Graphical Models

David Heckerman
Microsoft Research

Graphical models: Graphical representations of *conditional independence*

Helps with…

- Understanding
- Combining expert knowledge and data
- Making inference faster
- Learning cause and effect

Overview

- Basic concepts (probability and conditional independence)
- Dependency networks → undirected graphs
- Directed acyclic graphs (“Bayes nets”)
- Learning cause and effect from observational data
  - Application: HIV vaccine design

Probability

<u>$p(X=x|Y=y)$ has at least two meanings:</u>

- Bayesian: The belief of an individual that variable X takes on value x, given that Y=y
- Frequentist: The long run fraction that X=x when Y=y

Doesn't matter for this talk!

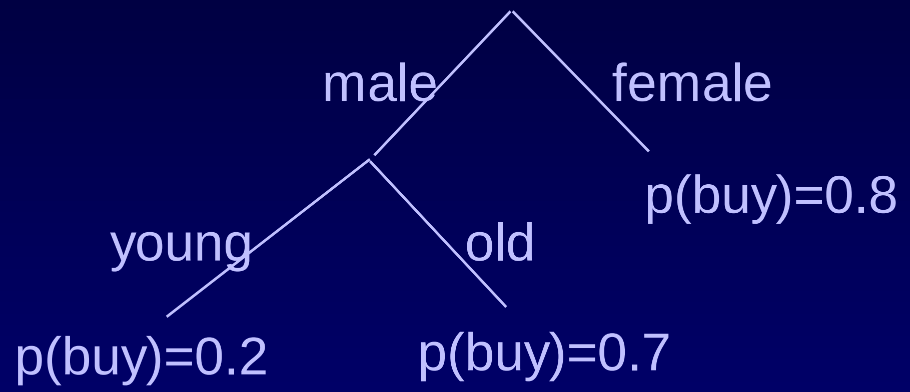
Conditional independence

Equivalent statements for variables X, Y, and Z:

- Y and Z are conditionally independent given X
- $Y \perp Z \mid X$
- $p(y|x,z) = p(y|x)$
- X is just as good a predictor of Y as X and Z
- $p(z|x,y) = p(z|x)$
- X is just as good a predictor of Z as X and Y

Conditional independence

buy	age	sex	month born
y	old	male	jan
y	old	female	mar
n	young	male	nov
n	young	female	july
n	young	female	august
…	…	…	…



p(buy | sex, age, month born) = p(buy | sex, age)

buy ⊥ month born | sex, age

Identifying conditional independence

- From personal belief (Bayesian)
- From data (Bayesian or frequentist)
  - Cross validation (example later)
  - Bayesian methods
  - Penalized likelihood
  - Others?

Overview

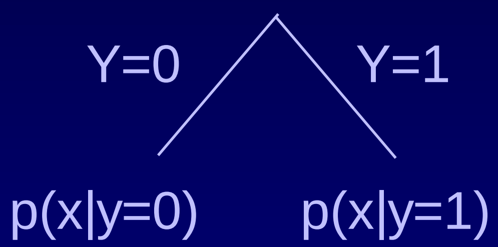
- Basic concepts (probability and conditional independence)
- Dependency networks → undirected graphs
- Directed acyclic graphs (“Bayes nets”)
- Learning cause and effect from observational data
  - Application: HIV vaccine design

Application: Data exploration

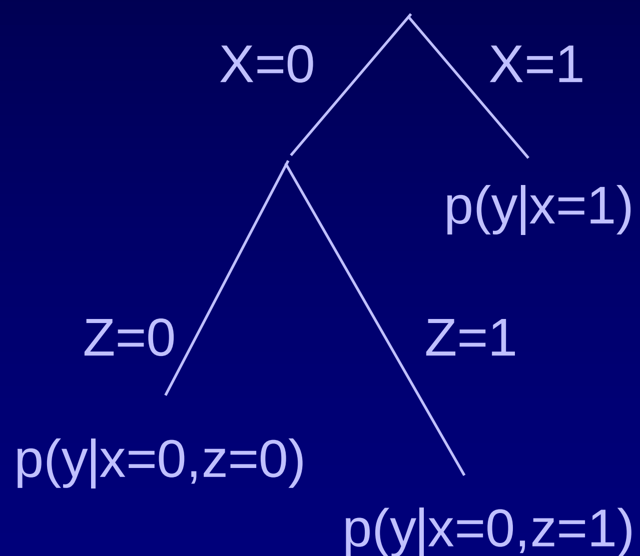
- Suppose you have thousands of variables and you're not sure about the interactions among those variables
- Build a classification/regression model for each variable, using the rest of the variables as inputs

Example with three variables X, Y, and Z

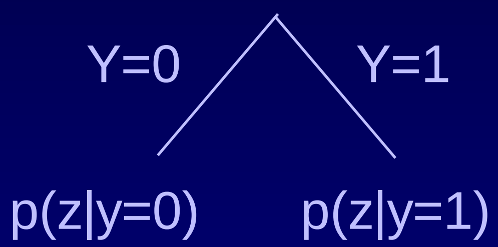
Target: X
Inputs: Y,Z



Target: Y
Inputs: X,Z



Target: Z
Inputs: X,Y

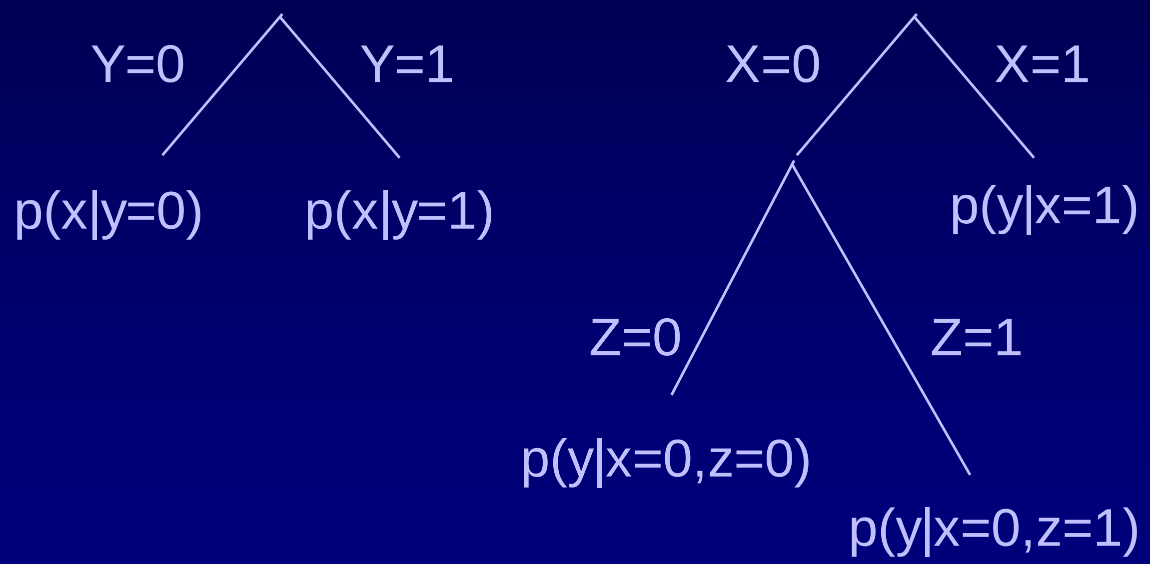


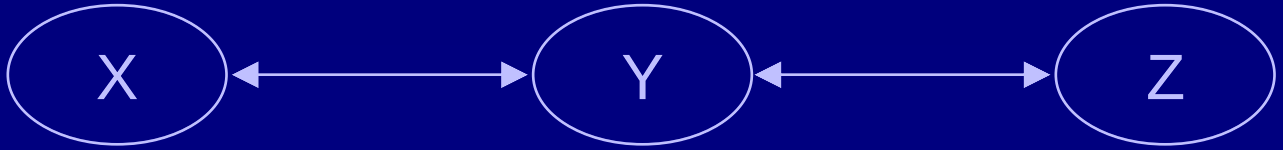
Summarize the trees with a single graph

Target: X
Inputs: Y,Z

Target: Y
Inputs: X,Z

Target: Z
Inputs: X,Y





Dependency Network

- Build a classification/regression model for every variable given the other variables as inputs
- Construct a graph where
  - Nodes correspond to variables
  - There is an arc from X to Y if X helps to predict Y
- The graph along with the individual classification/regression model is a “dependency network”

  (Heckerman, Chickering, Meek, Rounthwaite, Cadie 2000)

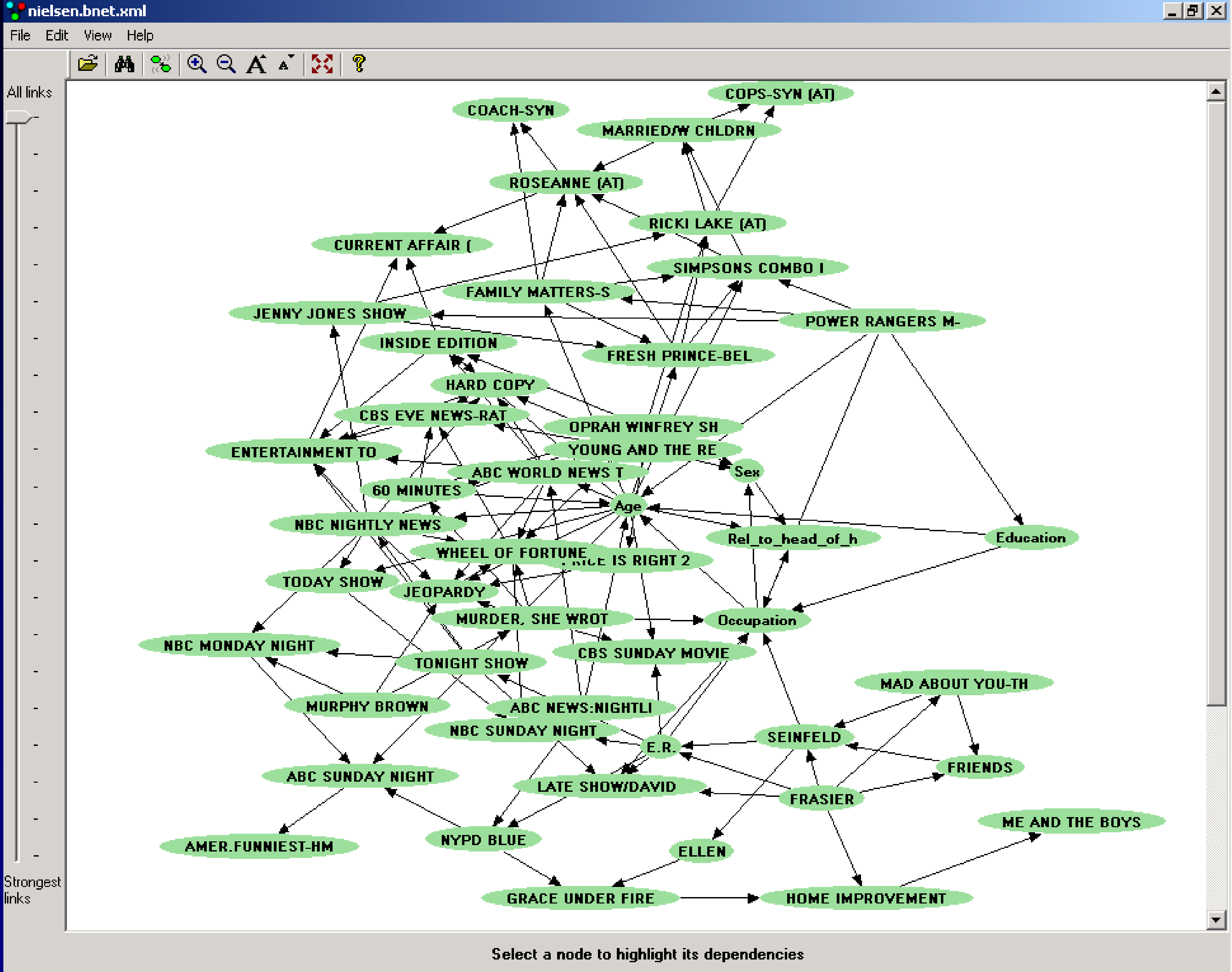
Example: TV viewing

Nielsen data: 2/6/95-2/19/95

	Age	Show1	Show2	Show3	
viewer 1	73	y	n	n	
viewer 2	16	n	y	y	...
viewer 3	35	n	n	n	
	etc.				

