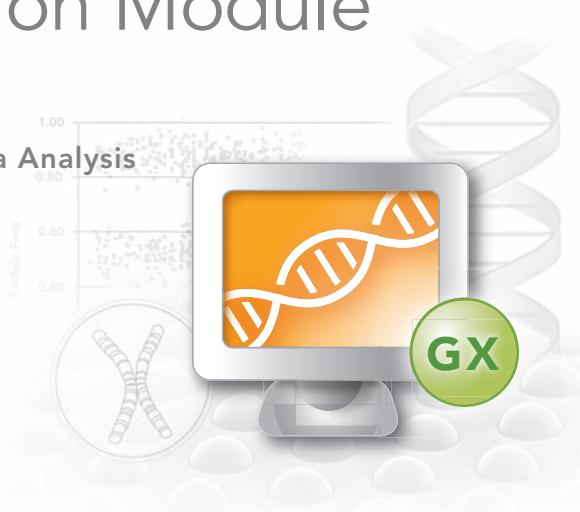


# BeadStudio Gene Expression Module User Guide

A Modular Tool for Illumina Data Analysis

FOR RESEARCH ONLY

VERSION 3





# Notice

This publication and its contents are proprietary to Illumina, Inc., and are intended solely for the contractual use of its customers and for no other purpose than to operate the system described herein. This publication and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina, Inc.

For the proper operation of this system and/or all parts thereof, the instructions in this guide must be strictly and explicitly followed by experienced personnel. All of the contents of this guide must be fully read and understood prior to operating the system or any parts thereof.

**FAILURE TO COMPLETELY READ AND FULLY UNDERSTAND AND FOLLOW ALL OF THE CONTENTS OF THIS GUIDE PRIOR TO OPERATING THIS SYSTEM, OR PARTS THEREOF, MAY RESULT IN DAMAGE TO THE EQUIPMENT, OR PARTS THEREOF, AND INJURY TO ANY PERSONS OPERATING THE SAME.**

Illumina, Inc. does not assume any liability arising out of the application or use of any products, component parts or software described herein. Illumina, Inc. further does not convey any license under its patent, trademark, copyright, or common-law rights nor the similar rights of others. Illumina, Inc. further reserves the right to make any changes in any processes, products, or parts thereof, described herein without notice. While every effort has been made to make this guide as complete and accurate as possible as of the publication date, no warranty or fitness is implied, nor does Illumina accept any liability for damages resulting from the information contained in this guide.

Illumina's BeadXpress Reader application software is subject to the licenses set forth in the applicable purchase and license agreement.

Illumina, Sentrix, Array of Arrays, BeadArray, DASL, Infinium, GoldenGate, BeadXpress, VeraCode, Oligator, IntelliHyb, iSelect, CSPro, and Making Sense Out of Life are registered trademarks or trademarks of Illumina, Inc. Other brand and product names mentioned herein may be trademarks or registered trademarks of their respective owners.

© 2007 Illumina, Inc. All rights reserved.



# Revision History

Revision	Date
Rev. B	February 2007
Rev. A	May 2006



# Table of Contents

Notice .....	iii
Revision History.....	v
Table of Contents .....	vii
List of Figures .....	xi
List of Tables.....	xv
<b>Chapter 1      Overview .....</b>	<b>1</b>
Introduction.....	2
Audience and Purpose .....	3
Installing the Gene Expression Module .....	3
Gene Expression Module Workflow .....	6
<b>Chapter 2      Creating a New Project .....</b>	<b>9</b>
Introduction.....	10
Creating a Project .....	11
Starting the Gene Expression Module .....	11
Selecting an Assay Type.....	12
Choosing a Project Location.....	13
Selecting Your Project Data .....	15
Defining a Groupset.....	18
Defining the Analysis Type and Parameters.....	24
Creating a Mask File.....	29
Applying a Group Layout File.....	30

<b>Chapter 3</b>	<b>Viewing Your Data.</b>	<b>35</b>
Introduction . . . . .	36	
Scatter Plots . . . . .	36	
Control Panel . . . . .	39	
Tools Menu . . . . .	41	
Context Menu . . . . .	43	
Finding Items . . . . .	46	
Viewing Marked Data . . . . .	54	
Viewing Marked Data in a Web Browser . . . . .	55	
Saving Marked Data in a Text File . . . . .	57	
Showing Item Labels in a Scatter Plot . . . . .	59	
Other Scatter Plot Functions . . . . .	60	
Bar Plots . . . . .	61	
Bar Plot Context Menu . . . . .	63	
Heat Maps . . . . .	64	
Heat Map Tools Menu . . . . .	66	
Heat Map Context Menu . . . . .	67	
Cluster Analysis Dendrograms . . . . .	68	
Similarities and Distances . . . . .	69	
Analyze Clusters . . . . .	70	
Dendrogram Context Menu Selections . . . . .	73	
View the Sub-Tree List Directly in the Dendrogram . . . . .	74	
Copy/Paste Clusters . . . . .	74	
From Scatter Plot to Dendrogram . . . . .	74	
From Dendrogram to Scatter Plot . . . . .	76	
Control Summary Reports . . . . .	78	
For the DirectHyb Assay . . . . .	78	
For the DASL Assay . . . . .	81	
Image Viewer . . . . .	84	
Selecting an Image to View . . . . .	85	
Displaying Overlay Cores . . . . .	88	
Changing Image Appearance . . . . .	89	
<b>Chapter 4</b>	<b>Normalization and Differential Analysis.</b>	<b>91</b>
Introduction . . . . .	92	
Normalization Methods & Algorithms . . . . .	92	
Background . . . . .	93	
Average . . . . .	93	
Rank Invariant . . . . .	93	
Cubic Spline . . . . .	94	

Differential Expression Algorithms .....	95
Illumina Custom .....	95
Mann-Whitney .....	98
T-Test .....	98
Detection P-Value .....	98
Whole Genome BeadChips .....	98
Focused Arrays .....	99
DASL Products .....	99
<b>Chapter 5 Generating a Final Report .....</b>	<b>101</b>
Introduction.....	102
Generating a Final Report.....	102
Viewing a Final Report .....	105
<b>Chapter 6 User Interface Reference .....</b>	<b>107</b>
Introduction.....	108
Detachable Docking Windows .....	108
Bar or Line Plot: Group Gene Profile .....	109
Samples Table .....	110
Group Gene Profile .....	112
Group Probe Profile .....	116
Sample Gene Profile .....	120
Sample Probe Profile .....	124
Control Gene Profile.....	128
Control Probe Profile .....	130
Control Summary .....	132
Project Window .....	134
Log Window .....	134
Main Window Menus .....	136
Context Menus .....	140
Introduction.....	144
Data Section .....	144
Sample Sheet Template .....	146
Sample Sheet Example .....	146
Introduction.....	147
Experiment Parameters.....	148
Scatter Plot .....	148
Finding Genes.....	149
Clustering Selected Genes .....	149
Filtering Selected Genes.....	150

Selected List .....	150
Select by Detection .....	151
Show R2.....	151
Dendrogram Changes.....	151
Introduction .....	153
Frequently Asked Questions .....	153

# List of Figures

Figure 1	BeadStudio Application Unzipping.....	3
Figure 2	Illumina BeadStudio Installation, Modules Tab.....	4
Figure 3	Accept License Agreement.....	5
Figure 4	Illumina BeadStudio Installation .....	5
Figure 5	Installation Progress Dialog Box.....	6
Figure 6	Gene Expression Analysis Workflow.....	7
Figure 7	Project Wizard - Welcome.....	12
Figure 8	Project Wizard - Gene Expression Assay Type .....	13
Figure 9	Project Wizard - Project Location .....	14
Figure 10	Project Wizard - Project Data Selection, Repository.....	15
Figure 11	Project Wizard - Project Data Selection, Sentrix Array Products	16
Figure 12	Project Wizard - Project Data Selection, Selected Samples ...	18
Figure 13	Project Wizard - Groupset Definition, Assigning a Groupset Name .....	19
Figure 14	Project Wizard - Groupset Definition, Selecting a Sentrix Array Product .....	20
Figure 15	Project Wizard - Groupset Definition, Selecting Samples.....	21
Figure 16	Project Wizard - Groupset Definition, Defining Project Groups	22
Figure 17	Project Wizard - Groupset Definition, New Project .....	23
Figure 18	Project Analysis Type and Parameters .....	24
Figure 19	BeadStudio Gene Expression Analysis Results .....	28
Figure 20	Group Layout File Example.....	30
Figure 21	Open Dialog Box .....	31
Figure 22	Samples Grouped Based on Group Layout File .....	32
Figure 23	Groups Expanded.....	33
Figure 24	Plot Columns Dialog Box .....	37
Figure 25	Scatter Plot .....	38
Figure 26	Scatter Plot Tools Menu .....	41
Figure 27	Scatter Plot Context Menu .....	43
Figure 28	Find Items Tool .....	46
Figure 29	Find Items Dialog .....	47
Figure 30	Zoom in to See Selected Genes.....	49

Figure 31	Gene Properties, Data Tab . . . . .	50
Figure 32	Gene Properties, Manifest Tab . . . . .	51
Figure 33	NCBI Website . . . . .	52
Figure 34	NCBI Record for the Selected Gene . . . . .	53
Figure 35	Gene Properties, Ontology Tab . . . . .	54
Figure 36	Scatter Plot Context Menu, Marked List Options . . . . .	55
Figure 37	BeadStudio Scatter Plot Output Data Dialog Box. . . . .	56
Figure 38	Marked Data Shown in a Web Browser. . . . .	57
Figure 39	Save Marked Genes List As . . . . .	57
Figure 40	BeadStudio Scatter Plot Output Data Dialog Box. . . . .	58
Figure 41	Saving Marked Data in a Text File . . . . .	59
Figure 42	Selecting a Label for a Scatter Plot . . . . .	59
Figure 43	Showing Item Labels for Marked Data in a Scatter Plot . . . . .	60
Figure 44	Bar Plot of Sample Probe Profile . . . . .	61
Figure 45	Plot Settings Dialog Box . . . . .	62
Figure 46	Bar Plot With User-Selected Attributes . . . . .	63
Figure 47	Bar Plot Context Menu . . . . .	63
Figure 48	Creating a Heat Map . . . . .	65
Figure 49	Heat Map . . . . .	66
Figure 50	Heat Map Tools Menu . . . . .	66
Figure 51	Heat Map Context Menu . . . . .	67
Figure 52	Dendrogram Similarity Example . . . . .	70
Figure 53	Dendrogram, Showing Nodes . . . . .	70
Figure 54	Cluster Analysis Dialog Box . . . . .	71
Figure 55	Dendrogram . . . . .	72
Figure 56	Dendrogram with Context Menu . . . . .	72
Figure 57	Zooming In to View a Sub-Tree List . . . . .	74
Figure 58	Selecting a Region . . . . .	75
Figure 59	Copying & Pasting . . . . .	76
Figure 60	Selecting a Sub-Tree . . . . .	77
Figure 61	Copying a Sub-Tree . . . . .	78
Figure 62	Control Summary Report . . . . .	79
Figure 63	Housekeeping Controls Secondary Graph . . . . .	80
Figure 64	Control Summary Context Menu . . . . .	81
Figure 65	Control Summary Report . . . . .	82
Figure 66	Contamination Controls Secondary Graph . . . . .	83
Figure 67	Control Summary Context Menu . . . . .	84
Figure 68	BeadStudio View Image . . . . .	85
Figure 69	Image Viewer . . . . .	86
Figure 70	Overlay Cores . . . . .	88
Figure 71	Image Control Pane . . . . .	89
Figure 72	Creating a Final Report . . . . .	102

Figure 73	Final Report Dialog Box . . . . .	103
Figure 74	Selecting Options for the Final Report . . . . .	103
Figure 75	Saving the Final Report. . . . .	104
Figure 76	Naming the Final Report. . . . .	104
Figure 77	Final Report. . . . .	105
Figure 78	Gene Expression Module Default View . . . . .	108
Figure 79	Bar Plot: Group Gene Profile . . . . .	109
Figure 80	Samples Table. . . . .	110
Figure 81	Group Gene Profile Window. . . . .	112
Figure 82	Group Probe Profile Window . . . . .	116
Figure 83	Sample Gene Profile Window. . . . .	120
Figure 84	Sample Probe Profile Window . . . . .	124
Figure 85	Control Gene Profile Window. . . . .	128
Figure 86	Control Probe Profile Window . . . . .	130
Figure 87	Control Summary Window, Direct Hyb. . . . .	132
Figure 88	Control Summary Window, DASL. . . . .	133
Figure 89	Project Window. . . . .	134
Figure 90	Log Window . . . . .	134
Figure 91	Sample Sheet Example. . . . .	146



# List of Tables

Table 1	Scatter Plot Control Panel Functions & Descriptions .....	39
Table 2	Scatter Plot Tools Menu Item Descriptions.....	41
Table 3	Scatter Plot Context Menu Item Descriptions.....	43
Table 4	Bar Plot Context Menu Item Descriptions.....	64
Table 5	Heat Map Tools Menu Item Descriptions.....	67
Table 6	Heat Map Context Menu Item Descriptions.....	68
Table 7	Dendrogram Context Menu Selections .....	73
Table 8	Image Viewer Features .....	86
Table 9	Samples Table Columns .....	110
Table 10	Group Gene Profile Window Columns .....	113
Table 11	Group Gene Profile Window Per-Group Columns .....	115
Table 12	Group Probe Profile Window Columns.....	117
Table 13	Group Probe Profile Window Per-Group Columns.....	119
Table 14	Sample Gene Profile Window Columns .....	121
Table 15	Sample Gene Profile Window Per-Sample Columns .....	123
Table 16	Sample Probe Profile Window Columns.....	124
Table 17	Sample Probe Profile Window Per-Sample Columns .....	127
Table 18	Control Gene Profile Window Columns .....	129
Table 19	Control Gene Profile Window Per-Sample Columns .....	129
Table 20	Control Probe Profile Window Columns.....	131
Table 21	Control Probe Profile Window Per-Sample Columns .....	131
Table 22	Log Window Selections & Functions.....	135
Table 23	Main Menu Selections & Functions.....	136
Table 24	Bar Plot: Group Gene Profile Window Context Menu .....	140
Table 25	Other Tabbed Window Context Menu .....	140
Table 26	Project Window Context Menu.....	141
Table 27	Log Window Context Menu .....	141
Table 28	Data Section, Optional and Required Columns .....	144
Table 29	Experiment Parameters, 1.5 vs. 2.X .....	148
Table 30	Scatter Plots, 1.5 vs. 2.X .....	148
Table 31	Tools   Find Genes, 1.5 vs. 2.X .....	149
Table 32	Tools   Cluster Selected Genes, 1.5 vs. 2.X.....	149

Table 33	Filter Selected Genes, 1.5 vs. 2.X .....	150
Table 34	Selected List, 1.5 vs. 2.X .....	150
Table 35	Select by Detection, 1.5 vs. 2.X. ....	151
Table 36	Show R2, 1.5 vs. 2.X .....	151
Table 37	Frequently Asked Questions .....	153

# Chapter 1

## Overview

### Topics

- 2 Introduction
- 3 Audience and Purpose
- 3 Installing the Gene Expression Module
- 6 Gene Expression Module Workflow

## Introduction

The BeadStudio Gene Expression Module is a tool for analyzing gene expression data from scanned microarray images generated by the Illumina BeadArray Reader. You can use the resulting BeadStudio output files with most standard gene expression analysis programs.

The BeadStudio Gene Expression Module enables two types of data analysis:

- ▶ Gene Analysis—quantifying gene expression signal levels.
- ▶ Differential Analysis—determining whether gene expression levels have changed between two experimental groups.

You can perform these analyses on individual samples or on groups of samples treated as replicates.

The BeadStudio Gene Expression Module reports experiment performance based on built-in controls that accompany each experiment.

In addition, this module includes the following tools, which provide a quick, visual means for exploratory analysis:

- ▶ Line Plots
- ▶ Bar Graphs
- ▶ Scatter Plots
- ▶ Histograms
- ▶ Dendograms
- ▶ Heat Maps
- ▶ Samples Table
- ▶ Image Viewer
- ▶ Illumina Genome Viewer (IGV)
- ▶ Illumina Chromosome Browser (ICB)

## Audience and Purpose

This guide is written for researchers who want to use the BeadStudio Gene Expression Module to analyze data generated by performing Illumina's DirectHyb or DASL® assays.

This guide includes procedures and user interface information specific to the BeadStudio Gene Expression Module. For information about the BeadStudio Framework, the common user interface and functionality available in all BeadStudio Modules, refer to the *BeadStudio Framework User Guide*, Part # 11204578.

## Installing the Gene Expression Module

To install the BeadStudio Gene Expression Module application on your computer:

1. Put the BeadStudio CD into your CD drive.
2. Do one of the following:
  - If the Illumina BeadStudio Installation screen (Figure 2) appears, continue to step 3.
  - If the CD does not load automatically, double-click the **BeadStudio<version>.exe** icon in the BeadStudio directory of the CD you received.

The BeadStudio application suite unzips (Figure 1).

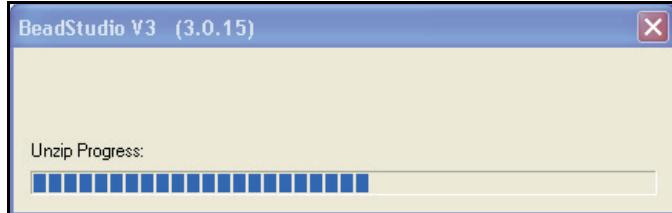
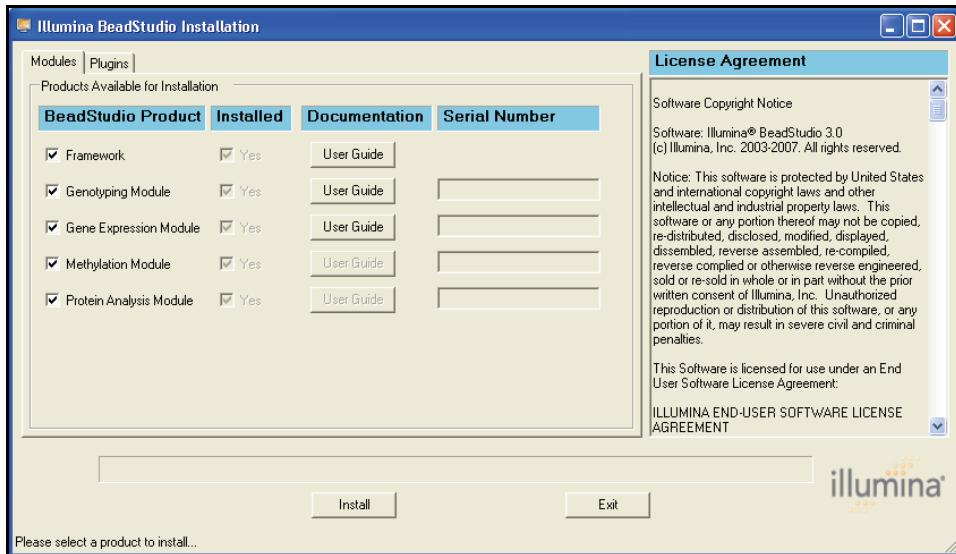


Figure 1 BeadStudio Application Unzipping

The **Illumina BeadStudio Installation** dialog box appears (Figure 2).



**Figure 2** Illumina BeadStudio Installation, Modules Tab

3. Read the software license agreement in the right-hand side of the **Illumina BeadStudio Installation** window.
4. In the BeadStudio Product area, select **Gene Expression Module**.



**NOTE**

Select additional BeadStudio modules if you have licenses for additional BeadStudio modules and want to install them now.

5. In the **Serial Number** area, enter your serial number for the Gene Expression module.



**NOTE**

Serial numbers are in the format #####-#####-##### and can be found on an insert included with your BeadStudio CD.

6. **[Optional]** Enter the serial numbers for additional BeadStudio modules if you have licenses for additional BeadStudio modules and want to install them now.

7. Click **Install**.

The **Accept License Agreement** dialog box appears (Figure 3).



Figure 3 Accept License Agreement

8. Click **Yes** to accept the software license agreement.

The Gene Expression Module is installed on your computer, along with any additional BeadStudio modules or plug-in algorithms you selected (Figure 4).

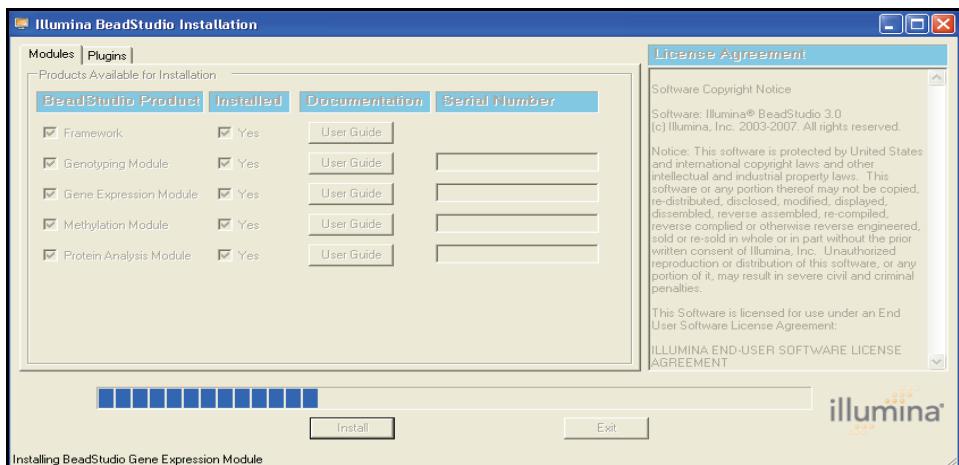


Figure 4 Illumina BeadStudio Installation

The **Installation Progress** dialog box notifies you that installation is complete (Figure 5).

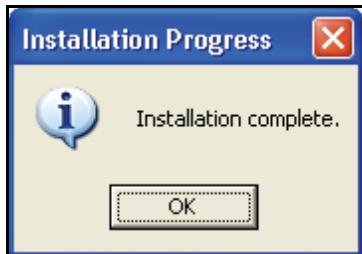


Figure 5 Installation Progress Dialog Box

9. Click **OK**.
10. In the **Illumina BeadStudio Installation** dialog box (Figure 4), click **Exit**.

You can now start a new project using the BeadStudio Gene Expression Module.

See Chapter 2, *Creating a New Project* for information about starting a new Gene Expression project.

## Gene Expression Module Workflow

The basic workflow for gene expression analysis using Illumina's BeadStudio Gene Expression Module is summarized in Figure 6.

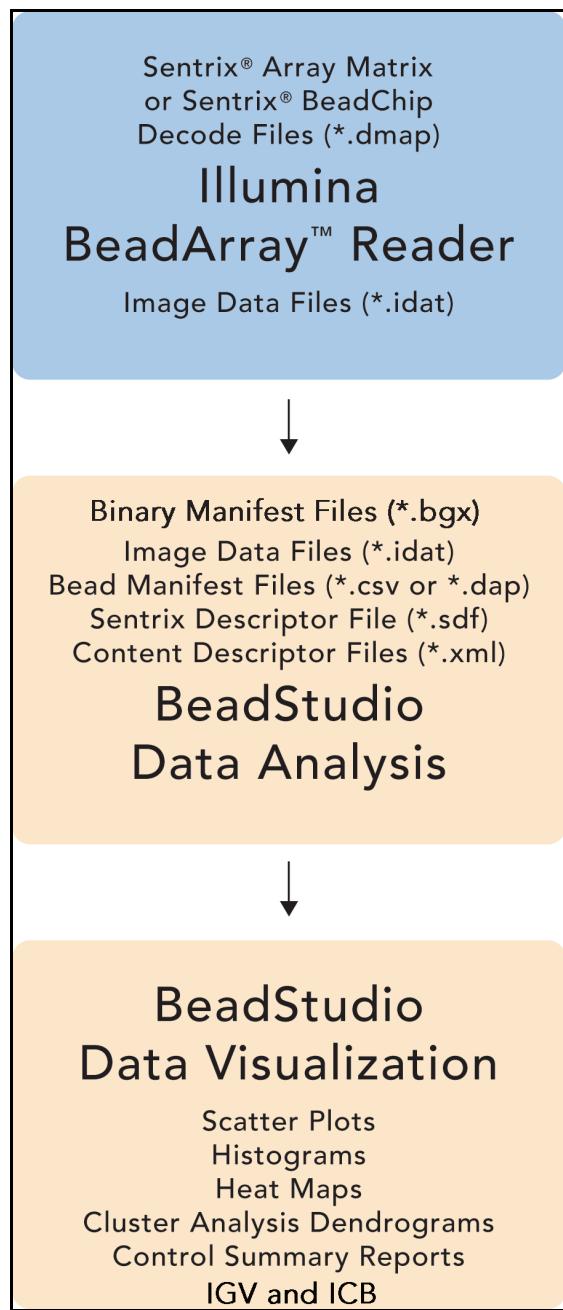


Figure 6 Gene Expression Analysis Workflow



# Chapter 2

# Creating a New Project

## Topics

- 10 Introduction
- 11 Creating a Project
  - 11 Starting the Gene Expression Module
  - 12 Selecting an Assay Type
  - 13 Choosing a Project Location
  - 15 Selecting Your Project Data
  - 18 Defining a Groupset
  - 24 Defining the Analysis Type and Parameters
  - 29 Creating a Mask File

## Introduction

Using the intensity files produced by the BeadArray Reader, BeadStudio's gene analysis tool produces data tables containing:

- ▶ Probe and gene lists
- ▶ Associated signal intensities (normalized or raw)
- ▶ Information about the system controls

In addition, BeadStudio's differential analysis tools can produce data tables displaying the probability that a gene's signal has changed between two samples or groups of samples.

BeadStudio's data visualization tools can create more sophisticated plotting analyses, including:

- ▶ Scatter Plots
- ▶ Cluster Analysis Dendograms
- ▶ Bar Graphs
- ▶ Line Graphs
- ▶ Histograms

To run a gene expression analysis, you first must create a BeadStudio project. In a project, you define one or more groupsets, one or more groups (sample sets that can be compared against each other for the purpose of identifying differences in gene expression), and one or more analyses. For more information about groupsets and groups, see *Defining a Groupset* later in this chapter.

In the simplest experiment, each group may have only one sample. However, if your experiment includes replicate samples, you can assign these to the same group.

In a project and within a group, BeadStudio averages the values for each gene across the samples, and its algorithms automatically take advantage of the beadtype replicates to provide accurate estimates of mRNA abundance. This translates into a highly sensitive determination of detection and differential expression.

The following section, *Creating a Project*, provides step-by-step instructions for the following tasks:

- ▶ Defining a project
- ▶ Creating groupsets and groups
- ▶ Applying normalization and differential expression algorithms
- ▶ Analyzing and visualizing your data

The BeadStudio Project Wizard guides you through creating a project, while the BeadStudio main page provides a starting-point from which you can carry out the same functions independently.

## Creating a Project

Follow the instructions in this section to create a project using the BeadStudio Project Wizard.

### Starting the Gene Expression Module

1. Start the BeadStudio Gene Expression Module by clicking **File | New Project | Gene Expression**.  
The BeadStudio Project Wizard—Welcome dialog box appears (Figure 7).

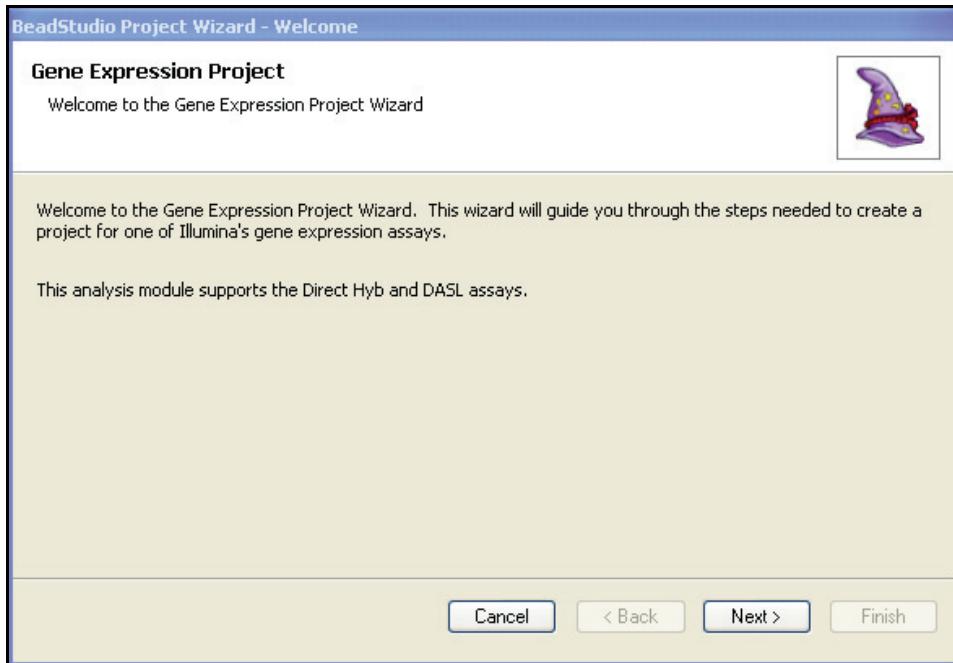


Figure 7 Project Wizard - Welcome

2. Click **Next** to advance to the BeadStudio Project Wizard—Gene Expression Assay Type dialog box.

#### Selecting an Assay Type

In the Gene Expression Assay Type dialog box (Figure 8), perform the following steps to select an assay type:

1. Specify an assay type by selecting **Direct Hyb** or **DASL**.

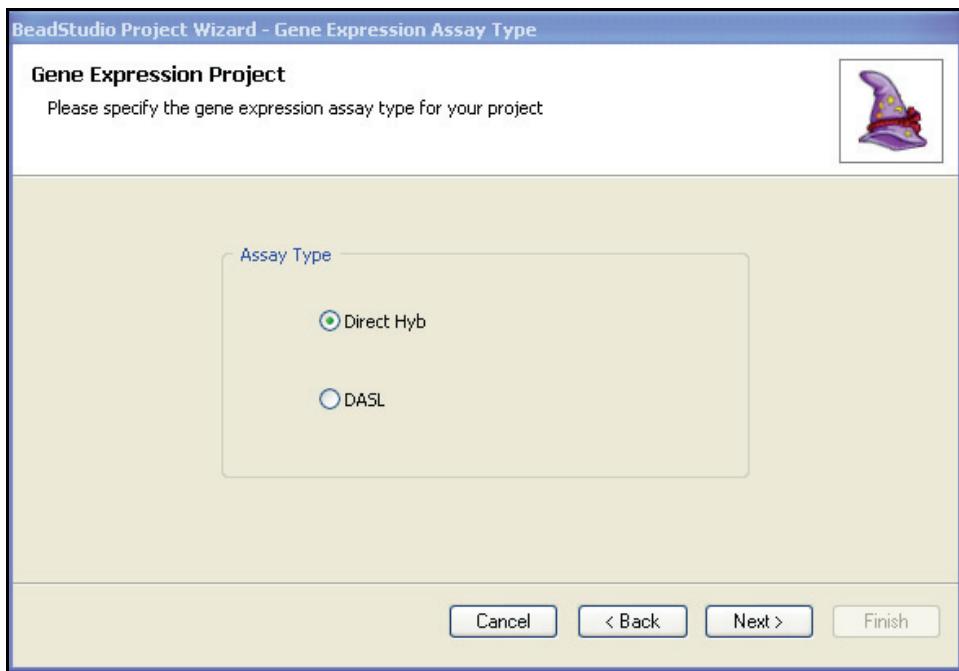


Figure 8 Project Wizard - Gene Expression Assay Type

2. Click **Next** to advance to the BeadStudio Project Wizard—Project Location dialog box.

### Choosing a Project Location

In the BeadStudio Project Wizard—Project Location dialog box (Figure 9), perform the following steps to choose a project location:

1. Browse to the location where you want to save your project.
2. Enter a name for your project.  
The full path for your project appears beneath the name you enter.

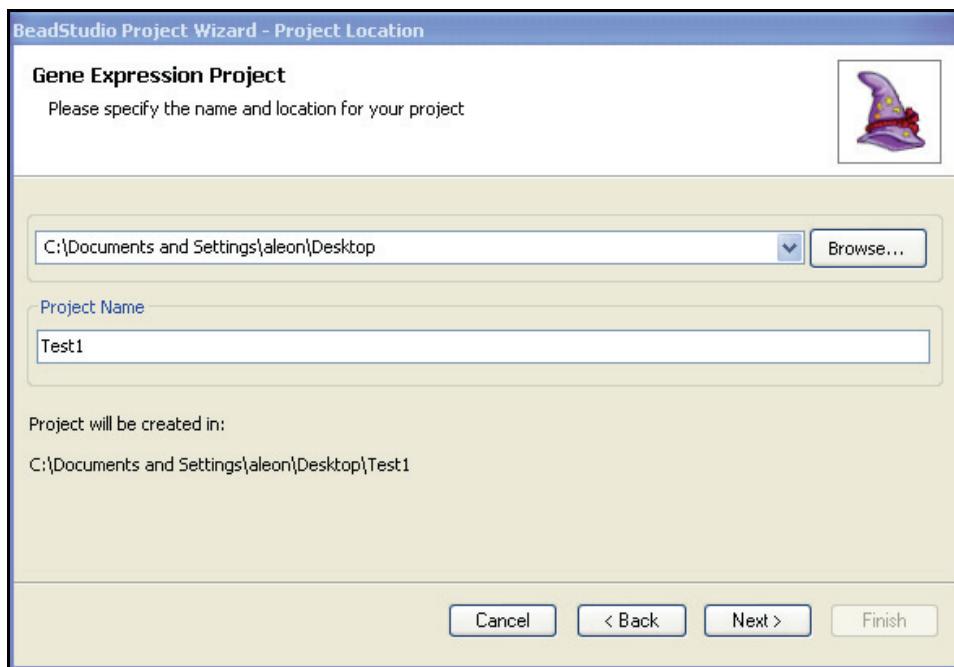


Figure 9 Project Wizard - Project Location

3. Click **Next** to advance to the BeadStudio Project Wizard—Project Data Selection dialog box.

## Selecting Your Project Data

In the BeadStudio Project Wizard—Project Data Selection dialog box (Figure 10), perform the following steps to select your project data:

1. In the **Repository** dropdown menu, browse to the project repository folder where your intensity data (\*.idat file) is stored.

