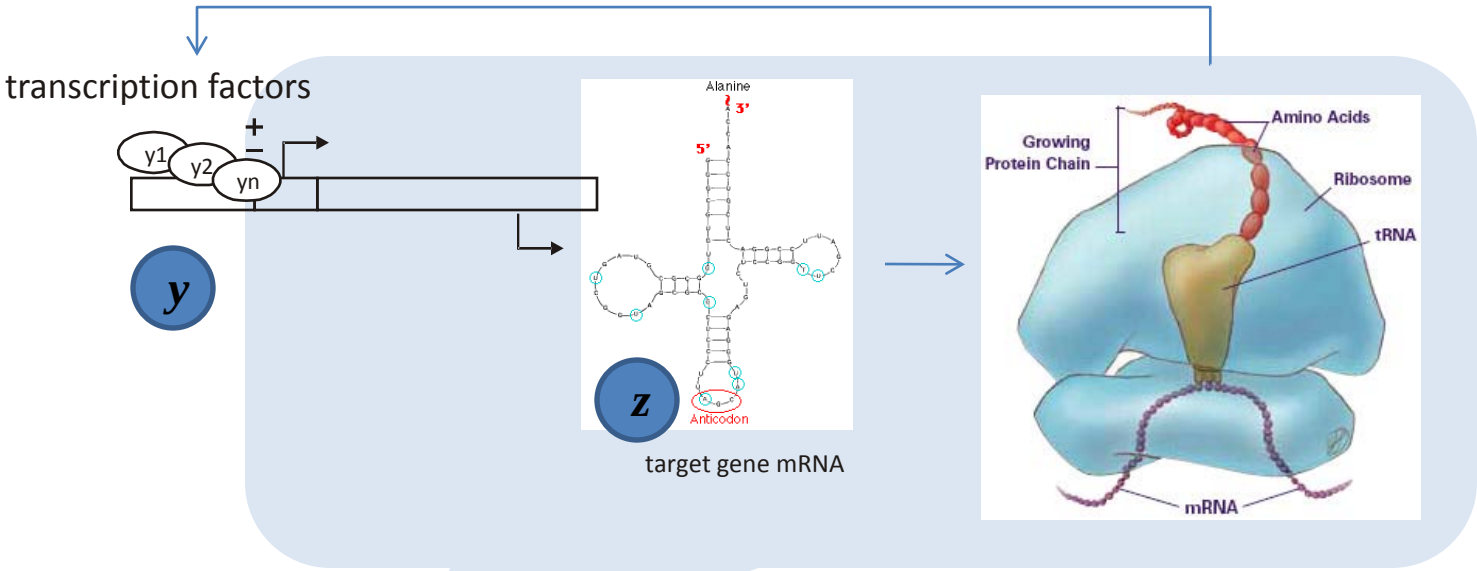
Inference of active genetic networks from microarray and ChIP-on-chip experiments by evolutionary modeling.

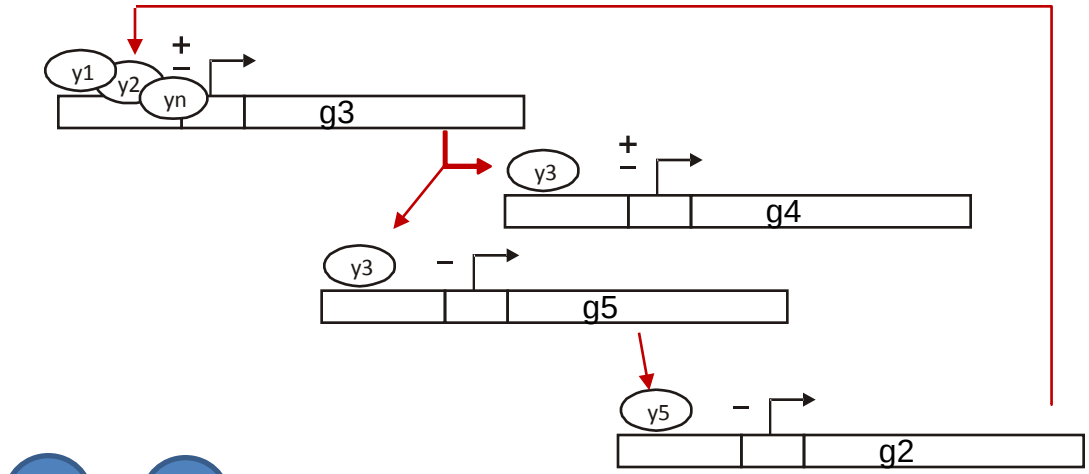
Cuong C. To, J. Vohradský
Laboratory of bioinformatics
Institute of Microbiology, ASCR, Prague.

