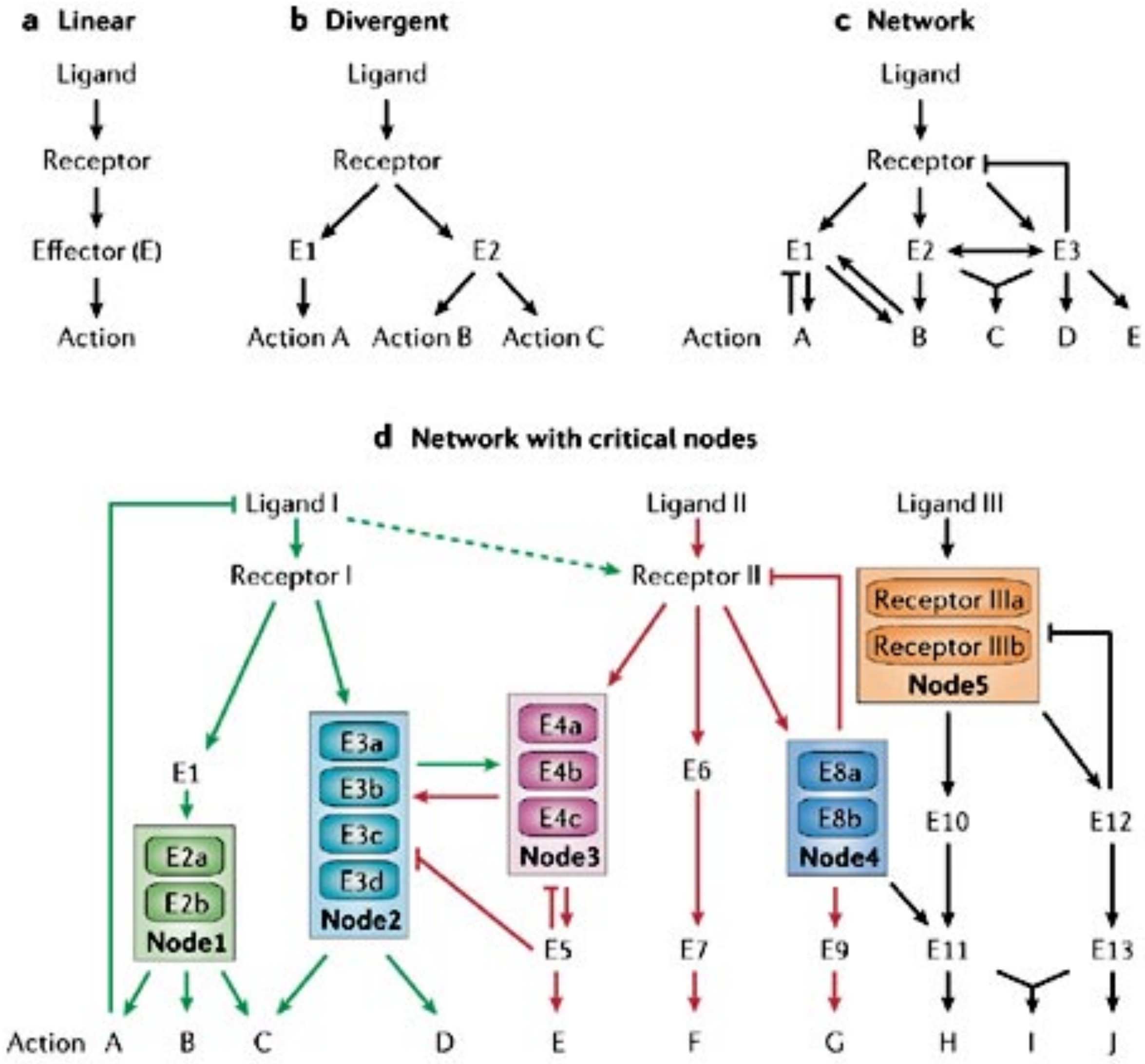
Biomedical Informatics
3184 Graves Hall 333 W. 10th Ave. Columbus, OH 43210 614-292-4778 (phone) 614-292-7659 (fax)

Introduction to Biological Networks

Kun Huang
Department of Biomedical Informatics
The Ohio State University

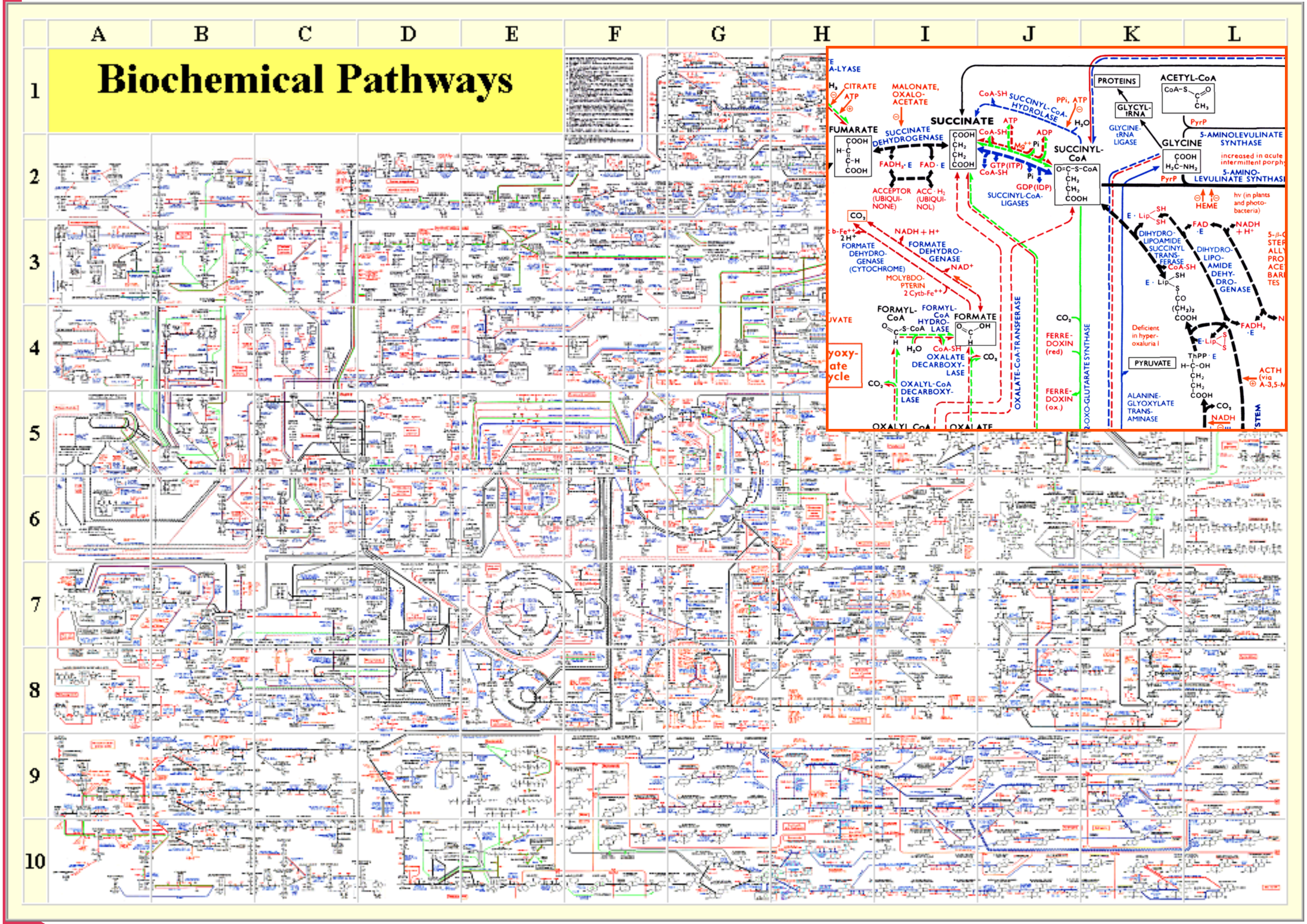
Outline

- Introduction to biological networks
  - Different kind of networks
- How to build networks
  - Experimental methods
  - Computing methods
    - Example 1 – Gene co-expression
  - Informatics methods
    - Example 2 – STRING
- Databases and resources
- Analyzing and mining the networks

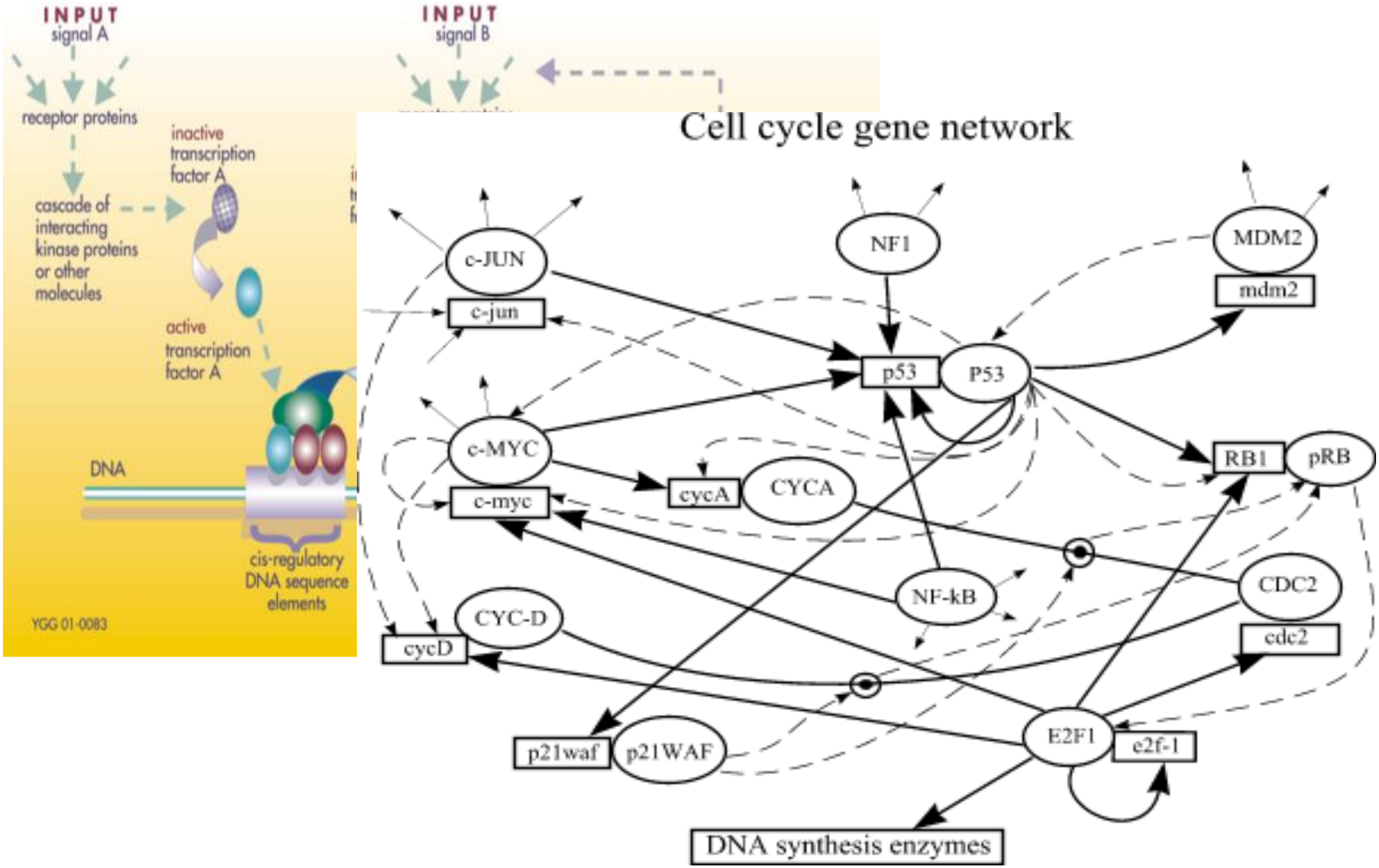


Nature Reviews | Molecular Cell Biology

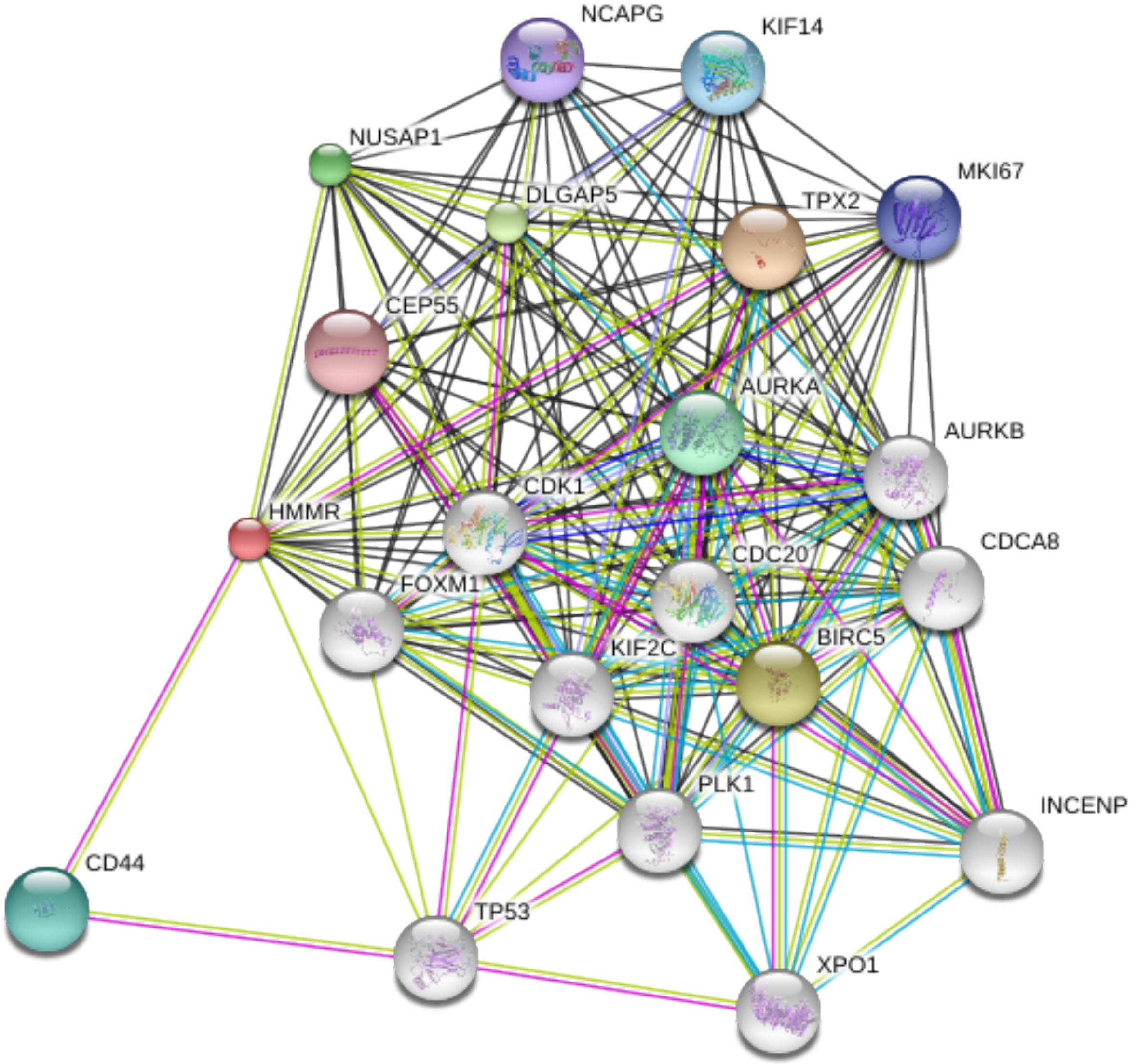
Taniguchi *et al.* *Nature Reviews Molecular Cell Biology* **7**, 85–96 (February 2006) | doi:10.1038/nrm1837


Biochemical Pathways
A
B
C
D
E
F
G
H
I
J
K
L
1
2
3
4
5
6
7
8
9
10
CITRATE
MALONATE, OXALOACETATE
SUCCINATE
FUMARATE
SUCCINATE DEHYDROGENASE
SUCCINYL-CoA HYDROLASE
SUCCINYL-CoA
SUCCINYL-CoA-LIGASES
PROTEINS
ACETYL-CoA
GLYCYL-tRNA
GLYCINE-tRNA LIGASE
GLYCINE
5-AMINOLEVULINATE SYNTHASE
5-AMINOLEVULINATE SYNTHASE
HEME
FORMATE DEHYDROGENASE (CYTOCHROME)
FORMATE DEHYDROGENASE
FORMYL-CoA
FORMYL-CoA HYDROLASE
FORMATE
OXALATE DECARBOXYLASE
OXALYL-CoA DECARBOXYLASE
OXALATE-CoA-TRANSFERASE
2-OXO-GLUTARATE SYNTHASE
FERREDOXIN (red)
FERREDOXIN (ox.)
DIHYDROLIPOAMIDE SUCCINYL TRANSFERASE
DIHYDROLIPOAMIDE DEHYDROGENASE
PYRUVATE
ALANINE-GLYOXYLATE TRANSAMINASE
Deficient in hyperoxaluria I
ACTH
OXALYL-CoA
OXALATE

