生物信息学与系统生物学

张世华

中国科学院数学与系统科学研究院

Q6: Aging and disease are known to be closely related. Can we see this relationship in the interactome?

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Disease-Aging Network Reveals Significant Roles of Aging Genes in Connecting Genetic Diseases

Jiguang Wang[1,2], Shihua Zhang[1], Yong Wang[1], Luonan Chen[3,4]*, Xiang-Sun Zhang[1]*

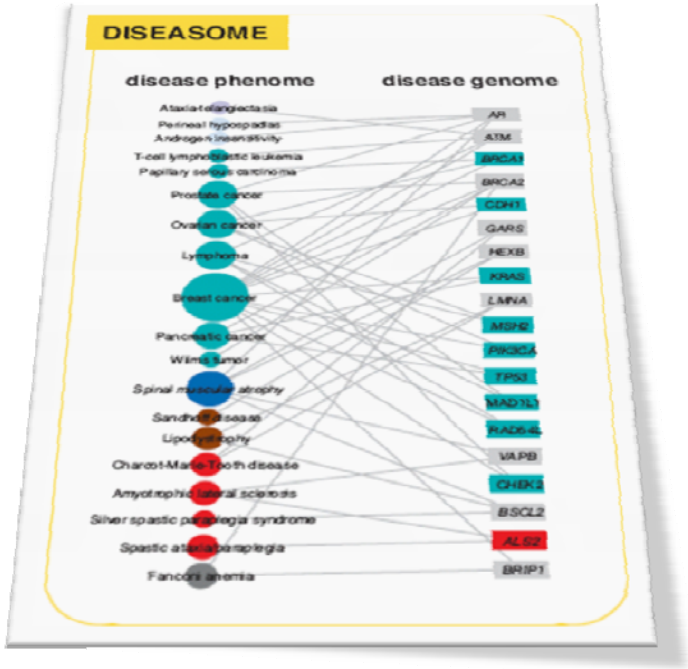
1 Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, **2** Graduate School of the Chinese Academy of Sciences, Beijing, China, **3** Institute of Systems Biology, Shanghai University, Shanghai, China, **4** Department of Electrical Engineering and Electronics, Osaka Sangyo University, Osaka, Japan

Abstract

One of the challenging problems in biology and medicine is exploring the underlying mechanisms of genetic diseases. Recent studies suggest that the relationship between genetic diseases and the aging process is important in understanding the molecular mechanisms of complex diseases. Although some intricate associations have been investigated for a long time, the studies are still in their early stages. In this paper, we construct a human disease-aging network to study the relationship among aging genes and genetic disease genes. Specifically, we integrate human protein-protein interactions (PPIs), disease-gene associations, aging-gene associations, and physiological system–based genetic disease classification information in a single graph-theoretic framework and find that (1) human disease genes are much closer to aging genes than expected by chance; and (2) diseases can be categorized into two types according to their relationships with aging. Type I diseases have their genes significantly close to aging genes, while type II diseases do not. Furthermore, we examine the topological characters of the disease-aging network from a systems perspective. Theoretical results reveal that the genes of type I diseases are in a central position of a PPI network while type II are not; (3) more importantly, we define an asymmetric closeness based on the PPI network to describe relationships between diseases, and find that aging genes make a significant contribution to associations among diseases, especially among type I diseases. In conclusion, the network-based study provides not only evidence for the intricate relationship between the aging process and genetic diseases, but also biological implications for prying into the nature of human diseases.

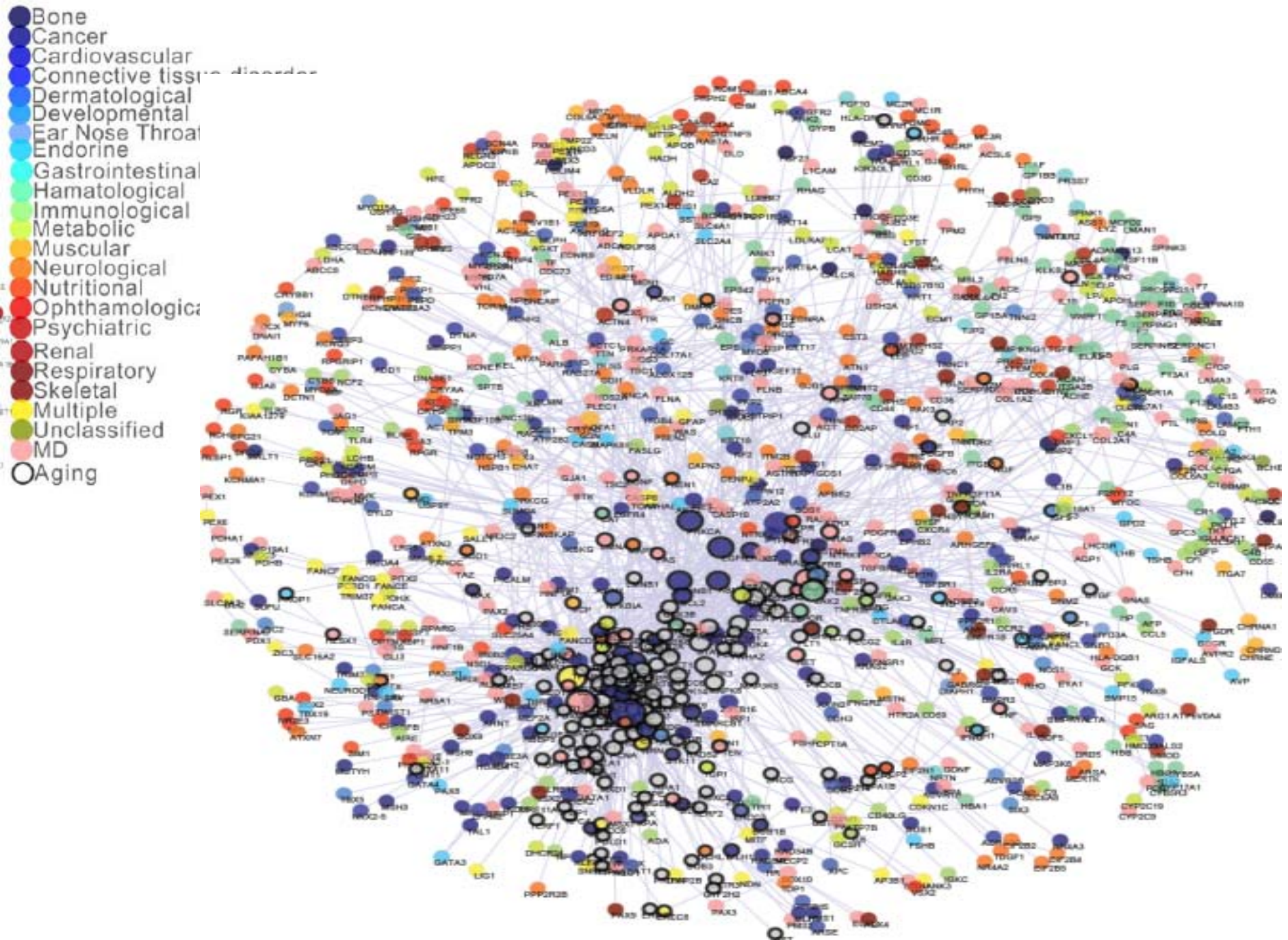
Aging

?



Disease

Association


Bone
Cancer
Cardiovascular
Connective tissu
Dermatological
Developmental
Ear Nose Throat
Endorine
Gastrointestinal
Hamatological
Immunological
Metabolic
Muscular
Neurological
Nutritional
Ophthamologica
Psychiatric
Renal
Respiratory
Skeletal
Multiple
Unclassified
MD
Aging

Results

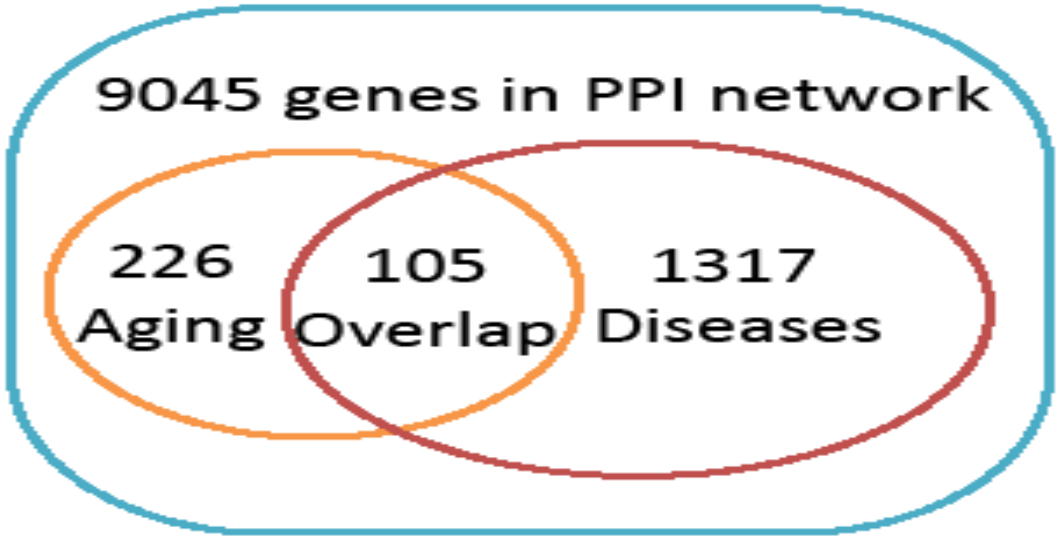
(1) Human disease genes are much closer to aging genes than expected by chance.

(2) Diseases can be categorized into two types according to their relationships with aging.

- ✓ Type I diseases have their genes significantly close to aging genes, while
- ✓ type II diseases do not.

(3) Aging genes make a significant contribution to associations among diseases.

Association?



Degree of aging genes	Average degree	Disease genes		
		Observed	Random	P-value
<20	9.38	2.51	1.99	7.3e-8
20-50	33.33	8.53	7.05	7.8e-7
50-100	69.27	17.49	14.52	1.9e-8
>100	139.81	33.86	28.82	1.4e-7

Results

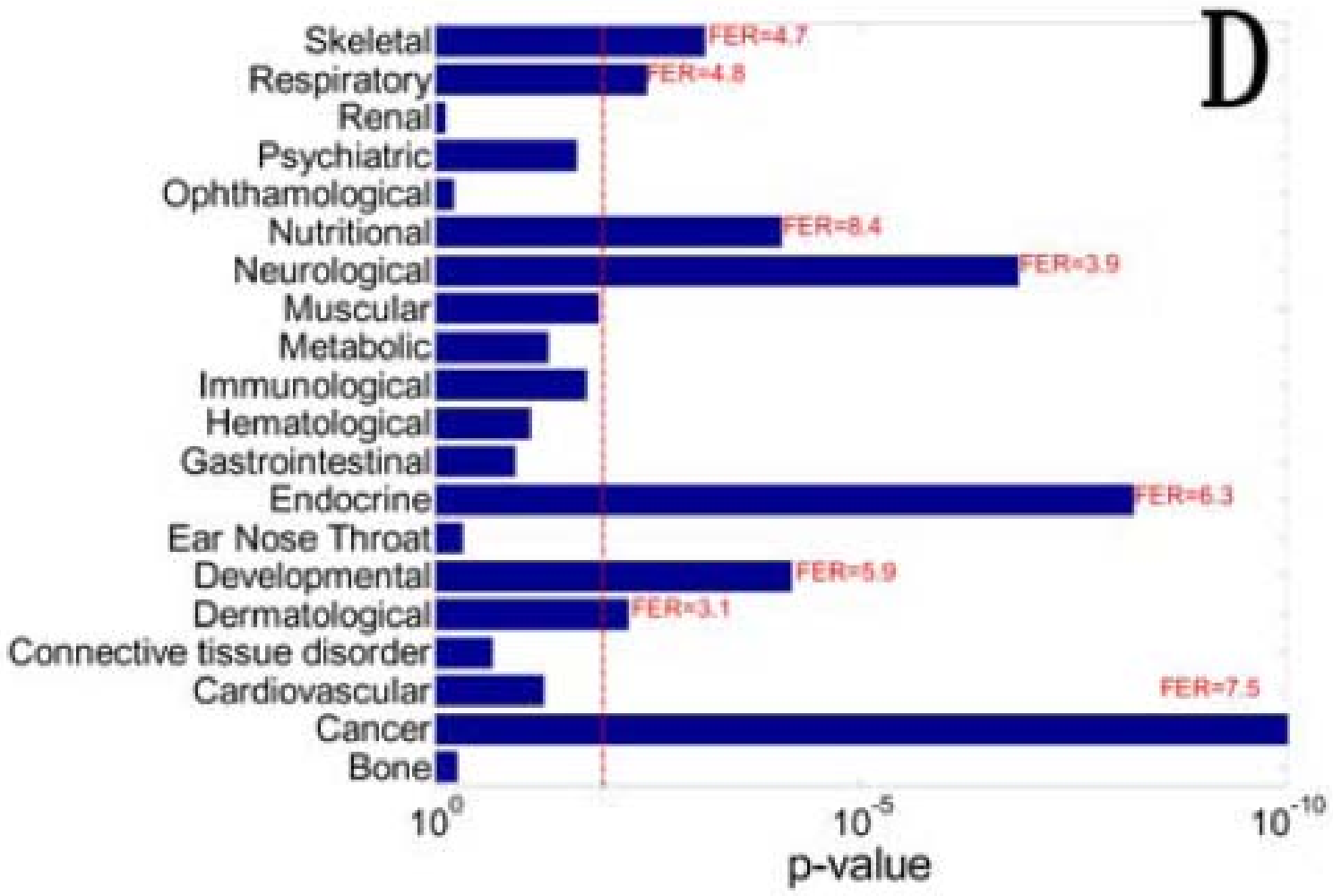
(1) Human disease genes are much closer to aging genes than expected by chance.

(2) Diseases can be categorized into two types according to their relationships with aging.

- ✓ Type I diseases have their genes significantly close to aging genes, while
- ✓ type II diseases do not.

(3) Aging genes make a significant contribution to associations among diseases.

Two types based on connection



Two types show functional diversity

Table 2. Different GOA enrichments of ARD and NARD.

GO-ID	ARD		NARD		Description
	p-value	#Genes	p-value	#Genes	
3676	1.4e-4	156	1.1e-10(under)	68	nucleic acid binding
5634	3.2e-13	193	2.2e-7(under)	79	nucleus
6139	5.0e-19	194	3.7e-03(under)	113	nucleobase, nucleoside, nucleotide and nucleic acid metabolic proc
5622	1.1e-9	411	>0.01	391	intracellular
16301	2.4e-8	63	>0.01	44	oxidoreductase activity
30528	5.3e-15	112	>0.01	49	transcription regulator activity
43170	3.4e-11	313	>0.01	295	macromolecule metabolic process
3824	>0.01	206	1.6e-8	282	catalytic activity
5478	>0.01	58	3.9e-10	101	transporter activity
9055	>0.01	12	8.3e-7	56	catabolic process
9056	>0.01	29	2.5e-5	85	biosynthetic process
9405	>0.01	2	7.6e-7	20	cell surface
9929	>0.01	11	2.9e-7	60	ion transmembrane transporter activity
15075	>0.01	36	8.5e-6	37	channel activity
5941	>0.01	1	4.6e-4	6	unlocalized protein complex
16740	>0.01	76	1.2e-5	129	hydrolase activity
16787	>0.01	88	1.9e-5	20	lyase activity
16874	>0.01	13	1.4e-7	113	cell differentiation

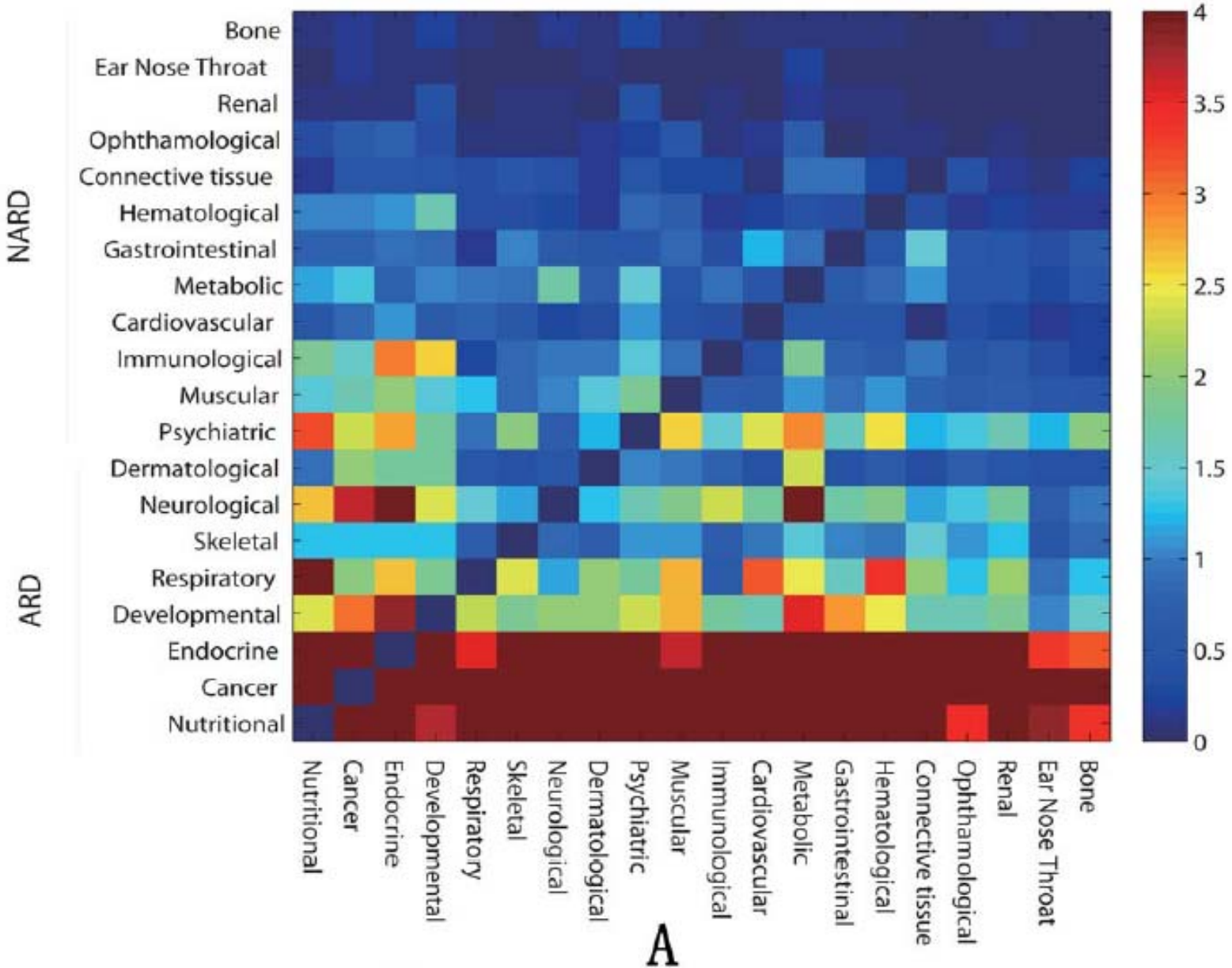
Results

(1) Human disease genes are much closer to aging genes than expected by chance.

