

## Lecture II

# Identification of Key Causal Regulators in Gene Networks

**Bin Zhang, PhD**

**Jun Zhu, PhD**

Department of Genetics & Genomic Sciences

Icahn Institute of Genomics and Multiscale Biology

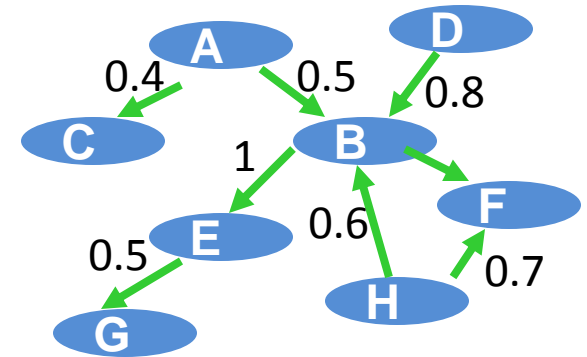
Icahn School of Medicine at Mount Sinai, New York, USA

Email: [bin.zhang@mssm.edu](mailto:bin.zhang@mssm.edu)

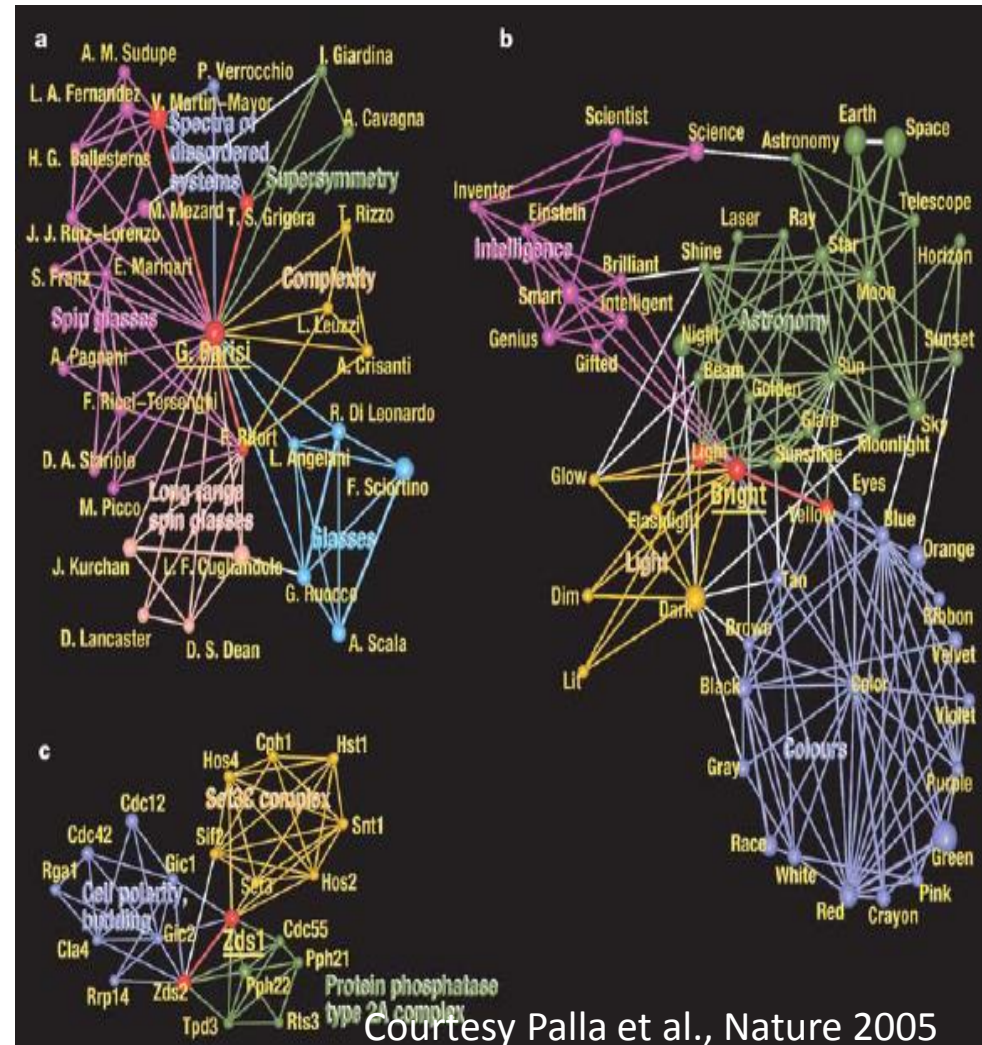
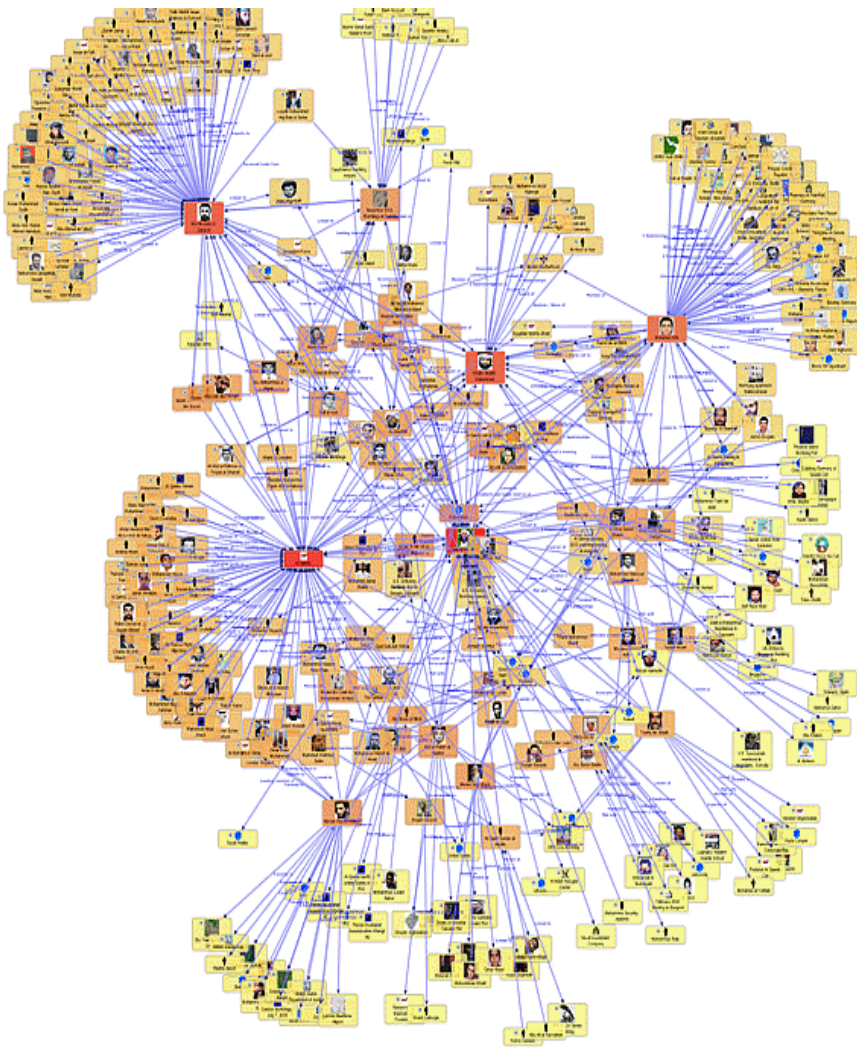
Web: <http://research.mssm.edu/multiscalenetwork>

# Basic Network Concepts

- Node
- Link
  - Directed or undirected
  - Weighted or unweighted
- Connectivity
  - degree, outdegree, indegree
- Centrality (<http://en.wikipedia.org/wiki/Centrality>)
  - degree, betweenness, closeness, eigenvector (PageRank)
- Module
  - a group of highly interacting nodes

