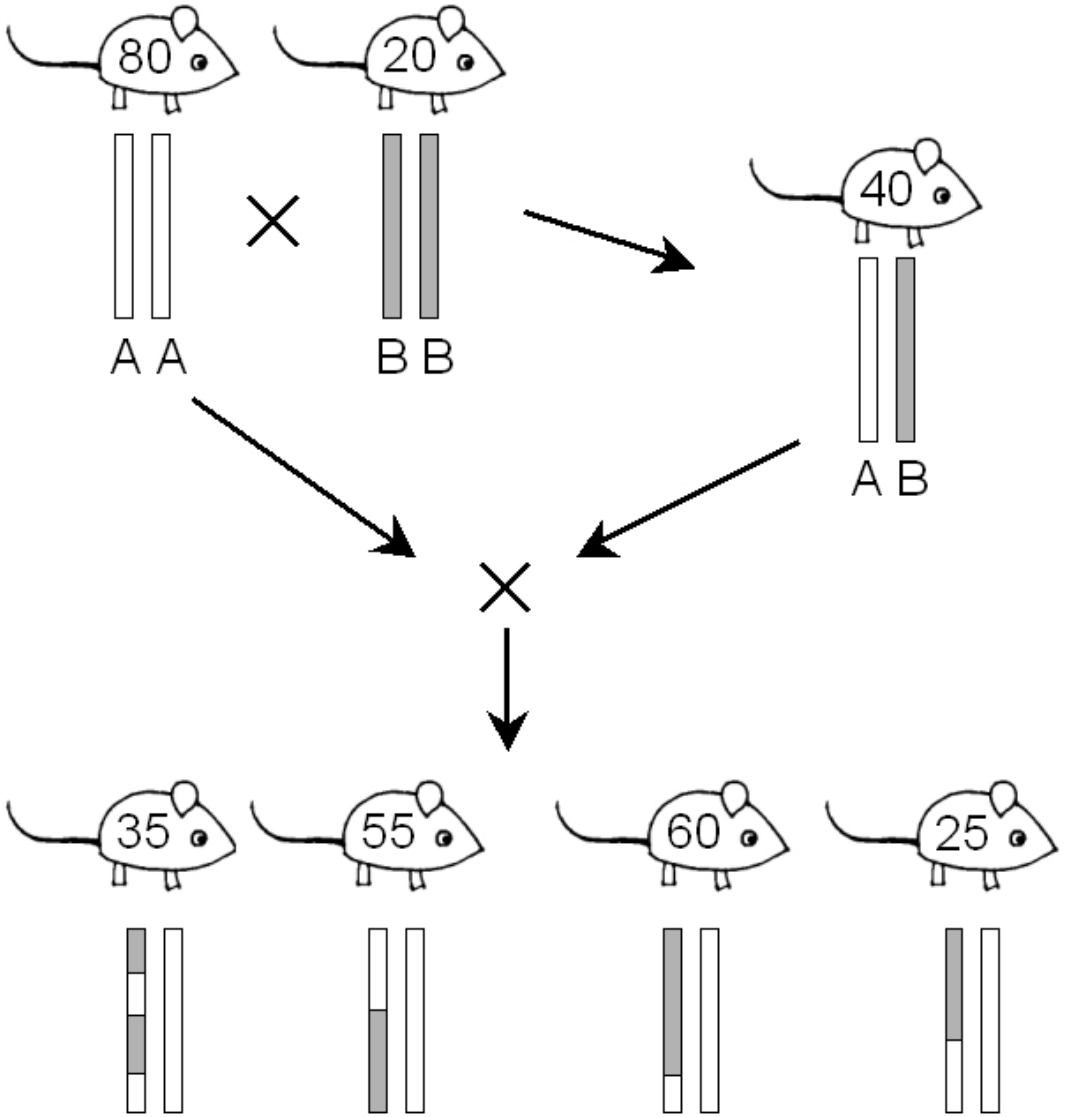
A causal gene network with genetic variations incorporating biological knowledge and latent variables

Jee Young Moon

Department of Statistics
University of Wisconsin-Madison

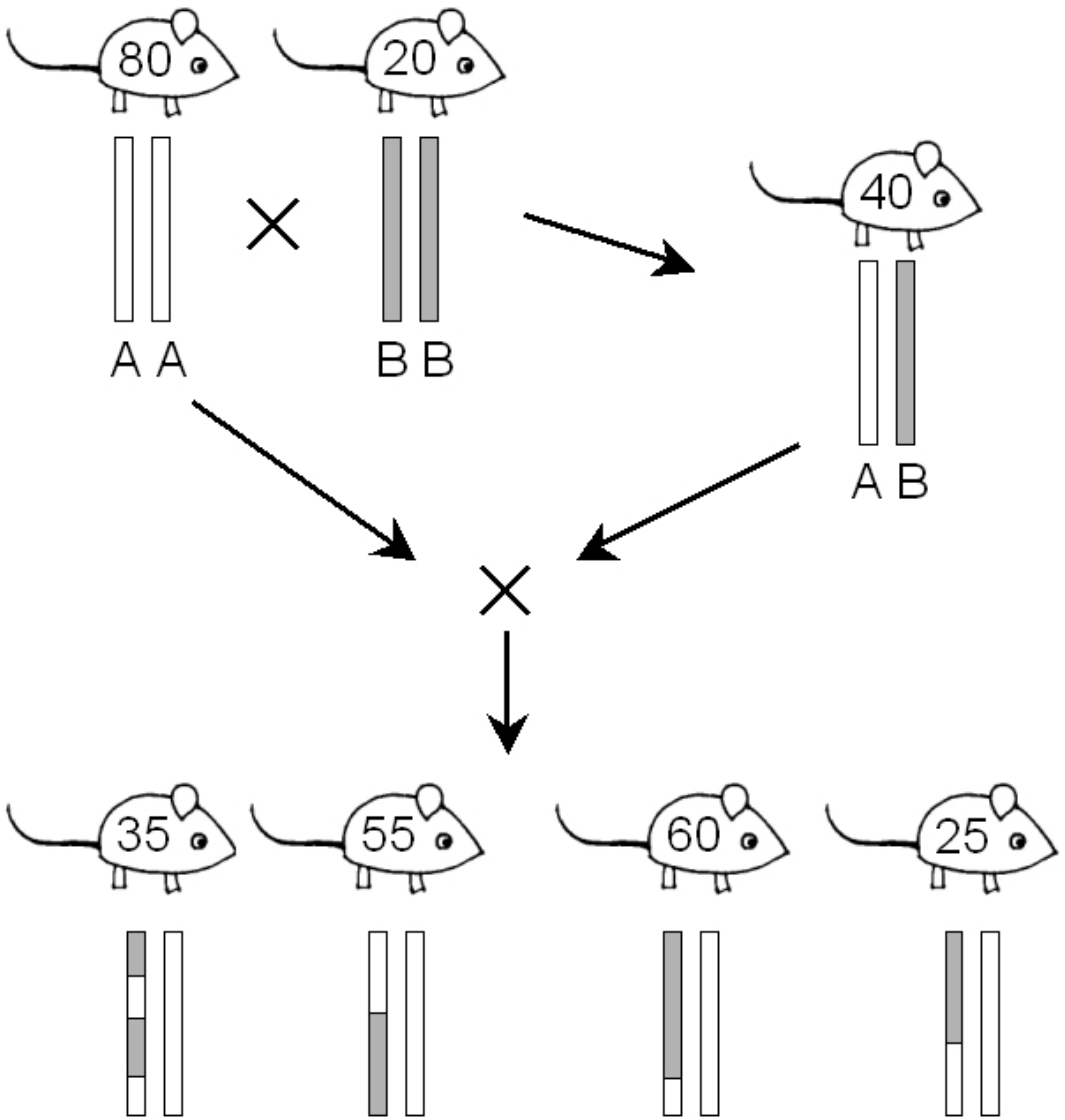
July 24, 2012

Experimental Cross Study



[Broman, K. W (2001) Lab Animal.]

Experimental Cross Study



[Broman, K. W (2001) Lab Animal.]

