Affymetrix® GeneChip® Targeted Genotyping Analysis Software User's Guide

Version 1.6

P/N 702126 Rev. 3

For research use only.
Not for use in diagnostic procedures.

Contents

Chapter 1

PREFACE

How to Use This Guide

Purpose of This Guide

The *Affymetrix GeneChip® Targeted Genotyping Analysis Software User Guide* provides you with instructions on how to:

- Import and process the data collected from Affymetrix GeneChip® Universal Tag Arrays
- Generate genotypes
- Export genotypes

Audience

This guide is intended for individuals who will:

- Routinely process and perform quality control checks on array data
- Generate the final genotypes for a study

User Attention Words

Two user attention words appear in this document. Each word implies a particular level of observation or action as described below:

NOTE: **Provides information that may be of interest or help, but is not critical to the use of the product.**

IMPORTANT: **Provides information that is necessary for proper software operation.**

About This Software

Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS) provides a full set of tools designed to help you generate and manage the highest quality genotypes using the Affymetrix Targeted Genotyping System. It allows you to:

- Administer users, array definitions, and assay panel files, and protocols.
- Design experiments and track samples through the genotyping process.

- Generate genotypes from the data collected by the Affymetrix GeneChip® Scanner 3000 7G 4C (Scanner 3000 7G 4C.)
- Manage the processed data and genotypes stored in the TG Post-Amp Workstation database

Chapter 2

About Affymetrix® GeneChip® Targeted Genotyping Analysis Software

Genotyping with Affymetrix GeneChip® Targeted Genotyping Analysis Software

Process Overview

Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS) is used to process the .cel file data collected from Affymetrix GeneChip® Universal Tag Arrays (Universal Tag Arrays or arrays) and generate genotypes. A high-level summary of the workflow from assay to genotype is as follows.

1. Perform the Targeted Genotyping Protocol
2. Hybridize samples onto arrays
3. Scan the arrays and collect data
4. Import data into GTGS
5. Process the data
6. Assign a status of pass or fail to each experiment (one experiment = one array = one sample)
7. Generate genotypes

Scanning Arrays

What Happens When Arrays are Scanned

As part of the Targeted Genotyping Protocol, processed samples are hybridized onto arrays. After hybridization, the arrays are stained and washed. During this process, four different dyes are added to each sample – one for each nucleotide. Once stained and washed, the arrays are loaded onto the Affymetrix GeneChip® Scanner 3000 7G 4C (Scanner 3000 7G 4C).

The laser scans each array four times to sequentially measure the fluorescence from each dye. Data from the scans is stored in four image files referred to as .dat files. (the filenames for these images include the extension .dat). For GCOS users, the .dat files are accessible from the Image Data folder on the Instrument Control Workstation that controls the scanner (Figure 2.1). For AGCC users, .dat files can be inspected using Command Console Viewer.

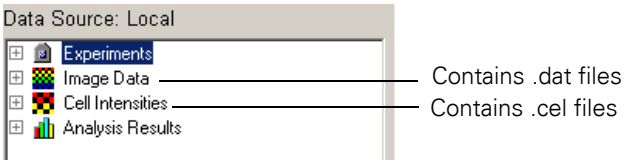


Figure 2.1 GCOS location of .dat and .cel Files on the Instrument Control Workstation

When opened, .dat files look similar to the example shown in Figure 2.2.

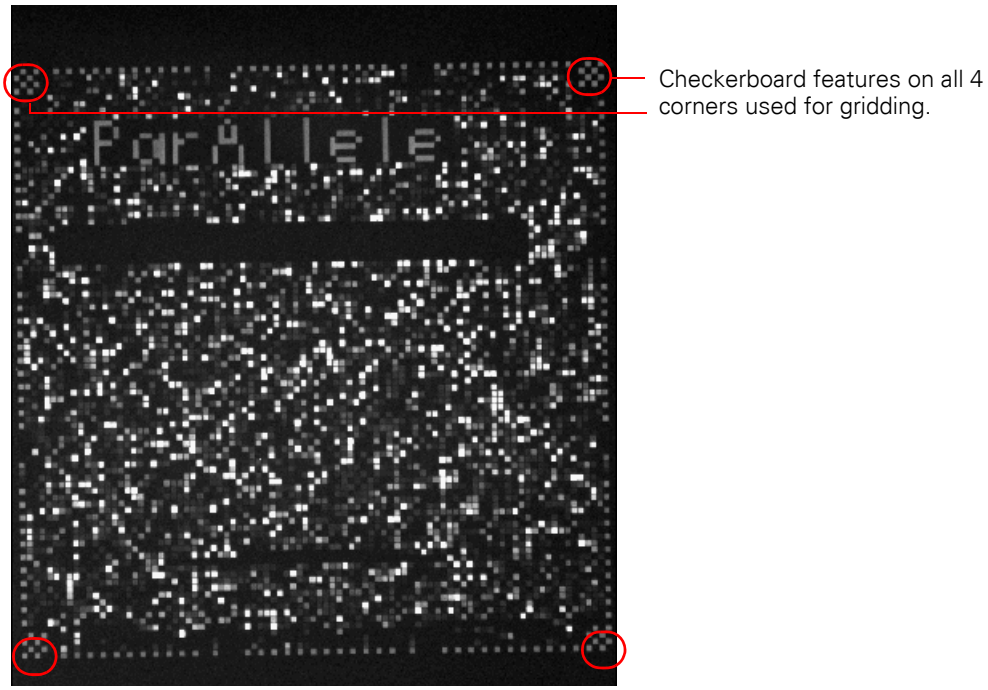


Figure 2.2 Example of a .dat File

Each .dat file contains the data for one channel:

Table 2.1 DAT File Channel Designations

.dat File Channel Designation	Channel	Nucleotide
A	A	Adenine
B	C	Cytosine
C	G	Guanine
D	T	Thymine

Once a .dat file has been created, AGCC or GCOS automatically converts it into a .cel file, also referred to as an intensity file (Figure 2.3).

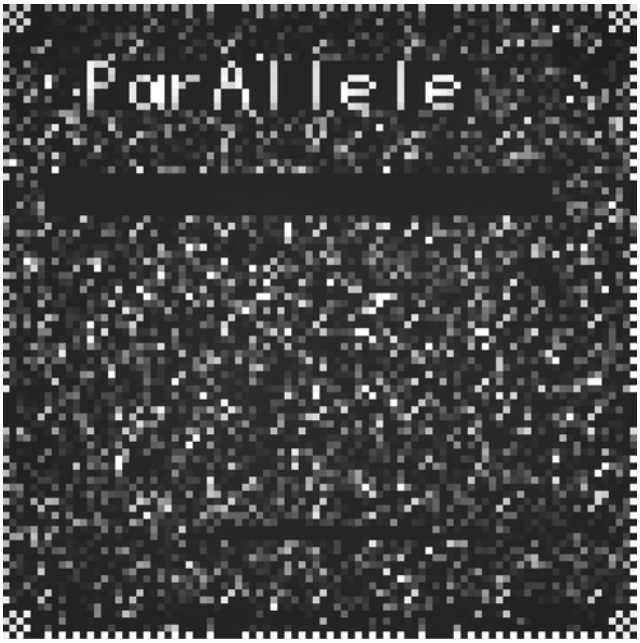


Figure 2.3 Example of a .cel File

Each square in .dat and .cel images is generically referred to as a feature. Two basic types of features exist on every array:

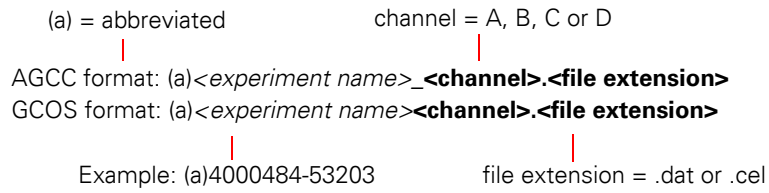
- Assay features (used for genotyping)
- Control features (used for gridding and other functions; shown in Figure 2.2)

The process of converting .dat to .cel files is referred to as gridding. A grid is superimposed on the .dat array image such that each element of the grid brackets a feature. Once the grid has been applied, AGCC or GCOS computes one intensity value for each cell, and saves these values in a .cel file. Thus, while .cel files look very similar to .dat files, the file size is much smaller.

Only the .cel files are read by GTGS. The data contained in these files is used to generate genotypes.

File Naming Convention for .dat and .cel Files

The file naming convention for .dat and .cel files is similar. The experiment name, generated by the tracking component of GTGS, is an abbreviated version of the array barcode (the last digits):



AGCC .cel File Example — Four .cel files for each array

(a)4000484-53203_A.cel
(a)4000484-53203_B.cel
(a)4000484-53203_C.cel
(a)4000484-53203_D.cel

GCOS .cel File Example — Four .cel files for each array

(a)4000484-53203A.cel
(a)4000484-53203B.cel
(a)4000484-53203C.cel
(a)4000484-53203D.cel

About .dat and .cel File Storage

The .dat and .cel files are usually stored on the Instrument Control Workstation.

IMPORTANT: **We strongly recommend that you periodically back up .dat and .cel files to a different computer. For AGCC users, copy or move data files directly from the file system folders used by Command Console. For GCOS users, use the GCOS Data Transfer Tool to copy files to another location.**

For more information on data storage, see Chapter 7, *Data Storage and Server Maintenance*.

Array Data Processing and Cluster Genotyping

About Experiments

An experiment is a test that utilizes one SNP assay panel to genotype one sample. An assay panel is a group of assays. Only one experiment is hybridized onto each Universal Tag Array. As described under *Scanning Arrays* on page 9, the data collected while each array is scanned is stored in .dat files which are then used to generate .cel files.

GTGS reads and processes the array data stored in .cel files in order to measure data quality and generate genotypes. Besides the array data, GTGS also maintains basic tracking information for each experiment including the sample name and processing history.

Workflow Overview from Assay to Genotype

The diagram below shows:

- The software modules used throughout the workflow
- The purpose of each module

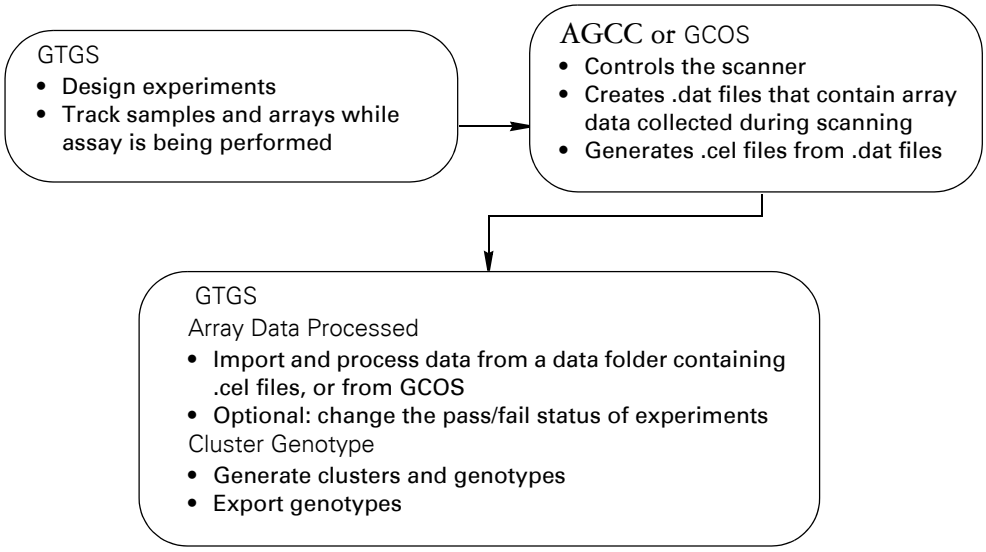


Figure 2.4 Workflow from Assay to Genotype

GTGS first reads the array data stored in the .cel files. The software processes the data to normalize it and to measure its quality against a set of quality control metrics. As a result of this initial processing, each experiment is assigned a status: Pass or Fail. The results of this operation are stored in the Experiment QC Summary table (Figure 2.5).

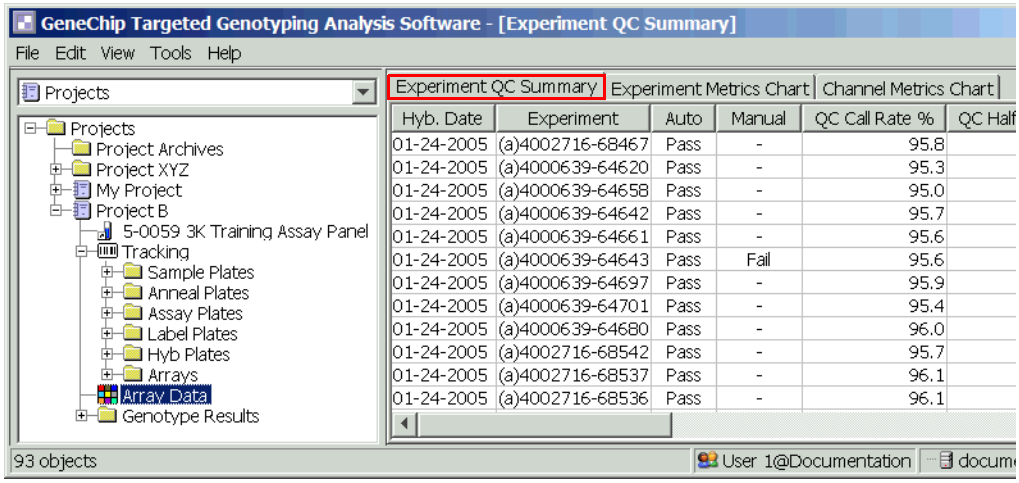


Figure 2.5 Experiment QC Summary Table

All experiments with the status Pass are included in the next operation, cluster genotyping.

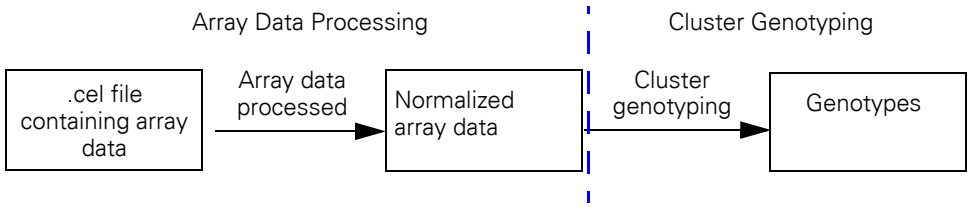


Figure 2.6 Workflow from Array Data Processing to Cluster Genotyping

The cluster genotyping operation groups the normalized assay data from each experiment based on its signal distribution. Each assay is clustered independently as shown in Figure 2.7 below.

Based upon the overall signal distribution of the samples clustered, one to three clusters can be identified:

- Homozygous for allele 1
- Heterozygous
- Homozygous for allele 2

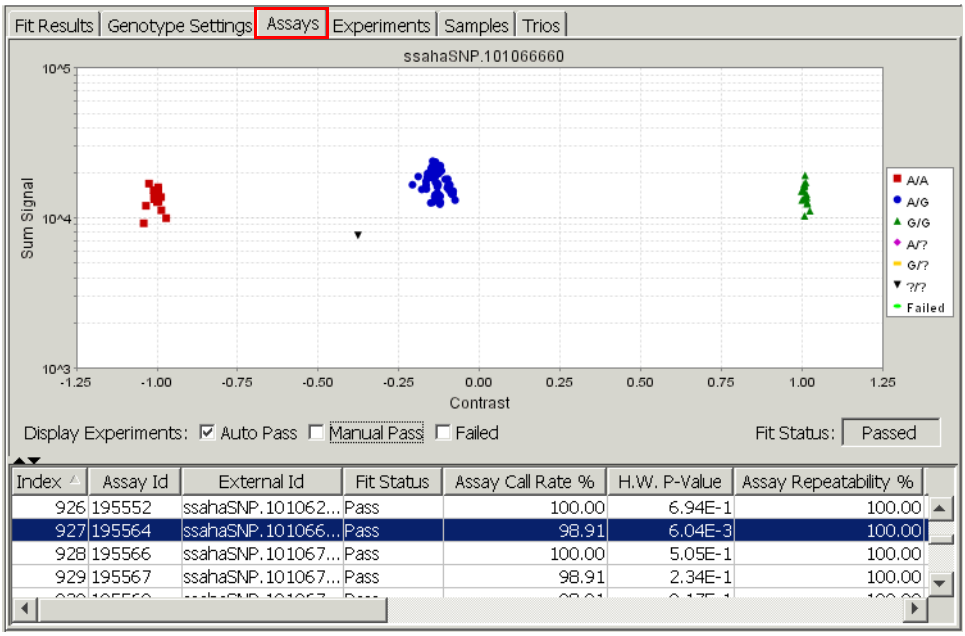


Figure 2.7 Cluster Example

Genotypes are assigned automatically based on the degree to which the data from each assay belongs to a particular cluster. The genotypes are then saved to the GTGS database, and can be exported as a file.

Data from multiple experiments are required for clustering. The fraction of passing assays, overall data completeness, and data accuracy increases as more experiments from unique samples are clustered. Therefore, to ensure that quality genotypes are called, we recommend collecting data from at least 80 experiments using unique samples before clustering.

About Projects

How Projects Are Used and Organized

All of the information related to a particular study is stored in what is referred to as Project. In GTGS, projects are identified by a *Notebook* icon (Figure 2.6).

Projects are used to:

- Store the sample tracking information, array data and genotype results for samples that have been tested with a specific assay panel
- Group the experiments that you want to genotype as a set

Data from individual experiments run with the same assay panel can be grouped and genotyped as a set. As shown in the figure below, you can do this by:

1. Creating a top-level folder.
2. Creating multiple projects within that folder with the same assay panel designation.
3. Grouping select experiments in each project.

For instructions on how to create a project, refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide*.

In this example, Project XYZ has two subfolders: Phase I and Phase II.
The Phase I folder contains two projects: Study A and Study B.

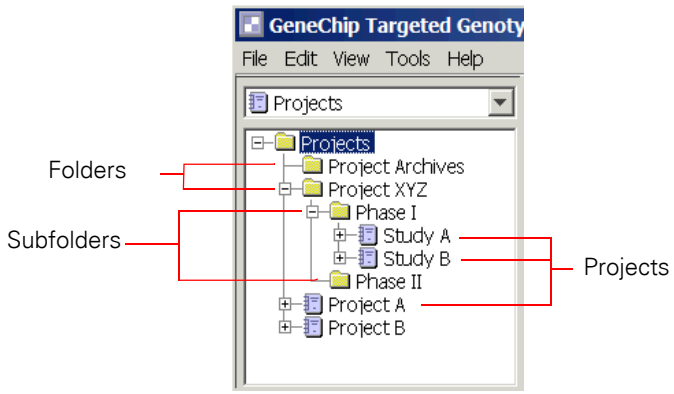


Figure 2.8 Creating Folders and Projects

Assay Panel Icon

The assay panel icon (test tube; Figure 2.9) indicates the assay panel associated with a particular project. The name associated with the assay panel icon is the filename of the the assay panel that was imported into GTGS when the project was created.

All experiment data in a particular project must be associated with the same assay panel. For more information on assay panel files, see ***About Assay Panels*** in the following section.

If you select the assay panel icon, information about the panel is displayed in the right pane. This information includes:

- The type of array that must be used when running samples with this particular assay panel (Properties tab)
- Information on each assay in the panel including the chromosome and position within the chromosome.

Sample and Sample Tracking Information

Within each project, sample and sample tracking and data analysis information can be viewed in a variety of ways. Depending upon where you are when running the protocol or processing data, sample information can be viewed by selecting:

- The Tracking folder
- Any of the plate folders within the Tracking folder
- The Array Data icon
- Any of the cluster icons in the Genotype Results folder (Figure 2.9).

While the various stages of the Targeted Genotyping Protocol are being performed, samples are moved from one plate to another, then to arrays. Each type of plate is organized in a subfolder with a corresponding name. See *About Sample and Sample Tracking Information* on page 21 for more information.

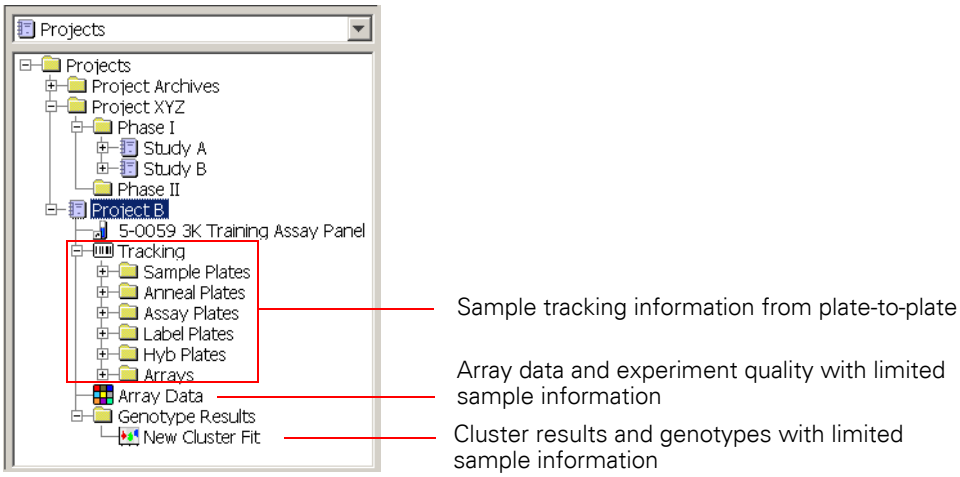


Figure 2.9 Project Icons and Folders

Data Quality Information

By clicking the Array Data icon, you can view various tables and charts that provide summary and detailed information on the data quality of each experiment. These tables and charts are generated prior to cluster genotyping, when the data is first imported. Information you can view includes:

- Experiment QC Summary table
- Experiment Metrics Charts
- Channel Metrics Charts

You can view the details for individual experiments, including an experiment summary, channel details, and array view. To view, click the Array Data icon, then right-click a specific experiment in the Experiment QC Summary table and select View Experiment Details. For more information on these tables and charts, see Chapter 4, *About Charts and Tables*.

Genotyping Information

Cluster genotyping results are accessed via the Genotype Results folder (Figure 2.9). If cluster genotyping has been performed on data for a particular project, the Genotype Results folder will contain at least one cluster icon. Results include tables and graphs that summarize the quality of the genotyping. Each time you cluster data for a particular project, all of the project experiments with the status Passed are genotyped.

Renaming and Moving Folders and Projects

You can also rename and move folders and projects.

To move folders or projects:

1. Right-click the project or folder name.
2. Select **Move Project …** (or **Move Folder …**) as appropriate.
3. Select the new location in the Move Project window, then click **OK**.

To rename folders or projects:

1. Right-click the project or folder name.
2. Select **Rename**.
3. Enter a new name in Rename window, then click **OK**.

About Assay Panels

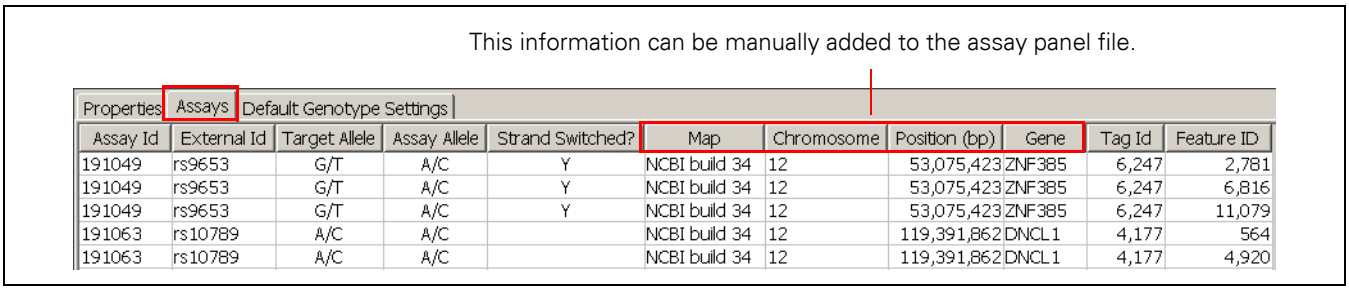
Viewing Assay Panel Information

To view information about an assay panel:

1. Open a project and select the assay panel name.
2. Click any of the tabs in the right pane (described below.)

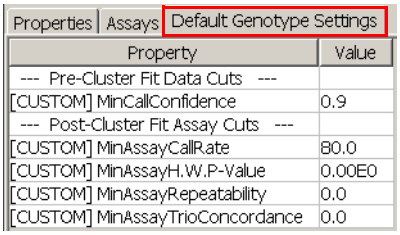
   The information available for assay panels is organized under three tabs:

   - **Properties tab**: Includes the assay panel name, description, and type of array to be used with that panel.
   - **Assays tab**: Includes the information shown Figure 2.10 including Assay ID, External ID, and Target Allele. You can add map, chromosome, position (bp) and gene information to the assay panel file and import it into GTGS at any time. See ***Adding or Updating Map, Chromosome, Position and Gene Details*** on page 20 for instructions.
   - **Default Genotype Settings tab**: Includes the information shown in Figure 2.11.



Assay Id	External Id	Target Allele	Assay Allele	Strand Switched?	Map	Chromosome	Position (bp)	Gene	Tag Id	Feature ID
191049	rs9653	G/T	A/C	Y	NCBI build 34	12	53,075,423	ZNF385	6,247	2,781
191049	rs9653	G/T	A/C	Y	NCBI build 34	12	53,075,423	ZNF385	6,247	6,816
191049	rs9653	G/T	A/C	Y	NCBI build 34	12	53,075,423	ZNF385	6,247	11,079
191063	rs10789	A/C	A/C		NCBI build 34	12	119,391,862	DNCL1	4,177	564
191063	rs10789	A/C	A/C		NCBI build 34	12	119,391,862	DNCL1	4,177	4,920

Figure 2.10 Assay Panel Information – Assays Tab



Property	Value
--- Pre-Cluster Fit Data Cuts ---	
[CUSTOM] MinCallConfidence	0.9
--- Post-Cluster Fit Assay Cuts ---	
[CUSTOM] MinAssayCallRate	80.0
[CUSTOM] MinAssayH.W.P-Value	0.00E0
[CUSTOM] MinAssayRepeatability	0.0
[CUSTOM] MinAssayTrioConcordance	0.0

Figure 2.11 Assay Panel Information – Default Genotype Settings Tab

Changing Assay Panel Descriptions

To change an assay panel description:

1. Right-click the assay panel name and select **Properties**.
2. Enter a new description in the Description field.
3. Click **Save**.

Adding or Updating Map, Chromosome, Position and Gene Details

You can add or update the map, chromosome, position (bp) and gene information for your assay panel at any time (Figure 2.10).

To add or update map, chromosome, position, and gene information:

1. Export the assay panel file by right-clicking on the relevant assay panel.
2. Open your assay panel file with Microsoft® Office Excel®.

   The assay panel file is also located on the CD-ROM included with the first shipment of GeneChip® SNP Kits you received.
3. Do one of the following:

   A. Edit the file to include map, chromosome, position and gene details.

   B. Modify the existing map, chromosome, position and gene details.
4. Save and close the file.
5. Open GTGS.
6. In the left pane, open the menu and select **Assay Panels**.
7. Right-click the assay panel name and select **Update Assay Information From File**.
8. Locate the assay panel file that you modified and click **Open**.

This operation will also update the annotation information for saved genotype results. The next time you export genotype results, the updated annotations will be included in the file.

About Sample and Sample Tracking Information

Within each project, sample and sample tracking information can be viewed in a variety of ways. Depending upon where you are when running the protocol or processing data, sample information can be viewed by selecting:

- The Tracking folder
- Any of the plate folders within the Tracking folder
- The Array Data icon
- Any of the cluster icons in the Genotype Results folder (Figure 2.9).

While the various stages of the Targeted Genotyping Protocol are being performed, samples are moved from one plate to another, then to arrays. Each plate is organized in its own tracking subfolder. To view sample tracking information a plate or array, expand the Sample, Anneal, Assay, Label, Hyb or Arrays folder and select a plate or array. Then select a particular plate or array, and view the information displayed in the right pane of the window.

For Sample Plates, you can view information such as the sample plate barcode and where each sample was located on the plate. For each Anneal, Assay, Label and Hyb Plate, you can see who created the plate and when, the plate from which the samples were loaded, and the plate or array to which the samples were transferred for the next stage of the protocol.

To view all of the tracking information in one table, select the All Experiments tab (Figure 2.12).

