Network based data analysis and literature exploration

Nicos Angelopoulos

Bioinformatics (Wessels) Group

overview

a tiny (Bayesian) network from literature

gene target identification

visualising PubMed citations

Cell

Network Crosstalk Dynamically Changes during Neutrophil Polarization

Chin-Jen Ku,[1,3] Yanqin Wang,[1,3] Orion D. Weiner,[2] Steven J. Altschuler,[1,*] and Lani F. Wu[1,*]
[1]Department of Pharmacology, Green Center for Systems Biology, Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[2]Cardiovascular Research Institute and Department of Biochemistry, University of California, San Francisco, San Francisco, CA 94158, USA
[3]These authors contributed equally to this work
*Correspondence: steven.altschuler@utsouthwestern.edu (S.J.A.), lani.wu@utsouthwestern.edu (L.F.W.)
DOI 10.1016/j.cell.2012.03.044

Cell *149*, 1073–1083, May 25, 2012

importance

concentrates on basic components of a system

illustrates what Bayesian networks CAN represent

causal links/networks can be inferred from interventions

likelihood functions are an unbiased way to do this

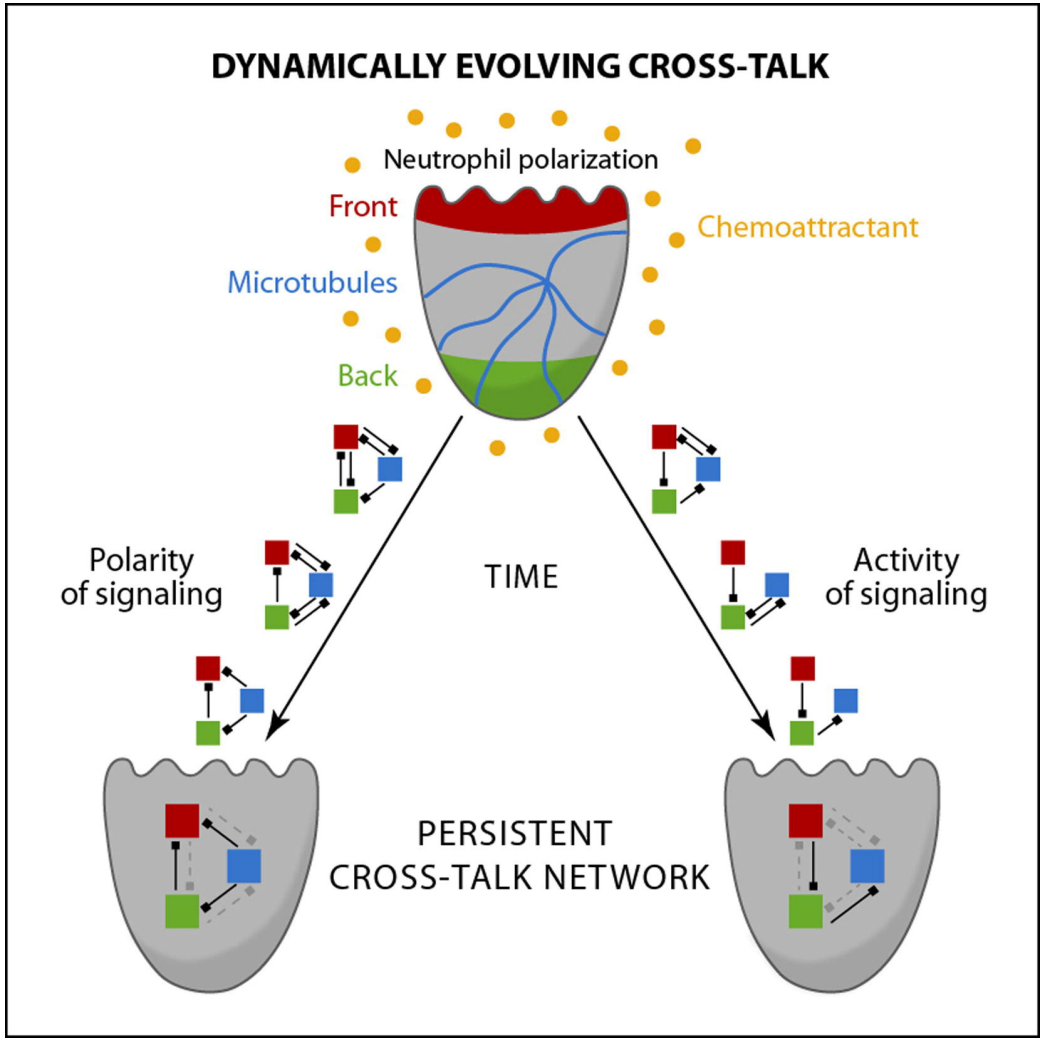
neutrophils

- produced in the bone marrow and circulate in the blood
- are an abundant type of white blood cells
- respond to infection and attack bacteria and other foreign invaders directly
- in the presence of f-Met-Leu-Phe (fMLP), neutrophils polarise within 1-2 minutes

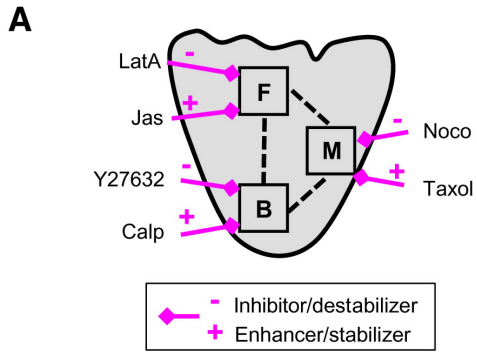
highlights

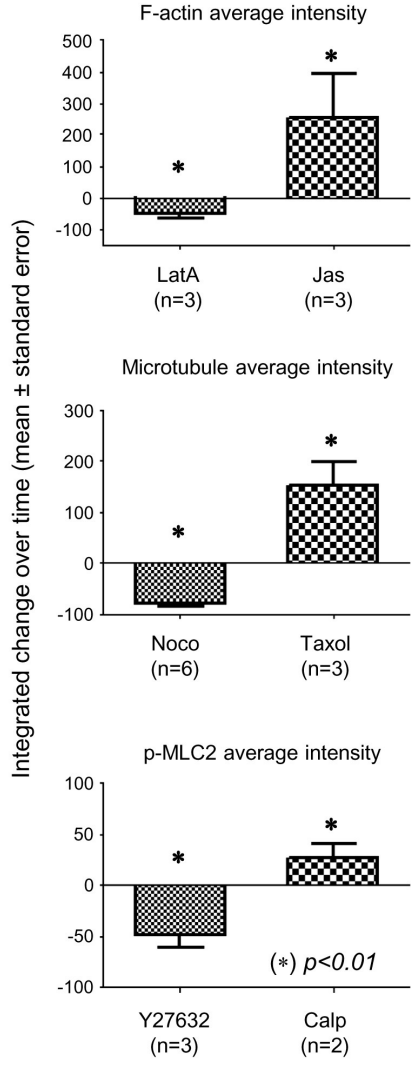
- 3 components system; measure intensity and polarity
- perturbations to signaling modules in stimulated neutrophils reveal crosstalk network
- network crosstalk dynamically changes during the neutrophil polarization process
- signalling activity and polarity are shaped by distinct patterns of crosstalk
- persistent crosstalk networks flow in opposite directions for intensity and polarity

