GenomeStudio™ Genotyping Module v1.0 User Guide

An Integrated Platform for Data Visualization and Analysis

FOR RESEARCH ONLY

ILLUMINA PROPRIETARY
Part # 11318815

illumina®

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 of fitness is implied, nor does Illumina accept any liability for damages resulting from the information contained in this guide.

Revision History

Title	Part Number	Revision	Date
GenomeStudio Genotyping Module v1.0 User Guide	#11317113	Rev. A	November 2008
BeadStudio Genotyping Module v3.2 User Guide	#11284301	Rev. A	December 2007
BeadStudio Genotyping Module User Guide	#11207066	Rev. C	February 2007
BeadStudio Genotyping Module User Guide	#11207066	Rev. B	March 2006
BeadStudio Genotyping Module User Guide	#11207066	Rev. A	December 2005

Table of Contents

List of Figures

List of Tables

Chapter 1
Overview

Topics

Introduction

This user guide describes Illumina's GenomeStudio™ v1.0 Genotyping Module. The GenomeStudio 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 GenomeStudio 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 GenomeStudio Genotyping Module.

For information about the GenomeStudio Framework, the common user interface and functionality available in all GenomeStudio Modules, refer to the *GenomeStudio Framework User Guide*.

Installing the Genotyping Module

To install the GenomeStudio Genotyping Module:

1. Put the GenomeStudio CD into your CD drive.

   If the Illumina GenomeStudio Installation screen appears (Figure 2), continue to Step 3.

   If the CD does not load automatically, double-click the *GenomeStudio<version>.exe* icon in the **GenomeStudio** folder on the CD.

   The GenomeStudio application suite unzips (Figure 1).

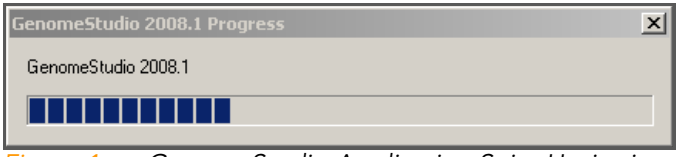


Figure 1 GenomeStudio Application Suite Unzipping

The Illumina GenomeStudio Installation dialog box appears (Figure 2).

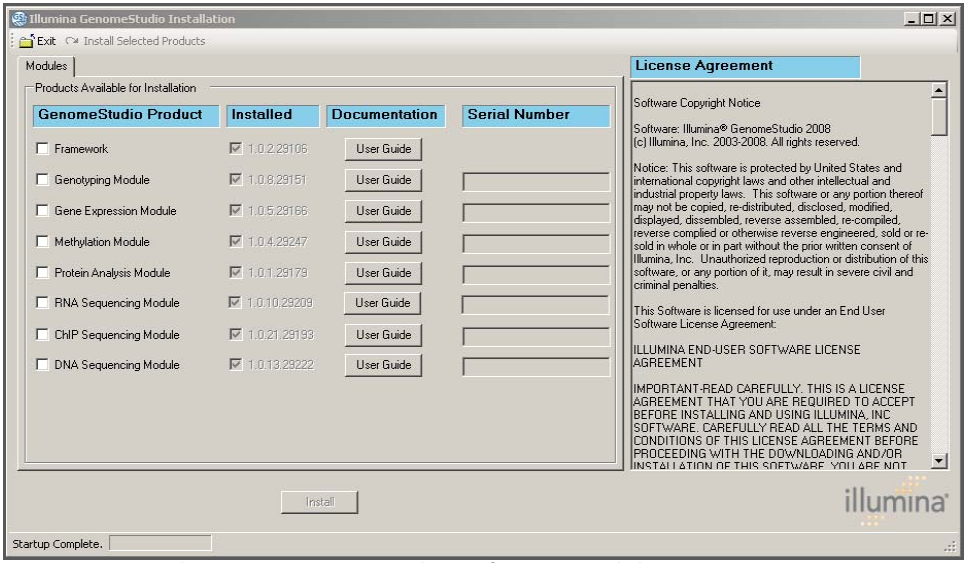


Figure 2 Selecting GenomeStudio Software Modules

2. Read the software license agreement in the right-hand side of the Illumina GenomeStudio Installation dialog box.
3. In the GenomeStudio Product area, select **Genotyping Module**.

NOTE

The GenomeStudio Framework works in conjunction with GenomeStudio software modules. Select the Framework and one or more GenomeStudio modules to install, and have your serial number(s) available.

4. 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 GenomeStudio CD.

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

6. Click **Install**.

   The Software License Agreement dialog box appears (Figure 3).

Figure 3 License Agreement

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

   The GenomeStudio Framework and Genotyping Module are installed on your computer, along with any additional GenomeStudio modules you selected (Figure 4).

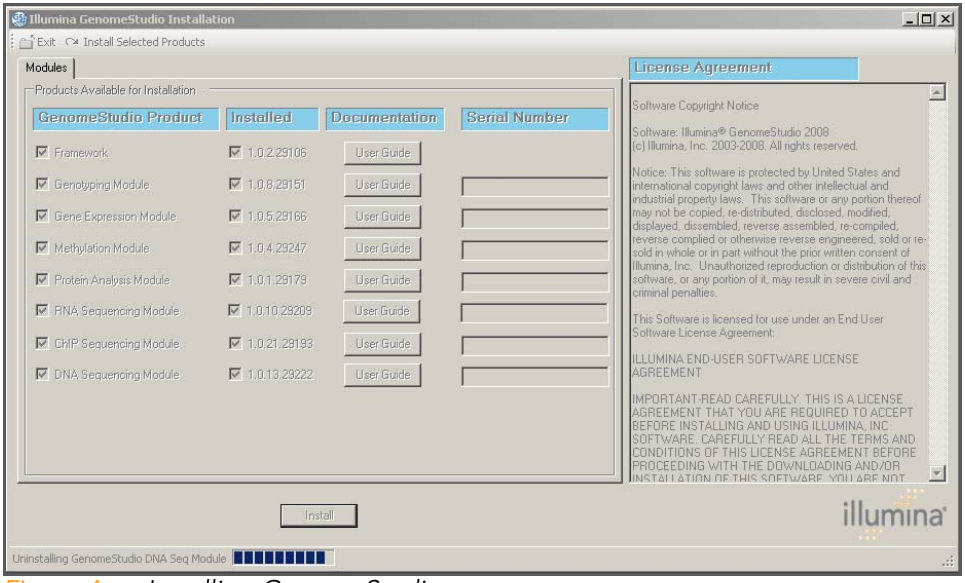


Figure 4 Installing GenomeStudio

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

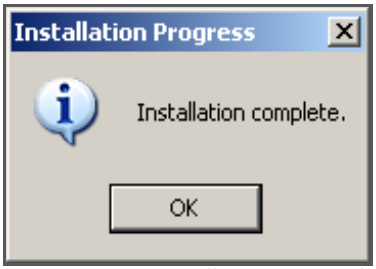


Figure 5 Installation Complete

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

   You can now start a new GenomeStudio project using any GenomeStudio module you have installed.

   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 GenomeStudio Genotyping Module is shown in Figure 6.

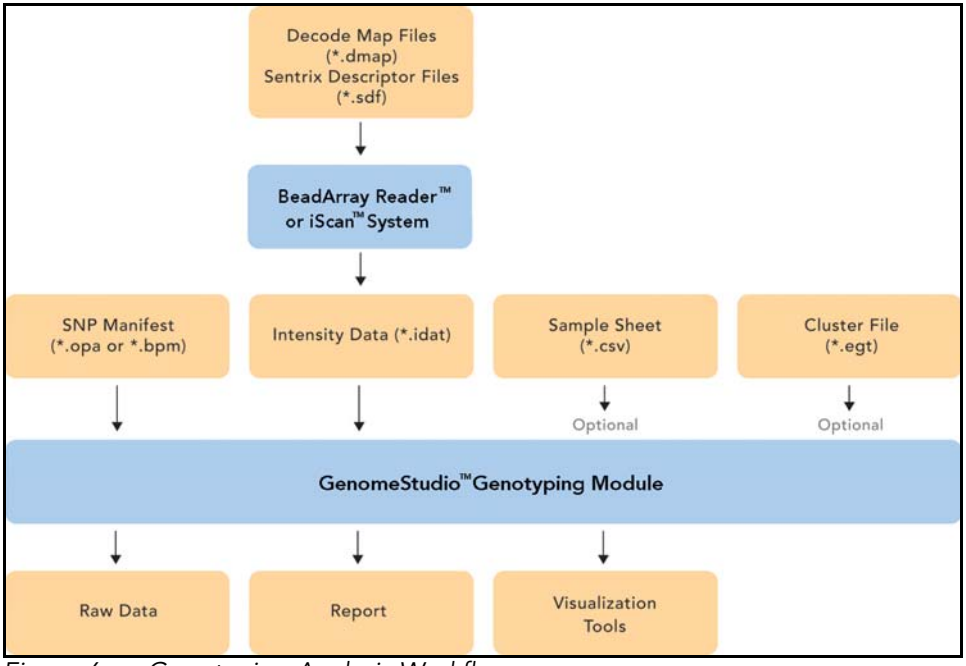


Figure 6 Genotyping Analysis Workflow

Chapter 2
Creating a New Project

Topics

Introduction

The New Project Wizard offers an easy way to start a new project from within any GenomeStudio 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** | **GenomeStudio.**
   - Double-click the GenomeStudio icon 

   on the desktop.

   The GenomeStudio application launches and the **Start** page appears.
2. On the GenomeStudio Start page (Figure 7), do one of the following:
   - In the New Project pane, click **Genotyping**.

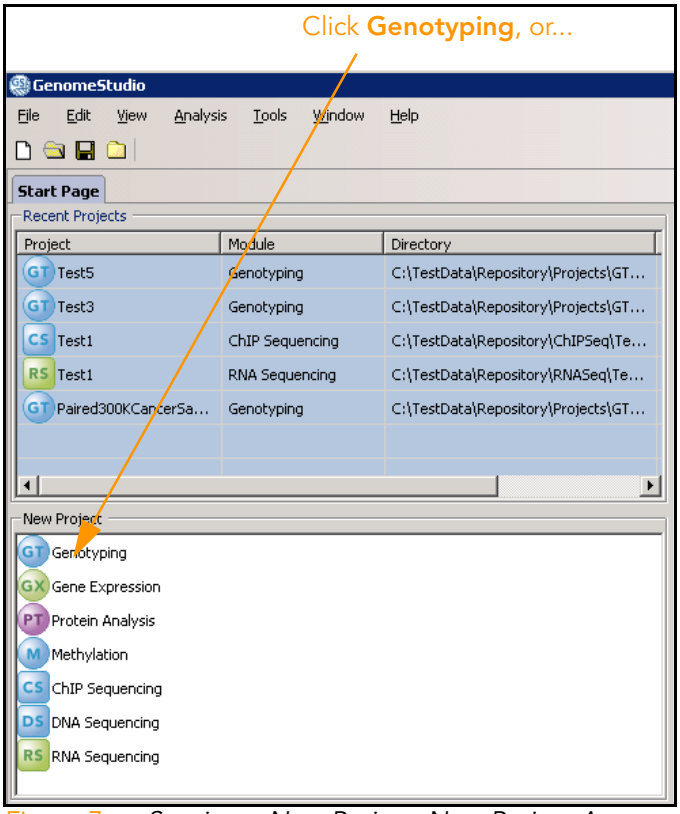


Figure 7 Starting a New Project, New Project Area

- Select **File** | **New Project** | **Genotyping** (Figure 8).

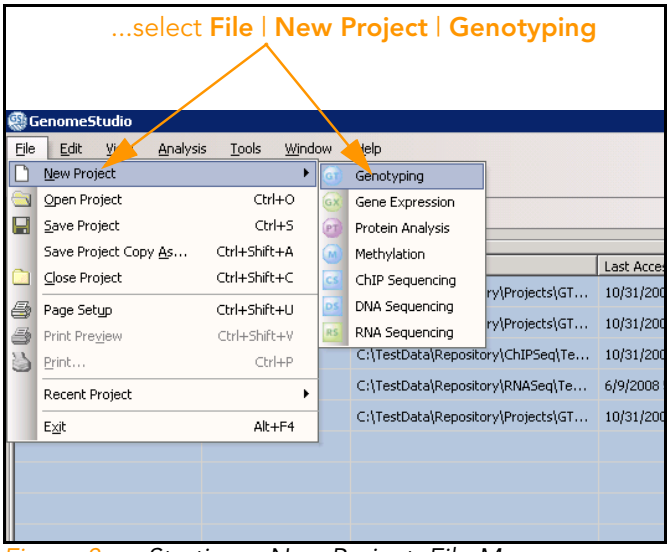


Figure 8 Starting a New Project, File Menu

The GenomeStudio Project Wizard - Welcome dialog appears (Figure 9).

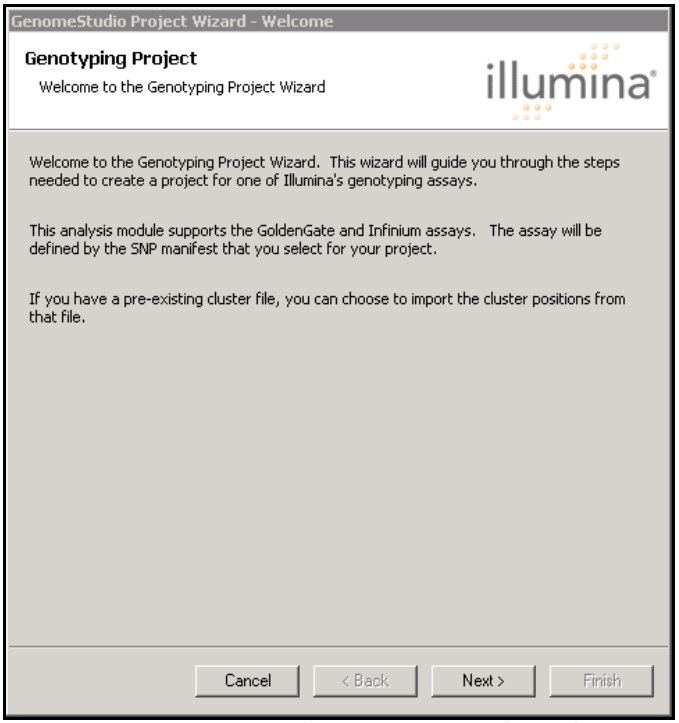


Figure 9 GenomeStudio Project Wizard - Welcome

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

Choosing a Project Name and Location

In the GenomeStudio Project Wizard - Project Location dialog (Figure 10), 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).

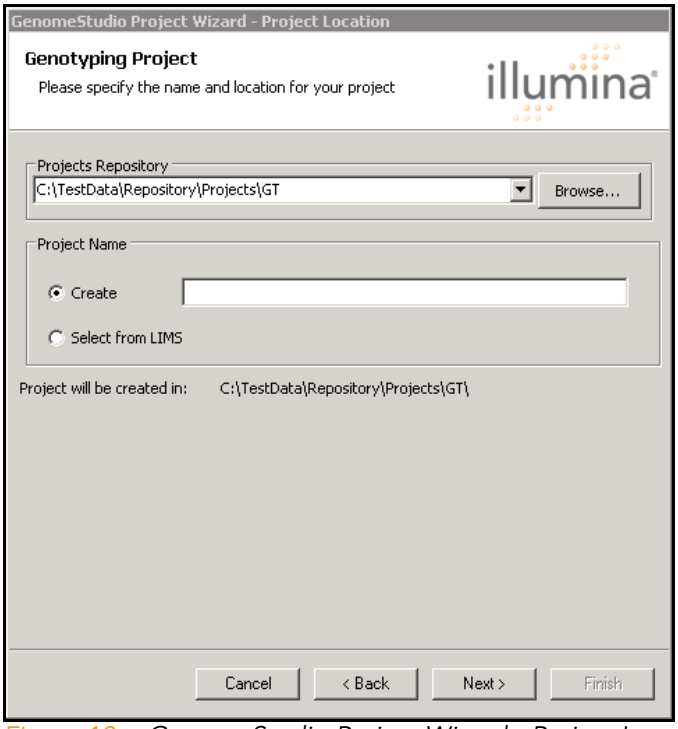


Figure 10 GenomeStudio Project Wizard - Project Location

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.
  - **c.** Continue to *Loading Sample Intensities Outside of LIMS* on page 19.

Selecting a Project From LIMS

To select a project from LIMS:

1. In the GenomeStudio Project Wizard - Project Location dialog (Figure 10), choose **Select from LIMS**.
2. The GenomeStudio Project Wizard - Select LIMS Project dialog appears (Figure 11).

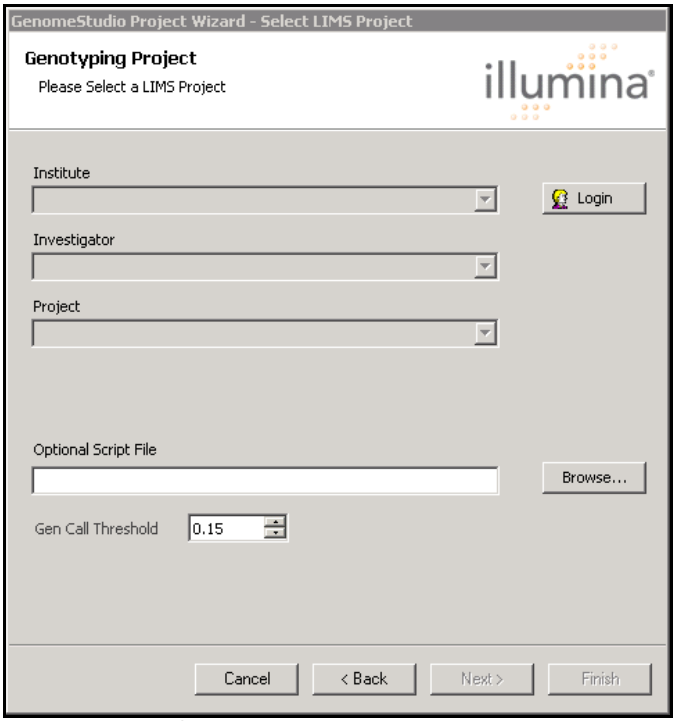


Figure 11 Select LIMS Project

3. Click **Login** to access the Login Infinium LIMS dialog.
4. Select the **Setup** tab (Figure 12).

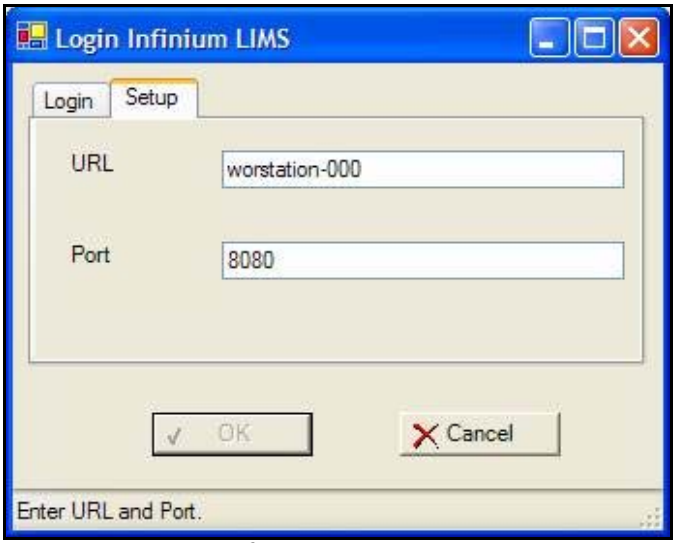


Figure 12 Login Infinium LIMS - Setup

5. In the Setup tab, enter the following:
   - URL
   - Port Number
6. Select the **Login** tab (Figure 13).

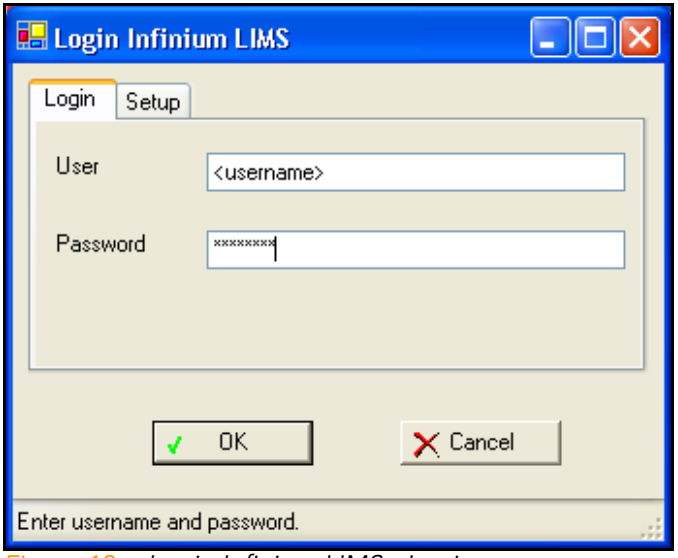


Figure 13 Login Infinium LIMS - Login

7. Enter your username and password.
8. Click **OK**.

   The Login Infinium LIMS dialog closes.

   You are returned to the **Select LIMS Project** dialog (Figure 14).
9. On the Select LIMS Project dialog, make the following selections from the dropdown menus:
   - **Institute**
   - **Investigator**
   - **Project**

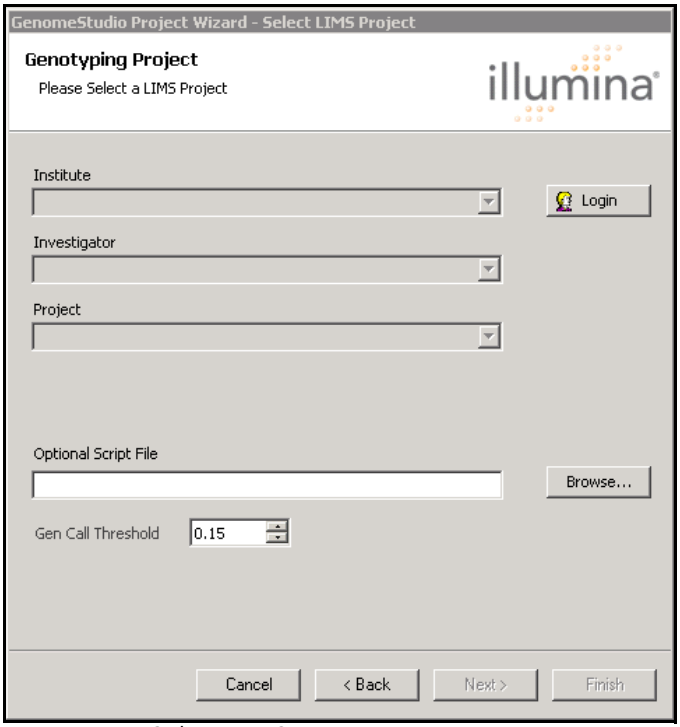


Figure 14 Select LIMS Project

If you have loaded information for a pre-existing project, the warning shown in Figure 15 appears.

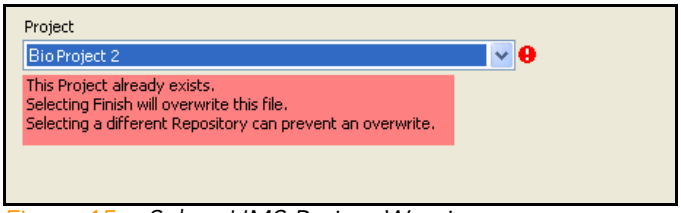


Figure 15 Select LIMS Project Warning

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

10. Click **Finish**.

    The Select Target Dates dialog appears (Figure 16).

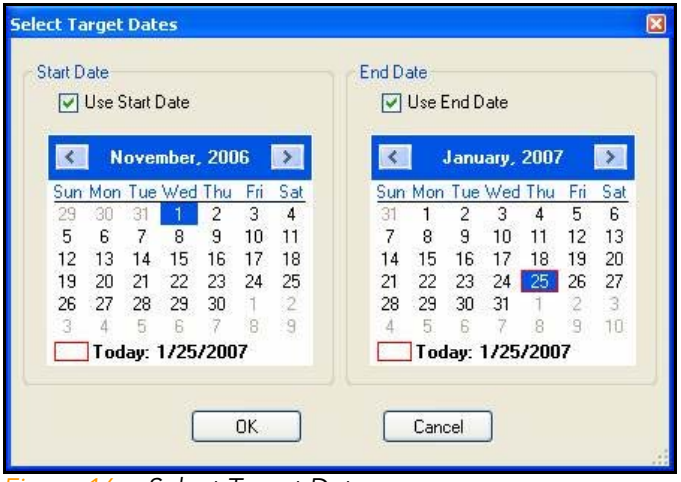


Figure 16 Select Target Dates

11. [Optional] Select **Use Start Date** and choose a start date in the calendar on the left (Figure 17).

12. [Optional] Select **Use End Date** and choose an end date in the calendar on the right (Figure 17).

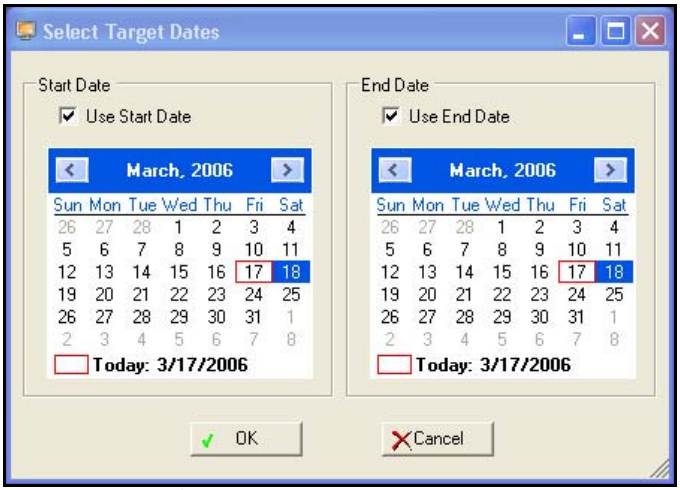


Figure 17 Selecting Target Dates

13. Click **OK**.

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

A heritability and reproducibility errors dialog appears (Figure 18).

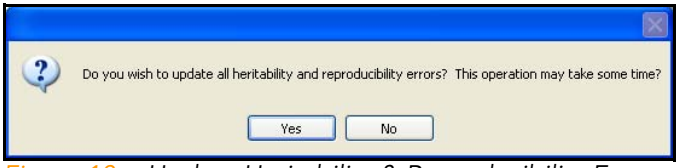


Figure 18 Update Heritability & Reproducibility Errors

If you click **Yes**, the Evaluating Heritability status bar appears (Figure 19) and heritability and reproducibillity are calculated.

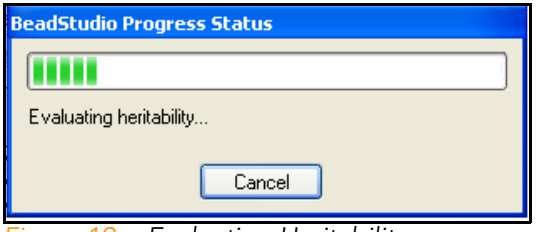


Figure 19 Evaluating Heritability

SNP data are saved, and the Sample Requeue Status Change message appears (Figure 20).

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

If the data from the GenomeStudio project and the LIMS database are the same, the Sample Requeue Status Change dialog displays the message "No updates were required."

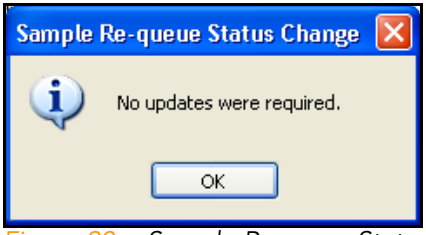


Figure 20 Sample Requeue Status

14. Click **OK**.

The project you selected loads from LIMS and displays in the GenomeStudio 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 19)
- Loading samples by selecting directories that contain intensity data files (page 23).

