Gene Regulatory Networks: Modeling and Intervention

Ritwik Kumar Layek

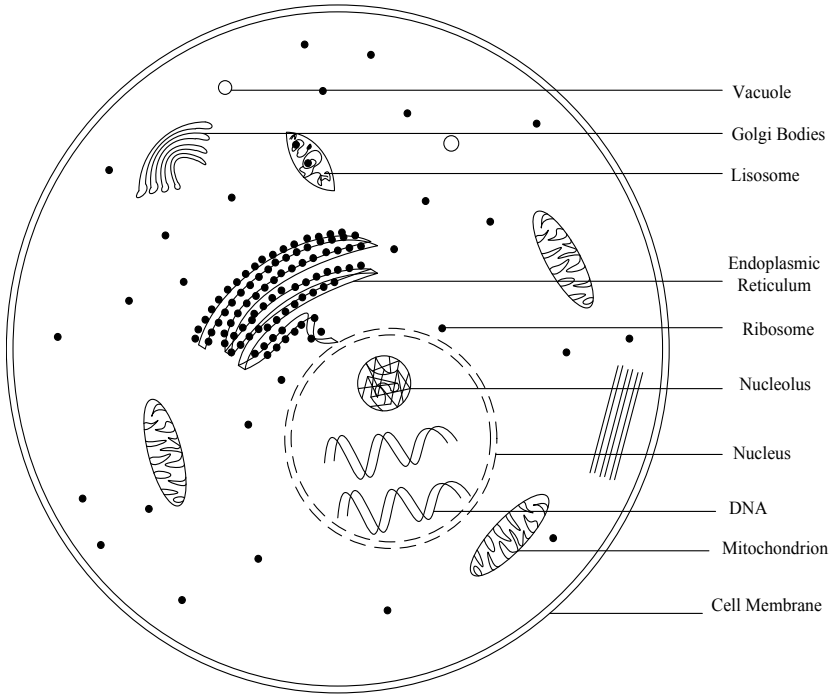
Department of Electronics and Electrical Engineering
Indian Institute of Technology Kharagpur, West Begal, India

October 30, 2013

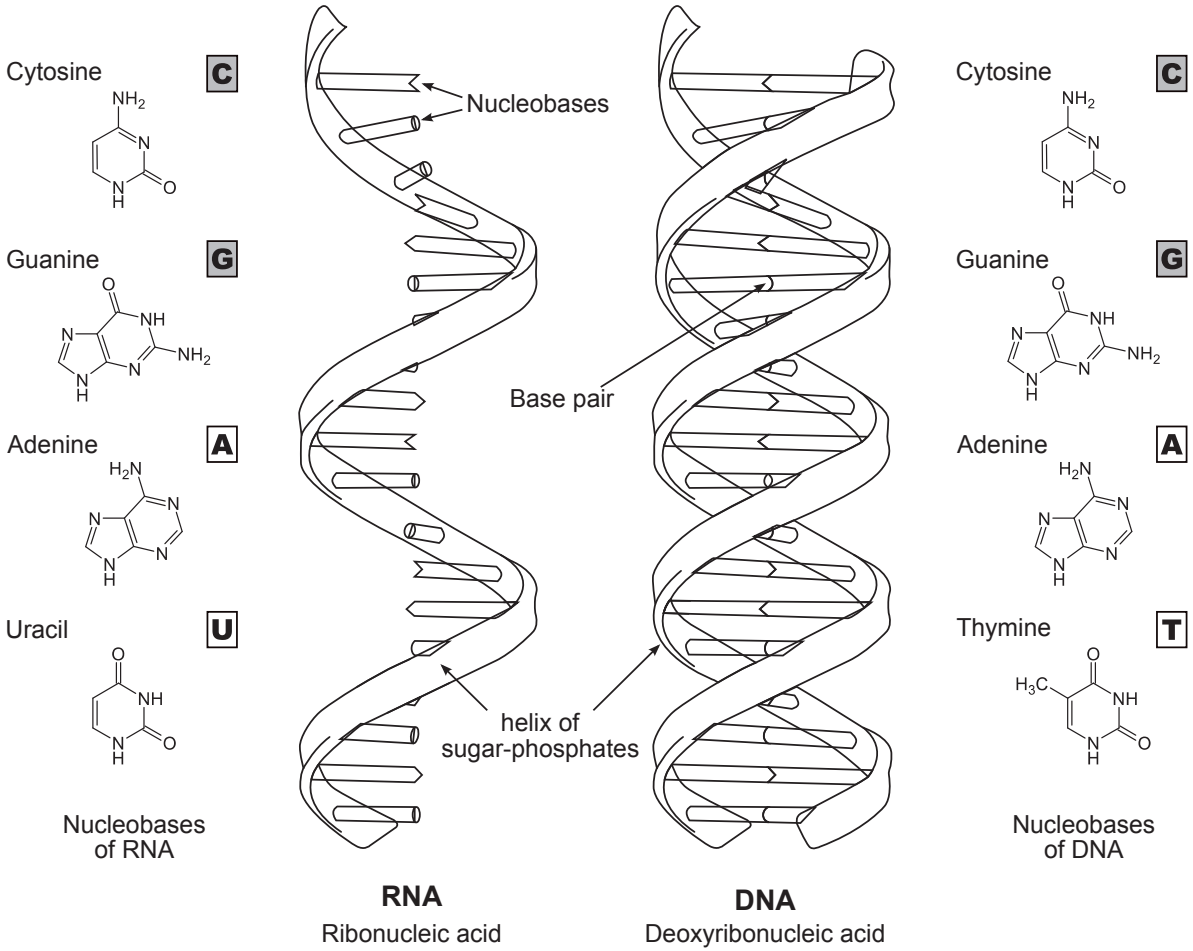
Outline

The Cell

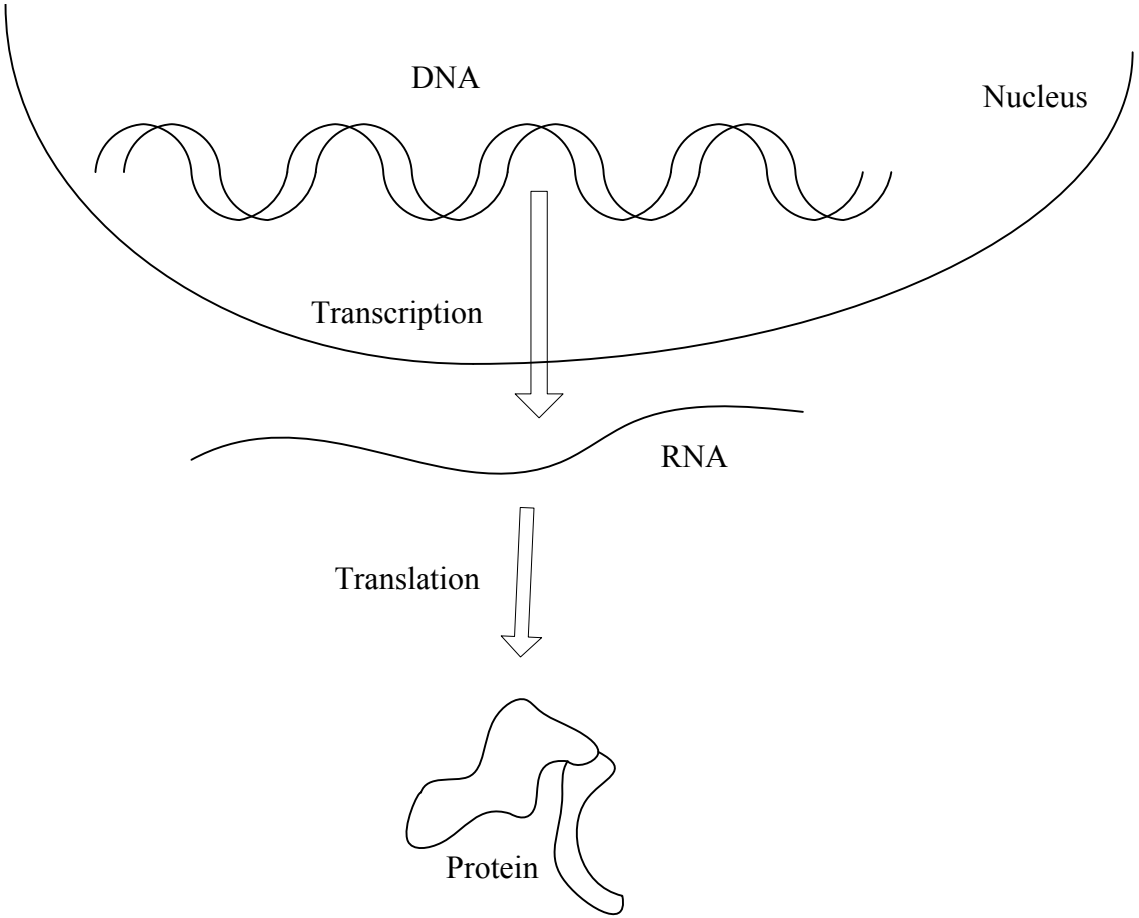
- Cell is the unit of life.
- The cell works as a massive signal processor.
- Dynamics of cell shows robust regulatory control.