$\Rightarrow$ Genotypes
Gene expressions
Phenotypes of interest

Outline

1. Bayesian network with genetic variations and biological knowledge
   - Background
   - Encoding of biological knowledge
   - Model (QTLnet-prior)
   - Implementation
   - Simulations
   - Yeast cell cycle analysis
2. A genetic network with latent variables
   - Motivation for latent variables
   - Introduction of ancestral graph
   - Model
   - Property and Theorems
   - Algorithm - MCMC
   - Simulations
   - Conclusion
3. Future Research Plan

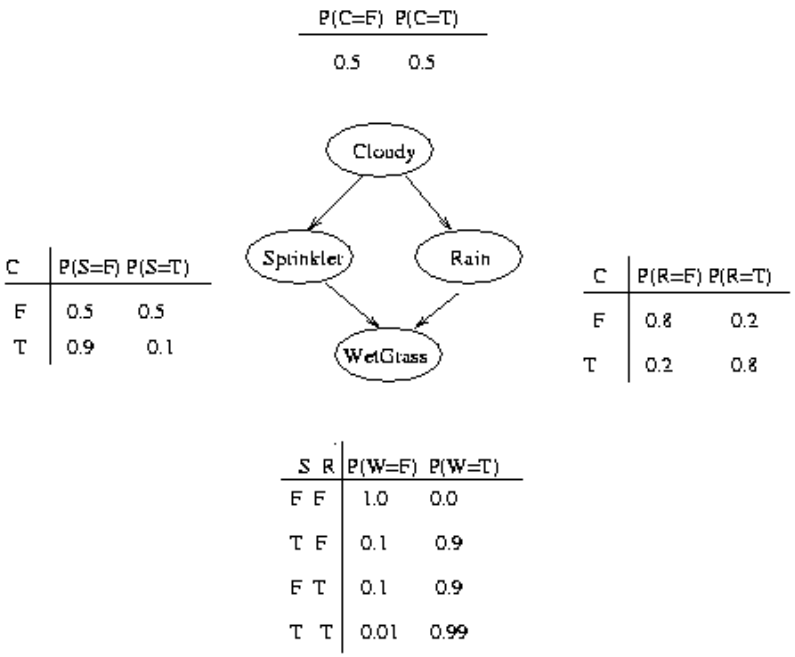
I. A Bayesian network with genetic variations and biological knowledge

Background

A Bayesian Network is a probabilistic graphical model whose conditional independence is represented by a directed acyclic graph (DAG), G.

Background

A Bayesian Network is a probabilistic graphical model whose conditional independence is represented by a directed acyclic graph (DAG), G.



picture from `http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html`

u $\longrightarrow$ **v**

technical definition: Y_v is conditionally dependent on Y_u

interpretation: Y_v is causally dependent on Y_u.

Properties of Bayesian network

Local directed Markov property Each variable is independent of its nondescendant variables conditional on its parent variables.

$$Y_t \perp Y_{V \setminus de(t)} | Y_{pa(t)} \quad \text{for all } t \in V$$

where $de(t)$ is the set of descendants of t, $pa(t)$ is the set of parents of t, V is the set of all nodes in a DAG G, and $Y_{pa(t)} = \{Y_i : i \in pa(t)\}$.

Properties of Bayesian network

Local directed Markov property Each variable is independent of its nondescendant variables conditional on its parent variables.

$$Y_t \perp Y_{V \setminus de(t)} | Y_{pa(t)} \quad \text{for all } t \in V$$

where $de(t)$ is the set of descendants of t, $pa(t)$ is the set of parents of t, V is the set of all nodes in a DAG G, and $Y_{pa(t)} = \{Y_i : i \in pa(t)\}$.

Factorization theorem.

$$\begin{aligned} P(Y_1, \ldots, Y_T) &= \prod_{t=1}^{T} P(Y_t | Y_{t-1}, \ldots, Y_1) \\ &= \prod_{t=1}^{T} P(Y_t | Y_{pa(t)}) \end{aligned}$$

Previous Work

- Friedman et al. (2000): a Bayesian network from microarray data with time-series measurements
- Chaibub Neto et al. (2010): a Bayesian network of phenotypes and causal QTLs
- Werhli and Husmeier (2007): a Bayesian network of phenotypes adjusted by prior Biological knowledge
- Zhu et al. (2008): Incorporate genetic variation and biological knowledge. But, network is constructed by piecewise merging.

Likelihood equivalence

Network Structure	Joint Likelihood
$G_Y^1 = Y1 \rightarrow Y_2 \rightarrow Y_3$	$P(Y_3\|Y_2)P(Y_2\|Y_1)P(Y_1) = P(Y_3\|Y_2)P(Y_2, Y_1)$
$G_Y^2 = Y1 \rightarrow Y_2 \leftarrow Y_3$	$P(Y_2\|Y_3, Y_1)P(Y_1)P(Y_3)$
$G_Y^3 = Y1 \leftarrow Y_2 \rightarrow Y_3$	$P(Y_2)P(Y_3\|Y_2)P(Y_1\|Y_2) = P(Y_3\|Y_2)P(Y_2, Y_1)$

G_Y^1 and G_Y^3 are likelihood equivalent.

Previous Work

- Friedman et al. (2000): a Bayesian network from microarray data with time-series measurements
- Chaibub Neto et al. (2010): a Bayesian network of phenotypes and causal QTLs
- Werhli and Husmeier (2007): a Bayesian network of phenotypes adjusted by prior Biological knowledge
- Zhu et al. (2008): Incorporate genetic variation and biological knowledge. But, network is constructed by piecewise merging.

causal QTL

Phenotypes are causally dependent on QTLs.

1. In Biology, genotypes influence phenotypes, not the other way. $Q \rightarrow Y$.
2. Alleles are randomized during meiosis.

Extended Network Structure	Joint Likelihood
$G^1 = Q \rightarrow Y1 \rightarrow Y_2 \rightarrow Y_3$	$P(Y_3\|Y_2)P(Y_2\|Y_1)P(Y_1\|Q)P(Q)$
$G^3 = Q \rightarrow Y1 \leftarrow Y_2 \rightarrow Y_3$	$P(Y_2)P(Y_3\|Y_2)P(Y_1\|Y_2,Q)P(Q)$

Adding QTL can distinguish G^1 and G^3 by likelihoods.

Previous Work

- Friedman et al. (2000): a Bayesian network from microarray data with time-series measurements
- Chaibub Neto et al. (2010): a Bayesian network of phenotypes and causal QTLs
- Werhli and Husmeier (2007): a Bayesian network of phenotypes adjusted by prior Biological knowledge
- Zhu et al. (2008): Incorporate genetic variation and biological knowledge. But, network is constructed by piecewise merging.

Biological knowledge

If $P(u \rightarrow v) > P(u \leftarrow v)$ by prior biological knowledge,

Biological knowledge

If $P(u \rightarrow v) > P(u \leftarrow v)$ by prior biological knowledge,
and $P(Y|u \rightarrow v) = P(Y|u \leftarrow v)$,

Biological knowledge

If $P(u \rightarrow v) > P(u \leftarrow v)$ by prior biological knowledge,
and $P(Y|u \rightarrow v) = P(Y|u \leftarrow v)$,
then posterior $P(u \rightarrow v|Y) > P(u \leftarrow v|Y)$.

- Transcription factor binding
- Protein-protein interaction

Previous Work

- Friedman et al. (2000): a Bayesian network from microarray data with time-series measurements
- Chaibub Neto et al. (2010): a Bayesian network of phenotypes and causal QTLs
- Werhli and Husmeier (2007): a Bayesian network of phenotypes adjusted by prior Biological knowledge
- Zhu et al. (2008): Incorporate genetic variation and biological knowledge. But, network is constructed by piecewise merging.

Encoding of Biological Knowledge, B

B is a matrix of number of phenotypes $\times$ number of phenotypes.

Encoding of Biological Knowledge, B

B is a matrix of number of phenotypes $\times$ number of phenotypes.
Transcription factor and DNA binding Suppose we have a p-value about whether a transcription factor binds to a certain DNA location. As in Bernard and Hartemink (2005), we assume

$$P_\lambda(P_{ij} = p | G(i,j) = 1) = \frac{\lambda e^{-\lambda p}}{1 - e^{-\lambda}}$$
$$P_\lambda(P_{ij} = p | G(i,j) = 0) = 1.$$

We assume $P(G_{i,j} = 1) = P(G_{i,j} = 0) = 1/2$. Then, the presence of an edge after observing p-value is

$$P(G(i,j) = 1 | P_{ij} = p) = \frac{1}{\lambda_H - \lambda_L} \int_{\lambda_L}^{\lambda_H} \frac{\lambda e^{-\lambda p}}{\lambda e^{-\lambda p} + (1 - e^{-\lambda})} d\lambda$$

$B(i,j) := P(G(i,j) = 1 | P_{ij} = p)$.

Encoding protein-protein interaction A Bayes classifier by Jansen et al. (2003) to combine heterogeneous interaction data.

$$O_{posterior} = \frac{P(pos|f_1, \ldots, f_L)}{P(neg|f_1, \ldots, f_L)} = O_{prior} \times LR$$
$$= \frac{P(pos)}{P(neg)} \times \frac{P(f_1, \ldots, f_L|pos)}{P(f_1, \ldots, f_L|neg)}.$$

$P(f_1, \ldots, f_L|pos)$ is obtained in the positive gold standard.