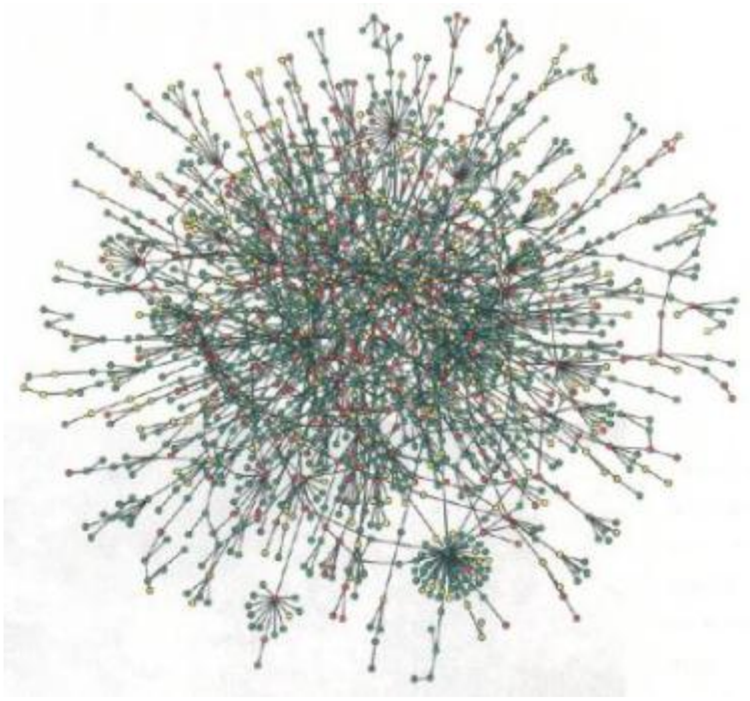
Gene Regulatory Network

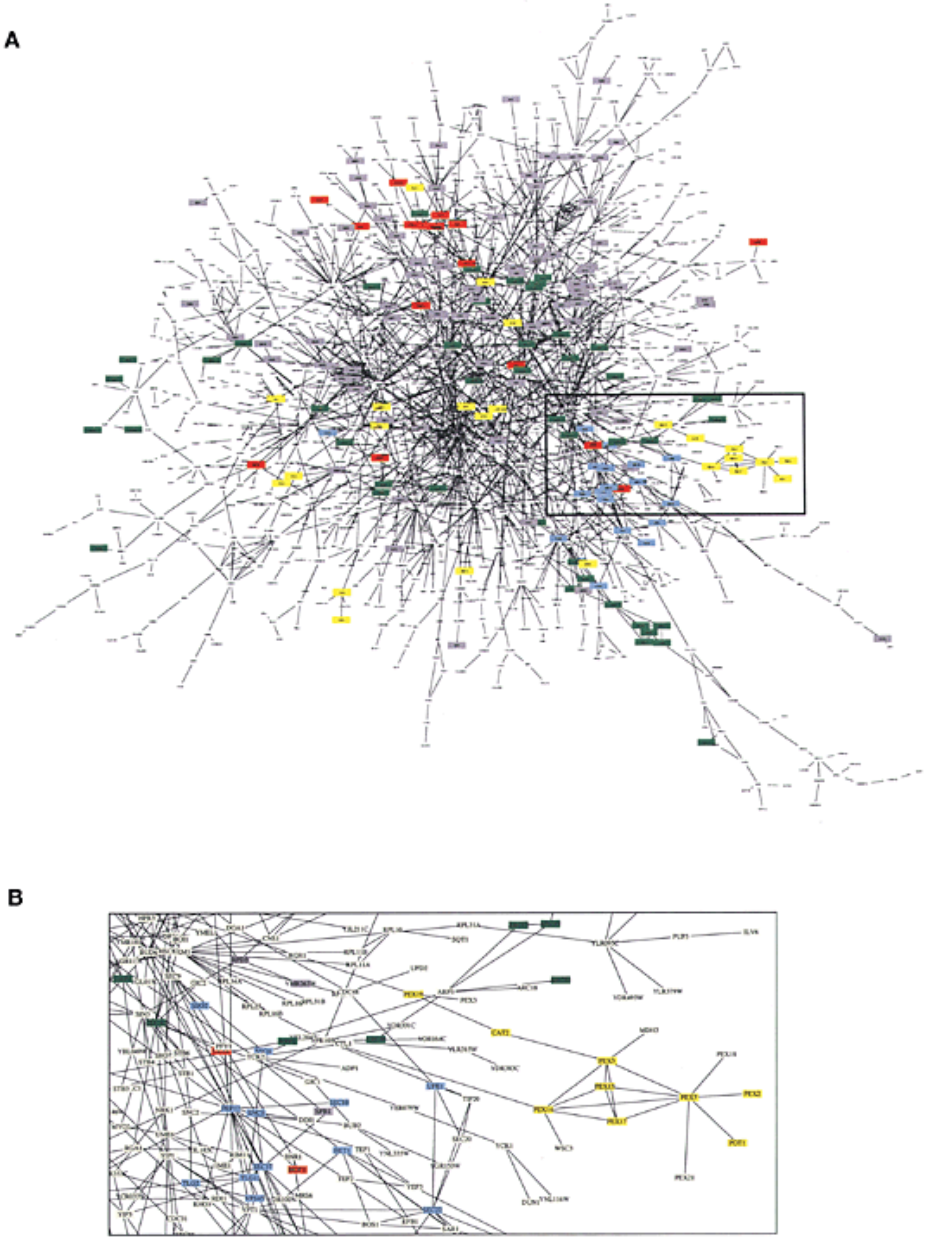


Protein-Protein Interaction Network



Wexner Medical Center



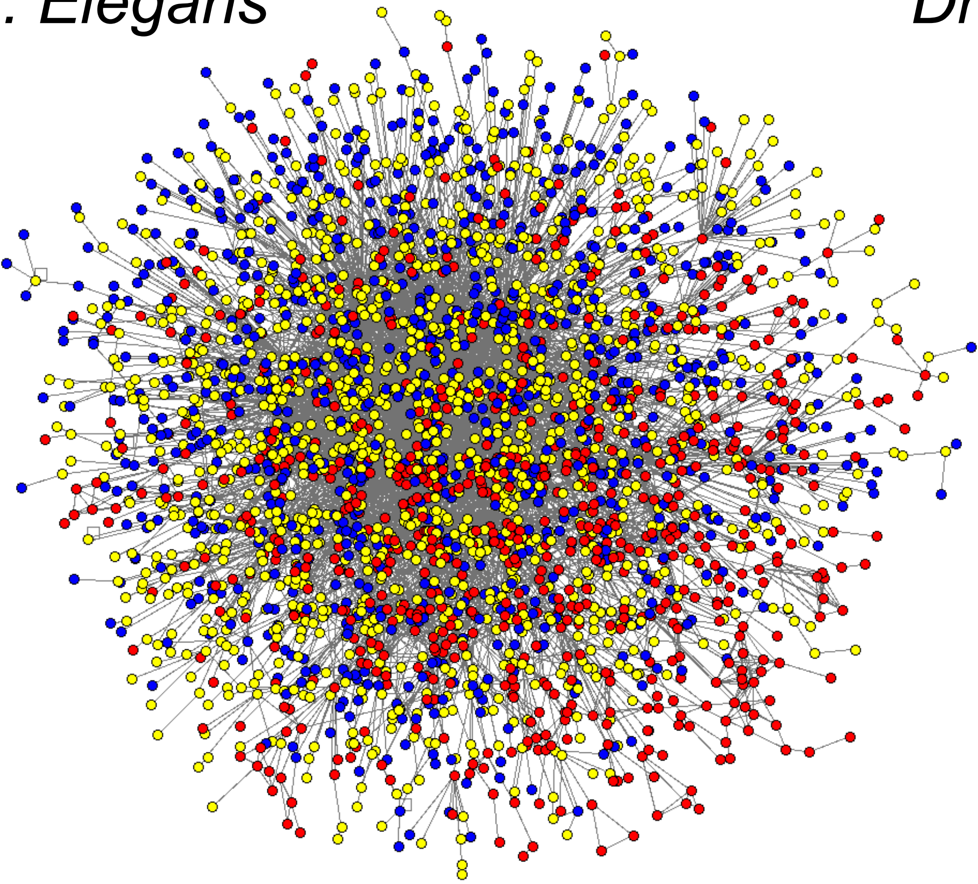


Nature Biotechnology 18, 1257 - 1261 (2000) doi:10.1038/82360

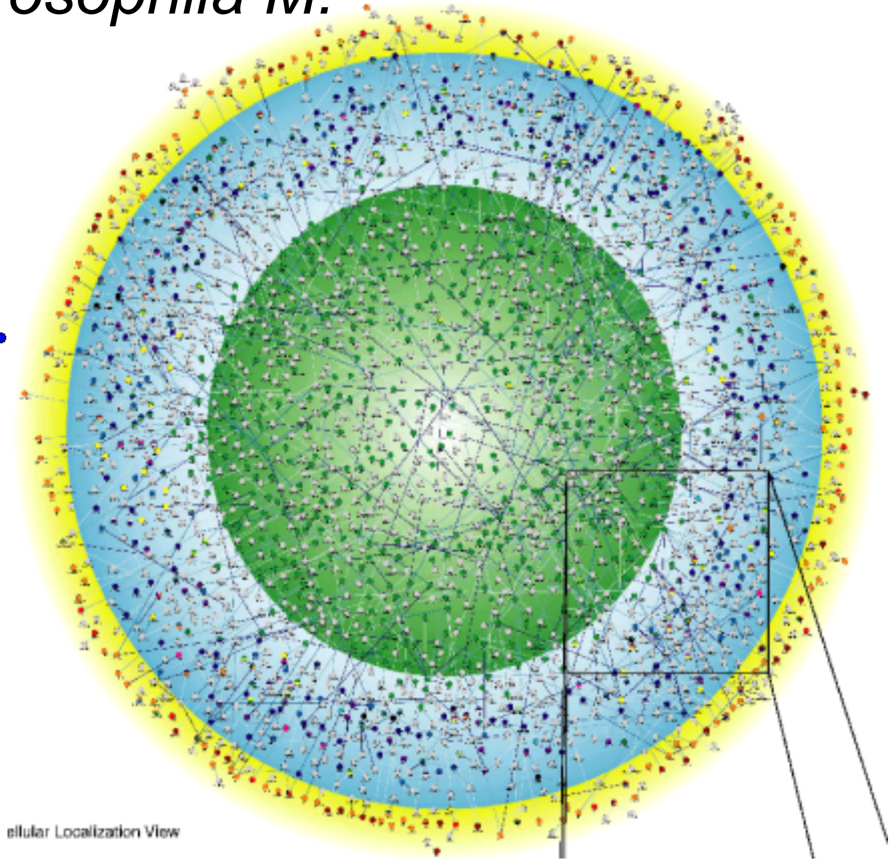
A network of protein−protein interactions in yeast

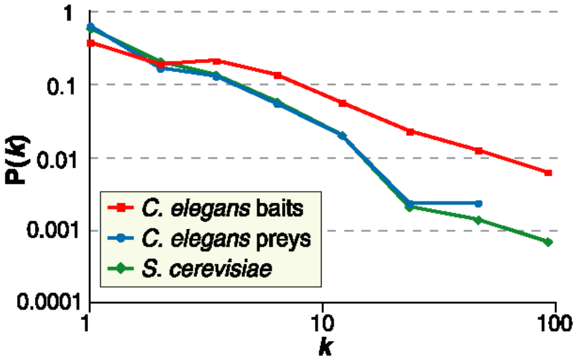
Benno Schwikowski, Peter Uetz & Stanley Fields

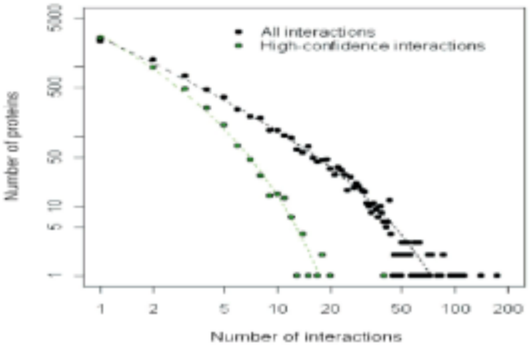
C. Elegans



Drosophila M.

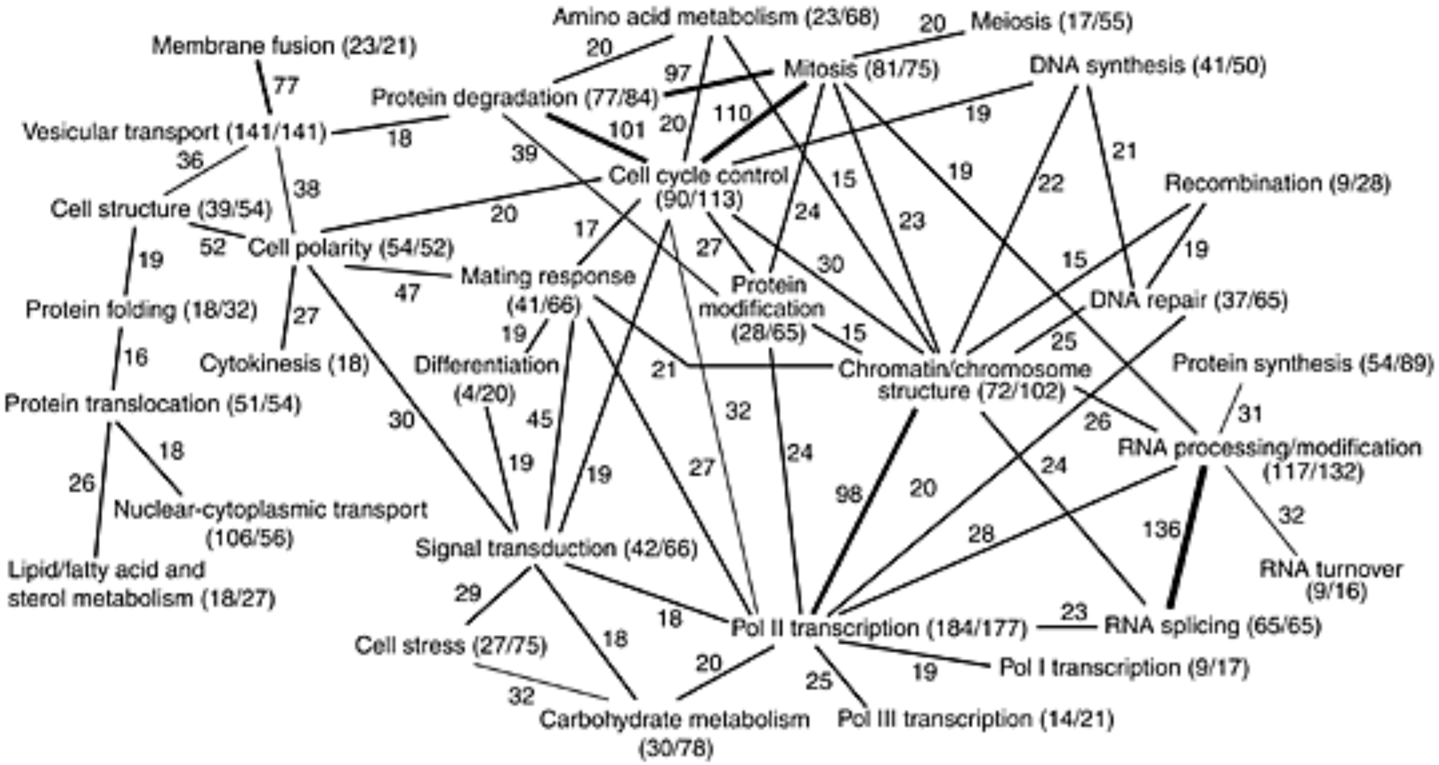






Li *et al.* Science 2004

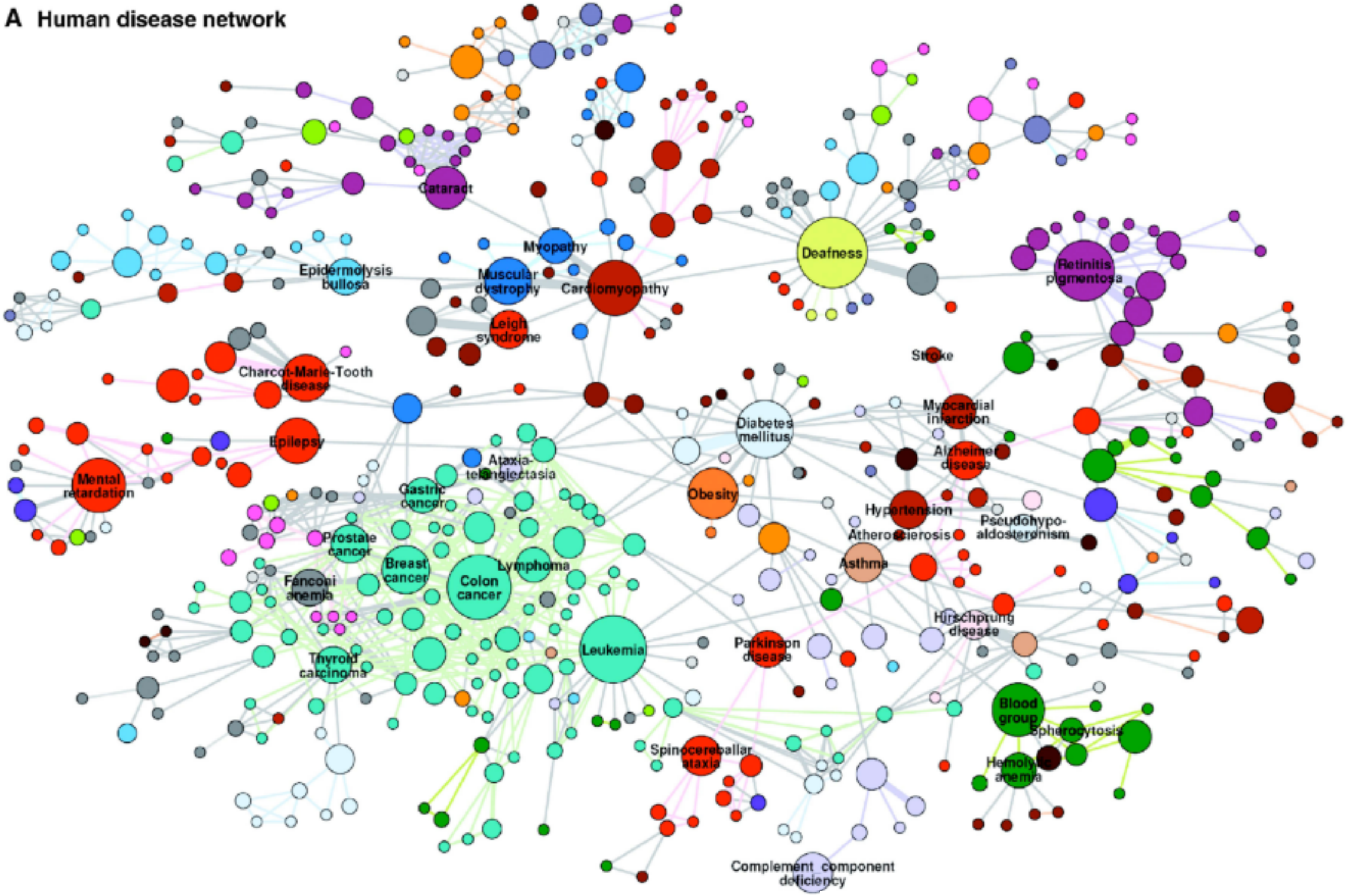
Giot *et al.* Science 2003



Nature Biotechnology 18, 1257 - 1261 (2000) doi:10.1038/82360

A network of protein−protein interactions in yeast

Benno Schwikowski, Peter Uetz & Stanley Fields



Molecular Systems Biology 3; Article number 124; doi:10.1038/msb4100163
Citation: Molecular Systems Biology 3:124

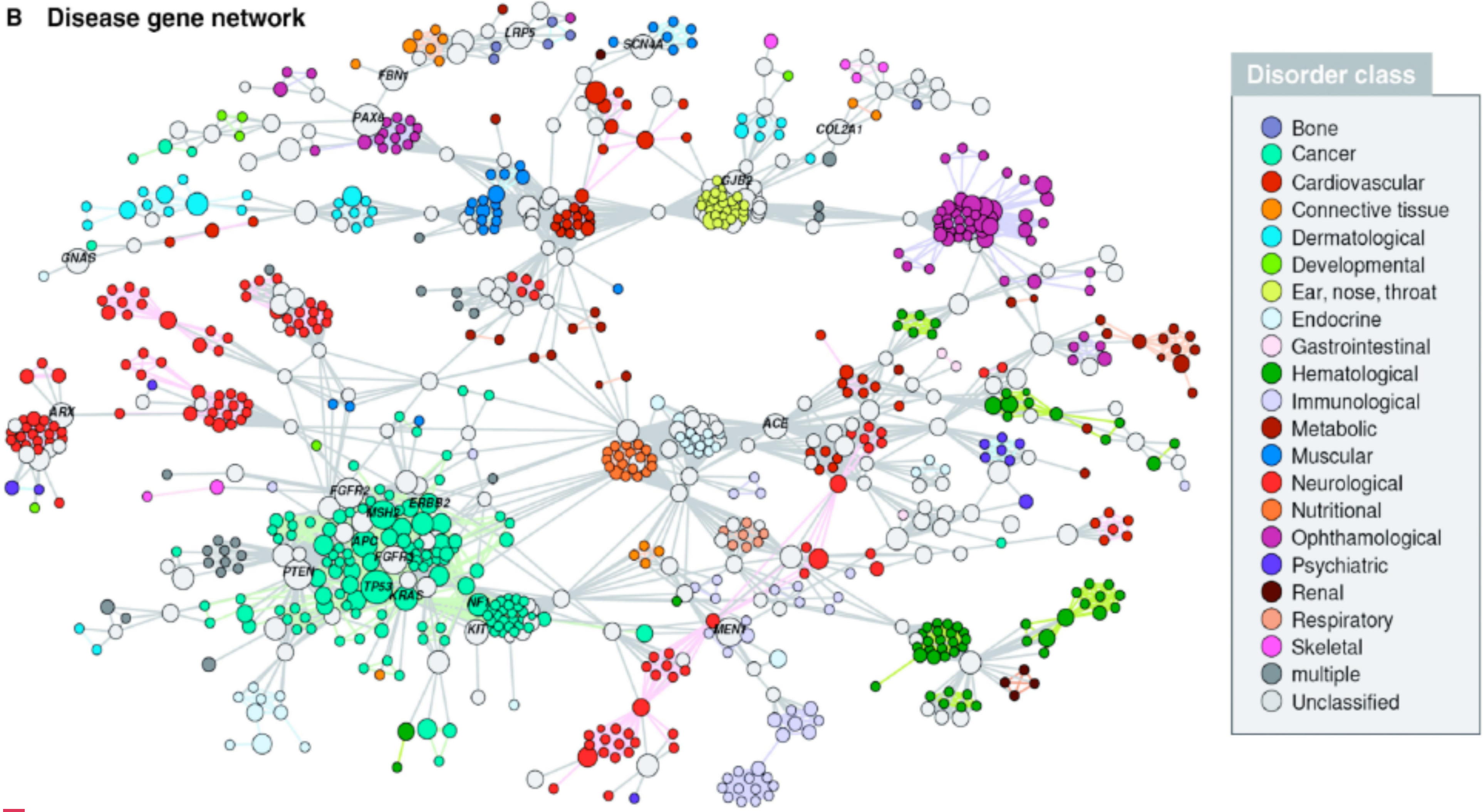
www.molecularsystemsbiology.com

PERSPECTIVE

Human disease classification in the postgenomic era: A complex systems approach to human pathobiology

