

BeadStudio Genotyping 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.

The BeadStudio Genotyping Module described herein is covered by U.S. Patent No. 7,035,740 and pending patent applications.

© 2007 Illumina, Inc. All rights reserved.

Revision History

Revision	Date
Rev. C	February 2007
Rev. B	March 2006
Rev. A	December 2005

Table of Contents

Chapter 1	Overview	1
Introduction	2
Audience and Purpose	2
Installing the Genotyping Module	2
Genotyping Module Workflow	6
Chapter 2	Creating a New Project	7
Introduction	8
Starting the New Project Wizard	8
Choosing a Project Name and Location	10
Creating a Project	11
Selecting a Project From LIMS	12
Loading Sample Intensities Outside of LIMS	18
Using a Sample Sheet	18
Selecting Directories	22
Importing Cluster Positions	24
Chapter 3	Generating Clusters	27
Introduction	28
Automatically Excluding Samples	28
Manually Excluding Samples	29
Running the Clustering Algorithm	30
Reviewing Clusters	31
Editing Clusters	32
Redefining the Cluster	32
Excluding Samples	32
Shifting the Cluster Location	33
Changing the Cluster Height/Width	33
Exporting the Cluster File	34

Chapter 4 Viewing Your Data	37
Introduction	38
SNP Graph.....	38
Cartesian and Polar Coordinates.....	39
Normalization	40
Adjusting Axes.....	41
Selecting Samples.....	41
Tips	42
Marking Samples.....	42
Displaying the Legend	46
Excluding Samples	47
Plotting Excluded Samples	48
Customizing the SNP Table.....	49
Shading Call Regions.	51
Viewing the Controls Dashboard.....	53
Exporting Controls Data	54
Viewing the Contamination Dashboard	56
Chapter 5 Analyzing Your Data	57
Introduction	58
Importing Phenotype Information.....	58
Estimating the Gender of Selected Samples.....	60
Editing the Properties of Selected Samples	62
Analyzing Paired Samples	64
Using Concordance Features	65
Exporting Allele Calls	65
Importing Allele Calls	66
Concordance Calculations	66
Using Column Plug-Ins	66

Chapter 6	Using the Illumina Genome Viewer	69
Introduction.	70	
Viewing Data in the Illumina Genome Viewer.	70	
Getting Data Files.	71	
Plotting Data.	73	
Viewing Data in the Illumina Chromosome Browser.	78	
Navigating the Illumina Chromosome Browser	79	
Viewing Gene Information	80	
Plotting Sample Columns	82	
Viewing Project Manifest SNPs.	82	
Chapter 7	Performing LOH and Copy Number Analysis . . .	83
Introduction.	84	
B Allele Frequency	84	
Log R Ratio	86	
Algorithm Plug-ins	88	
Using Autobookmarking Algorithms.	89	
Using Column Plug-Ins	93	
Chapter 8	Generating Reports	95
Introduction.	96	
Final Report.	96	
DNA Report	104	
Locus Summary Report	107	
Locus x DNA Report	113	
Reproducibility and Heritability Report.	115	
Error Types	119	

Chapter 9	User Interface Reference	123
Introduction	124	
Detachable Docking Windows	125	
Graph Window	125	
Data Table	129	
Samples Table	139	
Project Window	145	
Log Window	146	
Main Window Menus	147	
Graph Window Toolbar	153	
Table Windows Toolbar	154	
Context Menus	156	
Appendix A	Sample Sheet Guidelines	161
Introduction	162	
Manifests Section	162	
Data Section	163	
Redos and Replicates	164	
Sample Sheet Template	164	
Sample Sheet Example	165	
Appendix B	Troubleshooting Guide	167
Introduction	168	
Frequently Asked Questions	168	

List of Figures

Figure 1	BeadStudio Application Unzipping.....	3
Figure 2	Illumina BeadStudio Installation, Modules	3
Figure 3	Illumina BeadStudio Installation, Plug-ins.....	4
Figure 4	Accept License Agreement.....	5
Figure 5	Illumina BeadStudio Installation	5
Figure 6	Installation Progress	5
Figure 7	Genotyping Analysis Workflow.....	6
Figure 8	Starting a New Project, New Project Area	8
Figure 9	Starting a New Project, File Menu	9
Figure 10	BeadStudio Project Wizard - Welcome.....	10
Figure 11	BeadStudio Project Wizard - Project Location	11
Figure 12	Select LIMS Project	12
Figure 13	Login Infinium LIMS - Setup	13
Figure 14	Login Infinium LIMS - Login	13
Figure 15	Select LIMS Project	14
Figure 16	Select LIMS Project Warning.....	15
Figure 17	Select Target Dates.....	15
Figure 18	Selecting Target Dates	16
Figure 19	Update Heritability & Reproducibility Errors.....	16
Figure 20	Evaluating Heritability.....	16
Figure 21	Sample Requeue Status	17
Figure 22	Loading Sample Intensities.....	18
Figure 23	Loading Sample Intensities Using a Sample Sheet.....	19
Figure 24	Cluster Positions	20
Figure 25	Loading Sample Intensities, Selecting Dirs.s w/ Intensity Files.	22
Figure 26	Loading Sample Intensities, Selecting Dirs. w/ Intensity Files ..	23
Figure 27	Cluster Positions	25
Figure 28	Auto-Exclude Samples	28
Figure 29	Exclude Selected Samples	29
Figure 30	Cluster all SNPs.....	30
Figure 31	Cluster All SNPs	30
Figure 32	Clustering Progress.....	30

Figure 33	Reviewing Clusters	31
Figure 34	Editing Clusters	33
Figure 35	Export Cluster Positions Selected	34
Figure 36	Save Cluster Positions	35
Figure 37	SNP Graph	38
Figure 38	P-C Error (Left), Reproducibility Error (Right)	39
Figure 39	P-C Error and Reproducibility Error Highlighted in SNP Graph	39
Figure 40	Polar Coordinates (Left) & Cartesian Coordinates (Right)	40
Figure 41	Normalization Turned Off (Left) & Normalization Turned On (Right)	40
Figure 42	SNP Graph, Selected Samples Shown in Yellow	42
Figure 43	Configure Marks	43
Figure 44	Naming a Mark	43
Figure 45	Selecting a Color for a Mark	44
Figure 46	Displaying Marked Samples	45
Figure 47	Displaying the Legend	46
Figure 48	Legend Displaying Mark Name	46
Figure 49	Excluding Selected Sample	47
Figure 50	Project Properties	48
Figure 51	Column Chooser	50
Figure 52	Shaded Call Regions	51
Figure 53	Controls Dashboard	53
Figure 54	Exporting Controls Data	54
Figure 55	Saving the Controls Dashboard	55
Figure 56	Contamination Dashboard	56
Figure 57	Importing Phenotype Information	59
Figure 58	Phenotype Information File	59
Figure 59	Selected Samples	60
Figure 60	Samples Table Context Menu	61
Figure 61	Populating the Gender Column	61
Figure 62	Selected Samples	62
Figure 63	Samples Table Context Menu	63
Figure 64	Sample Properties	63
Figure 65	SNP Graph Showing Paired Samples	65
Figure 66	Select Column Plug-In Form	67
Figure 67	Illumina Genome Viewer	71
Figure 68	Plot Selector	73
Figure 69	Data Selector	75
Figure 70	Selected Plot	76
Figure 71	Series	77
Figure 72	Illumina Chromosome Browser	79
Figure 73	Exon Listbox	80

Figure 74	Viewing Gene Exons in Detail.....	81
Figure 75	Displaying Project Manifest SNPs in the ICB	82
Figure 76	Theta vs. B Allele Frequency.....	85
Figure 77	Log R Ratio	87
Figure 78	Selecting a Report Plug-In	88
Figure 79	Illumina Genome Viewer.....	89
Figure 80	Autobookmark Analysis	90
Figure 81	Selecting Samples for Analysis	91
Figure 82	Selecting Chromosomes for Analysis	92
Figure 83	Autobookmark Analysis	92
Figure 84	Analysis is Complete.....	93
Figure 85	Select Column Plug-In Form	94
Figure 86	Report Type	97
Figure 87	Final Report Format	98
Figure 88	Final Report - Standard Format Options.....	99
Figure 89	Final Report - Matrix Format Options.....	100
Figure 90	Final Report - 3rd Party Options.....	101
Figure 91	Destination	102
Figure 92	Sample Final Report	103
Figure 93	DNA Report Selected	104
Figure 94	Destination	105
Figure 95	Sample DNA Report	106
Figure 96	Locus Summary Report Selected	108
Figure 97	Destination - Locus Summary	109
Figure 98	Sample Locus Summary Report	110
Figure 99	Locus x DNA Selected	113
Figure 100	Destination - Locus x DNA	114
Figure 101	Reproducibility and Heritability.....	116
Figure 102	View Reproducibility and Heritability	116
Figure 103	Sample Reproducibility and Heritability Report	117
Figure 104	BeadStudio Genotyping Module Default View.....	124
Figure 105	SNP Graph	125
Figure 106	Sample Graph.....	126
Figure 107	Errors Table.....	127
Figure 108	SNP Graph Alt.....	129
Figure 109	Full Data Table	130
Figure 110	SNP Table	133
Figure 111	Paired Samples Table	137
Figure 112	Samples Table.....	140
Figure 113	Project Window.....	146
Figure 114	Log Window	146
Figure 115	Sample Sheet Example.....	165

List of Tables

Table 1	Illumina Genome Viewer Toolbar Buttons & Functions	72
Table 2	DNA Report - File Column Descriptions.....	106
Table 3	Locus Summary Report - File Column Descriptions	110
Table 4	Locus by DNA Report - File Column Descriptions	115
Table 5	Reproducibility and Heritability Report - File Column Desc. .	117
Table 6	Reproducibility Errors	119
Table 7	Parent-Child Heritability Errors	119
Table 8	Parent-Parent-Child Heritability Errors	120
Table 9	Errors Table Columns	127
Table 10	Full Data Table Columns.....	130
Table 11	Full Data Table Per-Sample Subcolumns	131
Table 12	SNP Table Columns	133
Table 13	Paired Samples Table Columns	137
Table 14	Paired Samples Table Per-Pair Subcolumns	138
Table 15	Samples Table Columns	140
Table 16	Samples Table Per-Manifest Subcolumns.....	143
Table 17	Log Window Options	146
Table 18	File Menu Functions	147
Table 19	Edit Menu Functions.....	149
Table 20	View Menu Functions	150
Table 21	Analysis Menu Functions.....	150
Table 22	Tools Menu Functions.....	152
Table 23	Windows Menu Functions.....	152
Table 24	Help Menu Functions	153
Table 25	Graph Window Toolbar Buttons & Functions.....	153
Table 26	Table Windows Toolbar Buttons & Functions.....	155
Table 27	Graph Window Context Menu	156
Table 28	Full Data Table Context Menu	157

Table 29	SNP Table Context Menu	158
Table 30	Samples Table Context Menu.....	158
Table 31	Error Table Context Menu.....	160
Table 32	Data Section, Required and Optional Columns	163
Table 33	Frequently Asked Questions	168

Chapter 1

Overview

Topics

- 2 Introduction
- 2 Audience and Purpose
- 2 Installing the Genotyping Module
- 6 Genotyping Module Workflow

Introduction

This manual describes Illumina's BeadStudio 3.0 Genotyping Module. The BeadStudio Genotyping Module is used to analyze data collected using Illumina's GoldenGate® and Infinium® genotyping assays.

Audience and Purpose

This guide is written for researchers who want to use the BeadStudio Genotyping Module to analyze data generated by performing Illumina's GoldenGate or Infinium assays.

This guide includes procedures and user interface information specific to the BeadStudio Genotyping 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 Genotyping Module

To install the BeadStudio Genotyping 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.
- The BeadStudio application suite unzips (Figure 1).

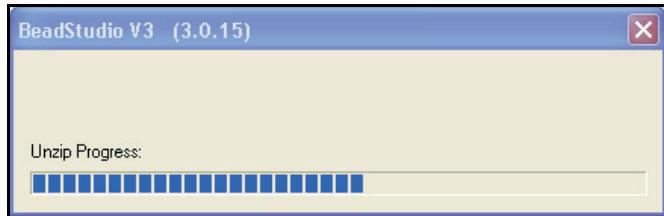


Figure 1 BeadStudio Application Unzipping

The **Illumina BeadStudio Installation** screen appears (Figure 2).

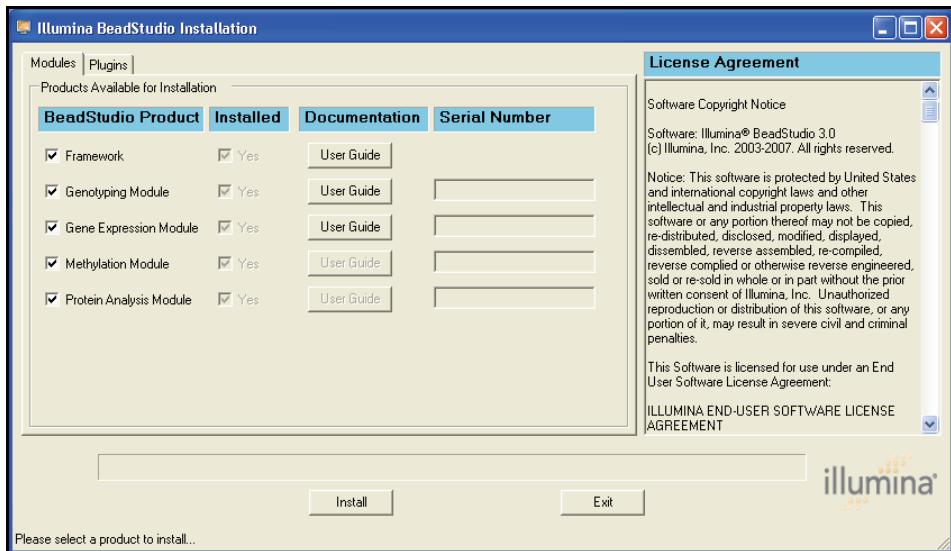


Figure 2 Illumina BeadStudio Installation, Modules

3. Read the software license agreement in the right-hand side of the **Illumina BeadStudio Installation** screen.
4. In the **BeadStudio Product** area, select **Genotyping 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 Genotyping 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. **[Optional]** Install analysis algorithm plug-ins by performing the following steps:
- Click the **Plug-ins** tab (Figure 3).
The **Plug-ins** tab lists the analysis algorithm plug-ins you can choose to install with BeadStudio.
 - Select the plugins you want to install.

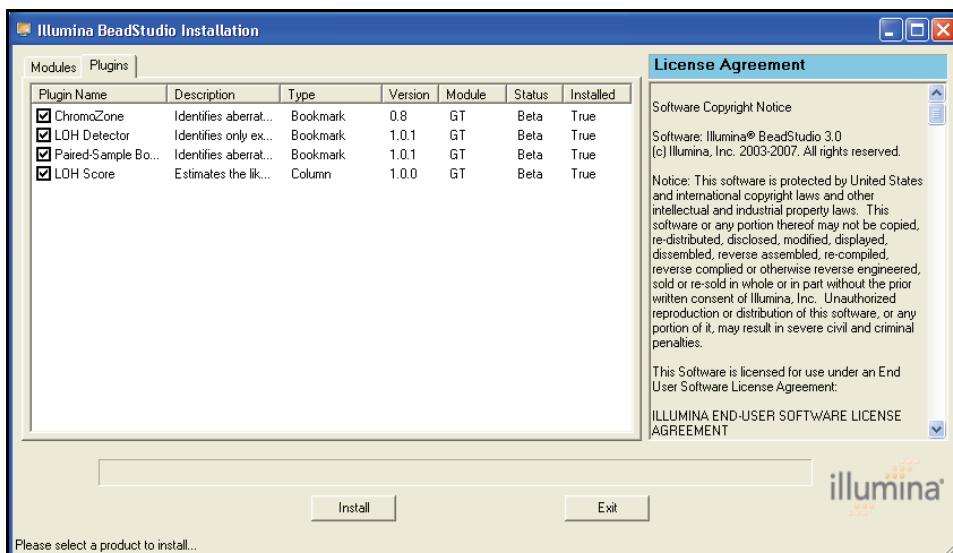


Figure 3 Illumina BeadStudio Installation, Plug-ins

8. Click **Install**.

The **Accept License Agreement** dialog box appears (Figure 4).



Figure 4 Accept License Agreement

9. Click **Yes** to accept the software license agreement. The Genotyping Module is installed on your computer, along with any additional BeadStudio modules or plug-in algorithms you selected (Figure 5).

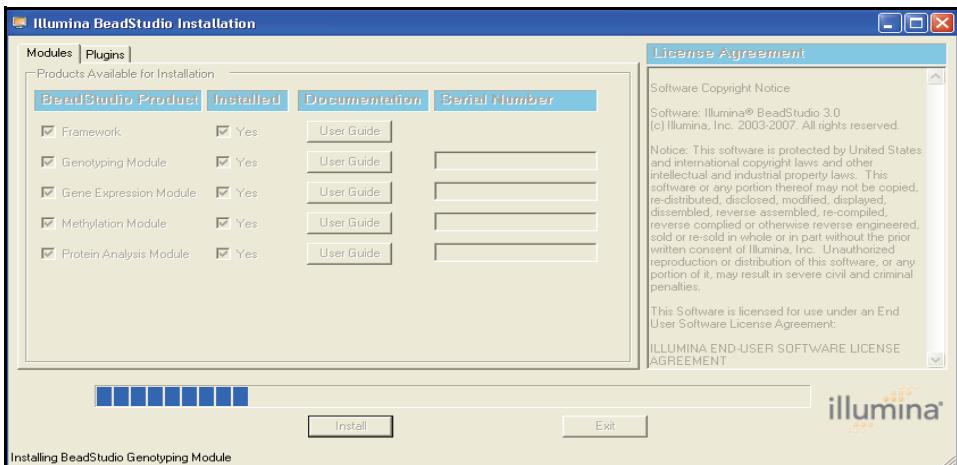


Figure 5 Illumina BeadStudio Installation

The **Installation Progress** dialog box notifies you that installation is complete (Figure 6).

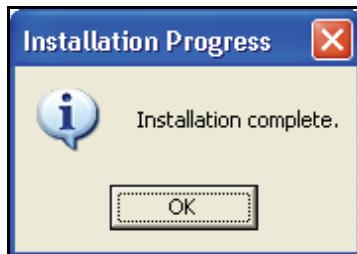


Figure 6 Installation Progress

10. Click **OK**.

11. On the **Illumina BeadStudio Installation** screen (Figure 5), click **Exit**.

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

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

Genotyping Module Workflow

The basic workflow for genotyping analysis using Illumina's BeadStudio Genotyping Module is summarized in Figure 7.

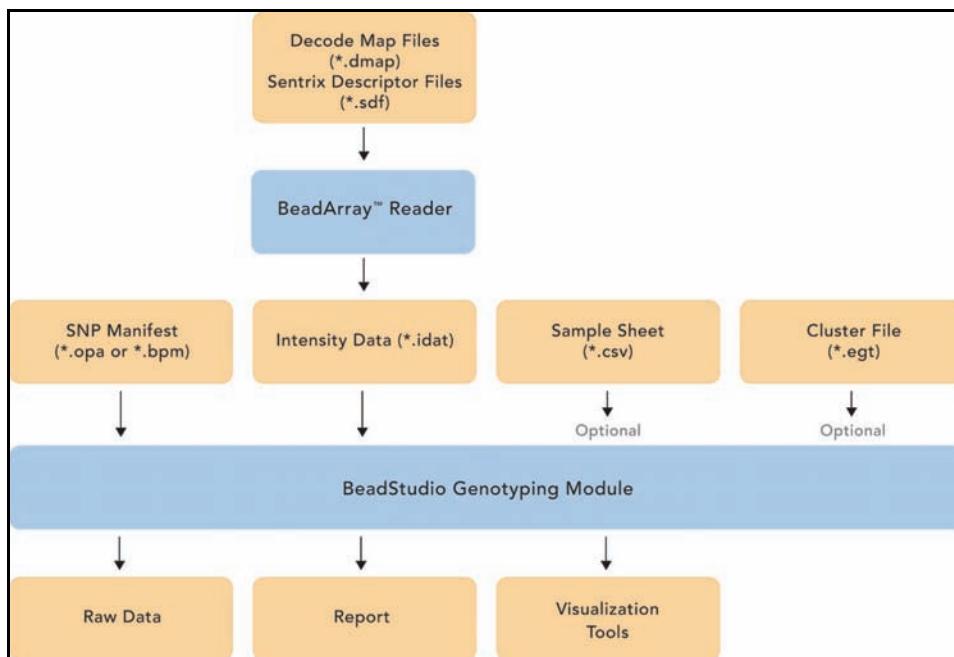


Figure 7 Genotyping Analysis Workflow

Chapter 2

Creating a New Project

Topics

- 8 Introduction
- 8 Starting the New Project Wizard
- 10 Choosing a Project Name and Location
- 11 Creating a Project
- 12 Selecting a Project From LIMS
- 18 Loading Sample Intensities Outside of LIMS
- 18 Using a Sample Sheet
- 22 Selecting Directories
- 24 Importing Cluster Positions

Introduction

The New Project Wizard offers an easy way to start a new project from within any BeadStudio module you install. The following sections describe how to use the New Project Wizard to begin a new genotyping project. Follow the same instructions to create projects that allow you to perform LOH or copy number analyses.

Starting the New Project Wizard

To create a new genotyping project:

1. Do one of the following:
 - Select **Start | Program Files | Illumina | BeadStudio.**

- Double-click the BeadStudio icon on the desktop.



The BeadStudio application launches and the **Start** page appears.

2. On the BeadStudio **Start** page, do one of the following:
 - In the **New Project** pane, click **Genotyping**.

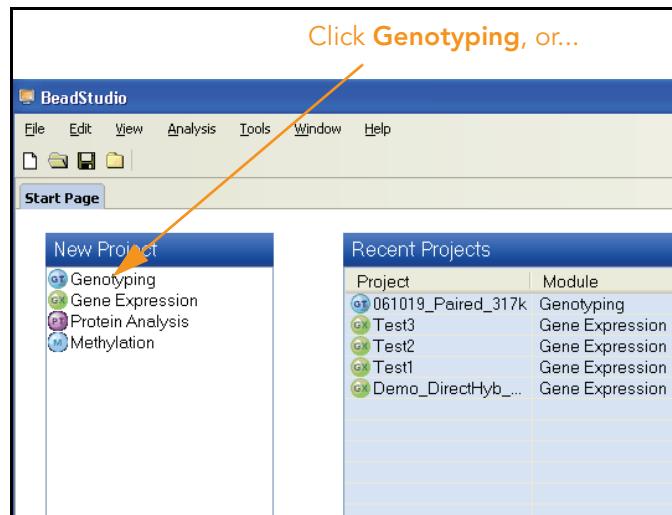


Figure 8 Starting a New Project, New Project Area

- Select **File | New Project | Genotyping** (Figure 9).

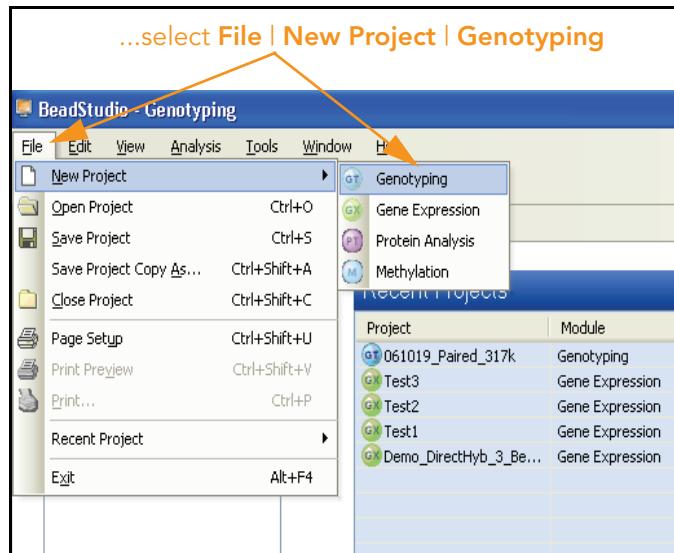


Figure 9 Starting a New Project, File Menu

The **BeadStudio Project Wizard - Welcome** dialog box appears (Figure 10).

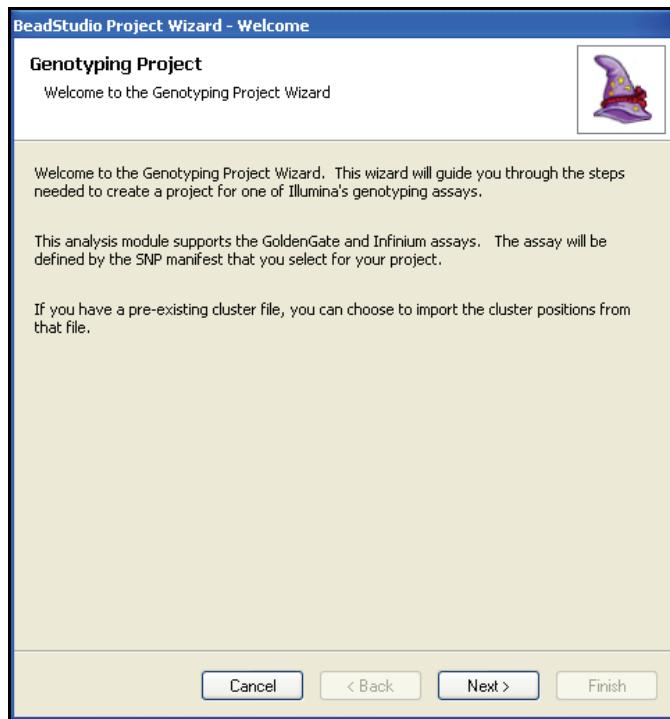


Figure 10 BeadStudio Project Wizard - Welcome

3. Click **Next** to advance to the **Project Location** dialog box.

Choosing a Project Name and Location

In the **BeadStudio Project Wizard - Project Location** dialog box (Figure 11), you must choose a project repository (the directory where you will store your projects). Each project is saved in a subdirectory that is given the same name as the project. All project-related files are saved within each project's subdirectory. The main project file is given a *.bsc file extension.

Additionally, you can choose whether you want to create a new project or whether you want to select an existing project from the Laboratory Information Management System (LIMS).

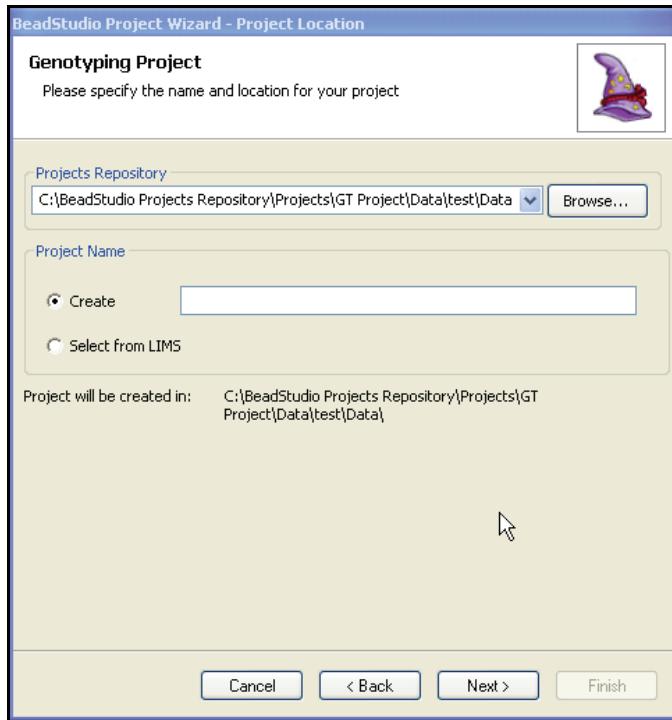


Figure 11 BeadStudio Project Wizard - Project Location

Creating a Project

To create a new project:

1. Browse to the project repository where you want to store your project.
2. Choose one of the following options:
 - If you want to select a project from LIMS, continue to [Selecting a Project From LIMS](#).
 - If you want to load sample intensities outside of LIMS, perform the following steps:
 - a. Type a name for your project in the **Project Name** text box.
 - b. Click **Next** to advance to the **Loading Sample Intensities** dialog box.
 - c. Continue to [Loading Sample Intensities Outside of LIMS](#) on page 18.

Selecting a Project From LIMS

To select a project from LIMS:

1. In the **BeadStudio Project Wizard - Project Location** dialog box (Figure 11), choose **Select from LIMS**.
2. The **BeadStudio Project Wizard - Select LIMS Project** dialog box appears (Figure 12).

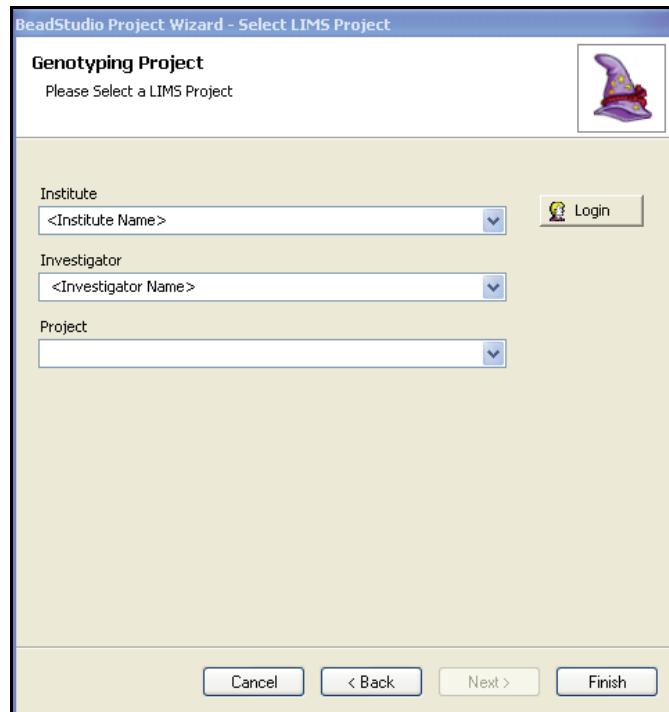


Figure 12 Select LIMS Project

3. Click the **Setup** tab (Figure 13).

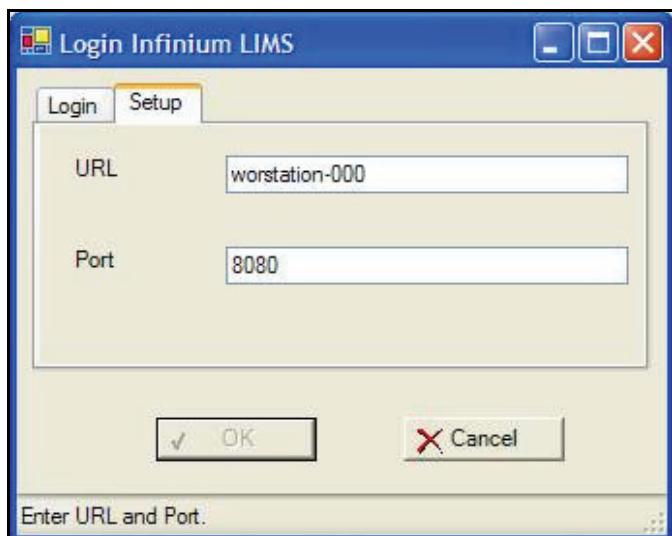


Figure 13 Login Infinium LIMS - Setup

4. In the **Setup** tab, enter the following:
 - URL
 - Port Number
5. Select the **Login** tab (Figure 14).

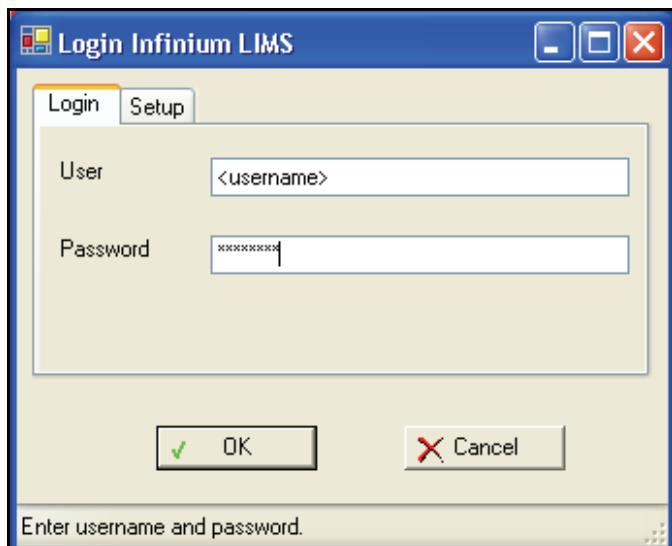


Figure 14 Login Infinium LIMS - Login

6. Enter your username and password.

7. Click **OK**.

The **Login Infinium LIMS** dialog box closes.

