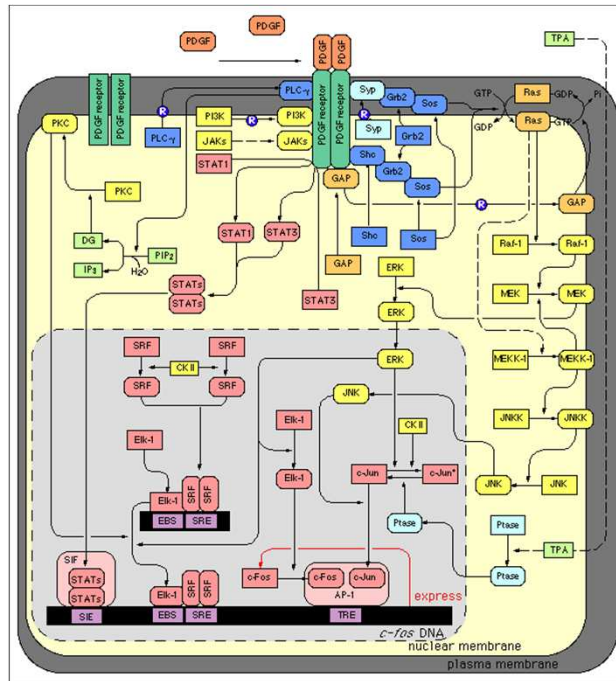


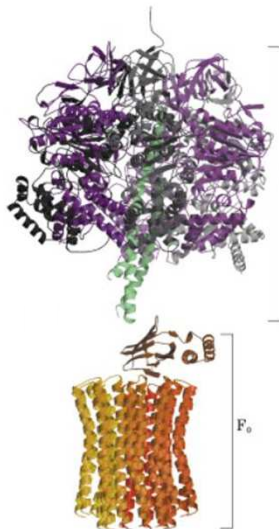
A typical genetic network



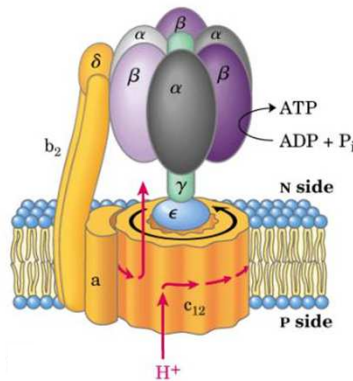
Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Contacts between proteins define protein interaction networks

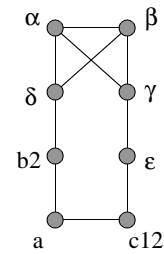
X-ray structure of ATP synthase



Schematic version



Network representation



Total set = protein complex
Sum of direct + indirect interactions

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

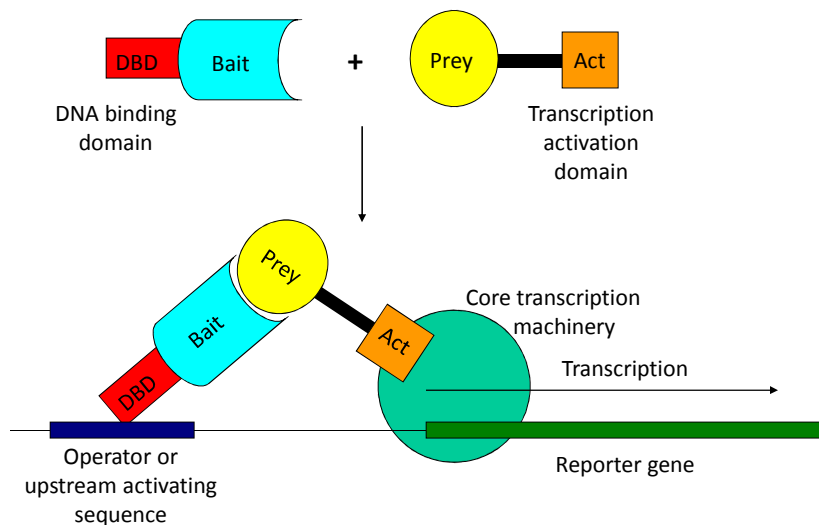
Let's look at some of the types of interaction data in more detail.

Some of these capture physical interactions, some genetic, some informational or logical.

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Pairwise protein interactions

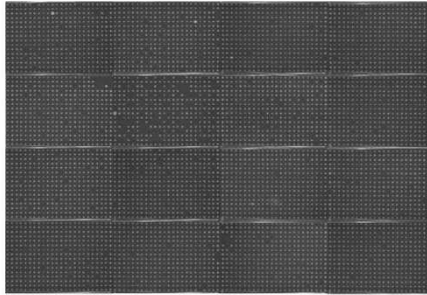
In general, purifying proteins one at a time, mixing them, and assaying for interactions is far too slow & laborious. We need something faster! Hence, high-throughput screens, e.g. yeast two-hybrid assays



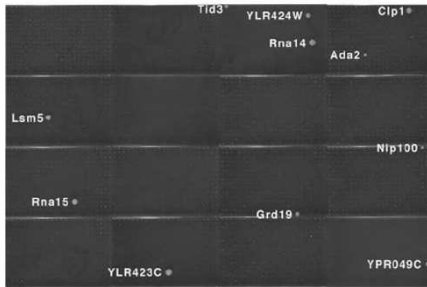
Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