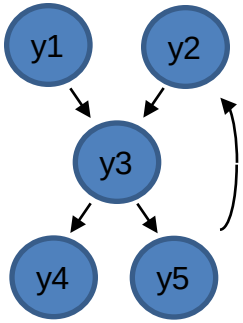
symbolic representation of gene expression



$$\frac{dz(t)}{dt} = f(t, y_{i=1..n}, w) - k_d z(t)$$

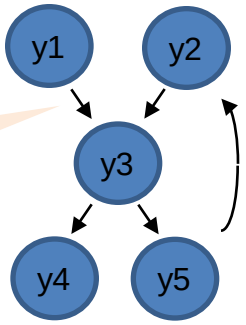
genetic network

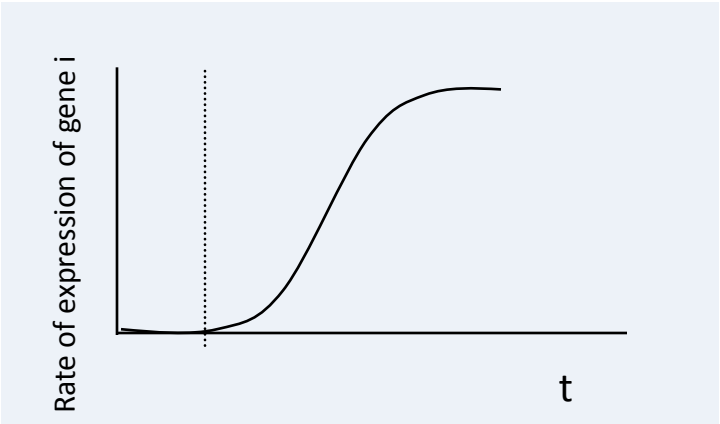




numerical model of gene expression

$$\frac{dz_i(t)}{dt} = \frac{k_1}{1+\exp(-\sum w_{i,j} y_j + b_i)} - k_d z_i(t)$$

