 | 24.354 | 63.085 | -1.438 | 52.631 | -3.128 | -1.698 | -2.574 | -1.188 | 0.504 | -0.522 |
| 11 | -1.618 | 0.186 | -0.054 | 1.196 | 2.787 | 8.131 | 12.668 | 41.058 | 76.323 | 32.737 | 16.954 | -0.376 | -3.265 | 0.575 | 0.160 | 0.121 | 0.091 |
| 10 | 4.031 | 2.401 | 3.598 | 6.297 | 7.255 | 10.582 | 5.686 | -29.344 | 43.293 | -6.746 | 27.116 | -0.452 | 0.066 | 0.433 | 0.205 | -0.453 | -0.688 |
| 9 | 1.515 | 2.374 | 3.740 | 4.440 | 8.103 | 12.772 | 12.448 | 14.424 | 45.284 | 16.353 | 58.690 | 5.942 | 2.550 | -0.851 | 0.088 | -2.133 | -1.190 |
| 8 | 0.693 | 2.635 | 1.586 | 1.492 | 7.467 | 3.035 | 15.279 | -23.029 | 65.835 | -40.806 | 42.841 | 10.222 | 3.888 | 3.987 | 0.170 | -2.146 | -2.589 |
| 7 | 0.778 | 2.096 | 1.154 | 2.103 | 6.557 | 3.309 | 9.712 | -16.883 | 51.660 | -25.789 | 46.193 | 3.134 | 2.661 | 6.421 | 0.591 | -1.317 | -4.095 |
| 6 | 0.866 | -1.625 | 1.156 | 0.517 | 6.418 | 2.126 | 6.597 | -33.233 | 38.380 | 28.862 | 29.817 | 0.475 | -0.008 | 1.491 | -0.498 | -2.102 | -0.315 |
| 5 | 3.104 | 1.721 | 2.600 | 0.827 | 3.370 | -7.618 | 11.408 | -33.038 | 49.387 | -1.760 | 56.628 | 7.129 | -2.601 | 2.540 | 0.537 | -3.642 | -3.424 |
| 4 | 1.023 | -4.176 | -2.734 | -0.457 | 10.280 | -10.089 | -7.889 | -37.395 | 44.649 | 2.696 | 65.498 | 15.834 | 1.335 | 7.550 | 0.307 | -3.733 | -0.545 |
| 3 | -3.857 | -0.795 | -1.407 | 1.506 | 7.086 | -1.291 | -8.114 | -50.828 | 65.311 | -18.844 | 39.592 | -3.819 | 2.623 | -2.522 | -0.387 | -4.126 | -0.715 |
| 2 | -5.860 | -1.289 | -4.252 | 0.215 | 1.370 | 2.451 | -0.682 | -36.576 | 32.338 | -20.239 | 10.954 | -1.307 | 0.409 | 1.522 | -4.939 | -0.090 | -1.412 |
| 1 | -5.761 | 5.488 | -4.735 | 12.372 | 7.468 | 15.919 | -35.058 | -32.376 | 47.344 | -8.836 | 26.688 | -12.632 | -33.351 | 23.934 | 5.317 | -3.435 | -3.390 |
| 0 | 1.458 | -2.533 | -2.940 | -1.930 | 6.293 | 9.087 | -10.644 | 14.931 | 60.572 | 10.519 | 41.828 | -0.500 | 0.248 | -0.014 | 2.446 | -0.614 | 0.794 |
| embeddings | -2.905 | 0.707 | 4.800 | 5.773 | 5.332 | -1.487 | 0.904 | -0.386 | 7.242 | 6.091 | 11.238 | 16.226 | -1.650 | 1.032 | -2.435 | 0.429 | 0.244 |

# TRIGGER EXTRACTION

| | ĠThe | Ġformation | Ġof | Ġa | Ġstable | Ġtetrameric | Ġ | [unused1] | - | [unused1] | Ġcomplex | , | Ġwhich | Ġrequired | Ġthe | Ġdimerization | Ġdomain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 22 | 1.531 | 3.358 | 4.568 | 6.362 | 2.490 | 5.186 | 9.769 | 9.920 | 9.712 | 10.513 | 7.224 | 3.558 | 0.717 | 0.392 | 0.453 | 0.432 | 0.357 |
| 21 | 0.166 | 2.214 | 3.321 | 8.529 | 1.309 | 6.305 | 31.516 | 34.723 | 30.925 | 39.330 | 12.245 | 0.316 | 0.337 | 0.307 | 0.248 | 0.253 | 0.290 |
| 20 | -0.600 | 1.099 | 1.568 | 5.060 | 0.626 | 1.849 | 25.754 | 39.092 | 50.523 | 69.007 | 7.805 | -3.590 | 0.050 | 0.077 | -0.109 | -0.139 | 0.020 |
| 19 | -1.008 | 0.586 | 0.588 | 1.735 | 0.354 | 1.537 | 19.917 | 36.754 | 46.165 | 77.417 | 7.517 | -7.077 | 0.158 | 0.214 | 0.109 | 0.164 | 0.437 |
| 18 | -0.524 | -0.441 | -1.075 | -1.297 | -0.280 | -0.317 | 9.072 | 58.563 | 27.092 | 75.014 | -0.713 | -3.248 | -0.170 | 0.086 | 0.310 | 0.155 | 0.159 |
| 17 | -0.422 | -0.426 | -0.451 | -0.341 | -0.667 | -1.011 | 2.003 | 74.995 | 8.566 | 65.026 | -0.847 | -2.921 | -0.146 | 0.017 | 0.161 | -0.123 | -0.090 |
| 16 | -0.732 | 2.142 | 1.444 | 0.787 | 0.078 | -2.099 | 4.496 | 73.162 | 29.060 | 59.009 | 2.967 | -1.683 | 0.104 | -0.451 | -0.003 | 0.040 | -0.058 |
| 15 | -1.702 | 1.460 | 2.146 | 2.287 | 2.549 | 2.247 | 3.443 | 59.614 | 60.360 | 45.958 | 21.414 | -1.110 | -0.469 | -0.144 | -0.173 | 0.095 | -0.116 |
| 14 | -2.352 | 1.425 | 1.776 | 2.583 | 4.649 | 7.468 | 7.370 | 55.175 | 45.217 | 58.128 | 32.410 | -1.677 | -1.243 | -1.132 | -0.490 | 0.127 | -0.339 |
| 13 | -3.934 | -0.383 | 0.963 | 3.722 | 4.673 | 7.945 | 10.543 | 35.450 | 83.759 | 25.693 | 26.498 | -2.475 | -1.211 | 0.328 | -0.527 | 0.006 | -0.497 |
| 12 | -4.516 | 0.183 | 1.168 | 4.041 | 7.488 | 11.607 | 17.723 | 24.354 | 63.085 | -1.438 | 52.631 | -3.128 | -1.698 | -2.574 | -1.188 | 0.504 | -0.522 |
| 11 | -1.618 | 0.186 | -0.054 | 1.196 | 2.787 | 8.131 | 12.668 | 41.058 | 76.323 | 32.737 | 16.954 | -0.376 | -3.265 | 0.575 | 0.160 | 0.121 | 0.091 |
| 10 | 4.031 | 2.401 | 3.598 | 6.297 | 7.255 | 10.582 | 5.686 | -29.344 | 43.293 | -6.746 | 27.116 | -0.452 | 0.066 | 0.433 | 0.205 | -0.453 | -0.688 |
| 9 | 1.515 | 2.374 | 3.740 | 4.440 | 8.103 | 12.772 | 12.448 | 14.424 | 45.284 | 16.353 | 58.690 | 5.942 | 2.550 | -0.851 | 0.088 | -2.133 | -1.190 |
| 8 | 0.693 | 2.635 | 1.586 | 1.492 | 7.467 | 3.035 | 15.279 | -23.029 | 65.835 | -40.806 | 42.841 | 10.222 | 3.888 | 3.987 | 0.170 | -2.146 | -2.589 |
| 7 | 0.778 | 2.096 | 1.154 | 2.103 | 6.557 | 3.309 | 9.712 | -16.883 | 51.660 | -25.789 | 46.193 | 3.134 | 2.661 | 6.421 | 0.591 | -1.317 | -4.095 |
| 6 | 0.866 | -1.625 | 1.156 | 0.517 | 6.418 | 2.126 | 6.597 | -33.233 | 38.380 | 28.862 | 29.817 | 0.475 | -0.008 | 1.491 | -0.498 | -2.102 | -0.315 |
| 5 | 3.104 | 1.721 | 2.600 | 0.827 | 3.370 | -7.618 | 11.408 | -33.038 | 49.387 | -1.760 | 56.628 | 7.129 | -2.601 | 2.540 | 0.537 | -3.642 | -3.424 |
| 4 | 1.023 | -4.176 | -2.734 | -0.457 | 10.280 | -10.089 | -7.889 | -37.395 | 44.649 | 2.696 | 65.498 | 15.834 | 1.335 | 7.550 | 0.307 | -3.733 | -0.545 |
| 3 | -3.857 | -0.795 | -1.407 | 1.506 | 7.086 | -1.291 | -8.114 | -50.828 | 65.311 | -18.844 | 39.592 | -3.819 | 2.623 | -2.522 | -0.387 | -4.126 | -0.715 |
| 2 | -5.860 | -1.289 | -4.252 | 0.215 | 1.370 | 2.451 | -0.682 | -36.576 | 32.338 | -20.239 | 10.954 | -1.307 | 0.409 | 1.522 | -4.939 | -0.090 | -1.412 |
| 1 | -5.761 | 5.488 | -4.735 | 12.372 | 7.468 | 15.919 | -35.058 | -32.376 | 47.344 | -8.836 | 26.688 | -12.632 | -33.351 | 23.934 | 5.317 | -3.435 | -3.390 |
| 0 | 1.458 | -2.533 | -2.940 | -1.930 | 6.293 | 9.087 | -10.644 | 14.931 | 60.572 | 10.519 | 41.828 | -0.500 | 0.248 | -0.014 | 2.446 | -0.614 | 0.794 |
| embeddings | -2.905 | 0.707 | 4.800 | 5.773 | 5.332 | -1.487 | 0.904 | -0.386 | 7.242 | 6.091 | 11.238 | 16.226 | -1.650 | 1.032 | -2.435 | 0.429 | 0.244 |