Joseph Loscalzo[1,2,*], Isaac Kohane[2,3]
and Albert-Laszlo Barabasi[4]

Yet, this classic diagnostic strategy has widely recognized shortcomings that reflect both a lack of sensitivity in identifying preclinical disease and a lack of specificity in



Molecular Systems Biology 3; Article number 124; doi:10.1038/msb4100163
Citation: Molecular Systems Biology 3:124

www.molecularsystemsbiology.com

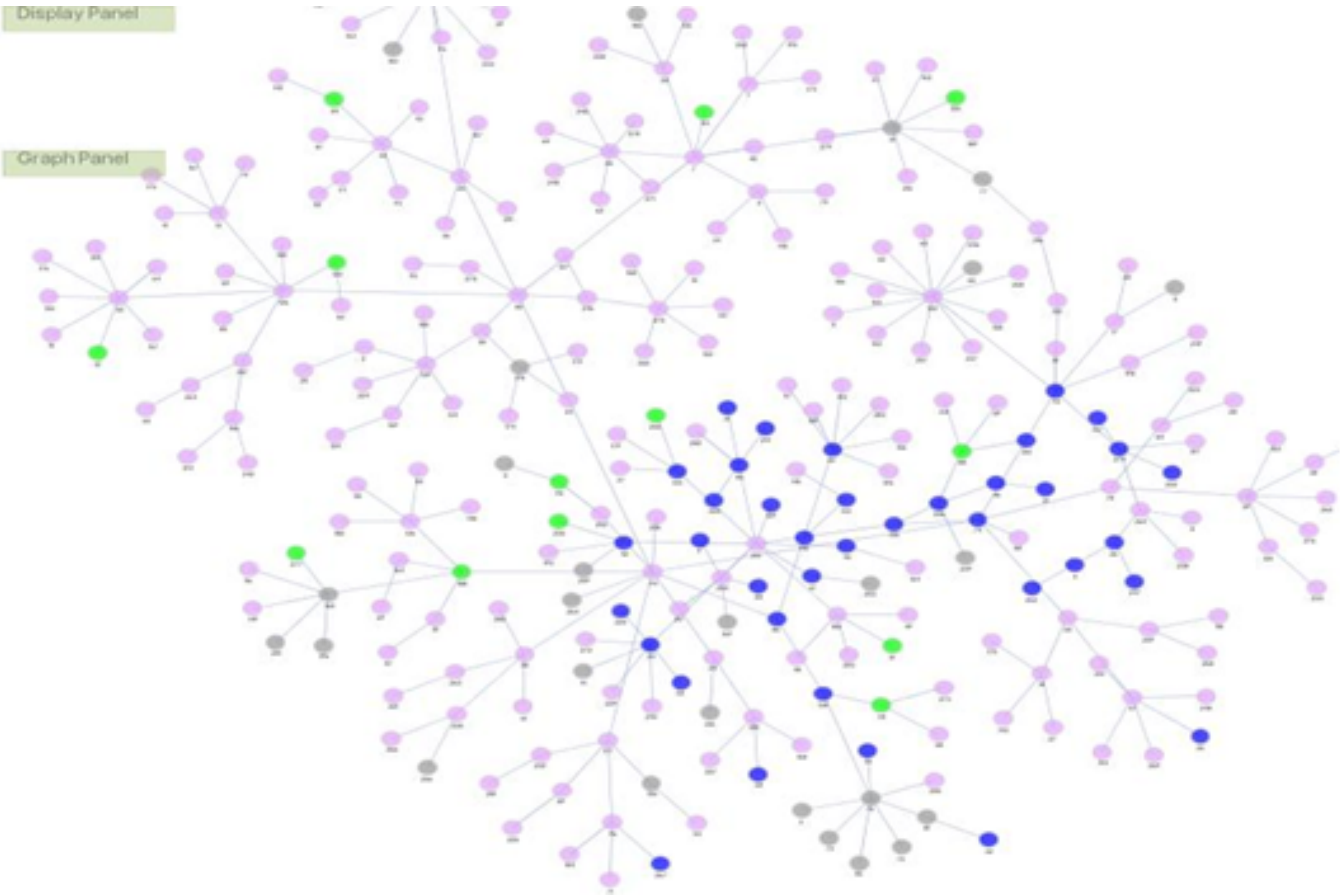
PERSPECTIVE

Human disease classification in the postgenomic era:
A complex systems approach to human pathobiology

Joseph Loscalzo[1,2,*], Isaac Kohane[2,3]
and Albert-Laszlo Barabasi[4]

Yet, this classic diagnostic strategy has widely recognized shortcomings that reflect both a lack of sensitivity in identifying preclinical disease and a lack of specificity in

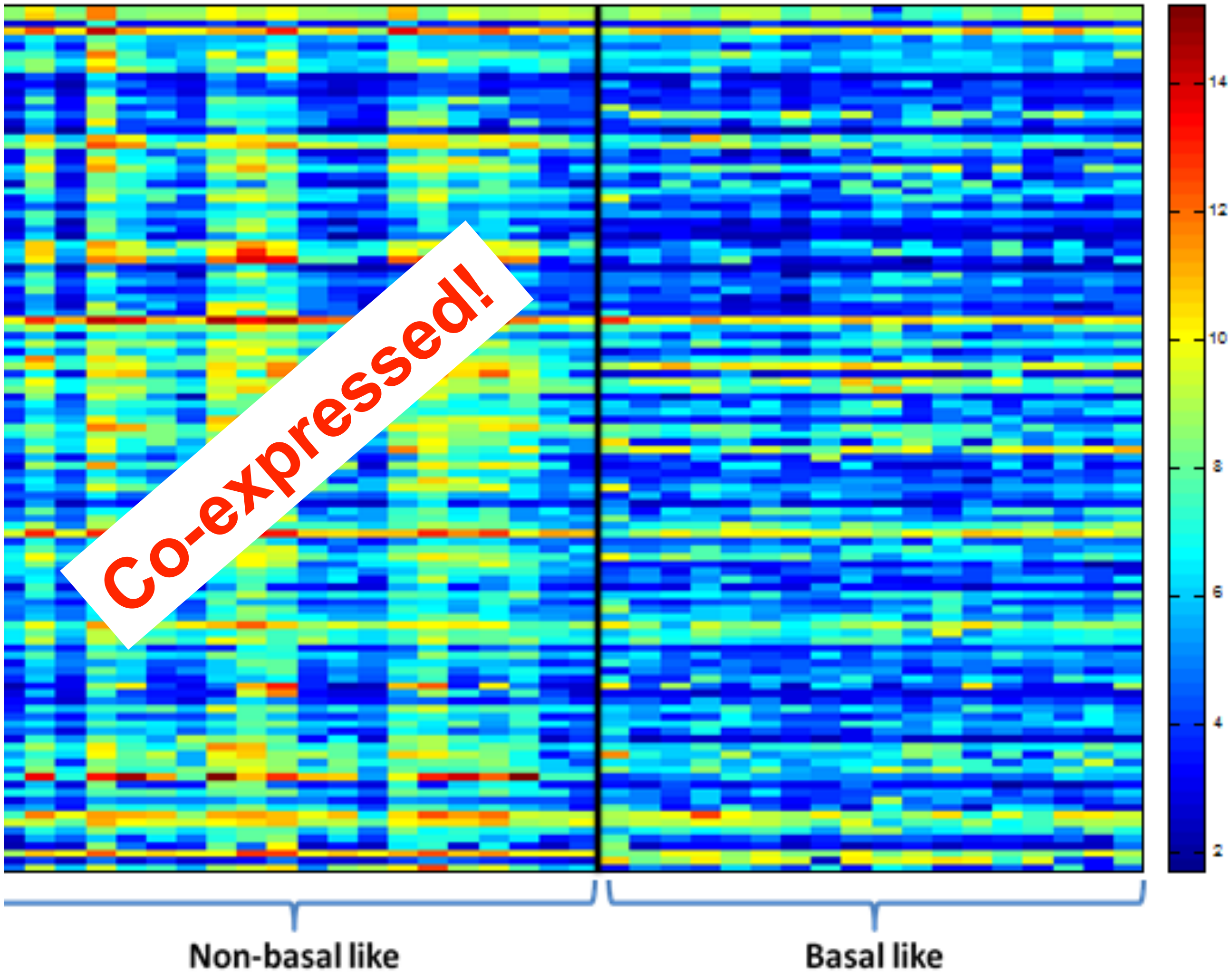
Article Network



Outline

- Introduction to biological networks
  - Different kind of networks
- How to build networks
  - Experimental methods (e.g., Y2H)
  - Computing methods
    - Example 1 – Gene co-expression
  - Informatics methods
    - Example 2 – STRING
- Databases and resources
- Analyzing and mining the networks

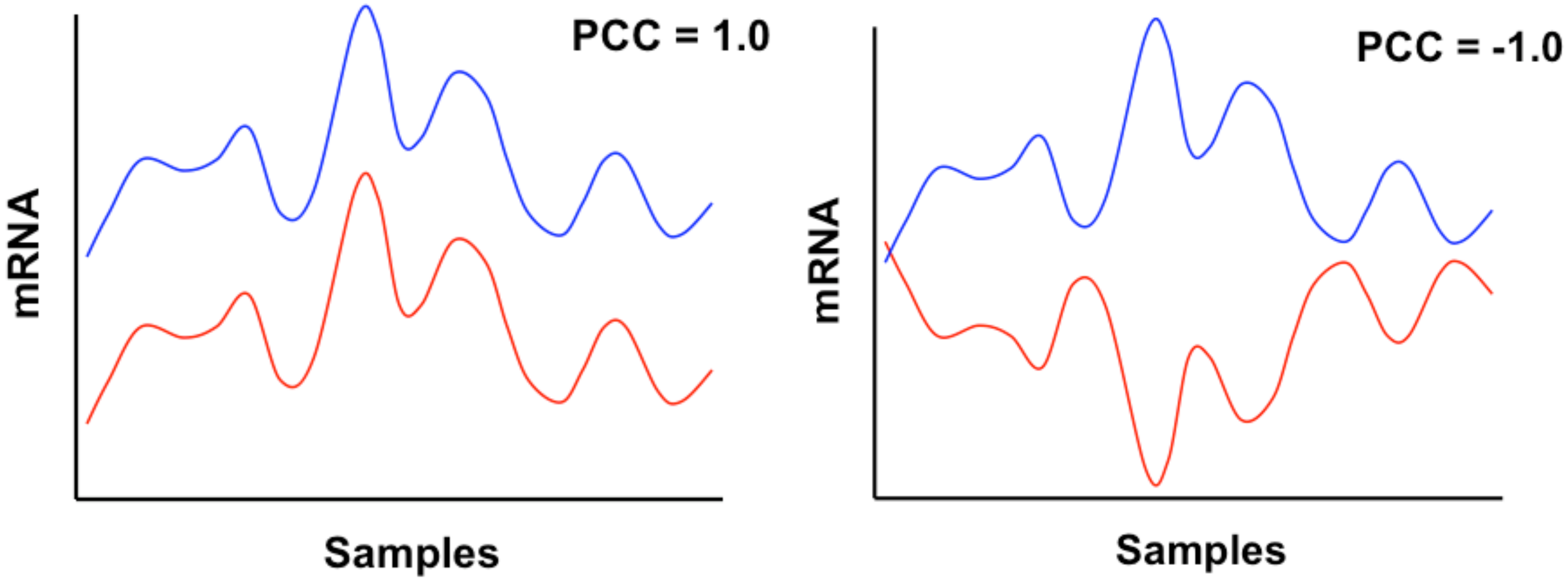
Gene Co-Expression



Co-expressed!

Kalluru et al, ICIBM, 2012.

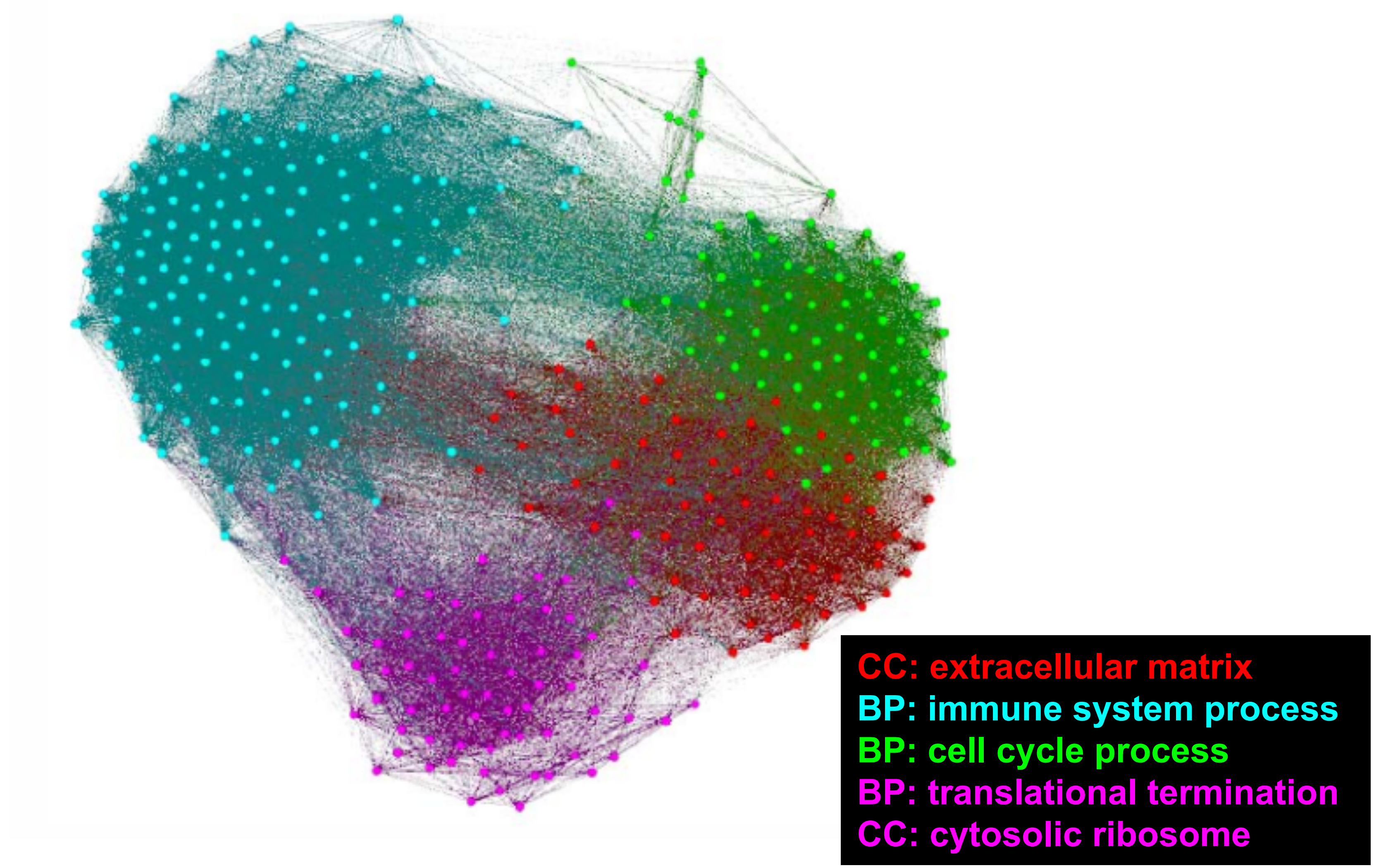
Gene Co-Expression



Q: What could be the mechanisms and implications for co-expression?

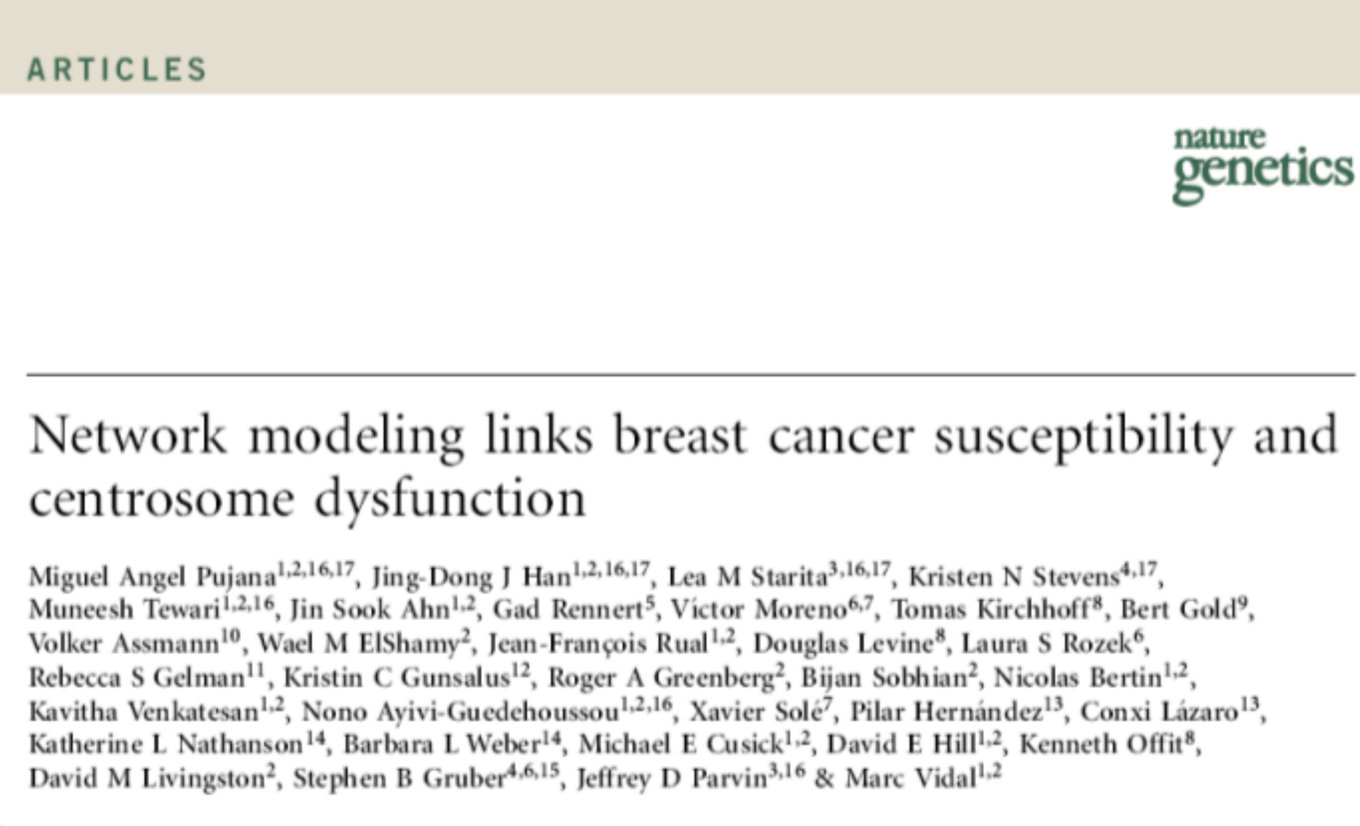
Wexner Medical Center

Gene Co-Expression Network



Y. Xiang, D. Fuhry, K. Kaya, R. Jin, U. Catalyurek, K. Huang *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2012, Volume 1, Number 3, Pages 103-116.