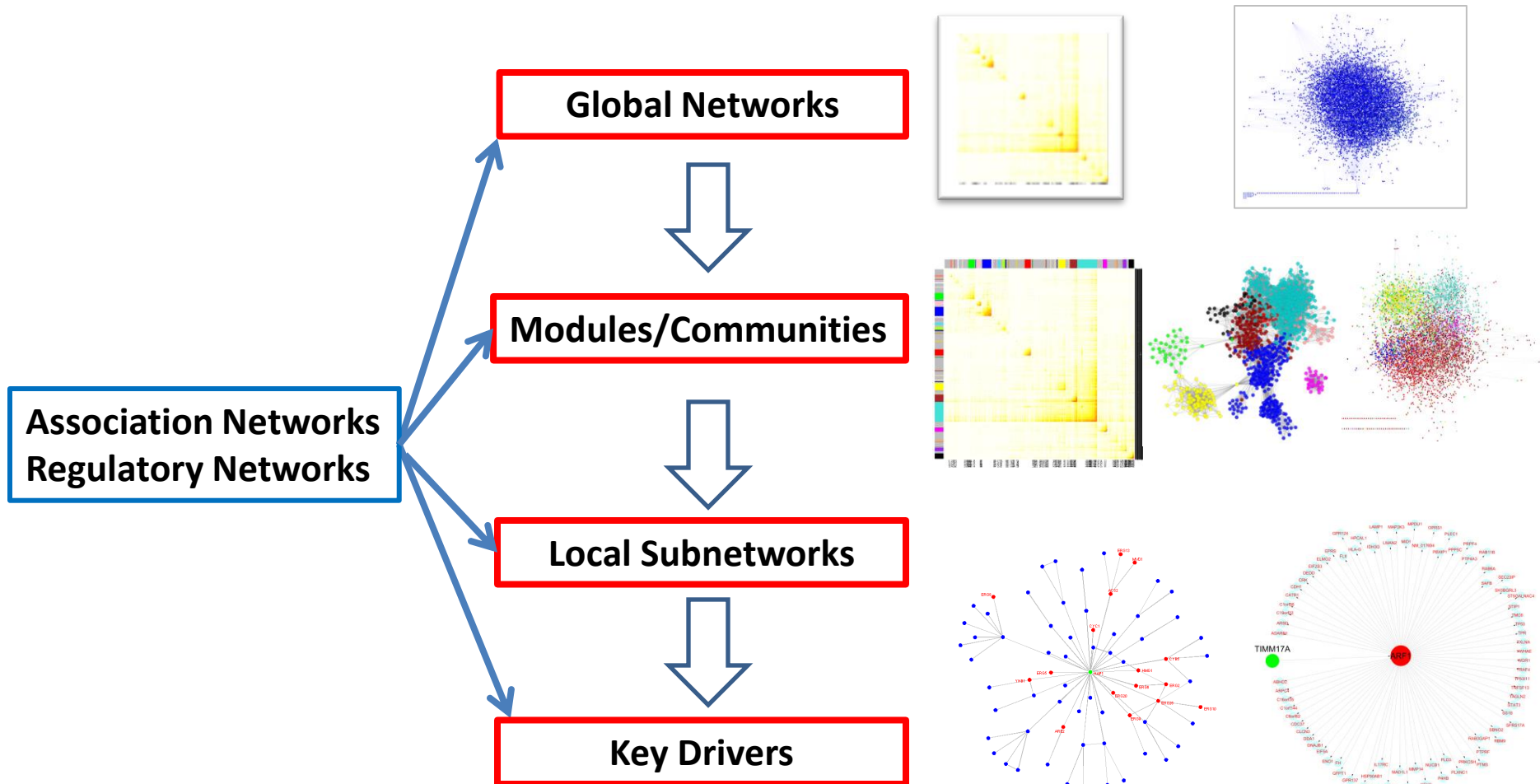


# What can we know from networks?



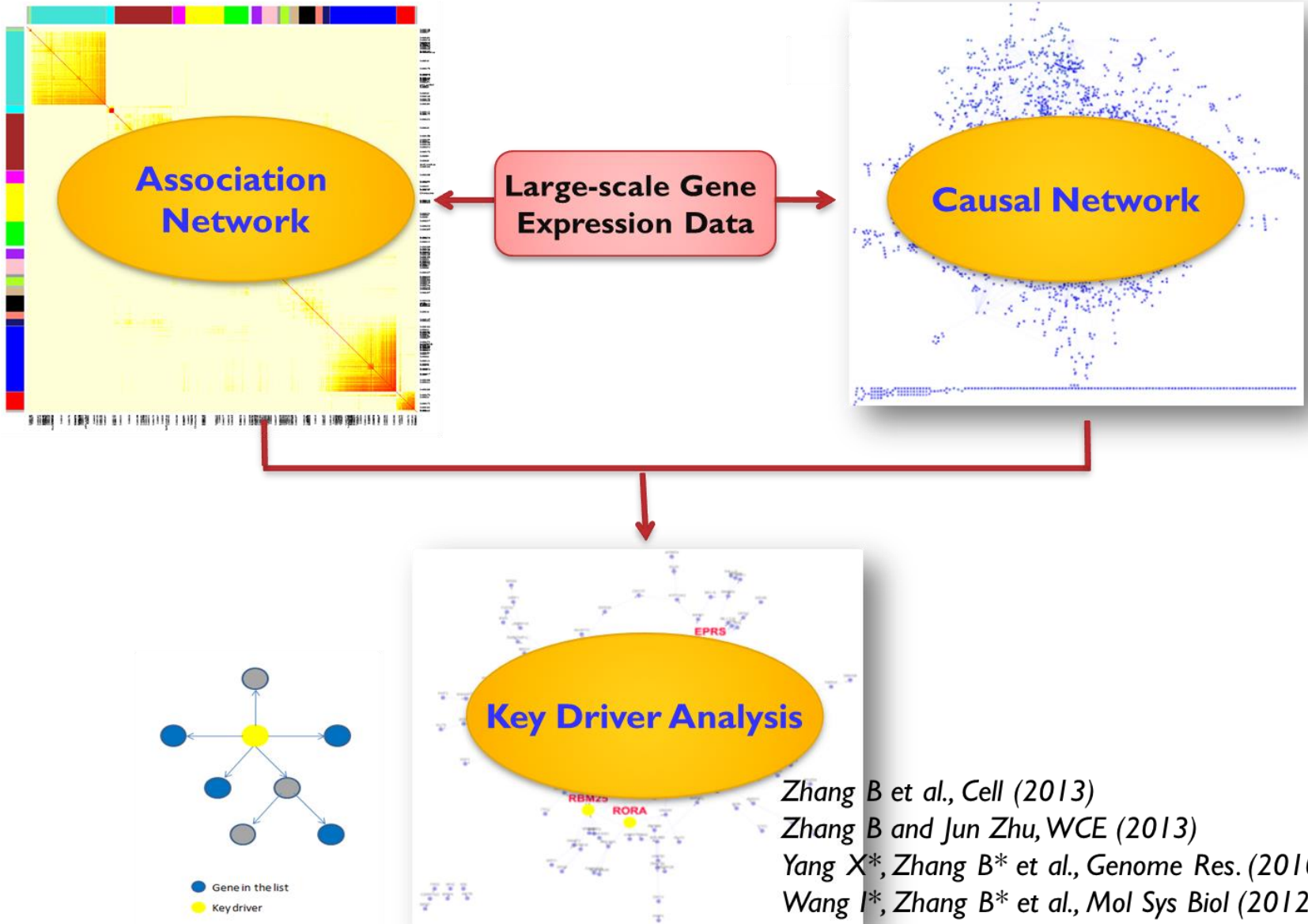
Courtesy Palla et al., Nature 2005

# Multiple Scales of Gene Networks

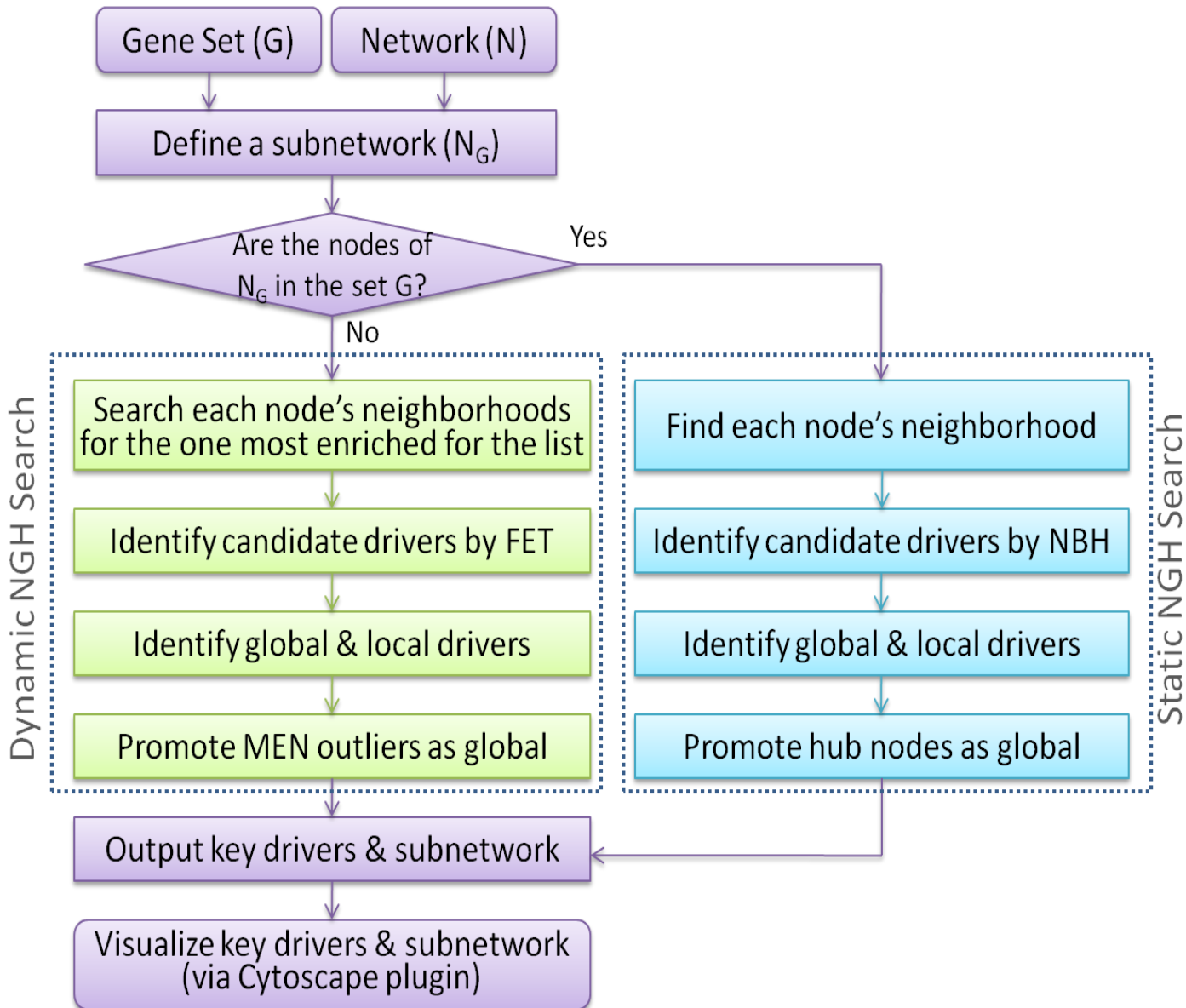




# Multiscale Gene Network Analysis



# Key Driver Analysis



# Dynamic Neighborhood Search (DNS)

- **Input**

- a set of nodes (**G**)
- a directed/undirected network (**N**)

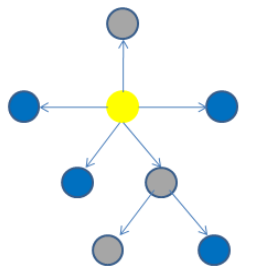
- **Procedure**

- Generate a sub-network  $N_G$ , defined as the set of nodes in  $N$  that are no more than  $h$ -layers away from the nodes in **G**.
- Search the  $h$ -layer neighborhood ( $h=1,..,H$ ) for each gene in  $N_G$  ( $HLN_{g,h}$ ) for the optimal  $h^*$ , such that

$$ES_{h^*} = \max(ES_{h,g}) \forall g \in N_G, h \in \{1..H\}$$

where  $ES_{h,g}$  is the computed enrichment statistic for  $HLN_{g,h}$

- A node becomes a candidate driver if its HLN is significantly enriched for the nodes in **G**
- Candidate drivers without any parent node (i.e., root nodes in directed networks) are designated as global drivers and the rest are local drivers.



● Gene in the list  
● Key driver

Enrichment Test

# Static Neighborhood Search (SNS)

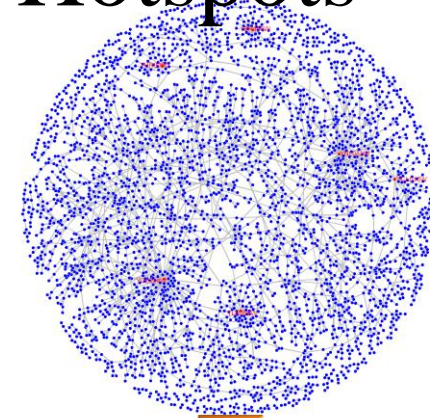
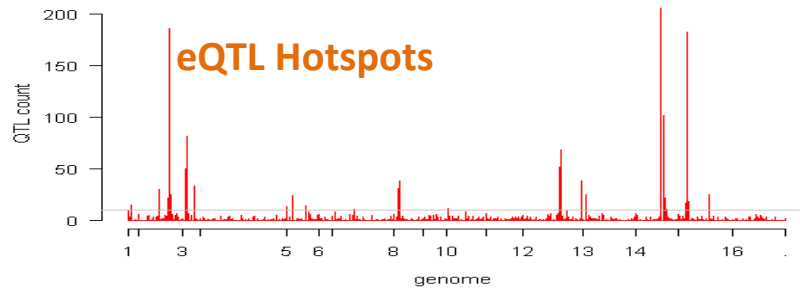
- **Input**
