An Introduction to Bayesian Networks

Master Seminar "Biological Networks"
03.02.2023

Medizinische Fakultät Heidelberg

Program of the day

Medizinische Fakultät Heidelberg

- 10.15am - 12.30am: lecture "Introduction to bayesian networks"
- 1.30pm: start of the practical part (download the RMarkdown on Moodle!)
- 3.30pm - 4.30pm: discussion of the practical part; debrief

Content of the presentation

- A **brief introduction** to the theory behind Bayesian networks (some slides from a presentation by M. Scutari)

- An **example of chromatin network** reconstruction in Neuroblastoma

- **Practical example / demo**
  - Reconstruction the BN of T-cell signaling pathway (Sachs et al., 2005)
  - Reconstruction of the BN for chronic lymphocytic leukemia patients

- Presentation, data, R Markdown scripts can be found here on Moodle!

Biological networks

undirected edge

- Protein A interacts with Protein B
- Disease A and disease B are comorbid
- Gene A is co-expressed with gene B

directed edge

- kinase A phosphorilates protein B
- TF A regulates gene B
- condition A (social status) influences condition B (disease risk)

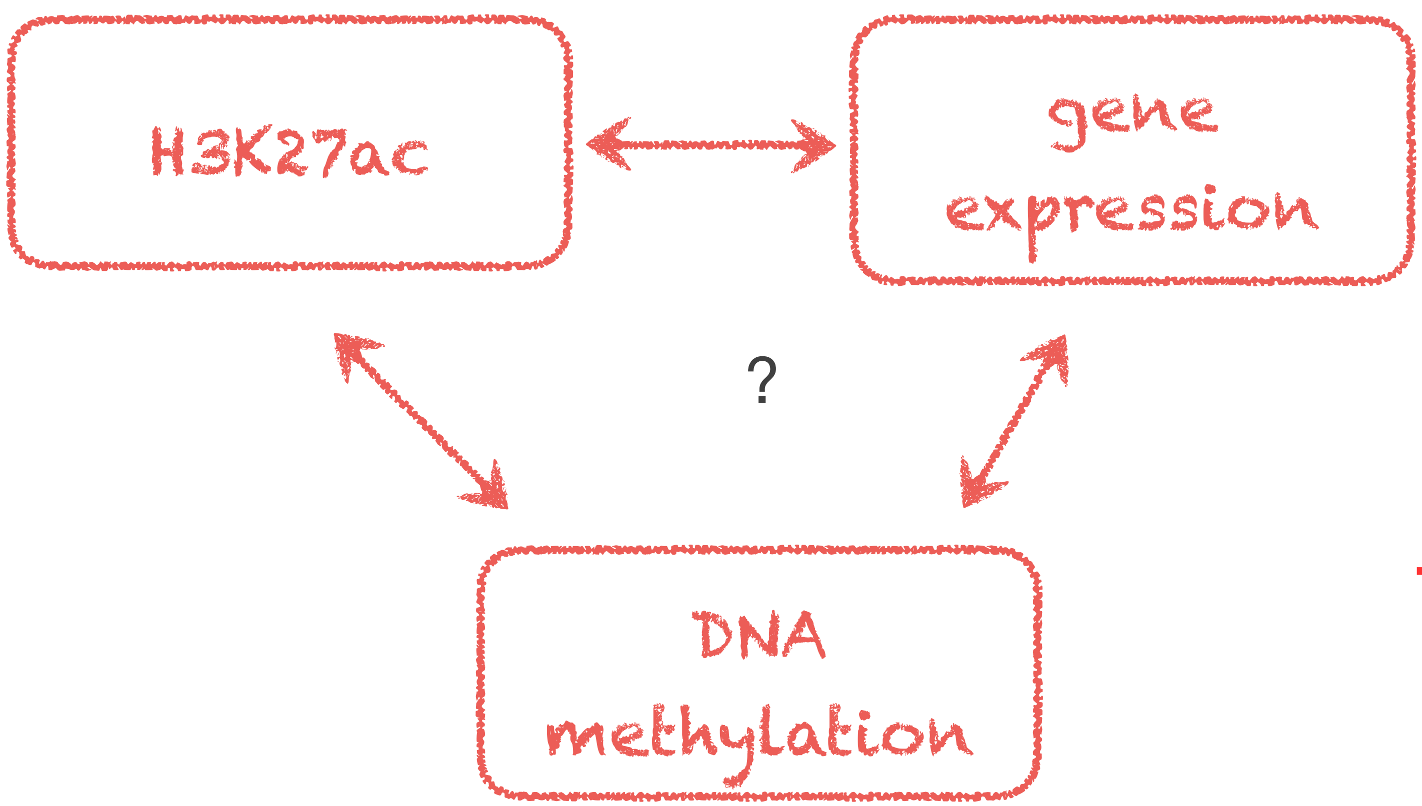
Causal relationships can be represented by directed graphs
Directed graph <u>can</u> represent causal relationships

Modelling chromatin networks

Medizinische Fakultät Heidelberg

- Most genomic analysis is based on **correlation** between features
  → ***"correlomics"***

- Can we go beyond towards oriented/causal networks ??

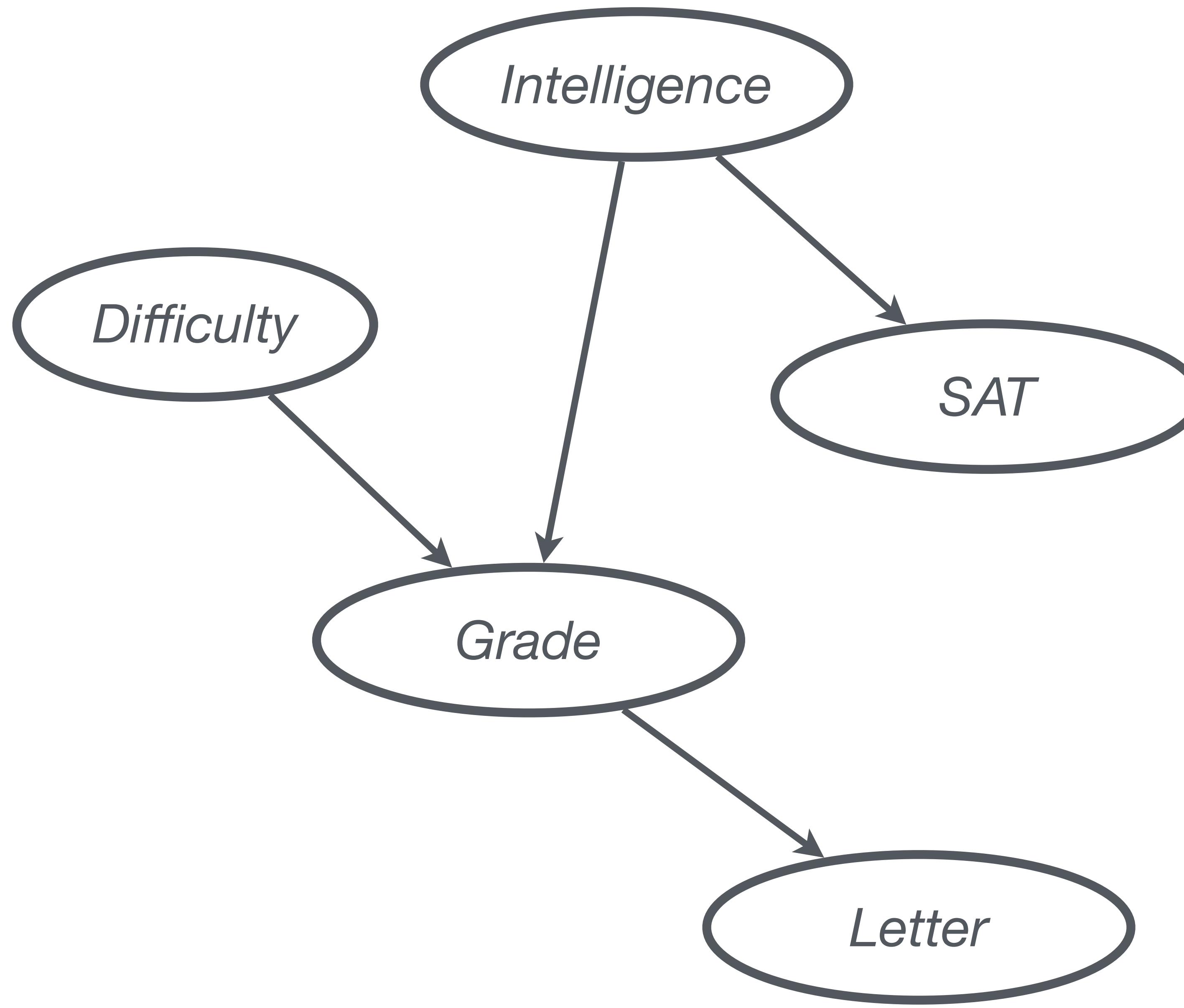


→ ***Bayesian Networks***

What are bayesian networks ?

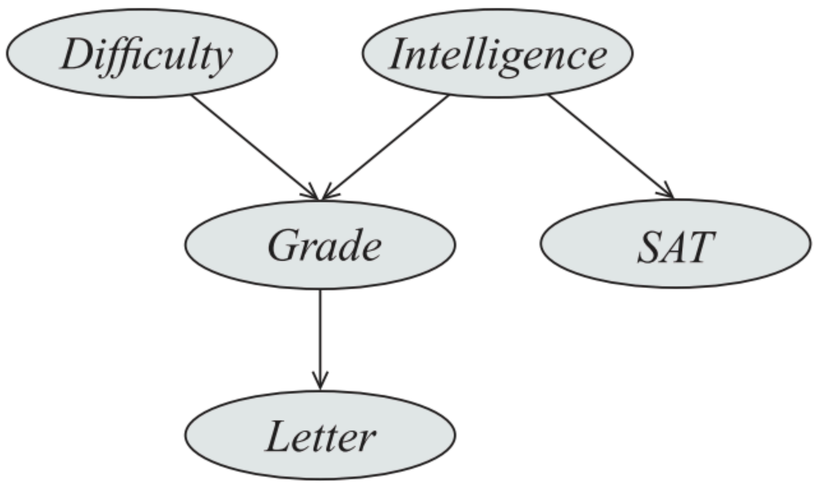
Medizinische Fakultät Heidelberg

- **Student model**
  - *Grade* at the exam
  - *Difficulty* of the exam
  - *Intelligence* of the student
  - *Scholastic assessment test (SAT) score*
  - *Recommendation letter*



What are bayesian networks ?

- **Student model**
  - ***Grade*** at the exam depends on the ***Difficulty*** of the exam and the ***Intelligence*** of the student
  - Intelligence influences the ***SAT score***
  - Grade at the exam influences how good the ***recommendation letter*** will be.



- Network of influences and conditional (in)dependences!
- Not necessarily **causal** networks !
- Need **interventional data** (perturbations) to turn a BN into a causal network

  Is Grade the direct and only influence on Recommendation letter?

Basic concepts in (bayesian) statistics

101 Bayesian statistics

Joint Probability Distribution: $P(A, B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$

Bayes Formula:

$$\boxed{P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}}$$

$P(A \mid B)$: **"*Conditional probability of A given B*"**

101 Bayesian statistics

- Random variable A, B are **independent** if

$$P(A,B) = P(A)\cdot P(B)$$

Example: probability to get *P(HT)* in a series of 2 coin throws

- From the Bayes formula, we thus have, if *X* and *Y* are **independent**:

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

- Marginalizing out a variable:

$$P(A) = \sum_B P(A,B) = \sum_B P(A\,|\,B)\ P(B)$$

101 Bayesian statistics

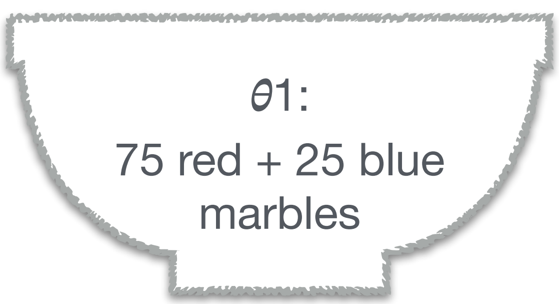
- Often, one random variable represents the **observed data *D***, the other represents the **model *θ* :**

posterior probability

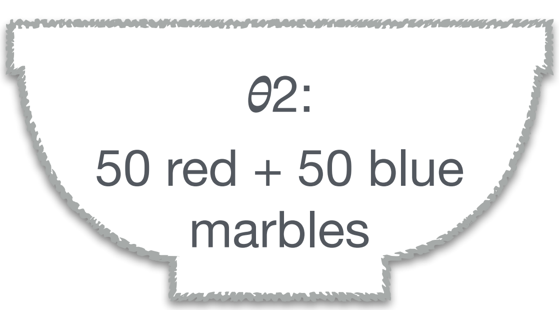
$$P(D\,|\,\theta) = \frac{P(\theta\,|\,D)\cdot P(D)}{P(\theta)}$$

likelihood of the data given the model

prior



Prior: $P(\theta 1) = 0.5\ ;\ P(\theta 2) = 0.5$



- 1 red marble sampled (observed data D): probability that is was sampled from bowl M1?

$P(D\,|\,\theta 1) = 0.75 \qquad P(D\,|\,\theta 2) = 0.5$

$$P(\theta 1\,|\,D) = \frac{P(D\,|\,\theta 1)P(\theta 1)}{P(D)}$$

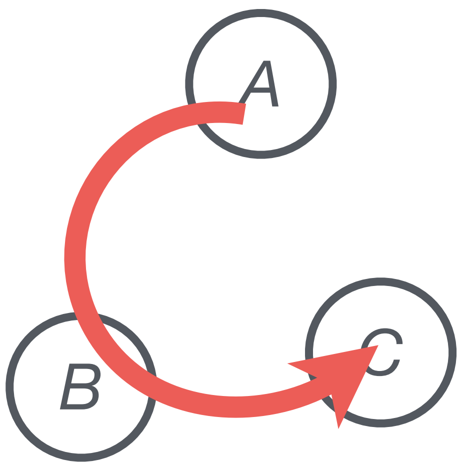
$$= \frac{P(D\,|\,\theta 1)P(\theta 1)}{P(D\,|\,\theta 1)P(\theta 1) + P(D\,|\,\theta 2)P(\theta 2)} = \frac{0.75\cdot 0.5}{0.75\cdot 0.5 + 0.5\cdot 0.5} = 0.6$$

101 Bayesian statistics

Medizinische Fakultät Heidelberg

Joint probability:
$$
\begin{aligned}
P(A,B,C) &= P(A)\cdot P(B\,|\,A)\cdot P(C\,|\,A,B)\\
&= P(A)\cdot P(C\,|\,A)\cdot P(B\,|\,A,C)\\
&= P(C)\cdot P(A\,|\,C)\cdot P(B\,|\,A,C)\\
&= P(C)\cdot P(B\,|\,C)\cdot P(A\,|\,B,C)
\end{aligned}
$$

(+ 2 other equations)



Bayes Formula:
$$
\boxed{P(A\,|\,B,C) = \frac{P(B\,|\,A,C)\cdot P(A\,|\,C)}{P(B\,|\,C)}}
$$

Conditional independence:
$$
P(A\,|\,B,C) = P(A\,|\,C)
$$

does **not** mean that *P(A|B) = P(A) !!*

"*A is **conditionally** independent of B given C*"

Conditional (in)dependence

Medizinische Fakultät Heidelberg

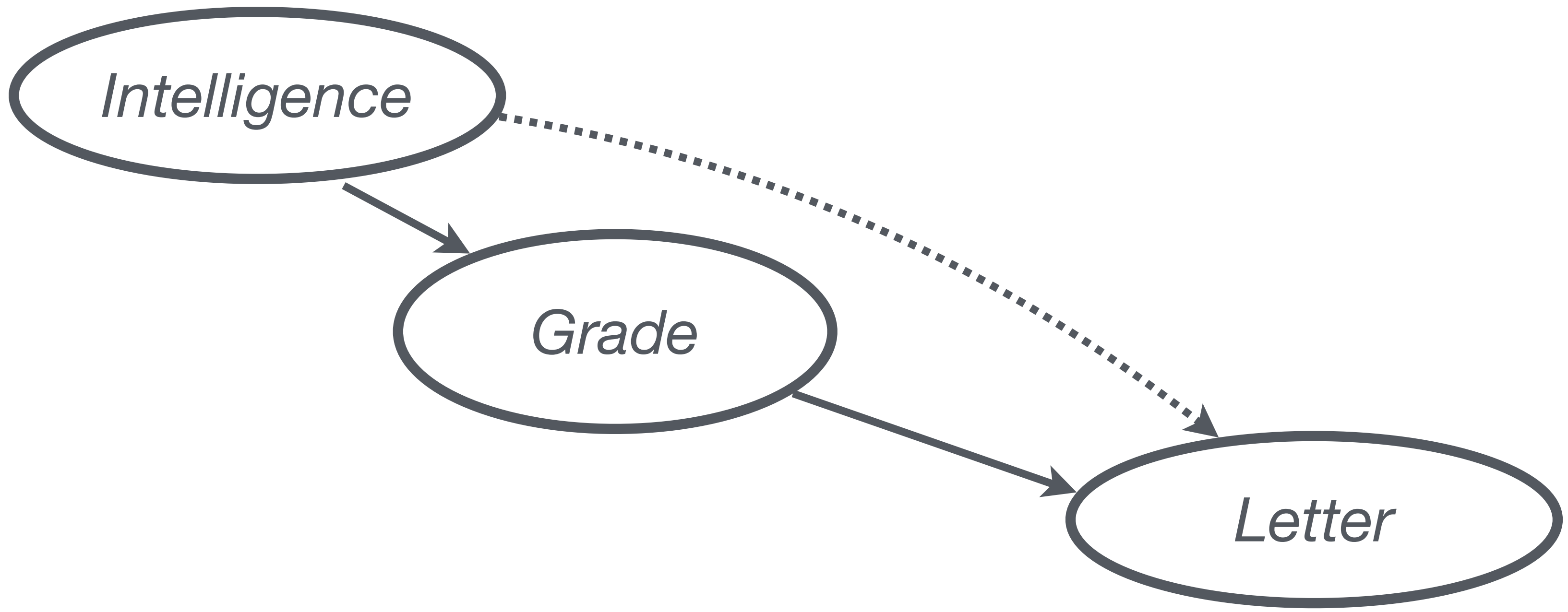
- Body size (event S) and richness of vocabulary (event V) are dependent (smaller people are usually children...)

$$S \not\perp V$$

- But knowing that the persons are above 18 years (event A) makes the two independent

$$S \perp V \,|\, A$$

101 Bayesian statistics



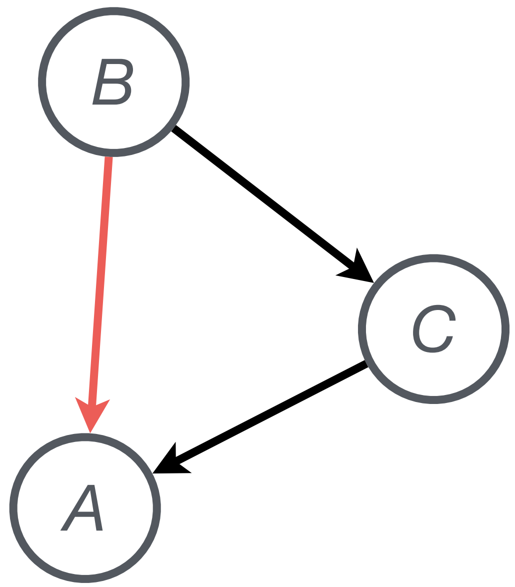
- Intelligence and Letter are not independent!

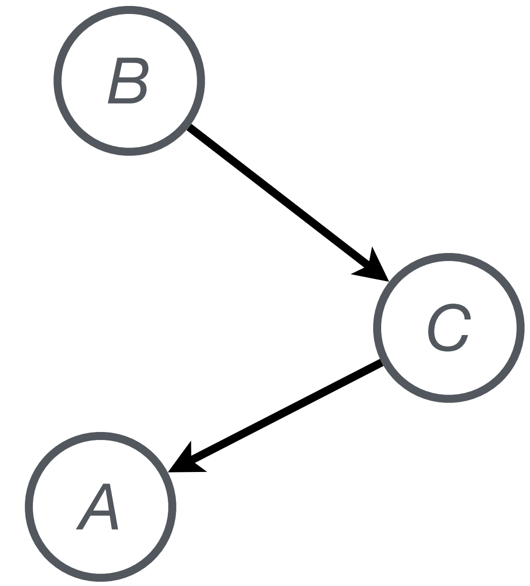
$$P(L \mid I) \neq P(L)$$

- But if I know the Grade, then the Intelligence will no longer influence the Letter!

$$P(L \mid I, G) = P(L \mid G)$$

101 Bayesian statistics





$$P(A \mid B, C) \neq P(A \mid C)$$

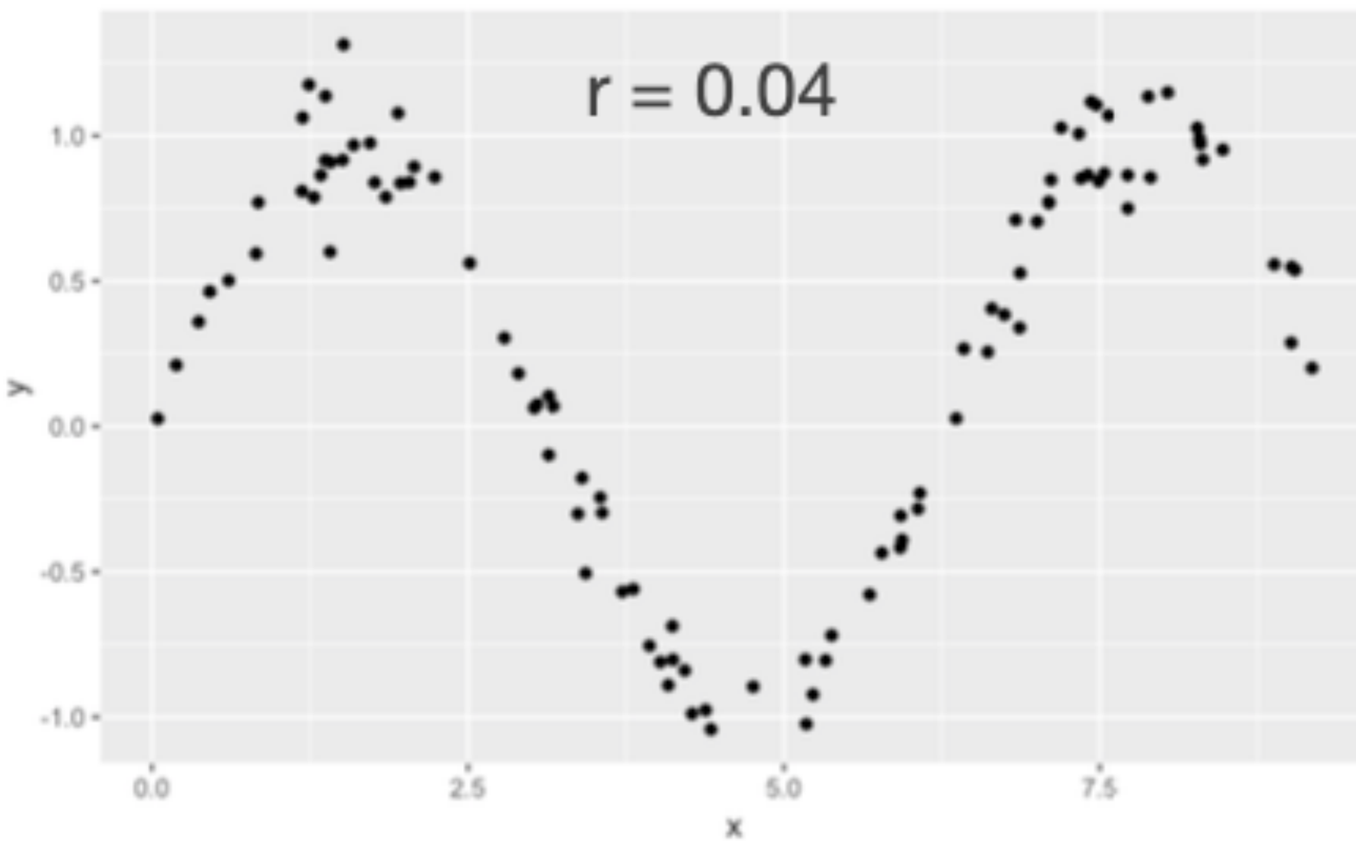
A and *B* are NOT conditionnally independent!

$$P(A \mid B, C) = P(A \mid C)$$

A and *B* are conditionnally independent!