Using a Sample Sheet

To load intensities using a sample sheet:

1. In the Loading Sample Intensities dialog, select **Use sample sheet to load sample intensities** (Figure 21).

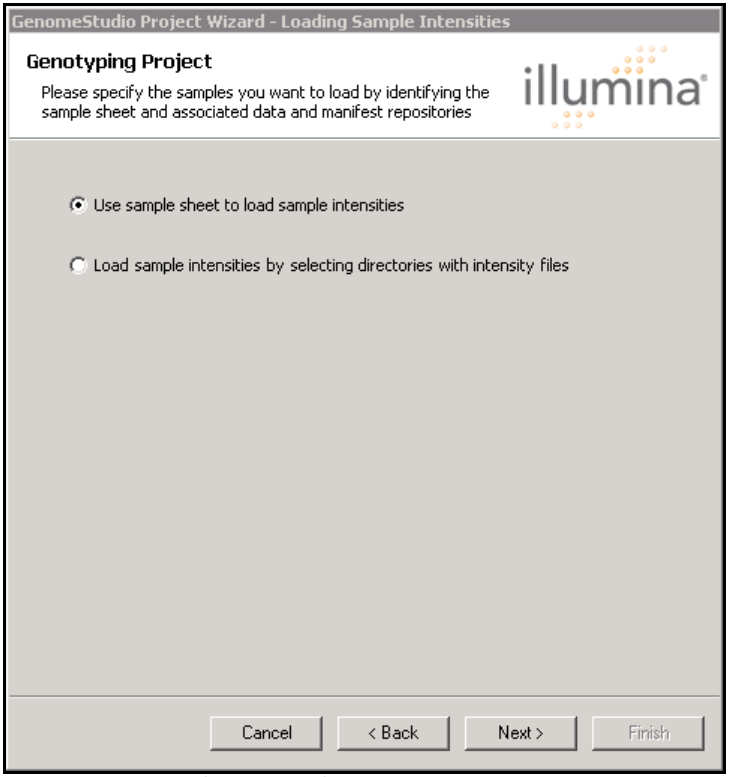


Figure 21 Loading Sample Intensities

2. Click **Next**.

   The Loading Sample Intensities dialog appears (Figure 22).

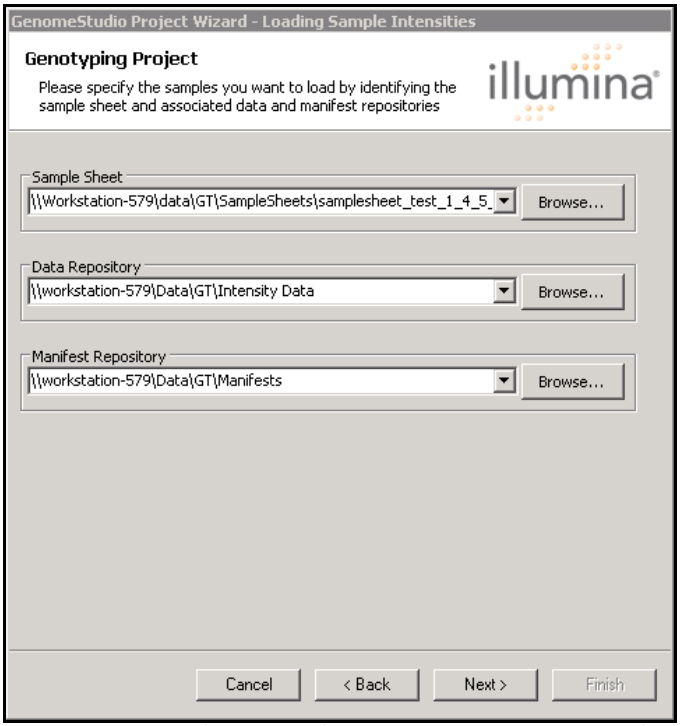


Figure 22 Loading Sample Intensities Using a Sample Sheet

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 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 manifest is contained in the sample sheet, and the GenomeStudio Genotyping Module needs to know where to find it.

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 appears (Figure 23).

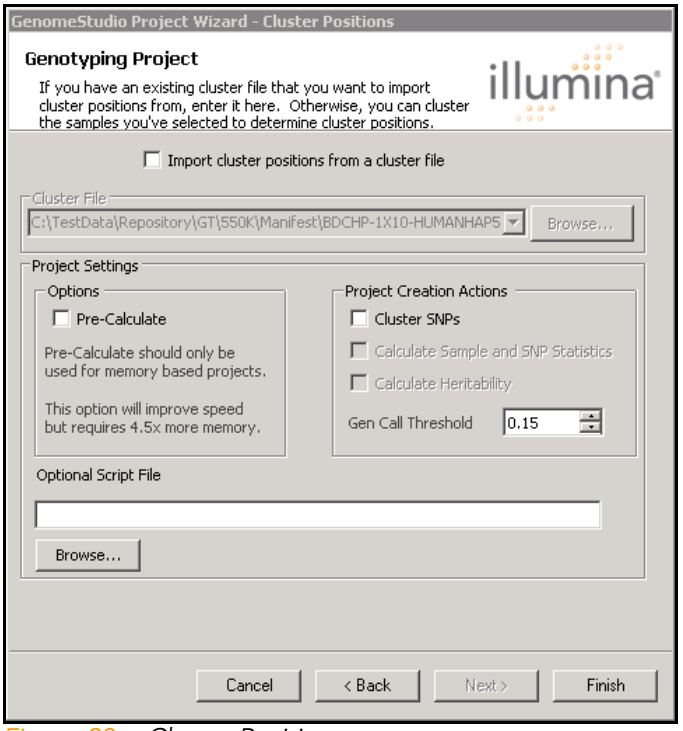


Figure 23 Cluster Positions

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.

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 more of the following options:
   - **Calculate Sample and SNP Statistics**
   - **Calculate Heritability**
   - **Gen Call Threshold**

NOTE

Illumina recommends that you use a GenCall Score cutoff of 0.15 for Infinium products and 0.25 for GoldenGate products.

After loading intensity data using a sample sheet, continue to *Importing Cluster Positions* on page 25.

Selecting Directories

To load intensities by selecting directories:

1. In the Loading Sample Intensities dialog, select **Load Sample Intensities by Selecting Directories with Intensity Files** (Figure 24).

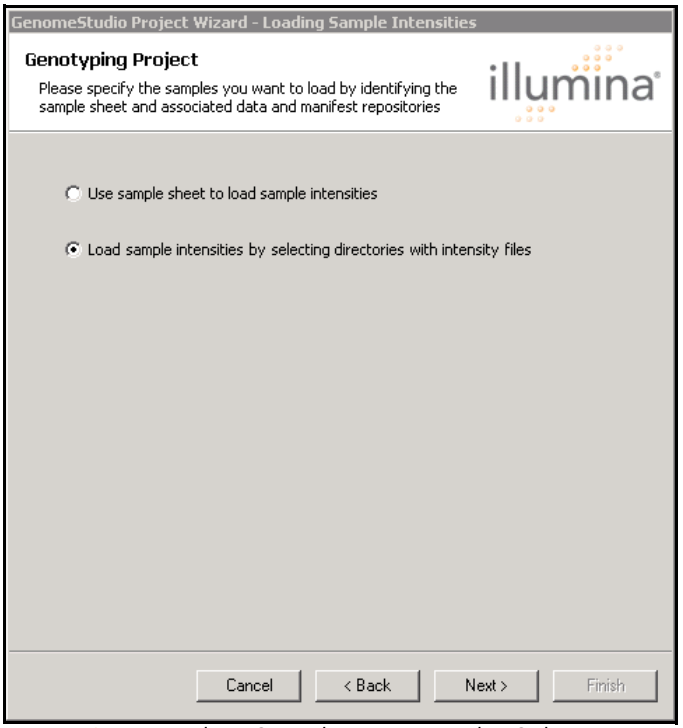


Figure 24 Loading Sample Intensities by Selecting Directories with Intensity Files

2. Click **Next**.

   The Loading Sample Intensities dialog appears (Figure 25).

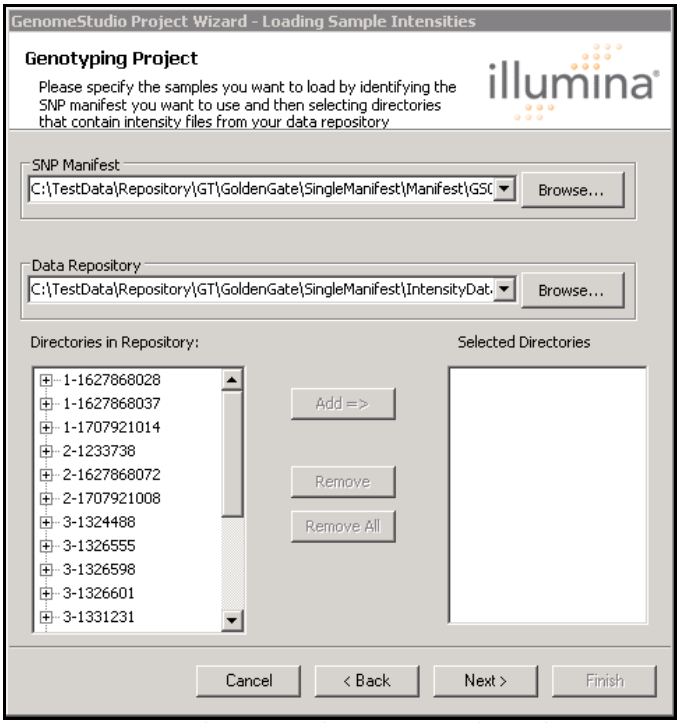


Figure 25 Loading Sample Intensities by Selecting Directories with Intensity Files

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.

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.

Importing Cluster Positions

The Cluster Positions dialog is the final screen of the GenomeStudio Project Wizard (Figure 26). From this screen, you can import a cluster file (*.egt file) and choose to use these cluster definitions to call genotypes for your samples.

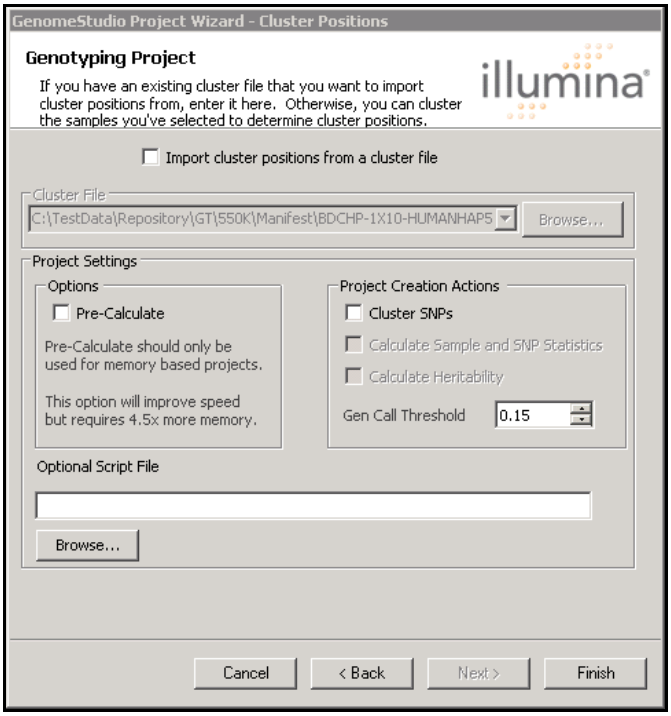


Figure 26 Cluster Positions

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 more of the following options:

- **Calculate Sample and SNP Statistics**
- **Calculate Heritability**
- **Gen Call Threshold**

Illumina recommends that you use a GenCall Score cutoff of 0.15 for Infinium products and 0.25 for GoldenGate products.

5. Click **Finish** to complete the wizard.
   The Genotyping Module loads your intensity files.

If you loaded a cluster file, go to Chapter 3,

If you did not load a cluster file, continue to Chapter 4,

Chapter 3
Viewing Your Data

Topics

Introduction

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

For more information about the various elements of the GenomeStudio user interface, such as windows, tables, and columns, see Chapter 8, *User Interface Reference*.

SNP Graph

The SNP Graph (Figure 27) 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 27).

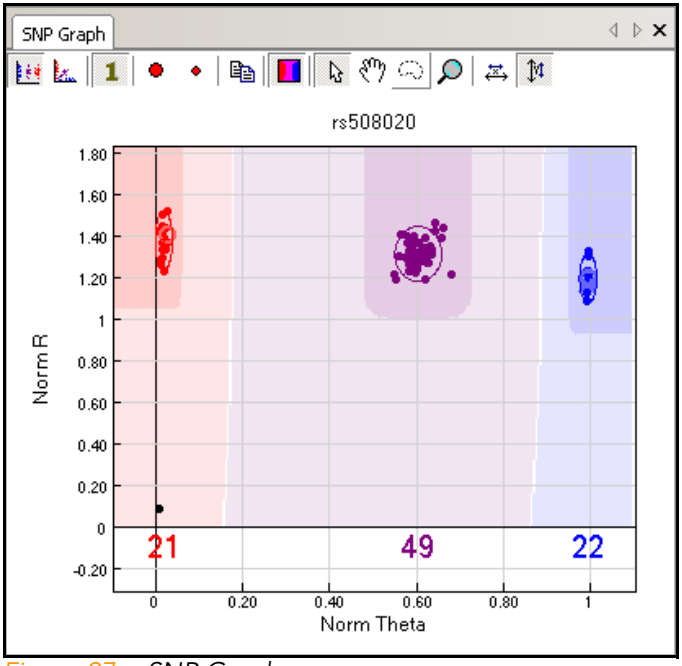


Figure 27 SNP Graph

Shading Call Regions

GenCall Score is a quality metric that indicates the reliability of each genotype call. The GenCall Score is a value between 0 and 1 assigned to every called genotype. Genotypes with lower GenCall scores are located further from the center of a cluster and have a lower reliability.

GenCall Scores are calculated using information from the clustering of the samples. To get a GenCall Score, each SNP is evaluated based on the following characteristics of the clusters:

- angle
- dispersion
- overlap
- intensity

There is no global interpretation of a GenCall Score, as the score depends on the clustering of your samples at each SNP, which is affected by many different variables including the quality of the samples and the loci.

A 50% GenCall Score refers to the 50th percentile GenCall Score in a particular distribution of GenCall Scores.

A 50% GenCall Score for a DNA sample represents the 50th percentile rank for all GenCall Scores for that sample.

Similarly, a 50% GenCall Score for a particular locus represents the 50th percentile rank for all GenCall scores for that locus.

In a genotyping project, samples are displayed in three distinct shaded areas based on their genotype calls. The size of the shaded area is defined by the GenCall Score cutoff.

Select  **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 a lower threshold for valid calls within GenomeStudio, perform the following steps:

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

NOTE

Illumina recommends that you use a GenCall Score cutoff of 0.15 for Infinium products and 0.25 for GoldenGate products.

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 28).

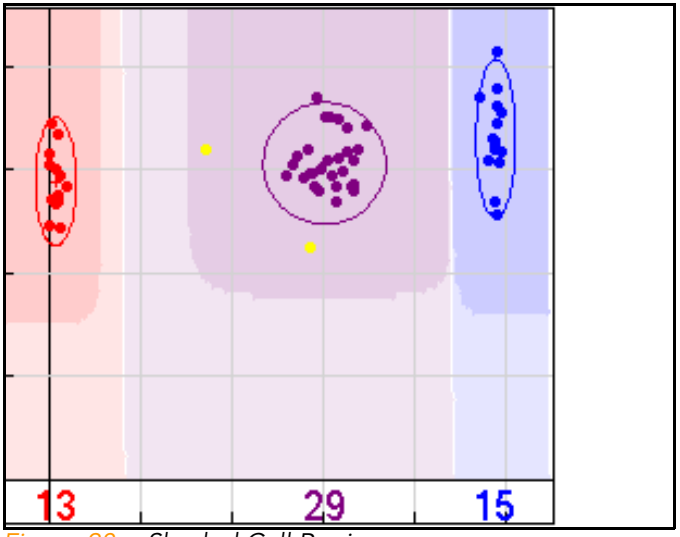


Figure 28 Shaded Call Regions

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.

SNP Graph Error Display

In the SNP Graph, if there are any P-C (parent-child) or P-P-C (parent-parent-child) 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 29).

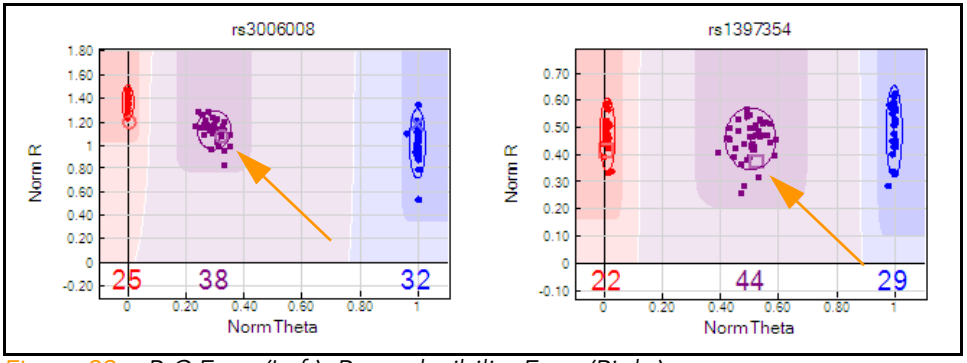


Figure 29 P-C Error (Left), Reproducibility Error (Right)

If you click an error entry in the Errors table, the associated samples are highlighted in yellow in the SNP Graph (Figure 30).

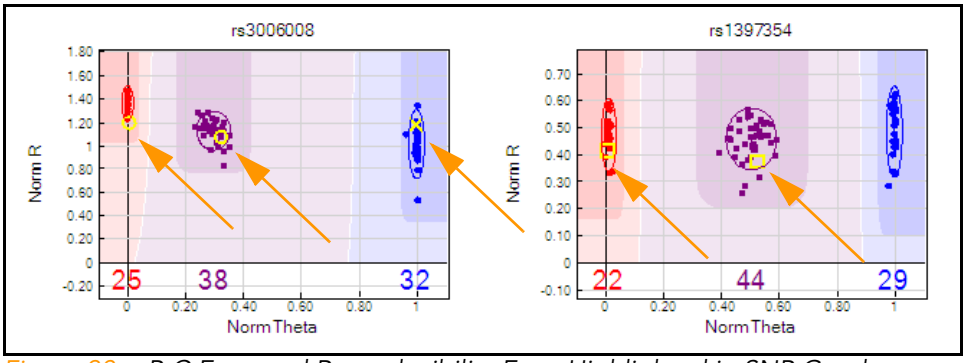


Figure 30 P-C Error and Reproducibility Error Highlighted in SNP Graph

Cartesian and Polar Coordinates

You can view the SNP Graph in either polar or Cartesian coordinates (Figure 31).

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*A+B*B)).

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

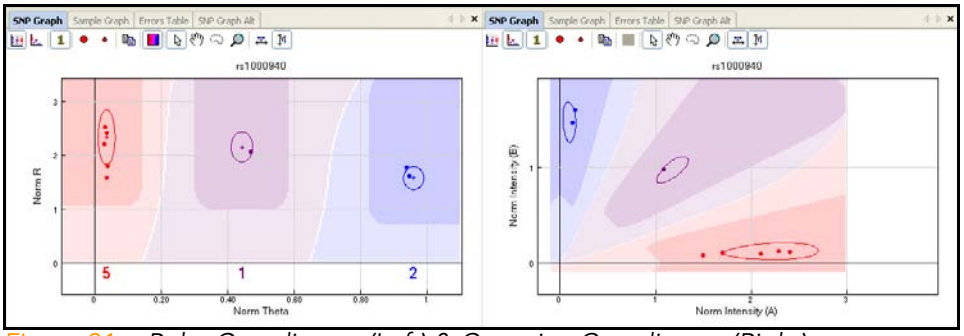


Figure 31 Polar Coordinates (Left) & Cartesian Coordinates (Right)

Normalization

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

Click **1** **Normalization** to turn normalization on or off.

Figure 32 shows a sample graph, in polar coordinates, with normalization turned off (left), and with normalization turned on (right):

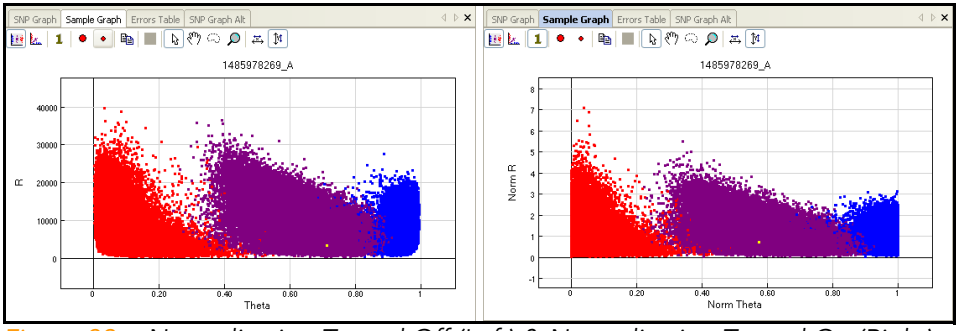


Figure 32 Normalization Turned Off (Left) & Normalization Turned On (Right)

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 33).

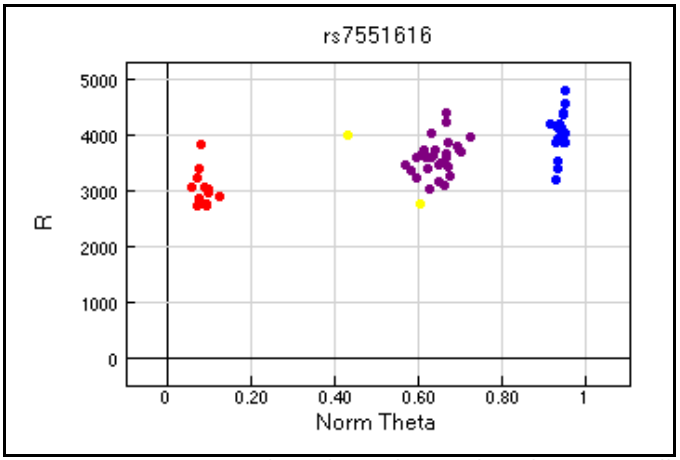


Figure 33 SNP Graph, Selected Samples Shown in Yellow

- 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 Marks** from the context menu.
   The Configure Marks dialog appears (Figure 34).

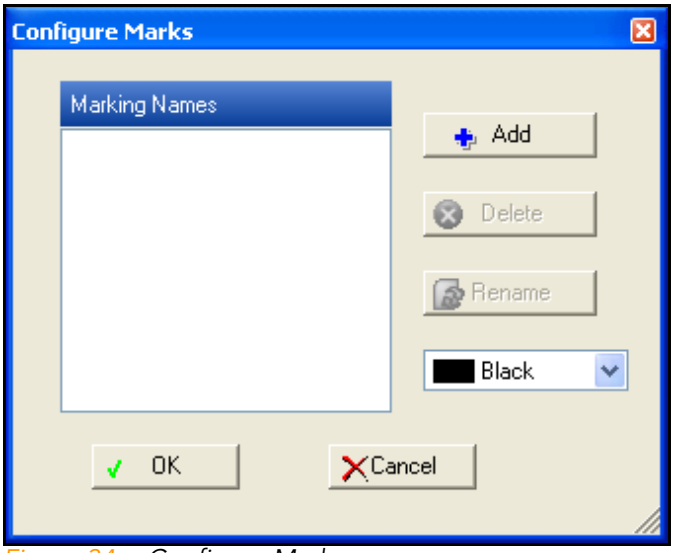


Figure 34 Configure Marks

2. Click **Add** to create a new mark.
   The Select Mark Name dialog appears (Figure 35).

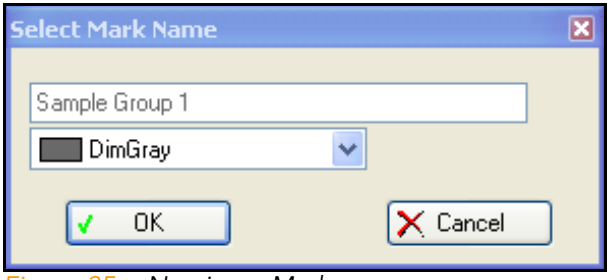


Figure 35 Naming a Mark

3. Give your mark a color by selecting a color from the pulldown menu (Figure 36).

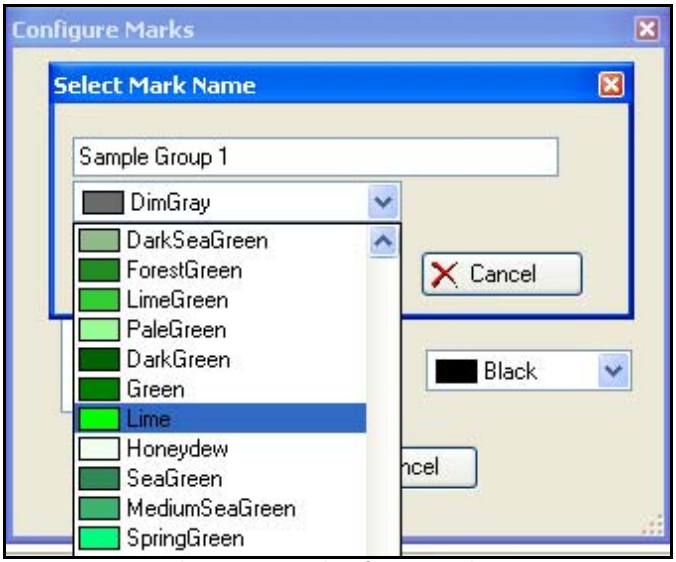


Figure 36 Selecting a Color for a Mark

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 37).

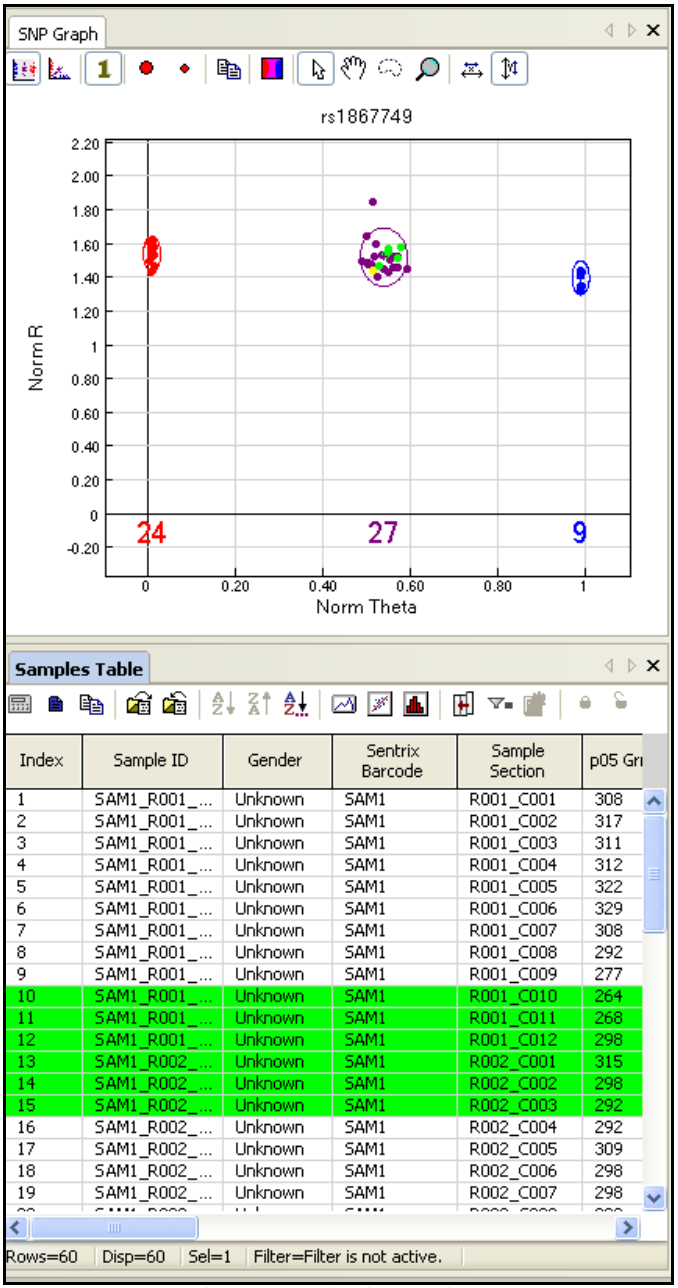


Figure 37 Displaying Marked Samples

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 38).