DNA and RNA: The Helices of Life



Gene, Protein and Central Dogma of Molecular Biology



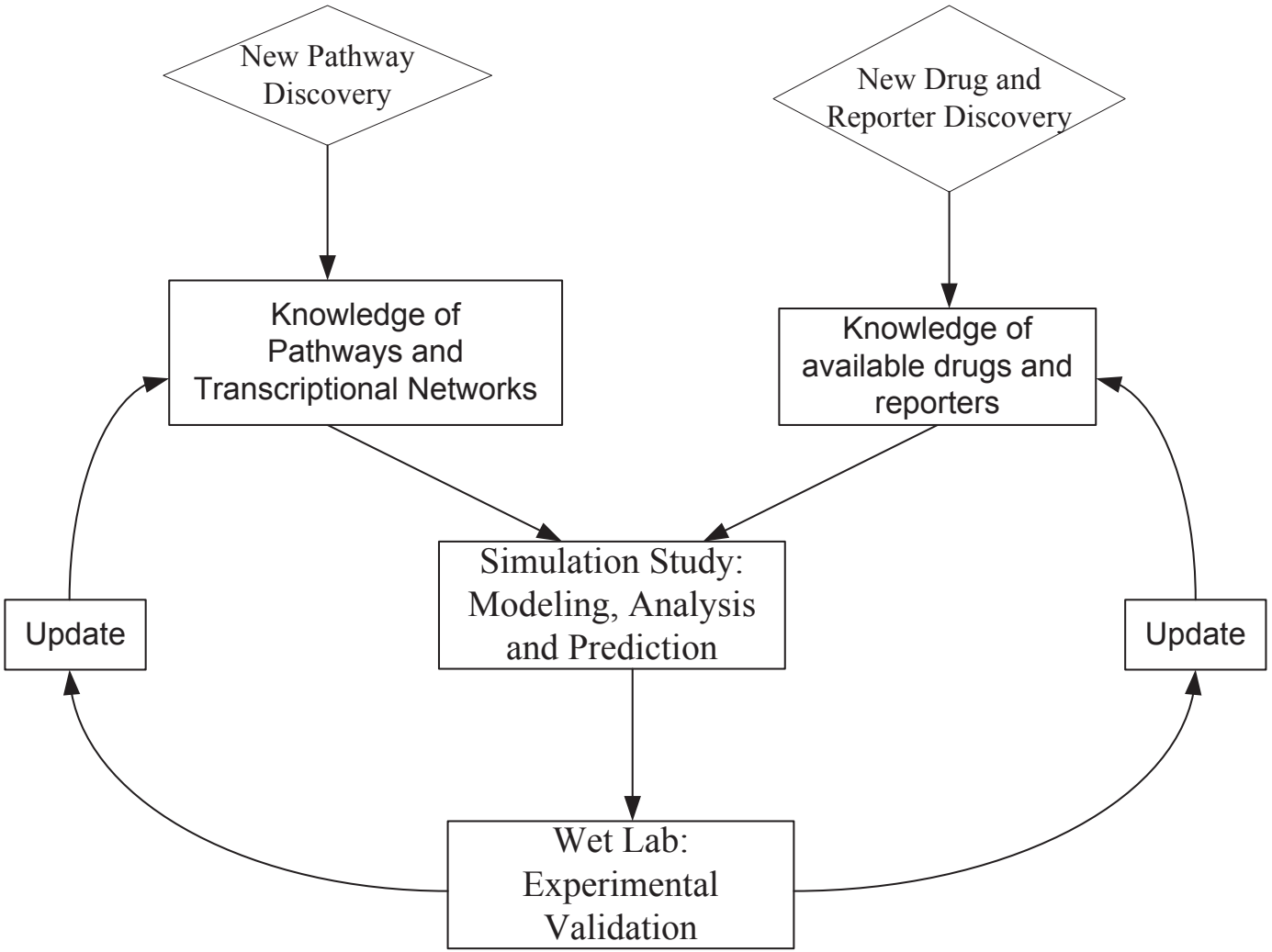
Transcription Factor, Promoter and Genetic Regulation

- Transcription Factor (TF): Small proteins which binds the genomic DNA and initiates transcription.
- Promoter: The segment of the DNA where the transcription factor binds.
- Activator: This TF enhances the transcriptional activity.
- Inhibitor: This TF represses the transcriptional activity.
- Logic gate type behavior: For multiple TFs, the combined interaction can be like logic gates.

Genetics and Epigenetics

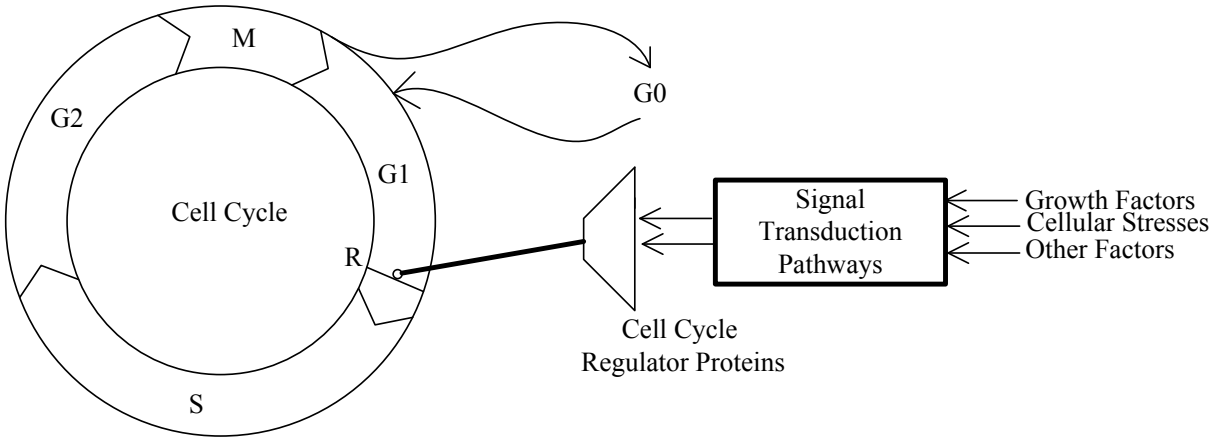
- Gene: Sequence of DNA which encodes protein.
- Genetics: Study of the gene sequences and the effects of any alterations (mutations) in those genes.
- Epigenetics: Study of genetic expressions and regulatory behavior.
- All cells in an organism have identical genes (except mutant cells of course), but epigenetic programming plays the bigger role.
- Dynamical system modeling is required to describe the epigenetic interactions.

Modeling and Control



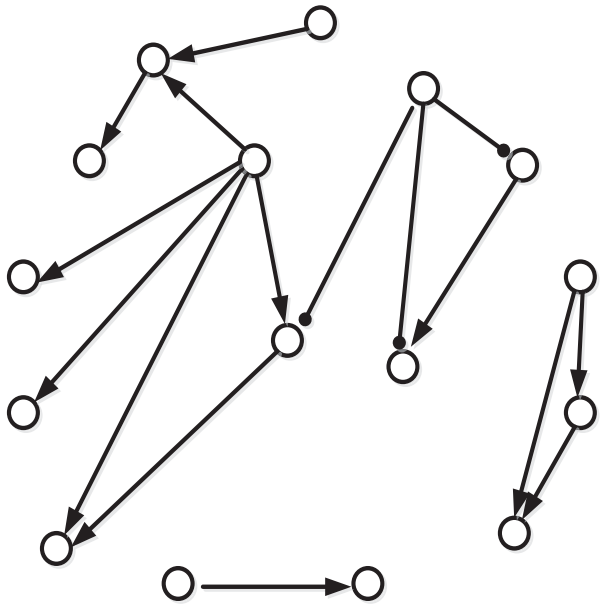
Cell Cycle: Restriction Point

- Mitosis: Growth of a multi-cellular organism is achieved by Mitosis.
- Growth factors begin Mitosis.
- Differentiated cells divide only by special requests (example, PDGF)