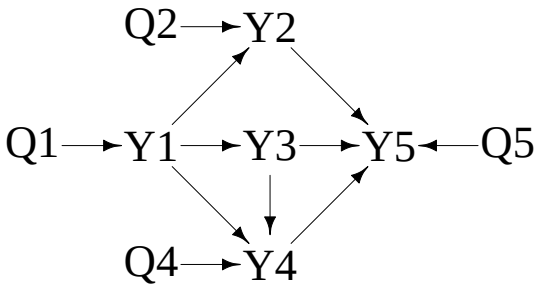
$$B(i,j) = B(j,i) := \frac{O_{posterior}}{1 + O_{posterior}} = P(pos|f_1, \ldots, f_L).$$

Our Model - QTLnet-prior

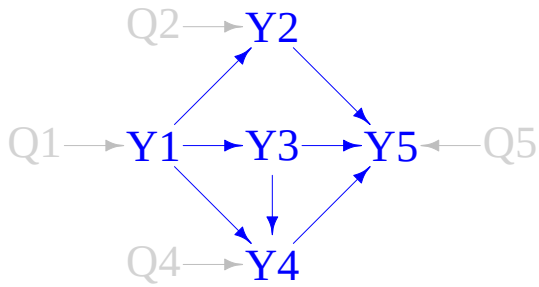
We incorporate both causal QTLs and biological knowledge to infer a Bayesian network of phenotypes.

$$
\begin{aligned}
P(G, W|Y, X, B) &\propto P(Y|G, W, X, B)P(G, W|X, B) \\
&= P(Y|G, X)P(G_Y, W|X, B)P(G_{Q\to Y}|X, B) \\
&= P(Y|G, X)P(G_Y, W|B)P(G_{Q\to Y}|X) \\
&= P(Y|G, X)P(G_Y|W, B)P(W|B)P(G_{Q\to Y}|X)
\end{aligned}
$$

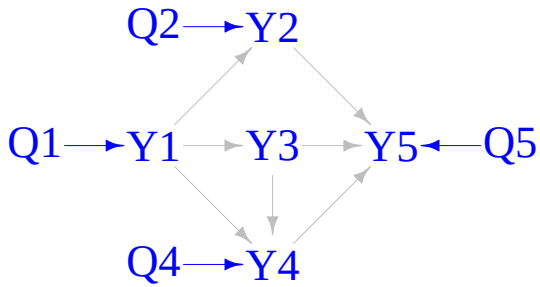
G	a Bayesian network of phenotypes and causal QTLs
G_Y	a subgraph of G composed of phenotype nodes and edges between phenotypes
$G_{Q\to Y}$	a subgraph of G composed of phenotypes and causal QTL nodes and edges from QTL to phenotypes
B	a matrix of biological knowledge
W	weight of biological knowledge
Y	expression data
X	genetic variation information


Q2
Y2
Q1
Y1
Y3
Y5
Q5
Q4
Y4

G


Q2
Y2
Q1
Y1
Y3
Y5
Q5
Q4
Y4

G_Y


Q2
Y2
Q1
Y1
Y3
Y5
Q5
Q4
Y4

$G_{Q \to Y}$

A Bayesian network of phenotypes with causal QTLS, $P(Y|G, X)$

We assume the following family of distribution for phenotypes by Chibub Neto et. al (2010)

$$y_{ti} = \mu^*_{ti} + \sum_{v \in pa(t)} \beta_{tv} y_{vi} + \epsilon_{ti}, \quad \epsilon_{ti} \sim N(0, \sigma^2_t)$$

where $\mu^*_{ti} = \mu_t + X_i \; diag(\gamma_t) \; \theta_t$, μ_t is the overall mean for a trait t,
θ_t is a column vector of all genetic effects,
X_i is a row vector for individual i from X,
β_{tv} is the partial regression coefficients relating phenotype t with phenotype v,
ϵ_{ti} is the associated independent normal error term.

Joint likelihood is obtained by multiplying all the likelihoods for all traits by factorization theorem.
Marginal likelihood is

$$P(Y|G, X) = \int P(Y|G, X, \theta_G) P(\theta_G|G) d\theta_G.$$

Prior on phenotype network structures, $P(G_Y|B, W)$

Assume a Gibbs distribution for the network structure to integrate biological knowledge from Werhli and Husmeier (2007).

$$P(G_Y|B, W) = \frac{\exp(-W\mathcal{E}(G_Y))}{Z(W)}, \quad G_Y \in \mathcal{DAG}$$

$$\text{where } \mathcal{E}(G_Y) = \sum_{i,j=1}^{T} |B(i,j) - G_Y(i,j)|.$$

where B is an encoding of biological knowledge ranging from 0 to 1 and G_Y is an adjacency matrix. $G_Y(i,j) = 1$ means the presence of the directed edge from node i to j.
W controls the contribution of biological knowledge.

- $W \to \infty$: prior on network structures peaks at the biological knowledge
- $W \to 0$: influence of knowledge gets negligible. Uniform distribution

Prior on biological knowledge weights, $P(W|B)$ and Prior on genetic architectures, $P(G_{Q \rightarrow Y})$

$$P(W|B) \sim \exp(-W)$$
$$P(G_{Q \rightarrow Y}) \sim \textit{Uniform}$$

Markov Chain Monte Carlo Sampling

1. Sample a new network structure of phenotypes G_Y^{new} from a network structure proposal distribution $R(G_Y^{new} | G_Y^{old})$.

Markov Chain Monte Carlo Sampling

1. Sample a new network structure of phenotypes G_Y^{new} from a network structure proposal distribution $R(G_Y^{new}|G_Y^{old})$.
2. Given a network structure of phenotypes G_Y^{new}, sample a new genetic architecture $G_{Q\rightarrow Y}$ from an architecture proposal distribution $R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})$.

Markov Chain Monte Carlo Sampling

1. Sample a new network structure of phenotypes G_Y^{new} from a network structure proposal distribution $R(G_Y^{new}|G_Y^{old})$.
2. Given a network structure of phenotypes G_Y^{new}, sample a new genetic architecture $G_{Q\rightarrow Y}$ from an architecture proposal distribution $R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})$.
3. Accept the new extended network structure G^{new} composed of G_Y^{new} and $G_{Q\rightarrow Y}^{new}$ given the biological knowledge weights W with a probability $A_G = \min\{1, \frac{P(Y|G^{new},X)P(G_Y^{new}|B,W)P(G_{Q\rightarrow Y}^{new})}{P(Y|G^{old},X)P(G_Y^{old}|B,W)P(G_{Q\rightarrow Y}^{old})} \frac{R(G_Y^{old}|G_Y^{new})R(G_{Q\rightarrow Y}^{old}|G_{Q\rightarrow Y}^{new})}{R(G_Y^{new}|G_Y^{old})R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})}\}$.

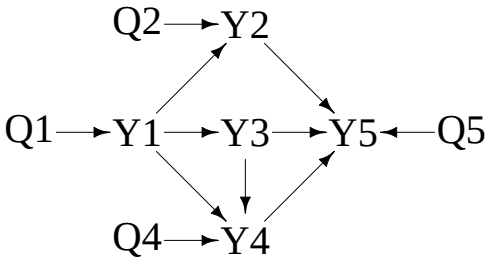
Markov Chain Monte Carlo Sampling

1. Sample a new network structure of phenotypes G_Y^{new} from a network structure proposal distribution $R(G_Y^{new}|G_Y^{old})$.
2. Given a network structure of phenotypes G_Y^{new}, sample a new genetic architecture $G_{Q\rightarrow Y}$ from an architecture proposal distribution $R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})$.
3. Accept the new extended network structure G^{new} composed of G_Y^{new} and $G_{Q\rightarrow Y}^{new}$ given the biological knowledge weights W with a probability
$$A_G = \min\{1, \frac{P(Y|G^{new},X)P(G_Y^{new}|B,W)P(G_{Q\rightarrow Y}^{new})}{P(Y|G^{old},X)P(G_Y^{old}|B,W)P(G_{Q\rightarrow Y}^{old})} \frac{R(G_Y^{old}|G_Y^{new})R(G_{Q\rightarrow Y}^{old}|G_{Q\rightarrow Y}^{new})}{R(G_Y^{new}|G_Y^{old})R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})}\}.$$
4. For each biological knowledge k,
   1. Sample a new W_k^{new} of biological knowledge weight from a proposal distribution, $R(W_k^{new}|W_k^{old})$.
   2. Accept the new biological weight W_k^{new} given phenotype network G_Y with a probability
   $$A_{W_k} = \min\{1, \frac{P(G_Y|W_k^{new},W_{-k}^{old},B)}{P(G_Y|W^{old},B)} \frac{P(W_k^{new}|B)}{P(W_k^{old}|B)} \frac{R(W_k^{old}|W_k^{new})}{R(W_k^{new}|W_k^{old})}\}.$$

