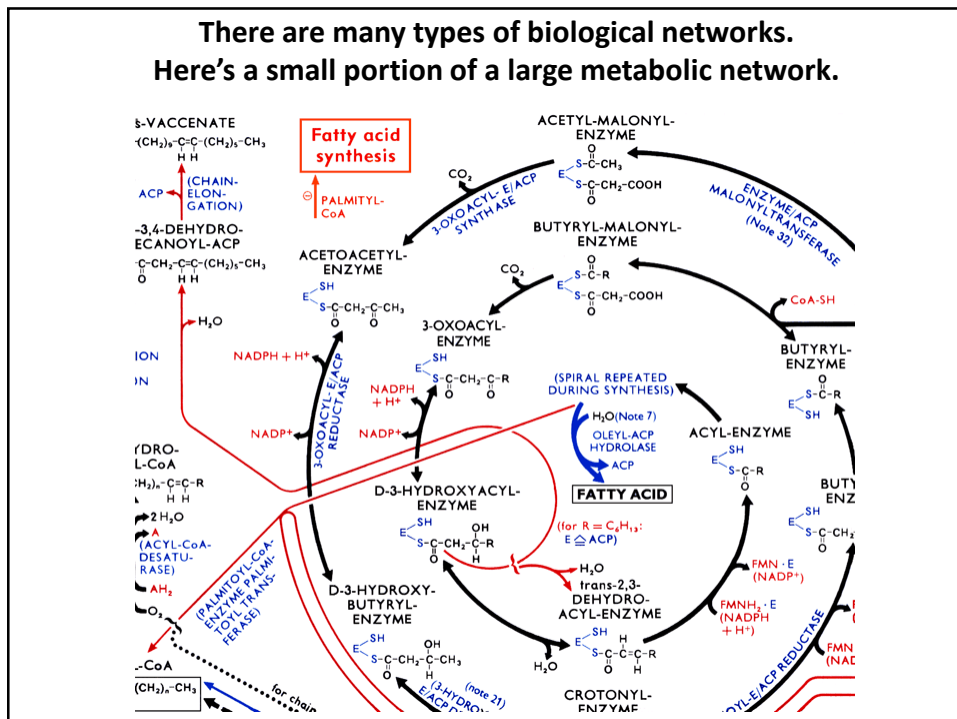


Network biology (& predicting gene function)

BCH394P/364C Systems Biology / Bioinformatics

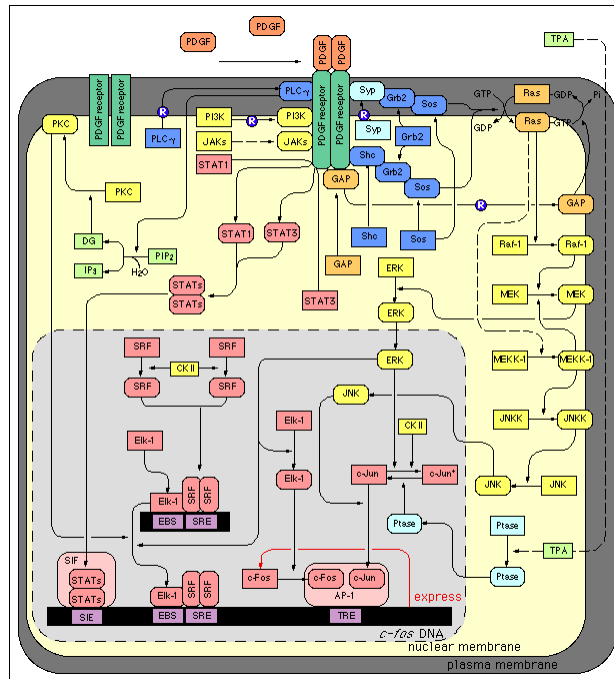
Edward Marcotte, Univ of Texas at Austin

1



2

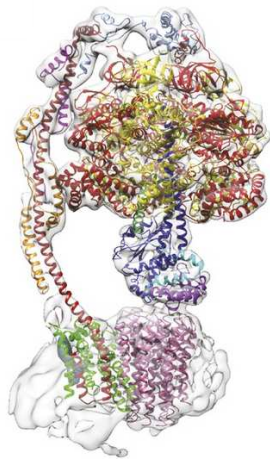
A typical genetic network



3

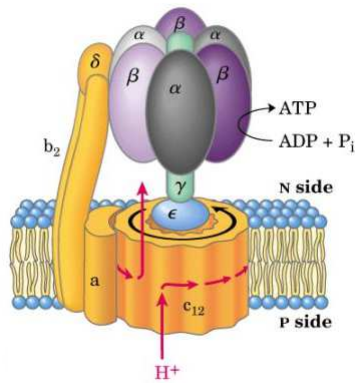
Contacts between proteins define protein interaction networks

CryoEM structure of ATP synthase



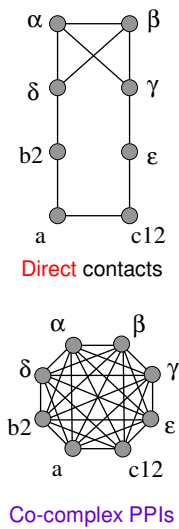
eLife 2015;4:e10180

Schematic version



Total set = protein complex
Sum of direct + indirect interactions

Network representation



4

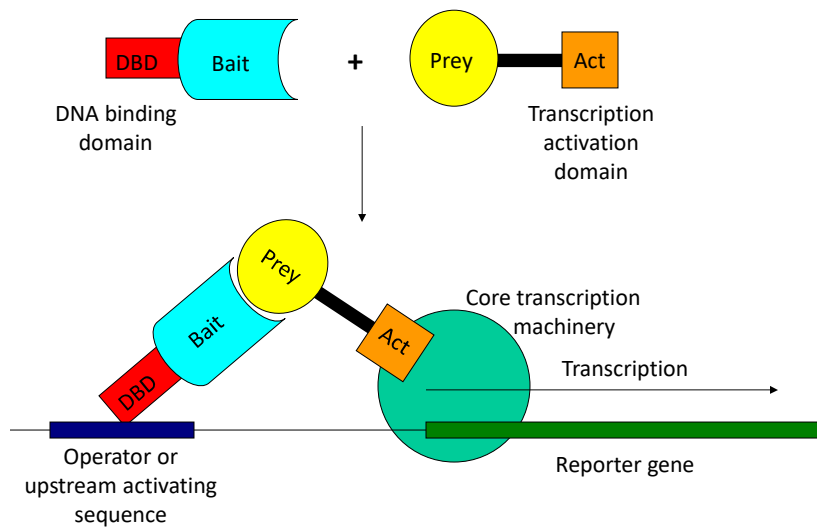
Let's look at some of the types of interaction data in more detail.

