Fall 2001 Genomics Exam #2
Genomic Variations and Expressions

There is no time limit on this test, though I have tried to design one that you should be able to complete within 5 hours, except for typing. There are three pages for this test, including this cover sheet. You are <u>not allowed discuss the test with anyone</u> until all exams are turned in at 9:30 am on Friday October 26. **EXAMS ARE DUE AT CLASS TIME ON FRIDAY OCTOBER 26**. You <u>may</u> use a calculator, a ruler, your notes, the book and the internet. However, you are not allowed to obtain and read journal articles as a part of your investigations. These questions are taken from the research literature and I do not want you to simply find the papers and read the answers. This is the Honor Code at its finest.

The **answers to the questions must be typed on a separate sheet of paper** unless the question specifically says to write the answer in the space provided. If you do not write your answers in the appropriate location, I may not find them. You may want 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.

-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)?

You must answer 3 of the questions for each paper. Every one must answer the last question. Every question is worth 10 points.

**This lists the key points.
No attempt was made to write paragraph answers.**

Jaccoud et al.
1) Summarize figure 2. Print a copy (color or B&W) and point to specific spots that illustrate your interpretations.
- one microarray of 96 spots replicated 4 times
- each spot was made from genomic representation of 9 cultivars of rice
- two cultivars were used to create green (IR64) or red (Millin) genomic representations as probes

Part A
- yellow spots represent DNA shared by both cultivars
- green spots represent DNA found in IR64 but not Millin
- red spots represent DNA found in Millin but not IR64
- black spots are DNA found in other cultivars but not in either IR64 or Millin representation.

Part B
This histogram shows the ratios of bound probe for green to red. If the two were equally bound, the ratio is 1:1 and colored yellow. These are the predominant ratios as shown by the frequency of each spot (Y-axis). The two tails represent the green spots (right side) or the red spots (left side) and fold differences of > 2.9 or less than –2.9. However, the color of the bars does not match the description in the figure legend. If we go by the colors, then there are more Millin-specific targets (73) on this array than IR64 –specific targets (31).

2) Explain the significance of figure 3. Don't simply say that it validates the method. Explain why and support your explanation by citing specifics from the figure.'
- left half used probe F4 and right half used probe F8
- top half is genomic DNA and bottom half is genomic representation
- All four panels are Southern Blots with four different cultivars as source of DNA
- We are trying to validate DArT so we should compare top and bottoms of each half.
- the probes were chosen since they appeared to recognize cultivar-specific DNA. This was due to a difference in restriction fragment length polymorphism (RFLP)

- We can see probe F4 only bound one in the representative blot but the genomic DNA blot bound in all 4 lanes. This validates those spots that were one color or the other.
- We see probe F8 bound all four representations but at different intensities. This validates the not yellow but not green or red spots on the DArT microarray. In the genomic blot, F8 bound all four lanes with equal intensities, though a RFLP is evident.

3) Summarize figure 5 Explain how the data in panel A were derived and used to produce the dendrogram.
- 28 cultivar-specific spots were used from the MspI DArT. (see figure 4B)
- each cultivar was tested for each of the 28 spots
- Either the cultivar bound to the spot and was scored with a 1 or it did not bind and was scored with a 0.
- Cluster software was used to place those cultivars next to each other if they had similar pattern of 1's and 0's.
- A rooted tree was drawn (should have been unrooted) to show the relatedness of the 9 cultivars.
- The tree has two major parts – 3 cultivars on one side and 6 on the other. This matches with previous knowledge of 3 Indica and 6 Japanica cultivars.

4) Using figure 6 as a starting place, design a way to use this approach to detect and identify biological weapons.
- You would want to create a diversity panel using human DNA and every possible biological weapon. Human would be spotted on one side and each different genome would be spotted in its own region.
- Two labeled probes would be made. 1 from humans and colored red. The other from a mixture of human and unknown sample all of which would be labeled green.
- If there were only yellow spots, either the sample was not hazardous or was a new reagent. The only DNA detected was human in a 1:1 ratio.
- If there were any green spots, this would indicate the presence of the appropriate (location on the DArT would tell us which species) biological weapon.

