Computational Molecular Biology and Bioinformatics

Systems Biology

Malay Bhattacharyya

Assistant Professor

Machine Intelligence Unit
Indian Statistical Institute, Kolkata

January, 2022

Basics

What is systems biology?

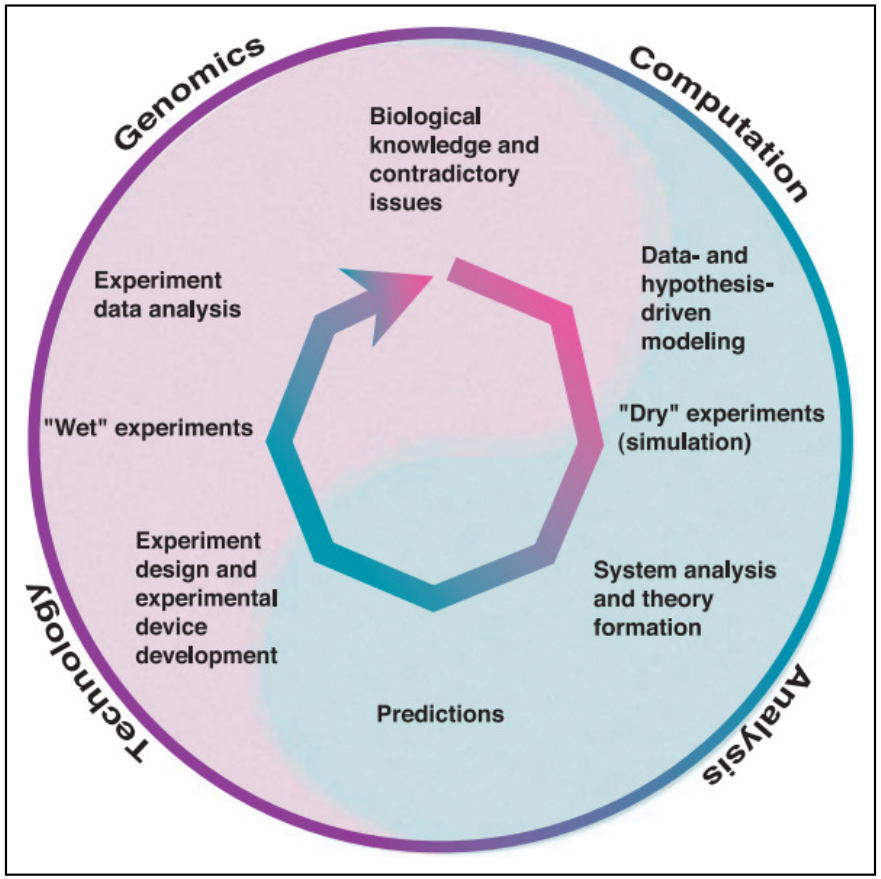
Systems biology concerns the system-level analysis that focuses on complex interactions within biological systems.

Why system-level analysis is necessary?

To understand the biology happening in vitro, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism.

The approaches towards systems biology

Systems biology approaches are mainly hypothesis-driven and follows a cyclic workflow as follows:



Biological insights from networks

Different network properties help to gain new biological insights.

Property	Nature	Biological interpretation
Order	High	Large-scale activity
	Low	Small-scale activity
Size	High	Large-scale activity
	Low	Small-scale activity
Path	Large	Indirect effect
	Small	Direct effect
Cycle	Large	Rare alternating path
Girth	Large	Sparse association
Degree	High	Acting as functional hub
Degree distribution	Power-law	Hubs are significant
Clustering coefficient	High	Dense associativity
	Low	Sparse associativity
Eccentricity	High	Hardly reachable
	Low	Easily reachable from anywhere
Radius	High	Sparse association
Diameter	Low	Dense association
Betweenness centrality	High	Mediator of many signals
Connected component	A single component	A functional module

Introduction

Differential patterns might appear at various levels of a network:

- **Differential node:** A node that shows differential pattern. E.g., a differentially expressed biomolecule.
- **Differential arc:** An arc that shows differential pattern. E.g., a pair of differentially co-expressed biomolecules.
- **Differential network:** A network that shows differential pattern. E.g., a co-expression network in control against the disease.

Note: The differential patterns might occur between disease types/subtypes, tissue types, different phenotypes, etc.