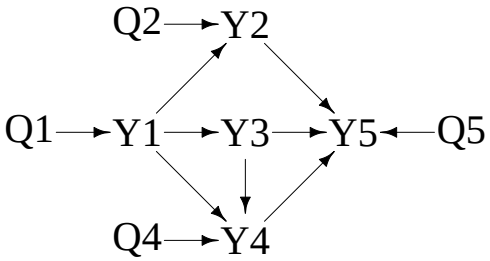
Markov Chain Monte Carlo Sampling

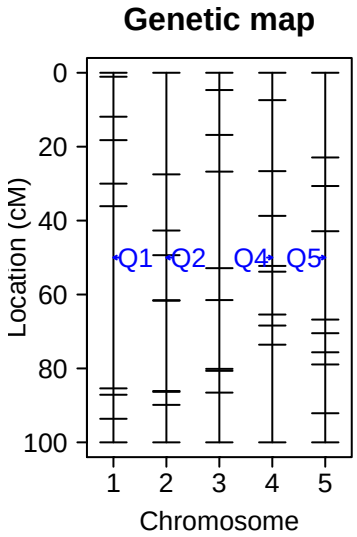
1. Sample a new network structure of phenotypes G_Y^{new} from a network structure proposal distribution $R(G_Y^{new}|G_Y^{old})$.
2. Given a network structure of phenotypes G_Y^{new}, sample a new genetic architecture $G_{Q\rightarrow Y}$ from an architecture proposal distribution $R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})$.
3. Accept the new extended network structure G^{new} composed of G_Y^{new} and $G_{Q\rightarrow Y}^{new}$ given the biological knowledge weights W with a probability
$$A_G = \min\{1, \frac{P(Y|G^{new},X)P(G_Y^{new}|B,W)P(G_{Q\rightarrow Y}^{new})}{P(Y|G^{old},X)P(G_Y^{old}|B,W)P(G_{Q\rightarrow Y}^{old})} \frac{R(G_Y^{old}|G_Y^{new})R(G_{Q\rightarrow Y}^{old}|G_{Q\rightarrow Y}^{new})}{R(G_Y^{new}|G_Y^{old})R(G_{Q\rightarrow Y}^{new}|G_{Q\rightarrow Y}^{old})}\}.$$
4. For each biological knowledge k,
   1. Sample a new W_k^{new} of biological knowledge weight from a proposal distribution, $R(W_k^{new}|W_k^{old})$.
   2. Accept the new biological weight W_k^{new} given phenotype network G_Y with a probability
   $$A_{W_k} = \min\{1, \frac{P(G_Y|W_k^{new},W_{-k}^{old},B)}{P(G_Y|W^{old},B)} \frac{P(W_k^{new}|B)}{P(W_k^{old}|B)} \frac{R(W_k^{old}|W_k^{new})}{R(W_k^{new}|W_k^{old})}\}.$$
5. Iterate the steps 1-4 until the chain converges.

Simulations

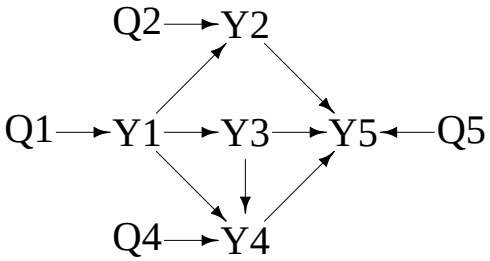


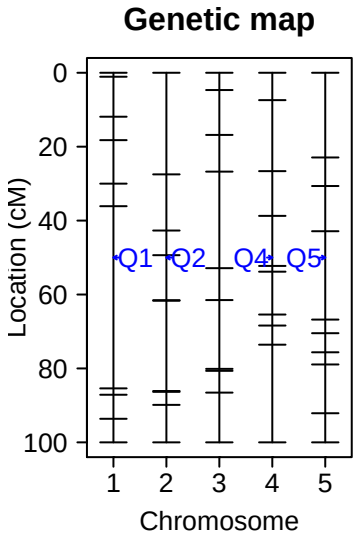
Simulations





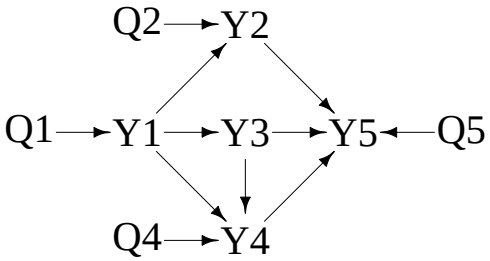
Simulations

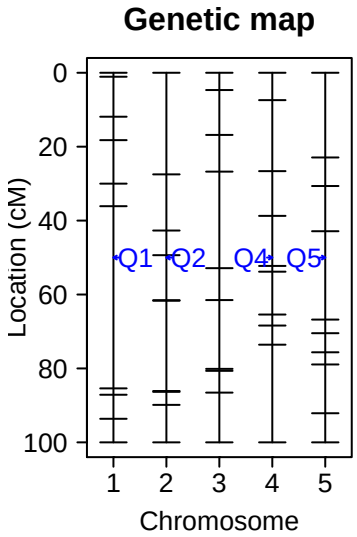




$\longrightarrow$ X for 500 mice in F2 population

Simulations

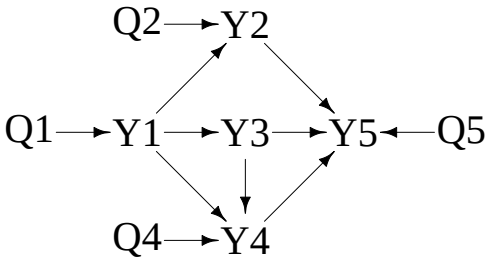


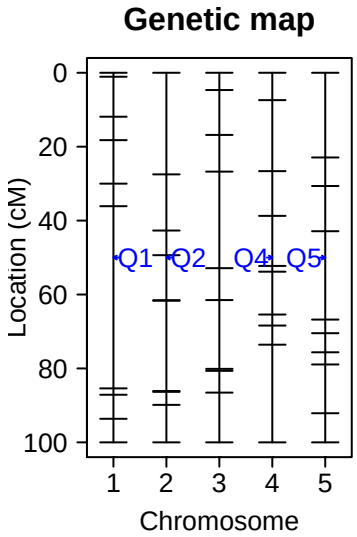


$\longrightarrow$ X for 500 mice in F2 population

$$\xrightarrow[\substack{\theta_{add} \sim U[0,0.5] \\ \theta_{dominance} \sim U[0,0.25] \\ \beta_{tv} \sim U[-0.5,0.5]}]{} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix}$$ for 500 mice.

Simulations

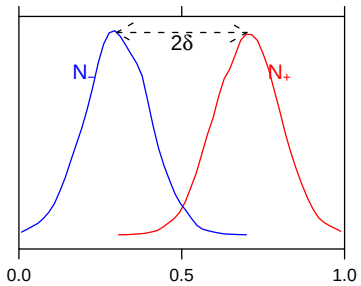




$\longrightarrow$ X for 500 mice in F2 population

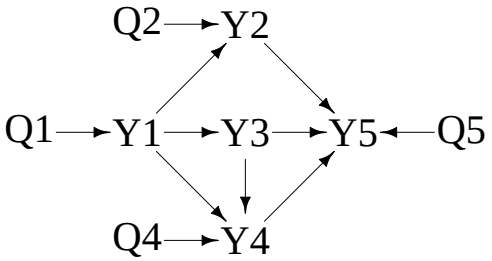
$$\xrightarrow[\substack{\theta_{add} \sim U[0,0.5] \\ \theta_{dominance} \sim U[0,0.25] \\ \beta_{tv} \sim U[-0.5,0.5]}]{} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix}$$ for 500 mice.

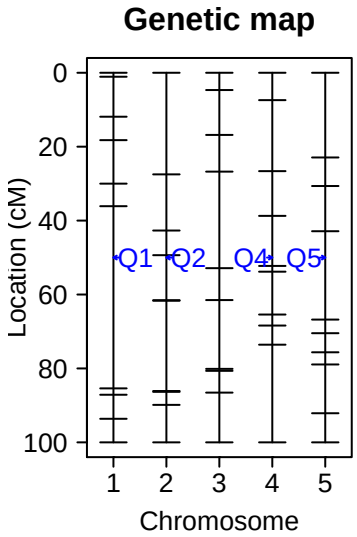
$$G_Y = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} & 1 & 1 & 1 & 0 \\ 0 & & 0 & 0 & 1 \\ 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & 0 & \end{pmatrix} \end{array}$$



$$\xrightarrow[Blue \sim N_-]{Red \sim N_+}$$ B $= \begin{pmatrix} & x & x & x & x \\ x & & x & x & x \\ x & x & & x & x \\ x & x & x & & x \\ x & x & x & x & \end{pmatrix}$

Simulations

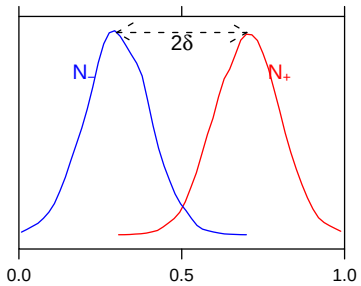




$\longrightarrow$ X for 500 mice in F2 population

$\xrightarrow[\theta_{dominance} \sim U[0,0.25]]{\theta_{add} \sim U[0,0.5]}$ $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix}$ for 500 mice.

