

Emergence of gene regulatory networks under functional constraints

Marcin Zagórski

Institute of Physics, Jagiellonian University, Kraków, Poland

in collaboration with

Zdzisław Burda, Andre Krzywicki and Olivier C. Martin

30 November 2012, Leipzig, CompPhys12



**INNOVATIVE
ECONOMY**
NATIONAL COHESION STRATEGY



Foundation for Polish Science

EUROPEAN UNION
EUROPEAN REGIONAL
DEVELOPMENT FUND



Genome size

Number of protein-coding genes

E. coli ~ 4000



Yeast ~ 6000

Fruit fly ~ 14000

Human ~ 23000

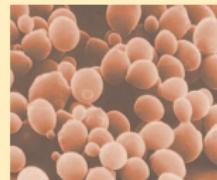
Genome size

Number of protein-coding genes

E. coli ~ 4000



Yeast ~ 6000



Fruit fly ~ 14000

Human ~ 23000

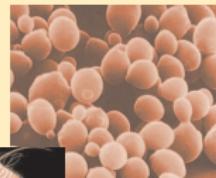
Genome size

Number of protein-coding genes

E. coli ~ 4000



Yeast ~ 6000



Fruit fly ~ 14000



Human ~ 23000

Genome size

Number of protein-coding genes

E. coli ~ 4000



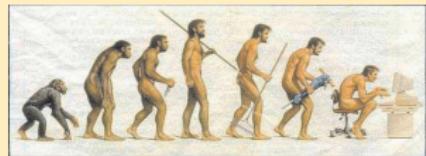
Yeast ~ 6000



Fruit fly ~ 14000

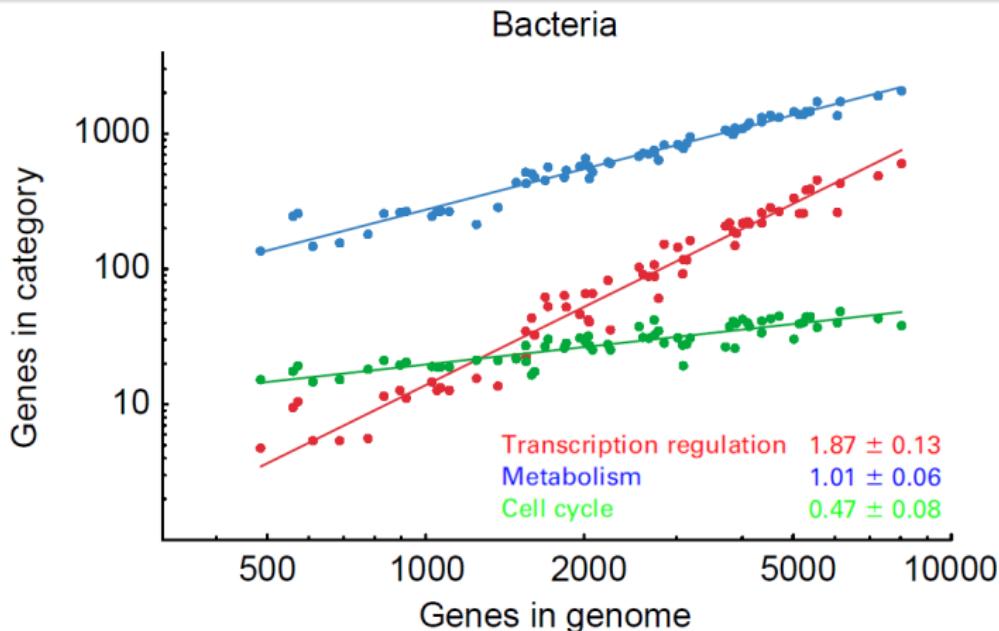


Human ~ 23000



Interactions between genes

Number of regulatory genes grows faster than linearly with the total number of genes.

