Inferring Protein-Signaling Networks II

Lectures 15 – Nov 16, 2011
CSE 527 Computational Biology, Fall 2011

Instructor: Su-In Lee
TA: Christopher Miles

Monday & Wednesday 12:00-1:20
Johnson Hall (JHN) 022

Closing the Loop

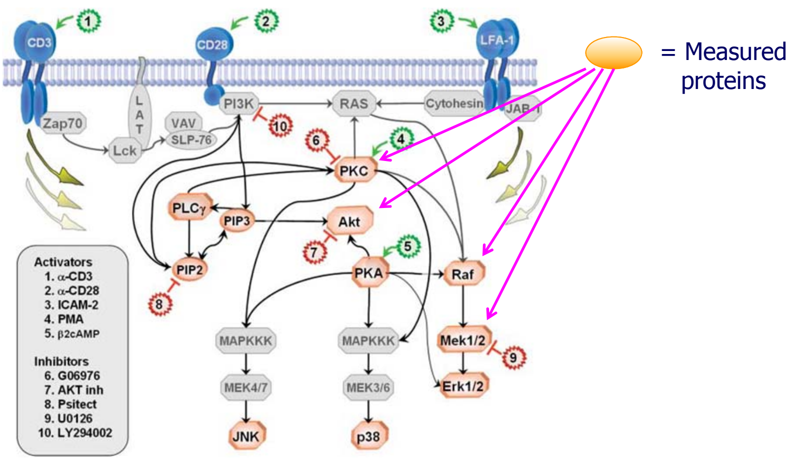
- Thank you for your participation to the survey!
- Things that helped you
  - A very diverse set of topics
  - Well-organized
  - "Who is with me?"
  - Quality of the slides
- Things that did not help you
  - Lack of depth
    - Intended to be achieved through problem sets and projects
  - HW problems
    - Needs to improve clarity

Outline

- Inferring the protein-signaling network
  - Computational problem
    - Structure score
    - Structure search
  - Interventional modeling
  - Evaluation of results
  - Conclusion
- Key learning points
  - Structure learning of Bayesian network
  - Intervention modeling
  - Evaluation of the inferred biological network

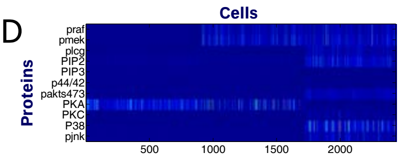
Inferring Signaling Networks

- Signaling networks are full of post-translational regulatory mechanisms (e.g. phosphorylation)

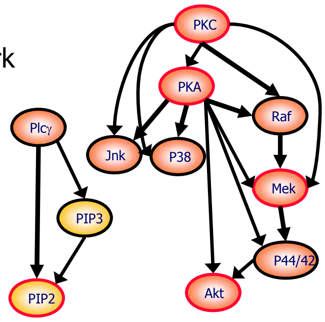


Computational Problem

- Given data D



- Goals
  - Infer the causal interaction network

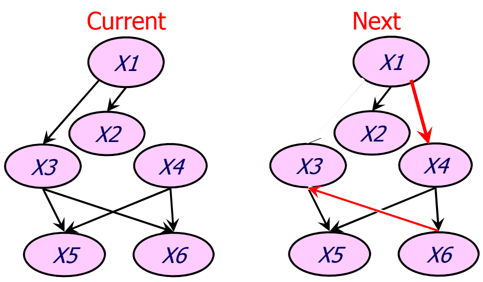


- Structure learning of Bayesian network
  - General machine learning problem
  - Applicable to different areas in network biology

5

Structure Search

- Score-based learning algorithm
  - Given a set of all possible structures and the scoring function, we aim to select the highest scoring network structure.
- Greedy hill-climbing search
  - Make one edge change which maximizes the graph score



Add an edge
Remove an edge
Reverse an edge

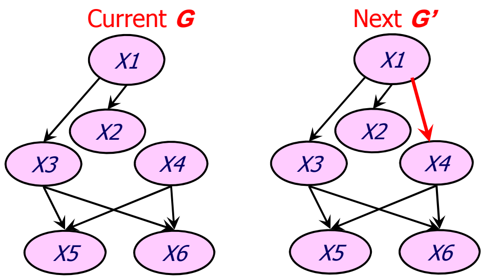
*Constrained to non-cyclic modifications

- Importance of score decomposition

6

Score Decomposition

- Greedy hill-climbing search
  - Make one edge change which maximizes the graph score



Add an edge

Compare G and G' in terms of the structure score

- If score decomposability holds,
  - Score of G = $S_1(X_1, PaX_1)+S_2(X_2, PaX_2)+...+S_6(X_6, PaX_6)$, where $S_i\,(X_i, PaX_i)$ is a "FamScore" for node X_i
  - When G→G', we only need to re-compute $S_4(X_4, PaX_4)$
- How about commonly used structure scores?

