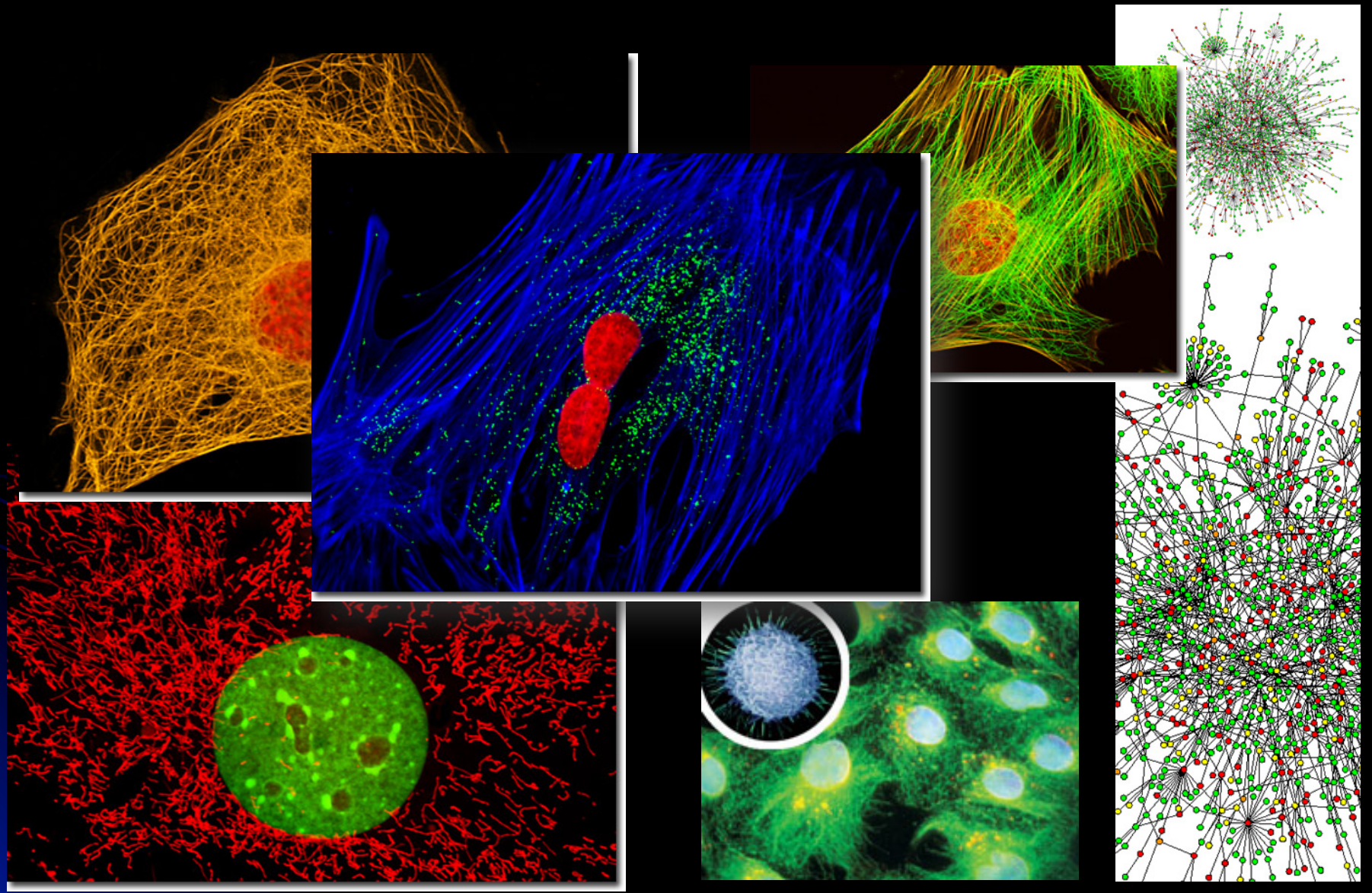
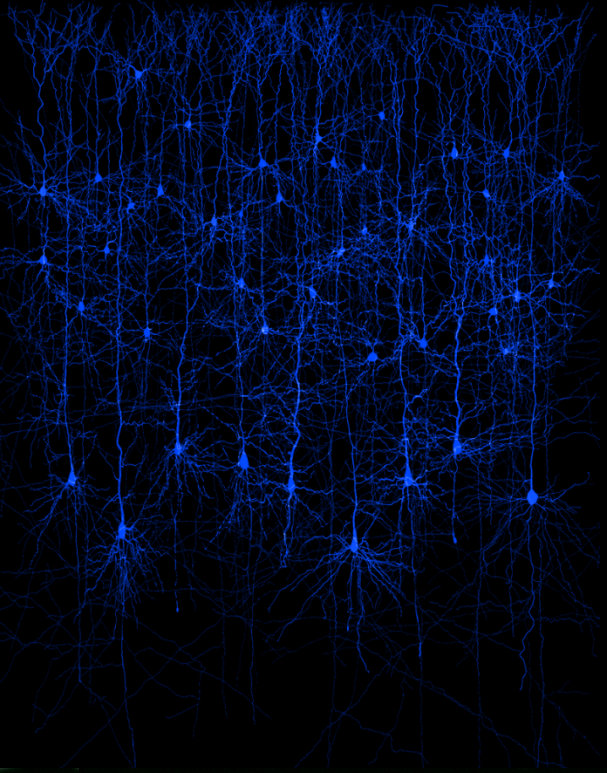
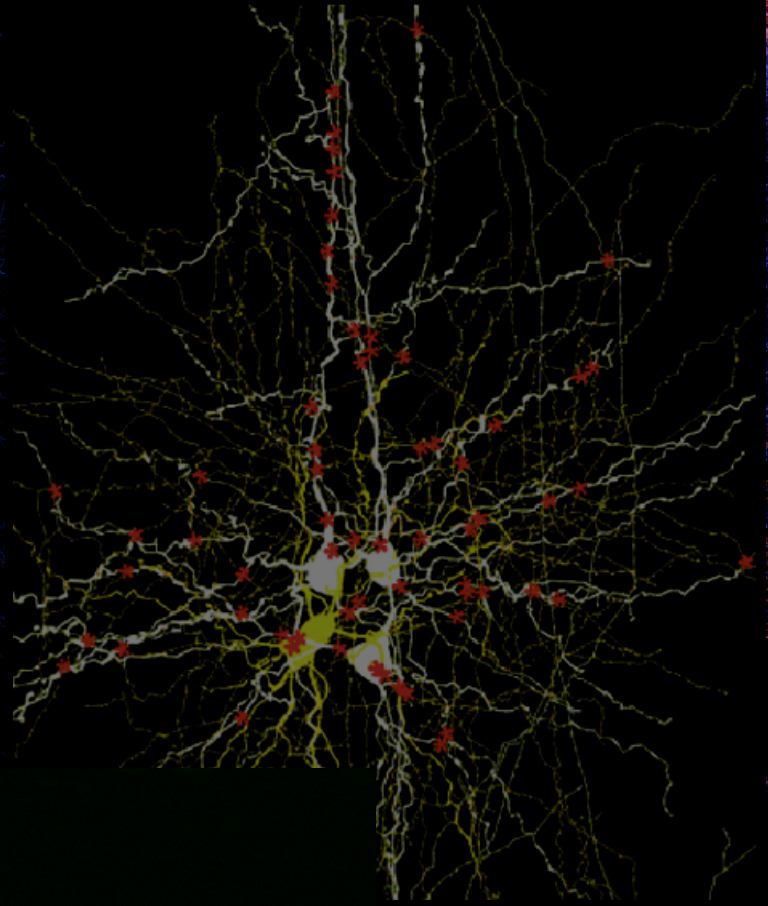
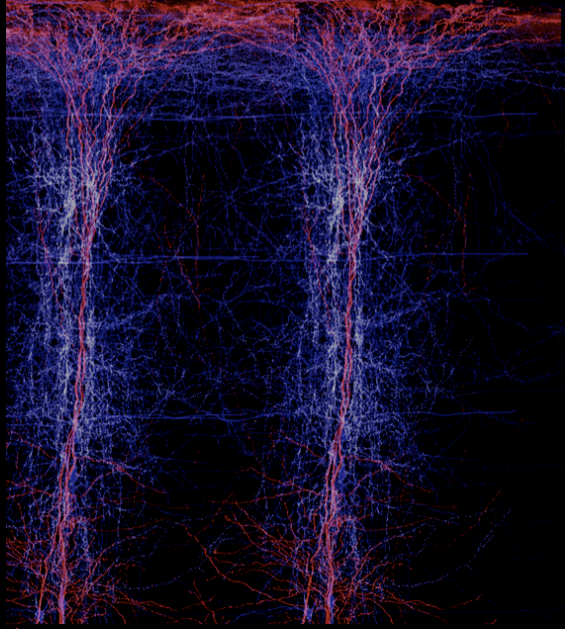
Probabilistic Boolean Networks