The relations of conditional independence constrain the structure of the network!

Testing (in)dependence



correlation is not a good measure of independence...

Testing conditional independence

Medizinische Fakultät Heidelberg

- **Mutual information** between A and B

$$I(A,B) = P(A,B)\ \log\left(\frac{P(A,B)}{P(A)P(B)}\right)$$

P(A,B) = P(A)P(B) if independent
→ I(A,B) = 0

- **Conditional mutual information** between A and B given C:

$$I(A,B\,|\,C) = P(A,B,C)\ \log\left(\frac{P(A,B,C)\,P(C)}{P(A,C)\,P(B,C)}\right)$$

Testing conditional independence

Medizinische Fakultät Heidelberg

- How to compute the conditional mutual information in discrete cases?

in practice: $$I(A,B\,|\,C) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_C} \frac{n_{ijk}}{n} \log\left(\frac{n_{ijk}\,n_{++k}}{n_{i+k}\,n_{+jk}}\right)$$

C = 0

	A=1	A=0
B=1	90	30
B=0	30	70

C = 1

	A=1	A=0
B=1	20	10
B=0	80	90

$n_{011} = 10$

$n_{0+1} = 100$

$n_{+11} = 30$

$n_{++1} = 200$

$n = 400$

Medizinische Fakultät Heidelberg

Questions ?

Test yourself !

What is the probability that someone smokes, if he has bronchitis?

$$P(S) = (0.5, 0.5)$$

$P(B \mid S)$	$S = \text{no}$	$S = \text{yes}$
$B = \text{no}$	0.7	0.4
$B = \text{yes}$	0.3	0.6

$$P(S=1 \mid B=1) = \frac{P(B=1 \mid S=1)P(S=1)}{P(B=1)} = \frac{P(B=1 \mid S=1)P(S=1)}{P(B=1 \mid S=1)P(S=1) + P(B=1 \mid S=0)P(S=0)}$$

$$= \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = 0.67$$

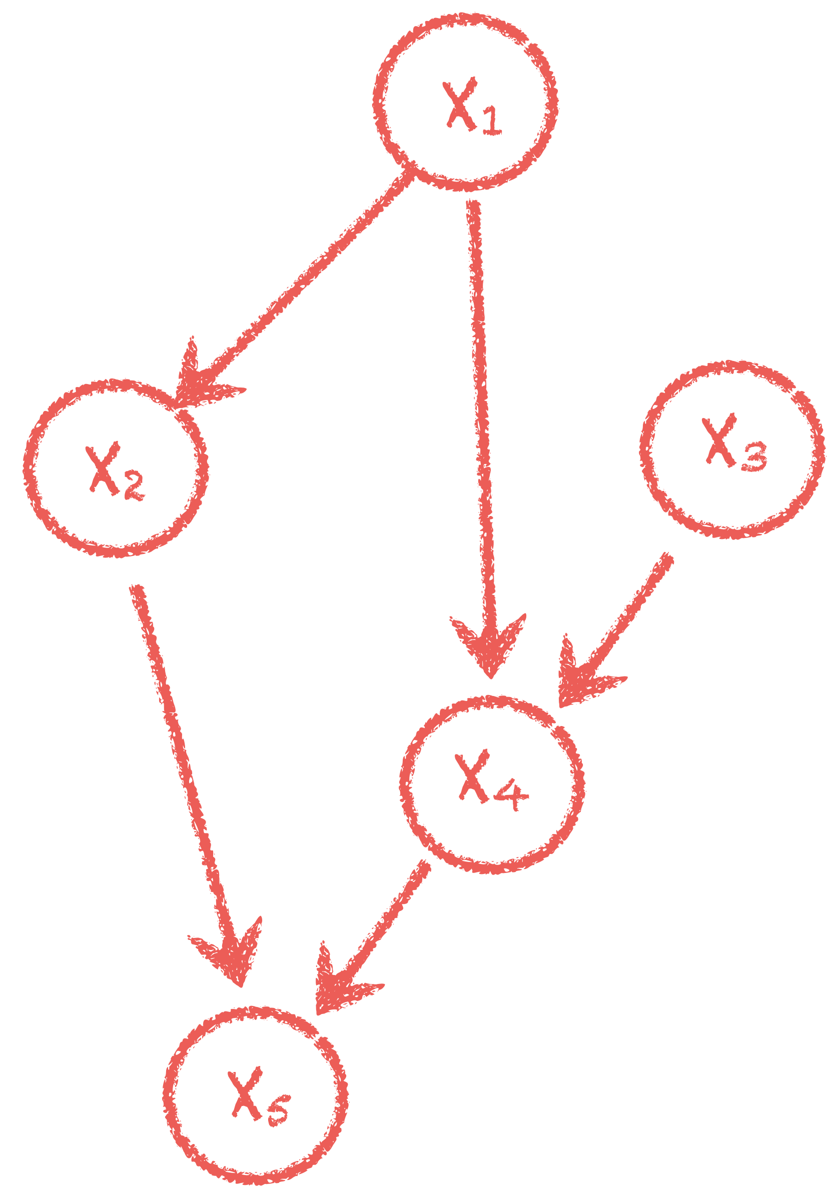
Medizinische Fakultät Heidelberg

Definition of Bayesian Networks

What are bayesian networks ?

Medizinische Fakultät Heidelberg

- **Graph** $G = (V,A)$ with nodes V and edges A
- each node v_i is a **random variable** X_i
- Property of the graph:
  Directed Acyclic Graph (DAG)
- no cycle: you cannot get back to your starting point following the edges



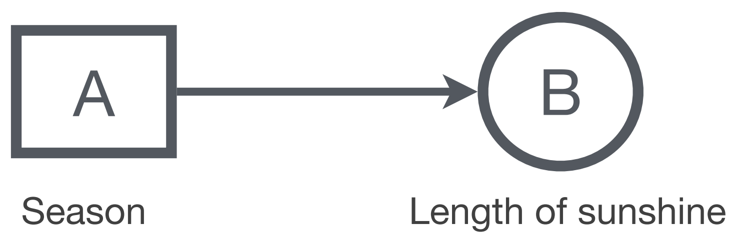
[Friedman, 2004,Friedman et al., 2000]

Conditional probabilities

$$P(B\,|\,A)$$

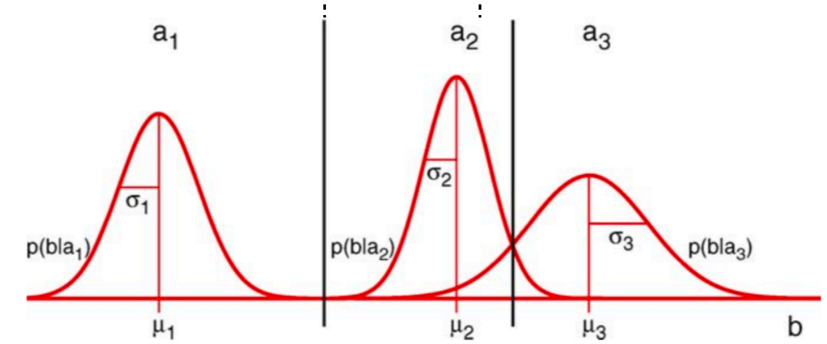
	B = no	B = yes
A = good	0.3	0.7
A = bad	0.88	0.12

- A is a **discrete** variable; B is a **discrete** variable P(B|A) given as **conditional probability table**

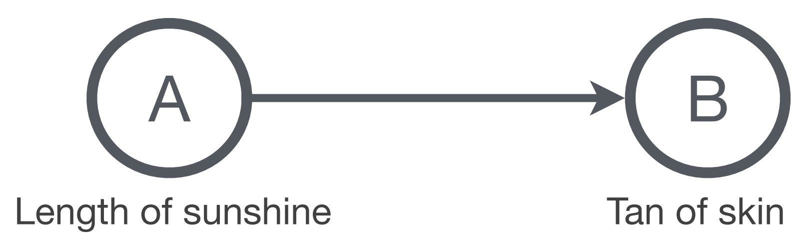


$$P(B\,|\,A)$$

	Expec	Variance
A = winter	μ1	σ1
A = spring	μ2	σ2
A = summer	μ3	σ3

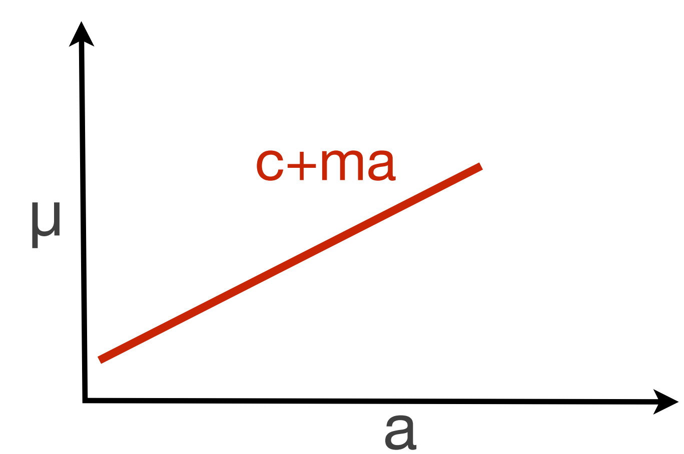


- A is a **discrete** variable; B is a **continuous** variable P(B|A) as **multiple Gaussian distributions**



$$P(B\,|\,A)$$

- A is a **continuous** variable; B is a **continuous** variable $P(B\,|\,A) \sim \mathcal{N}(c + ma, \sigma^2)$



What are bayesian networks ?

Medizinische Fakultät Heidelberg

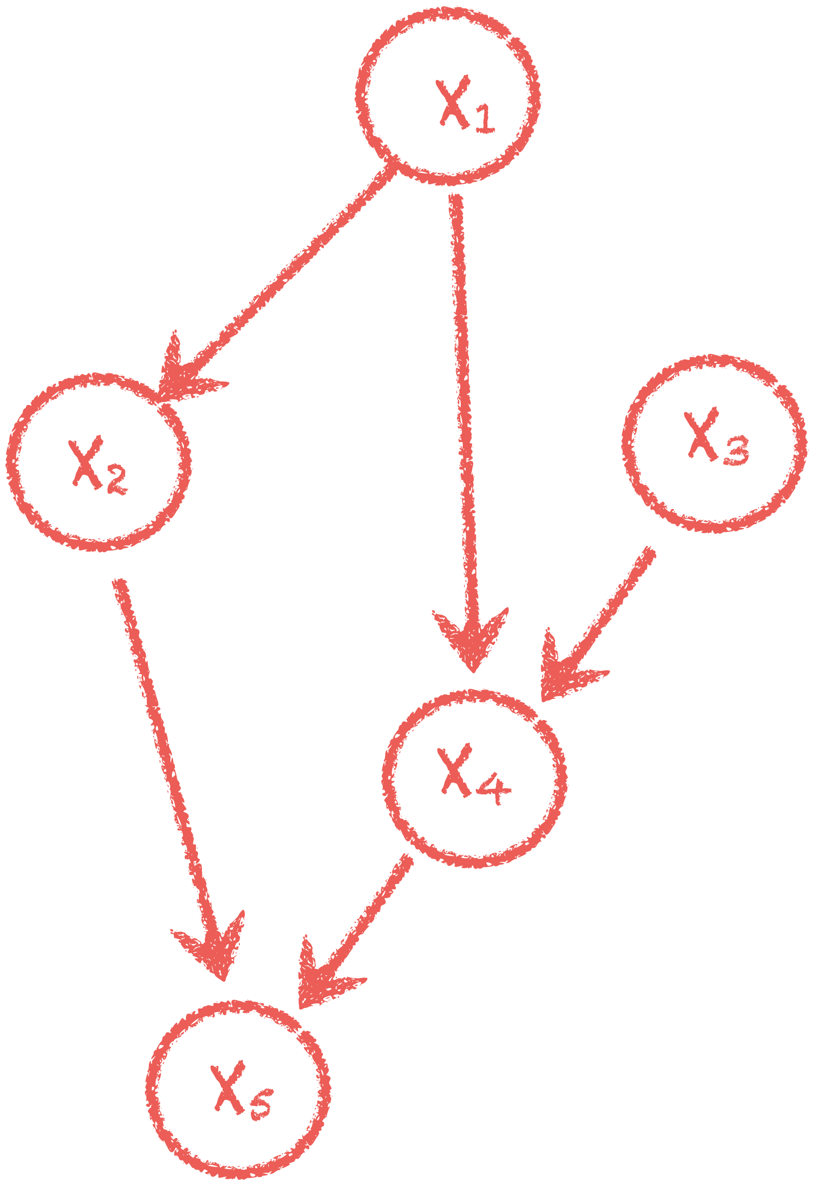
- We want to compute the **joint probability distribution (JPD)**

$$P(X_1, X_2, \ldots, X_n)$$

- Using **conditional independence**, we can write the JPD as :

$$\begin{aligned} P(X_1, X_2, ..., X_n) &= \prod_{i=1}^{n} P(X_i | X_{j \neq i}) \\ &= \prod_{i=1}^{n} P(X_i | pa(X_i)) \end{aligned}$$

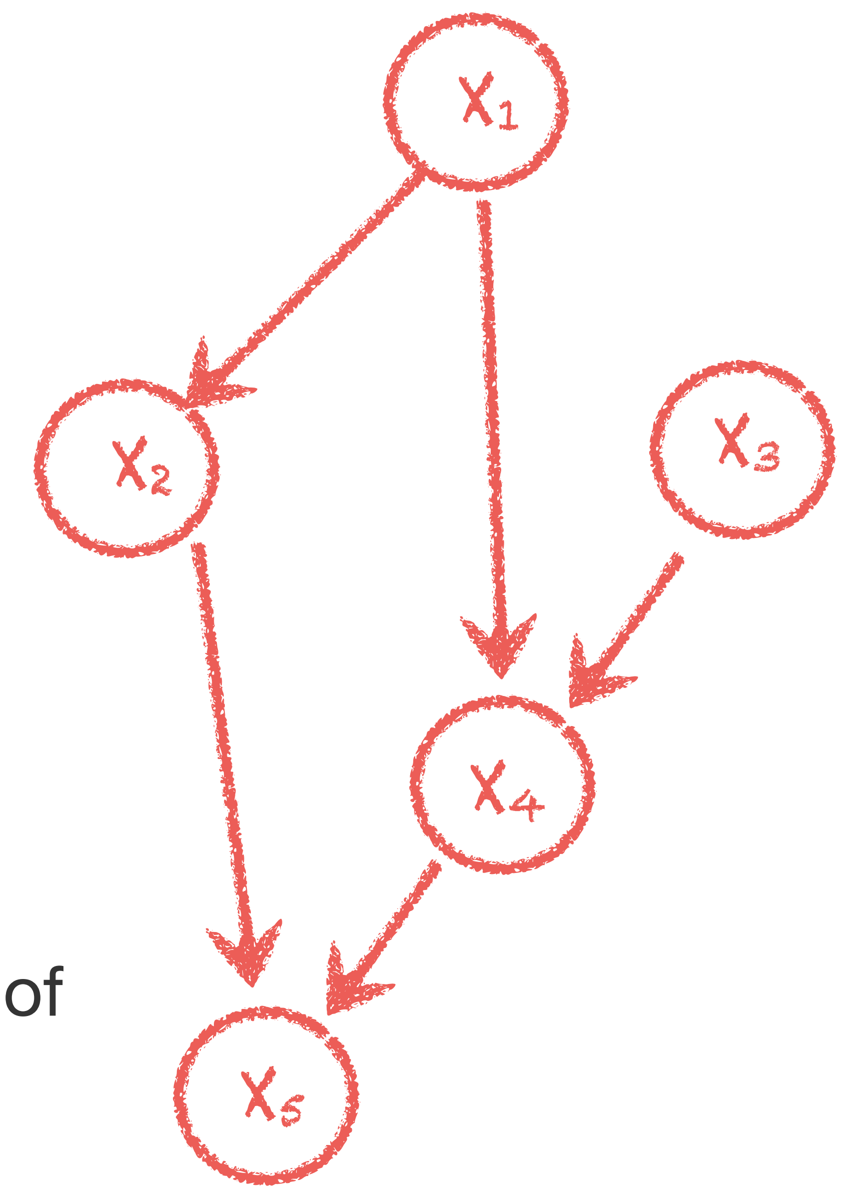
parents of X_i



[Friedman, 2004,Friedman et al., 2000]

Conditional (in)dependence

- **Conditional (in)dependence**: knowing the state of some nodes makes others nodes (in)dependent
- **Conditional independence**
  X_3 has an indirect effect on X_5; but knowing the state of X_4 makes X_5 and X_3 independent
  (if I know X_4, then X_5 does not give me additional information
  ***X_4 d*-separates X_3 and X_5** $(X_3 \perp X_5 \mid X_4)$



- **Conditional dependence**
  X_1 and X_3 are independent; but knowing the state of X_1 AND X_4 gives me additional information on X_3

[Friedman, 2004,Friedman et al., 2000]

Serial connection

- Good recommendation letter: student is probably smart! $(L \not\perp I)$
- if I know that the grade is A: knowing that the student is dumb will not given me any further indication on quality of the letter! $(L \perp I \mid G)$
- the middle node **d-separates** the 2 external ones ("blocks the flow of information")
- the state of a node only depends on its parents

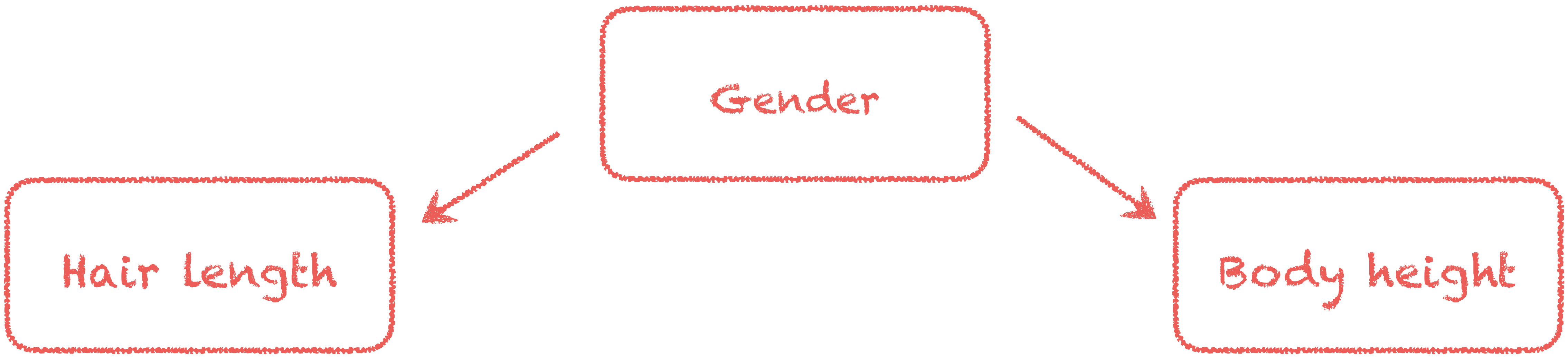
$$P(I, G, L) = P(L \mid G) \cdot P(G \mid I) \cdot P(I)$$

Converging connection

- sportiness and academic achievements are independent $(S \perp A)$
- but if I know that someone was admitted to college and is very sporty, this lowers the belief in high academic achievements.
- ***"v-structure"*** $(S \not\perp A \,|\, C)$

$$P(S, C, A) = P(S) \cdot P(A) \cdot P(C \,|\, S, A)$$

Diverging connection

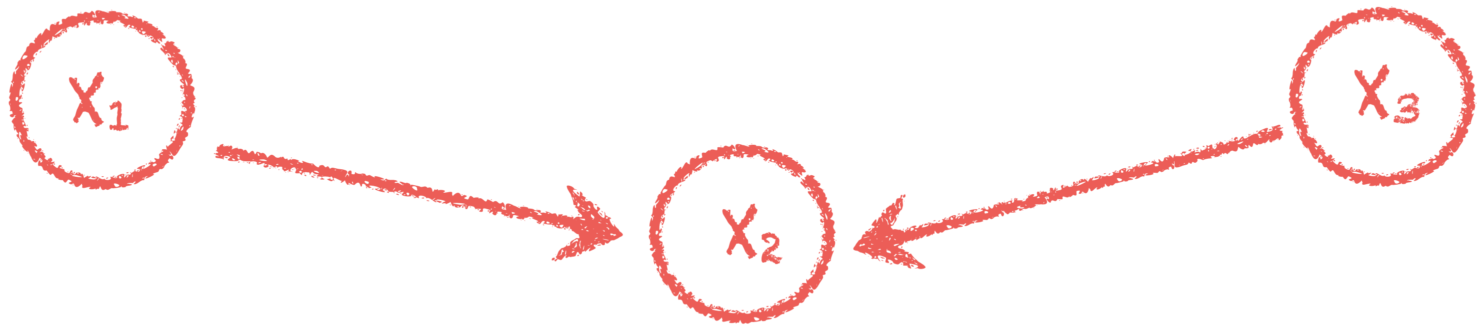


- If gender is unknown, knowing the hair length (L) influences the belief on the body height (H) (through gender!) $(L \not\perp H)$

