



# Using Networks to Integrate Omic and Semantic Data: Towards Understanding Protein Function on a Genome Scale

Biomarker Data Analysis  
2008.10.01, 9:25-9:50

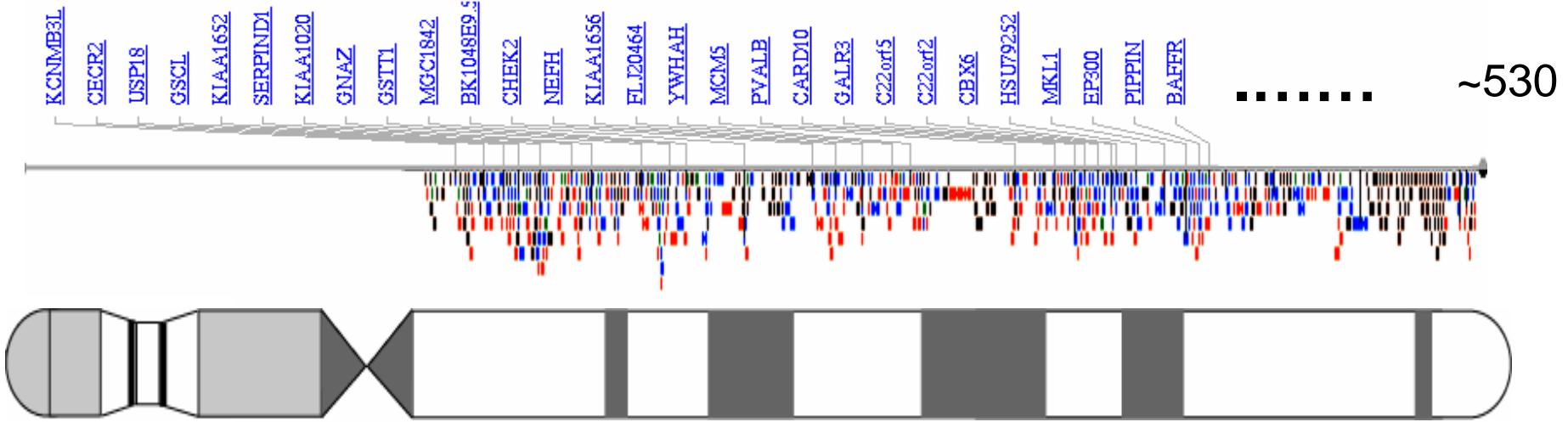
Mark B Gerstein  
Yale (Comp. Bio. & Bioinformatics)

**Slides downloadable from  
[Lectures.GersteinLab.org](http://Lectures.GersteinLab.org)**

(Please read permissions statement.)  
Paper references mostly from [Papers.GersteinLab.org](http://Papers.GersteinLab.org).  
See [streams.gerstein.info](http://streams.gerstein.info) on photos & images

(Very Short Nets talk: predess + SANDY [!IBIOMARKER])

# The problem: Grappling with Function on a Genome Scale?



- 250 of ~530 originally characterized on chr. 22  
[Dunham et al. Nature (1999)]
- >25K Proteins in Entire Human Genome (with alt. splicing)

# Traditional single molecule way to integrate evidence & describe function

EF2\_YEAST

**Descriptive Name:**  
Elongation Factor 2

**Lots of references**  
to papers

**Summary sentence describing function:**  
This protein promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome.

The screenshot shows a Microsoft Internet Explorer window displaying the UniProt Protein Viewer for the entry EF2\_YEAST. The page includes the UniProt logo and navigation links like Home, About UniProt, Getting Started, Searches/Tools, Databases, and Support/Documentation. The main content area displays general information, protein description, references, and comments. In the 'General information' section, it lists the entry name as EF2\_YEAST, primary accession number P32324, and various dates of release and modification. The 'Protein description' section identifies it as Elongation factor 2 with synonym EF-2. The 'References' section contains one entry with details about the nucleotide sequence and its publication. The 'Comments' section provides a summary sentence describing the protein's function: "This protein promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome." Other sections like Sequence and Domains are partially visible at the bottom.

# Some obvious issues in scaling single molecule definition to a genomic scale

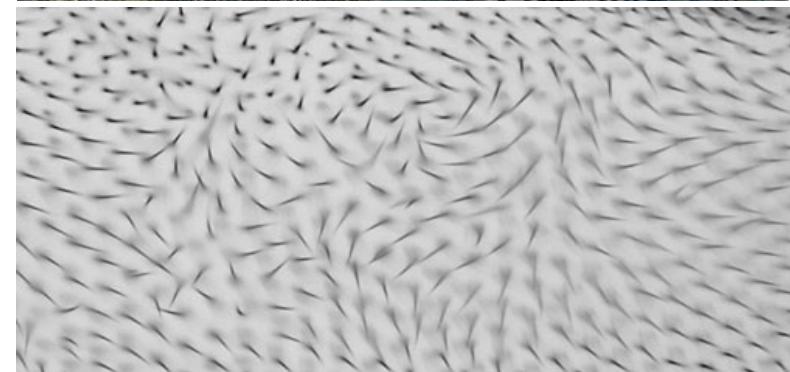
- Fundamental complexities
  - ◊ Often >2 proteins/function
  - ◊ Multi-functionality:  
2 functions/protein
  - ◊ Role Conflation:  
molecular, cellular, phenotypic

# Some obvious issues in scaling single molecule definition to a genomic scale

- Fundamental complexities
  - ◊ Often >2 proteins/function
  - ◊ Multi-functionality:  
2 functions/protein
  - ◊ Role Conflation:  
molecular, cellular, phenotypic
- Fun terms... but do they scale?....
  - ◊ **Starry night** (P Adler, '94)

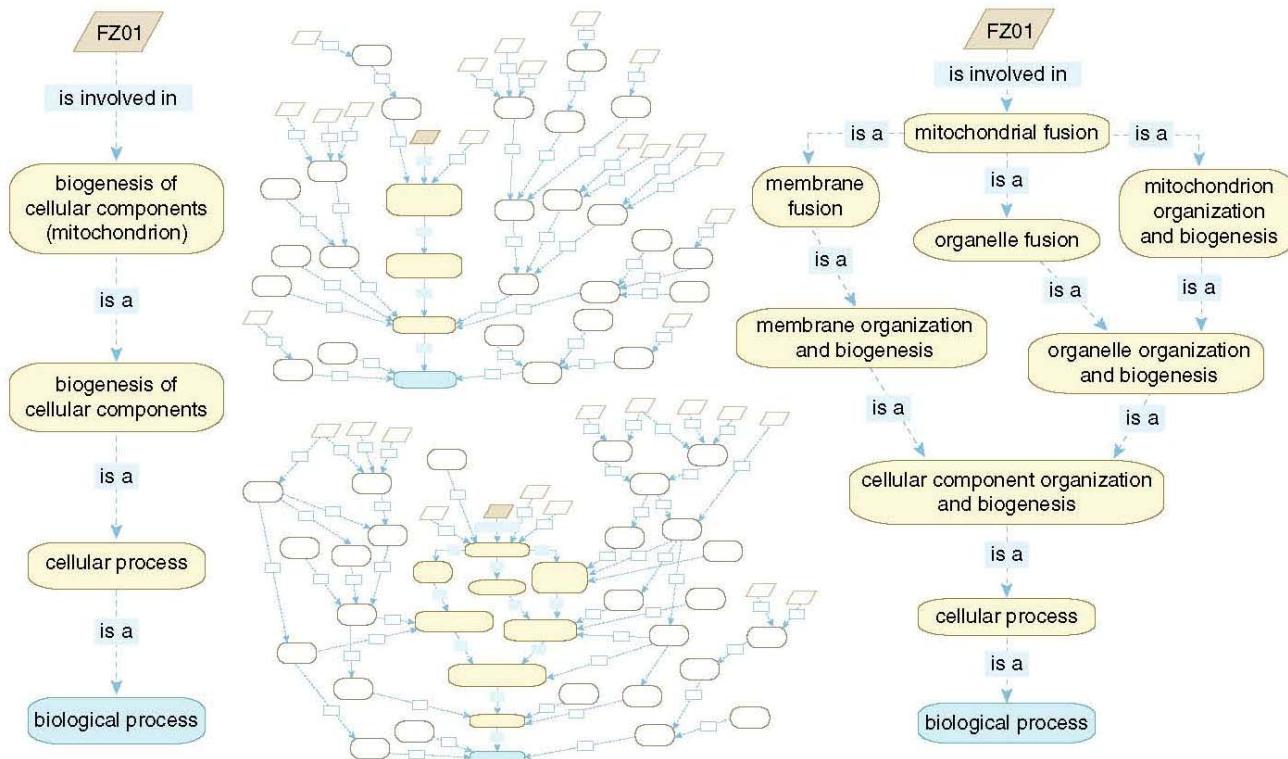
[tinman.nikunnakki.info](http://tinman.nikunnakki.info)

literature	pop culture	cars and driving
amontillado	sonic hedgehog	Sunday driver
luciferase	tribbles	limo
thor	pokemon	18wheeler
tigger	kojak	long island expressway
ariadne	groucho	
malvolio	maggie	
tinman	glass-bottom boat	
food	swiss cheese	people
chardonnay	moonshine	scott of the antarctic
		evander
		van gogh
		cleopatra



[Seringhaus et al. *Genome Biology* (2008)]

# Hierarchies & DAGs of controlled-vocab terms but still have issues...



**MIPS (Mewes et al.)**

**GO (Ashburner et al.)**

[Seringhaus & Gerstein, Am. Sci. '08]

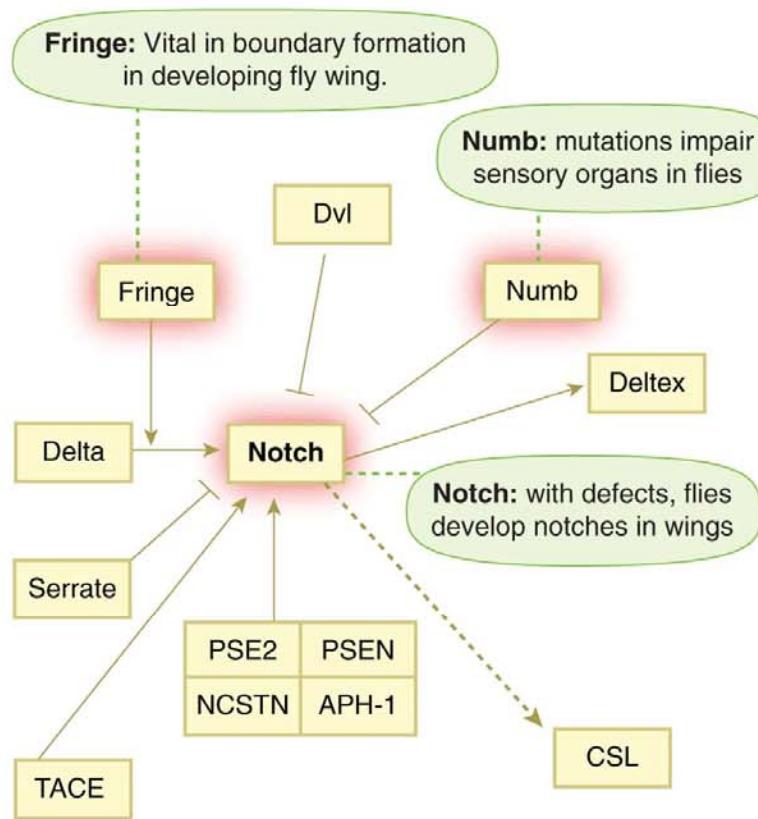
# Towards Developing Standardized Descriptions of Function

- Subjecting each gene to standardized expt. and cataloging effect
  - ◊ KOs of each gene in a variety of std. conditions => phenotypes
  - ◊ Std. binding expts for each gene (e.g. prot. chip)
- Function as a vector

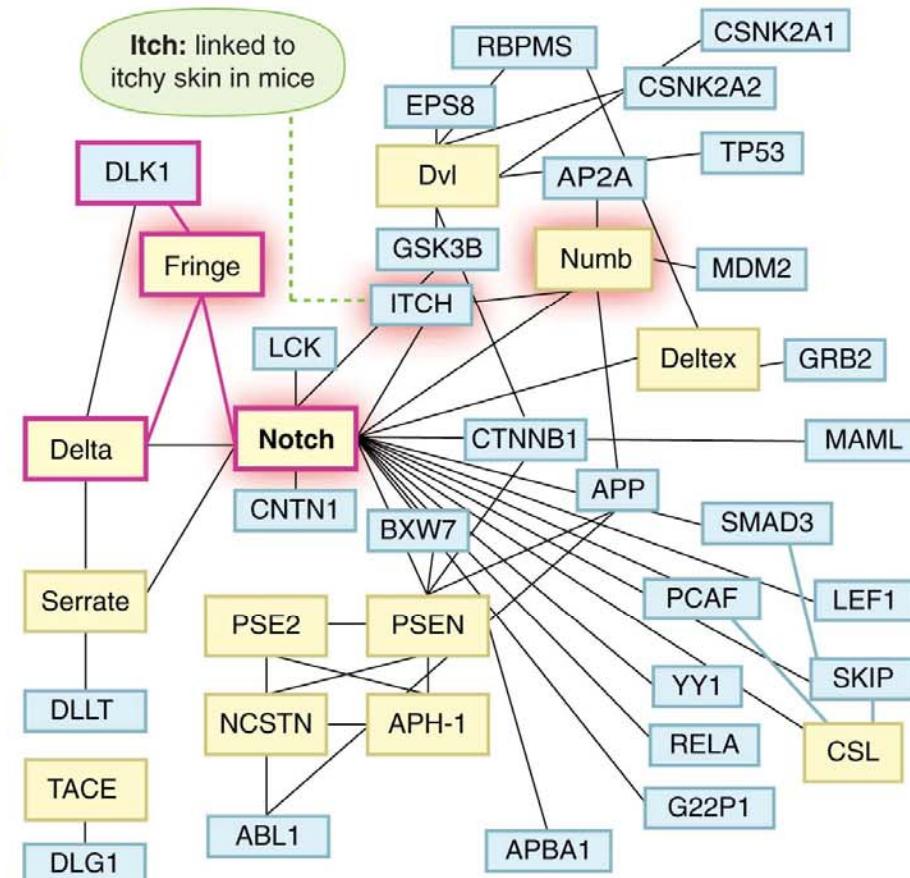
	nucleic acids	small molecules	proteins								
	DNA	RNA	ATP	Metal	CoA	NAD	.....	G protein	CDC28	Calmodulin	.....
protein 1	1.0	0	0	0	0	0	.....	0	0	0	.....
protein 2	0	0.9	0	0	0	0	.....	0	0	0	.....
protein 3	1.0	0	1.0	0	0	0	.....	0	0	0	.....
protein 4	0	0	0	0	0.8	0	.....	0	0	1.0	.....
protein 5	1.0	0	0	0	0	0	.....	0	0.9	0	.....
protein 6	0.9	0					.....				.....
protein 7	0	0.8					.....				.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