~400 shows, ~3000 viewers

Goal: exploratory data analysis (acausal)


nielsen.bnet.xml
File Edit View Help
All links
Strongest links
COPS-SYN (AT)
COACH-SYN
MARRIED/W CHLDRN
ROSEANNE (AT)
RICKI LAKE (AT)
CURRENT AFFAIR (
SIMPSONS COMBO I
FAMILY MATTERS-S
JENNY JONES SHOW
POWER RANGERS M-
INSIDE EDITION
FRESH PRINCE-BEL
HARD COPY
CBS EVE NEWS-RAT
OPRAH WINFREY SH
ENTERTAINMENT TO
YOUNG AND THE RE
Sex
ABC WORLD NEWS T
60 MINUTES
Age
NBC NIGHTLY NEWS
Rel_to_head_of_h
Education
WHEEL OF FORTUNE
IS RIGHT 2
TODAY SHOW
JEOPARDY
MURDER, SHE WROT
Occupation
NBC MONDAY NIGHT
TONIGHT SHOW
CBS SUNDAY MOVIE
MAD ABOUT YOU-TH
MURPHY BROWN
ABC NEWS:NIGHTLI
NBC SUNDAY NIGHT
E.R.
SEINFELD
FRIENDS
ABC SUNDAY NIGHT
LATE SHOW/DAVID
FRASIER
ME AND THE BOYS
AMER.FUNNIEST-HM
NYPD BLUE
ELLEN
GRACE UNDER FIRE
HOME IMPROVEMENT
Select a node to highlight its dependencies

A bit of history

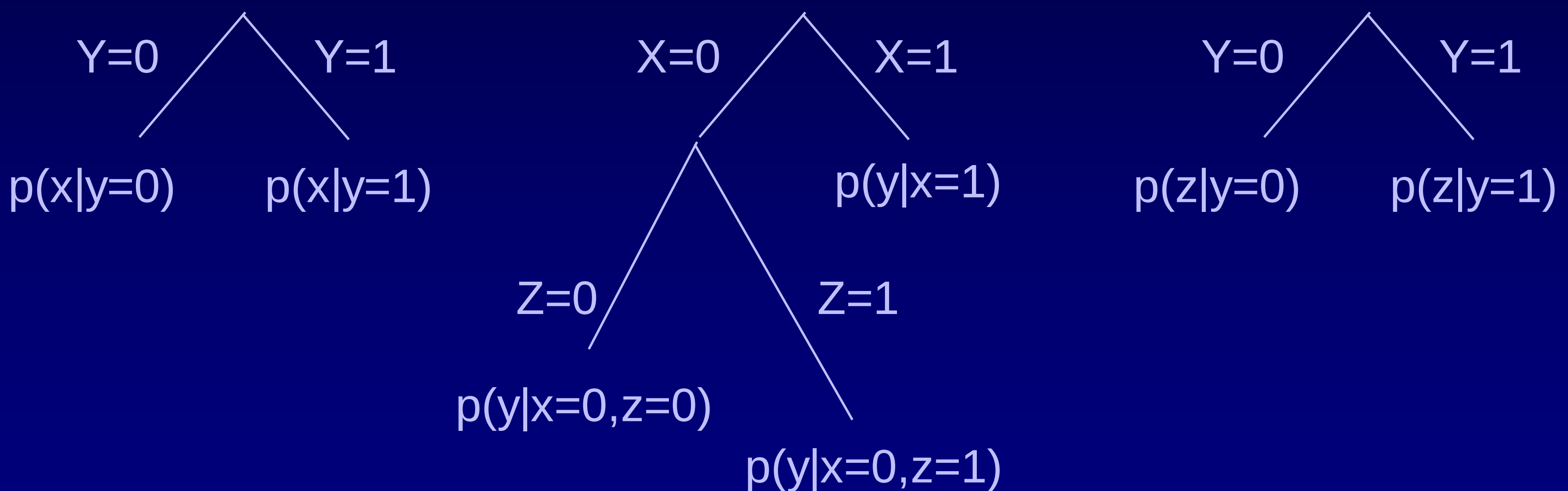
- Julian Besag (and others) invented dependency networks (under another name) in the mid 1970s
- But they didn't like them, because they could be inconsistent

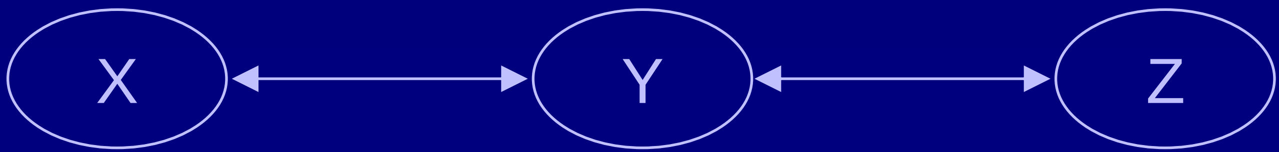
A consistent dependency network

Target: X
Inputs: Y,Z

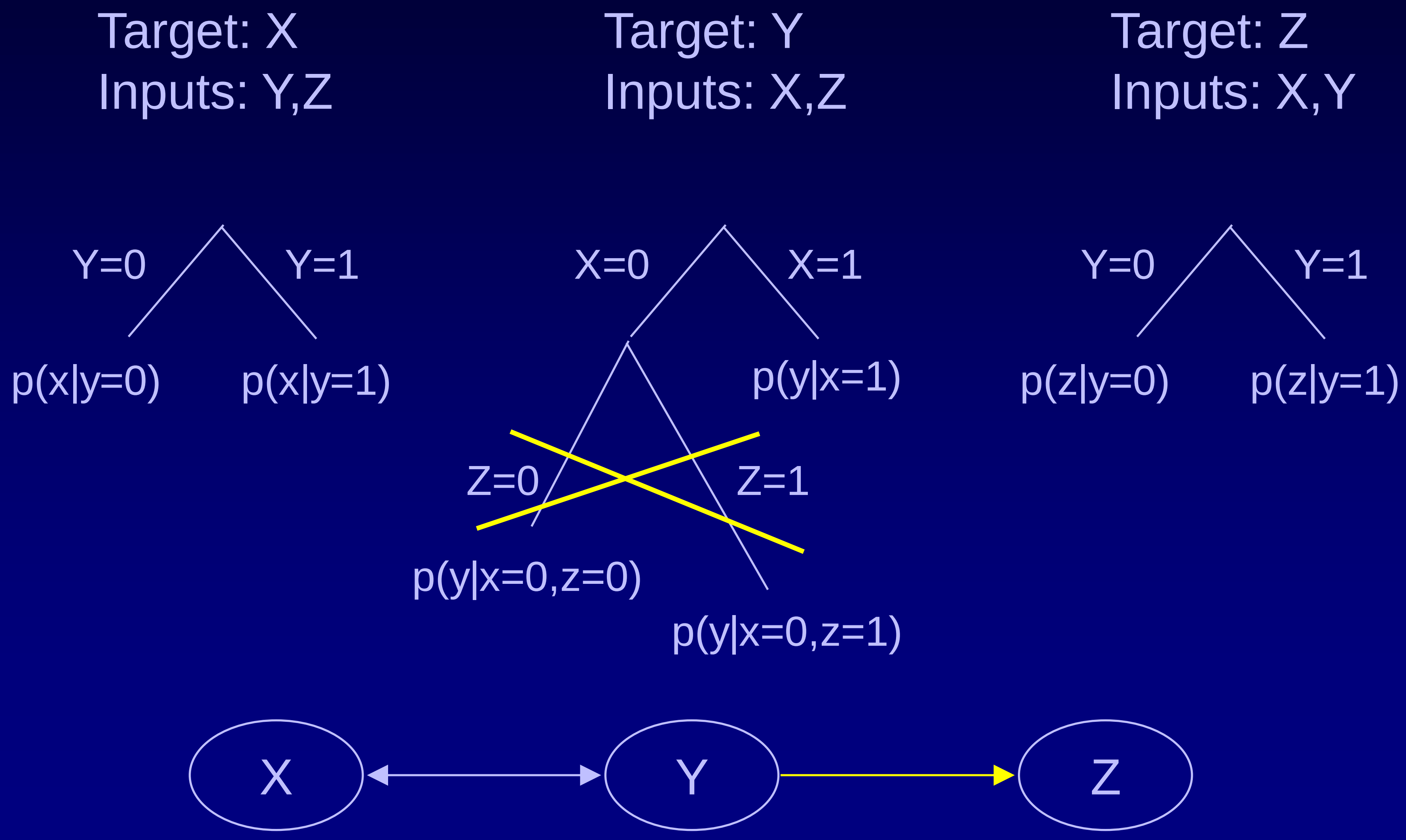
Target: Y
Inputs: X,Z

Target: Z
Inputs: X,Y





An inconsistent dependency network



A bit of history

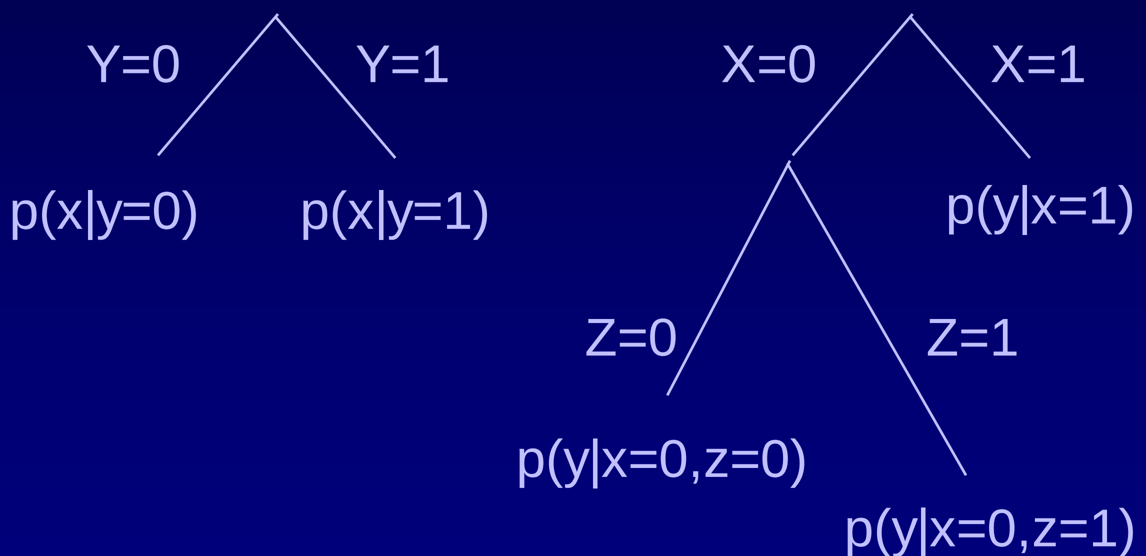
- Julian Besag (and others) invented dependency networks (under the name "Markov graphs") in the mid 1970s
- But they didn't like them, because they could be inconsistent
- So they used a property of consistent dependency networks to develop a new characterization of them

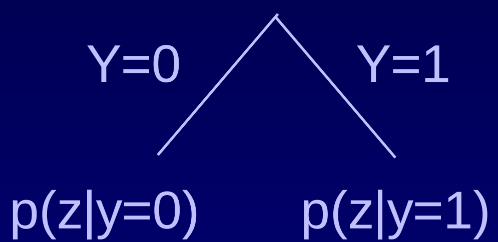
Conditional independence

Target: X
Inputs: Y,Z

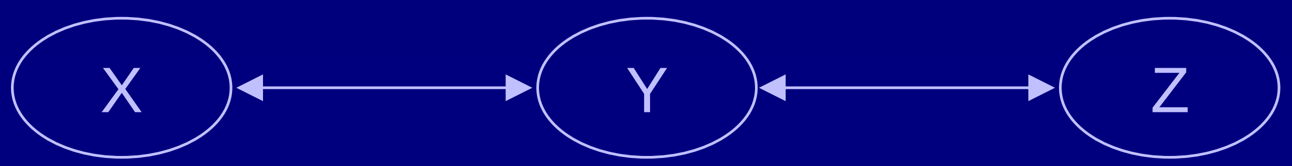
Target: Y
Inputs: X,Z

Target: Z
Inputs: X,Y





$$X \perp Z \mid Y$$



Conditional independence in a consistent dependency network

Each variable is independent of all other variables given its immediate neighbors

Hammersley-Clifford Theorem (Besag 1974)

- Given a set of variables which has a positive joint distribution
- Where each variable is independent of all other variables given its immediate neighbors in some graph G
- It follows that

“clique potentials”

