

A complex network diagram with numerous grey circular nodes and thin grey lines connecting them, representing protein-protein interactions. The nodes are distributed across the slide, with some clusters and many long-range connections.

Protein-protein interactions and functional genomics

Pablo Minguez (pablo.minguez@embl.de)
Structural and Computational Biology Unit
EMBL Heidelberg
<http://www.bork.embl.de/~paniagua/>

Outline

- ✓ Protein-protein interactions
 - ✓ ppis are important
 - ✓ Resources
 - ✓ Detection methods
- ✓ Interactome (networks)
 - ✓ Description
 - ✓ Graph theory
- ✓ Types of biological networks
- ✓ Functional profiling using ppi data
 - ✓ Scenario
 - ✓ In-house interactomes
 - ✓ Tools @ Babelomics
 - ✓ Snow
 - ✓ NetworkMiner

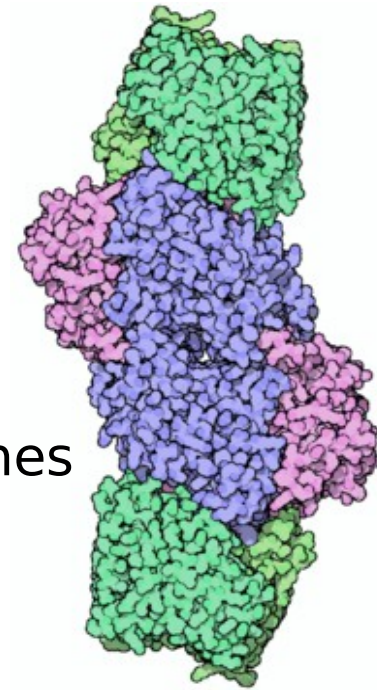
Outline

- ✓ **Protein-protein interactions**
 - ✓ ppis are important
 - ✓ Detection methods
 - ✓ Resources
- ✓ Interactome (networks)
 - ✓ Description
 - ✓ Graph theory
- ✓ Types of biological networks
- ✓ Functional profiling using ppi data
 - ✓ Scenario
 - ✓ In-house interactomes
 - ✓ Tools @ Babelomics
 - ✓ Snow
 - ✓ NetworkMiner

Ppis are a central point at almost every level of cell function

A ppi is a physical interactions between two proteins.

- ✓ **Transient (weak)** – eg. signalling cascades
 - ✓ **Stable (strong)** – eg. protein complexes
- **Structure** of subcellular organelles (structural proteins)
 - **Transport** machinery across the various biological membranes (nuclear pore importins)
 - **Packing** the chromatine (histones)
 - **Signal transduction** (important in many diseases, eg. cancer)
 - **Regulation** of gene expression (transcription factors)
 - **Protein modifications** (kinases)



Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions?

From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ BIND - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>

Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions?

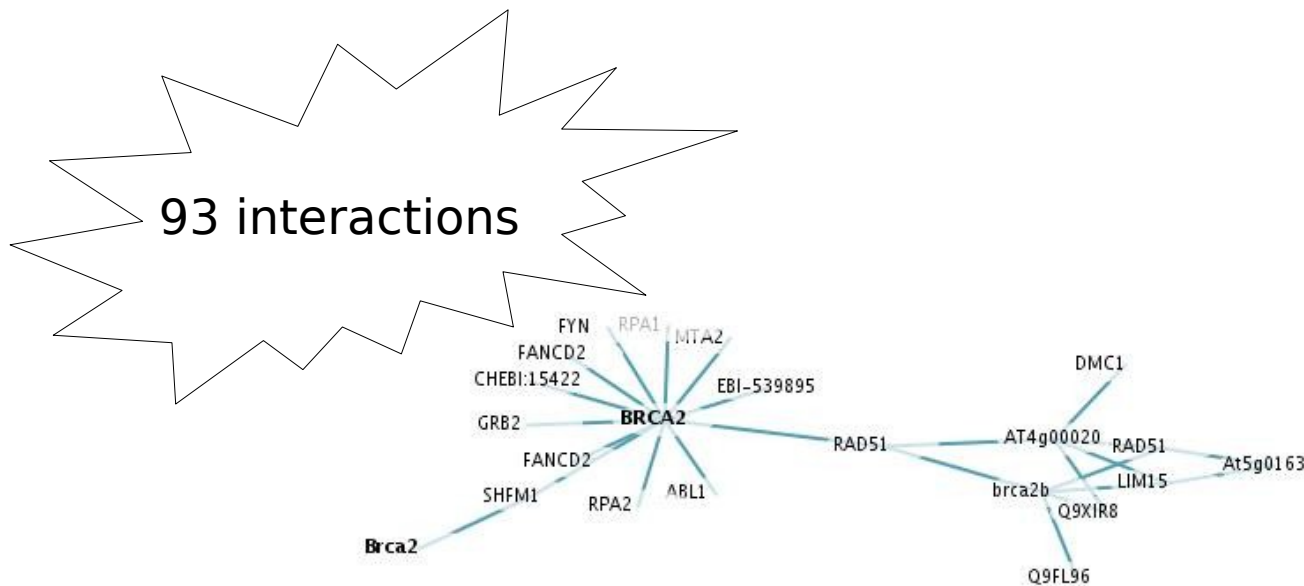
From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ BIND - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>

> 93 binary interactions were found in IntAct.
> Your query also matches 854 interaction evic
> Your query also matches 23 interaction evide

Previous 1-30 of 93 Next 30 | Expo

	Name molecule A	Links molecule A	Name molecule B
1	CG30169 PY UN	UniProt Dasty2	RecA protein homolog BA UN
2	Fanconi anemia group D1 protein NE IN	UniProt Dasty2	atp SM NE EN
3	NE IN		SM NE EN
4	NE IN		SM NE EN
5	NE IN		SM NE EN
6	NE IN		SM NE EN
7	Fanconi anemia	UniProt	Adapter protein GRB2



Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions?

From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ BIND - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>



24 interactions

PROTEIN INTERACTORS
BRCA1
CDC2
Cyclin A2
Cyclin dependent kinase 2
Cyclin E1
Fanconi anemia protein E
Fanconi anemia, complementation group D2
Filamin A
p53
RAD51
Replication factor A protein 1
DSS1
STAT5A
Polo like kinase
Mitotic checkpoint kinase MAD3L
SMAD3
High mobility group protein 20B
PCAF

Information about a single protein


BRCA2 is your favourite protein, where do you find information about its interactions?







From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ BIND - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>



6 interactions

 **BRCA2: Breast cancer type 2 susceptibility protein (P51587)**
6 partner(s) found in MINT.
Your query also matches **920** interaction(s) from other databases.

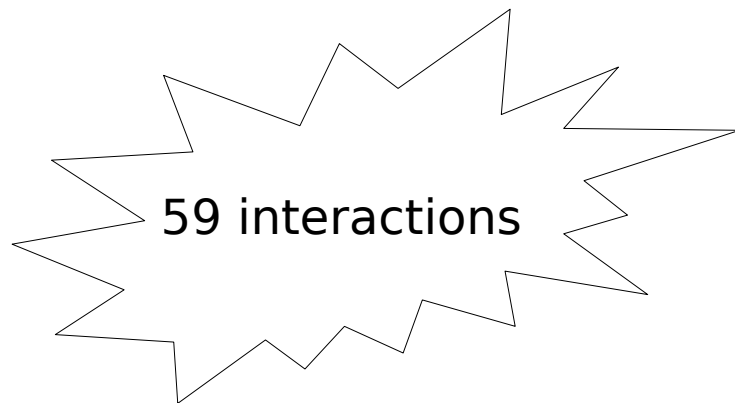
protein	evidences	score	direct.	ass.	coloc.	enz.	complex.	HT.
x  DMC1 Homo sapiens (Q14565)	11	0.93	9	2			1	
x  PALB2 Homo sapiens (Q86YC2)	4	0.82	3	1				
x  RAD51 Homo sapiens (Q06609)	7	0.71	4	3			1	
 ABL1 Homo sapiens (P00519)	1	0.28	1					1
 FYN Homo sapiens (P06241)	1	0.28	1					1
x  RAD23A Homo sapiens (P54725)	1	0.28		1				1

Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions?

From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ **BIND** - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>



59 interactions

Mol A Short Label	Mol B Short Label
FANCG	BRCA2
BRCA2	hsFLNa
FANCD2	BRCA2
BRCA2	RAD51
BRCA1	BRCA2
SMAD3	BRCA2
BRCA2	DSS1
CDK2/CCNA2	BRCA2
BRCA2	RAD51
BRCA2	RAD51
BRCA2	SHFDG1
SHFDG1	BRCA2
Brh2	Rad51
RAD51	RAD51
BRCA2	RAD51
RAB163	RAD51
Pol II	BRCA2 promoter

Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions?

From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ BIND - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>



5 interactions

Protein: BRCA2 protein				
Binary Complex				
DIP			Cross Reference	
Interaction	Interactor(s)	Links	PIR	SWISSPROT
DIP:57452E	DIP:29383N	●	---	---
DIP:40108E	DIP:462N	●	I58295	Q06609
DIP:40109E	DIP:368N	●	DNHU53	P04637
DIP:76301E	DIP:5971N	●	A58881	P38398
DIP:103802E	DIP:38427N	●	---	Q86YC2

Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions?

From:

- ✓ IntAct - <http://www.ebi.ac.uk/intact>
- ✓ HPRD - <http://www.hprd.org>
- ✓ MINT - <http://mint.bio.uniroma2.it>
- ✓ BIND - <http://bond.unleashedinformatics.com>
- ✓ DIP - <http://dip.doe-mbi.ucla.edu>
- ✓ BioGrid - <http://thebiogrid.org>



37 interactions

Switch View: **Summary** Sortable Table

Displaying **37** total unique interactors

RAD51 | HRAD51, RECA, BRCC5, HsT16930, HsRad51, RAD51A

RAD51 homolog (RecA homolog, E. coli) (*S. cerevisiae*)

HMG20B | SMARCE1r, HMGXB2, FLJ26127, SOXL, PP7706, HMGX2, BRAF25, BRAF35, pp8857

SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E, member 1-related

BRCA1 | RNF53, IRIS, BRCC1, PSCP, PNCA4, BRCAI, BROVCA1

breast and ovarian cancer susceptibility protein 1

State of art of ppi databases

A Overlap of ppis

HPRD (36,617)									
BIND (6,621)	4,903								
DIP (1,067)	801	264							
MINT (11,367)	8,690	1463	379						
Reactome (5,960)	538	207	67	102					
IntAct (10,244)	8,031	1167	283	7,362	173				
MIPS (346)	307	294	28	65	14	43			
PDZ Base (101)	93	19	0	60	0	5	3		
	HPRD (36,617)	BIND (6,621)	DIP (1,067)	MINT (11,367)	Reactome (5,960)	IntAct (10,244)	MIPS (346)	PDZ Base (101)	

B

Overlap of proteins

HPRD (9,427)									
BIND (3,887)	3,414								
DIP (804)	755	537							
MINT (4,975)	4,719	2218	562						
Reactome (970)	733	453	164	497					
IntAct (4,614)	4,421	1969	473	3795	497				
MIPS (405)	396	390	146	303	78	262			
PDZ Base (115)	114	64	10	99	1	54	16		
	HPRD (9,427)	BIND (3,887)	DIP (804)	MINT (4,975)	Reactome (970)	IntAct (4,614)	MIPS (405)	PDZ Base (115)	

No one has a complete coverage of the known ppis
(neither of the complete interactome)

Information about a single protein

BRCA2 is your favourite protein, where do you find information about its interactions? **Choose your favourite!!**

Switch View: Summary Sortable Table

Displaying **37** total unique interactors

RAD51 | HRAD51, RECA, BRCC5, HsT169
RAD51 homolog (RecA homolog, E. coli) (S. ce

PROTEIN INTERACTORS

> **93** binary interactions were found in IntAct.
> Your query also matches **854** interaction evic
> Your query also matches **23** interaction evide

Previous 1-30 of 93 Next 30 | Expo

	Name molecule A	Links molecule A	Name molecule B
1	CG30169 		RecA protein homolog

Protein: BRCA2 protein

Binary Comp

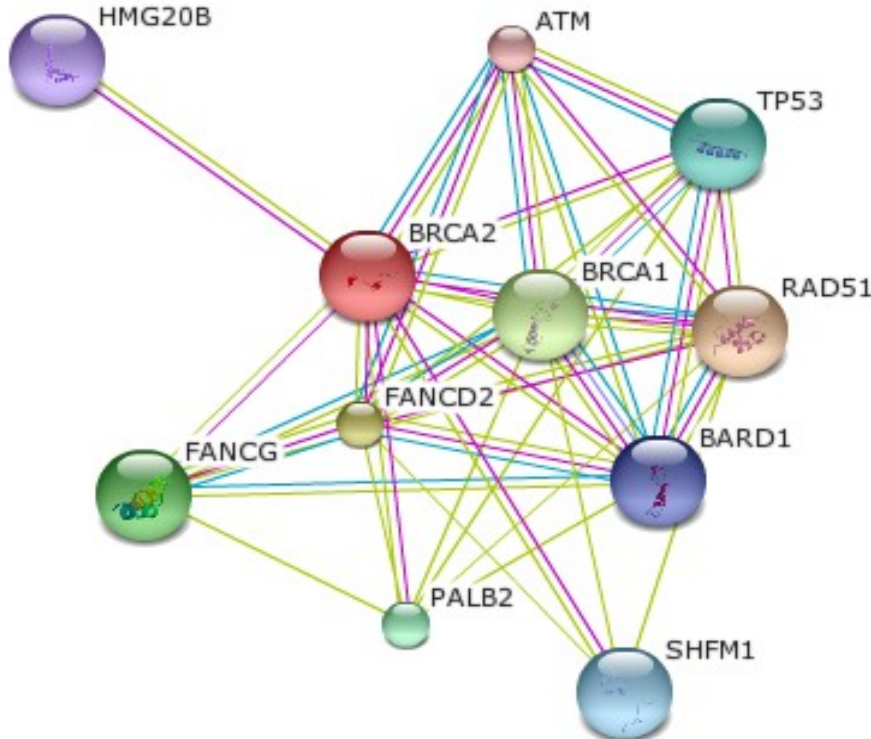
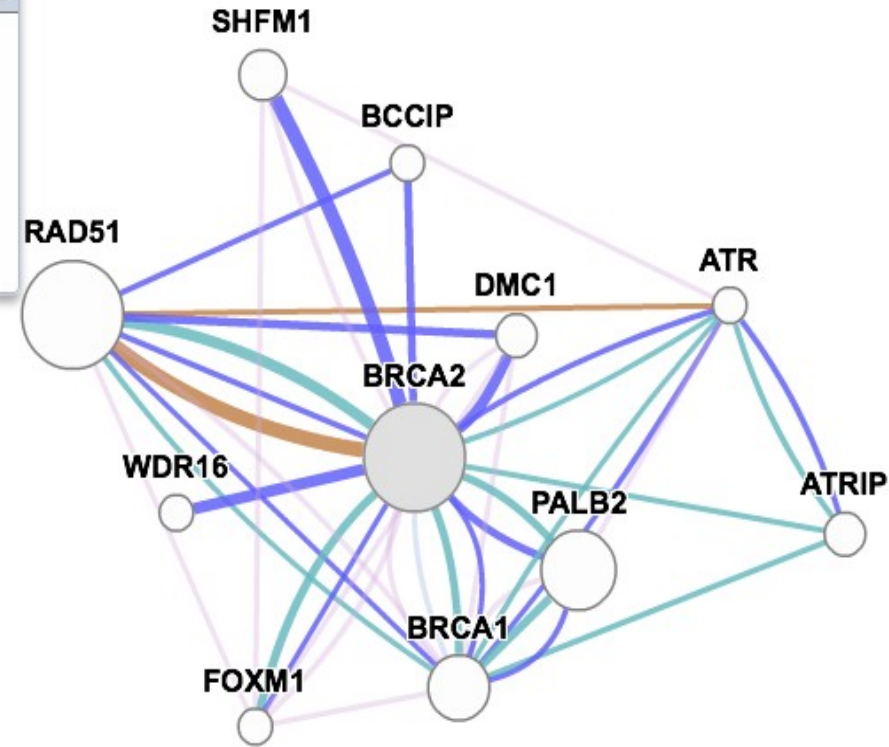
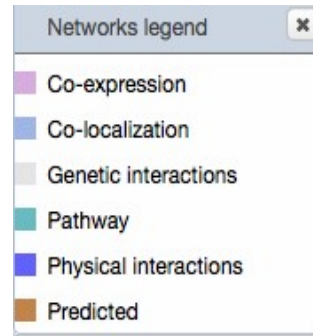
BRCA2: Breast cancer type 2 susceptibility protein (P51587)
6 partner(s) found in MINT.
Your query also matches **920** interaction(s) from other databases.

