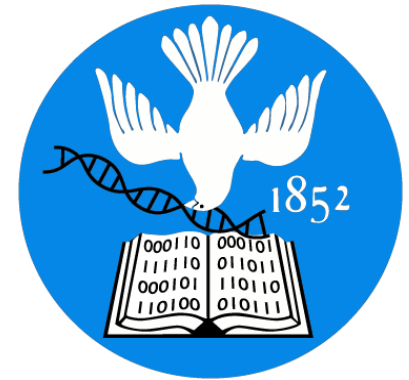




Tufts
UNIVERSITY



Regulatory Hub Identification in Yeast

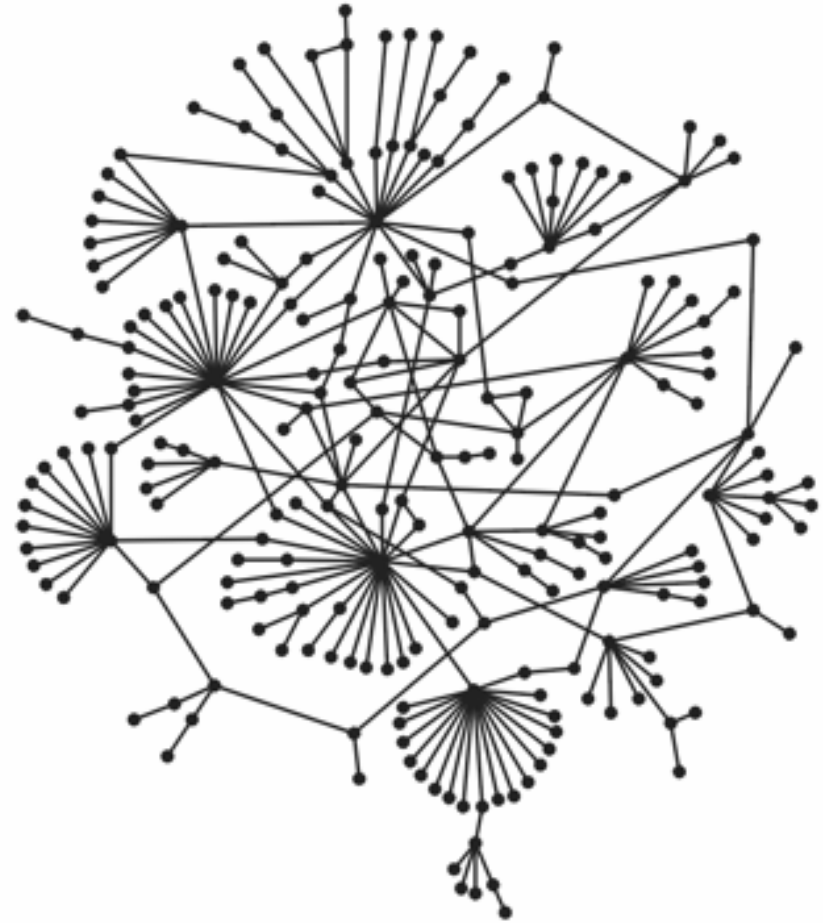
Andrew Fox, Donna K. Slonim

Tufts University, Dept. of Computer Science, Bioinformatics and Computational Biology

A Focus on Network Hubs

A major goal in interpreting whole-genome datasets such as gene expression data is finding key proteins that control the biological functions relevant to a given experiment.