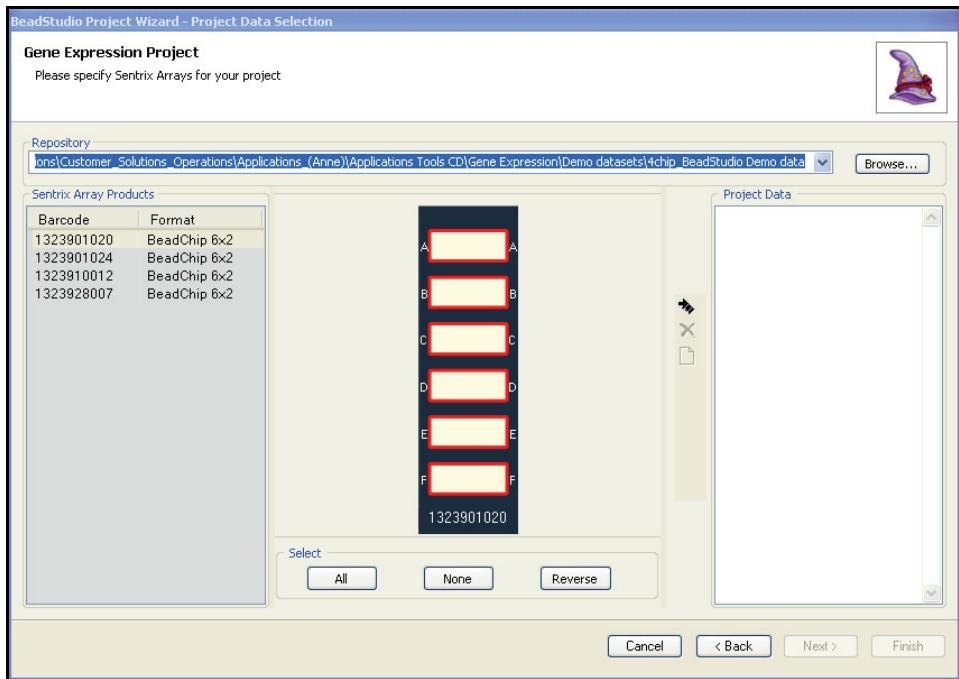


Figure 10 Project Wizard - Project Data Selection, Repository

2. In the Sentrix Array Products pane, select a Sentrix Array Product that you want to include in your project.

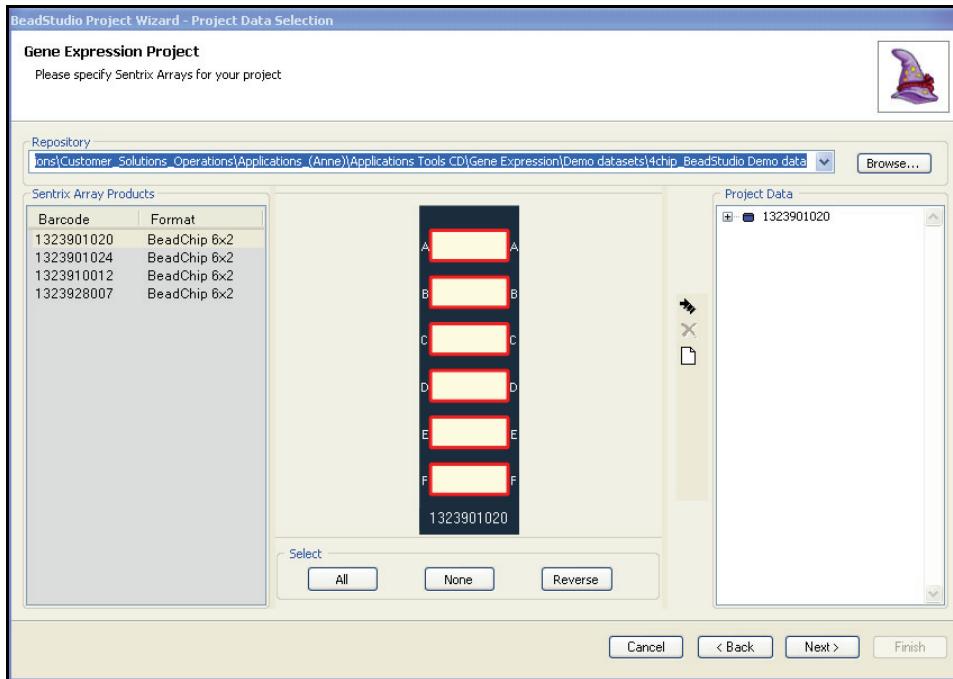


Figure 11 Project Wizard - Project Data Selection, Sentrix Array Products

All of the samples for each Sentrix Array Product are selected by default.



**NOTE**

The image of the Sentrix Array Product will be different for different products.

3. To change the selected samples, use the **All**, **None**, and **Reverse** buttons in the Select area.
  - To select a single sample, click the sample (on the image of the Sentrix Array Product).
  - To select multiple samples, press and hold **Ctrl** and click each sample you want to select.
  - To select all samples, click **All**.
  - To select no samples, click **None**.
  - To select the reverse of the samples currently selected (for example, samples 1, 2, and 3 are currently selected, but you want to select samples 4, 5, and 6), click **Reverse**.
4. Click  to add the selected samples to your project.

The selected samples appear under the name of the Sentrix Array Product in the Project Data pane.
5. **[Optional]** Click  (to the left of the group symbol) to display the list of samples chosen for the current project.

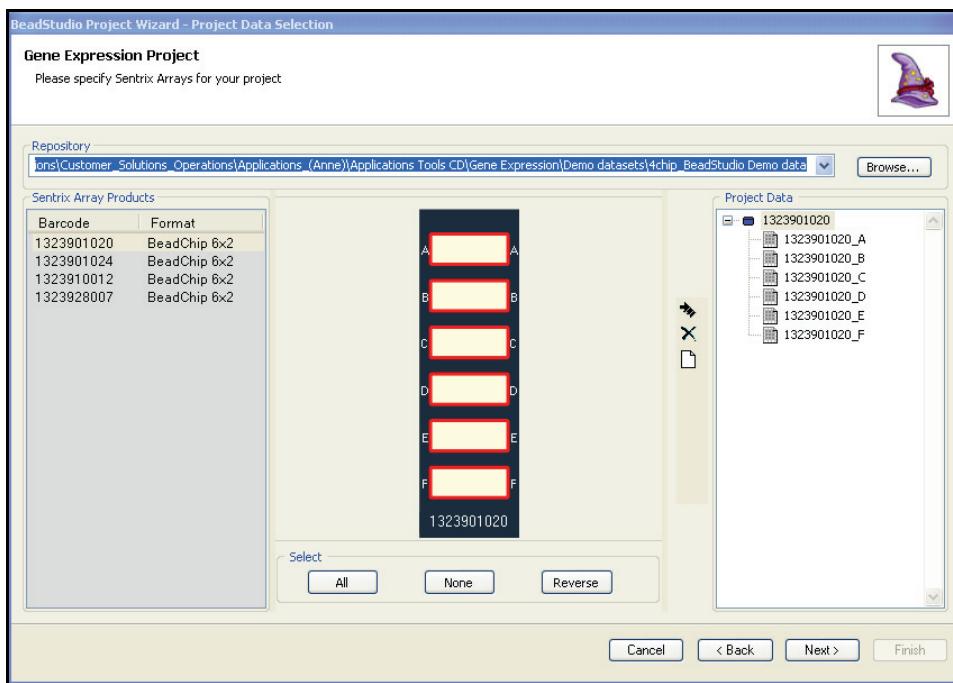


Figure 12 Project Wizard - Project Data Selection, Selected Samples

- Click **Next** to advance to the BeadStudio Project Wizard—Groupset Definition dialog box (Figure 13).

### Defining a Groupset

BeadStudio projects are structured in a hierarchical manner:

- A **project** includes one or more groupsets.
- A **groupset** is a collection of one or more groups that you choose to analyze simultaneously.
- A **group** is a set of arrays that share a functional relationship (e.g., replicates, zero time points, reference group). Within a groupset, an array can be included in more than one group, or it can be analyzed individually.

There are two types of analysis: **gene analysis** and **differential expression analysis**. You perform an analysis on a single groupset at a time.

Perform the following steps to define a groupset for your project.

1. Assign a name to your groupset by doing one of the following:
  - Click **New** and type a name for your new groupset.
  - Click **Existing** and choose the groupset you want from the dropdown menu.

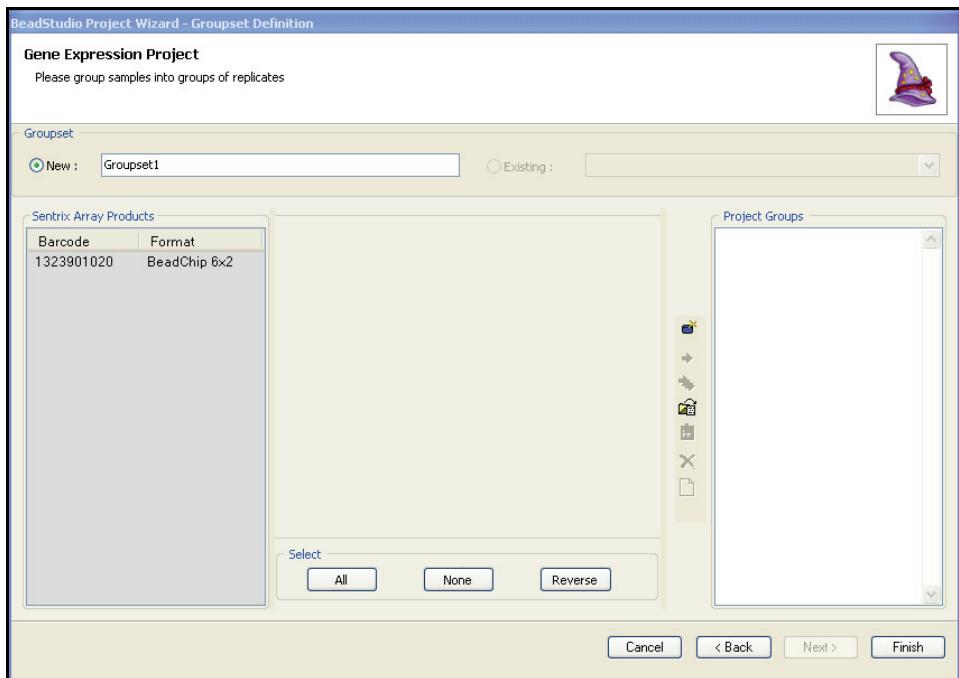


Figure 13 Project Wizard - Groupset Definition, Assigning a Groupset Name

2. In the Sentrix Array Products pane, select the Sentrix Array Product that contains the samples you want to assign to a groupset.

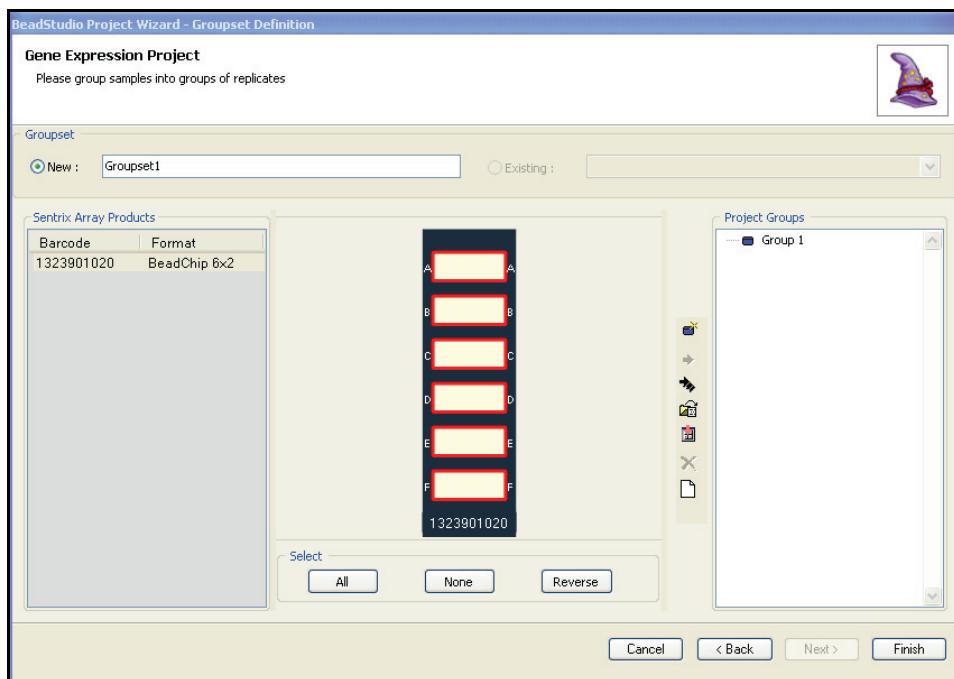


Figure 14 Project Wizard - Groupset Definition, Selecting a Sentrxi Array Product

3. Use the **All**, **None**, and **Reverse** buttons to select the specific samples you want to assign to a groupset.

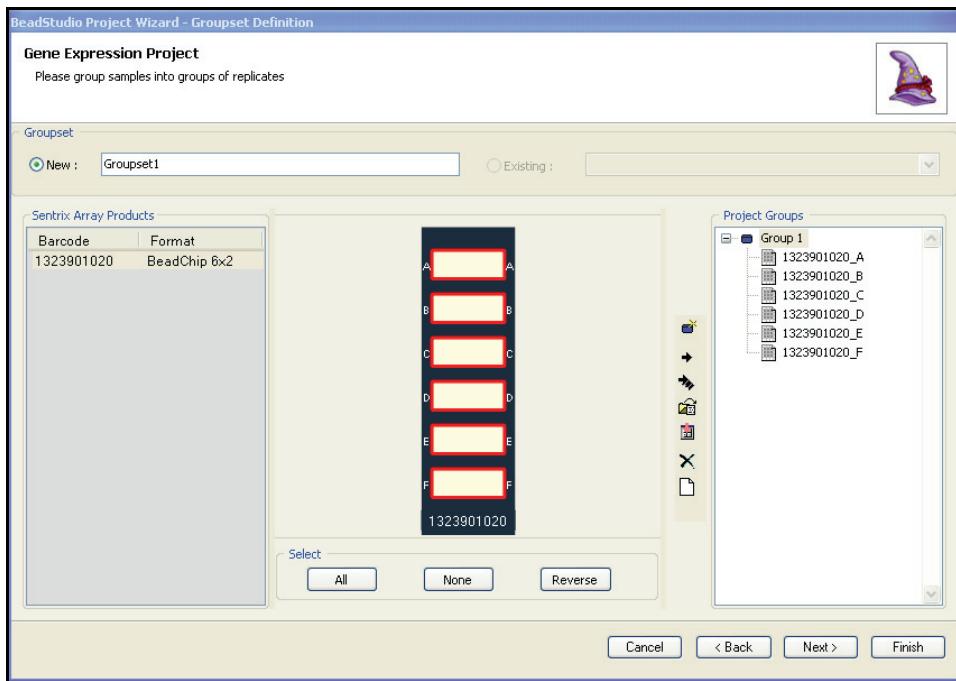


Figure 15 Project Wizard - Groupset Definition, Selecting Samples

4. Do any of the following to define project groups in a groupset:

To...	Click...
Create a new group	
Add selected samples to a group	
Create a group for each selected sample	
Load data from a sample sheet	

To...	Click...
Apply a group layout file	
Remove selected groups and samples from the groupset	
Remove all groups and samples from the project	

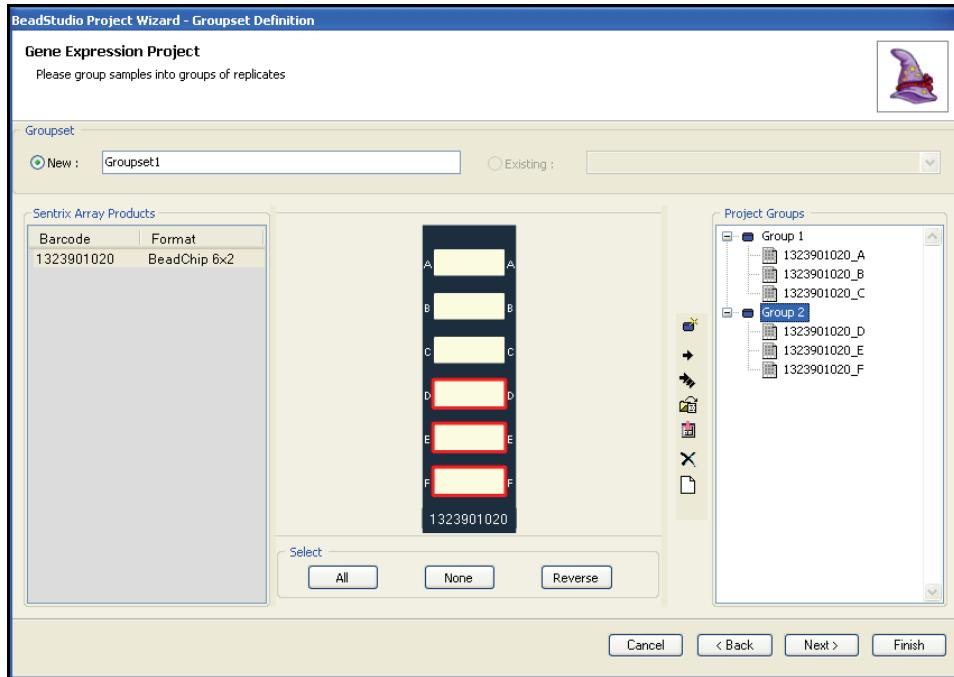


Figure 16 Project Wizard - Groupset Definition, Defining Project Groups

5. [Optional] Click **Finish** to finish building the groupset for your project.

For a Direct Hyb project, a data table containing the manifest data for your groupset appears in the main window (Figure 17).

The screenshot shows the BeadStudio interface for a 'Gene Expression - Test1' project. The main window displays a data table titled 'Manifest: Human\_WG-6.csv'. The table has columns: SEARCH\_KEY, TARGET, PROBEID, GID, TRANSCRIPT, and ACCESSION. The data table contains approximately 47,000 rows of manifest information. To the right of the main window is a 'Project' pane titled 'Name: Test1' which lists 'Manifests', 'Data', 'Group Sets', and 'Miscellaneous'. Below the main window is a 'Log' pane showing project logs and statistics: Last = 2:12 PM, Errors = 0, Warn = 0, Info = 11. A message at the bottom of the log pane says 'Project is ready'.

**Figure 17** Project Wizard - Groupset Definition, New Project

For DASL, no data table is displayed.



DASL is a universal array, so there is no up-front association between the data and the manifest file.

- Click **Next** to advance to the Project analysis type and parameters dialog box (Figure 18).

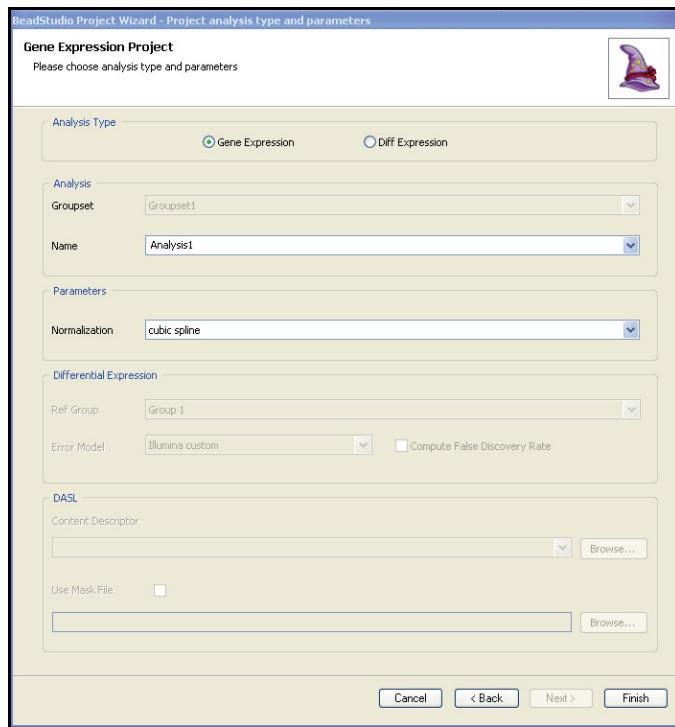


Figure 18 Project Analysis Type and Parameters

## Defining the Analysis Type and Parameters

Perform the following steps to run an analysis:

1. In the **Analysis Type** area, select **Gene Expression** or **Diff Expression**.
2. In the **Parameters** area, select the normalization method you want to use:
  - Average
  - Background
  - Cubic spline
  - None
  - Rank invariant

For information about normalization methods, see *Normalization Methods & Algorithms* on page 92.

3. If you are performing a differential expression analysis, select a **Reference Group** and **Error Model** in the **Differential Expression** area.
4. If you want to compute the false discovery rate, select the **Compute false discovery rate** checkbox.

**NOTE**

If you select **Compute false discovery rate**, the p-values (in the p-value column in the table) are adjusted accordingly. The p-values are not adjusted if you do not select **Compute false discovery rate**.

If you are performing a gene expression analysis using a Direct Hyb assay, you will not be able to select additional options because they do not apply to gene expression analysis projects.

5. Click **Finish** to run your analysis.
6. Select a content descriptor file (\*.xml), or a binary manifest file (\*.bgx).

**NOTE**

Text-based content descriptor files (\*.xml) and manifest files (\*.csv) are the primary annotation files used with earlier versions of Illumina Gene Expression products.

Newer Illumina Gene Expression products may also use binary manifest files (\*.bgx), which are more compact. Binary manifest files contain additional annotation information, such as chromosomal position, which allows you to view gene expression data in the Illumina Genome Viewer (IGV) and the Illumina Chromosome Browser (ICB).

## Selecting a Content Descriptor or Binary Manifest File

In order to use a text-based content descriptor file and corresponding manifest file in an analyses, the names must be identical to the content descriptor name embedded within the intensity files (\*.idat).

In order to use a binary manifest file in an analysis, it must adhere to Illumina naming conventions.

For example, for the HumanWG-6 v2 BeadChip, one possible binary manifest name is:

HumanWG-6\_V2\_0\_R1\_11223189\_B.bgx

Where:

- ▶ HumanWG-6 is the **product name**
- ▶ v2 is the **major version number** of the product
- ▶ 0 is the **minor version number** of the product
- ▶ R1 is the **annotation revision** of the product
- ▶ 11223189\_B is the **product identifier**

Only Sentrix Array products with identical product names and major version numbers can be combined in a single analysis.

### Example 1

HumanWG-6\_V2\_0\_R1\_11223189\_B.bgx

**can** be combined with

HumanWG-6\_V2\_1\_R2\_11223190\_C.bgx.

### Example 2

HumanWG-6\_V2\_0\_R1\_11223189\_B.bgx

**cannot** be combined with

HumanRef-8\_V2\_0\_R1\_11223189\_B.bgx

or with

RatRef-12\_V2\_0\_R1\_11223189\_B.bgx.

### Example 3

HumanWG-6\_V1\_0\_R1\_11223189\_B.bgx

**cannot** be combined with

HumanWG-6\_V2\_0\_R1\_11223189\_B.bgx.

7. **[Optional, DASL Only]** Select the **Use Mask File** checkbox to choose the mask file you want to use. For more information about mask files, see *Creating a Mask File* on page 29.

**NOTE**

It is possible that not all assay probes are functional, due to design or synthesis. You can use a mask file to filter out these non-functional probes from your data analysis. For more information about mask files, see *Creating a Mask File* on page 29.

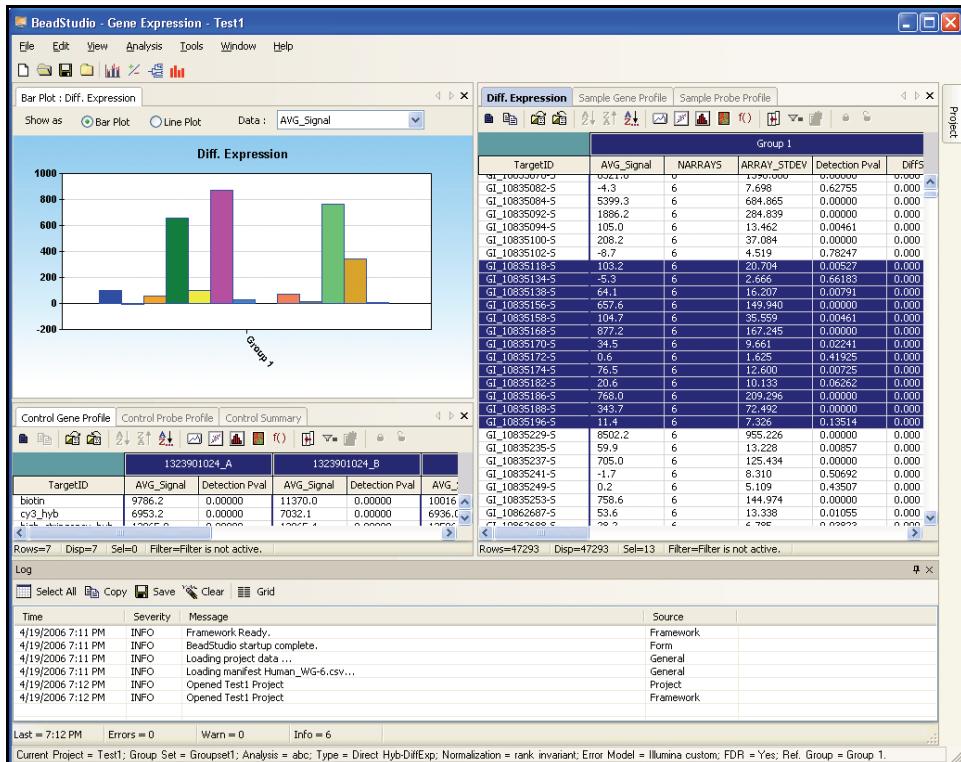
8. Click **Finish**.

BeadStudio begins to run your analysis. A progress bar indicates the completion level.

**NOTE**

The time it takes for your analysis to be processed depends on the number of samples, groups, and groupsets, and on the type of analysis you wish to perform

Your BeadStudio Gene Expression project appears (Figure 19).



**Figure 19** BeadStudio Gene Expression Analysis Results

For information about the Gene Expression Module graphical user interface, see Chapter 6, User Interface Reference.

## Creating a Mask File

If genomic DNA was used in a DASL Assay to verify probe performance, you can select probes that should be excluded from further analysis. Because all probes are designed to be intraexonic, all probes should be detectable when genomic DNA is used as a sample. Therefore, the Detection score reported in the `XXXXXX_gene_probe_profile` file can be used as an objective measure of probe performance on genomic DNA.

Illumina recommends excluding probes that have a Detection score of less than 1 on genomic DNA. However, you may define your own exclusion criteria.

In DASL mode, the `XXXXXX_gene_probe_profile` file contains an additional column named **0/1**. By default, this column will display a 1 in all rows containing probe-level data.

To exclude a probe:

1. Change the 1 to 0 in the **0/1** column.
2. Delete all the data columns between the **Probe\_ID** column and the **0/1** column.
3. Save the file as a \*.csv file in the same repository where the Content Descriptor file is stored. The file need not conform to a naming convention.



### NOTE

Any \*.csv file present in the same repository as Content Descriptor files will appear in the **Experiment Parameters** pulldown menu. To avoid confusion, Illumina recommends using separate repositories for Content Descriptor and for SAM/BeadChip data.

## Applying a Group Layout File

BeadStudio provides an optional alternative method for creating large numbers of groups in complex experiments that reduces project set-up time. The  **Apply Group Layout File** option allows you to apply a group layout file you previously created in Excel (or a similar application) to a single BeadChip or SAM.

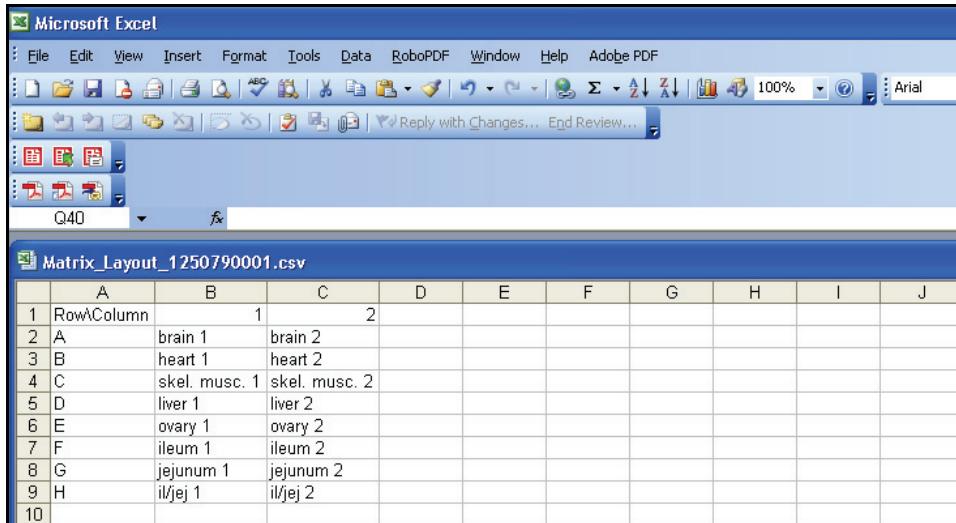


### NOTE

Group layout files must be saved in \*.csv format.

BeadStudio creates groups according to the specifications in the group layout file, and adds selected samples to those groups.

Perform the following steps to create a group layout file (Figure 20).



	A	B	C	D	E	F	G	H	I	J
1	Row\Column		1	2						
2	A	brain 1	brain 2							
3	B	heart 1	heart 2							
4	C	skel. musc. 1	skel. musc. 2							
5	D	liver 1	liver 2							
6	E	ovary 1	ovary 2							
7	F	ileum 1	ileum 2							
8	G	jejunum 1	jejunum 2							
9	H	il/jej 1	il/jej 2							
10										

Figure 20 Group Layout File Example

1. In the **Groups** area, click  **Apply Group Layout**.

The **Open** dialog box appears (Figure 21).

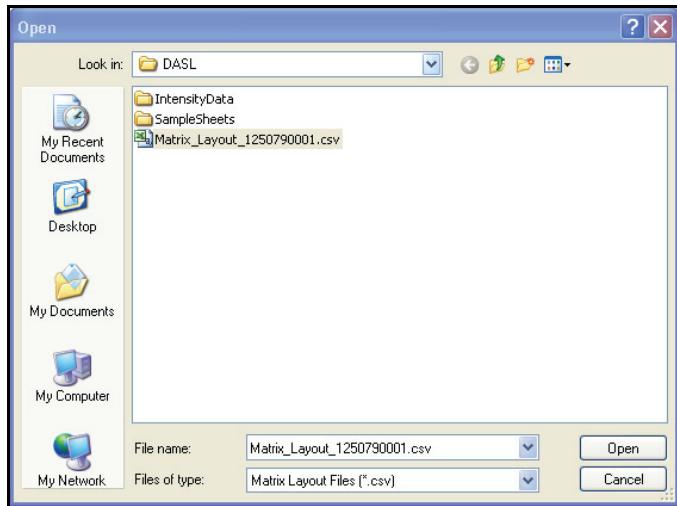


Figure 21 Open Dialog Box

2. Navigate to your group layout file and click **Open**.  
Samples on the selected BeadChip or Sentrix Array Matrix are mapped into groups according to the specifics of the group layout file you applied.

The groups are displayed in the Groups area (Figure 22).

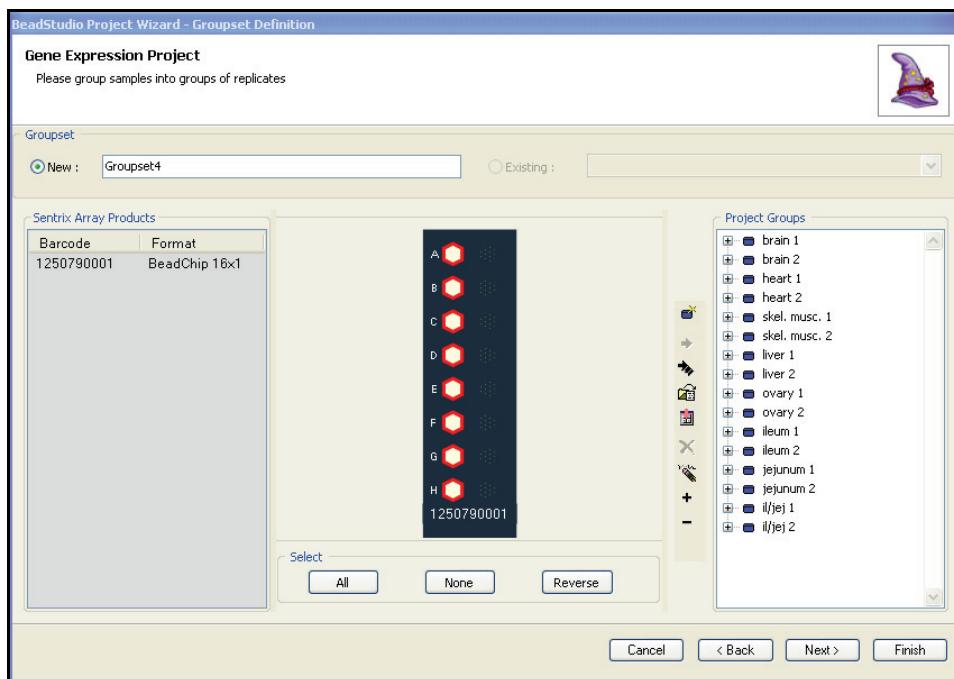


Figure 22 Samples Grouped Based on Group Layout File

3. [Optional] To display the samples in a group, click the plus sign to the left of that group (Figure 23).

