

# Bayesian Networks

Malachy Campbell

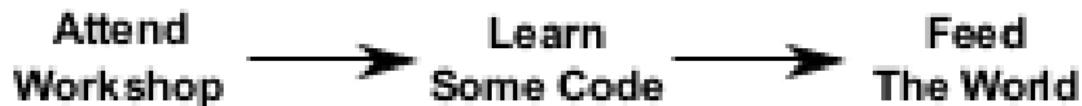
11/15/2018

## Introduction

- ▶ A Bayesian network (BN) is a graph that depicts the probabilistic dependancies between random variables

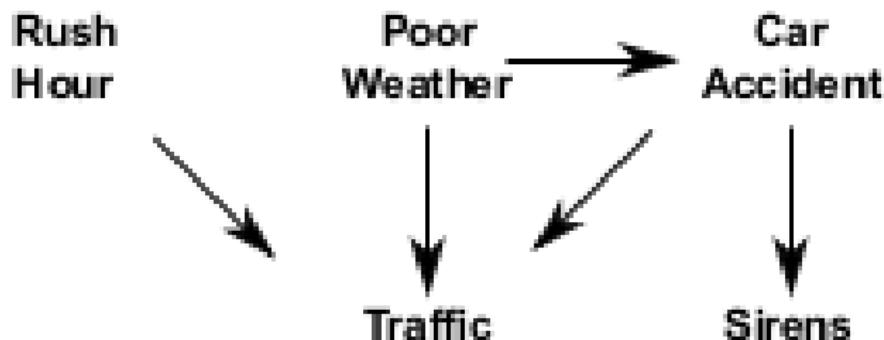
## Introduction

- ▶ A Bayesian network (BN) is a graph that depicts the probabilistic dependancies between random variables



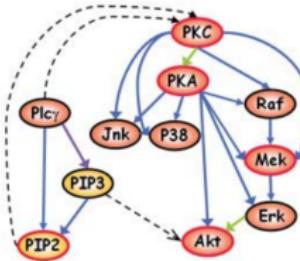
## Introduction

- ▶ A Bayesian network (BN) is a graph that depicts the probabilistic dependancies between random variables



# Introduction

## A Model inference result



● Phospho-Proteins  
● Phospho-Lipids  
● Perturbed in data

Expected	15/17
Reported	17/17
Reversed	1
Missed	3

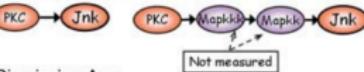
## B Features of approach

- a • Direct phosphorylation:



● Phospho-Proteins  
● Phospho-Lipids  
● Not measured

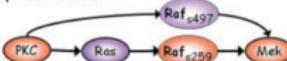
- b • Indirect signaling



- c • Dismissing Arcs



- d • Site-specific Arcs



## Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data

Karen Sachs,<sup>1,\*</sup> Omri Perez,<sup>2,\*</sup> Dana Pe'er,<sup>2,\*</sup>

Douglas A. Lauffenburger,<sup>1</sup> Garry P. Nolan<sup>1</sup>

Machine learning was applied for the automated derivation of causal influences in cellular signaling networks. This derivation relied on the simultaneous measurement of multiple perturbations and their influences on the state of thousands of individual primary human immune system cells. Perturbing these cells with molecular interventions drove the ordering of connection between pathways (represented when Bayesian networks inferred causal relationships) dramatically altered most of the traditionally reported signaling relationships and predicted novel interpathway network causalities, which we verified experimentally. Reconstruction of network models from physiologically relevant primary single cells might be applied to understanding native-state tissue signaling biology, complex drug actions, and dysfunctional signaling in disease cells.

each other (that is, the dependence of each biomolecule in the pathway on other biomolecules) in the form of an influence diagram. These models can be automatically derived from large-scale multiparameter perturbations of individual cells using a newly found computational procedure termed network inference. Although the relationships are statistical in nature, they can sometimes be interpreted as causal when new interventional data are used; for example, with the use of kinase-specific inhibitors (6, 7).