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

5

Pairwise protein interactions

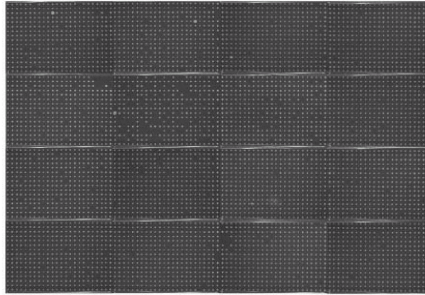
In general, purifying proteins one at a time, mixing them, and assaying for interactions is far too slow & laborious. We need something faster! Hence, high-throughput screens, e.g. yeast two-hybrid assays



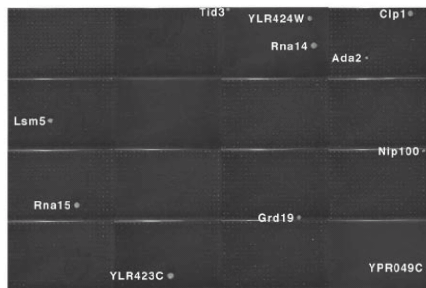
6

High-throughput yeast two-hybrid assays

Haploid yeast cells expressing activation domain-prey fusion proteins



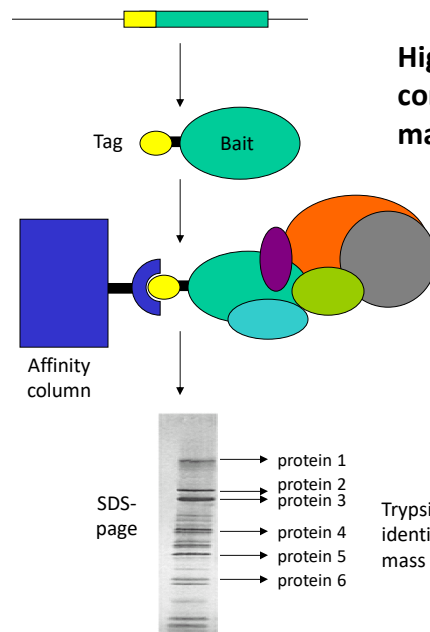
Diploid yeast probed with DNA-binding domain-Pcf11 bait fusion protein



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

7

Protein complexes



High-throughput complex mapping by mass spectrometry

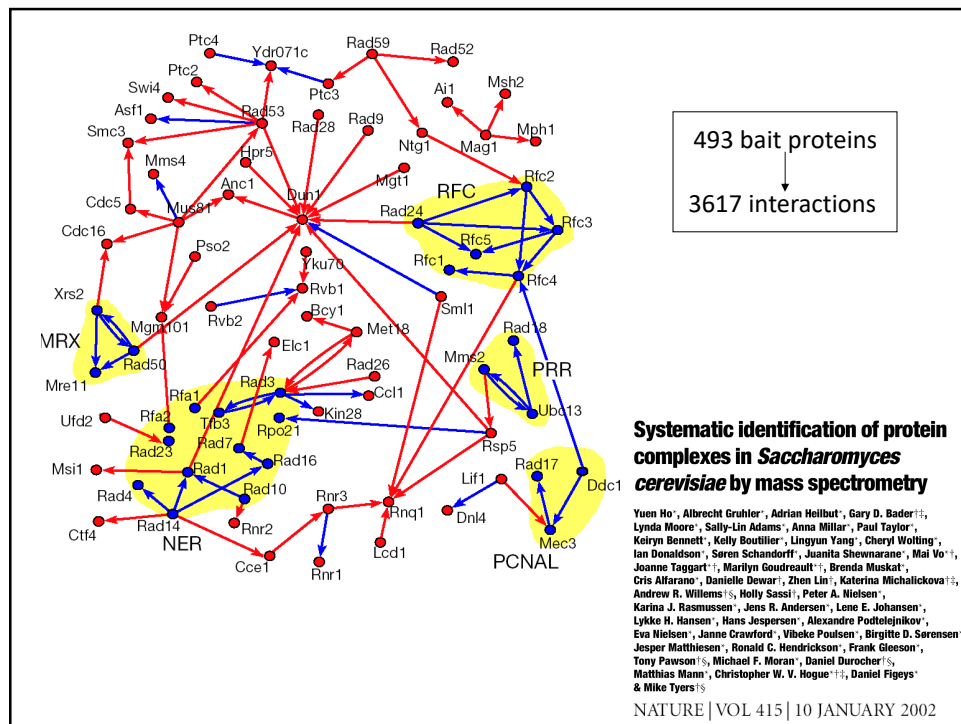
Affinity column

SDS-page

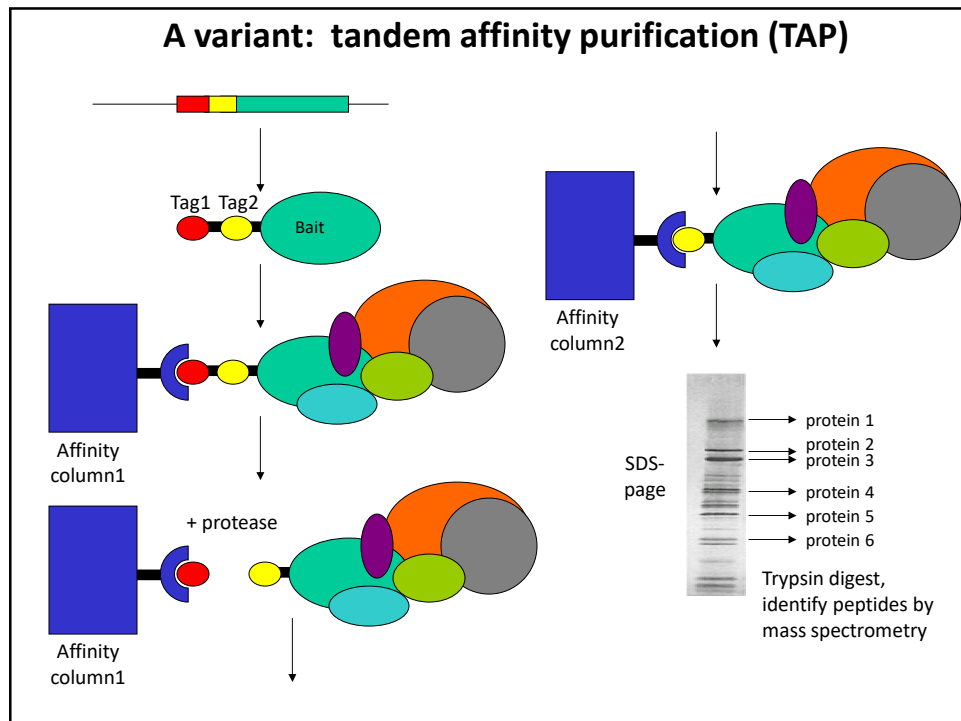
protein 1
protein 2
protein 3
protein 4
protein 5
protein 6

