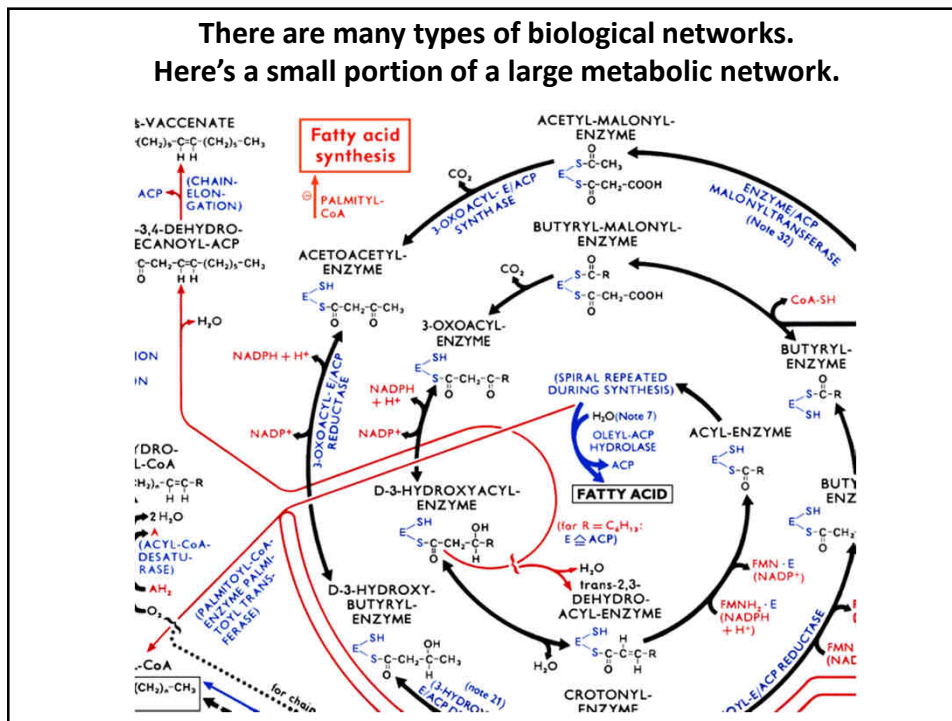
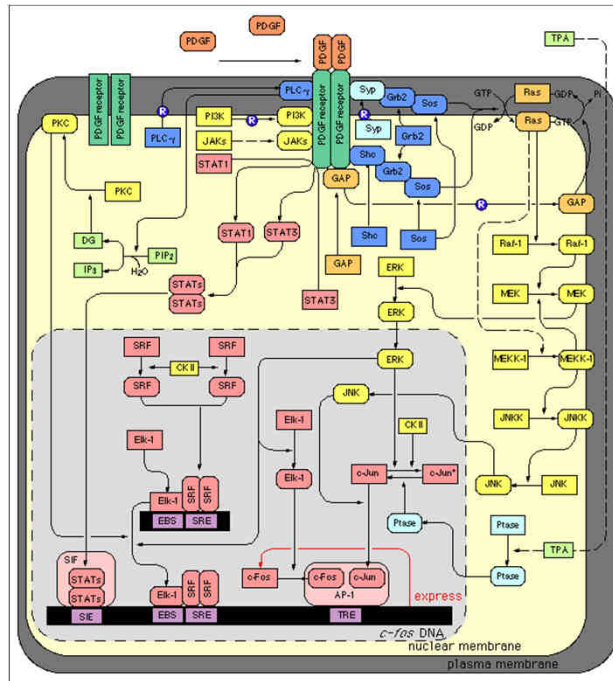


Network biology (& predicting gene function)

BCH339N Systems Biology / Bioinformatics – Spring 2016
Edward Marcotte, Univ of Texas at Austin

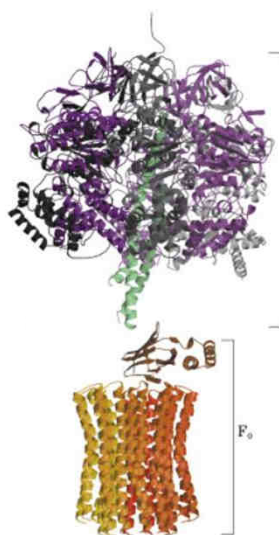


A typical
genetic
network

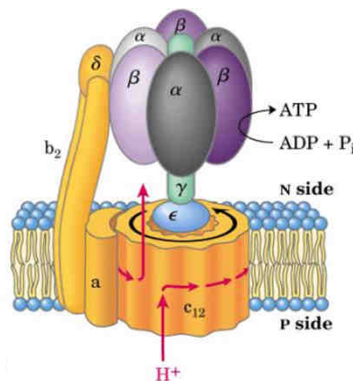


Contacts between proteins define protein interaction networks

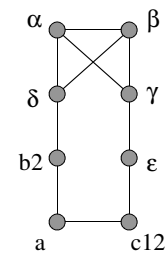
X-ray structure
of ATP synthase



Schematic
version



Network
representation



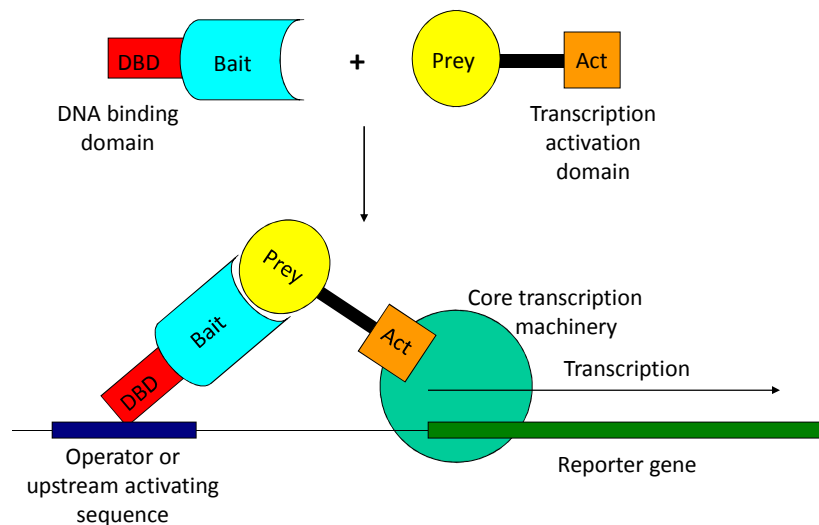
Total set = protein complex
Sum of **direct** + **indirect**
interactions

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.