**Interaction Vectors** [Lan et al, IEEE 90:1848]

# Networks (Old & New)



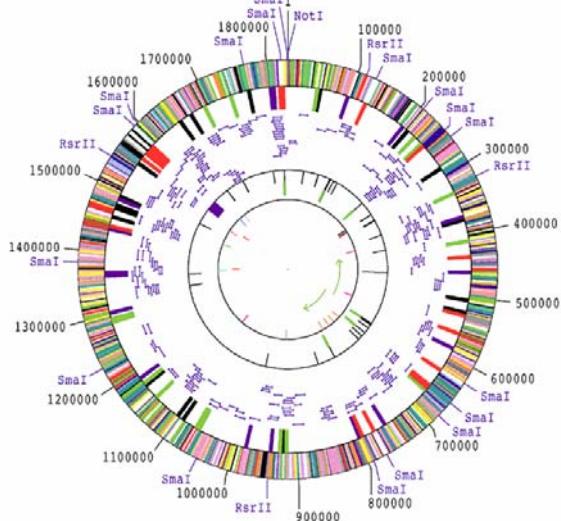
Classical KEGG pathway



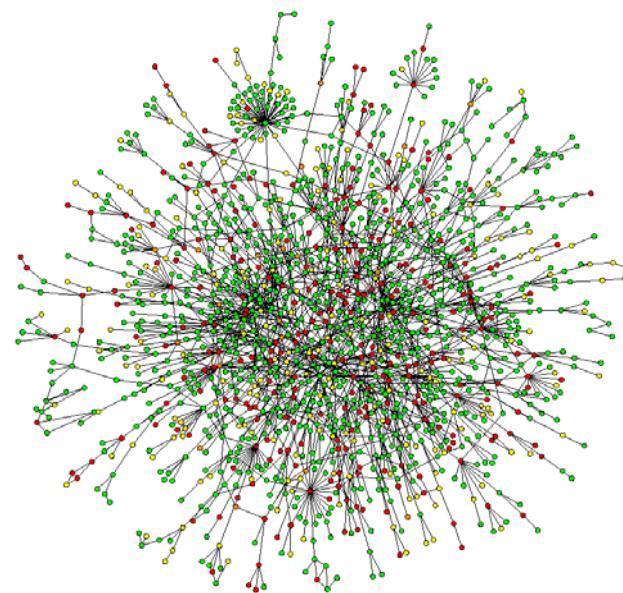
Same Genes in High-throughput Network

[Seringhaus & Gerstein, Am. Sci. '08]

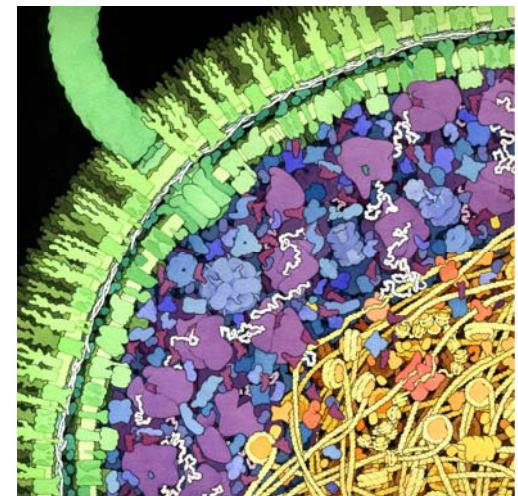
# Networks occupy a midway point in terms of level of understanding



1D: Complete Genetic Partslist



~2D: Bio-molecular Network Wiring Diagram

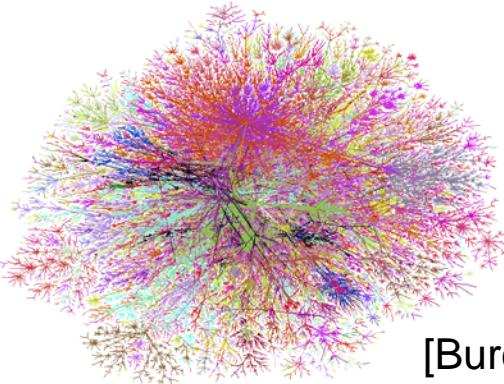


3D: Detailed structural understanding of cellular machinery

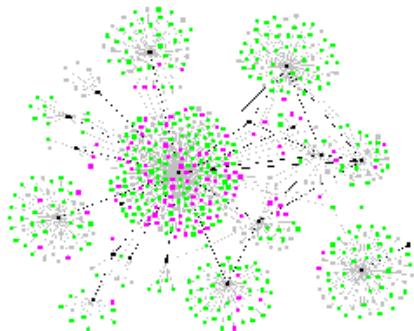
[Fleischmann et al., Science, 269 :496]

[Jeong et al. Nature, 41:411]

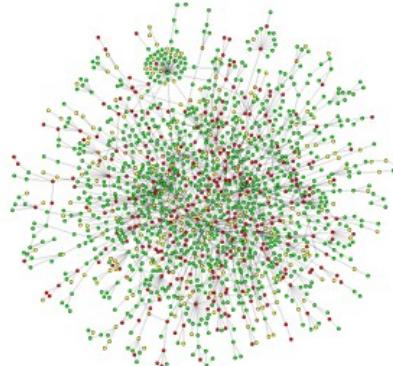
# Networks as a universal language



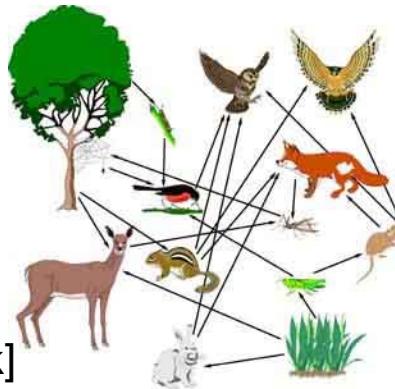
Internet  
[Burch & Cheswick]



Disease  
Spread  
[Krebs]



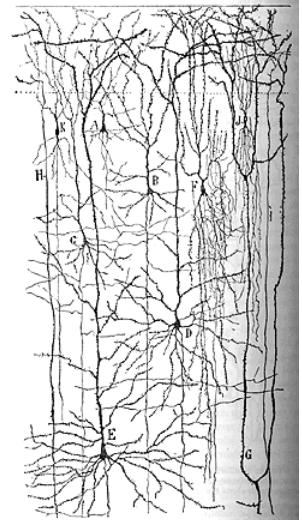
Protein  
Interactions  
[Barabasi]



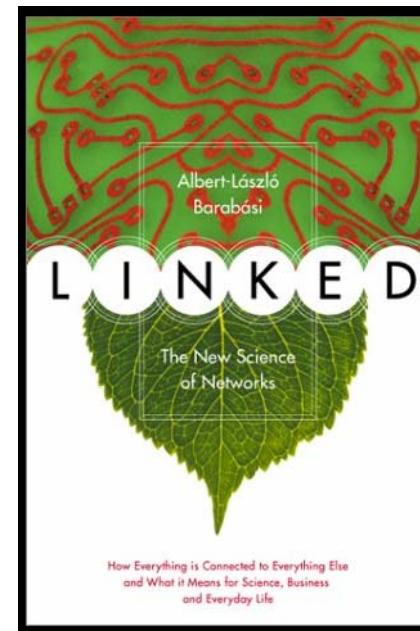
Food Web



Electronic  
Circuit

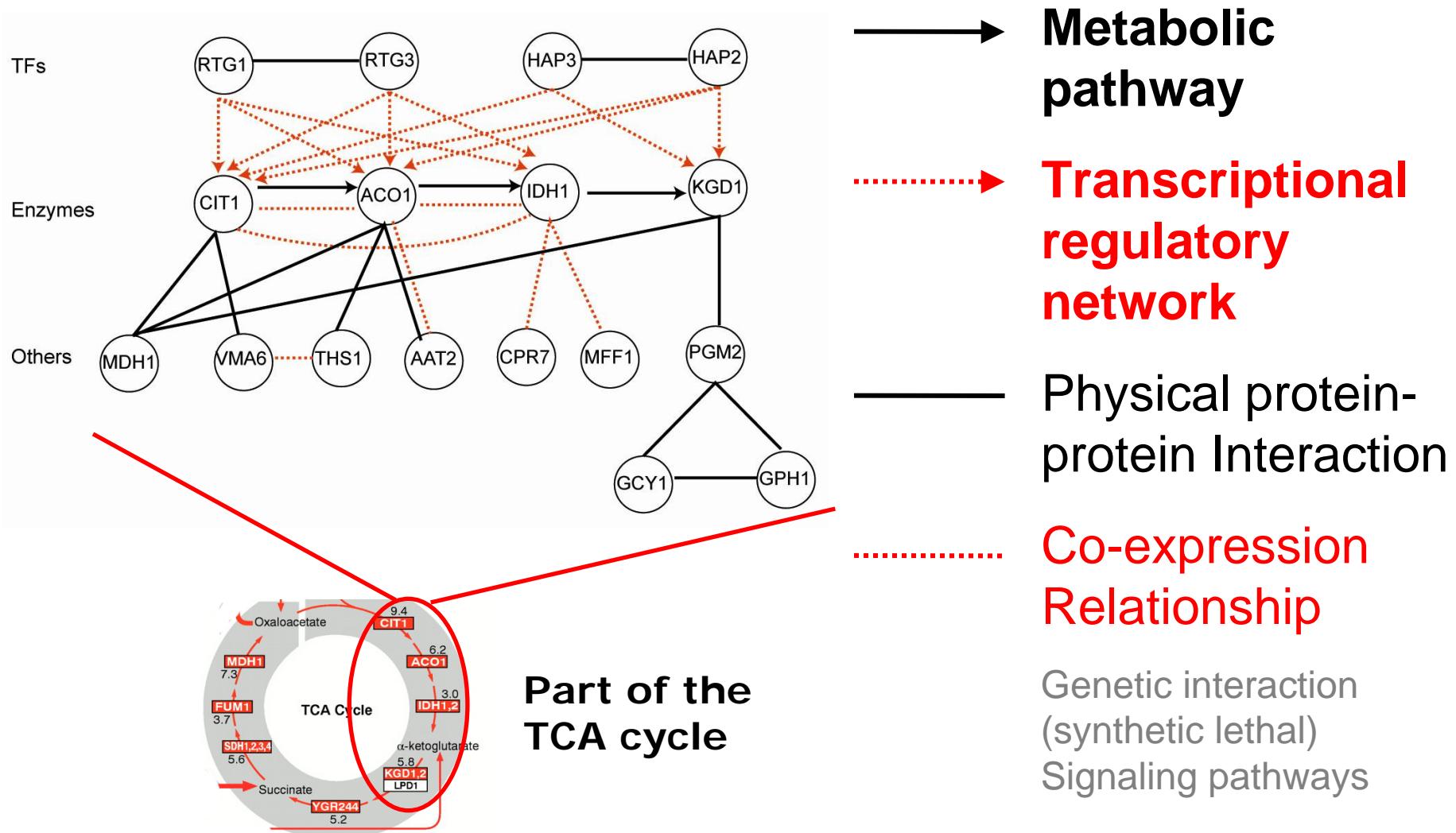


Neural Network  
[Cajal]



Social Network

# Combining networks forms an ideal way of integrating diverse information



# Standardized Phenotypic Functions for Each Gene: Predicting Essential Genes



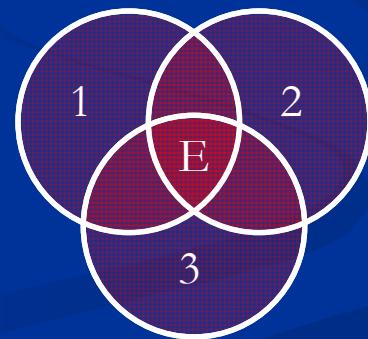
pers. photo, see  
[streams.gerstein.info](http://streams.gerstein.info)

# What are essential genes?

- Not all genes are necessary
  - Redundancy
  - Condition-specific
  - Function not vital
- Essential genes are those *necessary for survival*
  - Deletion / disablement is lethal
- Attractive drug targets
  - Disable one gene, kill the pathogen
- ‘Essentiality’ depends on growth environment
  - Normal definition: Laboratory growth, rich medium

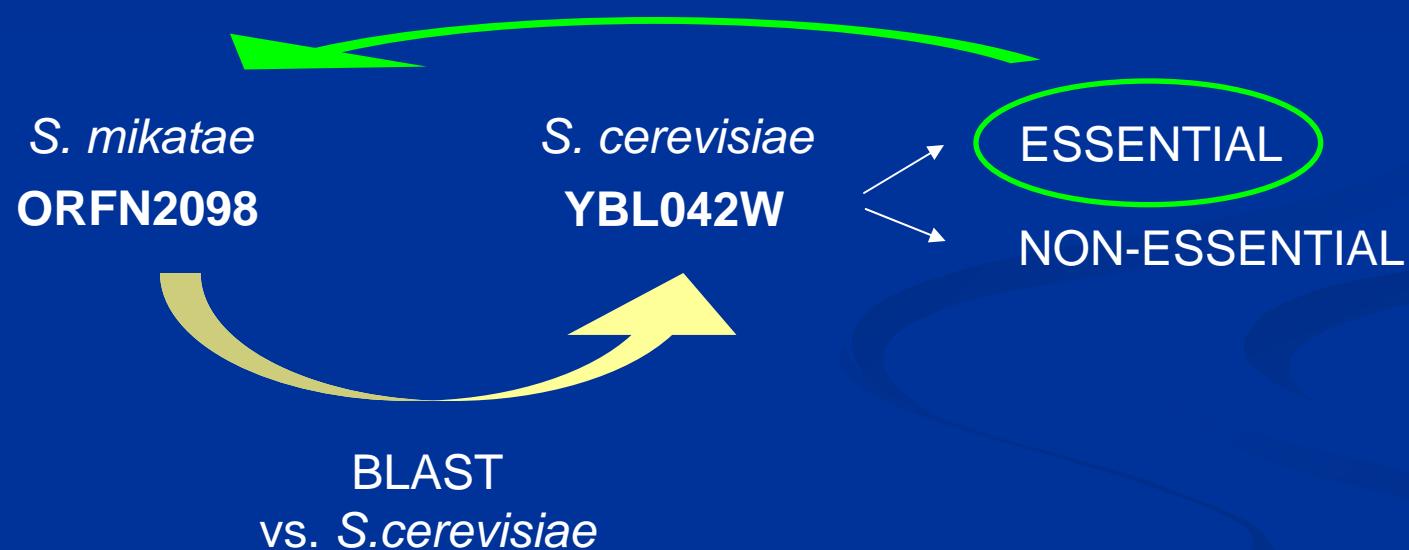
# Identifying essential genes

- Large-scale experimental screens
  - PCR deletion, transposon disruption, RNAi
  - Model organisms
- Homology mapping
  - Map findings to similar genes in other species
- Conservation as essentiality



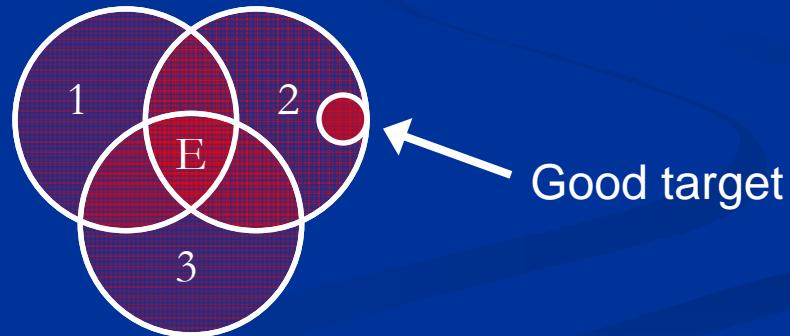
# Estimating essentiality by homology

Gene from close relative *Saccharomyces mikatae*



# The trouble with homology

- Sequence homology has limitations
  - Some fraction of essential genes are unique to a given organism
- For drug-design purposes:
  - Don't want to attack genes that are broadly conserved!



