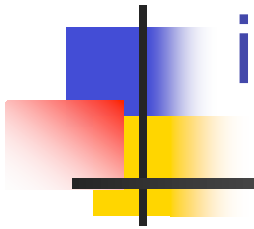
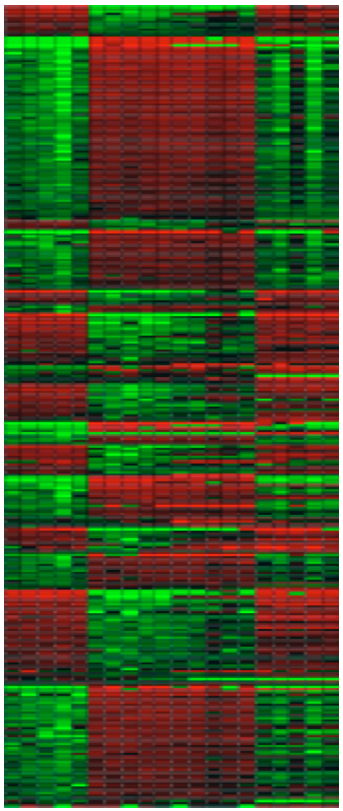
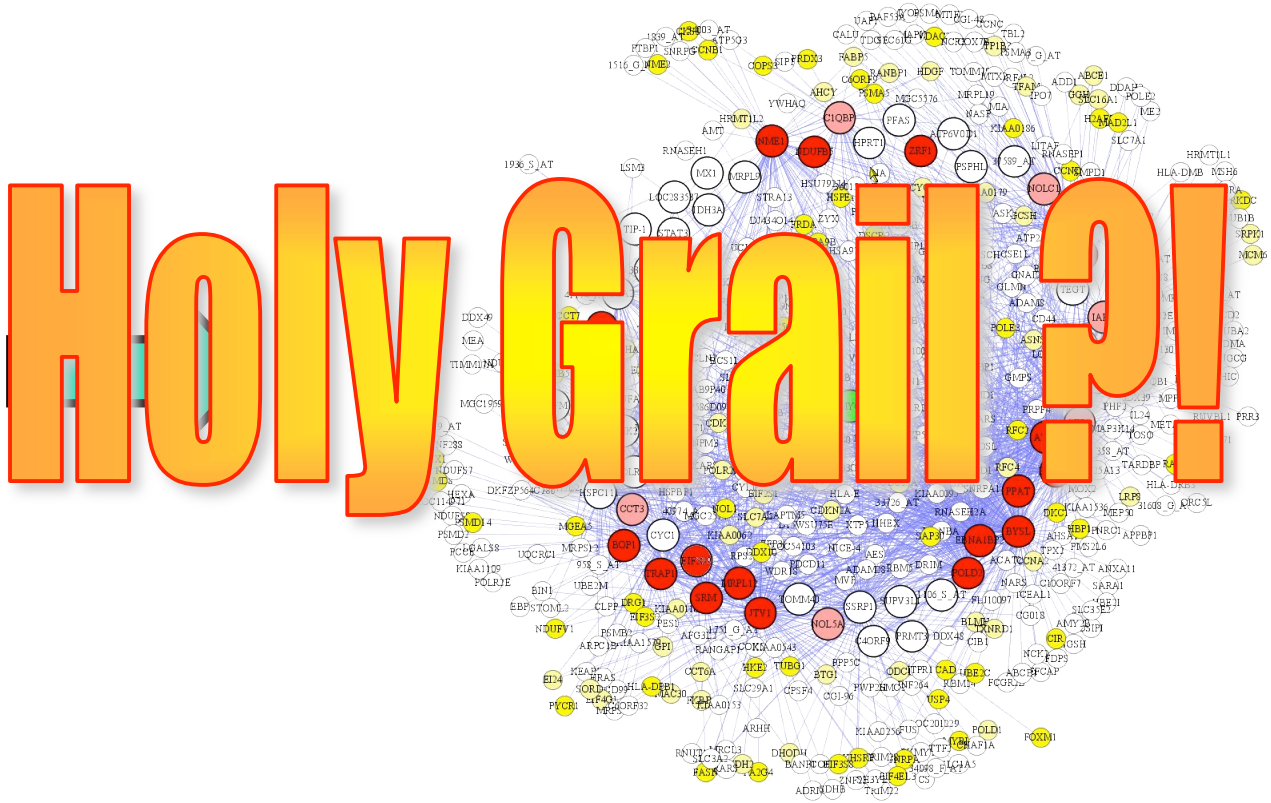


Genome-wide discovery of modulators of transcriptional interactions in human B lymphocytes

