Fall 2003 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 7. **EXAMS ARE DUE AT CLASS TIME ON FRIDAY NOVEMBER 7**. 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 WILL BE OUT OF TOWN NOV. 4-9 SO DO NOT WAIT TO ASK ME QUESTIONS.

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. Turn in the hard copy by the deadline to Mrs. Hartsell in the Biology Office or slide your test under my door.

Suggestion: Start with the MAGIC Tool problem first while I am still in town. This will allow you to ask questions before I fly to Hungary (no email access for me)

-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)?

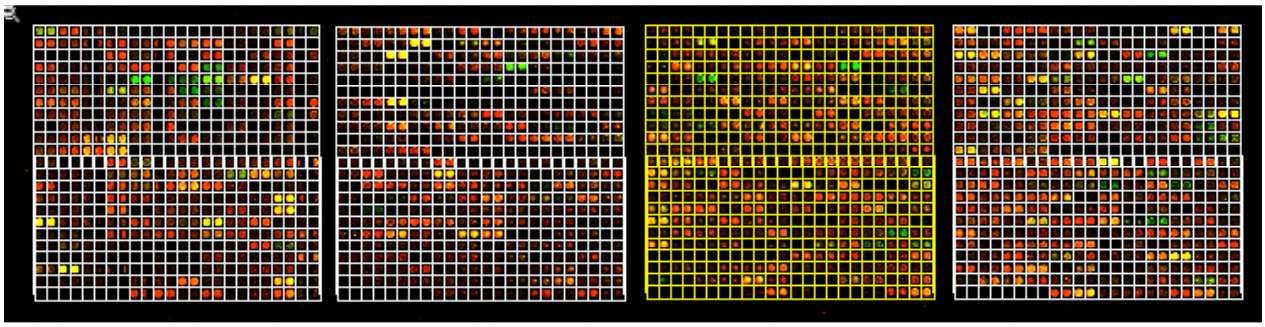
I have provided guidelines for good answers. I have not spelled out complete answers so these should not be used by future students as models for complete answers.

25 pts.
1) Go to this web site <www.bio.davidson.edu/courses/genomics/Exams/2003/QTL.pdf>. Read the abstract to get a general sense of this paper's findings. You do not have to understand this paper completely to accomplish your task.

Task: Summarize this paper using your own words and not quotes from the original paper. Each time you make a statement in your summary, support your statement with a figure. You should describe how the figure you have chosen supports your statement. The goal is to write a user-friendly version of the abstract with figures to support the claims.