Interactions	protein	evidences	score	direct.	ass.	coloc.	enz.	complex.	HT.
DIP:57452E	DMC1 Homo sapiens (Q14565)	11	0.93	9	2			1	
DIP:40108E	PALB2 Homo sapiens (Q86YC2)	4	0.82	3	1				
DIP:40109E	RAD51 Homo sapiens (Q06609)	7	0.71	4	3			1	
DIP:76301E	ABL1 Homo sapiens (P00519)	1	0.28	1					1
DIP:103802E	FYN Homo sapiens (P06241)	1	0.28	1					1
DIP:38427N	RAD23A Homo sapiens (P54725)	1	0.28		1				1

Information about a single protein

Some resources that collect (and predict) interactions

- ✓ STRING - <http://string-db.org/>
- ✓ GeneMania - <http://genemania.org>
- ✓ APID - <http://bioinfow.dep.usal.es/apid>
- ✓ GeneCards - <http://www.genecards.org>



Your Input:

● BRCA2 Breast cancer susceptibility protein 2 (Homo sapiens)

Predicted Functions

● RAD51 DNA repair protein Rad51 (Homo sapiens)
 ● FANCD2 Fanconi anemia complementation group D2 (Homo sapiens)
 ● BRCA1 Breast cancer susceptibility protein 1 (Homo sapiens)
 ● FANCG Fanconi anemia complementation group G (Homo sapiens)
 ● PALB2 Partner of BRCA1 and BRCA2 (Homo sapiens)
 ● TP53 Cellular tumor protein p53 (Homo sapiens)
 ● SHFM1 SHFM1 (Homo sapiens)

); Involved in double-strand
 checkpoint activation (3418 aa)

Neighbor	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
----------	-------------	--------------	--------------	-------------	-----------	------------	------------	-------

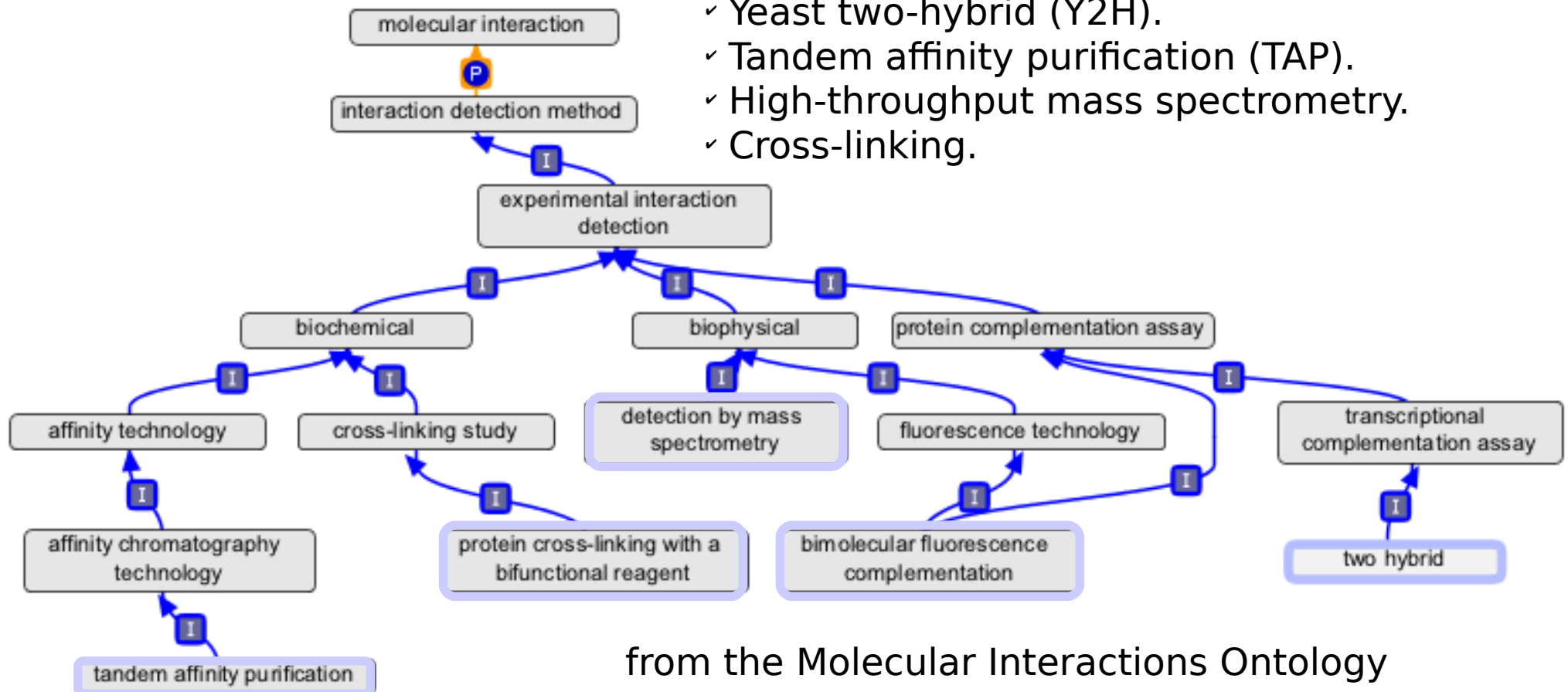
in a common DNA damage r [...] (340 aa)				●	●	●		0.999
: of chromosomal stabil [...] (1471 aa)				●	●	●		0.999
central role in D [...] (1863 aa)				●	●	●		0.999
DNA repair protein th [...] (622 aa)				●	●	●		0.998
a localization and s [...] (1188 aa)				●	●	●		0.996
ntigen NY-CO-13); Acts [...] (393 aa)				●	●	●		0.996

Techniques to explore protein-protein interactions

Non-screening techniques: Co-immunoprecipitation, fluorescence resonance energy transfer, Dual polarization interferometry.

But most of the data we find in the databases come from **high-throughput techniques:**

- ✓ Bio-molecular fluorescence complementation.
- ✓ Yeast two-hybrid (Y2H).
- ✓ Tandem affinity purification (TAP).
- ✓ High-throughput mass spectrometry.
- ✓ Cross-linking.

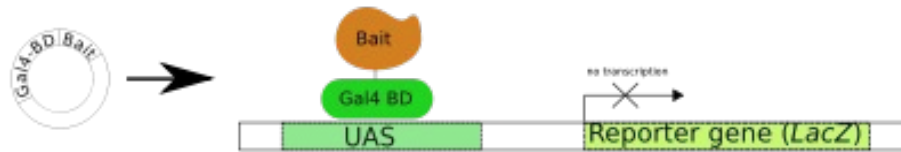


from the Molecular Interactions Ontology

Yeast two hybrid experiments



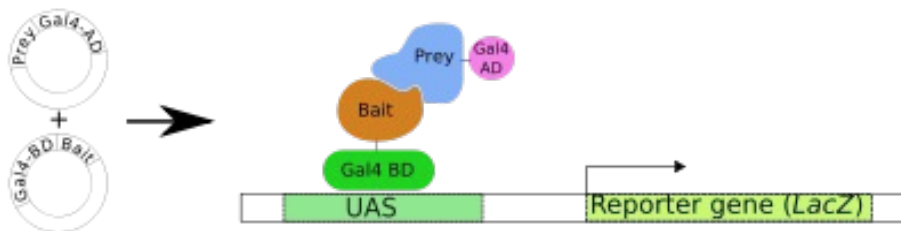
A. Regular transcription of the reporter gene



B. One fusion protein only (Gal4-BD + Bait) - no transcription



C. One fusion protein only (Gal4-AD + Prey) - no transcription



D. Two fusion proteins with interacting Bait and Prey

A. Gal4 transcription factor gene produces two domain protein (BD and AD), which is essential for transcription of the reporter gene (LacZ).

B,C. Two fusion proteins are prepared: Gal4BD+Bait and Gal4AD+Prey. None of them is usually sufficient to initiate the transcription (of the reporter gene) alone.

D. When both fusion proteins are produced and Bait part of the first interact with Prey part of the second, transcription of the reporter gene occurs.

If bait catch the prey (interaction) a reporter gene is expressed.

State of art of ppi interaction data

Issues with ppi data:

- ✓ High-throughput experiments **do not overlap**.
 - ✓ Methods do not reach **saturation** point.
 - ✓ Lack of **accuracy** or coverage of some of them.
- ✓ **False positives**, non specific interactions (y2h, 50%).
- ✓ **Bias** in the functional categories each of the methods detect.
 - ✓ datasets based on purified complexes predict relatively few interactions for proteins involved in transport.
 - ✓ y2h fails in detecting proteins involved in translation.
 - ✓ y2h cannot detect interactions requiring three or more proteins or those depending on post-translational modifications.
 - ✓ Interactions occurring outside the nucleus are not detected by y2h.

Ppi annotation

Standard format to submit ppis

PSI-MI: HUPO Proteomics Standards Initiative: <http://www.psidev.info/>

Defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification.

Controlled Vocabulary: Molecular Interactions Ontology

[Term]

id: **MI:0018**

name: **two hybrid**

def: "The classical two-hybrid system is a method that uses transcriptional activity as a measure of protein-protein interaction. It relies on the modular nature of many site-specific transcriptional activators (GAL 4) , which consist of a DNA-binding domain and a transcriptional activation domain. The DNA-binding domain serves to target the activator to the specific genes that will be expressed, and the activation domain contacts other proteins of the transcriptional machinery to enable transcription to occur. The two-hybrid system is based on the observation that the two domains of the activator need to be non-covalently brought together by the interaction of any two proteins. The application of this system requires the expression of two hybrid. Generally this assay is performed in yeast cell, but it can also be carried out in other organism."

[PMID:10967325, PMID:12634794, PMID:1946372]

related_synonym: "2-hybrid" []

related_synonym: "2H" []

related_synonym: "2h" []

related_synonym: "classical two hybrid" []

related_synonym: "Gal4 transcription regeneration" []

related_synonym: "two-hybrid" []

related_synonym: "yeast two hybrid" []

exact_synonym: "2 hybrid" []

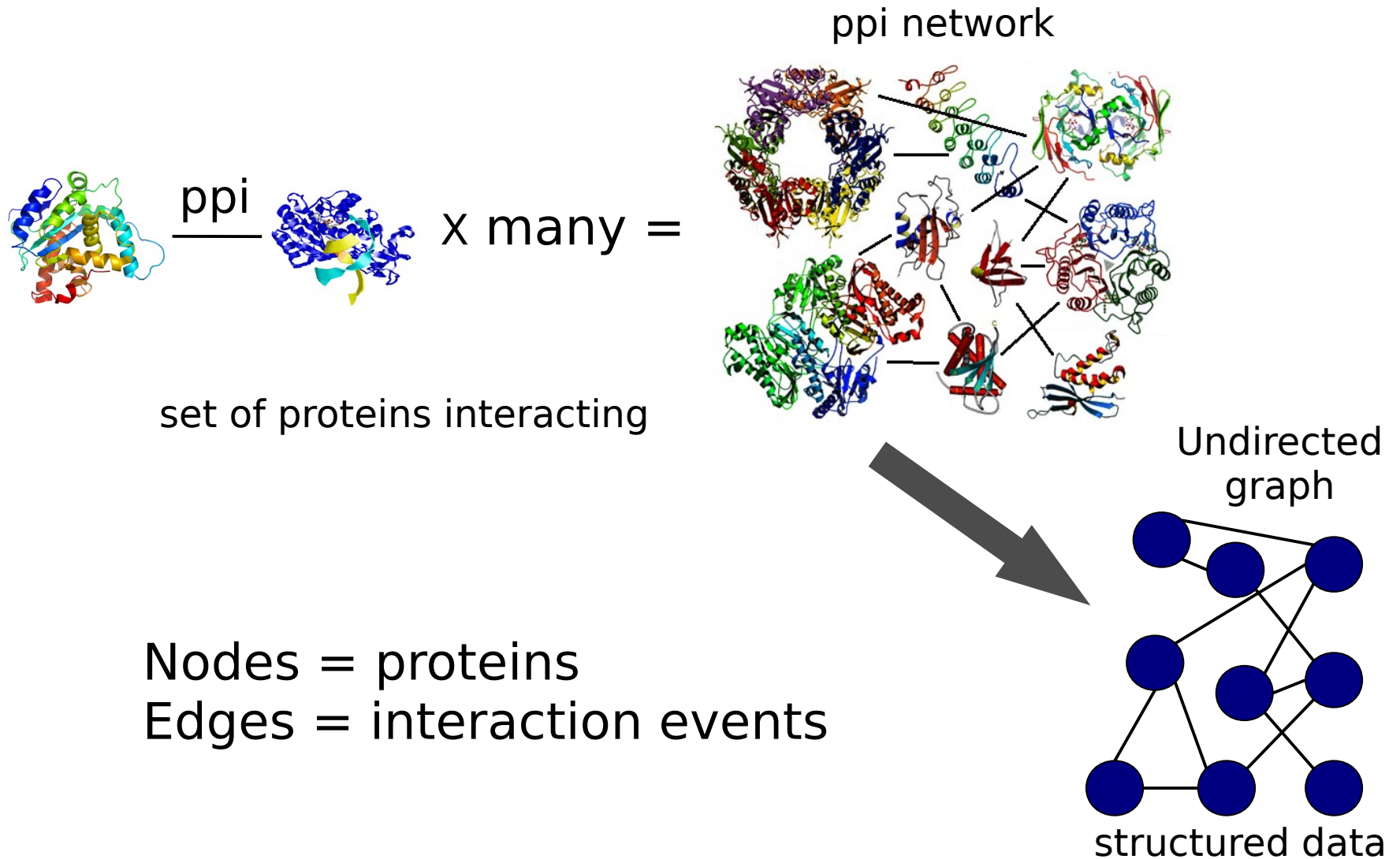
is_a: **MI:0232 ! transcriptional complementation assay**

