networks in molecular biology

Wolfgang Huber

networks in molecular biology

Regulatory networks:

components = gene products

interactions = regulation of transcription, translation, phosphorylation...

Metabolic networks:

components = metabolites, enzymes

interactions = chemical reactions

Interaction networks:

components = proteins

interactions = ability to form a complex

... and so on

What might people mean with the term „network“?

Representation of experimental data

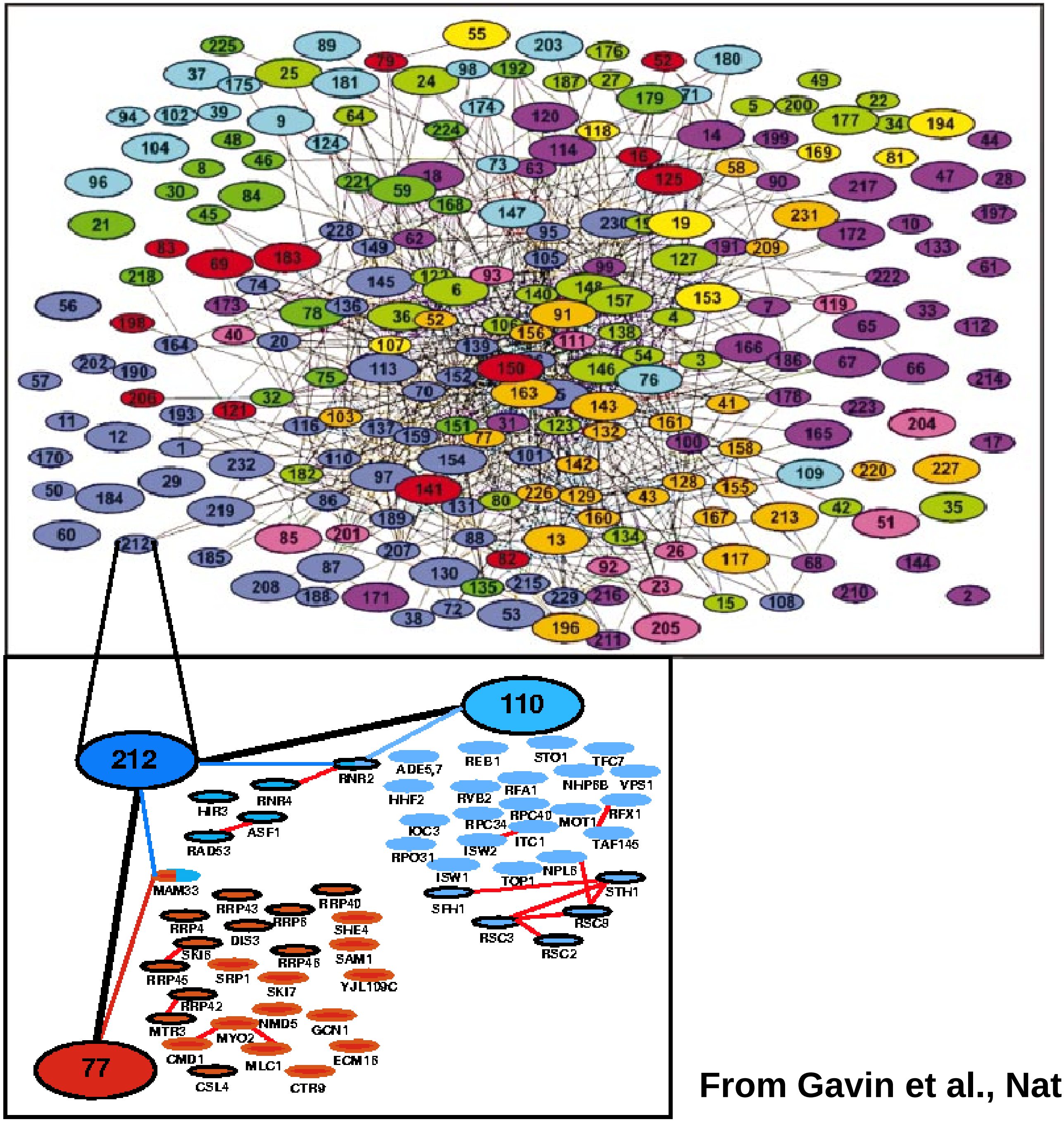
a convenient way to visualize experimentally observed protein-protein interactions or correlated occurences of events

Map

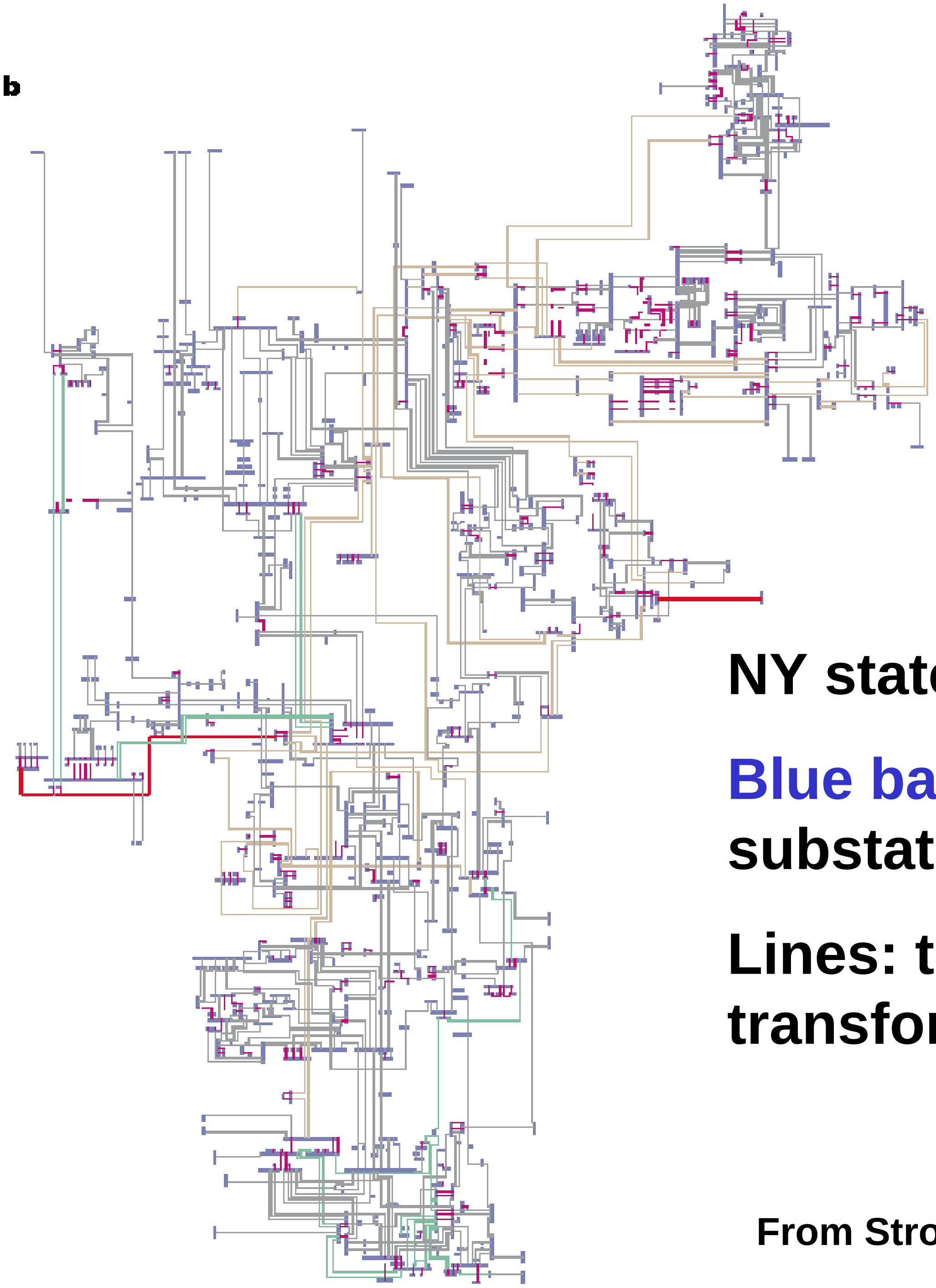
a visual tool to navigate through the world of gene products, proteins, domains, etc.

