Characterize gene expression profiles from multiple myeloma patients by using consistent graph modeling

Bertrand Miannay

MeForBio team, IRCCyN
Team 11, CRCNA

Directeur de thèse : Olivier Roux (IRCCyN, Nantes)
Co-encadrants : Stéphane Minvielle (CRCNA, Nantes)
Co-encadrant : Carito Guziolowski (IRCCyN, Nantes)

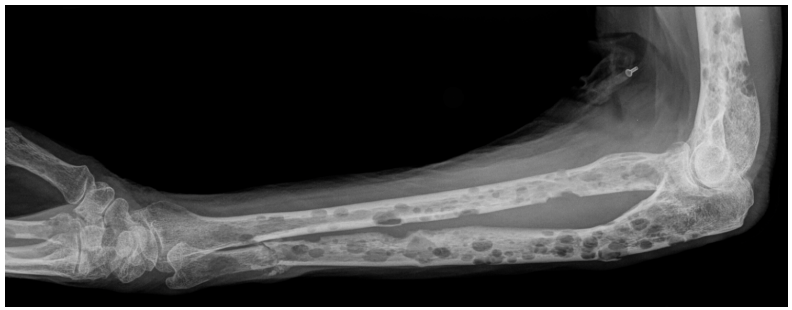
12 décembre 2015

Sommaire

Sommaire

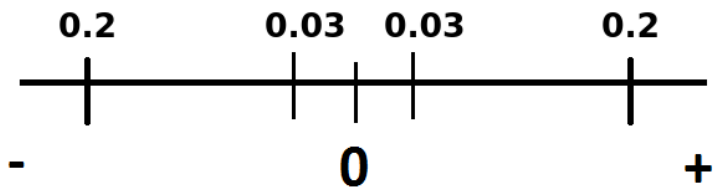
Multiple Myeloma

- Neoplasm of plasma cells.
- 1% of cancers in France.
- Incurable disease despite considerable progress in treatment.
- Characterized by a profound intra- and inter-individual heterogeneity[Magrangeas et al., 2013].
- Need to optimize discovery of clinically relevant anti-MM agents.



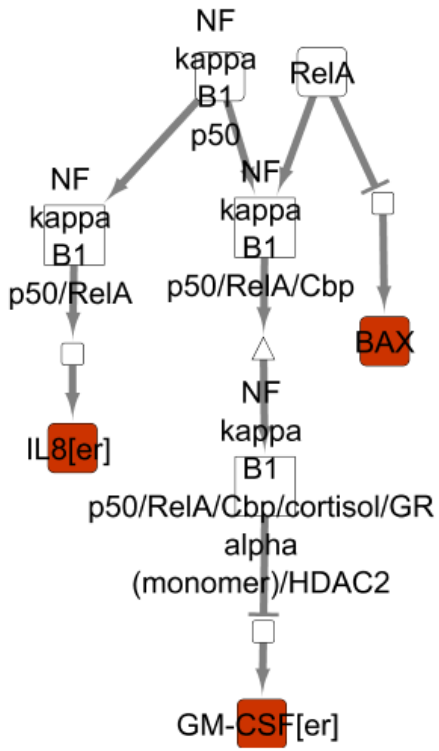
Transcriptomic data

- Genes expression profiles generated :
  - Using Affymetrix Human Exon1.0 chips.
  - With 50ng total RNA from highly purified bone marrow CD138+ plasma cells.
  - Otained from MM patients of the IFM (Intergroupe Francophone du Myélome) centers and healthy donors .
- Data from 611 individuals.
  - 9 Normal Individuals (NI).
  - 602 patients with Multiple Myeloma (MM).
- Discretization to identify variant (over/under-expressed) and unvariant genes with respect to NI



Regulatory networks

- PID-NCI (Pathway Interaction Database) [Schaefer et al., 2009]
  - 18154 nodes (proteins, genes, complexes, transcription, reaction, etc.) and 29936 edges (activation, inhibition, complexation, etc.)
  - 634 genes (with transcription events upstream)



