BIOC4004 - Industrial Biochemistry
Lecture 22 - Wed Mar 24, 04

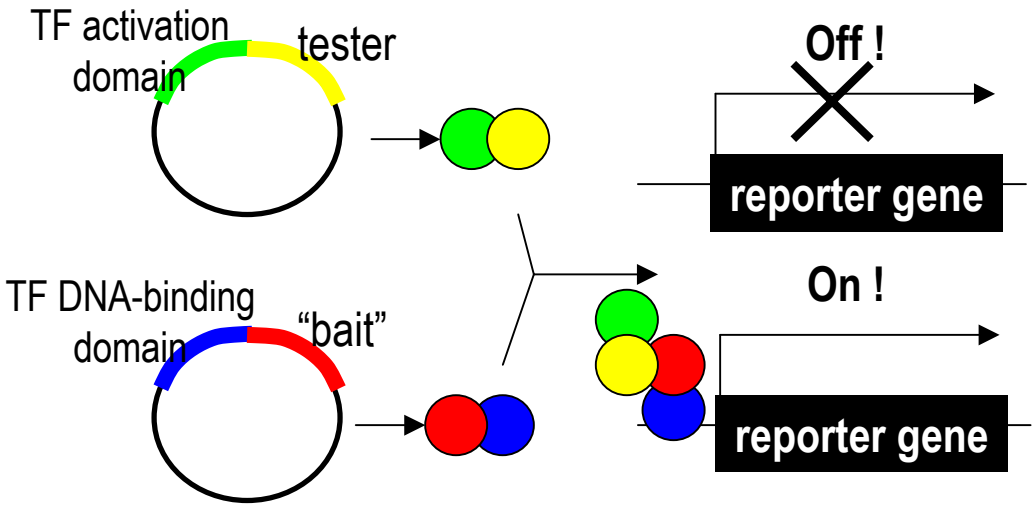
Topics for the Day:

- Functional Genomics
  - revisiting yeast two-hybrid assay
    … genomics style
  - RNA interference and functional genomics
  - Clinical Genomics

Yeast Two-hybrid Assay, genomics style (Pt.1)

- A fundamental question regarding all the genomics data:
  - What do the unknown proteins do ?
  - Maybe we can get some info from their interaction with other proteins

Yeast Two-hybrid Assay



- used to seek protein-protein interactions
- test a library of tester proteins vs a "bait" fusion protein
- interaction brings DNA-binding and activation domains together: activation
- look to see which tester proteins activate transcription

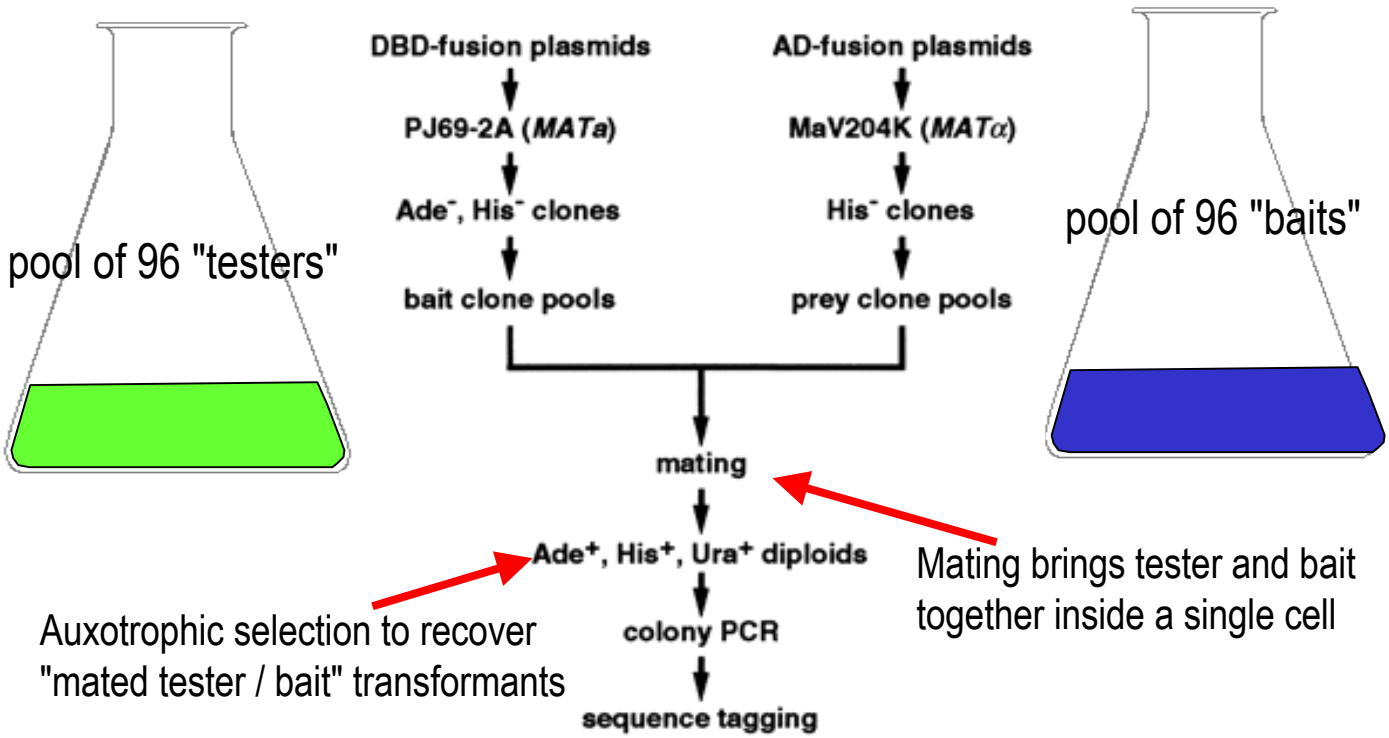
Genome-wide detection of protein-protein interactions (ie. the "interactome")