Predictive Model

like an electronic circuit: a complete description of causal connections that allows to predict and engineer the behavior of a biological system, like that of a radio receiver



From Gavin et al., Nature 415, 141 (2002)

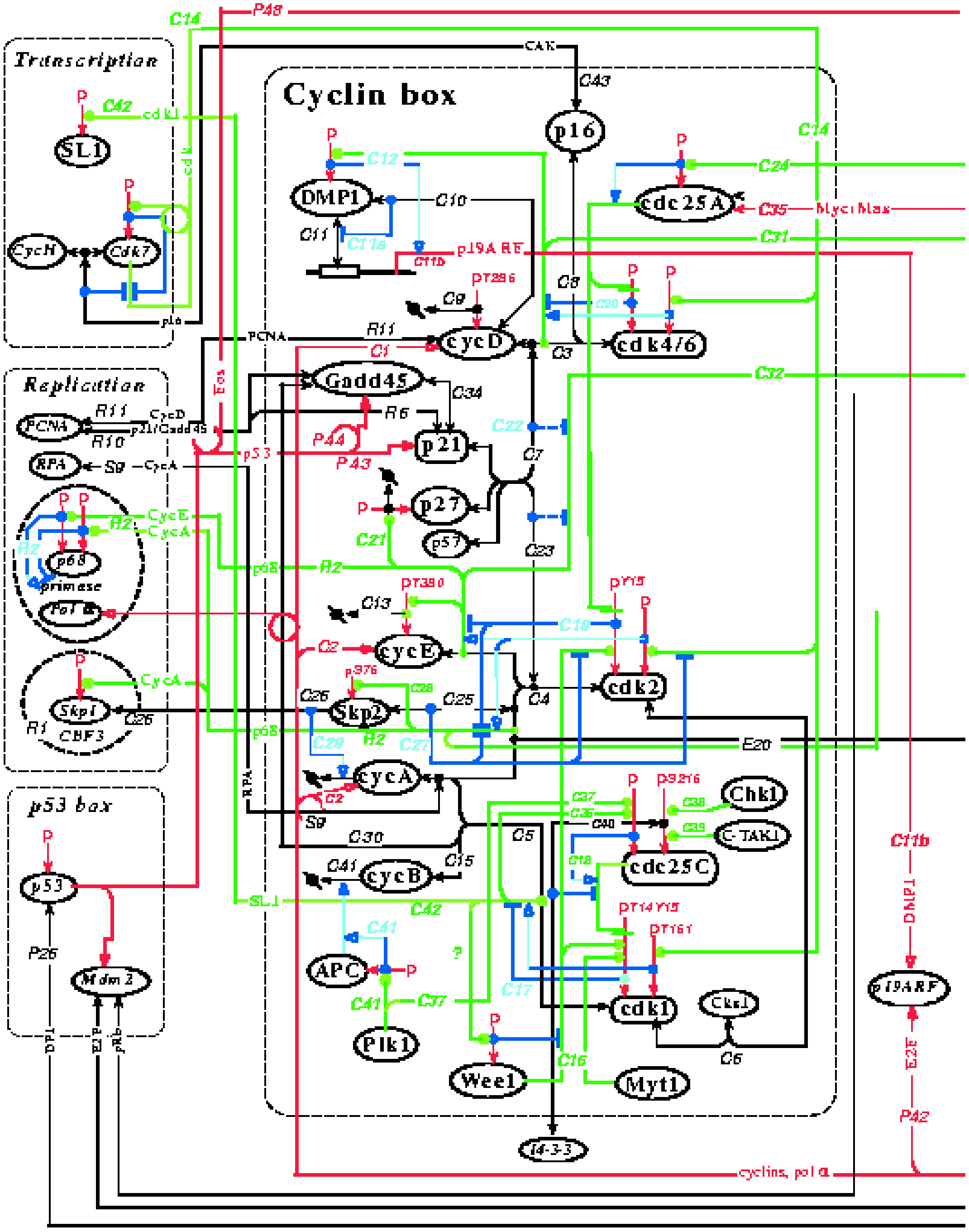


NY state electric power grid

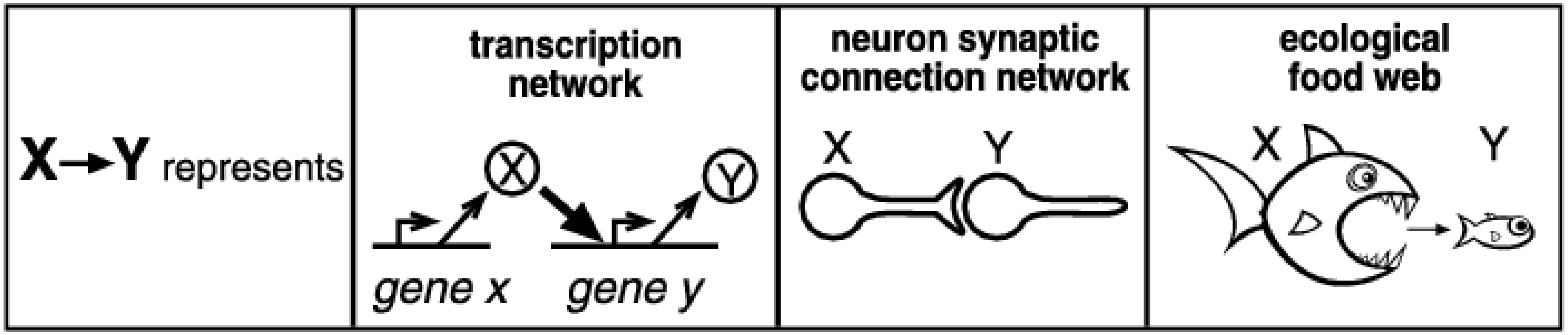
Blue bars: generators, substations

Lines: transmission lines, transformers

From Strogatz, Nature 410, 268 (2001)



Molecular interaction network for mammalian cell cycle regulation


X→Y represents
transcription network
X
Y
gene x
gene y
neuron synaptic connection network
X
Y
ecological food web
X
Y

Aspects of complicatedness

- Structural complexity (topology)
- Evolution over time
- Connection diversity: weights, directions, function
- Dynamical complexity: nodes themselves can already be complicated dynamical systems
- Node diversity

All of these complications can influence each other

Mathematical tools

- o Structural complexity: graph theory
- o Dynamical complexity: calculus, theory of dynamical systems, chaos theory

- o Connection diversity, node diversity : differential equations, graphical networks
- o Evolution over time: few ideas

All together: computer simulation, data analysis, ... very little hard results, but lots of excitement.

Simple network topologies

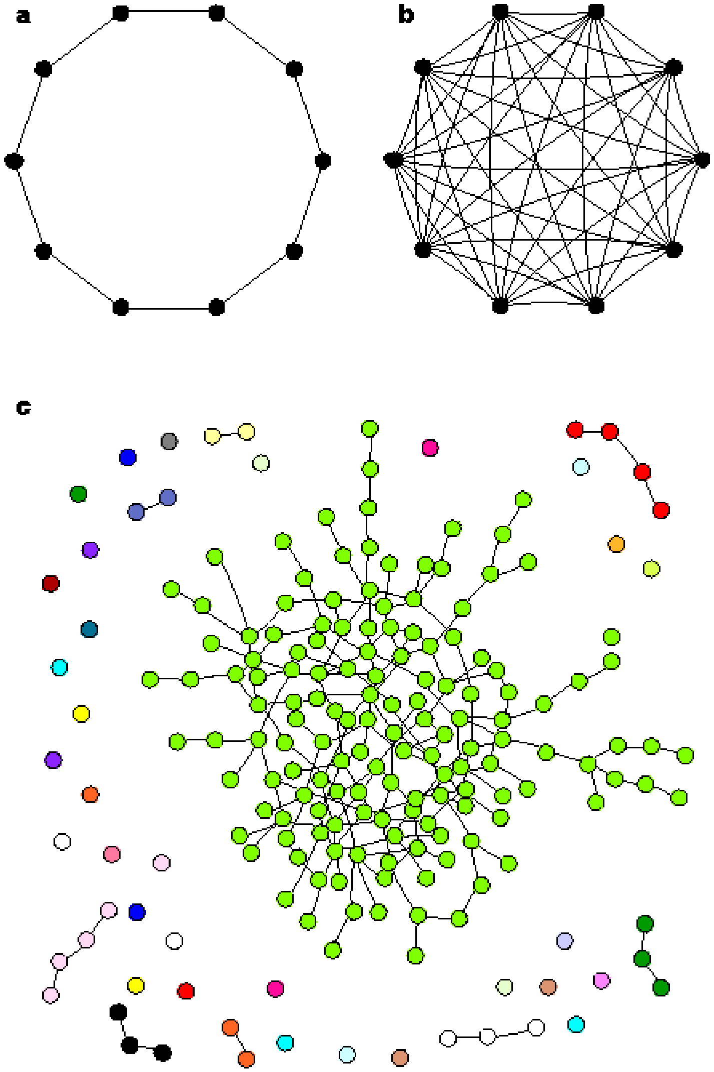
Regular nearest neighbour: 1d, 2d, 3d, ...