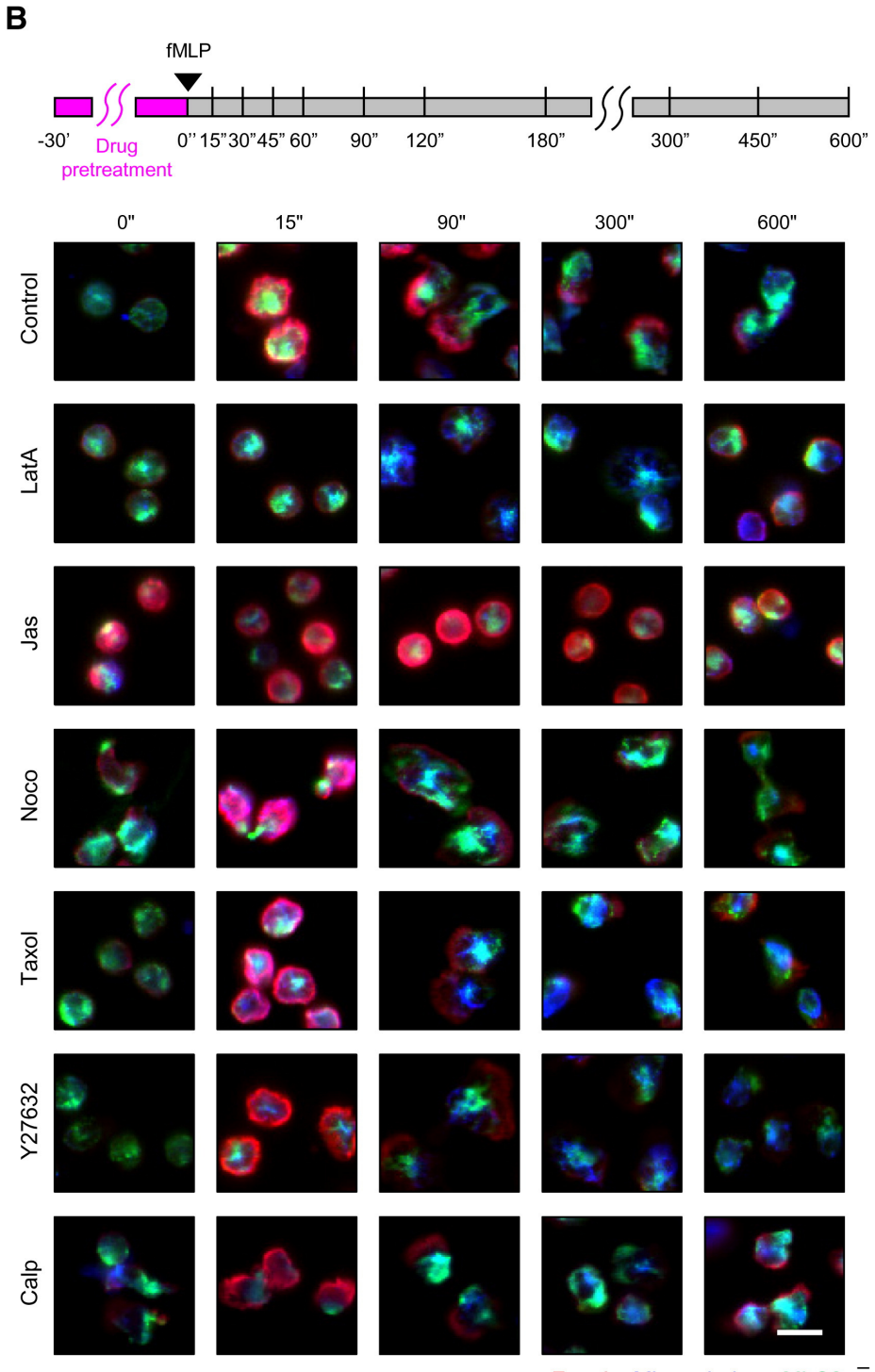
abstract


DYNAMICALLY EVOLVING CROSS-TALK
Neutrophil polarization
Front
Chemoattractant
Microtubules
Back
Polarity
of signaling
TIME
Activity
of signaling
PERSISTENT
CROSS-TALK NETWORK

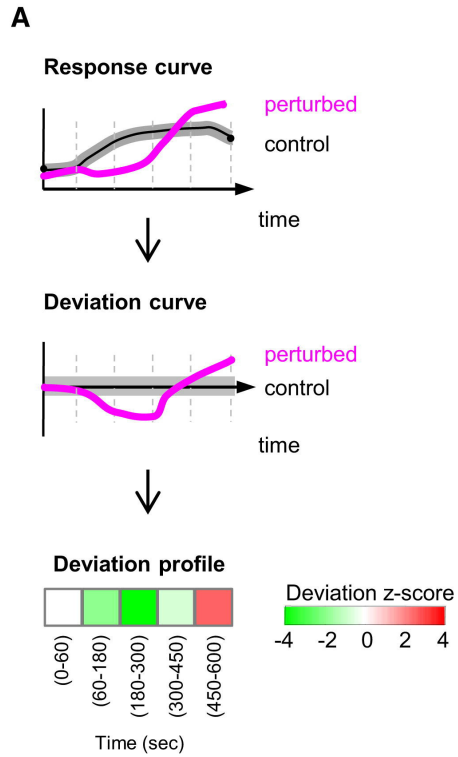
drugs

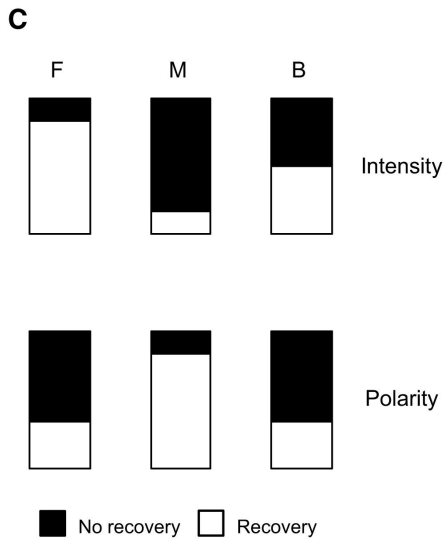

A
LatA
Jas
Y27632
Calp
F
M
B
Noco
Taxol
- Inhibitor/destabilizer
+ Enhancer/stabilizer

