Fall 2001 Genomics Final Exam (#3)
Proteomics and Genomic Circuits

There is no time limit on this test, though I have tried to design one that you should be able to complete within 6 hours, except for typing. There are eight pages in this test, including this cover sheet. You are not allowed discuss the test with anyone until all exams are turned in at 2:00 pm on Wednesday December 12. **EXAMS ARE DUE AT 2 PM ON WEDNESDAY DECEMBER 12**. You may use a calculator, a ruler, your notes, the book and the internet. However, you are not allowed to obtain and read journal articles as a part of your investigations. These questions are taken from research literature and I do not want you to simply find the papers and read the answers. This is the Honor Code at its finest.

The **answers to the questions must be typed on a separate sheet of paper** unless the question specifically says to write the answer in the space provided. If you do not write your answers in the appropriate location, I may not find them. You may want to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain.

Note Bene In the following questions, I have removed figures from papers and will ask you questions. You **MAY NOT** hunt for these papers and read the full papers. I will provide you with the necessary background information and then ask you questions about some figures that came from each paper.

-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please print):

Write out the full pledge and sign:

How long did this exam take you to complete (excluding typing)?

6 pts

1) Explain how the SELDI method Rick Drake uses is better than either the ESI or MALDI method we discussed in class. You may search any source you want to help you remember what SELDI is, but your answer must include how Rick takes advantage of it in his prostate proteomics research.

Case 1

2) **BACKGROUND**

A group of investigators have created a system for correlating transcriptomes with protein-protein interactions. First, they tabulated all the expression data and clustered genes according to their expression profiles (figure 1a, right side). On each row of the interaction pair table are listed two proteins that have been reported to interact directly.

a

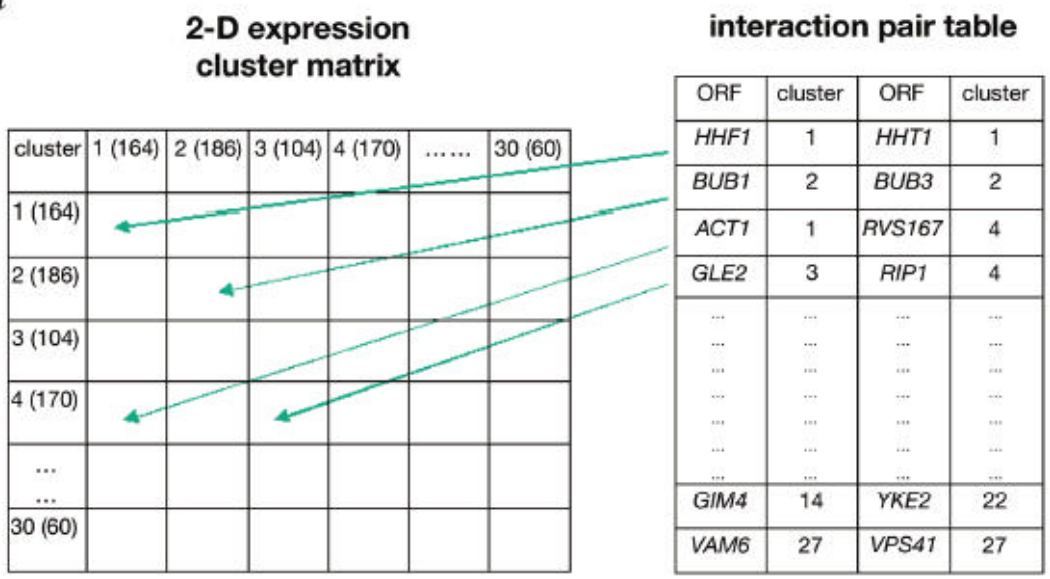


Fig. 1 A general strategy for transcriptome–interactome correlation mapping. ***a***, The 2-D matrix on the left shows all pair-wise combinations between the clusters of an expression profiling experiment. The numbers assigned to each cluster are indicated on the corresponding rows and columns of the matrix along with the number of genes each cluster contains (in parenthesis). The table on the right shows protein interaction pairs together with the expression cluster to which the corresponding genes belong. For each interaction pair, an arrow points to its corresponding square in the 2-D expression-cluster matrix.