Outline

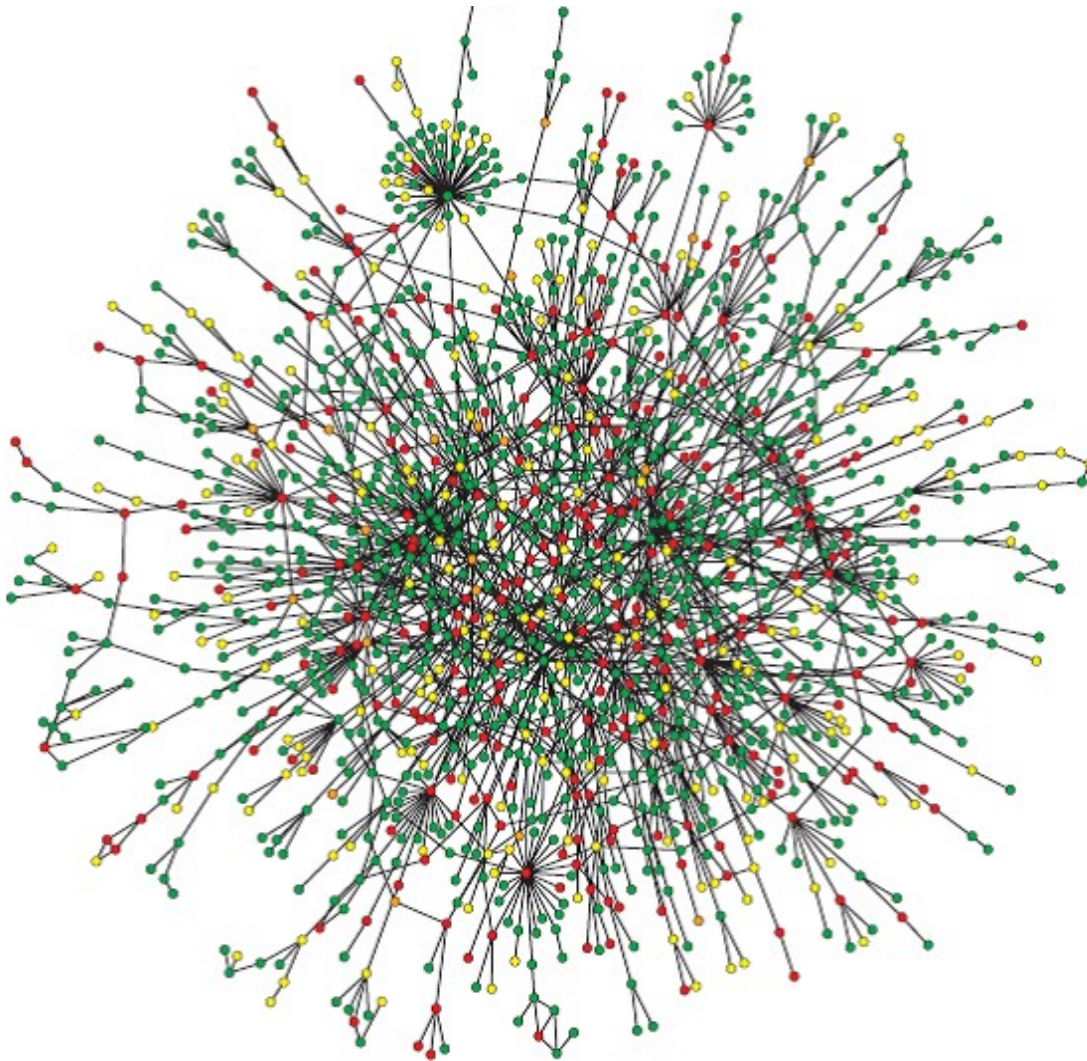
- ✓ Protein-protein interactions
 - ✓ ppis are important
 - ✓ Resources
 - ✓ Detection methods
- ✓ **Interactome (networks)**
 - ✓ Description
 - ✓ Graph theory
- ✓ Types of biological networks
- ✓ Functional profiling using ppi data
 - ✓ Scenario
 - ✓ In-house interactomes
 - ✓ Tools @ Babelomics
 - ✓ Snow
 - ✓ NetworkMiner

The interactome

Ppis are defined by pairwise interactions that leads to a complete net of interactions, the interactome.



The interactome

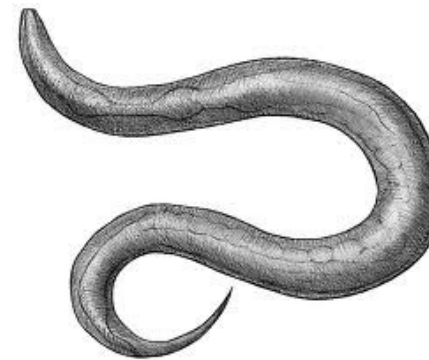


The interactions are the real source of complexity of the cell.



20,935 protein coding genes

650,000 predicted interactions

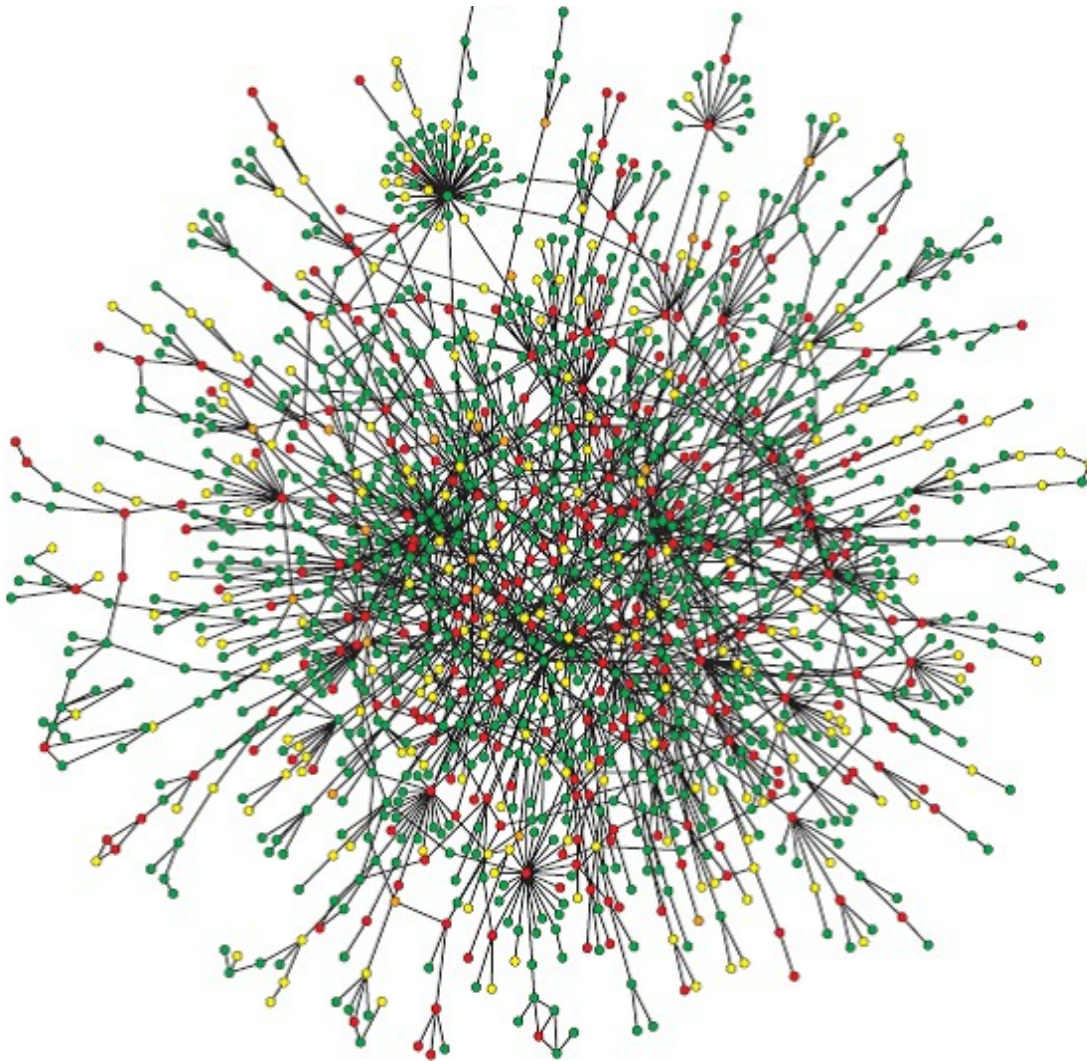


20,389 protein coding genes

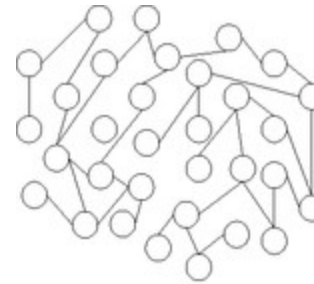
240,000 predicted interactions

So, ppi information can be the clue to explain your phenotype!

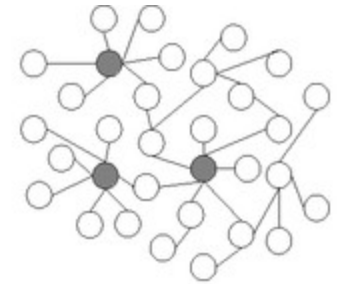
The interactome and graph theory



Scale-free network instead of random network



(a) Random network



(b) Scale-free network

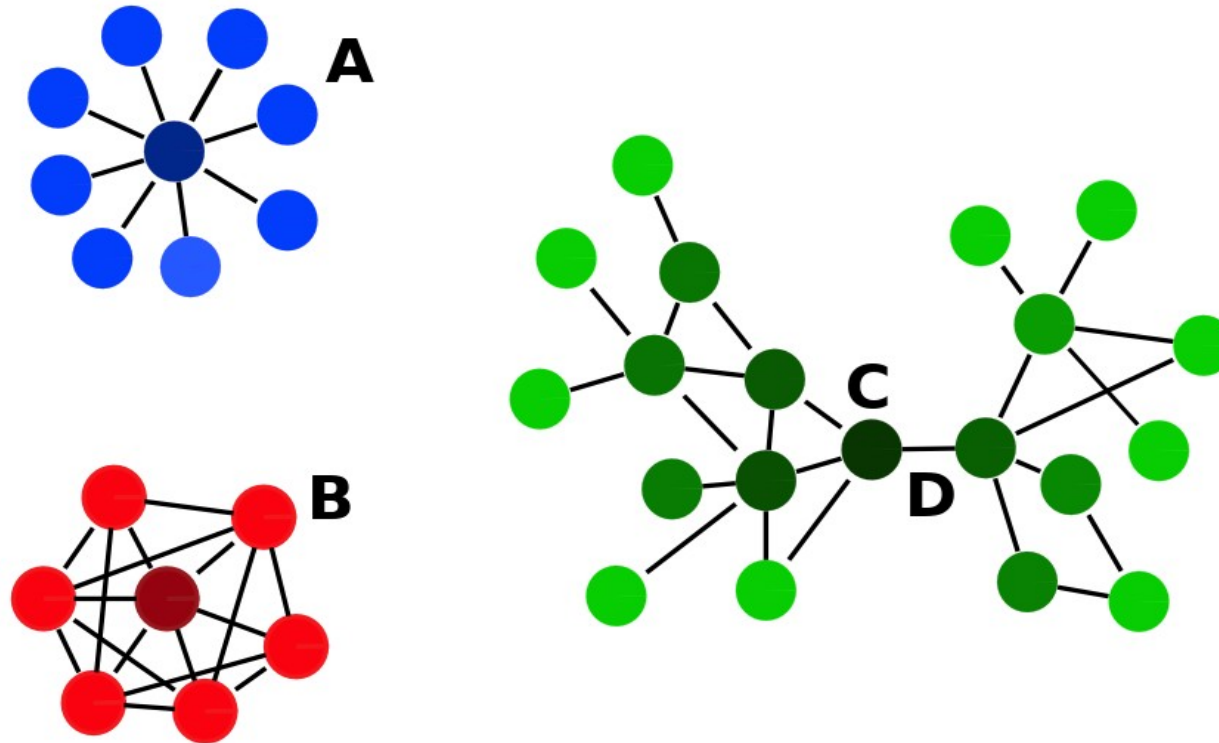
They are defined by a connections degree, number of connections of a node, distribution that approximates to a power law

$$P(k) \sim ck^{-\gamma}$$

Being $2 < \gamma < 3$

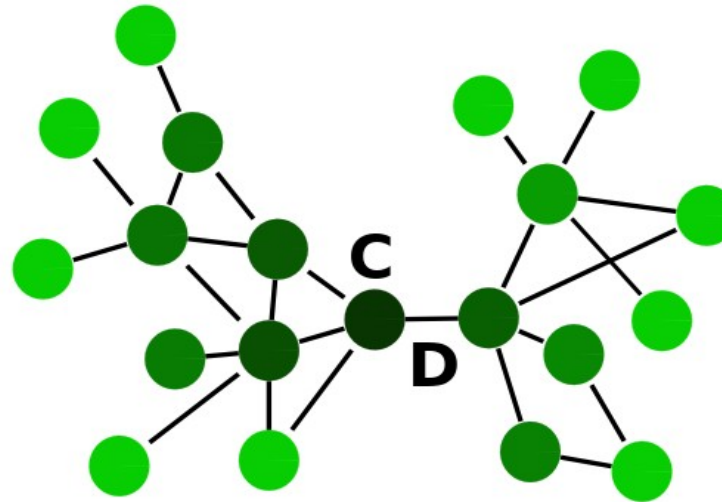
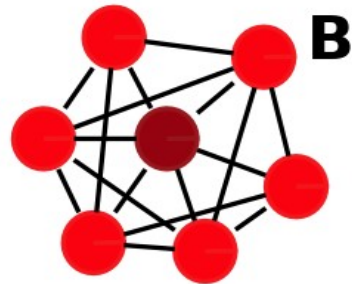
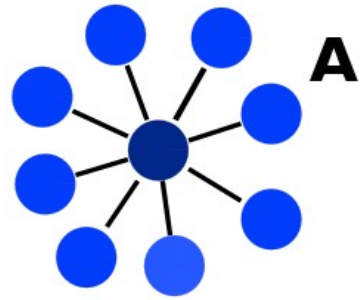
This indicates that the network has a low number of highly connected nodes. There are a few proteins that connect the whole interactome.

Graph theory - topological parameters



Degree (connectivity or connections): number of edges connected to a node. Nodes with high degree are called hubs.