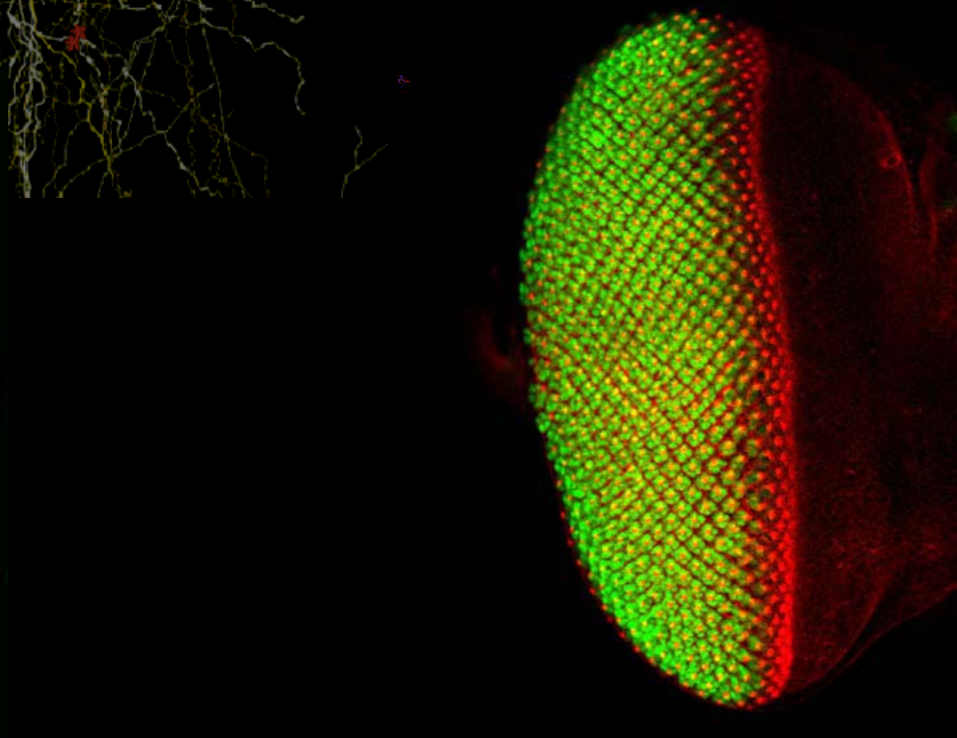
Ilya Shmulevich

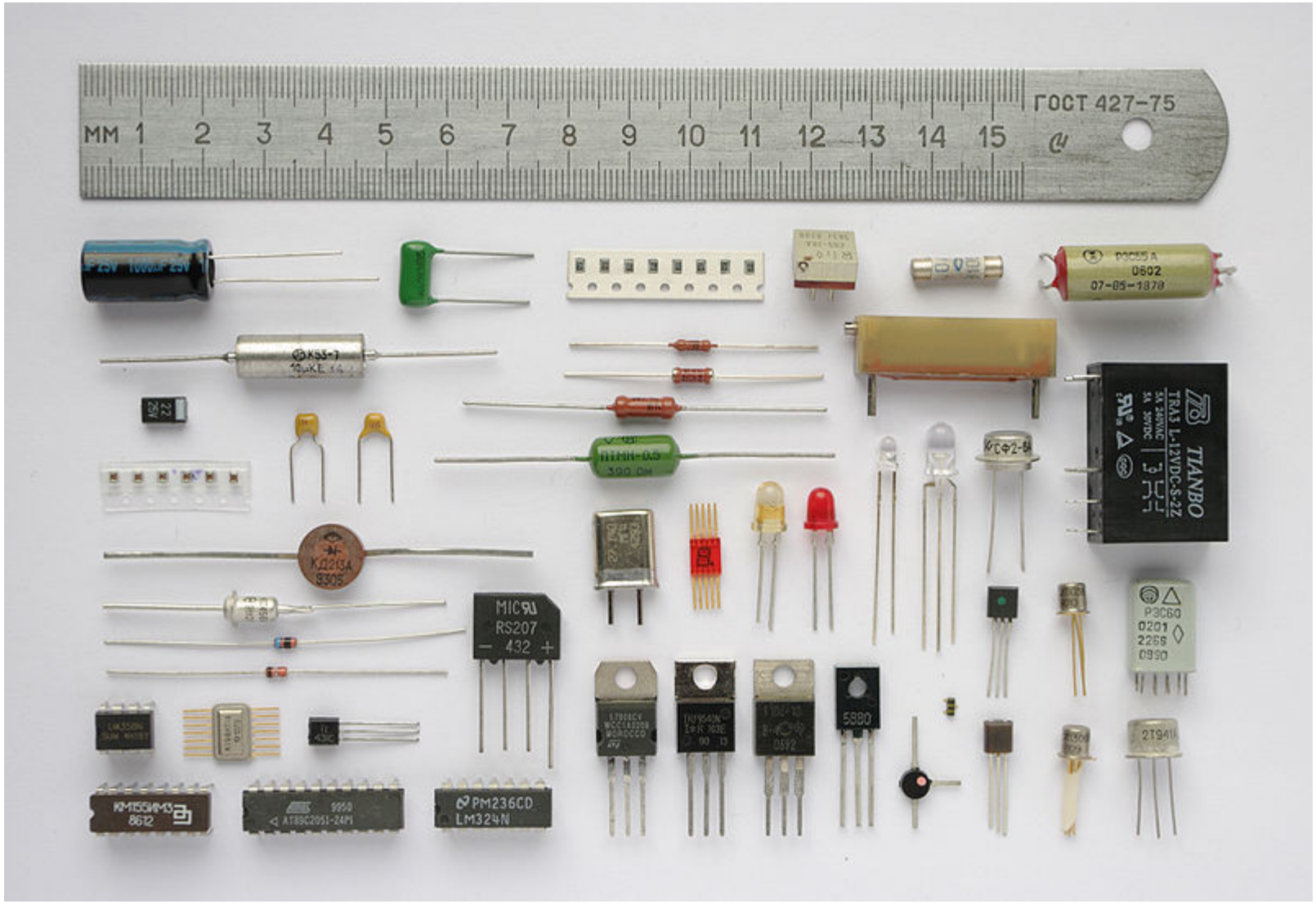


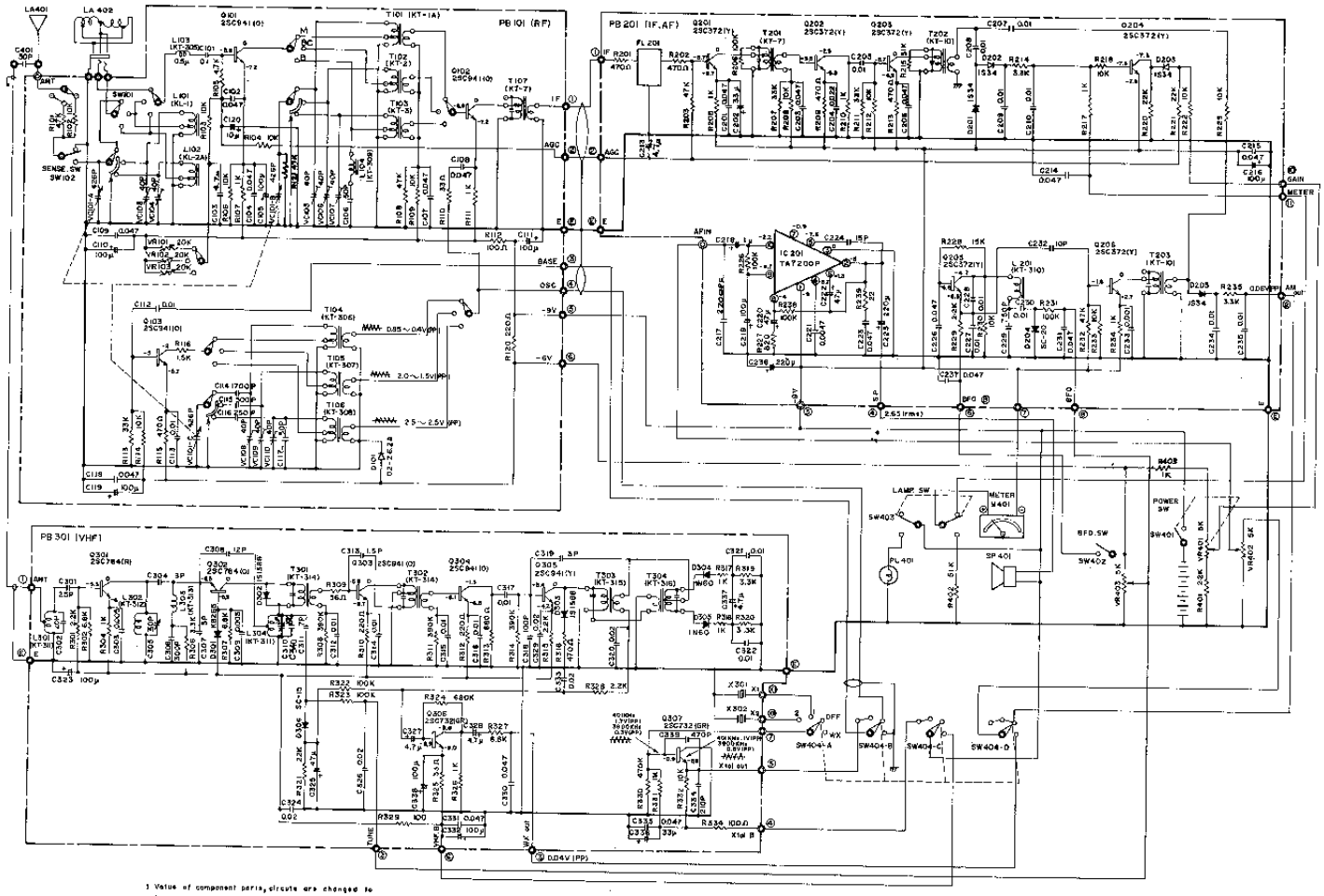




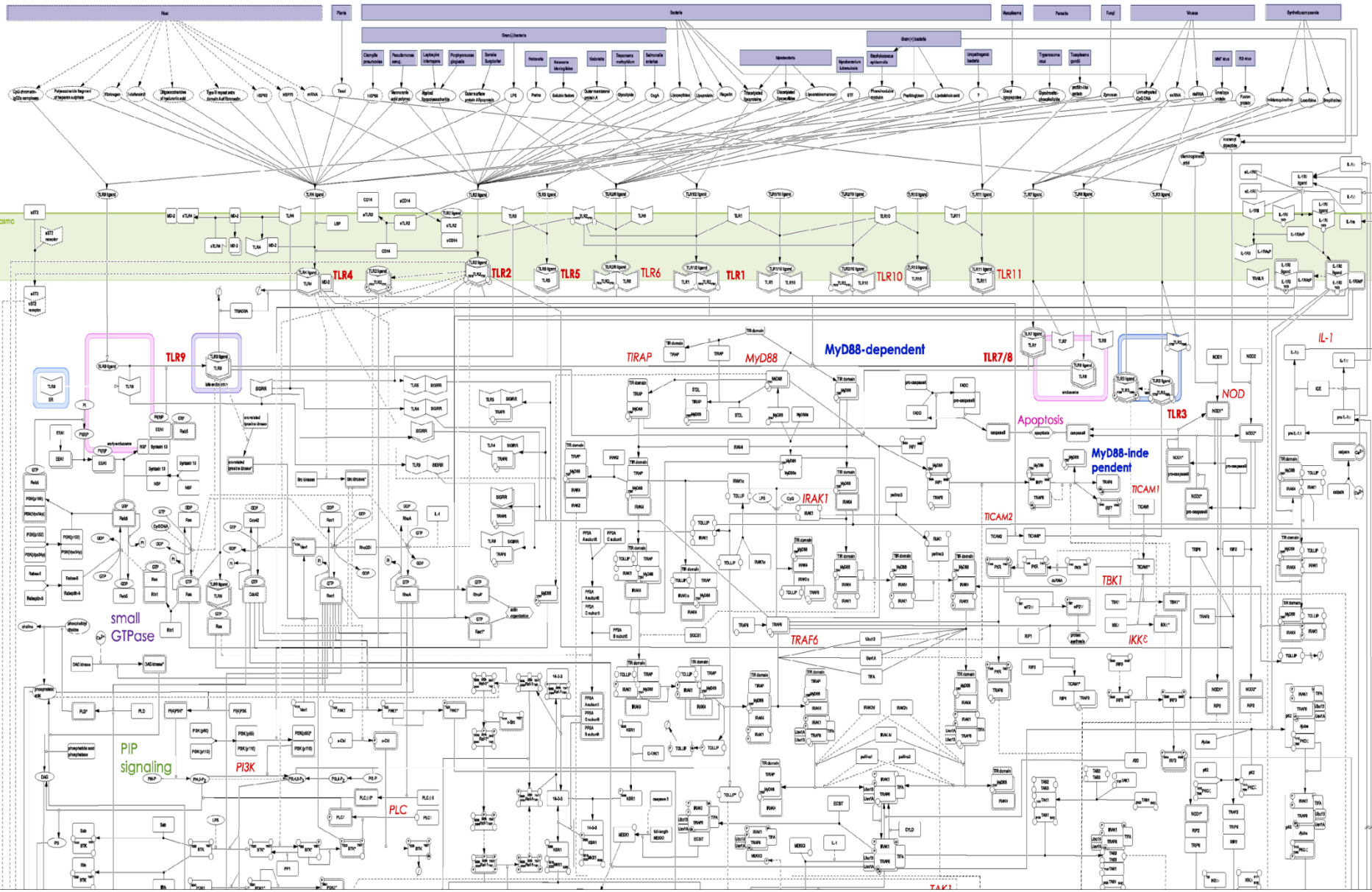





ГОСТ 427-75
MM 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
TIANBO
TRA3 L-12VDC-S-2Z
MIC
RS207
PM236CD
LM324N
0602
07-05-1870

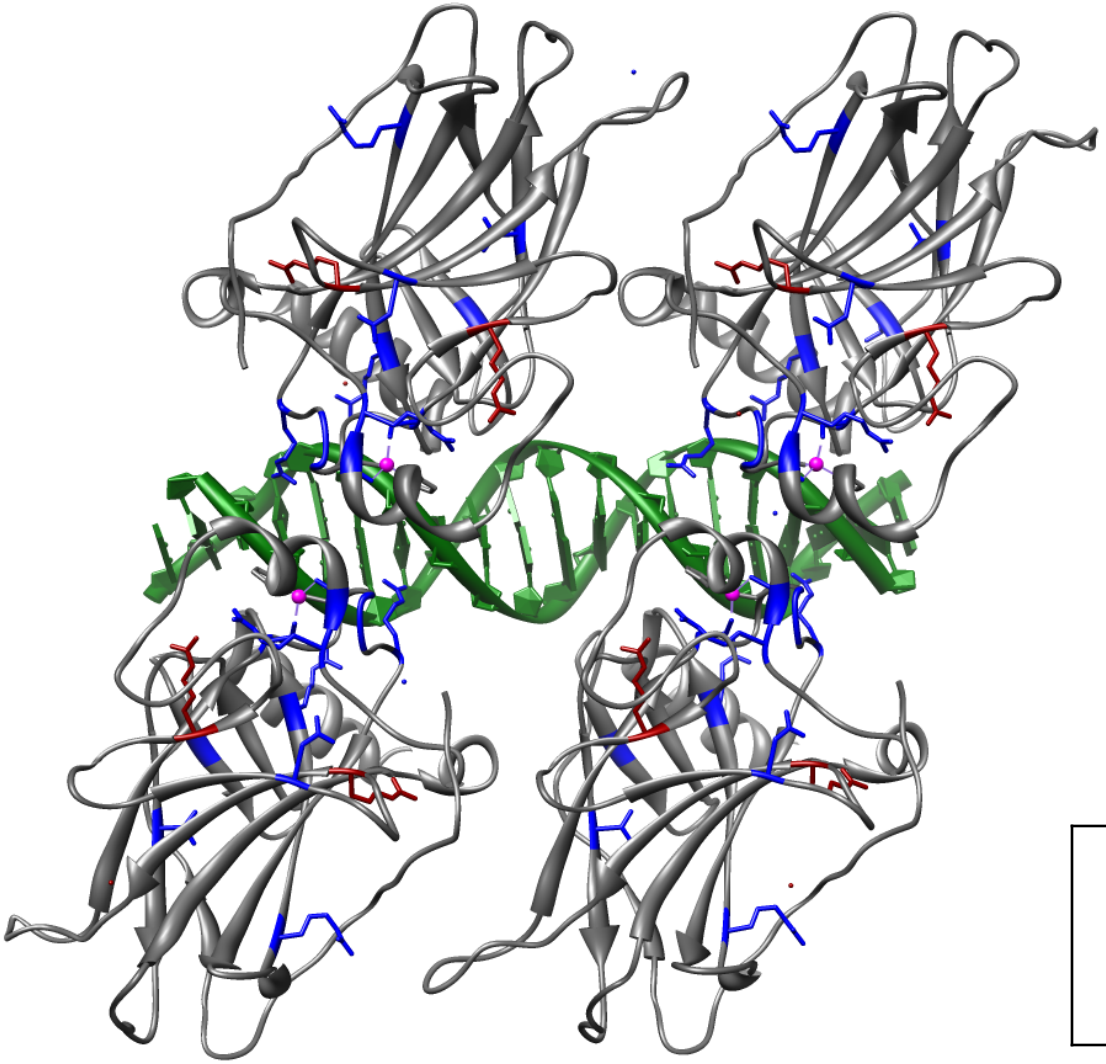


1 Value of component parts, circuits are changed to


Host
Plants
Bacteria
Gram(+) bacteria
Gram(-) bacteria
Mycobacteria
Mycoplasma
Parasite
Fungi
Viruses
Synthetic compounds
Plasma
TLR4
TLR2
TLR5
TLR6
TLR1
TLR10
TLR11
TLR9
TLR7/8
TLR3
MyD88-dependent
MyD88-independent
TIRAP
MyD88
IRAK1
TRAF6
TICAM1
TICAM2
TBK1
IKKε
NOD
Apoptosis
IL-1
small GTPase
PIP signaling
PI3K
PLC

COAD – TP53 Mutations

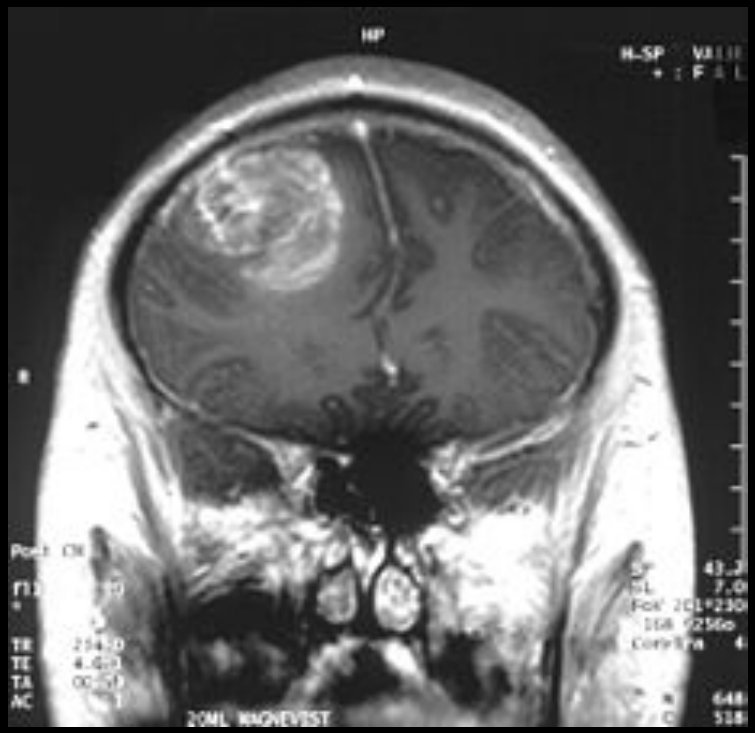
Transactivation Domain | DNA Binding Domain | Tetramerization Domain