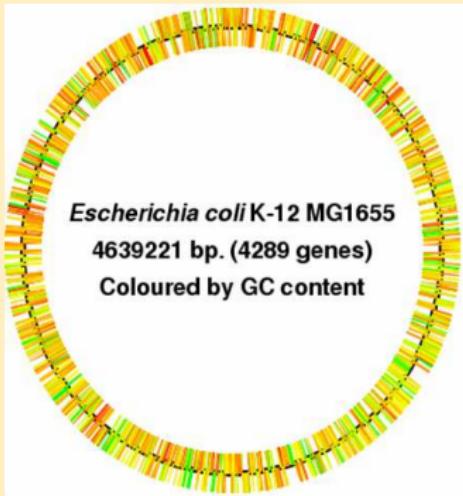


Source: E. van Nimwegen, Trends Genet. 19, 479 (2003).

Genetic information

Genome sequencing does not bring all essential information.

Change of paradigm

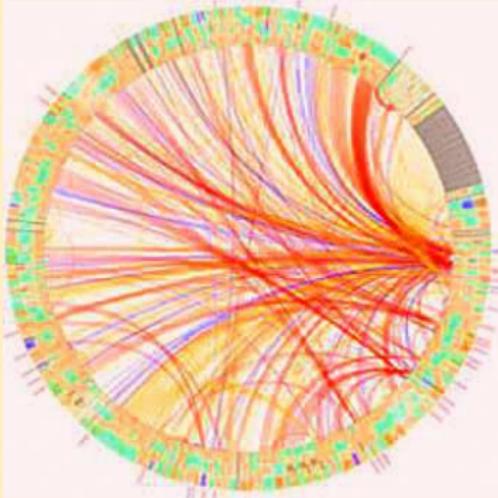
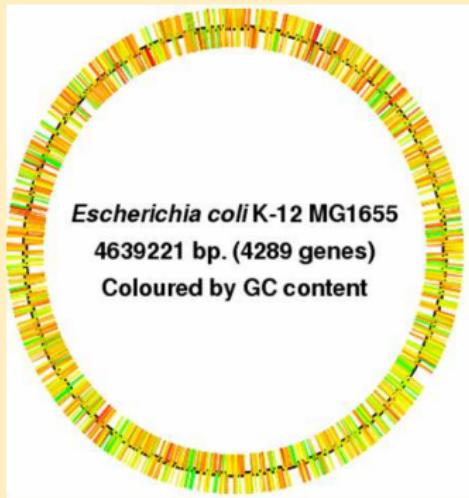


Source: <http://xbase.ac.uk>

Genetic information

Gene regulation plays an important role.

Change of paradigm

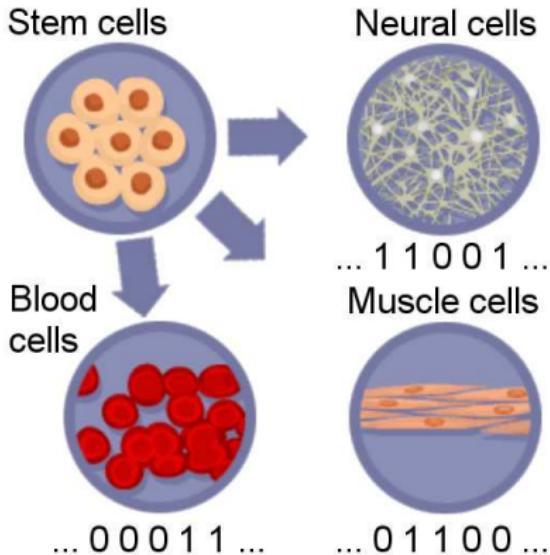


Source: <http://xbase.ac.uk>

Gene regulation

Cell differentiation

- different gene expression patterns correspond to different tissues
- many regulatory mechanisms
- transcription factors (TFs) - promote/block other genes