  - a set of nodes (**G**)
  - a directed/undirected network (**N**)
- **Procedure**
  - Compute the size of the h-layer neighborhood (HLN) for each node.
  - Let  $\mu$  be the sizes of HLN and  $d$  be the out-degrees for all the nodes
  - Nodes with HLN sizes greater than  $\bar{\mu} + \sigma(\mu)$  are considered as candidate drivers
  - Candidate drivers without any parent node (i.e., root nodes in directed networks) are designated as global drivers and the rest are local drivers.
  - Promote hub nodes nodes with out-degrees above  $\bar{d} + 2\sigma(d)$  as global drivers



# Application I: Key Drivers of eQTL Hotspots

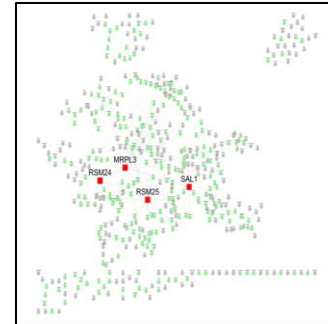
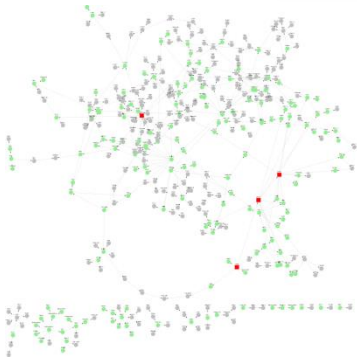
- A genotypic and expression data from a yeast cross of 112 segregants constructed from the BY and RM strains of *S. cerevisiae*
- Expression quantitative trait loci (eQTL) analysis identified 13 chromosomal regions harboring a large number of eQTL, i.e., eQTL hot spots.
- A Bayesian network reconstructed by integrating genotypic, gene expression, protein-protein interaction and transcription factor binding site (TFBS) data remains the most predictive (Zhu J, Zhang B et al., Nat Genet 2008)
- We apply DNS to identify key drivers of the eQTL hotspots, as oppose to the original static search.

