Ridge estimation of the VAR(1) model and its time series chain graph from multivariate time course omics data

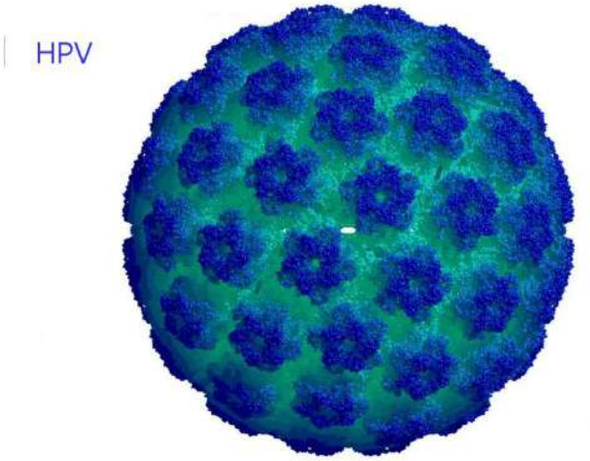
Viktorian Miok, Saskia Wilting and Wessel van Wieringen

Epidemiology & Biostatistics
VUmc

SMPGD
12-02-2016, Lille
France

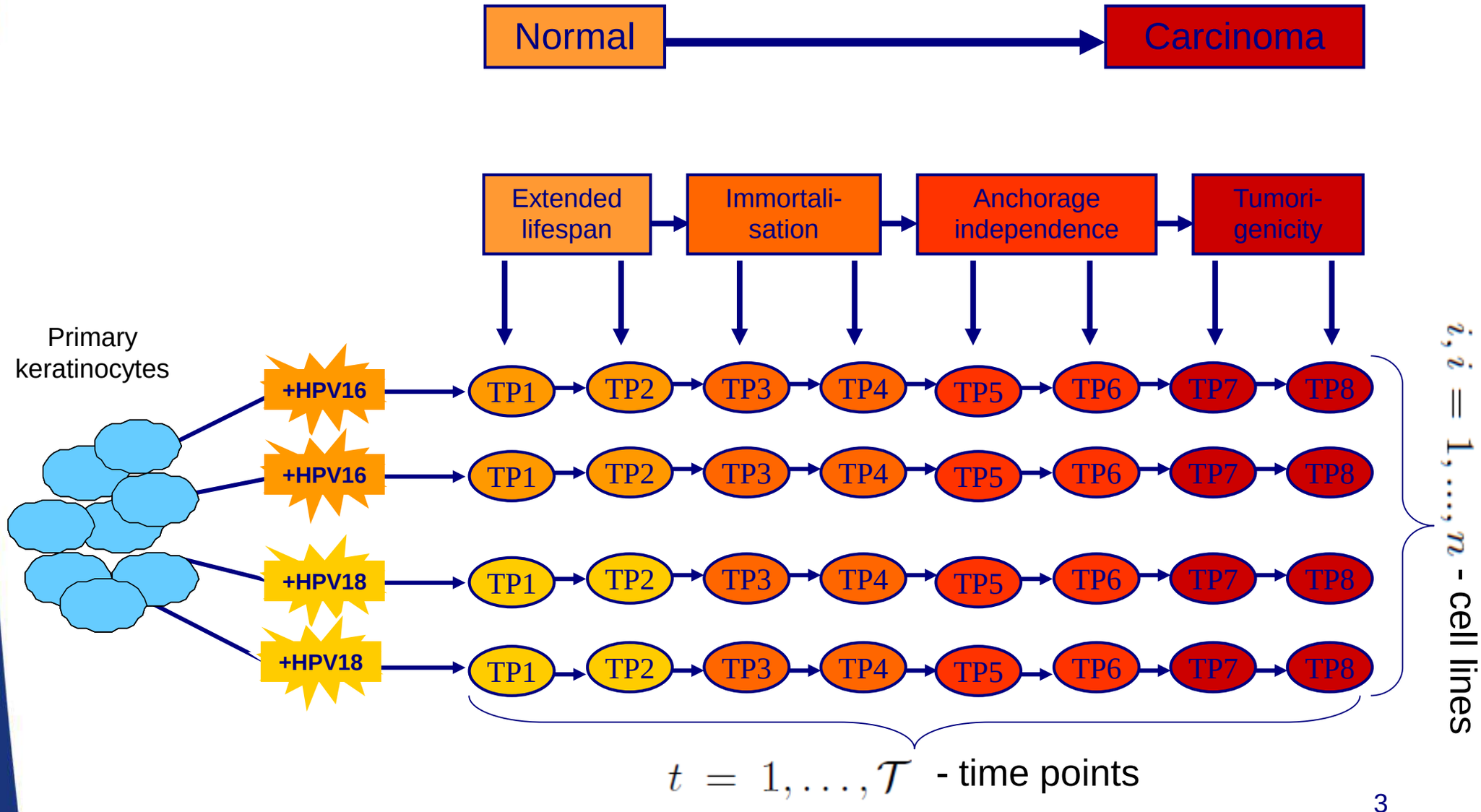
VU VRIJE UNIVERSITEIT AMSTERDAM

Cervical cancer study



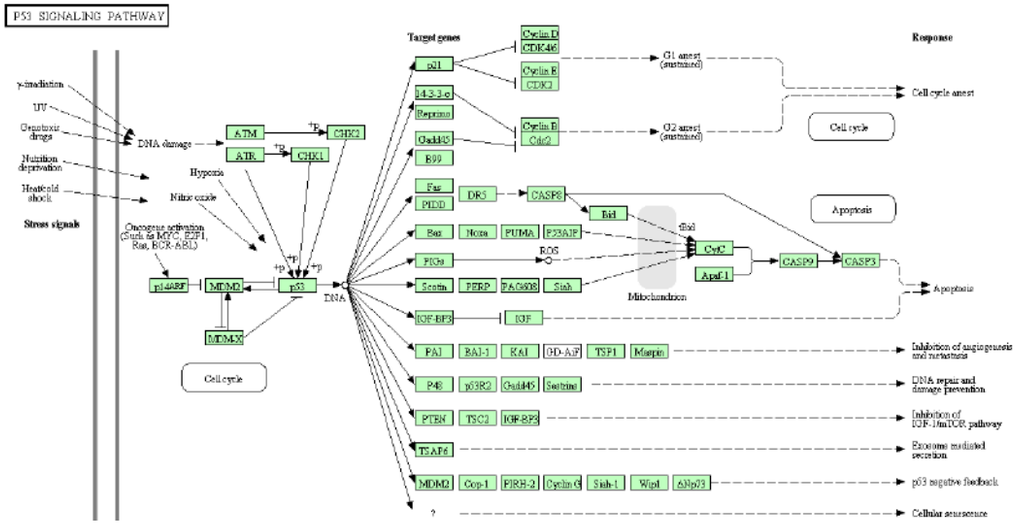
- Second most common cancer in women worldwide.
- Caused by HPV virus, in 70% cases HPV16 and HPV18.
- Cell line model – in vitro model system of HPV-induced transformation.
- Aim: reconstruct interaction among the genes.