# Approach

1. Gather attributes to characterize essential genes
2. Develop a computational predictor based on these attributes  
 $S. cerevisiae$  (learn)  $\rightarrow S. mikatae$  (predict)
3. Assess predictive success  
Homology, permutation, knockouts

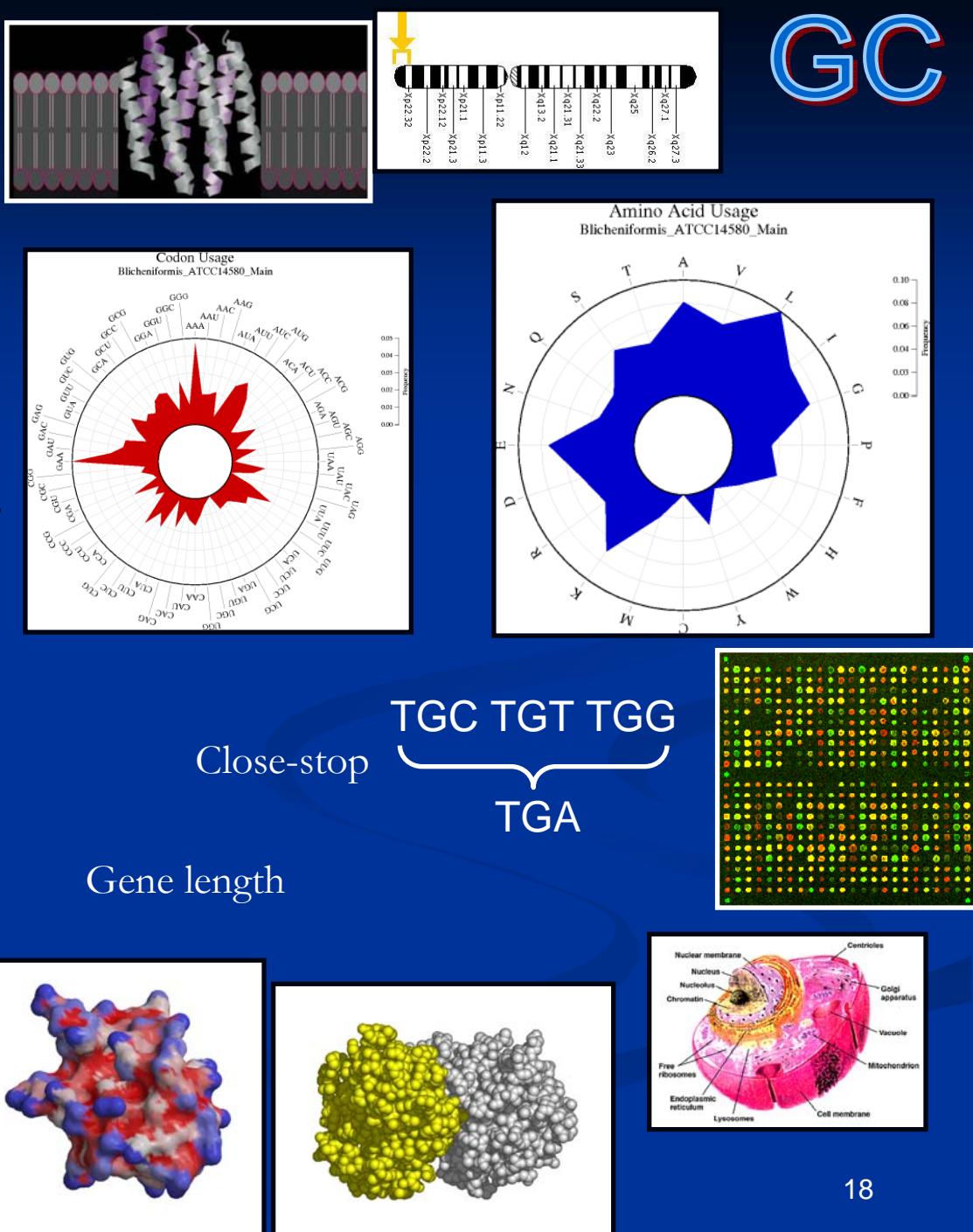
# Group features into 3 classes

## Sequence

- Gene
  - GC content, codon usage, length, etc.
- Genome
  - Chromosome position, duplication

## Predicted from Sequence

- TM helix, localization, predicted hydrophobicity



# 28 features in 3 classes

	NAME	DESCRIPTION	TYPE
Gene Seq	CAI	codon adaptation index	real
	Nc	effective number of codons	real
	GC	GC content	real
	L_aa	length of putative protein in amino acids	integer
	CLOSE_STOP_RATIO	percentage of codons one third-base away from a stop codon	real
	RARE_AA_RATIO	percentage rare amino acids in translated ORF	real
	BLAST_yeast	# BLAST hits in yeast genome (measure of duplication)	integer
	BLAST_closeyeast	hits in # of 6 close yeast species	integer
Genome Seq	BLAST_proks	hits in # of 5 prokaryotic species	integer
	CHR	located on chromosome (number)	integer
	CHR_POSN	Position relative to end of chromosome	real
	INTRON	gene has intron?	binary
	Hydro	hydrophobicity score	real
Predicted from Seq	TM_HELIx	Number of predicted transmembrane helices (TMHMM 2.0)	integer
	PRED_mitochondria	predicted subcellular localization: mitochondria	binary
	PRED_cytoplasm	predicted subcellular localization: cytoplasm	binary
	PRED_er	predicted subcellular localization: endoplasmic reticulum	binary
	PRED_nucleus	predicted subcellular localization: nucleus	binary
	PRED_vacuole	predicted subcellular localization: vacuole	binary
	PRED_other	predicted subcellular localization: any other compartment	binary
	mRNA_EXPR	normalized mRNA expression in copies/cell	real
Experimental Data	INTXN_PARTNERS	protein-protein interaction partners	integer
	EXPT_mitochondria	experimental subcellular localization: mitochondria	binary
	EXPT_cytoplasm	experimental subcellular localization: cytoplasm	binary
	EXPT_er	experimental subcellular localization: endoplasmic reticulum	binary
	EXPT_nucleus	experimental subcellular localization: nucleus	binary
	EXPT_vacuole	experimental subcellular localization: vacuole	binary
	EXPT_other	experimental subcellular localization: any other compartment	binary

# 14 features (unstudied genomes)

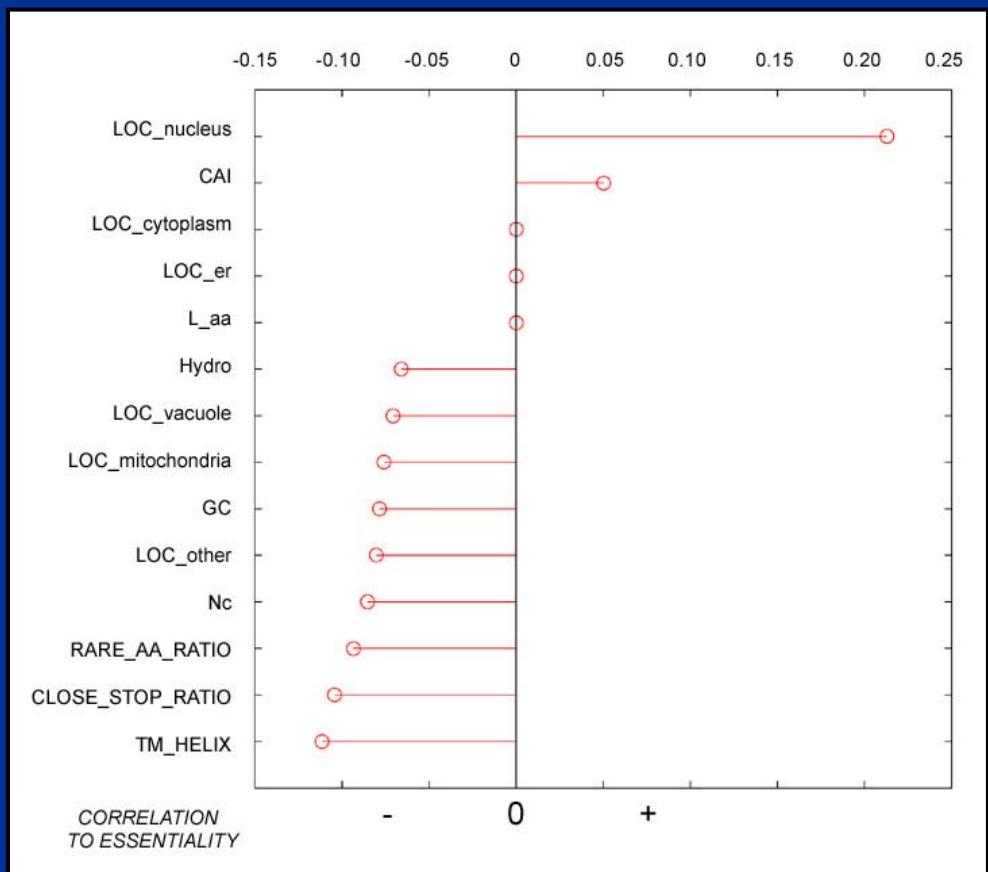
	NAME	DESCRIPTION	TYPE
Gene Seq	CAI	codon adaptation index	real
	Nc	effective number of codons	real
	GC	GC content	real
	L_aa	length of putative protein in amino acids	integer
	CLOSE_STOP_RATIO	percentage of codons one third-base away from a stop codon	real
	RARE_AA_RATIO	percentage rare amino acids in translated ORF	real
Genome Seq	BLAST_yeast	# BLAST hits in yeast genome (measure of duplication)	integer
	BLAST_closeyeast	hits in # of 6 close yeast species	integer
	BLAST_proks	hits in # of 5 prokaryotic species	integer
	CHR	located on chromosome (number)	integer
	CHR_POSN	Position relative to end of chromosome	real
	INTRON	gene has intron?	binary
Predicted from Seq	Hydro	hydrophobicity score	real
	TM_HELIx	Number of predicted transmembrane helices (TMHMM 2.0)	integer
	PRED_mitochondria	predicted subcellular localization: mitochondria	binary
	PRED_cytoplasm	predicted subcellular localization: cytoplasm	binary
	PRED_er	predicted subcellular localization: endoplasmic reticulum	binary
	PRED_nucleus	predicted subcellular localization: nucleus	binary
	PRED_vacuole	predicted subcellular localization: vacuole	binary
	PRED_other	predicted subcellular localization: any other compartment	binary
Experimental Data	mRNA_EXPR	normalized mRNA expression in copies/cell	real
	INTXN_PARTNERS	protein-protein interaction partners	integer
	EXPT_mitochondria	experimental subcellular localization: mitochondria	binary
	EXPT_cytoplasm	experimental subcellular localization: cytoplasm	binary
	EXPT_er	experimental subcellular localization: endoplasmic reticulum	binary
	EXPT_nucleus	experimental subcellular localization: nucleus	binary
	EXPT_vacuole	experimental subcellular localization: vacuole	binary
	EXPT_other	experimental subcellular localization: any other compartment	binary

# Correlating features to essentiality

- Pair-wise correlational analysis
  - Each feature vs. essentiality (*Nomogram*)

Not looking at co-variation or higher-order effects

3 features have pair-wise correlations that were not significant ( $p > 0.05$ ), and were set to zero



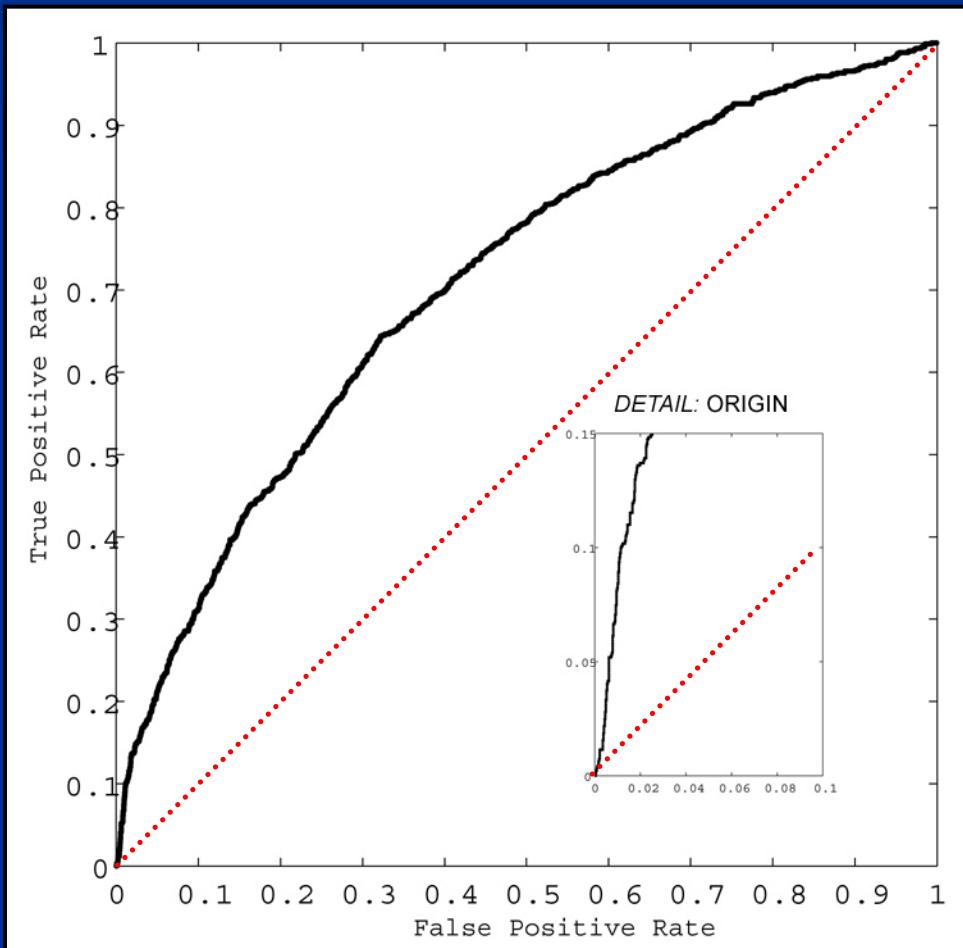
# Simple & Complex Classifier Design

- Simple
  - Naive Bayes  
(just adding votes of each feature in nomogram)
  
- Complex: hybrid machine learning system
  - Combination of Naïve Bayes, decision trees, & logistic regression
  - Naive Bayes classifier
  - decision stump boosted using Adaboost (Freund and Schapire 1996)
  - random forest (Breiman 2001)
  - alternating decision tree with 10 boosting iterations (Freund and Mason 1999)
  - C4.5 decision tree (Quinlan 1993)
  - multinomial logistic regression model with ridge estimator (le Cessie and van Houwelingen 1992)



# ROC curve: classifier performance on *S cerevisiae*

Receiver-Operating Characteristic curve



- Cost-benefit analysis of a binary classifier as discrimination threshold is varied
- *Ideal predictor*: Top left
- *Random predictor*: Diagonal
- Area of interest: Near origin
  - Don't need to identify *every* essential gene
- Adjust cost matrix to discourage false positives

# Classifier applied to *S. mikatae*

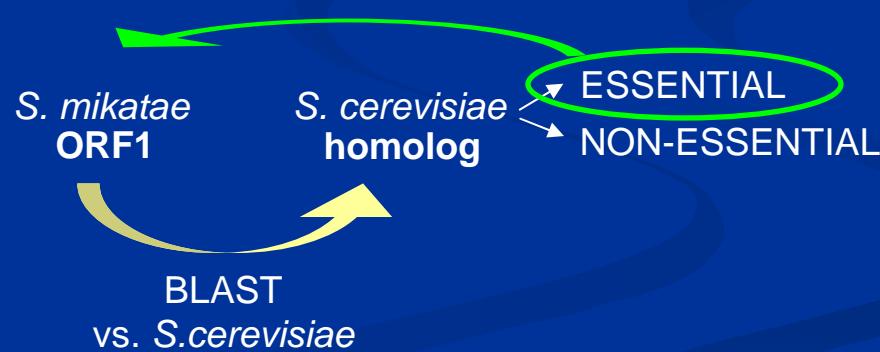
ORF	mitochondria	cytoplasm	er	nucleus	vacuole	other	CAI	Nc	GC	L_aa	Hydro	CLOSE_STOP_RATIO	RARE_AA_RATIO	TM_HELI	ESS
ORFN:16615	0	1	0	0	0	0	0.10	61	0.43	306	-0.01	0.07	0.06	1	?
ORFN:286	0	0	0	0	0	0	0.15	61	0.35	83	-0.01	0.04	0.08	5	?
ORFN:7697	0	0	0	0	0	0	0.06	61	0.41	109	0.63	0.11	0.07	1	?
ORFN:9220	0	0	0	0	0	0	0.11	61	0.44	57	0.66	0.09	0.10	1	?
ORFN:9266	0	0	0	0	0	0	0.09	61	0.44	137	-0.91	0.06	0.07	0	?
ORFN:388	0	0	0	0	0	1	0.12	61	0.46	51	0.64	0.04	0.06	0	?
ORFN:16740	0	0	0	0	0	0	0.12	61	0.43	145	0.54	0.03	0.03	2	?
ORFN:16767	0	1	0	0	0	0	0.08	61	0.44	190	-0.09	0.06	0.09	0	?
ORFN:450	0	0	0	0	0	0	0.10	61	0.42	200	-0.05	0.10	0.09	0	?
ORFN:454	0	0	0	0	0	0	0.07	61	0.42	124	0.38	0.08	0.08	1	?
ORFN:467	0	0	0	0	0	0	0.11	61	0.40	195	0.40	0.06	0.08	0	?
ORFN:16785	0	0	0	0	0	0	0.14	61	0.35	180	-0.53	0.06	0.06	0	?
ORFN:470	0	0	0	0	0	0	0.12	61	0.42	238	-0.15	0.08	0.06	0	?
ORFN:16814	0	0	0	0	0	0	0.08	61	0.40	71	-0.45	0.06	0.07	0	?