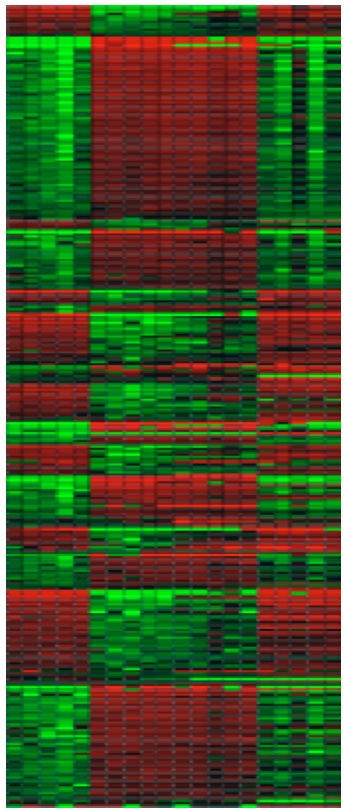
Ilya Nemenman
(JCSB/Columbia → CCS-3/LANL & SFI)
Kai Wang, Nilanjana Banerjee,
Adam Margolin, Andrea Califano
(JCSB/Columbia)

Reconstructing cellular interactions

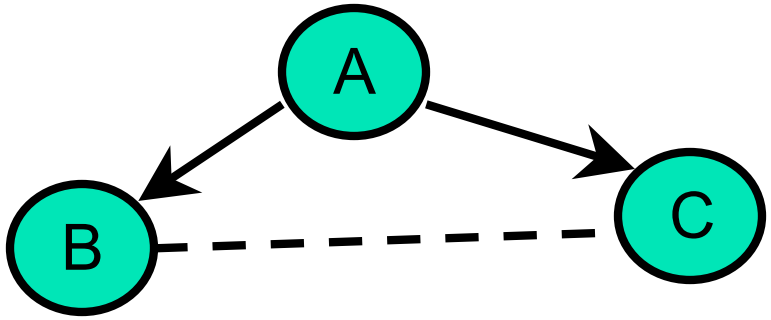




Reconstruction algorithms: The curse of "percent correct"

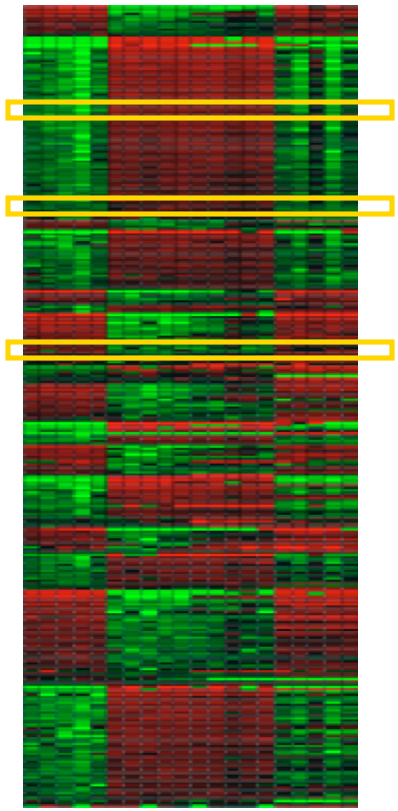


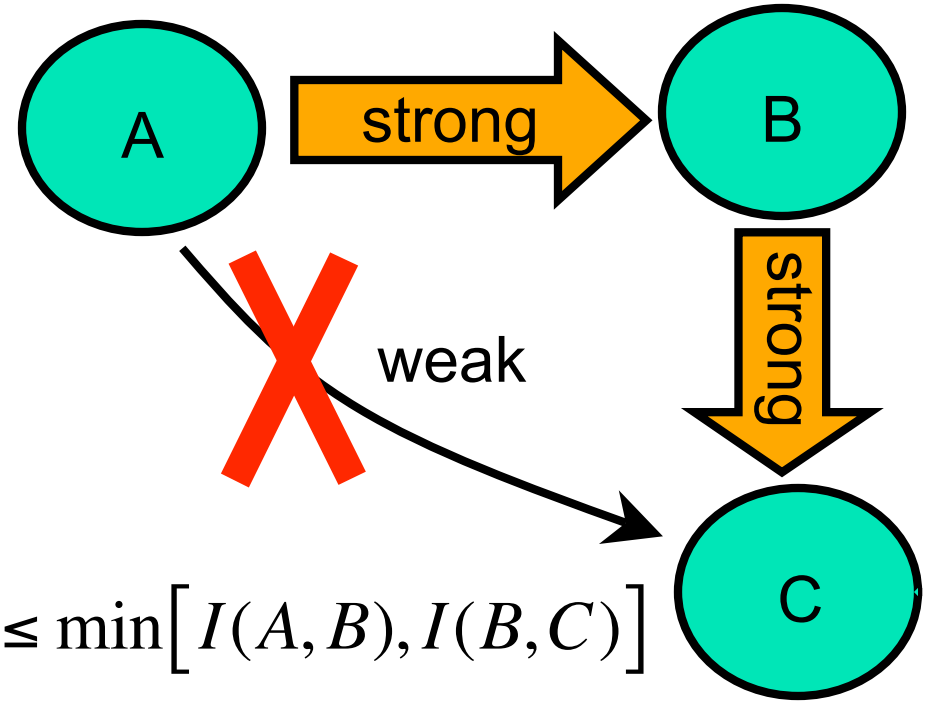
	Stat	Co	GM	Biochem.
Small data requirements	✖✔	✔	✖✔	✖
Robustness to fluct.	✔	✔	✖✔	✖
Computational complexity	✖	✔	✖	✖✔
Conditional interactions	✔	✖✔	✔	✖✔
Reparam inv., non-param.	✖✔	✖✔	✖✔	✔
Irreducibility	✔	✖	✔	✖



Influenciomics

ARACNE
(Data Processing Inequality, DPI)

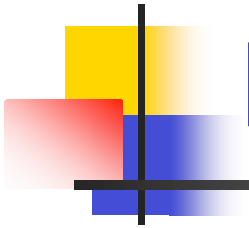




$$I(A,C) \le \min\left[I(A,B), I(B,C)\right]$$

Reparm. inv.; small sample; low complexity.