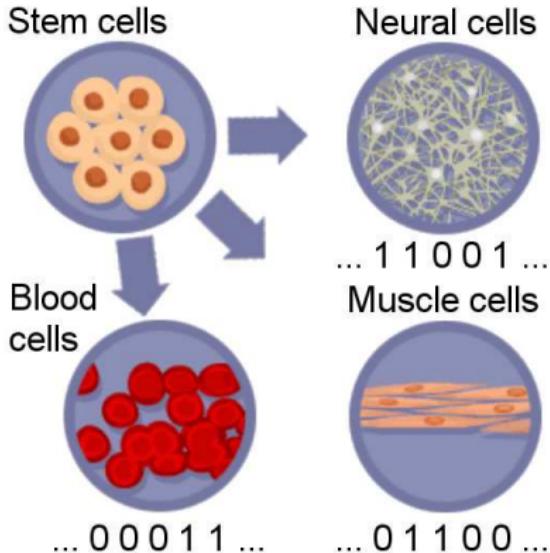
Cell-division cycle

- gene expression patterns correspond to different phases

Gene regulation

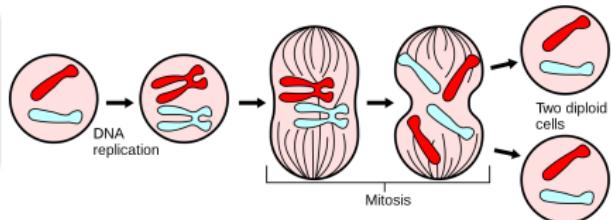
Cell differentiation

- different gene expression patterns correspond to different tissues
- many regulatory mechanisms
- transcription factors (TFs) - promote/block other genes



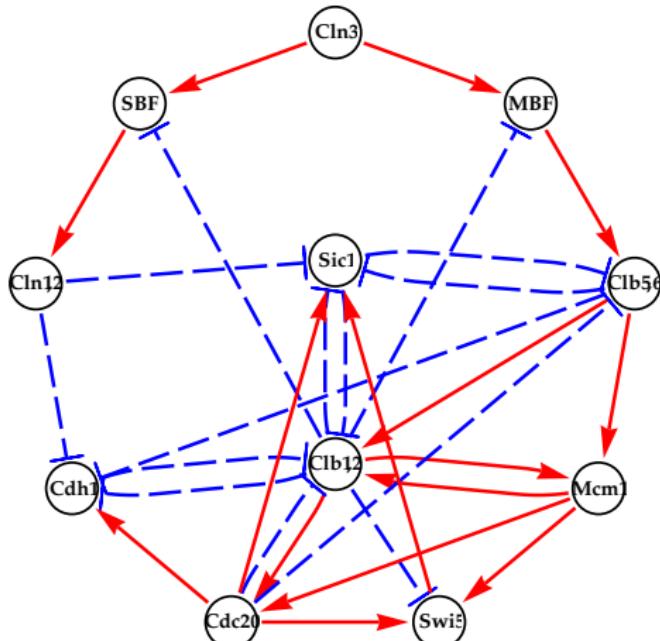
Cell-division cycle

- gene expression patterns correspond to different phases



Baker's yeast cell-cycle

Regulatory network and gene expression pattern



← Time

1	0	0	0	0	0	0	0	0	1	1
0	1	1	0	0	0	0	0	0	1	1
0	1	1	1	0	0	0	0	0	1	1
0	1	1	1	0	0	0	0	0	0	0
0	1	1	1	1	0	0	0	0	0	0
0	1	1	1	1	1	1	0	0	0	0
0	0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	1	1	1	1	0	0
0	0	0	0	0	0	1	1	1	1	0
0	0	0	0	0	0	0	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1

Source: F. Li, T. Long, Y. Lu, Q. Ouyang and C. Tang, PNAS 101, 4781 (2004).

Baker's yeast cell-cycle

Regulatory network and gene expression pattern



←	Time	1	0	0	0	0	0	0	0	0	1	1
		0	1	1	0	0	0	0	0	0	1	1
		0	1	1	1	0	0	0	0	0	1	1
		0	1	1	1	0	0	0	0	0	0	0
		0	1	1	1	1	0	0	0	0	0	0
		0	1	1	1	1	1	1	0	0	0	0
		0	0	0	1	1	1	1	1	0	0	0
		0	0	0	0	0	1	1	1	1	1	0
		0	0	0	0	0	0	1	1	1	1	0
		0	0	0	0	0	0	0	1	1	1	1
		0	0	0	0	0	0	0	0	1	1	1
		0	0	0	0	0	0	0	0	0	1	1