Once you have performed the *Hyb on Arrays* operation, finished sample and array tracking information can be viewed by clicking the Tracking icon for your project and selecting the Finished Experiments tab in the right pane of the window. Sample information available includes the:

- Experiment: a truncated version of the array barcode.
- Sample Name: the name of the sample loaded onto the array displayed in the Experiment column
- Sample Location: the sample plate barcode and well location the sample

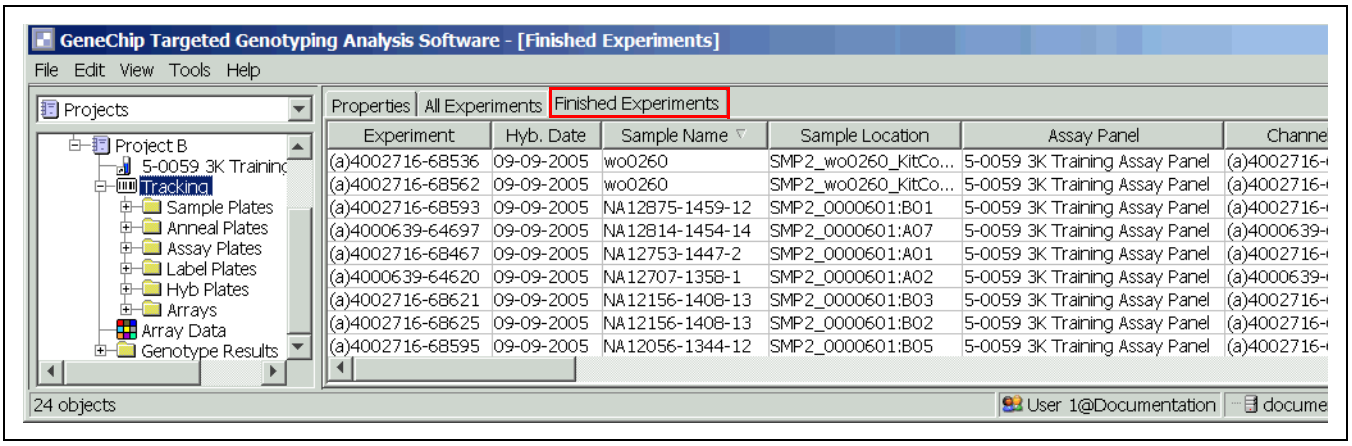


Figure 2.12 Sample Information Available After Performing the Hyb on Arrays Operation

Experiments that still require .cel data from GCOS can be viewed by selecting the Array Data icon (Figure 2.13):

- **Hyb Date**: the date the sample was hybridized onto the array listed in Experiment column
- **Experiment** (typically the last 12 digits of the array barcode for a particular sample)
- **Gender**: displayed if imported via a Sample Info file
- **Sample Name**: the name assigned to the sample and imported into GTGS via a Sample Plate file
- **Sample Location**: the barcode of the sample plate and the well location from which the sample was taken

Refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide* for more information on Sample Info and Sample Plate files.

Once you import and process the data collected by the scanner, the remaining columns in the Experiment QC Summary table are populated (except for the Manual column).

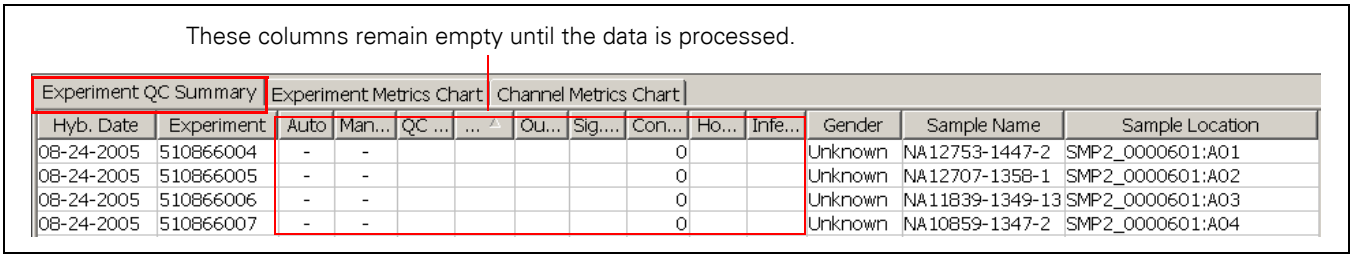


Figure 2.13 Sample and Array Information Displayed in the Experiment QC Summary Table

Chapter 3

ANALYZING DATA

Data Analysis Workflow

The flow chart below describes the workflow to follow when analyzing data. An overview of sample tracking and array data collection is included (Figure 3.1).

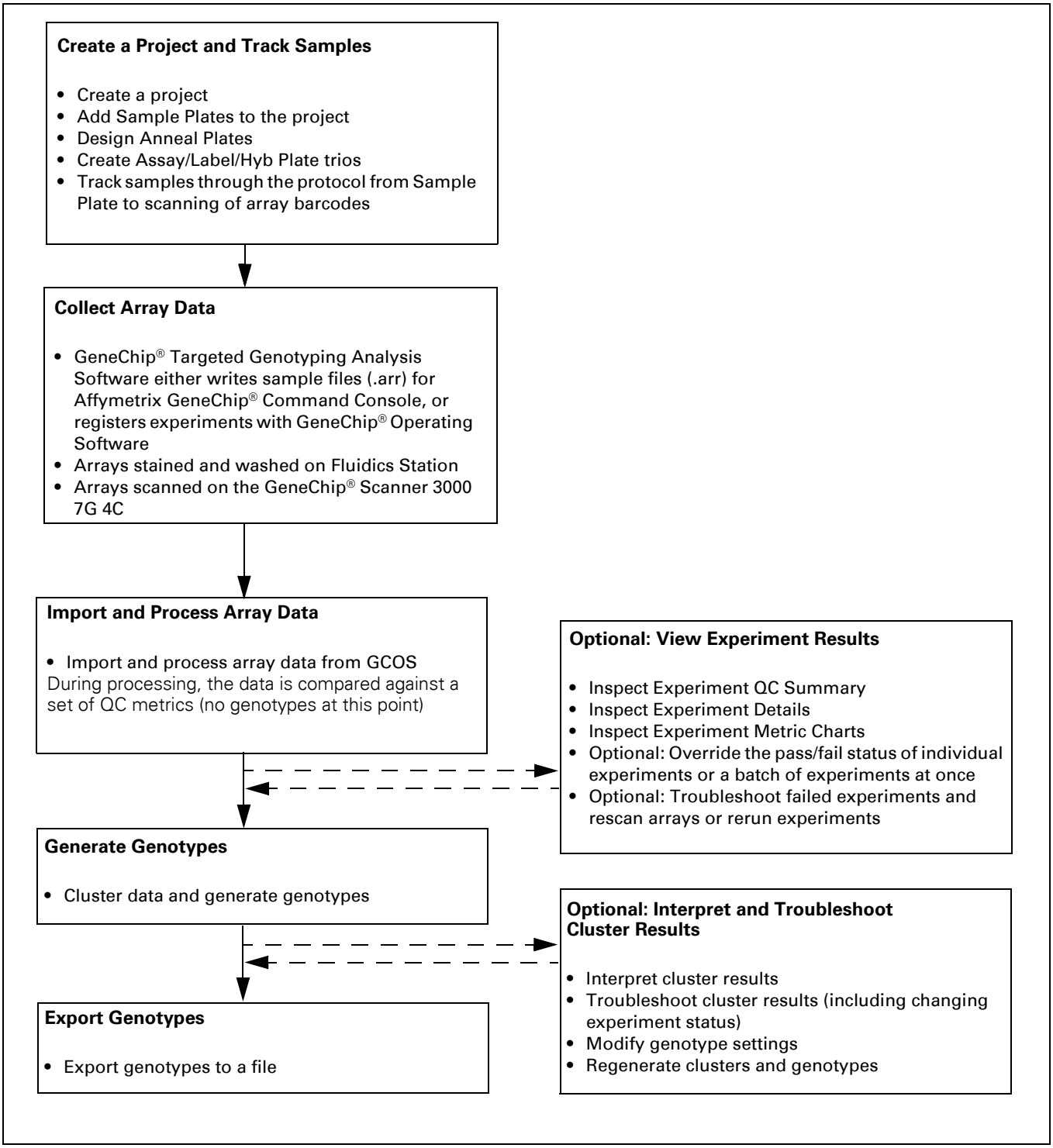


Figure 3.1 Workflow from Assay to Genotype

Creating Projects and Tracking Samples

For information on creating projects, performing the Targeted Genotyping Protocol, and tracking samples, refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide*.

Importing and Processing Array Data

Process Overview

Once arrays have been scanned, you must import the data collected by the scanner into Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS) and process it.

About Sample and Array Tracking Information

Sample and array tracking information is automatically transferred to the Array Data section of the project when the normal tracking steps are followed for the Targeted Genotyping Protocol.

If necessary, sample and array tracking information can be:

- Copied from one project to another.
- Imported from a file.

See *Importing Sample and Array Tracking Information from a File* on page 26.

For more details on sample tracking information, refer to *About Sample and Sample Tracking Information* on page 21.

Copying Sample and Array Tracking Information Between Projects

The following operation allows you to copy a subset of the sample and array tracking information between projects. The information is copied from the Tracking section of one project to the Array Data section of project in which you are working.

> **IMPORTANT:** **The following procedure should NOT be done for GCOS data while in AGCC mode, or AGCC data while in GCOS mode. See *AGCC and GCOS Compatibility Modes* *on page 101* for more information and alternate recommendations.**

To copy sample and array tracking information between projects:

1. Expand the folder for the project into which the data will be copied.
2. Right-click the **Array Data** icon and select **Import Experiment List**.
3. Under the **Select From Database** tab, select the project that contains the sample and array tracking information you want to copy.
4. Click **Import**, then click **Save**.

Importing Sample and Array Tracking Information from a File

You may want to import sample and array tracking information from a file if you want to:

- Copy tracking information from one project to another.
- Import tracking information not managed by your copy of GTGS, you must import the sample and array tracking information from a tab-delimited text file.

To import sample and array tracking information from a file:

1. Expand the project folder.
2. Right-click the **Array Data** icon (Array Data) and select **Import Experiment List**.
   The Import Experiment List window opens.
3. Click the **Load from File** tab.
4. Click the **Browse** button.
5. Locate and select the experiment list filename (saved as a tab-delimited text file, Figure 3.2).

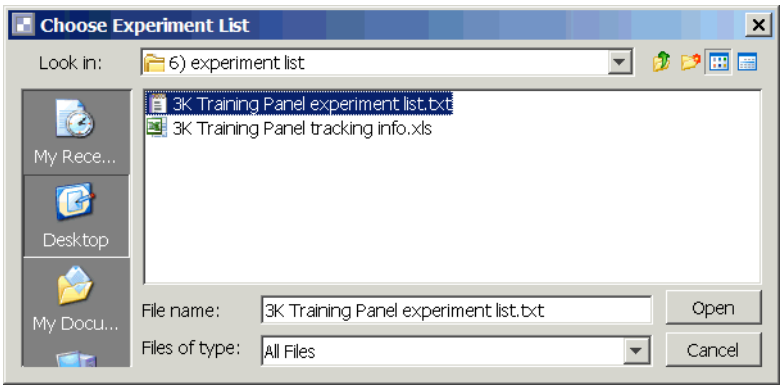


Figure 3.2 Importing an Experiment List from a File

6. Click **Open**.
7. Click the **Import** button; then click **Save**.
   The number of experiments imported is displayed in the Results field. Any errors that occurred are displayed in the Problems window (Figure 3.3).

NOTE: As you process additional samples for the same project, you must repeat this procedure to import the new tracking information into GTGS.

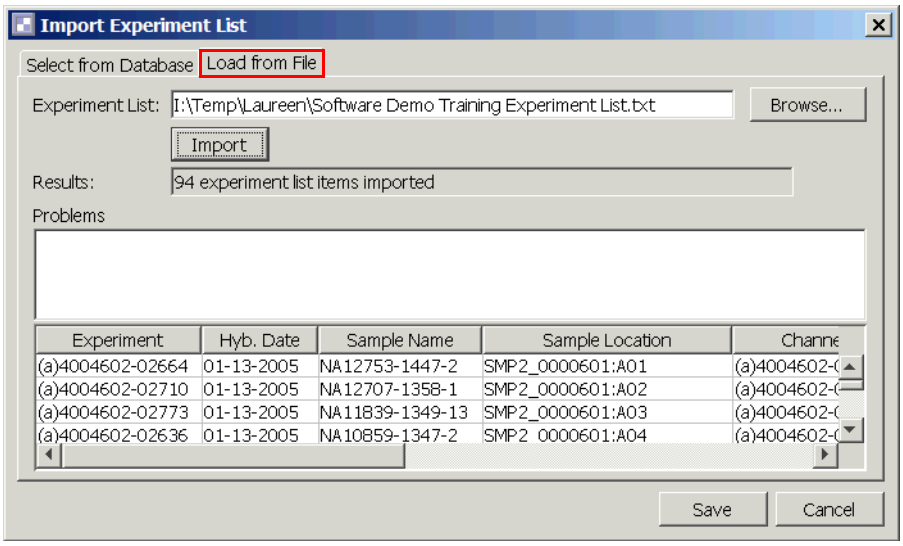


Figure 3.3 Import Experiment List Window After Importing Experiment List from a File

Viewing Sample and Array Tracking Information

Sample and array tracking information is viewed by selecting the Array Data icon for your project. It is displayed in the right pane under the Experiment QC Summary tab (Figure 3.4.) Notice that many of the fields in this table are empty. These fields remain empty until the experiment data is imported from Affymetrix GeneChip® Command Console (AGCC) or GeneChip® Operating Software (GCOS) and the data is processed.

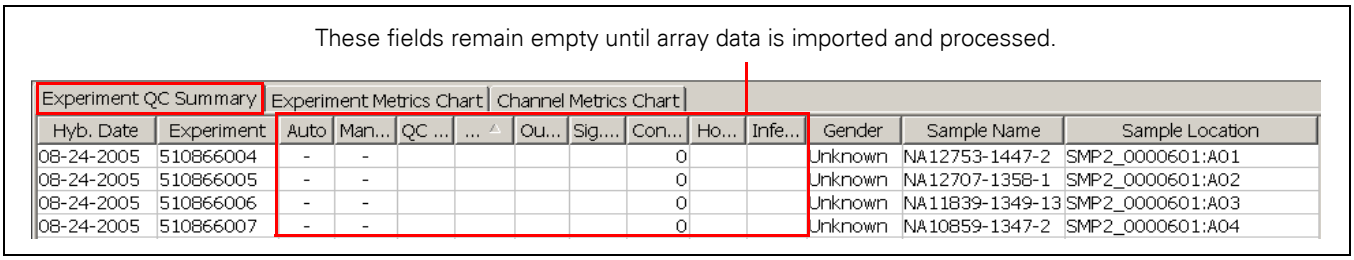


Figure 3.4 Viewing Sample and Array Tracking Data

Importing and Processing Array Data

About Importing and Processing Array Data

GTGS imports and processes the array data stored in .cel files directly from folders or from GCOS. As described below, the Import Experiment Data window (Figure 3.6) allows you to view the progress of this operation.

The Ready to Process table shows all of the tracked experiments in the project that have not yet been imported. As each line is read, GTGS extracts .cel file data from the specified location based on the filenames in the channel columns of the table.

The Messages field displays the data processing status. As data for each experiment is successfully loaded, it is saved to the database and is displayed in the Processed field. If an experiment cannot be loaded, a message is displayed in the Messages box.

> **IMPORTANT: Genotyping is not performed during this step. The purpose of the processing step is to normalize the data, calculate quality control metrics, and prepare the data for genotyping. As such, data processing can take some time.**

If the processing operation is cancelled before completion, the experiments that have already been processed remain in the database.

Array data is imported either directly from the GCOS server, or from a directory. To import and process array data:

- Directly from the GCOS server, follow the instructions listed under ***Importing Array Data From GCOS*** on page 28. This option is only available when GTGS is in GCOS compatibility mode. See *AGCC and GCOS Compatibility Modes on page 101* for more information.
- From a directory, follow the instructions listed under ***Importing Array Data From a Directory*** on page 29.

Importing Array Data From GCOS

To import and process array data from GCOS:

1. In GTGS, expand your project folder.
2. Right-click the **Array Data icon** (Array Data) and select **Import Experiment Data**.

   The Import Experiment Data window appears.
3. Select the GCOS **Server** radio button (Figure 3.6 on page 29).

   The information displayed in the GCOS Server field is the name of the workstation that manages the scanner data.
4. Click **Process**.

   The array data .cel files are imported from the Instrument Control Workstation and the data is processed. The progress of this operation is displayed in the Messages field.
5. When done, click **Load Rescans**.

   This action causes GTGS to read the list of .cel files in the scanner database. If the software finds that all 4 channels for a particular experiment have been rescanned, it will add a new experiment to the Ready to Process field.

   The .cel files for rescans are identified by a filename that includes the extension _<number><channel>. For example, if an array with the barcode 4004602-02693 has been scanned three times, the .cel filenames for the third scan would be (a)4004602-02693_3A, (a)4004602-02693_3B, (a)4004602-02693_3C, and (a)4004602-02693_3D.
6. Click **Process** to import and process any rescans.

7. Make a note of any experiments that were skipped.

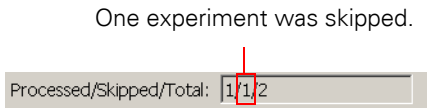


Figure 3.5 Skipped Experiments

See *About Skipped Experiments* on page 32 for more information.

8. Optional: Scroll through the list of processed experiments to preview the processing results.
9. Click **Close**.

   Notice that the previously empty columns in the Experiment QC Summary table in the right pane have been populated (Figure 3.6).

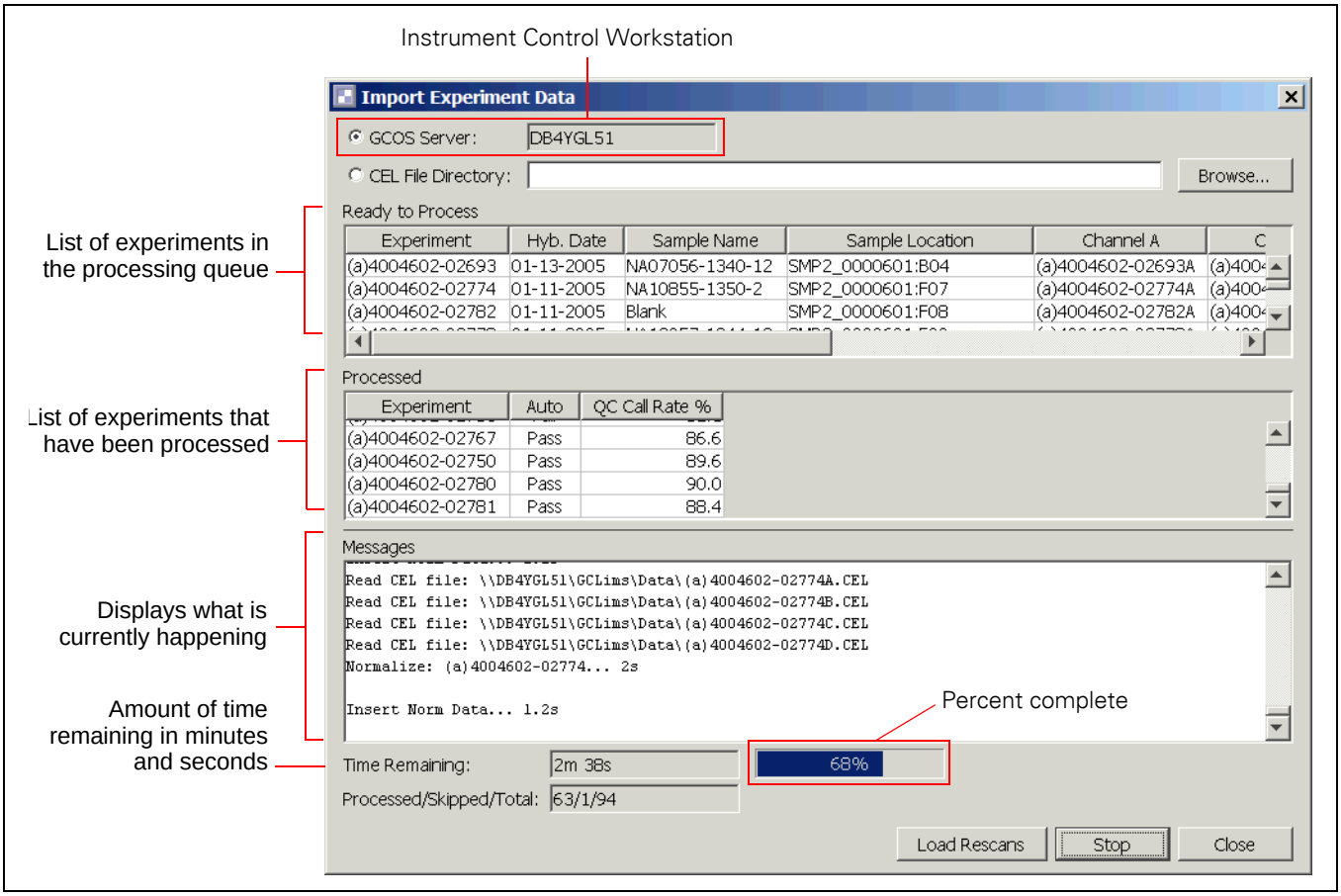


Figure 3.6 In Progress .cel File Import

Importing Array Data From a Directory

If the .cel files you want to access are not on the GCOS server, you must specify the directory in which the .cel files are located. For example, the directory may be where AGCC creates the .cel files for the current project.

To import and process array data from a directory:

1. Expand the project folder.
2. Right-click the **Array Data icon** (Array Data) and select **Import Experiment Data**.

   The Import Experiment Data window opens.
3. Select the **CEL File Directory** radio button (Figure 3.7.)

4. Click the **Browse** button and locate the folder that contains the .cel files.
5. Select the folder name; do not open the folder.

> **NOTE:** **If open the .cel file directory, it will appear empty. This behavior is normal. In this window, you can only select folders, not files.**

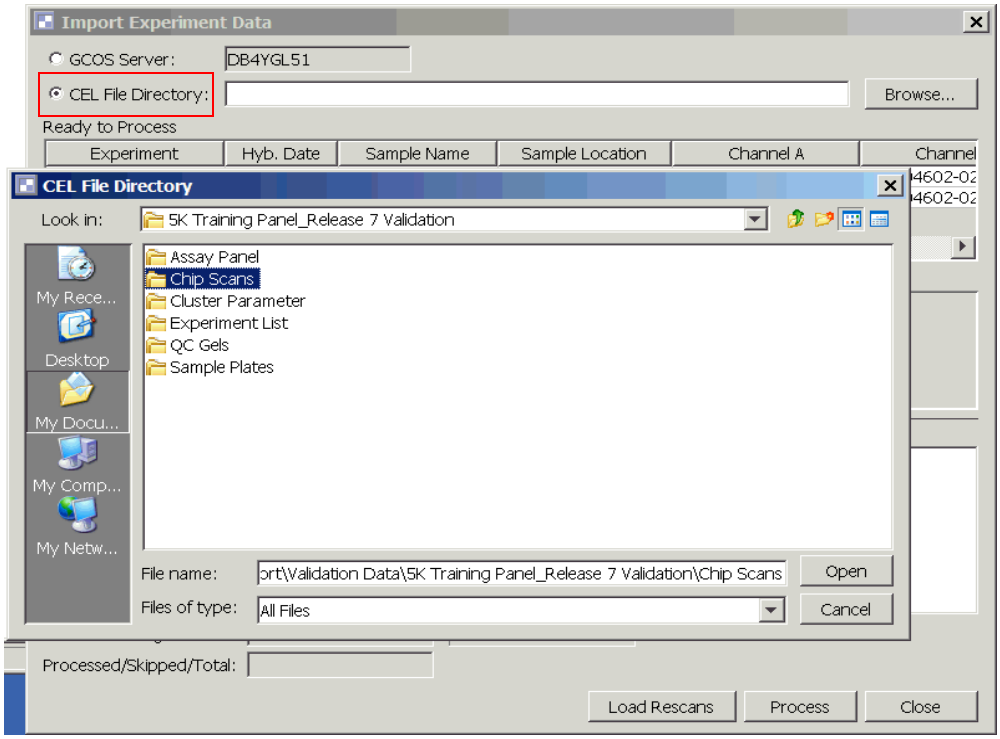


Figure 3.7 Importing .cel Files From a Directory

6. Click **Open**; then click **Process**.
7. When done, click **Load Rescans**.

   This action causes GTGS to look in the target directory for up to nine scans of each experiment currently in the project. If the software finds that all 4 channels for a particular experiment have been rescanned, it will add a new experiment to the Ready to Process field.

   The .cel files for rescans are identified by a filename that includes the extension _<number><channel>. For example, if an array with the barcode 4004602-02693 has been scanned three times, the .cel filenames for the third scan would be (a)4004602-02693_3A, (a)4004602-02693_3B, (a)4004602-02693_3C, and (a)4004602-02693_3D.
8. Click **Process** to import and process any rescans.
9. Make a note of any experiments that were skipped.

   See *About Skipped Experiments* on page 32 for more information.
10. Optional: Scroll through the list of processed experiments to preview the processing results.
11. Click **Close**.

    Notice that the previously empty columns in the Experiment QC Summary table in the right pane have been populated.

NOTE: **GTGS records the current .cel file location for each experiment. If you change the location of these files, the software may ask you to update the directory. However, this will not prevent you from genotyping the data, since a normalized copy of the data is saved in the GTGS database.**

Updating the Location of CEL Files

When data is imported, the .cel file location for each experiment is recorded by GTGS. If the location of the .cel files for a particular project is changed, you will have to update the path GTGS in order to view information such as Experiment Details. If the location is not updated, an error message stating that the .cel data cannot be found is displayed (Figure 3.8).

To update the location of .cel files for a project:

1. Open the project in GTGS.
2. Right-click the Array Data icon
3. Right-click on the row an experiment in the Experiment QC Summary table and select **View Experiment Details**.

   If GTGS cannot find the .cel file for this experiment, the error message shown in (Figure 3.8) is displayed (Summary Tab).

1. Click the **Change Location** button.
2. Browse to the directory containing the .cel files.

   Note that since the browser filters the display to only show directories, you will not see the CEL files.
3. Click **Open**; then click **Apply**.
4. Click **Next**; then click **Prev** to load the .cel file data.

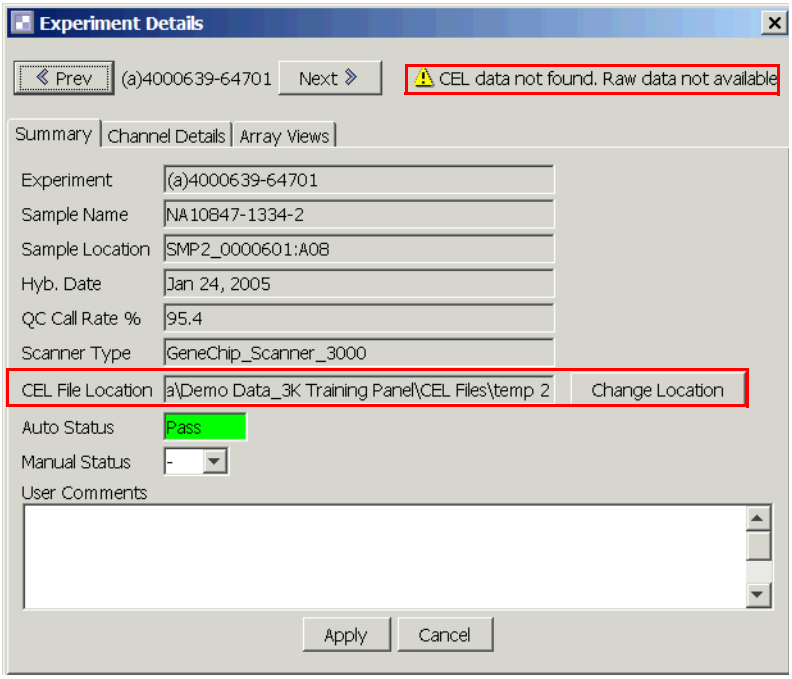


Figure 3.8 Updating CEL File Location

About Skipped Experiments

When you are finished importing and processing .cel files, the Import Experiment Data window may look something like the one in Figure 3.9. This illustration contains an example of what you will see when one or more .cel files for a particular experiment are not imported.

To determine why .cel files were not imported:

1. Look at the **Processed/Skipped/Total** field.

   In Figure 3.9 you can see that the .cel files for one experiment were not imported. The software imported and processed the .cel files for 93 out of 94 experiments.

2. Look at the top field, **Ready To Process**, for the experiment name.

3. In the Messages field, scroll through the list and locate the Error messages that pertain to skipped experiments (see Figure 3.9 and Figure 3.10.)

   In Figure 3.10, the error message indicates that .dat and .cel files for channel D could not be found. A possible cause would be that this particular channel has not yet been scanned.