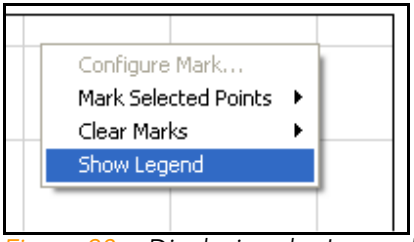


Figure 38 Displaying the Legend

2. Select **Show Legend**.

   The legend appears, and includes the name of your mark (Figure 39).

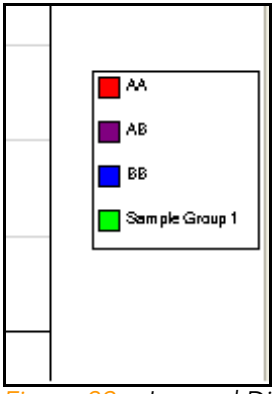


Figure 39 Legend Displaying Mark Name

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. GenomeStudio allows you to 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 40).

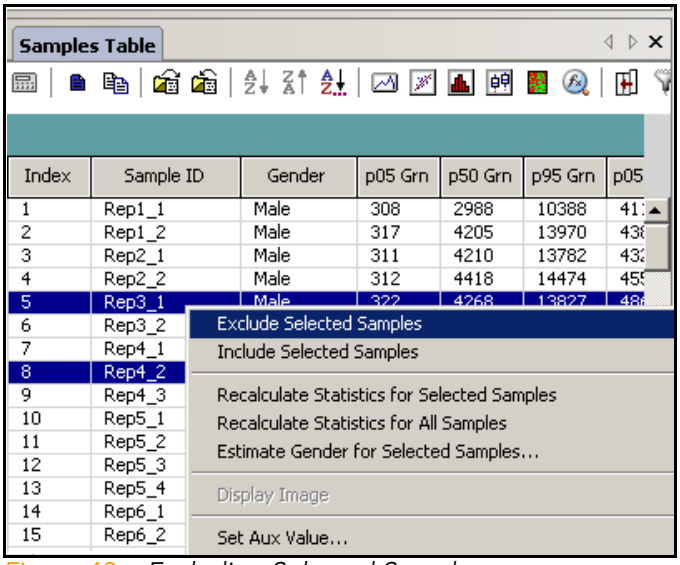


Figure 40 Excluding Selected Sample

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

You can use the SNP 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 SNP Graph.

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 appears (Figure 41).

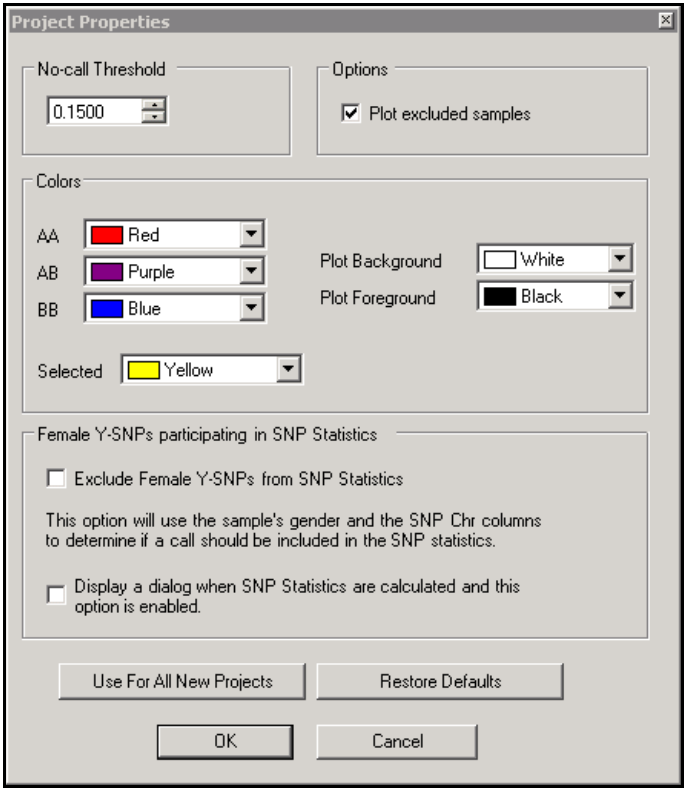


Figure 41 Project Properties

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 appears (Figure 41).

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. See Chapter 8 for descriptions of the columns.

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

   The Column Chooser appears (Figure 42).

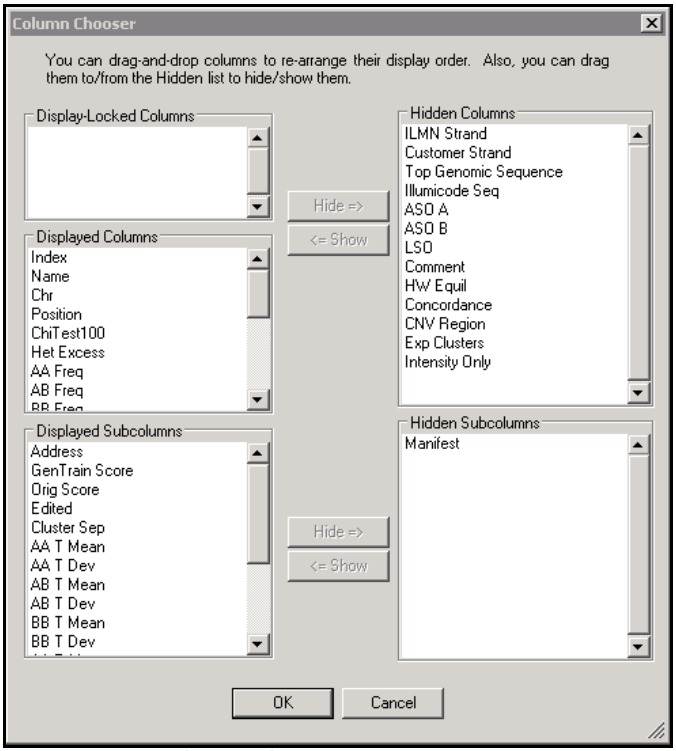


Figure 42 Column Chooser

2. In the Column Chooser dialog, 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.

Viewing the Controls Dashboard

To view a graphic report displaying system controls information:

- Select **Analysis** | **View Controls Dashboard**.

  The Controls window appears (Figure 43).

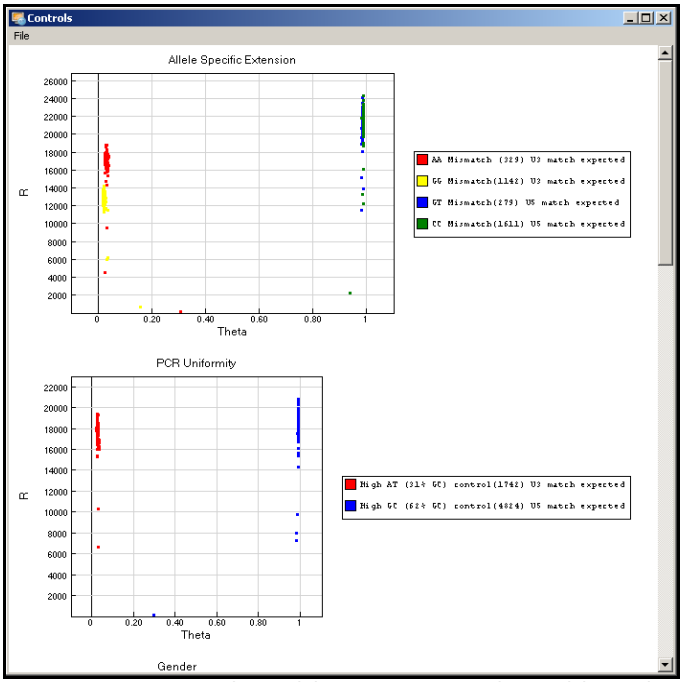


Figure 43 Example GoldenGate Controls Dashboard

NOTE

Excluded samples are not displayed in the Controls dashboard.

For further information about these controls, please refer to the assay manual for your specific application.

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 44).

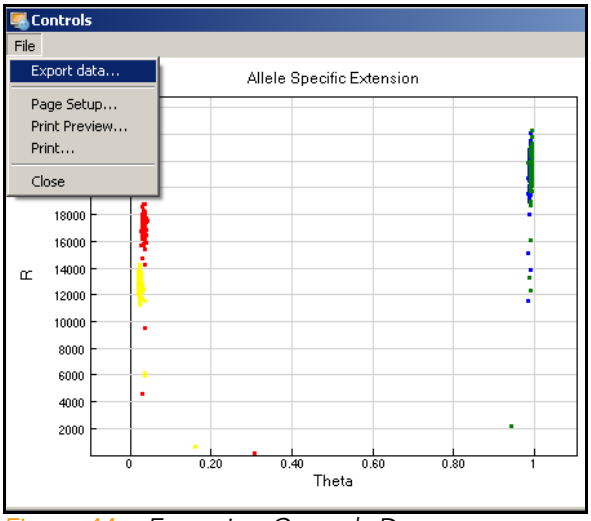


Figure 44 Exporting Controls Data

The Save As dialog appears (Figure 45).

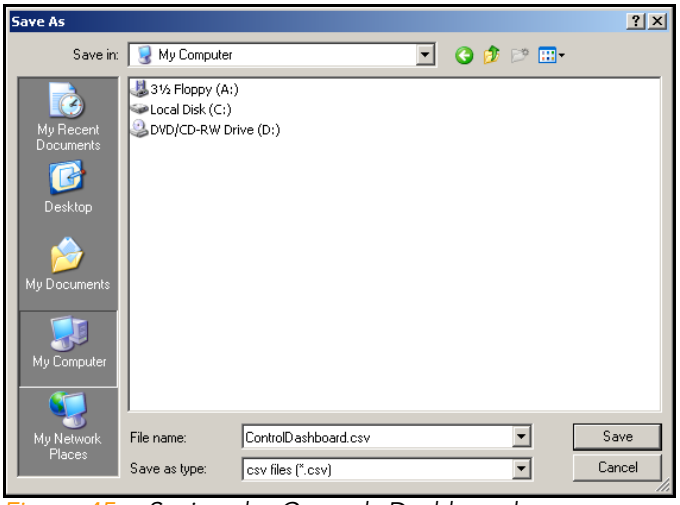


Figure 45 Saving the Controls Dashboard

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 46).

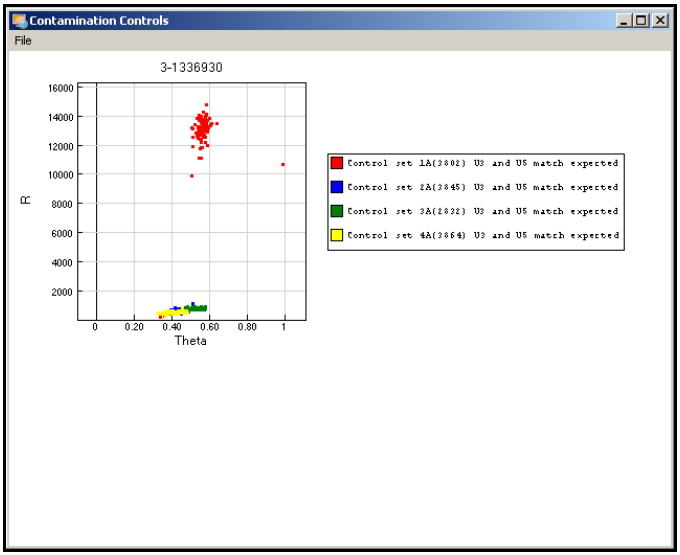


Figure 46 Contamination Dashboard

NOTE

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

Chapter 4
Generating Clusters

Topics

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 GenomeStudio Genotyping Module can automatically determine the cluster positions of the genotypes. If certain SNPs have one or two clusters that lack representation, the GenomeStudio 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.

Running the Clustering Algorithm

1. To run the clustering algorithm, do one of the following:
   - Select **Analysis** | **Cluster All SNPS**.
   - Click **Cluster all SNPS**  (Figure 47).

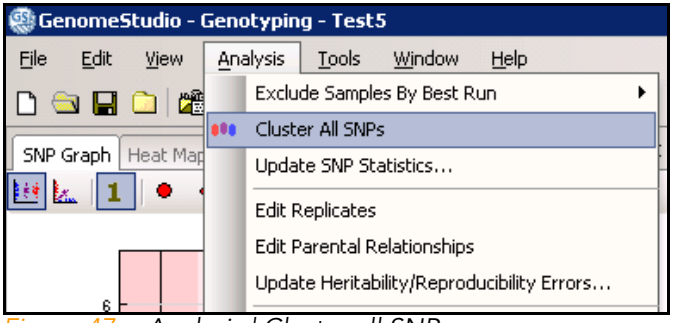


Figure 47 Analysis | Cluster all SNPs

NOTE

Using this feature clusters all SNPs based on the samples in your project.

The clustering algorithm runs, and the GenomeStudio Progress Status bar appears (Figure 48).

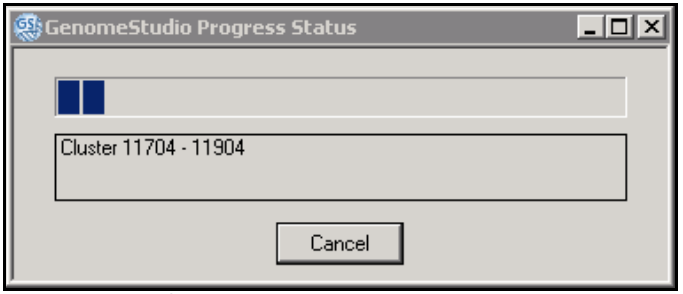


Figure 48 Clustering Progress

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

Reviewing Clusters

To review clusters:

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

The GenomeStudio 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 35.

To shade the calling regions:

- Click **Shade Calling Regions**.
  The calling regions are shaded in the SNP Graph (Figure 49).

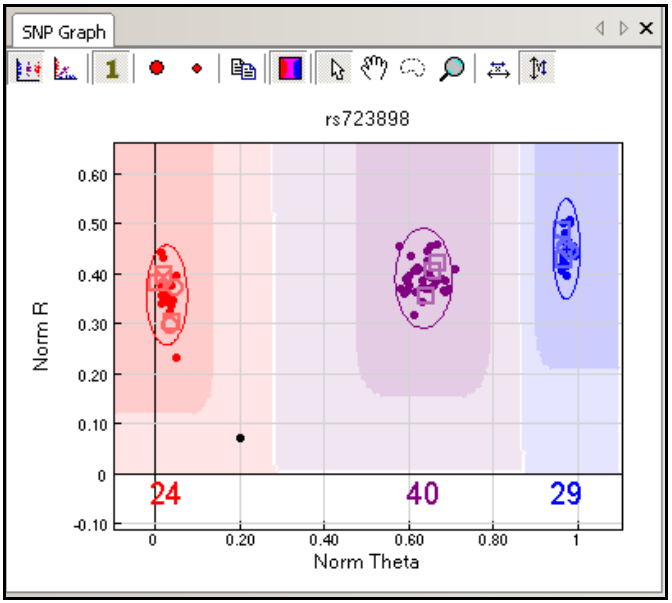


Figure 49 Reviewing Clusters

For more information about shading call regions, see *Shading Call Regions* on page 31.

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 126.

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 50).

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.

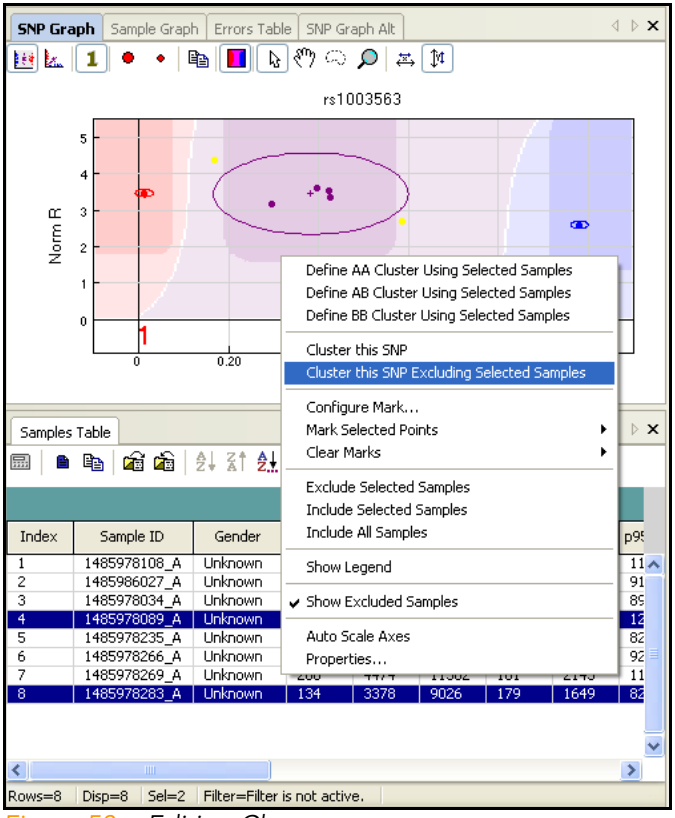


Figure 50 Editing Clusters

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 51).

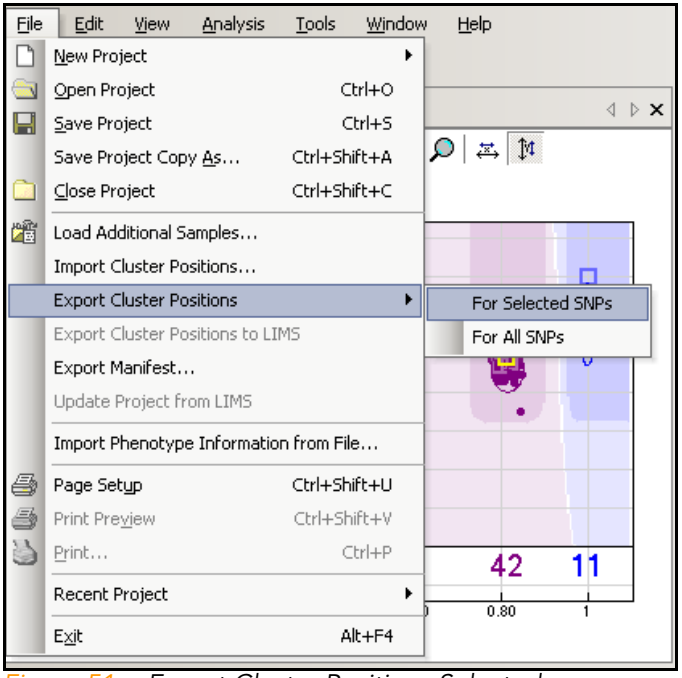


Figure 51 Export Cluster Positions Selected

2. Choose whether you want to export clusters **For Selected SNPs** (for SNPs you selected) or **For All SNPs** (for all SNPs in this project).

   The Save Cluster Positions dialog appears (Figure 52).

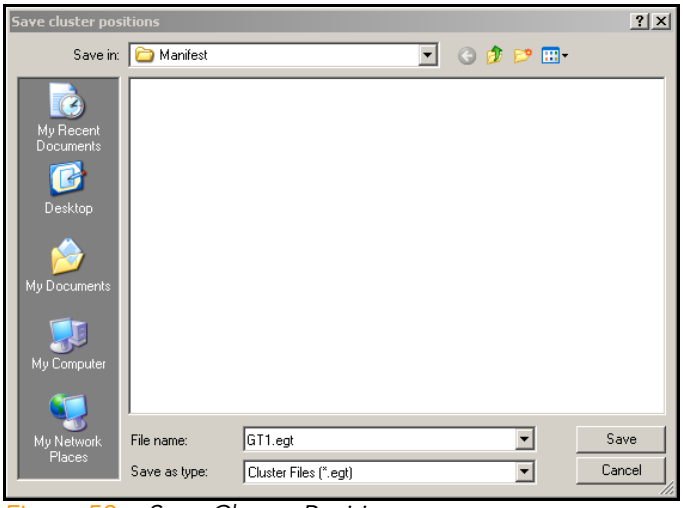


Figure 52 Save Cluster Positions

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 5
Analyzing Your Data

Topics

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

The columns listed above are the only columns you can import into a GenomeStudio genotyping project using a phenotype information file.

Additional columns present in a phenotype information file will not be imported into the GenomeStudio project.

To import phenotype information from a file:

1. Select **File** | **Import Phenotype Information From File**.
   The Import Phenotype File window appears (Figure 53).

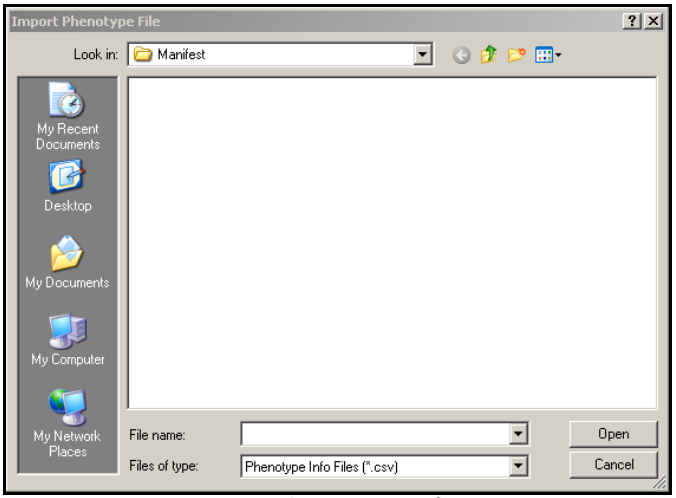
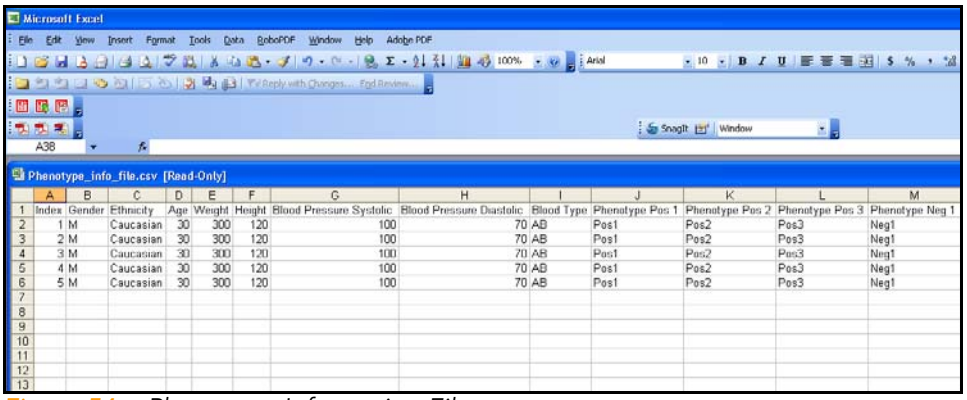


Figure 53 Importing Phenotype Information

2. Browse to a *.csv phenotype information file from which you want to import information (Figure 54).



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	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 54 Phenotype Information File

3. Select **Open**.
   Information from the phenotype information file you selected is imported into GenomeStudio 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 GenomeStudio to estimate gender.

   The selected samples are highlighted in dark blue. Note that the Gender column of each sample contains "Unknown" (Figure 55).

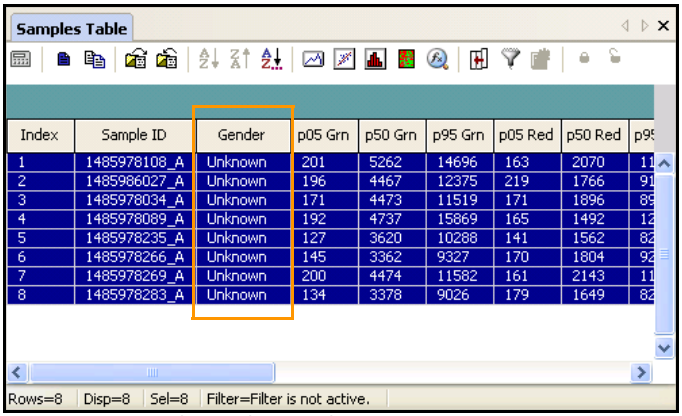


Figure 55 Selected Samples

1. Right-click anywhere on the selected samples.

   The context menu appears (Figure 56).

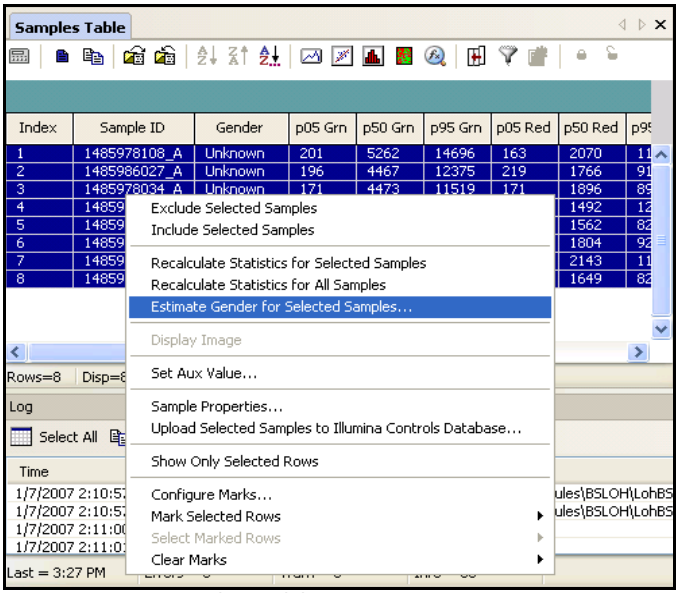


Figure 56 Samples Table Context Menu

2. Select **Estimate Gender for Selected Samples**.

   The Would you like to populate the Gender column... dialog appears (Figure 57).

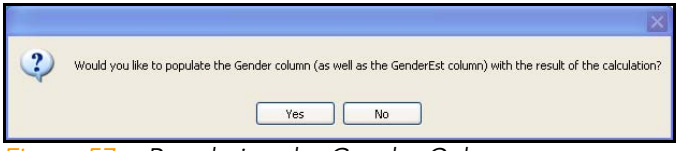


Figure 57 Populating the Gender Column

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 58).

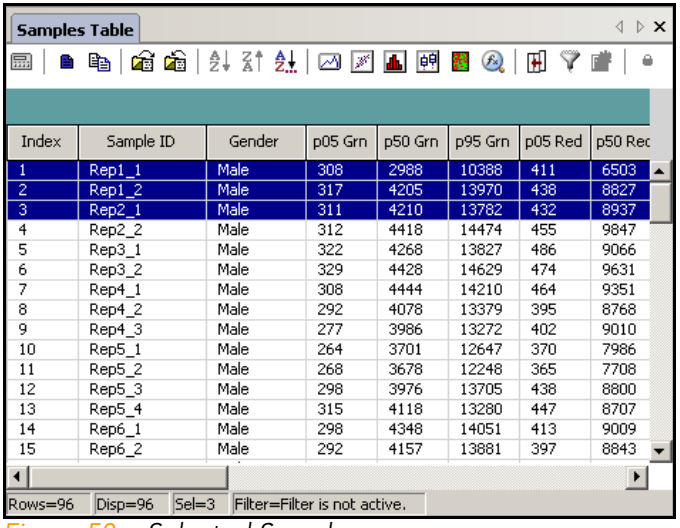


Figure 58 Selected Samples

2. Right-click anywhere on the selected samples.
3. The context menu appears (Figure 59).

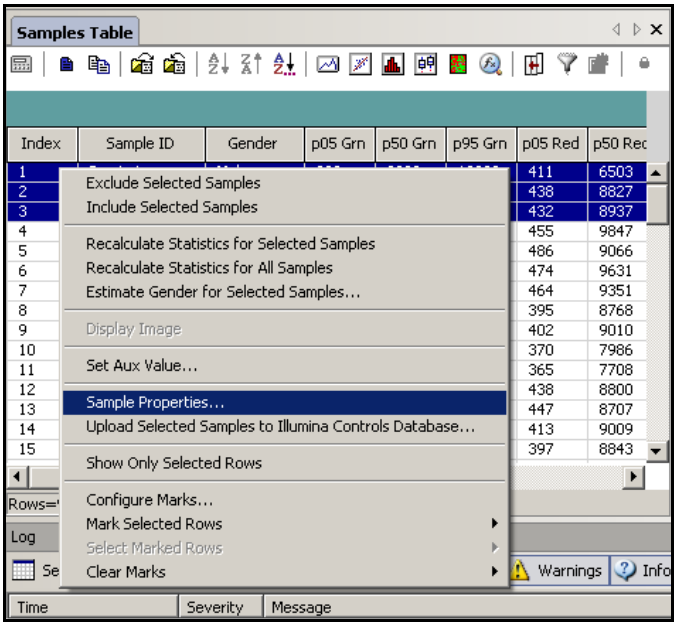


Figure 59 Samples Table Context Menu

4. Select **Sample Properties**.
   The Sample Properties window appears (Figure 60).

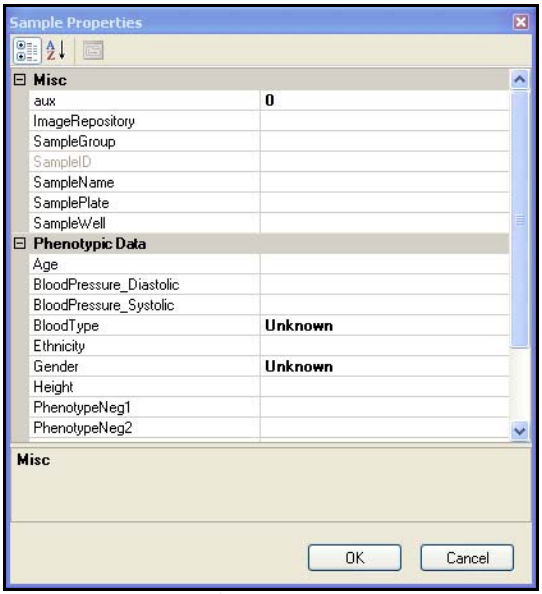


Figure 60 Sample Properties

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. GenomeStudio includes a Paired Sample 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 Sample 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 61 shows an aqua line designating a paired sample subject and reference.

