

Update of the ENCODE network

6/16/2011

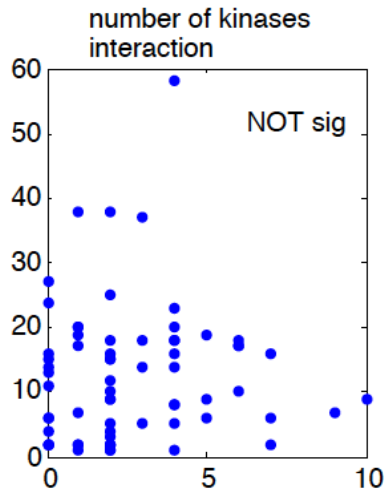
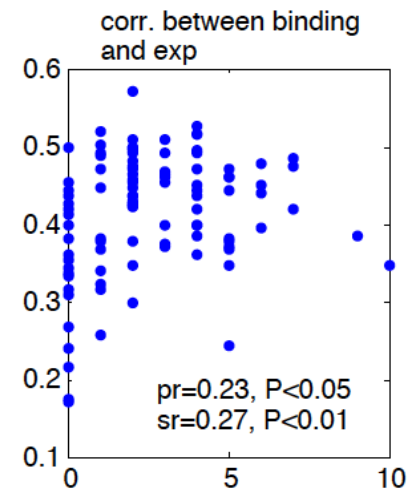
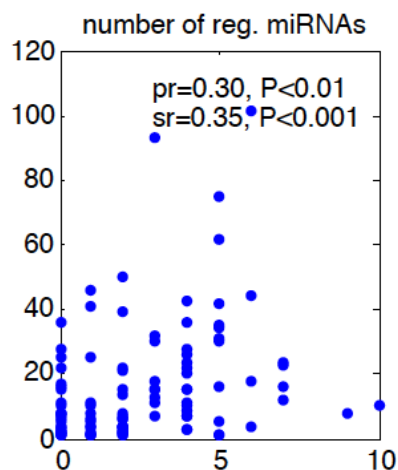
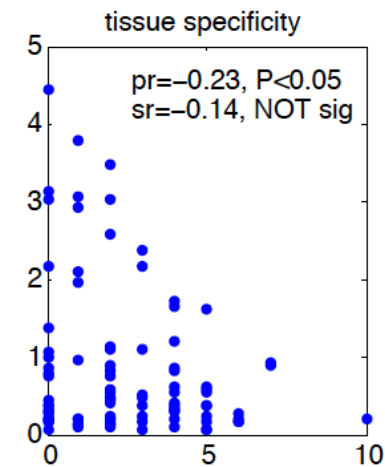
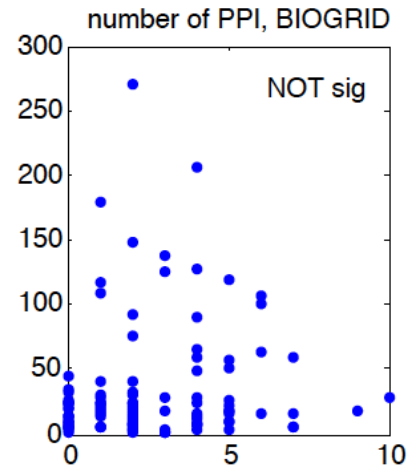
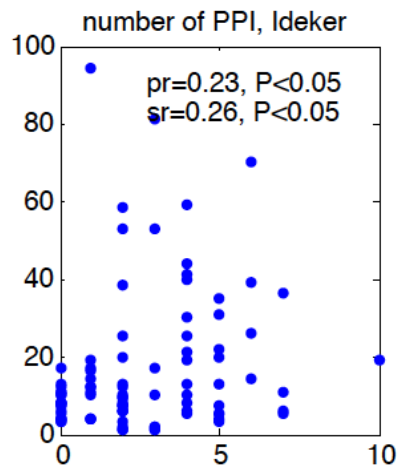
7/6/2011

7/28/2011

The Network

- ⊗ 428 ChIP-Seq tracks, with 119 unique transcription factors, in different cell lines or under different conditions
- ⊗ Target identification based on Chao's probabilistic framework, with q-value 0.01
- ⊗ 7 TFs are removed: CTCF, EP300, TAF1, TAF7, TBP, WRNIP1, XRCC4
- ⊗ TF network: 112 TFs, and 309 edges (compare with the previous version: 67 TFs, 1000 edges)
- ⊗ Properties of TFs: number of physical interacting partners (Ideker, BIOGRID), tissue specificity, number of regulating miRNAs, correlation of binding signals vs expression, number of kinase interactions, ... (structural properties: disorder etc, evolutionary ...)