High-throughput yeast two-hybrid assays

Haploid yeast cells expressing activation domain-prey fusion proteins



Diploid yeast probed with DNA-binding domain-Pcf11 bait fusion protein

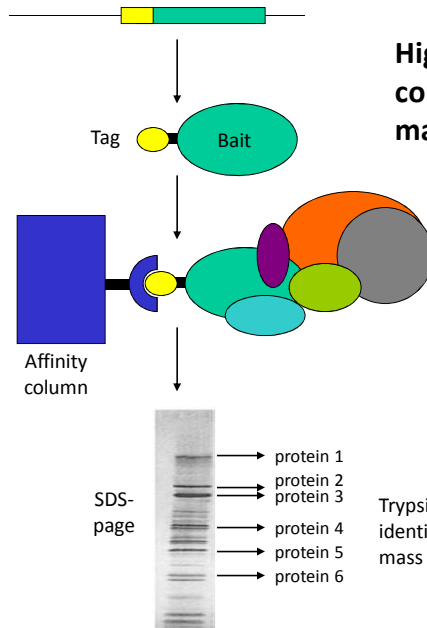


Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

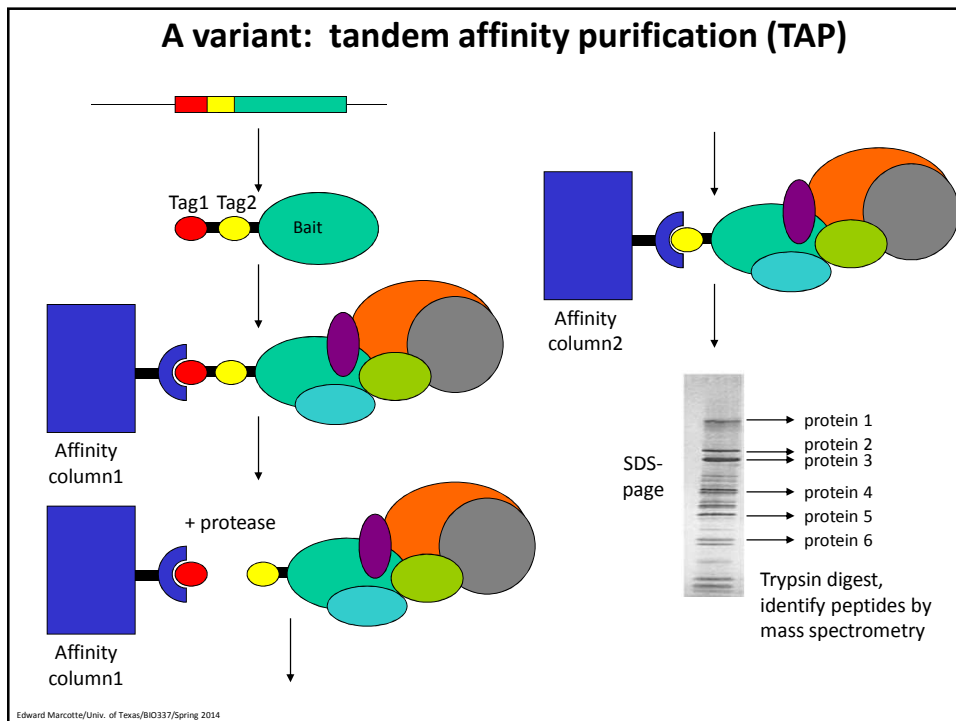
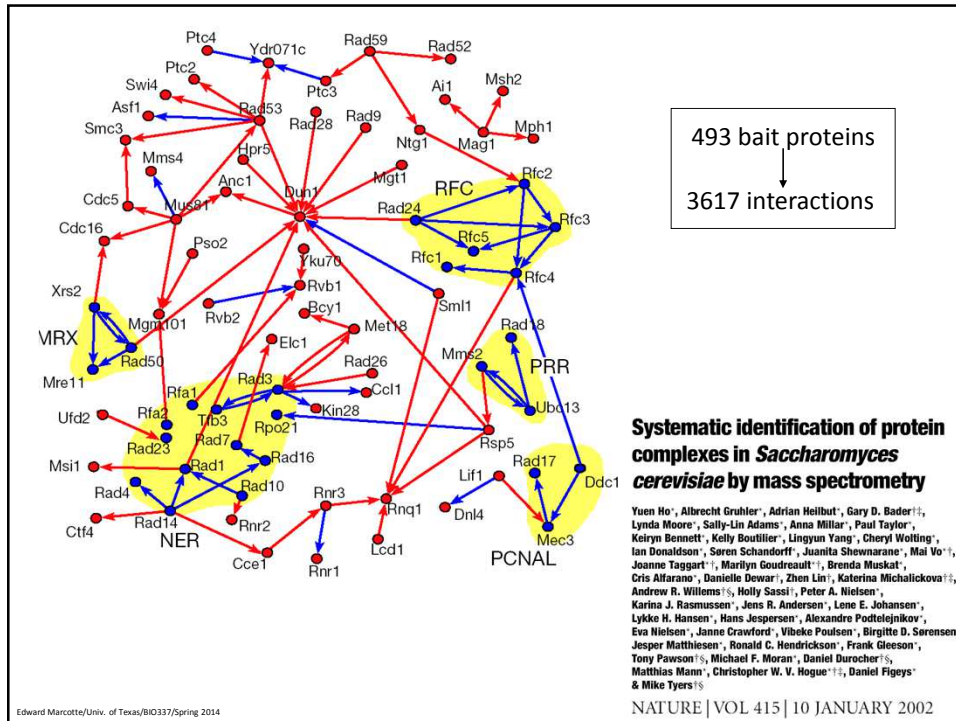
Uetz, Giot, et al. *Nature* (2000)

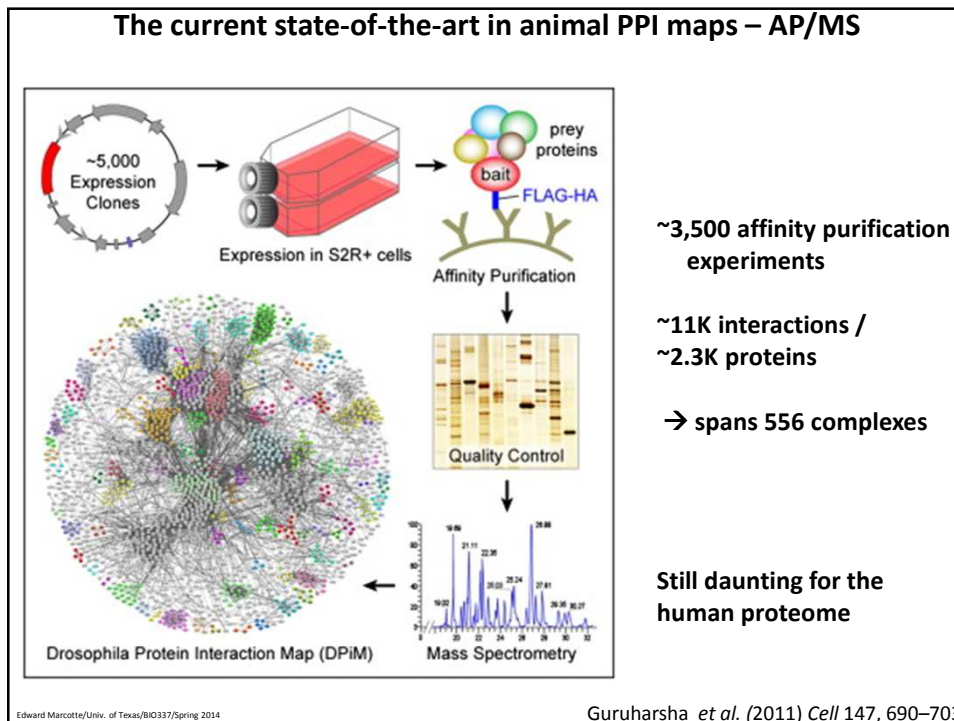
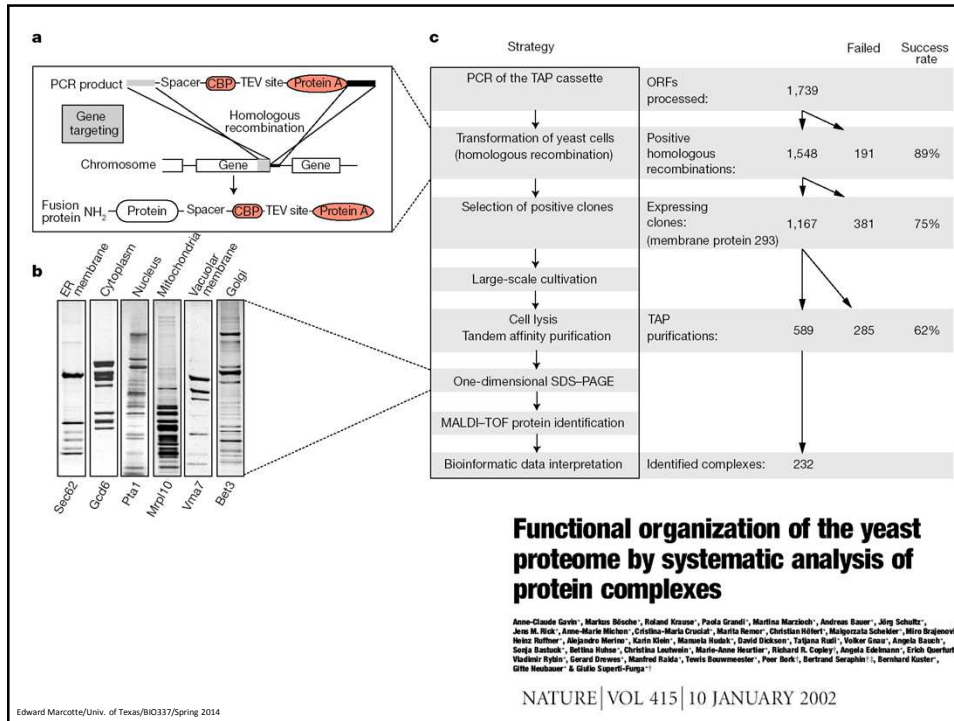
Protein complexes