Transcriptional Network

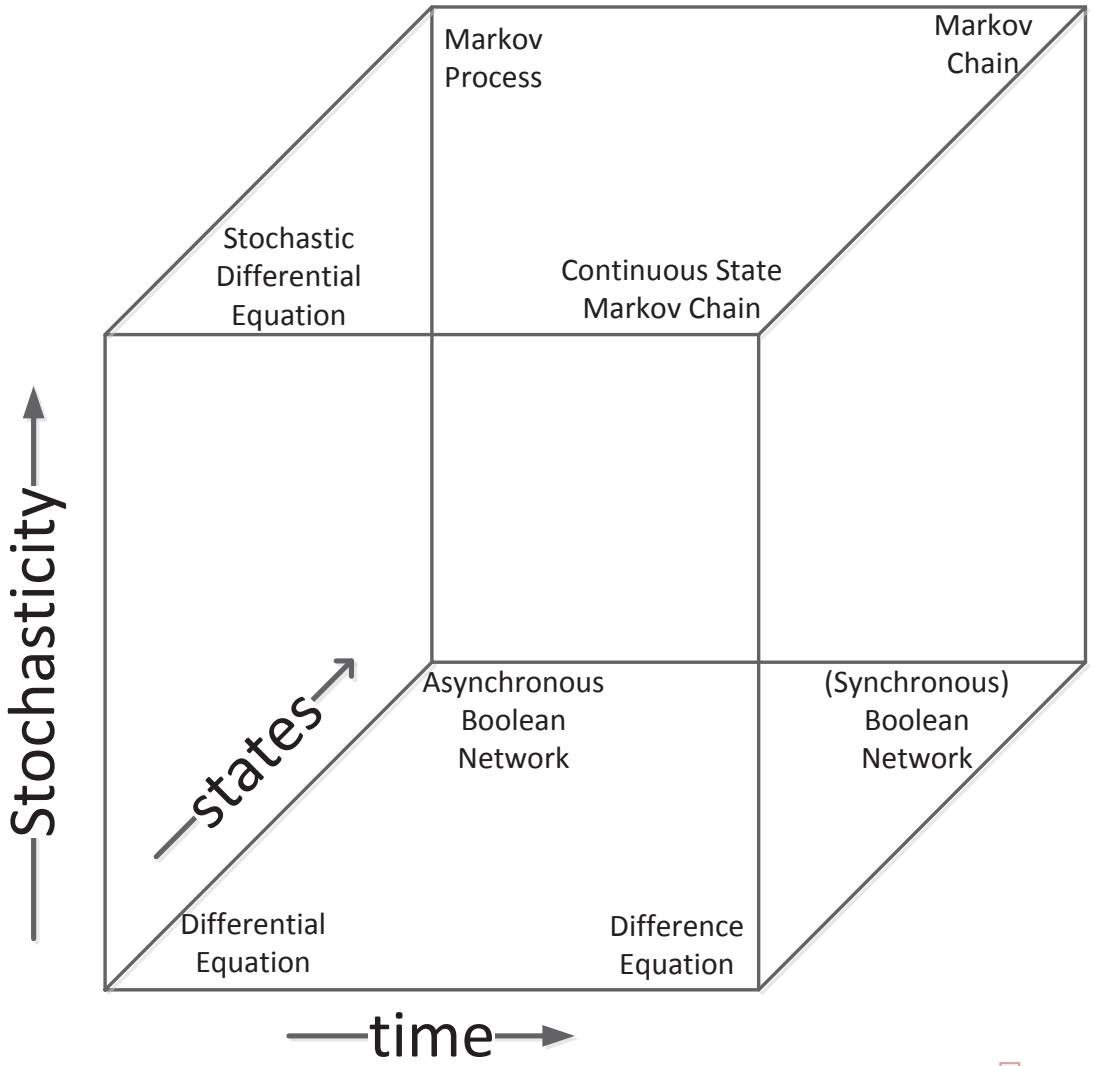
- Protein synthesis is slow.
- The transcriptional network is usually shallow in terms of the signalling depth.
- The amino acids are reused by protein synthesis-degradation cycle.



Signal Transduction Pathways

- Fast protein-protein interaction.
- The signal transduction pathways can be deep in terms of the signalling depth.
- They work in between transcriptional interactions.
- Transduction pathways can be modeled as combinatorial Boolean Network.
- Combining signal transduction with transcriptional (and post transcriptional) activity can lead to a better understanding of the system dynamics.

Dynamical System Modeling: Different Paradigms



Chemical Master Equation

- Stochastic Differential Equation with quantized state space.
- Markov Process
- CME can be derived from basic chemical kinetics and statistical physics [1].
- Consider a volume Ω with N different species of molecules $\{\psi_1, \psi_2, \psi_3, \ldots, \psi_N\}$.
- The counts of the molecules are $\{x_1, x_2, x_3, \ldots, x_N\}$. Define the state (x) as this n-tuple.

Chemical Master Equation: Parameters

- M elementary chemical reactions $\{R_1, R_2, R_3, \ldots, R_M\}$
- a_μ: Propensity function for the reaction R_μ
- $a_\mu(t)dt$ = Probability that the reaction R_μ will happen in the time interval $(t, t + dt)$
- $R_\mu \Longrightarrow (x \longleftarrow x + s_\mu)$
- Stoichiometric Matrix: $S = [s_1, s_2, \ldots, s_M]$

Chemical Master Equation: Problem Formulation

- Suppose current state $x(t) = x$
- Given x, $a_\mu(x)$ is computed for all R_μ
- $R_\mu : \psi_i \xrightarrow{c_\mu} \phi$, $a_\mu(x) = c_\mu x_i$
- $R_\mu : \psi_i + \psi_j \xrightarrow{c_\mu} \phi$, $a_\mu(x) = c_\mu x_i x_j$
- $R_\mu : \psi_i + \psi_i \xrightarrow{c_\mu} \phi$, $a_\mu(x) = c_\mu x_i(x_i - 1)/2$

Chemical Master Equation: Problem Formulation

- $p(x,t)$: Probability that the chemical system will be in state x at time t
- CME: time evolution of $p(x,t)$

$$\begin{aligned} p(x,t+dt) &= p(x,t)[1-\sum_{\mu=1}^{M} a_\mu(x)dt] \\ &+ \sum_{\mu=1}^{M} p(x-s_\mu,t)a_\mu(x-s_\mu)dt \end{aligned}$$

$$\begin{aligned} \frac{p(x,t+dt)-p(x,t)}{dt} &= -p(x,t)\sum_{\mu=1}^{M} a_\mu(x) \\ &+ \sum_{\mu=1}^{M} p(x-s_\mu,t)a_\mu(x-s_\mu) \end{aligned}$$

Chemical Master Equation: Example

- $A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$
- $x = [X_A, X_B, X_C]$
- $s_1 = [-1, -1, 1]$, $s_2 = [1, 1, -1]$
- $a_1 = k_1 X_A X_B$, $a_2 = k_2 X_C$
- Initial condition: $x_0 = [4, 3, 0]$
- State space: $\chi = \{(4, 3, 0), (3, 2, 1), (2, 1, 2), (1, 0, 3)\}$
- Probability space:
  $\wp = \{p(4, 3, 0), p(3, 2, 1), p(2, 1, 2), p(1, 0, 3)\}$

Chemical Master Equation: Example

- $\frac{dp(4,3,0)}{dt} = k_2 p(3,2,1) - 12k_1 p(4,3,0)$
- $\frac{dp(3,2,1)}{dt} = 12k_1 p(4,3,0) + 2k_2 p(2,1,2) - (6k_1 + k_2)p(3,2,1)$
- $\frac{dp(2,1,2)}{dt} = 6k_1 p(3,2,1) + 3k_2 p(1,0,3) - (2k_1 + 2k_2)p(2,1,2)$
- $\frac{dp(1,0,3)}{dt} = 2k_1 p(2,1,2) - 3k_2 p(1,0,3)$
- These four equations will govern the stochastic dynamics of the system.

Stochastic Simulation Algorithm

- Objective: Dynamics $x(t)$ not of $p(x, t)$
- Define $p(\tau, \mu|x, t)$: Probability (given $x(t) = x$) that the next reaction will occur in the time frame $[t + \tau, t + \tau + d\tau]$ and the reaction will be R_μ
- Time gap to the next reaction: τ
- Index of the next reaction: μ

SSA

- $P_0(\tau|x,t)$: Probability that given $x(t) = x$, no reaction occurs in the interval $[t, t+\tau)$
- $p(\tau, \mu|x,t)d\tau = P_0(\tau|x,t)a_\mu(x)d\tau$
- $P_0(\tau + d\tau|x,t) = P_0(\tau|x,t)[1 - \sum_{\mu=1}^{M} a_\mu(x)d\tau]$
- $\frac{dP_0(\tau|x,t)}{d\tau} = -P_0(\tau|x,t)\sum_{\mu=1}^{M} a_\mu(x) = -P_0(\tau|x,t)a_0(x)$
- Where, $a_0(x) = \sum_{\mu=1}^{M} a_\mu(x)$
- $p(\tau|x,t) = e^{-a_0(x)\tau}$
- $p(\tau, \mu|x,t) = a_\mu(x)e^{-a_0(x)\tau}$
- Initial condition: $P_0(0|x,t) = 1$
- $p(\tau, \mu|x,t) = [a_0(x)e^{-a_0(x)\tau}][\frac{a_\mu(x)}{a_0(x)}] = f(\tau)g(\mu)$

Monte Carlo Inversion Method

- $f(\tau)$: Probability that next reaction will happen between time interval $[t+\tau]$ and $[t+\tau+d\tau]$.
- $g(\mu)$: Probability that given there is a reaction between time interval $[t+\tau]$ and $[t+\tau+d\tau]$, the reaction will be R_μ
- Notice that τ and μ are conditionally independent given (x,t).
- Draw two uniform random variables r_1 and r_2 in the interval $[0,1]$.
- $\tau = \frac{1}{a_0(x)} ln(\frac{1}{r_1})$
- $\sum_{j=1}^{\mu-1} a_j(x) \leq r_2 a_0(x) < \sum_{j=1}^{\mu} a_j(x)$

Inversion Theory

- $x \sim p(x) \Rightarrow p(x')dx' =$ probability that $x' \leq x \leq x' + dx'$
- $F(x) = \int_{-\infty}^{x} p(x')dx'$
- $F(x_0)$: Probability that $x \leq x_0$
- Draw a random number $r \sim U(0,1)$. Find x which satisfies $F(x_0) = r$ or $x_0 = F^{-1}(r)$. We need to prove that $P(x_0 = x) = p(x)dx$
- Proof: $P(x \leq x_0 \leq x + dx) = P\{F(x) \leq r \leq F(x + dx)\} = F(x + dx) - F(x)$.

The Algorithm

1. Initialize at $t = t_0$ and $x = x_0$
2. Assuming $x(t) = x$, evaluate $a_\mu(x)$, $a_0(x)$
3. Generate τ, μ according to the Monte Carlo Inversion Method
4. $t \leftarrow t + \tau$, $x \leftarrow x + s_\mu$
5. Return to step 2.

Markov Process and Markov Chain

- The CME is actually a Markov Process with huge state space.
- It is possible to reduce the state space by thresholding the protein concentrations.
- Many reactions follow Hill function kind of characteristics.
- Sampling the time axis reduces the Markov Process to Markov Chain. (Nyquist rate in Biology!)

