ETH zürich

Nested Effects Models and perturbation experiments

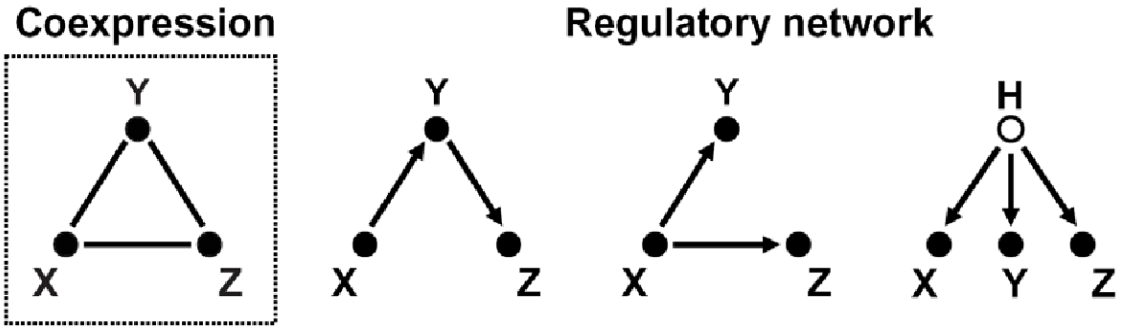
Martin Pirkl and Niko Beerenwinkel

Outline

- Interventions and perturbations
- Nested Effects Models
- Network inference
- Time series data
- Single-cell data

Interventions and perturbations

Learning from interventions



- Correlation or conditional independence can not, in general, completely resolve the network topology.

- Interventions:

		silencing		
		X	***Y***	***Z***
	X	1	0	0
affects	***Y***	1	1	0
	Z	1	1	1

Perturbation experiments

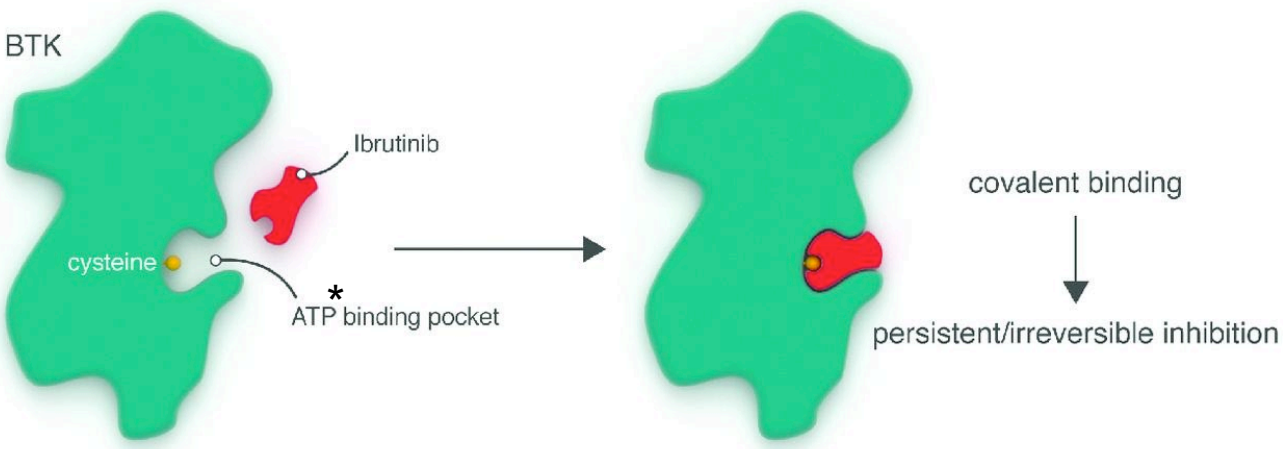
In each experiment, specific molecules are removed and a phenotype is measured.

Experiment 1	Experiment 2	Experiment 3	…	Experiment *h*
Knock-down: -/control	Knock-down: Gene A	Knock-down: Gene B	…	Knock-down: Gene X
Readout: ✓	Readout: ✗	Readout: ✗	…	Readout: ✓

The readout can be of different types. E.g.,

- Survival: is the gene essential for cell viability
- Gene expression: the expression of which genes goes up/down compared to cells without a knock-down

Drug inhibitors



Knock-out of the Bruton tyrosine kinase (BTK). Ibrutinib is able to dock in the binding socket. This leads to inhibition of the protein activity.

Wiestner et al., 2015

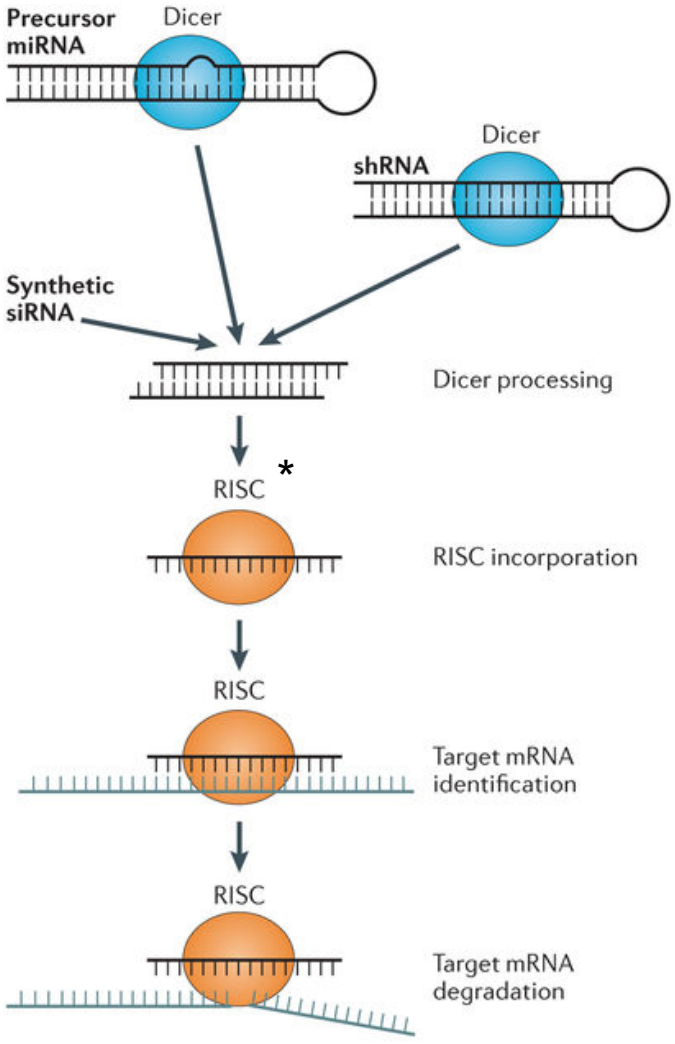
*Adenosine triphosphate

RNAi: siRNA knock-down

RNA interference (RNAi) includes methods which directly attack mRNA molecules with complementary RNA strands.

Micro RNAs (miRNAs) are small interfering RNAs (siRNAs) and naturally occurring. They are important for gene regulation by inhibiting a family of genes.

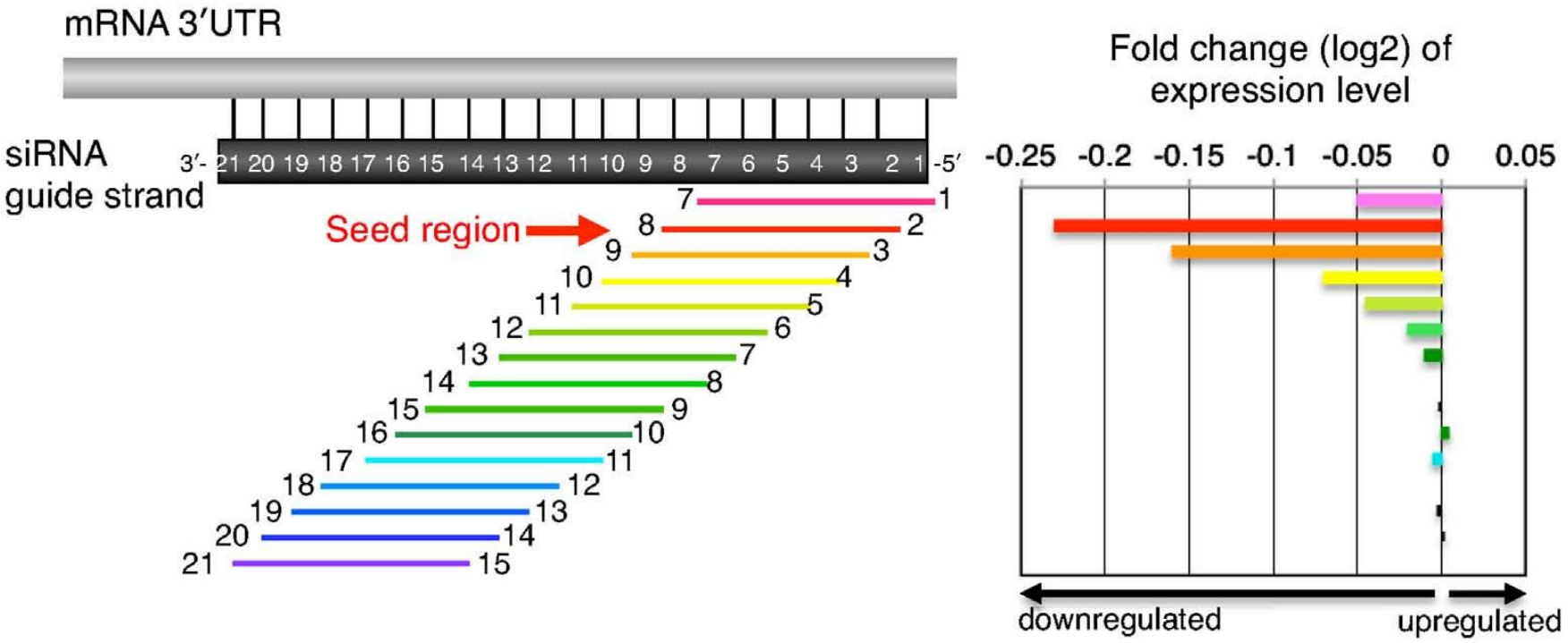
Synthetic siRNAs and small hairpin RNAs (shRNAs) can be artificially introduced into the cell. They use a similar mechanism as miRNAs to target specific genes and achieve a knock-down.