Maximum Likelihood

D = Data G=Graph
θ = CPT values for each node X
θ_{ijk} = P(Xi=k | Parents(Xi)=j)
N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

- Find G that maximizes:

  P(Data=D | Graph=G, Θ_{MLE})

K=#discrete levels of X

N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

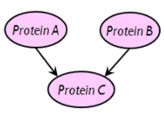
$$= \prod_{i=1}^{\#proteins} \prod_{j=1}^{\#parent\ states} \left(\prod_{k=1}^{K} \theta_{ijk}^{N_{ijk}} \right)$$

θ_{ijk} = P(Xi=k | Parents(Xi)=j)

$$\Theta_{ijk}^{ML} = N_{ijk} / \sum_k N_{ijk}$$

Conditional Probability Table (CPT)

[# of (0,0,0)] / [# of (0,0,*)]



A	B	P(C=0|Pa)		P(C=1|Pa)	
0	0	Θ_{C10}	+	Θ_{C11}	= 1
0	1	Θ_{C20}	+	Θ_{C21}	= 1
1	0	Θ_{C30}	+	Θ_{C31}	= 1
1	1	Θ_{C40}	+	Θ_{C41}	= 1

k, j, θ_{ijk}

D = [(0,0,0), (0,0,1),
(1,1,1), (1,0,0),
(1,1,1), (1,0,1),
:
(1,0,0), (1,0,1)]

Structure Score

- Bayesian score

P(Data=D | Graph=G)

$$= \int \underbrace{P(D|G,\theta)}_{\text{Multinomial}} \; \underbrace{P(\theta|G)}_{\text{Dirichlet prior} \sim \text{Dir}(\alpha)} \; d\theta$$

P(D, θ | G)

D = Data G=Graph
θ = CPT values for each node X
θ_{ijk} = P(Xi=k | Parents(Xi)=j)
N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

$$P(\theta_{ij1},\dots,\theta_{ijK}) \sim \prod_{k=1}^{K} \theta_{ijk}^{\alpha_{ijk}-1}$$

K=#discrete levels of X

N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

Dirichlet normalizer

$$P(D|G) = \prod_{i=1}^{\#proteins} \prod_{j=1}^{\#parent\ states} \int_{\Theta_{ij}} \left(\prod_{k=1}^{K} \theta_{ijk}^{N_{ijk}} \right) \left(\frac{1}{B(\alpha_{ij})} \prod_{k=1}^{K} \theta_{ijk}^{\alpha_{ijk}-1} \right) d\theta_{ij}$$

Θ_{ij} = Simplex $\{\sum_k \theta_{ijk} = 1\}$

θ_{ijk} = P(Xi=k | Parents(Xi)=j)

D. Heckerman. A Tutorial on Learning with Bayesian Networks. 1999, 1997, 1995.
G. Cooper E. Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning, 9, 309-347. 1992.

Structure Score

D = Data G=Graph
θ = CPT values for each node X
θ_{ijk} = P(Xi=k | Parents(Xi)=j)
N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

Dirichlet normalizer

$$B(\alpha) = \int_{\Delta^K} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} d\theta = \frac{\prod_{k=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{k=1}^{K} \alpha_i)}$$

$$P(D|G) = \prod_{i=1}^{\#proteins} \prod_{j=1}^{\#parent\ states} \frac{1}{B(\alpha_{ij})} \int_{\Theta_{ij}} \left(\prod_{k=1}^{K} \theta_{ijk}^{N_{ijk}+\alpha_{ijk}-1} \right) d\theta_{ij}$$

$$P(D|G) = \prod_{i=1}^{\#proteins} \prod_{j=1}^{\#parent\ states} \frac{B(\alpha_{ij} + N_{ij})}{B(\alpha_{ij})}$$

$$P(D|G) = \prod_{i=1}^{\#proteins} \prod_{j=1}^{\#parent\ states} \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ijk})}{\Gamma(\sum_{k=1}^{K} \alpha_{ijk} + N_{ijk})} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}$$