(2) Diseases can be categorized into two types according to their relationships with aging.

- ✓ Type I diseases have their genes significantly close to aging genes, while
- ✓ type II diseases do not.

(3) Aging genes make a significant contribution to associations among diseases.


NARD
ARD
Bone
Ear Nose Throat
Renal
Ophthamological
Connective tissue
Hematological
Gastrointestinal
Metabolic
Cardiovascular
Immunological
Muscular
Psychiatric
Dermatological
Neurological
Skeletal
Respiratory
Developmental
Endocrine
Cancer
Nutritional
4
3.5
3
2.5
2
1.5
1
0.5
0
Nutritional
Cancer
Endocrine
Developmental
Respiratory
Skeletal
Neurological
Dermatological
Psychiatric
Muscular
Immunological
Cardiovascular
Metabolic
Gastrointestinal
Hematological
Connective tissue
Ophthamological
Renal
Ear Nose Throat
Bone
A

Q7: Regarding to **evolution principles**, is the subnetwork and the whole interactome the same?

TF subnetwork vs whole network

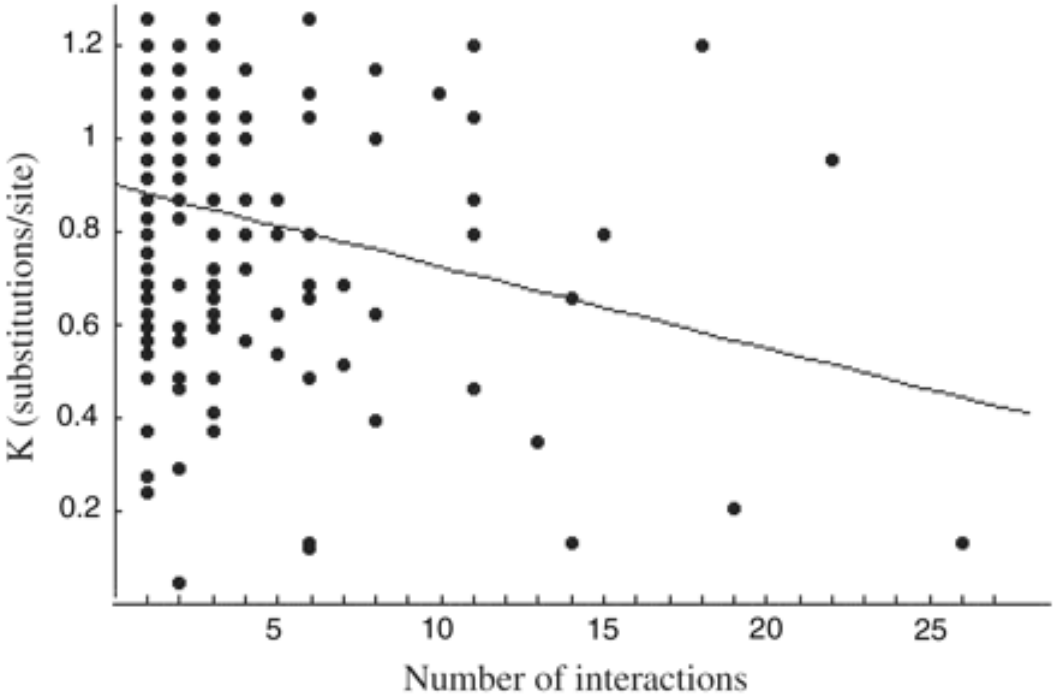
- We study evolutionary principles in the network of an important subset of proteins, the transcription factors (TFs).

- TFs are important regulators of cellular processes at the transcriptional level.

- The interactions and coordinated actions of multiple TFs in the TF network provide a primary mechanism for achieving fine-tuned transcriptional control in eukaryotes.

Well-known result

Hubs in the *S. cerevisiae* protein-protein interaction network tend to evolve more slowly than non-hubs



A protein's number of interaction partners exerts some influence on its evolutionary rate, most likely due to increased structural co-evolutionary constraints imposed by protein-protein interaction (negative selection) .

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science 296: 750-752.

Surprising findings

- Hubs in the yeast TF network tend to evolve more quickly than non-hubs;

- This result holds for all four major types of TF hubs:
  - Interaction hubs that interact with many other TFs
  - Regulatory in-degree hubs that are regulated by many TFs
  - Regulatory out-degree hubs that regulate many TFs
  - Co-regulatory hubs that jointly regulate target genes (TGs) with many other TFs.

Y. Wang, E. Franzosa, X.S. Zhang, and Y. Xia. *Nucleic Acids Research*, 38(18): 5959–5969, 2010.

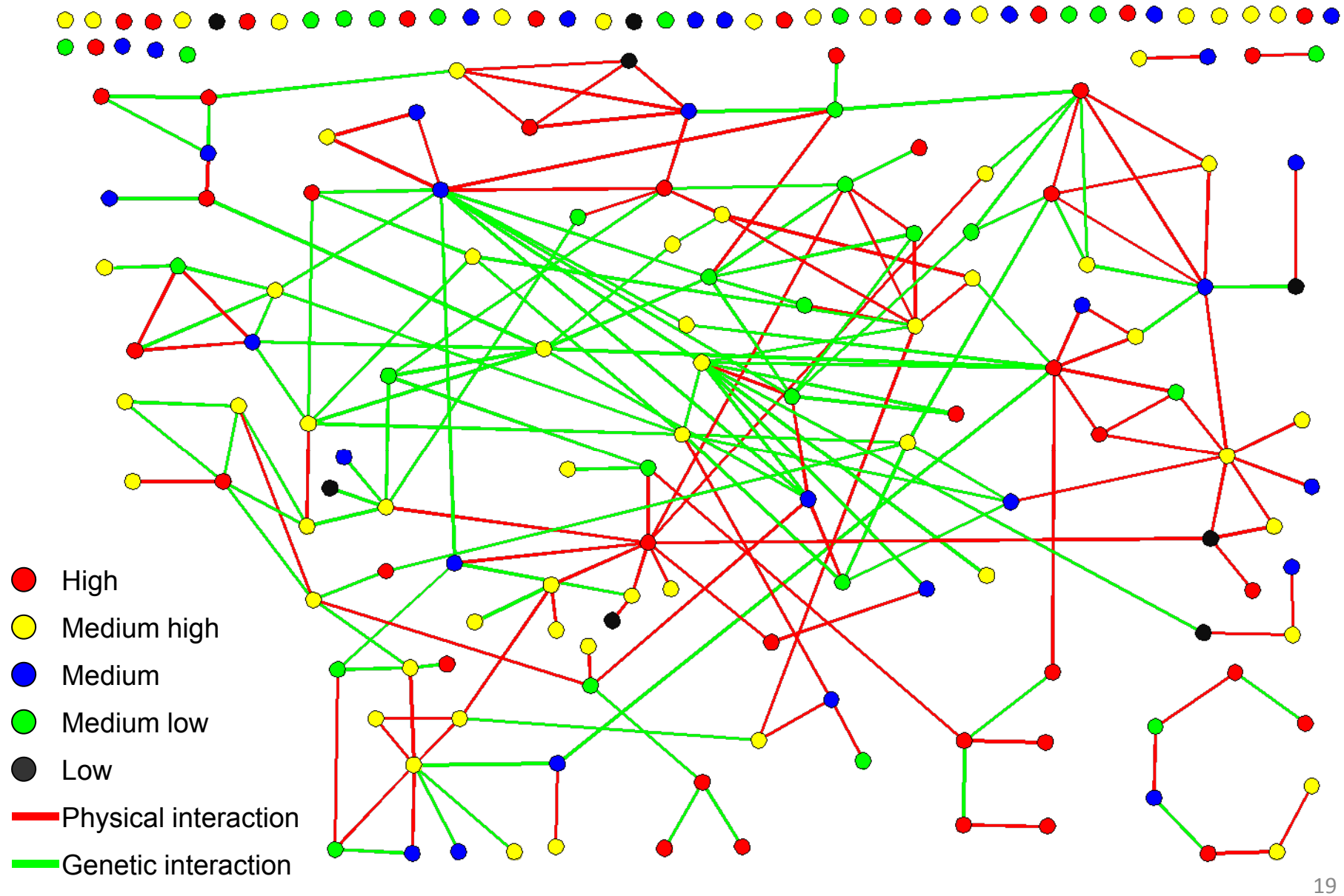
TF networks

- We collected 174 yeast TFs and assembled the whole-genome TF network based on three types of associations:
  - protein-protein interactions among TFs (forming the TF interactome)
  - transcriptional regulatory relationships among TFs (forming the TF transcriptional regulatory network)
  - joint regulation of target genes among TFs (forming the TF co-regulatory network)

Y. Wang, E. Franzosa, X.S. Zhang, and Y. Xia. *Nucleic Acids Research*, 38(18): 5959–5969, 2010.

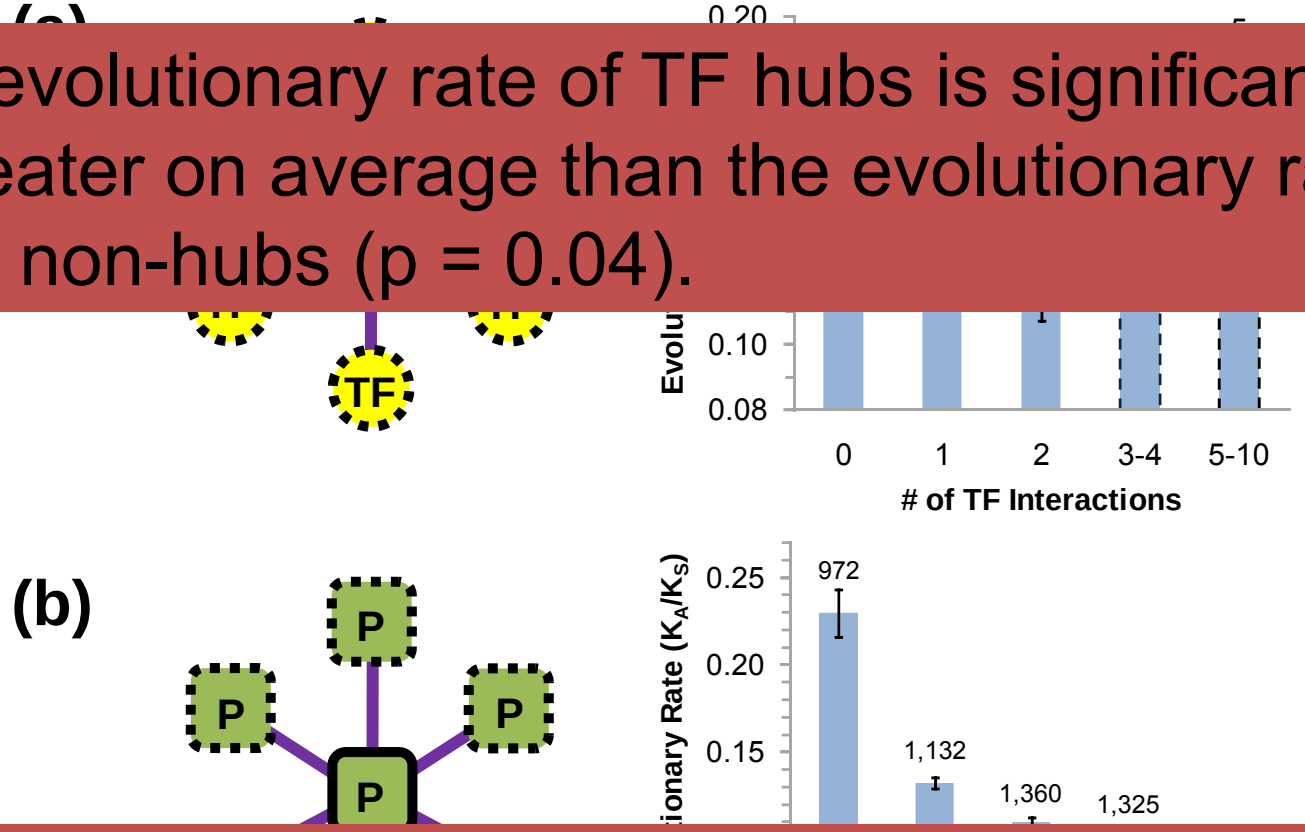
Evolutionary rate

- **Evolutionary rate** was measured as the K_A/K_S ratio calculated over alignments between the coding sequences of *S. cerevisiae* and their orthologs in *S. paradoxus* (the closest related yeast with a sequenced genome).
  - K_A/K_S is the ratio of the rate of non-synonymous substitutions (K_A) to the rate of synonymous substitutions (K_S), and serves as an approximate measure of the strength of sequence selection acting on a protein (factoring out mutational background and translational selection).
  - Smaller K_A/K_S values are associated with heightened purifying selection (reduced evolutionary rate), while larger values are associated with neutral or adaptive evolution (increased evolutionary rate).


High
Medium high
Medium
Medium low
Low
Physical interaction
Genetic interaction

TF interaction hubs evolve fast

The evolutionary rate of TF hubs is significantly greater on average than the evolutionary rate of TF non-hubs (p = 0.04).

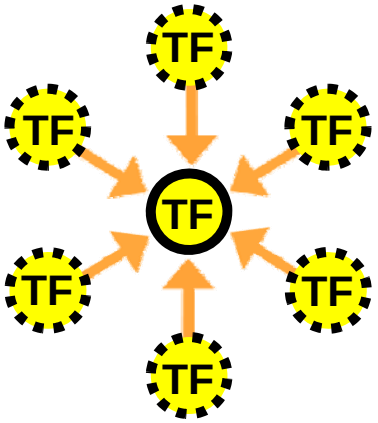


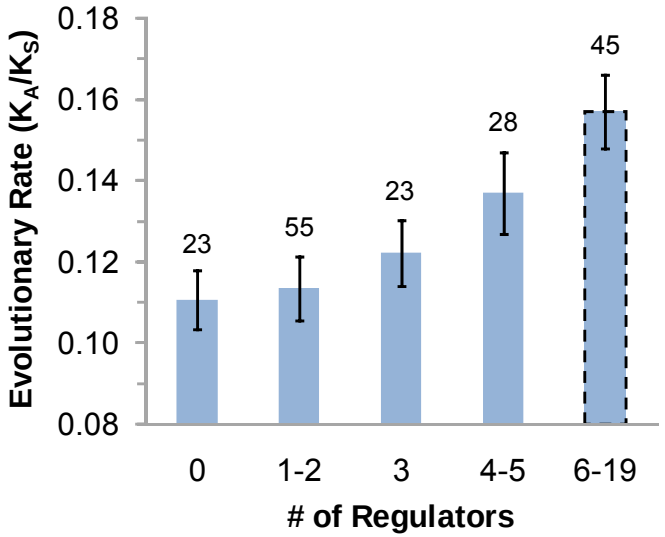
The mean of these sampled correlations between protein evolutionary rate and generic protein-protein interactions is significantly different from the observed correlation between TF evolutionary rate and TF-TF interactions ($p < 1.0 \times 10^{-6}$).

TF interaction hubs evolve fast

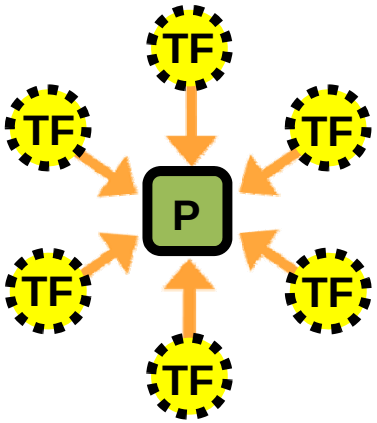
We conclude that TF-TF interactions and generic protein-protein interactions evolve in very different ways: hubs in the protein interactome tend to evolve more slowly than non-hubs, whereas hubs in the TF interactome tend to evolve more quickly than non-hubs.