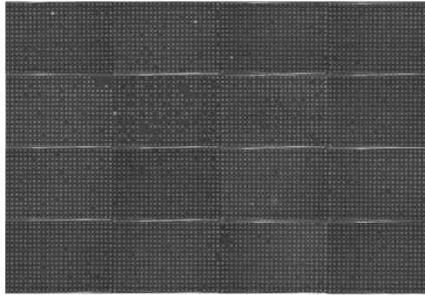
Pairwise protein interactions

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

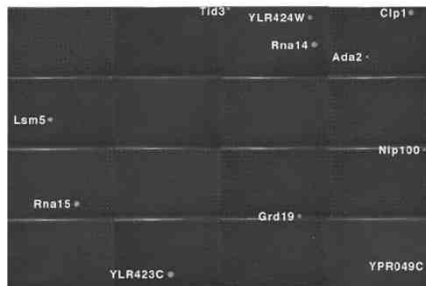


High-throughput yeast two-hybrid assays

Haploid yeast cells expressing activation domain-prey fusion proteins



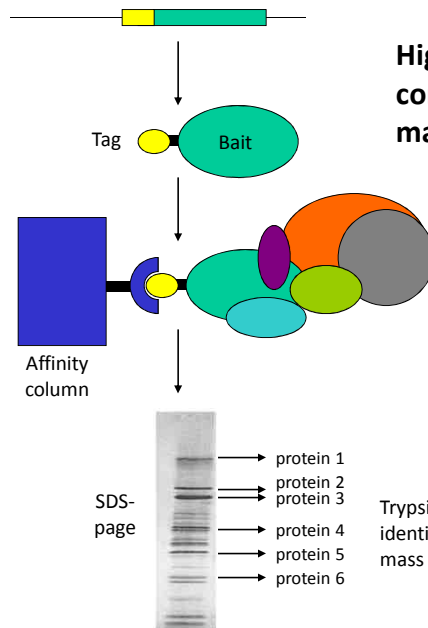
Diploid yeast probed with DNA-binding domain-Pcf11 bait fusion protein

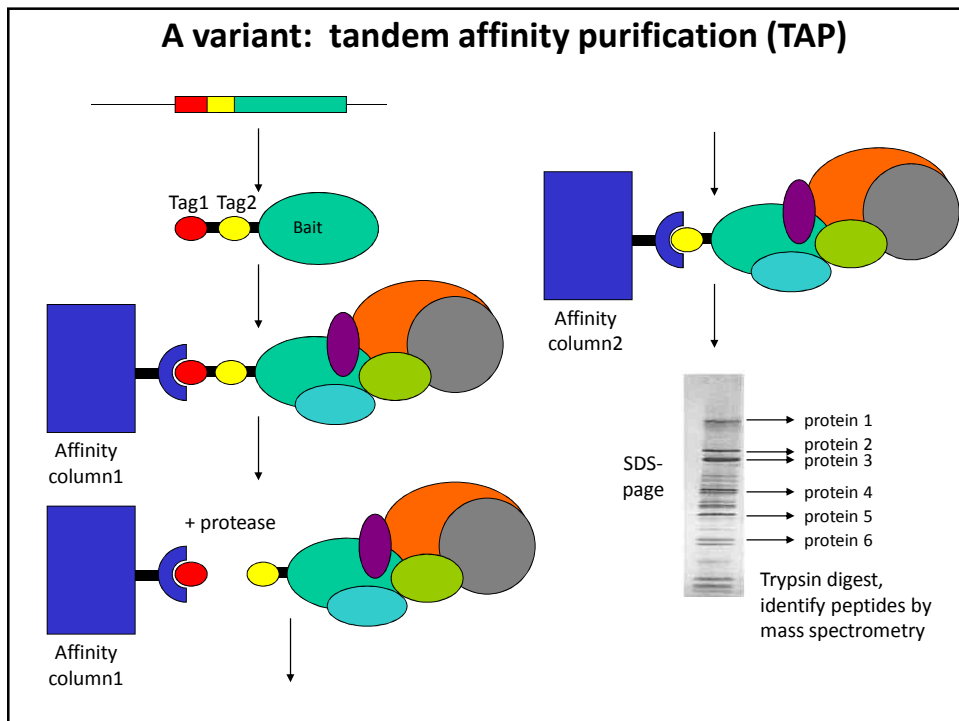
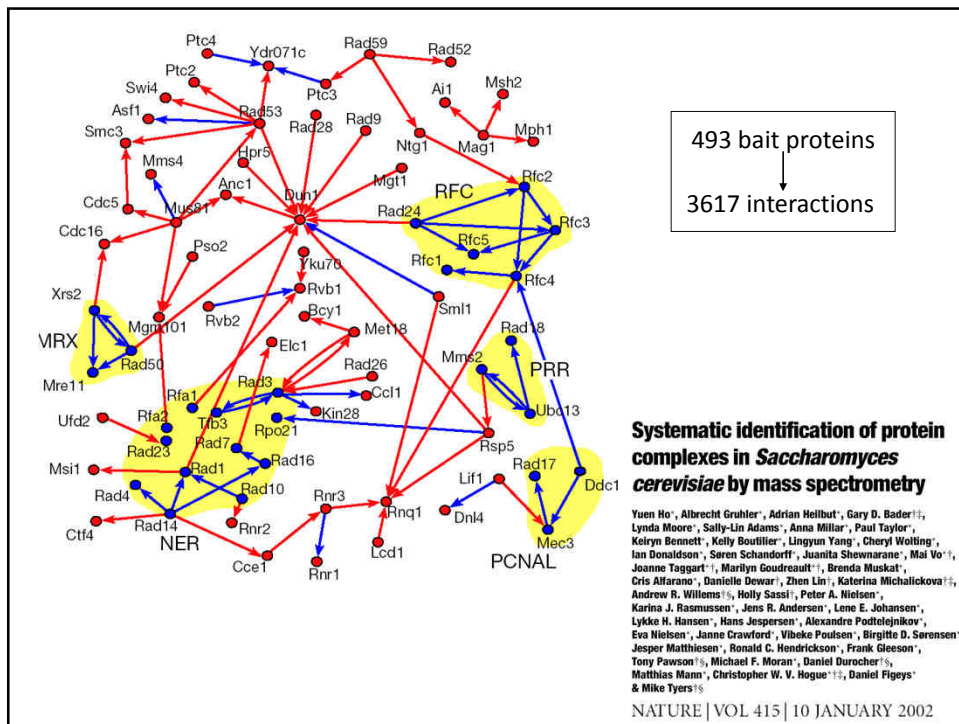