Markov Chain Properties

- Memoryless: $P(x_{n+1} = x | x_n = y, x_{n-1} = z, \ldots, x_0 = w) = P(x_{n+1} = x | x_n = y)$
- Stationarity: $P(x_{n+1} = x | x_n = y) = P(x_n = x | x_{n-1} = y)$
- n-step transition: $p_{ij}^{(n)} = P(x_n = j | x_0 = i)$
- Chapman Kolmogorov Equation: $: p_{ij}^{(n+m)} = \sum_{k=1}^{M} p_{ik}^{(n)} p_{kj}^{(m)}$

Markov Chain Properties..

- Marginal distribution: $P(x_n = x) = \sum_{k=1}^{M} p_{ij}^{(n)} P(x_0 = i)$
- Accessibility: $P(x_n = j | x_0 = i) = p_{ij}^{(n)} > 0$ for some $n \geq 0$. We say $i \rightarrow j$
- Communicate: If $i \rightarrow j$ and $j \rightarrow i \Rightarrow i \leftrightarrow j$
- Irreducible: single communicating class.
- Periodicity (period k): $k = gcd\{n : P(x_n = i | x_0 = i) > 0\}$
- Aperiodic: period $= 1$

Markov Chain Properties..

- Hitting time: $T_i = inf\{n \geq 1 : x_n = i | x_0 = i\}$
- Transient: $P(T_i < \infty) < 1$ or $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$
- Recurrent: $P(T_i < \infty) = 1$ or $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$
- Absorbing state: If $p_{ii} = 1$ and $p_{ij} = 0 \ \forall \ i \neq j$
- Ergodic: A state is ergodic if it is positive recurrent and aperiodic. If all states of an irreducible MC is ergodic then the MC is ergodic.
- Steady state distribution: $\pi_j = \sum_{i=i}^{M} \pi_i p_{ij}$

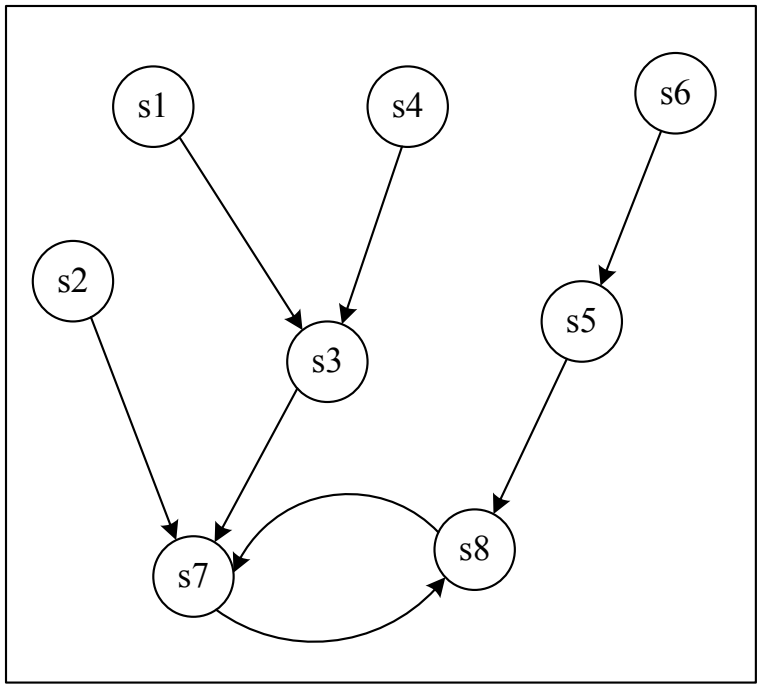
Boolean Network

- A Boolean Network(BN) $B(V, F)$ is defined by a set of nodes $V = \{x_1, x_2, \ldots, x_n\}$ and a set of predictor functions $F = \{f_1, f_2, \ldots, f_n\}$.
- A Boolean function $f_i(V)$ determines the value of x_i in the next time instance.
- The Boolean rules generates the dynamics of the system.
- In BN, every state transits to an unique state.
- Some states are transient, whereas others are called the attractors.

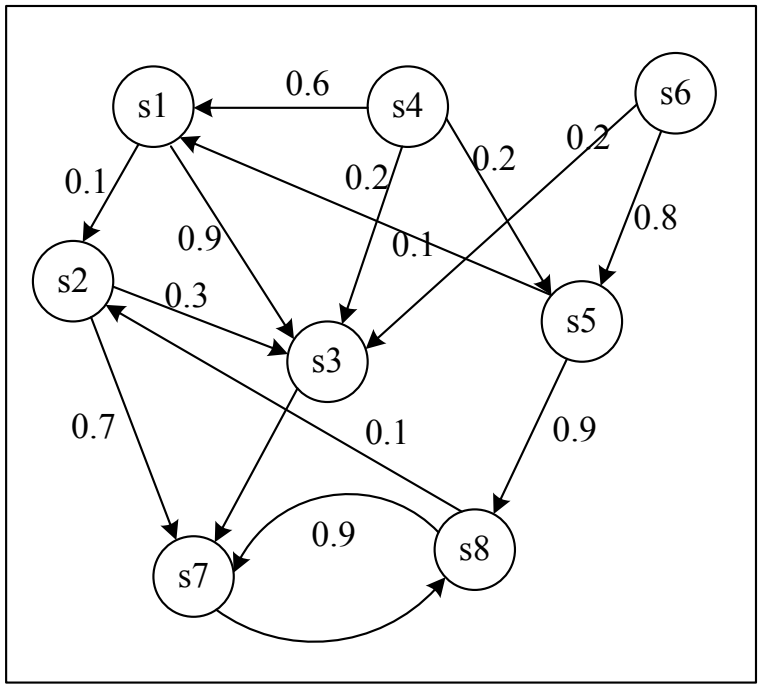
Probabilistic Boolean Network

- A Probabilistic Boolean Network(PBN) $B(V, F)$ is defined by a set of nodes $V = \{x_1, x_2, \ldots, x_n\}$ and a set of predictor functions $F = \{F_1, F_2, \ldots, F_n\}$.
- A set of Boolean functions $F_i(V) = f_i^j(V) \forall j = 1, 2, \ldots, l_i$ determines the value of x_i in the next time instance.
- The Boolean function $f_i^j(V)$ is chosen with selection probability P_i^j such that $\sum_{j=1}^{l_i} P_i^j = 1$
- In PBN, every state transits to different states according to a definite transition probability.

State Transition Diagram



(a) Boolean Network (BN)



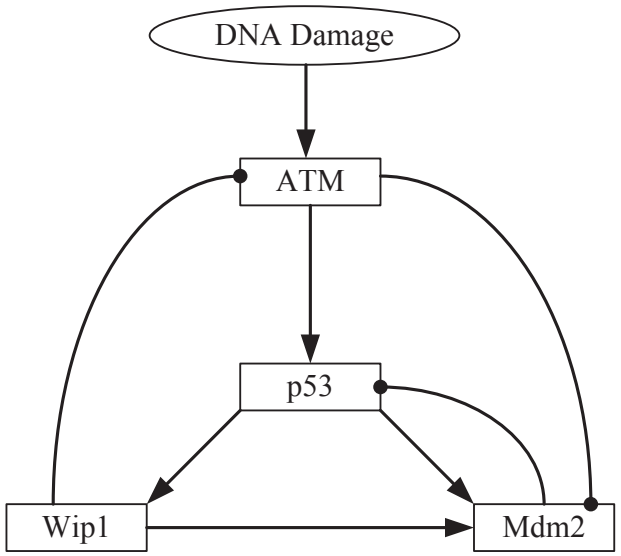
(b) Probabilistic Boolean Network (PBN)

Markov Chain and PBN

- It is evident that PBN is the structural representation of the Markov Chain.
- Inferring the constituent BNs, it is possible to know the 'Transition Probability Matrix' of the Markov Chain.
- Markov Chain is the missing link between the high level modeling of pathways and the low level modeling of molecular biochemistry.
- Markov Chain is also the most suitable model to design finite horizon optimal control in a biological framework.

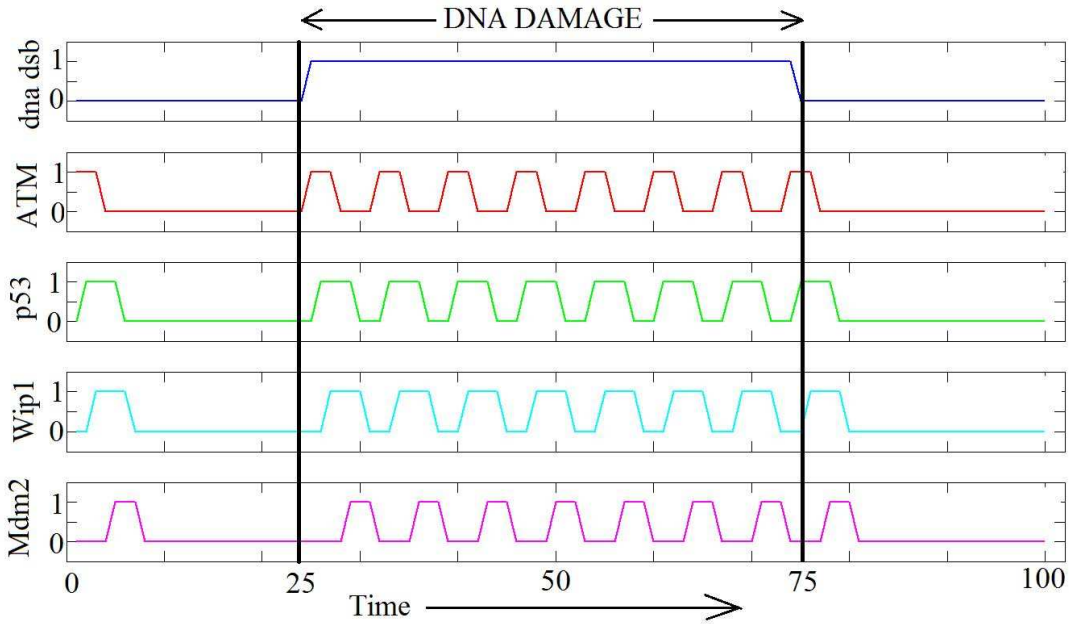
From Pathways to Networks[2]

- If the simplest model can capture the time dynamics try it first.
- p53 is an important transcription factor controlling the cell cycle and apoptosis.
- 50% of human cancers have p53 defect.