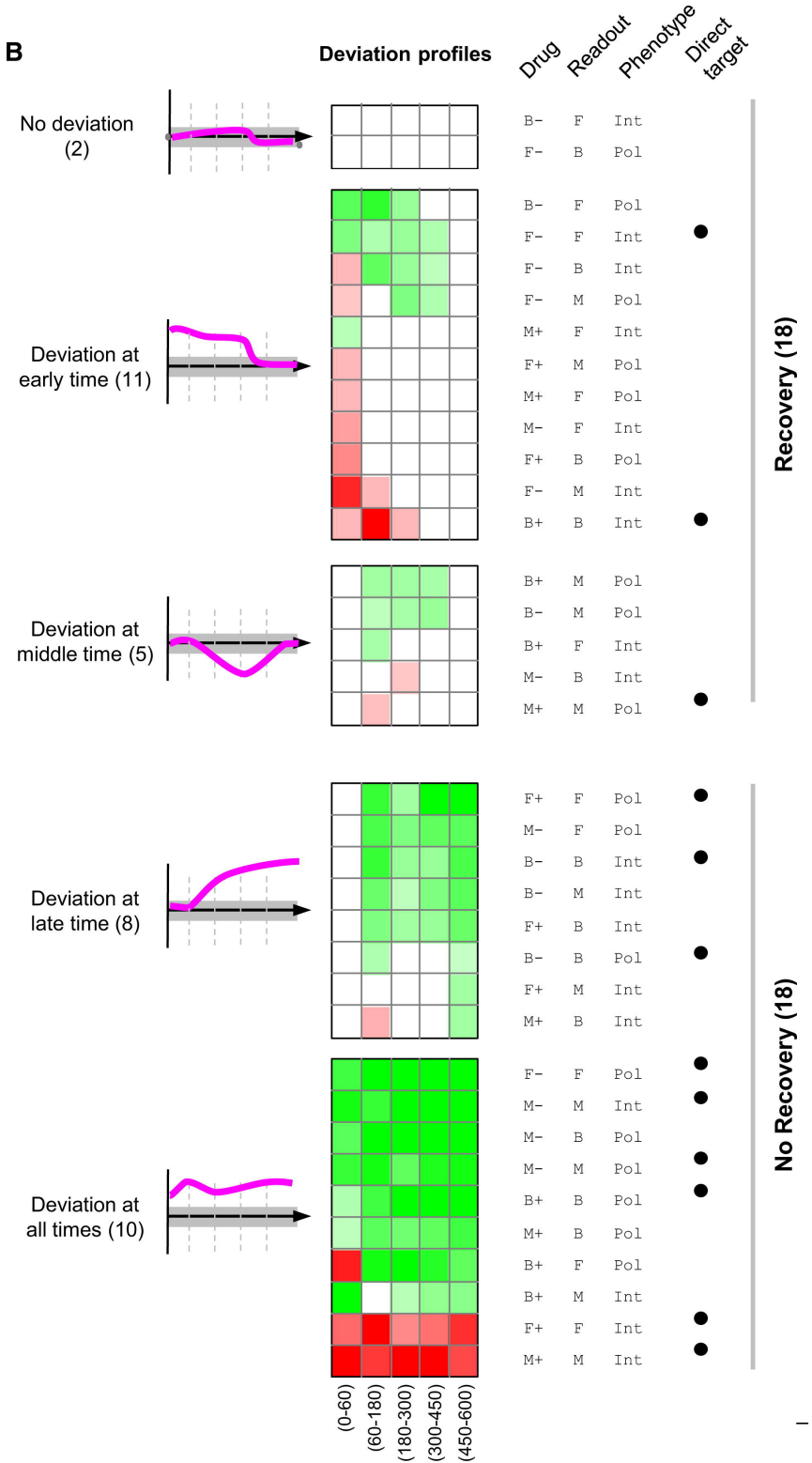

C
F-actin average intensity
Integrated change over time (mean ± standard error)
LatA (n=3)
Jas (n=3)
Microtubule average intensity
Noco (n=6)
Taxol (n=3)
p-MLC2 average intensity
Y27632 (n=3)
Calp (n=2)
(*) p<0.01


B
fMLP
-30'
Drug pretreatment
0'' 15" 30" 45" 60"
90"
120"
180"
300"
450"
600"
0"
15"
90"
300"
600"
Control
LatA
Jas
Noco
Taxol
Y27632
Calp
F-actin Microtubule p-MLC2