Trypsin digest, identify peptides by mass spectrometry

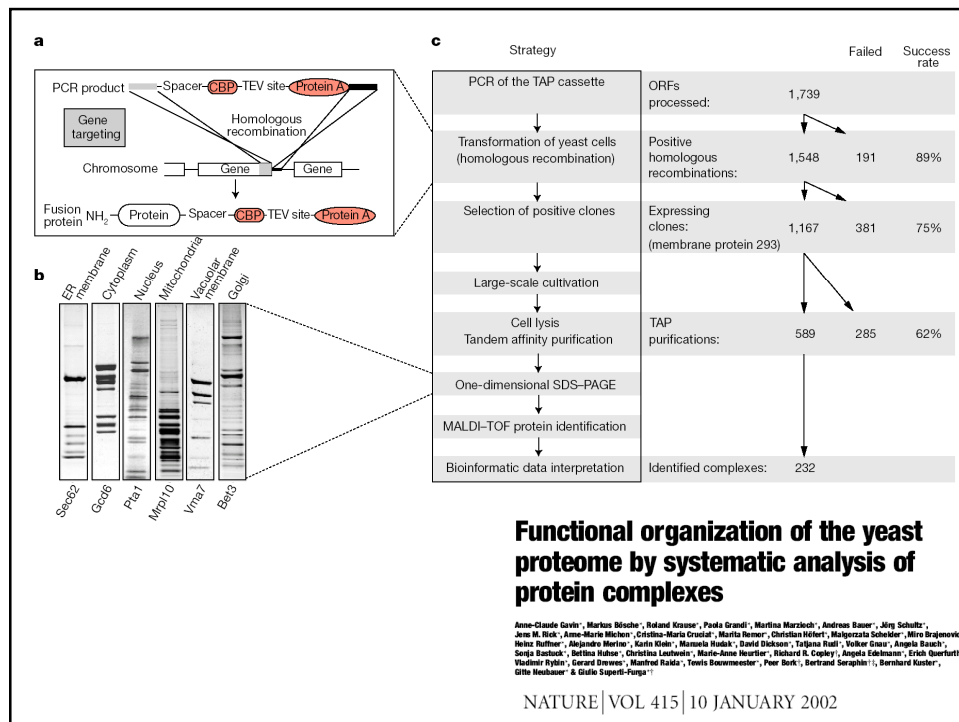
8



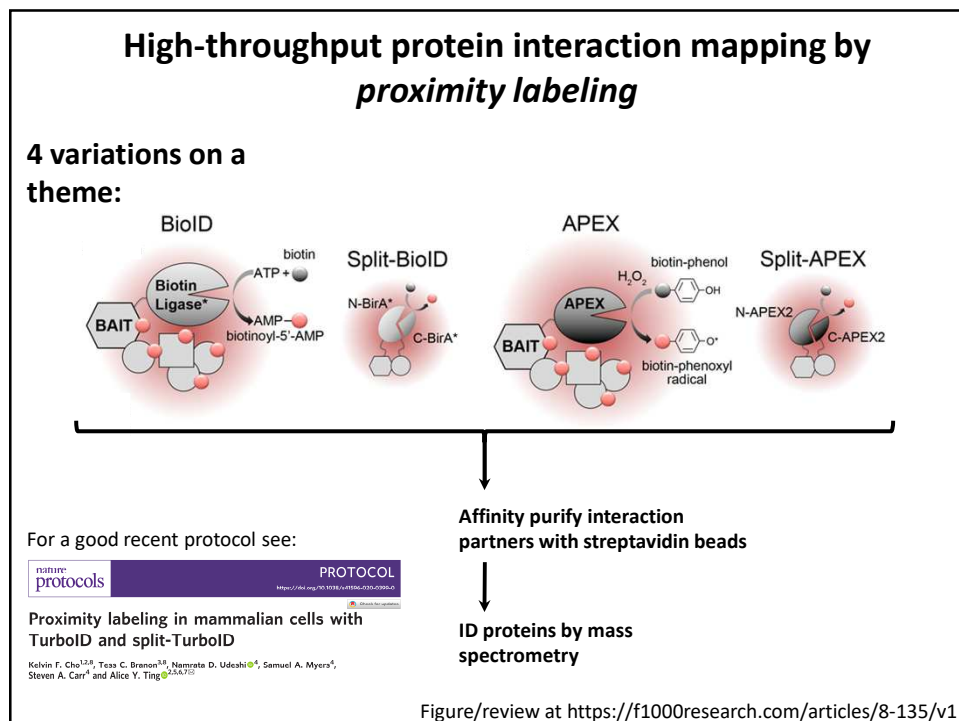
9



10

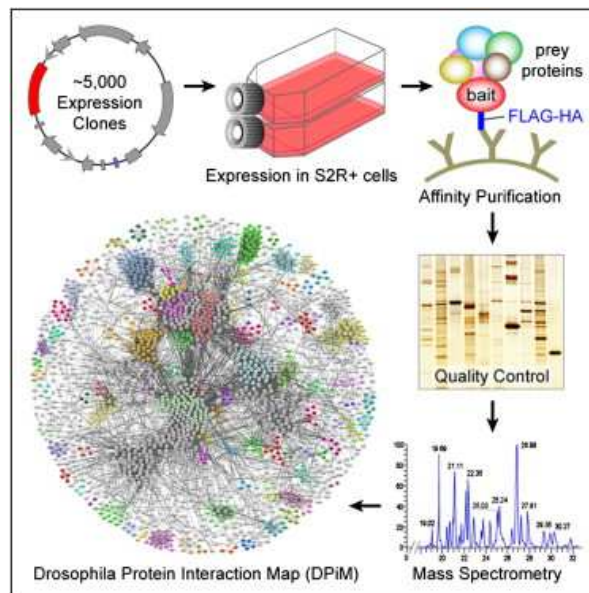


11



12

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



~3,500 affinity purification experiments

~11K interactions /
~2.3K proteins

→ spans 556 complexes

Still daunting for the
human proteome, but...

Guruharsha *et al.* (2011) *Cell* 147, 690–703

13

The current state-of-the-art in human PPI maps – Y2H

Human ORFeome (v9.1) → now ~90% of the protein-coding genes!

Screened *all x all* (150M pairs!) in 9 Y2H assays

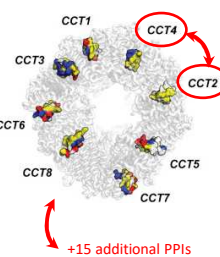
52,569 PPIs involving 8,275 proteins

<https://www.nature.com/articles/s41594-017-0016-2>



Y2H captures pairwise PPIs that can form when the proteins are expressed out of biological context (e.g., as fusion proteins in a yeast cell nucleus). It can reveal directly contacting proteins but often misses those that require additional molecular context or higher order assemblies,
← the exocyst e.g. the CCT complex →

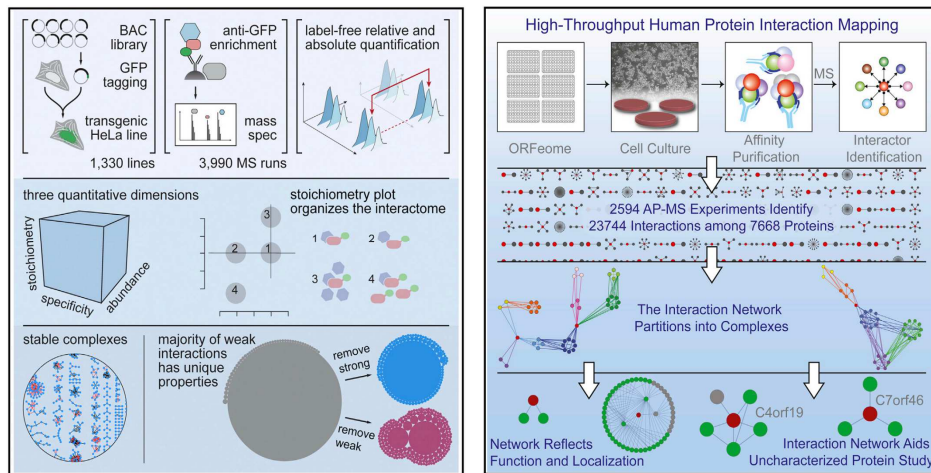
[https://www.cell.com/cell/fulltext/S0092-8674\(14\)01369-5](https://www.cell.com/cell/fulltext/S0092-8674(14)01369-5)



Luck *et al.*, A reference map of the human protein interactome, *bioRxiv*, posted April 10, 2019
<https://www.biorxiv.org/content/10.1101/605451v1>, published *Nature*, April 8, 2020

14

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



Hein *et al.*, *Cell* (2015) 163:712-23.