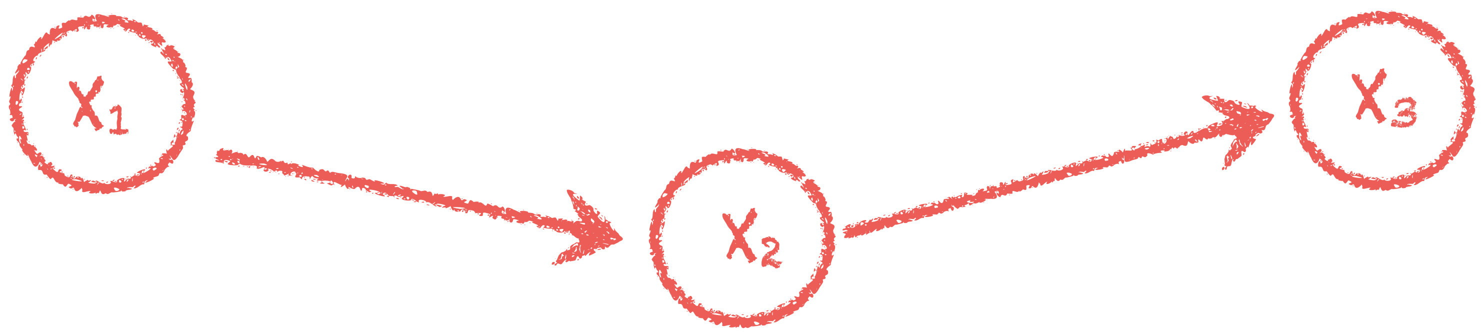
- If gender is known (man), then the length of his hair (L) gives no additional information on the body height (H)! $(L \perp H \mid G)$

$$P(G, L, H) = P(L \mid G) \cdot P(H \mid G) \cdot P(G)$$

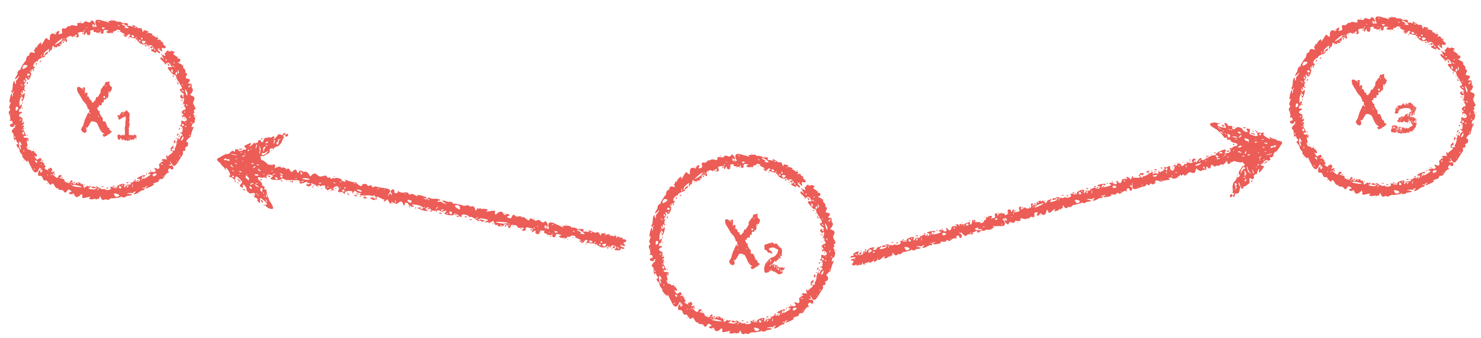
Equivalence



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



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



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

These 2 networks have same probability → *equivalence*

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

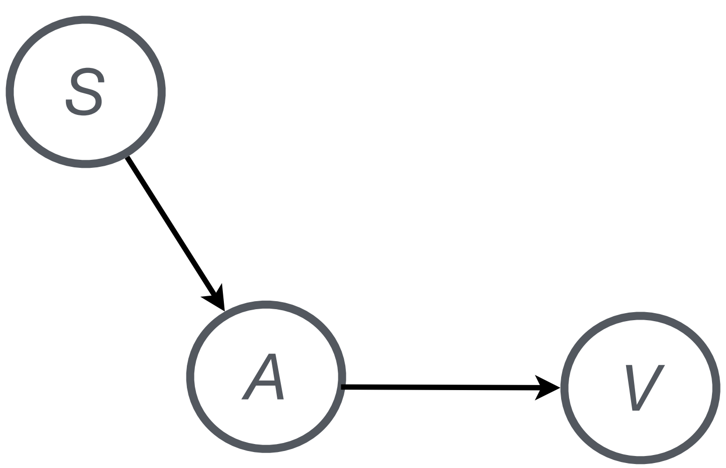
Medizinische Fakultät Heidelberg

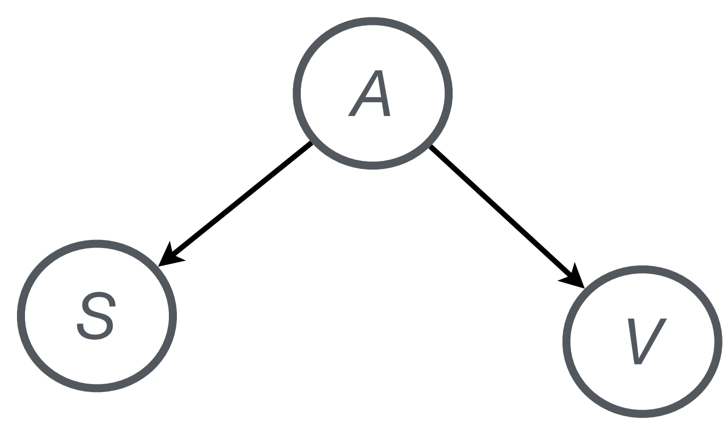
Questions ?

Test yourself!

What would be the network structure relating body size, age and vocabulary?

$S \not\perp V \qquad\qquad S \perp V \mid A$





Medizinische Fakultät Heidelberg

Usage Scenarios

Usage scenarios of BN

Medizinische Fakultät Heidelberg

1. **Inference**
   We know the graph structure of the BN and the parameters (i.e. the conditional probabilities corresponding to the edges)
   → we can perform **inference**, i.e. compute the joint probability of a certain configuration of the random variables
2. **Parameter learning**
   We know the structure of the BN and some parameters
   → we must perform **parameter learning** using **training data** to determine all parameters of the model
3. **Structure learning**
   We do NOT know the structure of the BN
   → we must perform **structure learning** using **training data** to obtain the most likely graph structure and parameters of the BN

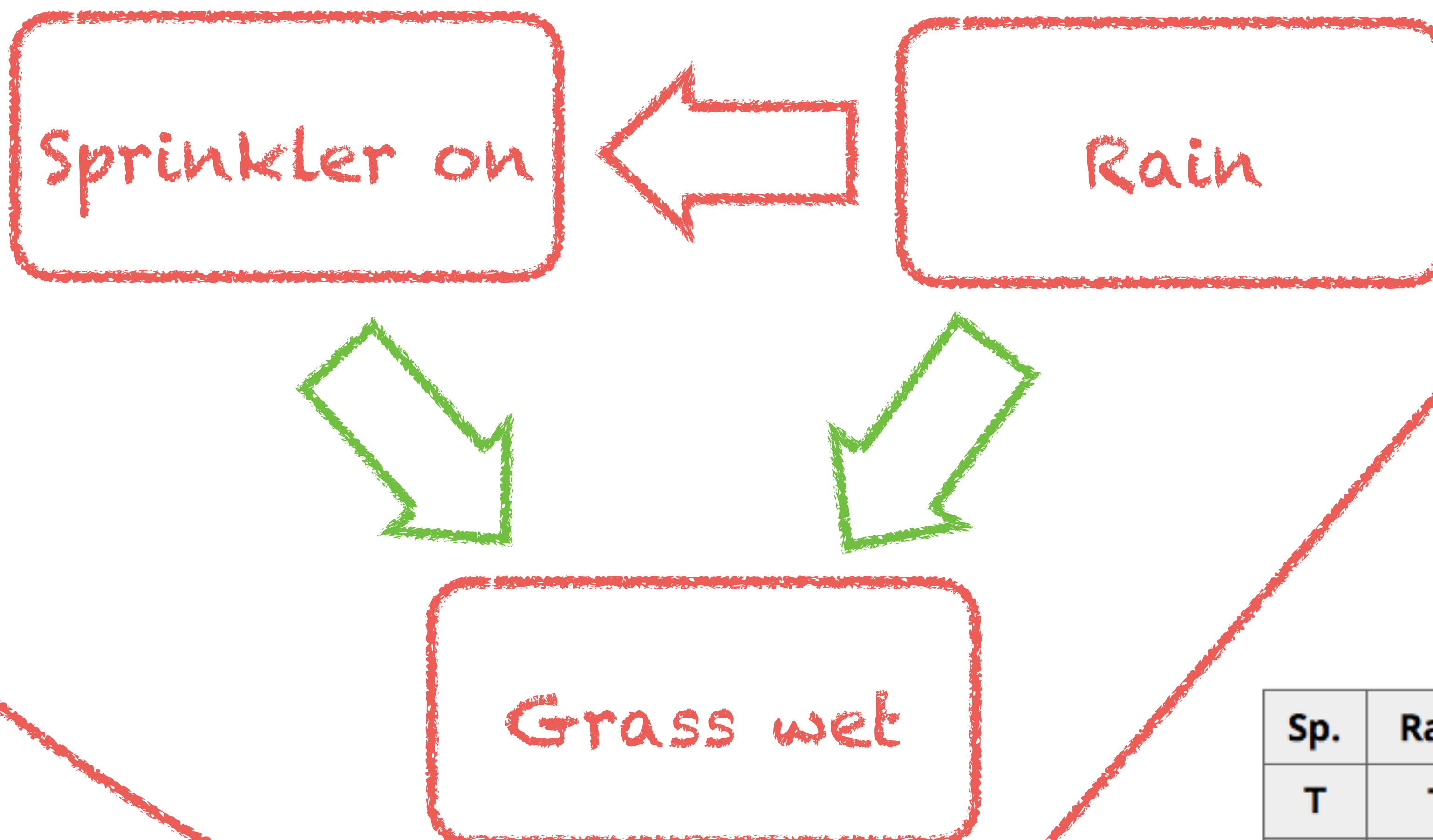
1.Inference

Sprinkler

Rain	T	F
T	0.01	0.99
F	0.4	0.6

Rain

T	F
0.2	0.8



Grass wet

Sp.	Rain	T	F
T	T	0.99	0.01
F	T	0.8	0.2
T	F	0.9	0.1
F	F	0	1

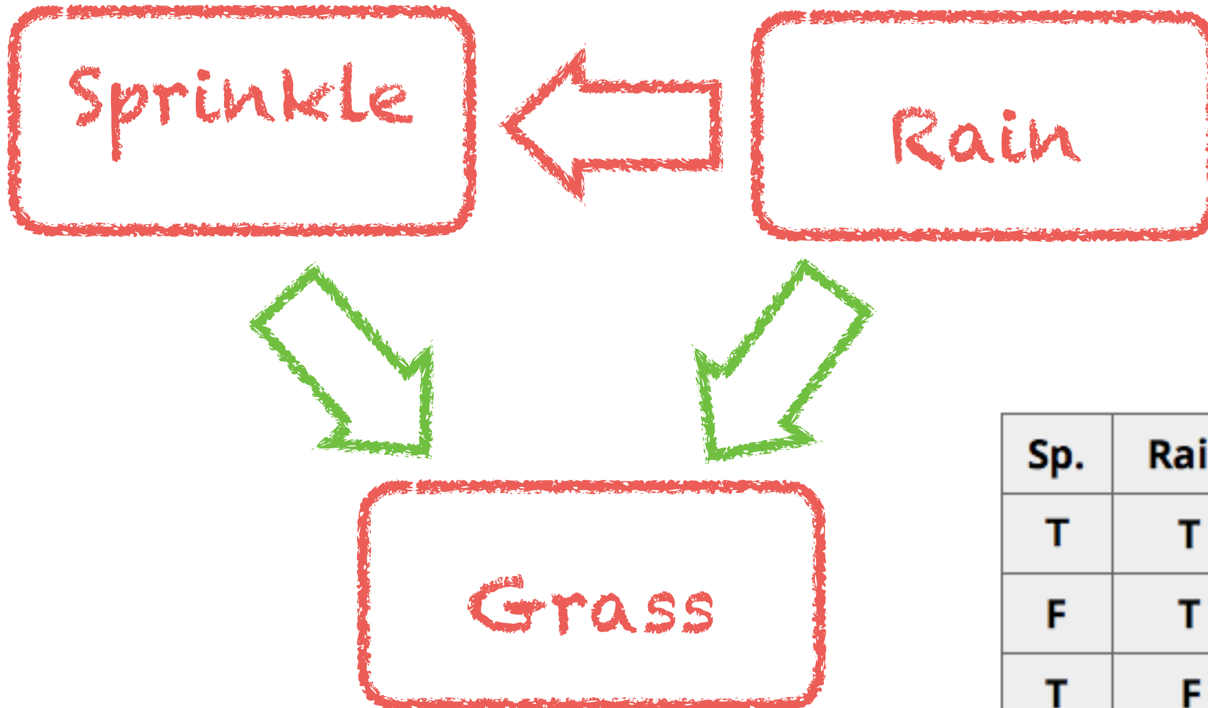
Joint probability

$$P(W,S,R) = P(W|S,R) \times P(S|R) \times P(R)$$

1.Inference

Sprinkler

Rain	T	F
T	0.01	0.99
F	0.4	0.6



Rain

T	F
0.2	0.8

Grass wet

Sp.	Rain	T	F
T	T	0.99	0.01
F	T	0.8	0.2
T	F	0.9	0.1
F	F	0	1

Joint probability:

$$P(W, S, R) = P(W|S, R) \times P(S|R) \times P(R)$$

What is the probability that the grass is wet, the sprinkler on and it rains ?

$$\begin{aligned} P(W=1, S=1, R=1) &= P(W=1|S=1, R=1)P(S=1|R=1)P(R=1) \\ &= 0.99 \times 0.01 \times 0.2 \\ &= 0.00198 \end{aligned}$$

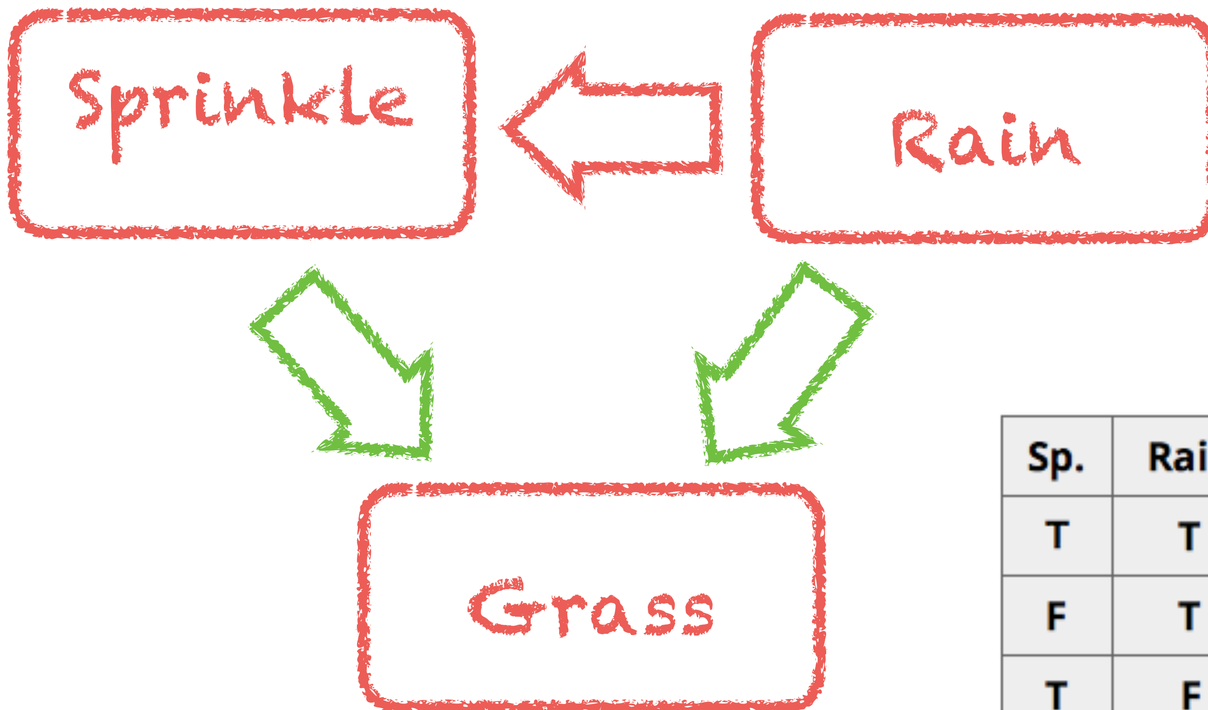
1.Inference

Rain

T	F
0.2	0.8

Sprinkler

Rain	T	F
T	0.01	0.99
F	0.4	0.6



Grass wet

Sp.	Rain	T	F
T	T	0.99	0.01
F	T	0.8	0.2
T	F	0.9	0.1
F	F	0	1

What is the probability that the grass is wet given that the sprinkler is on ?

→ **sum over marginal variable rain (unobserved variable)**

$$
\begin{aligned}
P(W=1|S=1) &= \frac{1}{P(S=1)} \sum_{r\in 0,1} P(W=1,S=1,r) && (1)\\
&= \frac{\sum_{r\in 0,1} P(W=1|S=1,r)P(S=1|r)P(r)}{\sum_{r\in 0,1} P(S=1|r)P(r)} && (2)\\
&= \frac{0.99\cdot 0.01\cdot 0.2 + 0.9\cdot 0.4\cdot 0.8}{0.01\cdot 0.2+0.4\cdot 0.8} && (3)\\
&= 0.9 && (4)
\end{aligned}
$$

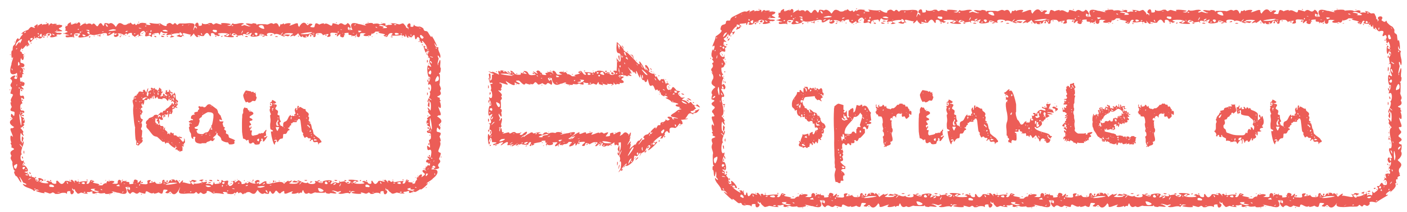
2. Parameter learning

Medizinische Fakultät Heidelberg

- Suppose we have the **structure** of the network and **some parameters**; other parameters are missing and have to be learned from observed data

θ_1

Rain: T	Rain: F
0.2	0.8



Rain \ Sprinkler	T	F
T	θ_2	
F		

	Day 1	Day 2	Day 3	Day 4	Day 5
R	Y	Y	N	N	Y
S	Y	N	N	Y	Y

Rain \ Sprinkler	T	F	
T	n_1	n_2	m_1
F	n_3	n_4	m_2

- Procedure: Maximize the **likelihood of the observed data** given the parameters

$$L(\theta) \equiv P(D \mid \theta) \qquad \theta_{ML} = \text{argmax } L(\theta)$$

2. Parameter learning

Medizinische Fakultät Heidelberg

Rain

θ_1

T	F
0.2	0.8

Rain $\Rightarrow$ Sprinkler on

Sprinkler

Rain	T	F
T	α	$1-\alpha$
F	β	$1-\beta$

θ_2

$$L(\theta) \equiv P(D\,|\,\theta) = \prod_{i=1}^{N} P(s_i\,|\,r_i,\theta)\cdot P(r_i,\theta)$$

$$= \left(\prod_{i=1}^{N} P(s_i\,|\,r_i,\theta_2)\right)\cdot\left(\prod_{i=1}^{N} P(r_i,\theta_1)\right)$$

$$= \left(\alpha^{n_1}\,(1-\alpha)^{n_2}\,\beta^{n_3}\,(1-\beta)^{n_4}\right)\cdot\left(0.2^{m_1}\,0.8^{m_2}\right)$$

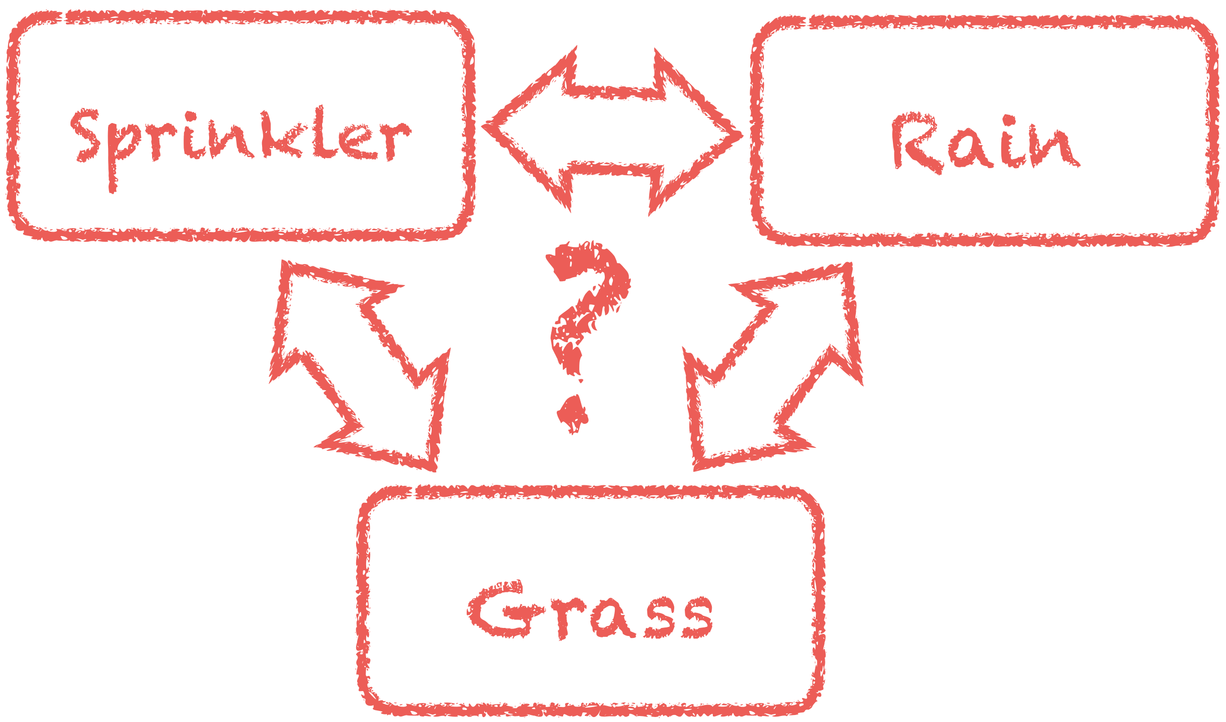
Sprinkler

Rain	T	F	
T	n_1	n_2	m_1
F	n_3	n_4	m_2

Solution : $\alpha = \dfrac{n_1}{m_1}$ $\beta = \dfrac{n_3}{m_2}$ ***Maximum Likelihood Solution***

3. Structure learning

What is the most likely network given the observed data ?