Gene Co-Expression Network to Infer Disease Gene

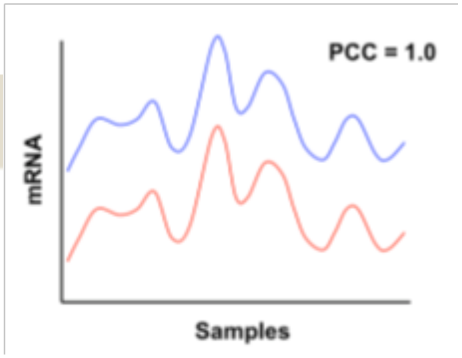


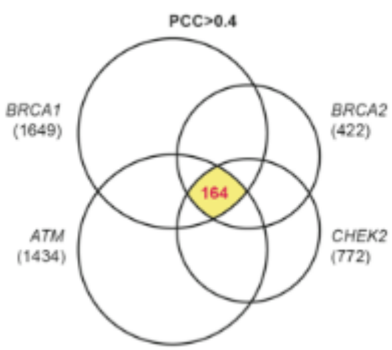
ARTICLES

nature genetics

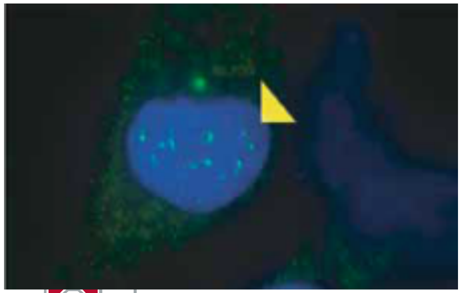
Network modeling links breast cancer susceptibility and centrosome dysfunction

Miguel Angel Pujana[1,2,16,17], Jing-Dong J Han[1,2,16,17], Lea M Starita[3,16,17], Kristen N Stevens[4,17], Muneesh Tewari[1,2,16], Jin Sook Ahn[1,2], Gad Rennert[5], Víctor Moreno[6,7], Tomas Kirchhoff[8], Bert Gold[9], Volker Assmann[10], Wael M ElShamy[2], Jean-François Rual[1,2], Douglas Levine[8], Laura S Rozek[6], Rebecca S Gelman[11], Kristin C Gunsalus[12], Roger A Greenberg[2], Bijan Sobhian[2], Nicolas Bertin[1,2], Kavitha Venkatesan[1,2], Nono Ayivi-Guedehoussou[1,2,16], Xavier Solé[7], Pilar Hernández[13], Conxi Lázaro[13], Katherine L Nathanson[14], Barbara L Weber[14], Michael E Cusick[1,2], David E Hill[1,2], Kenneth Offit[8], David M Livingston[2], Stephen B Gruber[4,6,15], Jeffrey D Parvin[3,16] & Marc Vidal[1,2]

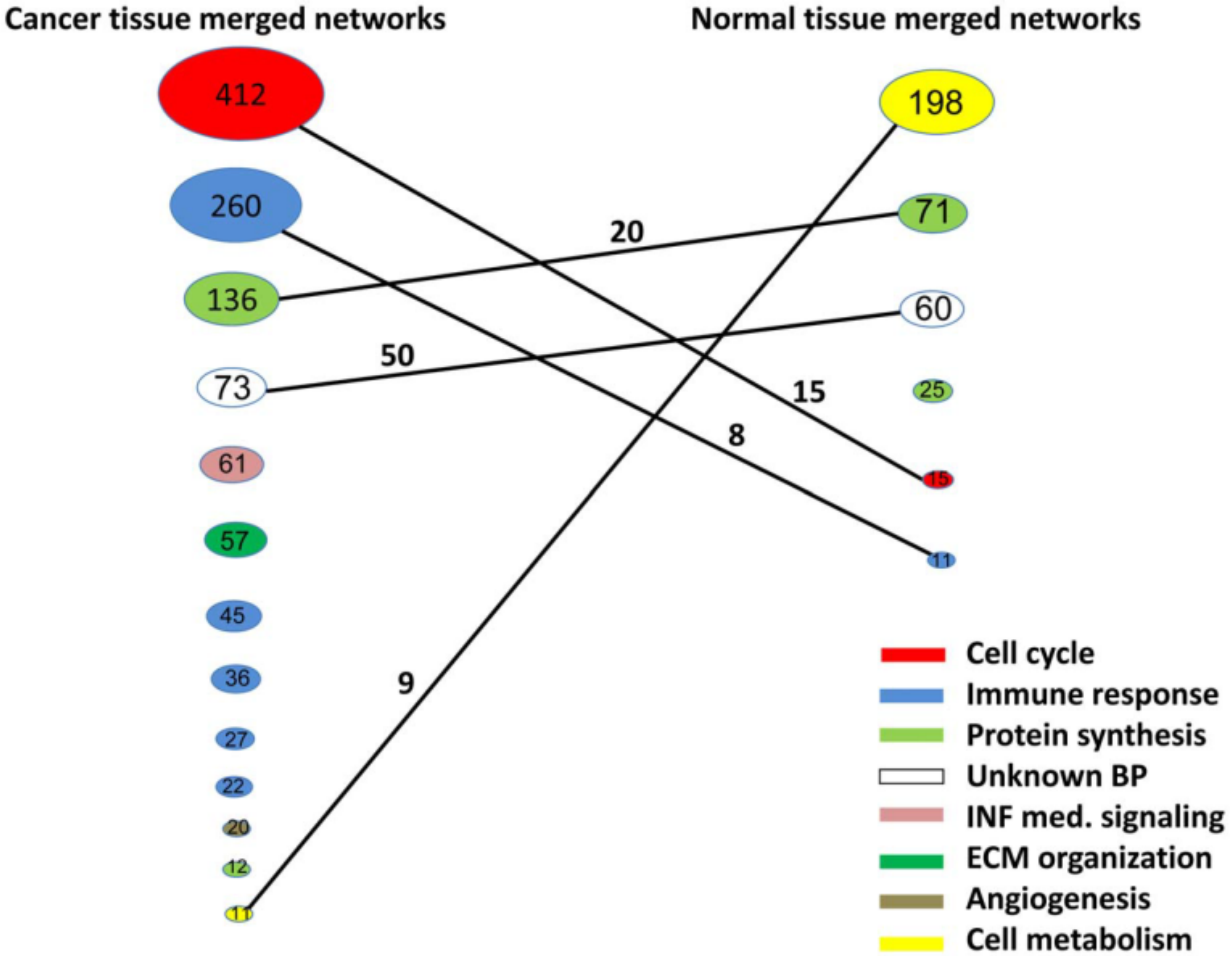




HMMR siRNA

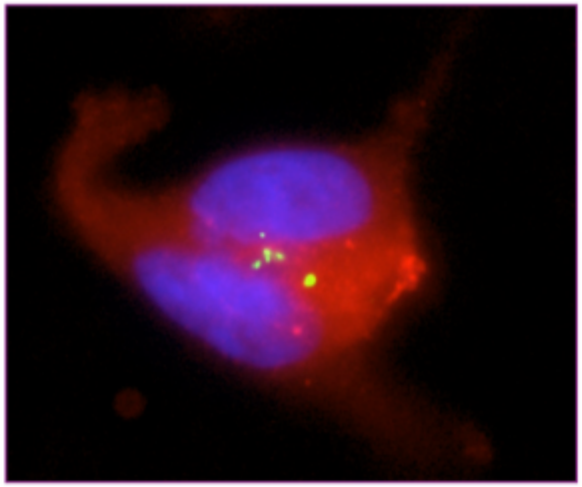


Compare Frequent Co-expression Networks

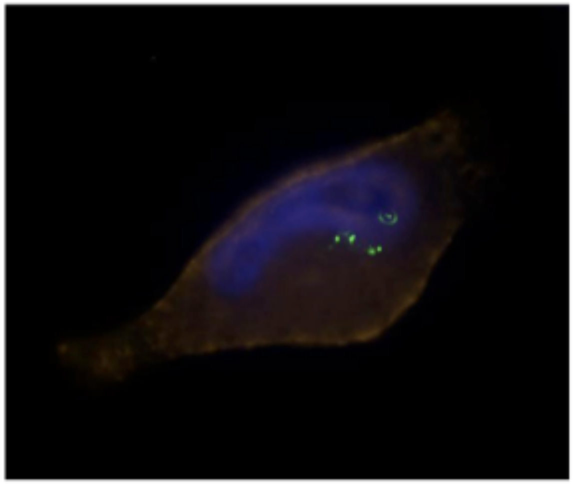


Zhang J, Lu K, Xiang Y, Islam M, et al. (2012) PLoS Comput Biol 8(8): e1002656.

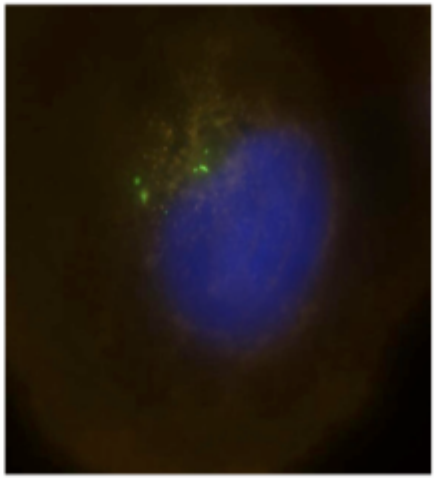
Predict New Gene Functions



siRNA: KIAA0101



siRNA: KIF23

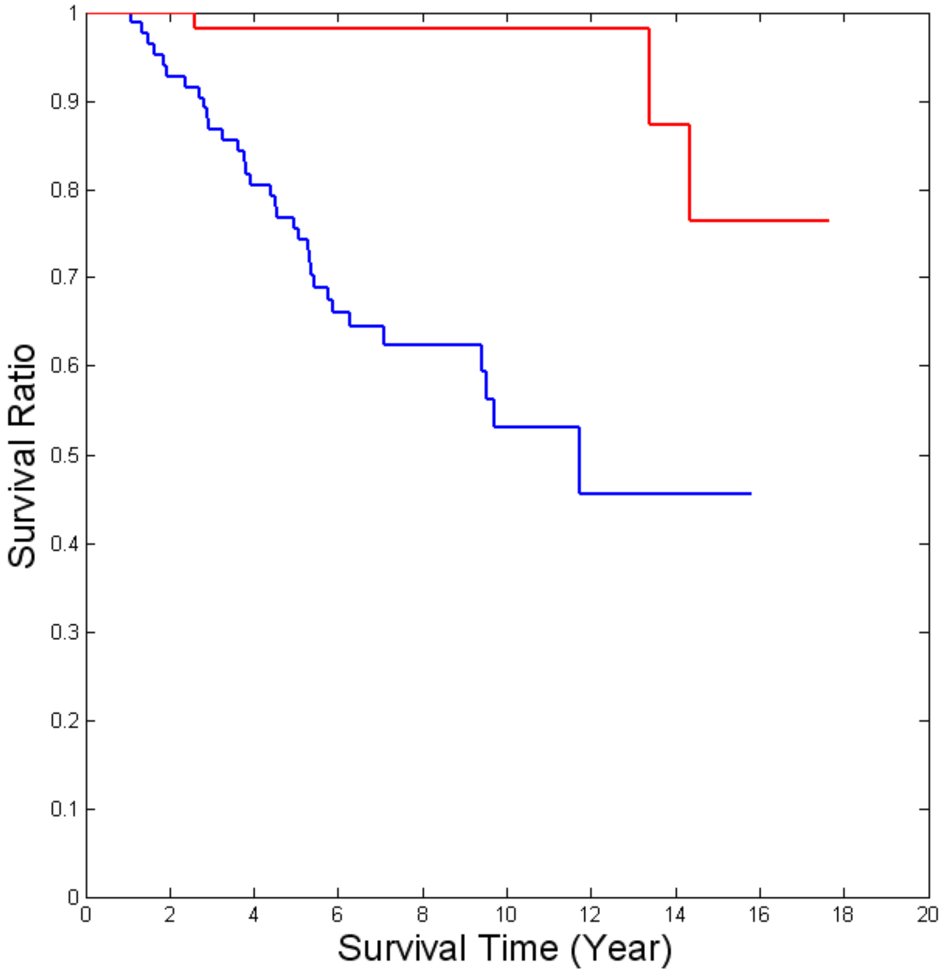


siRNA: KIF14

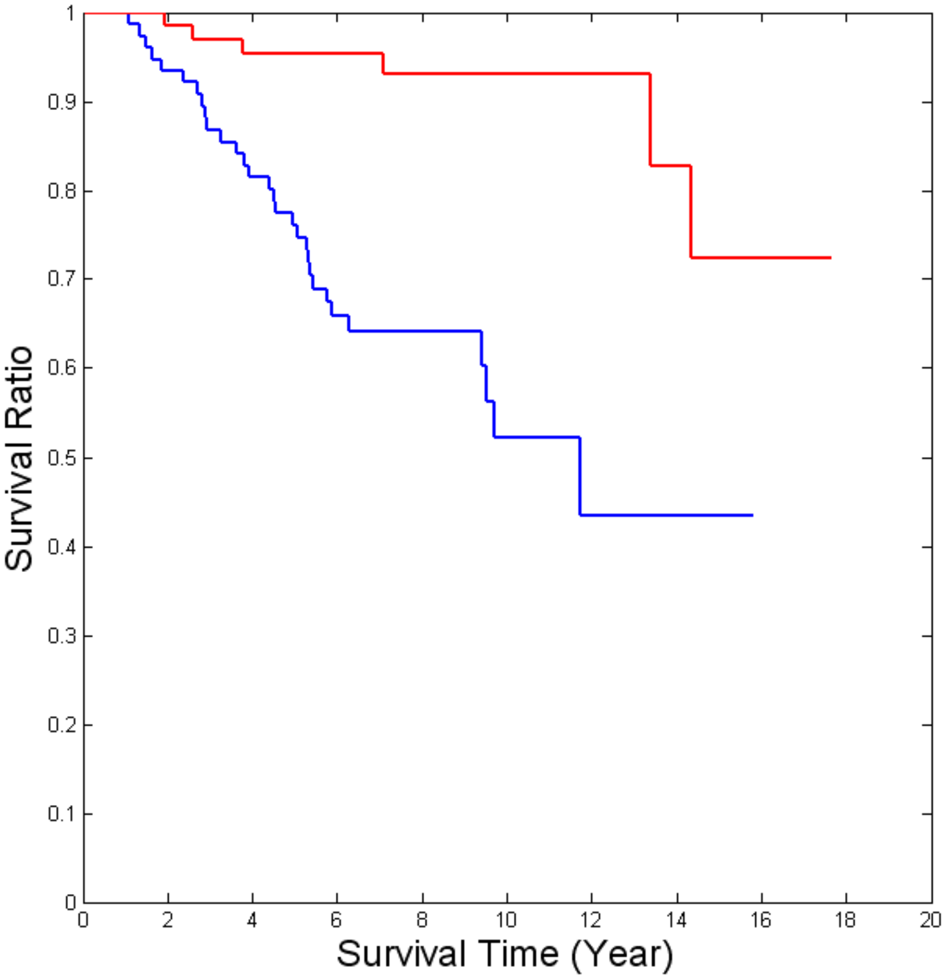
Predict Survival Outcome

vant Veer 70 Genes --Lymph Node Positive Samples p=0.00026302 N_1=55 N_2=89

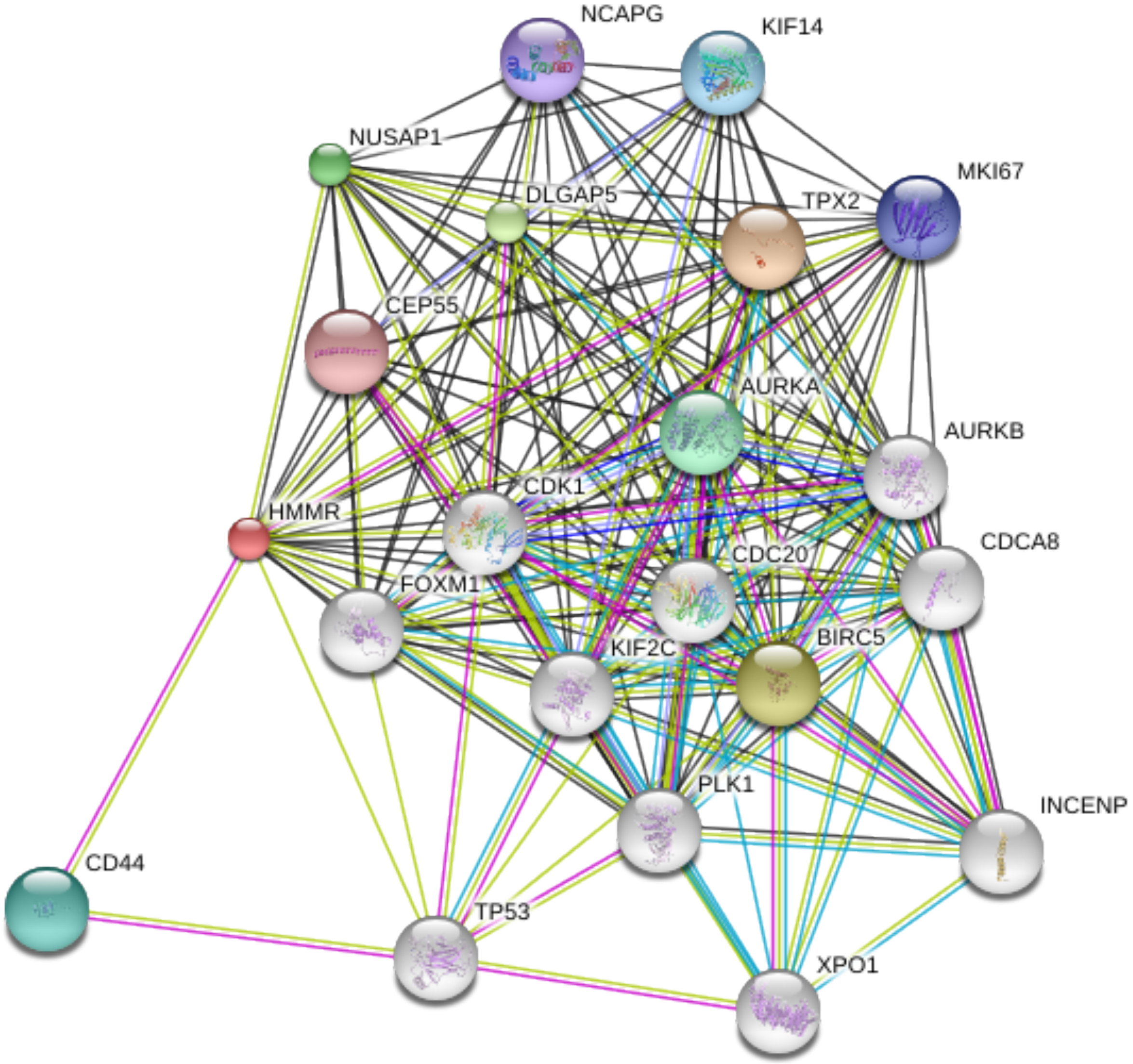
List2--Lymph Node Positive Samples p=1.4822e-006 N_1=83 N_2=60



List3--Lymph Node Positive Samples p=1.4977e-005 N_1=76 N_2=67

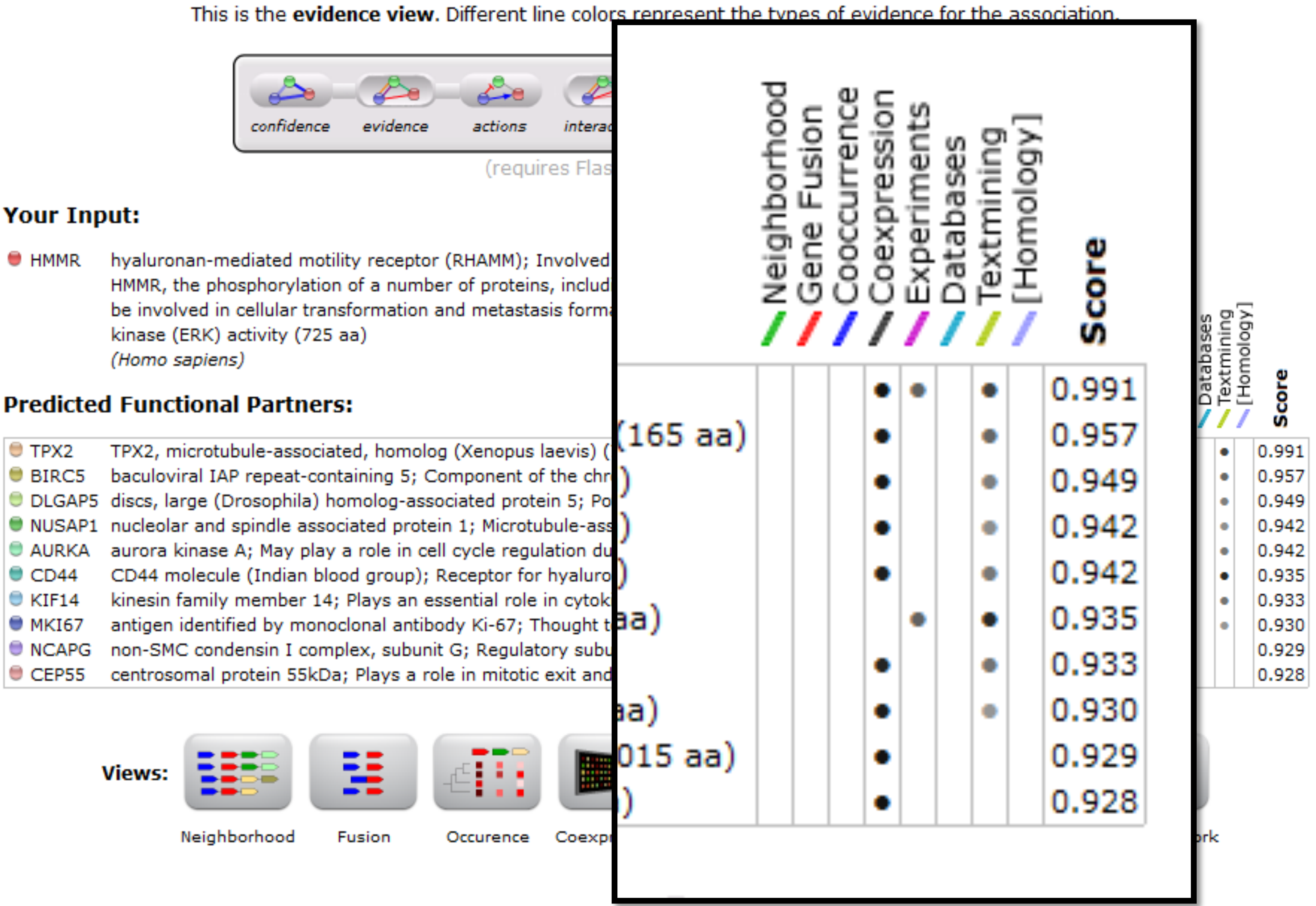


Protein-Protein Interaction Network



Wexner Medical Center

Protein-Protein Interaction Network - STRING



This is the **evidence view**. Different line colors represent the types of evidence for the association.