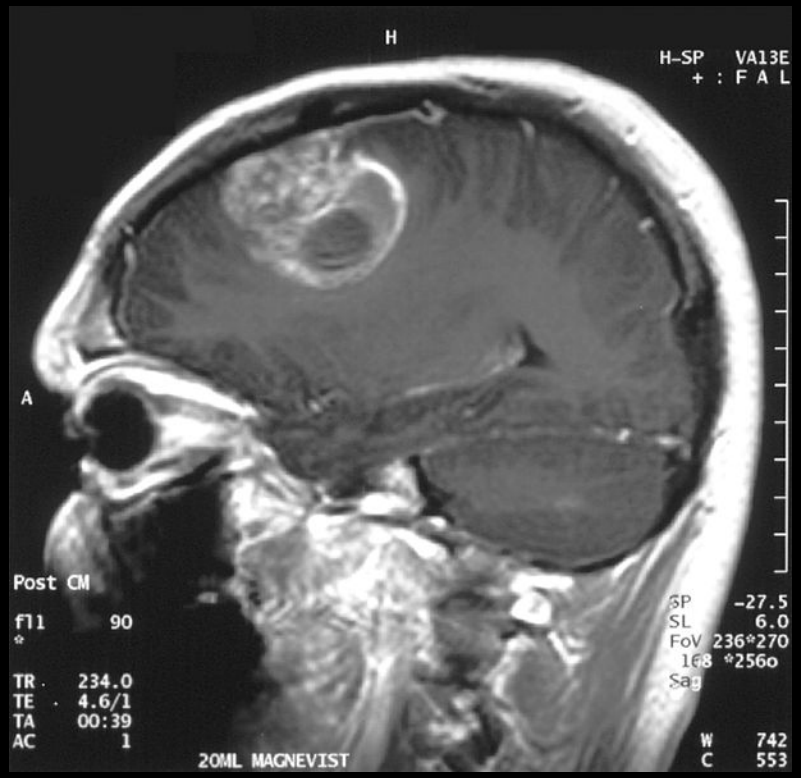


Missense mutation

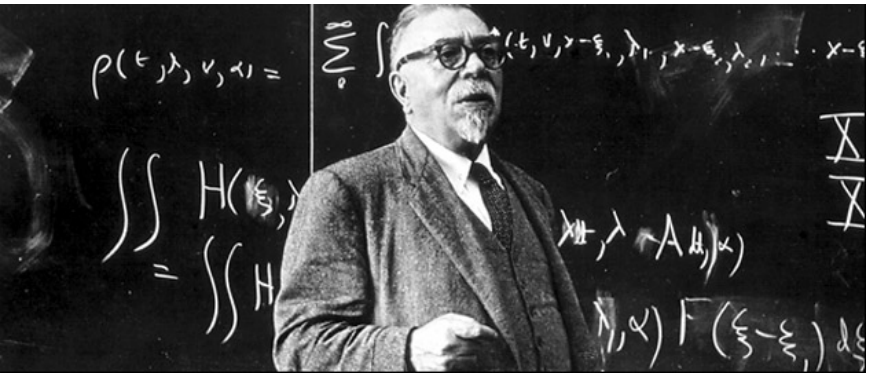
Nonsense mutation

The Cancer Genome Atlas

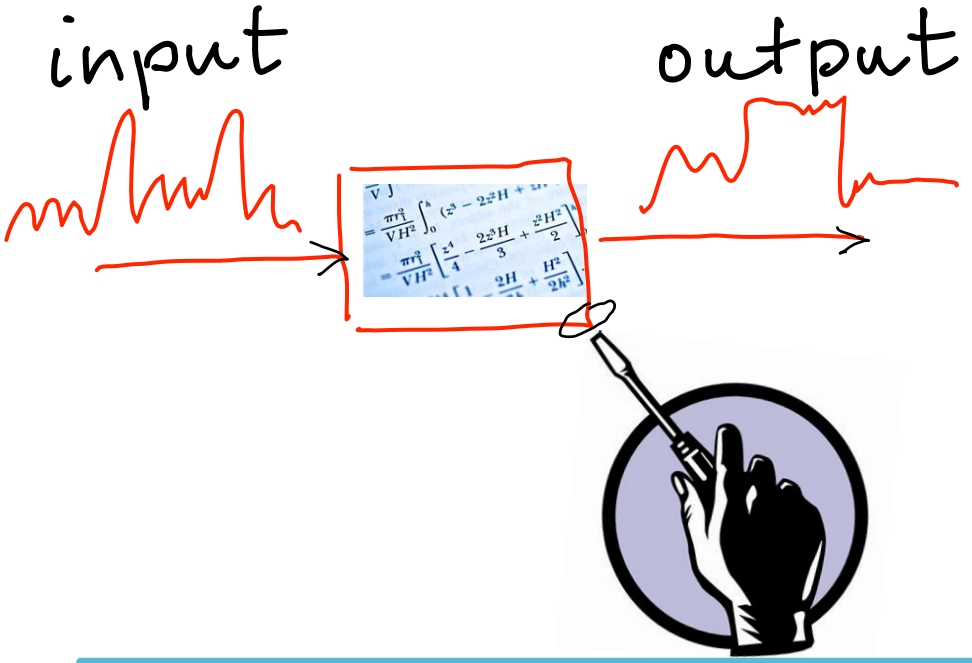

HP
20ML MAGNEVIST

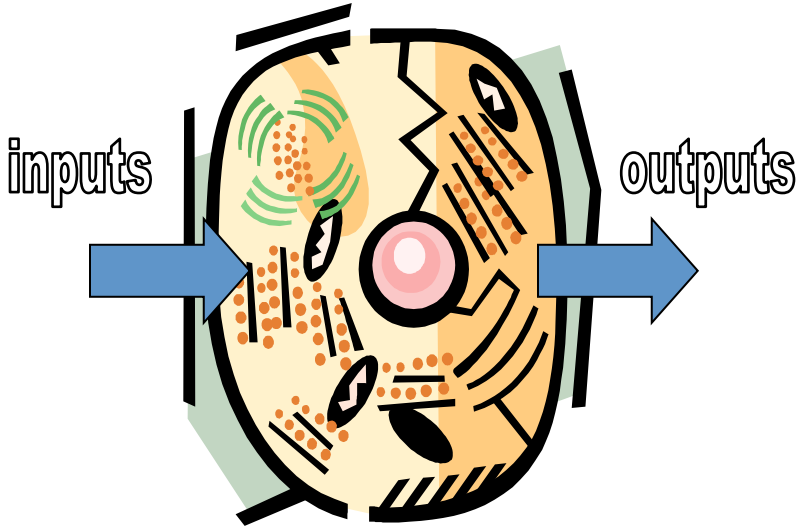

H
H-SP VA13E
+ : F A L
A
Post CM
f11 90
TR 234.0
TE 4.6/1
TA 00:39
AC 1
SP -27.5
SL 6.0
FoV 236*270
168 *256o
Sag
W 742
C 553
20ML MAGNEVIST

Need system model for understanding and control



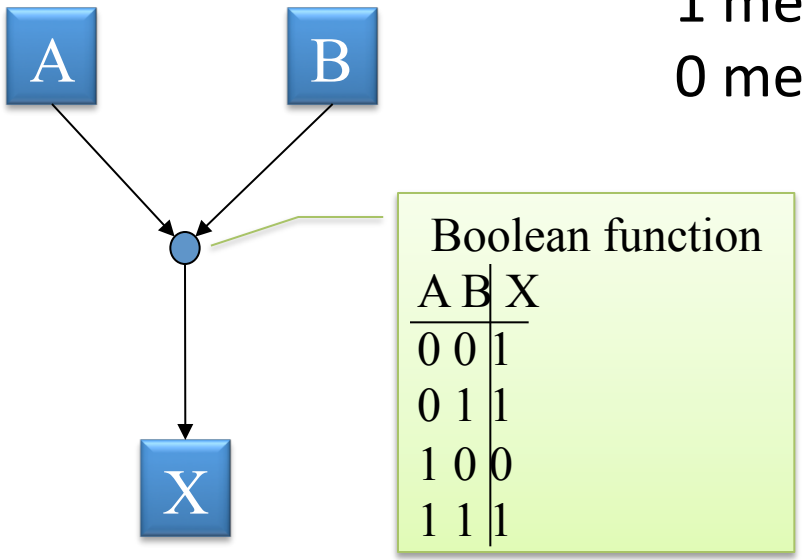
Norbert Wiener (1894 –1964)





"Many perhaps do not realize that the present age is ready for a significant turn in the development toward far greater heights than we have ever anticipated. The point of departure may well be the recasting and unifying of the theories of control and communication in the machine and in the animal on a statistical basis." *Norbert Wiener*, **1949**

Basic Structure of Boolean Networks

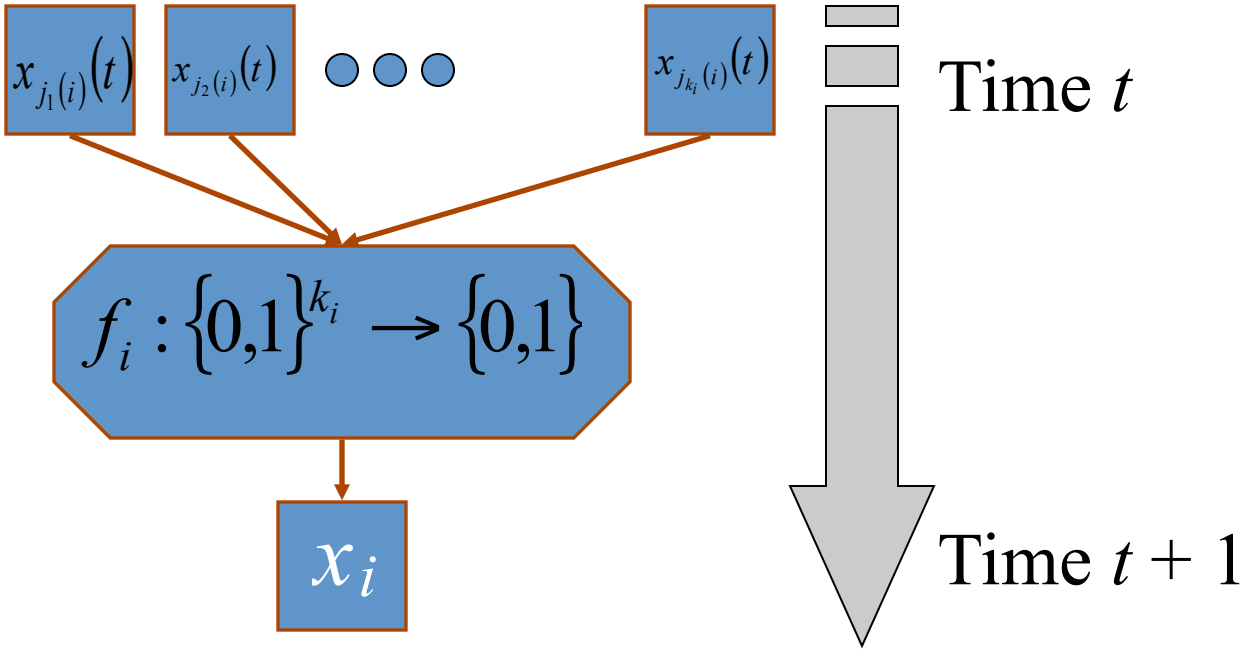


1 means *active/expressed*
0 means *inactive/unexpressed*

In this example, two genes (A and B) regulate gene X. In principle, any number of "input" genes are possible. Positive/negative feedback is also common (and necessary for homeostasis).

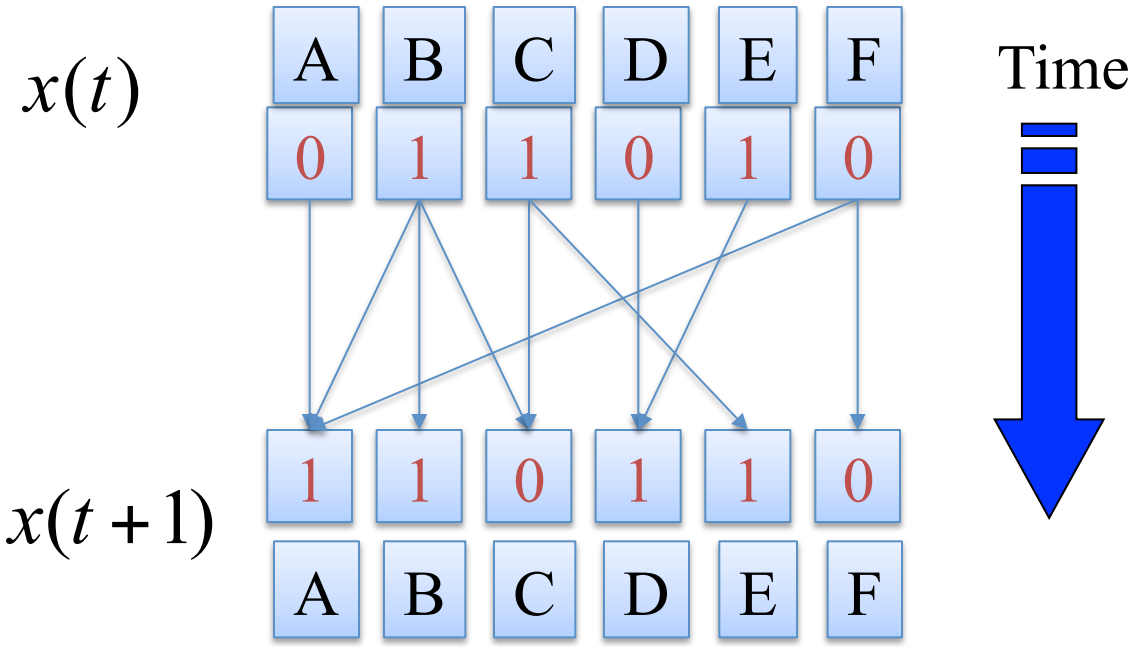
Boolean networks

A Boolean network contains n elements (genes) $x_1, \ldots, x_n$

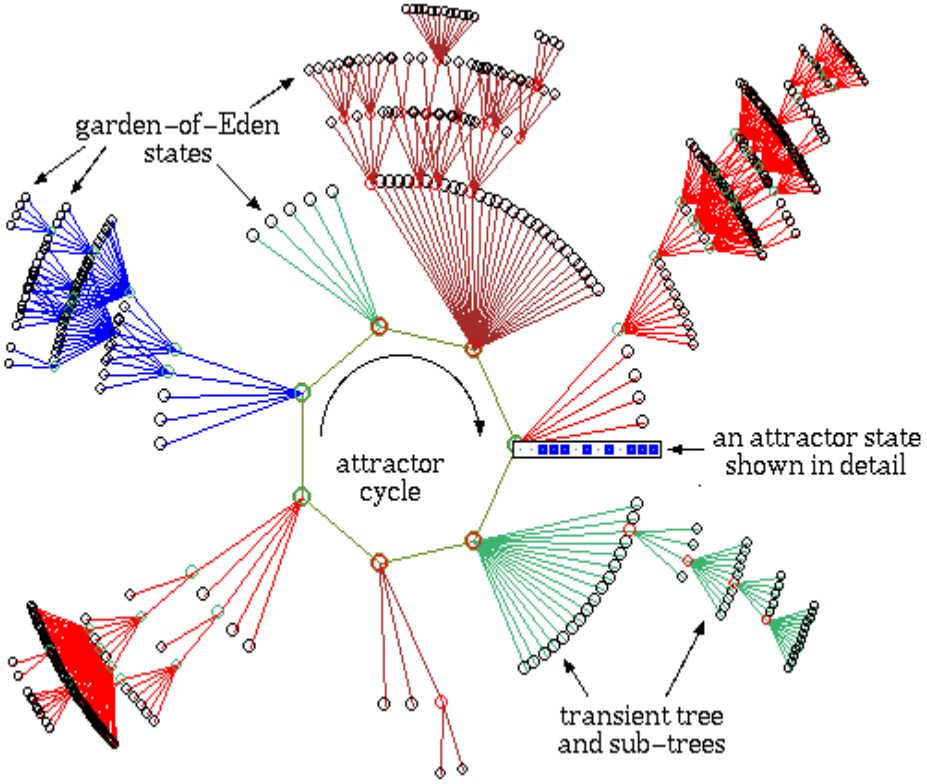


$$x_i(t+1) = f_i(x_{j_1(i)}(t), x_{j_2(i)}(t), \ldots, x_{j_{k_i}(i)}(t))$$

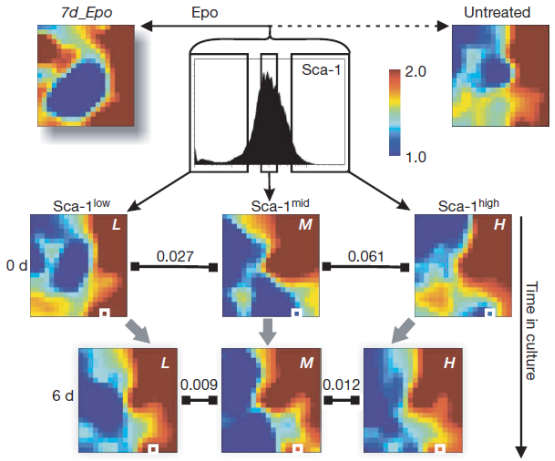
Dynamics of Boolean Networks



State Space of Boolean Networks



- **equate cellular states with attractors**
- attractor states are <u>stable under small perturbations</u>
  - most perturbations cause the network to flow back to the attractor.
  - <u>some genes are more important</u> and changing their activation can cause the system to transition to a different attractor.



recent findings provide experimental evidence for the existence of attractors in real regulatory networks

Implications for biology

This stability is physiologically important – it allows the cell to maintain its functional state within the tissue even under perturbations.

