Fall 2002 Genomics Mega-Problem Set #2
Genomic Variations and DNA Microarrays

There is no time limit on this test, though I have tried to design one that you should be able to complete within 5 hours, except for typing and web searches. There are 4 pages for this test, including this cover sheet. You have 1 week to complete this problem set (October 30 – November 6).You are not allowed discuss the test with anyone until all ~~exams~~ problem sets are turned in at 11:30 am on Wednesday November 6. **PROBLEM SETS ARE DUE AT CLASS TIME ON WEDNESDAY NOVEMBER 6**. You may use a calculator, a ruler, your notes, the book and the internet. This is a challenging test, so do NOT put it off too long. You may take it in as many blocks of time as you need to.

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 might want 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. If you are asked to print out any pages, you do not have to print in color, though it is permitted. Please staple your printed pages near your typed answers and label which prints go with which problems.

-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)?

This exam is based on a paper by Michelle Arbeitman (Is it an omen that her name is work man in German?) et al. I will send each of you a PDF copy of this paper. In addition, I will send you an additional PDF file. Finally, you must use the web resources at this URL http://flygenome.yale.edu/Lifecycle (this URL is case sensitive because it runs on a UNIX web server).

Part of this test will evaluate your ability to interpret data. The other part will test your ability to mine data. I have intentionally not provided you with directions on how to mine this dataset. That is what I want each of you to figure out based on what we have done in class. Good luck and I hope you enjoy the challenge.

10 points

1) In the abstract from this paper (Hum Mol Genet 2002 Sep 15;11(19):2289-95), you will read a lot of jargon. Please explain this abstract in plain English so that Bio111 students could understand it.

You were supposed to find this abstract and read it. This paper is different from the one by Arbeitman et al. Here is the link
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12217957&dopt=Abstract

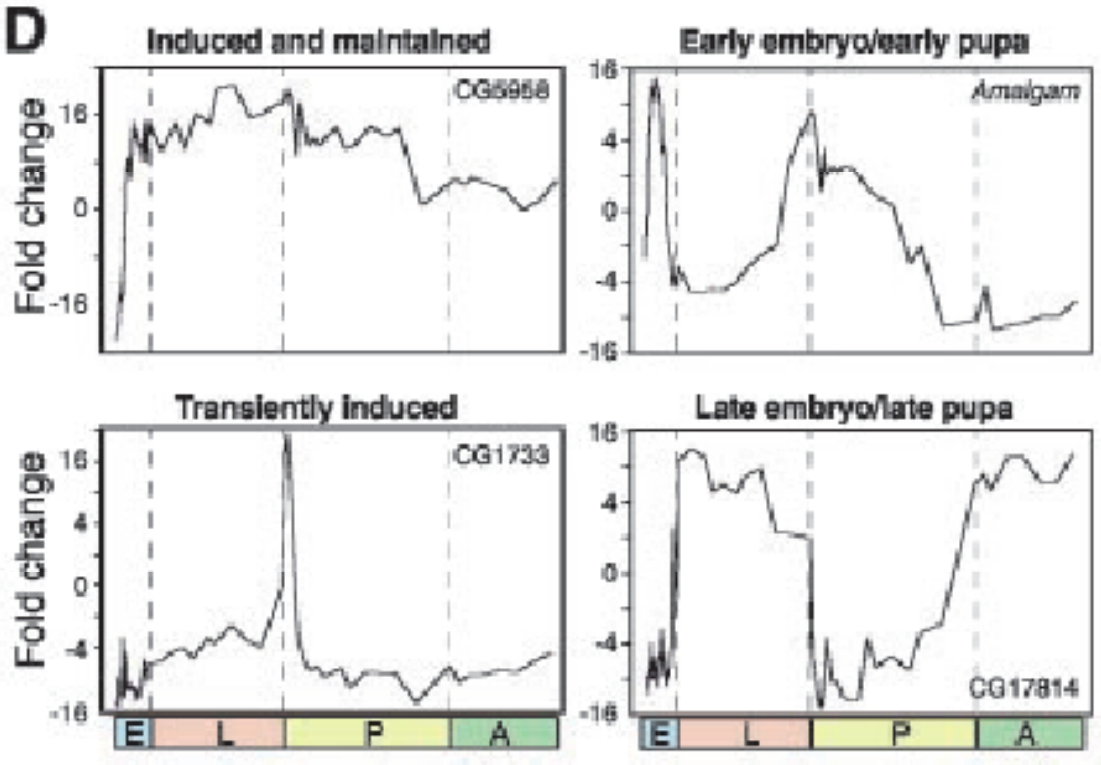
The key points you were to explain in Bio111 language were:
This is a multi-gene trait
The gene of interest was identified by Linkage disequilibrium.
There were 4 SNPs, one coding and 3 in the 3'UTR.
There were two common haplotypes.
The haplotype including the missense mutation was not linked to the decreased bone density.
One particular haplotype of the three 3' UTR did correlate with the weakened bones.
This haplotype accounted for about 1.2% of the variation in the population.