High-throughput complex mapping by mass spectrometry

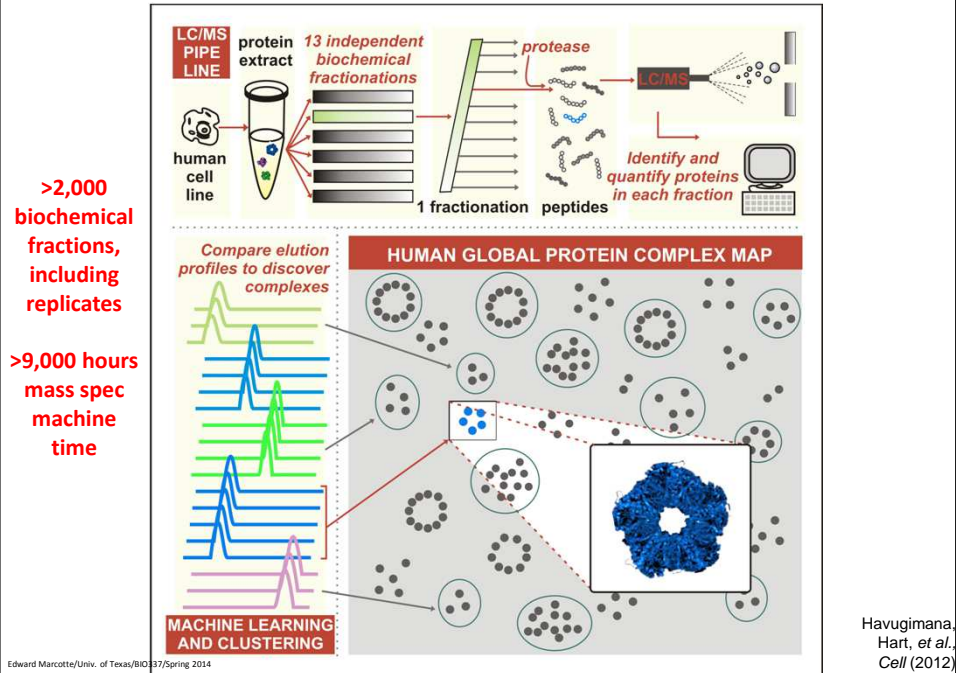


Edward Marcotte/Univ. of Texas/BIO337/Spring 2014



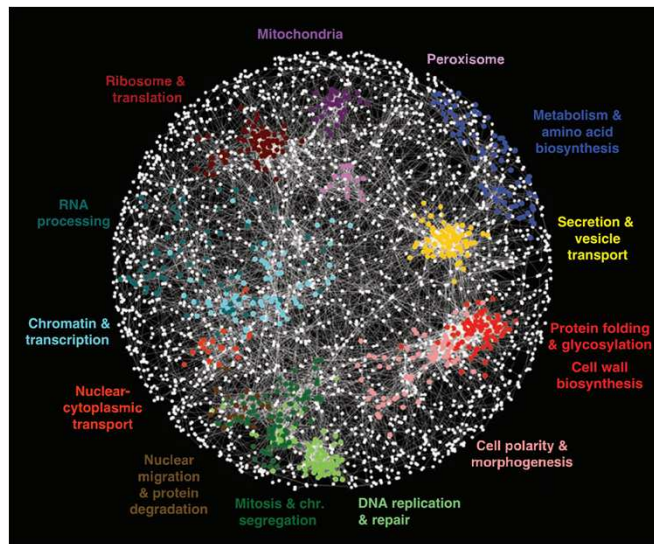


The current state-of-the-art in animal PPI maps – co-fractionation/MS



Genetic interactions

5.4 million gene-gene pairs assayed for synthetic genetic interactions in yeast



Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Costanzo *et al.*, *Science* 327: 425 (2010)

Comparative genomics

Functional relationships between genes impose subtle constraints upon genome sequences. Thus, genomes carry intrinsic information about the cellular systems and pathways they encode.