# MODEL

**RegulaTome corpus:**

- **10499** regulatory relationships
    - **2294** Regulation
    - **2131** Positive Regulation
    - **1920** Negative Regulation
    - **521** Regulation of Gene Expression
    - … **40** different types

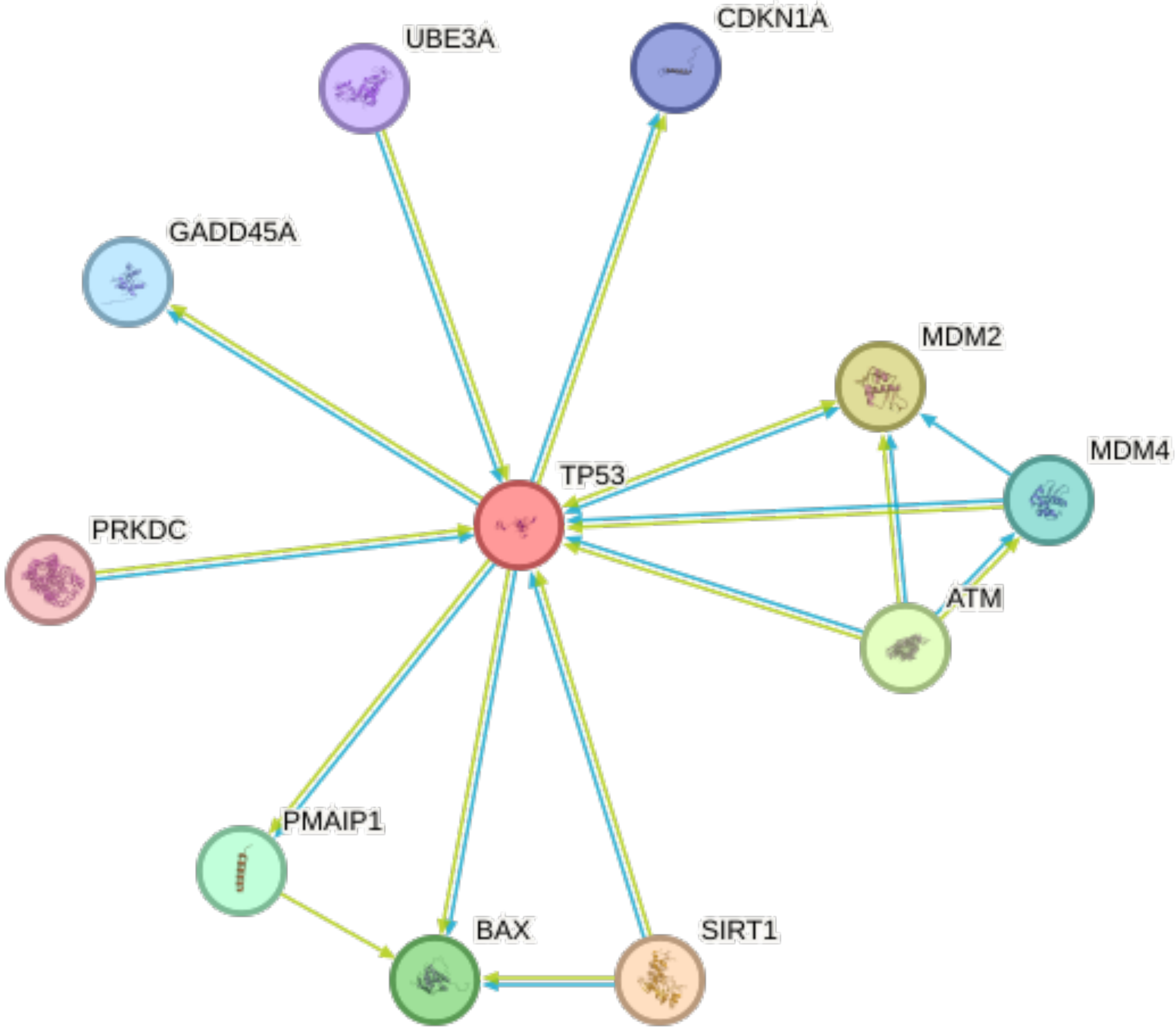
# MODEL

Model run:

- Documents with at least 1 pair
    - 3 mln abstracts
    - 3 mln full-text
- 1200 mln pairs
    - 18.4 mln with at least one positive label
    - < 100k high confidence labels