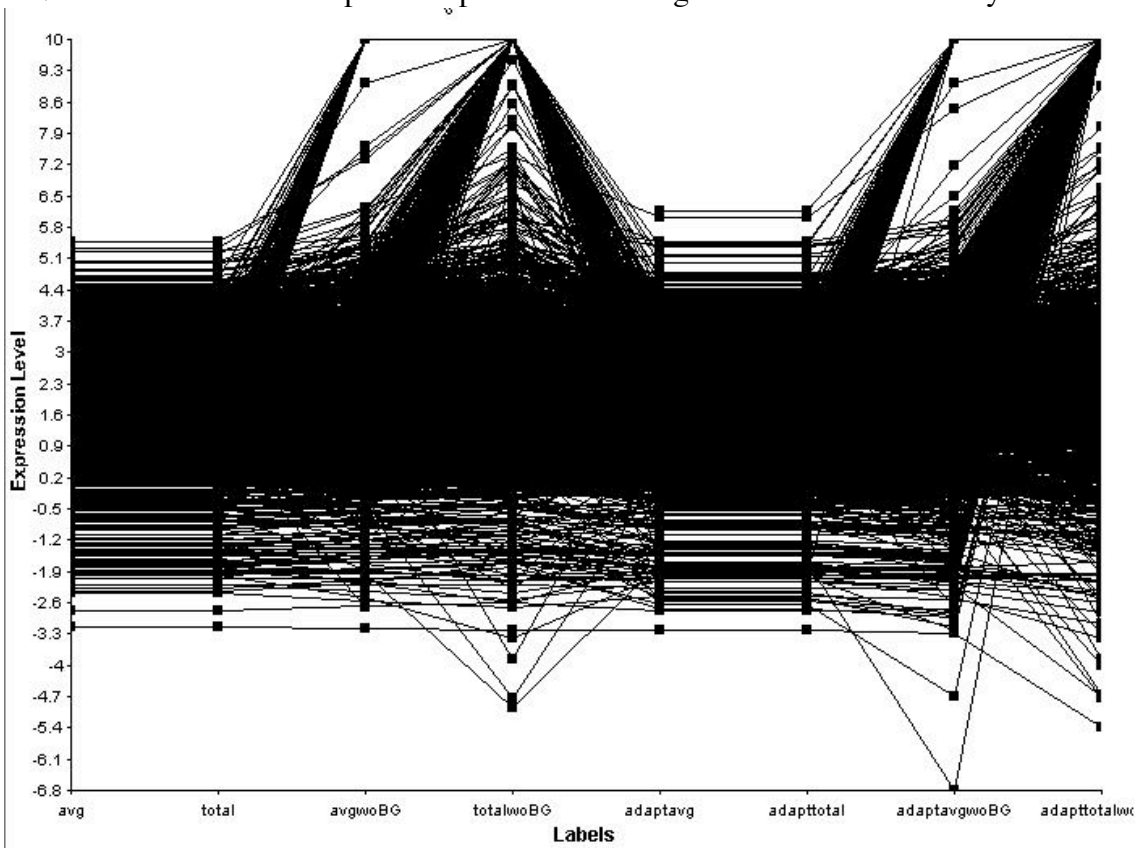
I was looking for a series of screen shots from the paper that illustrated the main points as provided in the abstract. Some key features included:
Allele Q has an A while wt (q) has a G at one base that created the QTN (figure 1).
This region is normally methylated in liver but not muscle (figure 2A).
It is paternally expressed (cite paper and textbook).
The Q promoter works better than the q version (figure 2C).
The Q or methylated q does not bind with a repressor protein but q does (figure 2b).
In vivo, Q alleles produce more mRNA than q alleles (figure 3).
A single base change in one intron can alter a QTL which is a surprise since we tend to think of QTL as too complex for such simple causes.