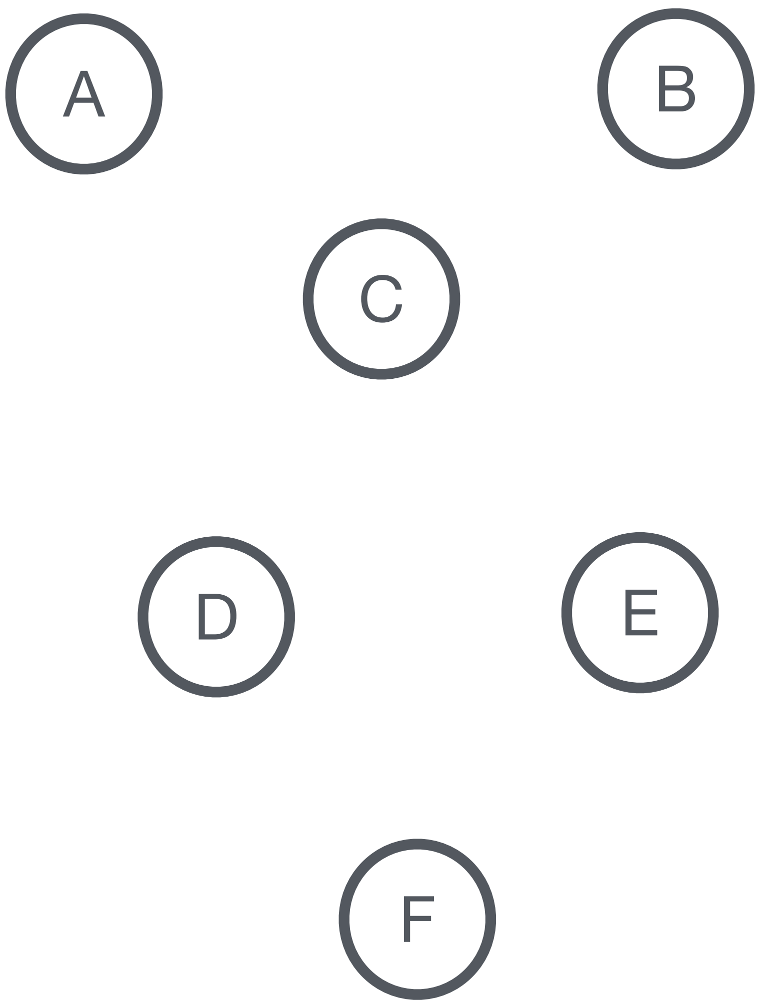


Day	Rain	Spr.	Grass wet ?
1	yes	no	no
2	no	no	no
3	no	yes	yes
4	yes	no	yes
5	...	...	...

Important assumption : all observations are sampled from the same random variable !
However, (unobserved) confounding variables could violate this assumption (influence of seasons ?)

3. Structure learning

Medizinische Fakultät Heidelberg



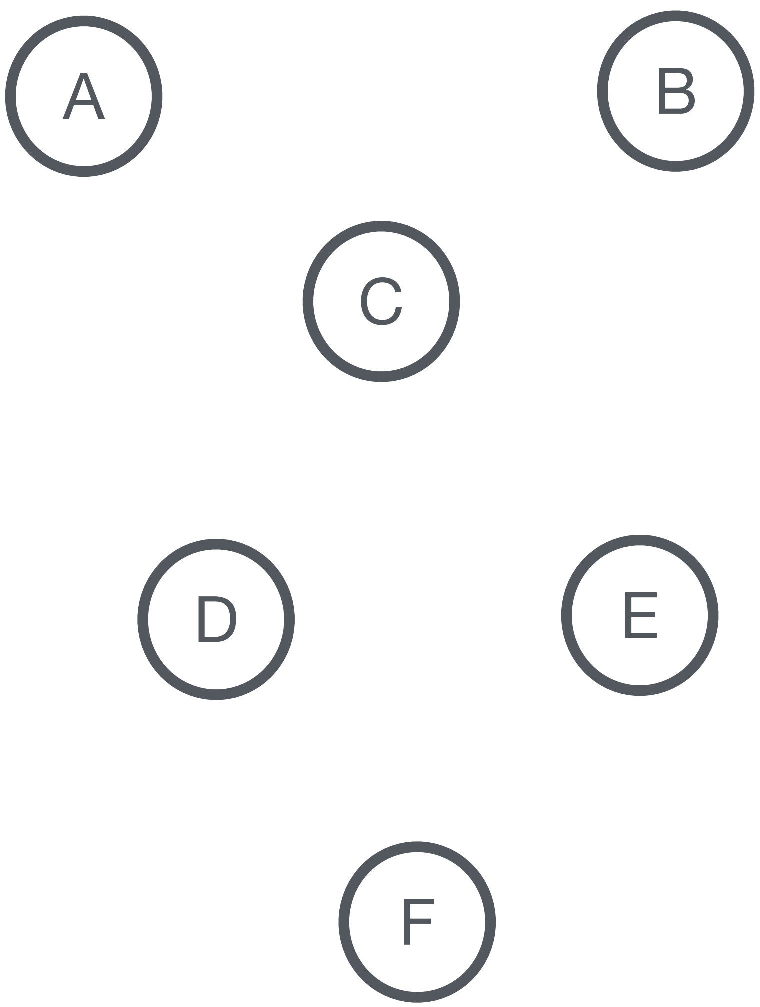
$\Longleftarrow$

Conditional independence tests

$$p_{A \perp\!\!\!\perp B}$$
$$p_{A \perp\!\!\!\perp E}$$
$$\vdots$$
$$p_{A \perp\!\!\!\perp E | C}$$
$$p_{B \perp\!\!\!\perp E | D}$$
$$p_{A \perp\!\!\!\perp B | C}$$
$$p_{E \perp\!\!\!\perp F | D}$$
$$\vdots$$
$$p_{F \perp\!\!\!\perp E | CD}$$

[M. Scutari]

3. Structure learning



$\Longleftarrow$

Graphical separation

$$
\begin{aligned}
&A \perp\!\!\!\perp_G B \\
&A \perp\!\!\!\perp_G D \mid C \\
&B \perp\!\!\!\perp_G D \mid C \\
&A \perp\!\!\!\perp_G E \mid C \\
&B \perp\!\!\!\perp_G E \mid C \\
&D \perp\!\!\!\perp_G E \mid C \\
&C \perp\!\!\!\perp_G F \mid D \\
&\ldots
\end{aligned}
$$

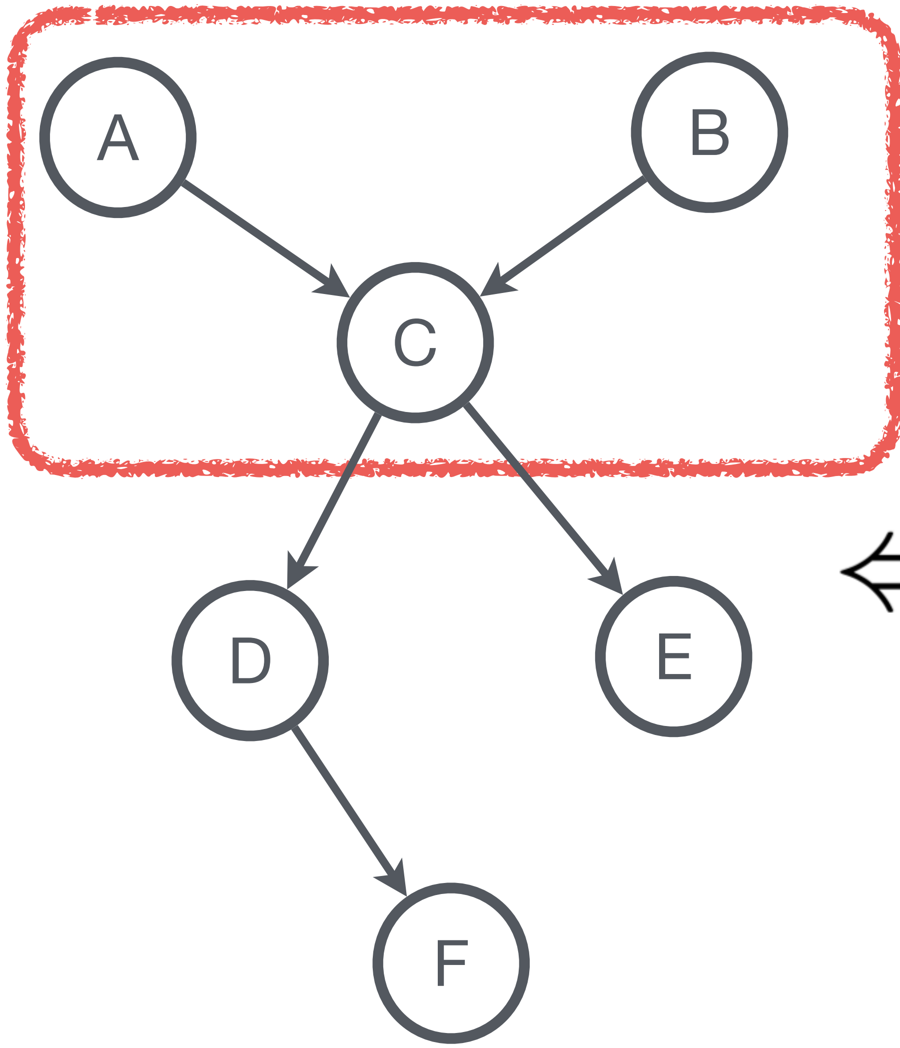
$\Longleftarrow$

Conditional independence tests

$$
\begin{gathered}
p_{A \perp\!\!\!\perp B} > \alpha \\
p_{A \perp\!\!\!\perp E} > \alpha \\
\vdots \\
p_{A \perp\!\!\!\perp E|C} > \alpha \\
p_{B \perp\!\!\!\perp E|D} > \alpha \\
p_{A \perp\!\!\!\perp B|C} < \alpha \\
p_{E \perp\!\!\!\perp F|D} > \alpha \\
\vdots \\
p_{F \perp\!\!\!\perp E|CD} > \alpha
\end{gathered}
$$

[M. Scutari]

3. Structure learning



$\Longleftarrow$

Graphical separation

$$A \perp\!\!\!\perp_G B$$
$$A \perp\!\!\!\perp_G D \mid C$$
$$B \perp\!\!\!\perp_G D \mid C$$
$$A \perp\!\!\!\perp_G E \mid C$$
$$B \perp\!\!\!\perp_G E \mid C$$
$$D \perp\!\!\!\perp_G E \mid C$$
$$C \perp\!\!\!\perp_G F \mid D$$
$$\ldots$$

$\Longleftarrow$

Conditional independence tests

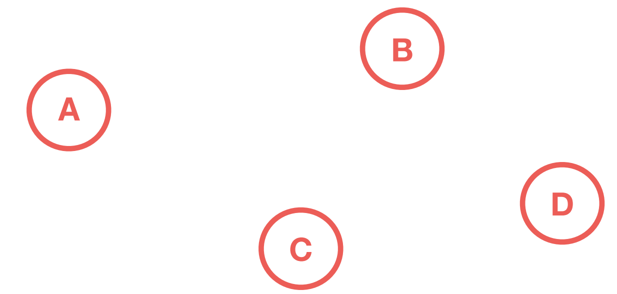
$$\boxed{p_{A \perp\!\!\!\perp B} > \alpha}$$
$$p_{A \perp\!\!\!\perp E} > \alpha$$
$$\vdots$$
$$p_{A \perp\!\!\!\perp E \mid C} > \alpha$$
$$p_{B \perp\!\!\!\perp E \mid D} > \alpha$$
$$\boxed{p_{A \perp\!\!\!\perp B \mid C} < \alpha}$$
$$p_{E \perp\!\!\!\perp F \mid D} > \alpha$$
$$\vdots$$
$$p_{F \perp\!\!\!\perp E \mid CD} > \alpha$$

[M. Scutari]

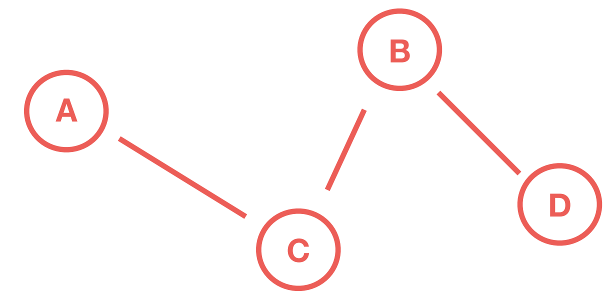
3. Structure learning

- ***"constrained based methods"***

1. identify the pairs of nodes which **cannot be made conditional independent** (e.g. through partial correlation)

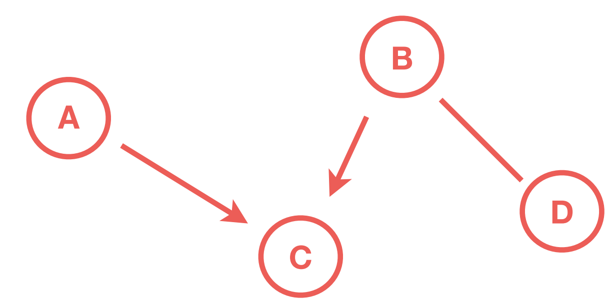


2. relate these nodes by **undirected** edge



3. if C does not *d*-separate A and B, then form a *v*-structure

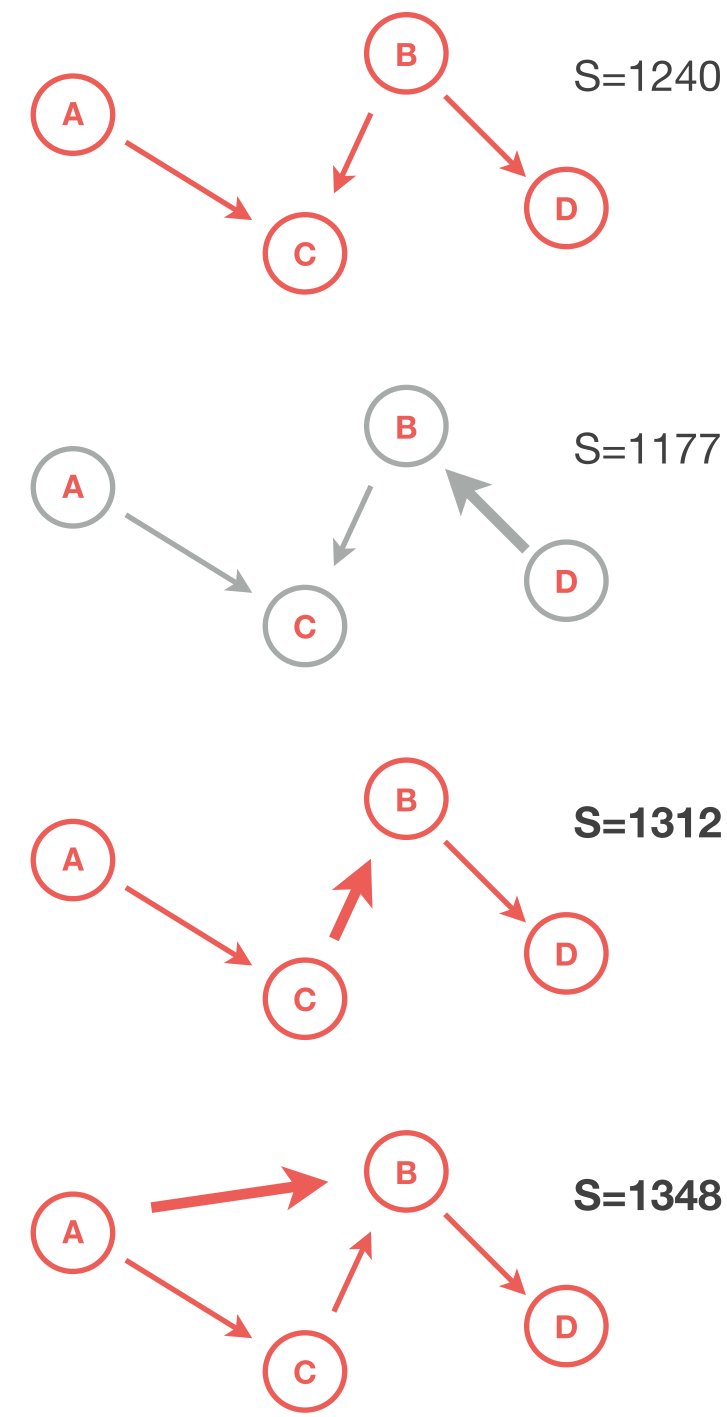
4. apply heuristics to (possibly) direct the still undirectedarcs
(no cycles!)



3. Structure learning

Medizinische Fakultät Heidelberg

- ***"score based methods"***
  - assign a **likelihood score** to each possible network
  - select the network with the highest likelihood score
- heuristics needed to handle the exponential number of possible networks !
  - hill-climbing : modify the current network slightly and check if this improves the score
  - improvements to **avoid local optima**:
    - *tabu search*: allow search to proceed around local optimum, **avoiding previous tested solutions**
    - *simulated annealing*: several random initialization, allow steps which degrade the score



3. Structure learning

- Akaike Information Criterion (*AIC*) / Bayesian information criterion (BIC)

$$BIC = \sum_{i=1}^{n} \log P(X_i | pa(X_i)) - \frac{d}{2} \log(n)$$

n = number of nodes

Likelihood of data, given learned parameters

d = number of edges, complexity penalty

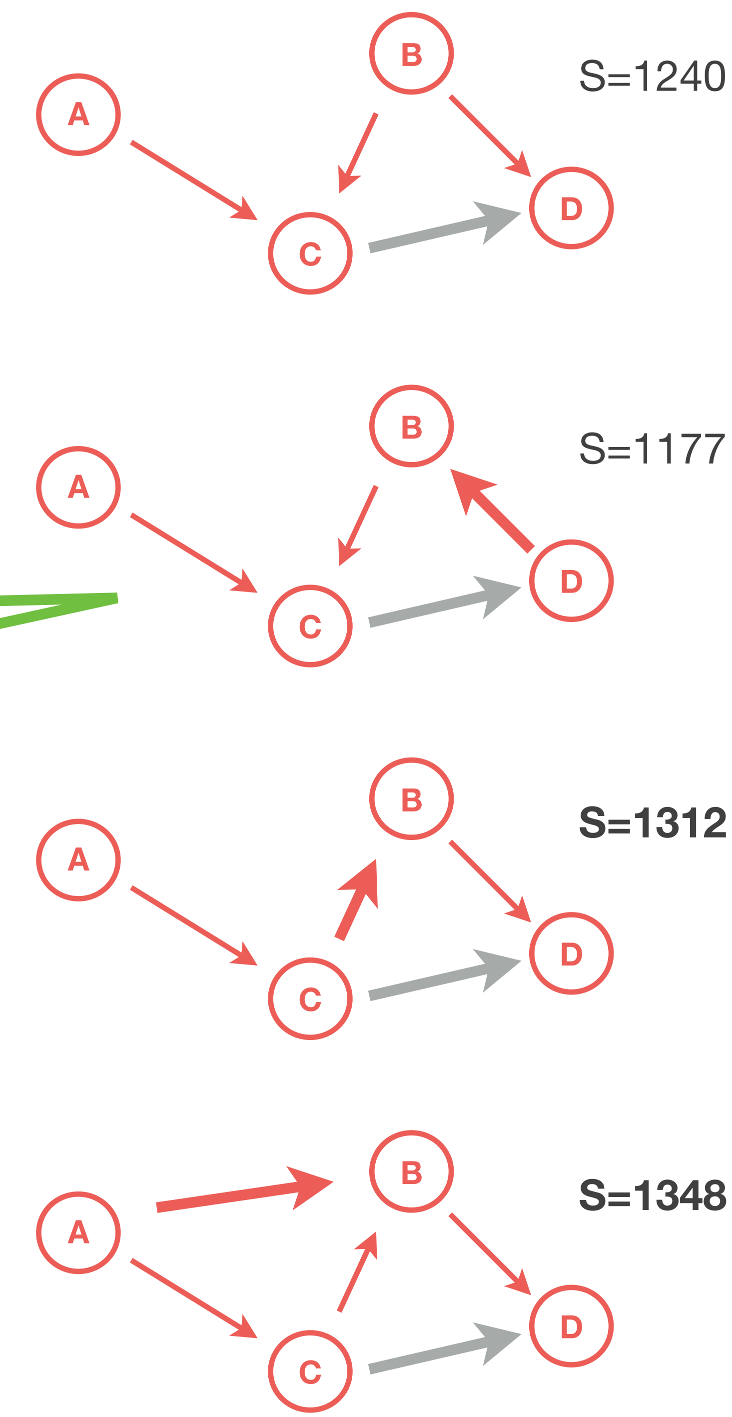
- **TABU search**
  - Optimize score by adding / removing / redirecting arcs
  - Prohibit the last *k* changes ("memory effect")
  - Try additional m steps when hitting a maximum (avoid local optima)

- **Bootstrapping** :
  - randomly sample a subset of observations *N* times
  - average the network
  - determine strength of edge / direction as the proportion of observations

3. Structure learning

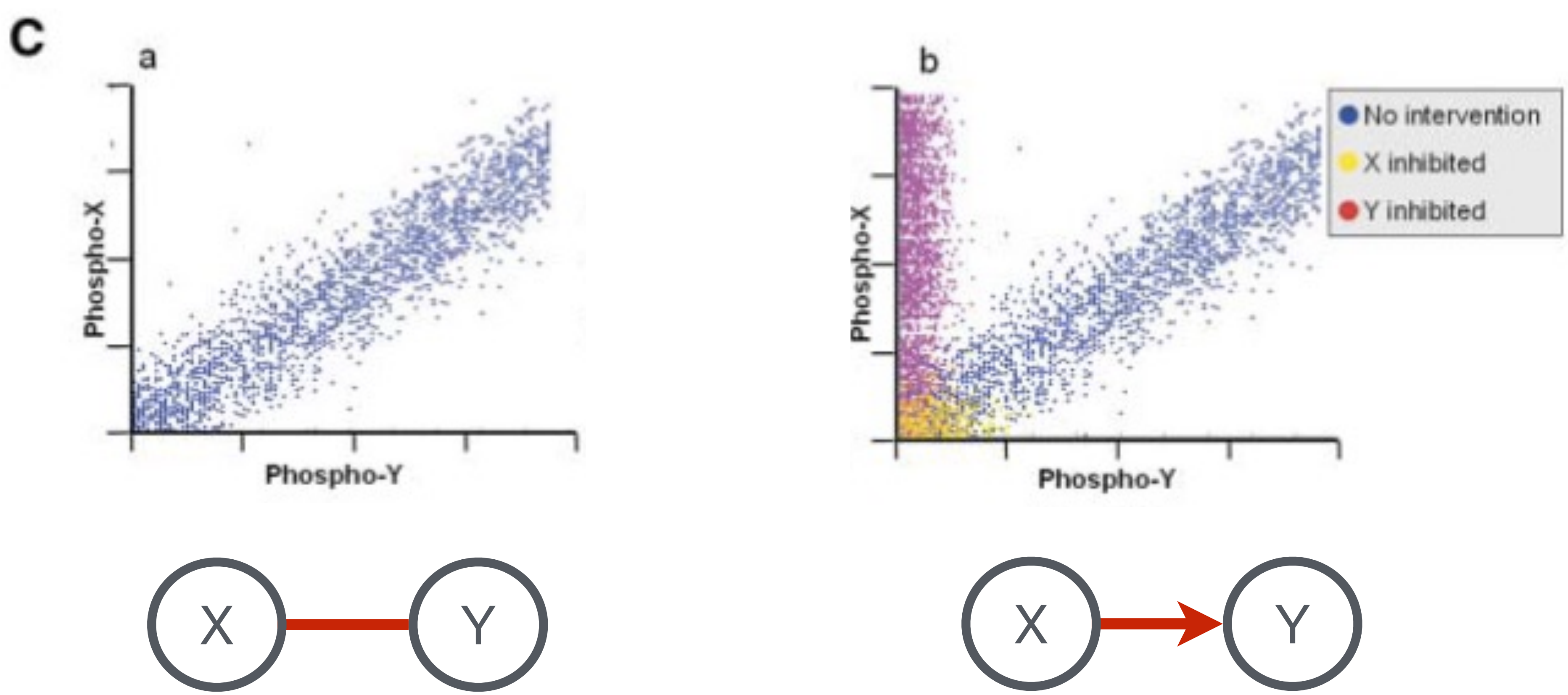
- some edges can be
  - **whitelisted**: they should be present in the final network, based on literature evidence, etc…
  - **blacklisted**: these should NOT appear in the network (unrealistic relations)

this move is not possible as it would create a cycle B → C → D



Using intervention data

- A strong correlation between X and Y does not give indication whether X → Y, Y → X or none of both
- We can use **intervention data** to resolve the dependency, by altering the state of X and checking the effect on Y, and vice versa

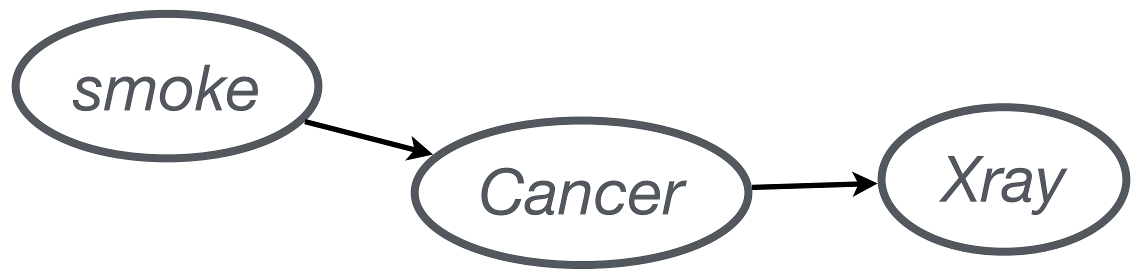


[Sachs et al., 2005]

Questions ?