- Yeast has 6000 proteins:
  checking all possible interactions would require 36 million tests

- Uetz et. al. (2000)
  - made a library containing each of yeast's 6000+ ORFs fused to GAL4 activator domain
  - tested ~ 200 ORFs fused to the GAL4 DNA-binding domain
  - tested one "tester" at a time….how inefficient !!!!!
- CuraGen
  - pooled the 6000 ORFs to speed up the testing process

Yeast Two-hybrid Assay, genomics style (Pt.2)

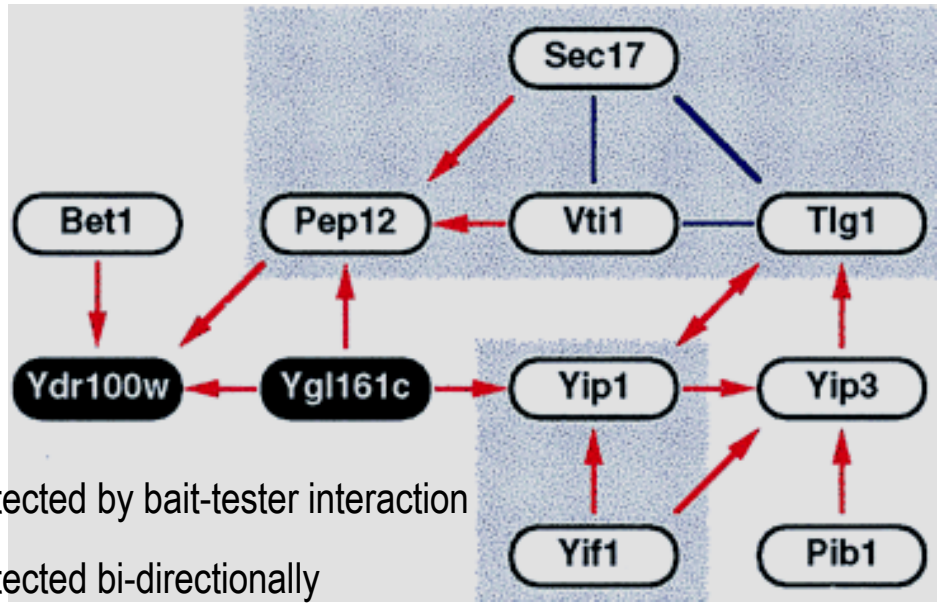
- **Ito et al. (2001)**
  - made a library of testers and a library of baits
  - split each library into pools of 96 clones
  - tested pools of testers vs pools of baits
  - tested 3900 different combinations of tester / bait pools
    - accounts for ~ 36 million single protein combinations



- picked > 100000 colonies per pair of pools tested to cover 96 x 96 potential interactions
- transferred onto plates with substrate for reporter gene
- tested three different promoter/reporter gene combos
  - looked for colonies that gave activation in all three reporter genes
- amplified sequences from "positive" clones
  - BLAST sequence against yeast Dbase to identify genes
- detected ~ 4500 interactions
  - 841 showed up three or more times

Yeast Two-hybrid Assay, genomics style (Pt.3)

