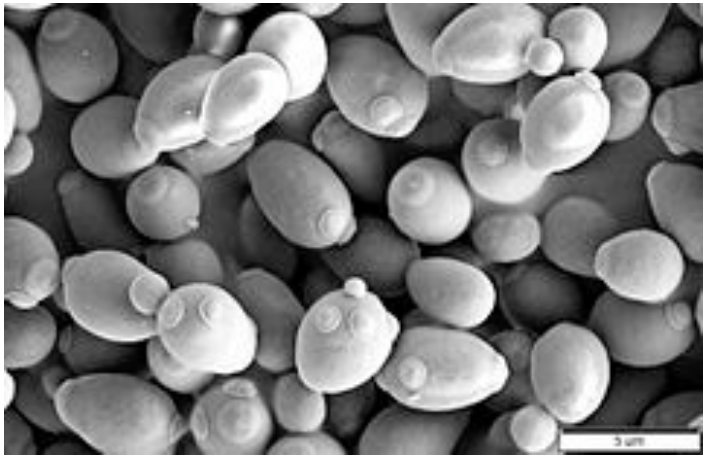
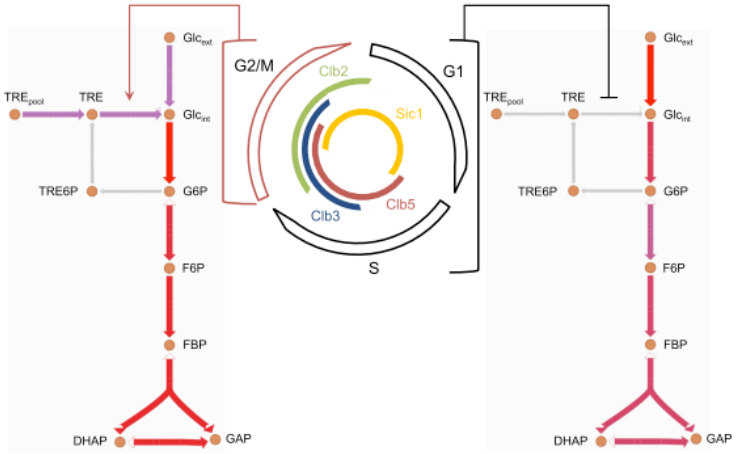




GEMMER

GEnome-wide software for Multi-scale Modeling data Extraction and Representation

(Mondeel, Crémazy and Barberis, in preparation)

Multi-scale modeling

- the term multi-scale is to be used when all of these aspects (functional, temporal and spatial separation) are present and key to understanding a single overarching biological process.
- "… multi-scale models is a composition of two or more single scale models representing the same phenomenon (or its parts) at different levels of descriptions" (Castiglione et al., 2014).

Why GEMMER

- For multi-scale modeling purposes it is useful to:
  - Find *proven* interactions between proteins
  - Know what the function of these proteins is
  - The function indicates some characteristic time-scale
  - Know what types of interactions these are: physical, genetic or regulation?
  - Know the expression levels and localization of the proteins
  - How many experiments, publications and methods have shown the interactions
  - Visualize these in a network
  - Export the data for later use

What GEMMER provides

- GEMMER aids (modeling) research by providing
  - Effortless data-integration from 3 databases
    - SGD, CYCLoPS and the GFP database
  - Interaction networks with localization and expression information incorporated
    - Centered around 1 or more proteins
  - Data visualization using interactive D3 (*not 3D!)* drawings
    - With various filtering possibilities
  - Data export for interaction networks to SVG & Excel

What GEMMER provides: visualization

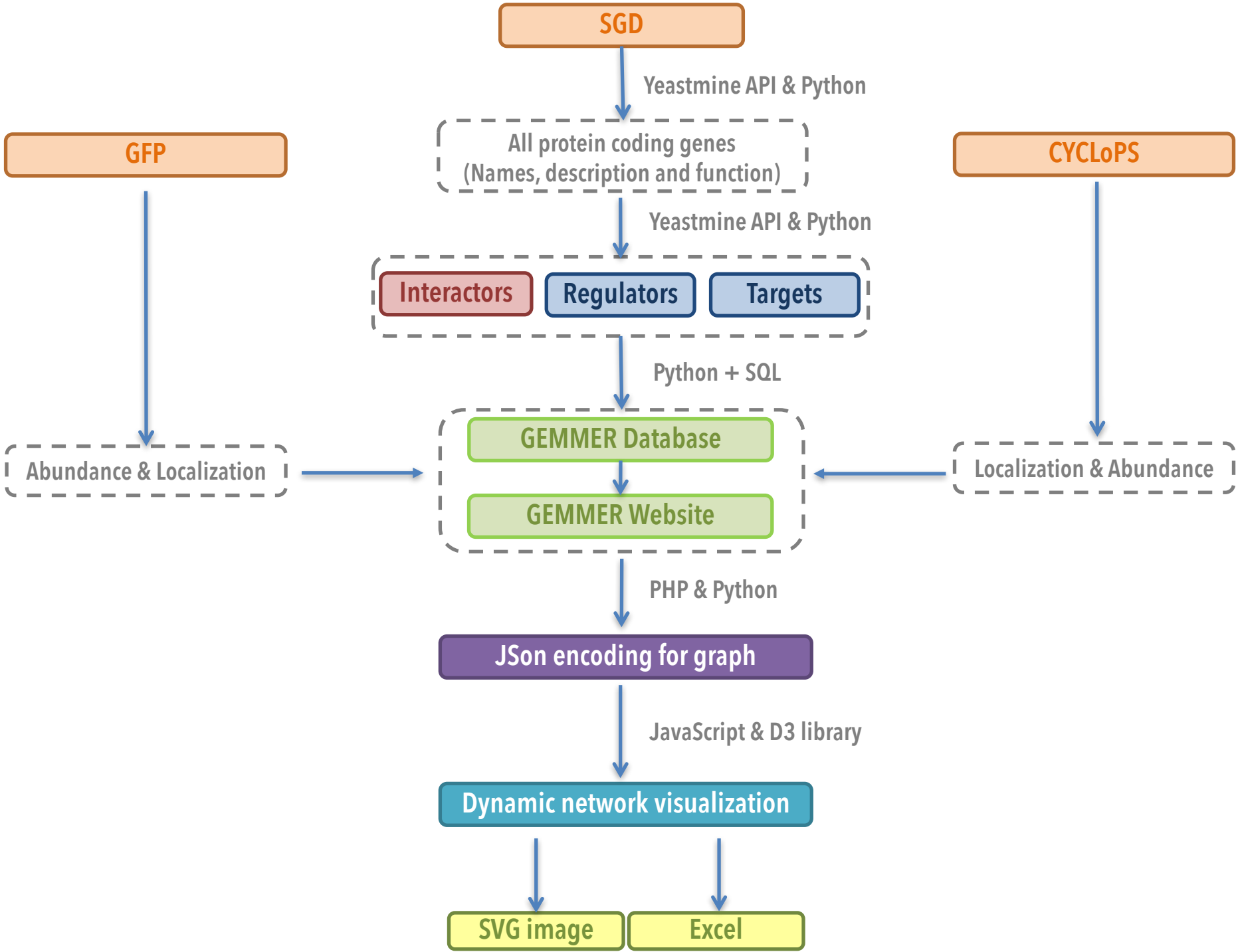
- Visualization options available in GEMMER
  - Interaction networks joining up to 3 or more genes
  - Clustering
    - Cluster the network on its dominant compartment (CYCLoPS expression) or function
  - Coloring
    - Color the nodes based on function through GO annotations or dominant compartments
  - Filtering
    - Decide what the lower bounds on the number of experiments, publications or experimental methods is
    - Select only genes that are expressed in specific compartments according to either CYCLoPS or GFP

