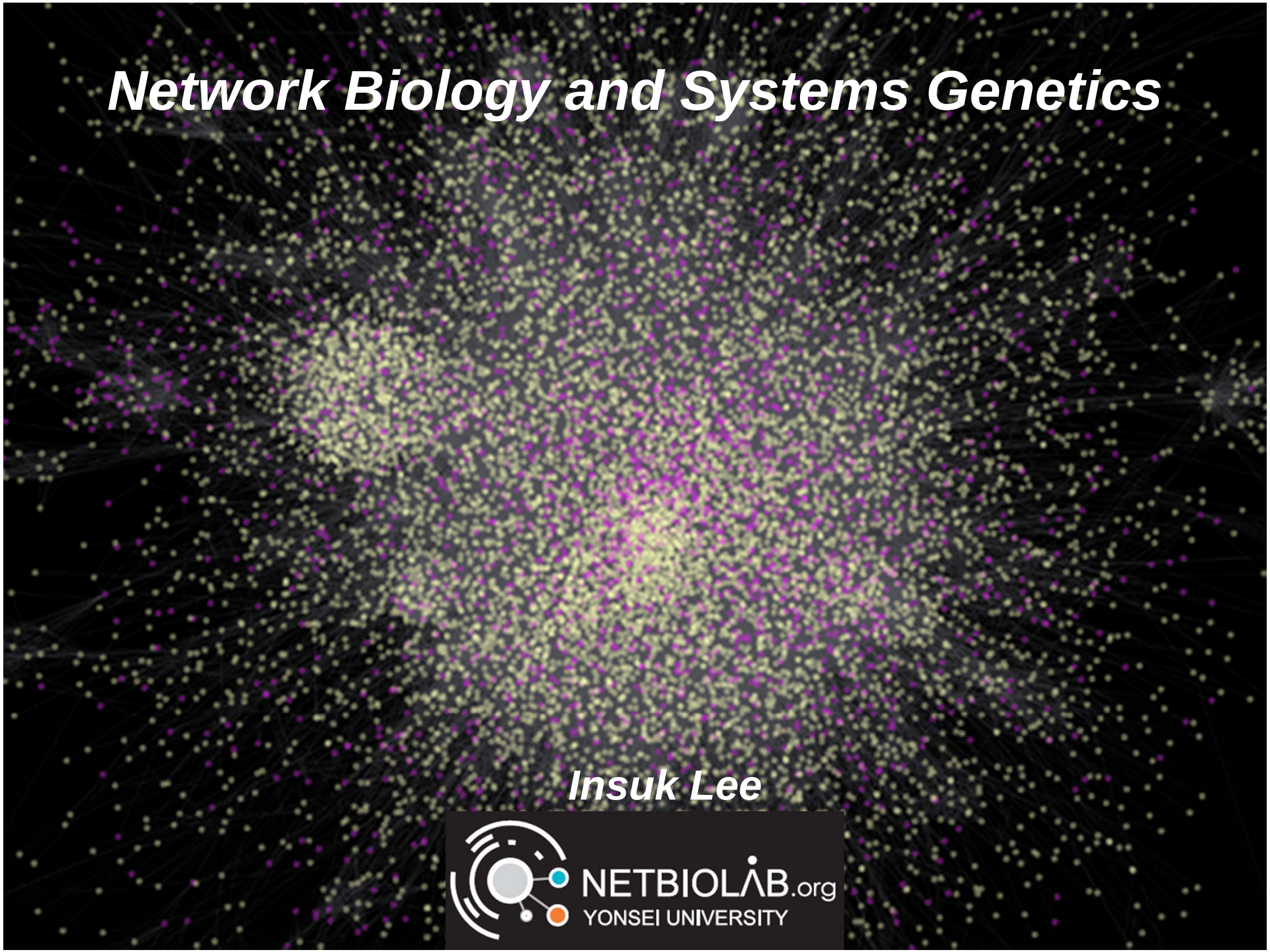
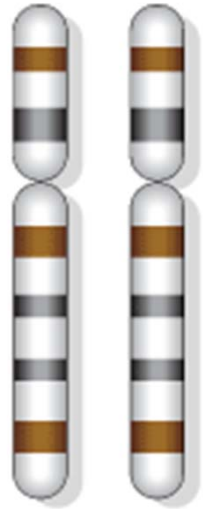

Network Biology and Systems Genetics
Insuk Lee
NETBIOLAB.org
YONSEI UNIVERSITY

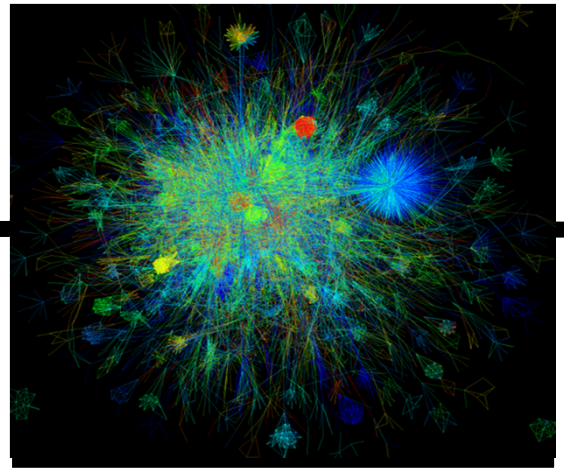
Genotype-Phenotype Association

NETBIOLAB.org
YONSEI UNIVERSITY

One of our ultimate goals in biological research is manipulating important phenotypes by rational gene perturbation.



Genotype



Phenotype

Gene network would help us to map links between genotype and complex phenotypes.

So we will discuss about

(1) How to construct a gene network

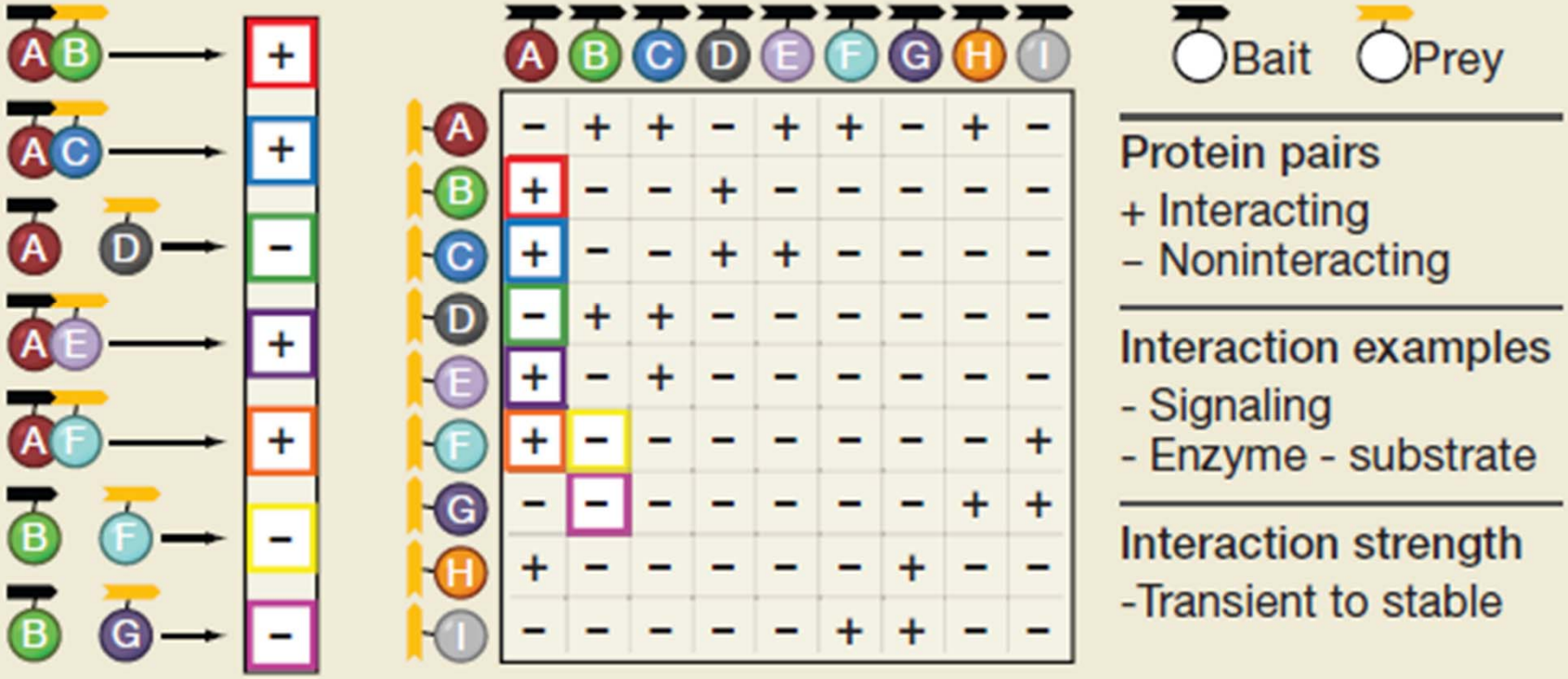
(2) How to use a gene network to map gene-phenotype association

Mapping functional links between genes

1. **Protein-protein interaction**
2. **Genetic interactions**
3. **Genome context**
4. **Co-expression**
5. **Associalogs**

Binary interactions

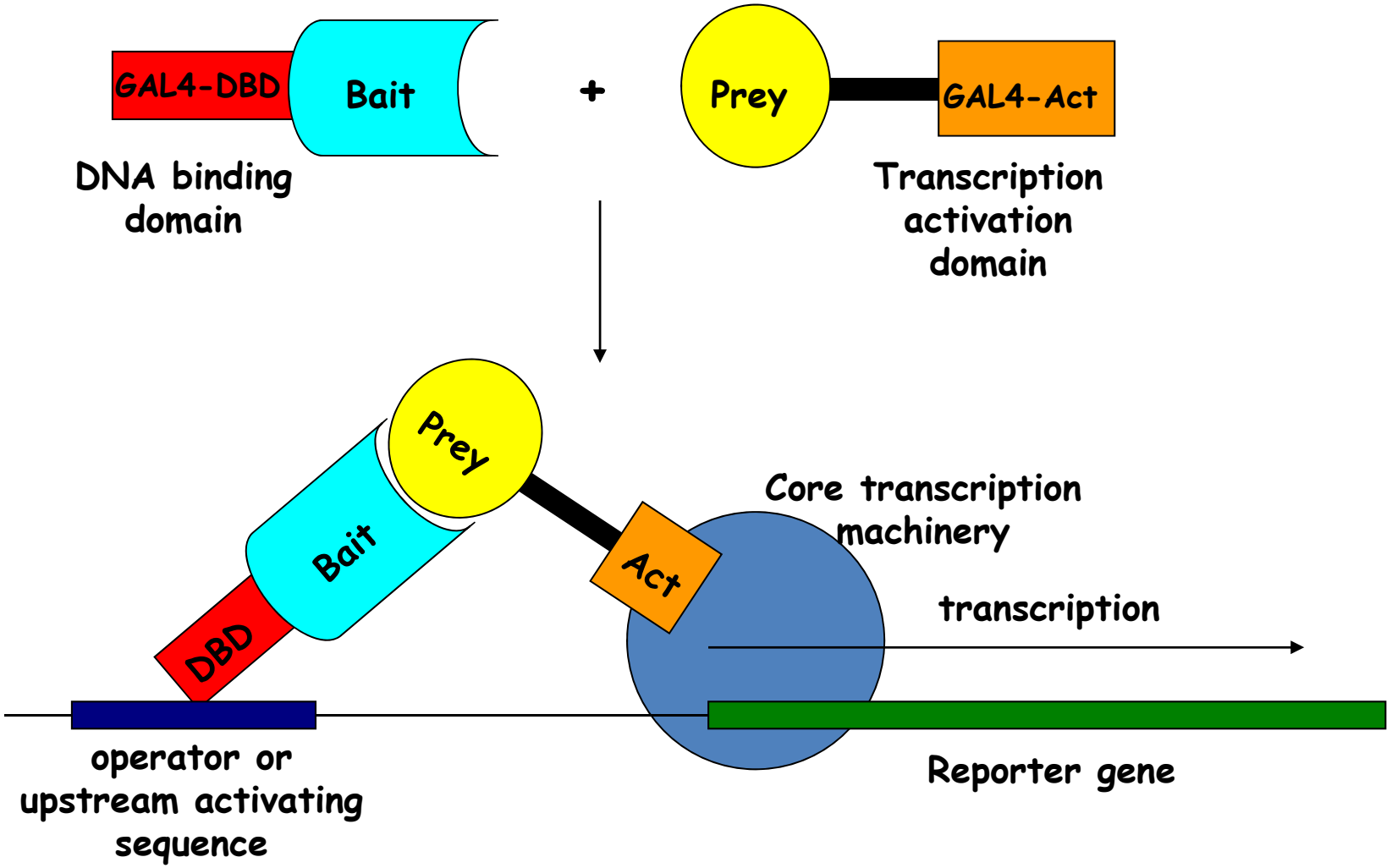
Methods	Split proteins	Assay/Readout
Yeast two-hybrid	Transcription factor, ubiquitin	Transcription
Protein fragment complementation assay	Dehydrofolate reductase	Antibiotic resistence
	GFP or YFP	Fluorescence



Cell 144: SnapShot (2011)

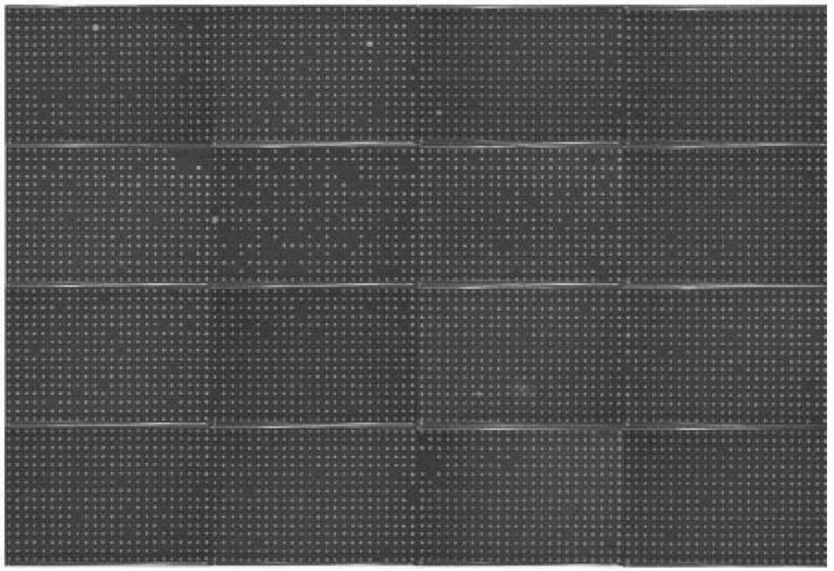
Experimental determination of PPI

Yeast two-hybrid (Y2H): *Nature 340:245 (1989)*

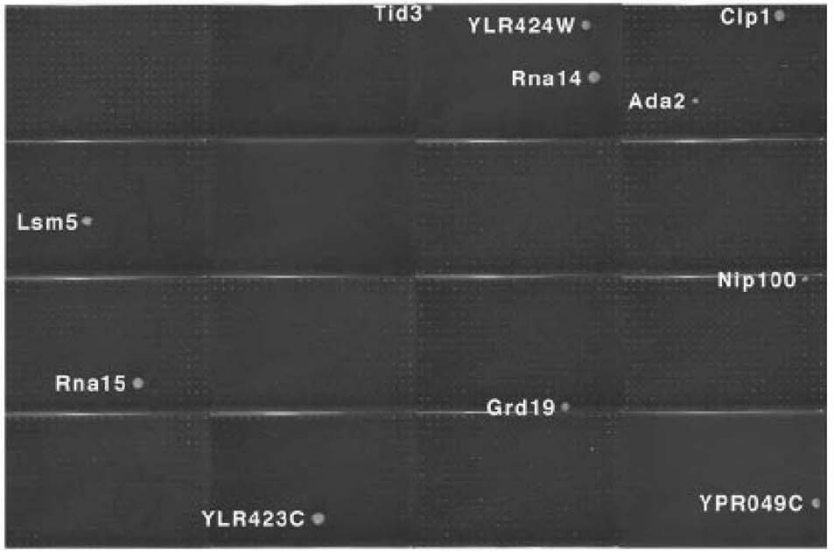


High-throughput yeast two-hybrid by Protein array: Using double transformation, one-by-one assay *(Uetz et al. Nature 2000)*

Array of haploid yeast cells expressing activation domain-prey fusion proteins



Two hybrid positive diploid yeast (on selective media) probed with DNA-binding domain-Pcf11 bait fusion protein

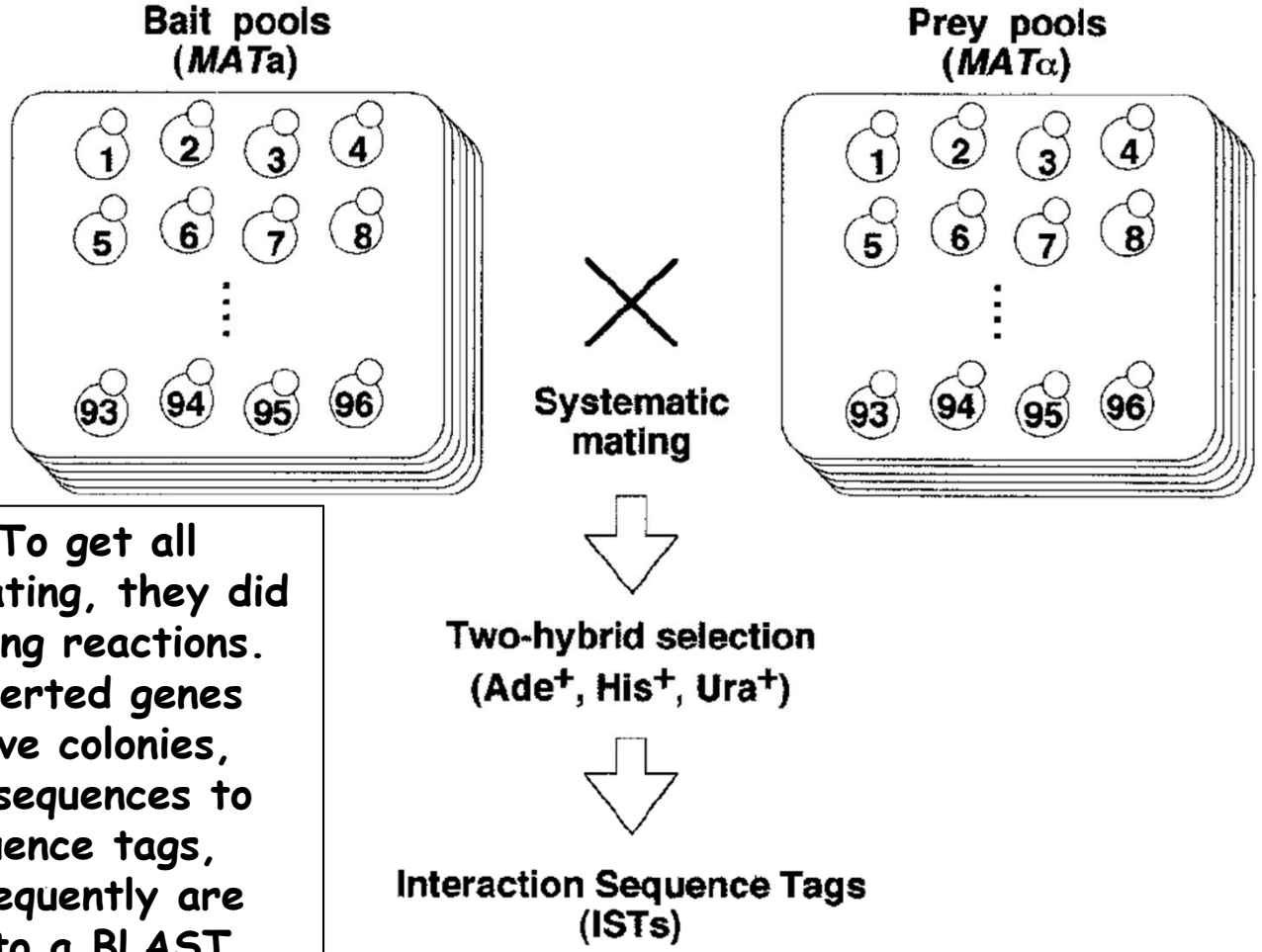


A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*

Peter Uetz*†, Loic Giot*‡, Gerard Cagney†, Traci A. Mansfield‡, Richard S. Judson‡, James R. Knight‡, Daniel Lockshon†, Vaibhav Narayan‡, Maithreyan Srinivasan‡, Pascale Pochart‡, Alia Qureshi-Emili†§, Ying Li‡, Brian Godwin‡, Diana Conover†§, Theodore Kalbfleisch‡, Govindan Vijayadamodar‡, Meijia Yang‡, Mark Johnston†‖, Stanley Fields†§ & Jonathan M. Rothberg‡

NATURE | VOL 403 | 10 FEBRUARY 2000

High-throughput Y2H by Pooled mating *(Ito et al. PNAS, 2001)*



62 pools. To get all possible mating, they did 3,844 mating reactions. Amplify inserted genes from positive colonies, then read sequences to obtain sequence tags, which subsequently are subjected to a BLAST search to decode interactions. Uetz also did pooled mating assay.

A comprehensive two-hybrid analysis to explore the yeast protein interactome

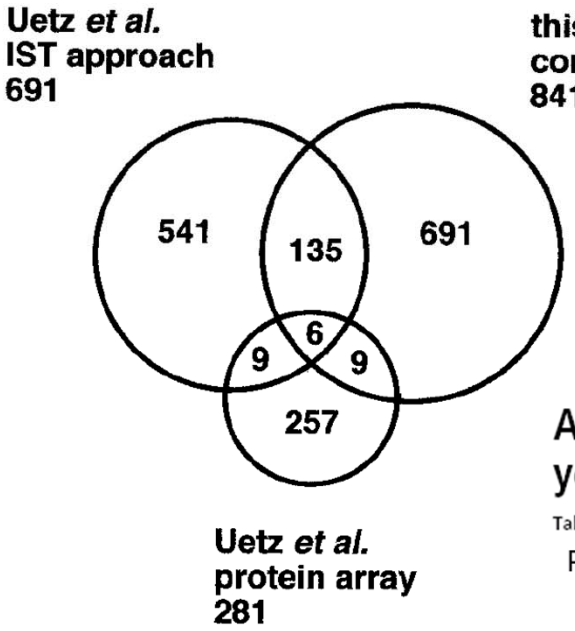
Takashi Ito*†, Tomoko Chiba*, Ritsuko Ozawa‡, Mikio Yoshida§, Masahira Hattori‡, and Yoshiyuki Sakaki‡¶

PNAS | April 10, 2001 | vol. 98 | no. 8 | 4569–4574

Y2H maps are not much overlapped.