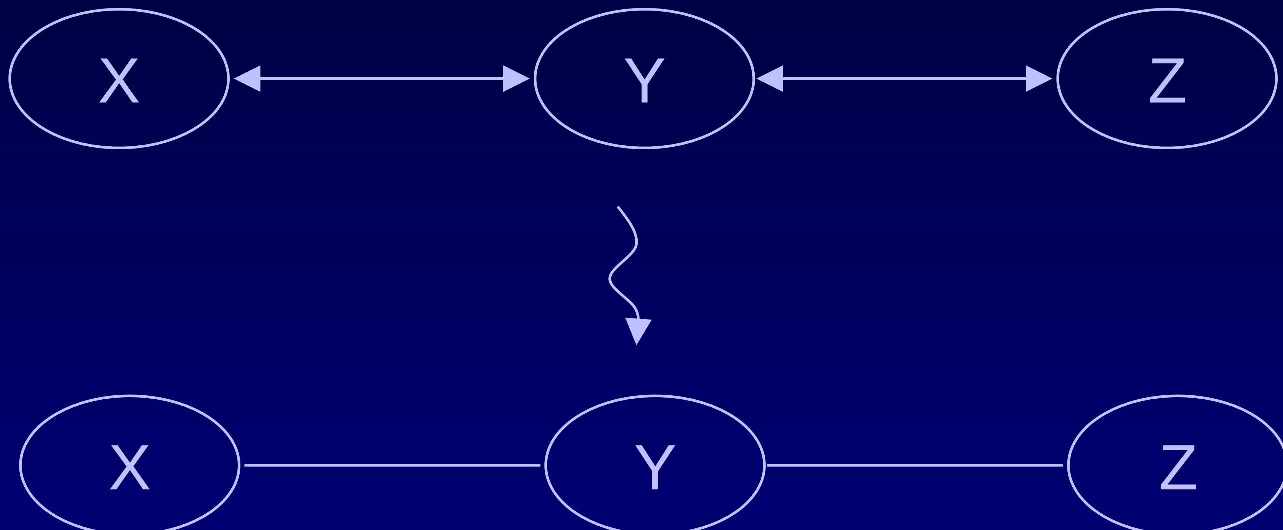
$$p(\mathbf{x}) = \prod_{i=1}^{N} f_i\ (\mathbf{c}_i)$$

where c_1, c_2, …, c_n are the maximal cliques in the graph G.

Example

$$p(x,y,z) = f_1(x,y)\ f_2(y,z)$$

Consistent dependency networks: Directed arcs not needed



$$p(x,y,z) = f_1(x,y)\ f_2(y,z)$$

A bit of history

- Julian Besag (and others) invented dependency networks (under the name "Markov graphs") in the mid 1970s
- But they didn't like them, because they could be inconsistent
- So they used a property of consistent dependency networks to develop a new characterization of them
- "Markov Random Fields" aka "undirected graphs" were born

Inconsistent dependency networks aren't that bad

- They are *almost consistent* because each classification/regression model is learned from the same data set (can be formalized)
- They are easy to learn from data (build separate classification/regression model for each variable)
- Conditional distributions (e.g., trees) are easier to understand than clique potentials

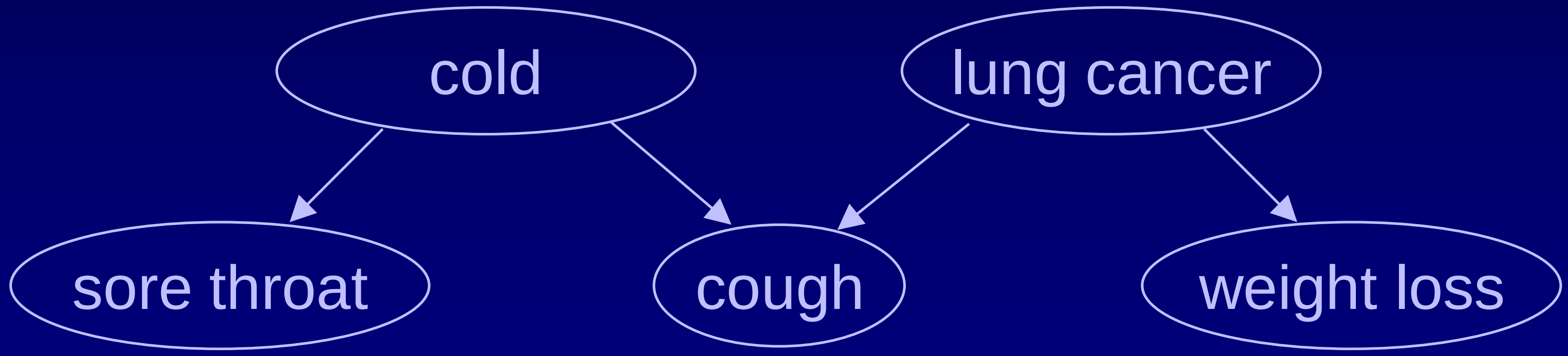
Inconsistent dependency networks aren't that bad

- They are *almost consistent* because each classification/regression model is learned from the same data set (can be formalized)
- They are easy to learn from data (build separate classification/regression model for each variable)
- Conditional distributions (e.g., trees) are easier to understand than clique potentials

- Over the last decade, has proven to be a very useful tool for data exploration

Shortcomings of undirected graphs

- **Lack a generative story (e.g., Lat Dir Alloc)**
- **Lack a causal story**

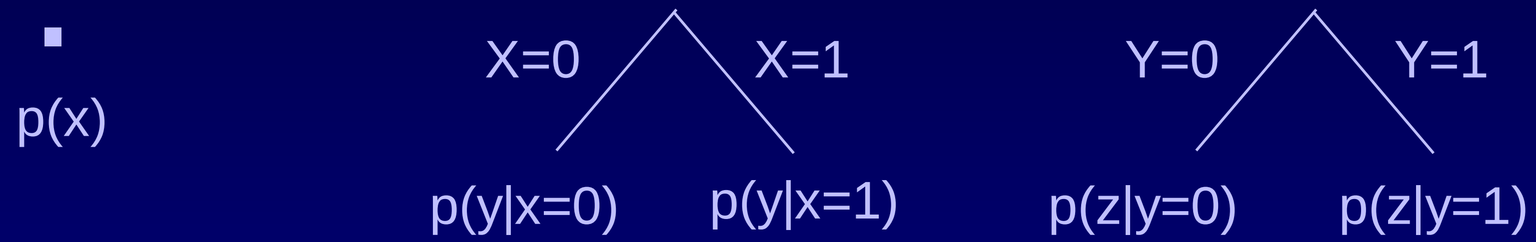


Solution: Build trees in some order

1. Target: X
Inputs: none

2. Target: Y
Inputs: X

3. Target: Z
Inputs: X,Y

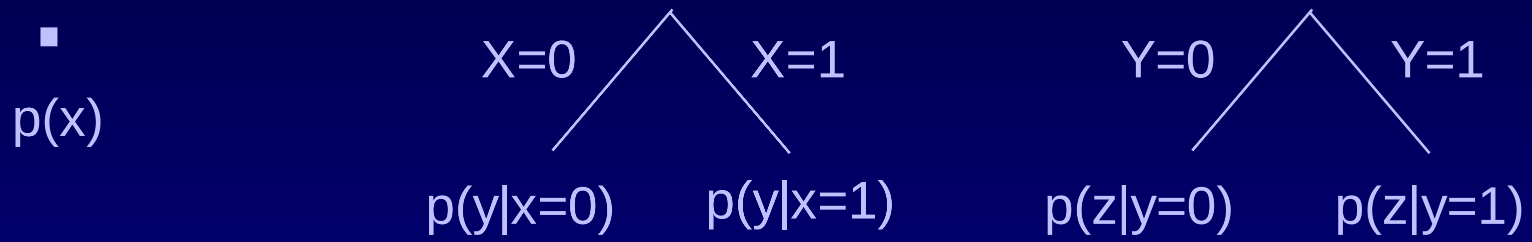


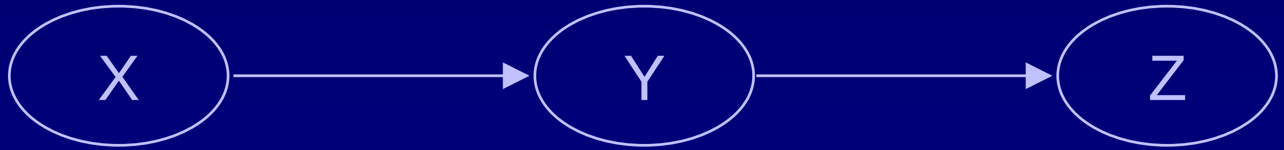
Solution: Build trees in some order

1. Target: X
Inputs: none

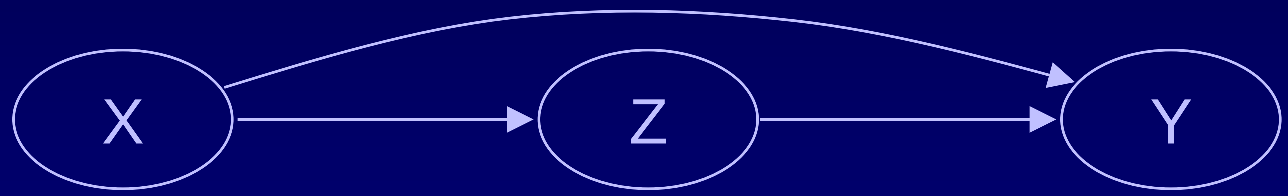
2. Target: Y
Inputs: X

3. Target: Z
Inputs: X,Y





Some orders are better than others



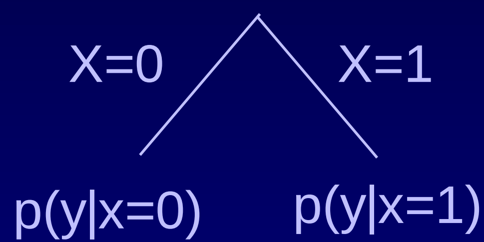
- Prior, often causal knowledge (Bayesian)
- Infer better orderings from data (Bayesian and frequentist)
  - Try random orders
  - Monte-Carlo methods
  - Greedy search

Joint distribution is easy to obtain

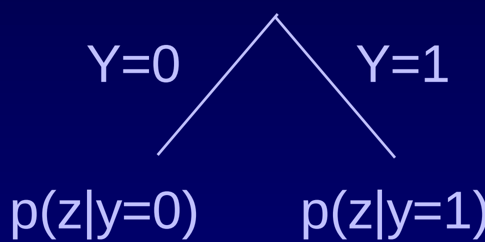
1. Target: X
Inputs: none

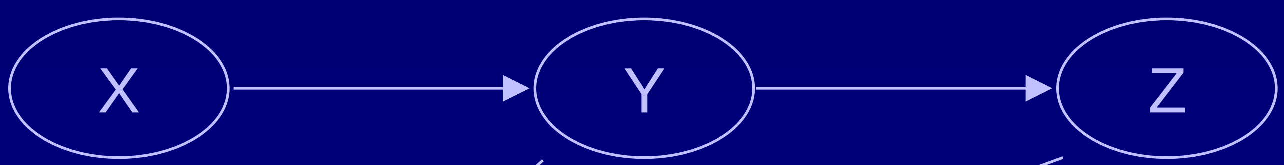
p(x)

2. Target: Y
Inputs: X



3. Target: Z
Inputs: X,Y





$$p(x)\ p(y \mid x)\ p(z \mid x, y) = p(x, y, z)$$

Directed Acyclic Graphical (DAG) models

$$p(x_1,\ldots,x_n) = \prod_i p(x_i \mid x_1,\ldots,x_{i-1})$$
$$= \prod_i p(x_i \mid \mathrm{parents}(x_i))$$

- Many inventors: Wright 1921; Good 1961; Howard & Matheson 1976, Pearl 1982

- Pearl developed them in the context of expert systems (where an individual provided the independencies and probabilities), hence the commonly used term "Bayesian Network"

Graphical models: Graphical representations of *conditional independence*

Helps with…

- Understanding
- Combining expert knowledge and data
- Making inference faster
- Learning cause and effect