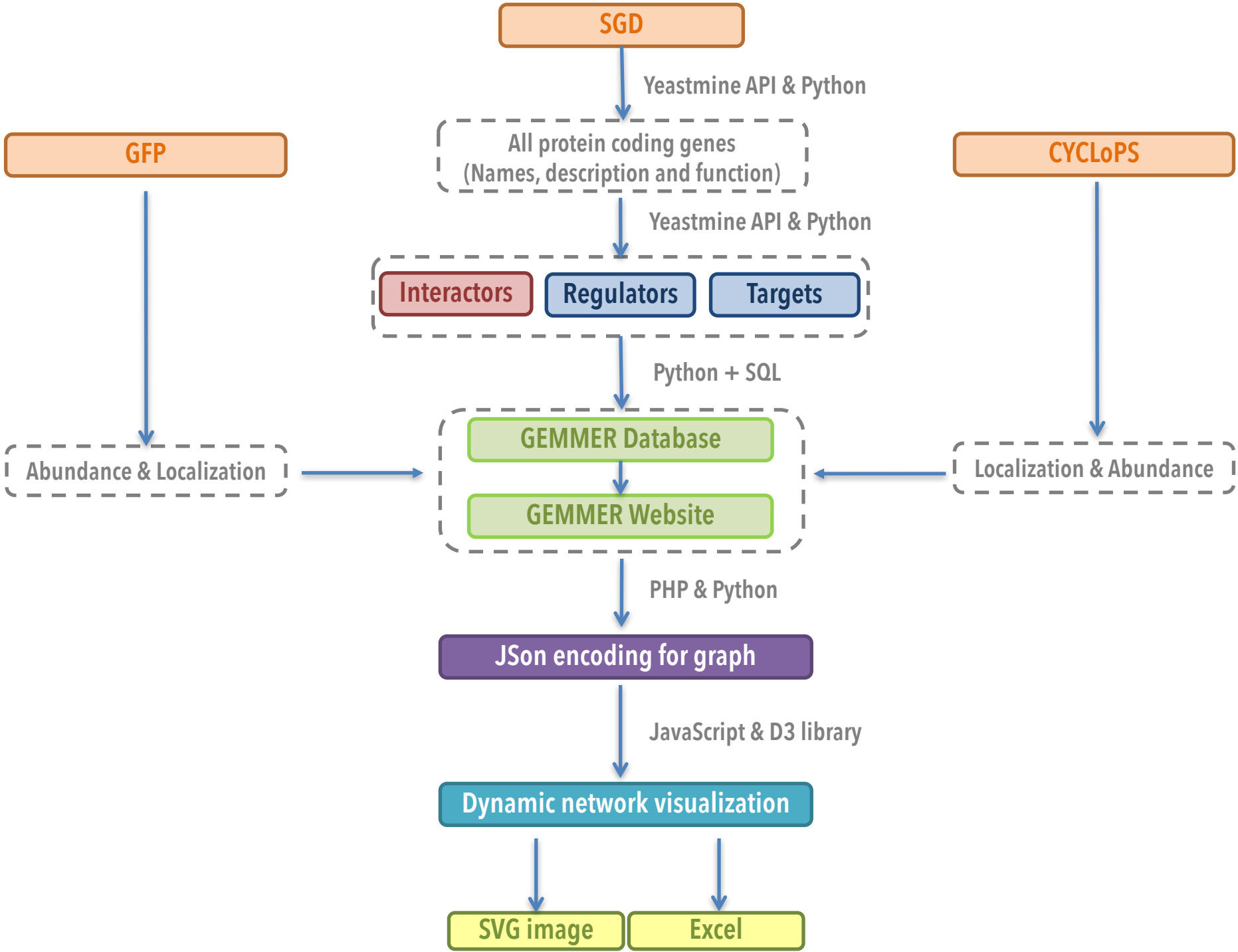
Comparison to existing tools

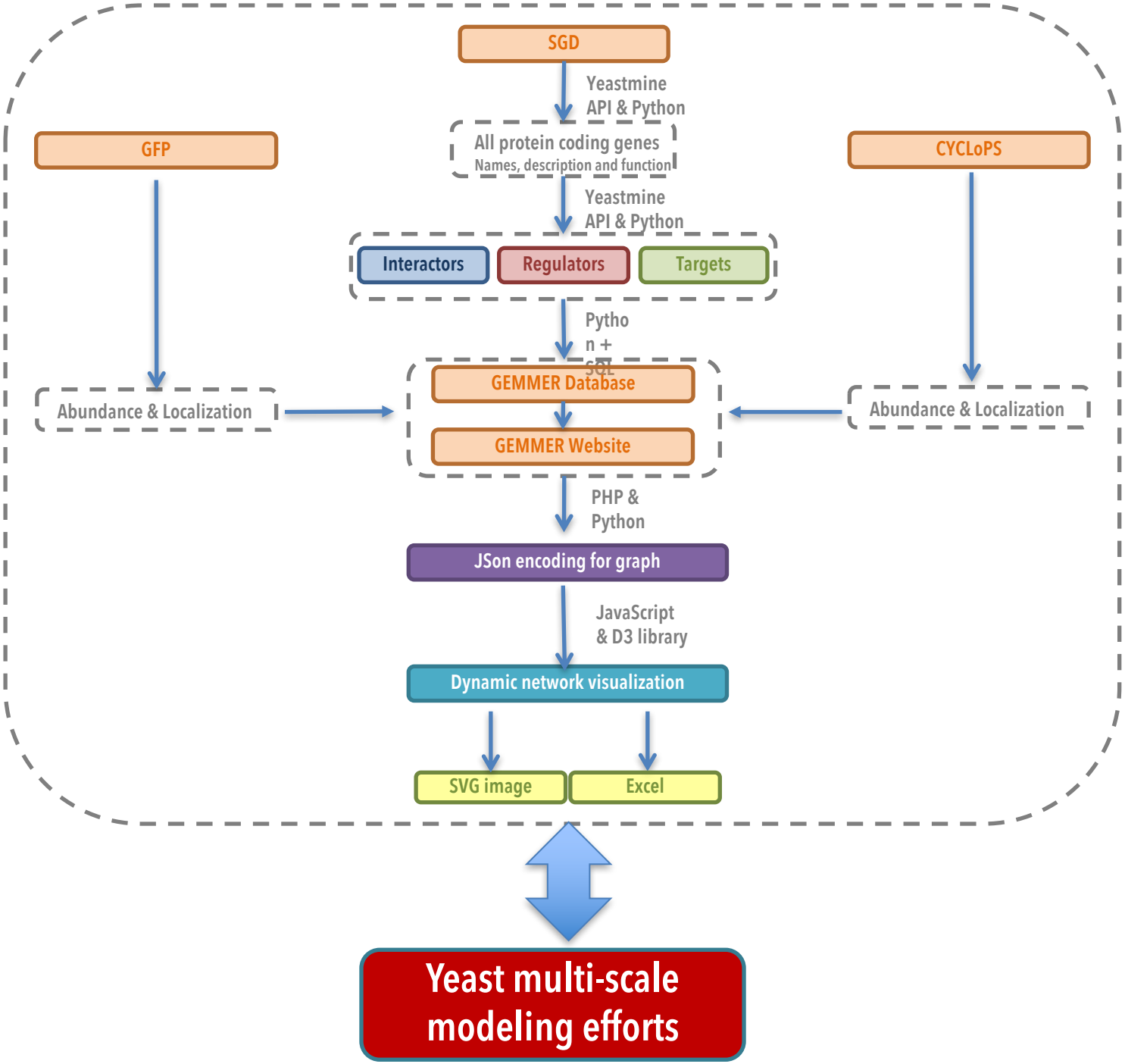
Features	GEMMER	STRING	BioGRID	APID	Arabidopsis Interactions Viewer	IntACT
Yeast specific and genome-wide	X					
Node size proportional to interactions in the visualized network	X		X			
Edge size proportional to interactions in the visualized network	X		X	X	X	
Coloring nodes by their GO-term annotated function: cell cycle, metabolism, signal transduction, DNA replication	X	X		Manually	X	
Visualizing both genetic and physical interactions and being able to filter out one of the two	X	X	X			
Showing interactions only if at least n experiments have shown an interaction	X	X*		X		
> n unique experimental methods	X					
multi-node	X					
Clustering nodes on function (GO-term)	X		X ***	X	X	
Incorporating localization information from the CYCLoPs database (Koh et al. 2015) and the GFP database (Ghaemmaghami et al. 2002) and being able to colour/cluster on it	X					
SVG + Excel export	X	X	X		X	

How GEMMER works (technical)

- GEMMER is built on common open-source software
  - Runs on a VPS (virtual private server)
  - Website based on Bootstrap framework
  - All information integrated in an SQL database
  - Python scripts perform the interfacing with the SQL database to generate networks
  - PHP to serve results to the user
  - Visualizations based on D3js (Data-Driven Documents Javascript)

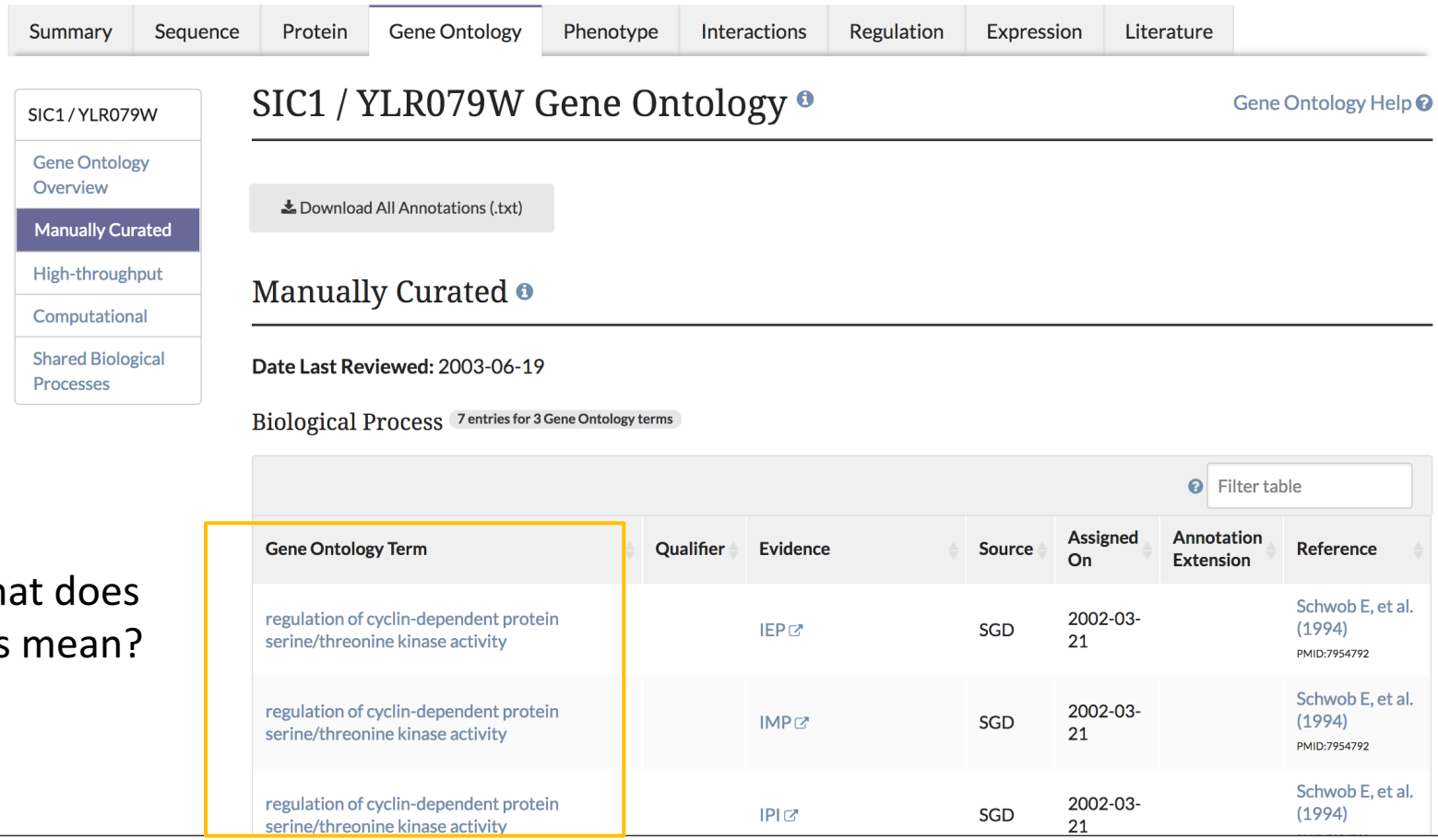

SGD
Yeastmine API & Python
All protein coding genes
(Names, description and function)
GFP
CYCLoPS
Yeastmine API & Python
Interactors
Regulators
Targets
Python + SQL
GEMMER Database
GEMMER Website
Abundance & Localization
Localization & Abundance
PHP & Python
JSon encoding for graph
JavaScript & D3 library
Dynamic network visualization
SVG image
Excel


SGD
Yeastmine API & Python
All protein coding genes
(Names, description and function)
GFP
CYCLoPS
Yeastmine API & Python
Interactors
Regulators
Targets
Python + SQL
GEMMER Database
GEMMER Website
Abundance & Localization
Localization & Abundance
PHP & Python
JSon encoding for graph
JavaScript & D3 library
Dynamic network visualization
SVG image
Excel


SGD
Yeastmine
API & Python
GFP
All protein coding genes
Names, description and function
CYCLoPS
Yeastmine
API & Python
Interactors
Regulators
Targets
Pytho
n +
SQL
GEMMER Database
Abundance & Localization
GEMMER Website
Abundance & Localization
PHP &
Python
JSon encoding for graph
JavaScript
& D3 library
Dynamic network visualization
SVG image
Excel
Yeast multi-scale
modeling efforts

Function prediction

- What does a specific protein do functionally?
- This is a ***hard*** question
- **GEMMER ”solves” this by using SGD’s GO annotations**

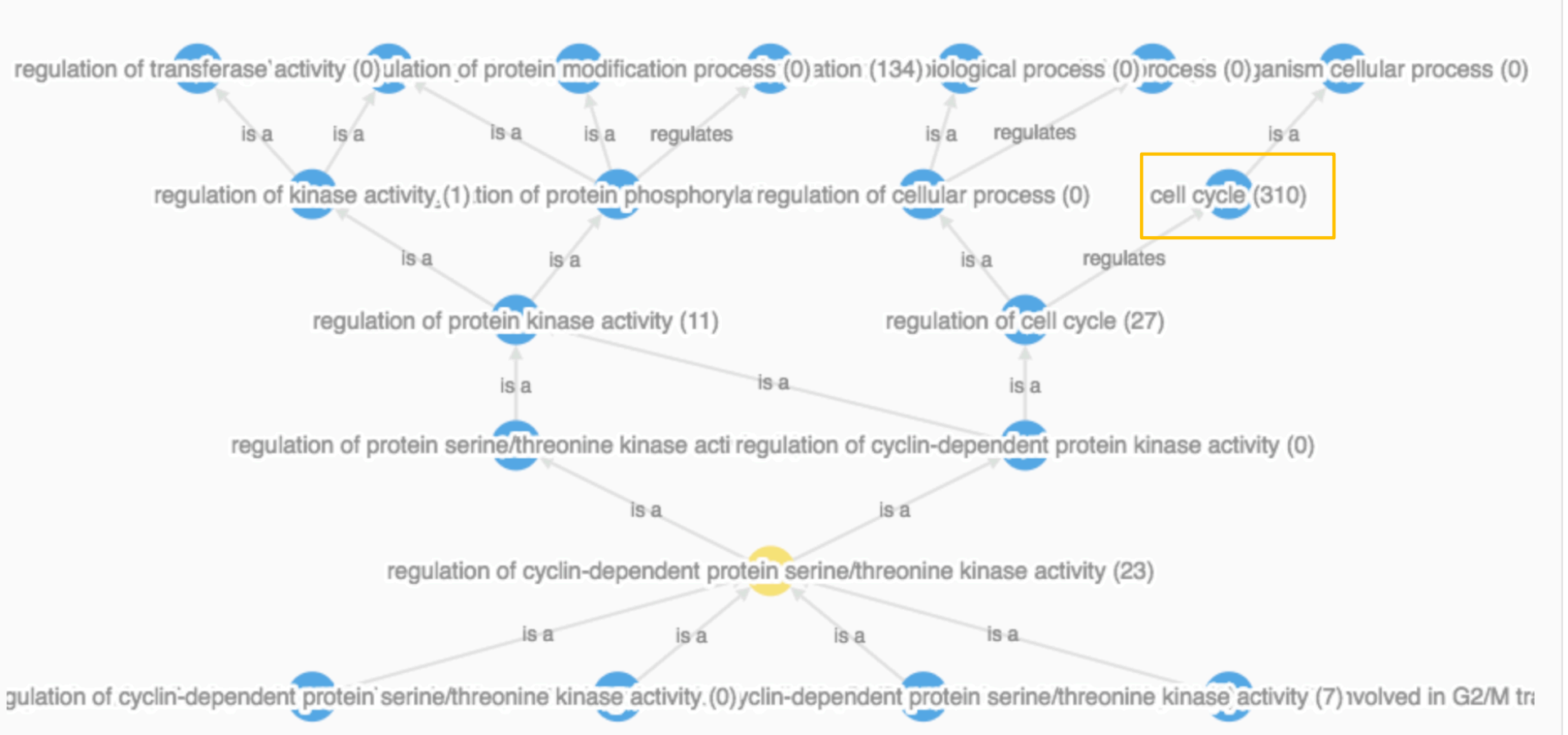


What does this mean?