Wittrup et al., 2015

*RNA induced silencing complex

siRNA on- and off-target effects

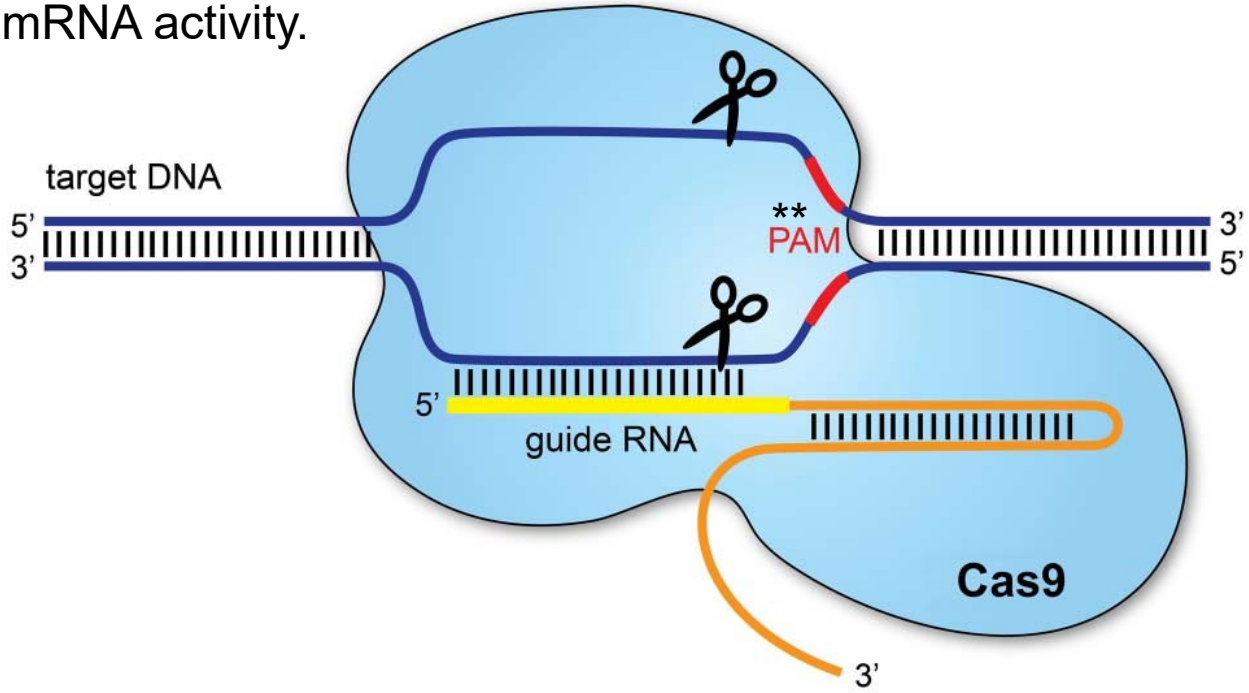


Depending on the sequence specificity (seed region) these knock-downs have off-target effects: not only is the targeted mRNA suppressed, but also other similar mRNAs.

Ui-Tei et al., 2015

CRISPR*/Cas genome editing

The CRISPR/Cas-system is used to directly edit the DNA of the genome. The identification of the targeted area works similar to RNAi. A small guiding RNA (sgRNA) with a complementary sequence is used together with the Cas9 protein to cut out pieces of DNA, which can lead to a knock-out of the gene. Contrary to RNAi this can result in the total depletion of the mRNA and not just a knock-down with reduced mRNA activity.

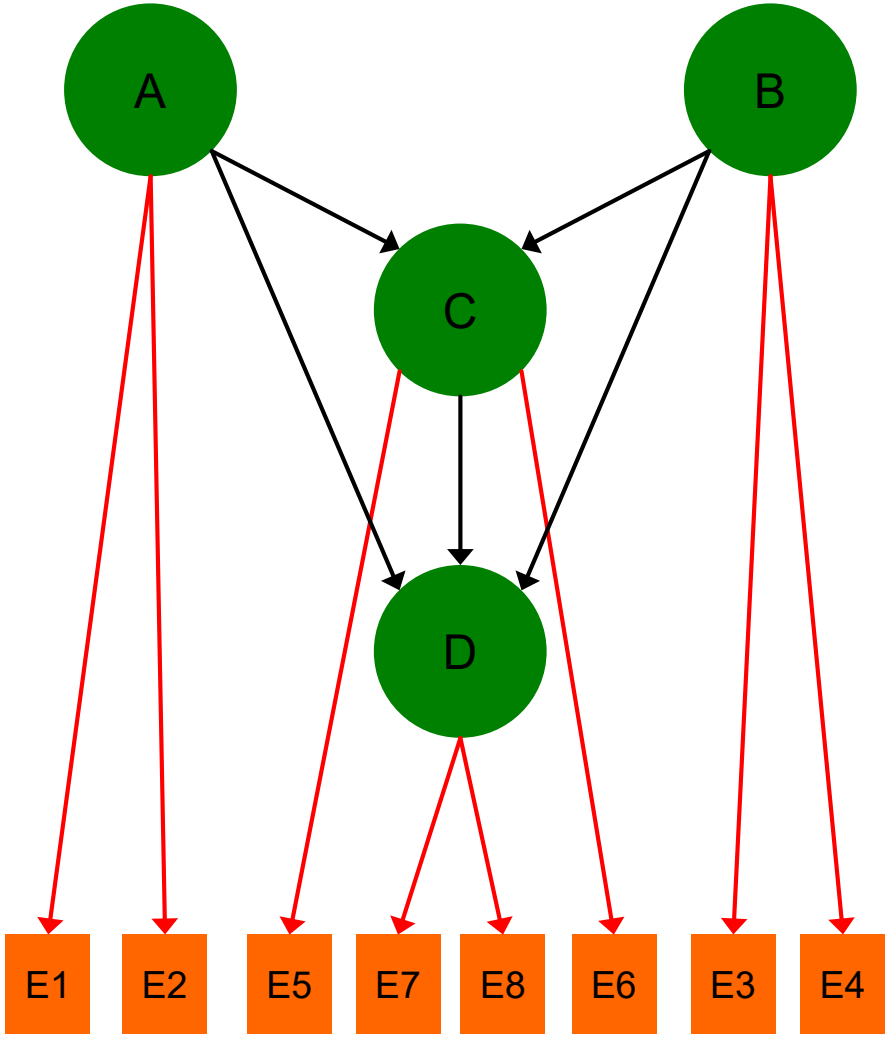


Redman et al., 2016

*Clustered Regularly Interspaced Short Palindromic Repeats **protospacer adjacent motif

Nested Effects Models

Nested Effects Models



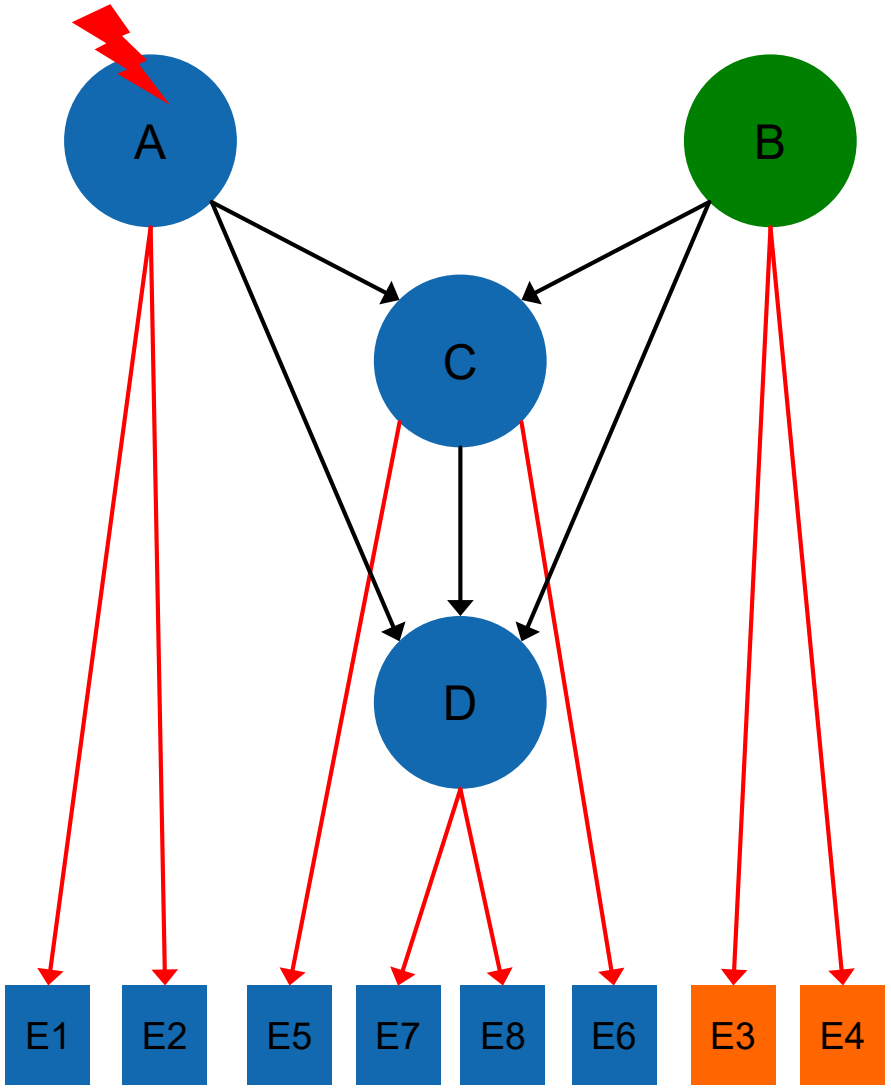
Unobserved, perturbed signaling genes (S-genes $i = 1, \ldots, n$)

