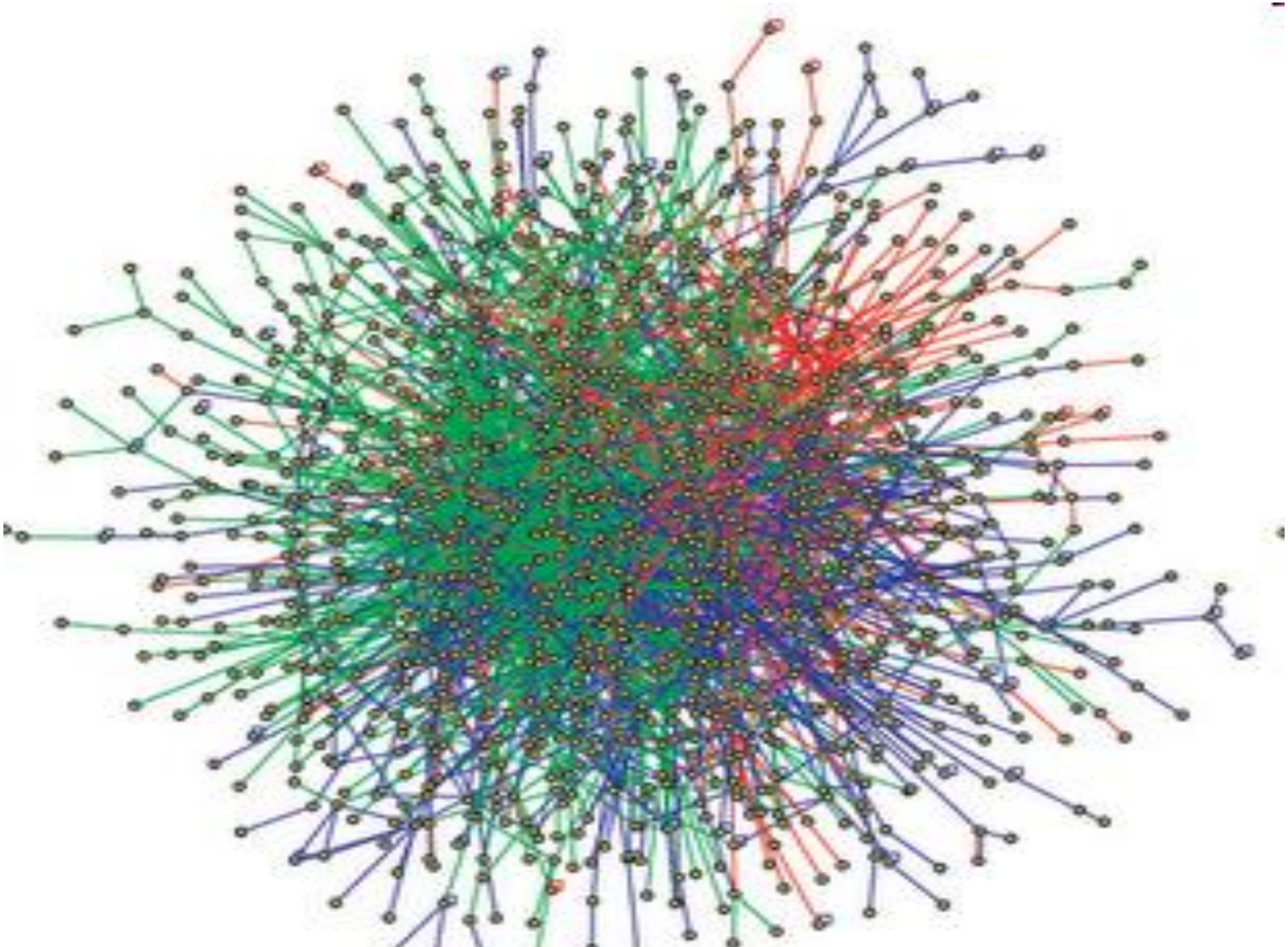
Networks

Day 6

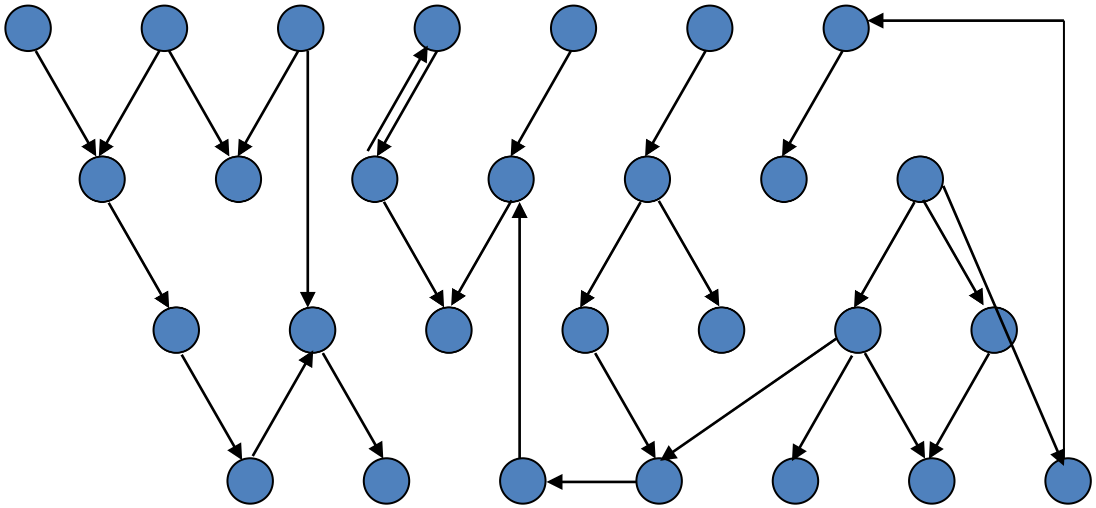
Human interactome

~130,000 binary interactions, most of which remain to be maped



Networks

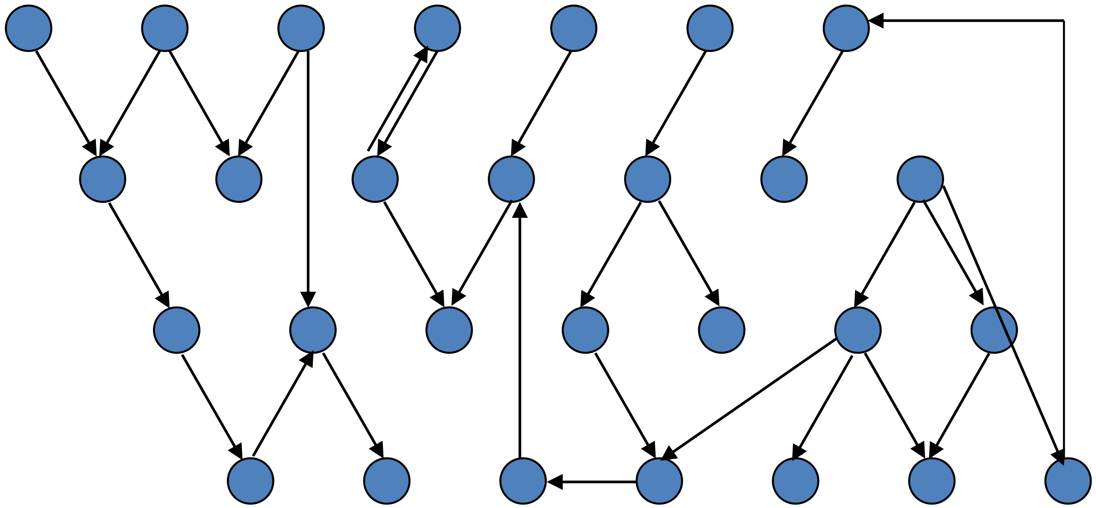
$G = (V, E)$



	Transcription network	**Signaling network**	**Metabolic network**
nodes *V* : ●	transcription factor genes	(genes of) signaling components	metabolites
edges (arcs) *E* : →	TF expression & transcriptional regulation	transmission of a signal	enzymatically catalyzed conversions

Networks

$G = (V, E)$



	Transcription network	**Signaling network**	**Metabolic network**
nodes V : ●	transcription factor genes	(genes of) signaling components	metabolic enzyme genes
edges (arcs) E : →	TF expression & transcriptional regulation	transmission of a signal	forwarding a metabolite

Networks

The sources for the existing knowledge:

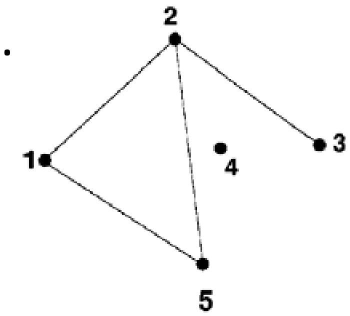
Transcription network: **TRANSFAC®**

Signaling network: **TRANSPATH®**

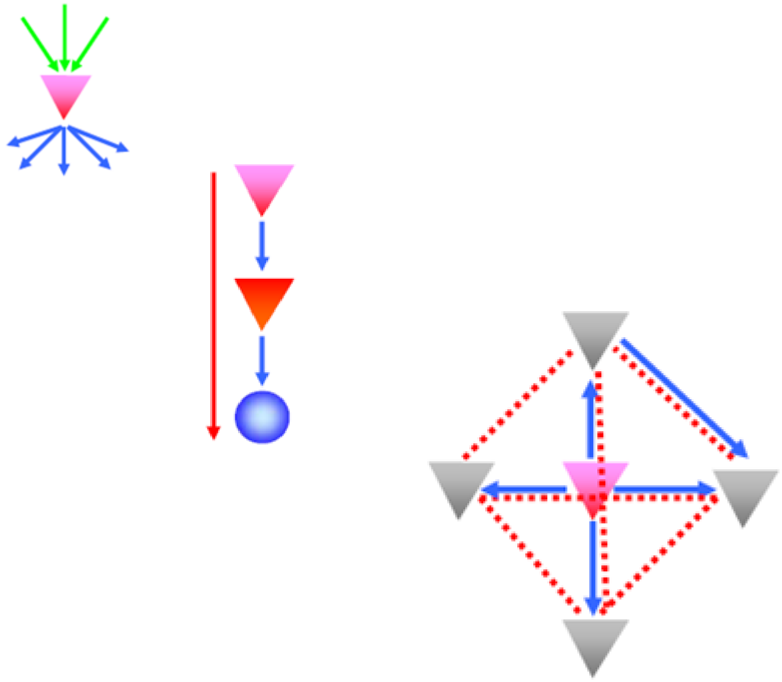
Metabolic network: **KEGG**

Graphs

Graph - the object consisting of vertices and edges joining them.
G={V, E}, where V- set of vertices (n), E – set of edges (m)
Graphs are cyclic and acyclic ,
sparse if m~n, **dense** m~n^2, **complete** m=n^2
directed and undirected