Linkages can be found from aspects of gene context, including:

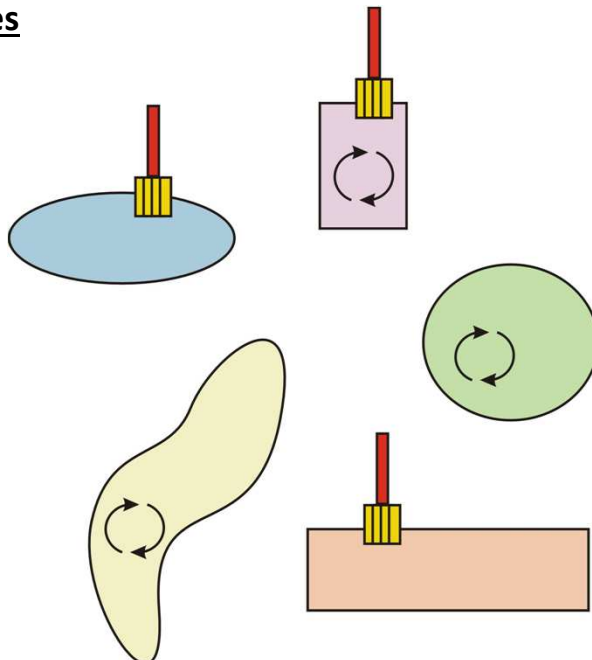
- Distances between sequence elements
- Order of sequences
- Variation in order between organisms
- Regulatory sequences near genes
- Gene content of an organism
- Variation in gene content between organisms
- Fusions between genes from different organisms

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Phylogenetic profiles

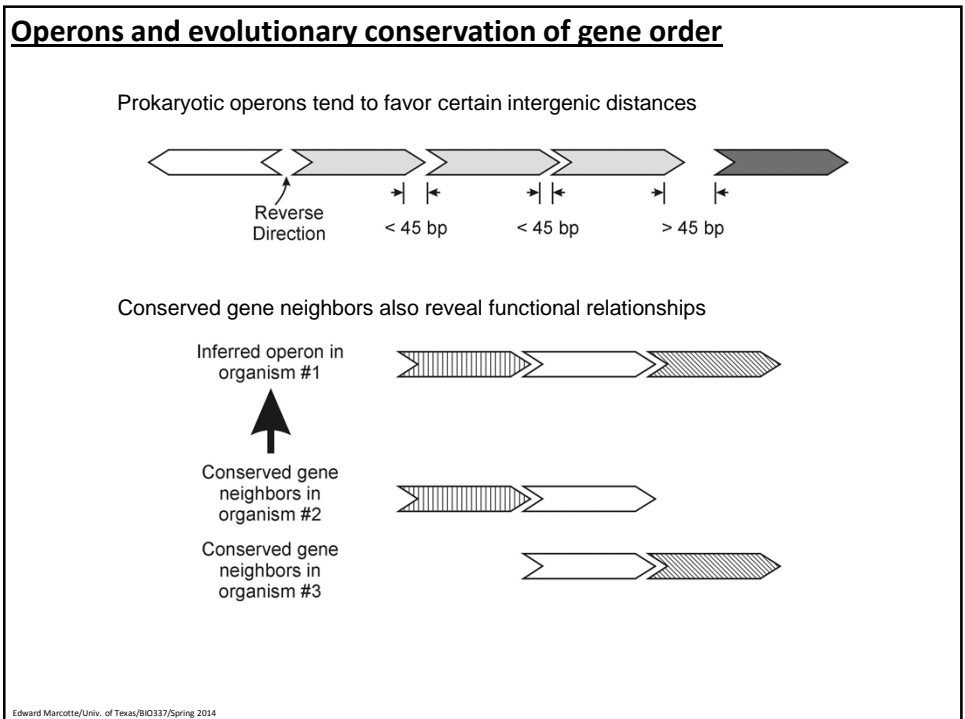
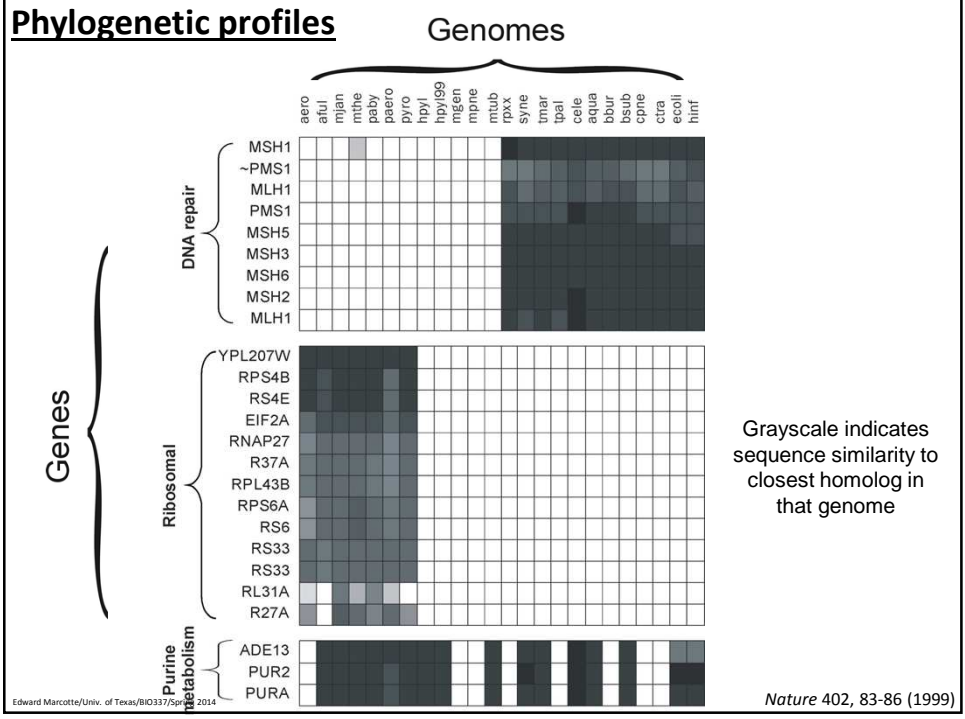
Organisms with e.g. a flagellum have the necessary genes; those without tend to lack them.

Specific trends of gene presence/absence thus inform about biological processes.

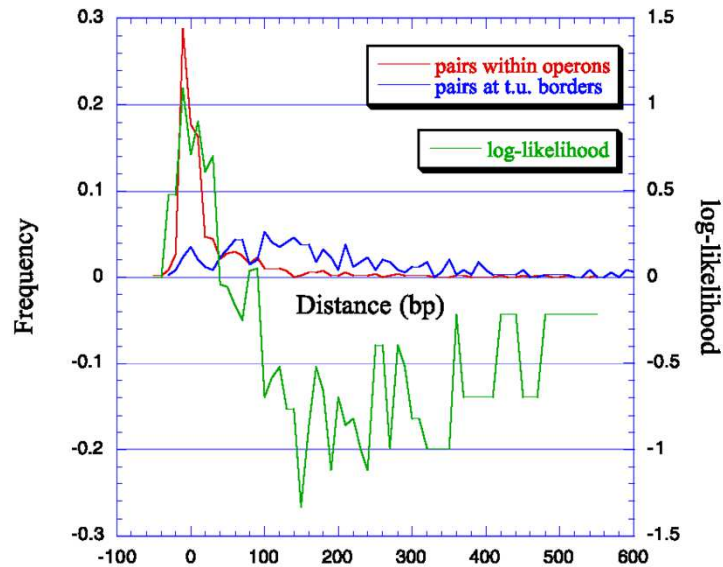


Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

PNAS 96, 4285-4288 (1999)