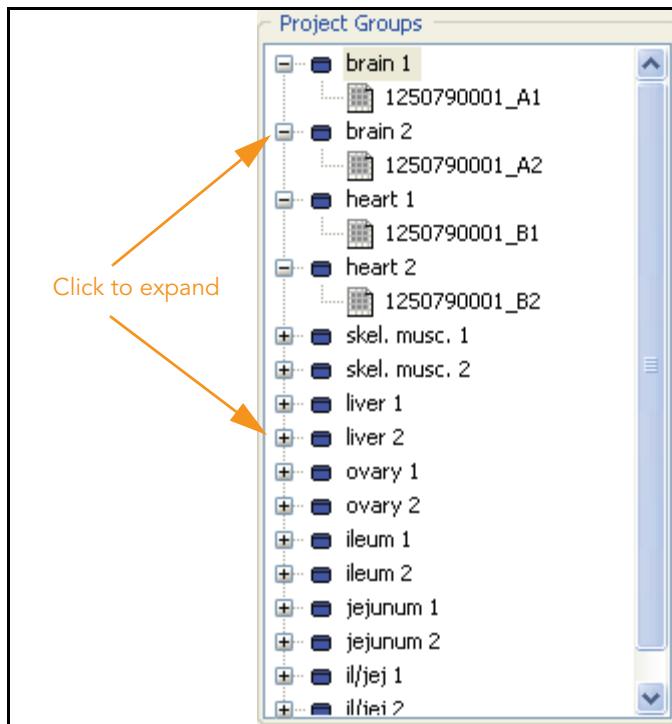


Figure 23 Groups Expanded



**NOTE**

Each BeadChip or Sentrix Array Matrix layout must be saved in a separate \*.csv file.



# Chapter 3

# Viewing Your Data

## Topics

- 36 Introduction
- 36 Scatter Plots
- 61 Bar Plots
- 64 Heat Maps
- 68 Cluster Analysis Dendrograms
- 74 Copy/Paste Clusters
- 78 Control Summary Reports
- 84 Image Viewer

## Introduction

This chapter describes the data visualization functions of the BeadStudio Gene Expression Module, which are used to create and display:

- ▶ Scatter Plots
- ▶ Bar Plots/Line Plots
- ▶ Heat Maps
- ▶ Cluster Analysis Dendograms
- ▶ Control Summary Reports
- ▶ Histograms
- ▶ Images

Use these tools to explore the data you create using the Gene Analysis or Differential Expression Analysis tools (described in Chapter 4, *Normalization and Differential Analysis*).

In BeadStudio v3, the Gene Expression Module also includes the Illumina Genome Viewer (IGV), the Illumina Chromosome Browser (ICB), and the Illumina Sequence Viewer (ISV). For more information about these tools, see the BeadStudio Framework User Guide, Part # 11204578.

## Scatter Plots

Once gene analysis and/or differential analysis have been completed, you can create scatter plots.

To create a scatter plot:

1. Click  **Scatter Plot**.

The **Plot Columns** dialog box appears (Figure 24).

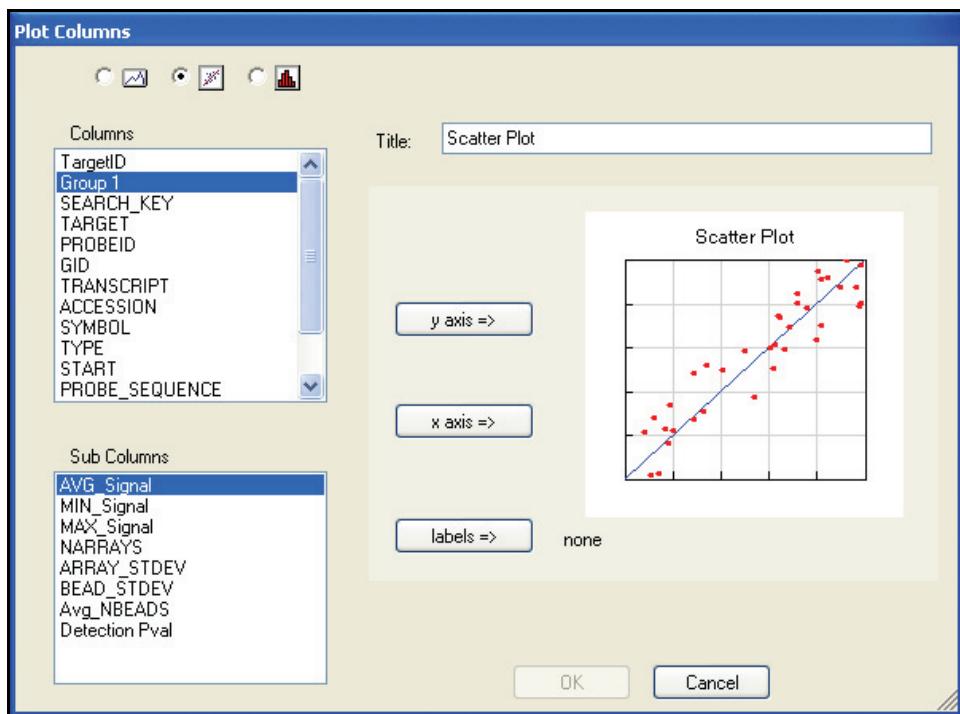


Figure 24 Plot Columns Dialog Box

2. In the **Plot Columns** dialog box, select options from the **Columns** and **Sub Columns** areas for:
  - **Y-axis**
  - **X-axis**
  - **Labels**

You can choose any subcolumn that contains numerical data for the axes.

The label you chose appears when you position your cursor over a point in the scatter plot.

3. Click **OK** to create and display the scatter plot (Figure 25).

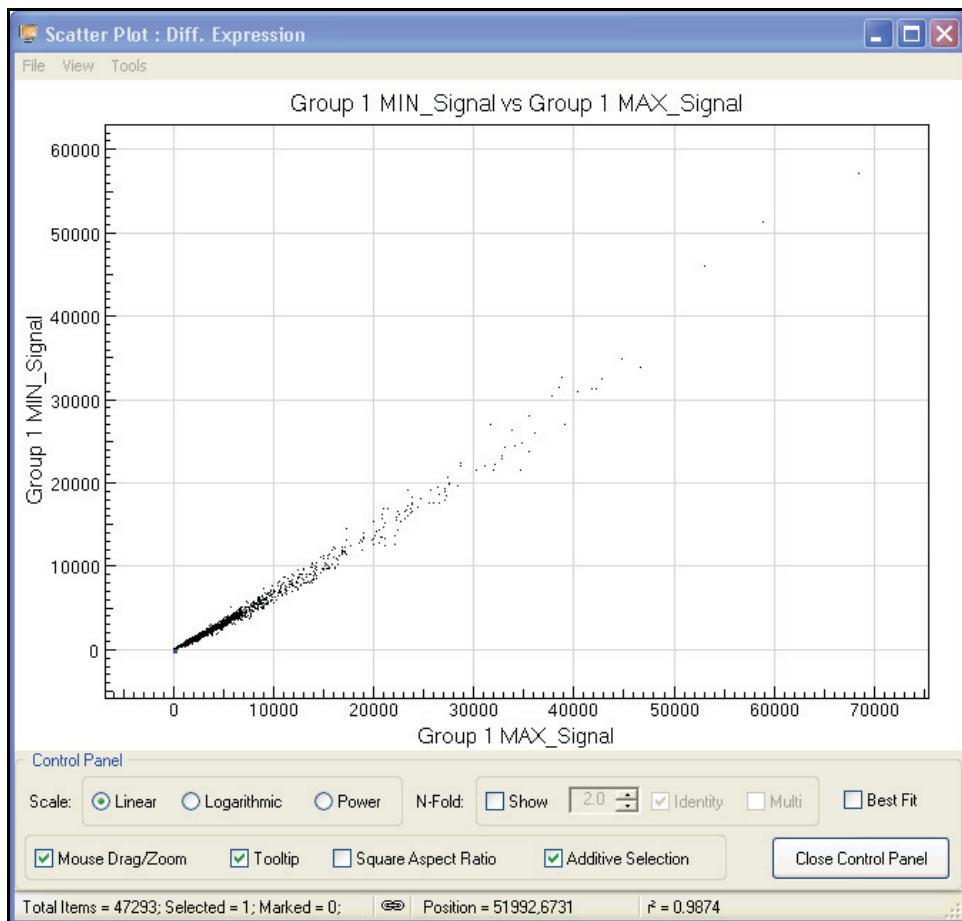


Figure 25 Scatter Plot

The **Control Panel** appears (lower portion of Figure 25).

## Control Panel

Table 1 describes scatter plot **Control Panel** functions.

**Table 1** Scatter Plot Control Panel Functions & Descriptions

Function	Item	Description
Scale	Linear	When enabled, X and Y axes are on a linear scale.
	Logarithmic	When enabled, X and Y axes are on a logarithmic scale.
	Power	When enabled, X and Y axes are on an nth root scale, where n is an odd number from 3 to 9. This allows visual separation of negative values from positive values.
N-Fold	Show	When checked, shows n-fold lines and allows you to select the fold value.
	N-fold setting selector	When <b>Show</b> is checked, allows you to select the fold change.
	Identity	When checked, BeadStudio displays the identity line in bold red color. If a gene is on this line, its X and Y intensities are equal.
	Multi	When checked, BeadStudio displays additional incremental fold-change regions.

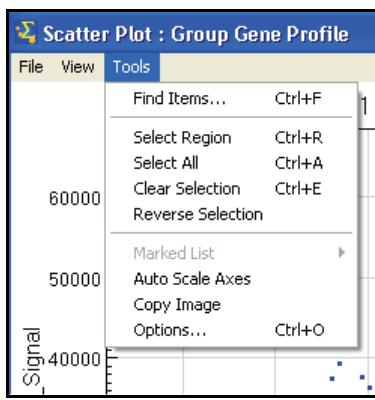
**Table 1** Scatter Plot Control Panel Functions & Descriptions (continued)

Function	Item	Description
Options	<b>Best Fit</b>	When checked, presents the Scatter Plot in the optimal fit for the genes of interest. The linear equation is displayed in Control Panel next to $R^2$ values.
	<b>MouseDrag /Zoom</b>	When checked, allows you to drag and zoom in or out using the mouse wheel. If your mouse does not have a wheel: <ul style="list-style-type: none"> <li>• Press and hold the <b>Shift</b> key while clicking the left mouse button.</li> <li>• Drag to create a rectangle around an area to zoom in on.</li> <li>• Release <b>Shift</b> and the mouse button to zoom.</li> <li>• To return to normal view, select Scatter Plot Tools   <b>Auto Scale Axes</b>.</li> </ul>
	<b>Tooltip</b>	When checked, the scatter plot displays the label you chose.
	<b>Square Aspect Ratio</b>	When checked, the X axis scale is equal to Y axis scale.
	<b>Additive Selection</b>	When checked, any new gene selection will be added to the scatter plot, along with previous selections. When not checked, any new selection replaces the previous selection(s).
	<b>Close Control Panel</b>	When clicked, closes the Control Panel.
Status Bar	<b>Total Items =</b>	The number of total items visible in the Scatter Plot.
	<b>Selected =</b>	The number of selected genes in the Scatter Plot.
	<b>Marked =</b>	The number of marked items visible in the Scatter Plot.
	<b>Linking</b>	 When clicked, toggles synchronization of marking and selection of the data shown in the scatter plot with data in the table.
	<b>Position</b>	Displays current X/Y position of gene (mouse pointer) on the Scatter Plot.

**Table 1** Scatter Plot Control Panel Functions & Descriptions (continued)

Function	Item	Description
	R <sup>2</sup>	Square of the correlation coefficient. Note: If the scatter plot is in linear scale, the R <sup>2</sup> value is calculated in linear space; if the scatter plot is in logarithmic scale, R <sup>2</sup> is calculated in log space.

4. To use additional plot tools, open the **Tools** menu.

**Figure 26** Scatter Plot Tools Menu

## Tools Menu

Table 2 describes various scatter plot tools.

**Table 2** Scatter Plot Tools Menu Item Descriptions

Tool Name	Description
Find Items	Opens the Find Items dialog box, from which you can enter a list of items separated by commas, or load a search item list from a text file.
Select Region	Converts the cursor to a crosshair tool, which you can use to draw a boundary around any region in the scatter plot. All genes within the boundary are selected.
Select All	Selects all genes in the scatter plot. Genes are displayed in the currently-selected color.
Clear Selection	Clears any previous selections.

**Table 2** Scatter Plot Tools Menu Item Descriptions

Tool Name	Description
Reverse Selection	Reverses the current selection (selects genes that are unselected and clears genes that are selected).
Marked List	<p>Includes operations you can perform on genes you mark in the scatter plot:</p> <ul style="list-style-type: none"> <li>▶ <b>View in Web Browser</b>—Displays a list of the marked genes in a web browser.</li> <li>▶ <b>Save in Text File</b>—Allows you to save genes in a file in a location you specify.</li> <li>▶ <b>Show Item Labels</b>—Shows item labels, if you applied a label when you created the scatter plot.</li> </ul>
Select by Detection	<p>Allows you to select points in a scatter plot based on a detection p-value cut-off.</p> <ul style="list-style-type: none"> <li>▶ Both Samples—Uses the same cut-off for both samples.</li> <li>▶ Sample X—Uses the cut-off only for the X-axis sample.</li> <li>▶ Sample Y—Uses the cut-off only for the Y-axis sample.</li> </ul>
Select by Diff Score	Allows you to select points in a scatter plot based on a Diff Score cut-off for the sample chosen as the Y-axis.
Auto Scale Axes	Automatically scales the X and Y axes of the scatter plot.
Copy Image	Copies the current image to the clipboard.
Options	<p>Opens the Scatter Plot dialog box, in which you can set the following parameters:</p> <ul style="list-style-type: none"> <li>▶ <b>Axes</b>—Displays the minimum and maximum X and Y axis values. When Square Aspect Ratio is not checked, you can set new X and Y axis values.</li> <li>▶ <b>Labels</b>—Allows you to choose font properties for the scatter plot title and axes.</li> <li>▶ <b>Data Points</b>—Allows you to select a point size and style for the Scatter Plot data points.</li> <li>▶ <b>Scale</b>—Allows you to select a power (3, 5, 7, or 9) for the Power setting.</li> <li>▶ <b>Colors</b>—Allows you to set colors for: <ul style="list-style-type: none"> <li>• Axes</li> <li>• Background</li> <li>• Grid</li> <li>• Data Points</li> <li>• Selection</li> </ul> </li> </ul>

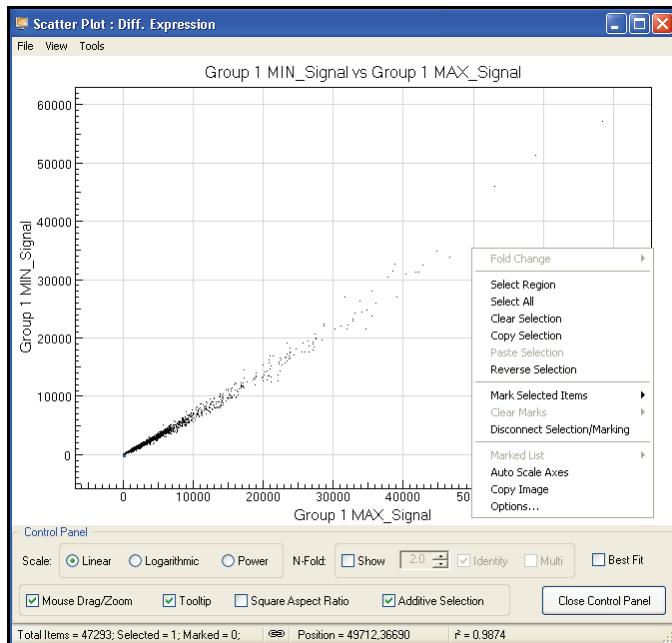
## Context Menu

The context menu contains options that can be applied to the selected project.

To view the scatter plot context menu:

- Right-click anywhere in the scatter plot.

The scatter plot context menu appears (Figure 27).



**Figure 27** Scatter Plot Context Menu

Table 3 lists context menu items and their functions.

**Table 3** Scatter Plot Context Menu Item Descriptions

Item	Description
<b>Fold Change</b>	If fold change lines are present, displays the fold change limits for the current cursor location. Allows you to select or deselect all genes inside the fold change.
<b>Select Region</b>	Allows you to select a region that contains samples of interest.
<b>Select All</b>	Allows you to select all samples.

**Table 3** Scatter Plot Context Menu Item Descriptions (continued)

Item	Description
<b>Clear Selection</b>	Clears your selection.
<b>Copy Selection</b>	Copies your selection to the clipboard.
<b>Paste Selection</b>	Pastes the contents of the clipboard to the current location.
<b>Reverse Selection</b>	Allows you to select the samples that were previously unselected.
<b>Mark Selected Items</b>	Allows you to mark items of interest.
<b>Clear Marks</b>	Clears your marks.
<b>Disconnect Selections/Marking</b>	Disconnects synchronization between the graph and the table.
<b>Marked List</b>	<p>Includes operations you can perform on genes you mark in the scatter plot:</p> <ul style="list-style-type: none"> <li>▶ <b>View in Web Browser</b>—Displays a list of the marked genes in a web browser.</li> <li>▶ <b>Save in Text File</b>—Allows you to save genes in a file in a location you specify.</li> <li>▶ <b>Show Item Labels</b>—Shows item labels, if you applied a label when you created the scatter plot.</li> </ul>
<b>Auto Scale Axes</b>	When selected, automatically scales the Scatter Plot X and Y axes.
<b>Copy Image</b>	When selected, places the Scatter Plot image on the clipboard.

**Table 3** Scatter Plot Context Menu Item Descriptions (continued)

Item	Description
Options	<p>Opens the Scatter Plot dialog box, in which you can set the following parameters:</p> <ul style="list-style-type: none"><li>▶ <b>Axes</b>—Displays the minimum and maximum X and Y axis values. When Square Aspect Ratio is not checked, you can set new X and Y axis values.</li><li>▶ <b>Labels</b>—Allows you to choose font properties for the scatter plot title and axes.</li><li>▶ <b>Data Points</b>—Allows you to select a point size and style for the Scatter Plot data points.</li><li>▶ <b>Scale</b>—Allows you to select a power (3, 5, 7, or 9) for the Power setting.</li><li>▶ <b>Colors</b>—Allows you to set colors for:<ul style="list-style-type: none"><li>• Axes</li><li>• Background</li><li>• Grid</li><li>• Data Points</li><li>• Selection</li></ul></li></ul>

## Finding Items

The BeadStudio Gene Expression Module provides a path to gene property information, including gene ID, intensities, and gene ontology information.

To find items in a scatter plot:

- From the menu bar, select **Tools | Find Items** (Figure 28).

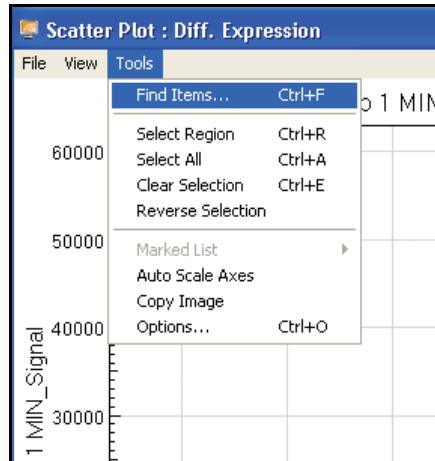


Figure 28 Find Items Tool

- In the **Find Items** dialog box (Figure 29), select specific items based on the following fields in the manifest, which includes GenBank database information:
  - Accession
  - Array Address ID
  - Chromosome
  - Definition
  - GI
  - Ontology Component
  - Ontology Function
  - Ontology Process
  - Probe Chr Orientation
  - Probe Coordinates
  - Probe ID
  - Probe Sequence
  - Probe Start

- Probe Type
- Search Key
- Source
- Source Reference ID
- Species
- Symbol
- Synonyms
- Transcript

In the **Search in** pane, select the manifest column you want to search.

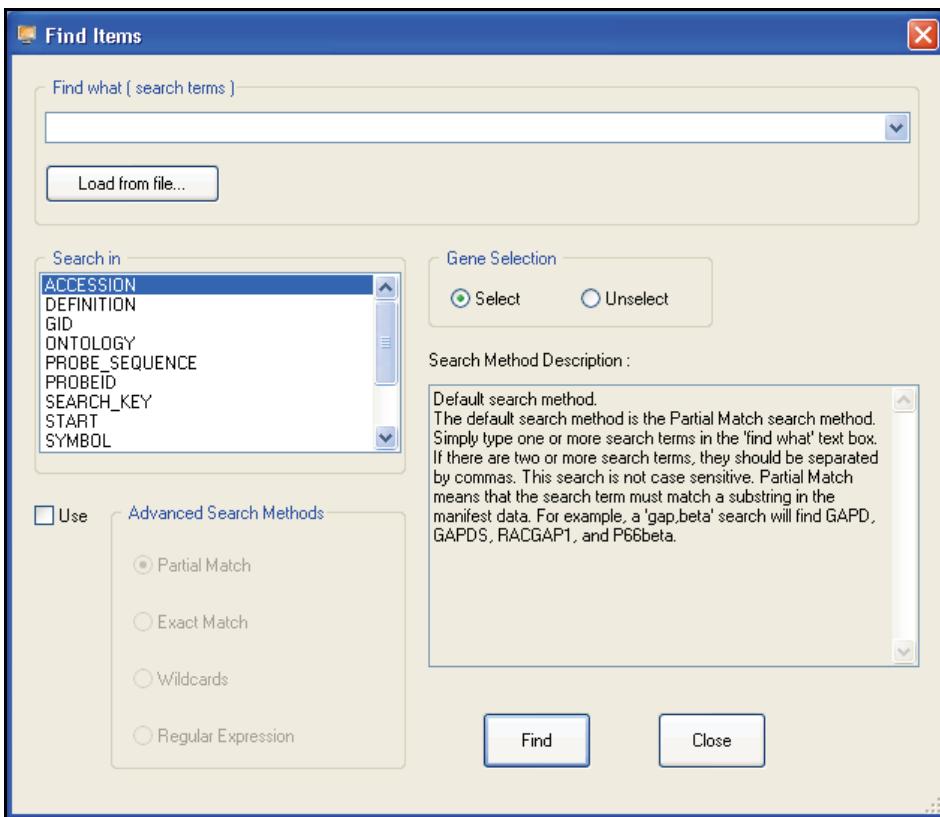


Figure 29 Find Items Dialog

3. In the **Find what (search terms)** text field, enter the search text.



## NOTE

By default, searches are partial. For example, if you search the word 'VEGF' in the **Symbol** field, the search will return not only VEGF, but also VEGFB and VEGFC.

Multiple search terms can be used, separated by commas.

Search terms can also be loaded from a text file. The file should have each term on a separate line.

4. Do one of the following:
  - Click **Select** to select found genes.
  - Click **Unselect** to clear found genes that were previously selected.
5. Click **Find** to return to the scatter plot with the identified genes highlighted.  
For more advanced search options, click **Use**, to the left of the **Advanced Search Methods** pane.  
Advanced search methods are described in the **Search Method Description** area (Figure 29).  
The scatter plot displays the selected gene (Figure 30).
6. **[Optional]** Use the mouse wheel to zoom in for a magnified view.

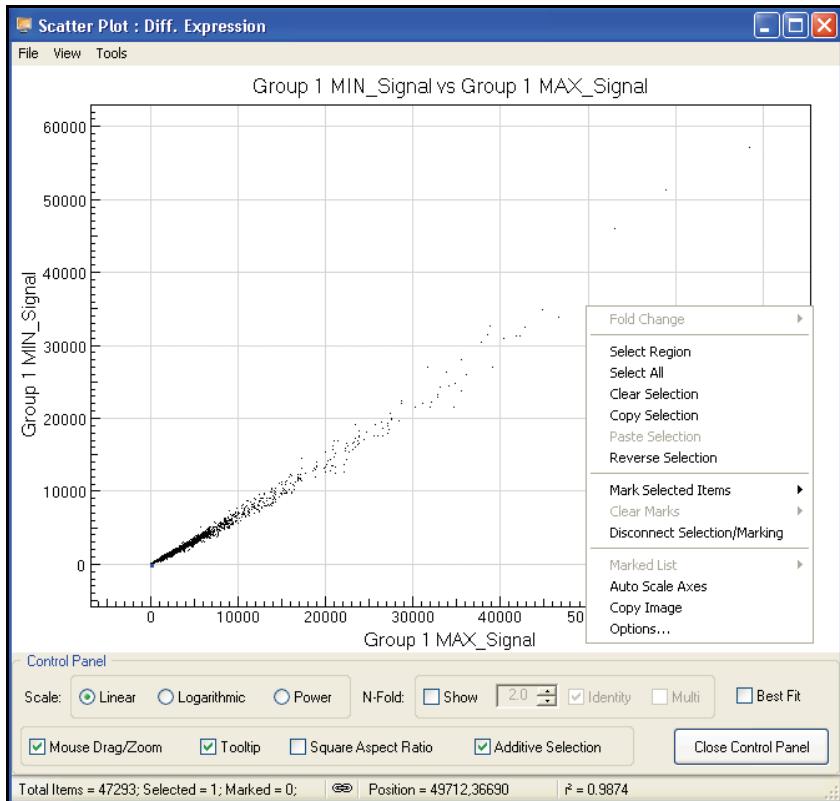


Figure 30 Zoom in to See Selected Genes

7. To display the **Gene Properties** dialog box,
  - a. Right-click the selected gene (Figure 30).
  - b. Click **Gene Symbol** in the context menu.

The **Gene Properties** dialog box appears (Figure 31).

The following paragraphs illustrate the functions of the **Gene Properties** dialog box.

### Data Tab

Figure 31 illustrates the **Data** tab of the **Gene Properties** dialog box.

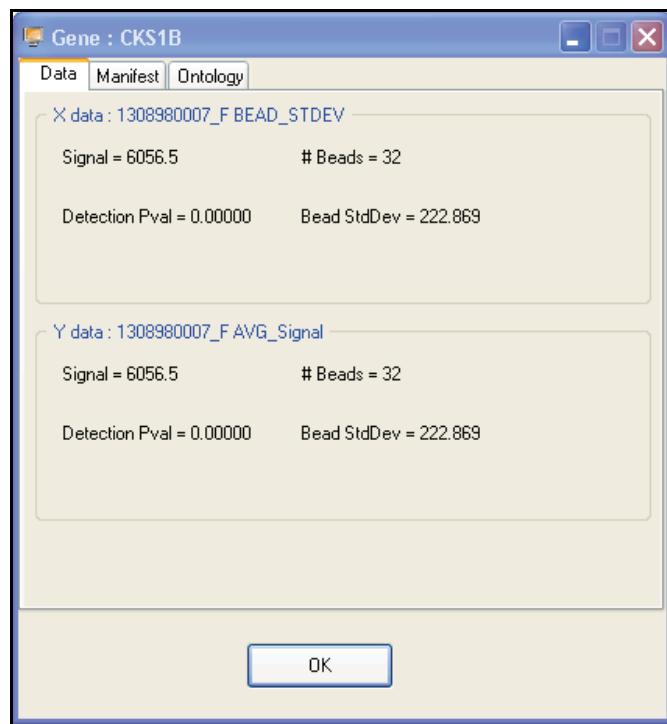


Figure 31 Gene Properties, Data Tab

## Manifest Tab

Figure 32, Figure 33, and Figure 34 illustrate functions of the **Manifest** tab of the **Gene Properties** dialog box.

1. Click the **Accession** link (Figure 32).

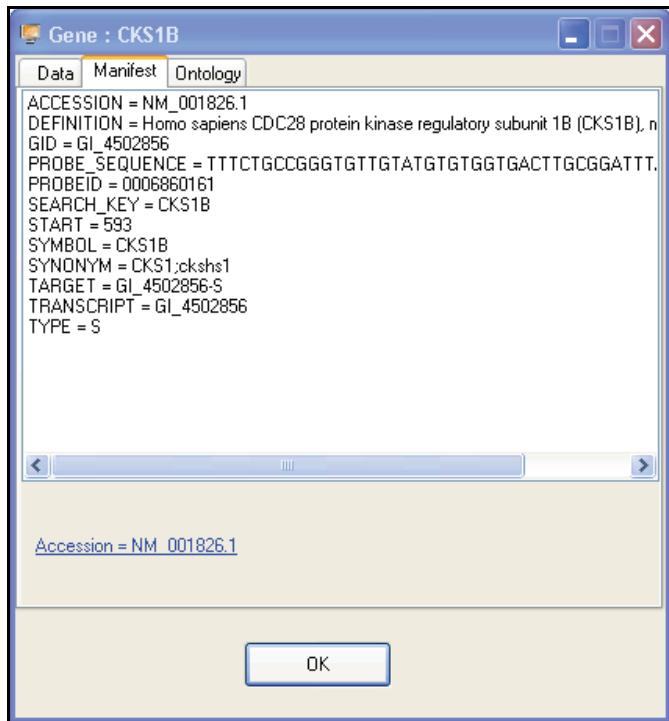


Figure 32 Gene Properties, Manifest Tab

BeadStudio jumps to the National Center for Biotechnology Information (NCBI) website (Figure 33) where you can view the record for the selected gene (Figure 34).

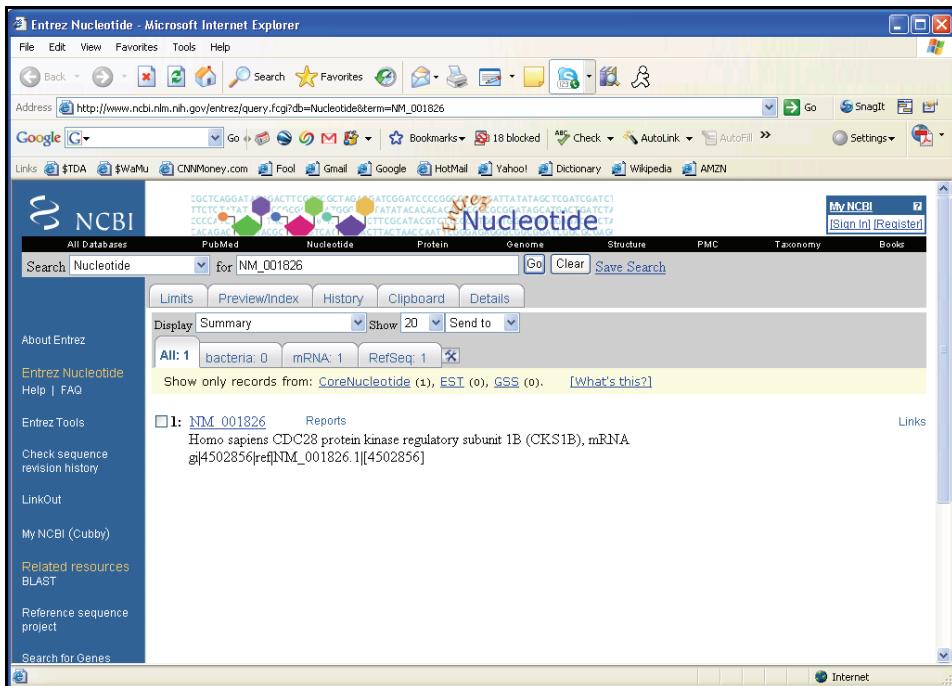


Figure 33 NCBI Website

2. Click the gene name to see the GenBank record (Figure 34).

The screenshot shows a Microsoft Internet Explorer window displaying the NCBI Sequence Viewer. The URL in the address bar is <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=4502856>. The main content area shows the GenBank record for NM\_001826. The record details the gene's definition as Homo sapiens CDC28 protein kinase regulatory subunit 1B (CKS1B), its locus (NM\_001826), and its version (NM\_001826.1 GI:4502856). It also lists keywords, source (Homo sapiens), organism (Homo sapiens), reference (1 bases 1 to 717), authors (Yao, Z.P., Zhou, M., Kelly, S.E., Seeliger, M.A., Robinson, C.V. and Itzhaki, L.S.), and title (Activation of ubiquitin ligase SCF(Skp2) by Cks1: insights from...). The page includes a sequence viewer at the top with a DNA sequence snippet.

Figure 34 NCBI Record for the Selected Gene

## Ontology Tab

Figure 35 illustrates the **Ontology** tab of the **Gene Properties** dialog box. This tab provides a quick reference to NCBI gene ontology information.