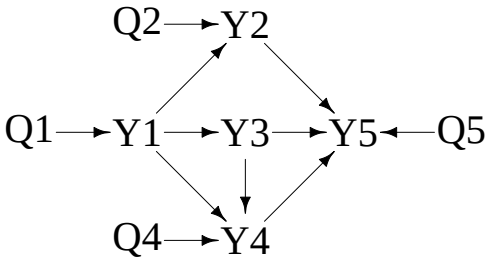
$$G_Y = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} & 1 & 1 & 1 & 0 \\ 0 & & 0 & 0 & 1 \\ 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & 0 & \end{pmatrix} \end{array}$$

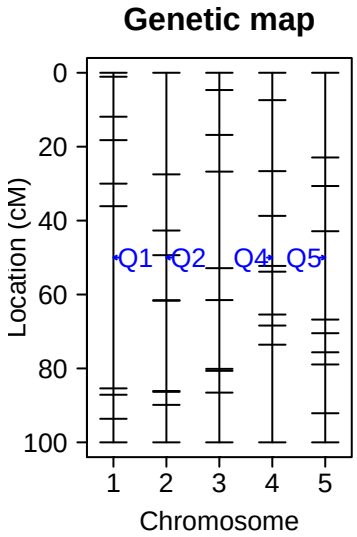


$\xrightarrow[Blue \sim N_-]{Red \sim N_+}$ B $= \begin{pmatrix} & x & x & x & x \\ x & & x & x & x \\ x & x & & x & x \\ x & x & x & & x \\ x & x & x & x & \end{pmatrix}$

X, Y, B are simulated for 100 times

Simulations

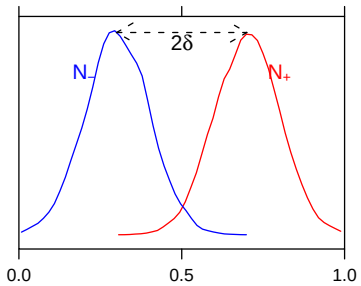




$\longrightarrow$ X for 500 mice in F2 population

$$\xrightarrow[\substack{\theta_{add} \sim U[0,0.5] \\ \theta_{dominance} \sim U[0,0.25] \\ \beta_{tv} \sim U[-0.5,0.5]}]{} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix}$$ for 500 mice.

$$G_Y = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} & 1 & 1 & 1 & 0 \\ 0 & & 0 & 0 & 1 \\ 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & 0 & \end{pmatrix} \end{array}$$



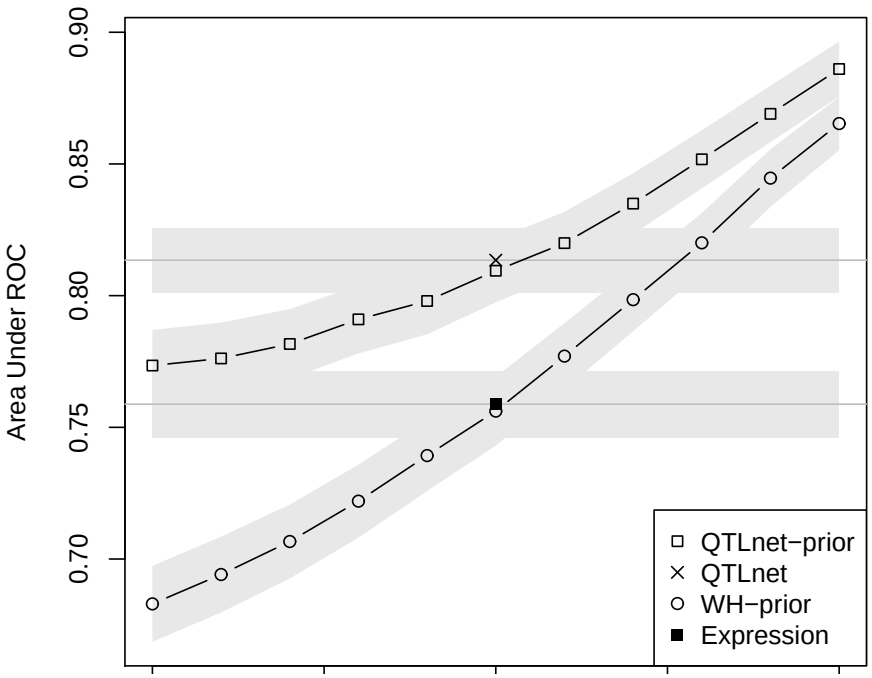
$\xrightarrow[Blue \sim N_-]{Red \sim N_+}$ B $= \begin{pmatrix} & x & x & x & x \\ x & & x & x & x \\ x & x & & x & x \\ x & x & x & & x \\ x & x & x & x & \end{pmatrix}$

X, Y, B are simulated for 100 times

for each $\delta \in \{\pm 0.5, \pm 0.25, \pm 0.2, \pm 0.15, \pm 0.1, \pm 0.05, 0\}$.

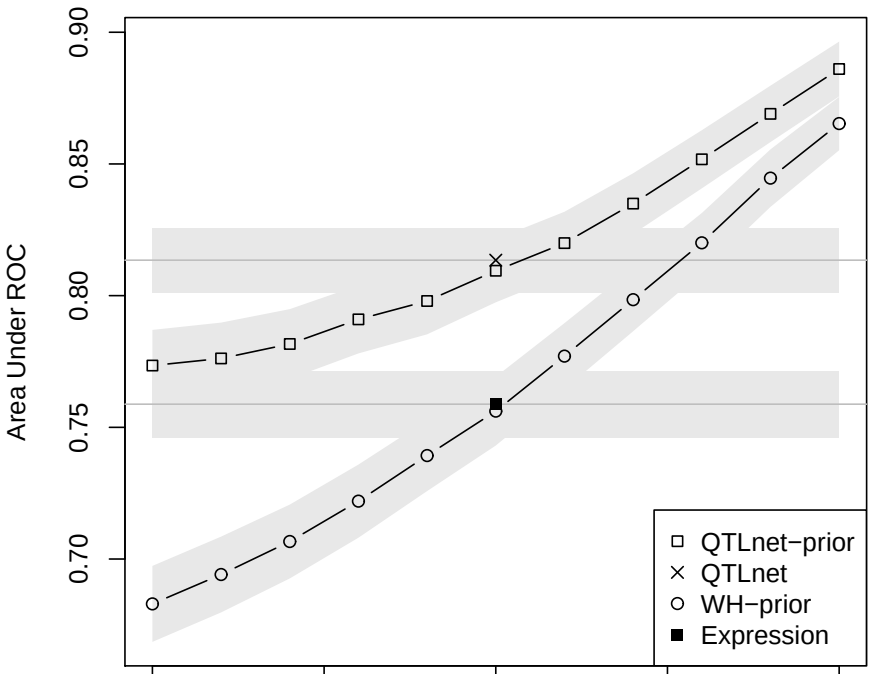
Method	Genetic Variation Information	Biological Knowledge
QTLnet-prior	YES	YES
QTLnet	YES	NO
WH-prior	NO	YES
Expression	NO	NO