Again, such observations can be turned into pairwise scores:



Operons in *Escherichia coli*: Genomic analyses and predictions

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014 | Heladia Salgado^{1*}, Gabriel Moreno-Hagelsieb^{1*}, Temple F. Smith¹, and Julio Collado-Vides^{1†}

To summarize so far:

Data about gene interactions comes many sources but is dominated by several major ones:

- mRNA co expression. Historically microarrays & ESTs, increasingly RNAseq. Typically very high coverage data.
- Comparative genomics. Available for free for all organisms (typically phylogenetic profiles & operons)
- Protein interactions, especially co-complex interactions from mass spectrometry
- Genetic interactions (more so matching profiles of interaction partners than the interactions themselves)
- Transfer from other species

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

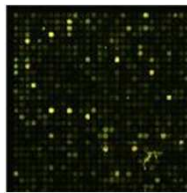
More abstractly, we might consider all of these as indicating “functional linkages” between genes

- Protein-protein interactions
- Participating in consecutive metabolic reactions
- Sharing genetic interactors
- Forming the same protein complex
- Giving rise to similar mutational phenotypes
- Exhibiting similar biological function
and so on...

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

These sorts of data can be combined into functional gene networks

Including measurements of...



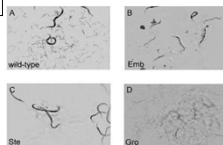
Gene expression
(RNA-seq/arrays)



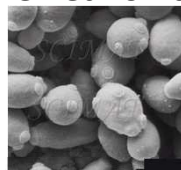
Protein expression
and interactions
(Mass spectrometry)

AAACTGCATCGA
ATCGCGCATCGC
AGCTCTAGCTCCC...

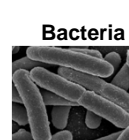
Gene organization
(Genome sequences)



Gene-gene interactions
(Genetic assays)



Yeast



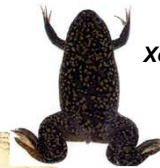
Bacteria

C. elegans
nematodes



Mouse

Sanger Ctr

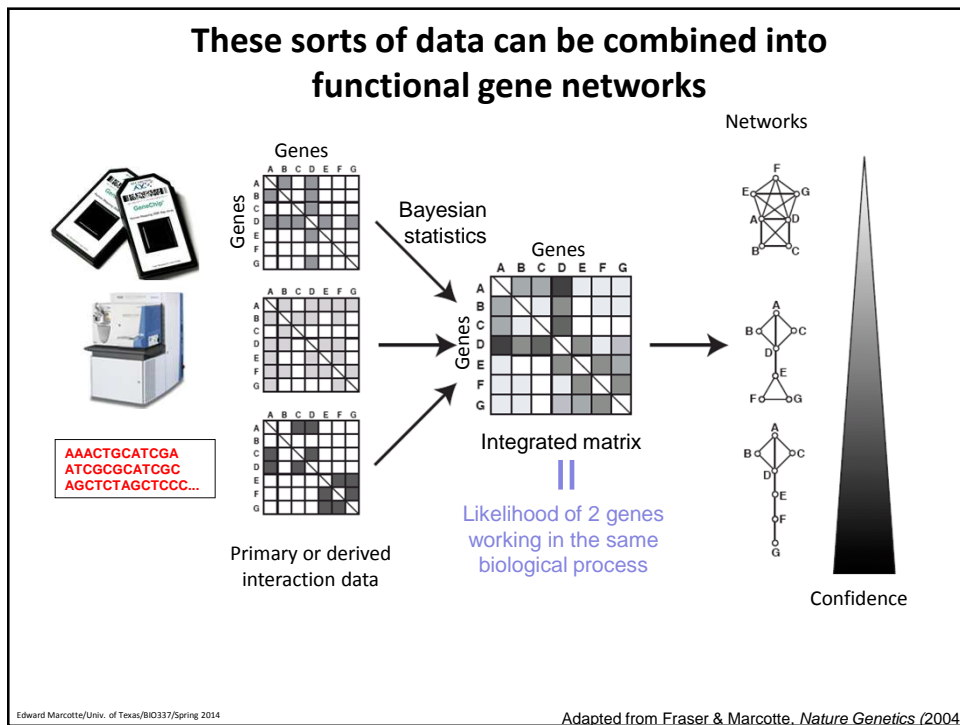
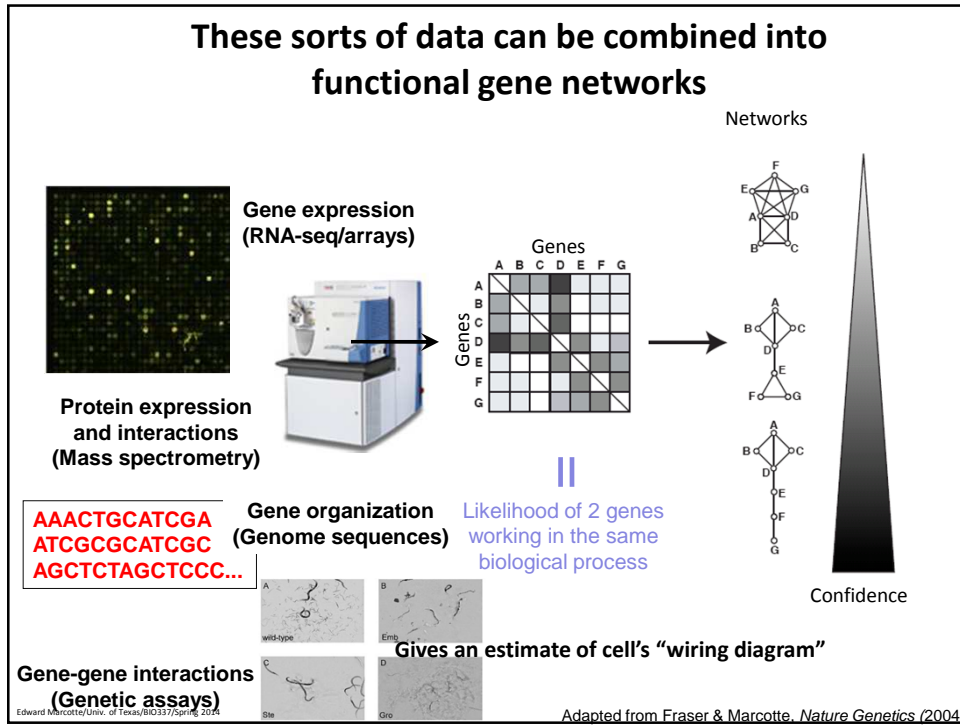