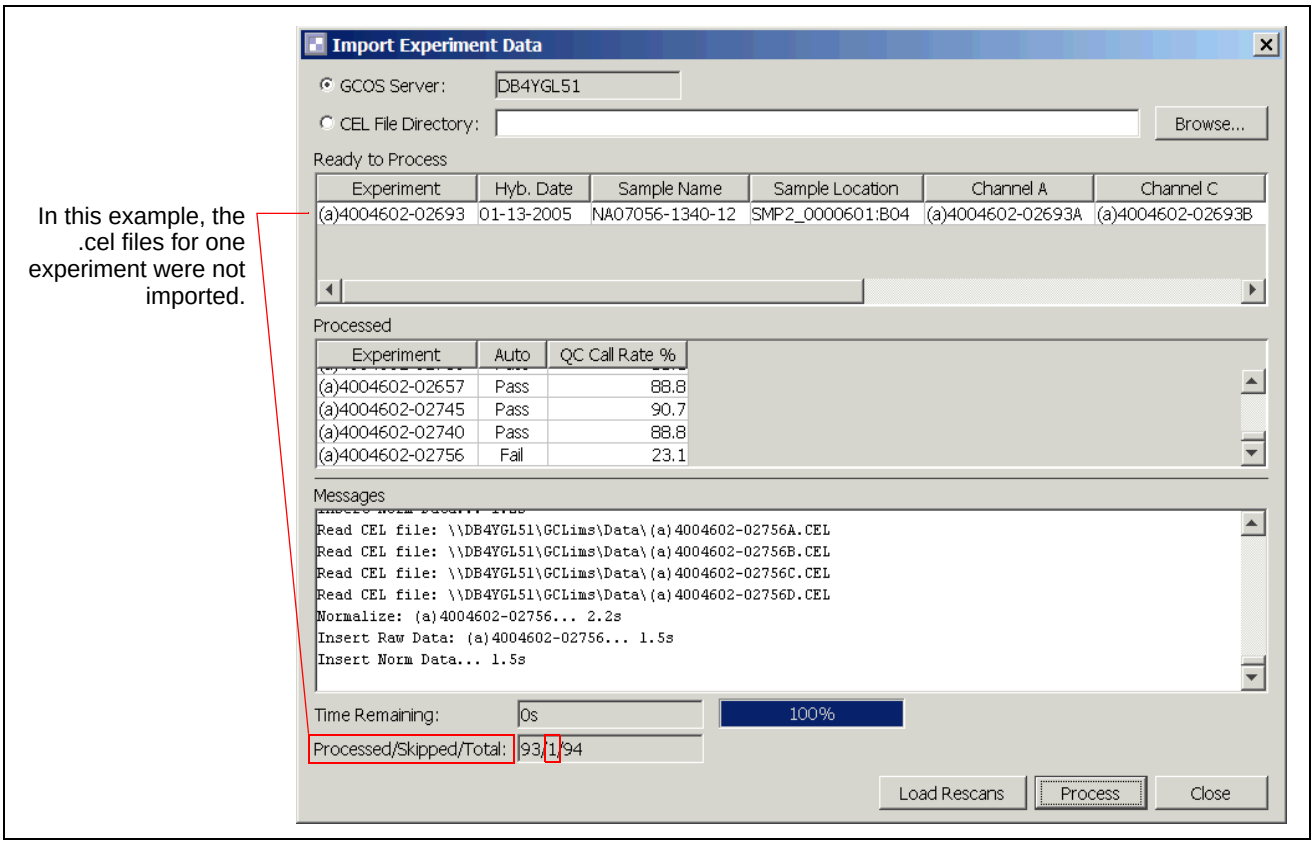


Figure 3.9 .cel File Import and Process Complete

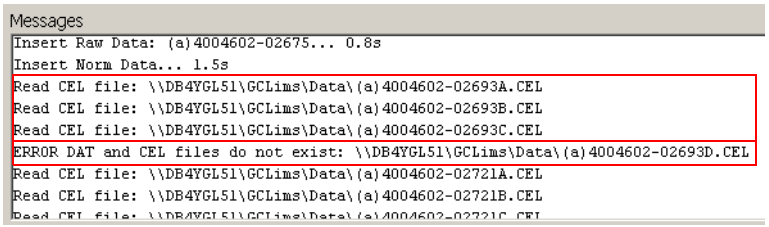


Figure 3.10 Error Message for .cel File Import

4. Troubleshoot any error messages based on the guidelines listed under *Troubleshooting Error Messages in the Messages Field* on page 33.

Troubleshooting Error Messages in the Messages Field

Table 3.1 Troubleshooting Error Messages When Importing Experiment Data

Error Message	Possible Causes	Solution
ERROR DAT and CEL files do not exist: <*path to .cel file*>	• The entire array was not scanned. • One or more channels of an array were not scanned.	• Wait for the array to finish scanning. • Rescan the array.
	The wrong folder was specified; therefore, the software could not find the .cel files.	Repeat the instructions under *Importing Array Data From a Directory* on page 29 and specify the correct folder.
	If you process GCOS .cel files while in AGCC mode, or process AGCC .cel files while in GCOS mode, there may be a mismatch between the expected and actual file names.	See *AGCC and GCOS Compatibility Modes* on page 101.
ERROR DAT file exists but CEL file does not exist: <*path to .cel file*>	• A gridding error occurred. • A software error occurred.	Manually grid the array. Refer to the troubleshooting chapter of the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide* for more information.

About Processed Data

You can view the results of experiment processing by:

- Scrolling through the list in the Processed field
- Viewing the Experiment QC Summary

Notice that columns that were previously empty have been populated (except for the Manual column; Figure 3.11). Each experiment has been assigned a pass or fail status. At this point, genotypes have not yet been determined.

Auto	Manual	QC Call Rate %	QC Half Rate %	Outlier Rate %	Sig./Bkgd.	Controls CV %	Hom %	Inferred Gender
Pass	-	95.8	3.79	0.102	156	4.16	68.66	Unknown
Pass	-	95.3	4.18	0.0567	131	4.76	68.96	Unknown
Pass	-	95.0	4.78	0.0340	149	3.88	68.73	Unknown

Figure 3.11 Columns Propagated When Array Data is Processed by GTGS

What You Can Do Now

At this point in the process, you can:

- **Continue processing and adding more experiment data to this project.**

  If you have not yet run at least 80 unique samples, we recommend that you continue processing samples and importing the data into GTGS.

  IMPORTANT: For meaningful results, we recommend you collect and process data from at least 80 unique samples before you perform any preliminary cluster genotyping.

- **Review your data before generating genotypes.**

  Before generating genotypes, you can review the data that you have imported and processed. While reviewing your data you may want to:

  - Compare the Inferred Gender against the known Gender for each sample. If the genders do not match, a sample tracking error may have occurred. (see *Comparing Known Gender with the Inferred Gender* on page 87 for more information).
  - Change the pass/fail status of experiments (see *Changing the Status of an Experiment* on page 97 for instructions).
  - Delete experiment data.
  - Look for trends in experiment data quality that occur over time. Trends can indicate:

    Assay performance variations in the lab.

    Sample quality variations.

    Refer to Chapter 5, *Reviewing, Working With, and Troubleshooting Data* for more information and instructions.

- **Generate and export genotypes.**
  - If you have processed at least 80 unique samples, you may want to generate a preliminary cluster. Continue to *Cluster Genotyping* on page 37.
  - If all of your samples have been run and all of the data has been processed in GTGS, you can perform final cluster genotyping and export your genotypes. Continue to *Generating and Exporting Genotypes* on page 35 for instructions.

Generating and Exporting Genotypes

About Generating Genotypes

Generating genotypes using GTGS is commonly referred to clustering or cluster genotyping. Clustered data are data that are similar to each other. In graphical plots, clustered data appears bunched together. An essential feature of SNP genotyping is that measured signals for a SNP (one assay) do not distribute uniformly. Instead, the signals distribute in clumps that generally denote 1 of only 3 calls: AA, AB, or BB. Cluster genotyping is genotyping that assigns calls to data by:

1. Determining the cluster locations for each SNP across a set of samples.
2. Then assigning each data point to a cluster.

When experiments are initially processed, the data is loaded into a database and is measured against a set of quality control metrics. The QC Call Rate reported in the Experiment QC Summary table is one such preliminary metric. QC Call Rate is only an estimate. It is based on fixed thresholds that treat all assays equally.

Differences in biochemistry among the assays means that some assays generate better quality data. GTGS adapts to the expected variation among assays, resulting in more accurate genotyping.

The cluster genotyping operation automatically performs the following steps:

1. **Validation of the individual assays for an experiment**: assays that fail certain quality control metrics are removed from the experiment.
2. **Validation of entire experiments**: experiments that fail certain quality control metrics are removed from the data set.
3. **Assay clustering:** each assay is automatically inspected across all validated experiments, and the locations of up to three clusters for each assay are found.
4. **Preliminary genotyping:** each validated data point for an assay (Figure 3.12) is assigned to one of up to three possible clusters (homozygous for allele 1, heterozygous, or homozygous for allele 2), but only if the assignment is unambiguous.
5. **Validation of genotyped assays:** an entire assay can be removed from the data set if it has:
   - A high proportion of experiments did not generate genotypes for an assay
   - An unusual distribution of genotypes among the clusters
   - Low repeatability in reporting a genotype for different experiments using the same sample
   - Low Mendelian consistency

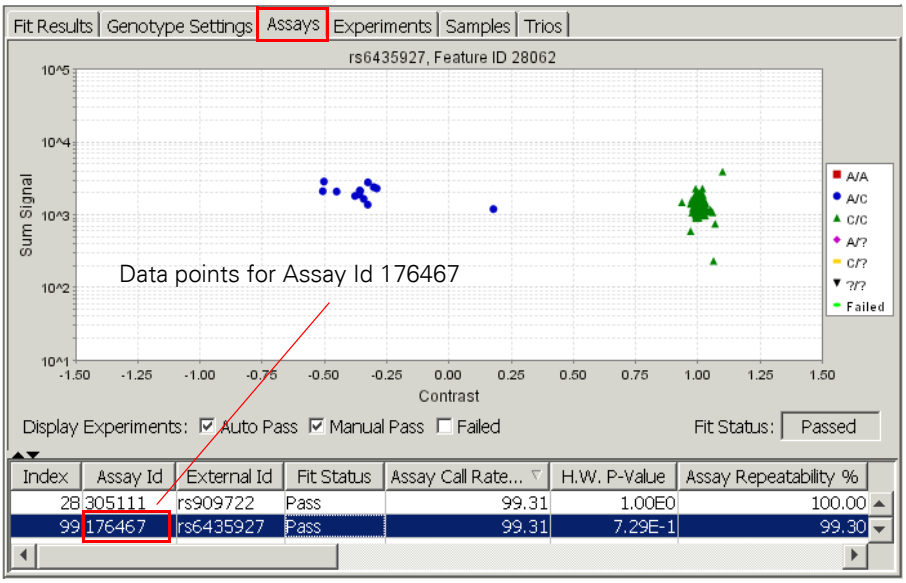


Figure 3.12 Data Points for One Assay

The genotype settings used for the clustering operation are provided in the form of a data file. The settings in this file are specific to each assay panel. The file is located on a CD ROM that is included when you receive the first GeneChip® SNP Kits for a particular assay panel.

The assay clustering step generates reliable clusters if high quality data is used to determine the cluster locations of each assay. Because of this, GTGS determines cluster locations based on validated experiments only – experiments automatically assigned the status of Pass during data processing. If you manually change the status of an experiment from Fail to Pass (using the Manual Status box), the experiment is **not** used to determine cluster locations. It will, however, be genotyped (see Table 3.2).

Table 3.2 How Experiment Status Effects Clustering and Genotyping

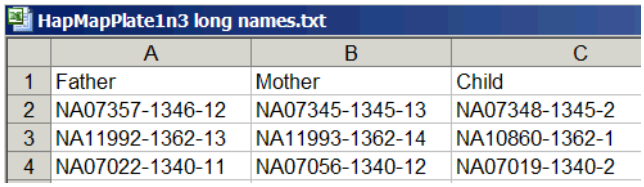
Auto Status	Manual Status	Experiment used to determine cluster locations?	Experiment Genotyped?
Pass	–/ Pass	Yes	Yes
Pass	Fail	No	No
Fail	Pass	No	Yes
Fail	–/ Fail	No	No

NOTE: In general, clustering on a larger set of experiments will generate higher quality results. Genotypes calculated from a set of experiments where most experiments are manually passed are of marginal value.

Cluster Genotyping

Creating Trio Name Files

If you are doing checks for Mendelian consistency, you will need to import a Trio Name file as part of the cluster genotyping operation. Trio Name files are tab-delimited text files that include the sample name for the mother, father and child of each trio (Figure 3.13). Create these files using a text editor such as Microsoft® Notepad or Microsoft® Office Excel® (saved as a tab-delimited text file).



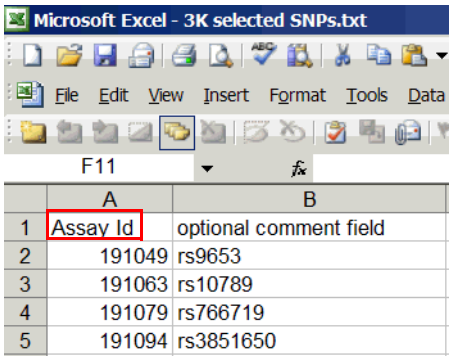
HapMapPlate1n3 long names.txt

	A	B	C
1	Father	Mother	Child
2	NA07357-1346-12	NA07345-1345-13	NA07348-1345-2
3	NA11992-1362-13	NA11993-1362-14	NA10860-1362-1
4	NA07022-1340-11	NA07056-1340-12	NA07019-1340-2

Figure 3.13 Trio Name File Format

Creating Assay Subset Files

To specify a subset of assays to be clustered, you must create a tab-delimited text file that contains the assay ids. Only the assay id is required. All other information is optional (Figure 3.14).



	A	B
1	Assay Id	optional comment field
2	191049	rs9653
3	191063	rs10789
4	191079	rs766719
5	191094	rs3851650

Figure 3.14 Assay Subset File

Performing Cluster Genotyping

To perform cluster genotyping:

1. Right-click the **Genotype Results** folder for your project and select **Perform Cluster Fit Genotyping** (Figure 3.15).

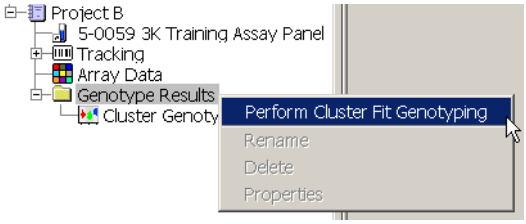


Figure 3.15 Perform Cluster Fit Genotyping

The Cluster Fits window appears.

2. Enter a name for the new cluster fit.

   Cluster name parameters (Figure 3.16).:

   - Maximum length 50 characters.
   - Cannot contain these characters: \ / : ; * ? " < > | ‘ , { } [ ]

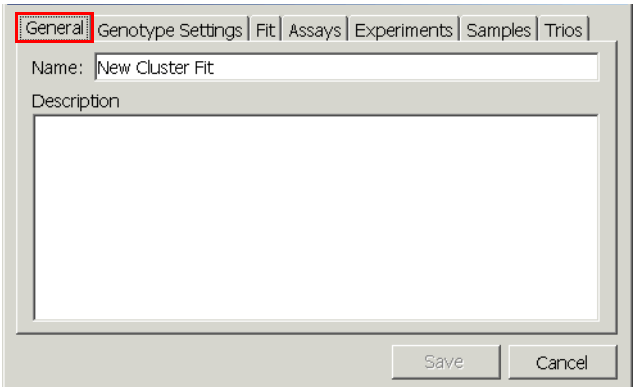


Figure 3.16 Cluster Fits Window

3. Optional: Enter a description for the cluster fit.
4. Optional: Under the **Genotype Settings** tab:

   A. If you are doing a Mendelian consistency check, import a Trio Names file.

   If not imported, the MinAssayTrioConcordance cut is meaningless.

   1) Click the **Browse** button associated with the **Trio Names File** field.

   2) Locate and select the file.

   3) Click **Set**.

   You must still click Set even if the correct path is displayed when you select the Genotype Settings tab.

   See *Creating Trio Name Files* on page 37 for more information.

   B. If you want to genotype a subset of assays for a particular assay panel, import an Assay ID Subset file.

   This action will decrease the amount of time it takes to cluster.

   1) Click the **Browse** button associated with the **Optional Assay ID Subset File** field.

   2) Locate and select the file.

3) Click **Set**.

You must still click Set even if the correct path is displayed when you select the Genotype Settings tab.

See *Creating Assay Subset Files* on page 37 for more information.

C. If you want to change any of the genotype settings:

1) Select the value you want to change.

2) Change the value and press **Enter**.

3) To restore the default values, click **Restore Defaults**.

A description of the value is displayed. The description includes a valid range of values that you can enter **(**Figure 3.17**)**.

NOTE: Any changes made are applied to this cluster fit only. The default values are the ones contained in the Genotype Settings file loaded from the CD-ROM that came the first shipment of GeneChip SNP Kits for this assay panel.

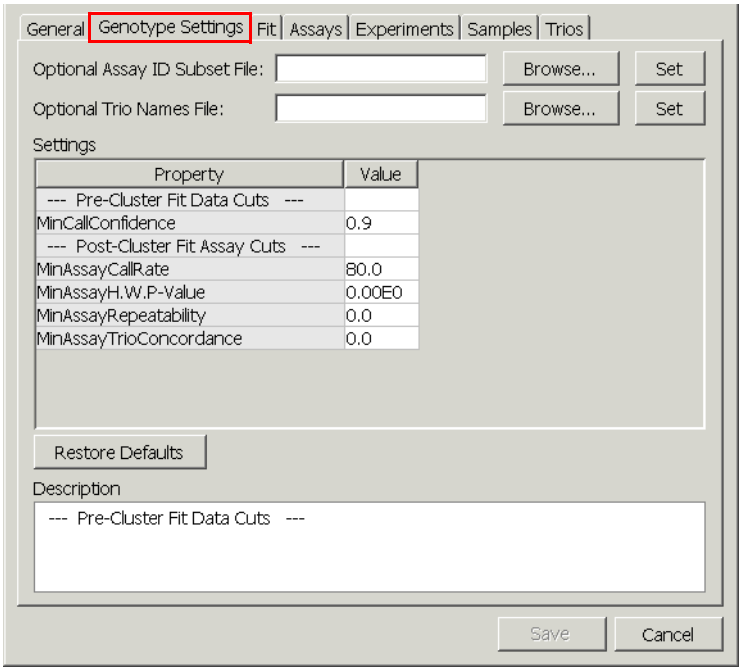


Figure 3.17 Cluster Fit Window – Genotype Settings Tab

5. Click the **Fit** tab.

6. Optional: To automatically save cluster and genotype results to the database, select the **Autosave Genotypes to Database** check box.

Advantage to selecting: If the clustering operation will take several hours or overnight, you can select this check box and know that the cluster results will be saved to the database and not lost.

Advantage to deselecting: You can review the cluster genotyping results first, then decide whether or not to save them to the database (Figure 3.18).

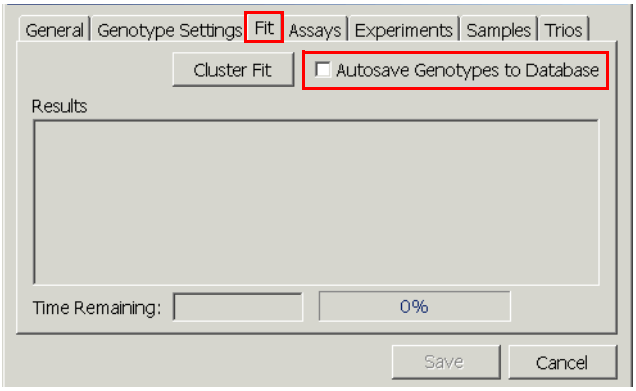


Figure 3.18 Autosave Genotypes to Database Option

7. Click the **Cluster Fit** button.
   The clustering operation begins.

> **NOTE: Clustering may take some time depending upon the amount of data you are processing.**

8. If you:
   - Selected **Autosave Genotypes to Database** before clustering, proceed to the next step.
   - Did not select **Autosave Genotypes to Database** before clustering, then do one of the following:
     Click **Cancel** to discard the cluster results.
     Click **Save** to save your results to the database.

> **NOTE: You may want to review the cluster results as shown in the Fit, Assays, Experiments and Samples tabs before saving the results to the database.**

9. Click **OK**.

Viewing Cluster Genotyping Results

To view the cluster genotyping results:

1. Expand the **Genotype Results** folder for your project (Figure 3.19).
2. Select the name of the cluster.

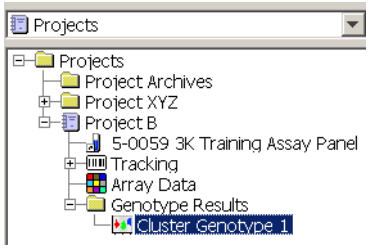


Figure 3.19 Selecting a Cluster to View Genotype Results

3. Select the Fit Results, Genotype Settings, Assays, Experiments, Samples and Trios tabs to view the cluster genotyping results (Figure 3.20).

Fit Results | Genotype Settings | Assays | Experiments | Samples | Trios

Property	Value
# Experiments	92
# Genotyped Experiments	91
# Unique Samples	84
# Genotyped Unique Samples	83
# Trios	24
# Assays	2,938
# Passed Assays	2,918
Passed Assays %	99.32
Failed by Low Call Rate %	0.68
Failed by Low H.W. P-Value %	0.00
Failed by Low Repeatability %	0.00
Failed by Low Trio Concordance %	0.00
# Repeated Sample Discordances	44
# Repeated Sample Concordances	40,419
Cluster Fit Repeatability %	99.89
# Trio Discordances	144
# Trio Concordances	65,253
Cluster Fit Trio Concordance %	99.78
Completeness %	98.75
Low Call Confidence Cut %	0.83

Figure 3.20 Viewing Cluster Genotyping Results

What You Can Do Now

At this point, you can do the following:

- Continue processing and adding more experiment data to this project. Any time you add more experiment data to a project, you must generate a new cluster to include the new data.
- Export your genotypes.
  - See *Exporting Genotypes for an Entire Project* on page 42.
- Review your cluster genotyping results.
  - Refer to Chapter 5, *Reviewing, Working With, and Troubleshooting Data* for more information.

Exporting Genotypes for an Entire Project

See *Genotype Output Formats and Call Codes* on page 44 for a description of output types.

To export genotypes for an entire project:

1. Expand the **Genotype Results** folder.
2. Right-click the cluster fit name and select **Export Genotype Results, Entire Project . . .**
   The Export Genotype Results window appears.
3. Open the **Output Type** menu and select one of the following **(**Figure 3.21**)**:
   - Genotypes Standard Format
   - Genotypes Long Format
   - Genotypes Short Format
   - Assay Allele Signals
   - Genotype Settings

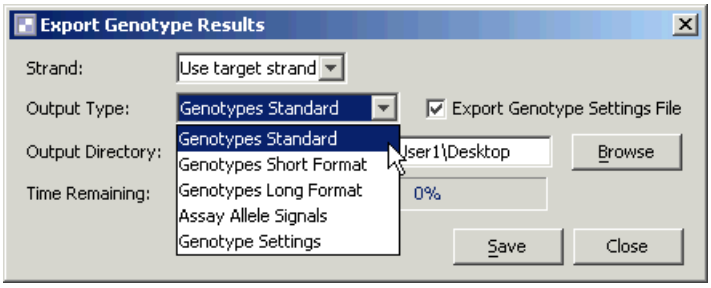


Figure 3.21 Export Genotype Results – Output Types Available

4. Click the **Browse** button and select a location for the output file.
5. Optional: To export the genotype settings at the same time, select the Export Genotype Settings File check box.
   Genotype settings are saved as a separate file.
6. Click **Save**.
7. When the save is complete, click OK, then click **Close**.

Exporting Genotypes for One Experiment

To export genotypes for one experiment:

1. Expand the **Genotype Results** folder.
2. Select the cluster name.
3. Select the **Experiments** tab.
4. In the table, right-click the experiment of interest and select **Export Genotype, This Experiment, Passed Assays**

   The Select Directory window appears.
5. Select a location for the file.
6. Click **Open**.

Exporting Genotypes for One Assay

To export genotypes for one assay only:

1. Expand the **Genotype Results** folder.
2. Select the cluster name.
3. Select the **Assays** tab.
4. In the table, right-click the experiment of interest and select **Export Genotype, This Assay, Passed Experiments**.

   The Select Directory window appears.
5. Select a location for the file.
6. Click **Open**.

Viewing Exported Genotype Files

We recommend opening and viewing exported genotype files using Microsoft Office Excel, unless exporting to the Genotypes Long format.

To view exported genotype files:

1. Right-click the filename.
2. Select **Open With → Microsoft Office Excel**.

Genotype Output Formats and Call Codes

Output Formats

Formats available for exporting genotype data and settings are:

- Standard – one assay per row (Figure 3.22)
- Short – one genotype per row (Figure 3.23)
- Long – one genotype per row (Figure 3.24)
- Assay Allele Signals – one genotype per row (Figure 3.25)

  The assay allele signals output includes the signals used to calculate the genotypes. For example, you can use this information to recreate the scatterplots shown in the assays or experiments tabs on a saved cluster result.
- Genotype Settings (Figure 3.26)

  Genotype settings are the full set of settings used to generate the genotypes

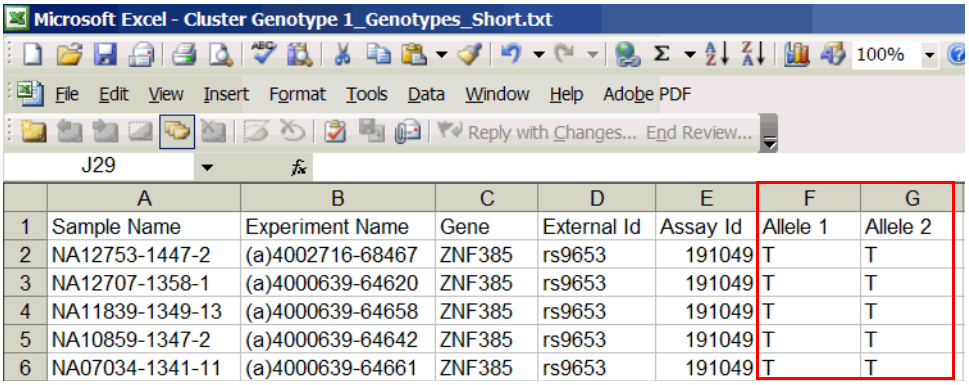
Example of Standard Format

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1									Experiment Name		(a)4002716-68467	(a)4000639-64620	(a)4000639-64643	(a)
2									Status		Auto Pass	Auto Pass	Fail	Au
3									Sample Location		SMP2_0000601:A01	SMP2_0000601:A02	SMP2_0000601:A06	SM
4									Sample Name		NA12753-1447-2	NA12707-1358-1	NA07055-1341-12	NA
5									Sample Repeat Count		1	1	0	
6									Experiment Call Rate %		99.2117889	98.32076765	0	
7	Assay Index	Assay Id	External Id	Target Allele	Genome Map	Chrom Name	Chrom Position	Gene	Assay Call Rate %					
8	0	191049	rs9653	G/T	NCBI build 34	12	53075423	ZNF385	100		2	2	5	
9	1	191063	rs10789	A/C	NCBI build 34	12	119391862	DNCL1	97.8021978		1	1	5	
10	2	191079	rs766719	G/T	NCBI build 34	12	16780580	-	100		2	2	5	
11	3	191094	rs3851650	A/C	NCBI build 34	12	115254812	-	100		1	2	5	

Figure 3.22 Genotype Output Formats – Standard

Example of Short Format

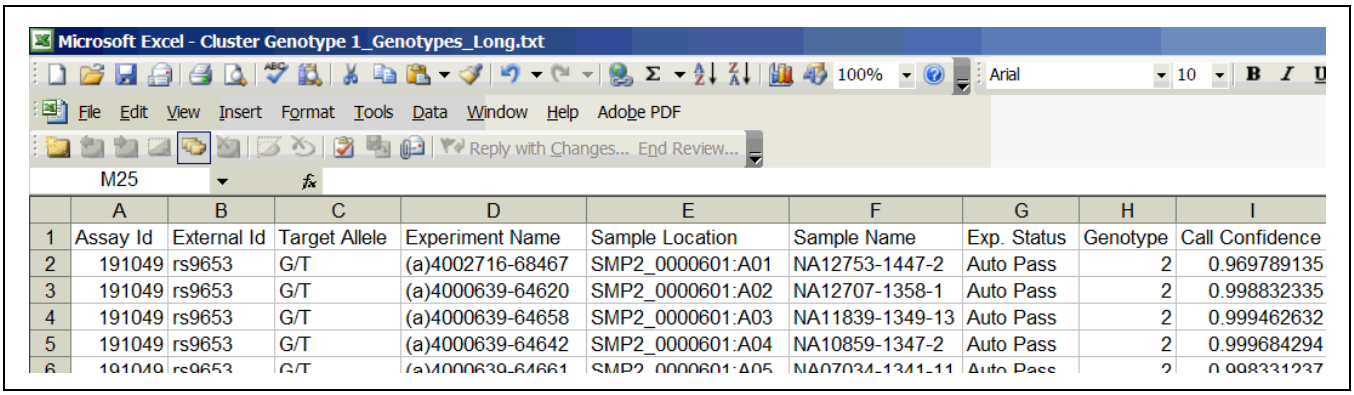
The short format includes the letter codes for genotypes as shown in the figure below.