The largest network reconstructed from interaction data

Dataset	Proteins in the largest network/total proteins (%)	Interactions in the largest network/total interactions (%)
Conventional studies*	1,003/1,858 (54)	1,504/2,209 (68)
This study†		
Core data	417/797 (52)	544/806 (67)
All data	2,838/3,278 (87)	4,224/4,475 (94)

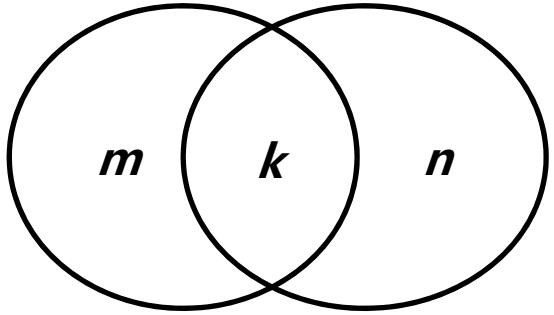


A comprehensive two-hybrid analysis to explore the yeast protein interactome

Takashi Ito*†, Tomoko Chiba*, Ritsuko Ozawa‡, Mikio Yoshida§, Masahira Hattori‡, and Yoshiyuki Sakaki‡¶

PNAS | April 10, 2001 | vol. 98 | no. 8 | 4569–4574

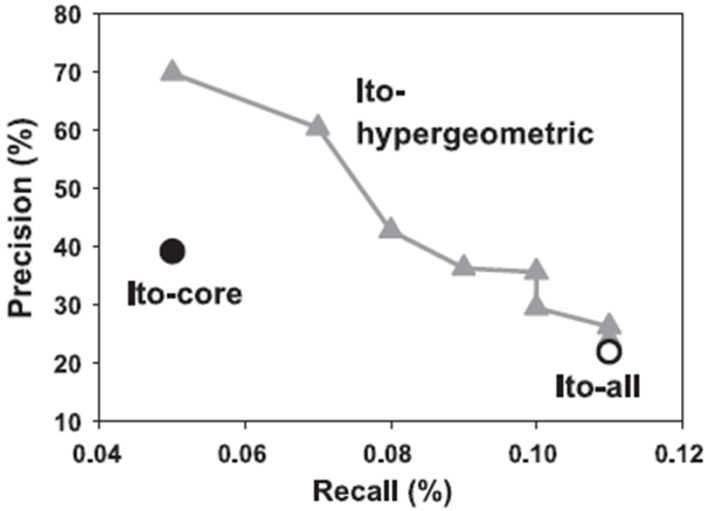
Modeling PPI by hypergeometric distribution



$$p(\#\,\text{interactions} \geq k \mid n, m, N) = \sum_{i=k}^{\min(n,m)} p(i \mid n, m, N)$$

where

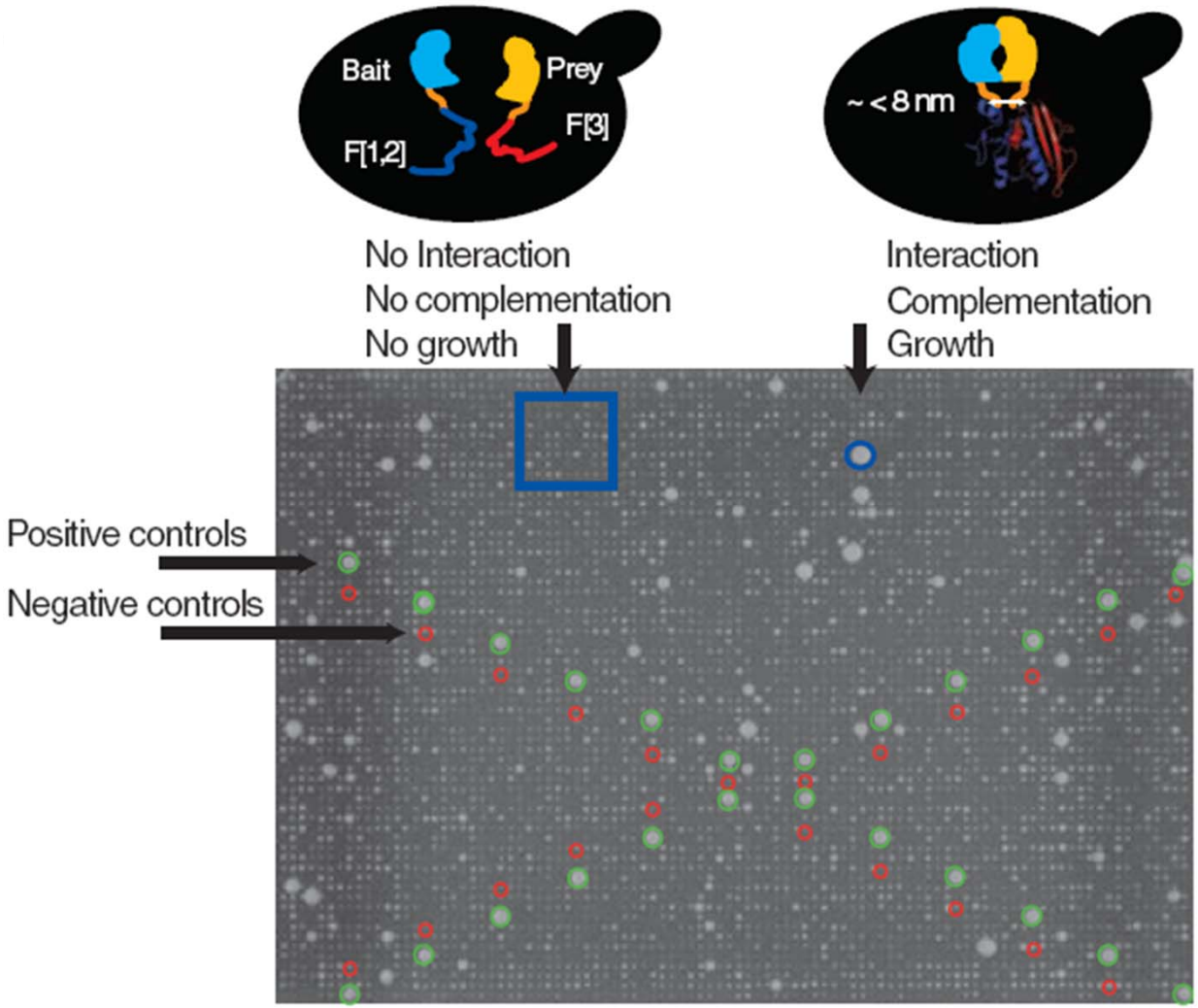
$$p(i \mid n, m, N) = \frac{\binom{n}{i}\binom{N-n}{m-i}}{\binom{N}{m}}$$



where k = the number of times the interaction between A and B is observed, n and m are the total number of interactions for proteins A and B, and N is the total number of interactions observed in the entire data set.

Protein-fragment complementation assay (PCA) *Science 320:1465*

Two proteins of interest are fused to *complementary fragments of a reporter protein*. If the proteins of interest interact physically, the reporter fragments are brought together and fold into their native structure, thus reconstituting the reporter activity of the PCA.



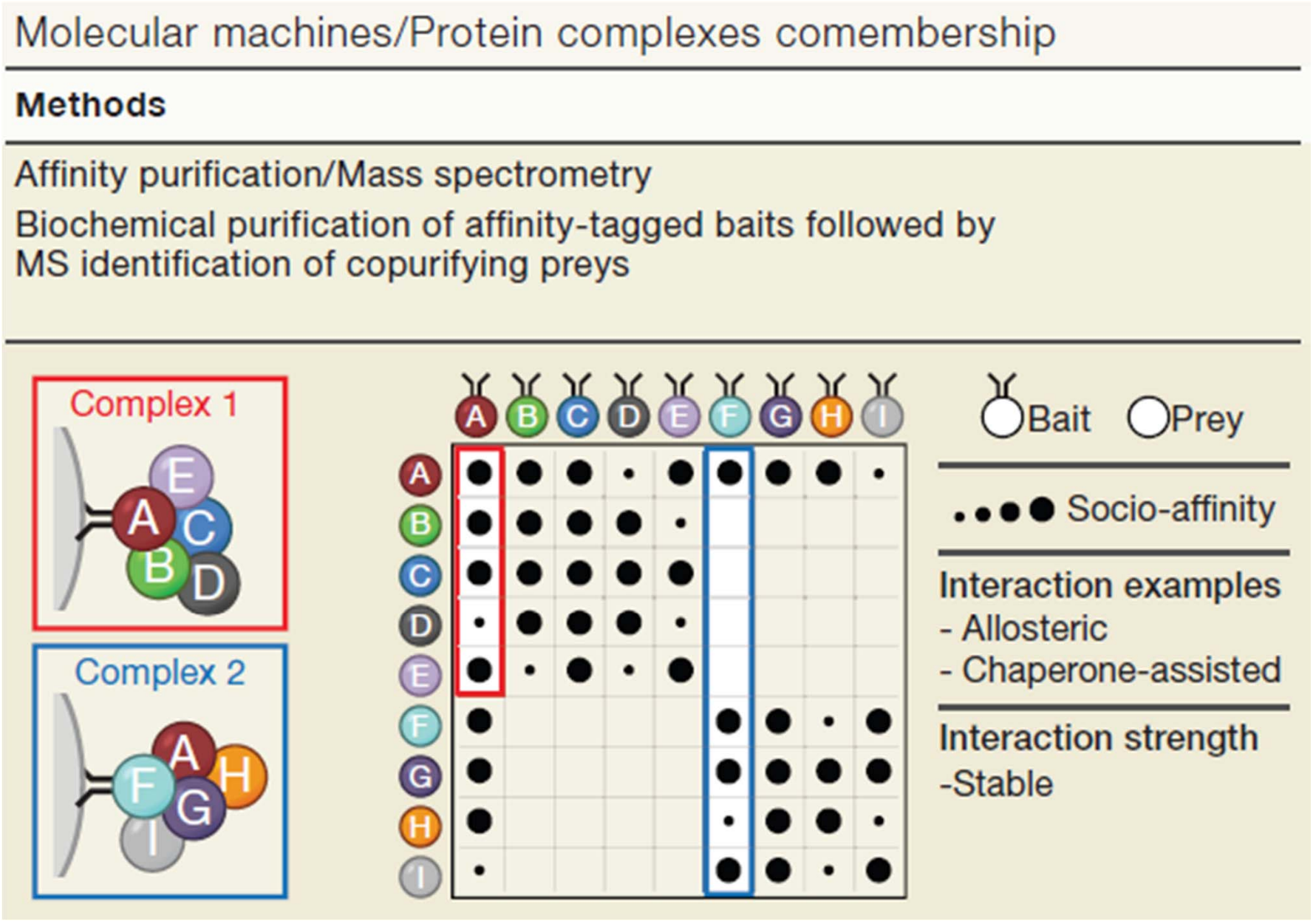
- Neither Y2H nor TAP-MS measures interactions between proteins in their ***natural cellular context***, and are not easily amenable to studying protein complexes that are transiently associated or dynamic under different conditions, that do not survive *in vitro* purification, or that cannot be transported to the nucleus.

- ***PCA provides a simple direct means for the detection of PPIs in vivo***, and do so with endogenously expressed full-length proteins in their native post-translationally modified states and cellular location.

- *Survival-selection assay* based on a mutant of Murine dihydrofolate reductase (mDHFR) that is insensitive to the DHFR inhibitor methotrexate but retains full catalytic activity and allows detection of PPIs with as few as 25-100 complexes per cell.

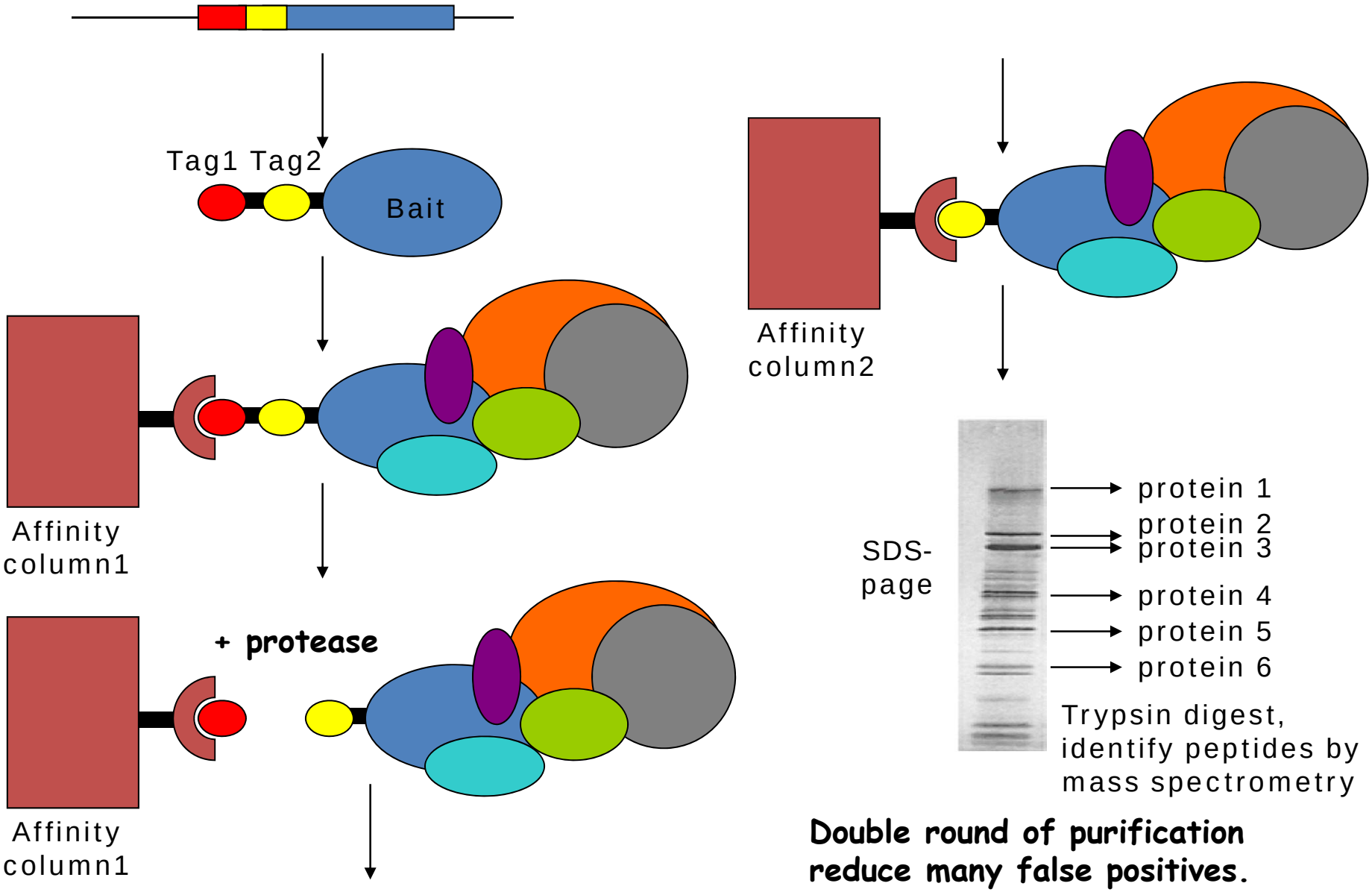
- Tarassov *et al.* identified 2770 interactions among 1124 endogeneously expressed yeast proteins. Most were not known by other previous studies.

- However, precision-recall analysis of PCA shows generally worse performance than Y2H (*personal analysis results*).



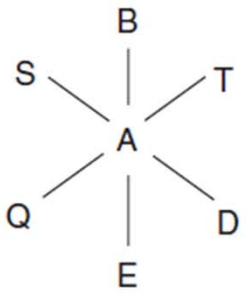
Cell 144: SnapShot (2011)

High-throughput complex mapping by affinity-purification followed by mass spectrometry (AP-MS)

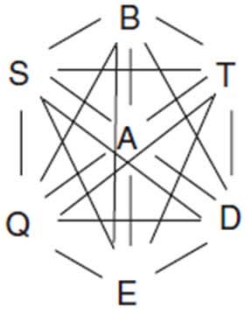


Double round of purification reduce many false positives.

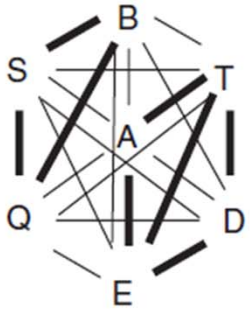
Modeling binary interactions from protein complexes (*BMC bioinformatics 8:236*)



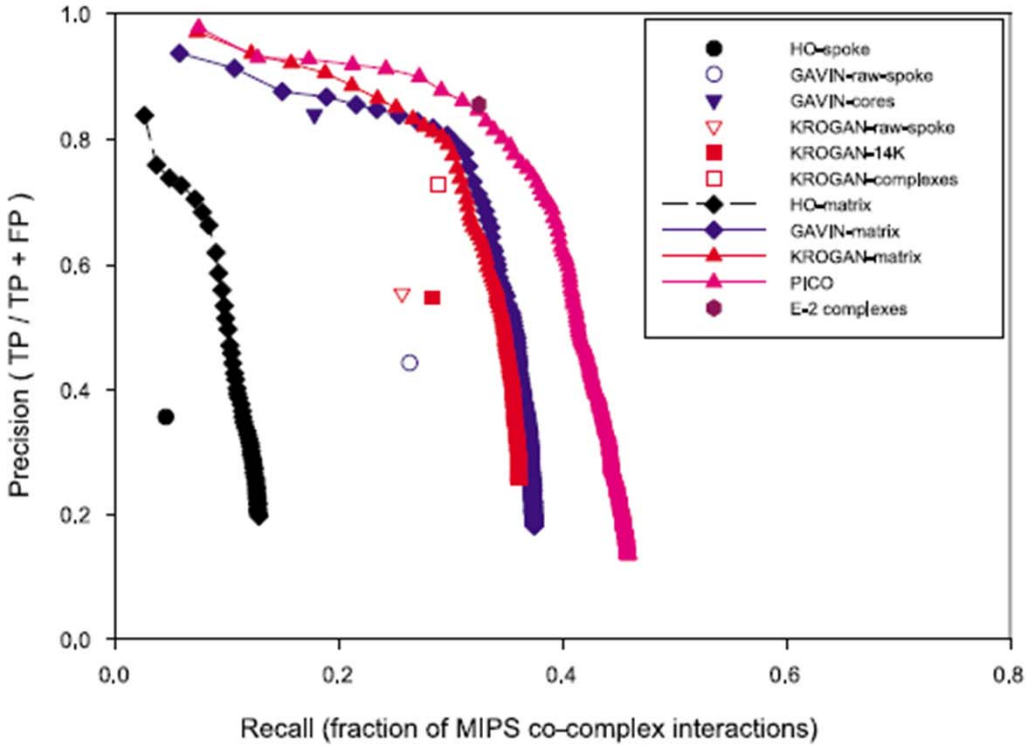
'Spoke' Model



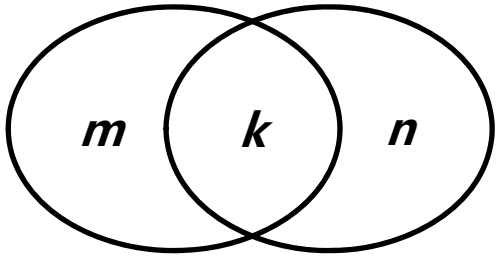
'Matrix' Model



Weighted Model



Use hypergeometric probability



$$p(\#\,\text{interactions} \geq k \mid n, m, N) = \sum_{i=k}^{\min(n,m)} p(i \mid n, m, N)$$

where

$$p(i \mid n, m, N) = \frac{\binom{n}{i}\binom{N-n}{m-i}}{\binom{N}{m}}$$

where k = the number of times the interaction between A and B is observed, n and m are the total number of interactions for proteins A and B, and N is the total number of interactions observed in the entire data set. When applied

"biologically significant"

∨

~All protein interactions are functional interactions

Not all functional interactions are physical interactions

Many other biological data also can support functional interactions.

- Genetic interaction by synthetic lethal screen
- Genome context relationship with many sequenced genomes
- Co-expression across array of transcriptome profiles
- Many more…

Functional interaction from genetic (epistatic) interaction

Definitions of epistasis

Nature Reviews Genetics 5:618, 9:855, Genetics 149:1167

1. From the ***Mendelian (classical geneticist) viewpoint*** (by Willian Beteson 1909):

- The ***action of one locus mask the allelic effects of another locus***, in the same way that completely dominant alleles mask the effects of the recessive allele at the same locus.
- Epistasis translates directly to "***standing upon***".
- Frequently genes interact with one another, ***distorting simple Mendelian ratios*** and sometimes leading to novel phenotypes.

Interaction Type	*A-B-*	*A-bb*	*aaB-*	*aabb*
Classical ratio	9	3	3	1
Dominant epistasis	12		3	1
Recessive epistasis	9	3	4	
Duplicate genes with cumulative effect	9	6		1
Duplicate dominant genes	15			1
Duplicate recessive genes	9	7		
Dominant & recessive interaction	13		3	

Some unusual segregation ratios. Arrows join genotypes with similar phenotypes.

Mendelian epistasis in the vulval differentiation pathway of *C. elegans*.

- The effect of lin-39 is masked by the effect of lin-26, and thus lin-26 is 'epistatic to', and upstream of, lin-39.
- Similarly, lin-39 is epistatic to let-23.

Journal of Biology 8:35 (2009)

2. From the ***Statistical (population) geneticist viewpoint*** (by R. A. Fisher 1918):

- ***Any statistical deviation from the additive*** (or multiplicative depending on scale) combination of two loci in their effects on a phenotype (***epistatic deviation***).
- This is more inclusive than Beteson's definition because many forms of gene interaction can lead to epistatic deviations.

Formal representation of epistasis

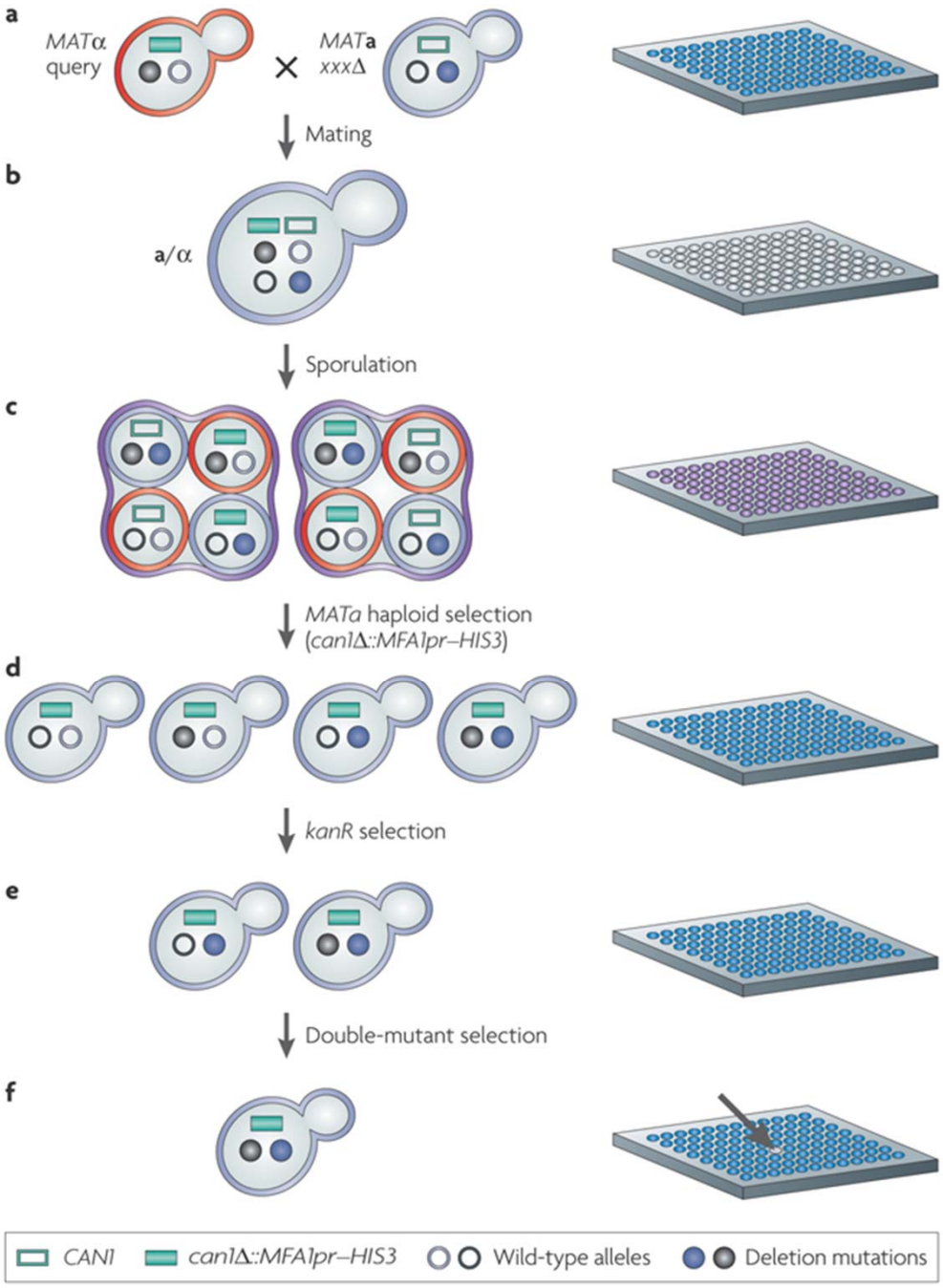
$\varepsilon = W_{ab} - W_a * W_b$

Where W_a, W_b, and W_{ab} represent the ***fitness (or growth rates)*** relative to wild-type organisms with mutation A, with mutation B, and with both mutations, respectively.