Graph theory - topological parameters



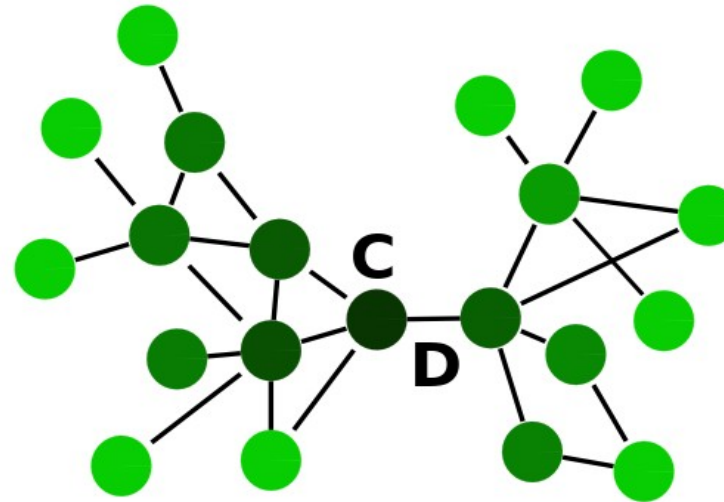
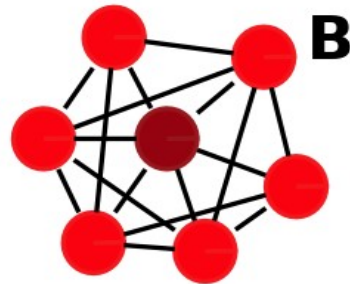
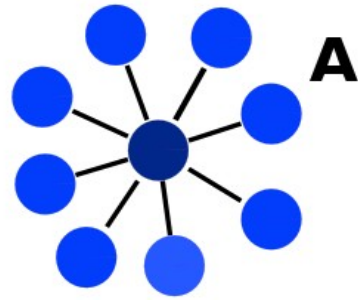
Betweenness: A measure of centrality of a node, it is defined by:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$\sigma_{st}(\mathbf{V})$ is the number of shortest paths that pass through node V

σ_{st} is total number of shortest paths in the graph.

Graph theory - topological parameters

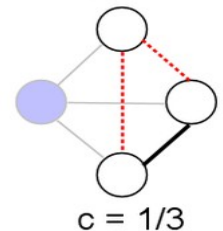


Clustering coefficient (of a node): A measure of how interconnected the neighbours of that node are.

Proportion of links between the nodes within its neighbourhood divided by the number of links that could possibly exist between them.

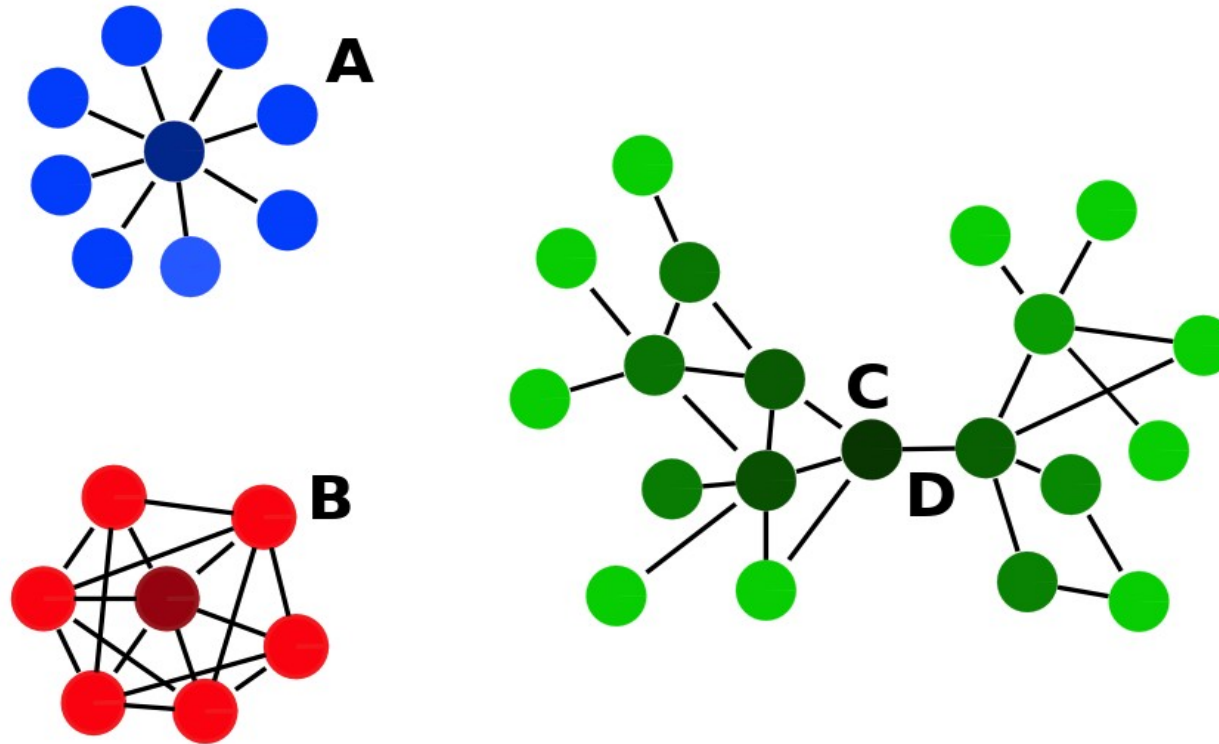
$$C_i = \frac{2e_i}{n_i(n_i - 1)}$$

e_i is the number of edges among the nodes connected to node i
 n_i is the number of neighbours of node i



To differentiate between star-shaped nets and more interconnected nets.

Graph theory - topological parameters



Shortest path: The path with less edges that connects two nodes.

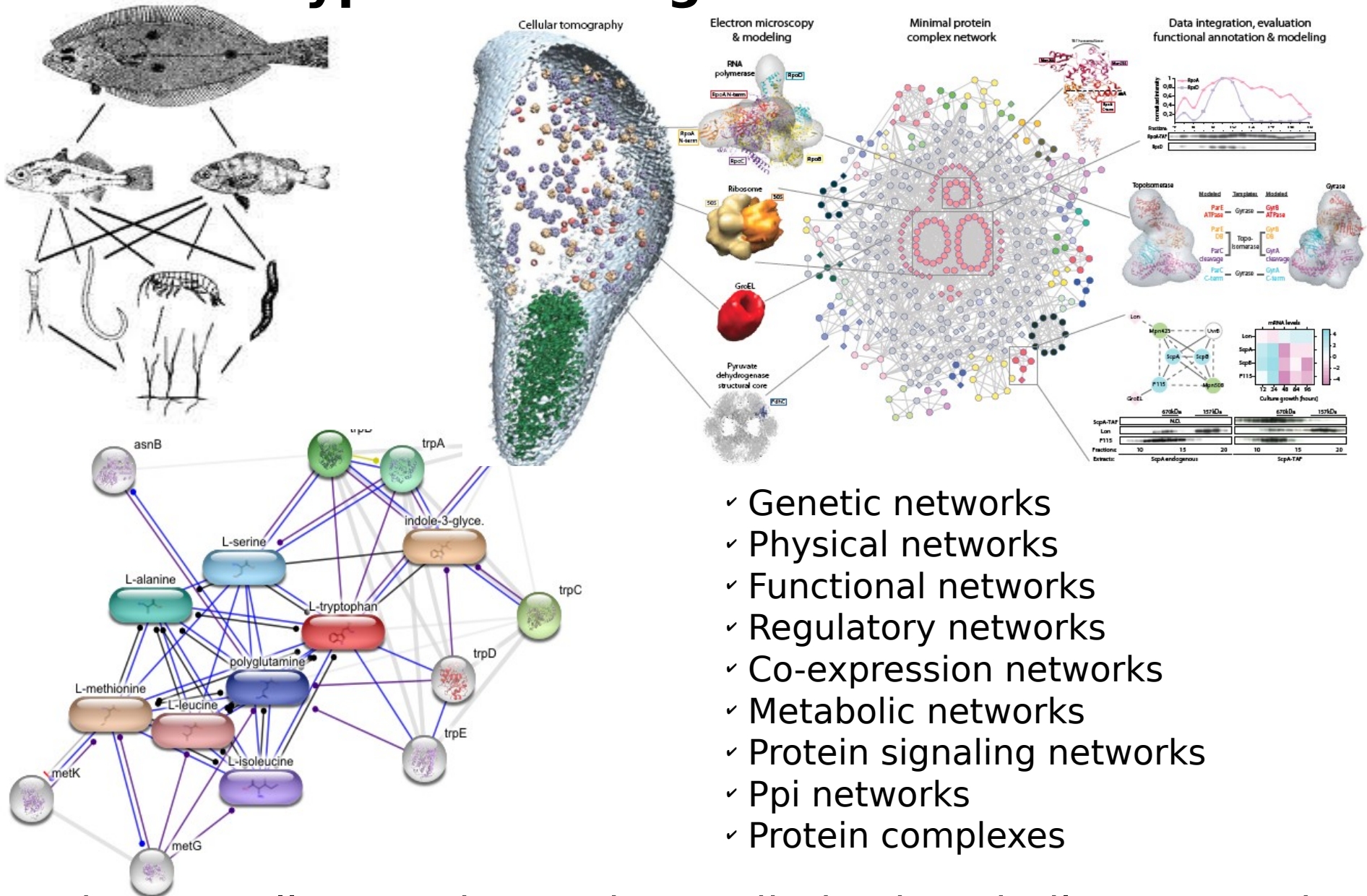
Component: A group of nodes connected among them.

Bicomponent: A group of nodes connected to other group of nodes by only an edge. The edge that joins two bicomponents is called articulation point.

Outline

- ✓ Protein-protein interactions
 - ✓ ppis are important
 - ✓ Resources
 - ✓ Detection methods
- ✓ Interactome (networks)
 - ✓ Description
 - ✓ Graph theory
- ✓ **Types of biological networks**
- ✓ Functional profiling using ppi data
 - ✓ Scenario
 - ✓ In-house interactomes
 - ✓ Tools @ Babelomics
 - ✓ Snow
 - ✓ NetworkMiner

Types of biological networks



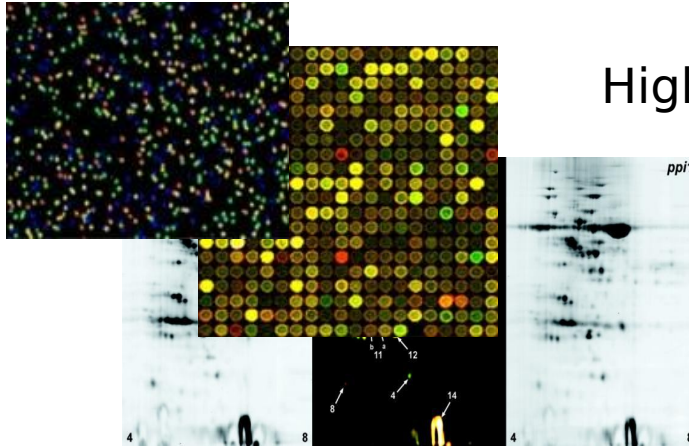
- ✓ Genetic networks
- ✓ Physical networks
- ✓ Functional networks
- ✓ Regulatory networks
- ✓ Co-expression networks
- ✓ Metabolic networks
- ✓ Protein signaling networks
- ✓ Ppi networks
- ✓ Protein complexes

Good news! All networks can be studied using similar approaches

Outline

- ✓ Protein-protein interactions
 - ✓ ppis are important
 - ✓ Resources
 - ✓ Detection methods
- ✓ Interactome (networks)
 - ✓ Description
 - ✓ Graph theory
- ✓ Types of biological networks
- ✓ **Functional profiling using ppi data**
 - ✓ Scenario
 - ✓ In-house interactomes
 - ✓ Tools @ Babelomics
 - ✓ Snow
 - ✓ NetworkMiner

Functional profiling of genome scale experiments using ppi data



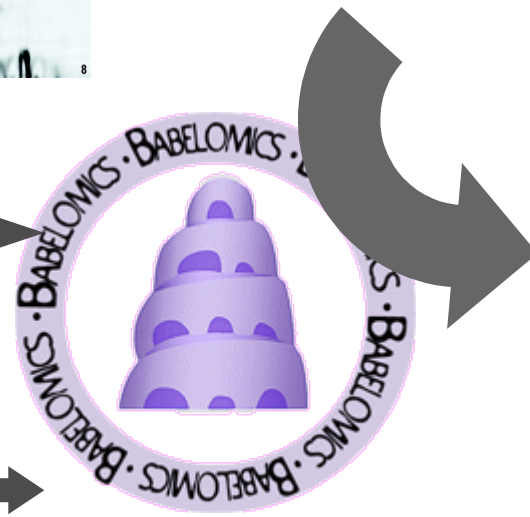
High-throughput experiment

- ✓ Gene selection: clustering, differential expression
- ✓ Gene sorting: differential expression

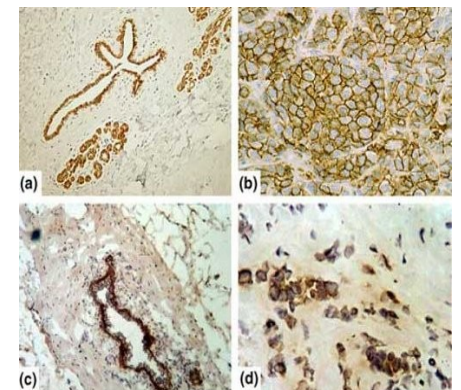
- ✓ GO terms
- ✓ KEGG pathways
- ✓ Interpro motifs
- ✓ ...