Function prediction

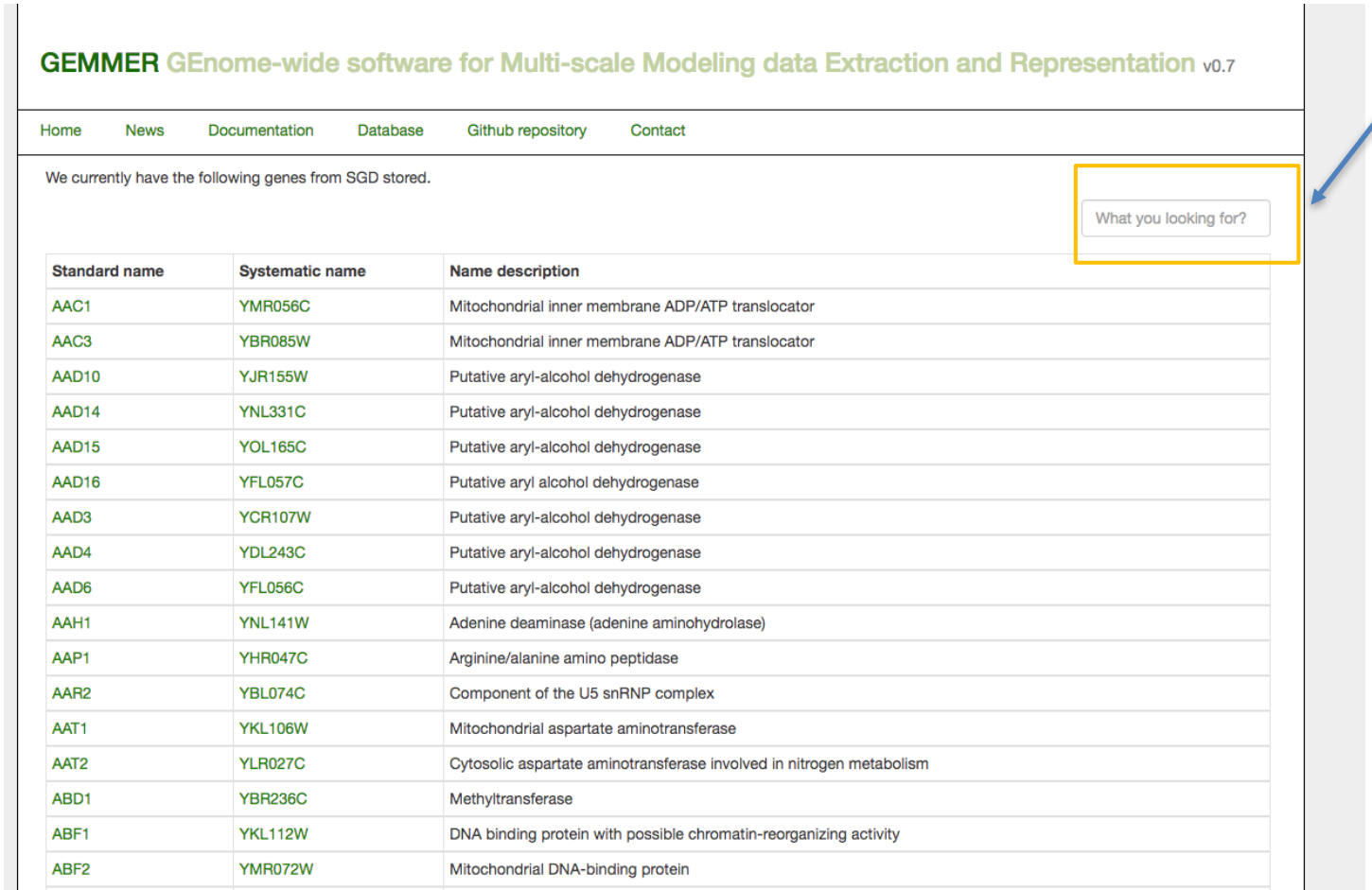
- GEMMER uses SGD GO annotations to predict functionality of a Gene in the **categories**:
  - Cell cycle
  - Cell division
  - Signal transduction
  - DNA replication
  - Metabolism
- This is done by assigning all annotated GO terms to one of these as a **parent** of the more specific GO term
- The two most abundantly annotated of these 5 are saved as GO term 1 and GO term 2

GO annotations exist in trees with “children” and “parents”



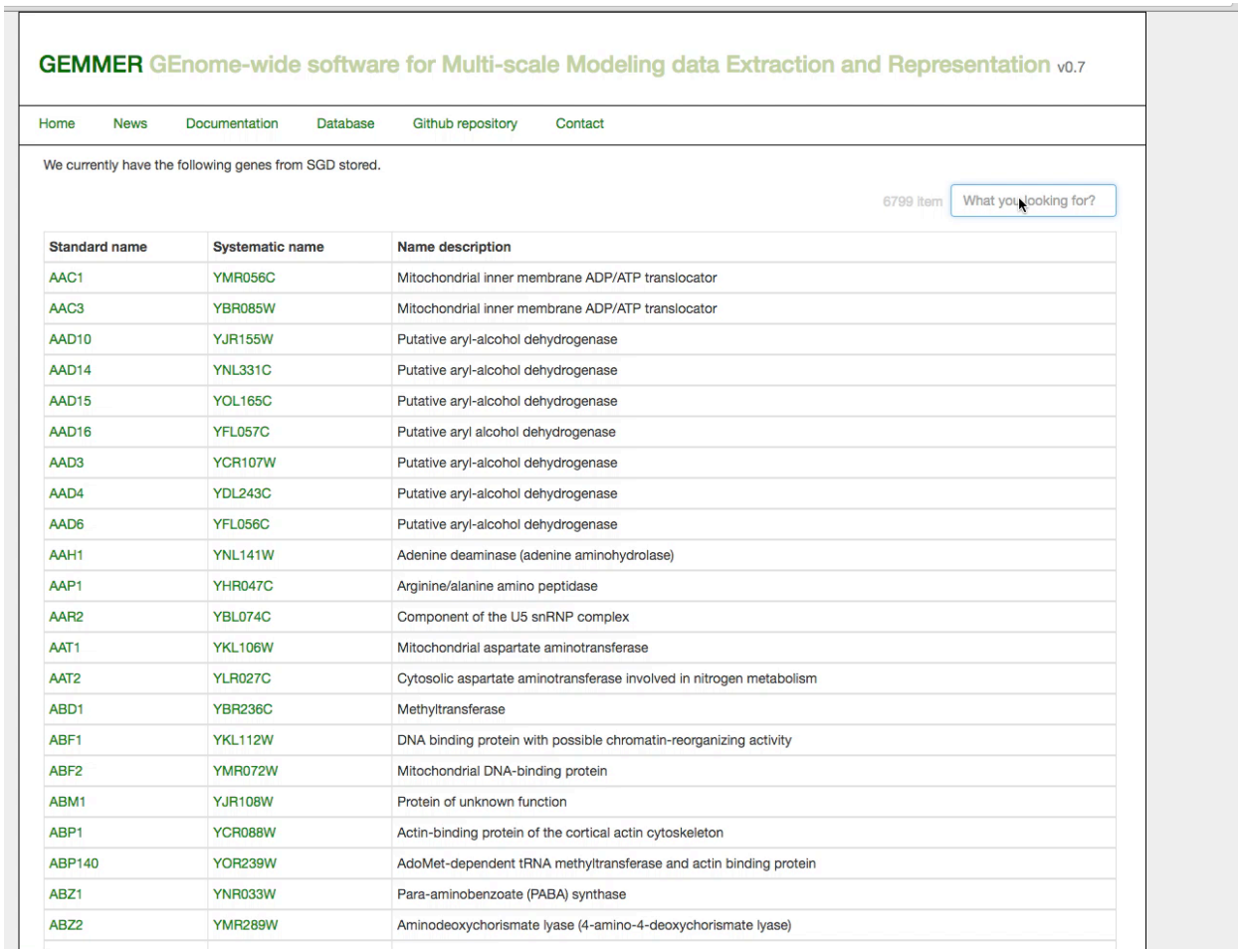
The database

Aside from visualization all info in the database can be searched and interacted with



Standard name	Systematic name	Name description
AAC1	YMR056C	Mitochondrial inner membrane ADP/ATP translocator
AAC3	YBR085W	Mitochondrial inner membrane ADP/ATP translocator
AAD10	YJR155W	Putative aryl-alcohol dehydrogenase
AAD14	YNL331C	Putative aryl-alcohol dehydrogenase
AAD15	YOL165C	Putative aryl-alcohol dehydrogenase
AAD16	YFL057C	Putative aryl alcohol dehydrogenase
AAD3	YCR107W	Putative aryl-alcohol dehydrogenase
AAD4	YDL243C	Putative aryl-alcohol dehydrogenase
AAD6	YFL056C	Putative aryl-alcohol dehydrogenase
AAH1	YNL141W	Adenine deaminase (adenine aminohydrolase)
AAP1	YHR047C	Arginine/alanine amino peptidase
AAR2	YBL074C	Component of the U5 snRNP complex
AAT1	YKL106W	Mitochondrial aspartate aminotransferase
AAT2	YLR027C	Cytosolic aspartate aminotransferase involved in nitrogen metabolism
ABD1	YBR236C	Methyltransferase
ABF1	YKL112W	DNA binding protein with possible chromatin-reorganizing activity
ABF2	YMR072W	Mitochondrial DNA-binding protein

Database: live search on name & description (video demo)



GEMMER GEnome-wide software for Multi-scale Modeling data Extraction and Representation v0.7

Home | News | Documentation | Database | Github repository | Contact

We currently have the following genes from SGD stored.

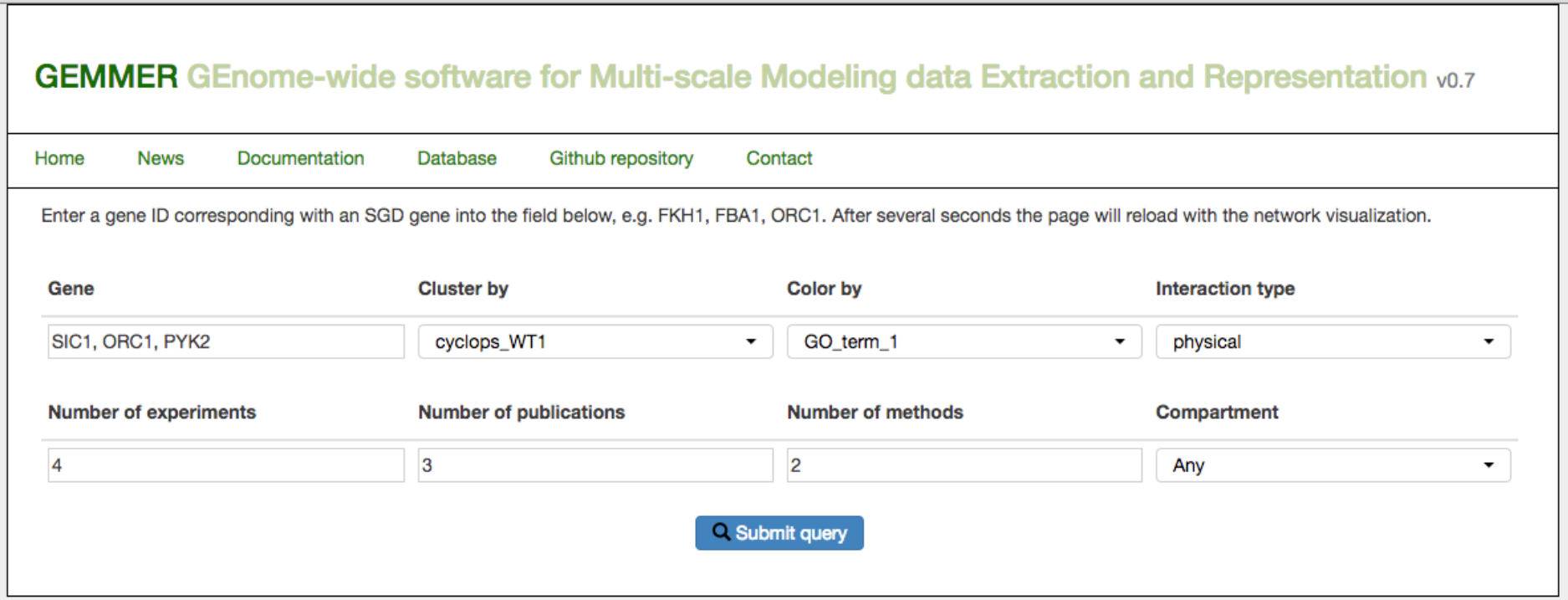
6799 Item | What you looking for?