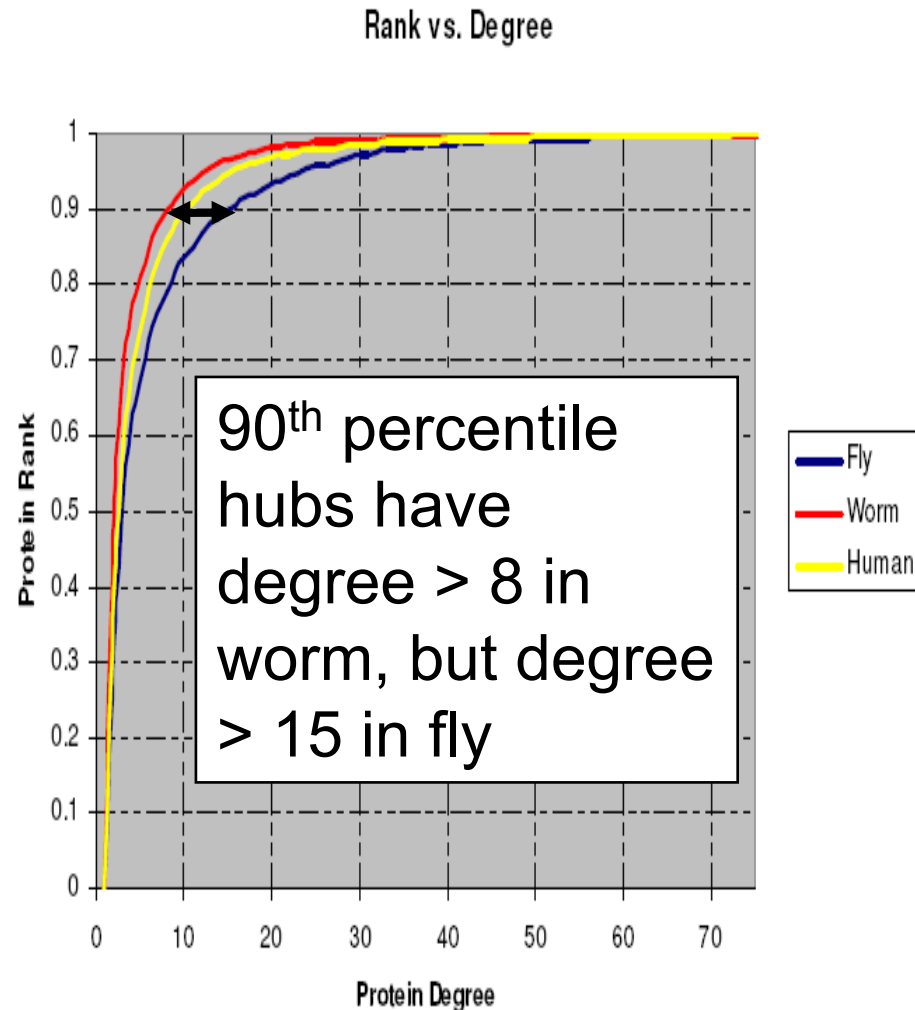
We focus our study on network hubs as these proteins are critical to the structure of many biological networks.



Many biological networks are scale-free, having many leaves and few hubs.

A Consistent Hub Definition Across Species

- We normalize out species-specific distribution properties by ranking:
- $\text{Rank}(\mathbf{x}) = [\% \text{ of proteins in the same species that have a lower degree than } \mathbf{x}]$
- Hub Threshold: T
- $\text{Hub}_T(\mathbf{p}) \equiv (\text{Rank}(\mathbf{p}) \geq T)$