You are returned to the **Select LIMS Project** dialog box (Figure 15).

8. On the **Select LIMS Project** dialog box, make the following selections from the dropdown menus:

- **Institute**
- **Investigator**
- **Project**

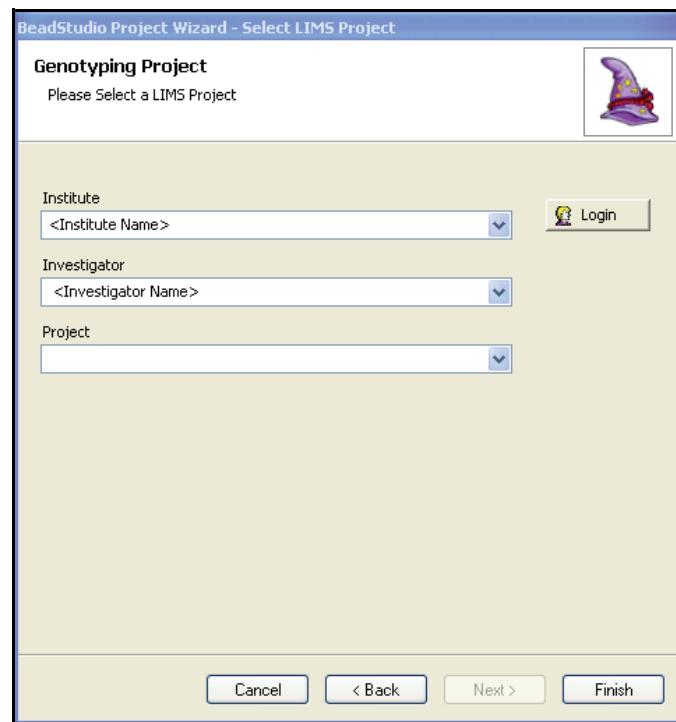


Figure 15 Select LIMS Project

If you have loaded information for a pre-existing project, the warning shown in Figure 16 appears.

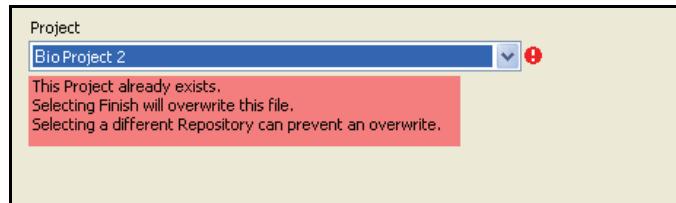


Figure 16 Select LIMS Project Warning

If you do not want to overwrite existing projects files, select different options in the **Select LIMS Project** dialog box.

9. Click **Finish**.

The **Select Target Dates** dialog box appears (Figure 17).



Figure 17 Select Target Dates

10. [Optional] Select **Use Start Date** and choose a start date in the calendar on the left (Figure 18).

11. [Optional] Select **Use End Date** and choose an end date in the calendar on the right (Figure 18).

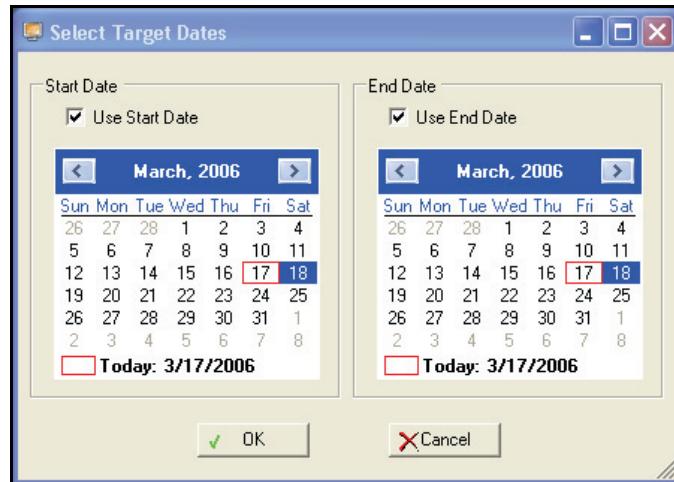


Figure 18 Selecting Target Dates

12. Click OK.

The manifests load, the clusters are imported, and the SNP statistics are calculated.

The **Do you wish to update all heritability and reproducibility errors?** dialog box appears (Figure 19).

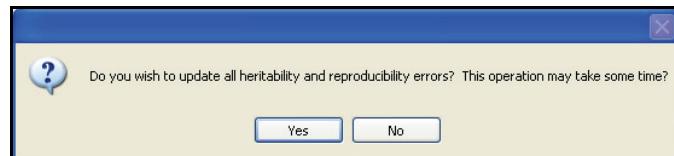


Figure 19 Update Heritability & Reproducibility Errors

If you click **Yes**, the **Evaluating Heritability** status bar appears (Figure 20) and heritability and reproducibility are calculated.



Figure 20 Evaluating Heritability

SNP data are saved, and the **Sample Requeue Status Change** message appears (Figure 21).

This message indicates whether any sample statuses have changed between the BeadStudio project and the LIMS database. If sample statuses are updated, this is reflected in BeadStudio.

If the data from the BeadStudio project and the LIMS database are the same, the **Sample Requeue Status Change** dialog box displays the message “No updates were required.”



Figure 21 Sample Requeue Status

13. Click OK.

The project you selected loads from LIMS and displays in the BeadStudio Genotyping Module.

Loading Sample Intensities Outside of LIMS

If you are not using a LIMS database for loading intensity data, you have two options for loading data outside of LIMS control:

- ▶ Loading sample intensities using a sample sheet (page 18)
- ▶ Loading samples by selecting directories that contain intensity data files (page 22).

Using a Sample Sheet To load intensities using a sample sheet:

1. In the **Loading Sample Intensities** dialog box, select **Use sample sheet to load sample intensities** (Figure 22).

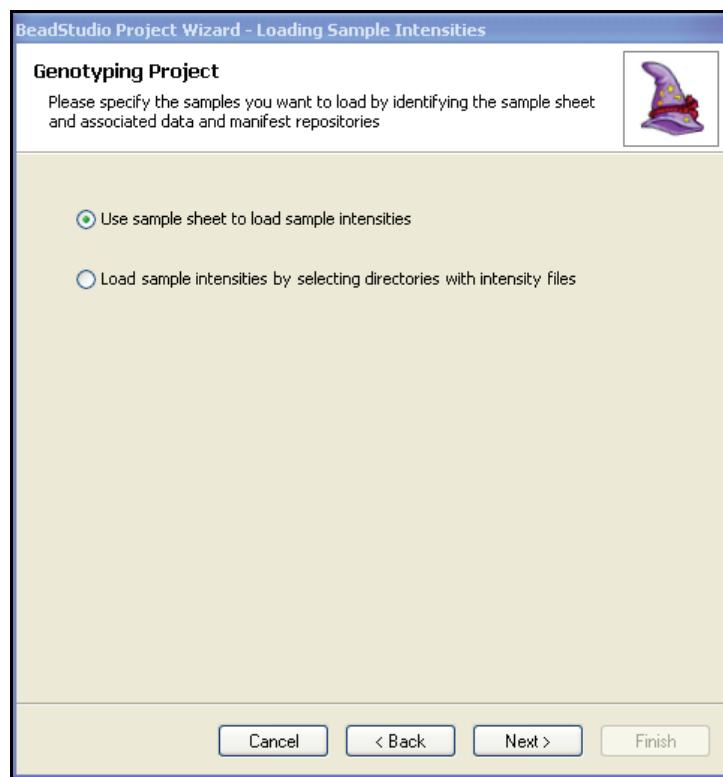


Figure 22 Loading Sample Intensities

2. Click **Next**.

The **Loading Sample Intensities** dialog box appears (Figure 23).

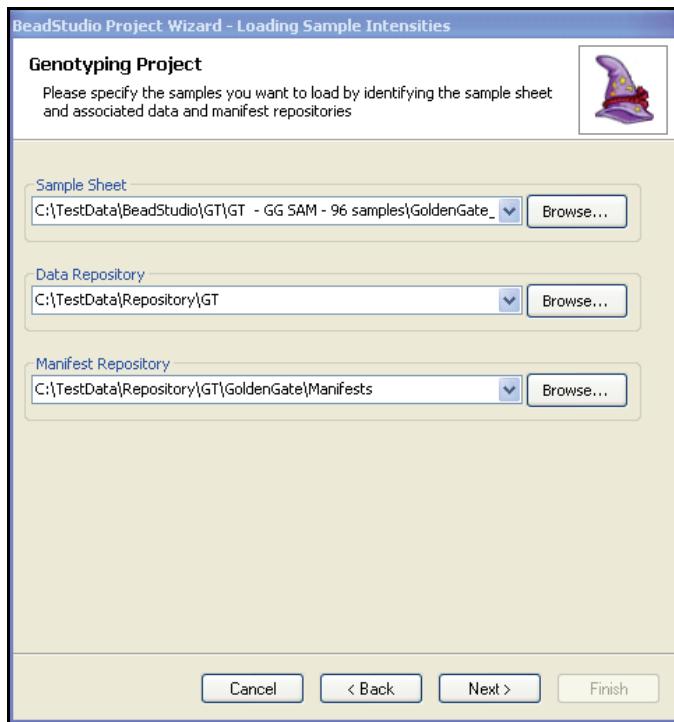


Figure 23 Loading Sample Intensities Using a Sample Sheet

3. Browse to select the following items:

- **Sample Sheet**
- **Data Repository**
- **Manifest Repository**

The **Sample Sheet** is a comma-delimited text file (.csv file). Its format is described in Appendix A of this document..

The **Data Repository** is the directory that contains the directories that contain your intensity (*.idat) files.

The **Manifest Repository** is the directory that contains your SNP manifests. This directory is necessary because the name(s) of the SNP manifests are contained in the sample sheet, and the BeadStudio Genotyping Module needs to know where to find them.

To select a sample sheet, data repository, and manifest repository:

1. Browse to the locations of your sample sheet, data repository, and manifest repository.
2. Click **Next**.

The **Cluster Positions** dialog box appears (Figure 24).

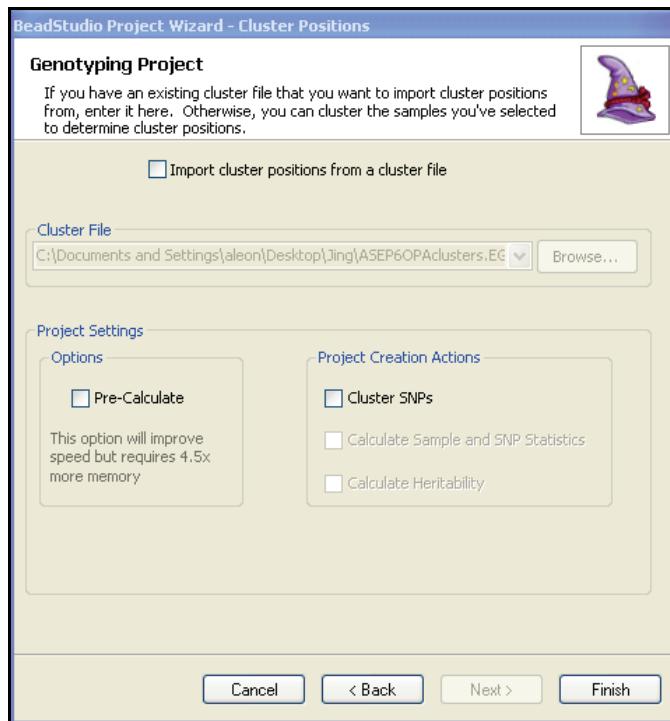


Figure 24 Cluster Positions

The number of samples that can be loaded into physical memory varies depending upon many factors, including how many other programs are running on your computer simultaneously, and the configuration of your virtual memory.

Use the following guidelines for a computer with the recommended minimum 2 GB of physical memory:

For HumanHap300 data:

- Approximately 200 samples of HumanHap300 SNP data can be loaded using memory-based storage.

- If you want to load more than 200 samples of HumanHap300 data, leave the **Precalculate** checkbox cleared to optimize memory.
- If you want to load fewer than 200 samples of HumanHap300 data, you may want to select **Precalculate** to optimize calculation speed.

For HumanHap550 data:

- ▶ Approximately 150 samples of HumanHap550 SNP data can be loaded using memory-based storage.
 - If you want to load more than 150 samples of HumanHap 550 data, leave the **Precalculate** checkbox cleared to optimize memory.
 - If you want to load fewer than 150 samples of HumanHap550 data, you may want to select **Precalculate** to optimize calculation speed.

3. In the **Project Settings** area, choose one of the following options:

- Select **Precalculate** if you expect the number of samples and SNPs to fit within the physical memory of your computer, and you want to increase calculation speed.
- Leave the **Precalculate** checkbox cleared if you do not expect the number of samples and SNPs you want to load to fit within the physical memory of your computer.



NOTE

You must choose whether to enable precalculation in a project at the time the project is created. You cannot change this option later in an existing project.

4. **[Optional]** In the **Project Creation Actions** area, select the following option for your project:

- **Cluster SNPs**

If you choose to cluster all SNPs, you may also select one or both of the following options:

- **Calculate Sample and SNP Statistics**
- **Calculate Heritability**

After loading intensity data using a sample sheet, continue to *Importing Cluster Positions* on page 24.

Selecting Directories

To load intensities by selecting directories:

1. In the Loading Sample Intensities dialog box, select **Load Sample Intensities by Selecting Directories with Intensity Files** (Figure 25).

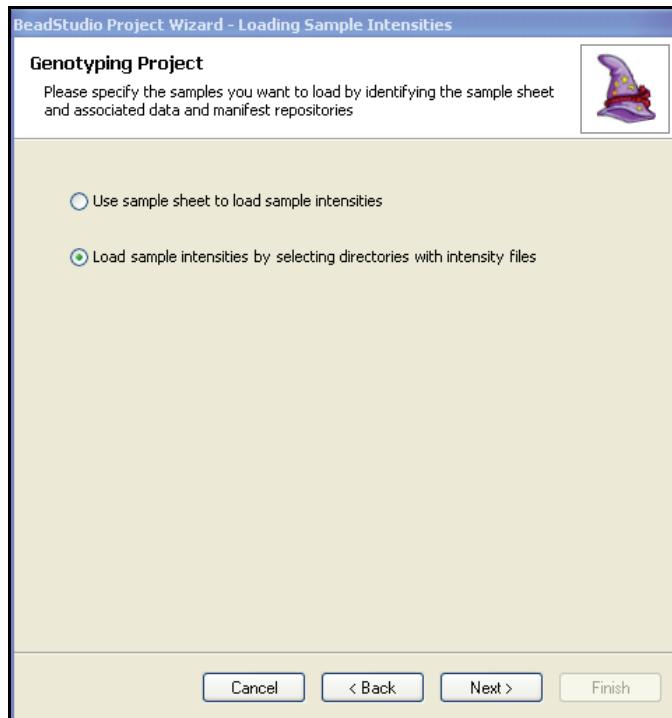


Figure 25 Loading Sample Intensities by Selecting Directories with Intensity Files

2. Click **Next**.

The **Loading Sample Intensities** dialog box appears (Figure 26).

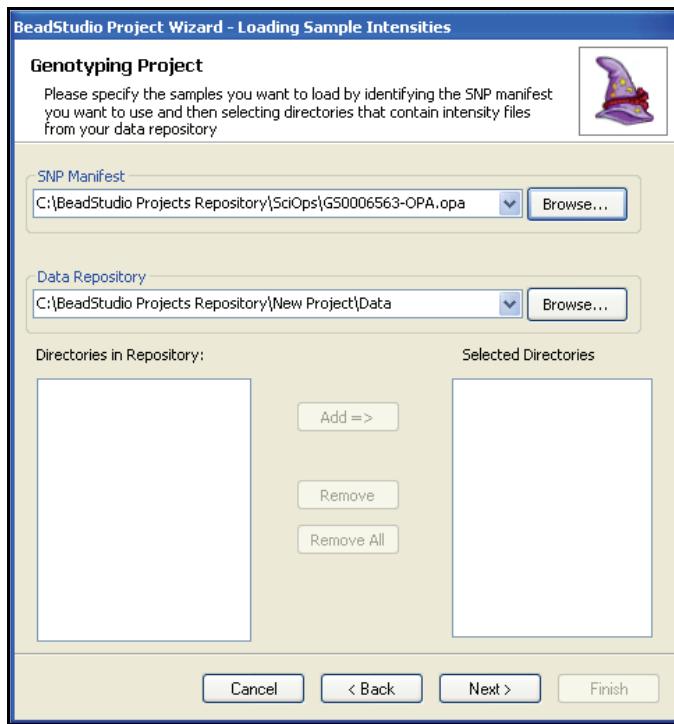


Figure 26 Loading Sample Intensities by Selecting Directories with Intensity Files

3. Select the following items:
 - **SNP Manifest**—an *.opa file for GoldenGate assays, or a *.bpm file for Infinium assays. The SNP manifest contains the mapping between bead-type identifier and SNP.
 - **Data Repository**—the directory that contains subdirectories with intensity files. When you change the entry in the data repository field, the **Directories in Repository** list box is populated with the directories contained in your repository.

To select the intensity files you want to load:

1. Browse to the SNP manifest and data repository you want to use.
2. Click on one or more directories in the **Directories in Repository** list box.

3. Click **Add** to add the directories to the project.

The directories appear in the **Selected Directories** listbox as you choose them.

All intensity files (*.idat files) contained within the selected directories are loaded and added to the project.

**NOTE**

If you are using LIMS, if the manifest name contained in the *.idat file does not match the name of the manifest you have loaded, that intensity file will be skipped.

4. Click **Next** to advance to the **Cluster Positions** dialog box.

Importing Cluster Positions

The **Cluster Positions** dialog box is the final screen of the BeadStudio Project Wizard (Figure 27). From this screen, you can import a cluster file and choose to use these cluster definitions to call genotypes for your samples.

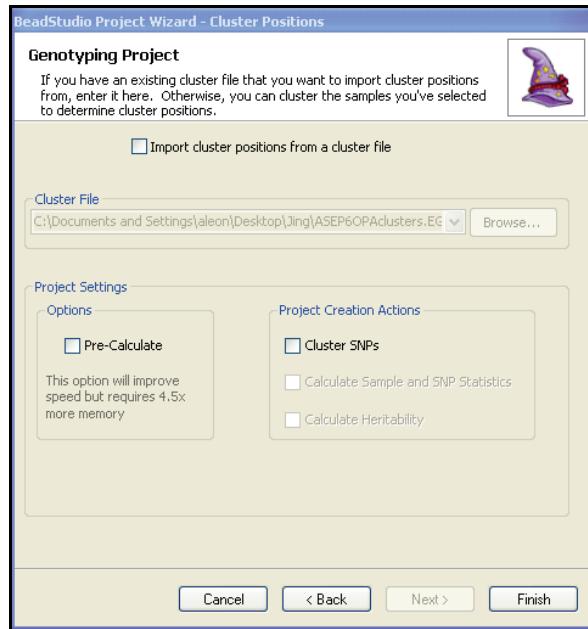


Figure 27 Cluster Positions

To import a cluster file:

1. Select **Import cluster positions from a cluster file**.
2. Browse to the cluster file you want to use.



NOTE

If you do not want to import a cluster file, clear the **Import cluster positions from a cluster file** checkbox and the **Cluster File** text field.

3. Select **Precalculate** if you want to optimize your project for speed based on the memory capabilities of your computer.
4. **[Optional]** In the **Project Creation Actions** area, select the following option for your project:
 - **Cluster SNPs**If you choose to cluster all SNPs, you may also select one or both of the following options:
 - **Calculate Sample and SNP Statistics**
 - **Calculate Heritability**

5. Click **Finish** to complete the wizard.

The Genotyping Module loads your intensity files.

If you did not load a cluster file, continue to Chapter 3,
Generating Clusters.

If you loaded a cluster file, go to Chapter 4,
Viewing Your Data.

Chapter 3

Generating Clusters

Topics

- 28 Introduction
- 28 Automatically Excluding Samples
- 29 Manually Excluding Samples
- 30 Running the Clustering Algorithm
- 31 Reviewing Clusters
- 32 Editing Clusters
 - 32 Redefining the Cluster
 - 32 Excluding Samples
 - 33 Shifting the Cluster Location
 - 33 Changing the Cluster Height/Width
- 34 Exporting the Cluster File

Introduction

Illumina's assays require cluster locations in order to generate the most accurate genotype calls. This is because the locations of the heterozygote and homozygotes for each SNP, though reproducible, can vary from SNP to SNP.

Given a population of samples that exhibit the three genotypes for every SNP, the BeadStudio Genotyping Module can automatically determine the cluster positions of the genotypes. If certain SNPs have one or two clusters that lack representation, the BeadStudio Genotyping Module can estimate the missing cluster positions.

One common question is: How large does the population of samples need to be? This depends on the minor allele frequency of the SNPs. The lower the minor allele frequency, the more samples are required to achieve representation of all clusters. A population of 100 or more samples is typically recommended.

Automatically Excluding Samples

Some samples may be of poor quality in some regard (e.g., they may not have hybridized well). In this case, you would not want to include them in your clustering.

To automatically exclude samples, select **Analysis | Auto-Exclude Samples** (Figure 28). The BeadStudio Genotyping Module evaluates each sample and determines whether it should be excluded.

The criteria the BeadStudio Genotyping Module uses to determine whether to exclude a sample are based on SNP intensities for the sample.

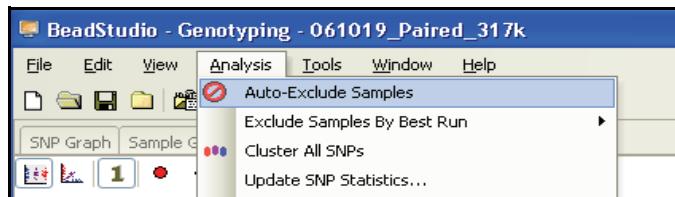


Figure 28 Auto-Exclude Samples

Manually Excluding Samples

You can also manually include or exclude samples. To manually exclude samples, perform the following steps:

1. In the **Samples Table** or **SNP Graph**, select the sample(s) you want to exclude.
2. Right-click to bring up the context menu.
3. Select **Exclude Selected Samples** (Figure 29).

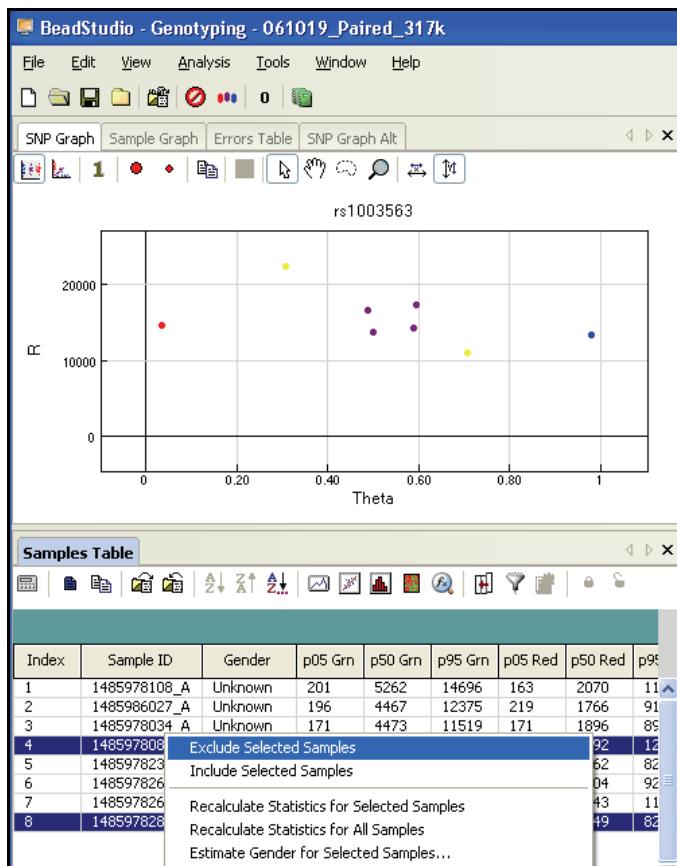


Figure 29 Exclude Selected Samples

You can use the **Sample Graph** to evaluate sample quality. If you click on a sample in the samples table, all of the SNPs for that sample are plotted in the **Sample Graph**.

Running the Clustering Algorithm

1. To run the clustering algorithm, do one of the following:

- Select **Analysis | Cluster All SNPs** (Figure 30)
- Click **Cluster all SNPs** .

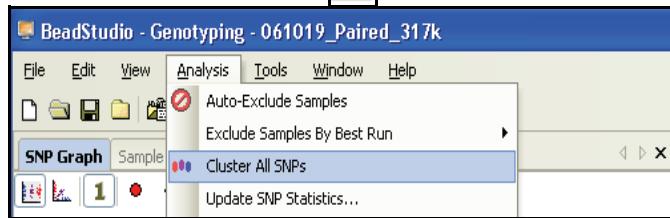


Figure 30 Cluster all SNPs

2. The **Are you sure you want to recluster all SNPs?** message appears (Figure 31).

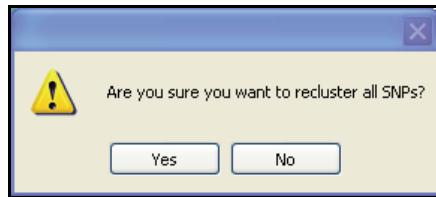


Figure 31 Cluster All SNPs

3. The clustering algorithm runs, and the **BeadStudio Progress Status** bar appears (Figure 32).

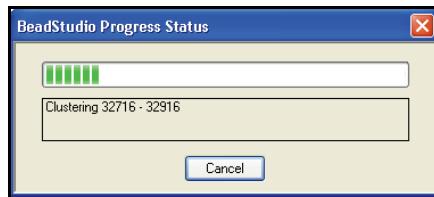


Figure 32 Clustering Progress

When the **BeadStudio Progress Status** bar disappears, your samples have been reclustered.

Reviewing Clusters

To review clusters:

- Click  **Normalization** to view normalized data (recommended).

The BeadStudio Genotyping Module displays the cluster ovals that represent the location of the clusters with two standard deviations.

For more information about normalization, see *Normalization* on page 40.

To shade the calling regions:

- Click  **Shade Calling Regions**.

The calling regions are shaded in the **SNP Graph** (Figure 33).

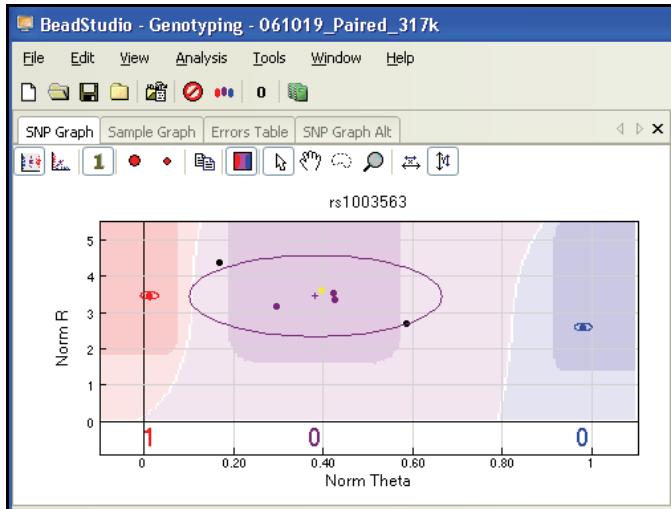


Figure 33 Reviewing Clusters

For more information about shading call regions, see *Shading Call Regions* on page 51.

Samples are colored according to their genotype call. Samples in the lighter shaded regions fall below the user-specified Call Score Threshold set in **Tools | Options | Project**, and are colored black to indicate that they are classified as "No Calls."

Note that you do not have to review all of your SNPs. You can sort by GenTrain score in the **SNP Table** and only review those SNPs that have the poorest clustering. Alternatively, if you have entered reproducibility or heritability relationships, you can sort by heritability or reproducibility errors (**Rep**, **P-C**, **P-P-C**) in the **SNP Table** and review only SNPs that exhibit errors.

For more information about sorting, see *Data Table* on page 129.

Editing Clusters

If, after reviewing the clustering of a SNP, you feel that the loaded cluster file or automated algorithm did not accurately calculate the cluster positions, you can manually edit the cluster locations in various ways.

Redefining the Cluster

To redefine the cluster using samples you select:

1. Select samples in the graph.
2. Right-click to display the context menu.
3. Select **Define AB (or AA, or BB) cluster using selected samples**.

The cluster's location and size are calculated based on the samples you have selected. The remaining samples are reclustered.

Excluding Samples

To exclude samples in the current graph:

1. Select samples in the graph.
2. Right-click to display the context menu.
3. Select **Cluster this SNP excluding selected samples** (Figure 34).

Shifting the Cluster Location

To shift the cluster location:

1. Press and hold the **Shift** key.
2. Click near the center of the cluster.

The  **move cursor** appears.

3. Drag the cluster to a new location.

Changing the Cluster Height/Width

To change the height or width of a cluster:

1. Press and hold the **Shift** key.
2. Click near the edge of an oval.

The  or  **resizing cursor** appears.

3. Drag the edge of the oval to reshape the cluster.

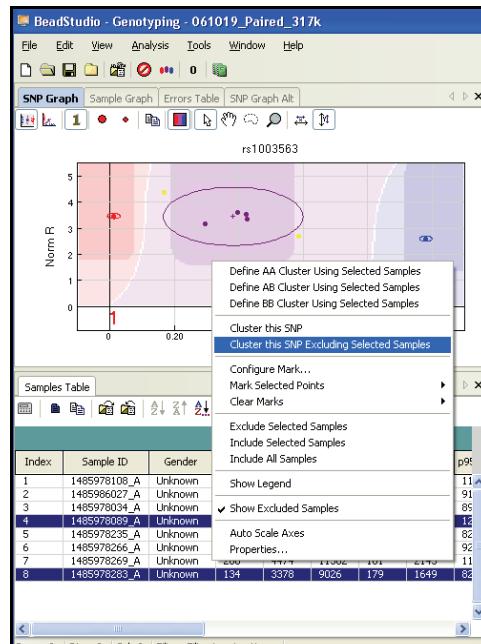


Figure 34 Editing Clusters

The clustering algorithm runs, excluding the samples you selected.

Exporting the Cluster File

You can export a cluster file any time after clustering.

To export the cluster file:

1. Select **File | Export Cluster Positions** (Figure 35).

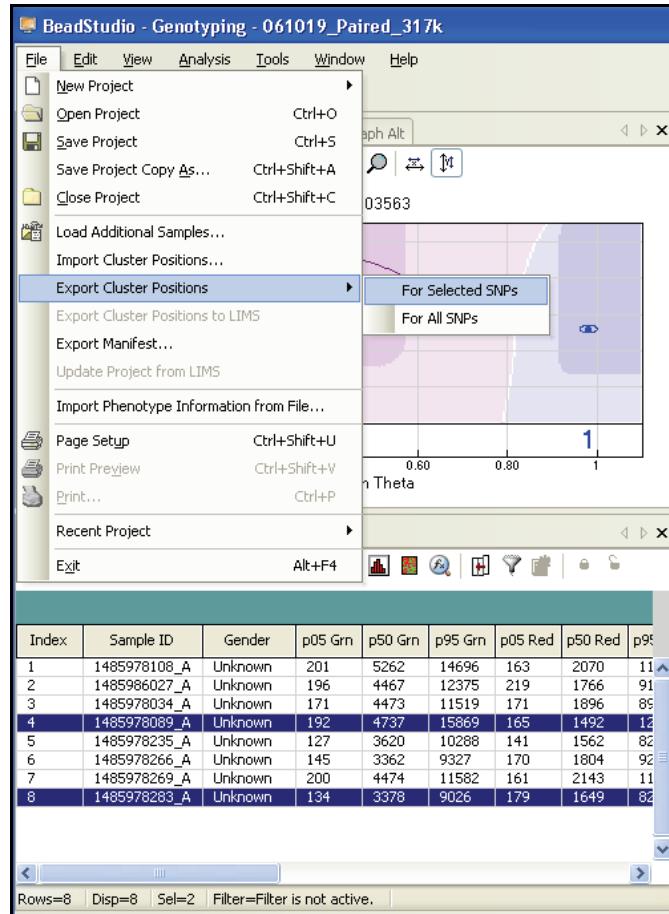


Figure 35 Export Cluster Positions Selected

2. Choose whether you want to export clusters **For Selected SNPs** or **For All SNPs**.

The **Save Cluster Positions** dialog box appears (Figure 36).