ROC curves

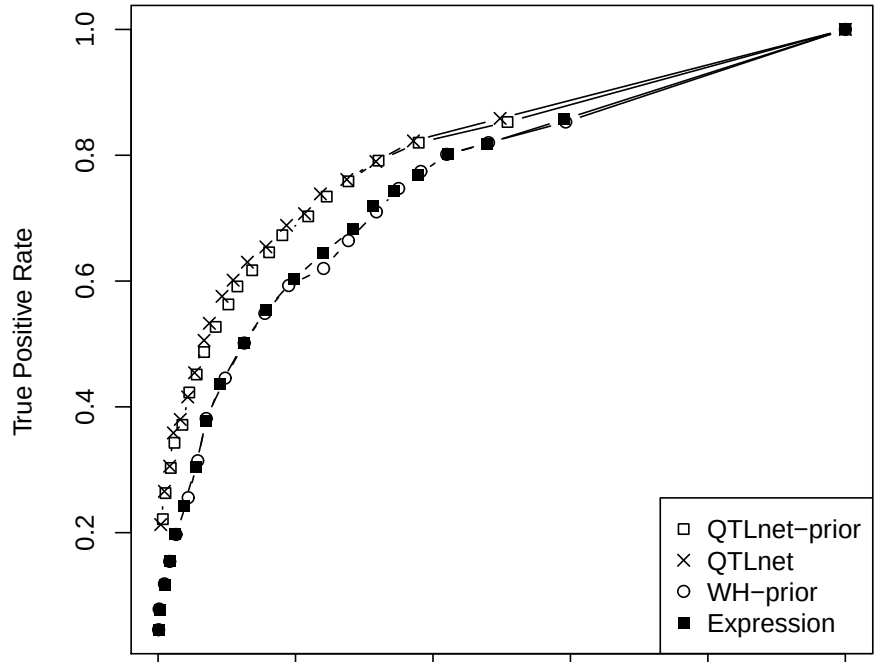


Area under ROC curves

ROC curves

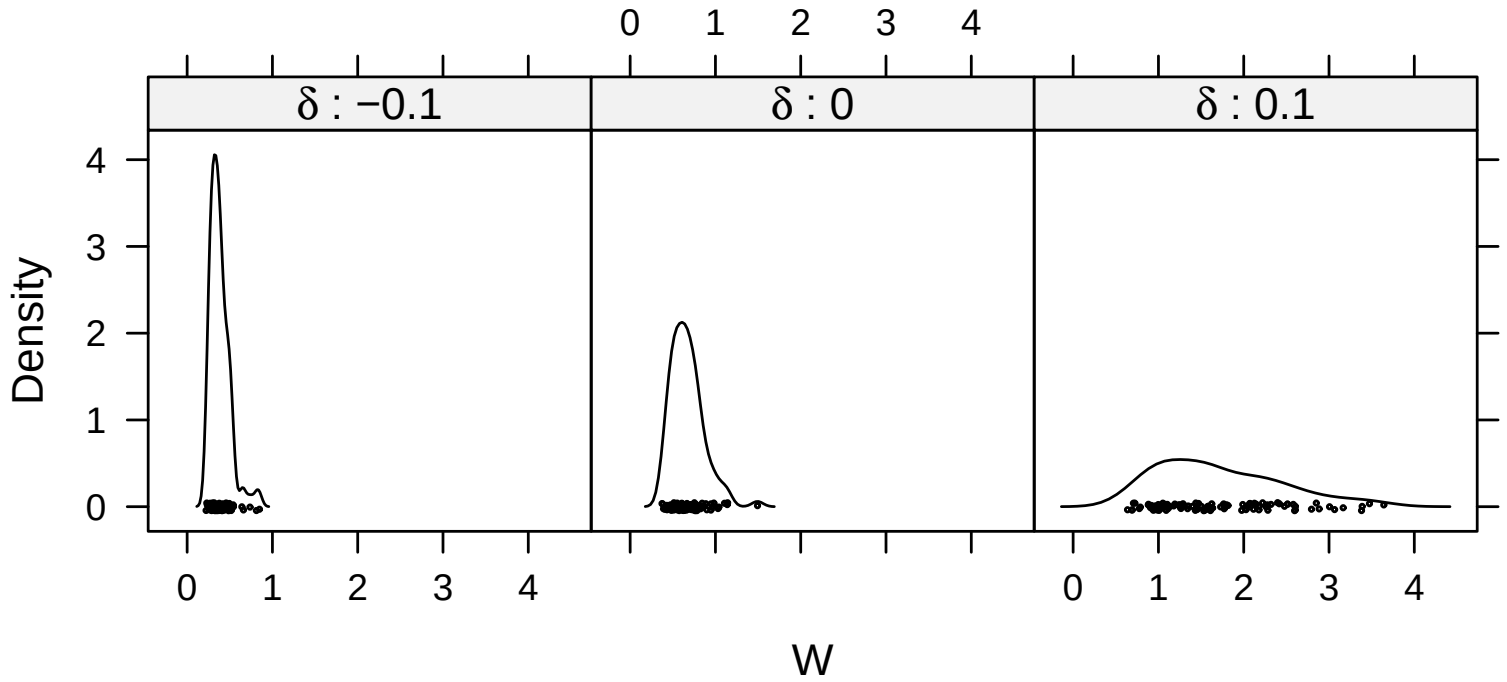


Area under ROC curves



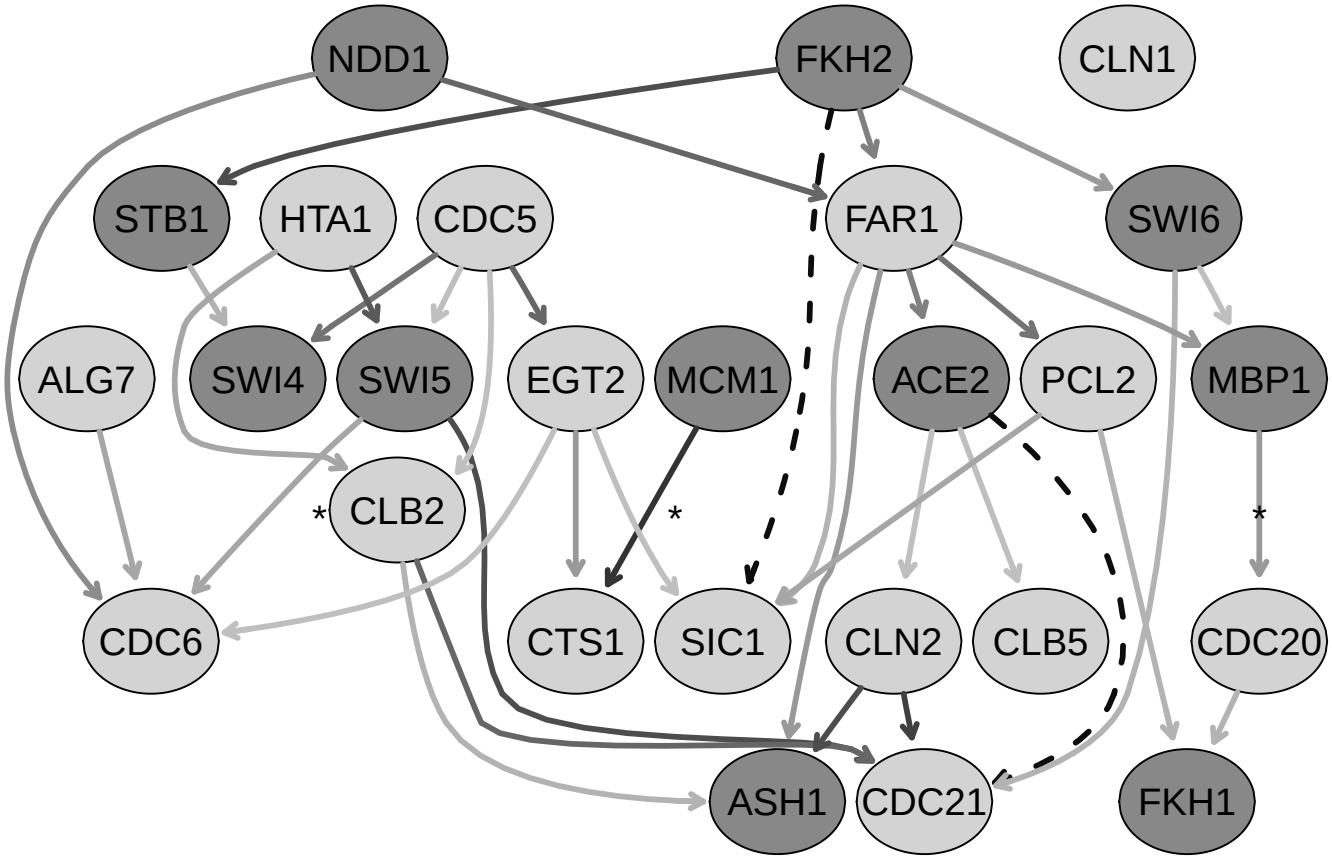
QTLnet-prior, WH-prior ($\delta = 0$) vs. QTLnet, Expression

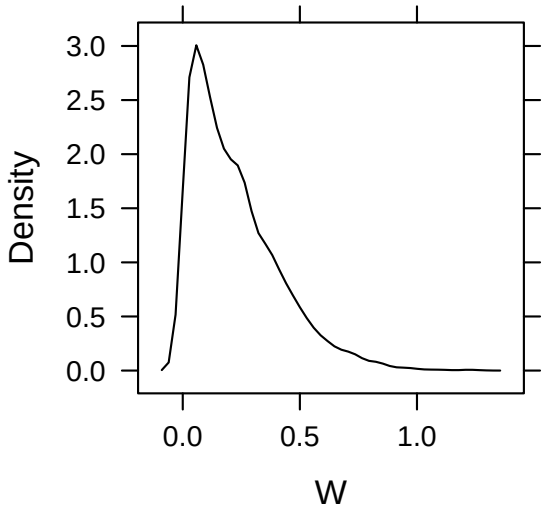
Convergence of W



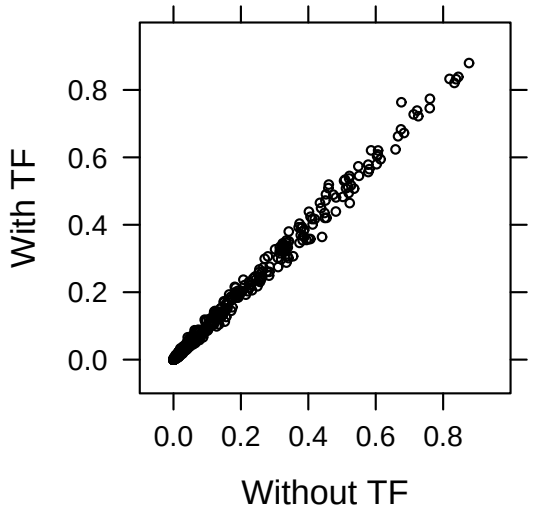
The distribution of median W of posterior sample by QTLnet-prior inference.