Degree Conservation

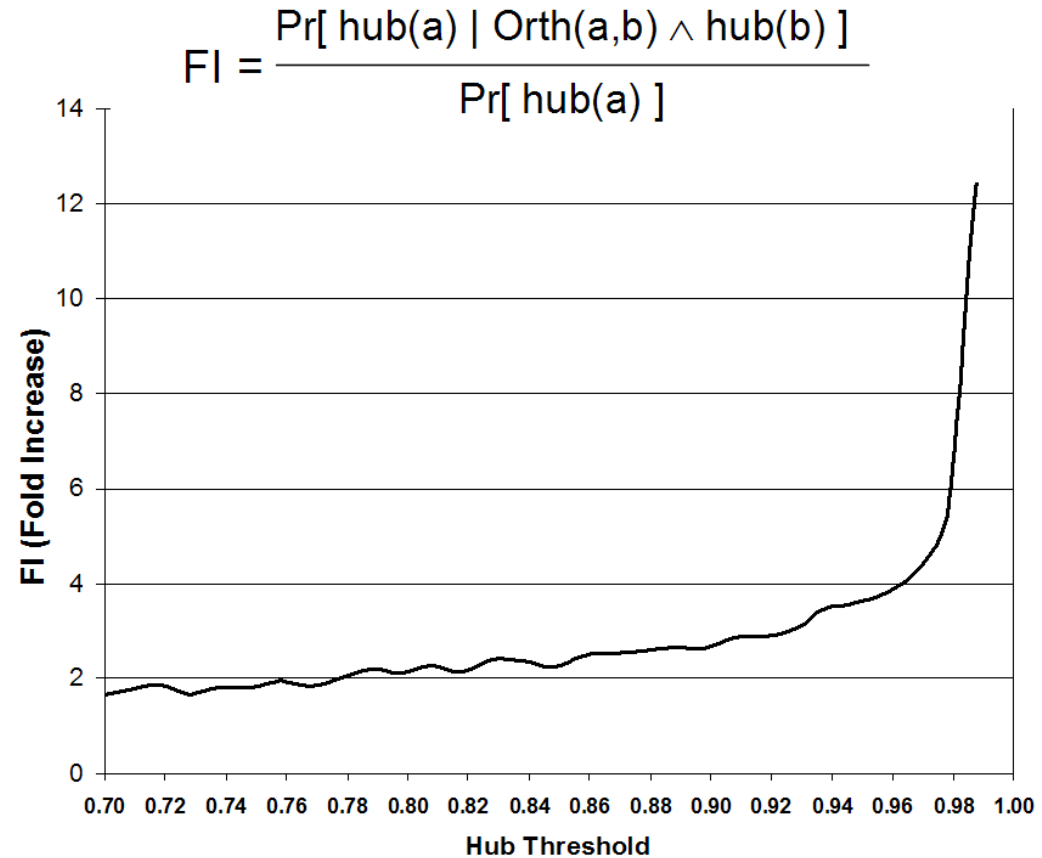
Functional network modules are generally not well conserved.^[1,2] However we intuitively expect high-degree proteins of regulatory and/or functional importance to maintain their high degree of connectivity as they evolve.

Definition: DegreeConserved_T(a,b)

$$= \text{Orthologs}(a,b) \wedge \text{Hub}_T(a) \wedge \text{Hub}_T(b)$$

Hubs are Preferentially Degree-Conserved

- Proteins have a significantly higher probability of being a hub if their ortholog is a hub
- *Non-linearity*: The highest degree hubs show a significantly higher level of conservation
- 4-species results consistent with 2-species results
- Conservation is preferential for hubs but not universal among hubs

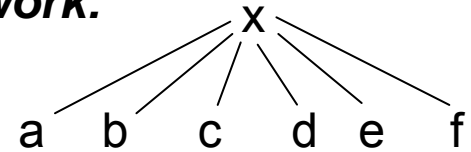


Data Sources: [7,8,9,10]

Which Hubs are Degree-Conserved?

- Han^[3] used average correlation of hub-target expression to detect distinct hub types:

PPI Network:



	Process 1	Process 2	Process 3
<ul style="list-style-type: none"> <u>Party Hubs</u> always bind to a large number of their targets (e.g. protein complexes) 	<pre> graph TD X((X)) --- a((a)) X --- b((b)) X --- c((c)) X --- d((d)) X --- e((e)) X --- f((f)) </pre>	<pre> graph TD X((X)) --- a((a)) X --- b((b)) X --- c((c)) X --- d((d)) X --- e((e)) X --- f((f)) </pre>	<pre> graph TD X((X)) --- a((a)) X --- b((b)) X --- c((c)) X --- e((e)) X --- f((f)) </pre>
<ul style="list-style-type: none"> <u>Date Hubs</u> bind to various smaller target <i>subsets</i> under different cell conditions / processes 	<pre> graph TD X((X)) --- a((a)) X --- e((e)) X --- f((f)) </pre>	<pre> graph TD X((X)) --- d((d)) X --- e((e)) </pre>	<pre> graph TD X((X)) --- a((a)) X --- b((b)) X --- c((c)) </pre>

One hypothesis is that Date hubs, being intermodular^[2] and transient, direct processes and represent the majority of the conserved hubs. Experimentally, however, we observe no *clear* difference in the level of degree-conservation between Date and Party hubs.