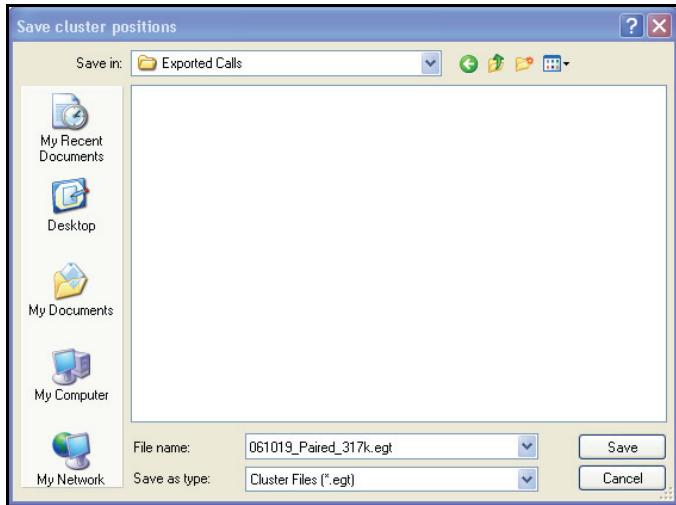


Figure 36 Save Cluster Positions

3. Browse to the location where you want to save your cluster position file.
4. Click **Save**.

The cluster file is assigned a default name based on the name of the project. However, you can choose to save your file with a different name.

Your exported cluster positions are saved as an *.egt cluster file, and are available to be imported into a different project.

Chapter 4

Viewing Your Data

Topics

- 38 Introduction
- 38 SNP Graph
- 39 Cartesian and Polar Coordinates
- 40 Normalization
- 41 Adjusting Axes
- 41 Selecting Samples
 - 42 Tips
 - 42 Marking Samples
 - 47 Excluding Samples
 - 48 Plotting Excluded Samples
 - 49 Customizing the SNP Table
 - 51 Shading Call Regions
 - 53 Viewing the Controls Dashboard
 - 54 Exporting Controls Data
 - 56 Viewing the Contamination Dashboard

Introduction

This chapter describes how to use graphs and tables to display, mark, and edit your data in the BeadStudio Genotyping Module.

SNP Graph

The **SNP Graph** (Figure 37) displays all samples for the currently-selected SNP in the **SNP Table** and in the **Full Data Table**.

Samples are colored according to their genotype. If you view a **SNP Graph** in polar coordinates, with normalization and call region shading turned on, the cluster ovals, call region shading, and number of samples in each cluster are also displayed (Figure 37).

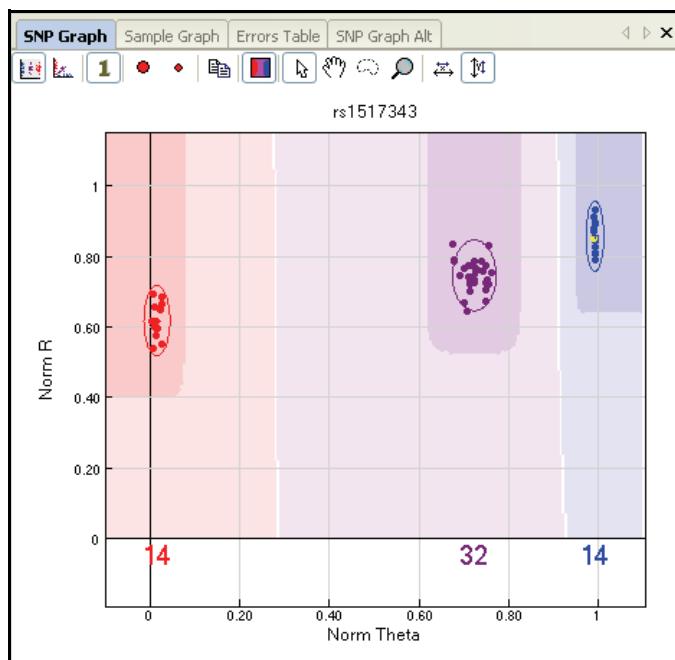


Figure 37 SNP Graph

In the **SNP Graph**, if there are any P-C or P-P-C errors in your data, the child appears as an "X" and the parent appears as an "O." Samples with reproducibility errors appear in the **SNP Graph** as squares (Figure 38).

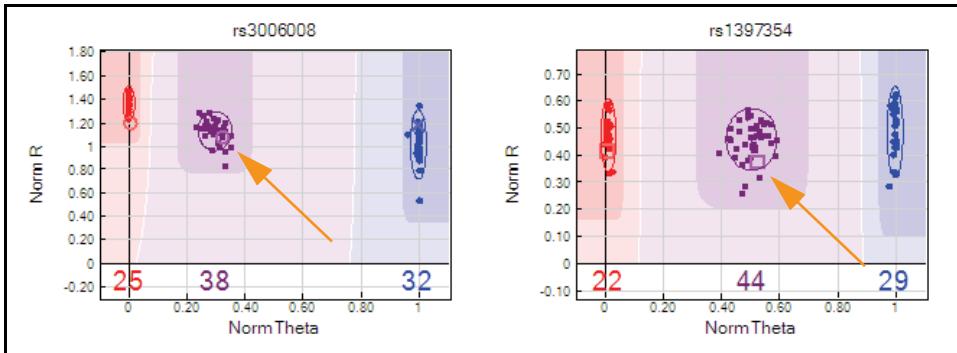


Figure 38 P-C Error (Left), Reproducibility Error (Right)

If you click an error entry in the **Errors** table, the associated samples are highlighted in yellow in the **SNP Graph** (Figure 39).

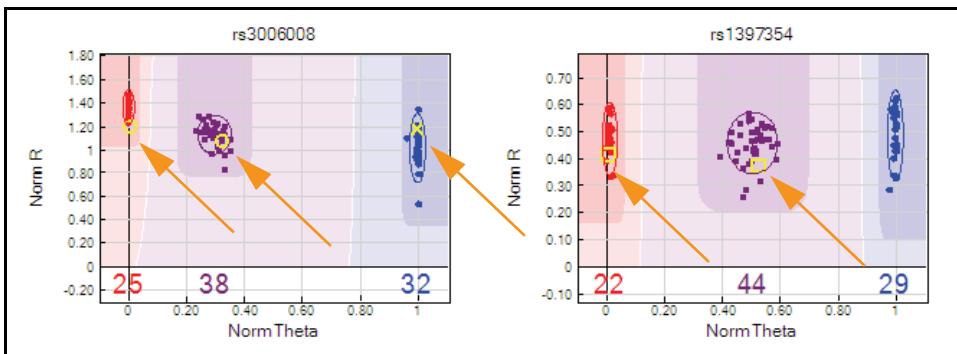


Figure 39 P-C Error and Reproducibility Error Highlighted in SNP Graph

Cartesian and Polar Coordinates

You can view the **SNP Graph** in either polar or Cartesian coordinates (Figure 40).

Cartesian coordinates use the X-axis to represent the intensity of the A allele and the Y-axis to represent the intensity of the B allele.

Polar coordinates use the X-axis to represent normalized theta (the angle deviation from pure A signal, where 0 represents pure A signal and 1.0 represents pure B signal), and the Y-axis to represent the distance of the point to the origin.

The Manhattan distance ($A+B$) is used rather than the Euclidian distance ($\sqrt{A^2+A^2B^2}$).

- ▶ Select  to display the plot in polar coordinates.
- ▶ Select  to display the plot in Cartesian coordinates.

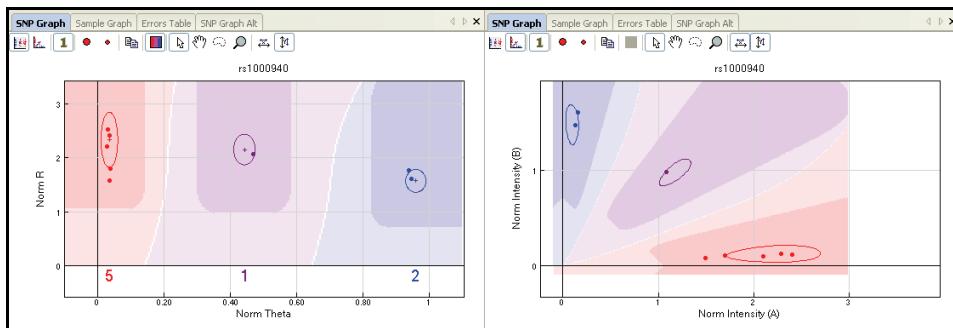


Figure 40 Polar Coordinates (Left) & Cartesian Coordinates (Right)

Normalization

You can view the SNP Graph in either normalized or raw format.

Click  **Normalization** to turn normalization on or off.

Figure 41 shows a sample graph, in polar coordinates, with normalization turned off (left), and with normalization turned on (right):

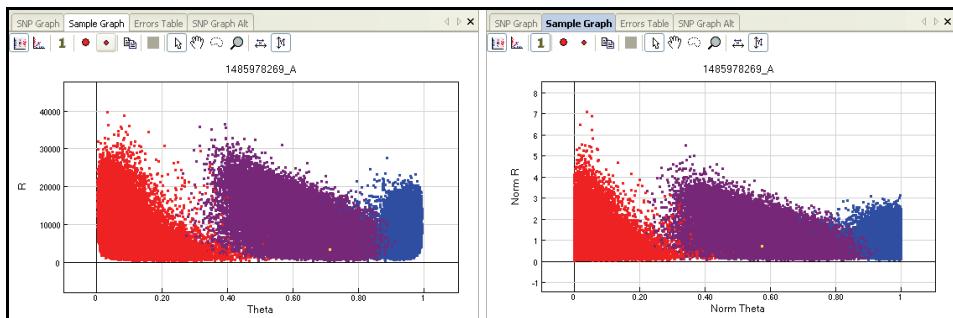


Figure 41 Normalization Turned Off (Left) & Normalization Turned On (Right)

Adjusting Axes

- ▶ To zoom in and out on the graphs:

Click  **Zoom Mode**.

In zoom mode you can:

- Click the left mouse button to zoom in.
- Click the right mouse button to zoom out.

Alternatively, using your mouse wheel you can:

- Roll up to zoom in.
- Roll down to zoom out.

- ▶ To change an axis:

Position your cursor over an axis and use the mouse wheel.

- ▶ To scroll along an axis:

Click, hold, and drag over an axis.

- ▶ To view different SNPs on the same scale:

Turn off  **Auto-Scale X-axis** or  **Auto-Scale Y-axis**.

Selecting Samples

You can select samples in the SNP Graph in a variety of ways:

- ▶ In  **Default Mode**, click-and-drag on the graph to draw a rectangle. When you release the button, all points in the rectangle are selected.
- ▶ In  **Lasso Mode**, click-and-drag on the graph to draw a region. When you release the button, all points in the shape you have drawn are selected.
- ▶ For the SNP Graph, selecting rows in the Samples Table selects the corresponding samples in the SNP Graph.
- ▶ To select additional samples without losing your original selection, press and hold the **Ctrl** button and click additional samples in the Samples Table.

The selected samples are shown in yellow by default (Figure 42).

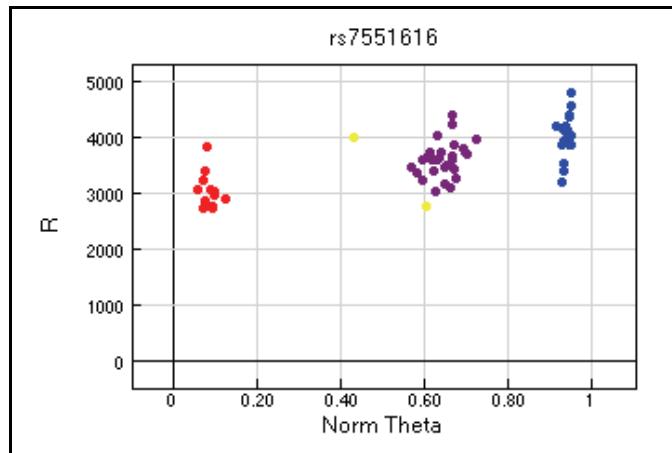


Figure 42 SNP Graph, Selected Samples Shown in Yellow

Tips

- ▶ To temporarily transfer to **Pan Mode**:
Position the cursor over an empty region of the genoplot (not over a cluster), then press and hold the **Shift** key.
- ▶ To temporarily transfer to **Lasso Mode**:
Press and hold the **Z** key.

Marking Samples

After you have selected samples, you may choose to mark them in a particular color. Mark colors are persistent, which means that the mark colors remain when you select a different SNP. Marks overwrite the default genotyping colors.

To mark selected samples:

1. Right-click on the graph and select **Configure Mark** from the context menu.

The **Configure Marks** dialog box appears (Figure 43).

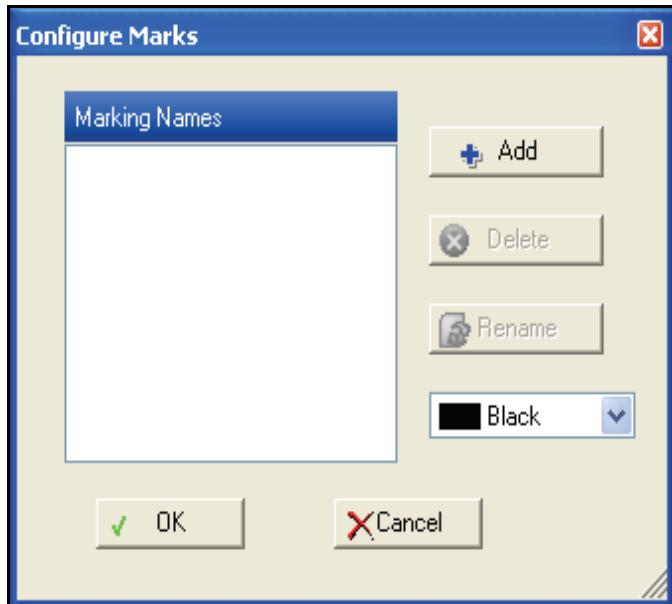


Figure 43 Configure Marks

2. Click **Add** to create a new mark.

The **Select Mark Name** dialog box appears (Figure 44).



Figure 44 Naming a Mark

3. Give your mark a color by selecting a color from the pulldown menu (Figure 45).

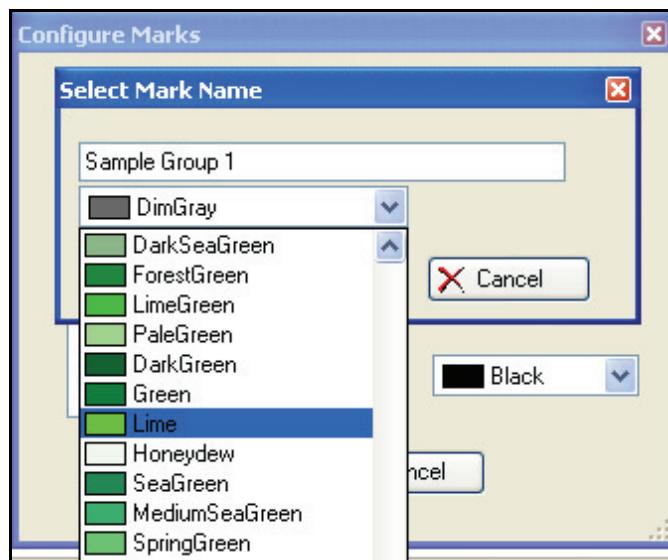


Figure 45 Selecting a Color for a Mark

4. Enter a name for your mark in the text field.
5. Click **OK**.

The selected samples appear in the **SNP Graph** and in the **Samples Table** in the color you chose (Figure 46).

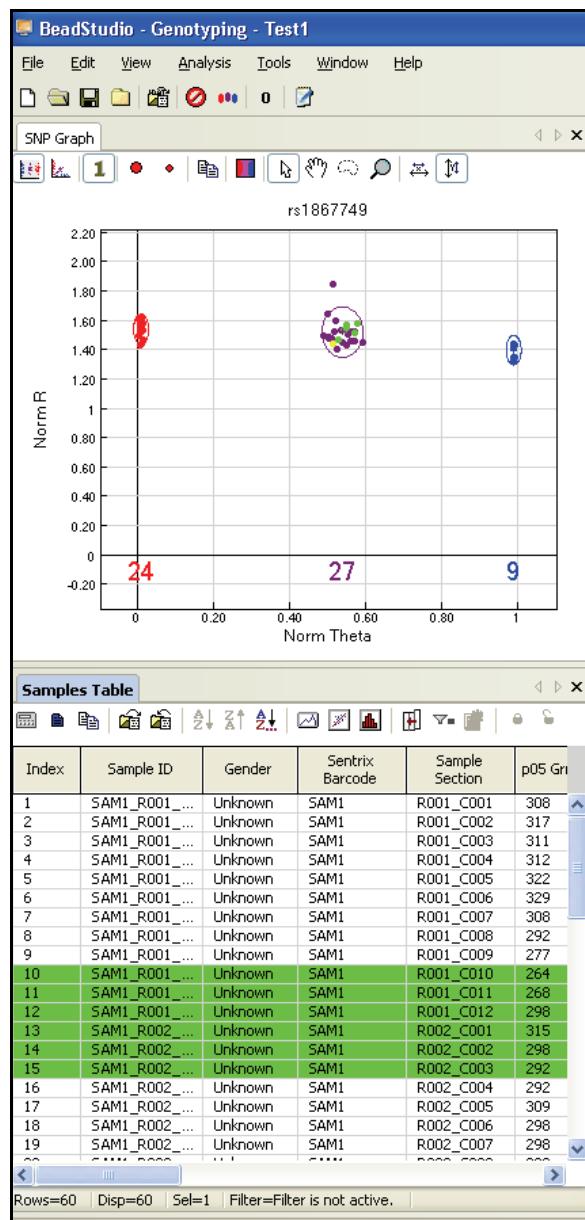


Figure 46 Displaying Marked Samples

Displaying the Legend

Perform the following steps to display the legend in the **SNP Graph** or **Sample Graph**.

1. Right-click in the graph.

The context menu appears.

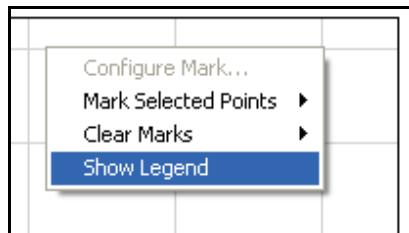


Figure 47 Displaying the Legend

2. Select **Show Legend**.

The legend appears, and includes the name of your mark.

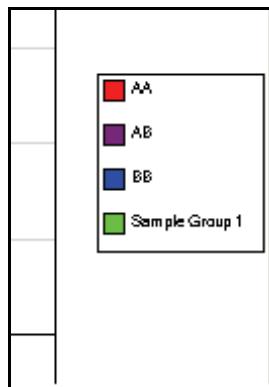


Figure 48 Legend Displaying Mark Name

Excluding Samples

To exclude one or more samples from your sample group:

1. Decide on one or more samples to exclude.
2. Right-click on the sample(s) in the **Samples Table**.
The context menu appears.
3. In the context menu (Figure 49), select **Exclude Selected Sample**.

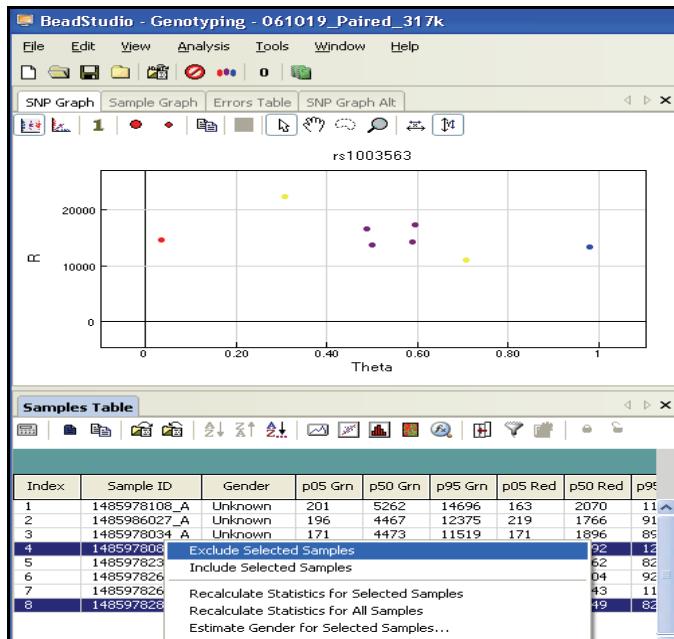


Figure 49 Excluding Selected Sample

The sample(s) you selected are excluded from your sample group.

Plotting Excluded Samples

If you have excluded one or more samples from your sample group, you may still want to plot them in the genoplot.

To plot excluded samples in the genoplot:

1. Select **Tools | Options | Project**.

The **Project Properties** dialog box appears (Figure 50).

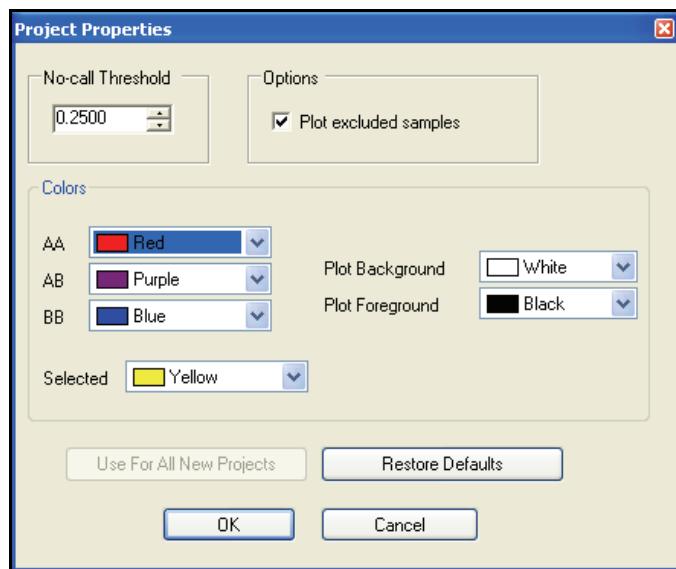


Figure 50 Project Properties

2. In the **Options** area, select the **Plot excluded samples** checkbox.
3. Click **OK**.

The excluded samples are plotted in the genoplot.

Alternatively, you can choose to plot excluded samples in the genoplot by right-clicking in the genoplot and choosing **Include All Samples** from the context menu.

To remove excluded samples from the genoplot:

1. Go to **Tools | Options | Project**.

The **Project Properties** dialog box appears (Figure 50).

2. In the **Options** area, clear the **Plot excluded samples** checkbox.

3. Click **OK**.

The excluded samples are removed from the genoplot.

Alternatively, you can choose to remove excluded samples from the genoplot by right-clicking in the genoplot and choosing **Exclude Selected Samples** from the context menu.

Customizing the SNP Table

Using the **Column Chooser**, you can select the columns you want to display in the **SNP Table** and arrange the columns in any order you want to display them.

1. In the **SNP Table**, click  **Column Chooser**.

The **Column Chooser** appears (Figure 48).

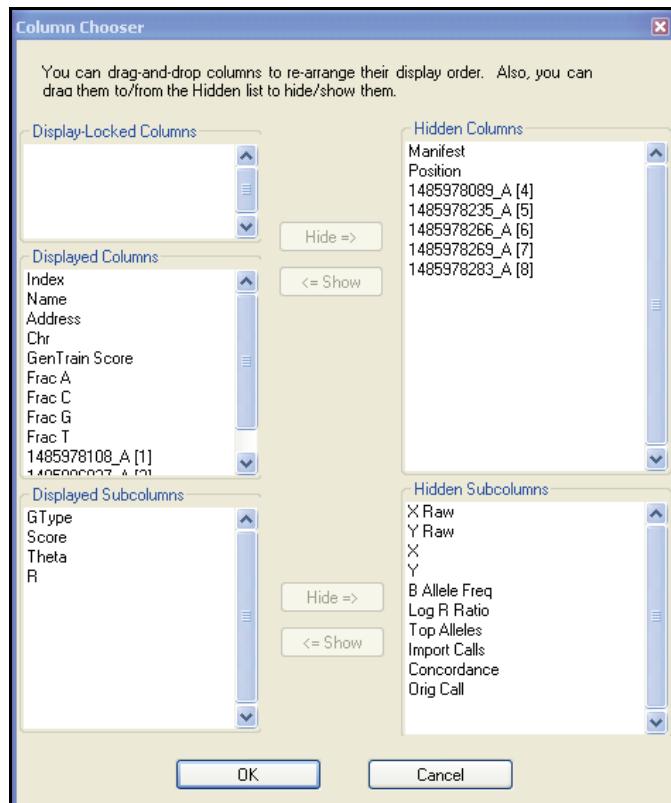


Figure 51 Column Chooser

2. In the **Column Chooser** dialog box, click to select a column that you want to display.
3. Click **Show**.
The column you selected is moved to the **Displayed Columns** list or the **Displayed Subcolumns** list.
Alternately, you can select and drag a column to the **Displayed Columns** list.
4. To change a column's position in the table, click to select a column, then drag the column header up or down in the displayed column list.
5. Click **OK** to display columns in their new positions.
Alternatively, click **Cancel** to retain columns in their current positions.

Shading Call Regions

Toggle  **Shade Call Regions** in the graph window toolbar to apply color to the genoplot calling regions in the graph window. These shaded regions correspond to the no-call threshold.

To set the lower threshold for valid calls within BeadStudio, perform the following steps:

1. Select **Tools | Options | Project**.
2. In the **No-Call Threshold** area, select a lower limit for valid calls within BeadStudio. The default is 0.25.

By default, samples lying within the **dark red** region are called AA; samples lying within the **dark purple** region are called AB; and samples lying within the **dark blue** region are called BB (Figure 52).

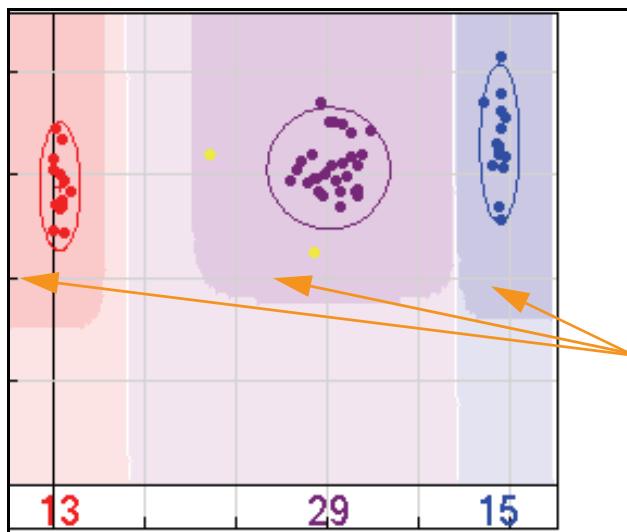


Figure 52 Shaded Call Regions



NOTE

Shading of clusters is toggled off by default, and is available for the polar graph only.

To change the colors for cluster calls:

1. Go to **Tools | Options | Projects**.
2. In the **Colors** area, use the dropdown menus to change the default colors for the AA, AB, and BB genotypes as well as for selected samples, plot foreground, and plot background.
3. Click **OK**.

The clusters display with the assigned colors.

To restore default colors to clusters and plot properties:

1. Go to **Tools | Options | Projects**.
2. Click **Restore Defaults**.
3. Click **OK**.

The default cluster and plot colors are restored.

Viewing the Controls Dashboard

To view a graphic report displaying system controls information:

- Select **Analysis | View Controls Dashboard** (Figure 53).
The **Controls** window appears.

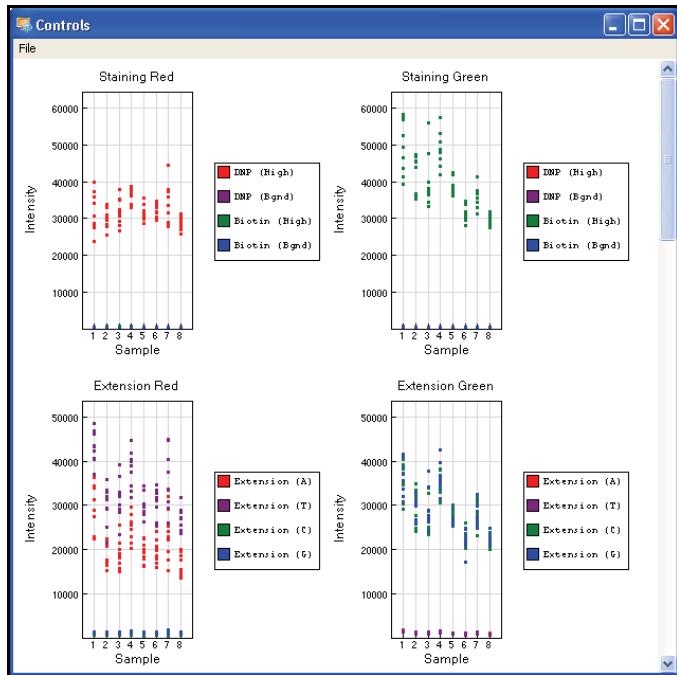


Figure 53 Controls Dashboard



NOTE

Excluded samples are not displayed in the Controls dashboard.

Exporting Controls Data

You may want to view a controls data file if you are interested in the numerical details of the data shown in the controls dashboard.

To export controls data, perform the following steps:

1. In the controls dashboard, select **File | Export Data** (Figure 54).

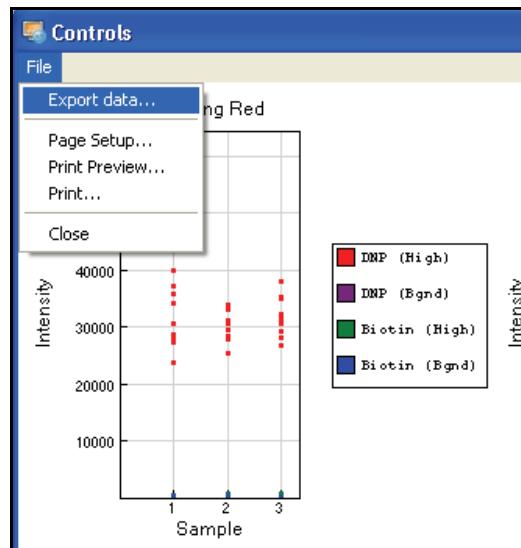


Figure 54 Exporting Controls Data

The **Save As** dialog box appears (Figure 55).

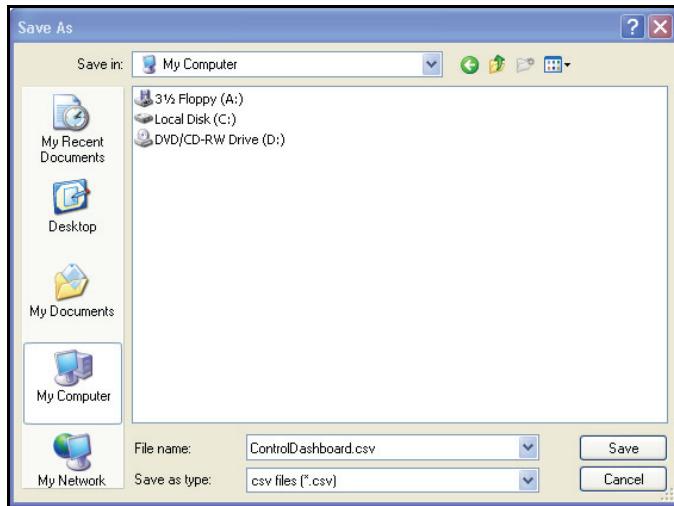


Figure 55 Saving the Controls Dashboard

2. Browse to the location where you want to save your file.
3. Type a name for your file in the **File Name** text field.
4. Click **Save**.

The exported controls dashboard file is saved as a *.csv file in the location you specified.

Viewing the Contamination Dashboard

To view a graphic report displaying contamination information:

- ▶ Select **Analysis | View Contamination Dashboard**.

The **Contamination Controls** window appears (Figure 56).

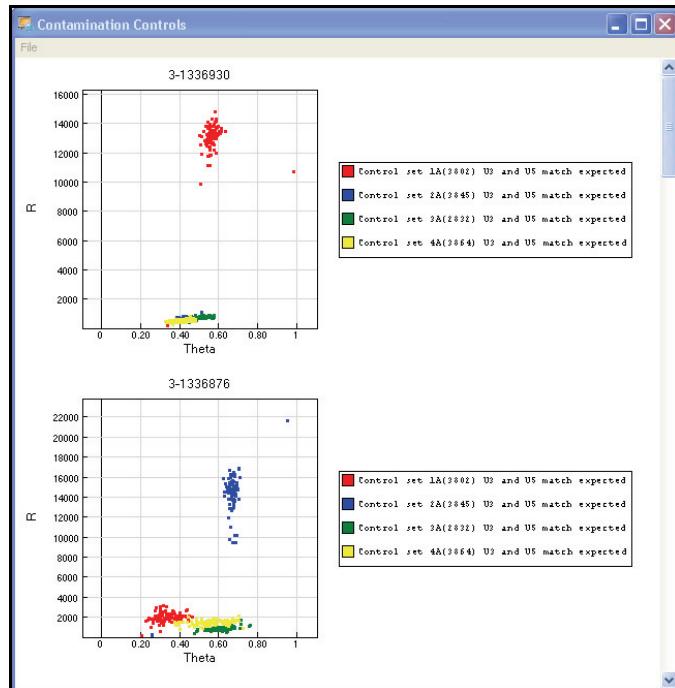


Figure 56 Contamination Dashboard



NOTE

The Contamination Dashboard applies only to GoldenGate data. There is no Contamination Dashboard for Infinium data.