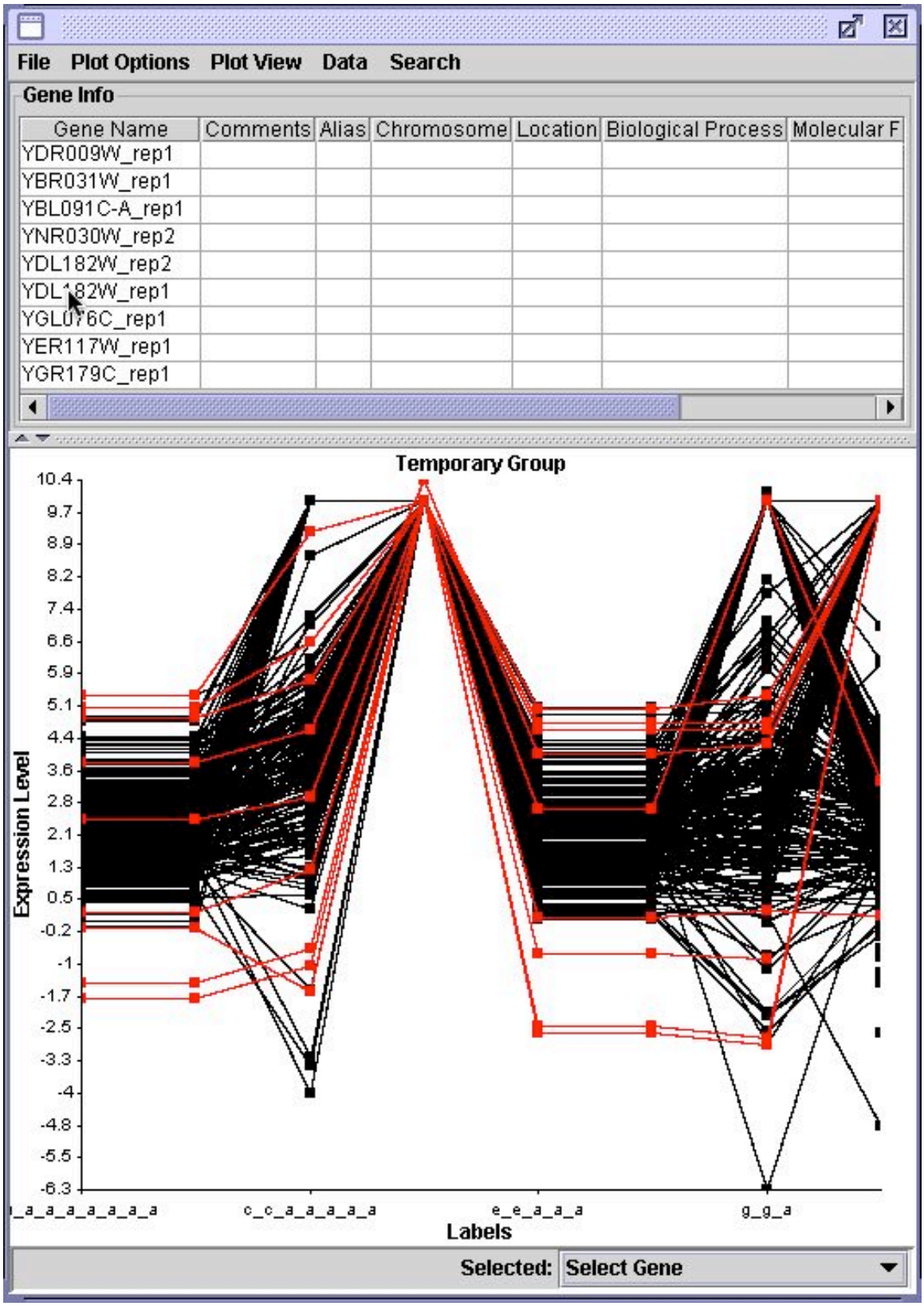
25 pts.
2) This question requires you to use MAGIC Tool. Launch MAGIC Tool properly so you can get the full RAM capacity by using MAGIC_launch.
a. Download these two tiff files:
<http://gcat.davidson.edu/MAGIC/exam2_red.tif>
<http://gcat.davidson.edu/MAGIC/exam2_green.tif>
b. Download the gene list
<http://gcat.davidson.edu/MAGIC/genelists/examGene_List.txt>.
c. Address and grid all four grids. The description for addressing the grids can be found here: <www.bio.davidson.edu/projects/GCAT/workshop.html> under the heading "Addressing these data". You have been given a subset of the files used in this workshop, but the printing pattern is the same as shown in the screen shot.
d. Produce a single expression file that includes these four grids measured eight ways:
fixed circle average
fixed circle total
fixed circle average background subtracted
fixed circle total background subtracted
adaptive circle average
adaptive circle total
adaptive circle average background subtracted
adaptive circle total background subtracted