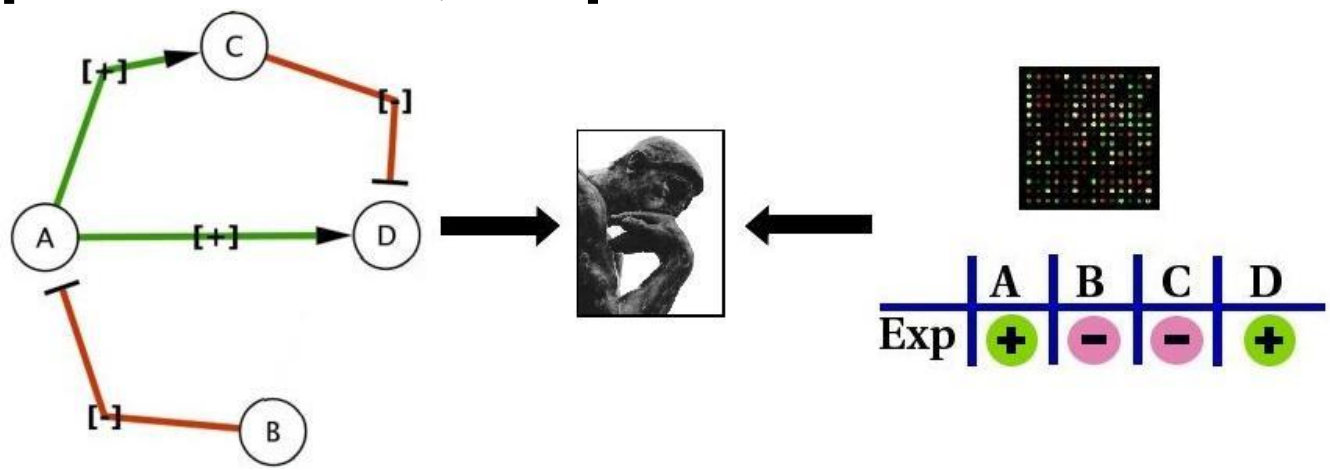
Aims

- Confront regulatory networks with logical reasoning to expression profiles.
- Infer specific molecular profiles among MM patients.
- Identify therapeutic targets.
- Identifye key nodes' and their up/down-regulation.

Sommaire

Graph's coloration

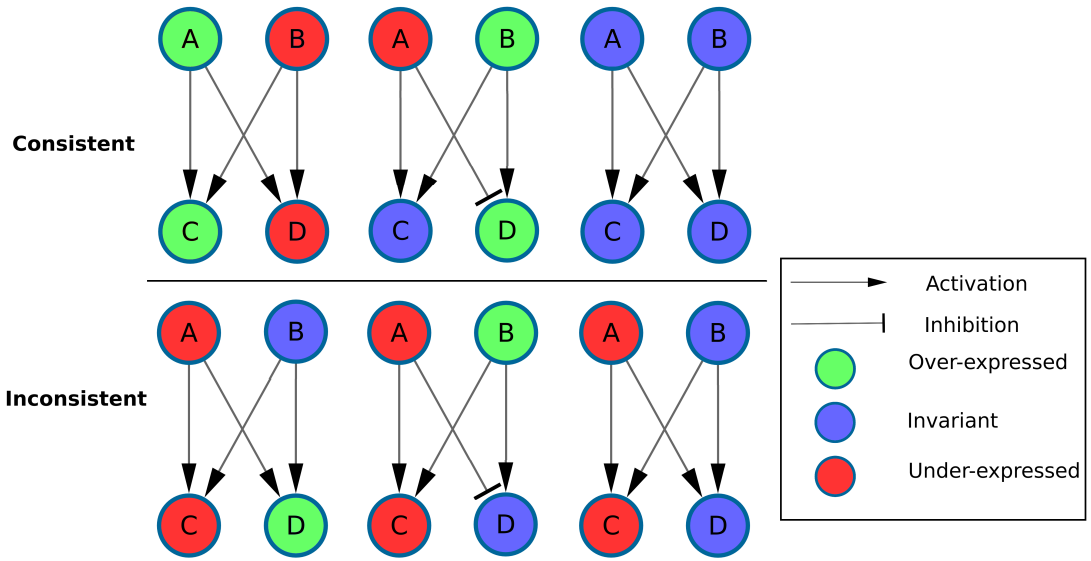
- Confrontation between graph's topology and observations [Guziolowski et al., 2012]



- Input :
  - Simple oriented graph {+, -}
  - Observations data : +, -, 0
- Output :
  - Unobserved nodes' coloration : +,-,0