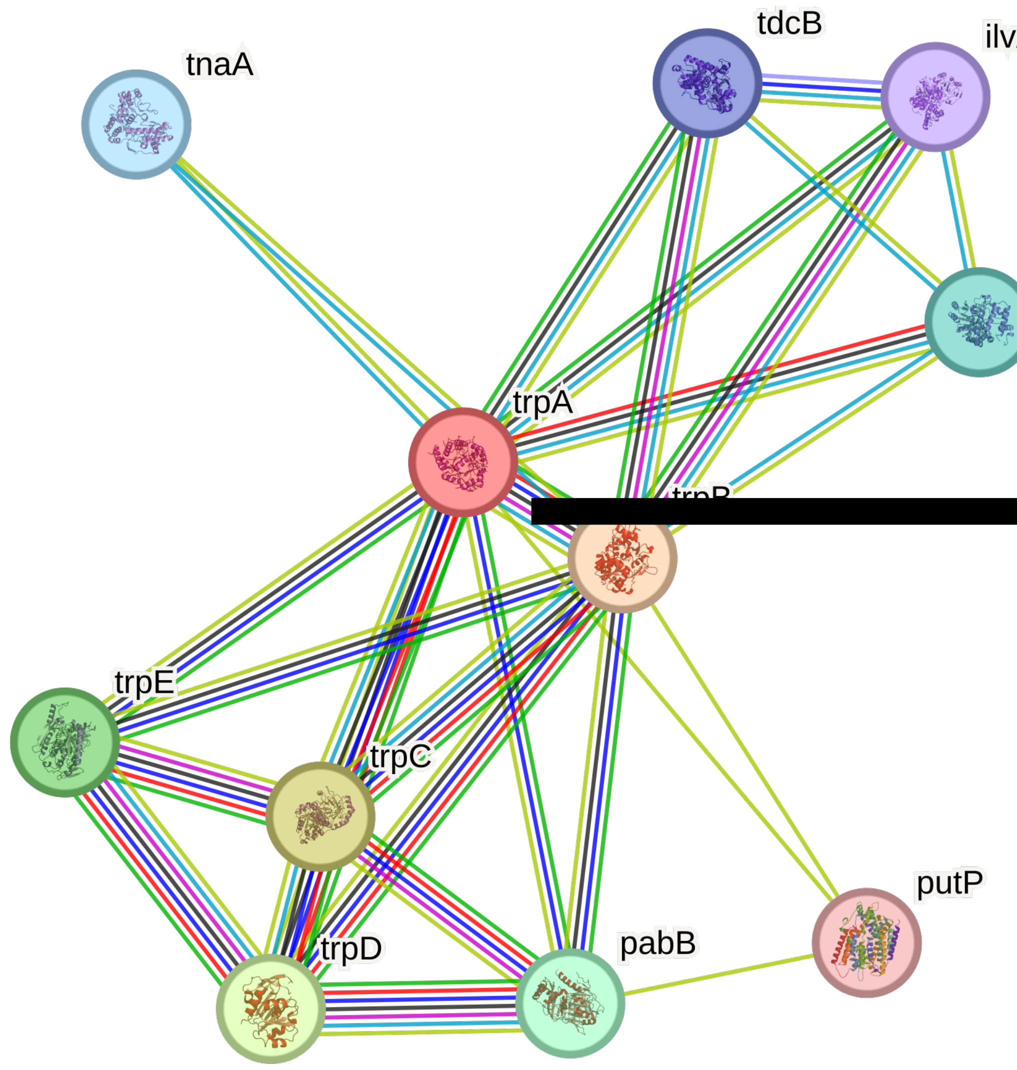
# REGULATION

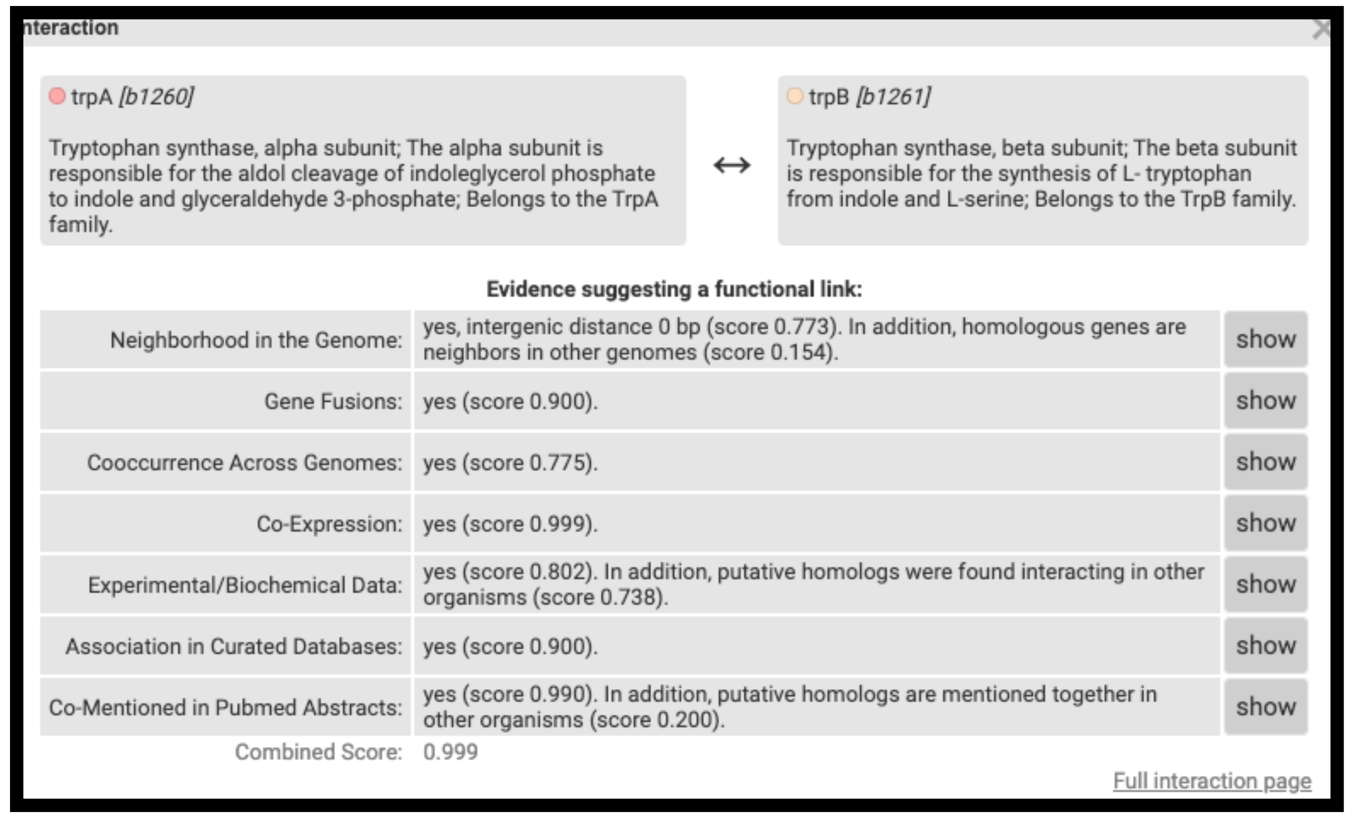
## Available
## Later This Year

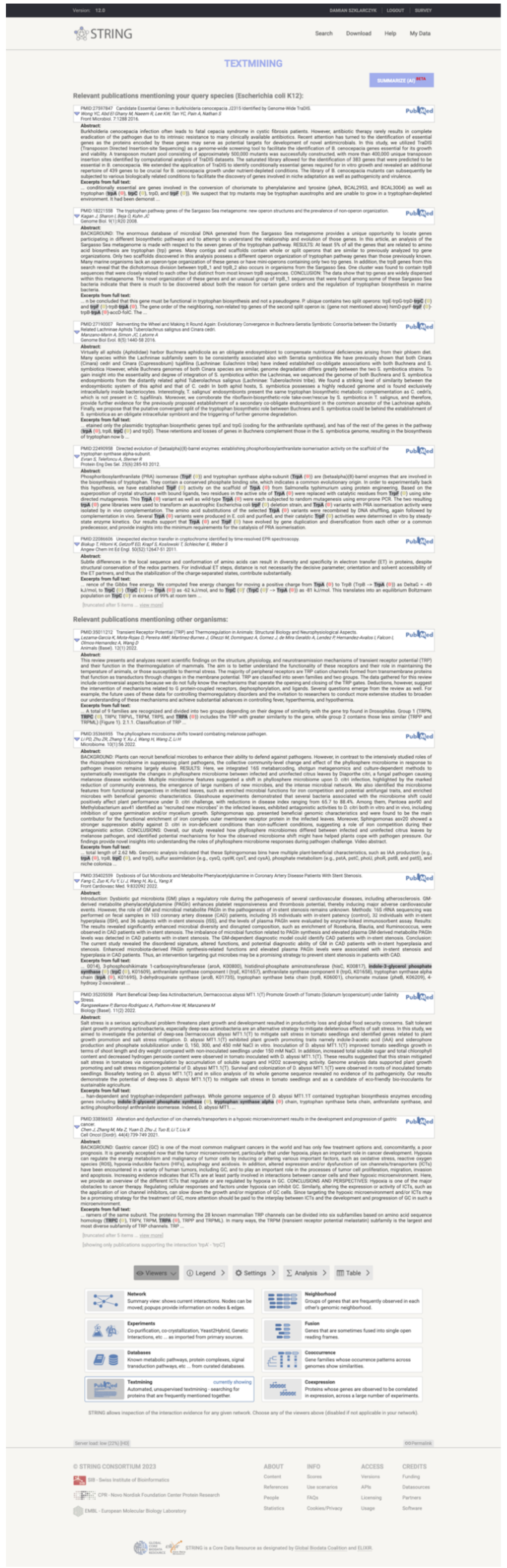


# STRING

## Textmining viewer



# STRING

## Textmining viewer

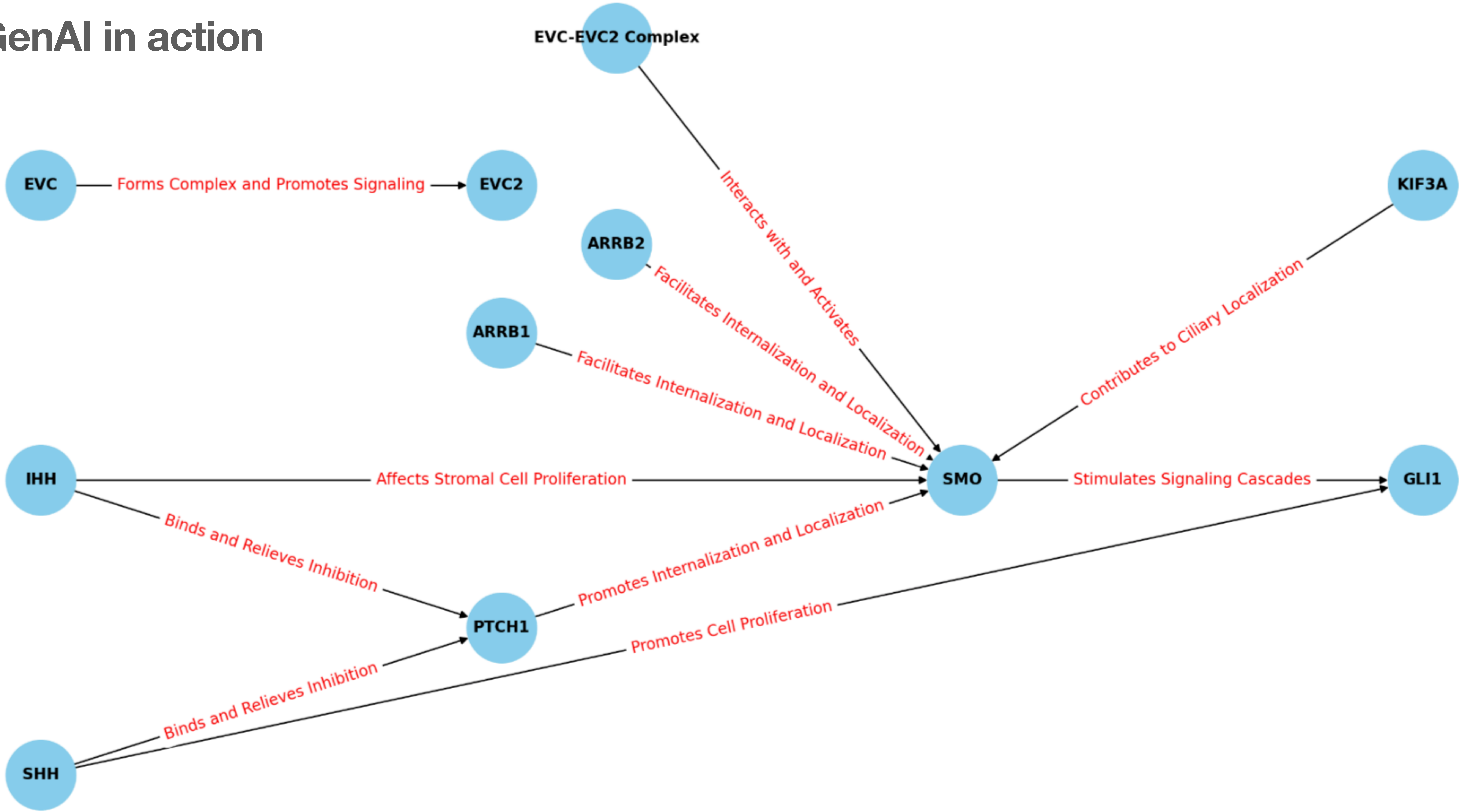
PMID:35022438 **Sonic Hedgehog** () acts as a macrophage chemoattractant during regeneration of the gastric epithelium.
*Chakrabarti J, Dua-Awereh M, Schumacher M, Engevik A, Hawkins J, Helmrath MA, Zavros Y*
NPJ Regen Med. 7(1):3 2022.

PubMed

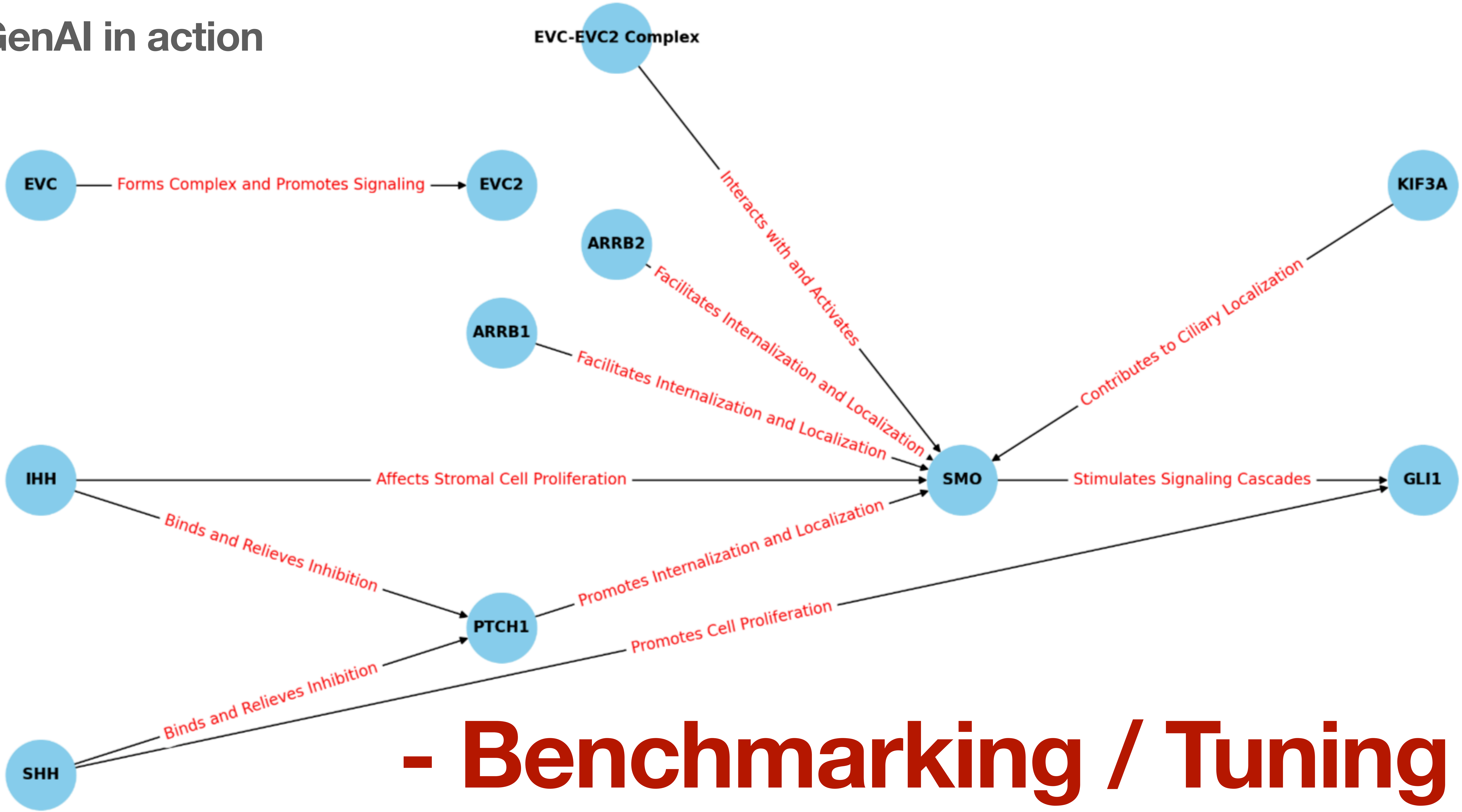
**Abstract:**
**Sonic Hedgehog** () (**Shh** ()), secreted from gastric parietal cells, contributes to the regeneration of the epithelium. The recruitment of macrophages plays a central role in the regenerative process. The mechanism that regulates macrophage recruitment in response to gastric injury is largely unknown. Here we tested the hypothesis that **Shh** () stimulates macrophage chemotaxis to the injured epithelium and contributes to gastric regeneration. A mouse model expressing a myeloid cell-specific deletion of **Smoothened** () (LysMcre/+;Smof/f) was generated using transgenic mice bearing loxP sites flanking the **Smo** () gene (**Smo** () loxP) and mice expressing a Cre recombinase transgene from the Lysozyme M locus (LysMCre). Acetic acid injury was induced in the stomachs of both control and LysMcre/+;Smof/f (SmoKO) mice and gastric epithelial regeneration and macrophage recruitment analyzed over a period of 7 days