	A	B	C	D	E	F	G
1	Sample Name	Experiment Name	Gene	External Id	Assay Id	Allele 1	Allele 2
2	NA12753-1447-2	(a)4002716-68467	ZNF385	rs9653	191049	T	T
3	NA12707-1358-1	(a)4000639-64620	ZNF385	rs9653	191049	T	T
4	NA11839-1349-13	(a)4000639-64658	ZNF385	rs9653	191049	T	T
5	NA10859-1347-2	(a)4000639-64642	ZNF385	rs9653	191049	T	T
6	NA07034-1341-11	(a)4000639-64661	ZNF385	rs9653	191049	T	T

Figure 3.23 Genotype Output Formats – Short

Example of Long Format



	A	B	C	D	E	F	G	H	I
1	Assay Id	External Id	Target Allele	Experiment Name	Sample Location	Sample Name	Exp. Status	Genotype	Call Confidence
2	191049	rs9653	G/T	(a)4002716-68467	SMP2_0000601:A01	NA12753-1447-2	Auto Pass	2	0.969789135
3	191049	rs9653	G/T	(a)4000639-64620	SMP2_0000601:A02	NA12707-1358-1	Auto Pass	2	0.998832335
4	191049	rs9653	G/T	(a)4000639-64658	SMP2_0000601:A03	NA11839-1349-13	Auto Pass	2	0.999462632
5	191049	rs9653	G/T	(a)4000639-64642	SMP2_0000601:A04	NA10859-1347-2	Auto Pass	2	0.999684294
6	191049	rs9653	G/T	(a)4000639-64661	SMP2_0000601:A05	NA07034-1341-11	Auto Pass	2	0.998331237

Figure 3.24 Genotype Output Formats – Long

Example of Assay Allele Signals Output

1	Assay Id	External Id	Experiment Name	Sample Location	Sample Name	Exp. Status	Assay Status	Assay Allele	Is Assay Reverse Complemented from Target SNP?
2	191049	rs9653	(a)4002716-	SMP2_0000601:A01	NA12753-1447-2	Auto Pass	Pass	A/C	TRUE
3	191049	rs9653	(a)4000639-	SMP2_0000601:A02	NA12707-1358-1	Auto Pass	Pass	A/C	TRUE
4	191049	rs9653	(a)4000639-	SMP2_0000601:A03	NA11839-1349-13	Auto Pass	Pass	A/C	TRUE
5	191049	rs9653	(a)4000639-	SMP2_0000601:A04	NA10859-1347-2	Auto Pass	Pass	A/C	TRUE

Measured Genotype for Assay Allele	A Normalized Signal	C Normalized Signal	G Normalized Signal	T Normalized Signal	Contrast	Signal Sum
0	7747.2056	292.53174	35.734016	-21.03104	-0.86343	8039.737
0	8399.323	-59.862076	-43.79608	81.44976	-1.03021	8339.461
0	8300.163	-34.280968	48.506363	-0.014109	-1.02257	8265.882
0	8744.991	24.794477	21.638216	162.45306	-0.98942	8769.785

Figure 3.25 Example of Assay Allele Signals Output

Example of Genotype Settings Output

	A
1	SoftwareVersion=1.5.0.0
2	Assay Panel=5-0059 3K Training Assay Panel
3	# --- General Fit Settings ---
4	SampleNormType=1
5	#--- Replicate Settings ---
6	ReplicateMethod=1
7	MinRepConcordanceRatio=0.5
8	MinRepDiscordanceRatio=0.34
9	# --- Cluster Fit Settings ---
10	ClusterAlgorithmVersion=0
11	SigmaConst=0.01
12	UseWeights=1
13	# --- Pre-Cluster Fit Data Cuts ---
14	MinSignalToNoise=3.0
15	MinSignalToBgnd=1.5
16	# MaxPixelCV is not a configurable setting.
17	MaxPixelCV=200.0
18	# --- Post-Cluster Fit Data Cuts ---
19	MinCallConfidence=0.9
20	MinClusterSignalFraction=0.2
21	MinHetToHalfRatio=2.0
22	# --- Post-Cluster Fit Assay Cuts ---
23	MinAssayCallRate=80.0
24	MinAssayH.W.P-Value=0.0
25	MinAssayRepeatability=0.0
26	MinAssayTrioConcordance=0.0

Figure 3.26 Example of Genotype Settings Output

Genotype Call Codes

The table below contains the genotype call codes you should use to determine your genotypes.

Table 3.3 Genotype Call Codes

Genotype Call Codes				
Allele/Call Code	**0**	**1**	**2**	**5**
AC	AA	AC	CC	NO CALL
AG	AA	AG	GG	NO CALL
AT	AA	AT	TT	NO CALL
CG	CC	CG	GG	NO CALL
CT	CC	CT	TT	NO CALL
GT	GG	GT	TT	NO CALL
General	**0**	**1**	**2**	**5**
XY	XX	XY	YY	NO CALL

Chapter 4

ABOUT CHARTS AND TABLES

Before You Start

Purpose of this Chapter

The purpose of this chapter is to provide a description of:

- The charts and tables included in Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS)
- The different ways in which you can manipulate and utilize the information in these charts and tables

NOTE: This chapter does not provide instructions for processing, analyzing and exporting data. For instructions, refer to **Chapter 3,** *Analyzing Data*.

Tables and Charts for Reviewing Array Data

About the Quality Control Metrics

The following quality control metrics are used to determine the status automatically assigned to each experiment by GTGS. The status can be Pass or Fail. To assign the status of Pass, the threshold for each metric must be met. If any threshold is not met, the experiment is automatically assigned the status Fail. Failed experiments are excluded from the genotyping operation (Table 4.1).

Table 4.1 Quality Control Metrics

Quality Control Metric	Description	Threshold
QC Call Rate %	Estimated percentage of assays that can be clearly genotyped. The higher the number, the more assays that will be genotyped.	Must be ≥ 80%
QC Half Rate %	Estimated percentage of assays that might be genotyped. The lower the number, the more likely it is that marginal assays will generate a genotype.	Must be ≤10%
Signal to Background	Median ratio of assay allele/non-allele channel signals (median signal to background ratio) The higher the value, the better the signal strength.	Must be ≥ 20
Controls CV% (Controls coefficient of variation)	Percent coefficient of variation of control feature signals.	Must be ≤30%
If any threshold is not met, the experiment is automatically assigned the status Fail.		

Experiment QC Summary

About the Experiment QC Summary

Prior to importing scanner data from Affymetrix Genechip® Command Console (AGCC) or GeneChip® Operating Software (GCOS), the following sample and array information can be viewed by selecting the Array Data icon (Figure 4.1):

- **Hyb Date**: the date the sample was hybridized onto the array listed in Experiment column
- **Experiment** (typically the last 12 digits of the array barcode for a particular sample)
- **Gender**: displayed if imported via a Sample Info file
- **Sample Name**: the name assigned to the sample and imported into GTGS via a Sample Plate file
- **Sample Location**: the barcode of the sample plate and the well location from which the sample was taken

Refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide* for more information on Sample Info and Sample Plate files.

When arrays are scanned, the data is captured in four files referred to as *.cel files*. Once you import and process this data from GCOS, the remaining columns in the Experiment QC Summary table are populated (except for the Manual column). These columns report how the assay performed when measured against a set of quality control metrics. Refer to *About the Quality Control Metrics* on page 48 for definitions that do not appear below.

- **Auto**: the status of the experiment automatically assigned by GTGS (status can be Pass or Fail)
- **Manual**: users can manually pass/fail experiments (see *Generating and Exporting Genotypes* on page 35 for instructions)
- **QC Call Rate %**
- **QC Half Rate %**
- **Outlier Rate %**
- **Signal/Background**
- **Controls CV %**
- **Hom %**:
- **Inferred Gender**: the sample gender as called by GTGS (only available for some assay panels)

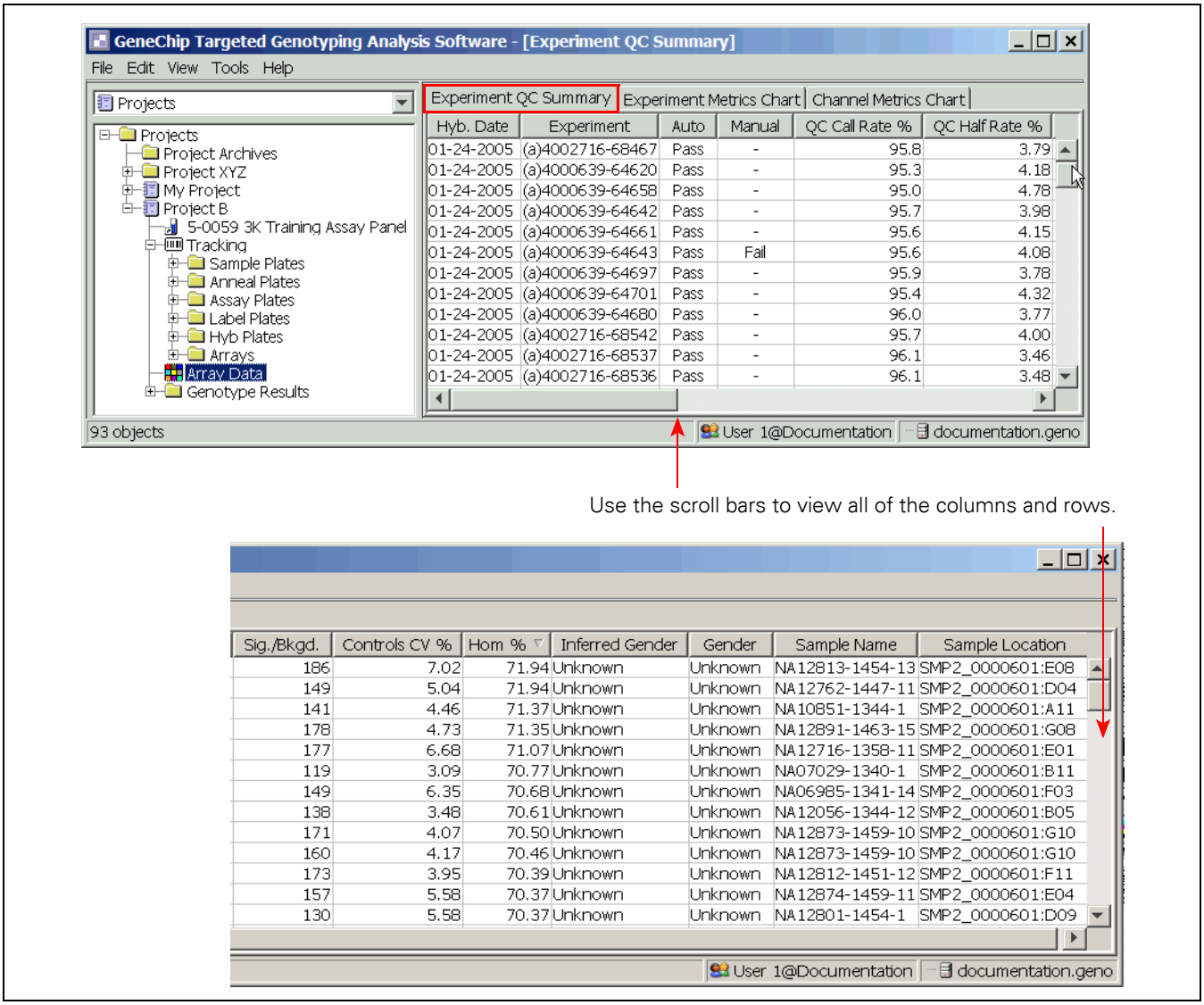


Figure 4.1 Example of Experiment QC Summary Table

From the Experiment QC Summary you can also:

- View the details of specific experiments
- Delete experiments from the Array Data icon
- Export the experiment list
- Export the Experiment QC Summary table

Instructions are provided below.

Viewing the Experiment QC Summary

To view an Experiment QC Summary:

1. In the left pane of GTGS, expand your project.
2. Select the **Array Data** icon.
3. In the right pane, click the **Experiment QC Summary** tab.
4. Use the scroll bars to view all of the columns and rows.

Viewing Experiment Details

To view the details of an experiment:

1. In the left pane of GTGS, expand your project.
2. Select the **Array Data** icon.
3. Right-click anywhere in the row of the experiment you want to view.
4. Select **View Experiment Details**.
   See *Experiment Details* on page 57 for more information.

Deleting an Experiment

To delete an experiment:

1. In the left pane of GTGS, expand your project.
2. Select the **Array Data** icon.
3. Right-click anywhere in the row of the experiment you want to view.
4. Select **Delete**.
5. Click **Yes**.

Exporting Experiment Lists

Experiment lists are exported as tab-delimited text files.

To export an experiment list:

1. In the left pane of GTGS, expand your project.
2. Right-click the **Array Data** icon and select **Export Experiment List**.
3. Enter a file name in the File name field.
4. Open the **Save In** menu and select a location for the file.
5. Click **Save**.

Experiment Metrics Chart

About the Experiment Metrics Chart

The Experiment Metrics Chart (Figure 4.2) shows trendline plots denoting experiment performance against a set of quality control metrics. You can use this chart (and the Channel Metrics Chart, on page 55) as a quick way to:

- Assess the overall quality of your data for a particular project or for multiple projects.
- Quickly identify any unusual experiments in a set.

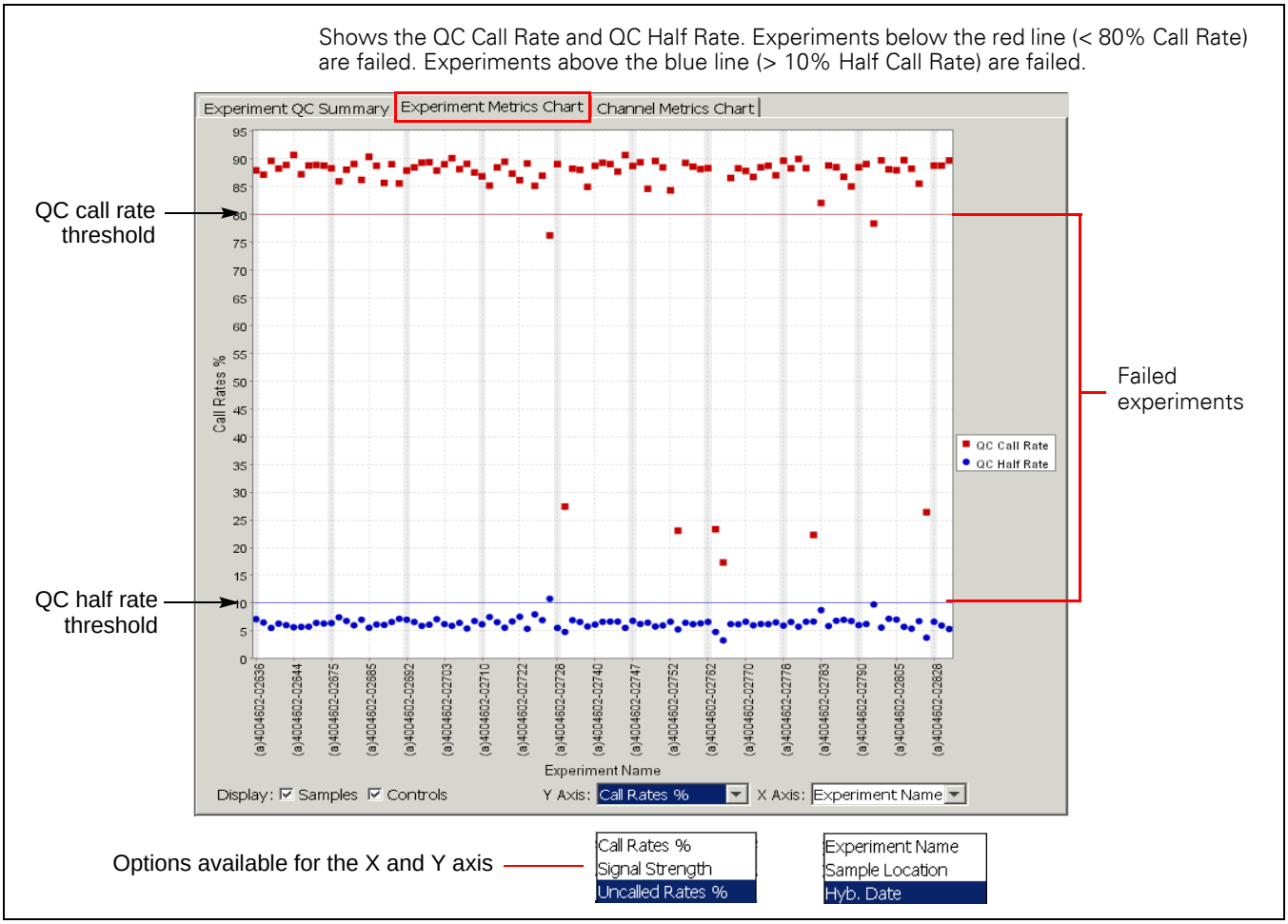


Figure 4.2 Experiment Metrics Chart Showing Call Rates by Hyb Date

Horizontal lines indicate thresholds for experiments of the same color. These thresholds are used to determine whether an experiment is automatically passed or failed prior to genotyping. Refer to ***About the Quality Control Metrics*** on page 48 for more information on thresholds.

During initial data processing, each assay on an array is evaluated against the quality control metrics and will fall into one of the following categories:

- Will be genotyped (QC Call Rate)
- Might be genotyped (QC Half Rate)
- Will not be genotyped, due to:
  - Weak signal (Low Signal Rate)
  - Strong non-allele signal (High Background Rate)

- Unreliable measurements (Rejected Outliers Rate)

NOTE: The percentages you see when looking at the QC Call Rate, QC Half Rate and Uncalled Rate for a particular assay will total 100% (Figure 4.3).

For experiment (a)4000484-01397:

Call Rate = ~ 90%
Half Call Rate = ~ 5%
Uncalled Rate due to low signal = ~ 3%
Uncalled Rate due to high background = ~ 2%
100%

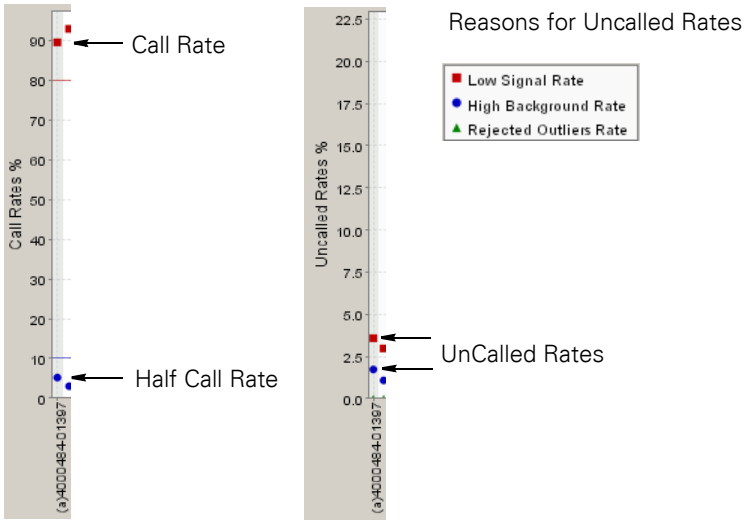


Figure 4.3 Call Rates in the Experiment Metrics Chart

Viewing Experiment Metrics Charts

To view an Experiment Metrics Chart:

1. Click one of the following:
   - The **Array Data** icon for a particular project.
   - Any folder that contains projects.

   The chart will display all of the data for each project contained within that folder and all subfolders and projects (Figure 4.4).

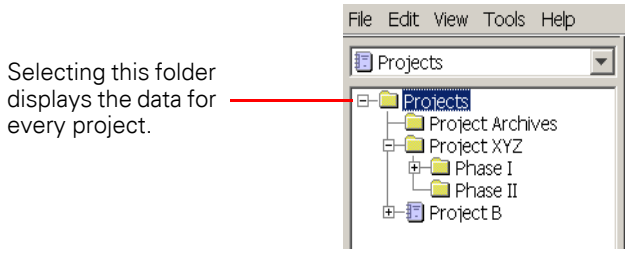


Figure 4.4 Viewing Experiment Metrics Charts for Multiple Projects

2. In the right pane, click the **Experiment Metrics Chart** tab.
3. Open the X and Y axis drop-down menus to change the information displayed.
   Display options are listed under *Display Options for the Experiment Metrics Chart* below.
4. Display samples or controls only by selecting and deselecting the **Samples** and **Controls** check boxes.

Display Options for the Experiment Metrics Chart

Below are the various display options available for Experiment Metrics Charts. See *About the Quality Control Metrics* on page 48 for more information.

- Y Axis options are:
  - Call Rates % (QC Call Rate and QC Half Rate)
  - Signal Strength (Signal to background)
  - Uncalled Rates %
- X Axis options are:
  - Experiment Name – typically an abbreviated version of the array barcode
  - Sample Location – location of the sample on the original sample plate
  - Hyb Date – date the assay was hybridized onto the array
- Display Samples/Controls check boxes
  Select or deselect check boxes to display experiments that are samples only, controls only, or samples and controls.

Channel Metrics Chart

About the Channel Metrics Chart

The Channel Metrics Chart (Figure 4.5) shows trendline plots denoting experiment performance against a set of quality control metrics. You can use this chart (and the Experiment Metrics Chart (Figure 4.14 on page 52) as a quick way to:

- Assess the overall quality of your data for a particular project or for multiple projects.
- Quickly identify any unusual experiments in a set.

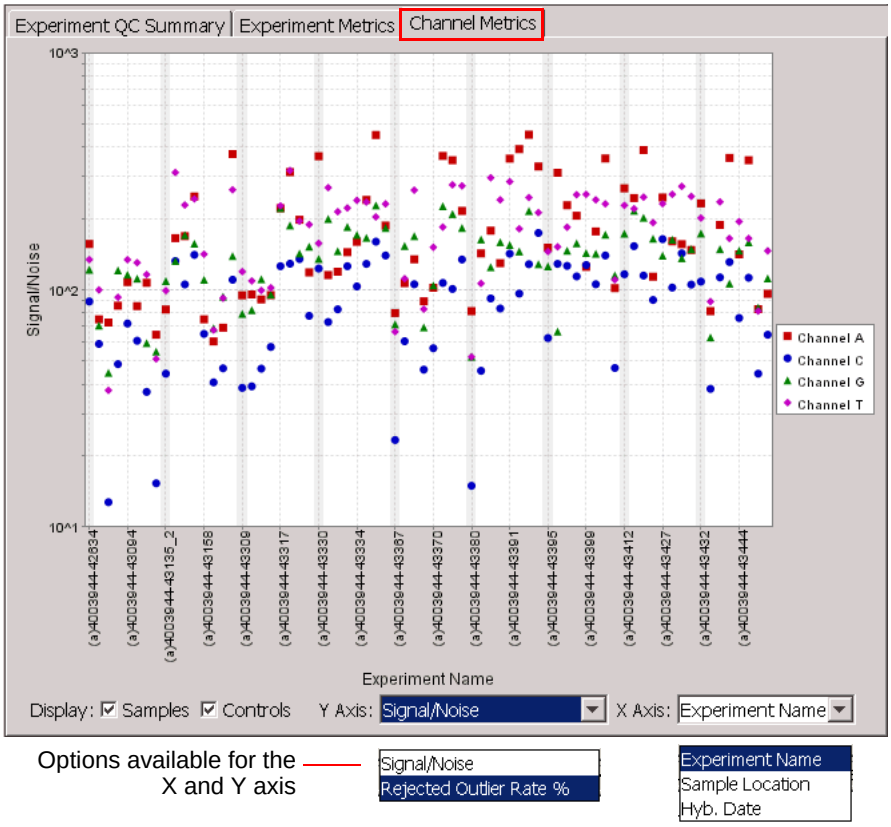


Figure 4.5 Channel Metrics Chart

Horizontal lines indicate thresholds for experiments of the same color. These thresholds are used to determine whether an experiment is automatically passed or failed prior to genotyping. Refer to *About the Quality Control Metrics* on page 48 for more information on thresholds.

Channel metrics are calculated for each individual channel. They summarize information about the data quality of the individual nucleotide channels that were scanned. Each channel corresponds to a specific nucleotide (A, C, G or T). As such, you will see four data points for each experiment. Because experiments are not passed or failed based on channel metrics, no thresholds are associated with this chart.

Viewing a Channel Metrics Chart

To view a Channel Metric:

1. Click one of the following:
   - The Array Data icon for a particular project.
   - Any folder that contains projects.

The chart will display all of the data for each project contained within that folder and all subfolders and projects.

2. In the right pane, click the **Channel Metrics Chart** tab.
3. Open the X and Y axis drop-down menus to change the information displayed.
   Display options are listed under Figure 4.6 below.
4. Display samples or controls only by selecting and deselecting the **Samples** and **Controls** check boxes.

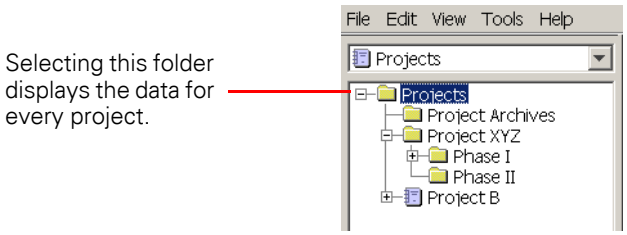


Figure 4.6 Viewing Channel Metrics Charts for Multiple Projects

Display Options for the Channel Metrics Chart

Below are the various display options available for Channel Metrics Charts. See *About the Quality Control Metrics* on page 48 for more information.

- Y Axis options are:
  - **Signal to Noise** – Median signal-to-noise for a particular channel (assay features/standard deviation of dimmest features). Noise is calculated across the entire array. The higher the value, the stronger the signal strength relative to noise.
  - **Rejected Outlier Rate %** – Percentage of assays failed due to unreliable measurements in that channel. The lower the number, the less rejected outliers.
- X Axis options are:
  - **Experiment Name** – typically an abbreviated version of the array barcode
  - **Sample Location** – location of the sample on the original sample plate
  - **Hyb Date** – date the assay was hybridized onto the array
- Display Samples/Controls check boxes
  Select or deselect check boxes to display experiments that are samples only, controls only, or samples and controls.

Experiment Details

About Experiment Details

From the Experiment QC Summary (on page 49), you can view detailed information for each experiment. This information is organized under three tabs: Summary, Channel Details, and Array Views. Each tab is described in more detail below. Reviewing these details can help you troubleshoot failed experiments.

Viewing Experiment Details

To view experiment details:

1. Expand the project folder.
2. Click the **Array Data** icon for the project.
3. Click the **Experiment QC Summary** tab.
4. Right-click anywhere in the row of the experiment of interest, and select **View Experiment Details**.
5. Select a tab in the Experiment Details window (Figure 4.7).

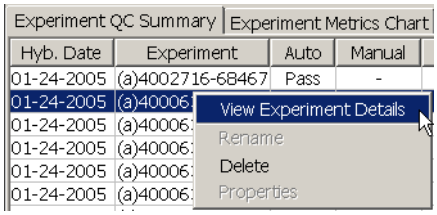


Figure 4.7 Viewing Experiment Details

About the Summary Tab

The Summary tab displays general experiment information as shown in Figure 4.8. It also allows you to manually change the status of an experiment and add comments. See *Changing the Status of an Experiment* on page 97 for instructions.

Also included in this window is the path to the directory where the .cel file data for each experiment is stored. If the files are moved, you will have to update the path in GTGS in order to view information such as Array Views in Experiment Details. To update the .cel file path, see *CEL Data Not Found – Updating the Location of CEL Files* on page 99.

By clicking the <Prev or Next> button, you can scroll through your experiments in the order they are listed in the Experiment QC Summary table.

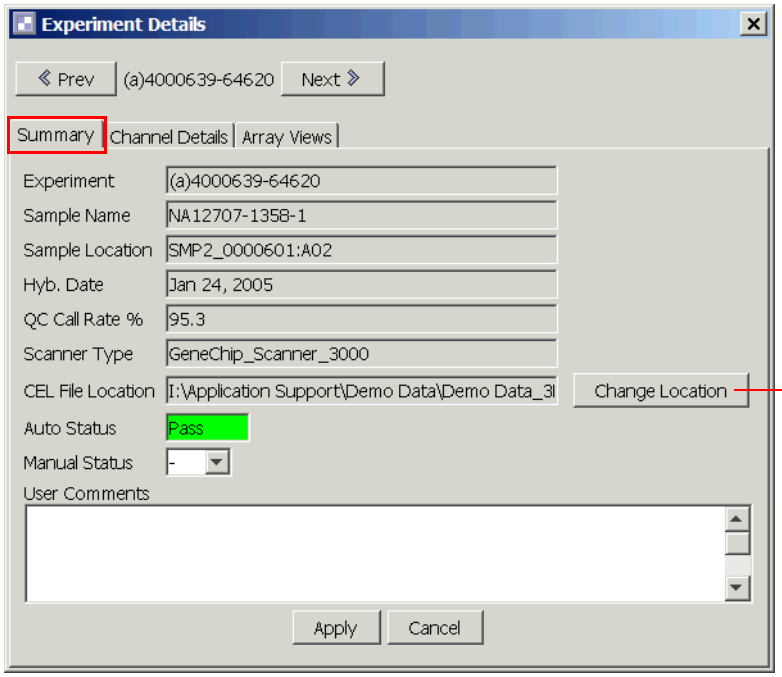


Figure 4.8 Experiment Details – Summary Tab

About the Channel Details Tab

The Channel Details tab displays metrics for each channel of an experiment (Figure 4.9).