Signaling pathway Φ (adjacency matrix)

Observed effect reporters (E-genes $j = 1, \ldots, m$)

E-gene regulation Θ with $\theta_{i,j} = 1$, if E-gene *j* is regulated by S-gene *i* (parsimony assumption: at most one parent per E-gene)

Nested Effects Models



Unobserved, perturbed signaling genes (S-genes $i = 1, \ldots, n$)

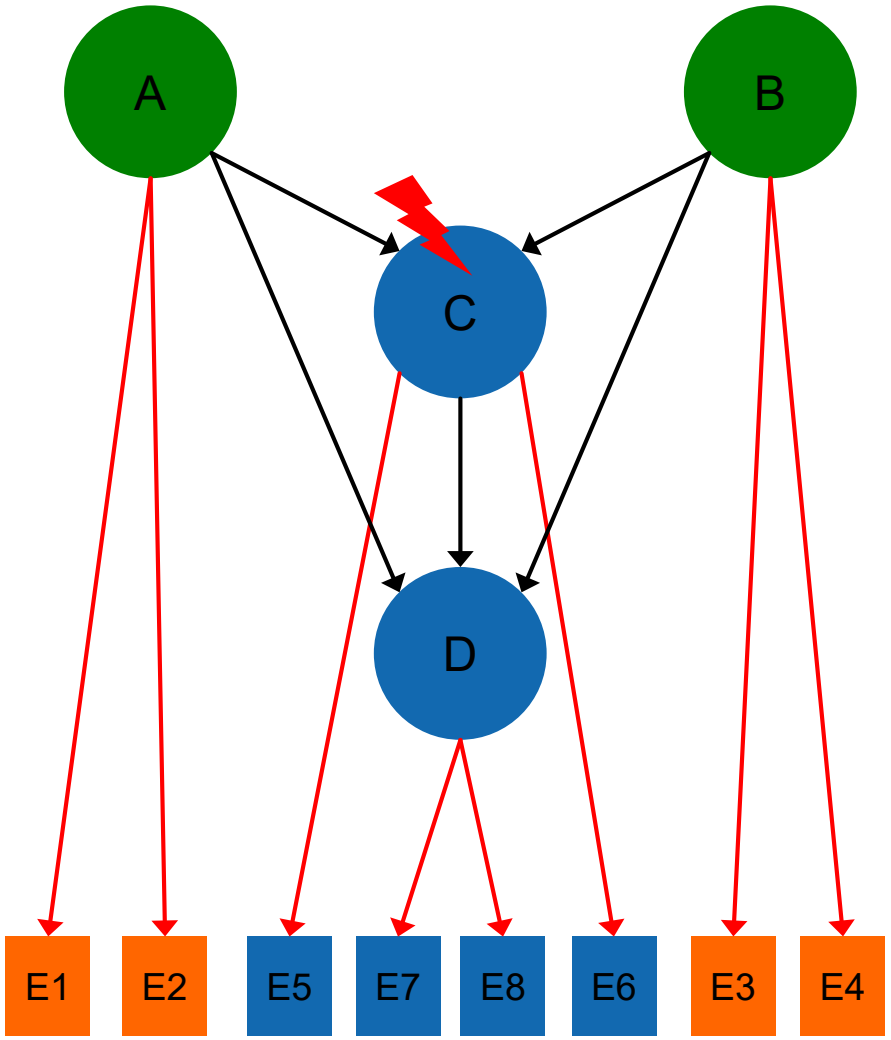
Signaling pathway Φ (adjacency matrix)

Observed effect reporters (E-genes $j = 1, \ldots, m$)

E-gene regulation Θ with $\theta_{i,j} = 1$, if E-gene *j* is regulated by S-gene *i* (parsimony assumption: at most one parent per E-gene)

A knock-down of A is propagated via C and D and reaches E-genes controlled by both, but not the E-genes controlled by B.

Nested Effects Models



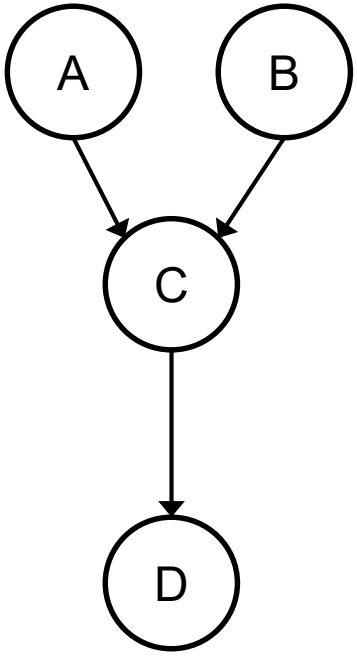
Unobserved, perturbed signaling genes (S-genes $i = 1, \ldots, n$)

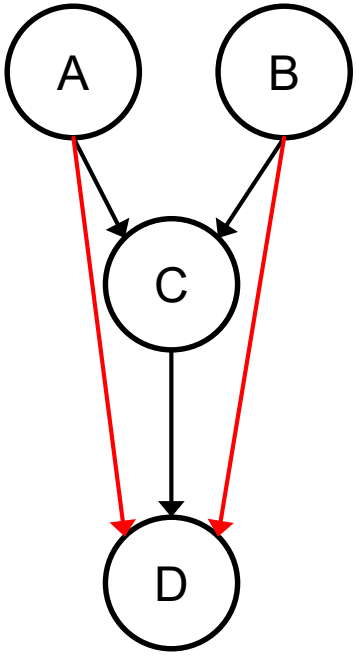
Signaling pathway Φ (adjacency matrix)

Observed effect reporters (E-genes $j = 1, \ldots, m$)

E-gene regulation Θ with $\theta_{i,j} = 1$, if E-gene *j* is regulated by S-gene *i* (parsimony assumption: at most one parent per E-gene)

Identifiability w.r.t. transitive closure





$\Phi =$

	A	B	C	D
A	1	0	1	0
B	0	1	1	0
C	0	0	1	1
D	0	0	0	1

$\Phi^2 =$

	A	B	C	D
A	1	0	2	1
B	0	1	2	1
C	0	0	1	2
D	0	0	0	1

Subset relationships are transitive.

We cannot say, if the transitive edges (feed-forward loops) are true or not, even with perfect data.

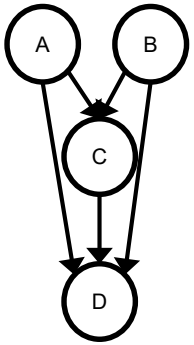
Computing transitive closure:

1. $\text{diag}(\Phi) \leftarrow (1, \ldots, 1)$
2. $\Phi \leftarrow \Phi^n$, s.t. $\Phi^n = \Phi^{n+1}$
3. $\Phi_{ii'} \leftarrow \max(\Phi_{ii'}, 1)$, all i, i'

or alternatively:

1. $\Phi \leftarrow \Phi + \Phi^2 + \Phi^3 + \ldots$
   (path lengths: 1, 2, 3, …)
2. $\Phi_{ii'} \leftarrow \max(\Phi_{ii'}, 1)$, all i, i'

Expected data



$\Phi =$

	A	B	C	D
A	1	0	1	1
B	0	1	1	1
C	0	0	1	1
D	0	0	0	1

$\Theta =$

	E1	E2	E3	E4	E5	E6	E7	E8
A	1	1	0	0	0	0	0	0
B	0	0	1	1	0	0	0	0
C	0	0	0	0	1	1	0	0
D	0	0	0	0	0	0	1	1

Expected effect pattern for the signalling pathway Φ and the E-gene attachments Θ:

$F = \Phi\Theta =$

	E1	E2	E3	E4	E5	E6	E7	E8
A	1	1	0	0	1	1	1	1
B	0	0	1	1	1	1	1	1
C	0	0	0	0	1	1	1	1
D	0	0	0	0	0	0	1	1

Each row is the expected effect pattern for a knock-down.

For example, $f_{A,E8} = (1\ 0\ 1\ 1)(0\ 0\ 0\ 1)^T = 1$, because A connects to D and D to E8.

Discrete case

Data can often be binarized to effect (1) or no effect (0). Assuming one perturbation experiment for each S-gene, the binarized data matrix $D = (e_{ji})$ has entries $e_{ji} = 1$ if S-gene *i* had an effect on E-gene *j*, and $e_{ji} = 0$ otherwise. False positives and false negatives occur with probability α and β, respectively.

