SBEAMS-Microarray
Demo Tutorial

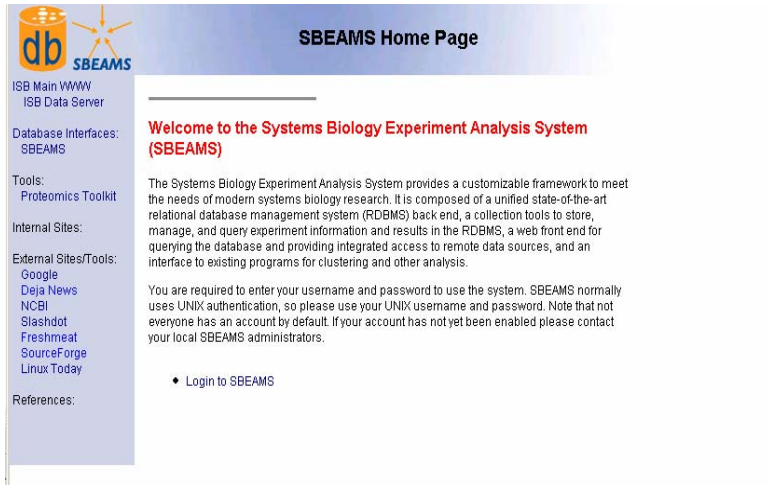
Bruz Marzolf

April 12, 2006

INSTITUTE FOR
Systems
Biology
Revolutionizing science. Enhancing life.

Getting Into *SBEAMS*

- Login:
  - Open the following page in your web browser: http://www.sbeams.org/sbeams
  - Click on the link **Login to SBEAMS**
  - There are 10 demo accounts, named ‘demo01’ through ‘demo10’. Enter one of these as your username.
  - Enter ‘sbeamsdemo’ in the password field.
- Go to Microarray Module
  - On the left menu bar click on **Microarray** to enter *SBEAMS-Microarray*

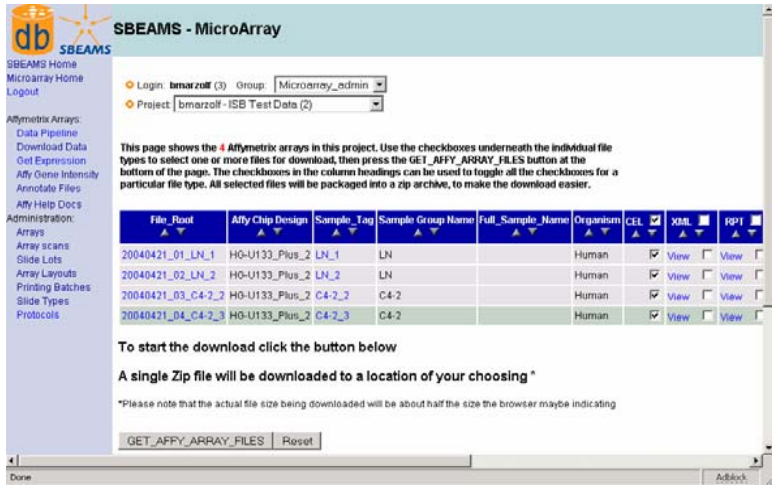


Basic Data Access

- The following three interfaces allow rapid access to data
  - **Data Download** provides a means to see QC information and download the raw CEL files for analysis with other software
  - **Simple Query** allows searching for a particular gene or genes with graphic visualization of the output
  - **Advanced Query** provides more search criteria and outputs in tabular format that can be exported and further analyzed.