Experiment


Normal
Carcinoma
Extended lifespan
Immortali-sation
Anchorage independence
Tumori-genicity
Primary keratinocytes
+HPV16
+HPV16
+HPV18
+HPV18
TP1
TP2
TP3
TP4
TP5
TP6
TP7
TP8
$i, i = 1, \ldots, n$ - cell lines
$t = 1, \ldots, \mathcal{T}$ - time points

Pathways

- Pathway are defined using repositories: **KEGG, GO, Reactome**…



- Problems with repositories:
  - Incomplete
  - Mostly well-known pathways
  - Loosely defined

- Reconstruction of the p 53 signaling pathway in cervical cancer

Model

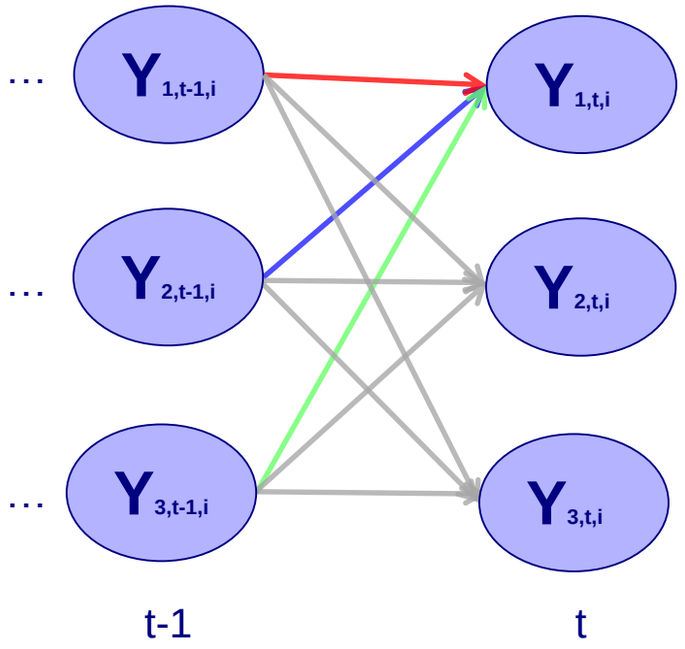
$\mathbf{Y}_{*,t,i}$ - mRNA gene expression

VAR(1) process: $\mathbf{Y}_{*,t,i} = \mathbf{A}\mathbf{Y}_{*,t-1,i} + \varepsilon_{*,t,i}$

First gene: $\mathbf{Y}_{1,t,i} = a_{1,1}\mathbf{Y}_{1,t-1,i} + a_{1,2}\mathbf{Y}_{2,t-1,i} + a_{1,3}\mathbf{Y}_{3,t-1,i} + \varepsilon_{1,t,i}$

node, a gene

edge, two gene interaction



Estimators

Ridge penalized log-likelihood: $\log[L(\mathbf{Y}; \mathbf{A}, \boldsymbol{\Omega}_\varepsilon)] - \lambda_\omega \|\boldsymbol{\Omega}_\varepsilon - \boldsymbol{\Omega}_0\|_2^2 - \lambda_a \|\mathbf{A} - \mathbf{A}_0\|_2^2$

VAR(1) log-likelihood function

Ridge penalties

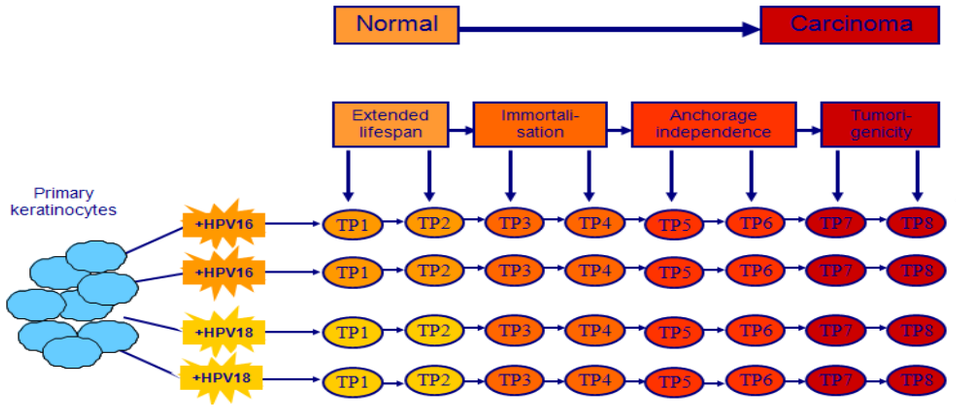
$\mathbf{A}_0, \boldsymbol{\Omega}_0$ - target matrices