Huttlin *et al.*, *Cell* (2015) 162:425-440

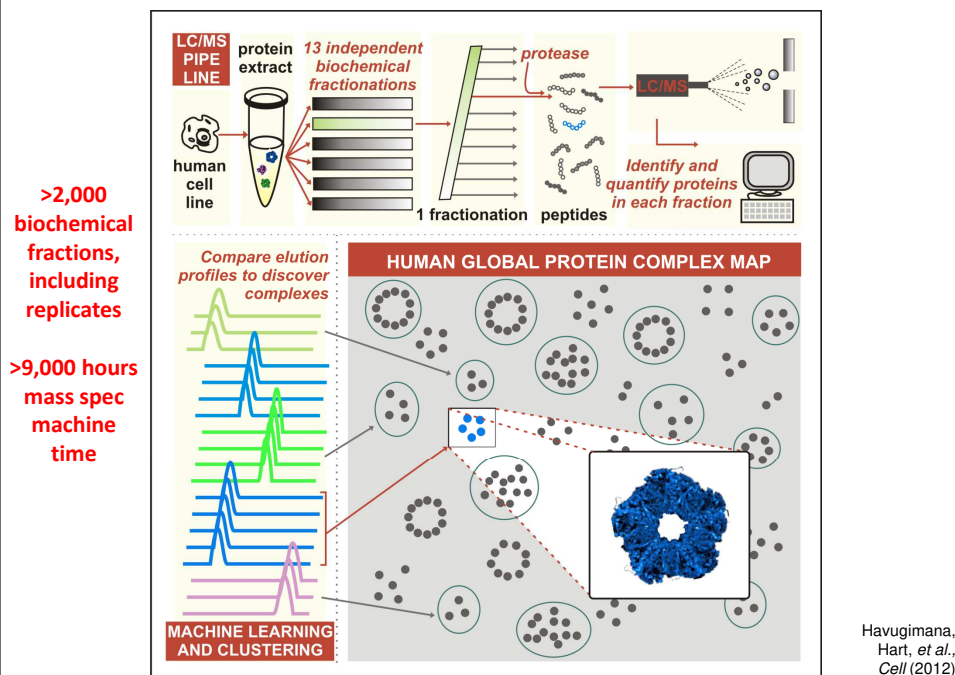
Huttlin *et al.*, *Nature* (2017) 545:505-509

Huttlin *et al.*, *bioRxiv* (2020)

Just in the past 5 years, >16K affinity purification/mass spec experiments on tagged human proteins expressed in cell lines

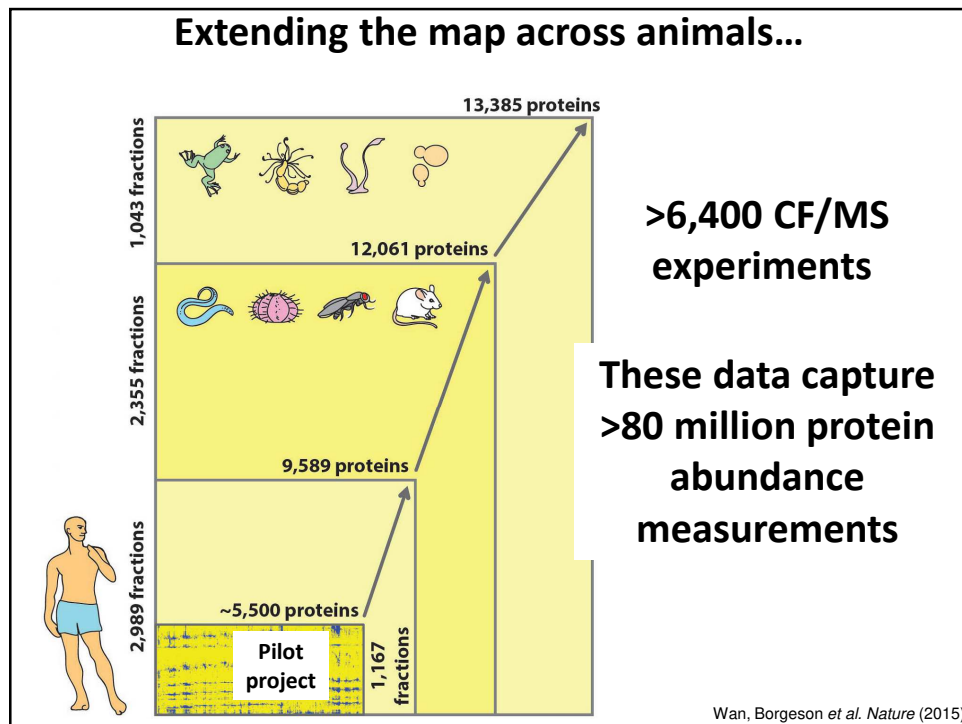
15

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

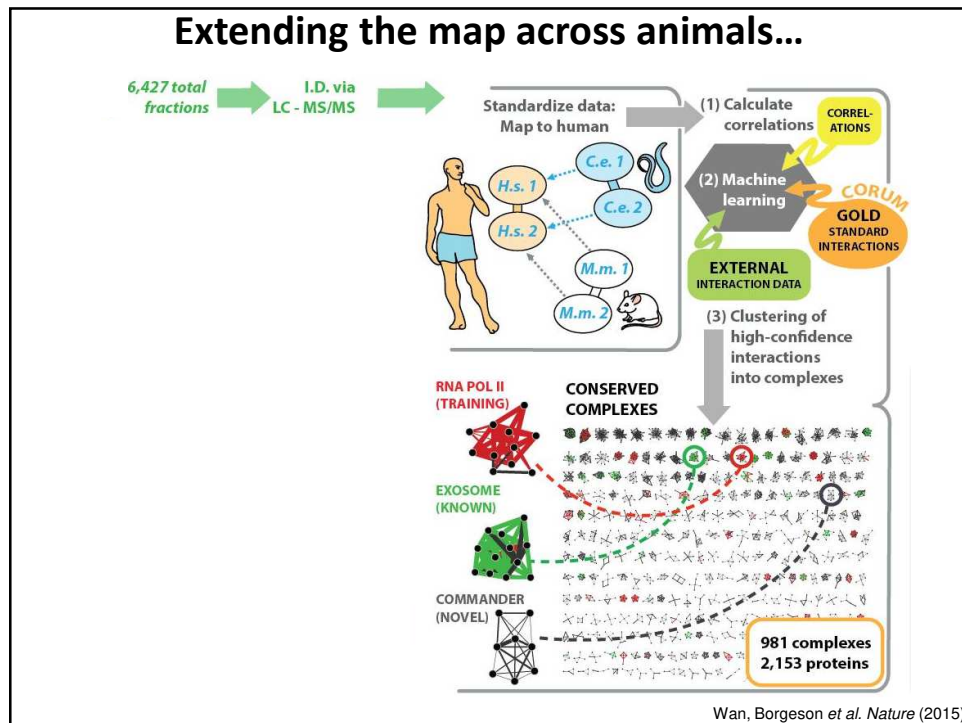


Havugimana,
Hart, *et al.*
Cell (2012)