Nevertheless, cells do switch states, e.g. from quiescence to growth, usually when certain genes are affected by extracellular signals.

The cell "translates" such signals into specific alterations of genes/proteins.

- cell surface receptors are "wired to master switches" and are good targets for manipulation.

Implications for biology

- Hysteresis: a change in the system's state caused by a stimulus is not changed back after the stimulus is withdrawn.
  - Networks have this kind of "memory".
  - It may also account for the fact that adaptive changes are often preserved through many cell division generations.
  - Stability and hysteresis could explain inheritance of gene expressions.

Tumorigenesis

- Disturbance of the balance between attractors could be caused by mutations affecting the "wiring" or activation of important genes.
  - for example, stabilizing the growth state could lead to tumorigenesis.
  - such mutations change the size of the *basins of attraction*.
  - since the state space is finite, an increase of one basin of attraction leads to a decrease of another, say, differentiation.

Drug discovery

Most research has focused on the *linear paradigm*.

- manipulation of individual molecular targets

Robustness of attractor states explains why single-gene perturbations have had little success on the macroscopic level.

Rethinking the "functions" of genes… to regulate the dynamics of attractors.

Biological networks can often be modeled as logical circuits from well-known local interaction data in a straightforward way.

- This is clearly one of the advantages of the Boolean network approach.

Though logical models may sometimes appear obvious and simplistic, compared to detailed kinetic models of biomolecular reactions, they may help to understand the dynamic key properties of a regulatory process.

Further, a Boolean network model can be formulated as a coarse-grained limit of the more detailed differential equations model for a system (Davidich and Bornholdt, 2008).

They may also lead the experimentalist to ask new question and to test them first *in silico*.

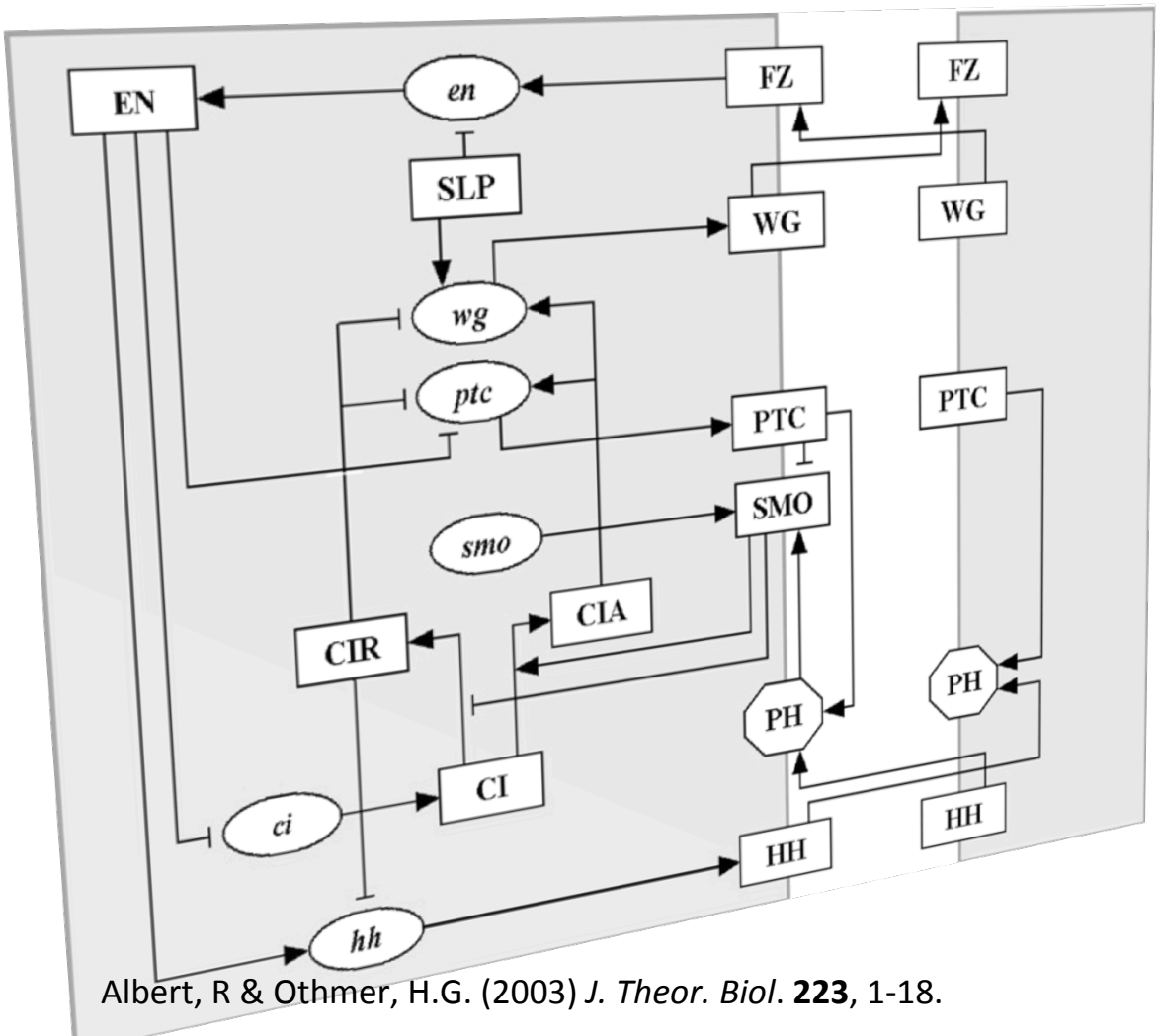
The segment polarity network of the fruit fly

The segment polarity genes represent the last step in the hierarchical cascade of gene families initiating the segmented body of the fruit fly *Drosophila melanogaster*

The stable maintenance of the segment polarity gene expression pattern is a crucial requirement in the further development of the embryo.

The dynamics of the best characterized segment polarity genes have been studied in different modeling approaches in order to understand the stability and robustness of this pattern formation

The network of interactions between the segment polarity genes

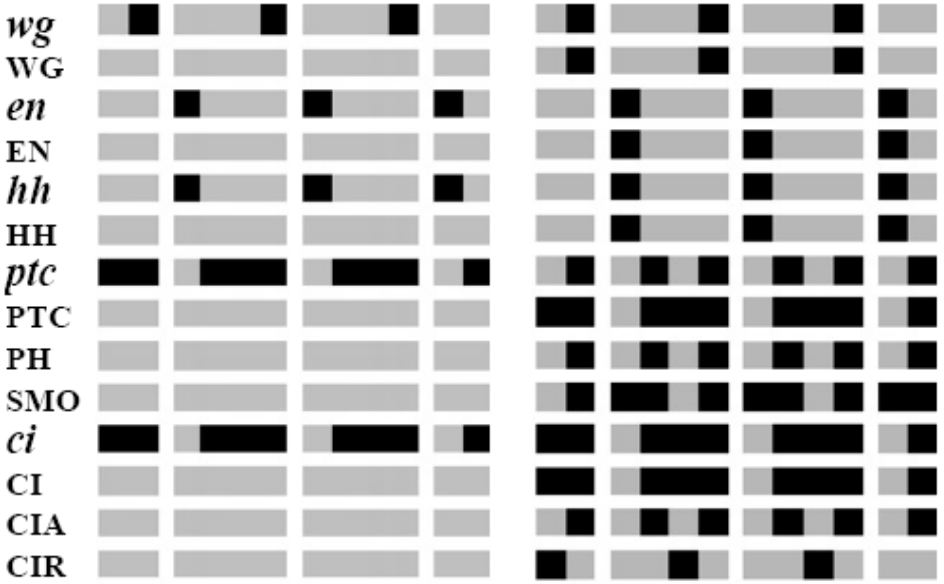


This model correctly reproduces the characteristic expression patterns observed in the wild-type development.

Further, distinct `knock down' or `overexpression' phenotypes can be simulated through the model by fixing the state of the particular gene.

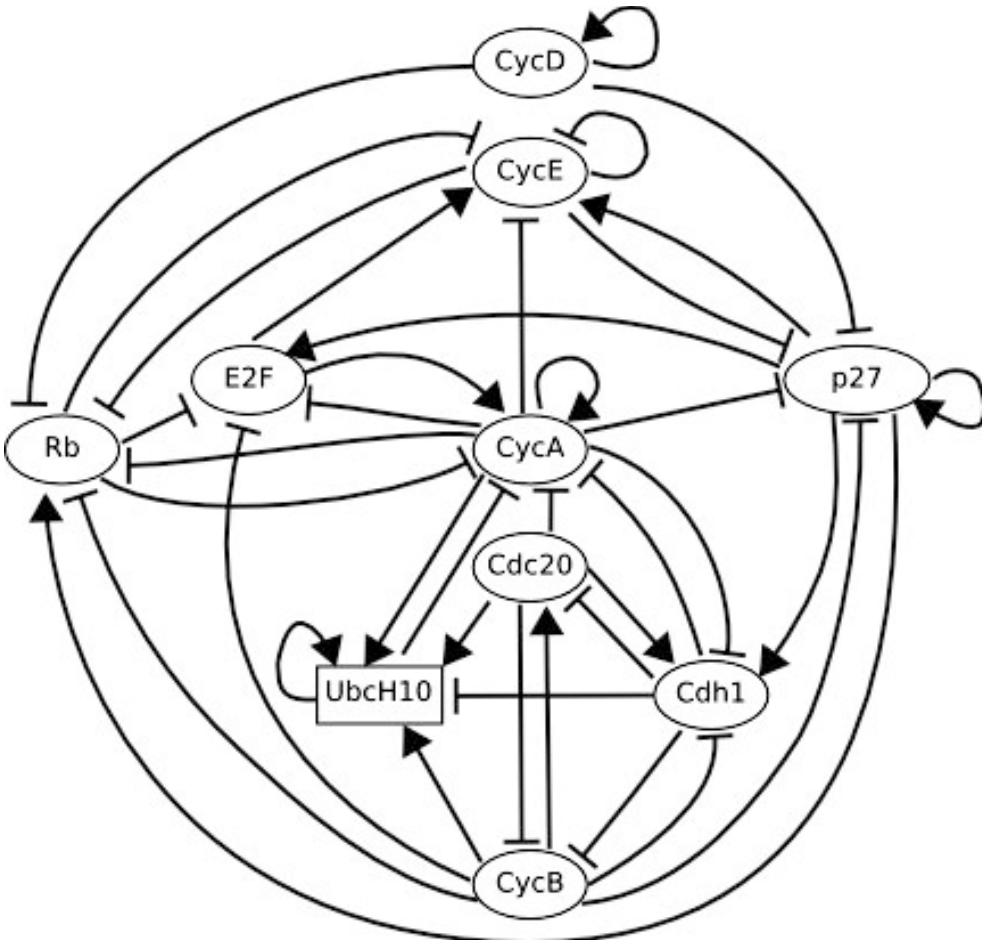
Albert, R & Othmer, H.G. (2003) *J. Theor. Biol*. **223**, 1-18.

Boolean rules and dynamics



Node	Boolean updating function
SLP_i	$SLP_i^{t+1} = SLP_i^t = \begin{cases} 0 & \text{if } i \bmod 4 = 1 \text{ or } i \bmod 4 = 2 \\ 1 & \text{if } i \bmod 4 = 3 \text{ or } i \bmod 4 = 0 \end{cases}$
wg_i	$wg_i^{t+1} = (CIA_i^t \text{ and } SLP_i^t \text{ and not } CIR_i^t) \text{ or } [wg_i^t \text{ and } (CIA_i^t \text{ or } SLP_i^t) \text{ and not } CIR_i^t]$
WG_i	$WG_i^{t+1} = wg_i^t$
en_i	$en_i^{t+1} = (WG_{i-1}^t \text{ or } WG_{i+1}^t) \text{ and not } SLP_i^t$
EN_i	$EN_i^{t+1} = en_i^t$
hh_i	$hh_i^{t+1} = EN_i^t \text{ and not } CIR_i^t$
HH_i	$HH_i^{t+1} = hh_i^t$
ptc_i	$ptc_i^t = CIA_i^{t+1} \text{ and not } EN_i^t \text{ and not } CIR_i^t$
PTC_i	$PTC_i^{t+1} = ptc_i^t \text{ or } (PTC_i^t \text{ and not } HH_{i-1}^t \text{ and not } HH_{i+1}^t)$
PH_i	$PH_i^t = PTC_i^t \text{ and } (HH_{i-1}^t \text{ or } HH_{i+1}^t)$
SMO_i	$SMO_i^t = \text{not } PTC_i^t \text{ or } HH_{i-1}^t \text{ or } HH_{i+1}^t$
ci_i	$ci_i^{t+1} = \text{not } EN_i^t$
CI_i	$CI_i^{t+1} = ci_i^t$
CIA_i	$CIA_i^{t+1} = CI_i^t \text{ and } (SMO_i^t \text{ or } hh_{i-1}^t \text{ or } hh_{i+1}^t)$
CIR_i	$CIR_i^{t+1} = CI_i^t \text{ and not } SMO_i^t \text{ and not } hh_{i\pm1}^t$

Control of the mammalian cell cycle



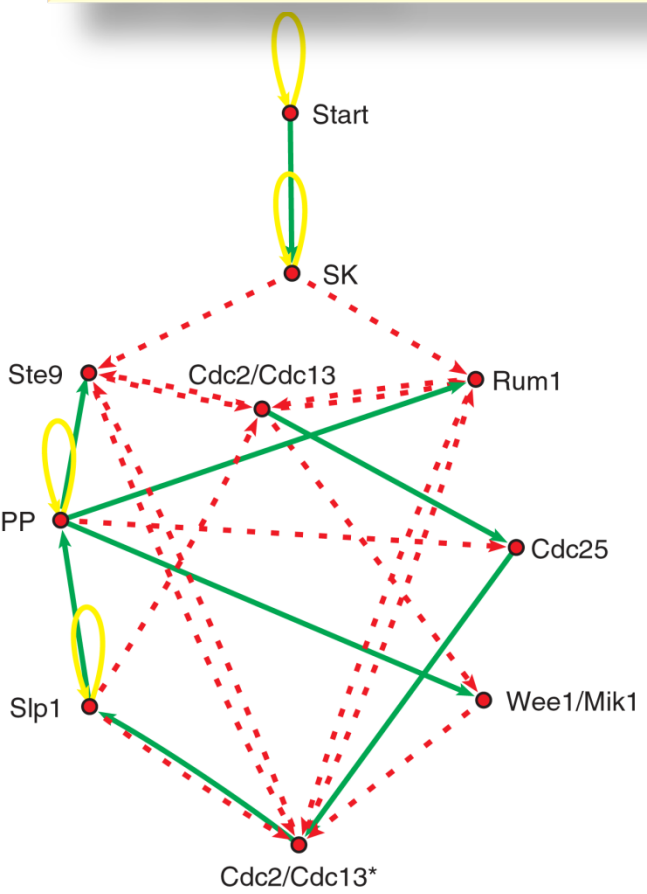
Under the simplistic synchronous updating scheme the network gives rise to two different attractors, a stable state, matching the quiescent cell state (G0) when growth factors are lacking and a complex dynamical cycle representing the cell cycle when the cyclin D complex is present.