$\varepsilon = 0$ for ***no epistasis***

$\varepsilon < 0$ for ***aggravating, negative, synergistic interaction, synthetic sick, synthetic lethal*** interactions

$\varepsilon > 0$ for ***alleviating, positive, antagonistic, buffering, partial suppressor*** interactions



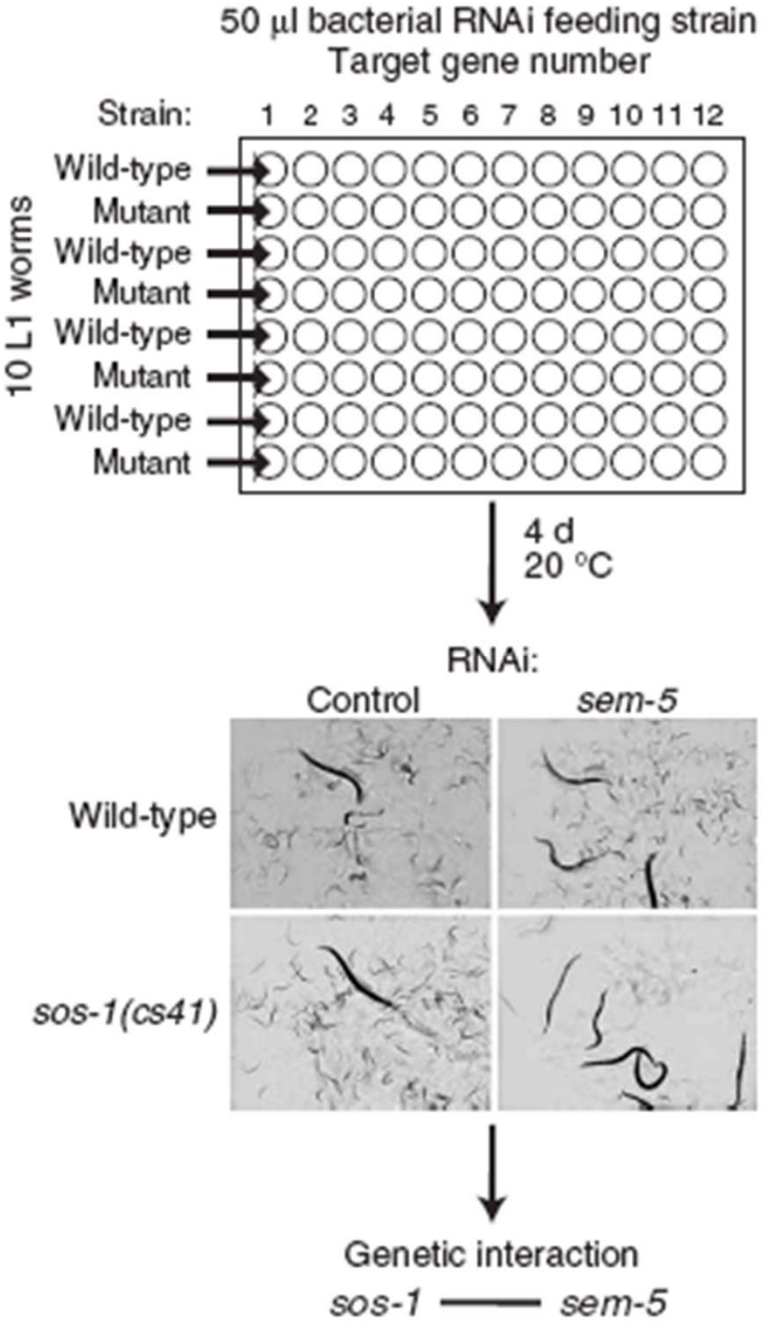
SGA *(Science 294:2364)*

a. **A query mutation is first introduced into a haploid starting strain, of mating type MATα, and then crossed to the array of gene-deletion mutants of the opposite mating type, MATa.**

b. **Sporulation of resultant diploid cells leads to the formation of double-mutant meiotic progeny. The MATα strain carries a reporter, *MFA1pr-HIS3*, that is only expressed in MATa meiotic progeny, which ensures that carryover of the diploid parental strain and/or conjugation of meiotic progeny does not give rise to false-negative interactions.**

c-f. **Both query mutation and the gene-deletion mutations were linked to dominant selectable markers to allow for selection of double mutants.**
Double mutants with slow growth are synthetic sick/lethal partner candidate.

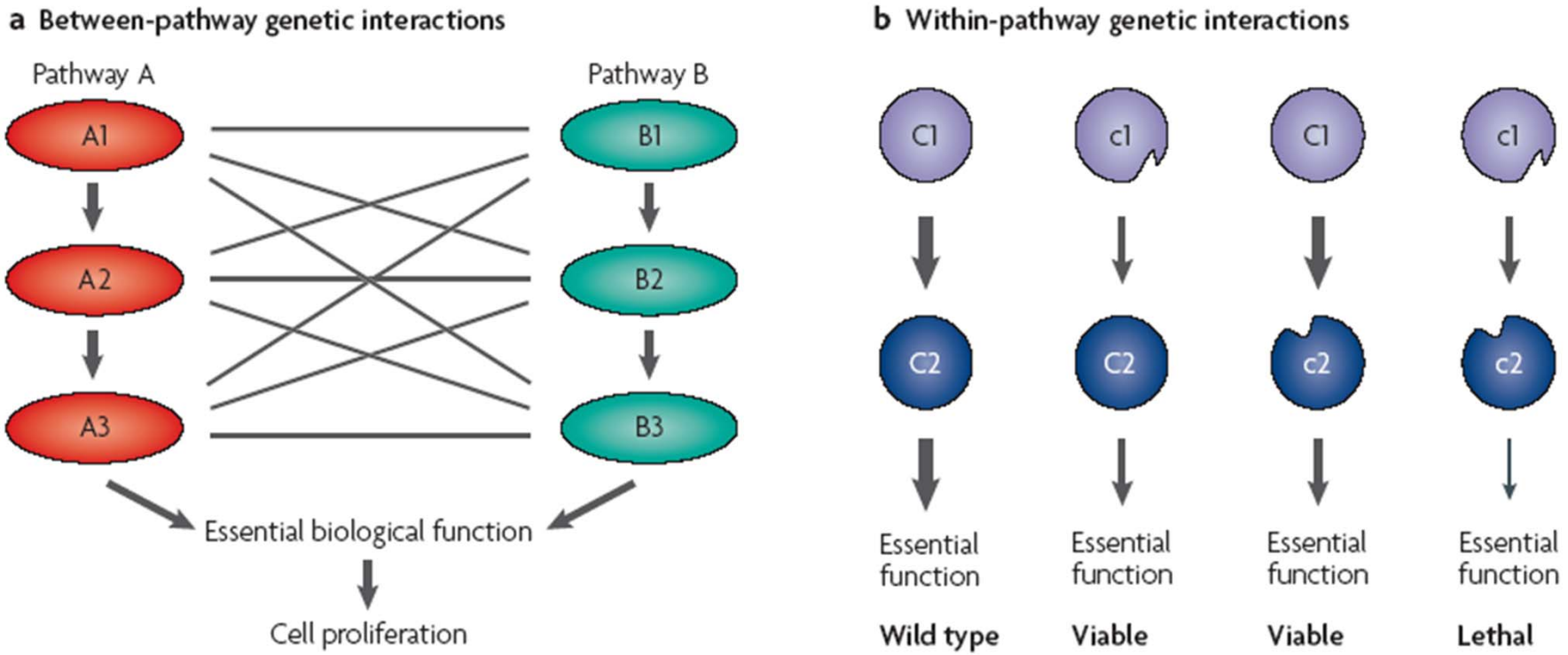
Limit: cannot test essential genes, false positives (~50%)



Genetic interaction screen for *C. elegans* using RNA interference (Nature Genetics 38:896)

- Target genes are knock down using RNAi by bacterial feeding in the background of defective query gene.
- Lehner *et al*. screen ~1750 RNAi library genes for signaling pathway components for 37 query strains, so tested ~65,000 pairs, and identified ~350 genetic interactions. All 37 query genes function in signaling pathways that are mutated in human diseases including components of the EGF/Ras, Notch and Wnt pathways.
- They identified a class of highly connected '***hub***' genes: inactivation of these genes can enhance the phenotypic consequences of mutation of many different genes. These hub genes all encode chromatin regulators, and their activity as genetic hubs seems to be conserved across animals.

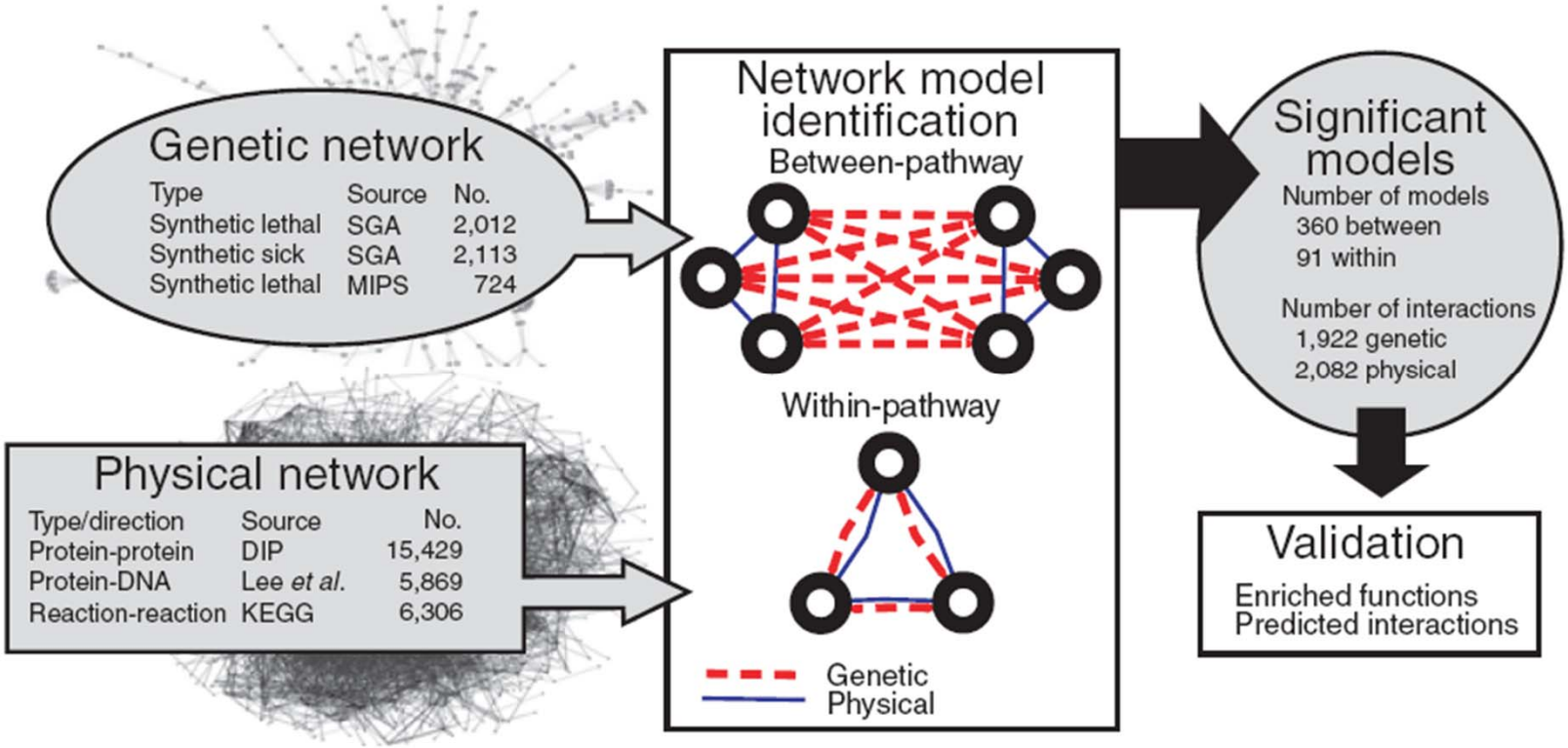
Mechanistic interpretation of genetic interactions



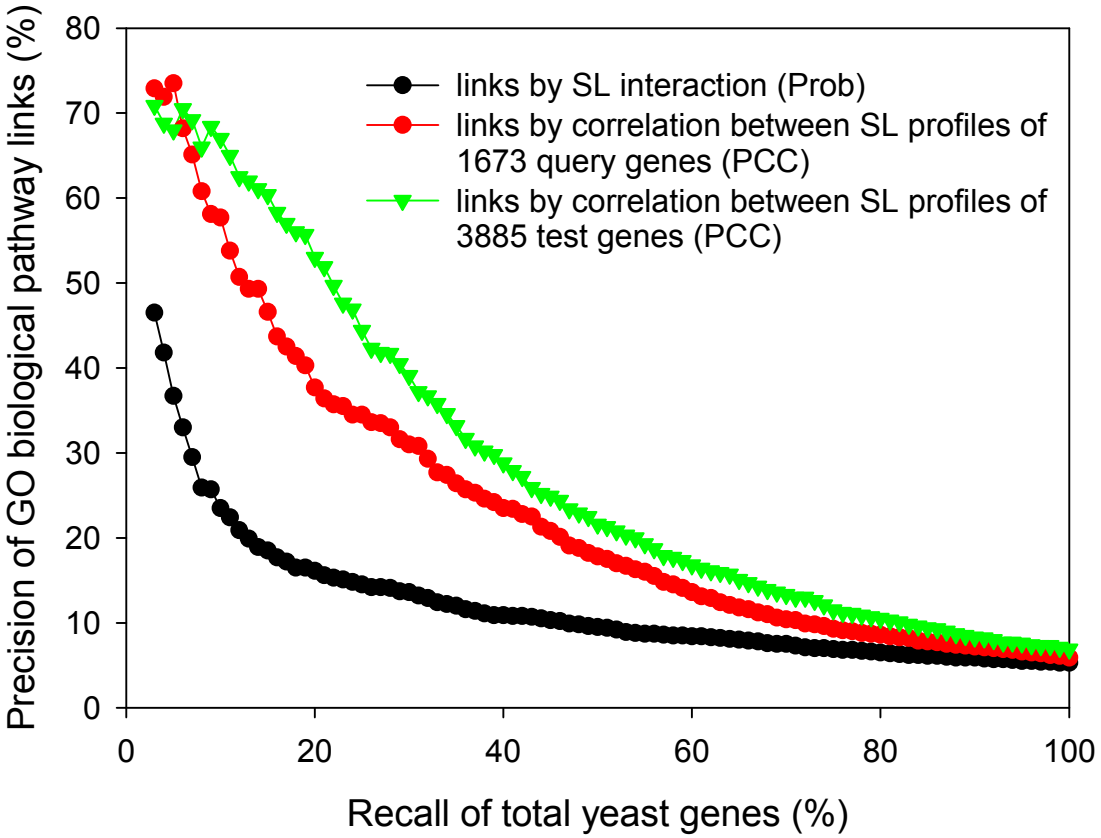
- **Possible mechanisms depend on the characteristics of the interacting alleles. The common interpretation is that the genes function in parallel pathways that impinge on a shared essential function. This is often referred to as the '*between-pathway model*' and typically reflects bidirectional genetic redundancy, in that each pathway compensates for defects in the other.**
- **Conversely, in the '*within-pathway model*', synthetic lethality indicates that both gene function in the same essential pathway, the function of which is diminished by each mutation.**
- **It has been demonstrated that *positive genetic interactions* can identify pairs of genes for *within-pathway* (*Cell 123:507, Nature 446:806*), whereas *negative genetic interactions* exist for *between-pathway* (*Science 303:808*).**

Modeling genetic interactions using protein physical interaction map
(By Kelley and Ideker, Nature Biotechnology 23:561)

- Used both between-pathway and within-pathway models. Here, 'pathway' is loosely defined as any densely connected set of proteins in the physical network. This method can explain ~40% of known genetic interactions that time, and *between-pathway explanations are better than within-pathway explanations*.



Pathway links by direct genetic link
vs.
Pathway links by similarity between genetic interactors



- Why? Many direct genetic interactions are between pathways, thus they do not support functional association.

Discovery of functional interaction from Genome sequences

Genomes carry ***intrinsic information*** about the cellular systems and pathways they encode. This information can be revealed by comparative genomics.

1 genome ---> can model the genes

>1 genome --> can model
the functions of the genes
gene, pathway & organismal evolution
genomic/organismal diversity
molecular characteristics of speciation
etc...

Methods for using comparative genomics for discovering pathways:

(1) Analyzing gene fusions

"Inferring protein interactions from genome sequences on the basis of the observation that some pairs of interaction proteins have homologs in another organism fused into a single protein chain" (*Nature 285:751*)

(2) Analyzing gene phylogenies

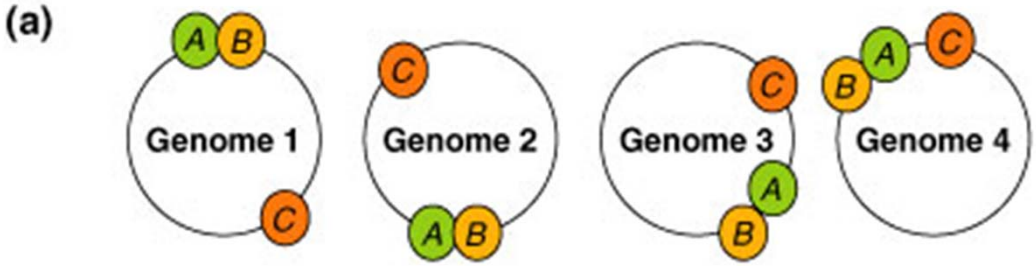
"Proteins that function together in a pathway or structural complex are likely to evolve in a correlated fashion. During evolution, all such functionally linked proteins tend to be either preserved or eliminated in a new species." (*PNAS 96:4285*)

(3) Analyzing operons (Conserved gene neighbors)

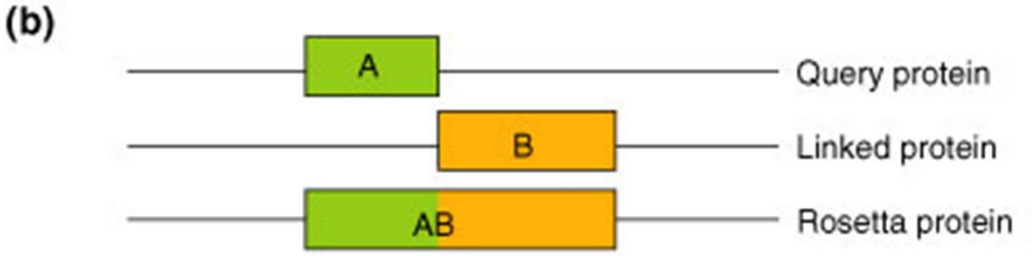
"One of the most striking features of prokaryotic gene clusters is that typically they are composed of functionally related genes." (*PNAS 96:2896*)

Genes --> comparisons between the genes from different organisms --> discovery of pathways --> integration of the pathways for all of the genes of a single organism --> "global" view of pathway

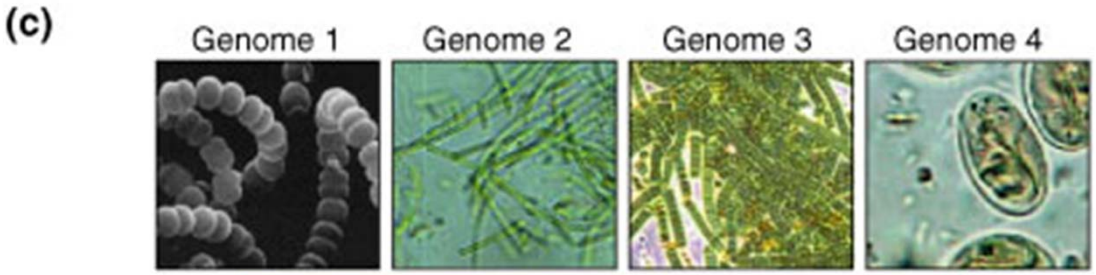
Genome context approaches



Gene neighbors



Gene fusion (Rosetta stone proteins)

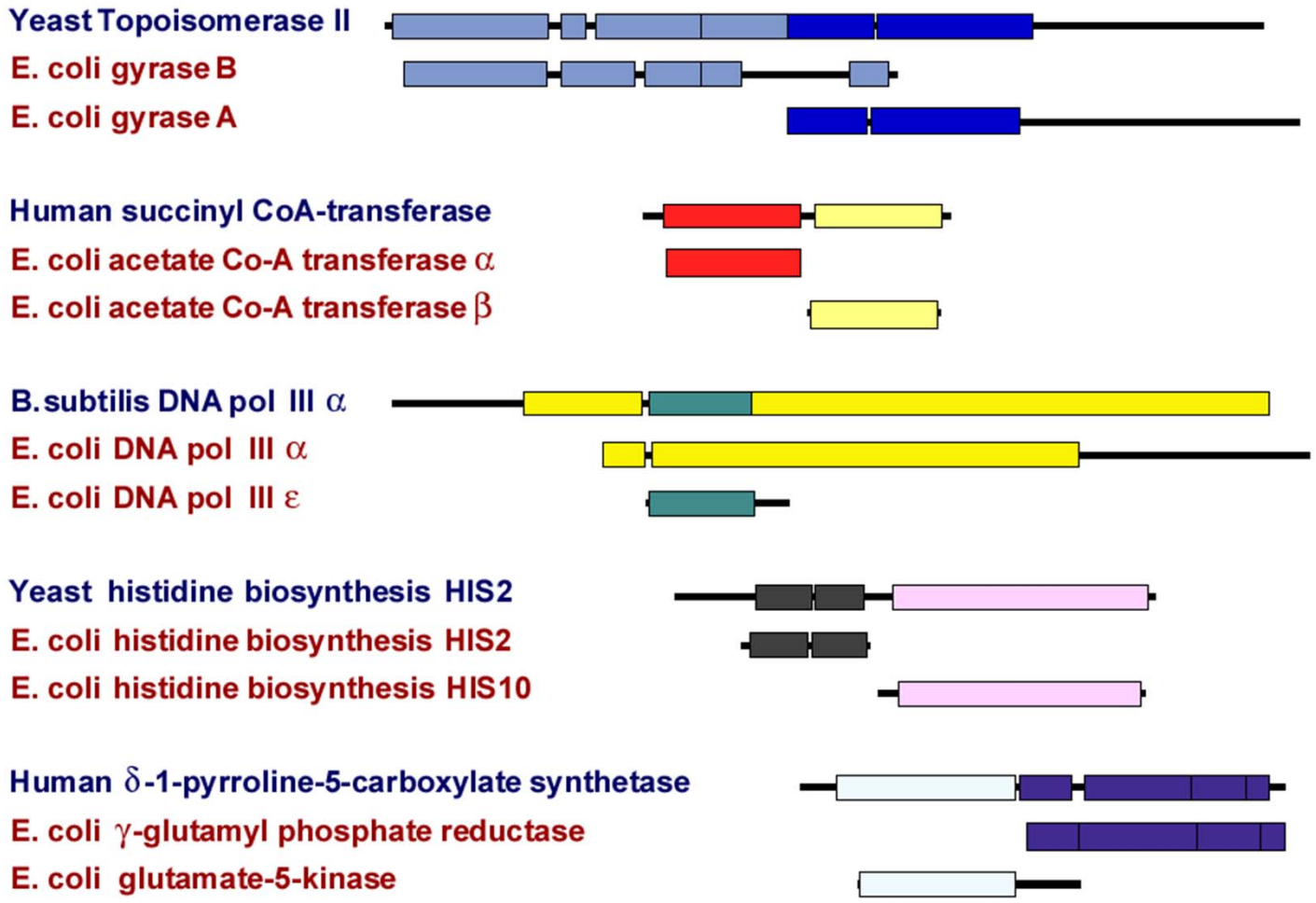


Protein A	1	1	1	0
Protein B	1	1	1	0
Protein C	1	0	1	1
Protein D	1	1	0	1

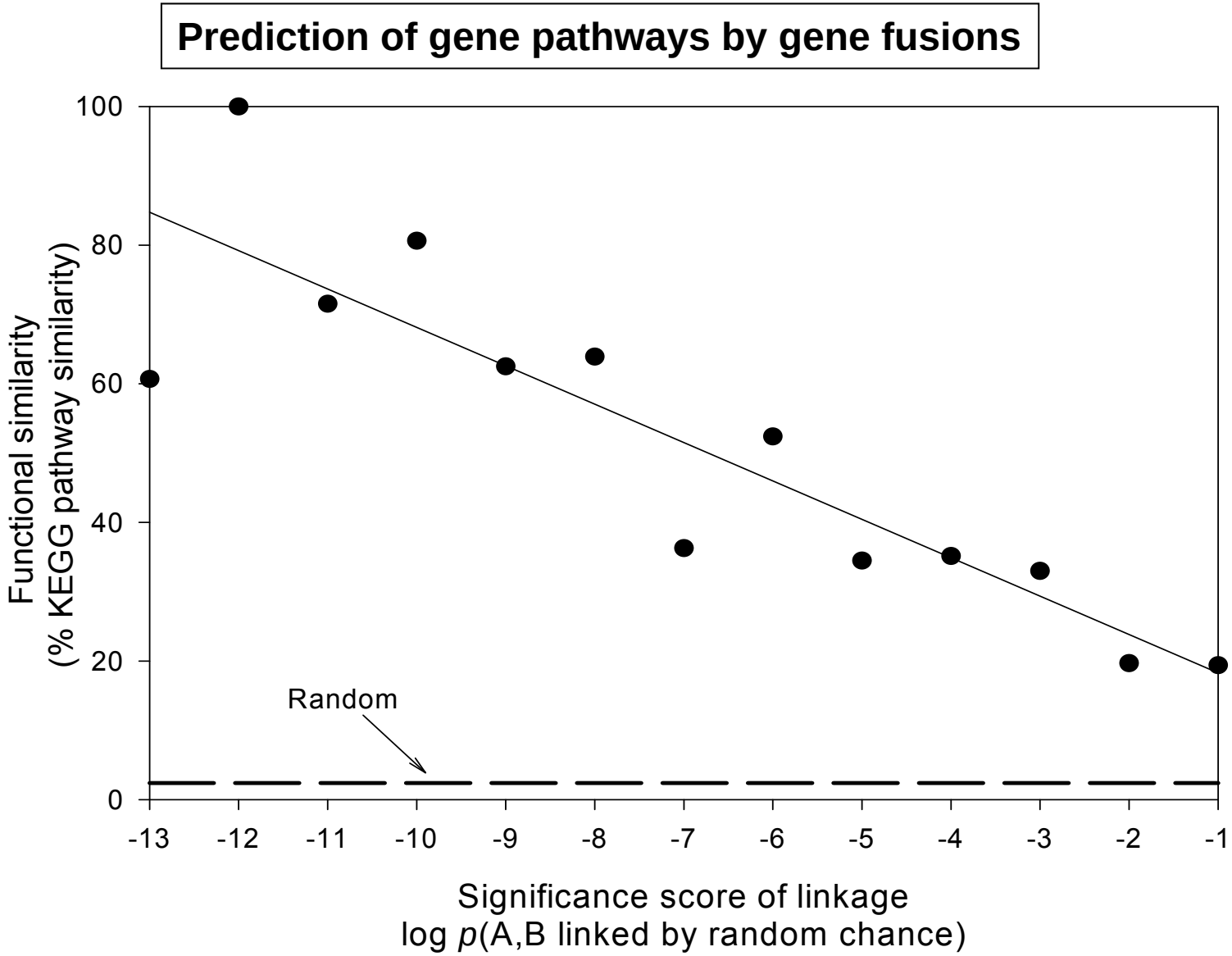
Phylogenetic profiling

1. Gene Fusion

Some pairs of interaction proteins have homologs in another organism fused into a single protein chain.