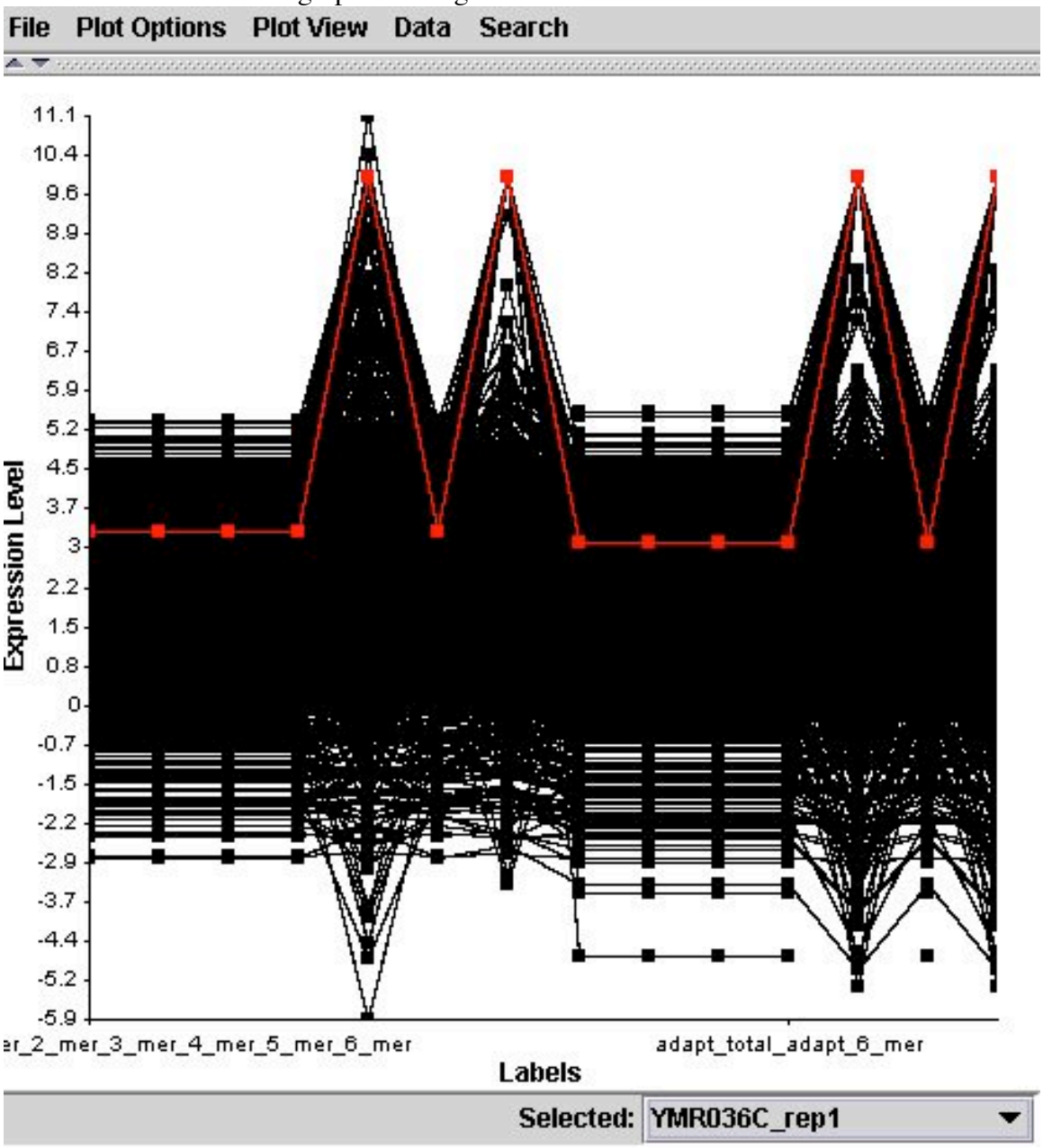
Answers are extracted from students' submissions:

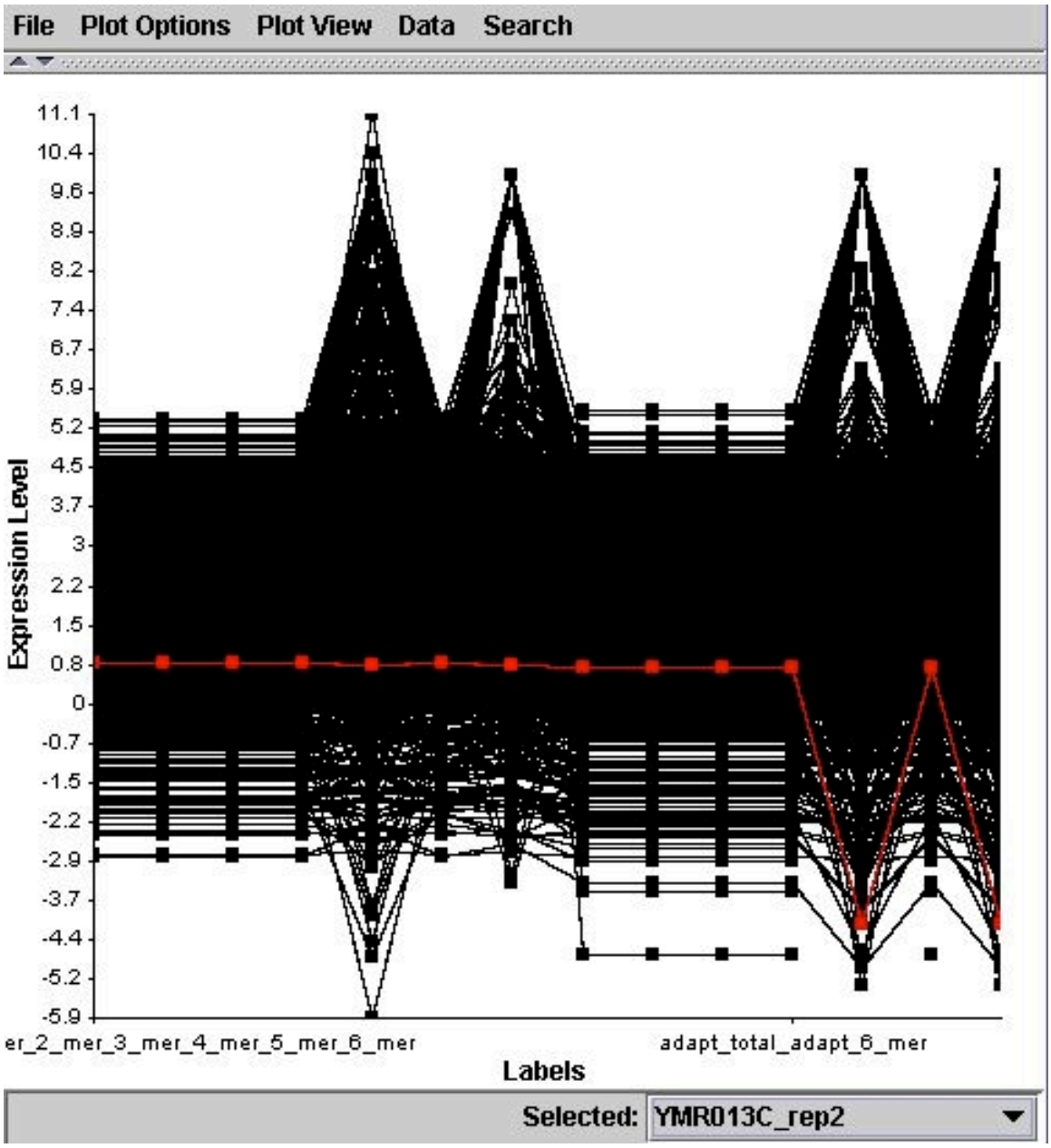
e. View the unclustered expression profiles of these genes measured all 8 ways.