Questions 2 & 3 do not need data mining from the web site to answer them correctly.

10 points

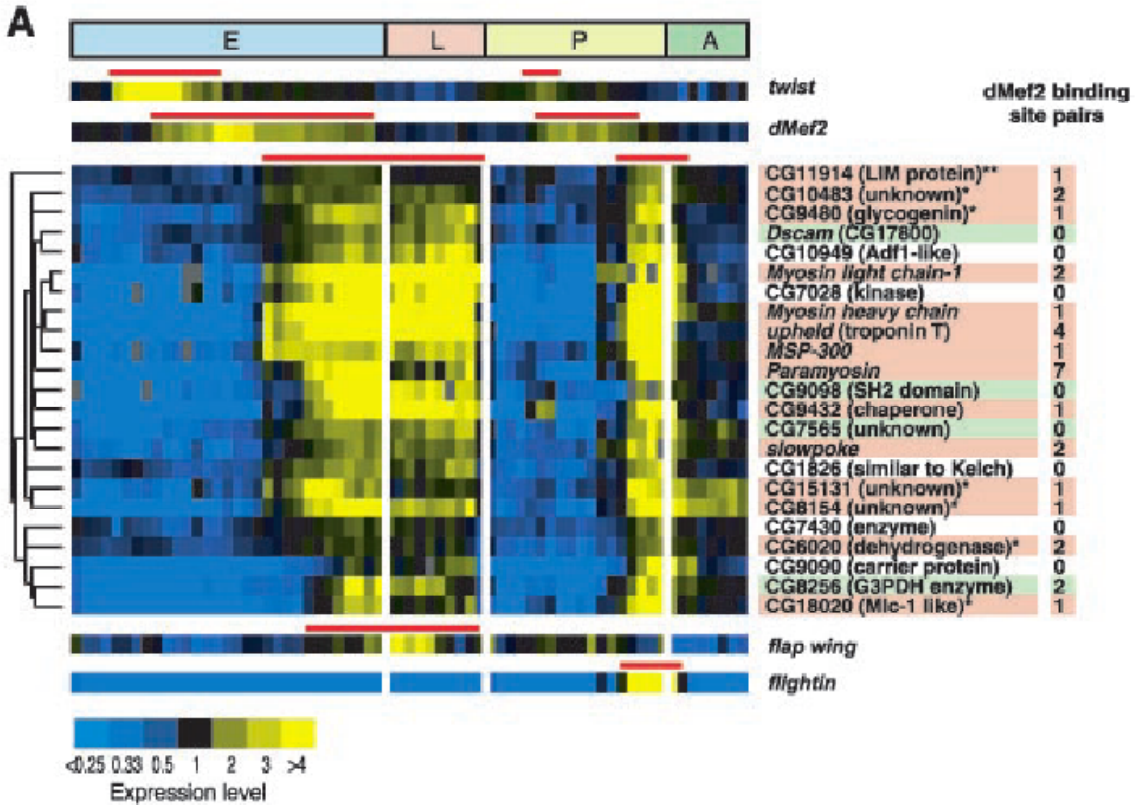
2) Interpret figure 1D. Do not simply quote the paper back to me. I want you to use your own words to explain what is happening in each of those four panels. You may use screen shots as a part of your answer.

You needed to comment on the two 2-wave trends and how genes Amalgam and CG17814 were two examples of these common temporal patterns.
Not that a single wave could be long lasting (CG5958) or a simple spike (cg1733).