# Key Drivers of eQTL Hotspots

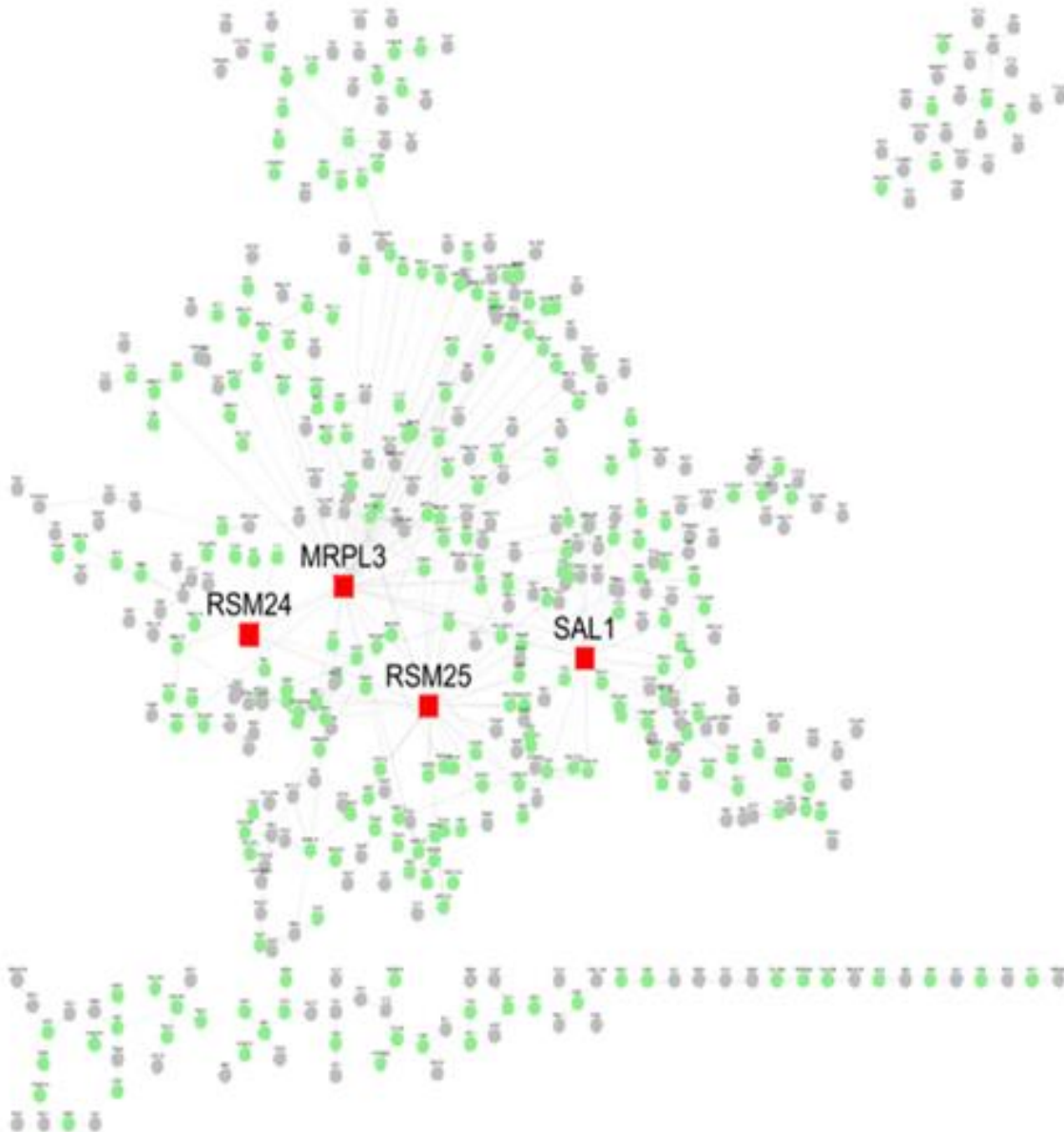


## Key Driver Analysis

eQTL	Hotspot	Hotspot base-pair	the original KDA (Zhu, Zhang et al. 2008)	KDA L1	KDA L2	KDA L3
hotspot	chr.	position				
2	2	560000	TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, AMN1	TBS1, ARA1, CSH1, SUP45, CNS1, PWP2	TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, ENP2, NOP7	TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, NMD3, RPF1
4	3	1.00E+05	LEU2, ILV6, NFS1, CIT2, MATALPHA1	LEU2, BAP2, OAC1	BAP2, LEU2, OAC1, RTG3	LEU2, BAP2, OAC1, RTG3
5	3	230000	MATALPHA1			
6	5	130000	URA3	URA3	URA3	URA3
7	8	130000	GPA1	GPA1	GPA1	GPA1
8	12	680000	HAP1	HAP1	HAP1	HAP1
9	12	107000	YRF1-4, YRF1-5, YLR464W	YRF1-4	YRF1-4	YRF1-4
11	14	503000	SAL1, TOP2	SAL1, RSM24, RSM25	SAL1, RSM24, RSM25, MRPL3	SAL1, RSM24, RSM25, MRPL3
12	15	180000	PHM7	TFS1, PHM7, TKL2,  YGR052W	PHM7, TFS1, YGR043C, HXT7, TKL2, GDB1, YGR052W	PHM7, TFS1, YGR043C, PIL1, TKL2, HXT7, YGR052W, GDB1
10	13	70000		GCV1	GCV1	GCV1
13	15	590000		ATP5	ATP20	ATP5, ATP20



# New Key Drivers for eQTL Hotspot 11



- SAL1 is the only one predicted by the original approach
- KDA identified three new regulators, RSM24, RSM25 and MRPL3
- 98 genes on the Hotspot 11 are the downstream of SAL1 ( $p < 7e-95$ )
- 142 genes on the Hotspot 11 are the downstream of RSM25 ( $p < 1.42e-114$ )

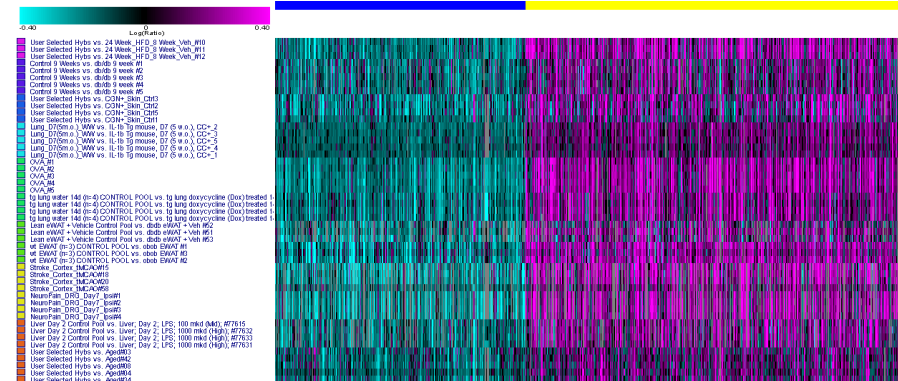
# Key Drivers of eQTL Hotspot 12



- PHM7 is the only regulator identified by the original approach.
- KDA uncovered 6 new regulators.
- The neighborhoods of TFS1, YGR043C, TKL2 and YGR052W are more significantly enriched for the genes links to the hot spot than that of PHM7.

# Application II: Inflammatome Driver Genes

Disease	Model	Species	Tissue profiled	# of Cases	# of Controls	# of Total Arrays
Asthma	OVA	Mouse	Lung	5	4	9
COPD	IL-1b Tg	Mouse	Lung	5	3	8
Fibrosis	TGFb Tg	Mouse	Lung	4	4	8
Atherosclerosis	ApoE KO HFD	Mouse	Aorta	3	3	6
Diabetes	db/db	Mouse	Adipose	3	3	6
Diabetes	db/db	Mouse	Islet	5	5	10
Obesity	ob/ob	Mouse	Adipose	3	3	6
Multiple	LPS	Rat	Liver	4	4	8
Stroke	MCAO	Rat	Brain	4	4	8
Neuropathic pain	Chung	Rat	DRG	4	4	8
Inflammation pain	CGN	Rat	Skin	4	5	9
Sarcopenia	Aged vs. Young	Rat	Muscle	5	5	10



Consistency of up- and down-regulated genes in 12 disease models.