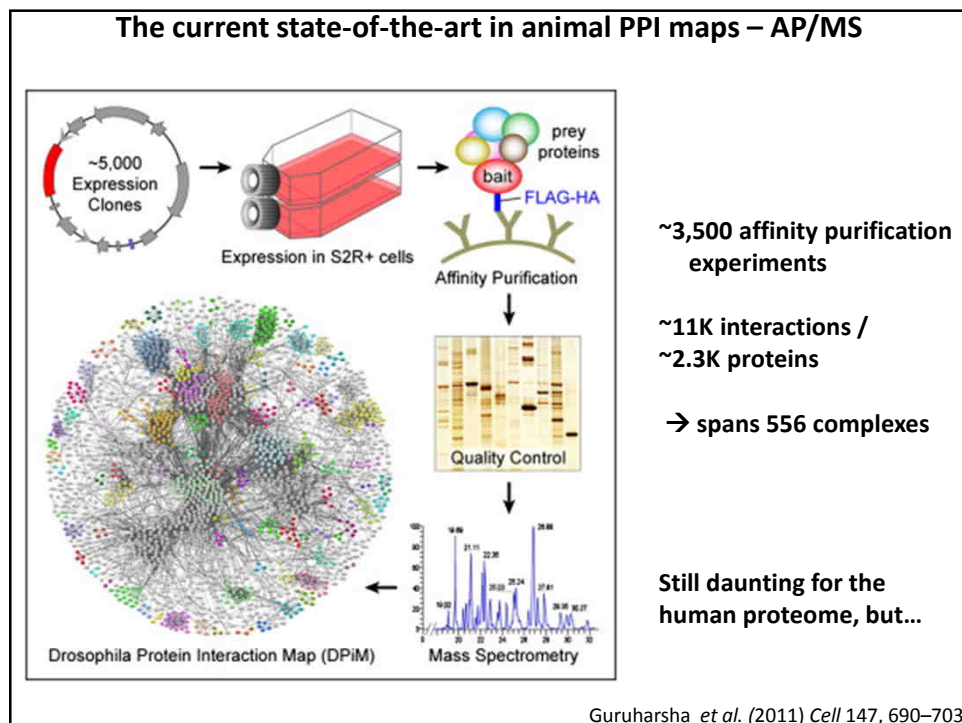
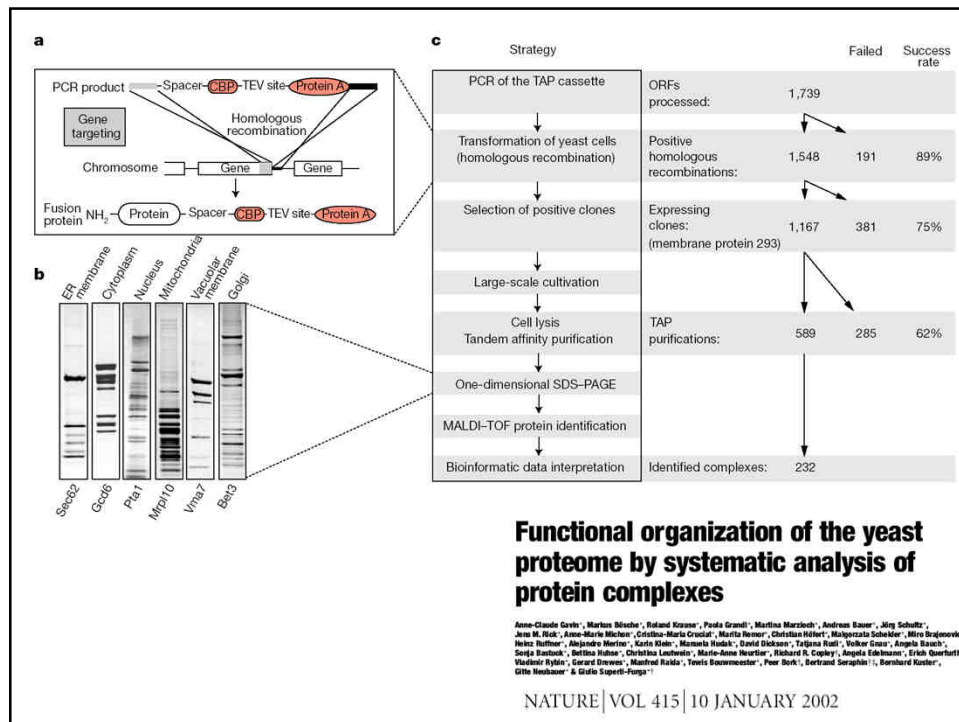
Uetz, Giot, *et al. Nature* (2000)

