Biological molecular networks

An introduction

Key concepts in networks

- **What** are molecular networks?
  - Graph-theoretic representation
  - Different types of networks
- **Why** are they useful for interpreting biological data?
  - Analysis of global network properties
  - Data integration
  - Hypothesis generation
- **How** can we learn networks from high-throughput data?
  - Classes of methods for expression-based network inference
  - Probabilistic graphical models

Understanding a cell as a system

- *Measure: identify the parts of a system*
  - Parts: different types of bio-molecules
    - Genes, proteins, metabolites
  - High-throughput assays to measure these molecules
- *Model: how these parts are put together*
  - Clustering
  - **Network inference and analysis**

Representing molecular networks

- Molecular networks typically represented by **graphs**
- Vertices/Nodes = a molecular part
- Edges = connections or interactions between parts
- Edges can have directionality, signs, and/or weight
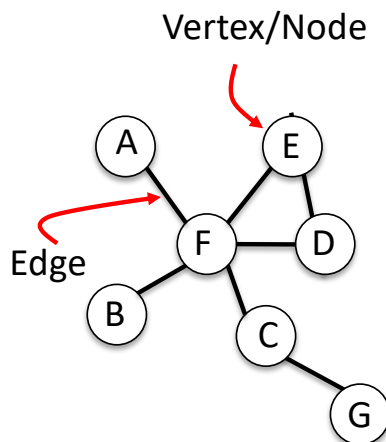
Different types of molecular networks

- Depends on what
  - the vertices represent
  - the edges represent
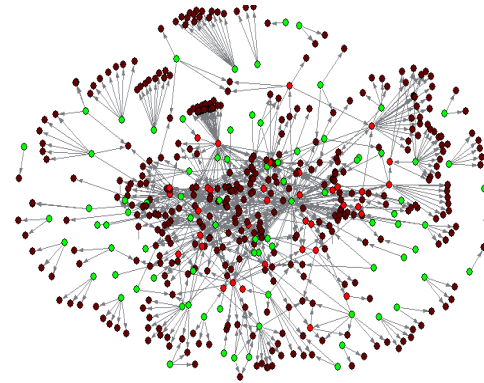 - whether edges directed or undirected

Transcriptional regulatory networks

Nodes: Regulatory protein like a TF, or target gene
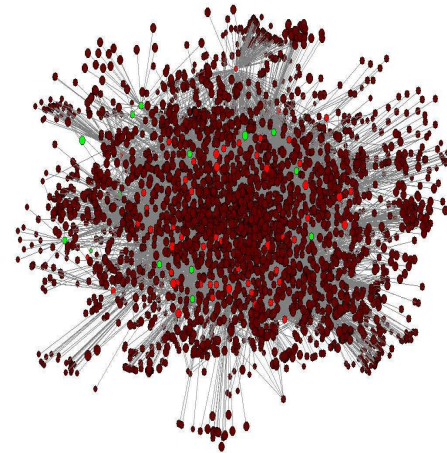Edges: TF A regulates gene C

Transcription factors
(TFs)