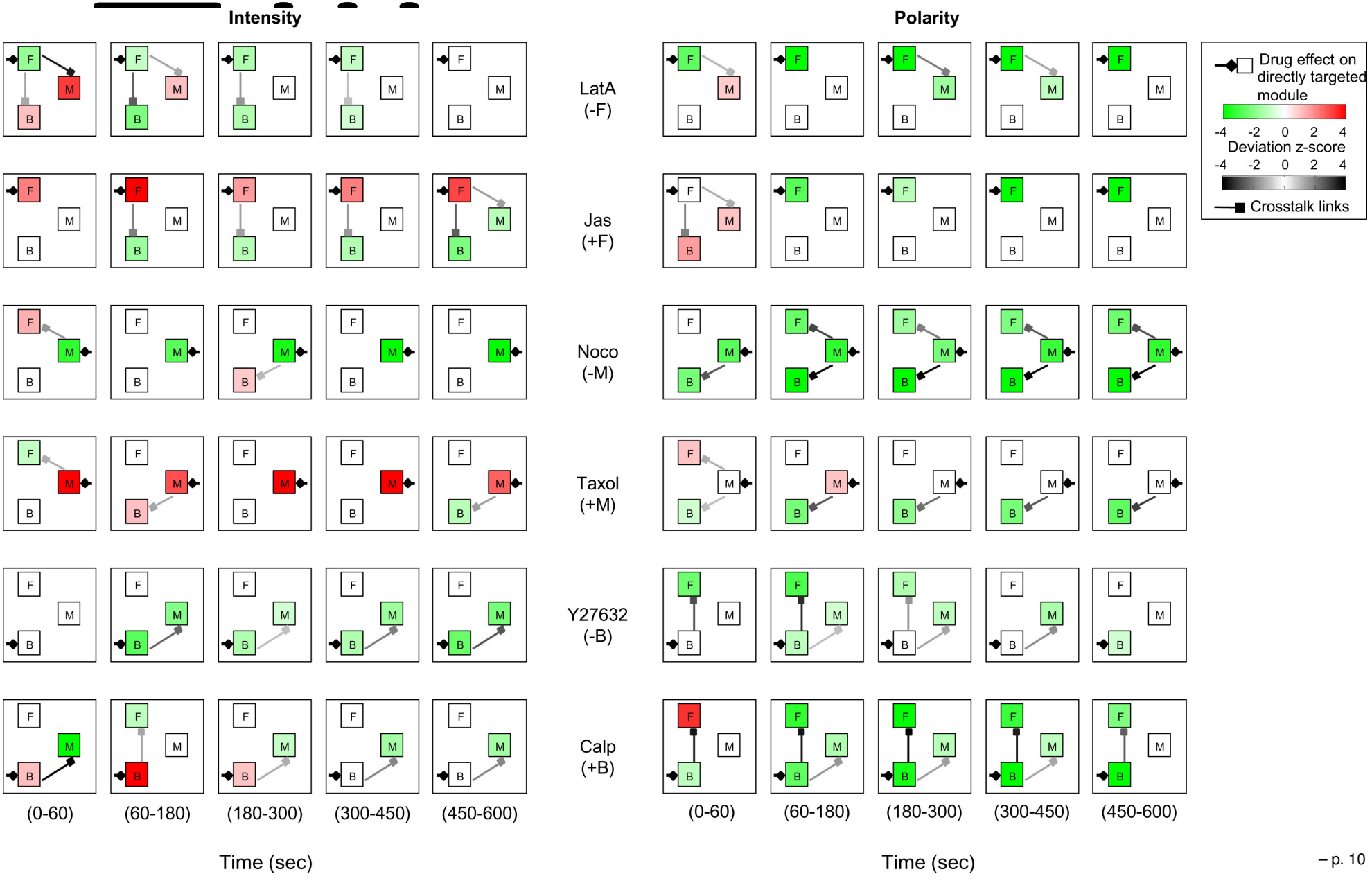
discretise


A
Response curve
perturbed
control
time
Deviation curve
perturbed
control
time
Deviation profile
Deviation z-score
-4 -2 0 2 4
(0-60)
(60-180)
(180-300)
(300-450)
(450-600)
Time (sec)

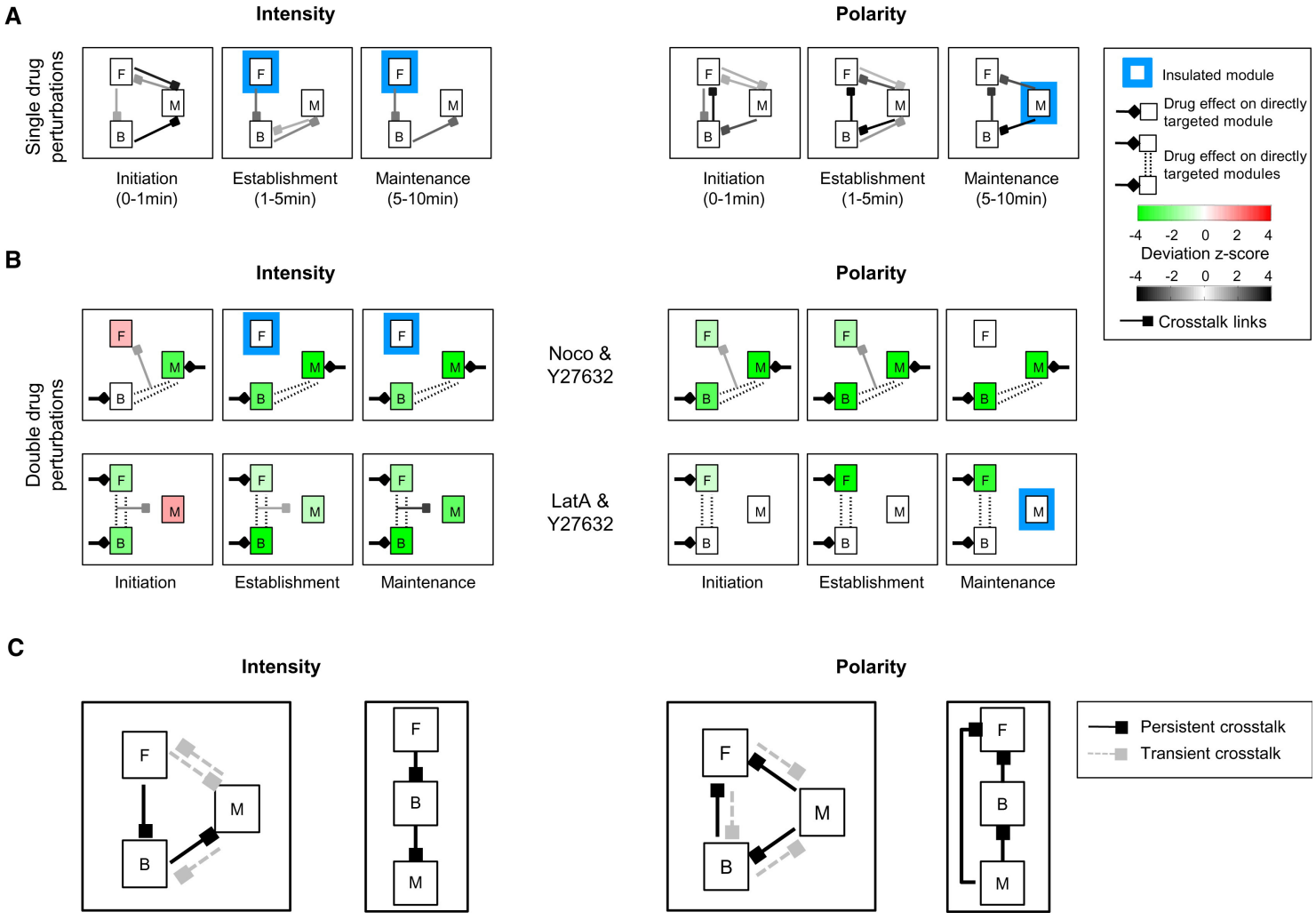

C
F
M
B
Intensity
Polarity
No recovery
Recovery


B
Deviation profiles
Drug
Readout
Phenotype
Direct target
No deviation (2)
Deviation at early time (11)
Deviation at middle time (5)
Deviation at late time (8)
Deviation at all times (10)
Recovery (18)
No Recovery (18)
(0-60)
(60-180)
(180-300)
(300-450)
(450-600)

point causality



persistent causality



BN learning with interventions

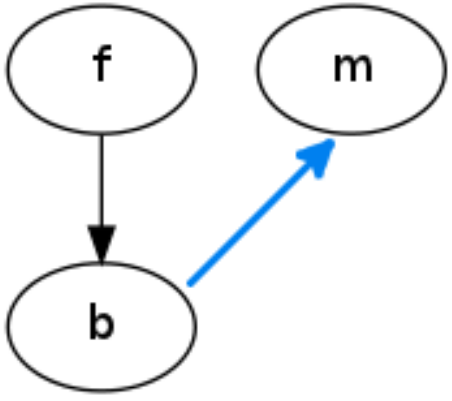
Take one:

- flatten time
- all time points of an intervened node are "set"
- 3 variables, 30 measurements per phenotype

