Spring 2011 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 about the same amount of time as the first one. You are not allowed to read any papers to help with this exam. There are 6 pages, including this cover sheet, for this test. There are no Discovery Questions on this exam. You are not allowed discuss the test with anyone until all exams are turned in at 9:30 am on Wednesday March 23. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE AT 9:30 am ON WEDNESDAY MARCH 23**. You may use a calculator, a ruler, your notes, the book, and the internet. You may take this exam in as many blocks of time as you want. Submit your electronic version before 9:30 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 PAPERS FOR THIS EXAM. RELY ON 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.

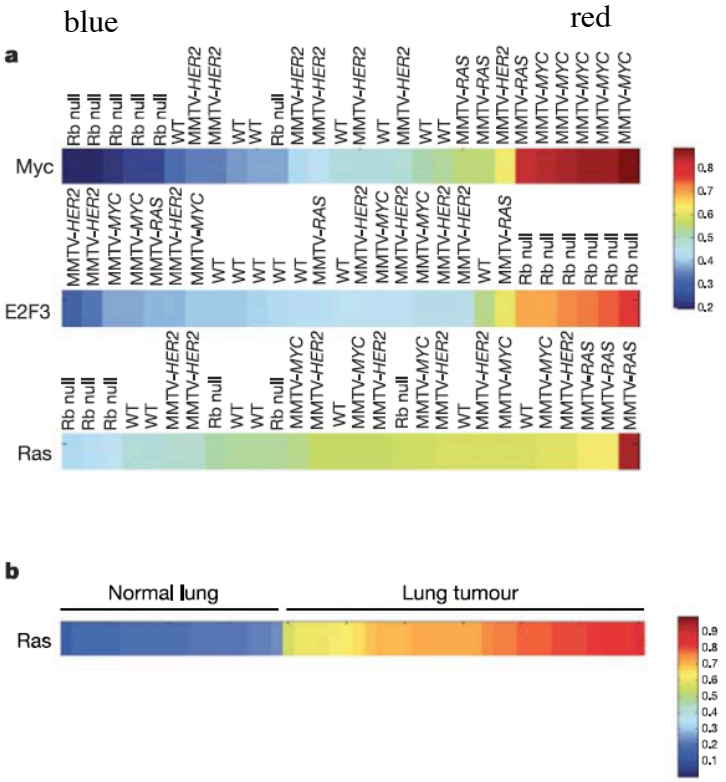
Name (please type):

Write out the full pledge and sign (electronic signature is ideal):

How long did this exam take you to complete?

16 points

1) Investigators engineered human cells to over-express three oncogenes (Ras, Her2 or Myc) or they deleted the tumor suppressor Rb gene (both alleles). In panel a) they used Affy DNA microarrays to measure the activity of three different oncogenic pathways (Myc, E2F3 or Ras) and the color-coded each engineered cell line based on the probability of the 3 distinct pathways to be activated. In panel b) non-engineered cells isolated from human patients were analyzed as in a).



a) What can you conclude from the data in part b? Limit your answer to a 3 sentence maximum.

b) How could you use these data to personalize cancer treatment for any cancer type? 3 sentence maximum.

c) Use NCBI to find a new drug (not FTS) that might work especially well on cancers with activated Ras pathways. What drug did you identify? You must include a screenshot of your findings and the URL of the NCBI resource that you used to reach this answer.
http://www.ncbi.nlm.nih.gov/pubmed/19010291

d) Find a visual pathway map that shows the relationship between Rb and E2F3. In 3 sentences or less, explain the data in the figure used in this question given what you found for your pathway map. Be sure to include a screen shot and the URL for the pathway map you find.
http://www.genome.jp/kegg-bin/show_pathway?hsa05223