Protein complexes

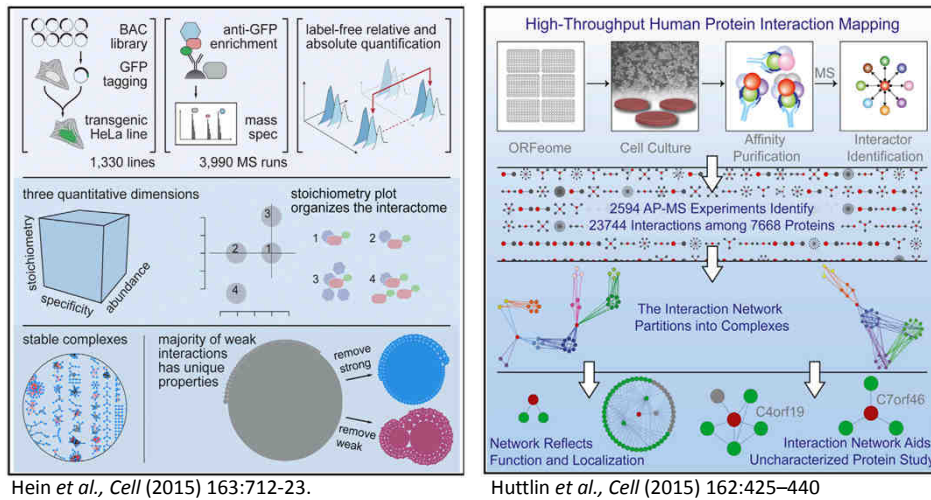
High-throughput complex mapping by mass spectrometry





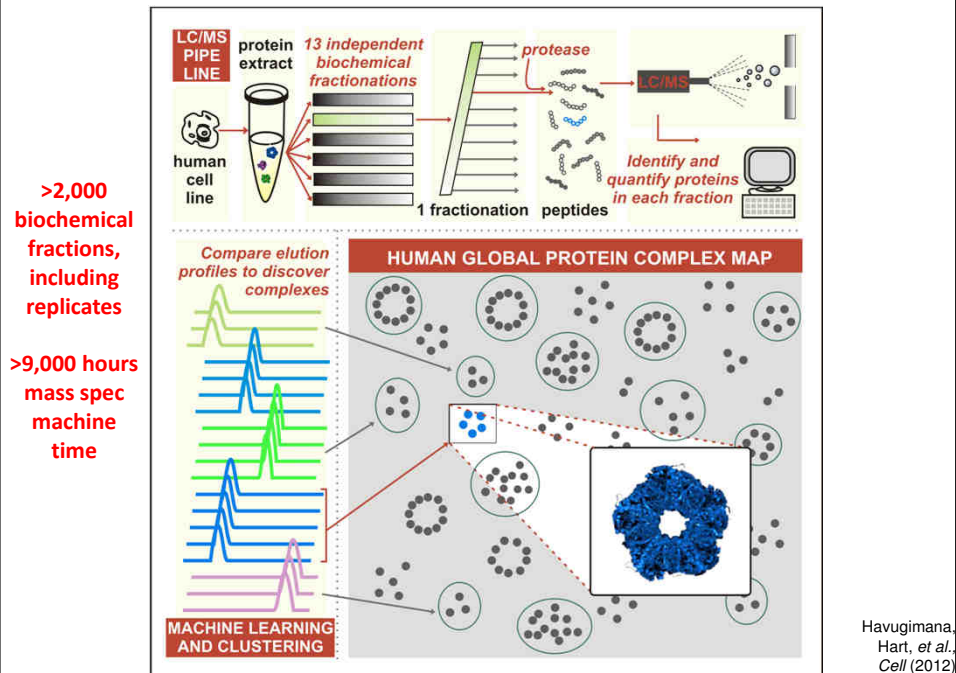


The current state-of-the-art in human PPI maps – large scale AP/MS

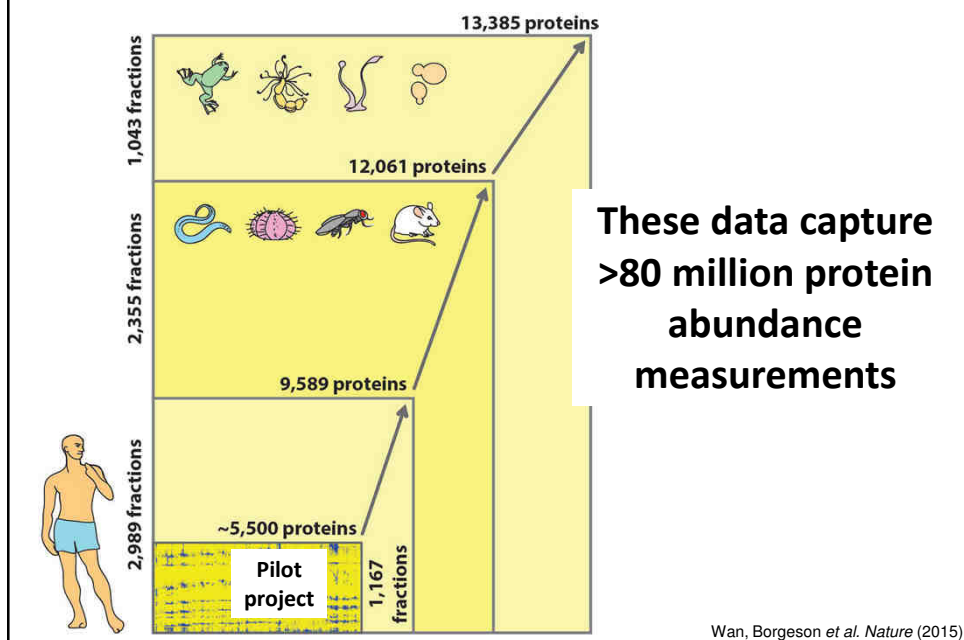


Just in the past year, nearly 4K affinity purification experiments on tagged human proteins expressed in cell lines