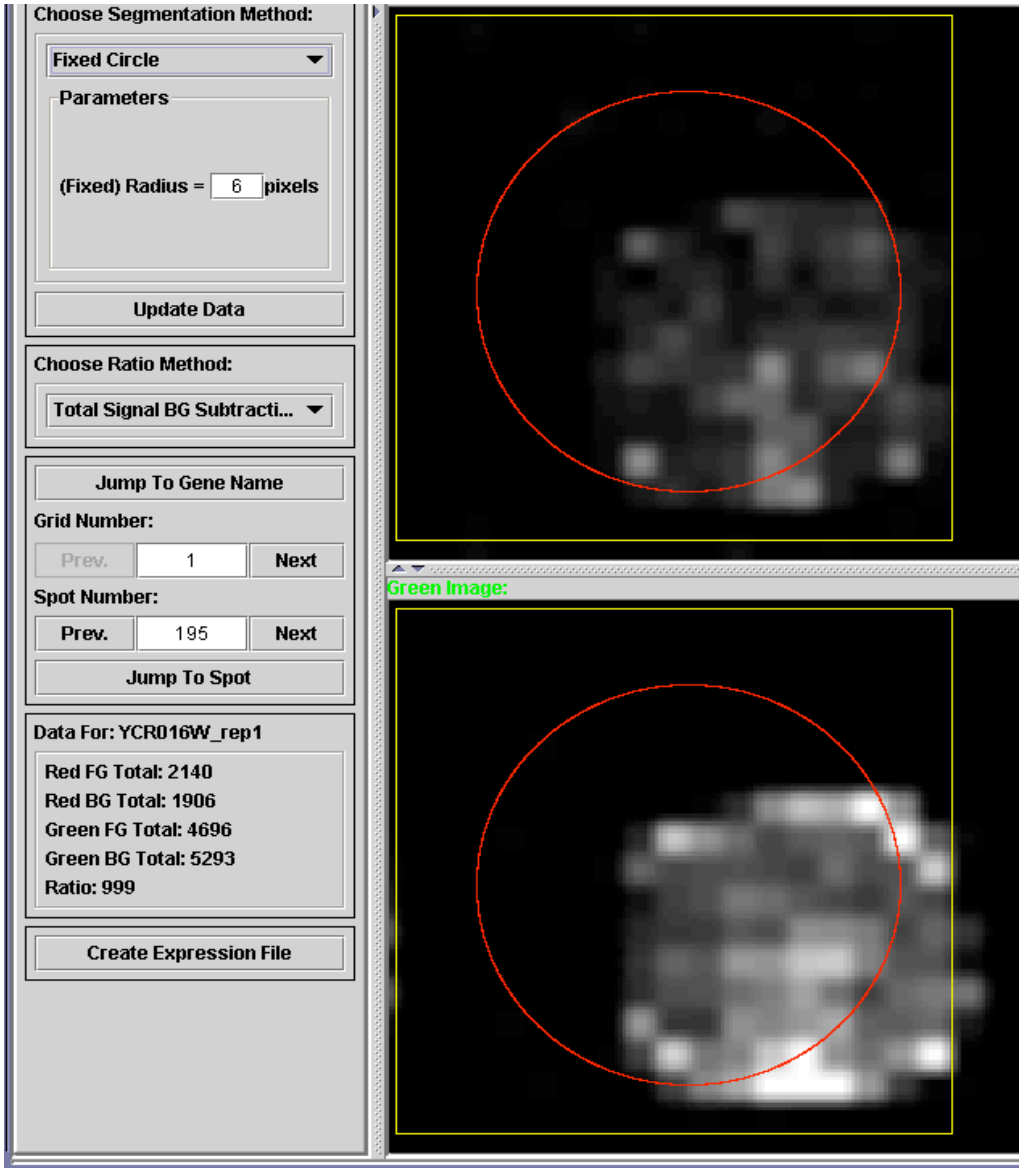

File
Plot Options
Plot View
Data
Search
Gene Info
Gene Name
Comments
Alias
Chromosome
Location
Biological Process
Molecular F
YDR009W_rep1
YBR031W_rep1
YBL091C-A_rep1
YNR030W_rep2
YDL182W_rep2
YDL182W_rep1
YGL076C_rep1
YER117W_rep1
YGR179C_rep1
Temporary Group
Expression Level
10.4
9.7
8.9
8.2
7.4
6.6
5.9
5.1
4.4
3.6
2.8
2.1
1.3
0.5
