

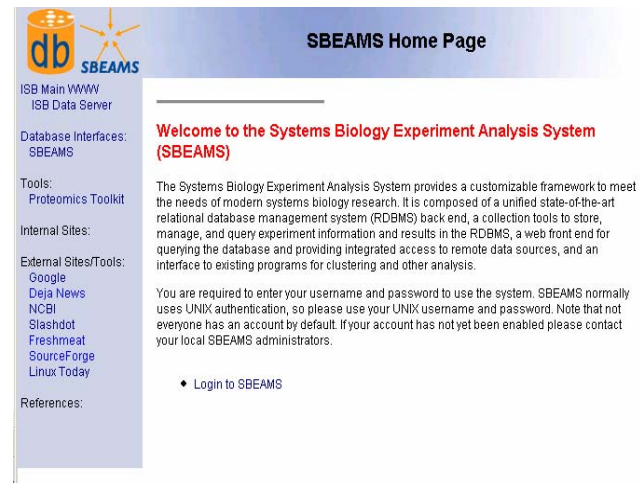
SBEAMS-Microarray **Demo Tutorial**

Bruz Marzolf

April 12, 2006

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*



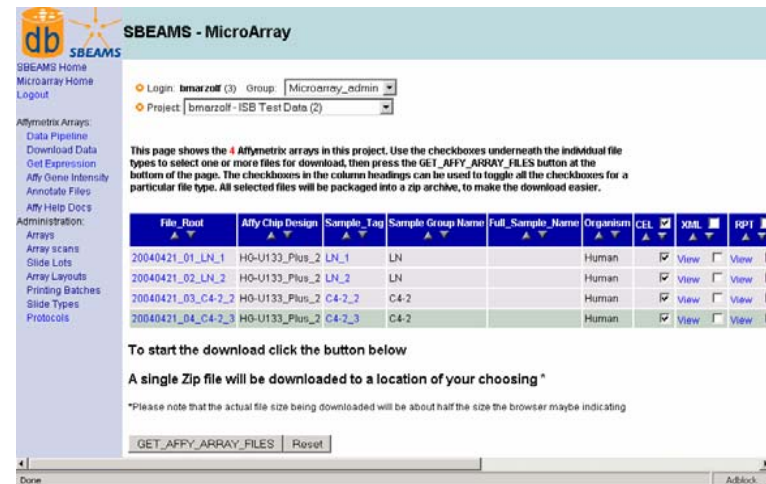
The screenshot shows the SBEAMS Home Page. On the left is a navigation menu with categories: ISB Main WWW, ISB Data Server, Database Interfaces: SBEAMS, Tools: Proteomics Toolkit, Internal Sites, External Sites/Tools (listing Google, Deja News, NCBI, Slashdot, Freshmeat, SourceForge, Linux Today), and References. The main content area has the title 'SBEAMS Home Page' and a welcome message: 'Welcome to the Systems Biology Experiment Analysis System (SBEAMS)'. Below this is a paragraph describing the system's architecture and a note about authentication requirements. A 'Login to SBEAMS' link is visible at the bottom of the main content area.

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.



The screenshot shows the SBEAMS - MicroArray web interface. The page title is "SBEAMS - MicroArray". The user is logged in as "bmarzolf (3)" with the group "Microarray_admin". The project selected is "bmarzolf - ISB Test Data (2)".

The page displays 4 Affymetrix arrays. The table below shows the details of these arrays:

File_Root	Affy Chip Design	Sample_Tag	Sample Group Name	Full_Sample_Name	Organism	CEL	XML	RPT
20040421_01_LN_1	HO-U133_Plus_2	LN_1	LN		Human	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20040421_02_LN_2	HO-U133_Plus_2	LN_2	LN		Human	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20040421_03_C4-2_2	HO-U133_Plus_2	C4-2_2	C4-2		Human	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20040421_04_C4-2_3	HO-U133_Plus_2	C4-2_3	C4-2		Human	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

To start the download click the button below

A single Zip file will be downloaded to a location of your choosing *

*Please note that the actual file size being downloaded will be about half the size the browser maybe indicating

GET_AFFY_ARRAY_FILES | Reset

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

SBEAMS - MicroArray

SBEAMS Home
Microarray Home
Logout

Login: **bmarzolf** (3) Group:
Project:

Affymetrix Arrays:
[Data Pipeline](#)
[Download Data](#)
[Get Expression](#)
[Affy Gene Intensity](#)
[Annotate Files](#)
[Affy Help Docs](#)

Administration:
[Arrays](#)
[Array scans](#)
[Slide Lots](#)
[Array Layouts](#)
[Printing Batches](#)
[Slide Types](#)
[Protocols](#)