Your Input:

- HMMR hyaluronan-mediated motility receptor (RHAMM); Involved HMMR, the phosphorylation of a number of proteins, includ be involved in cellular transformation and metastasis forma kinase (ERK) activity (725 aa) *(Homo sapiens)*

Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	**Score**
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis) (				●	●		●		0.991
BIRC5	baculoviral IAP repeat-containing 5; Component of the chr				●			●		0.957
DLGAP5	discs, large (Drosophila) homolog-associated protein 5; Po				●			●		0.949
NUSAP1	nucleolar and spindle associated protein 1; Microtubule-ass				●			●		0.942
AURKA	aurora kinase A; May play a role in cell cycle regulation du				●			●		0.942
CD44	CD44 molecule (Indian blood group); Receptor for hyaluro					●		●		0.935
KIF14	kinesin family member 14; Plays an essential role in cytok				●			●		0.933
MKI67	antigen identified by monoclonal antibody Ki-67; Thought t				●			●		0.930
NCAPG	non-SMC condensin I complex, subunit G; Regulatory subu				●					0.929
CEP55	centrosomal protein 55kDa; Plays a role in mitotic exit and				●					0.928

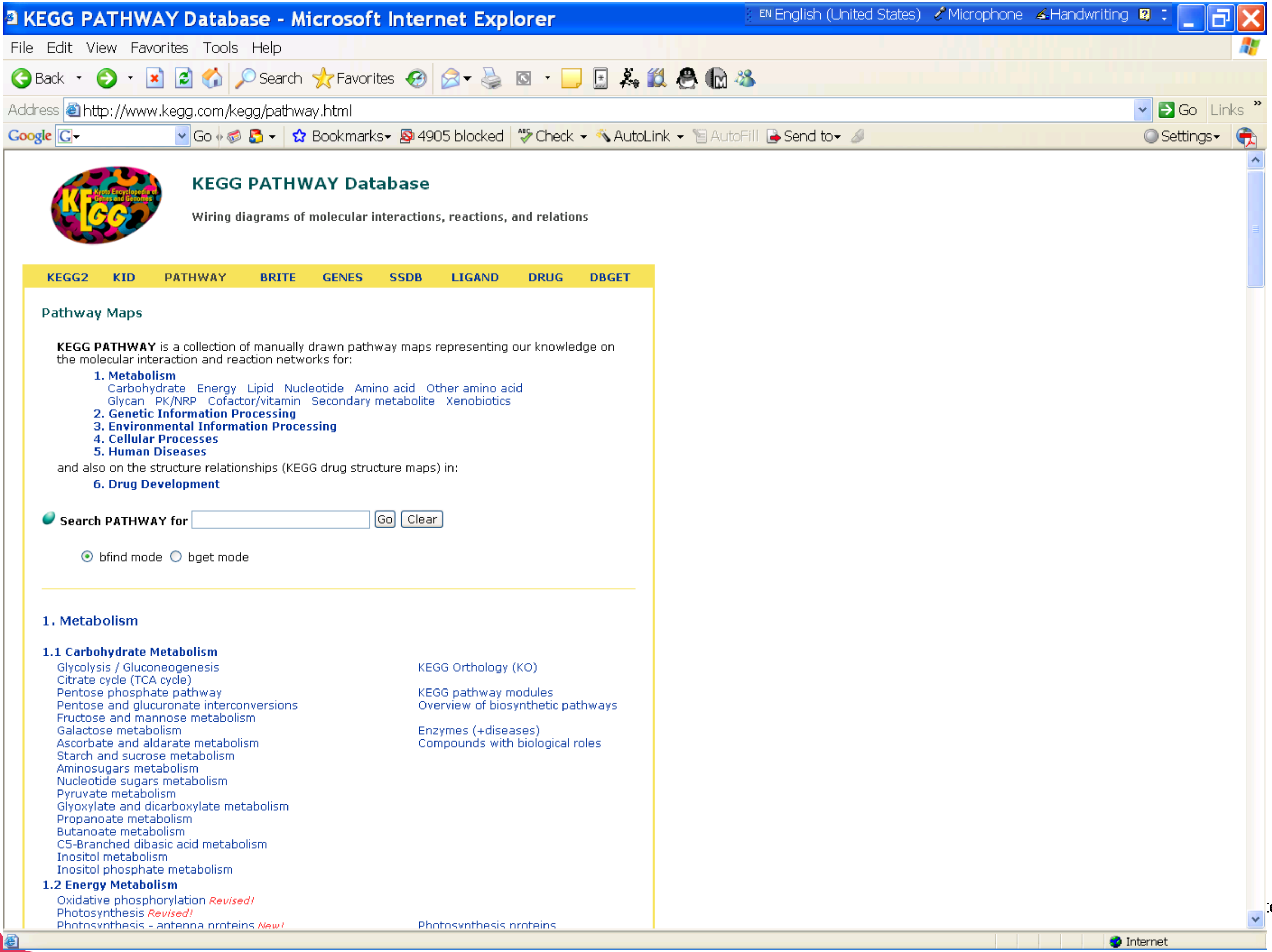
Views:

Databases and Resources

- KEGG, BioCarta
- MIMI – integrated many different PPI and molecular interaction databases
- STRING – multiple evidence based integration
- Ingenuity Pathway Analysis (commercial)
- MetaCore (commercial)
- Many others – http://www.pathguide.org

Outline

- Introduction to biological networks
  - Different kind of networks
- How to build networks
  - Experimental methods
  - Computing methods
    - Example 1 – Gene co-expression
  - Informatics methods
    - Example 2 – STRING
- **Databases and resources**
- Analyzing and mining the networks



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 KID PATHWAY BRITE GENES SSDB LIGAND DRUG DBGET

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

1. **Metabolism**
   Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid
   Glycan PK/NRP Cofactor/vitamin Secondary metabolite Xenobiotics
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

6. **Drug Development**

Search PATHWAY for [ ] Go Clear

● bfind mode ○ bget mode

---

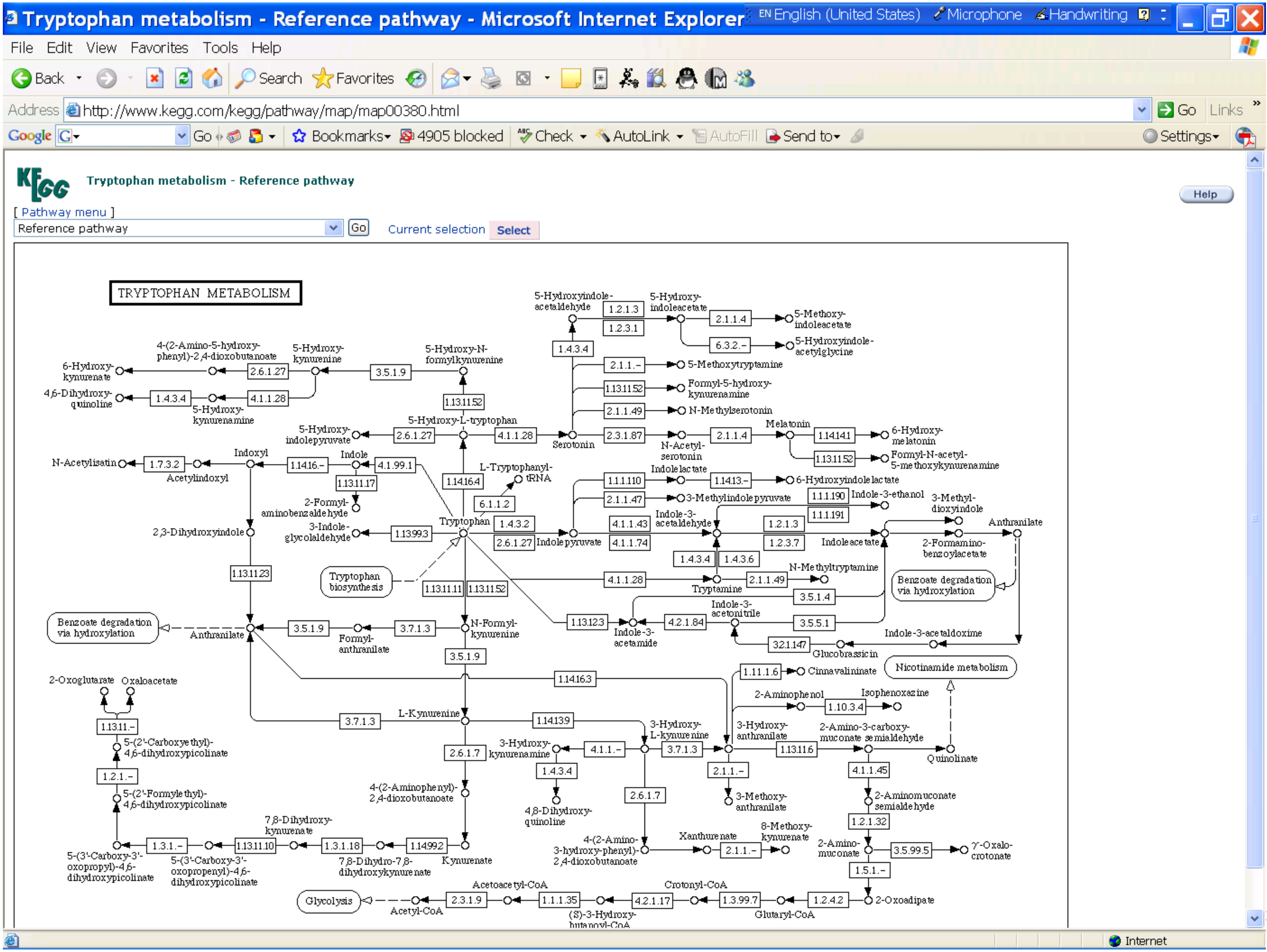
1. Metabolism

1.1 Carbohydrate Metabolism

Glycolysis / Gluconeogenesis	KEGG Orthology (KO)
Citrate cycle (TCA cycle)	
Pentose phosphate pathway	KEGG pathway modules
Pentose and glucuronate interconversions	Overview of biosynthetic pathways
Fructose and mannose metabolism	
Galactose metabolism	Enzymes (+diseases)
Ascorbate and aldarate metabolism	Compounds with biological roles
Starch and sucrose metabolism	
Aminosugars metabolism	
Nucleotide sugars metabolism	
Pyruvate metabolism	
Glyoxylate and dicarboxylate metabolism	
Propanoate metabolism	
Butanoate metabolism	
C5-Branched dibasic acid metabolism	
Inositol metabolism	
Inositol phosphate metabolism	

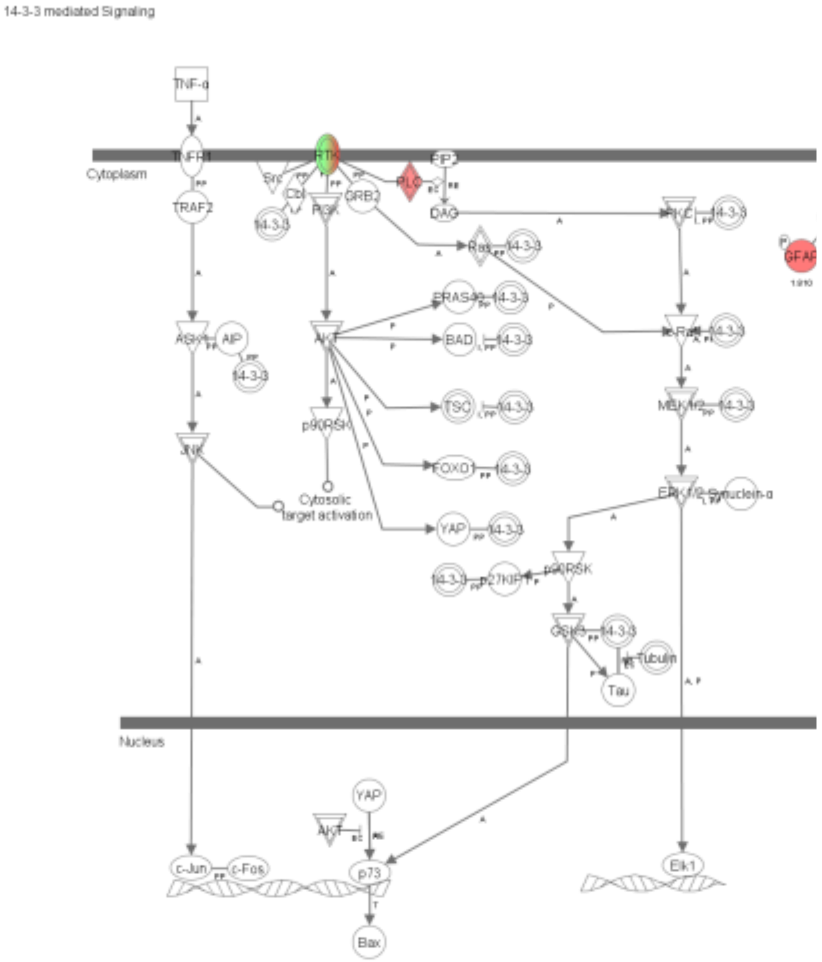
1.2 Energy Metabolism

Oxidative phosphorylation *Revised!*	
Photosynthesis *Revised!*	
Photosynthesis - antenna proteins *New!*	Photosynthesis proteins


Tryptophan metabolism - Reference pathway - Microsoft Internet Explorer
EN English (United States)
Microphone
Handwriting
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Address
http://www.kegg.com/kegg/pathway/map/map00380.html
Go
Links
Google
Go
Bookmarks
4905 blocked
Check
AutoLink
AutoFill
Send to
Settings
KEGG
Tryptophan metabolism - Reference pathway
Help
[Pathway menu]
Reference pathway
Go
Current selection
Select
TRYPTOPHAN METABOLISM
5-Hydroxyindole-acetaldehyde
1.2.1.3
1.2.3.1
5-Hydroxy-indoleacetate
2.1.1.4
5-Methoxy-indoleacetate
6.3.2.-
5-Hydroxyindole-acetylglycine
1.4.3.4
2.1.1.-
5-Methoxytryptamine
1.13.11.52
Formyl-5-hydroxy-kynurenamine
2.1.1.49
N-Methylserotonin
2.3.1.87
2.1.1.4
Melatonin
1.14.14.1
6-Hydroxy-melatonin
1.13.11.52
Formyl-N-acetyl-5-methoxykynurenamine
Serotonin
N-Acetyl-serotonin
4-(2-Amino-5-hydroxy-phenyl)-2,4-dioxobutanoate
5-Hydroxy-kynurenine
5-Hydroxy-N-formylkynurenine
6-Hydroxy-kynurenate
2.6.1.27
3.5.1.9
4,6-Dihydroxy-quinoline
1.4.3.4
4.1.1.28
5-Hydroxy-kynurenamine
1.13.11.52
5-Hydroxy-L-tryptophan
5-Hydroxy-indolepyruvate
2.6.1.27
4.1.1.28
N-Acetylisatin
1.7.3.2
Indoxyl
Acetylindoxyl
Indole
1.14.16.-
4.1.99.1
1.13.11.17
1.14.16.4
L-Tryptophanyl-tRNA
6.1.1.2
Indolelactate
1.1.1.110
1.14.13.-
6-Hydroxyindolelactate
2.1.1.47
3-Methylindolepyruvate
1.1.1.190
Indole-3-ethanol
3-Methyl-dioxyindole
1.1.1.191
2-Formyl-aminobenzaldehyde
3-Indole-glycolaldehyde
1.13.99.3
2,3-Dihydroxyindole
Tryptophan
1.4.3.2
2.6.1.27
Indolepyruvate
4.1.1.43
4.1.1.74
Indole-3-acetaldehyde
1.2.1.3
1.2.3.7
Indoleacetate
Anthranilate
2-Formamino-benzoylacetate
1.4.3.4
1.4.3.6
N-Methyltryptamine
1.13.11.23
Tryptophan biosynthesis
1.13.11.11
1.13.11.52
4.1.1.28
Tryptamine
2.1.1.49
3.5.1.4
Benzoate degradation via hydroxylation
1.13.12.3
Indole-3-acetamide
4.2.1.84
Indole-3-acetonitrile
3.5.5.1
Benzoate degradation via hydroxylation
Anthranilate
3.5.1.9
Formyl-anthranilate
3.7.1.3
N-Formyl-kynurenine
3.5.1.9
3.2.1.147
Glucobrassicin
Indole-3-acetaldoxime
1.14.16.3
1.1.1.6
Cinnavalininate
Nicotinamide metabolism
2-Oxoglutarate
Oxaloacetate
1.13.11.-
5-(2'-Carboxyethyl)-4,6-dihydroxypicolinate
1.2.1.-
5-(2'-Formylethyl)-4,6-dihydroxypicolinate
3.7.1.3
L-Kynurenine
1.14.13.9
3-Hydroxy-L-kynurenine
2-Aminophenol
Isophenoxazine
1.10.3.4
3-Hydroxy-anthranilate
2-Amino-3-carboxy-muconate semialdehyde
2.6.1.7
3-Hydroxy-kynurenamine
4.1.1.-
3.7.1.3
1.13.11.6
Quinolinate
1.4.3.4
2.1.1.-
4.1.1.45
4,8-Dihydroxy-quinoline
2.6.1.7
3-Methoxy-anthranilate
2-Aminomuconate semialdehyde
4-(2-Aminophenyl)-2,4-dioxobutanoate
1.2.1.32
7,8-Dihydroxy-kynurenate
8-Methoxy-kynurenate
5-(3'-Carboxy-3'-oxopropyl)-4,6-dihydroxypicolinate
1.3.1.-
5-(3'-Carboxy-3'-oxopropenyl)-4,6-dihydroxypicolinate
1.13.11.10
1.3.1.18
7,8-Dihydro-7,8-dihydroxykynurenate
1.14.99.2
Kynurenate
4-(2-Amino-3-hydroxy-phenyl)-2,4-dioxobutanoate
Xanthurenate
2.1.1.-
2-Amino-muconate
3.5.99.5
γ-Oxalo-crotonate
1.5.1.-
Glycolysis
Acetyl-CoA
2.3.1.9
Acetoacetyl-CoA
1.1.1.35
(S)-3-Hydroxy-butanoyl-CoA
4.2.1.17
Crotonyl-CoA
1.3.99.7
Glutaryl-CoA
1.2.4.2
2-Oxoadipate
Internet