G. Cooper E. Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning, 9, 309-347. 1992.

Structure Score

D = Data G=Graph
θ = CPT values for each node X
θ_{ijk} = P(Xi=k | Parents(Xi)=j)
N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

$$P(D|G) = \prod_{i=1}^{\#proteins} \underbrace{\prod_{j=1}^{\#parent\ states} \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ijk})}{\Gamma(\sum_{k=1}^{K} \alpha_{ijk} + N_{ijk})} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}}$$

$$\mathrm{FamScore}(X_i, Pa_i|D) = \log \prod_{j=1}^{\#parent\ states} \frac{\Gamma(\sum_{k=1}^{K} \alpha_{ijk})}{\Gamma(\sum_{k=1}^{K} \alpha_{ijk} + N_{ijk})} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}$$

$$\mathrm{Score}(G|D) = \sum_{i=1}^{\#proteins} \mathrm{FamScore}(X_i, Pa_i|D) + \log P(G)$$

Decomposability!

$$\mathrm{Score}(G|D) = \log P(D|G) + \log P(G)$$

Bayesian Score

D = Data G=Graph
θ = CPT values for each node X
θ_{ijk} = P(Xi=k | Parents(Xi)=j)
N_{ijk} = # times Xi=k and Parents(Xi)=j in the Data

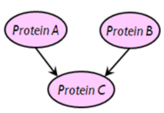
- P(Data=D | Graph=G)
  = ∫ P(D|G,θ) P(θ|G) dθ

$$= \prod_{i=1}^{\#proteins} \prod_{j=1}^{\#parent\ states} \int_{\Theta_{ij}} \left(\prod_{k=1}^{K} \theta_{ijk}^{N_{ijk}} \right) \left(\frac{1}{B(\alpha_{ij})} \prod_{k=1}^{K} \theta_{ijk}^{\alpha_{ijk}-1} \right) d\theta_{ij}$$

$$\Rightarrow \Theta_{ijk}^{BS} = (N_{ijk} + \alpha_{ijk}) / \textstyle\sum_k (N_{ijk} + \alpha_{ijk})$$

"Imaginary" counts

[# of (0,0,0)+ α_{C10}] / [# of (0,0,*) + α_{C10}+ α_{C11}]



Conditional Probability Table (CPT)

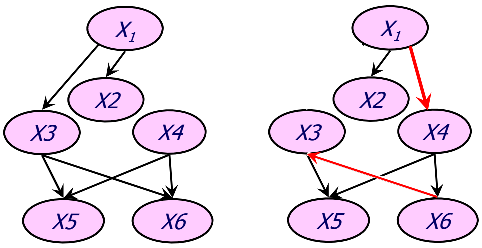
A	B	P(C=0|Pa)		P(C=1|Pa)	
0	0	Θ_{C10}	+	Θ_{C11}	= 1
0	1	Θ_{C20}	+	Θ_{C21}	= 1
1	0	Θ_{C30}	+	Θ_{C31}	= 1
1	1	Θ_{C40}	+	Θ_{C41}	= 1

j, k, θ_{ijk}

Structure Learning Algorithm

- Greedy hill-climbing search
  - Make one edge change which maximizes the graph score

Current ***G*** Next ***G'***



Add an edge Update $S_4(X_4, PaX_4)$ when $G \rightarrow G'$

Remove an edge Update $S_3(X_3, PaX_3)$ when $G \rightarrow G'$

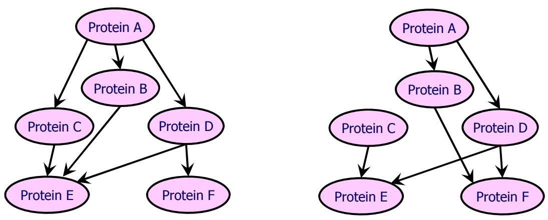
Reverse an edge Update $S_3(X_3, PaX_3)$ when $G \rightarrow G'$

- Repeat to make one edge changes until the score no longer increases (find local maxima)

13

Model Averaging

- Generate N graphs using bootstrapped data

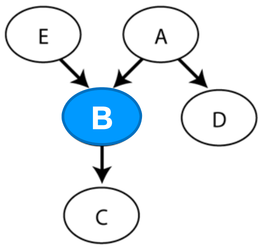


- Confidence(feature f) = # graphs with feature f / # graphs
- Select a confidence threshold

14

Causal Networks

- Bayes net is NOT a causal net



Conditional Independence

(A ⊥ E)
(B ⊥ D | A,E)
(C ⊥ A,D,E | B)
(D ⊥B,C,E | A)
(E ⊥A,D)

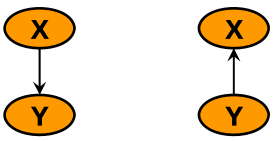
- Does structure learning of the Bayesian network reveal causal relationships?
  - Not necessarily…

15

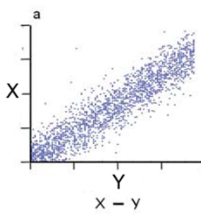
Pe'er D. Bayesian Network Analysis of Signaling Networks: A Primer. Science STKE. April 2005.