Standard name	Systematic name	Name description
AAC1	YMR056C	Mitochondrial inner membrane ADP/ATP translocator
AAC3	YBR085W	Mitochondrial inner membrane ADP/ATP translocator
AAD10	YJR155W	Putative aryl-alcohol dehydrogenase
AAD14	YNL331C	Putative aryl-alcohol dehydrogenase
AAD15	YOL165C	Putative aryl-alcohol dehydrogenase
AAD16	YFL057C	Putative aryl alcohol dehydrogenase
AAD3	YCR107W	Putative aryl-alcohol dehydrogenase
AAD4	YDL243C	Putative aryl-alcohol dehydrogenase
AAD6	YFL056C	Putative aryl-alcohol dehydrogenase
AAH1	YNL141W	Adenine deaminase (adenine aminohydrolase)
AAP1	YHR047C	Arginine/alanine amino peptidase
AAR2	YBL074C	Component of the U5 snRNP complex
AAT1	YKL106W	Mitochondrial aspartate aminotransferase
AAT2	YLR027C	Cytosolic aspartate aminotransferase involved in nitrogen metabolism
ABD1	YBR236C	Methyltransferase
ABF1	YKL112W	DNA binding protein with possible chromatin-reorganizing activity
ABF2	YMR072W	Mitochondrial DNA-binding protein
ABM1	YJR108W	Protein of unknown function
ABP1	YCR088W	Actin-binding protein of the cortical actin cytoskeleton
ABP140	YOR239W	AdoMet-dependent tRNA methyltransferase and actin binding protein
ABZ1	YNR033W	Para-aminobenzoate (PABA) synthase
ABZ2	YMR289W	Aminodeoxychorismate lyase (4-amino-4-deoxychorismate lyase)

The user-interface

Interaction with the GEMMER database for building visualizations happens through this form

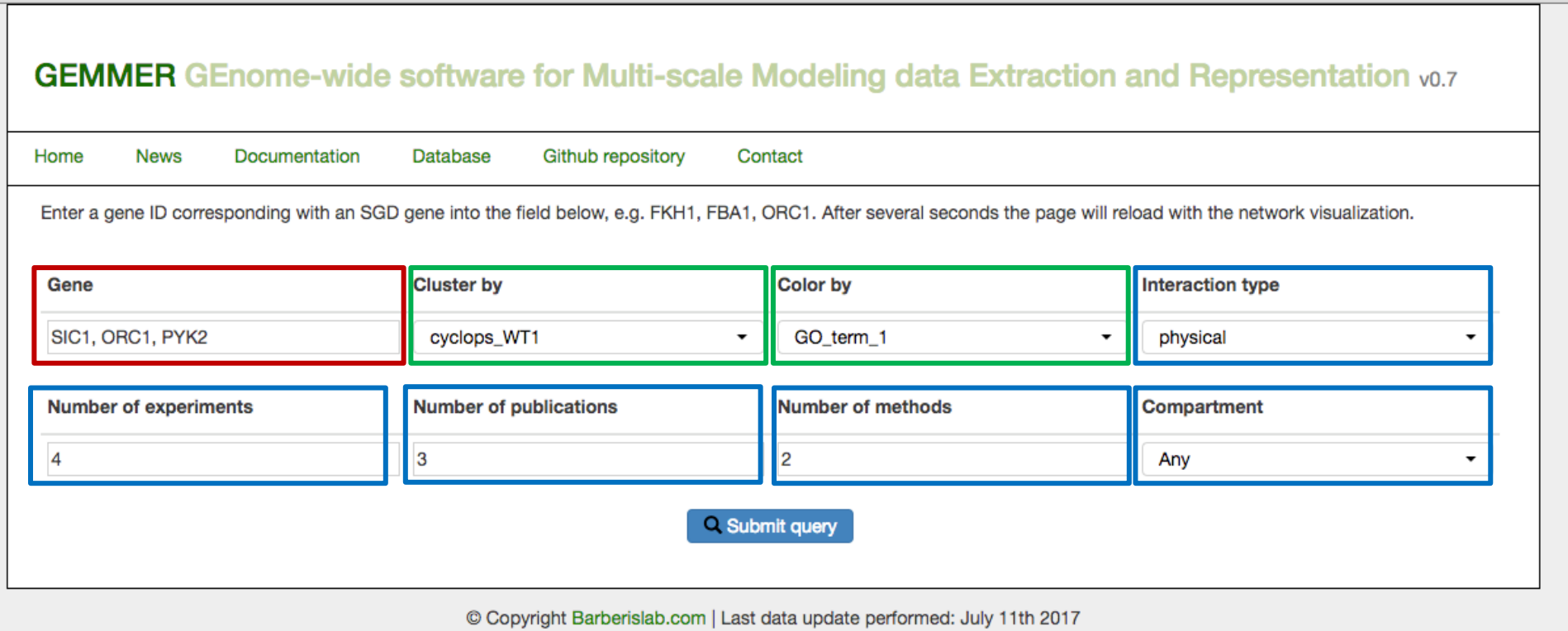
The user has various options: protein selection, visualization options and filters



The user-interface

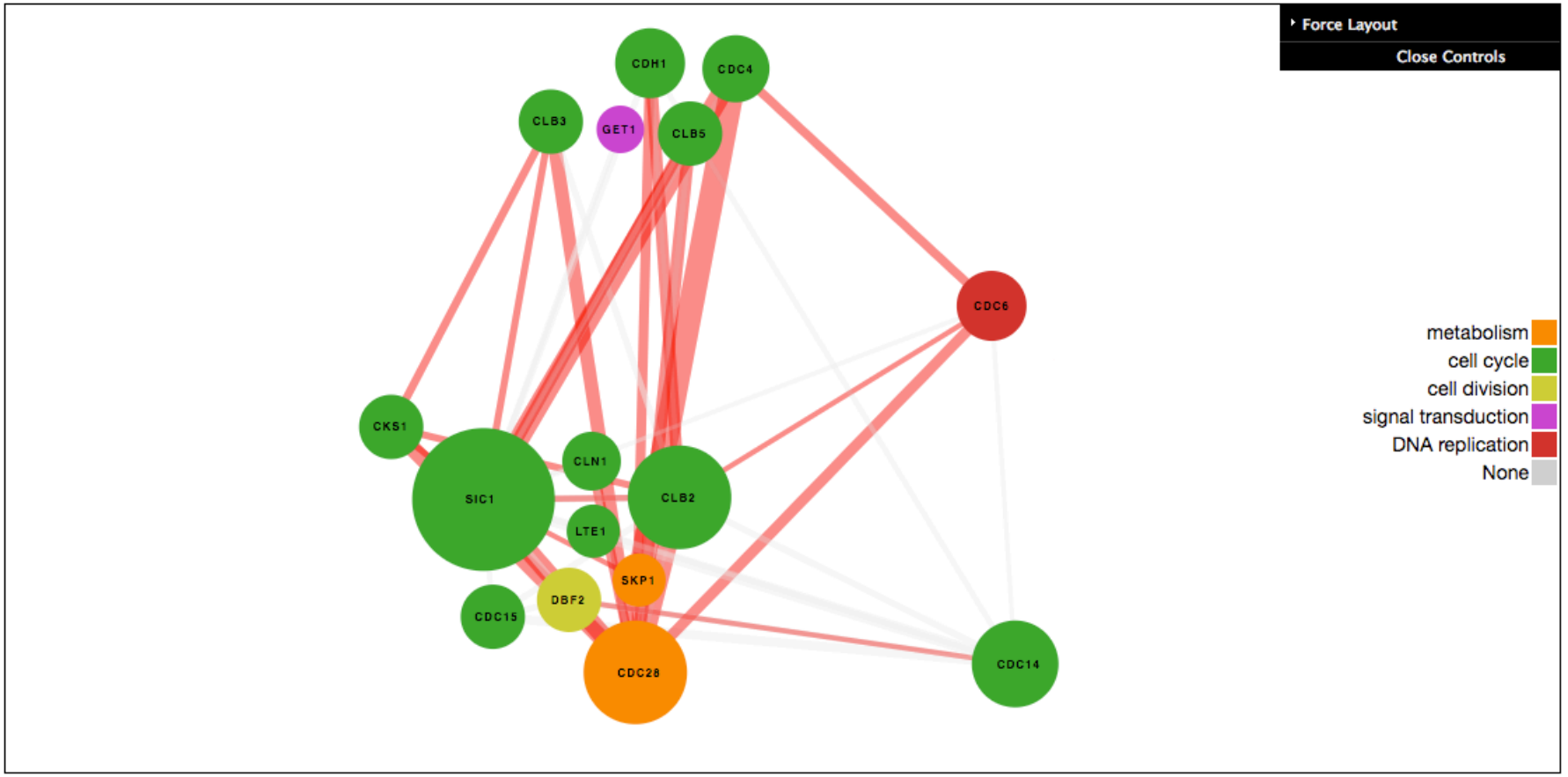
Interaction with the GEMMER database for building visualizations happens through this form

The user has various options: protein selection, visualization options and filters



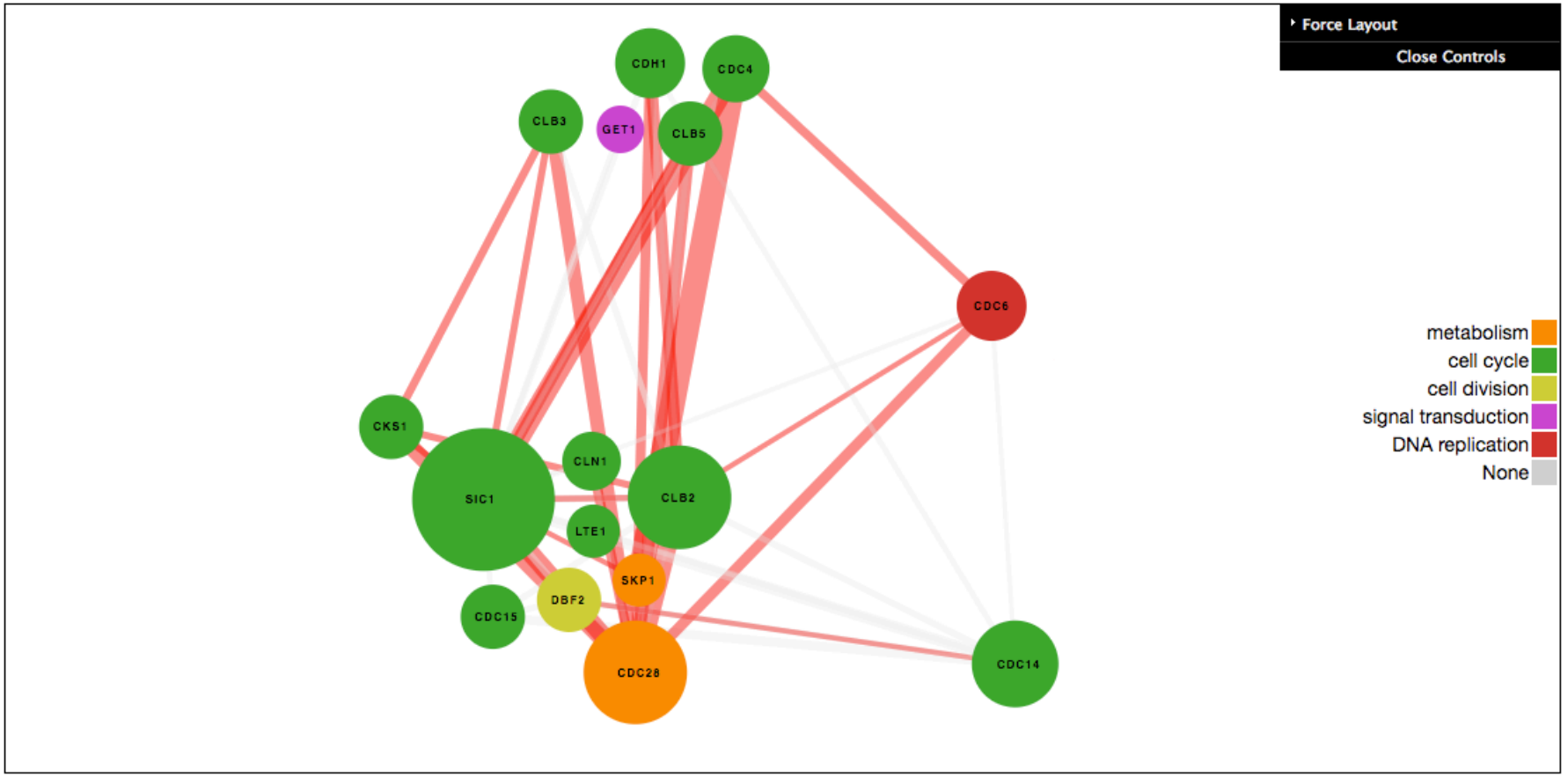
Example visualization: 1 protein (Sic1)

Any interaction type, minimally 4 experiments in at least 3 papers with at least 2 different methods.