Yeast cell cycle analysis





The posterior distribution of weight W of TF-binding



Comparison of posterior probability of every possible edge

Conclusion

- When the prior knowledge is correct, the performance (area under ROC curve, proportion of recovered edges) is improved by prior knowledge. QTL mapping does not improve the performance.
- When the prior knowledge is incorrect, QTL mapping is important.
- When the prior knowledge is noninformative, we lose some power, but not too much.

References

- Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000) Using bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620.
- Chaibub Neto, E., Keller, M. P., Attie, A. D., and S., Y. B. (2010) Causal graphical models in systems genetics: A unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. *Ann. Appl. Stat.*, 4(1).
- Werhli, A. V. and Husmeier, D. (2007) Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge. *Statistical Applications in Genetics and Molecular Biology*, 6.
- Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., Bumgarner, R. E., and Schadt, E. E. (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet*, 40(7):854–861.

II. A causal gene network with genetic variations and latent variables

Motivation for latent variables

1. There could be unmeasured variables in a network.
2. Inference of a network may be done on a subset of candidate variables.

Motivation for latent variables

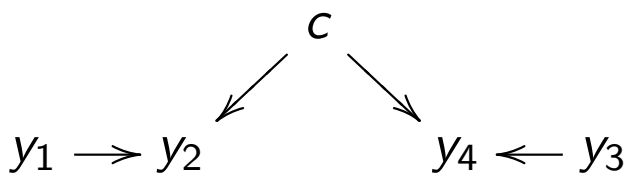
1. There could be unmeasured variables in a network.
2. Inference of a network may be done on a subset of candidate variables.

If the true network is a Bayesian network of measured and unmeasured variables, there may not exist a Bayesian network of measured variables.

Motivation for latent variables

1. There could be unmeasured variables in a network.
2. Inference of a network may be done on a subset of candidate variables.

If the true network is a Bayesian network of measured and unmeasured variables, there may not exist a Bayesian network of measured variables.

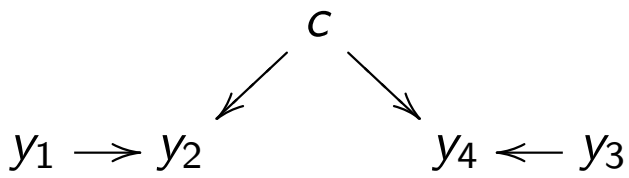


Let $\{y_1, y_2, y_3, y_4\}$: observed, c : unmeasured.

Motivation for latent variables

1. There could be unmeasured variables in a network.
2. Inference of a network may be done on a subset of candidate variables.

If the true network is a Bayesian network of measured and unmeasured variables, there may not exist a Bayesian network of measured variables.



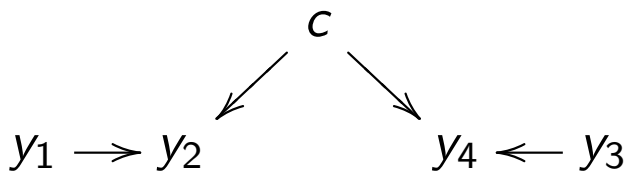
Let $\{y_1, y_2, y_3, y_4\}$: observed, c : unmeasured.
Conditional independence relations of observed variables:

$$y_2 \not\perp y_4 | \{y_1, y_3\}$$
$$y_1 \not\perp y_2 \qquad y_3 \not\perp y_4$$
$$y_1 \perp y_4 \qquad y_3 \perp y_2$$
$$y_1 \not\perp y_4 | y_2 \qquad y_3 \not\perp y_2 | y_4$$

Motivation for latent variables

1. There could be unmeasured variables in a network.
2. Inference of a network may be done on a subset of candidate variables.

If the true network is a Bayesian network of measured and unmeasured variables, there may not exist a Bayesian network of measured variables.



Let $\{y_1, y_2, y_3, y_4\}$: observed, c : unmeasured.
Conditional independence relations of observed variables:

$$y_2 \not\perp y_4 | \{y_1, y_3\}$$
$$y_1 \not\perp y_2 \qquad y_3 \not\perp y_4$$
$$y_1 \perp y_4 \qquad y_3 \perp y_2$$
$$y_1 \not\perp y_4 | y_2 \qquad y_3 \not\perp y_2 | y_4$$

Then, $y_1 \rightarrow y_2 \leftarrow y_4$ and $y_3 \rightarrow y_4 \leftarrow y_2$.

Introduction of ancestral graph

An ancestral graph is a graph whose vertexes are connected by at most one of undirected (—), directed ($\rightarrow$) or bidirected ($\leftrightarrow$) edges.

- Bidirected ($\leftrightarrow$) edges are associated with marginalization.
- Undirected (—) edges are associated with conditioning.

Introduction of ancestral graph

An ancestral graph is a graph whose vertexes are connected by at most one of undirected (—), directed ($\rightarrow$) or bidirected ($\leftrightarrow$) edges.

- Bidirected ($\leftrightarrow$) edges are associated with marginalization.
- Undirected (—) edges are associated with conditioning.

An ancestral graph holds the following conditions:

- there are no directed cycles;
- whenever there is an edge $x \leftrightarrow y$, then there is no directed path from x to y, or from y to x;
- if there is an undirected edge x — y then there are no vertex v such that $v \leftrightarrow x$, $v \leftrightarrow y$, $v \rightarrow x$, or $v \rightarrow y$.

Model

Y_{ti} be the phenotype for individual i and trait t.
Each phenotype is modeled as follows:

$$Y_{ti} = \mu^*_{ti} + \sum_{v \in pa(t)} \beta_{tv} Y_{vi} + \epsilon_{ti},$$

where $pa(t) = \{v : v \to t\}$ and $\mu^*_{ti} = \mu_t + X_i \; diag(\gamma_t) \; \theta_t$ is the QTL effect.

Model

Y_{ti} be the phenotype for individual i and trait t.
Each phenotype is modeled as follows:

$$Y_{ti} = \mu^*_{ti} + \sum_{v \in pa(t)} \beta_{tv} Y_{vi} + \epsilon_{ti},$$

where $pa(t) = \{v : v \to t\}$ and $\mu^*_{ti} = \mu_t + X_i \; diag(\gamma_t) \; \theta_t$ is the QTL effect.

$$\epsilon \sim N_T(0, \Omega),$$

where

$\Omega(t, s) = 0$ iff there is no bidirected edge between t and s.

QTLs to distinguish Markov equivalent directed ancestral graphs.

Theorem

Consider a class of Markov equivalent directed maximal ancestral graphs $\mathcal{G}_{\mathcal{Y}}$. Let Y_1 and Y_2 be any two adjacent nodes in the graphs in $\mathcal{G}_{\mathcal{Y}}$. Assume that for each such pair there exists at least two variables, Q_1 directly affecting Y_1 but not Y_2 and Q_2 directly affecting Y_2 but not Y_1. Let $\mathcal{G}$ represent the class of extended graphs. Then the graphs in $\mathcal{G}$ are not Markov equivalent.

QTLs to distinguish Markov equivalent directed ancestral graphs.

Theorem

Consider a class of Markov equivalent directed maximal ancestral graphs $\mathcal{G}_{\mathcal{Y}}$. Let Y_1 and Y_2 be any two adjacent nodes in the graphs in $\mathcal{G}_{\mathcal{Y}}$. Assume that for each such pair there exists at least two variables, Q_1 directly affecting Y_1 but not Y_2 and Q_2 directly affecting Y_2 but not Y_1. Let $\mathcal{G}$ represent the class of extended graphs. Then the graphs in $\mathcal{G}$ are not Markov equivalent.

$Y_1 \rightarrow Y_2$, $Y_1 \leftarrow Y_2$ and $Y_1 \leftrightarrow Y_2$

QTLs to distinguish Markov equivalent directed ancestral graphs.

Theorem

Consider a class of Markov equivalent directed maximal ancestral graphs $\mathcal{G}_{\mathcal{Y}}$. Let Y_1 and Y_2 be any two adjacent nodes in the graphs in $\mathcal{G}_{\mathcal{Y}}$. Assume that for each such pair there exists at least two variables, Q_1 directly affecting Y_1 but not Y_2 and Q_2 directly affecting Y_2 but not Y_1. Let $\mathcal{G}$ represent the class of extended graphs. Then the graphs in $\mathcal{G}$ are not Markov equivalent.

$Y_1 \rightarrow Y_2$, $Y_1 \leftarrow Y_2$ and $Y_1 \leftrightarrow Y_2$
$Q_1 \rightarrow Y_1 \rightarrow Y_2$ versus $Q_1 \rightarrow Y_1 \leftarrow Y_2$ or $Q_1 \rightarrow Y_1 \leftrightarrow Y_2$

QTLs to distinguish Markov equivalent directed ancestral graphs.