The probability of observing effect *j* under perturbation *i* given $(f_{ij}) = F = \Phi\Theta$ is then

$$P(e_{ji} = 0 \mid f_{ij} = 0) = 1 - \alpha$$
$$P(e_{ji} = 1 \mid f_{ij} = 0) = \alpha$$
$$P(e_{ji} = 0 \mid f_{ij} = 1) = \beta$$
$$P(e_{ji} = 1 \mid f_{ij} = 1) = 1 - \beta$$

$D =$

	A	B	C	D
E1	1	0	0	0
E2	1	0	1	0
E3	0	1	0	1
E4	0	1	0	0
E5	1	0	1	0
E6	1	1	1	0
E7	1	0	1	1
E8	1	1	1	1

We estimate α and β from control experiments independently of the NEM.

Experiments with replicates

If experiments do not correspond 1-to-1 to S-genes, we denote the observed data by $\widetilde{D} = (\tilde{e}_{jg})$, for all experiments $g = 1, \ldots, h$. The S-gene perturbed in experiment g is denoted $S(g)$. For example, if g_1 and g_2 are replicates of the knock-down of the same S-gene i, then $S(g_1) = S(g_2) = i$.

$$P(\tilde{e}_{jg} = 0 \mid f_{S(g)j} = 0) = 1 - \alpha$$
$$P(\tilde{e}_{jg} = 1 \mid f_{S(g)j} = 0) = \alpha$$
$$P(\tilde{e}_{jg} = 0 \mid f_{S(g)j} = 1) = \beta$$
$$P(\tilde{e}_{jg} = 1 \mid f_{S(g)j} = 1) = 1 - \beta$$

$\widetilde{D}$ =

	1	2	3	4	…	h
E1	1	1	0	0	…	0
E2	1	1	0	1	…	0
E3	0	0	1	0	…	1
E4	0	0	1	0	…	0
E5	1	1	0	1	…	0
E6	1	1	1	1	…	0
E7	1	0	0	1	…	1
E8	1	1	1	1	…	1
S =	1	1	2	3	…	n

Marginal likelihood

The marginal likelihood for scoring a network with *n* S-genes, *m* E-genes and *h* knock-down experiments is computed by

$$P(\tilde{D} \mid \phi) = \int P(\tilde{D} \mid \phi, \theta) P(\theta \mid \phi) d\theta$$

$$\overset{(1,2)}{=} \prod_{j=1}^{m} \int P(\tilde{D}_{j\cdot} \mid \phi, \theta_{\cdot j}) P(\theta_{\cdot j} \mid \phi) d\theta_{\cdot j}$$

$$\overset{(3,4)}{=} \frac{1}{n^m} \prod_{j=1}^{m} \sum_{i=1}^{n} P(\tilde{D}_{j\cdot} \mid \phi, \theta_{ij} = 1)$$

$$\overset{(1)}{=} \frac{1}{n^m} \prod_{j=1}^{m} \sum_{i=1}^{n} \prod_{g=1}^{h} P(\tilde{e}_{jg} \mid f_{S(g)j}, \theta_{ij} = 1)$$

(1) Conditional independence of observations:

$$P(\tilde{D} \mid \phi, \theta) = \prod_{j=1}^{m} P(\tilde{D}_{j\cdot} \mid F_{\cdot j}) = \prod_{j=1}^{m} \prod_{g=1}^{h} P(\tilde{e}_{jg} \mid f_{S(g)j})$$

(2) Conditional independence of E-genes:

$$P(\theta \mid \phi) = \prod_{j=1}^{m} P(\theta_{\cdot j} \mid \phi)$$

(3) Uniform prior for E-genes:

$$P(\theta_{ij} = 1 \mid \phi) = \frac{1}{n} \; \forall i, j$$

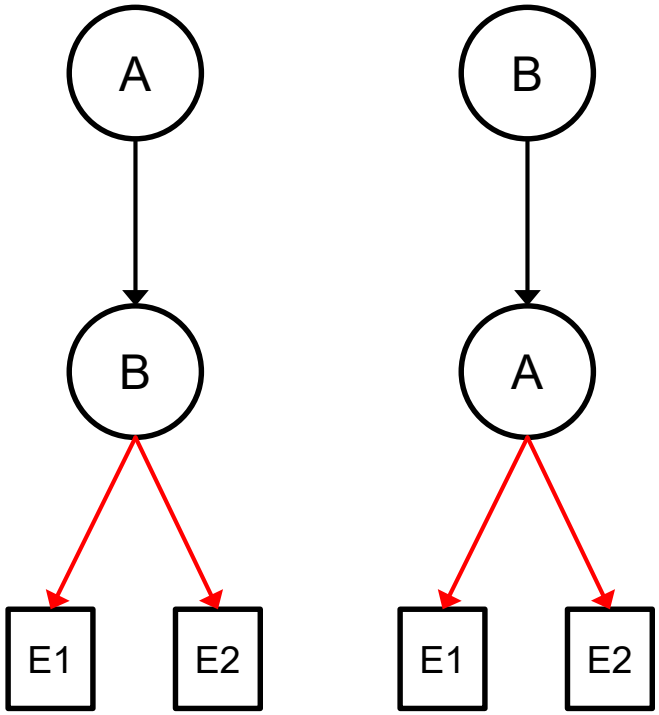
(4) Parsimony assumption

with $\tilde{D}_{j\cdot}$ as the *j*-th row of $\tilde{D}$, and $\theta_{\cdot j}$ and $F_{\cdot j}$ the *j*-th column of θ and F, respectively.

Undistinguishable S-genes

E-genes E1 and E2 react to both knock-downs. However, we do not observe E-genes reacting to the knock-down of A and not B and vice versa. The hierarchy of the S-genes cannot be resolved. This is usually depicted by a bi-directional edge.
The marginal likelihood for the bi-directional edge is higher than for the two uni-directional ones.

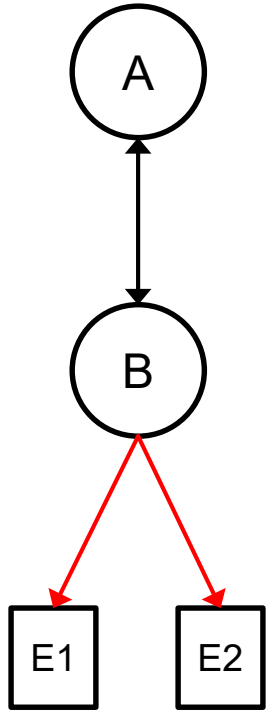
$$D = \begin{array}{c|cc} & A & B \\ E1 & 1 & 1 \\ E2 & 1 & 1 \end{array}$$



$$\phi_1 = \begin{array}{c|cc} & A & B \\ A & 1 & 1 \\ B & 0 & 1 \end{array} \qquad \phi_2 = \begin{array}{c|cc} & A & B \\ A & 1 & 0 \\ B & 1 & 1 \end{array}$$

$$\theta_1 = \begin{array}{c|cc} & E1 & E2 \\ A & 0 & 0 \\ B & 1 & 1 \end{array} \qquad \theta_2 = \begin{array}{c|cc} & E1 & E2 \\ A & 1 & 1 \\ B & 0 & 0 \end{array}$$

$$\phi_1\theta_1 = \phi_2\theta_2 = F = \begin{array}{c|cc} & E1 & E2 \\ A & 1 & 1 \\ B & 1 & 1 \end{array}$$



Continuous case

For continuous gene expression values d_{ji}, we can compute log-likelihood ratios for each E-gene *j* and S-gene *i* as

$$r_{ji} = \log \frac{P(d_{ji} \mid \mathbf{1})}{P(d_{ji} \mid N)}$$

where *N* is the null model predicting no effects and **1** the alternative model predicting only effects. Replicates of the same S-gene knock-down are usually summarized in the probabilities.

The log likelihood ratio matrix $R = (r_{ji})$ has values $r_{ji} > 0$ if S-gene *i* had an effect on E-gene *j*, and $r_{ji} \leq 0$ if S-gene *i* had no effect.

E.g., the R-package *limma* is designed for differential gene expression analysis (large difference between control and knock-down corresponds to an effect). The algorithm computes the probability *p* that a gene is differentially expressed and

$$r = \log \frac{p}{1-p}.$$

Log-likelihood ratio

Let *N* be the null model, which does not predict any effect. For fixed α and β, we compute the log likelihood ratio for a given model (Φ, Θ) by

$$\log P(D \mid \phi, \theta) - \log P(D \mid N) = \sum_{i,j} \log \frac{P(d_{ji} \mid f_{ij})}{P(d_{ji} \mid N)}$$

$$= \sum_{i,j} \begin{cases} r_{ji}, & \text{if } f_{ij} = 1 \\ 0, & \text{if } f_{ij} = 0 \end{cases} = \sum_{i,j} f_{ij} r_{ji} = \sum_{i} (FR)_{ii} = \text{trace}(FR)$$

We use the log-likelihood ratio for model optimization, since

$$\log P(D \mid \phi, \theta) = \text{trace}(FR) + C = \text{trace}(\phi\theta R) + C$$

Estimation of Θ given Φ

It is easier to estimate the E-gene attachments Θ for a given pathway Φ, instead of jointly optimizing both parameters.

Given Φ we compute the optimal attachment of each E-gene to an S-gene by $P = (p_{ji}) = R\Phi$ and set $\theta_{ij} = 1$ if $p_{ji} = \max\{p_{ji'} : i' = 1, \ldots, n\}$ (maximum a posteriori).