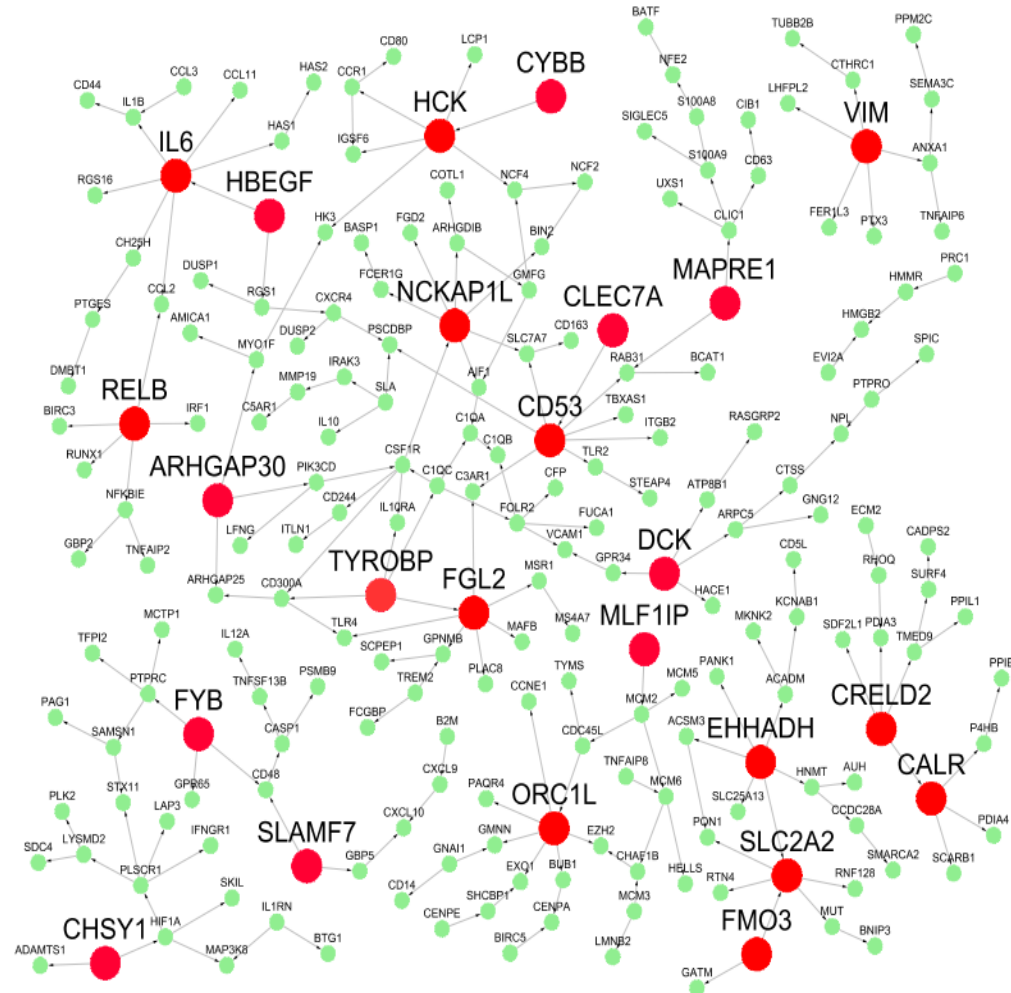
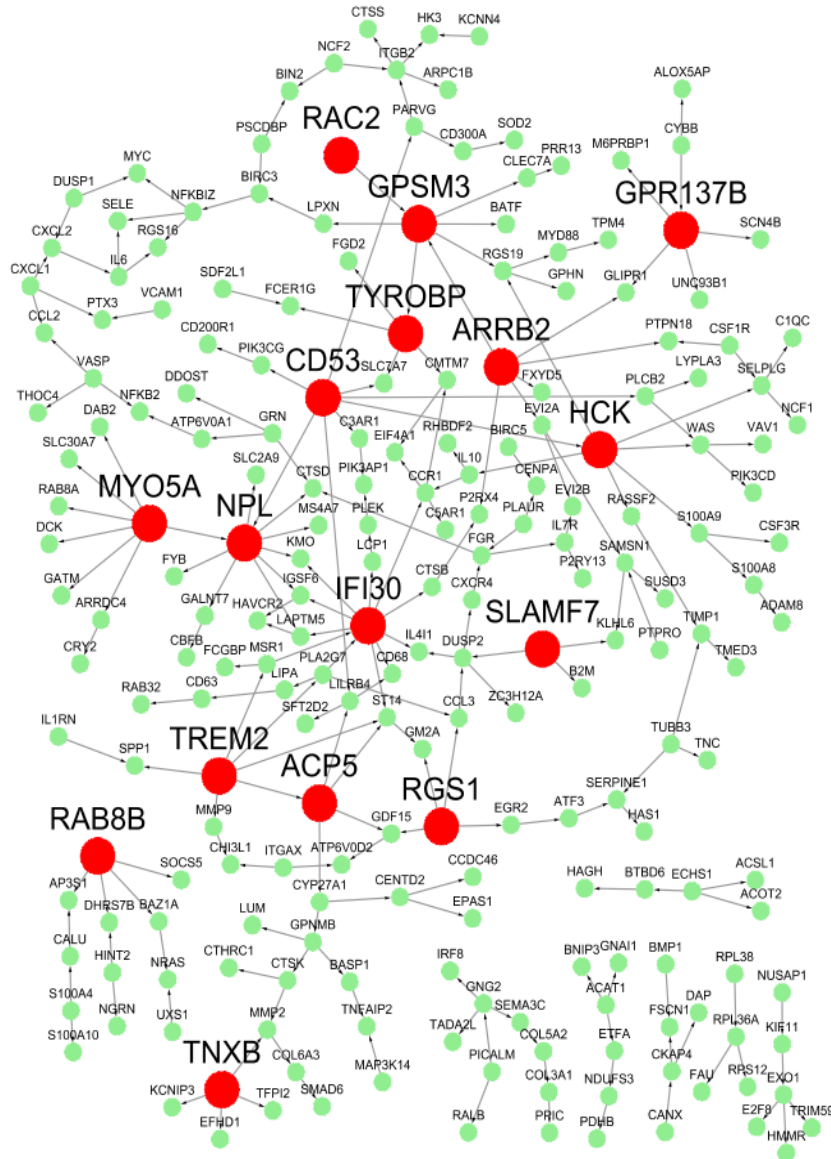
Up-regulated	Accumulated # of gene	Down-regulated	Accumulated # of gene
All 12 models	83	All 12 models	36
>= 11 models	303	>= 11 models	171
>= 10 models	614	>= 10 models	412
>= 9 models	939	>= 9 models	639
>= 8 models	1,193	>= 8 models	810
>= 7 models	1,357	>= 7 models	925

Gene Ontology Categories enriched in the Inflammatome Signature.

Similar Set: Up-regulated	Expectation	Overlap	Set	Similar Set: Down-regulated	Expectation	Overlap	Set
inflammatory response	4.76E-61	208	704	transmission of nerve impulse	3.32E-11	78	639
leukocyte activation	2.13E-32	164	704	Valine, leucine & isoleucine	1.34E-08	18	42
regulation of immune response	1.44E-25	84	260	degradation	4.03E-06	68	661
cytokine production	6.10E-18	85	335	carboxylic acid metabolic process	1.30E-05	31	198
Chemotaxis	4.97E-16	74	284	cofactor metabolic process	9.18E-05	57	554
humoral immune response	3.25E-14	69	271	generation precursor	0.000122	16	65
				metabolites/energy			
				fatty acid catabolic process			

Wang IM\*, Zhang B\* et al. (2012)  
Systems Analysis of Eleven Rodent Disease Models Reveals an Inflammatome Signature and Key Drivers.  
*Molecular Systems Biology* 8:594