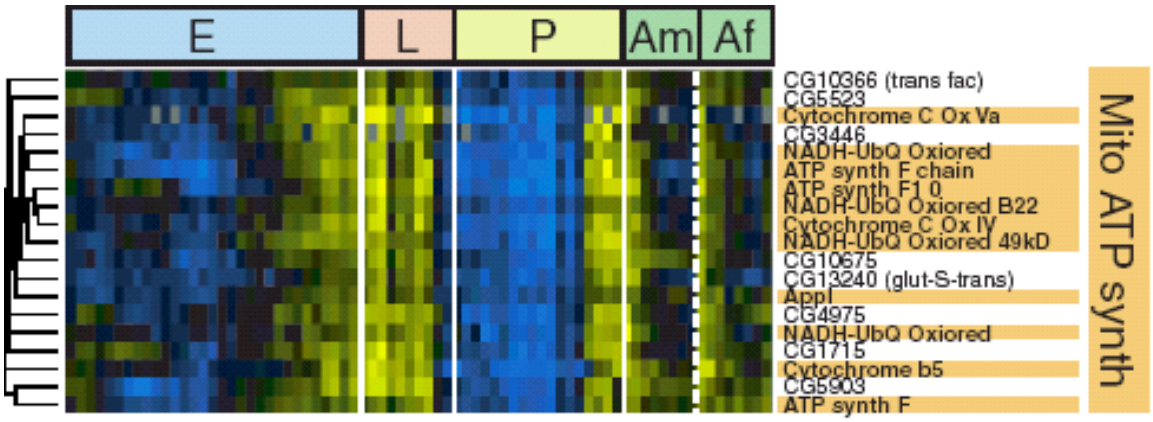
10 points

3) Interpret figure 4A. Do not simply quote the paper back to me. I want you to use your own words to explain what is happening in this panel. You may use screen shots as a part of your answer.

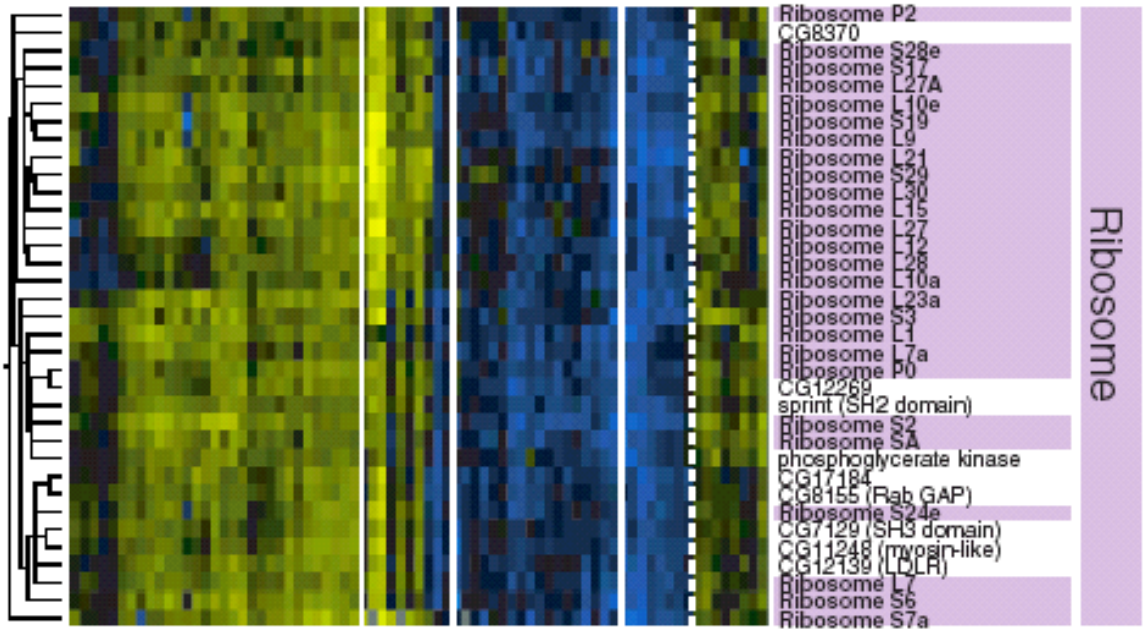


You needed to:
Address the sequential induction of twist, dMef2 and then the cluster of genes.
The significance of green, vs. pink, vs. white genes.
The number of dMef2 binding site pairs and the correlation to expression.
The significance of flightin and flap wing (below main cluster)
The difference in duration of the first wave and the second wave of gene expression.

