

Analysis of large gene expression datasets using bioinformatics

Russell Johnson

week #11 (29 April)

Techniques (such as microarray hybridization) that can measure the expression of many genes at a time (sometimes all the genes in organism) generate huge amounts of data. In order to properly understand and make use of this huge explosion in available data, we need appropriate tools for observing and analyzing the results of large-scale gene expression experiments.

Two commonly used methods for displaying gene expression data are scatter plots and e-northern. A scatter plot is very handy for looking at the expression of a LARGE number of genes, and comparing them under just TWO different conditions. An e-northern can be used for looking at the expression of a smaller group of genes under many different conditions. We will talk about how these work at the beginning of lab.

The HuGE index

This index is not only HUGE in the sense that it is very big, but it is also a great place to get information on Human Gene Expression. Today you will use information from the HuGE index to generate scatter plots. These plots will provide very helpful information about how specific genes are expressed in different human cells.

1. First, you will need to go to the HuGE Index web site (www.hugeindex.org). From this “Welcome” webpage, click on the blue circle that says “Databases”. From there, click on the “To HuGE Index Database” link. From this page, click on the “Scatterplot” link. This will take you to the main Scatterplot page.
2. On the left side choose Organ Average and brain. On the right side choose Organ Average and colon. Click on the Scatterplot button. If all goes well, a scatterplot will be produced for you and will open up in a new tab. By touching any one of the individual dots, you can get information about what gene it represents and numerical values for the amount of gene expression in both brain and colon tissue. *Print out a copy of the scatterplot for your notebook. Find 5-6 genes that are highly expressed in both brain and colon. Label the dots corresponding to these genes on your scatterplot. Have you heard of any of these genes before? Are you surprised that they are highly expressed in both tissues. Find 2-3 genes that are much more highly expressed in brain than in colon. Label the dots corresponding to these genes on your scatterplot. Find 2-3 genes that are much more highly expressed in colon than in brain. Label the dots corresponding to these genes on your scatterplot.*
3. Go back to the main scatterplot page and create a scatterplot using two different organs (not brain or colon). *Do the same things that you did with the brain:colon scatterplot.*

The Botany Array Resource for Arabidopsis Functional Genomics

The BAR provides a wealth of information on gene expression in Arabidopsis. Today you will use BAR to generate e-northern. These will show in a nice graphical form how genes are regulated in response to development or environmental stimuli.

1. First, you will need to go to the BAR web site (www.bar.utoronto.ca). In this initial webpage, you can see a list (on the left) of all the wonderful things that you can do at the BAR site. Touching each of the items on the list (with the mouse!) will give you a sample picture of the kind of output you would get from that particular type of analysis. Click on the “e-Northern with Expression Browser” link. This will take you to the main e-northern page.
2. In the dataset box, choose the AtGenExpress – Stress Series dataset. In the research areas box, select Cold stress, Osmotic Stress, Salt Stress, and Drought Stress. In the tissue types box, choose Shoot. Select all of the time points.
3. In the AGI IDs box, you will need to enter the AGI (Arabidopsis Genome Initiative) identification (or “locus tag”) for each of the genes whose expression you want to analyze. You can enter up to 125 genes. Listed below are 10 genes that you should use for this analysis:

<u>Name</u>	<u>locus tag</u>
<i>ABI2</i>	At5g57050
<i>ABI3</i>	At3g24650
<i>ABI4</i>	At2g40220
<i>ABI5</i>	At2g36270
<i>ABF1</i>	At1g49720
<i>ABF2</i>	?
<i>ABF3</i>	?
<i>ABF4</i>	?
<i>LCR70</i>	?
<i>RD29B</i>	?

For the five genes whose locus tag is not listed here, you will have to be creative and think of some way to find the appropriate locus tags.

4. Once you have entered all the necessary information, click on the Submit button at the bottom of the page. Now click on View Graphical Representation of Cluster Data. A new window will now open up showing so much information it will make your head spin.
5. The e-northern itself will be in the middle of the display. The e-northern should consist of 500 tiny rectangles (10 genes x 50 treatments), each representing the mRNA abundance for a certain gene under a certain condition. Pale yellow rectangles indicate the basal level of gene expression, while red rectangles represent increased gene expression. Above the e-northern, you are given information about each of the 50 different treatments, such as what kind of stress was given, how long after the treatment the sample was collected etc. At the right of each row is information about the gene being analyzed in that row. To the right of the “functional class” information, you will see that a “tree” has been constructed. In this tree, the genes are classified based on their overall pattern of expression (in this case, under stress conditions). Genes with

very similar expression patterns are grouped together, while those with very different patterns are placed farther apart in the “taxonomy”.

6. *Print out a copy of your e-northern together with the accompanying information (using the color printer in Olin 236 [137.146.166.37]) and include it in your notebook. Use the information to answer the following questions:*

a) *Which of the genes you analyzed is most highly induced under cold stress?*

b) *Which one is most highly induced under drought stress?*

c) *ABI2 is induced by both cold stress and salt stress? How does the pattern of ABI2 induction differ under these two different stresses?*

d) *What gene has the pattern of expression that is most similar to that of ABI5?*

e) *If a gene encoded a transcription factor, it would be grouped in the “DNA or RNA binding” classification. Which of the 10 genes you analyzed belong to this group?*

7. Carry out second e-northern “experiment”, testing the expression of genes under some different conditions. You could use the same genes you have already used above, or you may choose completely different ones. It might be useful to investigate the expression of Arabidopsis gene(s) that you have experience with or some interest in. You can also generate a list of many Arabidopsis gene “locus tags” using the “Random ID List Generator” at the main BAR website.

8. *Print out a copy of your second e-northern together with the accompanying information (using the color printer in Olin 236 [137.146.166.37]) and include it in your notebook. What interesting information can you obtain from this second e-northern?*