Interactomics and Proteomics

1. Interactomics

The field of interactomics is concerned with interactions between genes or proteins. They can be genetic interactions, in which two genes are involved in the same functional pathway (leading to a particular phenotype), or physical interactions, in which there is direct physical contact between two proteins (or between protein and DNA).

2. Complementation groups

Using forward saturation genetics, you may recover several independent mutants with the same (or similar) phenotype. There are two possibilities:

a) mutations are in the same gene
b) mutations are in different genes involved in the same pathway.

Scenario (b) can be tested genetically with a complementation test:

- Cross two homozygous mutants and observe heterozygous offspring phenotypes
- Mutations in the same gene will not complement -> offspring have mutant phenotype
- Mutations in different genes will complement -> offspring have wild-type phenotype
- Do pairwise crosses for all mutants to identify complementation groups
- Typically each complementation group represents a different gene
- If many mutations are recovered in the same genes, this implies saturation

3. Modifier screens

This is a way to find new genes involved in the pathway leading to a mutant phenotype. It can also identify epistatic interactions. Forward saturation mutagenesis is performed in the background of a particular mutant phenotype, then new mutants that alter the phenotype of the original mutant are identified. Types of mutations that can be uncovered:

Back mutations – a second mutation in the same gene that corrects the previous mutation and restores phenotype to wild-type. This is rare; modifying mutations are typically in other genes.

Enhancers – mutations that make the original mutant phenotype more extreme

Suppressors – mutations that make the original mutant phenotype less extreme

Example: Drosophila Bar eye. The *Bar* mutation in Drosophila causes flies to have narrow eyes that are shaped like a bar. *Enhancer of Bar* mutations cause the eyes to be even narrower, while *suppressor of Bar* mutations cause the eye to be wider and more like wild-type.

Epistasis = interactions between genes (non-additive).
A mutation in one gene allows a mutation in a second gene to be revealed.
The double mutant has a unique phenotype or one more extreme than expected from the two single mutations.

Example: Drosophila eye color. Drosophila eyes contain two color pigments, brown and red. When both are present, the eyes are dark red (wild-type). The *vermilion* mutation prevents production of the brown pigment, resulting in eyes that are bright red. The *brown* mutation

prevents production of the red pigment, resulting in eyes that are brown. When both pigments are missing, the eyes are white.

Example: synthetic lethals. Two mutations are viable individually, but lethal in combination. This has been extensively studied in yeast and is thought to be caused by redundancy of function of some genes.

In yeast, over 80% of the 6200 predicted genes are non-essential when knocked out individually. Large-scale array crosses of each single knockout to every other knockout can reveal interactions through synthetic lethals, however there are ≈ 20 million pairwise combinations.

A large-scale screen for genetic interactions in yeast was carried out in 2010. This covered not only synthetic lethals, but also interactions that were not lethal, but affected growth in combination. About 170,000 interactions were identified. Interacting genes tended to be of similar biological processes and formed coherent clusters. See:

Costanzo et al. (2010) The genetic landscape of a cell. Science 327: 425-431.

4. Protein interactions

The above "classical genetics" methods identify genes that interact in the generation of a particular phenotype, but the genes or their protein products may not physically interact with each other. Some possibilities are that the mutations affect:

a) different steps in same enzymatic pathway
b) steps in different enzymatic pathways, the products of each required for phenotype
c) *cis* or *trans* acting factors in gene expression
d) different components of a signal transduction pathway
e) different peptides that physically interact with each other
f) many more possibilities

Many of the above types of interactions are difficult to test on a large, genomic scale. Probably the most progress has been made in the detection of (e), protein-protein interactions. Some approaches:

Affinity chromatography– purify a target protein by passing through a chromatography column. Then see what other proteins come out with the target. This approaches detects direct or indirect interactions (protein complexes). Some methods:

- Co-immunoprecipitation: requires that specific antibodies to the target protein be linked to the column. This causes the target protein (and all proteins bound to it) to be retained in the column.

- Biotin-affinity chromatography: biotin is chemically linked to the target protein. The column contains the protein streptavidin, which has a very high binding affinity for biotin. Thus, the target protein (and all proteins bound to it) is retained in the column. This method is good for large-scale applications; commercial streptavidin columns are available.

- GST-fusion chromatography – a recombinant DNA encoding the target protein is joined to DNA encoding GST (glutathione S-transferase). The resulting fusion protein is run through a column containing glutathione. The strong binding between GST and glutathione causes the

target protein (and all proteins bound to it) to be retained in the column. An advantage of this method is that one does not need to start with purified target protein, only the gene encoding it. This method is good for large-scale applications; commercial glutathione columns are available.

- Direct purification – this approach typically requires much work to purify individual protein complexes by chromatography or sedimentation. An advantage is that you don't have to start with a particular target protein.