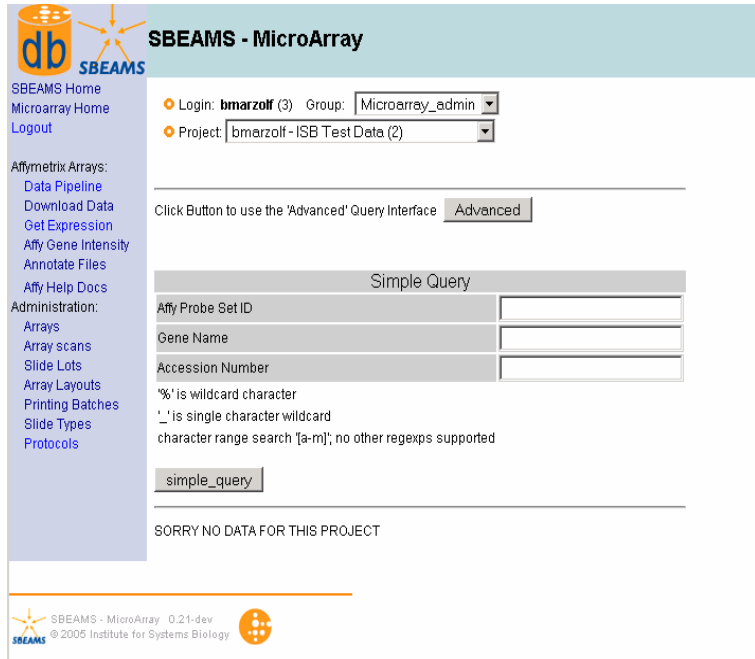
Download Array Data

- To download or view array files from a particular project choose the **Download Data** link in the left menu.
  1) Select the project ‘bmarzolf – ISB Test Data’ from the drop down list located at the top of the page.
  2) From one of the rows, click the link to **View** under the RPT column. This displays the GCOS report file on screen. Go back one page in your browser to return to the **Data Download** page.
  3) Download all the XML files by first unchecking the checkbox at the top of the CEL column, and then checking the checkbox at the top of the XML column. Finally, with all the files selected, click the button **Get_affy_array_files**
  4) Select a location to save the zip file containing all the information selected for downloading.



Simple Query

- Choose the **Affy Gene Intensity** link on the left menu to access the Simple Query page
- If ‘bmarzolf – ISB Test Data’ is not your current project, select it from the Project drop-down selector at the top of the page.
- Enter ‘pcdhb%’ in the **Gene Name** field
- Click **Simple Query**
- The results will display as a table of grayscale squares representing signal, with borders indicating detection call



Advanced Query

- Choose the **Affy Gene Intensity** link on the left menu to access the Simple Query page, then click the **Advanced** button at the top of the page
- Select the following criteria:
  - Make sure the ‘bmarzolf – ISB Test Data’ is selected.
  - Choose all of the arrays in **Affy Array Constraints**
  - In the **Signal** field, enter ‘> 25000’
  - In **Data Columns to Display**, choose ‘Signal Intensity’ and ‘Detection Call’
  - In **Display Options**, choose ‘Show all data if one condition meets criteria’ and ‘Pivot Array Samples as columns’
  - Click **Query**
  - Results will display in tabular format. Links at the bottom of the page allow you to save results out into Excel , XML, tab-separated or comma-separated formats.

Analysis Pipeline

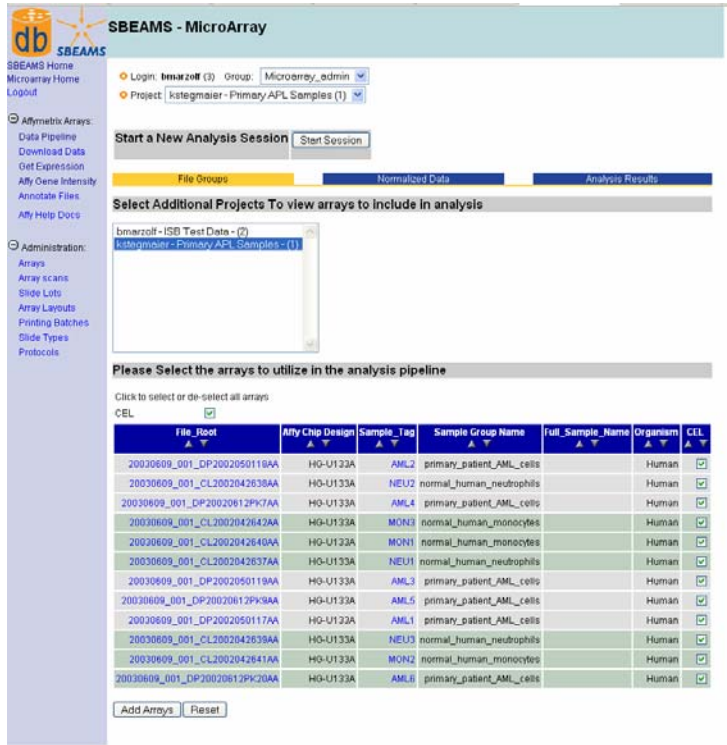
- There are three steps in the analysis pipeline in ***SBEAMS-Microarray***
  - **Selecting and Grouping Arrays**
  - **Normalization**
  - **Differential Expression Statistics**