Chapter 5

Analyzing Your Data

Topics

- 58 Introduction
- 58 Importing Phenotype Information
- 60 Estimating the Gender of Selected Samples
- 62 Editing the Properties of Selected Samples
- 64 Analyzing Paired Samples
- 65 Using Concordance Features
 - 65 Exporting Allele Calls
 - 66 Importing Allele Calls
 - 66 Concordance Calculations
- 66 Using Column Plug-Ins

Introduction

Use the procedures in the following sections to analyze your data.

Importing Phenotype Information

A phenotype information file is a *.csv file you can create and import into a project if you want include sample-related phenotype information.

A phenotype information file must contain an **Index** column that corresponds to the **Index** column in the **Samples Table**.

You can also optionally include the following columns in a phenotype information file:

- ▶ Gender
- ▶ Ethnicity
- ▶ Age
- ▶ Weight
- ▶ Blood Pressure Systolic
- ▶ Blood Pressure Diastolic
- ▶ Blood Type
- ▶ Phenotype Pos 1
- ▶ Phenotype Pos 2
- ▶ Phenotype Pos 3
- ▶ Phenotype Neg 1
- ▶ Phenotype Neg 2
- ▶ Phenotype Neg 3



NOTE

The columns listed above are the only columns you can import into a BeadStudio genotyping project using a phenotype information file. Additional columns present in a phenotype information file will not be imported into the BeadStudio project.

To import phenotype information from a file:

1. Select **File | Import Phenotype Information From File.**

The **Import Phenotype File** window appears (Figure 57).

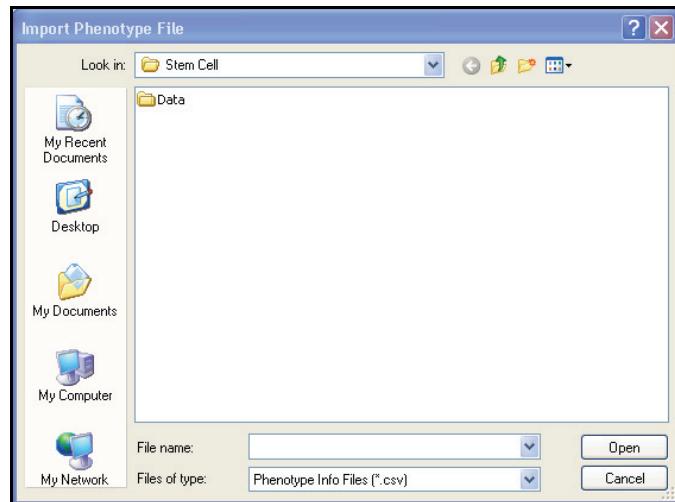


Figure 57 Importing Phenotype Information

2. Browse to a *.csv phenotype information file from which you want to import information (Figure 58).

A	B	C	D	E	F	G	H	I	J	K	L	M
Index	Gender	Ethnicity	Age	Weight	Height	Blood Pressure Systolic	Blood Pressure Diastolic	Blood Type	Phenotype Pos 1	Phenotype Pos 2	Phenotype Pos 3	Phenotype Neg 1
2	1 M	Caucasian	30	300	120	100		70 AB	Pos1	Pos2	Pos3	Neg1
3	2 M	Caucasian	30	300	120	100		70 AB	Pos1	Pos2	Pos3	Neg1
4	3 M	Caucasian	30	300	120	100		70 AB	Pos1	Pos2	Pos3	Neg1
5	4 M	Caucasian	30	300	120	100		70 AB	Pos1	Pos2	Pos3	Neg1
6	5 M	Caucasian	30	300	120	100		70 AB	Pos1	Pos2	Pos3	Neg1
7												
8												
9												
10												
11												
12												
13												

Figure 58 Phenotype Information File

3. Select **Open.**

Information from the phenotype information file you selected is imported into BeadStudio and displayed in the **Samples Table**.

Estimating the Gender of Selected Samples

To estimate gender for selected samples:

1. In the **Samples** table, select the samples for which you want BeadStudio to estimate gender.

The selected samples are highlighted in dark blue. Note that the **Gender** column of each sample contains “Unknown” (Figure 59).

Index	Sample ID	Gender	p05 Grn	p50 Grn	p95 Grn	p05 Red	p50 Red	p95 Red
1	1485978108_A	Unknown	201	5262	14696	163	2070	11
2	148596027_A	Unknown	196	4467	12375	219	1766	91
3	1485978034_A	Unknown	171	4473	11519	171	1896	89
4	1485978089_A	Unknown	192	4737	15869	165	1492	12
5	1485978235_A	Unknown	127	3620	10288	141	1562	82
6	1485978266_A	Unknown	145	3362	9327	170	1804	92
7	1485978269_A	Unknown	200	4474	11582	161	2143	11
8	1485978283_A	Unknown	134	3378	9026	179	1649	82

Figure 59 Selected Samples

1. Right-click anywhere on the selected samples.

The context menu appears (Figure 60).

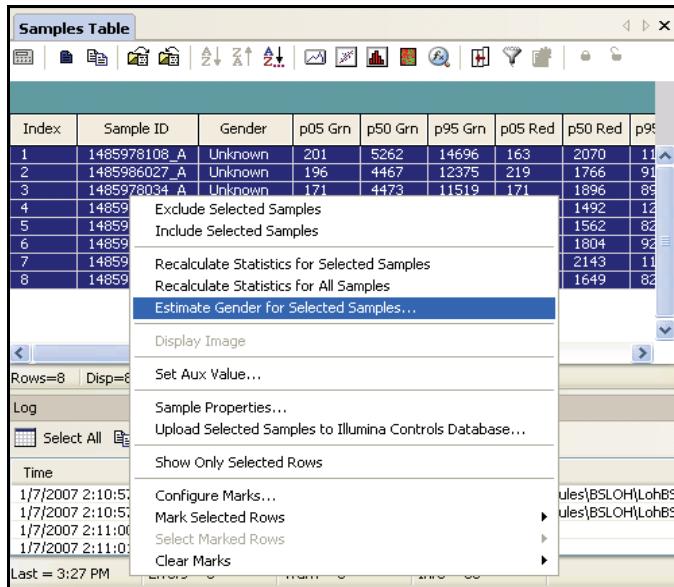


Figure 60 Samples Table Context Menu

2. Select Estimate Gender for Selected Samples.

The **Would you like to populate the Gender column...** dialog box appears (Figure 61).

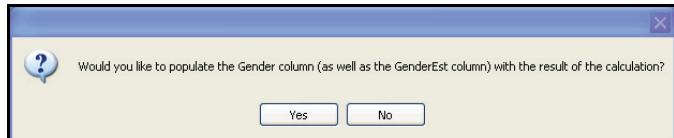


Figure 61 Populating the Gender Column

3. Choose one of the following:

Yes—the **Gender** and **Gender Est** columns of the **Samples Table** are populated with the estimated gender for the samples you selected.

No—only the **Gender Est** column of the samples table is populated with the estimated gender for the samples you selected.

Editing the Properties of Selected Samples

To edit the properties of selected samples:

1. In the **Samples** table, select one or more samples to edit.
The selected samples are highlighted in dark blue (Figure 62).

The screenshot shows a software window titled "Samples Table". The main area is a grid table with columns labeled: Index, Sample ID, Gender, p05 Grn, p50 Grn, p95 Grn, p05 Red, p50 Red, and p95 Red. There are 8 rows of data. Rows 1, 2, 3, and 4 are highlighted in dark blue, indicating they are selected. The data is as follows:

Index	Sample ID	Gender	p05 Grn	p50 Grn	p95 Grn	p05 Red	p50 Red	p95 Red
1	1485978108_A	Unknown	201	5262	14696	163	2070	11
2	1485986027_A	Unknown	196	4467	12375	219	1766	91
3	1485978034_A	Unknown	171	4473	11519	171	1896	89
4	1485978089_A	Unknown	192	4737	15869	165	1492	12
5	1485978235_A	Unknown	127	3620	10288	141	1562	82
6	1485978266_A	Unknown	145	3362	9327	170	1804	92
7	1485978269_A	Unknown	200	4474	11582	161	2143	11
8	1485978283_A	Unknown	134	3378	9026	179	1649	82

Figure 62 Selected Samples

2. Right-click anywhere on the selected samples.
3. The context menu appears (Figure 63).

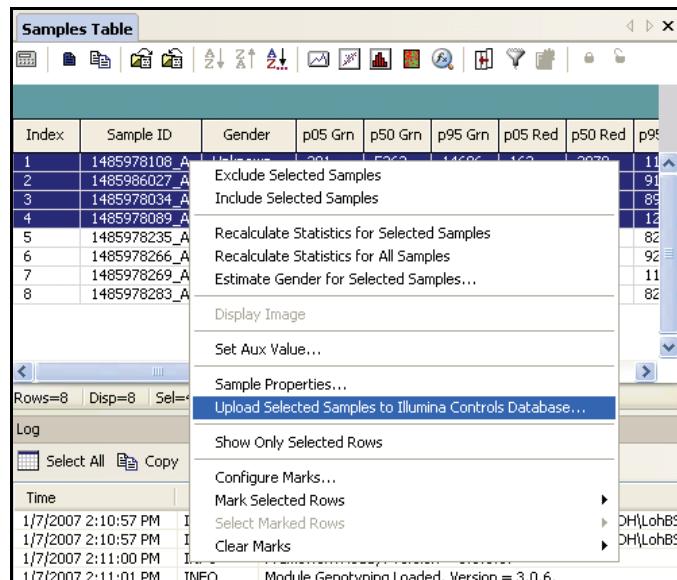


Figure 63 Samples Table Context Menu

4. Select Sample Properties.

The **Sample Properties** window appears (Figure 64).

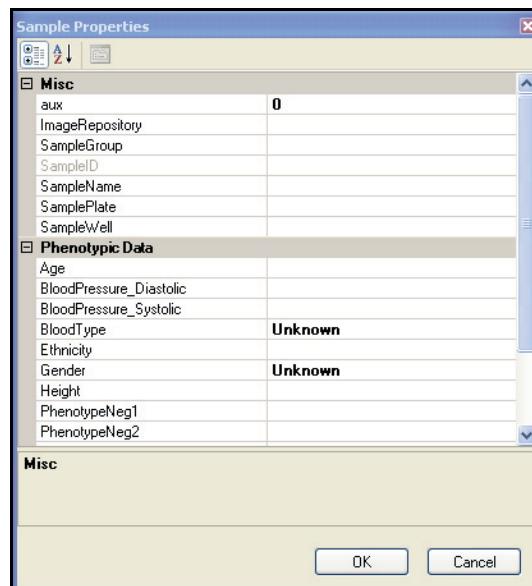


Figure 64 Sample Properties

5. Click in the right-hand column of any properties you want to edit and type new values.

6. Click **OK**.

The updated column properties are displayed in the **Samples** table.

**NOTE**

To change the path to images displayed in the Image Viewer, edit the Image Repository property.

Analyzing Paired Samples

Paired sample data can be useful for analyzing chromosomal aberrations. BeadStudio includes a **Paired Samples Table** with columns that show the differences in various statistical measures between a pair of samples (a subject sample and a reference sample).

Paired samples can be created in two ways:

- ▶ by designating subject-and-reference pairs in the sample sheet used to create a project
- ▶ by designating subject-and-reference samples using the paired samples editor

Once you designate paired samples, the pairs appear in the **Paired Samples Table**.

When paired sample data are loaded in the **Paired Sample Table**, certain features are enabled. These include the following:

- ▶ **Analysis | Calculate Paired Sample LOH/CN Scores**
- ▶ In the **SNP Graph**, graphical elements indicate which samples are paired. Figure 65 shows an aqua line designating a paired sample subject and reference.

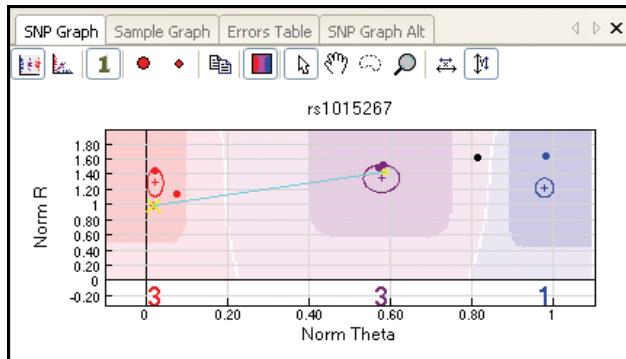


Figure 65 SNP Graph Showing Paired Samples

- ▶ In the IGV, paired sample data becomes available for plotting and autobookmarking.

Using Concordance Features

Use the concordance features described in the following sections to compare data from different different projects.

Exporting Allele Calls

If you want to compare the allele calls in your current project to allele calls in another project, you can export the allele calls from your current project and import them into other projects.



To export allele calls and import them into another project, the sample names in each project must be the same. Allele calls for sample names that do not match will not be compared.

To export allele calls from your current project:

1. Select **Analysis | Export Allele Calls**.
The **Export Allele Calls** dialog box appears.
 2. Browse to the directory where you want to save the allele calls from your current project.
 3. Click **OK**.
- The allele calls are saved to the directory you designated.

Importing Allele Calls

If you have previously exported and saved allele calls from a project, you can import these saved allele calls into a different project to calculate concordance.

To import allele calls into a project:

1. Select **Analysis | Import Allele Calls**.

The **Import Allele Calls** dialog box appears.

2. Browse to the location where you previously saved allele calls that you exported from a different project.

The files available to import are listed in the **Files Found** section of the **Import Directory** area.

3. Click **OK**.

The allele calls are imported. They populate the **Import Calls** column in the **Full Data Table**, and concordance is calculated.

Concordance Calculations

Concordance calculations appear in two locations:

- In the **Full Data Table**, in the **Concordance** subcolumn.
- In the **Samples Table**, in the **Concordance** column.



NOTE

Columns showing concordance are not visible by default. To display these columns, use the **Column Chooser**.

Using Column Plug-Ins

BeadStudio v3 gives you the option to install column plug-ins as part of the BeadStudio install process, or to create custom column plug-in algorithms. These plug-ins are used to create custom subcolumns in the **Full Data Table**. This open plug-in architecture allows you to add to the standard features available in BeadStudio.

Before you can create a new subcolumn, you must first make column plug-ins available to BeadStudio.

To make column plug-ins available to BeadStudio, do one of the following:

- If the column plug-in has an install program:

Run the install program.

The column plug-in is installed in the correct directory and is now available to BeadStudio.

- If the column plug-in does not have an install program:

Copy the dll file for the column plug-in to the following directory:

C:\Program Files\Illumina\BeadStudio 2.0\Plugins

The column plug-in is now available to BeadStudio.

To create a subcolumn based on a column plug-in:

1. Select Analysis | Create Plug-In Column.

The **Select Column Plug-In Form** dialog box appears.

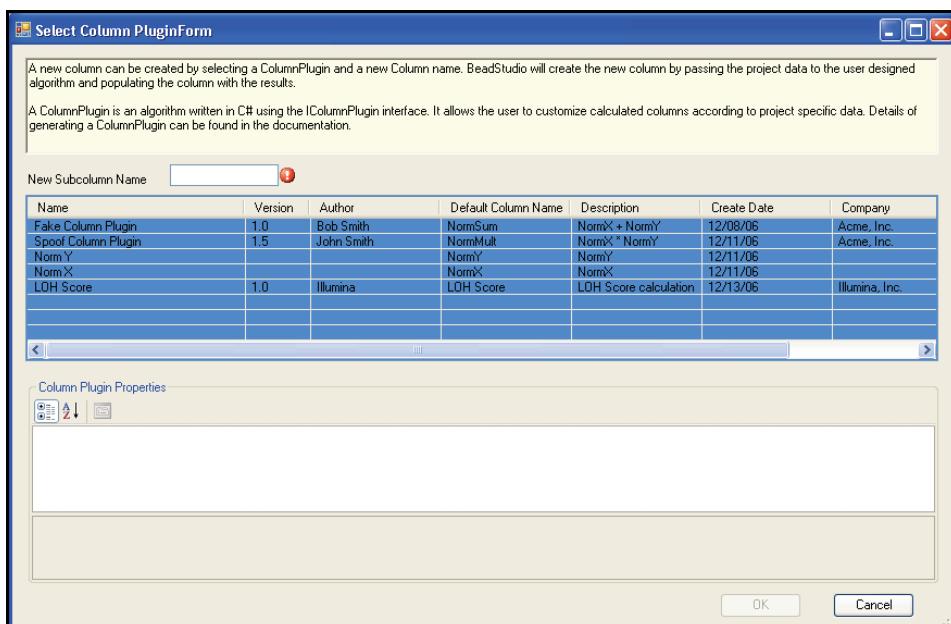


Figure 66 Select Column Plug-In Form

2. In the column plug-ins table, select a row from the list of available column plug-ins.
3. **[Optional]** Type a new name for the subcolumn in the New Subcolumn Name text field.
4. **[Optional]** To edit any pre-defined properties, click in the right-hand column of the **Column Plug-In Properties** table and enter new values.
5. Click **OK**.

The new subcolumn is created and appears in the **Full Data Table**.

Chapter 6

Using the Illumina Genome Viewer

Topics

- 70 Introduction
- 70 Viewing Data in the Illumina Genome Viewer
 - 71 Getting Data Files
 - 73 Plotting Data
- 78 Viewing Data in the Illumina Chromosome Browser
 - 79 Navigating the Illumina Chromosome Browser
 - 80 Viewing Gene Information
 - 82 Plotting Sample Columns
 - 82 Viewing Project Manifest SNPs

Introduction

This chapter describes how to use the Illumina Genome Viewer (IGV) and its components (the Illumina Chromosome Browser [ICB], the Chromosome Heat Map, and data tracks) to view and analyze your genotyping data. You can access the ICB, Chromosome Heat Map, and data tracks directly through the Illumina Genome Viewer.

Viewing Data in the Illumina Genome Viewer

The IGV allows you to visualize your data on a genome-wide scale.

Launch the IGV by selecting **Analysis | Show Genome Viewer**.

Figure 67 below shows the IGV main window. In **Chromosome Slide Show** mode, the IGV can display up to four plots at a time over one to four consecutive chromosomes. In **Whole Genome** mode, it can display up to two plots at a time for all chromosomes.

In addition to using the IGV, you can browse your data at the chromosomal level using the ICB and/or the **Zoom Plot**.

Launch the ICB by double-clicking on a chromosome in the Genome Viewer (Figure 67, #1). For more information about the ICB, see *Viewing Data in the Illumina Chromosome Browser* on page 78.

Launch the Zoom Plot by double-clicking on any plot in the IGV (Figure 67, #2). For more information about the Zoom Plot, see *Zooming Into a Region* on page 78.

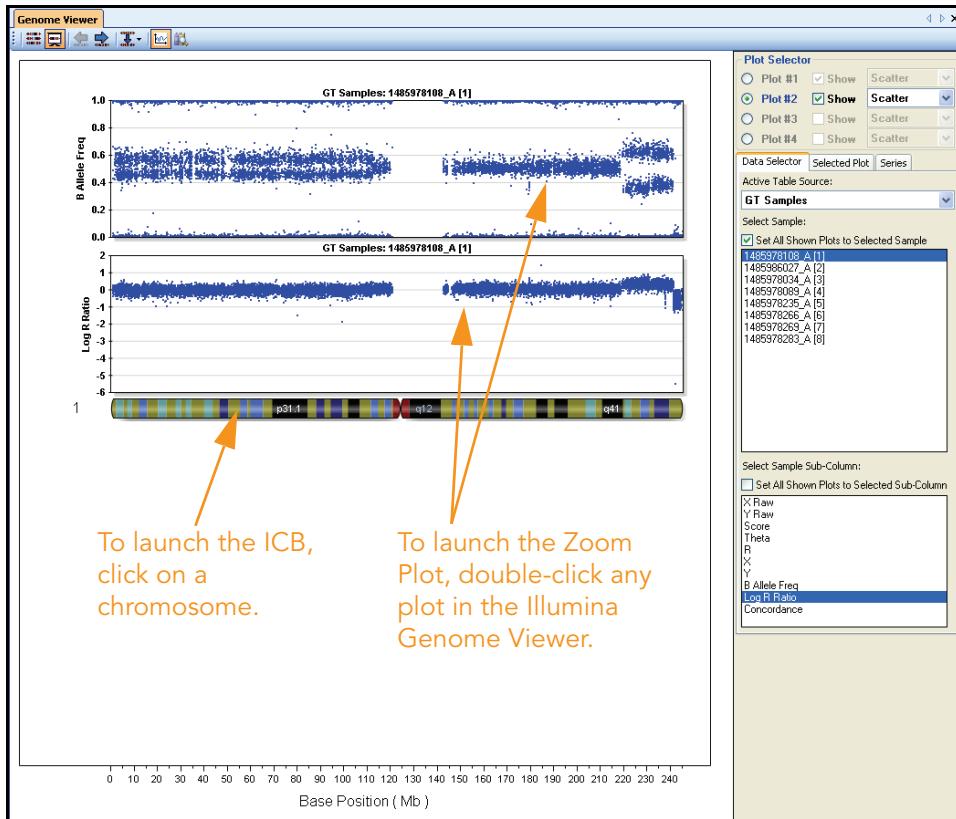


Figure 67 Illumina Genome Viewer

Getting Data Files

The IGV is installed with various genomes from the UCSC GoldenPath database.

If you want genomes or builds other than those installed with the IGV, download additional genome annotation files using the following procedure:

1. Go to: <http://hgdownload.cse.ucsc.edu/downloads.html>.
2. Click the species of the genome you want to download.
3. In your BeadStudio installed directory structure, navigate to the folder that contains your Genome Viewer genome files.

The default location is:

C:\Program Files\BeadStudio 2.0\Modules\GenomeViewer\Genomes

4. Create a new folder under the path in Step 3. The folder name should be descriptive of the genome and build; for example, "Human 2004 (hg17)."

Table 1 Illumina Genome Viewer Toolbar Buttons & Functions

Button	Description
	Edit User Preferences —displays the Preferences dialog box.
	Whole Genome View Mode —shows all genome chromosomes with all enabled plots shown above each chromosome. Whole Genome View Mode is the default IGV mode.
	Chromosome Slide Show Mode —enables the slide show mode which allows the user to view from one to four chromosomes at the same time.
	View Previous Slide —displays the previous chromosome slide. This button is only enabled if the Chromosome Slide Show Mode toolbar toggle button is active.
	View Next Slide —displays the next chromosome slide. This button is only enabled if the Chromosome Slide Show Mode toolbar toggle button is active.
	Jump to a Specific Chromosome —displays a menu that allows you to jump to a particular chromosome slide.
	Show Plot Selector Panel —toggles between showing and hiding the Plot Selector Panel. This button is only enabled if the Chromosome Slide Show Mode toggle button is active.
	Toggle Bookmark —toggles between displaying bookmarks and hiding bookmarks.

Plotting Data

When you launch the Illumina Genome Viewer from the **Analysis** menu, it has access to the sample columns in the **Full Data Table**. If any paired samples have been created, it also has access to the sample columns in the **Paired Sample Table**. The samples and per-sample subcolumns appear in the **Plot Selector** panel to the right of the IGV main window. Figure 68 shows the **Plot Selector** panel.

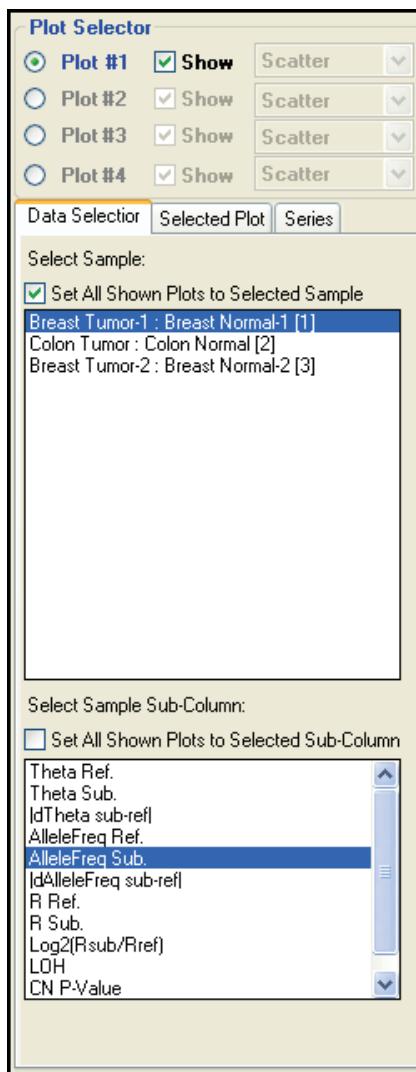


Figure 68 Plot Selector

Choosing Options in the Plot Selector Panel

The radio buttons labeled **Plot 1** to **Plot 4** allow you to select which plot to edit. The checkboxes labeled **Show** toggle the particular plot's visibility (leaving this checkbox unchecked hides the plot).

In Figure 68, Plot 2 is selected and the data series that is plotted in plot #2 will be the allele frequency for the reference sample #1 as shown by the current selected items in the **Select Sample Sub-Column** and **Select Sample** list boxes, respectively.

To select the sample as the data source for all shown plots, select the **Set All Shown Plots to Selected Sample** checkbox in the **Select Sample** listbox. Use this feature to look at different subcolumns for the same sample and to quickly view the same subcolumns for different samples by changing the selected sample in the **Select Sample** list box.

Choosing Options in the Data Selector Tab

In the **Data Selector** tab, you can choose which samples to plot and which sample subcolumns to display.

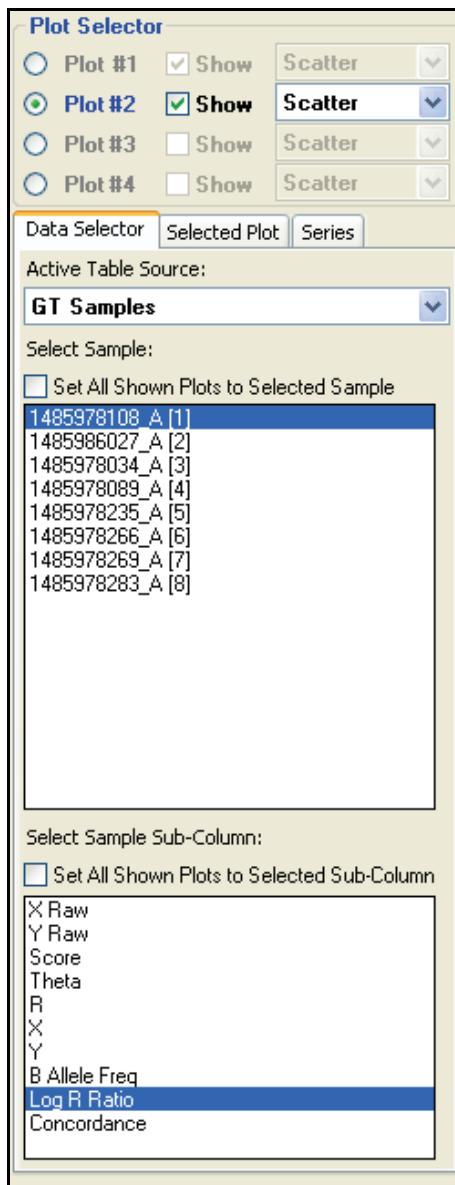


Figure 69 Data Selector

Choosing Options in the Selected Plot Tab

In **Selected Plot** tab, you can change the appearance of each plot. Select options for the following variables:

- ▶ **Title**—Plot title
- ▶ **Color**—Color of plot title
- ▶ **Height**—height of plot (small, normal, or large)
- ▶ **Label**—Y axis label
- ▶ **Min**—Minimum Y axis value
- ▶ **Max**—Maximum Y axis value
- ▶ **Step**—Y axis increment

In addition to changing the size of each plot, you can change the scale of the Y axis for each plot. If the **Auto-Scale Y Axis** checkbox is selected, each plot will be automatically scaled.

Click **Update Changes** to apply your changes to the plots.

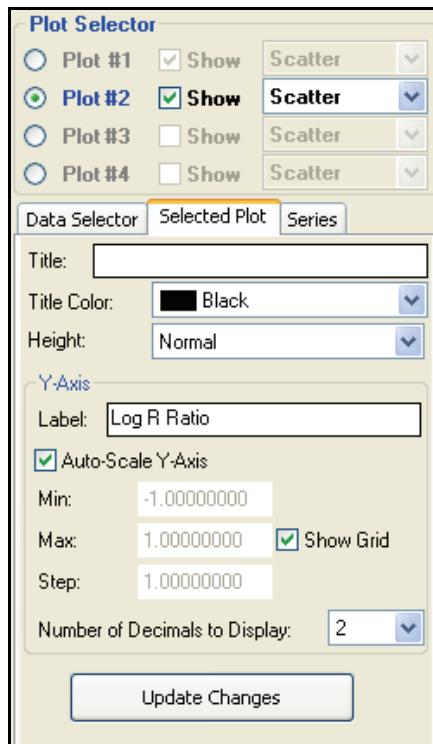


Figure 70 Selected Plot

Choosing Options in the Series Tab

In the **Series** tab, you can adjust the visual characteristics of each plot. Select options for the following variables:

- ▶ **Color**—color of the plot
- ▶ **Line Style**—solid, dash, or dot
- ▶ **Line Width**—width of the line in screen pixels

You can also add a smoothing series (a moving average or median) to each plot. This smoothing series can be drawn with a smoothing window as small as 5kb. The color, line style, and line width of the smoothing series are adjustable.

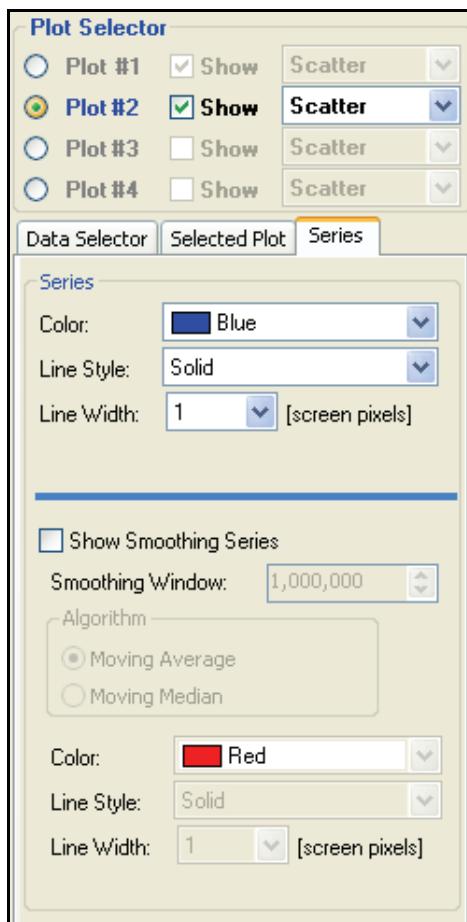


Figure 71 Series

Zooming Into a Region

The **Zoom Plot** allows you to zoom into a particular chromosomal region to view data in finer detail relative to known chromosomal annotation. Launch the Zoom Plot by double-clicking any plot in the IGV. The **Illumina Chromosome Browser** opens.

Viewing Data in the Illumina Chromosome Browser

The ICB allows you to explore data by chromosome and/or by gene. The following sections describe how to work with the ICB.

Launching the Illumina Chromosome Browser

Launch the ICB by double-clicking on a chromosome in the **Illumina Genome Viewer**.

Navigating the Illumina Chromosome Browser

Figure 72 shows the ICB main window.



Figure 72 Illumina Chromosome Browser

Viewing Gene Information

When selected, genes display in red (Figure 74). Information about the selected gene is shown in the **Gene ID** and **Gene Details** tabs at the bottom of the form. To zoom the current view to fit the whole gene, double-click the gene.

The **Gene ID** tab shows all available gene cross-reference names from various databases. A brief description and RefSeq Summary (if available) is also shown.

To open your default internet browser to the National Center for Biotechnology Information (NCBI) definition of that particular RefSeq gene:

Click the RefSeq ID label (shown in blue in the **Gene ID** tab of Figure 74).

The **Gene Details** tab shows all gene exons, gene strand types, transcription start/end positions, and coding region start/end positions for the currently selected gene.

To cause the view to fit a selected exon, double-click any exon shown in the exon listbox (Figure 73).