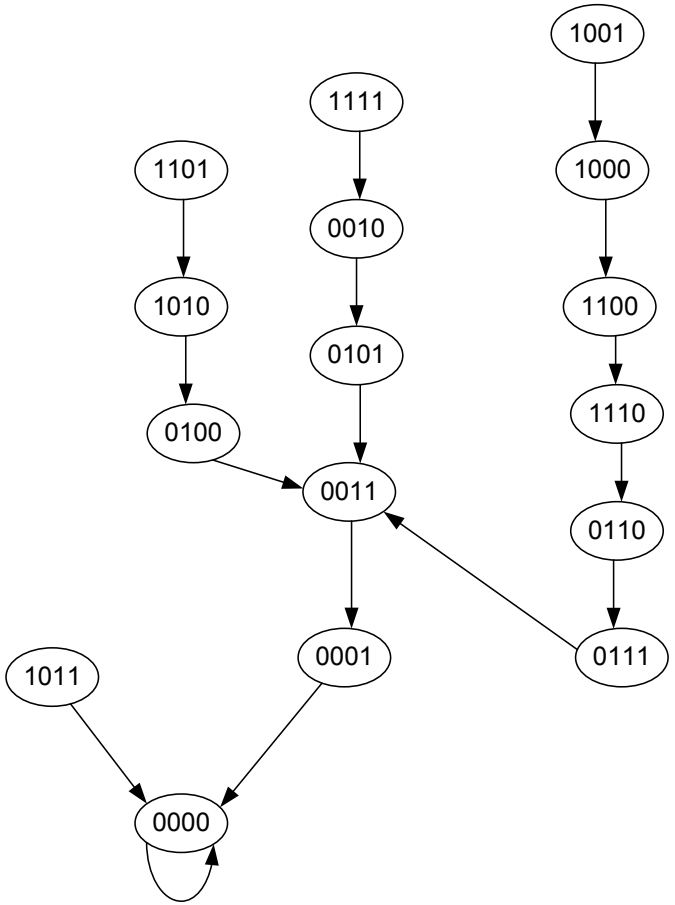


P53 Regulatory Network

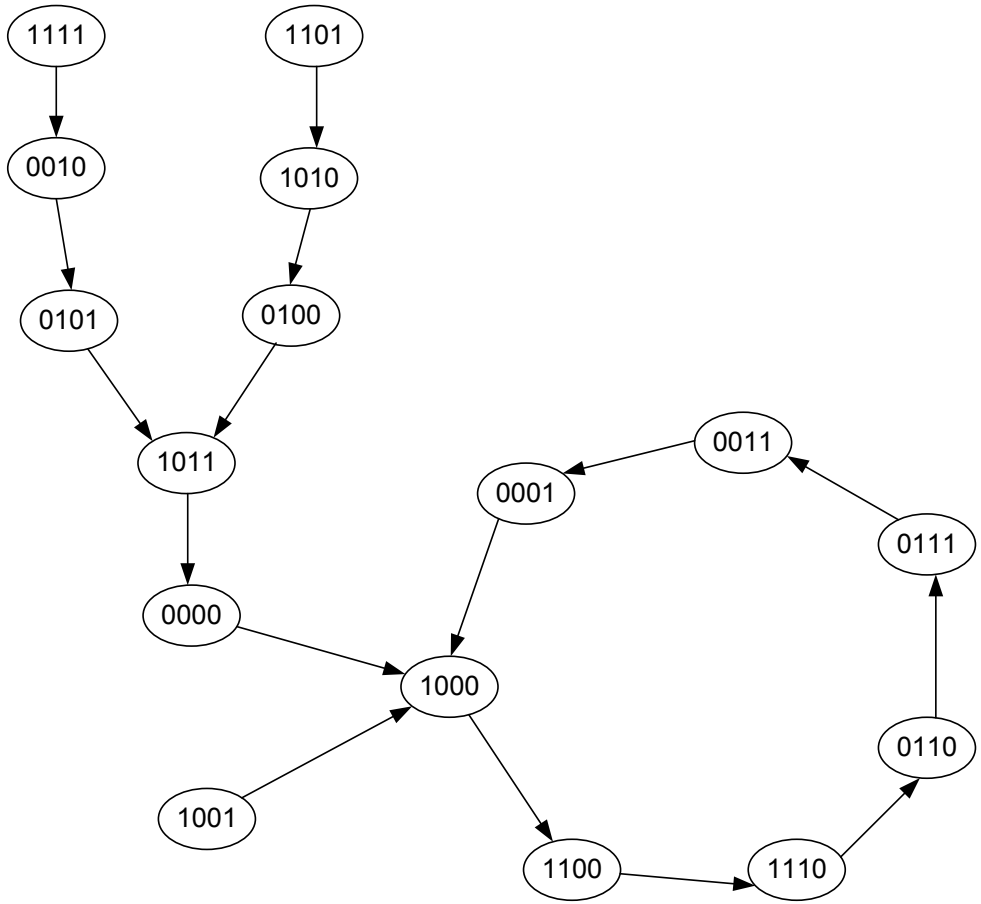
$$
\begin{aligned}
ATM_{next} &= \overline{Wip1}(ATM + dna\_dsb) \\
p53_{next} &= \overline{Mdm2}(ATM + Wip1) \\
Wip1_{next} &= p53 \\
Mdm2_{next} &= \overline{ATM}(p53 + Wip1).
\end{aligned} \tag{1}
$$



Steady state Behavior: No DNA damage

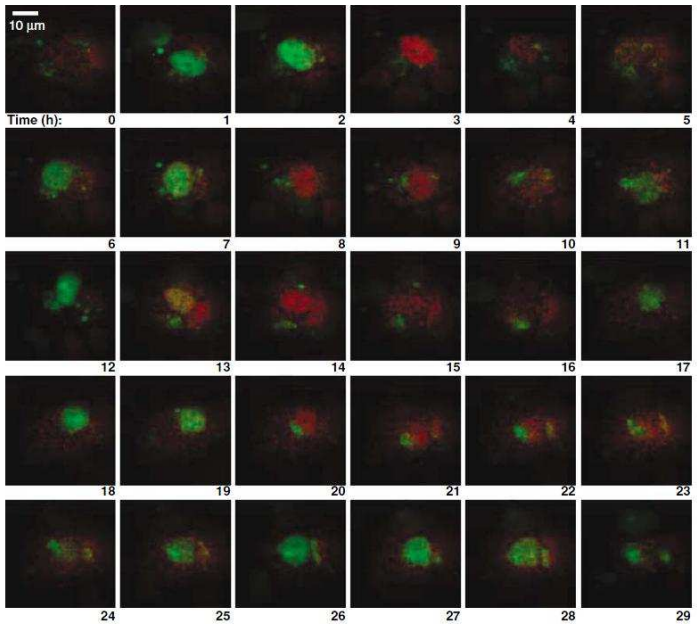


Steady state Behavior: DNA damage



Indirect validation of P53-Mdm2 oscillation

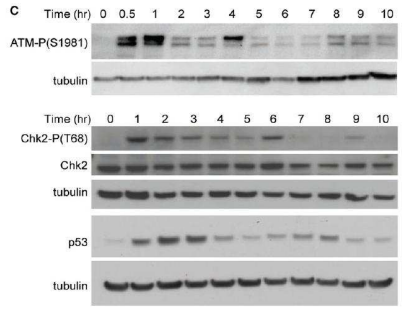
- Prolonged (29 hrs) oscillations in the nuclear levels of fluorescently tagged p53-CFP (mapped to green) and Mdm2-YFP (mapped to red) in individual MCF7, U280, cells following 5 Gy gamma irradiation .



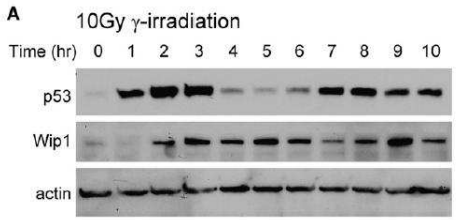
©Naama Geva-Zatorsky et al.

Indirect validation of P53-ATM-Wip1 oscillation

- Immunoblots of ATM-P(S1981), Chk2-P(T68), and p53 kinetics in MCF7 cells irradiated with 10Gy of g-irradiation.



- Immunoblots of p53 and Wip1 kinetics in MCF7 cells irradiated with 10Gy of g-irradiation.