All-to-all

Random graph

Scale free

Network topologies



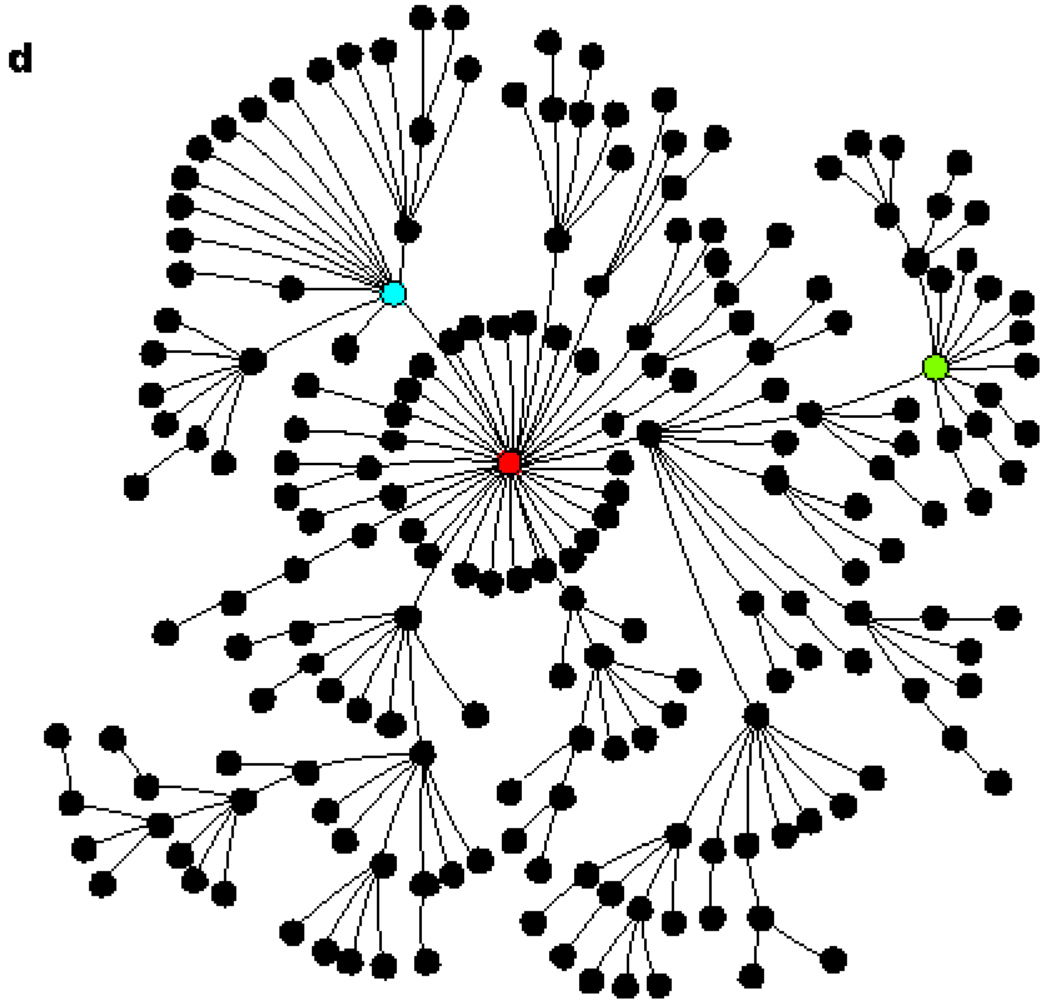
regular

all-to-all

Random graph

(after "tidy" rearrangement of nodes)

Network topologies



Scale-free

(Albert/Barabasi-model)

Graphs

Graph := set of nodes + set of edges

Edges can be

- **directed**
- **undirected**
- **weighted**

special cases: cycles, acyclic graphs, trees

Random Edge Graphs

n nodes, m edges

$p(i,j) = 1/m$

with high probability:

m < n/2: many disconnected components

m > n/2: one giant connected component: size ~ n.

(next biggest: size ~ log(n)).

degrees of separation: log(n).

Erdös and Rényi 1960

Some important concepts:

Small worlds

Clustering

Degree distribution

Motifs

Small word networks

typical path length („degrees of separation“) is short

Many examples:

- **Communications**
- **Epidemiology / infectious diseases**
- **Metabolic networks**
- **Scientific collaboration networks**
- **WWW**
- **Company ownership in Germany**
- **„6 degrees from Kevin Bacon“**

Clustering

Clique: every node connected to everyone else

Clustering coefficient:

$$c = \frac{\text{no. edges between first-degree neighbors}}{\text{maximum possible number of such edges}}$$

Random network: c=p

Real networks: c » p

Degree distributions

p(k) = fraction of nodes that have *k* edges

Random graph: p(k) = Poisson distribution with some parameter λ („scale“)

Many real networks: p(k) = power law,

$$p(k) \sim k^{-\gamma}$$

„scale-free“

(WWW: Yahoo, metabolic network: ATP)

Other distributions: exponential, Gaussian

Growth models for scale free networks

Start out with one node and continously add nodes, with preferential attachment to existing nodes, with probability ~ degree of target node.

$$\Rightarrow p(k) \sim k^{-3}$$

(Simon 1955; Barabási, Albert, Jeong 1999)

Modifications to obtain $\gamma \neq 3$:

Through different rules for adding or rewiring of edges, can tune to obtain any kind of degree distribution

What are the functional advantages („evolutionary fitness“) of scale free networks ?

Robustness: only a few hubs, so insensitive to the failure of most nodes

However: sensitive to the failure of hubs

Real networks

- tend to have power-law scaling (truncated)
- are ‚small worlds‘ (like random networks)
- have a high clustering coefficient independent of network size (like lattices and unlike random networks)