Daly et al.
1) Summarize figure 2. In your summary, just give me the main points, don't walk through every aspect of this large figure.
- top row shows the human chromosome region of 5q31.
- next are two rows of genes for the top and bottom strands of DNA
- Next is a row of tick marks that show individual SNPs.

- these are then expanded and we see the 11 blocks of DNA that appear to resist recombination. These blocks contain multiple SNPs which are inherited as haplotypes. 103 of the most common were used to study in detail.
- The dashed lines show the most common recombination events BETWEEN block.
- The sequence shown for each block is used to represent the 'signature sequence" for each block.
- b. shows the percentage of chromosomes tested (out of 258 European chromosomes) which retained the signature sequence.
- c These percentages show the frequency of each block in chromosomes not passed on to those in the study who had Crohn disease.
- d shows the recombination rate between blocks. You can see recombination is more frequent in some inter-block regions than others.

2) What significance does this paper have for linkage disequilibrium mapping of polygenic traits? Use the genes from figure 2 as examples in your answer, but you do not need to know the function of these genes. Just consider them generic genes that could contribute to a generic disease.

- This question was not graded since very few people were able to answer it correctly. If you did answer it correctly, I gave you +2 points.
- Here is the main point.
- Rather than looking for linkage for each SNP, we should look for all the blocks in the genome. Once these are defined, we could look for linkage of blocks and the disease phenotype. This would narrow and simplify our searching for linkage disequilibrium. As shown in figure 1, looking for LD using all possible SNPs is more difficult than using blocks.

3) Which block is the least conserved of those shown in figure 2. Given the data shown, speculate why this is the case.
Block number 10 is the least conserved. Two possible reasons are: 1) is overlaps with a gene that is able to function with a greater number of SNPs than other parts of this region of our genome 2) because it is flanked by more possible recombination events (6 on the left and 5 on the right) than any other block.

4) What is the next step after this paper? Summarize what you would do if you had the computer capacity to analyze all the data in TSC database.
1) collect as many SNPs as possible
2) sample DNA from as many different ethnic groups as possible
3) define block for all people and see if there are any ethnic differences.
4) Use the block system to look for LD between blocks and disease phenotypes.

4

Once block are identified, then look for allelic/SNP variations that may contribute to the disease.

Iyer et al.
1) Cite the data in figure 3 that support the claims that SBF and MBF can
a) activate different genes.
We need to look in columns 5 v. 17+18. We want to find loci that are yellow in column 5 and blue in 17/18 or vice versa. Examples include SRL1 (top one) and PUP3 (second from top on bottom group of genes).

b) activate the same genes.
For this, we need to find loci that are yellow in columns 5, 15/16 and 17/18. This is hard to find but a good example would be the gene called CLB6 near the bottom of the middle group of loci.

2) What is the significance of the genes marked by two dots in figure 3? How do these genes affect your interpretation of the figure as a whole.
The two dots indicate loci that are similar enough in sequence that they may bind to more than one feature on the array. This is problematic because they never tell us which other loci each double-dotted locus may bind to. We note that the middle group of loci has many of these. The bottom line is with the information provided to us in the paper, it is difficult to analyze any of the data. The entire paper is cast in doubt.