A Protein Interaction Network: vesicular transport



Detected by bait-tester interaction

Detected bi-directionally

Known interactions from literature not detected by screen

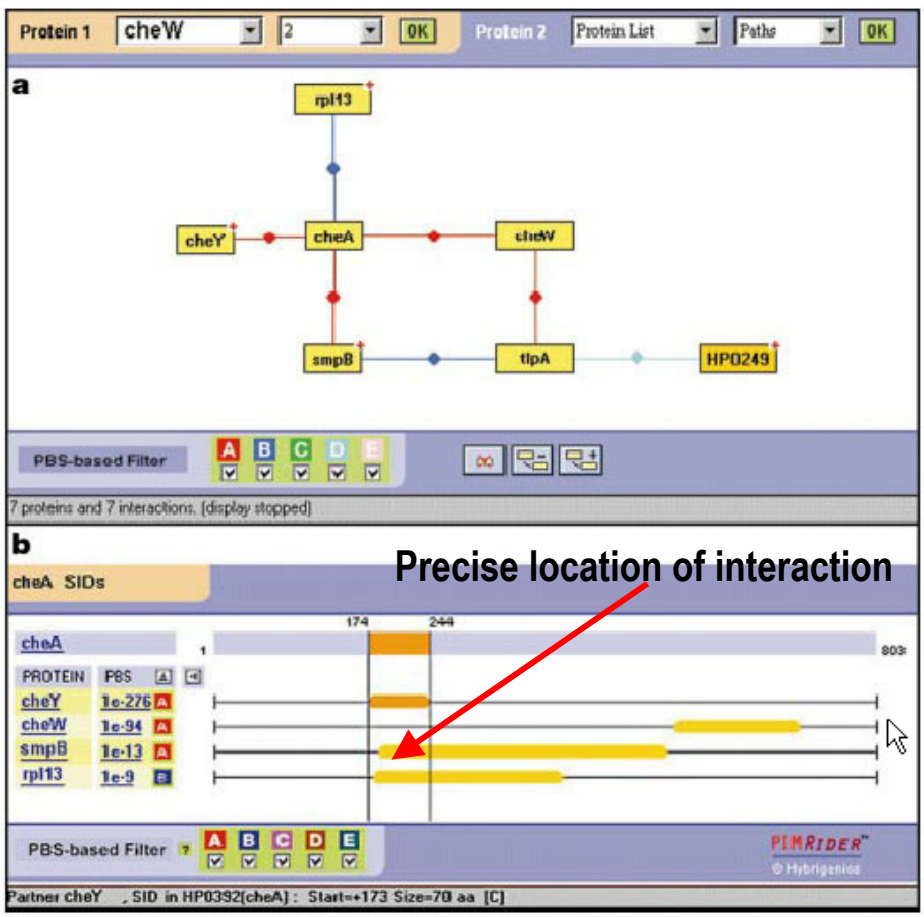
Proteins of unknown function

Previously described networks or complexes

- Problem with false positives
  - some proteins can activate without needing a partner
  - some proteins can not possibly interact if they are in different sub-cellular compartments
  - fluke interactions
- Problem with false negatives that are expected to interact
  - mutations during library construction leading to loss of interaction
  - some proteins require “activation” in order to interact with partner
  - incorrect folding in the wrong cellular compartment
  - some partners only interact through a "third-party"
- We expect ~ 50000 interactions, only 5000 detected so far
  - this is still far from comprehensive but it's a start….

Yeast Two-hybrid Assay, genomics style (Pt.4)

- **Rain et al. (2001)**
  - performed two-hybrid on protein fragments instead of full-length proteins
    - eliminate parts of the protein that do not interact
    - direct link to the functional region responsible for interaction
    - the fragments can be put back into cells to knock-out function
- How ?
  - Made random shotgun library of "tester" in yeast shuttle vector
    - Picked 2 million clones
  - Tested selected baits against the library
  - Positives were analyzed by sequencing
  - Discard clones that:
    - do not code for sequence
    - are in the wrong frame
  - Compared sequences to determine if different fragments belonged to same protein
    - the shared sequences indicate "intaracting domain"



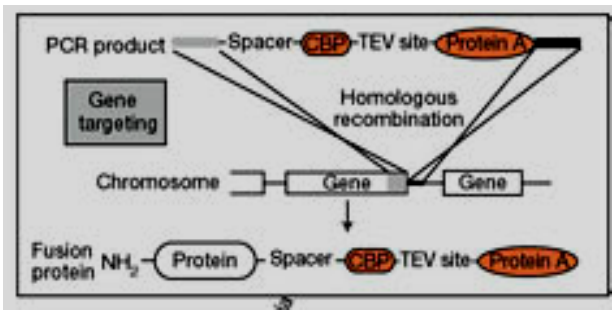
285 baits were tested

- ~14000 positives were sent for sequencing
- ~ 2700 interacting domains
  - ~1100 were non-ORFs
- Some data makes a lot of sense: interaction of operons
- Compared to known E.coli interaction data
  - good correlation
- can assign function to ORFs that are not too conserved sequence-wise
- ~1300 interactions detected (~3.4 per protein)

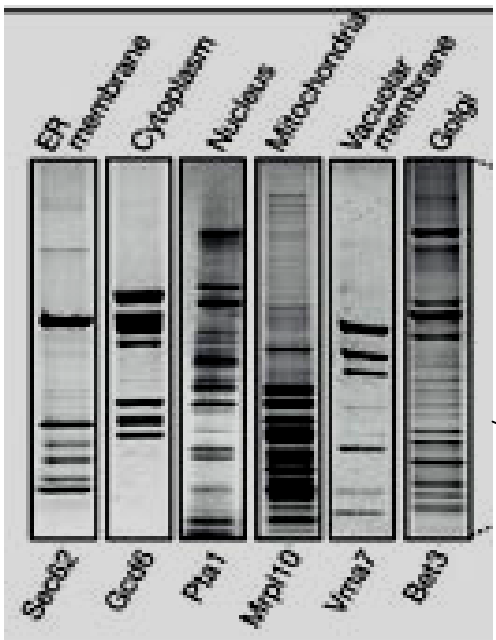
Yeast Two-hybrid Assay, genomics style (Pt.5)

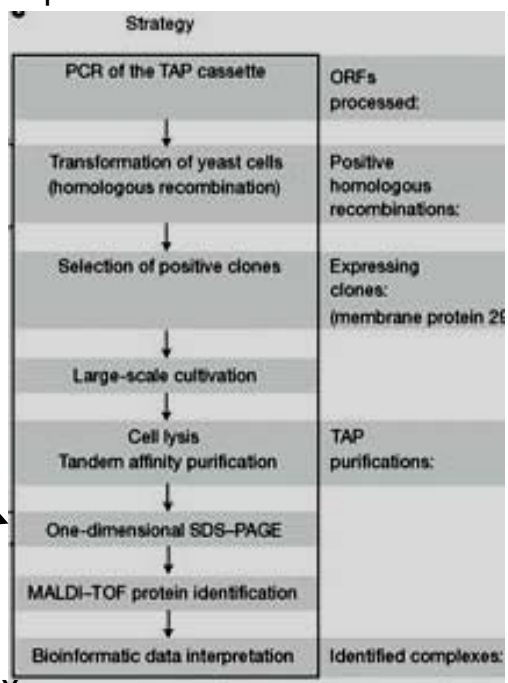
- **Gavin et al. (2001) (CellZome)**
  - didn't use two-hybrid assay at all !!!
  - They did use the concept of a fusion protein to study protein-protein interactions