Network motifs

:= pattern that occurs more often than in randomized networks

Intended implications

duplication: useful building blocks are reused by nature

there may be evolutionary pressure for **convergence** of network architectures

Network motifs

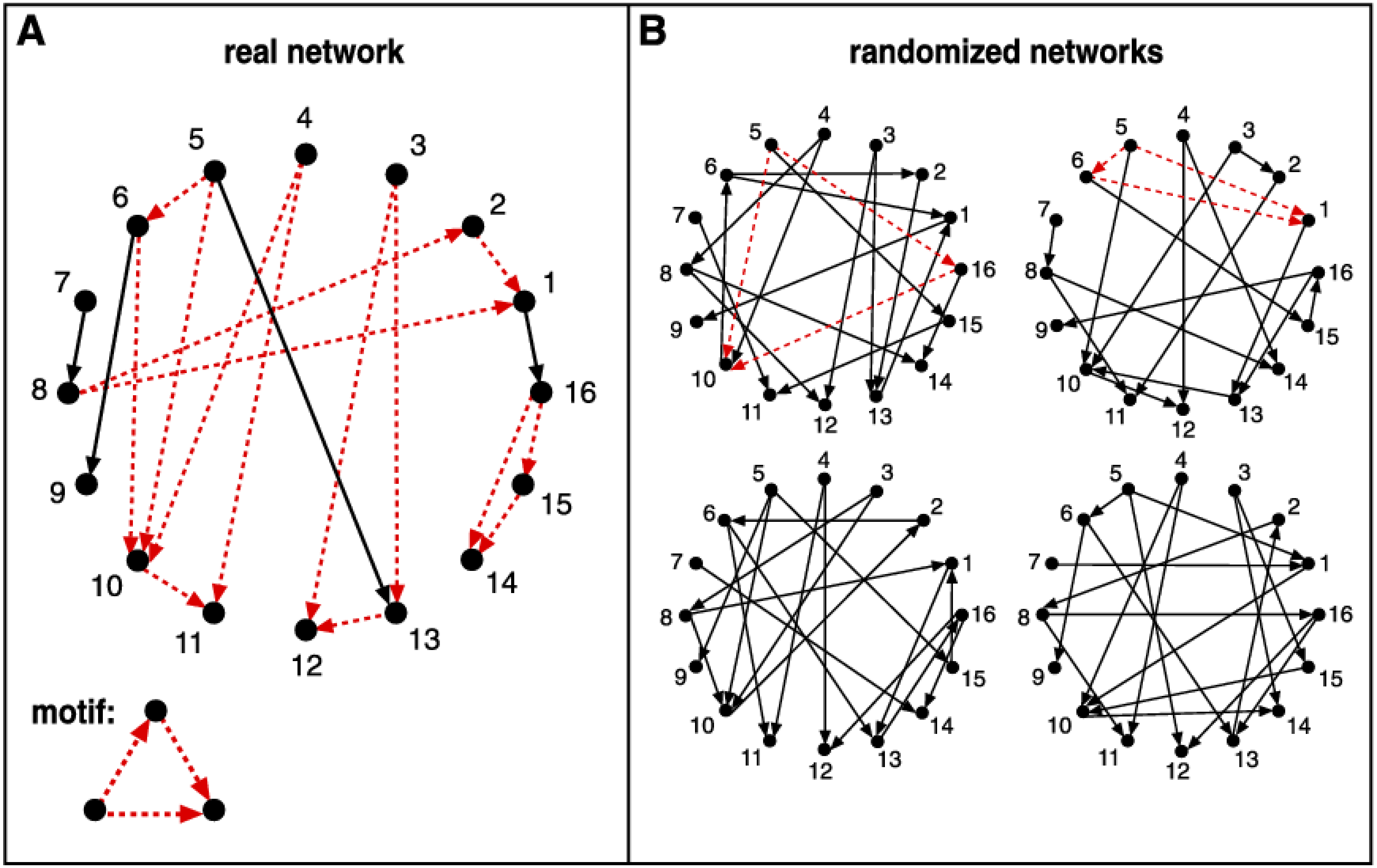
Starting point: graph with directed edges

Scan for n-node subgraphs (n=3,4) and count number of occurence

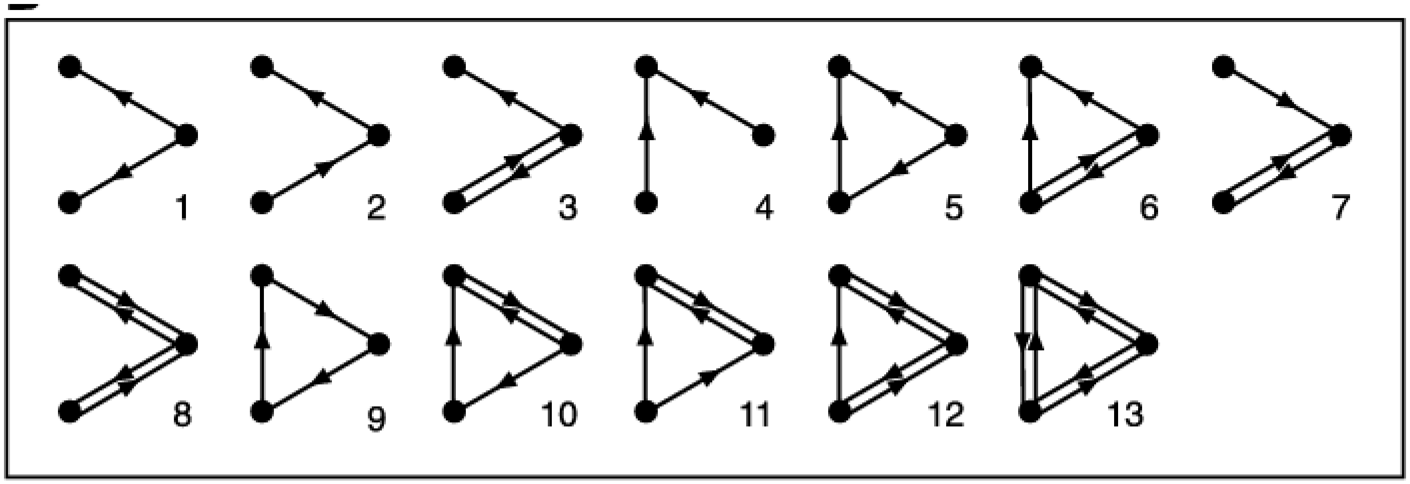
Compare to randomized networks

(randomization preserves in-, out- and in+out- degree of each node, and the frequencies of all (n-1)-subgraphs)