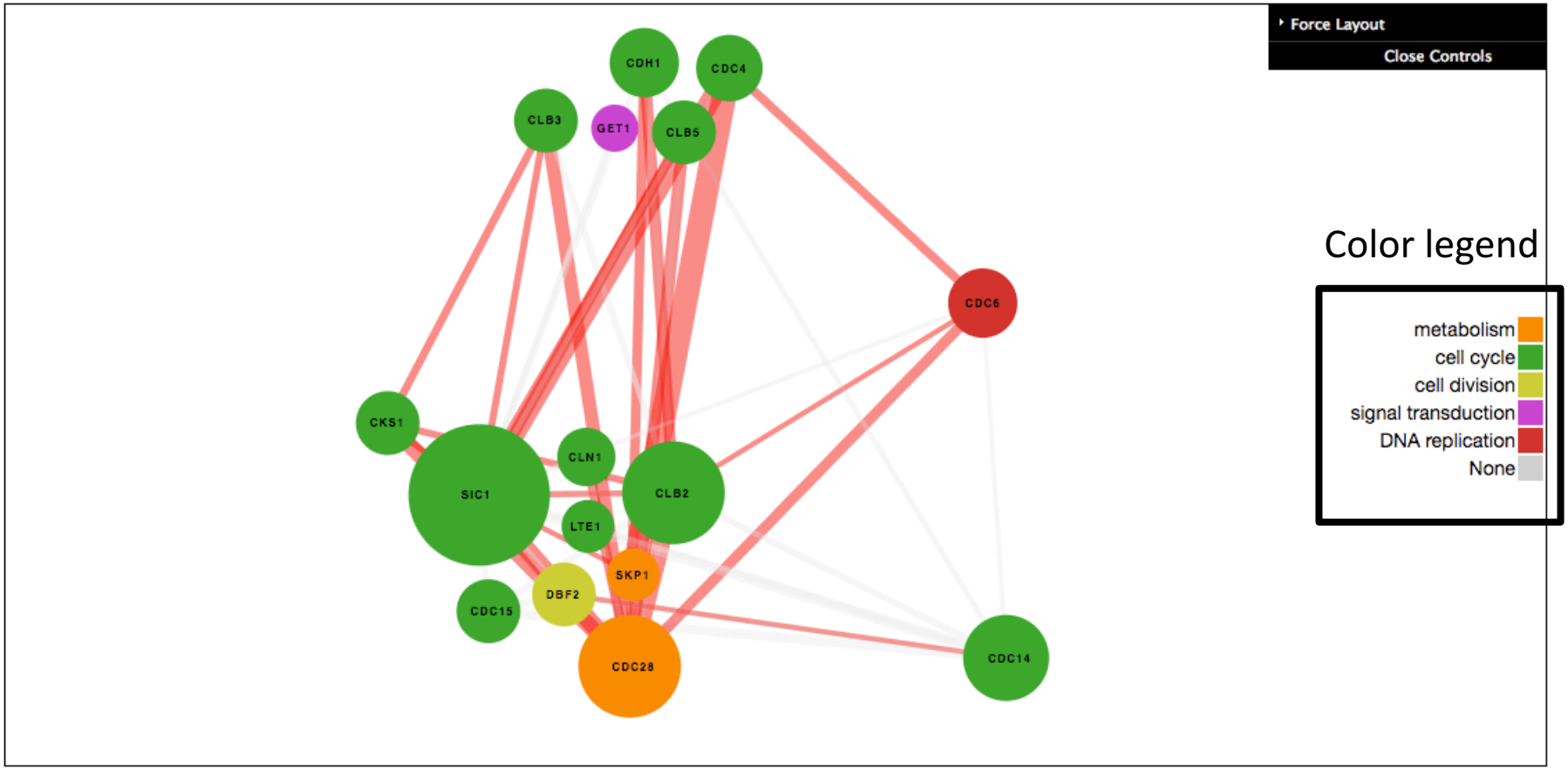
Example visualization: 1 protein (Sic1)

Note the clustering and coloring. Here the clustering indicates the compartment in the WT1 experiment of the CYCLoPS database where this protein is found most abundantly.
The coloring indicates its likely dominant function based on GO term annotation.



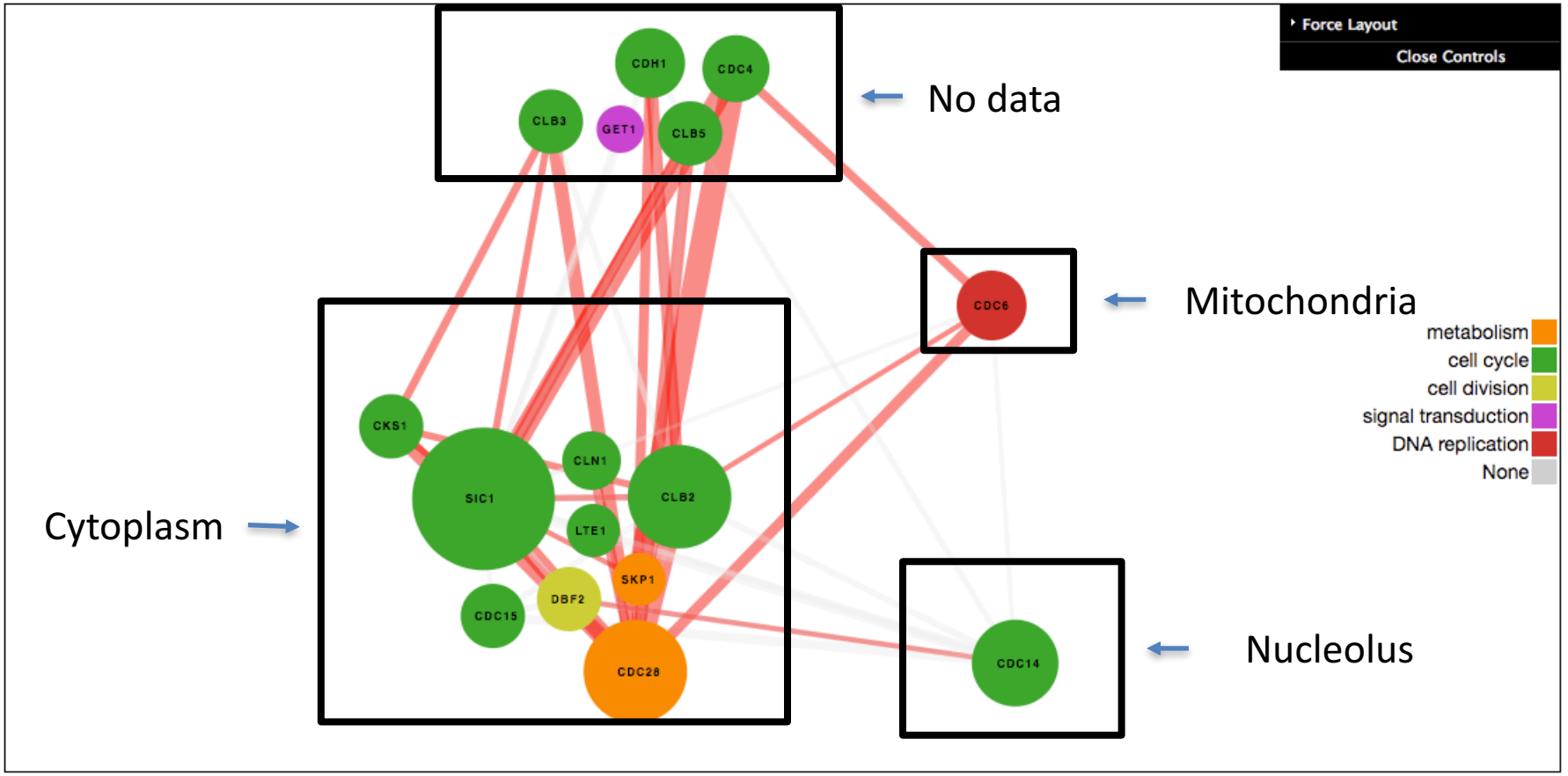
Example visualization: 1 protein (Sic1)

Note the clustering and coloring. Here the clustering indicates the compartment in the WT1 experiment of the CYCLoPS database where this protein is found most abundantly.
The **coloring** indicates its likely dominant function based on GO term annotation.



Example visualization: 1 protein (Sic1)

Note the clustering and colouring. Here the **clustering** indicates the compartment in the WT1 experiment of the CYCLoPS database where this protein is found most abundantly.
The coloring indicates its likely dominant function based on GO term annotation.

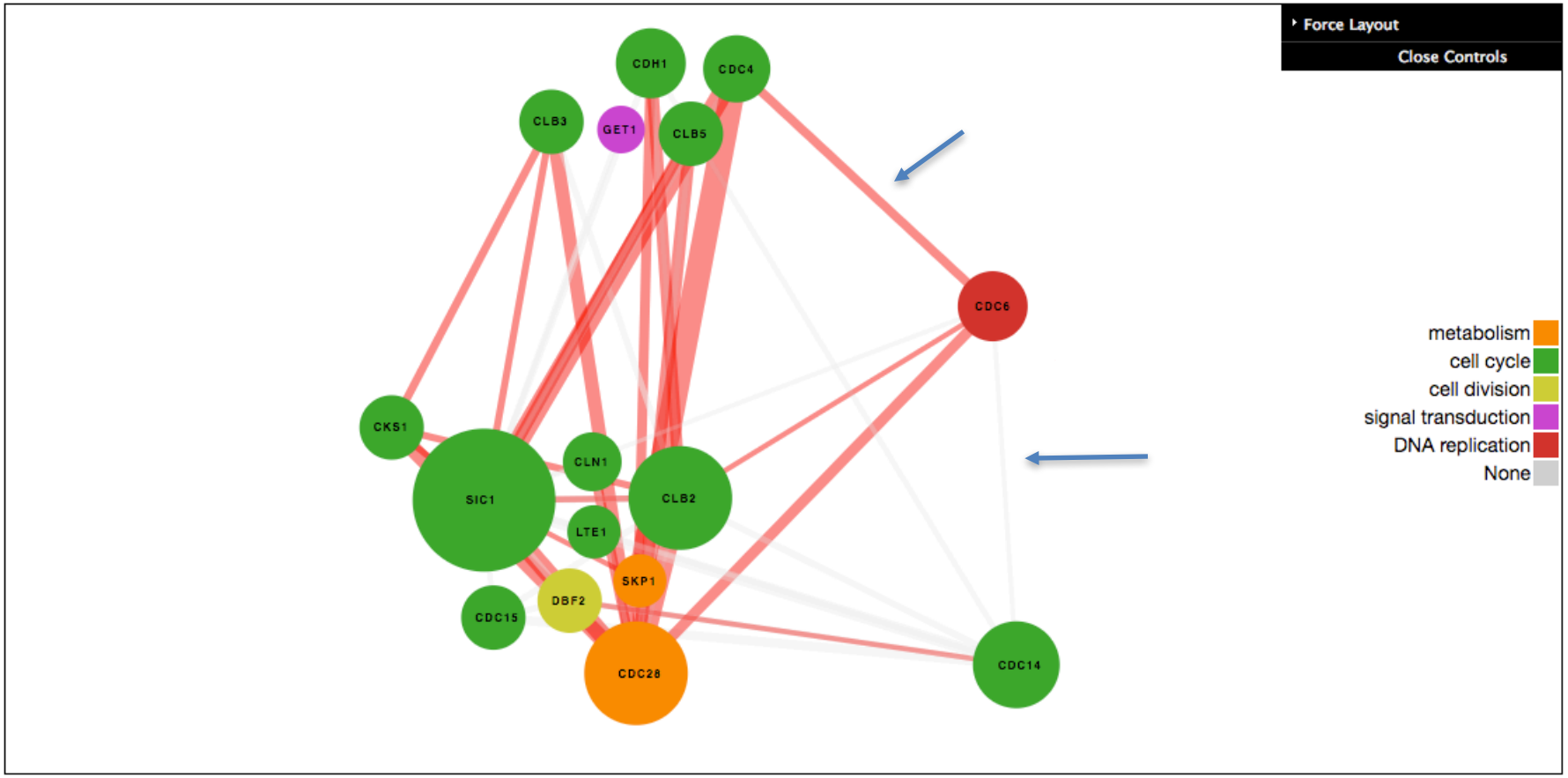


Example visualization: 1 protein (Sic1)

Interactions are colored according to the interaction type:

- Physical
- Genetic
- Regulatory

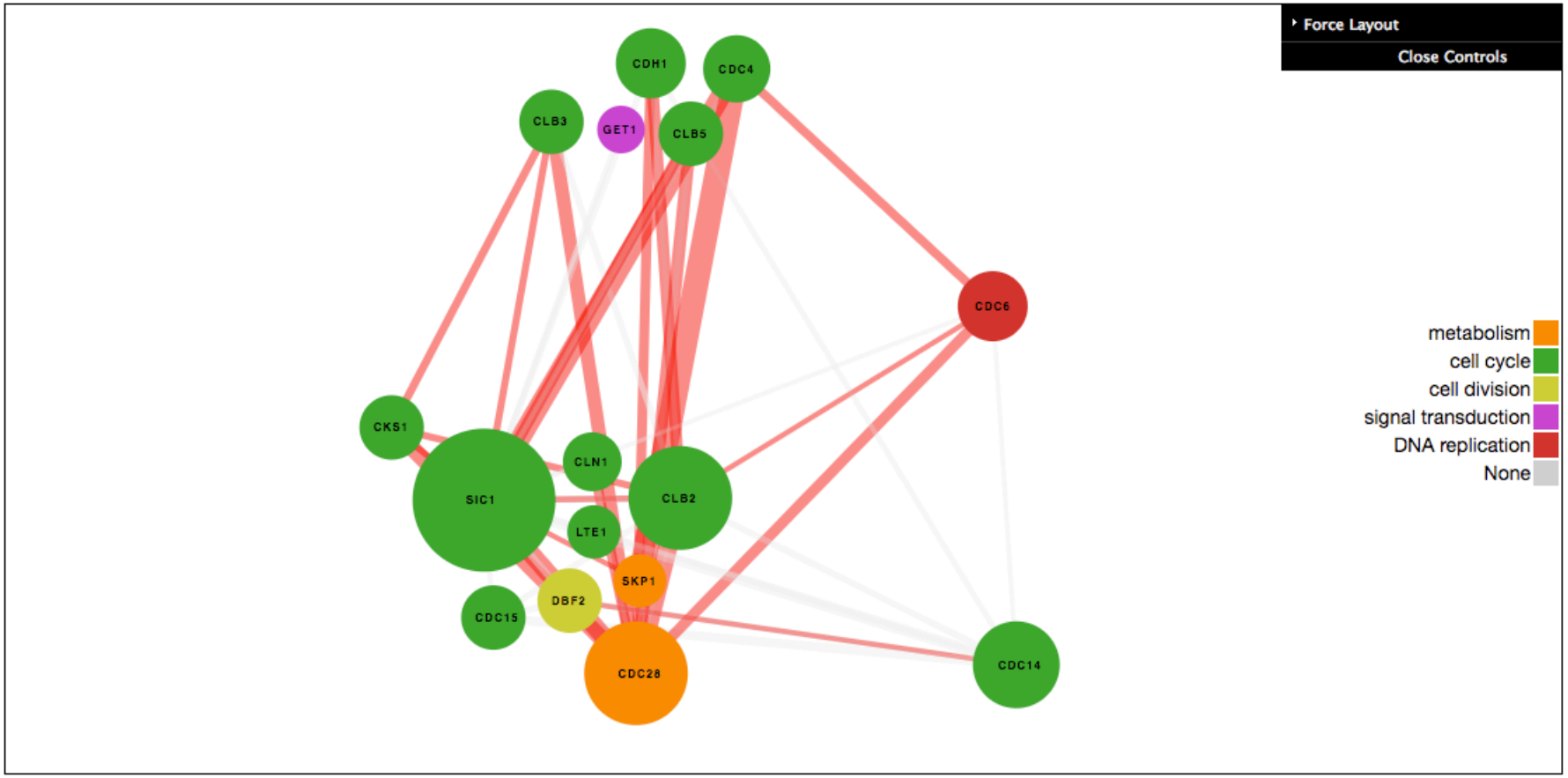
Thickness is proportional to the number of experiments supporting this interaction



Example visualization: 1 protein (Sic1)

Node size is proportional to the number of interactions in the current network.