The authors developed a type of Bayesian networks for the inference of signaling pathways from biological data sets. Bayesian networks can represent complex states of biological systems involving multiple interacting molecules, and their probabilistic nature can accommodate noise that is inherent in biological systems. They can describe direct molecular interactions as well as indirect influences that proceed through additional unobserved components, a property

# Introduction

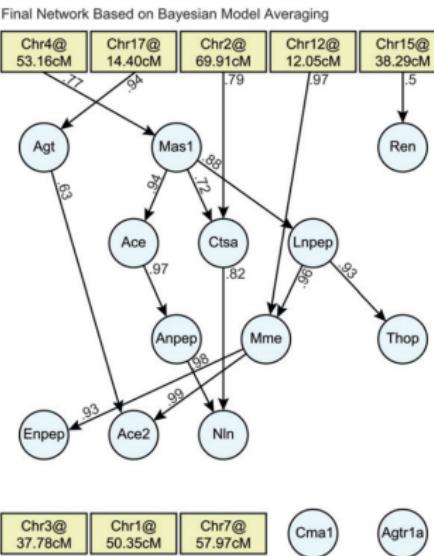


FIGURE 5.—A graphical representation of the final RAS network based on BMA. Edges were drawn if their probability exceeded 0.5.

Copyright © 2011 by the Genetics Society of America  
DOI: 10.1534/genetics.110.129278

## A Bayesian Framework for Inference of the Genotype–Phenotype Map for Segregating Populations

Rachael S. Hageman,\* Magalie S. Leduc,\* Ron Korstanje,\* Beverly Paigen\*  
and Gary A. Churchill<sup>\*,†</sup>

\*The Jackson Laboratory, Bar Harbor, Maine 04609 and <sup>†</sup>Southwest Foundation for Biomedical Research, San Antonio, Texas 78227  
Manuscript received September 16, 2010  
Accepted for publication January 2, 2011

# Introduction

- ▶ Plant breeding/genetics example: Given a dataset of many (10+) agronomic traits (plant height, nitrogen use efficiency, grain protein, grain carbohydrates/starch, leaf area index, radiation use efficiency, yield), how does selection for one (e.g. grain protein) impact another (e.g. grain yield)?
  - ▶ More on this later...

## (Very basic) Theory for BNs

- ▶ BNs represent a probability distribution as a probabilistic directed acyclic graph (DAG)

## (Very basic) Theory for BNs

- ▶ A graph ( $G = (V, A)$ ) is composed of nodes ( $V$ ) and vertices/arcs ( $A$ )
  - ▶ Nodes ( $u, v$ ): variables or traits
  - ▶ Arcs ( $a = (u, v)$ ): describe the relationships between nodes  $u$  and  $v$ 
    - ▶ If  $u$  and  $v$  are ordered then the arc is directed,  $u$  and  $v$  unordered then the arc is undirected.
- ▶ Undirected graphs: no ordering between nodes; Directed graphs: all arcs are ordered; Partially directed graphs: some arcs are ordered
- ▶ Paths: sequences of arcs connecting two nodes; passes through each arc only once and all arcs follow the same direction

## (Very basic) Theory for BNs

- ▶ **Chain rule:** Joint probability distribution of  $i$  random variables  $(x_1, x_2, x_3, x_4)$  is

$$P(x_1, x_2, x_3, x_4) = P(x_4|x_1, x_2, x_3)P(x_3|x_1, x_2)P(x_2|x_1)P(x_1)$$

- ▶ So, we can take a joint probability distribution and break it down into a bunch of conditional probabilities
- ▶ **Conditional independence:** Given some knowledge of  $C$ , provides no information on  $B$  occurring or  $A$  occurring

$$P(A, B|C) = P(A|C)P(B|C)$$

## (Very basic) Theory for BNs

- BN are a class of graphical models that represent the probabilistic dependencies between a set of random variables as a directed acyclic graph (no loops!). We can scale these concepts to a graphical model.