Performance?



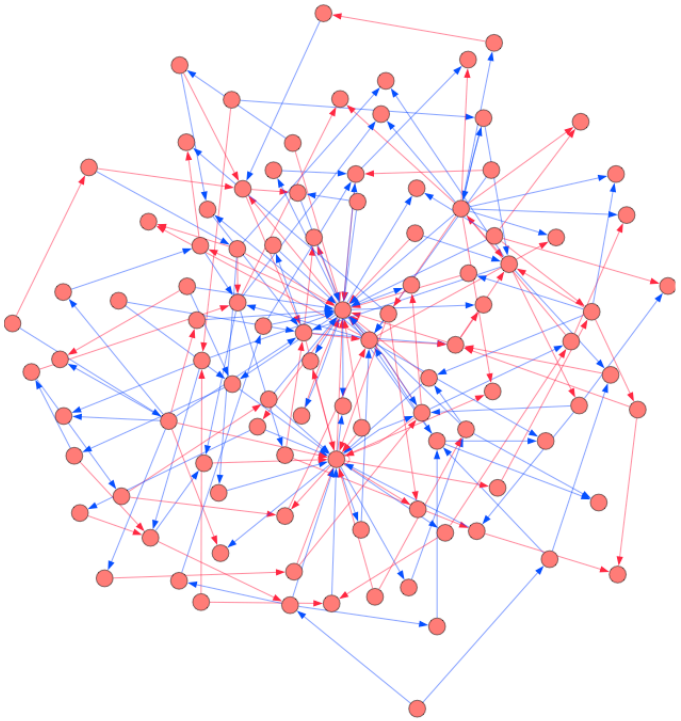
Performance: No false positives

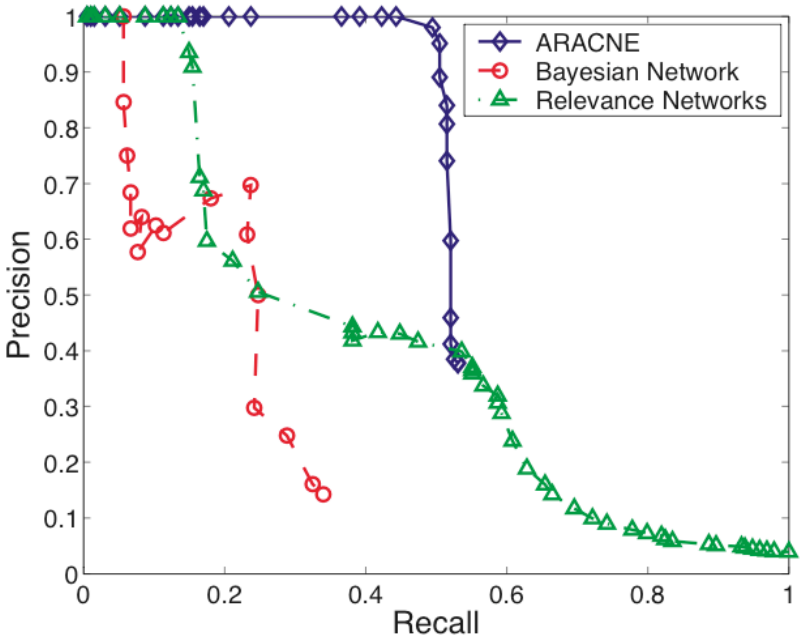
Theorem 1. If MIs can be estimated with no errors, then ARACNE reconstructs the underlying interaction network exactly, provided this network is a tree and has only pairwise interactions.

Theorem 2. The Chow-Liu maximum mutual information tree is a subnetwork of the network reconstructed by ARACNE.

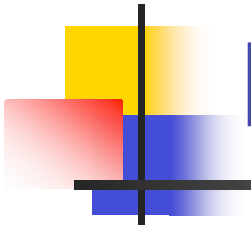
Theorem 3. Some sparseness/loopiness assumptions -- no false positives (no false negatives under stronger conditions).