- ✓ Text Mining bio-entities
- ✓ Expression in tissues profiles

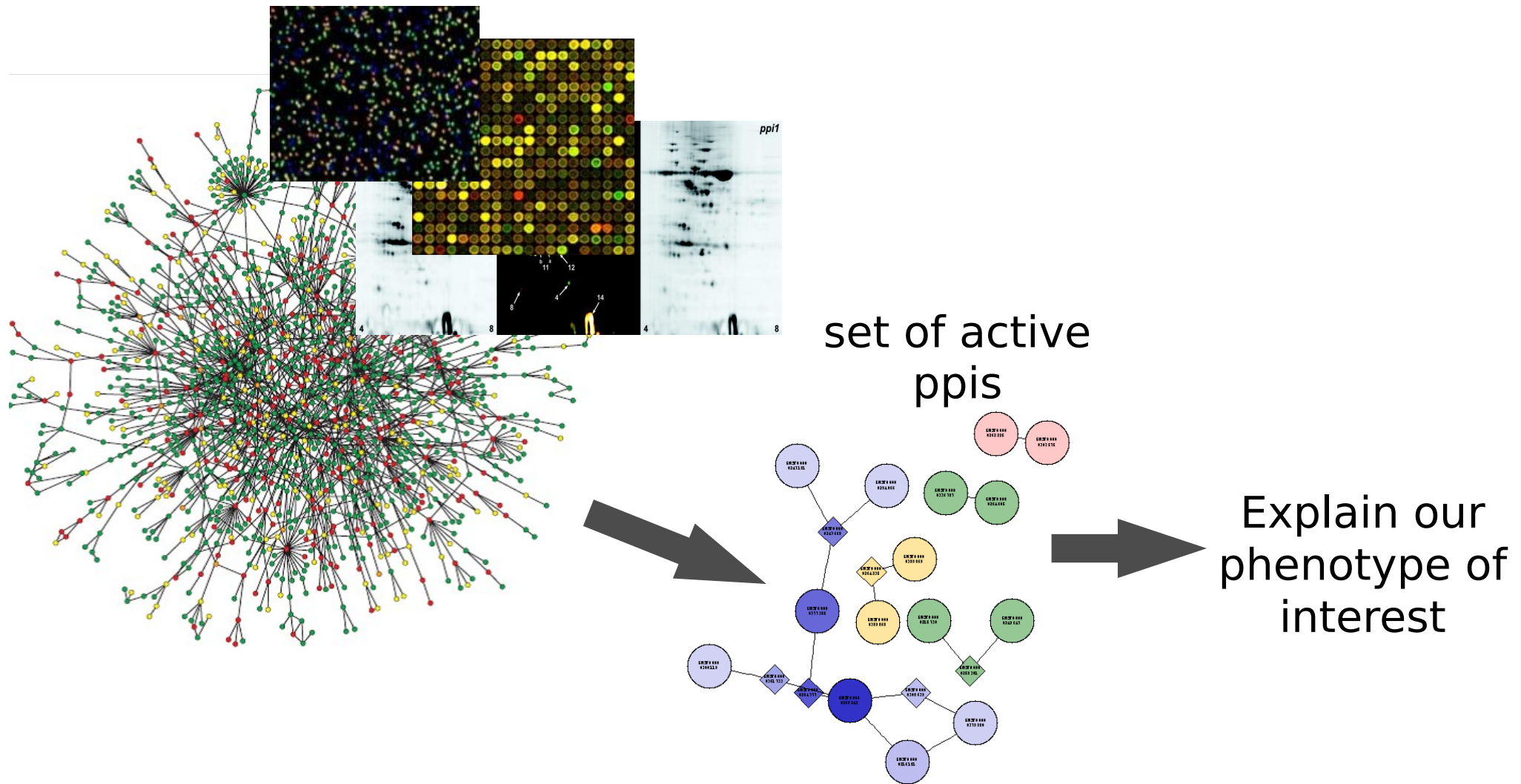
Protein-protein interactions



Functional profiling



Functional profiling of genome scale experiments using ppi data

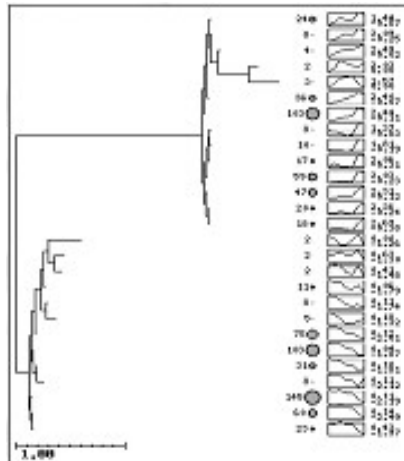


Interactome: Complete collection of protein-protein interactions in the cell.
Transcriptome determines the real interactome.

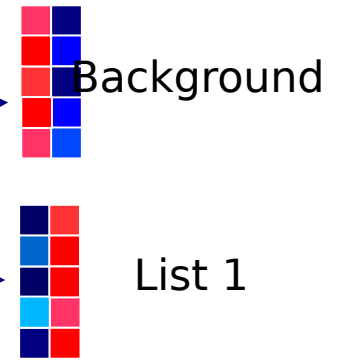
Functional profiling - Methods scenario

1

Find interesting patterns of expression (clustering)

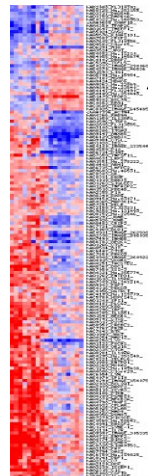


**Two steps methods
(functional classes enrichment)**



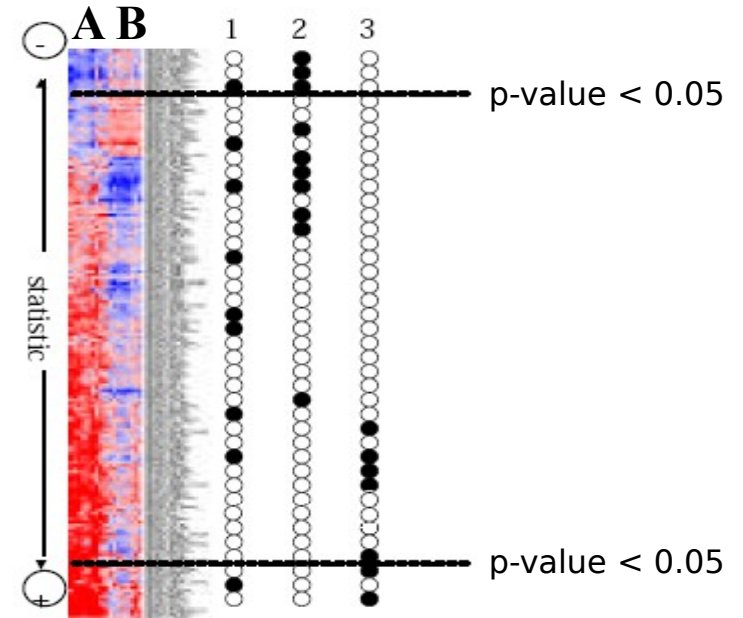
2

Get differentially expressed A B genes



p-value

**Threshold free methods
(gene set enrichment)**



Network-based tools @ Babelomics

**Two steps
methods
(functional
classes
enrichment)**



snow

<http://babelomics.bioinfo.cipf.es/snow.html>
<http://beta.babelomics.bioinfo.cipf.es/snow.html>
<http://snow.bioinfo.cipf.es>

**Threshold
free
methods
(gene set
enrichment)**



Network Miner

Finding the hidden network component in a ranked list of genes

<http://networkminer.bioinfo.cipf.es/>

Our interactomes

- Summary of **non redundant interactions** per species:

Type	Group	Ids	Ath	Bta	Dme	Dre	Eco	Hsa	Mmu	Sce
Physical	All	Proteins	6708	293	37528	177	14194	82909	7238	86646
Physical	All	Genes	6184	260	26733	126	13889	77966	7026	86271
Physical	Curated	Proteins	1590	21	20966	12	1644	21655	1080	32099
Physical	Curated	Genes	1512	18	16332	8	1638	21336	1056	32070

- Summary of **non redundant interactors** per species:

Type	Group	Ids	Ath	Bta	Dme	Dre	Eco	Hsa	Mmu	Sce
Physical	All	Proteins	3214	234	10208	109	2887	12119	4111	6093
Physical	All	Genes	2954	209	7723	72	2816	10782	3781	6021
Physical	Curated	Proteins	1251	34	7239	16	729	7333	1164	5394
Physical	Curated	Genes	1177	29	5989	11	725	7150	1122	5376

Abbreviation	Species
Ath	<i>Arabidopsis thaliana</i>
Bta	<i>Bos taurus</i>
Dre	<i>Danio rerio</i>
Dme	<i>Drosophila melanogaster</i>
Eco	<i>Escherichia coli (strain K12)</i>
Hsa	<i>Homo sapiens</i>
Mmu	<i>Mus musculus</i>
Sce	<i>Saccharomyces cerevisiae</i>

Databases: IntAct, MINT, BioGrid

Two interactomes per species: All interactions, Curated (ppis annotated with two different techniques).

Genes have a one to many relationship with proteins. There is not a real interactome for genes but our users usually come with gene lists.

Mapping genes (many potential proteins) into a proteins interactome may give a fake disperse network.

Snow

Studying networks in the omic work

Goal

To develop a methodology that may extract from lists of proteins/genes the ppi networks acting and evaluates whether they have importance in the cooperative behaviour of the list.

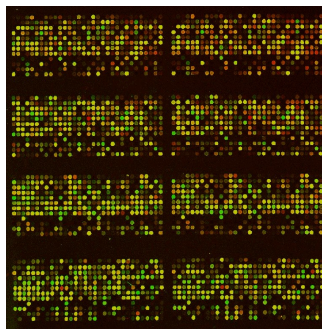
How we evaluate the cooperative behaviour of a list of proteins/genes in terms of its ppi network parameters?

Two different approximations

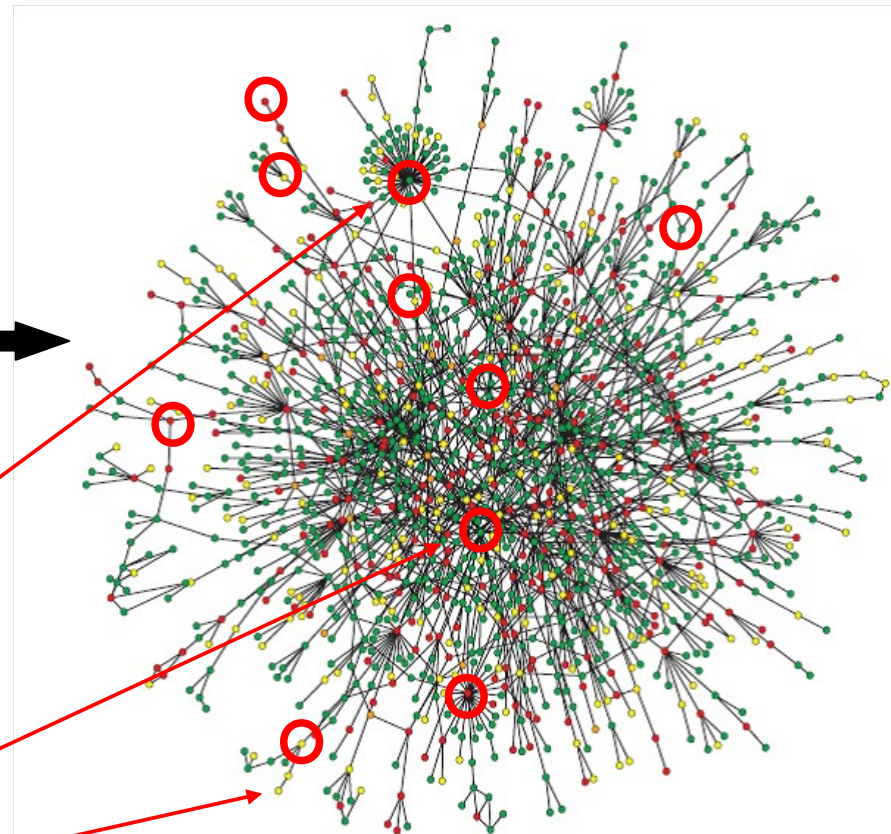
- ✓ Role in the complete interactome.
- ✓ Cooperative behaviour (Minimal Connected Network)

Role in the interactome

The list of proteins mapped into the complete interactome may provide clues about the importance of certain proteins.



M22382
Y00371
X15183
M90054
AA650377
U78027
U14972
M17887
M20472
AB015320



Hubs: nodes with high degree

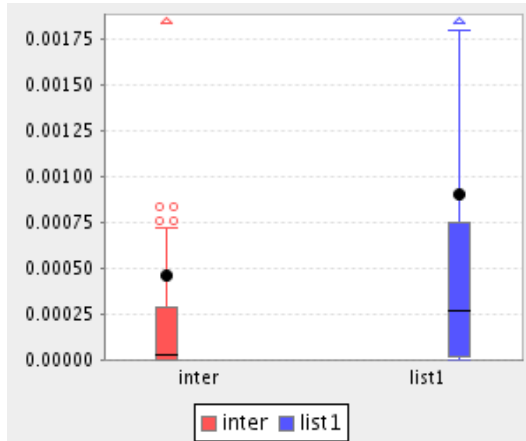
Nodes very central (high betweenness)

External proteins

Role in the interactome

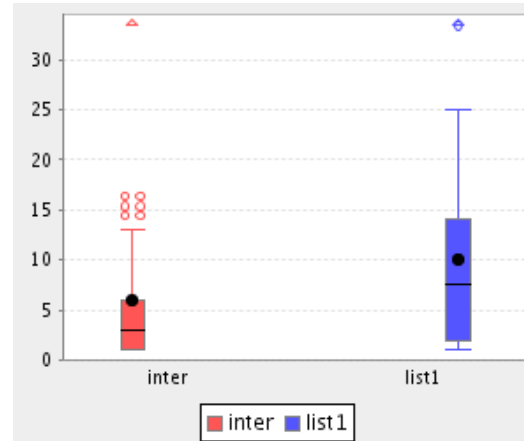
Comparison of parameters distributions of the lists versus complete interactome distributions applying a Kolmogorov-Smirnov test.

betweenness



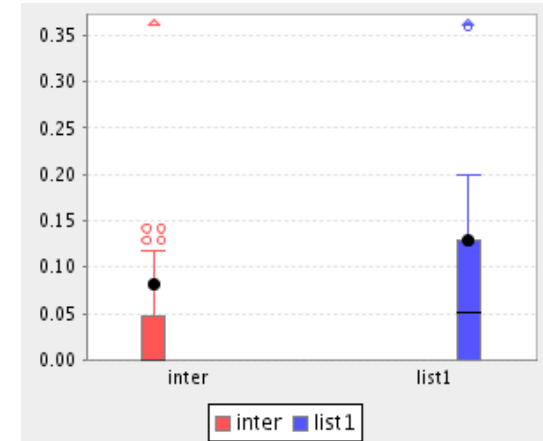
pval = 0.0058

connections degree



pval = 0.0013

clustering coefficient



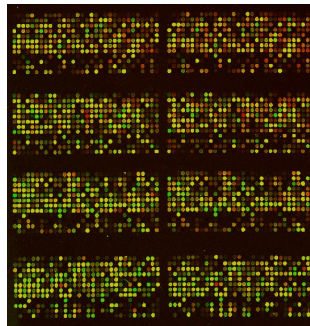
pval = 0.0055

Results indicate whether the set of prots/genes are collectively ...

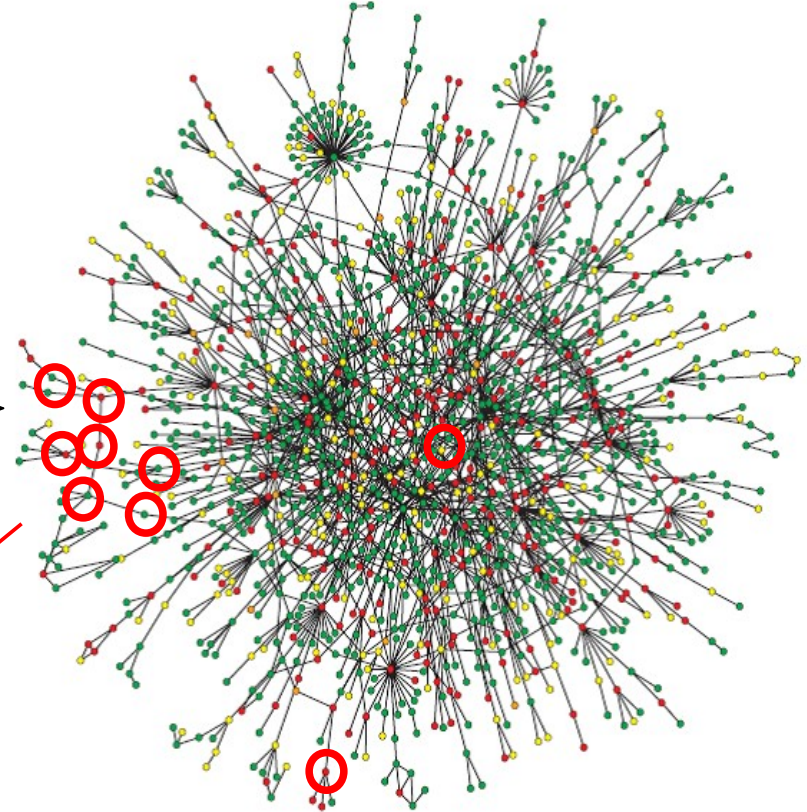
- ✓ In a central position in the interactome (betweenness).
- ✓ If they are highly connected (degree of connection).
- ✓ If they are in a very connected area (cluster coefficient).