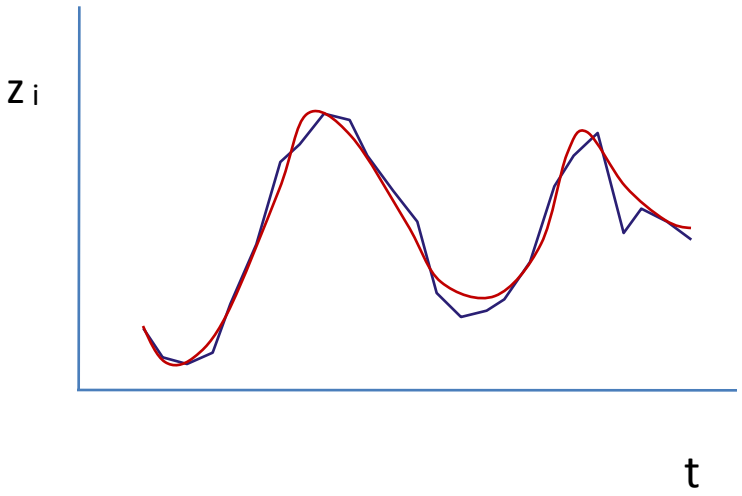




first order degradation

reverse engineering of genetic network

gene expression event is characterized by temporal profile of transcribed gene



reconstruct individual interactions

find k1,k2,w,b which minimize

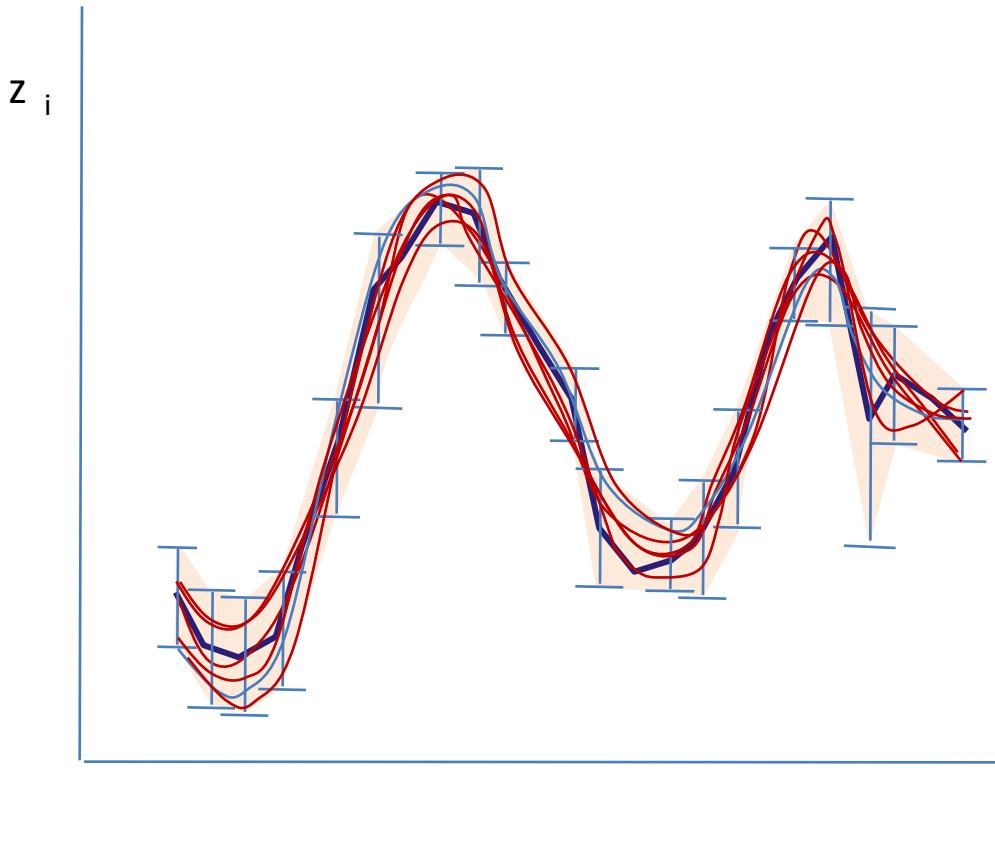
$$F = \min\left(\sum \sqrt{\left[\hat{\mathbf{z}}(t) - \mathbf{z}(t)\right]^2} \right)$$

computed

measured

$$\frac{dz_i(t)}{dt} = \frac{k_1}{1 + \exp(-\sum w_{i,j} y_j + b_i)} - k_d z_i(t)$$

confidence interval effect



reconstruct individual interactions

find k1,k2,w,b which minimize

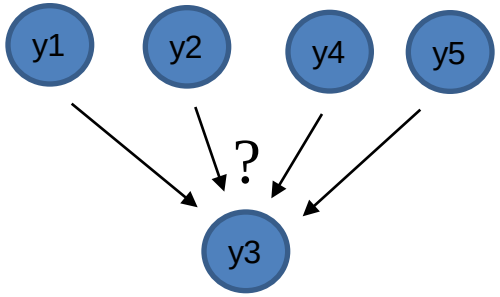
$$F = \min\left(\sum \sqrt{\left[\hat{\mathbf{z}}(t) - \mathbf{z}(t)\right]^2} \right)$$

computed

measured

and $\hat{z}$ is within confidence interval

such condition is satisfied for more than one regulator or combination of regulators