post-injury. Bone marrow-derived macrophages (BM-Moe) were collected from control and SmoKO mice. Human-derived gastric organoid/macrophage co-cultures were established, and macrophage chemotaxis measured. Compared to control mice, SmoKO animals exhibited inhibition of ulcer repair and normal epithelial regeneration, which correlated with decreased macrophage infiltration at the site of injury. Bone marrow chimera experiments using SmoKO donor cells showed that control chimera mice transplanted with SmoKO bone marrow donor cells exhibited a loss of ulcer repair, and transplantation of control bone marrow donor cells to SmoKO mice rescued epithelial cell regeneration. Histamine-stimulated **Shh** () secretion in human organoid/macrophage co-cultures resulted in macrophage migration toward the gastric epithelium, a response that was blocked with **Smo** () inhibitor Vismodegib. **Shh** ()-induced macrophage migration was mediated by AKT signaling. In conclusion, **Shh** () signaling acts as a macrophage chemoattractant via a **Smo** ()-dependent mechanism during gastric epithelial regeneration in response to injury.

**Excerpts from full text:**
... ite chamber (Fig. 9A). Collection of conditioned media from the chamber with organoid/ILC2 co-cultures revealed that histamine-induced significant secretion of **Shh** () that was blocked by **Smo** () inhibitor vismodegib (Fig. 9B). Similarly, IL-33 (Fig. 9C) and IL-13 Fig. 9D) secretion was induced by histamine, a response that was also blocked by vismod ...