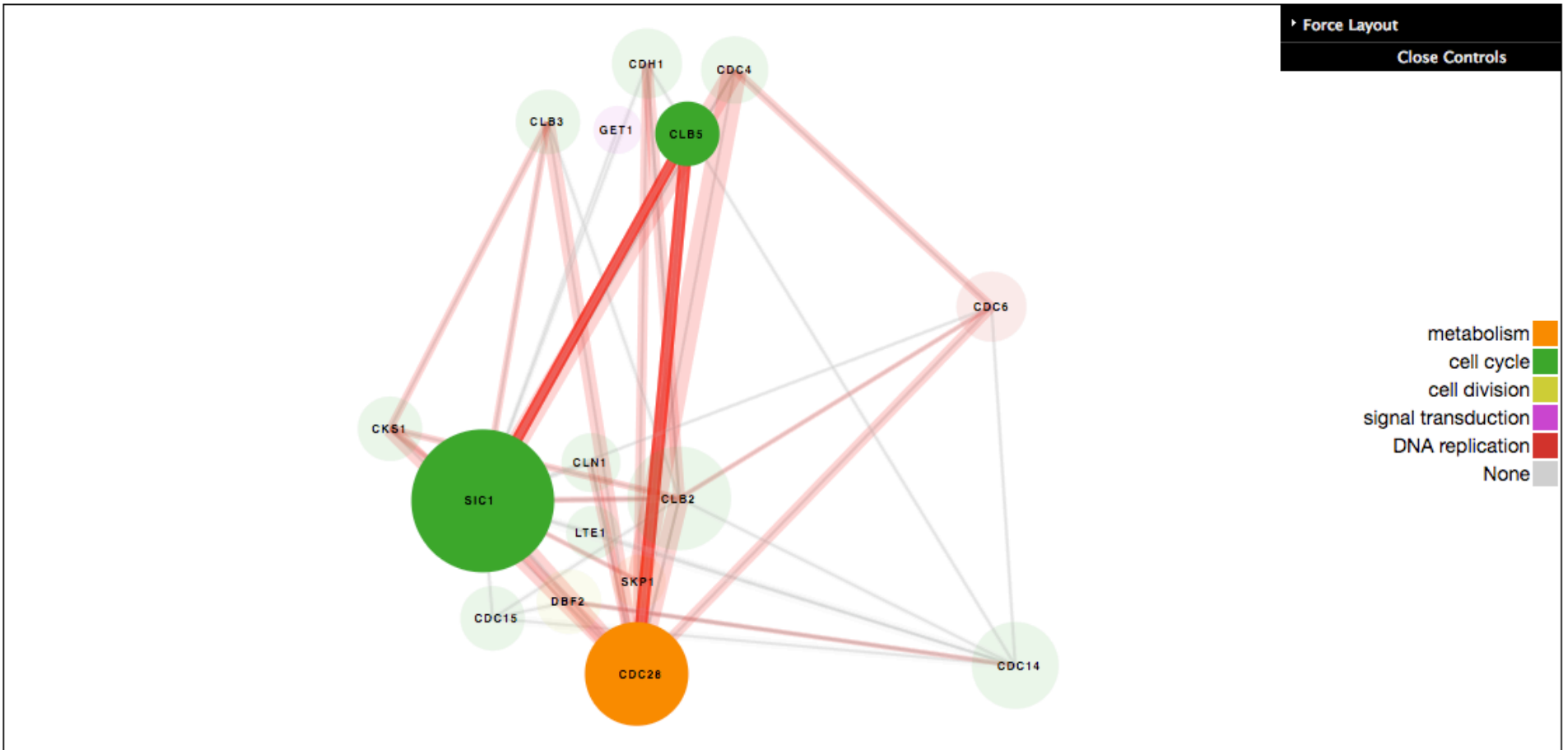
Notice that Sic1 is the biggest, since it is the *generator protein* for this network (it interacts with all nodes by design)



Example visualization: 1 protein (Sic1)

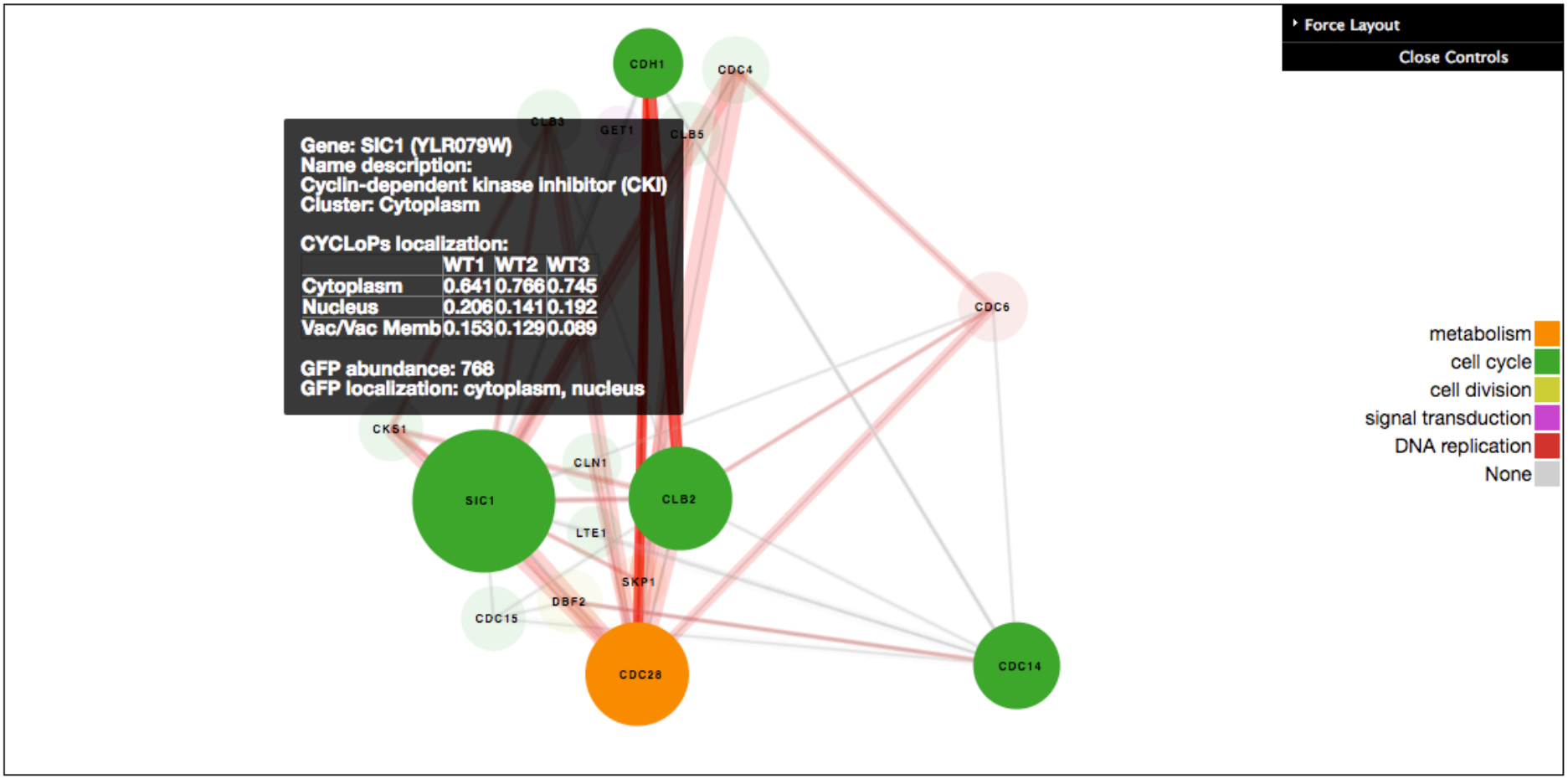
Clicking a node highlights its connections in the network

Focus on CLB5 for instance

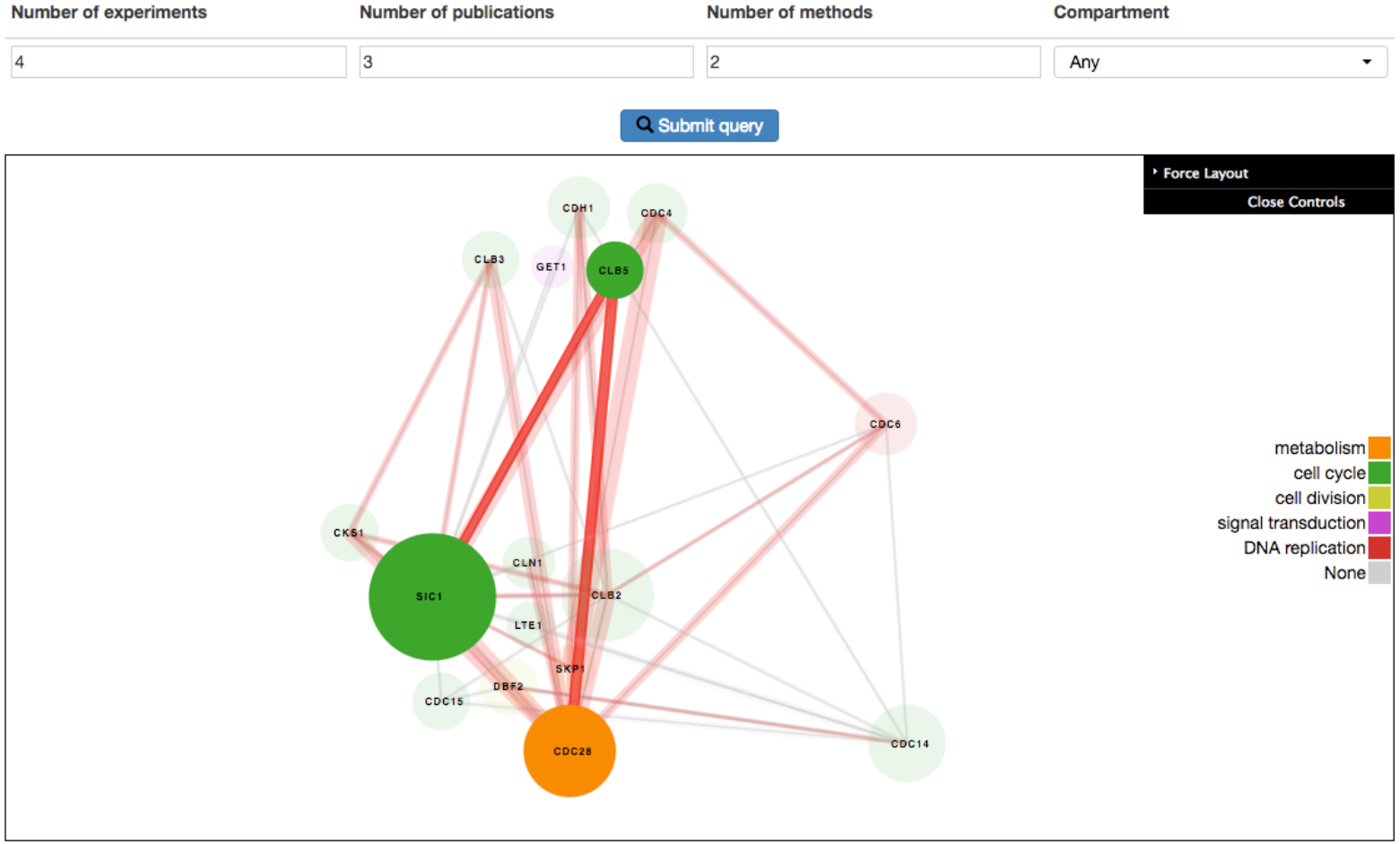


Example visualization: 1 protein (Sic1)