Exon Listbox				
Strand:	Positive	Exon	Start Position	End Position
Transcription Start:	40,089,902	1	40,089,902	40,090,150
Transcription End:	40,104,720	2	40,100,251	40,100,360
Coding Region Start:	40,090,057	3	40,093,155	40,093,159
Coding Region End:	40,104,333	4	40,093,975	40,100,099
Number of Exons:	14	5	40,100,235	40,100,314
		6	40,100,521	40,100,779
		7	40,101,385	40,101,456
		8	40,101,575	40,101,697
		9	40,101,850	40,101,934
		10	40,102,392	40,102,476

Figure 73 Exon Listbox

Figure 74 shows exon information for the currently-selected exon. If sequence files are available in the appropriate genome folders (and if the sequence will fit into the current view) the positive strand genomic sequence is displayed.

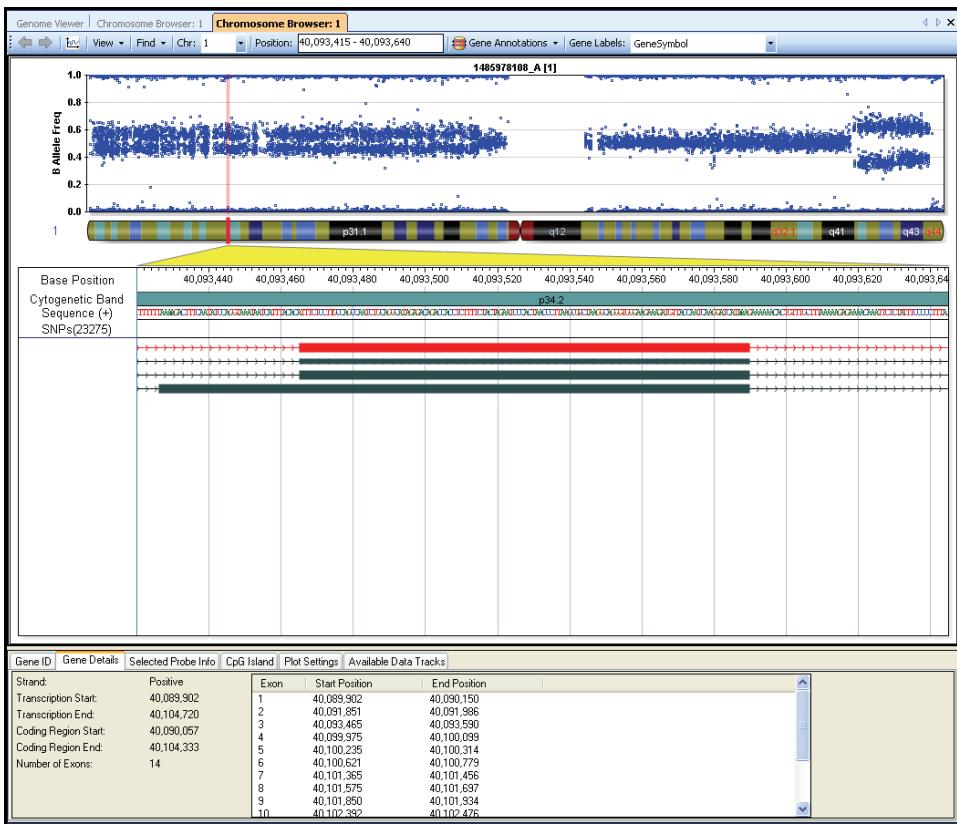


Figure 74 Viewing Gene Exons in Detail

Plotting Sample Columns

To plot sample columns from the **Full Data Table** or **Paired Samples Table** in the ICB:

Click  **Toggle Plot Display**.

The plot toggles to display mode and the **Plot Settings** tab at the bottom of the dialog box becomes active (Figure 5-11).

To hide the plot:

Click  **Toggle Plot Display** to toggle off.

You can select the sample and per-sample subcolumn to be plotted. The plot automatically updates when an item is selected in either the **Select Sample** or **Select Sample Sub-Column** listbox. You can set the plot type, y-axis properties, and series properties in the **Plot Settings** tab.

Viewing Project Manifest SNPs

To display SNPs in the current project manifest:

► Select **File | Load Manifest | SNPs**.

The SNPs in the current manifest are displayed (Figure 75).



Figure 75 Displaying Project Manifest SNPs in the ICB

Chapter 7

Performing LOH and Copy Number Analysis

Topics

- 84 Introduction
- 84 B Allele Frequency
- 86 Log R Ratio
- 88 Algorithm Plug-ins
- 89 Using Autobookmarking Algorithms
- 93 Using Column Plug-Ins

Introduction

BeadStudio provides visualization tools and detection algorithms to analyze both single and paired samples for loss of heterozygosity (LOH) and copy number (CN) changes.

In the BeadStudio Genotyping Module, the primary tool for displaying the results of LOH or CN analysis is the Illumina Genome Viewer (IGV). For more information about the IGV, see Chapter 5, Visualization Tools, in the BeadStudio Framework User Guide.

This chapter describes the tools you can use for LOH and copy number analysis:

- ▶ B allele frequency
- ▶ Log R ratio
- ▶ Algorithm plug-ins
 - Autobookmarking algorithms
 - Column plug-ins

B Allele Frequency

The **B Allele Freq** for a sample shows the theta value for a SNP, corrected for cluster position. Cluster positions are generated from a large set of normal individuals. The B Allele Frequency can also be referred to as "copy angle" or "allelic composition."

It is easier to visualize genotyping data for all SNPs within a chromosomal region using **B Allele Freq** rather than theta values. This is true because **B Allele Freq** exhibits less locus-to-locus variation than the **theta** values for a given sample.

The transformation of theta values to allele frequencies allows for improved measurements and better visualization of both LOH and copy number changes.

B allele freq is described by the following equation:

B allele freq:

$$\begin{aligned}&= 0 \text{ if } \theta < t_{AA} \\&= 0.5 * (\theta - t_{AA}) / (t_{AB} - t_{AA}) \text{ if } \theta < t_{AB} \\&= 0.5 + 0.5 * (\theta - t_{AB}) / (t_{BB} - t_{AB}) \text{ if } \theta < t_{BB} \\&= 1 \text{ if } \theta \geq t_{BB}\end{aligned}$$

where:

- ▶ t_{AA} = mean theta value of all genotypes in the AA cluster plotted in polar normalized coordinates
- ▶ t_{AB} = mean theta value of all genotypes in the AB cluster plotted in polar normalized coordinates
- ▶ t_{BB} = mean theta value of all genotypes in the BB cluster plotted in polar normalized coordinates

Figure 76 shows a comparison of plotting theta and B Allele Freq for the same sample on chromosome 5. The B Allele Freq plot exhibits less variation than the theta value plot. Notice the three clusters representing two homozygote clusters and one heterozygote cluster.

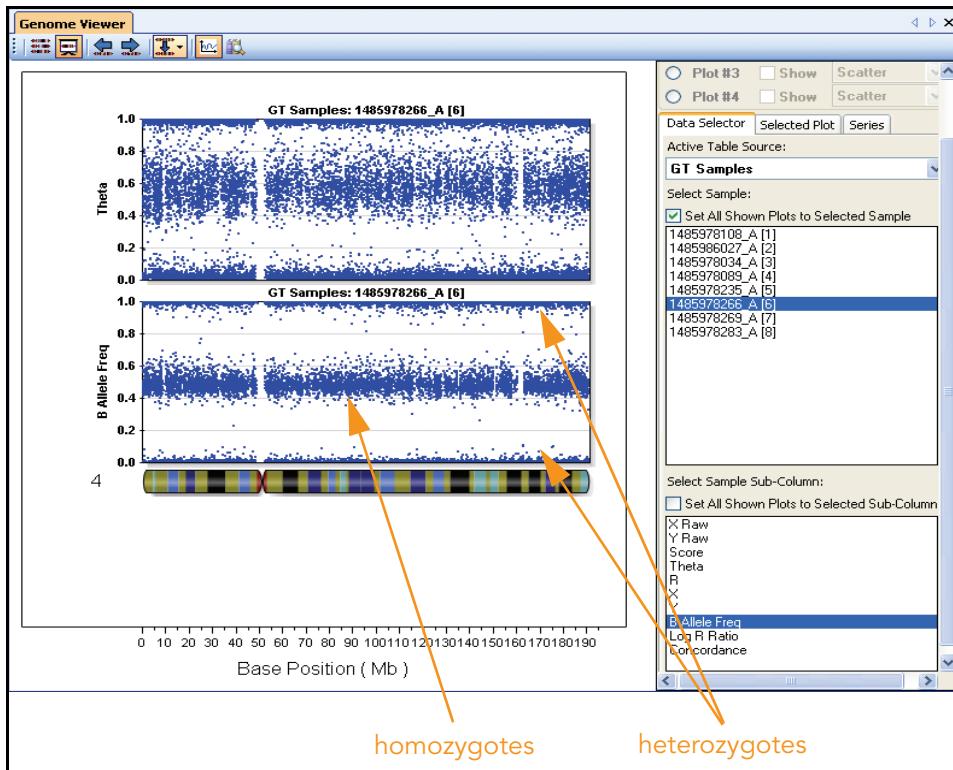


Figure 76 Theta vs. B Allele Frequency

Log R Ratio

The **Log R Ratio** subcolumn is based on normalized intensity data. In single-sample analysis mode, the Log R Ratio for a sample is the log (base 2) ratio of the normalized R value for the SNP divided by the expected normalized R value.

The expected R value is computed by linear interpolation of the R value at the SNP's theta value for a sample, relative to the R values of the surrounding clusters.

In paired-sample analysis mode, the Log R Ratio for a sample is the log (base 2) ratio of the normalized R value for the SNP from your subject sample divided by the normalized R value from your reference sample. In this case, the R values from the clusters are not used.

For example, if for a given sample and SNP with:

- A theta value of 0.2
- an AA cluster at theta = 0.1, R = 1.5
- an AB cluster at theta = 0.4, R = 2.5

The estimated R at theta for the sample is:

$$0.2 \text{ is } 1.5 + (0.2-0.1) * (2.5-1.5) / (0.4-0.1) = 1.83.$$

If the R value for the SNP is 1.6, the Log R Ratio is:

$$\log_2 (1.6/1.83) = -0.196.$$

Figure 77 shows an example of a log R ratio plot.

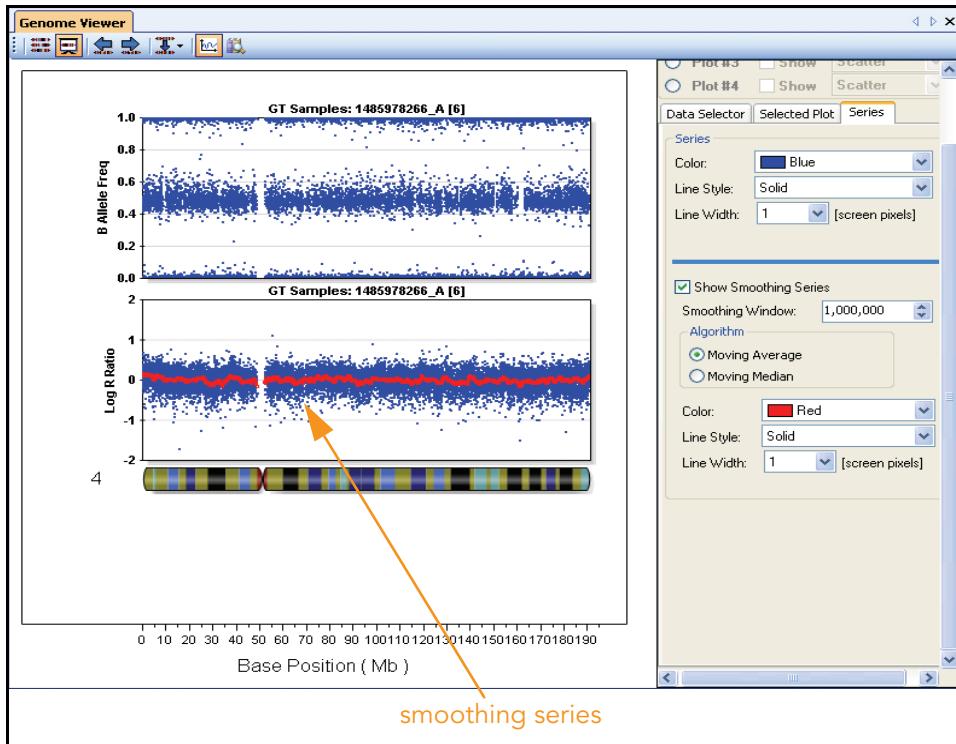


Figure 77 Log R Ratio

In Figure 77, a region of LOH is shown on the p arm of chromosome 5. This LOH event can also be demonstrated by a decrease in the log R ratio. The red line in the log R ratio plot indicates a smoothing series with a 200kb moving average window.

Algorithm Plug-ins

Illumina provides several algorithm plug-ins that you can use for LOH, copy number analysis, or other types of analysis. These plug-ins are included on your BeadStudio Genotyping Module CD and can be optionally installed when you install the BeadStudio software.

- ▶ **Autobookmarking plug-ins** are external code libraries that create bookmarks in the IGV based on data that appears in BeadStudio tables and chromosomal position information. You can access autobookmarking plug-ins from the IGV **Analysis** menu.
- ▶ **Plug-in columns** are external code libraries that create new subcolumns based on data that appears in BeadStudio tables. You can access column plug-ins by selecting **Analysis** | **Create Plug-In Column** from the BeadStudio Genotyping Module main window.
- ▶ **Report plug-ins** are customized reports provided by third parties. You can access report plug-ins in the report wizard. If report plug-ins are available, their names automatically appear in the **Custom Report** dropdown menu (Figure 78).

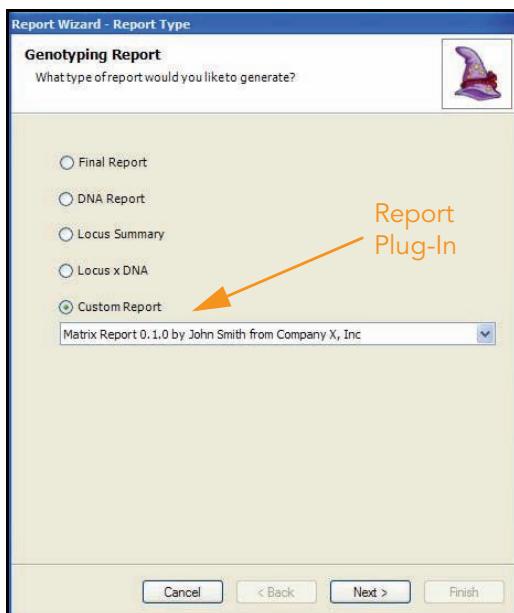


Figure 78 Selecting a Report Plug-In

Using Autobookmarking Algorithms

You can view the bookmarks created by an autobookmarking plug-in in the IGV, the ICB, and the Bookmark Viewer.

To apply autobookmarking algorithms to your data, perform the following steps:

1. After your data have been loaded into BeadStudio, select **Analysis | Genome Viewer** to launch the IGV.

The IGV appears (Figure 79).

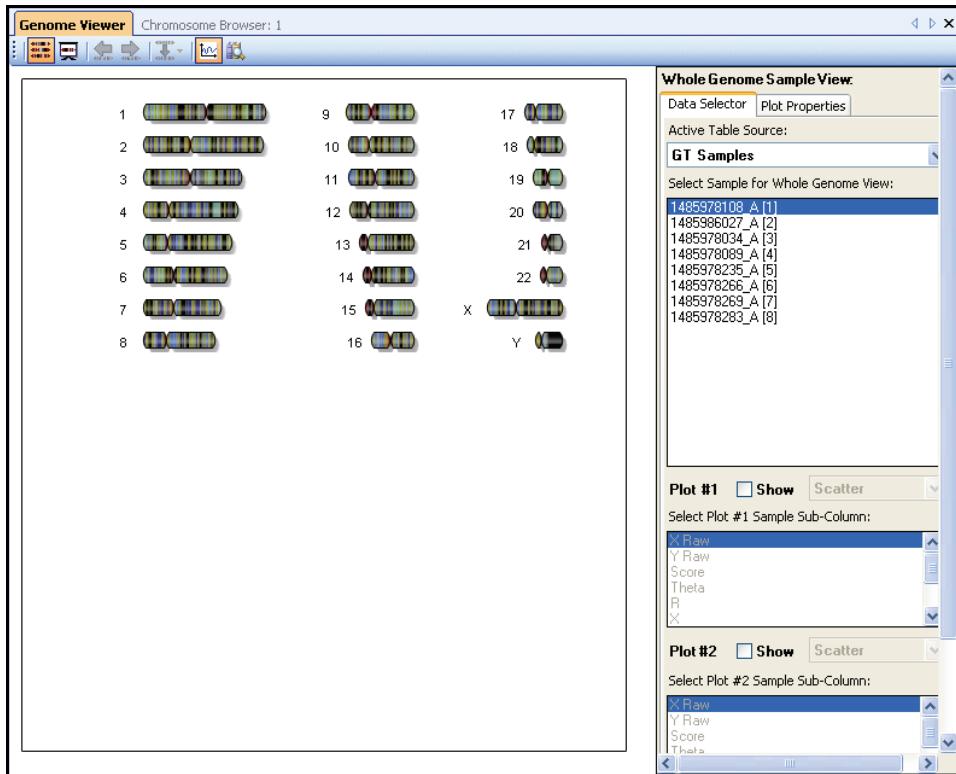


Figure 79 Illumina Genome Viewer

2. Select **Analysis | Run Autobookmark**.

The **Autobookmark Analysis** dialog box appears (Figure 80).

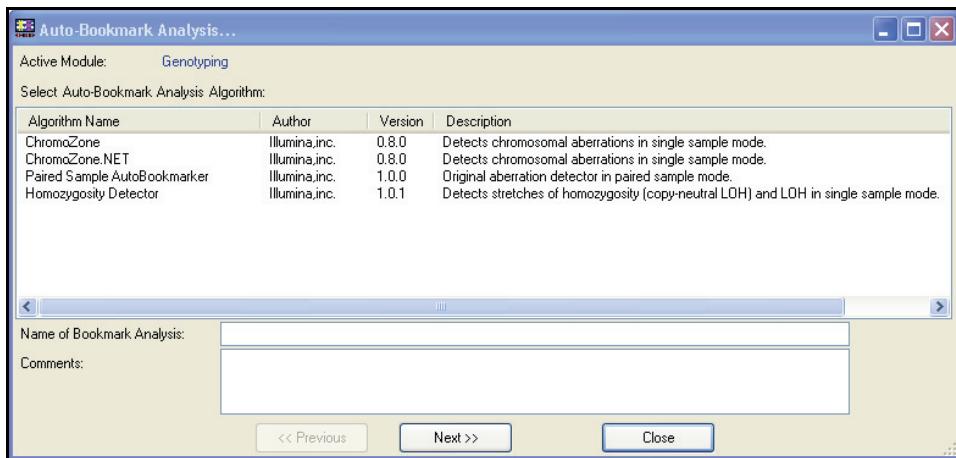


Figure 80 Autobookmark Analysis

The autobookmarking algorithms you have installed appear in the list of available algorithms.

3. Click an algorithm name to select an algorithm.
4. Enter a name for your bookmark analysis in the **Name of Bookmark Analysis** text field.
The bookmark analysis name will be visible in the **Data View** area under **Bookmark Analyses**.



NOTE

You can display the results of any bookmark analysis you have previously run by clicking its name in the **Bookmark Analyses** area.

5. **[Optional]** Enter comments in the **Comments** text field.
6. Click **Next** to advance to the next dialog box.
7. If the algorithm you want to use has editable properties, make selections from the available options.

**NOTE**

You may not be able to edit the input parameters of some algorithms supplied by Illumina.

If you cannot edit the input parameters, you will see the following message displayed in red, in the upper right-hand corner of the dialog box: Algorithm doesn't expose input parameters.

Continue to Step 8.

8. Click **Next.****9. Select the samples you want to include in this autobookmarking analysis.**

You can select all samples or any combination of samples provided that pairs are selected for the paired sample analysis (Figure 81).

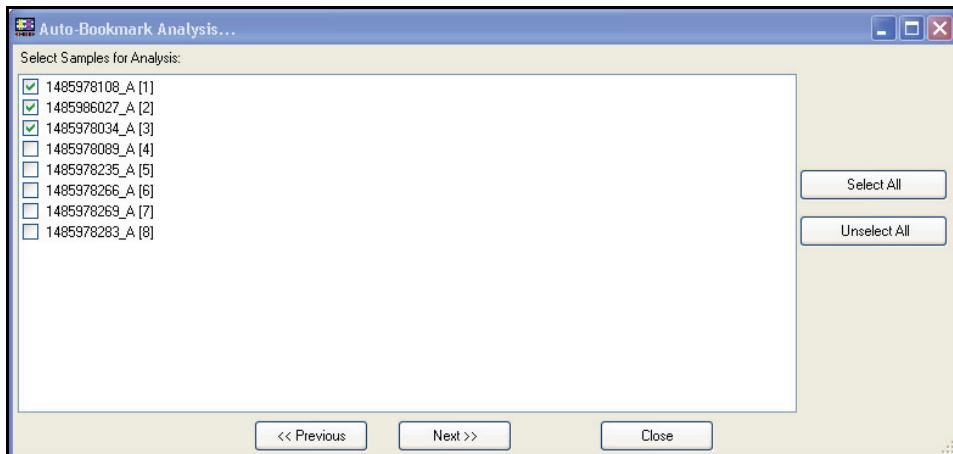


Figure 81 Selecting Samples for Analysis

10. Click **Next to advance to the next dialog box (Figure 82).**

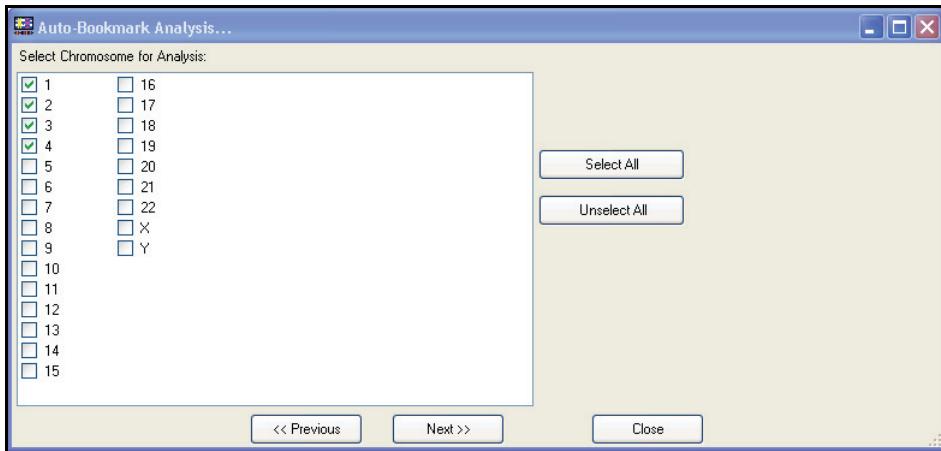


Figure 82 Selecting Chromosomes for Analysis

11. Select one or more chromosomes for analysis.

You can select all chromosomes or any combination of chromosomes.

12. Click **Next** to advance to the next dialog box (Figure 83).

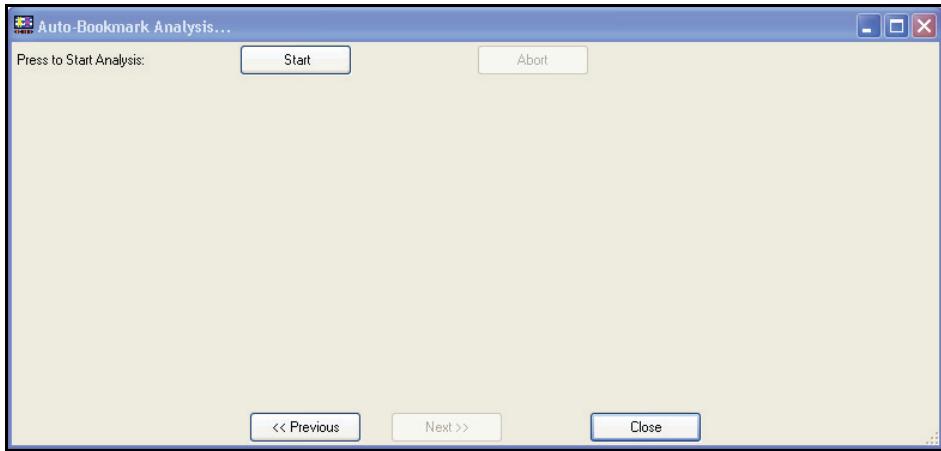


Figure 83 Autobookmark Analysis

13. Click **Start** to run the autobookmarking analysis.

The algorithm progress bar appears.

The **Algorithm Message Log** shows the progress as the algorithm is applied to your data.

14. When the analysis is complete, a message appears in the **Algorithm Message Log** (Figure 84).

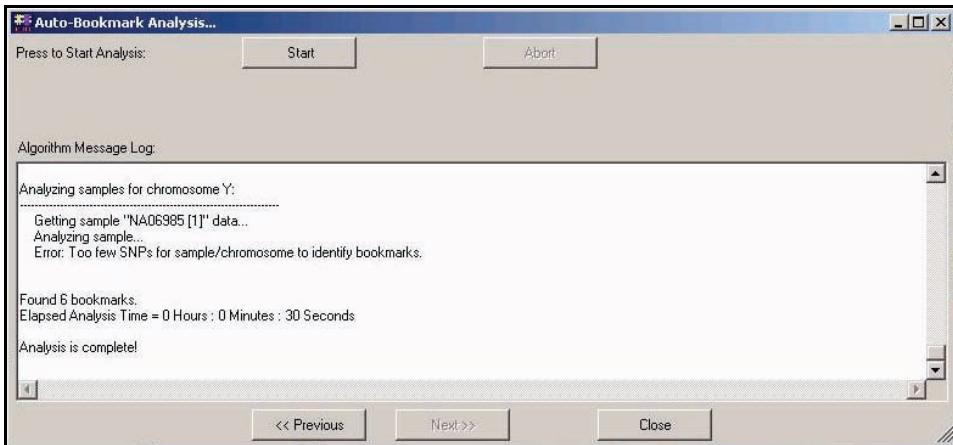


Figure 84 Analysis is Complete

15. Click **Close**.

Bookmarks appear in the IGV, the ICB, and the Bookmark Viewer.

Using Column Plug-Ins

All column plug-ins are accessed and run through the BeadStudio Genotyping Module main window. The results of applying the column plug-ins appear in the **Full Data Table**, the IGV, and the ICB.

To apply column plug-ins to your data, perform the following steps:

1. Select **Analysis | Create Plug-In Column**.

The **Select Column Plug-In Form** dialog box appears (Figure 85).

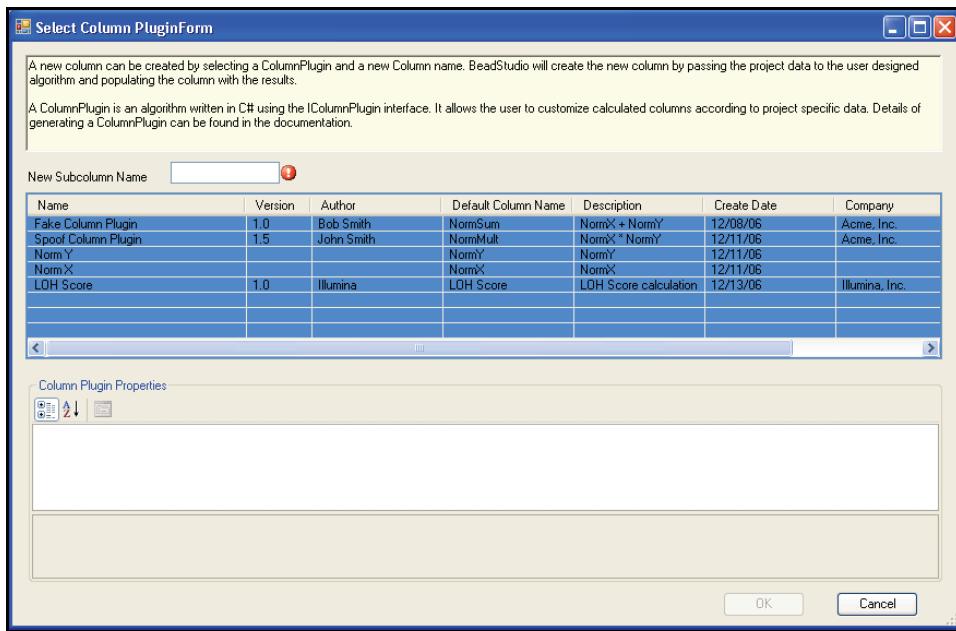


Figure 85 Select Column Plug-In Form

2. In the column plug-ins table, click to select a row from the list of available column plug-ins.
 3. **[Optional]** Type a name for the subcolumn in the **New Subcolumn Name** text field.
 4. **[Optional]** Edit the pre-defined properties of a column by clicking in the right-hand column of the **Column Plug-In Properties** table and entering new values.
 5. Click **OK**.
- The new subcolumn is created and appears in the **Full Data Table**. You can also view the results of applying this algorithm in available visualization tools.

Chapter 8

Generating Reports

Topics

- 96 Introduction
- 96 Final Report
- 104 DNA Report
- 107 Locus Summary Report
- 113 Locus x DNA Report
- 115 Reproducibility and Heritability Report
- 119 Error Types

Introduction

This chapter describes the BeadStudio Genotyping Module report types and how to generate each of these reports.

BeadStudio v3 includes a new Report Wizard, which streamlines the report creation process for the following report types:

- ▶ Final Report
- ▶ DNA Report
- ▶ Locus Summary Report
- ▶ Locus x DNA Report

In addition, if any report plug-ins are available, the name of the plug-in report automatically appears at the bottom of the report type list in the **Report Type** dialog box (Figure 86).

BeadStudio also allows you to manually create a Reproducibility and Heritability Report.



NOTE

The following sections describe the general process for creating reports. If your data includes zeroed SNPs or excluded samples, or if your data tables have been filtered, you may be presented with additional dialogs which allow you to filter the resulting report data.

Final Report

A Final Report is the final output of the BeadStudio Genotyping Module.

1. To generate a Final Report, run the Report Wizard by selecting **Analysis | Reports | Report Wizard**.

The **Report Type** dialog box appears (Figure 86).

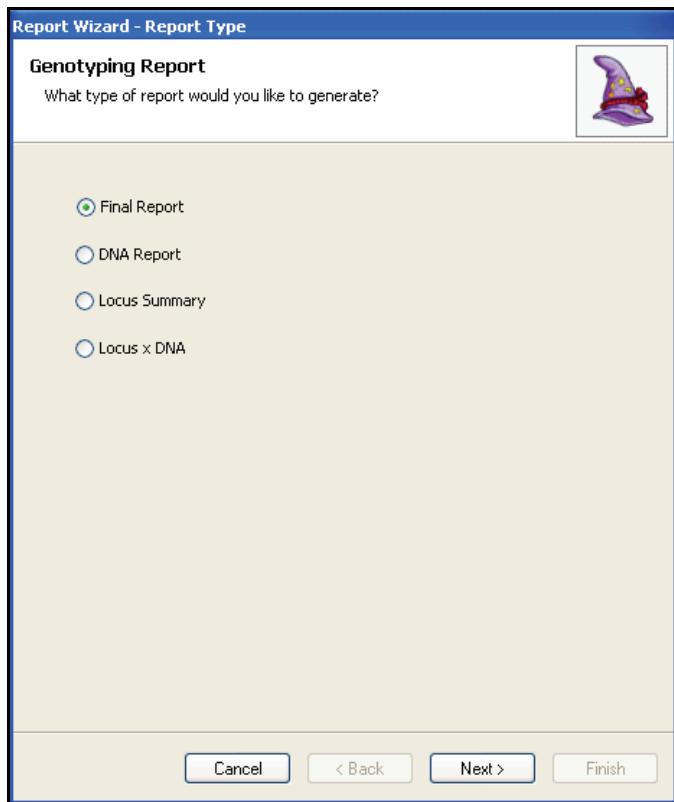


Figure 86 Report Type

Final Report is selected by default.

2. Click Next.

The **Final Report Format** dialog box appears (Figure 87).

