Fall 2004 Genomics Exam #2
Genomic Medicine and Sequencing Tools

There is no time limit on this test, though I don't want you to spend too much time on this. You know I work hard to design challenging tests, but not ones that are excessive. You do not need to read any additional papers other than the ones I send to you. There are three pages for this test, including this cover sheet. You are not allowed discuss the test with anyone until all exams are turned in at 11:30 am on Friday November 5. **EXAMS ARE DUE AT CLASS TIME ON FRIDAY 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 need to. NOTE: I leave town on November 5 and I want to take the tests with me to grade. Submit your paper and electronic versions before 11:30 am so I can take them with me.

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. 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. Print this test but make sure the screen shots are big enough to be seen easily. 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 DOWNLOAD ANY PAPERS FOR THIS EXAM. RELY ONLY ON THE FIGURES PROVIDED, YOUR EXPERIENCE AND YOUR SKILLS.

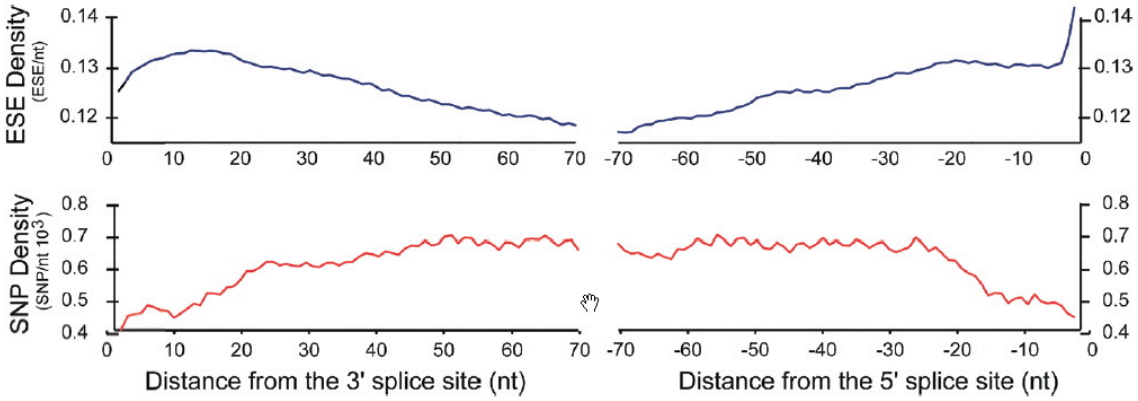
-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please print):

Write out the full pledge and sign:

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

20 pts.
1)



a) Interpret the figure above. Do not look for any online sources of information to answer part a.
b) Go to this AceView database and query for the human gene aqp1.

http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human

c) Use the "gene on genome" and "annotated mRNA" tabs at the top to help you navigate.
d) How many mRNAs are produced from this gene? Support your answer with an image.
e) Use the bDec03 version to answer the following questions.
f) How many exons are in this mRNA?
g) Choose one exon only, copy and paste its sequence into the form you see on this page.

http://genes.mit.edu/burgelab/rescue-ese/

Show the ESE you found with a screen shot.

h) Was the ESE you found above consistent with the data presented in the figure for part a? Explain your answer.

20 pts.
2) Several studies using microarrays have been performed on DLBCL. The figures below summarize much of this work.
a) Interpret figure 1 below.

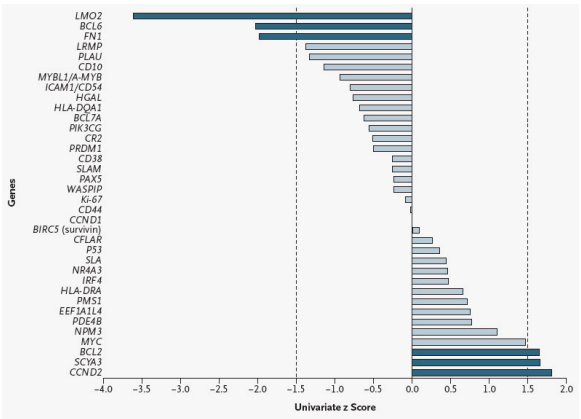
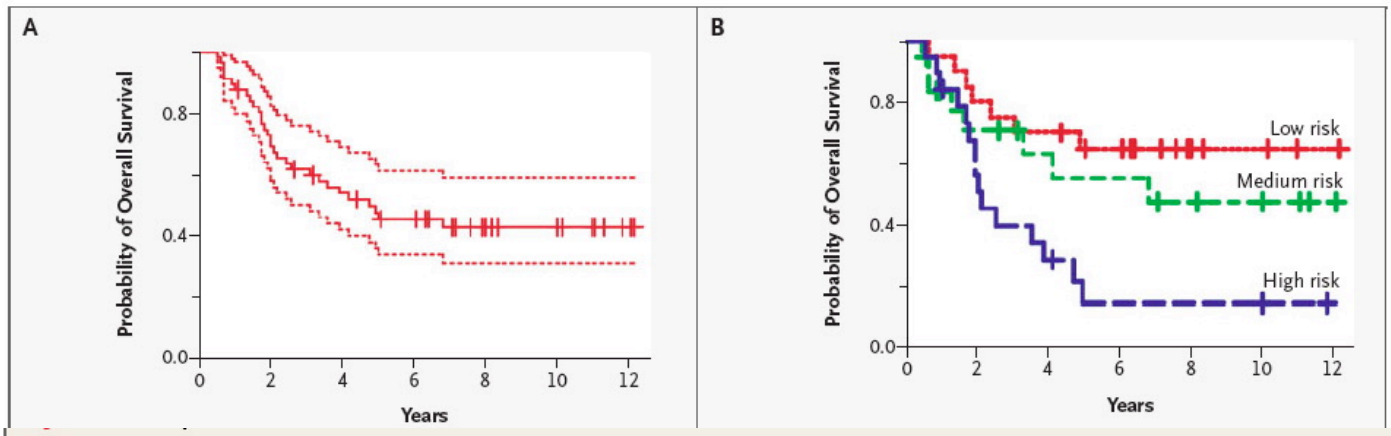