Starting a New Analysis Session

- Switch to the ‘kstegmaier – Primary APL Samples’ project using the drop-down list at the top of the page.
- Choose the **Data Pipeline** link on the left menu bar
- Click the **Start Session** button

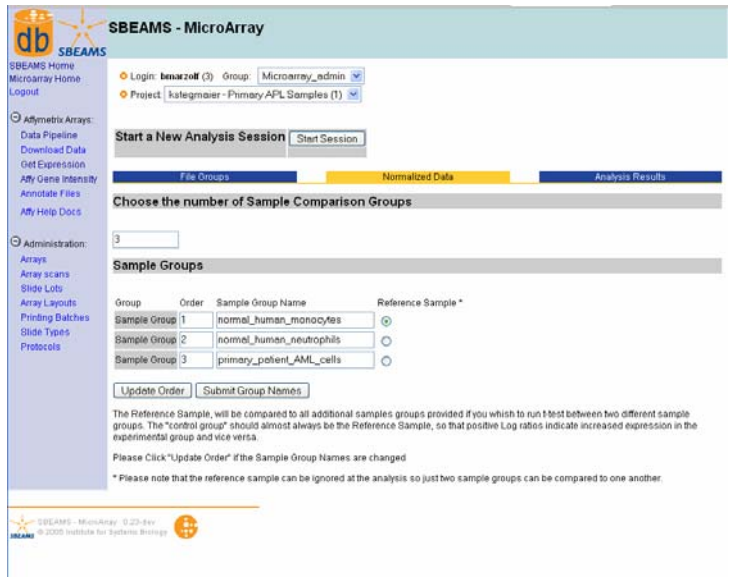
Selecting Arrays

- By default, all arrays in the project are checked. Uncheck the arrays with sample tags AML4, AML5 and AML6.
- Click the **Add Arrays** button at the bottom of the screen. The selected arrays will now appear at the top of the screen. There should be 9 arrays, 3 of each sample group (AML, MON, NEU)
- When shown your Current File Listing, click the **Continue File Grouping** button



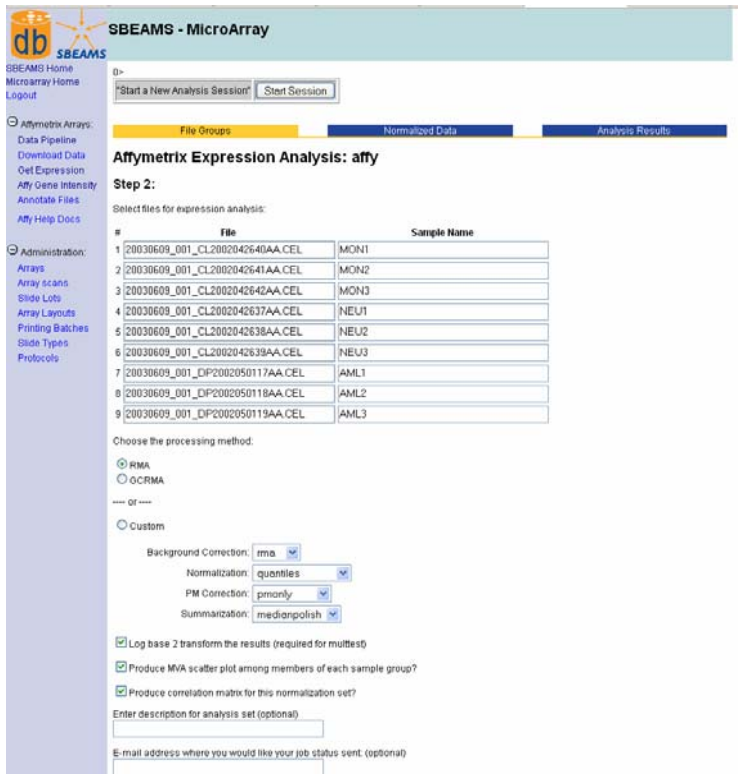
Choosing Sample Grouping

- The next page displayed allows sample groups to be renamed as desired, and choose a reference sample. The reference sample will be used with differential expression statistics later in the pipeline. Leave the default names and reference sample as they are.
- Click **Submit Group Names**
- On the next page that appears, make sure all the files are in the correct sample groups
- Click **Complete File Grouping**