Since we estimate Θ from P, we can speed up the computation of the log likelihood ratio:

$$\log P(D \mid \phi, \theta) - \log P(D \mid N) = \mathrm{trace}(\phi\theta R) = \mathrm{trace}(\theta R\phi) = \mathrm{trace}(\theta P).$$

We do not have to compute $F = \Phi\Theta$.

Network inference

Exhaustive and pairwise search

For a small number of S-genes (~<5) we can exhaustively search for the optimal pathway Φ by enumerating and scoring all possible networks. This becomes infeasible for more S-genes.

Pairwise search:
We optimize the pathway for every pair of S-genes *A, B*. We have to score the following combinations with the marginal log likelihood for the two S-gene model:

- *A* upstream of *B*
- *B* upstream of *A*
- *A* and *B* indistinguishable
- *A* and *B* unconnected

All pairwise relations are combined to one consensus pathway of all S-genes by graph merging.

All edges are treated independently and thus transitivity is not guaranteed.

Triplet search

The pairwise search has been extended to a triplet search, which has a higher accuracy at the cost of speed.

1. For *n* S-genes, we optimize all $\binom{n}{3}$ triplets. For each triplet, we have to score 29 unique networks.

2. We obtain the frequency of each edge as

$$f(A \rightarrow B) = \frac{1}{n-2} \sum_{C \notin \{A,B\}} 1\{(A \rightarrow B) \in M_{ABC}\}$$

with the triplet model M_{ABC} and indicator function 1{}.

We include edges with frequency $f(A \rightarrow B) > \lambda$, e.g. $\lambda = 0.5$.

Greedy search

1. Start with a random or specific network (e.g. fully disconnected)
2. Score each elementary change (added/deleted/reversed edge)
3. Execute the change which scores the highest
4. Repeat steps 2-3 until no change improves the score

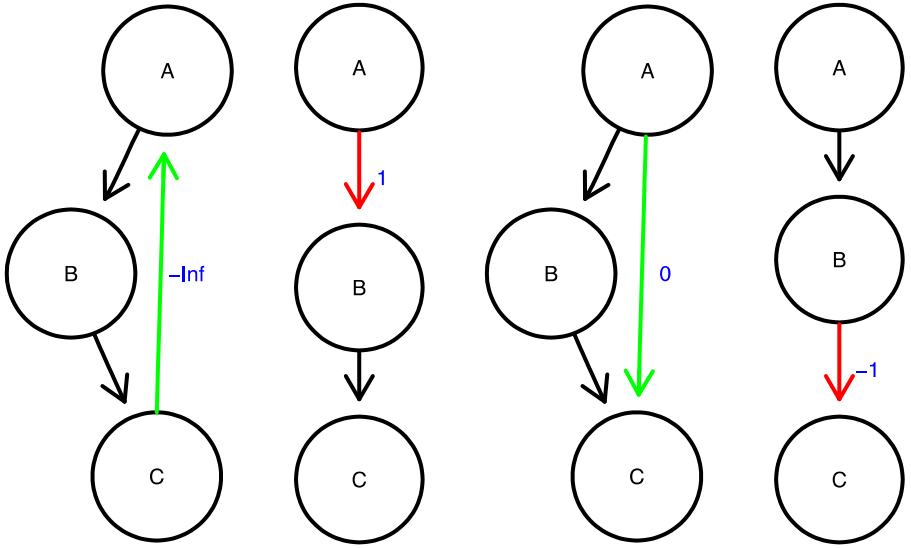
Greedy search is much faster than exhaustive search, but does not guarantee a global but only local optimum (restarts with different initial networks).

Black is the given graph.

Red is a deletion.

Green is an addition.

Blue are the scores.

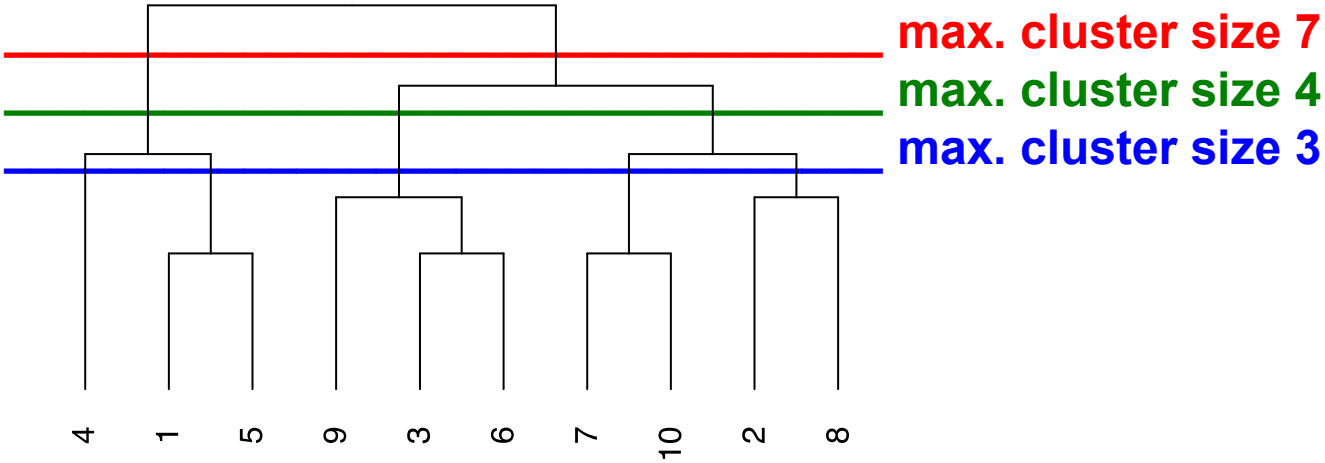


Module networks

Module networks is a more general divide and conquer approach.

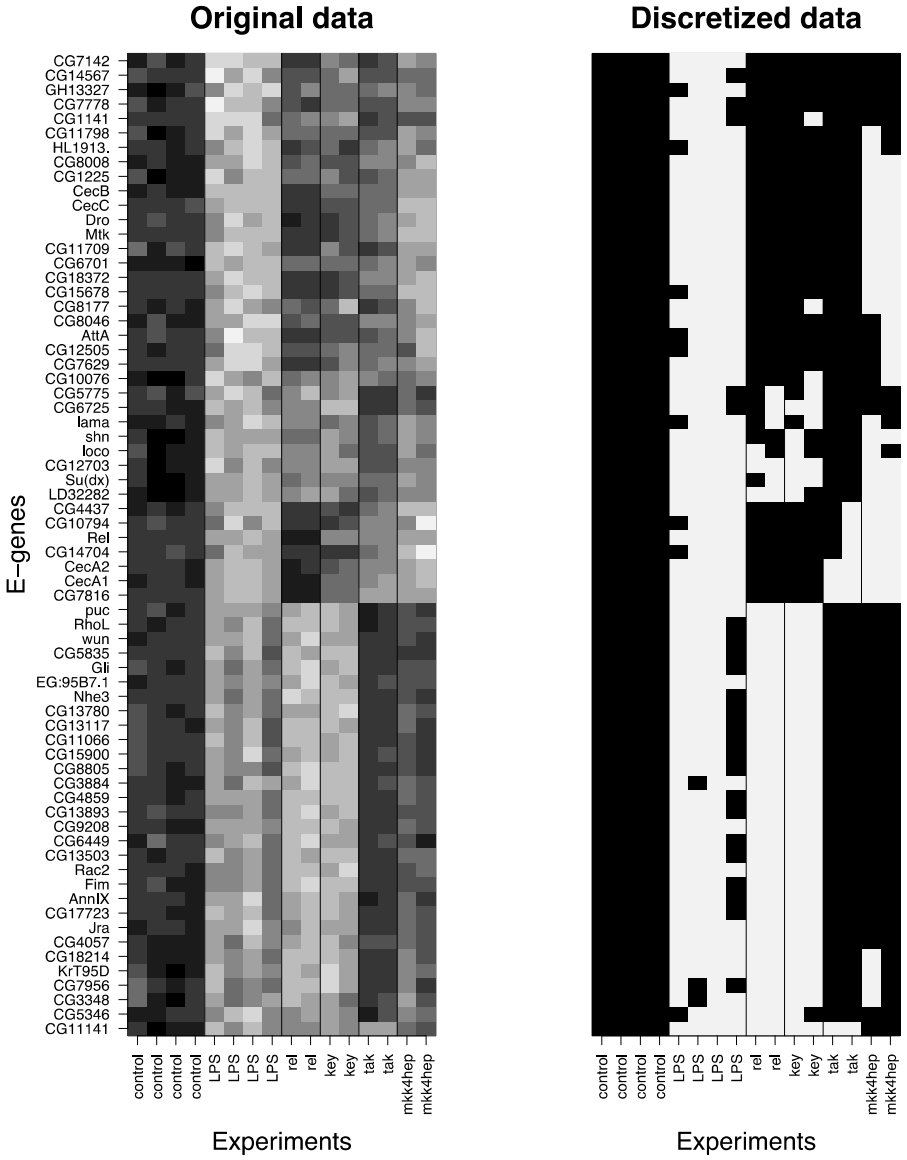
1. Hierarchically cluster all S-genes based on the effects/ratio matrix
2. Choose a cutoff such that the maximum number of S-genes per cluster is four
3. Exhaustively optimize the pathway for each cluster of S-genes
4. Link two pathways by searching for the optimal edge between two S-genes from different clusters and transitively close the joint pathway

The module approach can be further generalized by increasing the maximal cluster size to $n > 4$ and using the greedy instead of exhaustive search within each cluster.



