Fall 2008 Genomics Exam #2
Sequence Variations and Microarrays

There is no time limit on this test, though I don't want you to spend too much time on it. I have tried to design an exam that will take much less time that exams in the past. You do not need to read any additional papers other than the ones I send to you. There are 3 pages, including this cover sheet, for this test. You will have to answer 5 Discovery Questions and 2 original questions. I expect the original questions will take you more time than the Discovery Questions. You are not allowed discuss the test with anyone until all exams are turned in by 11:30 am on Wednesday November 5. **PAPER COPIES OF YOUR EXAM ANSWERS ARE DUE AT CLASS TIME ON WEDNESDAY NOVEMBER 5**. You may use a calculator, a ruler, your notes, the book, and the internet. You may take it in as many blocks of time as you want. Submit your paper and electronic versions before 11:30 am (eastern time zone:-).

The **answers to the questions must be typed in a Word file and emailed to me as an attachment**. Be sure to backup your test answers just in case (I suggest a thumb drive or other removable medium). You will need 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. Remember to explain your thoughts in your own words and use screen shots to support your answers. Screen shots without ***your*** words are worth very few points.

DO NOT READ or DOWNLOAD ANY NEW PAPERS FOR THIS EXAM. RELY ONLY ON THE FIGURES PROVIDED IN THIS EXAM, YOUR EXPERIENCE, AND YOUR SKILLS.

-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:
"On my honor I have neither given nor received unauthorized information regarding this work, I have followed and will continue to observe all regulations regarding it, and I am unaware of any violation of the Honor Code by others."

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

For this part of your exam, answer these 5 Discovery Questions. Be sure to support your answers with data and explain your logic. *10 points each*

I cannot post answers to these but am happy to talk about them.

Chapter 4, #20 Answer this question, then do your best to find the SNP in question. Which ethnic group has the greatest proportion of protective variation? Be sure to document your answer with screen shots and URLs as you go.

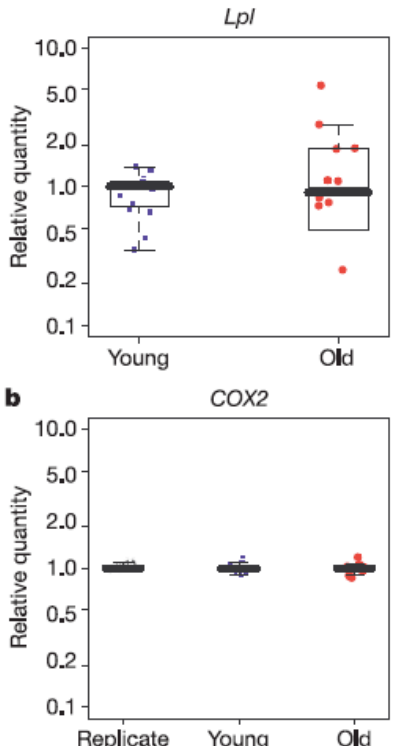
The hardest one was 4, 20. It was not possible to find this SNP in any database as far as I can tell.

Chapter 4, #21. From this answer, use GEO profiles to find an example of expression data indicating that this gene is induced in cutaneous malignant melanoma. Be sure to document your answer with screen shots and URLs as you go.

Go to the Clustering page from **Chapter 6 DQ #31**, but don't answer that question. Instead, find 4 repressed genes from the H_2O_2 experiment that cluster as a single gene with correlation coefficient of 0.7, three clusters with correlation of 0.97 and four clusters with correlation of 0.99. Which two genes were the most closely correlated? Be sure to document your answers with screen shots and numerical values.

Chapter 7 #57.
Chapter 6, #26. Do not read any papers for this question. I would like you to talk me through the logic you used to derive your answer. (this is the new question in place of the one I exposed in class)

1) *20 points*
a) Attached is a PDF file (called Figure 1.pdf) showing a new method to detect variation in gene expression in heart cells. Explain the overall method to a Bio111 student and analyze panel c for this same Bio111 student.
Wanted you to address the cells vs. cell equivalents. Also talk about amplification and qPCR. Keep in mind the audience was Bio111 so you needed to explain a bit more than some of you did.
b) Interpret as fully as you can the data in the figure to the right. It is from the same paper as part a) above. Do not look for this paper or read the abstract or the paper – read only what I have supplied. Young means heart cells from a young mouse and old means heart



cells from an old mouse. You may look up the genes from any source you want other than the paper.

Key points were the greater spread of data in old vs. young. It is not clear why this would happen. Also wanted you to comment on the differences of Cox2 and Lpl. Why is this so tight and Lpl so varied? Wanted you to link function to variance and wanted you to use a genomic database rather than Wikipedia or Google.

c) Design and experiment to determine if the global amplification introduces non-biologically relevant variation in the procedure.

Wanted you to compare before and after amplification of these two genes using qPCR or microarrays.