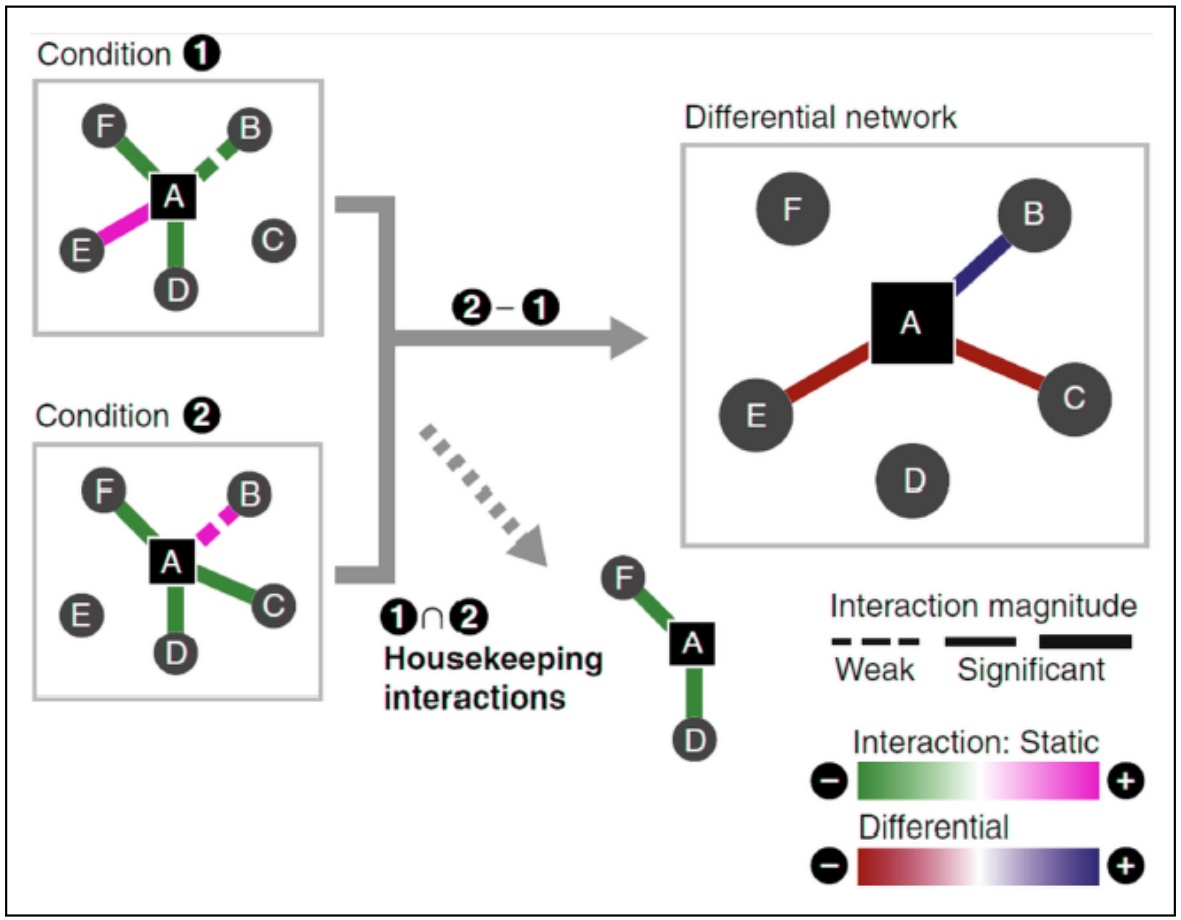
Types of differential networks

Differential networks can be built upon three different logics (symbolizing their differential patterns).

1. A network that comprises differential nodes (significantly vary between two separate conditions).
2. A network that comprises differential arcs (significantly vary between two separate conditions).
3. A network that comprises uncommon arcs between two networks (obtained from two separate conditions)

Fundamental idea

A differential network does not include housekeeping biomolecules (Ideker et al., *Mol Syst Biol* 8, 2012).



Approaches to constructing differential networks

Cross-species alignment of networks:

The biomolecular networks are aligned between two different species. E.g., the protein-protein interaction network of one organism can be compared against another organism to predict previously uncharacterized interactions or to identify evolutionarily conserved protein complexes.

Approaches to constructing differential networks

Dynamic mapping of interactions:

The alteration of physical interactions over two separate conditions are studied in this. E.g., the luminescence-based mammalian interactome mapping strategy is used to identify pairwise PPIs among a set of human factors in two different conditions.

Approaches to constructing differential networks

Cross-condition transcriptional regulation analysis:

The changes in transcriptional wiring that occur between two conditions are studied. E.g., how transcriptional regulation alters for genes for two different conditions can be studied to build networks.

Approaches to constructing differential networks

Cross-species transcriptional analysis:

The changes in transcriptional wiring are studied between two species. E.g., how transcriptional regulation alters between two different species can be studied to build networks.

Approaches to constructing differential networks

Cross-species genetic or chemogenetic analysis:

The alteration of genetic and chemogenetic interactions are studied between species.

Approaches to constructing differential networks

Cross-condition genetic interaction analysis:

The changes in genetic interactions that occur between two conditions are studied.

Approaches to constructing differential networks

Dynamic network mapping:

The quantitative changes in interaction with biomolecules are studied. E.g., the peak intensities for each peptide of a protein, which were combined into a weighted average intensity at each time point or condition. An intensity fold change was then calculated for each protein between two conditions, representing the change in interaction strength.

Approaches to constructing differential networks

Cross-condition epigenetic analysis:

The alterations in epigenetic behaviors that occur between two conditions are studied.

Tools for differential network analysis

- DNA (Gill et al., *BMC Bioinformatics* 11, 2010): Available as R code
- DDN (Zhang et al., *Bioinformatics* 27, 2011): Available as web server
- DINA (Gambardella et al., *Bioinformatics* 29, 2013): Available as web server

What is a regulatory network?