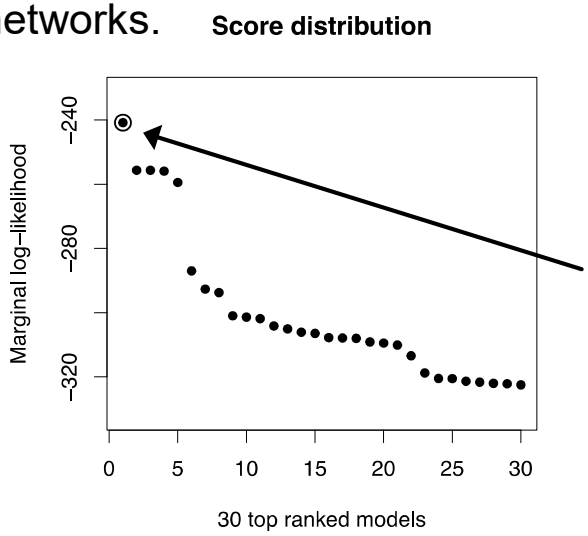
Application to innate immune response Drosophila

- Positive control (LPS stimulation)
- Negative control (no stimulation, no knock-down)
- Four different S-gene knock-downs (RNAi) during LPS stimulation: rel, key, tak, mkk4/hep
- Four replicates for each control and two for each knock-down
- Gene expression intensities from Affymetrix-microarrays

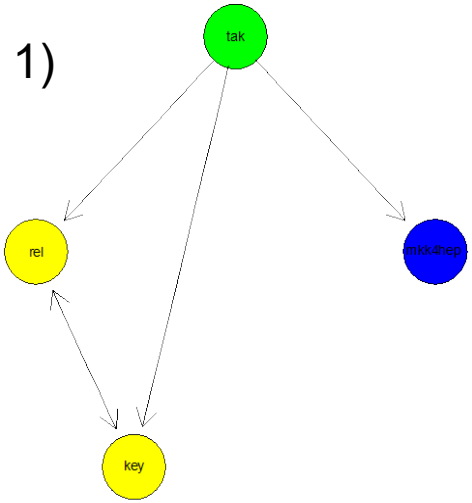


Application to innate immune response Drosophila

Exhaustive search on binarized data yields the following scores and five highest scoring networks.

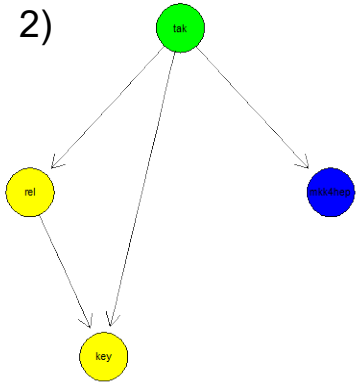


1)

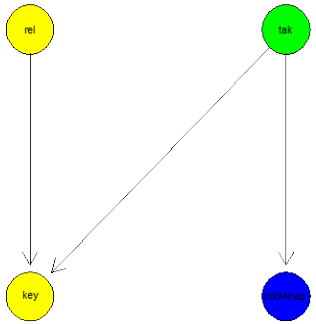


The highest scoring model agrees with literature findings. Rel and key are indistinguishable (both have similar effects). This is depicted by the bidirectional edge.

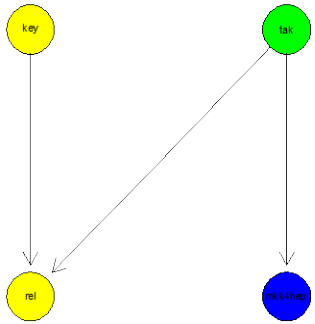
2)



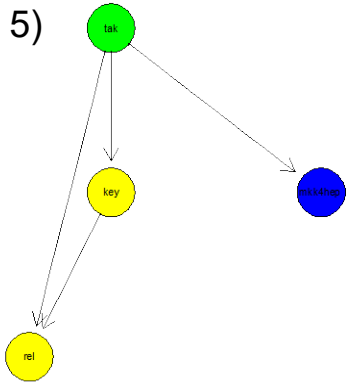
3)



4)



5)



Time series data

Time series of perturbation experiments

A time series of perturbation experiments consists of a data matrix for each of the *T* time points. We do not know if the difference between time points comes from false positives/negatives or if the network changed over time.

Time point 1

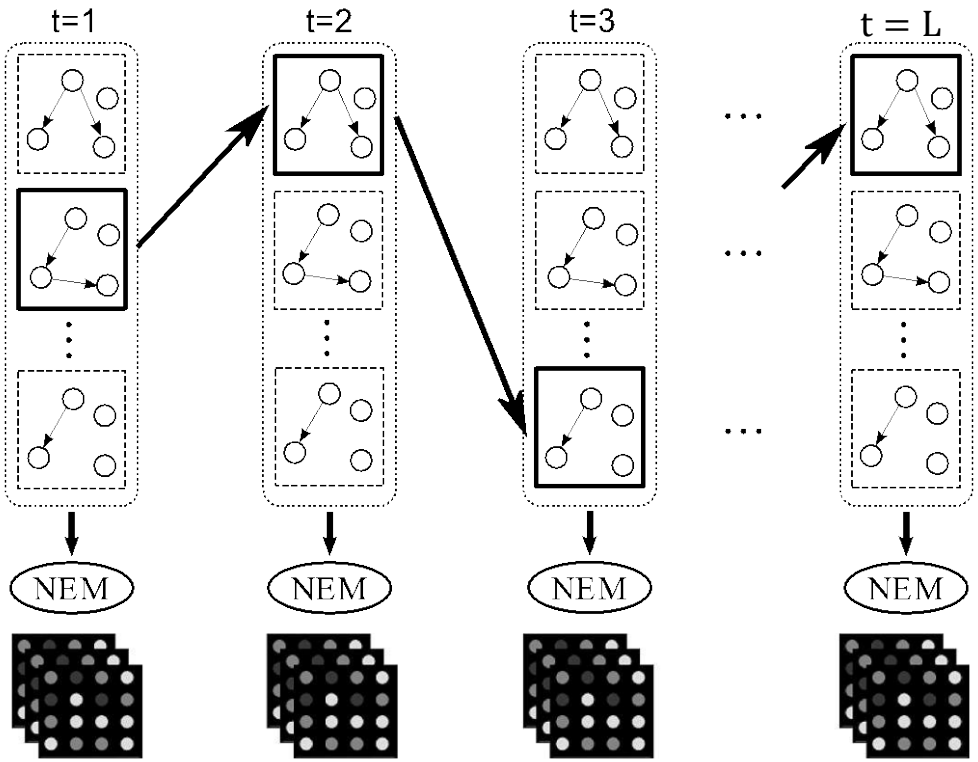
	A	B	C	D
E1	1	0	0	0
E2	1	0	1	0
E3	0	1	0	1
E4	0	1	0	0
E5	1	0	1	0
E6	1	1	1	0
E7	1	0	1	0
E8	1	1	1	0

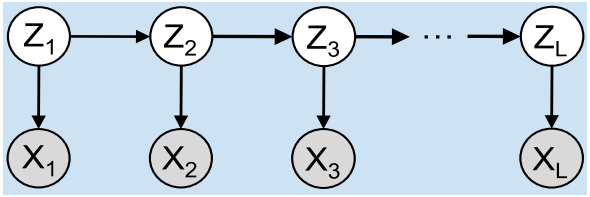
…

Time point *T*

	A	B	C	D
E1	1	1	0	0
E2	1	0	0	0
E3	1	1	0	0
E4	0	1	0	0
E5	1	1	1	1
E6	1	0	1	0
E7	1	1	1	1
E8	1	1	0	1

Hidden Markov Nested Effects Models (HM-NEM)





Compare to lecture on HMMs!

HM-NEMs model the evolution of a signaling pathway over time.

The state space in this HMM are the possible network structures with an initial distribution π (usually uniform) with initial probabilities for each network.

The emission probabilities E given the hidden state are defined by the corresponding NEM.

The transition probabilities T are defined between each network at time point t and $t + 1$.

Transition probabilities

The transition probabilities are based on network similarity. If a network is more similar to the current one, the transition probability is higher.
The probability to transition from network *u* to network *v* is calculated by

$$T_{uv} = P(\phi_{t+1} = v \mid \phi_t = u) = \frac{1}{C_u}(1-\lambda)^{s_{uv}}\lambda$$

with the network ϕ_t at time *t*, smoothness parameter *λ* in (0,1), the distance

$$s_{uv} = \|u - v\|_1 := \sum_i \sum_{i'} |u_{ii'} - v_{ii'}|$$

and the normalizing constant

$$C_u = \sum_w (1-\lambda)^{s_{uw}}\lambda\,.$$

Emission probabilities

The emission probabilities denote the probability to observe data $\widetilde{D}$ at any time point (homogeneous HMM), for a given network ϕ, and can be computed by the original NEM marginal likelihood formulation

$$E_{\phi\widetilde{D}} = P\left(\widetilde{D} \mid \phi\right) = \frac{1}{n^m}\prod_{j=1}^{m}\sum_{i=1}^{n}\prod_{g=1}^{h} P\left(\tilde{e}_{jg} \mid f_{S(g)j}, \theta_{ij} = 1\right).$$

Metropolis-Hastings within Gibbs sampler

Emission matrix E is known, but transition matrix T is dependent on λ, which is unknown. The Baum-Welch algorithm (lecture 3, slide 36) with complexity $O(LK^2)$ for one iteration is not feasible due to the large space of hidden states (networks ϕ), e.g., already more than 9×10^6 networks with seven nodes.