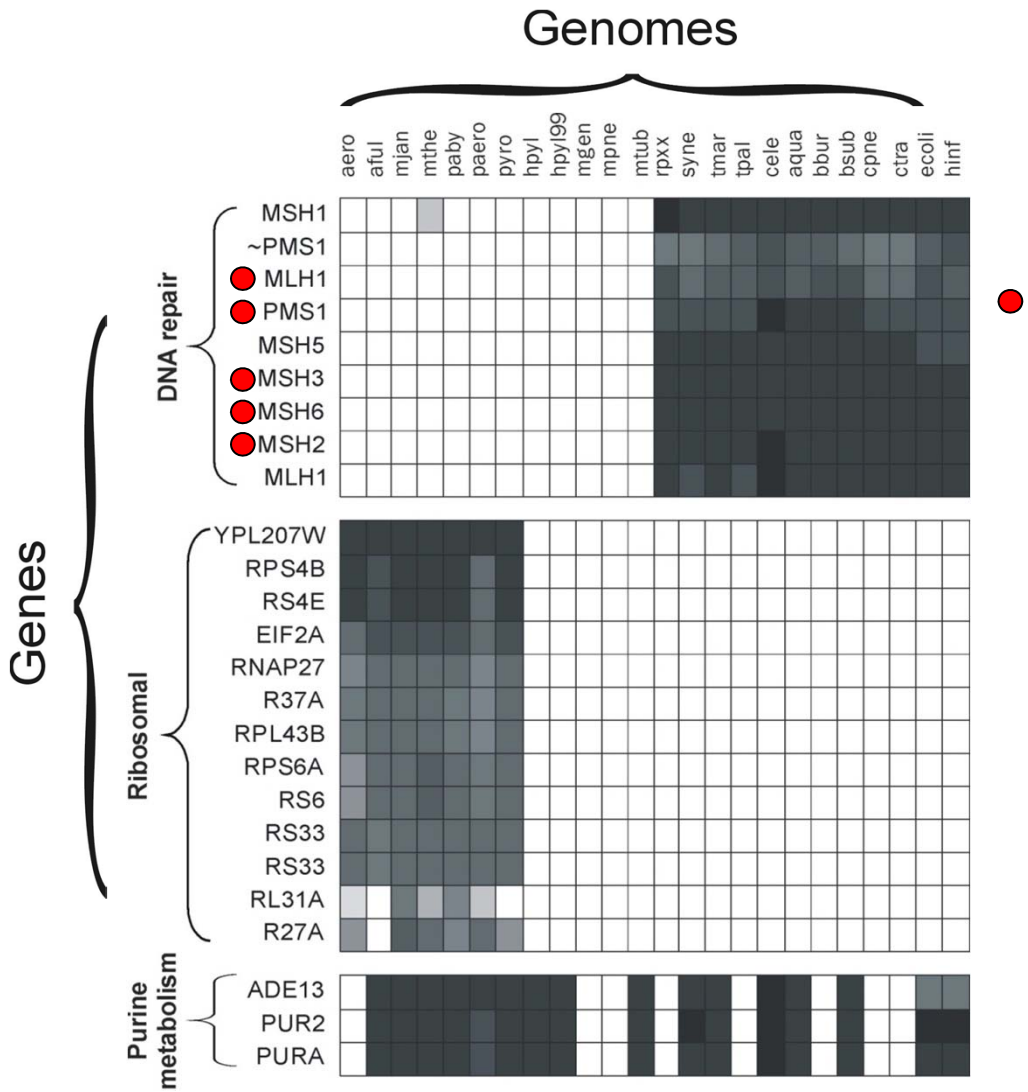


Science **285**, 751-753 (1999)


Prediction of gene pathways by gene fusions
Functional similarity
(% KEGG pathway similarity)
100
80
60
40
20
0
Random
-13
-12
-11
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
Significance score of linkage
log p(A,B linked by random chance)

2. Phylogenetic Profiling

During evolution, all such functionally linked proteins tend to be either preserved or eliminated in a new species (co-evolution).

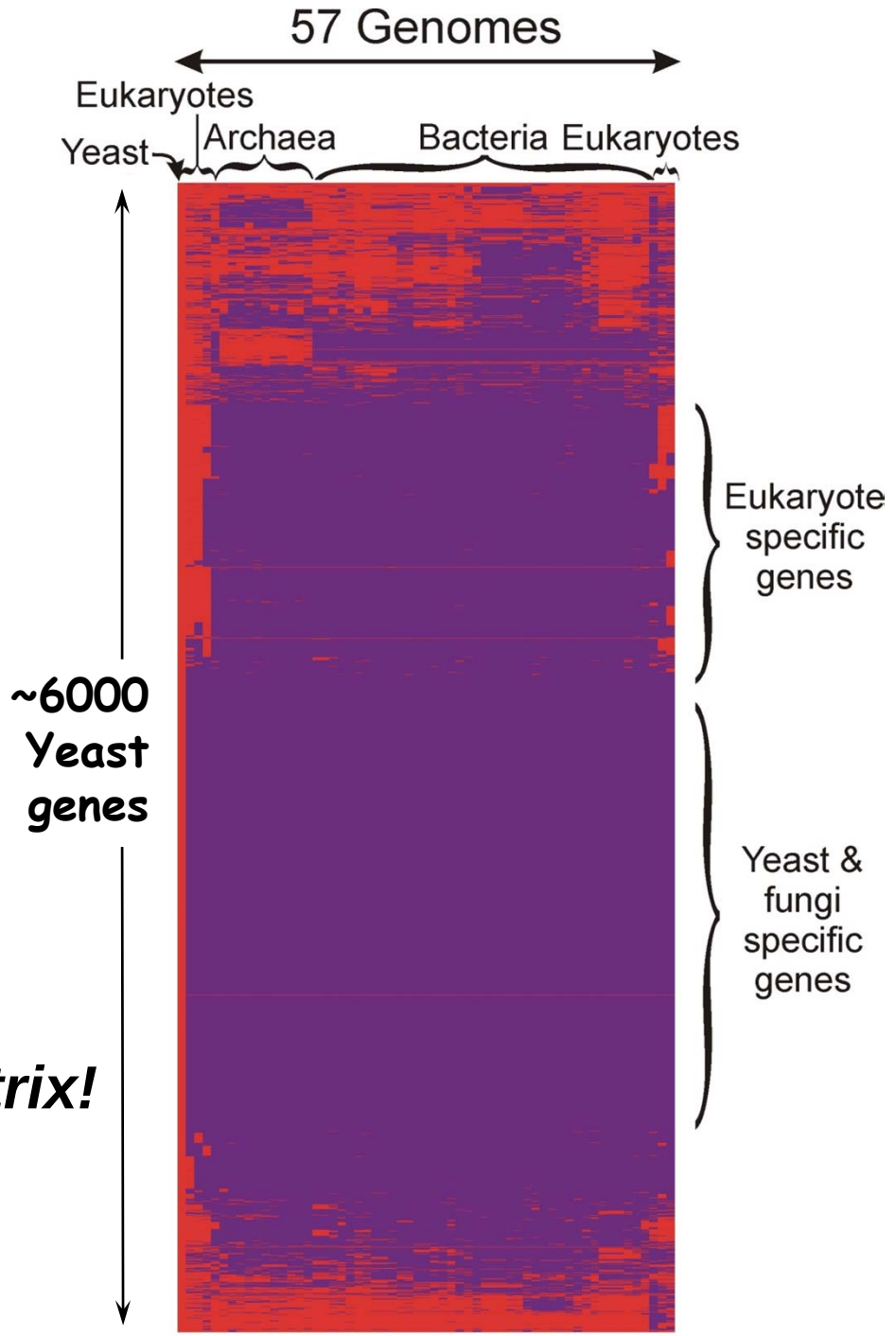


● = Implicated in human colorectal cancer

Nature **402**, 83-86 (1999)

Phylogenetic profiles of ~6000 genes of *Saccharomyces cerevisiae*

Red=Present
Purple=Absent



It is a very sparse matrix!

Estimating the significance of matching phylogenetic profiles

Profile 1: `000000000100000000000000000`
Profile 2: `000001000000000000000000000`

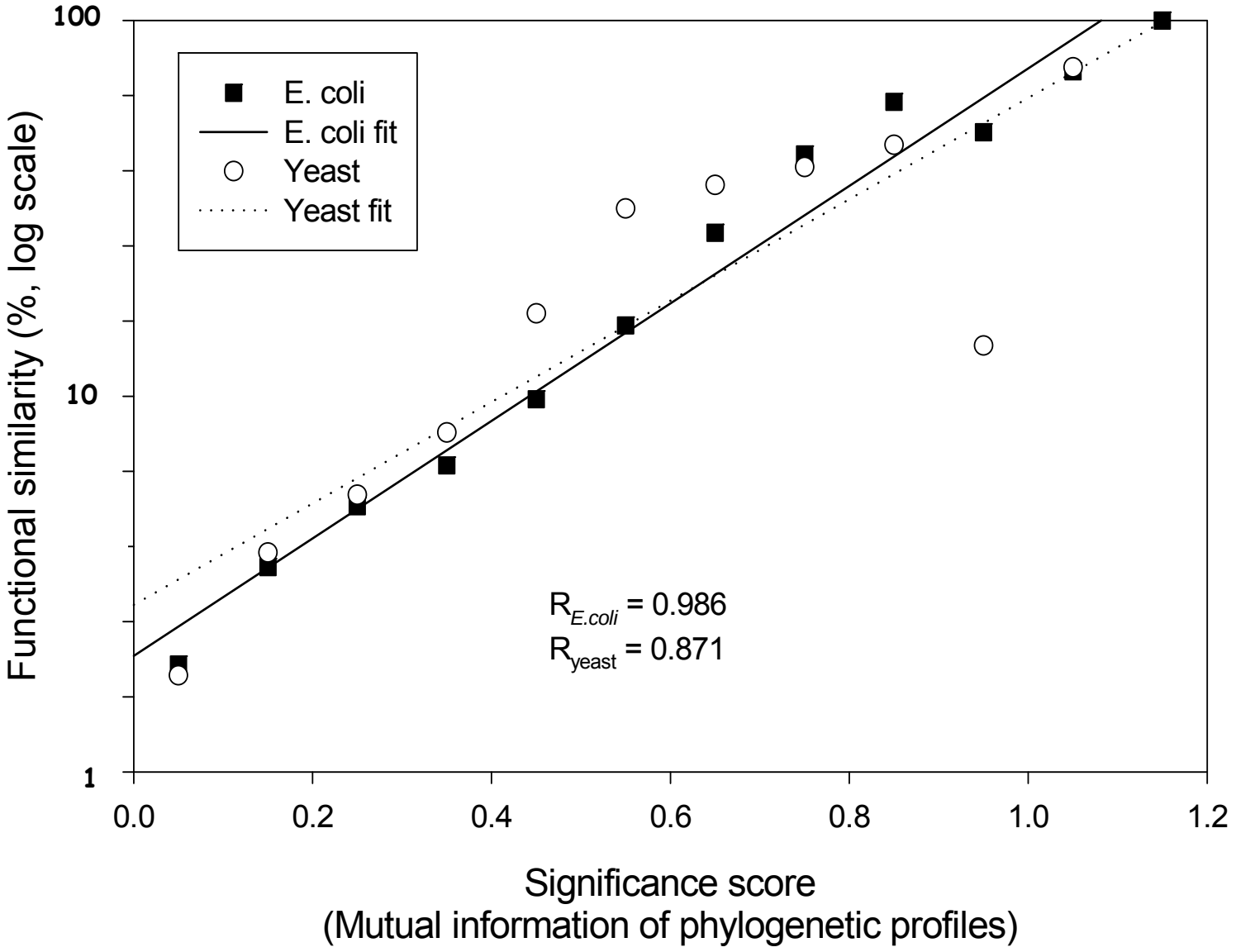
Profile 1: `000101110001100101100111100 11`
Profile 2: `000101111001100101000111100 11`

The distance between both pairs of profiles is **2 bits.**

However, the first pair is much less complex, therefore less informative, than the second pair.

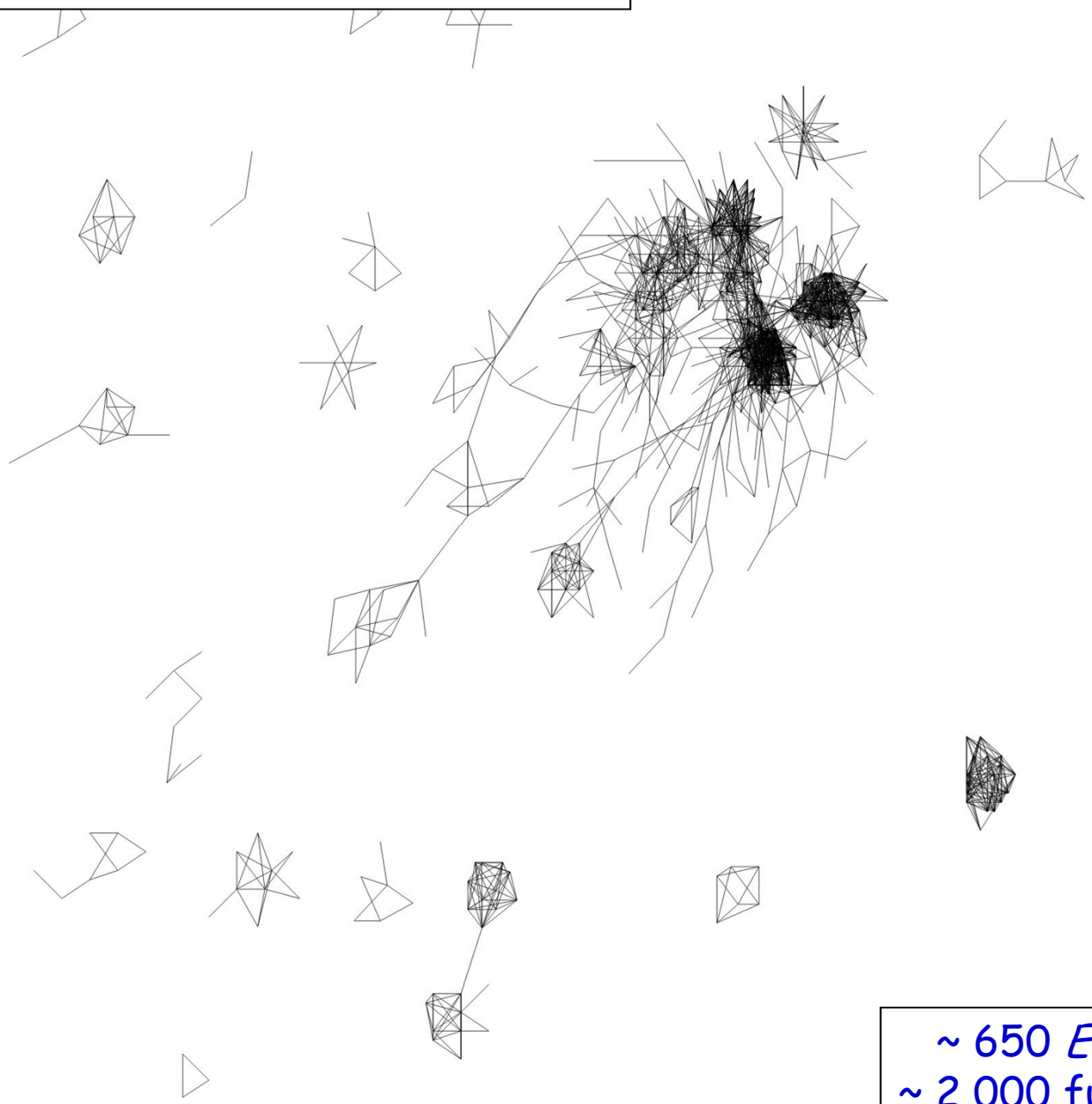
Mutual information = Entropy (Profile 1) + Entropy (Profile 2)
- Relative Entropy (Profile 1 and 2)