Different frameworks

Dynamical models of gene networks

Focusing on transcriptional dynamics

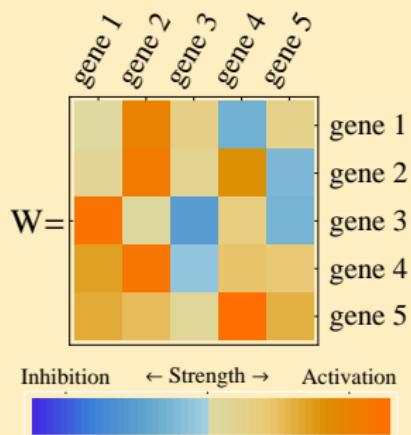
- Boolean networks (genes on-off) (e.g. Stuart Kauffman)
- Threshold networks (e.g. inspired by neural dynamics, Andreas Wagner)
- Differential equations with rates (e.g. Albert Goldbeter, John Tyson)
- Piece-wise linear input-output relations (e.g. Hidde de Jong)
- ...

Gene regulatory network model

Phenotype

$$\mathbf{S}(t) = (S_1(t), S_2(t), \dots, S_N(t))$$

Genotype



Transcriptional dynamics

$$\mathbf{S}(t+1) = G(\mathbf{S}(t), \mathbf{W})$$

$$\mathbf{S}(0) \xrightarrow{\mathbf{W}} \mathbf{S}(1) \xrightarrow{\mathbf{W}} \dots \xrightarrow{\mathbf{W}} \mathbf{S}(t)$$

Target pattern

$$\mathbf{S}^{target}(0) = 1\ 1\ 0\ 0\ 1$$

$$\mathbf{S}^{target}(1) = 0\ 1\ 1\ 0\ 0$$

$$\vdots$$
$$\mathbf{S}^{target}(T) = 0\ 0\ 0\ 1\ 1$$

Fitness

$$F(\mathbf{S}) = \exp(-f \sum_{t=0}^T |\mathbf{S}(t) - \mathbf{S}^{target}(t)|)$$

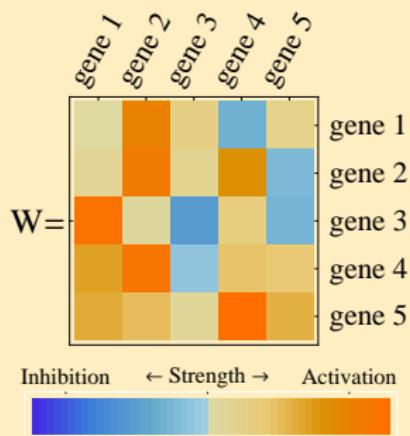
Source: Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, PNAS **108**, 17263 (2011).

Gene regulatory network model

Phenotype

$$\mathbf{S}(t) = (S_1(t), S_2(t), \dots, S_N(t))$$

Genotype



Transcriptional dynamics

$$\mathbf{S}(t+1) = G(\mathbf{S}(t), \mathbf{W})$$

$$\mathbf{S}(0) \xrightarrow{\mathbf{w}} \mathbf{S}(1) \xrightarrow{\mathbf{w}} \dots \xrightarrow{\mathbf{w}} \mathbf{S}(t)$$

Target pattern

$$\mathbf{S}^{target}(0) = 1\ 1\ 0\ 0\ 1$$

$$\mathbf{S}^{target}(1) = 0\ 1\ 1\ 0\ 0$$

⋮

$$\mathbf{S}^{target}(T) = 0\ 0\ 0\ 1\ 1$$

Fitness

$$F(\mathbf{S}) = \exp(-f \sum_{t=0}^T |\mathbf{S}(t) - \mathbf{S}^{target}(t)|)$$

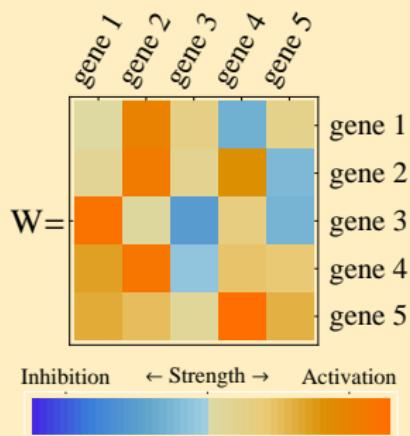
Source: Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, PNAS **108**, 17263 (2011).