Synthetic networks





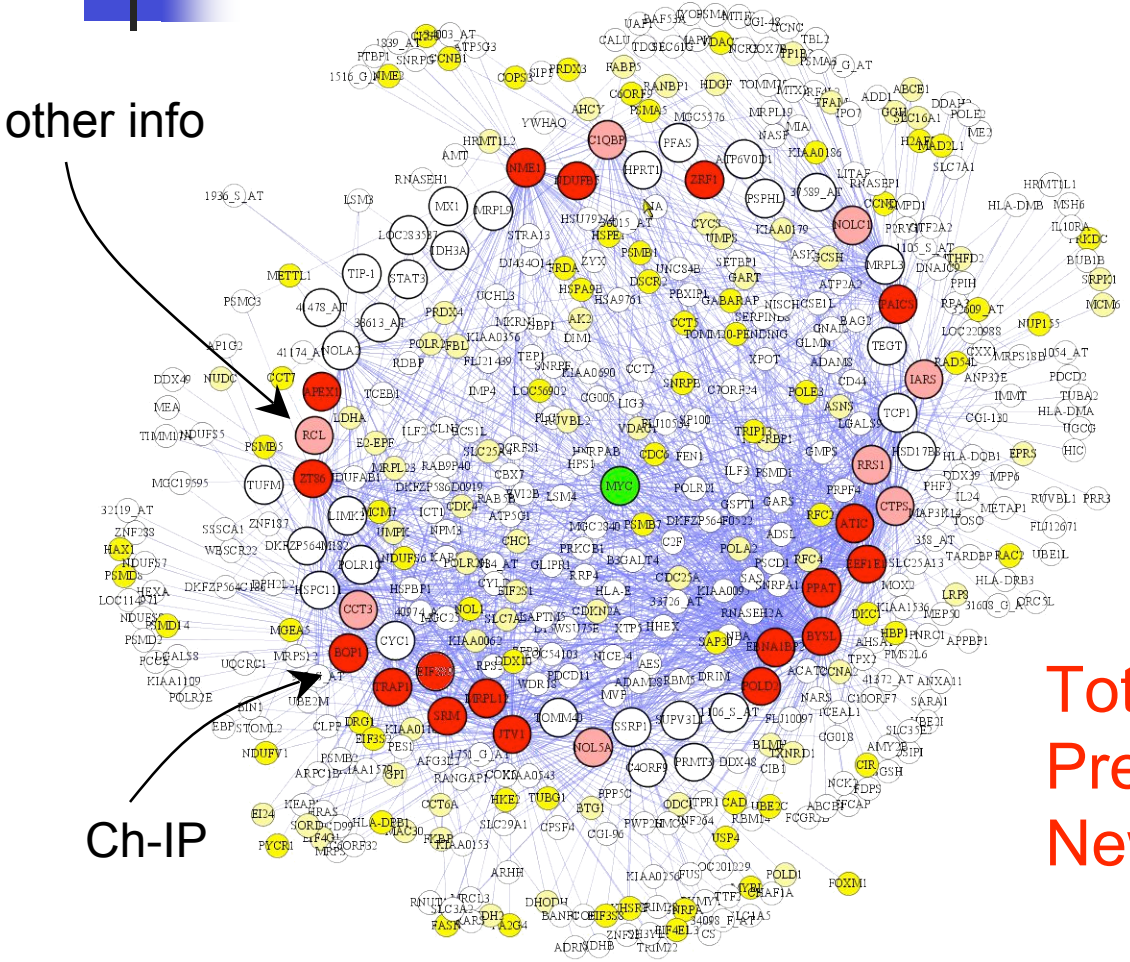
$$\frac{dx_i}{dt} = a_i \prod_j \frac{I_{0,j}^{\nu_j}}{I_j^{\nu_j} + I_{0,j}^{\nu_j}} \prod_j \left(1 + \frac{A_{0,j}^{\nu_j}}{A_j^{\nu_j} + A_{0,j}^{\nu_j}} \right) - b_i x_i$$



B-cell dataset

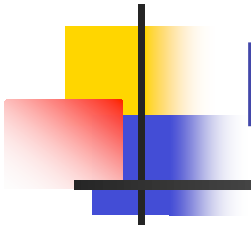
- ~400 arrays
- No dynamics
- ~250 naturally occurring, ~150 perturbed
- ~25 phenotypes (normal, tumors, experimental perturbations)
- Expression range due to differential expression in different phenotypes

c-MYC subnetwork



- Protooncogene,
- 12% background binding,
- one of top 5% hubs
- significant MI with 2000 genes

Total interactions: 56
Pre-known: 22
New Ch-IP validated: 11/12



Problem:

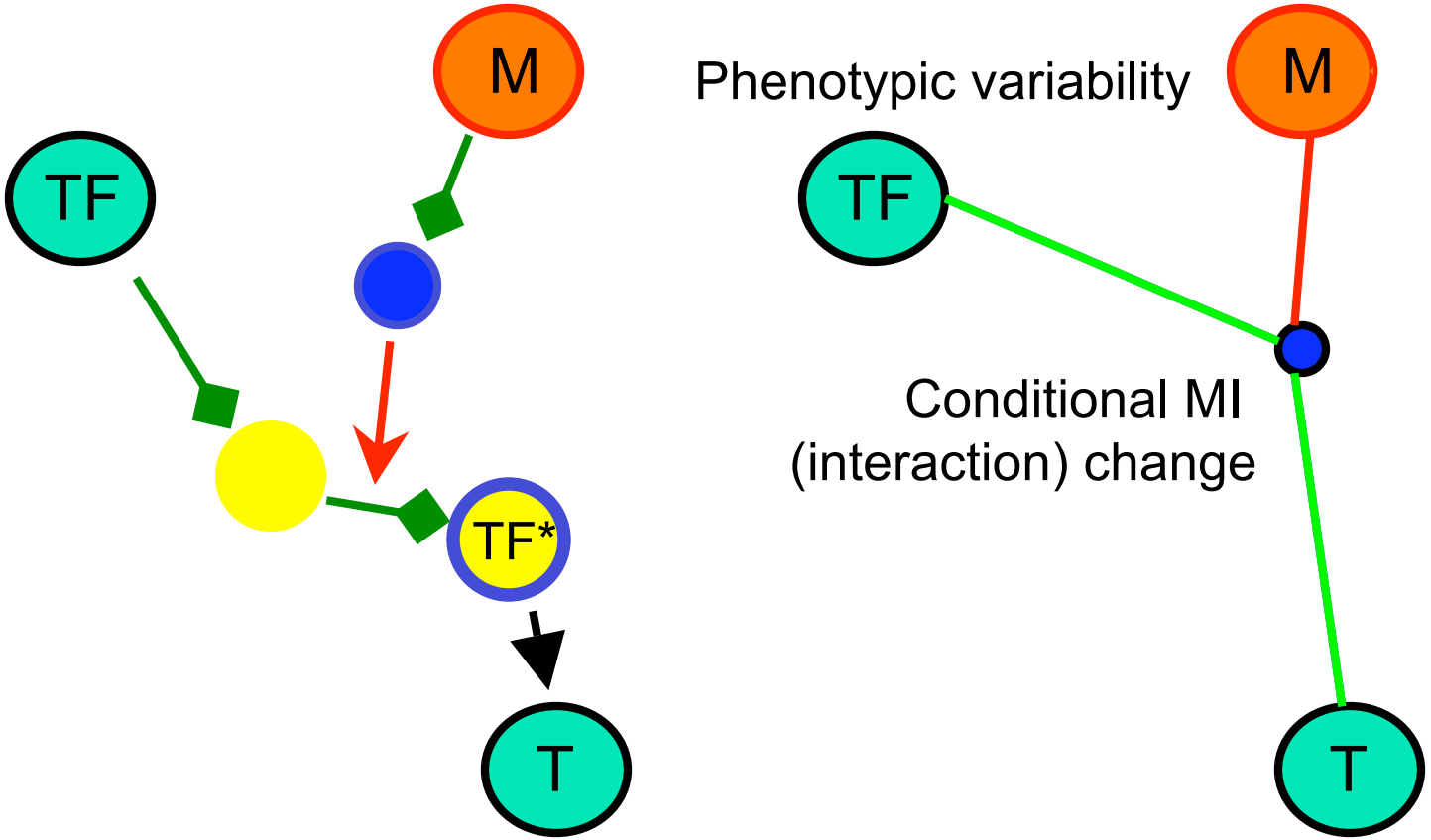
- Much of regulation in higher eukaryotes is post-transcriptional (e.g., splicing), and post-translational (e.g., phosphorylation, complex formation).
- Many mRNA (e.g., p53) constitutively expressed.

Can these be observed from mRNA expressions only?

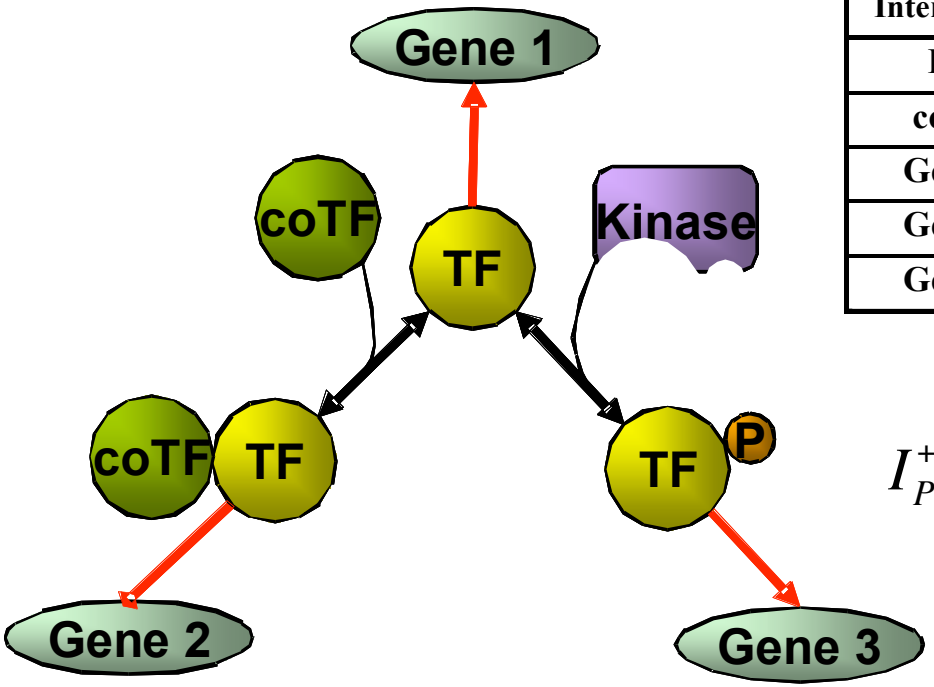
Solution:

Phenotypic and population variability (even in constitutively expressed genes) induces higher order dependencies between TFs, targets, and modulators.