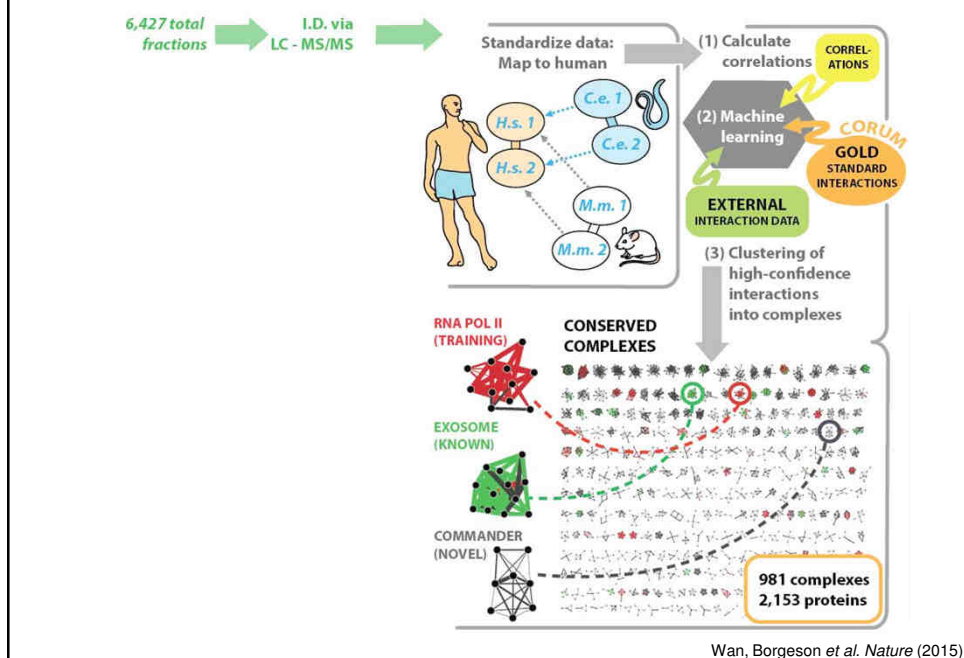
The current state-of-the-art in animal PPI maps – co-fractionation/MS



Now >6,400 mass spec experiments across animals

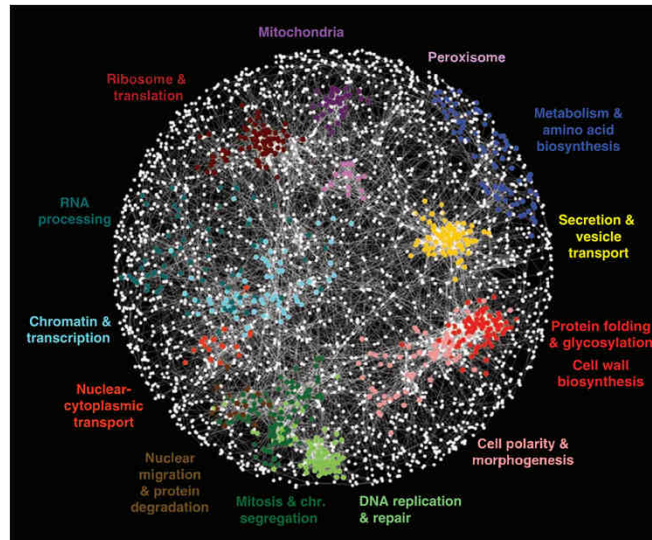


Extending the map across animals...



Genetic interactions

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



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.

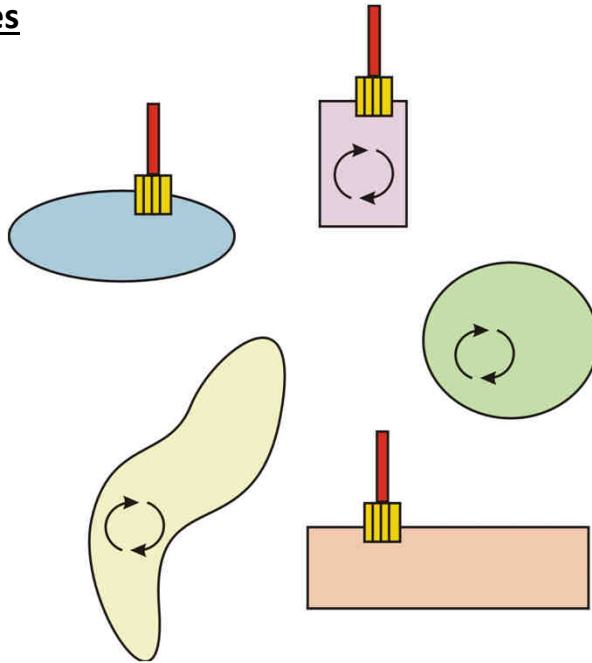
Linkages can be found from aspects of gene context, including:

- 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

Phylogenetic profiles

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

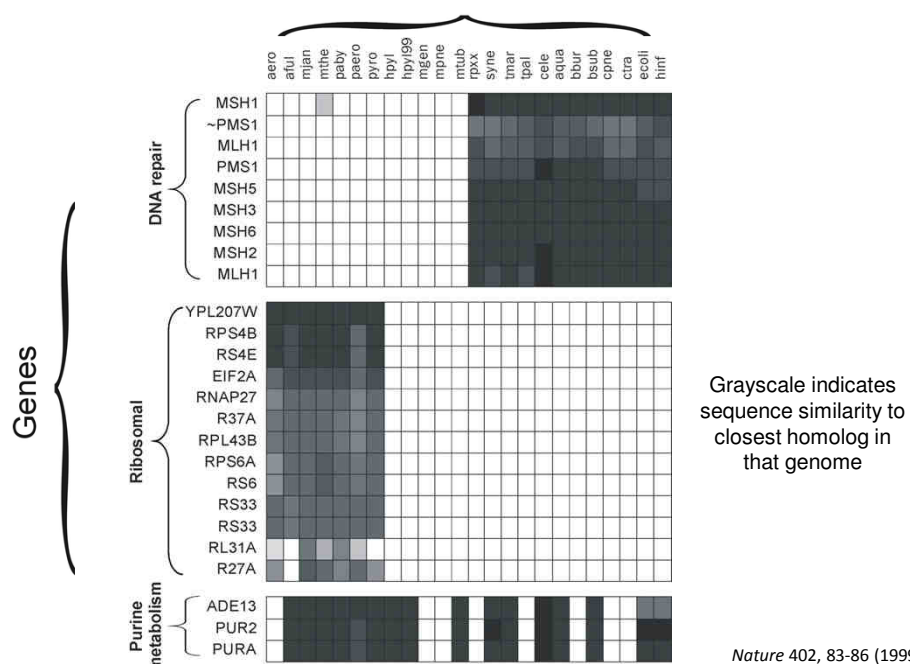
Specific trends of gene presence/absence thus inform about biological processes.