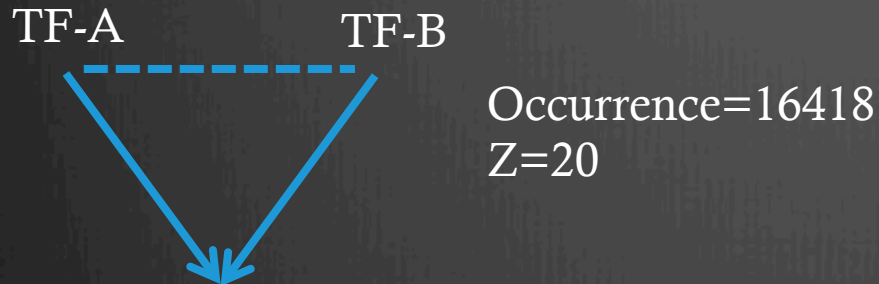
Correlating TF properties with network out-degree



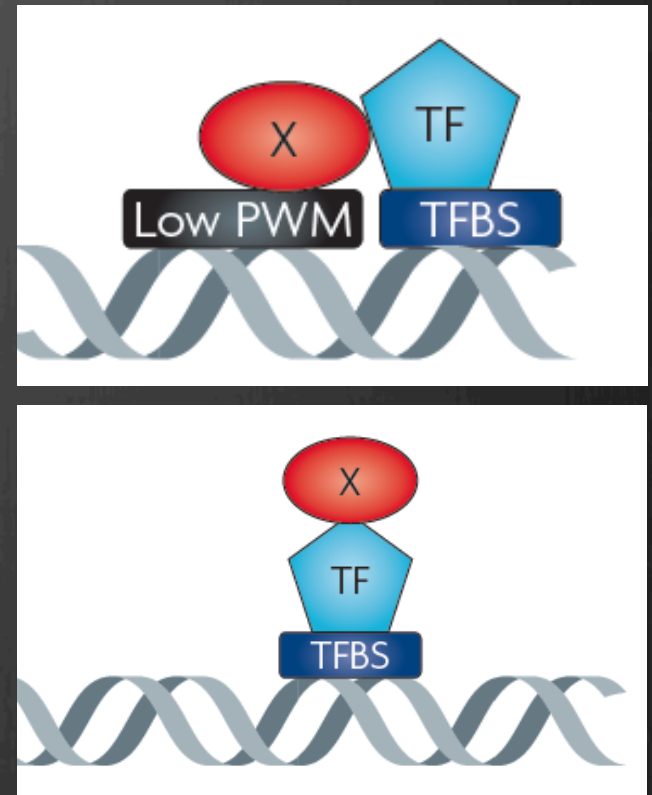
Remarks:

- ⊗ Previous version: K562, 67 TFs, 1000 edges, we didn't observe the correlations
- ⊗ NO correlation between TF properties and its in-degree (in-deg is not a genome-wide property)
- ⊗ Individual nodes -> pairwise correlation -> modules (motifs)
-> global (hierarchy)

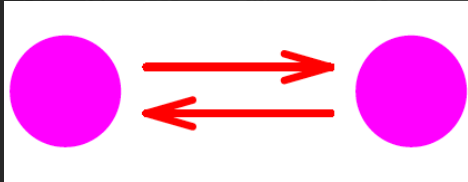
Integrate with PPI data



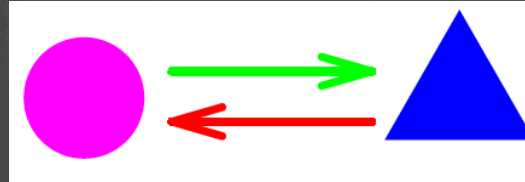
- ⊗ Look at individual motifs, distinguish the two scenarios
- ⊗ Using information content of the PWMs, identify master regulators, and tissue specific guys
- ⊗ The filtered network might do a better job for this analysis