Nontranscriptional modulation from mRNA expression



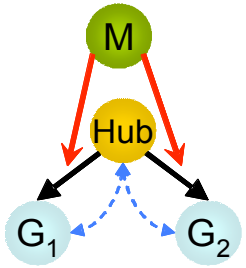
Numerical case study: Transistor modulation

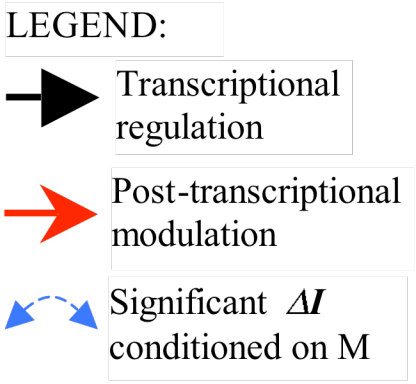


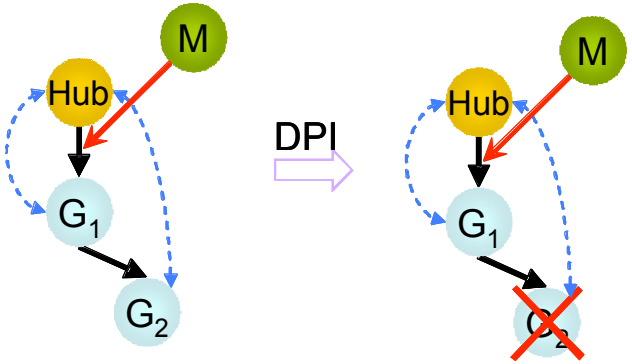
TF Interaction	I	I^{-}_{PK}	I^{+}_{PK}	ΔI_{PK}	I^{-}_{coTF}	I^{+}_{coTF}	ΔI_{coTF}
PK							
coTF							
Gene1	0.73	0.54	0.57		0.55	0.54	
Gene2						0.37	0.37
Gene3			0.35	0.34			

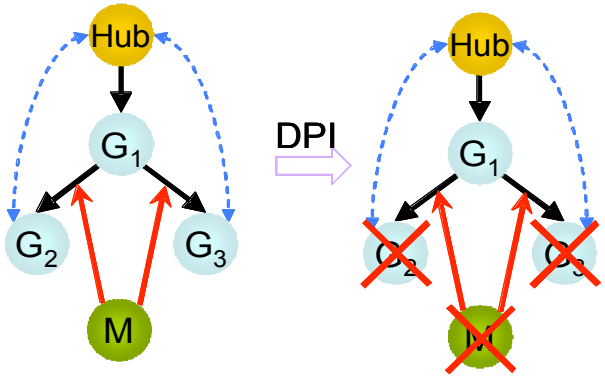
$$I^{+}_{PK}(\text{gene1}) = I(TF, \text{gene1} \mid PK \text{ high})$$