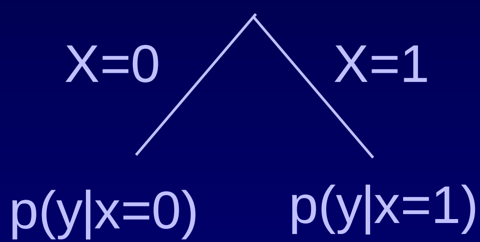
Inference

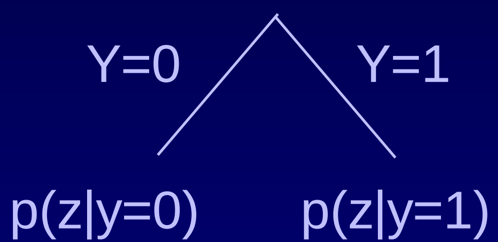
1. Target: X
Inputs: none

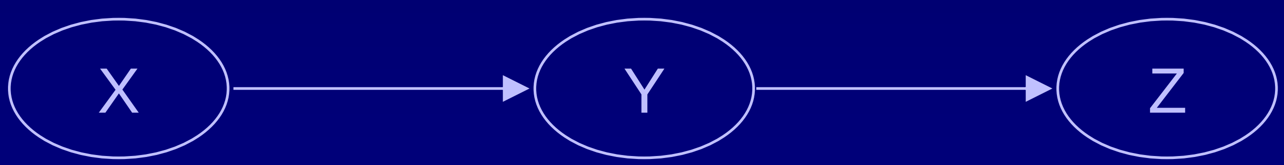
2. Target: Y
Inputs: X

3. Target: Z
Inputs: X,Y

▪
p(x)

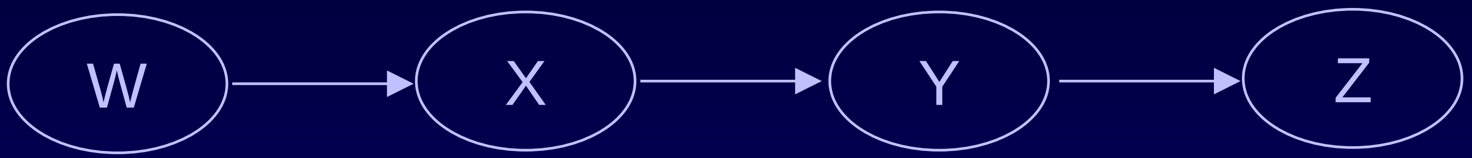






What is p(z|x=1)?

Inference: Example



$$p(z) = \sum_{w,x,y} p(w)\ p(x \mid w)\ p(y \mid x)\ p(z \mid y)$$

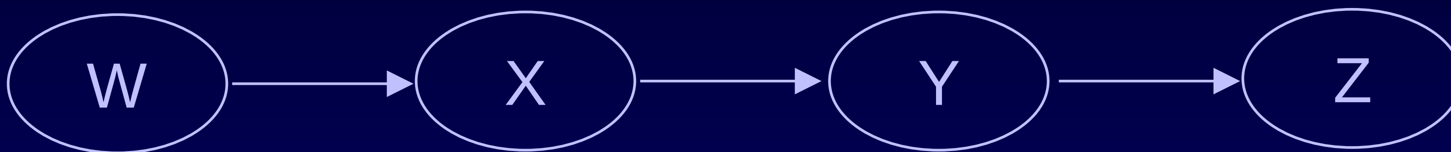
Inference: Example
("Elimination Algorithm")

$$W \rightarrow X \rightarrow Y \rightarrow Z$$

$$p(z) = \sum_{w,x,y} p(w)\ p(x \mid w)\ p(y \mid x)\ p(z \mid y)$$

$$= \sum_{w,x} p(w)\ p(x \mid w) \left(\sum_{y} p(y \mid x)\ p(z \mid y) \right)$$

Inference: Example
(“Elimination Algorithm”)



$$p(z) = \sum_{w,x,y} p(w)\ p(x \mid w)\ p(y \mid x)\ p(z \mid y)$$

$$= \sum_{w,x} p(w)\ p(x \mid w) \left(\sum_{y} p(y \mid x)\ p(z \mid y) \right)$$

$$= \sum_{w} p(w) \left(\sum_{x} p(x \mid w) \left(\sum_{y} p(y \mid x)\ p(z \mid y) \right) \right)$$

Inference

- **Exact methods for inference that exploit conditional independence are well developed (e.g., Shachter, Lauritzen & Spiegelhalter, Dechter)**

- **Exact methods fail when there are many cycles in the graph**
  - **MCMC (e.g., Geman and Geman 1984)**
  - **Loopy propagation (e.g., Murphy et al. 1999)**
  - **Variational methods (e.g., Jordan et al. 1999)**

Graphical models: Graphical representations of *conditional independence*

Helps with…

- Understanding
- Combining expert knowledge and data
- Making inference faster
- Learning cause and effect

Learning cause and effect

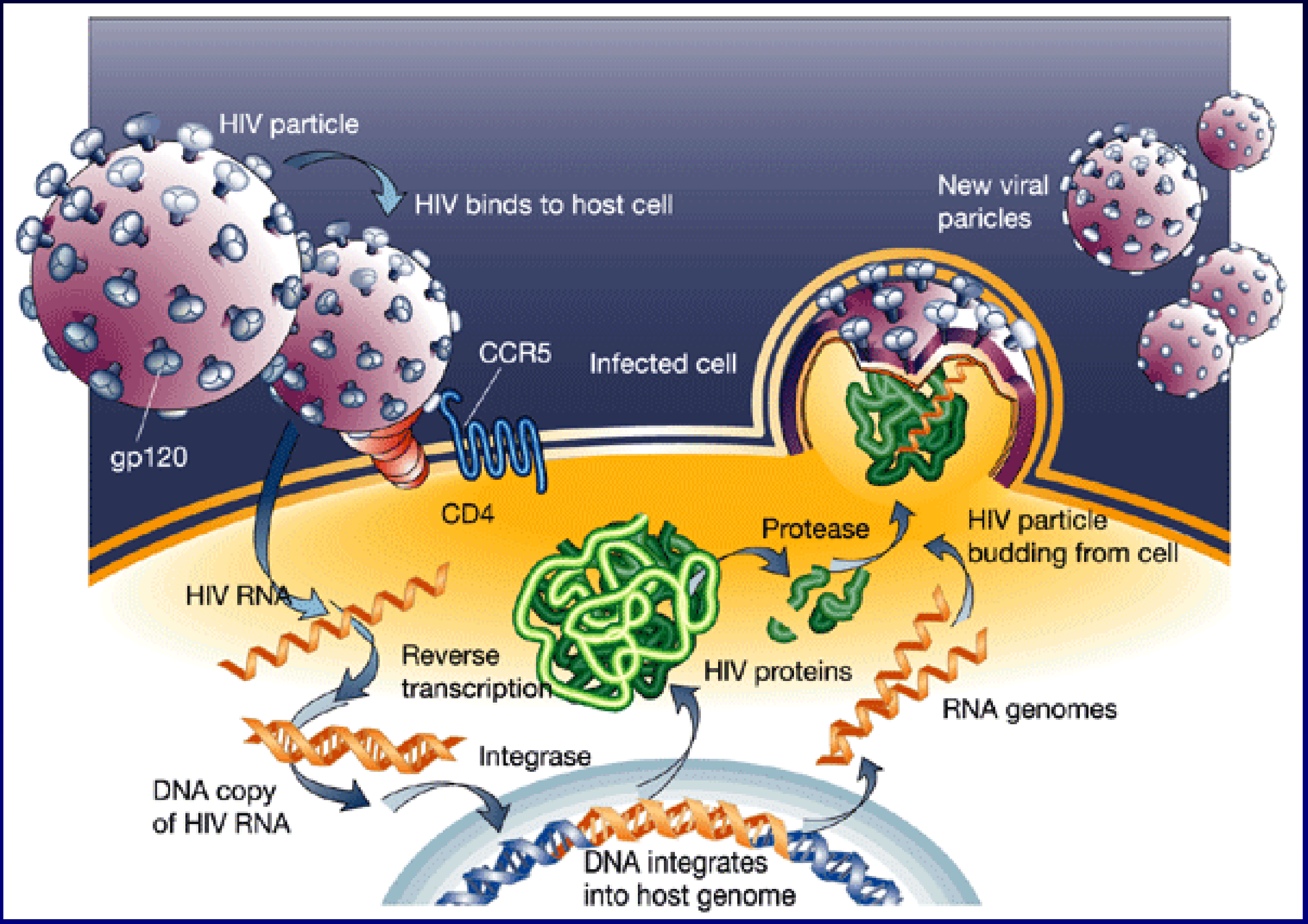
- The standard way:

  Manipulate X and see if p(Y) changes

- A new way:

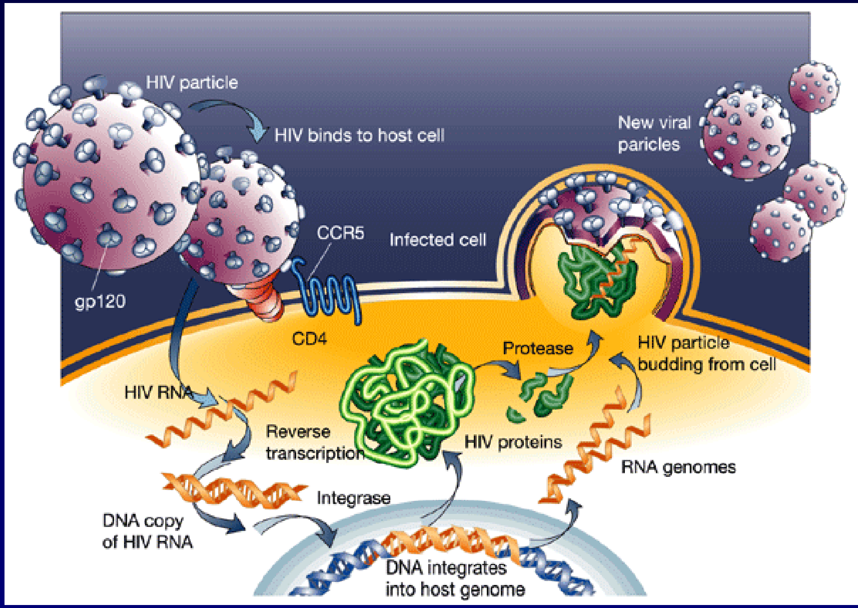
  Observe X, Y, and other variables and infer that X causes Y

HIV Life Cycle



Two arms of Adaptive Immune Response

- **Humoral arm (antibodies):** Recognize, neutralize and respond to free floating virus particles
- **Cellular arm (killer T cells):** Identify and destroy already infected cells