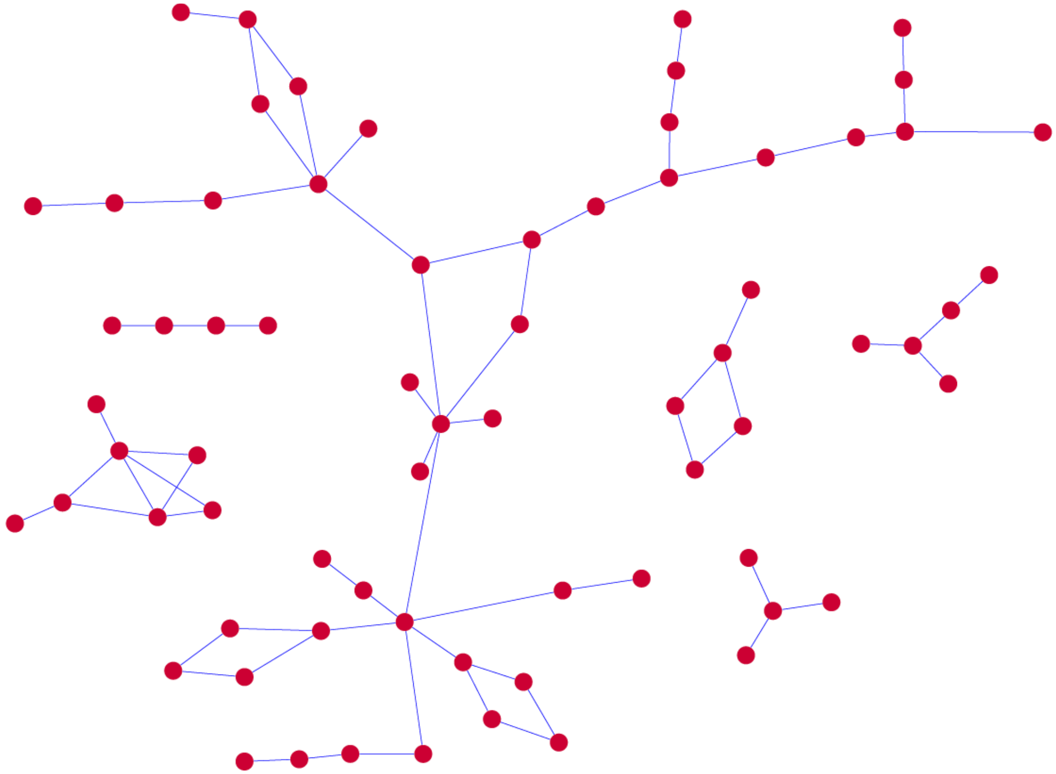


Characteristics of the topology



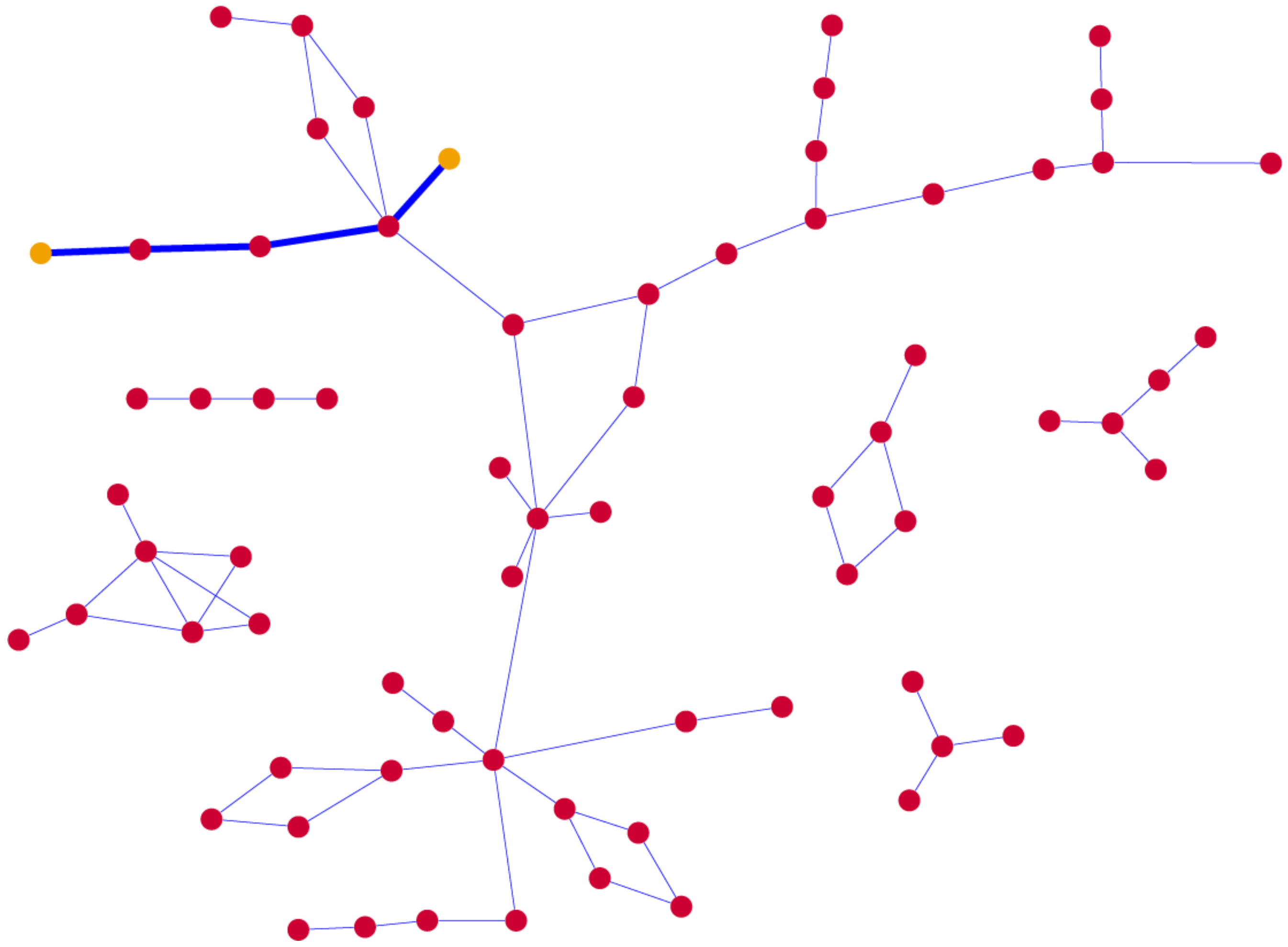
- Degree k_i
- Distribution of degree $P(k)$
- Average path length
- Network diameter
- Clustering coefficient