- How ?
  - Tandem Affinity Purification (TAP) using TAP tag:
    - Protein A (Staph aureus) --> 1st affinity tag(IgG sepharose)
    - Protease cleavage site --> release from 1st affinity tag
    - Calmodulin binding peptide --> 2nd affinity tag (Calmodulin sepharose)



  - Produced library of ~ 1500 TAP-tagged yeast proteins (in yeast !)
  - Purify TAP-tagged proteins along with whatever proteins associate with them



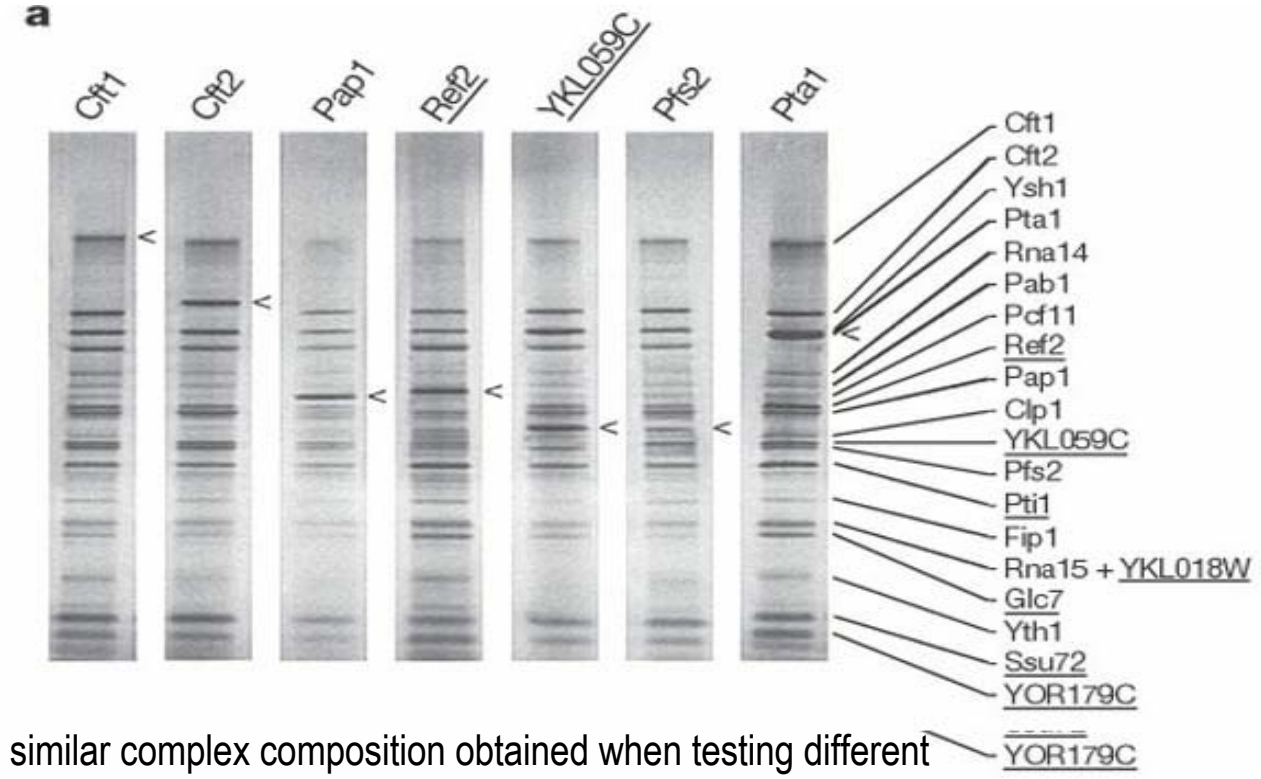


- Use MALDI-TOF MS to get identify proteins in complex
- Compared against yeast ORF database
  - detected interactions for ~80% of the proteins tested
  - average of 12 components per complex

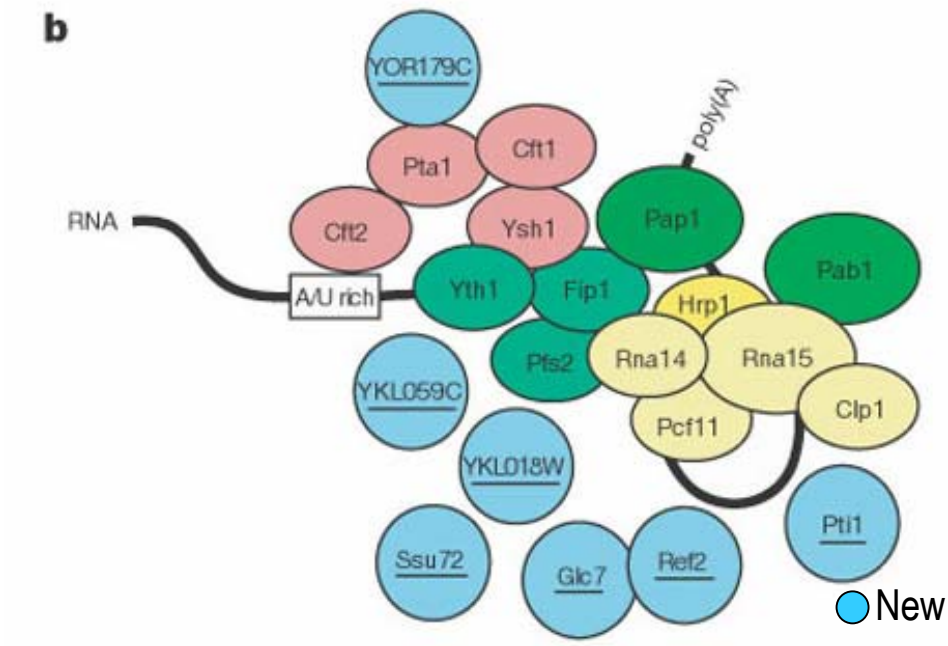
Yeast Two-hybrid Assay, genomics style (Pt.6)