So, they performed this correlation and then presented their results in a colorized format (figure 2). As you can see, they used data previously published and examined sporulation and cell stress in yeast.

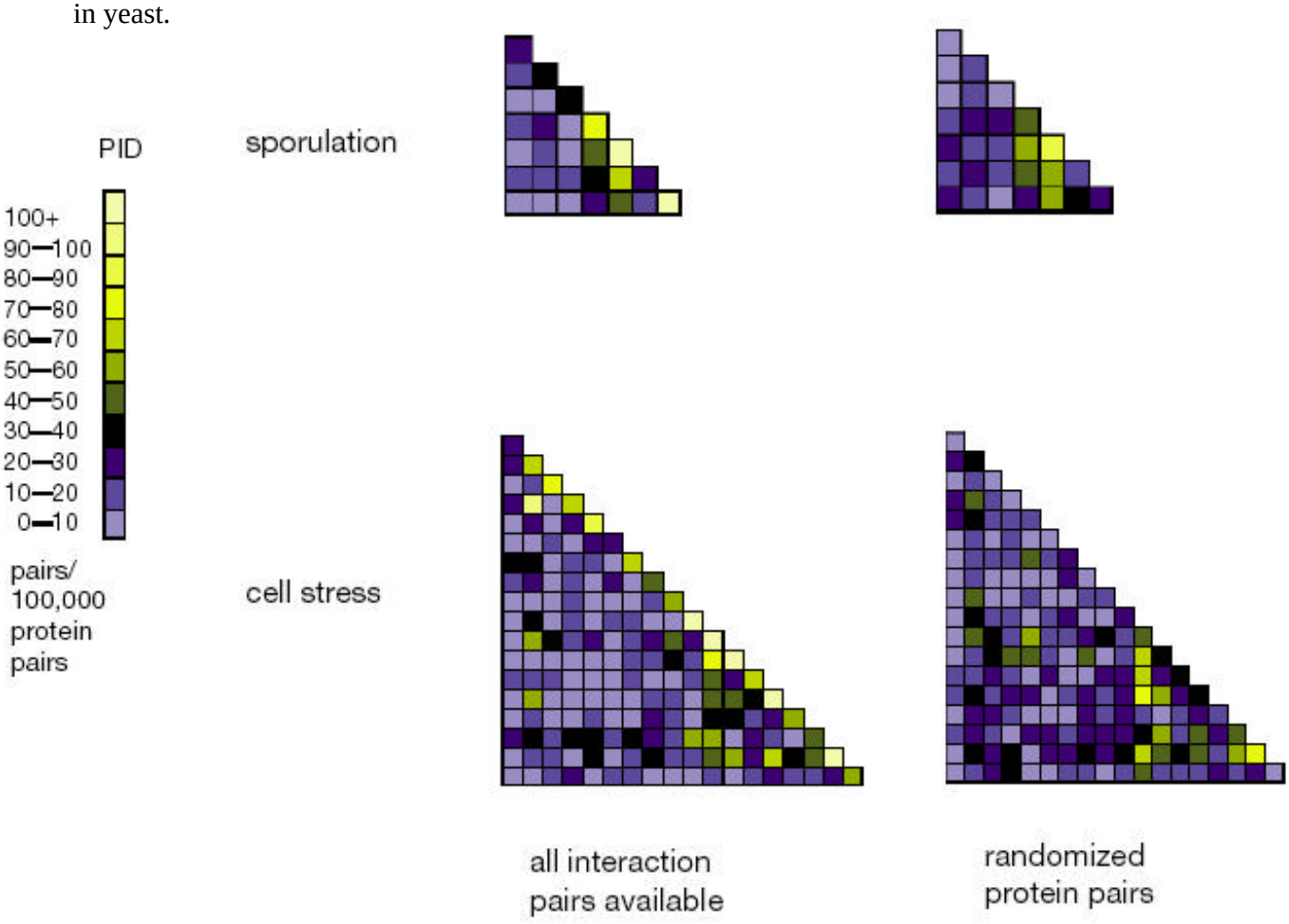


Figure 2. We calculated the protein interaction density (**PID**) for each square in the matrix as the ratio of interaction pairs assigned to the square (IP) over the total number of protein pairs possibly formed by combinations of the genes in the square (PP). PIDs are represented in the map by a color system, as indicated in the scale on the left side. Control maps can be generated by the same approach from randomized protein pairs (right side). The average PIDs from all intracluster squares (in the diagonal) and intercluster squares (outside the diagonal) can be calculated from the correlation maps. The unit of PID in each panel is interaction pairs/100,000 ORF pairs. We constructed transcriptome–interactome correlation maps using sproulation (top) and cell stress (bottom) expression–profiling clusters and protein interaction data from the literature.

The investigators compared the correlations of intracluster (observed) v. random protein clusters (expected) and the analysis is shown in Table 1.