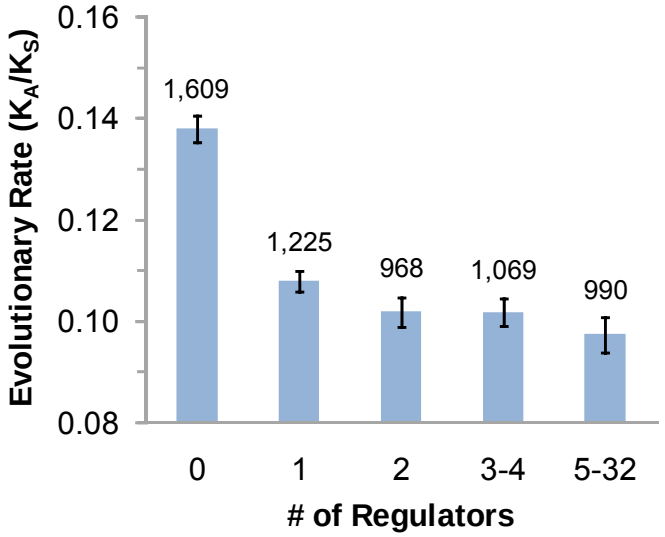
(a)


TF
TF
TF
TF
TF
TF
TF

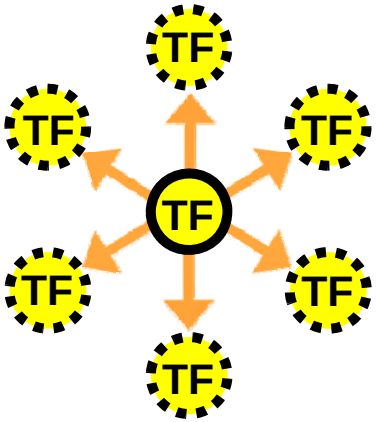

Evolutionary Rate (K_A/K_S)
0.18
0.16
0.14
0.12
0.10
0.08
23
55
23
28
45
0
1-2
3
4-5
6-19
# of Regulators

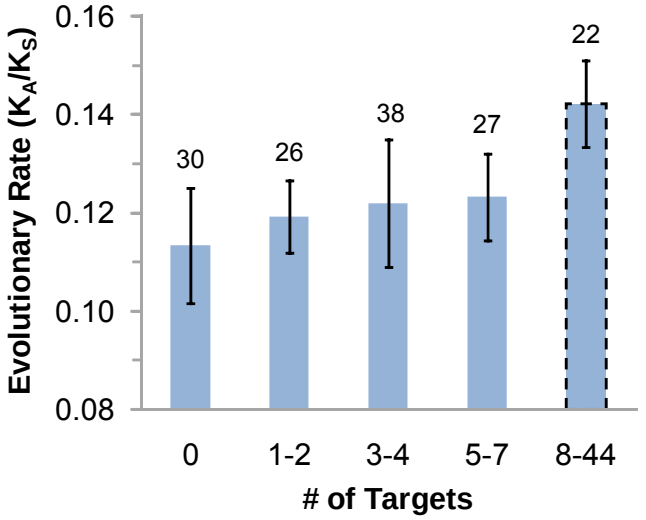
(b)


TF
TF
TF
P
TF
TF
TF

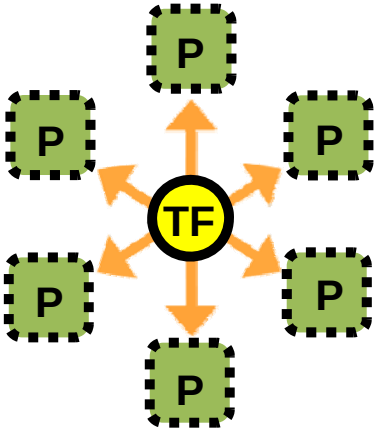

Evolutionary Rate (K_A/K_S)
0.16
0.14
0.12
0.10
0.08
1,609
1,225
968
1,069
990
0
1
2
3-4
5-32
# of Regulators

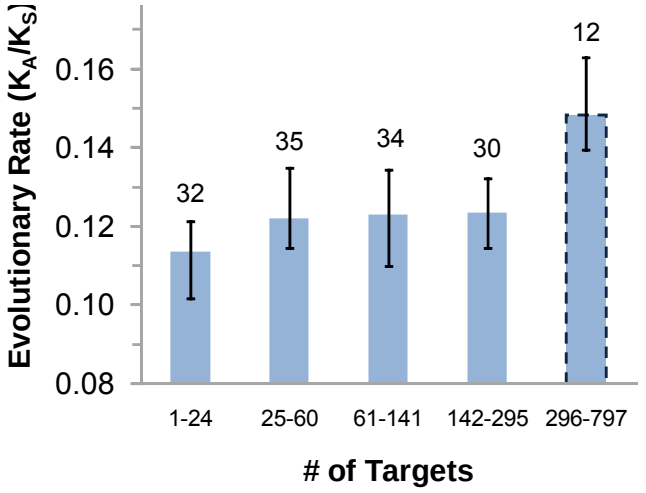
(a)

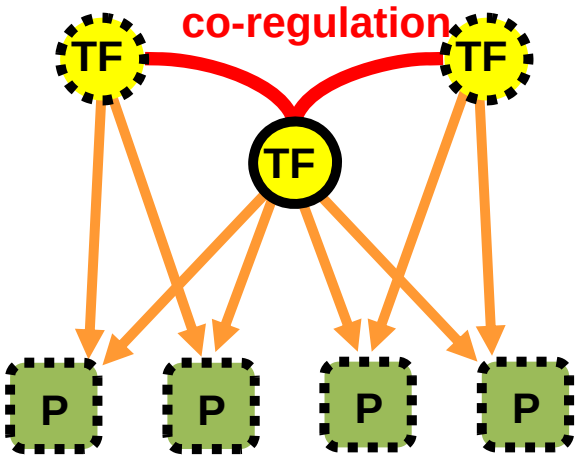

TF
TF
TF
TF
TF
TF
TF

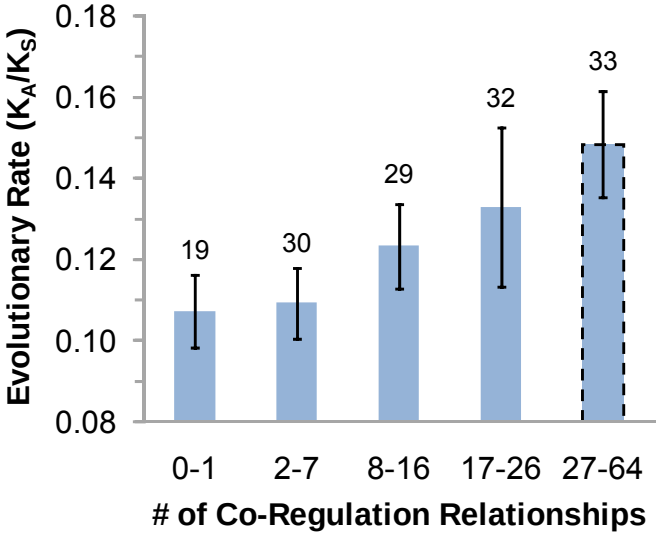

Evolutionary Rate (K_A/K_S)
0.16
0.14
0.12
0.10
0.08
30
26
38
27
22
0
1-2
3-4
5-7
8-44
# of Targets

(b)


P
P
P
TF
P
P
P

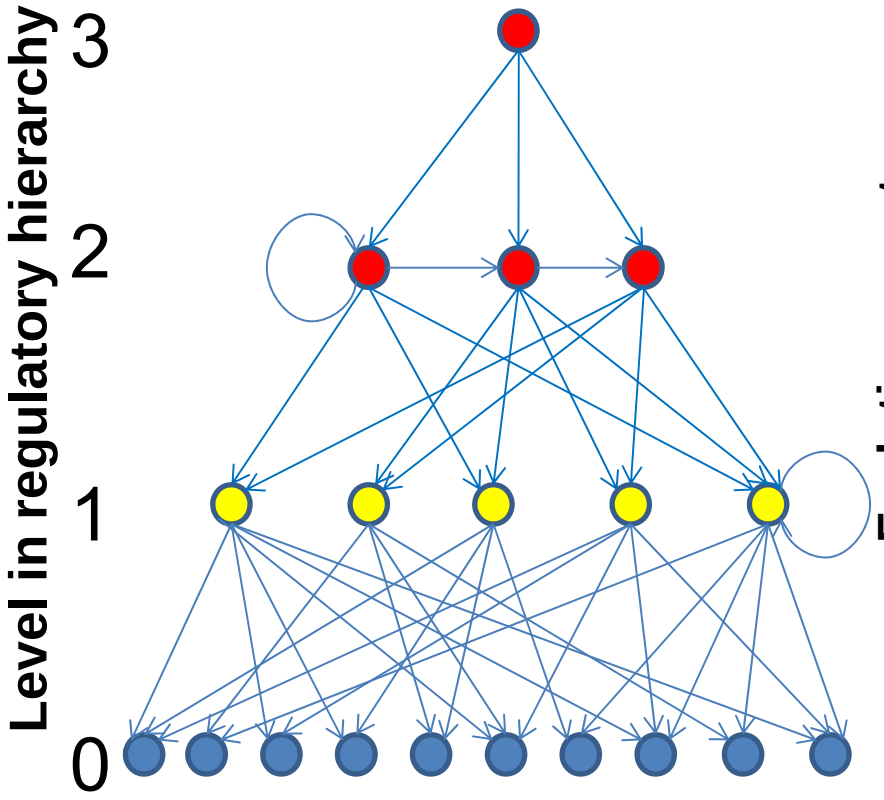

Evolutionary Rate (K_A/K_S)
0.16
0.14
0.12
0.10
0.08
32
35
34
30
12
1-24
25-60
61-141
142-295
296-797
# of Targets

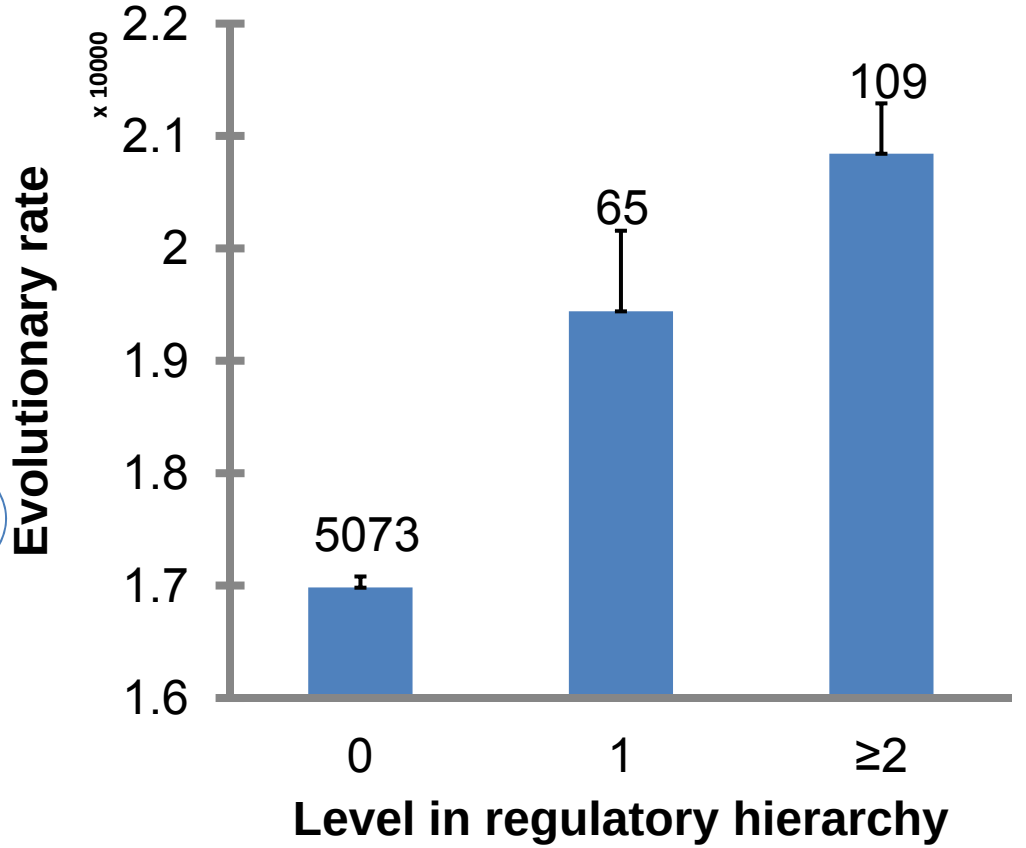

co-regulation
TF
TF
TF
P
P
P
P


Evolutionary Rate (K_A/K_S)
0.18
0.16
0.14
0.12
0.10
0.08
19
30
29
32
33
0-1
2-7
8-16
17-26
27-64
# of Co-Regulation Relationships

Network rewiring model

- We hypothesize that protein-protein interactions operate at a low level in the cellular network, and tend to be conserved during evolution.
- On the other hand, TF-TF associations operate at a high level in the cellular regulatory hierarchy, and tend to rewire during evolution.
- Protein-protein interactions are fundamental to the basic functions of a living cell; more interaction partners for a particular protein will lead to greater structural and functional constraint, resulting in negative selection.
- In contrast, TF-TF associations are more easily changed in evolution compared to protein-protein interactions. Positive selection acts to fix specific TF-TF associations that are beneficial to a particular organism in a particular environment. The rewiring of TF-TF associations also encourages adaptive TF evolution.


Level in regulatory hierarchy
3
2
1
0


Evolutionary rate
x 10000
2.2
2.1
2
1.9
1.8
1.7
1.6
5073
65
109
0
1
≥2
Level in regulatory hierarchy

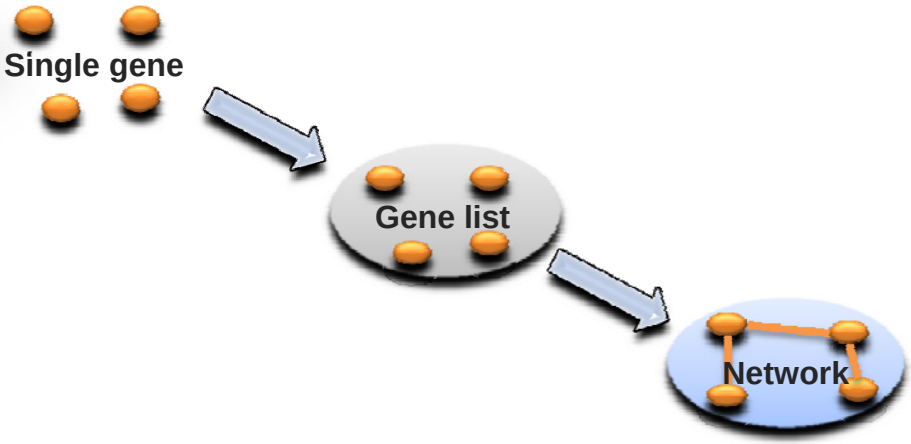
Lesson learned

- We observe that while generic protein hubs tend to evolve more slowly than non-hubs, TF hubs tend to evolve more quickly than TF non-hubs.

- We made the surprising finding that two of the most important interactome subnetworks, the TF interactome and the protein interactome, are fundamentally different in terms of their function and evolution.

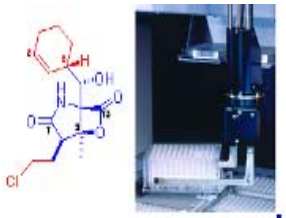
- Our work demonstrates a high degree of functional and evolutionary heterogeneity within biological networks, and highlights the rich insights that can be gained from modeling biomolecular subnetworks.

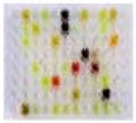
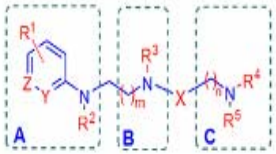
Network Ontology Analysis

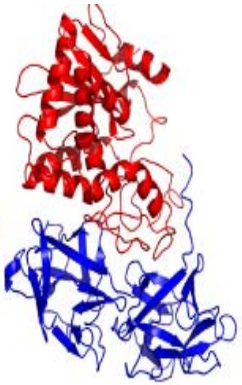


J.Wang, et al. NAR, 2011.