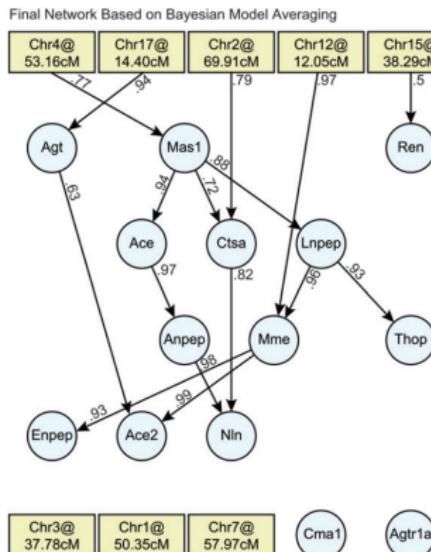
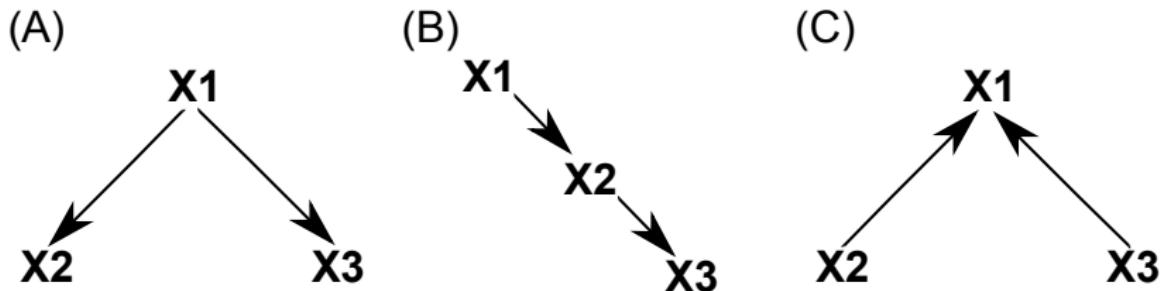


FIGURE 5.—A graphical representation of the final RAS network based on BMA. Edges were drawn if their probability exceeded 0.5.

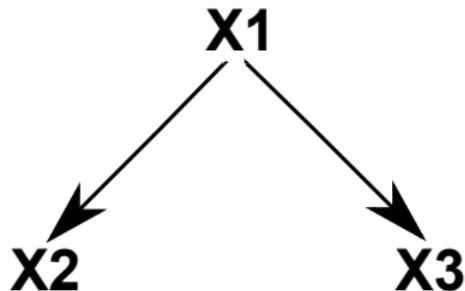
## (Very basic) Theory for BNs

- ▶ For a **divergent structure** (A) the joint distribution is  
 $P(X_1, X_2, X_3) = P(X_2|X_1)P(X_3|X_1)P(X_1)$ ; **serial** structure  
(B) the joint distribution is  
 $P(X_1, X_2, X_3) = P(X_3|X_2)P(X_2|X_1)P(X_1)$ ; **convergent** structure  
(C) the joint distribution is  
 $P(X_1, X_2, X_3) = P(X_2|X_1, X_3)P(X_1)P(X_3)$



## Bayesian Networks (BN)

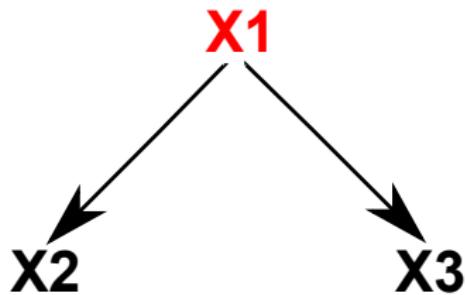
- ▶ **Example 1:** If we have no information for the divergent structure, are  $X_2$  and  $X_3$  independent?



$$P(X_2, X_3) = \sum_{X_1} P(X_2|X_1)P(X_3|X_1)P(X_1)$$

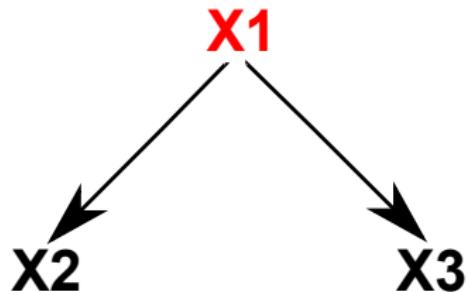
## Bayesian Networks (BN)

- ▶ **Example 2:** If we have information on X1 for the divergent structure, are X2 and X3 independent?



## Bayesian Networks (BN)

- ▶ **Example 2:** If we have information on  $X_1$  for the divergent structure, are  $X_2$  and  $X_3$  independent?