Table 1 • Statistical analysis

Data set	Total in map	Expected[a]	Observed[a]	*P* value[b]
cell cycle	670	25	117	9.8×10^{-43}
sporulation	309	46	115	1.1×10^{-21}
stress response	731	44	165	2.0×10^{-49}

[a] The observed number of groups (k) whose members belong to the same cluster is compared with the expected number, assuming the interaction groups are randomly distributed. [b] The probability for obtaining at least k observed groups in the intracluster region by chance is calculated for each data set using cumulative binomial distribution.

Finally, they examined some protein-protein interaction data in isolation (figure 3). From the interaction data alone, one model was created, but when the expression data was incorporated too, a different model was produced.

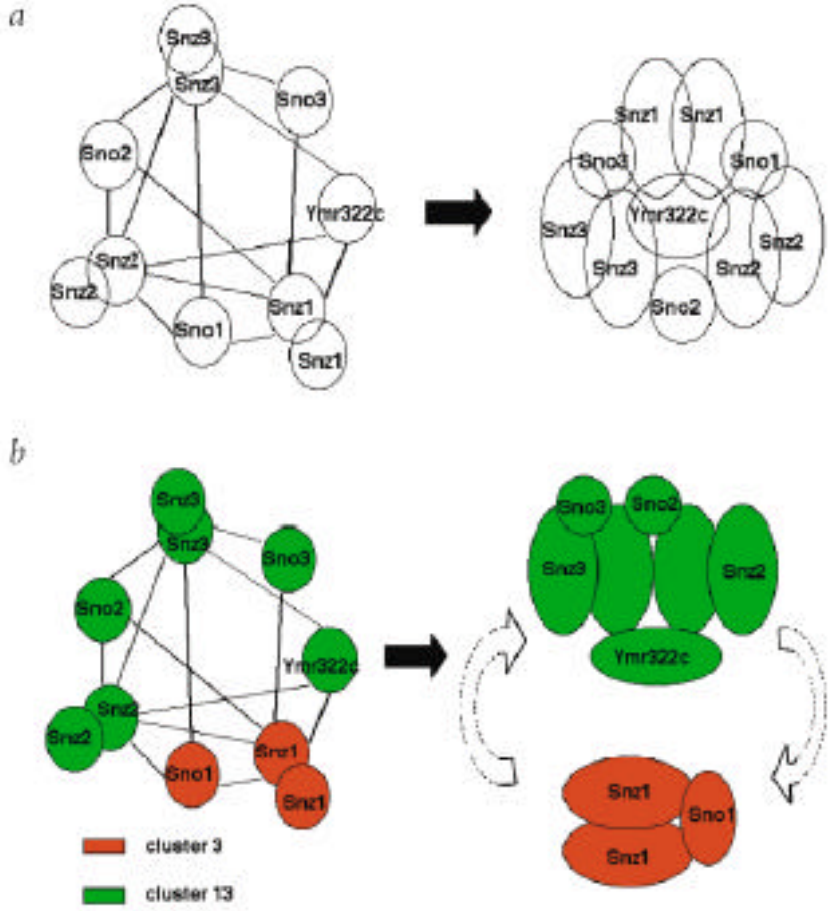


Figure 3. Model from the integration of transcriptome and interactome data. a, Interaction network obtained from the combined data set of protein–protein interactions. Circles represent proteins and lines represent two-hybrid interactions. b, Model obtained by taking into account expression profiles. Cluster numbers and colors refer to DNA microarray clusters.