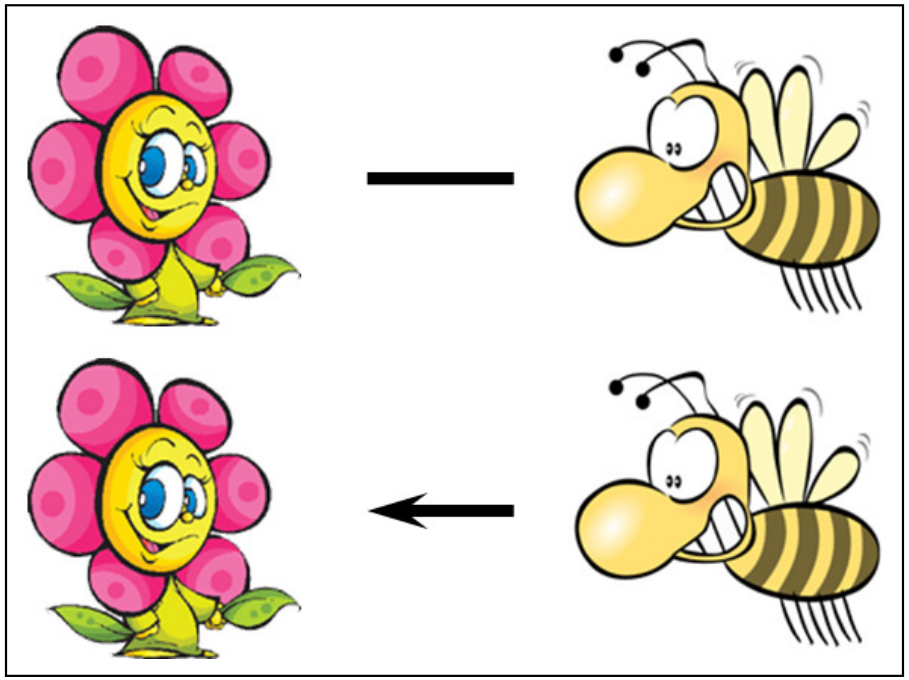
A regulatory network incorporates the regulatory information between the nodes it comprises.

Depending upon the biomolecule it comprises as its nodes, a biological regulatory network can be of different types:

- Gene regulatory network
- Protein regulatory network
- TF-gene regulatory network
- MicroRNA-gene regulatory network
- Viral-host regulatory network
- ... etc.

What is special in regulatory networks?

Regulatory networks are directed



Node pairs are distinguished based on additional information

Random Boolean networks

Definition (Boolean network)

A Boolean network $N = (V, A)$ is defined by a set of nodes V and a set of arcs A representing Boolean relations between the nodes.

Random Boolean networks (RBN) are generic models that can be used for modeling real-life systems (Kauffman, *OUP*, 1993).

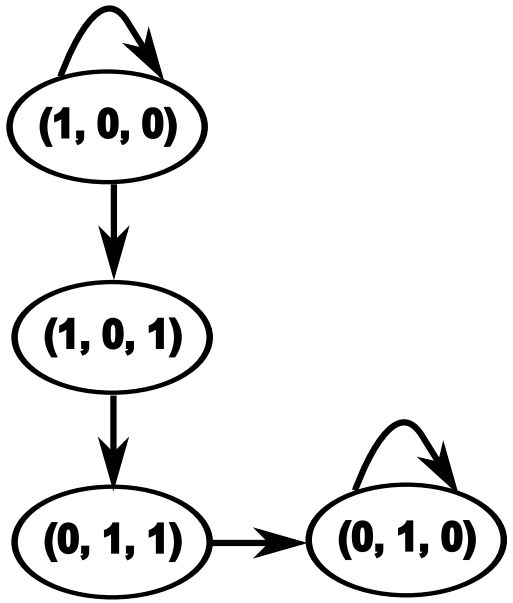
Consider an RBN having k number of nodes that can take the values zero (off) and one (on). The state of each node is determined by c connections coming from other (or the same) nodes. So, the state space of the RBN size becomes 2^k.

Note: Boolean networks are completely deterministic.

The state transition diagram

Considering the activity (on/off) of the three genes (shown previously) as a triplet state space (g_1, g_2, g_3) we can obtain the following state transition diagram:

Time point t	Time point (t+1)
(1, 0, 0)	(1, 0, 0)
(1, 0, 0)	(1, 0, 0)
(1, 0, 0)	(1, 0, 1)
(1, 0, 1)	(0, 1, 1)
(0, 1, 1)	(0, 1, 0)
(0, 1, 0)	(0, 1, 0)
(0, 1, 0)	(0, 1, 0)
(0, 1, 0)	(0, 1, 0)



Attractors in RBN

Definition (Attractor)

If a state space is repeated in an RBN the network is said to have reached an attractor.

Definition (Point attractor)

If the attractor consists of one state, it is called a point attractor or steady state.