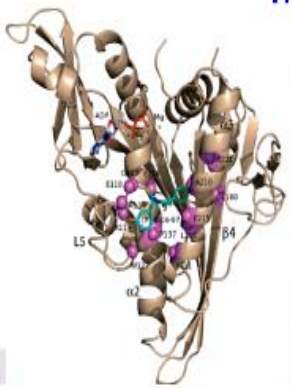
High-throughput data

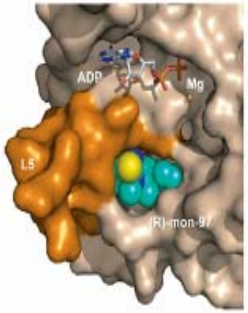


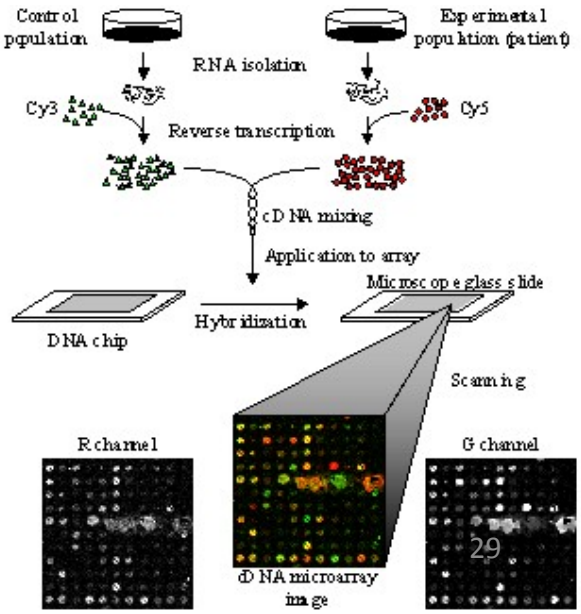



High Throughput Screening & Combinatorial Chemistry



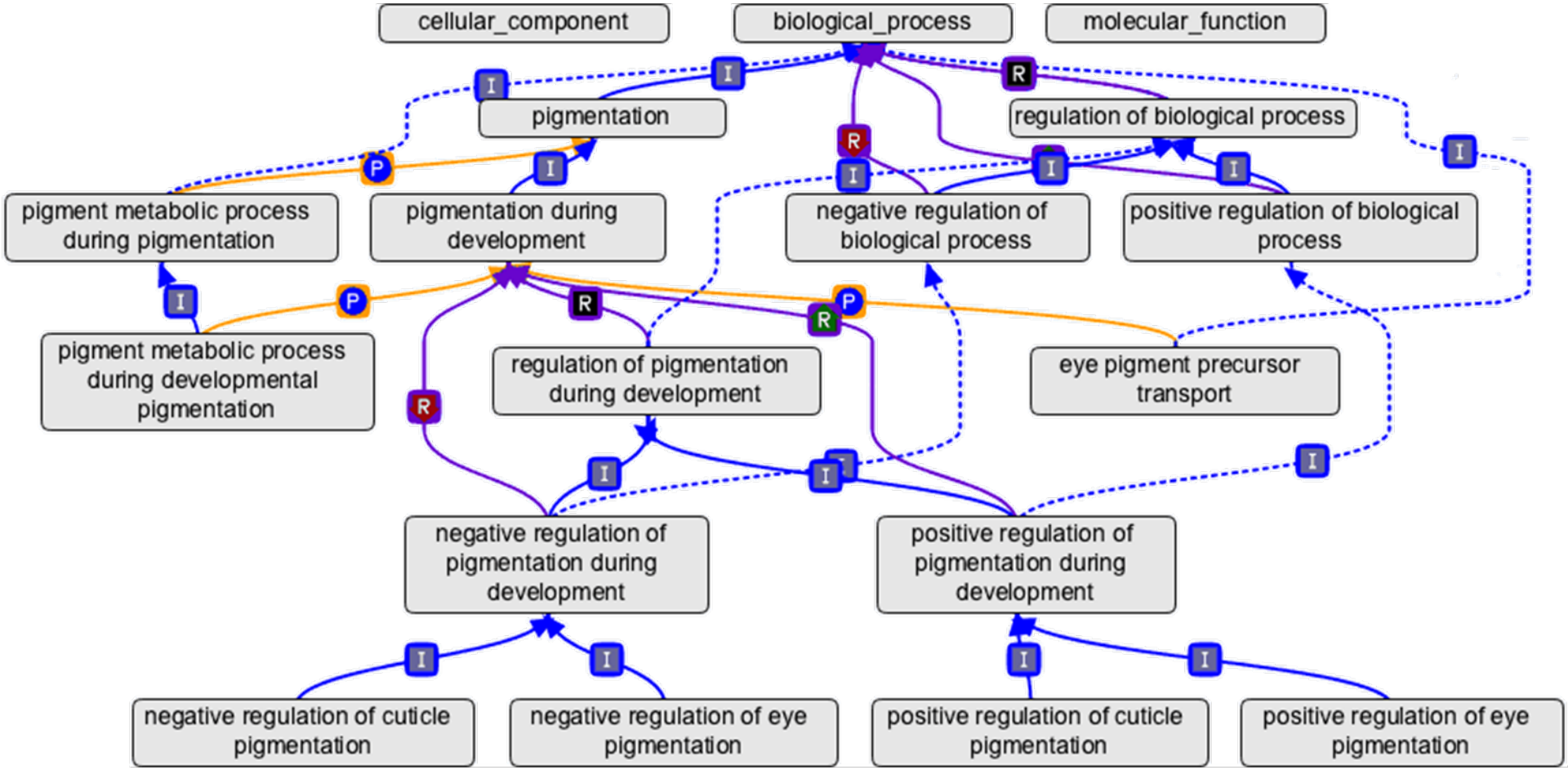




the Gene Ontology

Tool for the unification of biology

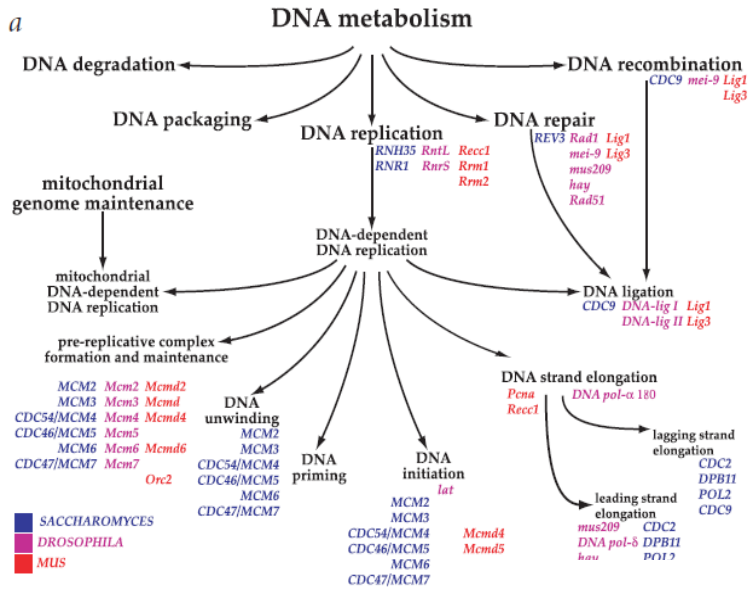


GO structure: directed acyclic graph

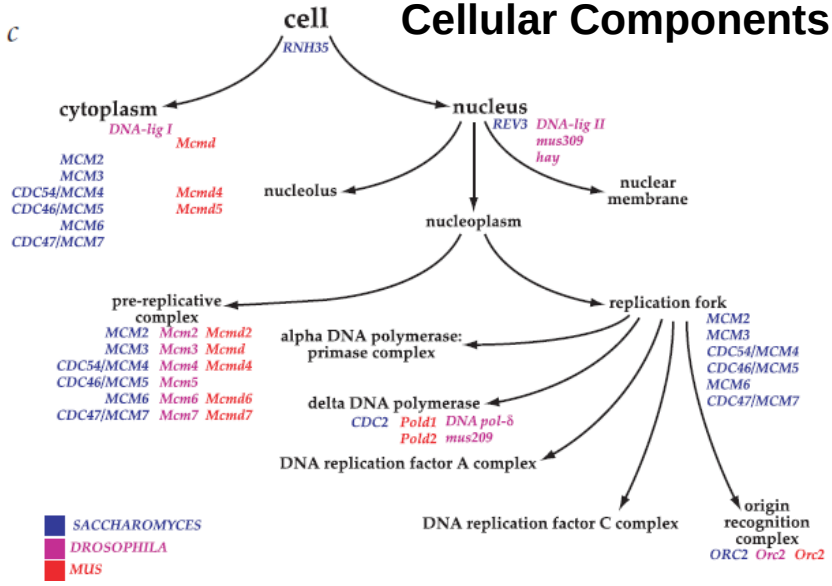
Nature genetics, 2000

the Gene Ontology

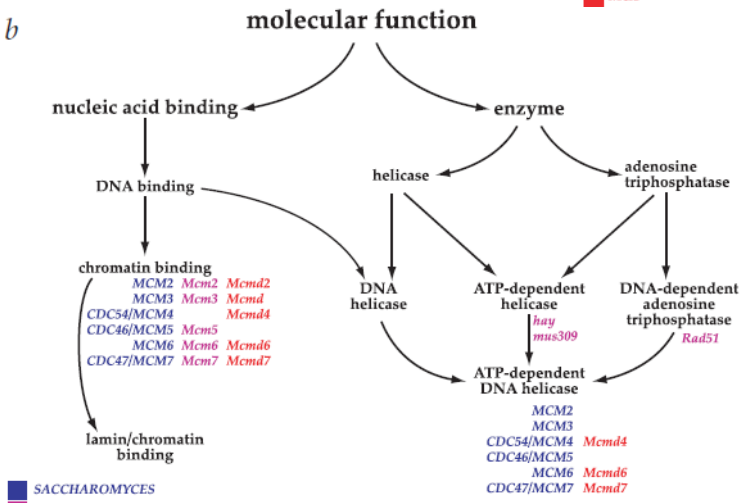
Tool for the unification of biology



Biological Processes



Cellular Components



Molecular Functions

GO structure: directed acyclic graph

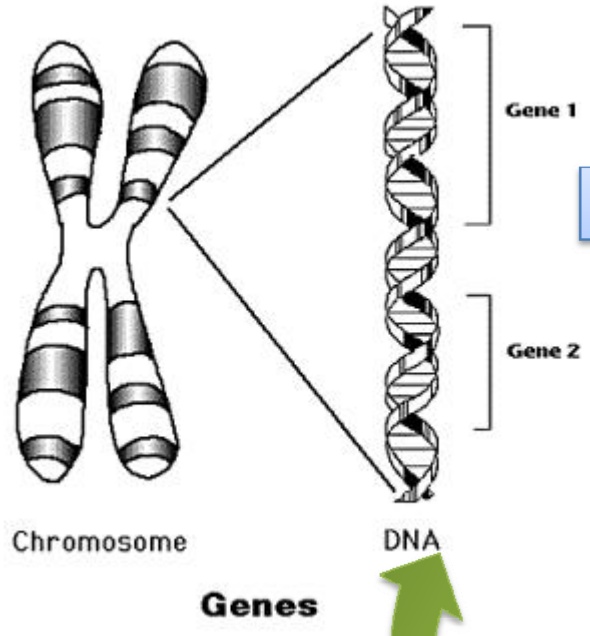
Nature genetics, 2000

Ontology?

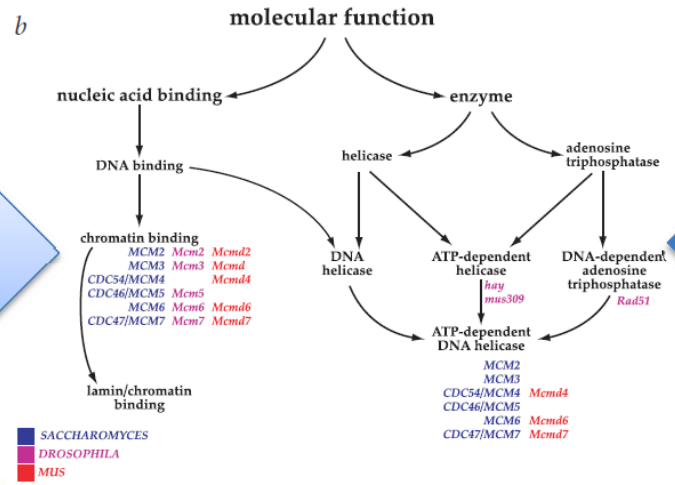
- 存在论（Ontology）是哲学的核心领域。顾名思义，存在论即关于“存在”的理论，是关于存在是什么以及存在如何存在的理论。存在论虽然是在17世纪才由德国经院学者郭克兰纽（Rudolphus Goclenius , 1547－1628）命名并由沃尔夫（Christian, Freiherr von Wolff, 1679－1754）加以完善并从理论上系统化，但就存在论这一学问而言，则是早已由古希腊哲学确定了其基本框架及理论内容的。事实上，存在论本身就是古希腊哲学的主题形态。

GO annotations

GO annotations are associations made between gene products and the GO terms that describe them



annotation



Up to October 26, 2010, there have been more than 2,753,338 annotations covering 48 species in GO database

Gene Ontology Tools

Consortium Tools

All tools, listed by category

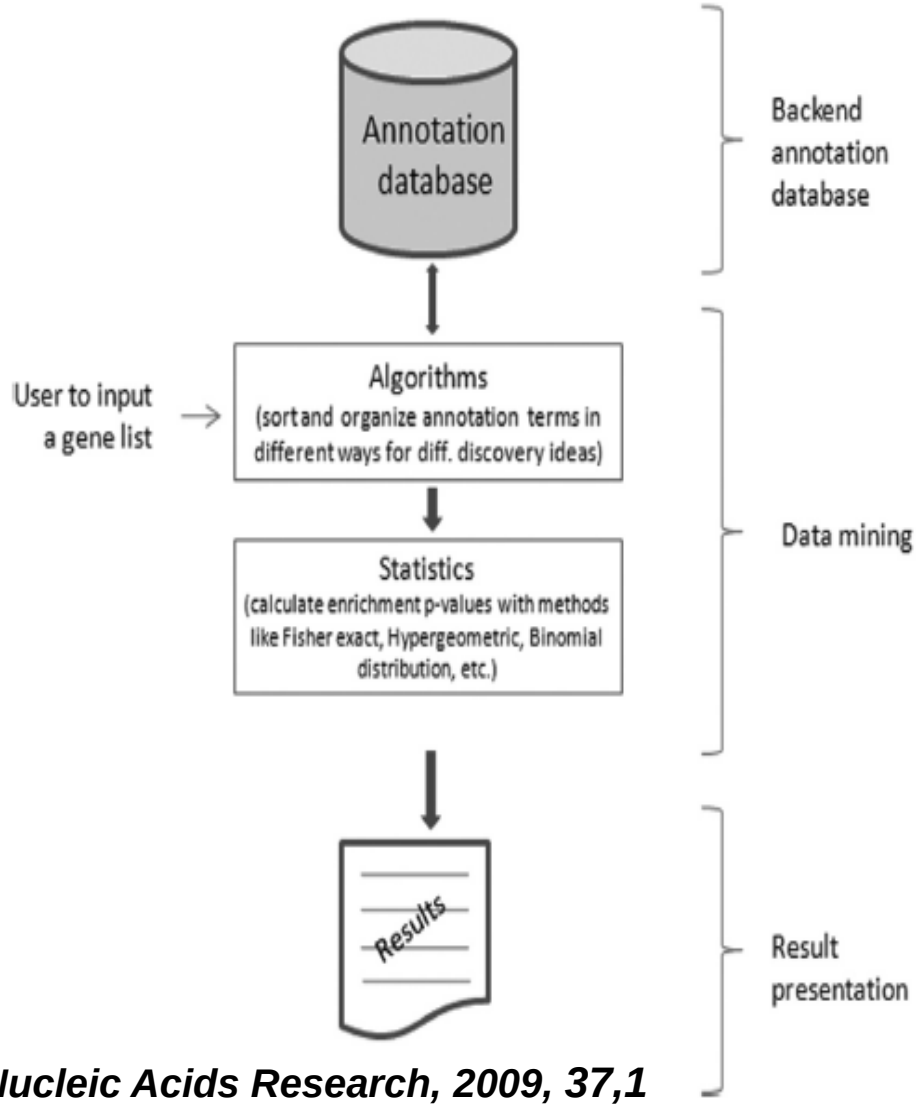
- Ontology or annotation browser
- Ontology or annotation search engine
- Ontology or annotation visualization
- Ontology or annotation editor
- Database or data warehouse
- Software library
- Statistical analysis
- Slimmer-type tool
- Term enrichment
- Text mining
- Protein interactions
- Functional similarity
- Semantic similarity
- Other analysis

All tools, alphabetical listing

Starred tools are those whose listings have been updated recently; unstarred tools may longer be active.

Agile Protein Interaction Data Analyzer: direct link to tool • entry in GO tools listings
agriGO*: direct link to tool • entry in GO tools listings
AmiGO*: direct link to tool • entry in GO tools listings
Avadis: direct link to tool • entry in GO tools listings
BiNGO*: direct link to tool • entry in GO tools listings
Bioconductor*: direct link to tool • entry in GO tools listings
Biomedical Logical Programming: direct link to tool • entry in GO tools listings
BioPerl: direct link to tool • entry in GO tools listings
Blast2GO: direct link to tool • entry in GO tools listings
CateGOrizer*: direct link to tool • entry in GO tools listings
CGAP GO browser: direct link to tool • entry in GO tools listings
ClueGO: direct link to tool • entry in GO tools listings
Cluster Assignment for Biological Inference: direct link to tool • entry in GO tools listings
Cluster Enrichment: direct link to tool • entry in GO tools listings
COBrA: direct link to tool • entry in GO tools listings
Comparative Toxicogenomics Database: direct link to tool • entry in GO tools listings
Database for Annotation, Visualization and Integrated Discovery*: direct link to tool • entry in GO tools listings
Db for Dummies!: direct link to tool • entry in GO tools listings
DynGO: direct link to tool • entry in GO tools listings
EASE: direct link to tool • entry in GO tools listings
ermineJ: direct link to tool • entry in GO tools listings
Exploratory Gene Association Networks: direct link to tool • entry in GO tools listings
Expression Profiler: direct link to tool • entry in GO tools listings
FatiGO: direct link to tool • entry in GO tools listings
Flash GViewer: direct link to tool • entry in GO tools listings
FuncAssociate: direct link to tool • entry in GO tools listings
FuncExpression: direct link to tool • entry in GO tools listings
FunCluster: direct link to tool • entry in GO tools listings
Functional Analysis of Transcriptional Networks: direct link to tool • entry in GO tools listings

Gene set enrichment analysis



Nucleic Acids Research, 2009, 37,1

Table 1. List of 68 enrichment tools

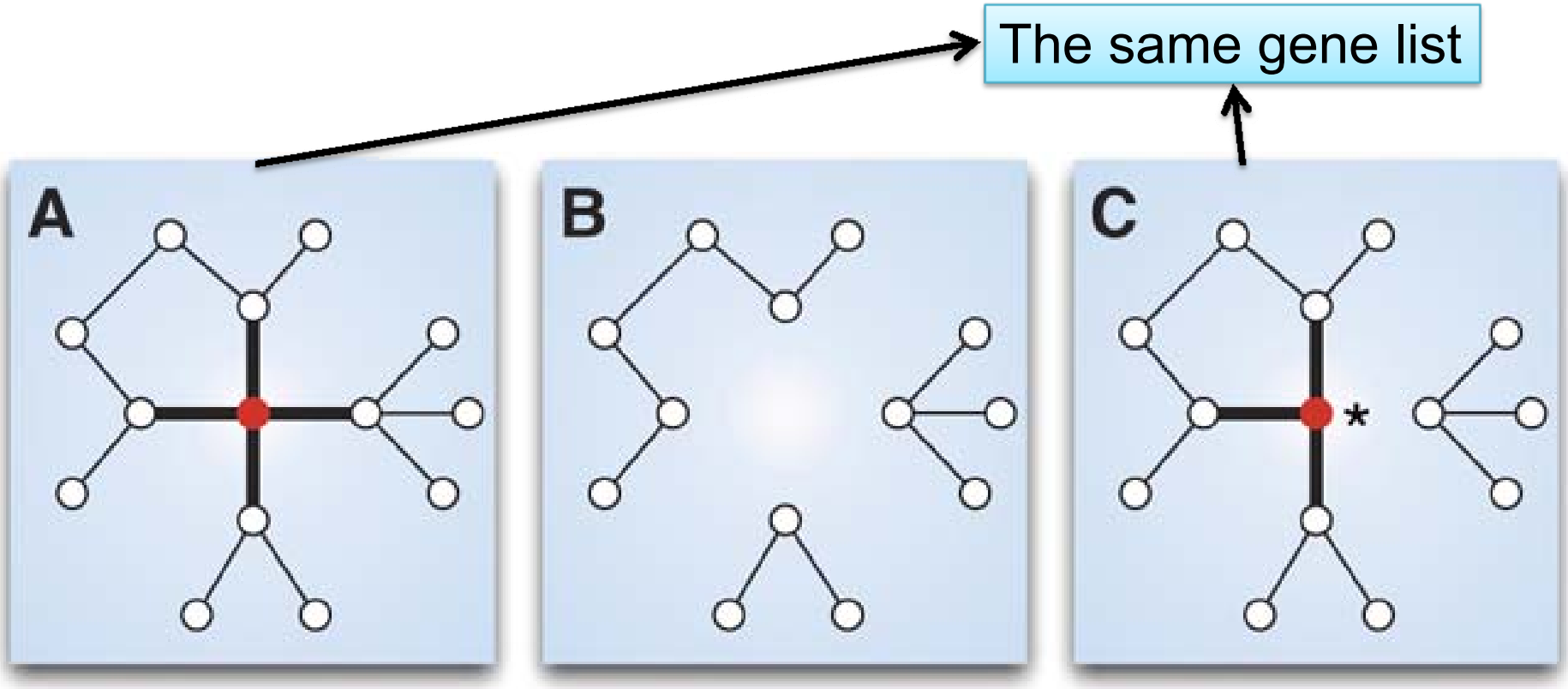
Enrichment tool name	Year of release	Key statistical method	Category
FunSpec	2002	Hypergeometric	Class I
Onto-express	2002	Fisher's exact; hypergeometic; binomial; chi-square	Class I
EASE	2003	Fisher's exact (modified as EASE score)	Class I
FatiGO (FatiWise/FatiGO+)	2003	Fisher's exact	Class I
FuncAssociate	2003	Fisher's exact	Class I
GARBAN	2003	Hypergeometric	Class I
GeneMerge	2003	Hypergeometric	Class I
GoMiner	2003	Fisher's exact	Class I
MAPPFinder	2003	*Z*-score; hypergeometric	Class I
CLENCH	2004	Hypergeometric; chi-square; binomial	Class I
GO::TermFinder	2004	hypergeometric	Class I
GOAL	2004	Permutation	Class I
GOArray	2004	Hypergeometric; *Z*-score; permutation	Class I
GOStat	2004	Fisher's exact; chi-squre	Class I
GoSurfer	2004	Chi-square	Class I
OntologyTraverser	2004	Hypergeometric; Fisher's exact	Class I
THEA	2004	Hypergeometric	Class I
BiNGO	2005	Hypergeometric; binomial	Class I
FACT	2005	Adopt GeneMerge and GO::TermFinder statistical modules	Class I
gfinder	2005	Fisher's exact	Class I
Gobar	2005	Hypergeometric	Class I
GOCluster	2005	Hypergeometric	Class I
GOSSIP	2005	Fisher's exact	Class I
L2L	2005	Binomial; hypergeometric	Class I
WebGestalt	2005	Hypergeometric	Class I
BayGO	2006	Bayesian; Goodman and Kruskal's gamma factor	Class I
eGOn/GeneTools	2006	Fisher's exact	Class I
Gene Class Expression	2006	*Z*-statistics	Class I
GOALIE	2006	Hidden Kripke model	Class I
GOFFA	2006	Fisher's inverse chi-square	Class I
GOLEM	2006	Hyerpgeometric	Class I
JProGO	2006	Fisher's exact; Kolmogorov–Smirnov test; student's *t*-test; Wilcoxon's test; hypergeometric	Class I
PageMan	2006	Fisher's exact; chi-square; Wilcoxon	Class I
STEM	2006	Hypergeometric	Class I
WEGO	2006	Chi-square	Class I
EasyGO	2007	Hypergeometric; chi-square; binomial	Class I
g:Profiler	2007	Hypergeometric	Class I
ProbCD	2007	Yule's Q; Goodman-Kruskal's gamma; Cramer's T	Class I
GOEAST	2008	Hypergeometric	Class I
GOHyperGAll	2008	Hypergeometric	Class I
CatMap	2004	Permutations	Class II
Godist	2004	Kolmogorov–Smirnov test	Class II
GO-Mapper	2004	Gaussian distribution; EQ-score	Class II
iGA	2004	Permutations; hypergeometric; *t*-test; *Z*-score	Class II
GSEA	2005	Kolmogorov–Smirnov-like statistic	Class II
MEGO	2005	*Z*-score	Class II
PAGE	2005	*Z*-score	Class II
T-profiler	2005	*t*-Test	Class II
FuncCluster	2006	Fisher's exact	Class II
FatiScan	2007	Fisher's Exact	Class II
FINA	2007	Fisher's exact	Class II
GAzer	2007	*Z*-statistics; permutation	Class II
GeneTrail	2007	Hypergeometric; Kolmogorov–Smirnov	Class II
MetaGP	2007	*Z*-score	Class II
Ontologizer	2004	Fisher's exact	Class III
POSOC	2004	POSET (a discrete math: finite partially ordered set)	Class III
topGO	2006	Fisher's exact	Class III
GO-2D	2007	Hypergeometric; binomial	Class III
GENECODIS	2007	Hypergeometric; chi-square	Class III
GOSim	2007	Resnik's similarity	Class III
PaIS	2008	Percent	Class III
ProfCom	2008	Greedy heuristics	Class III
GOTM	2004	Hypergeometric	Class I,II
ermineJ	2005	Permutations; Wilcoxon rank-sum test	Class I,II
DAVID	2003	Fisher's Exact (modified as EASE score)	Class I,III
GOToolBox	2004	Hypergeometric; Fisher's exact; Binomial	Class I,III
ADGO	2006	*Z*-statistic	Class II,III
FunNet	2008	Unclear	Unclear