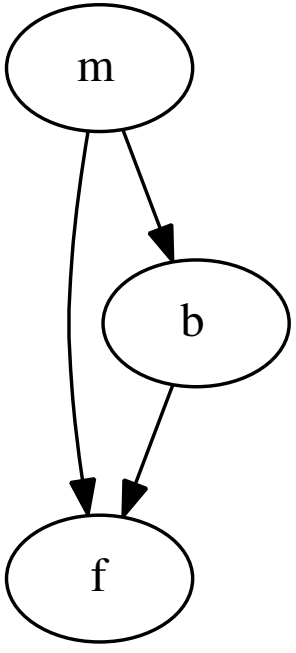
Likelihood:
unbiased function that estimates the fit of a model (here Bayesian network) to some data

$$P(M|D)$$

lo and behold

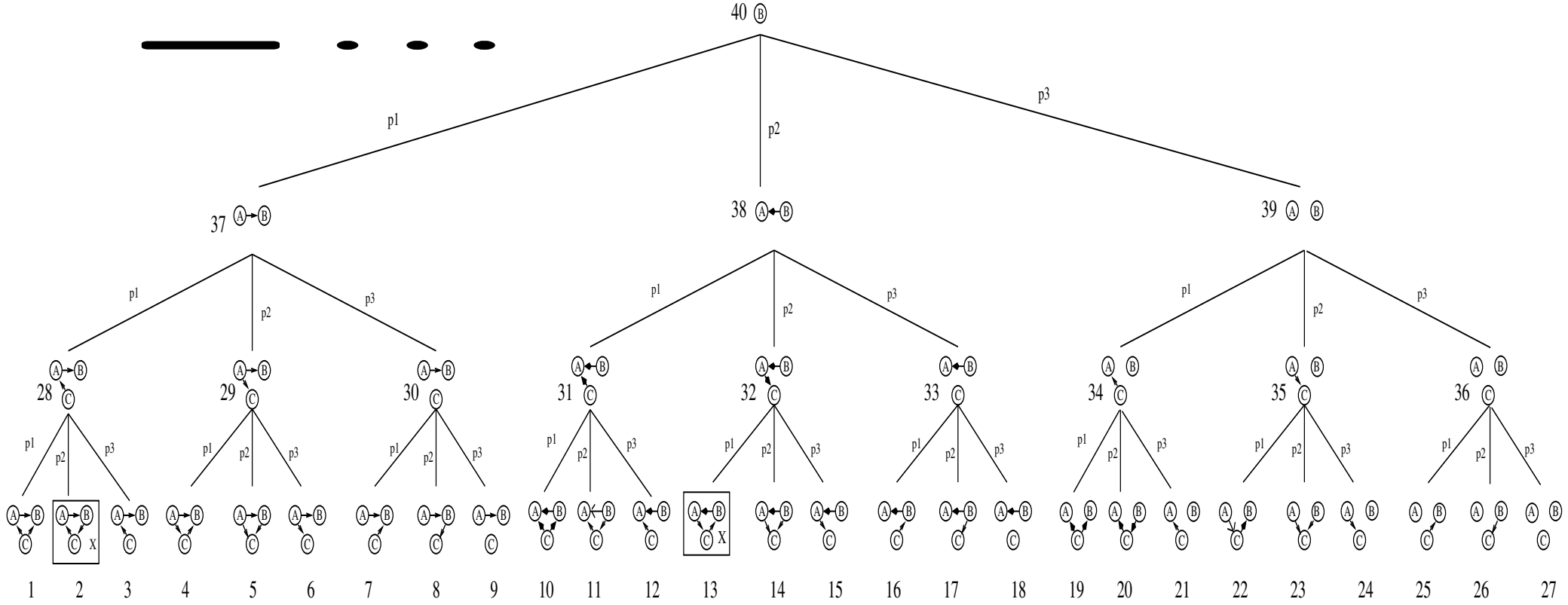


Intensity (Greedy 0.6)



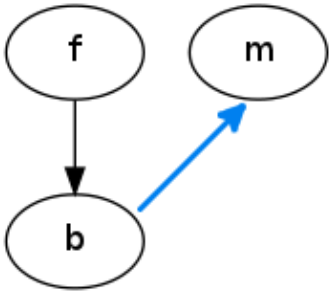
Polarity (SA 0.7)