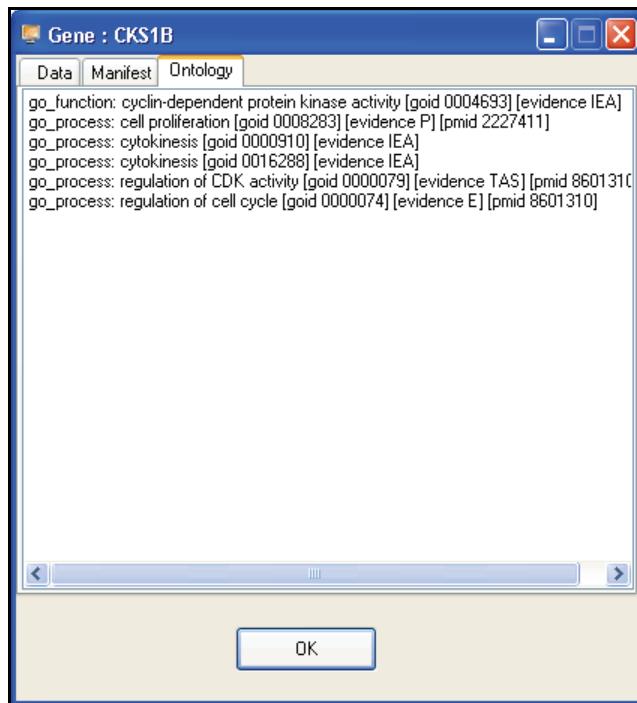


Figure 35 Gene Properties, Ontology Tab

## Viewing Marked Data

If you mark your data, you can use the scatter plot to view it in various ways.

For detailed information about marking data, refer to the *BeadStudio Framework User Guide*, Part # 11204578.

Once you have marked your data, you can select one of the following options from the **Marked List** option in the scatter plot context menu (Figure 36):

- ▶ **View in Web Browser**—displays your marked data in a web browser.
- ▶ **Save in Text File**—saves your marked data in a \*.txt file.
- ▶ **Show Item Labels**—displays item labels in the scatter plot.

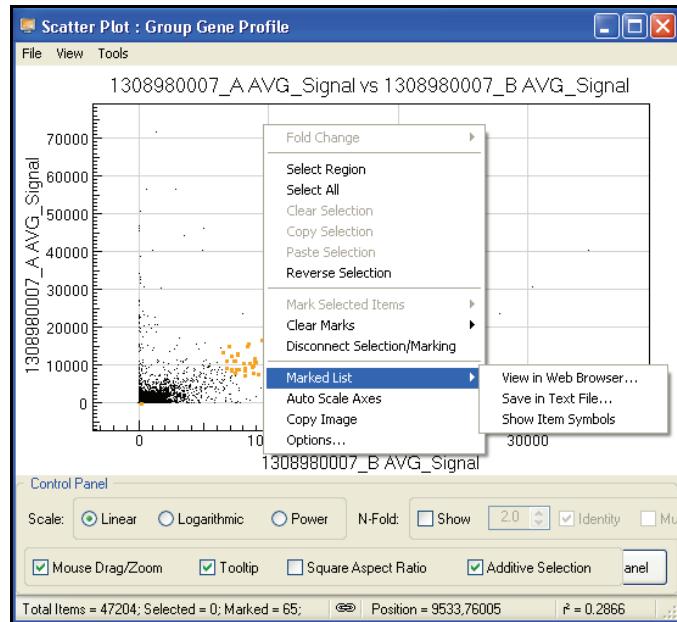


Figure 36 Scatter Plot Context Menu, Marked List Options

## Viewing Marked Data in a Web Browser

To view marked data in a web browser, do the following:

1. In the scatter plot context menu, select **Marked List | View in Web Browser**.  
The **BeadStudio Scatter Plot Output Data** dialog box appears (Figure 37).

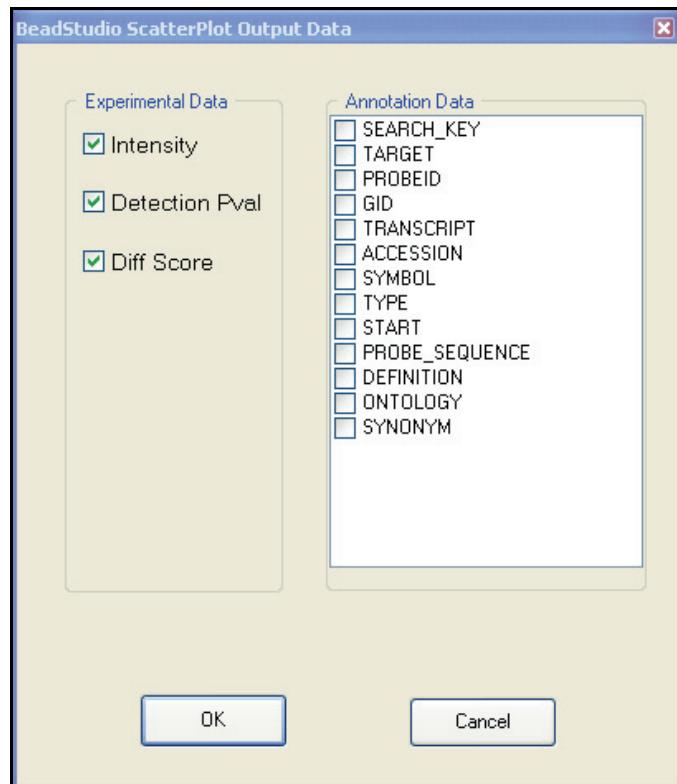


Figure 37 BeadStudio Scatter Plot Output Data Dialog Box

2. In the **Experimental Data** and **Annotation Data** areas, double-click selections for the data you want to include in the web browser.
3. Click **OK**.
4. The data is displayed in your default web browser (Figure 38).

TargetID	Signal_X	Signal_Y	DetectionPval_X	DetectionPval_Y	SEARCH_KEY	TARGET	PROBEID
GL_10092680-S	108.7	118.5	0.0197	0.0197	ADCK1	GL_10092680-S	0006020706
GL_10834975-S	8840.2	11027.3	0.0000	0.0000	GPX1	GL_10834975-S	0006200520
GL_10863926-S	11967.8	17038.1	0.0000	0.0000	FP1A	GL_10863926-S	0003610148
GL_11995467-S	18687.3	23866.6	0.0000	0.0000	IFTM3	GL_11995467-S	0003520025
GL_13376528-S	11377.4	8520.6	0.0000	0.0000	FLJ13621	GL_13376528-S	0003360609
GL_13904867-S	7962.8	10611.0	0.0000	0.0000	RPS10	GL_13904867-S	0006200609
GL_13904869-S	15723.1	15192.0	0.0000	0.0000	RPS5	GL_13904869-S	0007040315
GL_14141191-S	15031.2	17269.8	0.0000	0.0000	RPS14	GL_14141191-S	0001170064
GL_14277699-S	7066.7	8947.4	0.0000	0.0000	RPS12	GL_14277699-S	0006510594
GL_14670385-S	7518.9	13395.5	0.0000	0.0000	RPS21	GL_14670385-S	0003450129

Figure 38 Marked Data Shown in a Web Browser

## Saving Marked Data in a Text File

To save marked data in a text file, do the following:

1. In the scatter plot context menu, select **Marked List | Save in Text File**.

The **Save Marked Genes List As** screen appears (Figure 39).

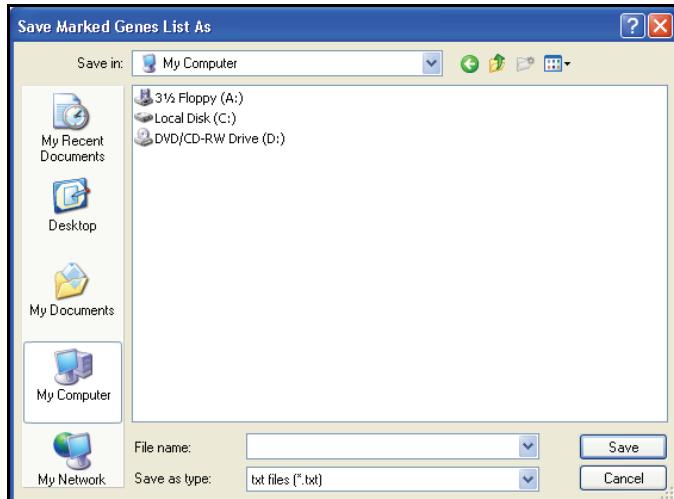


Figure 39 Save Marked Genes List As

2. Browse to the location where you want to save your file.
3. Type a name for your marked genes list in the **File name** text field.

4. Click **Save**.

The **BeadStudio Scatter Plot Output Data** dialog box appears (Figure 40).

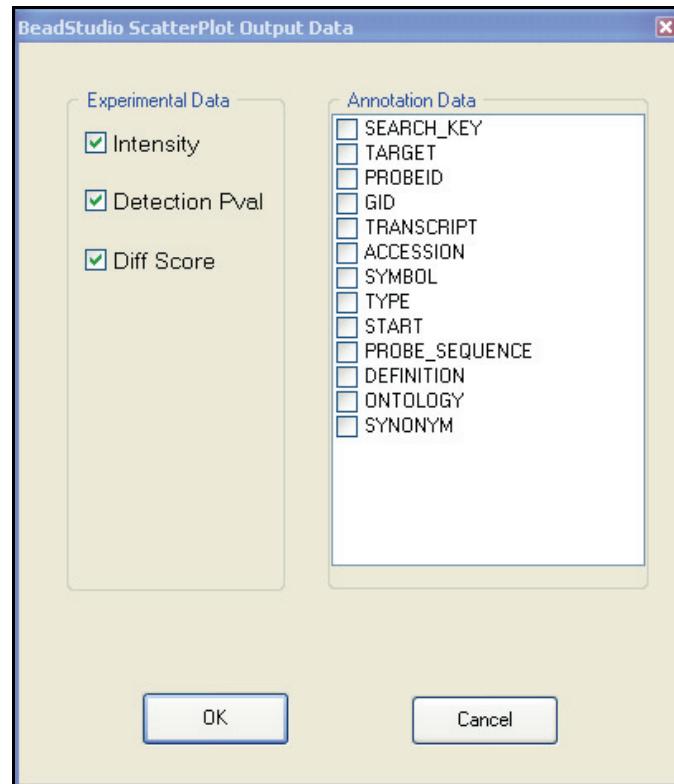


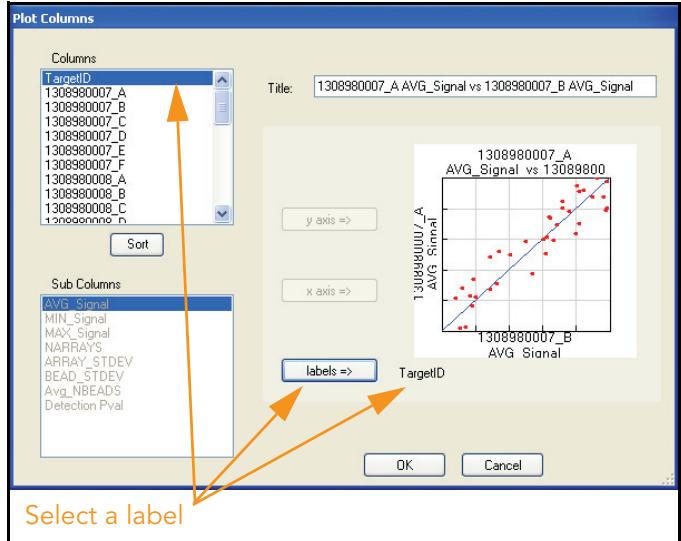
Figure 40 BeadStudio Scatter Plot Output Data Dialog Box

5. In the **Experimental Data** and **Annotation Data** areas, double-click selections for the data you want to include in the web browser.
6. Click **OK**.
7. The data is saved in a text file in the location you specified (Figure 41).

## Figure 41 Saving Marked Data in a Text File

## Showing Item Labels in a Scatter Plot

You can show item labels in a scatter plot if you assigned a label when you created the scatter plot (Figure 42).



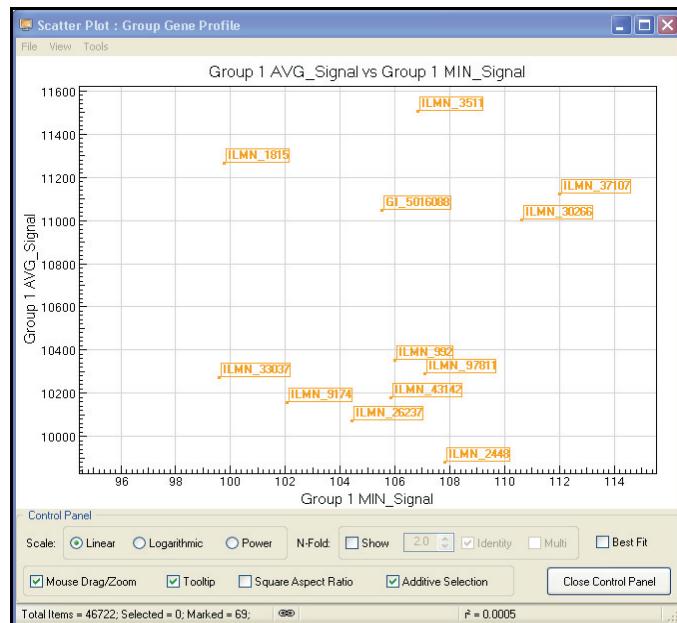
## Figure 42 Selecting a Label for a Scatter Plot

For more information about creating a scatter plot, see the Scatter Plots section of the BeadStudio Framework User Guide, Part # 11204578.

To show item labels in the scatter plot, do the following:

1. In the scatter plot context menu, select **Marked List | Show Item Labels**.

The item labels are displayed in the scatter plot (Figure 43).



**Figure 43** Showing Item Labels for Marked Data in a Scatter Plot

If the labels are too close together to read, you can change their appearance in the scatter plot.

- Using the mouse wheel, scroll up or down to change the resolution of the scatter plot.
- Click and drag to move the scatter plot.
- Shift-click to zoom into a particular region of the scatter plot.
- Control-click, hold, and move the mouse to select a specific gene or group of genes.

## Other Scatter Plot Functions

## Bar Plots

Once gene analysis or differential analysis is complete, you can create bar plots using BeadStudio data tables.

To create a bar plot:

1. Click  **Show Bar Plot**.

The bar plot appears.



**Figure 44** Bar Plot of Sample Probe Profile



### NOTE

If you want to view the same data in a line plot, select **Line Plot**.

2. In the **Data** dropdown menu (Figure 44), select the type of data to plot.



### NOTE

The graph only displays data that are shown in the table. To change which columns are displayed in the table, use the **Column Chooser** tool described in the BeadStudio Framework User Guide, Part # 11204578.

3. Right-click and select **Properties** from the context menu.  
The **Plot Settings** dialog box appears (Figure 45).

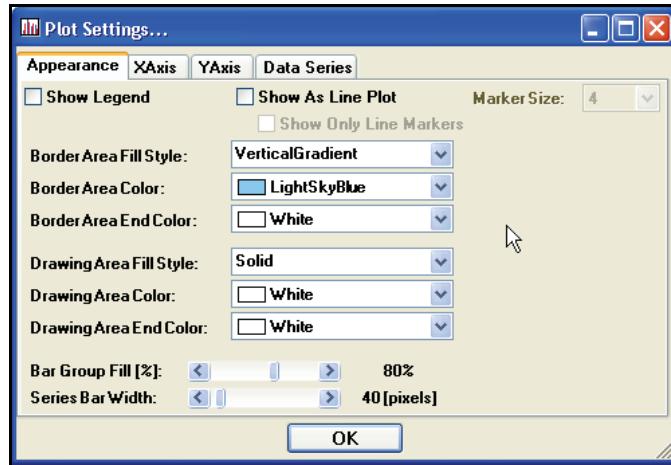
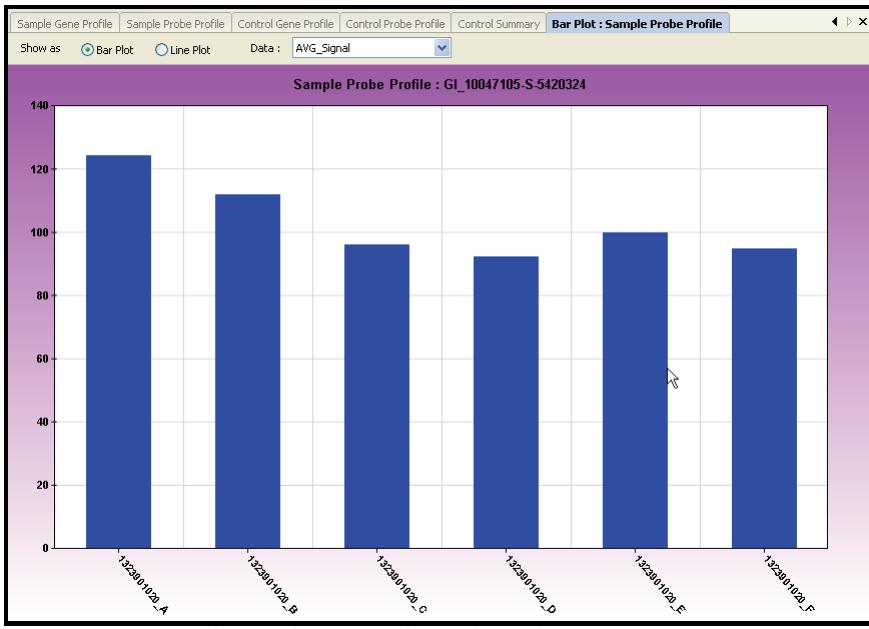


Figure 45 Plot Settings Dialog Box

4. Select attributes for the following aspects of the bar plot:
  - **Appearance**
  - **X-axis**
  - **Y-axis**
  - **Data series**
5. Click **OK**.

The bar plot appears with the attributes you chose (Figure 46).



**Figure 46** Bar Plot With User-Selected Attributes

### Bar Plot Context Menu

Right-click anywhere in the bar plot to view the context menu (Figure 46). The context menu contains features that can be applied to the selected project.



**Figure 47** Bar Plot Context Menu

Table 4 lists context menu items and their functions.

**Table 4** Bar Plot Context Menu Item Descriptions

Item	Description
Properties	Displays the <b>Plot Settings</b> dialog box, which allows you to change the characteristics of the bar plot.
Clear Selected Values	Clears the selected values.
Copy As	Copies the bar plot image to the clipboard in any of the following file formats: BMP, JPEG, PNG, GIF, or TIFF.

## Heat Maps

Once gene analysis or differential analysis has been completed, you can create heat maps using BeadStudio output files.

To create a heat map:

1. Click  **Column Chooser**.  
The **Column Chooser** dialog box appears.
2. Select rows and columns from the data table that you want to display in a heat map.
3. Click **OK**.  
The **Column Chooser** dialog box closes.
4. In the table toolbar, click  **Heat Map**.  
The **Plot Sample Subcolumns in a Heat Map** dialog box appears.

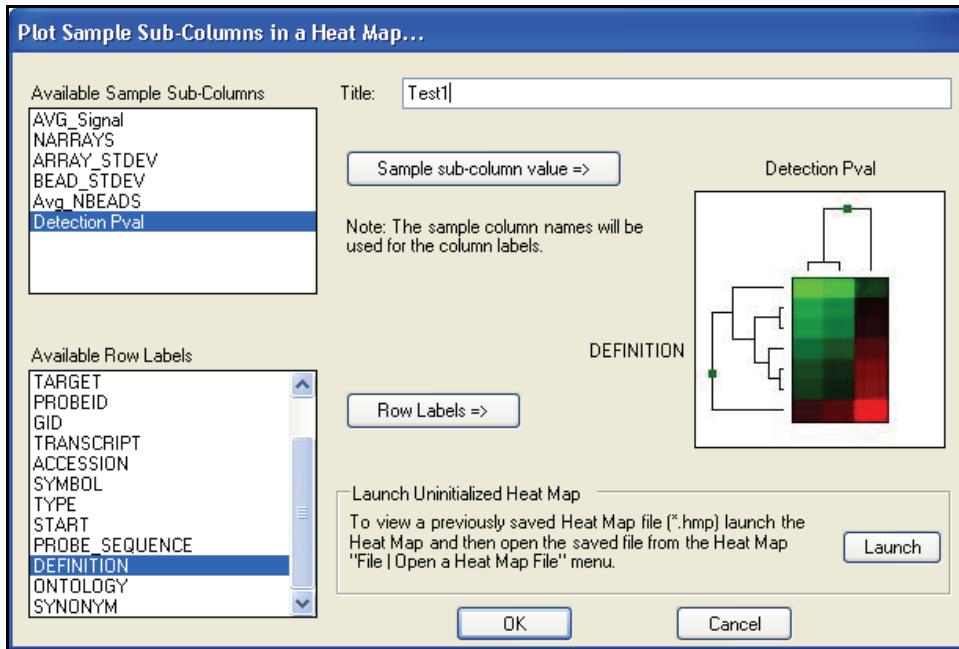


Figure 48 Creating a Heat Map

5. [Optional] Type a title in the **Title** text field.
6. Select an attribute in the **Available Sample Subcolumns** area.



Available sample subcolumns must contain plottable (numerical) data.

7. Select an attribute in the **Available Row Labels** area.
8. Click **OK** to create and display the heat map (Figure 49).

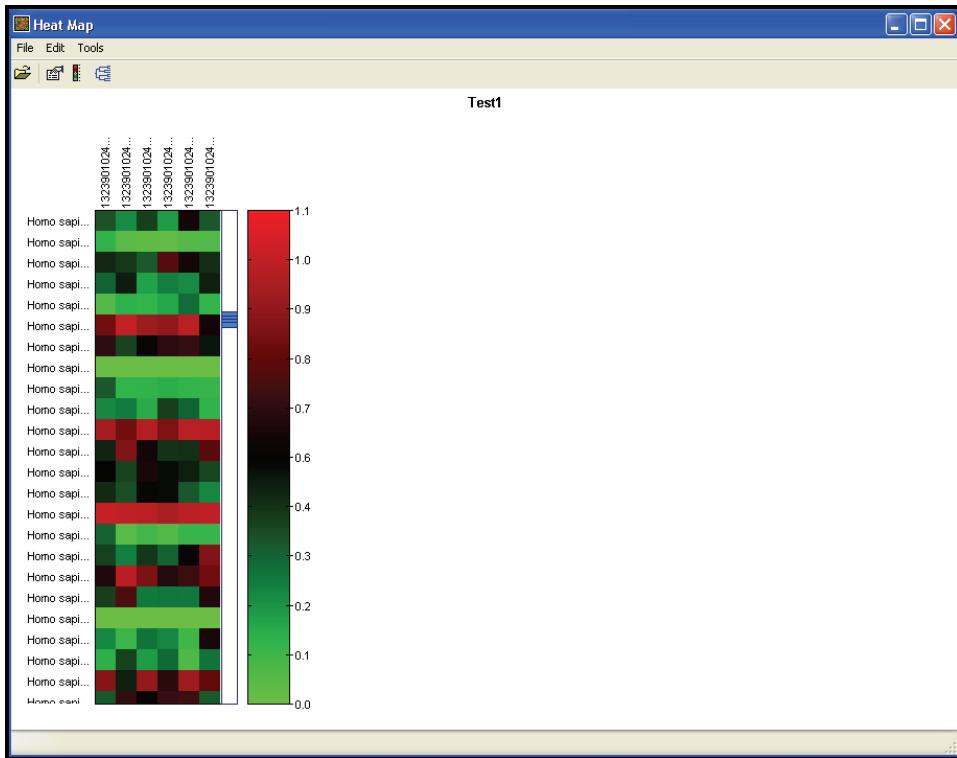


Figure 49 Heat Map

### Heat Map Tools Menu

9. To use additional heat map tools, click **Tools** on the menu bar.
10. Select **Cluster** or **Presentation Image Generator** (Figure 50).

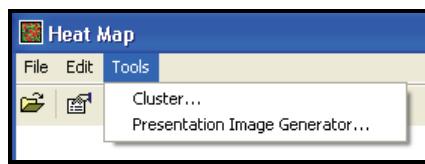


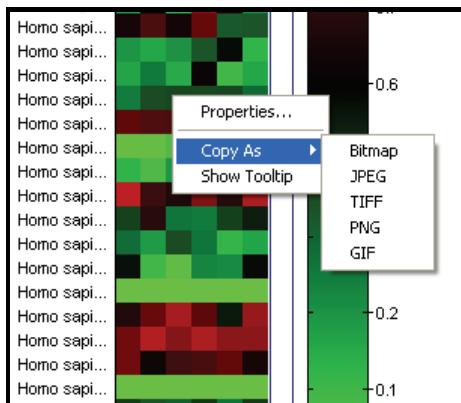
Figure 50 Heat Map Tools Menu

Table 5 describes the available heat map tools.

**Table 5 Heat Map Tools Menu Item Descriptions**

Tool Name	Description
Cluster	Displays the <b>Cluster Options</b> window, from which you can select whether to cluster rows or columns, as well as the hierarchical clustering method to use (COR, ACOR, Manhattan, or Euclidian). See <i>Cluster Analysis Dendograms</i> on page 68 for more information about clustering methods.
Presentation Image Generator	Displays the <b>Presentation Image Setup</b> window, from which you can set options for and generate a presentation image.

**Heat Map Context Menu** Right-click anywhere in the heat map to view the context menu (Figure 51). The context menu contains options that can be applied to the selected heat map.



**Figure 51 Heat Map Context Menu**

Table 6 lists describes map context menu items.

**Table 6** Heat Map Context Menu Item Descriptions

Item	Description
Properties	Displays the <b>Heat Map Properties</b> dialog box, from which you can alter the visual properties of the title, legend, rows, columns, and scroll bars of the heat map.
Copy As	Copies the heat map to the clipboard as any of the following image types: BMP, JPEG, TIFF, PNG, or GIF.
Show Tooltip	Displays a tooltip when the cursor is positioned over the heat map.

For more information about working with heat maps, see the *Heat Maps* section of the *BeadStudio Framework User Guide*, Part # 11204578.

## Cluster Analysis Dendrograms

Clustering is an analysis method used to group sets of objects into subsets or clusters. Objects assigned to the same cluster are more closely related to one another than to objects assigned to different clusters. In the context of gene expression, clustering can be used to answer two basic questions:

- ▶ Which genes show similar patterns of gene expression across a series of samples?

Knowing this is useful for identifying genes in common pathways, or genes that coordinately respond to the same stimuli.

- ▶ Which samples are most similar based on the expression levels of genes within them?

Knowing this is useful for identifying conditions that generate a common metabolic response. For example, in a toxicology study, if an unknown compound induces a pattern of expression similar to that induced by a panel of genotoxins, it is likely that the unknown is a genotoxin.

Mathematicians have devised dozens of clustering methods using different metrics to establish relationships between sets of values. In BeadStudio, clustering occurs using the nesting with average linkage method. BeadStudio offers four clustering metrics for calculating dissimilarities:

- ▶ **Correlation (COR)**—Computes the Pearson correlation using a  $1-r$  distance measure.
- ▶ **Absolute Correlation (ACOR)**—Computes the Pearson correlation using a  $1-|r|$  distance measure.
- ▶ **Manhattan**—Computes the distance between two points if a grid-like path is followed.
- ▶ **Euclidian**—Computes the shortest distance between two points.



#### NOTE

Illumina recommends using multiple clustering methods to validate results. Groupings with a true biological basis will usually show up regardless of the algorithm used.

## Similarities and Distances

There are several ways to compute the similarity of two series of numbers. The most commonly used similarity metric is the Pearson correlation. The Pearson correlation coefficient between any two series of numbers  $X = \{x_1, x_2, \dots, x_N\}$  and  $Y = \{y_1, y_2, \dots, y_N\}$  is defined as:

$$r = \frac{1}{N} \sum i = 1, N \left( \frac{x_i - \bar{x}}{\sigma_x} \right) \left( \frac{y_i - \bar{y}}{\sigma_y} \right)$$

Distance is then defined as  $1 - r$  for Correlation and  $1 - |r|$  for Absolute Correlation. BeadStudio also uses Manhattan ( $\Sigma |X_i - Y_i|$ ) and squared Euclidean ( $\Sigma (X_i - Y_i)^2$ ) distances.

BeadStudio presents the clustering information in the form of a dendrogram, a tree-like structure with branches that correspond to genes or samples, depending on how the analysis is run. The distance on the X axis establishes the similarity relationships among the genes or samples. For example, if the dendrogram plots the similarity of samples based on gene expression, samples C and D are very similar to each other, less similar to B, and even less similar to A (Figure 52).

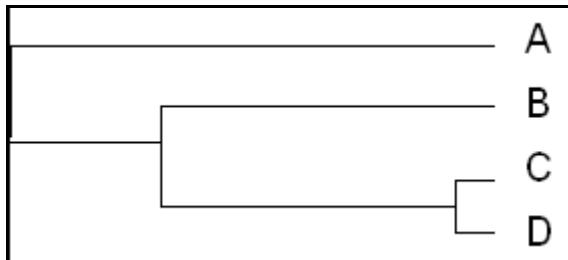


Figure 52 Dendrogram Similarity Example

After clustering, nodes are reordered starting near the top to ensure that node "ar" is closer to "B" than node "al", and node "bl" is closer to "A" than node "br" (Figure 53).

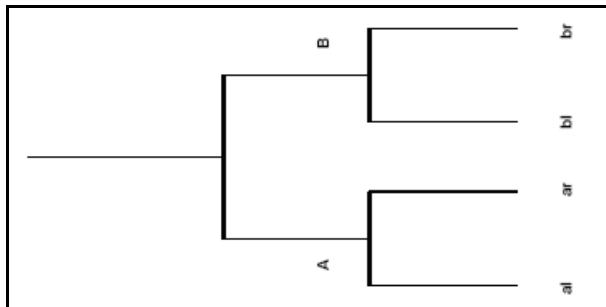


Figure 53 Dendrogram, Showing Nodes

## Analyze Clusters

To analyze clusters:

1. Click Run Cluster Analysis to open the cluster analysis tool.
2. In the **Cluster Analysis** dialog box (Figure 54), perform the following steps:
  - a. **Groups pane**—Highlight the group(s) whose clusters you wish to analyze. Select the **Sort** checkbox to sort the items in the Groups listbox alphabetically in ascending order.
  - b. **Cluster pane**—Click **Genes** or **Samples**.  
If you select **Genes**, the dendrogram displays a cluster of genes.  
If you select **Samples**, the dendrogram displays a cluster of samples.

**NOTE**

Clustering samples is much faster than clustering genes. Clustering thousands of genes can take hours.

- c. **Metric pane**—Select the metric you would like BeadStudio to use to calculate clusters.

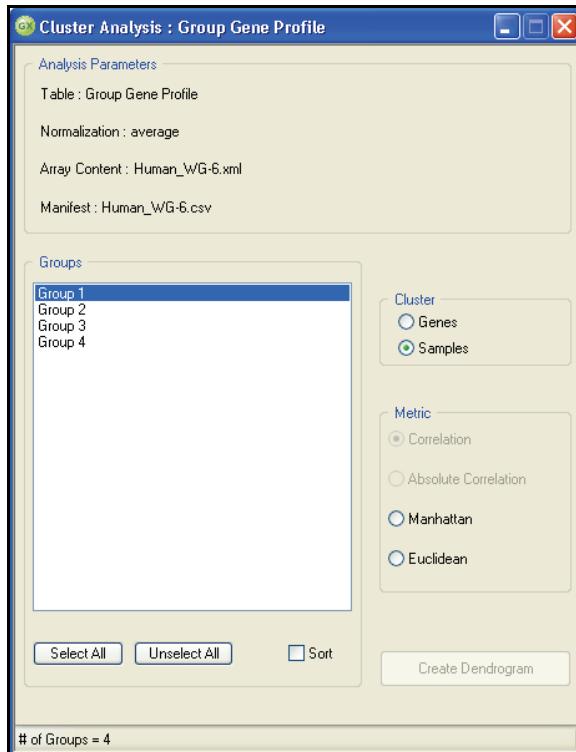


Figure 54 Cluster Analysis Dialog Box

- d. Click **Create Dendrogram** to view the graph (Figure 55).

**NOTE**

The scale at the bottom of the dendrogram shows dissimilarity between nodes. See *Similarities and Distances* on page 69.