Test yourself!

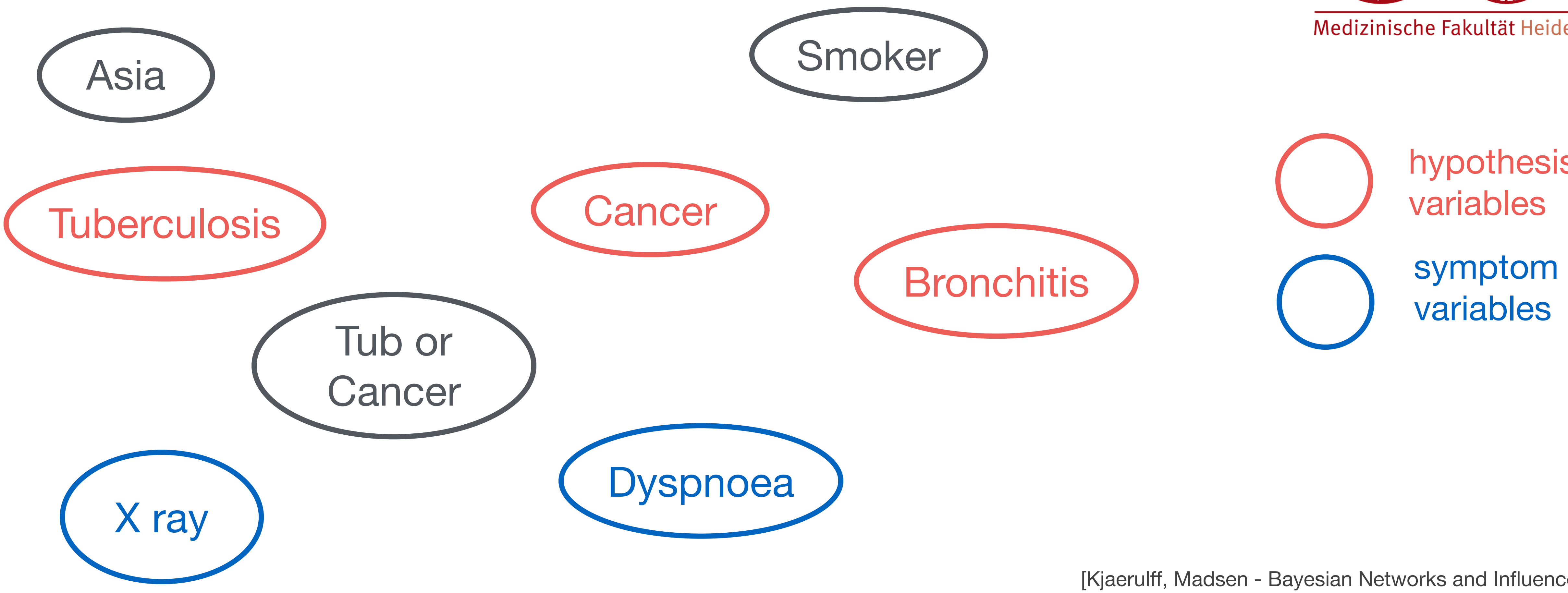
Does the smoking status influence the probability of Xray outcome?



Yes, if we do not know the cancer status
No, if we now the cancer status

Application 1: Diagnosis system
- inference -

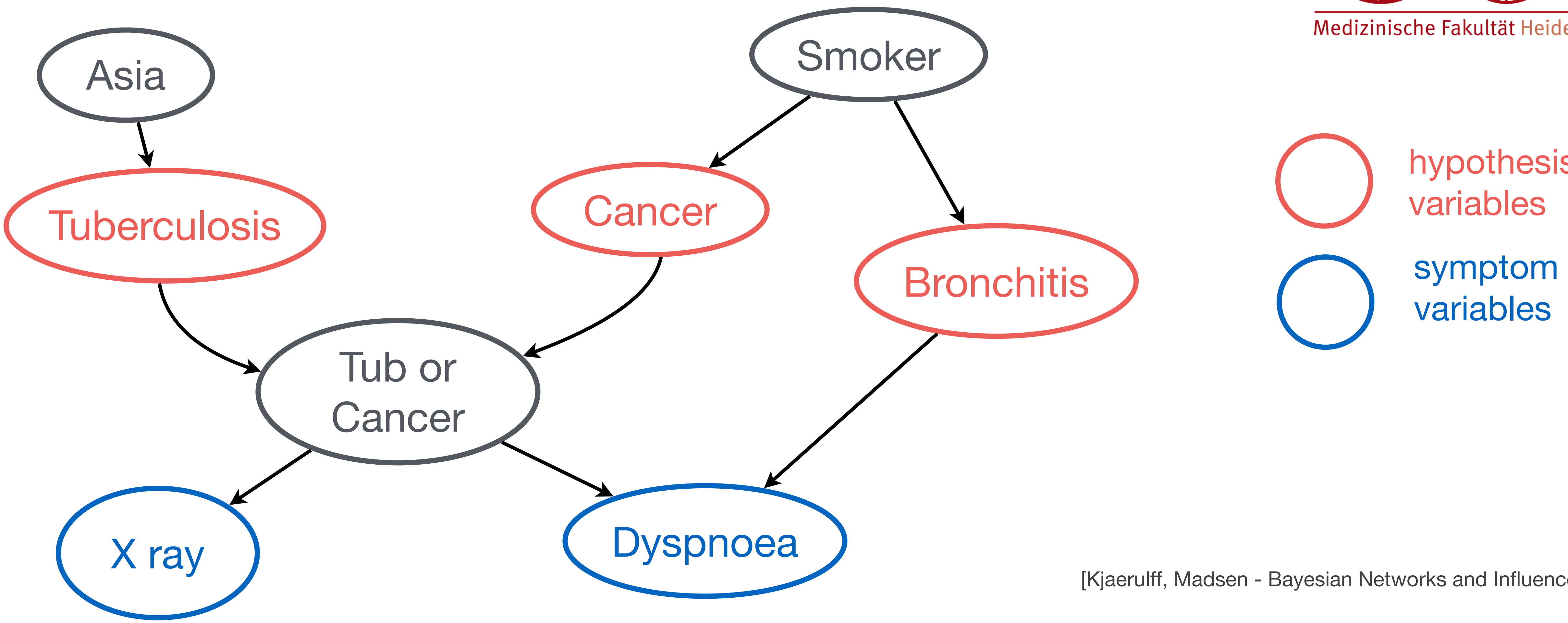
Diagnosis



[Kjaerulff, Madsen - Bayesian Networks and Influence

- A physician wants to diagnose cancer/tuberculosis/bronchitis using
  - presence of **symptoms** (X-ray positive / Dyspnoea)
  - clinical information (travel to Asia / smoking history)

Network structure

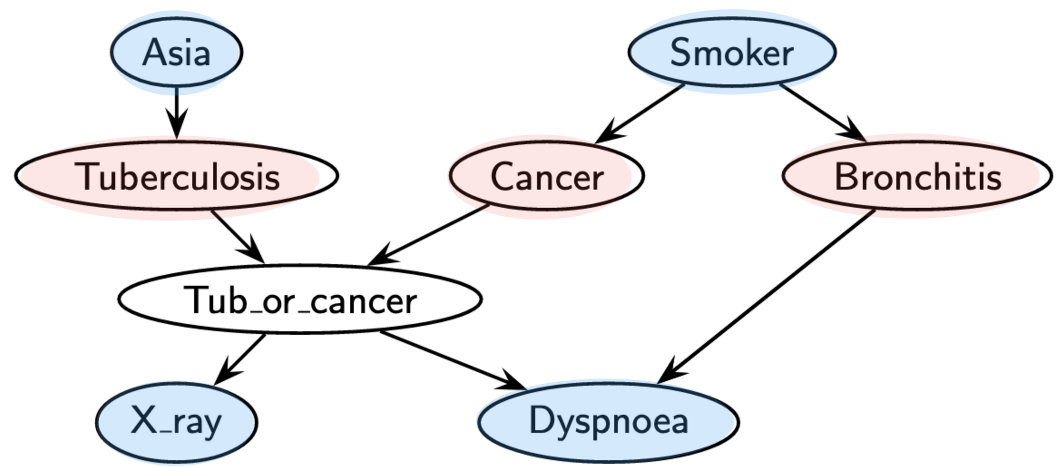


[Kjaerulff, Madsen - Bayesian Networks and Influence

- The structure is built from **prior knowledge** about the mechanisms and symptoms of each disease

Scenarios

Diagnosis:
predict the state of the hypothesis variables given evidence



Research of cause:
predict the probability of clinical state or symptoms given the hypothesis variables

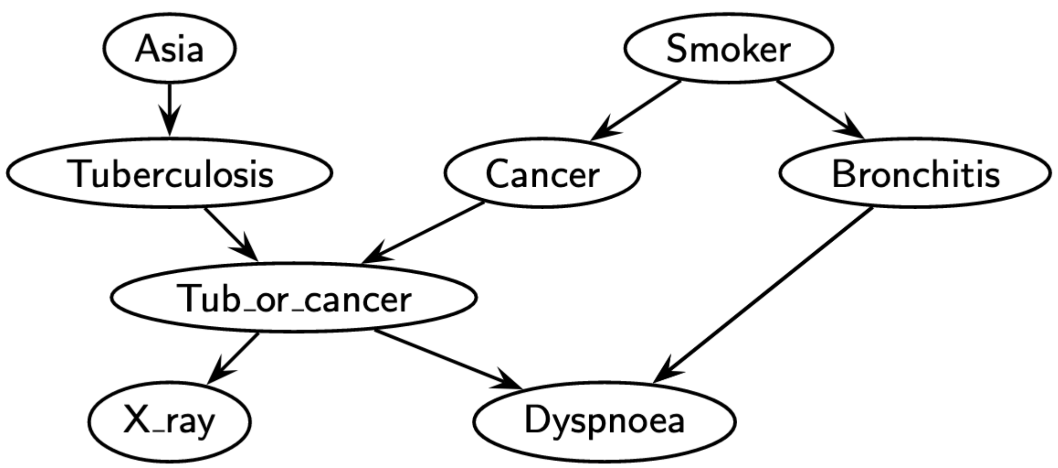
*"What is the probability that the patients has **bronchitis**, given that he went to **Asia** and has **dyspnoea**?"*

*"What is the probability that the patients went to **Asia**, given that he has **tuberculosis**?"*

Network parameters

Medizinische Fakultät Heidelberg

$$P(A) = (0.99, 0.01),\ P(S) = (0.5, 0.5)$$



	$B = \text{no}$		$B = \text{yes}$	
	$E = \text{no}$	$E = \text{yes}$	$E = \text{no}$	$E = \text{yes}$
$D = \text{no}$	0.9	0.3	0.2	0.1
$D = \text{yes}$	0.3	0.7	0.8	0.9

$P(L\|S)$	$S = \text{no}$	$S = \text{yes}$
$L = \text{no}$	0.99	0.9
$L = \text{yes}$	0.01	0.1

$P(B\|S)$	$S = \text{no}$	$S = \text{yes}$
$B = \text{no}$	0.7	0.4
$B = \text{yes}$	0.3	0.6

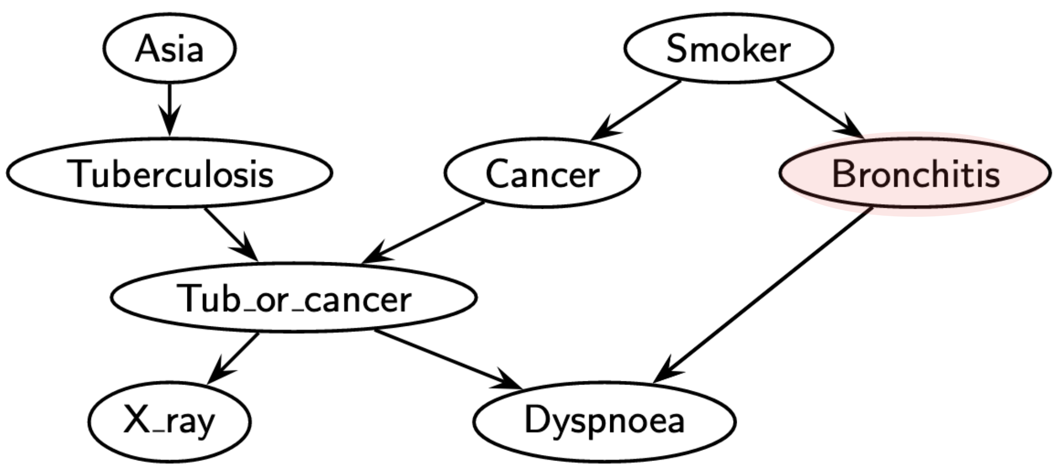
L = (lung) cancer
E = evidence (T or C)

$P(T\|A)$	$A = \text{no}$	$A = \text{yes}$
$T = \text{no}$	0.99	0.95
$T = \text{yes}$	0.01	0.05

$P(X\|E)$	$E = \text{no}$	$E = \text{yes}$
$X = \text{no}$	0.95	0.02
$X = \text{yes}$	0.05	0.98

Network parameters

$$P(A) = (0.99, 0.01),\ P(S) = (0.5, 0.5)$$



	B = no		B = yes	
	E = no	E = yes	E = no	E = yes
D = no	0.9	0.3	0.2	0.1
D = yes	0.3	0.7	0.8	0.9

P(L|S)	S = no	S = yes
L = no	0.99	0.9
L = yes	0.01	0.1

P(B|S)	S = no	S = yes
B = no	0.7	0.4
B = yes	0.3	0.6

P(T|A)	A = no	A = yes
T = no	0.99	0.95
T = yes	0.01	0.05

P(X|E)	E = no	E = yes
X = no	0.95	0.02
X = yes	0.05	0.98

- Probability of bronchitis?

$$P(B=1) = \sum_{S,L,D,E,A,T,X} P(B=1, S, L, D, E, A, T, X)$$

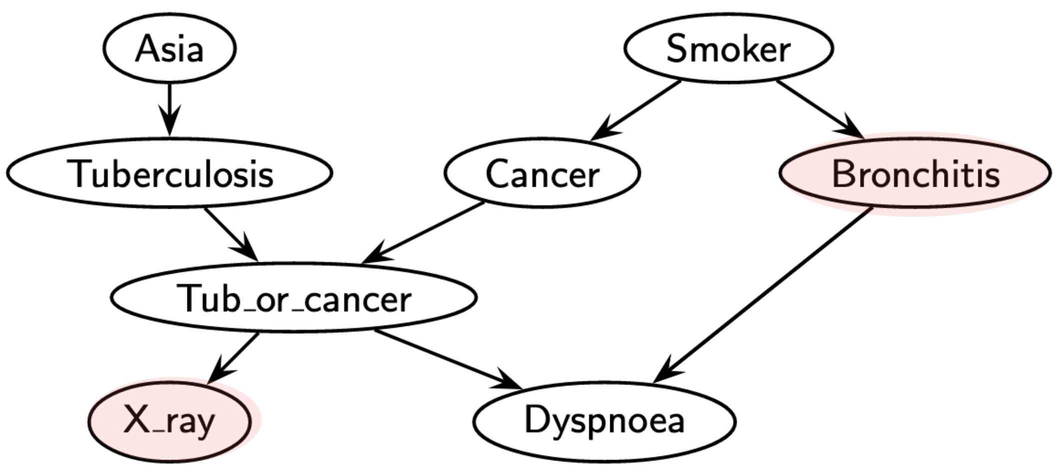
$$= \sum_{S,L,D,E,A,T,X} P(X\,|\,E)P(D\,|\,E, B=1)P(B=1\,|\,S)P(L\,|\,S)P(E\,|\,C,T)P(T\,|\,A)P(S)P(A)$$

$$= 0.45$$

Network parameters

Medizinische Fakultät Heidelberg

$$P(A) = (0.99, 0.01),\ P(S) = (0.5, 0.5)$$



	B = no		B = yes	
	E = no	E = yes	E = no	E = yes
D = no	0.9	0.3	0.2	0.1
D = yes	0.3	0.7	0.8	0.9

P(L|S)	S = no	S = yes
L = no	0.99	0.9
L = yes	0.01	0.1

P(B|S)	S = no	S = yes
B = no	0.7	0.4
B = yes	0.3	0.6

P(T|A)	A = no	A = yes
T = no	0.99	0.95
T = yes	0.01	0.05

P(X|E)	E = no	E = yes
X = no	0.95	0.02
X = yes	0.05	0.98

- Probability of bronchitis given that the X-ray is positive?

$$P(B=1\,|\,X=1) = \frac{1}{P(X=1)} \sum_{S,L,D,E,A,T} P(B=1, S, L, D, E, A, T, X=1)$$

$$= \frac{1}{P(X=1)} \sum_{S,L,D,E,A,T} P(X=1\,|\,E)P(D\,|\,E,B=1)P(B=1\,|\,S)P(L\,|\,S)P(E\,|\,C,T)P(T\,|\,A)P(S)P(A)$$

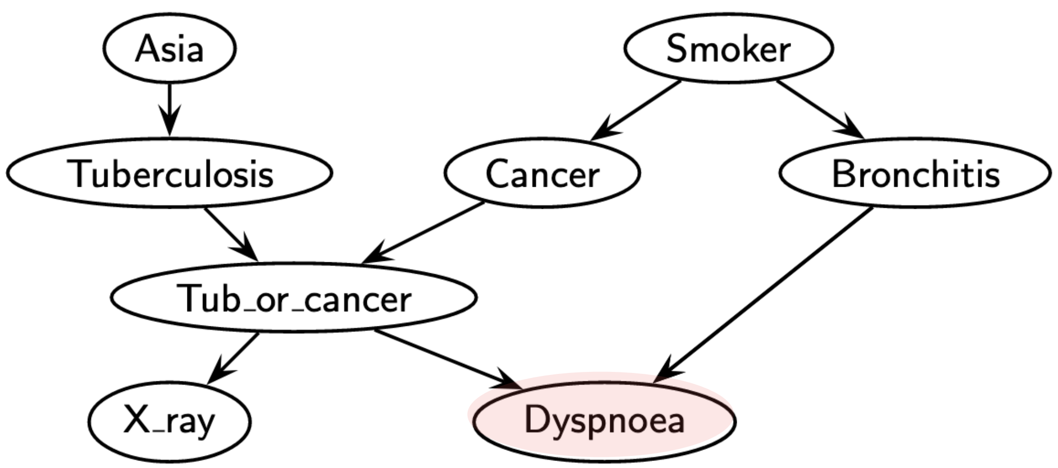
$$= 0.5$$

Why does the fact that the x-ray is positive increase the probability of bronchitis?

Network parameters

Medizinische Fakultät Heidelberg

$P(A) = (0.99, 0.01),\ P(S) = (0.5, 0.5)$



	B = no		B = yes	
	E = no	E = yes	E = no	E = yes
D = no	0.9	0.3	0.2	0.1
D = yes	0.3	0.7	0.8	0.9

P(L | S)	S = no	S = yes
L = no	0.99	0.9
L = yes	0.01	0.1

P(B | S)	S = no	S = yes
B = no	0.7	0.4
B = yes	0.3	0.6

P(T | A)	A = no	A = yes
T = no	0.99	0.95
T = yes	0.01	0.05

P(X | E)	E = no	E = yes
X = no	0.95	0.02
X = yes	0.05	0.98

- Probability of dyspnoea?