©Eric Batchelor et al. [3, 4]

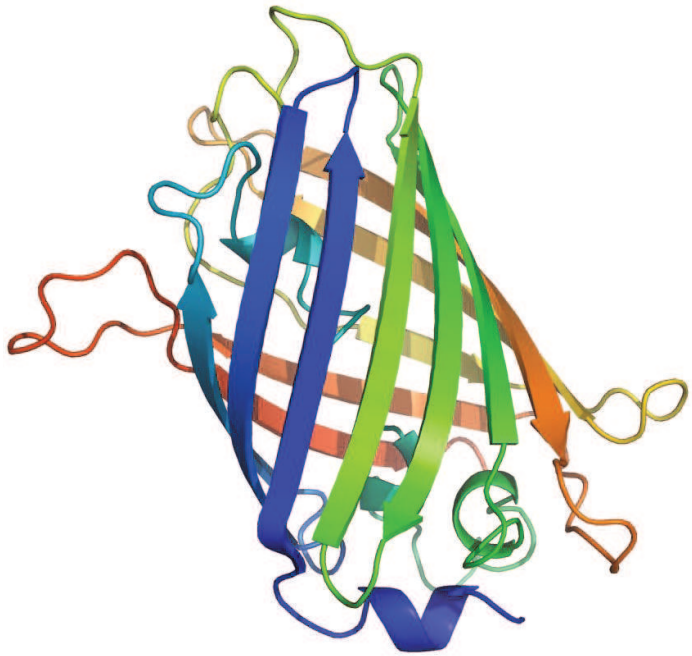
New Discovery!

- To find the missing piece of the puzzle

Green Fluorescent Protein

- New Microscope of the 21^{st} century biology!

Aequorea victoria



Mutation: God doesn't play dice?

- Any alteration in the gene sequence during DNA synthesis is called mutation.
- Mutation implies changed behavior of the gene product (protein).
- Mutation can lead to constitutive activation, constitutive deactivation or unchanged behavior of the protein in a particular pathway.
- Changed behavior due to mutation can be modeled as stuck-at-1, stuck-at-0 faults.
- Mutation is the reason for both evolution and cancer.

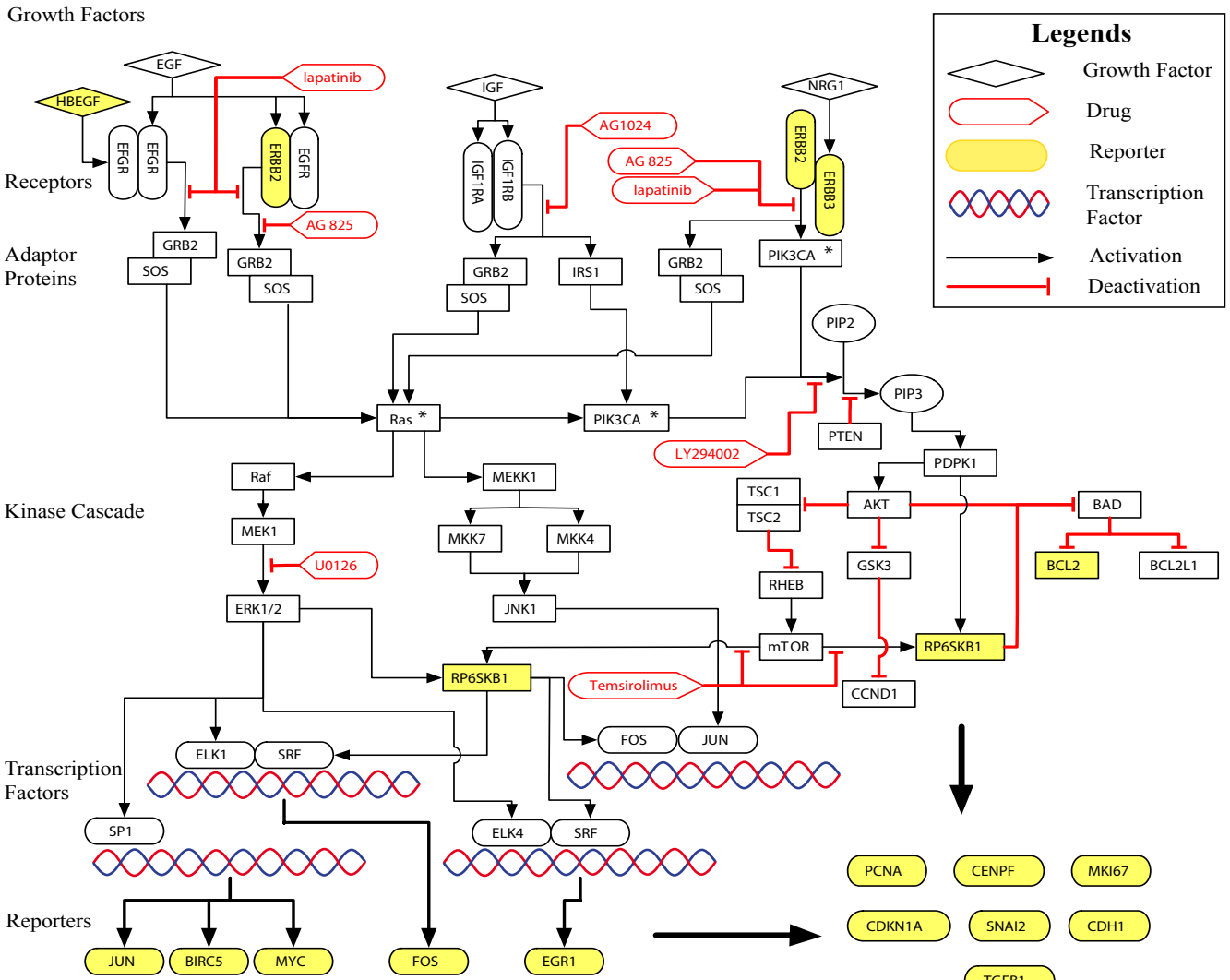
Cancer: Proliferation due to breakdown of the cell cycle

- Proliferation : Divide and conquer
- Apoptosis: Suicide is a crime
- Glycolysis: Save energy
- De-differentiation: Time machine
- Angiogenesis: Give me blood
- Invasion: Attack
- Increased motility: Move fast
- Decreased senescence: Forever young
- Metastasis: Immigrants
- Immortality: HeLa cell

Winning against Cancer: Modeling is the key

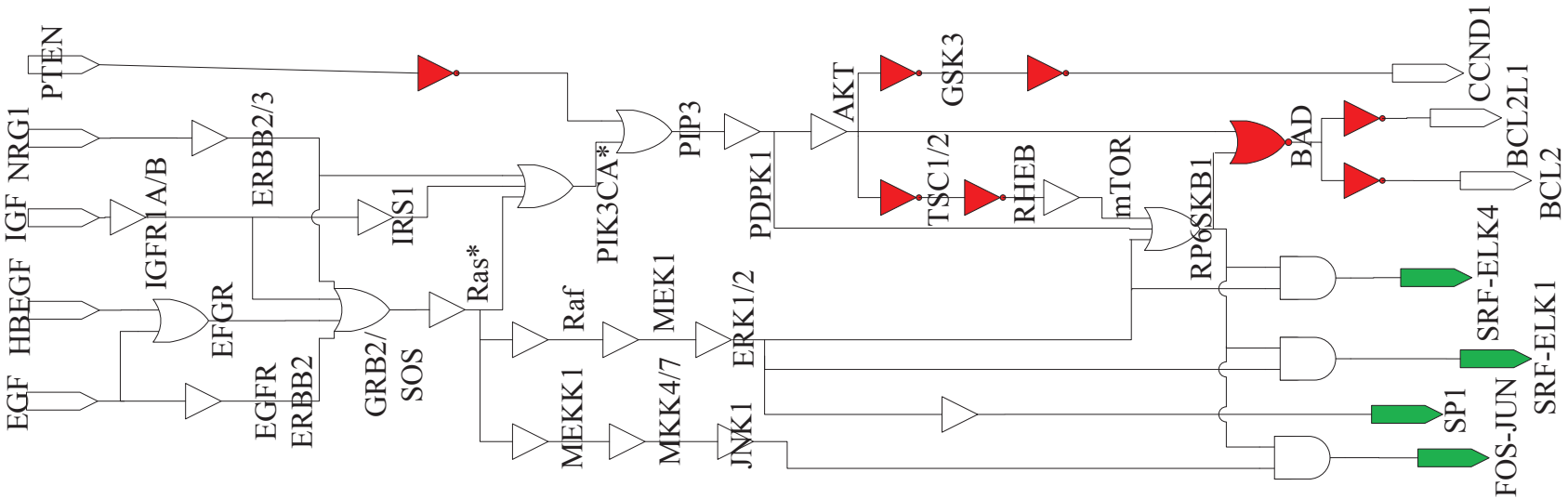
- Chemotherapy: Mass destruction
- Targetted Therapy: Model the mutations and design drugs the revert the steady state response.
- Modeling
- Fault Detection
- Reporter Design
- Drug Design