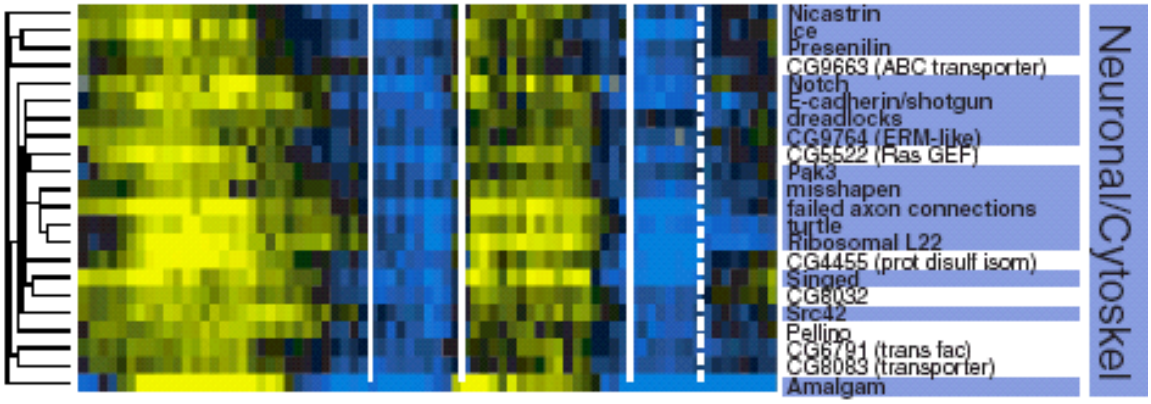
Questions 4 – 8 will require data mining to answer completely.
10 points
4) a. Interpret **supplemental** figure 3 as fully as you can. Do not comment on each gene but address only the three signature gene categories shown in figure 3B of the full paper.



Above: You need to comment on the two waves, the relative expression in females and males.



Above: You need to comment on the two waves: the extent of the first wave, the expression in females but not males as the second wave, and that eggs might be the reason why females needed the extra protein, compared to males.

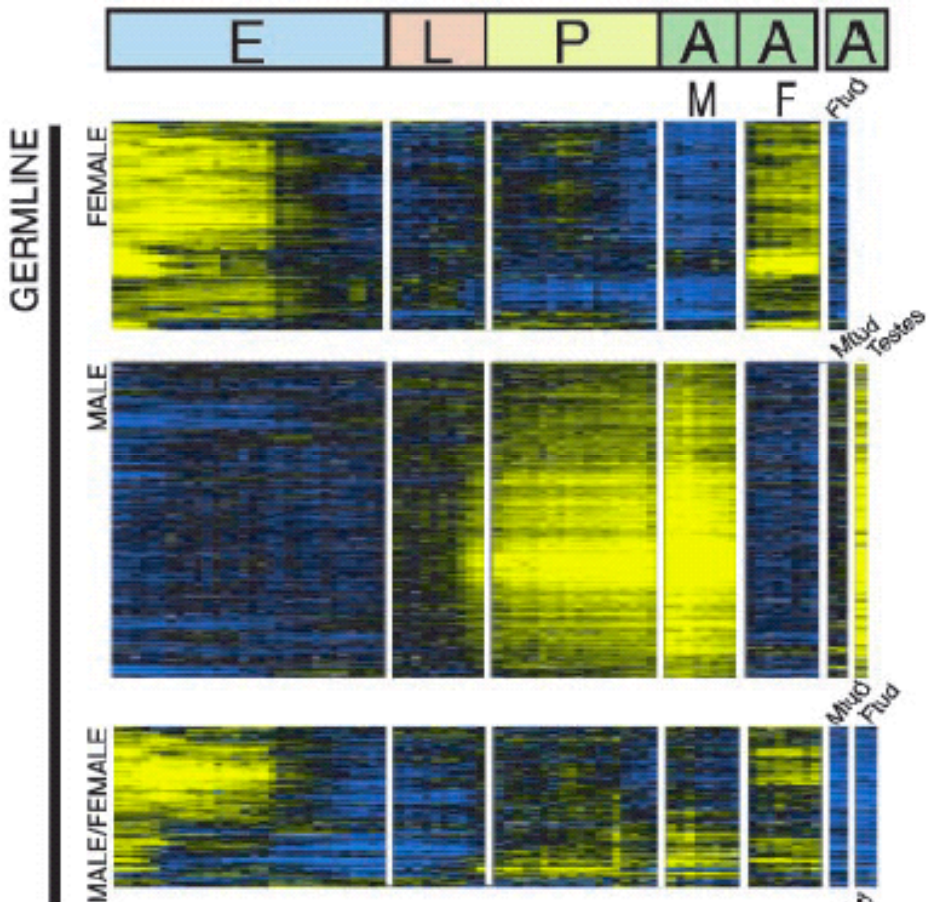


Above: You need to comment on the two waves, and the slight increase in females.