Upon hovering over a protein popup information appears



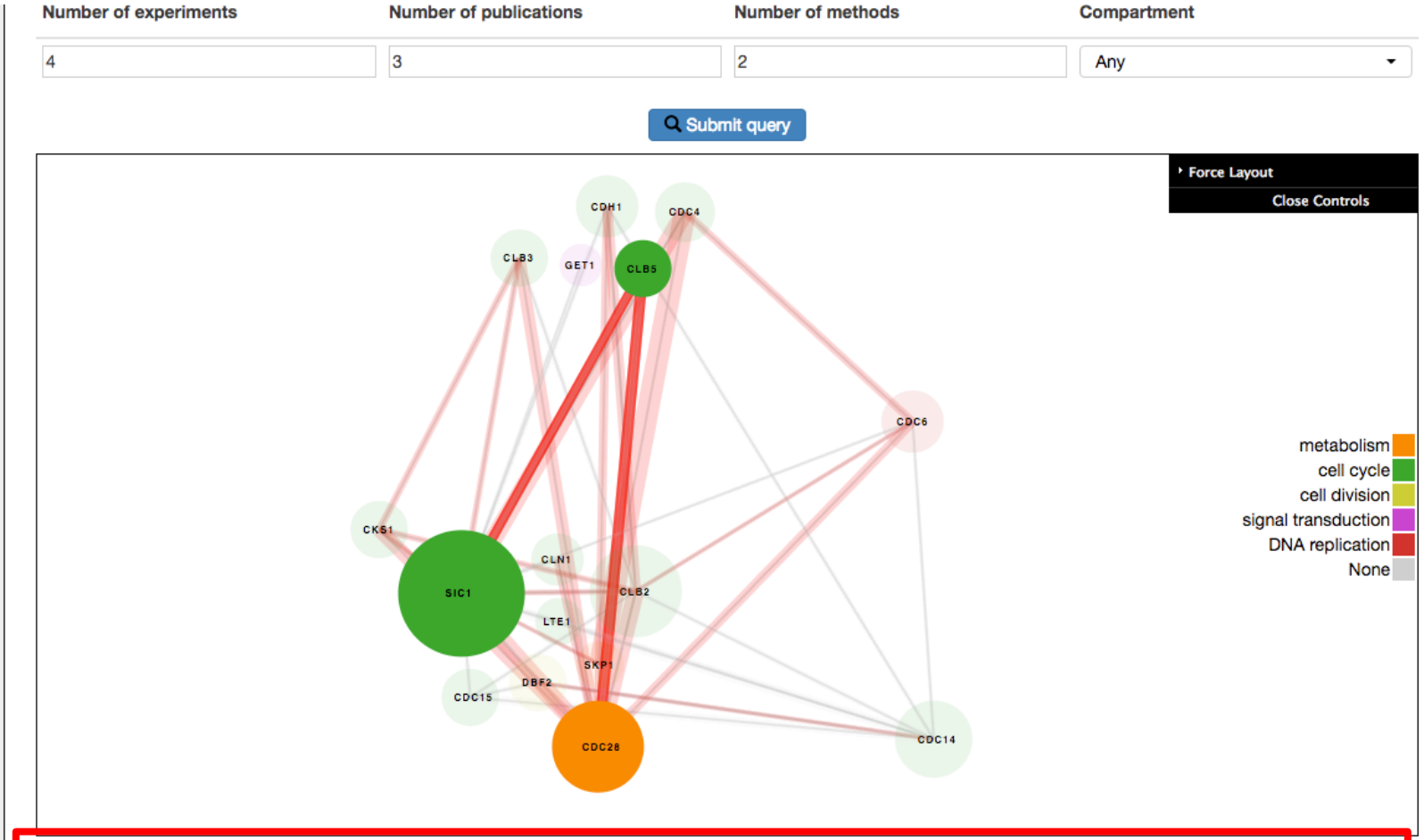
Exporting options: Excel and SVG



Export options

We provide the ability to download the image in SVG format by right-clicking "Download" and then clicking "Save link as". A formatted Excel workbook is also provided.
Download SVG | Download Excel workbook | Link to standalone HTML page

Exporting options: Excel and SVG



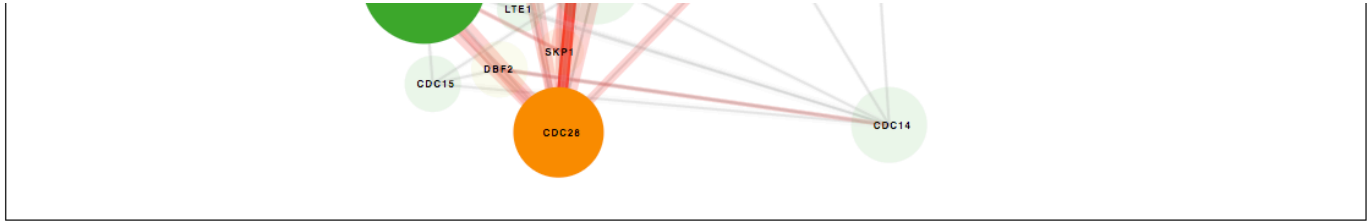
Export options

We provide the ability to download the image in SVG format by right-clicking "Download" and then clicking "Save link as". A formatted Excel workbook is also provided.
Download SVG | Download Excel workbook | Link to standalone HTML page

Tables of nodes and interactions

Below the export options are three tables with information regarding the network visualized above.

- The settings for this visualization
- Information on the nodes (proteins) of the network
- Information on the interactions



Export options

We provide the ability to download the image in SVG format by right-clicking "Download" and then clicking "Save link as". A formatted Excel workbook is also provided.
Download SVG | Download Excel workbook | Link to standalone HTML page

User input

Nodes

Interactions

source	target	type	evidence	#experiments	#publication	#methods
CDC14	CDC15	genetic	Dosage Rescue (8668128), Dosage Rescue (9613578), Dosage Rescue (9763445), Dosage Rescue (9885559), Phenotypic Enhancement (25049391), Synthetic Growth Defect (18927509), Synthetic Growth Defect (9613578), Synthetic Lethality (25049391), Synthetic Rescue (11914130), Synthetic Rescue (21036905)	10	8	5
CDC14	CDC6	genetic	Dosage Lethality (20980394), Dosage Rescue (12960422), Synthetic Growth Defect (11460169), Synthetic Growth Defect (12960422)	4	3	3
CDC14	CDH1	genetic	Phenotypic Enhancement (25049391), Phenotypic Suppression (22321970), Phenotypic Suppression (9885559), Synthetic Growth Defect (25049391), Synthetic Lethality (10732674)	5	4	4
CDC14	CLB2	genetic	Dosage Lethality (9763445), Negative Genetic (20093466), Phenotypic Enhancement (21784165), Phenotypic Suppression (23468650), Synthetic Lethality (10732674), Synthetic Rescue (23468650)	6	5	6
CDC14	DBF2	genetic	Dosage Rescue (10102376), Dosage Rescue (9613578), Dosage Rescue (9885559), Negative Genetic (20093466), Synthetic Growth Defect (18927509), Synthetic Growth Defect (9613578)	6	5	3

Tables of nodes and interactions

Below the export options are three tables with information regarding the network visualized above.

- The settings for this visualization
- Information on the nodes (proteins) of the network
- Information on the interactions

User input

	user input
Genes	SIC1
Cluster by	cyclops_WT1
Color by	GO_term_1
Interaction type	all
Minimal number of experiments	4
Minimal number of publications	3
Minimal number of methods	2
Compartment	Any

Nodes

Interactions

Tables of nodes and interactions

Below the export options are three tables with information regarding the network visualized above.

- The settings for this visualization
- Information on the nodes (proteins) of the network
- Information on the interactions

User input

Nodes

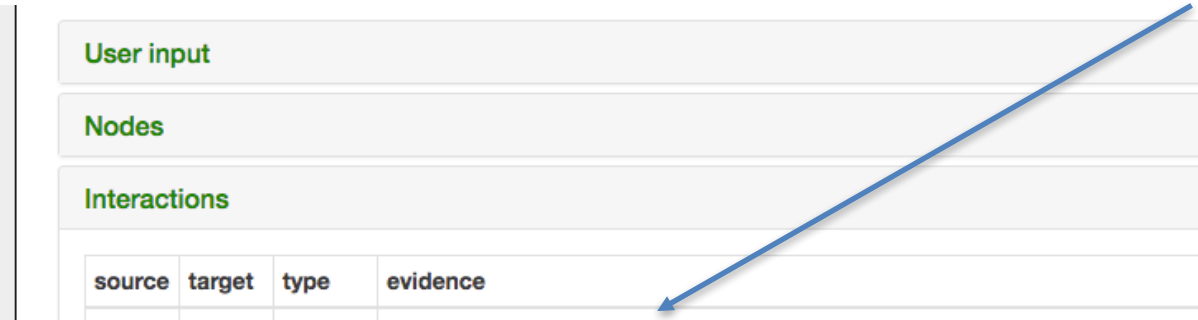
Standard name	Systematic name	Name description	GO term 1	GO term 2	GFP abundance	GFP localization	cyclops_raw
CDC14	YFR028C	Protein phosphatase required for mitotic exit	cell cycle	cell division	8550	nucleolus	WT1 / WT2 / WT3 Mito 0.133 NaN NaN Nucleolus 0.727 0.727 0.678 Nucleus 0.186 0.202 0.121 SpindlePole 0.216 0.270 0.441 Vac/Vac Memb NaN 0.057 NaN
CDC15	YAR019C	Protein kinase of the Mitotic Exit Network	cell cycle	cell division	238	ambiguous, spindle pole	WT1 / WT2 / WT3 Bud NaN NaN 0.103 Budneck NaN NaN 0.059 Cytoplasm 0.739 0.744 0.757 Endosome NaN 0.095 NaN Mito NaN 0.083 0.070 SpindlePole 0.188 0.079 0.141
CDC28	YBR160W	Cyclin-dependent kinase (CDK) catalytic subunit	metabolism	cell cycle	6670	cytoplasm, nucleus	WT1 / WT2 / WT3 Cytoplasm 0.555 0.493 0.562

Tables of nodes and interactions

Below the export options are three tables with information regarding the network visualized above.

- The settings for this visualization
- Information on the nodes (proteins) of the network
- **Information on the interactions**