Undirected graph

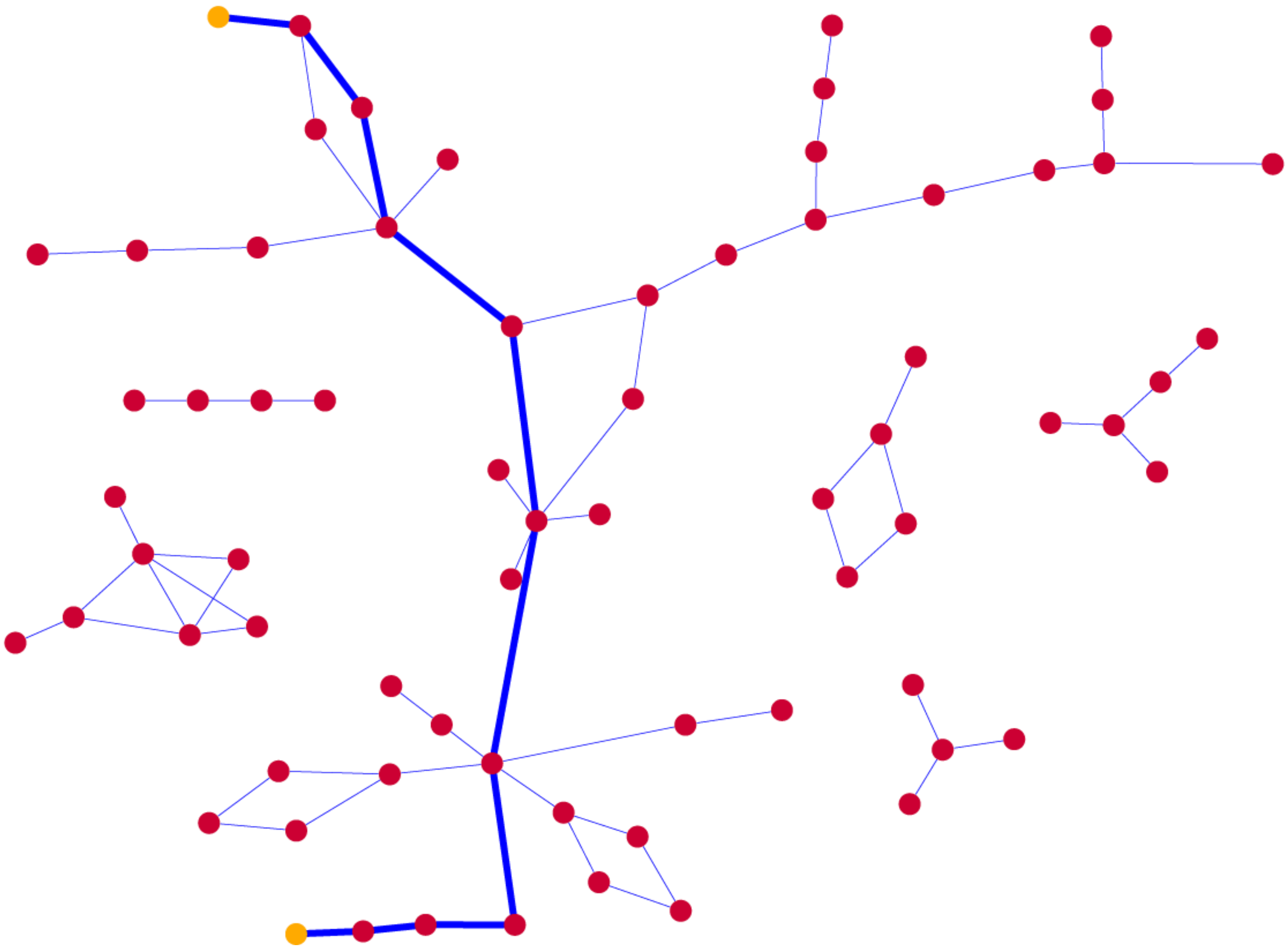


- G(V,E)
- |V| = 69
- |E| = 71

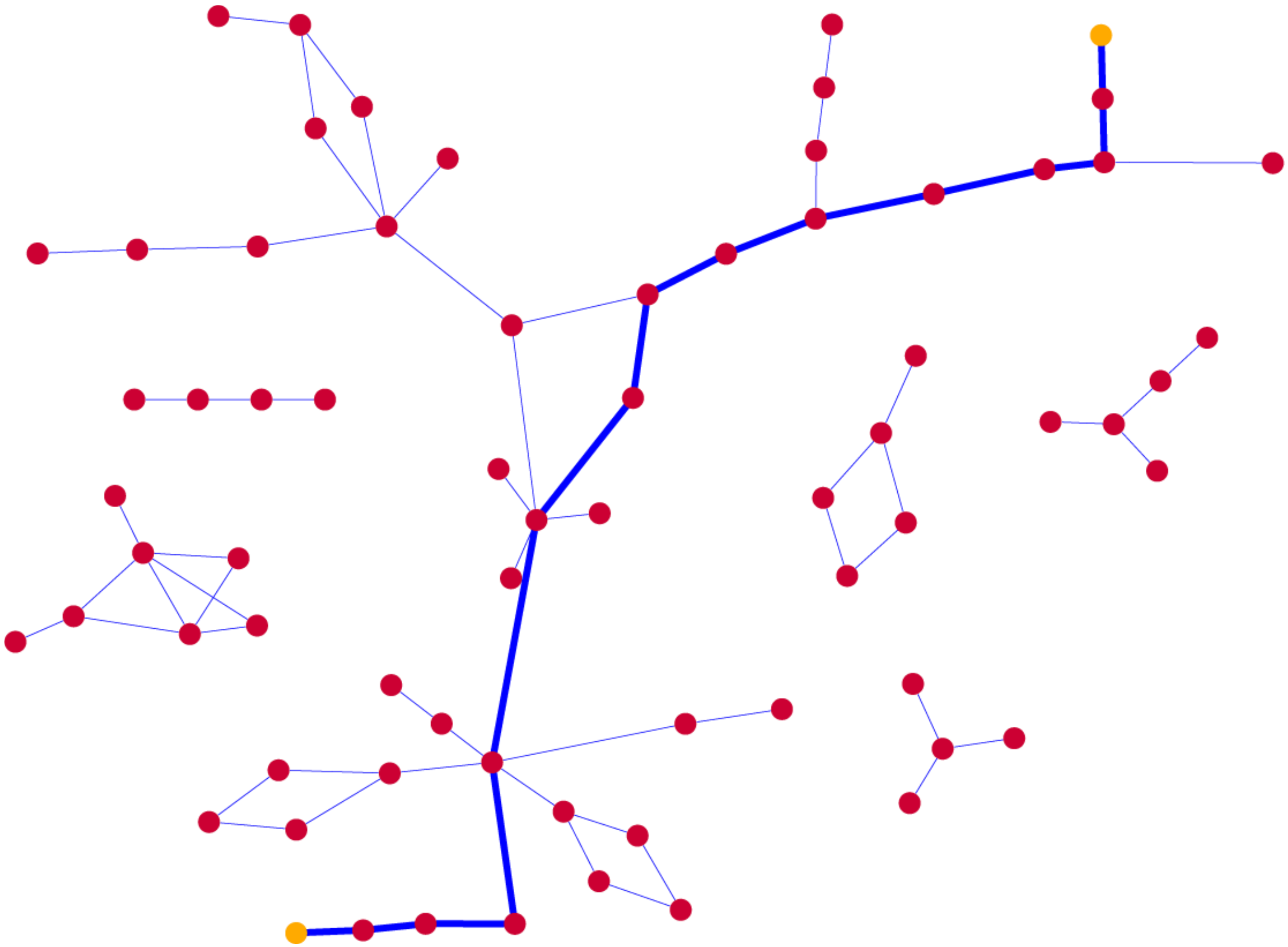
The shortest path between vertices



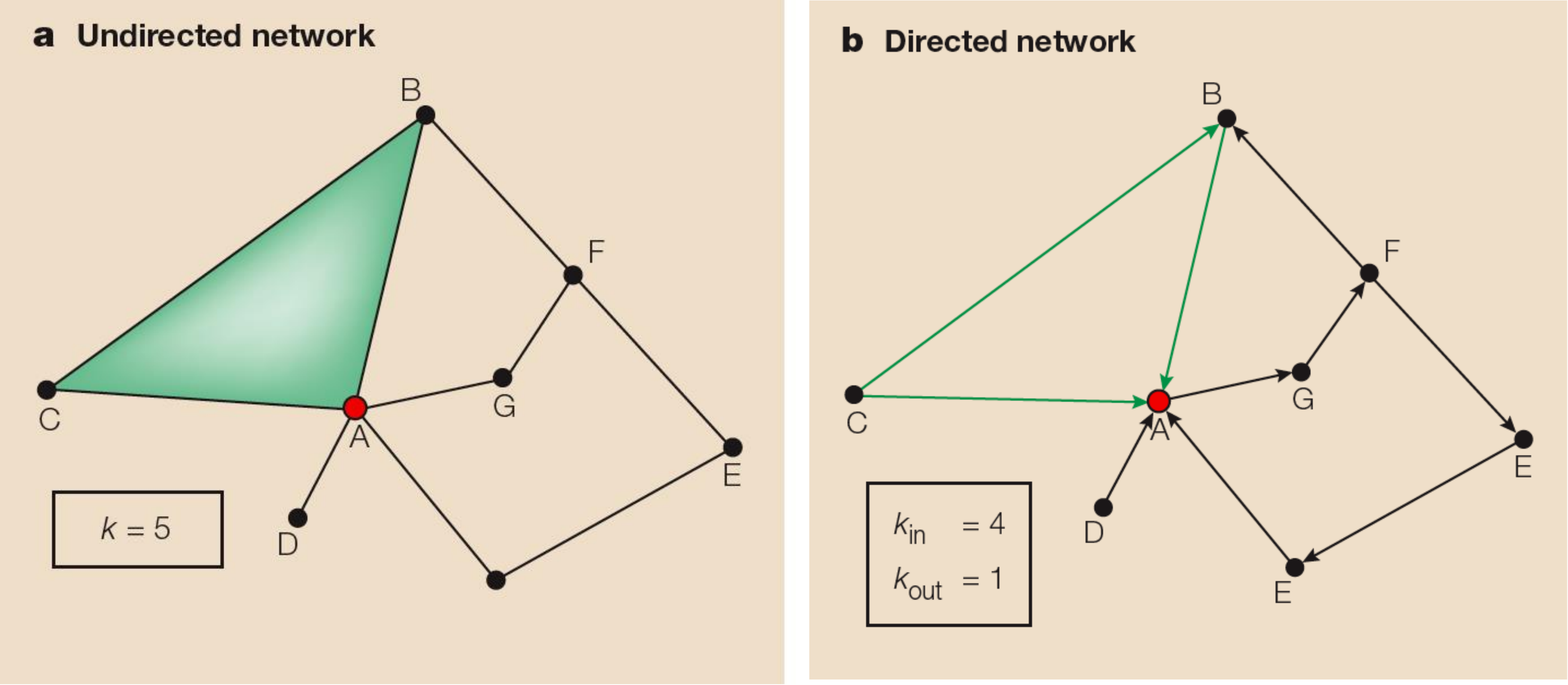
The shortest path between vertices



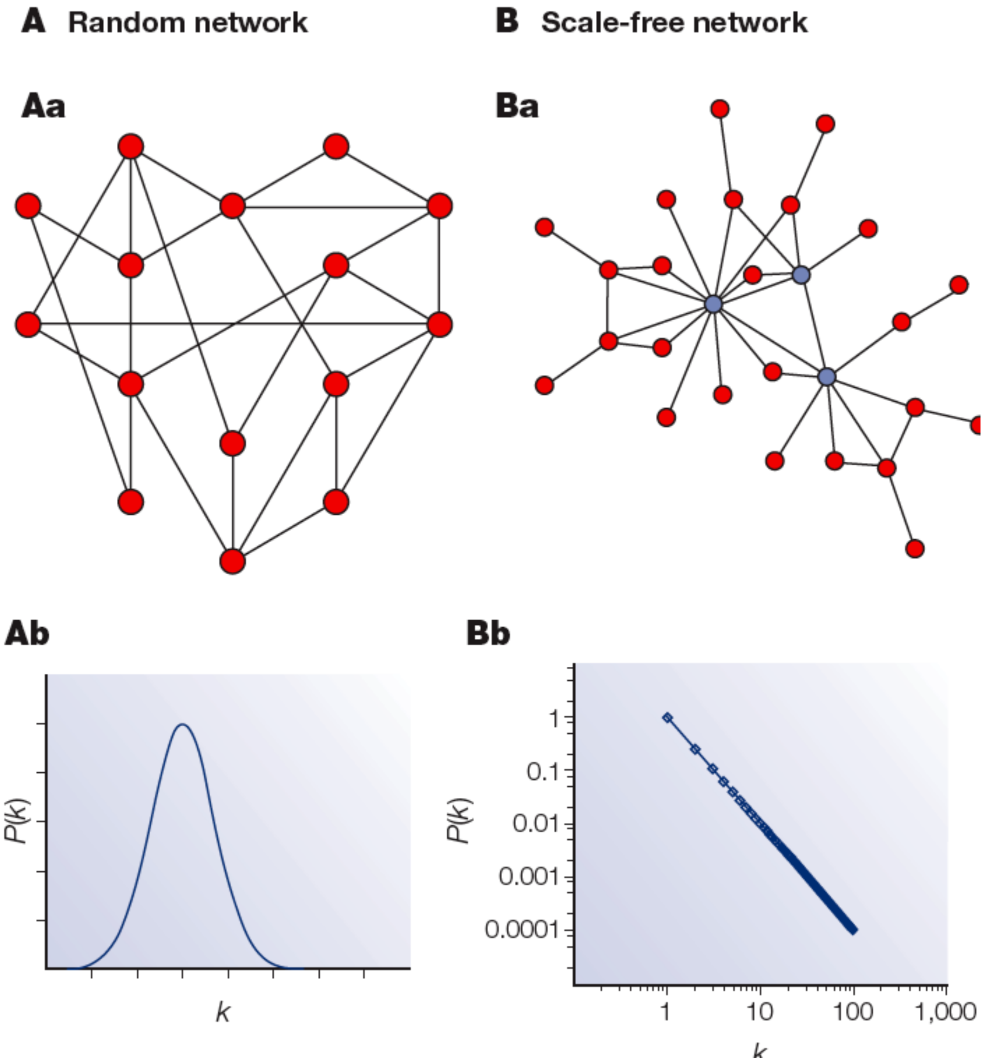
Diameter - the maximum distance in the network. Longest Shortest-Path



Characteristics of the network - the degree of vertex



Distribution of degree



P(k) is probability of each degree k, i.e fraction of nodes having that degree.

For random networks, P(k) is normally distributed.

For real networks the distribution is often a power-law:

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

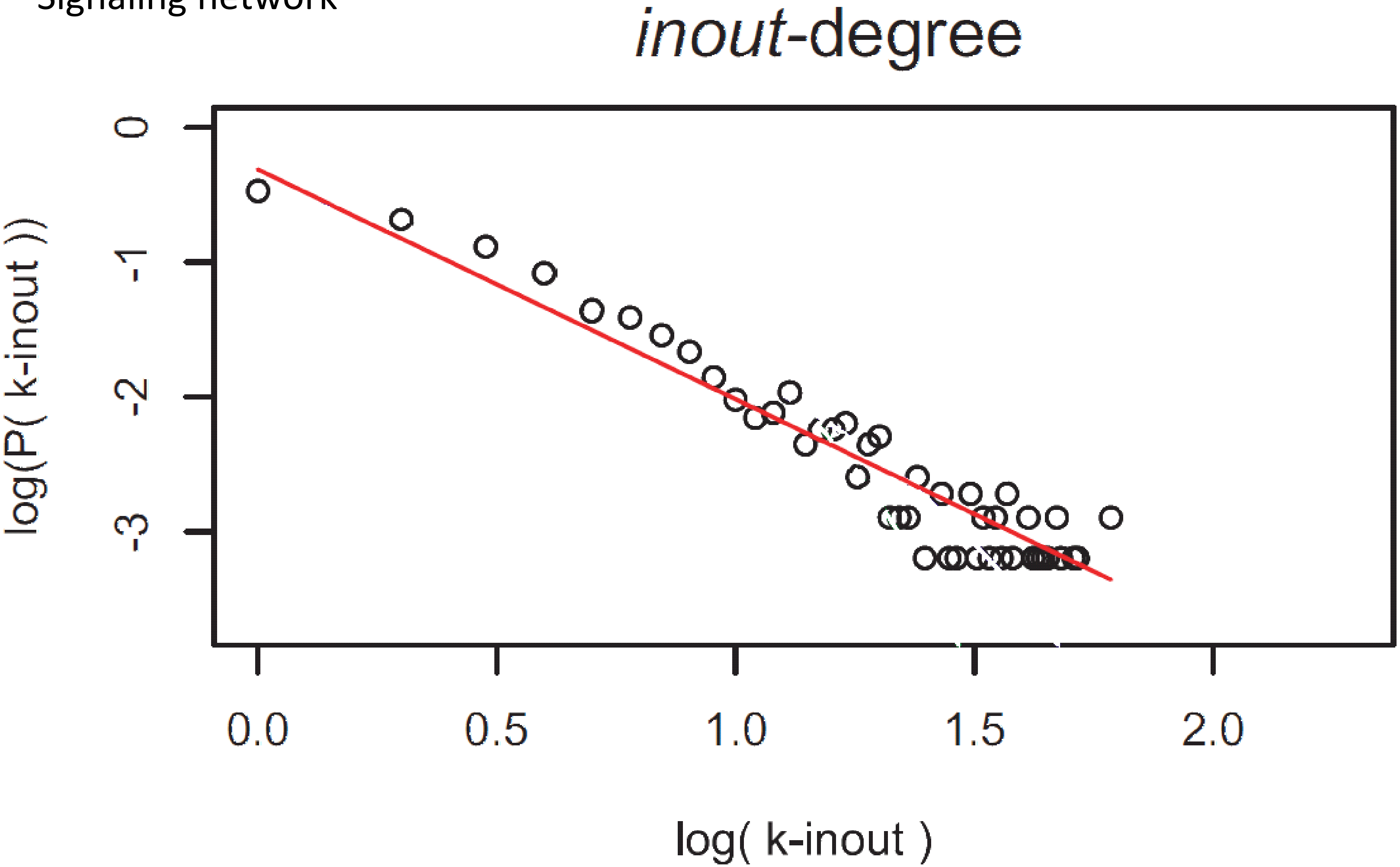
Such networks are said to be scale-free

Mammalian network analysis

Scale-freeness

$\gamma = 1.88$

Signaling network

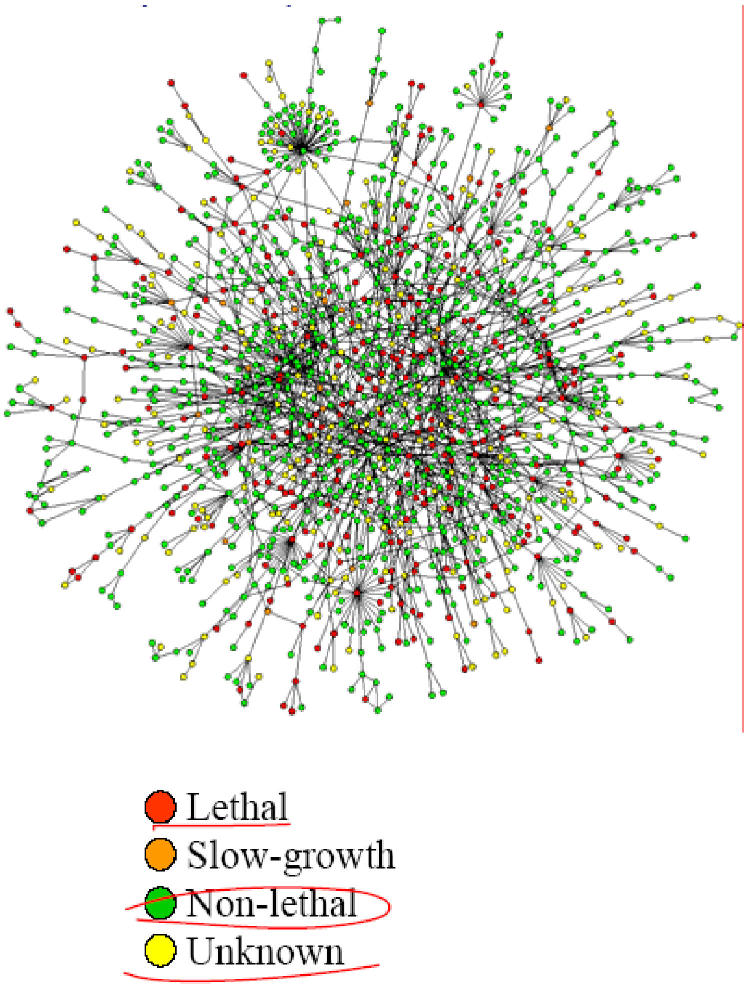


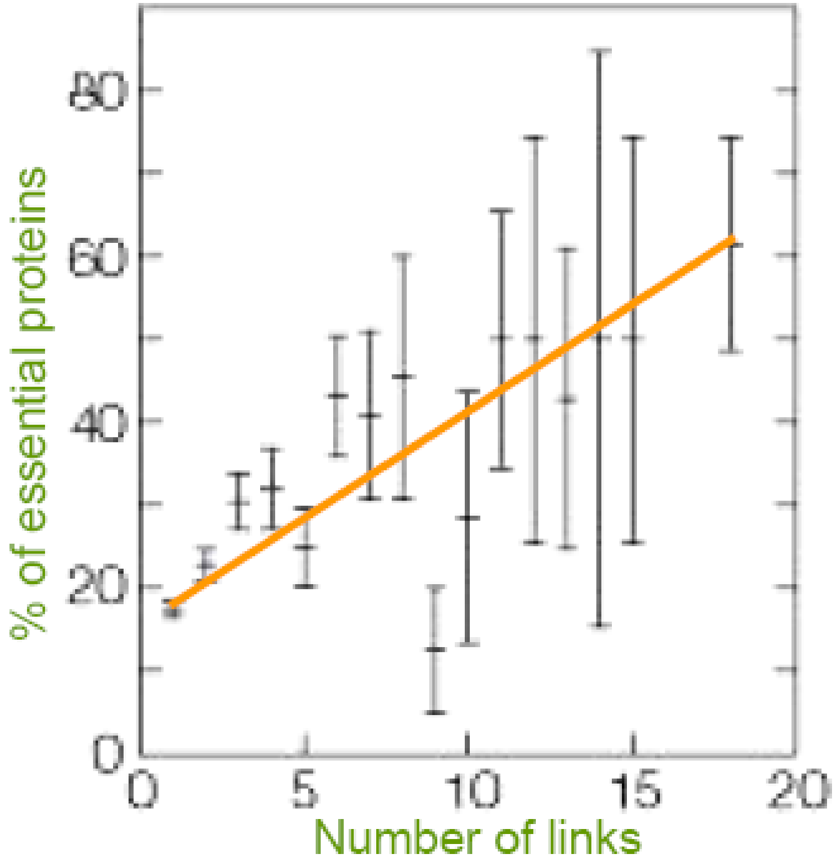
$$\boxed{P(k) \propto k^{-\gamma}}$$

Scalable networks are stable (Robust)

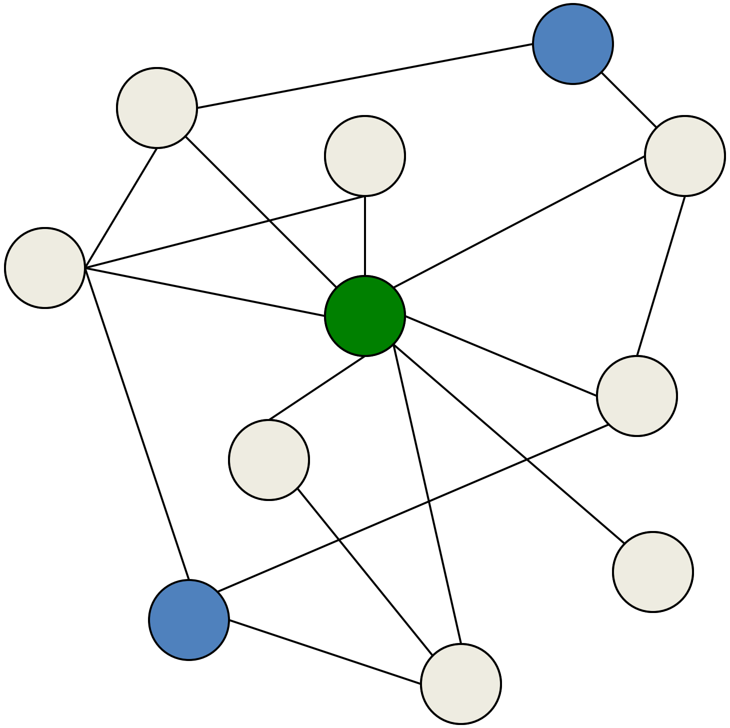
- Complex systems (cell, internet, social networks), are resilient to component failure
- Network topology plays an important role in this robustness
  - Even if ~80% of nodes fail, the remaining ~20% still maintain network connectivity
- *Attack vulnerability* if hubs are selectively targeted
- In yeast, only ~20% of proteins are lethal when deleted, and are 5 times more likely to have degree $k>15$ than $k<5$.

Lethality and connectedness are positively correlated





Clusters



Clustering coefficient of vertex i - is the ratio of the number of edges E_i neighbors to the number of possible edges that might exist in neighbors. Measure of network connectivity. Clustering coefficient for networks C - is the average ratio of all vertices C_i