Evaluating the cooperative behaviour

Your genes might be part of a network that explain the phenotype.



M22382
Y00371
X15183
M90054
AA650377
U78027
U14972
M17887
M20472
AB015320



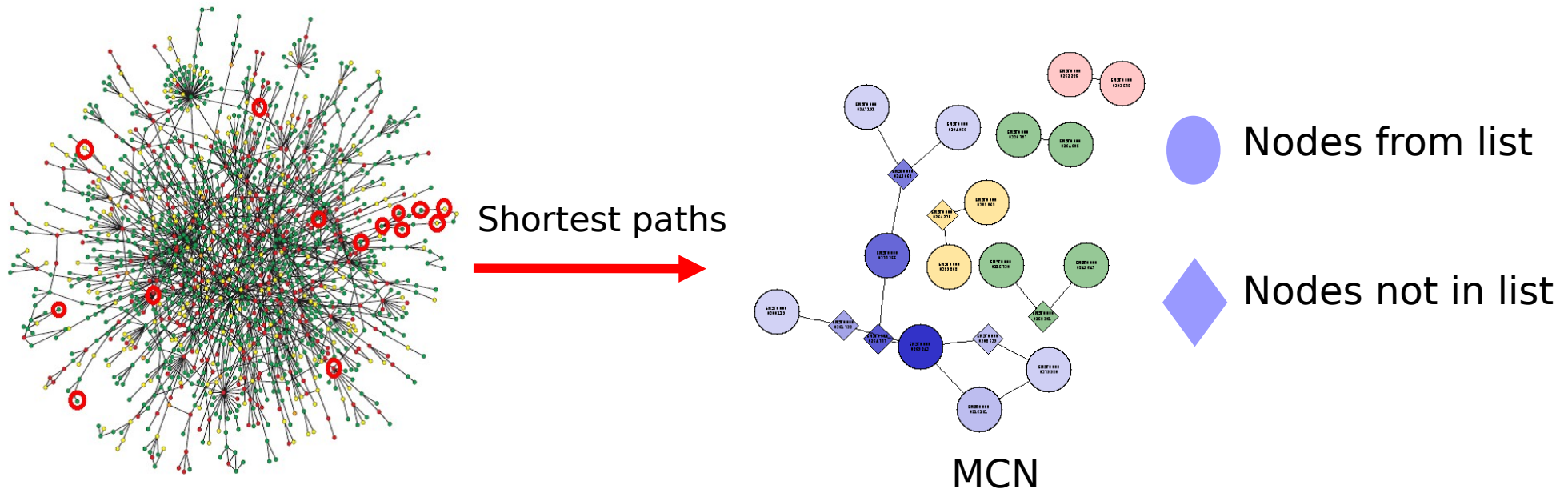
Proteins external in the complete interactome but they reveal an interesting subnetwork.

It seems we have found something about the cooperative behaviour of our list!!!

Evaluating the cooperative behaviour

Minimal connected network (MCN)

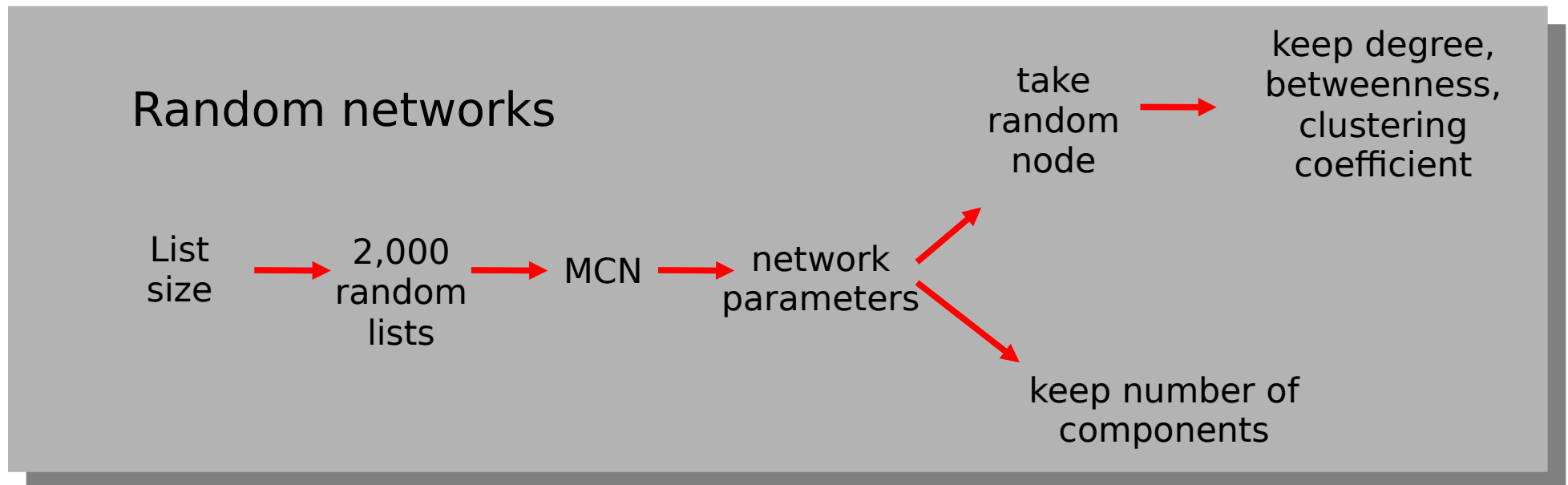
- ✓ Find all shortest paths for all the pairs of nodes.
- ✓ Accept paths that connect two proteins in the list either directly or through a predetermined number of not-in-list proteins (0-3).



Evaluating the MCN

Parameters to evaluate: degree, betweenness, clustering coefficient & number of components.

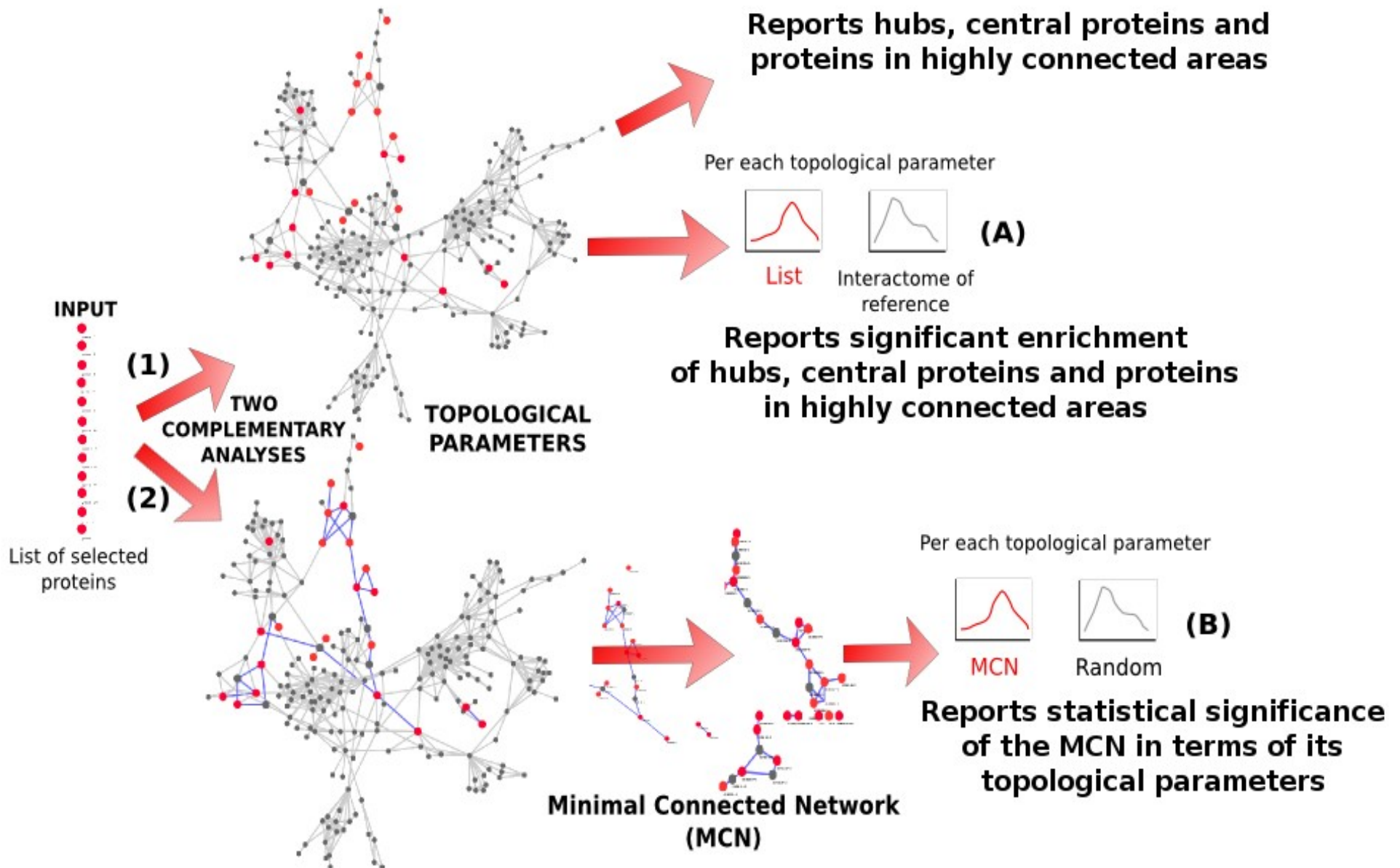
Comparison of parameters distributions of the network versus set of same size random distributions applying a Kolmogorov-Smirnov test.



You can use lists up to 200 nodes (prots/genes) with interactomic data (the list can be bigger).

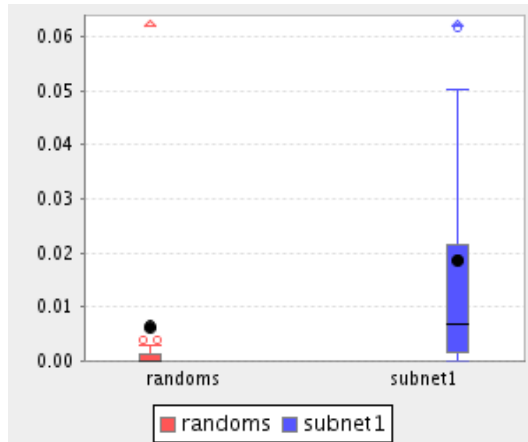
Snow

Studying networks in the omic work



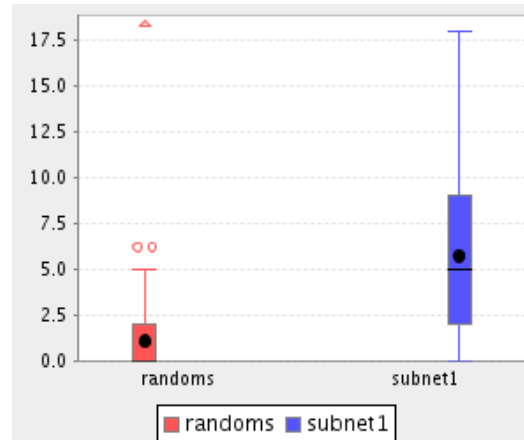
Evaluating the MCN

betweenness



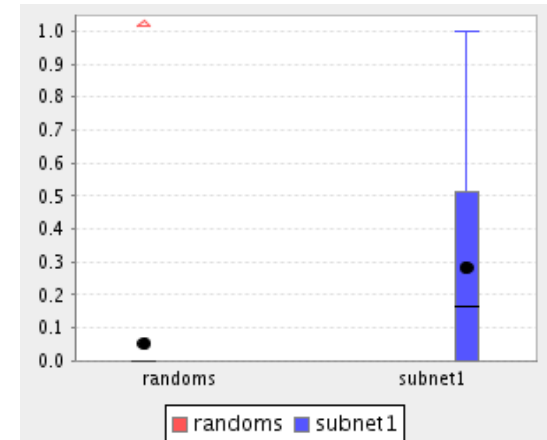
pval < 0.001

connections degree



pval < 0.001

clustering coefficient



pval < 0.001

Minimal Connected Network topological evaluation

- Number of components with more than 1 node : 1
- Number of components [95% confidence interval] : 12 [19, 34]

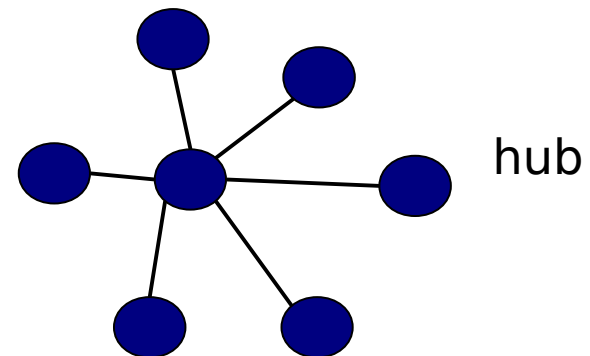
Evaluating the MCN

This tells us whether the complete list of proteins is ...

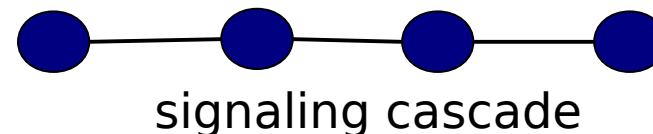
- ✓ More/less connected than a network coming from a random list (connections)
- ✓ Shape of the network in terms of centrality and internal connectivity (betweenness, clustering coefficient)
- ✓ Compact/Disperse network (n. of components)

Finding statistically significance in the different parameters points to different possible topologies of the network.

High Connections degree
Low Clustering coefficient



High Betweenness
Low Connections



Snow - web

BABELOMICS4

gene expression and functional profiling analysis suite

Upload data

Processing

Expression

Genomic

Functional analysis

Utilities

pablomiguez@gmail.com working on project *Cluestring* 8.60 Mb of 1.00 Gb (0.84)

Functional analysis

- Single enrichment analysis

- FatiGO

Provides significant over-representation of functional annotations by single enrichment analysis

- Marmite

Single enrichment analysis using text-mining derived annotations

- SNOW

PPI Network enrichment analysis. Finds subnetworks of protein-protein interactions with significant network parameters within a list of genes

Snow - web

Define your input data

- One list
 Comparing two lists

Select your data

[browse server](#) no data selected.