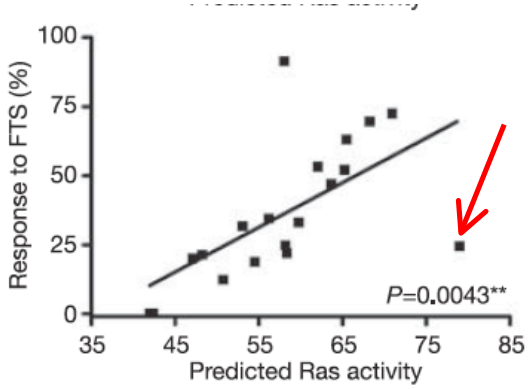
9 points

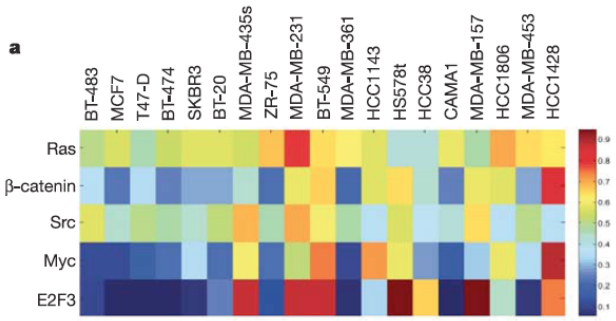
2) Another figure from the same paper looking at oncogene pathways and how to cure cancers.

a) Use the data from the figures below to determine which sample in the multi-color figure is the one pointed to by the red arrow. Just name the sample.

b) What conclusion(s) can you draw about personalized medicine based on the figures below? Limit your answer to 3 sentences or less.

c) What can you deduce given the outliers in the black and white graph on the left below? Limit your answer to 3 sentences or less.





12 points

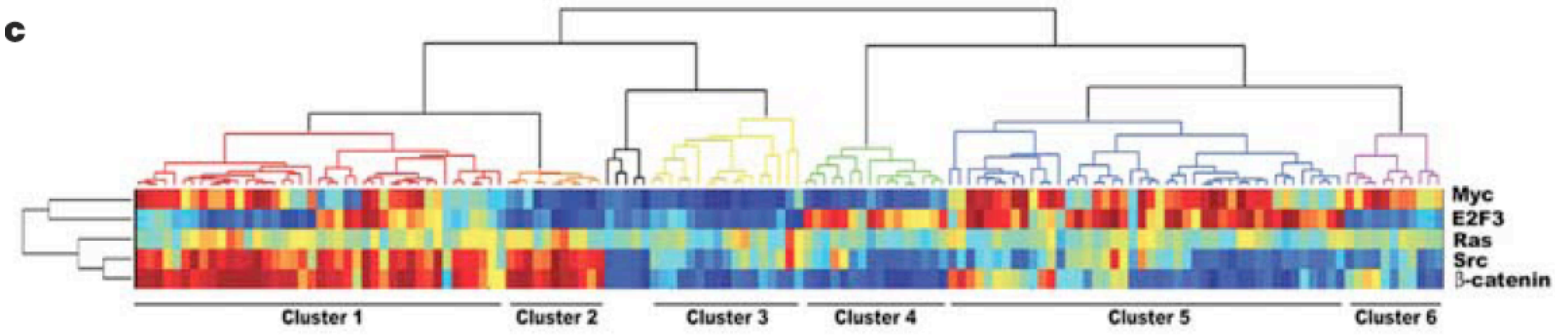
3) Look the third and final figures (below) relating DNA microarrays to cancer treatment. As before, the heat map was used to indicate the probability of pathway activation. In this case, they have taken breast cancer biopsies and run Affy chips on them to determine which pathways might be activated.