Types of Enrichment analysis

Table 2. Categorization of enrichment analysis tools

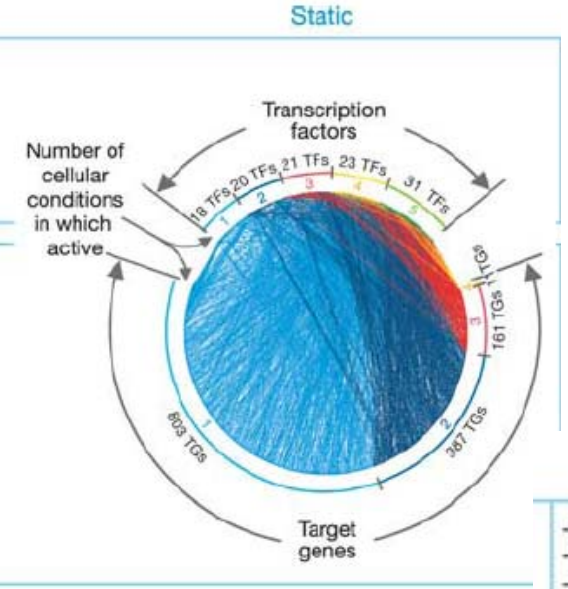
Tool category	Description	Indication and limitation	Sub-type of algorithms	Methods	Example tool
Class I: singular enrichment analysis (SEA)	Enrichment *P*-value is calculated on each term from the pre-selected interesting gene list. Then, enriched terms are listed in a simple linear text format. This strategy is the most traditional algorithm. It is still dominantly used by most of the enrichment analysis tools.	Capable of analyzing any gene list, which could be selected from any high-throughput biological studies/technologies (e.g. Microarray, ChIP-on-CHIP, ChIP-on-sequence, SNP array, EXON array, large scale sequence, etc.). However, the deeper inter-relationships among the terms may not be fully captured in linear format report.	Global reference background	Fisher's exact hypergeometric chi-square binomial	GoStat, GoMiner, GOTM, BinGO, GOtoolBox, GFinder, etc.
			Local reference background	Fisher's Exact hypergeometric chi-square binomial	DAVID, Onto-Express, GARBAN, FatiGO, etc.
			Neural network	Bayesian	BayGO
Class II: gene set enrichment analysis (GSEA)	Entire genes (without pre-selection) and associated experimental values are considered in the enrichment analysis. The unique features of this strategy are: (i) No need to pre-select interesting genes, as opposed to Classes I and II; (ii) Experimental values integrated into *P*-value calculation.	Suitable for pair-wide biological studies (e.g. disease versus control). Currently, may be difficult to be applied to the diverse data structures derived by a complex experimental design and some of the new technologies (e.g. SNP, EXON, Promoter arrays).	Based on ranked gene list	Kolmogorov–Smirnov-like	GSEA, CapMap, etc.
			Based on continuous gene values	*t*-Test permutation *Z*-score	FatiScan, ADGO, ermineJ, PAGE, iGA, GO-Mapper, GOdist, FINA, T-profiler, MetaGP, etc.
Class III: modular enrichment analysis (MEA)	This strategy inherits key spirit of SEA. However, the term–term/gene–gene relationships are considered into enrichment *P*-value calculation. The advantage of this strategy is that term–term/gene–gene relationship might contain unique biological meaning that is not held by a single term or gene. Such network/modular analysis is closer to the nature of biological data structure.	Capable of analyzing any gene lists, which could be selected from any high-throughput biological studies/technologies, like Class I. Emphasis on network relationships during analysis. 'Orphan' gene/term (with little relationships to other genes/terms), that sometimes could be very interesting, too, may be left out from the analysis.	Composite annotations	Measure enrichment on joint terms	ADGO, GeneCodis, ProfCom, etc.
			DAG Structure	Measure enrichment by considering parents-child relationships	topGO, Ontologizer, POSOC, etc.
			Global annotation relationship	Measure term–term global similarity with Kappa Statistics Czekanowski-Dice Pearson's correlation	DAVID, GoToolBox, etc.

Only gene list is not enough!

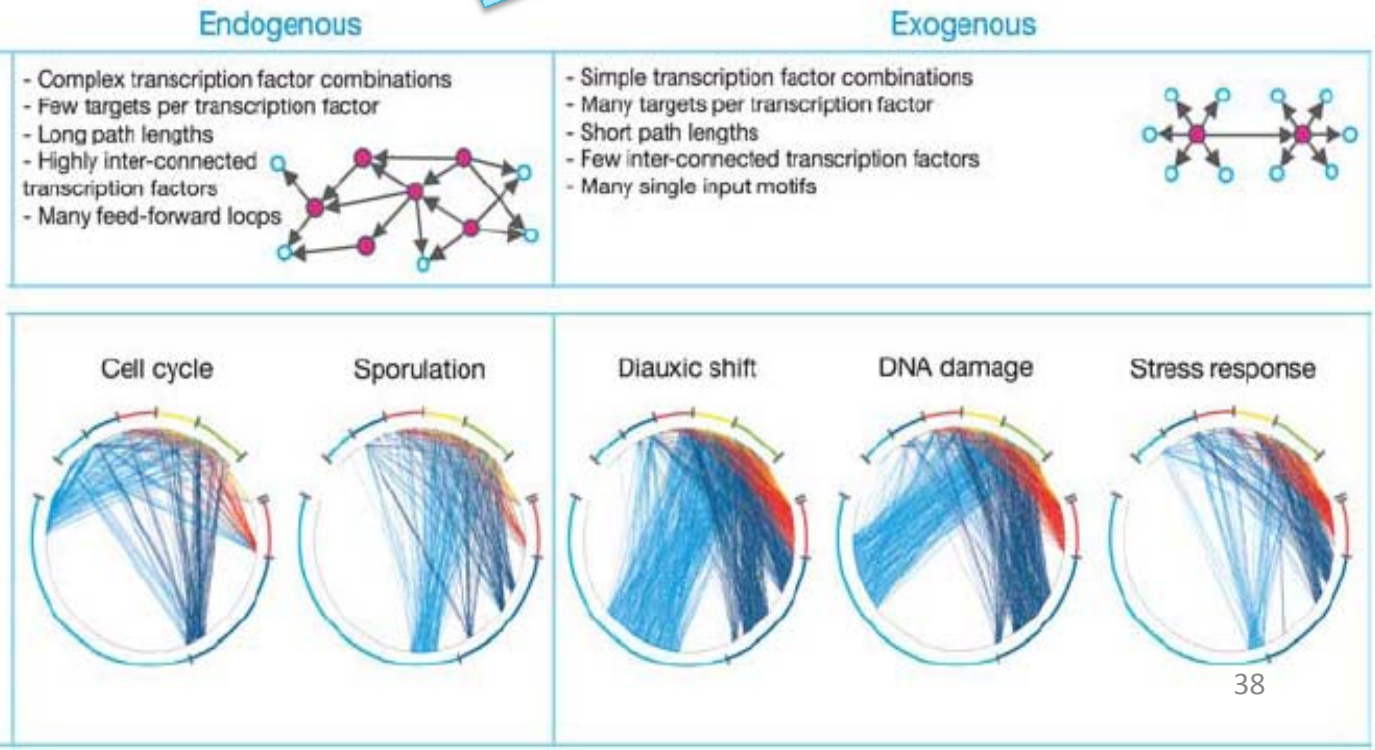


Molecular systems biology, 2009

Only gene list is not enough!



Regulatory network dynamics



Nature, 2004

Edge ontology

In fact, "edge ontology" or "arrow ontology" has been suggested by a forward-looking work.

Inspired by the gene ontology, Lu et al. aim to build a similar hierarchical term structure for edges.

However, edge ontology is still far from complete to describe the functional relationship in the network. In contrast, gene ontology has contained 32,862 terms and 2,753,338 annotations up to now.

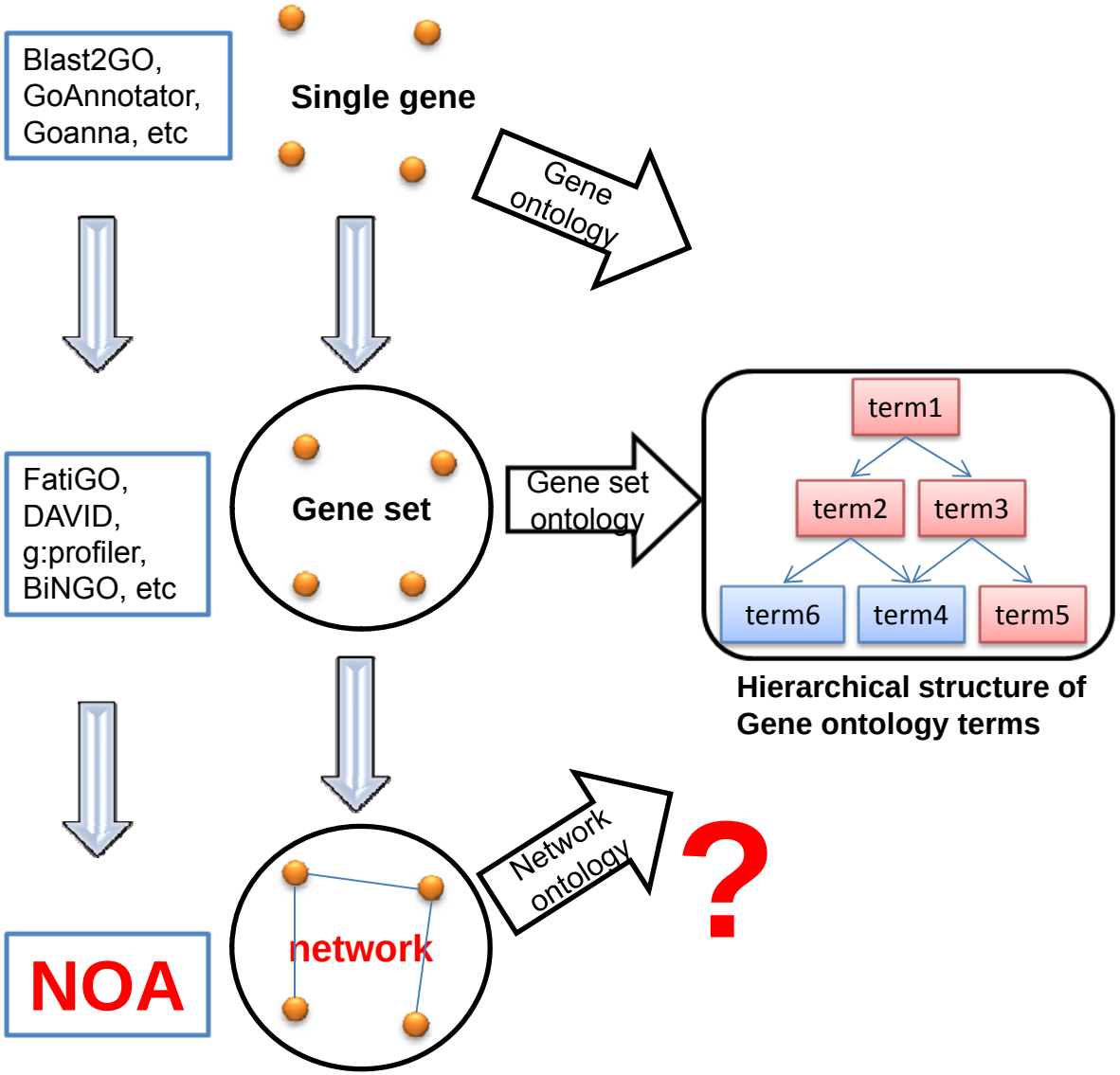
Table 1. A prototype of an edge ontology

	Direction (level I)	Type (level II)	Sub-type (level III)	Specification (level IV)
Interaction	Directed	Tagging of proteins[a]	Phosphorylation	Serine[d]
				Tyrosine[d]
				Other[d]
			Dephosphorylation	
			Ubiquitylation	
			Glycosylation	N-linked
				O-linked
			Methylation	
		Cleavage of proteins[a]		
		Translocation	Diffusion	
			Active transport	
		Conformational change		
		Chemical reaction[b]		
		Catalysis		
		Unknown/other		
	Undirected	Binding[c]		
		Complex association[c]		
		Binding or association[c]		
		Dissociation		
		Co-expression		
		Unknown/other		

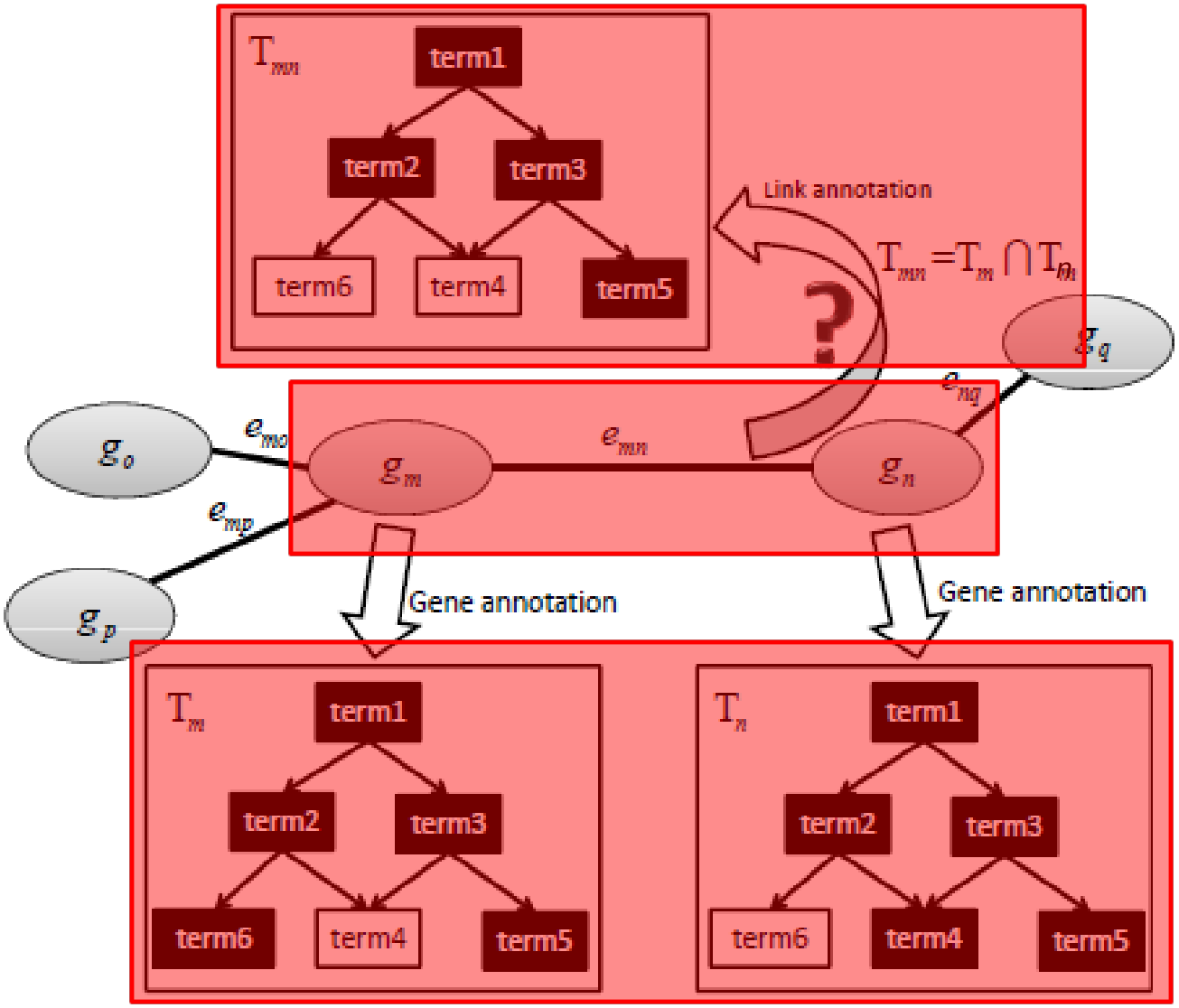
[a] describes 'inhibition' (where applicable).
[b] Double arrow describes reversible chemical reactions.
[c] 'Binding' describes direct physical interaction. 'Association' describes two proteins that are linked in the same complex but do not directly physically interact. 'Binding or association' describes the common scenario that arises in tandem affinity purification (TAP)-tagging experiments when the specific type of interaction is not known.
[d] Ser describes serine inhibition. A similarly annotated symbol can be used for tyrosine inhibition and so on.