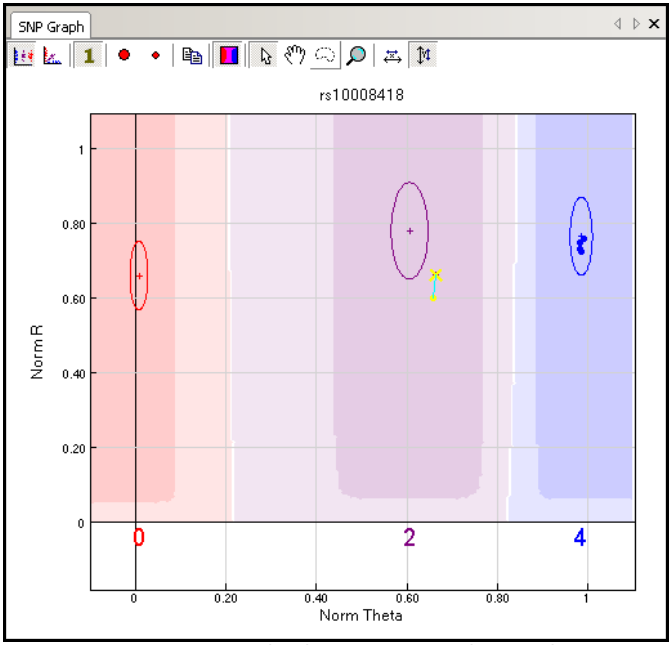


Figure 61 SNP Graph Showing Paired Samples

- 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 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 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 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.

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

Using Column Plug-Ins

You have the option to install column plug-ins as part of the GenomeStudio 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 GenomeStudio.

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

To make column plug-ins available to GenomeStudio, 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 GenomeStudio.

- 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\GenomeStudioGenomeStudio\Plugins

  The column plug-in is now available to GenomeStudio.

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

1. Select **Analysis** | **Create Plug-In Column**.
   The Select Column Plug-In Form dialog appears (Figure 62).

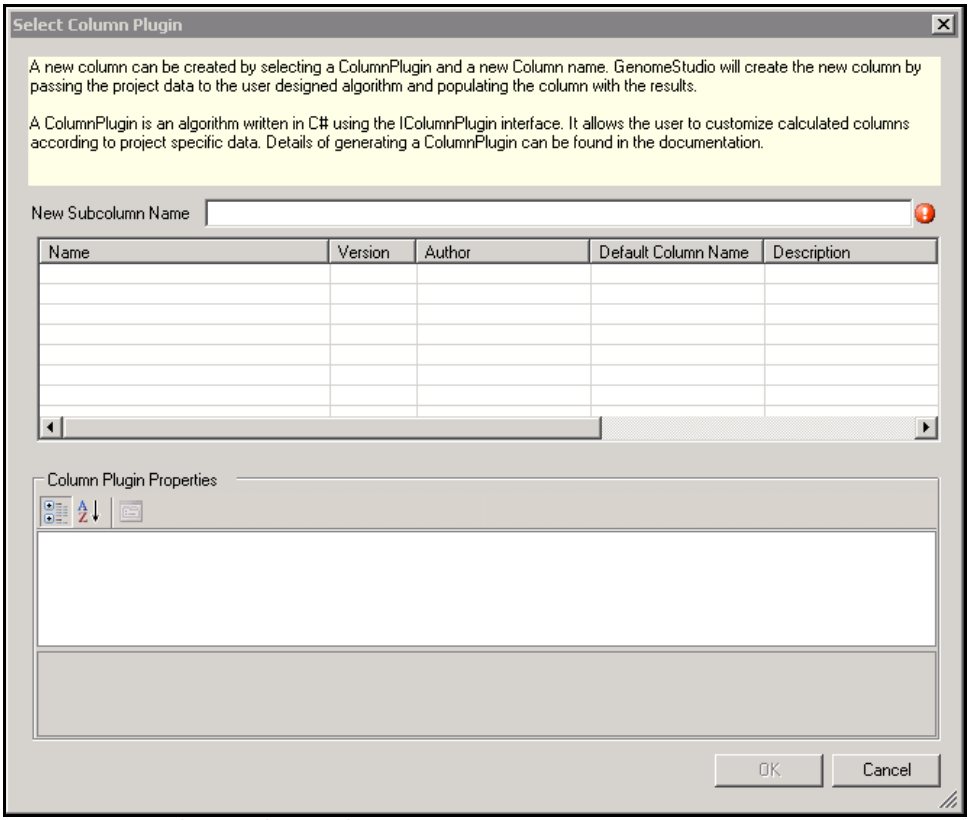


Figure 62 Select Column Plug-In Form

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
Generating Reports

Topics

Introduction

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

GenomeStudio includes a 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 report plug-ins are available, the name of the plug-in report automatically appears at the bottom of the report type list in the Custom Report dropdown menu (Figure 63).

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

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 a report that contains the allele calls of your samples. To generate a Final Report:

1. Run the Report Wizard by selecting **Analysis** | **Reports** | **Report Wizard**.

   The Report Type dialog appears (Figure 63).

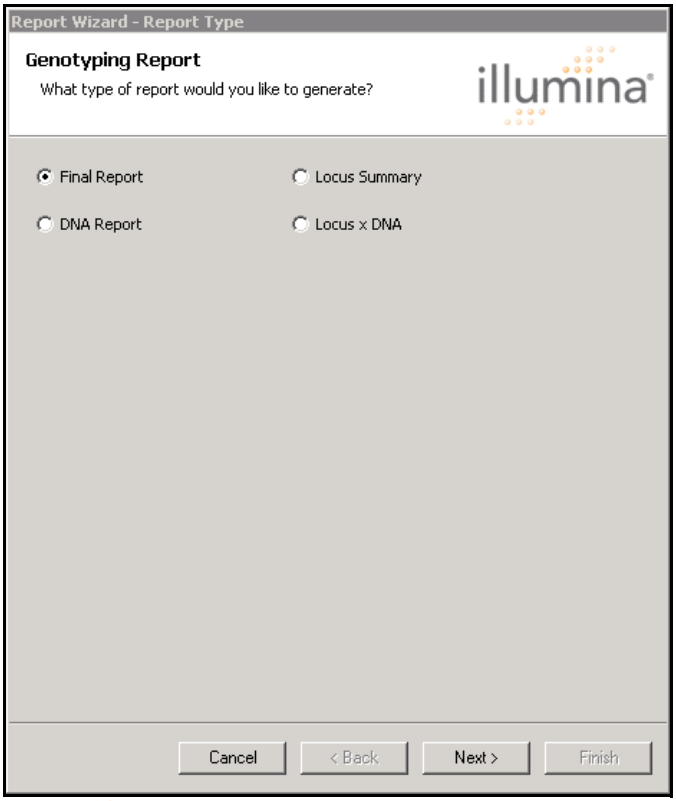


Figure 63 Report Type

Final Report is selected by default.

2. Click **Next**.
   The Included Samples dialog appears (Figure 64).

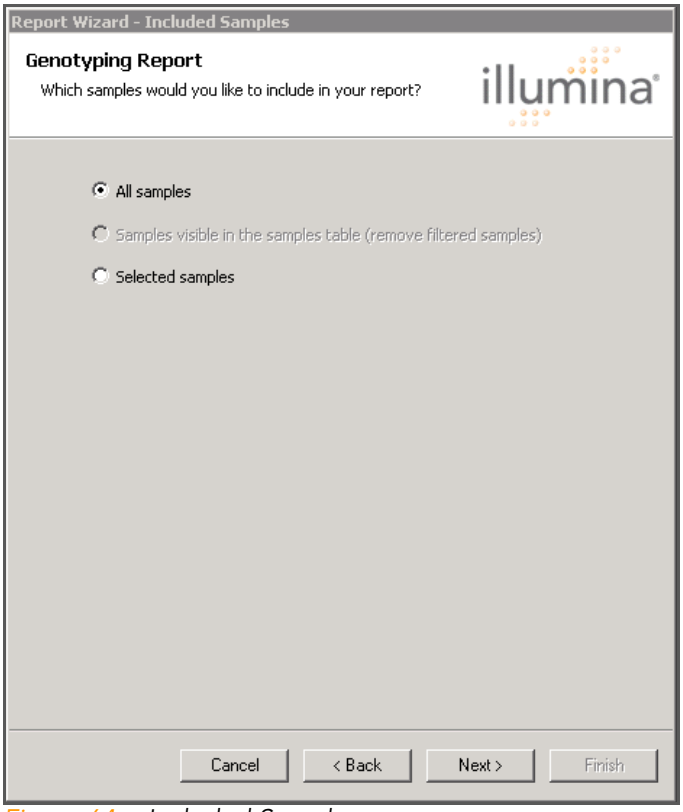


Figure 64 Included Samples

3. Select the samples you would like to include in this Final Report.
4. Click **Next**.
   The Final Report Format dialog appears (Figure 65).

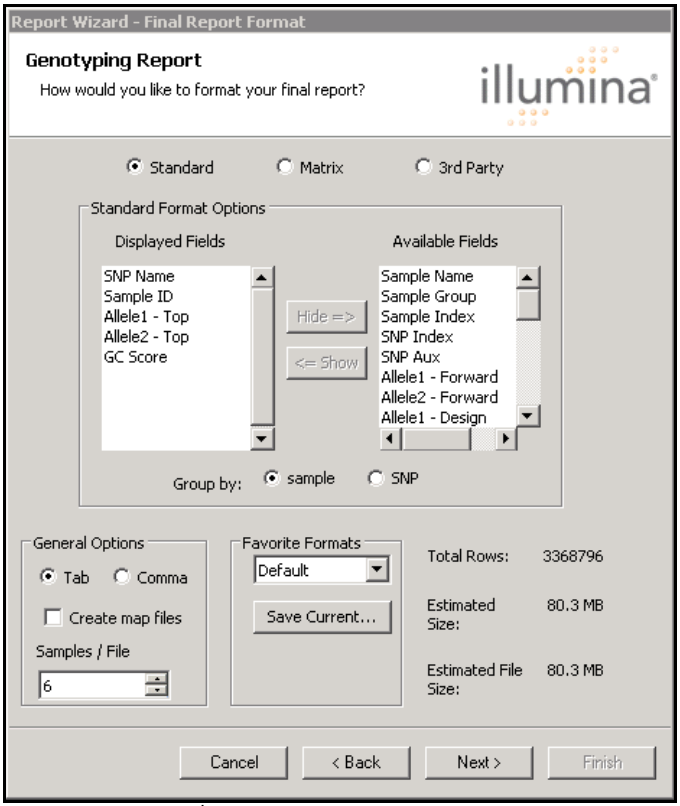


Figure 65 Final Report Format

5. Select a format for your Final Report:

   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 76.

   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 77.

   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 78.

Final Report - Standard Format

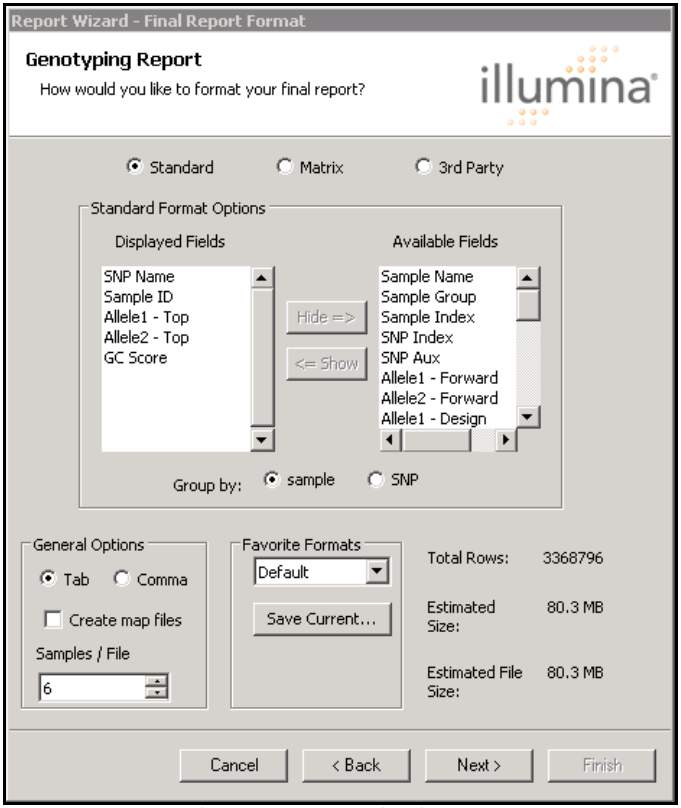


Figure 66 Final Report - Standard Format Options

a. 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.

b. Choose whether you want to group by sample or by SNP.

c. Continue to Step 6.

Final Report - Matrix Format

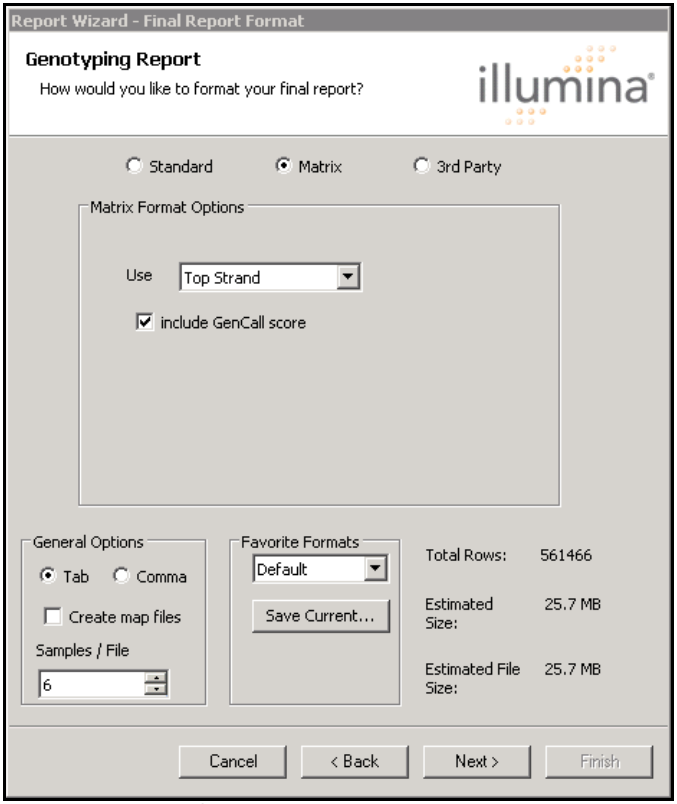


Figure 67 Final Report - Matrix Format Options

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 6.

Final Report - 3rd Party Options

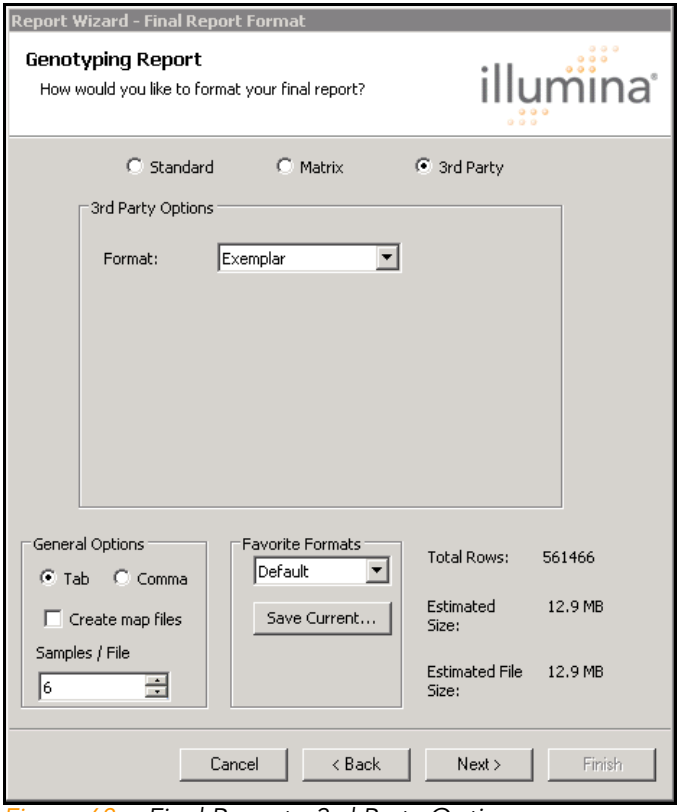


Figure 68 Final Report - 3rd Party Options

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

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

6. 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.

d. Select a favorite format: **Default** or **Default Small**

e. Click **Save Current** to save your current selections as the default selections when creating subsequent Final Reports.

7. Click **Next**.

The Destination dialog appears (Figure 69).

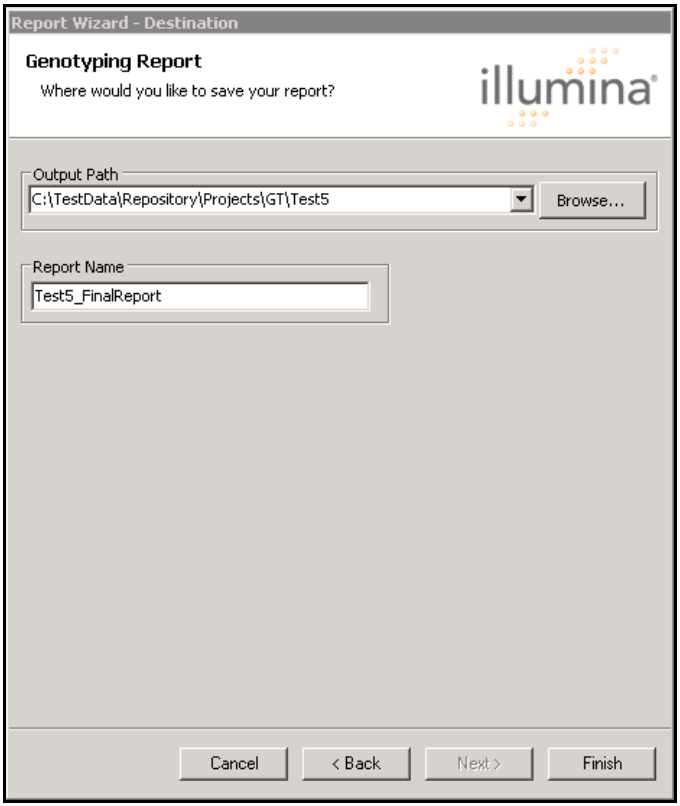


Figure 69 Destination

8. Click **Finish**.

The progress bar alerts you to the status of your report (Figure 70).

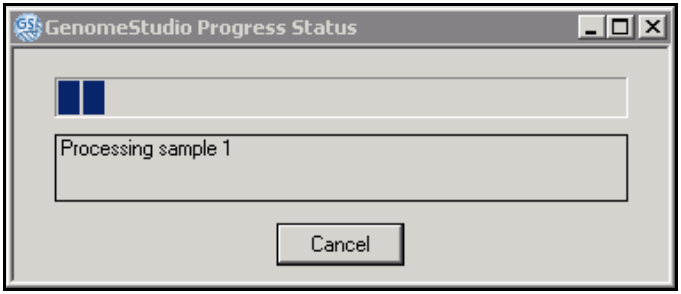
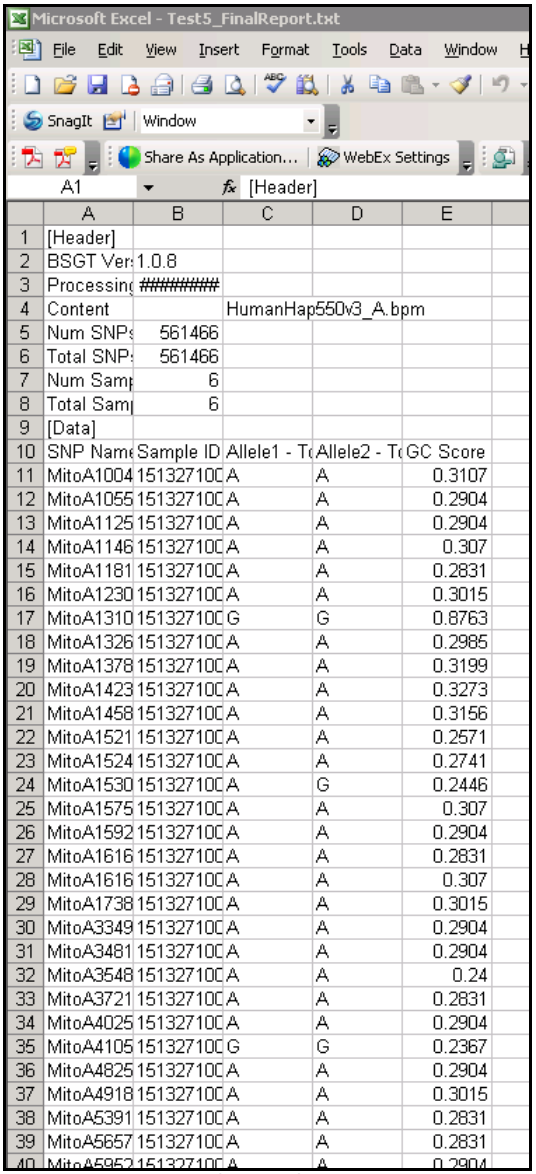


Figure 70 Report Progress

Your report is saved in the location you specified.



	A	B	C	D	E
1	[Header]				
2	BSGT Ver	1.0.8			
3	Processing	##########			
4	Content		HumanHap550v3_A.bpm		
5	Num SNPs	561466			
6	Total SNP	561466			
7	Num Samp	6			
8	Total Sam	6			
9	[Data]				
10	SNP Name	Sample ID	Allele1 - T	Allele2 - T	GC Score
11	MitoA1004	151327100	A	A	0.3107
12	MitoA1055	151327100	A	A	0.2904
13	MitoA1125	151327100	A	A	0.2904
14	MitoA1146	151327100	A	A	0.307
15	MitoA1181	151327100	A	A	0.2831
16	MitoA1230	151327100	A	A	0.3015
17	MitoA1310	151327100	G	G	0.8763
18	MitoA1326	151327100	A	A	0.2985
19	MitoA1378	151327100	A	A	0.3199
20	MitoA1423	151327100	A	A	0.3273
21	MitoA1458	151327100	A	A	0.3156
22	MitoA1521	151327100	A	A	0.2571
23	MitoA1524	151327100	A	A	0.2741
24	MitoA1530	151327100	A	G	0.2446
25	MitoA1575	151327100	A	A	0.307
26	MitoA1592	151327100	A	A	0.2904
27	MitoA1616	151327100	A	A	0.2831
28	MitoA1616	151327100	A	A	0.307
29	MitoA1738	151327100	A	A	0.3015
30	MitoA3349	151327100	A	A	0.2904
31	MitoA3481	151327100	A	A	0.2904
32	MitoA3548	151327100	A	A	0.24
33	MitoA3721	151327100	A	A	0.2831
34	MitoA4025	151327100	A	A	0.2904
35	MitoA4105	151327100	G	G	0.2367
36	MitoA4825	151327100	A	A	0.2904
37	MitoA4918	151327100	A	A	0.3015
38	MitoA5391	151327100	A	A	0.2831
39	MitoA5657	151327100	A	A	0.2831
40	MitoA5952	151327100	A	A	0.2904

Figure 71 Sample Final Report

DNA Report

The DNA Report is a comma-delimited text file (*.csv file) that includes the columns described in Table 1.

To generate a DNA Report:

1. Run the Report Wizard by selecting **Analysis** | **Reports** | **Report Wizard**.

   The Report Type dialog appears.

2. Select **DNA Report** (Figure 72).

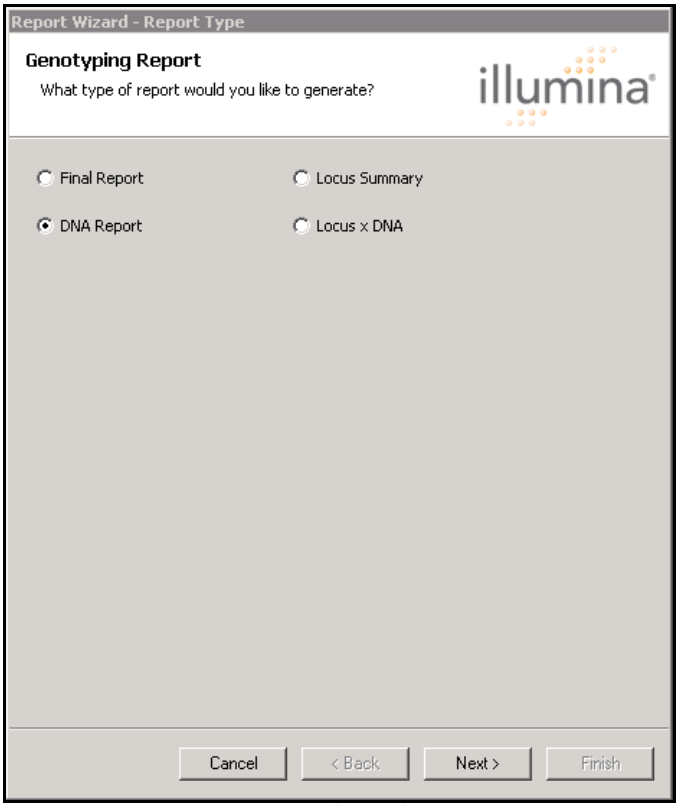


Figure 72 DNA Report Selected

3. Click **Next**.

   The Destination dialog appears (Figure 73).

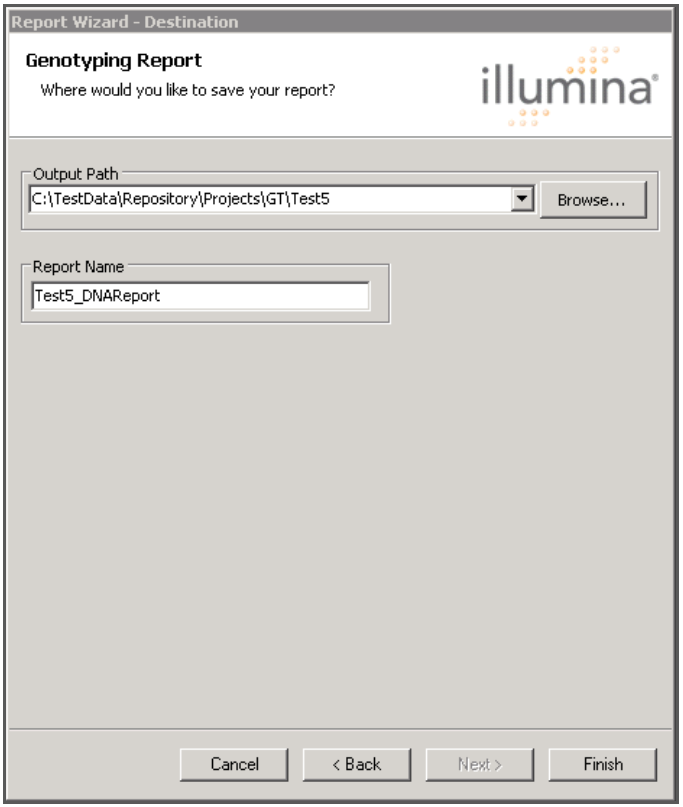


Figure 73 Destination

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 (Figure 74) is saved with the name and parameters you assigned to it in the location you specified.