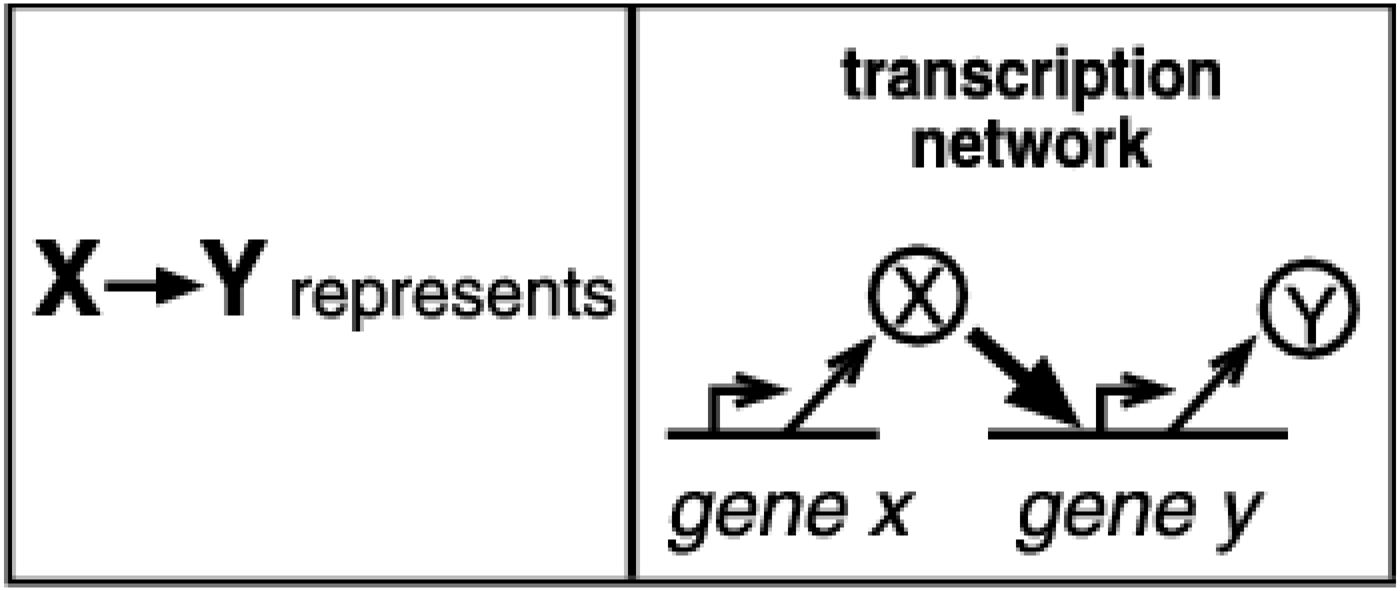
Schematic view of motif detection



All 3-node connected subgraphs



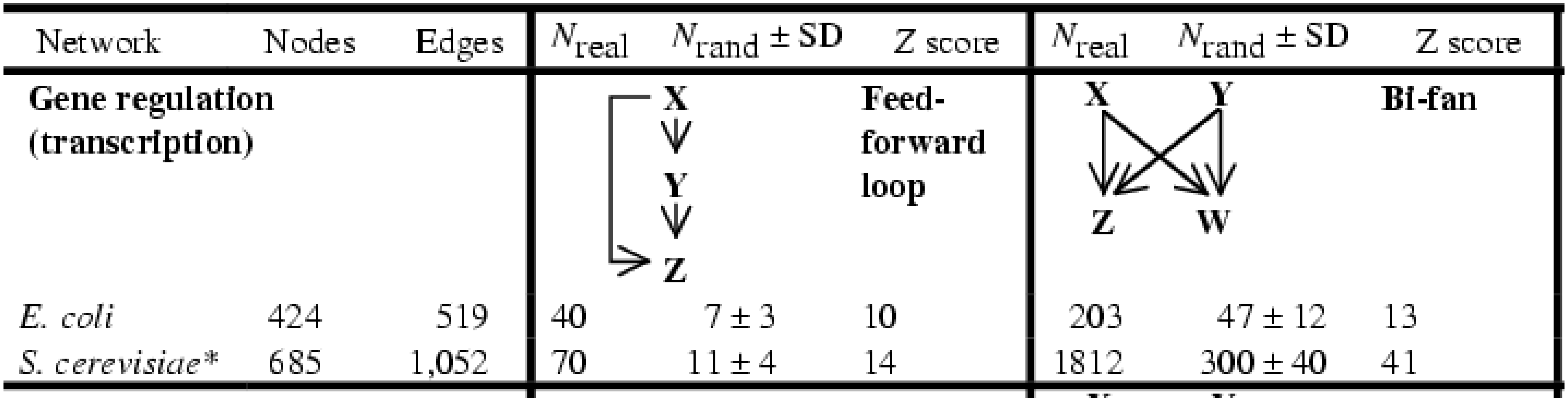
Transcription networks



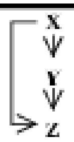
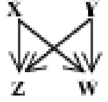
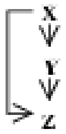
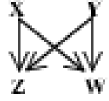
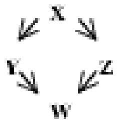
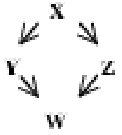
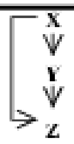
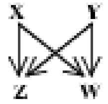
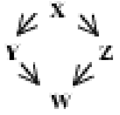
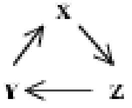
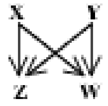
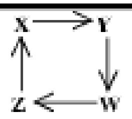
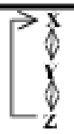
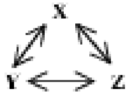
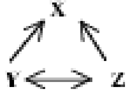
Nodes = transcription factors

Directed edge: X regulates transcription of Y

3- and 4-node motifs in transcription networks



Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41

Network	Nodes	Edges	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

System-size dependence

Extensive variable: proportional to system size. E.g. mass, diameter, number of molecules

Intensive variable: independent of system size. E.g. temperature, pressure, density, concentration

„Vanishing variable“: decreases with system size. E.g. Heat loss through radiation; in a city, probability to bump into one particular person

Alon et al.: In real networks, number of occurences of a motif is **extensive**. In randomized networks, it is non-extensive.