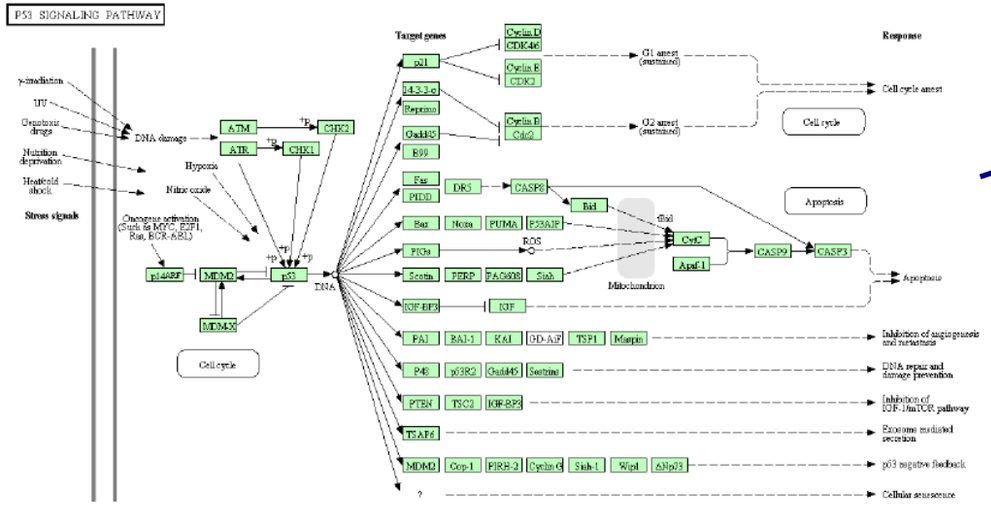
Maximization derive the estimators, in iterative procedure:

$$\text{vec}[\hat{\mathbf{A}}(\lambda)] = [\lambda_a \mathbf{I}_{p^2 \times p^2} + \hat{\boldsymbol{\Gamma}}(0) \otimes \boldsymbol{\Omega}_\varepsilon]^{-1} \{\lambda_a \text{vec}(\mathbf{A}_0) + \text{vec}[\boldsymbol{\Omega}_\varepsilon \hat{\boldsymbol{\Gamma}}(-1)]\}$$

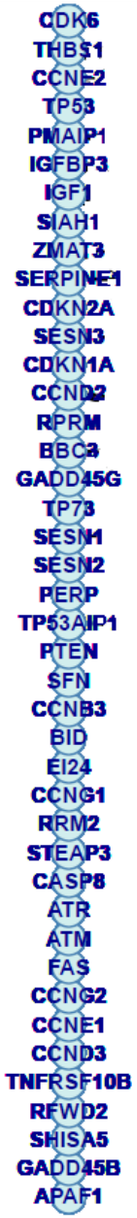
$$\hat{\boldsymbol{\Omega}}_\varepsilon(\lambda_\omega) = \left\{ \left[\lambda_\omega \mathbf{I}_{p \times p} + \frac{1}{4}(\mathbf{S}_\varepsilon - \lambda_\omega \boldsymbol{\Omega}_0)^2\right]^{1/2} + \frac{1}{2}(\mathbf{S}_\varepsilon - \lambda_\omega \boldsymbol{\Omega}_0) \right\}^{-1}$$

Data, model and network





Ridge estimation of VAR(1) model



t

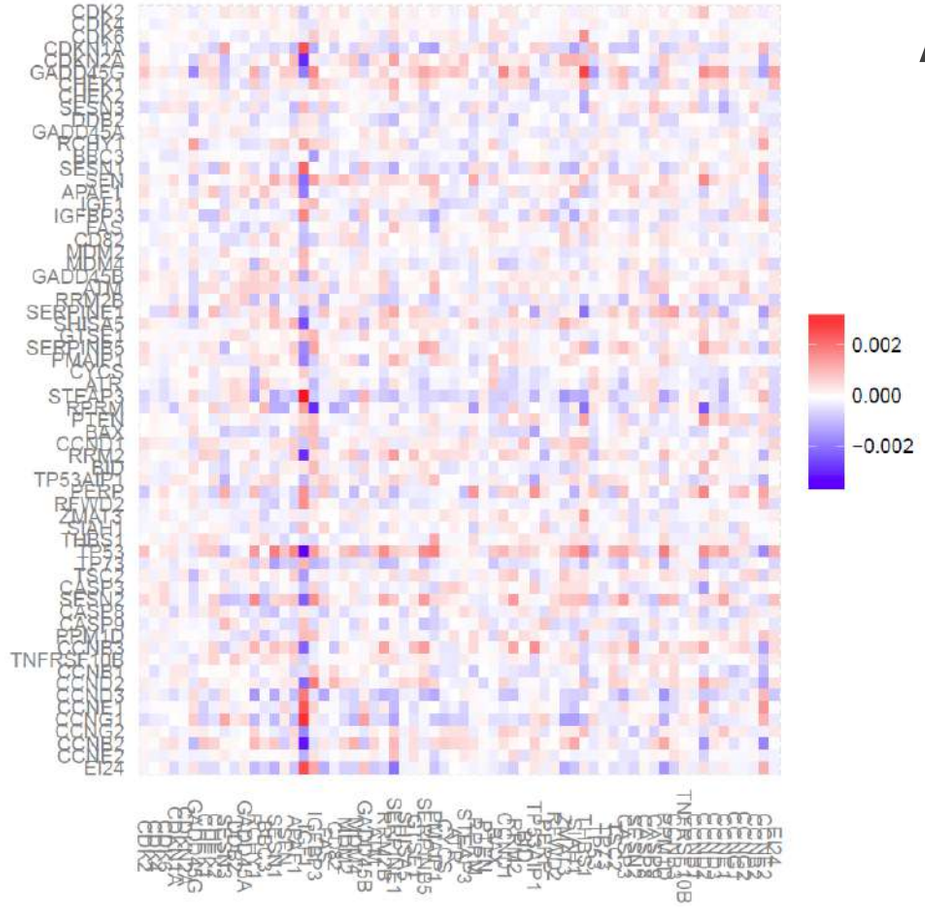
???

