Written Script for Demonstrating webbioc

Colin A. Smith

April 30, 2008

Introduction

This document is written for both new users and presenters to help provide a scripted walk through of the Bioconductor web interface. Because `webbioc` doesn't use R as the interface, this document will not use any R code and will instead give written queues for what to do in the web interface. For lack of straightforward and public sample data, this document relies on users to provide their own CEL files.

1 Uploading CEL Files

1. Go to the Upload Manager and start a new session by clicking the `Start New Session` button.

2. Note the session token at the bottom of the page. It is very important to record that token somewhere on the local computer so that you can return to the files you upload. To make things easier, click the `Save Cookie` button. That will store the upload manager token in the browser for a week, making many of the tools easier to use.

3. To upload files, click the top-left button to bring up a file selection dialog box. Choose one of the CEL files you want to process. Finally, click the upload file button.

4. Upload the rest of the CEL files that you want to process. To make the other tools easier to use, upload the files in the order you would like them listed. (The Upload Manager always lists files in the order in which they were created/uploaded.) Always upload files containing control samples before those containing experimental samples. Again, this helps later on.

1

2 Preprocessing Affymetrix Data

1. There are two ways to start preprocessing Affymetrix data:

   - Go to the affy page of the Bioconductor web interface. Enter the Upload Manager token and number of CEL files and click the `Next Step` button.
   - From the Upload Manager, check all the CEL files you want to process. Click the `affy` button at the bottom of the page.

2. If you want to change the order of the CEL files, use the popup menus in the File column. You may also specify a different sample names if the filenames are not descriptive enough or are too long.

3. For demonstration purposes, leave the processing method set to RMA. It is the fastest method and produces good expression measures. The custom methods can take from about 10 minutes to almost a day. Leave the other checkboxes alone as well.

4. Click the `Submit Job` button to start the processing. Your browser will be taken to a page with the job summary. It will automatically refresh until your job finishes. A fast computer with a reasonable number of CEL files shouldn't take any more than a couple of minutes.

3 Differential Expression and Multiple Testing

1. After the preprocessing finishes, click the `multtest` link and then click the `Next Step` button. (The upload manager token should already be filled in for you.)

2. Select the ExpressionSet that you just created which will be at the end of the list. Unless there are more than two types of experimental samples (usually control and experimental), leave the number of experimental classes as 2. Click the `Next Step` button.

3. As long as you followed previous recommendations and there are an equal number of control and experimental chips, the web interface should fill out the class label correctly for you. Otherwise, give all the control samples a label of 0 and all the experimental samples a label of 1. You may also chose to ignore any samples you wish.

4. Select the statistical test you want to use to detect differential expression. The first two t-tests tend to work well but the choice is yours.

5. Select the multiple testing procedure you want to use. The FDR methods are fairly effective when looking at the whole chip at once.

6. Leave every other setting at its default value. However, click the checkbox to copy the aafTable back to the upload manager. We will use that in the next step. Click `Submit Job` to start processing.

7. Once processing completes, look at the HTML file to see the top 100 most significantly differentially expressed genes. Also briefly look at the various plots produced.

   The M vs. A plot shows log fold change (M) vs. log overall expression (A). The red points show the same 100 genes in the HTML file. The Normal Q-Q plot compares the distribution of the t-statistics to the normal distribution. The wings on either end show deviation away from the normal distribution. The last plot shows how selective the multiple testing procedure is over a range of error rates.

4   Annotation with Biological Metadata

1. After you have finished looking at the diagnostic plots, click the `annaffy` link. Scroll towards the bottom of the page and click the `Load File for Annotation` button.

2. Select the aafTable that you just created which will be at the end of the list. Click the `Next Step` button.

3. First select the type of chip that was used in the microarray experiment.

4. Next select the aafTable and Data columns that you would like to include in the annotated table. Mac users should use the Command key to select multiple items. PC users should use the Control key. Rich data sources with good linked information include Probe, Description, LocusLink, Cytoband, PubMed, Gene Ontology, and Pathway.

5. Click the `Submit Job` and wait for the annotation to complete. If no annotation appears in the resulting HTML file, go back and check to make sure that you selected the correct chip. From the resulting table, try clicking on some of the links to see what sort of information they reveal. (Probe links require a free registration with Affymetrix.)

5   Searching Biological Medadata

1. Click the `annaffy search` link. First select the same chip that you were using before. Then select `Description` for the metadata type.

2. If you are a presenter, solicit the audience for a class of genes that you would like to search for. Try to select a word that you expect will appear somewhat frequently in the gene descriptions. Enter that single word in the text box.

3. Click the `Search Text` button to begin the search. Text searches are not completely optimized and may take longer than you expect. I fast computer will finish the text search in about 30 seconds.

4. Once the search completes, you can go to the HTML file to see the descriptions that were found. The text file contains a comma delimited list of just the probe ids that correspond to found genes.

5. If you wish you may copy and paste that list back into the `Test only these gene names` box in `multtest` and thus limit the scope of multiple testing.