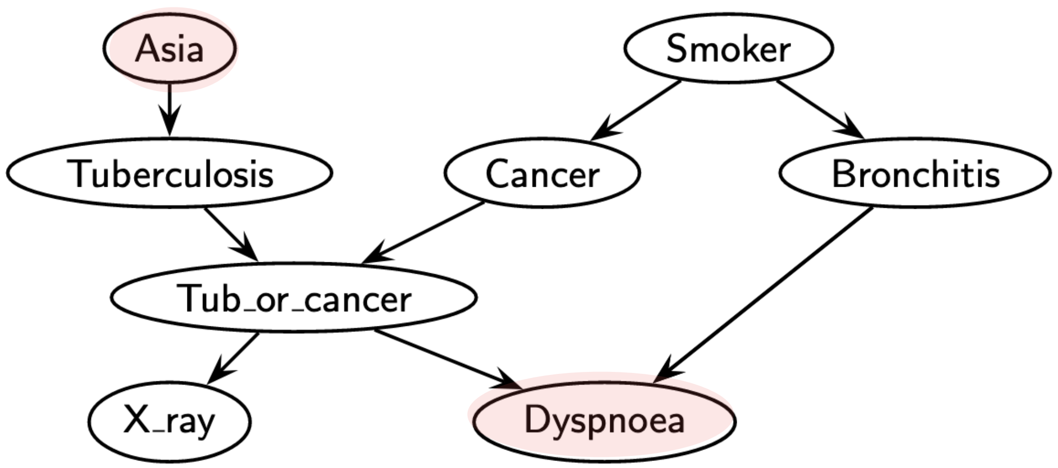
$$P(D=1) = \sum_{B,S,L,X,E,A,T} P(B,S,L,D=1,E,A,T,X)$$

$$= \sum_{B,S,L,X,E,A,T} P(X \mid E) P(D=1 \mid E,B) P(B \mid S) P(L \mid S) P(E \mid C,T) P(T \mid A) P(S) P(A)$$

$$= 0.44$$

Network parameters

$P(A) = (0.99, 0.01),\ P(S) = (0.5, 0.5)$



	B = no		B = yes	
	E = no	E = yes	E = no	E = yes
D = no	0.9	0.3	0.2	0.1
D = yes	0.3	0.7	0.8	0.9

P(L|S)	S = no	S = yes
L = no	0.99	0.9
L = yes	0.01	0.1

P(B|S)	S = no	S = yes
B = no	0.7	0.4
B = yes	0.3	0.6

P(T|A)	A = no	A = yes
T = no	0.99	0.95
T = yes	0.01	0.05

P(X|E)	E = no	E = yes
X = no	0.95	0.02
X = yes	0.05	0.98

- Probability of having travelled to Asia, given that the patient suffers from dyspnoea?

$$P(A=1 \,|\, D=1) = \frac{1}{P(D=1)} \sum_{B,S,L,X,E,T} P(B,S,L,D=1,E,A=1,T,X)$$

$$= 0.0104$$

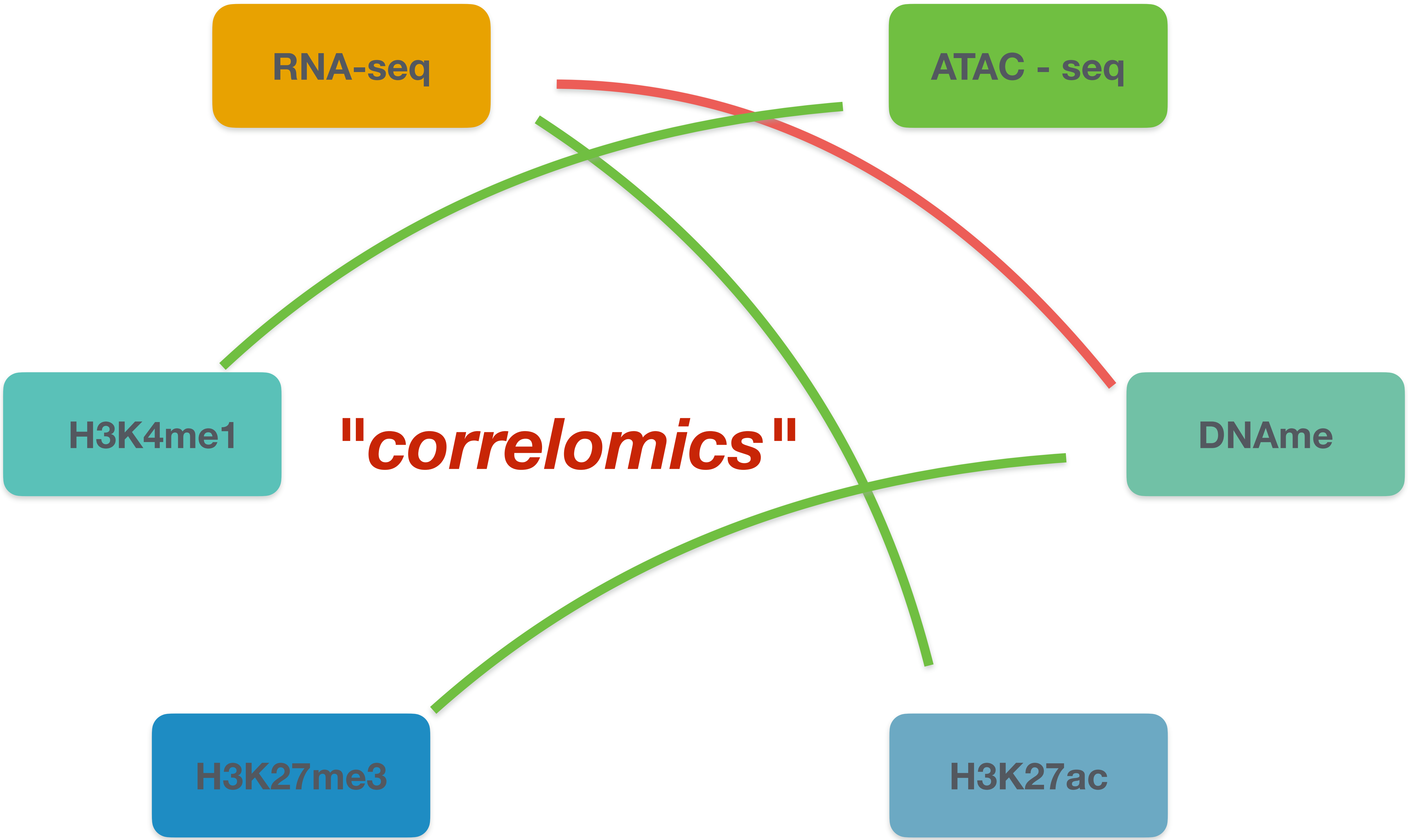
slightly higher than the overall probability of having travelled to Asia!

Medizinische Fakultät Heidelberg

Questions ?

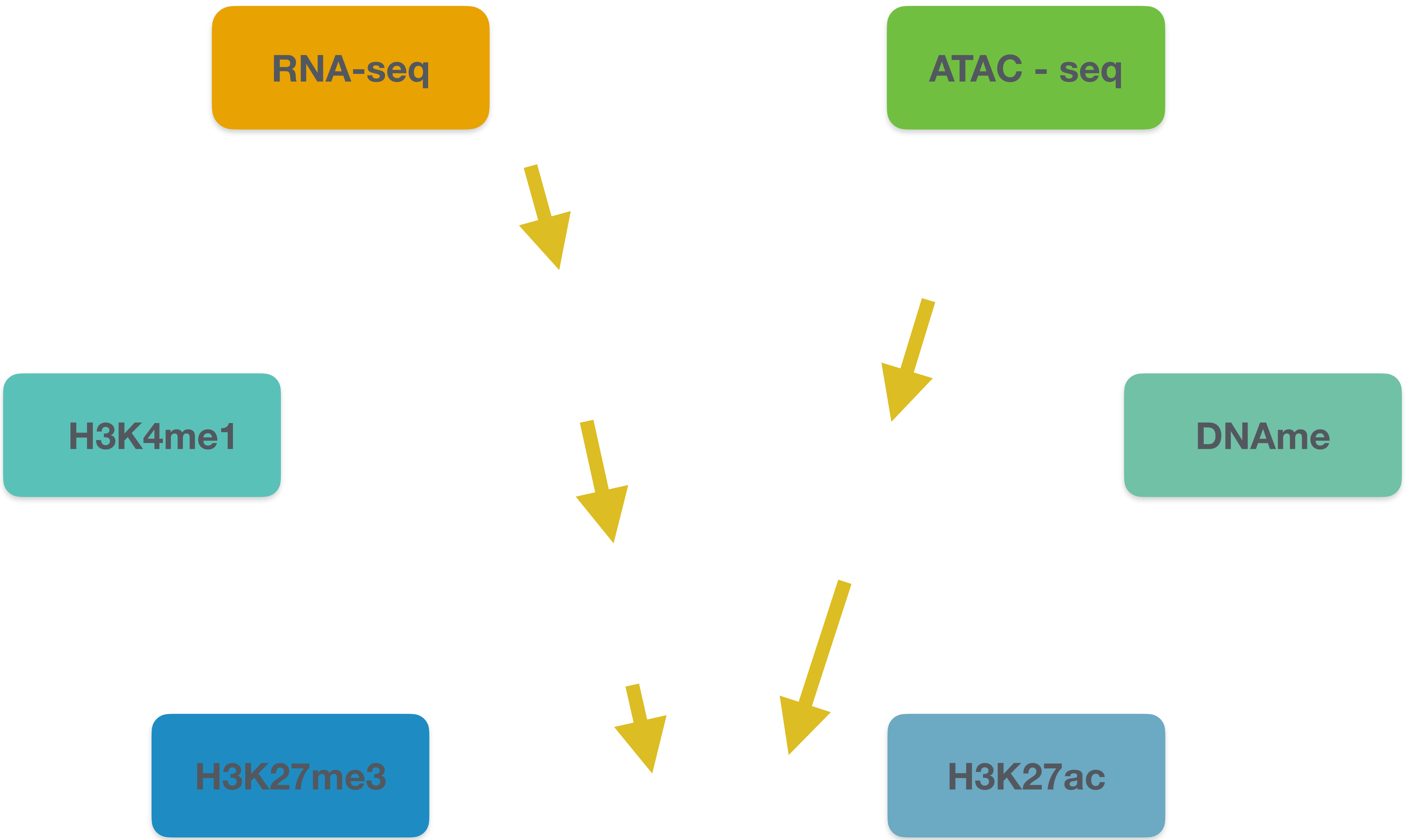
Application 2: Neuroblastoma epigenomics

Current challenges



Current challenges

Medizinische Fakultät Heidelberg



Genomics application

Medizinische Fakultät Heidelberg

DNA methylation

Gene expression

- Various neuroblastoma cell lines
- normal conditions / treated (inhibition)
- state at gene promoters represent the observations of the random variables

H3K27ac | **H3K4me1**

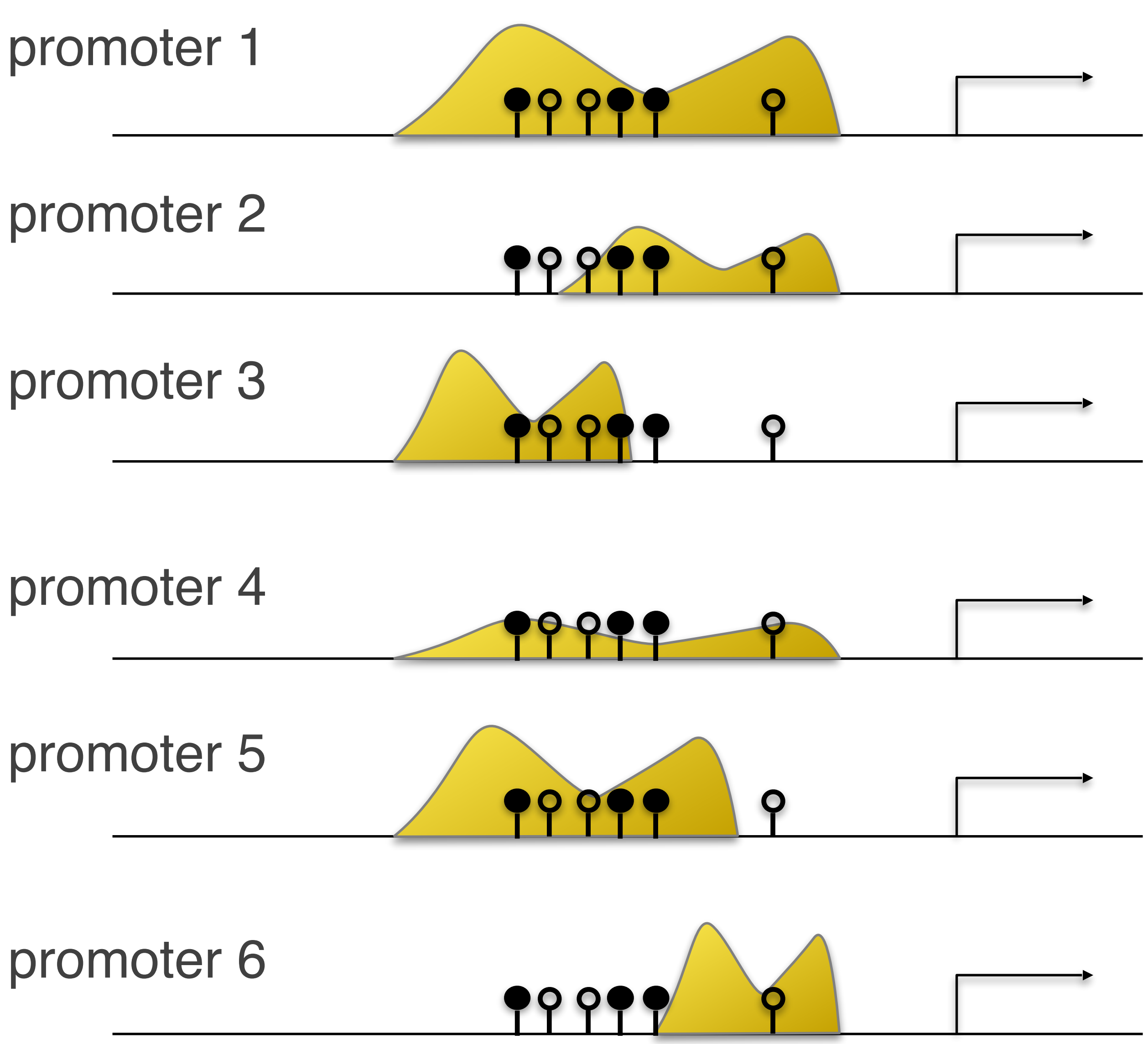
H3K4me3 | **H3K36me3**

H3K9me3 | **H3K27me3**

MYCN | **EZH2**

DNMT1 | **DNMT3**

Learning Network Structures



DNAme	K27ac
0.57	128.8
0.45	75.2
0.89	98.3
0.21	21.3
0.18	86.2
0.41	67.3

DNAme	K27ac
mid	5
mid	3
high	4
low	2
low	4
mid	3

3 states

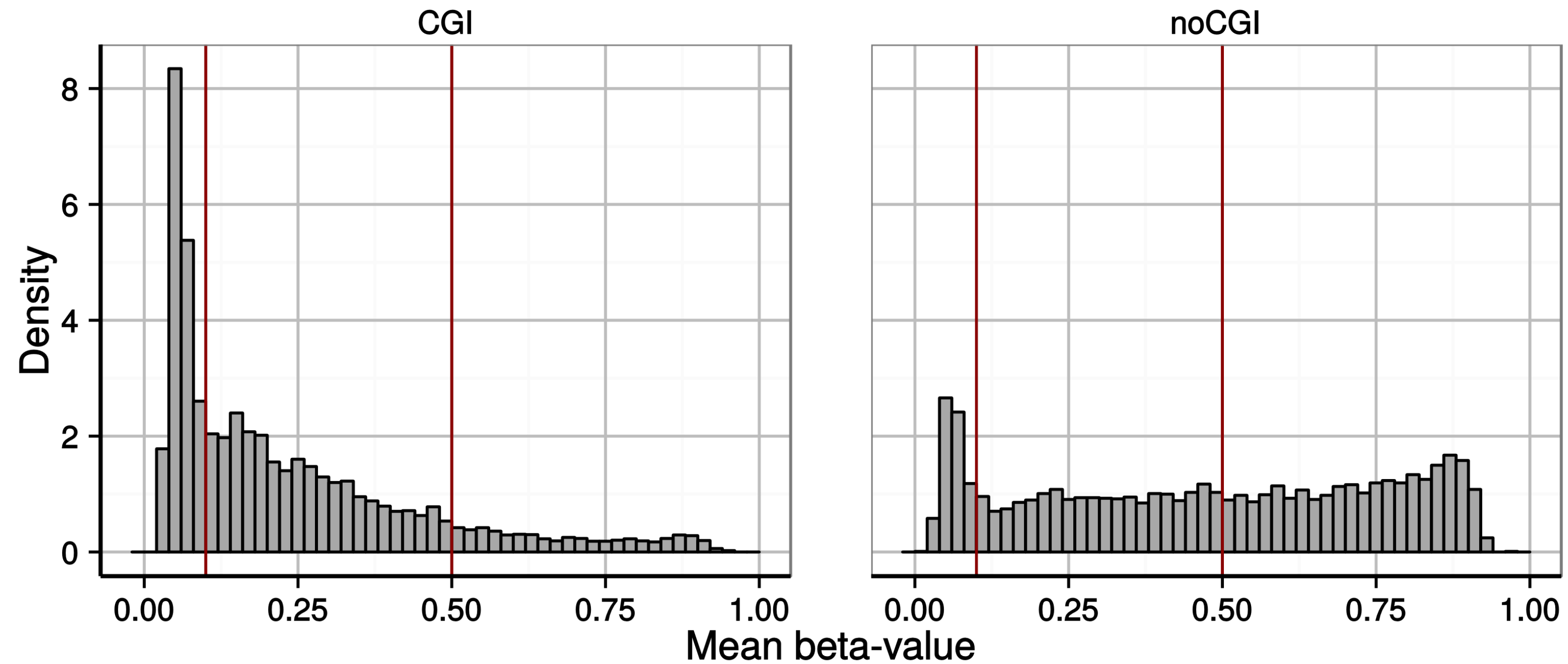
5 states

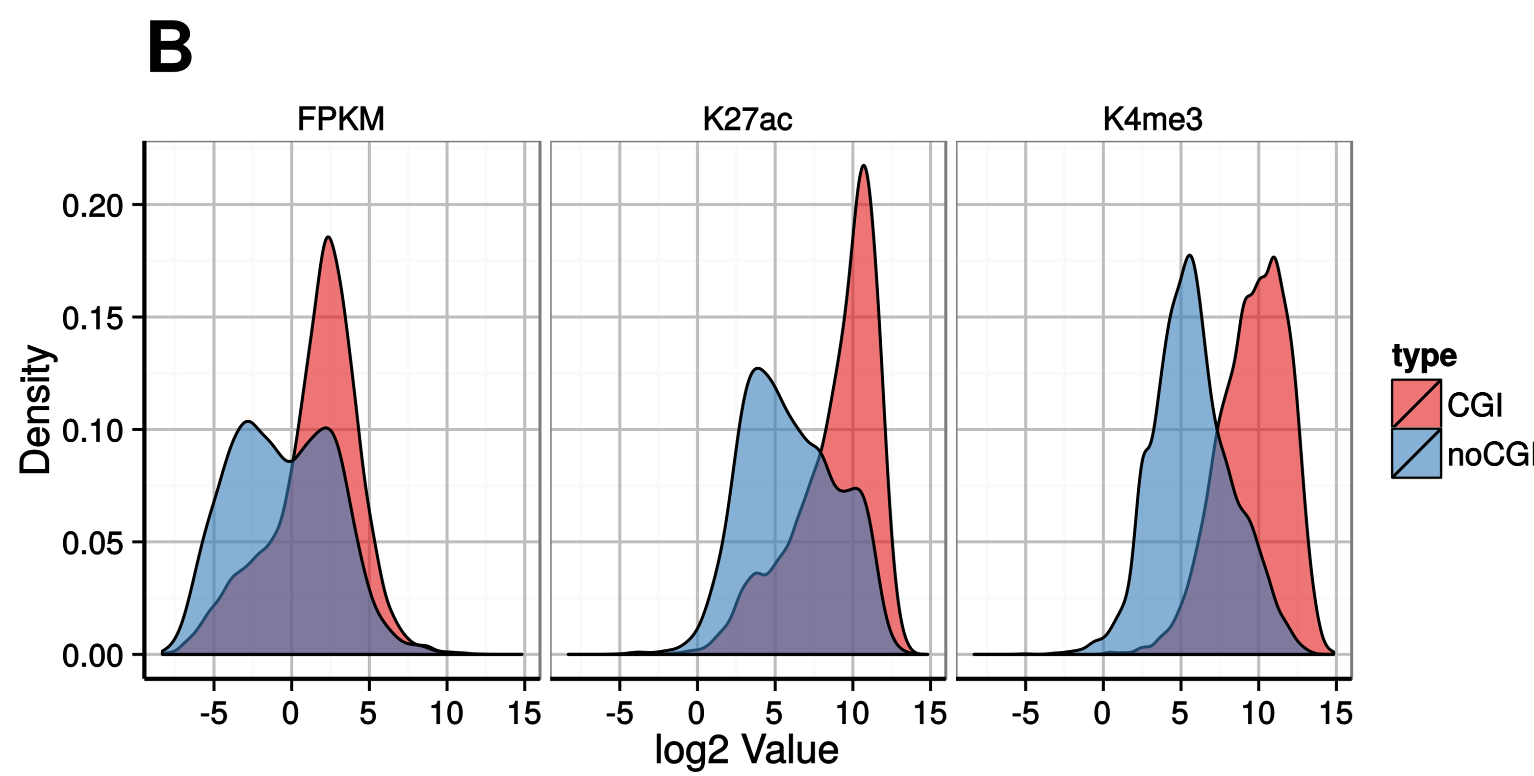
Discretisation strategies

- **Continuous data can be discretised** to circumvent the requirement for specific distributions (normal distribution)
- several discretisation strategies
  - ***naive discretisation***
    → define bins according to external evidence (low / mid / high)
  - ***quantile-based discretisation***
    → equally balanced levels
  - ***k-means based discretisation***
    → automatic definition of number of levels [Ckmeans.1d.dp, Wang et al. 2011]
  - ***mutual information preserving discretisation***
    → quantile-discretisation, then merging of levels such as to maintain the mutual information structure [Hartemink, 2005]

Different promoters have different distributions

- CpG-island overlapping
- non CpG-island overlapping





This might indicate that the observations correspond to different random variables

Bootstrapping

Medizinische Fakultät Heidelberg

- Network structures are learned over random subsets of observations (**Bootstrapping**)

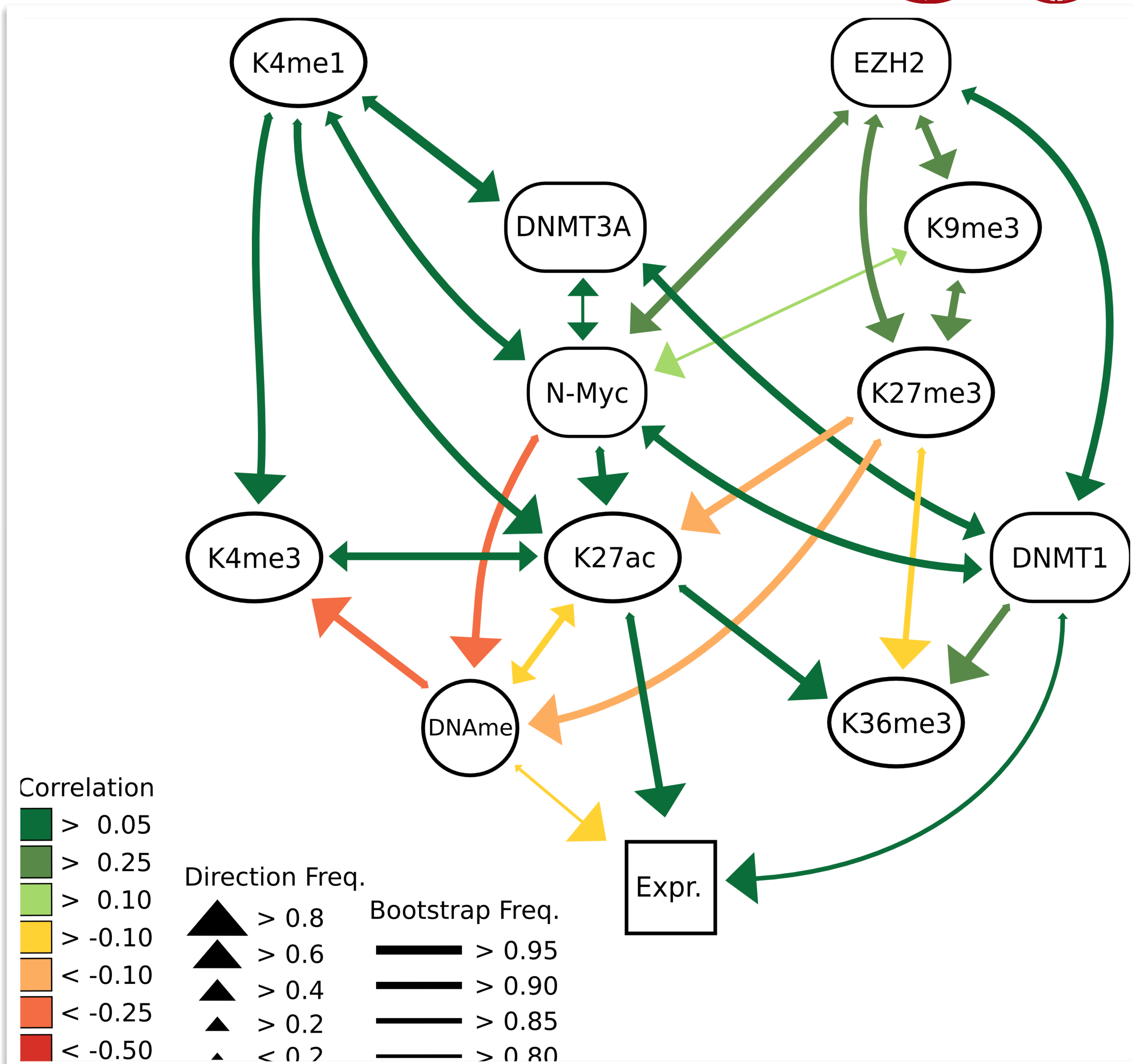
- Average the bootstrapped networks to determine
  - the **frequency** of an edge
  - the **direction** of the edge