Modeling Regulatory Hubs in Yeast

Average Pearson Correlation Coefficient:

- $\text{AvgPCC}(x) = \text{avg}_i \text{PCC}(x, g_i)$
 $g_i \in G = \{ \text{Interactors of } x \}$

		GE samples									
		s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
Genes	x										
	g1										
	g2										
	g3										
	g4										
	g5										
	g6										

We created a dataset of yeast hubs ($T=0.8$) and a training set of Regulatory hubs (REG) including Transcription Factors from the online TransFac database^[4] with extra regulatory hubs due to Han^[3]. We plotted these hubs by AvgPCC of their expression with their targets and found that:

- Regulatory Hubs were not linearly separable
- Date/Party Hubs were also not linearly separable
- Party hubs have a higher correlation measure than Date hubs, so a bimodal avgPCC distribution allows separation. But what if the distribution of AvgPCCs is not bimodal?

A New Statistical Measure Helps Identify Regulatory Hubs

Interactor Coexpression Variance (IC-Var):

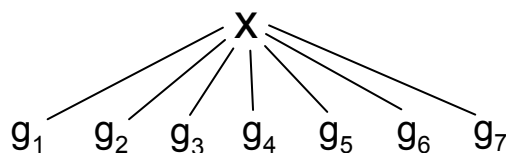
- $IC-Var(x) = \text{var}_i (\text{\#of coexpressed interactors of } x)$
 $i \in \{ s \mid x \text{ is expressed in sample } s \}$
- Party hubs should have lower IC-Variance but higher AvgPCC

Array Samples

	s1	s2	s3	s4	s5	s6	s7	s8	s9
x	1	1	1	1	1	1	1	1	1
g1	1	1	1	0	0	0	0	0	0
g2	1	1	1	0	0	0	0	0	0
g3	0	0	0	1	1	0	0	0	0
g4	0	0	0	1	1	0	0	0	0
g5	0	0	0	1	1	0	0	0	0
g6	0	0	0	1	1	0	0	0	0
g7	0	0	0	0	0	1	1	1	1
	2	2	2	4	4	1	1	1	1