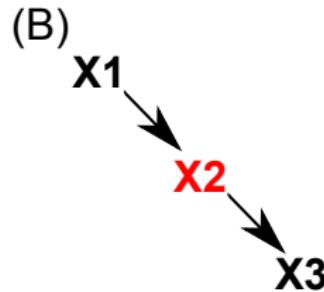
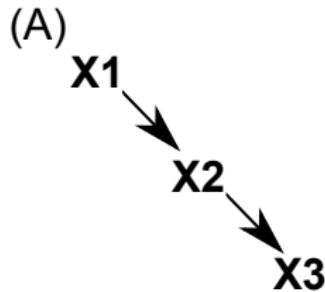


$$P(X_2, X_3 | X_1) = P(X_2 | X_1)P(X_3 | X_1)$$

# Bayesian Networks (BN)

## ► Example 3:

1. If we don't have information on X2 for the serial structure (A), are X1 and X3 independent?
2. If we do have information on X2 for the serial structure (B), are X1 and X3 independent?



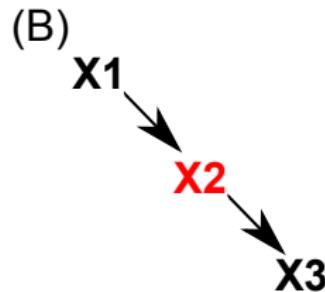
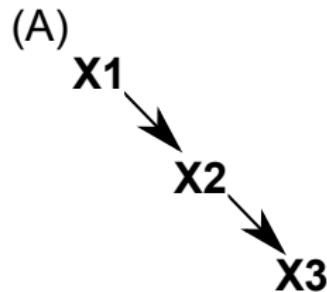
# Bayesian Networks (BN)

## ► Example 3:

1. If we don't have information on  $X_2$  for the serial structure (A), are  $X_1$  and  $X_3$  independent?

$$P(X_1, X_3) = \sum_{X_2} P(X_3|X_2)P(X_2|X_1)P(X_1)$$

2. If we do have information on  $X_2$  for the serial structure (B), are  $X_1$  and  $X_3$  independent?  $P(X_1, X_3|X_2) = P(X_3|X_2)P(X_1|X_2)$

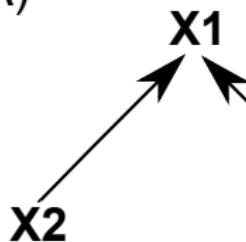


# Bayesian Networks (BN)

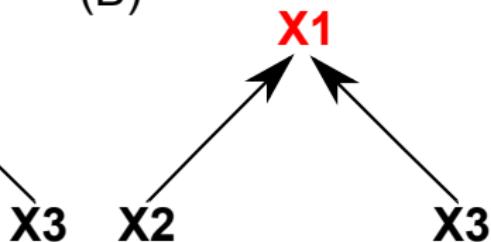
## ► Example 4:

1. If we don't have information on X1 for the convergent structure (A), are X2 and X3 independent?
2. If we do have information on X1 for the convergent structure (B), are X2 and X3 independent?

(A)



(B)



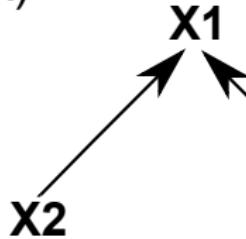
# Bayesian Networks (BN)

## ► Example 4:

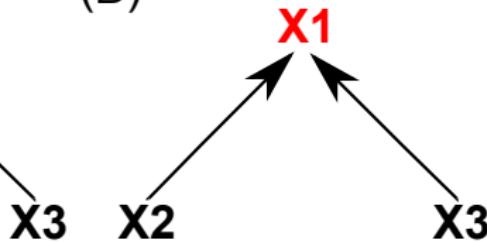
1. If we don't have information on  $X_1$  for the convergent structure (A), are  $X_2$  and  $X_3$  independent?  $P(X_2, X_3) = P(X_2)P(X_3)$
2. If we do have information on  $X_1$  for the convergent structure (B), are  $X_2$  and  $X_3$  independent?

$$P(X_1, X_3 | X_2) = \frac{P(X_3)P(X_1|X_2, X_3)}{P(X_1)}$$

(A)



(B)



## Bayesian Networks (BN): *direct-separation*

- ▶ If we scale this up, each node can be considered a subset of nodes in a DAG
- ▶ ***direct-separation (d-separation)***: Determines whether a set of A variables is independent of another set C, given a third set B. Reveals relationships and makes inference more efficient!
- ▶ *Formal definition:*
  - ▶ Consider only the cases where we have information. Given some information for a group of nodes (**B**), the two subsets of nodes **A** and **C** are conditionally independant if there is a node  $w$  that has (1) converging arrows and  $w$  nor any of its descendants are part of **B** (e.g. does not carry information), or if  $w$  does not have converging arrows and is a part of **B**