The order of activity switching (off or on) matches the available data, as well as the time plots published in the literature.

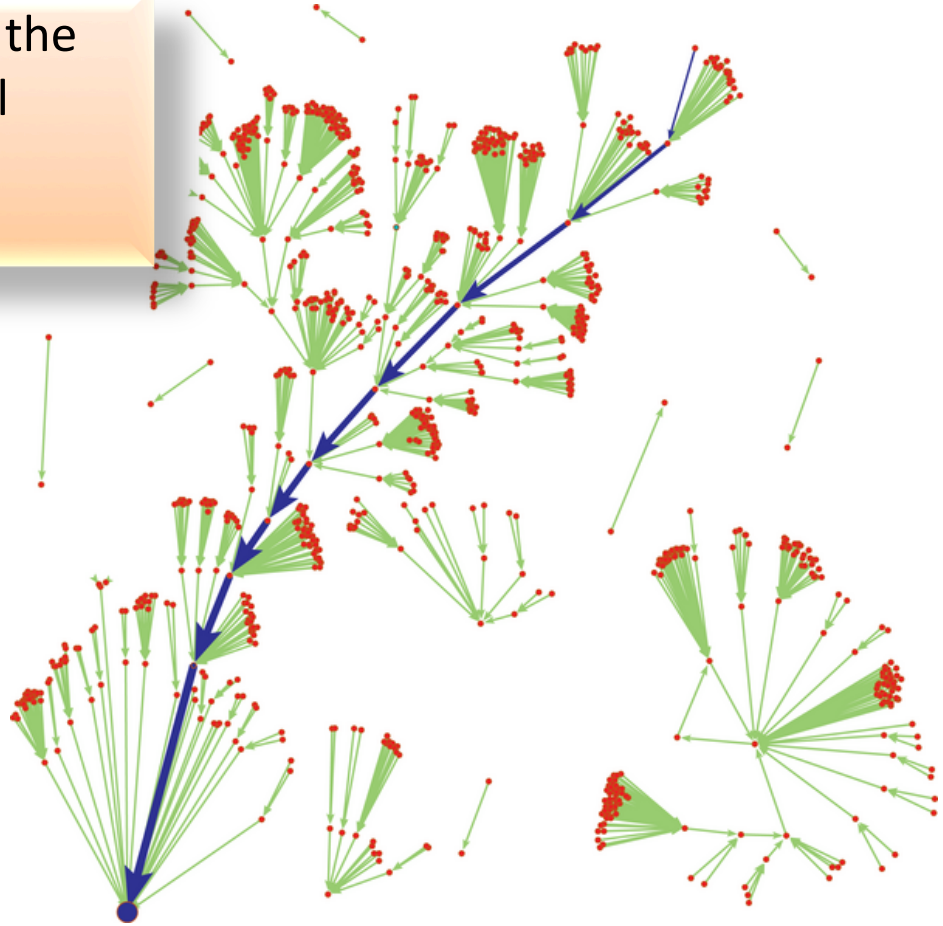
Bioinformatics **2006** 22(14):e124-e131

Model of the fission yeast cell cycle

The largest attractor attracts 73% of the entire state space, and the biological target state (G1) is robust to most perturbations.

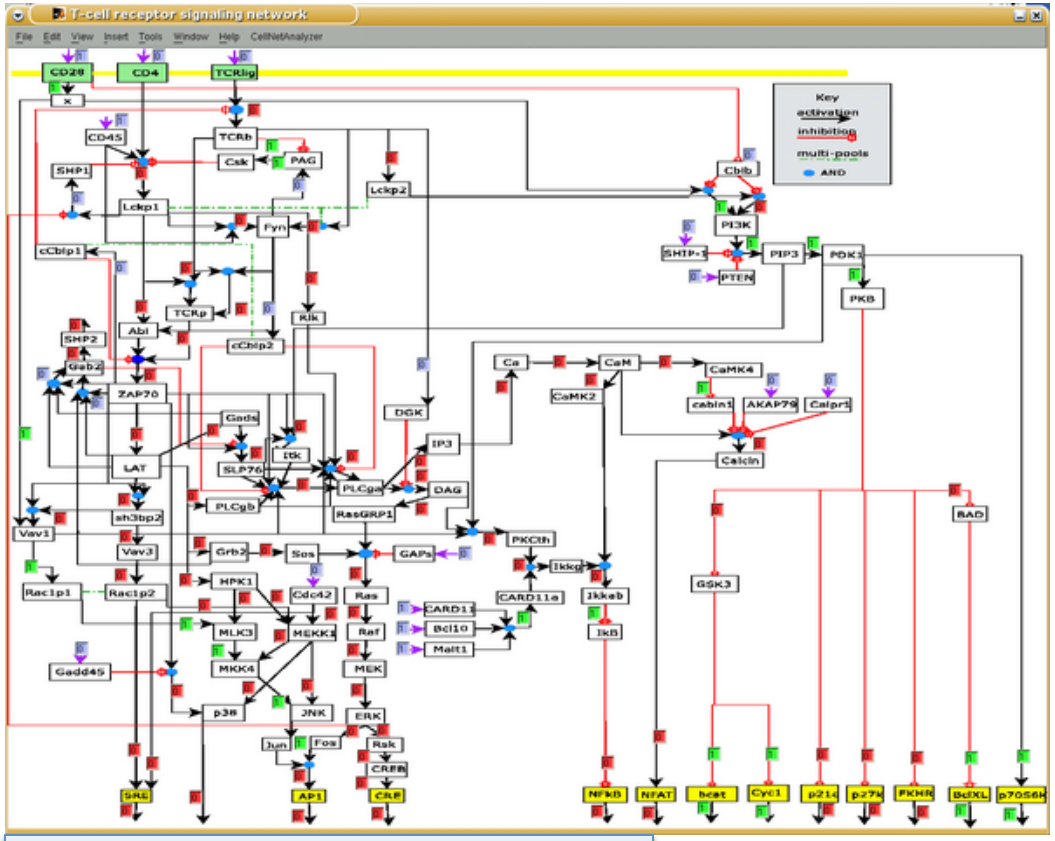


PLoS ONE 3(2): e1672



The cell cycle system for this organism is well understood in terms of established differential equation models

T-cell receptor signaling



94 nodes and 123 interactions

connected component comprising only 33 nodes

T cells recognize foreign peptides that are presented by antigen presenting cells by means of the T cell receptor (TCR) and costimulatory molecules (CD4/8/28).

This stimulus initiates a signaling cascade within the T cell that eventually influences its activation status.

The T cell response must be tightly regulated, as a response to a pathogen that is too weak endangers the whole organism, whereas an overshooting reaction may lead to autoimmune disorders.

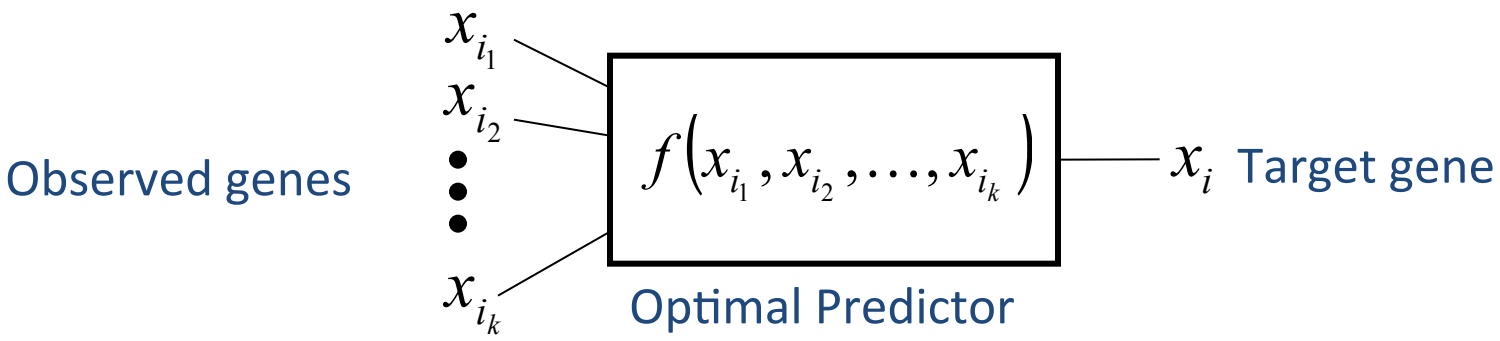
PLoS Computational Biology 3(8): e163

Model Inference from Gene Expression Data

Several approaches exist for Boolean networks:

- Coefficient of Determination
- Best-Fit Extensions
- Minimal Description Length (MDL)
- Mutual information
- Others…

COD Definition

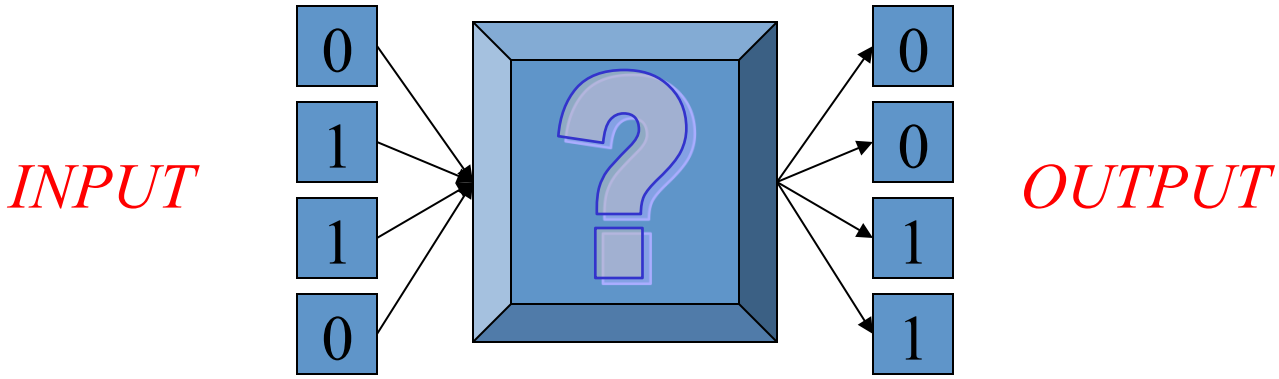


$$\theta = \frac{\varepsilon_i - \varepsilon_{opt}}{\varepsilon_i}$$

ε_i is the error of the best (constant) estimate of x_i in the absence of any conditional variables

ε_{opt} is the optimal error achieved by f

Best-Fit Extensions



- Measurement errors can arise in the data acquisition process or may be due to unknown latent factors.
- Best-Fit Extension paradigm can incorporate such inconsistencies.
- **Goal**: to establish a rule or in our case, network, that would make as few misclassifications as possible.

Best-Fit Problem Formulation

A partially defined Boolean function (pdBf) is defined by a pair of sets

$$T, F \subseteq \{0,1\}^n$$

A function f is called an extension of pdBf(T,F) if

$$T \subseteq T(f) \text{ and } F \subseteq F(f)$$

$$\text{where } T(f) = \{x \in \{0,1\}^n : f(x) = 1\}$$

$$F(f) = \{x \in \{0,1\}^n : f(x) = 0\}$$

We are also given positive weights $w(x), x \in T \cup F$

Define: $w(S) = \sum_{x \in S} w(x)$

Machine Learning, Vol. 52, pp. 147-167, 2003.

Best-Fit Problem Formulation

Then, the error size of function f is defined as

$$\varepsilon(f) = w(T \cap F(f)) + w(F \cap T(f))$$

Goal: Output subsets T^* and F^* such that $T^* \cap F^* = \varnothing$ and $T^* \cup F^* = T \cup F$

so that any extension f (from some class of functions) of pdBf(T^* ,F^*) has minimum error size.

Efficient algorithms for the Best-Fit problem have been developed.

Machine Learning, Vol. 52, pp. 147-167, 2003.

Constraints During Inference

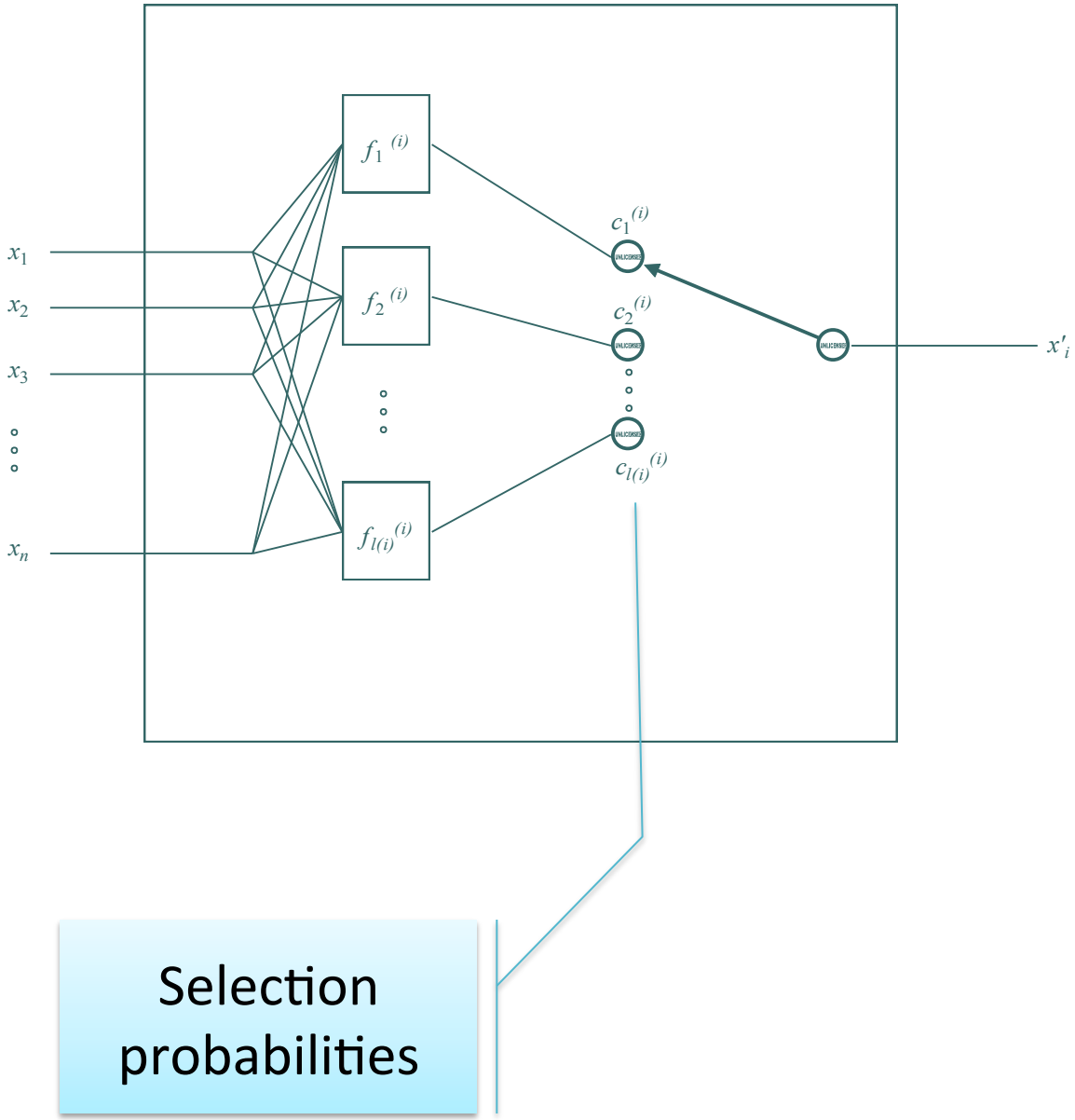
- Constraining the class of predictors can have advantages:
  - lessening the data requirements for reliable estimation;
  - incorporate prior knowledge of the class of functions representing genetic interactions;
  - certain classes of functions are more plausible from the point of view of evolution, noise resilience, network dynamics, etc.
- An example of such constraints are canalizing functions and Post classes

PNAS 100:19,10734-10739; PNAS 101:17102-17107; PNAS 100:14796-14799

Probabilistic Boolean Networks (PBNs)

- Share the appealing rule-based properties of Boolean networks.
- Robust in the face of uncertainty.
- Dynamic behavior can be studied in the context of Markov Chains.
  - Boolean networks are just special cases.
- Close relationship to (dynamic) Bayesian networks
  - Explicitly represent probabilistic relationships between genes.
  - Can represent the same joint probability distribution.
- Allow quantification of influence of genes on other genes.