t+1

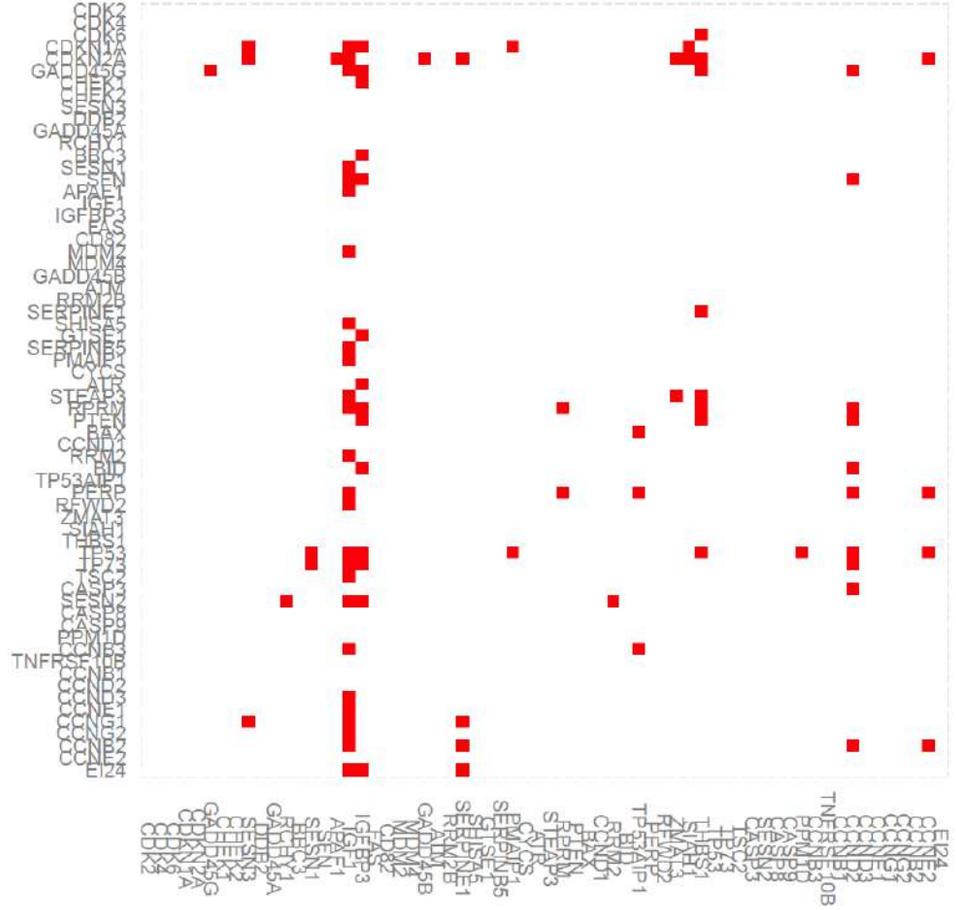
Sparse model parameters

- Penalty parameters estimated using leave-one-out cross-validated log likelihood
- Using local FDR non zero elements of interest are obtained

A:



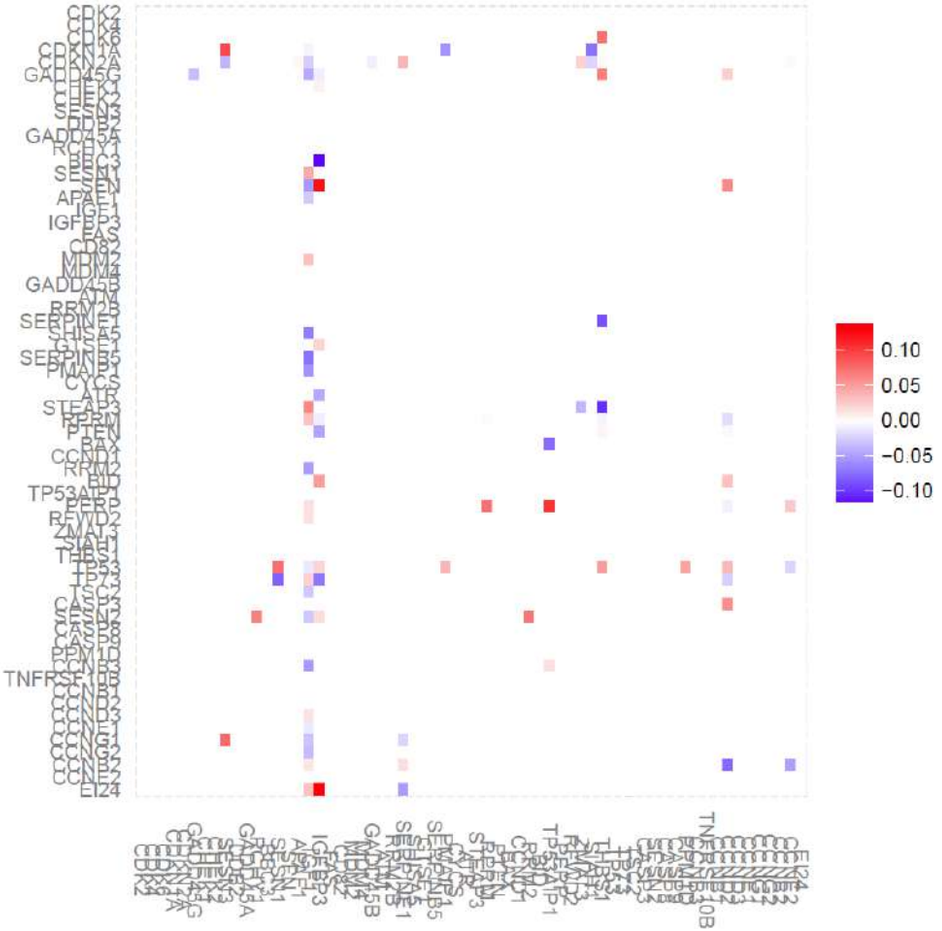
A:



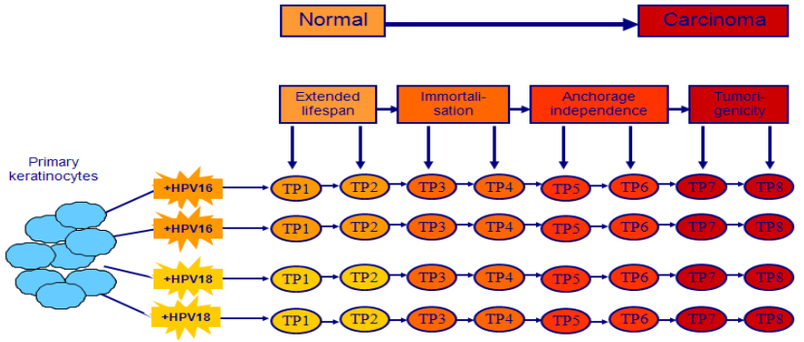
Prior knowledge

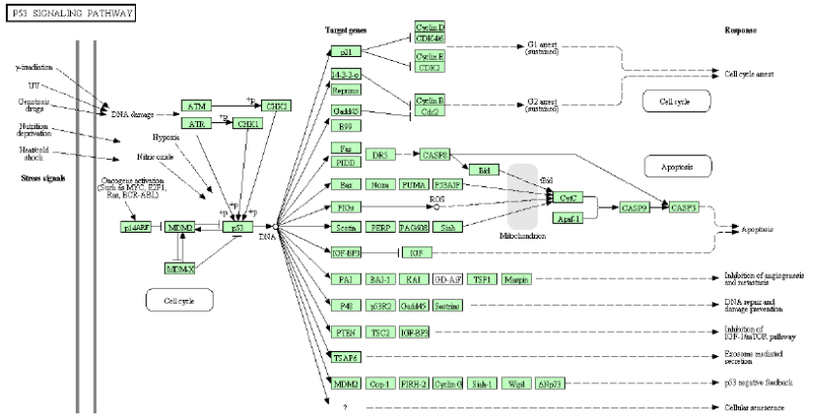
- Re-estimated parameters of **A**, taking the support of **A** into account
- Support of **A** formulated as linear constraints on **A**
- Less biased estimates of **A**

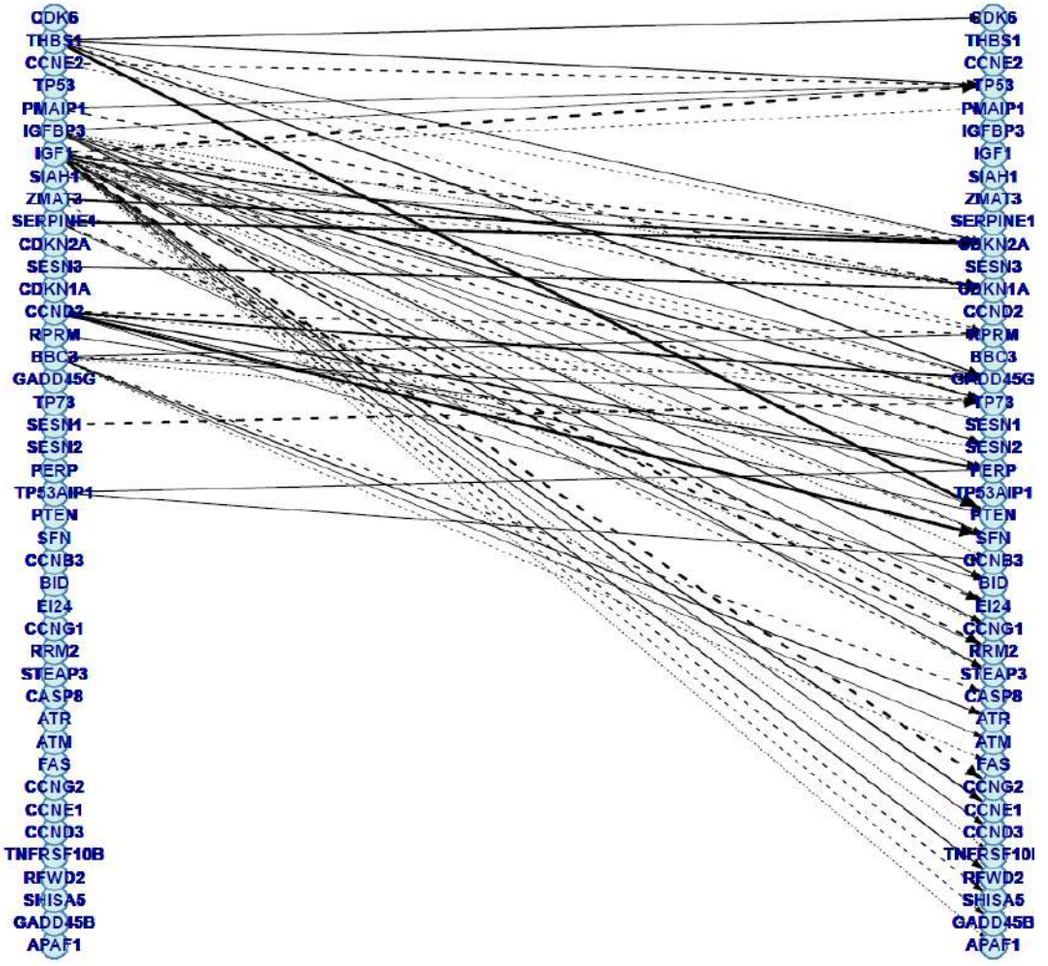
A:



Network







Ridge estimation of VAR(1) model

t

t+1

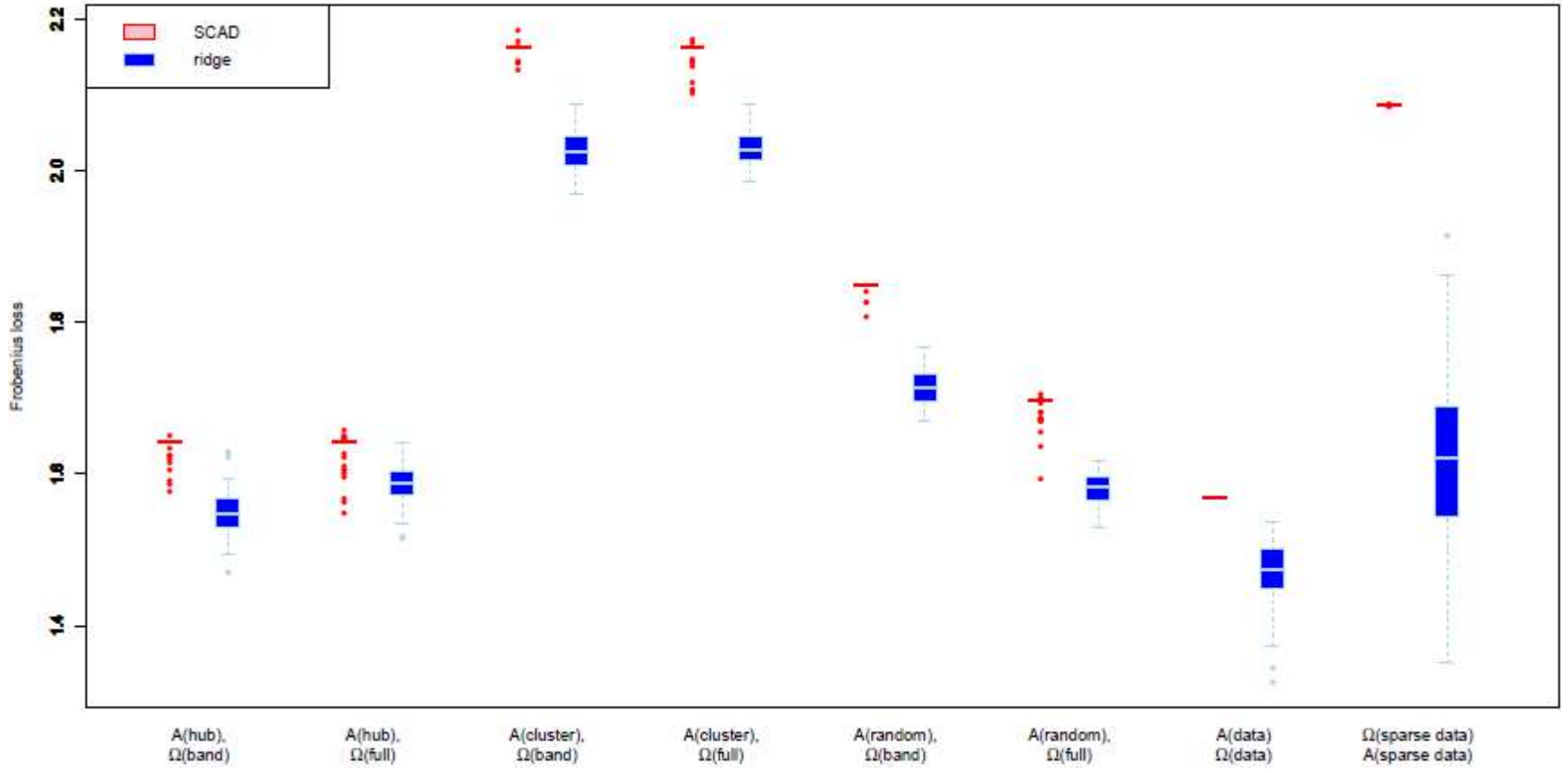
Comparison - ridge vs. SCAD

Data

- Simulated data (p=25, T=10, n=5)

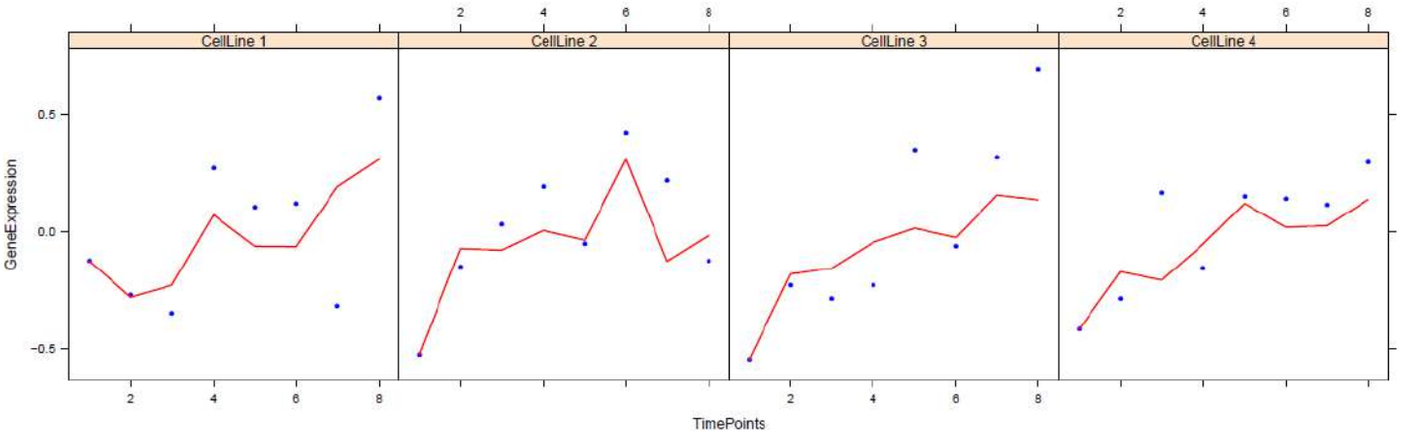
Frobenius loss of the estimates

- Optimal penalty parameters
- Reconstruction of time series chain graph

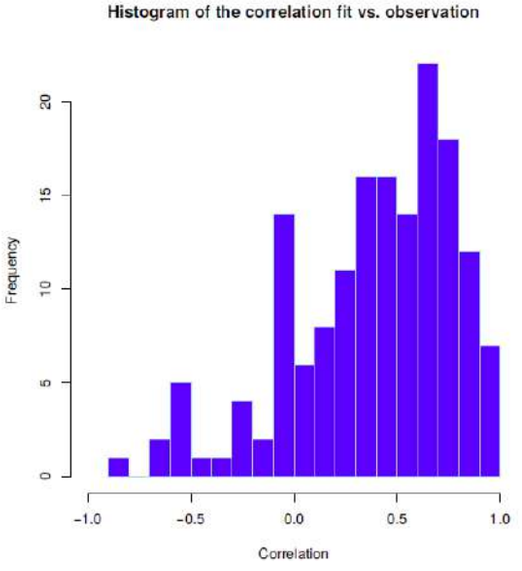


Model fit

- Fit of the TP73 gene:



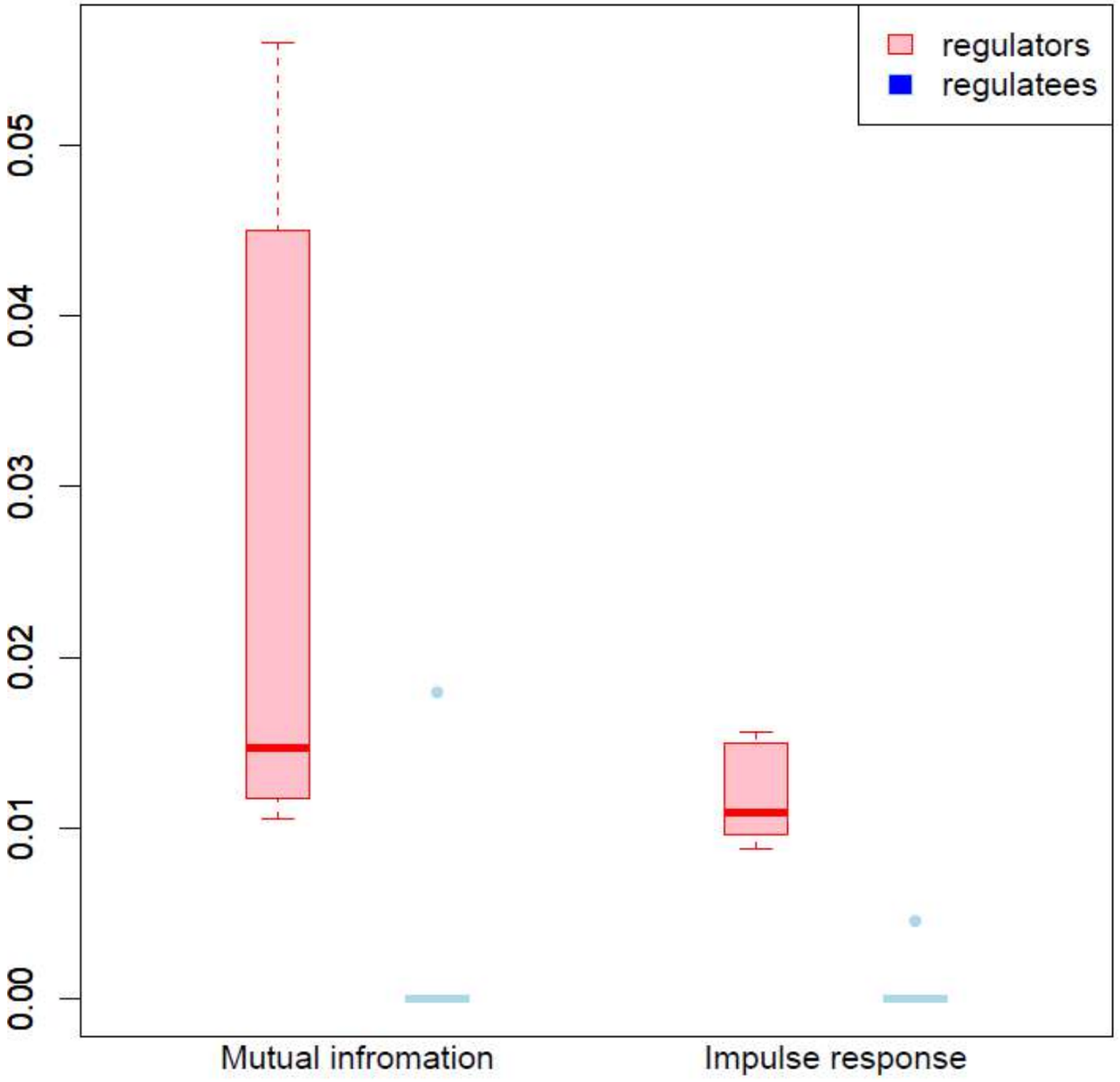
- Histogram of the fit vs. observation:



Node statistics

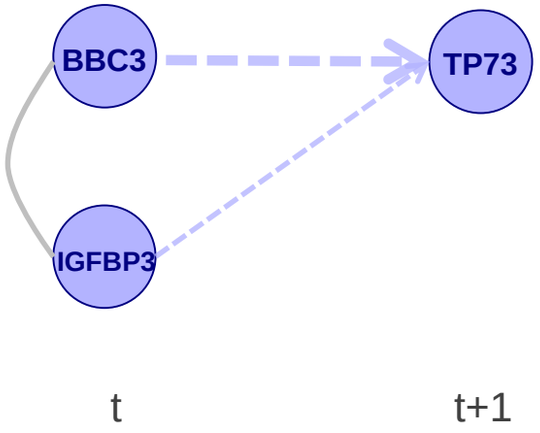
- Node statistics of the ‘regulators’ and ‘regulatees’

	$\deg^-(\mathbf{A})$	$\deg^+(\mathbf{A})$
BBC3	0	17
CCND2	0	12
IGF1	1	14
IGFBP3	0	16
THBS1	0	11
CCNG1	6	0
CDKN2A	12	0
SERPINE1	8	4
SESN2	8	0
STEAP3	9	0

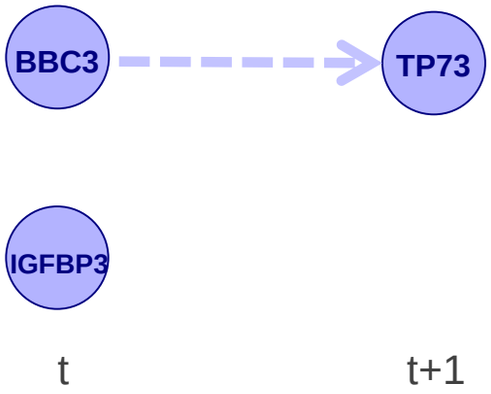


Path decomposition

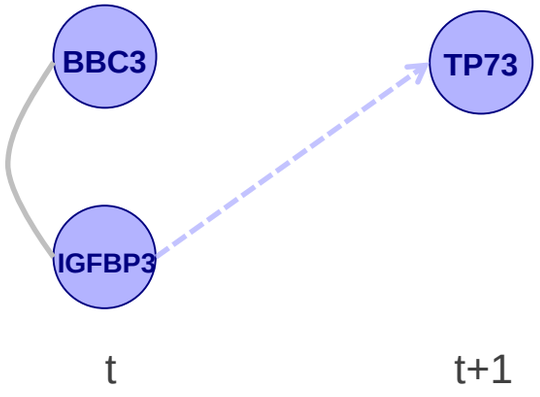
Contribution between BBC3 and TP73: -0.003168001