Gene regulatory network model

Phenotype

$$\mathbf{S}(t) = (S_1(t), S_2(t), \dots, S_N(t))$$

Genotype



Transcriptional dynamics

$$\mathbf{S}(t+1) = G(\mathbf{S}(t), \mathbf{W})$$

$$\mathbf{S}(0) \xrightarrow{\mathbf{w}} \mathbf{S}(1) \xrightarrow{\mathbf{w}} \dots \xrightarrow{\mathbf{w}} \mathbf{S}(t)$$

Target pattern

$$\mathbf{S}^{target}(0) = 1\ 1\ \textcolor{blue}{0}\ 0\ 1$$

$$\mathbf{S}^{target}(1) = \textcolor{blue}{0}\ 1\ 1\ 0\ 0$$

⋮

$$\mathbf{S}^{target}(T) = 0\ 0\ 0\ \textcolor{blue}{1}\ 1$$

Fitness

$$F(\mathbf{S}) = \exp(-f \sum_{t=0}^T |\mathbf{S}(t) - \mathbf{S}^{target}(t)|)$$

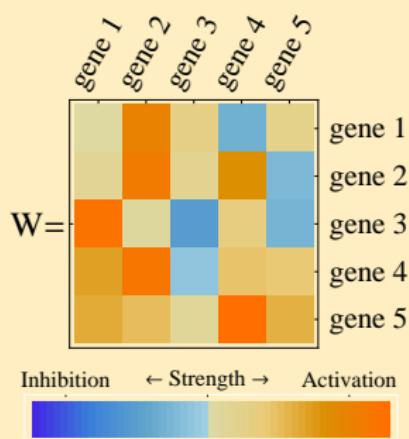
Source: Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, PNAS **108**, 17263 (2011).

Gene regulatory network model

Phenotype

$$\mathbf{S}(t) = (S_1(t), S_2(t), \dots, S_N(t))$$

Genotype



Transcriptional dynamics

$$\mathbf{S}(t+1) = G(\mathbf{S}(t), \mathbf{W})$$

$$\mathbf{S}(0) \xrightarrow{\mathbf{w}} \mathbf{S}(1) \xrightarrow{\mathbf{w}} \dots \xrightarrow{\mathbf{w}} \mathbf{S}(t)$$

Target pattern

$$\mathbf{S}^{target}(0) = 1 \ 1 \ 0 \ 0 \ 1$$

$$\mathbf{S}^{target}(1) = 0 \ 1 \ 1 \ 0 \ 0$$

⋮

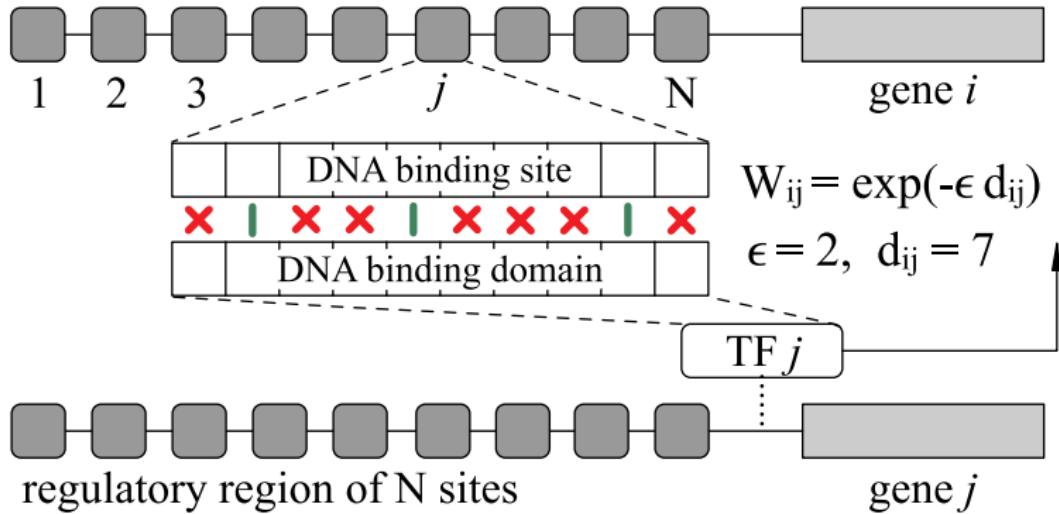
$$\mathbf{S}^{target}(T) = 0 \ 0 \ 0 \ 1 \ 1$$