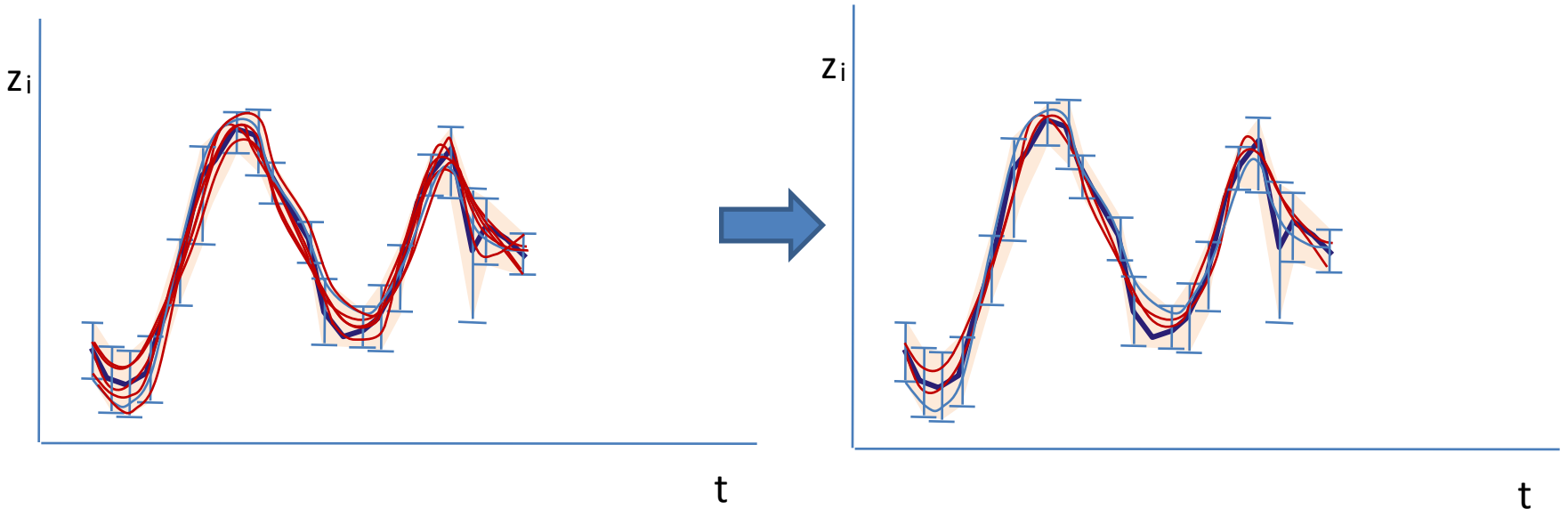
need for constraints

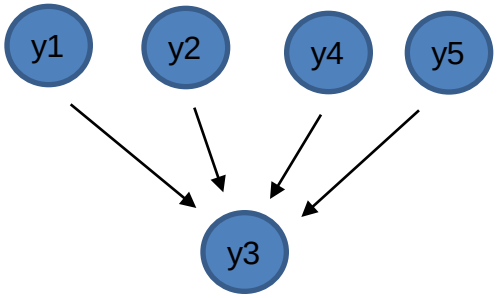
ChIP-on chip

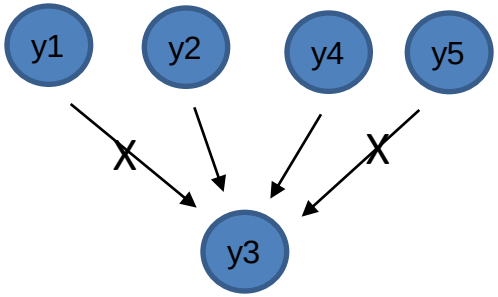
identify potential regulators binding to the target gene promoter

ChIP-on-chip interaction matrix

	r1	r2	r3	…	rm
g1					1
g2	1		1		
g3		1			
.					
.					
gn	1				







Alternative indistinguishable connections create set of alternative equivalent networks !