- **Channel Name**: A .cel filename. The file contains the data for a single channel (A, C, G or T). For more information on .cel files and filenames, see *Scanning Arrays* on page 9.
- **Channel Median S/N** (signal-to-noise): The typical signal strength of each channel.
- **# Rejected Outliers**: The number of features in each channel with unreliable measurements.

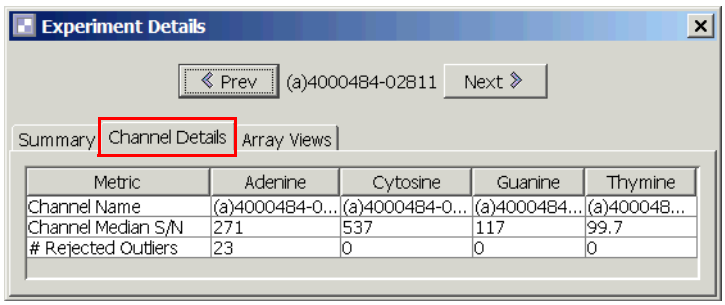


Figure 4.9 Experiment Details – Channel Details Tab

About the Array View Tab

The Array View tab (Figure 4.10) displays the array image (stored as four .cel files; one for each channel) for a single experiment. You can view the image as raw or processed data. Looking at the array image is a good first step when troubleshooting data quality. It allows you to see any obvious flaws that affected the experiment. Flaws include air bubbles, excess stain, or gridding failures. Refer to Chapter 5, *Reviewing, Working With, and Troubleshooting Data* for more information.

Each square in the image is generically referred to as a feature. Two basic types of features exist on every array:

- Assay features (used for genotyping)

- Control features (used for gridding and other functions)

For more information, see *Scanning Arrays* on page 9.

Use the Prev and Next buttons to sequentially move from one experiment to another. You might find it useful to sort experiments in the Experiment QC Summary table prior to looking at Array Views.

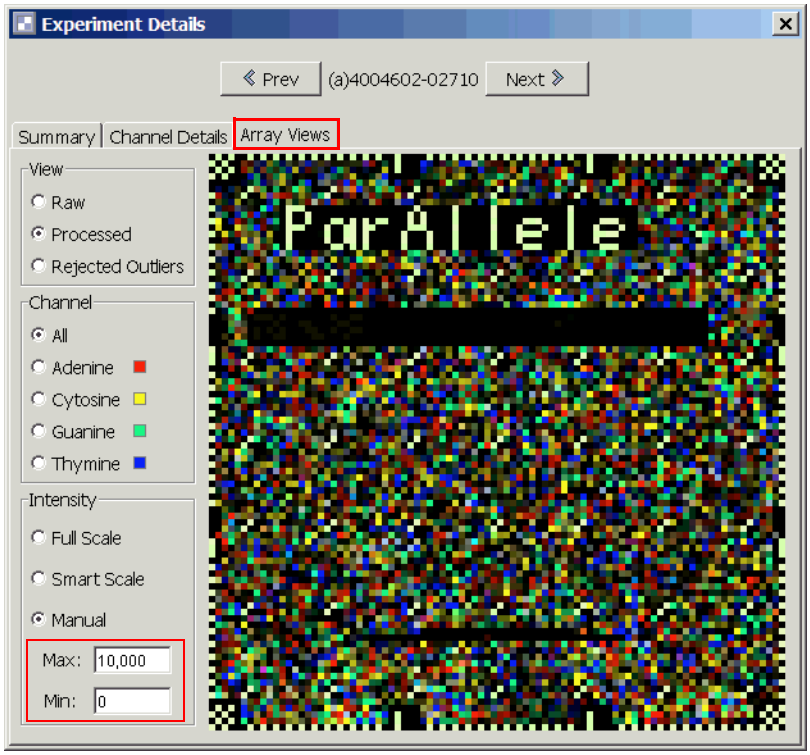


Figure 4.10 Experiment Details – Array Views Tab

View Options

The .cel file data can be viewed three ways: raw, processed, and by rejected outliers.

- **Raw:** feature intensities as reported in the .cel files.
- **Processed:** processed feature intensities used for genotyping.
- **Rejected Outliers:** features with unreliable measurements. In this view, colored features are rejected outliers; black features are not outliers and are genotyped (Figure 4.11).

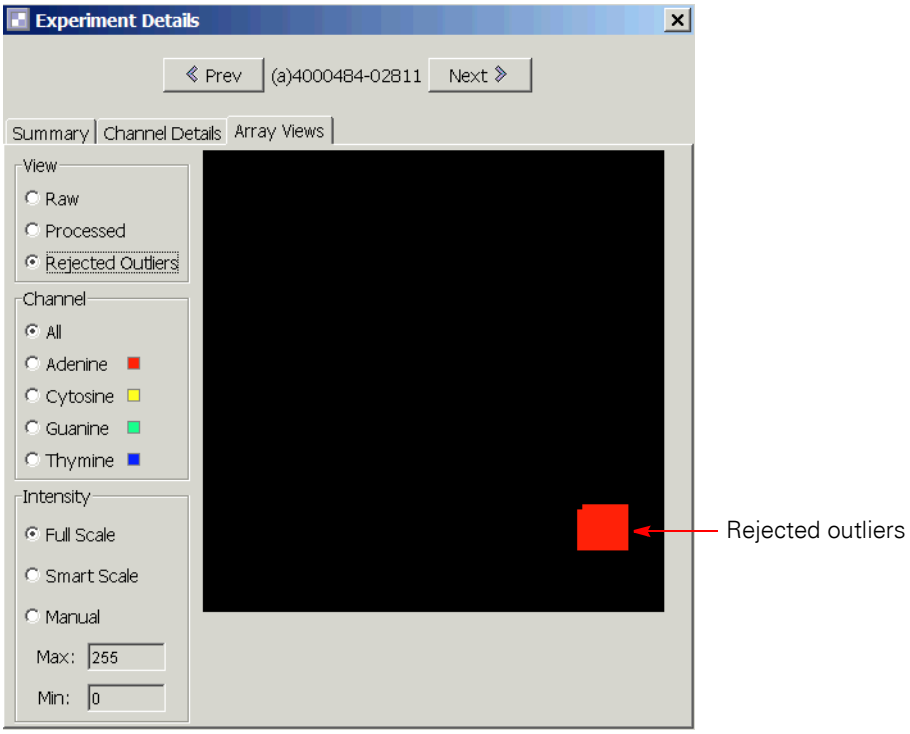


Figure 4.11 Arrays Views Tab – Rejected Outliers

Channel Box Options

You can view each channel individually by nucleotide, or you can view all four channels at once.

Intensity Box Options

Allows you to adjust the range of intensities displayed, from darkest to brightest.

- **Full scale:** shows the full range of intensities. The minimum intensity is displayed as black; the maximum intensity is displayed as the brightest color.
- **Smart scale:** displays a more narrow range of intensities. It sets the minimum to zero and the maximum to three times the average signal.
- **Manual:** allows you to select your own minimum and maximum values.

Tables and Charts for Reviewing Cluster Genotypes

Tables and Charts Available for Viewing Clusters

The following tools are available for viewing cluster results:

- Fit Results tab
- Genotype Settings tab
- Assays tab
- Experiments tab
- Samples tab
- Trios tab

Viewing Cluster Result Tables and Charts

To view cluster result tables and charts:

1. Expand the **Genotype Results** folder for your project.
2. Select the name of the cluster that contains the results you want to view.
3. In the right pane, click one of the following tabs:
   - Fit Results
   - Genotype Settings
   - Assays
   - Experiments
   - Samples
   - Trios

Fit Results Tab

The Fit Results tab displays a table that summarizes the results of the clustering operation. Refer to the *Glossary* for information on the terms displayed in the Property column (Figure 4.12).

Fit Results | Genotype Settings | Assays | Expe

Property	Value
# Experiments	92
# Genotyped Experiments	91
# Unique Samples	84
# Genotyped Unique Samples	83
# Trios	24
# Assays	2,938
# Passed Assays	2,918
Passed Assays %	99.32
Failed by Low Call Rate %	0.68
Failed by Low H.W. P-Value %	0.00
Failed by Low Repeatability %	0.00
Failed by Low Trio Concordanc...	0.00
# Repeated Sample Discordances	44
# Repeated Sample Concordances	40,419
Cluster Fit Repeatability %	99.89
# Trio Discordances	144
# Trio Concordances	65,253
Cluster Fit Trio Concordance %	99.78
Completeness %	98.75
Low Call Confidence Cut %	0.83

Figure 4.12 Genotype Results – Fit Results Tab

Genotype Settings Tab

The Genotype Settings tab displays a table that shows the settings used to generate the genotypes for a project (set of experiments). Refer to the ***Glossary*** for information on the terms displayed in the Property column (Figure 4.13).

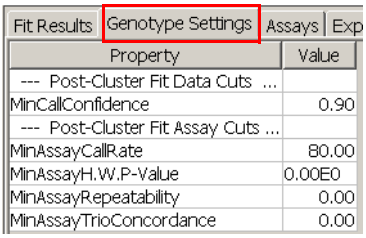


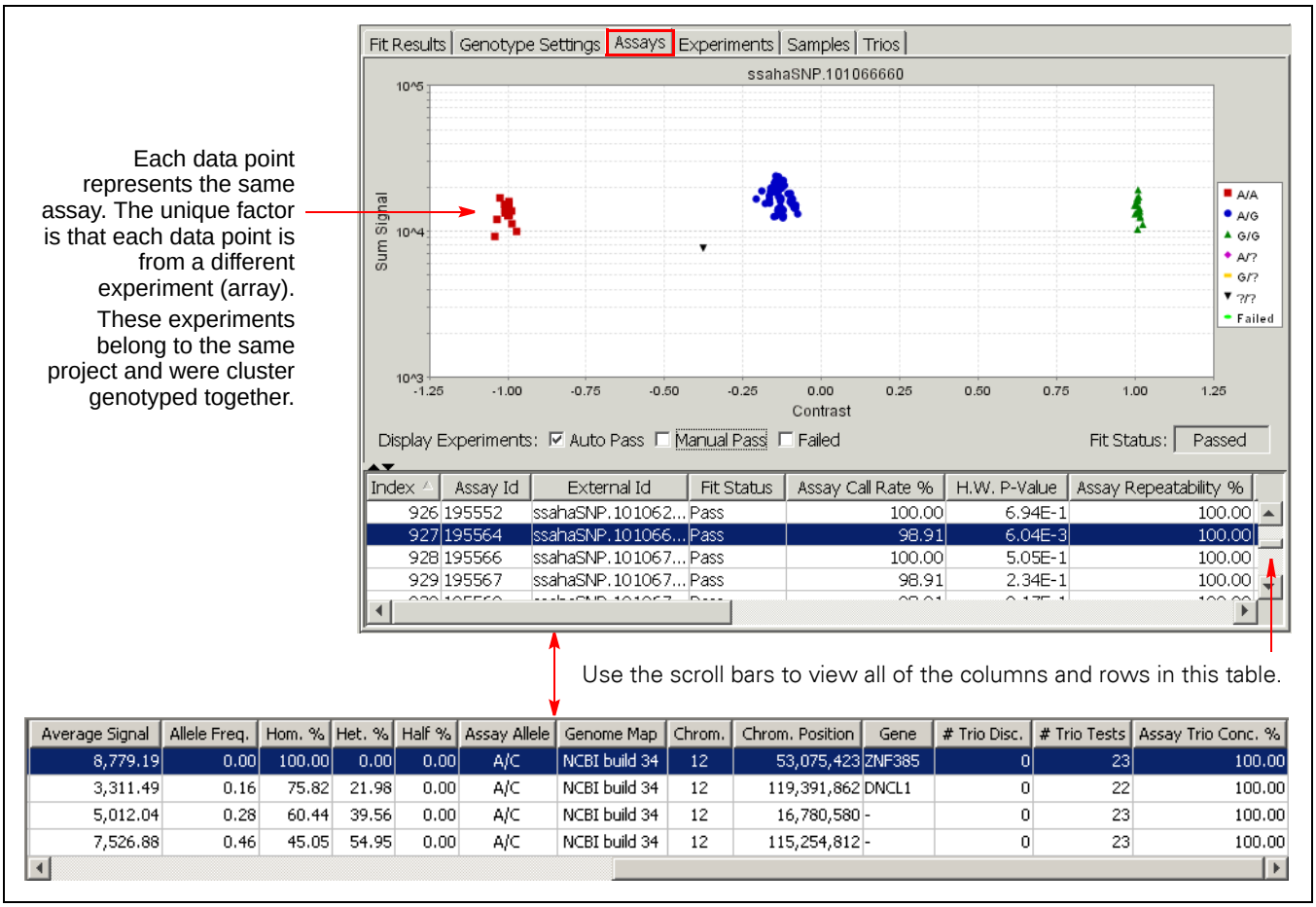
Figure 4.13 Genotype Results – Genotype Settings tab

Assays Tab

About the Assays Tab

The Assays Tab (Figure 4.14) includes:

- A chart that displays the cluster results for one assay. Each data point represents the assay performance for a different experiment (array). From this window, you can jump to the Experiment tab and view the experiment cluster results for each individual assay shown in the chart. See *Jumping to the Experiment Results for One Assay* on page 64.
- A table that includes information about each assay, such as Assay Id, Fit Status and Assay Call Rate %. Refer to the *Glossary* for information on the terms displayed in the table.



Index	Assay Id	External Id	Fit Status	Assay Call Rate %	H.W. P-Value	Assay Repeatability %
926	195552	ssahaSNP.101062...	Pass	100.00	6.94E-1	100.00
927	195564	ssahaSNP.101066...	Pass	98.91	6.04E-3	100.00
928	195566	ssahaSNP.101067...	Pass	100.00	5.05E-1	100.00
929	195567	ssahaSNP.101067...	Pass	98.91	2.34E-1	100.00

Average Signal	Allele Freq.	Hom. %	Het. %	Half %	Assay Allele	Genome Map	Chrom.	Chrom. Position	Gene	# Trio Disc.	# Trio Tests	Assay Trio Conc. %
8,779.19	0.00	100.00	0.00	0.00	A/C	NCBI build 34	12	53,075,423	ZNF385	0	23	100.00
3,311.49	0.16	75.82	21.98	0.00	A/C	NCBI build 34	12	119,391,862	DNCL1	0	22	100.00
5,012.04	0.28	60.44	39.56	0.00	A/C	NCBI build 34	12	16,780,580	-	0	23	100.00
7,526.88	0.46	45.05	54.95	0.00	A/C	NCBI build 34	12	115,254,812	-	0	23	100.00

Figure 4.14 Genotype Results – Assays Tab

Viewing the Cluster Results for One Assay (SNP)

To view the cluster results for a particular assay:

- Scroll through the list of assays.
- Click anywhere in the row of the assay you want to view.

Determining the Assay Experiment

To determine the experiment for a particular assay:

1. Move the cursor over the assay data point.
2. Hold the cursor still until a box appears.

   The box contains the genotype results for that assay for that particular experiment, as well as the experiment and sample names (Figure 4.15).

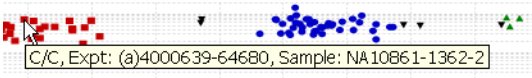


Figure 4.15 Viewing the Genotype, Experiment, and Sample Name for a Particular Assay

Jumping to the Experiment Results for One Assay

You can view the cluster results for the experiment in which a particular assay data point was included.

To view the cluster results for the experiment for a particular assay data point:

Right-click a data point and select **Display Experiment: <*experiment name*>**.

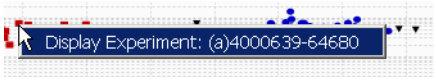


Figure 4.16 Jumping to the Cluster Results for an Experiment

The result of this action is shown below. In this example, you are now looking at the cluster results for experiment (a)4000639-64680, which is located under the Experiments tab (Figure 4.17). See *Experiments Tab* on page 65 for more information.

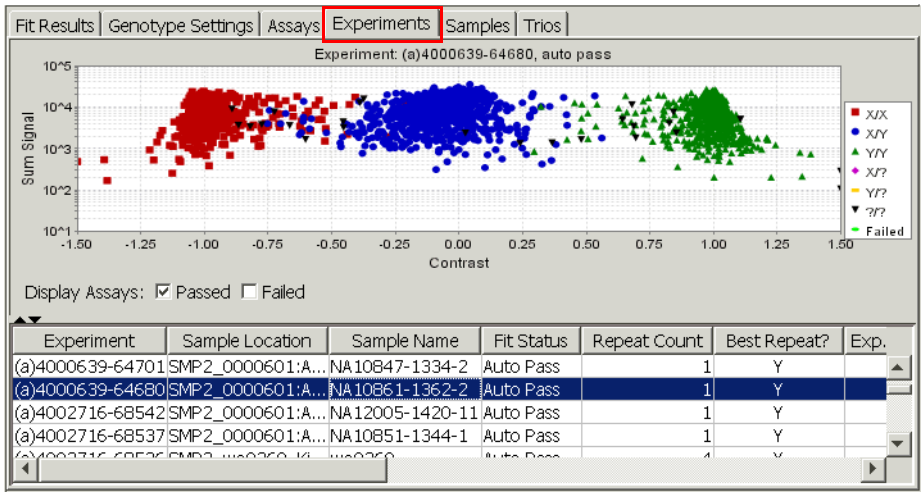


Figure 4.17 Experiments Tab

Experiments Tab

About the Experiments Tab

The Experiments Tab includes (Figure 4.18):

- A chart that displays the cluster results (signal distribution) for all of the assays for one experiment (array). Each data point in the chart represents one assay from the array.
  From this window, you can jump to the Assays tab and view the assay cluster results for a specific assay. See *Assays Tab* on page 63.
- A table that includes information about each experiment, such as Experiment name, Sample Location and Fit Status. Refer to the *Glossary* for information on the terms displayed in the table.

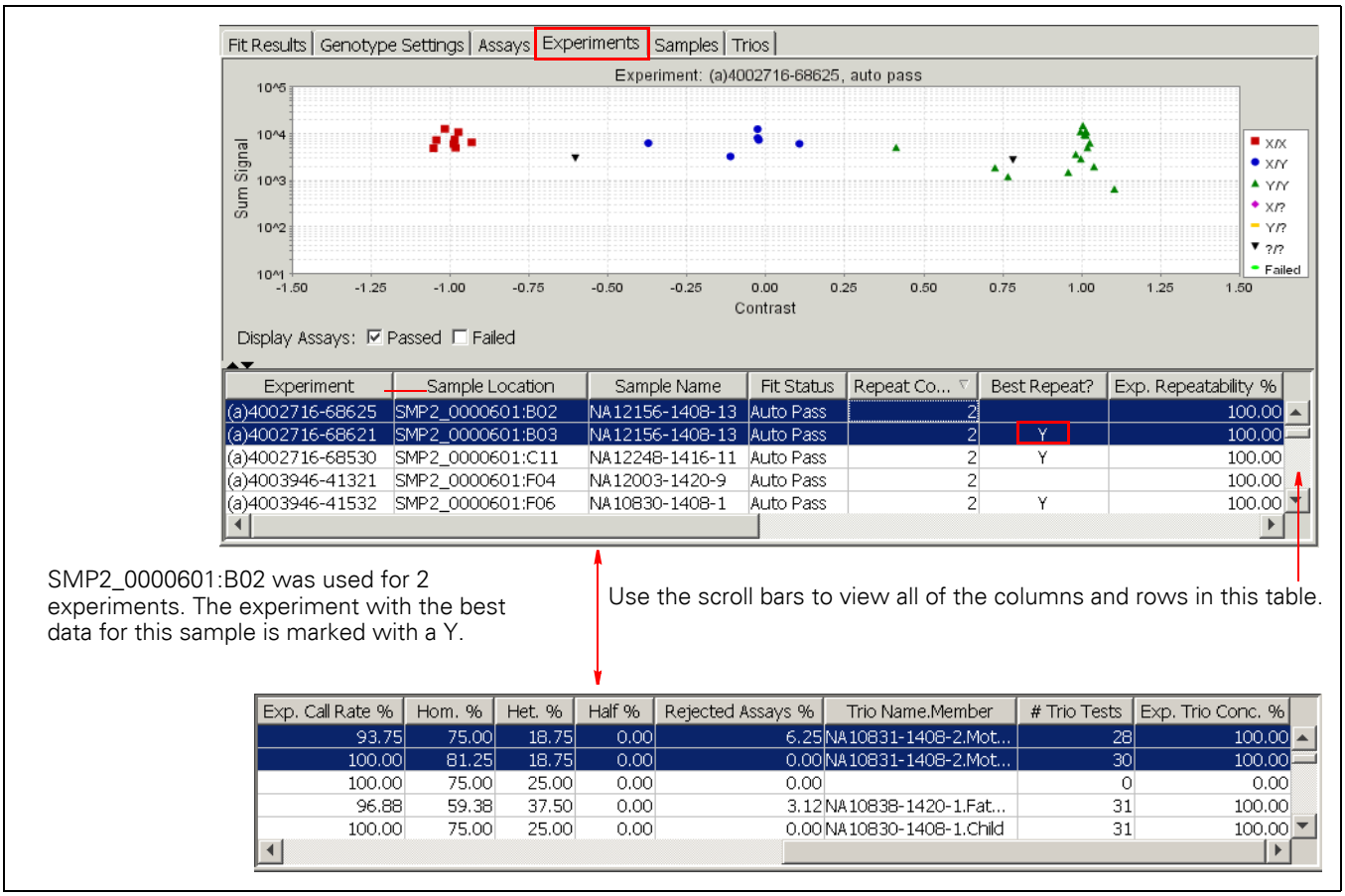


Figure 4.18 Genotype Results – Experiments Tab

Viewing the Cluster Results for One Experiment

To view the cluster results for a particular experiment:

1. Scroll through the list of experiments in the table.
2. Highlight the experiment you want to view.

Viewing Assay Information

To view general information for a particular assay:

1. Move the cursor over one of the data points.
2. Hold the cursor still until a box appears.

   The box contains the genotype results for that assay, the External Id, Assay Id and Assay Index. (Figure 4.19)

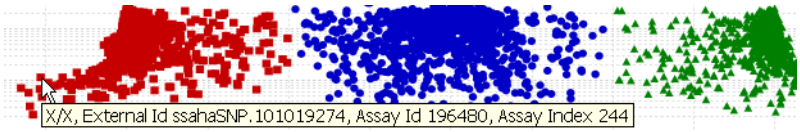


Figure 4.19 Experiments Tab – Viewing Assay Information

Jumping to the Assay Cluster Results for One Assay

You can view the cluster results for a specific assay in the experiment displayed.

To view the experiment cluster results for a particular assay:

1. Right-click a data point.
2. Select **Display Assay Index**: <assay index>.

Figure 4.20 Experiments Tab – Jumping to Assay Cluster Results

As a result of this action, you now see the cluster results (Assays tab) for this particular assay as it performed across a set of experiments. See *Assays Tab* on page 63 for more information.

Samples Tab

The Samples tab allows you to quickly see how many times a particular sample was tested. Sort on the column, Repeat Count, to list repeated samples first (Figure 4.21).

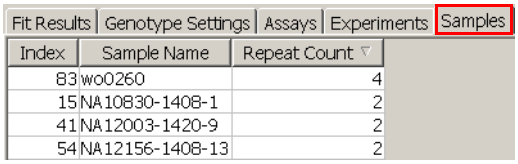


Figure 4.21 Genotype Results – Samples tab

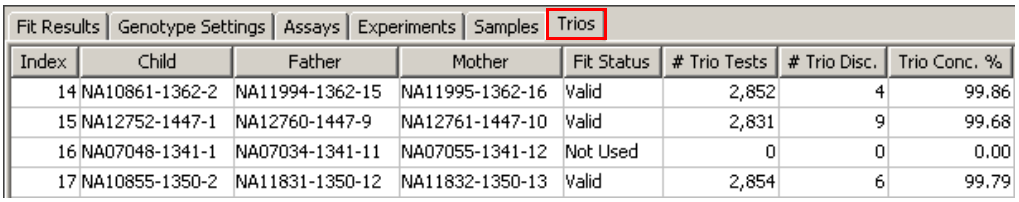
Trios Tab

The Trios tab displays the sample names for the trio members. It also reports the:

- Fit status: valid or invalid

- # Trio Tests: number of comparisons made

  Only the experiment with the highest QC Call Rate for each trio member is used in this calculation.
- # Trio Disc: the number discordances out of the number of comparisons made
- Trio Conc. (%): the concordance percent based on the number of comparisons made



Fit Results | Genotype Settings | Assays | Experiments | Samples | Trios

Index	Child	Father	Mother	Fit Status	# Trio Tests	# Trio Disc.	Trio Conc. %
14	NA10861-1362-2	NA11994-1362-15	NA11995-1362-16	Valid	2,852	4	99.86
15	NA12752-1447-1	NA12760-1447-9	NA12761-1447-10	Valid	2,831	9	99.68
16	NA07048-1341-1	NA07034-1341-11	NA07055-1341-12	Not Used	0	0	0.00
17	NA10855-1350-2	NA11831-1350-12	NA11832-1350-13	Valid	2,854	6	99.79

Figure 4.22 Genotype Results – Trios tab

Deleting Projects and Project-Related Information

Deleting a Project

To delete a project, you must delete the information in reverse order from which it was created.

For example, if a project has results in the Genotype Results folder, you must:

1. Delete each cluster genotype that is in the Genotype Results folder.
2. Delete all of the experiments listed under the Array Data icon in the Experiment QC Summary.

   Experiments may be deleted individually or all at once. If you are using GCOS, you may also want to go to GCOS Manager and delete all experiments associated with the project. If you are using AGCC, you may also want to go to the folder containing ARR and CEL files, and move or delete these as well.
3. Delete the arrays from the Arrays folder.
4. Delete each plate in the Hyb Plates folder.
5. Delete each plate in the Label Plates folder.
6. Delete each plate in the Assay Plates folder.
7. Delete each plate in the Anneal Plates folder.
8. Delete each plate in the Sample Plates folder.
9. Delete the project icon.

Deleting Cluster Genotype Results

To delete cluster genotype results:

1. Right-click the cluster name.
2. Select **Delete**.
3. Click **Yes**.

Deleting Experiments

Experiments can be deleted only if they are not included in any cluster genotyping results.

To delete experiments individually:

1. Click the **Array Data** icon for your project.
2. Click the Experiment QC Summary tab.
3. Right-click anywhere in the row of the experiment you want to delete and select **Delete**.

   If you cannot select Delete, the experiment is included in cluster genotype results. You must delete the cluster genotype results before you can delete the experiment.
4. Click **OK**.

To delete all experiments at once:

1. Right-click the Array Data icon for your project and select Delete All Experiments.
2. The status bar at the bottom of the the application window will show the deletion progress. Do not force quit the application during this operation, or all these experiments will be restored. Since restoring all the experiments takes some time, the application may be unresponsive until SQL Server has finished this task.

Deleting Arrays

You should delete arrays only after the associated Array Data experiments have been deleted.

To delete arrays:

1. Expand the arrays folder.
2. Right-click an array and select **Delete**.
3. Click **Yes**.
4. Repeat this procedure to continue deleting arrays.

Deleting Anneal, Assay, Label and Hyb Plates

You can delete plates as long as there are no arrays hybridized with samples that originated from these plates.

To delete these plates, you must delete them in reverse order starting with the Hyb Plate. For all plates, right-click the plate and select Delete.

Deleting Sample Plates

You can delete a sample plate as long as it is not included in a project. If referenced by any projects, you must first remove the sample plate from those projects, then delete the sample plate.

To delete a sample plate:

1. Delete any projects that reference the sample plate.

   See *Deleting a Project* on page 68.
2. Open the menu in the left pane of the GTGS window and select **Samples**.
3. Expand the **Sample Plates** folder.
4. Right-click the sample plate name and select **Delete**.

   You can also select **Edit** → **Delete**. If you cannot select Delete, the sample plate has been used in a project that still exists.
5. Click **Yes**.

Deleting Sample Info

If a sample name in the Sample Info table does not exist in any Sample Plate, you can remove these samples.

To delete Sample Info:

1. Open the menu in the left pane of the GTGS window and select **Samples**.
2. Right-click the Sample Info icon and select **Remove Unused Sample Info**.
3. Click **Yes**.

Deleting Assay Panels

You can delete an assay panel as long as it is not included in a project. If referenced by any projects, you must first delete those projects, then delete the assay panel.

To delete an assay panel:

1. Delete any projects that reference the assay panel.
   See ***Deleting a Project*** on page 68.
2. Open the menu in the left pane of the GTGS window and select **Assay Panels**.
3. Expand the **Assay Panels** folder.
4. Right-click the assay panel name and select **Delete**.
   You can also select **Edit** → **Delete**.
5. Click **Yes**.