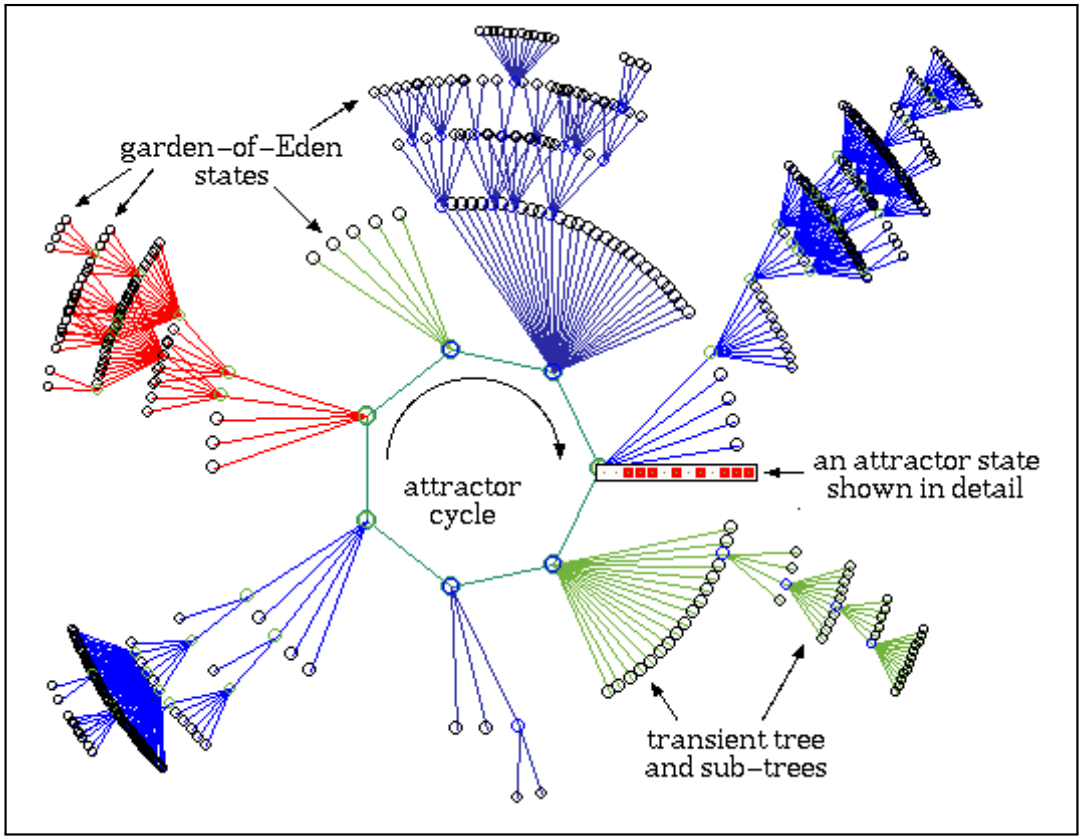
Definition (Cycle attractor)

If the attractor consists of two or more states, it is called a cycle attractor or state cycle.

Attractors in RBN

Definition (Basin of attraction)

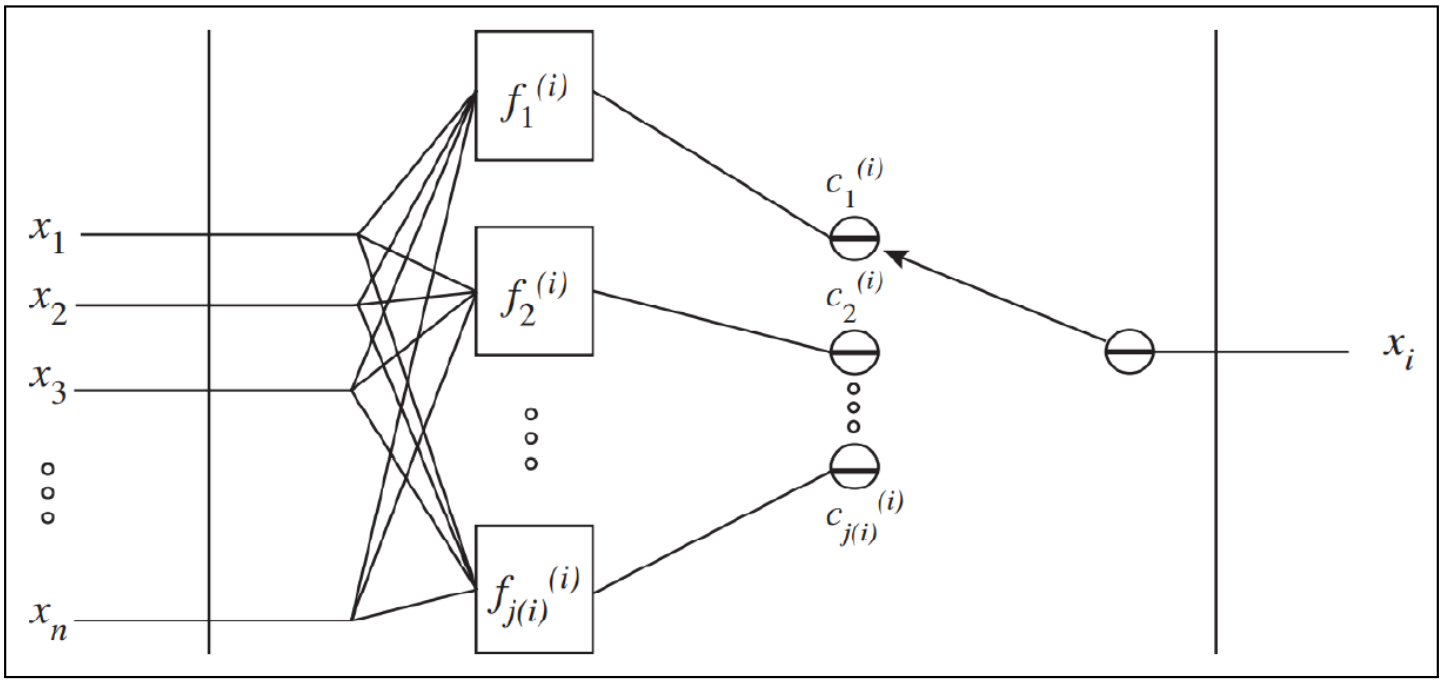
The set of states that flow towards an attractor is called the basin of attraction.