Causal Networks

- Simple example



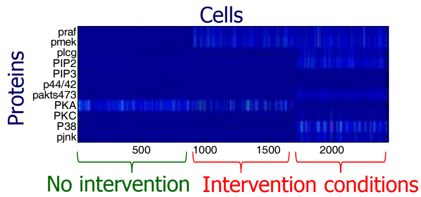
  - Given D={<0,0>,<0,1>,<0,0>,…,<1,1>}, we can compute the Bayesian scores of both graphs → very similar! (e.g. -6.46, -6.78)



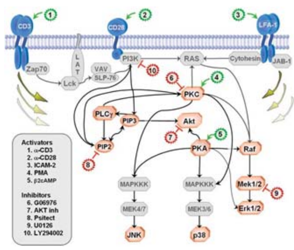
- Data **D’** containing “intervention” samples can improve the accurate inference of causal relationships
  - **D’** = { (0,0), (0,1), [do(X=1),1], (0,0), [do(X=0),0], [do(X=0),0], (1,1), …, (1,1) }

Observation vs Intervention Data

- Training data **D'**



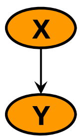
- In each intervention condition, add a chemical known to inhibit/ activate a certain protein



- How do we treat the data from intervention conditions?

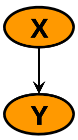
Intervention Modeling

- Let's say that the "real" network is:

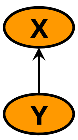


- Observational / interventional data would be like:
  - D'={ (0,0), (1,1), (0,0), (0,0), (1,1), (0,1), (1,1), (1,0), (0,0)… [do(X=0),0], [do(X=1),1], [do(X=0),0], [do(X=0),0], [do(X=1),1],… [1, do(Y=0)], [1, do(Y=1)], [0,do(Y=0)], [0,do(Y=1)], [0,do(Y=1)],… }

- How should the scores be computed in each case?
  - Let's consider ML score



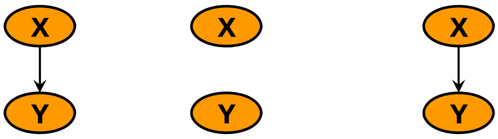
G_1



G_2

Intervention Modeling

- **D'**={ (0,0), (1,1), (0,0), (0,0), (1,1), (0,1), (1,1), (1,0), (0,0)…
  [do(X=0),0], [do(X=1),1], [do(X=0),0], [do(X=0),0], [do(X=1),1],…
  [1, do(Y=0)], [1, do(Y=1)], [0,do(Y=0)], [0,do(Y=1)], [0,do(Y=1)],…}
- When computing the structure score, the training samples in D are assumed to come from the same underlying network. However, do(X=0/1) means that X is forced to have value 0/1 and does not depend on it's parents.
  - We should treat the "intervention" samples differently account when computing the score.

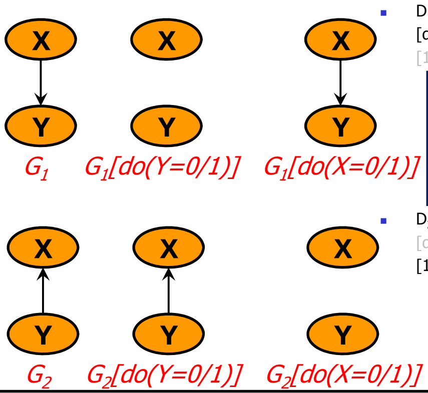


G_1 $G_1[do(Y=0/1)]$ $G_1[do(X=0/1)]$

  - When estimating MLE for Y's CPD, ignore the samples with [do(Y=0/1)]

Intervention Modeling

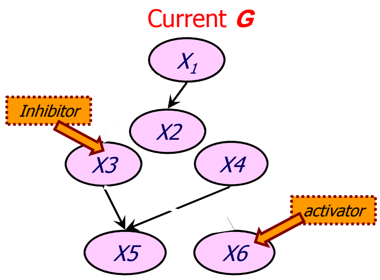
- **D'**={ (0,0), (1,1), (0,0), (0,0), (1,1), (0,1), (1,1), (1,0), (0,0)…
  [do(X=0),0], [do(X=1),1], [do(X=0),0], [do(X=0),0], [do(X=1),1],…
  [1, do(Y=0)], [1, do(Y=1)], [0,do(Y=0)], [0,do(Y=1)], [0,do(Y=1)],…}
- ML scores of G_1 & G_2 :
  - Which of G_1 and G_2 is likely to have a higher score?