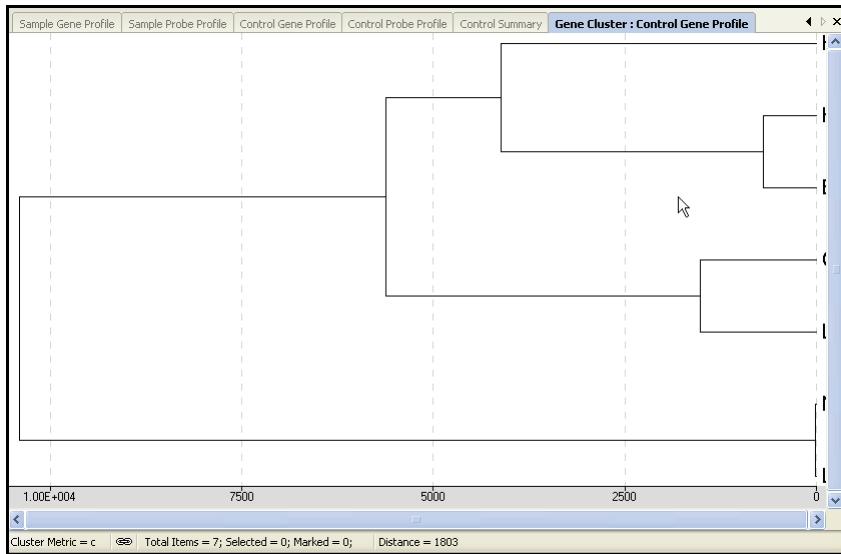


Figure 55 Dendrogram

3. Right-click in the dendrogram to view the context menu (Figure 56).

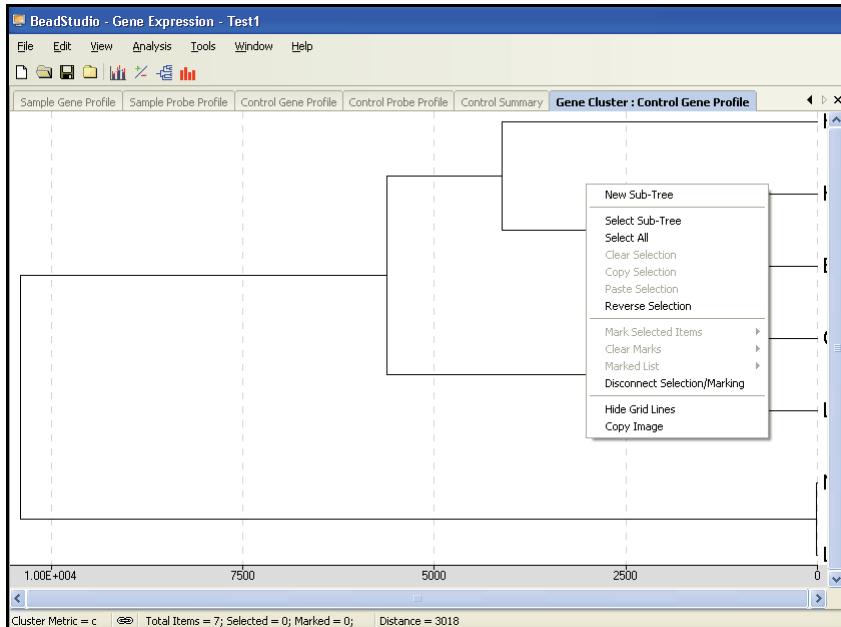


Figure 56 Dendrogram with Context Menu

## Dendrogram Context Menu Selections

Table 7 describes dendrogram context menu selections.

**Table 7** Dendrogram Context Menu Selections

Item	Description
New Sub-Tree	Displays the selected sub-tree in a new window. This feature is disabled when the cursor is outside of any tree.
Select Sub-Tree	Highlights the sub-tree in blue. This feature is disabled when the cursor is outside of any tree.
Select All	Selects all sub-trees.
Clear Selection	Clears any selection.
Copy Selection	Copies the current selection(s) to the clipboard.
Paste Selection	Pastes the current clipboard contents to a location you choose.
Reverse Selection	Reverses the last selection made.
Mark Selected Items	Marks the currently-selected items.
Clear Marks	Clears all marks.
Marked List	<p>Includes operations you can perform on genes you mark in the scatter plot:</p> <ul style="list-style-type: none"> <li>▶ <b>View in Web Browser</b>—Displays a list of the marked genes in a web browser.</li> <li>▶ <b>Save in Text File</b>—Allows you to save genes in a file in a location you specify.</li> <li>▶ <b>Show Item Symbols</b>—Shows item symbols.</li> </ul>
Disconnect Selection/Marking	Disconnects synchronization between the graph and the table.
Hide Grid Lines	Hides background grid lines.
Copy Image	Copies the current image to the clipboard.

## View the Sub-Tree List Directly in the Dendrogram

- ▶ To view the sub-tree list directly in the dendrogram, zoom in by using the mouse wheel. The sub-tree list appears to the right of the dendrogram (Figure 57).
- ▶ To resize the dendrogram, press **Ctrl** and the right or left arrow keys on your keyboard. The scale adjusts appropriately.
- ▶ To return the dendrogram to its default size, click the mouse button.

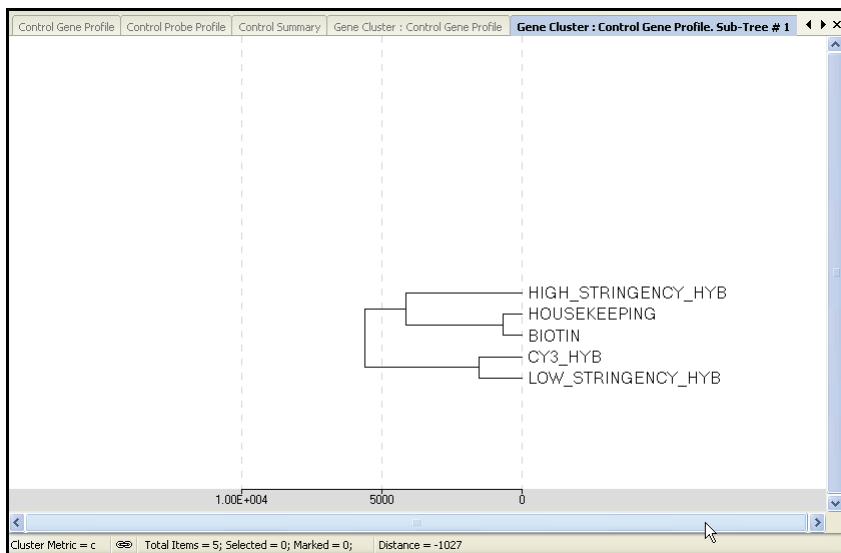


Figure 57 Zooming In to View a Sub-Tree List

## Copy/Paste Clusters

You can copy and paste gene clusters from a scatter plot to a dendrogram and vice versa. Refer to Figure 58 through Figure 61.

### From Scatter Plot to Dendrogram

To select genes within clusters that you want to copy from a scatter plot and paste into a dendrogram:

1. Select **Tools | Select Region**.
2. Using the crosshair tool, draw around the genes you wish to copy.

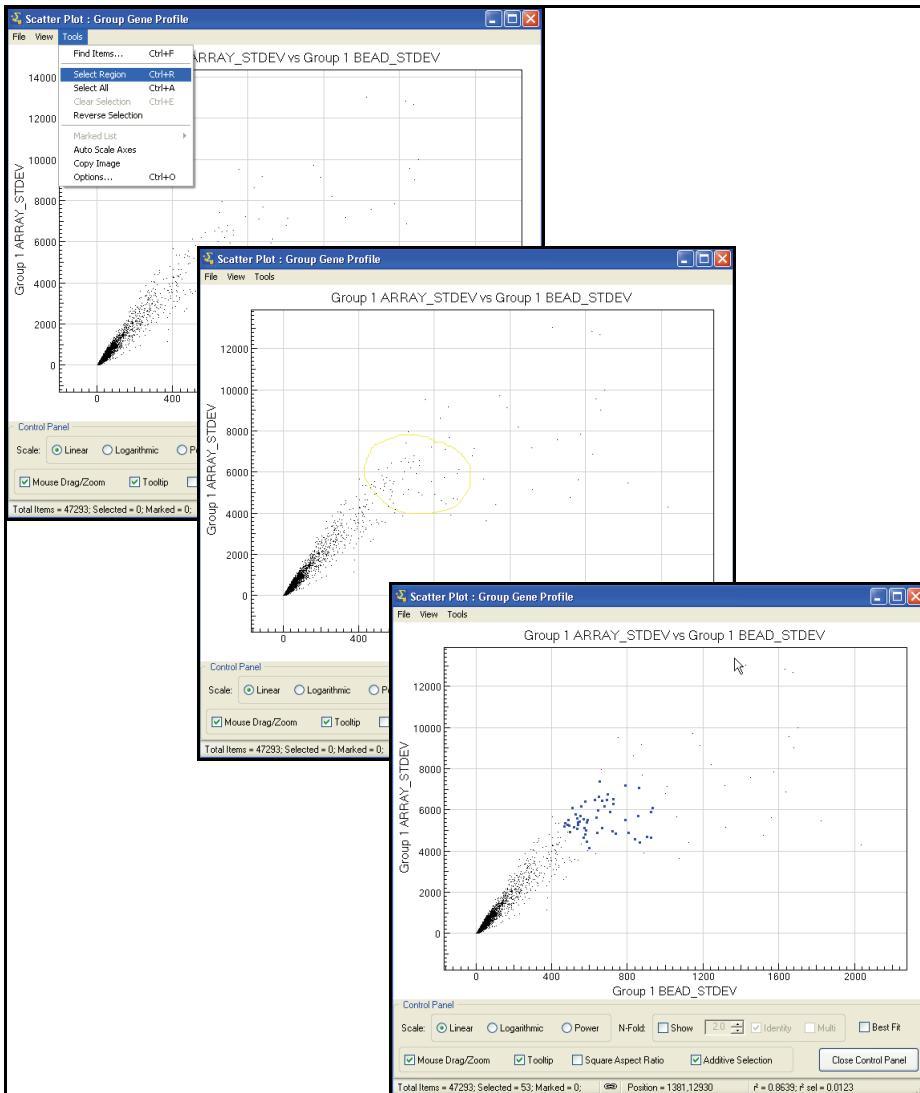
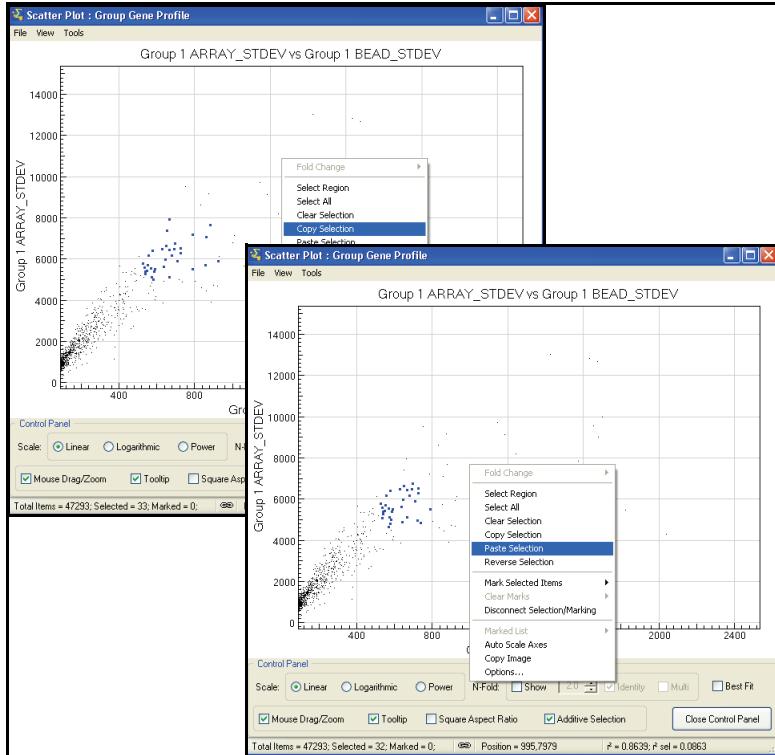


Figure 58 Selecting a Region



The selected genes will change color: to blue by default, or to the color you set in Scatter Plot Options.

3. To copy the selection to the clipboard, select **Tools | Copy Selection**.
4. To paste the selection into the dendrogram, select **Tools | Paste Selection**.



**Figure 59** Copying & Pasting

### From Dendrogram to Scatter Plot

To select clusters for copying from the dendrogram:

1. Position the cursor over the sub-tree you want to copy.
2. Right-click and click **Select Sub-Tree** from the context menu.



Click **inside** the sub-tree you want to select. The sub-tree you select appears in blue.

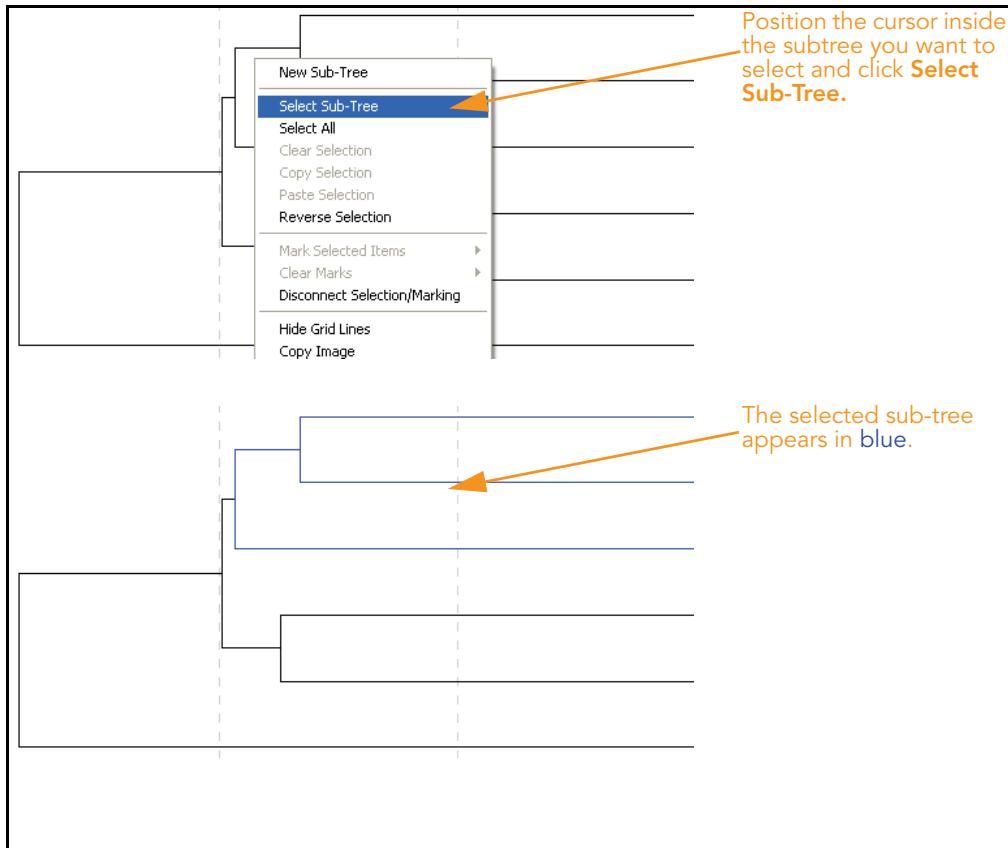


Figure 60 Selecting a Sub-Tree

3. Right-click and select **Copy Selection** from the context menu.

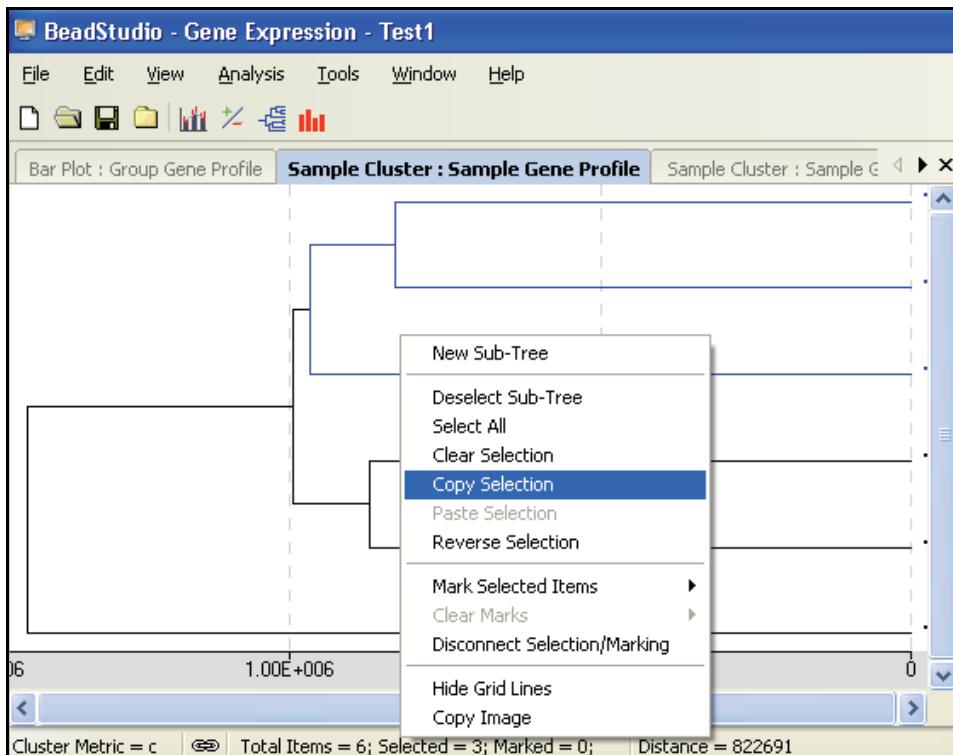


Figure 61 Copying a Sub-Tree

4. To paste your selection into a scatter plot, select **Tools | Paste Selection**.

The genes you selected in a subtree are pasted from the dendrogram into the scatter plot.

## Control Summary Reports

The following sections describe the Control Summary Reports for the Direct Hyb assay and the DASL assay.

- |  |  |
|--|--|
| <b>For the<br/>DirectHyb<br/>Assay</b> | BeadStudio displays a graphic Control Summary Report for selected samples based on the performance of the built-in controls (Figure 62). |
|--|--|

For more detailed information about the controls, see the *System Controls* appendix in the appropriate Illumina product guide.

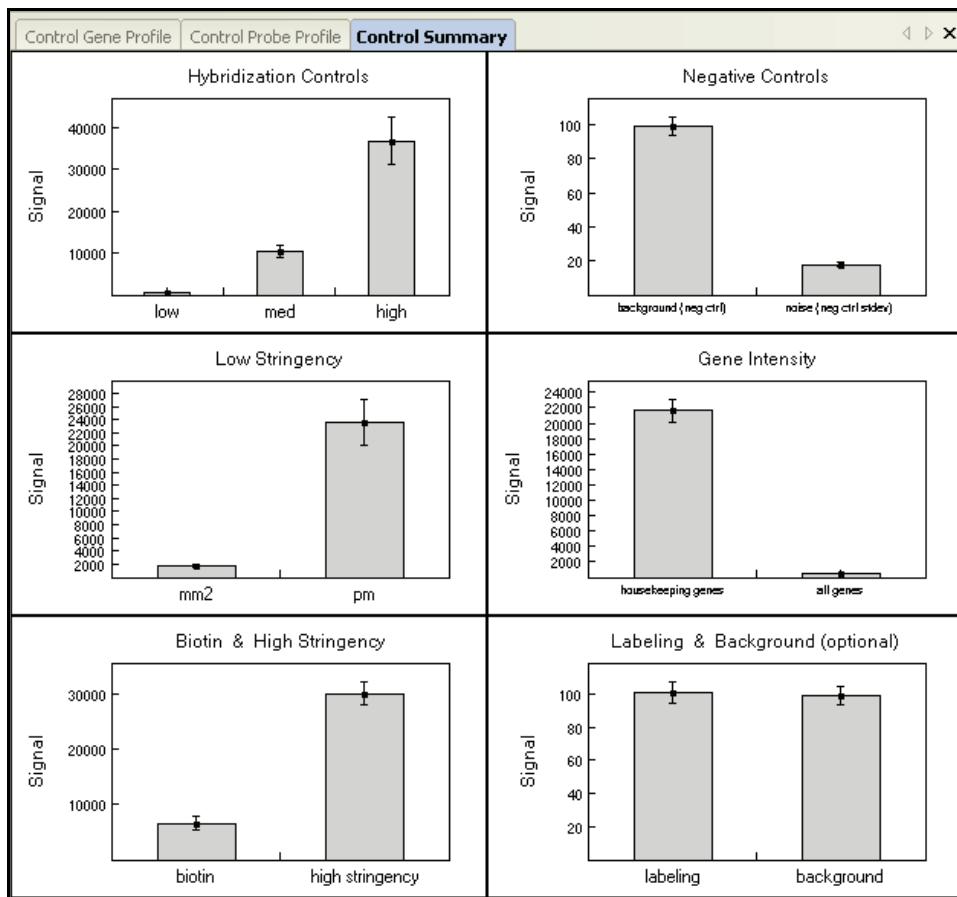


Figure 62 Control Summary Report

- To view secondary graph(s), click on a data point in any of the graphs shown in Figure 62. Each point in the secondary graph represents a sample.

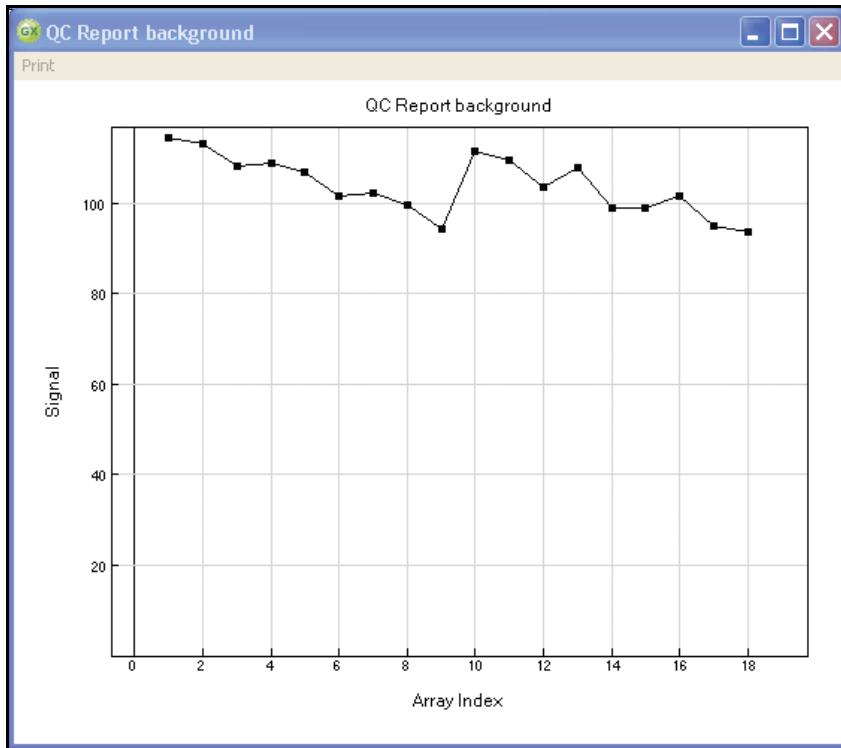


Figure 63 Housekeeping Controls Secondary Graph

6. To copy, change the page setup, or see a print preview, right click in any graph to use the context menu.
7. To print the graph, click **Print** in the menu bar (Figure 64).

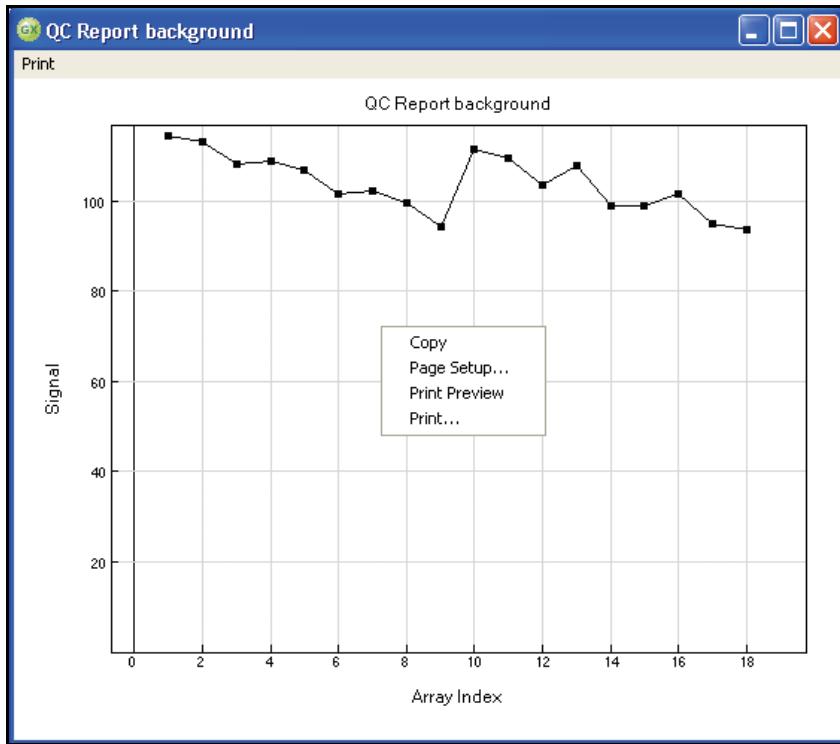


Figure 64 Control Summary Context Menu

### For the DASL Assay

BeadStudio can display a graphic Control Summary for the selected samples based on the performance of the built-in controls (Figure 65).

For more detailed information on the controls, see the *System Controls* appendix in the appropriate Illumina product guide.

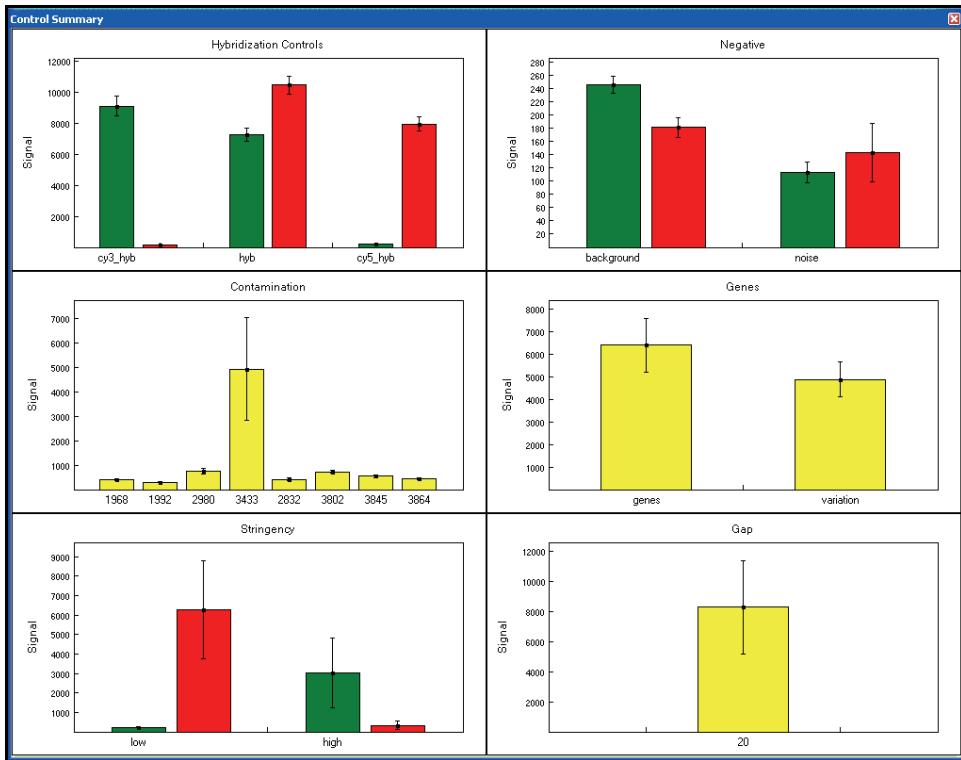


Figure 65 Control Summary Report

8. To view secondary graph(s), click on a data point in any of the graphs shown in Figure 65. Each point in the secondary graph represents a sample.

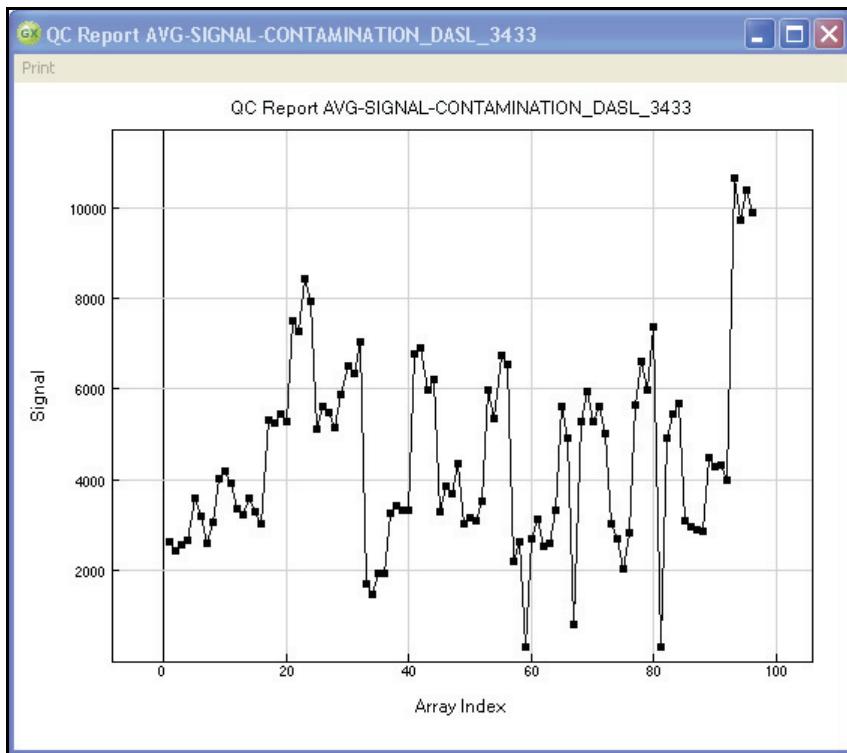


Figure 66 Contamination Controls Secondary Graph

9. To copy, change the page setup, or see a print preview, right click in any graph to use the context menu (Figure 67).
10. To print the graph, click **Print** in the menu bar.

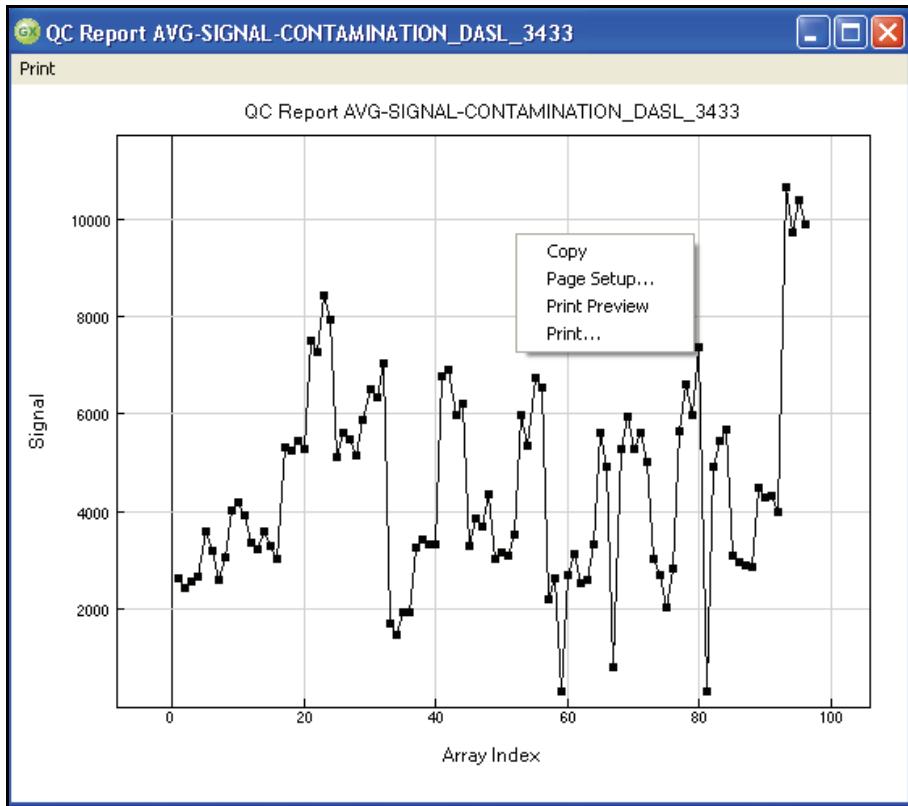


Figure 67 Control Summary Context Menu

## Image Viewer

You can visually inspect any sample using the BeadStudio Image Viewer. The Image Viewer allows you to see images for the purpose of determining whether or not you want to include a particular sample or samples in your experiment.

Using the Image Viewer, you can do the following things:

- ▶ See registration information for individual samples
- ▶ See registration for BeadStudio-processed images
- ▶ Adjust the contrast of an image
- ▶ Zoom in or out
- ▶ See pixel intensities

## Selecting an Image to View

To select an image to view, do the following:

1. Go to **Analysis | View Image**.

The **BeadStudio View Image** dialog box appears (Figure 68).

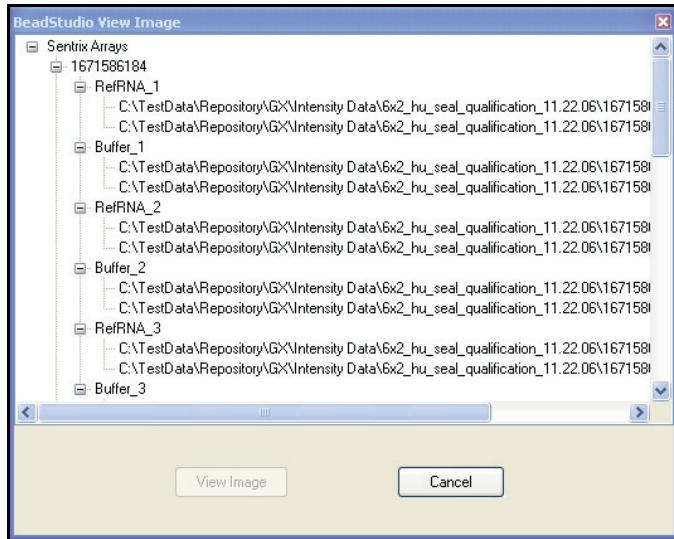


Figure 68 BeadStudio View Image

2. Click to select an image from the list of available images.
3. Click **View Image**.

The **Image Viewer** window appears (Figure 69).