Probabilistic Boolean networks

To overcome the inflexibility of 'deterministic behavior' of Boolean networks, we incorporate the notion of probability in the model.

Here, more than one possible function (known as predictor) defines a single node (Shmulevich et al., *Bioinformatics* 18, 2002).



Petri nets

Definition (Petri net)

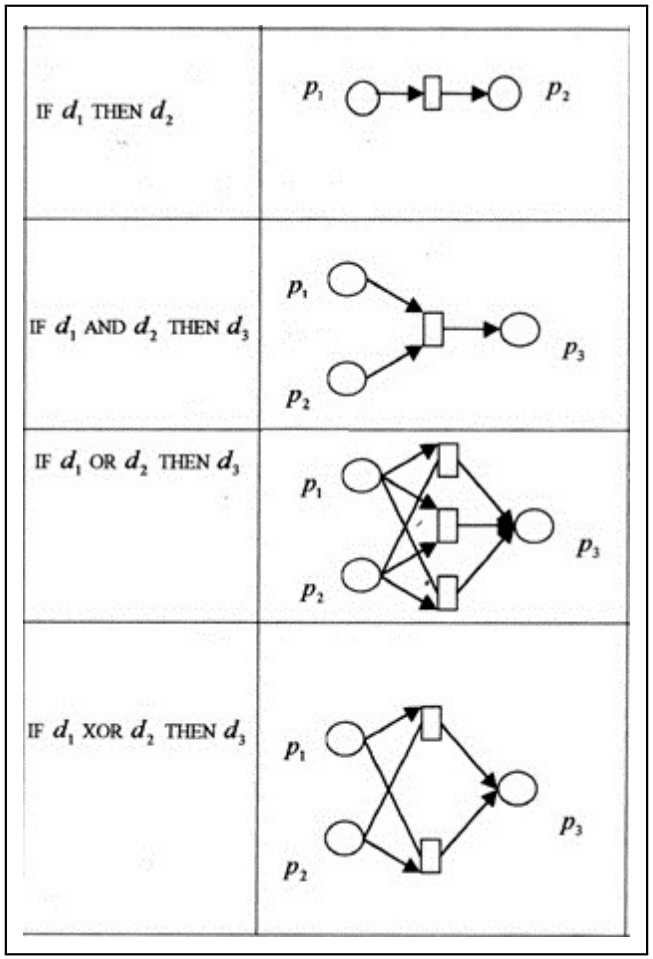
A Petri net $N = (V_1, V_2, A)$ is represented by two distinct sets of nodes V_1 and V_2 denoting the transitions and places, respectively, and a set A of directed arcs between the transitions and places.

- Transitions (e.g., the events that may occur) are signified by bars
- Places (e.g., the conditions) are signified by circles
- Directed arcs (describing which places are pre- and/or postconditions for which transitions) are signified by arrows

Note: A Petri net is also known as a place/transition net or P/T net.

Petri nets

Different types of logical relations can be represented through Petri net models.



Other models

Some other recent approaches are listed below (Karlebach et al., *Nat Rev Mol Cell Biol* 9, 2008):

- Finite state linear models
- Dynamic Boolean networks
- Probabilistic graphical models
- Stochastic networks