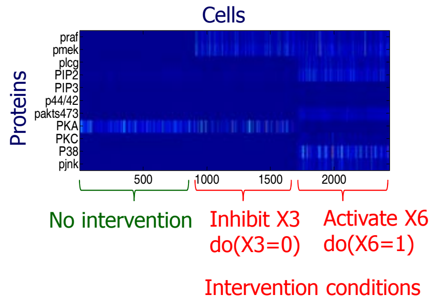


- D_1'={ (0,0), (1,1), (0,0), (0,0), (1,1), (0,1), (1,1), (1,0), (0,0)…
  [do(X=0),0], [do(X=1),1], [do(X=0),0], [do(X=0),0], [do(X=1),1],…
  [1, do(Y=0)], [1, do(Y=1)], [0,do(Y=0)], [0,do(Y=1)], [0,do(Y=1)],…}
- Dependency between X and Y can be modeled by G_1 better.
  - Intervention samples help!
- D_2'={ (0,0), (1,1), (0,0), (0,0), (1,1), (0,1), (1,1), (1,0), (0,0)…
  [do(X=0),0], [do(X=1),1], [do(X=0),0], [do(X=0),0], [do(X=1),1],…
  [1, do(Y=0)], [1, do(Y=1)], [0,do(Y=0)], [0,do(Y=1)], [0,do(Y=1)],…}

Intervention Modeling

- Compute the score with intervention data*
  - Modified FamScore S_i
  - Ignore counts for the intervened node

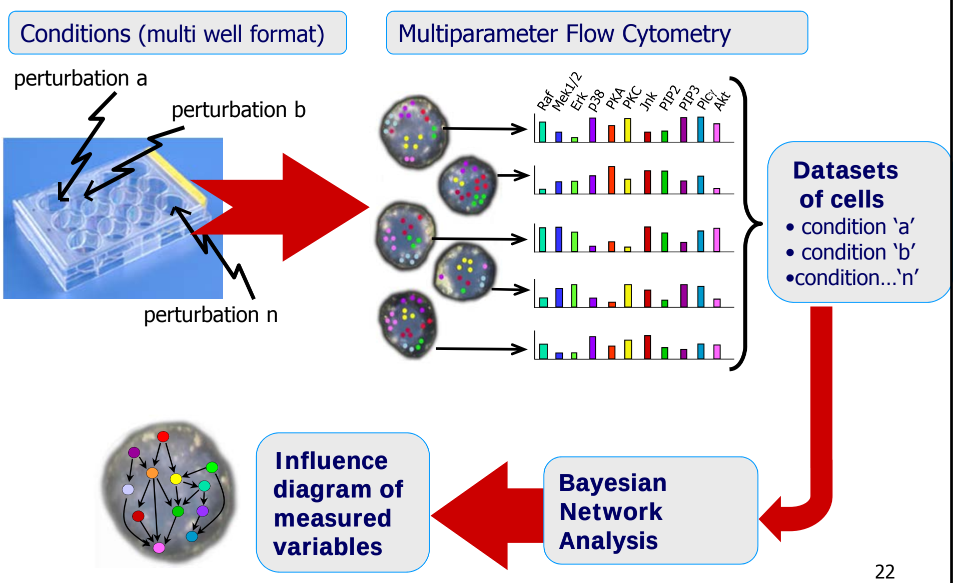




21

*G. Cooper E. Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning, 9, 309-347. 1992.