$$C_I = \frac{n_I}{\binom{k}{2}} = \frac{2n_I}{k \cdot (k-1)}$$

k: neighbors of *I*

n_I: edges between node *I*'s neighbors

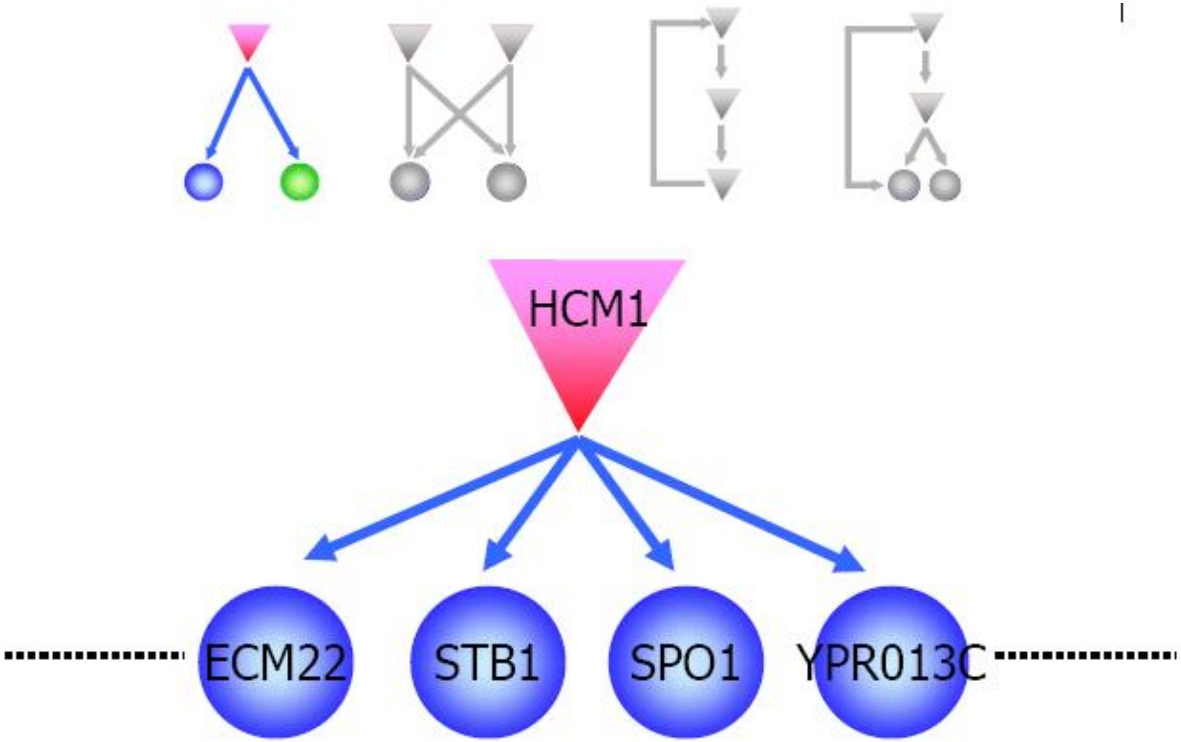
The center node has 8 (grey) neighbors

There are 4 edges between the neighbors

C = 2*4 /(8*(8-1)) = 8/56 = 1/7

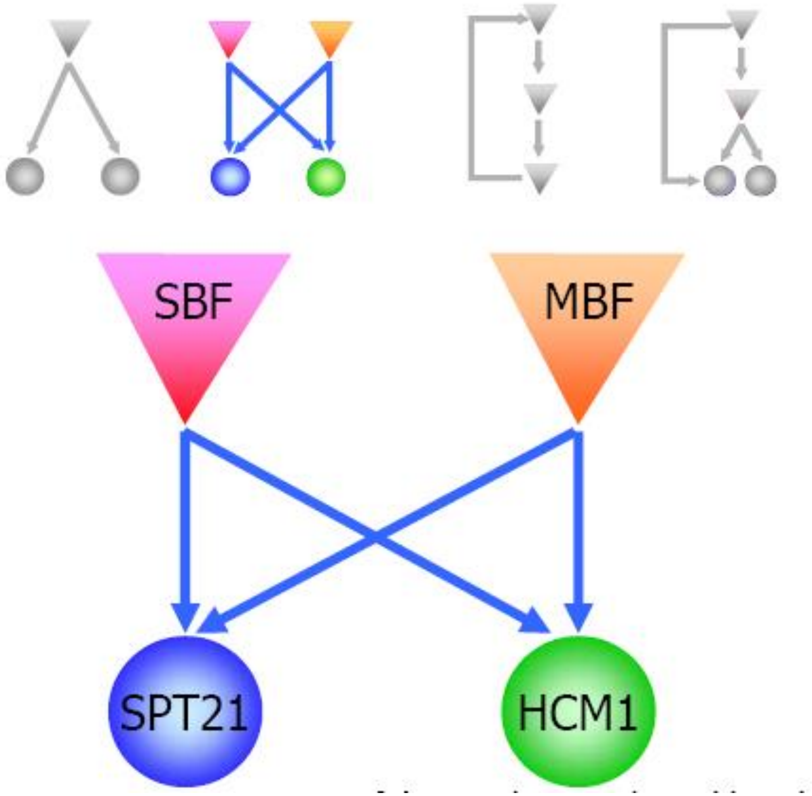
Regulatory motifs

Single input motifs



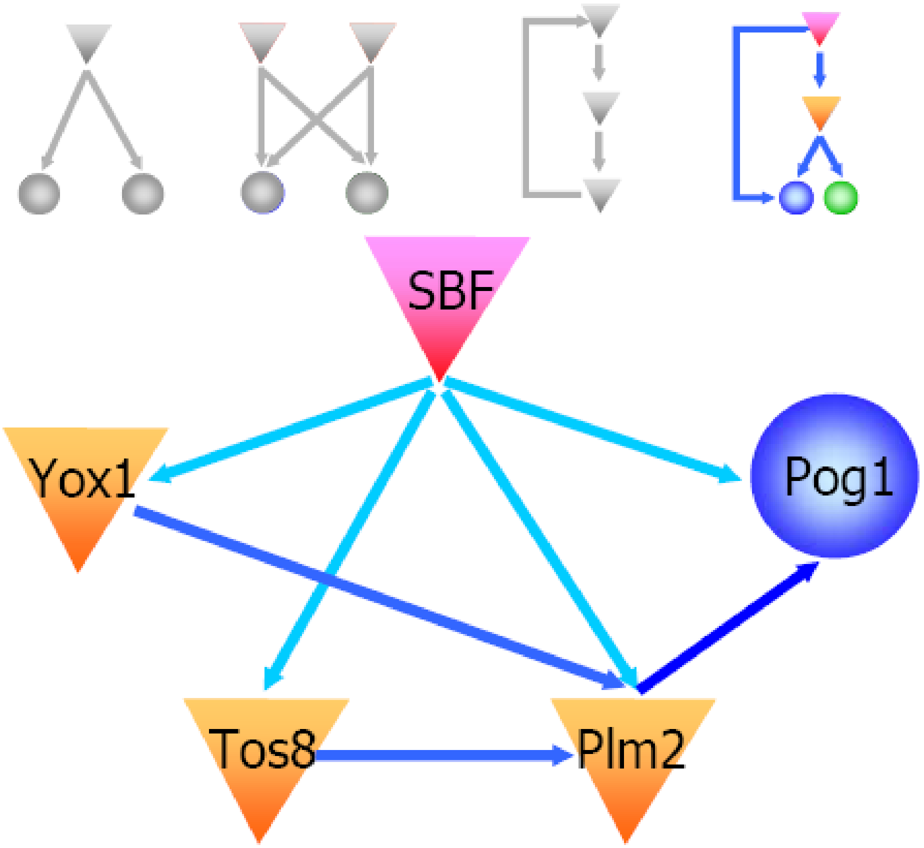
Regulatory motifs

Multiple output motifs



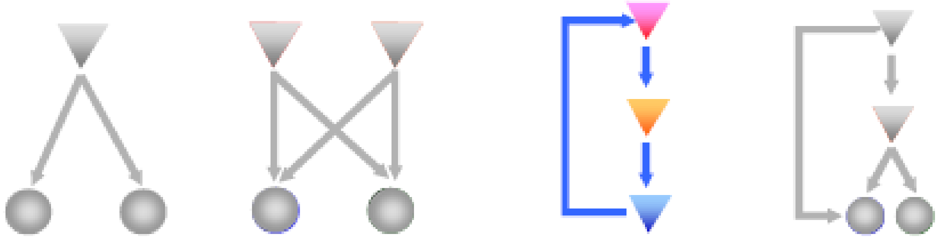
Regulatory motifs in the networks

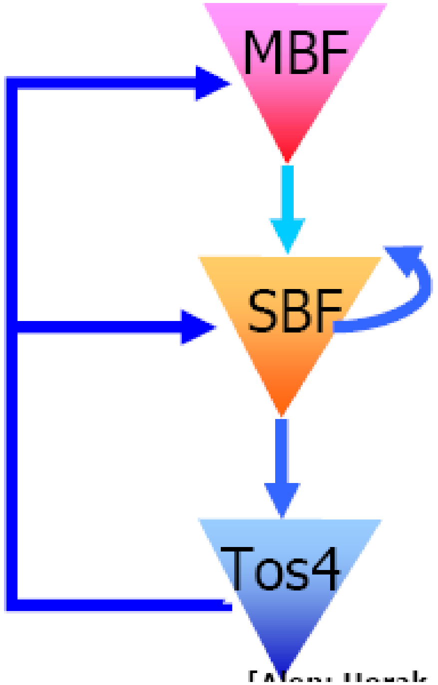
Feed-forward loops



Regulatory motifs

Feed-back loops

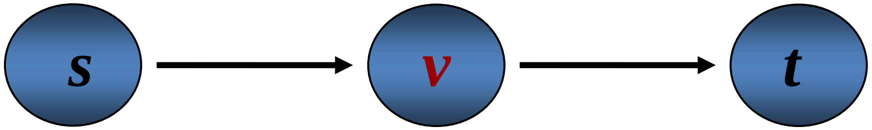




[Alon; Horak, Luscombe et al (2002), Genes &

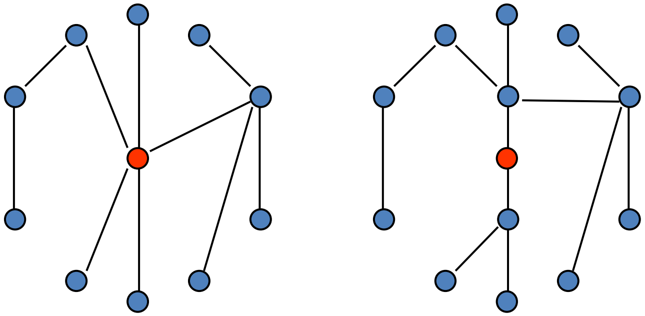
Network analysis: methodology

Betweenness centrality of a vertex *v*



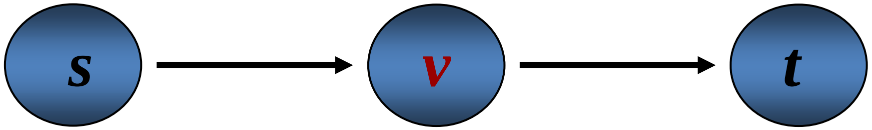
$$B_c(v) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq v \in V} \frac{\delta_{st}(v)}{\delta_{st}}$$

Fraction of those shortest paths between all pairs of vertices *s* and *t* that pass through vertex *v*. Allows to quantify how influential a given gene/protein in a whole network is.



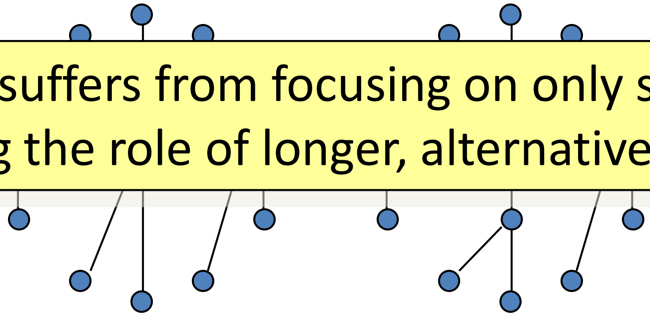
Network analysis: methodology

Betweenness centrality of a vertex *v*

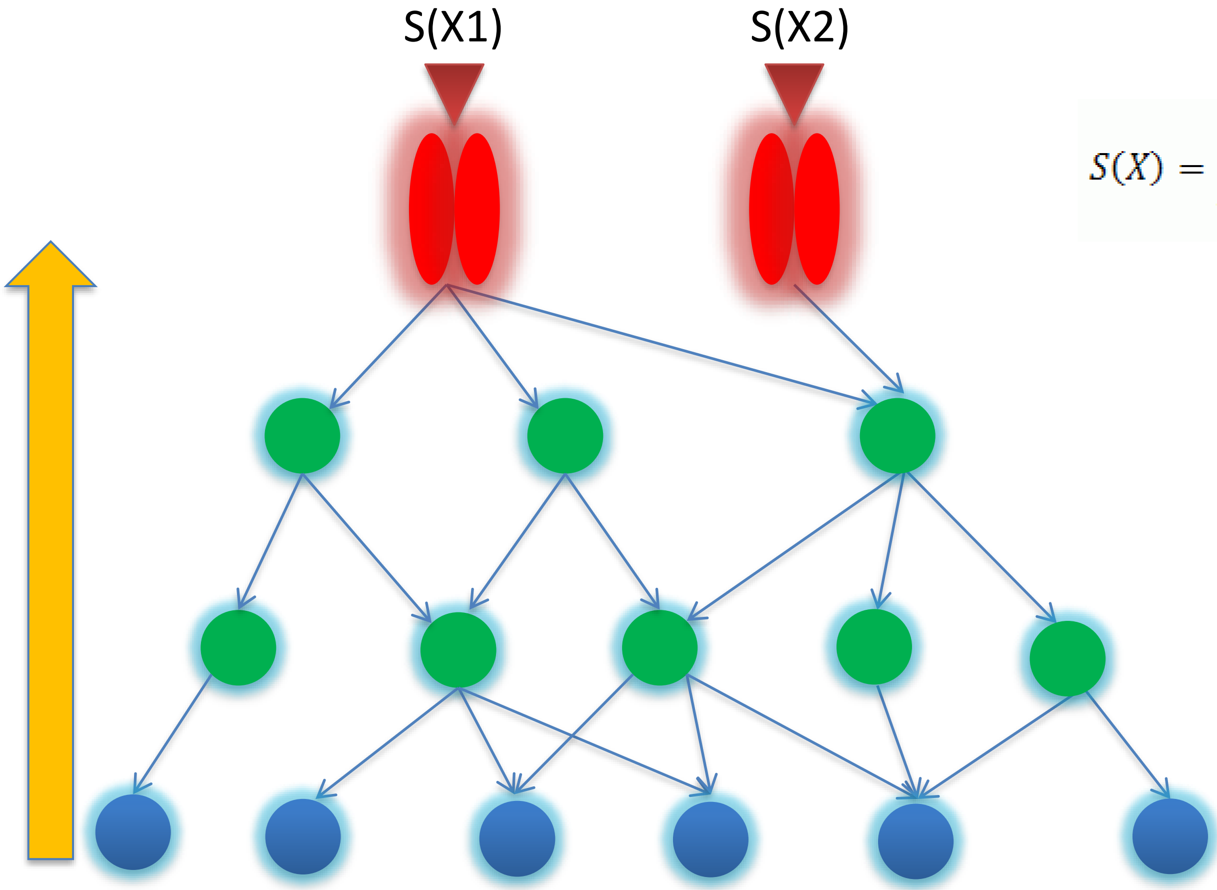


Betweenness centrality may point to important nodes in the network, e. g. proto-oncogenes in the transcriptional network.