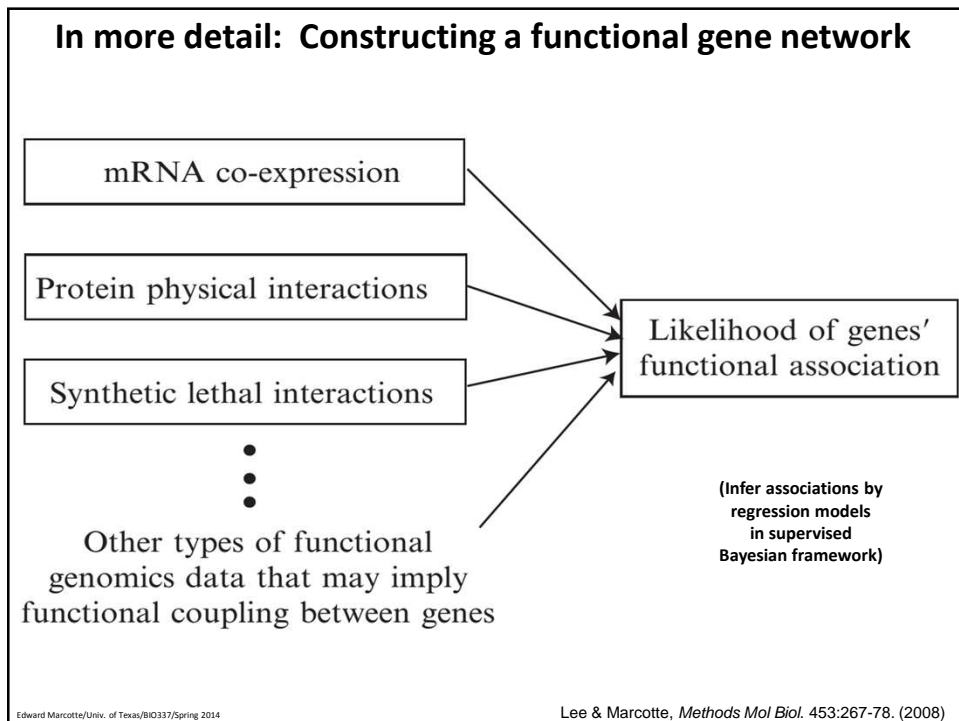
Xenopus
frogs



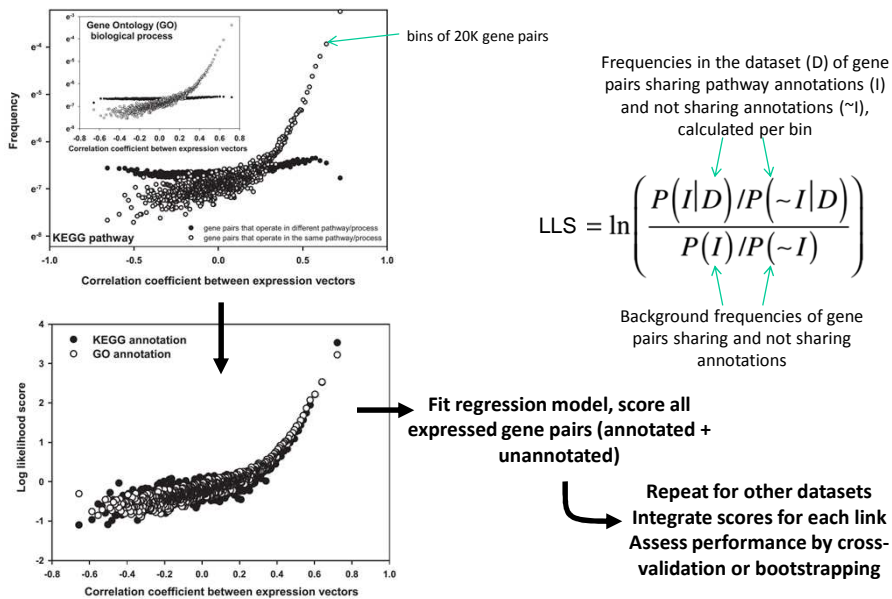
Humans

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014





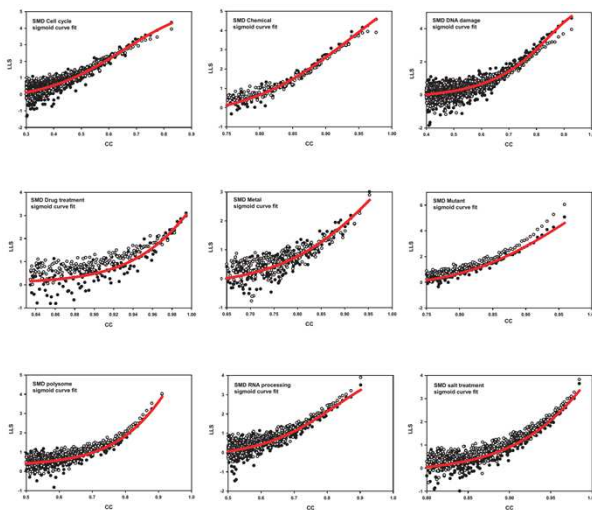
e.g., inferring functional linkages from mRNA co-expression across a given set of conditions



Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Lee *et al. Science*, 306:1555-8 (2004)

Typically calculated for many different sets of experiments sampling many different conditions...



Each represents a different set of mRNA abundance profiling experiments (here, DNA microarrays) interrogating a given set of conditions.

Different gene pairs may be correlated in each condition.

... and so on

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Lee *et al. Science*, 306:1555-8 (2004)

21 evidence types contribute to the HumanNet human gene network

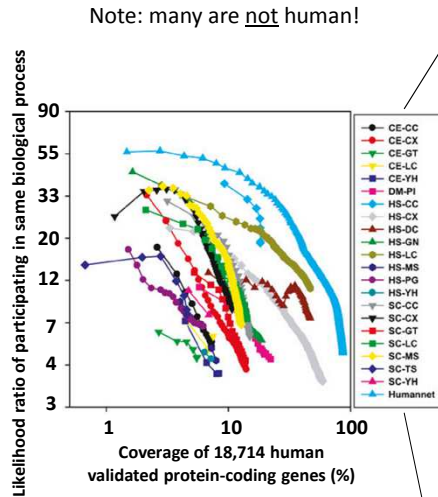


Table S1. Twenty-one different lines of evidence supporting HumanNet linkages.