16



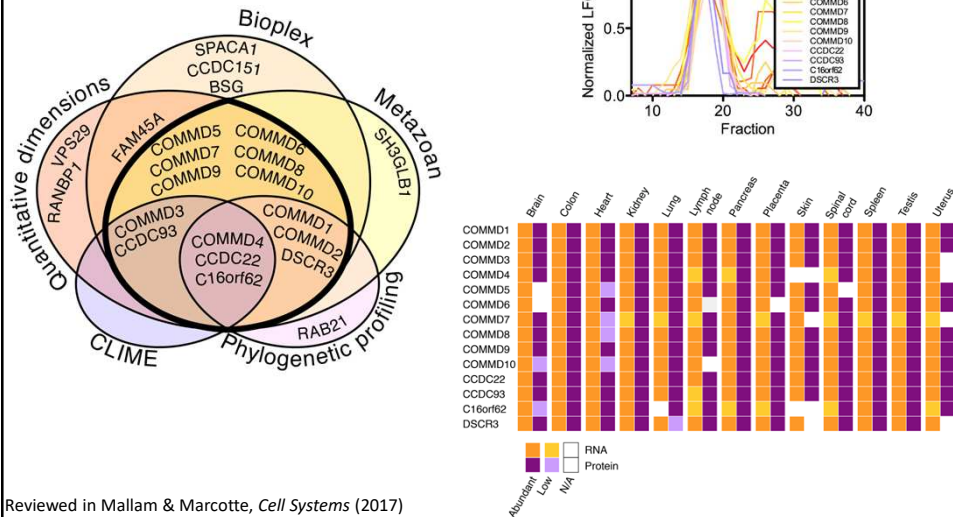
17



18

There are still lots of cellular machines left to find

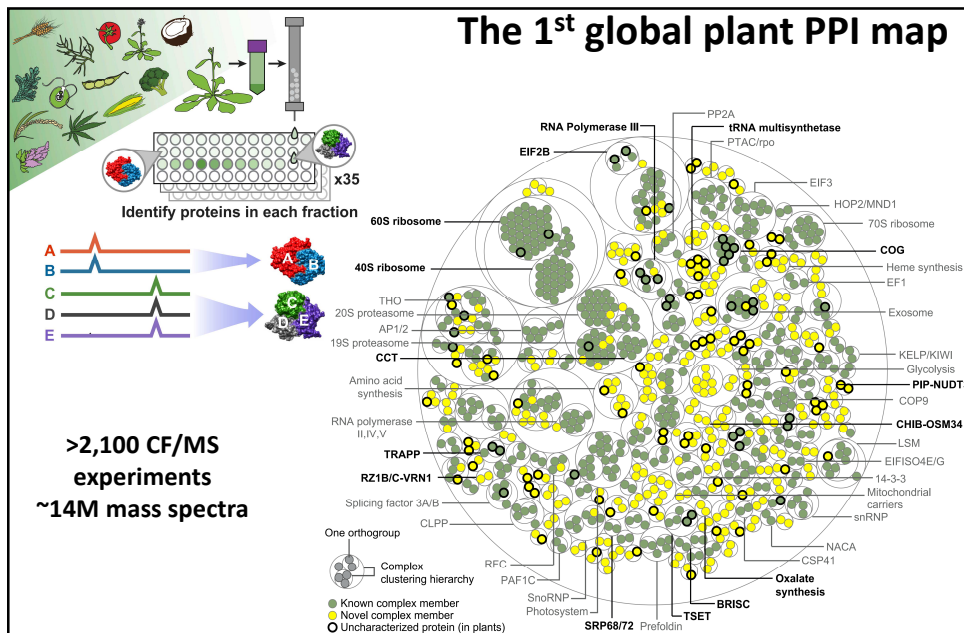
e.g. the “Commander” complex, found in all 3 large human PPI maps, a 600 kDa protein complex expressed in nearly every human cell type and tissue



Reviewed in Mallam & Marcotte, *Cell Systems* (2017)

19

The 1st global plant PPI map



A Pan-plant Protein Complex Map Reveals Deep Conservation and Novel Assemblies

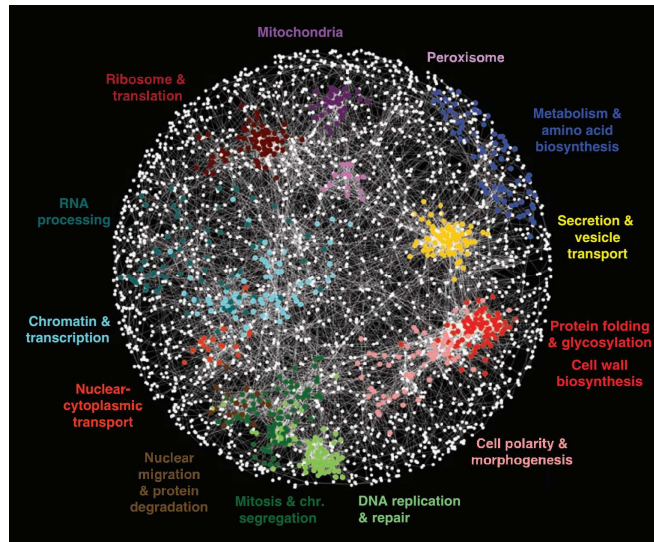
Claire D. McWhite,^{1,2} Ophelia Papoulas,^{1,2} Kevin Drew,¹ Rachael M. Cox,¹ Viviana June,¹ Oliver Xiaou Dong,^{3,4} Taejoon Kwon,⁵ Cuihong Wan,^{1,2} Mari L. Salmi,¹ Stanley J. Roux,¹ Karen S. Browning,¹ Z. Jeffrey Chen,¹ Ranaela C. Brown,^{1,2} and Edward M. Marcotte^{1,2}