# GenAI in action



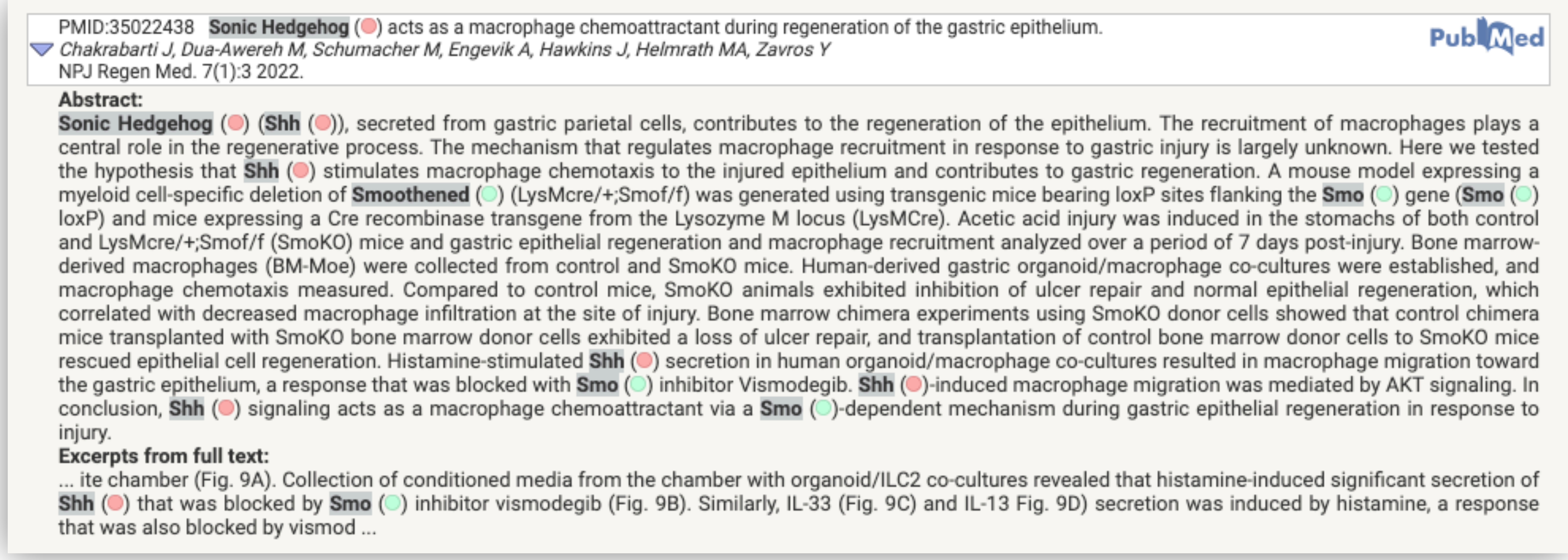
# GenAI in action




# - Benchmarking / Tuning

# STRING

## Textmining viewer

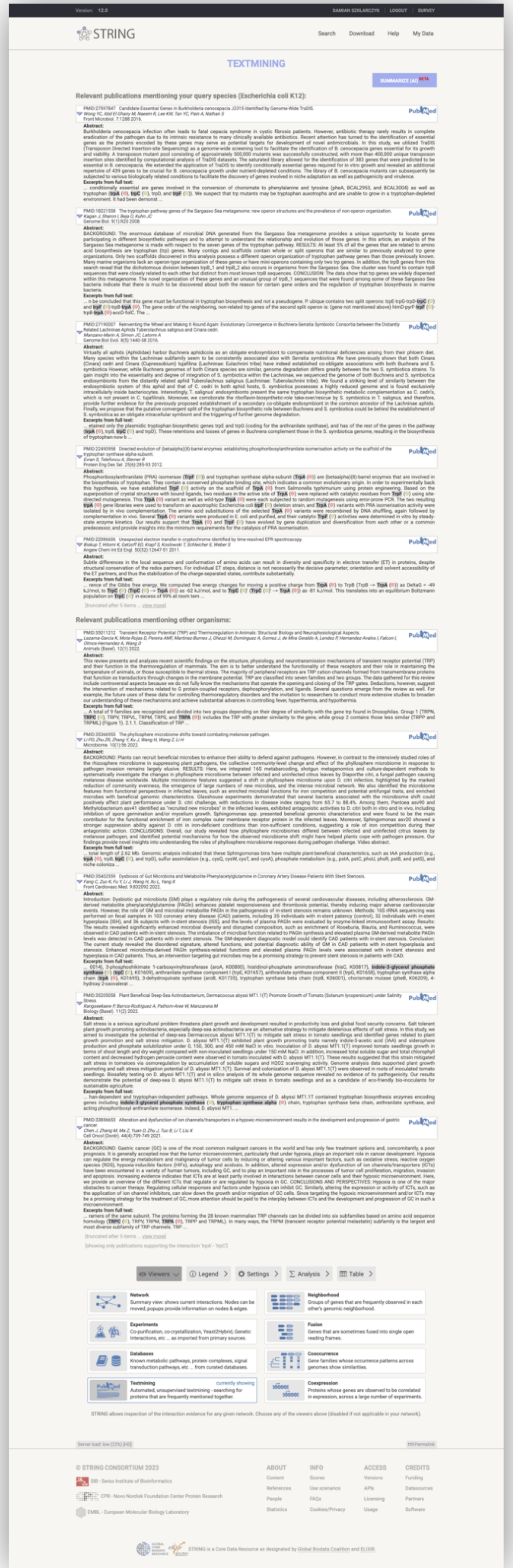


PMID:35022438 **Sonic Hedgehog** ( ) acts as a macrophage chemoattractant during regeneration of the gastric epithelium.
*Chakrabarti J, Dua-Awereh M, Schumacher M, Engevik A, Hawkins J, Helmrath MA, Zavros Y*
NPJ Regen Med. 7(1):3 2022.

PubMed

**Abstract:**
**Sonic Hedgehog** ( ) (**Shh** ( )), secreted from gastric parietal cells, contributes to the regeneration of the epithelium. The recruitment of macrophages plays a central role in the regenerative process. The mechanism that regulates macrophage recruitment in response to gastric injury is largely unknown. Here we tested the hypothesis that **Shh** ( ) stimulates macrophage chemotaxis to the injured epithelium and contributes to gastric regeneration. A mouse model expressing a myeloid cell-specific deletion of **Smoothened** ( ) (LysMcre/+;Smof/f) was generated using transgenic mice bearing loxP sites flanking the **Smo** ( ) gene (**Smo** ( ) loxP) and mice expressing a Cre recombinase transgene from the Lysozyme M locus (LysMCre). Acetic acid injury was induced in the stomachs of both control and LysMcre/+;Smof/f (SmoKO) mice and gastric epithelial regeneration and macrophage recruitment analyzed over a period of 7 days post-injury. Bone marrow-derived macrophages (BM-Moe) were collected from control and SmoKO mice. Human-derived gastric organoid/macrophage co-cultures were established, and macrophage chemotaxis measured. Compared to control mice, SmoKO animals exhibited inhibition of ulcer repair and normal epithelial regeneration, which correlated with decreased macrophage infiltration at the site of injury. Bone marrow chimera experiments using SmoKO donor cells showed that control chimera mice transplanted with SmoKO bone marrow donor cells exhibited a loss of ulcer repair, and transplantation of control bone marrow donor cells to SmoKO mice rescued epithelial cell regeneration. Histamine-stimulated **Shh** ( ) secretion in human organoid/macrophage co-cultures resulted in macrophage migration toward the gastric epithelium, a response that was blocked with **Smo** ( ) inhibitor Vismodegib. **Shh** ( )-induced macrophage migration was mediated by AKT signaling. In conclusion, **Shh** ( ) signaling acts as a macrophage chemoattractant via a **Smo** ( )-dependent mechanism during gastric epithelial regeneration in response to injury.