- **Gavin et al. (2001) (CellZome) - continued**

Validation of Polyadenylation complex



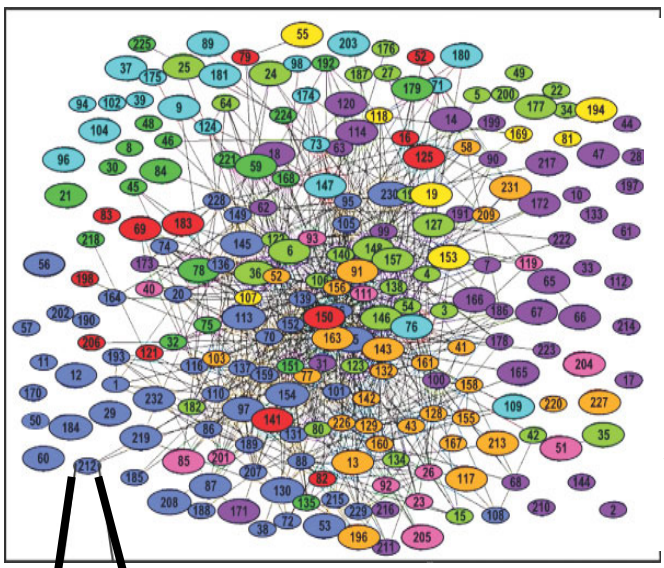
- similar complex composition obtained when testing different TAP-tagged proteins from the complex



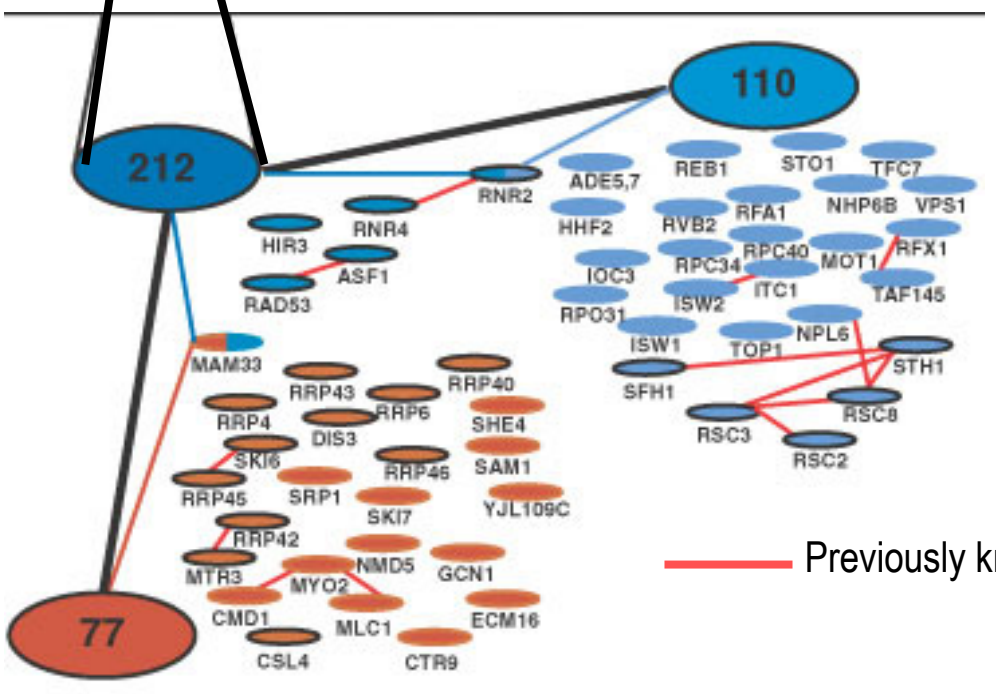
New components found

Yeast Two-hybrid Assay, genomics style (Pt.7)

Yeast Partial Interactome



red, cell cycle
dark green, signalling
dark blue, transcription, DNA maintenance, chromatin structure
pink, protein and RNA transport
orange, RNA metabolism
light green, protein synthesis and turnover
brown, cell polarity and structure
violet, intermediate and energy metabolism
light blue, membrane biogenesis and traffic.