Fraction of those shortest paths between all pairs of vertices *s* and *t* that pass through vertex *v*. Allows to quantify how influential a given TF-gene in a whole network is.



However, it suffers from focusing on only shortest path and ignoring the role of longer, alternative paths.

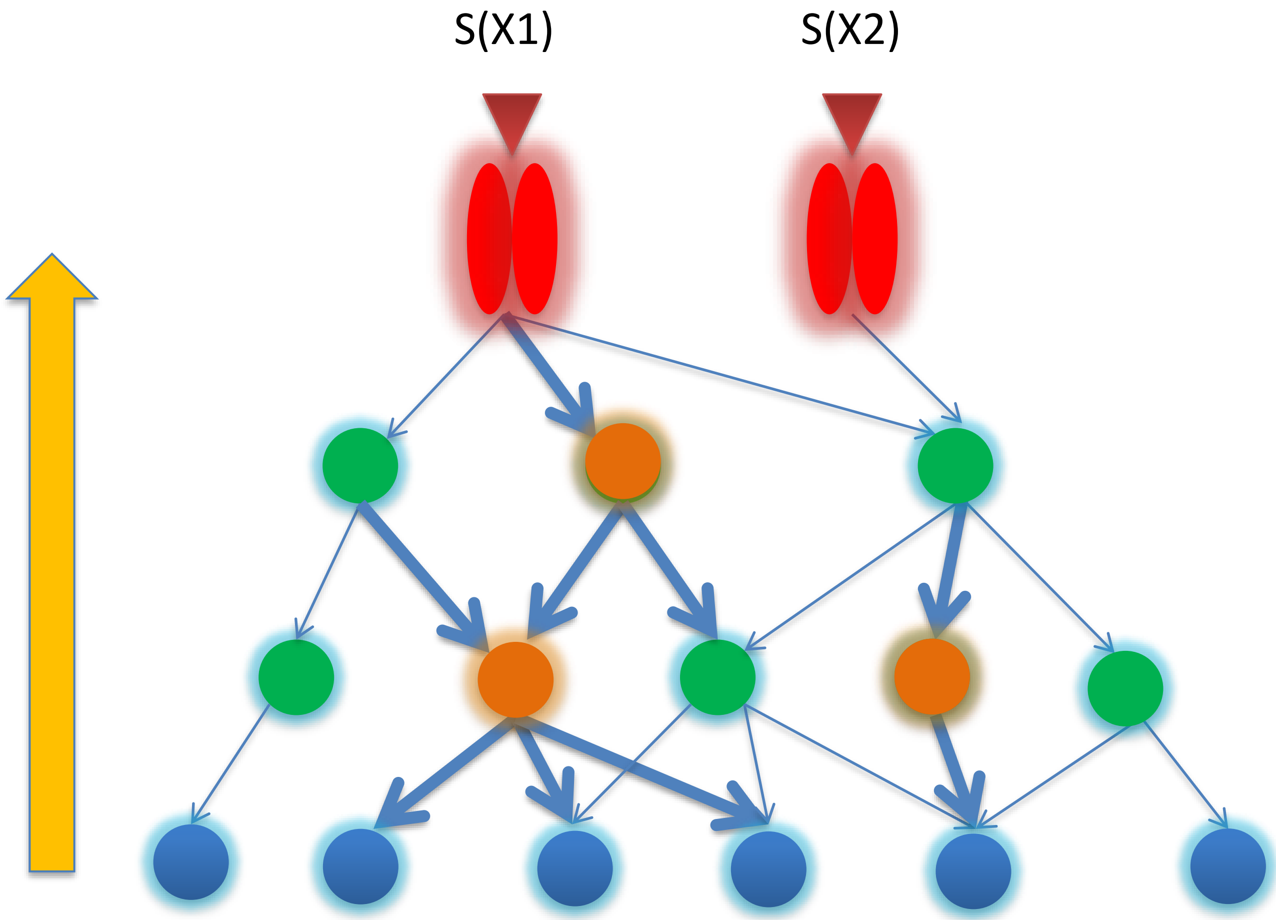


Score:

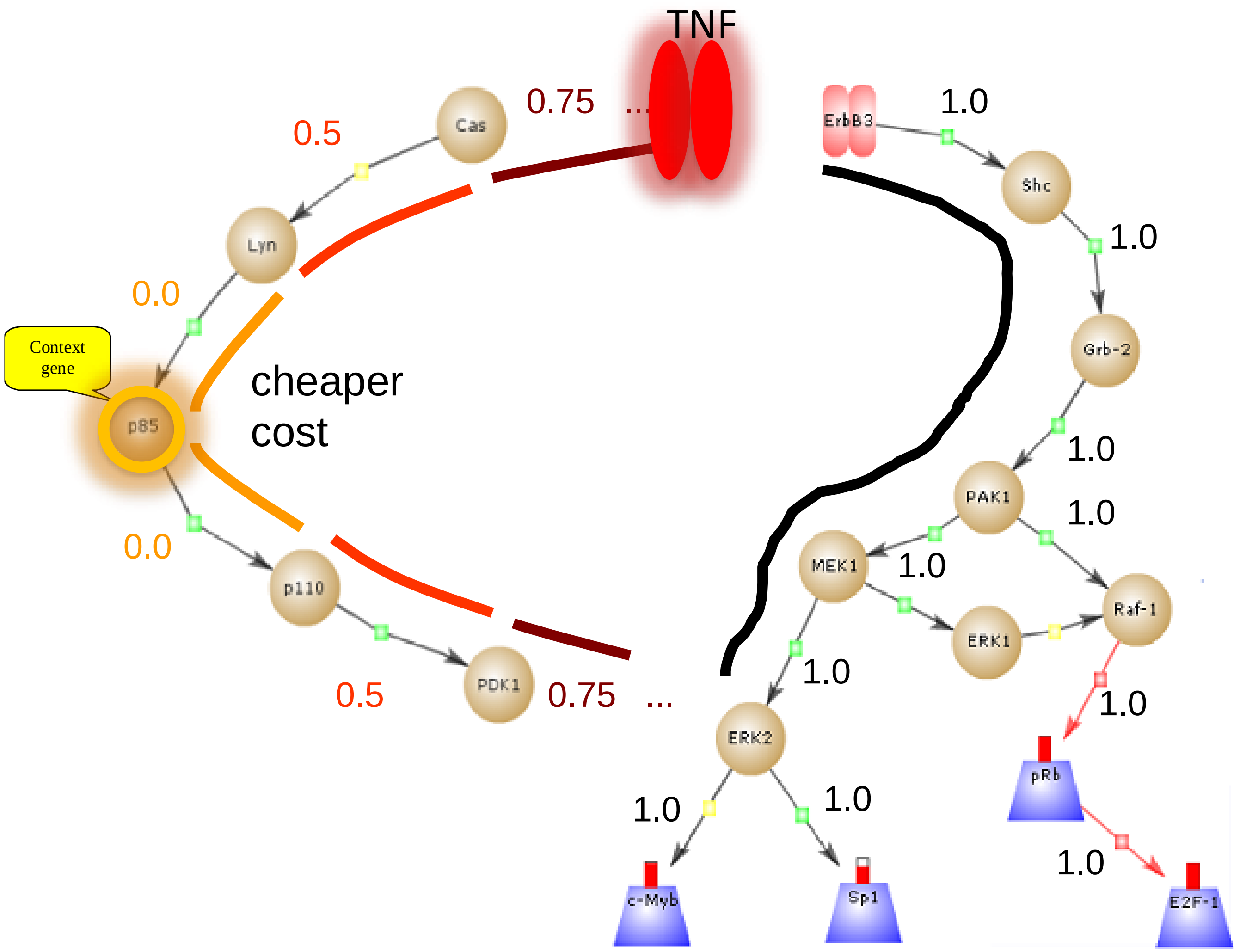
$$S(X) = \sum_{r=1}^{R} \frac{M(X,r)}{M_{max}(r)} \cdot \frac{1}{1 + pN(X,r)/N_{max}(r)}$$

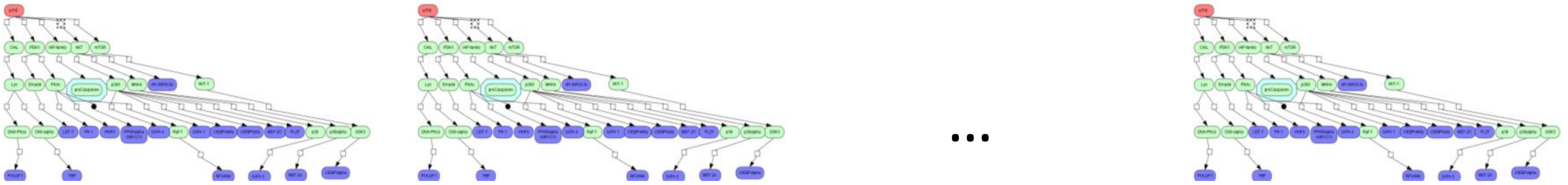
Where:
R - Max radius (input parameter)
p - Penalty (input parameter)
N(X,r) - total number of molecules reachable from key molecule X within the radius r.
$N_{max}(r)$ - maximal value of *N(X,r)* over all key molecules X found for this radius.
M(X,r) - sum of *w(X)* for all hits reachable from key molecule X within the radius r, where *w(X)* - weight of hit X.
$M_{max}(r)$ - maximal value of *M(X,r)* over all key molecules X found for this radius.

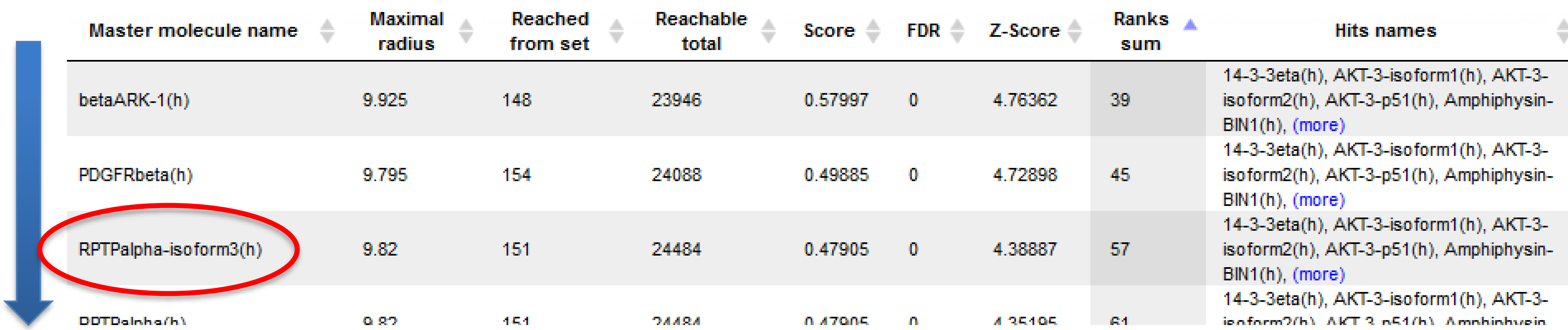
Use of context proteins (orange nodes) to calculate weights/costs of the edges.