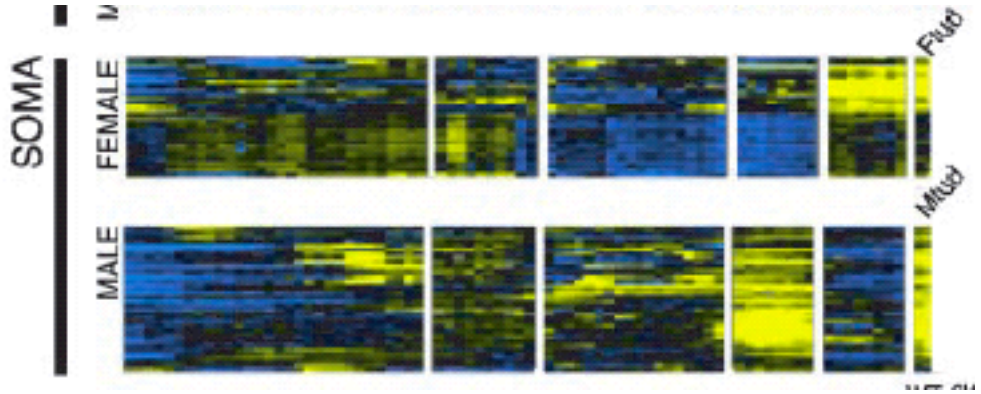
b. Chose one uncharacterized gene from one of these three sections and deduce its function based on available data. Support your conclusions with data.
Many options here. Minimum was to associate your unknown with nearby knowns. Better to do BLAST, or FlyBase analysis. Some of you found very good data to support a functional role.

10 points

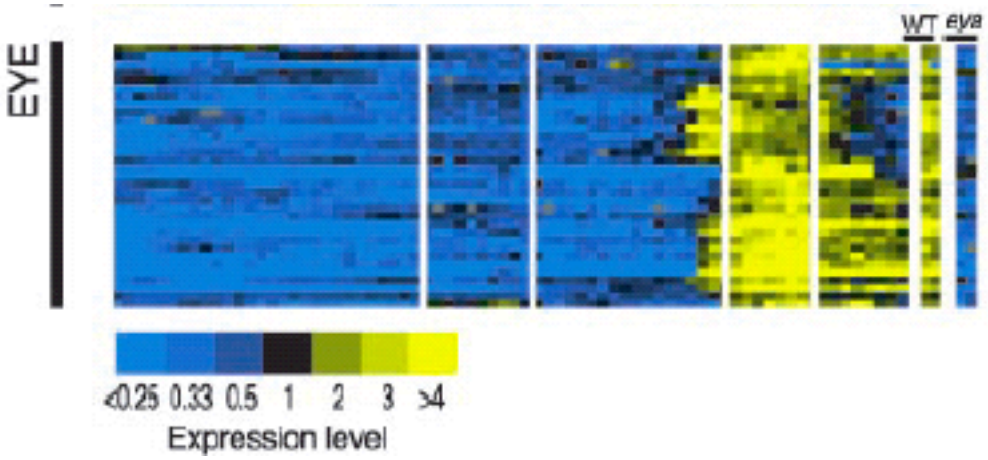
5) a. Interpret figure 5A-C. Address each of the six gene clusters individually.



Female: note two waves but second wave only in female gonads. Verified with F Tudor mutant strain.
Male: One large wave that begins in all pupae (maybe only males??) but in adults is found only in male gonads, as verified with last two columns.
Both: Not typical two wave class, and cluster seems less cohesive than most. Genes expressed in both gender's gonads as demonstrated by tudor lanes.



Male or Female specific expression of genes but not expressed in gonads since tudors were similar to wt adults.



Only adults express these genes which were clustered by correlation but tended to have many eye genes. Notice the wt v. eya strains had opposite expressions. Also notice the bulge into pupal stage (in top half) for some genes and related repression later in adulthood.

b. Fine the name of one gene for four out of the 6 categories (4 genes total) but your genes must not be mentioned in the paper. To get full credit, print out page(s) to document what you found.