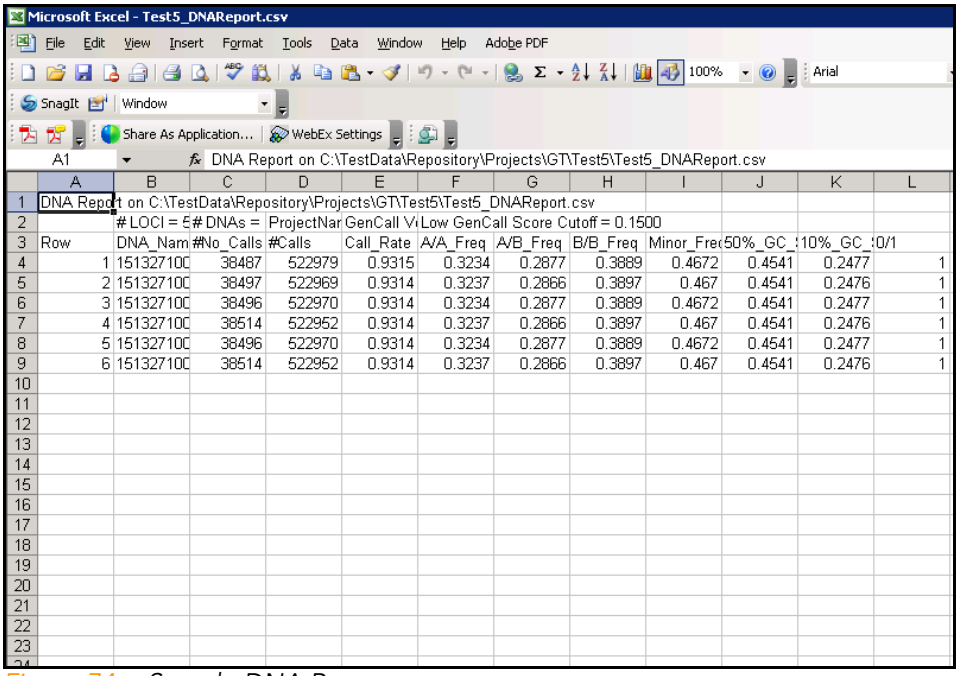


Figure 74 Sample DNA Report

Column Descriptions The DNA Report includes the columns described in Table 1.

Table 1 DNA Report - 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 1 DNA Report - 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 If the number of AA < number of BB for a sample, the frequency for the minor allele A for that sample is (2*AAs + ABs) for the sample divided by (2*AAs + ABs + BBs) for the sample across all loci.
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	A formula determines whether a sample is recommended for inclusion or exclusion. 0 = Remove 1 = Include

Locus Summary Report

The Locus Summary Report is a comma-delimited text file (.csv file) that includes the columns described in Table 2.

To generate a Locus Summary Report:

1. Run the Report Wizard by selecting **Analysis** | **Reports** | **Report Wizard**.

   The Report Type dialog appears.

2. Select **Locus Summary Report** (Figure 72).

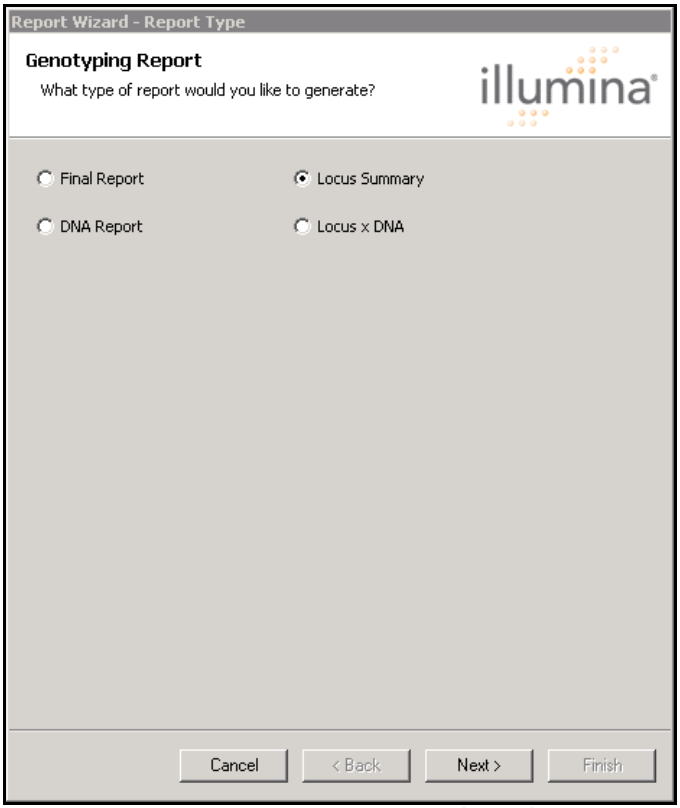


Figure 75 Locus Summary Report Selected

3. Click **Next**.

   The Destination dialog appears (Figure 76).

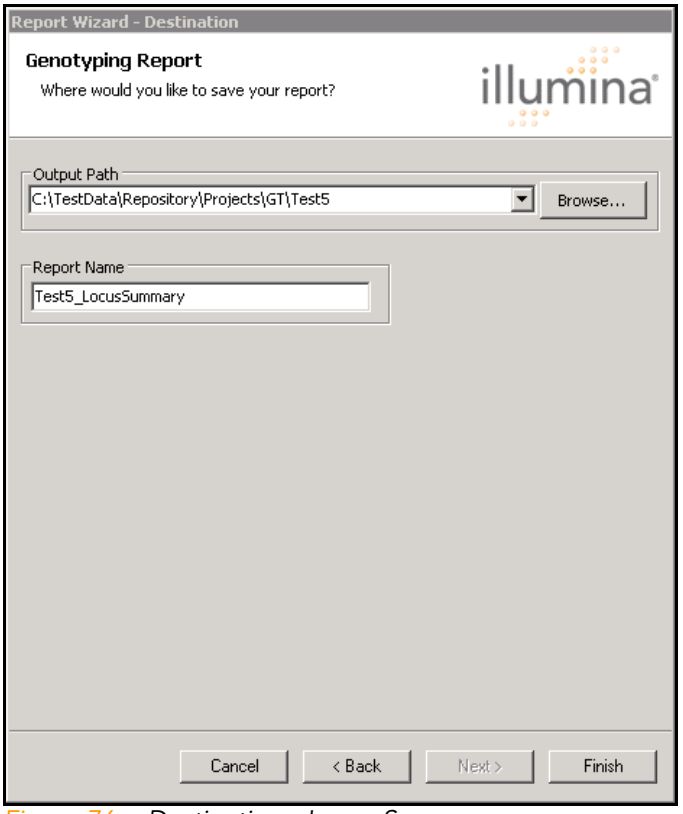
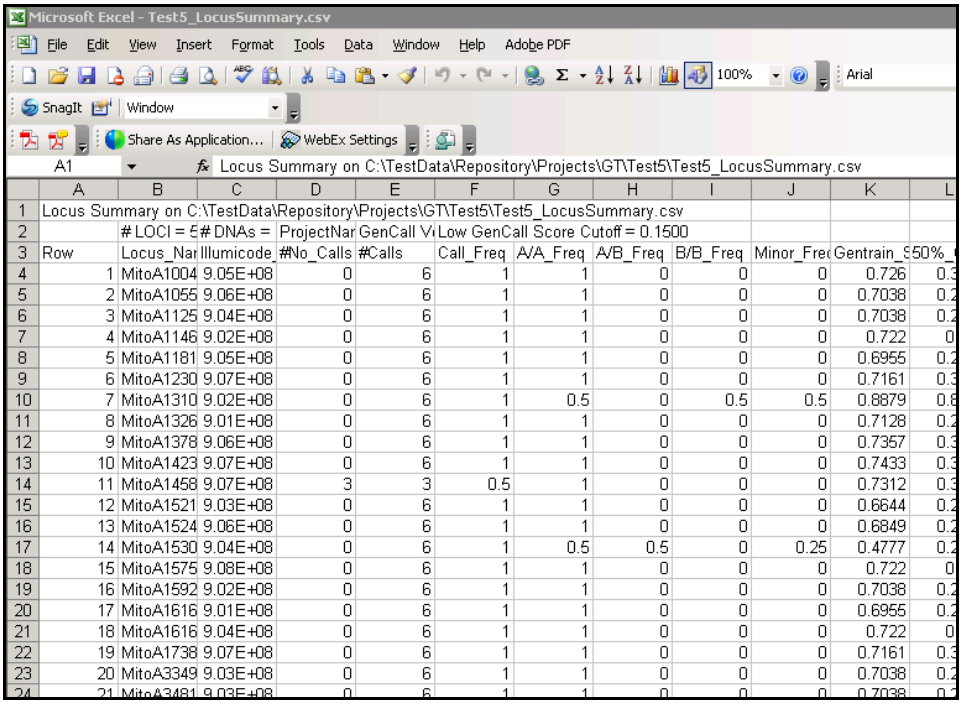


Figure 76 Destination - Locus Summary

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 (Figure 77) is saved with the name and parameters you assigned to it in the location you specified.



Microsoft Excel - Test5_LocusSummary.csv

File Edit View Insert Format Tools Data Window Help Adobe PDF

SnagIt Window

Share As Application... WebEx Settings

A1 Locus Summary on C:\TestData\Repository\Projects\GT\Test5\Test5_LocusSummary.csv

	A	B	C	D	E	F	G	H	I	J	K	L
1	Locus Summary on C:\TestData\Repository\Projects\GT\Test5\Test5_LocusSummary.csv											
2		# LOCI = 5	# DNAs =	ProjectNar	GenCall V	Low GenCall Score Cutoff = 0.1500						
3	Row	Locus_Nar	Illumicode_	#No_Calls	#Calls	Call_Freq	AA_Freq	AB_Freq	BB_Freq	Minor_Fre	Gentrain_S	50%_
4	1	MitoA1004	9.05E+08	0	6	1	1	0	0	0	0.726	0.3
5	2	MitoA1055	9.06E+08	0	6	1	1	0	0	0	0.7038	0.2
6	3	MitoA1125	9.04E+08	0	6	1	1	0	0	0	0.7038	0.2
7	4	MitoA1146	9.02E+08	0	6	1	1	0	0	0	0.722	0
8	5	MitoA1181	9.05E+08	0	6	1	1	0	0	0	0.6955	0.2
9	6	MitoA1230	9.07E+08	0	6	1	1	0	0	0	0.7161	0.3
10	7	MitoA1310	9.02E+08	0	6	1	0.5	0	0.5	0.5	0.8879	0.8
11	8	MitoA1326	9.01E+08	0	6	1	1	0	0	0	0.7128	0.2
12	9	MitoA1378	9.06E+08	0	6	1	1	0	0	0	0.7357	0.3
13	10	MitoA1423	9.07E+08	0	6	1	1	0	0	0	0.7433	0.3
14	11	MitoA1458	9.07E+08	3	3	0.5	1	0	0	0	0.7312	0.3
15	12	MitoA1521	9.03E+08	0	6	1	1	0	0	0	0.6644	0.2
16	13	MitoA1524	9.06E+08	0	6	1	1	0	0	0	0.6849	0.2
17	14	MitoA1530	9.04E+08	0	6	1	0.5	0.5	0	0.25	0.4777	0.2
18	15	MitoA1575	9.08E+08	0	6	1	1	0	0	0	0.722	0
19	16	MitoA1592	9.02E+08	0	6	1	1	0	0	0	0.7038	0.2
20	17	MitoA1616	9.01E+08	0	6	1	1	0	0	0	0.6955	0.2
21	18	MitoA1616	9.04E+08	0	6	1	1	0	0	0	0.722	0
22	19	MitoA1738	9.07E+08	0	6	1	1	0	0	0	0.7161	0.3
23	20	MitoA3349	9.03E+08	0	6	1	1	0	0	0	0.7038	0.2
24	21	MitoA3481	9.03E+08	0	6	1	1	0	0	0	0.7038	0.2

Figure 77 Sample Locus Summary Report

Column Descriptions The Locus Summary Report includes the columns described in Table 2.

Table 2 Locus Summary Report - 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

Table 2 Locus Summary Report - Column Descriptions

Column	Description
Call_Freq	Call frequency, or call rate, calculated as follows: #Calls/(#No_Calls + #Calls)
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 If the number of AA < number of BB for a sample, the frequency for the minor allele A for that sample is (2*AAs + ABs) for the sample divided by (2*AAs + ABs + BBs) for the sample across all loci.
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	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: $$(f_{AB} - 2pq)/(2pq + \varepsilon)$$ 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.
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

Table 2 Locus Summary Report - Column Descriptions

Column	Description
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
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 appears.

2. Select **Locus x DNA Report** (Figure 72).

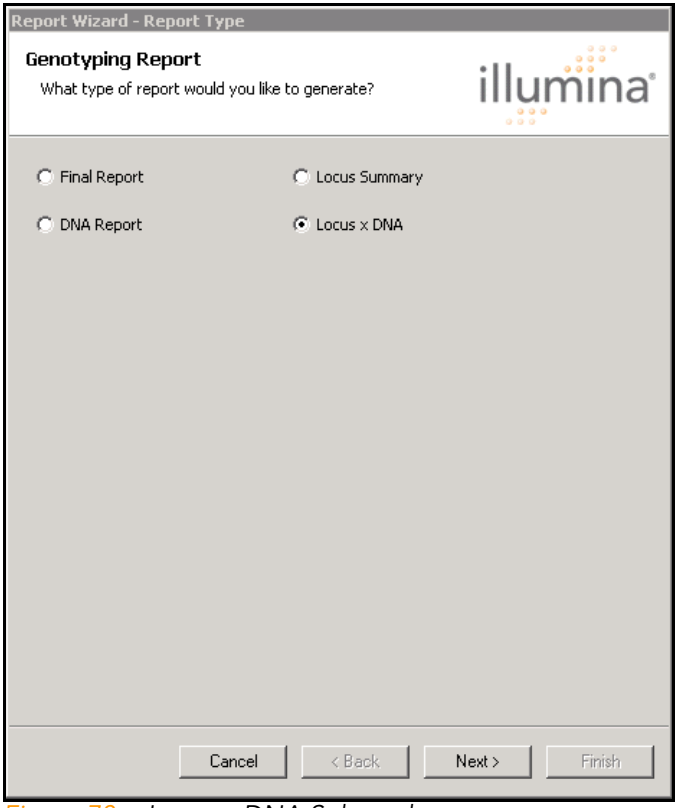


Figure 78 Locus x DNA Selected

3. Click **Next**.

   The Destination dialog appears (Figure 79).

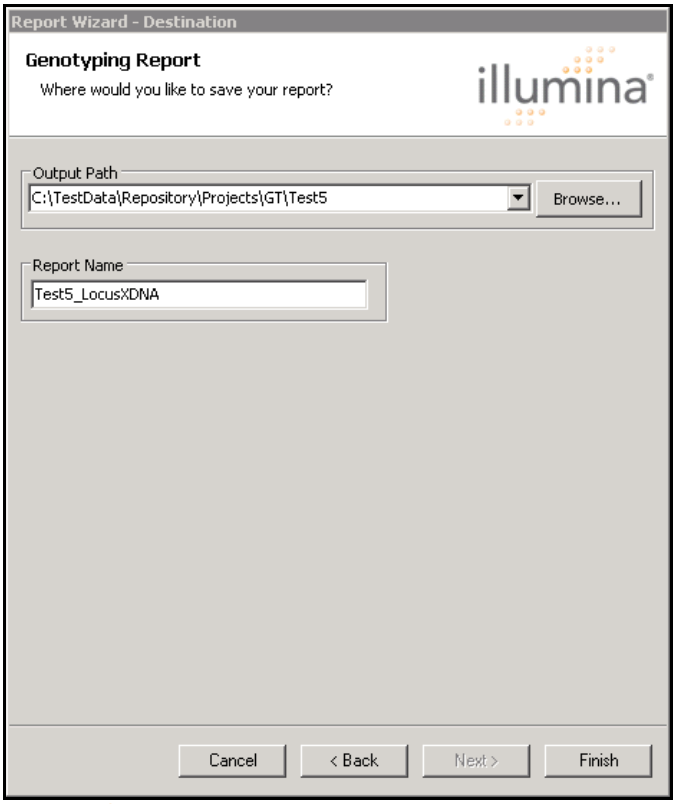
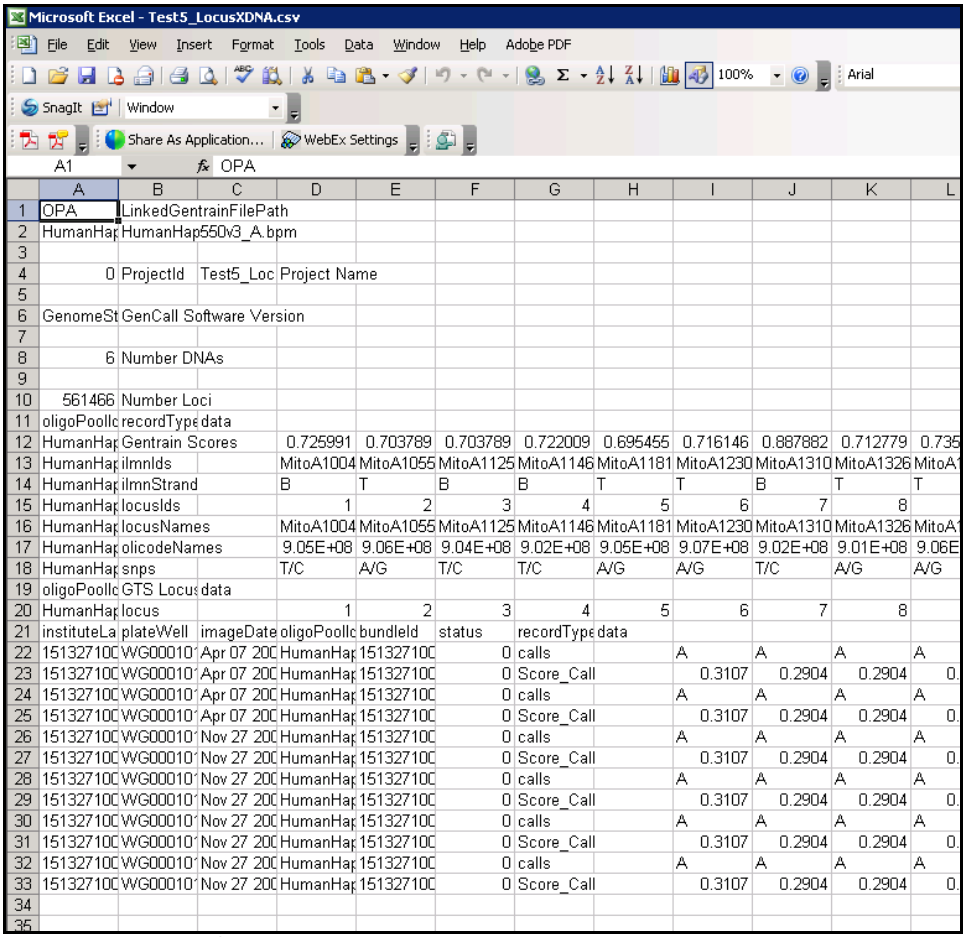


Figure 79 Destination - Locus x DNA

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 (Figure 80) is saved with the name and parameters you assigned to it in the location you specified.



Microsoft Excel - Test5_LocusXDNA.csv

File Edit View Insert Format Tools Data Window Help Adobe PDF

SnagIt Window

Share As Application... WebEx Settings

A1 OPA

	A	B	C	D	E	F	G	H	I	J	K	L
1	OPA	LinkedGentrainFilePath										
2	HumanHap	HumanHap550v3_A.bpm										
3												
4	0	ProjectId	Test5_Loc	Project Name								
5												
6	GenomeSt	GenCall Software Version										
7												
8	6	Number DNAs										
9												
10	561466	Number Loci										
11	oligoPoolIc	recordType	data									
12	HumanHap	Gentrain Scores		0.725991	0.703789	0.703789	0.722009	0.695455	0.716146	0.887882	0.712779	0.735
13	HumanHap	ilmnIds		MitoA1004	MitoA1055	MitoA1125	MitoA1146	MitoA1181	MitoA1230	MitoA1310	MitoA1326	MitoA1
14	HumanHap	ilmnStrand		B	T	B	B	T	T	B	T	T
15	HumanHap	locusIds		1	2	3	4	5	6	7	8	
16	HumanHap	locusNames		MitoA1004	MitoA1055	MitoA1125	MitoA1146	MitoA1181	MitoA1230	MitoA1310	MitoA1326	MitoA1
17	HumanHap	olicodeNames		9.05E+08	9.06E+08	9.04E+08	9.02E+08	9.05E+08	9.07E+08	9.02E+08	9.01E+08	9.06E
18	HumanHap	snps		T/C	A/G	T/C	T/C	A/G	A/G	T/C	A/G	A/G
19	oligoPoolIc	GTS Locus	data									
20	HumanHap	locus		1	2	3	4	5	6	7	8	
21	instituteLa	plateWell	imageDate	oligoPoolIc	bundleId	status	recordType	data				
22	151327100	WG00010	Apr 07 200	HumanHap	151327100	0	calls		A	A	A	A
23	151327100	WG00010	Apr 07 200	HumanHap	151327100	0	Score_Call		0.3107	0.2904	0.2904	0.
24	151327100	WG00010	Apr 07 200	HumanHap	151327100	0	calls		A	A	A	A
25	151327100	WG00010	Apr 07 200	HumanHap	151327100	0	Score_Call		0.3107	0.2904	0.2904	0.
26	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	calls		A	A	A	A
27	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	Score_Call		0.3107	0.2904	0.2904	0.
28	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	calls		A	A	A	A
29	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	Score_Call		0.3107	0.2904	0.2904	0.
30	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	calls		A	A	A	A
31	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	Score_Call		0.3107	0.2904	0.2904	0.
32	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	calls		A	A	A	A
33	151327100	WG00010	Nov 27 200	HumanHap	151327100	0	Score_Call		0.3107	0.2904	0.2904	0.
34												
35												

Figure 80 Sample Locus x DNA Report

Column Descriptions

The Locus x DNA Report is a comma-delimited text file (.csv file) that includes the columns described in Table 3.

Table 3 Locus x DNA Report - 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 GenomeStudio Genotyping Module.

To generate a Reproducibility and Heritability Report:

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

   The Reproducibility and Heritability dialog appears (Figure 81).

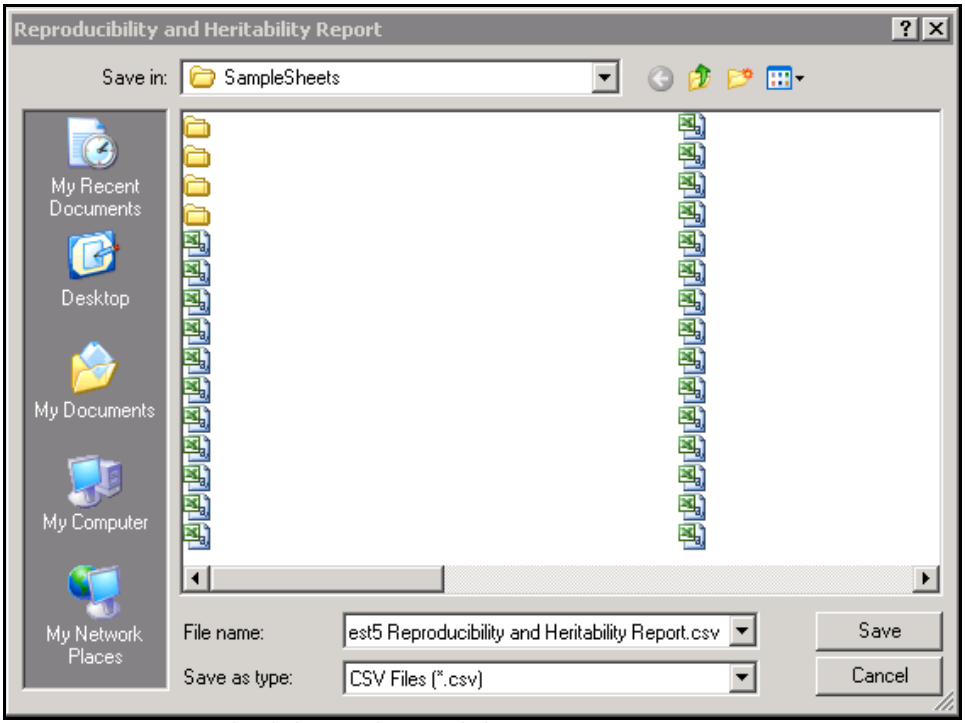


Figure 81 Reproducibility and Heritability

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 82).

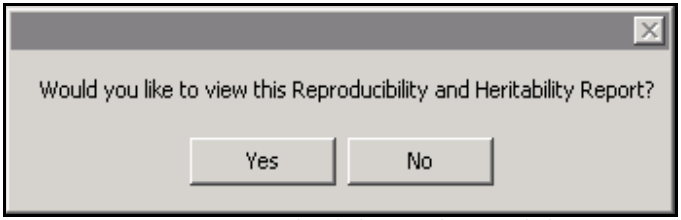
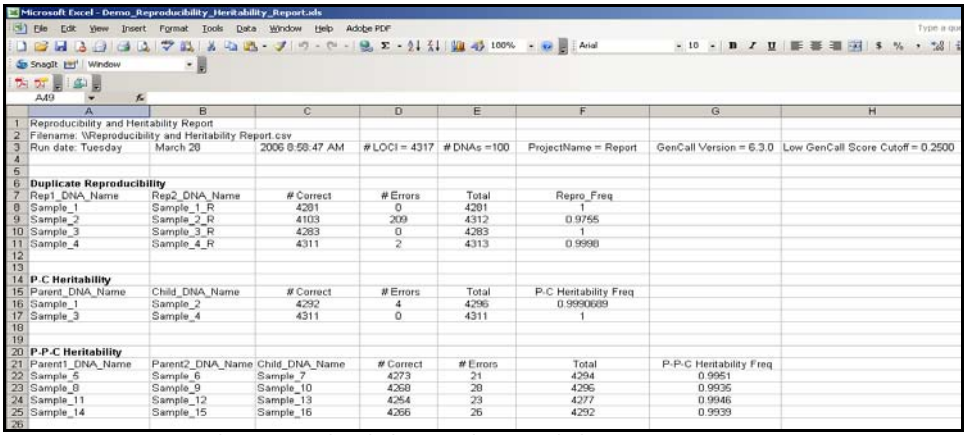


Figure 82 View Reproducibility and Heritability

5. Do one of the following:

- Click **Yes** to view the Reproducibility and Heritability Report.

  The Reproducibility and Heritability Report appears (Figure 83).



Microsoft Excel - Demo_Reproducibility_Heritability_Report.xls

	A	B	C	D	E	F	G	H
1	Reproducibility and Heritability Report							
2	Filename: \\Reproducibility and Heritability Report.csv							
3	Run date: Tuesday	March 28	2006 6:58:47 AM	# LOCI = 4317	# DNAs =100	ProjectName = Report	GenCall Version = 6.3.0	Low GenCall Score Cutoff = 0.2500
4								
5								
6	**Duplicate Reproducibility**							
7	Rep1_DNA_Name	Rep2_DNA_Name	# Correct	# Errors	Total	Repro_Freq		
8	Sample_1	Sample_1_R	4281	0	4281	1		
9	Sample_2	Sample_2_R	4103	209	4312	0.9755		
10	Sample_3	Sample_3_R	4283	0	4283	1		
11	Sample_4	Sample_4_R	4311	2	4313	0.9998		
12								
13								
14	**P-C Heritability**							
15	Parent_DNA_Name	Child_DNA_Name	# Correct	# Errors	Total	P-C Heritability Freq		
16	Sample_1	Sample_2	4292	4	4296	0.9990689		
17	Sample_3	Sample_4	4311	0	4311	1		
18								
19								
20	**P-P-C Heritability**							
21	Parent1_DNA_Name	Parent2_DNA_Name	Child_DNA_Name	# Correct	# Errors	Total	P-P-C Heritability Freq	
22	Sample_5	Sample_6	Sample_7	4273	21	4294	0.9951	
23	Sample_8	Sample_9	Sample_10	4268	28	4296	0.9935	
24	Sample_11	Sample_12	Sample_13	4254	23	4277	0.9946	
25	Sample_14	Sample_15	Sample_16	4266	26	4292	0.9939	
26								

Figure 83 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.