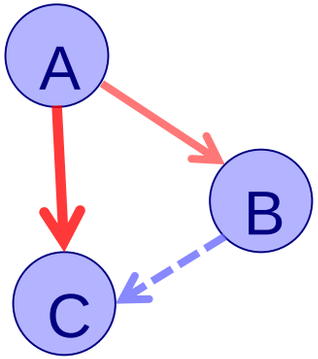
Contribution path 1: -0.002483485



Contribution paths 2: -0.0006845158

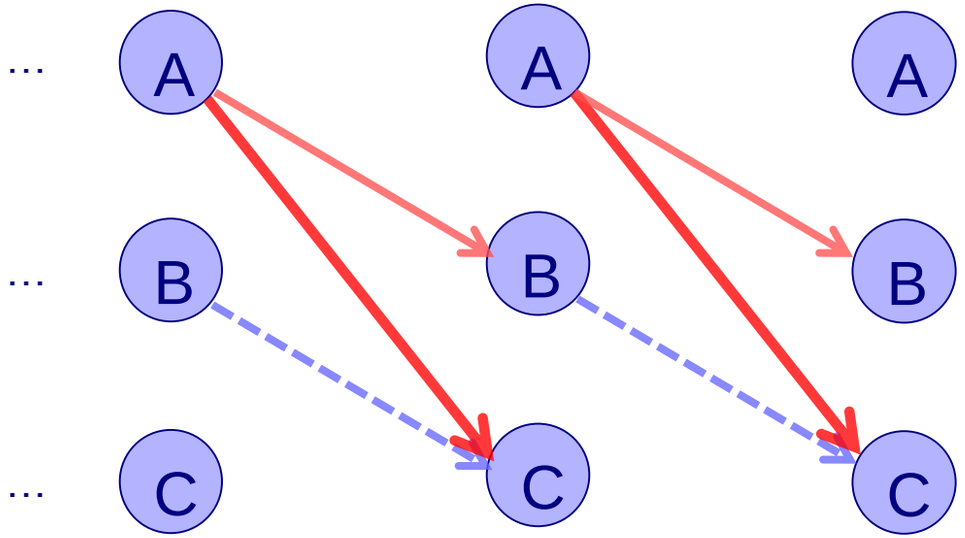


Feed-forward loop



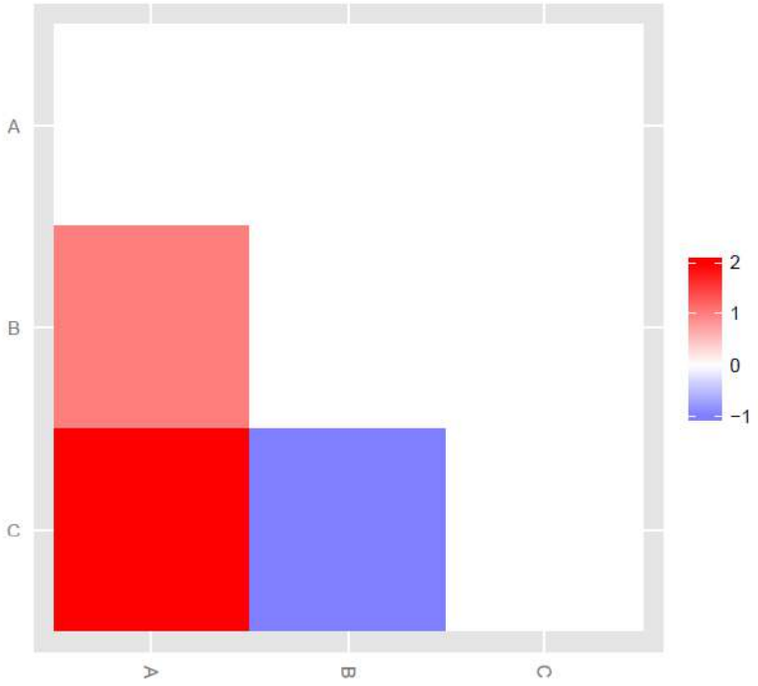
Feed-forward loop

$$
\begin{aligned}
A_t &= &&&&+\varepsilon_{A,t} \\
B_t &= &A_{t-1}&&&+\varepsilon_{B,t} \\
C_t &= &2A_{t-1} &- B_{t-1} &&+\varepsilon_{C,t}
\end{aligned}
$$



Feed-forward loop (unrolled)

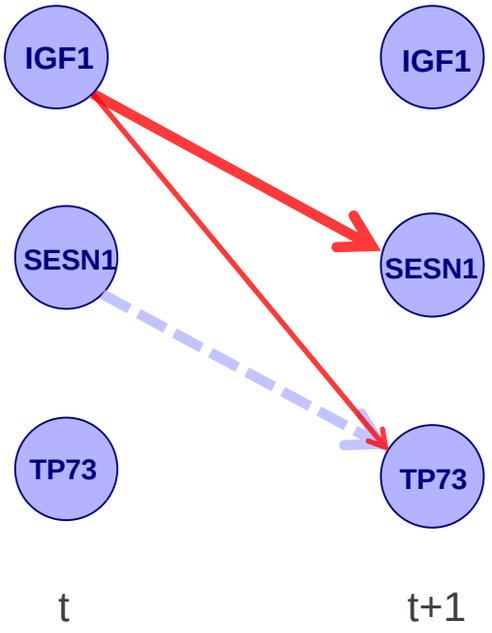
to

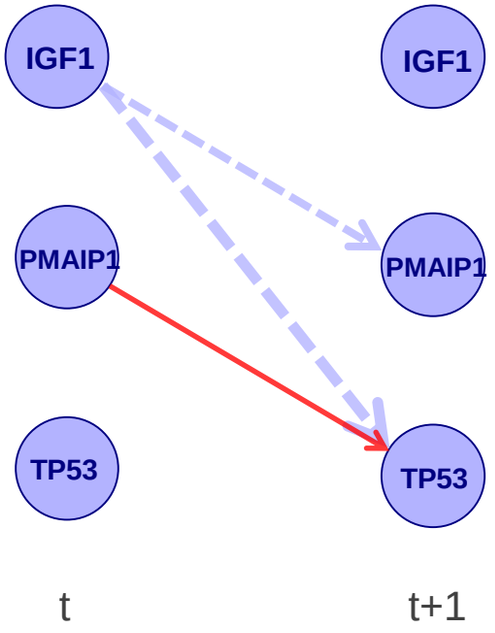


from

Motifs

- Building block of the pathways
- Carry out specific dynamic function
- Deregulation lead to cancer formation





Summary

- Reconstruction of the temporal interaction between genes
- Comparison in simulation study with SCAD, lasso based competitor
- Down-stream analysis presented on p53 signaling pathway
- Post estimation analysis: node statistics, impulse response, mutual information, path decomposition and motifs

Thank you for your attention!