Trends in Biochemical Sciences, 2007

Network-based gene ontology analysis

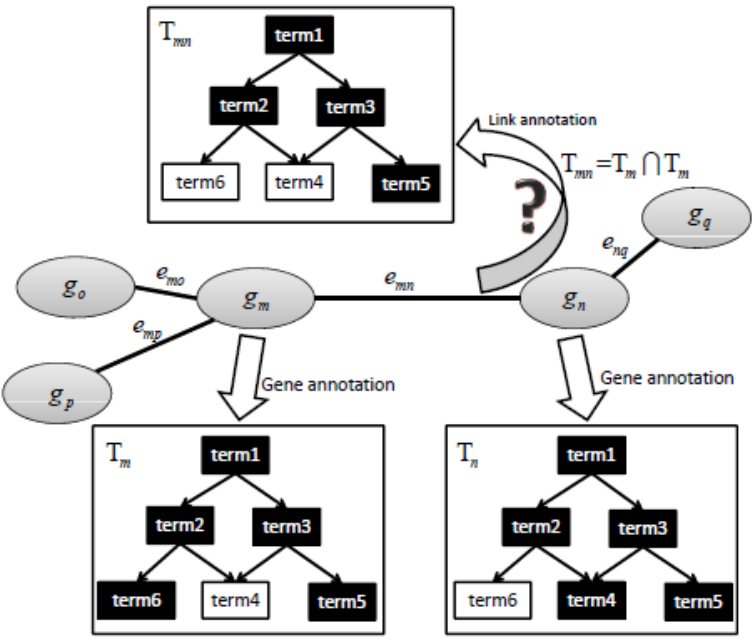


Link ontology



How to define the function of links based on gene annotation?

Diversity and Coverage



$$\mathcal{D}(\mathcal{T}(E)) = \sum_{e_{mn} \in E} D(e_{mn}) = \sum_{e_{mn} \in E} \sum_{t \in T_{mn}} \frac{2 - S(t, T_m) - S(t, T_n)}{2|T_{mn}|}$$

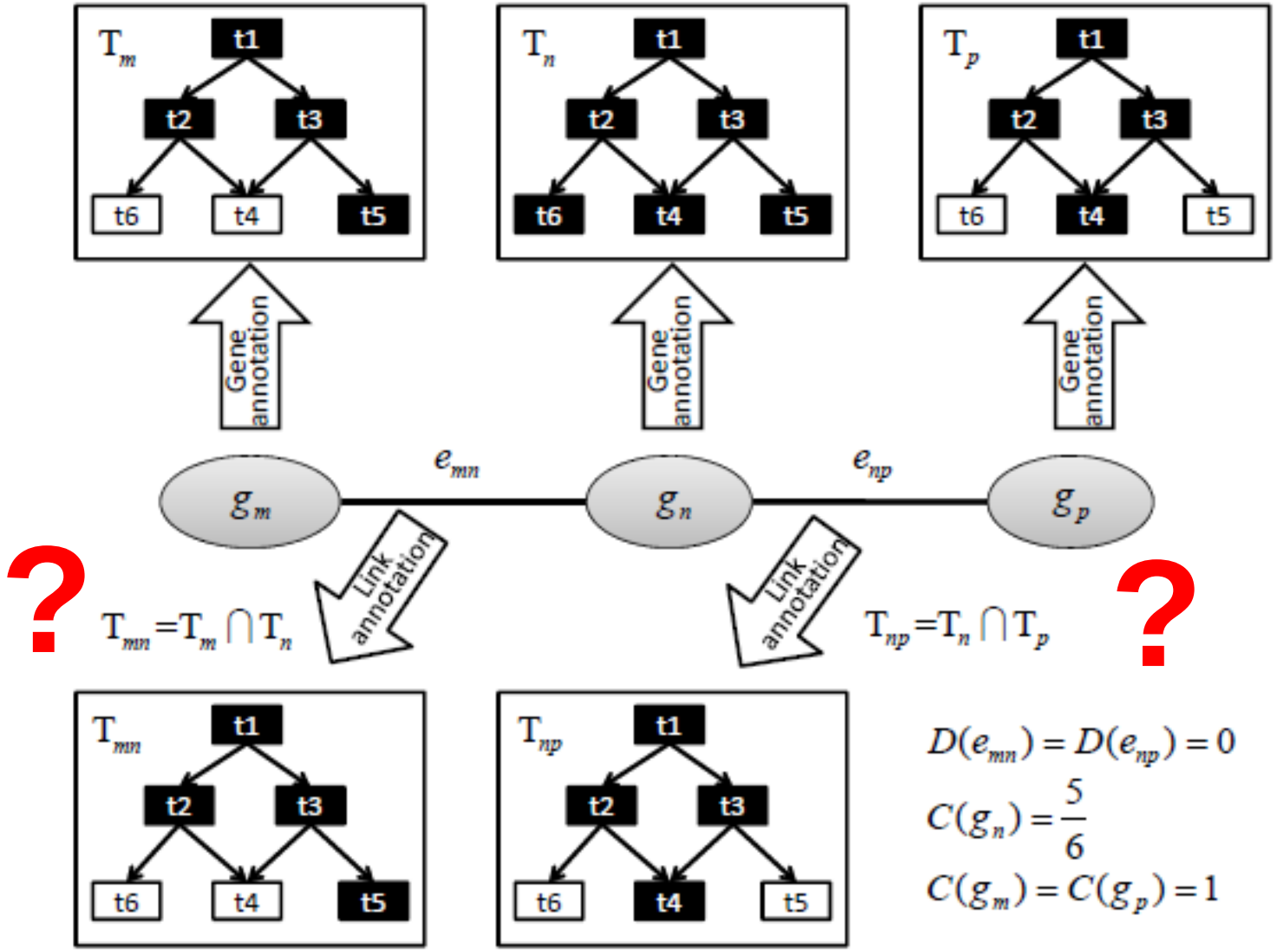
Diversity is the average $D(e_{mn})$, which represents the functional consistency of edge e_{mn} with both nodes g_m and g_n,

$$\mathcal{C}(\mathcal{T}(E)) = \sum_{g_m \in G} C(g_m) = \sum_{g_m \in V} \sum_{t \in T_m} \frac{S(t, \bigcup_{n:e_{mn} \in E} T_{mn})}{|T_m|}$$

Coverage is the average $C(g_m)$, which implies the coverage ratio of all functions on node g_m, covered by the functions of all edges connecting to g_m.

For a Network or Edge set,
A good assignment should have small ***diversity*** and large ***coverage***

An example

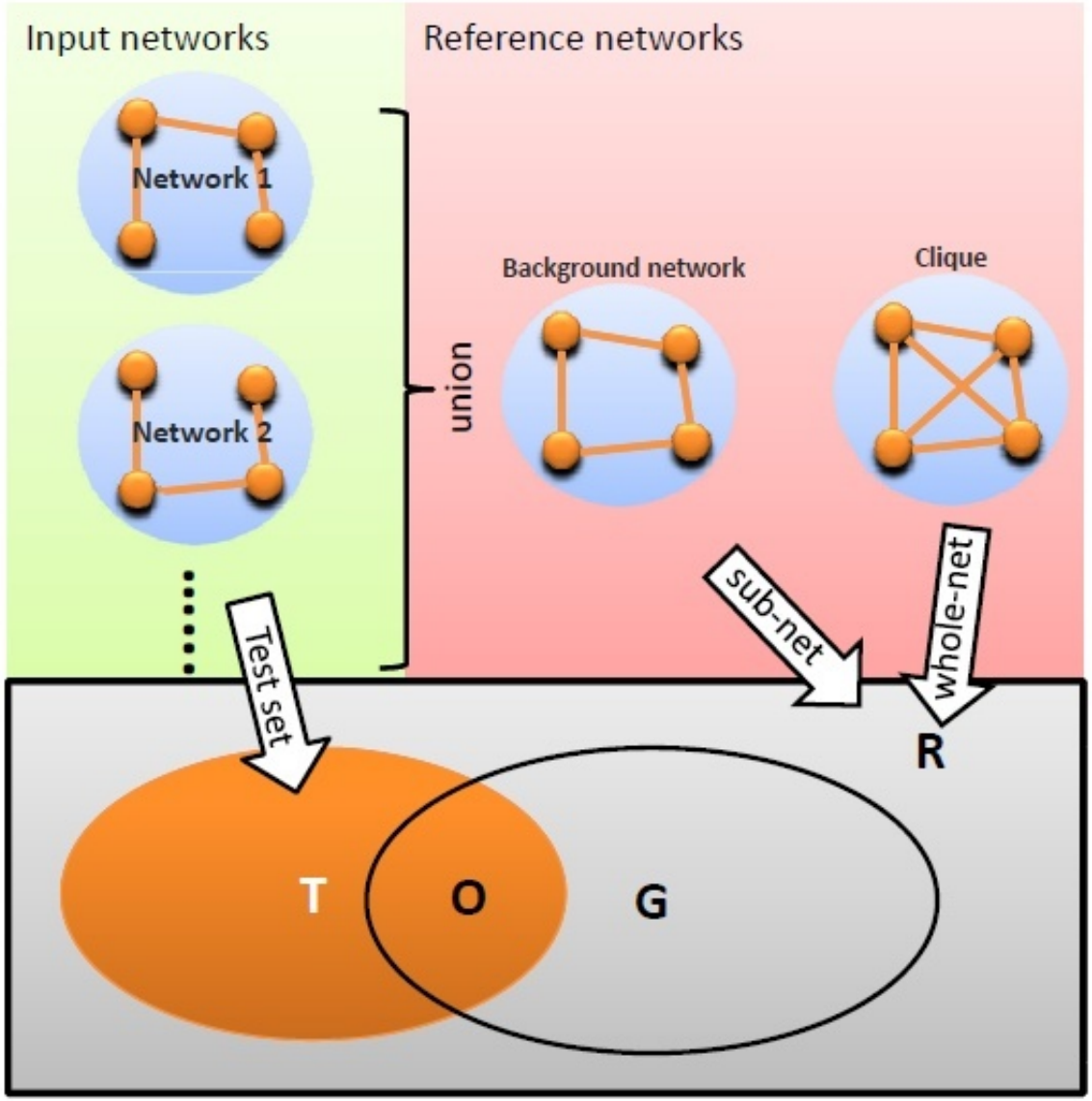


$$D(e_{mn}) = D(e_{np}) = 0$$

$$C(g_n) = \frac{5}{6}$$

$$C(g_m) = C(g_p) = 1$$

Network ontology



Based on the definition of link ontology, next we can further define **network ontology** via regarding the network as a set of links.

$$P(X \geq O) = \sum_{k=O}^{\min\{G,T\}} \frac{\binom{G}{k}\binom{R-G}{T-k}}{\binom{R}{T}}.$$

Network ontology

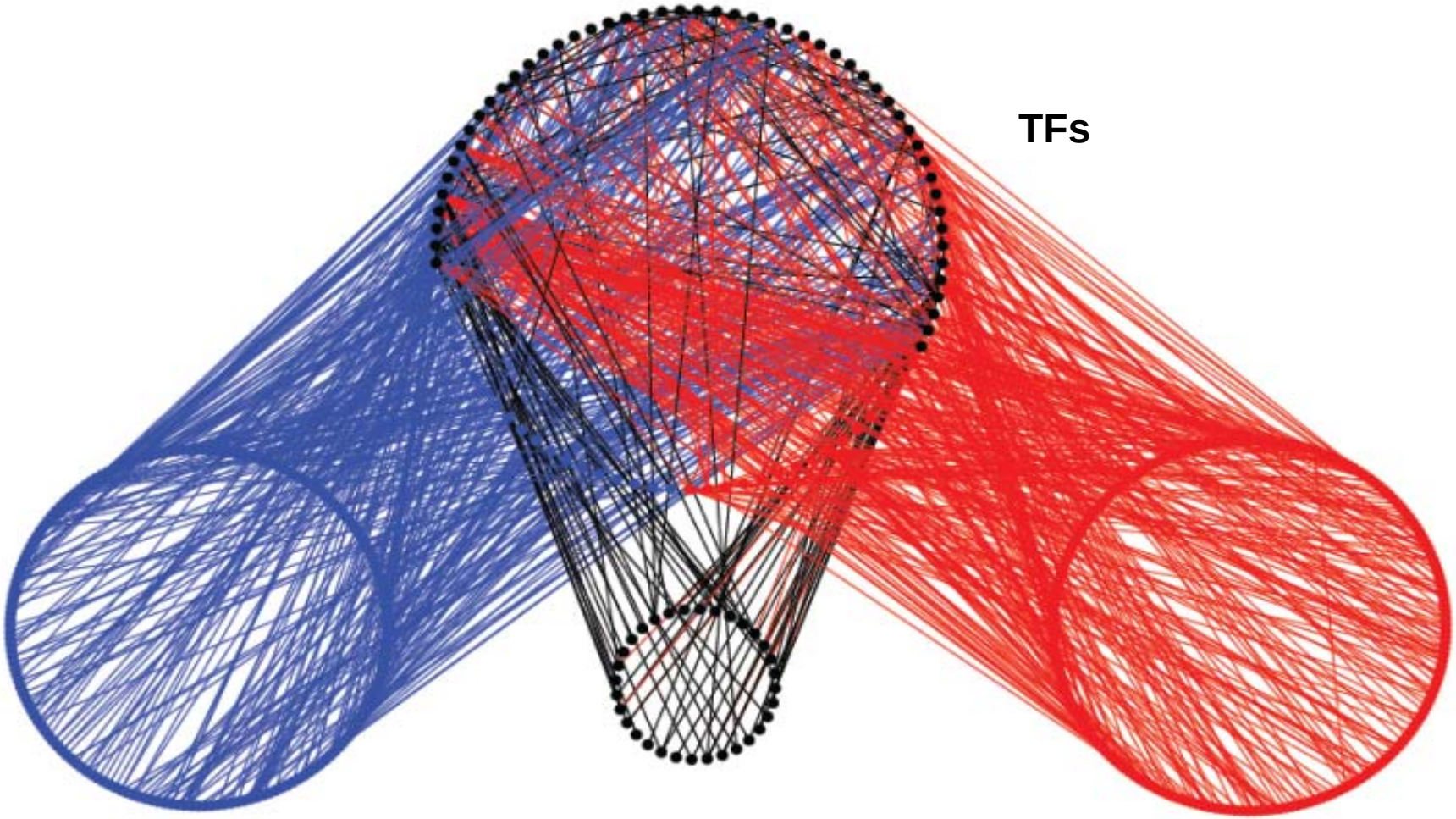
Table 1. Test set and reference set of the four types of GO analysis methods: whole-net NOA, sub-net NOA, whole-net gene list method, and sub-net gene list method.

		whole-net	sub-net
NOA	Test set	Link list	Link list
	Reference set	Clique	Background network
GLM	Test set	Gene list	Gene list
	Reference set	Yeast gene	Gene in background network

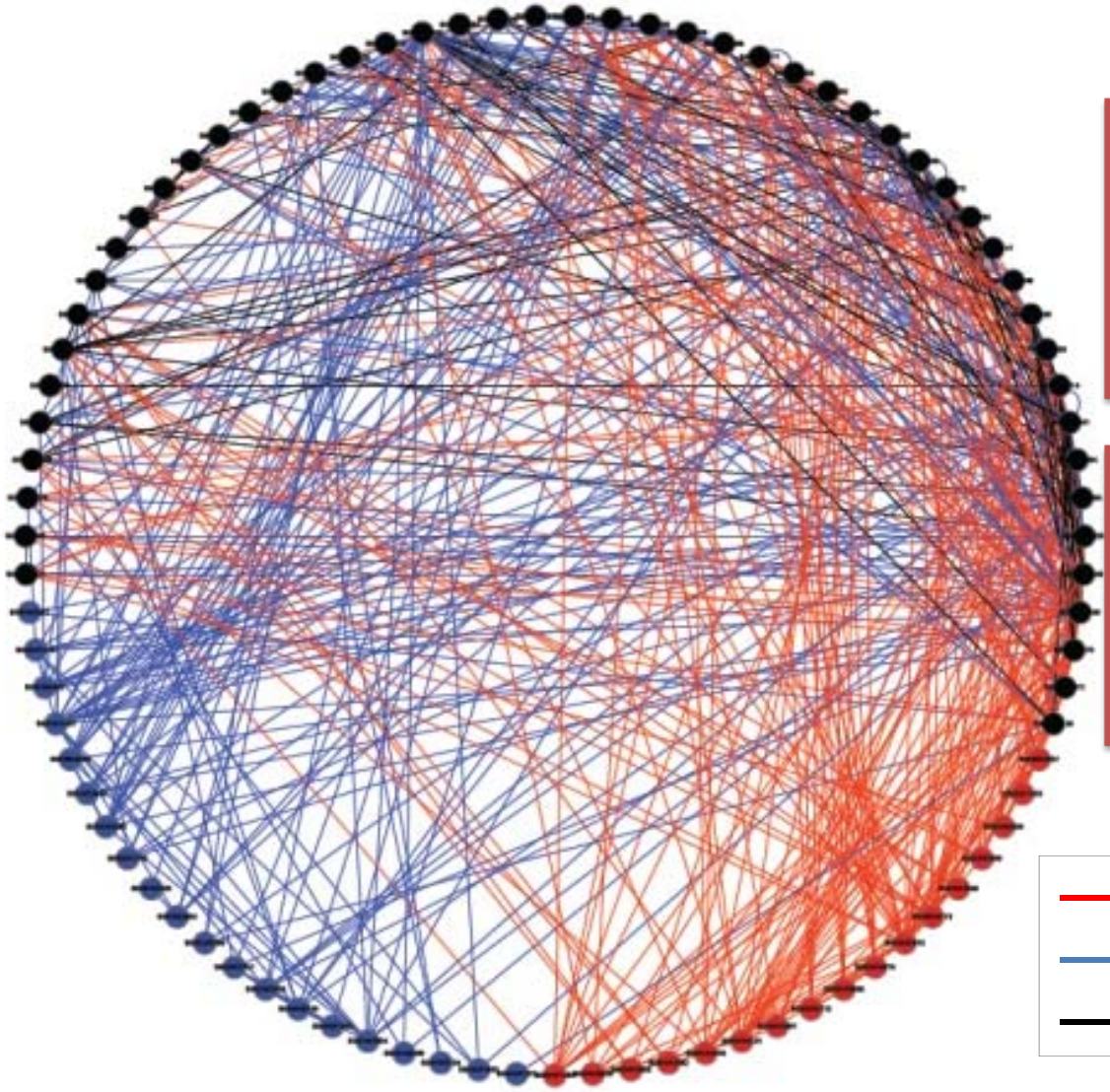
where GLM means gene list based method.