reverse engineering of genetic network

input – time series of gene expression (chips, qPCR)
constraints – ChIP-on-chip

n x n network

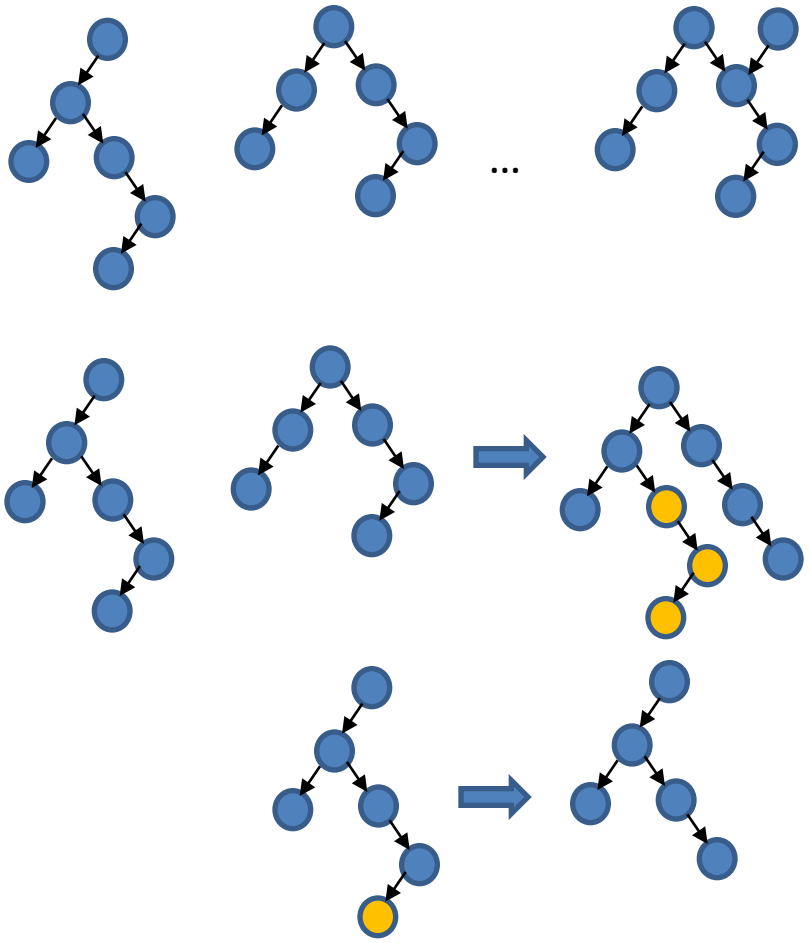
- all parameter optimization
  – possible only for small networks (5x5)
- optimization of parameters for individual interactions
  – computationally intensive (necessary to compute all regulator/target , multiple regulator/target combinations)
- evolutionary programming network reconstruction
  – computationally intensive, but fully unbiased

evolutionary programming network reconstruction

1. create a set of random networks
2. solve dif. equations and
3. compute fitness of each net

$$G = \sum_{i=1..k} \sqrt{\left[\hat{\mathbf{z}}_i(t) - \mathbf{z}_i(t)\right]^2}$$