Growth Factor Mediated Pathways



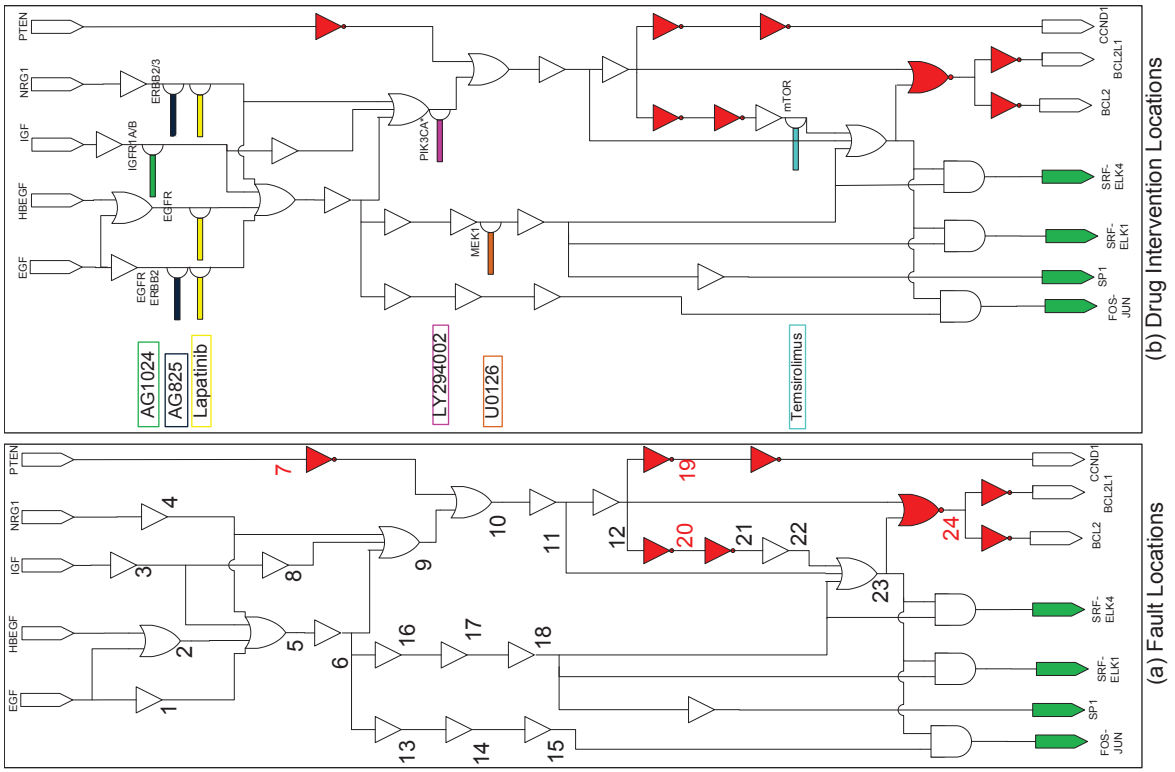
Growth Factor Mediated Pathways: Boolean Network

- The growth factor mediated signal transduction pathway is modeled as a combinatorial Boolean Network.
- These pathways don't show critical feedback loop.



Fault and Drug Design

- All the possible faults which can lead to proliferation are enumerated.
- All the available drugs are modeled as inhibitory control.



Single Fault Simulation

- Single fault simulation
- Equivalent fault classes

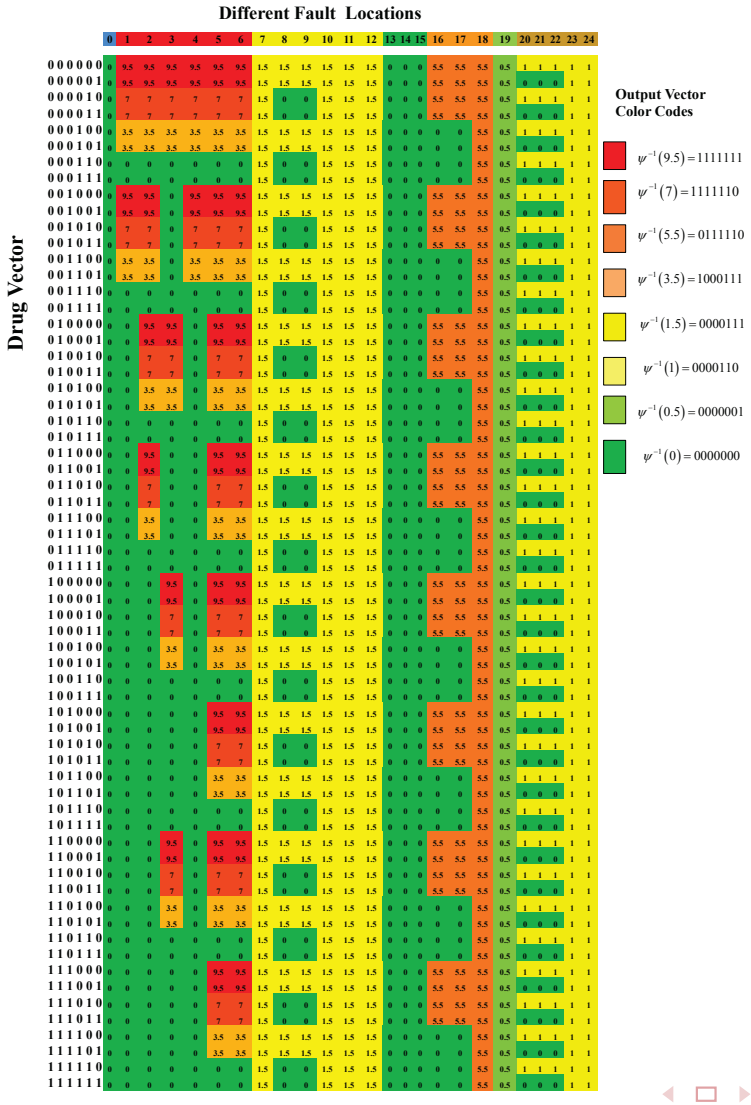
Outputs	No Fault	Fault Locations																						
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Fos-Jun	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
SRF-ELK1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
ERF-ELK4	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
BCL2	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	1	1	1
BCL2L1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	1	1	1
CCND1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0

(a) Single Fault Simulation

Output	Equivalent Fault Groups
1111111	1,2,3,4,5,6
0000111	7,8,9,10,11,12
0000000	0(No Fault),13,14,15
0111110	16,17,18
0000001	19
0000110	20,21,22,23,24

(b) Equivalent Faults for Input = 00001

Optimal Therapeutic Intervention



Importance of the Theoretical Result [5]

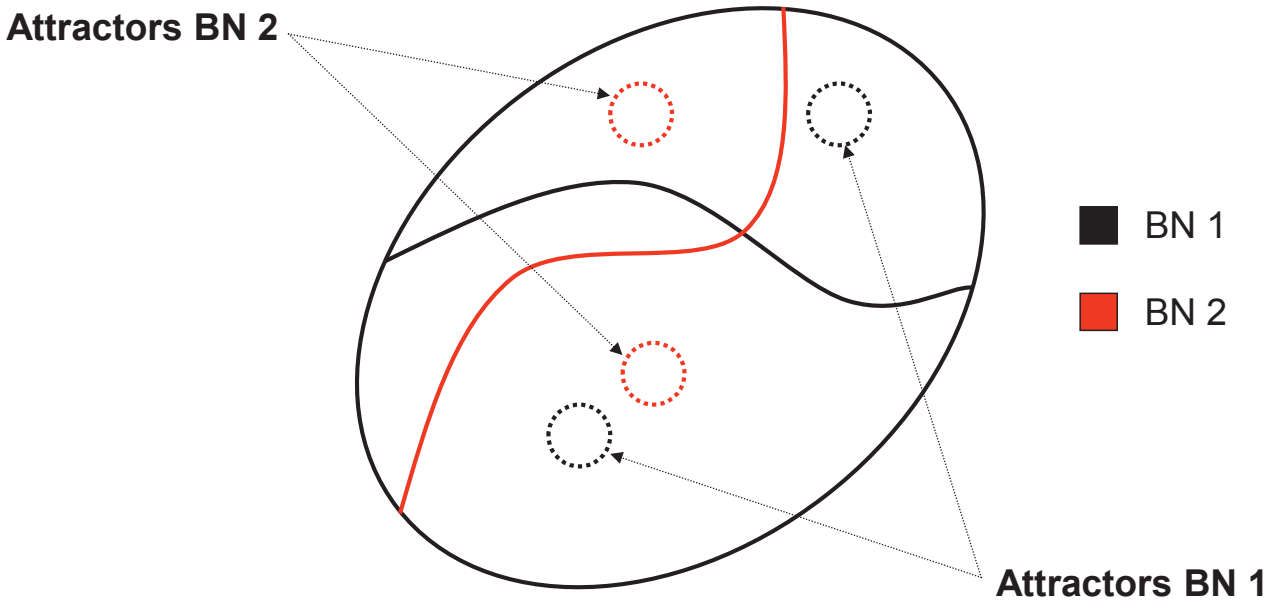
- The generalized fault detection approach can give us clue about reporter design.
- The drug simulation can eliminate the need of laboratory trials for many drug combinations.