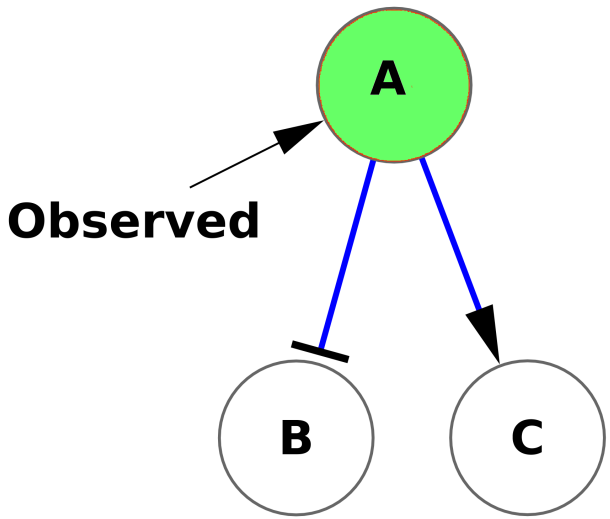
Consistency rules

- All inputs (nodes without predecessor) are consistent
- Variant nodes (+,-) have to be explained by, at least, one predecessor
- Unvariant nodes (0) have to be explained by :
  - Unvariant predecessors
  - Two opposite predecessors : One activator and one inhibitor