3-nodal BNs

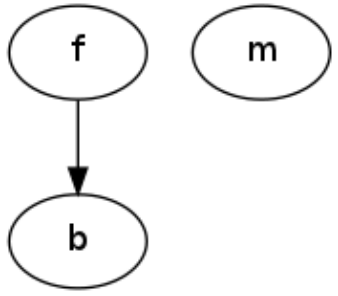


Angelopoulos & Cussens, UAI 2001

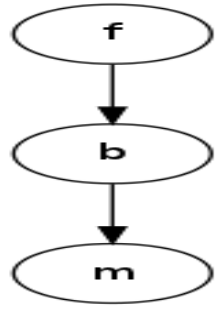
Intensity revisited



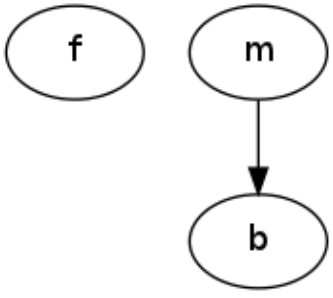
Greedy:0.6



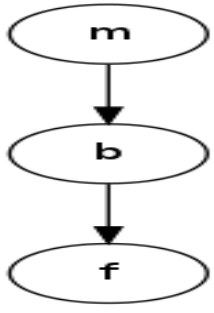
llhood:-51.3264



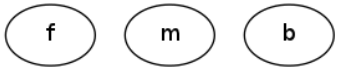
llhood:-52.8407



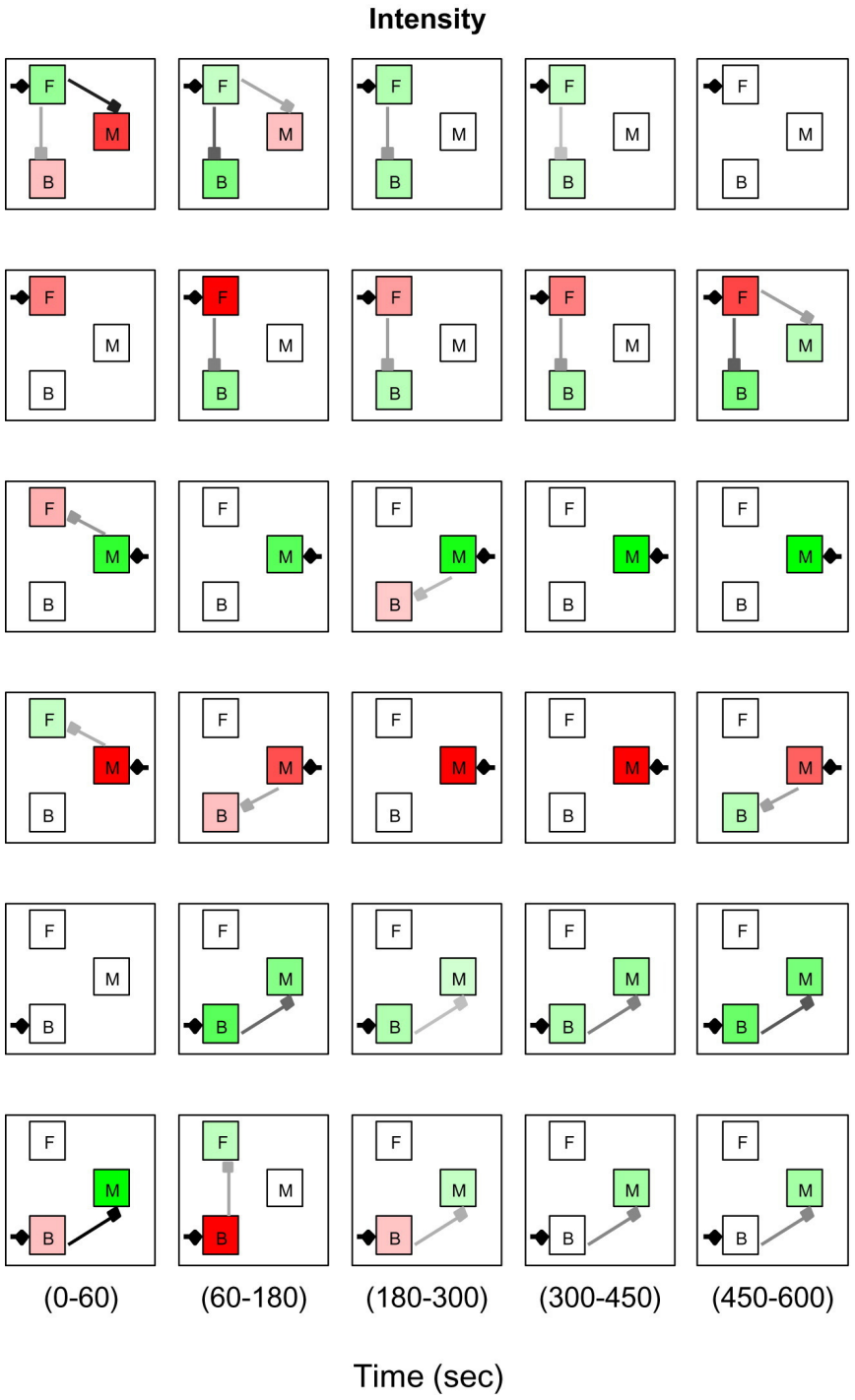
llhood:-52.2985

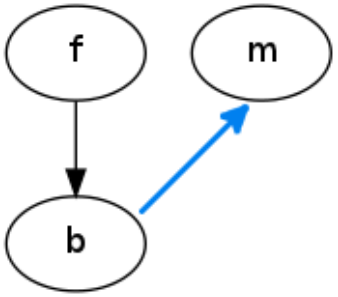


llhood: -56.4416

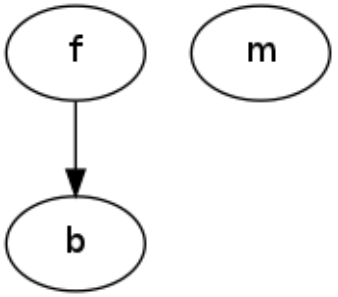


llhood: -51.8523

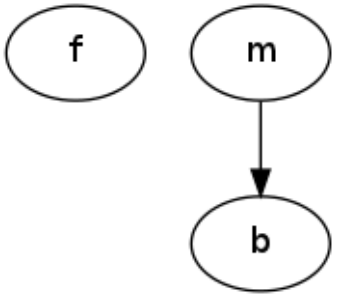


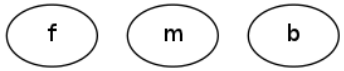


Greedy:0.6



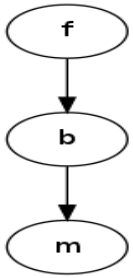
llhood:−51.3264

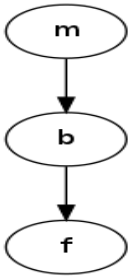




llhood:−51.8523

llhood:−52.2985





llhood:−52.8407

llhood:−56.4416

bottom line

abstracts to basic components of a system

a way forward to bridging The Gap

causal links/networks can be inferred from interventions

likelihood functions are an unbiased way to do this

motility screen

Co-analysis of (1) a motility screen of breast cancer cell lines and (2) the gene expression profiles of the panel

BC motility screen

Random Cell Migration of Human Breast Cancer Cell Lines.
(Withheld in this version as it is unpublished data.)

target identification

- cell motility screen: 45 cell lines
- adhesome library: 575 genes
- Rotterdam microarrays:
  39 cell lines, $22,283$ probes

Common

- adhesome: 519 genes, 33 cell lines
  14 basal, 19 luminal
- genome: 12466 genes

objective: find most influential genes for phenotype

methods used

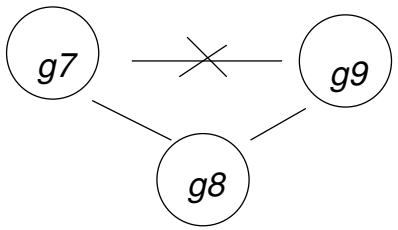
Linear regression: each gene tested separately

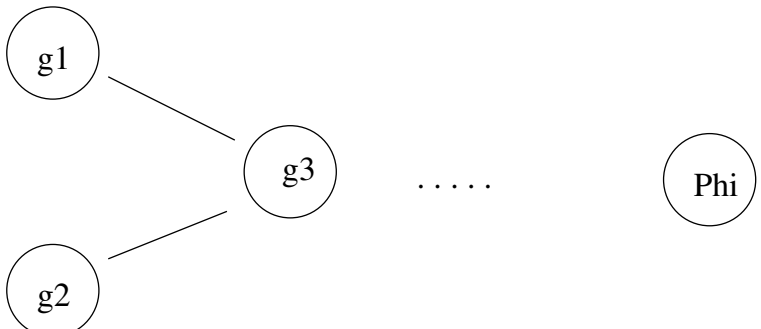
$$\bar{\phi} = \alpha \bar{g}_i + c$$

$$\bar{\phi} = \alpha \bar{g}_i + \beta \bar{t} + c$$

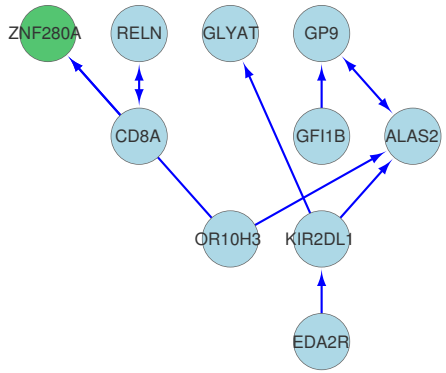
anova (analysis of variance)
provides a confidence on the fit: p-value

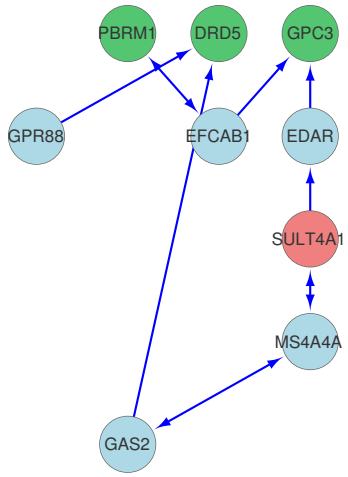
Do calculus: assume Gaussians and take into account parents in a constructed network.