ORF	mitochondria	cytoplasm	er	nucleus	vacuole	other	CAI	Nc	GC	L_aa	Hydro	CLOSE_STOP_RATIO	RARE_AA_RATIO	TM_HELI	ESS
ORFN:16615	0	1	0	0	0	0	0	61	0.43	306	-0.01	0.07	0.06	1	green
ORFN:286	0	0	0	0	0	0	0.15	61	0.35	83	-0.01	0.04	0.08	5	red
ORFN:7697	0	0	0	0	0	0	0.06	61	0.41	109	0.63	0.11	0.07	1	green
ORFN:9220	0	0	0	0	0	0	0.11	61	0.44	57	0.66	0.09	0.10	1	green
ORFN:9266	0	0	0	0	0	0	0.09	61	0.44	137	-0.91	0.06	0.07	0	red
ORFN:388	0	0	0	0	0	1	0.12	61	0.46	51	0.64	0.04	0.06	0	green
ORFN:16740	0	0	0	0	0	0	0.12	61	0.43	145	0.54	0.03	0.03	2	green
ORFN:16767	0	1	0	0	0	0	0.08	61	0.44	190	-0.09	0.06	0.09	0	green
ORFN:450	0	0	0	0	0	0	0.10	61	0.42	200	-0.05	0.10	0.09	0	green
ORFN:454	0	0	0	0	0	0	0.07	61	0.42	124	0.38	0.08	0.08	1	green
ORFN:467	0	0	0	0	0	0	0.11	61	0.40	195	0.40	0.06	0.08	0	red
ORFN:16785	0	0	0	0	0	0	0.14	61	0.35	180	-0.53	0.06	0.06	0	green
ORFN:470	0	0	0	0	0	0	0.12	61	0.42	238	-0.15	0.08	0.06	0	green
ORFN:16814	0	0	0	0	0	0	0.08	61	0.40	71	-0.45	0.06	0.07	0	green

Assigns essentiality score to each gene

RANK	<i>S.mikatae</i> ORF_ID	SCORE (0-1)
1	18373	1
2	20026	0.997
3	20713	0.986
4	10758	0.981
5	3115	0.974
6	12016	0.964
7	911	0.964
8	22659	0.962
9	17592	0.957
10	3944	0.953
11	3749	0.946
12	11901	0.944
13	14387	0.944
14	19319	0.944
15	2919	0.944
16	21382	0.943
17	4885	0.941
18	20238	0.936
19	20242	0.936
20	644	0.936
21	788	0.936
22	7883	0.936
23	17520	0.934
24	8795	0.934
25	1400	0.932
26	19369	0.932
27	13994	0.931
28	21002	0.929
29	21414	0.929
30	21484	0.929
...	...	...
3726	5507	0.198
...	...	...
3751	18487	0.193
...	...	...
3939	2579	0.128

- Assigned essentiality score (0-1)
- 3939 putative *S. mikatae* ORFs
- How does this compare to homology mapping?



RANK	<i>S.mikatae</i> ORF_ID	SCORE (0-1)	homolog in <i>S.cerevisiae</i>	BLAST vs. <i>S.cerevisiae</i> essential genes		
				1.00E-04	1.00E-20	1.00E-50
1	18373	1	Multiple	1	1	1
2	20026	0.997	YOR046C	1	1	1
3	20713	0.986	YOR341W	1	1	1
4	10758	0.981	YIL126W	1	1	1
5	3115	0.974	YDL031W	1	1	1
6	12016	0.964	YJL050W	1	1	1
7	911	0.964	YBR088C	1	1	1
8	22659	0.962	Multiple	1	1	1
9	17592	0.957	YMR308C	1	1	1
10	3944	0.953	YDR190C	1	1	1
11	3749	0.946	YDR101C	0	0	0
12	11901	0.944	YJL080C	0	0	0
13	14387	0.944	YLL004W	1	1	0
14	19319	0.944	YOL010W	1	1	1
15	2919	0.944	Multiple	1	1	1
16	21382	0.943	YPL235W	1	1	1
17	4885	0.941	YBR247C	1	1	1
18	20238	0.936	YOR117W	1	1	1
19	20242	0.936	YOR116C	1	1	1
20	644	0.936	YBR158W	0	0	0
21	788	0.936	YBR119W	0	0	0
22	7883	0.936	YGL078C	0	0	0
23	17520	0.934	YMR290C	1	1	1
24	8795	0.934	YGR145W	1	1	1
25	1400	0.932	YBL023C	1	1	1
26	19369	0.932	YOL021C	1	1	1
27	13994	0.931	YKR081C	1	1	1
28	21002	0.929	YOR259C	1	1	1
29	21414	0.929	YPL228W	1	1	1
30	21484	0.929	YPL211W	1	1	1
...	...	...	...	...	...	...
3726	5507	0.198	YDR525W-A	0	0	0
...	...	...	...	...	...	...
3751	18487	0.193	YNL101W	0	0	0
...	...	...	...	...	...	...
3939	2579	0.128	YDL173W	0	0	0

RANK	<i>S.mikatae</i> ORF_ID	SCORE (0-1)	homolog in <i>S.cerevisiae</i>	BLAST vs. <i>S.cerevisiae</i> essential genes			
				1.00E-04	1.00E-20	1.00E-50	
1	18373	1	Multiple	1	1	1	Predicted Essential
2	20026	0.997	YOR046C	1	1	1	Predicted Essential
3	20713	0.986	YOR341W	1	1	1	Predicted Essential
4	10758	0.981	YIL126W	1	1	1	Predicted Essential
5	3115	0.974	YDL031W	1	1	1	Predicted Essential
6	12016	0.964	YJL050W	1	1	1	Predicted Essential
7	911	0.964	YBR088C	1	1	1	Predicted Essential
8	22659	0.962	Multiple	1	1	1	Predicted Essential
9	17592	0.957	YMR308C	1	1	1	Predicted Essential
10	3944	0.953	YDR190C	1	1	1	Predicted Essential
11	3749	0.946	YDR101C	0	0	0	Disputed
12	11901	0.944	YJL080C	0	0	0	Disputed
13	14387	0.944	YLL004W	1	1	0	Predicted Essential
14	19319	0.944	YOL010W	1	1	1	Predicted Essential
15	2919	0.944	YMR105C	1	1	1	Predicted Essential
16	21332	0.944	YVR101C	1	1	1	Predicted Essential
17	4805	0.941	YBR247C	1	1	1	Predicted Essential
18	20238	0.936	YOR117W	1	1	1	Predicted Essential
19	20242	0.936	YOR116C	1	1	1	Predicted Essential
20	644	0.936	YBR158W	0	0	0	Disputed
21	788	0.936	YBR119W	0	0	0	Disputed
22	7883	0.936	YGL078C	0	0	0	Disputed
23	17520	0.934	YMR290C	1	1	1	Predicted Essential
24	8795	0.934	YGR145W	1	1	1	Predicted Essential
25	1400	0.932	YBL023C	1	1	1	Predicted Essential
26	19369	0.932	YOL021C	1	1	1	Predicted Essential
27	13994	0.931	YKR081C	1	1	1	Predicted Essential
28	21002	0.929	YOR259C	1	1	1	Predicted Essential
29	21414	0.929	YPL228W	1	1	1	Predicted Essential
30	21484	0.929	YPL211W	1	1	1	Predicted Essential
...	...	...	...	...	...	...	...
3726	5507	0.120	YDR525W_A	0	0	0	Predicted Non-Essential
...	...	...	...	...	...	...	...
3751	18487	0.120	YDL173W	0	0	0	Predicted Non-Essential
...	...	...	...	...	...	...	...
3939	2579	0.120	YDL173W	0	0	0	Predicted Non-Essential

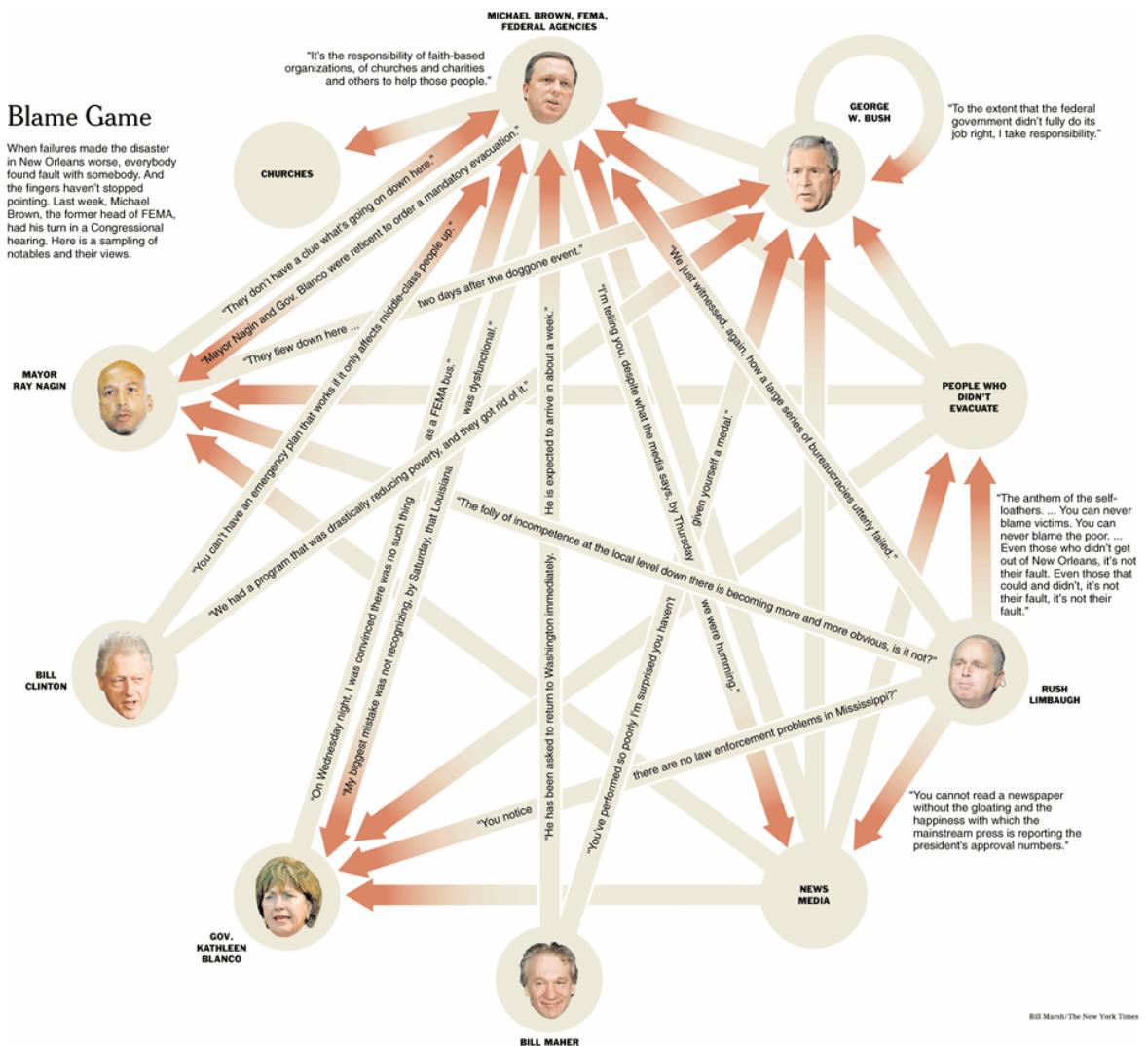
10 / 10 agree with homology mapping

25 / 30 agree with homology mapping

# Using Networks to Describe Function

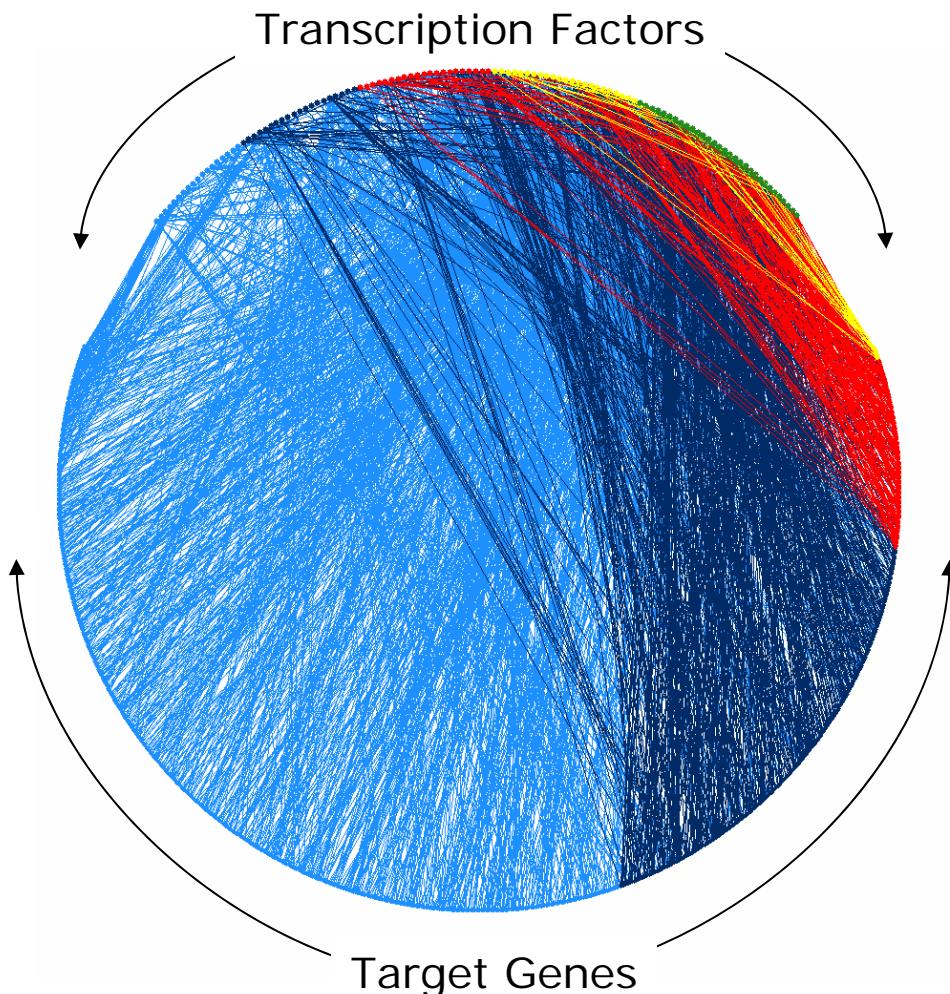
## Blame Game

When failures made the disaster in New Orleans worse, everybody found fault with somebody. And the fingers haven't stopped pointing. Last week, Michael Brown, the former head of FEMA, had his turn in a Congressional hearing. Here is a sampling of notables and their views.