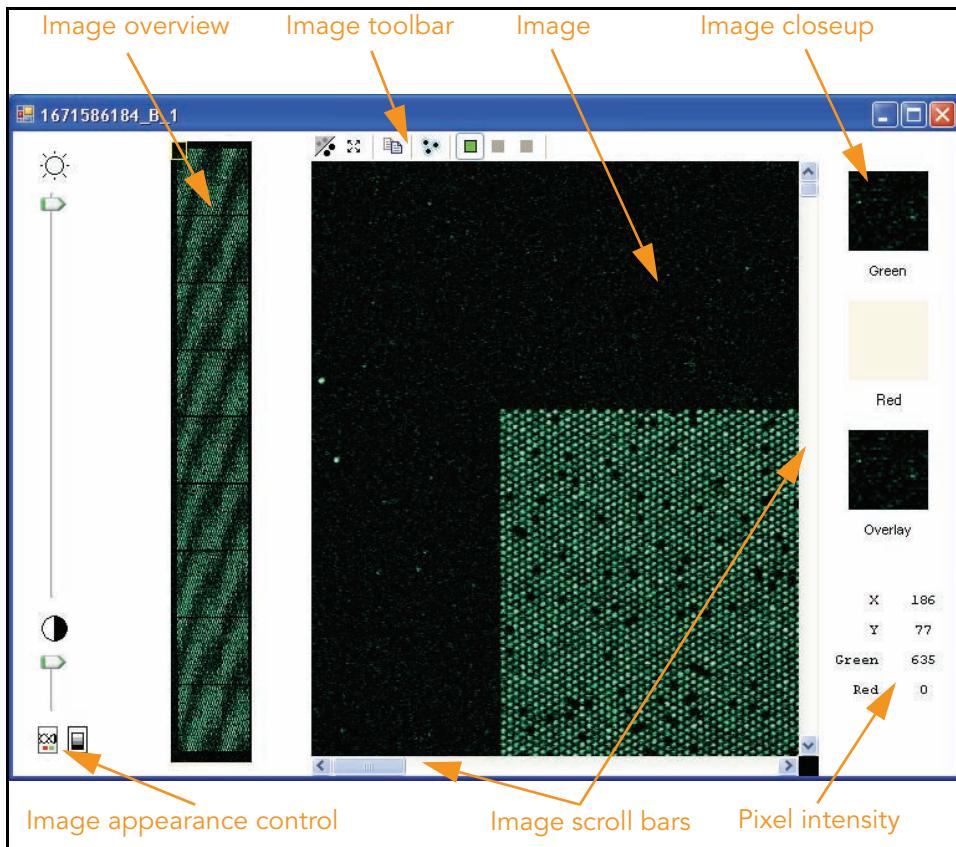


Figure 69 Image Viewer

Table 8 describes the features of the image viewer.

Table 8 Image Viewer Features

Feature	Description
Image overview pane	Displays a high-level view of the sample image. Use the mouse to move the yellow box, which determines the field of view.
Image toolbar	Click <b>Auto Contrast</b> to reset the image contrast to the default value.

**Table 8** Image Viewer Features (continued)

Feature	Description
Image toolbar (cont.)	 Click <b>Zoom Out</b> to return to the default image view.
	 Click <b>Copy to Clipboard</b> to save an image to the clipboard for pasting into another application.
	 Click <b>Overlay Cores</b> to verify successful registration during data extraction on the BeadArray Reader. See Overlay Cores on page x-x for details.
	 Click <b>Show Green Channel</b> to see the green channel only.
	 Click <b>Show Red Channel</b> to see the red channel only. (This option is disabled for monochrome Direct-Hyb images.)
	 Click <b>Overlay Channels</b> to see both green and red channels. (This option is disabled for monochrome Direct-Hyb images.)
Image pane	Allows detailed inspection of the sample image. Use the mouse wheel to control the zoom level. If your mouse lacks a wheel, zoom in to a region by pressing the <b>Shift</b> key and the left mouse button at the same time, then dragging to select the zoom area, then releasing. To zoom out, click <b>Zoom Out</b> on the Image toolbar.
Image closeup pane	Displays a closeup view of your image in the red and green color channels, and in a merged channel (overlay) view. The view region is determined by the location of your cursor on the image. Note: the red color channel is disabled for monochrome direct hyb images.
Image appearance control	Use these controls to change image brightness, contrast, and color balance. Note: these controls affect only the appearance of the image on the screen. They do not change the underlying image file.
Image scroll bars	Allow you to change the viewing region in the image pane.
Pixel intensity	Reports the X and Y coordinates of your mouse pointer on the image pane, along with the pixel intensity at that location.

## Displaying Overlay Cores

Click  **Overlay Cores** to display the image pane as shown in Figure 70. This feature allows you to verify that registration succeeded during data extraction on the BeadArray Reader.



### NOTE

The overlay cores feature is enabled only when viewing a single channel (green or red). This feature is not enabled when viewing both channels simultaneously.

Zoom in on a corner of the image to see blue circles overlaying the scanned image sample spots. Successful registration is indicated when the boundary of the blue-circle grid coincides with the sample pixel boundary.

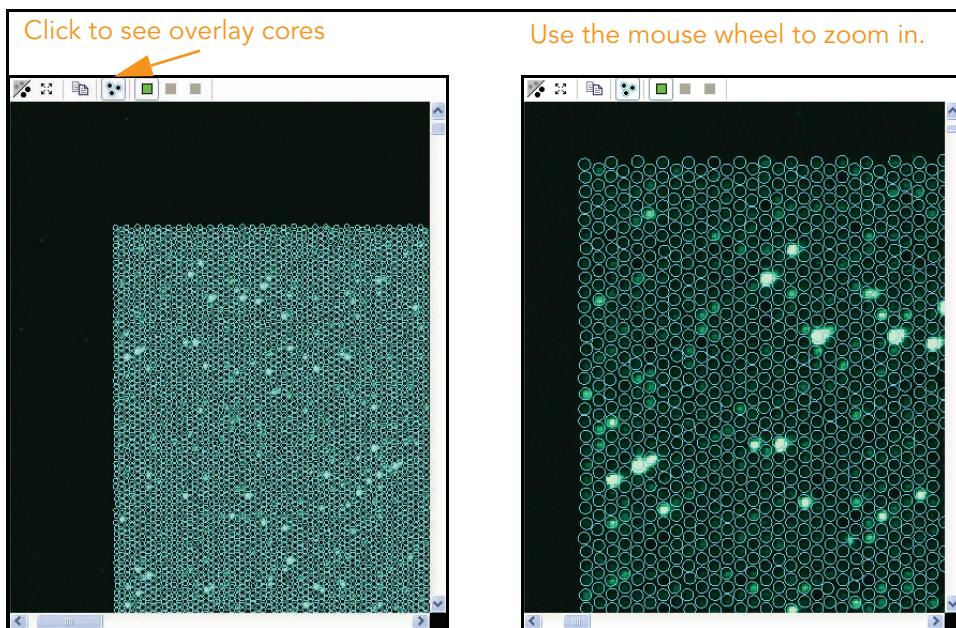


Figure 70 Overlay Cores



### NOTE

In rare cases, registration may fail. If this occurs, contact Illumina Customer Solutions.

## Changing Image Appearance

You can change the appearance of your images in two ways:

- ▶ **Brightness/Contrast** mode
- ▶ **Intensity Threshold** mode

Use the selection buttons at the bottom of the pane to select the mode you want to change.

The components of the image appearance pane are described in Figure 71.

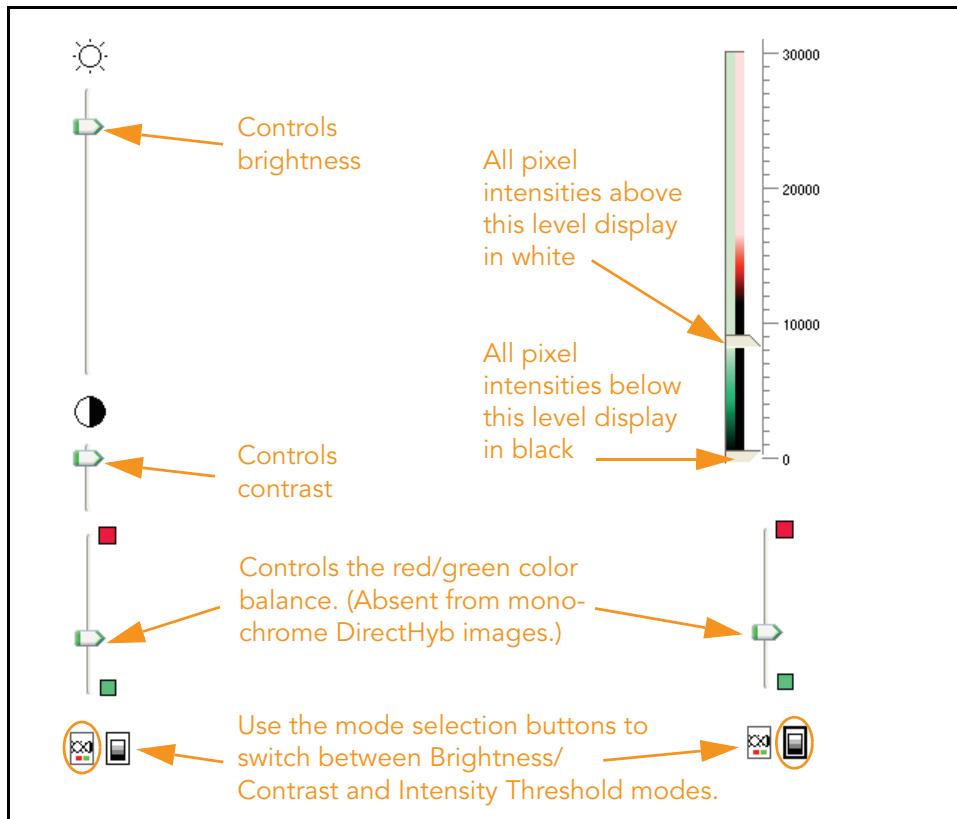


Figure 71 Image Control Pane



# Chapter 4

## Normalization and Differential Analysis

### Topics

- 92      Introduction
- 92      Normalization Methods & Algorithms
  - 93      Background
  - 93      Average
  - 93      Rank Invariant
  - 94      Cubic Spline
- 95      Differential Expression Algorithms
  - 95      Illumina Custom
  - 98      Mann-Whitney
  - 98      T-Test
  - 98      Detection P-Value
  - 98      Whole Genome BeadChips
  - 99      Focused Arrays
  - 99      DASL Products

## Introduction

This chapter describes the statistical algorithms used in BeadStudio gene expression analysis.

## Normalization Methods & Algorithms

Normalization algorithms adjust sample signals in order to minimize the effects of variation arising from non-biological factors. The BeadStudio Gene Expression Module offers four routines that are described in the following sections.

For all algorithms, normalization is computed with respect to a mathematically calculated “virtual” sample that represents averaged probe intensities across a group of samples.

In the cases of cubic spline and rank invariant normalizations, the virtual sample is computed based on the content of the reference group. If there is no reference group, the first group in the list of groups displayed in the **Project Group** pane is used for group analysis.

For SAM/BeadChip analysis, the virtual sample is computed based on the content of the first alphanumeric entry in the upper-left area of the **Matrix** pane. For the average methods, all samples in the experiment are averaged to produce the virtual sample.

The following sections contain detailed description of these normalization algorithms:

- ▶ Background
- ▶ Average
- ▶ Rank Invariant
- ▶ Cubic Spline

**Background**

Background is defined as the average signal intensity estimated from the negative control bead types. Outliers are removed using the median absolute deviation method.

Background is capable of minimizing the amount of variation in background signals between arrays. Background subtraction can be performed to remove signal due to nonspecific hybridization.

As a result, the expected signal for unexpressed targets is equal to zero. Half of the unexpressed targets are expected to have negative signals.

**Average**

Sample intensities are scaled by a factor equal to the ratio of average intensity of virtual sample to the average intensity of the given sample. Background is subtracted prior to the scaling.

Average normalization adjusts for differences in overall intensity between arrays and chips. Sample intensities are scaled by a factor so that the average signal of all samples becomes equal to the global average of all sample signals. Background subtraction is done prior to scaling, so half of the unexpressed targets are expected to have negative signals.

**Rank Invariant**

This method uses a rank invariant set of probes between a given sample and a virtual sample.

The rank invariant set is found as follows: we start by considering probes with intensities ranked between LowRank = 50<sup>th</sup> percentile and HighRank = 90<sup>th</sup> percentile. If the probe's

relative rank changes  $\frac{r_x - r_v}{r_v} \leq 0.05$  the probe is considered to be

rank invariant. If less than 2% of all probes in the region are identified as rank invariant, LowRank is gradually decreased until it reaches the 25<sup>th</sup> percentile.

Rank invariant normalization operates under the assumption that probes with similar ranking between samples have similar expression levels. This method minimizes the effects of additive and multiplicative factors. The subset of probes with low relative rank change is defined as follows.

1. The virtual reference sample is created by averaging the content of the reference group (in differential analysis) or the first group of the group set (in gene analysis).
2. All probes ranked between LowRank = 50<sup>th</sup> and HighRank = 90<sup>th</sup> percentiles are considered. If the change of rank relative to the virtual reference is less than 0.05, the probe is considered to be "rank invariant."
3. If less than 2% of all probes are picked as rank invariant, LowRank is gradually decreased until it reaches the 25<sup>th</sup> percentile.
4. Normalization coefficients are computed using iteratively re-weighted least squares. This method is applied after the background normalization method described above.

**Cubic Spline**

The method is similar to the one proposed by Workman et al.<sup>1</sup> The normalization uses quantiles of sample intensities to fit smoothing B-splines.

$$\text{Let } q_i = \frac{i - 0.5}{N}, i = 1, 2 \dots N$$

be a vector of N quantiles  $\left( N = \max\left(15, \frac{N_{\text{probes}}}{100}\right)\right)$

Here,  $N_{\text{probes}}$  is the number of probes represented on a sample.

For each sample, the vector of quantile intensities is computed. Similarly, quantiles for the "virtual" averaged sample after background subtraction are computed. Cubic B-spline is computed and used for interpolation. For points with intensities ranked outside the interval, linear extrapolation rather than cubic spline is used to avoid nonlinear effects outside the region of interpolation.

---

1. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild HH, Nielsen C, Brunak S, Knudsen S. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* 2002 Aug 30; 3(9):research0048. PMID: 12225587 [PubMed - indexed for MEDLINE]

Cubic spline normalization is capable of minimizing effects that cause nonlinear transformation of data (e.g., saturation). This is similar to the method proposed by Workman et. al. Just as in the rank invariant method, cubic spline normalization is complete with respect to the virtual reference sample. The assumption of this method is that all samples have similar distributions of transcript abundance. The transformation is computed using smoothing B-splines. The number of quantiles equals 1% of the number of probes, but cannot be lower than 15. For points outside of interpolation interval, linear extrapolation is used. This method is applied after the background normalization method described above.

**NOTE**

For the DASL assay, red and green channels are normalized independently.

## Differential Expression Algorithms

All algorithms compare a group of samples (referred to as the condition group) to a reference group. The comparison is done using the following error models:

- ▶ Illumina Custom
- ▶ Mann-Whitney
- ▶ T-Test

### Illumina Custom

This model assumes that target signal intensity ( $I$ ) is normally distributed among replicates corresponding to some biological condition. The variation has three components: sequence specific biological variation ( $\sigma_{\text{bio}}$ ), nonspecific biological variation ( $\sigma_{\text{neg}}$ ), and technical error ( $\sigma_{\text{tech}}$ ).

$$I = N(\mu, \sigma)$$

$$\sigma = \sqrt{\sigma_{\text{tech}}^2 + \sigma_{\text{neg}}^2 + \sigma_{\text{bio}}^2}$$

$$\sigma_{\text{tech}} = a + b < I >$$

Variation of nonspecific signal  $\sigma_{\text{neg}}$  is estimated from the signal of negative control sequences (using median absolute deviation). For  $\sigma_{\text{tech}}$ , we estimate two sets of parameters  $a_{\text{ref}}, b_{\text{ref}}$  and  $a_{\text{cond}}, b_{\text{cond}}$  for reference and condition groups respectively.

We estimate  $\sigma_{\text{tech}}$  using iterative robust least squares fit, which reduces the influence of highly variable genes. This implicitly assumes that the majority of genes do not have high biological variation among replicates. When this assumption does not hold we overestimate technical error by some averaged biological variation.

When groups contain biological replicates, we produce p-values using the following approach:

$$S_{\text{ref}} = (\max(s_{\text{ref}}, a_{\text{ref}} + b_{\text{ref}} I_{\text{ref}}))$$

$$S_{\text{cond}} = (\max(s_{\text{cond}}, a_{\text{cond}} + b_{\text{cond}} I_{\text{cond}}))$$

$$p = z \left( \frac{|I_{\text{cond}} - I_{\text{ref}}|}{\sqrt{\frac{S_{\text{ref}}^2 + S_{\text{neg}(\text{ref})}^2}{N_{\text{ref}}} + \frac{S_{\text{cond}}^2 + S_{\text{neg}(\text{cond})}^2}{N_{\text{cond}}}}} \right)$$

where  $S_{\text{ref}}$  and  $S_{\text{cond}}$  are standard deviations of probe signals.



#### NOTE

$N_{\text{ref}}$  and  $N_{\text{cond}}$  denote the number of samples in the reference and condition groups respectively.

We consider that standard deviations exceeding  $\sigma_{\text{tech}}$  reflects biological variation. However, we assume that estimates smaller than  $\sigma_{\text{tech}}$  are caused by random errors. Therefore, we use the larger of two estimates. Usage of  $\sigma_{\text{neg}}$  provides regularization for low abundance targets. Z is two-sided tail probability of standard normal distribution.

When reference and conditions groups contain one sample each, we can neither estimate sequence specific biological variation nor sample processing variation. Instead, we can only assess  $\sigma$  using bead type variation. Therefore, we penalize for that by a factor of 2.5 applied to parameter  $b$ . This factor was determined empirically from examination of real sample data.

$$p = z \left( \frac{|I_{cond} - I_{ref}|}{\sqrt{(a_{ref} + 2.5b_{ref}I_{ref})^2 + \sigma_{neg(ref)}^2 + (a_{cond} + 2.5b_{cond}I_{cond})^2 + \sigma_{neg(cond)}^2}} \right)$$

A Diff Score for a probe is computed as:

$$\text{DiffScore} = (10 \operatorname{sgn}(I_{cond} - I_{ref}) \log_{10}(p))$$

For the gene, Diff Scores of corresponding probes are averaged. In addition, concordance between probes is reported.

**NOTE**

For the DASL assay, the red and green channel signals are added together before computing diff scores.

**Mann-Whitney** This implementation produces exact p-value if:

$$\min(N_{\text{ref}}, N_{\text{cond}}) < 3$$

or

$$\max(N_{\text{ref}}, N_{\text{cond}}) < 22$$

Otherwise, normal approximation with continuity correction is used. Differential scores are computed as described for the Illumina Custom model (page 95).

**T-Test** When either the reference group or a condition group contains at least two samples, variance is estimated across replicate samples. Otherwise, variance is estimated from bead-to-bead variation. We use t-test with the assumption of equal variance. Differential scores are computed the same way as described for the Illumina Custom model (page 95).

## Detection P-Value

Illumina microarrays contain negative control probes that are thermodynamically equivalent to the regular probes, but lack specific targets in the transcriptome. Therefore, their signals report the amount of nonspecific cross-hybridization and other sources of noise. Detection p-value can be interpreted as the probability of seeing a certain signal level without specific probe-target hybridization. Thus, the higher the signal level, the lower the probability that it could be due to nonspecific sources.

**Whole Genome BeadChips** Since these products have a large number of negative control bead types ( $> 700$ ), BeadStudio uses a nonparametric method for the computation of detection of p-values. Probe signals are ranked relative to signals of negative controls. Detection p-value =  $1 - R/N$ , where  $R$  is the rank of the gene signal relative to negative controls and  $N$  is number of negative controls. For groups containing multiple samples, the following modification is used. Let  $m$  be the number of samples in the group. On the  $i^{\text{th}}$  sample, the  $g^{\text{th}}$  gene signal is converted to a Z value and the average Z value across all  $m$  samples is computed.

$$Z_g = \frac{1}{m} \sum_i Z_g^i$$

$$Z_g^i = \frac{I - \mu_{neg}^i}{\sigma_{neg}^i}$$

Here  $\mu_{neg}^i$  and  $\sigma_{neg}^i$  are the mean and standard deviation of signals of the negative controls on the  $i^{th}$  sample.  $i$  is the signal from gene  $g$ . The same transformation that is applied to  $i$  is also applied to the signals of negative controls. Detection p-values are computed based on the rank of the Z value of a probe relative to the Z values of the negative controls.

## Focused Arrays

Since these products contain a relatively small number of negative controls, BeadStudio uses normal distribution to model their signals. The detection p-value is given by:

$$p = Z \left( \frac{I_{probe} - \mu_{neg}}{\sigma_{neg}} \right)$$

For the gene represented by N probes, we use:

$$p = Z \left( \sqrt{N} \left| \frac{I_{gene} - \mu_{neg}}{\sigma_{neg}} \right| \right)$$

For groups containing multiple samples, the average Z value is computed as described in *Whole Genome BeadChips* on page 98. The detection p-value is then computed using the following formula:

$$p = Z \left( \frac{|Z_g|}{\sigma_{Z_{neg}}} \right)$$

where  $\sigma_{Z_{neg}}$  is the standard deviation of the average Z values of negative controls.

## DASL Products

Cy3 and Cy5 channel signals are added for the computation of detection p-values.



# Chapter 5

# Generating a Final Report

## Topics

- 102    Introduction
- 102    Generating a Final Report
- 105    Viewing a Final Report

## Introduction

BeadStudio v3 includes the ability to generate a final report. You can save a final report to view later, or to use with downstream, third-party software applications.

This chapter describes how to generate, save, and view a BeadStudio Gene Expression Module final report.

## Generating a Final Report

A final report is the final output of the BeadStudio Gene Expression Module.

To generate a final report, do the following:

1. Select **Analysis | Create Final Report** (Figure 72).

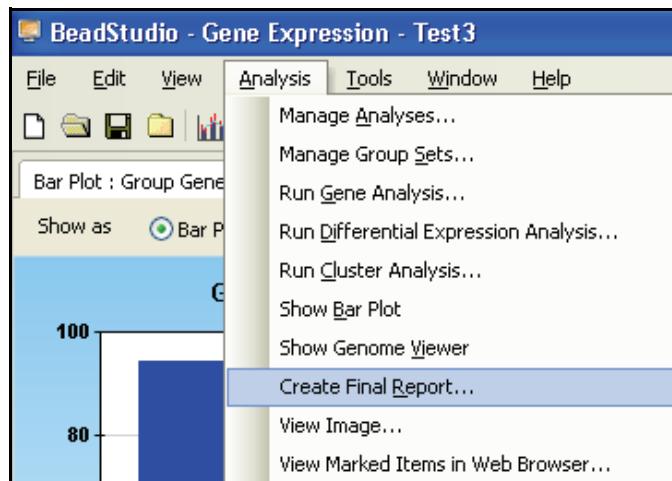
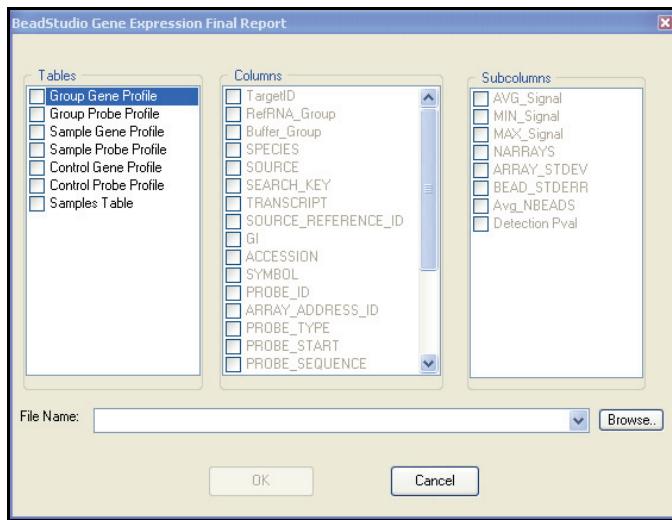


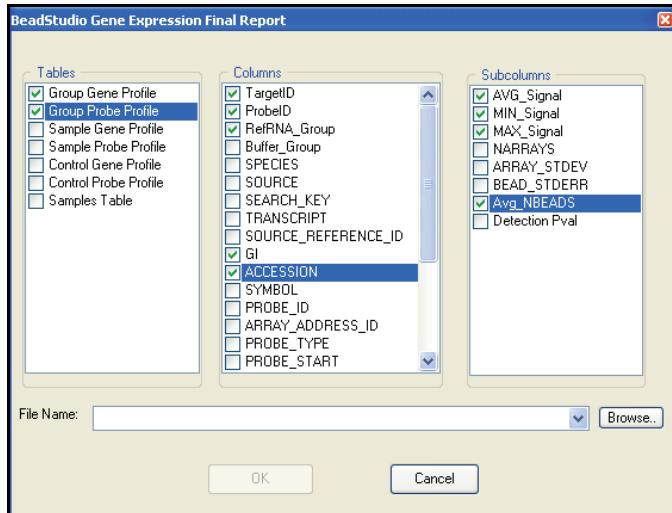
Figure 72 Creating a Final Report

The **BeadStudio Gene Expression Final Report** dialog box appears (Figure 73).



**Figure 73** Final Report Dialog Box

2. Select **Table**, **Columns**, and **Subcolumns** options to include in the final report (Figure 74):



**Figure 74** Selecting Options for the Final Report

3. Click **Browse..** to browse to the location where you would like to save the final report.

The **Final Report File** window appears (Figure 75).

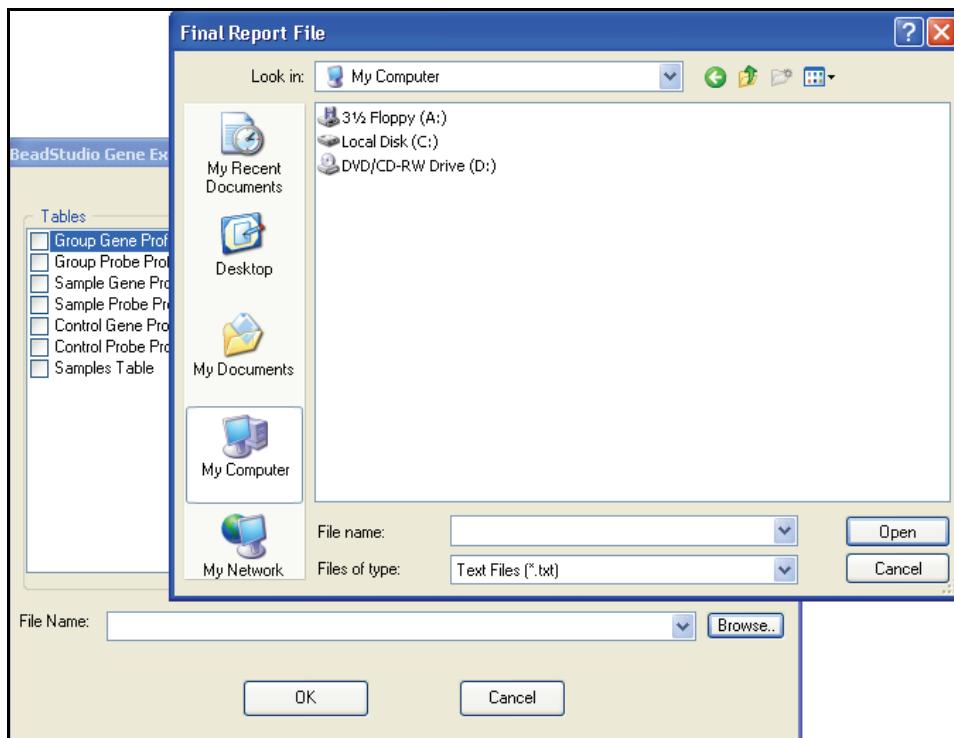


Figure 75 Saving the Final Report

4. In the **Filename** area, type a name for the final report (Figure 76).



Figure 76 Naming the Final Report

5. Click **Open** on the **Final Report File** dialog box.

The **Final Report File** dialog box closes.

6. Click **OK** on the **BeadStudio Gene Expression Final Report** dialog box.

The **BeadStudio Gene Expression Final Report** dialog box closes.

The final report is saved as a tab-delimited \*.txt file in the location you specified.

## Viewing a Final Report

- ▶ Open the final report in Excel or a similar spreadsheet program or text editor (Figure 77).

	A	B	C	D	E	F	G	H	I	J
	ProbeID	RefRNA_Group:AVG_Signal	RefRNA_Group:MIN_Signal	RefRNA_Group:MAX_Signal	RefRNA_Group:Avg_Nbeads	GI	ACCESSION			
1	Project	3.0.9								
2	BSGx Version	3.0.9								
3	Report Date	1/19/2007 14:00								
4	Project	Test3								
5	Group Set	Groupset3								
6	Analysis	Analysis3								
7	Normalization	none								
8	[Group Probe Profile]									
9	TargetID									
10	GI_14141192	2140735	4079.942	3737.702	4611.924	27.33333	14141192_NM_001013.2			
11	GI_14141192	7650170	3846.966	3882.413	4310.893	36.55566	14141192_NM_001013.2			
12	GI_14141192	6550170	1384.22	1143.22	1426.01	30.33333	14141192_NM_001013.2			
13	GI_25453469	1070519	23995.13	20610.07	20509.03	136.4444	25453469_NM_201402.4			
14	GI_25453469	2690379	23469.42	20160.88	27309.6	136.2222	25453469_NM_201402.4			
15	GI_34304116	4780009	15657.59	13773.2	17045.19	39.00889	34304116_NM_021000.2			
16	GI_4507728	4590356	2139.751	1866.2	2364.286	138.1111	4507728_NM_001069.1			
17	GI_4507728	7330238	797.9295	674.2806	914.6613	36.33333	4507728_NM_001069.1			
18	GI_4507744	4260048	6573.157	5732.129	7248.581	153	4507744_NM_003329.1			
19	GI_4507744	7610747	6068.231	5426.35	6937.771	52.77778	4507744_NM_003329.1			
20	GI_5016088	130424	10731.23	9402.636	11513.23	48.55566	5016088_NM_001101.2			
21	GI_5016088	5860528	31395.54	26825.89	34652.14	123.4444	5016088_NM_001101.2			
22	GI_7669301	1770001	9602.707	8529.003	10000.27	91.44444	7669301_NM_002066.2			
23	GI_76693491	7300041	16943.44	14175.38	18039.65	169.6666	76693491_NM_002048.2			
24	ILMN_10000	6980451	214.7381	195.5032	239.8465	38.55556	40317520_NM_007112.3			
25	ILMN_100000	2695054	98.26516	86.26795	105.8547	38.44444	59202801_AL027696			
26	ILMN_100007	7560397	104.575	91.19371	113.4277	39.11111	6462397_AW178383			
27	ILMN_10009	5360747	107.2034	91.74918	122.2501	39.22222	23690395_BU742987			
28	ILMN_10001	2600731	2522.17	2171.289	2699.756	42.22222	31543637_NM_01976.3			
29	ILMN_10010	3990110	127.1422	116.1113	144.2052	42	15937005_BI625456			
30	ILMN_10002	2120309	94.37944	84.26244	99.87687	42.22222	26212219_NM_175669.1			
31	ILMN_10028	6420474	143.5376	123.8457	160.8053	30	1970722_AA32646			
32	ILMN_10030	670095	96.74176	93.14203	106.469	40.66667	13746520_BI625456			
33	ILMN_10031	13037	103.8600	92.96019	113.300	38.66667	1970722_AA32646			
34	ILMN_10034	3600242	89.95807	70.00039	101.1399	32.00009	19930270_BU00211			
35	ILMN_10037	6480041	106.9544	94.26245	117.3792	33.44444	30757590_CD104406			
36	ILMN_10004	7610608	367.607	328.079	390.0716	40.66666	38327631_NM_001954.3			
37	ILMN_10005	1570434	619.8431	546.736	681.0746	42.11111	34304372_NM_031996.2			
38	ILMN_10054	1740092	98.06398	86.36703	102.1347	33.22222	51664217_CR73903			
39	ILMN_10059	3440520	127.7014	109.8731	142.5934	35.44444	6085802_AW117218			
40	ILMN_10006	6520451	120.2124	111.0949	143.398	38.55566	14249385_NM_032452.1			
41	ILMN_10075	780746	114.0733	101.9674	123.5387	45	21757375_AH094644			
42	ILMN_10079	150545	95.51662	88.78805	103.6666	38.66666	34526666_AK123118			

Figure 77 Final Report



# Chapter 6

# User Interface Reference

## Topics

- 108 Introduction
- 108 Detachable Docking Windows
  - 109 Bar or Line Plot: Group Gene Profile
  - 110 Samples Table
  - 112 Group Gene Profile
  - 116 Group Probe Profile
  - 120 Sample Gene Profile
  - 124 Sample Probe Profile
  - 128 Control Gene Profile
  - 130 Control Probe Profile
  - 132 Control Summary
  - 134 Project Window
  - 134 Log Window
- 136 Main Window Menus
- 140 Context Menus

## Introduction

This chapter explains how to use the detachable docking windows, main window menus, and context menus in the BeadStudio Gene Expression Module.

## Detachable Docking Windows

Detachable docking windows let you customize BeadStudio's user interface to suit your analysis needs.

Figure 78 shows the default view of the BeadStudio Gene Expression Module. Detachable docking windows are outlined in orange.