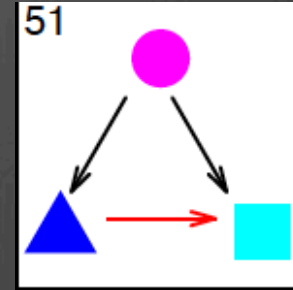
Enriched motifs



N=40, z=6.96

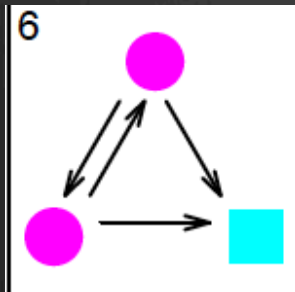


Not sig



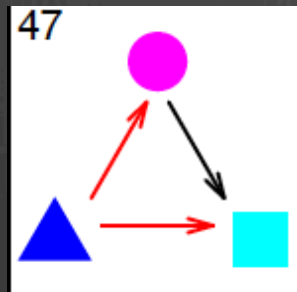
51

Not sig



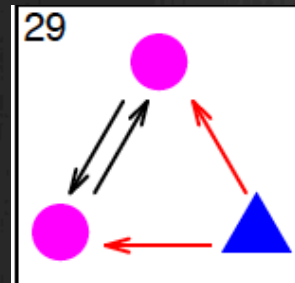
6

N=4203



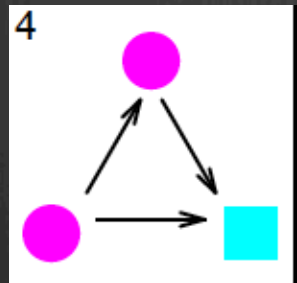
47

z=2.65



29

N=110

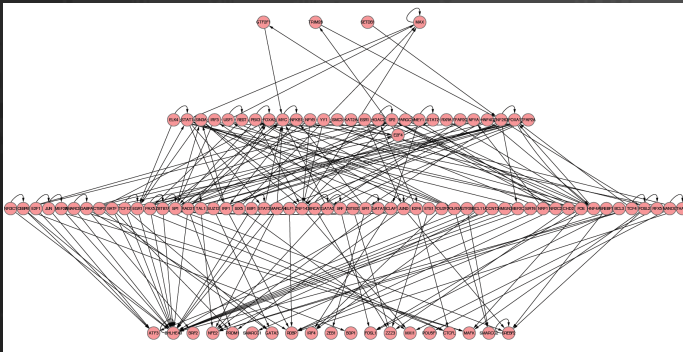


4

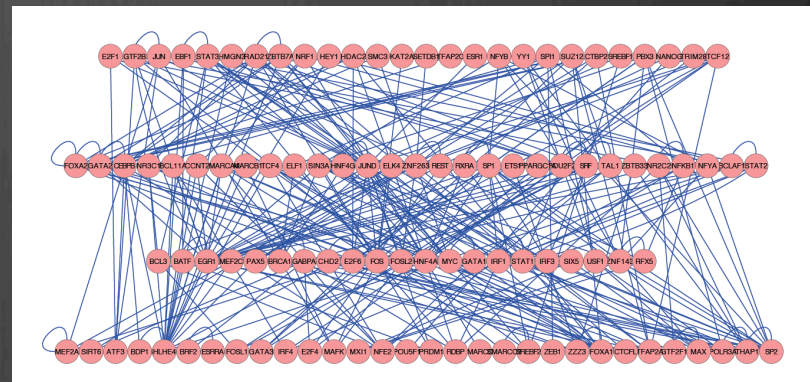
z=28

Zoo of Hierarchies

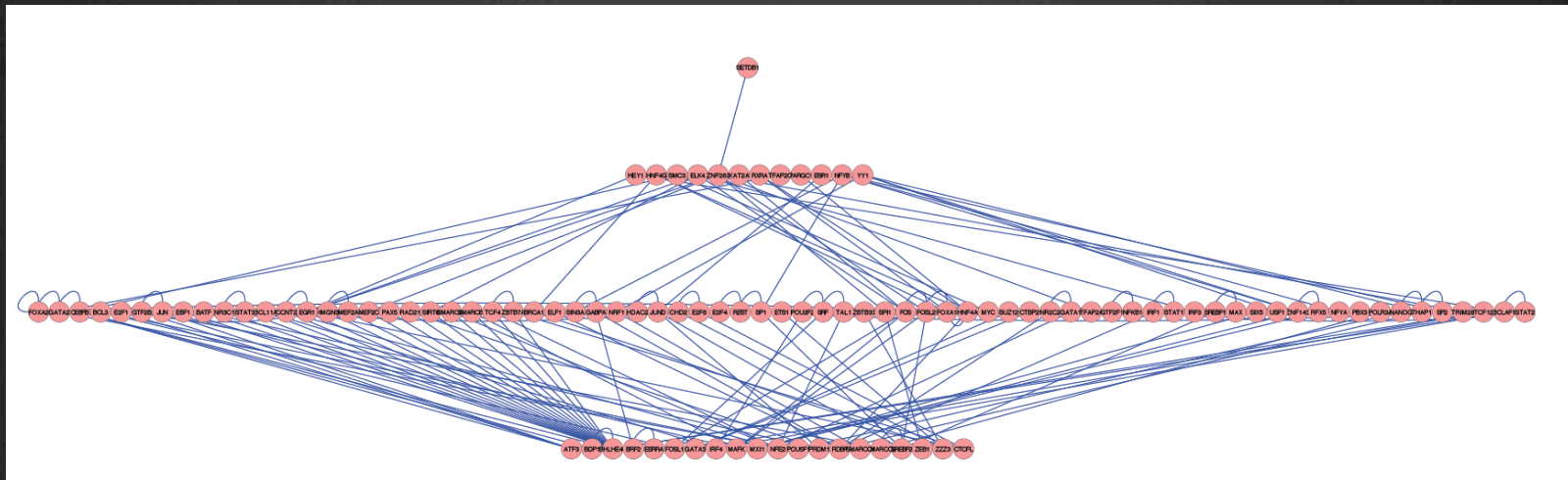
Leaf removal algorithm



Minimize feedback



Breadth first search



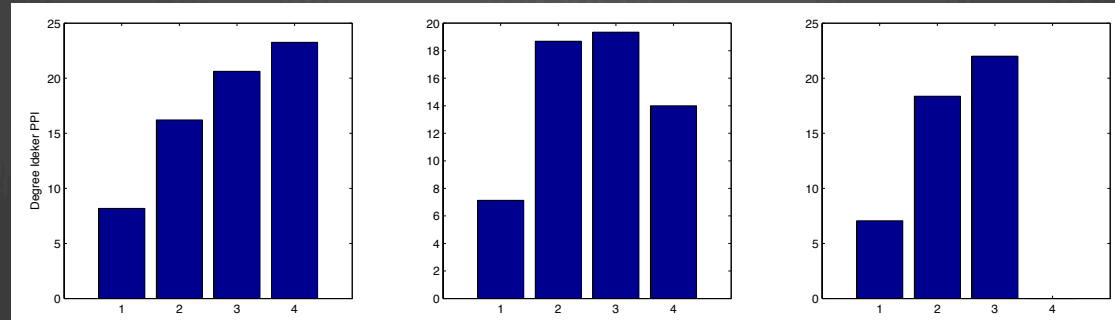