Data set	Description	# genes	# gene pairs
CE-CC	Co-citation of worm genes	1,370	12,928
CE-CX	Co-expression among worm genes	2,633	41,645
CE-GT	Worm genetic interactions	1,040	5,430
CE-LC	Literature-curated worm protein physical interactions	1,402	2,640
CE-YH	High-throughput yeast 2-hybrid assays among worm genes	1,561	3,254
DM-PI	Fly protein physical interactions	4,153	15,738
HS-CC	Co-citation of human genes	3,423	6,172
HS-CX	Co-expression among human genes	11,050	156,317
HS-DC	Co-occurrence of domains among human proteins	8,737	38,797
HS-GN	Gene neighborhoods of bacterial and archaeal orthologs of human genes	3,504	36,487
HS-LC	Literature-curated linkages from protein-protein interaction DBs (HPRD, BIND, BioGRID, IntAct, MINT) and Rual <i>et al.</i> , and pathway DB (Reactome)	8,783	56,505
HS-MS	Human protein complexes from affinity purification/mass spectrometry	1,485	3,575
HS-PG	Co-inheritance of bacterial and archaeal orthologs of human genes	1,170	18,868
HS-YH	High-throughput yeast 2-hybrid assays among human genes	1,358	1,365
SC-CC	Co-citation of yeast genes	2,798	31,353
SC-CX	Co-expression among yeast genes	2,001	48,423
SC-GT	Yeast genetic interactions	2,584	17,678
SC-LC	Literature-curated yeast protein physical interactions	2,661	17,280
SC-MS	Yeast protein complexes from affinity purification/mass spectrometry	2,382	65,986
SC-TS	Yeast protein interactions inferred from tertiary structures of complexes	859	6,270
SC-YH	High-throughput yeast 2-hybrid assays among yeast genes	1,292	1,801

Lee, Blom *et al.*, *Genome Research* 21:1109-21 (2011)

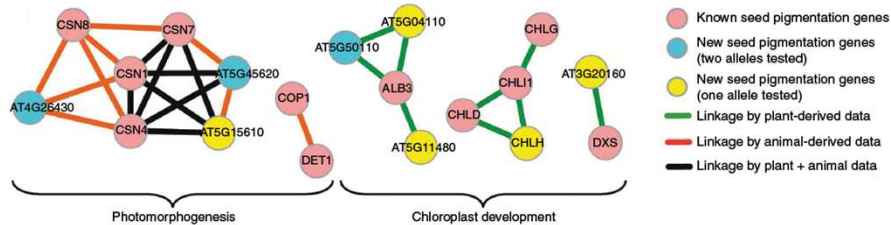
Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Evolutionary information is usually a key predictor— e.g., predictions for plant-specific traits often use fungal & animal data

For example, new seedling pigmentation genes...



...were predicted from both plant and animal data:

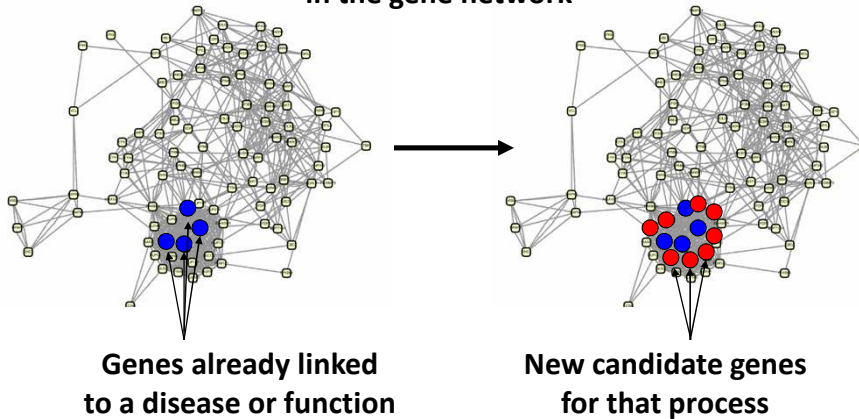


Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Lee, Ambaru, *et al.* *Nature Biotech* 28(2):149-156 (2010)

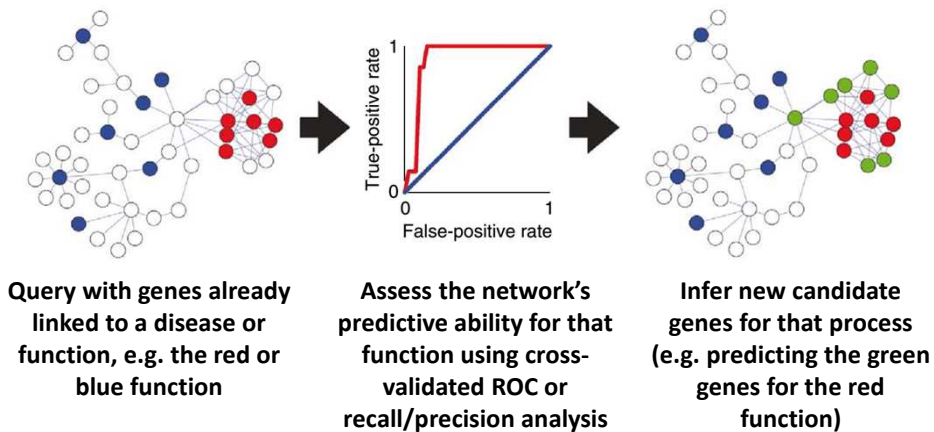
These networks are hypothesis generators.
 Given a gene, what other genes does it function with?
 What do they do?

**Guilt-by-association
 in the gene network**



Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

We can propagate annotations across the graph to infer new annotations for genes (network “guilt-by-association”, or GBA).
 Measuring how well this works on hidden, but known, functions gives us an idea how predictive it will be for new cases.



Lee, Ambaru *et al. Nature Biotechnology* 28:149-156 (2010)

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Calculating ROC curves

		Actual	
		P	N
Prediction	P'	True Positive	False Positive
	N'	False Negative	True Negative

Basic idea: sort predictions from best to worst, plot TPR vs. FPR as you traverse the ranked list

$$\text{TPR} = \text{TP} / P = \text{TP} / (\text{TP} + \text{FN})$$

= True Positive Rate
= Sensitivity, Recall

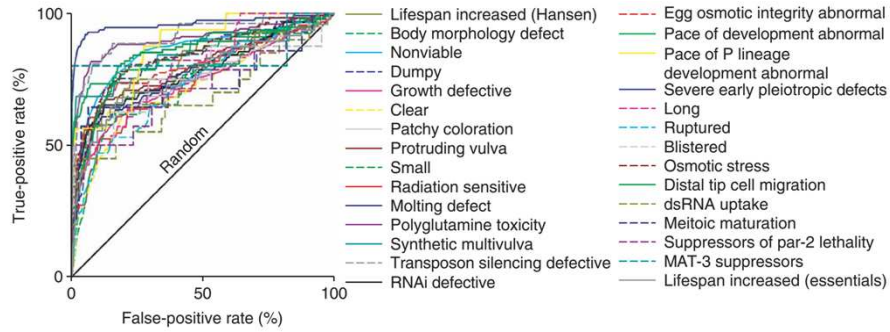
$$\text{FPR} = \text{FP} / N = \text{FP} / (\text{FP} + \text{TN})$$