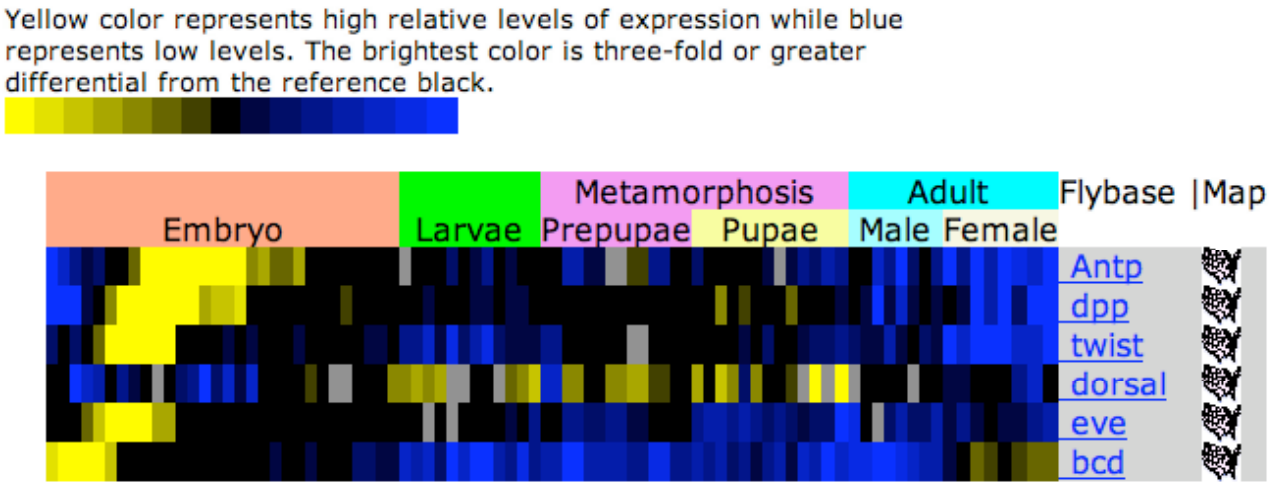
Two main ways to do this. You could either begin with the Ticker interface or the Cluster interface. Either way, you would want to verify that it was an appropriate gene by doing a FlyBase or similar search for functional role.

20 points

6) Here is a list of 6 genes.

Antp
dpp
twist
dorsal
eve
bcd

a. Do any of these genes function in the same signaling cascade? Support your answer with data/printouts.



I wanted you to generate a figure like this using the list of genes interface. From this figure, it appears that these genes were a part of a single cascade, with the exception of dorsal.
However, if you clicked on the Flybase link, or used other databases, you could deduce that more than one cascade was involved and their interactions were more complex than the simplistic view above.

b. Do any of these genes expression profiles differ from what was previously known about their expression patterns? Support your findings with data/printouts.
Yes, almost every one of them appears to contradict the literature. However, as several of you noted, the color coding is relative to a common pool of mRNA so blue only indicates repressed relative to the pool, which is not the same as no transcription.

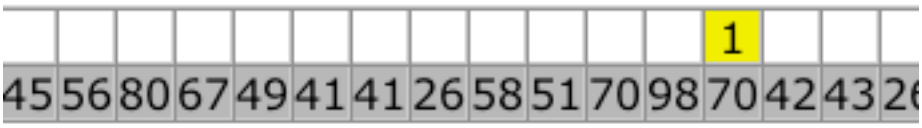
c. Tell me when three of these genes (you can choose which 3) are expressed at their highest and lowest levels. Support your data with printouts of what you found.
You needed to use the single gene interface and show the graphs. From this, you could get precise time of maximum expression. See answer 7d for an example.

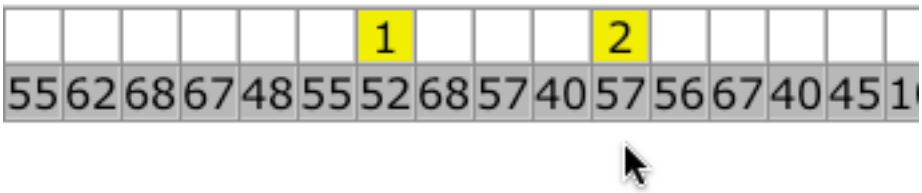
30 points

7) Mine the data to answer these questions:

a. Find a cluster of genes that has two genes that are very near each other on a chromosome and are these two genes are expressed at the same time. Print out critical web pages that you used. To receive full credit, you must tell me how you located the cluster you chose.

The way I thought of was to use the cluster interface and choose one cluster. Then below this you'd find a diagram of the chromosomal location for the genes in each cluster. We were looking for a 2 or greater above a single band of any chromosome.





b. Find a gene that is expressed early in embryogenesis that is associated with DLBCL. Tell me the name of the fly and human orthologs and what type of mutation(s) cause the disease.