Yeast 2-hybrid screen – this method detects direct protein-protein interactions. cDNA libraries are used to make fusions of "bait" and "prey" proteins to the DNA-binding domain (BD) and activator domain (AD) of a transcriptional activator. The recombinant proteins are expressed in yeast. If there is a bait-prey interaction, transcription of a reporter gene will be induced. Note that, although the assay is performed in yeast, it can be used to detect interactions between proteins from any organism.

False positives – bait alone activates transcription due to chance interactions

False negatives – normally interacting proteins change structure when fused to bait or prey and no longer interact

Can be used to make protein interaction maps. Examples:

YEAST
See Uetz *et al*. (2000) Nature 403: 623-627.
Yeast has about 6000 genes. Pairwise comparisons = $n(n-1)/2 \approx 18$ million

Yeast-2-hybrid (Y2H) method with two approaches:

i) Array Screening
a) Fused 6000 ORFs to activator domain (Prey) and transformed them into separate yeast cells. Grew cells in 384-well plates.

b) Fused 192 (half of 384) ORFs to DNA binding domain (Bait). Added each separately to the 6000 Prey cells. 192 x 6000 > 1 million interaction tests. Look for "positives" (cells expressing *His3* gene, which allows them to grow on media lacking histidine).

c) Independently repeat the experiment. Consider only interactions identified in both replicates to be "real".

d) 87 of the 192 had 1 or more interaction. Total = 281 interacting pairs.

ii) Library Screening
a) 6000 Prey cells pooled together and pools grown in 96-well plates

b) Same 6000 ORFs fused to DNA binding domain (Bait). Each separately added to the pooled Prey cells.

c) Cells expressing reporter gene selected and Prey sequenced.

d) 817 ORFs had 1 or more interaction. Total = 692 discrete interacting pairs (472 identified two or more times.

Comparison
Array method produced a higher percentage of interactions (87/192 = 45% vs. 817/6000 = 14%), but library method had much higher throughput.

DROSOPHILA
See Giot *et al*. (2003) Science 302: 1727-1736.
Drosophila has about 14,000 genes ≈ 100 million pairwise comparisons.

Used Y2H with two library screening methods:

i) cDNA Screen
a) Designed primers to 14,000 predicted ORFs, amplified from cDNA and cloned into Bait fusion vector

b) Cloned random cDNAs into Prey vectors.

c) Tested every Bait with random cDNA library Prey. Sequence Prey that gives reporter gene expression.

d) 5,200 ORFs had 1 or more interaction. Total = 10,782 interactions.
High confidence, non-self = 1,344 ORFs. Total = 1,673 interactions.

ii) Collection Screen
a) Cloned 14,000 ORFs into both Bait and Prey vectors

b) Each Bait was added to pooled Prey cells.

c) Cells expressing reporter gene selected and Prey sequenced.

d) 4,500 ORFs had 1 or more interaction. Total = 10,021 interactions.
High confidence, non-self = 3,402 ORFs. Total = 3,319 interactions.

Combined Total
7,048 ORFs had 1 or more interaction. Total = 20,405 interactions.
High confidence, non-self = 4,655 ORFs. Total = 4,677 interactions.

Only high confidence interactions used to make a protein interaction map.

Limitations of Y2H
– Autoactivation. Some Baits may activate transcription by themselves.

– Some Baits or Preys may be lethal when expressed in yeast.

– Only detects direct, pairwise protein-protein interactions. Some complexes may require 3 or more proteins. Does not detect protein-DNA or protein-RNA interactions.

– Interaction in yeast does not guarantee interaction in other cell/organism (proteins can only interact if they are expressed at same place/time).

– Fusion of proteins (Bait-binding domain or Prey-Activator domain) may alter protein structure/function.

5. Protein microarrays - proteins can be purified and immobilized on glass slides in a manner similar to DNA microarrays. The arrays can then be used to test for protein or small-molecule binding.

Typically, many different proteins are placed on an array, but only a single antibody, protein, protein co-factor, small molecule, *etc*. is used as a target for "hybridization" to the array. This approach can identify proteins that interact with the target.

Some problems with protein microarrays:

- unlike with DNA arrays, where it is relatively easy to do PCR or synthesize oligonucleotides on a large scale, it is difficult to produce large amounts of purified proteins to put on arrays.
- Do the proteins maintain their structure/function after attaching them to the array? Typically they need to be kept hydrated and cannot be dried and stored like DNA arrays.

6. 2D-PAGE (two-dimensional polyacrylamide gel electrophoresis)
Total proteins are separated by charge (isoelectric point), then by size on an acrylamide gel. The proteins are stained with a general silver or fluorescent stain. The result is a complex pattern of 1000's of spots that is semi-quantitative.

Good for "head-to-head" comparison of two samples, such as control vs. treatment or human vs. chimpanzee.

Quantitative change = size of spot (amount of protein)
Qualitative change = location of spot (size or charge of protein)