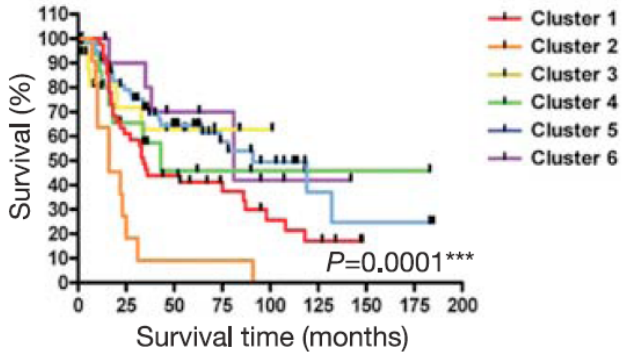
a) How many types of breast cancer are depicted in this figure? In two sentences or less, explain how you reached your number.

b) In two sentences or less, describe how the investigators cut the tree to generate their clusters.

c) Use a numbered list to tell me the molecular pathway(s) in the most aggressive breast cancer subtype.

d) Which breast cancer subtype is the least lethal? Describe the 5 molecular pathways for this type of cancer in 3 sentences or less.



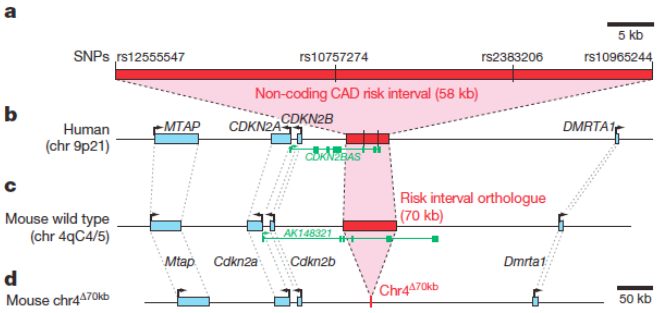


24 points

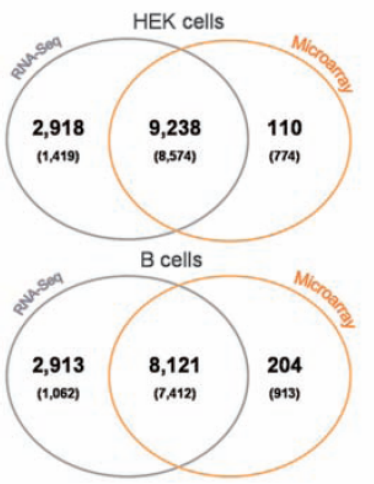
4) For this question, I am providing you with the abstract and one figure. The "questions" ask you to draw data that match the text. My goal is to measure how well you can understand the jargon-rich text and communicate the meaning in a non-verbal way. You may do this by hand and scan your images, or draw them electronically. I recommend you do a rough draft first. I need to be able to read your hand writing, so write neatly.

a) Draw a picture of well controlled experimental results showing the transcription levels of Cdkn2a and Cdkn2b in cardiac tissue.
b) Diagram what the abstract is describing about the experimentally produced heterozygous mice.
c) Diagram the experimental results showing what happens to aortic smooth muscle cells in homozygous deletion mice.
d) Draw a picture of the cardiac blood vessels in a homozygous mutant mice that explains why CAD results from the engineered deletion.
e) In words, speculate why this deletion causes the phenotype it does. Limit your answer to 3 sentences or less.

Sequence polymorphisms in a 58-kilobase (kb) interval on chromosome 9p21 confer a markedly increased risk of coronary artery disease (CAD), the leading cause of death worldwide[1,2]. The variants have a substantial effect on the epidemiology of CAD and other life-threatening vascular conditions because nearly one-quarter of Caucasians are homozygous for risk alleles. However, the risk interval is devoid of protein-coding genes and the mechanism linking the region to CAD risk has remained enigmatic. Here we show that deletion of the orthologous 70-kb non-coding interval on mouse chromosome 4 affects cardiac expression of neighbouring genes, as well as proliferation properties of vascular cells. Chr4$^{\Delta 70kb/\Delta 70kb}$ mice are viable, but show increased mortality both during development and as adults. Cardiac expression of two genes near the non-coding interval, *Cdkn2a* and *Cdkn2b*, is severely reduced in chr4$^{\Delta 70kb/\Delta 70kb}$ mice, indicating that distant-acting gene regulatory functions are located in the non-coding CAD risk interval. Allele-specific expression of *Cdkn2b* transcripts in heterozygous mice showed that the deletion affects expression through a *cis*-acting mechanism. Primary cultures of chr4$^{\Delta 70kb/\Delta 70kb}$ aortic smooth muscle cells exhibited excessive proliferation and diminished senescence, a cellular phenotype consistent with accelerated CAD pathogenesis. Taken together, our results provide direct evidence that the CAD risk interval has a pivotal role in regulation of cardiac *Cdkn2a/b* expression, and suggest that this region affects CAD progression by altering the dynamics of vascular cell proliferation.