A   B

Gene C

A   B

C

Directed,
weighted
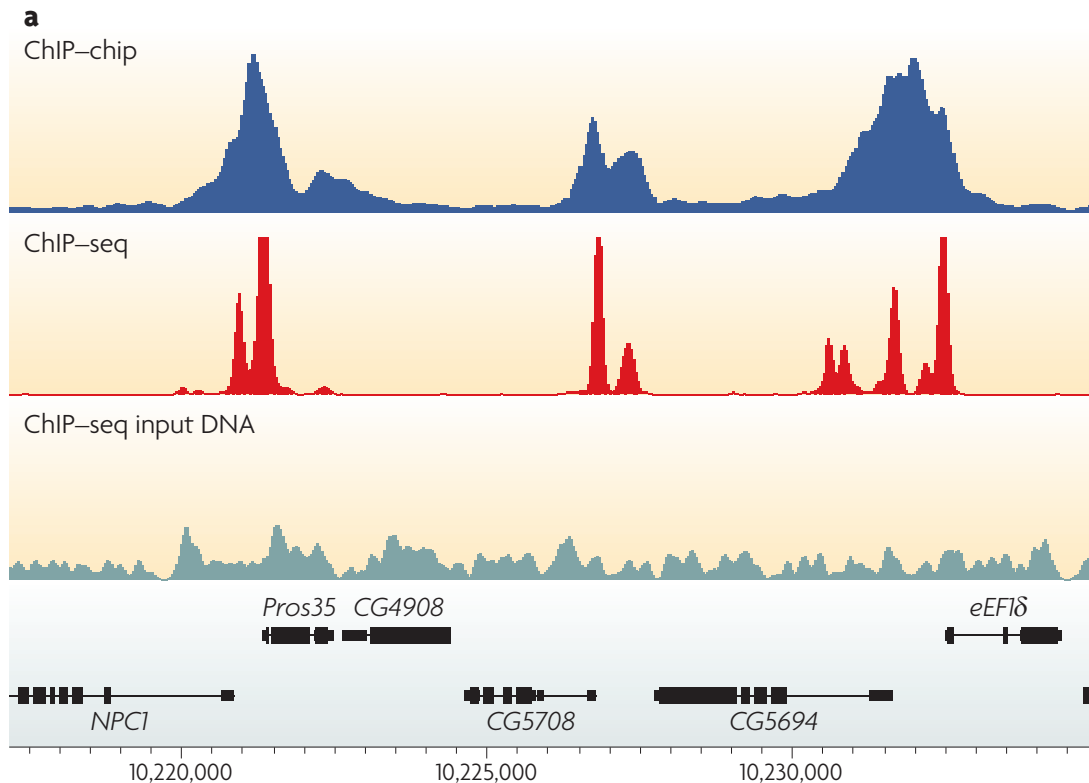
Vargas and Santillan, 2008
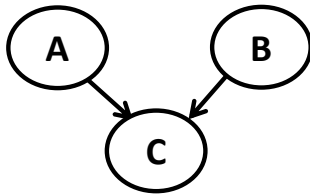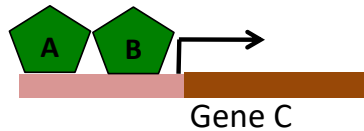
E. coli: 153 TFs and 1319 target genes

S. cerevisiae: 157 TFs and 4410 target genes
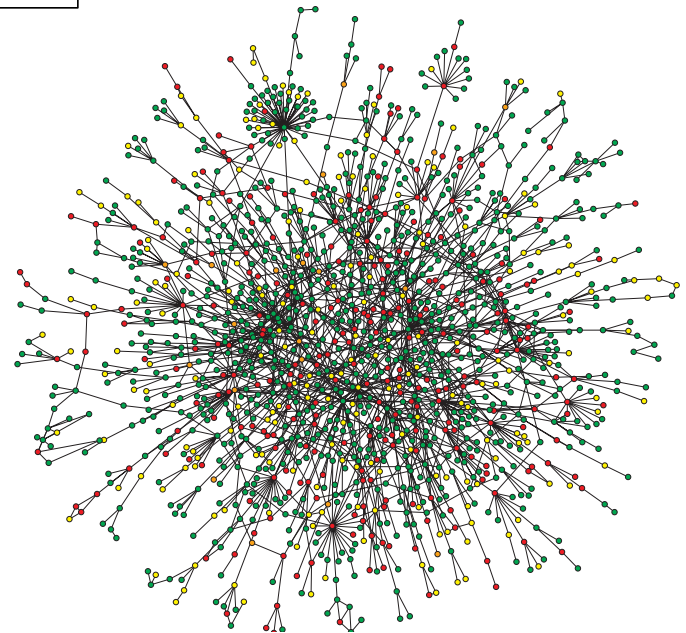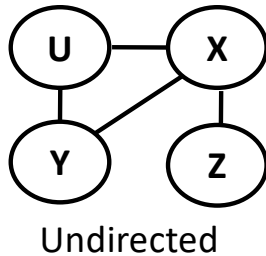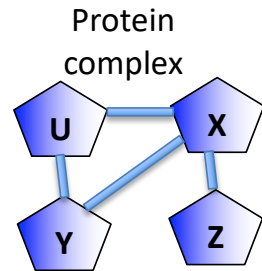
6