Column Descriptions and Examples

The following sections include Reproducibility Report column descriptions, and examples of the three main report sections:

- Duplicate Reproducibility
- Parent-Child Heritability
- Parent-Parent-Child Heritability

Duplicate Reproducibility Columns

Table 4 describes the columns of the Duplicate Reproducibility section of the Reproducibility and Heritability Report.

Table 4 Reproducibility and Heritability Report - Duplicate Reproducibility

Column	Description
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 loci with consistent replicate genotype comparisons
# Errors	Number of loci with inconsistent replicate genotype comparisons
Total	Number of total genotype comparisons (one genotype comparison per locus per replicate pair). Does not include genotypes with intensities that fall below the no-call threshold (low GenCall Score Cutoff). Equals (# Correct + # Errors).
Repro_Freq	Reproducibility frequency, calculated as sqrt(1 - error rate). The error rate does not include genotype calls that fall below the no-call threshold.

Table 5 is an example of the Duplicate Reproducibility section of a Reproducibility and Heritability Report.

Table 5 Example - Duplicate Reproducibility

Rep1 Genotype	Rep2 Genotype	# Correct	# Errors	Repro_Freq
AB	AB	1	0	1
AA	AB	0	1	0
AA	BB	0	1	0
AA	No call	0	0	NAN

Parent-Child Heritability Columns

Table 6 describes the columns of the Parent-Child Heritability section of the Reproducibility and Heritability Report.

Table 6 Reproducilbility and Heritability Report - P-C Heritability

Column	Description
Parent_DNA_Name	Name of the sample designated as parent in a P-C relationship.
Child_DNA_Name	Name of the sample designated as child in a P-C relationship.
# Correct	Number of loci with consistent Parent-Child genotype comparisons
# Errors	Number of loci with inconsistent Parent-Child genotype comparisons
Total	Number of total genotype comparisons (one genotype comparison per locus per Parent-Child pair). Does not include genotype comparisons with intensities that fall below the no-call threshold (low GenCall Score Cutoff). Equals (# Correct + # Errors).
PC_Heritability_Freq	Heritability frequency calculated as (# Correct / # Total)

Table 7 is an example of the Parent-Child Heritability section of a Reproducibility and Heritability Report.

Table 7 Example - Parent-Child Heritability

Parent Genotype	Child Genotype	# Correct	# Errors	P-C Heritability Freq
AA	BB	0	1	0
AA	AB	1	0	1
AA	No call	0	0	NAN

Parent-Parent-Child Heritability Columns

Table 8 describes the columns of the Parent-Parent-Child Heritability section of the Reproducibility and Heritability Report.

Table 8 Reproducilbility and Heritability Report - P-P-C Heritability

Column	Description
Parent1_DNA_Name	Name of the sample designated as parent #1 in a P-P-C relationship.
Parent2_DNA_Name	Name of the sample designated as parent #2 in a P-P-C relationship.
Child_DNA_Name	Name of the sample designated as child in a P-P-C relationship.
# Correct	Number of loci with consistent Parent1-Child ***and*** Parent2-Child genotype comparisons
# Errors	Number of loci with inconsistent Parent1-Child ***or*** Parent2-Child genotype comparisons
Total	Number total of loci that contribute to the trio heritability analysis. Does not include loci where Parent1, Parent2 or Child have genotypes with intensities that fall below the no-call threshold (low GenCall Score Cutoff).
P-P-C Heritability Freq	Heritability frequency calculated as (# Correct / # Total)

Table 9 is an example of the Parent-Parent-Child Heritability section of a Reproducibility and Heritability Report.

Table 9 Example - Parent-Parent-Child Heritability

Parent 1 Genotype	Parent 2 Genotype	Child Genotype	# Correct	# Errors	P-P-C Heritability Freq
AA	BB	AB	1	0	1
AA	AA	BB	0	1	0
AA	AB	BB	0	1	0
AA	No call	AB	0	0	NAN

Chapter 7

Performing LOH and Copy Number Analysis

Topics

Introduction

GenomeStudio 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 GenomeStudio 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 the GenomeStudio 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 plug-ins
  - CNV Analysis plug-ins
  - Column plug-ins
  - Report 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
= 0 if theta < tAA
= 0.5 * (theta - tAA) / (tAB - tAA) if theta < tAB
= 0.5 + 0.5 * (theta - tAB) / (tBB - tAB) if theta < tBB
= 1 if theta >= tBB

where:

- tAA = mean theta value of all genotypes in the AA cluster plotted in polar normalized coordinates
- tAB = mean theta value of all genotypes in the AB cluster plotted in polar normalized coordinates
- tBB = mean theta value of all genotypes in the BB cluster plotted in polar normalized coordinates

Figure 84 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.

B Allele Freq is set to NAN for loci included in the "IntensityOnly" category. These are markers such as non-polymorphic probes which do not provide genotypes, or SNP markers showing unusual clustering patterns during the standard clustering process.

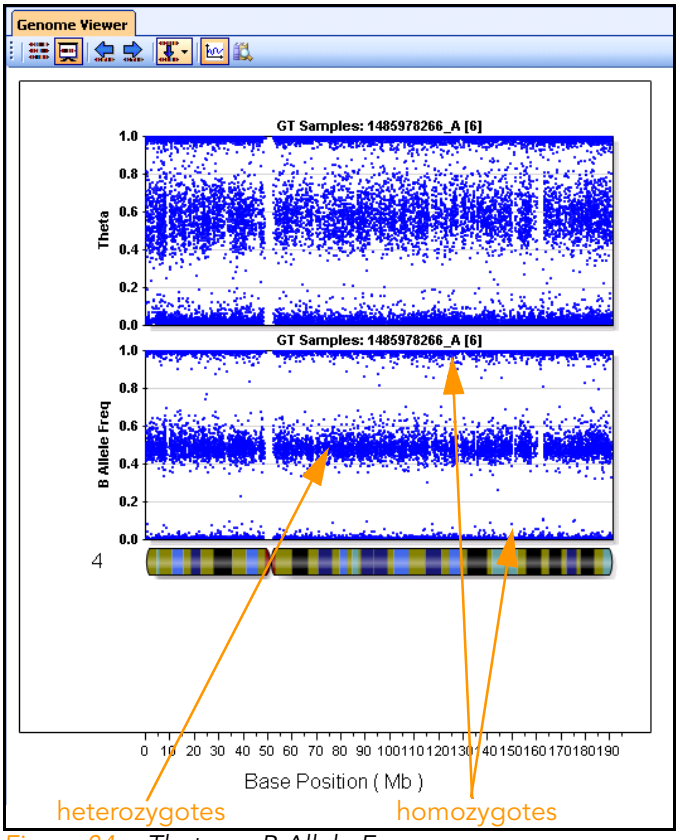


Figure 84 Theta vs. B Allele Frequency

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.

For loci included in GenomeStudio statistics such as Call Rate, 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.

Because no clusters are generated for loci in the "Intensity Only" category, the Log R Ratio for these loci is adjusted so that the expected R value is based on the weighted mean of the cluster itself. The Log R Ratio is displayed the same way for these loci as it is for loci included in GenomeStudio statistics in tools such as the IGV.

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 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 85 shows an example of a log R ratio plot.

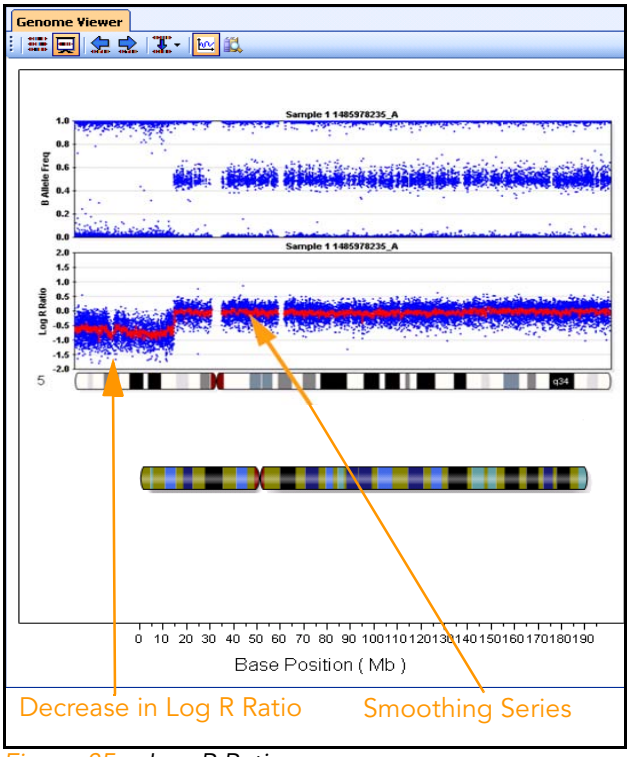


Figure 85 Log R Ratio

In Figure 85, a region of LOH is shown. This LOH event is 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.

CNV Analysis

GenomeStudio includes a CNV analysis workflow and related visualization tools which provide access to CNV algorithms and allow you to display algorithm results for all samples across the entire genome. CNV Analysis algorithms are provided as plug-ins by Illumina and our partners.

A CNV Analysis computes CNV Value and CNV Confidence for chromosomal regions in each sample. CNV Value usually represents an estimated copy number, while CNV Confidence is a relative score indicating confidence in the accuracy of the copy number estimate.

Creating a CNV Analysis

To create a CNV analysis:

1. Go to **Analysis** | **CNV Analysis**.
   The CNV Analysis dialog appears (Figure 86).

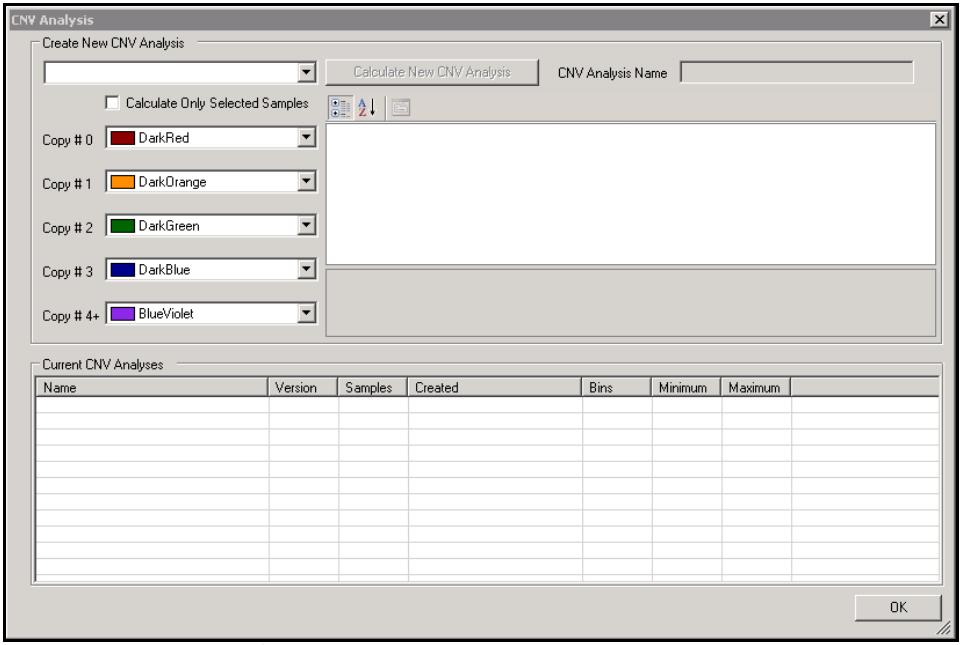


Figure 86 CNV Analysis

2. Select a CNV algorithm from the dropdown list.

You must have previously installed one or more CNV analysis plug-ins in order for them to appear in the dropdown list.

3. [Optional] Select the **Calculate Only Selected Samples** checkbox.
4. [Optional] Change the CNV Analysis name.
5. [Optional] Adjust the CNV Analysis input parameters.
6. Click **Calculate New CNV Analysis**.

   The CNV analysis begins, and a progress message appears.

   When the analysis is complete, the CNV Region Display appears (Figure 87).

   For more information about the CNV Region Display, see "Viewing the CNV Analysis Region Display" on page 109.
7. In the CNV Analysis dialog, click **OK**.

   The CNV Analysis dialog closes.

Selecting the Active CNV Analysis

To select the active CNV Analysis:

1. In the Current CNV Analyses area of the CNV Analysis dialog, select the CNV analysis you want to make active.
2. Click **OK**.

   The analysis you selected is now active.

   The active CNV analysis is the analysis used in the CNV Region Display and in the Full Data Table.

Deleting a CNV Analysis

To delete a CNV analysis:

- In the CNV Analysis dialog, right-click on the analysis you want to delete and select **Remove Analysis**.

  The analysis you selected is deleted from the list of available CNV analyses.

Viewing the CNV Analysis Region Display

The CNV Analysis Region Display is a heat map that shows copy number values for all samples across the genome. Samples are displayed on the X-axis and chromosomal position is displayed on the Y-axis.

To view the CNV Analysis Region Display:

1. In the GenomeStudio main window, select **Analysis** | **Show CNV Region Display**.

   The CNV Analysis Region Display appears (Figure 87).

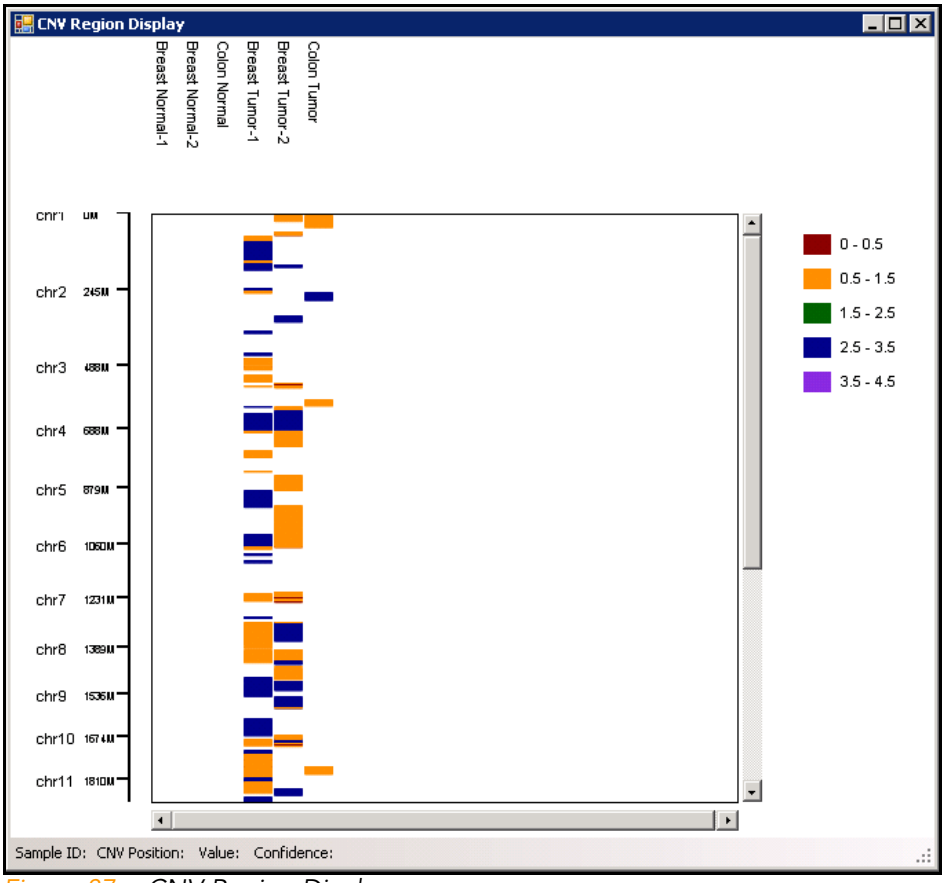


Figure 87 CNV Region Display

NOTE

The active CNV analysis appears in the CNV Analysis Region Display window.

The legend in the upper right of the CNV Region Display window shows the colors assigned to bins that represent copy number value ranges.

When you mouse over a region, information about that region displays in the status bar at the bottom of the window.

To view data at a higher resolution, use the mouse wheel to zoom in.

Viewing CNV Analysis Data in the Full Data Table

To view CNV analysis data in the Full Data Table:

1. In the Full Data Table, select **Column Chooser**.
   The Column Chooser dialog appears.
2. In the Hidden Subcolumns area, select **CNV Value** and **CNV Confidence**.
3. Click **Show**.
4. Click **OK**.
   The CNV Value and CNV Confidence Columns appear in the Full Data Table

CNV Value and CNV Confidence are calculated differently by each CNV algorithm. CNV Confidence may not be computed by some CNV algorithms.

Converting CNV Analysis Data into Bookmarks

To convert CNV analysis data into bookmarks:

1. In the IGV, select **View** | **CNV Analysis as Bookmarks**.
   The Display CNV Analysis dialog appears (Figure 88).

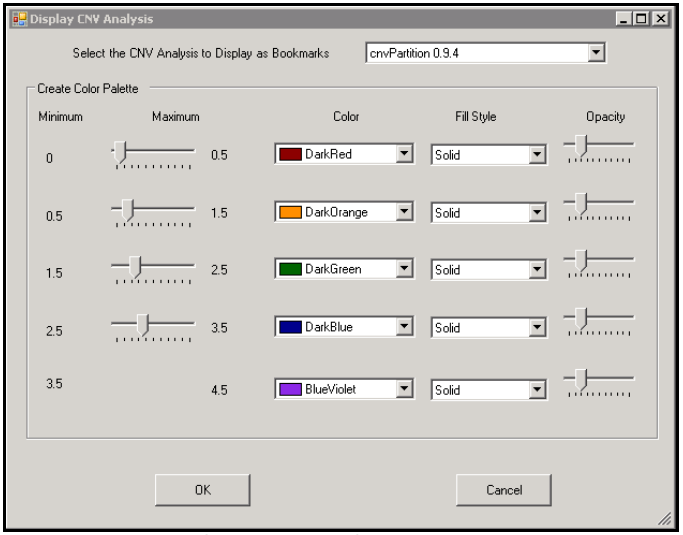


Figure 88 Display CNV Analysis

2. Select the CNV Analysis to display as bookmarks.

3. Click **OK**.

   The CNV analysis is converted into bookmarks and becomes the active bookmark analysis in the IGV and ICB.

Plug-ins

Illumina provides several types of plug-ins that you can use for LOH visualization, copy number analysis, or other types of analysis. Plug-ins are available from the GenomeStudio Portal. You can install one or more plug-ins after installing the GenomeStudio Framework and at least one software module.

- Autobookmarking plug-ins are external code libraries that create bookmarks in the IGV based on data that appears in GenomeStudio tables and on chromosomal position information. You can access autobookmarking plug-ins from the IGV Analysis menu.
- CNV Analysis plug-ins are external code libraries that create CNV Analyses in GenomeStudio. For more information about CNV analysis in GenomeStudio, see "CNV Analysis" on page 107.
- Column plug-ins are external code libraries that create new subcolumns based on data that appears in GenomeStudio tables. You can access column plug-ins by selecting Analysis | Create Plug-In Column from the GenomeStudio Genotyping Module main window.
- Report plug-ins are customized report formats provided by third parties.These plug-ins must be downloaded and installed in the correct directory before they are available in GenomeStudio.

Using Auto-bookmarking Plug-ins

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 GenomeStudio, select **Tools** | **Show Genome Viewer** to launch the IGV.

   The IGV appears, with the Add Favorite Data Plots form prominent (Figure 89).

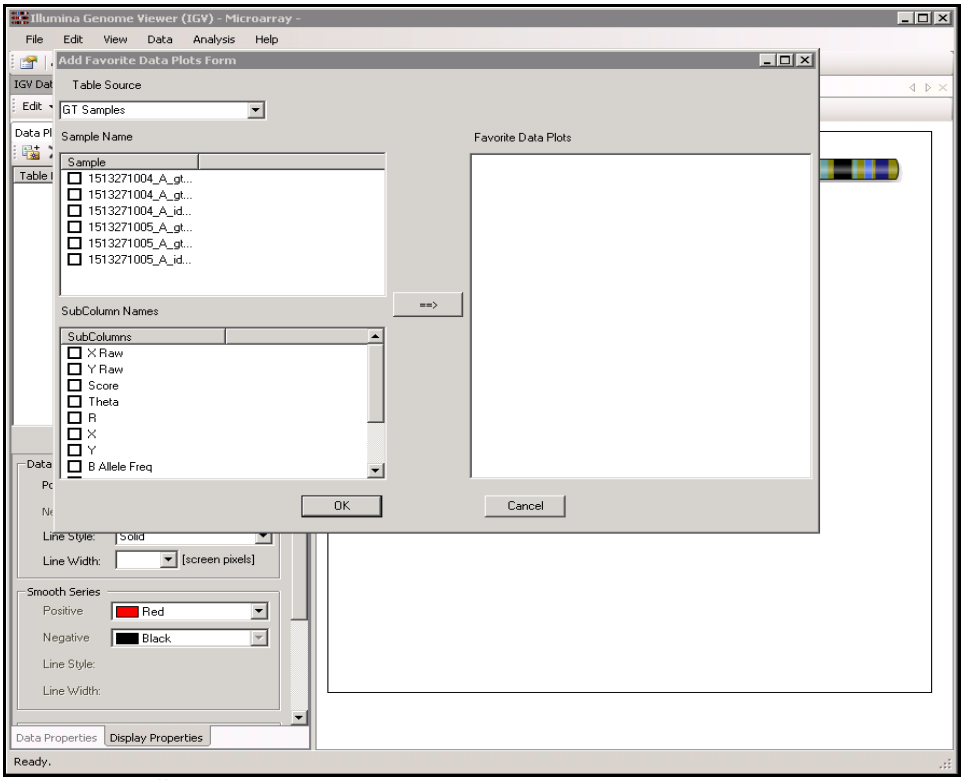


Figure 89 Illumina Genome Viewer

2. Select the data plots you want to view (Figure 90).

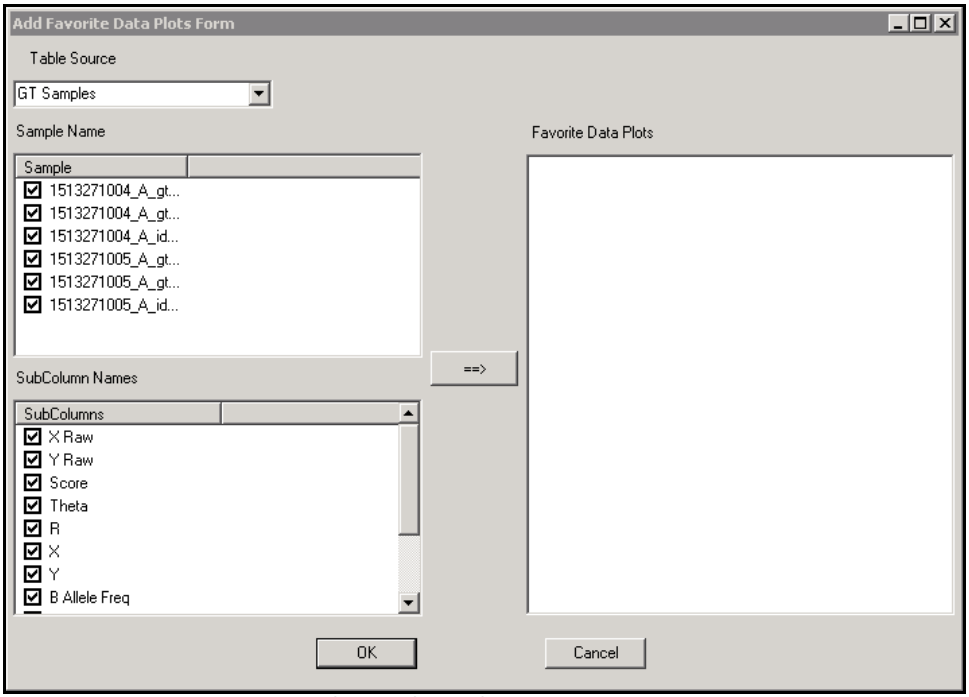


Figure 90 Favorite Data Plots Selected

3. Click **OK**.
   The IGV becomes prominent (Figure 91).

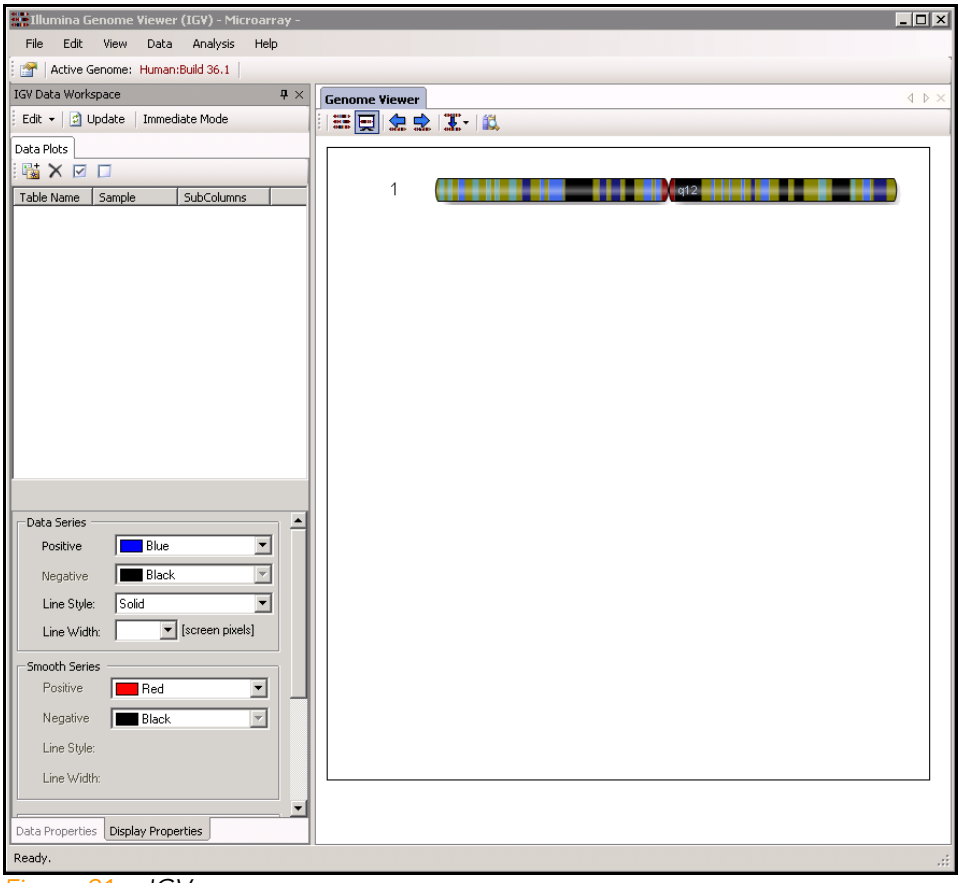


Figure 91 IGV

4. Select **Analysis** | **Run Autobookmark**.
   The **Autobookmark Analysis** dialog appears (Figure 92).

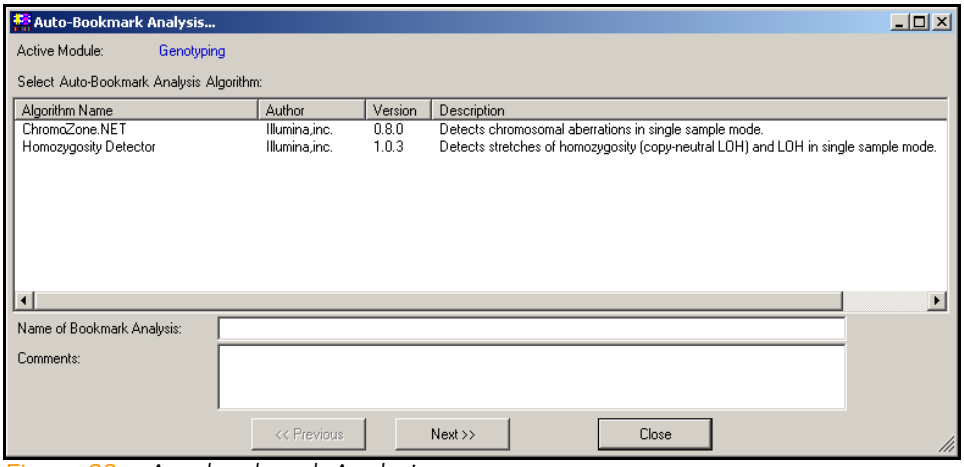


Figure 92 Autobookmark Analysis

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

5. Click an algorithm name to select an algorithm.
6. 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.

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

7. [Optional] Enter comments in the Comments text field.
8. Click **Next** to advance to the next dialog.
9. 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: Algorithm doesn't expose input parameters.