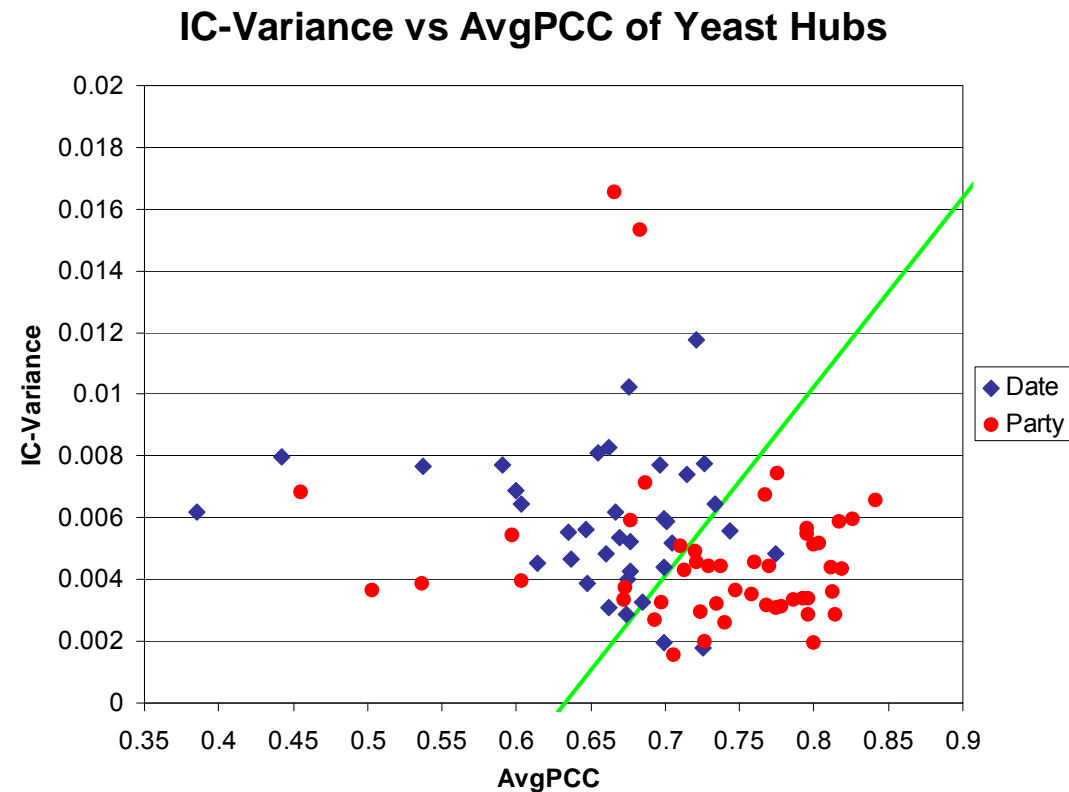
coexpressed interactors

PPI Network:



IC-Variance Helps To Separate Party/Date Hubs

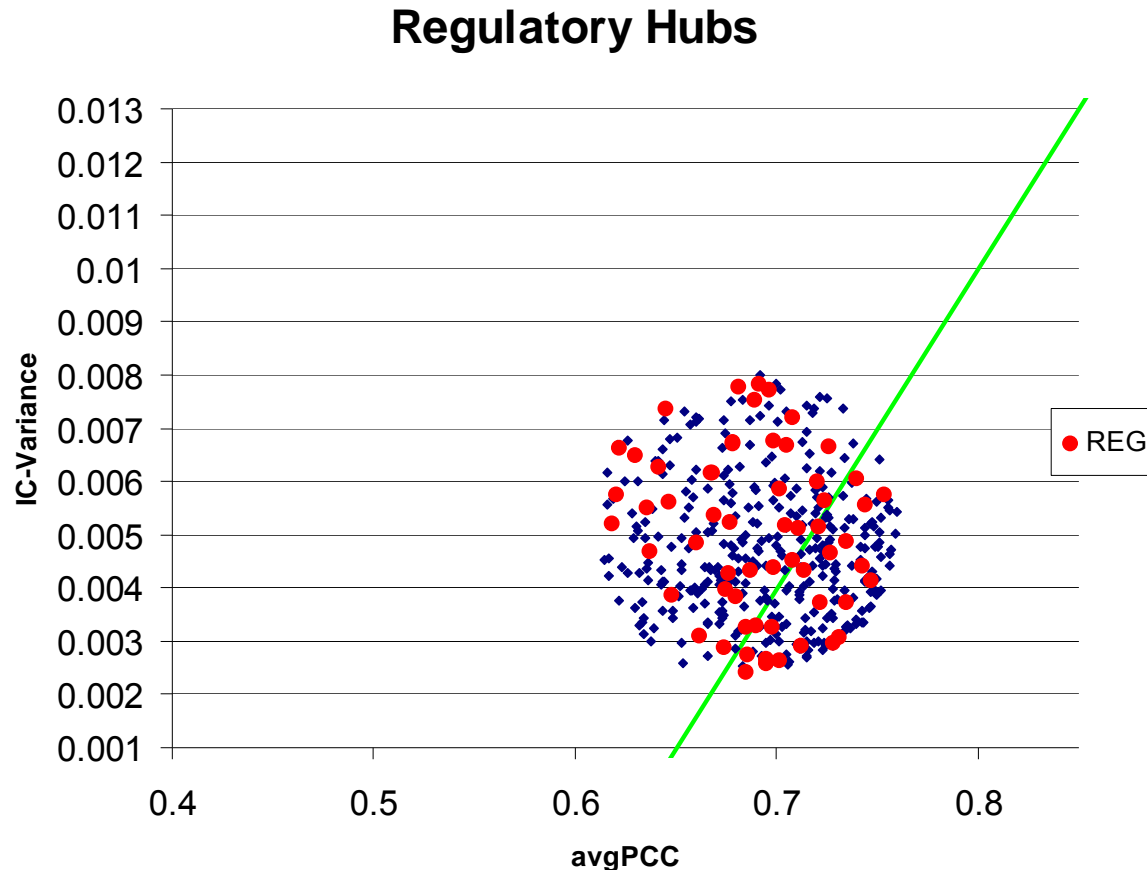
- Yeast hubs are more naturally separated in PCC-Variance space
- Regulatory hubs are **not** well separated, in fact most are close to the separating plane. The majority are Date hubs.