Fitness

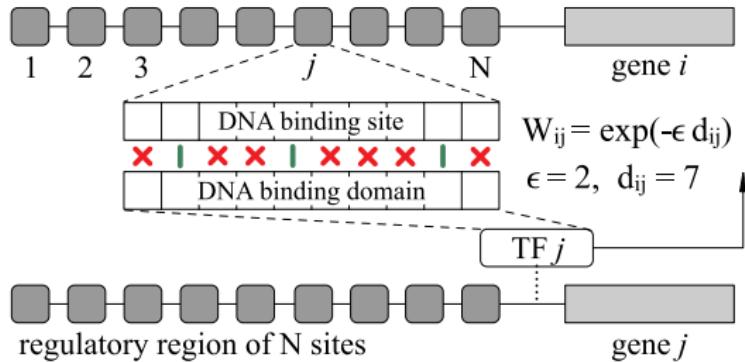
$$F(\mathbf{S}) = \exp(-f \sum_{t=0}^T |\mathbf{S}(t) - \mathbf{S}^{target}(t)|)$$

Source: Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, PNAS **108**, 17263 (2011).

Biophysical modelling of interactions



Biophysical modelling of interactions



The binding energy between two molecules is additive between facing elements with each mismatch contributing a penalty ϵ .

$$P_{ij} = \frac{n_j W_{ij}}{1 + n_j W_{ij}} \text{ where } n_j = n S_j(t)$$

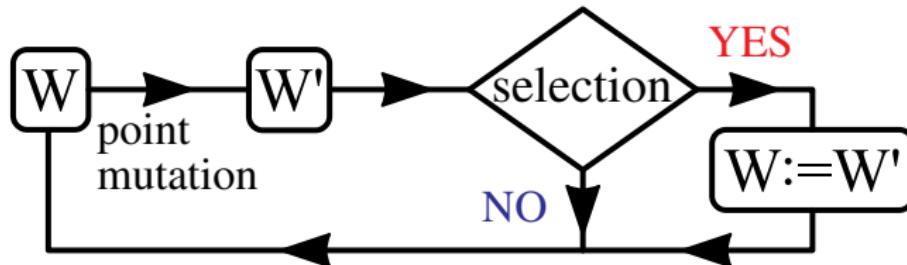
Source: U. Gerland, J. Moroz and T. Hwa, PNAS **99**, 12015 (2002).

Expression dynamics and MCMC sampling

$$S_i(t+1) = \underbrace{\left[1 - \prod_j (1 - P_{ij}(t)) \right]}_{\text{Activators}} \prod_{j'} (1 - P_{ij'}(t)) \underbrace{\quad}_{\text{Repressors}}$$

