MAPPFinder Analysis of Prostate Cancer Microarray Data

In this exercise, we will analyze a microarray dataset from human prostate cancer tumors to see which genes were turned on or turned off in the different comparisons made during the experiment.

Procedure

1. Launch the MAPPFinder program.

2. Make sure that the Gene Database for the human species is loaded. The name of the Gene Database appears at the bottom of the window and should read "C:\GenMAPP 2 Data\Gene Databases\Hs-Std_20070817.gdb". If the right database is not showing, go to *File > Choose Gene Database* and choose the correct one. (The Gene Databases are stored in the folder C:\GenMAPP 2 Data\Gene Databases\.)

3. Click on the button "Calculate New Results".

4. Click on "Find File" and choose the "Hs_ED_ProstateCancer_20060130.gex" file and click OK. (The Expression Datasets are stored in the folder C:\GenMAPP 2 Data\Expression Datasets\.)

5. Choose a Color Set and Criterion to filter the data. [I will assign in class].

6. Check the boxes next to "Gene Ontology" and "p value".

7. Click the "Browse" button and give a filename for your results.

8. Click "Run MAPPFinder". The analysis will take about 20 minutes. The computer may look like it is stalled, be patient.

9. When the results have been calculated, a Gene Ontology browser will open showing your results. All of the Gene Ontology terms that have at least 3 genes measured and a p value of less than 0.05 will be highlighted yellow. A term with a p value less than 0.05 is considered a "significant" result. Browse through the tree to see your results.

10. To see a list of the most significant Gene Ontology terms, click on the menu item "Show Ranked List". List the top five Gene Ontology terms below:

11. One of the things you can do in MAPPFinder is to find the Gene Ontology term(s) with which a particular gene is associated. First, in the main MAPPFinder Browser window, click on the button "Collapse the Tree". Then, for example you can search for the tumor suppressor gene TP53 (one of the most famous tumor suppressor genes). Type this identifier for the gene "P04637" into the MAPPFinder browser gene ID search field. The GO term(s) that has your TP53 associated with it will be highlighted in blue. List the GO terms below:

    The Nelson et al. (2003) review article also lists some cancer susceptibility genes. Find out the GO terms with which these genes are associated. Note: you will need to look up the genes listed in the article in the UniProt database to determine their correct identifier. List the GO terms. Are there any commonalities between these terms?

12. Click on one of the GO terms that are associated with one of the cancer susceptibility genes. A MAPP will open listing all of the genes (as boxes) associated with that GO term. Moreover, the genes on the MAPP will be color-coded with the gene expression data from the microarray experiment. Red means that the gene has increased in expression as compared to the primary tumor, blue means that the gene is decreased in expression as compared to the primary tumor, grey means that there is no change in gene expression, and white means that the gene was not measured in the experiment. List below the name of the GO term you clicked on and whether the expression of TP53 changed in the experiment. (NOTE: there appears to be a bug where the GenMAPP program will stall upon opening the MAPP file. The MAPP has been created and can be manually opened in GenMAPP from C:\GenMAPP 2 Data\MAPPs\HS GO and colored with an Expression Dataset.)

13. Make a copy of your results (XXX-CriterionX-GO.txt) file. Launch Microsoft Excel. Open your .txt file in Excel (you will need to "Show all files" and click "Finish" to the wizard that will open your file). This will show you the same data that you saw in the MAPPFinder Browser, but in tabular form. You will filter this list to show the top GO terms represented in your data. You will need to filter your list down to about 20 terms. Click on a cell in the row of headers for the data. Then go to the Data menu and click "*Filter* > *Autofilter*". Drop-down arrows will appear in the row of headers. You can now choose to filter the data. Click on the drop-down arrow for the column you wish to filter and choose "(Custom…)". A window will open giving you choices on how you want to filter.

    You must set these two filters:
    Z Score (in column N) greater than 2
    PermuteP (in column O) less than 0.05

    You will use these two filters depending on the number of terms you have:
    Number Changed (in column I) greater than or equal to 4 or 5
    AND less than 100
    Percent Changed (in column L) greater than or equal to 25-50%

14. Print out your filtered list from Excel. You will need to readjust the column sizes to get all of the relevant data to print onto one page. You will want to stretch column B to see the full names of the GO terms. You also need to show the data from columns A-C and I-P. You do not need to show the other columns so they can be deleted or shrunk so that everything fits on one page.

15. Are any of your filtered GO terms closely related to one another, meaning are they a direct child or parent to another term in the list? Return to the MAPPFinder Browser window to find out and state the relationship between terms below:

16. Look up the definitions for any GO terms that are unfamiliar to you. The "official" definitions for GO terms can be found at http://www.geneontology.org.

Assignment to turn in:

1. Turn in a print-out of your filtered list of GO terms in Excel.

2. Write a description of your results and how they might relate to the disease processes in prostate cancer for the data comparison you were making (about 1 page).

   - State the data comparison you made:
     primary tumor vs treated primary tumor
     primary tumor vs metastatic tumor
     primary tumor vs androgen independent tumor
     and whether you were looking at an increase or decrease in gene expression.

   - State the filters you performed (Z score, P value, Number changed, Percent changed)

   - do the Gene Ontology categories listed in your results make sense for the data comparison you made? Why or why not? What could be a biological explanation for your results?