Predicting Yeast Regulatory Hubs

Define Regulatory Cluster:

- Centroid = median(REG)
- Maximize number of REG hubs in cluster while minimizing cluster area
- Under this model blue points in the figure are predicted yeast regulatory hubs



Regulatory Hubs cluster in PCC-Variance space. A significant majority are Date hubs ($p < 5e-3$, Fisher's Exact Test)

Predicted Reg. Hubs are Enriched for Regulatory Functions

We tested for functional enrichment of the non-REG hubs in the cluster model using NCBI's DAVID webtool: (vs Yeast Genome)

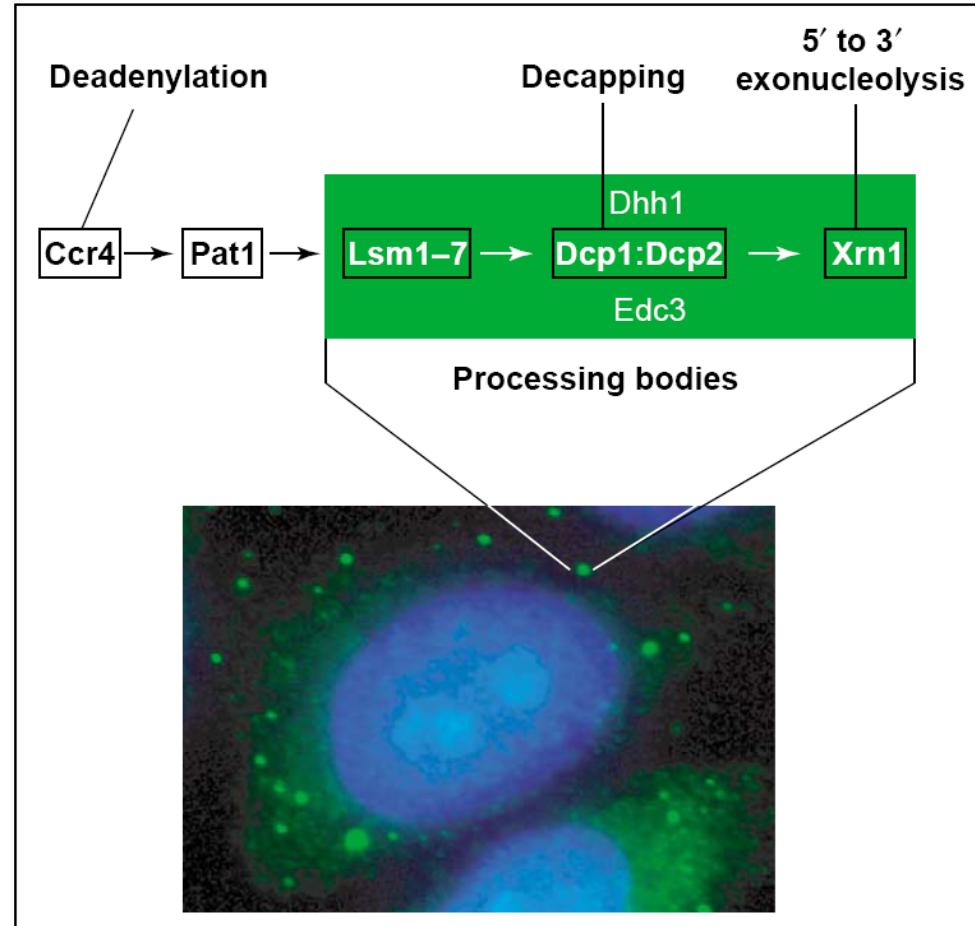
Regulatory Functions

	<u>p-val(Benjamini)</u>
1. chromatin remodeling	7.6e-10
2. nucleotide-binding	1.3e-08
3. hydrolase activity	2.8e-07
4. pyrophosphatase activity	3.0e-07
5. cellular macromolecule catabolism	3.3e-07
6. sh3 kinase domain	8.4e-06
7. response to endogenous stimulus	5.4e-04
8. response to DNA damage stimulus	8.6e-04
9. unfolded protein binding	1.4e-03
10. RNA localization	3.2e-03
11. protein kinase CK2 complex	5.5e-03
12. RNA catabolism	5.8e-03
13. nuclear mRNA splicing, via spliceosome	7.7e-03
14. Transcriptional Regulation from RNA polymerase I *	1.4e-02
15. regulation of protein biosynthesis	1.8e-02

* An additional 21 Transcription Factors not included in the Reg. training set were found using the cluster model

Regulating mRNA Degradation

- The process of Deadenylation, Decapping and Exonucleolysis regulates cellular mRNA degradation and hence contributes to the regulation of protein biosynthesis [5]
- Ccr4, Dcp1, Dcp2, Edc3, Xrn1 and 5 related Lsm genes were present in the regulatory cluster we identified.
- Also present was Sho1 which is a vital controller of the important Map Kinase signalling pathway. MAPK is involved in the action of most nonnuclear oncogenes[6]



Source: Fillman & Lykke-Andersen (COiCB, 2005) [5]

Yeast Regulatory Hubs are Degree-Conserved