PNAS 96, 4285-4288 (1999)

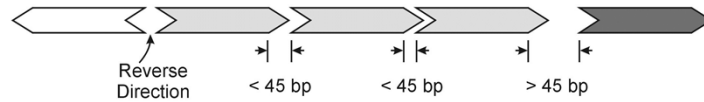
Phylogenetic profiles

Genomes

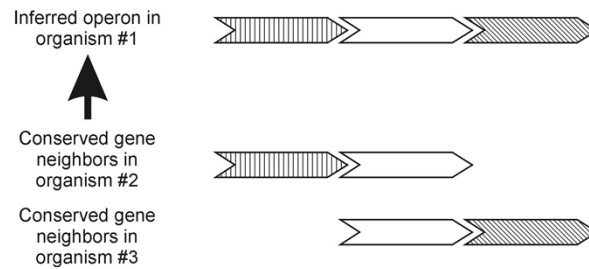


Operons and evolutionary conservation of gene order

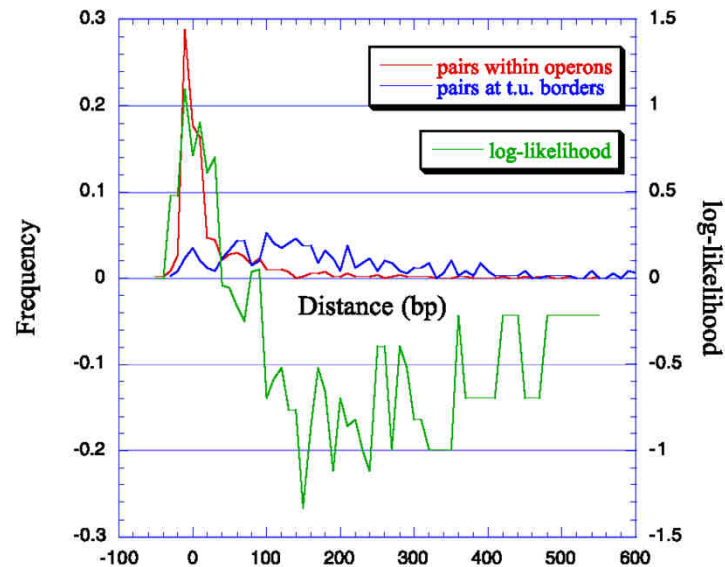
Prokaryotic operons tend to favor certain intergenic distances



Conserved gene neighbors also reveal functional relationships



Again, such observations can be turned into pairwise scores:



Operons in *Escherichia coli*: Genomic analyses and predictions

6452-6457 | PHAS | June 6, 2000 | vol. 97 | no. 12

Heladia Salgado¹, Gabriel Moreno-Hagelsieb¹, Temple F. Smith¹, and Julio Collado-Vides^{1,2}

To summarize so far:

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

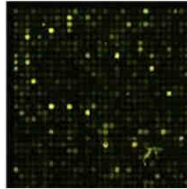
- mRNA co expression. Historically microarrays & ESTs, increasingly RNAseq. Typically very high coverage data.
- Comparative genomics. Available for free for all organisms (typically phylogenetic profiles & operons)
- Protein interactions, especially co-complex interactions from mass spectrometry
- Genetic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

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...

These sorts of data can be combined into functional gene networks

Including measurements of...



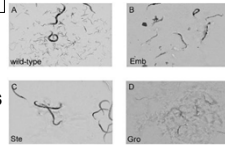
Gene expression
(RNA-seq/arrays)



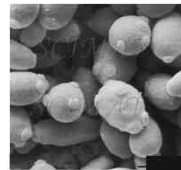
Protein expression
and interactions
(Mass spectrometry)

AAACTGCATCGA
ATCGCGCATCGC
AGCTCTAGCTCCC...

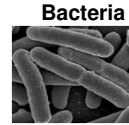
Gene organization
(Genome sequences)



Gene-gene interactions
(Genetic assays)



Yeast

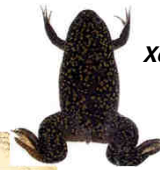


Bacteria

C. elegans
nematodes



Mouse

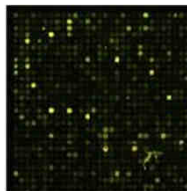


Xenopus
frogs



Humans

These sorts of data can be combined into functional gene networks



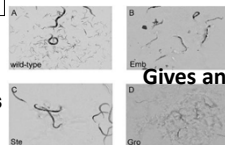
Gene expression
(RNA-seq/arrays)



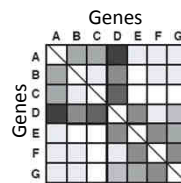
Protein expression
and interactions
(Mass spectrometry)

AAACTGCATCGA
ATCGCGCATCGC
AGCTCTAGCTCCC...