3) Print figure 4 (color or B&W is fine) and then draw very clear arrows that are labeled clearly that point to only one gene:
For each of these, we need to see particular colors in particular columns. The columns are G1 and the last four columns. Below, I have only supplied to color combinations we need to see in order to demonstrate what was requested.
a) that was induced during G1 and could have bound both transcription factors (based on sequence) but only bound SBF
red, yellow, yellow, yellow, blue
b) that was induced in G1 that should not bind SBF but did and was classified as a gene activated by both transcription factors.
red, blue, any color, yellow, any color
c) that was activated in G1 that should not have bound SBF (by sequence analysis) but did bind SBF. This gene must come from the group of genes characterized as an MBF-only binding target.
red, blue, any color, yellow, any color

5

4) Draw a graph (similar to the other graphs in figure 4b) of what you would predict to see during meiosis for targets of these two transcription factors. You can use two colors of ink if this helps with clarity. It does not have to be done on a computer, you can draw it by hand.
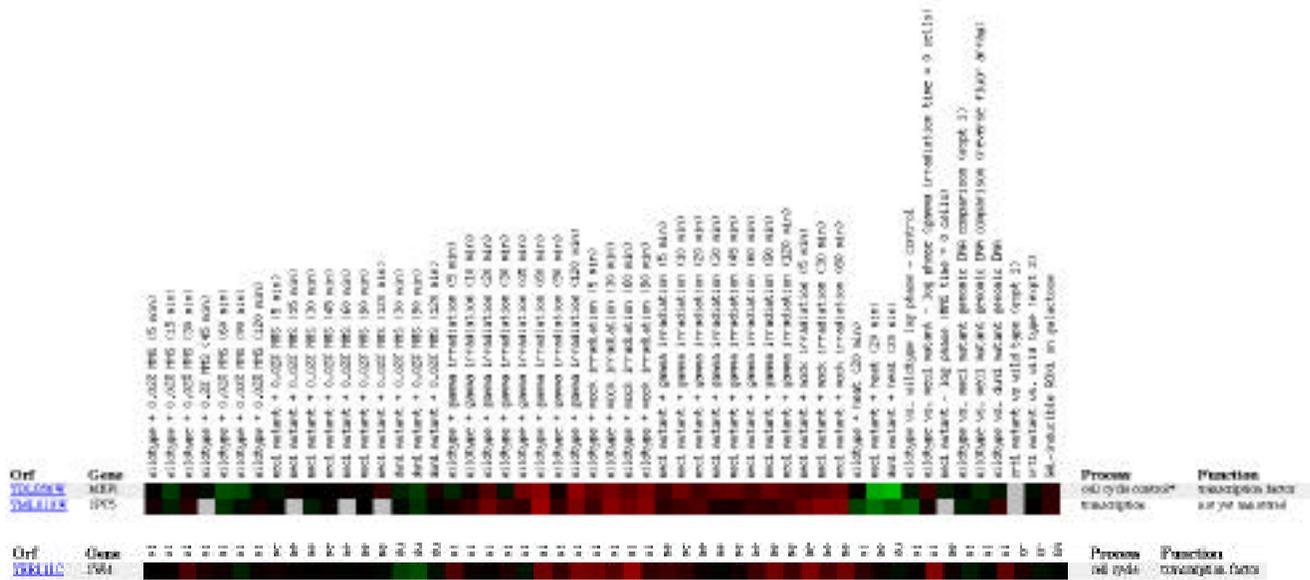
I wanted to make sure you had a peak during DNA replication for MBF. DNA replication occurs once during the cell cycle and not during meiosis. SBF should not have been peak during this time. I was just testing if you understood this output of data and read the paper to know when MBF would be up.

Answer this one.

10)

a) Use expression data online and determine which of these transcription factors/factor are/is more likely to be induced when the cells are experiencing mutagenic alterations of their genome. Support your answer with images from the database (static images are fine, no hyperlinks necessary).

MBF was induced to higher level and under more conditions than SBF. This is a judgment call and I accepted SBF if you compared the two and thought it was higher.

b)What gene or genes are coexpressed with your answer to part "a"?
 Only SPT5 is clustered with MBF4.

c) Is this an example of aneuploidy? Support your answer with data.

Probably not since Mbp1 is on chromosome 4 and SPT5 is on chromosome 13.