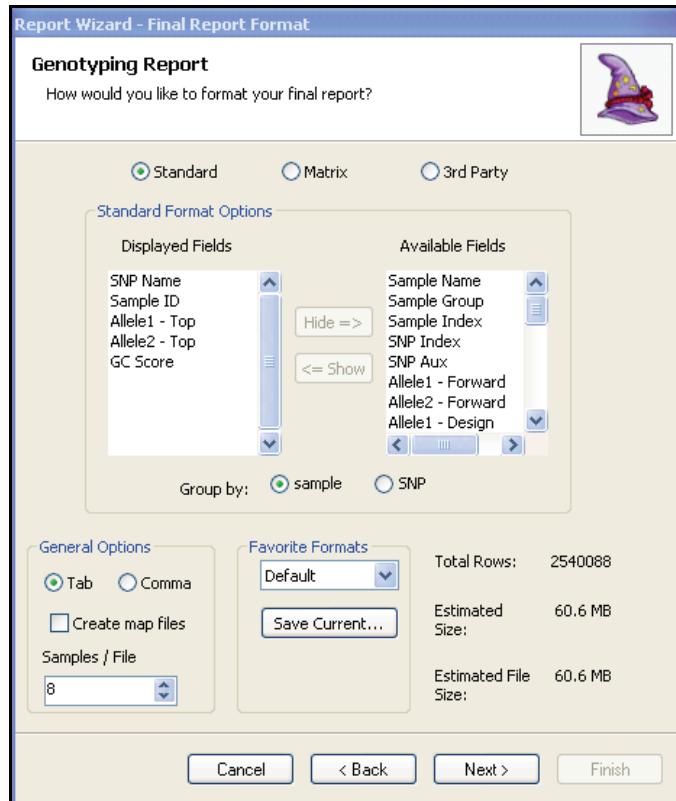


Figure 87 Final Report Format

3. Select one of the following options:
 - **Standard**—In **Standard** format, all data are presented in rows in the Final Report. You can choose the fields that will be included in a standard Final Report.
See *Final Report - Standard Format* on page 99.
 - **Matrix**—In **Matrix** format, rows represent SNPs and columns represent samples. You can choose to include the GenCall score or just output the genotypes.
See *Final Report - Matrix Format* on page 100.
 - **3rd Party**—In **3rd Party** format, you can specify the desired output style of the Final Report based on the target application for downstream analyses.
See *Final Report - 3rd Party Options* on page 101.

► Final Report - Standard Format

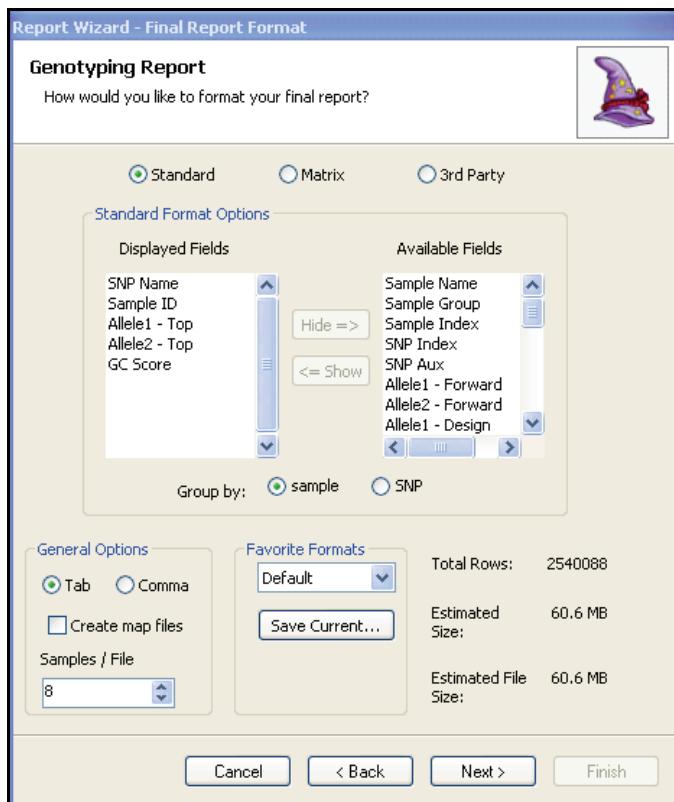


Figure 88 Final Report - Standard Format Options

- d. To select the fields included in your Final Report, select one or more fields from the **Available Fields** list and click **Show** to add them to the **Displayed Fields** List.
- e. Choose whether you want to group by sample or by SNP.
- f. Continue to Step 4.

► Final Report - Matrix Format

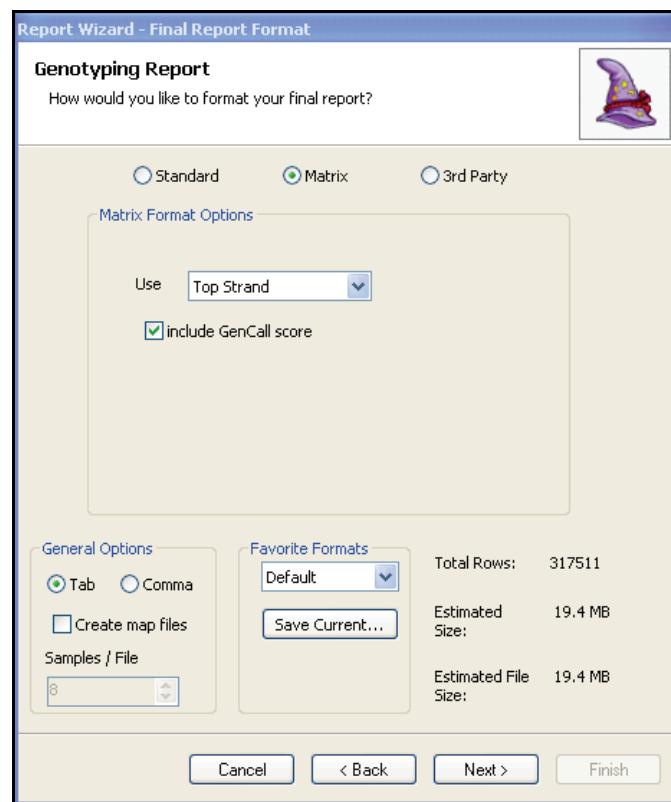


Figure 89 Final Report - Matrix Format Options

- a. In the **Use** dropdown menu, select one of the following options:
 - Top strand
 - Forward strand
 - Design strand
 - AB
- b. If you want to include GenCall scores in your Final Report, select **Include GenCall Score**.
- c. Continue to Step 4.

► Final Report - 3rd Party Options

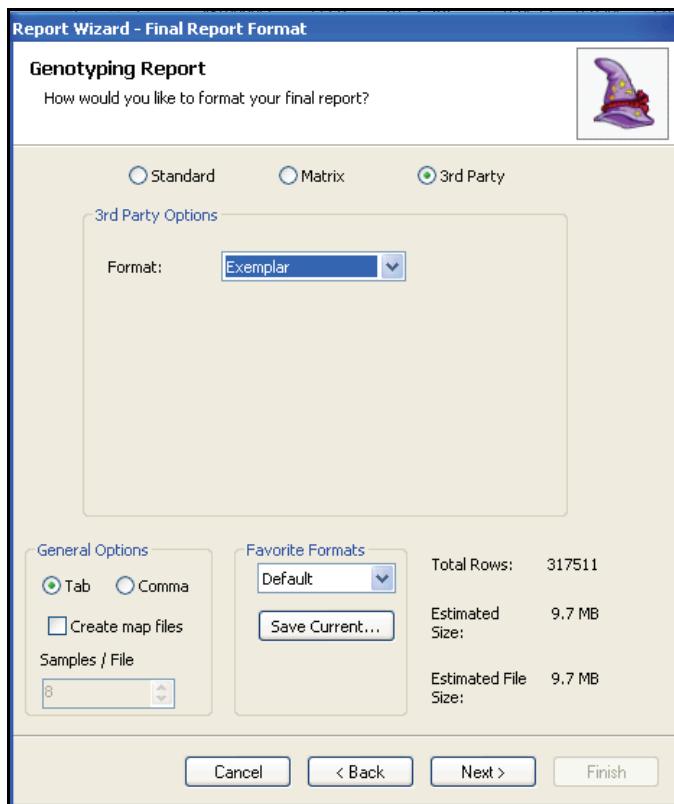


Figure 90 Final Report - 3rd Party Options

- a. Select a third party format for your Final Report from the 3rd Party Options **Format** dropdown menu.



NOTE

Currently-available 3rd party formats for Final Reports include Exemplar and GeneSpring.

4. In the **General Options** area, choose from among the following options:
 - Select **Tab** to create the Final Report in tab-delimited format, or select **Comma** to create the Final Report in comma-delimited format.
 - Select **Create map files** if you want to create map files.
 - Use the arrows to the right of **Samples / File** to specify the number of samples per file to include in the Final Report.
 - b. Select a favorite format: **Default** or **Default small**
 - c. Click **Save Current** to save your current selections as the default selections when creating subsequent Final Reports.
5. Click **Next**.

The **Destination** dialog box appears (Figure 91).



Figure 91 Destination

6. Browse to select an output path for your Final Report.

7. A report name is generated by default.

You can give your Final Report a different name by typing a name in the **Report Name** text field.

8. Click **Finish**.

Your Final Report is saved with the name and parameters you assigned to it in the location you specified.

	A	B	C	D	E	F
1	[Header]					
2	BSGT Version	2.2.20.34436				
3	Processing Date	3/17/2006 9:20				
4	Content		GS0006492-OPA			
5	Num SNPs	1536				
6	Total SNPs	1536				
7	Num Samples	60				
8	Total Samples	60				
9	[Data]					
10	SNP Name	Sample ID	Allele1 - Top	Allele2 - Top	GC Score	
11	rs1867749	SAM1_R001_C001	C	G	0.8237	
12	rs1397354	SAM1_R001_C001	G	G	0.8111	
13	rs2840531	SAM1_R001_C001	G	G	0.7827	
14	rs649593	SAM1_R001_C001	G	G	0.2586	
15	rs1517342	SAM1_R001_C001	G	G	0.7639	
16	rs1517343	SAM1_R001_C001	A	A	0.8332	
17	rs1868071	SAM1_R001_C001	G	G	0.765	
18	rs7611162	SAM1_R001_C001	C	C	0.8061	
19	rs911903	SAM1_R001_C001	A	G	0.7764	
20	rs753846	SAM1_R001_C001	G	G	0.8085	
21	rs2477703	SAM1_R001_C001	A	A	0.8577	
22	rs558912	SAM1_R001_C001	A	A	0.835	
23	rs3571116	SAM1_R001_C001	G	G	0.5707	
24	rs734999	SAM1_R001_C001	A	G	0.8752	
25	rs2377041	SAM1_R001_C001	A	G	0.7216	
26	rs7551616	SAM1_R001_C001	G	G	0.8574	
27	rs715494	SAM1_R001_C001	A	C	0.3153	
28	rs223201	SAM1_R001_C001	G	G	0.7982	
29	rs213006	SAM1_R001_C001	C	C	0.8631	
30	rs5203954	SAM1_R001_C001	G	G	0.8931	
31	rs874515	SAM1_R001_C001	C	G	0.2524	
32	rs10000021	SAM1_R001_C001	C	C	0.8395	
33	rs2030162	SAM1_R001_C001	G	G	0.662	
34	rs1489396	SAM1_R001_C001	G	G	0.7904	
35	rs742230	SAM1_R001_C001	A	G	0.8054	
36	rs718391	SAM1_R001_C001	C	C	0.7889	
37	rs667089	SAM1_R001_C001	T	T	0.8319	
38	rs8559	SAM1_R001_C001	A	A	0.858	
39	rs1970168	SAM1_R001_C001	A	G	0.7675	
40	rs2606418	SAM1_R001_C001	A	G	0.8896	
41	rs2027262	SAM1_R001_C001	A	A	0.6481	

Figure 92 Sample Final Report

DNA Report

To generate a DNA Report:

1. Run the Report Wizard by selecting **Analysis | Reports | Report Wizard**.
The **Report Type** dialog box appears.
2. Select **DNA Report** (Figure 93).

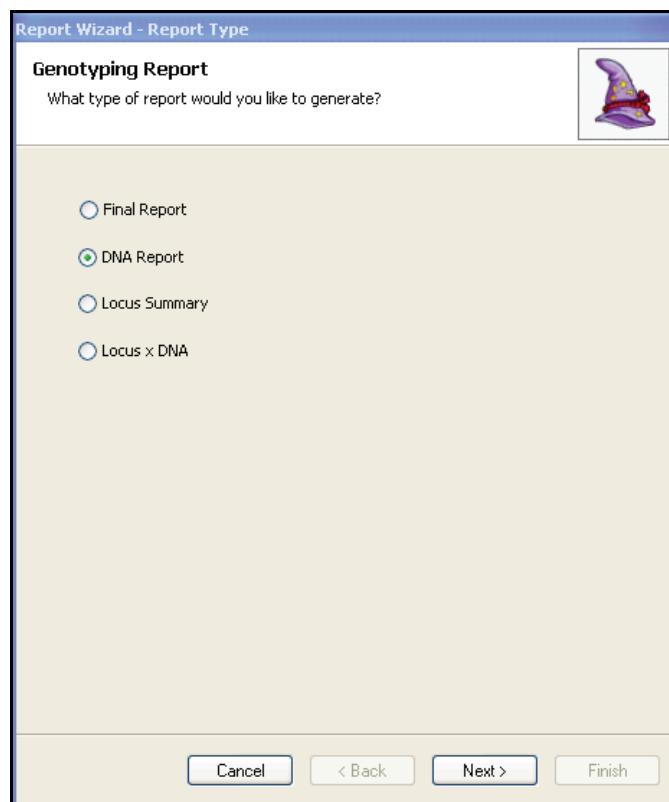


Figure 93 DNA Report Selected

3. Click **Next**.

The **Destination** dialog box appears (Figure 94).



Figure 94 Destination

4. Browse to select an output path for your DNA Report.
5. A report name is generated by default. You can give your DNA Report a different name by typing the name in the **Report Name** text field.
6. Click **Finish**.
Your DNA Report is saved with the name and parameters you assigned to it in the location you specified.

DNA Report on C:\Documents and Settings\user\Desktop\GT Project DNA REPORT.csv												
Row	DNA_Nam	#No_Calls	#Calls	Call_Freq	A/A_Freq	A/B_Freq	B/B_Freq	Minor_Fre	50%_GC	10%_GC	:0/1	
1	GS000280	47	1489	0.9694	0.2579	0.4681	0.274	0.4919	0.7611	0.4754	0	
2												
3												
4	GS000280	1	1535	0.9993	0.2664	0.4645	0.2691	0.4987	0.7883	0.6674	1	
5	GS000280	1	1535	0.9993	0.2697	0.4547	0.2756	0.4971	0.7876	0.669	1	
6	GS000280	1	1535	0.9993	0.2762	0.4241	0.2997	0.4883	0.7882	0.669	1	
7	GS000280	1	1535	0.9993	0.284	0.4502	0.2658	0.4909	0.7883	0.669	1	
8	GS000280	1	1535	0.9993	0.2877	0.4553	0.257	0.4847	0.787	0.6657	1	
9	GS000280	3	1533	0.998	0.2877	0.4553	0.257	0.4847	0.7883	0.6674	1	
10	GS000280	1	1535	0.9993	0.2612	0.4599	0.2788	0.4912	0.7883	0.669	1	
11	GS000280	3	1533	0.998	0.2838	0.4325	0.2838	0.5	0.7881	0.669	1	
12	GS000280	1	1535	0.9993	0.2945	0.4371	0.2684	0.487	0.7882	0.6675	1	
13	GS000280	1	1535	0.9993	0.2619	0.4508	0.2873	0.4873	0.7883	0.6692	1	
14	GS000280	1	1535	0.9993	0.2775	0.4456	0.2769	0.4997	0.7878	0.668	1	
15	GS000280	2	1534	0.9987	0.2627	0.4446	0.2927	0.485	0.7883	0.6675	1	
16	GS000280	1	1535	0.9993	0.2489	0.4678	0.2834	0.4827	0.7814	0.6351	1	
17	GS000280	2	1534	0.9987	0.2823	0.4335	0.2842	0.499	0.7884	0.669	1	
18	GS000280	3	1533	0.998	0.2825	0.4331	0.2844	0.499	0.7883	0.669	1	
19	GS000280	1	1535	0.9993	0.2775	0.4554	0.2671	0.4948	0.7883	0.669	1	
20	GS000280	2	1534	0.9987	0.2627	0.4628	0.2744	0.4941	0.7872	0.6679	1	
21	GS000280	2	1534	0.9987	0.2718	0.4381	0.2901	0.4909	0.788	0.6669	1	
22	GS000280	1	1535	0.9993	0.2782	0.4378	0.284	0.4971	0.7883	0.668	1	
23	GS000280	1	1535	0.9993	0.2749	0.4358	0.2893	0.4928	0.7883	0.669	1	
24	GS000280	2	1534	0.9987	0.2816	0.4439	0.2744	0.4964	0.7884	0.6692	1	
25	GS000280	2	1534	0.9987	0.2725	0.442	0.2855	0.4935	0.7884	0.6692	1	
26	GS000280	2	1534	0.9987	0.2725	0.4596	0.2679	0.4977	0.7883	0.669	1	

Figure 95 Sample DNA Report

The DNA Report is a comma-delimited text file (*.csv file) with the following columns:

Table 2 DNA Report - File Column Descriptions

Column Name	Description
Row	Row number
DNA_Name	DNA name
#No_Calls	Number of loci with GenCall scores below the call region threshold (Tools Options Flags)
#Calls	Number of loci with GenCall score above the call region threshold
Call_Freq	Call frequency, or call rate, calculated as follows: #Calls/(#No_Calls + #Calls)

Table 2 DNA Report - File Column Descriptions (continued)

Column Name	Description
A/A_Freq	Frequency of homozygote allele A calls
A/B_Freq	Frequency of heterozygote calls
B/B_Freq	Frequency of homozygote allele B calls
Minor_Freq	Frequency of the minor allele
50%_GC_Score	GenCall score at the 50% rank when scores are ranked for all loci
10%_GC_Score	GenCall score at the 10% rank when scores are ranked for all loci
0/1	Indicates whether the sample is to be included or removed, as follows: 0 Remove 1 Include

Locus Summary Report

To generate a Locus Summary Report:

1. Run the Report Wizard by selecting **Analysis | Reports | Report Wizard**.
The **Report Type** dialog box appears.
2. Select **Locus Summary Report** (Figure 93).

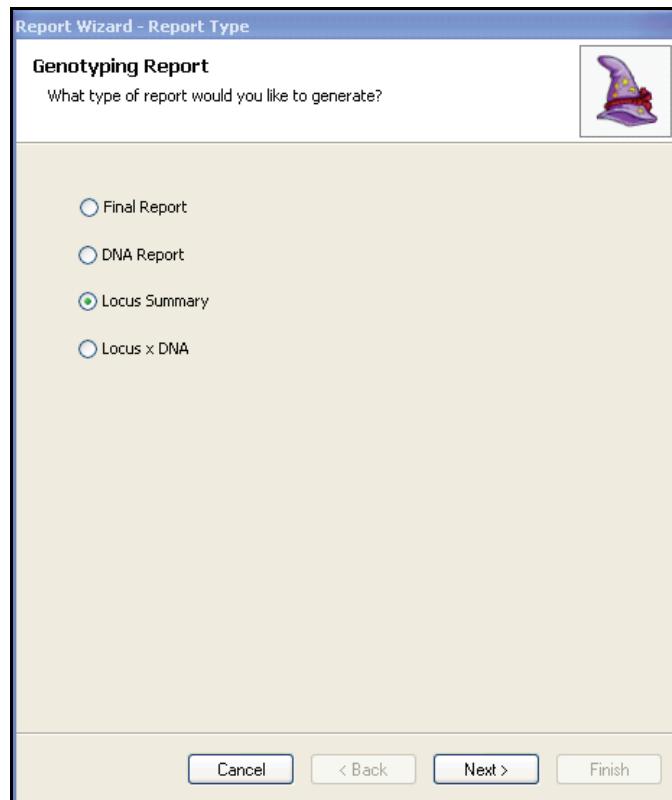


Figure 96 Locus Summary Report Selected

3. Click **Next**.
- The **Destination** dialog box appears.



Figure 97 Destination - Locus Summary

4. Browse to select an output path for your Locus Summary Report.
5. A report name is generated by default. You can give your Locus Summary Report a different name by typing the name in the **Report Name** text field.
6. Click **Finish**.

Your Locus Summary Report is saved with the name and parameters you assigned to it in the location you specified.

Locus Summary on C:\Documents and Settings\user\Desktop\GT Project.LOCUS_SUMMARY.csv													
Row	Locus_Nam	IllumiCode_Nam	#No_Calls	#Calls	Call_Freq	A/A_Freq	A/B_Freq	B/B_Freq	Minor_Freq	GenTrain_S50%	GC_10%	GC_110%	Het_Excess
1	rs1867749	3	0	60	1	0.367	0.433	0.15	0.367	0.8237	0.8237	0.8237	-0.0377
2	rs1397354	10	0	60	1	0.183	0.45	0.317	0.408	0.8928	0.8928	0.8928	-0.0336
3	rs2840531	21	0	60	1	0	0	0	0	0.7846	0.7846	0.7846	0
4	rs649593	23	0	60	1	0.083	0.517	0.35	0.342	0.6059	0.6059	0.6059	0.1806
5	rs1517342	27	0	60	1	0.233	0.5	0.217	0.467	0.7639	0.7639	0.7639	0.053
6	rs1517343	28	0	60	1	0.217	0.5	0.233	0.467	0.8332	0.8332	0.8332	0.053
7	rs1866071	30	0	60	1	0.117	0.533	0.3	0.383	0.765	0.765	0.765	0.1662
8	rs761162	31	0	60	1	0.3	0.483	0.167	0.408	0.8061	0.8061	0.8061	0.038
9	rs911903	33	0	60	1	0	0	0	0	0.7764	0.7764	0.7764	0
10	rs753646	36	0	60	1	0.25	0.467	0.233	0.467	0.8085	0.8085	0.8085	-0.0172
11	rs2477703	38	0	60	1	0.4	0.45	0.1	0.325	0.8755	0.8755	0.8755	0.0523
12	rs558912	40	0	60	1	0.133	0.417	0.4	0.342	0.8493	0.8493	0.8493	-0.0476
13	rs357116	41	0	60	1	0.217	0.55	0.183	0.458	0.5707	0.5707	0.5707	0.1593
14	rs734999	42	0	60	1	0.217	0.533	0.2	0.467	0.8752	0.8752	0.8752	0.1232
15	rs2377041	50	0	60	1	0.1	0.533	0.317	0.367	0.7216	0.7216	0.7216	0.1844
16	rs7551616	51	0	60	1	0.217	0.483	0.25	0.458	0.8574	0.8574	0.8574	0.0188
17	rs715494	54	0	60	1	0.15	0.433	0.367	0.367	0.8406	0.8406	0.8406	-0.0377
18	rs223201	56	0	60	1	0.05	0.35	0.55	0.225	0.7982	0.7982	0.7982	0.0192
19	rs213006	61	0	60	1	0.417	0.45	0.083	0.308	0.8631	0.8631	0.8631	0.0804
20	rs520354	62	0	60	1	0.267	0.417	0.317	0.475	0.8932	0.8932	0.8932	-0.1646
21	rs874515	63	0	60	1	0.55	0.433	0.017	0.233	0.8455	0.8455	0.8455	0.2112
22	rs1000021	66	0	60	1	0.433	0.4	0.167	0.367	0.8395	0.8395	0.8395	-0.1388
23	rs2030162	67	0	60	1	0.233	0.5	0.267	0.483	0.765	0.765	0.765	0.0011
24	rs1496206	69	0	60	1	0.4	0.45	0.15	0.375	0.7904	0.7904	0.7904	0.04

Figure 98 Sample Locus Summary Report

The Locus Summary Report is a comma-delimited text file (.csv file) with the following columns:

Table 3 Locus Summary Report - File Column Descriptions

Column	Description
Row	Row number
Locus_Name	Locus name from the Manifest
IllumiCode_Name	Locus ID from the Manifest
#No_Calls	Number of samples with GenCall score below the call region threshold (Tools Options Flags)
#Calls	Number of samples with GenCall score above the call region threshold
Call_Freq	Call frequency, or call rate, calculated as follows: #Calls/(#No_Calls + #Calls)

Table 3 Locus Summary Report - File Column Descriptions

Column	Description
A/A_Freq	Frequency of homozygote allele A calls
A/B_Freq	Frequency of heterozygote calls
B/B_Freq	Frequency of homozygote allele B calls
Minor_Freq	Frequency of the minor allele
GenTrain_Score	A number between 0 and 1 indicating how well the samples clustered for this locus
50%_GC_Score	GenCall score at the 50th percentile when scores are ranked for all samples
10%_GC_Score	GenCall score at the 10th percentile when scores are ranked for all samples
Het_Excess_Freq	<p>Heterozygote excess frequency, calculated as (Observed - Expected)/Expected for the heterozygote class. If f_{AB} is the heterozygote frequency observed at a locus, and p and q are the major and minor allele frequencies, then het excess is defined as:</p> $(f_{AB} - 2pq)/(2pq + \epsilon)$ <p>The ϵ value regularizes the estimation of heterozygote excess frequency. This reduces the variance of the estimation for cases of extremely low minor allele frequency.</p>
ChiTest_P100	Hardy-Weinberg p-value estimate calculated using genotype frequency. The value is calculated with 1 degree of freedom and normalized to 100 individuals.
Cluster_Sep	Cluster separation score
AA_T_Mean	Mean of the normalized theta angles for the AA genotype
AA_T_Std	Standard deviation of the normalized theta angles for the AA genotype
AB_T_Mean	Mean of the normalized theta angles for the AB genotype
AB_T_Std	Standard deviation of the normalized theta angles for the AB genotype

Table 3 Locus Summary Report - File Column Descriptions

Column	Description
BB_T_Mean	Mean of the normalized theta angles for the BB genotypes
BB_T_Std	Standard deviation of the normalized theta angles for the BB genotypes
AA_R_Mean	Mean of the normalized r-values for the AA genotypes
AA_R_Std	Standard deviation of the normalized r-values for the AA genotypes
AB_R_Mean	Mean of the normalized r-values for the AB genotypes
AB_R_Std	Standard deviation of the normalized r-values for the AB genotypes
BB_R_Mean	Mean of the normalized r-values for the BB genotypes
BB_R_Std	Standard deviation of the normalized r-values for the BB genotypes

Locus x DNA Report

To generate a Locus x DNA Report:

1. Run the Report Wizard by selecting **Analysis | Reports | Report Wizard**.
The **Report Type** dialog box appears.
2. Select **Locus x DNA Report** (Figure 93).

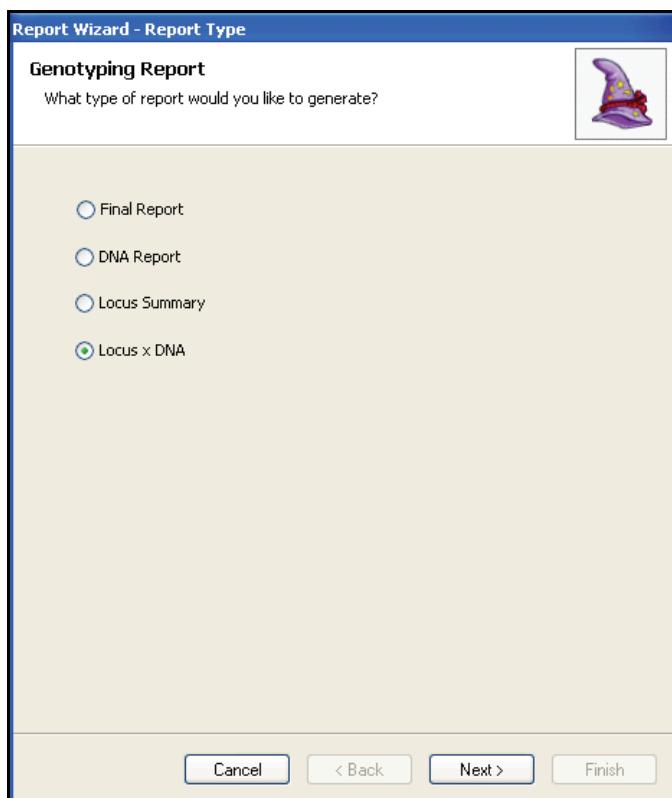


Figure 99 Locus x DNA Selected

3. Click **Next**.
The **Destination** dialog box appears.



Figure 100 Destination - Locus x DNA

4. Browse to select an output path for your Locus x DNA Report.
5. A report name is generated by default. You can give your Locus x DNA Report a different name by typing the name in the **Report Name** text field.
6. Click **Finish**.
7. Your Locus x DNA Report is saved with the name and parameters you assigned to it in the location you specified.

Table 4 Locus by DNA Report - File Column Descriptions

Column Name	Description
instituteLabel	Customer's unique sample ID for the DNA sample.
plateWell	Concatenation of the Sample Plate and Sample Well.
imageDate	Imaging date for that sample.
oligoPoolId	Name of the OPA (e.g., GS0001111-OPA)
bundleId	Identifier of the bundle which includes the array barcode + row + column + customer provided non-unique sample name.
status	Flag for whether or not these data came from the last run through Autogenopipe (0 = last run, >0 = older runs)
recordType	Identifies each row of data in the file as "calls" or "Score_Call". Each row of data in the file is for each DNA sample; there will be two rows of data for each DNA sample (one with "A", "B" or "H" = call and another with the corresponding Gencall score for that call)
data	Actual data (calls or scores) for each DNA sample and locus

Reproducibility and Heritability Report

The Reproducibility and Heritability Report is the error output of the BeadStudio Genotyping Module.

To generate a Reproducibility and Heritability Report:

1. Select **Analysis | Reports | Create Reproducibility and Heritability Report**.

The Reproducibility and Heritability dialog box appears.

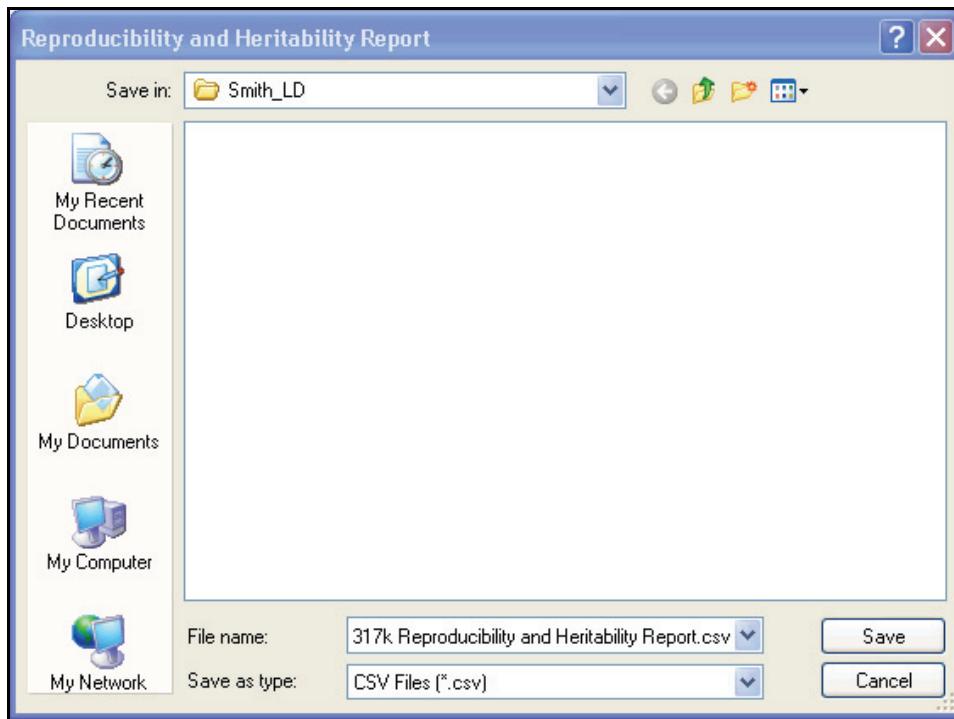


Figure 101 Reproducibility and Heritability

2. In the **File Name** text box, a default name appears for the report. You can leave the name as it is or make changes.
 3. In the **Save In** dropdown menu at the top of the screen or to the left of the main window, browse to the location where you would like to save the report.
 4. Click **Save** to save the report.
- The **View Reproducibility and Heritability Report** dialog box appears.

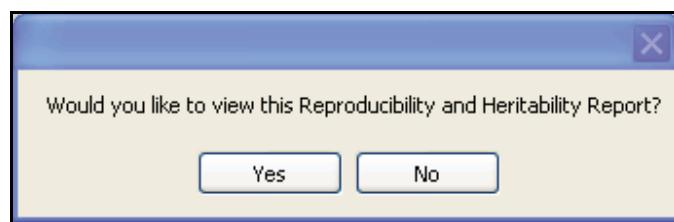


Figure 102 View Reproducibility and Heritability

5. Do one of the following:

- Click **Yes** to view the Reproducibility and Heritability Report. The Reproducibility and Heritability Report appears (Figure 103).