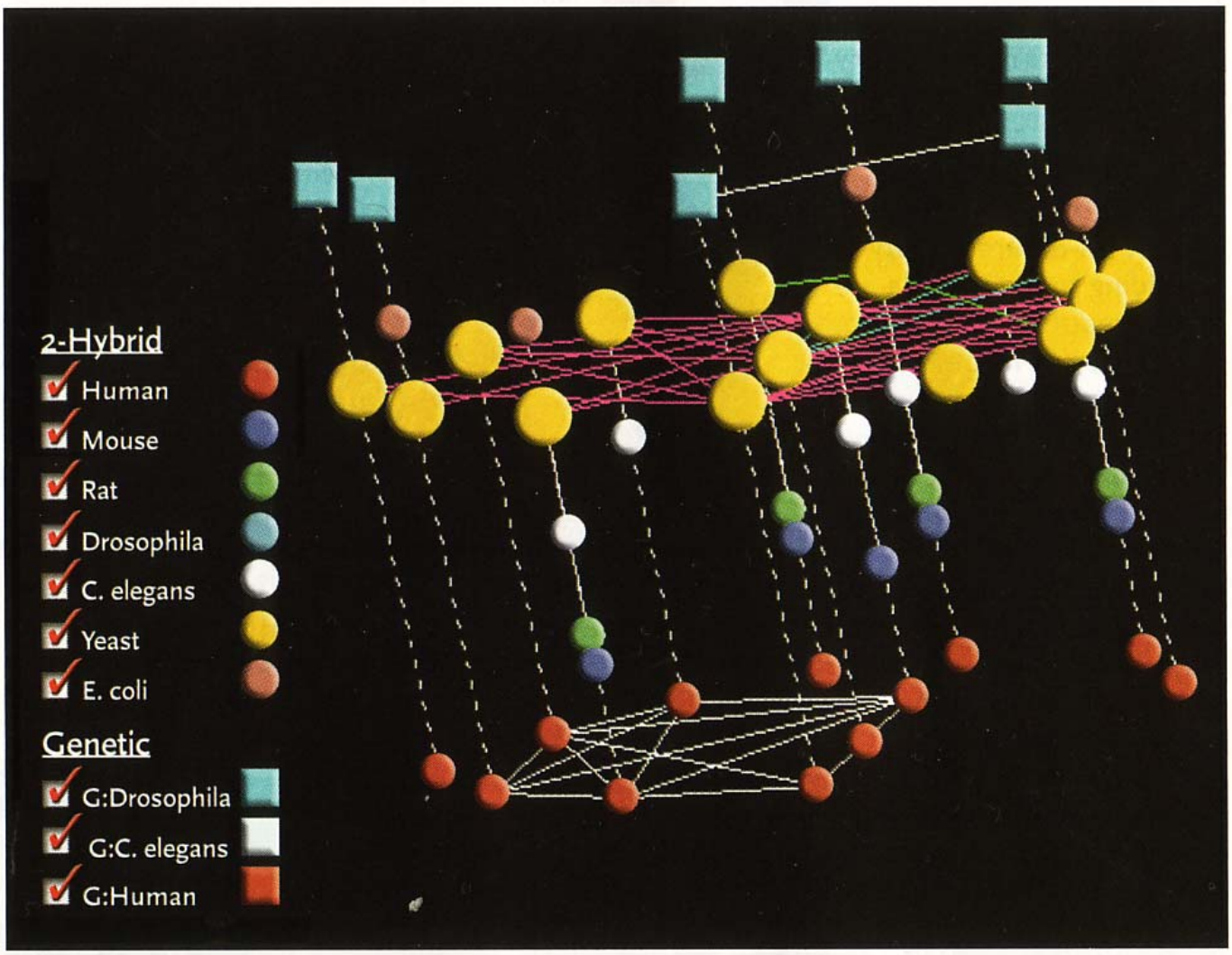


- Many complexes link to other complexes by shared components
- Connections in this network not only reflect physical interaction of complexes, but may also represent common regulation, localization, turnover or architecture
- The more connected a complex, the more central its position in the network.

Protein Networks and Evolution

• Yeast two-hybrid screens are notorious for producing false positives

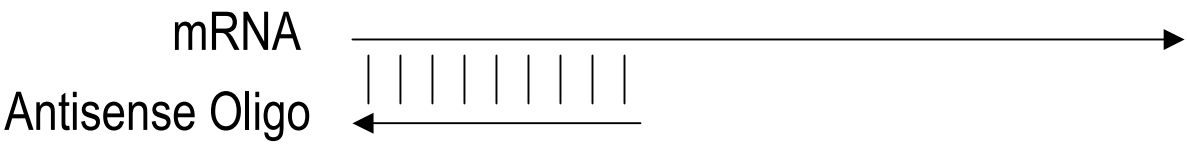
• As we do this work at the genome level, thousands of fake interactions can be generated:
    • need other methods to weed out false positives
• **One way**: generate interaction data from other organisms, look for conserved genes that also interact



BUDDING NEW RELATIONSHIPS: *Homolog view of the yeast pathway as compared to other species including human and* Drosophila

Gene Silencing using siRNAs

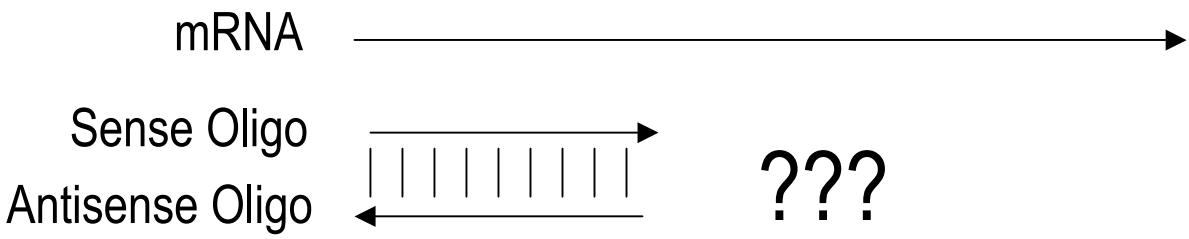
- **Ultimately, will need to validate protein function**
  - One method of validation is to produce gene knock-outs
  - Relatively easy to do in bacteria and some yeasts (homologous recombination)
  - Very difficult to generate knock-outs in higher organisms
    - same reasons as when we were discussing stable transfection
      - difficulty getting the DNA in
      - rare events
      - Etc…
  - Early on, Antisense technology was seen as a potential method