-0.2
-1
-1.7
-2.5
-3.3
-4
-4.8
-5.5
-6.3
c_c_a_a_a_a_a
e_e_a_a_a
g_g_a
Labels
Selected:
Select Gene

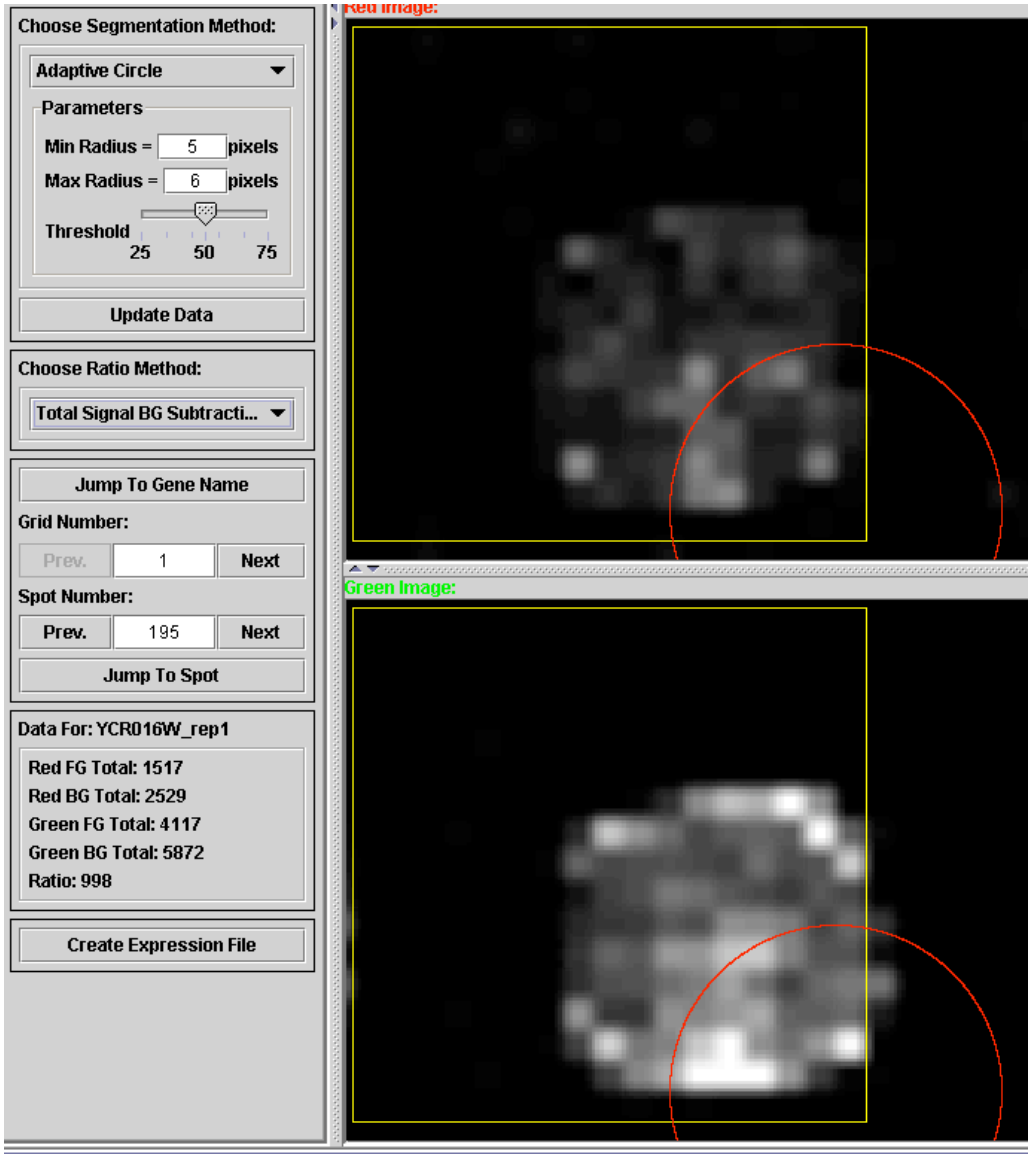
f. Identify any genes that appear to vary in their expression by a substantial amount when measured by one of these 8 ways. Name the genes you have identified and provide me with a screen shot of this graph showing the outliers.