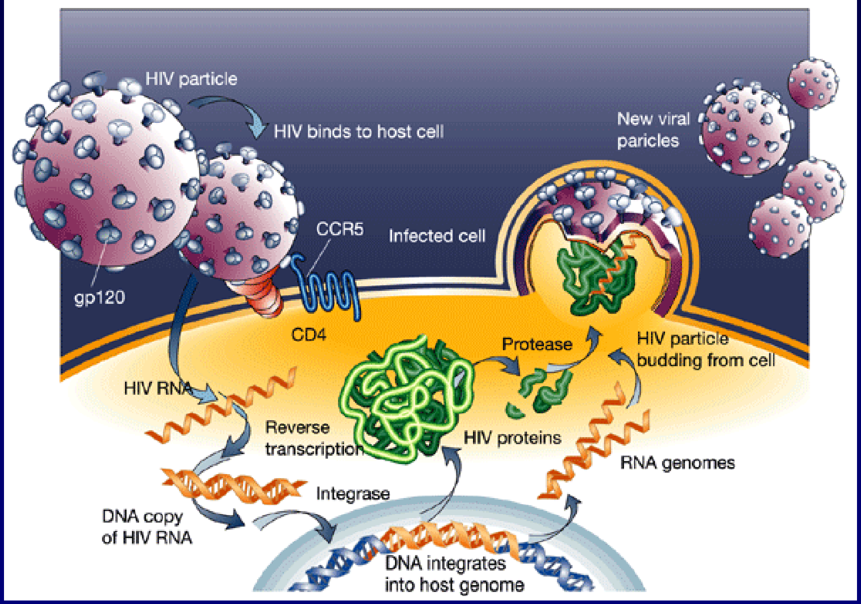


Two arms of Adaptive Immune Response

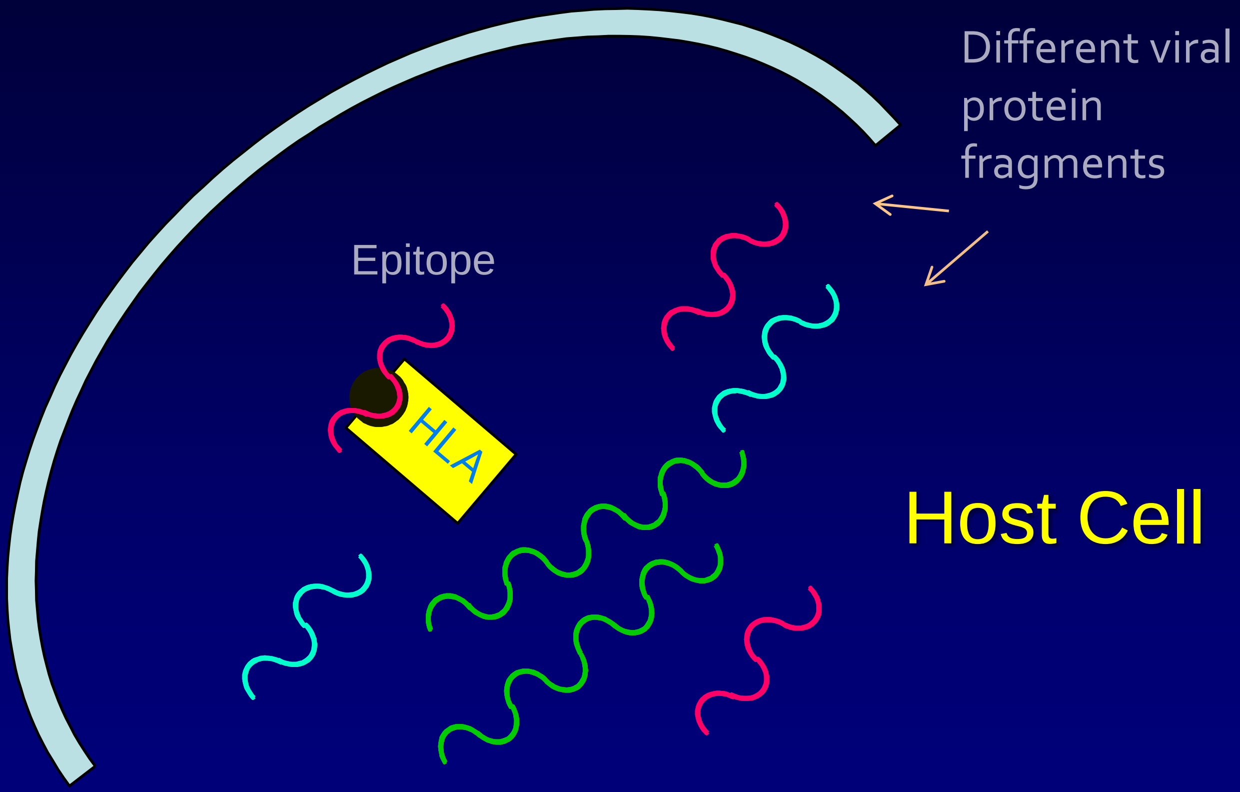
- **Humoral arm (antibodies):** Have been trying for 20 years without success
- **Cellular arm (killer T cells)**