Cell 181(2) P460-474.e14, April 16, 2020

20

Genetic interactions

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

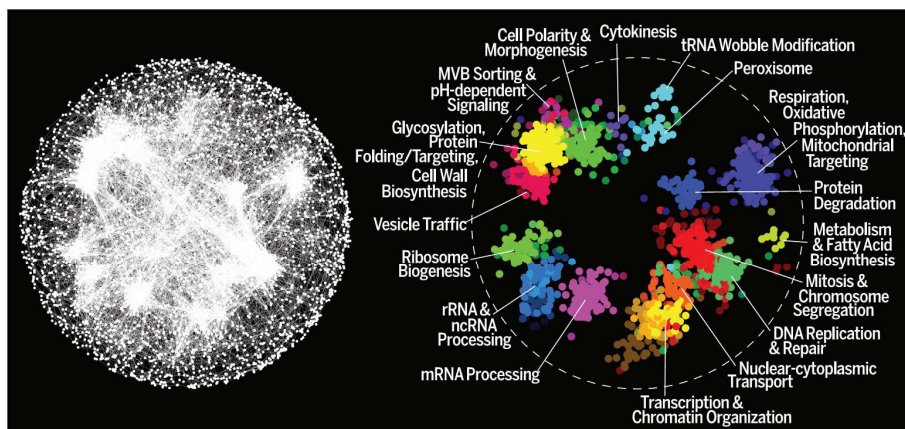


Costanzo *et al.*, *Science* 327: 425 (2010)

21

Genetic interactions, the 2016 version

23 million gene-gene pairs assayed for synthetic genetic interactions in yeast, identifying ~550,000 negative and ~350,000 positive genetic interactions

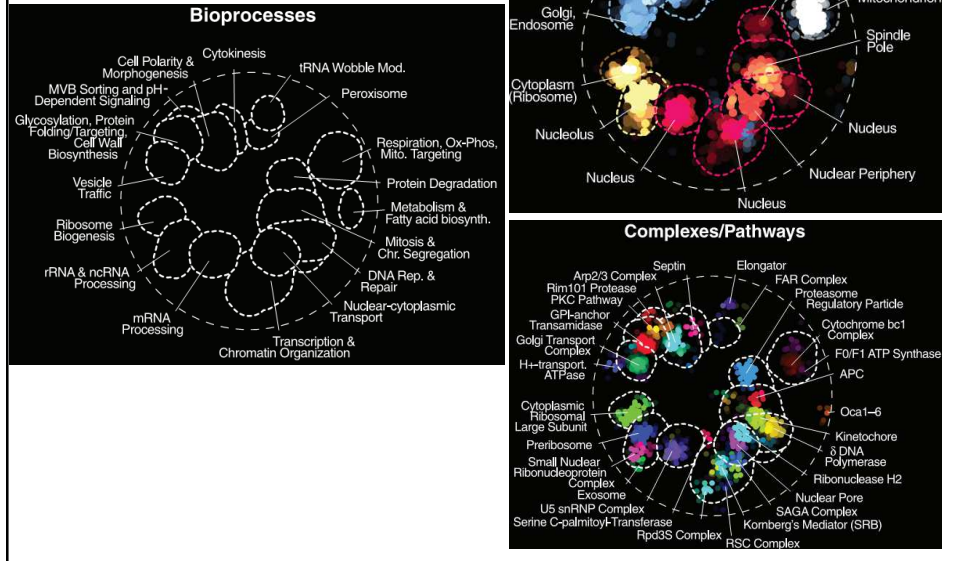


A global network of genetic interaction profile similarities. (Left) Genes with similar genetic interaction profiles are connected in a global network, such that genes exhibiting more similar profiles are located closer to each other, whereas genes with less similar profiles are positioned farther apart. **(Right)** Spatial

Costanzo *et al.*, *Science* 353: 1381 (2016)

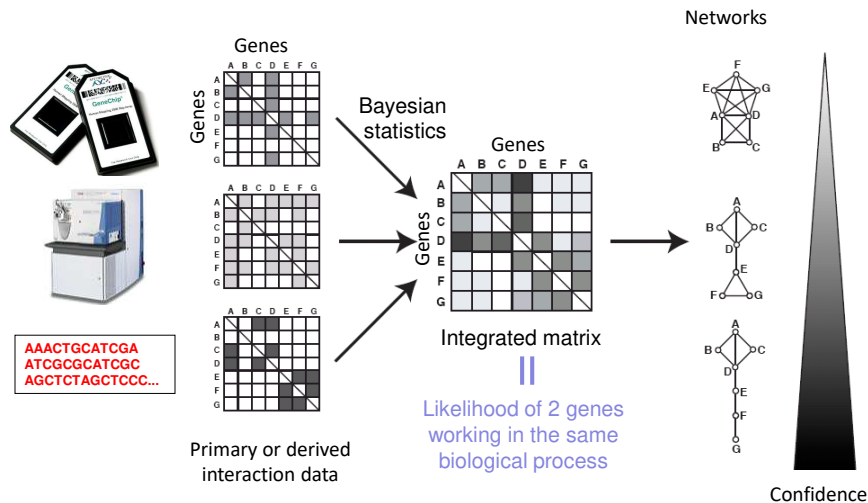
22

The global genetic interaction profile similarity network reveals a hierarchy of cellular function.