Prediction of gene function by phylogenetic profiles

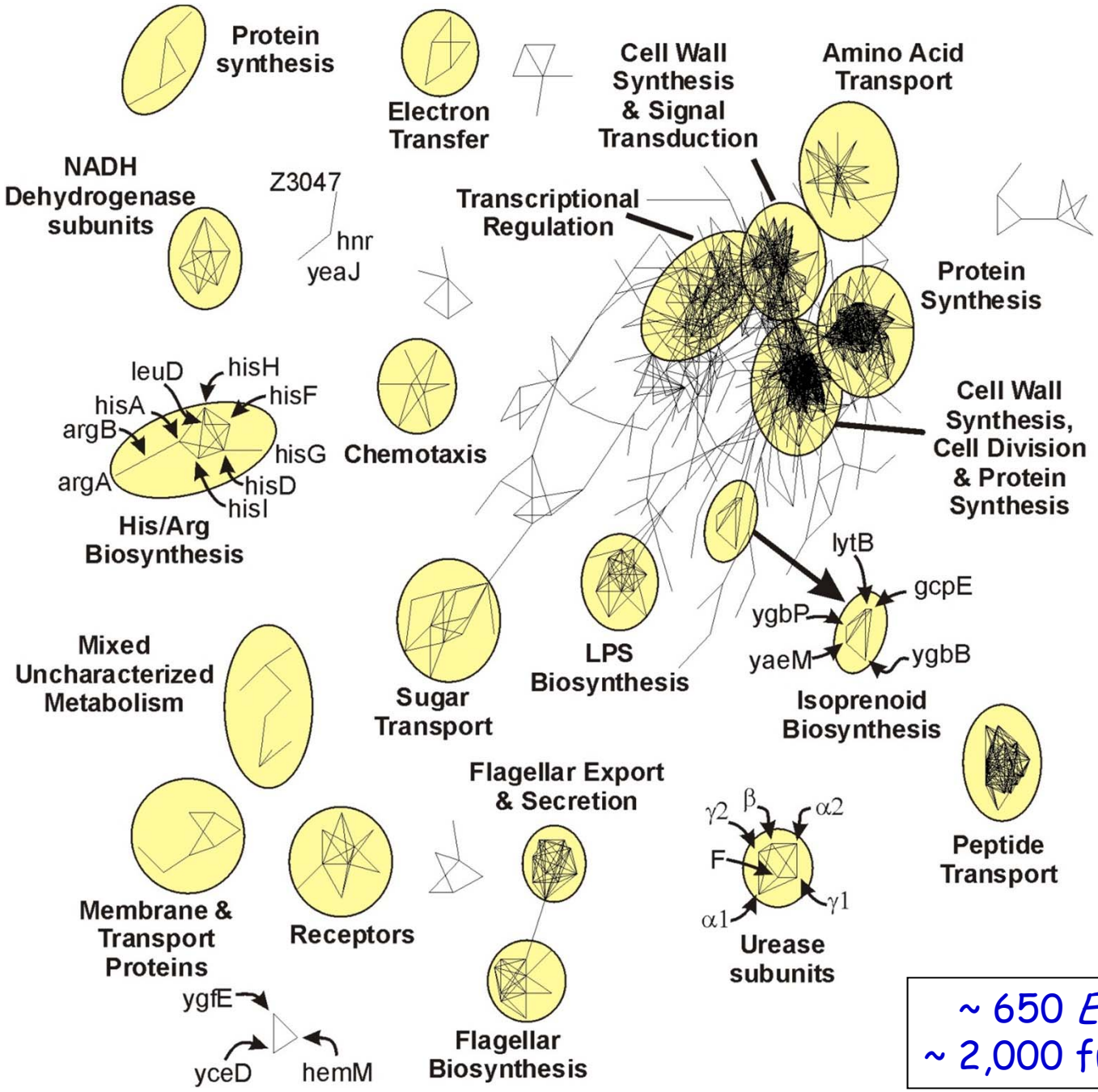


Date et al. NBT 2003

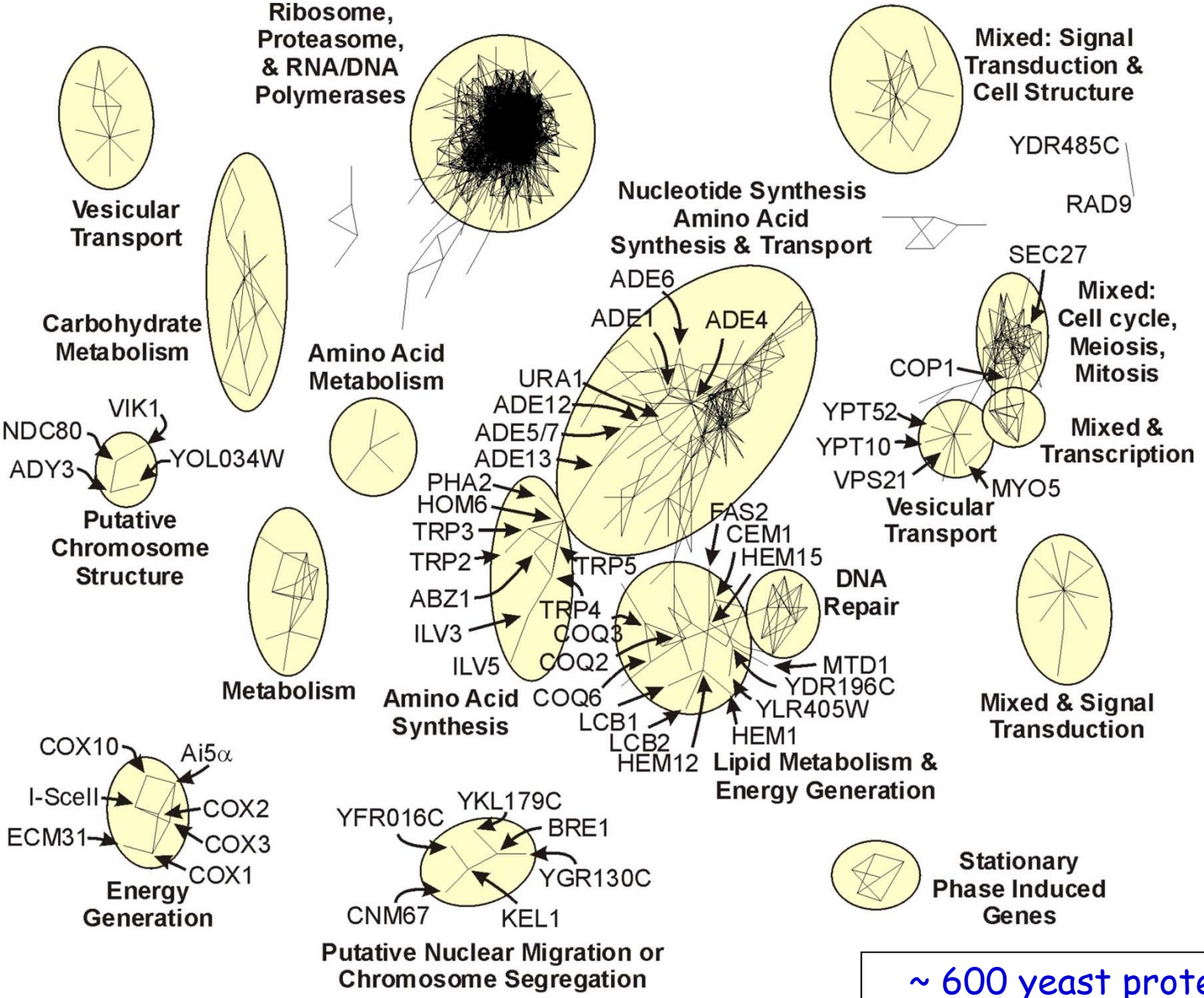
Network by Phylogenetic profile links



~ 650 *E. coli* proteins
~ 2,000 functional links

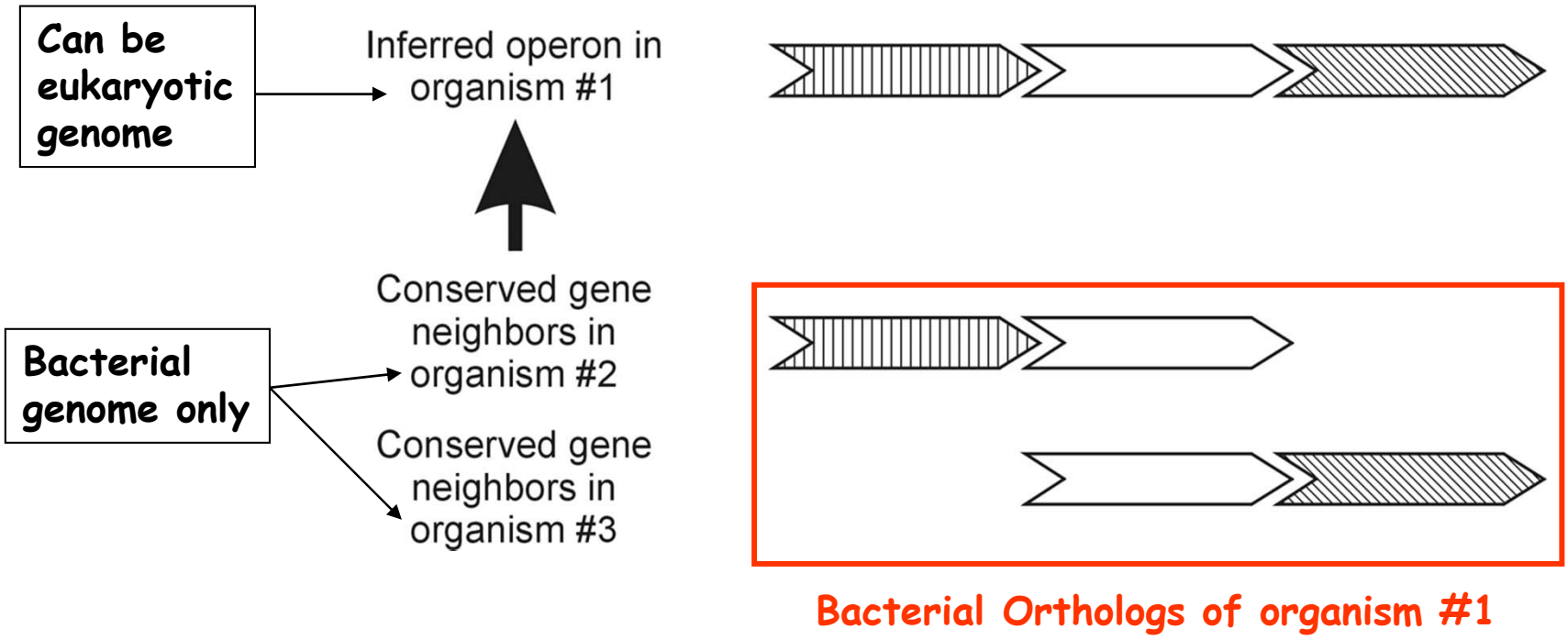


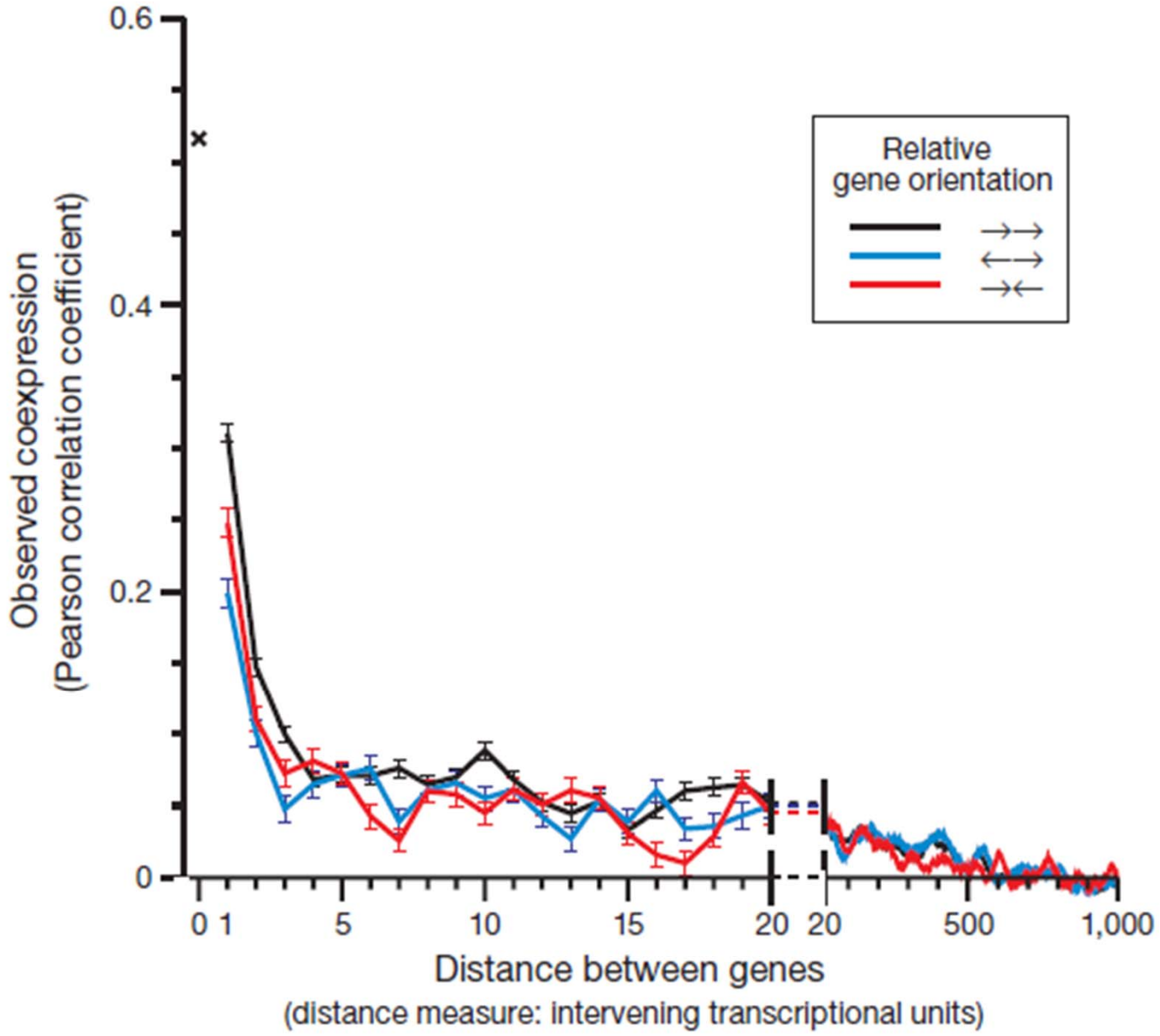
~ 650 *E. coli* proteins
~ 2,000 functional links


Ribosome, Proteasome, & RNA/DNA Polymerases
Mixed: Signal Transduction & Cell Structure
YDR485C
RAD9
Vesicular Transport
Nucleotide Synthesis Amino Acid Synthesis & Transport
SEC27
Mixed: Cell cycle, Meiosis, Mitosis
Carbohydrate Metabolism
Amino Acid Metabolism
ADE6
ADE1
ADE4
URA1
ADE12
ADE5/7
ADE13
COP1
YPT52
YPT10
VPS21
MYO5
Mixed & Transcription
Vesicular Transport
VIK1
NDC80
ADY3
YOL034W
Putative Chromosome Structure
PHA2
HOM6
TRP3
TRP2
TRP5
ABZ1
TRP4
ILV3
COQ3
ILV5
COQ2
COQ6
LCB1
LCB2
HEM12
FAS2
CEM1
HEM15
DNA Repair
MTD1
YDR196C
YLR405W
HEM1
Lipid Metabolism & Energy Generation
Metabolism
Amino Acid Synthesis
Mixed & Signal Transduction
COX10
Ai5α
I-SceII
COX2
ECM31
COX3
COX1
Energy Generation
YKL179C
YFR016C
BRE1
YGR130C
CNM67
KEL1
Putative Nuclear Migration or Chromosome Segregation
Stationary Phase Induced Genes
~ 600 yeast proteins
~ 7,000 functional links

3. Conserved gene neighbors in bacteria

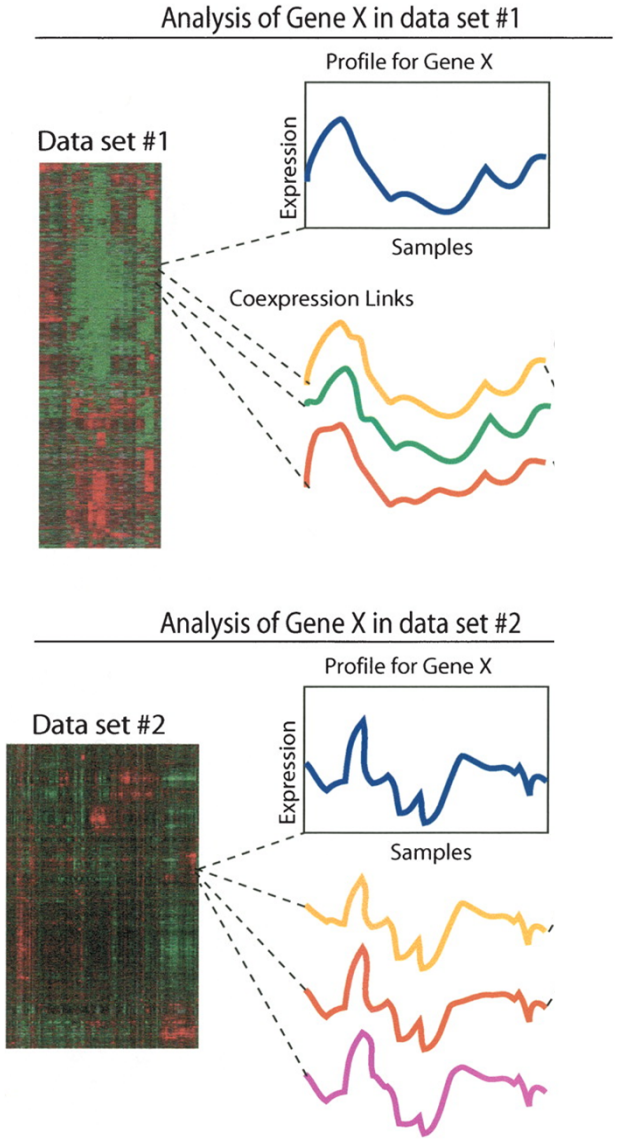
Prokaryotic gene clusters are composed of functionally related genes (Operon).





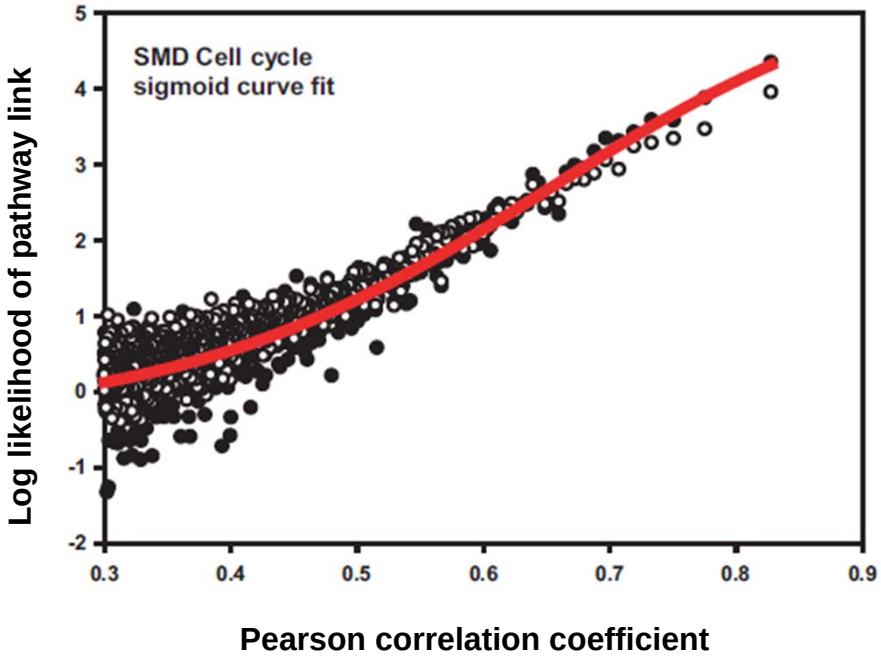
Nature Biotechnology 22:911

Inferring functional interaction from co-expression pattern



Lee H K et al. Genome Res. 2004;14:1085-1094

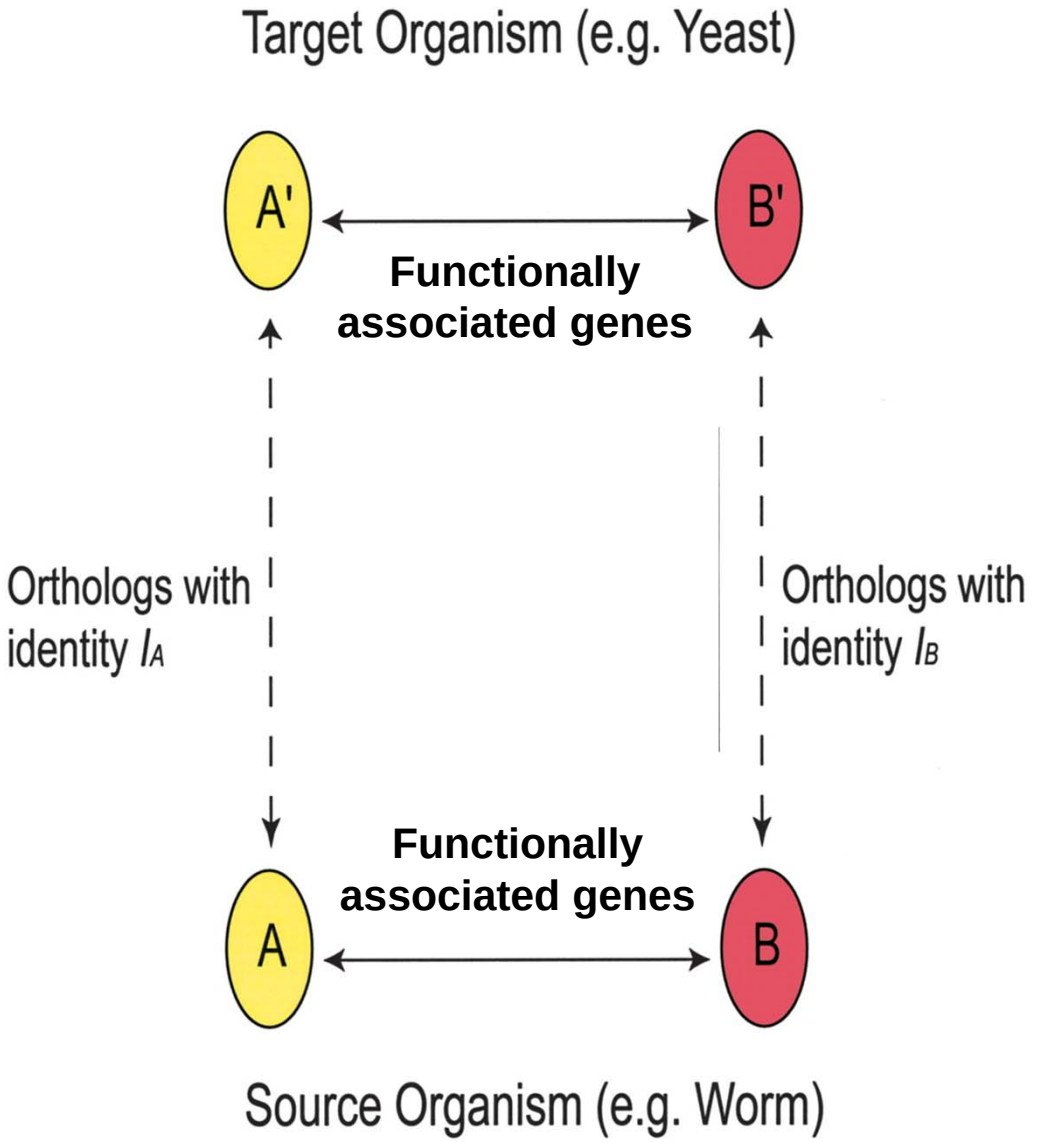
- Assumption: genes for same biological processes are under same transcriptional program.



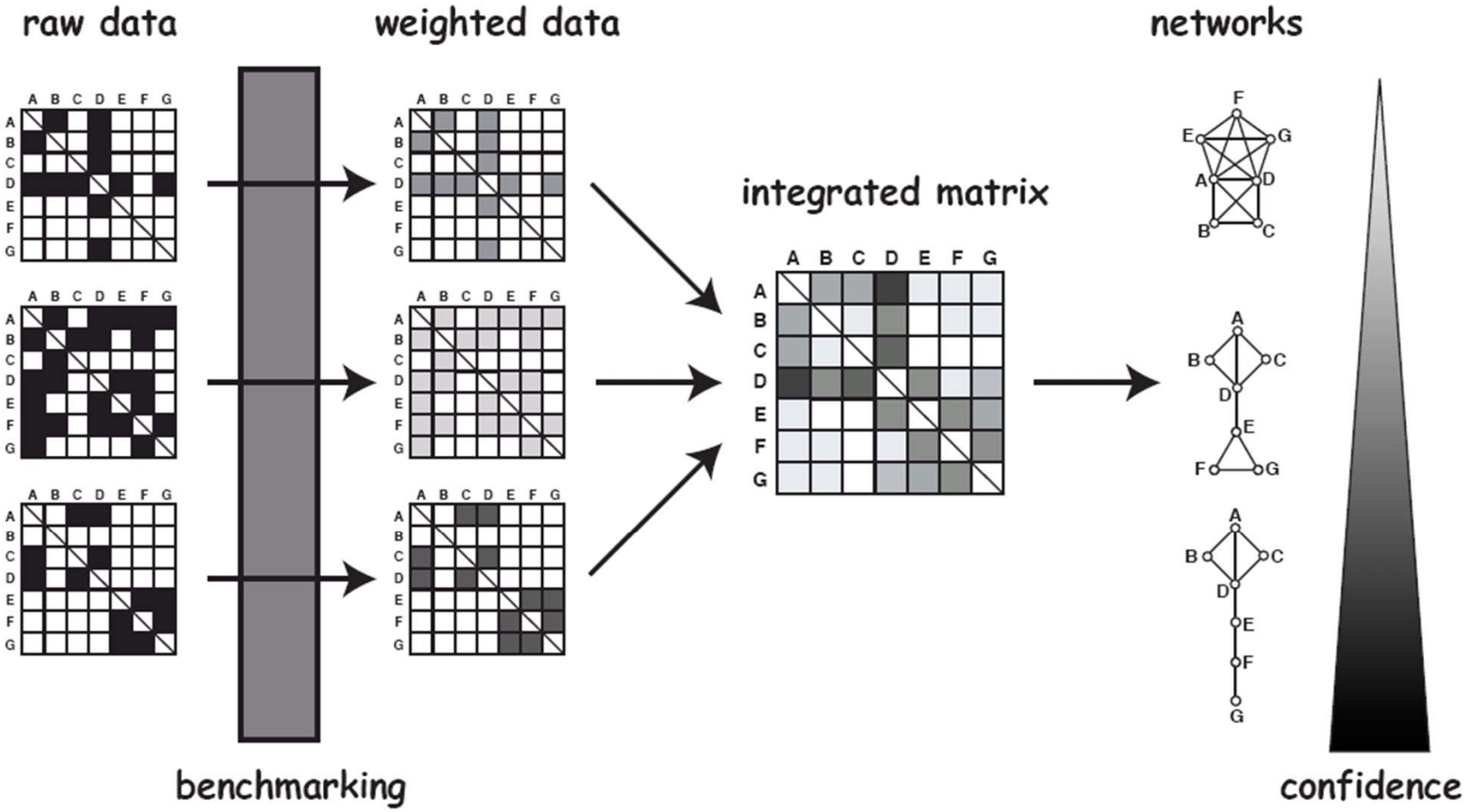
- With massive amount of microarray data, it turned out to be one of most powerful data for pathway modeling.