We make inference over the joint posterior of network structures $\phi_{1:L}$ and smoothness parameter λ

$$P\left(\phi_{1:L}, \lambda \mid \widetilde{D}_{1:L}\right) = \frac{P\left(\widetilde{D}_{1:L} \middle| \phi_{1:L}, \lambda\right) P(\phi_{1:L}, \lambda)}{P(\widetilde{D}_{1:L})}$$

and use a Gibbs sampler to draw samples iteratively from the two conditionals

$$P\left(\phi_{1:L} \mid \lambda, \widetilde{D}_{1:L}\right)$$
$$P(\lambda \mid \phi_{1:L}, \widetilde{D}_{1:L})$$

We need proposal distributions Q_{uv} for the network structure ϕ_t and $\tilde{Q}_{\kappa\tau}$ for λ.

Sampling ϕ, $P\left(\phi_{1:L} \mid \lambda, \widetilde{D}_{1:L}\right)$

We sample one hidden state at a time. Let $\phi_{-t} \coloneqq \{\phi_{t'} : t' \neq t\}$. We write the conditional probability for a single state as

$$P\left(\phi_t = u_t \mid \phi_{-t}, \widetilde{D}_{1:L}, \lambda\right)$$

$$\propto \begin{cases} P(\phi_{t+1} = u_{t+1} \mid \phi_t = u_t, \lambda) P\left(\widetilde{D}_t \mid \phi_t = u_t\right) & \text{if } t = 1 \\ P(\phi_t = u_t \mid \phi_{t-1} = u_{t-1}, \lambda) P(\phi_{t+1} = u_{t+1} \mid \phi_t = u_t, \lambda) P\left(\widetilde{D}_t \mid \phi_t = u_t\right) & \text{if } t = 2, \dots, L-1 \\ P(\phi_t = u_t \mid \phi_{t-1} = u_{t-1}, \lambda) P\left(\widetilde{D}_t \mid \phi_t = u_t\right) & \text{if } t = L \end{cases}$$

$$= \begin{cases} T_{u_t u_{t+1}} E_{u_t \widetilde{D}_t} & \text{if } t = 1 \\ T_{u_{t-1} u_t} T_{u_t u_{t+1}} E_{u_t \widetilde{D}_t} & \text{if } t = 2, \dots, L-1 \\ T_{u_{t-1} u_t} E_{u_t \widetilde{D}_t} & \text{if } t = L \end{cases}$$

Acceptance probability $A_{u_t u_t'}$

The acceptance ratio for a proposed graph u_t' graph given the current graph u_t is

$$a_{u_t u_t'} = \begin{cases} \dfrac{T_{u_t' u_{t+1}} E_{u_t' \widetilde{D}_t} Q_{u_t' u_t}}{T_{u_t u_{t+1}} E_{u_t \widetilde{D}_t} Q_{u_t u_t'}} & \text{if } t = 1 \\ \dfrac{T_{u_{t-1} u_t'} T_{u_t' u_{t+1}} E_{u_t' \widetilde{D}_t} Q_{u_t' u_t}}{T_{u_{t-1} u_t} T_{u\_t\, u_{t+1}} E_{u_t \widetilde{D}_t} Q_{u_t u_t'}} & \text{if } t = 2, \dots, L-1 \\ \dfrac{T_{u_{t-1} u_t'} E_{u_t' \widetilde{D}_t} Q_{u_t' u_t}}{T_{u_{t-1} u_t} E_{u_t \widetilde{D}_t} Q_{u_t u_t'}} & \text{if } t = L \end{cases}$$

The acceptance probability is

$$A_{u_t u_t'} = \min\left\{a_{u_t u_t'}, 1\right\}$$

Sampling λ, $P(\lambda \mid \phi_{1:L}, \tilde{D}_{1:L})$

The posterior probability of λ is given by the prior and the joint probability of hidden states and observed symbols

$$P\left(\lambda \mid \phi_{1:L}, \tilde{D}_{1:L}\right) \propto P\left(\phi_{1:L}, \tilde{D}_{1:L} \mid \lambda\right) P(\lambda) = \pi \prod_{t=2}^{L} P(\phi_t \mid \phi_{t-1}, \lambda) \prod_{t=1}^{n} P\left(\tilde{D}_t \mid \phi_t\right).$$

with a uniform prior.

The acceptance ratio for a proposed λ' given λ is

$$\tilde{a}_{\lambda\lambda'} = \frac{P\left(\lambda' \mid \phi_{1:T}, \tilde{D}_{1:L}\right) \tilde{Q}_{\lambda'\lambda}}{P\left(\lambda \mid \phi_{1:T}, \tilde{D}_{1:L}\right) \tilde{Q}_{\lambda\lambda'}}$$

and the probability

$$\tilde{A}_{\lambda\lambda'} = \min\{a_{\lambda\lambda'}, 1\}.$$

Metropolis-Hastings-within-Gibbs for HM-NEM (compare to lecture 6, slide 17)

- Start with random guess $\phi^{(0)}, \lambda^{(0)}$
- **For** n = 1, …, N

  Generate a new network $\phi^{(n)}$:

  For t = 1, …, L

  Propose $\phi_t^{(n)}$ by randomly flipping an edge in $\phi_t^{(n-1)}$

  Accept with probability $A_{\phi_t^{(n-1)}\phi_t^{(n)}}$, otherwise $\phi_t^{(n)} = \phi_t^{(n-1)}$

  end for

  Generate new smoothing parameter $\lambda^{(n)}$

  Accept with probability $A_{\lambda^{(n-1)}\lambda^{(n)}}$, otherwise $\lambda^{(n)} = \lambda^{(n-1)}$

end for

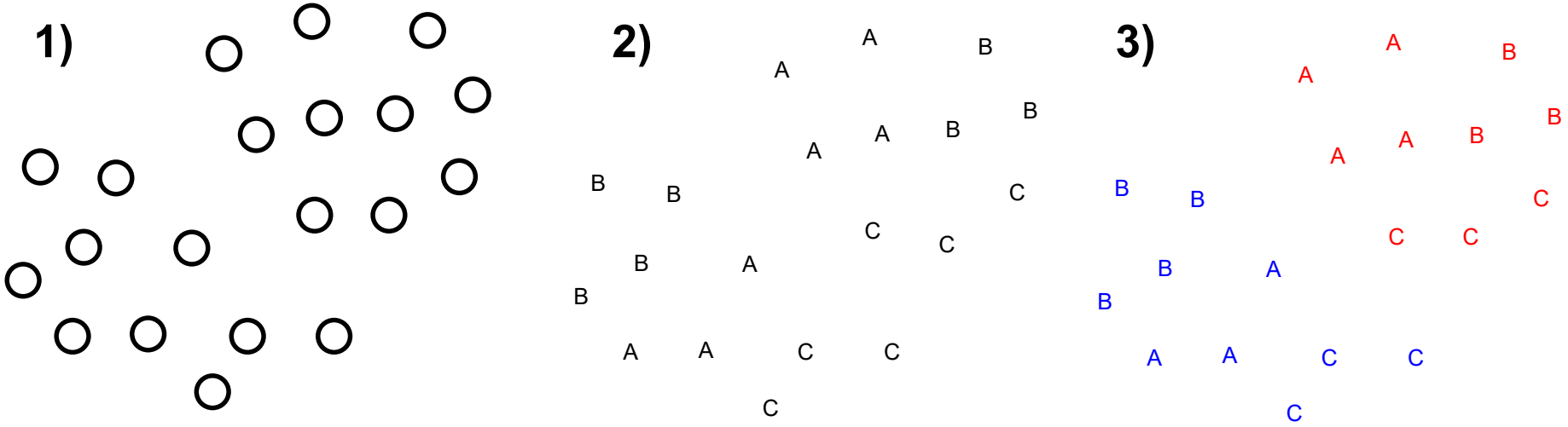
- Discard initial burn-in phase $\phi^{(0)}, \dots, \phi^{(N-R)}$
- Compute expectations

$$\widehat{\phi_t} = \frac{1}{R} \sum_{r=N-R+1}^{N} \phi_t^{(r)} \; \forall t = 1, \dots, L$$