Working With Tables

What You Can Do With Tables

Below is a list of what you can do with tables. Instructions for each action are included in this section.

- Export tables
- Find information in a particular column
- View exported tables
- Adjust column widths
- Resize windows
- Sort tables
- Customize the sort order of tables

Exporting Tables

Tables in GTGS can be exported and saved as tab-delimited .txt files.

To export a table:

1. Display the table.
2. Open **File** → **Export Table**.
   The Save As window appears.
3. Enter a name for the file in the File name field.
4. Open the **Save In** menu and select the location where you want the table saved.
5. Click **Save**.

Finding Information in Columns and in the Navigation Tree

You can search for information in table columns and in the navigation tree (left pane of the GTGS window).

To find information in columns or in the navigation tree:

1. Do one of the following:
   A. Open **Edit → Find** to search in the navigation tree or a column
   B. In the table you want to search, right-click the column heading and select **Find in column** (Figure 4.23).

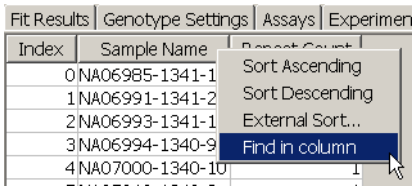
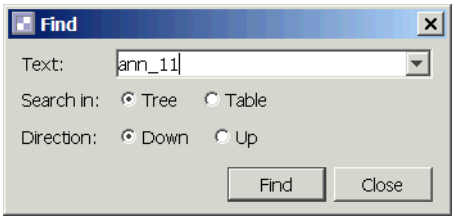


Figure 4.23 Find in Column Command

2. Enter the following:
   - The text you want to search for in the Text field.
   - Select Up or Down.
   - If using the Find box, select Tree or Table (Figure 4.24).
3. Click **Find**.

Find box when using Edit → Find



Search box when using Find in column

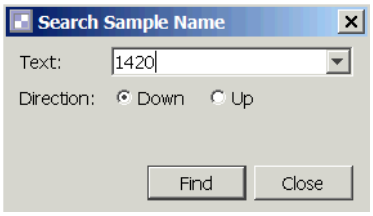


Figure 4.24 Entering and Selecting Find Parameters

Viewing Exported Tables

We recommend opening tables in Microsoft® Office Excel®.

To view an exported table in Microsoft Office Excel:

1. Right-click the file icon.
2. Select **Open With →Microsoft Office Excel**.

Adjusting Column Widths

To adjust the width of column in a table:

1. Move and hold the cursor between column names.
2. Click, hold and drag the column divider to the desired width (Figure 4.25).

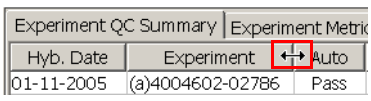


Figure 4.25 Adjusting Column Widths

Resizing Windows

All of the windows in GTGS can be resized. Note that some elements may not be displayed if the window is resized too small. For example, some of the columns may not be displayed.

To resize a window:

1. Move and hold the cursor on a corner or on the side of the window.

The cursor changes to a double-headed arrow as shown below (Figure 4.26).

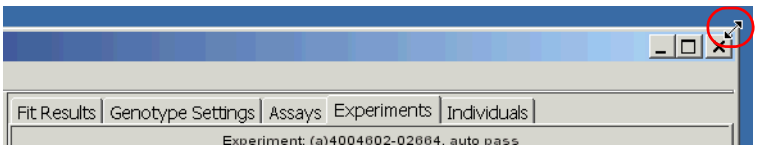


Figure 4.26 Resizing Windows

2. Click, hold and drag the window the desired size (Figure 4.27).

NOTE: Some windows display a chart and a table (for example, the Arrays tab display). As shown below, you can move the dividing line between the chart and table up or down to resize either element.

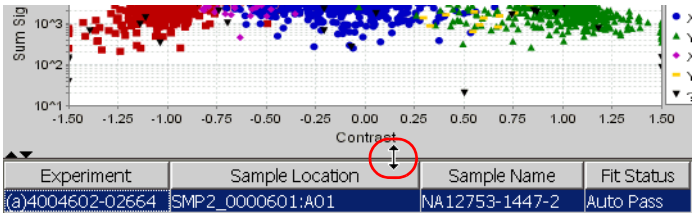


Figure 4.27 Resizing Windows

Sorting in Tables

The information displayed in tables can be sorted in ascending or descending order based on the column heading.

To sort the information in a table:

1. Click a column heading.
2. Click the column heading again to change the sort order (Figure 4.28).

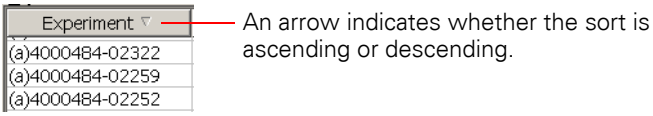


Figure 4.28 Sorting the Information in a Column

Customizing the Sort Order of a Table

You can customize the sort order for columns in tables that contain text or integers. This option is particularly useful when working with the Assays table (see *Assays Tab* on page 63).

NOTE: Customized sorts are not saved.

To customize the sort order of a table:

1. Export a table.
2. Open the table Microsoft Office Excel.
3. Set up a customized sort and save the table.
4. Read the table back into GTGS as follows:

   A. Right-click a column heading in GTGS and select **External Sort** …

The column you select must exist in the Microsoft Office Excel file you import (Figure 4.29).

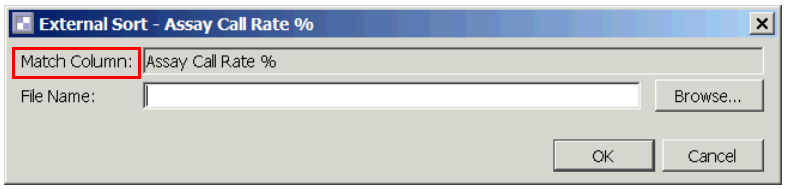


Figure 4.29 Customizing the Sort Order of a Table

B. Click the **Browse** button.

C. Locate and select the Microsoft Office Excel table.

D. Click **Open**.

E. Click **OK**.

The data in the GTGS table will be sorted in the same order as in the file. Table rows that do not exist in the file are moved to the bottom of the table.

Working With Charts

What You Can Do With Charts

Below is a list of what you can do with charts. Instructions for each action are included in this section.

- Resize windows
- View information for a particular assay
- Zoom in on an entire window or particular experiment
- Zoom out in a window or from a particular experiment
- Print a chart
- Save a chart as a graphic

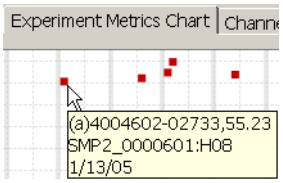
Resizing Windows

See ***Resizing Windows*** on page 72.

Viewing Information for a Particular Assay

To view information for a particular assay:

Hold the cursor over the feature for a few seconds. A small window appears (Figure 4.30).



Information displayed is:

- Experiment name
- X and Y axes coordinates
- Sample plate:well
- Hyb date

Figure 4.30 Viewing Information for a Particular Assay

Zooming In and Out

To zoom in on a window:

1. Do one of the following:
   - Click and drag the cursor from the top left to bottom right of the window.
   - Right-click in the chart to view more options. Select Zoom In; then select Both Axes, Horizontal Axis, or Vertical Axis (Figure 4.31).

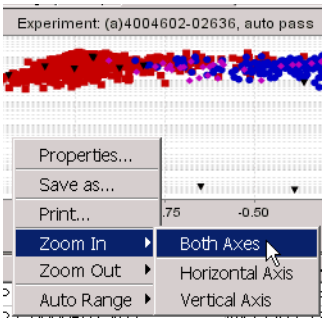
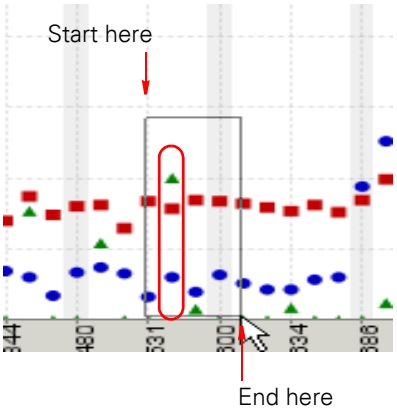


Figure 4.31 Zooming In

To zoom in on a particular experiment:

1. Click and drag the cursor to draw a box around the sample (Figure 4.32.)
2. Repeat if necessary to zoom in even further.

Step 1. Draw a box around the sample by clicking and dragging the cursor from the top left to the bottom right area around the sample. The sample we want to zoom in on is circled in red (a)4000484-02568.



Step 2. Repeat Step 1 if necessary to zoom in even further.

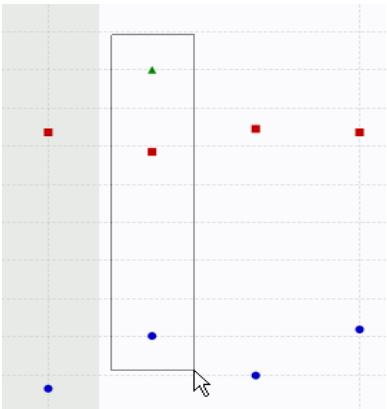


Figure 4.32 Zooming in on a Particular Sample – Method 1

To zoom out from a particular experiment or in a window:

1. Do one of the following:
   - Click and drag the cursor from the bottom right to the top left of the window.
   - Right-click in the chart to view more options.
     - Select **Zoom Out**; then select **Both Axes**, **Horizontal Axis**, or **Vertical Axis**.
     - Select **Autorange → Both Axes**. The chart will return to full scale.

Printing Charts

To print a chart:

1. Right-click and hold in the chart until a menu appears (Figure 4.33).
2. Select **Print**.
3. Specify options as desired in the Print Setup window.
4. Click **OK**.

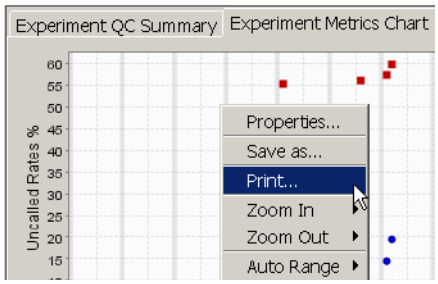


Figure 4.33 Printing Charts

Saving Charts as Graphics

To save a chart as a graphic:

1. Right-click and hold in the chart until a menu appears (Figure 4.33).
2. Select **Save as …**
3. Enter a filename for the graphic.
4. Select a location to store the graphic.
5. Click **Save**.
   The chart is saved as a .png file to the specified location.

Chapter 5

REVIEWING, WORKING WITH, AND TROUBLESHOOTING DATA

Reviewing Data Quality Prior to Genotyping

About Reviewing Data Quality

Reviewing the quality of processed data prior to cluster genotyping is optional. You can proceed directly to cluster genotyping once you have imported and processed your experiment data, or you can continue to process additional samples and add them to a project prior to cluster genotyping.

Reasons that you may want to review the data quality include the following:

- You might want to override the pass or fail status that is automatically assigned to each experiment by Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS)
- You can observe trends that may occur in data quality
- You may find data that is helpful for troubleshooting (for example, identifying assay performance issues or sample tracking errors)

About the Data Used for Cluster Genotyping

The essence of cluster genotyping is that data is compared across a set of experiments to obtain the most accurate and complete data. Therefore, the genotype calls for a particular experiment can be influenced by the other experiments in a set. For this reason, it is important to remove poor quality data from a set of experiments prior to cluster genotyping.

If enough experiments with poor quality data are not excluded, genotype call quality for the remaining experiments can be adversely effected. To help ensure that data quality is acceptable for cluster genotyping, GTGS automatically applies a set of quality control metrics to the data from each experiment. The purpose of this process is to identify and reject experiments with poor quality data by applying.

About Data Processing and Quality Control Metrics

Data processing occurs as the array data is extracted from .cel files and imported into GTGS. During processing, GTGS calculates and compares the results of each experiment against the following set of quality control metrics (see Table 5.1):

- QC Call Rate
- QC Half Rate
- Signal to Background (median ratio)
- Controls Coefficient of Variation (Controls CV %)

Based on this analysis, GTGS automatically assigns the status of Pass or Fail to each experiment. The threshold for each of these metrics must be met for an experiment to be assigned the status Pass. If any metric is not met, the experiment is assigned the status Fail.

Table 5.1 Quality Control Metrics

Quality Control Metric	Description	Threshold
QC Call Rate %	Percentage of assays clearly genotyped. (Signal in shaded areas of Figure 5.1.)	Must be ≥ 80%
QC Half Rate %	Percentage of assays marginally genotyped. (Signal in white areas of Figure 5.1.)	Must be ≤ 10%
Signal to Background	Median ratio of assay allele/non-allele channel signals (median signal to background ratio) As shown in Figure 5.1, calculated for each assay on an array as: (S1 + S2) / (B1 + B2)	Must be ≥ 20
Controls CV %	Percent coefficient of variation of control features across an array.	Must be ≤ 30%
Rejected Outliers	Assays that meet at least one of the following criteria: • Saturate the detector. • Are in a region of the array with an unusual baseline. • Have a low allele/non-allele signal (low signal/background) for a single feature. • Have unusually high variation in pixel intensities within a feature.	N/A

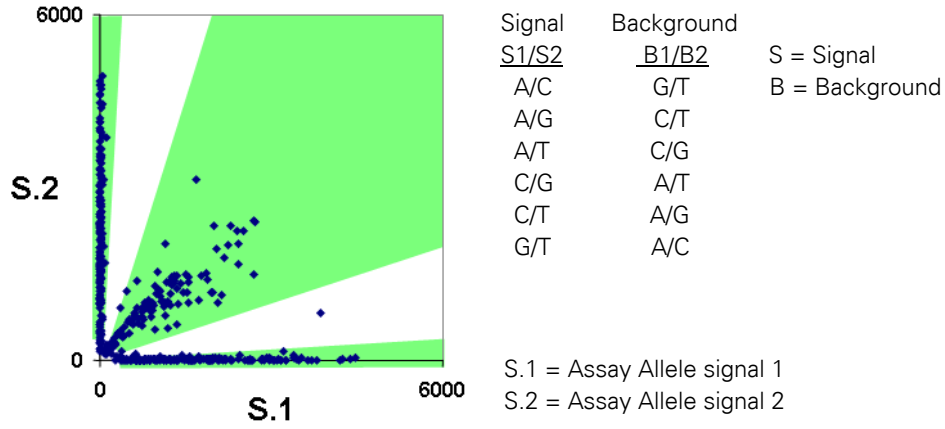


Figure 5.1 Normalized Data for One Experiment with QC Thresholds

Tools Available For Reviewing Your Data

The following tools are available in GTGS for reviewing your data prior to cluster genotyping:

- Experiment QC Summary
- Experiment Metrics Chart
- Channel Metrics Chart
- Experiment Details including:
  - Summary
  - Channel Details
  - Array View

For a description of these tools and how to work with them, refer to Chapter 4, *About Charts and Tables*.

Recommended Workflow for Data Review

Recommended Workflow

We recommend that you review your data in the order presented below.

1. Identify any experiments for which data was not imported (skipped experiments.) Troubleshoot the cause and import the data if it is available.
   See *Identifying Skipped Experiments* on page 81.
2. Determine if any arrays were scanned multiple times. Select the best scan.
   See *Working with Rescans* on page 84.
3. Review the pass/fail status of the most recent experiments in Experiment QC Summary.
   See *Reviewing the Pass/Fail Status of Experiments* on page 85.
4. If your assay panel is designed to identify gender, compare the inferred gender with the known gender by looking at the Experiment QC Summary.
   See *Comparing Known Gender with the Inferred Gender* on page 87.

Identifying Skipped Experiments

How To Identify Skipped Experiments

Two methods are available for identifying skipped experiments. You can:

- View the Import Experiment Data window as data from .cel files is being imported and processed.
- View the Experiment QC Summary table.

Viewing the Import Experiment Data Window

Experiment data is imported to GTGS and processed via the Import Experiment Data window. While the software is importing and processing data, check the Processed/Skipped/Total field to see if any experiments were skipped. In the example shown in Figure 5.2, one experiment was skipped.

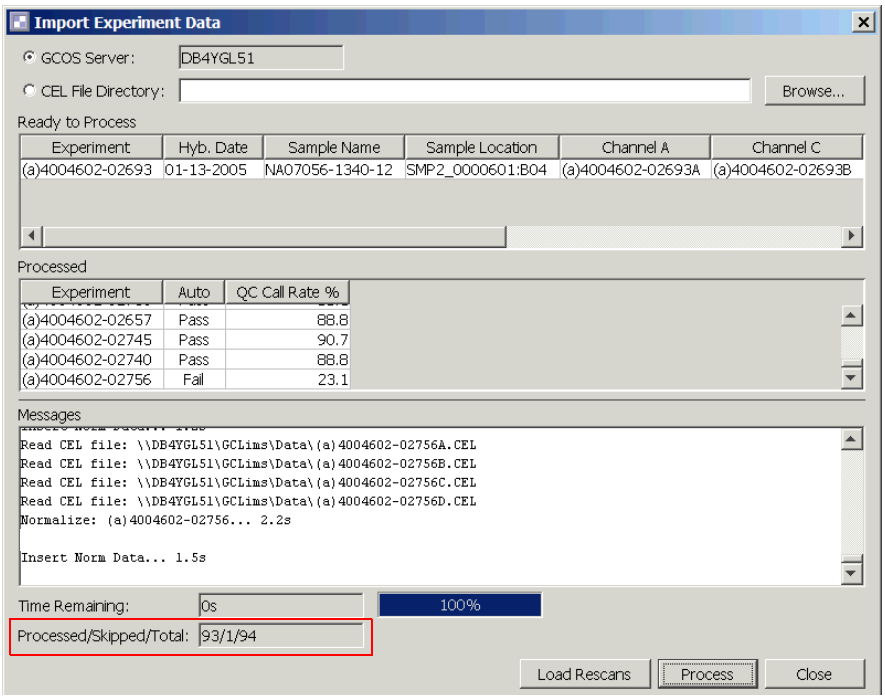


Figure 5.2 Identifying Skipped Experiments in the Import Experiment Data Window

For instructions on troubleshooting skipped experiments via the Import Experiment Data window, see *About Skipped Experiments* on page 32.

Viewing the Experiment QC Summary Window

If the Import Experiment Data window has already been closed, you can identify skipped experiments by looking at the Experiment QC Summary table. The columns that hold values for the QC metrics (QC Call Rate, QC Half Rate, and so on) will be empty as shown in Figure 5.3.

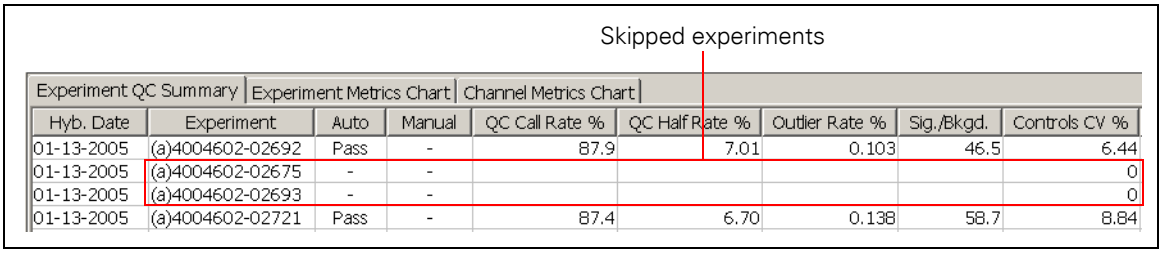


Figure 5.3 Identifying Skipped Experiments in the Experiment QC Summary Table

To troubleshoot skipped experiments found in the Experiment QC Summary table:

1. Select the Array Data icon for your project.
2. In the Experiment QC Summary table, look for rows where the QC Metric columns are empty (Figure 5.3).
3. Determine why the data was not imported as follows:

   A. Right-click the **Array Data** icon for the project and select **Import Experiment Data**.

The Import Experiment Data window appears.

B. Look at the **Ready to Process** field for experiments that were skipped.

The names of the experiments you identified in the Experiment QC Summary table will appear in the Ready to Process list (Figure 5.4).

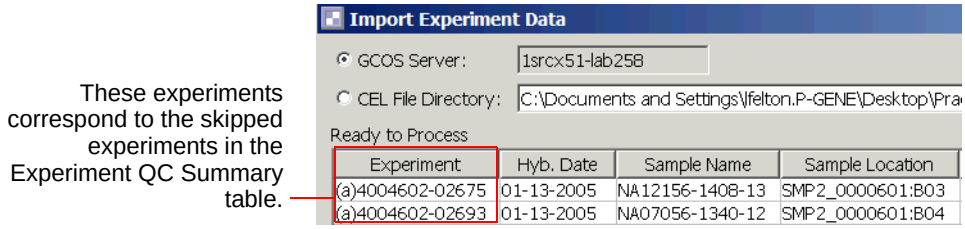


Figure 5.4 Troubleshooting Skipped Experiments

C. Look at the **Messages** field and do one of the following:

- If the **Messages** field is empty, click **Process**. The software will run through the list of experiments again and will display error messages related to the skipped experiments.
- If the Messages field is not empty, scroll through the messages for the ones related to the skipped experiments.

D. See *About Skipped Experiments* on page 32 for further instructions on troubleshooting these experiments.

Working with Rescans

Identifying Rescans

To identify rescans:

1. Select the **Array Data** icon for your project.
2. In the **Experiment QC Summary** table, sort by **Experiment Name**.
3. Scroll through the list of experiment names, looking for experiments with the extension _<*number*>.
   In the example below, (a)4003944-43367 and (a)4003944-43367_2 are scans of the same experiment. (a)4003944-43367_2 is the rescan (Figure 5.5).

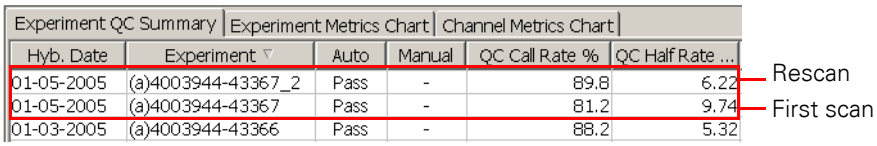


Figure 5.5 Identifying Rescans

Selecting the Best Scan

To select the best scan:

1. Look at the QC Call Rate for all scans of the same experiment.
2. Do one of the following:
   - Identify the experiment with the best QC Call Rate.
   - If QC Call Rates are identical for all scans, identify the scan with the lowest QC Half Call Rate.

   For the example above, the rescan (a)4003944-43367_2 has a higher QC Call Rate.
3. Manually fail or delete the experiments with lower data quality.
   Using the example above, you would either delete experiment (a)4003944-43367, or manually change its status to Fail. As such, only the data for the rescan is used for cluster genotyping.
   See *Changing the Status of an Experiment* on page 97 for instructions.

Reviewing the Pass/Fail Status of Experiments

Reviewing Experiment Status

For a single project, data will typically be imported into GTGS multiple times. As new data is imported and processed, the experiment information is added to the bottom of the Experiment QC Summary table.

To review the pass/fail status of experiments:

1. Click the **Array Data** icon for your project.
2. In the Experiment QC Summary table, sort on the **Hyb Date** column so that the latest imports are displayed first.
3. Look at the pass/fail status of the experiments you have not yet reviewed.
4. If all or most experiments passed, you can:
   - Continue collecting experiment data until all of the samples for a particular project have been processed.
   - Perform cluster genotyping.

   NOTE: Although preliminary cluster genotyping can be performed at any time, the results will be most meaningful if you cluster the data from at least 80 unique experiments. Perform final cluster genotyping when all of the data for a particular project has been imported and processed by GTGS.

   - Determine which experiments might benefit from rescanning.

     See *Identifying Arrays that Could Benefit from Being Rescanned* on page 85 for more information.
5. If too many experiments are failing, you can investigate the possible causes.

   See *Troubleshooting Experiment Failures* on page 89 for more information.

Identifying Arrays that Could Benefit from Being Rescanned

Identifying Arrays for Rescanning

Arrays that can benefit from rescanning are typically arrays that have high outlier rates. Causes for high outlier rates include:

- Excess stain on the array, which is seen as high or nonuniform baseline in the array image.
- Manually regridding of an array was not done properly.

To identify arrays for rescanning:

1. In the Experiment QC Summary, sort on Outlier Rate %.
2. View the range of values in the Outlier Rate column to determine the range for a particular project.
3. Note the experiments with high outlier rates (> 0.5%).
4. Return to the **Experiment QC Summary** table and open the **Array View** for each experiment you want to investigate as follows:
   A. Right-click anywhere in the row of the experiment and select **Display Experiment Details**.
   B. Click the **Array Views** tab.
   C. In the View box, select the **Rejected Outliers** button.
5. On the Instrument Control Workstation, open the .dat file for channels with a significant number of rejected outliers and look for the following:
   - A grid that was not applied properly (see Figure 5.7).
   - Areas of high or nonuniform baseline (Figure 5.6) caused by excess stain on the array.

6. If the grid was not applied properly:
   - **A.** Delete the experiment from GTGS.
   - **B.** Import the Experiment List from GTGS.
     The sample and array tracking information for this experiment must be added to GTGS again.
   - **C.** Manually realign the grid in GeneChip® Operating Software (GCOS or Command Console Viewer).
   - **D.** Right-click the .dat file and select **Recalculate CEL Intensity** to generate a new .cel file.
   - **E.** Import the new .cel file into GTGS and process the data.
   - **F.** Review the results.
7. If there are areas of high or nonuniform background, you can try rewashing and rescanning the array.
   Refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide* for instructions on rewashing and rescanning arrays.