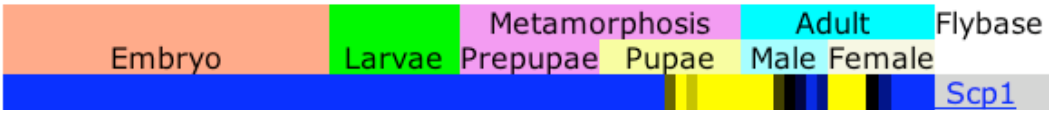
From the human fly ortholog link, you should have located this OMIM entry
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?109565

c. For the fly gene you located for part b., align the human and fly protein orthologs to demonstrate why these two are considered orthologous. Print out the data.

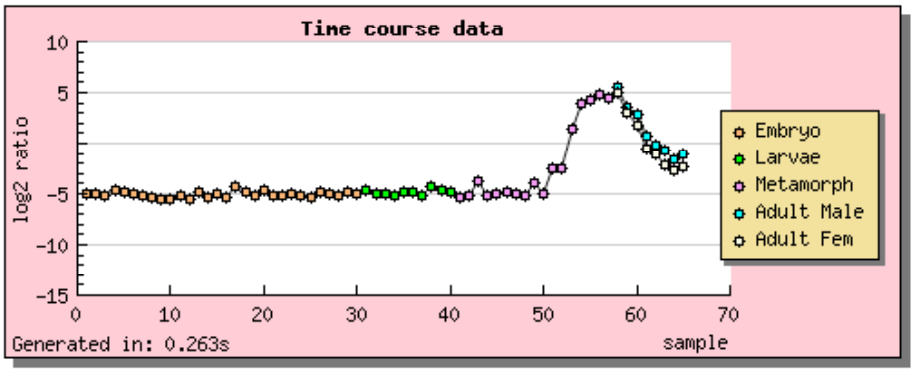
You should have used BLAST2p to generate a graph and a percent identity/similarity. Because there are more than one entry for the human ortholog, more than one answer was correct.

d. Which gene was the most strongly repressed during the embryo stage?

I did not intend for you to choose from the proteins in part c, but due to my poor wording, this was acceptable. However, I wanted you to do a general query to find the gene with the greatest repression. The answer is Scp1.