Click Button to use the 'Advanced' Query Interface

Simple Query

Affy Probe Set ID	<input type="text"/>
Gene Name	<input type="text" value="pcdhb%"/>
Accession Number	<input type="text"/>

% is wildcard character
_ is single character wildcard
character range search [a-m], no other regexps supported

SORRY NO DATA FOR THIS PROJECT

SBEAMS - MicroArray 0.21-dev
© 2005 Institute for Systems Biology

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

SBEAMS - MicroArray

SBEAMS Home
Microarray Home
Logout

Login: bmarzof (3) Group: Microarray_admin
Project: kistegmaier - Primary APL Samples (1)

Start a New Analysis Session | Start Session

File Groups | Normalized Data | Analysis Results

Select Additional Projects To view arrays to include in analysis

bmarzof - ISB Test Data - (7)
kistegmaier - Primary APL Samples - (1)

Please Select the arrays to utilize in the analysis pipeline

Click to select or de-select all arrays

CEL

File_Root	Affy Chip Design	Sample_Tag	Sample Group Name	Full_Sample_Name	Organism	CEL
20030609_001_DP2002050119AA	HG-U133A	AML2	primary_patient_AML_cells		Human	<input checked="" type="checkbox"/>
20030609_001_CL2002042638AA	HG-U133A	NEU2	normal_human_neutrophils		Human	<input checked="" type="checkbox"/>
20030609_001_DP200206129K7AA	HG-U133A	AML4	primary_patient_AML_cells		Human	<input checked="" type="checkbox"/>
20030609_001_CL2002042642AA	HG-U133A	MON3	normal_human_monocytes		Human	<input checked="" type="checkbox"/>
20030609_001_CL2002042640AA	HG-U133A	MON1	normal_human_monocytes		Human	<input checked="" type="checkbox"/>
20030609_001_CL2002042637AA	HG-U133A	NEU1	normal_human_neutrophils		Human	<input checked="" type="checkbox"/>
20030609_001_DP2002050119AA	HG-U133A	AML3	primary_patient_AML_cells		Human	<input checked="" type="checkbox"/>
20030609_001_DP200206129K8AA	HG-U133A	AML5	primary_patient_AML_cells		Human	<input checked="" type="checkbox"/>
20030609_001_DP2002050117AA	HG-U133A	AML1	primary_patient_AML_cells		Human	<input checked="" type="checkbox"/>
20030609_001_CL2002042639AA	HG-U133A	NEU3	normal_human_neutrophils		Human	<input checked="" type="checkbox"/>
20030609_001_CL2002042641AA	HG-U133A	MON2	normal_human_monocytes		Human	<input checked="" type="checkbox"/>
20030609_001_DP200206129K20AA	HG-U133A	AML6	primary_patient_AML_cells		Human	<input checked="" type="checkbox"/>

Add Arrays | Reset

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**

The screenshot shows the SBEAMS - MicroArray web interface. The page title is "SBEAMS - MicroArray". The user is logged in as "amazonf (3)" with the group "Microarray_admin". The project is "Kistegmaier-Primary/APL Samples (1)".

There are two main sections: "Start a New Analysis Session" and "Choose the number of Sample Comparison Groups".

The "Start a New Analysis Session" section has a "Start Session" button.

The "Choose the number of Sample Comparison Groups" section has a text input field containing the number "3".

Below this is the "Sample Groups" section, which contains a table with the following data:

Group	Order	Sample Group Name	Reference Sample *
Sample Group 1	1	normal_human_monocytes	<input checked="" type="radio"/>
Sample Group 2	2	normal_human_neutrophils	<input type="radio"/>
Sample Group 3	3	primary_patient_AML_cells	<input type="radio"/>

Below the table are two buttons: "Update Order" and "Submit Group Names".

Below the buttons is a paragraph of text: "The Reference Sample, will be compared to all additional samples groups provided if you wish to run t-test between two different sample groups. The 'control group' should almost always be the Reference Sample, so that positive Log ratios indicate increased expression in the experimental group and vice versa."

Below the paragraph is another paragraph: "Please Click 'Update Order' if the Sample Group Names are changed"

Below the paragraph is a footnote: "* Please note that the reference sample can be ignored at the analysis so just two sample groups can be compared to one another."

At the bottom of the page, there is a footer: "SBEAMS - MicroArray 0.23-Rev" and "© 2005 Institute for Systems Biology".