12 points



5) This question is also a learning moment.
To the right are two Venn diagrams showing the number of genes detected by either direct RNA sequencing (RNA-seq) or by DNA microarrays. HEK = human embryonic kidney cell line. Bigger, bold numbers indicate the number of genes measured by the appropriate method at least once while the smaller (numbers) indicate genes measured at least five times.

a) What trend do you notice when comparing the two methods?
b) As a skeptical scientist comparing a new method with a more familiar method, what additional data would you like to see before determining if the new method was valid or not.
c) What aspect of the data visualization is less than optimal if you are skeptical of this new method?

16 points

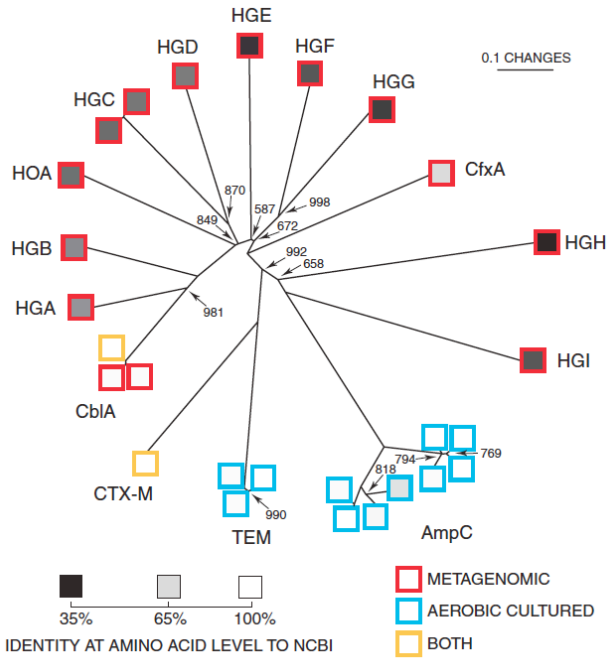
6)
rs=2298771:
a) tell me what gene this is, what sort of genomic event is being described, what phenotype is associated with this number, and the population distribution in 4 distinct groups. To receive credit, you must provide me with all links to the sites you used. Remember, do not read any papers for this – use only credible databases.

gq343003
b) Tell me what gene this is, what phenotype it is associated with, where the DNA came from.
c) Find a screen shot of the sequence logo for this class of protein and determine how well gq343003 matches this sequence logo.
http://www.expasy.org/cgi-bin/prosite/sequence_logo.cgi?ac=PS00146
d) What degree of amino acid identity does this protein have with its highest BLAST hit for a known species? Show me a screen shot of your best BLAST results of a known species.

11 points

7) This last question is pretty straight forward. Coincidentally, it is related to the bioinformatics talk this week!
a) What type of phylogenetic tree is shown below? What name is associated with this type of tree? (Same question worded two ways.)
b) To what family of molecules do AmpC, TEM, CTX-M and CblA belong? How did your reach this conclusion?

c) What is the medical consequence of this metagenomic study given that the HG_ series represent different examples from the same protein family? The DNA for the HG_ series was isolated from fecal samples taken from only two people.



Bonus Question (3 pts):

What innovation did the group invent that allowed them to test every exon in the human genome for expression using DNA microarrays?