Detecting protein-DNA interactions

- ChIP-chip and ChIP-seq binding profiles for transcription factors
- Determine the (approximate) locations in the genome where a protein binds

Protein-protein interaction networks

Vertices: Proteins
Edges: Protein U physically interacts with protein X

Protein complex

Undirected

Yeast protein interaction network

Barabasi et al. 2004

Metabolic networks

Vertices: Enzymes
Edges: Enzyme M and N share a metabolite

Proteins (enzymes)

Metabolites

Enzymes

Metabolites

Undirected,
weighted

Metabolic networks

Vertices: Enzymes
Edges: Enzyme M and N share a metabolite

Proteins (enzymes)

Enzymes

Metabolites

M

a — b

N

c

O

d

M — N — O

Undirected,
weighted

Metabolites

1.1.3.9

1.1.1.48
1.1.1.120   D-Galactono-
1,4-lactone

3.1.1.25   D-Galactonate   4.2.1.6   2-Dehydro-3-deoxy-
D-galactonate   2.7.1.58

UDPglucose

Enzyme: 4. Lyases
4.2 Carbon-oxygen lyases
4.2.1 Hydro-lyases
4.2.1.6 galactonate dehydratase

C00880

D-Galactonate

C00001

C01216

**2-Dehydro-3-deoxy-
D-galactonate**
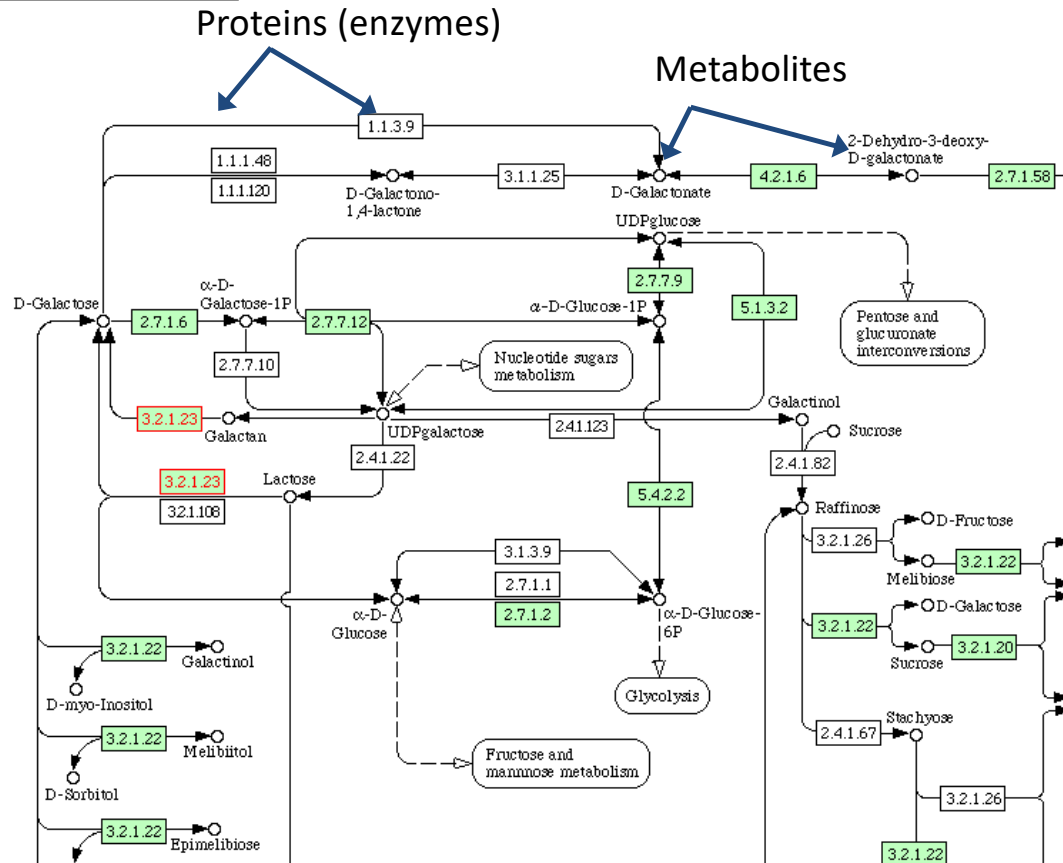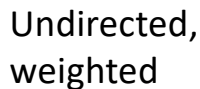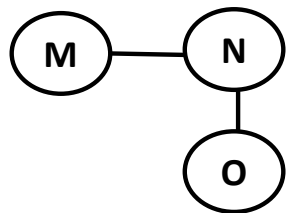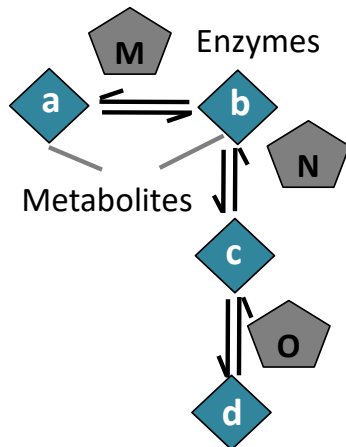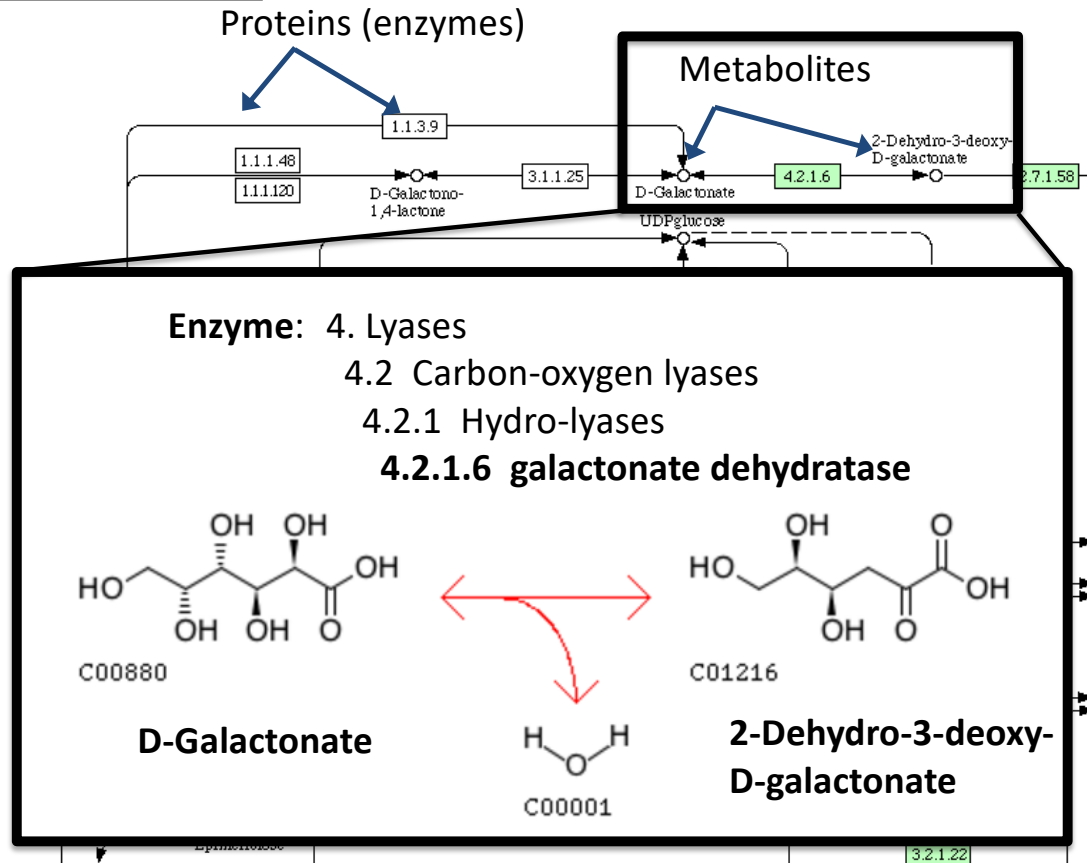
Overview of the *E. coli* Metabolic Pathway Map