Basic structure of PBNs



If we have several "good" competing predictors (functions) for a given gene and each one has "determinative power," don't put all our faith in one of them!

$x_1x_2x_3$	f_1^1	f_2^1	f_1^2	f_1^3	f_2^3
000	0	0	0	0	0
001	1	1	1	0	0
010	1	1	1	0	0
011	1	0	0	1	0
100	0	0	1	0	0
101	1	1	1	1	0
110	1	1	0	1	0
111	1	1	1	1	1
c_j^i	0.6	0.4	1	0.5	0.5

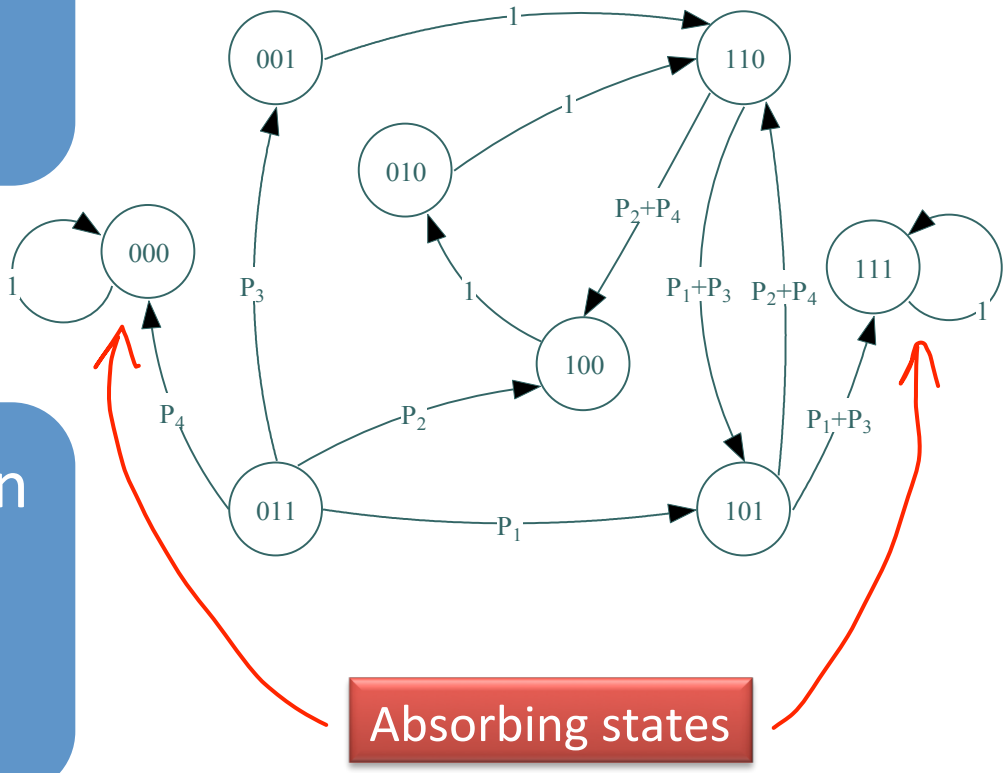
Selection probabilities

Dynamics of PBNs

Dynamics of PBNs can be studied using Markov Chain theory. From the Boolean functions, we can compute

- transition probabilities
- steady-state distribution (if it exists)

We can ask the question: "In the long run, what is the probability that some given gene(s) will be ON/OFF?"



Random Gene Perturbations

- Genes can sometimes change value with a small probability *p*.
  - The genome is not a closed system – genes can be activated/inhibited due to external variables

$$\text{Perturbation vector } \gamma \in \{0,1\}^n$$
$$\Pr\{\gamma_i = 1\} = E[\gamma_i] = p$$

Random Gene Perturbations

$$x' = \begin{cases} x \oplus \gamma, & \text{with probability } 1-(1-p)^n \\ \mathbf{f}_k(x_1, \ldots, x_n), & \text{with probability } (1-p)^n \end{cases}$$

- If no genes are perturbed, the standard network transition function will be used.
- Observation:
  - For $p > 0$, the Markov chain corresponding to the PBN is ergodic.
    - Thus, the steady-state distribution exists.
    - Convergence partially depends on p.

Transition Probabilities

Theorem 2 *Given a PBN* $G(V, F)$ *with genes* $V = \{x_1, \ldots, x_n\}$ *and a list* $F = (F_1, \ldots, F_n)$ *of sets* $F_i = \left\{ f_1^{(i)}, \ldots, f_{l(i)}^{(i)} \right\}$ *of Boolean predictors, as well as a gene perturbation probability* $p > 0$,

$$A(x, x') = \left(\sum_{i=1}^{N} P_i \left[\prod_{j=1}^{n} \left(1 - \left| f_{K_{ij}}^{(j)}(x_1, \ldots, x_n) - x'_j \right| \right) \right] \right) \times (1-p)^n + \\ + p^{\eta(x,x')} \times (1-p)^{n-\eta(x,x')} \times 1_{[x \neq x']},$$

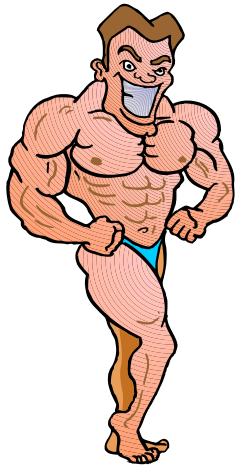
where $\eta(x, x') = \sum_{i=1}^{n} (x_i \oplus x'_i)$ *is the Hamming distance between vectors* x *and* x', P_i *is given in (2), and* $1_{[x \neq x']}$ *is an indicator function that is equal to 1 only when* $x \neq x'$.

It is possible to compute state transitions directly from the Boolean functions, the selection probabilities, and the perturbation probability.

Influence of genes

- In a Boolean function, some variables have greater "determinative power" on the output.
- *Influence* is defined in terms of the partial derivative of the Boolean function and the underlying joint probability distribution of the inputs (efficient spectral methods exist)
- PBNs naturally allow us to compute influences between (sets of) genes
  - genes with a high influence may make potentially good targets for intervention.

Example



$$f(x_1, x_2, x_3) = x_1 + x_2 \cdot x_3$$

Influence

$$\frac{\partial f(x)}{\partial x_j} = f(x^{(j,0)}) \oplus f(x^{(j,1)}),$$

where $\oplus$ is addition modulo 2 (exclusive OR) and

$$x^{(j,k)} = (x_1, \ldots, x_{j-1}, k, x_{j+1}, \ldots, x_n),$$

for $k = 0, 1$.

The *influence* of the variable x_j on the function f is the expectation of the partial derivative with respect to the distribution $D(x)$:

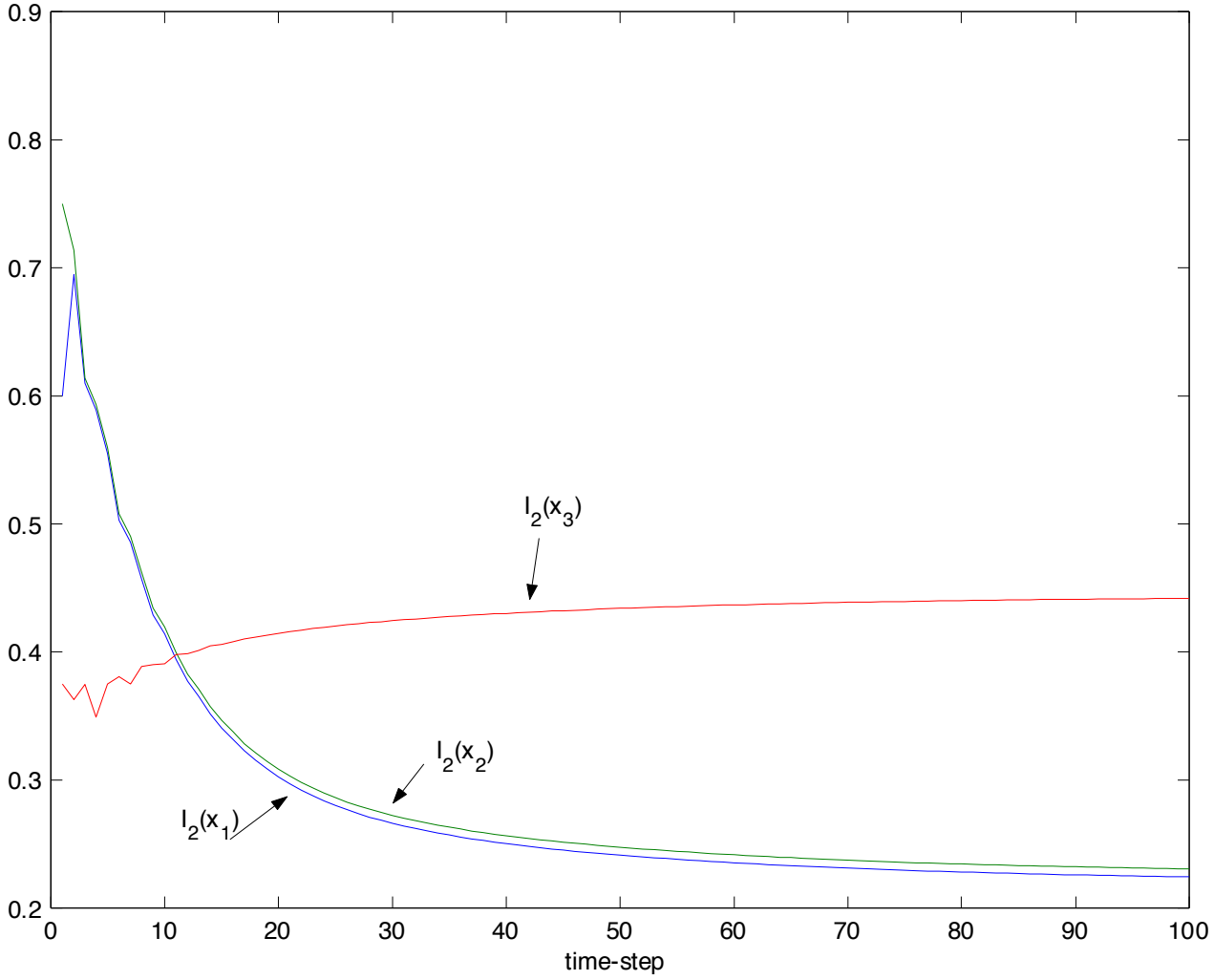
$$I_j(f) = E_D\left[\frac{\partial f(x)}{\partial x_j}\right]$$

$$= \Pr\left\{\frac{\partial f(x)}{\partial x_j} = 1\right\} = \Pr\{f(x) \neq f(x^{(j)})\}.$$

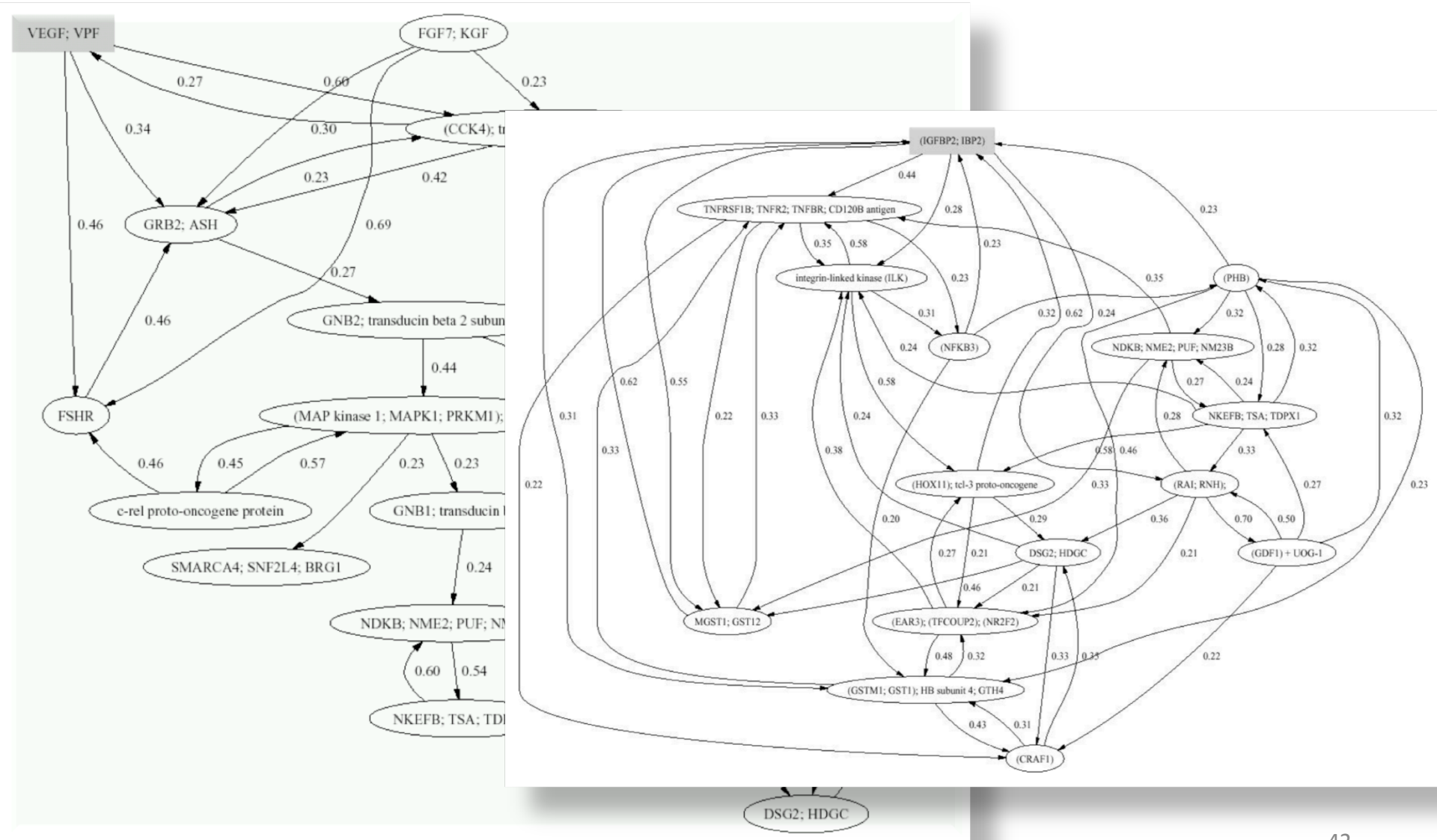
Influence and Sensitivity

- We can easily define the influence of a gene on another (set of) gene(s), in the PBN framework.
- We can also define the *sensitivity* of a gene (definition omitted here).
  - Biologically, this represents the stability, or in some sense, the "autonomy" of a gene.