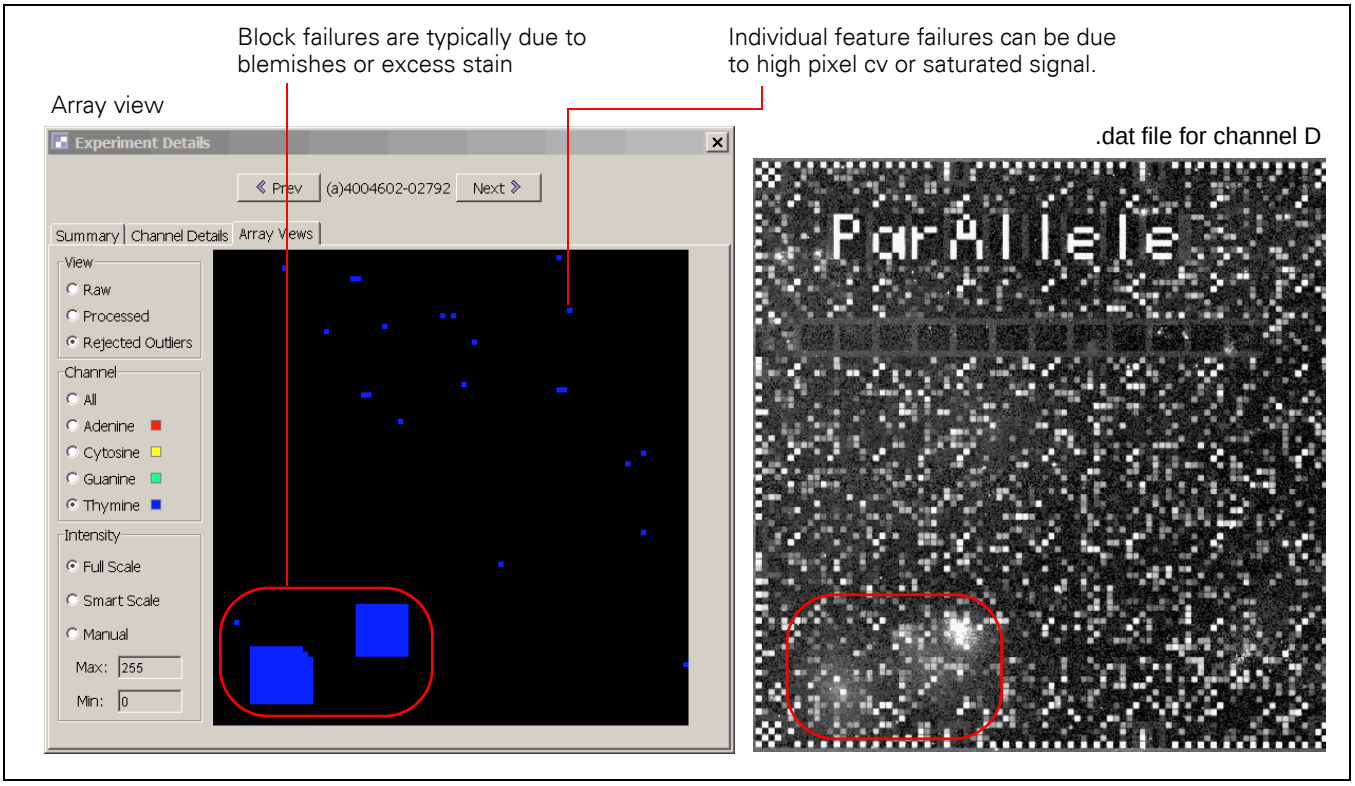


Figure 5.6 High Outlier Rate Caused by Excess Stain

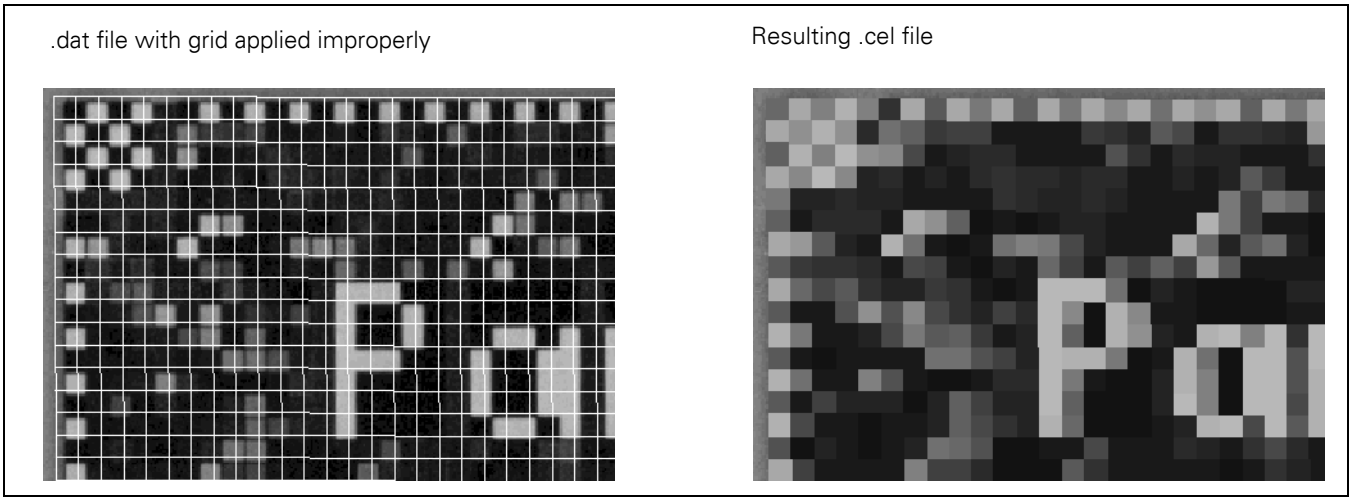


Figure 5.7 Example of High Outlier Rate Caused by Improper Gridding

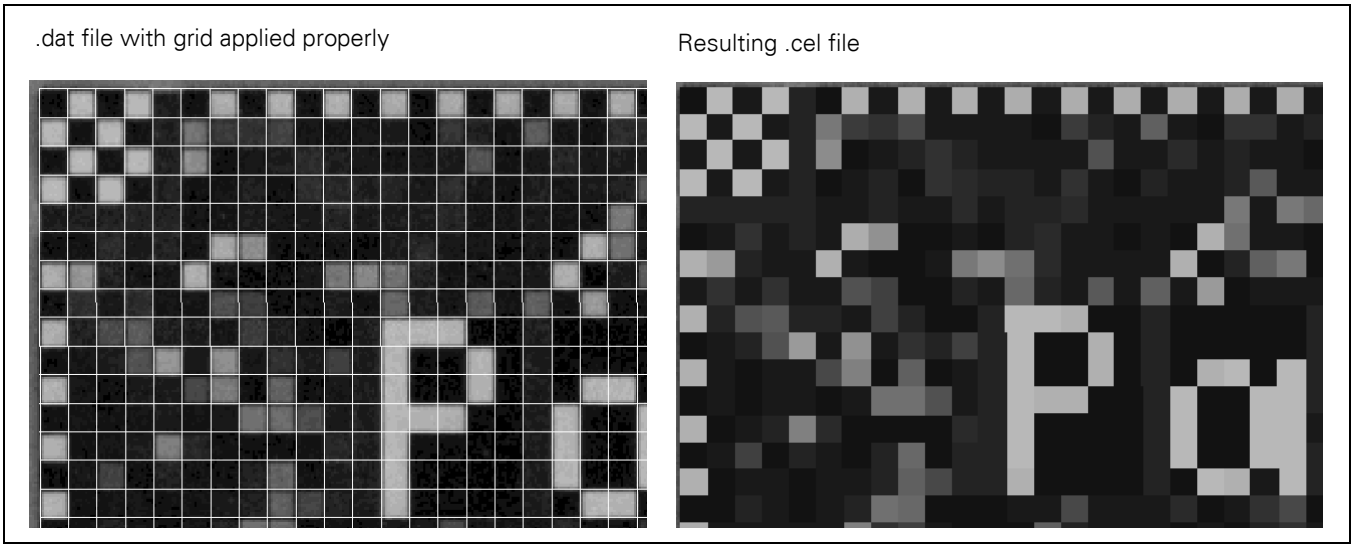


Figure 5.8 Example of Proper Gridding

Comparing Known Gender with the Inferred Gender

If your assay panel was designed to identify sample gender, you can compare the known gender with the inferred gender. To view this information, select the Array Data icon for your project and look at the Experiment QC Summary table.

The *Inferred Gender* column displays the gender called by GTGS based on the experiment results for a particular sample (Figure 5.9). Inferred gender is determined by special control assays that look for Y chromosome SNPs.

The *Gender* column displays the known gender that was assigned to a particular sample in the Sample Info file. Refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide* for more information about Sample Info files.

If the inferred gender is different from the known gender, a sample tracking error may have occurred. We recommend:

- Manually failing experiments with a gender mismatch

- Rerunning the samples

Experiment QC Summary | Experiment Metrics Chart | Channel Metrics Chart

Controls CV %	Hom %	Inferred Gender	Gender	Sample Name	Sample Location
16.3	80.31	Male	Male	NA10838	SMP_20050817_APPS-0...
13.1	80.58	Male	Male	NA10838	SMP_20050817_APPS-0...
28.0	80.31	Male	Male	NA10838	SMP_20050817_APPS-0...
14.3	79.66	Male	Male	NA10838	SMP_20050817_APPS-0...
49.5	79.83	Female	Female	NA12004	SMP_20050817_APPS-0...

Figure 5.9 Comparing the Inferred and Known Genders of Samples

Troubleshooting Experiment Failures

Most Common Types of Failures

The most common experiment failures are:

- **Sporadic failures**: Experiment failures that occur randomly with no apparent relation to each other.
- **Row or Plate failures**: Failure of an entire row or all of the samples from a plate (Assay, Label or Hyb).

Sporadic Failures

Sporadic failures are random experiment failures within a project that bear no apparent relation to each other. They are typically seen in the region indicated in Figure 5.10.

To view sporadic failures:

1. Select the **Array Data** icon and click **Experiment Metrics Chart** tab.
2. View the data sorted as follows:
   - Call Rates % by Hyb Date
   - Call Rates % by Sample Location

   Failures are most likely sporadic if they are not grouped in either view.
3. Optional: Rerun samples that have failed.
4. For more information on troubleshooting experiment failures, refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide.*

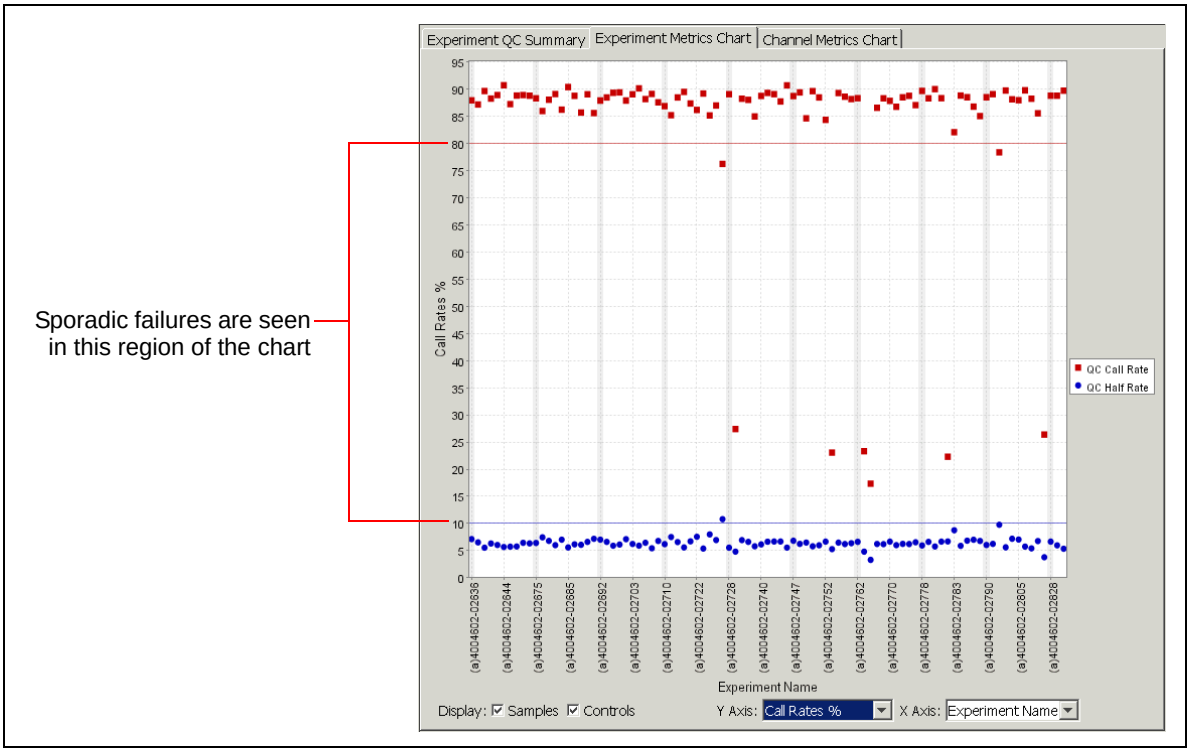


Figure 5.10 Example of Sporadic Failures

Row or Plate Failures

If most or all of the samples from a particular row of a plate fail, it is referred to as a plate failure (Figure 5.11). Plate failures are typically caused by errors made while performing the protocol.

To view row or plate errors:

1. Select the **Array Data** icon and click **Experiment Metrics Chart** tab.

2. View the data sorted as follows:
   - Call Rates % by Hyb Date
   - Call Rates % by Sample Location

   Plate failures are most likely grouped together.
3. Optional: Repeat the assay on samples that have failed.
4. For more information on troubleshooting experiment failures, refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide*.

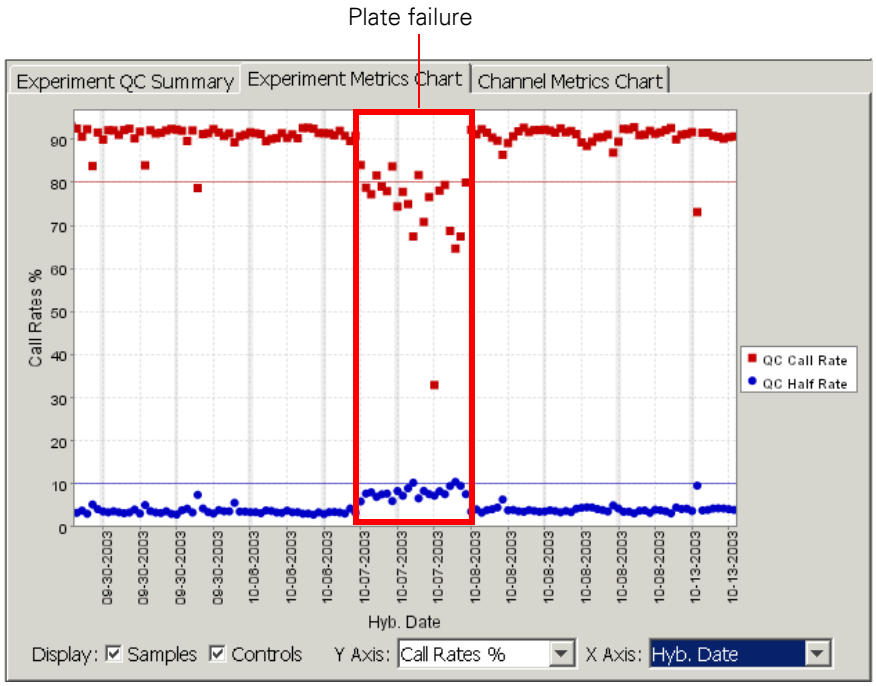


Figure 5.11 Example of Plate Failure

Reviewing, Troubleshooting, and Refining Cluster Genotyping Results

Reviewing Cluster Genotyping Results

To review your cluster genotyping results, look at the Fit Results for a particular cluster.

- If the specifications for your study were met, you can:
  - Export genotypes.
  - Investigate refining cluster genotyping results and generate a new cluster.

    See the following as appropriate for more information:
  - *Generating and Exporting Genotypes* on page 35.
  - *Refining Cluster Genotyping Results* on page 96.
- If the specifications were not met, you can:
  - Review cluster genotyping results to determine possible causes
  - Resolve any problems found
  - Generate a new cluster

    See the workflow described under *Troubleshooting Cluster Genotyping Results* on page 92.

Viewing Fit Results

To view Fit Results:

1. Expand your project; then expand the **Genotype Results** folder.
2. Highlight the cluster you want to review.
3. Click the **Fit Results** tab.

Fit Results | Genotype Settings | Assays | Expe

Property	Value
# Experiments	92
# Genotyped Experiments	91
# Unique Samples	84
# Genotyped Unique Samples	83
# Trios	24
# Assays	2,938
# Passed Assays	2,918
Passed Assays %	99.32
Failed by Low Call Rate %	0.68
Failed by Low H.W. P-Value %	0.00
Failed by Low Repeatability %	0.00
Failed by Low Trio Concordanc...	0.00
# Repeated Sample Discordances	44
# Repeated Sample Concordances	40,419
Cluster Fit Repeatability %	99.89
# Trio Discordances	144
# Trio Concordances	65,253
Cluster Fit Trio Concordance %	99.78
Completeness %	98.75
Low Call Confidence Cut %	0.83

Figure 5.12 Viewing Fit Results

Troubleshooting Cluster Genotyping Results

We recommend that you troubleshoot cluster genotyping results by following the workflow below in the order presented.

Cluster Genotyping Troubleshooting Workflow

The workflow is as follows:

- ***Were enough experiments run on unique samples?*** (See *Were enough experiments run on unique samples? on page 92*)
- ***Are there any unusual experiments that are skewing the Fit Results?*** (See *Are there any unusual experiments that are skewing the Fit Results? on page 93*)
- ***Adjust the genotype settings to improve certain Fit Result values.*** (See *Adjust the genotype settings to improve certain Fit Result values. on page 93*)

Were enough experiments run on unique samples?

To determine if enough experiments were run on unique samples:

1. Open the **Fit Results** for your cluster (see ***Viewing Fit Results*** on page 91).
2. Look at the value for **# Genotyped Unique Samples** to determine if enough experiments of unique samples were successfully run and genotyped:

> **IMPORTANT: Cluster genotyping results are best if at least 80 unique samples are successfully genotyped (Auto status is Pass).**

Example: The Fit Results table in Figure 5.13 shows that only 19 out of 20 experiments were genotyped. Of the 19, only 18 were unique samples.

As a result of the low number of experiments clustered, the Passed Assays % and Cluster Fit Repeatability are not as high as they could be. In this example, the Hardy-Weinberg cut (H.W.P-Value) was also applied. The cut failed almost 1% of the assays, some of them inappropriately. With a larger sample set, this cut is more appropriate.

Solution: Continue running more experiments for this project. When data from at least 80 unique samples has been collected and processed, generate a new cluster and view the Fit Results.

Fit Results | Genotype Settings | Assays | Experime

Property	Value
# Experiments	20
# Genotyped Experiments	19
# Unique Samples	19
# Genotyped Unique Samples	18
# Trios	0
# Assays	9704
# Passed Assays	9449
Passed Assays %	97.37
Failed by Low Call Rate %	1.81
Failed by Low H.W. P-Value %	0.81
Failed by Low Repeatability %	0.00
Failed by Low Trio Concordance %	0.00
# Repeated Sample Discordances	27
# Repeated Sample Concordances	18780
Cluster Fit Repeatability %	99.86
Completeness %	99.27
Low Call Confidence Cut %	0.46

Number of unique samples cluster genotyped is less than 80.

Figure 5.13 Example of Fit Results

Are there any unusual experiments that are skewing the Fit Results?

Unusual experiments are experiments that have an unusually strong influence on overall Cluster Fit Repeatability, Completeness, and Trio Concordance.

To determine if there are any unusual experiments that are skewing Fit Results:

1. Select a cluster.
2. Select the **Samples** tab, and sort the sample names by **Repeat Count**.

> **IMPORTANT:** **Cluster Fit Repeatability is most meaningful if at least 5 pairs of unique samples are repeated.**

3. Select the **Fit Results** tab.
4. Look at the value for **Cluster Fit Repeatability**.
   If the value is less than 99%, you may have:
   - Experiments with sample tracking errors
   - Experiments with poor data quality

   See *Identifying Experiments with Low Repeatability* on page 93 for more information.
5. Look at the value for **Completeness %**.
   If the value is below 96%, look for experiments with low call rates. See *Identifying Causes of Low Completeness* on page 94 for more information.
6. If trio testing is part of the cluster, look at the value for **Cluster Fit Trio Concordance.**
   If the value is less than desired, you can see if a particular trio of samples is skewing this metric. To do this:
   A. Select the **Experiments** tab.
   B. In the **Exp. Trio Conc. %** column, look for trios with a value lower than 99.5%.
      Ignore trios with a 0% value. These trios were not tested.

   If found, it is possible that at least one of the experiments for a trio member has been mistracked, or is really not part of the trio.
7. Optional: If you find unusual experiments that are degrading your overall metrics, you can manually fail these experiments and generate a new cluster.

Adjust the genotype settings to improve certain Fit Result values.

Before adjusting genotype settings and generating new clusters, be sure that you have:

- Successfully run and genotyped at least 80 unique samples.
- Determined there are no unusual experiments skewing the fit results.

You can adjust the genotype settings customized for your assay panel and generate a new cluster. See *Adjusting Genotype Settings* on page 96 for more information and instructions.

Identifying Experiments with Low Repeatability

To identify experiments with low repeatability:

1. Click the **Experiments** tab.
2. Sort on the **Exp. Repeatability %** column.
3. Identify experiments with the lowest repeatability.
4. Select one experiment by highlighting the experiment row, then sort by Sample Name.
   This will put repeated experiments of the same samples in adjacent rows.
5. Note the experiment name.
6. Click the **Array Data** icon and look at the **QC Call Rates** for that experiment.

Low repeatability can be caused by:

- Sample tracking errors
- Sample contamination
- Error made while performing the assay

Table 5.2 Troubleshooting Experiments with Low Repeatability

If …	Then …
Experiment repeatability is less than 90% and the QC Call Rate and QC Half Rates are good	A sample tracking error was most likely made. See *Sample Tracking Errors* on page 94.
Experiment repeatability is less than 99% but greater than 90%, the QC Call Rate is low, and the QC Half Rate is high	• an error may have been made while performing the assay. Repeat the experiment. • sample quality may be poor. • the sample may be contaminated. Refer to the *Affymetrix GeneChip® Scanner 3000 Targeted Genotyping System User Guide* for information on sample criteria.

If repeatability is low for only a few experiments, you can fail these as well before generating a new cluster. Doing this will ensure that your overall Fit Results are representative of the data set as a whole, and are not skewed by a few unusual experiments.

Sample Tracking Errors

In Figure 5.14, the highlighted experiments have a repeatability of less than 90%. Repeatability this low indicates that a different sample than the one recorded was hybridized on at least one of these arrays.

Experiment	Sample Location	Sample Name	Fit Status	Repeat Count	Best Repeat?	Exp. Repeatability %
(a)4000484-02674	SMP_124:D06	AMann	Auto Pass	2	Y	99.59
(a)4000484-02811	SMP_124:D06	AMann	Auto Pass	2		99.59
(a)4000484-02322	SMP_124:C09	BJones	Auto Pass	2	Y	76.25
(a)4000484-02784tr...	SMP_124:C09	BJones	Auto Pass	2		76.25

Figure 5.14 Example of Low Repeatability – Sample Tracking Error

Recommendation: If you suspect tracking errors, fail the suspect experiments in the Experiment QC Summary table and generate a new cluster. Try to determine how the sample tracking error occurred because other experiments processed alongside these may also be mistracked.

Identifying Causes of Low Completeness

To identify the causes of low completeness (average call rate):

1. Click the **Experiments** tab.
2. Sort on the **Exp. Call Rate %** column.
3. Locate the experiments with the lowest call rates (Figure 5.15).
4. Consider changing the experiment status to Fail if:
   - The call rate is very low and you have much better data for the same sample in another experiment.

- You manually passed an experiment that was failed by GTGS. These experiments will generally have the lowest call rates.

Experiment	Sample Location	Sa...	Fit Status	Repeat Count	Best Repeat?	Exp. Repeatability %	Exp. Call Rate %
(a)4000484-02775	SMP_124:A03	GBr...	Auto Pass	1	Y	0.00	97.72
(a)4000484-02498	SMP_0000001:...	PH...	Auto Pass	5		99.92	96.35
(a)4000484-02756	SMP_124:A02	FM...	Auto Pass	1	Y	0.00	95.84
(a)4000484-0233...	SMP_124:D07	RM...	Auto Pass	3		99.87	95.64
(a)4000484-02178	SMP_124:A10	BC...	Auto Pass	1	Y	0.00	95.18
(a)4000484-02370	SMP_124:C08	GJain	Auto Pass	1	Y	0.00	95.08
(a)4000484-02568	SMP_124:A09	ZC...	Auto Pass	1	Y	0.00	94.01
(a)4000484-0232...	SMP_124:B03	HD...	Auto Pass	1	Y	0.00	90.72
(a)4000484-0233...	SMP_124:D07	RM...	Auto Pass	3		99.89	89.65
(a)4000484-02686	SMP_124:B11	PH...	Auto Pass	1	Y	0.00	89.35
(a)4000484-0233...	SMP_124:D07	RM...	Fail	3		0.00	0.00

Figure 5.15 Example of Low Experiment Call Rate

Refining Cluster Genotyping Results

About Refining Clusters

Refining clusters is done by adjusting the genotype settings supplied with your assay panel and generating a new cluster.

> **IMPORTANT:** **Be aware that adjusting a genotype setting will impact several metrics. For example, increasing the stringency on MinCallConfidence will lower the assay call rate. This, in turn, will cause an assay to fail if it falls below the MinAssayCallRate threshold.**

Adjusting Genotype Settings

There are two types of genotype settings: default genotype settings for the assay panel and application genotype settings.

- **Default Genotype Settings**

  A customized set of default genotype settings is included with every assay panel. These settings are included on the CD-ROM supplied with the first GeneChip® SNP Kits you receive for a particular assay panel. Default settings are determined by ParAllele to yield good cluster results. These settings can be viewed by selecting an assay panel and viewing the Default Genotype Settings tab. They are preceded by the word CUSTOM .

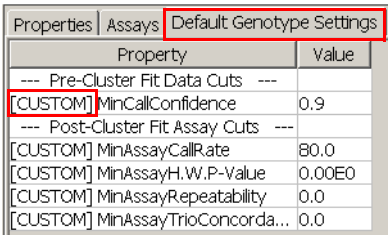


Figure 5.16 Default Genotype Settings tab

> **IMPORTANT:** **Always generate at least one cluster using the default genotype settings. You have the option of changing one or more of these settings to see how it affects your data.**

- **Application Genotype Settings**

  These settings are used if you do not import the settings from the CD-ROM supplied with the first GeneChip SNP Kits you received for a particular assay panel.

> **NOTE:** **We recommend always using the default genotype settings included with your assay panel.**

Guidelines for Adjusting Genotype Settings

The following genotype settings can be adjusted to refine your cluster genotyping results:

- MinCallConfidence
- MinAssayCallRate
- MinAssayH.W.P-Value
- MinAssayRepeatability
- MinAssay Trio Concordance

The impact of adjusting genotype settings on the following metrics is described in Table 5.2:

- Passed assays % and #

- Completeness %
- Cluster Fit Repeatability %
- Trio Concordance %

Table 5.3 Impact of Adjusting Genotype Settings

Genotype Settings	Fit Result Metric			
	Passed Assays %	Completeness %	Cluster Fit Repeatability %	Trio Concordance
MinCallConfidence – Increased	–	–	+ + +	+ + +
MinCallConfidence – Decreased	+	+	– – –	– – –
MinAssayCallRate – Increased	–	+ + +	+	+
MinAssayCallRate – Decreased	+	– – –	–	–
MinAssayH.W.P-Value – Increased	–	+ / –	+	+
MinAssayH.W.P-Value – Decreased	+	+ / –	–	–
MinAssayRepeatability – Increased	–	+	α	+
MinAssayRepeatability – Decreased	+	–	– –	– –
MinAssay Trio Concordance - Increased	–	+	+	β
MinAssay Trio Concordance - Decreased	+	–	– –	– –

– = slight decrease; – – = moderate decrease; – – – = major decrease
+ = slight increase; + + = moderate increase; + + + = major increase
α = metric not measuring true repeatability if MinAssayRepeatability > 0%
β = metric not measuring true concordance if MinAssayTrioConcordance > 0%

IMPORTANT: Population annotations in the Sample Info table are not used when calculating the H.W.P-Value. The MinAssayH.W.P-value cut is not applied to assays for X and Y chromosome SNPs.

For instructions on adjusting genotype settings, see *Cluster Genotyping* on page 37.

Manually Failing Experiments with Poorest Quality Data

To manually fail experiments with the poorest quality data:

1. Expand your project and select the **Array Data** icon.
2. Click the **Experiment QC Summary** tab.
3. Identify the experiments you want to fail.
   See *Changing the Status of an Experiment* on page 97.
4. Generate a new cluster.
   See *Cluster Genotyping* on page 37.

Changing the Status of an Experiment

To change the status of multiple experiments at once:

1. Expand your project folder.
2. Click the **Array Data** icon.

3. In the right pane, select one or multiple experiment rows whose status you would like to change. You may want to sort the table first or control click to select discontinuous rows. Right-click on the selected rows and select **Set Manual Status** (Figure 5.17).

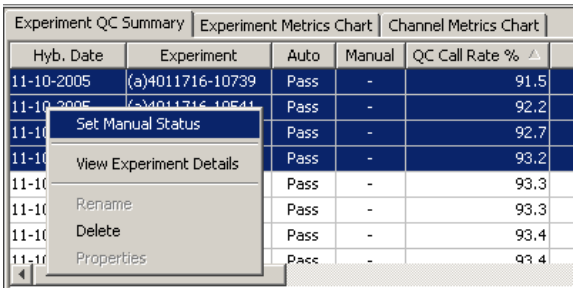


Figure 5.17 Select Manual Status

4. The Set Manual Status window appears under the Summary tab (Figure 5.18)..

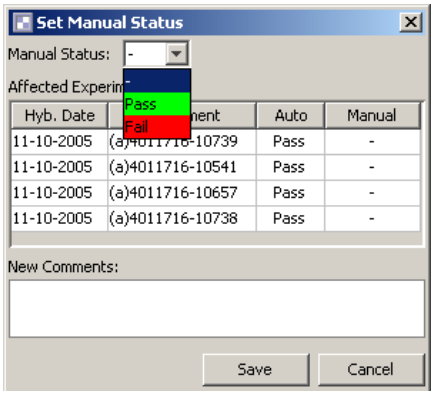


Figure 5.18 Set Manual Status: Pase/Fail

   A. Select a Manual Status of Pass, Fail, or - (reset to Auto-Status state).
   B. Optional: Enter comments in the New Comments field.
   C. Click **Save**.
   D. Close the window by clicking the box in the upper right corner of the window or click **Cancel**.

To change the status of a single experiment:

1. Expand your project folder.
2. Select the **Array Data** icon.
3. In the right pane, right-click anywhere in the row of the experiment you want to change the status of and select **Display Experiment Details**.

   The Experiment Details window appears.
4. Under the Summary tab:

   A. Open the **Manual Status** menu.
   B. Select a status.
   C. Optional: Enter comments in the User Comments field.
   D. Click **Apply**.
   E. Close the window by clicking the box in the upper right corner of the window or click **Cancel**.

The Manual column of the Experiment QC Summary table will reflect your change.

CEL Data Not Found – Updating the Location of CEL Files

When data is imported, the .cel file location for each experiment is recorded by GTGS. If the location of the .cel files for a particular project is changed, you will have to update the path GTGS in order to view information such as Experiment Details. If the location is not updated, an error message stating that the .cel data cannot be found is displayed (Figure 5.19).

To update the location of .cel files for a project:

1. Open the project in GTGS.
2. Right-click the Array Data icon
3. Right-click on the row an experiment in the Experiment QC Summary table and select **View Experiment Details**.

   If GTGS cannot find the .cel file for this experiment, the error message shown in (Figure 5.19) is displayed (Summary Tab).

1. Click the **Change Location** button.
2. Browse to the directory containing the .cel files.

   Note that since the browser filters the display to only show directories, you will not see the CEL files.
3. Click **Open**; then click **Apply**.
4. Click **Next**; then click **Prev** to load the .cel file data.