Image from the KEGG database

Signaling networks

Vertices: Enzymes and other proteins
Edges: Enzyme P modifies protein Q

Receptors

Directed

TF

Activators
1. α-CD3
2. α-CD28
3. ICAM-2
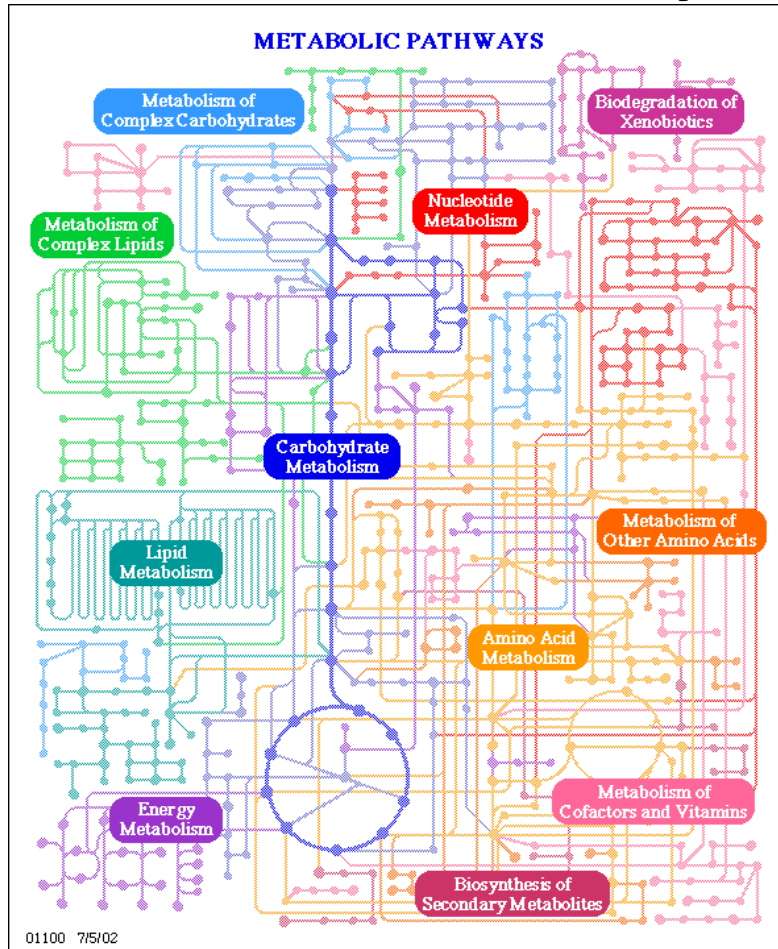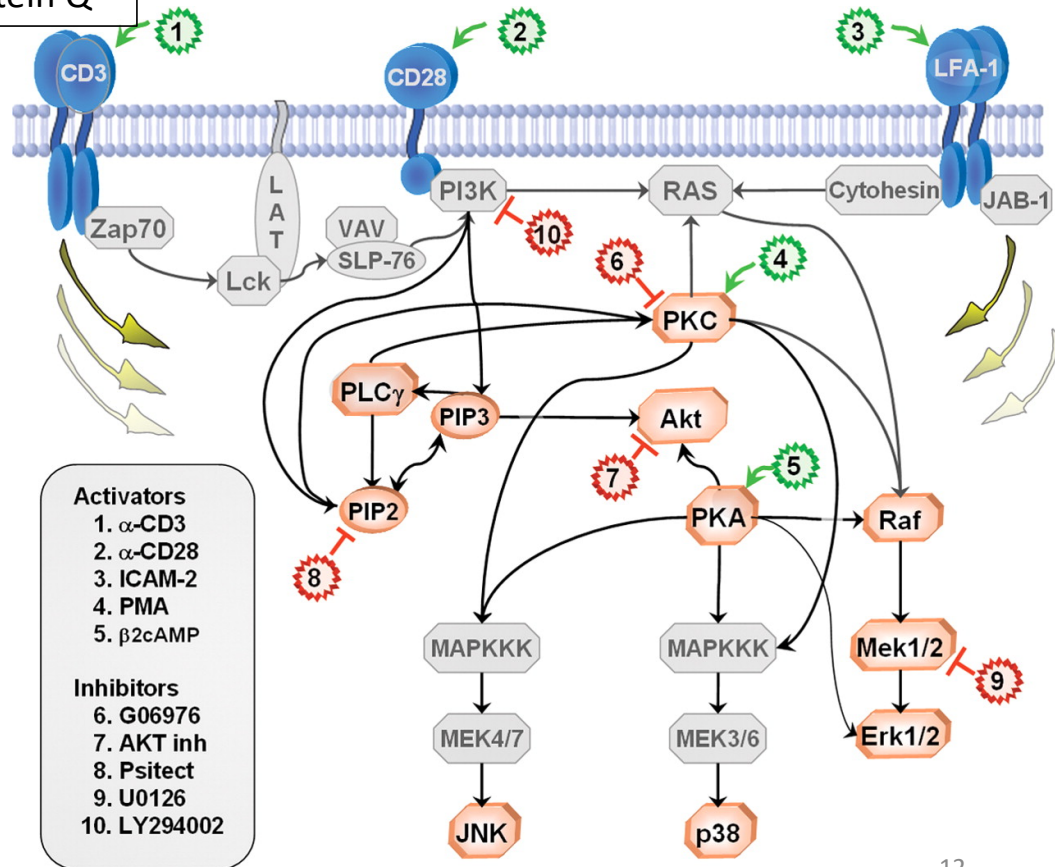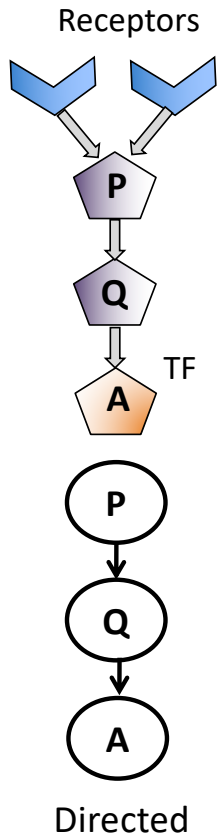4. PMA
5. β2cAMP