	A	B	C	D	E	F	G
1	Reproducibility and Heritability Report						
2	Filename: C:\Documents and Settings\<username>\Desktop\GT Project Reproducibility and Heritability Report.csv						
3	Run date: Thursday	December 06	2005 5:41:51 PM	# LOCI = 1536	# DNAs = 60	ProjectName = GT Project	GenCall Version = 5.2.0 Low Gen
4							
5	Duplicate Reproducibility						
6	Rep1_DNA_Name	Rep2_DNA_Name	# Correct	# Errors	Total	Repro_Freq	
7	GS0002800-DNA002-NA12156	GS0002800-DNA003-NA12156	1534	0	1534	1	
8	GS0002800-DNA001-NA06993	GS0002800-DNA012-NA06993	1533	1	1534	0.999674	
9							
10	P-C Heritability						
11	Parent_DNA_Name	Child_DNA_Name	# Correct	# Errors	Total	P-C Heritability Freq	
12							
13	P-P-C Heritability						
14	Parent1_DNA_Name	Parent2_DNA_Name	# Correct	# Errors	Total	P-P-C Heritability Freq	
15							

Figure 103 Sample Reproducibility and Heritability Report

- Click **No** if you do not want to view the Reproducibility and Heritability Report.

The Reproducibility and Heritability Report is saved at the location you specified, but it does not display. You can return to it later.

Table 5 describes the columns of the Reproducibility and Heritability Report.

Table 5 Reproducibility and Heritability Report - File Column Descriptions

Column	Description
Duplicate Reproducibility	
Rep1_DNA_Name	Name of the sample designated as replicate #1.
Rep2_DNA_Name	Name of the sample designated as replicate #2.
# Correct	Number of correct reproducibility calculations.
# Errors	Number of erroneous reproducibility calculations.

Table 5 Reproducibility and Heritability Report - File Column Descriptions

Column	Description
Total	Total number of SNPs used to calculate reproducibility.
Repro_Freq	Reproducibility frequency, calculated as $\sqrt{1 - \text{error rate}}$. The error rate does not include genotype calls that fall below the no-call threshold.
P-C Heritability	
Parent_DNA_Name	Sample_ID designated as parent in a P-C relationship.
Child_DNA_Name	Sample_ID designated as child in a P-C relationship.
# Correct	Number of correct P-C heritability calculations.
# Errors	Number of erroneous P-C heritability calculations.
Total	Total number of SNPs used to calculate P-C heritability.
PC_Heritability_Freq	P-C heritability frequency calculated as: $\frac{\# \text{ correct}}{\# \text{ errors} + \# \text{correct}}$
P-P-C Heritability	
Parent1_DNA_Name	Sample_ID designated as parent #1 in a P-P-C relationship.
Parent2_DNA_Name	Sample_ID designated as parent #2 in a P-P-C relationship.
Child_DNA_Name	Sample_ID designated as child in a P-P-C relationship.
# Correct	Number of correct P-P-C heritability calculations.
# Errors	Number of erroneous P-P-C heritability calculations.
Total	Total number of SNPs used to calculate P-P-C heritability.
P-P-C Heritability Freq	P-P-C heritability frequency calculated as: $\frac{\# \text{ correct}}{\# \text{ errors} + \# \text{correct}}$

Error Types

The following tables describe the possible error types that may be included in the Reproducibility and Heritability Report.

Table 6 Reproducibility Errors

GT1	GT2	Error
AA	AA	N
AA	AB	Y
AA	BB	Y
AB	AA	Y
AB	AB	N
AB	BB	Y
BB	AA	Y
BB	AB	Y
BB	BB	N

Table 7 Parent-Child Heritability Errors

GT Parent	GT Child	Error
AA	AA	N
AA	AB	N
AA	BB	Y
AB	AA	N
AB	AB	N
AB	BB	N
BB	AA	Y
BB	AB	N
BB	BB	N

Table 8 Parent-Parent-Child Heritability Errors

GT Parent 1	GT Parent 2	GT Child	Error
AA	AA	AA	N
AA	AA	AB	Y
AA	AA	BB	Y
AA	AB	AA	N
AA	AB	AB	N
AA	AB	BB	Y
AA	BB	AA	Y
AA	BB	AB	N
AA	BB	BB	Y
AB	AA	AA	N
AB	AA	AB	N
AB	AA	BB	Y
AB	AB	AA	N
AB	AB	AB	N
AB	AB	BB	N
AB	BB	AA	Y
AB	BB	AB	N
AB	BB	BB	N
BB	AA	AA	Y
BB	AA	AB	N
BB	AA	BB	Y
BB	AB	AA	Y

Table 8 Parent-Parent-Child Heritability Errors (continued)

GT Parent 1	GT Parent 2	GT Child	Error
BB	AB	AB	N
BB	AB	BB	N
BB	BB	AA	Y
BB	BB	AB	Y
BB	BB	BB	N

Chapter 9

User Interface Reference

Topics

- 124 Introduction
- 125 Detachable Docking Windows
 - 125 Graph Window
 - 129 Data Table
 - 139 Samples Table
 - 145 Project Window
 - 146 Log Window
- 147 Main Window Menus
- 153 Graph Window Toolbar
- 154 Table Windows Toolbar
- 156 Context Menus

Introduction

The BeadStudio Genotyping Module user interface provides tools for loading intensity files, running the clustering algorithm, browsing loci, and displaying them graphically. Figure 104 shows the default window configuration of the BeadStudio Genotyping Module.

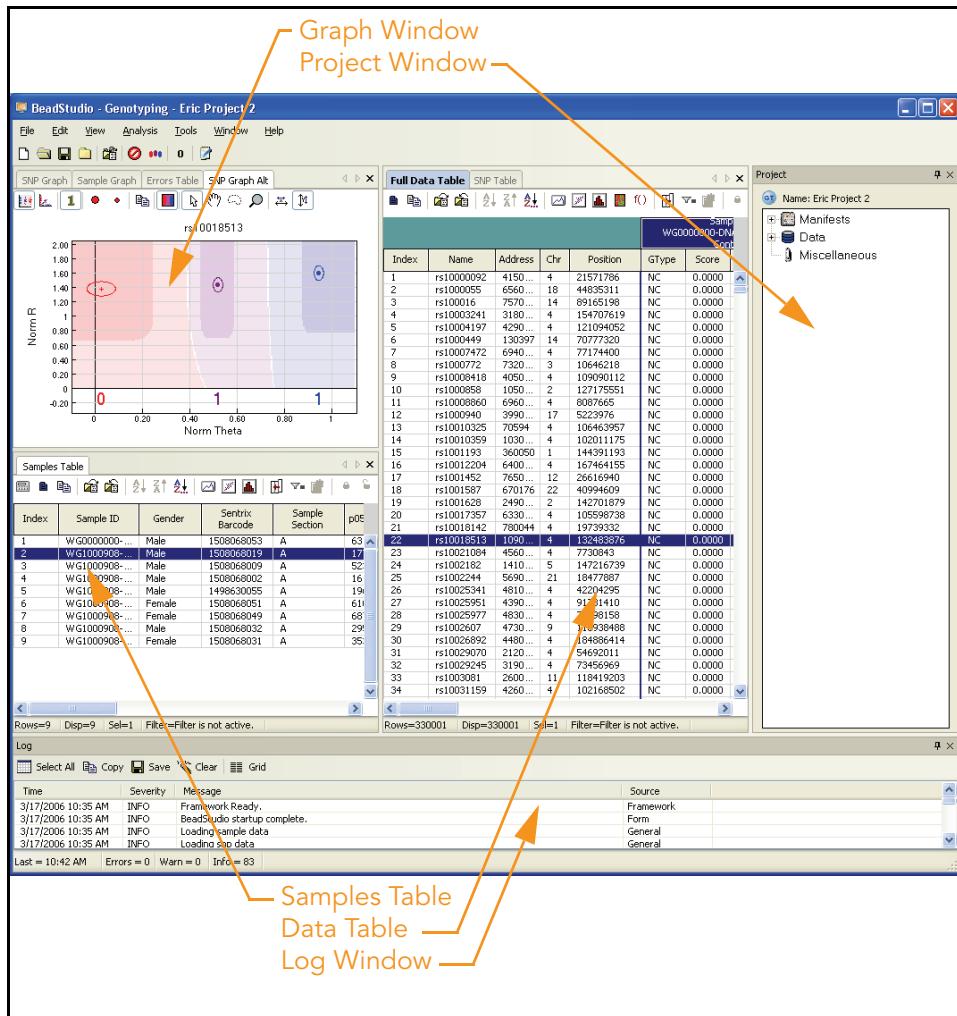


Figure 104 BeadStudio Genotyping Module Default View

Detachable Docking Windows

Detachable docking windows provide a flexible way to customize BeadStudio's user interface to suit your analysis needs.

The following sections describe each of the Genotyping Module's detachable docking windows and their component tabs.

Graph Window

The graph window contains the **SNP Graph** by default. In the graph window, you can toggle among the **SNP Graph**, the **Sample Graph**, the **Errors Table**, and the **SNP Graph Alt**.

SNP Graph

The **SNP Graph** plots all samples for the currently selected SNP in the **Full Data Table** or **SNP Table** (Figure 105).

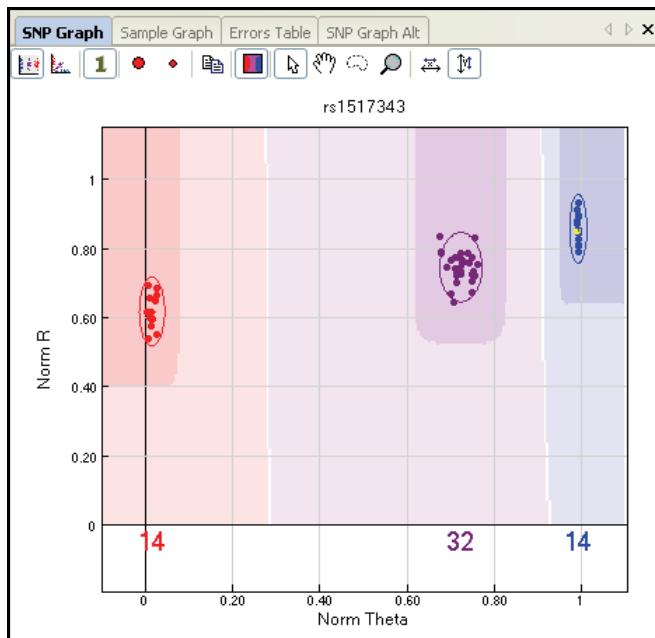


Figure 105 SNP Graph

Sample Graph

The **Sample Graph** (Figure 106) displays all SNPs for the currently-selected sample in the **Samples Table**. The SNPs are colored according to their genotype calls. Use the **Sample Graph** to evaluate sample quality.

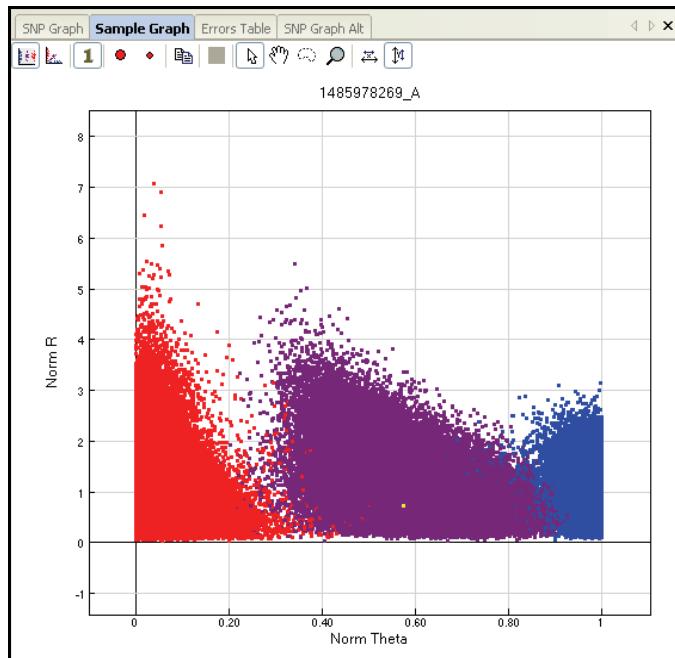


Figure 106 Sample Graph

Errors Table

The **Errors Table** (Figure 107) lists any reproducibility errors or parent-child heritability errors found in the data loaded into BeadStudio.

Errors Table								
Error Index	Error Type	Index 1	Name 1	GType 1	Index 2	Name 2	GType 2	Index 3
1	Rep	7	GS0002800-DNAA07-N...	AA	8	GS0002800-DNAA08-NA1...	AB	

Rows=1 | Disp=1 | Sel=0 | Filter=Filter is not active.

Figure 107 Errors Table

The columns in the **Errors Table** are listed and described in Table 9.

Table 9 Errors Table Columns

Column	Description	Type	Visible by Default?
Error Index	The row index of the error	integer	Y
Error Type	The type of error: <ul style="list-style-type: none"> • Rep—Reproducibility • P-C—Parent-Child heritability • P-P-C—Parent-Parent-Child heritability 	string	Y
Index 1	The sample index of the first sample involved in the error	integer	Y
Name 1	The sample id of the first sample involved in the error	string	Y
GType 1	For a parental relationship error, this is the genotype of parent 1. For a replicate error, this is the genotype of replicate 1.	string	Y
Index 2	The sample index of the second sample involved in the error	integer	Y
Name 2	The sample id of the second sample involved in the error	string	Y

Table 9 Errors Table Columns (continued)

Column	Description	Type	Visible by Default?
GType 2	For a parental relationship error, this is the genotype of parent 2. For a replicate error, this is the genotype of replicate 2.	string	Y
Index 3	The sample index of the child involved in a P-P-C error (not used for Rep or P-C errors)	integer	Y
Name 3	The sample id of the child involved in a P-P-C error (not used for Rep or P-C errors)	string	Y
GType 3	For a P-P-C relationship error, this is the genotype of the child.	string	Y
SNP Index	The index number of the SNP where the error occurred.	integer	Y
SNP Name	The name of the SNP where the error occurred.	string	Y

SNP Graph Alt

The **SNP Graph Alt** is an alternate SNP graph that you can display along with the **SNP Graph** to compare different views within BeadStudio.

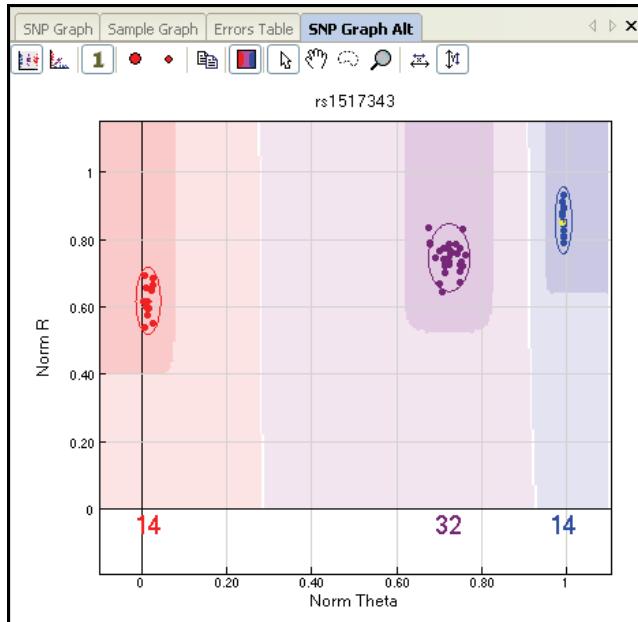


Figure 108 SNP Graph Alt

Data Table

The **Data Table** contains the **Full Data Table** by default. In the **Data Table**, you can toggle between the **Full Data Table** and the **SNP Table**.

Full Data Table

The **Full Data Table** (Figure 109) contains all data for every sample.

To sort the **Full Data Table** by any column:

1. Click the header of the column you want to use as a basis for sorting the table.
2. Do one of the following:

- ▶ Click  to sort by the column in ascending order.
- ▶ Click  to sort by the column in descending order.
- ▶ Click  to sort by multiple columns.

				Sample 1 G50002800-DNA001-NA12753				Sample 2 G50002800-DNA002-NA12707				Sample 3 G50002800-DNA003-NA11			
Index	Name	Address	Chr	Score	GType	Theta	R	Score	GType	Theta	R	Score	GType	Theta	
3	rs2840531	21	1	0.7827	BB	1.0000	1.4122	0.7846	BB	0.9988	1.7071	0.7846	AB	0.8430	
4	rs649593	23	1	0.2586	BB	0.9973	1.2602	0.6059	AA	0.0059	1.6464	0.6059	BB	0.9971	
5	rs1517342	27	2	0.7639	BB	0.9982	1.0482	0.7639	AB	0.7794	0.9870	0.7639	AB	0.7567	
6	rs1517343	28	2	0.8332	AA	0.0225	0.6481	0.8332	AB	0.7118	0.7242	0.8332	AB	0.7494	
7	rs1668071	30	2	0.7650	BB	1.0000	1.2050	0.7650	AB	0.6992	1.5524	0.7650	AB	0.6926	
8	rs761162	31	1	0.8061	BB	0.9983	1.7662	0.8061	AB	0.7677	1.5597	0.8061	AB	0.7912	
9	rs911903	33	1	0.7764	AB	0.7101	1.4167	0.7764	AA	0.0150	1.6473	0.7764	AB	0.6997	
10	rs753646	36	1	0.8085	BB	0.9978	1.4960	0.8085	AA	0.0060	1.6636	0.8085	BB	1.0000	
11	rs2477703	38	1	0.8577	AA	0.0201	1.2076	0.8755	AB	0.5612	1.3688	0.8755	BB	0.9864	
12	rs558912	40	2	0.8350	AA	0.0299	0.8065	0.8493	AB	0.6700	1.3485	0.8493	AB	0.6761	
13	rs357116	41	3	0.5707	BB	0.9911	1.4637	0.5707	AB	0.6619	1.6667	0.5707	AB	0.6495	
14	rs734999	42	1	0.8752	AB	0.3785	1.1030	0.8752	AA	0.0098	1.2481	0.8752	AA	0.0077	
15	rs2377041	50	1	0.7216	AB	0.6910	1.8170	0.7216	BB	0.9955	1.7182	0.7216	BB	0.9973	
16	rs7551616	51	1	0.8574	BB	0.9838	0.3443	0.8574	AB	0.5940	0.2793	0.8574	BB	0.9738	
17	rs715494	54	22	0.3153	AB	0.5578	0.9870	0.8406	AB	0.5715	1.4069	0.8406	AB	0.5860	
18	rs223201	56	1	0.7982	BB	0.9860	1.1782	0.7982	BB	0.9899	1.3355	0.7982	BB	0.9903	
19	rs213006	61	1	0.8631	BB	0.9977	1.0809	0.8631	AA	0.0094	1.7397	0.8631	AA	0.0082	
20	rs520354	62	2	0.8931	BB	0.9783	0.5257	0.8932	BB	0.9865	0.7193	0.7605	AB	0.6087	
21	rs874515	63	1	0.2524	AB	0.4872	0.9771	0.8455	AA	0.0498	1.4519	0.8455	AA	0.0560	
22	rs1000021	66	3	0.8395	AA	0.0007	1.9531	0.8395	AA	0.0000	1.7870	0.8361	AB	0.5892	
23	rs2030162	67	2	0.6620	BB	0.9962	1.1746	0.7650	AB	0.6798	1.5940	0.7650	AA	0.0595	
24	rs1489396	69	1	0.7904	BB	1.0000	1.3985	0.7904	AB	0.8194	0.9569	0.7904	BB	1.0000	
25	rs742230	70	1	0.8054	AB	0.6909	1.4618	0.8054	AB	0.6717	1.3864	0.8054	AA	0.0054	

Figure 109 Full Data Table

The annotation columns of the **Full Data Table** are listed and described in Table 10.

Table 10 Full Data Table Columns

Column	Description	Type	Visible by Default?
Index	The row index of the SNP	integer	Y
Name	The name of the SNP	string	Y
Address	The bead-type identifier	integer	Y
Chr	The chromosome of the SNP	string	Y
Manifest	The name of the manifest to which the SNP belongs	string	N
Position	The chromosomal position of the SNP	integer	N

Table 10 Full Data Table Columns (continued)

Column	Description	Type	Visible by Default?
GenTrain Score	Score for that SNP from the GenTrain clustering algorithm	float	Y
FRAC A	Fraction of the A nucleotide in the top genomic sequence	float	Y
FRAC C	Fraction of the C nucleotide in the top genomic sequence	float	Y
FRAC G	Fraction of the G nucleotide in the top genomic sequence	float	Y
FRAC T	Fraction of the T nucleotide in the top genomic sequence	float	Y

The per-sample subcolumns of the **Full Data Table** are listed and described in Table 11.

Table 11 Full Data Table Per-Sample Subcolumns

Column	Description	Type	Visible by Default?
GType	The genotype of this SNP for the sample.	string	Y
Score	The call score of this SNP for the sample.	float	Y
Theta	The normalized Theta-value of this SNP for the sample.	float	Y
R	The normalized R-value of this SNP for the sample.	float	Y
X Raw	The raw intensity of the A allele.	integer	N
Y Raw	The raw intensity of the B allele.	integer	N
X	The normalized intensity of the A allele.	float	N
Y	The normalized intensity of the B allele.	float	N

Table 11 Full Data Table Per-Sample Subcolumns (continued)

Column	Description	Type	Visible by Default?
B Allele Freq	The B allele theta value of this SNP for the sample, relative to the cluster positions. This value is normalized so that it is zero if theta is less than or equal to the AA cluster's theta mean, 0.5 if it is equal to the AB cluster's theta mean, or 1 if it is equal to or greater than the BB cluster's theta mean. It is linearly interpolated between 0 and 1.	float	N
Log R Ratio	The base-2 log of the normalized R value over the expected R value for the theta value (interpolated from the R-values of the clusters).	float	N
Top Alleles	The genotype of the Illumina-designated top strand.	string	N
Import Calls	Genotype calls for the given sample imported when the Import Allele Calls feature is used.	string	N
Concordance	The numeric correlation of the top allele call for a SNP in the current project with the imported allele call of a SNP from a different project.	integer	N
Orig Call	The genotype call of the SNP and sample at the time the project was originally clustered.	string	N

SNP Table

The **SNP Table** (Figure 110) shows statistics for each SNP.

SNP Table																
Index	Name	Chr	Position	Address	GenTrain Score	Orig Score	Edited	Cluster Sep	ChiTest P100	Het Excess	AA Freq	AB Freq	BB Freq	Call Freq		
1	rs1867749	2	1201109057	3	0.8237	0.8237	0	1.0000	0.7065	-0.0377	0.3860	0.4561	0.1579	0.9500	C	
2	rs1397354	2	215118936	10	0.8928	0.8928	0	0.9033	0.7369	-0.0336	0.1930	0.4737	0.3333	0.9500	C	
3	rs2840531	1	2155821	21	0.7846	0.7846	0	0.7327	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	C	
4	rs649593	1	3763525	23	0.6059	0.6059	0	0.8909	0.0707	0.1808	0.0877	0.5439	0.3684	0.9500	C	
5	rs1517342	2	169218217	27	0.7639	0.7639	0	0.7325	0.5964	0.0530	0.2456	0.5263	0.2281	0.9500	C	
6	rs1517343	2	169218519	28	0.8332	0.8332	0	0.7913	0.5964	0.0530	0.2281	0.5263	0.2456	0.9500	C	
7	rs1868071	2	30219358	30	0.7650	0.7650	0	0.9426	0.0964	0.1662	0.1228	0.5614	0.3158	0.9500	C	
8	rs761162	1	13733634	31	0.8061	0.8061	0	0.9213	0.7040	0.0380	0.3158	0.5088	0.1754	0.9500	C	
9	rs911903	1	46982589	33	0.7764	0.7764	0	0.8160	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	C	
10	rs753646	1	208765509	36	0.8085	0.8085	0	0.8141	0.8631	-0.0172	0.2632	0.4912	0.2456	0.9500	C	
11	rs2477703	1	2295405	38	0.8755	0.8755	0	1.0000	0.6009	0.0523	0.4211	0.4737	0.1053	0.9500	C	
12	rs588912	2	16975766	40	0.8493	0.8493	0	1.0000	0.6328	-0.0478	0.1404	0.4386	0.4211	0.9500	C	
13	rs357116	3	132966852	41	0.5707	0.5707	0	0.6749	0.1111	0.1593	0.2281	0.5789	0.1930	0.9500	C	
14	rs734999	1	2382629	42	0.8752	0.8752	0	0.9128	0.2181	0.1232	0.2281	0.5614	0.2105	0.9500	C	
15	rs2377041	1	2496311	50	0.7216	0.7216	0	0.8877	0.0652	0.1844	0.1053	0.5614	0.3333	0.9500	C	
16	rs7551616	1	0	51	0.8574	0.8574	0	0.4221	0.8500	0.0188	0.2281	0.5088	0.2632	0.9500	C	
17	rs15494	22	28106987	54	0.8406	0.8406	0	0.9181	0.7065	-0.0377	0.1579	0.4561	0.3860	0.9500	C	
18	rs223201	1	18055067	56	0.7982	0.7982	0	1.0000	0.8481	0.0192	0.0526	0.3684	0.5789	0.9500	C	
19	rs213006	1	21148075	61	0.8631	0.8631	0	1.0000	0.4215	0.0804	0.4386	0.4737	0.0877	0.9500	C	
20	rs520354	2	21234148	62	0.8932	0.8932	0	0.7859	0.0998	-0.1646	0.2667	0.4167	0.3167	1.0000	C	
21	rs874515	1	23348434	63	0.8455	0.8455	0	1.0000	0.0347	0.2112	0.5500	0.4333	0.0167	1.0000	C	
22	rs1000021	3	177788791	66	0.8395	0.8395	0	1.0000	0.1653	-0.1388	0.4333	0.4000	0.1667	1.0000	C	
23	rs2030162	2	169284578	67	0.7650	0.7650	0	0.8794	0.9911	0.0011	0.2333	0.5000	0.2667	1.0000	C	
24	rs1489396	1	78824847	69	0.7904	0.7904	0	0.7056	0.6892	-0.0400	0.4000	0.4500	0.1500	1.0000	C	
25	rs42230	1	24602592	70	0.8054	0.8054	0	0.9850	0.2394	-0.1176	0.3500	0.4333	0.2167	1.0000	C	

Figure 110 SNP Table

The **SNP Table** columns are listed and described in Table 12.

Table 12 SNP Table Columns

Column	Description	Type	Visible by Default?
Index	The row index of the SNP	integer	Y
Name	The name of the SNP	string	Y
Chr	The chromosome of the SNP	string	Y
Position	The chromosomal position of the SNP	integer	N
Address	The bead-type identifier for this SNP	integer	Y
GenTrain Score	A measure of the cluster quality for the SNP	float	Y
Orig Score	The original (unedited) GenTrain Score for the SNP	float	Y

Table 12 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
Edited	A flag indicating whether the SNP was edited after initial clustering positions were identified (1=> edited, 0=> unedited)	integer	Y
Cluster Sep	A measure of the cluster separation for the SNP that ranges between 0 and 1	float	Y
ChiTest 100	A normalized Hardy-Weinberg p value calculated using genotype frequency. The value is calculated with 1 degree of freedom and normalized to 100 individuals.	float	Y
Het Excess	A measure of the excess of heterozygotes for the SNP (based on Hardy-Weinberg Equilibrium). 0 indicates no excess of heterozygotes. Negative values indicate a deficiency of heterozygotes.	float	Y
AA Freq	The frequency of AA calls	float	Y
AB Freq	The frequency of AB calls	float	Y
BB Freq	The frequency of BB calls	float	Y
Call Freq	The overall call frequency	float	Y
Minor Freq	The minor allele frequency	float	Y
Aux	A user-set auxiliary value for the SNP	integer	Y
Rep Errors	The number of reproducibility errors for this SNP as allele comparisons between replicates.	integer	Y
P-C Errors	The number of parent-child heritability errors for the SNP compared among parent-child genotypes.	integer	Y
P-P-C Errors	The number of parent-parent-child heritability errors for the SNP compared among parent-parent-child genotypes.	integer	Y
AA T Mean	The theta value of the center of the AA cluster, in normalized polar coordinates	float	Y

Table 12 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
AA T Dev	The standard deviation in theta of the AA cluster, in normalized polar coordinates	float	Y
AB T Mean	The theta value of the center of the AB cluster, in normalized polar coordinates	float	Y
AB T Dev	The standard deviation in theta of the AB cluster, in normalized polar coordinates	float	Y
BB T Mean	The theta value of the center of the BB cluster, in normalized polar coordinates	float	Y
BB T Dev	The standard deviation in theta of the BB cluster, in normalized polar coordinates	float	Y
AA R Mean	The R value of the center of the AA cluster, in normalized polar coordinates	float	Y
AA R Dev	The standard deviation in R of the AA cluster, in normalized polar coordinates	float	Y
AB R Mean	The R value of the center of the AB cluster, in normalized polar coordinates	float	Y
AB R Dev	The standard deviation in R of the AB cluster, in normalized polar coordinates	float	Y
BB R Mean	The R value of the center of the BB cluster, in normalized polar coordinates	float	Y
BB R Dev	The standard deviation in R of the BB cluster, in normalized polar coordinates	float	Y
Manifest	The manifest from which this SNP was loaded	string	N
10% GC	The 10th percentile GenCall score over included samples for the SNP	float	N
50% GC	The 50th percentile GenCall score over included samples for the SNP	float	N
SNP	The nucleotide substitution for the SNP on the Illumina top strand	string	N

Table 12 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
ILMN Strand	Design strand designation	string	N
Customer Strand	Customer strand designation	string	N
Top Genomic Sequence	Sequence on the top strand around the SNP	string	N
Address 2	Bead type unidentified for the second allele (only used for Infinium I)	string	N
Comment	User-specified comment. (Right-click in the column to view the context menu to set this value)	string	N
Norm ID	The normalization ID for the SNP	integer	N
HW Equil	The Hardy-Weinberg Equilibrium score for the SNP	float	N

Paired Samples Table

The **Paired Samples Table** (Figure 111) shows statistics for paired samples.

Paired Sample Table									
Index	Name	SNP	Address	Chr	Position	Theta_Ref.	Theta_Sub.	dTheta_sub-ref	
166	rs101866933	[T C]	2060...	2	133472453	0.0089	0.6783	0.6694	
169	rs101866407	[A G]	3710...	2	213395771	0.0368	0.7628	0.7260	
170	rs10187298	[T C]	7040...	2	38579987	0.0090	0.0090	0.0000	
171	rs10188753	[A G]	3370...	2	208474929	0.9953	0.9928	0.0025	
172	rs10188985	[A G]	1690...	2	85852732	0.9829	0.9849	0.0020	
173	rs10189695	[T C]	4540...	2	115179293	1.0000	1.0000	0.0000	
174	rs10190695	[A G]	3990...	2	73470058	0.0796	0.6260	0.5463	
175	rs10191517	[A G]	540437	2	53715912	0.9911	0.9903	0.0008	
176	rs10192011	[A C]	6450...	2	181636454	0.0442	0.0476	0.0034	
177	rs10192997	[A G]	7100...	2	65836559	0.9958	0.9966	0.0008	
178	rs1019365	[A G]	5050...	2	231805524	0.1541	0.0824	0.0717	
179	rs10194776	[A G]	4210...	2	231805524	0.0140	0.0075	0.0065	
180	rs1019540	[A C]	2970...	16	59235686	0.0260	0.0207	0.0052	
181	rs1019661	[T G]	6020...	4	13864860	0.9899	0.9869	0.0030	
182	rs10202118	[A G]	2760...	2	51861975	0.0331	0.0290	0.0040	
183	rs10202986	[T C]	6450...	2	36317466	0.0113	0.5245	0.5132	
184	rs10203280	[T C]	2260...	2	56464354	0.0102	0.0110	0.0007	
185	rs10204599	[A G]	5910...	2	38263620	0.0114	0.6263	0.6149	
186	rs1020461	[T C]	5560...	12	77073654	0.0140	0.0075	0.0065	
187	rs1020603	[T C]	2940...	2	169207849	0.9948	0.9956	0.0008	
188	rs10208199	[A G]	5870...	2	33461857	0.9910	0.9785	0.0125	
189	rs10210526	[A G]	20554	2	129804947	0.9857	0.7410	0.2447	
190	rs10210670	[T G]	7040...	2	220702807	0.9896	0.6208	0.3688	
191	rs10211034	[T C]	2450...	2	150317745	0.9902	0.6177	0.3725	
192	rs10215031	[A G]	2900...	7	3797265	0.9966	0.7967	0.1999	
193	rs10215692	[T C]	3840...	7	24629272	0.0429	0.0470	0.0041	
194	rs10216611	[T G]	3460...	8	113350968	0.9967	0.9964	0.0003	
195	rs10216950	[A G]	6200...	8	76065158	0.0497	0.0588	0.0092	
196	rs1021702	[A G]	5390...	3	25034231	0.0236	0.6485	0.6249	
197	rs1021720	[T G]	4480...	9	8390610	0.9923	0.9975	0.0052	
198	rs1021879	[T C]	6980...	9	71152636	0.9755	0.4849	0.4906	
199	rs10219445	[T C]	1710...	12	100848543	0.0182	0.0109	0.0073	
200	rs10219721	[T G]	6330...	12	129802651	0.9726	0.9832	0.0106	
201	rs1022307	[T C]	5860...	11	83628884	0.9990	1.0000	0.0010	
202	rs10224537	[A G]	7550...	7	21413645	0.0257	0.5783	0.5525	
203	rs10225212	[T G]	1430...	7	121222626	0.9871	0.9959	0.0088	
204	rs1022549	[T C]	2900...	6	1659676	0.9673	0.6709	0.2964	
205	rs10225614	[T C]	3610...	7	110197378	0.0166	0.0079	0.0086	
206	rs10225674	[A C]	4590...	7	145487396	0.7313	0.5603	0.1710	
207	rs10226237	[T C]	20048	7	80637285	0.7985	0.7006	0.0979	
208	rs10226468	[T C]	1240...	7	42907176	0.9774	0.5184	0.4590	
209	rs10227002	[T G]	360709	7	47500847	0.9915	0.9909	0.0006	
210	rs10230697	[A G]	4230...	7	8407452	0.0009	0.0014	0.0005	

Figure 111 Paired Samples Table

The **Paired Samples Table** columns are listed and described in Table 13.