= False Positive Rate
= 1 - Specificity

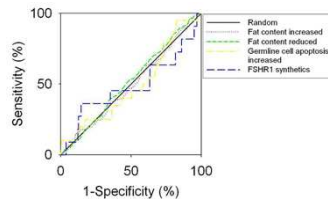
Also useful to plot Precision [= TP / (TP + FP)] vs. Recall (= TPR)

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

For example, predicting genes linked with worm phenotypes in genome-wide RNAi screens



Some very poorly predicted pathways:

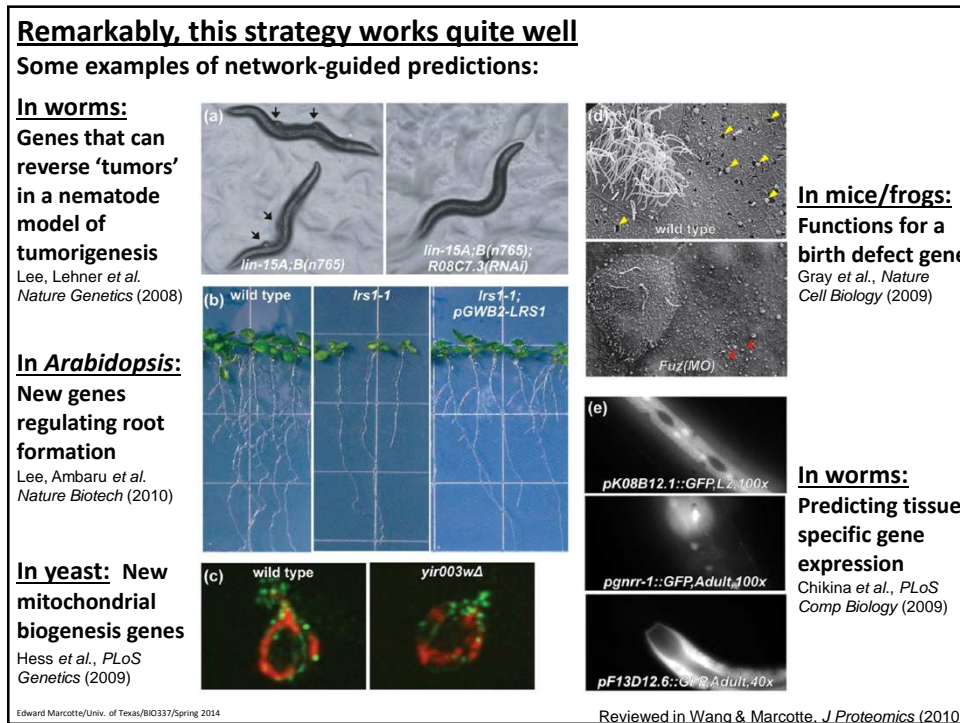
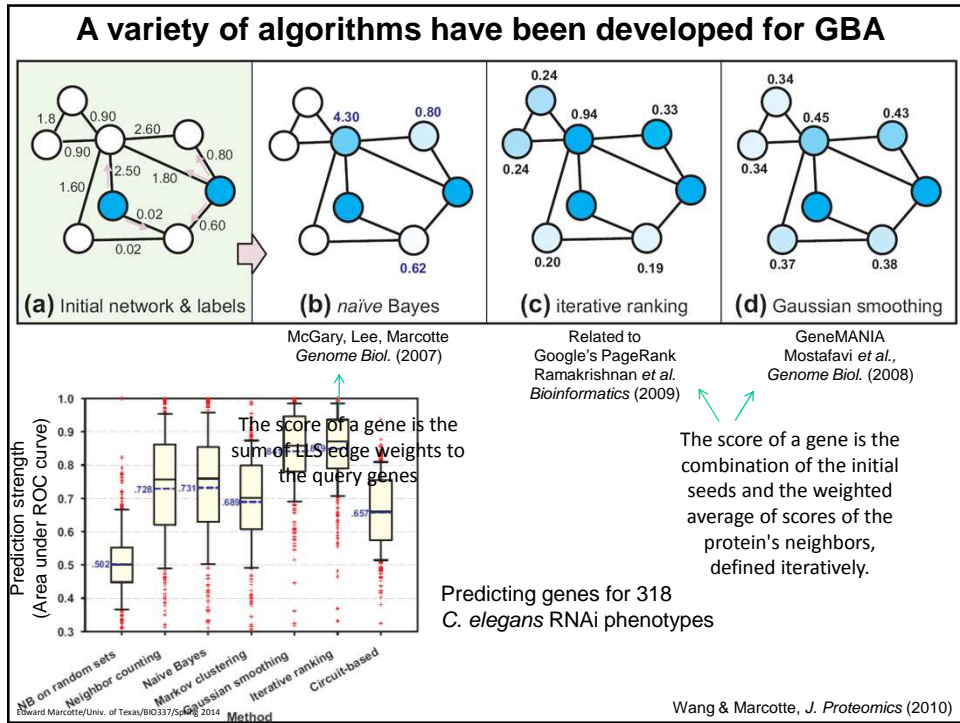


ROC analysis indicates the likely predictive power of the network for a system of interest.

A poor ROC → no better than random guessing.

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

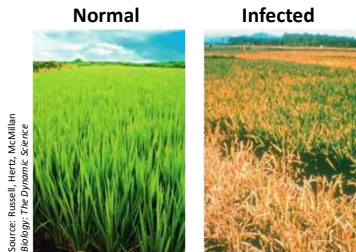
Lee, Lehner et al., Nat Genet, 40(2):181-8 (2008)



Applicable to non-model organisms:

In rice: Identifying genes regulating resistance to *Xanthomonas oryzae* infection ...

- Rice is the primary food source for >2 billion people worldwide
- >500 million tons rice/year are grown
 - Bacterial rice blight destroys up to 10-50% / year in Africa/Asia



Infection with *Xoo*



Lee, Seo, *et al.* PNAS 108:18548–18553 (2011)

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

Summary of the major themes

- Gene networks serve as general frameworks for studying gene function
- Functional gene networks can be (re)constructed based upon millions of experimental observations via integrating these data into statistical models of functional connectivity among genes
- Guilt-by-association in such a network allows association of genes with functions, and genes with phenotypic traits, even highly polygenic ones
- Functional gene networks have been constructed for yeast, *C. elegans*, *Arabidopsis*, rice, mice, humans, many prokaryotes, and many other organisms...

Edward Marcotte/Univ. of Texas/BIO337/Spring 2014

**Live demo of
functional networks
and Cytoscape**