Continue to Step 8.

10. Click **Next**.

11. 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 93).

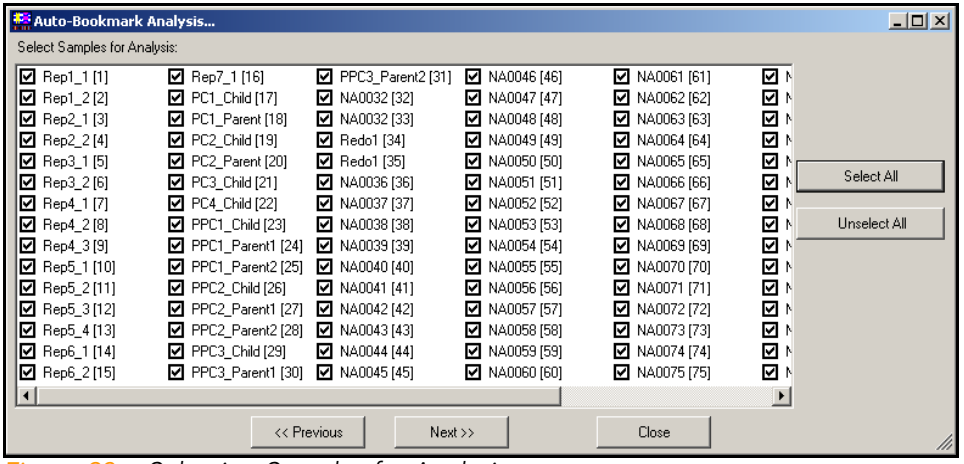


Figure 93 Selecting Samples for Analysis

12. Click **Next** to advance to the next dialog (Figure 94).

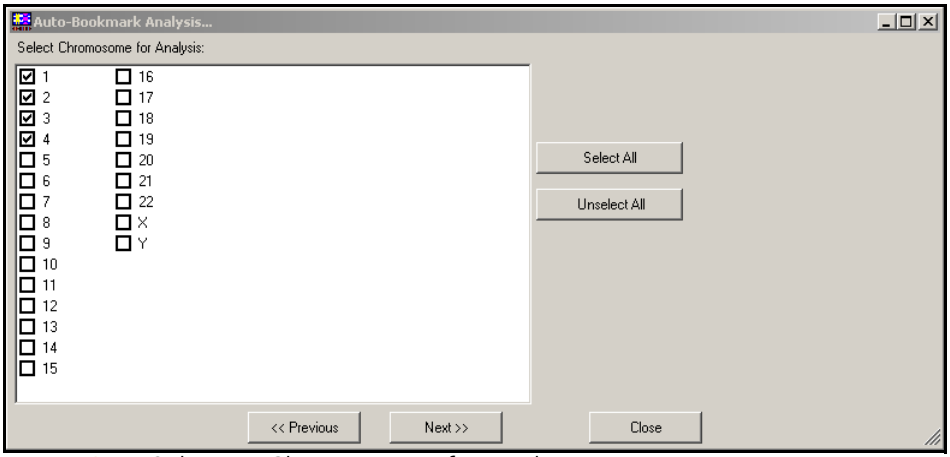


Figure 94 Selecting Chromosomes for Analysis

13. Select one or more chromosomes for analysis.

    You can select all chromosomes or any combination of chromosomes.

14. Click **Next** to advance to the next dialog (Figure 95).

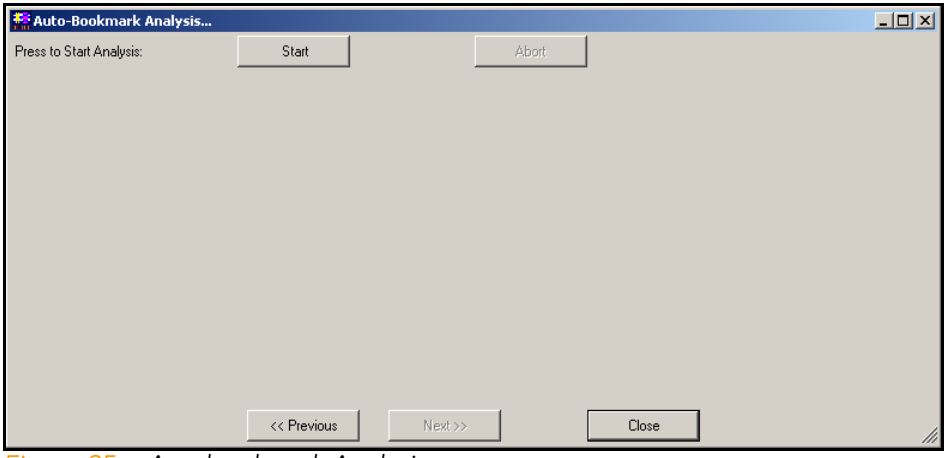


Figure 95 Autobookmark Analysis

15. 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.

16. When the analysis is complete, a message appears in the Algorithm Message Log (Figure 96).

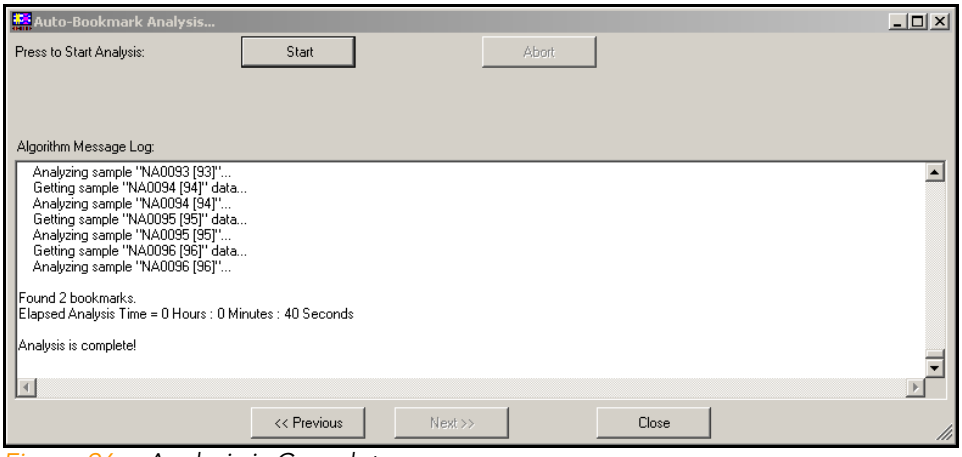


Figure 96 Analysis is Complete

17. Click **Close**.

    Bookmarks appear in the IGV, the ICB, the Bookmark Viewer, and the Full Data Table.

Using Column Plug-Ins

All column plug-ins are accessed and run through the GenomeStudio 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. In the GenomeStudio Genotyping Module main window, select **Analysis** | **Create Plug-In Column**.

   The Select Column Plug-In Form dialog appears (Figure 97).

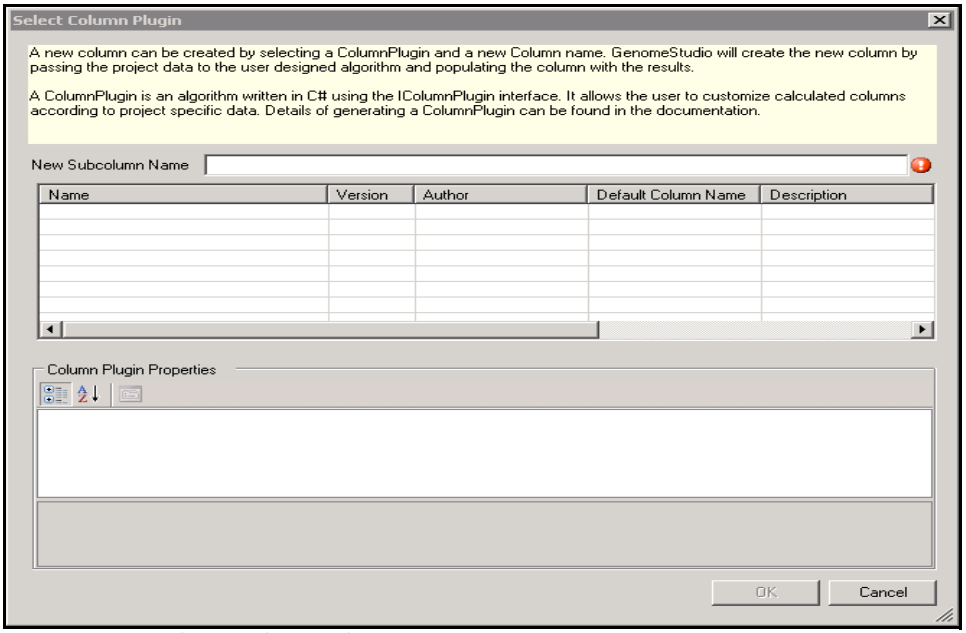


Figure 97 Select Column Plug-In Form

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
User Interface Reference

Topics

Introduction

The GenomeStudio Genotyping Module user interface provides tools for loading intensity files, running the clustering algorithm, browsing loci, and displaying them graphically. Figure 98 shows the default window configuration of the GenomeStudio Genotyping Module.

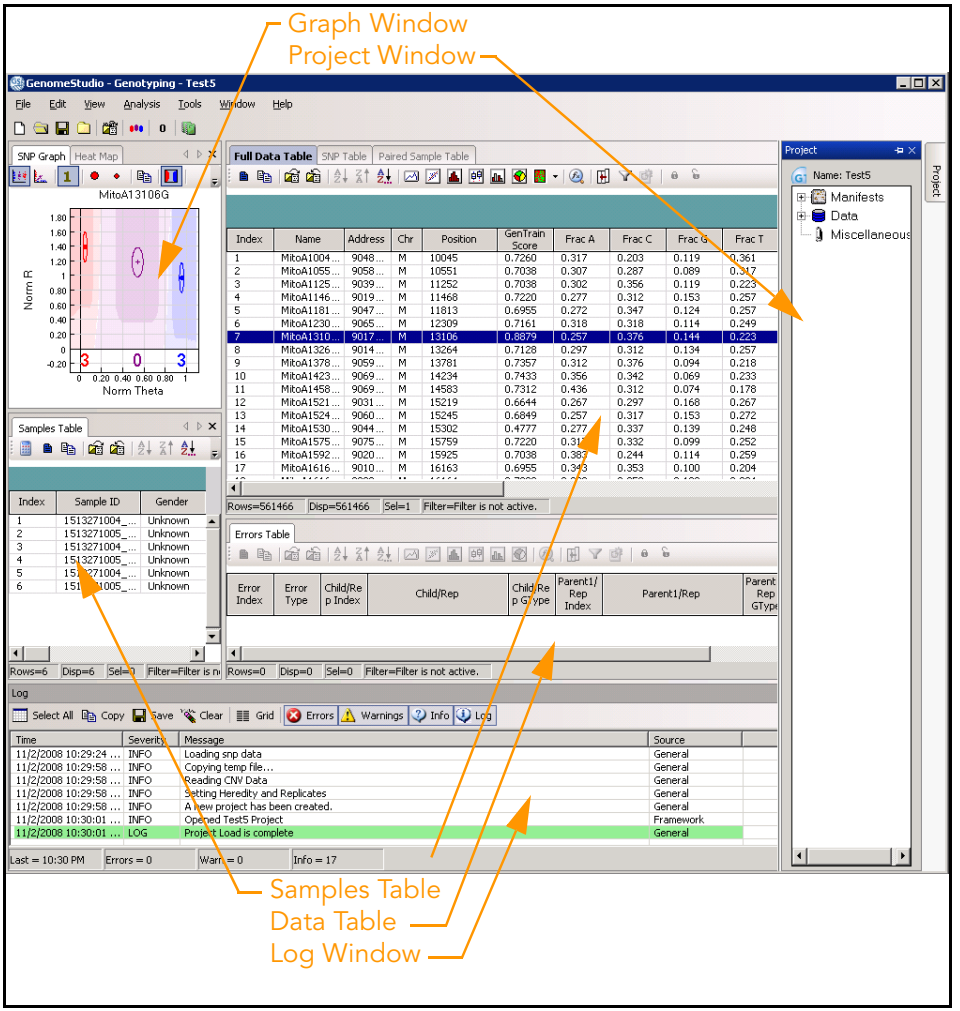


Figure 98 GenomeStudio Genotyping Module Default View

Detachable Docking Windows

Detachable docking windows provide a flexible way to customize GenomeStudio'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 99).

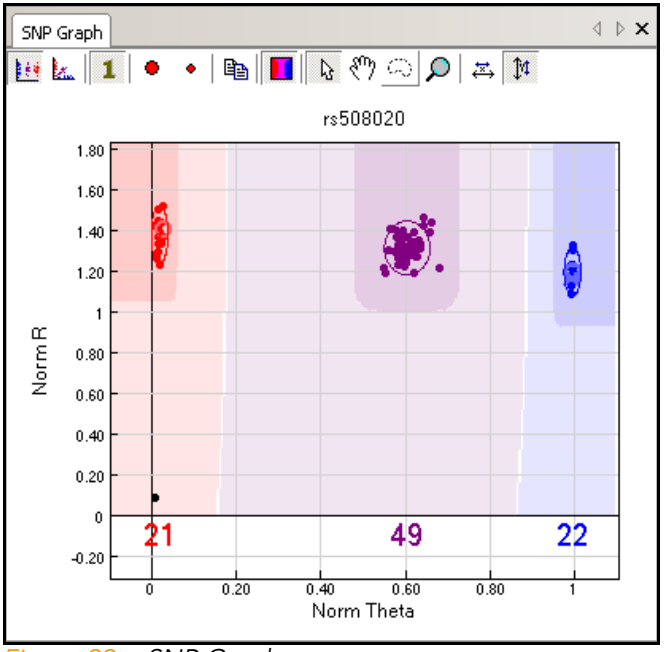


Figure 99 SNP Graph

Sample Graph

The Sample Graph (Figure 100) 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.

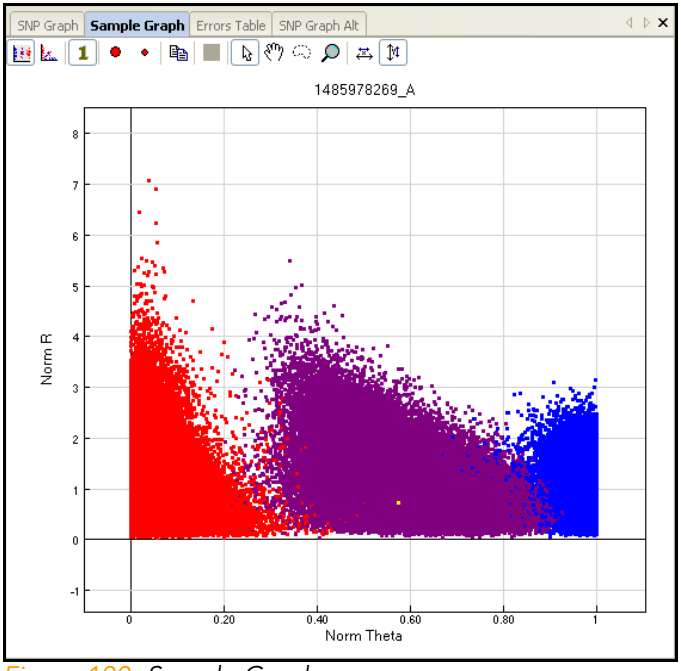


Figure 100 Sample Graph

Errors Table

The Errors Table (Figure 101) lists any reproducibility errors or parent-child heritability errors found in the data loaded into GenomeStudio.

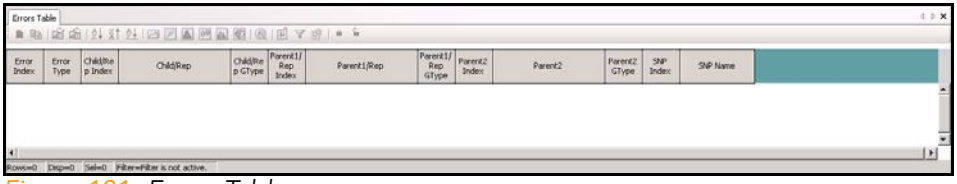


Figure 101 Errors Table

The columns in the Errors Table are listed and described in Table 10.

Table 10 Errors Table Columns

Column	Description	Type	Visible by Default?
Error Index	Row index of the error	integer	Y
Error Type	Type of error: Rep—Reproducibility P-C—Parent-Child heritability P-P-C—Parent-Parent-Child heritability	string	Y
Child/Rep Index	Sample index of the child sample involved in the error	integer	Y
Child/Rep	Sample ID of the child sample involved in the error	string	Y
Child/Rep GType	For a parental relationship error, the genotype of the child.	string	Y
Parent1/Rep Index	Sample index of the Parent1 sample involved in the error	integer	Y
Parent1/Rep	Sample ID of the Parent1 sample involved in the error	string	Y
Parent1/Rep GType	For a parental relationship error, the genotype of Parent1. For a replicate error, the genotype of replicate 1.	string	Y
Parent2 Index	Sample index of the Parent2 sample involved in the error	integer	Y
Parent2	Sample ID of the Parent2 sample involved in the error	string	Y
Parent2 GType	For a parental relationship error, the genotype of Parent2. For a replicate error, the genotype of replicate 2.	string	Y
SNP Index	Index number of the SNP where the error occurred.	integer	Y
SNP Name	Name of the SNP where the error occurred.	string	Y

SNP Graph Alt

The SNP Graph Alt (Figure 102) is an alternate SNP graph that you can display along with the SNP Graph to compare different views within GenomeStudio.

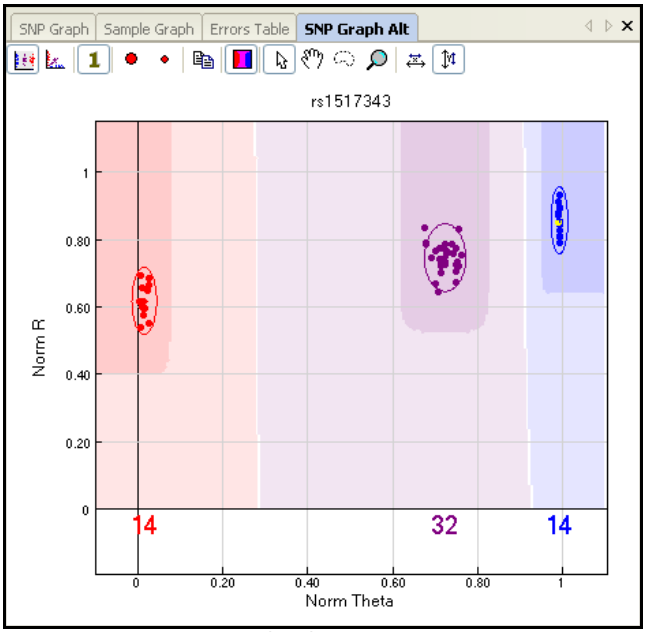


Figure 102 SNP Graph Alt

Data Table

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

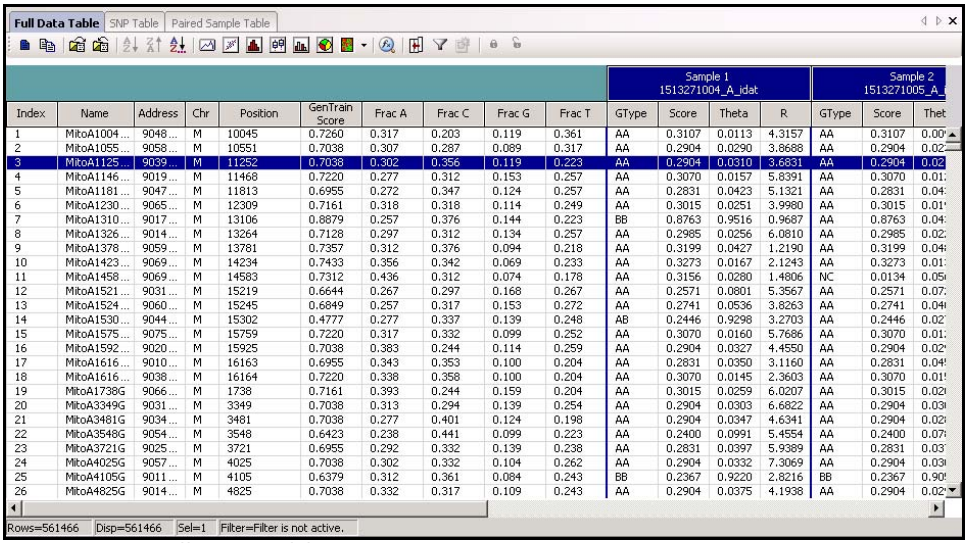
Full Data Table

The Full Data Table (Figure 103) 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.



Full Data Table | SNP Table | Paired Sample Table

										Sample 1 1513271004_A_idat				Sample 2 1513271005_A_		
Index	Name	Address	Chr	Position	GenTrain Score	Frac A	Frac C	Frac G	Frac T	GType	Score	Theta	R	GType	Score	Thet
1	MitoA1004...	9048...	M	10045	0.7260	0.317	0.203	0.119	0.361	AA	0.3107	0.0113	4.3157	AA	0.3107	0.00
2	MitoA1055...	9058...	M	10551	0.7038	0.307	0.287	0.089	0.317	AA	0.2904	0.0290	3.8688	AA	0.2904	0.02
3	MitoA1125...	9039...	M	11252	0.7038	0.302	0.356	0.119	0.223	AA	0.2904	0.0310	3.6831	AA	0.2904	0.02
4	MitoA1146...	9019...	M	11468	0.7220	0.277	0.312	0.153	0.257	AA	0.3070	0.0157	5.8391	AA	0.3070	0.01
5	MitoA1181...	9047...	M	11813	0.6955	0.272	0.347	0.124	0.257	AA	0.2831	0.0423	5.1321	AA	0.2831	0.04
6	MitoA1230...	9065...	M	12309	0.7161	0.318	0.318	0.114	0.249	AA	0.3015	0.0251	3.9980	AA	0.3015	0.01
7	MitoA1310...	9017...	M	13106	0.8879	0.257	0.376	0.144	0.223	BB	0.8763	0.9516	0.9687	AA	0.8763	0.04
8	MitoA1326...	9014...	M	13264	0.7128	0.297	0.312	0.134	0.257	AA	0.2985	0.0256	6.0810	AA	0.2985	0.02
9	MitoA1378...	9059...	M	13781	0.7357	0.312	0.376	0.094	0.218	AA	0.3199	0.0427	1.2190	AA	0.3199	0.04
10	MitoA1423...	9069...	M	14234	0.7433	0.356	0.342	0.069	0.233	AA	0.3273	0.0167	2.1243	AA	0.3273	0.01
11	MitoA1458...	9069...	M	14583	0.7312	0.436	0.312	0.074	0.178	AA	0.3156	0.0280	1.4806	NC	0.0134	0.05
12	MitoA1521...	9031...	M	15219	0.6644	0.267	0.297	0.168	0.267	AA	0.2571	0.0801	5.3567	AA	0.2571	0.07
13	MitoA1524...	9060...	M	15245	0.6849	0.257	0.317	0.153	0.272	AA	0.2741	0.0536	3.8263	AA	0.2741	0.04
14	MitoA1530...	9044...	M	15302	0.4777	0.277	0.337	0.139	0.248	AB	0.2446	0.9298	3.2703	AA	0.2446	0.02
15	MitoA1575...	9075...	M	15759	0.7220	0.317	0.332	0.099	0.252	AA	0.3070	0.0160	5.7686	AA	0.3070	0.01
16	MitoA1592...	9020...	M	15925	0.7038	0.383	0.244	0.114	0.259	AA	0.2904	0.0327	4.4550	AA	0.2904	0.02
17	MitoA1616...	9010...	M	16163	0.6955	0.343	0.353	0.100	0.204	AA	0.2831	0.0350	3.1160	AA	0.2831	0.04
18	MitoA1616...	9038...	M	16164	0.7220	0.338	0.358	0.100	0.204	AA	0.3070	0.0145	2.3603	AA	0.3070	0.01
19	MitoA1738G	9066...	M	1738	0.7161	0.393	0.244	0.159	0.204	AA	0.3015	0.0259	6.0207	AA	0.3015	0.02
20	MitoA3349G	9031...	M	3349	0.7038	0.313	0.294	0.139	0.254	AA	0.2904	0.0303	6.6822	AA	0.2904	0.03
21	MitoA3481G	9034...	M	3481	0.7038	0.277	0.401	0.124	0.198	AA	0.2904	0.0347	4.6341	AA	0.2904	0.02
22	MitoA3548G	9054...	M	3548	0.6423	0.238	0.441	0.099	0.223	AA	0.2400	0.0991	5.4554	AA	0.2400	0.07
23	MitoA3721G	9025...	M	3721	0.6955	0.292	0.332	0.139	0.238	AA	0.2831	0.0397	5.9389	AA	0.2831	0.03
24	MitoA4025G	9057...	M	4025	0.7038	0.302	0.332	0.104	0.262	AA	0.2904	0.0332	7.3069	AA	0.2904	0.03
25	MitoA4105G	9011...	M	4105	0.6379	0.312	0.361	0.084	0.243	BB	0.2367	0.9220	2.8216	BB	0.2367	0.90
26	MitoA4825G	9014...	M	4825	0.7038	0.332	0.317	0.109	0.243	AA	0.2904	0.0375	4.1938	AA	0.2904	0.02

Rows=561466 | Disp=561466 | Sel=1 | Filter=Filter is not active.

Figure 103 Full Data Table

The annotation columns of the Full Data Table are listed and described in Table 11.

Table 11 Full Data Table Columns

Column	Description	Type	Visible by Default?
Index	Row index of the SNP	integer	Y
Name	Name of the SNP	string	Y
Address	Bead-type identifier	integer	Y
Chr	Chromosome of the SNP	string	Y

Table 11 Full Data Table Columns (continued)

Column	Description	Type	Visible by Default?
Manifest	Name of the manifest to which the SNP belongs	string	N
Position	Chromosomal position of the SNP	integer	N
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 12.

Table 12 Full Data Table Per-Sample Subcolumns

Column	Description	Type	Visible by Default?
GType	Genotype of this SNP for the sample.	string	Y
Score	Call score of this SNP for the sample.	float	Y
Theta	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	Raw intensity of the A allele.	integer	N
Y Raw	Raw intensity of the B allele.	integer	N

Table 12 Full Data Table Per-Sample Subcolumns (continued)

Column	Description	Type	Visible by Default?
X	Normalized intensity of the A allele.	float	N
Y	Normalized intensity of the B allele.	float	N
B Allele Freq	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. B Allele Freq is linearly interpolated between 0 and 1, or set to NaN for loci categorized as "intensity only."	float	N
Log R Ratio	For loci included in GenomeStudio statistics: 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). For loci categorized as "intensity only": adjusted so that the expected R value is based upon the weighted mean of the cluster itself.	float	N
Top Alleles	Illumina-designated top strand genotype	string	N
Import Calls	Genotype calls for the given sample imported when the Import Allele Calls feature is used.	string	N
Concordance	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	Genotype call of SNP and sample at the time the project was originally clustered.	string	N
CNV Value	Estimate of copy number at individual locus	float	N
CNV Confidence	Level of confidence that the CNV value is correct, based on the CNV algorithm used	float	N

SNP Table

The SNP Table (Figure 104) shows statistics for each SNP.