Associalogs method (*Lee et al. Nature Genetics 2008*)

- Inferring functional links in the target organism by transferring information from other organisms' gene networks.
- Similar to Interolog method, but transfer not only protein-protein interactions but also functional association, which is much more comprehensive.
- Transferred associalogs from animals to plant can predict plant-specific pathways (*Lee et al. Nature Biotechnology 2010*)



Constructing a functional gene network



Fraser & Marcotte, *Nature Genetics* (2004)
Lee *et al*. *Science* (2004)

Standardization of data intrinsic scores by an Unified Score

Based on *Bayesian Likelihood*

Posterior Odds

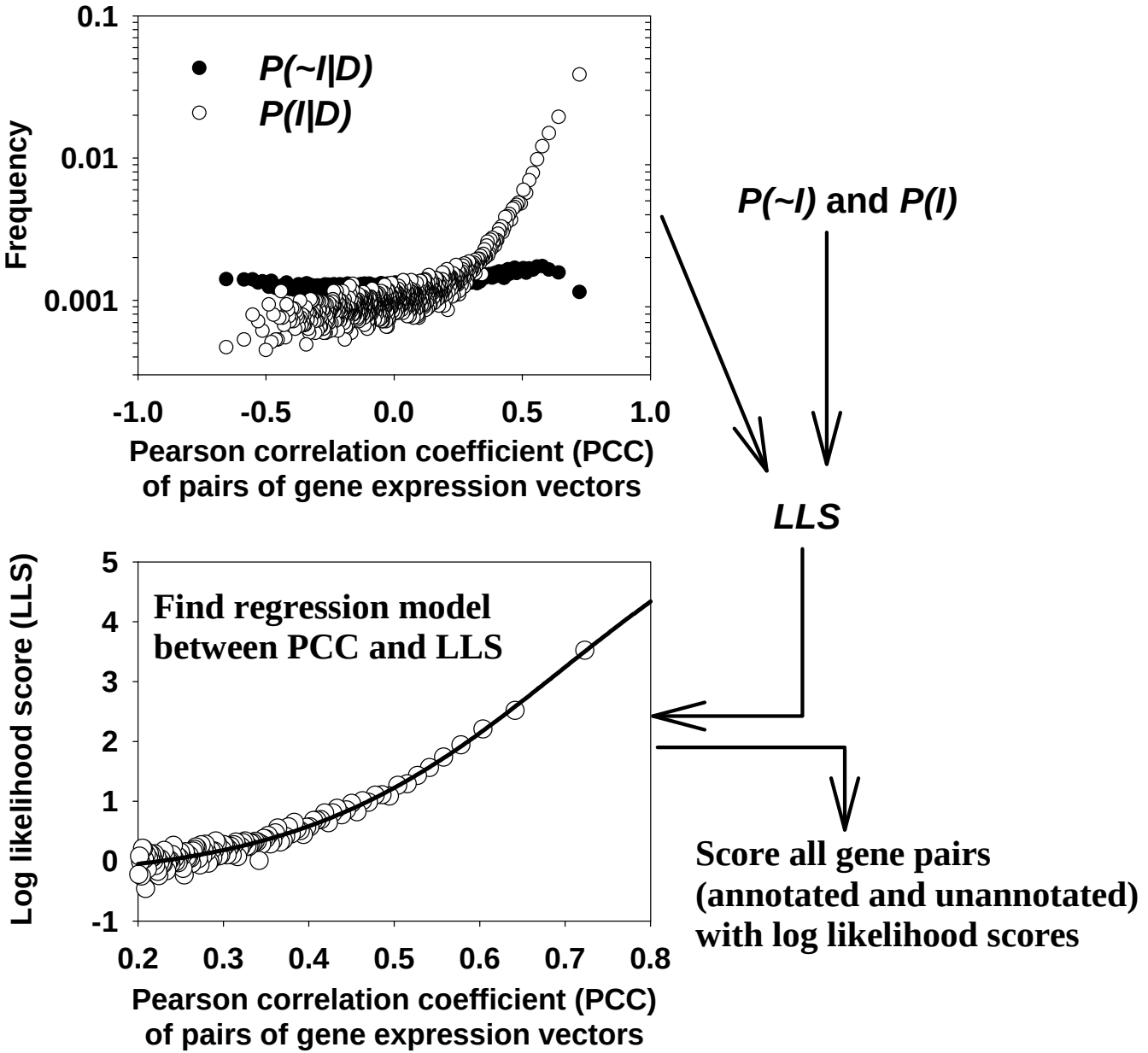
$$\text{Log Likelihood Scores } (LLS) = \ln\left(\frac{P(I|D)\ /\sim P(I|D)}{P(I)\ /\sim P(I)}\right)$$

Prior Odds

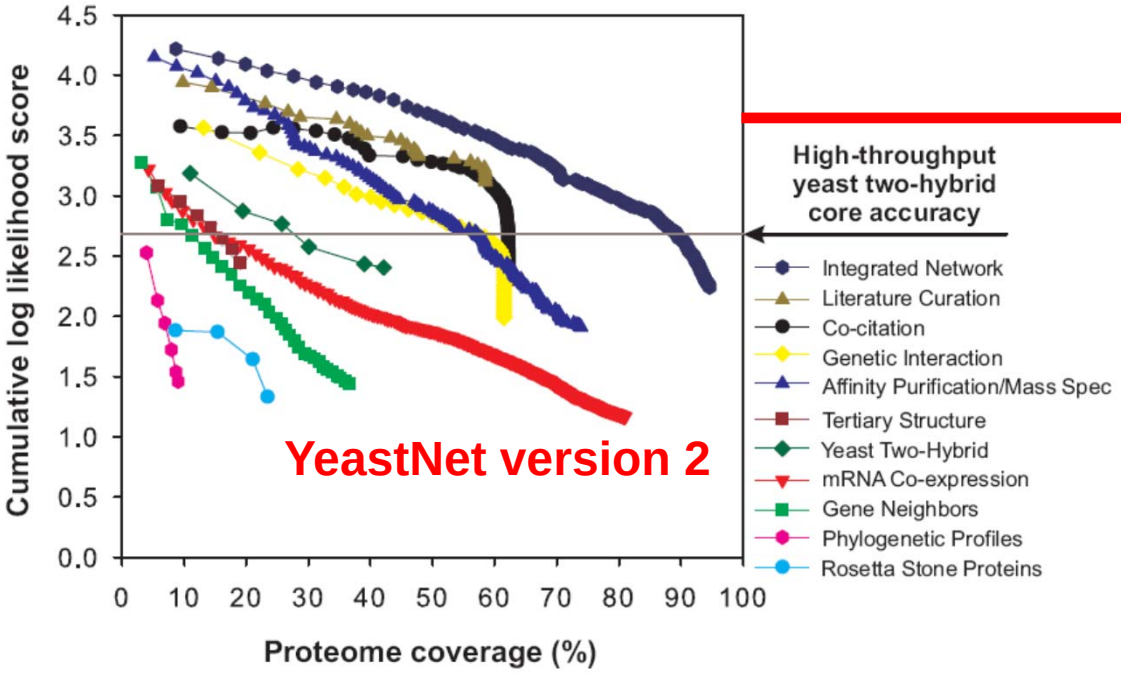
I: two genes interact each other (with at least one shared functional annotation)

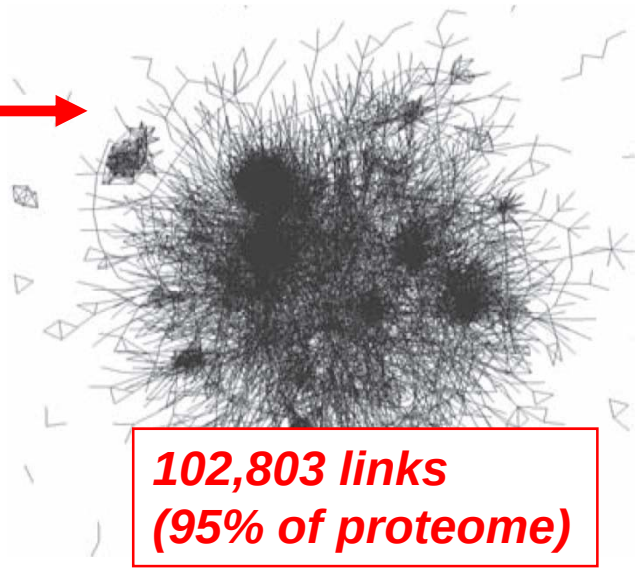
D: given data

If LLS = 0, equal to random chance

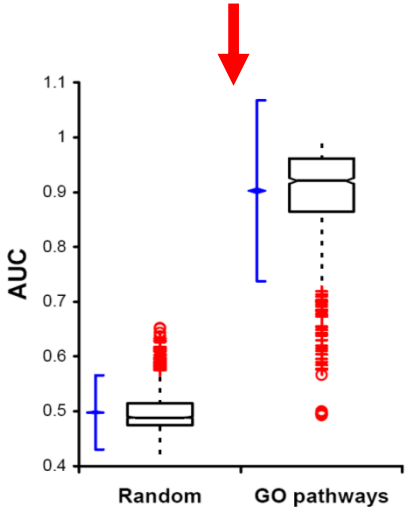

0.1
0.01
0.001
Frequency
P(~I|D)
P(I|D)
-1.0
-0.5
0.0
0.5
1.0
Pearson correlation coefficient (PCC)
of pairs of gene expression vectors
P(~I) and P(I)
LLS
5
4
3
2
1
0
-1
Log likelihood score (LLS)
Find regression model
between PCC and LLS
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Pearson correlation coefficient (PCC)
of pairs of gene expression vectors
Score all gene pairs
(annotated and unannotated)
with log likelihood scores

Integrating diverse functional genomics data produces a larger and more accurate network

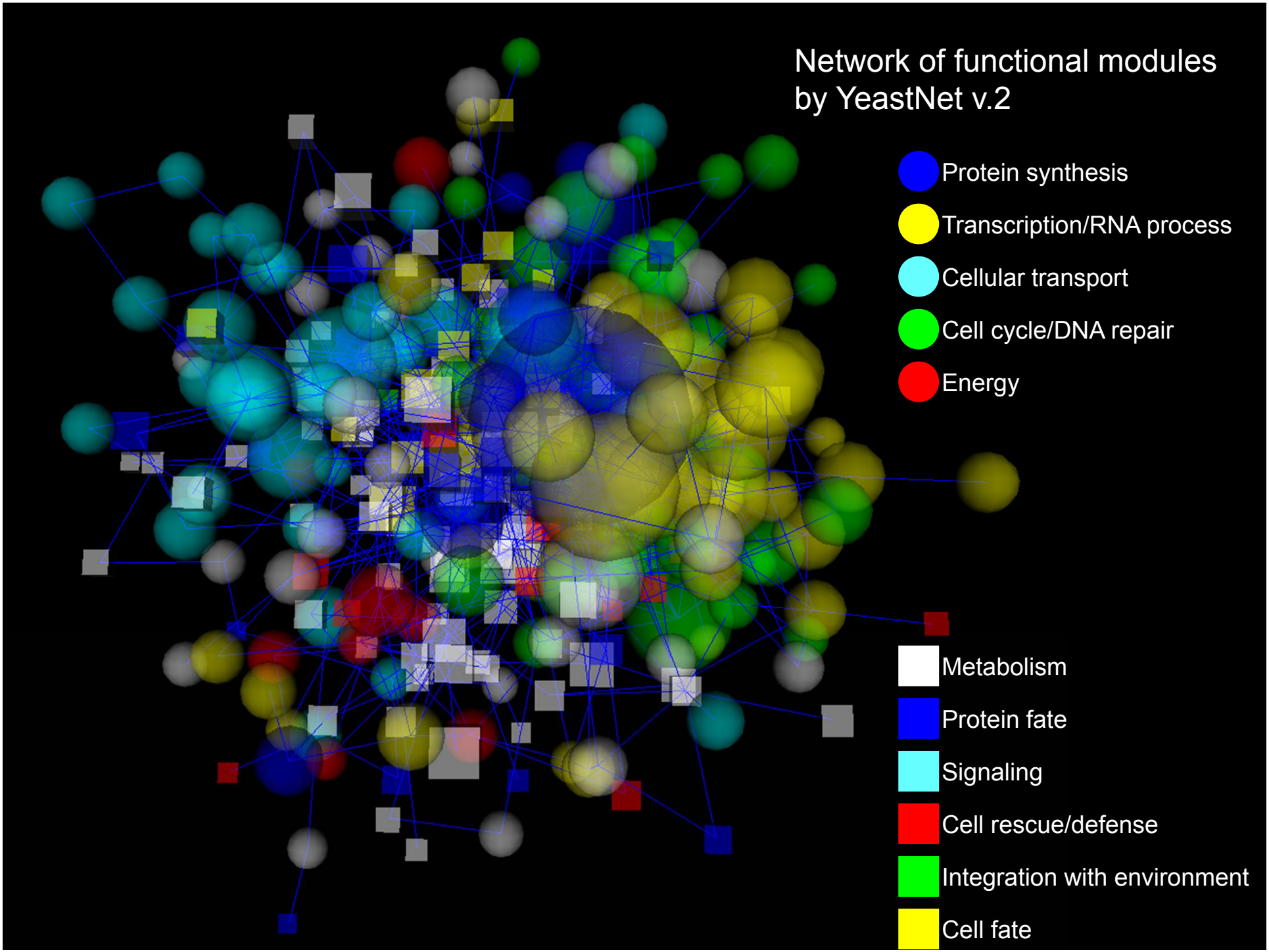




102,803 links (95% of proteome)



Lee *et al*. *Science* (2004)
Lee *et al*.. *PLOS One* (2007)

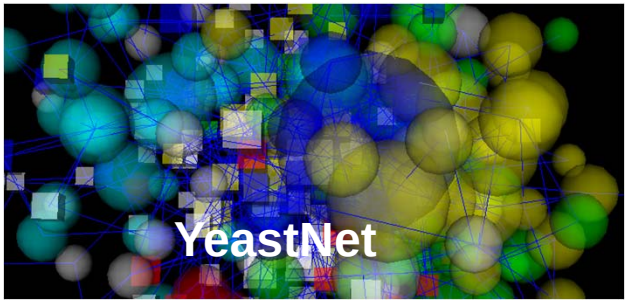

Network of functional modules
by YeastNet v.2
Protein synthesis
Transcription/RNA process
Cellular transport
Cell cycle/DNA repair
Energy
Metabolism
Protein fate
Signaling
Cell rescue/defense
Integration with environment
Cell fate

Network-guided discovery of new ribosomal biogenesis genes

NETBIOLAB.org
YONSEI UNIVERSITY

Large scale data:

Gene expression,
Yeast two-hybrid,
Protein complexes,
in silico prediction,
Synthetic lethals…
(~ 95% proteome)



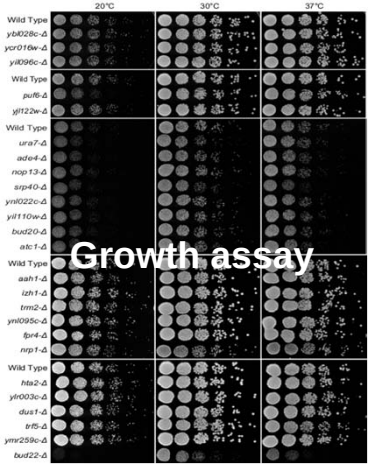
Candidates connected to ~150 known ribosome biogenesis genes

50 Essential genes

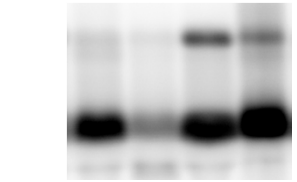
150 Nonessential genes

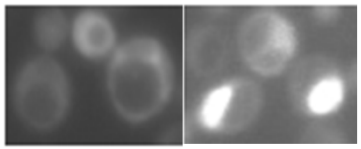
TetO7-promoter alleles
Gal1-promoter alleles

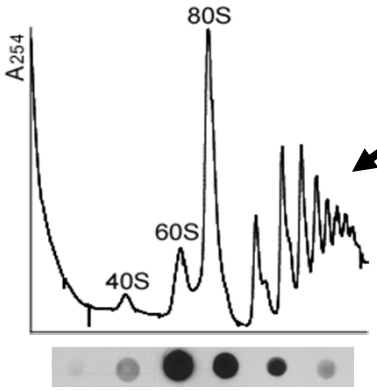
Haploid knockout strains



50
Slow growth



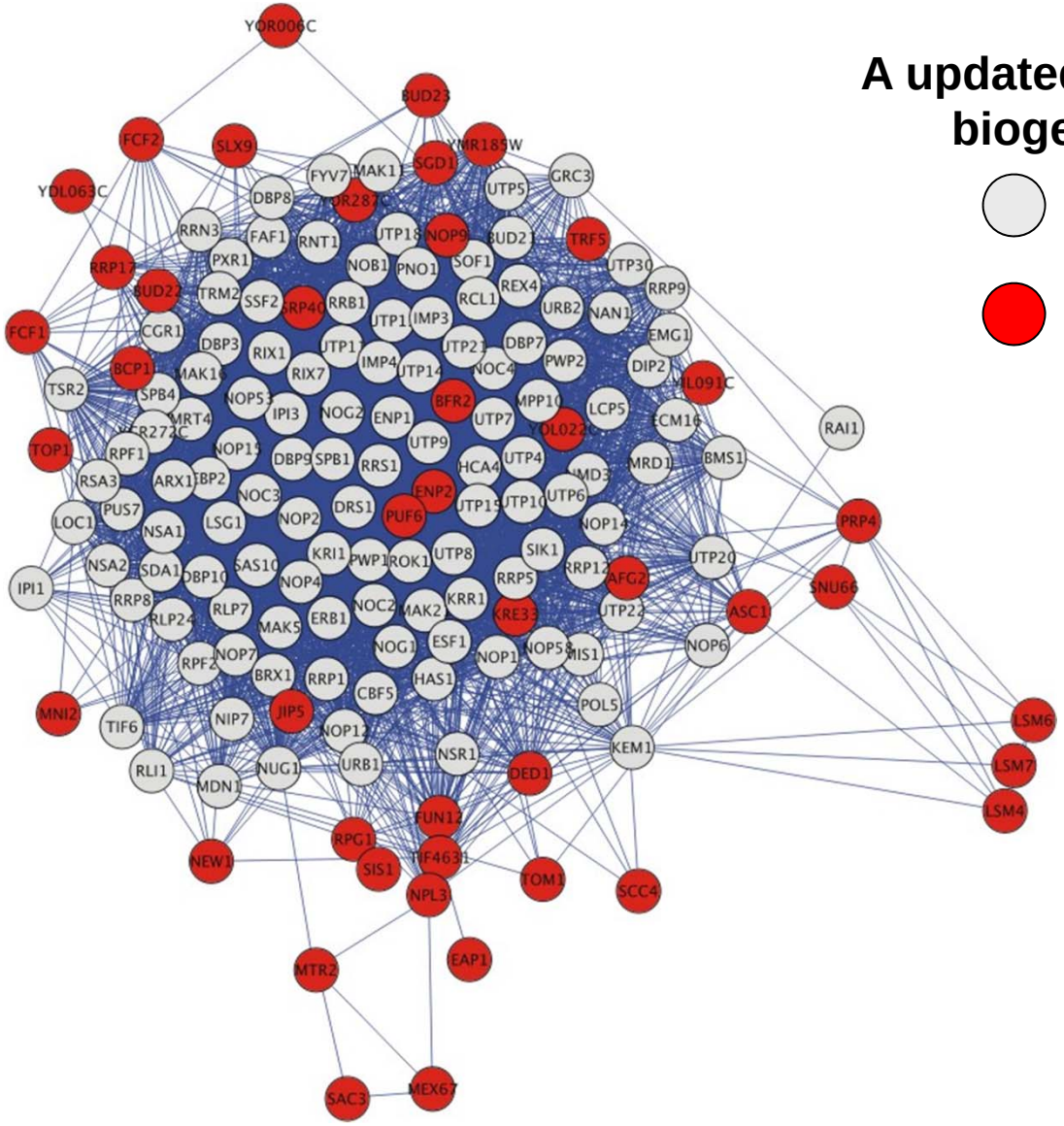




Li *et al. PLoS Biology* (2009)

Experimentally validated ~40 new ribosome biogenesis genes from 100 tested candidates

NETBIOLAB.org
YONSEI UNIVERSITY



A updated network of ribosomal biogenesis (RB) system

- Known RB genes (seeds)
- New RB genes

Li *et al. PLoS Biology* (2009)

"Network-guided focused genetics"

NETBIOLAB.org
YONSEI UNIVERSITY

Key Ideas

- *Guilt-by-association*: connected genes in a network are functionally associated
- *Seed & connections to seed*: select unknown genes connected to known seed genes
- *Focused test*: ONLY genes highly connected to seed
- *Reduces time, labor, and can rescue false negatives*
- *Easier interpretation* : functional clues from network neighbors

Gene network for Systems Genetics

What is Systems Genetics?

"Mendel's genetics has its focus on single-gene traits. However, phenotypic variation, including many human diseases, often results from multiple interactions among numerous genetic and environmental factors. Systems genetics seeks to understand this complexity by integrating the questions and methods of systems biology with those of genetics to solve the fundamental problem of interrelating genotype and phenotype in complex traits and diseases." (Nadeau and Dudley, Science 2011)

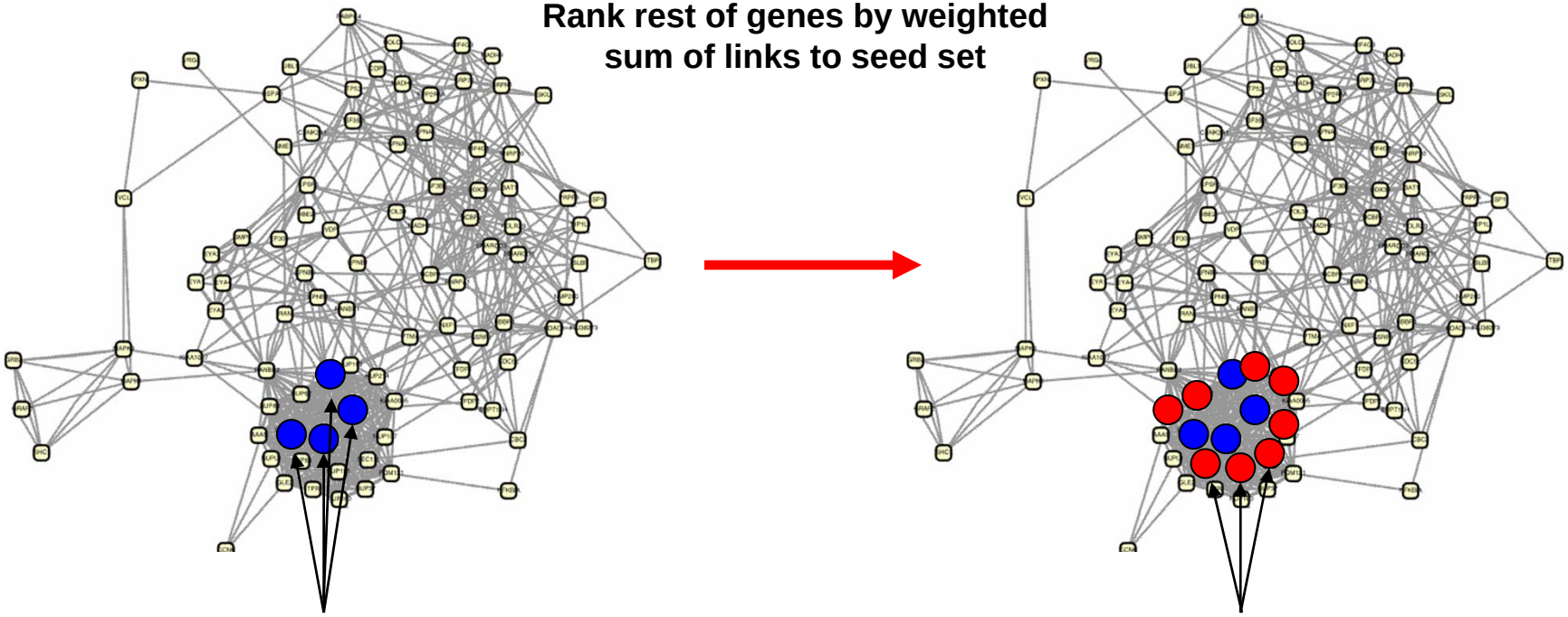
Therefore, Systems Genetics = Systems Biology + Genetics

Here, we have gene networks as a Systems Biology method.

Predict genes associated to a phenotype using seed genes and gene network

NETBIOLAB.org
YONSEI UNIVERSITY

Rank rest of genes by weighted sum of links to seed set



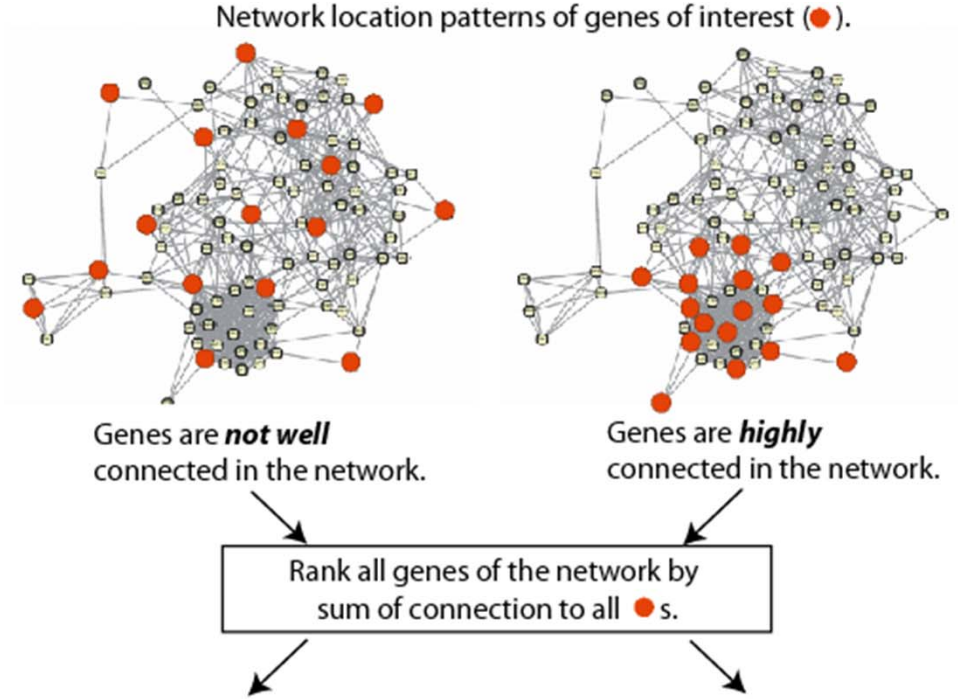
‘Seed’ genes that exhibit a phenotype of interest

New candidates for showing phenotype

McGary *et al. Genome Biology* (2007)

How do we measure predictive power of a network for a given phenotype?