**Excerpts from full text:**
... ite chamber (Fig. 9A). Collection of conditioned media from the chamber with organoid/ILC2 co-cultures revealed that histamine-induced significant secretion of **Shh** ( ) that was blocked by **Smo** ( ) inhibitor vismodegib (Fig. 9B). Similarly, IL-33 (Fig. 9C) and IL-13 Fig. 9D) secretion was induced by histamine, a response that was also blocked by vismod ...



normalized names

| | |
|---|---|
| **PMID:33078454** | ... re tested at 2x and 8x MIC of each compound. The enzymes involved in de novo tryptophan biosynthesis are encoded on the trp operon (5'-3'): trpE, Rv1610, trpC, trpB, trpA. To confirm that compounds 1, 2 and 3, target biological components of the tryptophan biosynthesis, and more specifically tryptophan synthase, the MIC of M ... |
| **PMID:31311884** | ... , representing 4% of the genome, five categories were equally represented with 3 genes each: energy metabolism (glgA, lpdA, and glmS), amino acid biosynthesis (trpA, trpB, and aroL), DNA replication and repair (nrdA, recA, and dnaQ), type III secretion (mcsC, CTL0085, and CTL0043), and translation (pheT, cysS, and thrS). Of ... |
| **PMID:22086606** | ... T to generate the RP state FAD*…TrpB *via FAD*...TrpA could basically still occur. However, without the terminal TrpC, the photo-generated radicals on FAD and TrpA/TrpB are not sufficiently spatially separated for stable charge separation on a longer time scale. Thus, efficient backward ET and radical recombination would l ... |

openAI chat (gpt-3.5-turbo-1106)

**PMID:33078454**

... re tested at 2x and 8x MIC of each compound. The enzymes involved in de novo tryptophan biosynthesis are encoded on the trp operon (5'-3'): trpE, Rv1610, trpC, trpB, trpA. To confirm that compounds 1, 2 and 3, target biological components of the tryptophan biosynthesis, and more specifically tryptophan synthase, the MIC of M ...

**PMID:31311884**

... , representing 4% of the genome, five categories were equally represented with 3 genes each: energy metabolism (glgA, lpdA, and glmS), amino acid biosynthesis (trpA, trpB, and aroL), DNA replication and repair (nrdA, recA, and dnaQ), type III secretion (mcsC, CTL0085, and CTL0043), and translation (pheT, cysS, and thrS). Of ...

**PMID:22086606**

... T to generate the RP state FAD*…TrpB *via FAD*...TrpA could basically still occur. However, without the terminal TrpC, the photo-generated radicals on FAD and TrpA/TrpB are not sufficiently spatially separated for stable charge separation on a longer time scale. Thus, efficient backward ET and radical recombination would l ...

openAI chat (gpt-3.5-turbo-1106)

trpA and trpB are subunits of the tryptophan synthase enzyme. Studies have reported inhibitors that compete with IGP for the active site of trpA, bind at the trpA:trpB interface, and interact with the hydrophobic intermolecular indole channel in trpB. Mutations in trpA have been identified in various Chlamydia trachomatis strains, suggesting its involvement in organotropism. The trpA:trpB communication domain and nonsynonymous mutations have been analyzed in clinical strains. Additionally, trpA and trpB are components of the tryptophan biosynthesis pathway and are part of the trp operon. Furthermore, there is an ET step between trpB and trpA in the tryptophan synthase enzyme.

GPT:

- Negligible cost for non programmatic access with user trigger
- We do not want GPT to inject knowledge
- Small models perform well at this task
- Don’t overwhelm the model with text
- Frustrating to read, as there is no source

trpA and trpB are subunits of the tryptophan synthase enzyme. Studies have reported inhibitors that compete with IGP for the active site of trpA, bind at the trpA:trpB interface, and interact with the hydrophobic intermolecular indole channel in trpB. Mutations in trpA have been identified in various Chlamydia trachomatis strains, suggesting its involvement in organotropism. The trpA:trpB communication domain and nonsynonymous mutations have been analyzed in clinical strains. Additionally, trpA and trpB are components of the tryptophan biosynthesis pathway and are part of the trp operon. Furthermore, there is an ET step between trpB and trpA in the tryptophan synthase enzyme.

trpA and trpB are subunits of the tryptophan synthase enzyme. Studies have reported inhibitors that compete with IGP for the active site of trpA, bind at the trpA:trpB interface, and interact with the hydrophobic intermolecular indole channel in trpB. Mutations in trpA have been identified in various Chlamydia trachomatis strains, suggesting its involvement in organotropism. The trpA:trpB communication domain and nonsynonymous mutations have been analyzed in clinical strains. Additionally, trpA and trpB are components of the tryptophan biosynthesis pathway and are part of the trp operon. Furthermore, there is an ET step between trpB and trpA in the tryptophan synthase enzyme.

Split into clauses

trpA and trpB are subunits of the tryptophan synthase enzyme.

Additionally, trpA and trpB are components of the tryptophan biosynthesis pathway and are part of the trp operon

Studies have reported inhibitors that compete with IGP for the active site of trpA

bind at the trpA:trpB interface, and interact with the hydrophobic intermolecular indole channel in trpB

Mutations in trpA have been identified in various Chlamydia trachomatis strains

suggesting its involvement in organotropism

The trpA:trpB communication domain and nonsynonymous mutations have been analyzed in clinical strains

Furthermore, there is an ET step between trpB and trpA in the tryptophan synthase enzyme.

embedding

**[-0.7, -2.8, 0.0, 9.9]**

**[6.6, -4.1, -6.0, 0.1]**

**[-0.2, 2.0, 0.1, 0.3]**

**[-0.2, 3.8, 2.0, -2.1]**

**[6.6, -4.1, -6.0, 0.1]**

**[-0.2, 2.0, 0.1, 0.3]**

**[-0.7, -2.8, 0.0, 9.9]**

PMID:33078454 The enzymes involved in de novo tryptophan biosynthesis are encoded on the trp operon (5'-3'): trpE, Rv1610, trpC, trpB, trpA

PMID:31311884 five categories were equally represented with 3 genes each: energy metabolism (glgA, lpdA, and glmS), amino acid biosynthesis (trpA, trpB, and aroL)

PMID:22086606 the photo-generated radicals on FAD and TrpA/TrpB are not sufficiently spatially separated for stable charge separation on a longer time scale.

embedding

[-0.2, 3.8, 2.0, -2.1]

[6.6, -4.1, -6.0, 0.1]

[-0.2, 2.0, 0.1, 0.3]

embedding

[-0.7, -2.8, 0.0, 9.9]

[6.6, -4.1, -6.0, 0.1]

[-0.2, 2.0, 0.1, 0.3]

[-0.2, 3.8, 2.0, -2.1]

[6.6, -4.1, -6.0, 0.1]

[-0.2, 2.0, 0.1, 0.3]

[-0.7, -2.8, 0.0, 9.9]

similarity

PMID:33078454 The enzymes involved in de novo tryptophan biosynthesis are encoded on the trp operon (5'-3'): trpE, Rv1610, trpC, trpB, trpA

PMID:31311884 five categories were equally represented with 3 genes each: energy metabolism (glgA, lpdA, and glmS), amino acid biosynthesis (trpA, trpB, and aroL)

PMID:22086606 the photo-generated radicals on FAD and TrpA/TrpB are not sufficiently spatially separated for stable charge separation on a longer time scale.

embedding

[-0.2, 3.8, 2.0, -2.1]

[6.6, -4.1, -6.0, 0.1]

[-0.2, 2.0, 0.1, 0.3]

embedding

[-0.7, -2.8, 0.0, 9.9]

[6.6, -4.1, -6.0, 0.1]

[-0.2, 2.0, 0.1, 0.3]

[-0.2, 3.8, 2.0, -2.1]

[6.6, -4.1, -6.0, 0.1]

[-0.2, 2.0, 0.1, 0.3]

[-0.7, -2.8, 0.0, 9.9]

similarity

trpA and trpB are subunits of the tryptophan synthase enzyme **(1,2)**. Studies have reported inhibitors that compete with IGP for the active site of trpA **(2)**, bind at the trpA:trpB interface, and interact with the hydrophobic intermolecular indole channel in trpB **(2)**. Mutations in trpA have been identified in various Chlamydia trachomatis strains **(1)**, suggesting its involvement in organotropism. The trpA:trpB communication domain and nonsynonymous mutations have been analyzed in clinical strains **(1,3)**. Additionally, trpA and trpB are components of the tryptophan biosynthesis pathway and are part of the trp operon **(1)**. Furthermore, there is an ET step between trpB and trpA in the tryptophan synthase enzyme **(1)**.