2) *30 points*

The figure on the next page displays transcriptome DNA microarray data from a study of a genetic disease called FPM where high FPM is considered very sick. Severity of phenotypes are displayed in colors along the right Y-axis. Genes are displayed along the X-axis.

a) Interpret this figure as completely as you can.

Needed to define axes, discuss the clustering of genes and patients, and the subdivision of high FPM.

b) Analyze the challenges for using DNA microarrays to categorize the FPM disease state of these people?

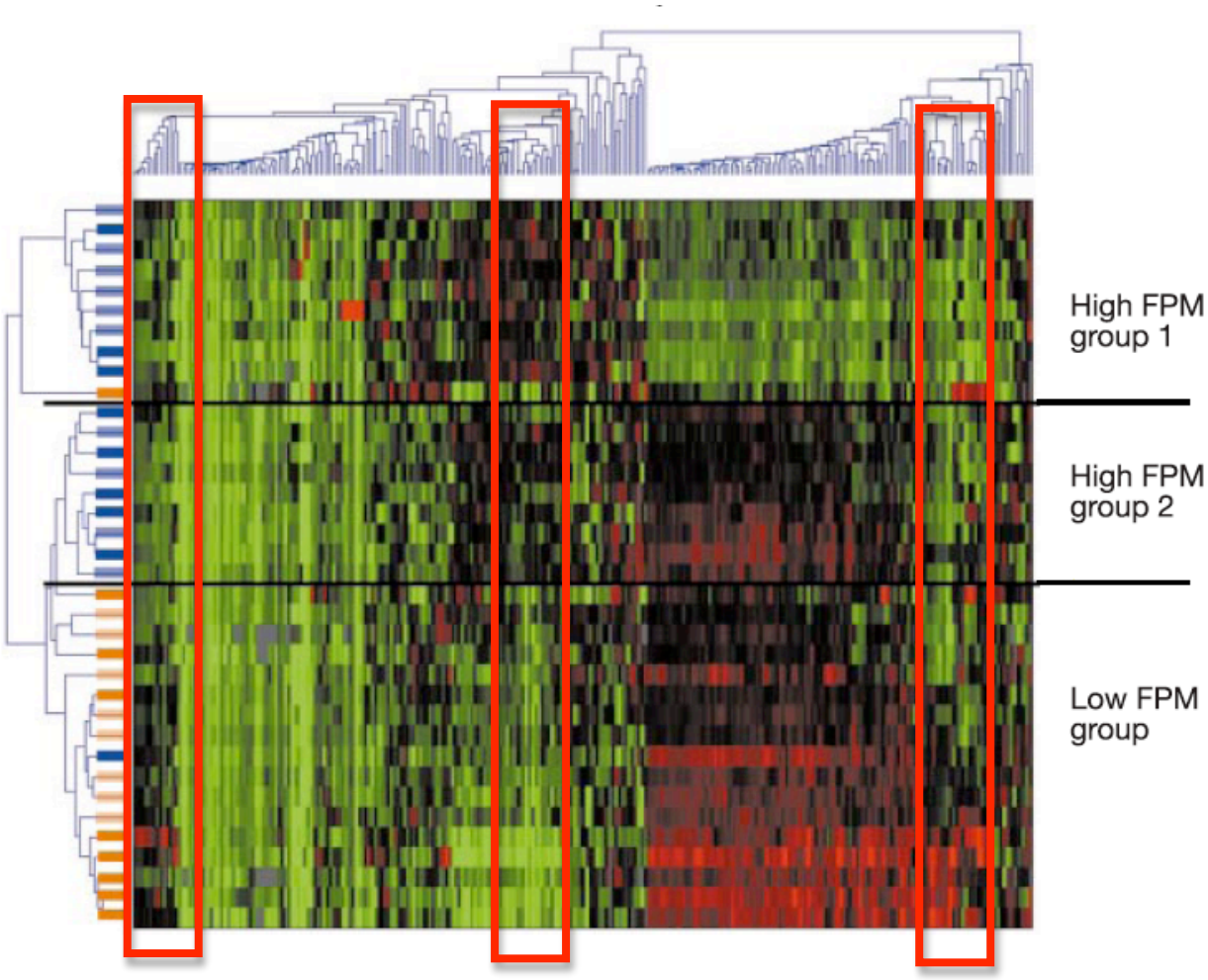
How do you subdivide when the clinical categories do not match overall gene categories?

c) Propose which genes/mRNAs you would use to distinguish high FPM patients from Low FPM people. You may draw shapes or use arrows to point to the genes you have chosen. Don't try to find gene names.

See my red boxes below. Other similar difference were acceptable.

d) Evaluate how they divided the people into three FPM groups.

Key point was the either arbitrary division of patients or the combination of symptoms with correlation coefficients.


High FPM group 1
High FPM group 2
Low FPM group