Immediate extension

- Multiple fault modeling.
- Fault in feedback circuit.
- Intervention design to rectify the effects of these faults.
- Boolean Satisfiability problem: NP complete

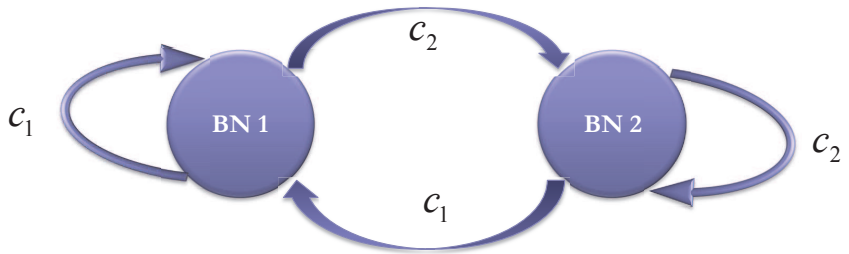
Basins of Attraction

- Control is required to change the basin of attraction



PBN revisited

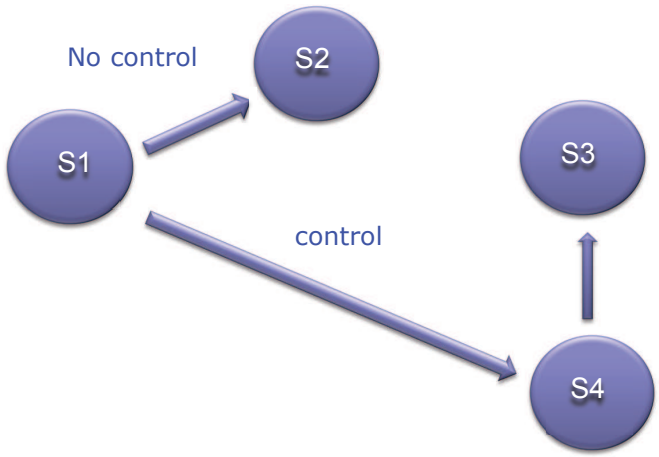
- The most appropriate model for regulatory network is Markov Chain/Process.
- PBN is the most convenient model having both mathematical rigor and structural elegance.
- Transition Probability Matrix of an 'Instantaneously Random PBN' can be computed by weighted sum of the constituent BN's TPM.



$$c_1 + c_2 = 1$$

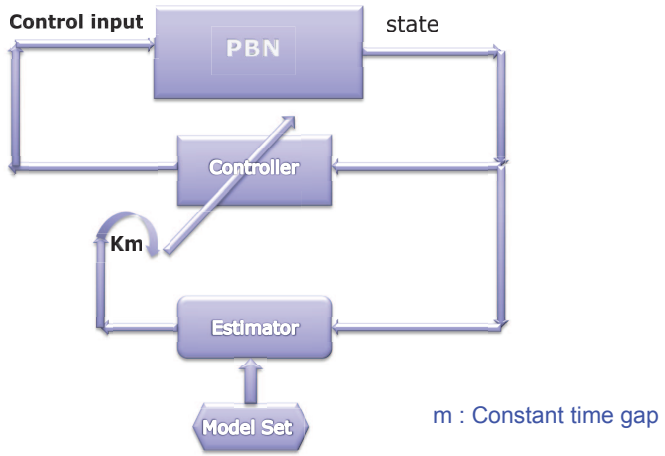
control in Markov Chain

- In PBN environment control (drug) has to be applied periodically in a finite/infinite horizon window.
- Inhibitory control forcefully transits certain states.
- Essentially the TPM is changed during the therapeutic intervention.
- The objective is to move the state from undesirable basin of attraction to the desirable basin of attraction.



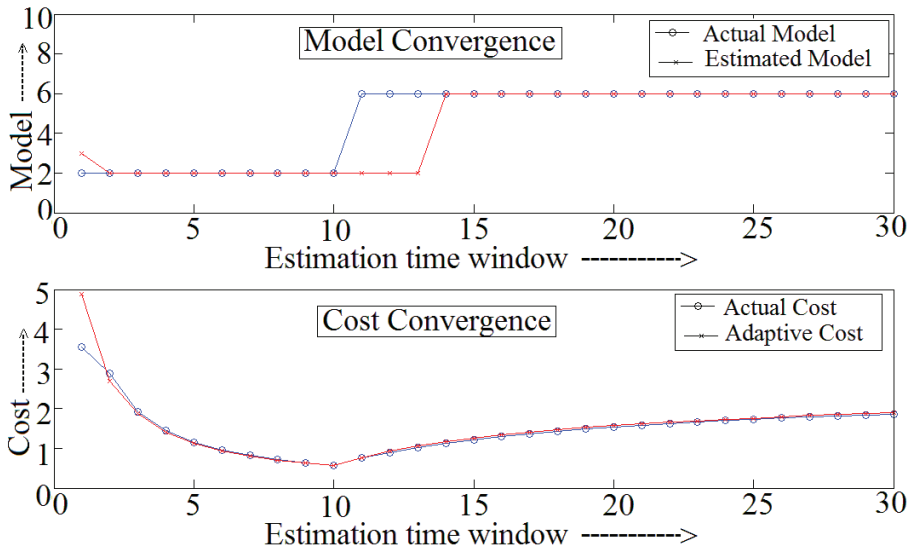
Adaptive Controller Design

- Assumption: All the constituent model BNs are known.
- The estimator will sense the context and deploy the control accordingly.

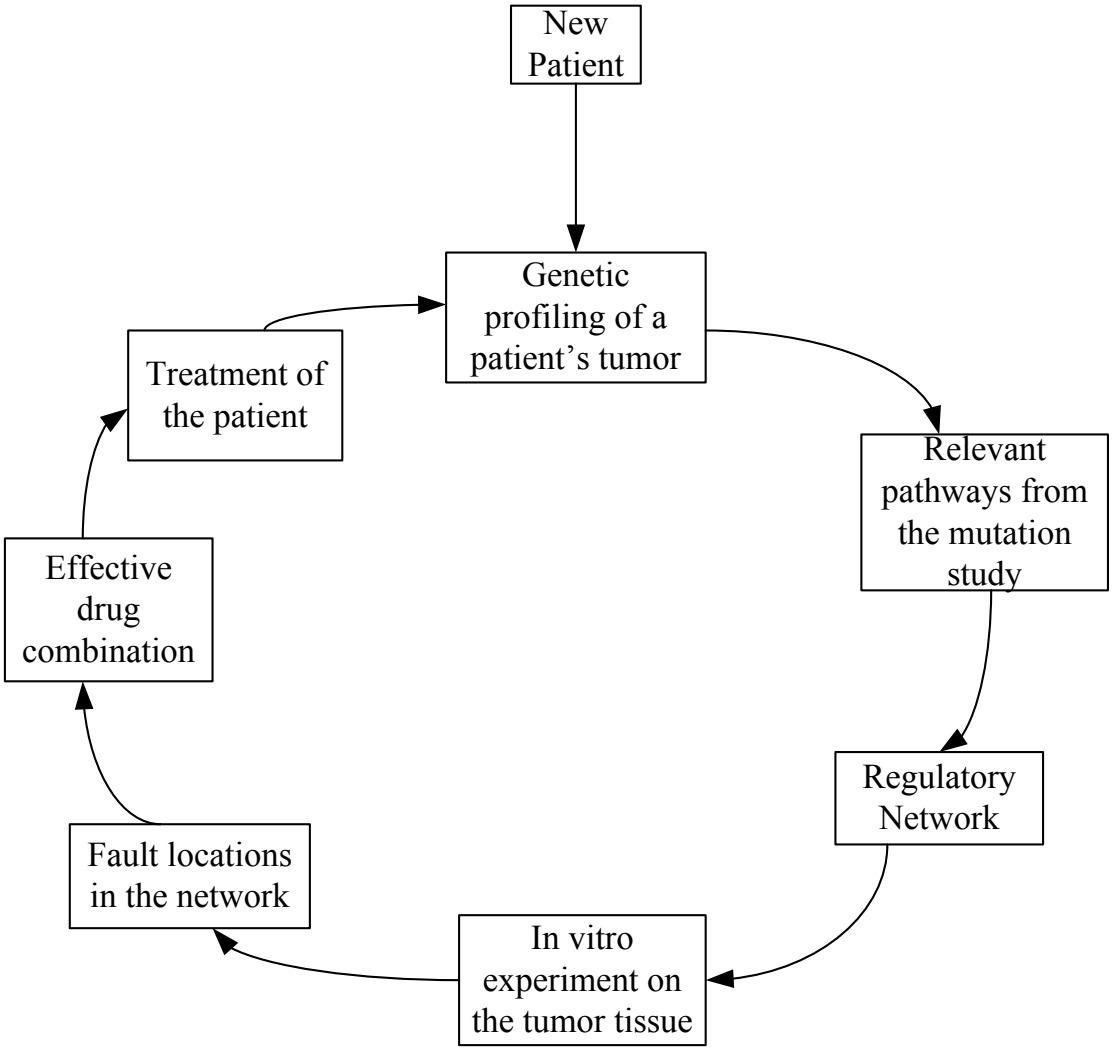


Simulation [7]

- Model tracking and cost tracking
- Almost sure convergence [6]



Personalized Medicine



Future Research

- Multiple fault detection in synchronous/asynchronous Boolean Network
- Multiple fault detection in Markov Chain/Markov Process
- Optimal observer design for detecting most of the faults
- Optimal drug design for Boolean Network/Markov Chain environment
- Wet Lab validation

Daniel T. Gillespie.
A general method for numerically simulating the stochastic time evolution of coupled chemical reactions.
Journal of Computational Physics, 22:403–434, 1976.

R. Layek, A. Datta, and E. R. Dougherty.
From biological pathways to regulatory networks.
Mol. BioSyst., 7:843–851, 2011.

E. Batchelor, A. Loewer, and G. Lahav.
The ups and downs of p53: Understanding protein dynamics in single cells.
Nature Reviews: Cancer, 9(5):371–377, May 2009.

E. Batchelor, C. S. Mock, I. Bhan, A. Loewer, and G. Lahav.
Recurrent initiation: A mechanism for triggering p53 pulses in response to dna damage.
Mol. Cell, 30(3):277–289, May 2008.


R. Layek, A. Datta, M. Bittner, and E. R. Dougherty.
Cancer therapy design based on pathway logic.
Bioinformatics, 27(4):548–555, 2011.


P. R. Kumar and W. Lin.
Optimal adaptive controller for unknown markov chains.
IEEE Trans. on Automatic Control, 27:765–774, 1982.


R. Layek, A. Datta, R. Pal, and E. R. Dougherty.
Adaptive intervention in probabilistic boolean networks.
Bioinformatics, 25(16):2042–2048, 2009.

Question??

Thank You