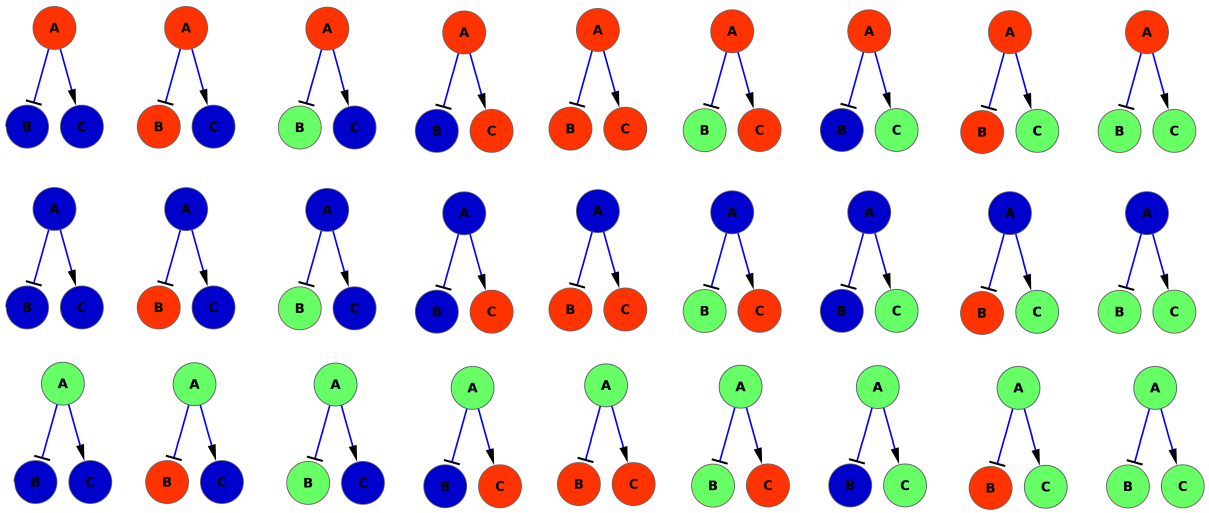
Example 1

- One graph with 3 nodes and 2 edges
- 1 observation : A = +



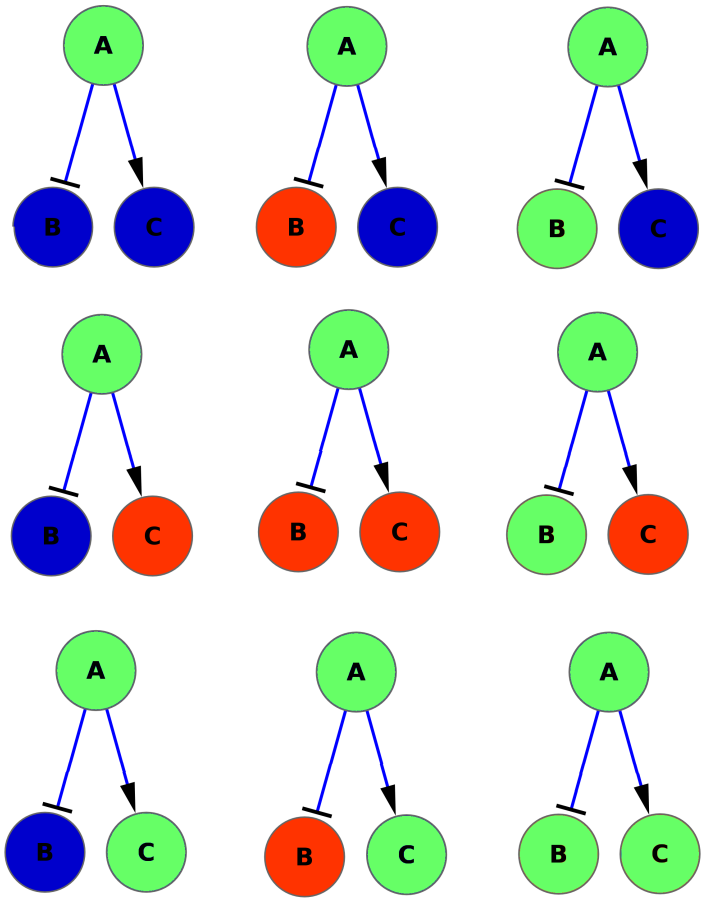
Example 1

- Instanciation of all graph's colorations :



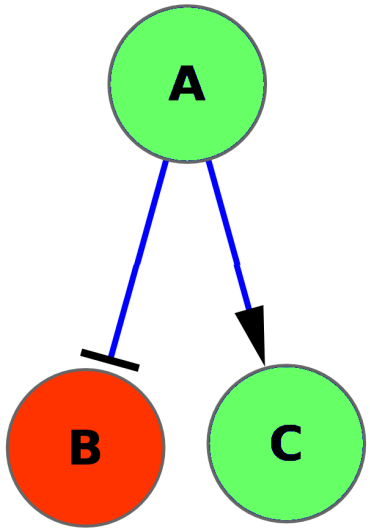
Example 1

- Reduction with the observed data (A = +) :



Example 1

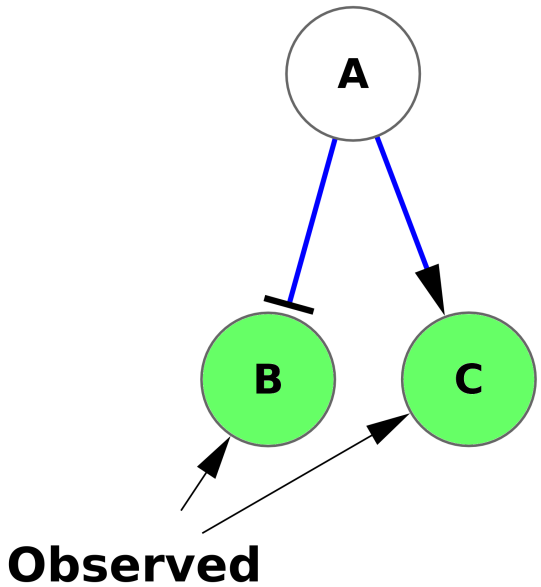
- Reduction with consistency rules :



- B can be predicted as "-" and C as "+"