Choosing Normalization Options

- Select GC-RMA as the processing method
- Leave Log base 2 selected, so that results are output in log2 scale
- Leave the checkboxes for MVA plots and correlation matrices selected. These are useful as diagnostic plots. (These require substantially more memory during analysis, and should be turned off to facilitate normalization of large numbers of arrays.)
- Name the analysis with a name of your choice
- Optionally enter an email address to be sent a message upon completion of normalization run.
- Click **Start Normalization**



Viewing Normalization Results

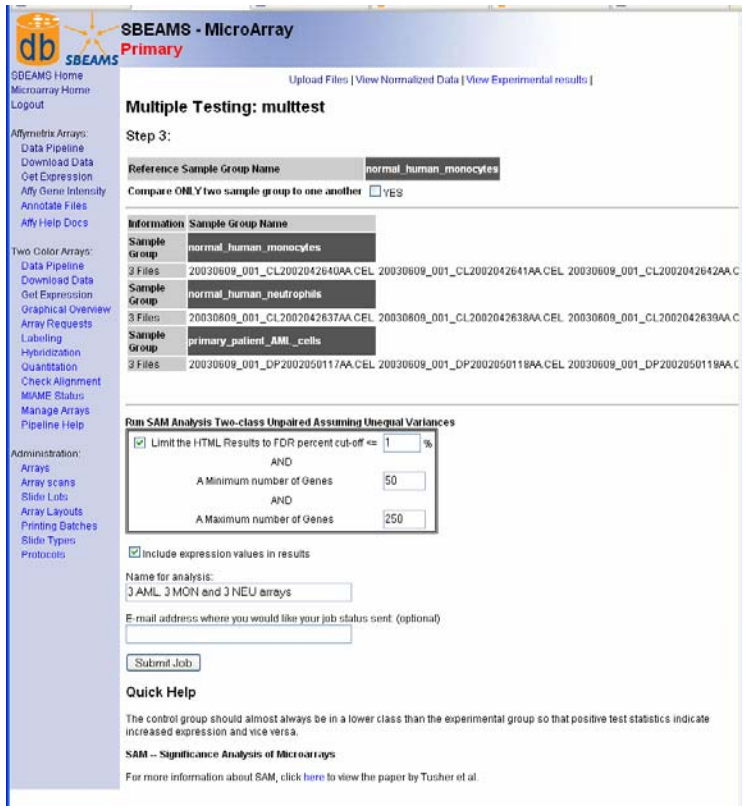
- Either return to the browser window where you performed the analysis, or click on the result link in the email sent to you by the pipeline.
  - Click on the link **Show Files** located at the top of the page
- View the normalized data file
  - Under the header Analysis Run Files→Data click the link **Show**
  - This load log2 expression values for all the genes and arrays just analyzed, although it's not very useful in the browser window.
- Download the normalized data
  - Under the header Analysis Run Files→Data click the link **Get**. These can be loaded into Excel or other spreadsheet software.

Returning to Normalization Results

- Make sure the ‘kstegmaier – Primary APL Samples’ project is selected
- Click **Data Pipeline** link on the left navigation menu
- Click the **Normalized Data** tab
- Click the **Show Files** Link for your normalization run

Performing a SAM Run

- Find the heading ‘Start Additional Analysis->Multiple t-test’
  - Click the link **Start Multtest**
- Leave the radio button selection at “SAM”
- Click the button **Next Step**
- When the run starts the reference sample will be compared to the other experimental sample groups.
- A single pairwise comparison out of a set of many groups may be performed by selecting the checkbox labeled ‘Compare ONLY two sample group to one another’
- Set SAM options for a FDR cutoff of 1%, minimum genes of 50 and maximum genes 250.
- Name your analysis with a title of your choice.
- Optionally add your e-mail address to be notified after the run is complete.
- Click **Submit Job**



Adding Data to GetExpression

- Once the SAM run from the previous step has completed, click the **Show Files** link under the ‘ Show analysis Data’ heading
- In your SAM or t-test result page, under Add Results to Get Expression→Add Data, choose **Add Data Link**
- Click **Check Condition Names** to see whether this condition already exists. If the name is unique proceed to the next step, otherwise go back to the previous step and create a new name
- Click the button **Upload Conditions**. Wait for the data to load.
- Click the link ‘Go to Get Expression Page’ **here** at the bottom of the page OR choose the **GetExpression** link on the left menu bar