  Today's talk

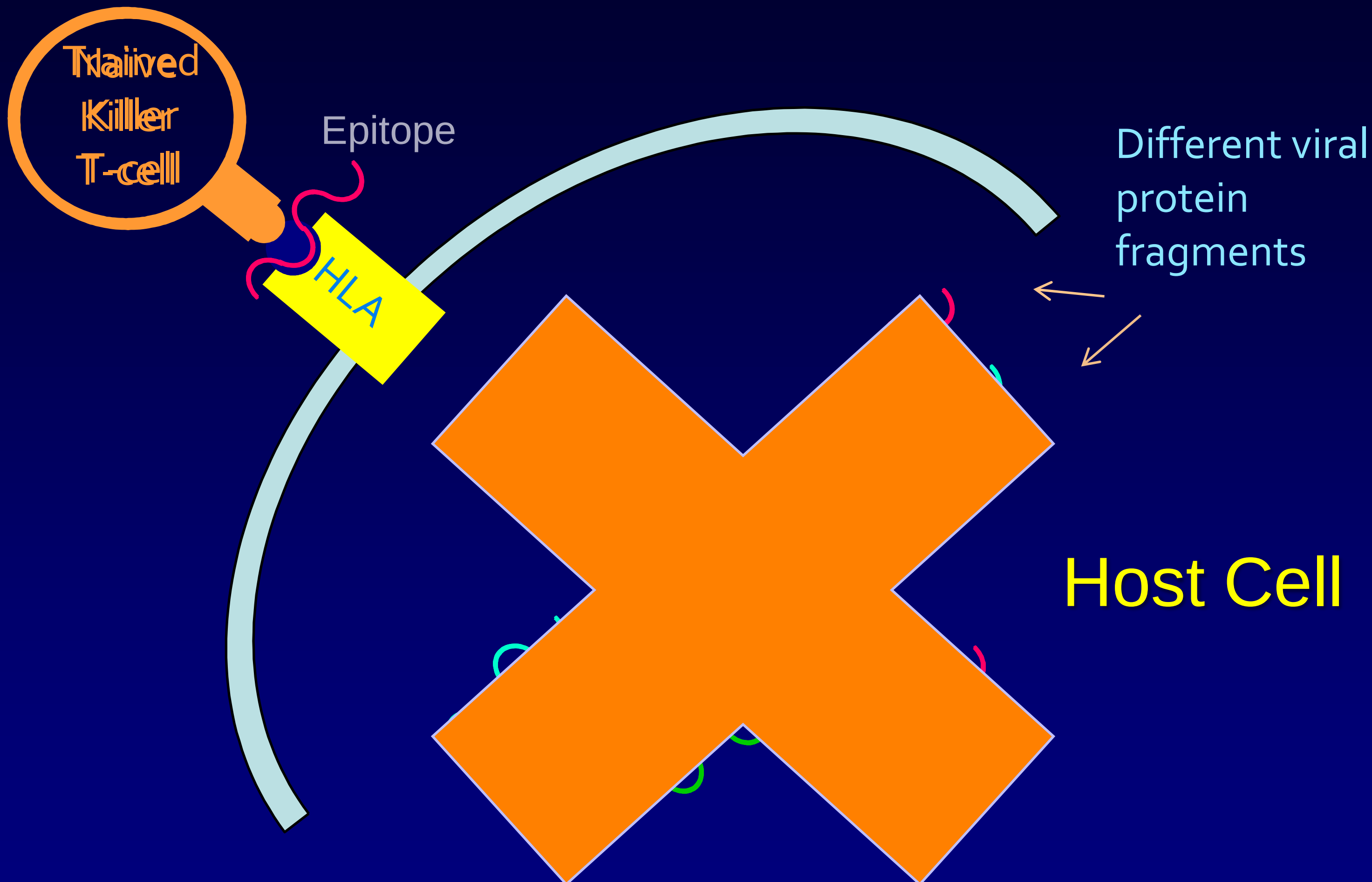
  Vector vs. immunogen



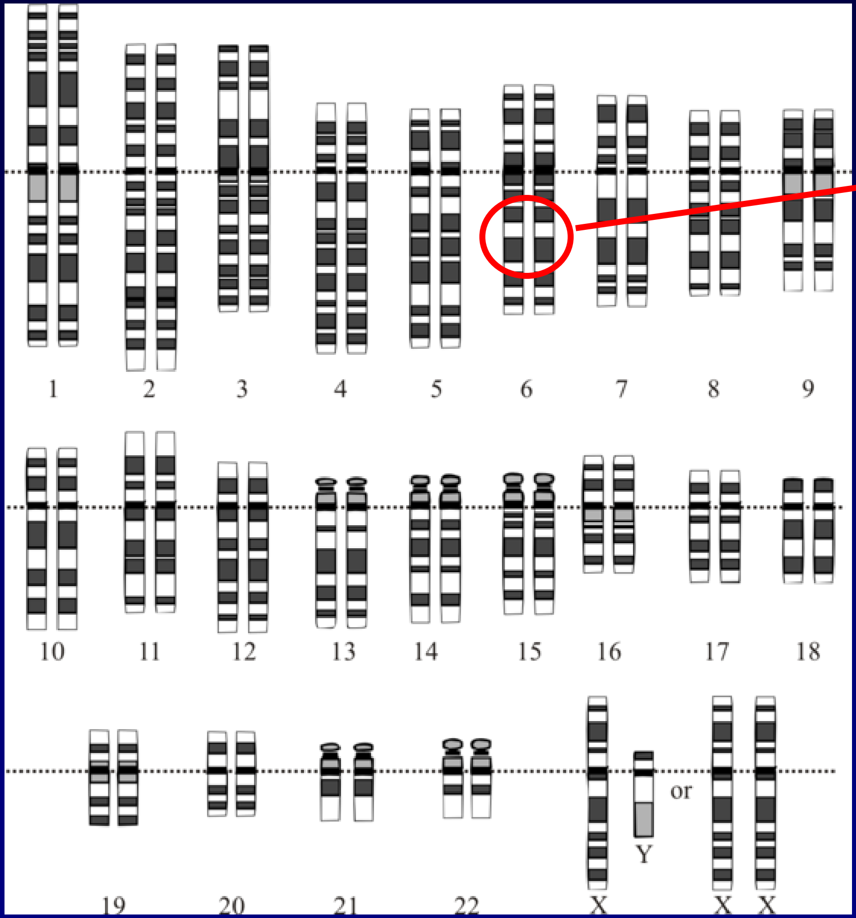
Cellular Arm Details

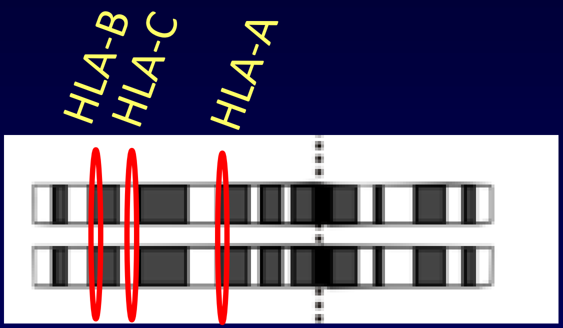


Cellular Arm Details



HLA variability

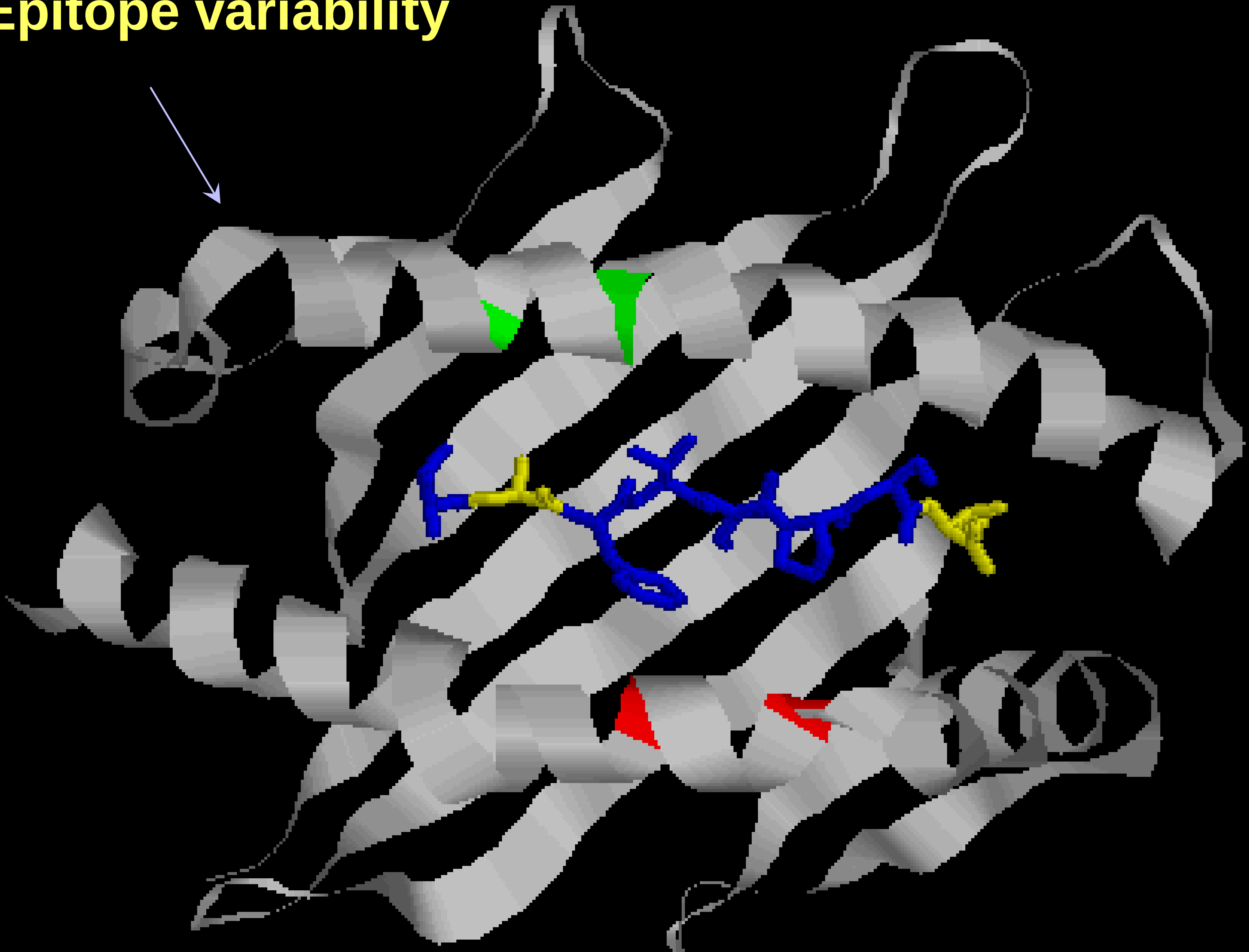




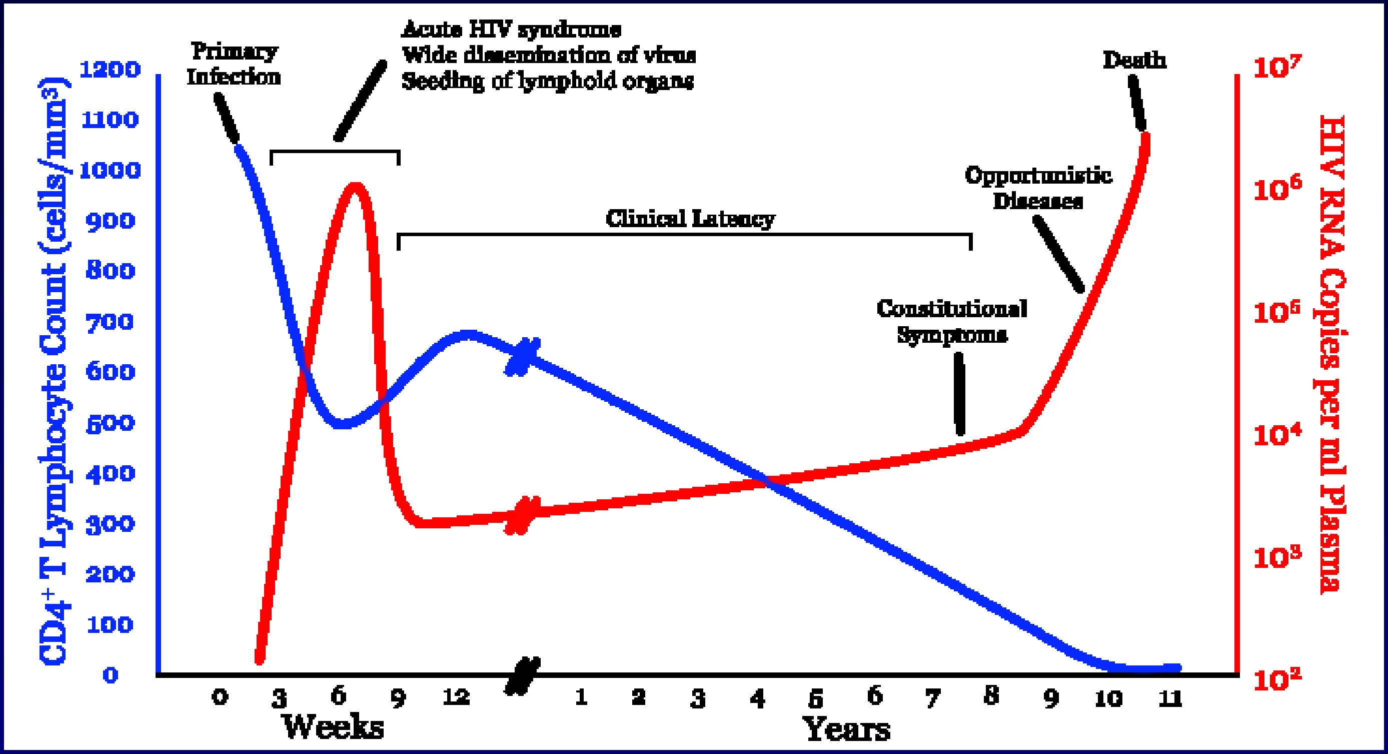
- Each person has up to 6 different HLA types:
(2 ‘A’, 2 ‘B’, 2 ‘C’)

- HLA region is most variable region of DNA - rare for two people to have the same HLA types

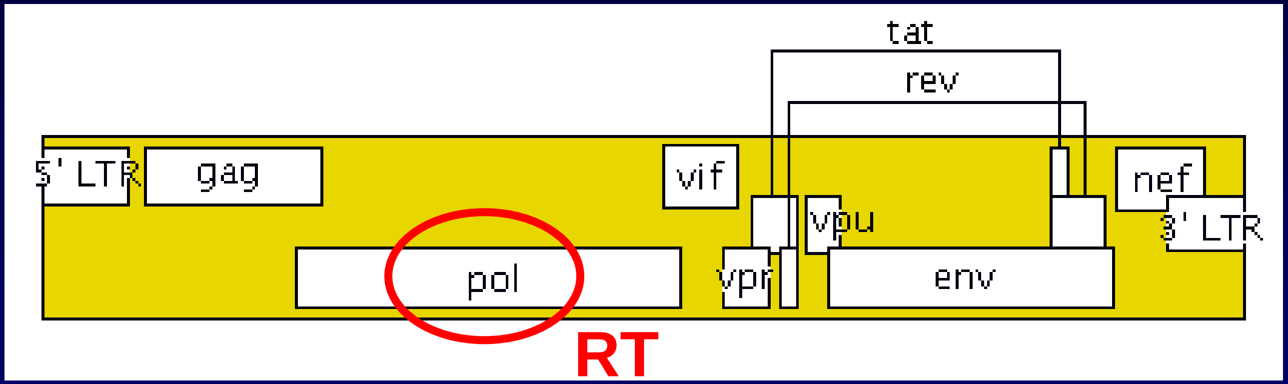
Epitope variability



HIV Disease Progression

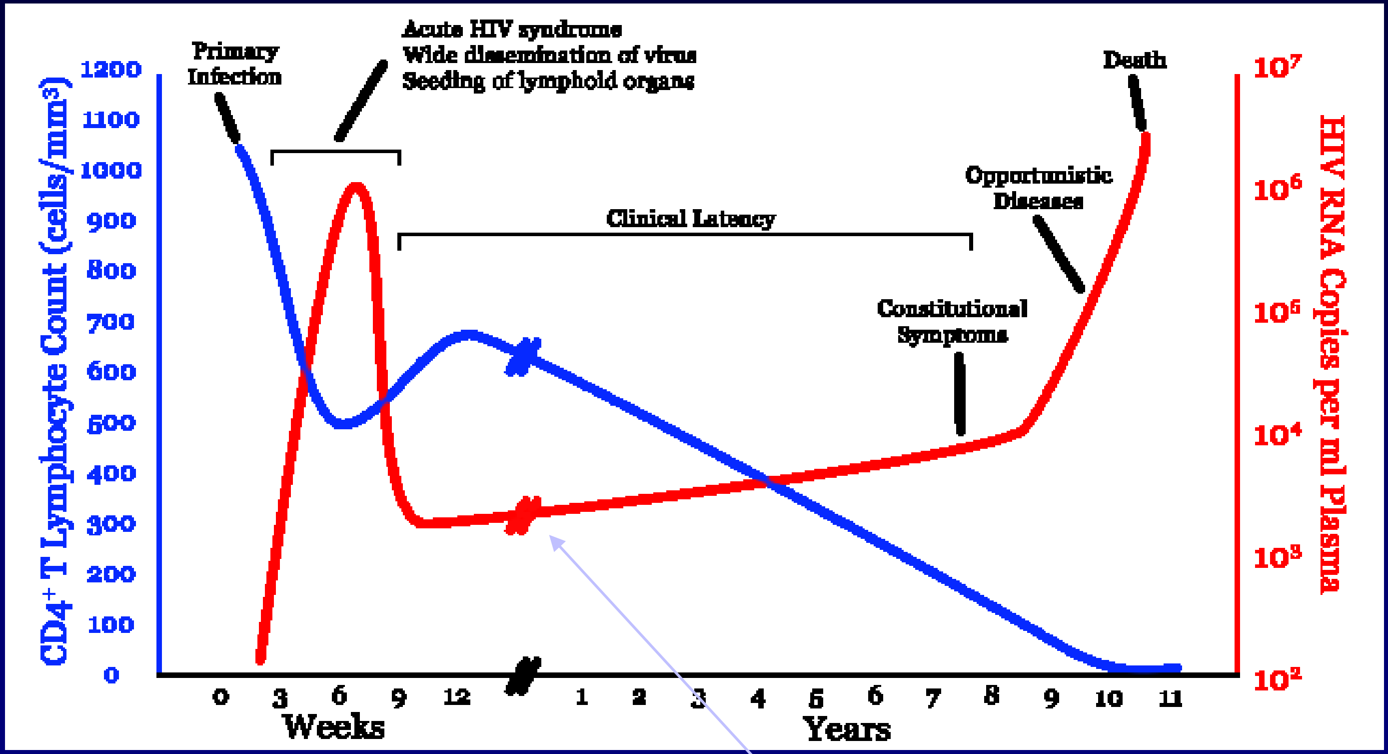


Why do our immune systems fail to control HIV?



- HIV mutates a lot
- If our immune system attacks an epitope, HIV can mutate an AA within (or near) that epitope to avoid the attack

HIV Disease Progression



Vaccine goal: keep viral load low

Some HLA types control HIV better than others

- Protective HLAs: B*57, B*27
- Non-protective HLAs: B*35

Possible causal models

- **Is there something intrinsic about protective HLAs that lead to better protection?**

$$\text{HLA} \longrightarrow \text{Viral control}$$

- **Is there something about the epitopes targeted that lead to better protection?**

$$\text{HLA} \rightarrow \text{Epitope} \rightarrow \text{Viral control}$$

- **Both of the above?**

$$\begin{tikzcd} \text{HLA} \arrow[r] \arrow[rr, bend left] & \text{Epitope} \arrow[r] & \text{Viral control} \end{tikzcd}$$

Learning causal models from data
(Pearl 1993; Spirtes, Glymour, Scheines, 1993)

- **Key assumption: Lack of cause implies conditional independence (causal Markov assumption)**

HLA → Epitope → Viral control

implies

HLA and Viral control and independent given Epitope

Learning causal models from data

- **Another example:**

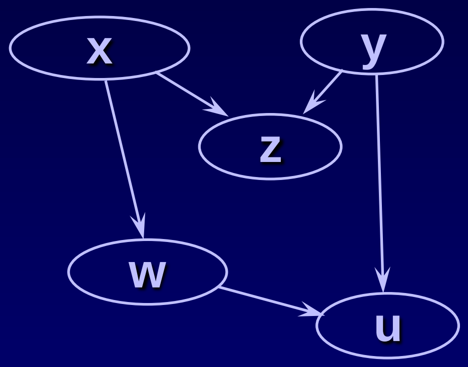
Epitope $\leftarrow$ HLA $\rightarrow$ Viral control

implies

Epitope and Viral control and independent given HLA

Learning causal models from data

Causal Markov assumption:



When the causal graph is interpreted as a DAG model,

the conditional independencies implied by the DAG model hold true in the joint distribution of the variables.

Learning causal models from data

Causal Markov assumption:

When the causal graph is interpreted as a DAG model,

the conditional independencies implied by the DAG model hold true in the joint distribution of the variables.