Enforcing irreducibility: ARACNE on a TF-hub

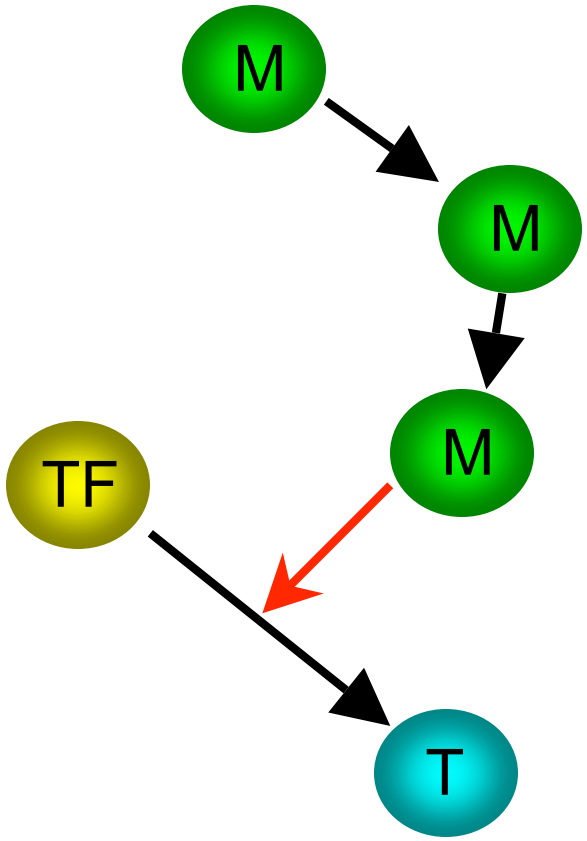








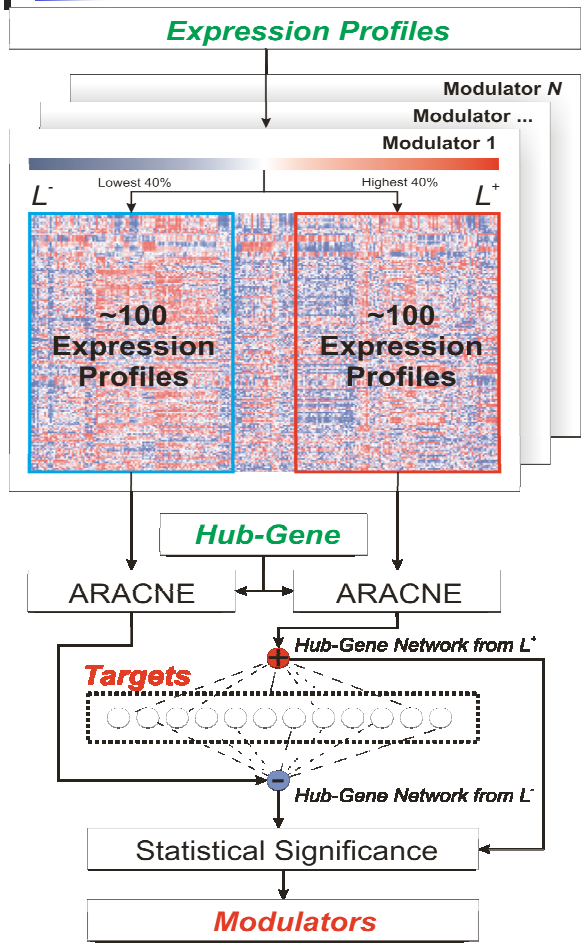
However:
No solution yet for…



Modulators are not irreducible.

Any suggestions?

Algorithm flowchart



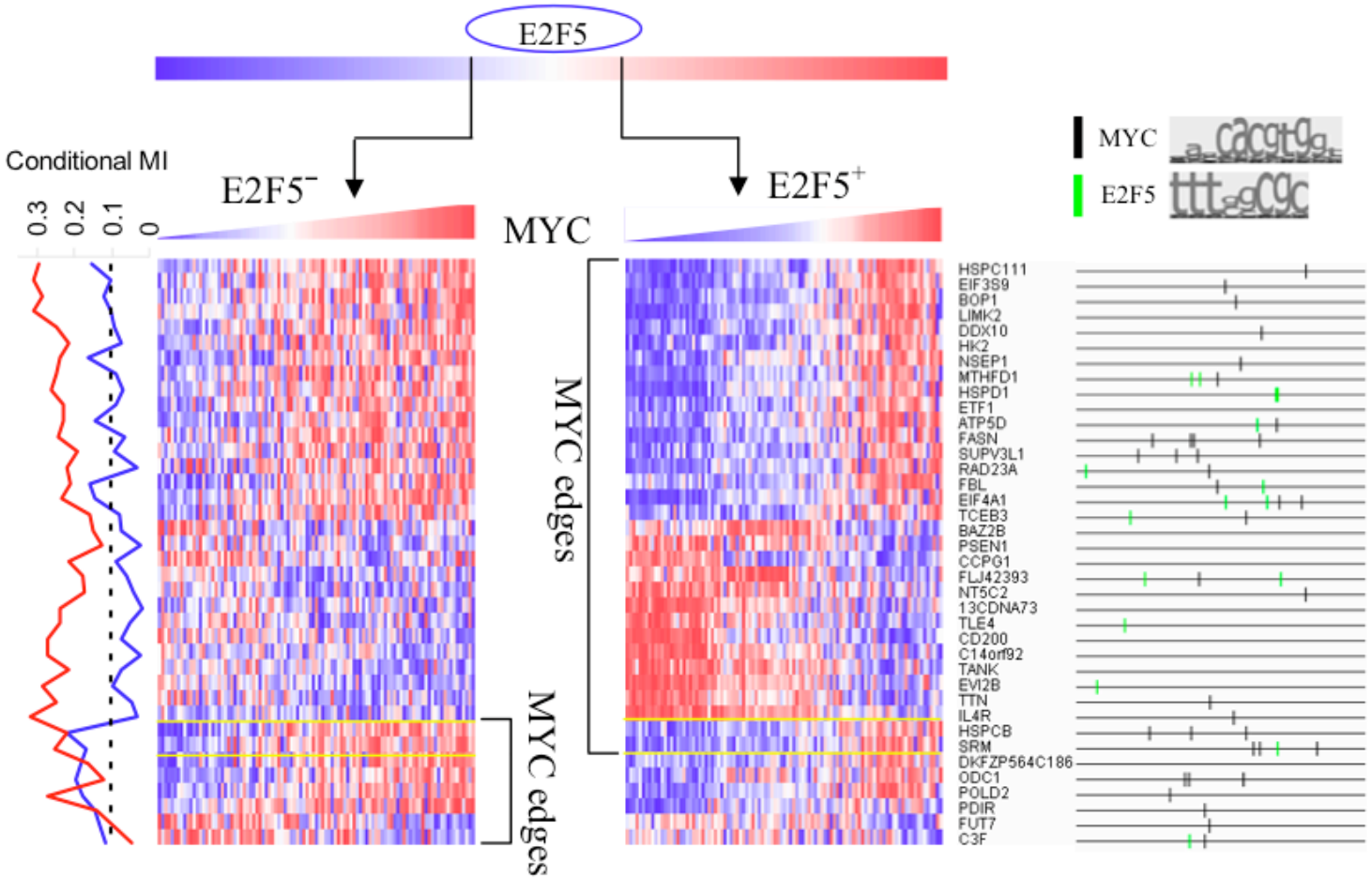
- Focus on a hub (c-MYC).
- Select modulators with σ> microarray noise (Tu et al., 2002) -- many signaling genes, constitutively expressed genes.
- Find modulators whose expression inflicts significant conditional MI changes for an ARACNE target in at least one conditional topology.
- No guarantee of modulator irreducibility.
- Guarantee of target irreducibility (after multiple hypothesis correction).