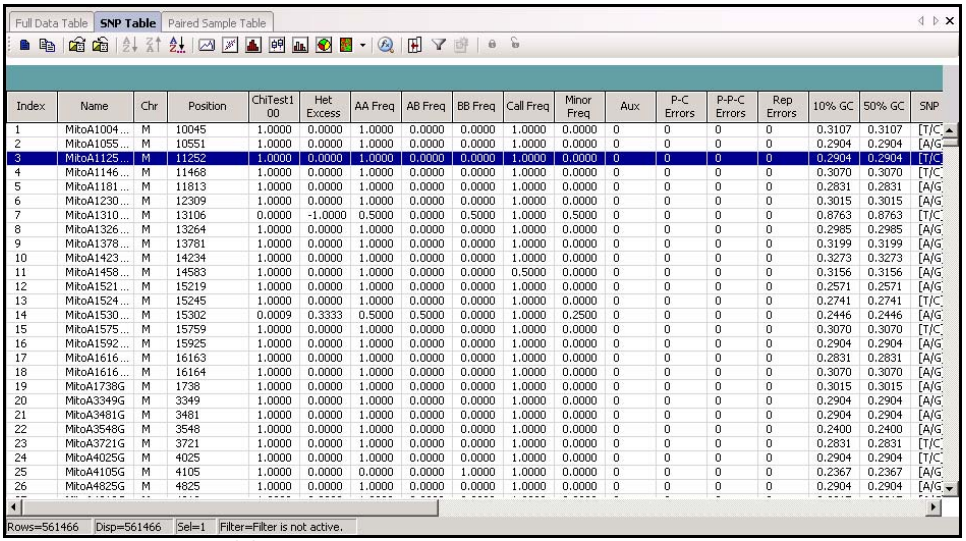


Figure 104 SNP Table

The SNP Table columns are listed and described in Table 13.

Table 13 SNP Table Columns

Column	Description	Type	Visible by Default?
Index	Row index of the SNP	integer	Y
Name	Name of the SNP	string	Y
Manifest	Manifest from which this SNP was loaded	string	N
Chr	Chromosome of the SNP	string	Y
Position	Chromosomal position of the SNP	integer	N
Address	Bead type identifier for this SNP	integer	Y
GenTrain Score	Measure of the cluster quality for the SNP	float	Y

Table 13 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
Orig Score	Original (unedited) GenTrain Score for SNP	float	Y
Edited	Flag indicating whether the SNP was edited after initial clustering positions were identified (1=> edited, 0=> unedited)	integer	Y
Cluster Sep	Measure of the cluster separation for the SNP that ranges between 0 and 1	float	Y
ChiTest 100	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	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	Frequency of AA calls	float	Y
AB Freq	Frequency of AB calls	float	Y
BB Freq	Frequency of BB calls	float	Y
Call Freq	Overall call frequency	float	Y
Minor Freq	Minor allele frequency If the number of AA < number of BB for a sample, the frequency for the minor allele A for that sample is (2*AAs + ABs) for the sample divided by (2*AAs + ABs + BBs) for the sample across all loci.	float	Y
Aux	User-set auxiliary value for the SNP	integer	Y
Rep Errors	Number of reproducibility errors for this SNP as allele comparisons between replicates.	integer	Y
P-C Errors	Number of parent-child heritability errors for the SNP compared among parent-child genotypes.	integer	Y

Table 13 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
P-P-C Errors	Number of parent-parent-child heritability errors for the SNP compared among parent-parent-child genotypes.	integer	Y
AA T Mean	Theta value of the center of the AA cluster, in normalized polar coordinates	float	Y
AA T Dev	Standard deviation in theta of the AA cluster, in normalized polar coordinates	float	Y
AB T Mean	Theta value of the center of the AB cluster, in normalized polar coordinates	float	Y
AB T Dev	Standard deviation in theta of the AB cluster, in normalized polar coordinates	float	Y
BB T Mean	Theta value of the center of the BB cluster, in normalized polar coordinates	float	Y
BB T Dev	Standard deviation in theta of the BB cluster, in normalized polar coordinates	float	Y
AA R Mean	R value of the center of the AA cluster, in normalized polar coordinates	float	Y
AA R Dev	Standard deviation in R of the AA cluster, in normalized polar coordinates	float	Y
AB R Mean	R value of the center of the AB cluster, in normalized polar coordinates	float	Y
AB R Dev	Standard deviation in R of the AB cluster, in normalized polar coordinates	float	Y
BB R Mean	R value of the center of the BB cluster, in normalized polar coordinates	float	Y
BB R Dev	Standard deviation in R of the BB cluster, in normalized polar coordinates	float	Y
SNP	Nucleotide substitution for the SNP on the Illumina top strand	string	N
ILMN Strand	Design strand designation	string	N

Table 13 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
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	Normalization ID for the SNP	integer	N
HW Equil	Hardy-Weinberg Equilibrium score for the SNP	float	N
Concordance	Measure of agreement between two genotypes from the same SNP locus	integer	N
CNV Region	SNPs and nonpolymorphic probes falling in known CNV regions. This column is automatically populated with information from the product manifest and may not be current because the number of known CNV regions is constantly changing. This column is for informational purposes only.	integer	Y
Exp Clusters	Number of expected clusters for a locus: 1 for nonpolymorphic probes 2 for mitochondrial DNA and Y loci 3 for any other loci This column is automatically populated with information from the product manifest. This column is for informational purposes only.	integer	Y

Table 13 SNP Table Columns (continued)

Column	Description	Type	Visible by Default?
Intensity Only	Indicates what type of information is available for a locus. 1 = Locus with intensity information only that is not included in GenomeStudio statistics such as Call Rate 0 = Locus with intensity and genotyping information that is included in GenomeStudio statistics such as Call Rate This column is automatically populated with information from the product manifest., but is also editable. This information has been determined based on HapMap samples and therefore may not apply to a different sample set of interest.	integer	Y

Paired Sample Table

The Paired Sample Table (Figure 105) shows statistics for paired samples.

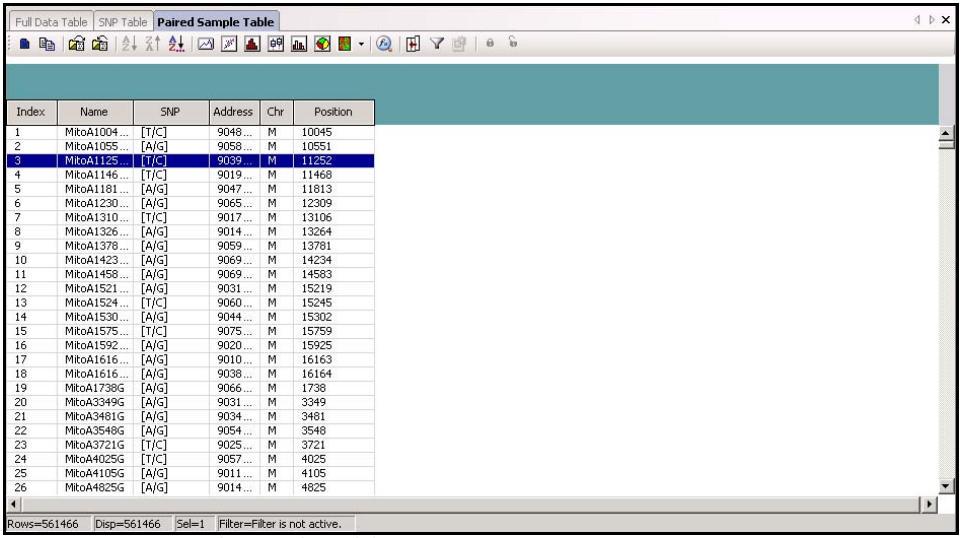


Figure 105 Paired Sample Table

The Paired Sample Table columns are listed and described in Table 14.

Table 14 Paired Sample Table Columns

Column	Description	Type	Visible by Default?
Index	Row index of the SNP	integer	Y
Name	Name of the SNP	string	Y
SNP	SNP	string	Y
Address	Bead-type identifier for the SNP	integer	Y
Chr	Chromosome of the SNP	string	Y
Position	Chromosomal position of the SNP	integer	N

The Paired Sample 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 Sample Table. The subcolumns are described in Table 17.

Table 15 Paired Sample 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
|dTheta sub-ref|	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

Table 15 Paired Sample Table Per-Pair Subcolumns (continued)

Column	Description	Type	Visible by Default?
|dAlleleFreq sub-ref|	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
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	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	Statistical confidence level between 0 and 1 indicating whether or not a copy number change has occurred. Values of approximately 1 indicate no copy number change. Values of approximately 0 indicate copy number change.	float	Y

Samples Table

The Samples Table (Figure 106) contains information for each DNA sample loaded into GenomeStudio. The Samples Table has the same column re-ordering properties as the SNP Table.

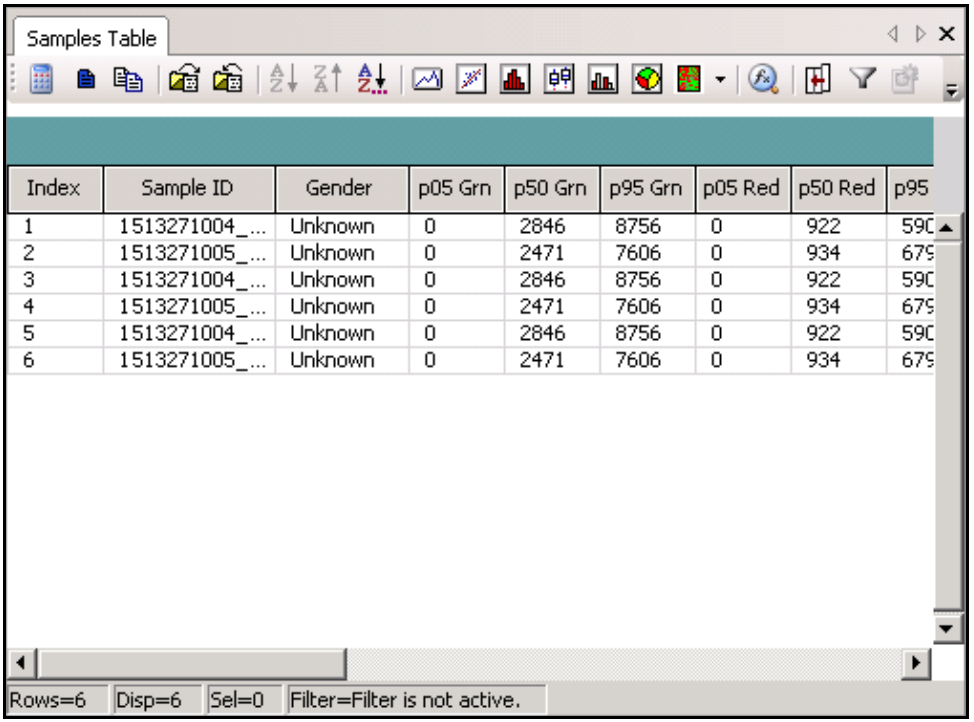


Figure 106 Samples Table

Table 16 Samples Table Columns

Column	Description	Type	Visible by Default?
Index	Row index of the sample	integer	Y
Sample ID	Sample identifier	string	Y
Gender	User-specified gender for the sample	string	Y
p05 Grn	5th percentile of A-allele intensity	integer	Y
p50 Grn	50th percentile of A-allele intensity	integer	Y
p95 Grn	95th percentile of A-allele intensity	integer	Y

Table 16 Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
p05 Red	5th percentile of B-allele intensity	integer	Y
p50 Red	50th percentile of B-allele intensity	integer	Y
p95 Red	95th percentile of B-allele intensity	integer	Y
p10 GC	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	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	Reproducibility error rate for this sample, calculated as 1 - sqrt(1 - errors/ 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	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	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 (expressed as a decimal) 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

Table 16 Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
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
Sample Name	Sample name	string	N
Sample Group	Sample group	string	N
Sample Plate	Sample plate	string	N
Sample Well	Well within the sample plate	string	N
Gender Est	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	Ethnicity of the individual from which this sample was acquired	string	N
Age	Age of the individual from which this sample was acquired	integer	N
Weight	Weight in kg of the individual from which this sample was acquired	string	N
Height	Height in meters of the individual from which this sample was acquired	string	N
Blood Pressure Systolic	Systolic blood pressure of the individual from which this sample was acquired	integer	N
Blood Pressure Diastolic	Diastolic blood pressure of the individual from which this sample was acquired	integer	N
Blood Type	Blood type of the individual from which this sample was acquired	string	N

Table 16 Samples Table Columns (continued)

Column	Description	Type	Visible by Default?
Phenotype Pos 1	Positive phenotype 1 of the individual from which this sample was acquired	string	N
Phenotype Pos 2	Positive phenotype 2 of the individual from which this sample was acquired	string	N
Phenotype Pos 3	Positive phenotype 3 of the individual from which this sample was acquired	string	N
Phenotype Neg 1	Negative phenotype 1 of the individual from which this sample was acquired	string	N
Phenotype Neg 2	Negative phenotype 2 of the individual from which this sample was acquired	string	N
Phenotype Neg 3	Negative phenotype 3 of the individual from which this sample was acquired	string	N
Comment	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
Tissue Source	Tissue source of the individual from which this sample was acquired	string	N
Calls	Number of loci on which this sample is being called	integer	N
No Calls	Number of loci on which this sample is not being called	integer	N
Excluded	1 = Sample is excluded 0 = Sample is included	integer	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 17.

Table 17 Samples Table Per-Manifest Subcolumns

Column	Description	Type	Visible by Default?
Sentrix ID	Barcode number of the Universal Array Product to which this sample was hybridized	string	Y
Sentrix Position	Section/bundle on the product	string	Y
Imaging Date	Date on which the product was scanned.	string	N
Scanner ID	ID of the scanner on which the product was scanned	string	N
PMT Green	Green PMT setting of the scanner on which the product was scanned	integer	N
PMT Red	Red PMT setting of the scanner on which the product was scanned	integer	N
Software Version	Version of the BeadScan software used to scan the product	string	N
User	User name of the person logged into the PC on which the product was scanned	string	N
p05 Grn	5th percentile of A-allele intensity	integer	N
p50 Grn	50th percentile of A-allele intensity	integer	N
p95 Grn	95th percentile of A-allele intensity	integer	N
p05 Red	5th percentile of B-allele intensity	integer	N
p50 Red	50th percentile of B-allele intensity	integer	N
p95 Red	95th percentile of B-allele intensity	integer	N

Table 17 Samples Table Per-Manifest Subcolumns (continued)

Column	Description	Type	Visible by Default?
p10 GC	10th percentile GenCall score over all SNPs for this sample. If displayed as 0.000, this column needs to be manually recalculated.	float	N
p50 GC	50th percentile GenCall score over all SNPs for this sample. If displayed as 0.000, this column needs to be manually recalculated.	float	N
Call Rate	Percentage of SNPs (expressed as a decimal) 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 155 of this manual.

Project Window

The Project window (Figure 107) identifies the manifest(s) loaded for your project and has a data section that identifies all of the Universal Array product barcodes used in your project. You can expand a barcode and view the samples loaded on that Universal 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.

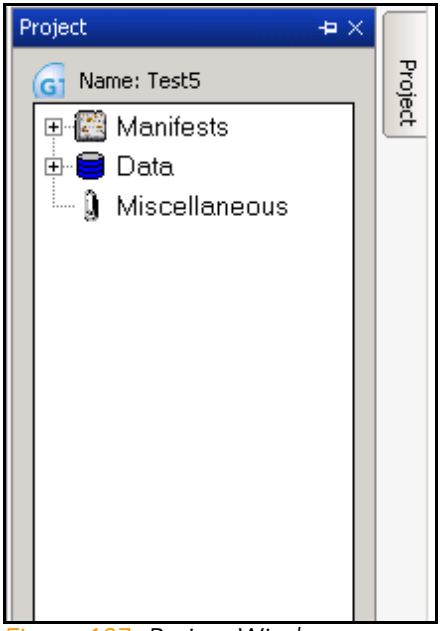


Figure 107 Project Window

Log Window

The Log window (Figure 108) is a simple console providing feedback on GenomeStudio processes. The Log window displays errors in red.

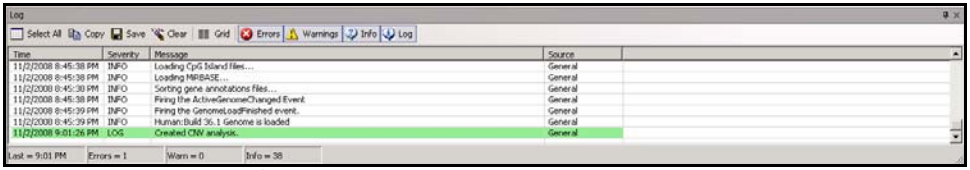


Figure 108 Log Window

Table 18 Log Window Options

Option	Function	Toolbar Button (if used)
Select All	Selects all log entries	
Copy	Copies log entries to the clipboard	
Save	Saves all log entries	
Clear	Clears all log entries	
Grid	Toggles the grid 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 GenomeStudio Genotyping Module's main window menus (and corresponding toolbar buttons).

Table 19 describes File Menu functions.

Table 19 File Menu Functions

Selection	Function	Toolbar Button (if used)
New Project	Opens a new project	
Open Project	Opens a previously saved project	
Save Project	Saves all current information in this project, so you can return to it later	
Save Project Copy As	Displays the **Save Project Copy As** dialog, 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 GenomeStudio 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.	

Table 19 File Menu Functions (continued)

Selection	Function	Toolbar Button (if used)
Export Cluster Positions	Allows you to export cluster position data to an *.egt file using the following options: ▸ 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.	
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, 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. Use this dialog to select print options for the currently displayed graph	
Recent Project	Allows you to select a project you have recently worked on	
Exit	Closes GenomeStudio	

Table 20 describes Edit Menu functions.

Table 20 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 21 describes View Menu functions.

Table 21 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 22 describes Analysis Menu functions.

Table 22 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 based on the samples in a project and determines the resulting genotype score for each locus. Clustering over-rides any cluster files that may have been used at project creation	
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	

Table 22 Analysis Menu Functions (continued)

Selection	Function	Toolbar Button (if used)
Reports	Allows you to create any of the following: ▸ Reproducibility and Heritability Report ▸ Final Report ▸ DNA Report ▸ Locus Summary Report ▸ Locus x DNA Report ▸ Custom Reports (if installed)	
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, 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, which allows you to select a directory from which to import allele calls	
Export Allele Calls	Displays the **Export Allele Calls** dialog, 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.	
Create Plug-in Column	Displays the **Select Column Plug-In Form** dialog, from which you can select an algorithm-based column plug-in. You can use the column plug-in to create a new subcolumn.	

Table 23 describes Tools Menu functions.

Table 23 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 | GenomeStudio	Opens the GenomeStudio **Options** window in which you can select GenomeStudio 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 24 describes Windows Menu functions.

Table 24 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 25 describes Help Menu functions.

Table 25 Help Menu Functions

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

Graph Window Toolbar

Table 26 lists GenomeStudio's Genotyping Module graph window toolbar buttons and their functions.

Table 26 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 representing an individual locus appear smaller on the screen.
	**Copy plot to clipboard**—Copies the current plot to the clipboard.
	**Shade call regions**—Applies colored shading to each cluster. ▸ 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.

Table 26 Graph Window Toolbar Buttons & Functions (continued)

Toolbar Button	Function(s)
	**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 27 lists and describes GenomeStudio's Genotyping Module Table Windows toolbar buttons and their functions.

Table 27 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 a line plot of the table data.
	**Scatter Plot**—Displays a scatter plot of the table data.
	**Histogram**—Displays a histogram of the table data.

Table 27 Table Windows Toolbar Buttons & Functions (continued)

Toolbar Button	Function(s)
	**Box Plot**—Displays a box plot of the table data.
	**Frequency Plot**—Displays a frequency plot of the table data.
	**Pie Chart**—Displays a pie chart of the table data.
	**Heat Map (Full Data Table only)**—Allows you to generate a new heat map or open an existing heat map.
	**New subcolumn**—Allows you to create a new subcolumn.
	Column Chooser—Displays the Column Chooser dialog box.
	**Filter Rows**—Displays the Filter Table Rows dialog box.
	**Clear Filter**—Removes the filter.

Context Menus

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

Table 28 describes graph window context menu selections.

Table 28 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.
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.

Table 28 Graph Window Context Menu (continued)

Selection	Description
Properties	Launches the **Graph Control Settings** dialog.

Table 29 describes Full Data Table context menu selections.

Table 29 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 30 describes SNP Table context menu selections.

Table 30 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.

Table 30 SNP Table Context Menu (continued)

Selection	Description
Clear Marks | <All>	Clears all marks.

Table 31 describes Samples Table context menu selections.

Table 31 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.
Sample Properties	Opens the **Sample Properties** dialog, 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.

Table 31 Samples Table Context Menu (continued)

Selection	Description
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 32 describes Error Table context menu selections.

Table 32 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

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 33 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 Universal Array Product that this sample was hybridized to for Manifest A.	R
SentrixPosition_A	The position within the Universal Array Product this sample was hybridized to for Manifest A (and similarly for _B, _C, etc. depending on how many manifests are used with your project).	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 33 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	• Figure 109 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.	

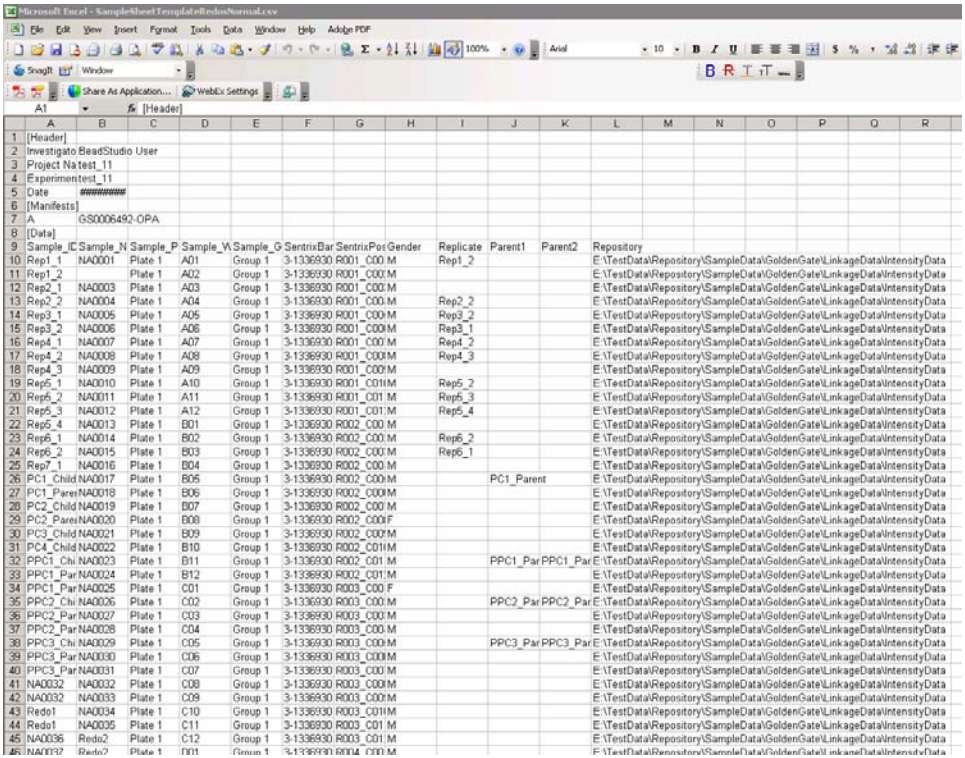
Redos and Replicates

Sample entries with the same Sample_ID are considered "redos" in the GenomeStudio 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 GenomeStudio 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 GenomeStudio CD. Use this template to create your own user-defined sample sheet.



	A	B	C	D	E	F	G	H	I	J	K	L
1	[Header]											
2	Investigato	BeadStudio User										
3	Project Na	test_11										
4	Experimen	test_11										
5	Date	########										
6	[Manifests]											
7	A	GS0006492-OPA										
8	[Data]											
9	Sample_ID	Sample_N	Sample_P	Sample_W	Sample_G	SentrixBar	SentrixPos	Gender	Replicate	Parent1	Parent2	Repository
10	Rep1_1	NA0001	Plate 1	A01	Group 1	3-1336930	R001_C00	M	Rep1_2			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
11	Rep1_2		Plate 1	A02	Group 1	3-1336930	R001_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
12	Rep2_1	NA0003	Plate 1	A03	Group 1	3-1336930	R001_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
13	Rep2_2	NA0004	Plate 1	A04	Group 1	3-1336930	R001_C00	M	Rep2_2			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
14	Rep3_1	NA0005	Plate 1	A05	Group 1	3-1336930	R001_C00	M	Rep3_2			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
15	Rep3_2	NA0006	Plate 1	A06	Group 1	3-1336930	R001_C00	M	Rep3_1			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
16	Rep4_1	NA0007	Plate 1	A07	Group 1	3-1336930	R001_C00	M	Rep4_2			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
17	Rep4_2	NA0008	Plate 1	A08	Group 1	3-1336930	R001_C00	M	Rep4_3			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
18	Rep4_3	NA0009	Plate 1	A09	Group 1	3-1336930	R001_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
19	Rep5_1	NA0010	Plate 1	A10	Group 1	3-1336930	R001_C01	M	Rep5_2			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
20	Rep5_2	NA0011	Plate 1	A11	Group 1	3-1336930	R001_C01	M	Rep5_3			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
21	Rep5_3	NA0012	Plate 1	A12	Group 1	3-1336930	R001_C01	M	Rep5_4			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
22	Rep5_4	NA0013	Plate 1	B01	Group 1	3-1336930	R002_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
23	Rep6_1	NA0014	Plate 1	B02	Group 1	3-1336930	R002_C00	M	Rep6_2			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
24	Rep6_2	NA0015	Plate 1	B03	Group 1	3-1336930	R002_C00	M	Rep6_1			E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
25	Rep7_1	NA0016	Plate 1	B04	Group 1	3-1336930	R002_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
26	PC1_Child	NA0017	Plate 1	B05	Group 1	3-1336930	R002_C00	M		PC1_Parent		E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
27	PC1_Paren	NA0018	Plate 1	B06	Group 1	3-1336930	R002_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
28	PC2_Child	NA0019	Plate 1	B07	Group 1	3-1336930	R002_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
29	PC2_Paren	NA0020	Plate 1	B08	Group 1	3-1336930	R002_C00	F				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
30	PC3_Child	NA0021	Plate 1	B09	Group 1	3-1336930	R002_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
31	PC4_Child	NA0022	Plate 1	B10	Group 1	3-1336930	R002_C01	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
32	PPC1_Chi	NA0023	Plate 1	B11	Group 1	3-1336930	R002_C01	M		PPC1_Par	PPC1_Par	E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
33	PPC1_Par	NA0024	Plate 1	B12	Group 1	3-1336930	R002_C01	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
34	PPC1_Par	NA0025	Plate 1	C01	Group 1	3-1336930	R003_C00	F				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
35	PPC2_Chi	NA0026	Plate 1	C02	Group 1	3-1336930	R003_C00	M		PPC2_Par	PPC2_Par	E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
36	PPC2_Par	NA0027	Plate 1	C03	Group 1	3-1336930	R003_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
37	PPC2_Par	NA0028	Plate 1	C04	Group 1	3-1336930	R003_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
38	PPC3_Chi	NA0029	Plate 1	C05	Group 1	3-1336930	R003_C00	M		PPC3_Par	PPC3_Par	E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
39	PPC3_Par	NA0030	Plate 1	C06	Group 1	3-1336930	R003_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
40	PPC3_Par	NA0031	Plate 1	C07	Group 1	3-1336930	R003_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
41	NA0032	NA0032	Plate 1	C08	Group 1	3-1336930	R003_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
42	NA0032	NA0033	Plate 1	C09	Group 1	3-1336930	R003_C00	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
43	Redo1	NA0034	Plate 1	C10	Group 1	3-1336930	R003_C01	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
44	Redo1	NA0035	Plate 1	C11	Group 1	3-1336930	R003_C01	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData
45	NA0036	Redo2	Plate 1	C12	Group 1	3-1336930	R003_C01	M				E:\TestData\Repository\SampleData\GoldenGate\LinkageData\IntensityData

Figure 109 Sample Sheet Example

Appendix B

Troubleshooting Guide

Topics

Introduction

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

Frequently Asked Questions

Table 34 lists frequently asked questions and associated responses.

Table 34 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.