Examples

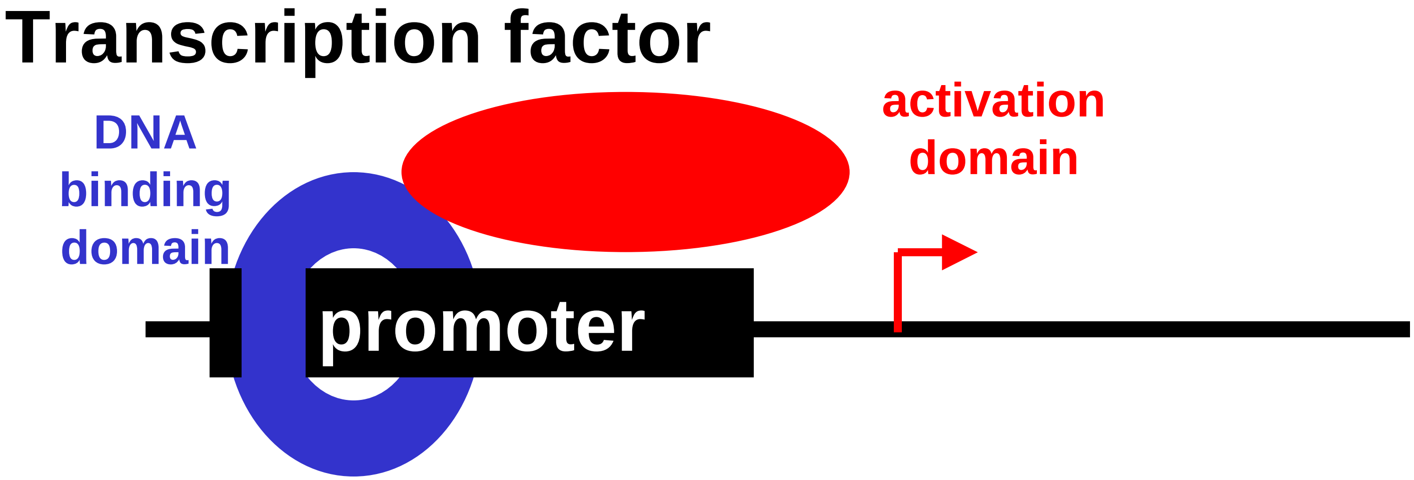
Yeast-2-Hybrid

Rosetta compendium

Regulatory networks

TAP: Tandem-Affinity-Purification

Two-hybrid screen

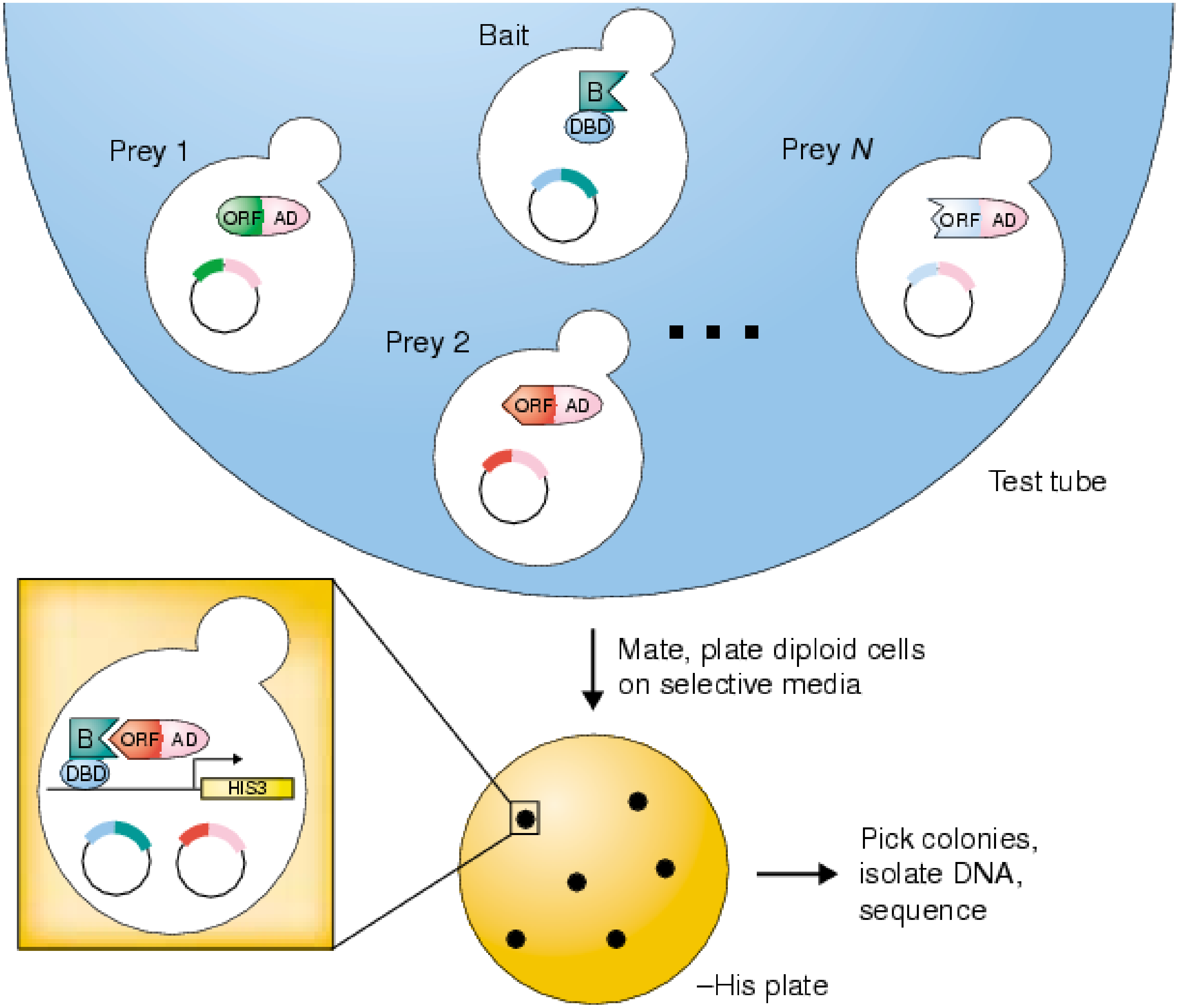


Idea:

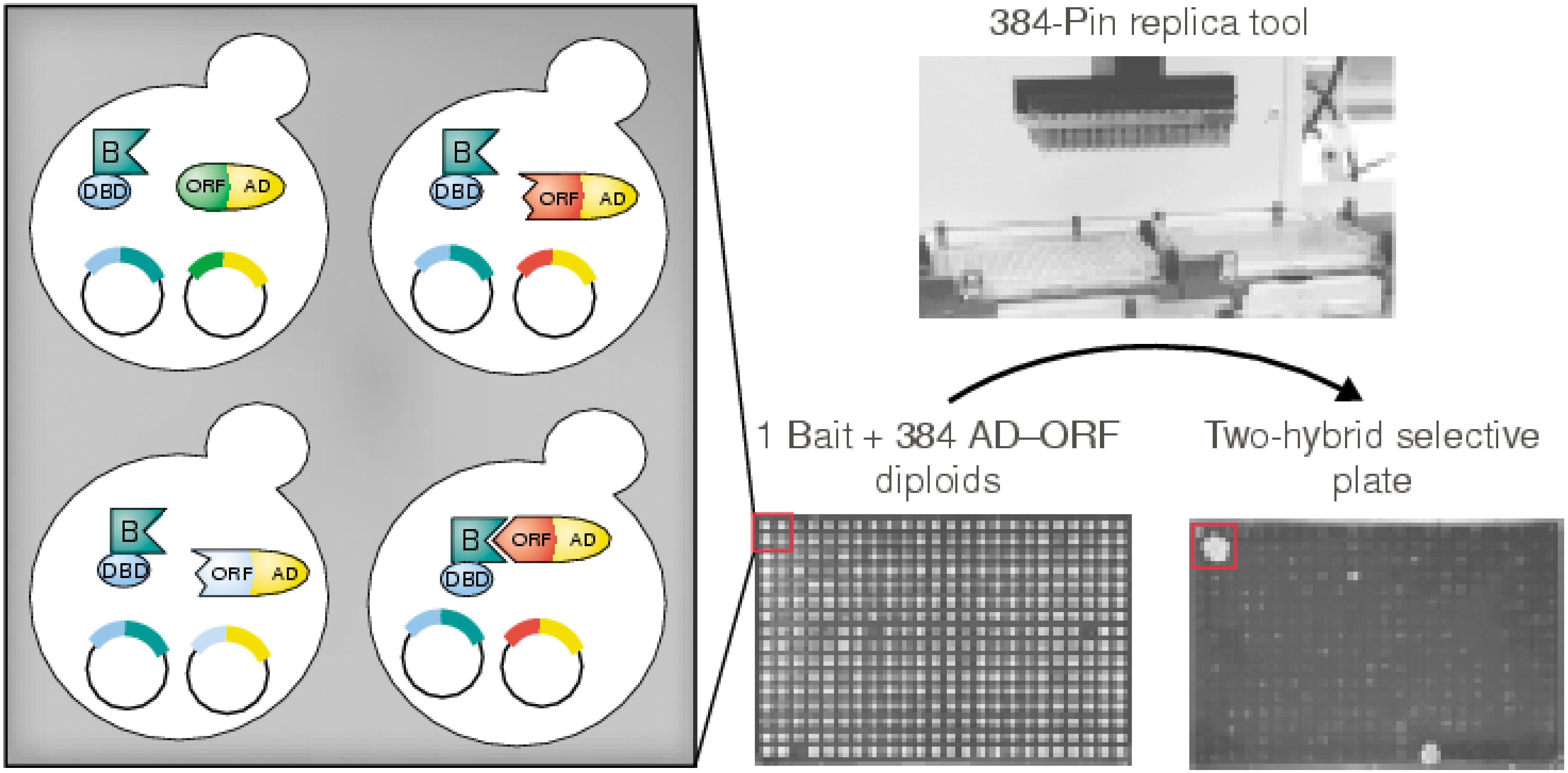
„Make potential pairs of interacting proteins a transcription factor for a reporter gene“

Two-hybrid screen

(a)



Two-hybrid arrays



Current Opinion in Chemical Biology

Colony array:

each colony expresses a defined pair of proteins

Table 2.

Two-hybrid array screens discussed in this paper.

Organism	Project	Proteins*	Assays*	Interactions*	Refs
Drosophila	Cell cycle proteins	13	45	19	[7]
C. elegans	Vulva development	29	841	8[†]	[9]
Mouse	Whole-genome pilot	~3500	$\sim 12\times 10^6$	145	[15••]
HCV	Whole genome	10	~100	0/3[‡]	[16]
Vaccinia	Whole genome	266	~64 000	37[§]	[17•]
Yeast	One by one array	192	~1 150 000	281	[18••]
Yeast	Pool by pool	~6000	~36 000 000	4549/841[‡]	[19,20••]
Yeast	Cell polarity	68	~408 000	191[#]	[10]
Yeast	Proteasome	31	~186 000	55	[12]

Sensitivity, specificity and reproducibility

Specificity – false positives: **the experiment reports an interaction even though is really none**