The key-node search is attracted to the paths containing context proteins

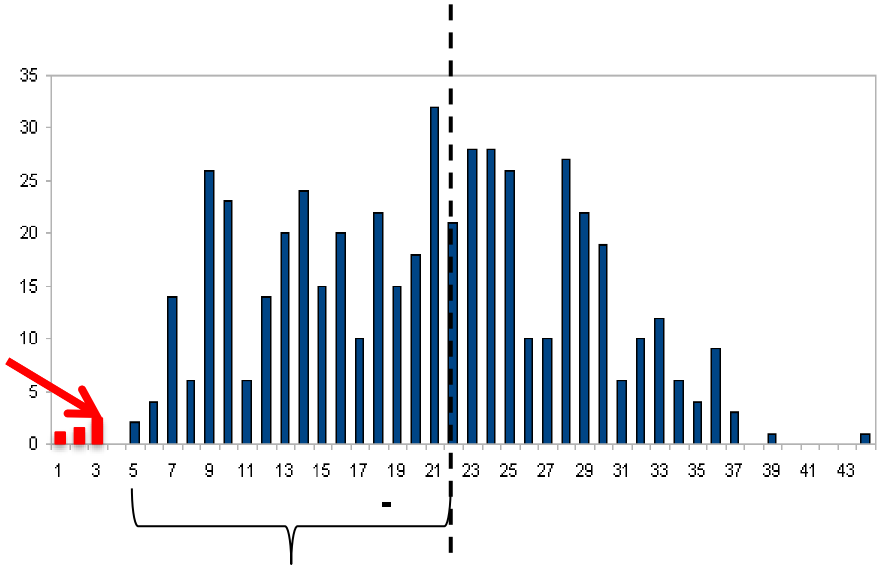






Master molecule name	Maximal radius	Reached from set	Reachable total	Score	FDR	Z-Score	Ranks sum	Hits names
betaARK-1(h)	9.925	148	23946	0.57997	0	4.76362	39	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
PDGFRbeta(h)	9.795	154	24088	0.49885	0	4.72898	45	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
RPTPalpha-isoform3(h)	9.82	151	24484	0.47905	0	4.38887	57	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
RPTPalpha(h)	9.82	151	24484	0.47905	0	4.35195	61	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin

Rank of the key node



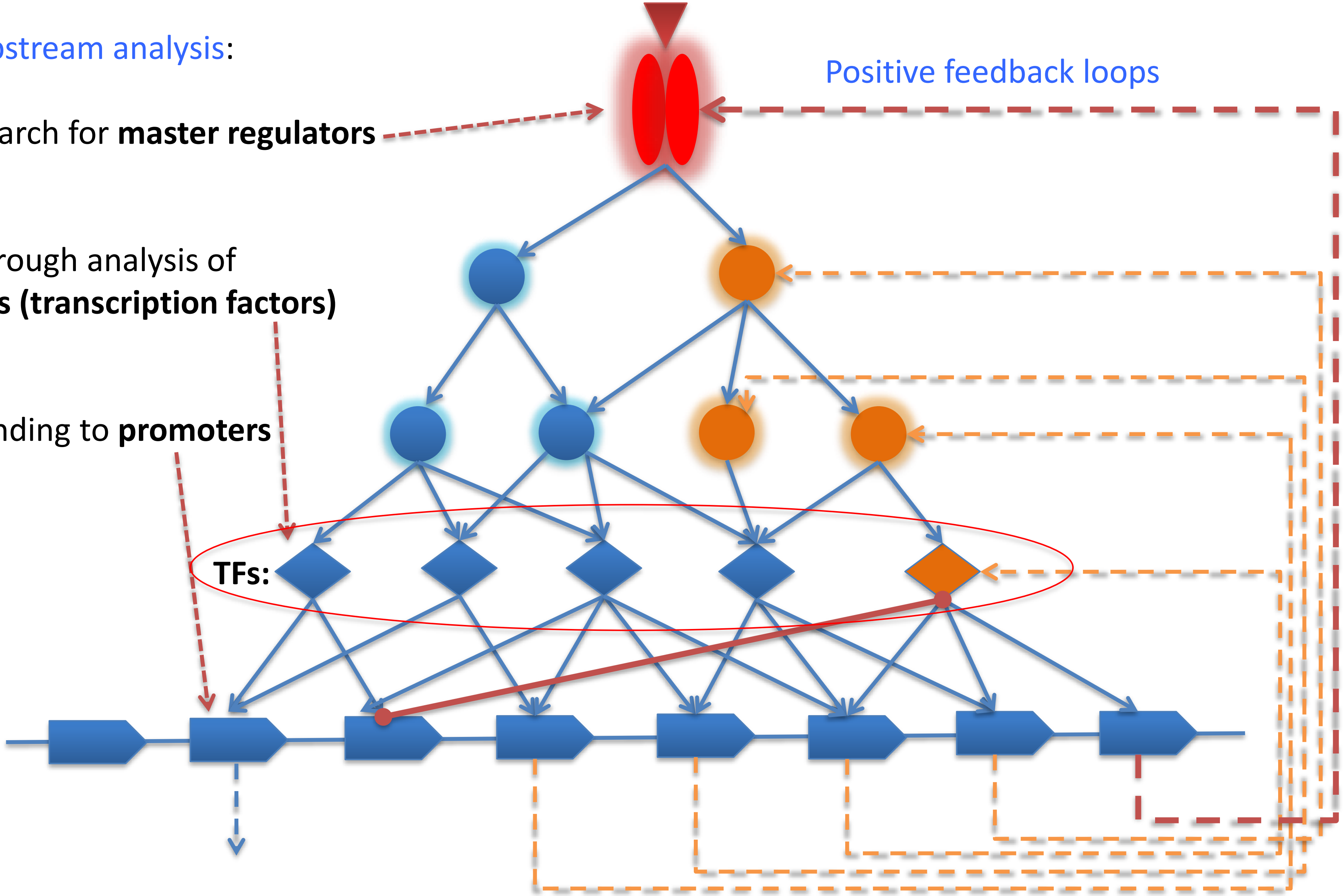
$$Z = \frac{X - \mu}{\sigma}$$

Upstream analysis:

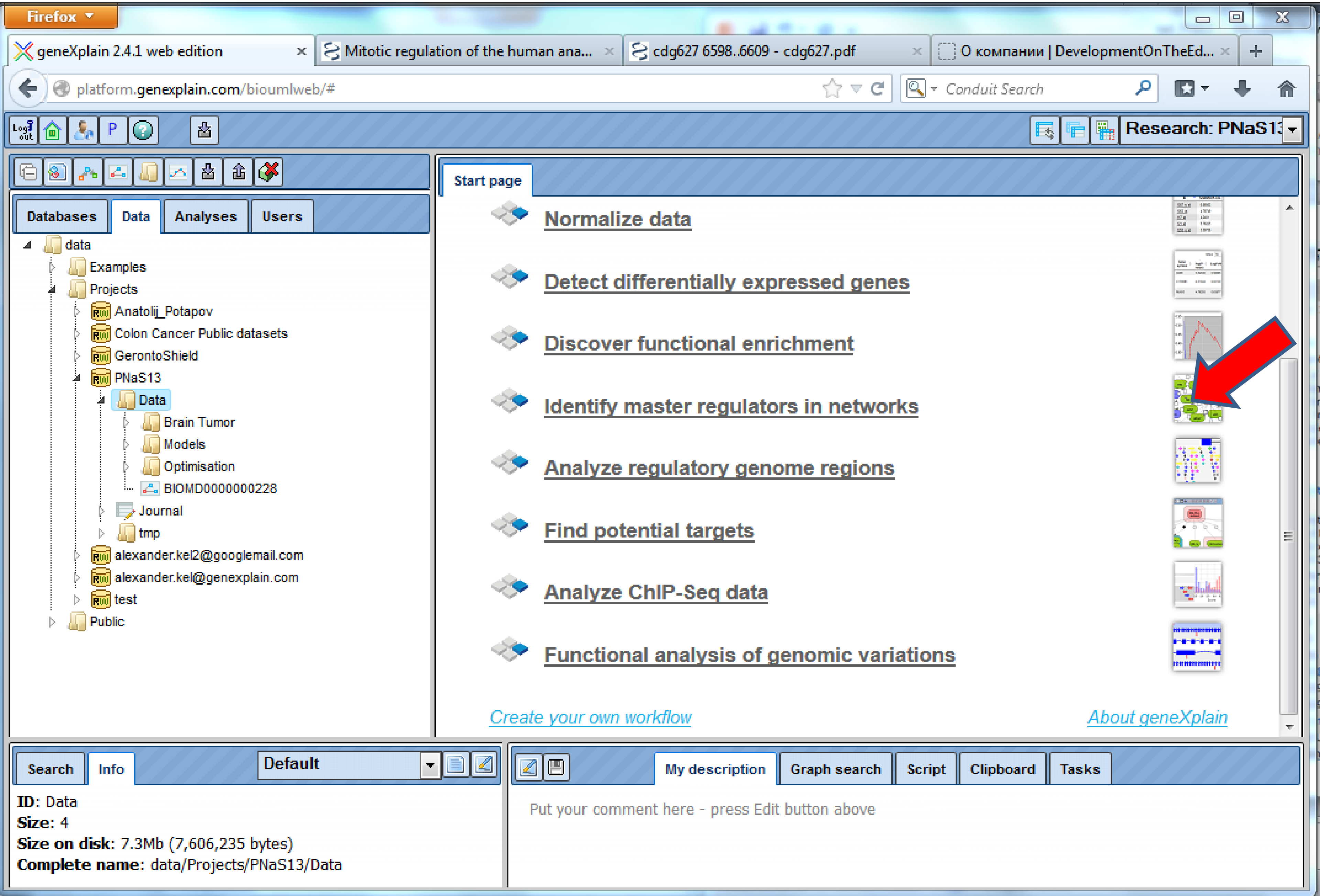
Search for **master regulators**

through analysis of
TFs (transcription factors)

binding to **promoters**



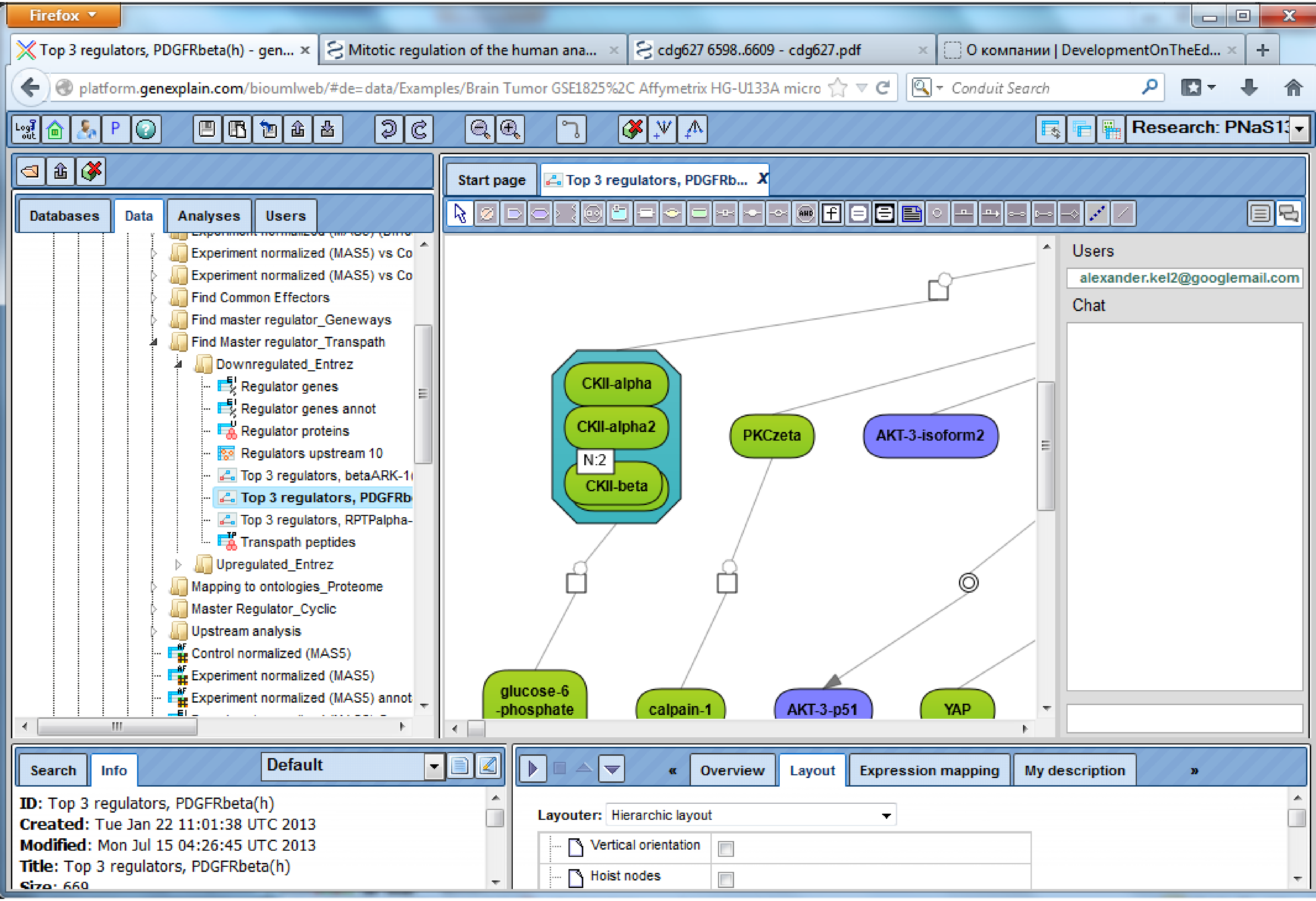
Master regulators in networks



ID: Data
Size: 4
Size on disk: 7.3Mb (7,606,235 bytes)
Complete name: data/Projects/PNaS13/Data