- Apply thresholds to build a consensus networks
  - strength > 0.8
  - direction > 0.7

	from	to	strength	direction
1	Raf	Mek	1.000	0.5640000
2	Raf	Plcg	0.260	0.4730769
3	Raf	PIP2	0.036	0.5277778
4	Raf	PIP3	0.020	0.4500000
5	Raf	Erk	0.004	0.5000000
6	Raf	Akt	0.008	1.0000000
7	Raf	PKA	0.268	0.4440299
8	Raf	PKC	0.014	0.5000000
9	Raf	P38	0.022	0.5909091
10	Raf	Jnk	0.004	0.2500000
11	Mek	Raf	1.000	0.4360000
12	Mek	Plcg	0.008	0.2500000
13	Mek	PIP2	0.004	0.2500000
14	Mek	PIP3	0.002	0.0000000
15	Mek	Erk	0.030	0.2000000
16	Mek	Akt	0.012	0.8333333
17	Mek	PKA	0.022	0.1818182
18	Mek	PKC	0.218	0.3990826
19	Mek	P38	0.030	0.7000000
20	Mek	Jnk	0.020	0.4000000

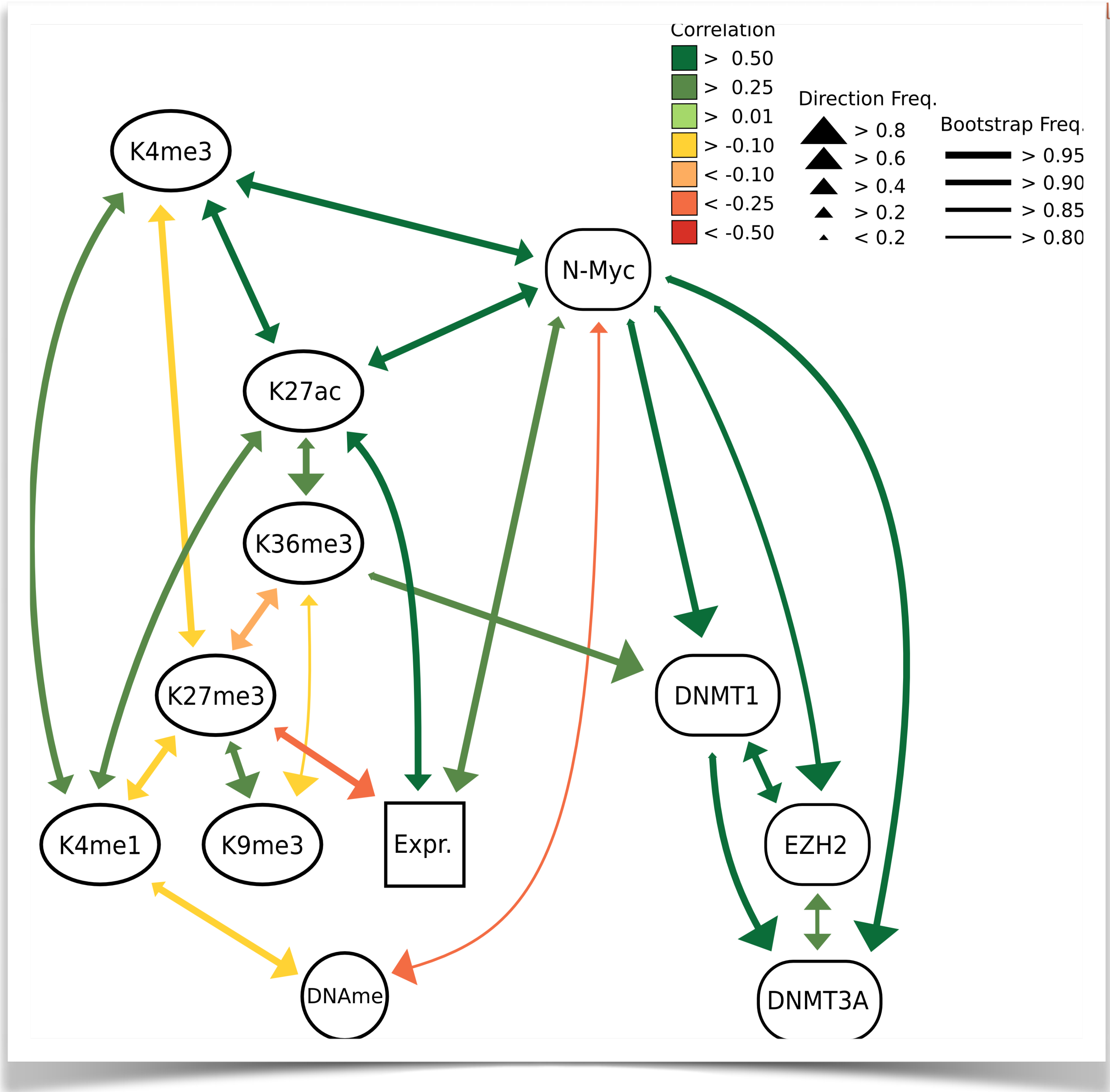
Promoter BN

- non-CGI Promoters
- (n = 5139)



Promoter BN

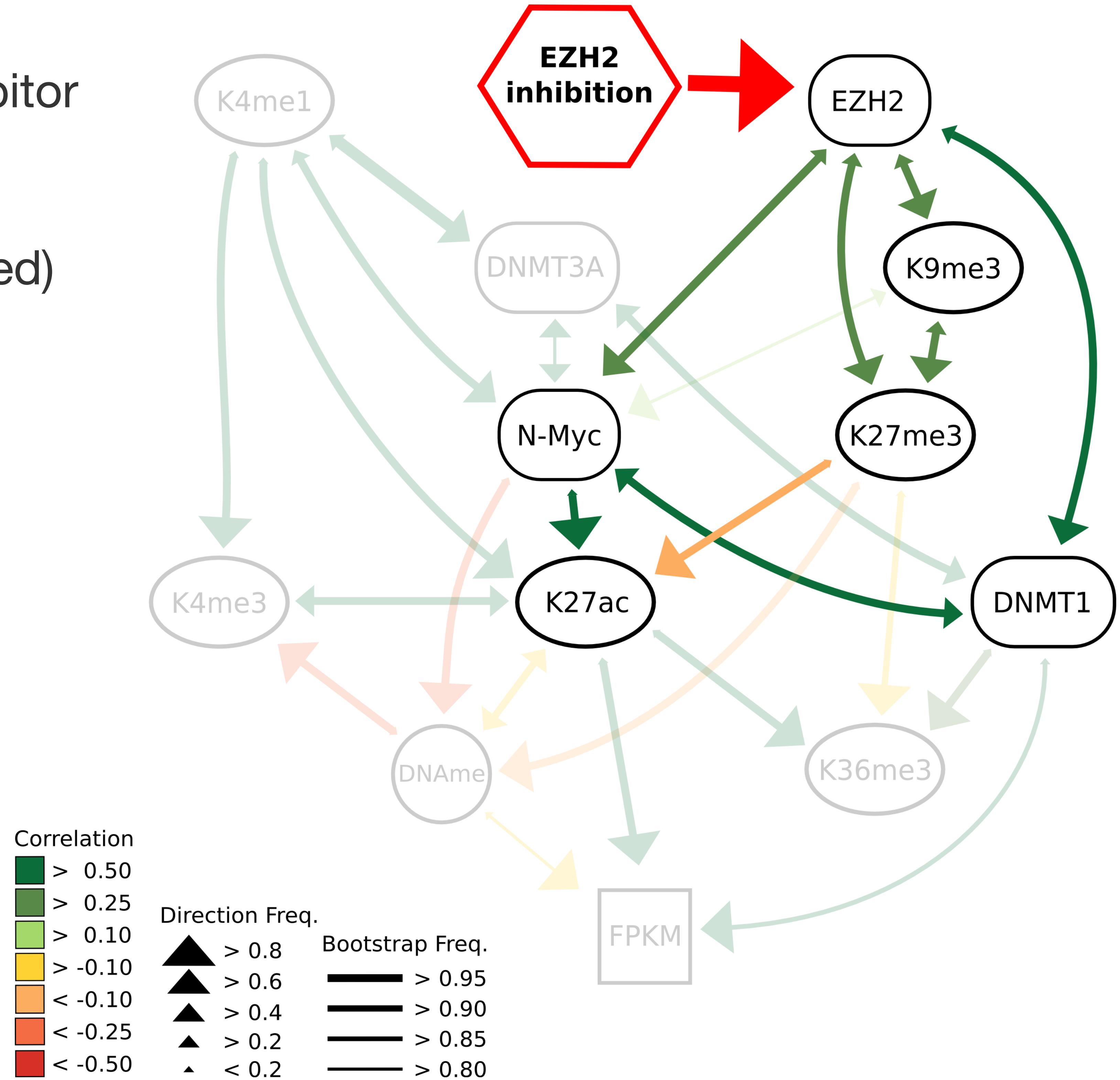
- CGI Promoters
- (n = 8906)



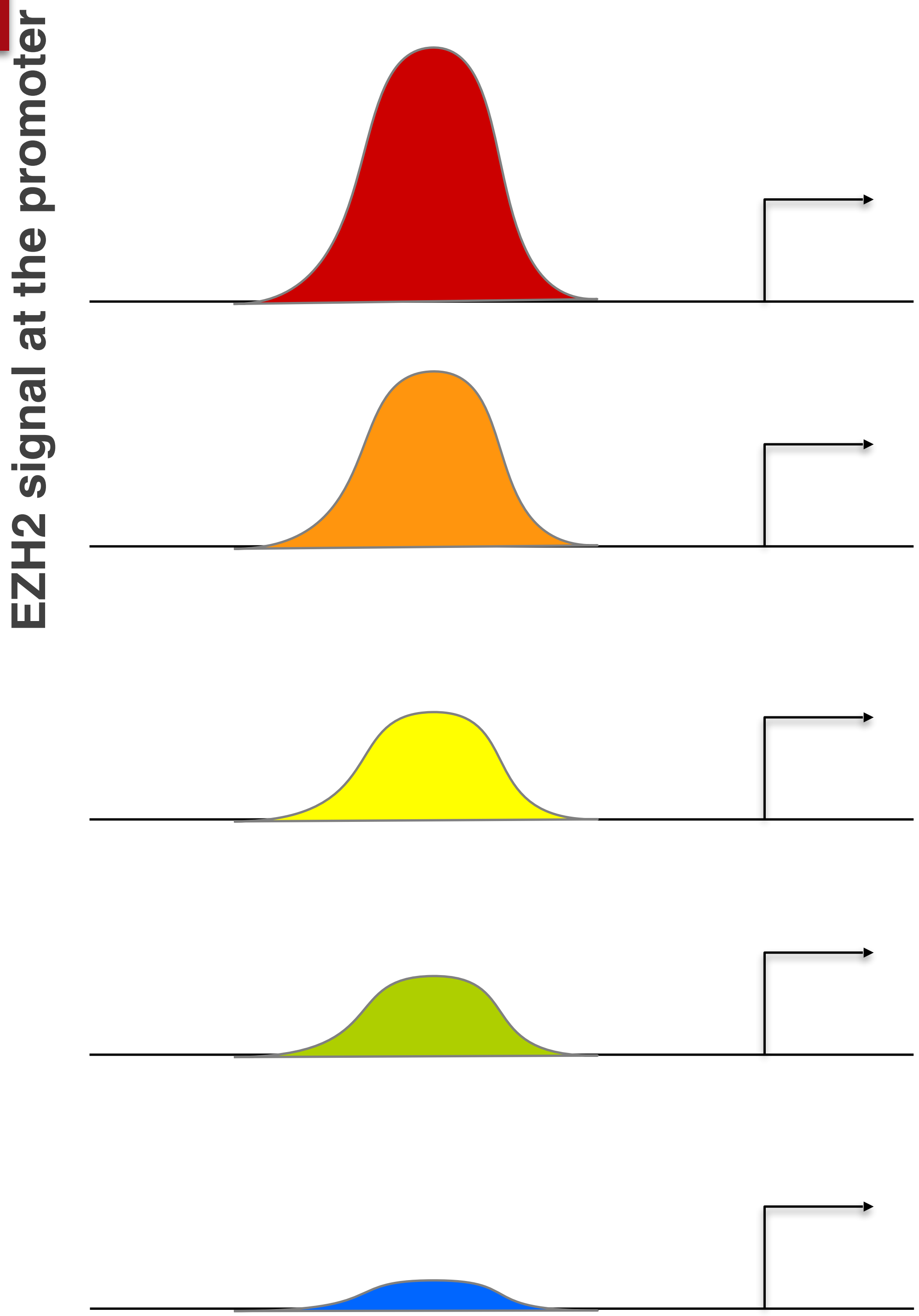
Predicting Interventions

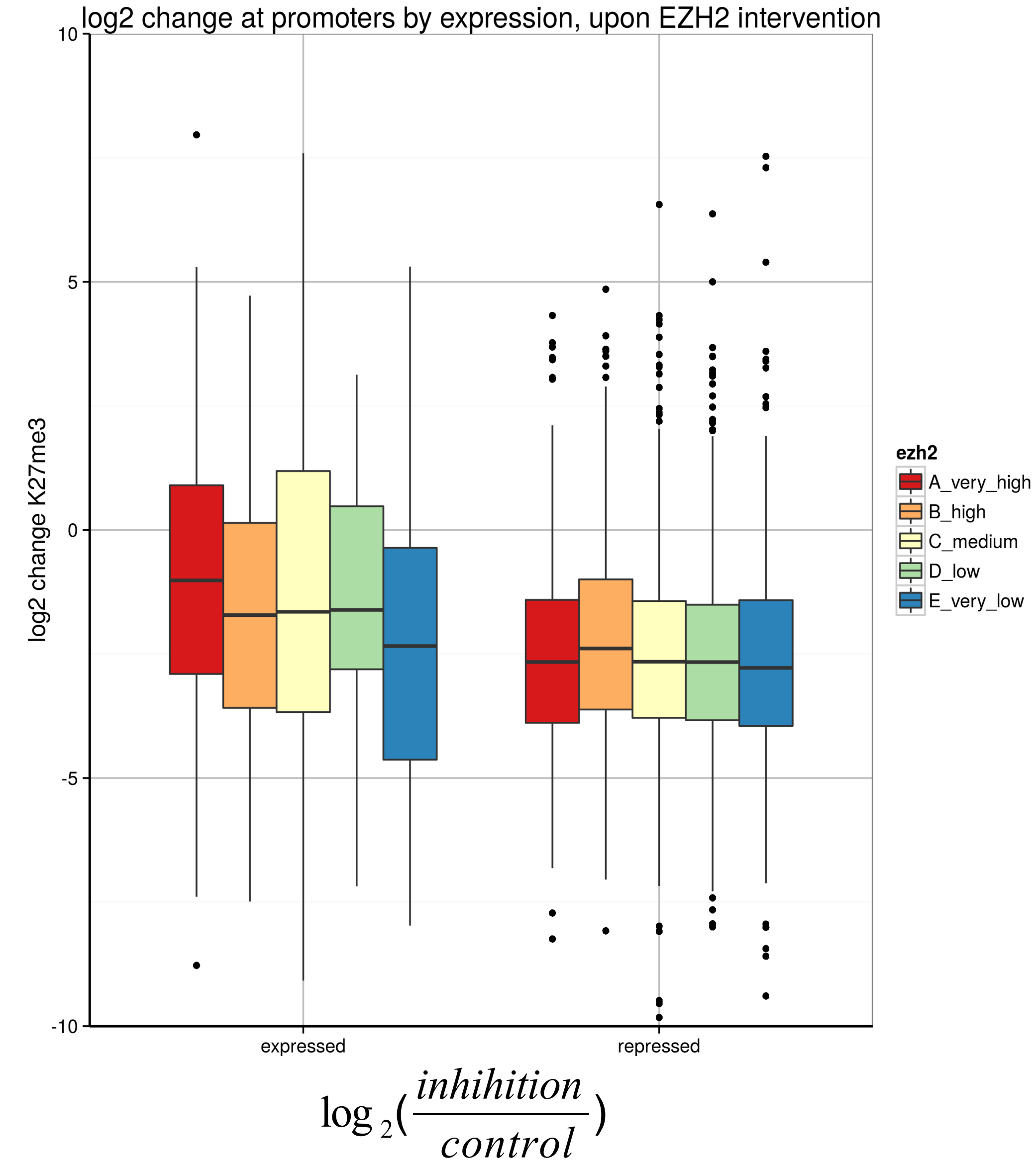
Medizinische Fakultät Heidelberg

- Small molecule EZH2 inhibitor
- Histone mark ChIP-seq,
- RNA-seq (control vs. treated)
- Same NB cell line Be(2)-C



Changes H3K27me3

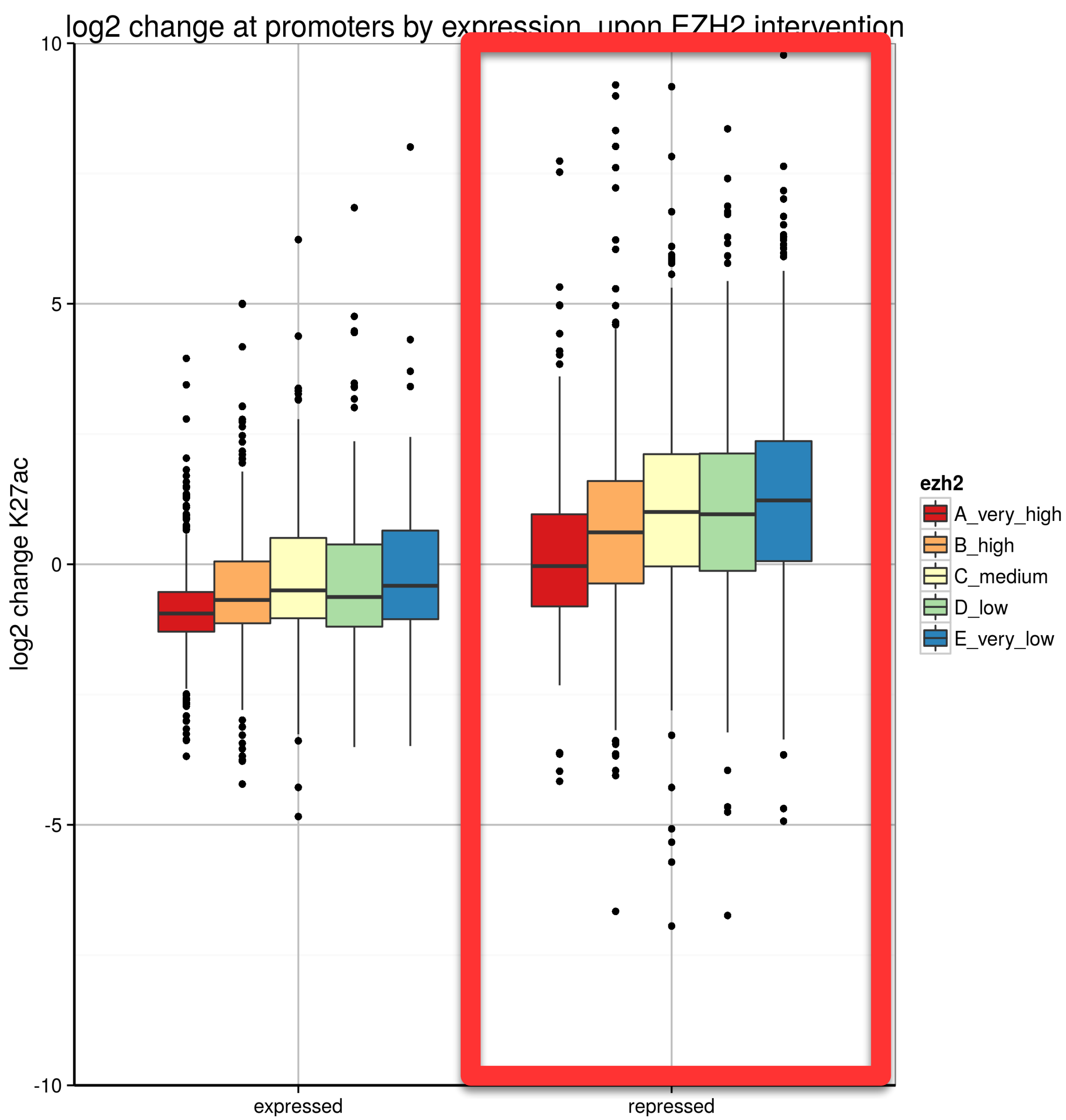




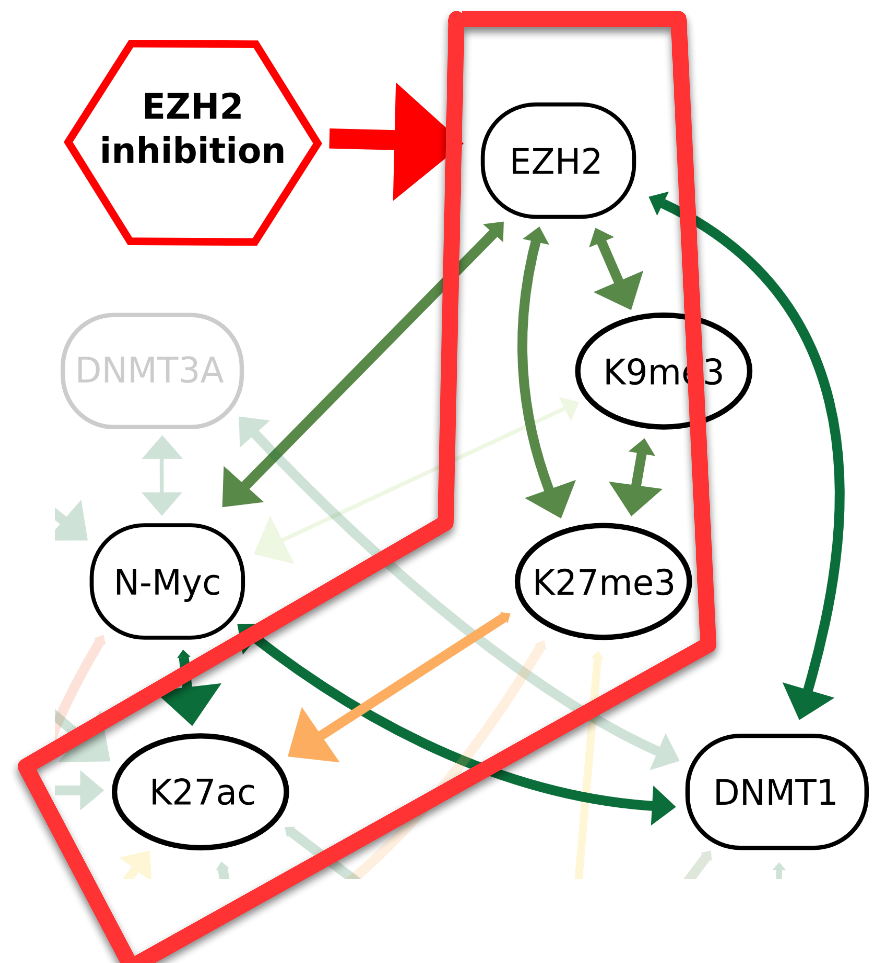
As expected, EZH2 inhibition reduces the overall H3K27me3 signal at the gene promoters