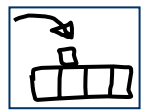
Query Data using GetExpression

- Select the ‘kstegmaier – Primary APL Samples’ project
- Within the Conditions menu box select both conditions created by SAM and uploaded in the previous steps
- In the ‘False Discovery Restraint’ field, enter “< 0.1”
- The Following columns MUST be selected in order to ensure the data can be loaded into Cytoscape:
  - Data Columns to Display:
    - Log 10 Ratio
    - False Discovery Rate
  - Display Options:
    - Show All Conditions if one condition meets criteria
    - Pivot Conditions as columns
- Click the **Query** button

Launch Cytoscape

- At the bottom of the page, click the link next to Download ResultSet in Format:→ **Cytoscape**
- When the web page come back, click the first three Links (The Boss must go first):
  - Boss
  - Network
  - Data Matrix Viewer
- Make sure all three programs are running: Gaggle Boss, Cytoscape and Data Matrix Viewer(DMV)

Load Expression Data into Cytoscape

- Load the Expression Data
  - Click on the DMV window
  - Click on the folder “Expression”
  - Click on the button to the right of “new” 
  - Within the DMV go up to the drop down menu currently labeled “None”, and select the first Condition in the list. This should make the circle (gene nodes) change color in Cytoscape
  - To play a, movie click the icon 

Selecting Genes in Cytoscape

- Manually Select the genes that both conditions share in common in Cytoscape
  - Go to the Cytoscape window
  - Click and drag to highlight the gene nodes. Use Shift if all of the genes cannot be selected at one time
  - Right click to bring up the ‘Node Browser’
  - Click the ‘Gene_name’ tab to see the names of the selected genes
  - Click **Dismiss** to close the ‘Node Browser’ window

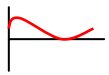
Using Filters in Cytoscape

- Use a filter to select all the genes that both conditions have in common
  - Click somewhere on the Cytoscape display away from any nodes to de-select the currently selected nodes
  - Click the **Grid** icon located at the top-middle of the Cytoscape menu bar
  - Find the tab **Topology 1**
  - Enter 2 in the field “Minimum Number of Neighbors”
  - Enter 1 in the field “Within Depth”
  - Click **Select**. This should have selected the nodes in common between the two conditions.
  - Click **Dismiss** to close the ‘Filters’ window

Graph Genes in the DMV

- Graph the Selected genes in the DMV to view the global expression profile
  - Go to the DMV, and click the ‘log 10 ratios’ tab. Select one gene in the list, then click the **Clear** button.
  - Go to the Cytoscape windows, where the genes of interest should still be highlighted. in Click the **Broadcast** buttonin the upper-right corner to broadcast these selected genes to the DMV.
  - Go to the DMV window.
  - It should indicate that 8 genes are selected.
  - Click the icon  This presents a graph of ratios between the two conditions

Find Similar Expression Profiles in the DMV

- Find Genes that have a similar expression profile
  - Make sure no genes are currently selected. Go to the 'log 10 ratios' tab and click the **Clear** button
  - Go back to the 'log 10 ratios (plot)' tab made in the previous step, click a gene name with an interesting profile.
  - Click the Broadcast icon located in the upper left hand part of the screen
  - Go back to the 'log 10 ratios' tab.
  - Click the icon 
  - Move the slider so the Threshold reads 99. Click **Select in Browser**, then **Dismiss**
  - Graph the data by clicking the icon 
  - To highlight the genes in Cytoscape click the Broadcast button, after first de-selecting any selected nodes in Cytoscape. Note that all the genes selected in the DMV my not be viewable in Cytoscape since the expression network is a small subset of all the expression data loaded from the GetExpression query.

Apply GO Annotation in Cytoscape

- Apply GO annotation
  - Click the ‘add annotation to nodes’ icon top-middle of the Cytoscape menu 
  - Open **Go, Molecular Function, Homo sapiens**
  - Click on level “4”
  - Click **Apply Annotation to All Nodes**
  - Open the folder on the right half of the screen and walk through the different GO levels. Look at the Cytoscape screen to see what becomes highlighted.

Summary

- SBEAMS Login
- Basic Data Access
  - Data Download
  - Simple Query
  - Advanced Query
- Analysis Pipeline
  - File Grouping and Normalization
  - Differential Expression Testing
  - Querying GetExpression
  - Cytoscape