Figure 1. **Univariate Analysis of Expression of 36 Genes with Overall Survival as a Dependent Variable.**
The genes are ranked on the basis of their predictive power (univariate z score), with a negative score associated with longer overall survival and a positive score associated with shorter overall survival. The dashed lines represent an absolute univariate z score of ±1.5. The prediction model is based on the weighted expression of six genes and is expressed

b) Interpret this figure below



Panel A shows Kaplan–Meier estimates of overall survival in the 66 patients with diffuse large-B-cell lymphoma, analyzed by quantitative reverse-transcriptase polymerase chain reaction with TaqMan probe-based assays. The dotted lines represent 95 percent confidence intervals. Panel B shows Kaplan–Meier curves for overall survival in the three groups (at low, medium, and high risk of death) as defined by a prediction

c) Given the different method for classifying risk of survival with DLBCL compared to the version in your genomics textbook, do you think this meta-analysis method is better, worse, or the same? Explain your answer to get full credit.

d) Go to the lymphoma search page
http://genome-www.stanford.edu/lymphoma/search.shtml

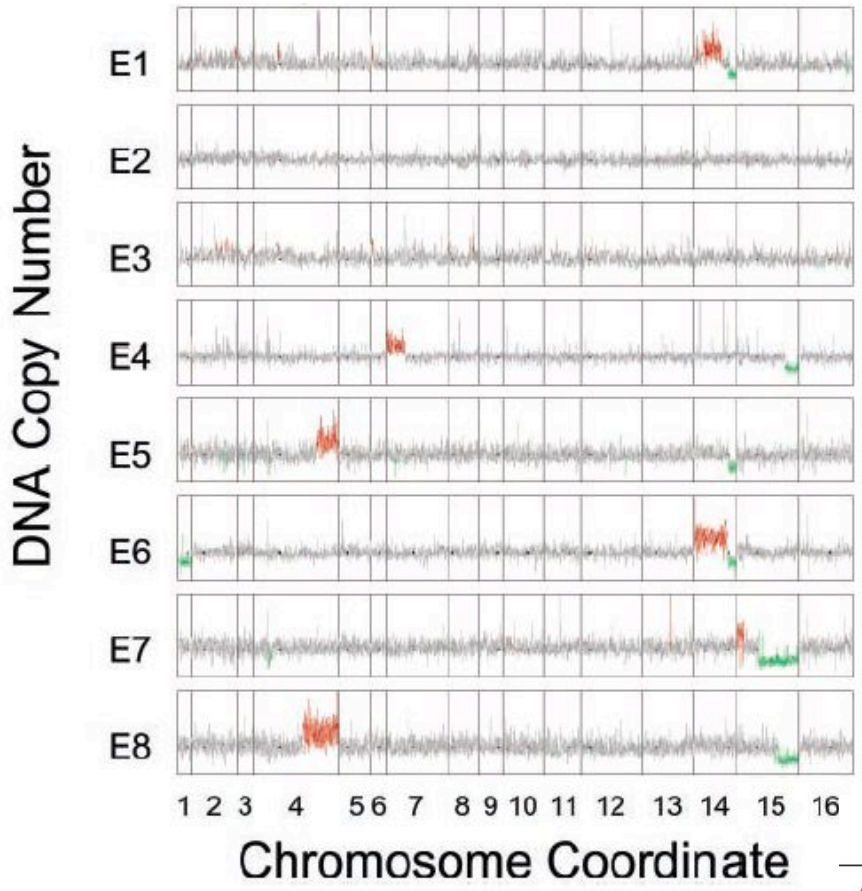
and see if you can validate any of the genes indicated as important from this meta-analysis.