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**

The screenshot shows the SBEAMS - MicroArray web interface. The page title is "SBEAMS - MicroArray". The main content area is titled "Affymetrix Expression Analysis: affy" and is in "Step 2". The user is prompted to "Select files for expression analysis:". A table lists 9 files with their corresponding sample names:

#	File	Sample Name
1	20030609_001_CL2002042640AA.CEL	MON1
2	20030609_001_CL2002042641AA.CEL	MON2
3	20030609_001_CL2002042642AA.CEL	MON3
4	20030609_001_CL2002042637AA.CEL	NEU1
5	20030609_001_CL2002042638AA.CEL	NEU2
6	20030609_001_CL2002042639AA.CEL	NEU3
7	20030609_001_DP2002050117AA.CEL	AML1
8	20030609_001_DP2002050118AA.CEL	AML2
9	20030609_001_DP2002050119AA.CEL	AML3

Below the table, the user is asked to "Choose the processing method:". The "RMA" option is selected. The "Custom" option is also available. The "Background Correction" is set to "rma", "Normalization" is "quantiles", "PM Correction" is "pmonly", and "Summarization" is "medianpolrsh". There are three checkboxes: "Log base 2 transform the results (required for multitest)", "Produce MVA scatter plot among members of each sample group?", and "Produce correlation matrix for this normalization set?". The first two are checked. There is a text input field for "Enter description for analysis set (optional)" and another for "E-mail address where you would like your job status sent (optional)".

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.

The screenshot displays the SBEAMS - MicroArray Primary web interface. The top navigation bar includes 'File Groups', 'Normalized Data', and 'Analysis Results'. The main content area is titled 'Analysis Run Info' and contains the following details:

- Edit Data:** [Edit Analysis Description](#)
- Parent Analysis Data:** [Edit Parent Analysis Description](#)
- Delete Analysis Run:** [Delete Analysis Run](#)
- User Description:** 3 AML, 3 MON and 3 NEU arrays
- File Names:** 20030809_001_CL2002042640AA.CEL, 20030809_001_CL2002042641AA.CEL, 20030809_001_CL2002042642AA.CEL, 20030809_001_CL2002042643AA.CEL, 20030809_001_CL2002042644AA.CEL, 20030809_001_CL2002042645AA.CEL, 20030809_001_DP2002050117AA.CEL, 20030809_001_DP2002050118AA.CEL, 20030809_001_DP2002050119AA.CEL
- Description:** Yes
- Sample Names:** =>MON1, MON2, MON3, NEU1, NEU2, NEU3, AML1, AML2, AML3
- Processing:** =>RMA

Below this, the 'Start Additional Analysis' section includes options for 'Multiple Test' and 'Process file to view in Mv'.

The 'Analysis Run Files' section contains two tables:

Data	Show File	Download File	Info
	Show	Get	Annotated expression values file

R Files	Show File	Download File	Info
---	Get		Tar Gzipped Archive of Analysis
Show	Get		Completed Job -- Html File
Show	Get		Completed Job -- Html File
Show	Get		XML file showing groupings
---	Get		R Binary affy library expression file
Show	Get		R Error File
Show	Get		R Script

The footer of the page includes the SBEAMS logo, the text 'SBEAMS - MicroArray 0.21 dev', and '© 2005 Institute for Systems Biology'.

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 Multitest**
- 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**

The screenshot shows the SBEAMS MicroArray Primary interface. The main heading is "Multiple Testing: multitest". Under "Step 3:", the "Reference Sample Group Name" is set to "normal_human_monocytes". The checkbox "Compare ONLY two sample group to one another" is unchecked. Below this, there is a table of sample groups:

Information	Sample Group Name
Sample Group	normal_human_monocytes
3 Files	20030609_001_CL2002042640AA.CEL 20030609_001_CL2002042641AA.CEL 20030609_001_CL2002042642AA.C
Sample Group	normal_human_neutrophils
3 Files	20030809_001_CL2002042637AA.CEL 20030809_001_CL2002042638AA.CEL 20030809_001_CL2002042639AA.C
Sample Group	primary_patient_AMI_cells
3 Files	20030609_001_DP2002050117AA.CEL 20030609_001_DP2002050118AA.CEL 20030609_001_DP2002050119AA.C

Below the table, the "Run SAM Analysis Two-class Unpaired Assuming Unequal Variances" section is visible. It includes a checkbox "Limit the HTML Results to FDR percent cut-off <= 1 %", a text input for "A Minimum number of Genes" set to 50, and a text input for "A Maximum number of Genes" set to 250. There is also a checkbox "Include expression values in results" which is checked. The "Name for analysis:" field contains "3 AML 3 MON and 3 NEU arrays". The "E-mail address where you would like your job status sent (optional)" field is empty. A "Submit Job" button is present.

Quick Help
The control group should almost always be in a lower class than the experimental group so that positive test statistics indicate increased expression and vice versa.
SAM - Significance Analysis of Microarrays
For more information about SAM, click [here](#) to view the paper by Tusher et al.

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

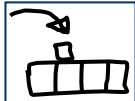

Query Data using GetExpression

- 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


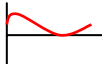
Using Filters in Cytoscape

- 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** button in 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 may 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