Differentiating direct and indirect regulation

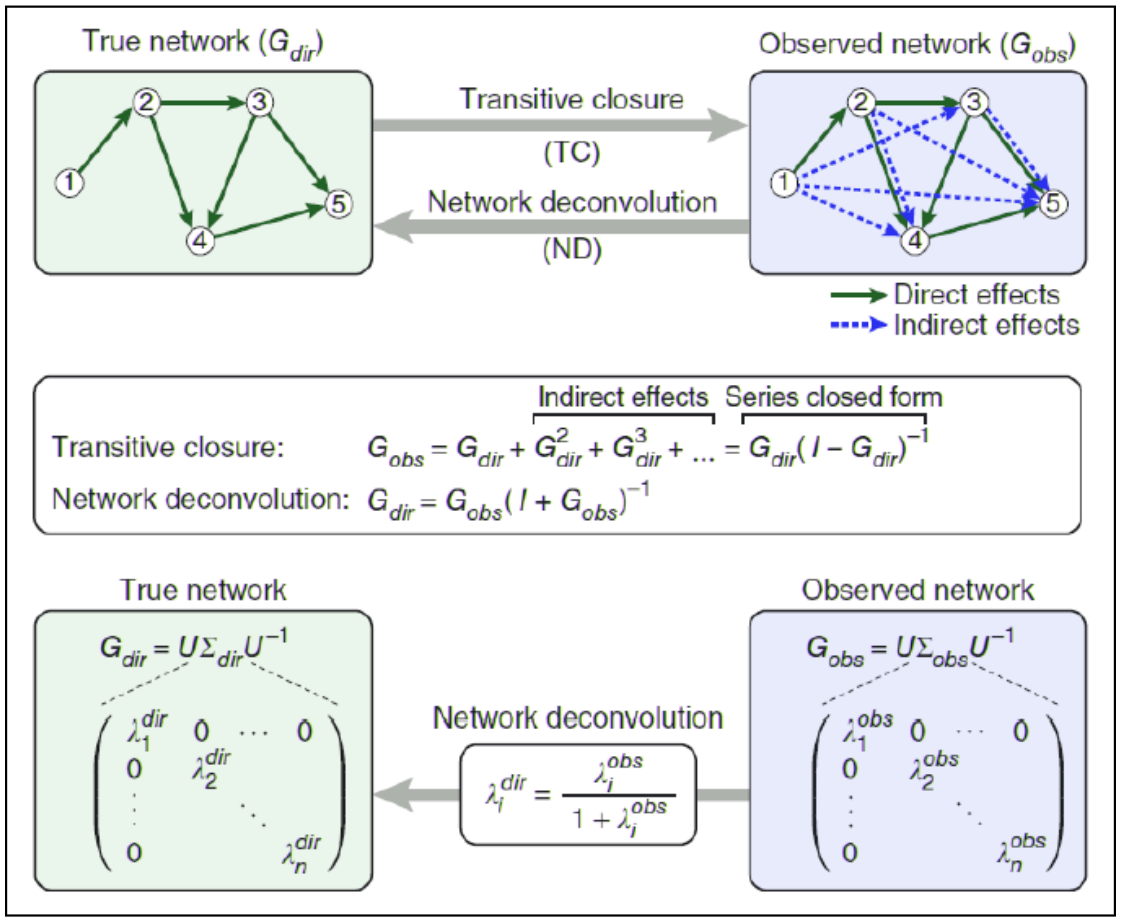
Regulatory networks comprise both direct and indirect effects. It is possible to remove the combined effect of indirect paths from such networks and recognize the direct regulations only.

Using network deconvolution we can reverse the effect of transitive information flow across all indirect paths and derive the true direct network.

This finally retains the direct effects (regulatory) in the network.

Differentiating direct and indirect regulation

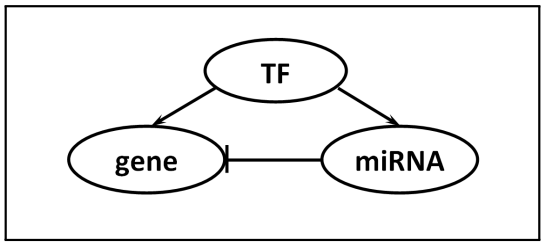
The network deconvolution works as follows (Feizi et al., *Nat Biotechnol* 31, 2013).



Incorporating diverse regulation information

Diverse regulation types might exist between the biomolecules:

1. Transcription factors to genes
2. MicroRNAs to genes
3. Transcription factors to microRNAs



What is known so far in human?

# Genes:	$\sim$ 30000
# Transcription factors:	$\sim$ 2000
# MicroRNAs:	$\sim$ 2500

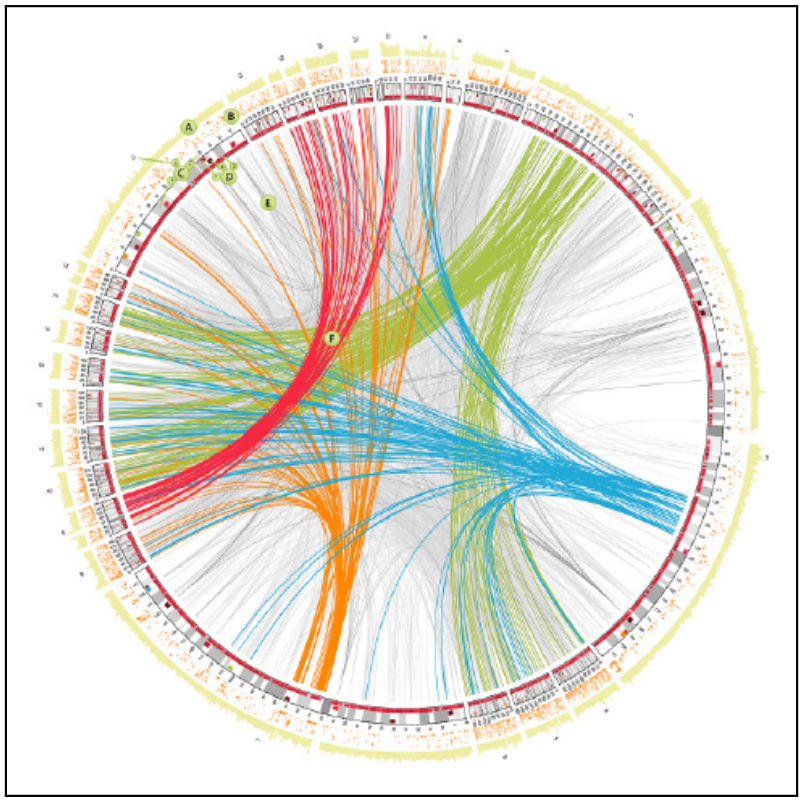
So, the complexity of 'transcription factor-gene-microRNA' inter-regulatory network becomes very high!

