

Fall 2009 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 than 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 are not allowed discuss the test with anyone until all exams are turned in by 11:30 am on Thursday November 5. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE AT 11:30 am ON THURSDAY 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 electronic versions before 11:30 am (eastern time zone:-). You will be penalized 1 letter grade for each 24 hour period your exam is submitted late.

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)?

20 points

1) Look at the paper by Simona Granata *et al.*

6 pts

a) Describe what figure 1 is telling you at the big picture level. Don't focus on individual genes, but tell me how to read this figure in general. Address the colors, the types of people and the dendrograms.

Looking for HD different from other two and CKD as a transition from HS to HD.

5 pts

b) Study the dendrogram in figure 1. Explain how they cut the tree to form the three groupings of patients and why this presents an interesting exception to the general trend presented.

Looking for the odd example of a CKD appearing in the HS cluster. Why and how? Why would CKD patients cluster with HS – stage of disease is HD treatment causing a change in genes??

9 pts

c) Study figure 3. Describe what processes are affected by the 11 genes highlighted in this paper and how these processes are affected for the patients in this study.

Looking for groups of effected processes.

20 points (5 pts each)

2) Now look at the paper by Legoffic *et al.*

a) Explain what figure 1 is telling you. Use your own words, *not* quotes or paraphrases from the paper. You may find it easier to use a screen shot and draw annotations to the figure so I know what you are describing.

Looking for step by step of A – D.

b) Summarize the main lessons you learn in figure 2?

gene turns on early, but not first step. Some noise too.

c) What is the point of figure 3? Summarize the results for this figure.

Validating method for measuring reg4 mRNA.

d) How could the results for this paper be converted into clinically useful information? Support your answer by linking the data in the paper to the clinical application.

reg4 is a surface protein. Might be able to quantify it on the surface of pancreas cells. Could be used to block action of this protein.

Identify the typo in figure2 for 2 bonus points.

None after all. Only axis wrong in Figure 4.

30 points

3) Look at the paper by Kuntz-Melcavage *et al.*

7 pts.

a) Explain the experimental design for this paper using figure 1 to help you.

Rats addicted to heroin, then half were withdrawn for 90 minutes one day after last dose while others were given the same setup 14 days after last dose. Licks on the spout were quantified.

7 pts.

b) What very difficult social problem is this paper addressing? Utilize figure 1 when answering this question.

Why addicts come back to the drug even after being away for a long period of time.

8 pts.

c) Make a list of the 6 genes in Figure 2. Provide a short (2 sentences or less) summary of each gene's function and then state what impact the experiment had on each gene's expression.

Dusp 5 up (1 day vs 14 days)

Dusp 6 no difference

RGS2 up some

BDNF up

Calb1 up

NPY no difference

8 pts.

d) Which gene in figure 3 would be the one you would choose for possible therapeutic intervention. Explain why you chose your target gene.

PDGF BB because it has the highest degree in the graph.

30 points

4) Now look at the paper by Gry *et al.*

5 pts

a) Summarize the research question and experimental design of this paper in your own words. Use figure 1 to supplement your summary.

Did cDNA and oligo expression profiles; did immunohistochemistry and proteins arrays, and quantified all pairwise comparisons for correlations. Used 23 human cell lines and over 1000 genes/proteins.

5 pts

b) Explain in words why the four genes in figure 2 have the correlation coefficients they do. I don't want you to describe the calculations, I want you to use the visual data to explain why $a > b > c > d$.

We are looking at the blue vs. red lines and how closely they match each other, or not. It is hard to visually validate $a > b$ because the impact of a few poor fits between red and blue.

8 pts

c) Use the same source of protein information they used in this paper to show me (with screen shots) the cellular component of 4 proteins in figure 2. Furthermore, I want you to name 2 tissues

that produce a lot of each protein and 2 tissues that produce very little or no detectable amounts of each protein.

Varied answers.

4 pts

d) Explain the significant finding summarized in figure 4?

1066 total analyzed

238 top left (cDNA – protein)

292 top right (oligo – protein)

678 bottom (RNA-RNA)

4 pts

e) Summarize the take home message in figure 5. Choose 2-3 cell types to use as examples in your summary. Please take a screen shot and draw an arrow pointing to your chosen genes in each tree.

The cell lines cluster similarly to each other with the DNA microarray data than they do to the proteome data. However, the proteome data had the higher Pearson correlation coefficient than the DNA microarray data which says that the transcriptome produced less homogeneous cell line data than the proteome data did for these 169 genes/proteins.

You can find several examples such as RH-30 and U-251MG that cluster together in the first two trees, but not the bottom tree.

4 pts

f) Describe the Y axis in figure 5, how it should be read, and why panel c has a different scale than panels a and b.

Correlations among cell lines for proteomes are closer to 1 (all 0.455 or higher) than correlations for DNA microarrays/transcriptomes which must be as low as 0.05 for the cDNA microarray and 0.01 for the oligo microarray.

This seems to be an artifact of the genes chosen. Why would the correlation be so low for microarrays compared to proteins? Regardless, the Y-axis makes it seem as if the clustering of cell lines for the microarray data is based on lower correlations than the protein tree.