$$\Delta I(g_{TF}, g_t \mid g_m) = \left| I(g_{TF}, g_t \mid g_m^+) - I(g_{TF}, g_t \mid g_m^-) \right| > 0$$

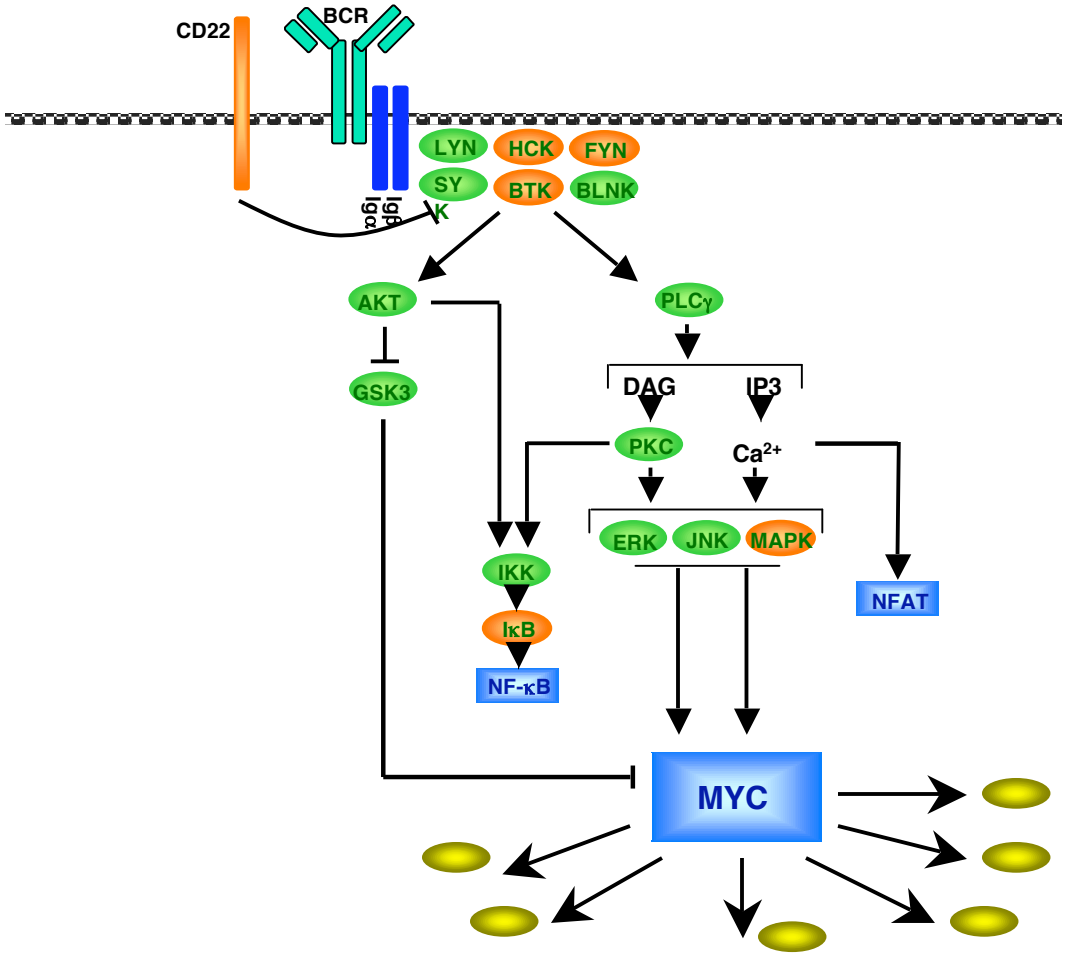
c-MYC modulators

- 1117 candidate modulators
- 100 modulators, 130 targets, 205 interactions
- GO enrichment of the modulator set: kinases, acyltransferases, TFs (all p<5%)
- Modulators in known MYC regulation pathways (e.g., BCR)
- TFs: 15/100, p=1e-6.
- 4/5 TF modulators (e.g., E2F5) with TRANSFAC signatures have binding sites in modulated targets promoter regions.
- Modulators with many (>=4) targets are not-specific (proteolisis, upstream signaling components, receptor signaling molecules).
- Modulators with few (1-2) effected targets are mostly co-TFs, interaction-specific.
- ~1/3 modulators are literature-validated.
- Biochemical validation of predictions in progress.

Example:
TF co-factor modulator



BCR pathway: Reducibility



predicted modulators

not in the candidate list

TF's not predicted

Protein complex

Targets