State-of-the-art knowledge

1. **TF→gene**
   - Past: Sequence match based on PWMs (TRANSFAC), Conservation analysis and TSS proximity analysis
   - Recent: Histone modification and DNase I cleavage
2. **miRNA⊣mRNA**
   - Past: Microarray, Reporter gene assay, Northern blot, Western blot, PCR, pSILAC, SILAC, prediction methods (MultiMiTar, TargetSpy, TargetScan, DIANA MicroT 3.0, TargetMiner, miRanda, PITA, PicTar), TarBase 6.0, etc.
   - Recent: Degradome-seq, CLIP-seq, etc.
3. **TF→miRNA**
   - Past: Curation (TransmiR), Sequence match (PuTmiR), etc.
   - Recent: TSS proximity

What is network dynamics?

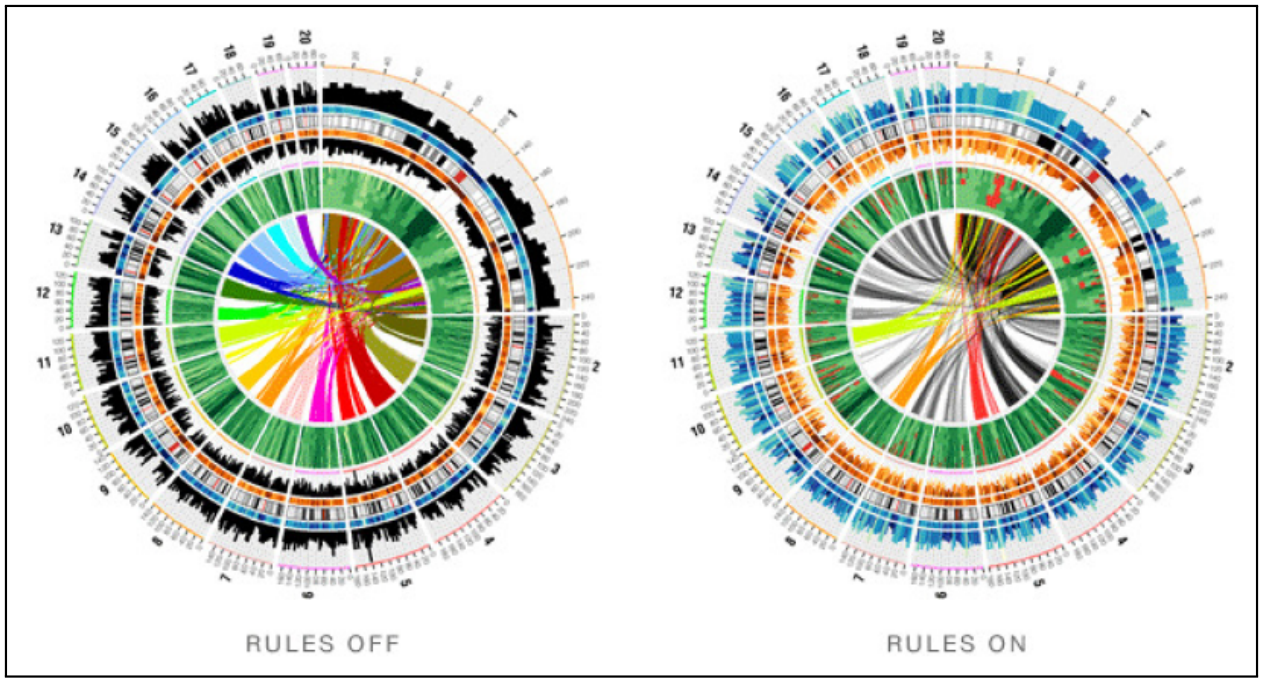
Network dynamics is the study of networks that change over time.



A network where interactions at different time points are shown with different colors

Regulatory dynamics

The dynamics of a network can be studied by considering the regulatory effect (on/off) of a single biomolecule.



Types of network dynamics

The endogenous and exogenous networks have distinguishable patterns (Luscombe et al., *Nature* 431, 2004).