Long-term Influence



Examples of PBN influences in human glioma

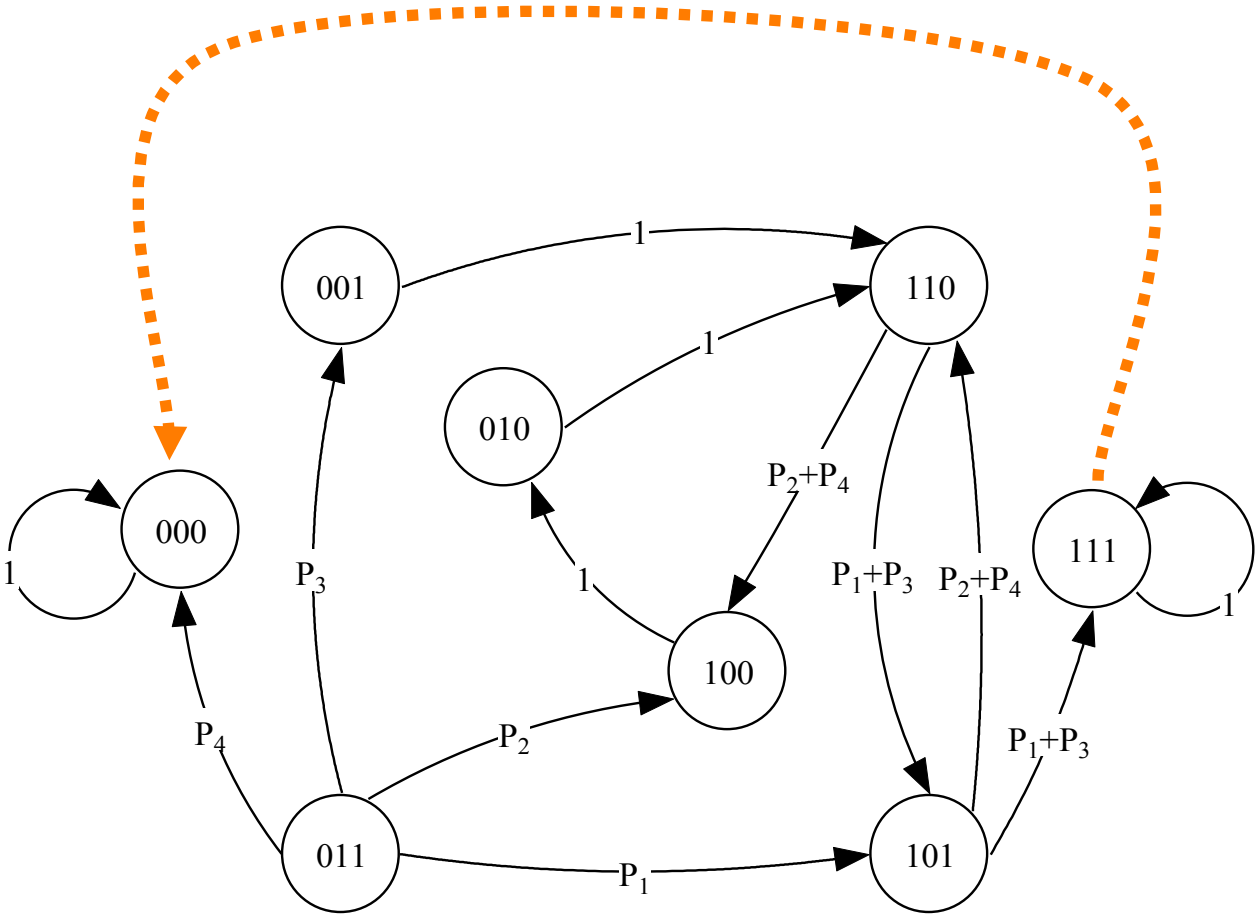


Bioinformatics, Vol. 20, No. 8, pp. 1241-1247, 2004.

Intervention

- One of the key goals of PBN modeling is the determination of possible intervention targets (genes) such that the network can be "persuaded" to transition into a desired state or set of states.
- Clearly, perturbation of certain genes is more likely to achieve the desired result than that of some other genes.
- Our goal, then, is to discover which genes are the best potential "lever points" in the sense of having the greatest possible impact on desired network behavior.

Example

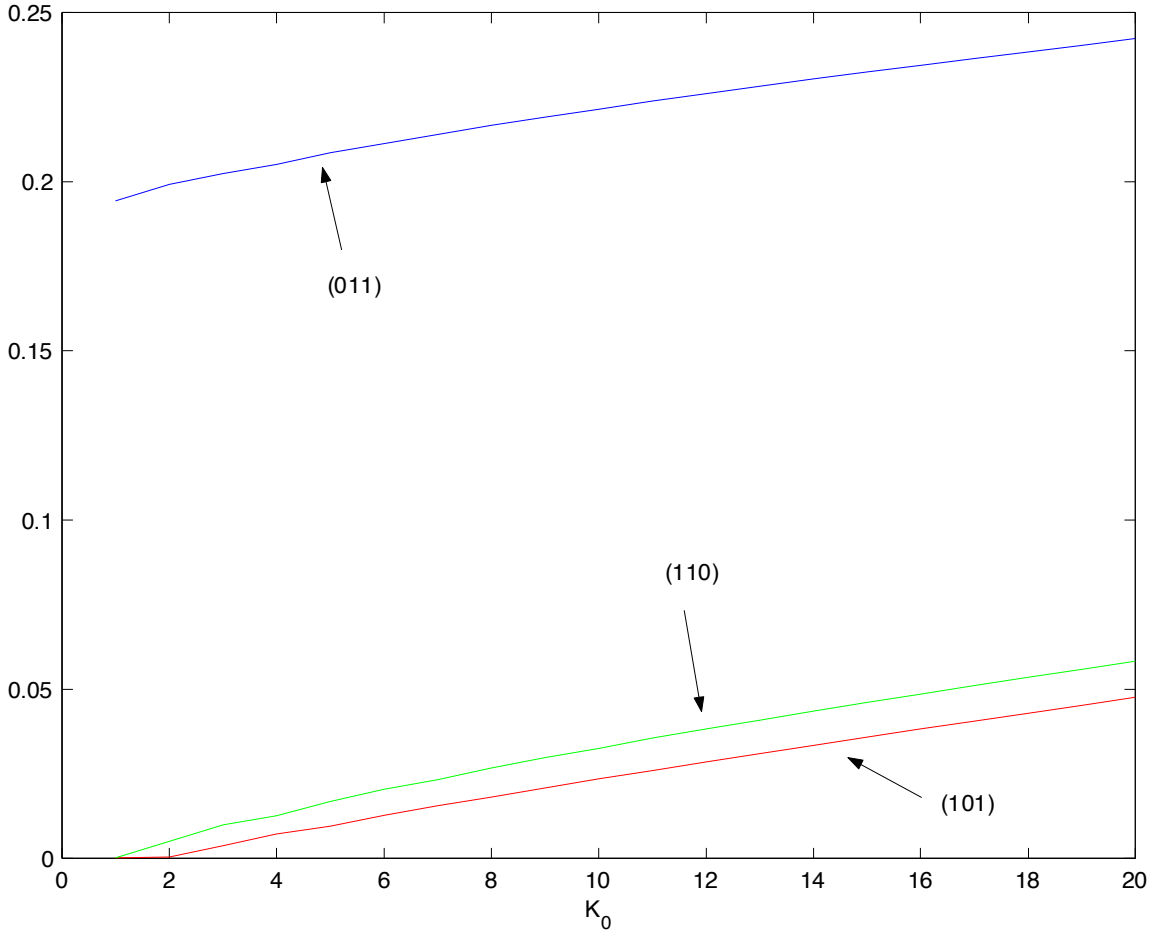


Clearly, the choice in this simple example should be gene x_1.

Intervention

- When $p > 0$, the entire Markov chain is ergodic and thus, every state will eventually be visited.
- Thus, the question of intervention should be posed in the sense of reaching a desired state as early as possible.
- We use first passage times.
- In biology, there are numerous examples when the (in)activation of one gene or protein can lead much quicker (or with a higher probability) to a certain cellular functional state or phenotype than the (in)activation of another gene or protein.

Same Example as Before



There are several possibilities: find the gene that

- minimizes the mean first passage time
- maximizes the probability of reaching a particular state before a certain fixed time
- minimizes the time needed to reach a certain state with a given fixed probability.

Sensitivity of Stationary Distributions to Gene Perturbations

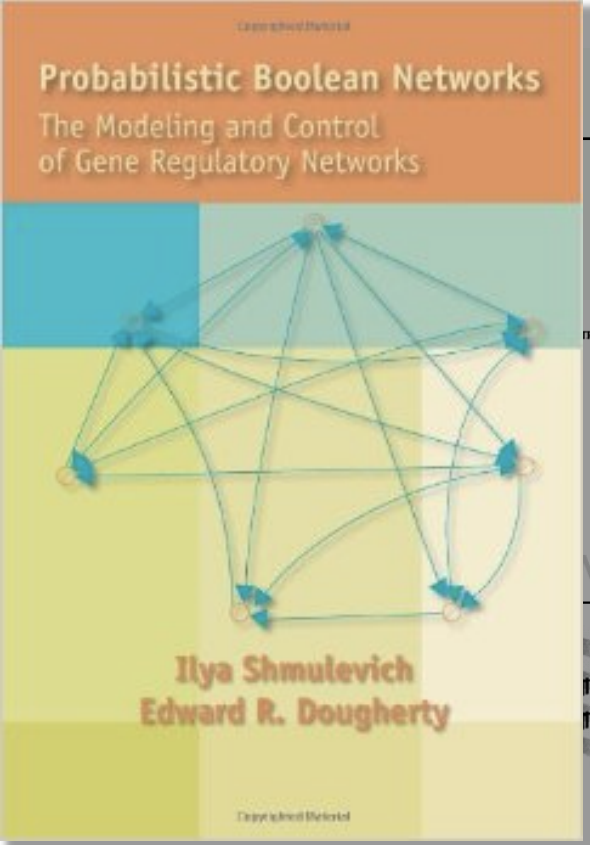
- What is the effect of perturbations on long-term network behavior?
- Similar problems have been addressed in perturbation theory of stochastic matrices.
- Using results by Cho & Meyer (2000), we can show…

Sensitivity Result

Theorem *Given a PBN $G(V,F)$ with an existing steady-state distribution, let π_y be a limiting probability of state y when $p=0$ (no perturbations) and let $\tilde{\pi}_y$ be the limiting probability of the same state when $0<p<1/2$. Then,*

$$\frac{|\pi_y - \tilde{\pi}_y|}{\pi_y} \leq (1-(1-p)^n) \max_{x \neq y} M(x,y).$$

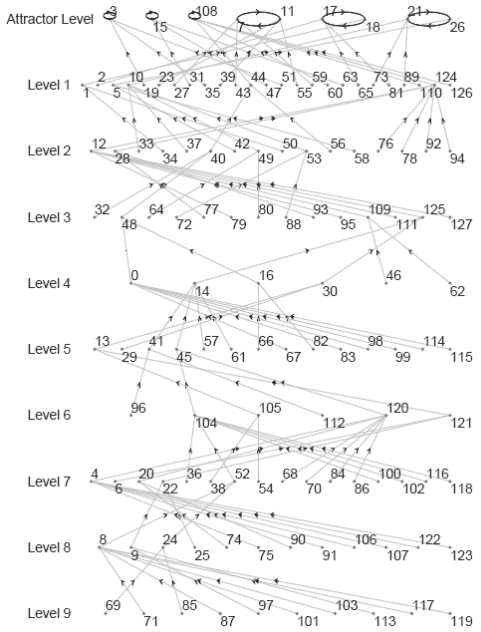
One important implication is that if a particular state of a PBN can be "easily reached" from other states, meaning that the mean first passage times are small, then its steady-state probability will be relatively unaffected by perturbations. Such sets of states, if we hypothesize them to correspond to some functional cellular states, are thus relatively insensitive to random gene perturbations.



BIOINFORMATICS
Vol. 18 no. 10 2002

Gene perturbation and intervention in probabilistic Boolean networks

Ilya Shmulevich[1,*], Edward R. Dougherty[2] and Wei Z



TRANSACTIONS ON SIGNAL PROCESSING, VOL. 54, NO. 6, JUNE 2006

Optimal Infinite-Horizon Control for Probabilistic Boolean Networks

Ranadip Pal, *Student Member, IEEE*, Aniruddha Datta, *Senior Member, IEEE*, and Edward R. Dougherty, *Member, IEEE*

NFORMATICS
Vol. 20 no. 6 2004,
DOI: 10.1093/bioinfc

External control in Markovian genetic regu networks: the imperfect information case

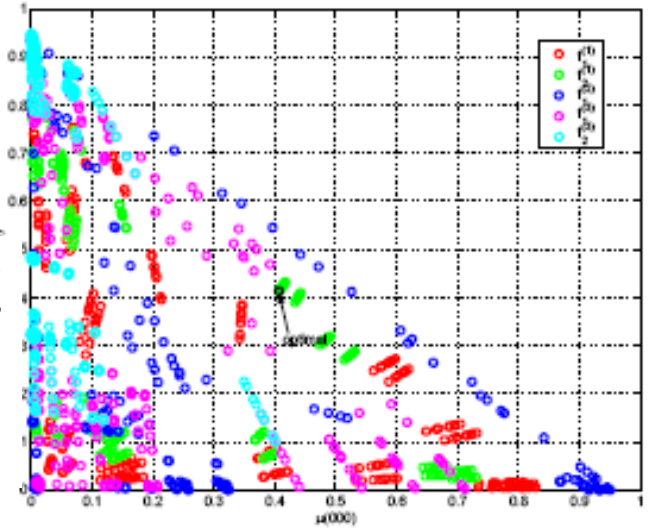
Aniruddha Datta[1], Ashish Choudhary[1], Michael L. Bittner[2] Edward R. Dougherty[1,3,*]

BIOINFORMATICS ORIGINAL PAPER
Vol. 21 no. 7 2005, pages 1211–1218
doi:10.1093/bioinformatics/bti131

Systems biology

Intervention in context-sensitive probabilistic Boolean networks

Ranadip Pal[1], Aniruddha Datta[1], Michael L. Bittner[2] and Edward R. Dougherty[1,3,*]



BIOINFORMATICS ORIGINAL PAPER
Vol. 22 no. 2 2006, pages 226–232
doi:10.1093/bioinformatics/bti765

Systems biology

Intervention in a family of Boolean networks

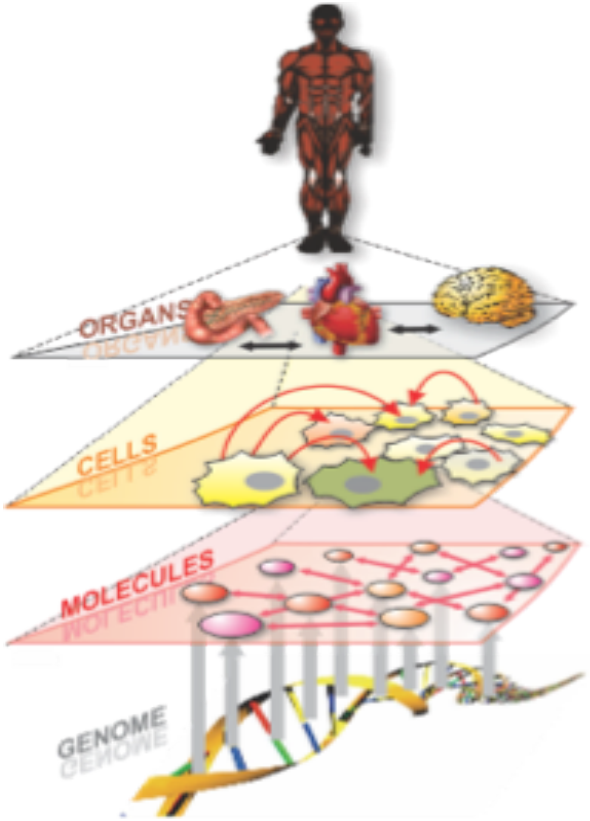
Ashish Choudhary[1], Aniruddha Datta[1], Michael L. Bittner[2] and Edward R. Dougherty[1,2,*]

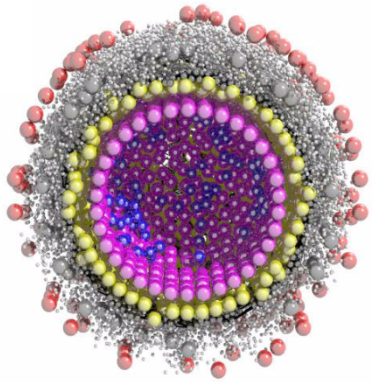
Biocellion

Accelerating Multicellular Biological System Simulation

- Individual Cell Behavior
- Cell-cell interaction
  1. Mechanical contact
  2. Diffusible molecules
- Cell-environment interaction
- Scale to billions of cells

Kang et al, Bioinformatics, 30(21): 3101-3108, 2014.