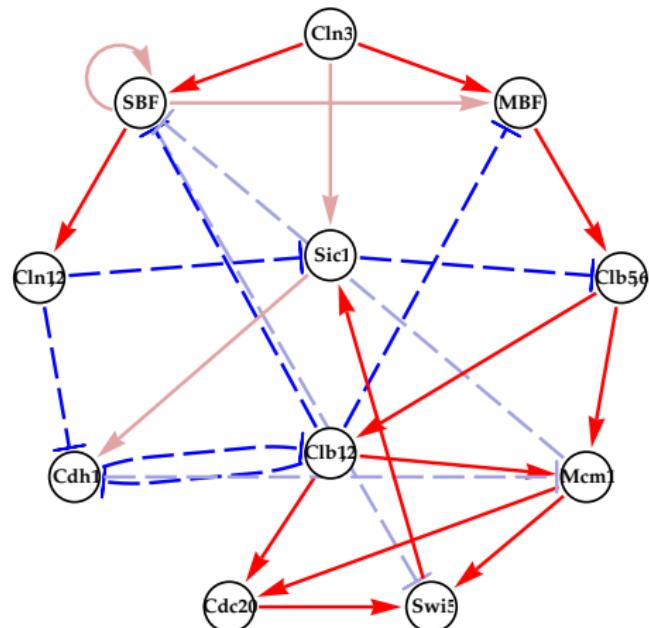
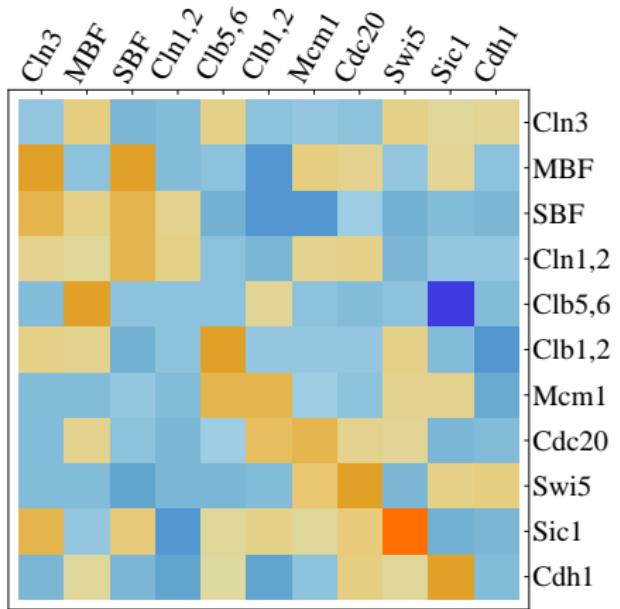
$$P_{ij}(t) = \frac{1}{1 + \exp(\varepsilon d_{ij} - \log n S_j(t))}$$

Metropolis sampling of space of viable genotypes.

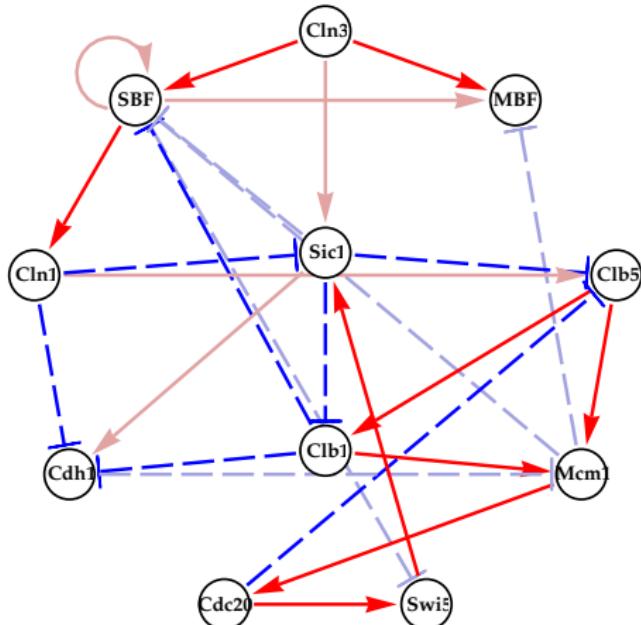
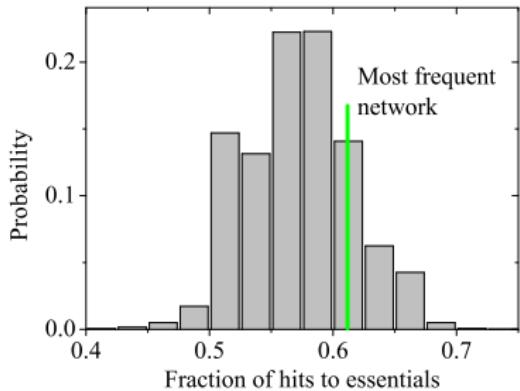
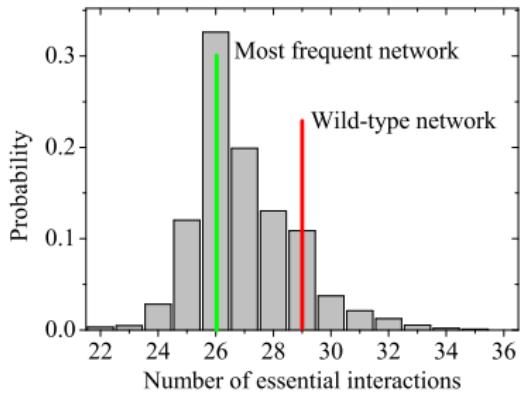


Network of essential interactions

Example of matrix \mathbf{W} with corresponding regulatory network for yeast cell-cycle expression pattern.