	Endogenous	**Exogenous**
TF combinations	complex	simple
Path length	long	short
TF outdegree	Low	High
Significant motifs	FFLs	single input
TF interconnectedness	High	Low

Topological properties of networks

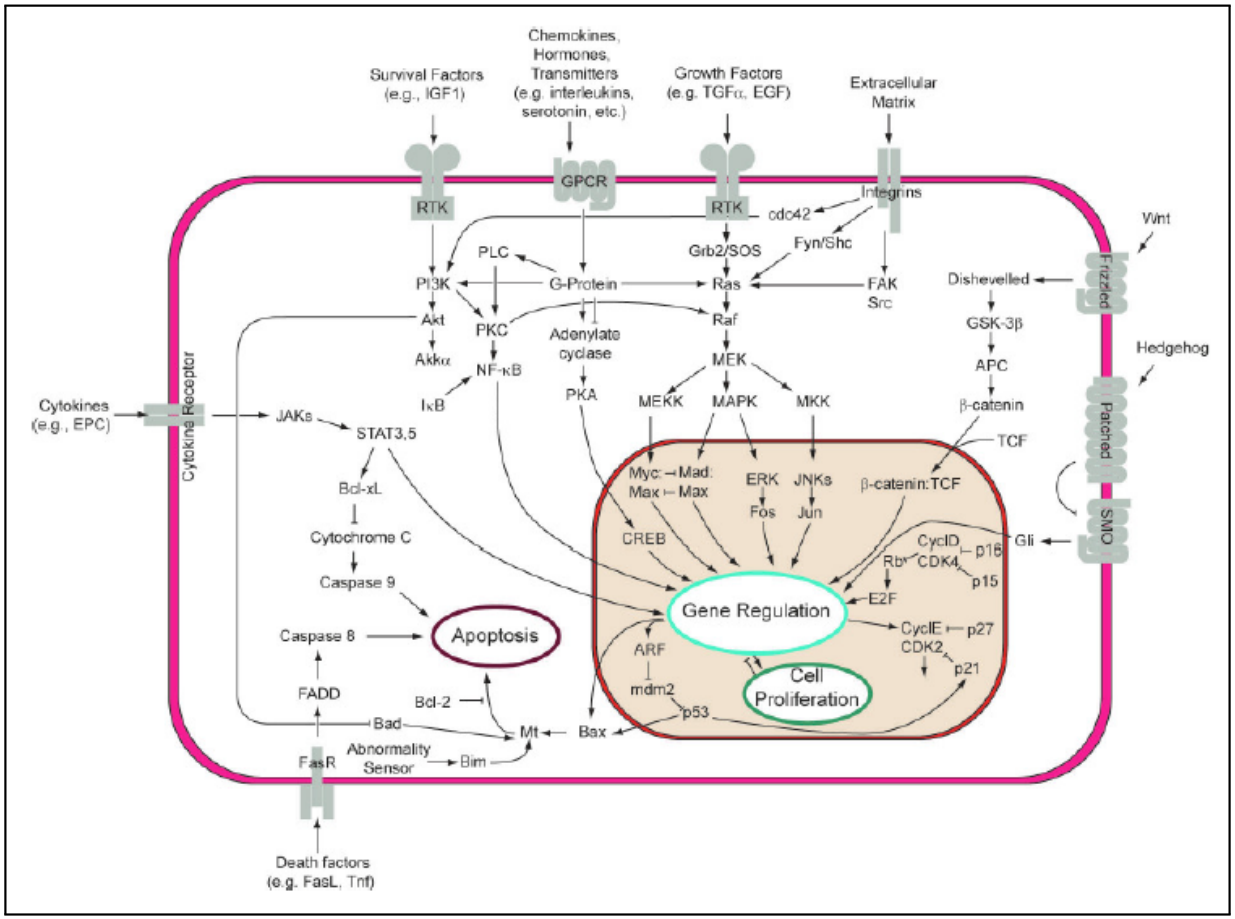
Analysis of network dynamics

Some tools for studying network dynamics are given below:

- SANDY
- GeneVis
- Circos
- CAVE

What is a pathway?

A biological pathway is a series of actions among molecules in a cell that leads to a certain product or a change in a cell.



Types of pathways

Some of the most common biological pathways are involved in metabolism, the regulation of gene expression and the transmission of signals.

Pathways are mainly of three types:

1. Metabolic pathways
2. Signal transduction pathways
3. Gene regulation pathways

Note: KEGG is an important resource of pathways.

Feedback inhibition in pathways

The steady state activity is maintained through feedback inhibition in a pathway. So pathways can be modeled with and without feedback inhibition to understand differential effects.

Partial and ordinary differential equations are mainly used to model feedback inhibition.

What is crosstalk?

Biological crosstalk refers to instances in which one or more components of one pathway affect another component in another pathway.

This can be achieved through a number of ways with the most common form being crosstalk between proteins of signaling cascades.

Crosstalk can be of different types as follows:

- Crosstalk Between Signaling Pathways
- Transmembrane Crosstalk
- Crosstalk in Lymphocyte Activation
- ... etc.

Modeling crosstalk of pathways

Crosstalk modeling helps to understand the effect of one pathway on another.

Some of the commonly used approaches are listed below:

- Flux balance analysis
- Petri nets

Hands-on

1. KEGG stands for Kyoto Encyclopedia of Genes and Genomes. Open the KEGG portal (`https://www.genome.jp/kegg`) and do the following.
   i) Search for the term 'SARS-CoV-2'.
   ii) Open the "SARS-CoV-2 S to AngII-AT1R-NOX2 signaling pathway" under the KEGG NETWORK.
   iii) Report the interesting things that you observe in the signaling pathway.