g. Go back to <u>one</u> of the spots you identified in f above and get a screen shot of this spot.


Choose Segmentation Method:
Fixed Circle
Parameters
(Fixed) Radius = 6 pixels
Update Data
Choose Ratio Method:
Total Signal BG Subtracti...
Jump To Gene Name
Grid Number:
Prev.
1
Next
Spot Number:
Prev.
195
Next
Jump To Spot
Data For: YCR016W_rep1
Red FG Total: 2140
Red BG Total: 1906
Green FG Total: 4696
Green BG Total: 5293
Ratio: 999
Create Expression File
Green Image:



h. Explain why this spot was an outlier by one or more ways of measuring the ratios.

The remaining questions all take advantage of the Open Access permitted through Public Library of Science (PloS). It also focuses on the life cycle of the parasite *Plasmodium*. To help you interpret figures from this paper, you might wan to visit <http://malaria.ucsf.edu/Figures.php> if you want to change from red/green to yellow/blue. You do not have to understand the entire paper in order to answer these questions, but I strongly recommend you read the whole paper.

30 pts.
3) Download this paper <http://malaria.ucsf.edu/paper.php>.

a. Summarize what is happening to the *Plasmodium* parasite by general category from figure 2. Pay special attention to panels B-M. You don't need to describe in great detail, just summarize the main activity that is tied to the life cycle of *Plasmodium*.

I wanted you to state how each major signature gene class spoke to a major step in the life cycle of *Plasmodium*. You needed to address each panel B-M and relate the category to where the parasite was in its life cycle. Two really interesting aspects were the glycolysis v. TCA clusters and the overlap of panels M and B.

b. What does this paper say about isozyme expression?
Some isozymes were clustered together as in panels B-M. However two exceptions were noted on page 90 where one was expressed in clusters you would have expected but the other one was not. This raises the issue of alternative roles or need for more specific gene regulation that can best be achieved by two separate promoters.

c. Summarize figure 3A and 3C.
Key points:
3A - no special association with gene regulation on the regular chromosomes. There is sequence divergence between the reference sequence and the strain of Plasmodium used in this study.
3C – plastid genome showed a lot of coordinated gene expression. But as one student noted, only a subset of genes were shown and we might wonder about the others.

d. Summarize figure 4A.
The goal was to identify proteins (via their genes) that might be destined for the parasite specific and essential organelle. These proteins would make good targets for new drugs.