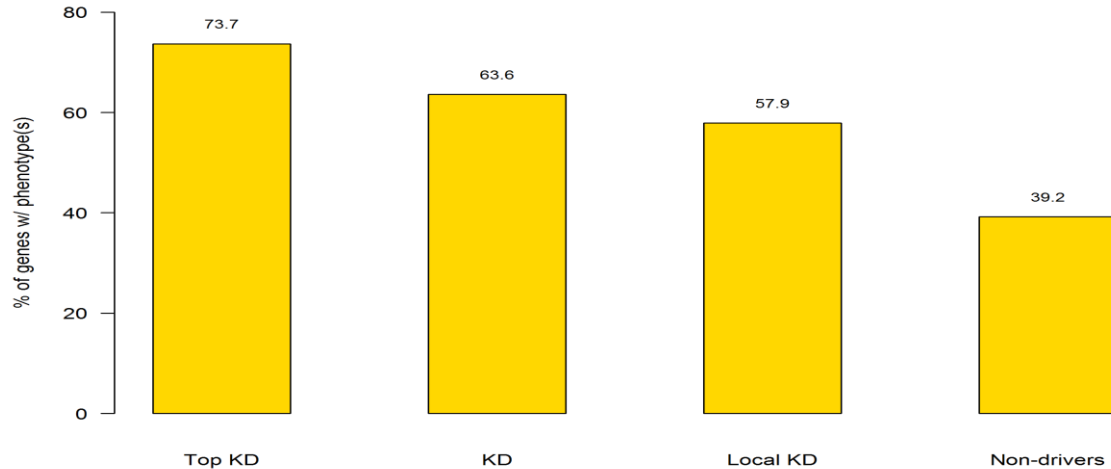
# Inflammatome Networks in Human Liver and Adipose



Wang IM\*, **Zhang B\*** et al. (2012) Systems Analysis of Eleven Rodent Disease Models Reveals an Inflammatome Signature and Key Drivers. *Molecular Systems Biology* 8:594



# Inflammatome Signature and Drivers versus MGI Phenotype Database

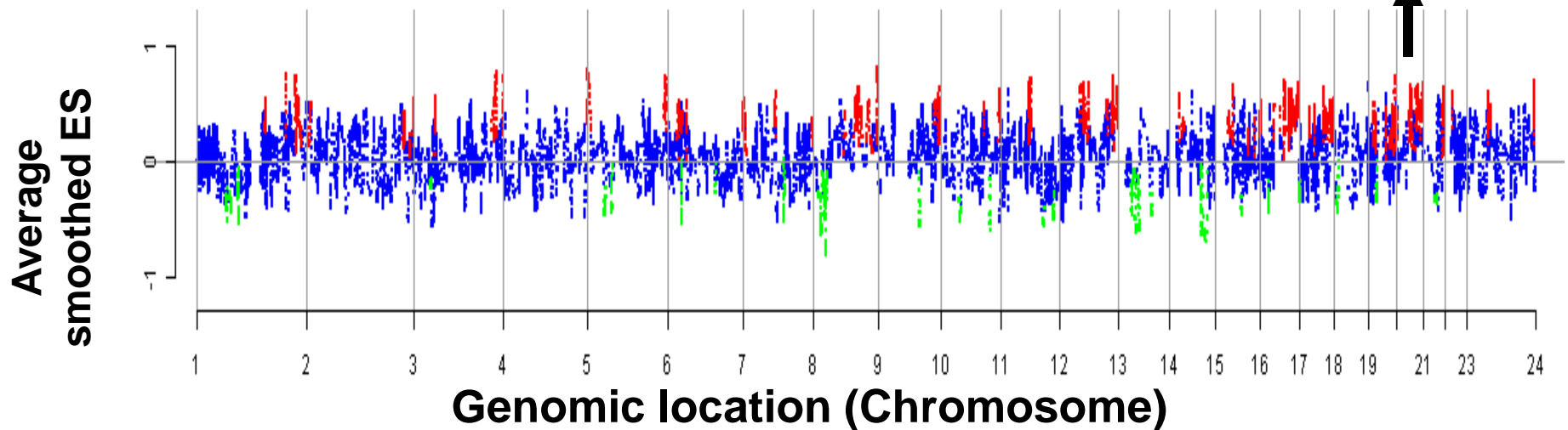
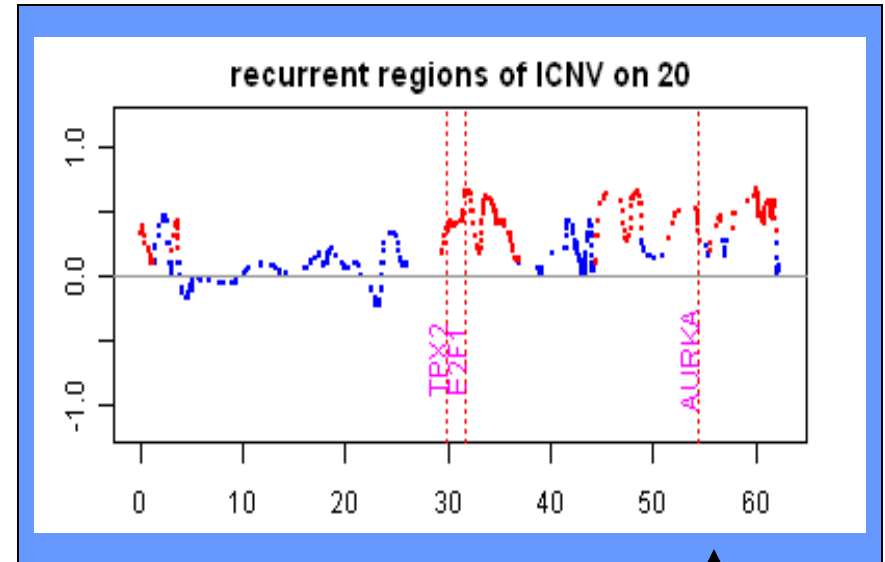


Group	No. of genes	No. of gene tested in the MGI phenotype database	No. of genes with MGI phenotype(s)	% tested genes with phenotype(s)
top 55 key drivers	55	19	14	73.7
key drivers	151	44	28	63.6
local drivers	212	57	33	57.9
non-drivers	2098	609	239	39.2

Mouse Genome Informatics database (MGI; <http://ftp.informatics.jax.org/pub/reports/index.html#pheno>)

# Application III: Drivers of Cancer Genomic Alterations

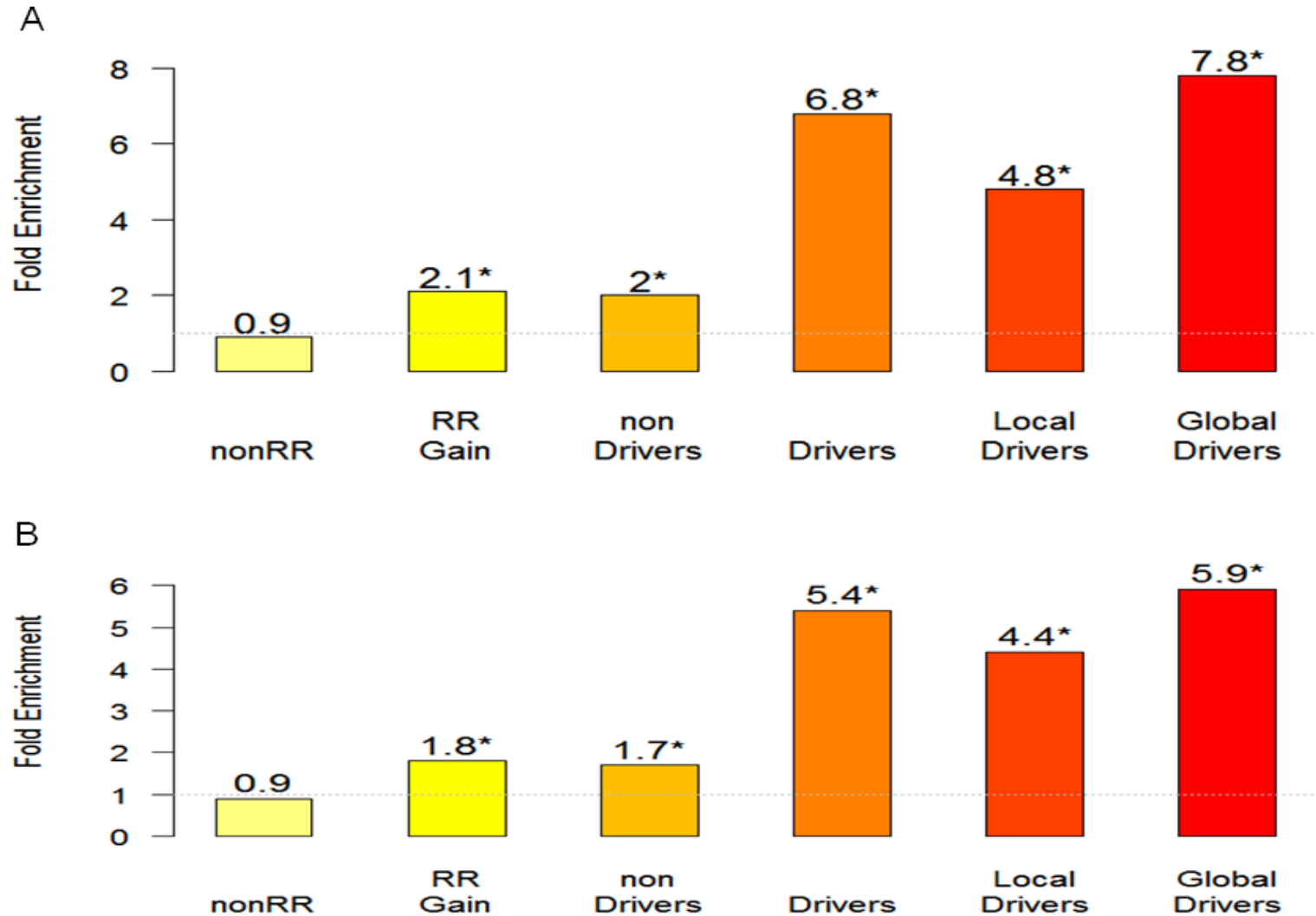
- Tumor outcome:
  - Good: free metastasis for  $\geq 5$  years
  - Bad: metastasis for  $< 5$  years
- 109 recurrent regions of inferred CNV identified from 4 data sets