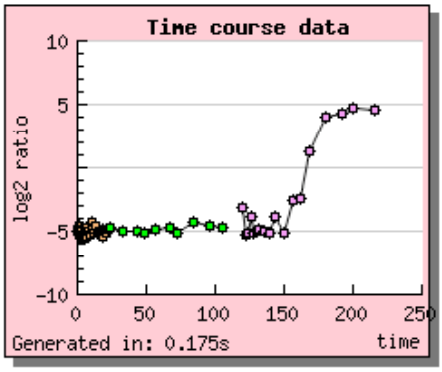


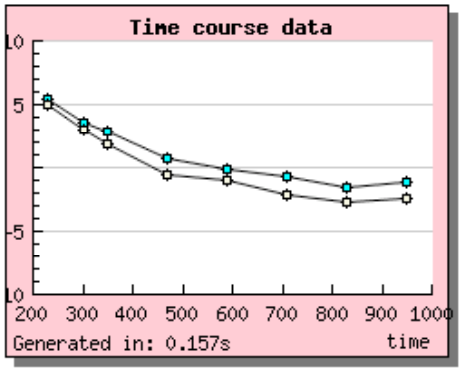
Graph: time course



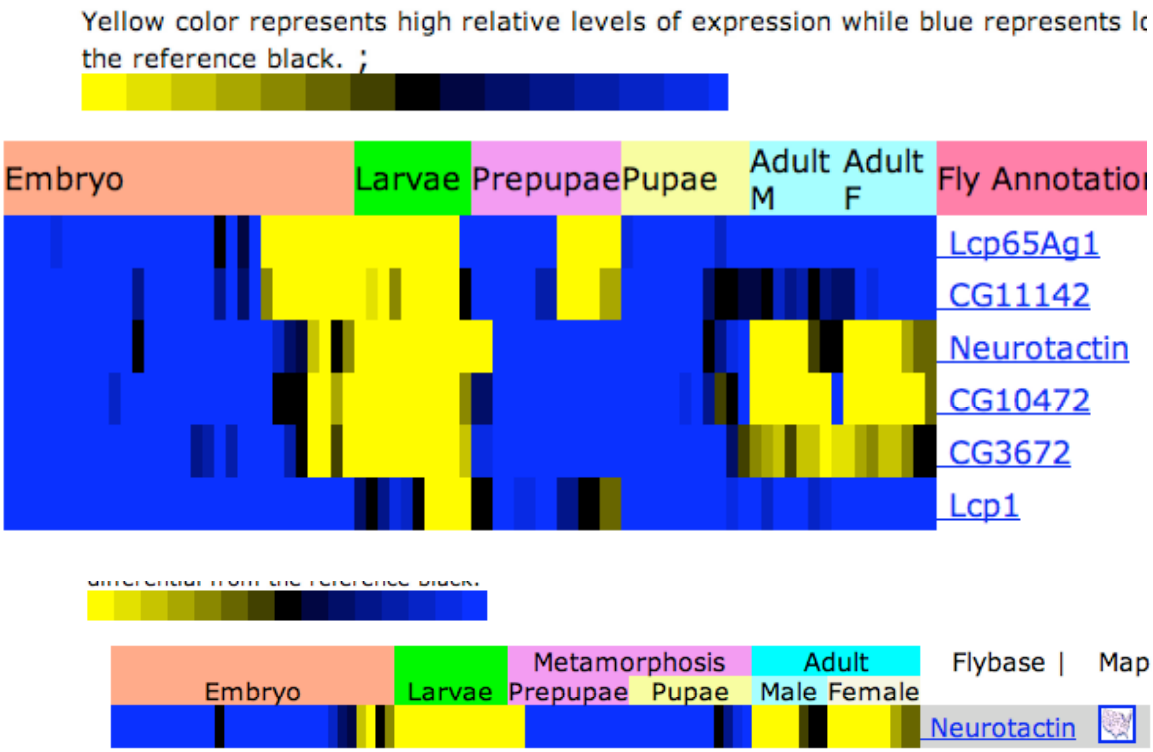
Gene info

Gene name:
Celera gen:C
Protein doma
Status: confi
Chromosome
Fbgn: FBgn0
EST: GH152
Function: liga
Mixed annota
(Penaeus sp.
SARCOPLASM

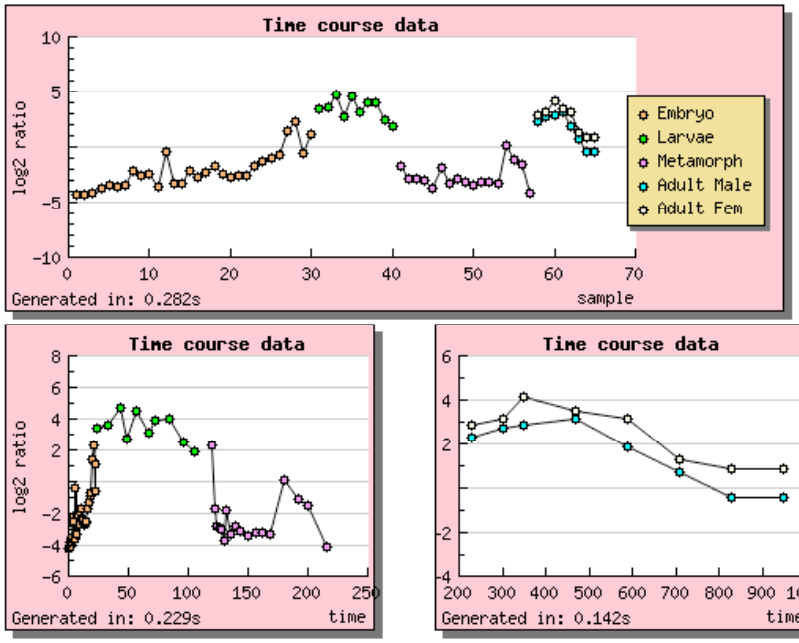




e. Find a gene that is induced at least 22 fold during larval stage, is expressed in a second wave at the typical time period as defined in the paper for larval genes expressed in two waves, and is expressed slightly higher in adult females than adult males. Show me screen shots of major intermediate points in your search.



Graph: time course



Gene info

Gene name:Neurotactin
Celera gen:CG9704
Protein domain: "CARBOXY
Status: confirmed
Chromosome band:
Fbgn: FBgn0004108
EST: LP05519
Function: cell adhesion
Mixed annotation: "axon, c

Doing a general query, you should have gotten the list of genes at the top. From this, you see that Neurotactin was the only gene that easily fit the description. There were other ways to find this gene, but this was the easiest way for most students. It was helpful if you verified this expression pattern with a single gene search to produce the graphs above.