Overview

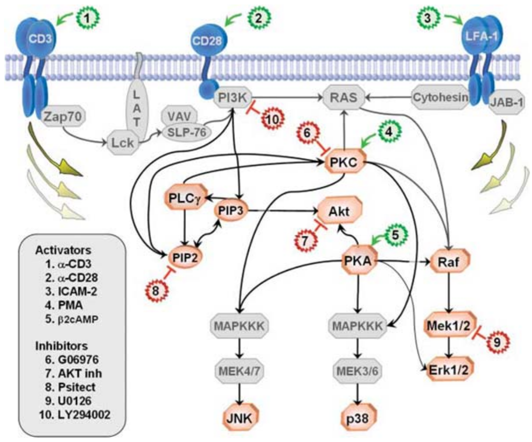


22

Source: Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

Signaling Networks – Example

- Classic signaling network and points of intervention
- Human T cell (white blood cell)

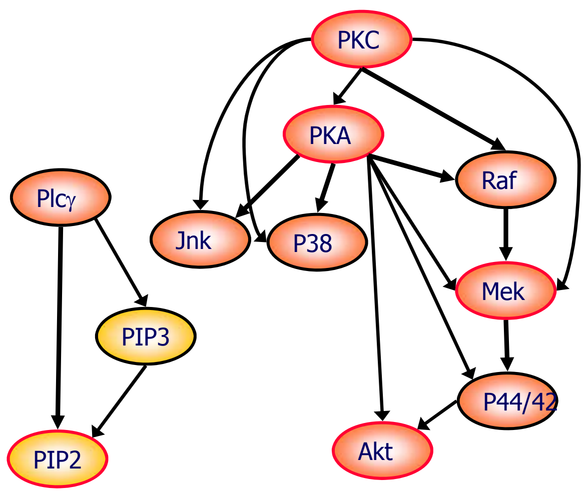


= Measured proteins

- Intervention conditions
  - Stimuli:
    - anti-CD3, anti-CD28, ICAM-2
  - Inhibitors to:
    - Akt, PKC, PIP3, Mek
  - Activators of:
    - PKC, PKA

Source: Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

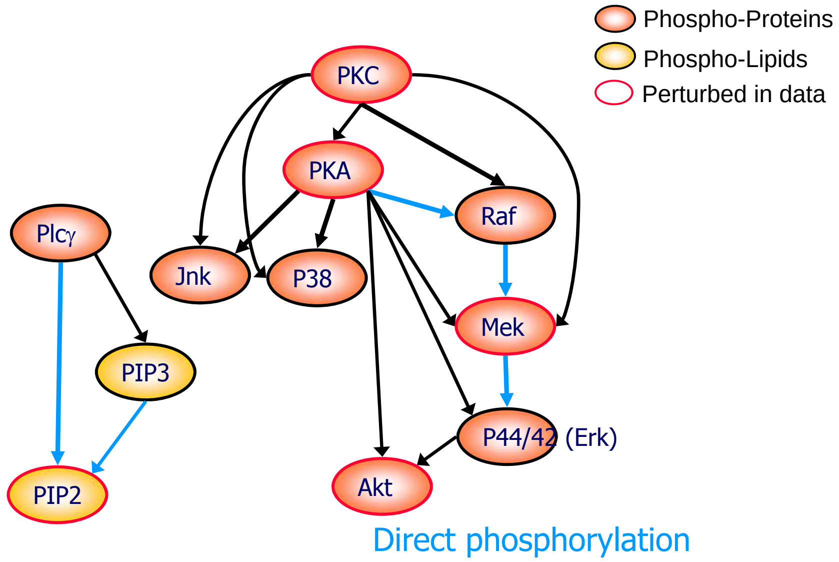
Inferred Network



Phospho-Proteins
Phospho-Lipids
Perturbed in data

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

Evaluation I

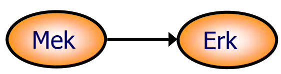


25

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

Features of Approach

- Direct phosphorylation:

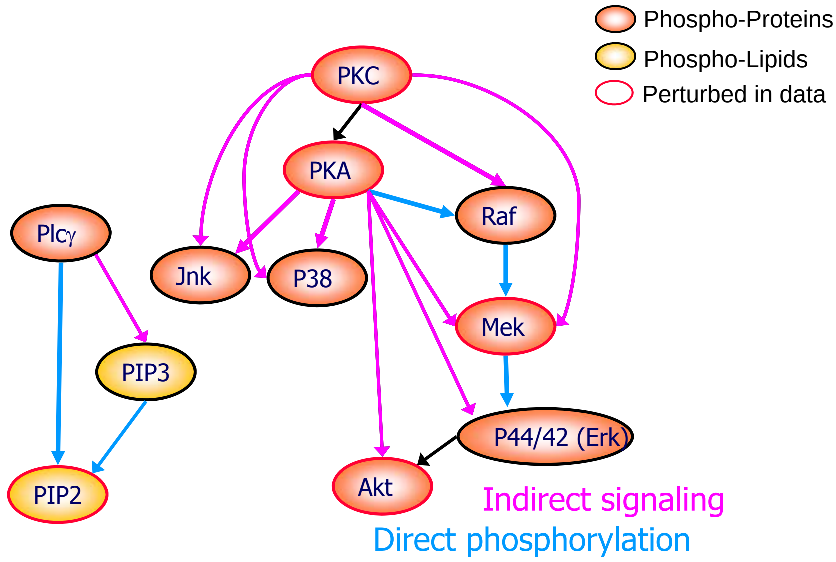


Difficult to detect using other forms of high-throughput data:
- Protein-protein interaction data
- Microarrays