## Bayesian Networks (BN): Markov blankets

- ▶ The **Markov blanket** for some node is the node's parents, children, and the any other nodes that share a child with the node.
  - ▶ Nodes in the Markov blanket  $d$ -separate the node from the rest of the network

## Learning Bayesian Networks (BN)

- ▶ Learning the **structure** of the DAG (model selection)
  - ▶ Unsupervised or supervised
- ▶ Learning the **parameters** of the DAG

# Structure Learning Algorithms

- ▶ Three classes:
  - ▶ **Constraint-based:** Learn network structure using conditional independence tests. Based off inductive causation (IC) algorithm proposed by Verma and Pearl (1991)
  - ▶ **Score based:** “Search-and-score” Evaluate candidate network structures using a scoring metric, and use an optimization procedure to identify the highest scoring structure that best fits the data.
  - ▶ **Hybrid:** Use constraint-based methods to **restrict** score-based algorithms to a subset of nodes. Score-based algorithms attempt to find the optimal network for the subset of nodes.

## IC algorithm (Constraint-based)

1. **Identify pairs of variables connected by arc.** Determine if pair  $X$ ,  $Y$ , are independent given a subset of variables  $\mathbf{X}$ . If not, join  $A$  and  $B$  with an undirected arc. Repeat for all pairs.
2. **Set direction of edges (e.g. find v-structures in graph)** in the graph from (1) ( $K$ ) using conditional independence tests. For all triplet structures  $X - Z - Y$ , determine if  $X$  and  $Y$  are independent given  $Z$ . If not, then set as a convergent structure  $X \rightarrow Z \leftarrow Y$
3. **Set direction for remaining arcs** according to these rules (1) if  $A \rightarrow B$ ,  $B$  and  $C$  are adjacent,  $A$  and  $C$  are not adjacent, and there is no directed arc at  $B$ , then set  $B - C$  as  $B \rightarrow C$ ; (2) if  $A$  and  $B$  are adjacent and there is a directed path from  $A$  to  $B$  then set  $A - B$  to  $A \rightarrow B$

## Constraint-based Structure Learning Algorithms

- ▶ Steps (1) and (2) of IC are not feasible because the number of combinations is far too high. Modern solutions only consider the local nodes  $G$

## Constraint-based Structure Learning Algorithms

- ▶ **PC algorithm** (Spirtes et al 2001): Similar to IC, but starts with a full undirected network ( $G$ ), but uses a backward selection process to thin size of  $G$ . Is  $X$  and  $Y$  independent? Is  $X$  and  $Y$  independent given first neighbor? Is  $X$  and  $Y$  independent given first and second neighbor? ...
- ▶ **Grow-Shrink algorithm** (Margaritis, 2003): Start with an empty set of variables ( $S$ ), and a variable ( $X$ ). The first growing phase adds variables to  $S'$  if they are dependent with  $X$  given the variable in  $S$ . In the second shrinking phase, variables that are not in the Markov blanket of  $X$  are tested for dependence given  $S$ .

## Score-based Structure Learning Algorithms

- ▶ “*Search-and-score*”: Evaluate candidate network structures using a scoring metric, use an optimization procedure to identify the highest scoring structure that best fits the data.
  - ▶ Scoring metrics: likelihood, log-likelihood, AIC, BIC, Bayesian Dirichlet equivalent, K2
- ▶ **Hill climbing/tabu search:** Starting with a network structure ( $G_0$ ) (can be empty) compute score ( $Score_0$ ); modify structure ( $G_1$ ) (add, remove, or reverse arc) and recompute score ( $Score_1$ ); if  $Score_1 > Score_0$  then repeat process with  $G_1$  until the score is no longer maximized

## Hybrid Structure Learning Algorithms

- ▶ Use constraint-based methods to **restrict** score-based algorithms to a subset of nodes. Score-based algorithms attempt to find the optimal network for the subset of nodes.
- ▶ *Max-min hill climbing*: Learns the **undirected network** first using the max-min parent children (MMPC) algorithm, and directs the edges using a hill-climbing approach.

## References