al Driver  
al Driver (validated)  
al Driver  
al Driver (validated)

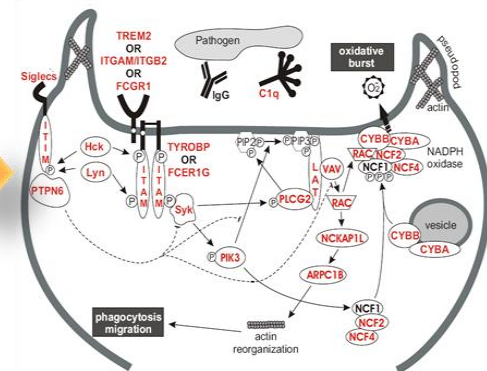
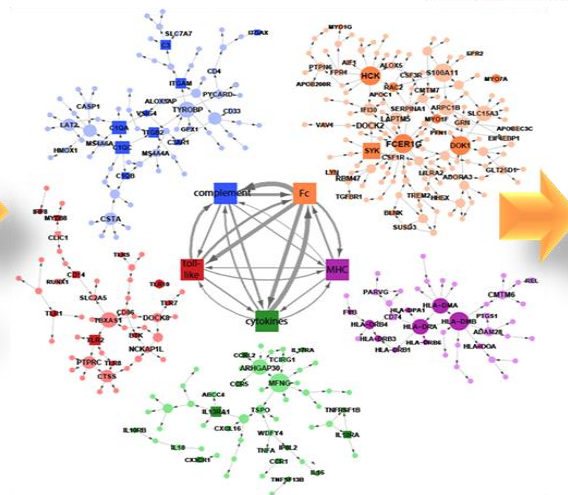
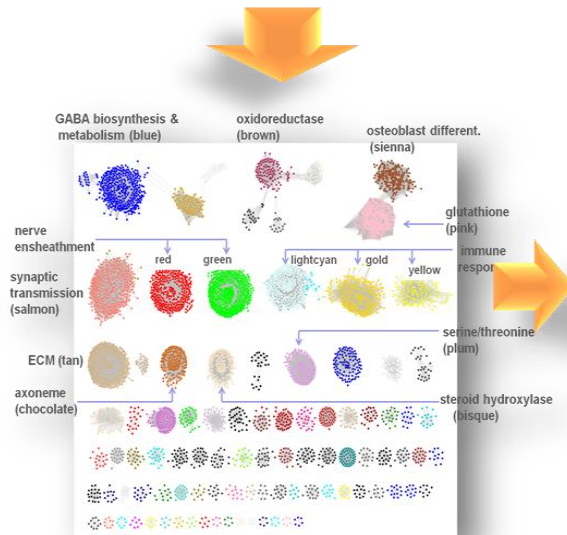
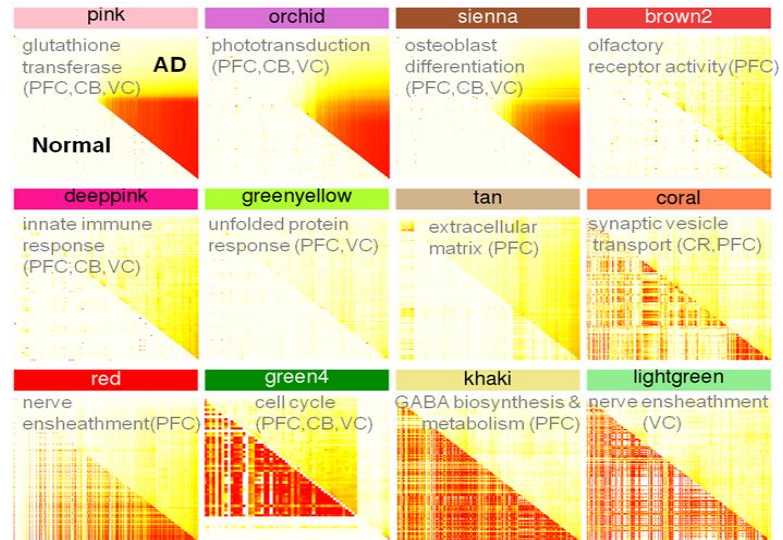
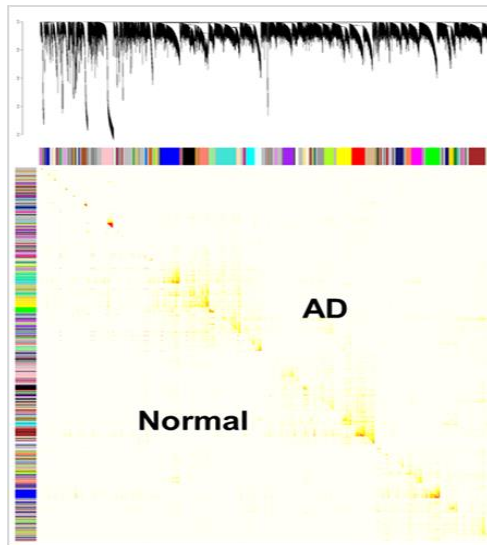
TPX2
AURKA
GIN52
COX4NB
TIMM17A
ACBD3
TK1
NUP93
BIRC5
KARS
NIP7
CIAPIN1
ARF1
SAP30BP
UBE2S
ARHGDI2A
RNASEH2A
SLC9A3R1
MRP57
PUF60
NOL11
CDC45L
YTHDF1
DHX38
HIST1H2BF
SLC25A19
HIST1H2BE
HIST1H2BH
ATP6V1C1
MTDH
YWHAZ
ERBB2