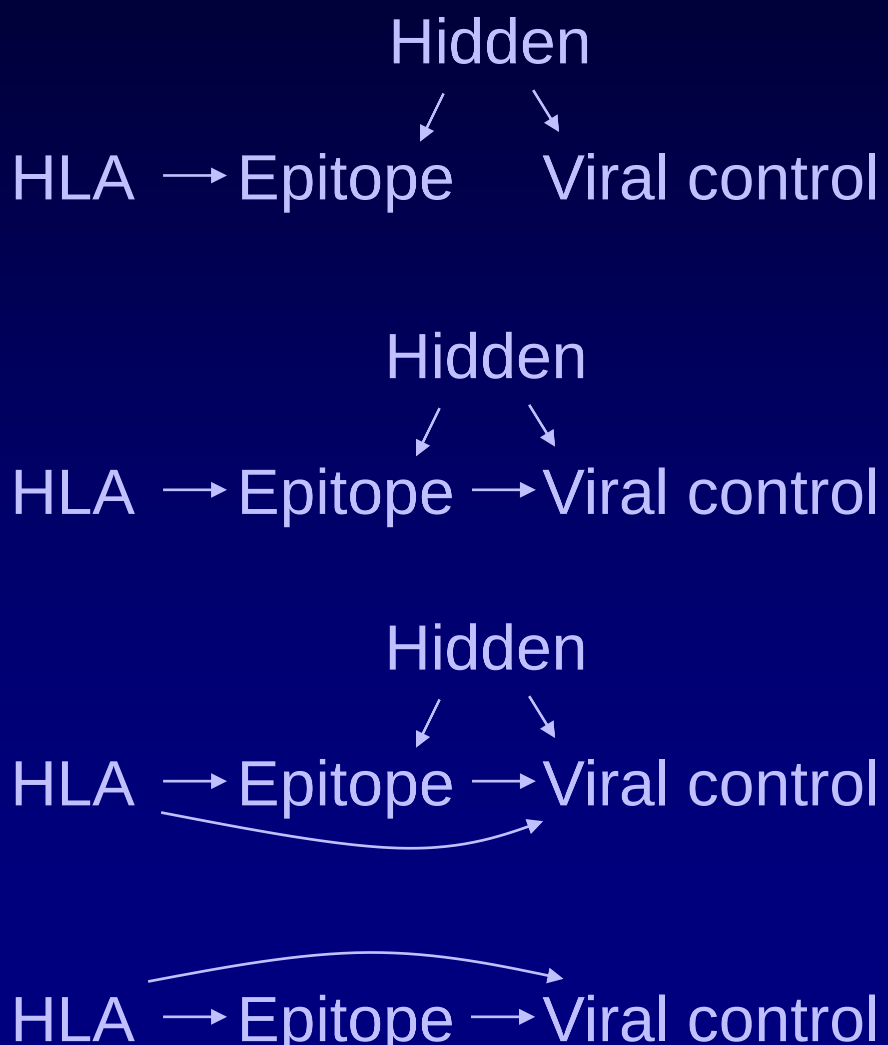
Excluding causal models

$$\begin{tikzcd} \text{HLA} \arrow[dr] & & \text{Viral control} \arrow[dl] \\ & \text{Epitope} & \end{tikzcd}$$

Implies HLA and Viral control are independent

Implies HLA and Viral control are independent

Unable to distinguish these causal models



these models have the same conditional independencies among the observed variables

Simplifying assumptions for our problem

- HLA is a root cause
- HLA causes Epitope

Implies HLA and Viral control are independent

Our possibilities

- **Epitope and Viral control independent given HLA**

$$\text{Epitope} \longleftarrow \text{HLA} \longrightarrow \text{Viral control}$$

- **HLA and Viral control independent given Epitope**

$$\text{HLA} \longrightarrow \text{Epitope} \longrightarrow \text{Viral control}$$

- **No independence**

$$\begin{tikzcd} \text{HLA} \arrow[r] \arrow[rr, bend left] & \text{Epitope} \arrow[r] & \text{Viral control} \end{tikzcd}$$

Details

Other assumptions:

- Faithfulness (conditional independence doesn't happen by accident)
- Causal model is not cyclic (e.g., Epitope ←→ Viral control)

- It's not the case that one model applies to one HLA/epitope and another causal model applies to another HLA/epitope

The Analysis

Subjects:

- Viremic controllers (n=148)
  - At least 3 x VL < 2000 for at least 12 months
  - Blips if infrequent and non-consecutive
- Chronic progressors (n=102)
  - Untreated VL > 10,000

Measuring Epitope:

- IFN-γ ELISpot assays
- Known list: Frahm et al., HIV Mol Imm, 2008; N=222

Predictive models:

- Logistic regression (L1 prior; LASSO)
- Evaluate predictive ability with ROC curves

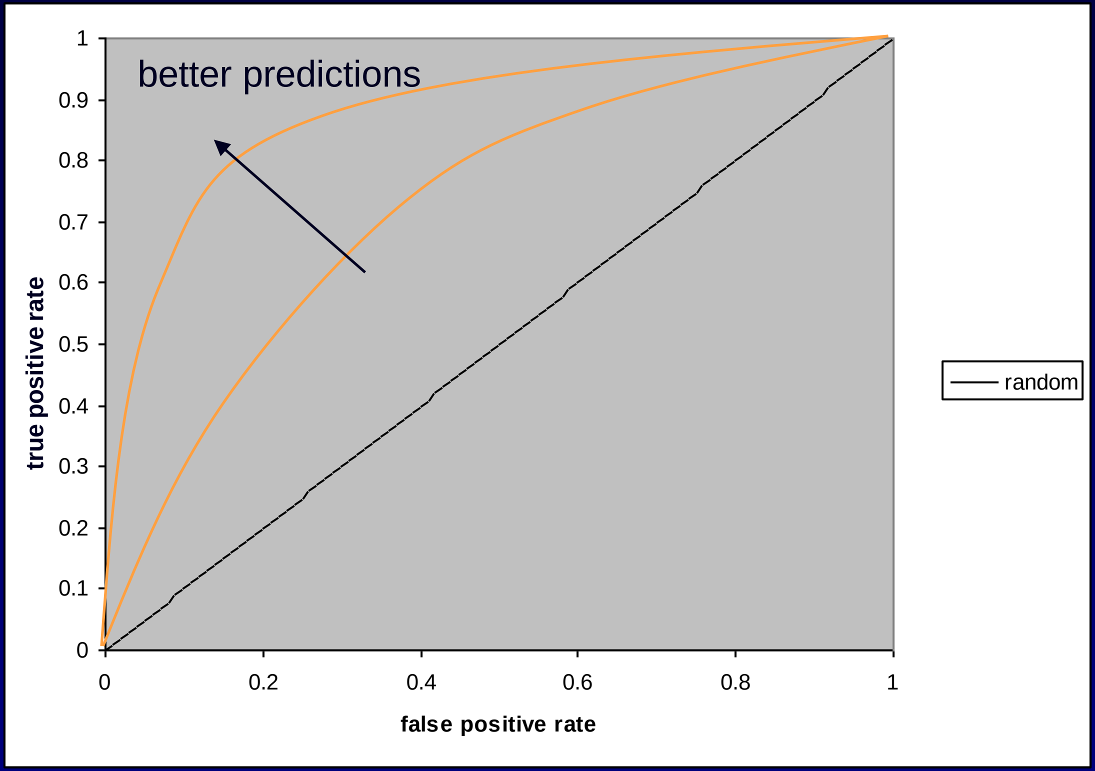
Logistic regression

$$\log \frac{p(y \mid \mathbf{x})}{1 - p(y \mid \mathbf{x})} = w_0 + \sum_{i=1}^{k} w_i x_i$$

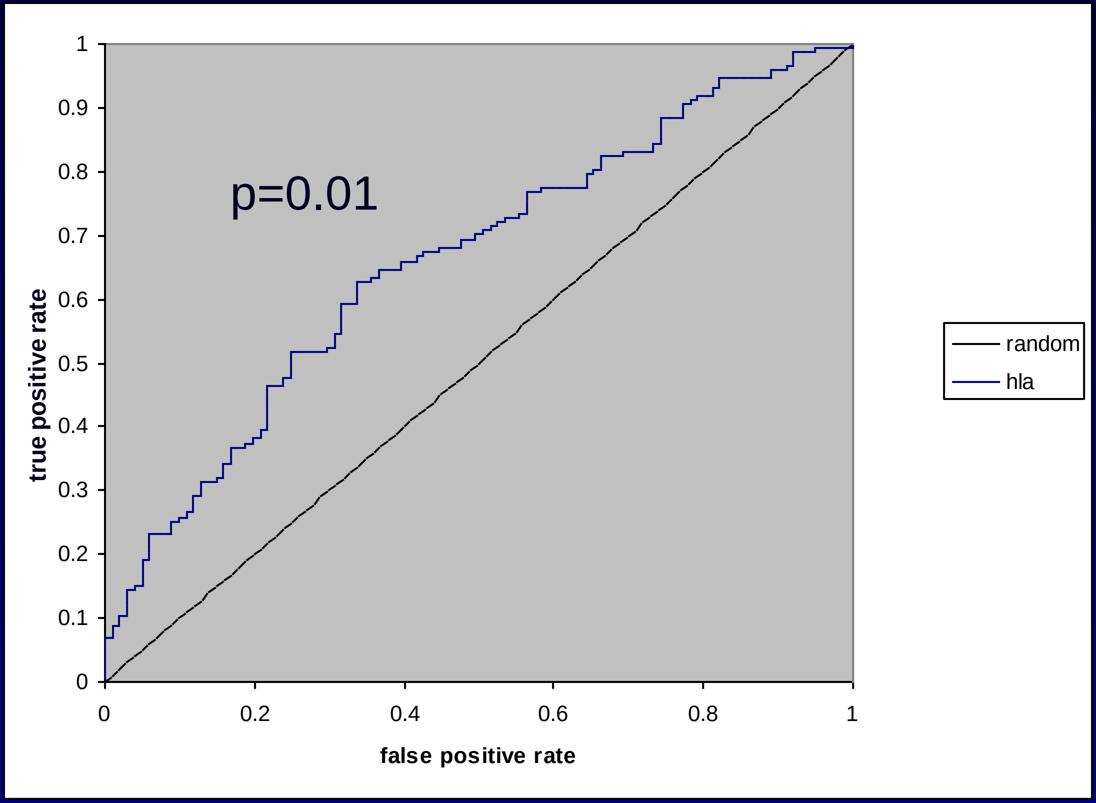
LASSO: MAP values for w are those that maximize

$$\log L - \alpha \sum_{i=1}^{k} \left| w_i \right|$$

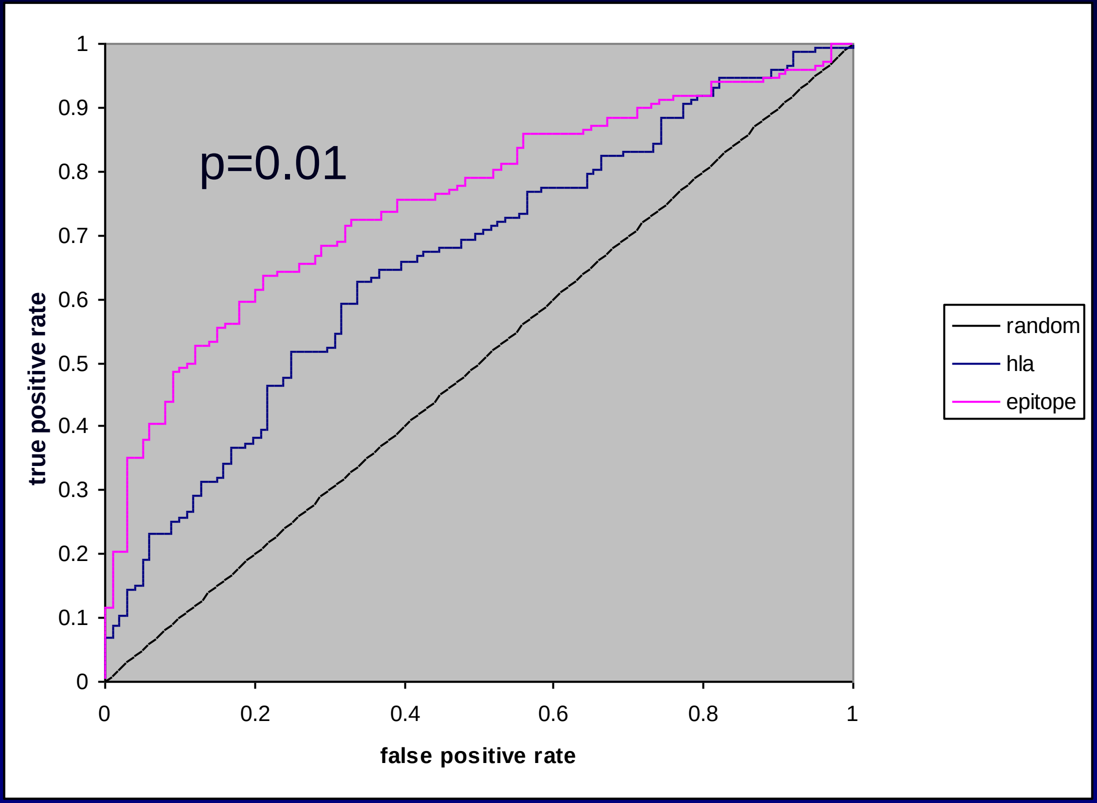
ROC curve:
A measure of predictive accuracy