- Further analysis of the regulatory cluster hubs reveals that they are preferentially degree-conserved compared to non-cluster hubs ($p < 2e-3$)*

Degree Conserved?	inCluster	outCluster
<i>Yes</i>	105	128
<i>No</i>	280	602
<i>Total</i>	385	730

* Fisher's Exact Test

Conclusions

- It is possible to identify a significantly large subset of regulatory hubs in yeast through statistical analysis of gene expression data and high throughput PPI data
- A significant majority of Regulatory Hubs are Date hubs
- Regulatory hubs are significantly better degree-conserved than non-regulatory hubs. All hubs are preferentially degree-conserved compared to non-hubs.

References

- [1] KR Brown & I Jurisica 2007, 'Unequal evolutionary conservation of human protein interactions in interologous networks', *Genome Biology*, **8**:R95,
- [2] Z Wang & J Zhang 2007, 'In Search of the Biological Significance of Modular Structures in Protein Networks', *PLoS Computational Biology* Vol. 3, No. 6, p1011-1021
- [3] J. Han et al 2004, 'Evidence for dynamically organized modularity in the yeast protein-protein interaction network', *Nature*, vol.430, p88-93.
- [4] V Matys et al 2003, 'TRANSFAC: transcriptional regulation, from patterns to profiles', *Nucleic Acids Res.*, 31, p374-378.
- [5] C Fillman & J Lykke-Andersen 2005, 'RNA decapping inside and outside of processing bodies', *Current Opinion in Cell Biology*, vol. 17, p326-331.
- [6] M Kanehisa et al 2006, 'From genomics to chemical genomics: new developments in KEGG', *Nucleic Acids Res.* 34, D354-357
- [7] L Giot et al 2003, 'A Protein Interaction Map of *Drosophila melanogaster*', *Science* **302**, p1727
- [8] S Li et al 2004, 'A Map of the Interactome Network of the Metazoan *C. elegans*', *Science* **303**, p540
- [9] *Human Protein Reference Database*, www.hprd.org
- [10] U Guldener et al 2006, "MPact: the MIPS protein interaction resource on yeast.", *Nucleic Acids Res.* 2006 Jan 1;34(Database issue):D436-41.