Pubmed links



User input

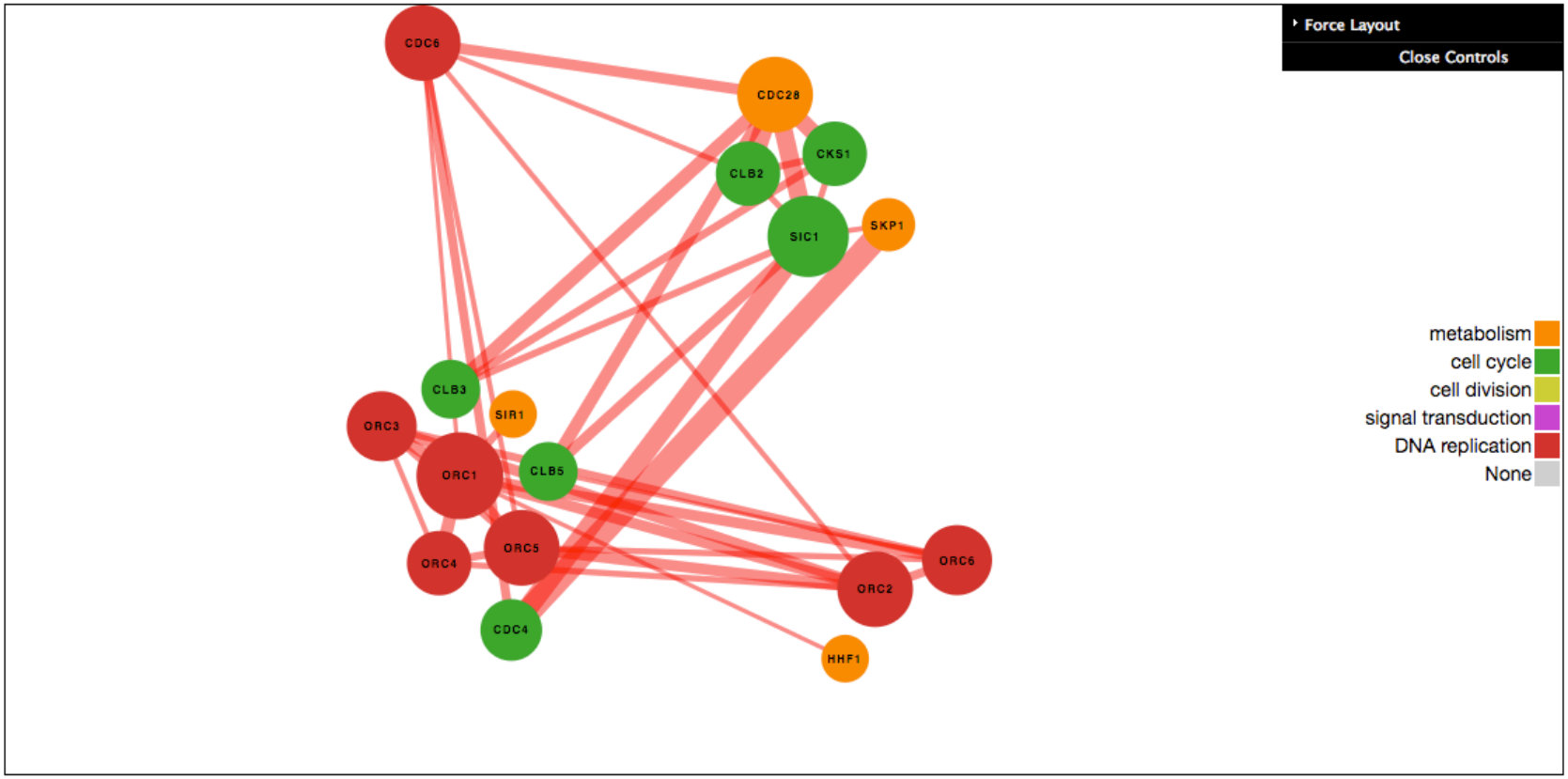
Nodes

Interactions

source	target	type	evidence	#experiments	#publication	#methods
CDC14	CDC15	genetic	Dosage Rescue (8668128), Dosage Rescue (9613578), Dosage Rescue (9763445), Dosage Rescue (9885559), Phenotypic Enhancement (25049391), Synthetic Growth Defect (18927509), Synthetic Growth Defect (9613578), Synthetic Lethality (25049391), Synthetic Rescue (11914130), Synthetic Rescue (21036905)	10	8	5
CDC14	CDC6	genetic	Dosage Lethality (20980394), Dosage Rescue (12960422), Synthetic Growth Defect (11460169), Synthetic Growth Defect (12960422)	4	3	3
CDC14	CDH1	genetic	Phenotypic Enhancement (25049391), Phenotypic Suppression (22321970), Phenotypic Suppression (9885559), Synthetic Growth Defect (25049391), Synthetic Lethality (10732674)	5	4	4
CDC14	CLB2	genetic	Dosage Lethality (9763445), Negative Genetic (20093466), Phenotypic Enhancement (21784165), Phenotypic Suppression (23468650), Synthetic Lethality (10732674), Synthetic Rescue (23468650)	6	5	6
CDC14	DBF2	genetic	Dosage Rescue (10102376), Dosage Rescue (9613578), Dosage Rescue (9885559), Negative Genetic (20093466), Synthetic Growth Defect (18927509), Synthetic Growth Defect (9613578)	6	5	3
CDC14	DBF2	physical	Affinity Capture-MS (20489023), Biochemical Activity (19221193), Biochemical Activity (20123997), Co-localization (10984431)	4	4	3

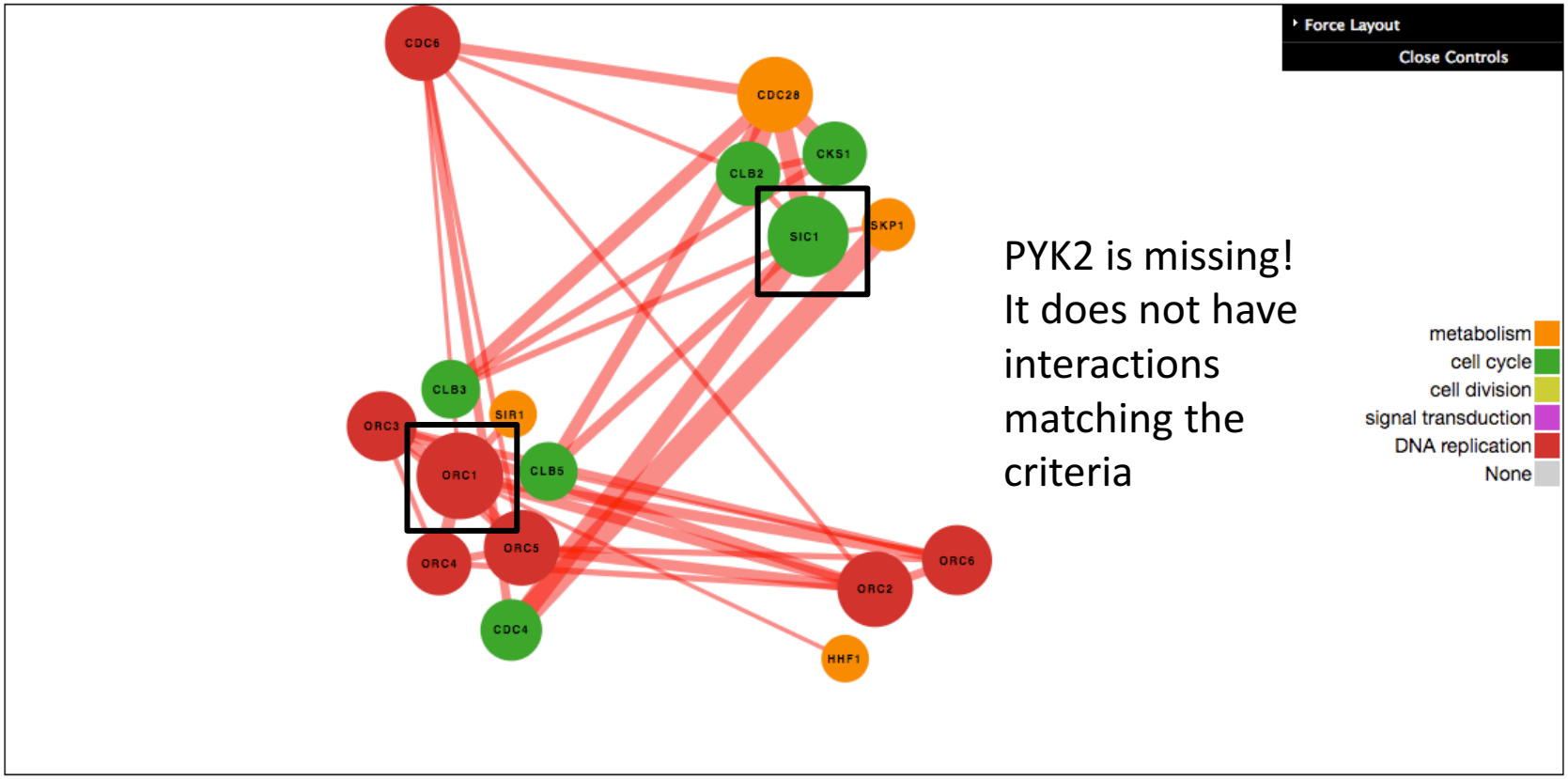
Example: 3 proteins: SIC1, ORC1 and PYK2

- GEMMER allows you to build a network with multiple proteins



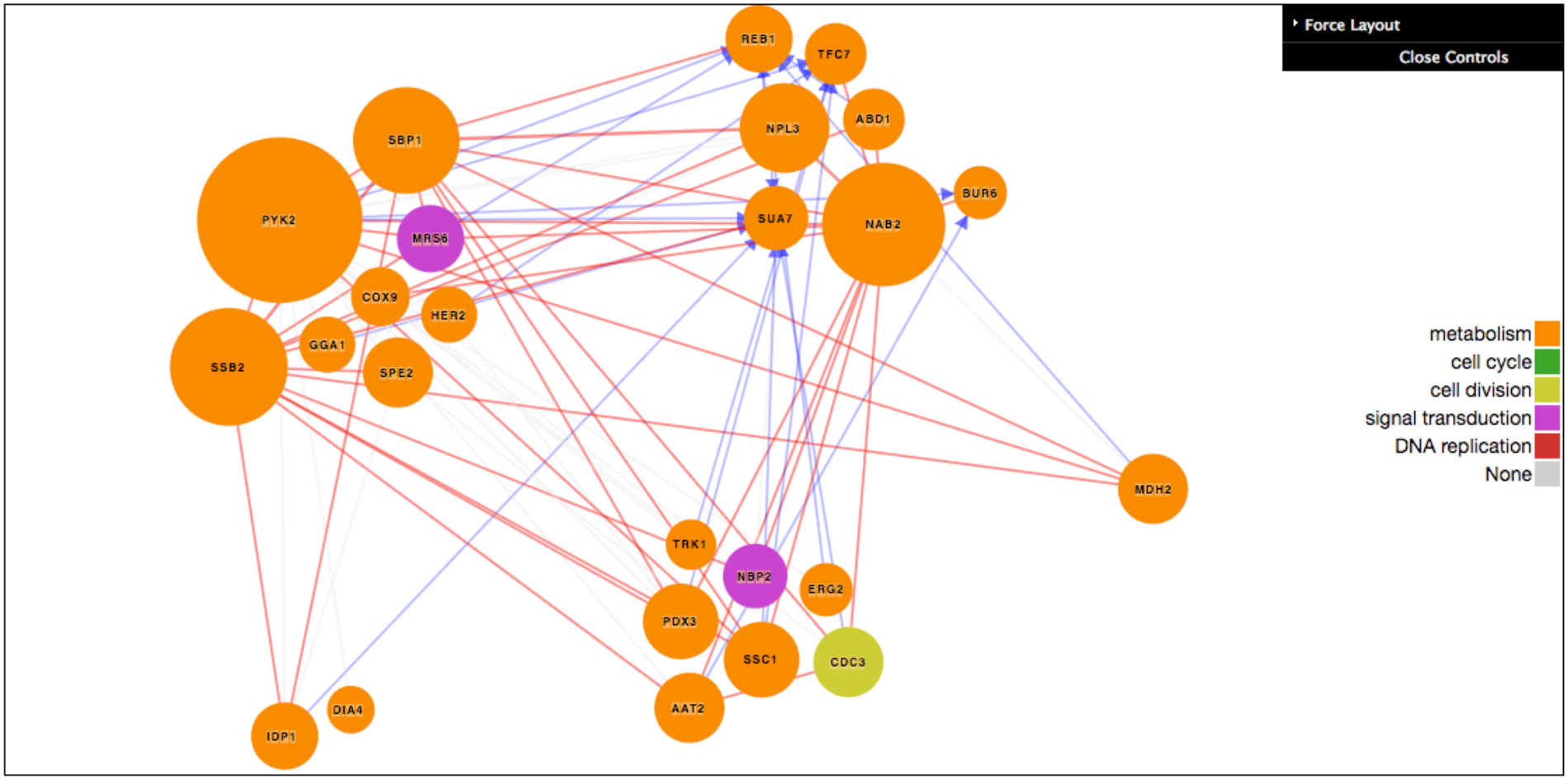
Example: 3 proteins: SIC1, ORC1 and PYK2

- GEMMER allows you to build a network with multiple proteins



Example: 3 proteins: SIC1, ORC1 and PYK2

- Lets just visualize PYK2 without limitations (all filters at 1)



Example: 3 proteins: SIC1, ORC1 and PYK2

- It did not show up previously because its interactions have limited validation! Compared to the rest of the network

Interactions

source	target	type	evidence	#experiments	#publications	#methods
AAT2	AAT2	physical	Reconstituted Complex (9655342)	1	1	1
AAT2	BUR6	regulation	chromatin immunoprecipitation-chip evidence (21329885)	1	1	1
AAT2	CDC3	physical	Affinity Capture-MS (11805837)	1	1	1
AAT2	NAB2	physical	Affinity Capture-RNA (19840948)	1	1	1
AAT2	PYK2	genetic	Positive Genetic (21623372)	1	1	1
AAT2	SSB2	physical	Co-purification (23332755)	1	1	1
ABD1	NAB2	physical	Affinity Capture-RNA (19840948)	1	1	1
ABD1	PYK2	genetic	Negative Genetic (20093466)	1	1	1
ABD1	REB1	regulation	computational combinatorial evidence (16522208)	1	1	1
ABD1	SSB2	physical	Co-purification (23332755)	1	1	1
BUR6	NAB2	physical	Affinity Capture-RNA (19840948)	1	1	1
CDC3	AAT2	physical	Affinity Capture-MS (11805837)	1	1	1
CDC3	CDC3	physical	Affinity Capture-Western (15282341)	1	1	1
CDC3	NAB2	physical	Affinity Capture-RNA (19840948)	1	1	1
CDC3	PYK2	genetic	Positive Genetic (20093466)	1	1	1
CDC3	SBP1	physical	Affinity Capture-RNA (23222640)	1	1	1
CDC3	SUA7	regulation	chromatin immunoprecipitation-chip evidence (21329885)	1	1	1

Example workflow for model-building

Still to come in GEMMER > 1.0

- Increase and improve clustering and coloring: Add GFP options, or ’None’ option
- Multiple layouts: more D3 layouts, hive plots etc.
- Generate mathematical analyses of the generated networks
- Implement of the periodic expression data from Tu et al. and the SCEPTRANS database
  - Dynamic visualization of level of expression during cell cycle progression
- Integration information from KEGG (especially for metabolic proteins)