Put your comment here - press Edit button above

Master regulators in networks



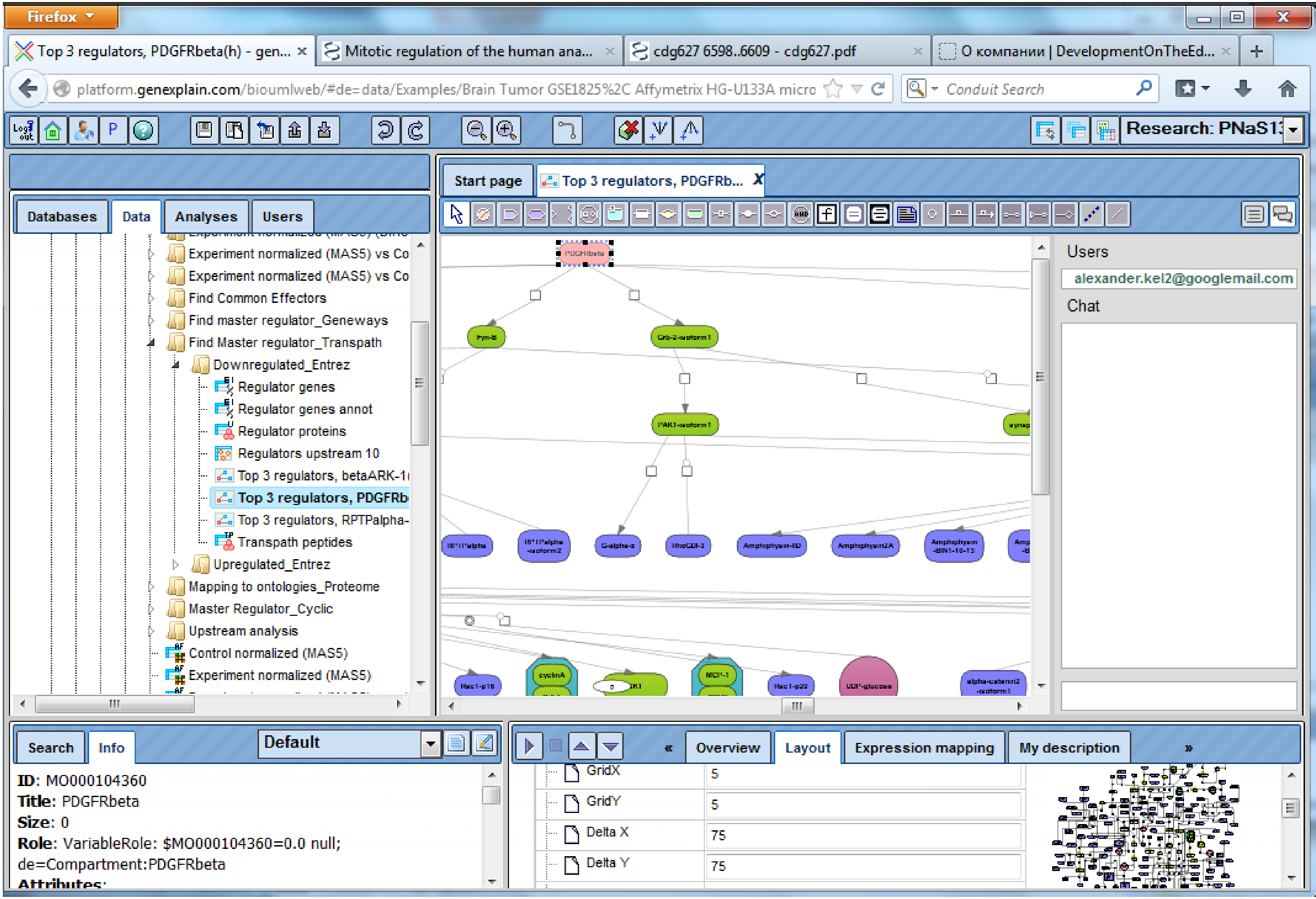
alexander.kel2@googlemail.com

ID: Top 3 regulators, PDGFRbeta(h)
Created: Tue Jan 22 11:01:38 UTC 2013
Modified: Mon Jul 15 04:26:45 UTC 2013
Title: Top 3 regulators, PDGFRbeta(h)
Size: 669

Layouter: Hierarchic layout

Vertical orientation ☐
Hoist nodes ☐

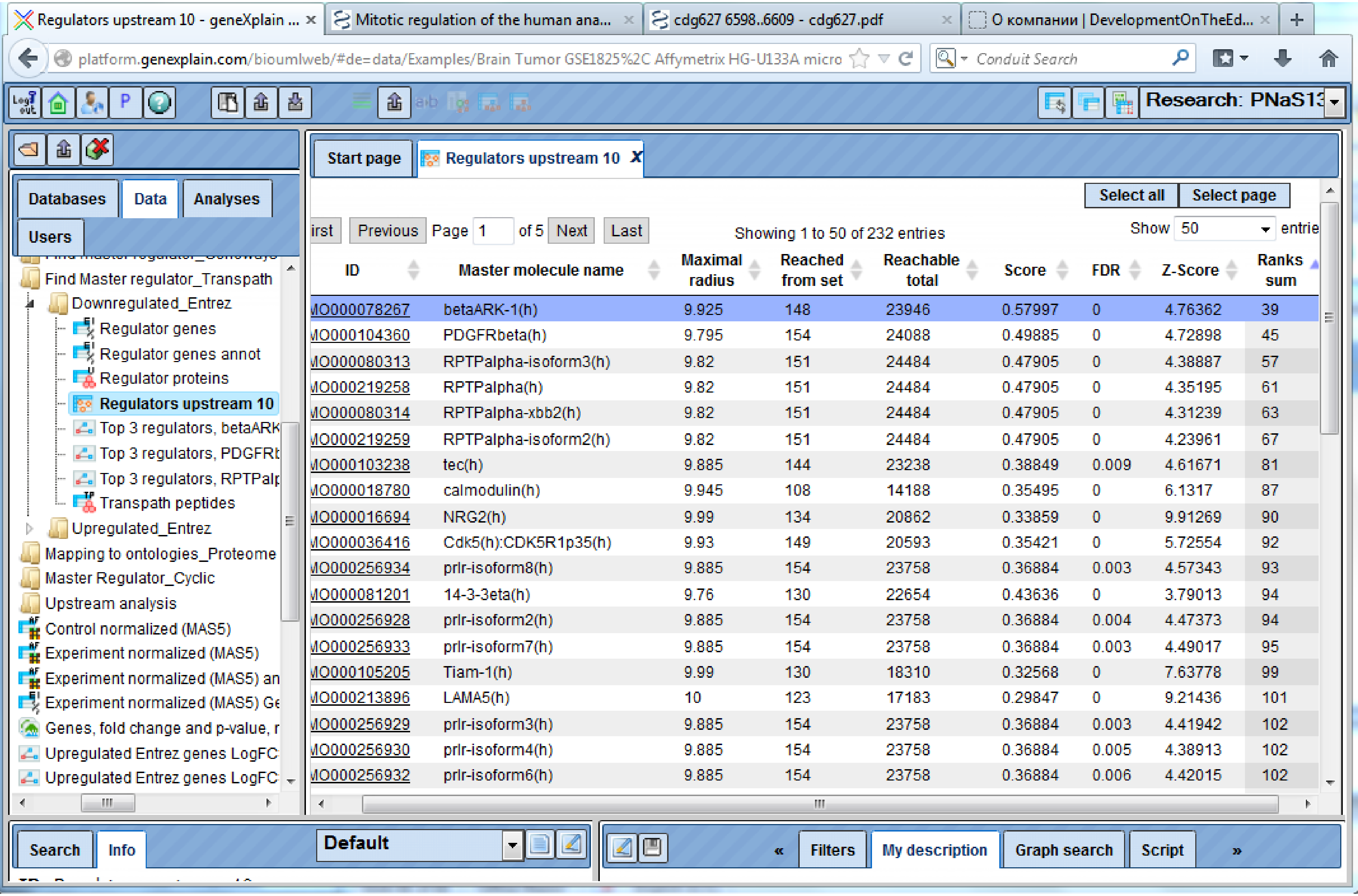
Master regulators in networks



ID: MO000104360
Title: PDGFRbeta
Size: 0
Role: VariableRole: $MO000104360=0.0 null;
de=Compartment:PDGFRbeta
Attributes:

GridX 5
GridY 5
Delta X 75
Delta Y 75

Master regulators in networks



Showing 1 to 50 of 232 entries

ID	Master molecule name	Maximal radius	Reached from set	Reachable total	Score	FDR	Z-Score	Ranks sum
MO000078267	betaARK-1(h)	9.925	148	23946	0.57997	0	4.76362	39
MO000104360	PDGFRbeta(h)	9.795	154	24088	0.49885	0	4.72898	45
MO000080313	RPTPalpha-isoform3(h)	9.82	151	24484	0.47905	0	4.38887	57
MO000219258	RPTPalpha(h)	9.82	151	24484	0.47905	0	4.35195	61
MO000080314	RPTPalpha-xbb2(h)	9.82	151	24484	0.47905	0	4.31239	63
MO000219259	RPTPalpha-isoform2(h)	9.82	151	24484	0.47905	0	4.23961	67
MO000103238	tec(h)	9.885	144	23238	0.38849	0.009	4.61671	81
MO000018780	calmodulin(h)	9.945	108	14188	0.35495	0	6.1317	87
MO000016694	NRG2(h)	9.99	134	20862	0.33859	0	9.91269	90
MO000036416	Cdk5(h):CDK5R1p35(h)	9.93	149	20593	0.35421	0	5.72554	92
MO000256934	prlr-isoform8(h)	9.885	154	23758	0.36884	0.003	4.57343	93
MO000081201	14-3-3eta(h)	9.76	130	22654	0.43636	0	3.79013	94
MO000256928	prlr-isoform2(h)	9.885	154	23758	0.36884	0.004	4.47373	94
MO000256933	prlr-isoform7(h)	9.885	154	23758	0.36884	0.003	4.49017	95
MO000105205	Tiam-1(h)	9.99	130	18310	0.32568	0	7.63778	99
MO000213896	LAMA5(h)	10	123	17183	0.29847	0	9.21436	101
MO000256929	prlr-isoform3(h)	9.885	154	23758	0.36884	0.003	4.41942	102
MO000256930	prlr-isoform4(h)	9.885	154	23758	0.36884	0.005	4.38913	102
MO000256932	prlr-isoform6(h)	9.885	154	23758	0.36884	0.006	4.42015	102