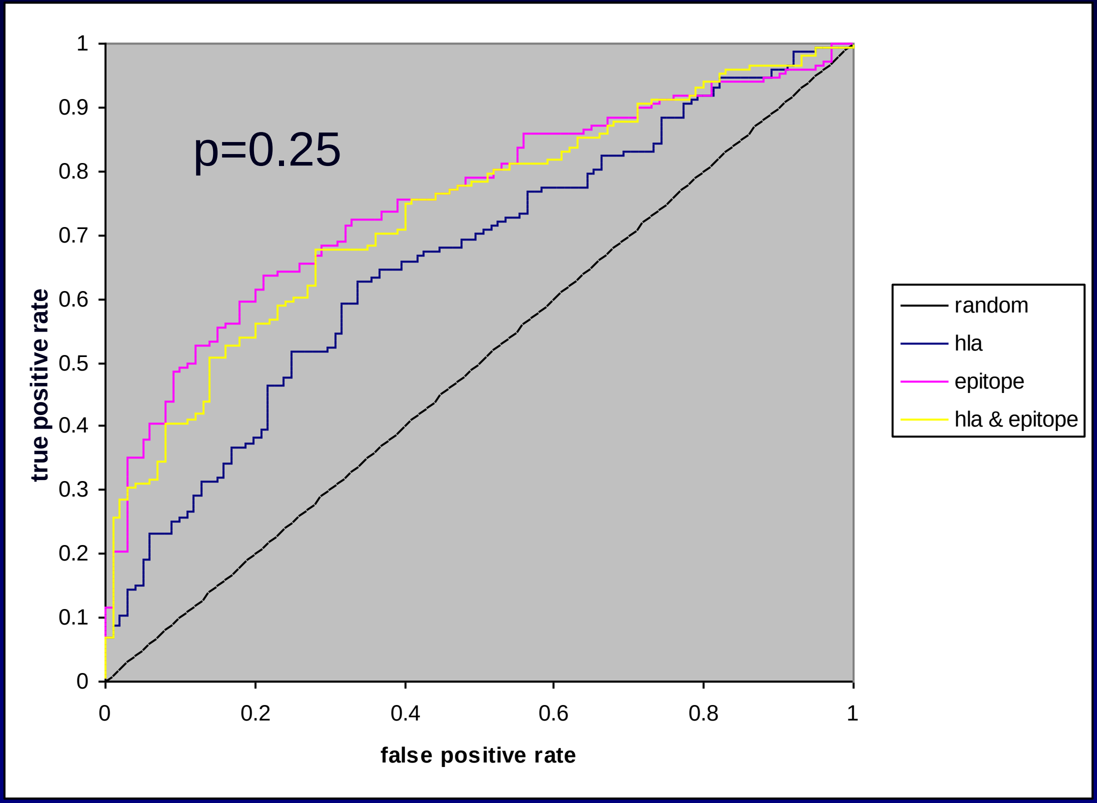
HLA predicts viral control better than random



Epitope responses predict viral control better than HLA



HLA adds no additional information over epitope responses



Inferring causal model from data

- **Epitope and Viral control independent given HLA**

$$\text{Epitope} \leftarrow \text{HLA} \rightarrow \text{Viral control}$$

- **HLA and Viral control independent given Epitope**

$$\text{HLA} \rightarrow \text{Epitope} \rightarrow \text{Viral control}$$

- **No independence**

$$\text{HLA} \rightarrow \text{Epitope} \rightarrow \text{Viral control}$$

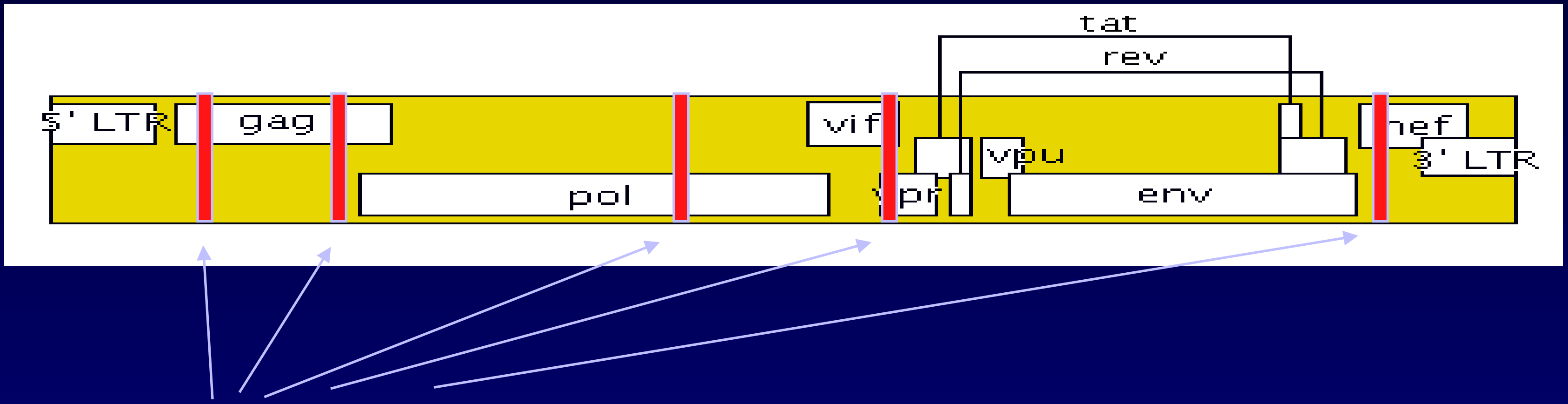
Targeting of which epitopes leads to viral control / chronic progression?

Use logistic regression with forward selection to identify the epitope specific CD8+ T cell responses that correlate with HIV control (q-value < 0.2)

Epitope specific CD8+ T cell responses associated with viral control--“good” epitopes

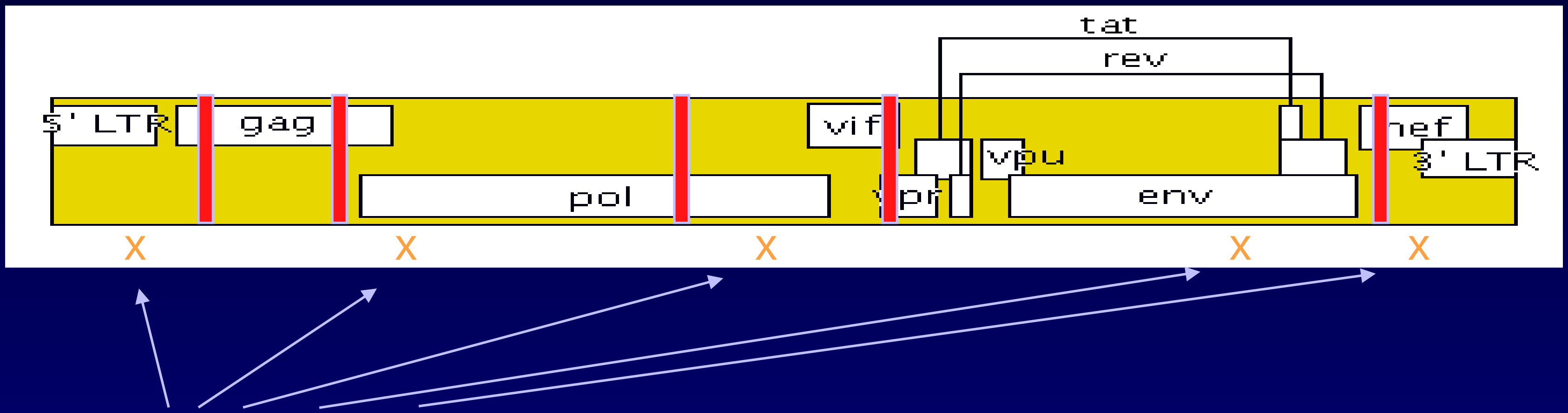
PEPTIDE	HLA	PROTEIN	P VALUE	Q VALUE	L1 WEIGHT
AW9	B*57	*Vpr*	1.2 10^{-7}	0	3.19
KK10	B*27	*p24*	<0.001	0.02	1.78
TW10	B*57	*p24*	0.001	0.02	0.87
HW9	B*57	*Nef*	0.002	0.08	1.11
DA9	B*14	*p24*	0.003	0.12	0.88
LV10	A*02	*Nef*	0.004	0.18	0.55

Consequences for vaccine design



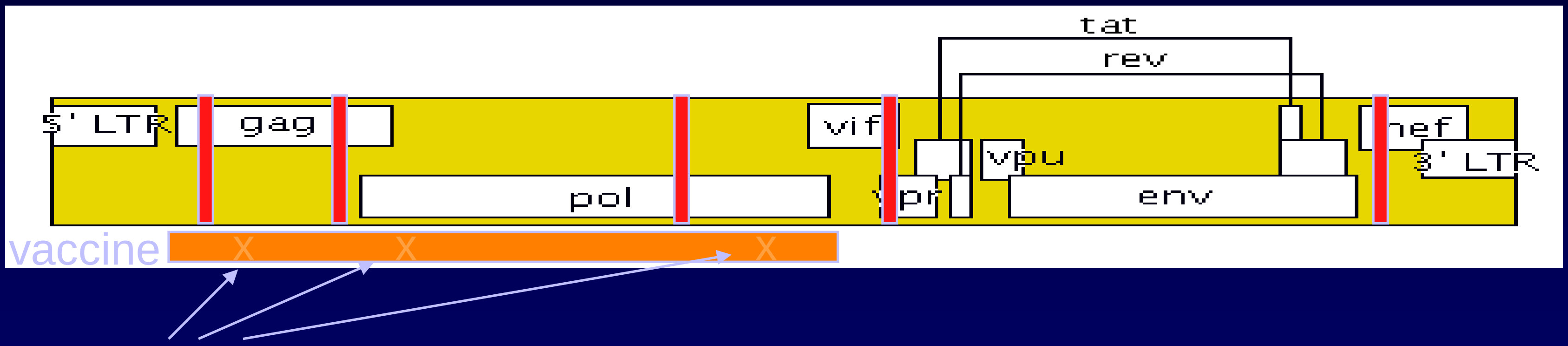
A few epitopes make HIV vulnerable

Consequences for vaccine design



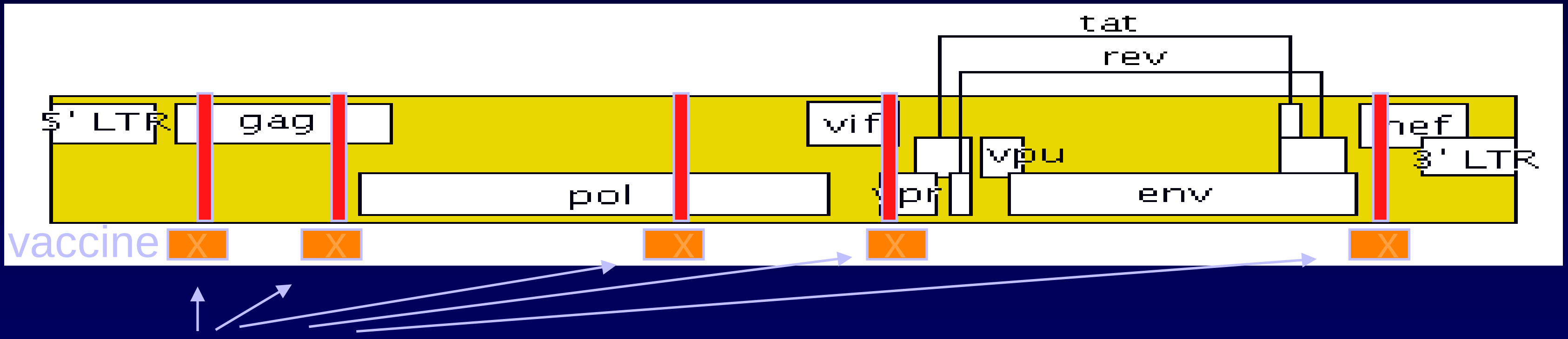
Left to its own devices, our immune system attacks at random epitopes (immunodominance)

Consequences for vaccine design



A “whole protein” vaccine does little to help the situation

Consequences for vaccine design



A focused vaccine can show immune system where to attack

Things I didn't have time to talk about

- Factor graphs, mixed graphs, etc.
- Relational learning: PRMs, Plates, PERs
- Bayesian methods for learning
- Variational methods
- Non-parametric distributions

To learn more

Tutorial on my home page

Main conferences:

- Uncertainty in Artificial Intelligence (UAI)
- Neural information Processing Systems (NIPS)