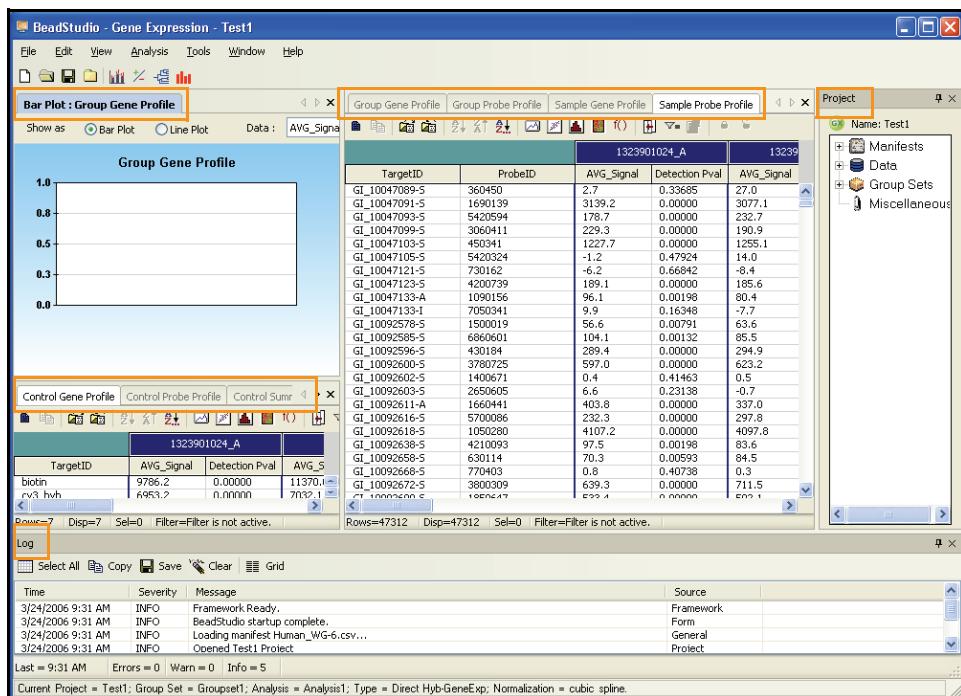


Figure 78 Gene Expression Module Default View

The following sections describe each of the Gene Expression Module's detachable docking windows.

## Bar or Line Plot: Group Gene Profile

Figure 79 shows an example of a Bar Plot: Group Gene Profile window. Click **Line Plot** (outlined in orange) to display the data in a line plot instead.

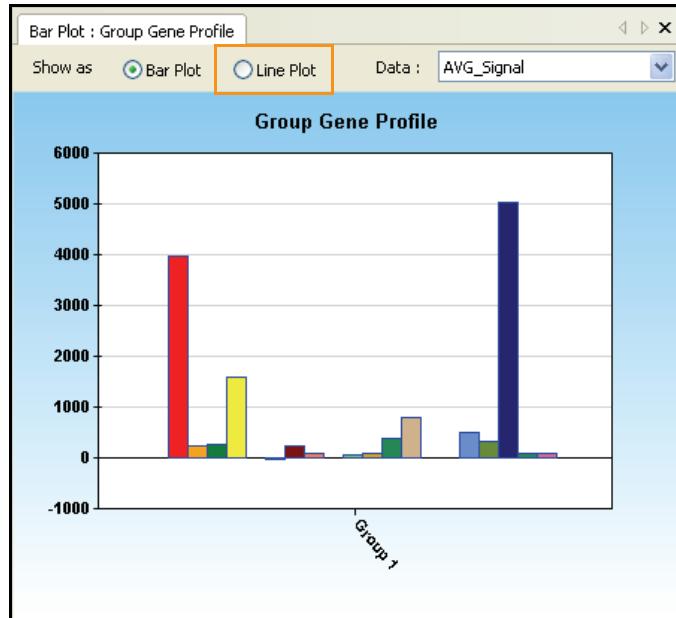


Figure 79 Bar Plot: Group Gene Profile

To customize the view of the bar or line plot, right click on the plot and make selections in the **Plot Settings** dialog box.

**Samples Table**

Figure 80 shows a sample **Samples Table** window.

The screenshot shows a software interface titled "Bar Plot : Group Gene Profile" with a tab labeled "Samples Table". The main area is a table with the following data:

Index	Sample ID	Sample Group	Sentrix Barcode	Sample Section
1	RefRNA_1	RefRNA_Group	1671586184	A
2	RefRNA_2	RefRNA_Group	1671586184	C
3	RefRNA_3	RefRNA_Group	1671586184	E
4	RefRNA_4	RefRNA_Group	1671586185	A
5	RefRNA_5	RefRNA_Group	1671586185	C
6	RefRNA_6	RefRNA_Group	1671586185	E
7	RefRNA_7	RefRNA_Group	1671586186	A
8	RefRNA_8	RefRNA_Group	1671586186	C
9	RefRNA_9	RefRNA_Group	1671586186	E
10	Buffer_1	Buffer_Group	1671586184	B
11	Buffer_2	Buffer_Group	1671586184	D
12	Buffer_3	Buffer_Group	1671586184	F
13	Buffer_4	Buffer_Group	1671586185	B
14	Buffer_5	Buffer_Group	1671586185	D
15	Buffer_6	Buffer_Group	1671586185	F
16	Buffer_7	Buffer_Group	1671586186	B
17	Buffer_8	Buffer_Group	1671586186	D
18	Buffer_9	Buffer_Group	1671586186	F

At the bottom of the window, there are status indicators: Rows=18, Disp=18, Sel=0, Filter=Filter is not active.

**Figure 80** Samples Table

The annotation columns of the **Samples Table** are described in Table 9.

**Table 9** Samples Table Columns

Column	Description	Type	Visible by Default?
Index	The row index of the sample	integer	Y
Sample ID	The sample identifier	string	Y
Sample Group	The group the sample belongs to, as defined in a sample sheet or in the <b>Groupset Definition</b> dialog box	string	Y
Sentrix Barcode	The barcode of the Sentrix Array Product	integer	Y
Sample Section	The section on the Sentrix Array product	string	Y

**Table 9** Samples Table Columns (continued)

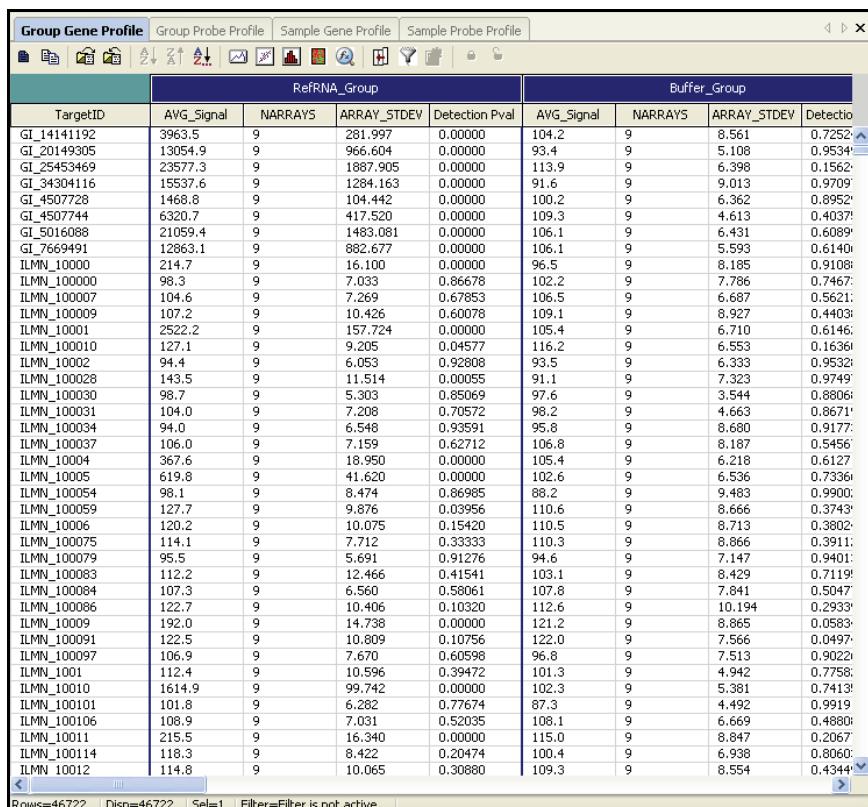
Column	Description	Type	Visible by Default?
<b>Detected Genes (0.01)</b>	The number of genes with a detection p-value of 0.01 or less.	integer	Y
<b>Detected Genes (0.05)</b>	The number of genes with a detection p-value of 0.05 or less	integer	Y
<b>Signal Average</b>	The average signal for the sample across all probes	integer	Y
<b>Signal P05</b>	The fifth percentile average intensity across all probes	integer	Y
<b>Signal P25</b>	The twenty-fifth percentile average intensity across all probes	integer	Y
<b>Signal P50</b>	The fiftieth percentile average intensity across all probes	integer	Y
<b>Signal P75</b>	The seventy-fifth percentile average intensity across all probes	integer	Y
<b>Signal P95</b>	The ninety-fifth percentile average intensity across all probes	integer	Y
<b>Sample_Well</b>	The well within the sample plate	string	Y
<b>Sample_Plate</b>	The sample plate	string	Y
<b>Pool_ID</b>	The bead pool ID	string	Y
<b>BIOTIN</b>	The average biotin control intensity across all probes	float	Y
<b>CY3_HYB</b>	The average CY3_HYB control intensity across all probes	float	Y
<b>HIGH_STRINGENCY_HYB</b>	The average HIGH_STRINGENCY_HYB control intensity across all probes	float	Y
<b>HOUSEKEEPING</b>	The average HOUSEKEEPING control intensity across all probes	float	Y
<b>LABELING</b>	The average LABELING control intensity across all probes	float	Y

**Table 9** Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
LOW_STRINGENCY_HYB	The average LOW_STRINGENCY_HYB control intensity across all probes	float	Y
NEGATIVE (background)	The average NEGATIVE (background) control intensity across all probes	float	Y
NOISE	The average noise control intensity across all probes	float	Y

## Group Gene Profile

Figure 81 shows a sample **Group Gene Profile** window.

**Figure 81** Group Gene Profile Window

The annotation columns of the **Group Gene Profile** window are described in Table 10.

**Table 10** Group Gene Profile Window Columns

Column	Description	Type	Visible by Default?
TargetID	Probe name. Also used as a key column for data import.	string	Y
SPECIES	The species of the BeadChip product	string	N
SOURCE	The database from which the annotation data was acquired	string	N
SEARCH_KEY	Gene identifier provided by the customer (for the DASL assay). Generally equivalent to SYMBOL (for Direct Hyb).	string	Y
TRANSCRIPT	RefSeq entry specifying an isoform (GI number).	string	N
SOURCE_REFERENCE_ID	Database accession number	string	N
GI	RefSeq entry identifier (GI number).	string	N
ACCESSION	RefSeq entry (NM or XM number).	string	N
SYMBOL	Gene name as reported in RefSeq.	string	Y
PROBE_ID	Illumina identifier for probe sequence.	integer	N
ARRAY_ADDRESS_ID	Internal ID used by Illumina software	integer	N

**Table 10** Group Gene Profile Window Columns (continued)

Column	Description	Type	Visible by Default?
PROBE_TYPE	A, I, or S <ul style="list-style-type: none"> <li>For transcripts with a single isoform, we design "-S" probes (S=single)</li> </ul> <p>For transcripts with multiple isoforms, we design two types of probes:</p> <ul style="list-style-type: none"> <li>"-I" (I=isoform-specific) are probes designed to query only one of multiple isoforms</li> <li>"-A" (A=all) are probes designed to query all known isoforms of that transcript</li> </ul>	string	N
PROBE_START	Coordinate in database entry where probe sequence begins	integer	N
PROBE_SEQUENCE	Sequence used as a probe on the array.	string	N
CHROMOSOME	The chromosome on which the probe is located	string	Y
PROBE_CHR_ORIENTATION	The DNA strand on which the probe is located (positive or negative)	string	N
PROBE_COORDINATES	The start and end positions of the probe on the chromosome	string	N
DEFINITION	Single-line description of gene in RefSeq	string	Y
ONTOLOGY_COMPONENT	The gene ontology (GO) component classification(s) for this probe	string	N
ONTOLOGY_PROCESS	The gene ontology (GO) process classification(s) for this probe	string	N
ONTOLOGY_FUNCTION	The gene ontology (GO) function classification(s) for this probe	string	N
SYNONYMS	Other names (aliases) for the same gene.	string	Y

The per-group columns of the **Group Gene Profile** window are described in Table 15.

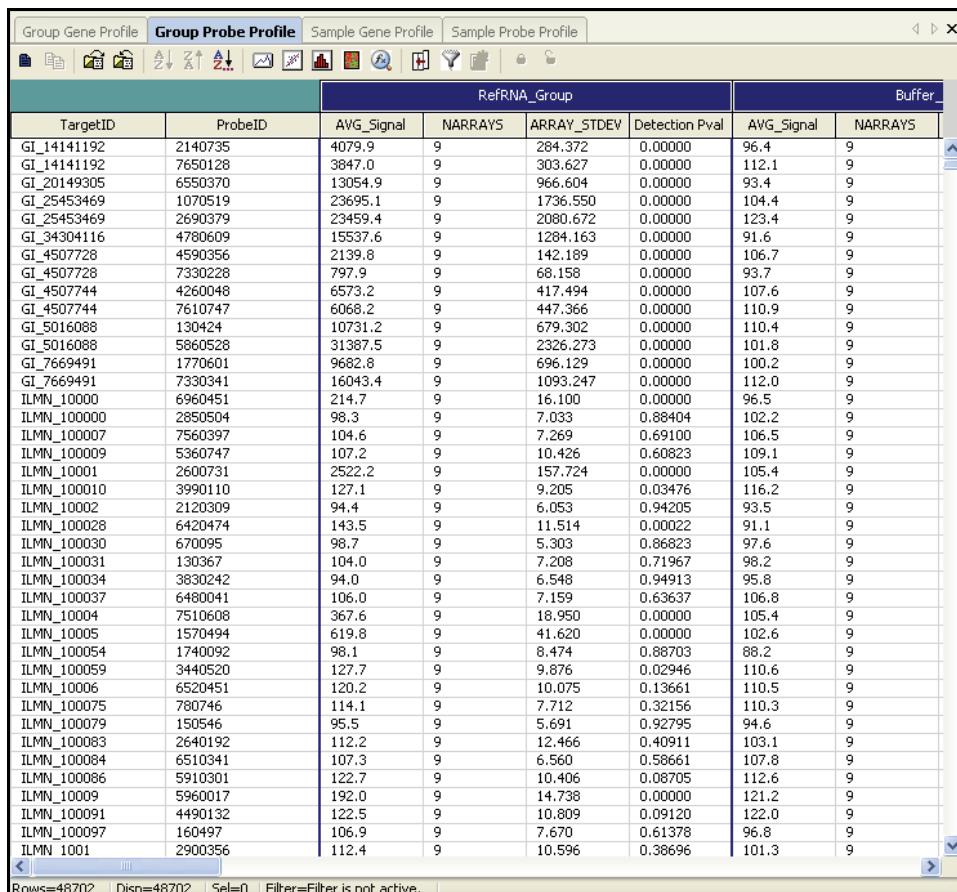
**Table 11** Group Gene Profile Window Per-Group Columns

Column	Description	Type	Visible by Default?
AVG_Signal	Average intensity of the bead type/target in the group.	float	Y
NARRAYS	Number of samples in the group.	integer	Y
ARRAY_STDEV	Standard deviation associated with sample-to-sample variability within the group. Undefined when the group contains a single sample.	float	Y
Detection Pval	P-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls.	float	Y
MIN_Signal	Minimum intensity of the bead type/target in the group.	float	Y
MAX_Signal	Maximum intensity of the bead type/target in the group.	float	Y
BEAD_STDERR	Average standard error associated with bead-to-bead variability for the samples in the group.	float	Y
Avg_NBEADS	Average number of beads per bead type representing probes for the gene.	integer	Y

In the case of groups containing only one sample, MIN, AVG, and MAX signals are equal.

## Group Probe Profile

Figure 82 shows a sample **Group Probe Profile** window.



The screenshot shows a software interface titled "Group Probe Profile". The window contains a table with two main sections: "RefRNA\_Group" and "Buffer". The "RefRNA\_Group" section has columns for TargetID, ProbeID, AVG\_Signal, NARRAYS, ARRAY\_STDEV, Detection Pval, and Buffer. The "Buffer" section has columns for AVG\_Signal and NARRAYS. The table lists numerous entries, mostly starting with "GI\_". The bottom of the window displays status information: "Rows=48702 Disp=48702 Sel=0 Filter=Filter is not active."

		RefRNA_Group				Buffer	
TargetID	ProbeID	AVG_Signal	NARRAYS	ARRAY_STDEV	Detection Pval	AVG_Signal	NARRAYS
GI_14141192	2140735	4079.9	9	284.372	0.00000	96.4	9
GI_14141192	7650128	3847.0	9	303.627	0.00000	112.1	9
GI_20149305	6550370	13054.9	9	966.604	0.00000	93.4	9
GI_25453469	1070519	23695.1	9	1736.550	0.00000	104.4	9
GI_25453469	2690379	23459.4	9	2080.672	0.00000	123.4	9
GI_34304116	4780609	15537.6	9	1284.163	0.00000	91.6	9
GI_4507728	4590356	2139.8	9	142.189	0.00000	106.7	9
GI_4507728	7330228	797.9	9	68.158	0.00000	93.7	9
GI_4507744	4260048	6573.2	9	417.494	0.00000	107.6	9
GI_4507744	7610747	6068.2	9	447.366	0.00000	110.9	9
GI_5016088	130424	10731.2	9	679.302	0.00000	110.4	9
GI_5016088	5860528	31387.5	9	2326.273	0.00000	101.8	9
GI_7669491	1770601	9682.8	9	696.129	0.00000	100.2	9
GI_7669491	7330341	16043.4	9	1093.247	0.00000	112.0	9
ILMN_10000	6960451	214.7	9	16.100	0.00000	96.5	9
ILMN_100000	2850504	98.3	9	7.033	0.88404	102.2	9
ILMN_100007	7560397	104.6	9	7.269	0.69100	106.5	9
ILMN_100009	5360747	107.2	9	10.426	0.60823	109.1	9
ILMN_10001	2600731	2522.2	9	157.724	0.00000	105.4	9
ILMN_100010	3990110	127.1	9	9.205	0.03476	116.2	9
ILMN_10002	2120309	94.4	9	6.053	0.94205	93.5	9
ILMN_100028	6420474	143.5	9	11.514	0.00022	91.1	9
ILMN_100030	670095	98.7	9	5.303	0.86823	97.6	9
ILMN_100031	130367	104.0	9	7.208	0.71967	98.2	9
ILMN_100034	3830242	94.0	9	6.548	0.94913	95.8	9
ILMN_100037	6480041	106.0	9	7.159	0.63637	106.8	9
ILMN_10004	7510608	367.6	9	18.950	0.00000	105.4	9
ILMN_10005	1570494	619.8	9	41.620	0.00000	102.6	9
ILMN_100054	1740092	98.1	9	8.474	0.88703	88.2	9
ILMN_100059	3440520	127.7	9	9.876	0.02946	110.6	9
ILMN_10006	6520451	120.2	9	10.075	0.13661	110.5	9
ILMN_100075	780746	114.1	9	7.712	0.32156	110.3	9
ILMN_100079	150546	95.5	9	5.691	0.92795	94.6	9
ILMN_100083	2640192	112.2	9	12.466	0.40911	103.1	9
ILMN_100084	6510341	107.3	9	6.560	0.58661	107.8	9
ILMN_100086	5910301	122.7	9	10.406	0.08705	112.6	9
ILMN_10009	5960017	192.0	9	14.738	0.00000	121.2	9
ILMN_100091	4490132	122.5	9	10.809	0.09120	122.0	9
ILMN_100097	160497	106.9	9	7.670	0.61378	96.8	9
ILMN_1001	2900356	112.4	9	10.596	0.38696	101.3	9

Figure 82 Group Probe Profile Window

The annotation columns of the **Group Probe Profile** window are described in Table 12.

**Table 12** Group Probe Profile Window Columns

Column	Description	Type	Visible by Default?
Target_ID	Probe name. Also used as a key column for data import.	string	Y
Probe_ID	Bead type.	integer	Y
SPECIES	The species of the BeadChip product	string	N
SOURCE	The database from which the annotation data was acquired	string	N
SEARCH_KEY	Gene identifier provided by the customer (for the DASL assay). Generally equivalent to SYMBOL (for Direct Hyb).	string	Y
TRANSCRIPT	RefSeq entry specifying an isoform (GI number).	string	N
SOURCE_REFERENCE_ID	Database accession number	string	N
GI	RefSeq entry identifier (GI number).	string	N
ACCESSION	RefSeq entry (NM or XM number).	string	N
SYMBOL	Gene name as reported in RefSeq.	string	Y
PROBE_ID	Illumina identifier for probe sequence.	integer	N
ARRAY_ADDRESS_ID	Internal ID used by Illumina software	integer	N

**Table 12** Group Probe Profile Window Columns (continued)

Column	Description	Type	Visible by Default?
PROBE_TYPE	A, I, or S <ul style="list-style-type: none"> <li>• For transcripts with a single isoform, we design "-S" probes (S=single)</li> </ul> For transcripts with multiple isoforms, we design two types of probes: <ul style="list-style-type: none"> <li>• "-I" (I=isoform-specific) are probes designed to query only one of multiple isoforms</li> <li>• "-A" (A=all) are probes designed to query all known isoforms of that transcript</li> </ul>	string	N
PROBE_START	Coordinate in database entry where probe sequence begins.	integer	N
PROBE_SEQUENCE	Sequence used as a probe on the array.	string	N
CHROMOSOME	The chromosome on which the probe is located	string	Y
PROBE_CHR_ORIENTATION	The DNA strand on which the probe is located (positive or negative)	string	N
PROBE_COORDINATES	The start and end positions of the probe on the chromosome	string	N
DEFINITION	Single-line description of gene in RefSeq.	string	Y
ONTOLOGY_COMPONENT	The gene ontology (GO) component classification(s) for this probe	string	N
ONTOLOGY_PROCESS	The gene ontology (GO) process classification(s) for this probe	string	N
ONTOLOGY_FUNCTION	The gene ontology (GO) function classification(s) for this probe	string	N
SYNONYMS	Other names for the same gene (aliases).	string	Y

The per-group columns of the **Group Probe Profile** window are described in Table 13.

**Table 13** Group Probe Profile Window Per-Group Columns

Column	Description	Type	Visible by Default?
AVG_Signal	Average intensity of the bead type/target in the group.	float	Y
NARRAYS	Number of samples in the group.	integer	Y
ARRAY_STDEV	Standard deviation associated with sample-to-sample variability within the group (undefined when the group contains a single sample).	float	Y
Detection Pval	P-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from negative controls.	float	Y
MIN_Signal	Minimum intensity of bead type/target in the group.	float	N
MAX_Signal	Maximum intensity of bead type/target in the group.	float	N
BEAD_STDERR	Average standard error associated with bead-to-bead variability for the samples in the group.	float	N
Avg_NBEADS	Average number of beads per bead type representing probes for the gene.	integer	N

In the case of groups containing only one sample, MIN, AVG, and MAX signals are equal.

## Sample Gene Profile

Figure 83 shows an example of a Sample Gene Profile window.

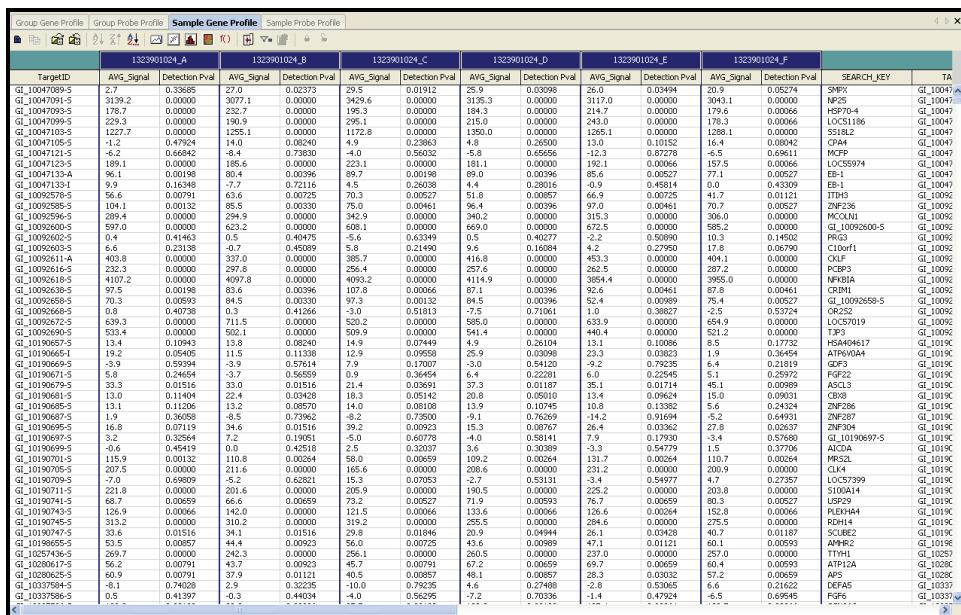


Figure 83 Sample Gene Profile Window

The annotation columns of the **Sample Gene Profile** window are described in Table 14.

**Table 14** Sample Gene Profile Window Columns

Column	Description	Type	Visible by Default?
TargetID	Probe name. Also used as a key column for data import.	string	Y
SPECIES	The species of the BeadChip product	string	N
SOURCE	The database from which the annotation data was acquired	string	N
SEARCH_KEY	Gene identifier provided by the customer (for the DASL assay). Generally equivalent to SYMBOL (for Direct Hyb).	string	Y
TRANSCRIPT	RefSeq entry specifying an isoform (GI number).	string	N
SOURCE_REFERENCE_ID	Database accession number	string	N
GI	RefSeq entry identifier (GI number).	string	N
ACCESSION	RefSeq entry (NM or XM number).	string	N
SYMBOL	Gene name as reported in RefSeq.	string	Y
PROBE_ID	Illumina identifier for probe sequence.	integer	N
ARRAY_ADDRESS_ID	Internal ID used by Illumina software	integer	N

**Table 14** Sample Gene Profile Window Columns (continued)

Column	Description	Type	Visible by Default?
PROBE_TYPE	<p>A, I, or S</p> <ul style="list-style-type: none"> <li>For transcripts with a single isoform, we design "S" probes (S=single)</li> </ul> <p>For transcripts with multiple isoforms, we design two types of probes:</p> <ul style="list-style-type: none"> <li>"-I" (I=isoform-specific) are probes designed to query only one of multiple isoforms</li> <li>"-A" (A=all) are probes designed to query all known isoforms of that transcript</li> </ul>	string	N
PROBE_START	Coordinate in database entry where probe sequence begins.	integer	Y
PROBE_SEQUENCE	Sequence used as a probe on the array.	string	N
CHROMOSOME	The chromosome on which the probe is located	string	Y
PROBE_CHR_ORIENTATION	The DNA strand on which the probe is located (positive or negative)	string	N
PROBE_COORDINATES	The start and end positions of the probe on the chromosome	string	N
DEFINITION	Single-line description of gene in RefSeq.	string	Y
ONTOLOGY_COMPONENT	The gene ontology (GO) component classification(s) for this probe	string	N
ONTOLOGY_PROCESS	The gene ontology (GO) process classification(s) for this probe	string	N
ONTOLOGY_FUNCTION	The gene ontology (GO) function classification(s) for this probe	string	N
SYNONYMS	Other names for the same gene (aliases).	string	Y

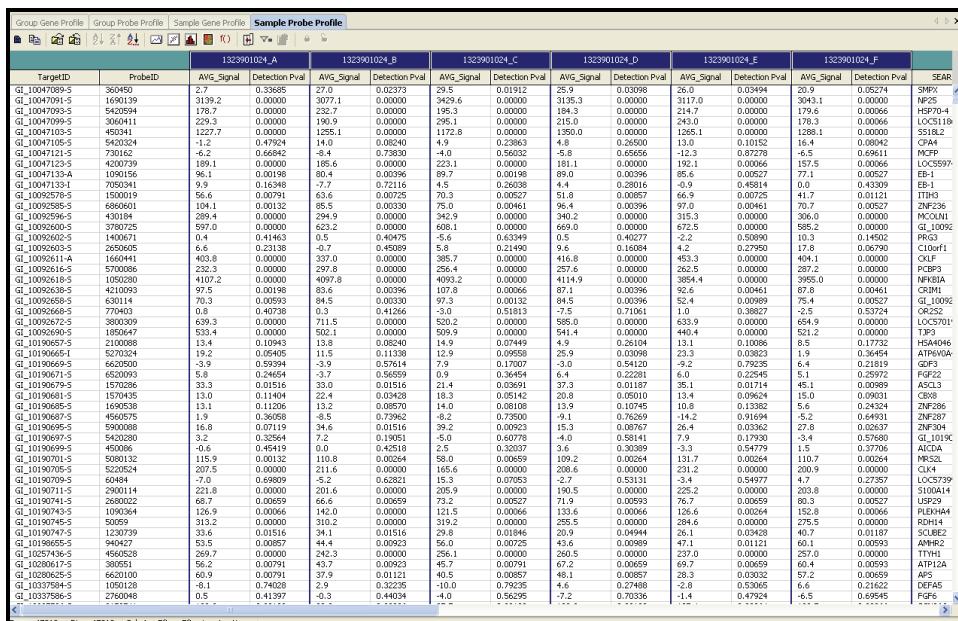
The per-sample columns of the **Sample Gene Profile** window are described in Table 15.

**Table 15** Sample Gene Profile Window Per-Sample Columns

Column	Description	Type	Visible by Default?
AVG_Signal	Average intensity of the bead type/target in the group.	float	Y
Detection_Pval	P-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls.	float	Y
NARRAYS	Number of samples in the group.	integer	N
ARRAY_STDEV	Standard deviation associated with sample-to-sample variability within the group. Undefined when the group contains a single sample.	string	N
BEAD_STDERR	Average standard error associated with bead-to-bead variability for the samples in the group.	float	N
Avg_NBEADS	Average number of beads per bead type representing probes for the gene.	integer	N

# Sample Probe Profile

Figure 84 shows an example of a **Sample Probe Profile** window.



**Figure 84** Sample Probe Profile Window

The annotation columns of the **Sample Probe Profile** window are listed and described in Table 16.

**Table 16** Sample Probe Profile Window Columns

Column	Description	Type	Visible by Default?
TargetID	Identifies the probe name. Also used as a key column for data import.	string	Y
ProbeID	Identifies the bead type.	integer	Y
SPECIES	The species of the BeadChip product	string	N
SOURCE	The database from which the annotation data was acquired	string	N

**Table 16** Sample Probe Profile Window Columns (continued)

Column	Description	Type	Visible by Default?
SEARCH_KEY	Gene identifier provided by the customer (for the DASL assay). Generally equivalent to SYMBOL (for Direct Hyb).	string	Y
TRANSCRIPT	RefSeq entry specifying an isoform (GI number).	string	N
SOURCE_REFERENCE_ID	Database accession number	string	N
GI	RefSeq entry identifier (GI number).	string	N
ACCESSION	RefSeq entry (NM or XM number).	string	N
SYMBOL	Gene name as reported in RefSeq.	string	Y
PROBE_ID	Illumina identifier for probe sequence.	integer	N
ARRAY_ADDRESS_ID	Internal ID used by Illumina software	integer	N
PROBE_TYPE	A, I, or S <ul style="list-style-type: none"> <li>• For transcripts with a single isoform, we design "-S" probes (S=single)</li> </ul> <p>For transcripts with multiple isoforms, we design two types of probes:</p> <ul style="list-style-type: none"> <li>• "-I" (I=isoform-specific) are probes designed to query only one of multiple isoforms</li> <li>• "-A" (A=all) are probes designed to query all known isoforms of that transcript</li> </ul>	string	N
PROBE_START	Coordinate in database entry where probe sequence begins.	integer	N
PROBE_SEQUENCE	Sequence used as a probe on the array.	string	N
CHROMOSOME	The chromosome on which the probe is located	string	Y

**Table 16** Sample Probe Profile Window Columns (continued)

Column	Description	Type	Visible by Default?
SEARCH_KEY	Gene identifier provided by the customer (for the DASL assay). Generally equivalent to SYMBOL (for Direct Hyb).	string	Y
TRANSCRIPT	RefSeq entry specifying an isoform (GI number).	string	N
SOURCE_REFERENCE_ID	Database accession number	string	N
GI	RefSeq entry identifier (GI number).	string	N
ACCESSION	RefSeq entry (NM or XM number).	string	N
SYMBOL	Gene name as reported in RefSeq.	string	Y
PROBE_ID	Illumina identifier for probe sequence.	integer	N
ARRAY_ADDRESS_ID	Internal ID used by Illumina software	integer	N
PROBE_TYPE	A, I, or S <ul style="list-style-type: none"> <li>• For transcripts with a single isoform, we design "-S" probes (S=single)</li> </ul> For transcripts with multiple isoforms, we design two types of probes: <ul style="list-style-type: none"> <li>• "-I" (I=isoform-specific) are probes designed to query only one of multiple isoforms</li> <li>• "-A" (A=all) are probes designed to query all known isoforms of that transcript</li> </ul>	string	N
PROBE_START	Coordinate in database entry where probe sequence begins.	integer	N
PROBE_SEQUENCE	Sequence used as a probe on the array.	string	N
CHROMOSOME	The chromosome on which the probe is located	string	Y

**Table 16** Sample Probe Profile Window Columns (continued)

Column	Description	Type	Visible by Default?
PROBE_CHR_ORIENTATION	The DNA strand on which the probe is located (positive or negative)	string	N
PROBE_COORDINATES	The start and end positions of the probe on the chromosome	string	N
DEFINITION	Single-line description of gene in RefSeq.	string	Y
ONTOLOGY_COMPONENT	The gene ontology (GO) component classification(s) for this probe	string	N
ONTOLOGY_PROCESS	The gene ontology (GO) process classification(s) for this probe	string	N
ONTOLOGY_FUNCTION	The gene ontology (GO) function classification(s) for this probe	string	N
SYNONYMS	Other names for the same gene (aliases).	string	Y

The per-sample columns of the **Sample Probe Profile** window are listed and described in Table 17.