Changes H3K27ac

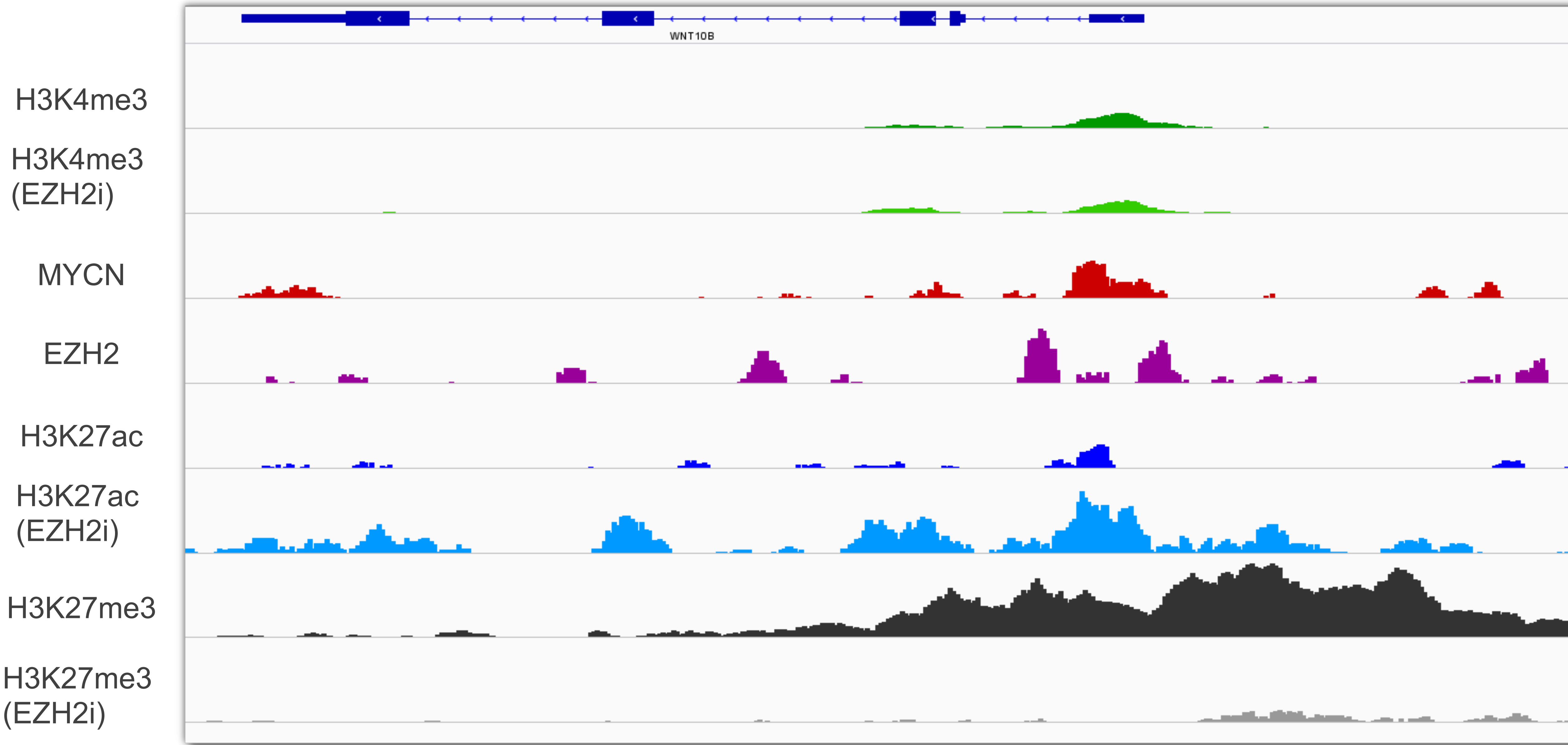
- **Increase of K27ac** at the promoter of repressed genes



$$\log_2\left(\frac{inhibition}{control}\right)$$

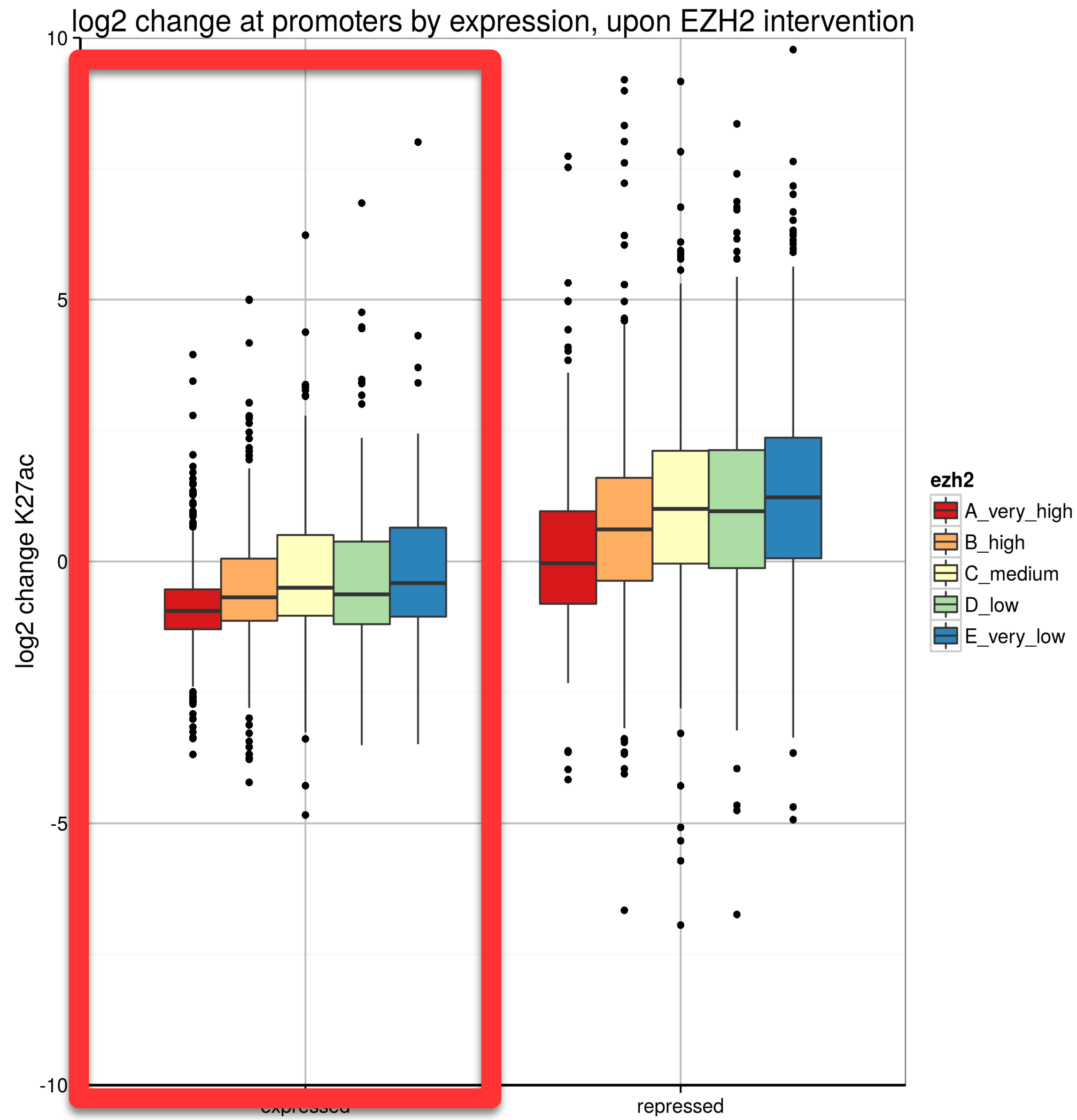


Changes in H3K27ac upon EZH2 inhibition

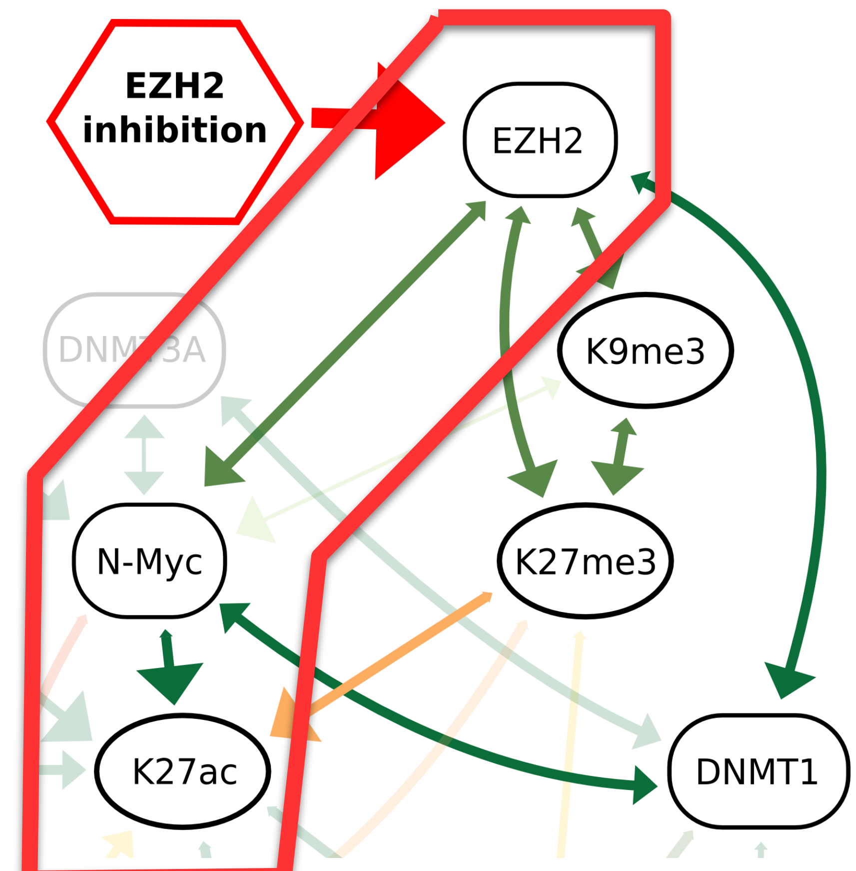


Changes H3K27ac

- EZH2 inhibition leads to a **reduction of K27ac** at the promoter of expressed genes

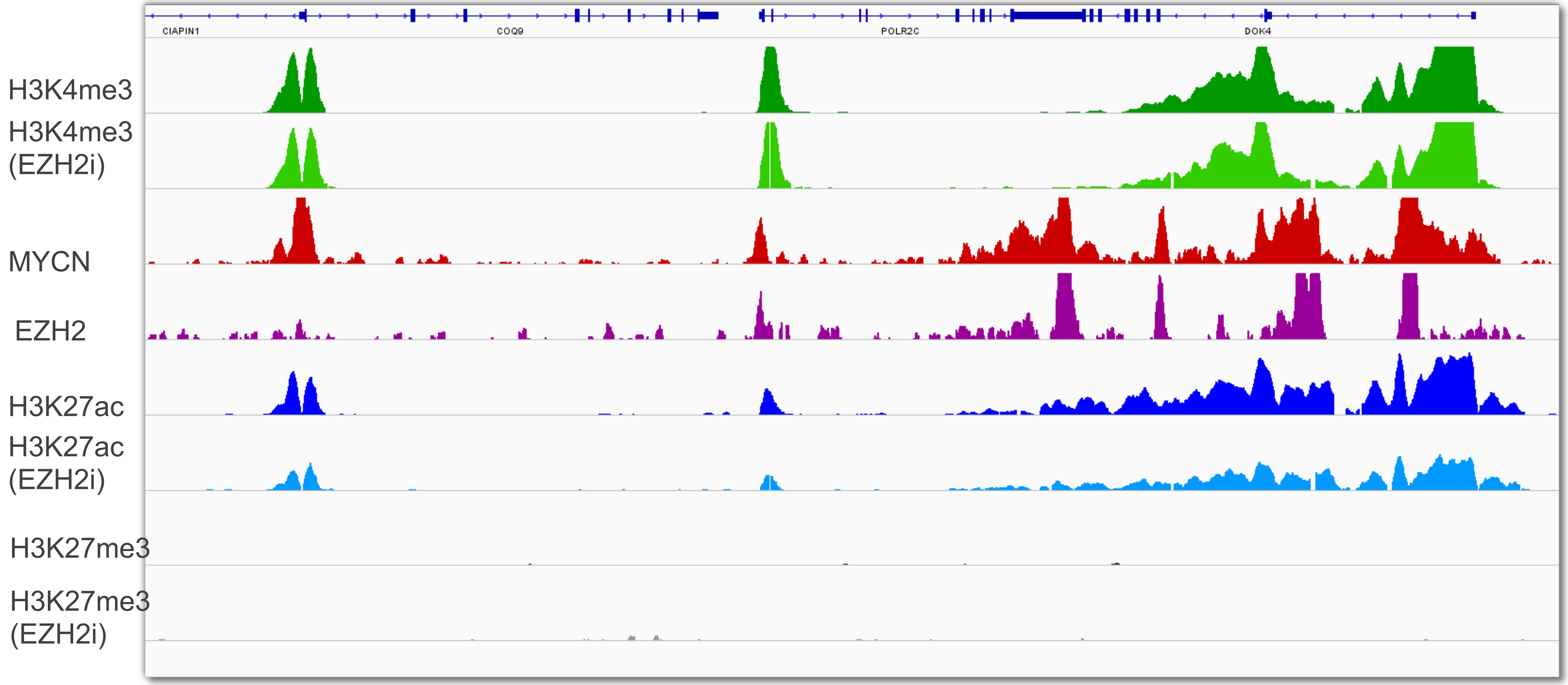


$$\log_2\left(\frac{inhihition}{control}\right)$$

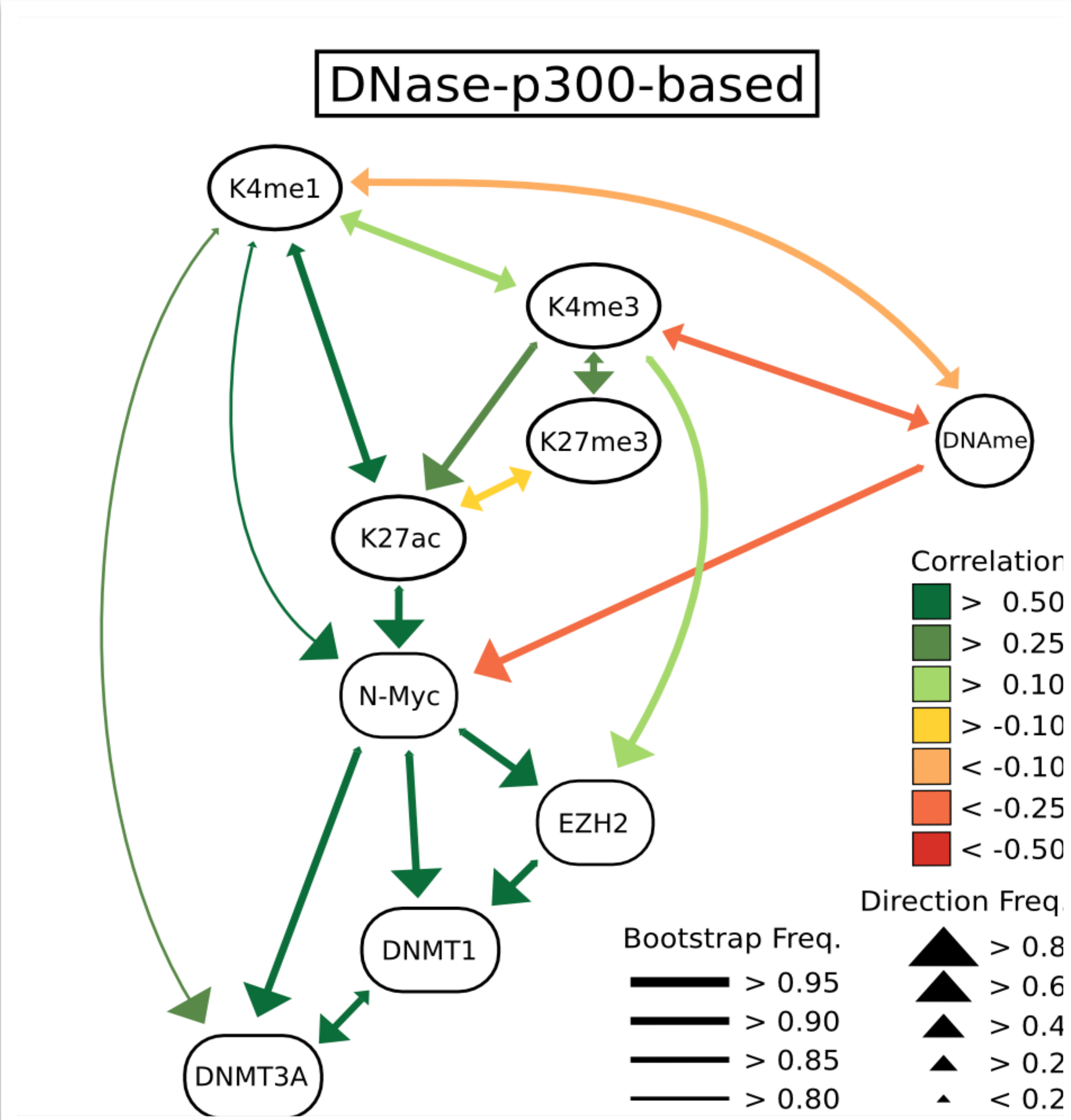


Changes in H3K27ac upon EZH2 inhibition

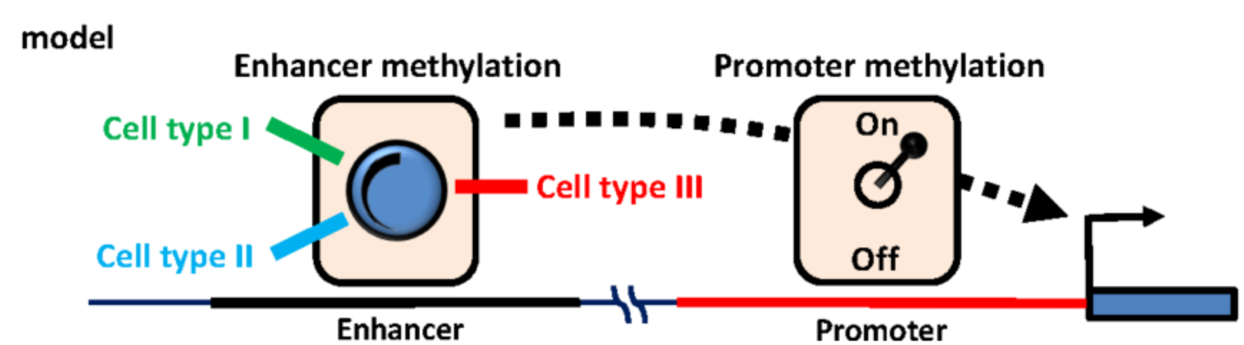
Medizinische Fakultät Heidelberg



Enhancer networks



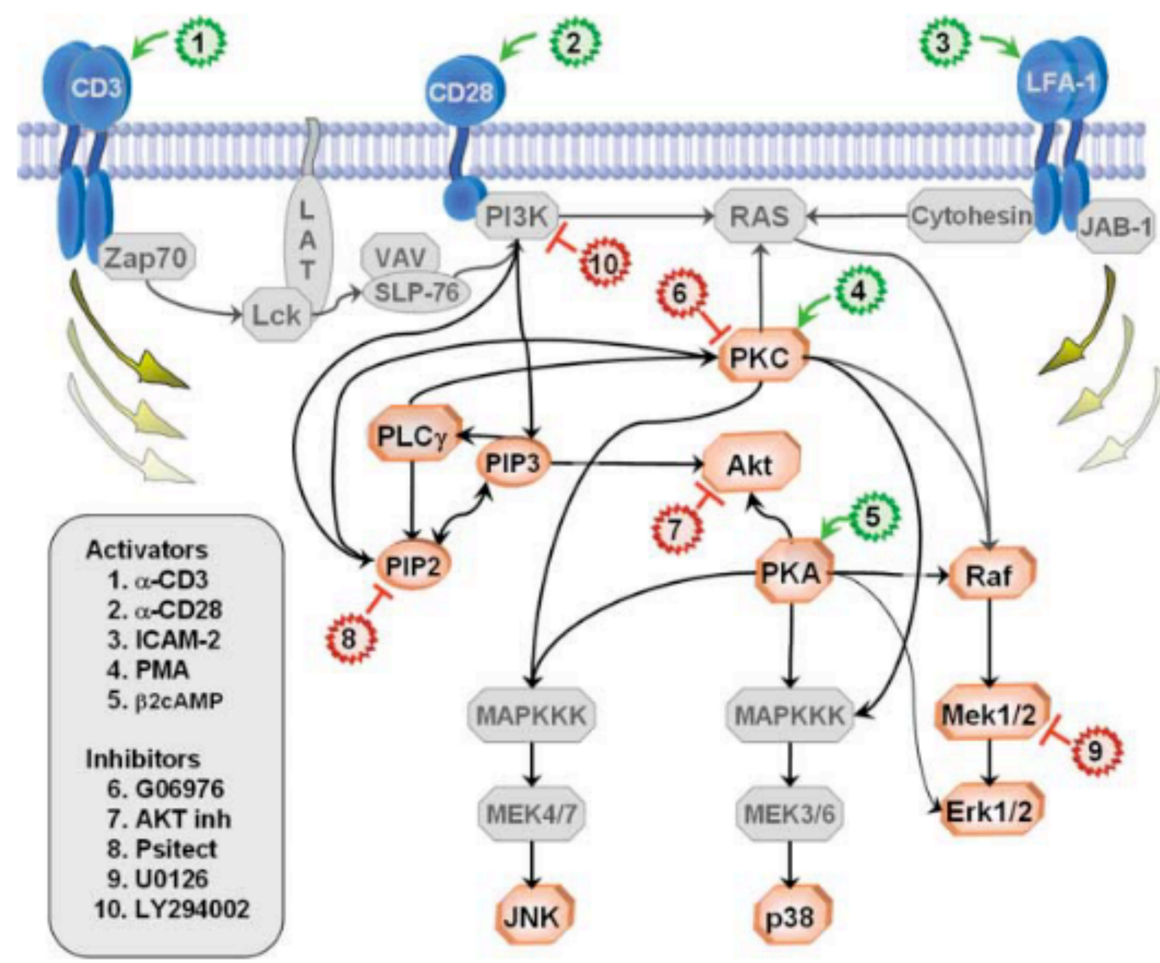
- Enhancers defined using DNAse/p300 in matching tissues / cell lines
- DNA methylation appears to play a more "active" role, compared to promoter networks



[Aran et al, Genome Biol. 2013]

About the tutorial

- Dataset of single-cell multicolor flow cytometry on 11 proteins of the MAPK pathway
- observational data + intervention data



[Sachs et al., 2005]

References

Books

- **Probabilistic Graphical Models** D. Koller & N. Friedman (MIT Press)
- **Causality** J. Pearl (Cambridge University Press)
- **Bayesian Networks in R** R. Nagarajan, M. Scutari, S. Lèbre (Springer)

Review papers

- **A primer on learning in Bayesian Networks for Computational Biology**, C. Needham et al. (PLOS Comp.Biol. 2007)
- **Inference in Bayesian Networks**, C. Needham et al. (Nature Biotech. 2006)
- **Learning Bayesian Networks in R**, S.Bottcher & C Dethlefsen

- **bnlearn tools** https://www.bnlearn.com/