Most frequent network in the ensemble



Network made of most frequent interactions

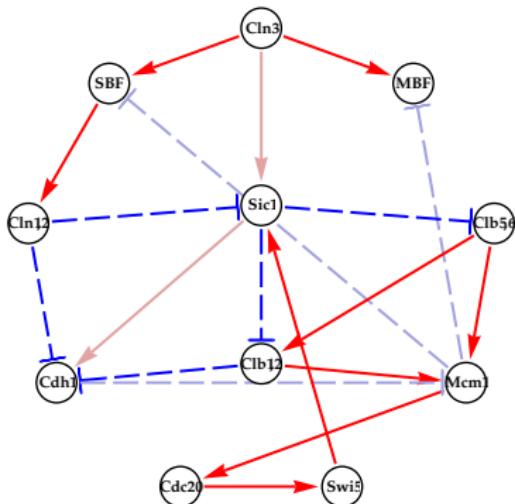
Edge usage

Activatory interactions tend to form a long feed forward cascade

	<i>Cln3</i>	<i>MBF</i>	<i>SBF</i>	<i>Cln1,2</i>	<i>Clb5,6</i>	<i>Clb1,2</i>	<i>Mcm1</i>	<i>Cdc20</i>	<i>Swi5</i>	<i>Sic1</i>	<i>Cdh1</i>
<i>Cln3</i>	0.04	0	0	0	0	0	0	0	0	0	0
<i>MBF</i>	1.	0.5	0.54	0.01	0	0	0	0	0	0	0
<i>SBF</i>	1.	0.51	0.52	0.01	0	0	0	0	0	0	0
<i>Cln1,2</i>	0	0.59	0.62	0	0	0	0	0	0	0	0
<i>Clb5,6</i>	0	0.39	0.4	0.38	0.06	0	0	0	0	0	0
<i>Clb1,2</i>	0	0	0	0	0.99	0.01	0	0	0	0	0
<i>Mcm1</i>	0	0	0	0	0.87	0.98	0	0	0	0	0
<i>Cdc20</i>	0	0	0	0	0.02	0.1	1.	0	0	0	0
<i>Swi5</i>	0.01	0	0	0.01	0.01	0.05	0.09	1.	0	0.01	0.01
<i>Sic1</i>	1.	0	0	0	0	0.01	0.01	0	1.	0	0
<i>Cdh1</i>	0	0	0	0	0	0	0	0.01	1.	0	0

Regulatory network

Interactions present in >60% of networks



Summary

Results

- much of the cell-cycle network is reproduced under functional constraints
- motifs found are compatible with the imposed functional capability
- GRNs are as *sparse* as possible to maintain the imposed expression pattern

Similar outcome for two other cell-cycles:

- fission yeast, Davidich and Borhholdt (PloS One, 2008)
- mammals, Faure et al. (Bioinformatics, 2006)

Acknowledgments

Project carried within the MPD programme “Physics of Complex Systems” of the Foundation for Polish Science and cofinanced by the European Development Fund in the framework of the Innovative Economy Programme.



**INNOVATIVE
ECONOMY**
NATIONAL COHESION STRATEGY



EUROPEAN UNION
EUROPEAN REGIONAL
DEVELOPMENT FUND



References

-  M. Zagorski, A. Krzywicki, O.C. Martin, *Edge usage, motifs and regulatory logic for cell cycling genetic networks*, to appear.
-  Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, *Motifs emerge from function in model gene regulatory networks*, PNAS **108**, 17263-17268 (2011).
-  Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, *Distribution of essential interactions in model gene regulatory networks under mutation-selection balance*, Phys. Rev. E **82**, 011908 (2010).

Thank you for your attention