Result I: NOA works well in dynamic networks

Network rewiring of yeast transcription regulatory



Network rewiring of Yeast transcription factor co-regulatory



We construct TF coregulatory networks via adding an edge between two TFs if they have at least one common target gene

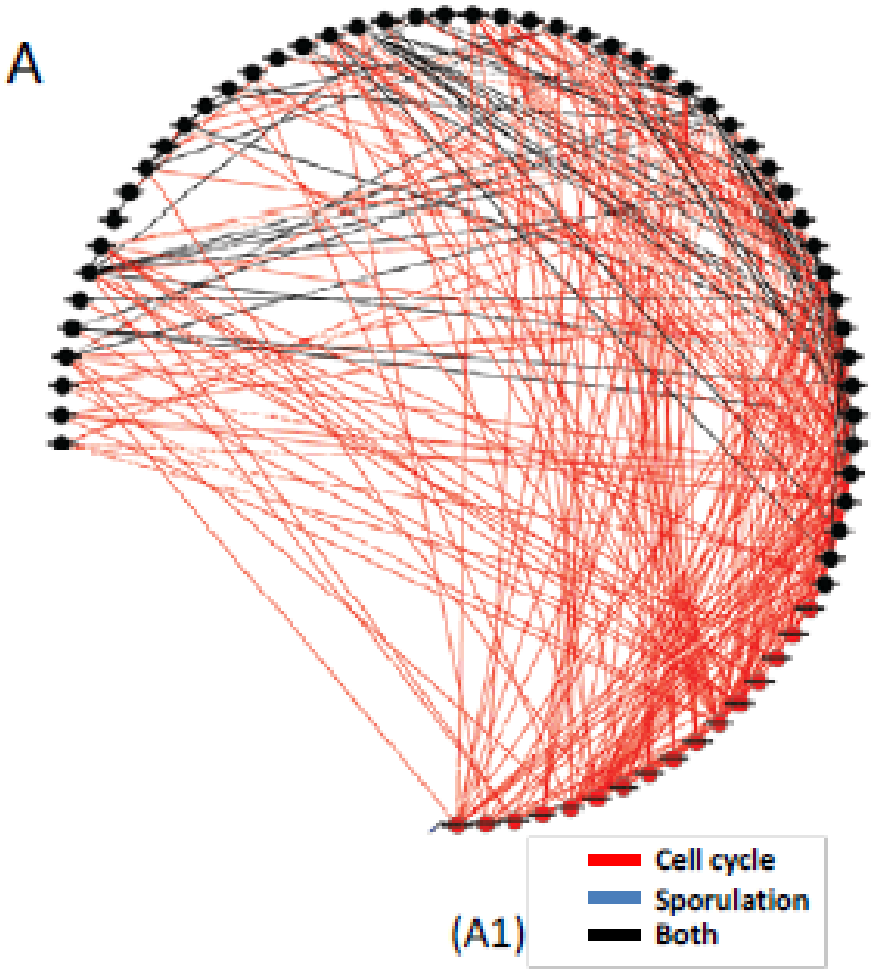
Most of nodes are the same in cell cycle and sporulation networks, but the links are significantly different (Network rewiring)

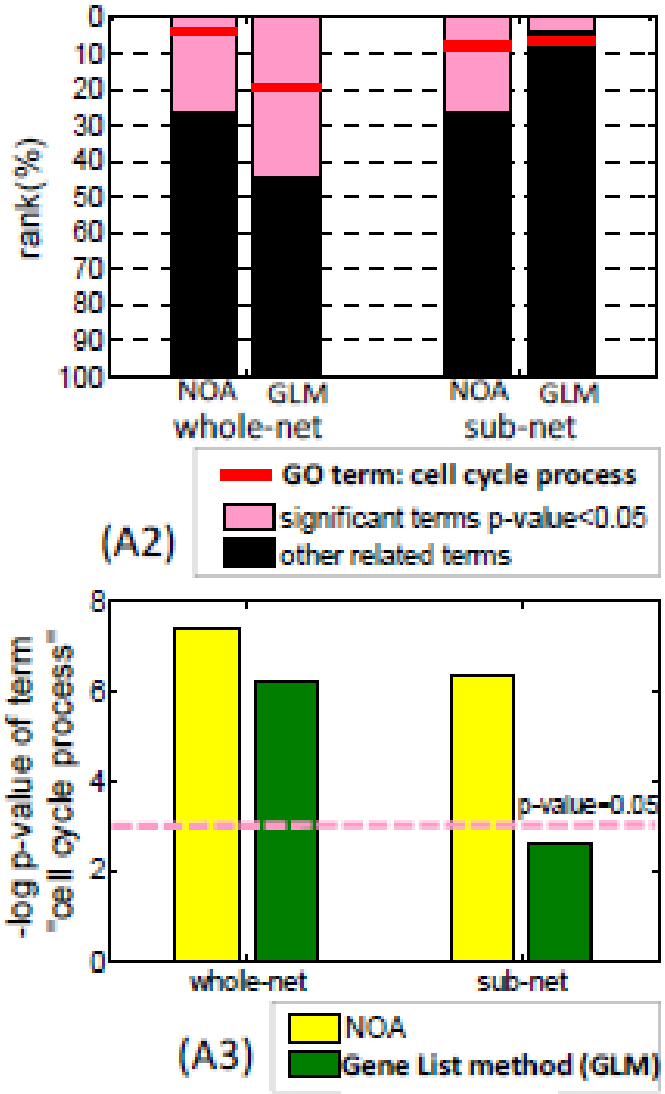
Cell cycle
Sporulation
Both

Results

	Whole-net cell cycle edge	Whole-net cell cycle node	sub-net cell cycle edge	sub-net cell cycle node	Whole-net Sporulation edge	Whole-net Sporulation node	sub-net Sporulation edge	sub-net Sporulation node
Rank	12	101	20	20	33	335	54	163
# significant terms	56	217	55	18	79	235	85	15
# terms	209	485	209	485	234	536	234	536
P-value	0.0006	0.0020	0.0017	0.0720	0.0029	0.1754	0.0107	0.3403

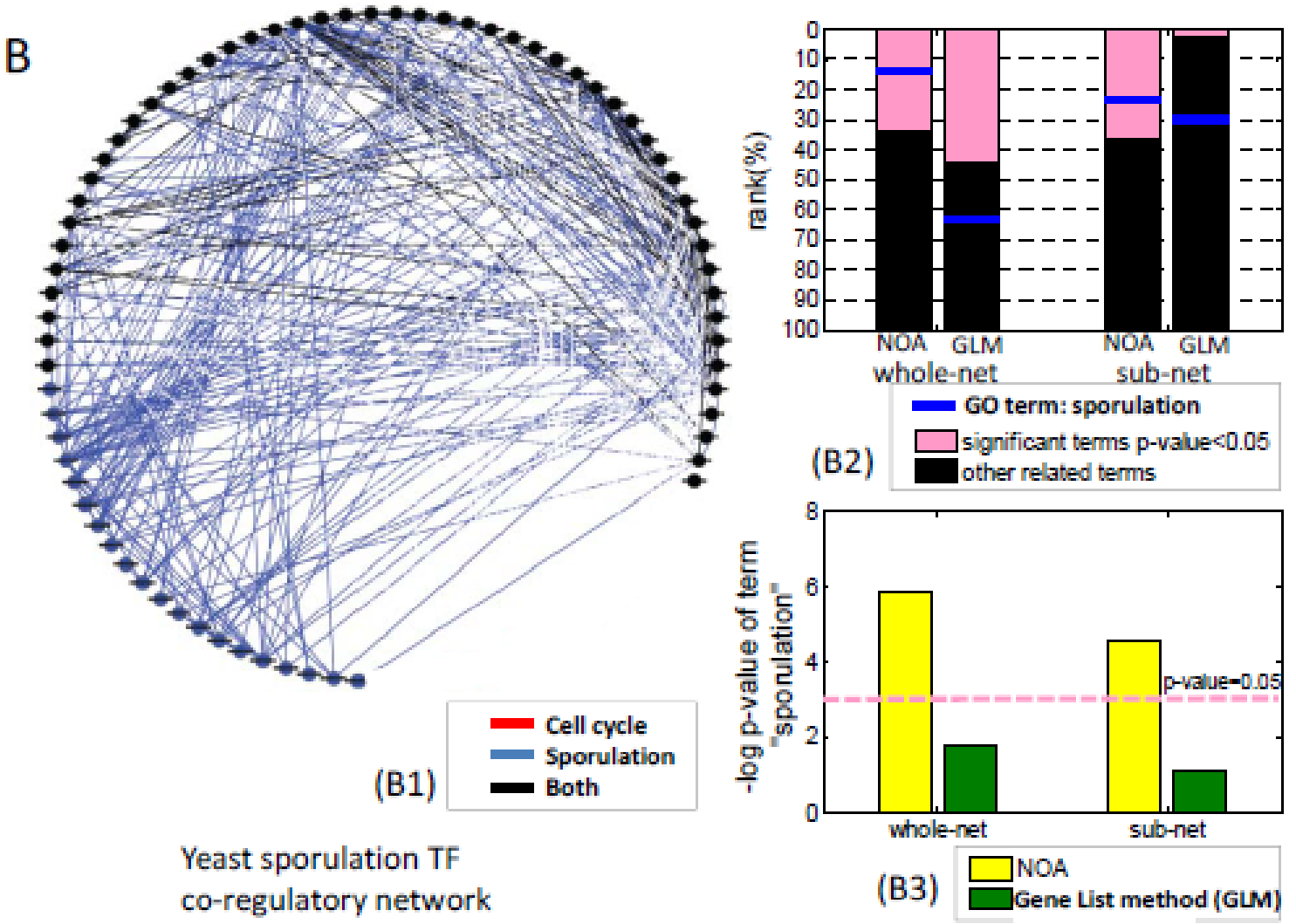
Comparing methods: NOA vs GLM





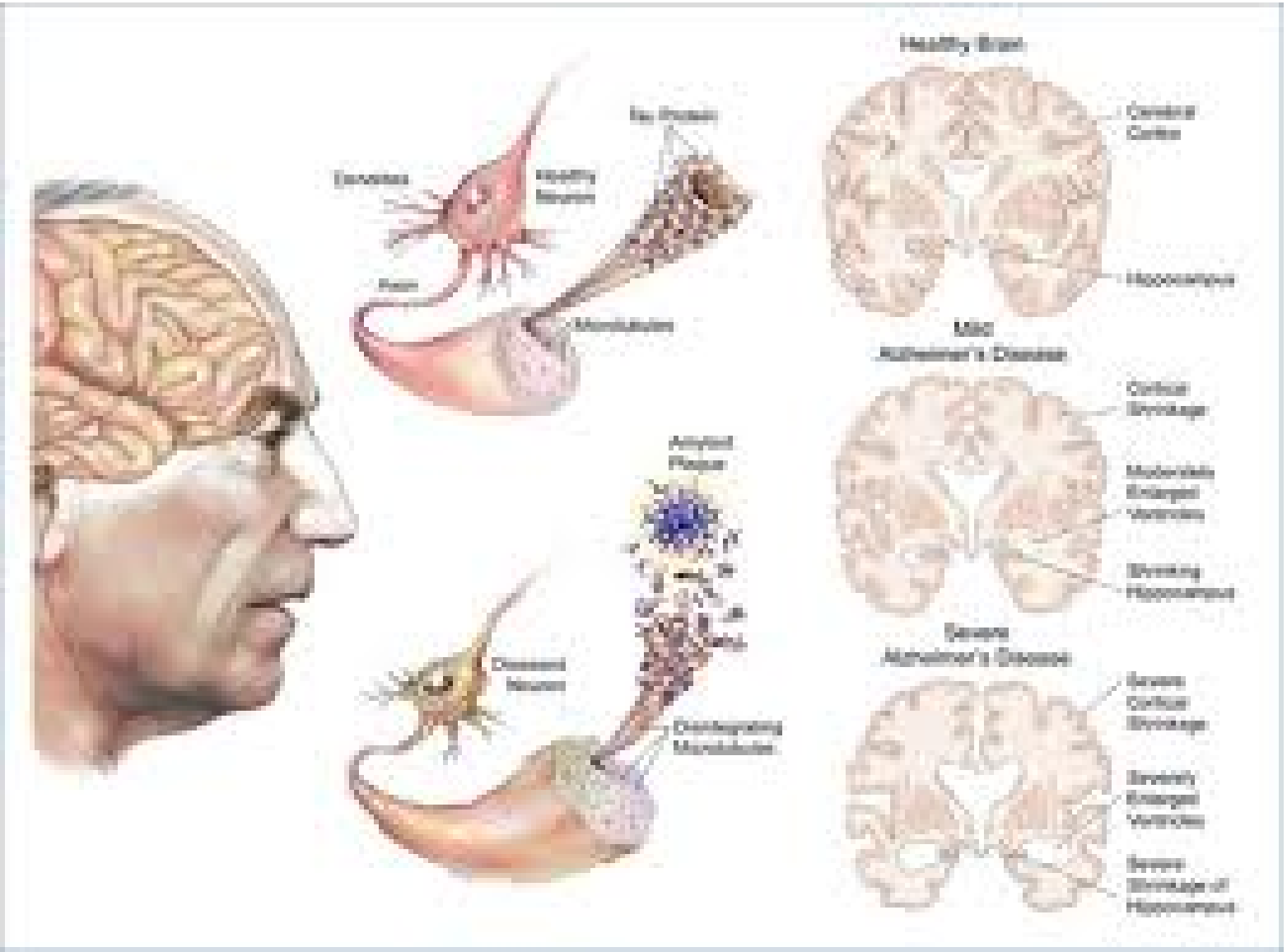
Yeast cell cycle TF co-regulatory network

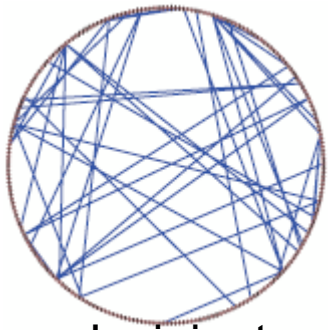
Comparing methods: NOA vs GLM



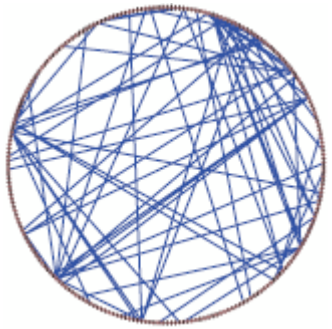
Yeast sporulation TF co-regulatory network

Different stages in Alzheimer's disease





Incipient



Moderate

Severe

Results by two types of methods

Type	GO term	Description (frequency in pathway/network)	P-value
incipient	GO:0006355	regulation of transcription, DNA-dependent (13/28)	2.67e-09
	GO:0045944	positive regulation of transcription from RNA polymerase II promoter (9/14)	1.01e-07
	GO:0007242	intracellular signaling cascade (6/8)	1.93e-05
moderate	GO:0006916	anti-apoptosis (11/22)	3.12e-06
	GO:0007165	signal transduction (16/51)	6.17e-06
	GO:0006355	regulation of transcription, DNA-dependent (11/28)	4.15e-05
severe	GO:0006355	regulation of transcription, DNA-dependent (16/28)	1.98e-09
	GO:0006629	lipid metabolic process (10/22)	7.59e-05
	GO:0045944	positive regulation of transcription from RNA polymerase II promoter (10/14)	5.82e-07

GLM

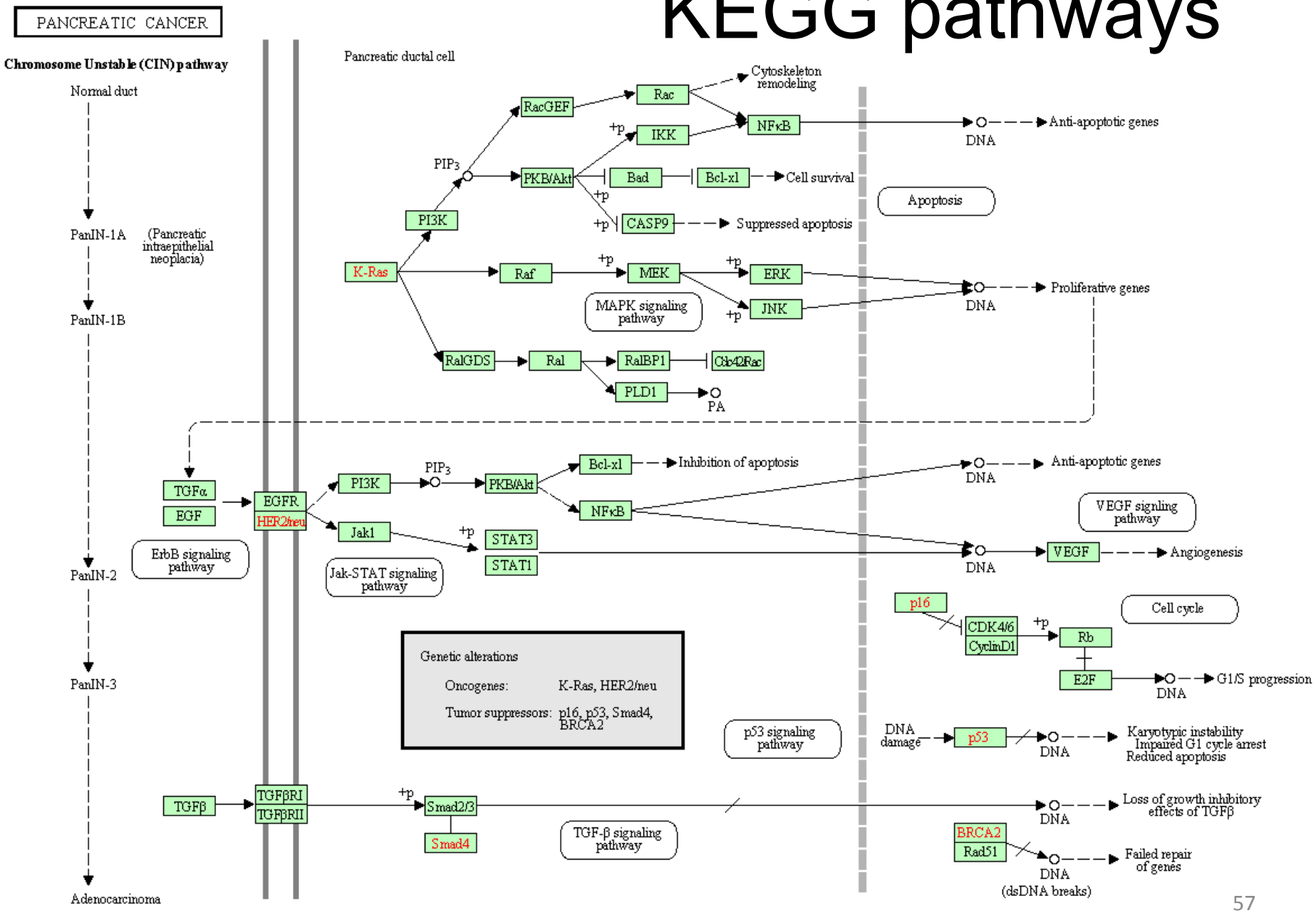
NOA

Network type	GO term (BP)	Description
Incipient	GO:0016192	Vesicle-mediated transport
	GO:0042325	Regulation of phosphorylation
	GO:0005979	Regulation of glycogen biosynthetic process
Moderate	GO:0043549	Regulation of kinase activity
	GO:0048589	Developmental growth
	GO:0006897	Endocytosis
Severe	GO:0015918	Sterol transport
	GO:0006915	Apoptosis
	GO:0006509	Membrane protein ectodomain proteolysis