STRING

Search Download Help My Data

# TEXTMINING

SUMMARIZE (AI) BETA

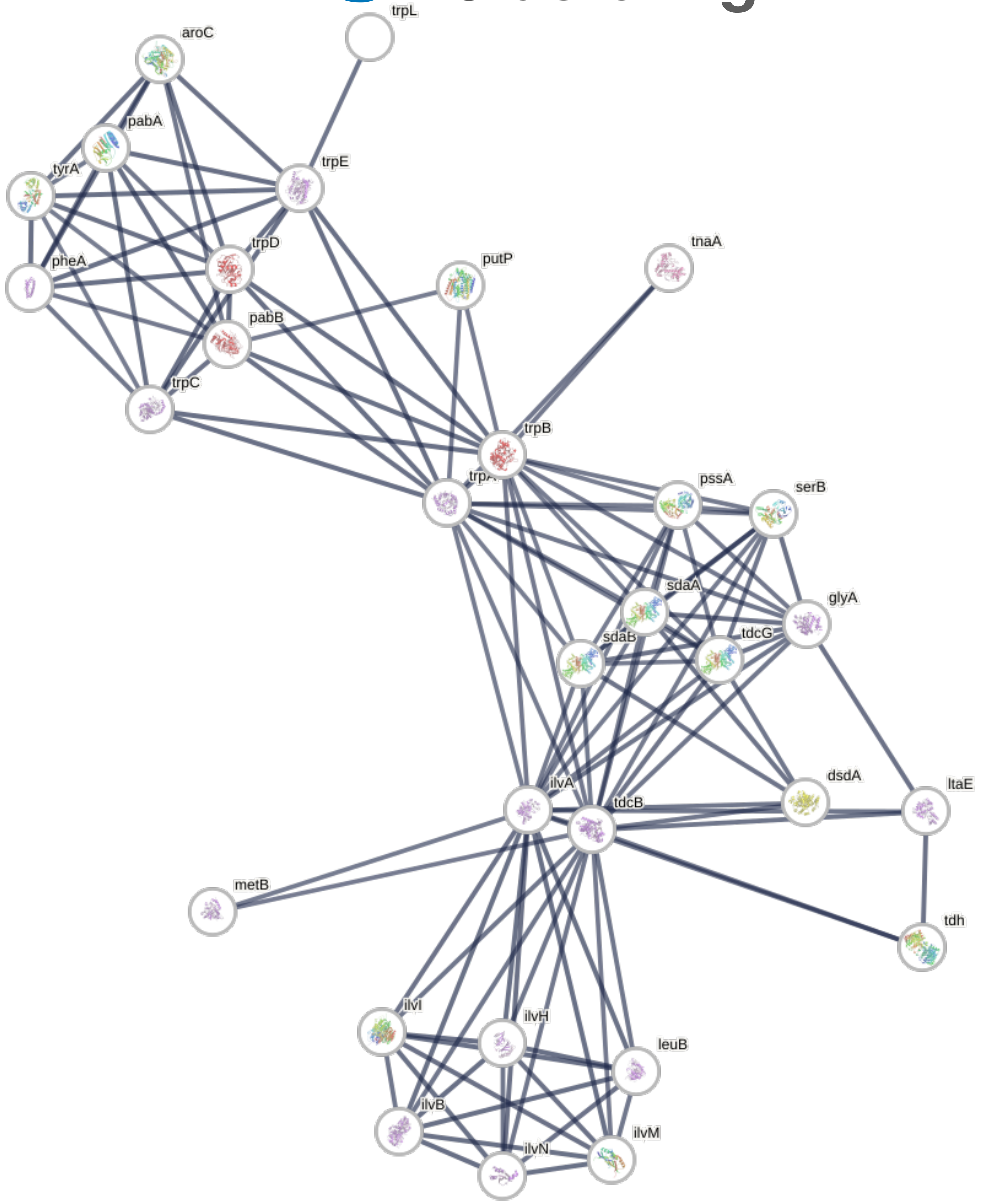
## Relevant publications mentioning your query species (Homo sapiens):

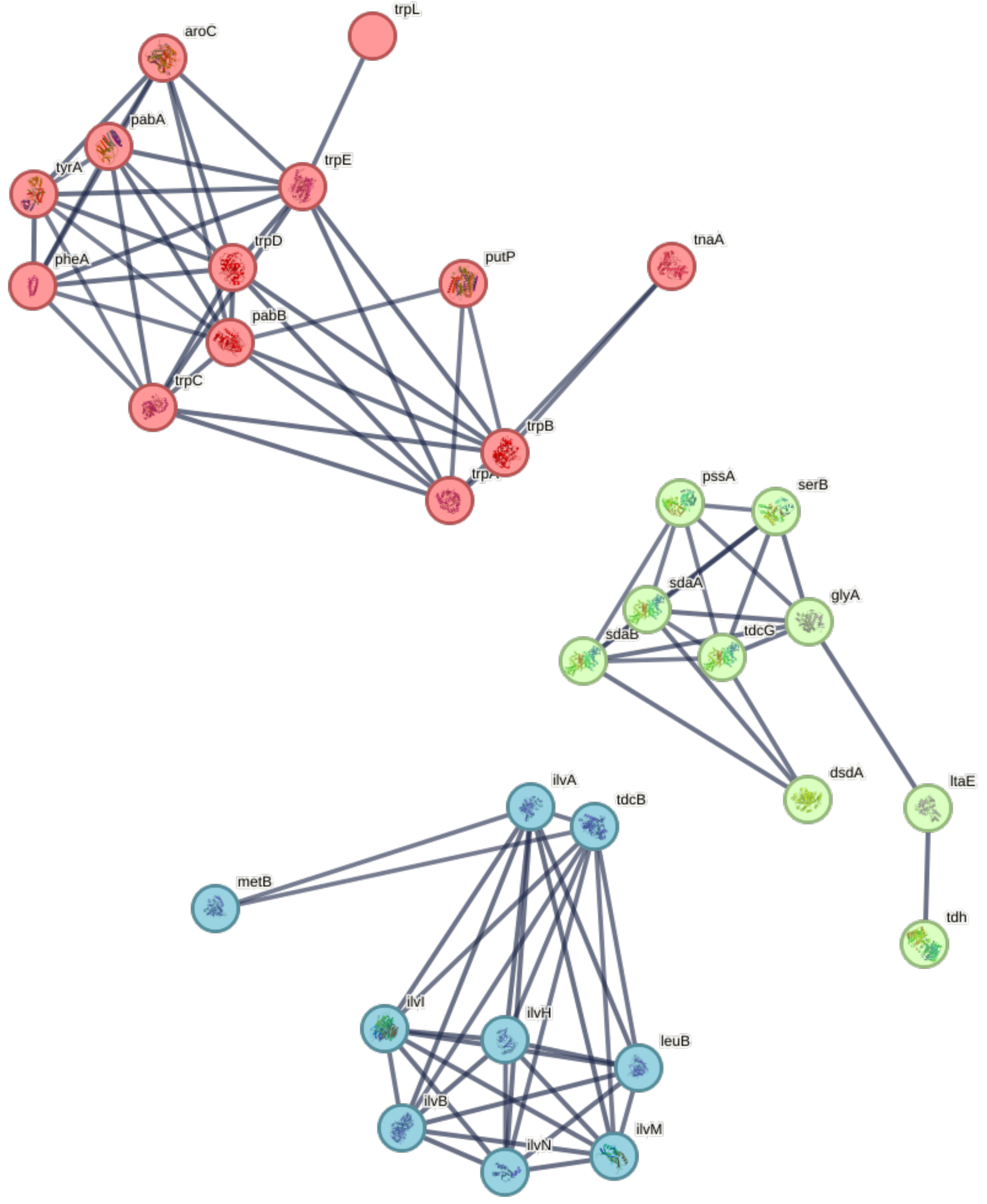
PMID:31756206 Expression profile of **sonic hedgehog** ( ) signaling-related molecules in basal cell carcinoma.
*Kim HS, Kim YS, Lee C, Shin MS, Kim JW, Jang BG*
PLoS One. 14(11):e0225511 2019.