Two hints at no extra charge:
Use BCL-6 instead of BCL6.
Use ttg-2 instead of LMO2

e) There is some ambiguity from the figure 1 for part a of this question. However, after doing these searches (part d), you should be able to answer this question now. For each z score, is it better for the patient's survival to have the best indicator genes induced or repressed. Explain how you determined your answer. Refer to figure 1 as necessary.

20 pts.
3) In the figure below, you see the expression ratio for every yeast for 8 evolved (E) strains. All the strain were grown in constant glucose-limited conditions for hundreds of generations. The 8 strains began their evolution as genetically identical isolates from the same parental strain.

a) Find the gene HXT6 in these graphs. Do the best you can to narrow the exact location and draw an arrow on this test where you think it is located. Explain the rationale you used to pinpoint HXT6's location. You must tell me where you found the information required for your answer in order to receive full credit.
b) Find a region of the genome that experienced a lot of aneuploidy. On this test, circle your region of choice.



Segments with altered copy number
Amp. of YNL telomere-YNL066W Del. of YNR002C-YNR telomere
Amp. of YGL telomere-YGL098W Del. of YOR290C-YOR telomere
Amp. of YDR327W-YDR telomere Del. of YNR002C-YNR telomere
Amp. of YNL telomere-YNL018C Del. of YNR002C-YNR telomere
Amp. of YOL telomere-YOL107W Del. of YOR012W-YOR telomere
Amp. of YDR211W-YDR telomere Del. of YOR193W-YOR telomere

c) Using the table to the right, find the exact location for one end of the aneuploidy for one of the strains in the table. Look at chromosome and describe any features that like good candidates for the site of the chromosomal rearrangement. You may want to compare what you found to a few other sites.

d) Find one location for the aneuploidy and search the Tup1 and Yap+++ databases from the DeRisi paper to see if you can verify similar sites of aneuploidy in the either mutant strain used by DeRisi. You do not have to find it exists, but you do have to document how you showed it does or does not exist in both mutant strains. I strongly encourage you to combine screen shots with your explanations.

20 pts.

4) First, I want to tell you this question is the most experimental one I have written in the test. If you try to just grunt it out and systematically do every permutation, you will waste WAY too much time. I strongly suggest you do a quick pass with the list on both sites, navigate the sites some to see what you can find, and grab a few screen shots to remind you of your options. Then, answer the questions below.

Go to these two web sites
http://transcriptome.ens.fr/ymgv/access_2.php

http://db.yeastgenome.org/cgi-bin/expression/expressionConnection.pl
to analyze the genes in this list.

ELM1
SRC1
FKH1
YOL128C
GSP2
WSC4
GPA1
YBL009W
YML125C
YGL101W
STU2
YOR073W
CIN8
YEL017W
YLR455W
STR3
Cdc28
CRC1
YRO2
MET17
STB1

a) Do these genes have anything in common? In other words, can you use their expression profiles to find any patterns? You do not have to consider all of these genes as a group. Feel free to pull out subsets that appear to be associated. To answer this question for full credit you must:
b) Show me the data you used to support your answer.
c) Explain the logic you used to produce your answer.
d) Compare the data produced by the two databases and decide which was more helpful.
e) Explain why one database was more helpful to you than the other.
f) Validate your conclusion with data from another site and show me the data.

20 pts.
5) This question has you work with an excel sheet to approximate what real microarray data analysis is like. You must download this excel file before you can answer any of these questions.

URL: http://www.bio.davidson.edu/courses/genomics/exams/2004/Exam2_04.xls

Do all your work on this excel file and submit it separately by email along with your final exam Word file. Print the two worksheets out, but make sure you limit the printouts to single pages (i.e. don't waste paper or time printing unnecessary pages.

a) On the 10 hour sheet, determine the color for each of the genes. To answer, simply copy and paste the color from the table into an appropriately labeled column. You may

create as many columns as you see fit. Be sure to label each column so I know what you are thinking.
b) Copy your 10 hour data and paste it into the appropriate column on the All Times sheet.
c) Finish the graph by including the data you added to the 10 hour column.
d) List two pairs of genes that have a correlation near 1.0. Explain your answer.
e) List one pair of genes that have a correlation exactly 1.0. Explain your answer.
f) List one pair of genes that have a correlation very near –1.0. Explain your answer.
g) List one pair of genes that have a correlation very near zero. Explain your answer.
h) What would have helped you answer these questions more efficiently? Explain your answer.