TF properties wrt levels

MF

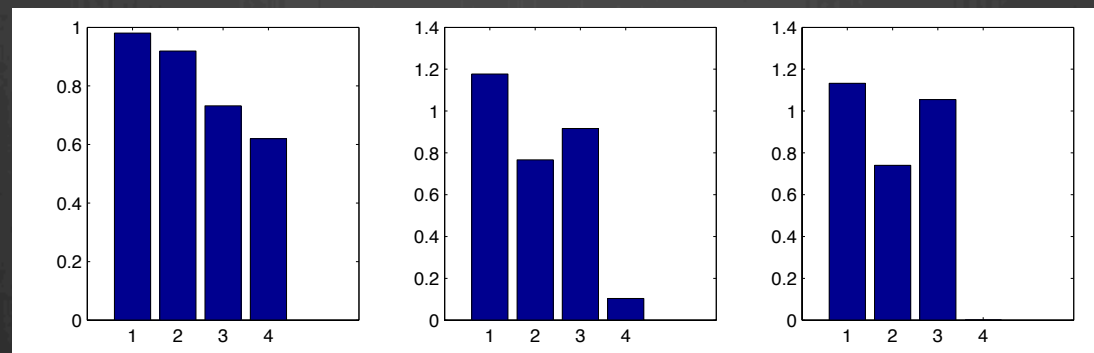
LR

BFS

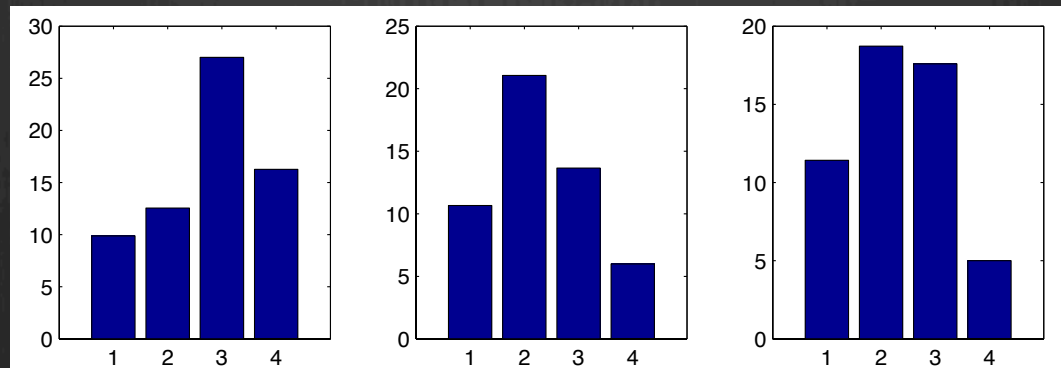
Number of PPI partners



Tissue specificity



Num. of miRNA regulators



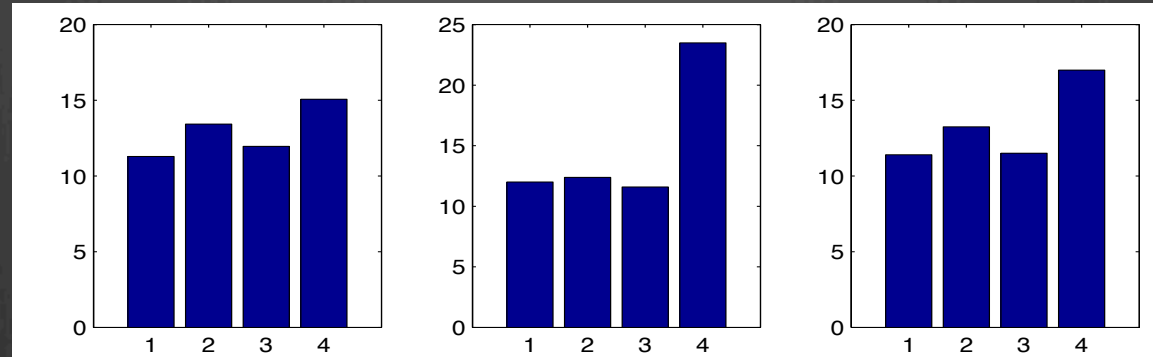
TF properties wrt levels

MF

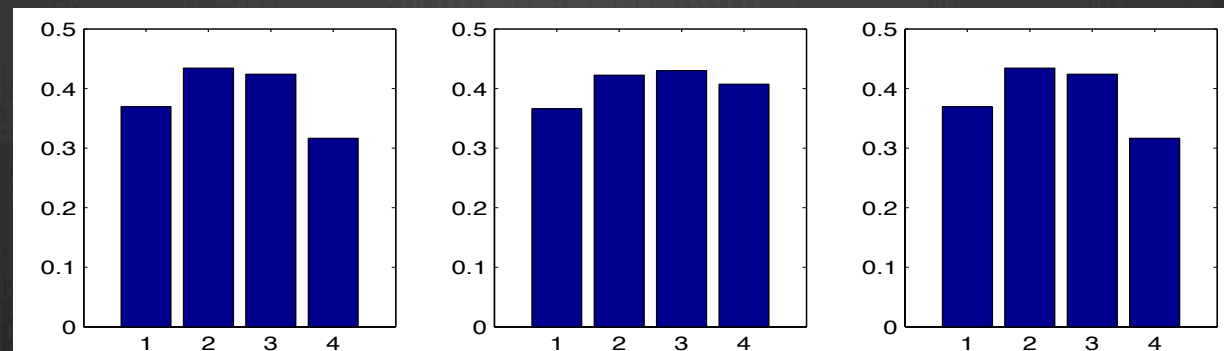
LR

BFS

Num of regulating kinases

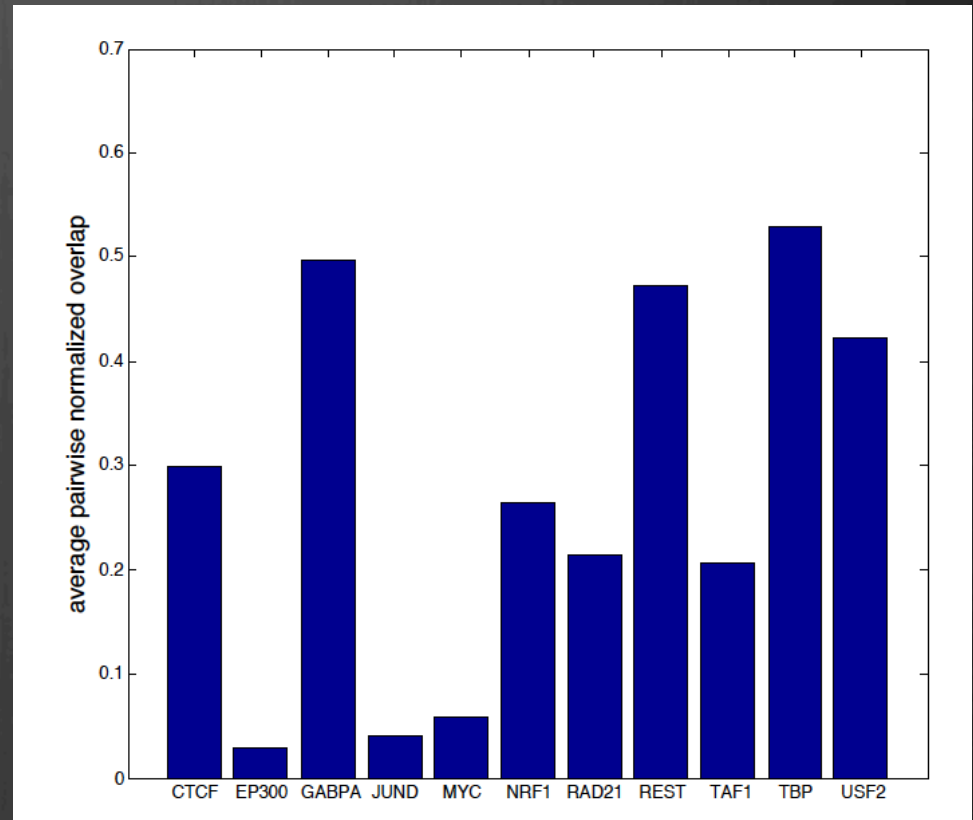


Corr. With exp



TF Dynamics

- ⊗ Only 11 TFs (CTCF, EP300, GABPA, JUND, MYC, NRF1, RAD2, REST, TAF1, TBP, USF2) with tracks in 5 major cell lines (Gm12878, H1hesc, Helas3, Hepg2, K562)