PubMed

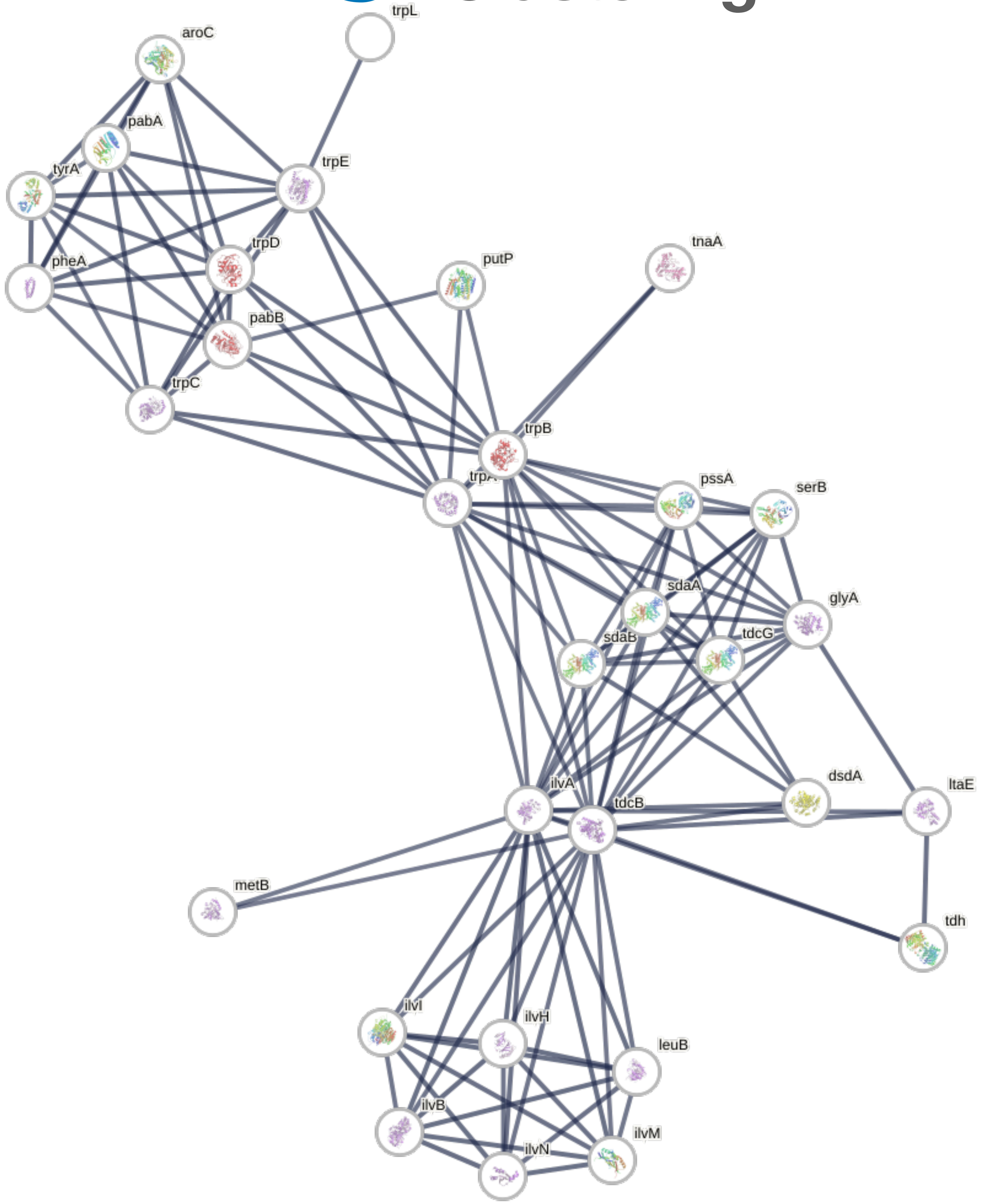
**Abstract:**
Basal cell carcinoma (BCC) is the most common human cancer, characterized by aberrant activation of the hedgehog (HH) signaling pathway resulting from mutations in the patched 1 (PTCH1) or smoothened (SMO) genes. In the present study, to uncover the expression profile of HH signaling-related molecules, we thoroughly examined the mRNA and protein expression levels of six molecules including GLI1, GLI2, PTCH1, **PTCH2** ( ), **SHH** ( ), and SMO in BCC and various other cutaneous tumors. Real-time PCR analysis demonstrated that BCC showed remarkably enhanced mRNA expression of all HH molecules, except SMO compared to other skin tumors. However, immunohistochemical analysis revealed that only GLI1 protein was specifically upregulated in BCC, while the other HH-related proteins did not show any significant differences between the tumors. Notably, other skin malignancies such as squamous cell carcinoma, sebaceous carcinoma, and malignant melanoma showed no GLI1 expression and there was no difference in GLI1 expression between the BCC subtypes. In addition, GLI1 and

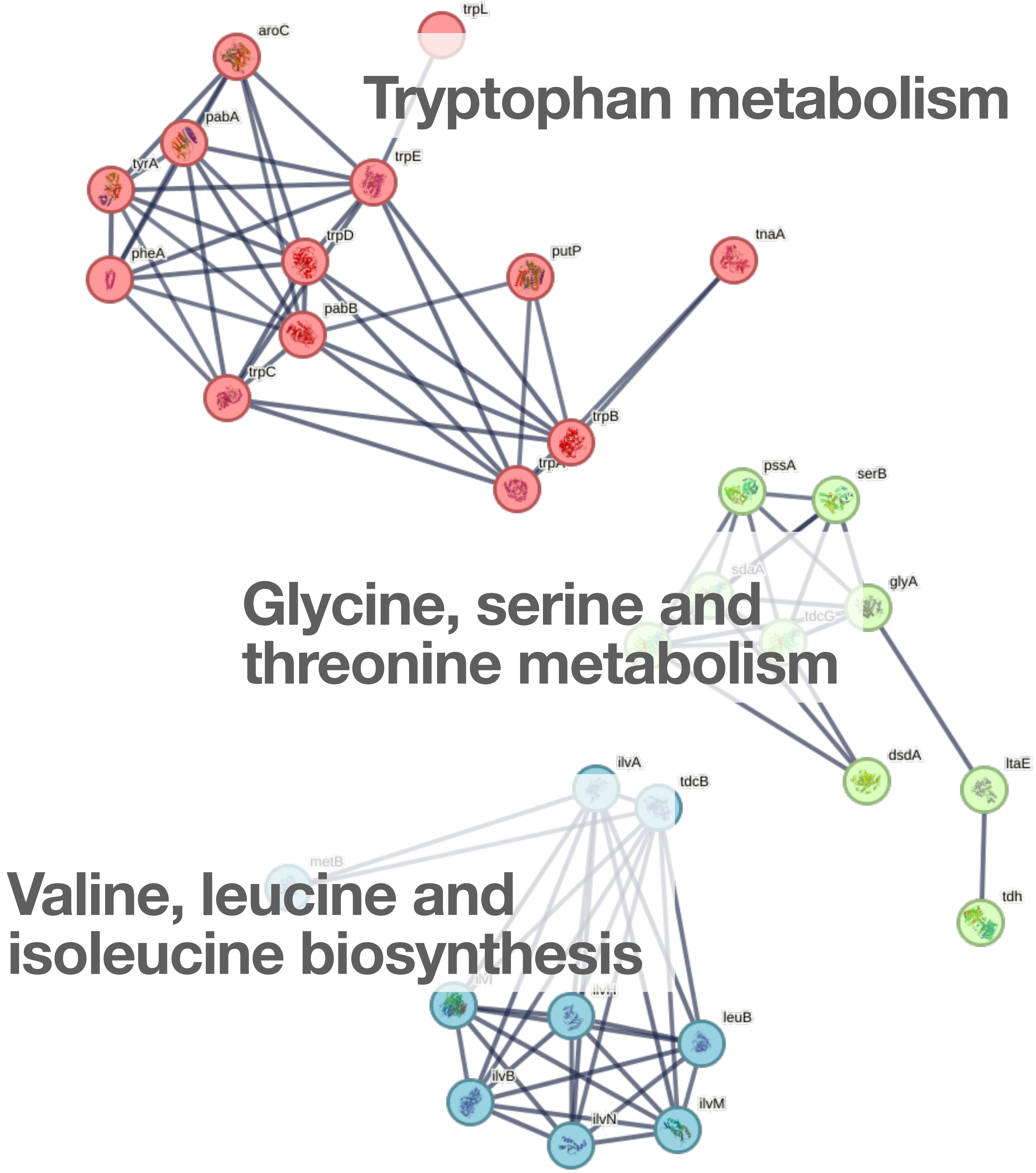
# STRING Clustering



# STRING Clustering



# MAIN ISSUES WITH GPT

## From scientific resource perspective in no particular order

- Reproducibility / Explainability

# MAIN ISSUES WITH GPT

## From scientific resource perspective in no particular order

- Reproducibility / Explainability
- Inferential Overreach

# MAIN ISSUES WITH GPT

## From scientific resource perspective in no particular order

- Reproducibility / Explainability
- Inferential Overreach
- Paper mills

# MAIN ISSUES WITH GPT

## From scientific resource perspective in no particular order

- Reproducibility / Explainability
- Inferential Overreach
- Paper mills
- Cuts in primary curation

# MAIN ISSUES WITH GPT

From scientific resource perspective in no particular order

- Reproducibility / Explainability
- Inferential Overreach
- Paper mills
- Cuts in primary curation
- Web crawlers

# ACKNOWLEDGMENTS


Swiss Institute of Bioinformatics

Universität Zürich[UZH]

Novo Nordisk Foundation Center for Protein Research

UNIVERSITY OF COPENHAGEN

UNIVERSITY OF TURKU

| UZH / SIB | KU / NNF CPR | Uni. Turku |
|---|---|---|
| Tao Fang | Katerina Nastou | Farrokh Mehryary |
| Radja Hachilif | Rebecca Kirch | Tomoko Ohta |
| Qingyao Huang | Dawei Hu | Sampo Pyysalo |
| Christian von Mering | Lars Juhl Jensen | |

# THANKS!