Species

Species

- Homo sapiens**
- Mus musculus*
- Drosophila melanogaster*
- Saccharomyces cerevisiae*
- Arabidopsis thaliana*
- Bos taurus*
- Escherichia coli (K12)*
- Own*

Nature of

- Tra
 Pro
 Ge

Snow - web

Side for statistical tests

Side for
Kolgomorov
test

- less
 greater

Random networks parameters

Number of randoms

500

1000

2000

Minimal connected generation parameters

Max. number of external proteins introduced

Calculate the number of bicomponents

Snow - web

Nature of your lists

- Transcripts
- Proteins
- Genes

Select interactome of reference

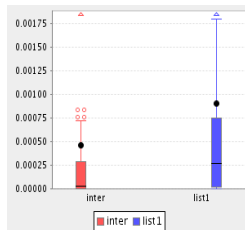
Select
group
interaction

- all ppis
- ppis detected by at least two methods (curated)

Snow - Results

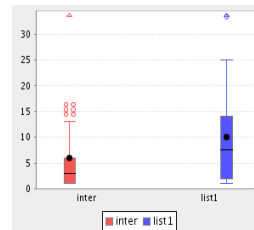
Role in the interactome

betweenness



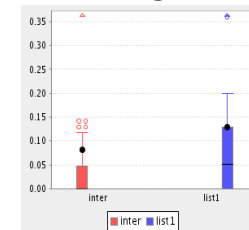
pval = 0.0058

connections degree



pval = 0.0013

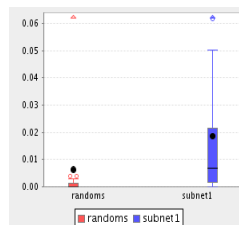
clustering coefficient



pval = 0.0055

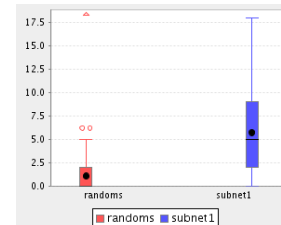
Evaluation of the MCN

betweenness



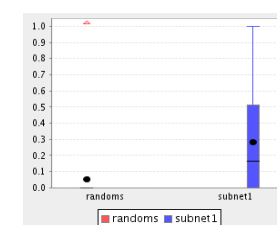
pval < 0.001

connections degree



pval < 0.001

clustering coefficient



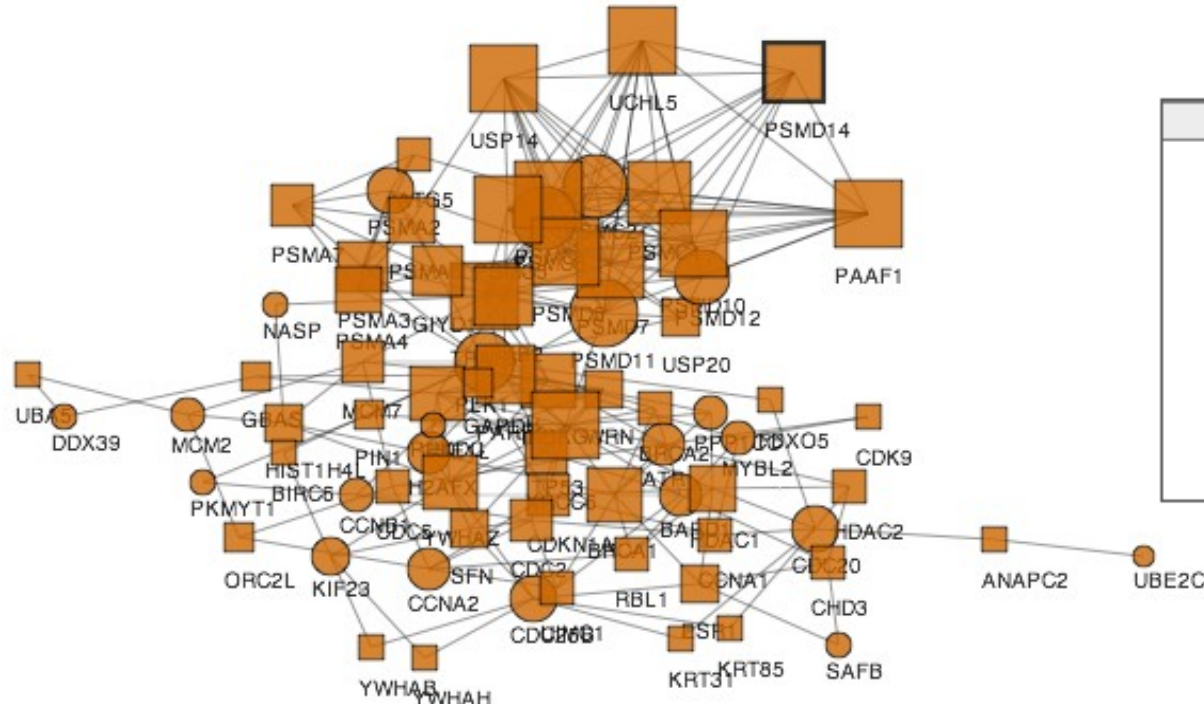
pval < 0.001

Information about components

Minimal Connected Network topological evaluation

- Number of components with more than 1 node : 1
- Number of components [95% confidence interval] : 12 [19, 34]

Snow - Network visualization



Information	
Id:	ENSG00000115233
Betweenness:	0.00
Connections:	10
Coefficient:	0.86
Symbol:	PSMD14
Ensembl:	ENSG00000115233
Uniprot:	Q00487

Better with Google Chrome!!!

Information
Search
Databases
Nodes editor
Layout
Clean network

▶ Genomic regions

▼ Functional annotation

Gene Ontology (GO)
Go biological process

- [GO:0006511](#): The chemical reactions and pathways resulting in the breakdown of a protein or peptide by hydrolysis of its peptide bonds, initiated by the covalent attachment of a ubiquitin moiety, or multiple ubiquitin moieties, to the protein.
- [GO:0031145](#): The chemical reactions and pathways resulting in the breakdown of a protein or peptide by hydrolysis of its peptide bonds, initiated by the covalent attachment of ubiquitin, with ubiquitin-protein ligation catalyzed by the anaphase-promoting complex, and mediated by the proteasome.
- [GO:0051436](#): Any process that stops, prevents or reduces the frequency, rate or extent of ubiquitin ligase activity during the mitotic cell cycle.
- [GO:0051437](#): Any process that activates, maintains or increases the rate of ubiquitin ligase activity during the mitotic cell cycle.
- [GO:0070536](#): A protein deubiquitination process in which a K63-linked ubiquitin chain, i.e. a polymer of ubiquitin formed by linkages between lysine residues at position 63 of the ubiquitin monomers, is removed from a protein.

Go cellular component

- [GO:0000502](#): A large multisubunit complex which catalyzes protein degradation. This complex consists of the barrel shaped proteasome core complex and one or two

Snow - Editing your network

Information Search Databases Nodes editor Layout Clean network

Search for any field in the nodes

apoptosis Search

Descriptions	Nodes	Edit
<input checked="" type="checkbox"/> GO:0006916: A process which directly inhibits any of the steps required for cell death by apoptosis.	<input type="checkbox"/> YWHAZ	Show or hide nodes: <input type="button" value="show"/> <input type="button" value="hide"/>
<input checked="" type="checkbox"/> GO:0042981: Any process that modulates the occurrence or rate of cell death by apoptosis.	<input type="checkbox"/> TRAF6	Change node shape: <input type="button" value="Square"/> <input type="button" value="Circle"/>
<input checked="" type="checkbox"/> GO:0006917: A process that directly activates any of the steps required for cell death by apoptosis.	<input type="checkbox"/> IKBKG	Change color nodes: <input type="text" value="#123456"/>
<input checked="" type="checkbox"/> GO:0008630: A cascade of processes initiated by the detection of DNA damage and resulting in the induction of apoptosis (programmed cell death).	<input type="checkbox"/> WRN	
<input checked="" type="checkbox"/> GO:0043065: Any process that activates or increases the frequency, rate or extent of cell death by apoptosis.	<input type="checkbox"/> BRCA1	
	<input type="checkbox"/> PRKDC	
	<input type="checkbox"/> SFN	
	<input type="checkbox"/> TP53	
	<input type="checkbox"/> PSMA3	

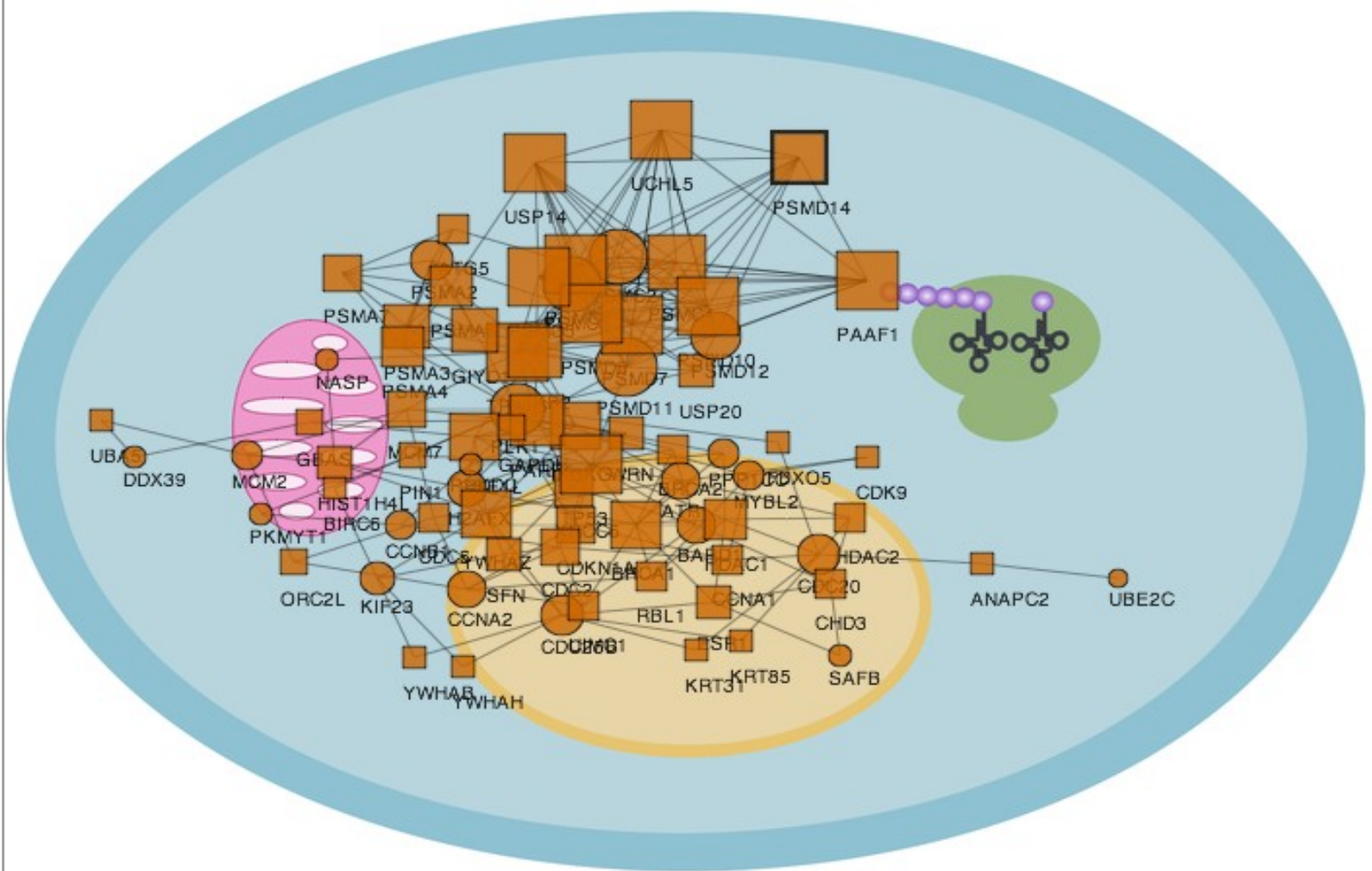
Information Search Databases Nodes editor Layout Clean network

Edit nodes by typing its id separated by comma, for instance: Q7Z4G1,P24530

Hide external nodes
 Show isolated nodes
 Show labels

<input type="text"/>	Show or hide nodes:	<input type="button" value="Show"/>	<input type="button" value="Hide"/>
	Change node shape:	<input type="button" value="Square"/>	<input type="button" value="Circle"/>
	Change color nodes:	<input type="text" value="#123456"/>	

Snow - Editing your network



Information Search Databases Nodes editor **Layout**

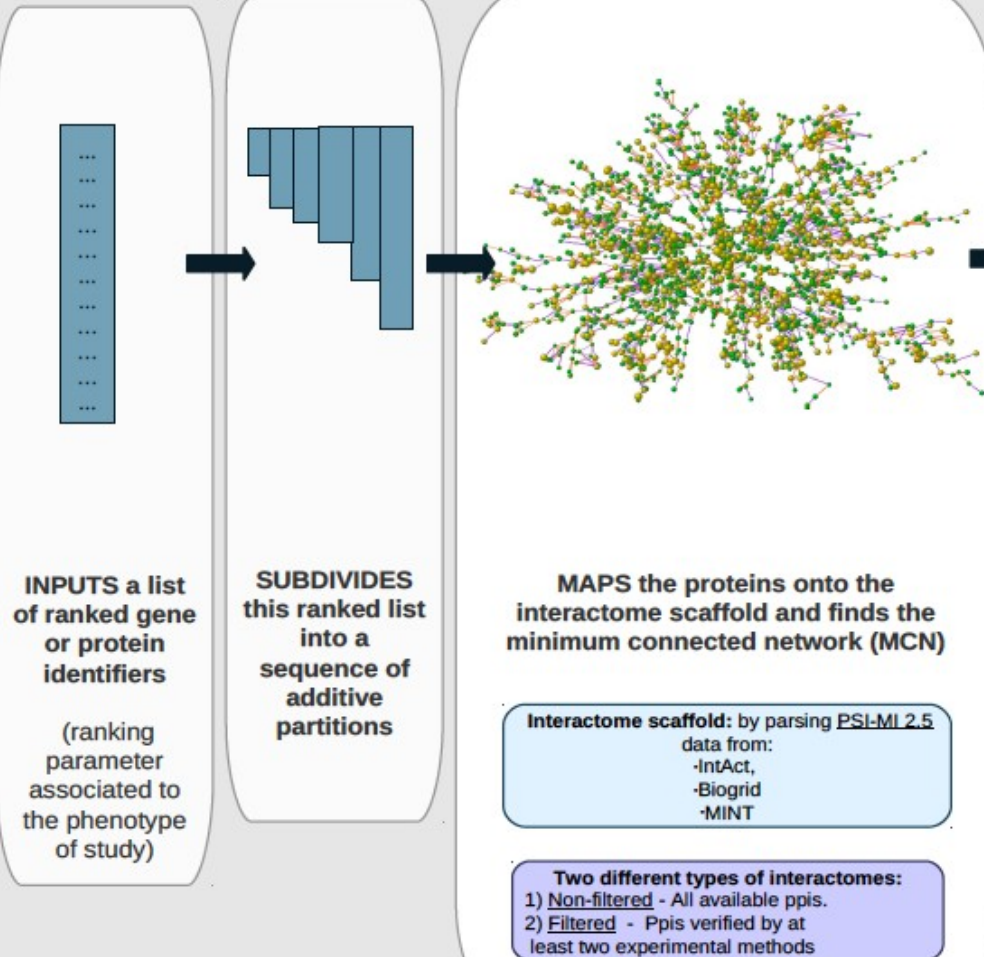
Select here your favourite layout

- Empty layout
- Full cell layout
- Partial cell layout

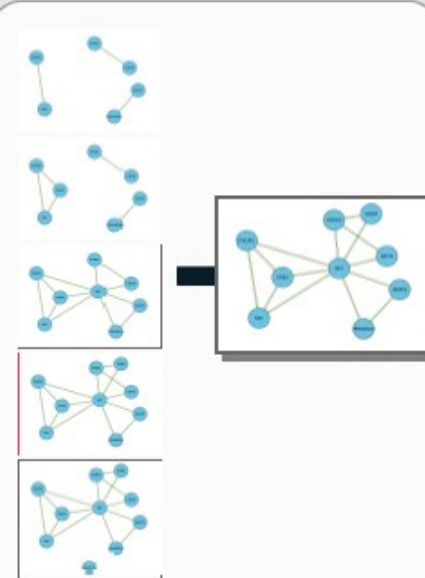
NetworkMiner

Threshold free methods (gene set enrichment)

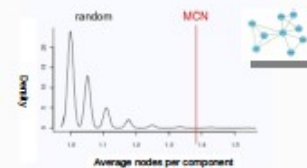
Network Miner Schema



List size / n. components



DETERMINES the partition with MCNs having the average connectivity per component significantly higher than their random expectations (by sampling randomly protein lists of the same size)



Advantages

1. Does not depend on an arbitrary threshold imposition
2. Finds groups of genes subject to the constraints of molecular interaction network.
3. Reports statistical significance of the MCN topology:
 - Connections degree
 - Betweenness
 - Clustering coefficient
 - Components
 - Bicomponents
4. MCN visualization that reports functional information about the MCN and its components.
 - Iterative exploration
 - Editable
 - Associated with external databases

NetworkMiner - web

Select your data

[browse server](#)

no data selected.