- ⊗ Quantify the average pairwise rewiring (intersection/union) of these TFs in the 5 cell lines



A brunch of things tried

- ⊗ Look at the targets of each TF, check how many of them are themselves TFs. There are TFs that tend to regulate other TFs, or tend to regulate non-TFs. Apparently, I cannot find significant differences
- ⊗ Many TFs in our network with zero kout actually regulate many other TFs that are not present in the network.
- ⊗ Significant correlation with out-degree is lost if we use TFs for K562 only

A bit more on the hierarchy

Statistics from the hierarchy

Level 3 ————— 43

Level 2 ————— 34

Level 1 ————— 42

level- level	3->2	3->1	2->1	1->2	1->3	2->3	3->3	2->2	1->1
number of edges	78	88	79	8	8	15	9+6	13+11	20+11

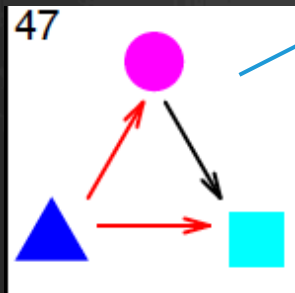
245 forward edges

31 backward edges

42 within-level edges +
28 auto-regulation

Mapping motifs to the hierarchy

- Auto-regulation is depleted at the top layer: 6 auto regulator out of 43 TFs (28 out of 119 in the whole system)
- MIR-TF-TF (59 in the network)



from which level to which level?

3->2	14
3->1	17
2->1	15
1->2	4
1->3	0
2->3	2

- TF-TF-FFL (123 of them)