Questions:
8 pts. each
2) Explain how you are supposed to read the data in figure 2.
3) Do you think their findings from this integration is biologically relevant? Support your answer with data.
4) Explain how their model of proteomic interactions change with the integration of DNA microarray data?

Case Number 2.
Do not explore the full web page associated with this publication.

Background:

Mike Snyder at Yale is pushing the envelop for protein microarrays (If you're looking for a good lab to work in, they just got a 15 million dollar grant!) They spotted out 5800 yeast proteins produced as fusion proteins with GST (GST is an epitope tag). They wanted to see if they could identify which of these spotted proteins could bind to different cellular components (both protein components and lipid components). For the lipids, they created liposomes made of **PC** (phosphatidylcholine) as the basic liposome. Then, for experimental liposomes, they doped each one with a different lipid variation of phosphatidylinositol (**PI**; one specific example is PI(4)P). Calmodulin is a protein. The number immediately after the "PI" tells you which carbon is phosphorylated and the P_X after the parentheses tells you how many phosphates were added to the lipid.

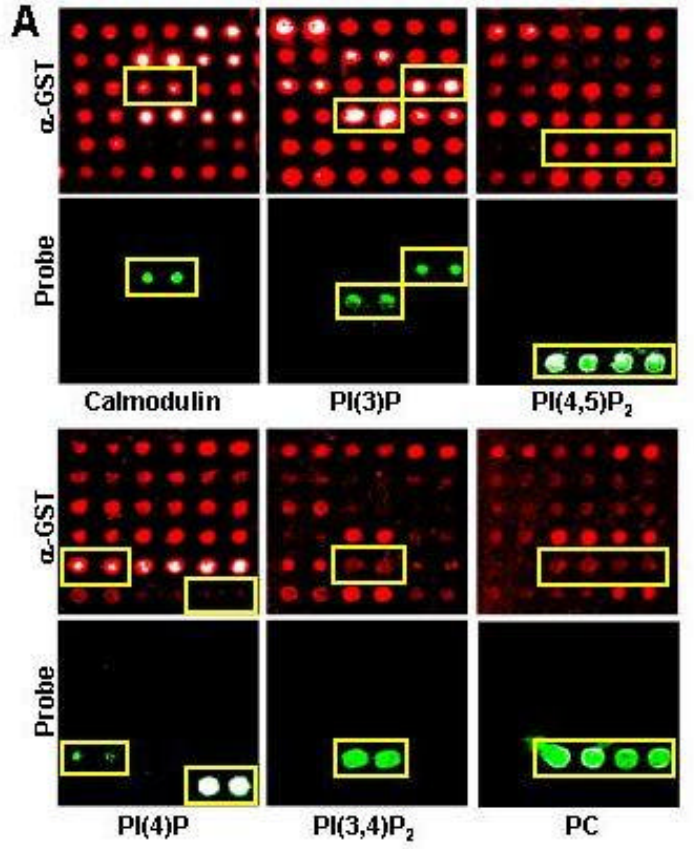
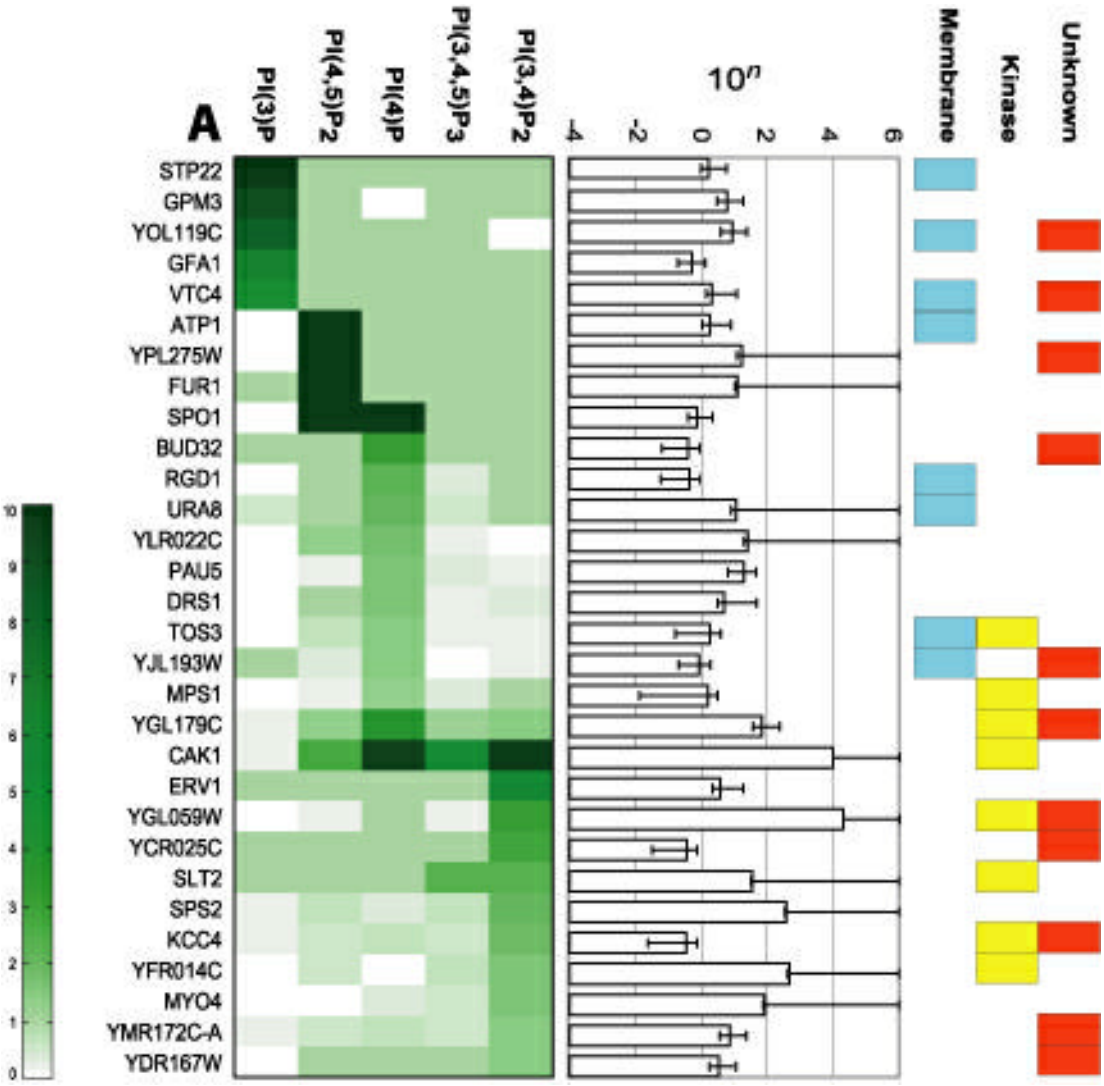


Figure 4. Examples of different assays on the proteome chips.
Positive protein targets are indicated by yellow boxes identified in six protein-liposomes interaction or calmodulin-binding assays. Each block contains 16X18 protein spots. The positive signals in duplicate (green) appear as horizontal pairs. The duplicate spotting serves as an internal control, which is important when the signals are weak relative to the background. A control proteome chip was probed with anti-GST antibodies. The upper panel shows the amounts of GST fusion proteins as detected by the anti-GST antibodies (red); corresponding proteins were probed for the specific activities shown in the lower panel. As demonstrated by the images, strong signals are often observed from samples that have little GST fusion protein; thus the binding is sensitive and specific.