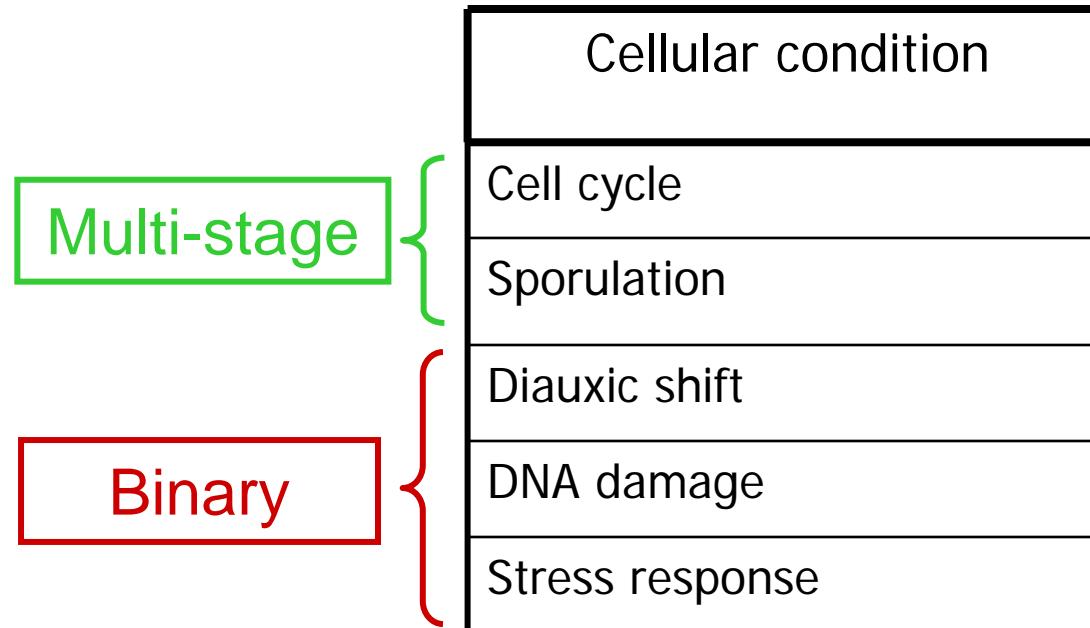
[NY Times, 2-Oct-2005]

# *Dynamic* Yeast TF network



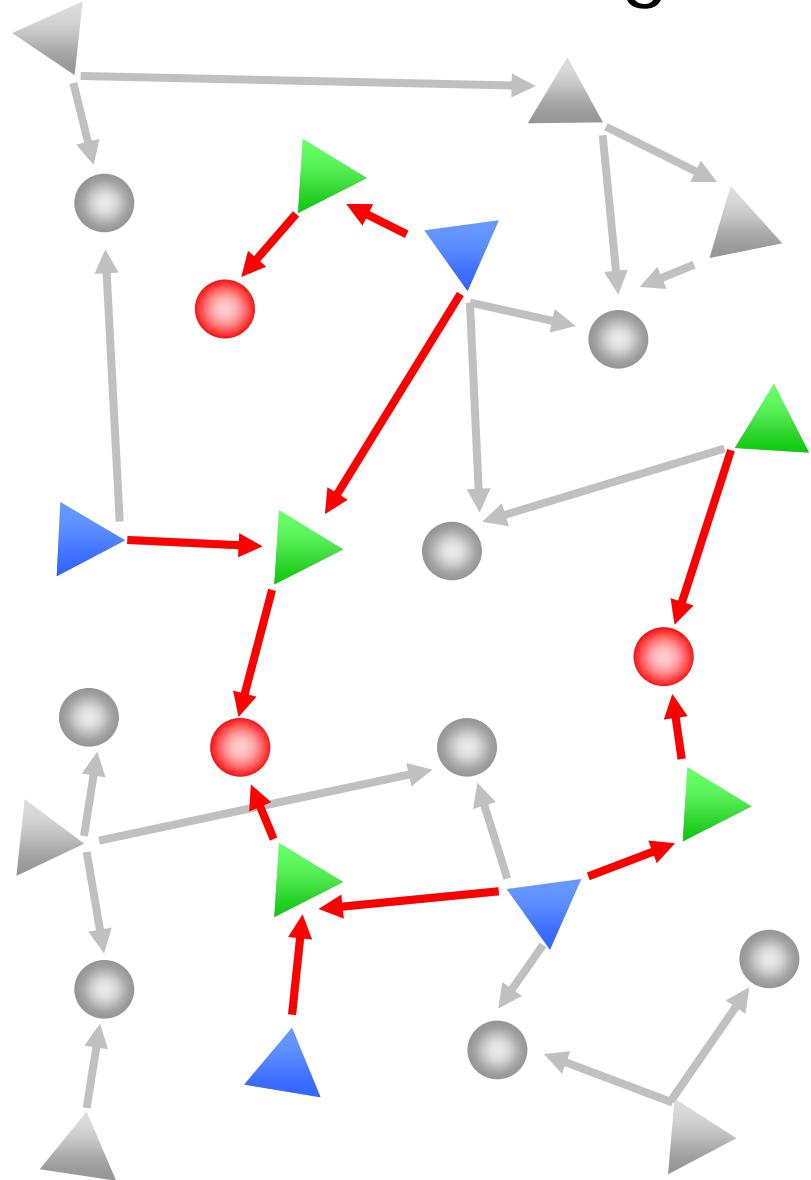
- Analyzed network as a static entity
- But network is *dynamic*
  - ◊ Different sections of the network are active under different cellular conditions
- Integrate gene expression data

# Gene expression data for five cellular conditions in yeast



[Brown, Botstein, Davis....]

# Backtracking to find active sub-network

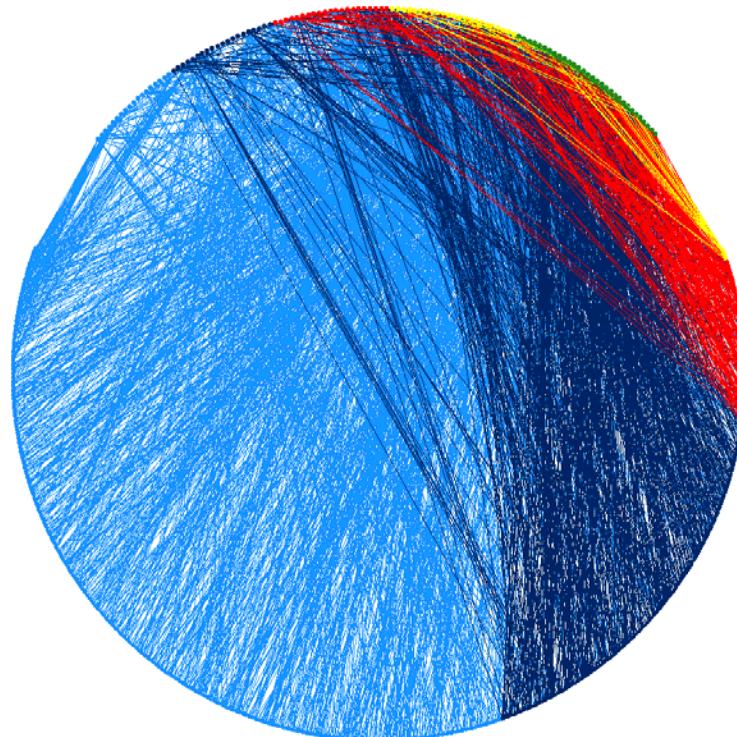


- Define differentially expressed genes
- Identify TFs that regulate these genes
- Identify further TFs that regulate these TFs

Active regulatory sub-network

# Network usage under different conditions

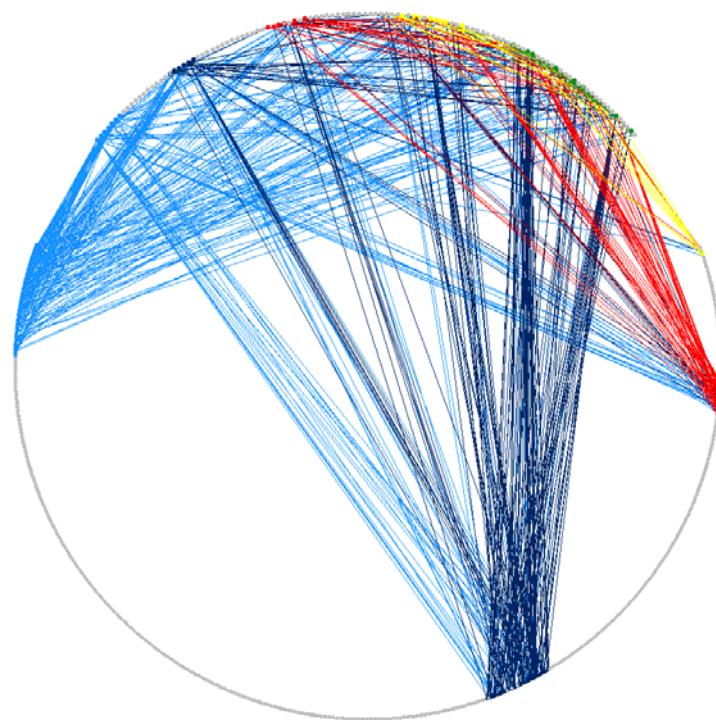
**static**



Luscombe et al. Nature 431: 308

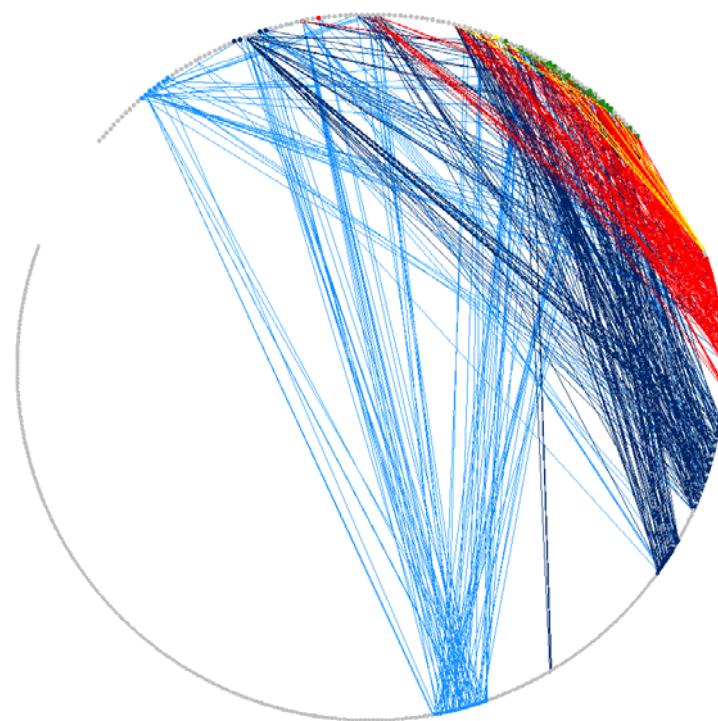
# Network usage under different conditions

**cell cycle**



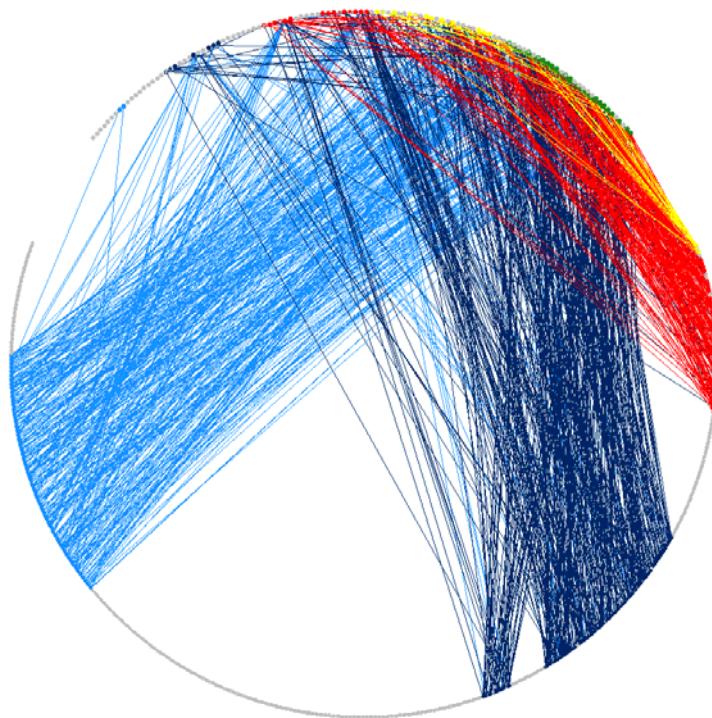
# Network usage under different conditions

## **sporulation**



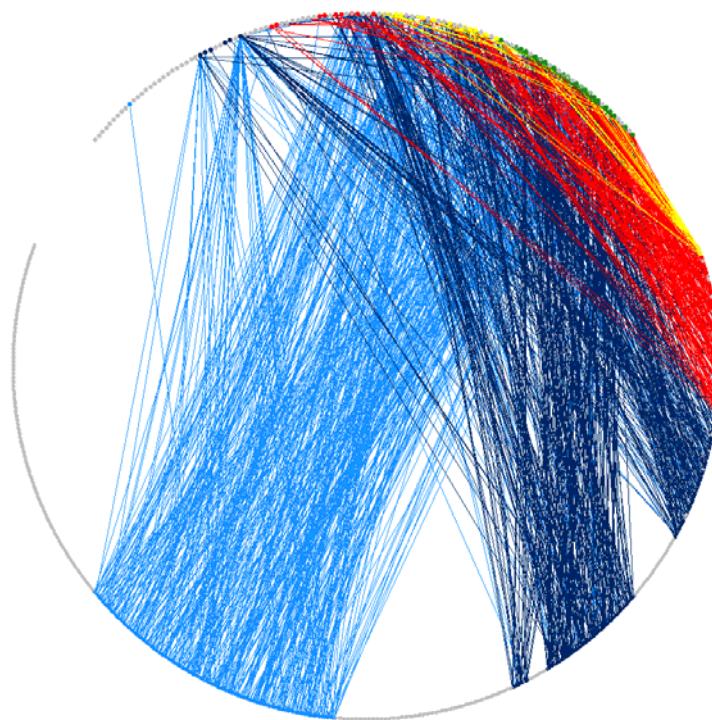
# Network usage under different conditions

**diauxic shift**



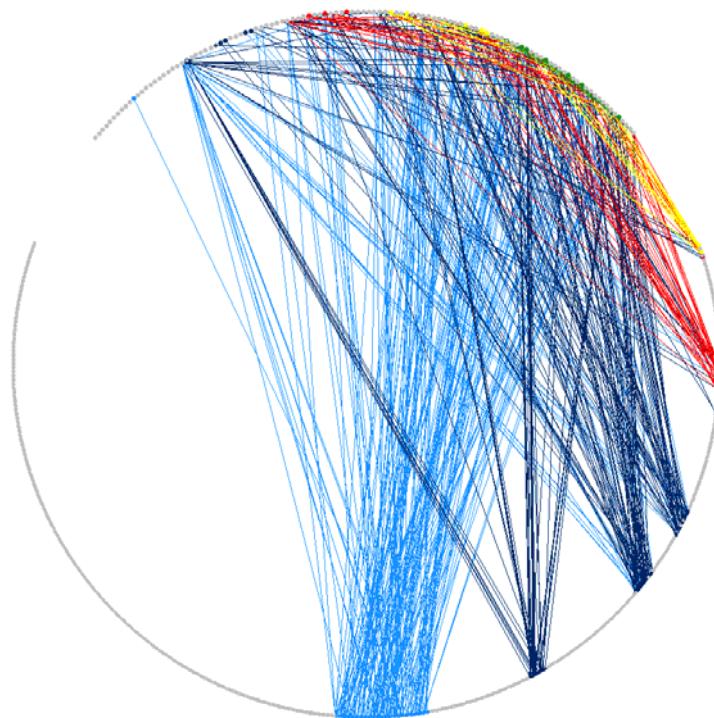
# Network usage under different conditions