**Table 17** Sample Probe Profile Window Per-Sample Columns

Column	Description	Type	Visible by Default?
AVG_Signal	Average intensity of the bead type/target in the group.	float	Y
Detection_Pval	P-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls.	float	Y
NARRAYS	Number of samples in the group.	integer	N
ARRAY_STDEV	Standard deviation associated with sample-to-sample variability within the group. Undefined when the group contains a single sample.	string	N

**Table 17** Sample Probe Profile Window Per-Sample Columns (continued)

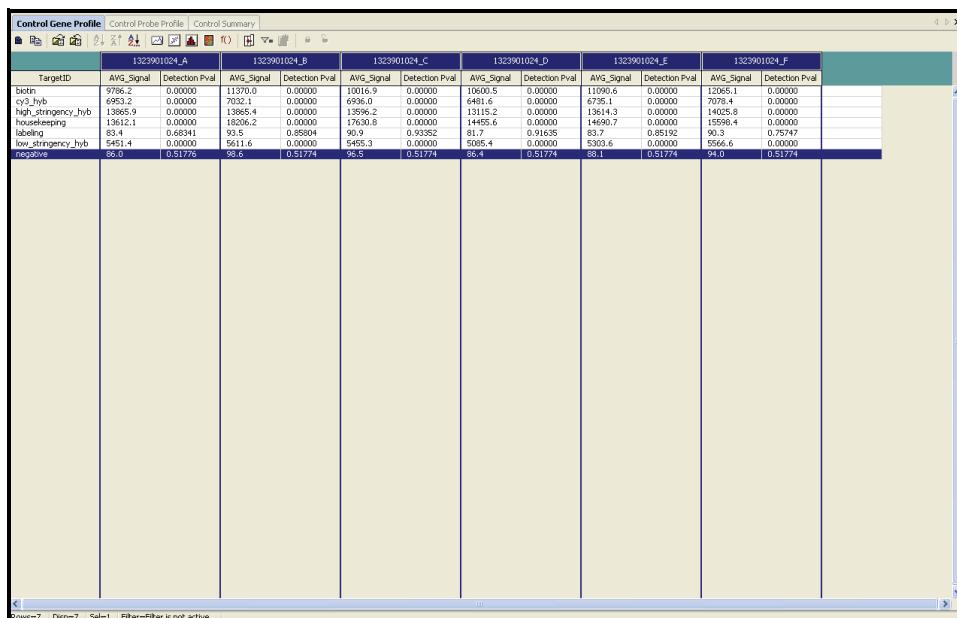
Column	Description	Type	Visible by Default?
BEAD_STDERR	Average standard error associated with bead-to-bead variability for the samples in the group.	float	N
Avg_NBEADS	Average number of beads per bead type representing probes for the gene.	integer	N



Groups (columns in Group tabs) are named according to the names given when defining groupsets, or from the sample sheet. Samples (columns in Sample tabs) are named according to the Sentrix Array product.

## Control Gene Profile

Figure 85 shows a sample **Control Gene Profile** window.

**Figure 85** Control Gene Profile Window

The annotation columns of the **Control Gene Profile** window are described in Table 18.

**Table 18** Control Gene Profile Window Columns

Column	Description	Type	Visible by Default?
TargetID	Probe name. Also used as a key column for data import.	string	Y

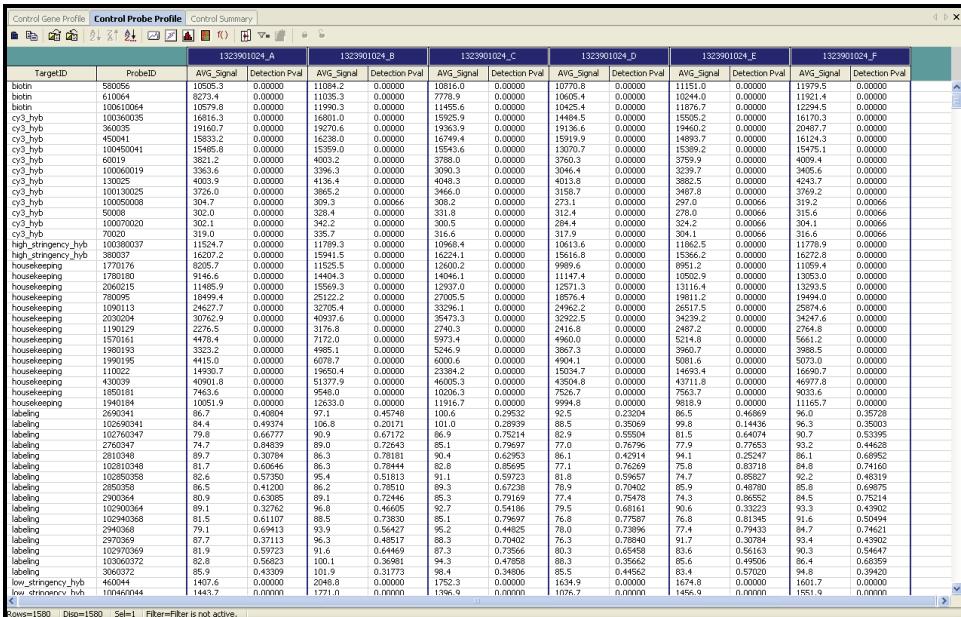
The per-sample columns of the Control Gene Profile window are described in Table 15.

**Table 19** Control Gene Profile Window Per-Sample Columns

Column	Description	Type	Visible by Default?
AVG_Signal	Average intensity of the bead type/target in the group.	float	Y
Detection_Pval	P-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls.	float	Y
NARRAYS	Number of samples in the group.	integer	N
ARRAY_STDEV	Standard deviation associated with sample-to-sample variability within the group. Undefined when the group contains a single sample.	string	N
BEAD_STDERR	Average standard error associated with bead-to-bead variability for the samples in the group.	float	N
Avg_NBEADS	Average number of beads per bead type representing probes for the gene.	integer	N

## Control Probe Profile

Figure 86 shows a sample Control Probe Profile window.



The screenshot displays a software interface titled "Control Probe Profile". The main area is a grid table with 15 columns and approximately 180 rows of data. The columns are labeled as follows:

- Row 1: TargetID, ProbeID, 1323901024\_A, 1323901024\_B, 1323901024\_C, 1323901024\_D, 1323901024\_E, 1323901024\_F
- Row 2: TargetID, ProbeID, AVG\_Signal, Detection\_Pval, AVG\_Signal, Detection\_Pval, AVG\_Signal, Detection\_Pval, AVG\_Signal, Detection\_Pval, AVG\_Signal, Detection\_Pval, AVG\_Signal, Detection\_Pval, AVG\_Signal, Detection\_Pval

The data in the grid includes various numerical values such as 1059.5, 0.00000, 11084.2, 0.00000, etc., for each target and probe combination. The interface has a standard Windows-style toolbar at the top and a status bar at the bottom indicating "Rows=1580 Disp=1580 Sel=1 Filter=Filter is not active."

Figure 86 Control Probe Profile Window

The annotation columns of the **Control Probe Profile** window are described in Table 20.

*Table 20* Control Probe Profile Window Columns

Column	Description	Type	Visible by Default?
TargetID	Probe name. Also used as a key column for data import.	string	Y
ProbeID	Bead type.	integer	Y

The per-sample columns of the **Control Probe Profile** window are listed and described in Table 15.

*Table 21* Control Probe Profile Window Per-Sample Columns

Column	Description	Type	Visible by Default?
AVG_Signal	Average intensity of the bead type/target in the group.	float	Y
Detection_Pval	P-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls.	float	Y
NARRAYS	Number of samples in the group.	integer	N
ARRAY_STDEV	Standard deviation associated with sample-to-sample variability within the group. Undefined when the group contains a single sample.	string	N
BEAD_STDERR	Average standard error associated with bead-to-bead variability for the samples in the group.	float	N
Avg_NBEADS	Average number of beads per bead type representing probes for the gene.	integer	N

## Control Summary

The Gene Expression Module displays a graphic **Control Summary** for the selected arrays based on the performance of the built-in controls.

Figure 87 shows a sample **Control Summary** window.

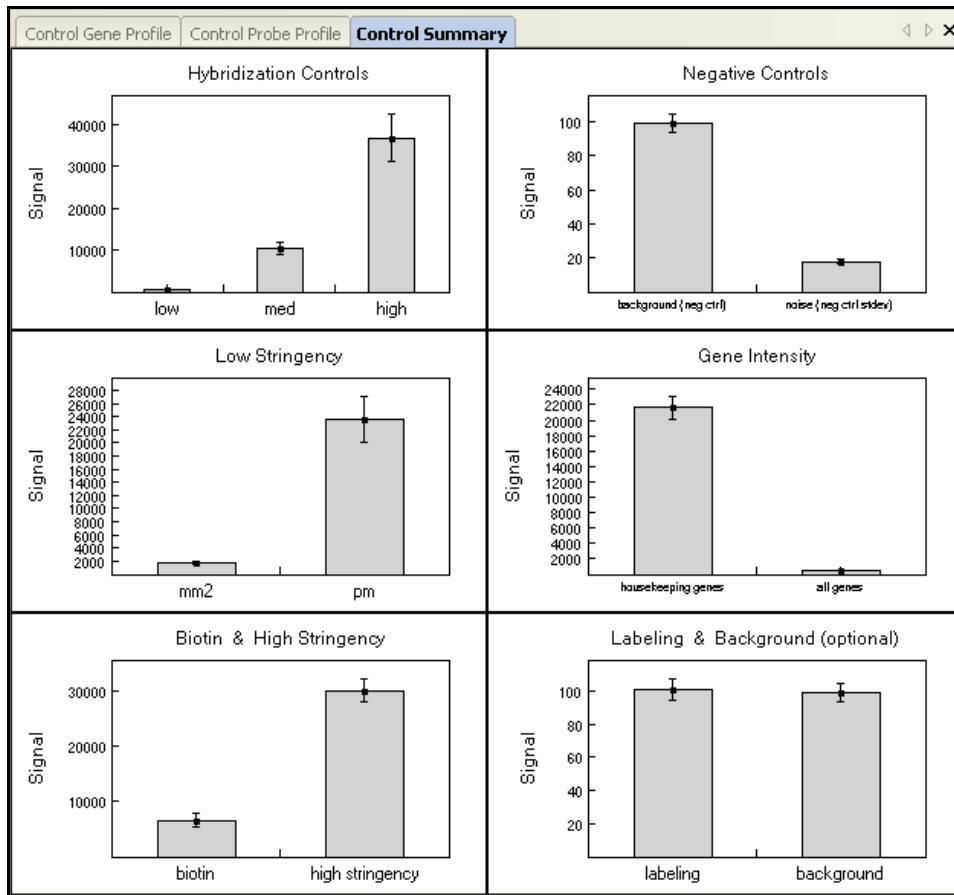


Figure 87 Control Summary Window, Direct Hyb

Figure 88 shows a sample **Control Summary** window for DASL.

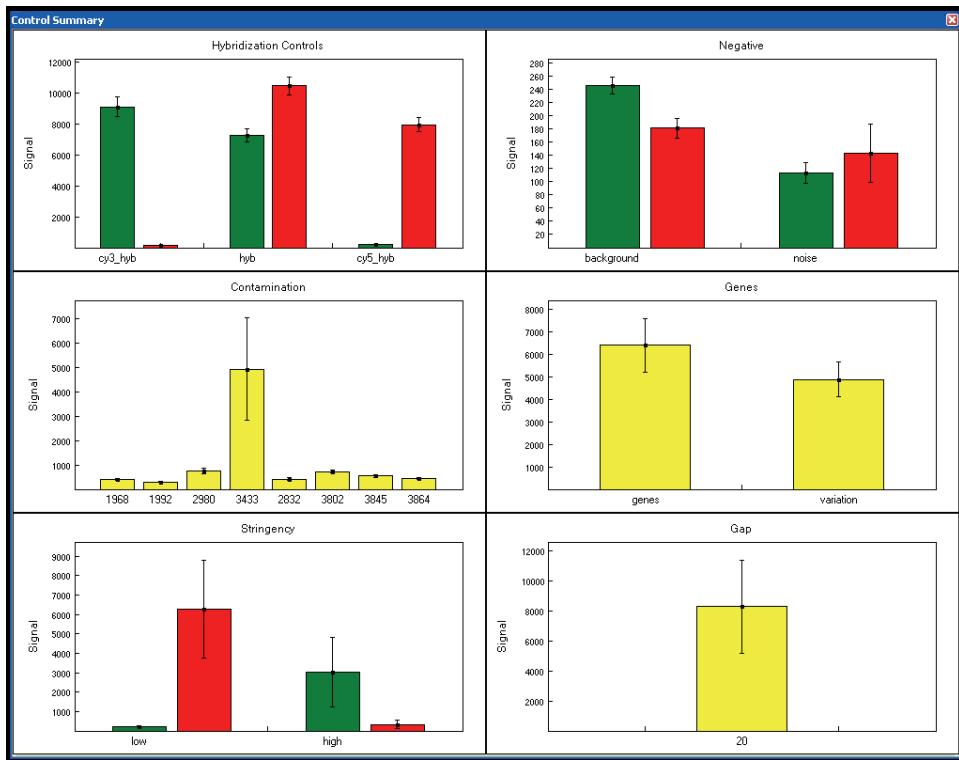


Figure 88 Control Summary Window, DASL

For detailed information about the controls, see the *System Controls* appendix in the appropriate Illumina product guide.

**Project  
Window**

The **Project** window (Figure 89) identifies the manifest(s) loaded for your project and has a data section that identifies all of the Sentrix Array Product barcodes used in your project. You can expand a barcode and view the samples loaded on that Sentrix Array Product.



Figure 89 Project Window

**Log Window**

The **Log** window (Figure 90) is a simple console that provides feedback on BeadStudio processes. The **Log** window displays any errors in red.

Log			
<input type="checkbox"/> Select All <input type="button" value="Copy"/> <input type="button" value="Save"/> <input type="button" value="Clear"/> <input type="button" value="Grid"/>			
Time	Severity	Message	Source
3/9/2006 9:23 AM	INFO	Framework Ready.	Framework
3/9/2006 9:23 AM	INFO	BeadStudio startup complete.	Form
3/9/2006 9:24 AM	INFO	Opened Test Project 1 Project	Project
3/9/2006 9:24 AM	INFO	Loading manifest Human_WG-6.csv...	General
3/9/2006 9:24 AM	INFO	Opened Test Project 1 Project	Framework

Figure 90 Log Window

**Table 22** Log Window Selections & Functions

Selection	Function	Toolbar Button (if used)
Select All	Selects all log entries.	
Copy	Copies log entries to the clipboard.	
Save	Saves all log entries.	
Clear	Clears all log entries.	
Grid	Toggles the grid off and on.	
Time	Displays the time the log entry was generated.	
Severity	Displays the severity of the log entry.	
Message	Displays the text description of the log entry.	
Source	Displays the source of the log entry.	

## Main Window Menus

Table 23 lists the selections available from the BeadStudio Gene Expression Module main window menus and corresponding toolbar buttons.

**Table 23** Main Menu Selections & Functions

Selection	Function	Toolbar Button (if used)
<b>File Menu</b>		
New Project	Opens a new project	
Open Project	Opens a previously saved project.	
Save Project	Saves all current information in this project, so you can return to it later.	
Save Project Copy As	Allows you to specify a file name and location where you would like to save a copy of the current project.	
Close Project	Closes the current project and returns you to the start window of the Gene Expression module.	
Export to GeneSpring GX Format	<b>Group Gene Profile</b> —Exports the Group Gene Profile to a *.txt file in a location you specify. <b>Group Probe Profile</b> —Exports the Group Probe Profile to a *.txt file in a location you specify. <b>Sample Gene Profile</b> —Exports the Sample Gene Profile to a *.txt file in a location you specify. <b>Sample Probe Profile</b> —Exports the Sample Probe Profile to a *.txt file in a location you specify.	
Manage Project Data	Opens the BeadStudio Project Wizard - Project Data Selection dialog box, from which you can specify the Sentrix Array Products to include in your project.	
Page Setup	Opens the Windows Page Setup dialog box, which you can use to set up the page properties and configure the printer properties.	

**Table 23** Main Menu Selections & Functions (continued)

Selection	Function	Toolbar Button (if used)
<b>Print Preview</b>	Opens the Print Preview window, which you can use to see how the selected graph will print.	
<b>Print</b>	Displays the print dialog box. Use this dialog box to select options for printing the currently displayed graph.	
<b>Recent Project</b>	Allows you to select a project you have recently worked on.	
<b>Exit</b>	Closes BeadStudio.	
<b>Edit Menu</b>		
<b>Cut</b>	Cuts the current selection.	
<b>Copy</b>	Copies the current selection to the clipboard.	
<b>Paste</b>	Pastes the current selection from the clipboard.	
<b>Delete</b>	Deletes the current selection.	
<b>Select All</b>	Selects all rows in the current table.	
<b>View Menu</b>		
<b>Save Default View</b>	Allows you to save the default view of the open project.	
<b>Restore Default View</b>	Restores the default view of the open project.	
<b>Save Custom View</b>	Allows you to save a custom view to use again later.	
<b>Load Custom View</b>	Allows you to load a previously-saved custom view.	
<b>Log</b>	Shows or hides the <b>Log</b> window.	
<b>Project</b>	Shows or hides the <b>Project</b> window.	

**Table 23** Main Menu Selections & Functions (continued)

Selection	Function	Toolbar Button (if used)
<b>Analysis Menu</b>		
Manage Analyses	Displays the <b>BeadStudio Manage Analyses</b> dialog box.	
Manage Group Sets	Displays the <b>BeadStudio Project Wizard—Groupset Definition</b> dialog box.	
Run Gene Analysis	Performs gene analysis for the current experiment.	
Run Differential Expression Analysis	Performs differential expression analysis for the current experiment.	
Run Cluster Analysis	Creates a dendrogram for the current experiment.	
Show Bar Plot	Creates a bar plot for the current experiment.	
Show Genome Viewer	Launches the Illumina Genome Viewer (IGV).	
Create Final Report	Displays the <b>BeadStudio Gene Expression Final Report</b> dialog, from which you can create a Final Report.	
View Image	Displays the <b>BeadStudio View Image</b> dialog, from which you can select an image to view.	
View Marked Items in Web Browser	Displays the <b>BeadStudio Gene Expression Web Browser</b> dialog, from which you can select columns and subcolumns to display in a web browser.	

**Table 23** Main Menu Selections & Functions (continued)

Selection	Function	Toolbar Button (if used)
<b>Tools Menu</b>		
Options	<p><b>Project</b>—Opens the Project Properties window, in which you can make changes to project settings.</p> <p><b>BeadStudio</b>—Opens the BeadStudio Options window, in which you can select BeadStudio options including the maximum number of project files and display attributes such as font name, size, and style.</p> <p><b>Module</b>—Opens the module Properties window, in which you can select file-based storage or memory-based storage.</p>	
<b>Windows Menu</b>		
	<p>This menu is populated with the windows that are currently available. Check marks indicate the windows that are currently displayed.</p>	
<b>Help Menu</b>		
About BeadStudio	<p>Opens the <b>About</b> box for the BeadStudio application, which contains:</p> <ul style="list-style-type: none"> <li>• Version information for the BeadStudio Framework as well as any BeadStudio modules you have installed</li> <li>• BeadStudio copyright information</li> <li>• Software copyright notice</li> </ul>	

## Context Menus

The following tables list BeadStudio Gene Expression Module context menu elements and descriptions.

Table 24 includes **Bar Plot: Gene Profile** window context menu elements and descriptions.

**Table 24** Bar Plot: Group Gene Profile Window Context Menu

Element	Description
Properties	Displays the Plot Settings dialog box, from which you can alter the visual properties of the bar plot.
Clear Selected Values	Clears selected values from the bar plot.
Copy As	Copies the bar plot to the clipboard as one of the following file types: BMP, JPEG, PNG, GIF, or TIFF.

Table 25 includes context menu elements and descriptions for other tabs.

**Table 25** Other Tabbed Window Context Menu

Element	Description
Show Only Selected Rows	Shows only selected rows.
View Image	Displays Image Viewer for selected samples.
Configure Marks	Allows you to configure the properties of your marks.
Mark Selected Rows   <Add New>	Creates a new mark and marks selected rows.
Select Marked Rows	Selects marked rows.
Clear Marks <All>	Clears all marks.

Table 26 includes **Project** window context menu elements and descriptions.

**Table 26** Project Window Context Menu

Element	Description
Project Window	<b>Export Project</b> —Exports a project.
	<b>Expand All</b> —Expands all project repositories in the <b>Project</b> window.
	<b>Collapse All</b> —Collapses all project repositories in the <b>Project</b> window.
	<b>Style</b> —Selects a style for your project. Available project styles include: Standard, Plain, Explorer, Navigator, Group, Office Light, and Office Dark.

Table 27 includes **Log** window context menu elements and descriptions.

**Table 27** Log Window Context Menu

Element	Description
Log	Toggles the <b>Log</b> window (visible/hidden).
Project	Toggles the <b>Project</b> window (visible/hidden).
Show All	Displays both the <b>Project</b> and <b>Log</b> windows.
Hide All	Hides both the <b>Project</b> and <b>Log</b> windows.



## Appendix A

# Sample Sheet Format

### Topics

- 144    Introduction
- 144    Data Section
- 146    Sample Sheet Template
- 146    Sample Sheet Example

## Introduction

The sample sheet is a comma delimited text file (\*.csv). It is divided into sections, indicated by lines with the section name enclosed by square brackets. The Data section is the only required section. You can also include a Header section, or other user-defined sections.

## Data Section

The first row of the Data section must indicate the column names of the data to follow. The columns can be in any order, and additional user-defined columns can be included in the file.

**Table 28** Data Section, Optional and Required Columns

Column	Description	Optional (O) or Required (R)
Sample_Name	For example, S12345. Name of the sample (used only for display in the table). BeadStudio assigns a default sample name, concatenating the sample plate and sample well names.	O
Sample_Well	For example, A01. The well within the sample plate for this sample. Used only for display in the table.	O
Sample_Plate	For example, XXXXXXXX-RNA. The barcode of the sample plate for this sample. Used only for display in the table.	O
Sample_Group	For example, Group_1 User-specified name of the sample group. Note: If Sample_Group is missing, BeadStudio creates one group with the name "Default Group."	R
Pool_ID	Not used for Direct Hyb.	
Sentrix_ID	For example, 1167988 SAM or BeadChip ID.	R

**Table 28** Data Section, Optional and Required Columns (continued)

Column	Description	Optional (O) or Required (R)
SentrixPosition	For example, R001_C001 for a SAM, A1 for a BeadChip. For SAMs, the SAM sample to which the sample is hybridized. For BeadChips, the section to which the sample is hybridized.	R
NOTES:	Figure 91 is an example of a sample sheet. Your sample sheet header may contain whatever information you choose. Your sample sheet may contain any number of columns you choose. Your sample sheet must be in a comma-delimited (*.csv) file format.	

## Sample Sheet Template

A template for a sample sheet is provided on the BeadStudio CD.

## Sample Sheet Example

	A	B	C	D	E	F	G
1	[Header]						
2	Investigator Name	<Name>					
3	Project Name	<Name>					
4	Experiment Name	Samples 1-72					
5	Date	3/29/2006					
6							
7	[Data]						
8	Sample_Name	Sample_Well	Sample_Plate	Sample_Group	Pool_ID	Sentrix_ID	Sentrix_Position
9	positive control 1	A01	Plate1	positives	GS0006501-DAP	1367517	R001_C001
10	positive control 2	A02	Plate1	positives	GS0006501-DAP	1367517	R001_C002
11	positive control 3	A03	Plate1	positives	GS0006501-DAP	1367517	R001_C003
12	positive control 4	A04	Plate1	positives	GS0006501-DAP	1367517	R001_C004
13	zero time 1	A05	Plate1	zeroes	GS0006501-DAP	1367517	R001_C005
14	zero time 2	A06	Plate1	zeroes	GS0006501-DAP	1367517	R001_C006
15	zero time 3	A07	Plate1	zeroes	GS0006501-DAP	1367517	R001_C007
16	zero time 4	A08	Plate1	zeroes	GS0006501-DAP	1367517	R001_C008
17	negative control 1	A09	Plate1	negatives	GS0006501-DAP	1367517	R001_C009
18	negative control 2	A10	Plate1	negatives	GS0006501-DAP	1367517	R001_C010
19	negative control 3	A11	Plate1	negatives	GS0006501-DAP	1367517	R001_C011
20	negative control 4	A12	Plate1	negatives	GS0006501-DAP	1367517	R001_C012
21	positive control 1	B01	Plate1	positive control 1	GS0006501-DAP	1367517	R002_C001
22	positive control 2	B02	Plate1	positive control 2	GS0006501-DAP	1367517	R002_C002
23	positive control 3	B03	Plate1	positive control 3	GS0006501-DAP	1367517	R002_C003
24	positive control 4	B04	Plate1	positive control 4	GS0006501-DAP	1367517	R002_C004
25	zero time 1	B05	Plate1	zero time 1	GS0006501-DAP	1367517	R002_C005
26	zero time 2	B06	Plate1	zero time 2	GS0006501-DAP	1367517	R002_C006
27	zero time 3	B07	Plate1	zero time 3	GS0006501-DAP	1367517	R002_C007
28	zero time 4	B08	Plate1	zero time 4	GS0006501-DAP	1367517	R002_C008
29	negative control 1	B09	Plate1	negative control 1	GS0006501-DAP	1367517	R002_C009
30	negative control 2	B10	Plate1	negative control 2	GS0006501-DAP	1367517	R002_C010
31	negative control 3	B11	Plate1	negative control 3	GS0006501-DAP	1367517	R002_C011
32	negative control 4	B12	Plate1	negative control 4	GS0006501-DAP	1367517	R002_C012
33	sample X 1	C01	Plate1	sample X 1	GS0006501-DAP	1367517	R003_C001
34	sample X 2	C02	Plate1	sample X 2	GS0006501-DAP	1367517	R003_C002
35	sample X 3	C03	Plate1	sample X 3	GS0006501-DAP	1367517	R003_C003
36	sample X 4	C04	Plate1	sample X 4	GS0006501-DAP	1367517	R003_C004
37	sample X 5	C05	Plate1	sample X 5	GS0006501-DAP	1367517	R003_C005
38	sample X 6	C06	Plate1	sample X 6	GS0006501-DAP	1367517	R003_C006

Figure 91 Sample Sheet Example

## Appendix B

# Upgrading from v1.5 to v2.X and Above

## Introduction

This appendix includes information of interest if you are upgrading from BeadStudio 1.5 to the BeadStudio Gene Expression Module, v2.X or above.

In the following sections, you will find information about these topics:

- ▶ Tasks that have changed
- ▶ Procedures that have changed
- ▶ Modifications to the software interface

## Experiment Parameters

Table 29 describes changes to the Experiment Parameters feature.

**Table 29** Experiment Parameters, 1.5 vs. 2.X

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
<p>1. Select <b>Tools   Visualize Existing Data</b>. 2. Choose the input file. 3. Create a scatter plot. 4. Click <b>OK</b>.</p> <p>The Experiment Parameters appear in the BeadStudio Scatter Plot dialog box.</p>	<p>Experiment parameters are always visible in the status bar at the bottom of the BeadStudio main window.</p> <p>In addition, you can:</p> <ul style="list-style-type: none"><li>• Select <b>Analysis   Manage Analyses</b>. A tree of your analysis data appears.</li><li>• Choose a groupset and analysis from the tree.</li></ul> <p>The analysis parameters appear in the Analysis Properties area.</p>

## Scatter Plot

Table 30 describes changes to the scatter plot.

**Table 30** Scatter Plots, 1.5 vs. 2.X

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
You can create a scatter plot for Average Intensity data.	You can create a scatter plot for data from any subcolumn in the corresponding table, and any column can be a label for a data point.

## Finding Genes

Table 31 describes changes to the **Tools | Find Genes** feature.

*Table 31 Tools | Find Genes, 1.5 vs. 2.X*

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
Tools   Find Genes	Tools   Find Items

## Clustering Selected Genes

Table 32 describes changes to the **Tools | Cluster Selected Genes** feature.

*Table 32 Tools | Cluster Selected Genes, 1.5 vs. 2.X*

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
Tools   Cluster Selected Genes	1. Select <b>Show Selected Rows</b> . 2. Go to the table. 3. <b>Show Only Selected Rows</b> . 4. <b>Analysis   Run Cluster Analysis</b> . or Click <b>Run Cluster Analysis</b> on the main toolbar.

## Filtering Selected Genes

Table 33 describes changes to the Filter Selected Genes feature.

**Table 33** Filter Selected Genes, 1.5 vs. 2.X

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
<p>1. Select genes.</p> <p>2. Right-click and filter selected or unselected genes.</p>	<p>1. Click <b>Show Scatter Plot</b>.</p> <p>2. Click <b>Select Region</b>.</p> <p>3. Click <b>Mark Selected Items   Add New</b> to create a new mark. The selected rows are marked but the unselected rows also appear in the scatter plot.</p> <p>or</p> <p>1. Select genes.</p> <p>2. Go to the table.</p> <p>3. Right click and select <b>Show Only Selected Rows</b> from the context menu.</p> <p>4. Close the Scatter Plot.</p> <p>5. Create a new Scatter Plot.</p> <p>Only the selected rows appear in the scatter plot.</p>

## Selected List

Table 34 describes changes to the Selected List feature.

**Table 34** Selected List, 1.5 vs. 2.X

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
<b>Tools   Selected List</b>	<p>When items are marked, a new context menu option is available:</p> <p><b>Marked List</b></p> <ul style="list-style-type: none"> <li>► <b>View in Web Browser</b></li> <li>► <b>Save in Text File</b></li> <li>► <b>Show Item Labels</b></li> </ul>

## Select by Detection

Table 35 describes changes to the Select by Detection feature.

*Table 35* Select by Detection, 1.5 vs. 2.X

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
Scatter Plot   Tools	<ol style="list-style-type: none"><li>1. Select whatever rows you want in the table. You can do this by filtering, or by sorting the detection p-value and manually selecting the rows with values you want to include.</li><li>2. Create a scatter plot.</li></ol>

## Show R<sup>2</sup>

Table 36 describes changes to the Show R<sup>2</sup> feature.

*Table 36* Show R<sup>2</sup>, 1.5 vs. 2.X

BeadStudio 1.5	BeadStudio 2.X Gene Expression Module
Toggle R <sup>2</sup> on and off	R <sup>2</sup> is always displayed in the status bar of the BeadStudio main window.

## Dendrogram Changes

Dendograms have changed in the BeadStudio Gene Expression module since BeadStudio 1.5. Changes to dendograms correspond to the scatter plot changes described in Scatter Plot on page 148 of this chapter.



# Appendix C

# Troubleshooting

## Introduction

Use this troubleshooting guide to assist you with any questions you may have about the BeadStudio Gene Expression Module.

## Frequently Asked Questions

Table 37 lists frequently asked questions and associated responses.

*Table 37* Frequently Asked Questions

#	Question	Response
1.	What is the difference between a group and a groupset?	A <b>group</b> is a collection of arrays combined according to experimental criteria (e.g., arrays hybridized to similar or replicated samples). A <b>groupset</b> is a collection of groups.
2.	What is the minimum statistically-significant detectable fold change of Gene Expression BeadChips?	1.35 fold for single replicates Much lower for multiple replicates
3.	Over what range of intensities can I detect the minimum significant detectable fold change?	The intensity range over which a fold change of ~1.35 are significantly distinguishable is >3 logs.
4.	How can I calculate a p-value from the diff score?	$p = 1/(10^{(diff\ score/10*(sgn(\mu_{cond} - \mu_{ref}))})}$

**Table 37** Frequently Asked Questions (continued)

#	Question	Response
5.	What do the designators "A," "S," and "I" mean in the manifest files?	<p>For transcripts with a single isoform, we design "-S" probes (S=single) For transcripts with multiple isoforms, we design two types of probes:</p> <ul style="list-style-type: none"> <li>• "-I" (I=isoform-specific) are probes designed to query only one of multiple isoforms</li> <li>• "-A" (A=all) are probes designed to query all known isoforms of that transcript</li> </ul>
6.	How do I determine which genes are accurately detected?	Filter the genes using the detection p-value. Setting detection at 0.01 means that you have a 1% false positive rate. 0.05 is a commonly-used cut-off.
7.	What do the column headers "GeneSymbol," "GID," and "Accession" reference on the gene list for Illumina's standard BeadChips, and where do the numbers come from?	Descriptions for all of the column headers can be found in the document "Bead Manifest Field Descriptors" located on the documentation CD included in the startup kit.
8.	Is it possible to get the data for each feature on the BeadChip?	Yes, this is known as bead-level data. Contact your Field Application Scientist (FAS) or Tech Support for assistance.
9.	What is a Diff Score and how can I use it to get the p-value I want?	<p>The Diff Score is a transformation of the p-value that provides directionality to the p-value based on the difference between the average signal in the reference group vs. the comparison group. The formula is:</p> $\text{DiffScore} = 10^{\text{sgn}(\mu_{\text{ref}} - \mu_{\text{cond}})} * \log_{10}(p)$ <p>For a p-value of 0.05, Diff Score = <math>\pm 13</math>      For a p-value of 0.01, Diff Score = <math>\pm 20</math>      For a p-value of 0.001, Diff Score = <math>\pm 30</math></p>