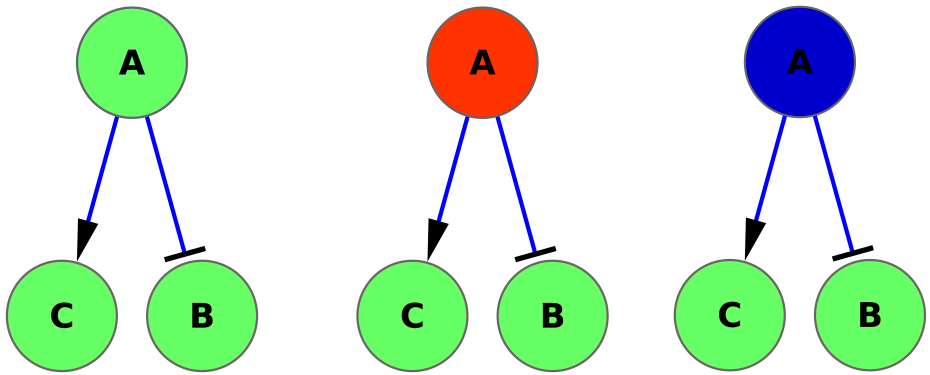
Example 2

- A graph with 3 nodes, 2 edges
- 2 observations : B,C = +



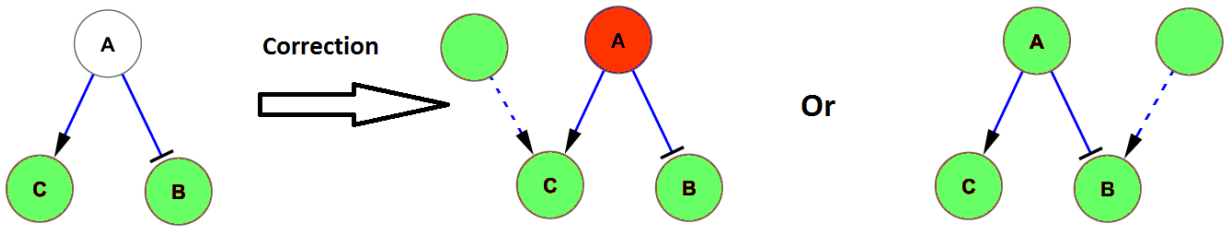
Example 2

- Instanciation then reduction with observed data :



Example 2

- Reduction with consistency rules : Empty solution
- Correction by adding influence : $|mcos| = 1$ [1]



- A can be predicted as + or - $\Rightarrow$ "CHANGE"

1. MCOS : minimal correction set

Sommaire

Graph generation and predictions

- Analysis with the 611 individuals with IGGY [Thiele et al., 2015] (Answer Set Programming)
- Generation of a graph from PID :
  - Method : Connecting the myeloma cell survival, proliferation and drug resistance factors(IL6, IGF1 et CD40)[Klein, 2010] to variant genes by shortest path**s**
  - Result : Graph with 621 nodes and 805 edges
    - Contains 557 genes
- Predictions after correction of nodes' signs for each patient

Graph generation and predictions

Signs	Observed		Predicted	
	NI (9)	MM (602)	NI (9)	MM (602)
+	34 %	38 %	26 %	40 %
-	34 %	51%	26 %	35 %
0	32 %	11 %	24 %	4 %
Change	0 %	0 %	7 %	8 %
Not+	0 %	0 %	2	1 %
Not-	0 %	0 %	3 % %	$<$ 1 %
?	0 %	0 %	13 %	12 %
total	2195 (244)	222299 (369)	5589 (377)	373842 (252)

How to compare predicted value

- Couple representation (node,sign) : 621x3 variables

prediction	(node,+)	(node,-)	(node,0)
+	1	0	0
-	0	1	0
0	0	0	1
CHANGE	1	1	0
Not+	0	1	1
Not-	1	0	1
?	1	1	1