26

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

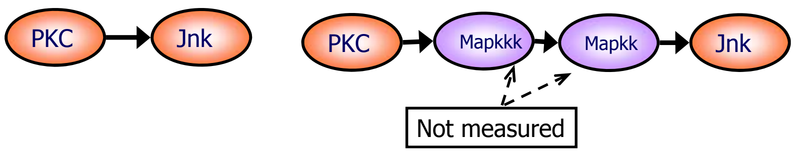
Evaluation II



Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

Features of Approach

- Indirect signaling



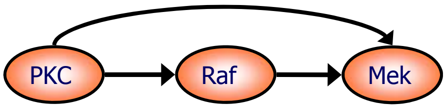
Indirect connections can be found even when the intermediate molecule(s) are not measured

- Explaining away

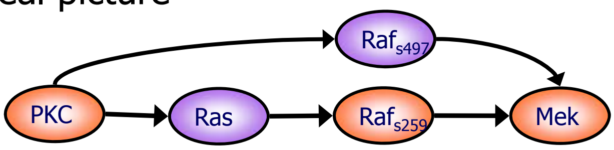
Unnecessary edges do NOT appear

Indirect signaling - Complex example

- Is this a mistake?

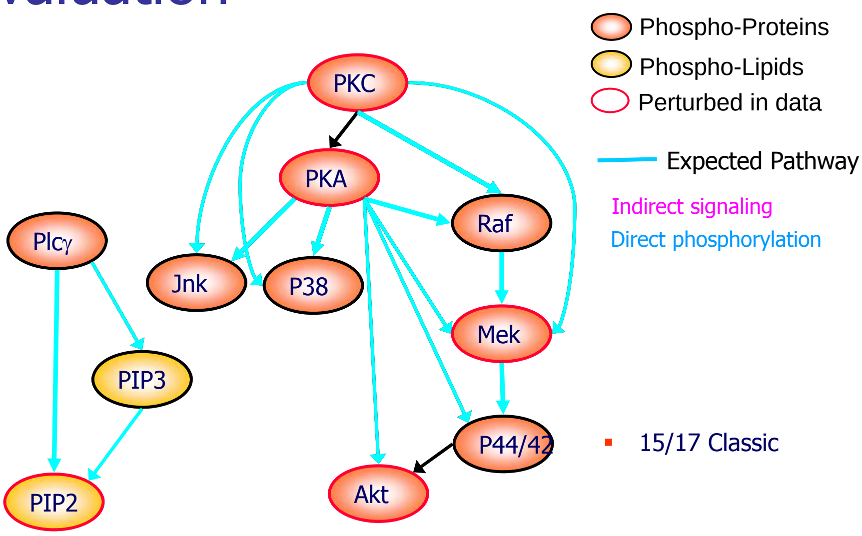


- The real picture



- Phoso-protein specific
- More than one pathway of influence

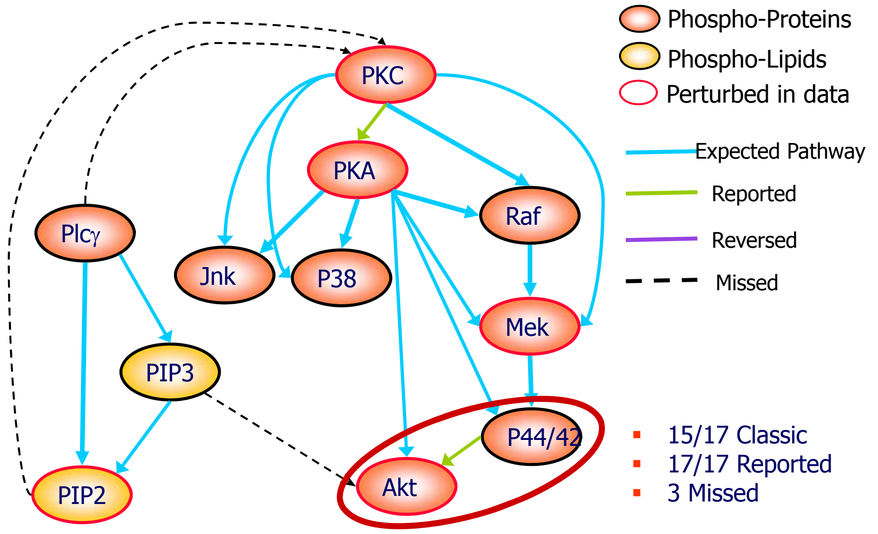
Evaluation



- 15/17 Classic

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Sachs et al. Science (2005).