4. create next generation by



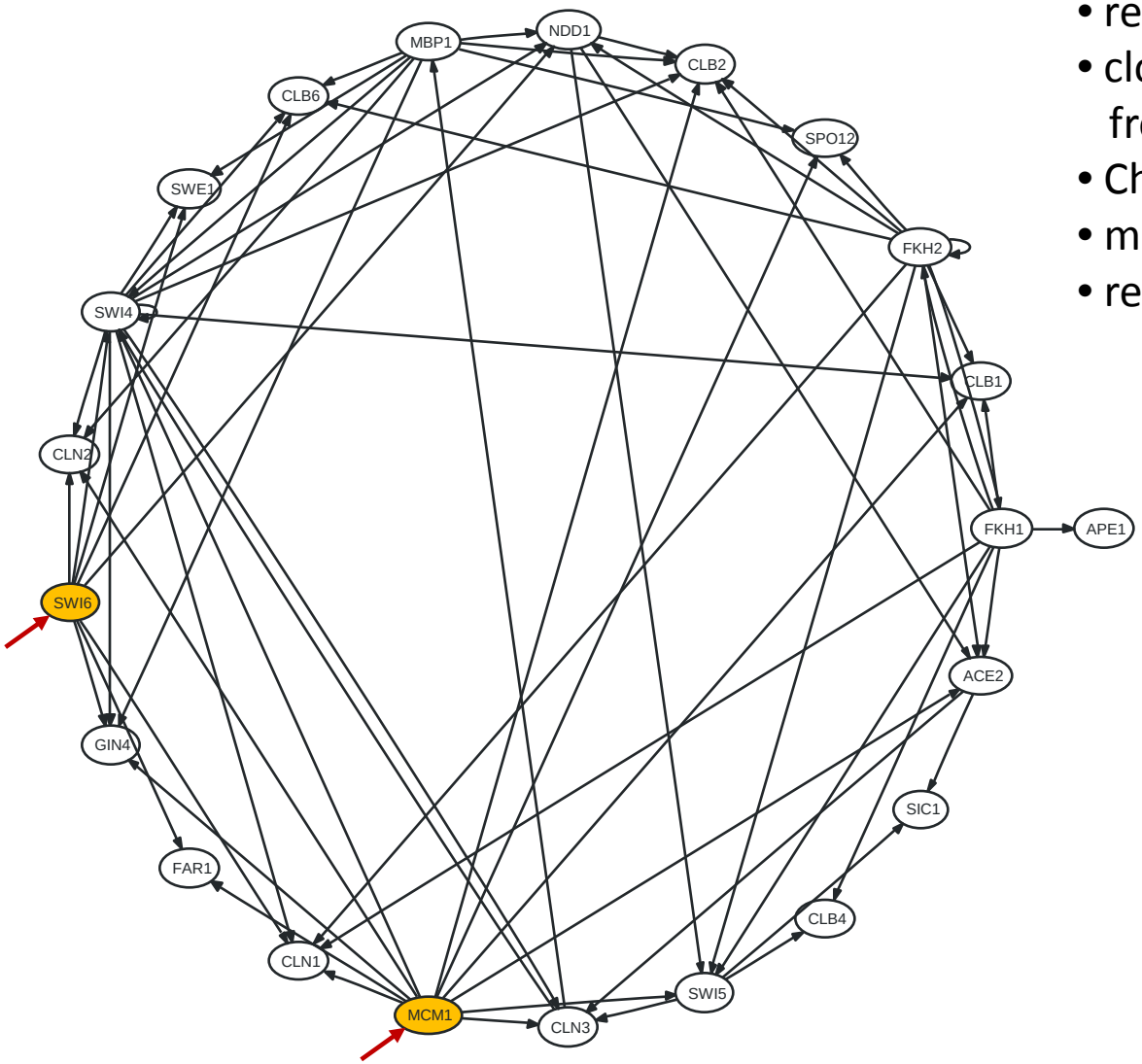
crosover

mutation

reproduction – simple copy

5. Go to 2. and repeat until convergence

Yeast cyclins network



- realtively small (22 genes)
- closed (only two genes controlled from outside)
- Chip-on-chip measurements exist
- microarray time series exist
- repeated measurements

cyclins network reconstruction

Step 1: Generate a population of networks with random connections among genes, Net(i) with $1 \le i \le 100$;
constraint – only connections given by ChIP-on-chip measurements were allowed.

Step 2: For each network compute parameters P={$\boldsymbol{W, b, k_1, k_2}$} using genetic algorithm.

Step 2-1. Generate population of random parameters $P_i(k)$, $1 \le k \le 500$ with max number of generations=500;
probability of crossover, mutation, and reproduction = 0.6, 0.3, and 0.1, respectively.