# Validation of Causal Drivers



# Application IV:

## Key Driver Genes of Alzheimer's Disease



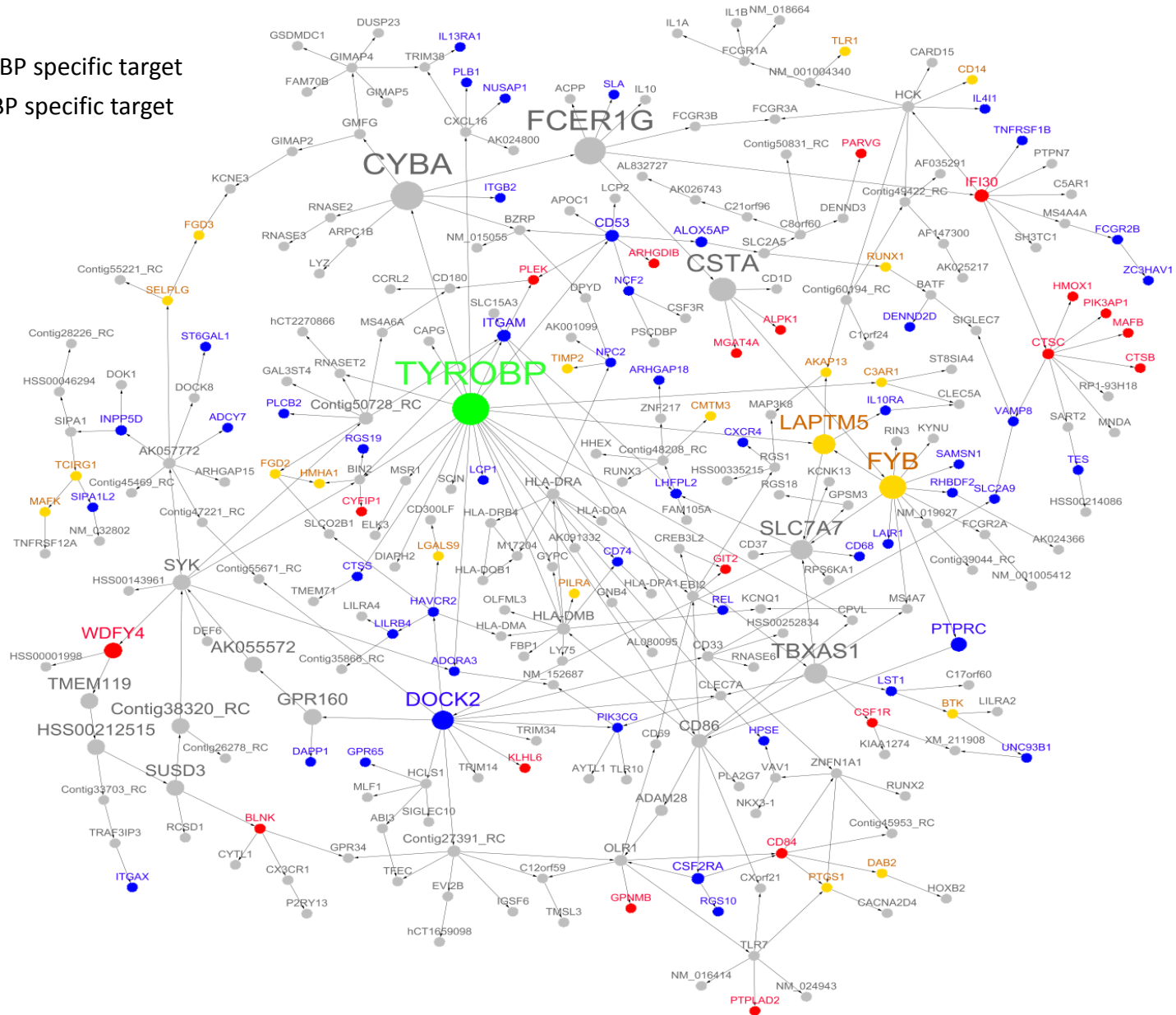
# Relevance of the Predicted Drivers to AD

- The 13 well known AD susceptibility genes (Bertram, McQueen et al. 2007)
- Seven (7) of the 13 known AD susceptibility genes were included in the multi-tissue network
- Three (3) of the 7 genes, CST3, PSEN1 and TF, are the predicted drivers, representing a 5.5-fold enrichment ( $P=1.06e-3$ ) while the rest four are not drivers, i.e., they are underrepresented in the non-driver genes (0.62 fold-enrichment,  $P=0.99$ ).
- The predicted drivers are 9 times more likely to be the known AD susceptibility genes than the non-driver genes.



# Validation of TYROBP Networks

- full-length TYROBP specific target
- truncated TYROBP specific target
- common target



# Summary

- An algorithm (KDA) was developed to identify key drivers of biological networks based on various centrality measurements
- Key regulators predicted by KDA appear to be biologically more important than non-drivers
- Many key regulators predicted by KDA have been validated at various stages in complex human diseases
- More comprehensive network analysis methods need be explored to further understand the complexity of biological networks and their underlying biology

# Acknowledgements

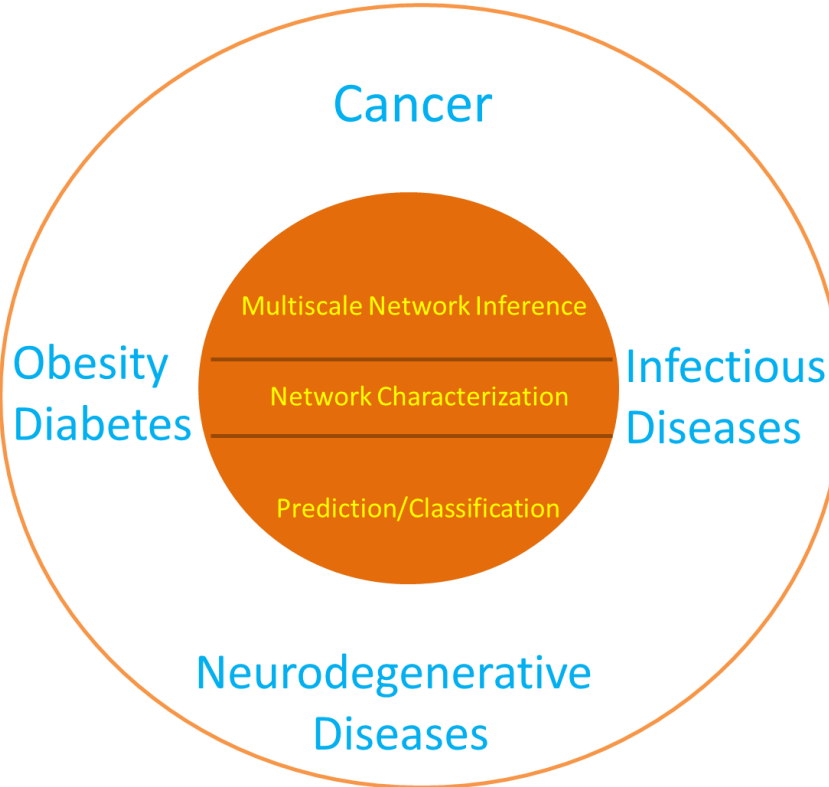
Sage Bionetworks

- Justin Guinney
- Dave A Henderson



# Multiscale Network Modeling Laboratory

<http://research.mssm.edu/multiscalenetwork>



## Openings for

Postdoctoral Fellow &  
Senior Scientist Positions in  
Computational Neuroscience &  
Cancer Biology

<http://research.mssm.edu/multiscalenetwork/Opportunities.html>



**Dr. Bin Zhang (PI)**  
Assoc. Professor

## Group Members



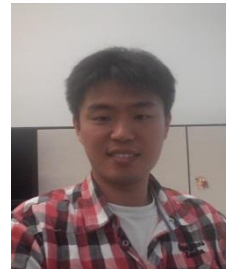
**Dr. Christian Forst**  
Assist. Professor



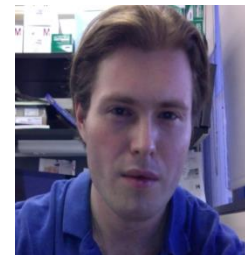
**Dr. Yongzhong Zhao**  
Senior Scientist



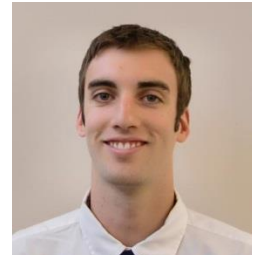
**Dr. Xiandong Lin**  
Scientist



**Dr. Won-min Song**  
Postdoc Fellow



**Igor Katsyv**  
PhD Student



**Andrew McKenzie**  
PhD Student

# Stay Connected with Us!



@IcahnInstitute