TABLE : Table of association between signs' projections and couples

Prediction analysis : Supervised learning

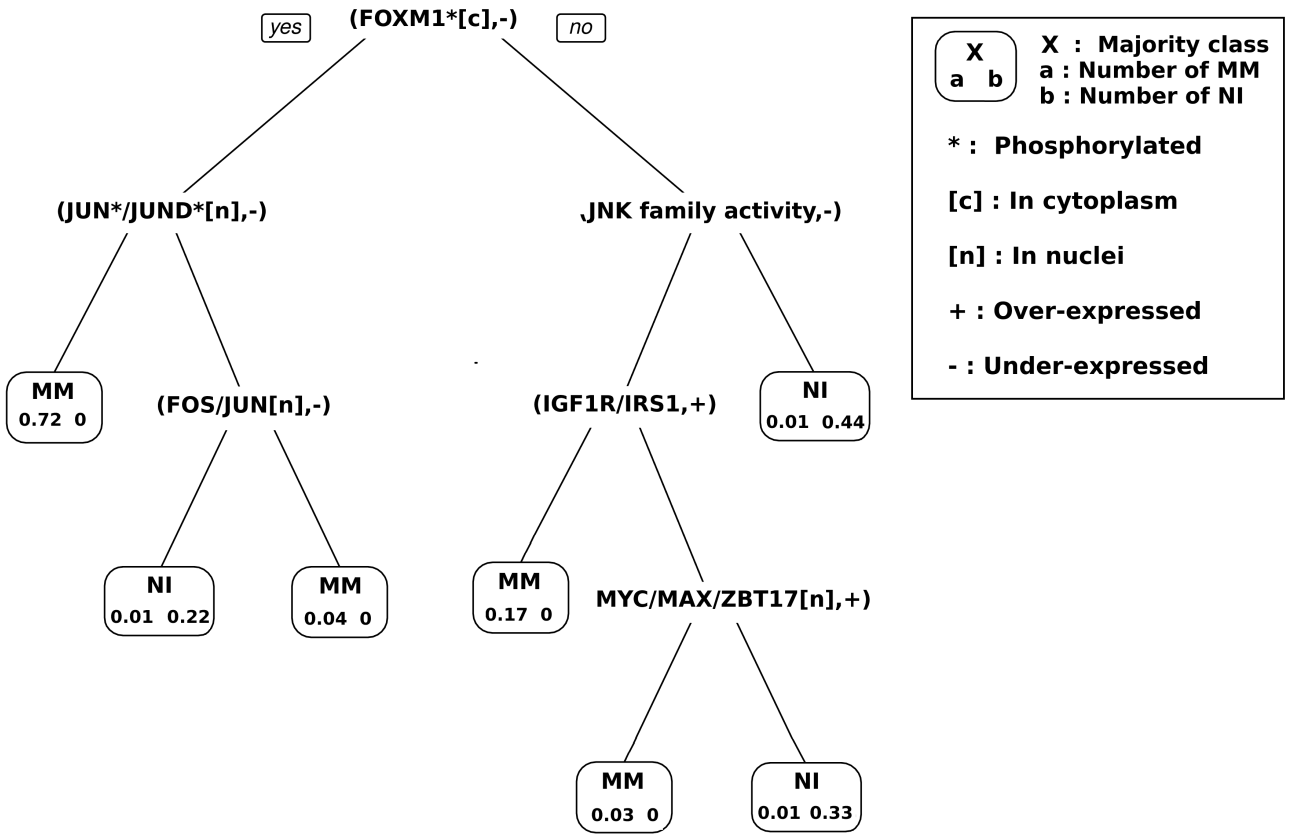


FIGURE : Supervised learning between NI and MM

Prediction analysis : Frequency analysis

node	sign	Frequency NI	Frequency MM	p.val (fisher)
FOXM1*[c]	-	0,222	0,776	0,0008
RB1/E2F1-3/DP[n]	+	0,333	0,829	0,0015
STAT-6[c]	-	0,333	0,822	0,0018
JNK family activity	-	0,444	0,885	0,0019
IL23 pathway	-	0,444	0,8787	0,0025

TABLE : Top 5 : Frequency results analysis

Sommaire

Conlusion

- Method :
  - Merge informations from expression data and regulatory network.
  - Deduct new informations from those data : Nodes and their sign.
  - Characterize distinct groups :
    - Identify Key nodes.
- In progress :
  - Simulation of therapeutic target effectiveness.

Merci de votre attention

Références I

Guziolowski, C., Kittas, A., Dittmann, F., and Grabe, N. (2012).

Automatic generation of causal networks linking growth factor stimuli to functional cell state changes.
The FEBS journal, 279(18) :3462–74.

Klein, B. (2010).

Positioning NK-kappaB in multiple myeloma.
Blood, 115(17) :3422–4.

Magrangeas, F., Avet-Loiseau, H., Gouraud, W., Lodé, L., Decaux, O., Godmer, P., Garderet, L., Voillat, L., Facon, T., Stoppa, a. M., Marit, G., Hulin, C., Casassus, P., Tiab, M., Voog, E., Randriamalala, E., Anderson, K. C., Moreau, P., Munshi, N. C., and Minvielle, S. (2013).

Minor clone provides a reservoir for relapse in multiple myeloma.
Leukemia, 27(2) :473–81.

Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., and Buetow, K. H. (2009).

PID : the Pathway Interaction Database.
Nucleic acids research, 37(Database issue) :D674–9.

Thiele, S., Cerone, L., Saez-Rodriguez, J., Siegel, A., Guziołowski, C., and Klamt, S. (2015).

Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies.
BMC bioinformatics, 16(1) :345.