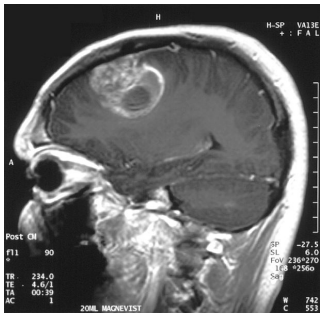
Seunghwa Kang, Simon Kahan, Nick Flann, Jason McDermott

Utah State University

Pacific Northwest NATIONAL LABORATORY

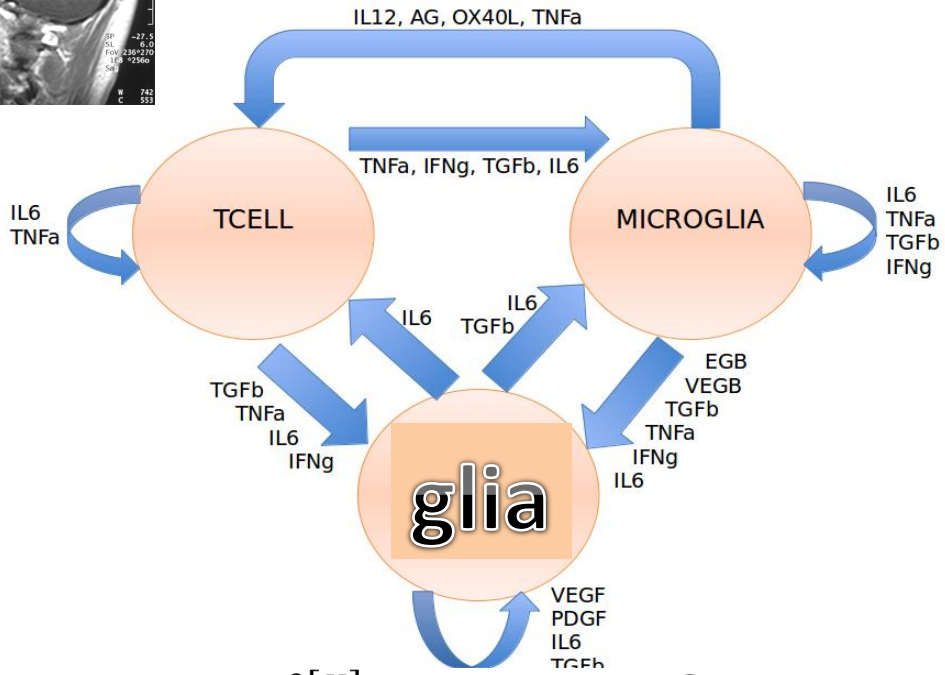
Institute for Systems Biology

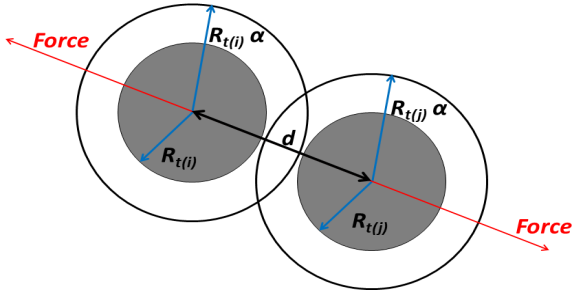
Revolutionizing Science. Enhancing Life.



Glioma

cell-cell repulsion, when two cells are to close, and cell-cell adhesion between pairs with specific intracellular states.

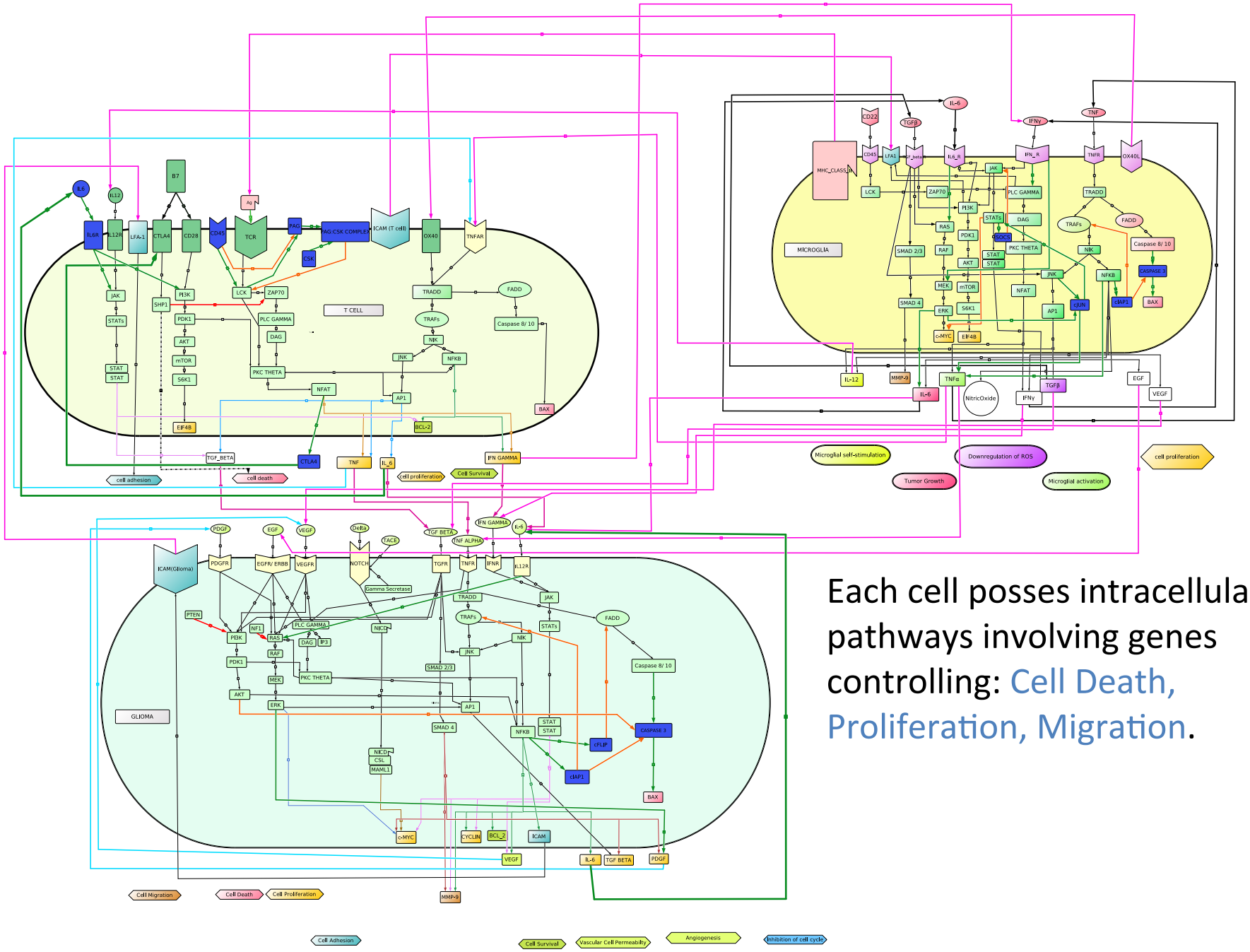




$$\frac{\partial[X]}{\partial t} = \nabla \cdot (D_X \nabla[X]) + \frac{S_X n_l}{h^3} - \gamma_X [X]$$

$$F_{ij} = \begin{cases} \frac{1}{2}(R_i + R_j - d) & , if\ d < R_i + R_j \\ \frac{1}{2}(R_i + R_j - d) e^{-\left(\frac{d}{R_i + R_j} - 1\right)^2} & , if\ d \geq R_i + R_j \end{cases}$$

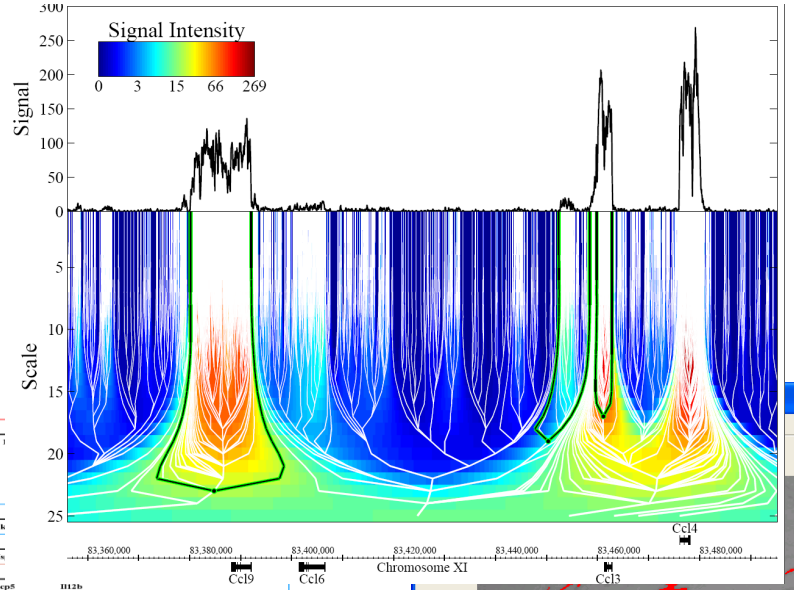
RamRup Sarkar, Boris Aguilar, Farzin Ghaffarizadeh

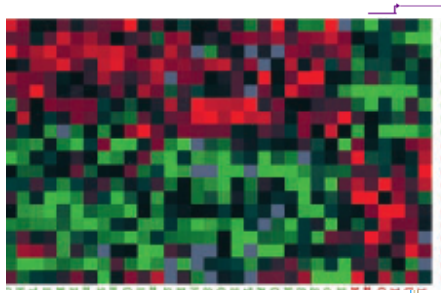


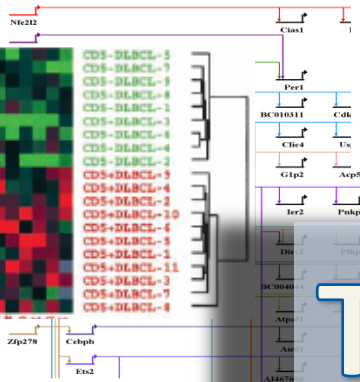
Each cell posses intracellular pathways involving genes controlling: Cell Death, Proliferation, Migration.

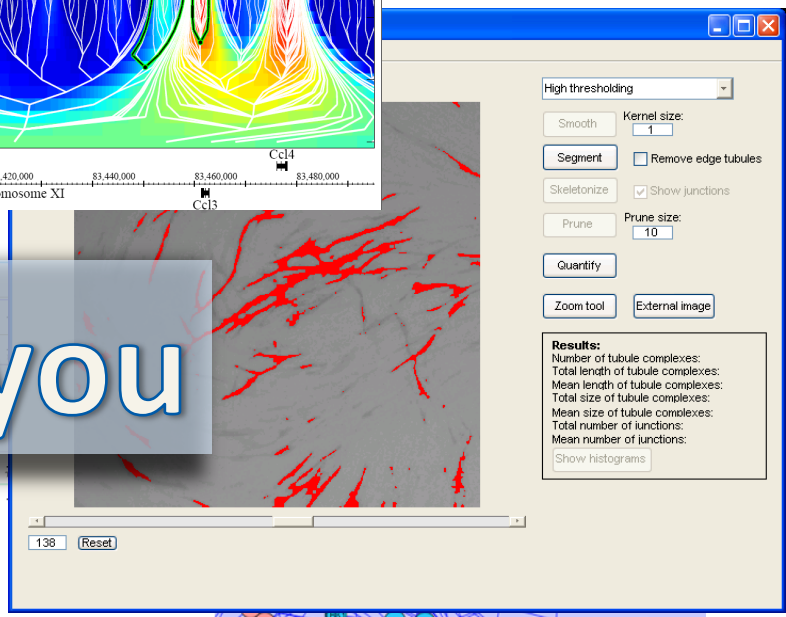
RamRup Sarkar, Boris Aguilar, Farzin Ghaffarizadeh

INSTITUTE FOR
Systems
Biology









Thank you

$$
\begin{aligned}
L &= \frac{\lim_{t\to\infty} P[X_{t-j}=x\,|X_0=h,\tau(t)=j]}{\lim_{t\to\infty} P[X_{t-j}\in B_k|X_0=h,\tau(t)=j]}\\
&= \frac{1}{\pi(B_k)}\lim_{t\to\infty} P[X_{t-j}=x|X_0=h,\tau(t)=j]\\
&= \frac{1}{\pi(B_k)}\lim_{t\to\infty}\sum_{i=1}^{m}\sum_{y\in B_i} P[X_{t-j}=x|X_{t-j-1}=y,\\
&\quad X_0=h,\tau(t)=j]\\
&\quad \times P[X_{t-j-1}=y|X_{t-j-1}\in B_i, X_0=h,\tau(t)=j]\\
&\quad \times P[X_{t-j-1}\in B_i|X_0=h,\tau(t)=j]\\
&= \frac{1}{\pi(B_k)}\sum_{i=1}^{m}\sum_{y\in B_i}\lim_{t'\to\infty} P[X_{t'}=x|X_{t'-1}=y,\\
&\quad X_0=h,\tau(t')=0]\\
&\quad \times \lim_{t'\to\infty} P[X_{t'-1}=y|X_{t'-1}\in B_i, X_0=h,\tau(t')=0]\\
&\quad \times \lim_{t'\to\infty} P[X_{t'-1}\in B_i|X_0=h,\tau(t')=0]\\
&= \frac{1}{\pi(B_k)}\sum_{i=1}^{m}\sum_{y\in B_i} P^*_y(x)\pi^*(y|B_i)\pi(B_i),
\end{aligned}
$$