## DNA damage

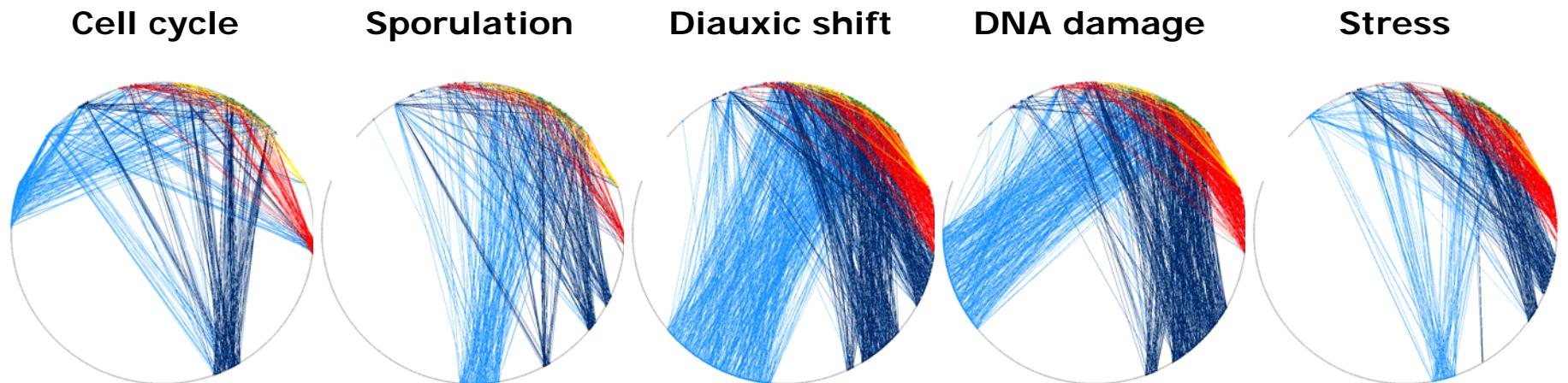


# Network usage under different conditions

## stress response



# Network usage under different conditions

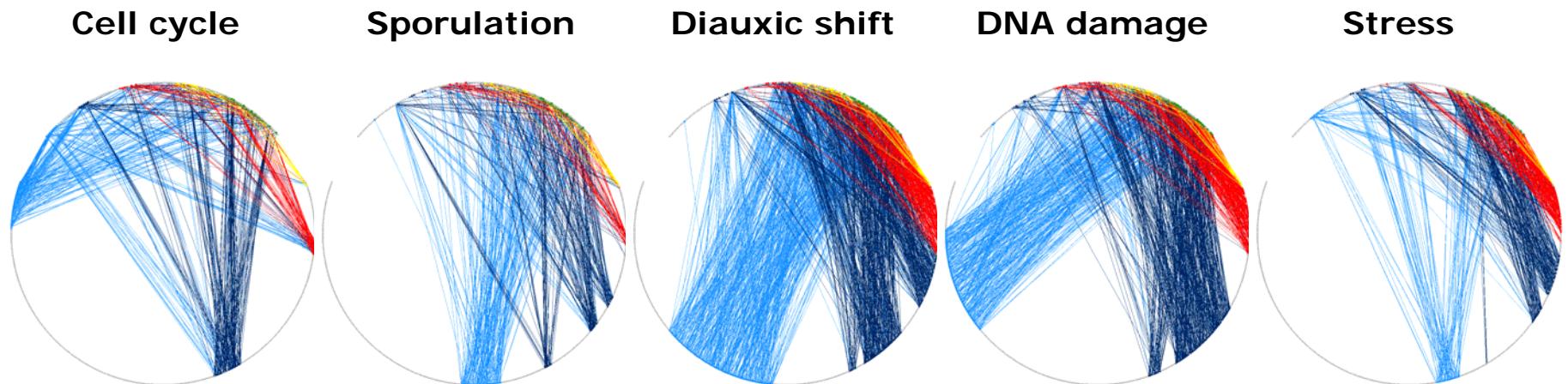


## SANDY:

- 1. Standard graph-theoretic statistics:**
  - Global topological measures
  - Local network motifs
- 2. Newly derived follow-on statistics:**
  - Hub usage
  - Interaction rewiring
- 3. Statistical validation of results**

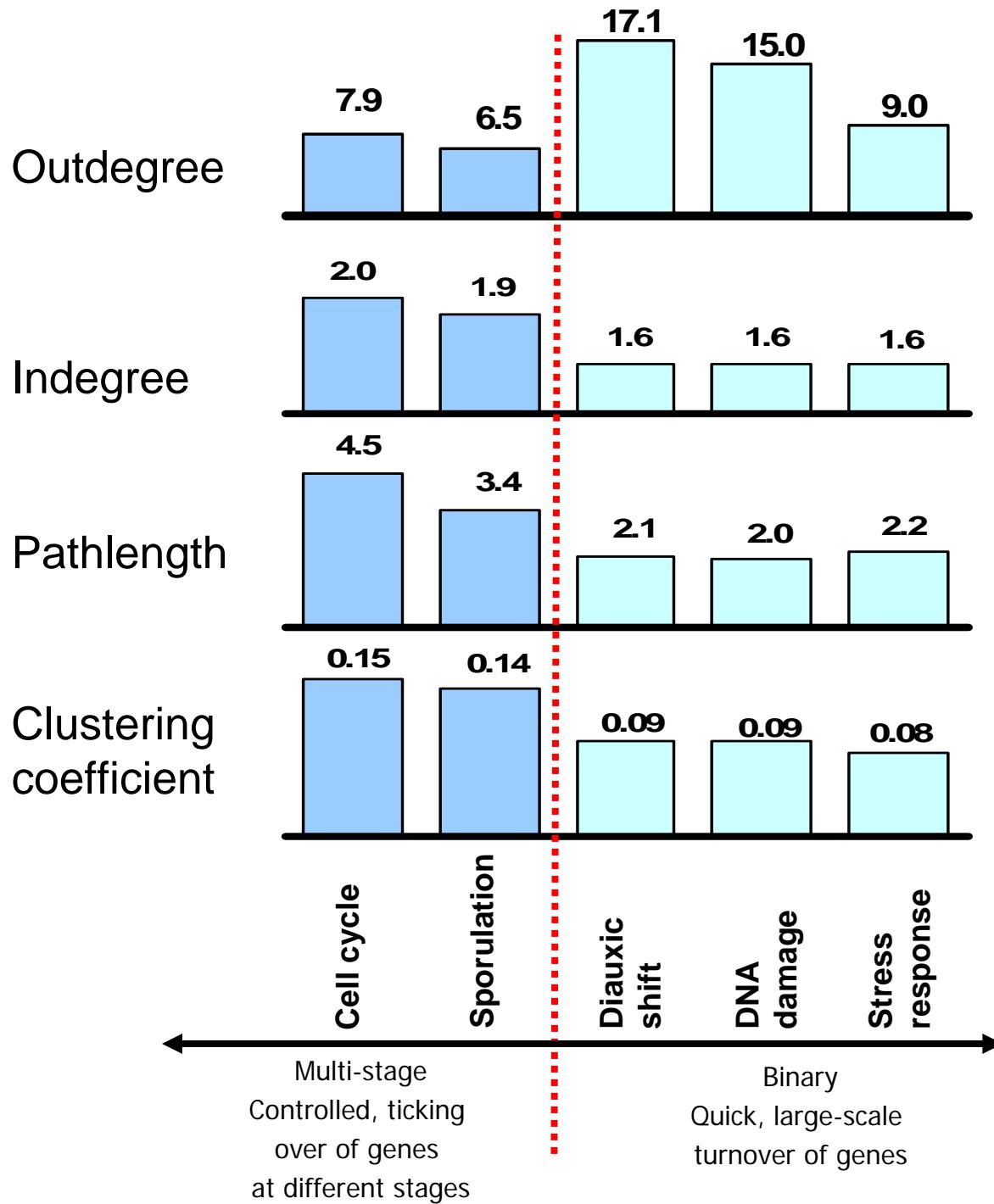
Luscombe et al. Nature 431: 308

# Network usage under different conditions



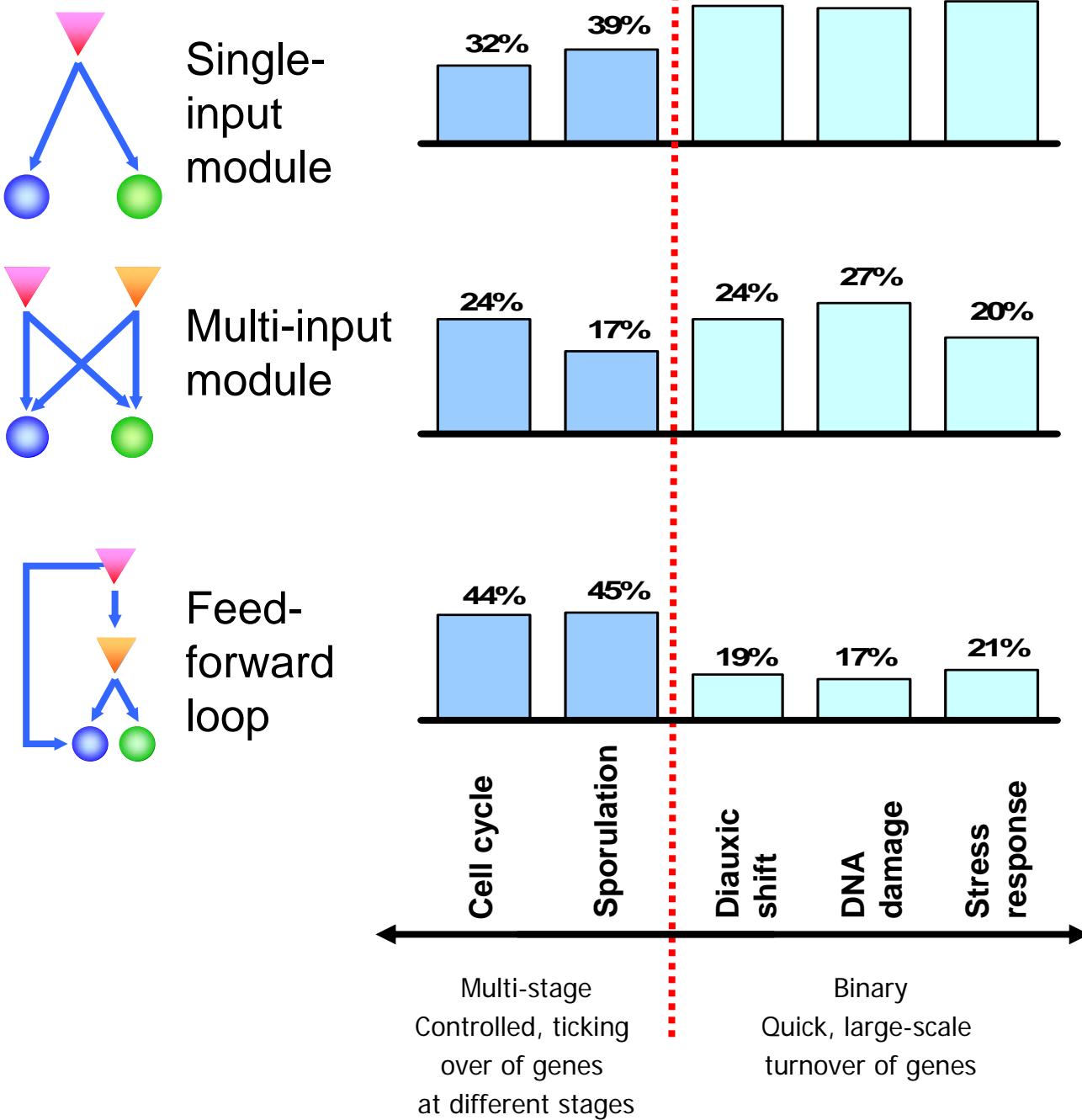
## SANDY:

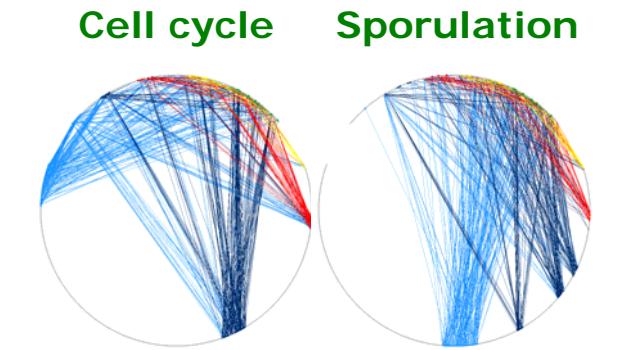
1. Standard graph-theoretic statistics:
  - Global topological measures
  - Local network motifs
2. Newly derived follow-on statistics:
  - Hub usage
  - Interaction rewiring
3. Statistical validation of results



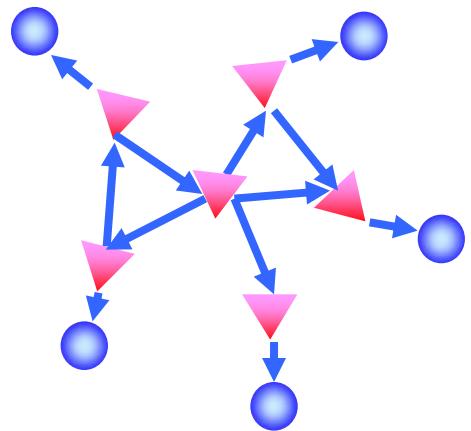
# Analysis of condition-specific subnetworks in terms of global topological statistics

# Analysis of condition-specific subnetworks in terms of occurrence of local motifs

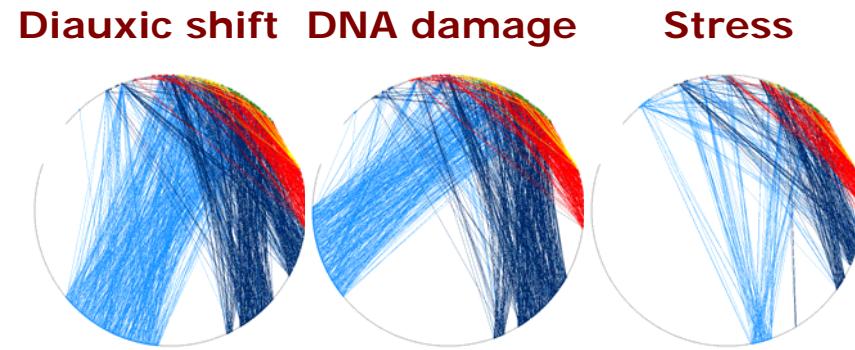




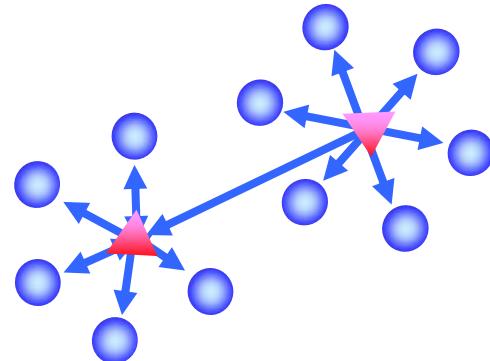
multi-stage conditions



less pronounced  
longer  
more  
complex loops (FFLs)