Gene organization
(Genome sequences)

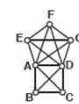


Gene-gene interactions
(Genetic assays)



Likelihood of 2 genes
working in the same
biological process

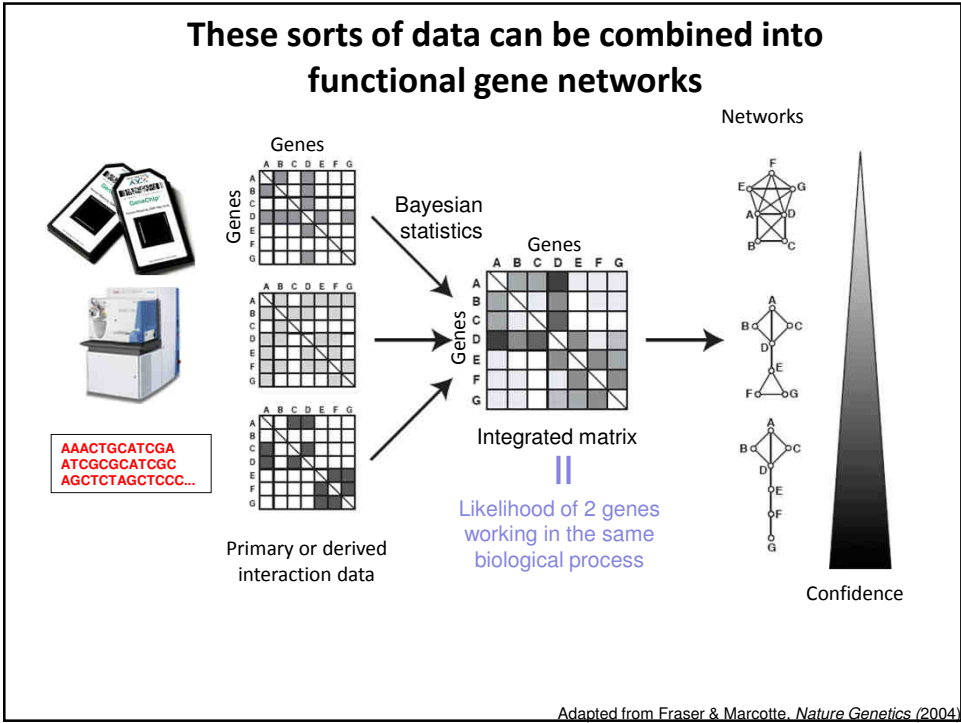
Networks



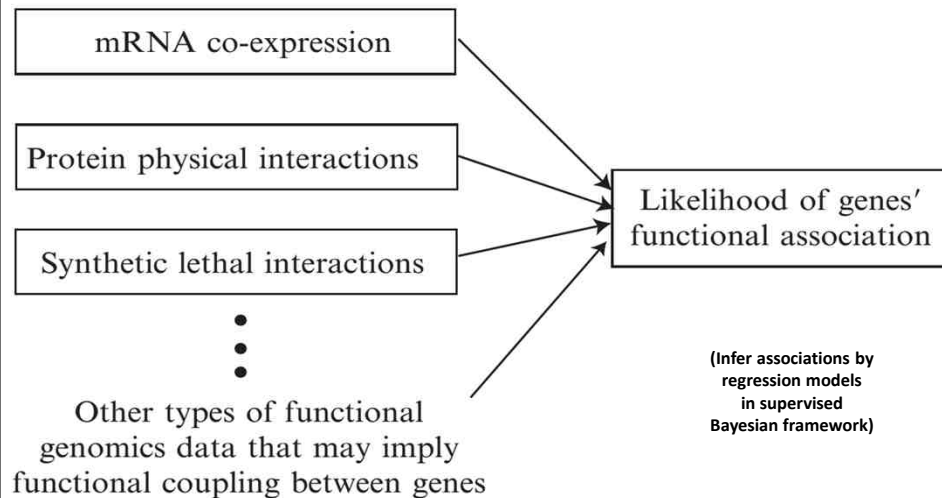
Confidence

Gives an estimate of cell's "wiring diagram"

Adapted from Fraser & Marcotte, *Nature Genetics* /2004



In more detail: Constructing a functional gene network



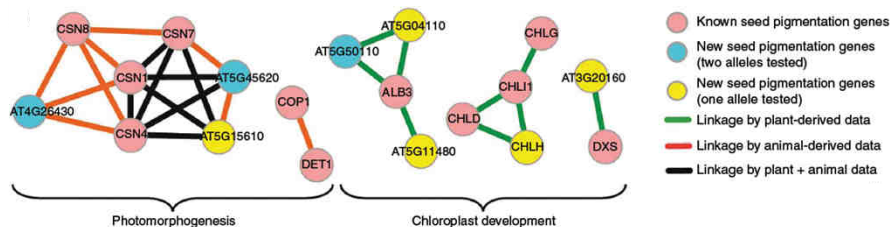
Lee & Marcotte, *Methods Mol Biol.* 453:267-78. (2008)

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...



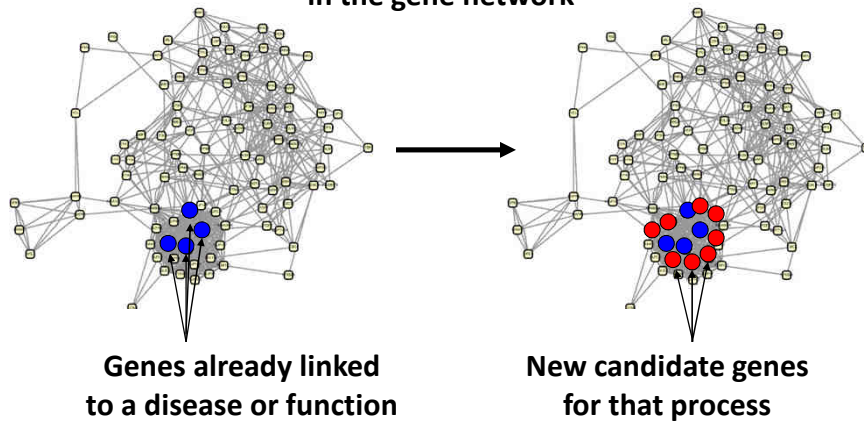
...were predicted from both plant and animal data:



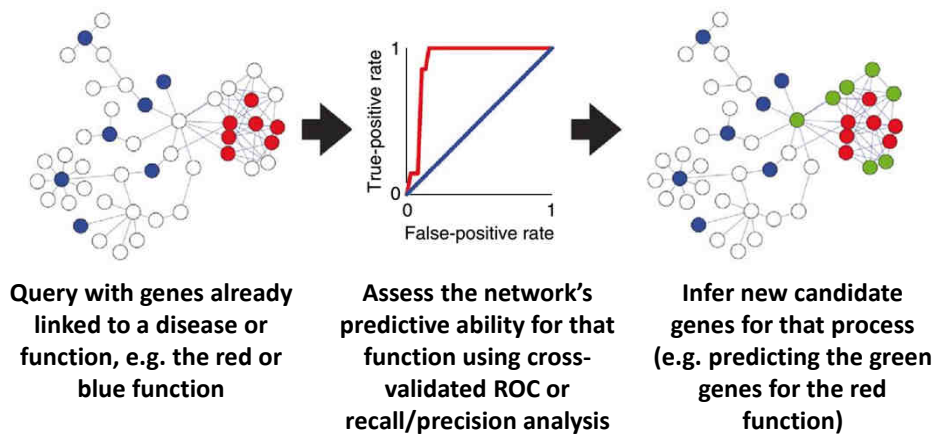
Lee, Ambaru, et al. *Nature Biotech* 28(2):149-156 (2010)

**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**



**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.**



Calculating ROC curves

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

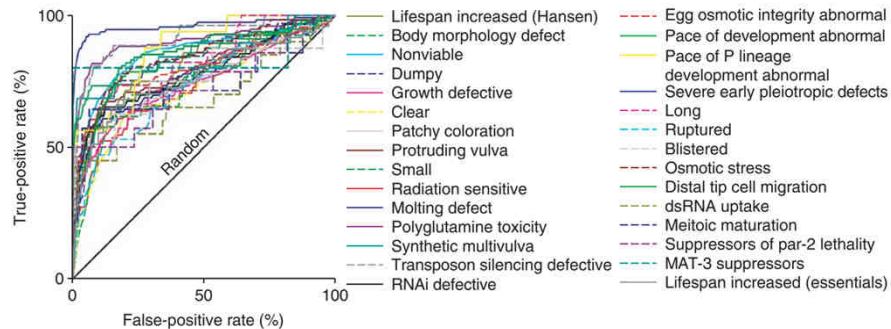
Basic idea: sort predictions from best to worst, plot TPR vs. FPR as you traverse the ranked list

$$\begin{aligned} \text{TPR} &= \text{TP} / P = \text{TP} / (\text{TP} + \text{FN}) \\ &= \text{True Positive Rate} \\ &= \text{Sensitivity, Recall} \end{aligned}$$

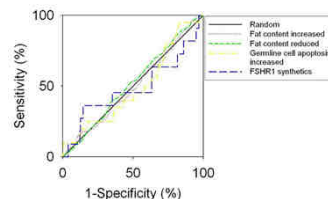
$$\begin{aligned} \text{FPR} &= \text{FP} / N = \text{FP} / (\text{FP} + \text{TN}) \\ &= \text{False Positive Rate} \\ &= 1 - \text{Specificity} \end{aligned}$$

Also useful to plot Precision [= $\text{TP} / (\text{TP} + \text{FP})$] vs. Recall (= TPR)

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 \rightarrow no better than random guessing.

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

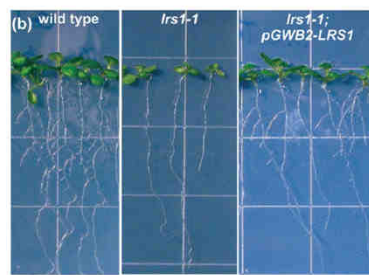
Remarkably, this strategy works quite well

Some examples of network-guided predictions:

In worms:

Genes that can reverse 'tumors' in a nematode model of tumorigenesis

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



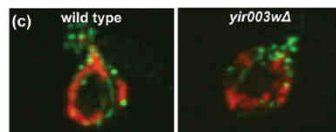
In Arabidopsis:

New genes regulating root formation

Lee, Ambaru *et al.*
Nature Biotech (2010)

In yeast: New mitochondrial biogenesis genes

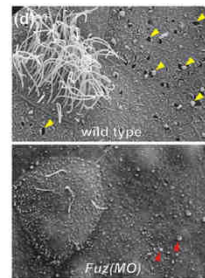
Hess *et al.*, *PLoS Genetics* (2009)



In mice/frogs:

Functions for a birth defect gene

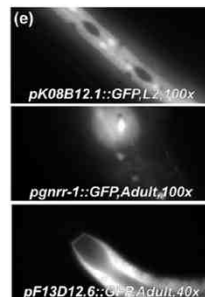
Gray *et al.*, *Nature Cell Biology* (2009)



In worms:

Predicting tissue specific gene expression

Chikina *et al.*, *PLoS Comp Biology* (2009)



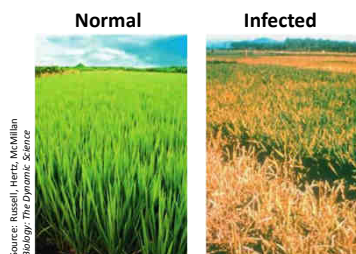
Reviewed in Wang & Marcotte, *J Proteomics* (2010)

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



Source: Russell, Hertz, McMillan
Biology: The Dynamic Science

Infection with *Xoo*

A new gene promoting resistance to rice blight



wild type resistant resistant + RNAi

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

Edward Marcotte/Univ. of Texas/BIOS337/Spring 2014

**Live demo of
functional networks
and Cytoscape**