Theorem

Consider a class of Markov equivalent directed maximal ancestral graphs $\mathcal{G}_{\mathcal{Y}}$. Let Y_1 and Y_2 be any two adjacent nodes in the graphs in $\mathcal{G}_{\mathcal{Y}}$. Assume that for each such pair there exists at least two variables, Q_1 directly affecting Y_1 but not Y_2 and Q_2 directly affecting Y_2 but not Y_1. Let $\mathcal{G}$ represent the class of extended graphs. Then the graphs in $\mathcal{G}$ are not Markov equivalent.

$Y_1 \rightarrow Y_2$, $Y_1 \leftarrow Y_2$ and $Y_1 \leftrightarrow Y_2$
$Q_1 \rightarrow Y_1 \rightarrow Y_2$ versus $Q_1 \rightarrow Y_1 \leftarrow Y_2$ or $Q_1 \rightarrow Y_1 \leftrightarrow Y_2$
$Q_1 \rightarrow Y_1 \rightarrow Y_2 \leftarrow Q_2$ versus $Q_1 \rightarrow Y_1 \leftarrow Y_2 \leftarrow Q_2$ versus $Q_1 \rightarrow Y_1 \leftrightarrow Y_2 \leftarrow Q_2$.

Search algorithms

1. Constraint-based search : Conditional independence tests for a pair of nodes, Removes edges, orient edges (FCI)
2. Likelihood-based search : Search over DMAG models by their likelihoods

Search algorithms

1. Constraint-based search : Conditional independence tests for a pair of nodes, Removes edges, orient edges (FCI)
2. Likelihood-based search : Search over DMAG models by their likelihoods

Markov equivalence of G_1 and $G_2 \Leftrightarrow$ Distribution equivalence of G_1 and G_2 in a (parametric) family F?

1. Markov equivalence: G_1 and G_2 represent the same set of conditional independence relations.
2. Distribution equivalence with respect to F: $\forall \theta_{G_1}$, there exists a θ_{G_2} such that $p(Y \mid \theta_{G_1}, G_1) = p(Y \mid \theta_{G_2}, G_2)$, and vice versa.
   They represent the same set of joint probability distributions.

Parametric Family

$$Y_{ti} = \mu^*_{ti} + \sum_{v \in pa(t)} \beta_{tv} Y_{vi} + \epsilon_{ti}$$

$\epsilon \sim N_T(0, \Omega)$

$\Omega(t, s) = 0$ iff there is no bidirected edge between t and s.

Property

A set of linear equations and correlated errors fall into a homogeneous conditional Gaussian (HCG) family.

Parametric Family

$$Y_{ti} = \mu^*_{ti} + \sum_{v \in pa(t)} \beta_{tv} Y_{vi} + \epsilon_{ti}$$

$\epsilon \sim N_T(0, \Omega)$

$\Omega(t, s) = 0$ iff there is no bidirected edge between t and s.

Property

A set of linear equations and correlated errors fall into a homogeneous conditional Gaussian (HCG) family.

A conditional Gaussian (CG) family : the joint distribution of continuous variables are Gaussian conditional on discrete variables.

Parametric Family

$$Y_{ti} = \mu^*_{ti} + \sum_{v \in pa(t)} \beta_{tv} Y_{vi} + \epsilon_{ti}$$

$\epsilon \sim N_T(0, \Omega)$

$\Omega(t, s) = 0$ iff there is no bidirected edge between t and s.

Property

A set of linear equations and correlated errors fall into a homogeneous conditional Gaussian (HCG) family.

A conditional Gaussian (CG) family : the joint distribution of continuous variables are Gaussian conditional on discrete variables.
A homogeneous conditional Gaussian (HCG) family: the covariance in the conditional Gaussian distribution is independent of discrete variable values.

Markov equivalence $\Leftrightarrow$ Distribution equivalence in a HCG family

Theorem

For two Markov equivalent DMAGs G_1 and G_2, G_1 and G_2 differ only by $t \rightarrow v$ in G_1 and $t \leftrightarrow v$ in G_2. In a Gaussian distribution family, suppose the recursive equations for G_1 regarding t and v is represented by

$$Y_t = \mu_t + B_t(Y_{pa(t)} - \mu_{pa(t)}) + \epsilon_t$$
$$Y_v = \mu_v + B_v(Y_{pa(v)\setminus\{t\}} - \mu_{pa(v)}) + b_{tv}(Y_t - \mu_t) + \epsilon_v$$

where $cov(\epsilon_t, \epsilon_v) = 0$. Then, the re-parametrization below for G_2 regarding t and v gives out the same joint probability to the joint probability of G_1.

$$Y_t = \mu_t^* + B_t^*(Y_{pa(t)} - \mu_{pa(t)}^*) + \epsilon_t^*$$
$$Y_v = \mu_v^* + B_v^*(Y_{pa(v)\setminus\{t\}} - \mu_{pa(v)}^*) + \epsilon_v^*$$

where

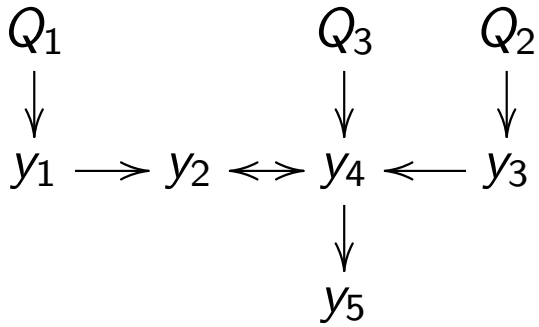
1. $B_v^* = B_v + b_{tv}B_t$
2. $var(\epsilon_v^*) = \sigma_v + b_{tv}^2\sigma_t$
3. $cov(\epsilon_v^*, \epsilon_{sp(v)}) = \sigma_{v,sp(v)} + b_{tv}\sigma_{t,sp(v)}$
4. $cov(\epsilon_t, \epsilon_v^*) = b_{tv}\sigma_t$
5. $B_t^* = B_t$
6. $Var(\epsilon_t^*) = \sigma_t$
7. $cov(\epsilon_t^*, \epsilon_{sp(t)}) = \sigma_{t,sp(t)}$.

Algorithm - MCMC

1. Divide a DMAG G_0 into bidirected graph G_0^B and directed graph and G_0^D.
2. Propose a new directed graph G^D from G_0^D by a DAG proposal distribution $R(G^D|G_0^D)$.
3. For each node, get a list of ancestors or descendants in G^D. Then, get a list of possible bidirected edges in $G \setminus G^D$. Propose new bidirected edges G^B by Bernoulli distribution for each possible bidirected edge with probability pB.
4. If $G = G^D \oplus G^B$ is not a maximal ancestral graph, make it to be maximal: $Max(G)$. Obtain several G^B and their proposal probabilities to become equivalent to $Max(G)$. Its proposal distribution is $R(Max(G)|G^D)$.
5. Accept the new DMAG $G_1 = Max(G)$ with a probability,

$$\min\{1, \frac{P(Y|G_1)}{P(Y|G_0)} \frac{R(G_0^D|G_1^D)R(G_0|G_0^D)}{R(G_1^D|G_0^D)R(G_0|G_1^D)}\}.$$

Simulations



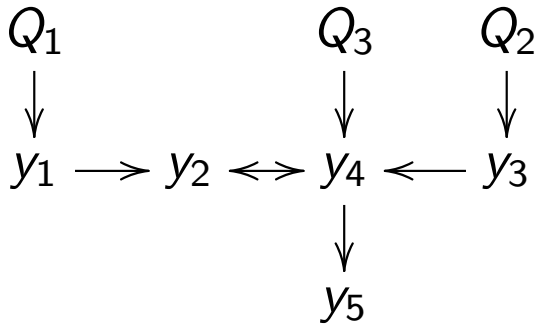
$\theta_{add} \sim U[0, 0.5]$

$\theta_{dominance} \sim U[0, 0.25]$

$\beta_{tv} \sim U[0.2, 0.5] \times Bernoulli((-1, 1))$

for 500 individuals.

Simulations



$\theta_{add} \sim U[0, 0.5]$

$\theta_{dominance} \sim U[0, 0.25]$

$\beta_{tv} \sim U[0.2, 0.5] \times Bernoulli((-1, 1))$

for 500 individuals.

Preliminary Result: The inferred skeleton has 1.35 edge difference to the true skeleton on average from 20 simulations.

Conclusion

- QTL can be included to distinguish Markov equivalent ancestral graphs.

Conclusion

- QTL can be included to distinguish Markov equivalent ancestral graphs.
- Our model is a homogeneous conditional Gaussian (HCG) family.

Conclusion

- QTL can be included to distinguish Markov equivalent ancestral graphs.
- Our model is a homogeneous conditional Gaussian (HCG) family.
- Distribution equivalence in a HCG family $\Leftrightarrow$ Markov equivalence.

References

- Richardson, T. S. and Spirtes, P. (2002) Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030.
- Ali, R. A., Richardson, T. S., Spirtes, P. (2009) Markov equivalence for ancestral graphs. *Annals of Statistics*, 37(5B):2808–2837.
- Lauritzen, S. L. Lauritzen. Graphical Models. Oxford University Press, 1996

Future Research Plan