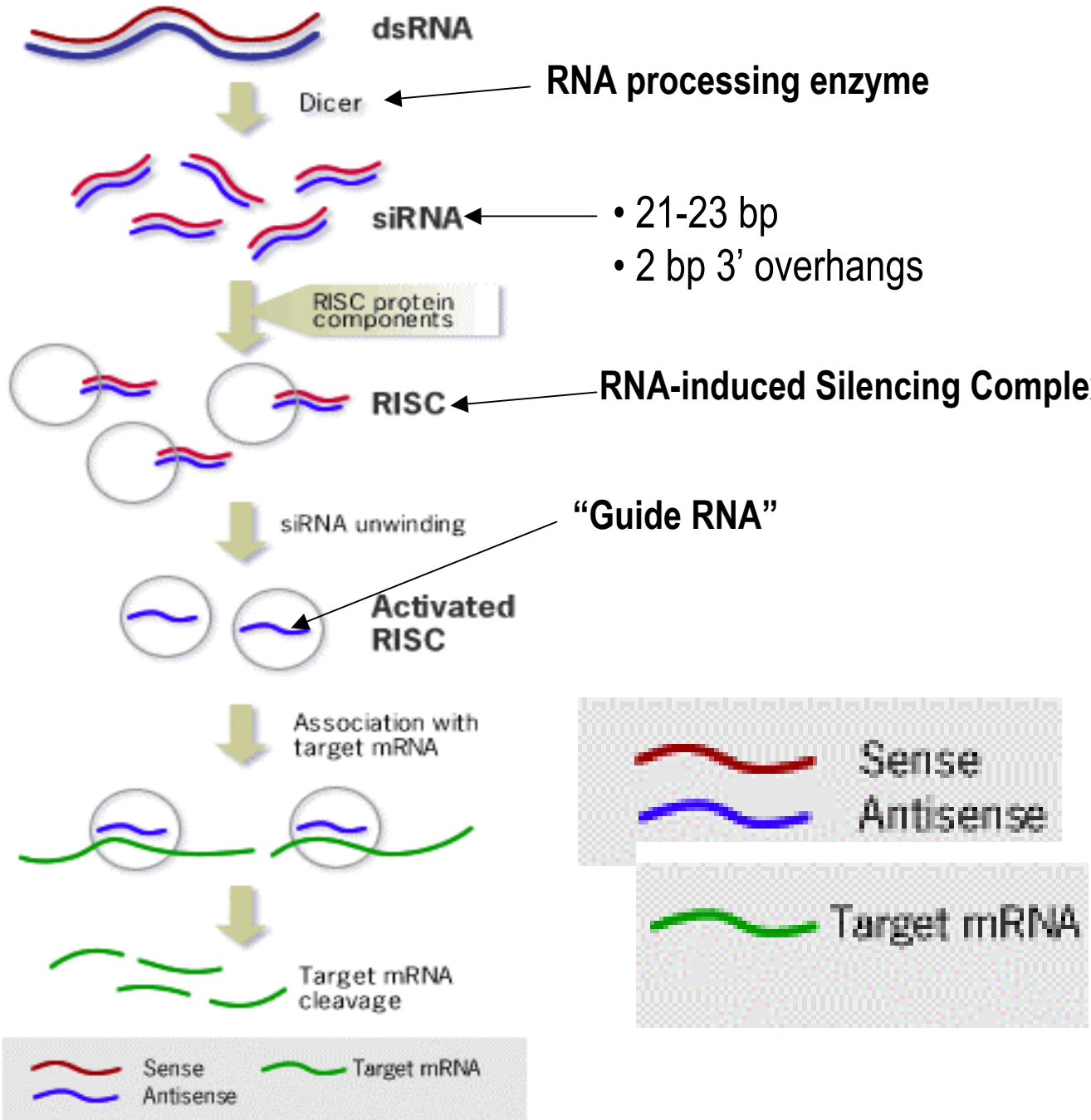
Impairment of translation…in theory

Fire and Mello (1998)

- trying RNA oligo antisense knock-outs
- test : anti-sense RNA oligo
  - → led to decrease
- control : anti-sense RNA oligo and sense RNA oligo
  - → led to <u>even bigger</u> decrease



Mechanism of RNA Interference



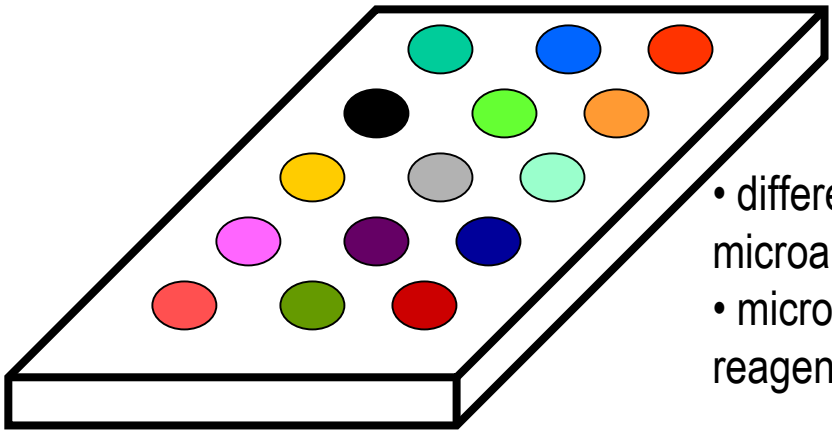
- highly conserved in evolution: plants, some fungi, animals
- perhaps a method of defense against RNA viruses
- however, *in vivo* silencing might thought to occur
    - (ie. used in gene regulation)