Table 13 Paired Samples Table Columns

Column	Description	Type	Visible by Default?
Index	The row index of the SNP	integer	Y
Name	The name of the SNP	string	Y

Table 13 Paired Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
SNP	The SNP	string	Y
Address	The bead-type identifier	integer	Y
Chr	The chromosome for the SNP	string	Y
Position	The chromosomal position of the SNP	integer	N

The **Paired Samples Table** also includes per-pair subcolumns, which are populated from the **Reference to Cluster** and **Reference** columns of the Sample Sheet. The pairing number (for example, Paired Sample 1) and sample names appear above the subcolumn list in the **Paired Samples Table**. The subcolumns are described in Table 16.

Table 14 Paired Samples Table Per-Pair Subcolumns

Column	Description	Type	Visible by Default?
Theta Ref.	Value of theta for the reference sample	float	Y
Theta Sub.	Value of theta for the subject sample	float	Y
IdTheta sub-refl	Absolute value of the difference between subject and reference theta values	float	Y
Allele Freq Ref.	Allele frequency of the reference sample	float	Y
Allele Freq Sub.	Allele frequency of the subject sample	float	Y
IdAlleleFreq sub-refl	Absolute value of the difference between subject and reference allele frequency values	float	Y
R Ref.	Value of R for the reference sample	float	Y
R Sub.	Value of R for the subject sample	float	Y

Table 14 Paired Samples Table Per-Pair Subcolumns (continued)

Column	Description	Type	Visible by Default?
Log2 (Rsub/Rref)	Log base 2 of the ratio of subject and reference R values	float	Y
GType Ref.	Genotype of the reference sample	string	Y
GType Sub.	Genotype of the subject sample	string	Y
LOH Score	The probability that there is loss of heterozygosity in a region of interest	float	Y
CN Estimate	Estimate of the actual copy number at an individual locus	float	Y
CN Shift	<p>Statistical confidence level between 0 and 1 indicating whether or not a copy number change has occurred.</p> <ul style="list-style-type: none"> Values of approximately 1 indicate no copy number change. Values of approximately 0 indicate a copy number change. 	float	Y

Samples Table

The **Samples Table** (Figure 112) contains information for each DNA sample loaded into BeadStudio. The **Samples Table** has the same column re-ordering properties as the **SNP Table**.

Samples Table										
Index	Sample ID	Gender	Sentrix Barcode	Sample Section	p05 Grn	p50 Grn	p95 Grn	p05 Red	p50 Red	p95 Red
1	GS0002800-D...	Unknown	SAM1	R001_C001	308	2988	10388	411	6503	13917
2	GS0002800-D...	Unknown	SAM1	R001_C002	317	4205	13970	438	8827	17451
3	GS0002800-D...	Unknown	SAM1	R001_C003	311	4210	13782	432	8937	16963
4	GS0002800-D...	Unknown	SAM1	R001_C004	312	4418	14474	455	9847	19211
5	GS0002800-D...	Unknown	SAM1	R001_C005	322	4268	13827	486	9066	17878
6	GS0002800-D...	Unknown	SAM1	R001_C006	329	4428	14629	474	9631	18809
7	GS0002800-D...	Unknown	SAM1	R001_C007	308	4444	14210	464	9351	18041
8	GS0002800-D...	Unknown	SAM1	R001_C008	292	4078	13379	395	8768	16880
9	GS0002800-D...	Unknown	SAM1	R001_C009	277	3986	13272	402	9010	17468
10	GS0002800-D...	Unknown	SAM1	R001_C010	264	3701	12647	370	7986	15314
11	GS0002800-D...	Unknown	SAM1	R001_C011	268	3678	12248	365	7708	14822
12	GS0002800-D...	Unknown	SAM1	R001_C012	298	3976	13705	438	8800	17340
13	GS0002800-D...	Unknown	SAM1	R002_C001	315	4118	13280	447	8707	17055
14	GS0002800-D...	Unknown	SAM1	R002_C002	298	4348	14051	413	9009	17634
15	GS0002800-D...	Unknown	SAM1	R002_C003	292	4157	13881	397	8843	17636
16	GS0002800-D...	Unknown	SAM1	R002_C004	292	4065	13527	410	8520	17104
17	GS0002800-D...	Unknown	SAM1	R002_C005	309	3750	12730	436	7618	15764
18	GS0002800-D...	Unknown	SAM1	R002_C006	298	4111	13494	406	8498	16854
19	GS0002800-D...	Unknown	SAM1	R002_C007	298	4278	13750	414	9268	17691
20	GS0002800-D...	Unknown	SAM1	R002_C008	288	4163	13469	410	8613	17310
21	GS0002800-D...	Unknown	SAM1	R002_C009	276	3899	13081	400	8612	16871
22	GS0002800-D...	Unknown	SAM1	R002_C010	274	3678	12401	386	8147	15776
23	GS0002800-D...	Unknown	SAM1	R002_C011	279	4142	13456	401	8649	16890
24	GS0002800-D...	Unknown	SAM1	R002_C012	290	4162	13752	443	9386	18023
25	GS0002800-D...	Unknown	SAM1	R003_C001	314	4189	13676	438	9066	17611

Figure 112 Samples Table

Table 15 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
Gender	The user-specified gender for the sample	string	Y
p05 Grn	The 5th percentile of the A-allele intensity	integer	Y
p50 Grn	The 50th percentile of the A-allele intensity	integer	Y
p95 Grn	The 95th percentile of the A-allele intensity	integer	Y
p05 Red	The 5th percentile of the B-allele intensity	integer	Y

Table 15 Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
p50 Red	The 50th percentile of the B-allele intensity	integer	Y
p95 Red	The 95th percentile of the B-allele intensity	integer	Y
p10 GC	The 10th percentile GenCall score over all SNPs for this sample. If displayed as 0.000 , this column needs to be manually recalculated.	float	Y
p50 GC	The 50th percentile GenCall score over all SNPs for this sample. If displayed as 0.000 , this column needs to be manually recalculated.	float	Y
Rep Error Rate	The reproducibility error rate for this sample, calculated as $1 - \sqrt{1 - \text{errors}/\text{max_possible_errors}}$. Errors and max_possible_errors do not include genotype calls that fall below the no-call threshold. If displayed as 0.000 , this column needs to be manually recalculated.	float	Y
PC Error Rate	The parent-child heritability error rate for the sample. If displayed as 0.000 , this column needs to be manually recalculated.	float	Y
PPC Error Rate	The parent-parent-child heritability error rate for the sample. If displayed as 0.000 , this column needs to be manually recalculated.	float	Y
Call Rate	Percentage of SNPs whose GenCall score is greater than the specified threshold.	integer	N
Aux	Arbitrary integer you can use to differentiate and/or sort samples. Use the context menu to set this value by right-clicking anywhere in the Samples Table .	integer	N
Genotype	Genotype for this sample for the SNP currently selected in the SNP Table .	integer	N
Score	GenCall score for this sample for the SNP currently selected in the SNP Table .	integer	N

Table 15 Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
Sample Name	The sample name	string	N
Sample Group	The sample group	string	N
Sample Plate	The sample plate	string	N
Sample Well	The well within the sample plate	string	N
Gender Est	The estimated gender of the individual from which the sample was acquired	string	N
Requeue Status	Displays a note ("Needs Requeue") if the sample is marked to be requeued, otherwise this column is blank.	string	N
Concordance	Concordance across all SNPs for this sample	float	N
Ethnicity	The ethnicity of the individual from which this sample was acquired	string	N
Age	The age of the individual from which this sample was acquired	integer	N
Weight	The weight in kg of the individual from which this sample was acquired	string	N
Height	The height in meters of the individual from which this sample was acquired	string	N
Blood Pressure Systolic	The systolic blood pressure of the individual from which this sample was acquired	integer	N
Blood Pressure Diastolic	The diastolic blood pressure of the individual from which this sample was acquired	integer	N
Blood Type	The blood type of the individual from which this sample was acquired	string	N
Phenotype Pos 1	The positive phenotype 1 of the individual from which this sample was acquired	string	N

Table 15 Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
Phenotype Pos 2	The positive phenotype 2 of the individual from which this sample was acquired	string	N
Phenotype Pos 3	The positive phenotype 3 of the individual from which this sample was acquired	string	N
Phenotype Neg 1	The negative phenotype 1 of the individual from which this sample was acquired	string	N
Phenotype Neg 2	The negative phenotype 2 of the individual from which this sample was acquired	string	N
Phenotype Neg 3	The negative phenotype 3 of the individual from which this sample was acquired	string	N
Comment	A user-defined field in which you can record custom comments. This field maintains a list of all previously-entered comments. You can access comments from the context menu by right-clicking from within the column.	string	N

The samples table also includes per-manifest subcolumns. The manifest name (for example, HumanHap300) appears above the subcolumn list in the **Samples Table**. The subcolumns are described in Table 16.

Table 16 Samples Table Per-Manifest Subcolumns

Column	Description	Type	Visible by Default?
Sentrix ID	The barcode number of the Sentrix Array Product to which this sample was hybridized	string	Y
Sentrix Position	The section/bundle on the Sentrix Array product	string	Y
Imaging Date	The date on which the Sentrix Array product was scanned.	string	N

Table 16 Samples Table Per-Manifest Subcolumns (continued)

Column	Description	Type	Visible by Default?
Scanner ID	The ID of the scanner on which the Sentrix Array product was scanned	string	N
PMT Green	The green PMT setting of the scanner on which the Sentrix Array product was scanned	integer	N
PMT Red	The red PMT setting of the scanner on which the Sentrix Array product was scanned	integer	N
Software Version	The version of the BeadScan software used to scan the Sentrix Array product	string	N
User	The username of the person logged into the PC on which the Sentrix Array product was scanned	string	N
p05 Grn	The 5th percentile of the A-allele intensity	integer	N
p50 Grn	The 50th percentile of the A-allele intensity	integer	N
p95 Grn	The 95th percentile of the A-allele intensity	integer	N
p05 Red	The 5th percentile of the B-allele intensity	integer	N
p50 Red	The 50th percentile of the B-allele intensity	integer	N
p95 Red	The 95th percentile of the B-allele intensity	integer	N
p10 GC	The 10th percentile GenCall score over all SNPs for this sample. If displayed as <i>0.000</i> , this column needs to be manually recalculated.	float	N
p50 GC	The 50th percentile GenCall score over all SNPs for this sample. If displayed as <i>0.000</i> , this column needs to be manually recalculated.	float	N
Call Rate	Percentage of SNPs whose GenCall score is greater than the specified threshold.	float	N

Context Menu LIMS Options

The following LIMS options are available in the **Samples Table** context menu if you are logged into LIMS:

- ▶ LIMS Actions
 - Update Project From LIMS
 - Send Requeue to LIMS
 - Set to Needs Requeue
 - Clear Needs Requeue
- ▶ Export Cluster Positions to LIMS
- ▶ Update Project from LIMS

For more information about the LIMS options available from the **Samples Table** context menu, see *Context Menus* on page 156 of this manual.

Project Window

The **Project** window (Figure 113) 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 by clicking the + to its left. Double-clicking a sample brings up the **Image Viewer**, which displays the corresponding array image if the image is available in the same directory as the intensity files.

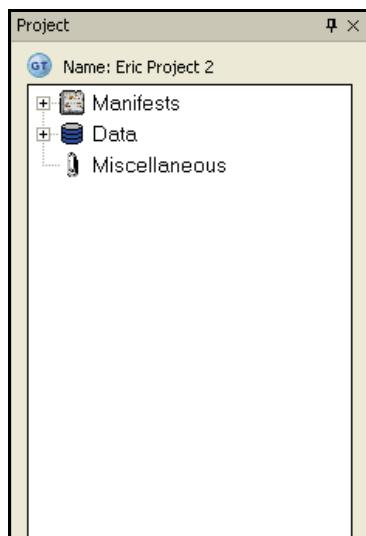


Figure 113 Project Window

Log Window

The **Log** window (Figure 114) is a simple console providing feedback on BeadStudio processes. The **Log** window displays 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="checkbox"/> 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 114 Log Window

Table 17 Log Window Options

Option	Function	Toolbar Button (if used)
Select All	Selects all log entries	

Table 17 Log Window Options (continued)

Option	Function	Toolbar Button (if used)
Copy	Copies log entries to the clipboard	
Save	Saves all log entries	
Clear	Clears all log entries	
Grid	Toggles the grid on and off	
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

The following tables list the selection available from the BeadStudio Genotyping Module's main window menus (and corresponding toolbar buttons).

Table 18 describes **File** Menu functions.

Table 18 File Menu Functions

Selection	Function	Toolbar Button (if used)
New Project	Opens a new project	

Table 18 File Menu Functions (continued)

Selection	Function	Toolbar Button (if used)
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	Displays the Save Project Copy As dialog box, in which you can specify a file name and location to save a copy of the current project that does not include currently-excluded samples.	
Close Project	Closes the current project and returns to the start screen of the Genotyping Module.	
Load Additional Samples	Opens the BeadStudio Project Wizard to the Loading Sample Intensities page, which allows you to use a sample sheet to load sample intensities, or load sample intensities by selecting directories with intensity files.	
Import Cluster Positions	Opens to the last directory used to load clusters, so that you can choose a data file from which to import cluster positions.	
Export Cluster Positions	Allows you to export cluster position data to an *.egt file using the following options: <ul style="list-style-type: none"> • For selected SNPs—allows you to export cluster position data for selected SNPs only. • For all SNPs—allows you to export cluster position data for all SNPs. 	
Export Cluster Position to LIMS	Displays a list from which you can choose to export cluster positions data to LIMS.	
Export Manifest	Allows you to export a manifest as a *.csv file.	
Update Project from LIMS	Allows you to update the project from LIMS.	

Table 18 File Menu Functions (continued)

Selection	Function	Toolbar Button (if used)
Import Phenotype Information from File	Allows you to import phenotype information for your samples from a file.	
Page Setup	Opens the Windows Page Setup dialog box, which you can use to set up the page properties and configure the printer properties	
Print Preview	Opens the Print Preview window, from which you can preview how the selected graph will print	
Print	Displays the Print dialog box. Use this dialog box to select print options for the currently displayed graph	
Recent Project	Allows you to select a project you have recently worked on	
Exit	Closes BeadStudio	

Table 19 describes **Edit** Menu functions.

Table 19 Edit Menu Functions

Selection	Function	Toolbar Button (if used)
Cut	Cuts the current selection	
Copy	Copies the current selection to the clipboard	
Paste	Pastes the current selection from the clipboard	
Select All	Selects all rows and visible columns in the current table	

Table 20 describes View Menu functions.

Table 20 View Menu Functions

Selection	Function	Toolbar Button (if used)
Save Current View	Allows you to save the window configuration of the open project	
Restore Default View	Restores the default window configuration	
Save Custom View	Allows you to save a custom window configuration	
Load Custom View	Allows you to load a previously-saved window configuration	
Log	Shows or hides the Log window	
Project	Shows or hides the Project window	

Table 21 describes **Analysis** Menu functions.

Table 21 Analysis Menu Functions

Selection	Function	Toolbar Button (if used)
Auto Exclude Samples	Automatically evaluates each sample and determines its suitability for inclusion based on overall intensity. Excludes under-performing samples.	
Exclude Samples by Best Run	Samples that have been processed more than once appear in the Samples table multiple times. These samples can be identified by their matching Sample IDs. Using Exclude Samples by Best Run , only the sample with the highest GC10 or GC50 score for each particular sample ID will be included. The other samples with that sample ID will be excluded.	
Cluster All SNPs	Initiates clustering or reclustering and determines the resulting genotype score for each locus	

Table 21 Analysis Menu Functions (continued)

Selection	Function	Toolbar Button (if used)
Update SNP statistics	Updates SNP statistics	
Edit Replicates	Allows you to edit, include, or exclude replicates for a sample	
Edit Parental Relationships	Allows you to edit, include, or exclude P-C and P-P-C relationships for a sample	
Update Heritability/ Reproducibility Errors	Updates replicate, P-C, and P-P-C heritability information in various columns and reports.	
Reports	Allows you to create any of the following: ▶ Reproducibility and Heritability Report ▶ Final Report ▶ DNA Report ▶ Locus Summary Report ▶ Locus x DNA Report	
View Controls Dashboard	Displays the controls dashboard.	
View Contamination Dashboard	Displays the contamination controls dashboard for GoldenGate data.	
Paired Sample Editor	Displays the Paired Sample Editor dialog box, from which you can edit the list of paired samples.	
Calculate Paired Sample LOH/CN	Calculates LOH and copy number-related scores for paired samples.	
Show Genome Viewer	Displays the Illumina Genome Viewer	
Import Allele Calls	Displays the Import Allele Calls dialog box, which allows you to select a directory from which to import allele calls	
Export Allele Calls	Displays the Export Allele Calls dialog box, which allows you to select a directory to which you want to export allele calls	
Remove Imported Allele Calls	Removes imported allele calls from the project.	

Table 21 Analysis Menu Functions (continued)

Selection	Function	Toolbar Button (if used)
Create Plugin Column	Displays the Select Column Plug-In Form dialog box, from which you can select an algorithm-based column plug-in. You can use the column plug-in to create a new subcolumn.	

Table 22 describes **Tools** Menu functions.

Table 22 Tools Menu Functions

Selection	Function	Toolbar Button (if used)
Options Project	Displays the Project Properties window in which you can make changes to project settings.	
Options BeadStudio	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.	
Options Module	Allows you to select storage and memory options.	

Table 23 describes **Windows** Menu functions.

Table 23 Windows Menu Functions

Selection	Function	Toolbar Button (if used)
The Window menu is populated with a list of available windows to display. Windows marked with a check mark are currently displayed.		

Table 24 describes **Help** Menu functions.

Table 24 Help Menu Functions

Selection	Function	Toolbar Button (if used)
About BeadStudio	Brings up the About box for your currently-installed BeadStudio modules, which contains version information and the Software Copyright Notice.	

Graph Window Toolbar

Table 25 lists BeadStudio's Genotyping Module graph window toolbar buttons and their functions.

Table 25 Graph Window Toolbar Buttons & Functions

Toolbar Button	Function(s)
	Polar coordinates —Displays locus using polar coordinates.
	Cartesian coordinates —Displays locus using Cartesian coordinates.
	Plot normalized values —Allows you to toggle normalization on or off in the SNP Graph .
	Make dots larger —Makes each dot representing an individual locus appear larger on the screen.
	Make dots smaller —Makes each dot representing an individual locus appear smaller on the screen.
	Copy plot to clipboard —Copies the current plot to the clipboard.

Table 25 Graph Window Toolbar Buttons & Functions (continued)

Toolbar Button	Function(s)
	Shade call regions —Applies colored shading to each cluster. <ul style="list-style-type: none"> ▶ Loci falling within the dark shaded region of each color are considered to be within the call range (above the GenCall Score threshold). ▶ Loci displayed within the light shaded region of each color are considered to be outside of the call range.
	Default mode —Toggle this button on to activate an arrow cursor that allows you to select samples in the graph window with a rectangle.
	Pan mode —Toggle this button on, then drag the graph in the direction you want.
	Lasso mode —Toggle this button on to draw a lasso to select samples in the graph window.
	Zoom mode —Toggle this button on to zoom in or out in the graph window. When toggled on, the cursor changes to a +, allowing you to zoom in to the graph. Pressing the Ctrl key on your keyboard while in this mode allows you to zoom out.
	Automatically scale X-axis —Automatically scales the X-axis (for the currently displayed graph only).
	Automatically scale Y-axis —Automatically scales the Y-axis (for the currently displayed graph only).

Table Windows Toolbar

Table 26 lists and describes BeadStudio's Genotyping Module Table Windows toolbar buttons and their functions.

Table 26 Table Windows Toolbar Buttons & Functions

Toolbar Button	Function(s)
	Calculate —(Samples Table only) Calculates all samples. This button only appears if there are samples that need to be calculated.
	Select all Rows —Highlights all the rows in the table.
	Copy to Clipboard —Copies the selected columns or rows to the clipboard.
	Export to File —Exports the selected item(s) to a file.
	Import Columns —Imports sample data from a file you specify.
	Sort Column (Ascending) —Sorts columns in the sample table in ascending order.
	Sort Column (Descending) —Sorts columns in the sample table in descending order.
	Sort by Column(s) —Allows you to sort the sample table data by a column or columns you select.
	Line Plot —Displays the Plot Columns window with line graph selected.
	Scatter Plot —Displays the Plot Columns window with scatter plot selected.
	Histogram —Displays the Plot Columns window with Histogram selected.
	Heat Map (Full Data Table Only) —Displays the Plot Sample Sub-Groups in a Heat Map window.

Table 26 Table Windows Toolbar Buttons & Functions (continued)

Toolbar Button	Function(s)
	New subcolumn - Allows you to create new subcolumn.
	Column Chooser - Displays the Column Chooser window.
	Filter Rows - Displays the Filter Table Rows window.
	Clear Filter - Removes the filter.

Context Menus

The tables in this section describe context menu selections for the BeadStudio Genotyping Module.

Table 27 describes graph window context menu selections.

Table 27 Graph Window Context Menu

Selection	Description
Define AA cluster using selected SNP	Uses the selected sample(s) to determine the size and position of the AA genotype cluster.
Define AB cluster using selected SNP	Uses the selected sample(s) to determine the size and position of the AB genotype cluster.
Define BB cluster using selected SNP	Uses the selected sample(s) to determine the size and position of the BB genotype cluster.
Cluster this SNP	Determines cluster locations and score for each locus.
Cluster this SNP Excluding Selected Samples	Determines the cluster locations for each locus except those you have excluded.

Table 27 Graph Window Context Menu (continued)

Selection	Description
Configure Mark.	Marks selected samples in a color you choose.
Mark Selected Points - <Add New>	Allows you to create a new mark.
Clear Marks - <All>	Clears all marks.
Exclude Selected Samples	Excludes selected samples from the genoplot.
Include Selected Samples	Includes selected samples in the genoplot.
Show Legend	Displays the genoplot marks legend.
Show Excluded Samples	Shows excluded samples.
Auto Scale Axes	Automatically scales the axes.
Properties	Launches the Graph Control Settings dialog box.

Table 28 describes Full Data Table context menu selections.

Table 28 Full Data Table Context Menu

Selection	Description
Show Only Selected Rows	Shows only selected rows in the Full Data Table.
Configure Marks	Configures 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 29 describes SNP Table context menu selections.

Table 29 SNP Table Context Menu

Selection	Description
Cluster Selected SNP	Clusters a selected SNP.
Zero Selected SNP	Zeroes a selected SNP.
Set Aux Value	Sets the aux value of a SNP.
Show Only Selected Rows	Shows only selected rows in the SNP Table.
Configure Marks	Configures 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 30 describes Samples Table context menu selections.

Table 30 Samples Table Context Menu

Selection	Description
Exclude Selected Sample	Excludes the selected sample
Include Selected Sample	Includes the selected sample
Recalculate Statistics for Selected Sample	Recalculates statistics for selected samples
Recalculate Statistics for All Samples	Recalculates statistics for all samples.
Estimate Gender for Selected Samples	Estimates gender for the selected samples.
Display Image	Image will be displayed only if you have access to the *.idat file, the *.locs (locus) file, the *.xml file, and either the *.jpg or *.tif image file for the sample or sample section.
Set Aux Value	Sets the aux value of a sample.

Table 30 Samples Table Context Menu (continued)

Selection	Description
Sample Properties	Opens the Sample Properties dialog box, from which you can change values for sample data, such as sample group, sample name, gender, and phenotype properties, or change the path to associated image files.
Upload Selected Samples to Illumina Controls Database	Allows you to upload selected samples to the Illumina Controls Database.
LIMS Actions - Contains a subset of actions related to LIMS. The LIMS Actions menu option and its related suboptions are only available if you are logged into LIMS.	Update Project from LIMS —Updates the current project with the most recent information available in the LIMS database.
	Send Requeue to LIMS —Sends information about a requeued sample to the LIMS database.
	Set to Needs Requeue —Adds a note in the Requeue Status column for a sample that this sample needs to be requeued.
	Clear Requeue —Clears the requeue note in the Requeue Status column for a sample.
Show Only Selected Rows	Shows only selected rows in the Samples Table.
Configure Marks	Configures 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 31 describes **Error Table** context menu selections.

Table 31 Error Table Context Menu

Selection	Description
Show Only Selected Rows	Configures the Samples Table to show only selected rows.
Edit Replicates	Edits replicates.
Edit Parental Relationships	Edits parental relationships.
Configure Marks	Allows you to configure 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 from the table.

Appendix A

Sample Sheet Guidelines

Topics

- 162 Introduction
- 162 Manifests Section
- 163 Data Section
- 164 Redos and Replicates
- 164 Sample Sheet Template
- 165 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 required sections are the Manifests and Data sections. You can also include a Header section, or any other user-defined sections.

Manifests Section

The Manifests section contains two columns. The first column is populated by A, B, C, etc. The second column is populated by the name of the manifest file corresponding to manifest A, B, C, etc.

For example,

[Manifests]

A, GS0006492-OPA
B, GS0006493-OPA
C, GS0006494-OPA
D, GS0006495-OPA

Data Section

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

Table 32 Data Section, Required and Optional Columns

Column	Description	Optional (O) or Required (R)
Sample_ID	Sample identifier (used only for display in the table).	R
Sample_Name	Name of the sample (used only for display in the table).	O
Sample_Plate	The barcode of the sample plate for this sample (used only for display in the table).	O
Sample_Well	The well within the sample plate for this sample (used only for display in the table).	O
SentrixBarcode_A	The barcode of the Sentrix array product that this sample was hybridized to for Manifest A.	R
SentrixPosition_A	The Sentrix position within the Sentrix array product that this sample was hybridized to for Manifest A (and similarly for _B, _C, etc. depending on how many manifests your project has).	R
Gender	Male, Female, or Unknown.	O
Sample_Group	A group, if any, that this sample belongs to (used for exclusion in the Final Report Wizard).	O
Replicates	The Sample_ID of a sample that is a replicate to this sample (used in reproducibility error calculations).	O
Parent1	The Sample_ID of the first parent for this sample.	O
Parent2	The Sample_ID of the second parent for this sample.	O

Table 32 Data Section, Required and Optional Columns (continued)

Column	Description	Optional (O) or Required (R)
Path	Directory where your data are stored.	O
Reference	Used for paired sample analysis. Populate this column with the sample ID of the reference sample.	O
NOTES	<ul style="list-style-type: none"> • Figure 115 is an example sample sheet • Your sample sheet header may contain any, and as much, information as 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. 	

Redos and Replicates

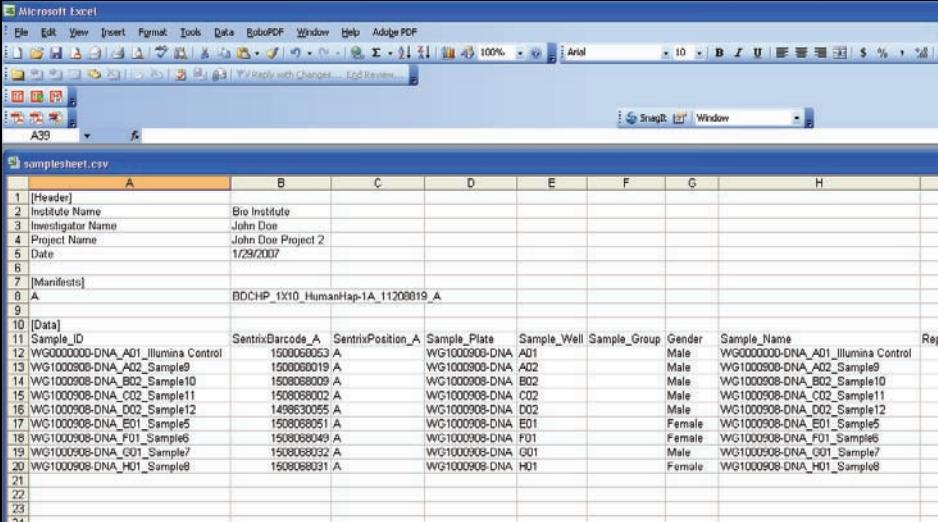
Sample entries with the same Sample_ID are considered "redos" in the BeadStudio Genotyping Module. When you generate the Final Report, you have the option to keep data for the best run of a redo set. If you want to keep data for all redos in the Final Report, it is best to make each Sample_ID unique in the Sample Sheet.

If a Replicate is specified for a Sample_ID occurring more than two times in the Sample Sheet (considered a redo), the BeadStudio Genotyping Module by default forms one replicate pair with the next occurrence of that Sample_ID.

Sample Sheet Template

A template for a sample sheet is provided on your BeadStudio CD. Use this template to create your own user-defined sample sheet.

Sample Sheet Example



The screenshot shows a Microsoft Excel window displaying a CSV file named "samplesheet.csv". The spreadsheet contains three main sections: Header, Manifests, and Data.

	A	B	C	D	E	F	G	H
1	[Header]							
2	Institute Name	Bio Institute						
3	Investigator Name	John Doe						
4	Project Name	John Doe Project 2						
5	Date	1/29/2007						
6								
7	[Manifests]							
8	A	BDCHP_1X10_HumanHap1A_11208019_A						
9								
10	[Data]							
11	Sample_ID	SentrixBarcode_A	SentrixPosition_A	Sample_Plate	Sample_Well	Sample_Group	Gender	Sample_Name
12	WG00000000-DNA_A01_Illumina Control	1500060050	A	WG1000900-DNA	A01		Male	WG00000000-DNA_A01_Illumina Control
13	WG1000908-DNA_A02_Sample9	1500060019	A	WG1000900-DNA	A02		Male	WG1000908-DNA_A02_Sample9
14	WG1000908-DNA_B02_Sample10	1508068009	A	WG1000908-DNA	B02		Male	WG1000908-DNA_B02_Sample10
15	WG1000908-DNA_C02_Sample11	1508068002	A	WG1000908-DNA	C02		Male	WG1000908-DNA_C02_Sample11
16	WG1000908-DNA_D02_Sample12	1498630055	A	WG1000908-DNA	D02		Male	WG1000908-DNA_D02_Sample12
17	WG1000908-DNA_E01_Sample5	1508068051	A	WG1000908-DNA	E01		Female	WG1000908-DNA_E01_Sample5
18	WG1000908-DNA_F01_Sample6	1508068049	A	WG1000908-DNA	F01		Female	WG1000908-DNA_F01_Sample6
19	WG1000908-DNA_G01_Sample7	1508068032	A	WG1000908-DNA	G01		Male	WG1000908-DNA_G01_Sample7
20	WG1000908-DNA_H01_Sample8	1508068031	A	WG1000908-DNA	H01		Female	WG1000908-DNA_H01_Sample8
21								
22								
23								
24								

Figure 115 Sample Sheet Example

Appendix B

Troubleshooting Guide

Topics

- 168 Introduction
- 168 Frequently Asked Questions

Introduction

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

Frequently Asked Questions

Table 33 lists frequently asked questions and associated responses.

Table 33 Frequently Asked Questions

#	Question	Response
1	What is a SNP Manifest?	A SNP Manifest is a file containing the SNP-to-beadtype mapping, as well as all SNP annotations. For the GoldenGate® assay, this is an OPA file in *.opa format. For the Infinium® assay, this is a *.bpm file in binary format. You can always export your manifest information to *.csv format by selecting File Export Manifest .
2	What information does a cluster file contain?	The cluster file contains the mean (R) and standard deviation (theta) of the cluster positions, in normalized coordinates, for every genotype, for every SNP. The cluster file also includes cluster score information, as well as the allele frequencies from the training set used to generate the cluster file.