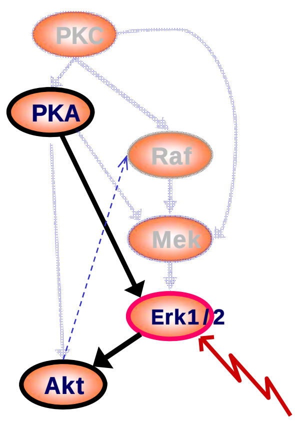
Evaluation



- 15/17 Classic
- 17/17 Reported
- 3 Missed

31

Prediction



Erk1/2 unperturbed

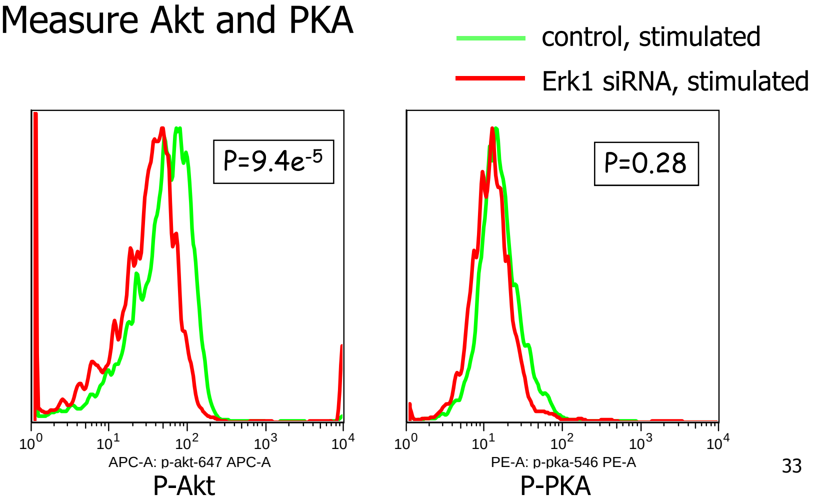
- Erk influence on Akt previously reported in colon cancer cell lines

Predictions:

- Erk1/2 influences Akt
- While correlated, Erk1/2 does not influence PKA

Validation

- SiRNA on Erk1/Erk2
- Select transfected cells
- Measure Akt and PKA



33

Conclusions

- Many limitations
  - Interventions
  - Flow cytometry (4-12 proteins, no time series data)
  - Bayesian networks (no feedback loops)
- Advantages
  - In vivo measurement
  - No a priori knowledge
  - Enablers of accurate inference
    - Network intervention
    - Sufficient numbers of single cells

34

What We've Covered So Far

Lecture notes

Lecture 1: Course logistics, short intro to molecular biology, example project topics [PPT] [PDF]

Lecture 2: Introduction to Bayesian networks for computational biology [PPT] [PDF]
Lecture 3: Maximum Likelihood Estimation, Expectation Maximization [PPT] [PDF]

ML basics (Bayesian networks, MLE, EM)

Lecture 4: Genetic basics, QTL mapping, Association studies [PPT] [PDF]
Lecture 5: QTL mapping, haplotypes [PPT] [PDF]
Lecture 6: Haplotype reconstruction [PPT] [PDF]
Lecture 7: Disease association studies [PPT] [PDF]
Lecture 8: Linkage analysis [PPT] [PDF]

Genetics (association studies, phasing, linkage analysis)

Lecture 9: Inferring transcriptional regulatory networks I [PDF] [PPT]
Lecture 10: Inferring transcriptional regulatory networks II [PDF] [PPT]
Lecture 11: Advanced topics in inferring regulatory networks [PDF]
Lecture 12: Regulatory motif finding I [PDF]
Lecture 13: Regulatory motif finding II [PDF]
Lecture 14: Inferring the signaling networks I [PDF]
Lecture 15: Inferring the signaling networks II [PDF]

Systems biology (gene regulation, gene interaction)

- What next?

35

Sequence Analysis (5 lectures)

- Sequencing techniques
- Sequence alignment
- Comparative genomics

36