Here is the output from a subset of proteins and several lipids. The color scale indicates the PI/PC ratio of binding. Each protein is listed on the Y-axis and the different lipids are listed on the X-axis (along the top). The 10^n graph shows the signal intensity compared to background for each of the spots (averaged over all the experiments with 95% confidence interval shown). The colored boxes on the far right show us what is known from the literature about location (blue = membrane bound/associated), cellular role (yellow = kinase), and red/orange = unknown ORF.



Questions:

8 pts. each

5) What advantage(s) do you see in this protein microarray method compared to yeast two-hybrid.

6) Which protein seems to be the least specific of this group?

7) Which protein had the highest specificity for PI(3)P?

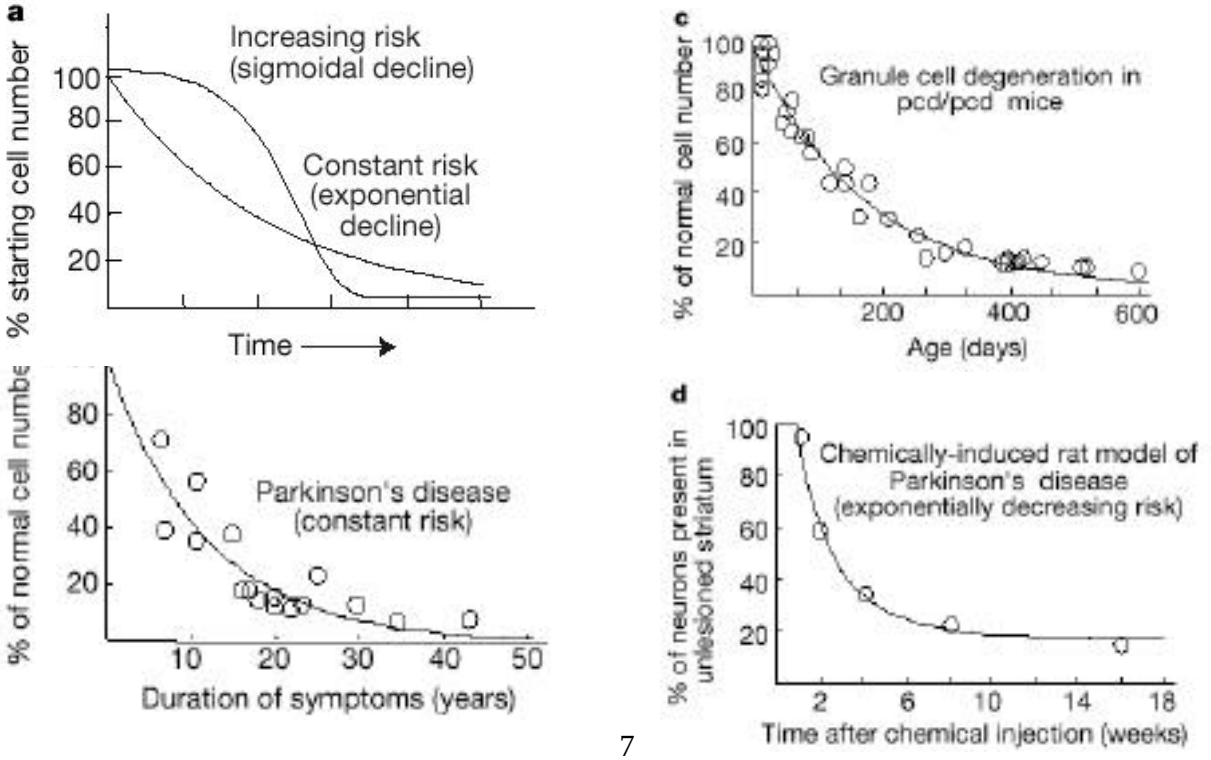
8) If CAK1 is a kinase, what do you think its substrate is? Its product?