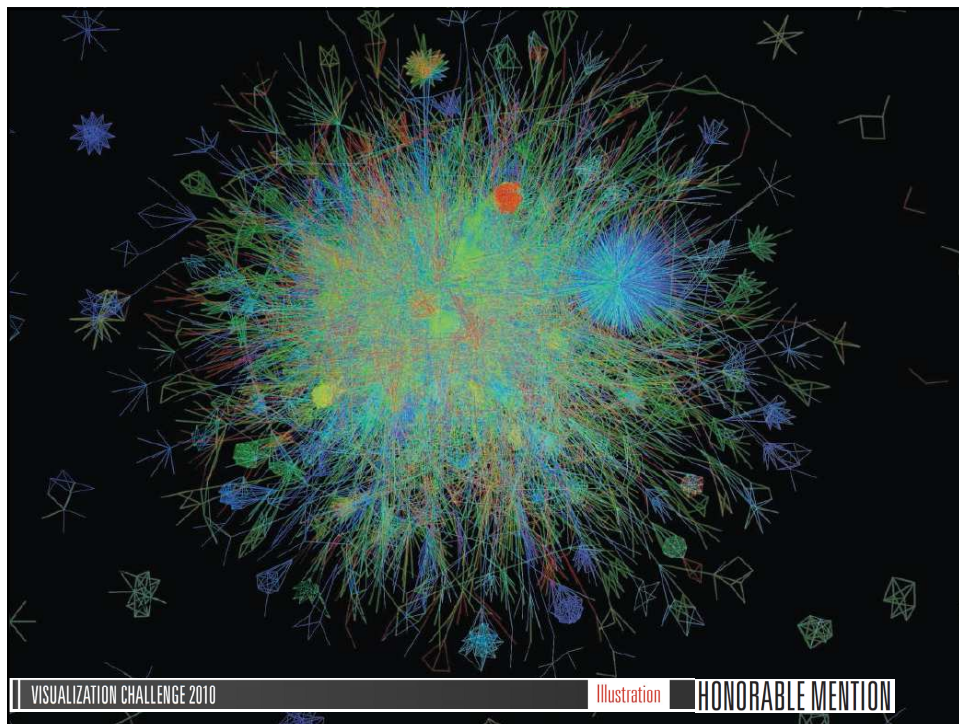
23

These sorts of data can be combined into functional gene networks

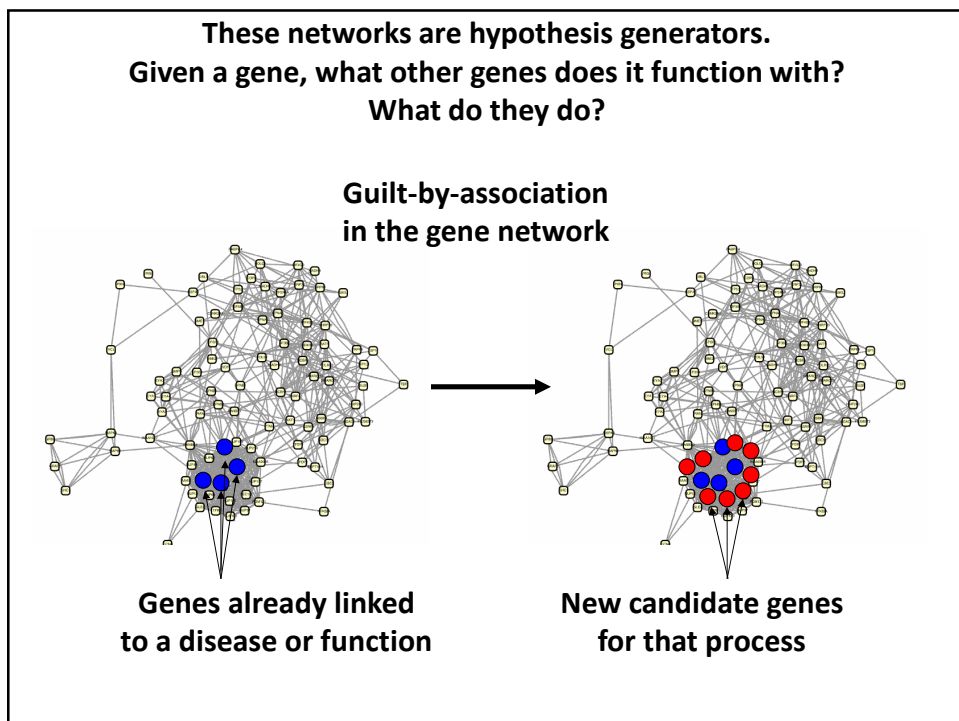


Adapted from Fraser & Marcotte, *Nature Genetics* (2004)

24

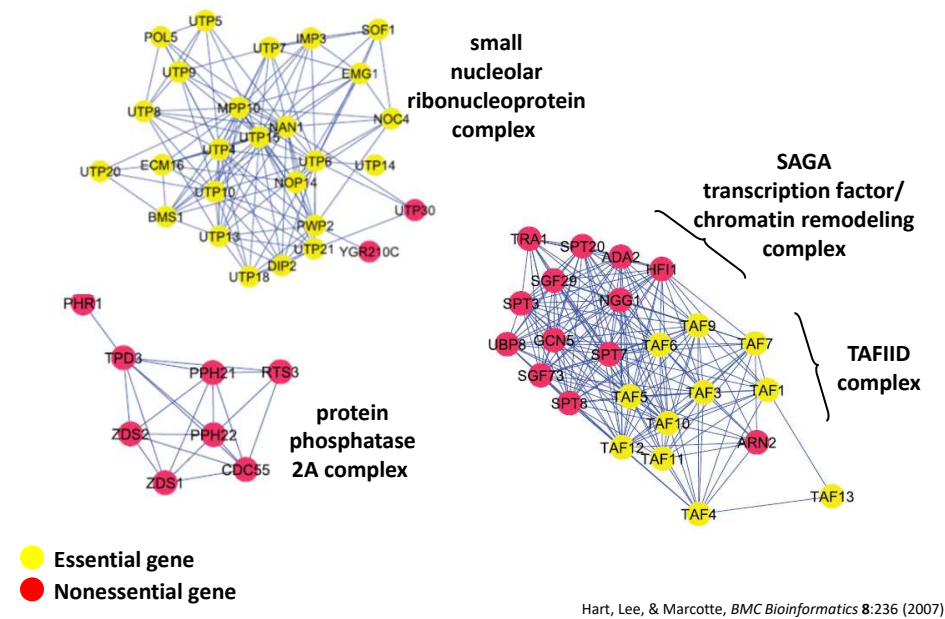


25



26

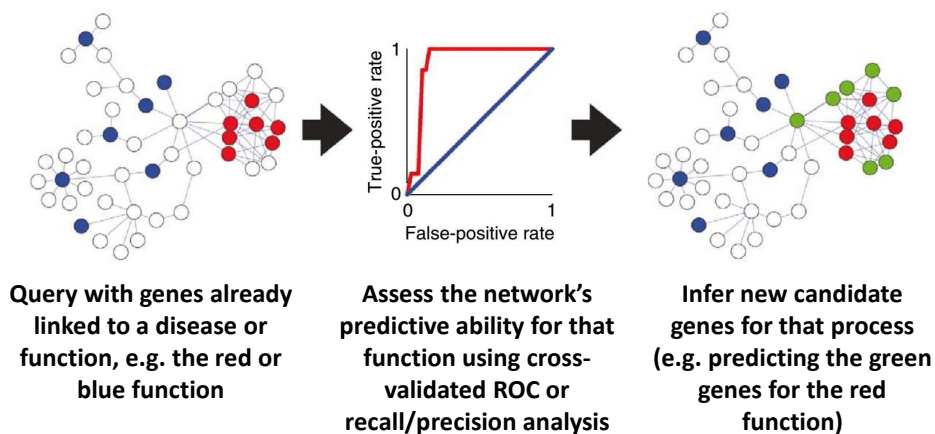
Gene networks frequently reflect functions, pathways, & phenotypes, e.g., lethality in yeast is linked to the molecular machine, not the gene



27

We can propagate annotations across the graph to infer new annotations for genes (network “guilt-by-association”, or GBA).

Testing how well this works on hidden, but known, cases let’s us measure how predictive it will be for new cases.