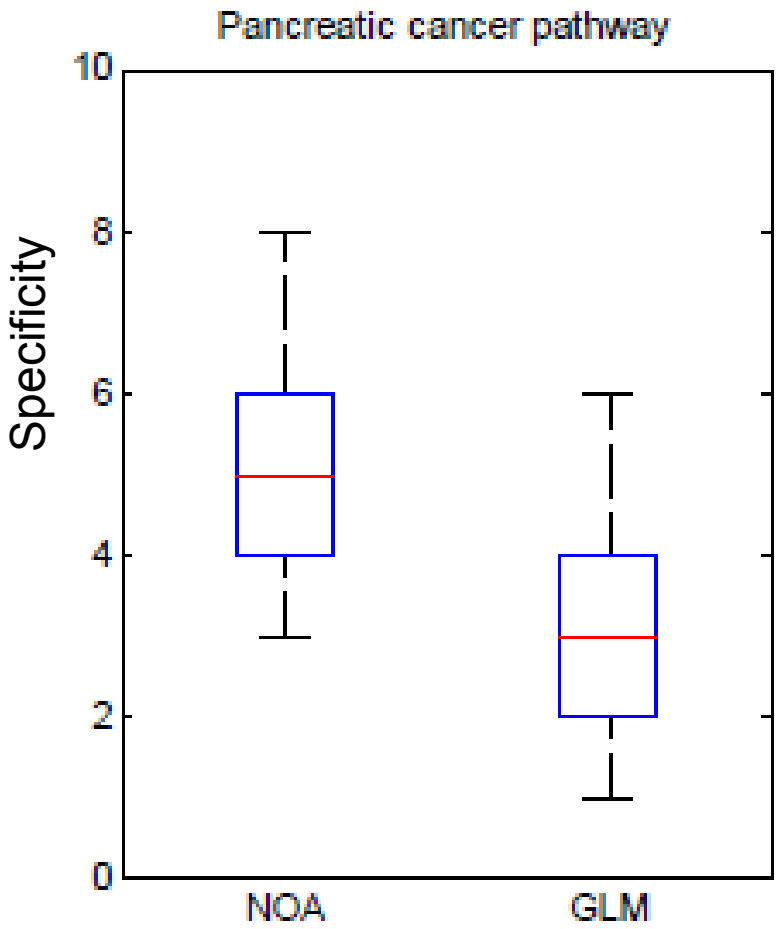
where AD means Alzheimer's disease and BP means biological process

Result II: NOA can identify more specific and meaningful functional terms in static network

KEGG pathways



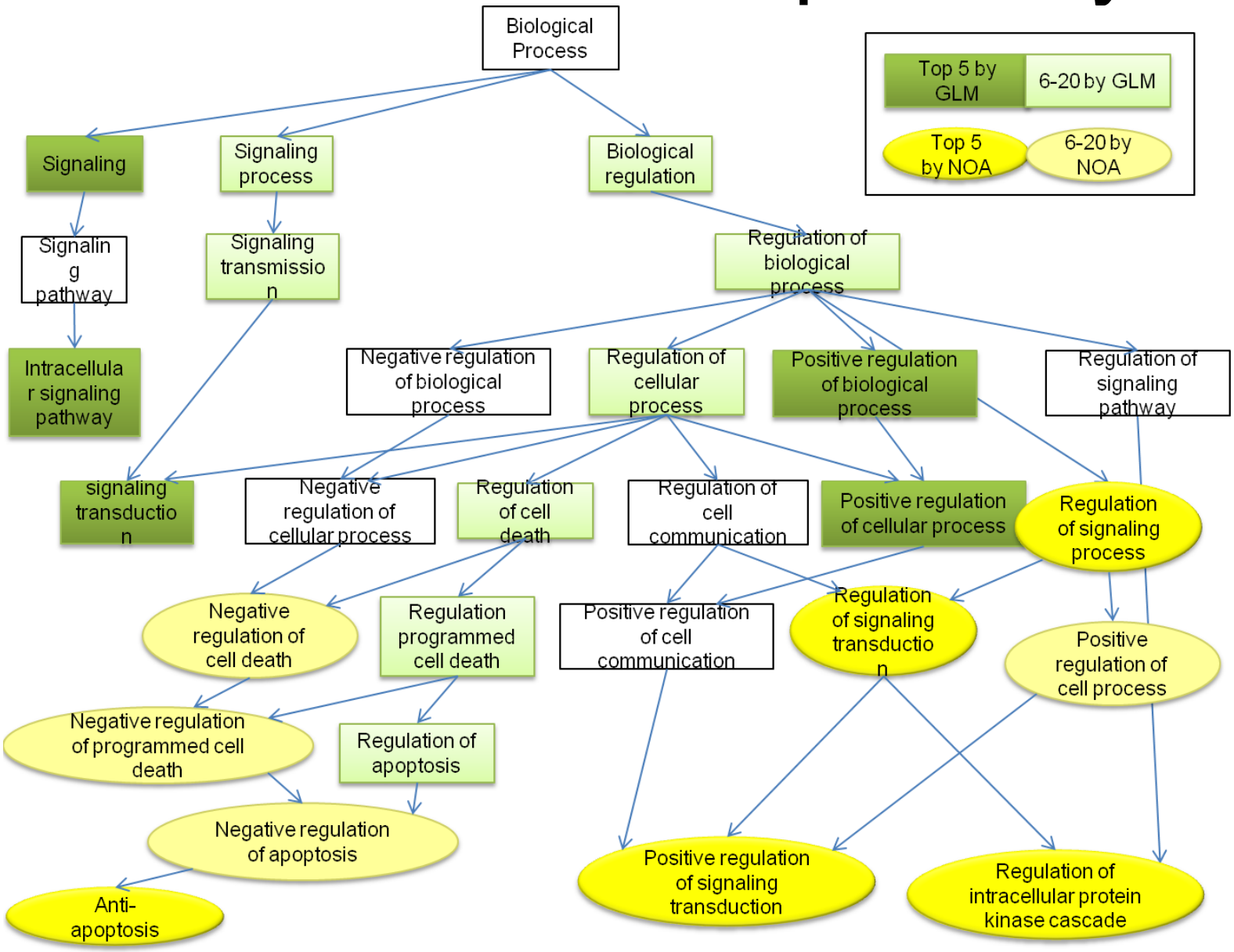
Specificity of different methods



Level of terms reported by NOA and Gene List Methods (GLM) in GO structure.

Roughly, terms reported by NOA have deeper level than GLM.

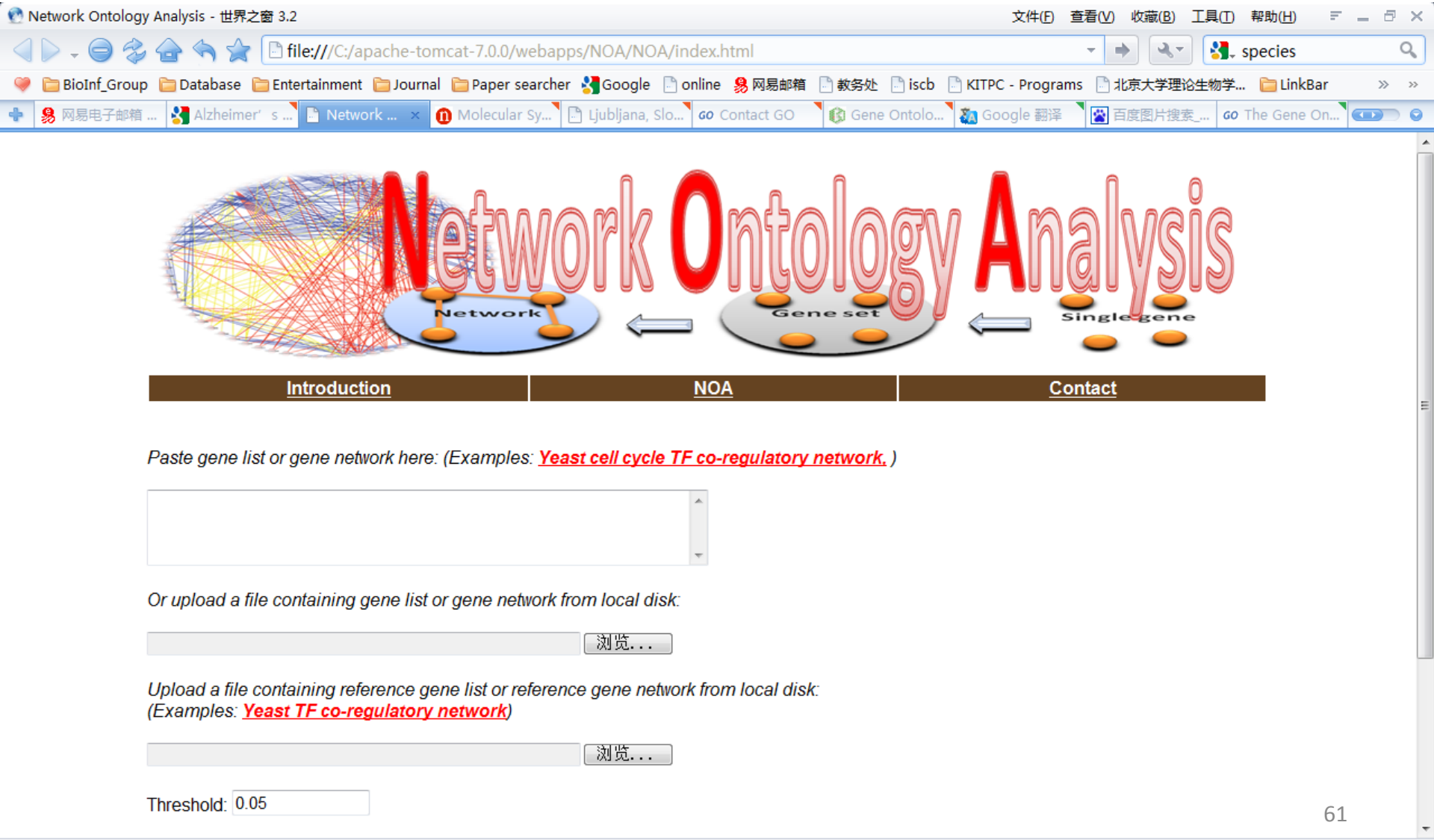
Results in KEGG pathway



Webserver

- Input:
  - Species
  - Edges of a network
    - Upload a file
    - Directly paste in blank field
  - Cutoff
- Output:
  - Enriched GO terms by NOA and GLM
  - Corresponding p-values
  - Corresponding edges (or nodes)

Index page


Network Ontology Analysis - 世界之窗 3.2
文件(F) 查看(V) 收藏(B) 工具(T) 帮助(H)
file:///C:/apache-tomcat-7.0.0/webapps/NOA/NOA/index.html
species
BioInf_Group Database Entertainment Journal Paper searcher Google online 网易邮箱 教务处 iscb KITPC - Programs 北京大学理论生物学... LinkBar
Network Ontology Analysis
Network
Gene set
Single gene
Introduction
NOA
Contact
Paste gene list or gene network here: (Examples: Yeast cell cycle TF co-regulatory network.)
Or upload a file containing gene list or gene network from local disk:
浏览...
Upload a file containing reference gene list or reference gene network from local disk:
(Examples: Yeast TF co-regulatory network)
浏览...
Threshold: 0.05

Reports

Introduction | NOA | Contact

Parameters explain.
R: Number of genes in reference set.
T: Number of genes in test set.
G: Number of genes annotated by given term in reference set.
O: Number of genes annotated by given term in test set.

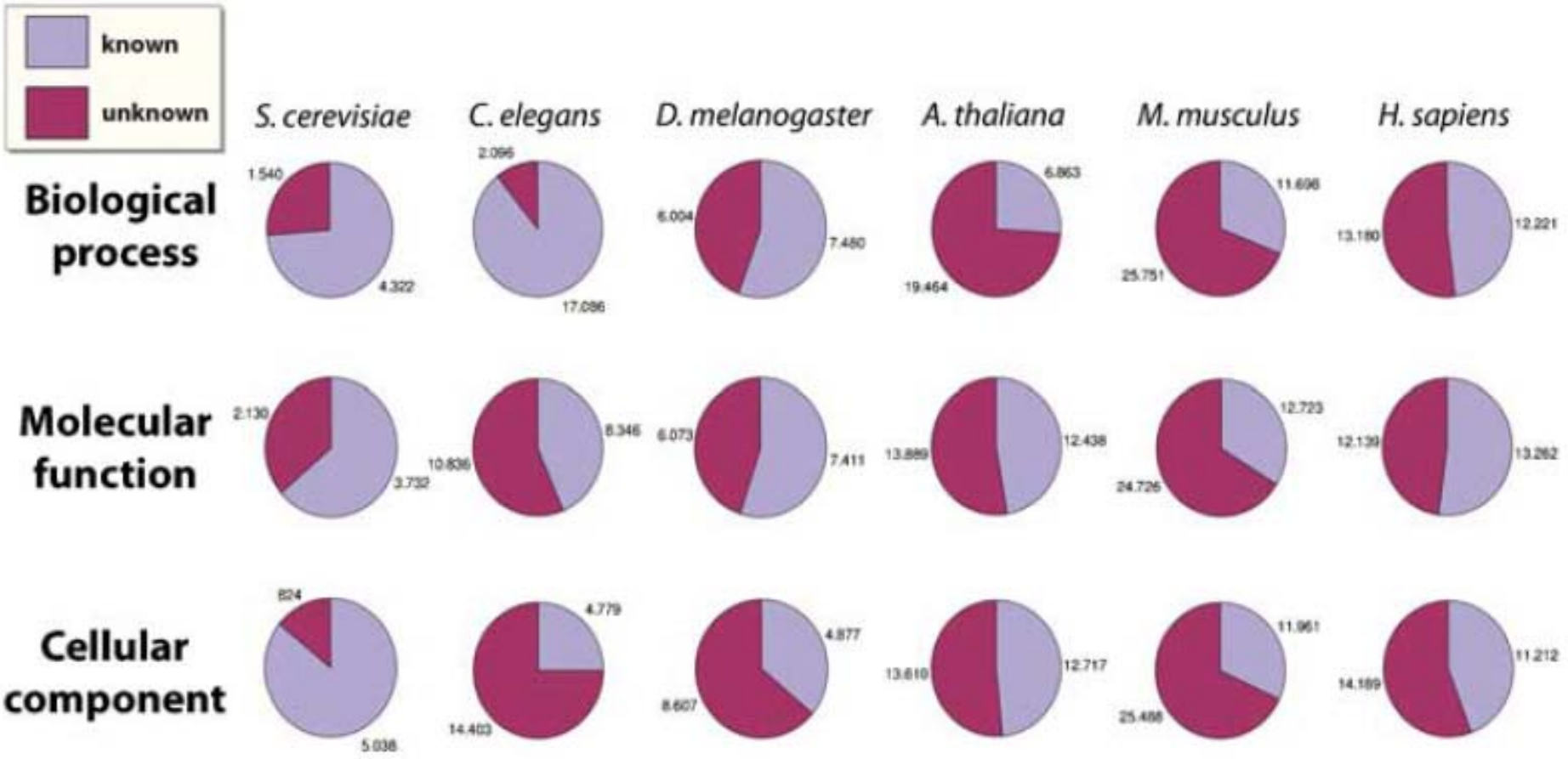
Biological Process

GO: term	p-value	corrected p-value	R	T	G	O	Term name
GO:0019752	5.2E-6	0.0010	2211	319	15	10	carboxylic acid metabolic process
GO:0042180	5.2E-6	0.0010	2211	319	15	10	cellular ketone metabolic process
GO:0043436	5.2E-6	0.0010	2211	319	15	10	oxoacid metabolic process
GO:0044106	5.2E-6	0.0010	2211	319	15	10	cellular amine metabolic process
GO:0044281	5.2E-6	0.0010	2211	319	15	10	small molecule metabolic process
GO:0006082	5.2E-6	0.0010	2211	319	15	10	organic acid metabolic process
GO:0006519	5.2E-6	0.0010	2211	319	15	10	cellular amino acid and derivative metabolic process
GO:0006520	5.2E-6	0.0010	2211	319	15	10	cellular amino acid metabolic process
GO:0009308	5.2E-6	0.0010	2211	319	15	10	amine metabolic process
GO:0006790	8.6E-6	0.0018	2211	319	6	6	sulfur metabolic process

Cell Component

GO: term	p-value	corrected p-value	R	T	G	O	Term name
GO:0000785	2.0E-4	0.0077	2211	319	66	21	chromatin

Discussion



Annotations of genes are far from complete

NOA is an important step towards annotating functions on a biological system since it actually offers a novel way to infer edge function additional with gene function.

Molecular Systems Biology, 2007

Take-home messages

- Network is powerful
- Network is a new platform
- Network can be dangerous
- More stories in network can be expected, but we need to ask a good question first!!!