NETBIOLAB.org
YONSEI UNIVERSITY

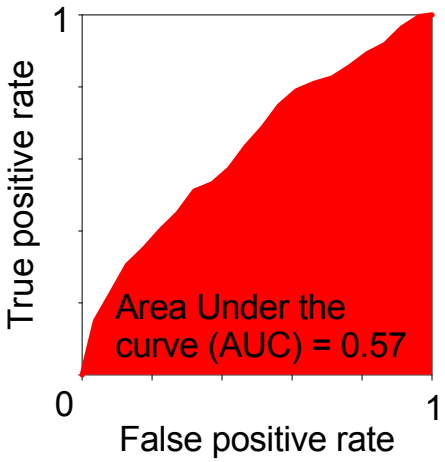


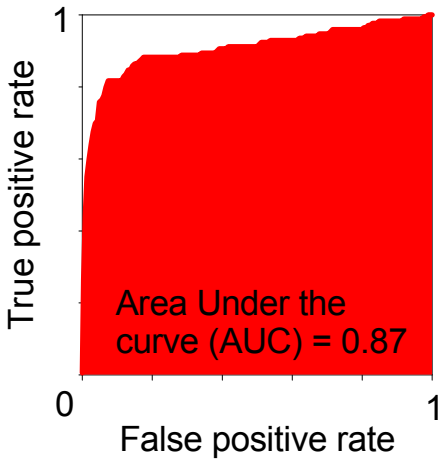
AUC **Predictive power**

1.0 **High**

0.6 **Low**

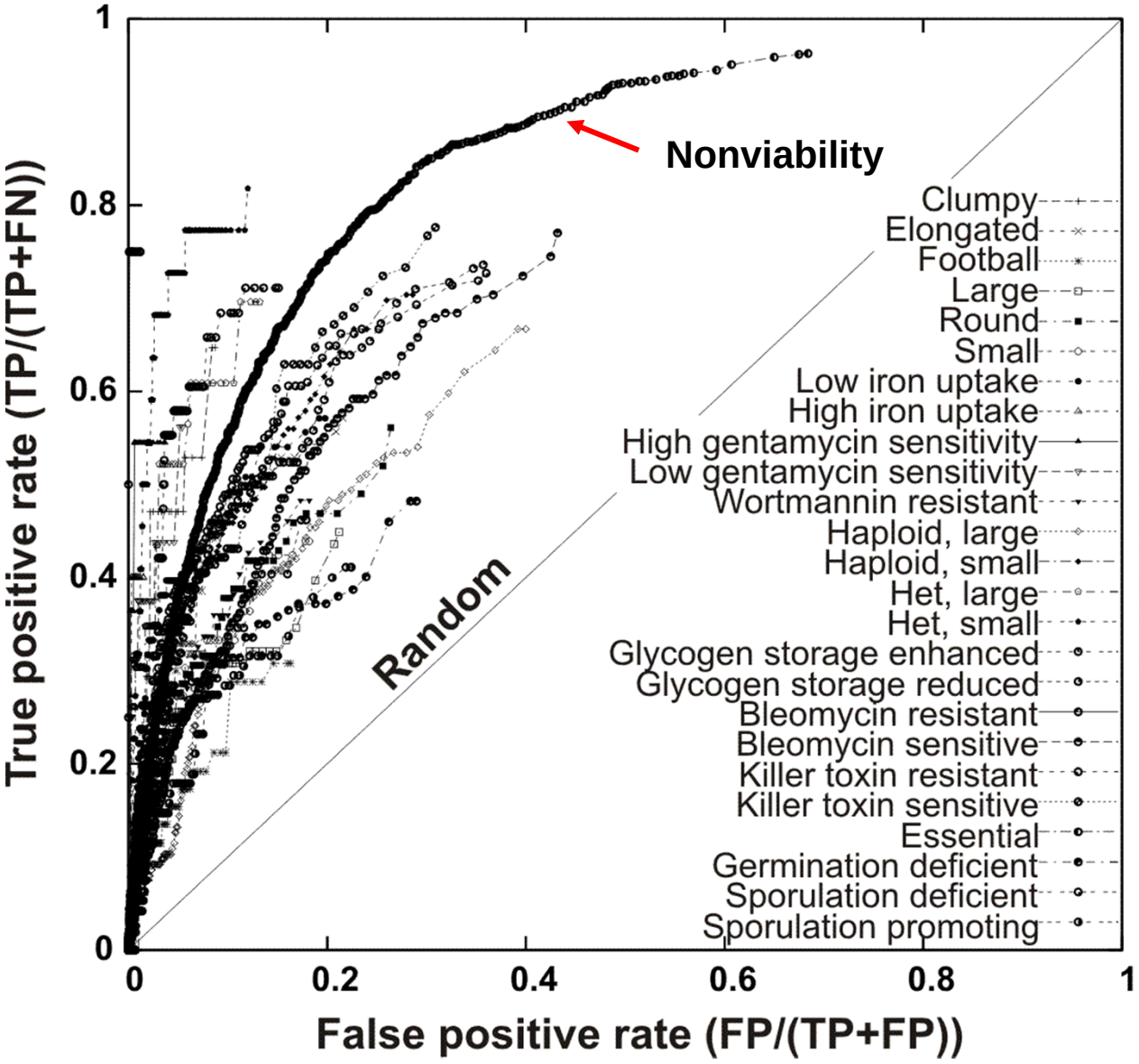
0.5 **= random**



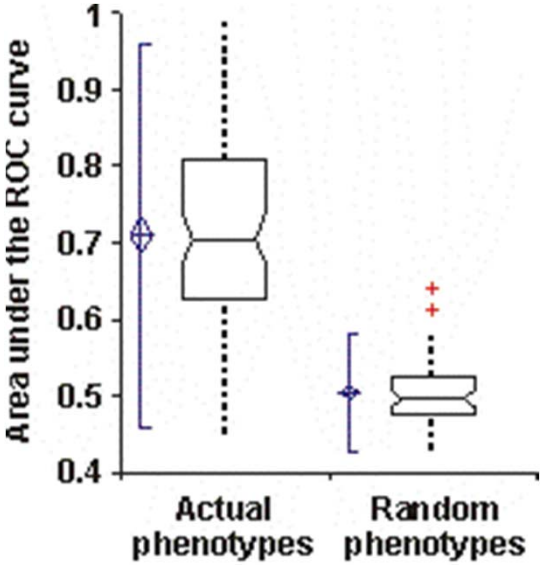


Specific yeast knockout phenotypes can be predicted.

NETBIOLAB.org
YONSEI UNIVERSITY



Tested for 100 KO phenotypes from Literature



McGary *et al.*
Genome Biology (2007)

Predictive network for animals and plants?

NETBIOLAB.org
YONSEI UNIVERSITY

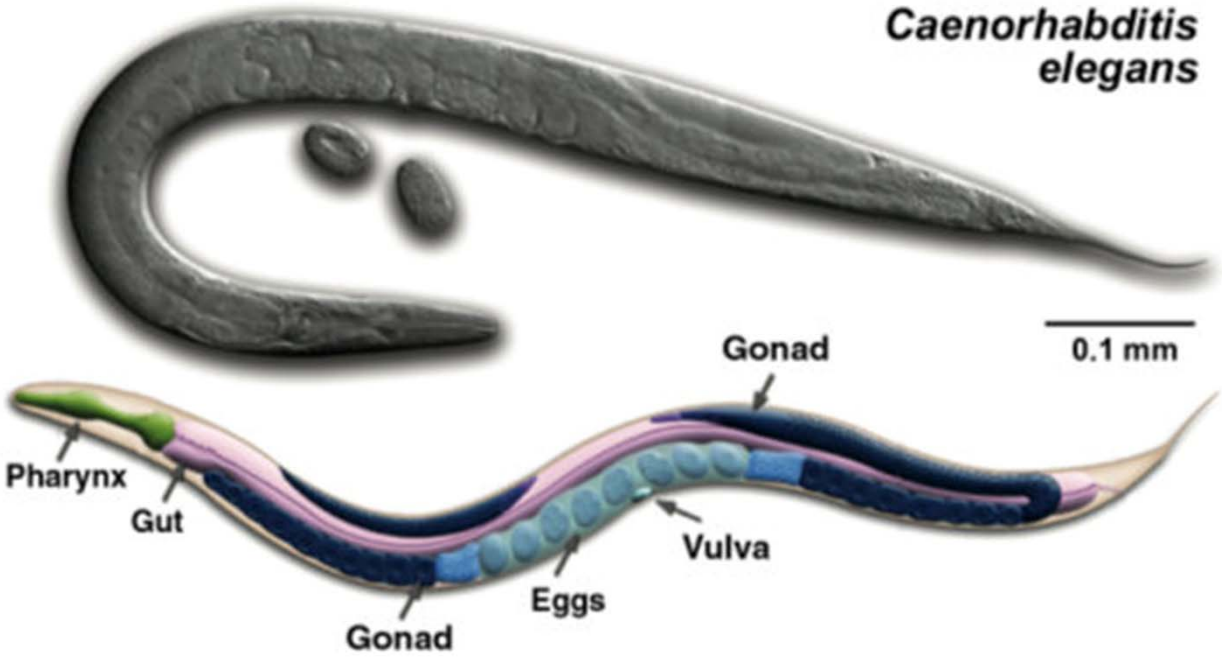
Expected to be very difficult. Why?

- **Much larger genome**
  - **Combinatorial Explosion of the number of gene pairs**
  - **~18M tests for yeast vs. ~200M tests for human**
  - **Do we have ~10fold more data?**
- **Multiple cell/tissue types**
  - **single cell for yeast vs. 100 trillion (M of M) cells for human**
  - **~200 known distinct cell types for human**
  - **But we have a single integrated network model for all cell types**
  - **Many raw data are not cell/tissue type specific (e.g., Y2H)**
- **Would animal or plant networks be equally predictive?**

Tested in C. elegans (Worm). Why?

NETBIOLAB.org
YONSEI UNIVERSITY

- **Only 959 cells for a whole body (for adult hermaphrodite)**
- **High-throughput gene silencing by bacterial feeding RNA interference (RNAi)**



WormNet: *C. elegans* probabilistic gene network

NETBIOLAB.org
YONSEI UNIVERSITY

Protein physical interactions
- yeast two-hybrid interactions from fly, worm, human
- yeast affinity purification/mass spectrometry interactions

Co-expression across 855 *C. elegans* DNA microarray experiments

Linkages from computational genetics
- coinheritance of genes (phylogenetic profiles)
- location of homologs in the same bacterial operons (Gene neighbors)

Genetic interactions
- ~4,000 C. elegans genetic interactions

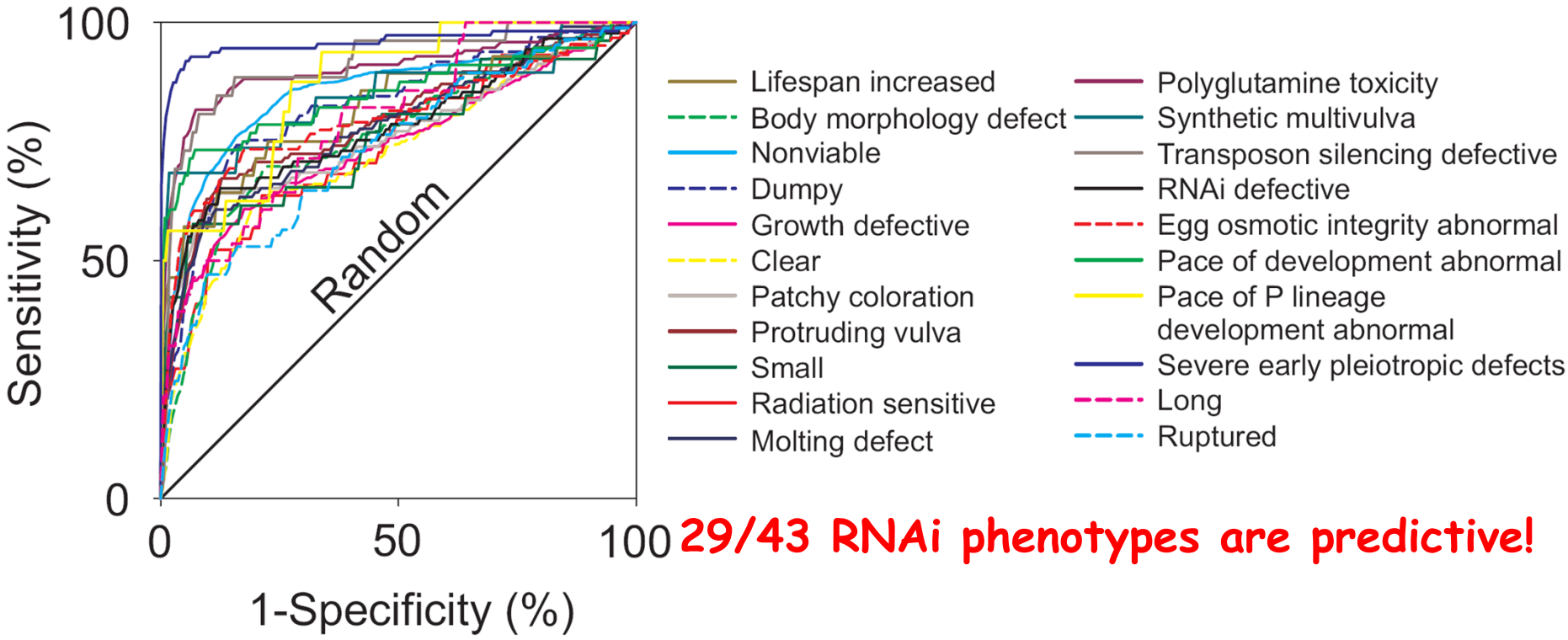
Yeast probabilistic gene network (YeastNet v.2, *PLoS One* 2007)

Literature mining for *C. elegans* interactions

Analyzed >20 millions experimental observations

Version 2: 999,367 links / 15,139 genes (~75% of proteome)

Specific C. elegans RNAi phenotypes can be predicted.



29/43 RNAi phenotypes are predictive!

Functionally linked genes in an animal network therefore also tend to exhibit related loss-of-function phenotypes

Lee, Lehner *et al. Nature Genetics* (2008)

Massive amount of Genetics data are publicly available.

Do we understand inheritance of complex phenotypes now?

NETBIOLAB.org
YONSEI UNIVERSITY

By 04/13/2011, 862 GWAS papers published, 4306 trait-associated SNPs (A catalog of GWAS)

How much trait can we explain by the trait-associated SNPs?

Phenotype	Number of GWAS loci	Proportion of heritability explained (%)*
Type 1 diabetes	41	~60
Fetal haemoglobin levels	3	~50
Macular degeneration	3	~50
Type 2 diabetes	39	20–25
Crohn's disease	71	20–25
LDL and HDL levels	95	20–25
Height	180	~12

(Nature 470:187)

Where the <u>missing inheritance of complex traits</u> come from?

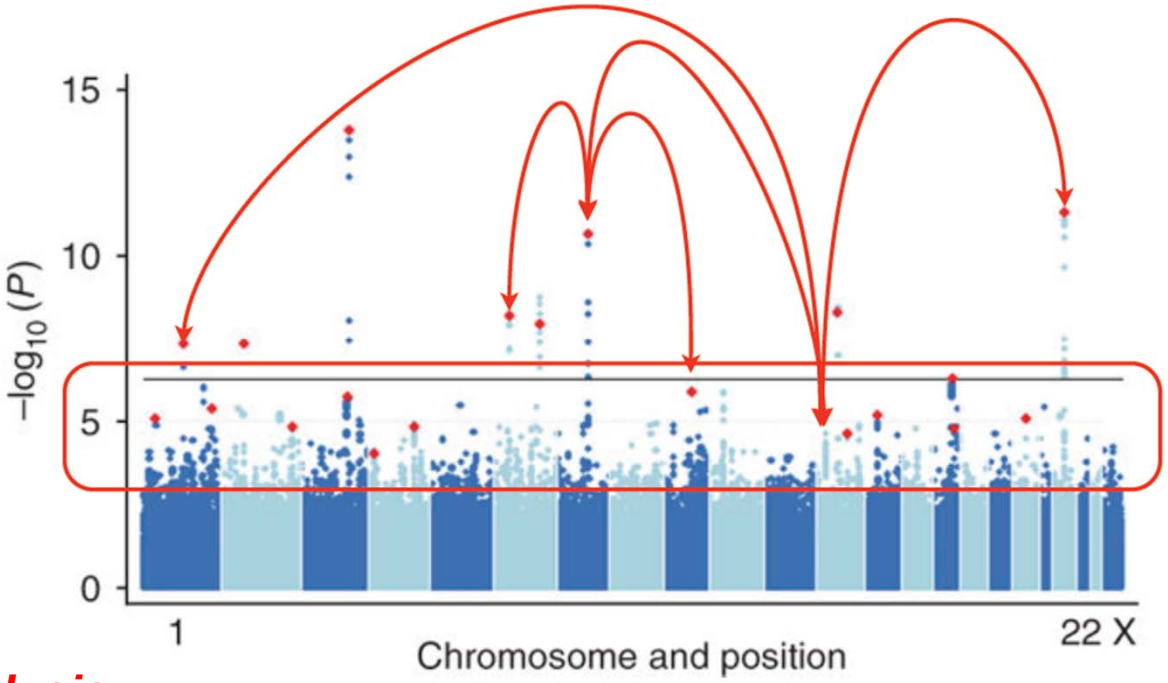
1. Lack of statistical power to detect weak genetic penetration per each SNP.
2. Lack of considering polygenic effect for traits (Epistatic interactions).

Thus, the next challenge in genetics of complex traits is (1) improving statistical power to identify more trait-associated genes, (2) mapping epistatic interactions. But How?

Boosting GWAS signal by HumanNet

NETBIOLAB.org
YONSEI UNIVERSITY

- **It is very hard to pass the statistical test with *Bonferroni correction*.**
- **Many minor contributors are below the threshold.**
- **Can we boost GWAS signal by pathway relations?**

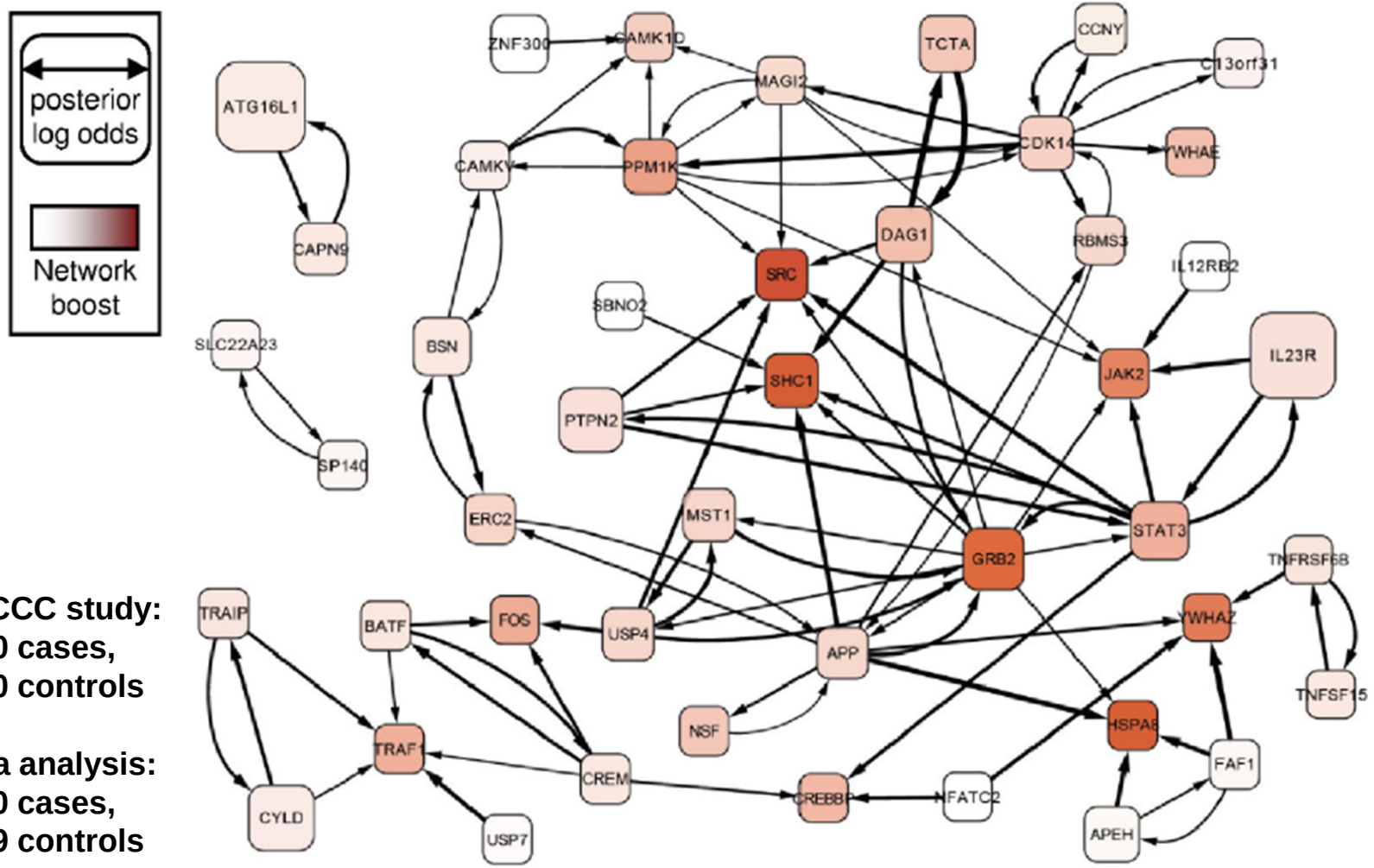


Analysis:

- **Boost original *p-values* from WTCCC(Welcome Trust Case Control Consortium) 2007 study**
- **Validate boosted genes by newly identified genes by meta analysis with larger sample (Barrett et al. 2008, Zeggini et al. 2008)**

Lee, Blom, et al. Genome Research 2011

Validation by meta GWAS data: Crohn's disease



WTCCC study:
2000 cases,
3000 controls

Meta analysis:
3230 cases,
4829 controls

- **Original study identified IL23R, PTPN2, ATG16L1**
- **STAT3, JAK2, GRB2, SHC1 are strongly boosted.**