Inhibitors
6. G06976
7. AKT inh
8. Psitect
9. U0126
10. LY294002

Sachs *et al.*, 2005, Science

12

Computational problems in networks

- Analysis of network properties
  - Degree distributions
  - Network motifs
  - Network modules
  - Highly connected nodes and relationship to lethality
- Network applications
  - Interpretation of gene sets
  - Identification of condition-specific subnetworks or pathways
  - Using networks to infer functions of a gene
- Network reconstruction
  - Infer the structure and parameters of networks
  - We will examine this problem in the context of "expression-based network inference"

Network reconstruction

- Given
  - A set of measurements associated with network nodes
  - Typically measurements are mRNA levels
- Do
  - Infer which nodes interact
- Algorithms for network reconstruction can vary based on their meaning of interaction
  - Similarity
  - Mutual information
  - Predictive ability

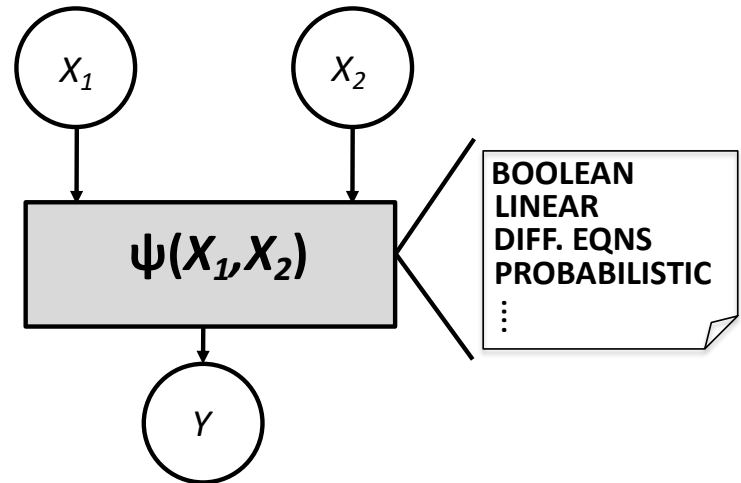# Computational methods to infer networks