9) If a protein bound to one of these lipids but does not have a blue box next to it, what can you conclude?

Case 3

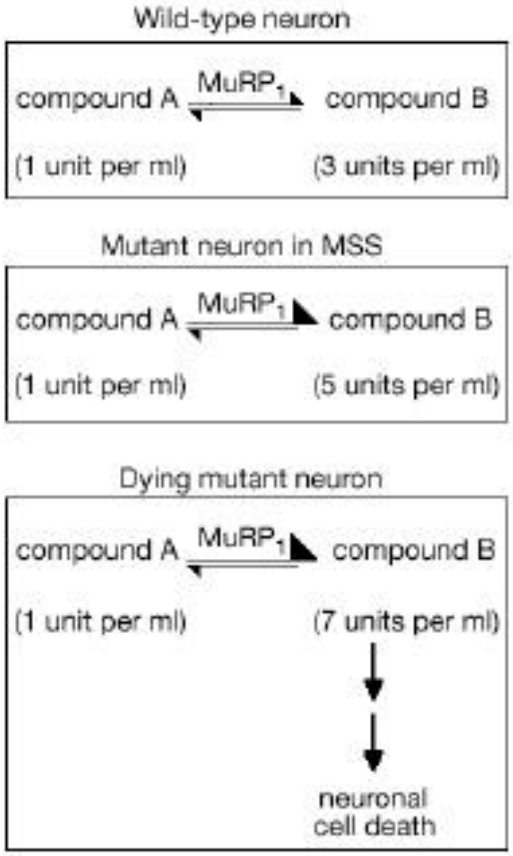
Background:

Abstract: In genetic disorders associated with premature neuronal death, symptoms may not appear for years or decades. This delay in clinical onset is often assumed to reflect the occurrence of age-dependent cumulative damage. For example, it has been suggested that oxidative stress disrupts metabolism in neurological degenerative disorders by the cumulative damage of essential macromolecules. A prediction of the cumulative damage hypothesis is that the probability of cell death will increase over time. Here we show in contrast that the kinetics of neuronal death in 12 models of photoreceptor degeneration, hippocampal neurons undergoing excitotoxic cell death, a mouse model of cerebellar degeneration, Parkinson's and Huntington's diseases are all exponential and better explained by mathematical models in which the risk of cell death remains constant or decreases exponentially with age. These kinetics argue against the cumulative damage hypothesis; instead, the time of death of any neuron is random. Our findings are most simply accommodated by a 'one-hit' biochemical model in which mutation imposes a mutant steady state on the neuron and a single event randomly initiates cell death. This model appears to be common to many forms of neurodegeneration and has implications for therapeutic strategies.



I have selected three real graphs (plus one stylized graph to show you the two possible graphs for the two competing hypotheses). The three real graphs illustrate a subset of the investigators' data published in their paper. The *pcd/pcd* mice are a genetic model for neuronal cell loss in the eye.

The investigators summarized their thoughts about the One Hit Hypothesis in a model (see below).



One Hit Model: The exponential kinetics of cell death in inherited neuronal degenerations suggest the existence of a mutant steady state (**MSS**) in which the risk of cell death is increased. In wild-type neurons, a reaction, catalyzed for example by the enzyme **MuRP1**, is associated with a concentration of compound **B** of 3 units per mL. In a mutant neuron in the MSS, the MuRP1 activity changes in response to the mutation, so that the concentration of compound B is increased to 5 units per ml. Random increases in the concentration of compound B to 7 units per mL will trigger cell death.

Questions:

10) How could there be a decreasing risk of cell death over time? Explain this using the One Hit Model above. Keep Rick Drake's public lecture in mind, though attendance is not necessary to explain the model.

11) For the constant risk cells, what are the consequences on any given cell if its neighbor dies? Explain your answer.

12) Explain how a cell can go from 5 units of compound B to 7 units if there are no additional mutations within any cells in the organism.