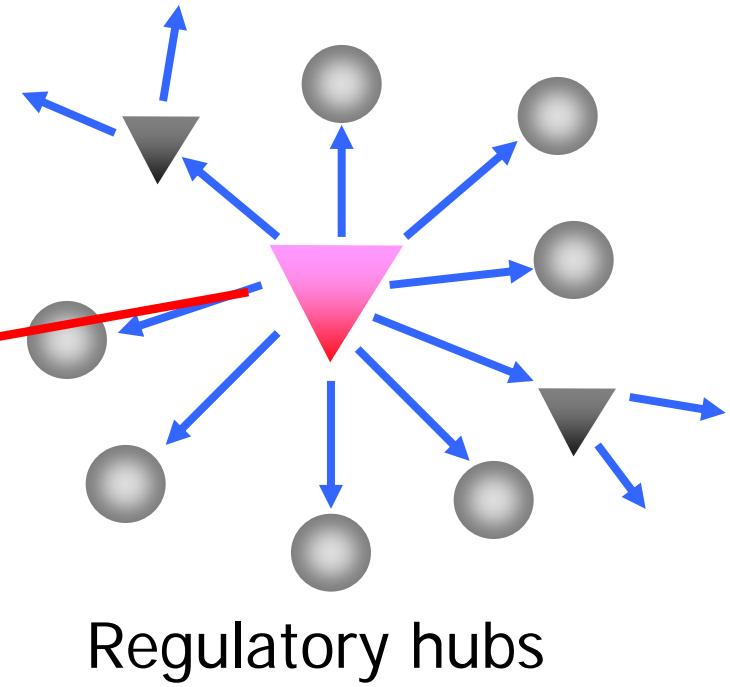
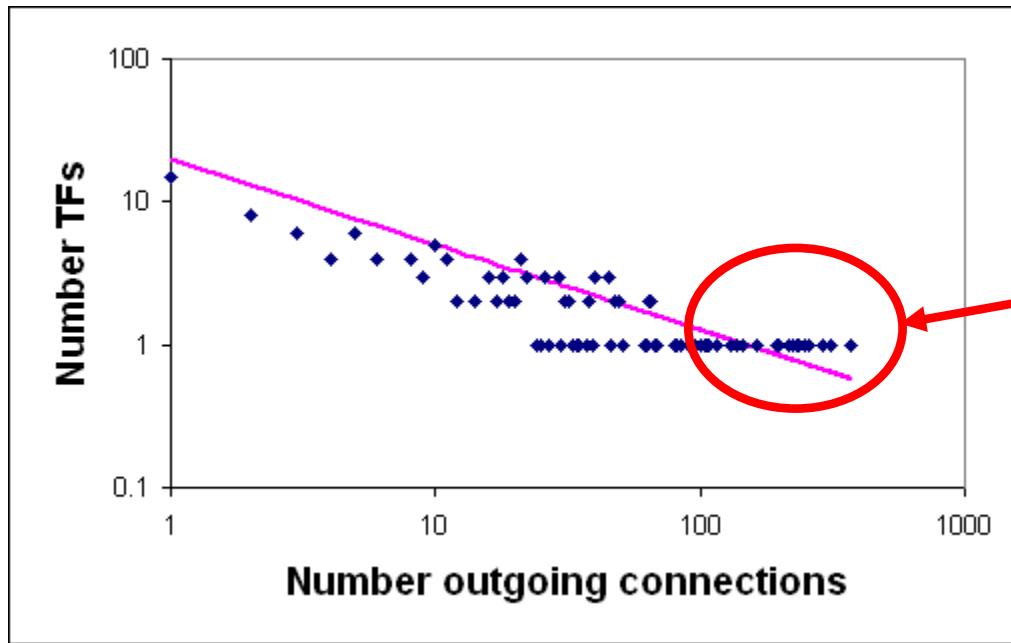
binary conditions



## Summary

Hubs	more pronounced
Path Lengths	shorter
TF inter-regulation	less
Motifs	simpler (SIMs)

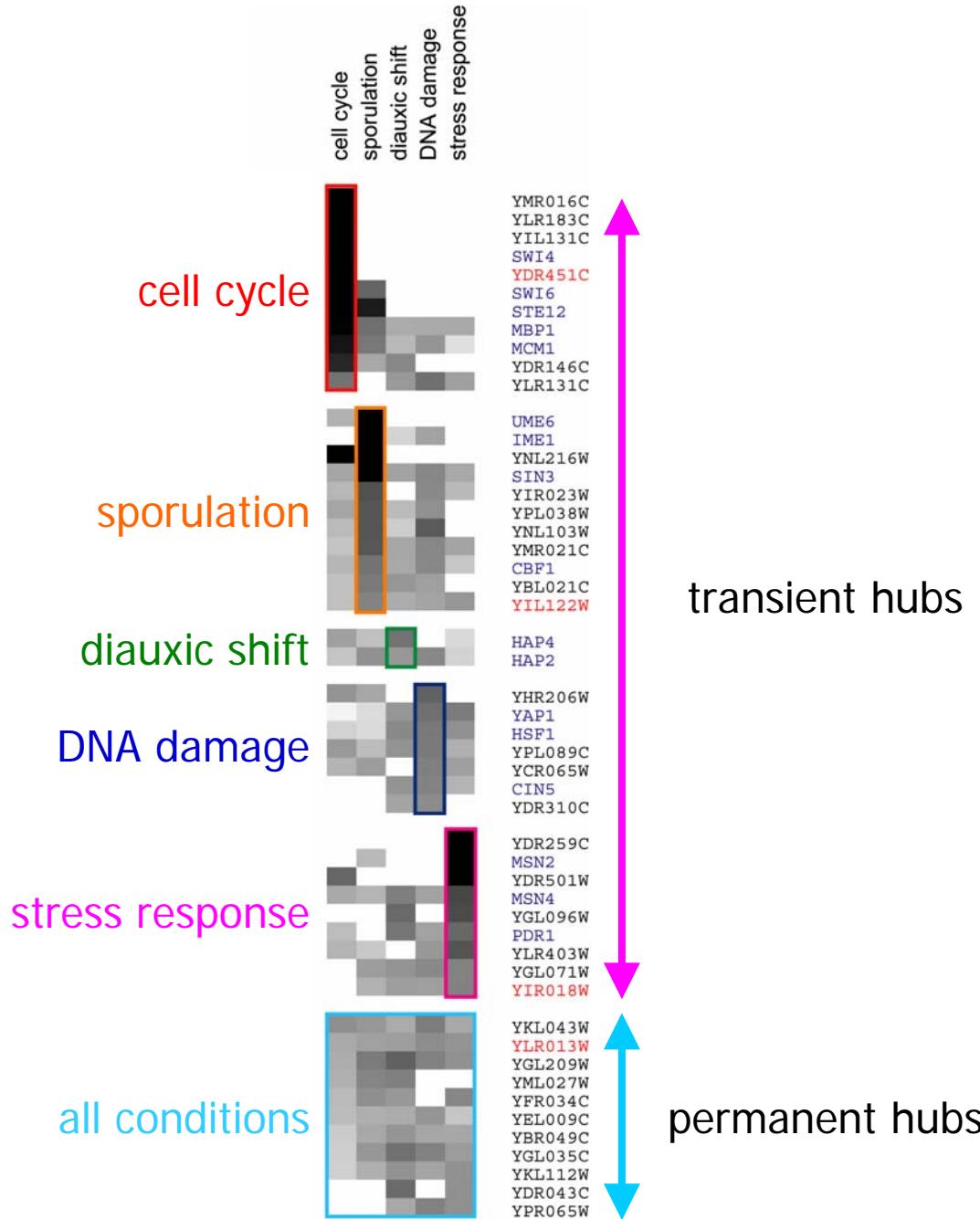
# Transient Hubs



Regulatory hubs

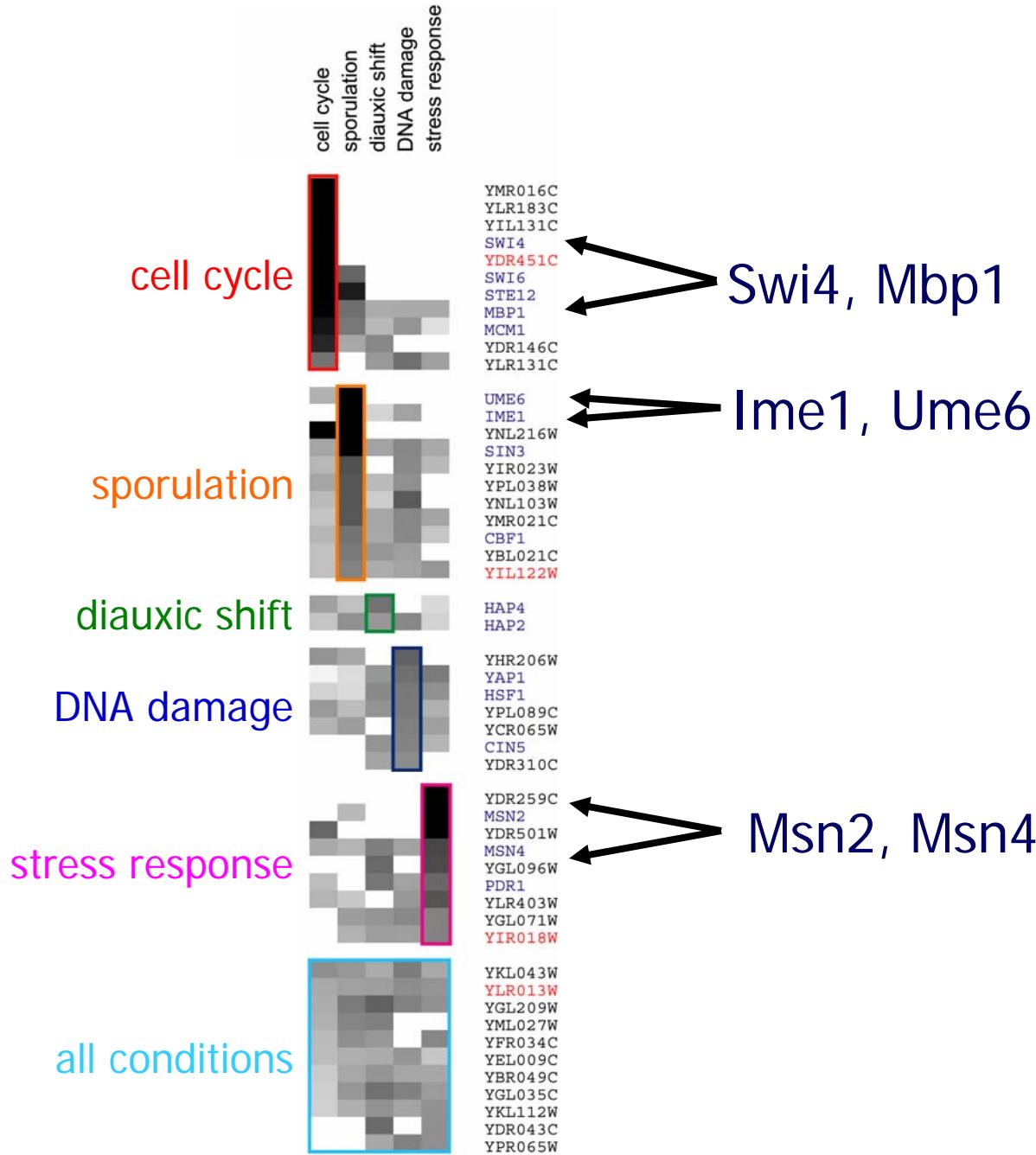
- Questions:
  - ◊ Do hubs stay the same or do they change over between conditions?
  - ◊ Do different TFs become important?
- Our Expectations
  - ◊ Literature:
    - Hubs are permanent features of the network regardless of condition
  - ◊ Random networks (sampled from complete regulatory network)
    - Random networks converge on same TFs
    - 76-97% overlap in TFs classified as hubs (*i.e.* hubs are permanent)

Luscombe et al. Nature 431: 308

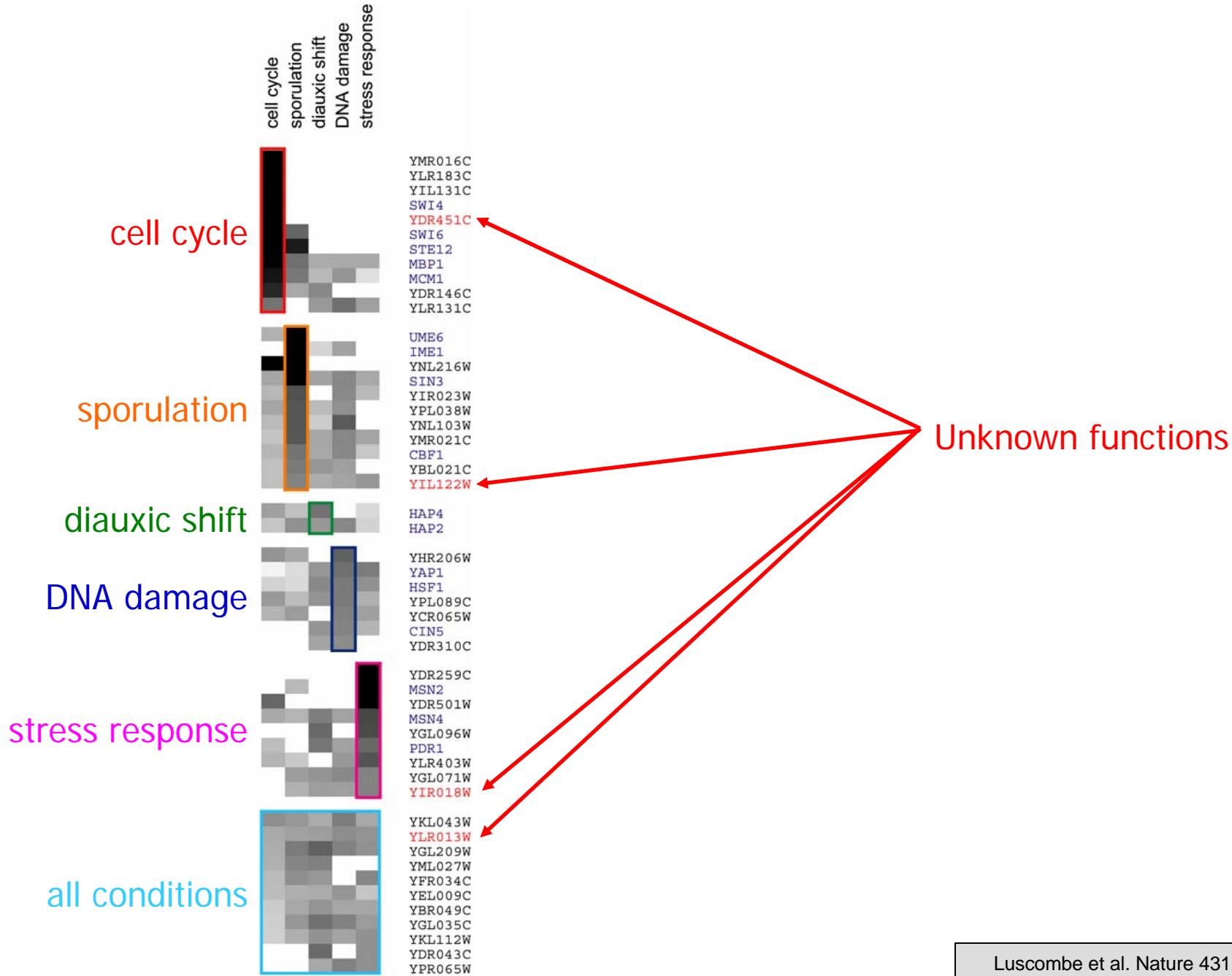


- Some permanent hubs
  - ◊ house-keeping functions
- Most are transient hubs
  - ◊ Different TFs become key regulators in the network
- Implications for condition-dependent vulnerability of network

Luscombe et al. Nature 431: 308



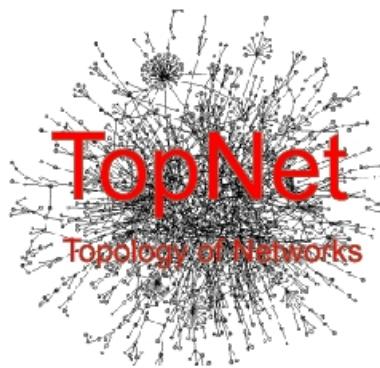
Luscombe et al. Nature 431: 308



Luscombe et al. Nature 431: 308

# Conclusions

- Developing Standardized Descriptions of Protein Function
  - ◊ Integration of Information for Prediction and Description
- Prediction of Essentiality
  - ◊ Build classifier to predict essentiality from diverse features in *S. cerevisiae*
  - ◊ Apply it to *S. mikatae* & show performance is as good as homology
- Describe Network Dynamics in Yeast Reg. Net
  - ◊ Merge expression data with net
  - ◊ Active network markedly different in different conditions
  - ◊ Identify transient hubs associated with particular conditions
  - ◊ Use these to annotate genes of unknown func.