Step 2–2: solve Eq.2 to calculate the value of fitness function G for each P(k) of network Net(i).

Step 2–3: Update parameters by reproduction and crossover to create new generation of parameters $P_{i+1}(k)$.

Step 2–4: Loop Step 2–2 and 2–3 until convergence or predefined number of generation is reached.

Step 2–5: Parameters P(m) giving a minimal value of G are selected as the best approximation of Net(i).

Step 3: Using reproduction, crossover and mutation operations create new generation of Net(i).

Step 4: Loop Step 2 and 3 until no improvement in fitness G or preset number of generations is reached.

Step. 5: Sort all networks according to increasing values of fitness G.

maximum number of generations was set to 500; probability of crossover, mutation, and reproduction were set to 0.6, 0.3, and 0.1, respectively.

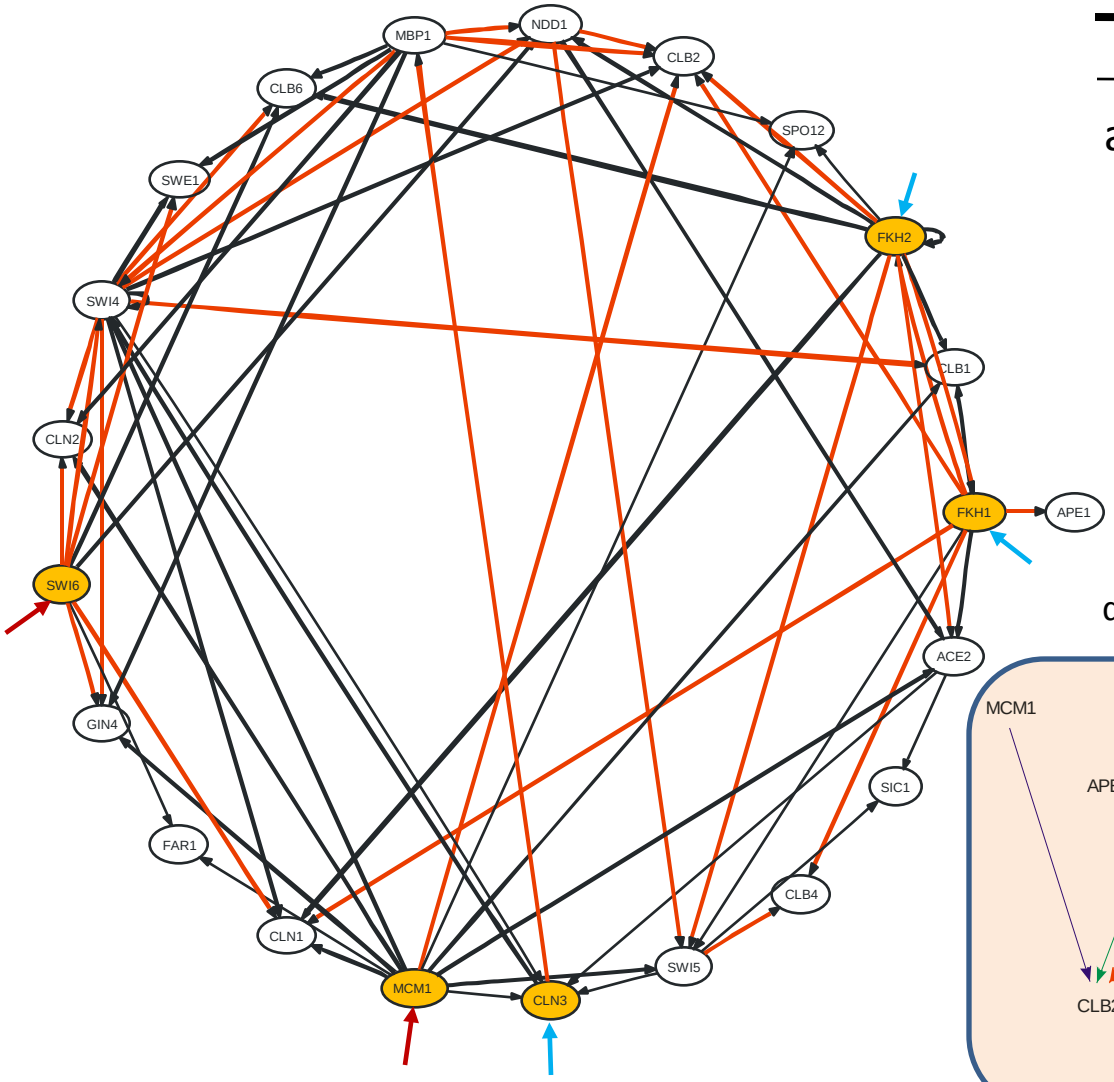
cyclins network reconstruction -results

7 networks satisfying goodness of fit criteria – reconstructed expression profile within confidence interval

4 genes (CLN3, SPO12, SIC1, FAR1) – reconstruction not possible

set of minimal number of vertices for each connection occurring in any of the 7 networks form **minimal network**.

cyclins minimal network



minimal network

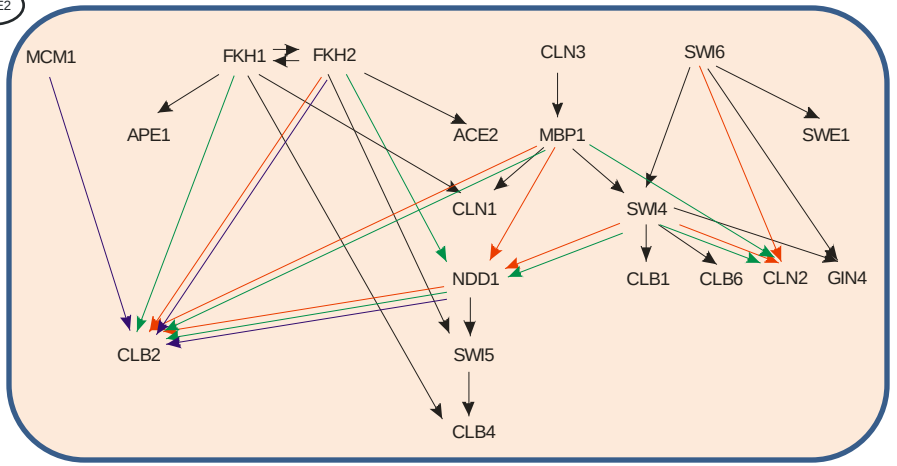
alternative connections

ChIP-on-chip only

all ChIP-on-chip network

causal relations

different colors – alternative connections



common principles