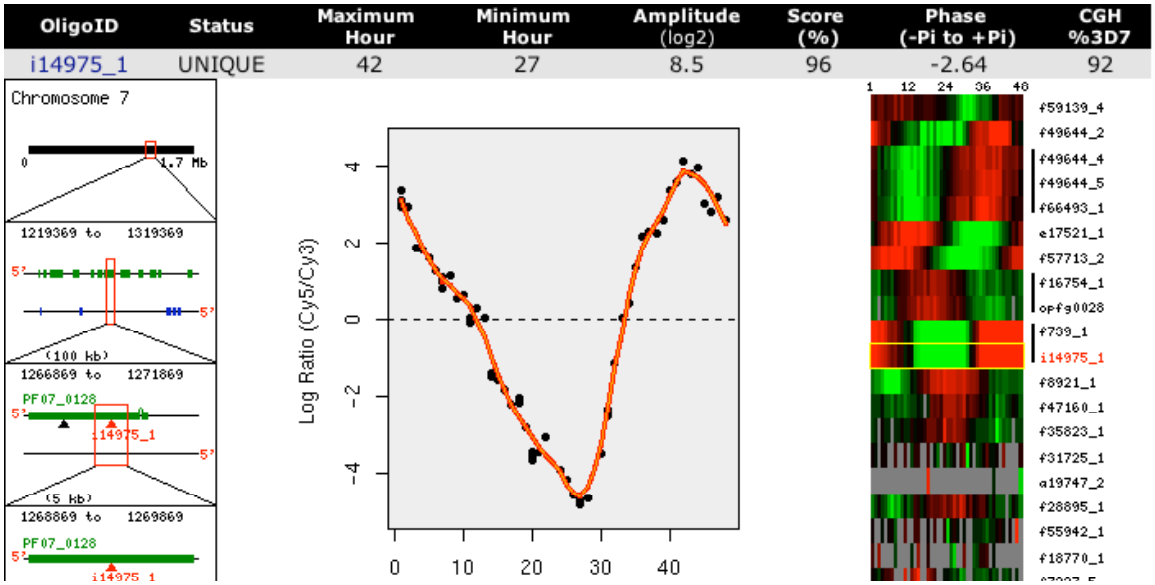
e. Summarize figure 5.
This figure showed known and prospective proteins that might make good vaccine candidates due to their timing of expression (when the parasite is infecting new RBC's) and probably expression on the cell surface.

20 pts.
4) Go to this interactive dataset <http://malaria.ucsf.edu/>. Get familiar with this interface. You should not need to use the tutorial so try to figure it out by yourself if you can (just to save time for most). However, you may use the tutorial if you want.

a. Show me the expression profile (both ways of visualizing it) for one of the known seven vaccine candidate proteins. I want the image that looks similar to figure 1 E. You can choose which one to display. Be sure and tell me which gene you chose.

Search by gene name is the easiest.

EBA175



b. Screen shot a list with expression profiles of the 6 genes with the greatest fold change in expression between induction and repression that would make good candidates for vaccines. Validate your choices by identifying at least two known candidate vaccine genes described in the paper. To validate, list the names of all the genes you selected and explain how you homed in on your genes.

Restrict the timing (induced between 35-48 hours) and ask for genes with the greatest amplitude (8 or greater).

HOME — DeRisi Lab Malaria Transcriptome Database — October 24, 2003

Oligo ID	PlasmoDB ID	Uniqueness	Maximum Hour	Minimum Hour	Amplitude	Score (%)	IDC Expression
a31914_2	PFA0440w	UNIQUE	42	25	8.1	95	
i10472_1	PFL2520w	UNIQUE	42	26	8.2	95	
i14975_1	PF07_0128	UNIQUE	42	27	8.5	96	
j53_21	PF10_0138	UNIQUE	41	25	8.3	95	
e24991_1	PFE0080c	UNIQUE	39	21	8.1	97	
n150_50	PF14_0102	UNIQUE	38	21	8.3	97	

EBA175

RAP1

3 Bonus Points: submit some constructive feedback to the authors of this web database. To get the points, you must show me a readable screen shot of your comments.

If you go to the web site now, you will see three new features have been added based on the feedback you provided to them. I am very impressed. Dr. C.