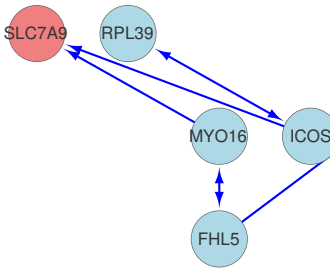


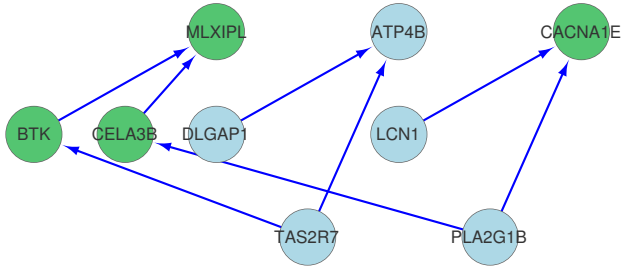


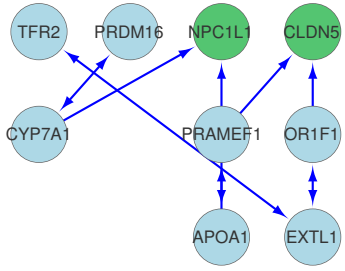
network ex. sparse



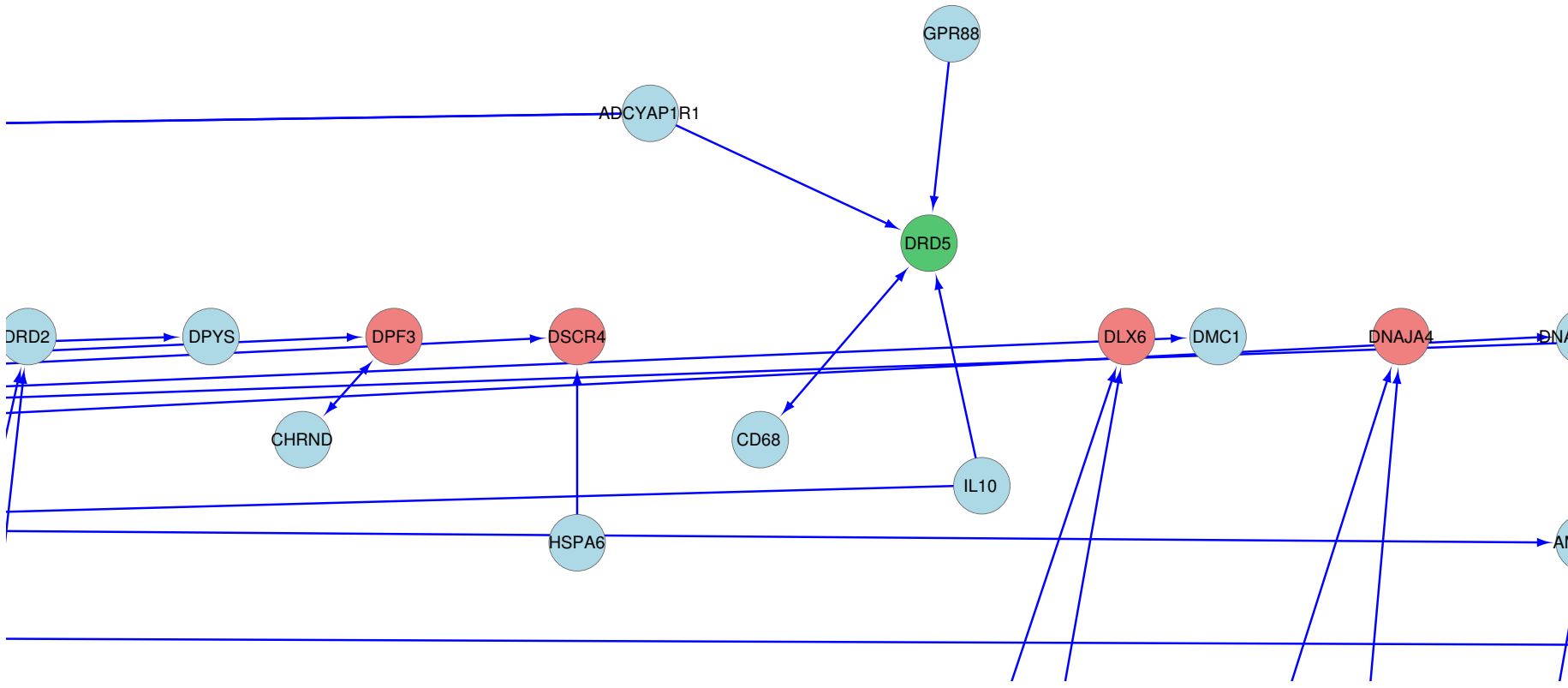








network ex. dense



significant hits (LR)

linear regression, typed (33) cell lines :

	p< 0.01	p< 0.05	q< 0.01	q< 0.05	tested
library	93	154	39	90	519
gwide	1391	2905	93	847	12,465
common	93	154	9	62	

choices

$$\{LR, PC\} \times \{lib, gen, lgr\} \times \{typ, all, bas, lum\}$$

Stratify

top-tier

- $LR \times \{lib, gen\} \times typ$
- $PC \times adh \times all$

lower-tier

- $LR \times \{lib, gen\} \times \{bas, lum\}$
- $PC \times \{gnm, lgr\} \times all$

Total ? Positive + Negative ? Controls !?
Measure of success ?

bottom line

We provide 2 different ways in which to prioritise between genes that are all relevant (adhesome library).

Among the top hits of the genome-wide search we identified genes that could easily be included in the adhesome library.

or is it ?

new microarray set

39 common cell lines 114 samples

18 Basal lines 7 Basal A 11 Basal B

21 Luminal

focus:

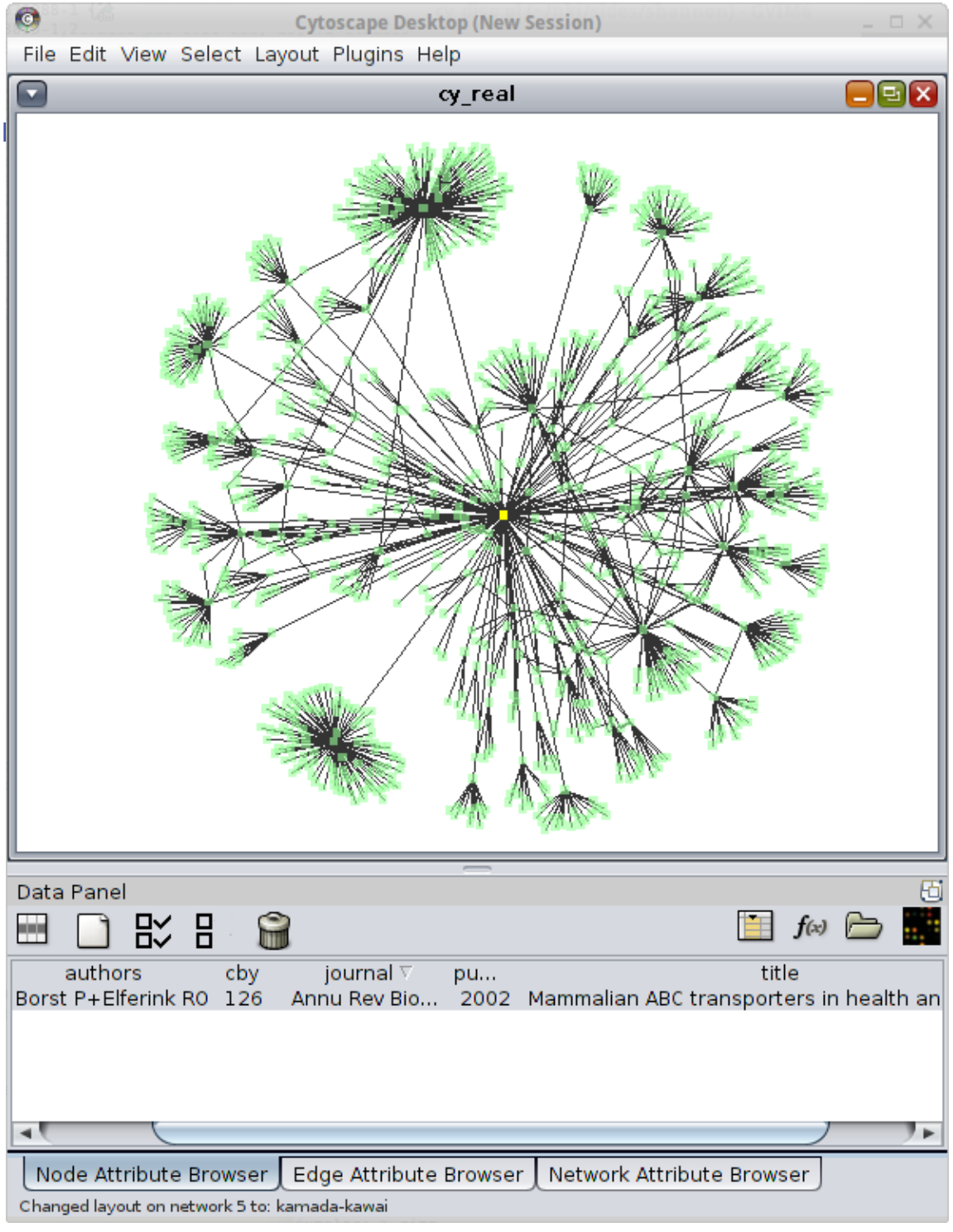
LR to select new genome-wide candidates, use those to build networks to select top candidates.

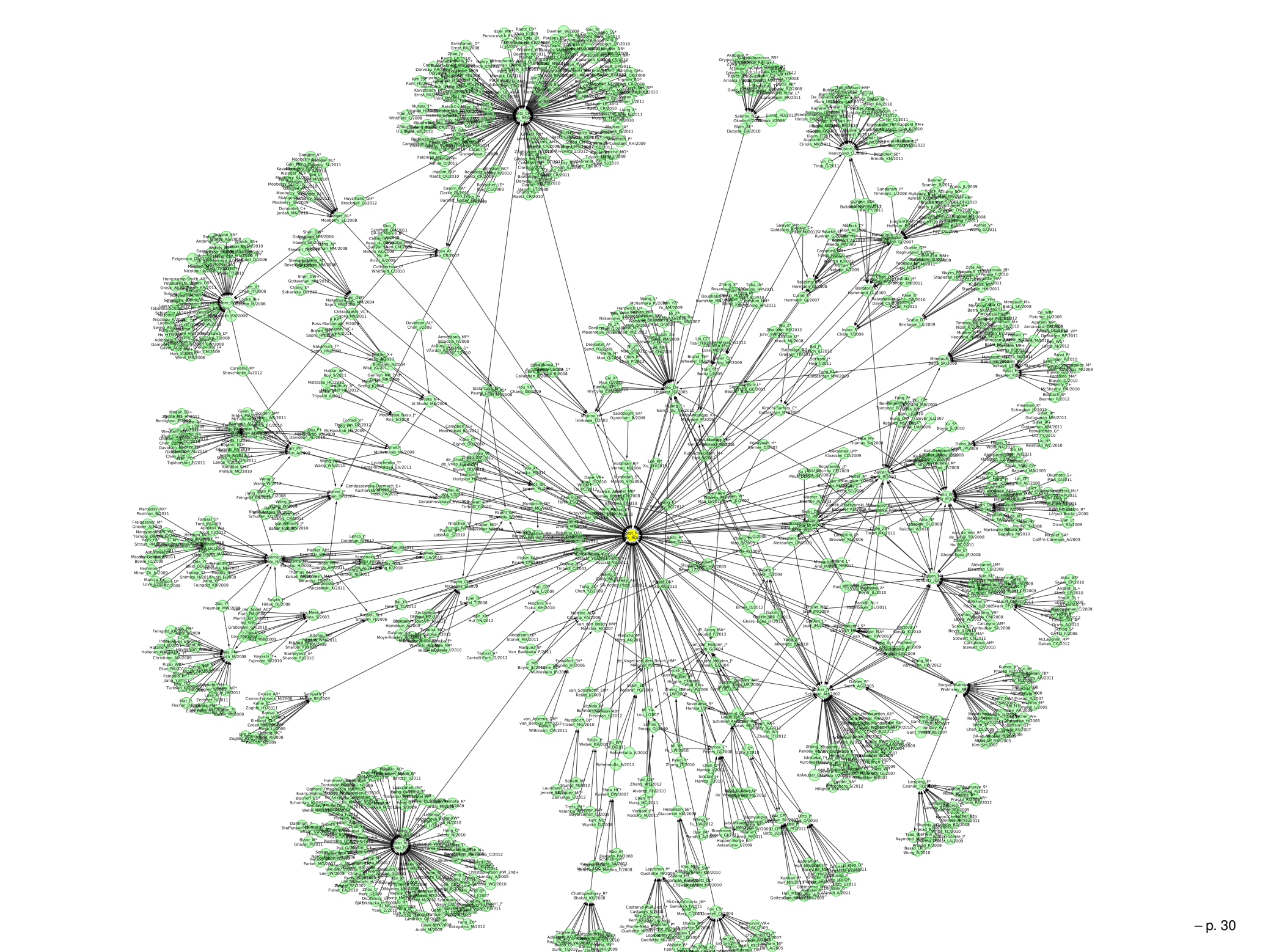
pubGraph

from a seed set of papers collect and display all papers citing them (recursively)

- interrogate PubMed
- store incremental local copies in a variety of formats
- visualise and graphs and extract information interactively

Mammalian ABC transpo





bottom line

Can be used to quickly and efficiently navigate literature in a specific area.

Identify hubs.

Connect pdfs for immediate browsing.

Thanks

Bioinformatics Group

Sander Canisius

Andreas Schlicker

Lodewyk Wessels

Leiden

Vasiliki-Maria Rogkoti

Sylvia Le Devedec

Bob van de Water

Erasmus MC

Karin Legerstee

Adriaan Houtsmuller

Marcel Smid

John Martens

Ubrecht Institute

Emma Spanjaard

Johan de Rooij