Filter your list

Select filter

- Whole list
- Subset of the list

Number items

items with ranking parameter below (ascending list) or above (descending list)

Order your list

Select order

- Ascending
- Descending

Select interactome

- Species Homo sapiens
 Saccharomyces cerevisiae
 Drosophila melanogaster
 Escherichia coli (K12)
- Level curated (ppis detected by two methods)

NetworkMiner - web

Type of Identifier in the list

Proteins/Transcripts
 Genes

Network generation parameters

Max. number of external proteins introduced

Select threshold of significance (p-value)

Significant value:

Job

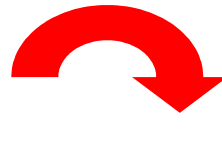
Job name:

Job description

NetworkMiner - Results

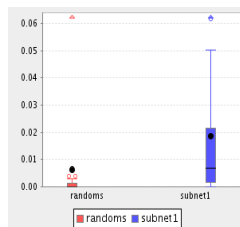
Results

- All results : [result_all.txt](#)
- pval of MCN chosen : 0
- size of MCN chosen : **197**



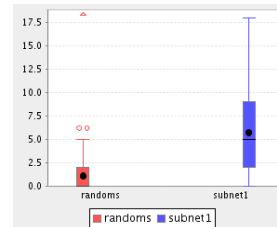
Evaluation of the MCN

betweenness



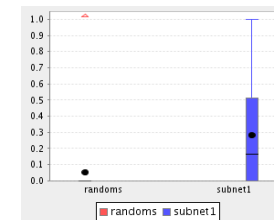
pval < 0.001

connections degree



pval < 0.001

clustering coefficient



pval < 0.001

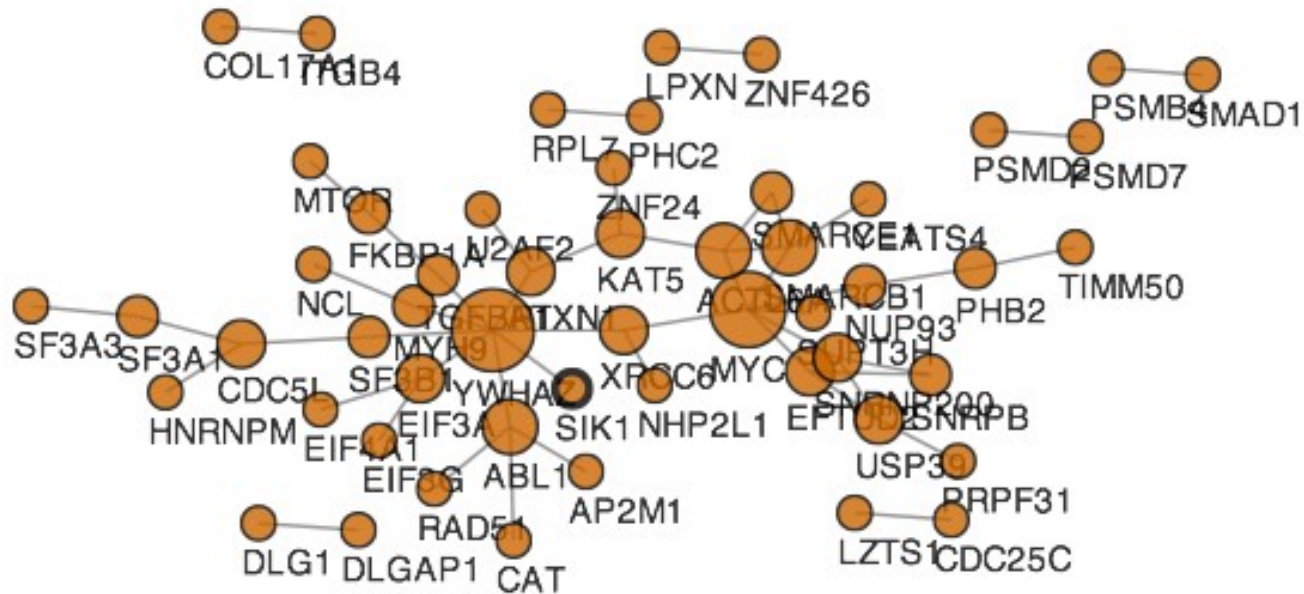
Information about components

Minimal Connected Network topological evaluation

- Number of components with more than 1 node : 1
- Number of components [95% confidence interval] : 12 [19, 34]

NetworkMiner - Results

Information	
Id:	MYB
Betweenness:	0.0
Connections:	1
Coefficient:	0.0
Symbol:	SIK1
Ensembl:	ENSG00000142178
Uniprot:	P57059



Exercises

Snow On-line examples

SNOW

▼ Online examples (test the form with example data)

- Example 1: Downregulated in fibroblasts from old individuals, compared to young
- Example 2: Upregulated by induction of exogenous BRCA1 in EcR-293 cells



NetworkMiner On-line examples

Network Miner

▼ Online examples (test the form with example data)

- Example 1: Genes_up_in_control_Vs_case_Hirschsprung_disease
- Example 2: Essential_genes_in_cancer_cell_line_K562



Exercises

2. Some datasets to run SNOW






<http://bioinfo.cipf.es/babelomicstutorial/>

Here are several examples of lists of genes selected to differentiate two samples in microarray experiments. The description of the experiment is given.

The SNOW parameters used to perform the analyses were:

- Interactome of reference: ppis detected by two methods.
- Maximum number of external proteins: 1
- Nature of the lists: Genes

Download the lists and perform your own SNOW analyses choosing same or different parameters. For a reference we give the results pages as you will obtain them, have a look at them and compare them with SNOW results using different parameters taking into account that results shown here may have been run with different version of ppi data.

Example number	Dataset	Description
2.1	 brca1_overexp_up	Upregulated by induction of exogenous BRCA1 in EcR-293 cells
2.2	 brca1_overexp_dn	Downregulated by induction of exogenous BRCA1 in EcR-293 cells
2.3	 serum_fibroblast_cellcycle	Cell-cycle dependent genes regulated following exposure to serum in a variety of human fibroblast cell lines
2.4	 ageing_brain_dn	Age-downregulated in the human frontal cortex
2.5	 brca1_sw480_up	Up-regulated by infection of human colon adenocarcinoma cells (SW480) with Ad BRCA1 versus Ad LacZ control

Exercises

Some old stuff ...

Exercises

A case of study

BRCA1 transcriptionally regulates genes involved in breast tumorigenesis

Piri L. Welch^{*}, Ming K. Lee^{*}, Rachel M. Gonzalez-Hernandez^{*†}, Daniel J. Black^{*}, Mamatha Mahadevappa[‡], Elizabeth M. Swisher^{*}, Janet A. Warrington[‡], and Mary-Claire King^{*§}

Departments of ^{*}Medicine and Genome Sciences and [†]Pathology, University of Washington, Health Sciences Room K-160, Seattle, WA 98195-7720; and [‡]Affymetrix, Inc., Santa Clara, CA 95051

Communicated by Arno G. Motulsky, University of Washington, Seattle, WA, March 28, 2002 (received for review December 3, 2001)

Loss of function of *BRCA1* caused by inherited mutation and tissue-specific somatic mutation leads to breast and ovarian cancer. Nearly all *BRCA1* germ-line mutations involve truncation or loss of the C-terminal BRCT transcriptional activation domain, suggesting that transcriptional regulation is a critical function of the wild-type gene. The purpose of this project was to determine whether there is a link between the role of *BRCA1* in transcriptional regulation and its role in tumor suppression. We developed a cell line (in which *BRCA1* can be induced) and used microarray analysis to compare transcription profiles of epithelial cells with low endogenous levels of *BRCA1* vs. transcription profiles of cells with 2–4-fold higher induced levels of expression of *BRCA1*. At these levels of expression, *BRCA1* did not induce apoptosis. Undirected cluster analysis of six paired experiments revealed 373 genes, the expression of which was altered significantly and consistently by *BRCA1* induction. Expression of 62 genes was altered more than 2-fold. *BRCA1*-regulated genes associated with breast tumorigenesis included the estrogen-responsive genes *MYC* and cyclin D1, which are overexpressed in many breast tumors; *STAT1* and *JAK1*, key components of the cytokine signal transduction pathway; the extracellular matrix protein laminin 3A; *ID4*, an inhibitor of DNA-binding transcriptional activators, which in turn negatively regulates *BRCA1* expression; and the prohormone stanniocalcin, expression of which is lost in breast tumor cells. Coordinated expression of *BRCA1* with *ID4* and with stanniocalcin was confirmed in primary breast and ovarian tumors.

17). Overexpression of *BRCA1* induces genes in the apoptotic pathway (18, 19). Increased expression of *BRCA1* leads to repression of estrogen receptor (ER)-mediated transcription (20–22).

Our purpose in this project was to determine whether there is a link between the role of *BRCA1* in transcriptional regulation and its role in tumor suppression by identifying transcriptional targets of *BRCA1* that are involved in breast tumorigenesis. We developed an epithelial cell line in which *BRCA1* could be induced at modest levels and then used microarray technology to investigate changes in the cellular transcription profile in response to induction of *BRCA1*. In six replicate experiments, after induction of *BRCA1* expression levels of 373 genes were altered consistently, 62 of them at least 2-fold. Among those implicated in breast tumorigenesis are cyclin D1, *JAK1* and *STAT1*, *MYC*, and *ID4*. These experiments also revealed that *BRCA1* induction was highly correlated with expression of the extracellular matrix protein laminin A3 (*LAMA3*) and with stanniocalcin (*STC1*), a prohormone whose loss may serve as a marker of breast and ovarian cancer.

Materials and Methods

Generation of Cell Lines with Inducible *BRCA1* Expression. A full-length *BRCA1* cDNA was assembled from partial cDNA clones kindly provided by I. M. Verma (The Salk Institute, La Jolla, CA). Sequence-verified, full-length wild-type *BRCA1* was cloned downstream of the ecdysone-inducible promoter in pIND (Invitrogen).

EcR-293 (Invitrogen) is a human embryonal kidney epithelial

Exercises

A case of study

A list of 168 genes induced by over-expression of BRCA1 as described in Welch et al., 2002 which doesn't include BRCA1

The MCN has a large component (79 genes) where BRCA1 (ENSG00000012048) is included

A classical enrichment analysis gave no significant over-representation of Gene Ontology terms or KEGG pathways in such component compared with the rest of the genome showing the multi-functionality of BRCA1

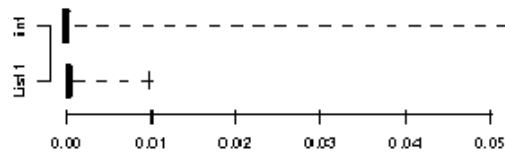
Exercises

A case of study

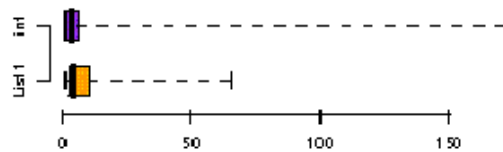
Network parameters evaluation

List's role within complete interactome

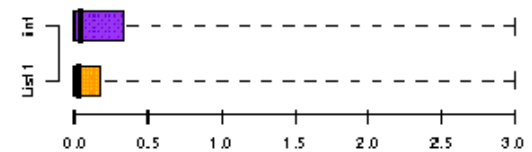
Betweenness:
List1 > Interactome pval=0



Connections:
List1 > Interactome pval=0.0017

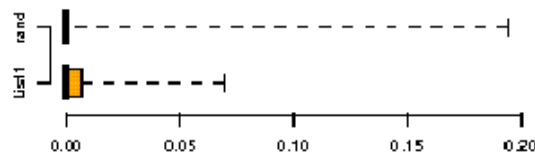


Clustering Coeff:
List1 < Interactome pval=0.0294

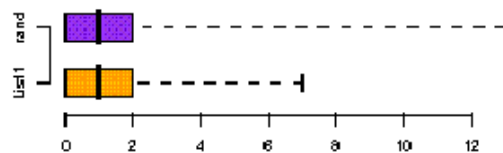


Minimal Connected Network topological evaluation

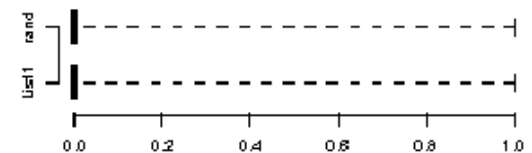
Betweenness:
List1 > Random pval=0



Connections:
List1 > Random pval=1e-04



Clustering Coeff:
List1 > Random pval=0.6873



Number of components [95% confidence interval]:
Number of components with more than 1 node:
Number of Bicomponents:
Articulation points:

List1: 59 [60-120]
List1: 9
List1: 68
List1: 70

Exercises

A case of study

- **BRCA1** presents a high centrality measure (betweenness)
- **BRCA1** interacts directly with genes such as:
 - STAT1 (its activation by kinase JAK1, interaction shown in the results, mediate cytokine and growth-factor signal transduction pathways),
 - HDAC1 (Histone deacetylase involved in the control of cell proliferation and differentiation),
 - UBE2D3 (Ubiquitin-conjugating enzyme E2D 3, responsible for the ubiquitination of the tumor-suppressor protein p53)
 - CD2 (involved in induction of mitosis).
- The most central genes are:
 - GRB2 (ENSG00000177885), STAT1, HTATIP, BRCA1 (ENSG00000012048), CREB1, IQGAP1, ANXA2, CALM2 and EGFR (ENSG00000146648)
 - being BRCA1 and EGFR two genes not included in the original list but introduced by calculating the MCN.