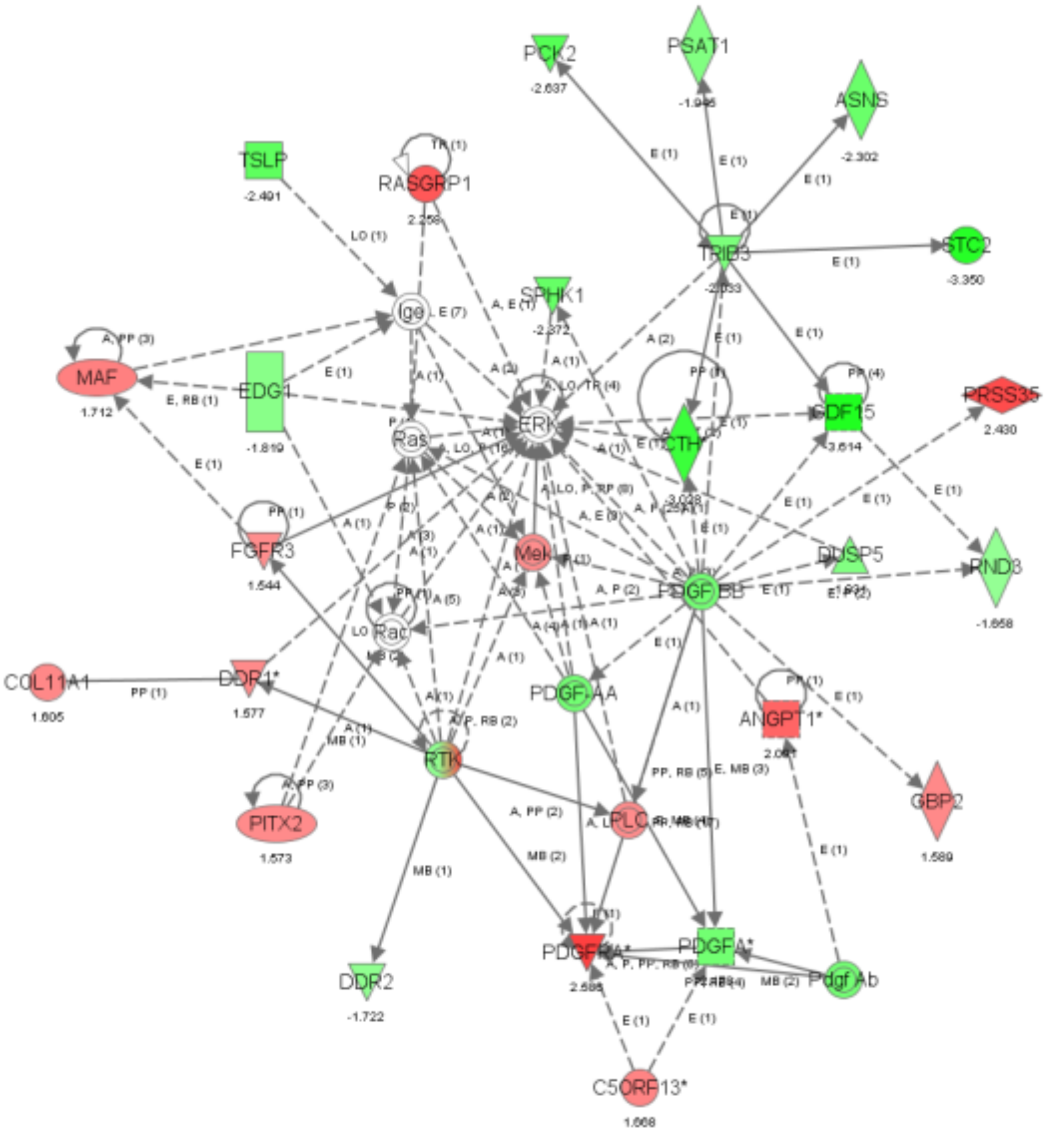
er

Ingenuity Pathway Analysis (IPA)



© 2000-2008 Ingenuity Systems, Inc. All rights reserved.

Network 1 : SeanL_RMAestimates_Results - 2008-07-14 08:24 AM : SeanL_RMAestimates_Results.xls



© 2000-2008 Ingenuity Systems, Inc. All rights reserved.

Pathway – What’s out there?

Complete Listing of All Pathguide Resources

Pathguide contains information about ~~222~~ 325 biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

Protein-Protein Interactions

3DID - 3D interacting domains

ABCdb - Archaea and Bacteria ABC transporter database

AfCS - Alliance for Cellular Signaling Molecule Pages Database

AllFuse - Functional Associations of Proteins in Complete Genomes

ASEdb - Alanine Scanning Energetics Database

Outline

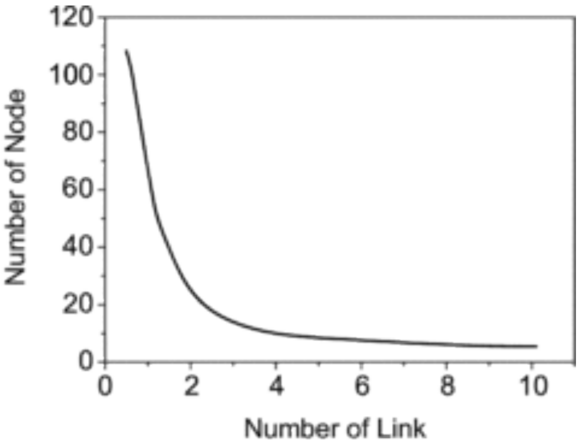
- Introduction to biological networks
  - Different kind of networks
- How to build networks
  - Experimental methods
  - Computing methods
    - Example 1 – Gene co-expression
  - Informatics methods
    - Example 2 – STRING
- Databases and resources
- **Analyzing and mining the networks**

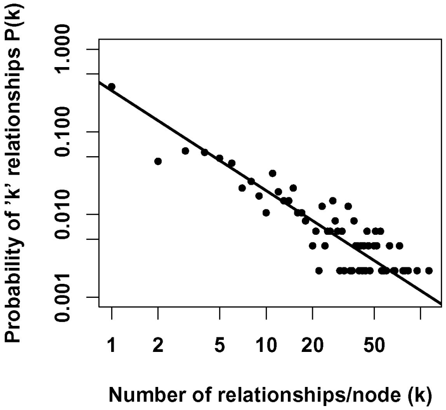
What can we learn from networks?

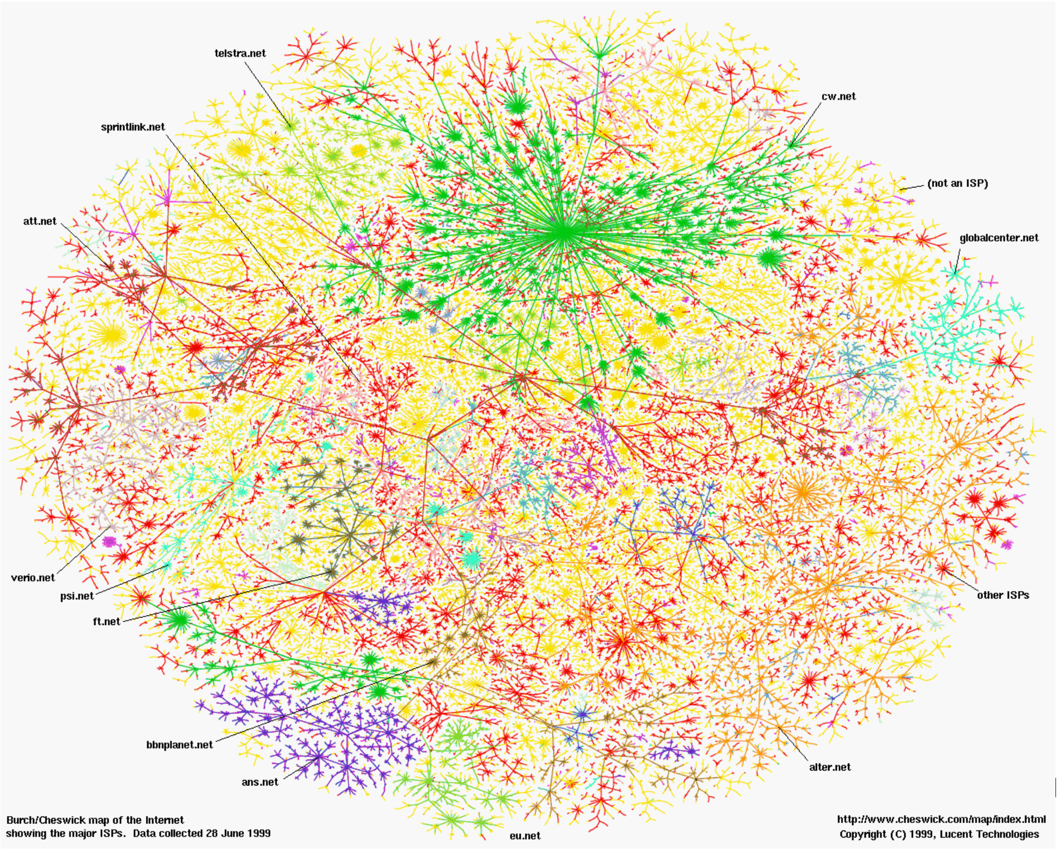
- Relationships
- Community information
- Logical or dynamic information
- (Sometimes) 1 + 1 > 2 in networks
- Need new theories from physics, mathematics, statistics, …

Power Law
Small World
Rich Get Richer (preferential attachment)
Self-similarity

HUBS!



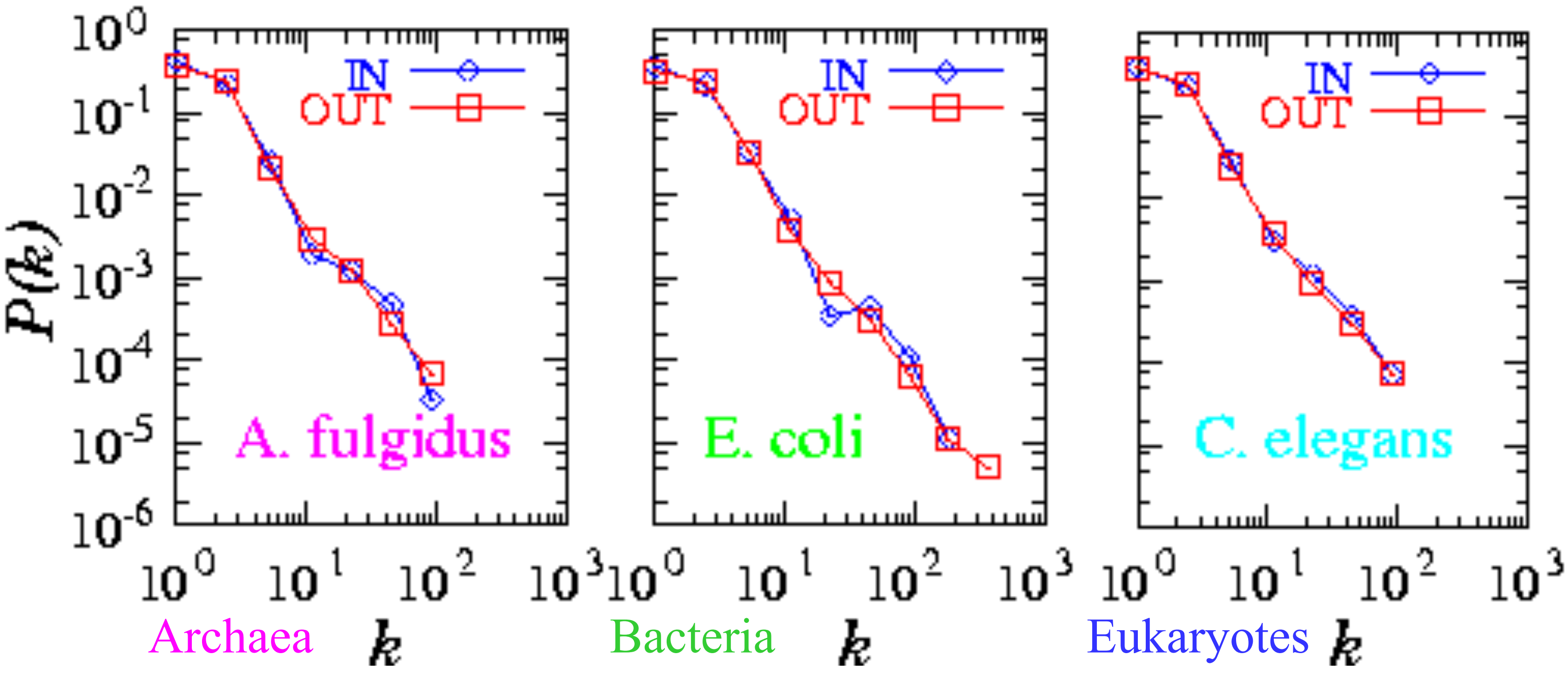




Burch/Cheswick map of the Internet
showing the major ISPs. Data collected 28 June 1999

http://www.cheswick.com/map/index.html
Copyright (C) 1999, Lucent Technologies

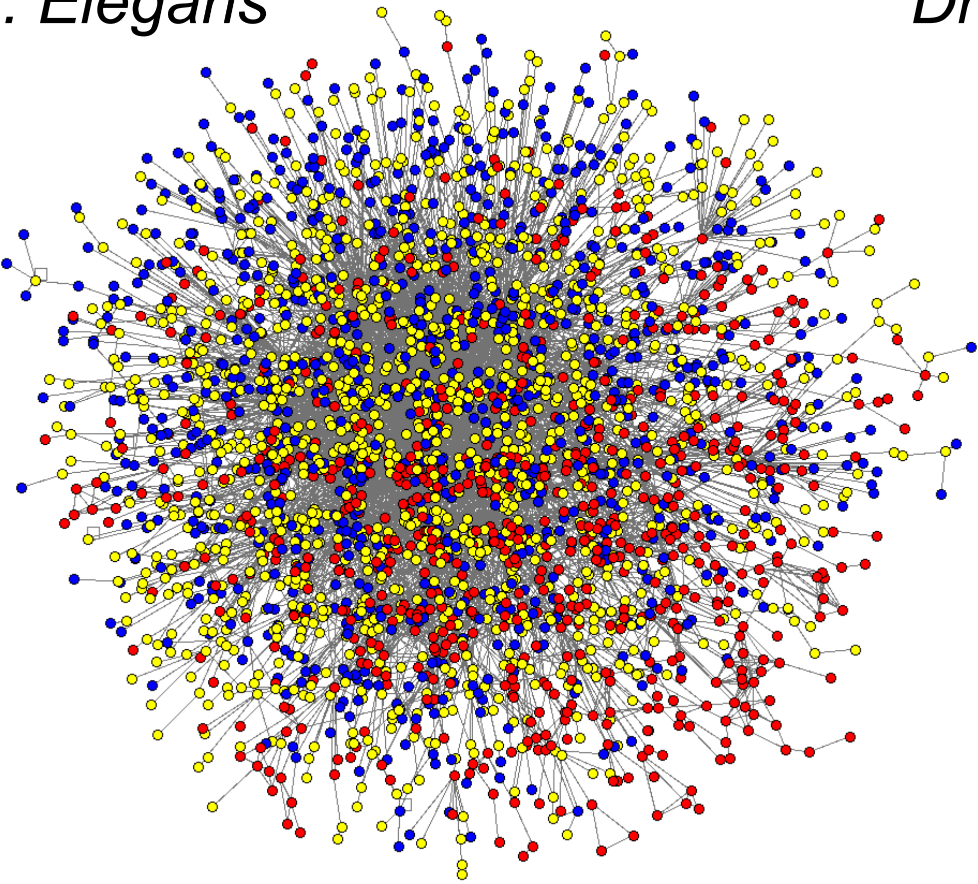
Metabolic network



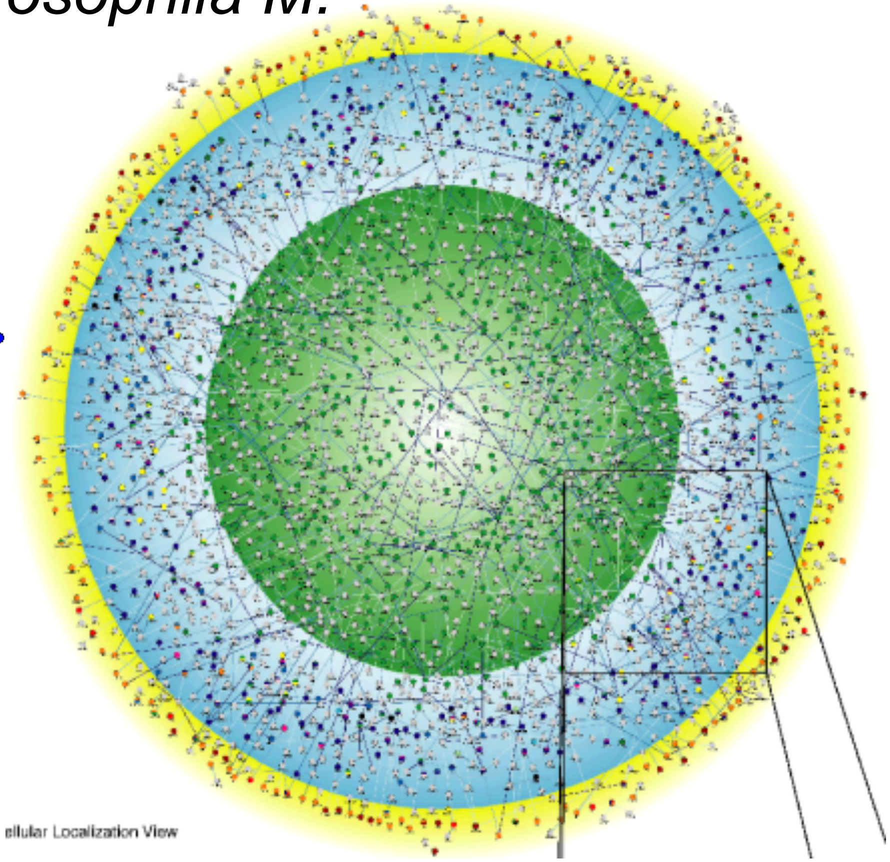
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, **407** 651 (2000)

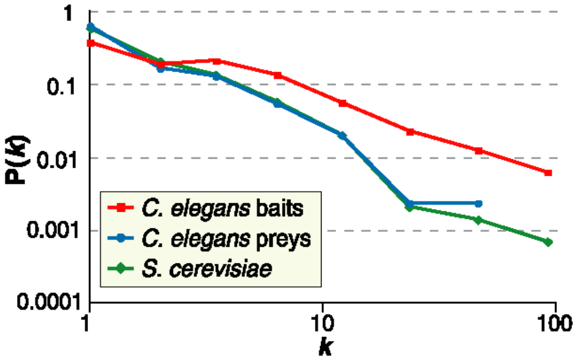
Organisms from all three domains of life are ***scale-free*** networks!