- networks derived from kinetic measurements is smaller than predicted from ChIP-on-chip experiments => ChIP-on-chip network is only a potential network.

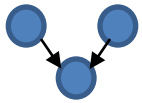
- It is always possible to identify a set of minimal networks, i.e. equivalent networks with minimal number of vertices which still fit experimental data.

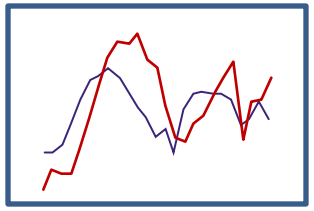
- One or two regulators are sufficient to correctly interpret experimental data.

- Although a single regulator can be found which interprets experimental data, multiple regulators will form more robust control, especially for the case of activator and repressor pair.

- If more regulators of one target gene satisfy the data confidence interval criterion, such case cannot be neglected even if a simpler mode of control can be found.

- Regulators controlling target gene expression in pairs usually act so that one is activator and second repressor, their gene expression profiles have similar shape but are mutually shifted in time.





future perspectives

- compute all combinations of interactions within the cyclins network given by ChIP-on-chip matrix for one, two and three regulators per gene and compare with the GP results.

- compute cell cycle networks for all genes given by ChIP-on-chip matrix.

acknowledgments

This work was supported by grants from the Czech Science Foundation, No. 310/07/1009, EC Integrated Project ActinoGEN, LSHM-CT-2004-005224.

Cuong C. To