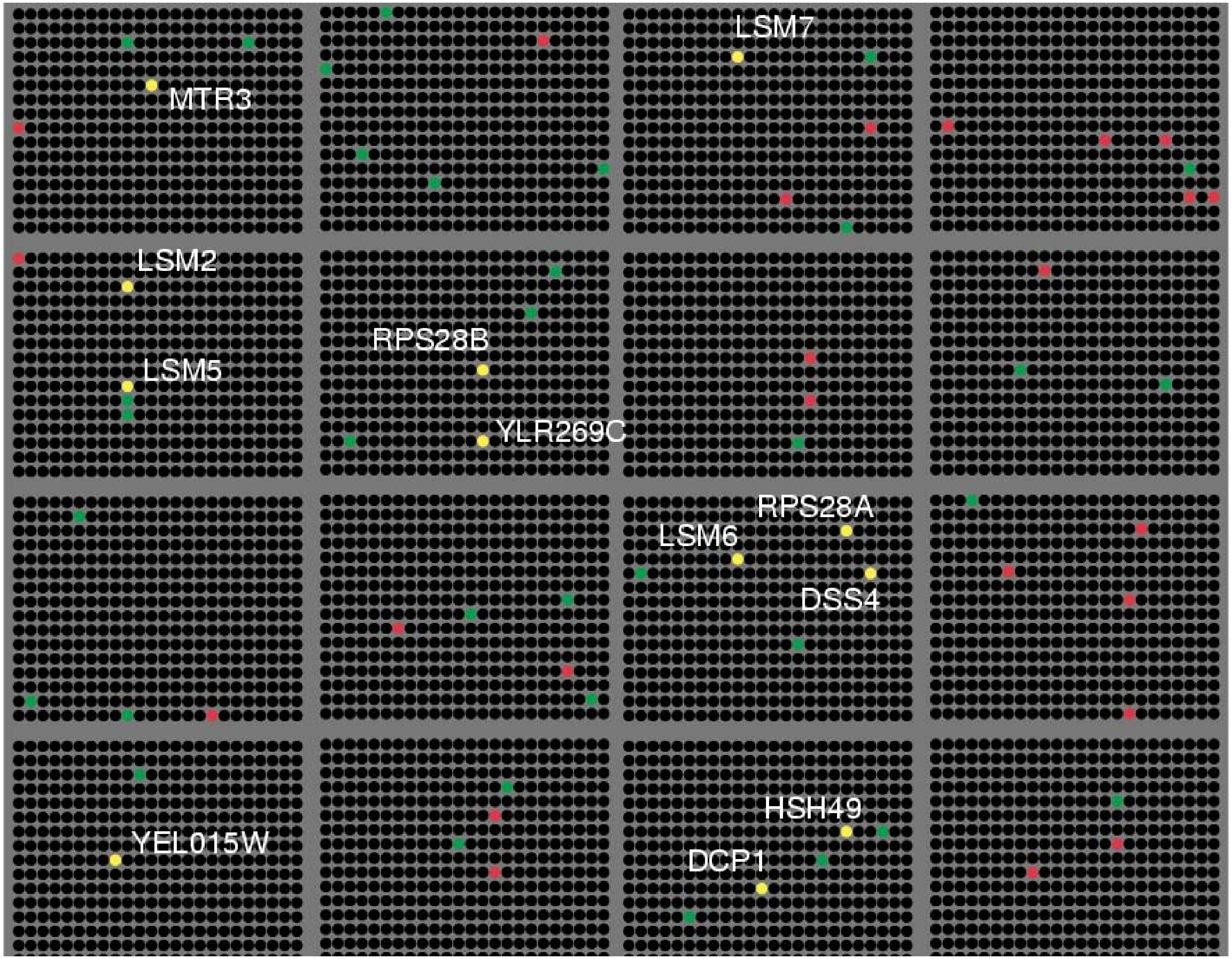
Sensitivity – false negatives: **the experiment reports no interaction even though is really one**

Problem: **what is the objective definition of an interaction?**

(Un)reproducibility: **the experiment reports different results when it is repeated**

„The molecular reasons for that are not really understood...“ (Uetz 2001)

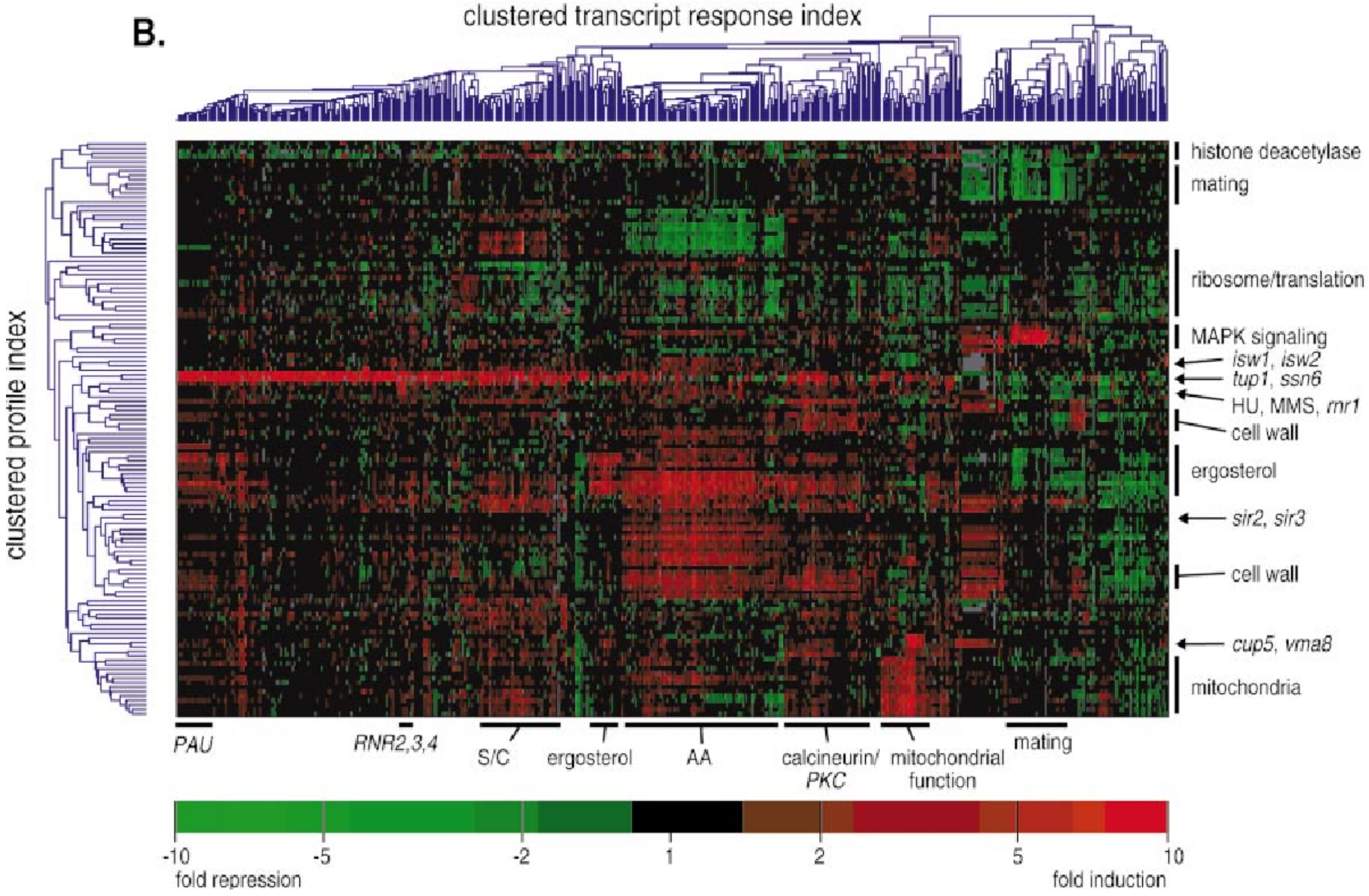
Reproducibility



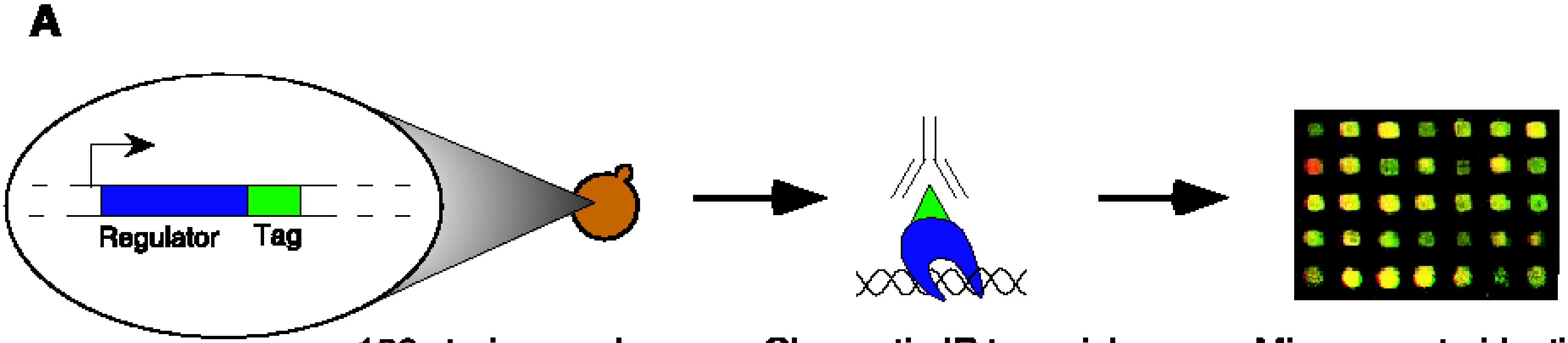
Rosetta compendium

568 transcript levels

300 mutations or chemical treatments



Transcriptional regulatory networks from "genome-wide location analysis"