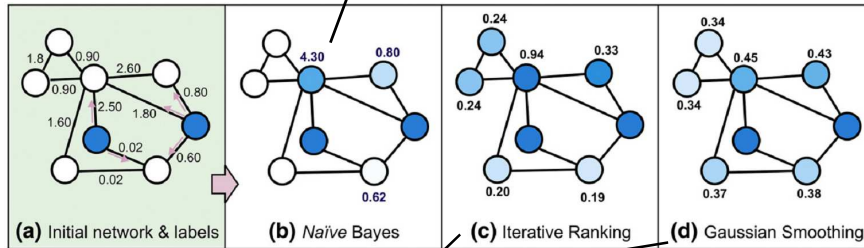
Lee, Ambaru et al. *Nature Biotechnology* 28:149-156 (2010)

28

Numerous algorithms exist for network GBA

Naïve Bayes assigns scores to neighboring nodes based on edges

Similar to Google's personalized PageRank



Network diffusion algorithms start with initial annotations and the graph topology, then propagate initial scores across the network, e.g. Gaussian smoothing tries to find scores:

$$f^{final} = \underset{f}{\operatorname{argmin}} \alpha \sum_i (f_i - f_i^0)^2 + (1 - \alpha) \sum_i \sum_j w_{ij} (f_i - f_j)^2$$

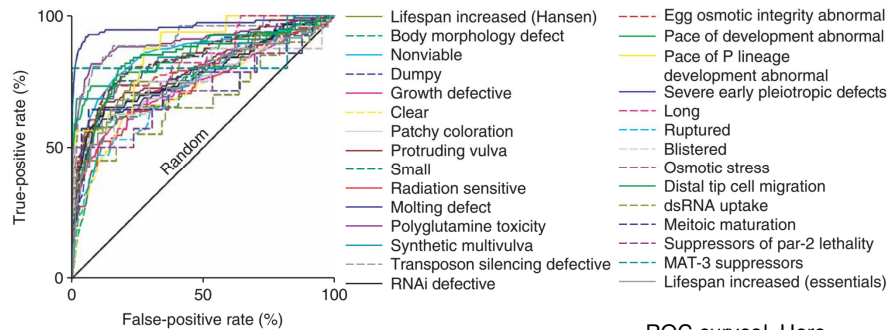
minimizing the difference between final and initial scores of a protein

& between a protein's score and that of each of its neighbors

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

29

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



ROC curves! Here, indicating the likely predictive power of the network for a system of interest, independent of how big the system is.

A poor ROC → no better than random guessing.

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

30

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)

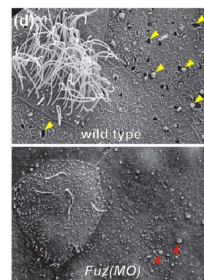
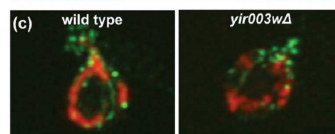
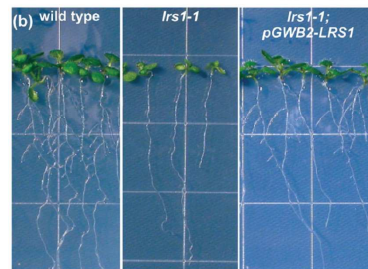
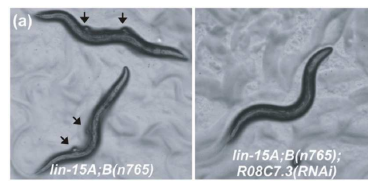
In Arabidopsis:

New genes regulating root formation

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

In yeast: New mitochondrial biogenesis genes

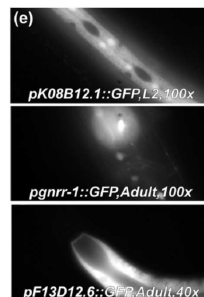
Hess *et al.*, *PLoS Genetics* (2009)



In mice/frogs:

Functions for a birth defect gene

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



In worms:

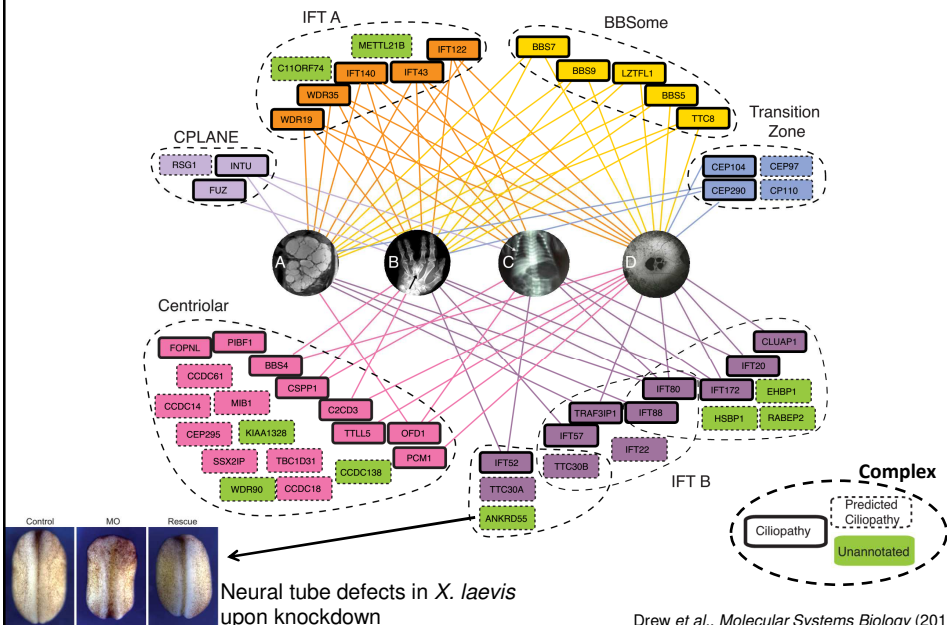
Predicting tissue specific gene expression

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

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

31

We use this approach routinely in the lab, e.g. an example of predicting new ciliopathy genes from protein complexes



32

The New York Times

Scientists Identify 69 Drugs to Test Against the Coronavirus

Two dozen of the medicines are already under investigation. Also on the list: chloroquine, a drug used to treat malaria.



A worker checking the production of chloroquine phosphate in China last month. There has been "anecdotal evidence" that chloroquine, a drug used to treat malaria, might work against the coronavirus. FeatureChina, via Associated Press



By Carl Zimmer

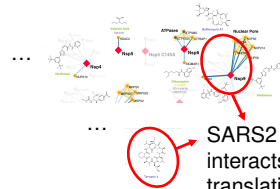
Published March 22, 2020 Updated April 9, 2020



Nearly 70 drugs and experimental compounds [may be effective in treating the coronavirus](#), a team of researchers reported on Sunday night.

A SARS-CoV-2 protein interaction map reveals targets for drug repurposing

Nature | Vol 583 | 16 July 2020



SARS2 Nsp9 interacts with translation initiation factors; drugs vs eIFs show activity



So, they tested a few more....

33

Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A

White *et al.*, *Science* **371**, 926–931 (2021)

Cancer drug derived from sea squirts outperforms remdesivir in COVID-19 preclinical models

By Angus Liu · Jan 27, 2021 07:40am

activity coronavirus COVID-19 drug repurposing



Scientists found PharmaMar's myeloma drug plitidepsin, derived from sea squirts, might be a more potent anti-COVID-19 drug than Gilead's remdesivir. (Quantitative Biosciences Institute)

→ Phase III clinical trials + data from compassionate use



Plitidepsin as a successful rescue treatment for prolonged viral SARS-CoV-2 replication in a patient with previous anti-CD20 monoclonal antibody-mediated B cell depletion and chronic lymphocytic leukemia

[Published: 10 January 2022](#)

[P. Guisado-Vasco et al. Journal of Hematology & Oncology 15, Article number: 4 \(2022\)](#)



PharmaMar shares soar after drug study suggests efficacy against Omicron

January 11, 2022
7:22 AM CST

Reuters

34

**Live demo of
STRING, BioGRID,
GeneMania,
functional networks
and Cytoscape**

35