RNA Interference in Mammalian Cells

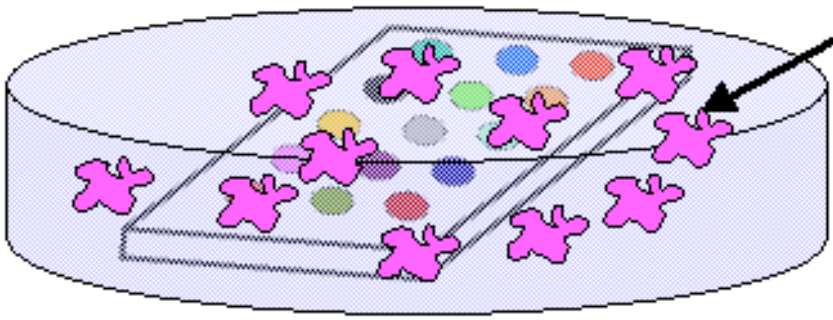
- RNA interference was elucidated in plants, fungi, and insects
- Long double-stranded RNAs lead to strong antiviral response in mammalian cells
    - until mechanism was elucidated, people used long ds-RNAs
    - now we know the best size is 21-23 bp
- Small RNAs (less than 30 bp) do not produce antiviral response
- When tested, small double stranded RNAs did lead to silencing in mammalian cells
- Effectiveness varies, but can be > 90% reduction in mRNA and protein
- How is it performed ??? Transient transfection
    - direct introduction of in vitro synthesized small ds-RNAs
    - in vivo expression of ds-RNA:
        - plasmid expressing sense and antisense sequences
        - plasmid expressing RNA with self-complementarity
            - processed into siRNA-like molecule

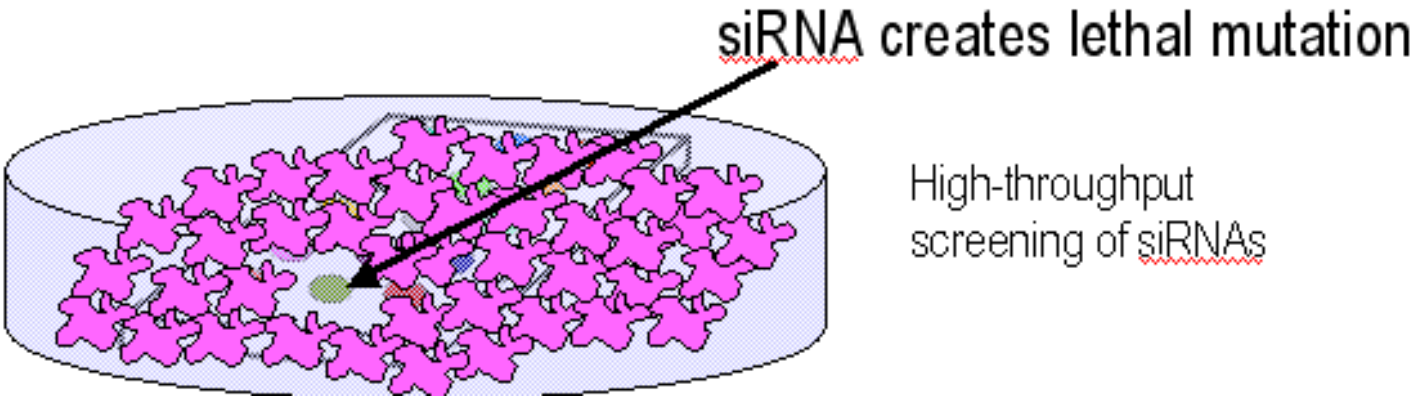
- Significance ?
    - powerful tool for functional genomics
    - relatively easy way to perform gene knock-outs
    - can be used to investigate loss-of-function mutations
    - could eventually be investigated for therapeutic purposes

Functional Genomics using siRNA Microarrays



- different siRNAs are printed onto a microarray
- microarray surface: transfection reagent polymer

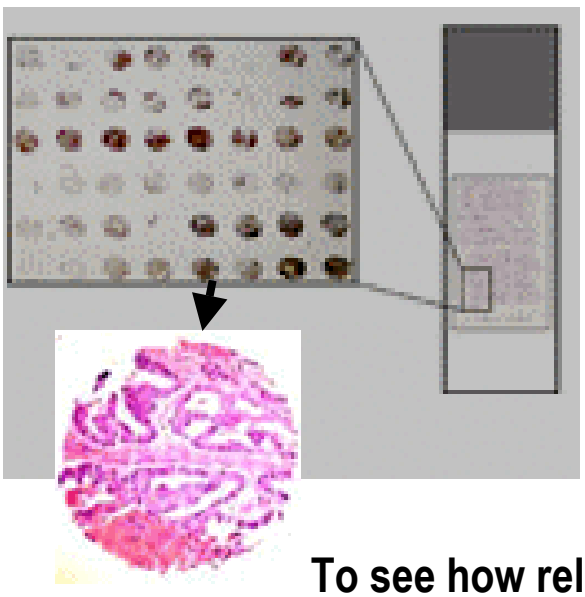




Clinical Genomics using Tissue Microarrays

- Let's say, on the basis of microarrays or proteomics, you have found…
    - genes that appear to be up-regulated in cancerous tissue
- Need to validate the relevance of your findings in a clinical sense
    - ie. You think it's important….but is it truly ?

- Tissue Arrays contain micro-sections of hundreds of biopsy samples
    - arrayed in order
    - lots of tissue arrays generated from the same biopsy samples (ie. each array gets the same biopsy samples)



- High throughput histological work
    - histochemical staining
    - immunohistochemistry
    - *In situ* hybridization

To see how relevant your protein is to cancer, develop antibody to it, see if the protein is detected in cancerous biopsies but not in normal biopsies !