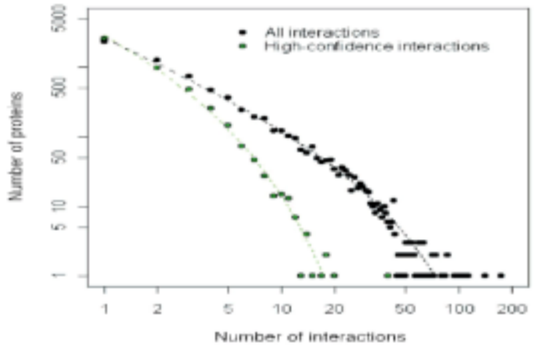
C. Elegans



Drosophila M.







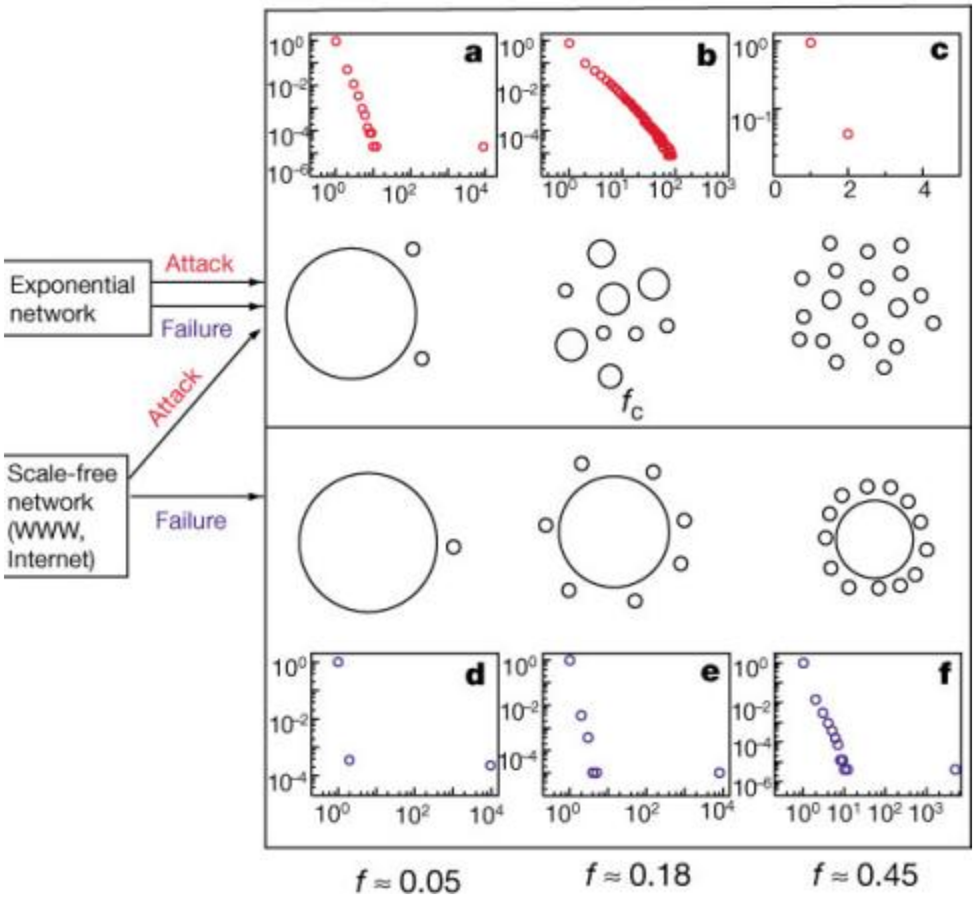
Li *et al.* Science 2004

Giot *et al.* Science 2003

Implications: Hubs and Robustness

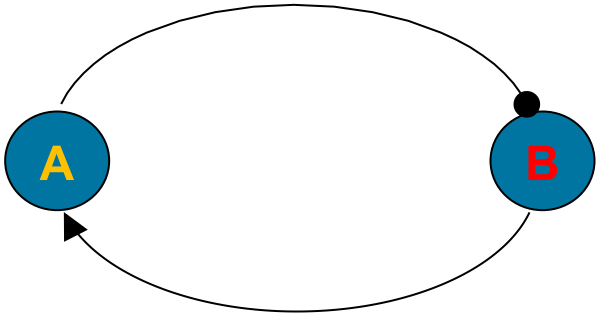
Complex systems maintain their basic functions even under errors and failures

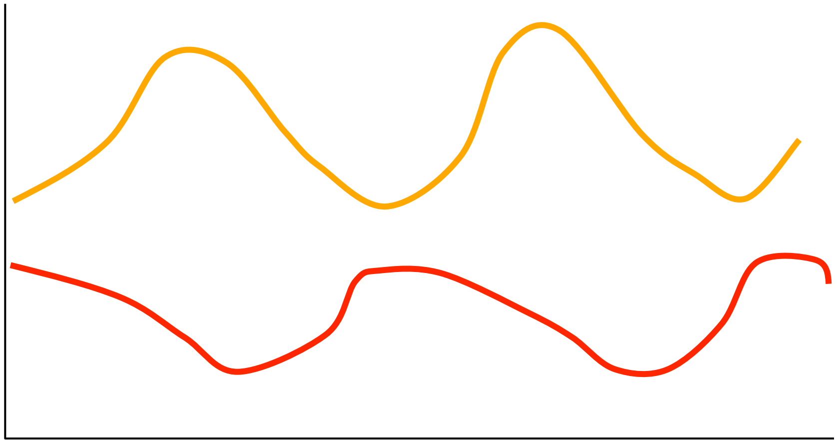
(cell → mutations; Internet → router breakdowns)

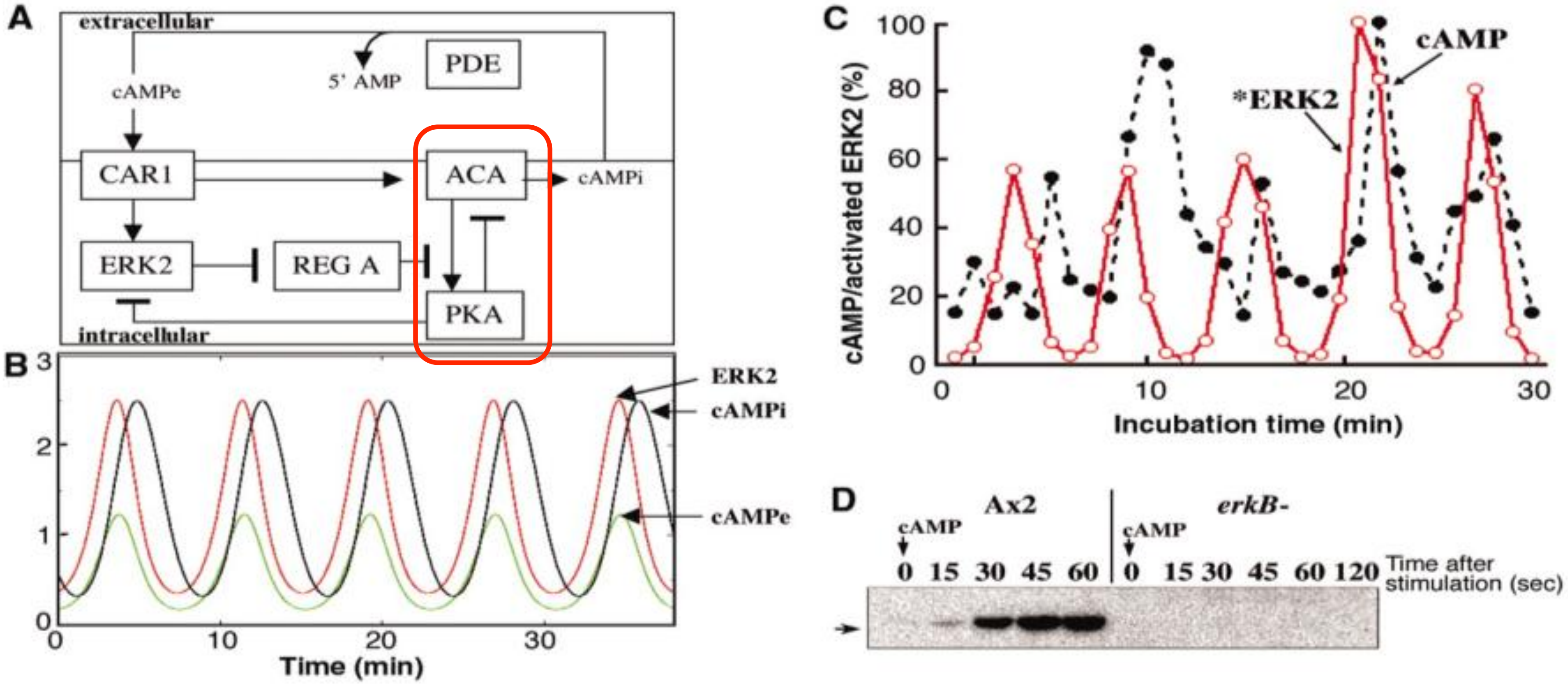


R. Albert, H. Jeong, A.L. Barabasi, Nature **406** 378 (2000)

Oscillation

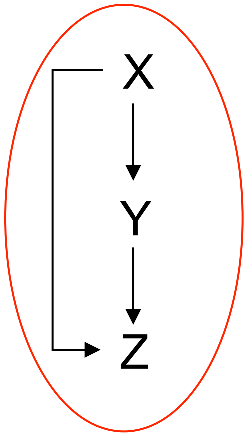


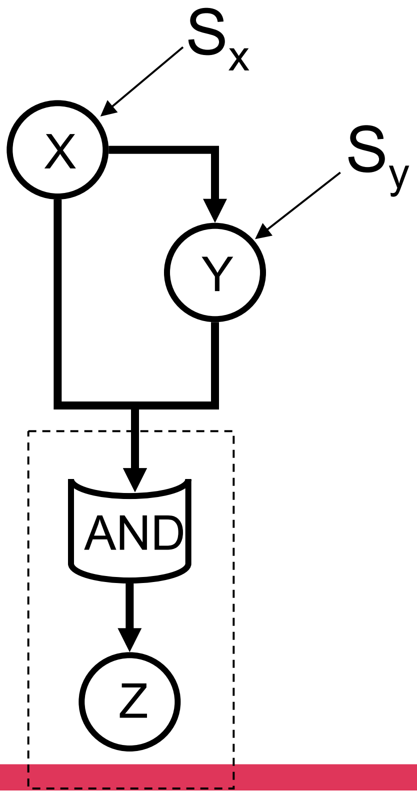


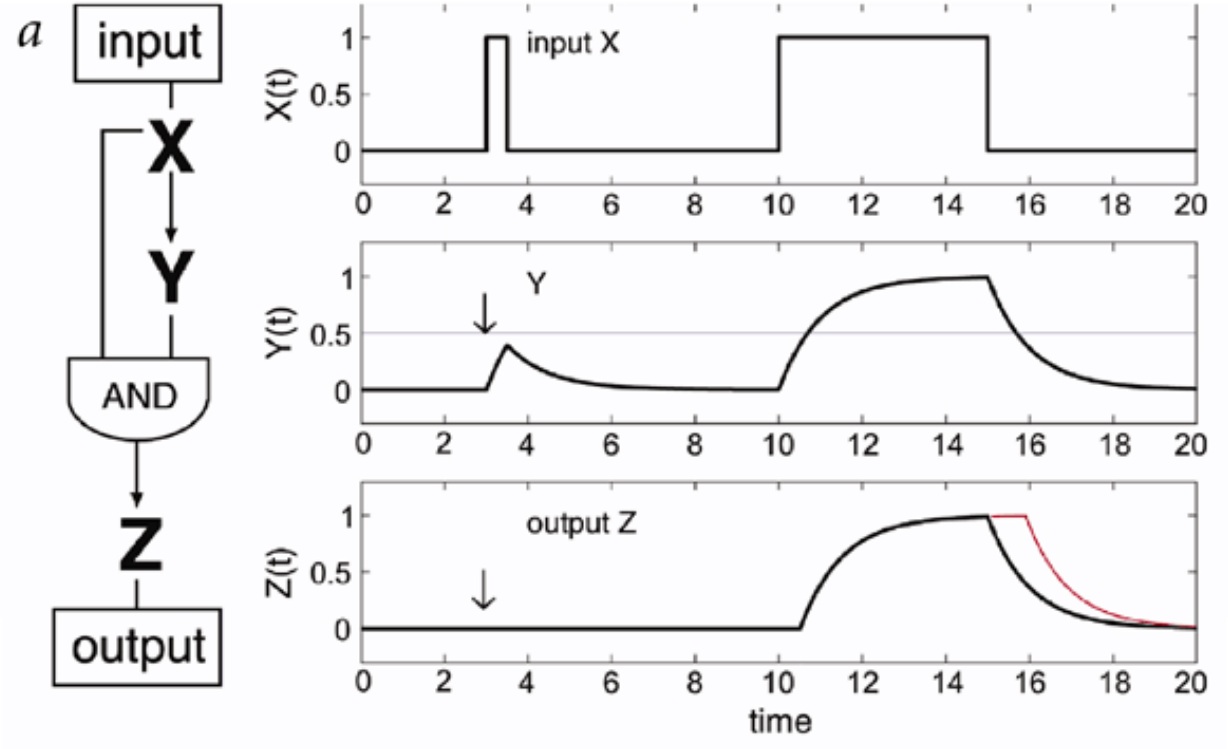


Maeda et al., Science, 304(5672):875-878, 2004

Coherent **Feed Forward Loop (FFL)**





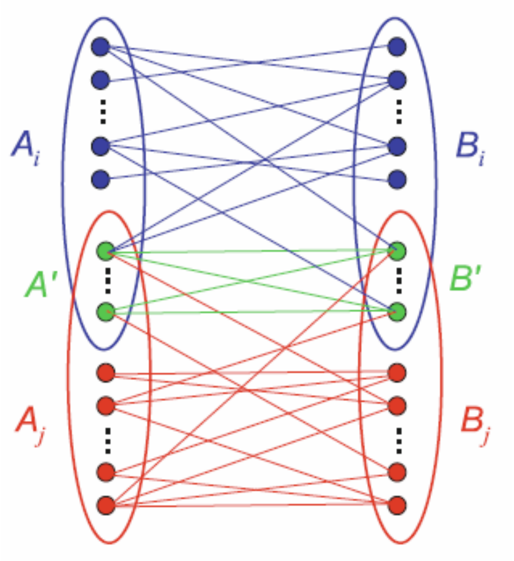


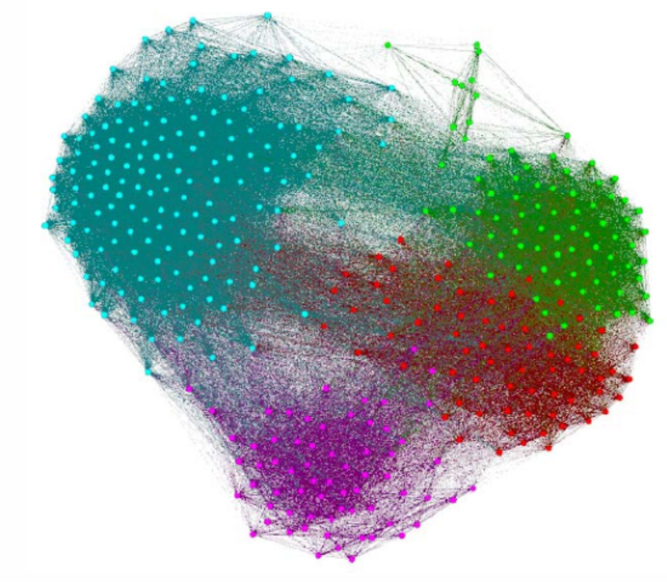
Nature Genetics **31**, 64 - 68 (2002)
Network motifs in the transcriptional regulation network of *Escherichia coli*
Shai S. Shen-Orr, Ron Milo, Shmoolik Mangan & Uri Alon

Noise (low-pass) filter

Graph Pattern Mining - Clique Enumeration and Merging

- **NetMerge guarantees subnetwork densities.**
- **NetMerge runs very fast.**



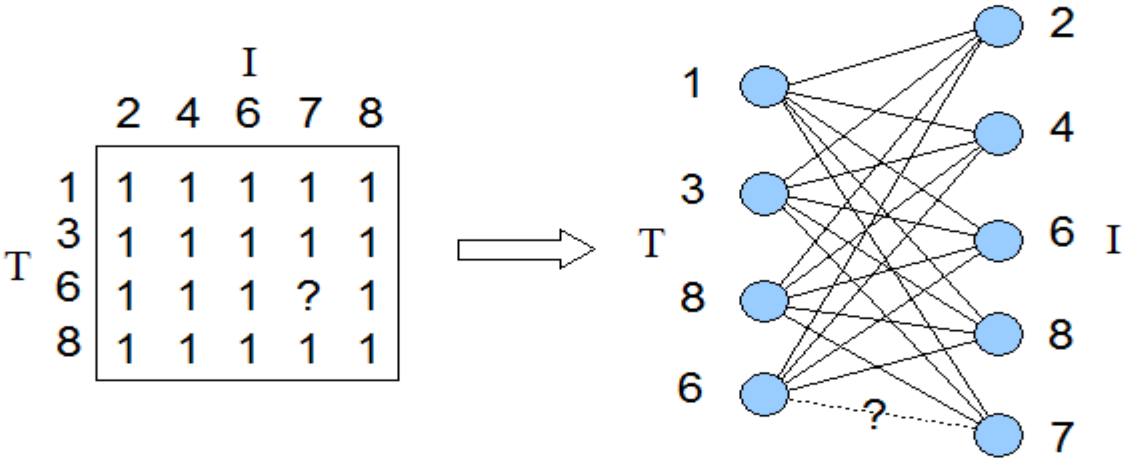


CC: extracellular matrix
BP: immune system process
BP: cell cycle process
BP: translational termination CC: cytosolic ribosome

Y. Xiang, D. Fuhry, K. Kaya, R. Jin, U. Catalyurek, K. Huang, "Merging Network Patterns: A General Framework to Summarize Biomedical Network Data", *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2012, Volume 1, Number 3, Pages 103-116.

Bipartite Graphs

- **Disease Gene Prioritization (Using Graph Pattern Mining Approaches)**



(a) (b)

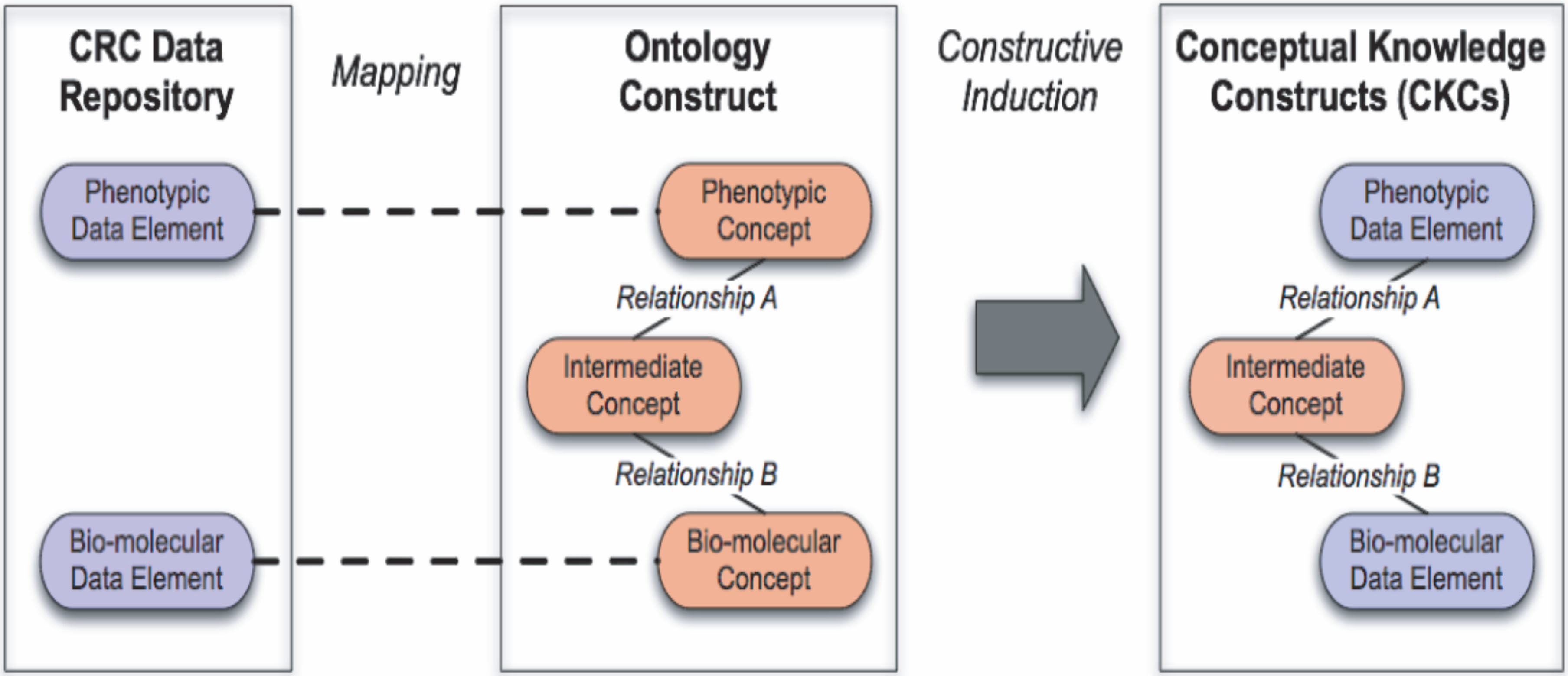
$$\mathcal{F}(i,j) = \max_{T\times I\in S(i,j)} (|T|-1)*(|I|-1)$$

Y. Xiang, P.R.O. Payne, K. Huang, "Transactional Database Transformation and Its Application in Prioritizing Human Disease Genes", IEEE-ACM Transactions on Computational Biology and Bioinformatics (TCBB). 2012, Vol 9, Issue 1, pp 294-304.

