

**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 four 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)?

Dr. Campbell's Genomics Exam #2 – Fall 2002

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.

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.

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.

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.
b. Chose one uncharacterized gene from one of these three sections and deduce its function based on available data. Support your conclusions with data.

10 points

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

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.
- b. Do any of these genes expression profiles differ from what was previously known about their expression patterns? Support your findings with data/printouts.
- 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.

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