Single-cell data

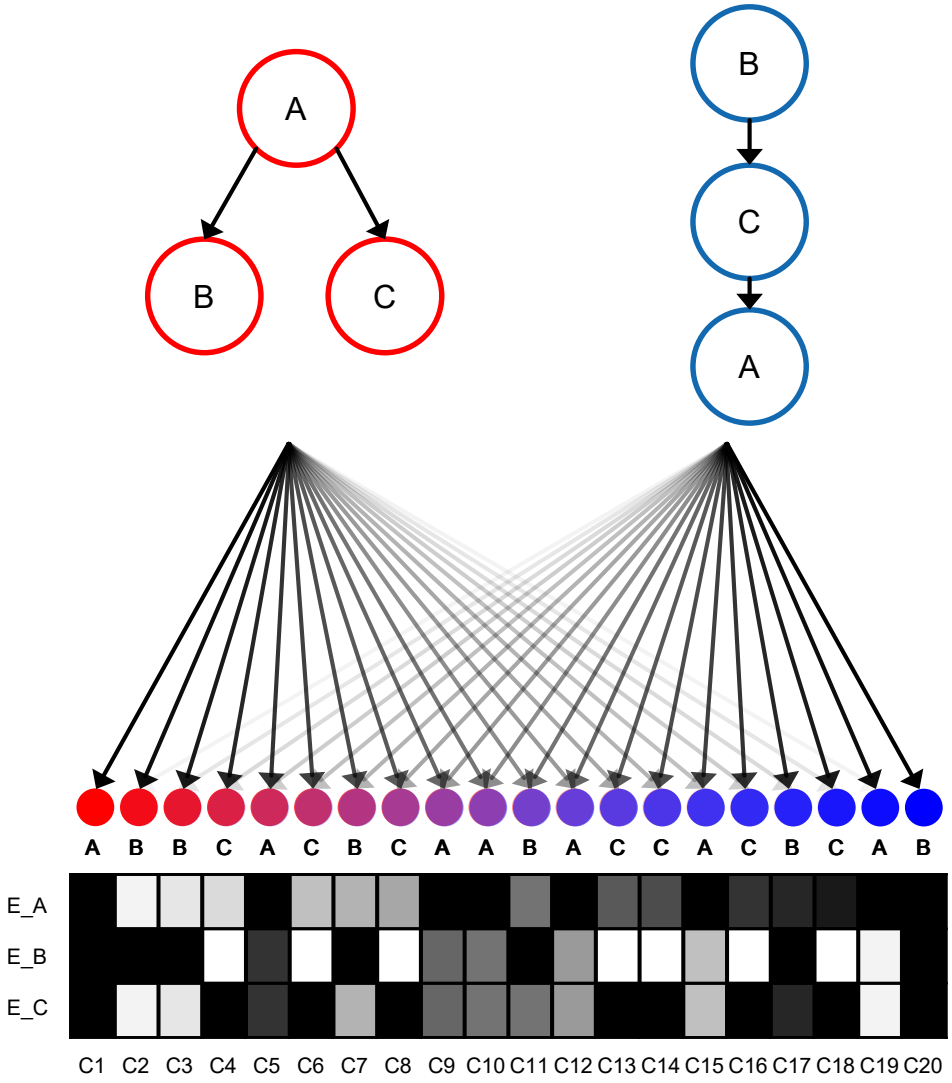
Mixture Nested Effects Models (M&NEM)

- Cell populations can be very heterogeneous.
- New technologies allow the measurement of gene expression data of single cells (scRNA-seq) from perturbation experiments (CRISPR).
- Higher resolution: sub-populations harbor different causal networks of the same genes.



- 2D projection of single-cell RNA-seq (1), annotated with knock-downs (known, 2) and clustered (unknown, 3).
- The underlying causal network and the cell clustering are unknown.
- We can cluster cells and infer a network for each cluster or we can simultaneously infer the clusters and networks. ➔ M&NEM

Schematic example



Two networks generate data with three E-genes and 20 cells.

The color gradient shows the probability for each cell to have been generated by either the blue or the red network.

Cell 1 was perturbed by a knock-down of *A*, all E-genes show an effect (black) and the cell fits perfectly (100%) to the red network.

Cell 11 was perturbed by a knock-down of *B*. The E-genes of *A* and *C* show a mild reaction to the knock-down and the cell fits almost equally well to either network.

Maximum likelihood

Parameters:
Network parameters $(\boldsymbol{\Phi,\Theta}) = (\Phi_k,\Theta_k)_{k=1,\ldots,K}$
Mixture weights $\pi = (\pi_1,\ldots,\pi_K)$ in $[0,1]^K$

The assignment of each cell c is modelled by the hidden variable z_{kc}, with $z_{kc} = 1$ if cell c was generated by NEM component k.

The mixture weights are defined by $\pi_k = P(z_{kc} = 1)$ with $\sum_{k=1}^{K} \pi_k = 1$.

Usually, several cells have been perturbed by a knock-down of S-gene i. We call $\rho = (\rho_{ic})$ the cellular perturbation map. We have $\rho_{ic} = 1$ if cell c has been perturbed by a knock-down of S-gene i.

We calculate the log likelihood ratio of a model to the Null model for component k and all cells by

$$L_k = (l_{k,cd}) = \rho^T \Phi_k \Theta_k R.$$

We have stored the log likelihood ratios for all cells in the diagonal of L_k.

Log-likelihood of the mixture model

We compute the log-likelihood ratio to the Null model *N* of a mixture by

$$\log P(D \mid \boldsymbol{\phi}, \boldsymbol{\theta}) - \log P(D \mid N) = \operatorname{trace}\left(\log \sum_{k=1}^{K} \pi_k \exp(L_k)\right).$$

We optimize this likelihood with an implementation of the Expectation Maximization algorithm.

E-step

Given a parametrization π, $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}$ we calculate L_k (the log likelihood ratio) and the responsibilities

$$\gamma(z_{kc}) = P(z_{kc} = 1 \mid d_{.c}) = \frac{\pi_k \exp(l_{k,cc})}{\sum_{s=1}^{K} \pi_s \exp(l_{s,cc})}$$

which we store in the $K \times C$ matrix $\Gamma = (\gamma_{kc})$ with entries in [0, 1].

M-step

Given the responsibilities we update the mixture weights

$$\pi_k = \frac{\sum_{c=1}^{C} \gamma_{kc}}{\sum_{s=1}^{K} \sum_{c=1}^{C} \gamma_{sc}}$$

and we optimize the S-gene network Φ_k and the E-gene attachments Θ_k with respect to

$$R_k = \left(r_{jc}\gamma_{kc}\right)$$

The optimization of Φ_k and Θ_k is done with any of the previous inference methods for standard NEMs.

Other extensions

- Dynamic networks to infer signaling rates and feed forward loops (Anchang et al., PNAS 2009)
- Dynamic networks including cycles (Fröhlich et al., Bioinformatics 2011)
- Probabilistic pathway activation in single cells (Siebourg-Polster et al., PLOS Computational Biology 2015)
- Boolean formulation allowing for combinatorial interventions with knock-ins and the inclusion of literature knowledge (Pirkl et al., Bioinformatics 2016)
- Analysis of large double knock-out screens (Pirkl et al., PLOS Computational Biology 2017)
- Mixture model on E-genes (Sverchkov et al., RECOMB 2018)
- Accounting for siRNA off-target effects (Srivatsa et al., ISMB 2018, Bioinformatics 2018)

Summary

- Nested Effects Models (NEMs) are designed to learn dependencies among genes from interventions (RNAi, CRISPR).
- The idea is to consider the nested structure of intervention effects.
- The likelihood involves (few) hidden signaling genes and (many) observed effect genes.
- Every effect is linked to exactly one signaling gene, but the connectivity is unknown and usually integrated out or estimated.
- NEMs are identifiable up to their transitive closure.
- Hidden Markov NEMs model temporal evolution of causal networks from time series data.
- Mixture NEMs simultaneously infer sub-populations and networks from single-cell perturbation data.

Notation

- S-genes: $i, i' = 1, \ldots, n$
- E-genes: $j = 1, \ldots, m$
- Experiments: $g = 1, \ldots, h$ with $S(g) = i$ the perturbed S-gene i, e.g., $S(1) = S(2) = 4$ means that in experiments 1 and 2 S-gene 4 has been perturbed, i.e., they are replicates.
- Data $D = (e_{ji})$: simple case with one experiment per S-gene
- Data $D = (\tilde{e}_{jg})$: general case with more experiments per S-gene
- Continuous data $R = (r_{ji})$
- Timepoints $t = 1, \ldots, L$
- Mixture components: $k = 1, \ldots, K$
- Cells $c = 1, \ldots, C$
- Perturbation matrix $\rho = (\rho_{ic})$

References

- Wiestner, A. (2015). The role of B-cell receptor inhibitors in the treatment of patients with chronic lymphocytic leukemia. Haematologica Dec 2015, 100 (12) 1495-1507; DOI: 10.3324/haematol.2014.119123.
- Wittrup, A., Lieberman, J. (2015). Knocking down disease: a progress report on siRNA therapeutics. Nature Reviews Genetics volume 16, pages 543–552 (2015).
- Ui-Tei, K., Nishi, K., Takahashi, T., Nagasawa, T. (2012). Thermodynamic control of small RNA-mediated gene silencing. Frontiers in Genetics, Vol. 3, 101.
- Redman, M., King, A., Watson, C., King, D. (2016). What is CRISPR/Cas9?. Archives of Disease in Childhood - Education and Practice Published Online First: 08 April 2016.
- Markowetz, F., Bloch, J., and Spang, R. (2005). Non-transcriptional pathway features reconstructed from secondary effects of rna interference. *Bioinformatics*, 21(21), 4026–4032.
- Markowetz, F., Kostka, D., Troyanskaya, O. G., and Spang, R. (2007). Nested effects models for high-dimensional phenotyping screens. *Bioinformatics*, 23(13), i305–i312.
- Froehlich, H., Fellmann, M., Sueltmann, H., Poustka, A., Beissbarth, T. (2007). Large scale statistical inference of signaling pathways from RNAi and microarray data. BMC BioinformaticB0078:386.
- NEM Biconductor R-package '*nem*'. (2018). http://www.bioconductor.org.
  - Extensions on Bioconductor: '*mnem*', '*epiNEM*' , '*bnem*', '*nempi*'.
- Wang, X., Yuan, K., Hellmayr, C., Liu, W., and Markowetz, F. (2014). Reconstructing evolving signalling networks by hidden markov nested effects models. *Ann. Appl. Stat.*, 8(1), 448–480.
- Pirkl, M., Beerenwinkel, N. (2018). Single cell network analysis with a mixture of Nested Effects Models. ECCB 2018, Bioinformatics 2018.