## - an automated web tool

tYNA

(vers. 2 :  
"TopNet-like  
Yale Network Analyzer")

**tYNA**

Getting started API WSDL Download tYNA Installation guide Plugins for Cytoscape Contact Known problems

You are logged in as kevin Logout

List Owned Biological networks with (Attribute name) = (Attribute value) List

**Workspace manager**

Load an existing network

Load Into workspace 0 Categorized by Nil Load

Current working networks in your workspaces:

Workspace 0: statFilter(degrees, geq, 1, value, neighbors=false, intersection("Uetz 2000 yeast two hybrid", "Ito 2001 yeast two hybrid"))

Workspace 1: (empty)

Workspace 2: (empty)

Workspace 3: (empty)

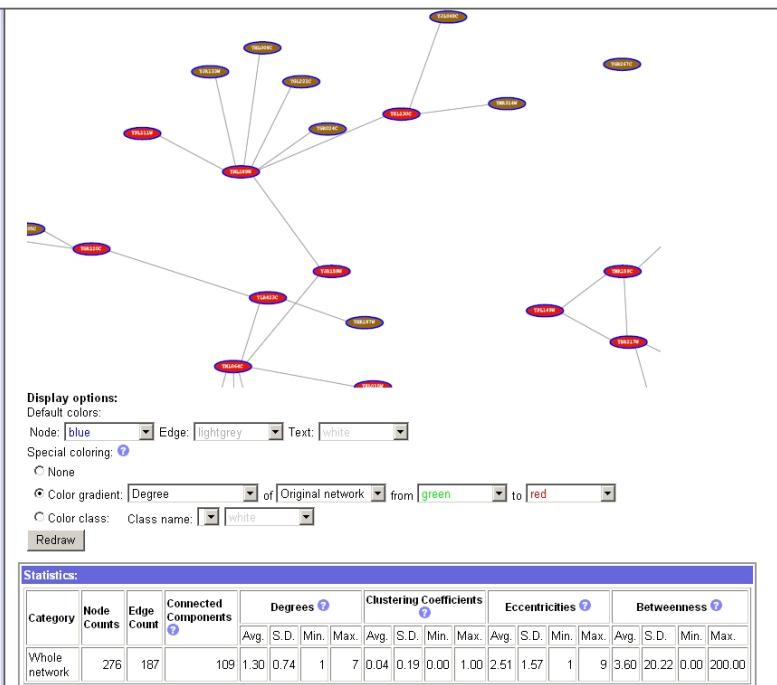
Multiple network analysis

**Networks in database (upload download)**

ID	Name	Creator	Creation date
14	Uetz 2000 yeast two hybrid	kevin	21-Feb-06
15	Ito 2001 yeast two hybrid	kevin	21-Feb-06
16	Ho 2002 pull down	kevin	21-Feb-06
17	Gavin 2002 pull down	kevin	21-Feb-06
18	Jansen 2003 PPI	kevin	21-Feb-06
19	MIPS yeast PPI	kevin	21-Feb-06
21	BIND yeast data	kevin	21-Feb-06
22	DIP yeast data	kevin	21-Feb-06
23	Kim 2006 structural interaction	kevin	21-Feb-06
24	Han 2004 FYI data	kevin	21-Feb-06
25	Luscombe 2004 regulatory	kevin	21-Feb-06

**Categories in database (upload download)**

ID	Name	Creator	Creation date
Whole network	276	187	109 1.30 0.74 1 7 0.04 0.19 0.00 1.00 2.51 1.57 1 9 3.60 20.22 0.00 200.00

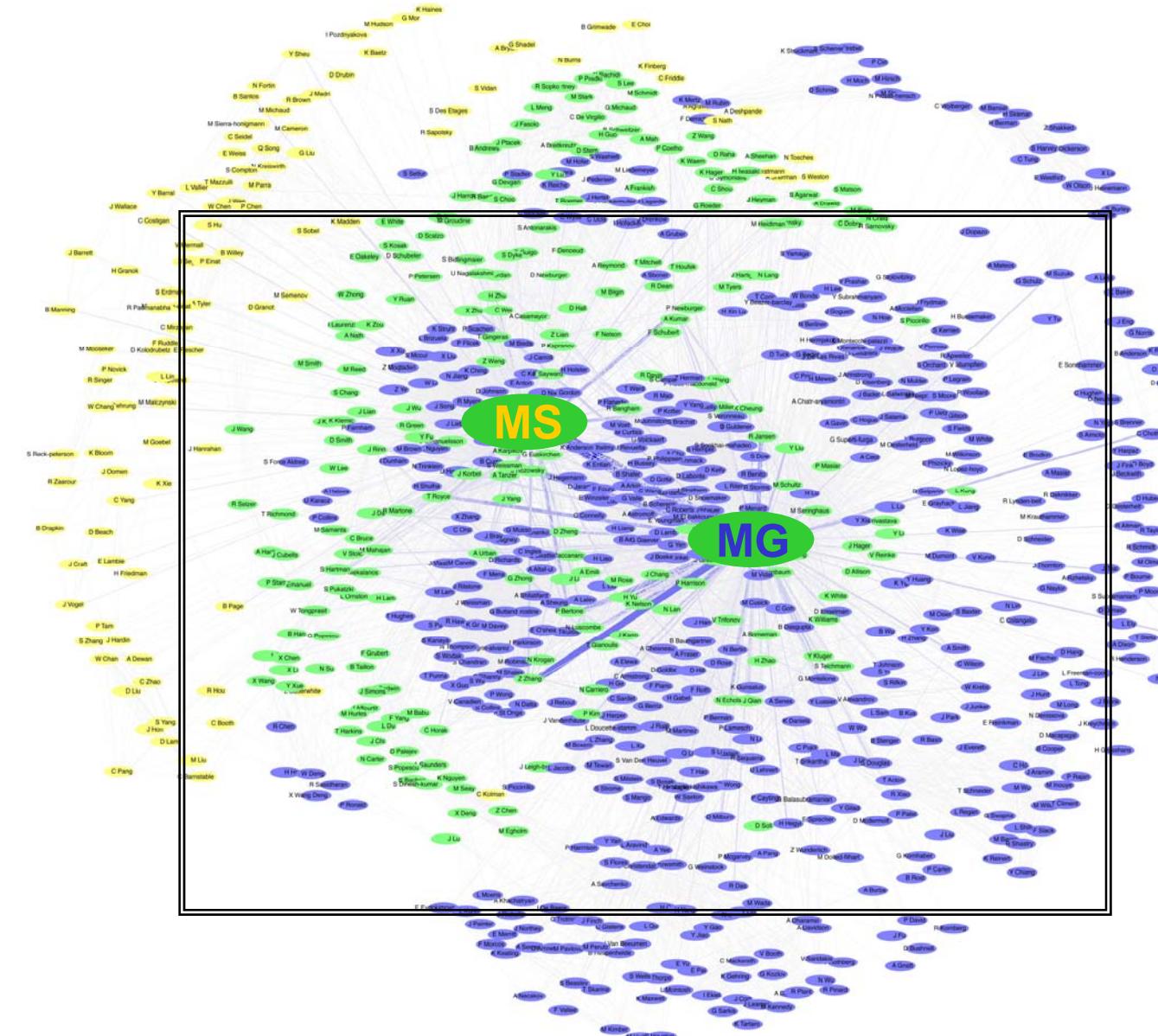


Normal website + Downloaded code (JAVA)  
+ Web service (SOAP) with Cytoscape plugin

[Yu et al., NAR (2004); Yip et al. Bioinfo. (2006);  
Similar tools include Cytoscape.org, Idekar, Sander et al]

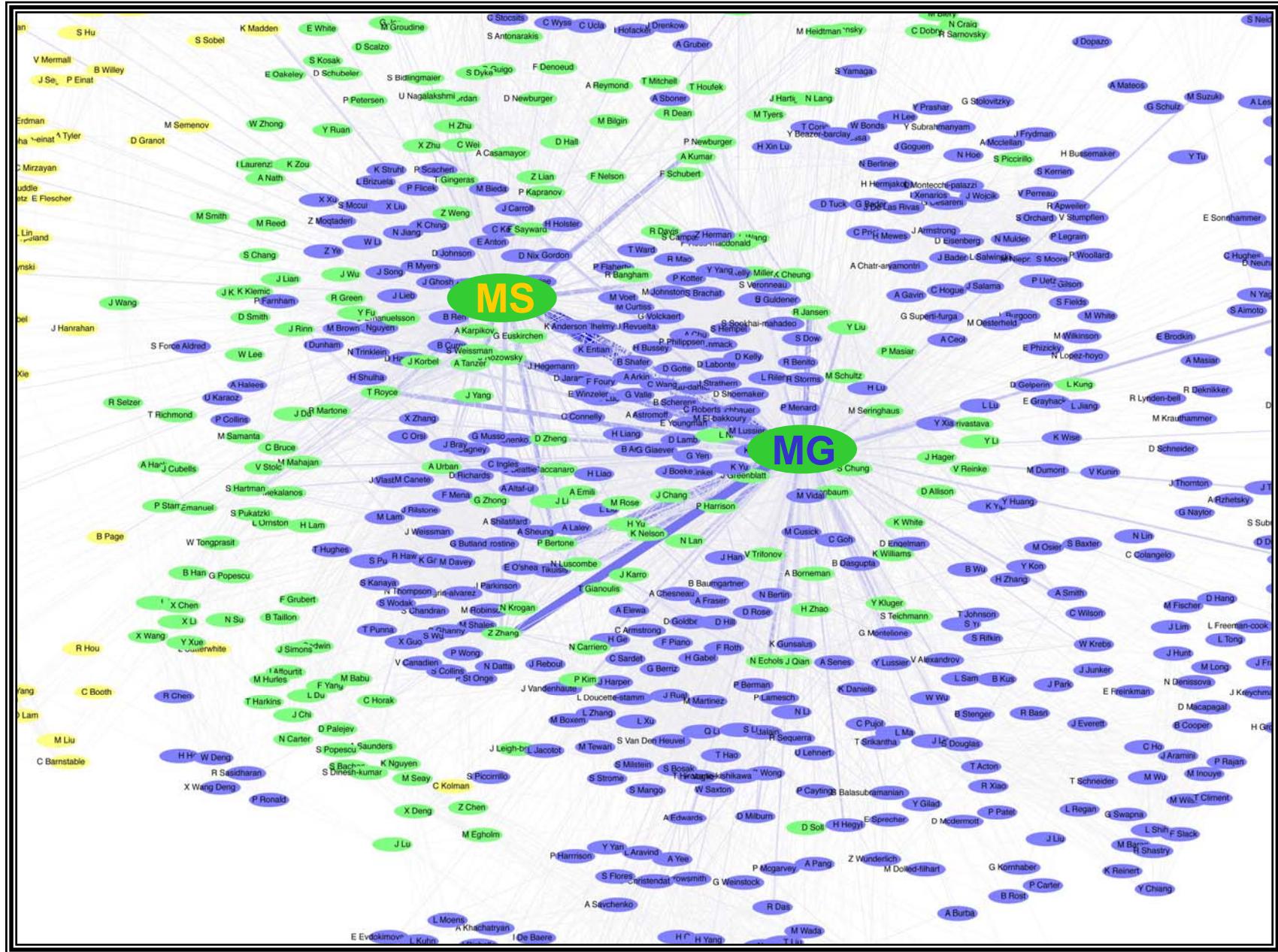
# Acknowledgements

## TopNet.GersteinLab.org



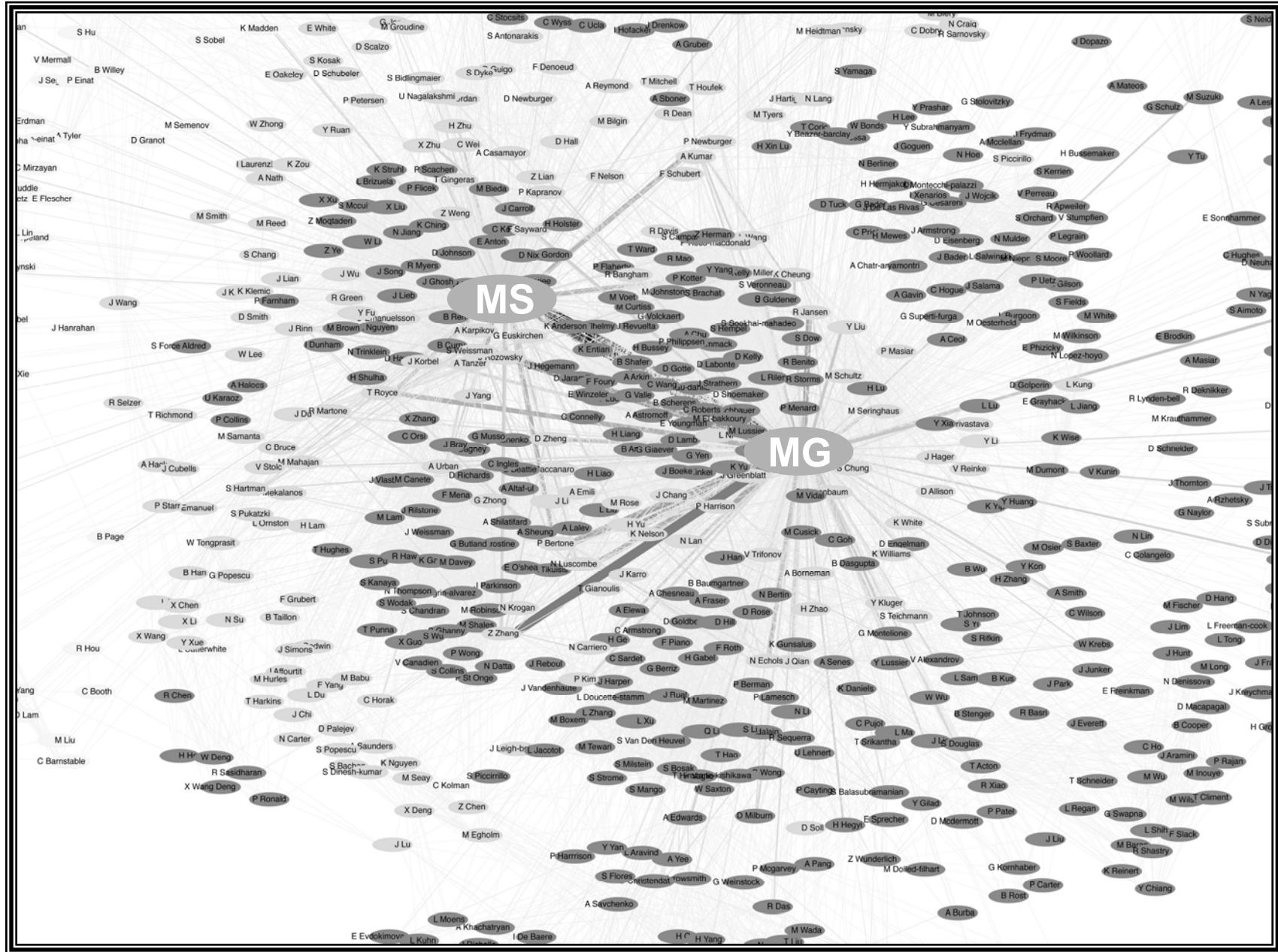
# Acknowledgements

## TopNet.GersteinLab.org



# Acknowledgements

## TopNet.GersteinLab.org



NIH, NSF, Keck

## Acknowledgements TopNet.GersteinLab.org

Job opportunities currently  
for postdocs & students