Bipartite Graphs

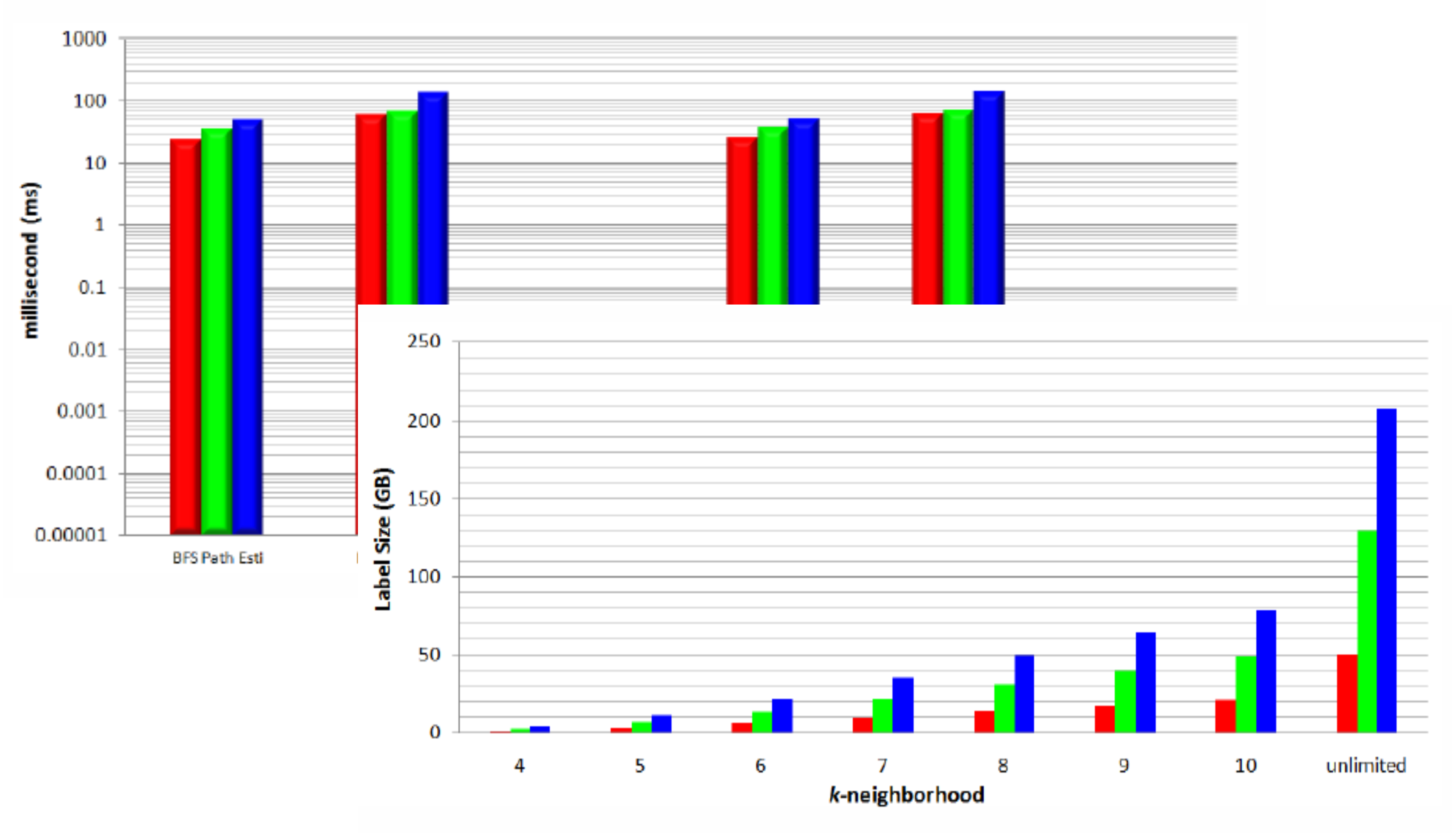
- Can you think of any biomedical applications that can be formulated using bi-partite (or multi-partite) graphs?

Graph Indexing for Unified Medical Language System (UMLS): Conceptual Knowledge Construct



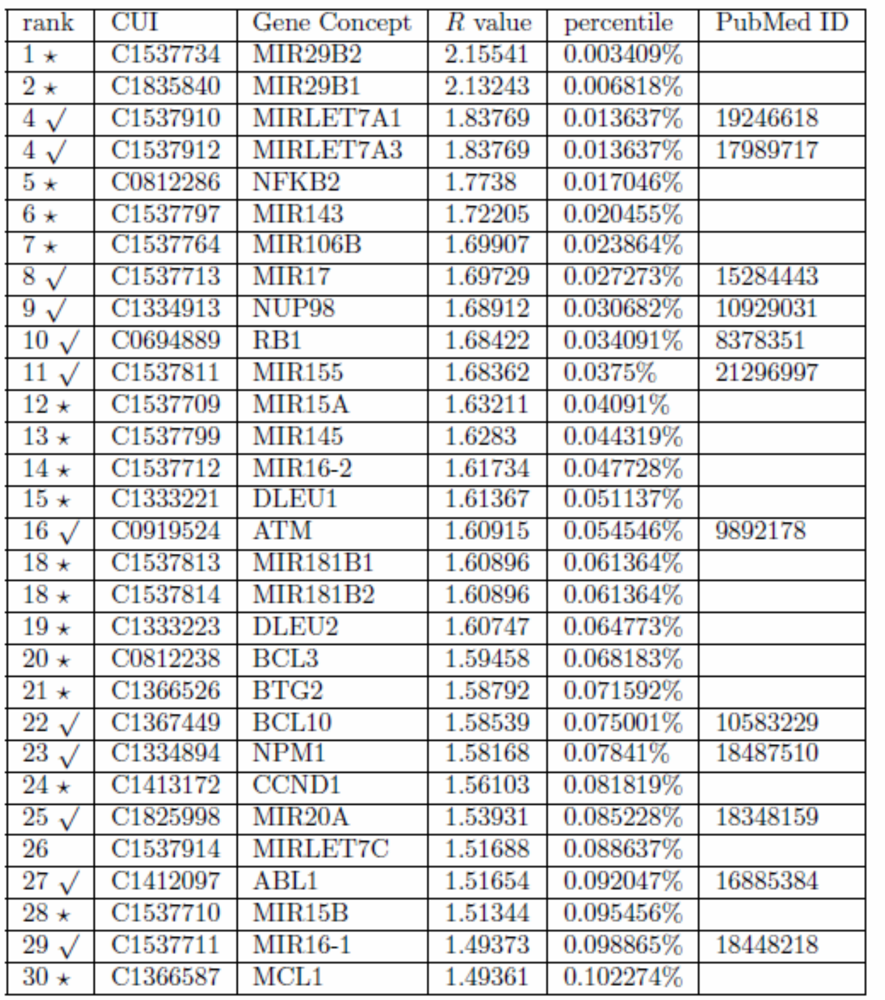
Graph Indexing: UMLS

- **kDLS *vs* BFS or DFS**

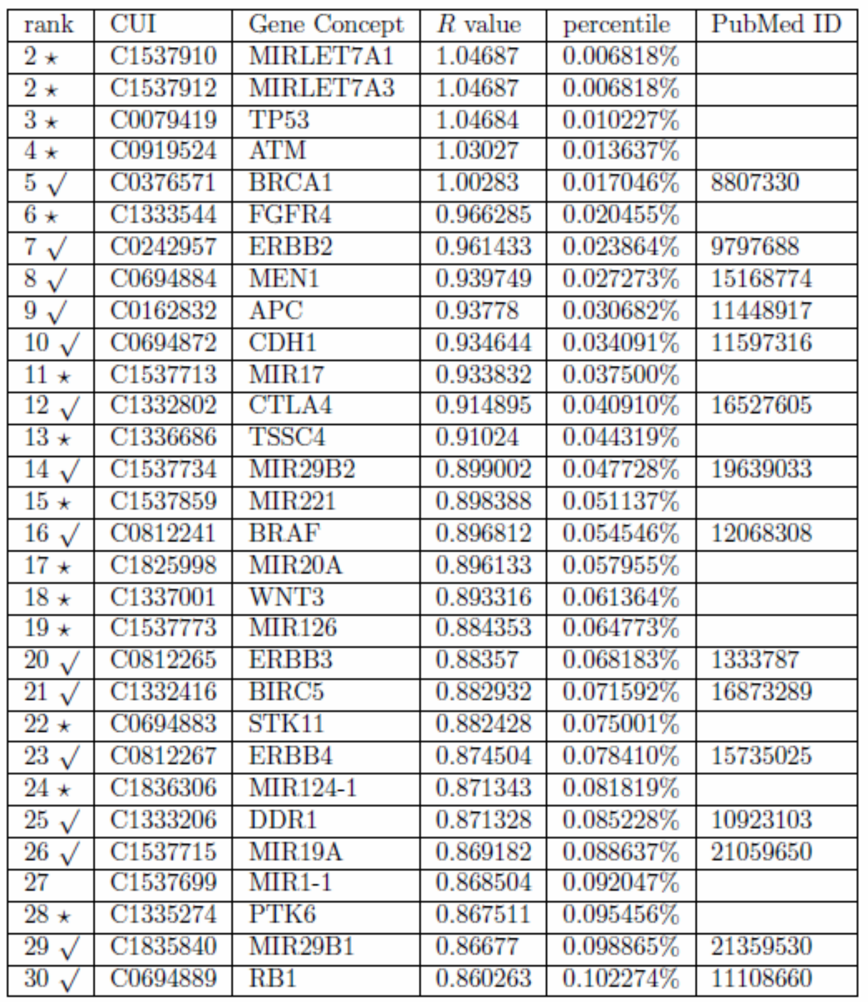


Y. Xiang, K. Lu, S. L. James, T. B. Borlawsky, K. Huang, P.R.O. Payne, *Journal of Biomedical Informatics.* 2012, Vol 45, Issue 2, pp 323-336.

Network-based Ranking



rank	CUI	Gene Concept	R value	percentile	PubMed ID
1 ★	C1537734	MIR29B2	2.15541	0.003409%	
2 ★	C1835840	MIR29B1	2.13243	0.006818%	
4 √	C1537910	MIRLET7A1	1.83769	0.013637%	19246618
4 √	C1537912	MIRLET7A3	1.83769	0.013637%	17989717
5 ★	C0812286	NFKB2	1.7738	0.017046%	
6 ★	C1537797	MIR143	1.72205	0.020455%	
7 ★	C1537764	MIR106B	1.69907	0.023864%	
8 √	C1537713	MIR17	1.69729	0.027273%	15284443
9 √	C1334913	NUP98	1.68912	0.030682%	10929031
10 √	C0694889	RB1	1.68422	0.034091%	8378351
11 √	C1537811	MIR155	1.68362	0.0375%	21296997
12 ★	C1537709	MIR15A	1.63211	0.04091%	
13 ★	C1537799	MIR145	1.6283	0.044319%	
14 ★	C1537712	MIR16-2	1.61734	0.047728%	
15 ★	C1333221	DLEU1	1.61367	0.051137%	
16 √	C0919524	ATM	1.60915	0.054546%	9892178
18 ★	C1537813	MIR181B1	1.60896	0.061364%	
18 ★	C1537814	MIR181B2	1.60896	0.061364%	
19 ★	C1333223	DLEU2	1.60747	0.064773%	
20 ★	C0812238	BCL3	1.59458	0.068183%	
21 ★	C1366526	BTG2	1.58792	0.071592%	
22 √	C1367449	BCL10	1.58539	0.075001%	10583229
23 √	C1334894	NPM1	1.58168	0.07841%	18487510
24 ★	C1413172	CCND1	1.56103	0.081819%	
25 √	C1825998	MIR20A	1.53931	0.085228%	18348159
26	C1537914	MIRLET7C	1.51688	0.088637%	
27 √	C1412097	ABL1	1.51654	0.092047%	16885384
28 ★	C1537710	MIR15B	1.51344	0.095456%	
29 √	C1537711	MIR16-1	1.49373	0.098865%	18448218
30 ★	C1366587	MCL1	1.49361	0.102274%	



rank	CUI	Gene Concept	R value	percentile	PubMed ID
2 ★	C1537910	MIRLET7A1	1.04687	0.006818%	
2 ★	C1537912	MIRLET7A3	1.04687	0.006818%	
3 ★	C0079419	TP53	1.04684	0.010227%	
4 ★	C0919524	ATM	1.03027	0.013637%	
5 √	C0376571	BRCA1	1.00283	0.017046%	8807330
6 ★	C1333544	FGFR4	0.966285	0.020455%	
7 √	C0242957	ERBB2	0.961433	0.023864%	9797688
8 √	C0694884	MEN1	0.939749	0.027273%	15168774
9 √	C0162832	APC	0.93778	0.030682%	11448917
10 √	C0694872	CDH1	0.934644	0.034091%	11597316
11 ★	C1537713	MIR17	0.933832	0.037500%	
12 √	C1332802	CTLA4	0.914895	0.040910%	16527605
13 ★	C1336686	TSSC4	0.91024	0.044319%	
14 √	C1537734	MIR29B2	0.899002	0.047728%	19639033
15 ★	C1537859	MIR221	0.898388	0.051137%	
16 √	C0812241	BRAF	0.896812	0.054546%	12068308
17 ★	C1825998	MIR20A	0.896133	0.057955%	
18 ★	C1337001	WNT3	0.893316	0.061364%	
19 ★	C1537773	MIR126	0.884353	0.064773%	
20 √	C0812265	ERBB3	0.88357	0.068183%	1333787
21 √	C1332416	BIRC5	0.882932	0.071592%	16873289
22 ★	C0694883	STK11	0.882428	0.075001%	
23 √	C0812267	ERBB4	0.874504	0.078410%	15735025
24 ★	C1836306	MIR124-1	0.871343	0.081819%	
25 √	C1333206	DDR1	0.871328	0.085228%	10923103
26 √	C1537715	MIR19A	0.869182	0.088637%	21059650
27	C1537699	MIR1-1	0.868504	0.092047%	
28 ★	C1335274	PTK6	0.867511	0.095456%	
29 √	C1835840	MIR29B1	0.86677	0.098865%	21359530
30 √	C0694889	RB1	0.860263	0.102274%	11108660

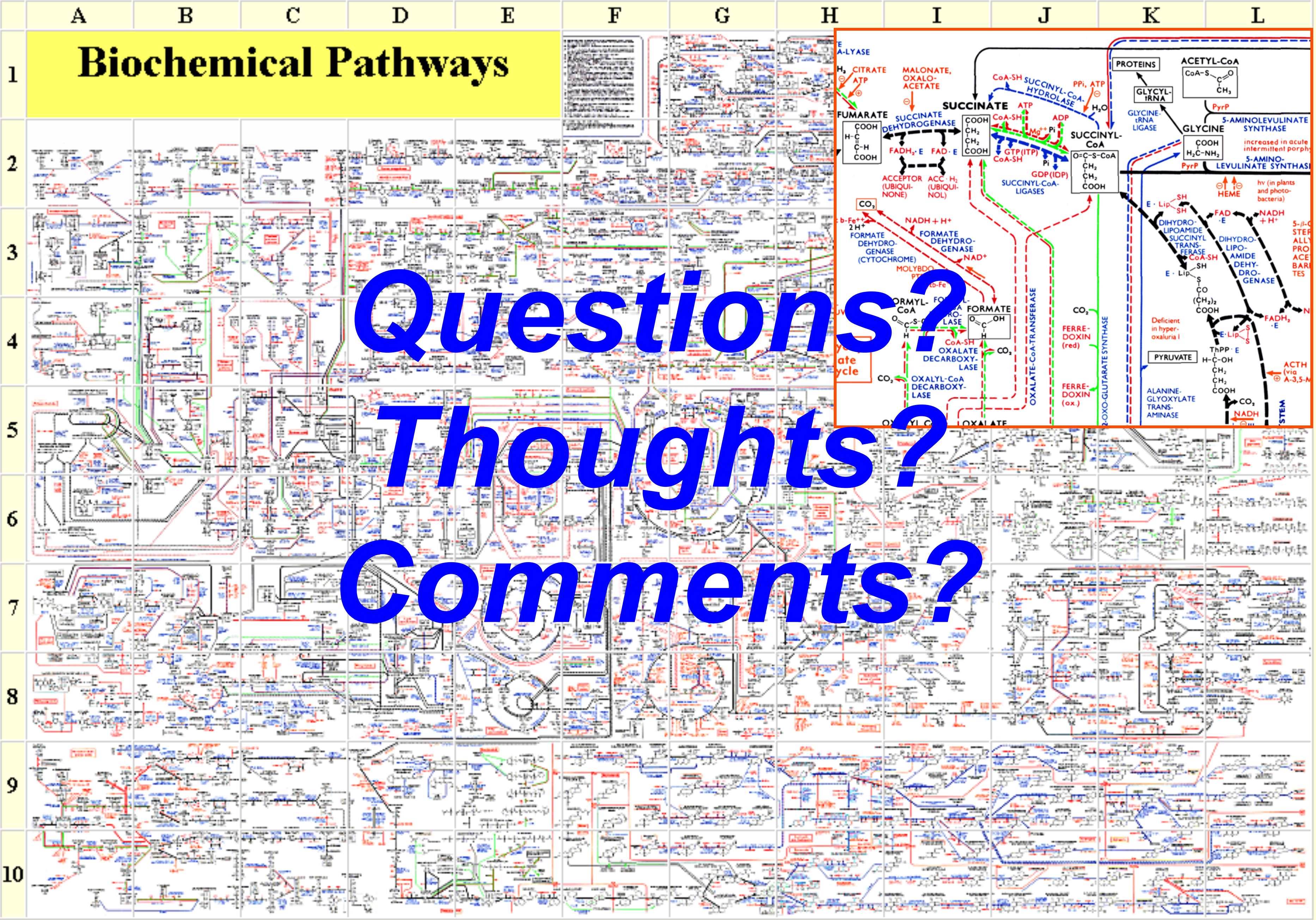
Chronic Lymphocytic Leukemia (CLL)

Breast Carcinoma

Y. Xiang, K. Lu, S. L. James, T. B. Borlawsky, K. Huang, P.R.O. Payne, *Journal of Biomedical Informatics.* 2012, Vol 45, Issue 2, pp 323-336.

Summary

- Network modeling is an effective way of studying biomedical systems.
- Many existing resources for biomedical networks.
- Network may exhibit complicated behavior and requires new theory and algorithms.



Questions?
Thoughts?
Comments?