106 strains, each with a tagged regulator

Chromatin IP to enrich promoters bound by regulator in vivo

Microarray to identify promoters bound by regulator in vivo

regulator := a transcription factor (TF) or a ligand of a TF

tag: c-myc epitope

106 microarrays

samples: enriched (tagged-regulator + DNA-promoter)

probes: cDNA of all promoter regions

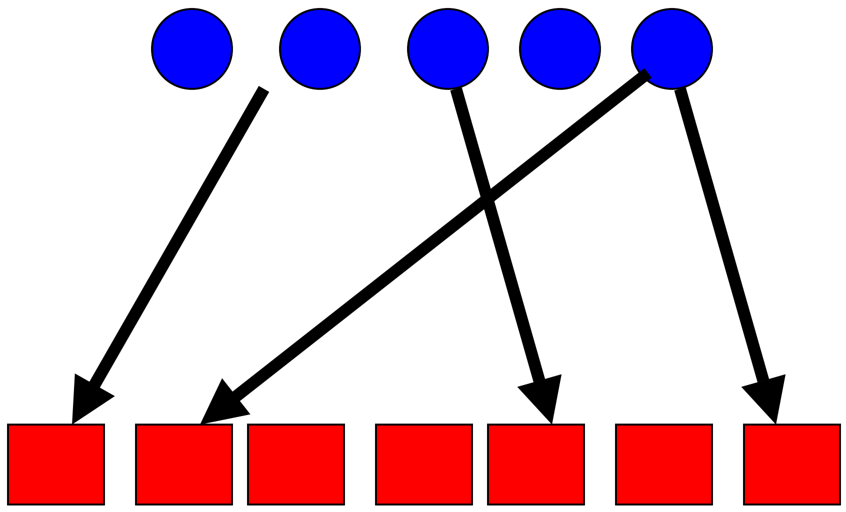
spot intensity ~ affinity of a promotor to a certain regulator

Transcriptional regulatory networks bipartite graph

106 regulators (TFs)

6270 promoter regions

							1
		1					
			1				
						1	
				1			
	1						
						1	

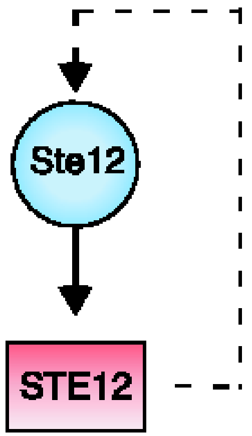


regulators

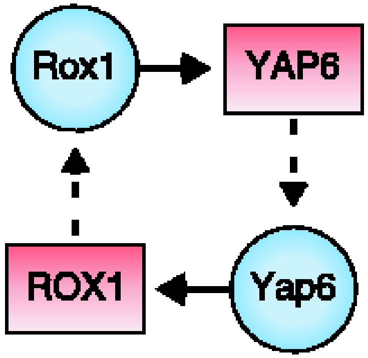
promoters

Network motifs

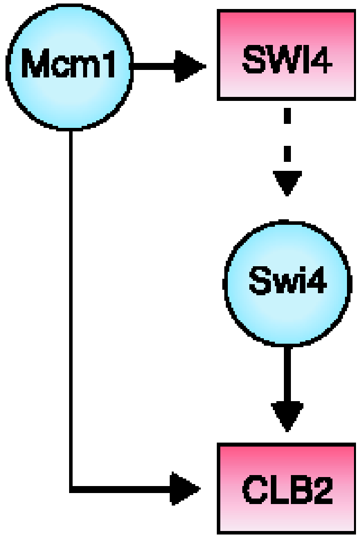
Autoregulation



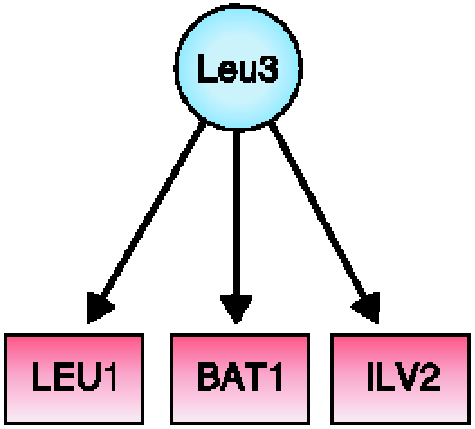
Multi-Component Loop



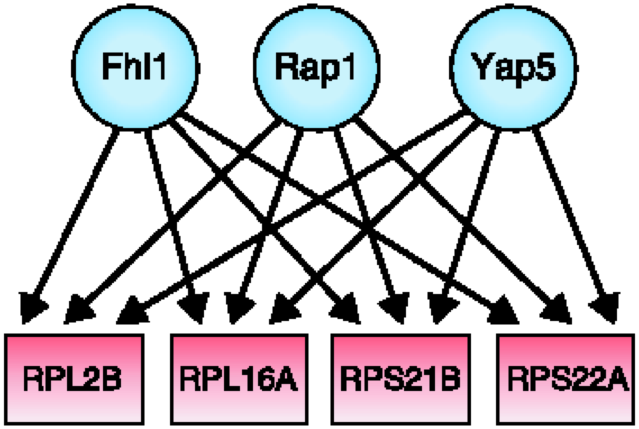
Feedforward Loop



Single Input Motif

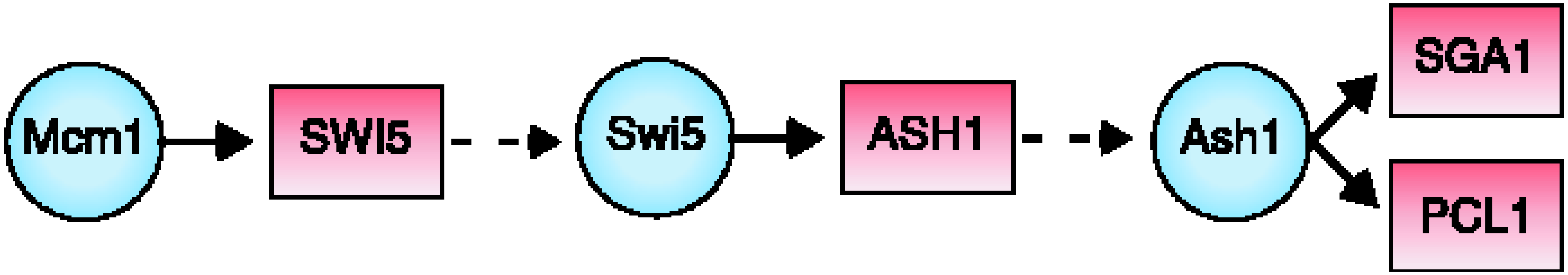


Multi-Input Motif



Network motifs

Regulator Chain



Literature

Exploring complex networks, Steven H Strogatz, Nature 410, p. 268, 8 Mar 2001

Statistical Mechanics of Complex Networks, Reka Albert and Albert-László Barabási, arXiv:cond-mat/0106096, 6 Jun 2001

Network Motifs: Simple Building Blocks of Complex Networks, R. Milo et al., Science 298, 824-827 (2002)

Two-hybrid arrays, Peter Uetz, Current Opinion in Chemical Biology 6, 57-62 (2001)

Transcriptional Regulatory Networks in Saccharomyces Cerevisiae, TI Lee et al., Science 298, 799-804 (2002)

Functional organization of the yeast proteome by systematic analysis of protein complexes, AC Gavin et al., Nature 415, 141 (2002)

Functional discovery via a compendium of expression profiles, TR Hughes et al., Cell 102, 109-126 (2000)