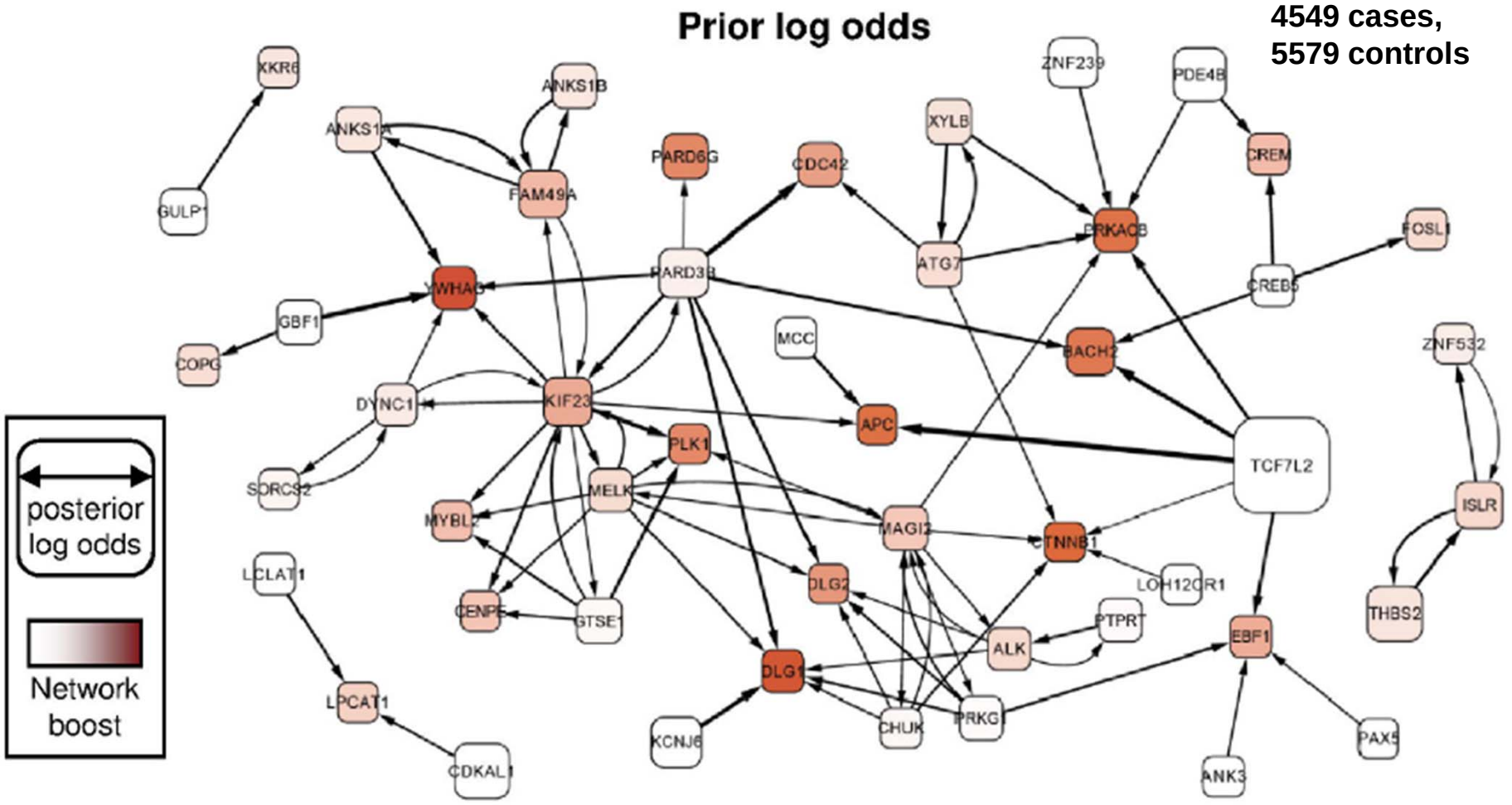
Validation by meta GWAS data: type 2 diabetes

NETBIOLAB.org
YONSEI UNIVERSITY

- **Original GWAS identified TCF7L2**
- **BACH2, PARD6G, PARD3B, CDC42 are strongly boosted.**

WTCCC study:
2000 cases,
3000 controls

Meta analysis:
4549 cases,
5579 controls

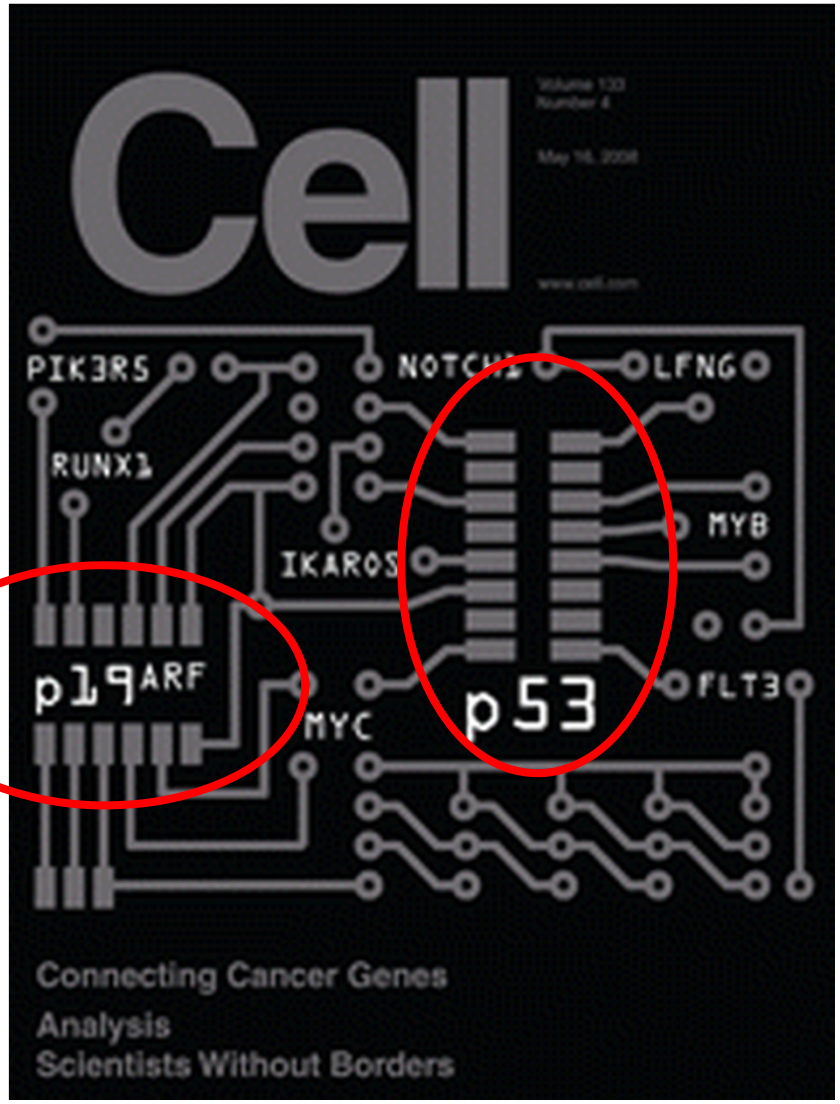


Complex diseases are due to complex networks of disease related genes.

NETBIOLAB.org
YONSEI UNIVERSITY

“Genetic interactions to two major tumor suppressor genes, p53 and p19ARF construct the network of genes that are likely to cooperate in tumorigenesis.”

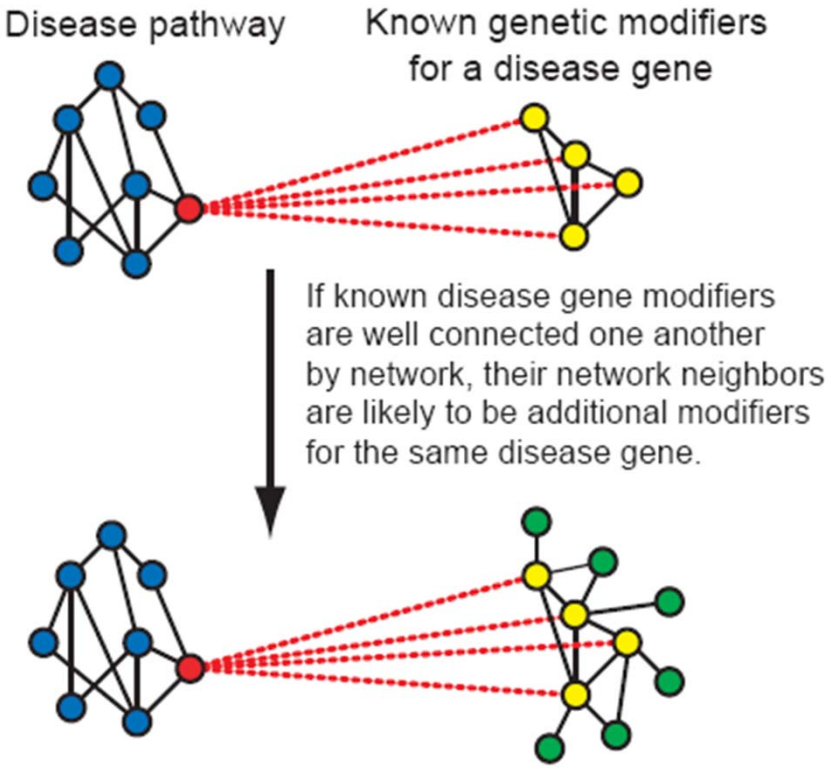
***Cell*, May 16, 2008**

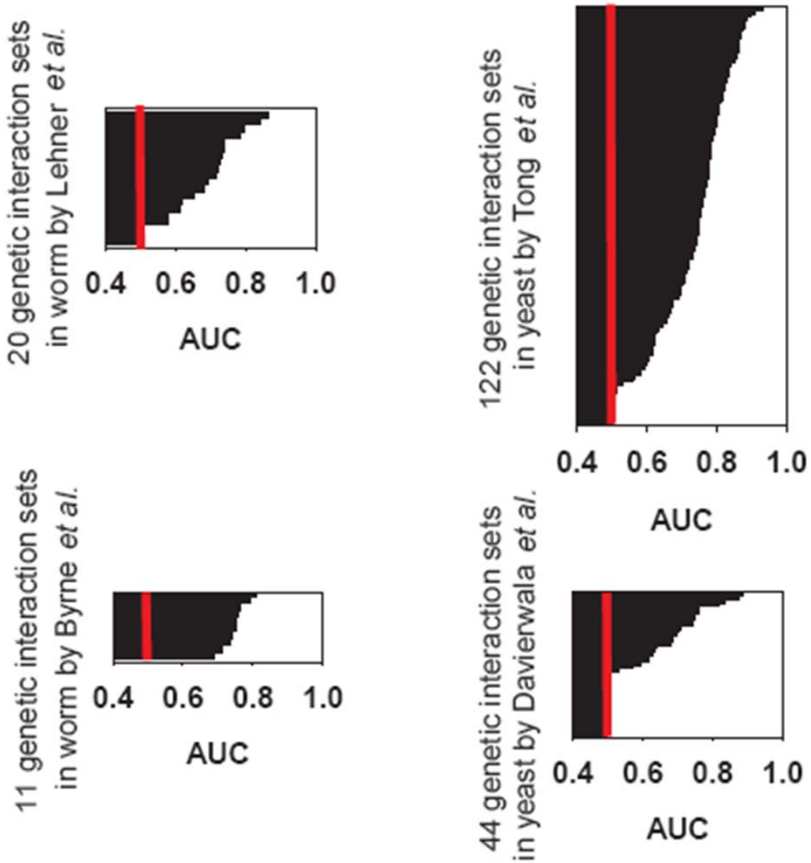


Modifiers of the same gene are interconnected.

NETBIOLAB.org
YONSEI UNIVERSITY

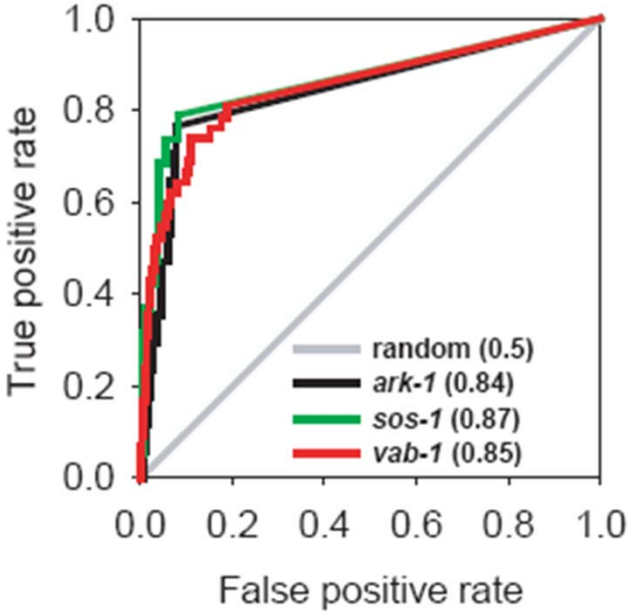
Hypothesis:
genetic modifiers for the same disease gene tend to participate in the same pathway.





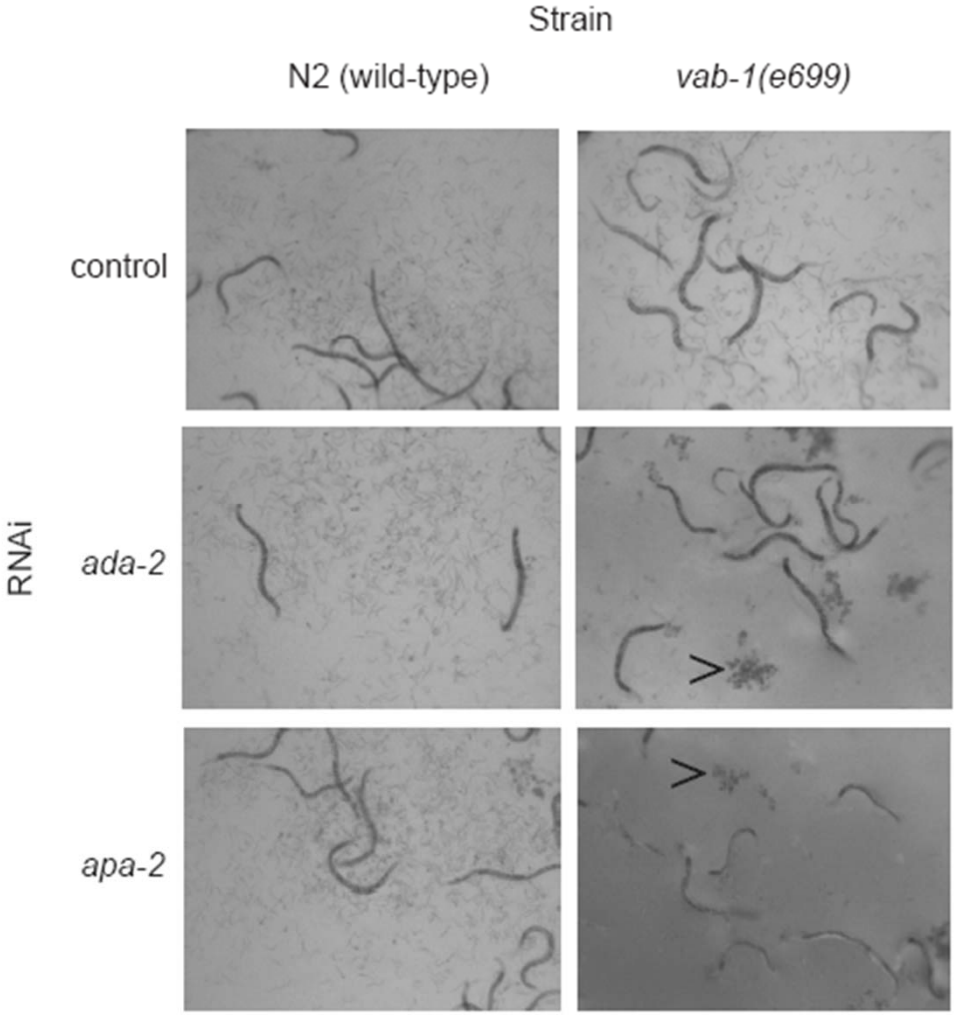
Lee *et al. Genome Research* (2010)

Identification of genetic modulators for three disease-related genes in worm



Tested ~90 additional candidates for each.

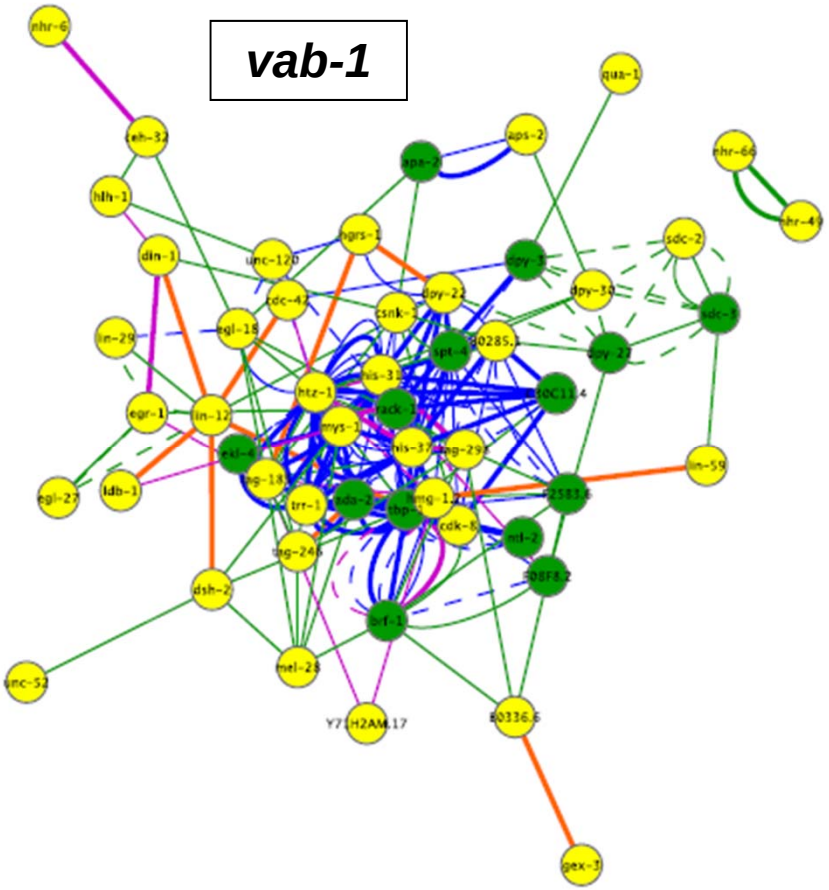
Compared to original semi-random screen, it shows >7 fold enrichment.

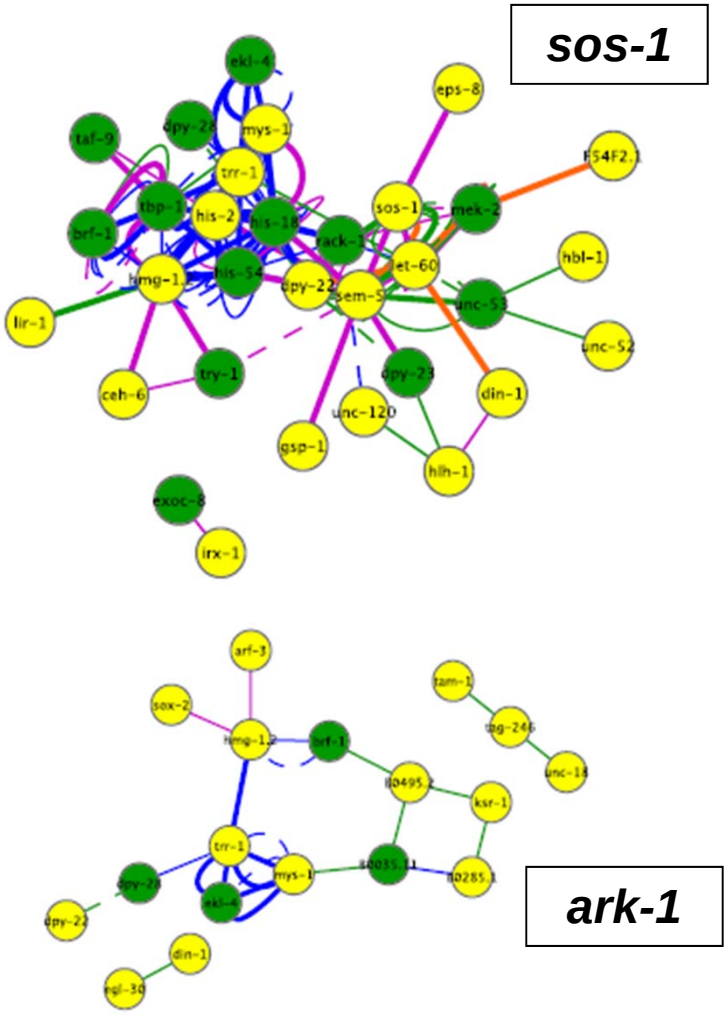


Lee *et al. Genome Research* (2010)

Pathways modulating disease genes in worm

NETBIOLAB.org
YONSEI UNIVERSITY





Lee *et al. Genome Research* (2010)

Acknowledgements

Science (2004)
PLOS One (2007)

Nature Genetics (2008), *Genome Research* (2010)

WormNet
Probabilistic Functional Gene Network of *Caenorhabditis elegans*

Genome Research (2011)

HumanNet
Probabilistic Functional Gene Network of *Homo sapiens*

Nature Biotechnology (2010), *Nature Protocols* (2011)

AraNet
Probabilistic Functional Gene Network of *Arabidopsis thaliana*

Submitted

RiceNet
Probabilistic Functional Gene Network of *Oryza sativa*

Collaborators

Edward Marcotte,	University of Texas
Arlen Johnson,	University of Texas
Andrew Fraser,	University of Toronto
Ben Lehner,	EMBL-CRG, Spain
Matthew Hurles,	Sanger Institute, UK
Sangjun Ha,	Yonsei University
Dongryul Lee,	Cha Medical school
Sue Rhee,	Carnegie Institution
Pamela Ronald,	UC Davis
Philip Benfey,	Duke University
Yongsun Bahn,	Yonsei University
Sangsun Yoon,	Yonsei Medical school

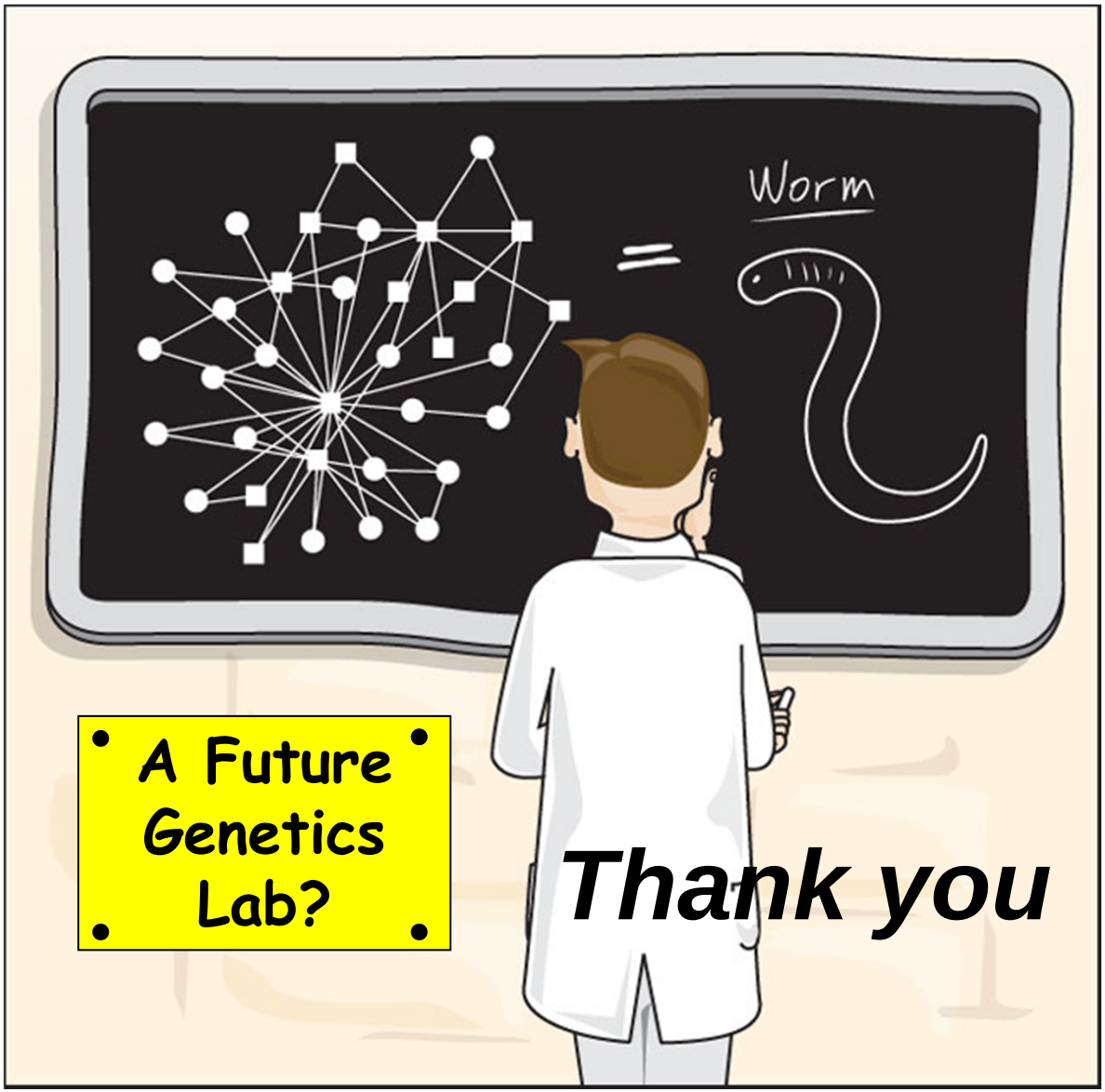
Funding Agencies

Korean National Research Foundation

POSCO TJ Park Science Fellowship

Gene Network for Genetics Research!

NETBIOLAB.org
YONSEI UNIVERSITY



Adapted from Nature Methods

March 2008