- We will focus on transcriptional regulatory networks
- These networks are often inferred from gene expression data
- Many methods to do network inference
  - We will focus on probabilistic graphical models

Modeling a regulatory network

Hot1 **Sko1**

HSP12

Hot1 **Sko1**

Hot1 regulates
HSP12

↕

HSP12 is a target
of Hot1

HSP12
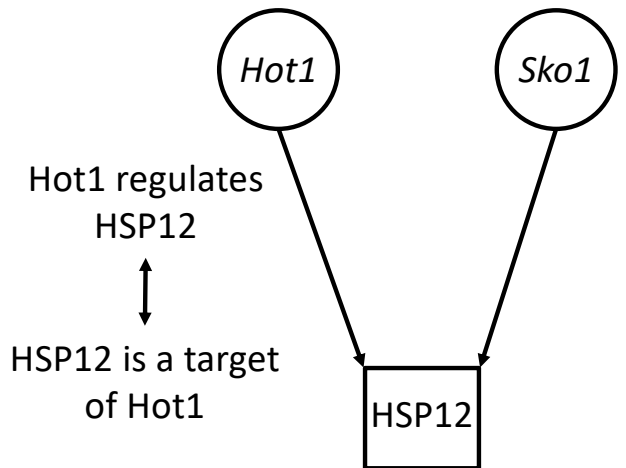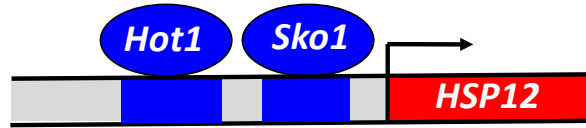
Structure

Who are the regulators?

X_1 X_2

$\psi(X_1, X_2)$

**BOOLEAN
LINEAR
DIFF. EQNS
PROBABILISTIC**
⋮

Y

Function

How they determine expression levels?

Mathematical representations of networks

X_1
X_2 Input expression of neighbors

f

X_3 Output expression of node
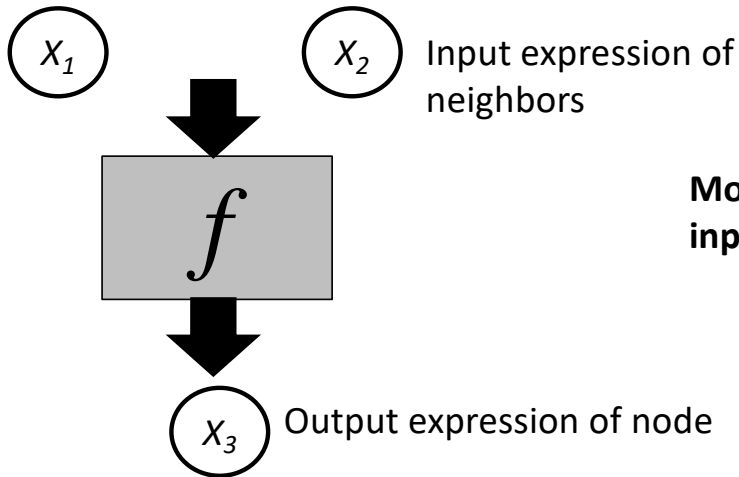
Models differ in the function that maps input system state to output state

Boolean Networks

Input		Output		
X1	X2	**X1**	**X2**	**X3**
		0	0	0
		0	1	1
		1	0	1
		1	1	1

X3

Differential equations

$$\frac{dX_3(t)}{dt} = \kappa\ g(X_1(t), X_2(t)) - rX_3(t)$$

Rate equations

Probabilistic graphical models

$$P(X_3|X_1, X_2) = N(X_1 a + X_2 b, \sigma)$$

Probability distributions

Summary

- Many types of molecular networks
  - Transcriptional regulatory
  - Protein-protein interaction
  - Metabolic
  - Signaling
- Variety of computational tasks associated with networks
- Our focus: reconstruction of transcriptional regulatory networks
- Variety of approaches to modeling transcriptional regulatory networks
- Our focus: probabilistic graphical models