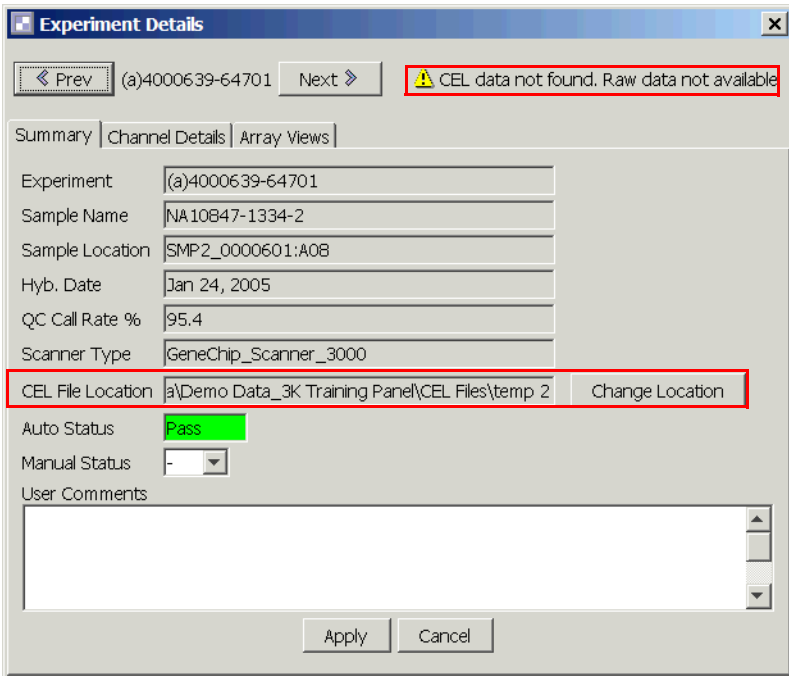


Figure 5.19 Updating CEL File Location

Chapter 6

AGCC AND GCOS COMPATIBILITY MODES

IMPORTANT: **To avoid problems, do not switch GTGS between GCOS and AGCC compatibility modes in the middle of a project. The compatibility mode should not be frequently switched.**

About GCOS and AGCC compatibility modes

As of version 1.6, GTGS is compatible with Affymetrix GeneChip® Command Console (AGCC) instrument control software:

- GTGS batch creates sample files (.arr), which are required before arrays can be washed or scanned
- GTGS reads AGCC intensity files (.cel)

GTGS is backwards compatible with Affymetrix GeneChip Operating Software (GCOS), so you don't have to immediately transition to AGCC.

However, due to differences in how GTGS communicates with GCOS and AGCC, and due to differences in how AGCC and GCOS name the data files, GTGS is not simultaneously fully compatible with both systems. Therefore, GTGS must be explicitly set for either GCOS mode or AGCC mode.

Compatibility differences between GCOS and AGCC

The AGCC naming convention for .dat and .cel files includes an underscore before the channel code, i.e. (a)1234567-12345_A.cel. This differs from the GCOS convention, which does not include an underscore. When the software is configured to NOT communicate with GCOS, both the Tracking operation **Hyb on Arrays** and the Array Data operation **Import Experiment List → from Project** will pass an experiment list to the Array Data section that assumes that .cel file names include an underscore before the channel code. Additionally, when using the **Load Rescans** feature as part of the **Import Experiment Data** operation, the expected rescan file names depend on whether GTGS is currently in AGCC mode or GCOS mode.

The naming convention difference can lead to issues when reading GCOS .cel files after GTGS has been configured to be compatible with AGCC, or when reading AGCC .cel files after GTGS has been configured to be compatible with GCOS. For more information, see *Recovering from a failed import of .cel data on page 103*.

Determining the compatibility mode

In the Projects tree, expand a project and right-click on the Tracking node (Figure 6.1).

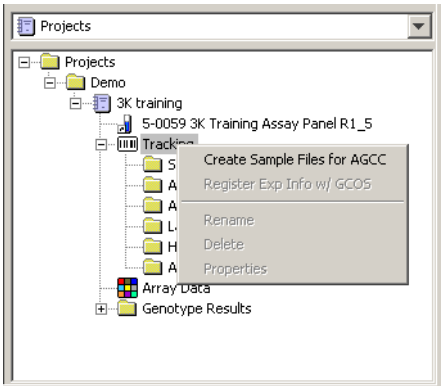


Figure 6.1 The Tracking Node

If you have eabled **Create Sample Files for AGCC**, then you have configured GTGS to be compatible with AGCC. If instead you have enabled **Register Exp Info w/ GCOS**, then you have configured GTGS to be compatible with GCOS.

Changing the compatibility mode

The GTGS setting that defines the GCOS/AGCC compatibility mode is accessible only by directly editing the configuration file **GCOS.properties** on the server computer (generally the Post-Amp computer). Only a user with the "write" privilege for this file can change this setting. This file is in the folder "C:\Program Files\Apache Software Foundation\Tomcat 6.0\webapps\geno\config".

To view or change the GCOS/AGCC compatibility mode,

1. Open GCOS.properties with Microsoft Notepad (Figure 6.2)..

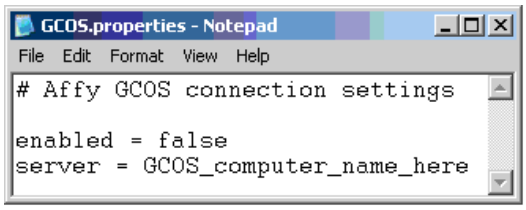


Figure 6.2 GCOS.properties

2. Edit the "server" and "enabled" fields as needed
   - **A.** GCOS mode when "enabled = true". This mode should be selected when tracking and importing GCOS data. GTGS will communicate with the GCOS computer whose name is in the "server" line. GTGS will register experiments with GCOS at the end of each Hyb plate's **Hyb on Arrays** operation. The **Register Expts w/ GCOS** operation will be visible when selecting the Tracking node, for users who wish to (re)register all experiments with GCOS. GTGS will assume that .cel files do not have an underscore between the experiment name and the channel identifier.
   - **B.** AGCC mode when "enabled = false". This mode should be selected when tracking and importing AGCC data. At the end of each **Hyb on Arrays** operation, GTGS will ask the user to select a location to save the .arr ("sample") files used by AGCC. The user should put these .arr files in an AGCC data root subfolder in order for AGCC to find the files when washing and scanning the associated arrays. The **Create Sample Files for AGCC** operation will be visible when selecting the Tracking node, for users who wish to (re)register all experiments with AGCC.

3. Save the file, and restart GTGS.

Recovering from a failed import of .cel data

Resolving a DAT and CEL Error

While **Import Experiment Data** is processing files, you may get the message *ERROR DAT and CEL files do not exist: <path to CEL file>*

To resolve this error:

1. Verify that you have selected the correct folder containing your .cel files
2. Verify that these files have been created in the expected folder
3. If you process GCOS .cel files while in AGCC mode, or process AGCC .cel files while in GCOS mode, there may be a mismatch between the expected and actual file names. For example, GTGS may expect there to be an underscore "_" before the channel code when there isn't one, or vice versa. The resolution is to correct the experiment list information maintained in the Array Data node of the affected project. Choose one of the following options that better suits your needs:

Reprocess data from an existing project in a new project

NOTE: This option assumes that the Experiment QC Summary table in your original project shows QC results, which indicates the .cel data was properly loaded.

1. From the original project, right-click on **Array Data** and select **Export Experiment List.** Save this file.
2. In the destination project, right-click on **Array Data** and select **Import Experiment List → From File**. Select the file you just exported.
3. In the destination project, right-click on **Array Data** and select **Import Experiment Data**. Select the current location of the reference .cel files, and import the data.

Manually edit the Experiment List:

1. From the project that reports this error, right-click on **Array Data** and select **Export Experiment Lis**t. Save this file and open in an editor like Microsoft Excel or Notepad.
2. The columns "Data File Ch A, Data File Ch C, Data File Ch G, and Data File Ch T" contain the expected .cel names for the A.cel, B.cel, C.cel, and D.cel files, respectively. Manually edit these names to match the actual file names for the associated experiments, and save this file.
3. As a test, create a new project, right-click on **Array Data**, and select **Import Experiment List → From File**. Load this experiment list. Then right-click on **Array Data** and select **Import Experiment Data**. If you can successfully load CEL data and see QC results for these experiments in the Experiment QC Summary table, you know you have the correct information in the Experiment List. If there are errors, edit the list as necessary.
4. Once you've corrected the .cel names in the experiment list file, it is safe to delete unprocessed experiments with incorrect .cel names in the original project, and replace with the correct information. If NO experiments have been processed, or if you don't mind deleting and reprocessing all experiments in the original project, then right-click on **Array Data** and select **Delete All Experiments**. If you only wish to delete unprocessed experiments in your project, then select relevant experiment rows in the Experiment QC Summary table, right-click, and select **Delete Experiment**. You will be asked to confirm deletion for each experiment.
5. Finally, from the original project, right click on **Array Data** and select **Import Experiment List → From File**. Load the corrected experiment list. Only experiments not already listed in this project's Experiment QC Summary table will be loaded. Then right-click on **Array Data** and select **Import Experiment Data**.

Chapter 7

DATA STORAGE AND SERVER MAINTENANCE

Where DAT and CEL Files are Stored

When generated, the .jpg, .dat and .cel files are stored on the Instrument Control Workstation.

NOTE: If you have a GeneChip® Operating Software (GCOS) server, the .jpg, .dat, .cel files can optionally be stored on the remote GCOS Server (depending on the workflow configuration). For details, please refer to the *Affymetrix GeneChip® Operating Software With AutoLoader User's Guide*, Version 1.4 (701439-5) and the *Affymetrix® GCOS Server Addendum, Versions 1.2 and 1.3* (701748). If you have Affymetrix GeneChip® Command Console (AGCC), you can easily move the data files to another location.

Data Backup and Storage Recommendations

Periodic Backups

We recommend that you backup your project data to a different, secure location of your choice every one to two weeks.

Data Storage Recommendations

At the end of a project, we recommend consolidating the data and information listed below, and storing it in a secure location. Doing so will ensure the integrity of the entire analysis, and will allow you to completely recover your data if necessary.

- .jpg, .dat and .cel files (on the Instrument Control Workstation).

  NOTE: For GCOS users, use the Affymetrix Data Transfer Tool to archive these files. Refer to the GCOS User's Guide for more information and instructions on using the Data Transfer Tool. For AGCC users, you can easily move the data files to another location using Windows Explorer.

- Sample plate files (wherever they were created).
- Assay panel file (on the CD ROM included with your assay panel).
- Finished experiment list from Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS; Figure 7.1).
- For each cluster generated for a project, the table displayed under the Experiments tab (Figure 7.2).

This table includes the exact list of experiments used for clustering, and the pass/fail status of each experiment.

NOTE: This information is different from the Experiment QC Summary table that you see when you click the Array Data icon.

- Exported genotype files, including the genotype settings.

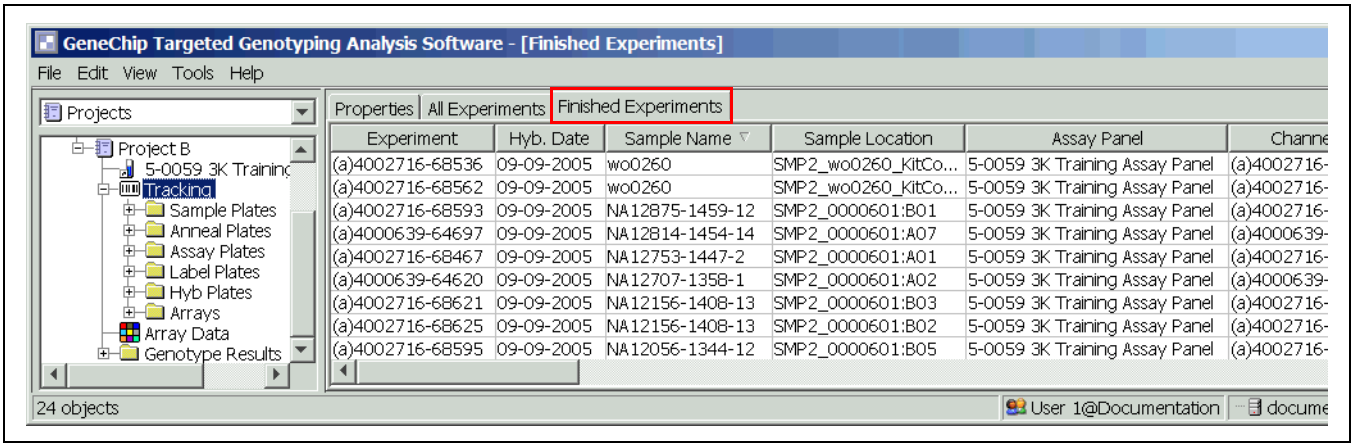


Figure 7.1 Finished Experiments Table

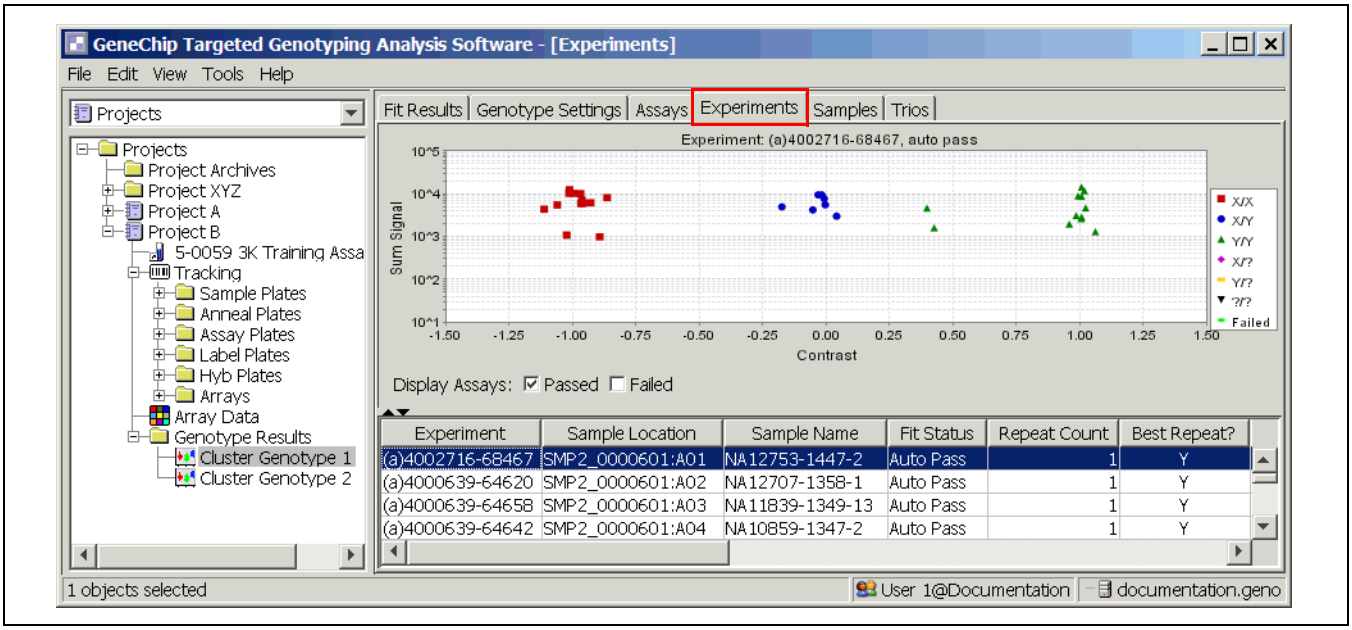


Figure 7.2 Experiments Table

Server Maintenance

The following batch jobs and computer maintenance recommendations are to be performed on the TG Post-Amp Workstation.

Set up Database Job Failure Notification for SQL 2000

You can set up popup notification on the database server to notify you when a database job fails. Instructions are listed below. If you do not wish to setup database popup notification, we recommend that you periodically check the error logs in the event viewer (see *Check the Logs* on page 116).

WARNING: To set up database popup notification, the messenger service must be enabled. Messenger can be a potential security issue. Spammers have found a way to use it to deliver messages to the desktop. To prevent unwanted messages from popping up on the database server, do not use Internet Explorer to surf the net on the computer that is running the database.

Activate Messenger

To activate messenger service to enable Net Send:

1. Open **Start** → **Settings** → **Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.
3. In the left pane, expand **Services and Applications**.
4. Select **Services** and look for **Messenger** in the right pane.
5. Right-click the Messenger service and select Properties (Figure 7.3).

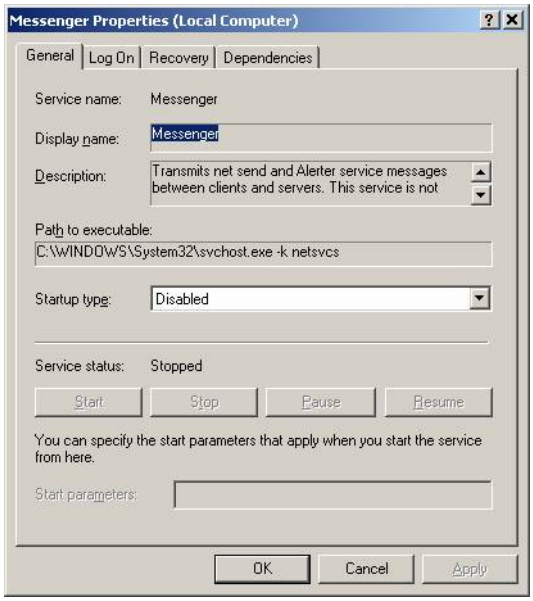


Figure 7.3 Activate Messenger

6. Select **Automatic** from the Startup type list, and then click **Apply**.
7. Click the **Start** button.
8. Click **OK**.

Create an Operator in Microsoft® SQL Server 2000

To create an operator in SQL server:

1. Open **Start** → **Settings** → **Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.
3. In the left pane, expand **Services and Applications**.
4. Expand **Microsoft SQL Servers**; then expand **(local) (Windows NT)**.
5. Expand **Management**; then expand **SQL Server Agent**.
6. Right-click **Operators** and select **New Operator** (Figure 7.4).

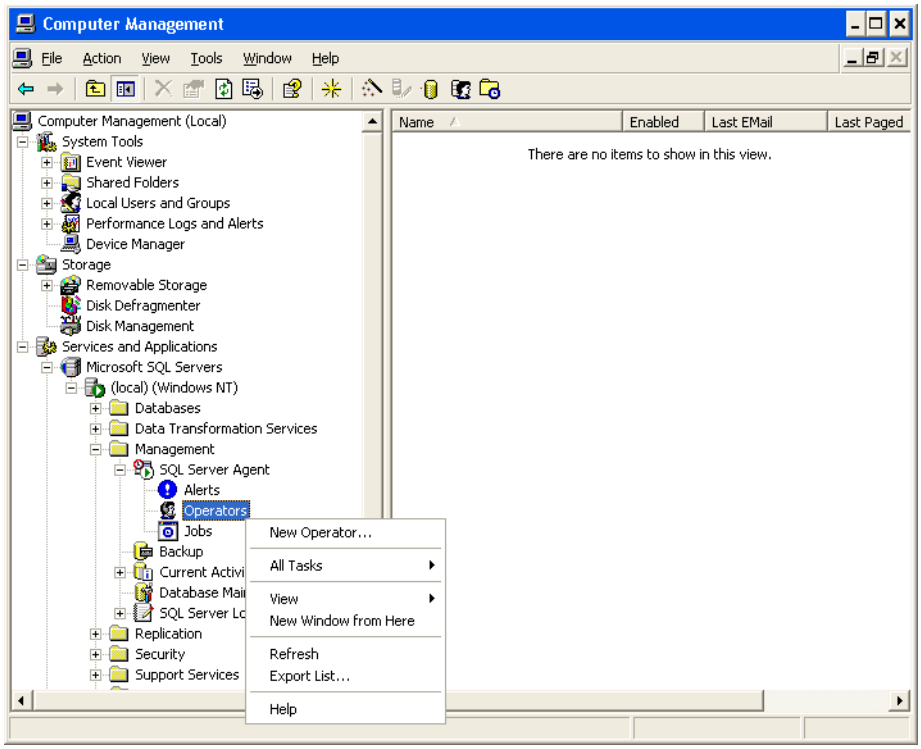


Figure 7.4 Creating an Operator in SQL Server

7. In the Name box, type the name of the operator: **AFFXUser**.
8. In the Net send address box, enter the name of the computer (Figure 7.5).

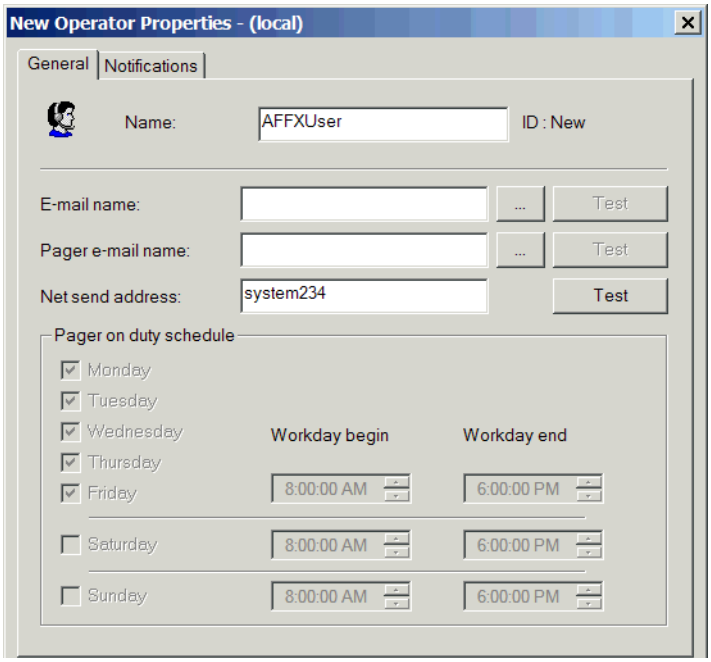


Figure 7.5 New Operator Properties Window

9. Click the **Test** button to make sure a Messenger Service window pops up (Figure 7.6).

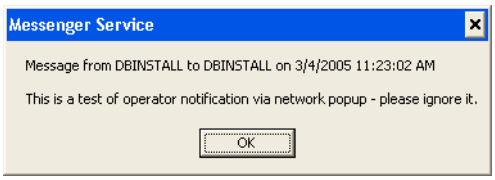


Figure 7.6 Messenger Service Window

Enable Message Popup for Database Jobs in SQL 2000

1. Open **Start** → **Settings** → **Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.
3. In the left pane, expand **Services and Applications**.
4. Expand **Microsoft SQL Servers**; then expand **(local) (Windows NT)**.
5. Expand **Management**; then expand **SQL Server Agent**.
6. Select **Jobs**.
7. Right-click the first job in the list and select **Properties**.
8. Select the **Notifications** tab (Figure 7.7).
9. Check **Net send operator:** and select **AFFXUser** from the drop down box next to it. Be sure that **When the job fails** is selected.

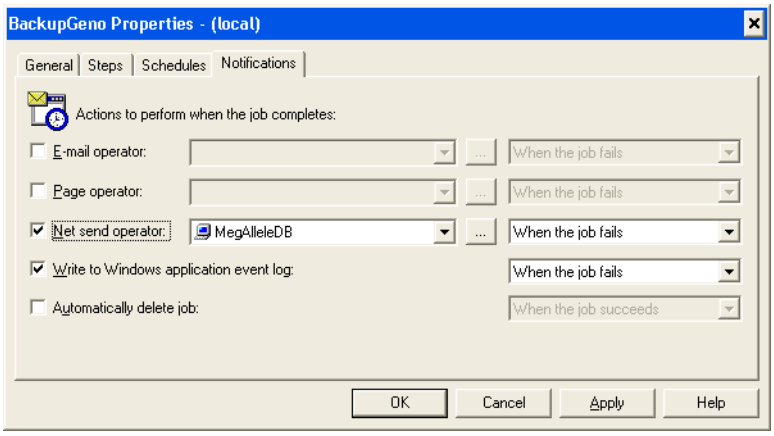


Figure 7.7 BackupGeno Properties Window

10. Click **OK**.

11. Repeat steps 6 through 9 for each job in the list of jobs.

Set up Database Job Failure Notification for SQL 2005

You can set up popup notification on the database server to notify you when a database job fails. Instructions are listed below. If you do not wish to setup database popup notification, we recommend that you periodically check the error logs in the event viewer (see *Check the Logs* *on page 116*).

Activate Messenger

To activate messenger service to enable Net Send:

1. Click **Start → Settings → Control Panel**.
2. Click **Administrative Tools**, then click **Computer Management.**
3. In the left pane, expand **Services and Applications.**
4. Select **Services** and look for Messenger in the right pane.
5. Right-click the **Messenger** service and select **Properties** (Figure 7.8).

NOTE: To set up database popup notification, the messenger service must be enabled. Messenger can be a potential security issue. Spammers have found a way to use it to deliver messages to the desktop. To prevent unwanted messages from popping up on the database server, do not use Internet Explorer to surf the net on thecomputer that is running the database.

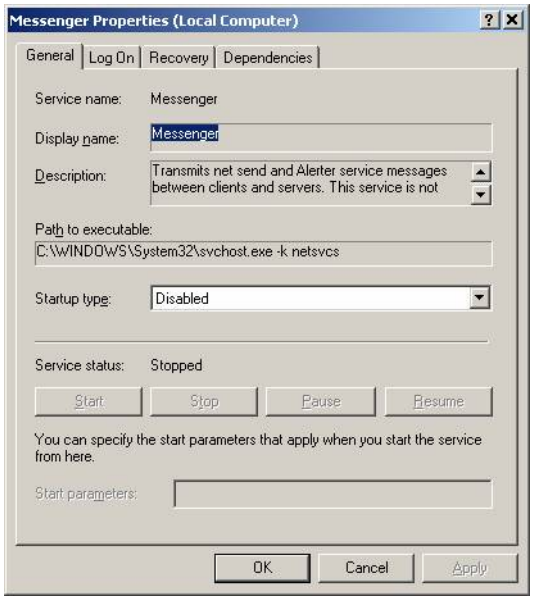


Figure 7.8 Activate Messenger

6. Select **Automatic** from the Startup type list, and then click **Apply**.
7. Click the **Start** button.
8. Click **OK**.

Create an Operator in Microsoft® SQL Server 2005

1. Click **Start → Programs → Microsoft SQL Server 2005 → SQL Server Management Studio** (Figure 7.9).
2. In the left pane, expand the local machine name.
3. Expand **SQL Server Agent**.
4. Right-click **Operators** and select **New Operator**.

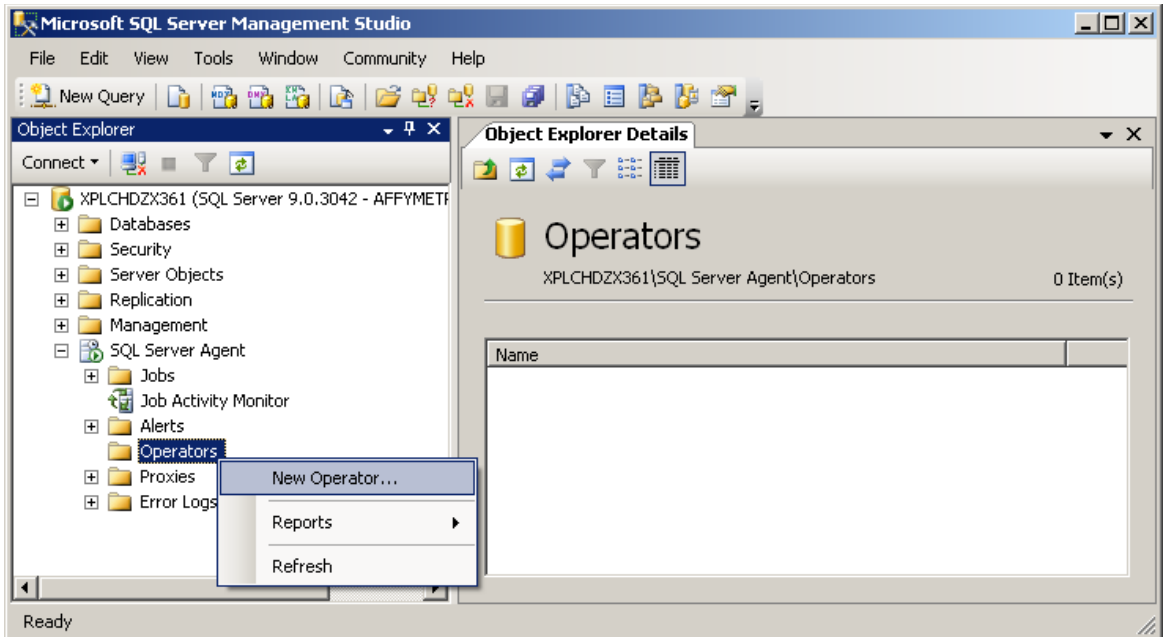


Figure 7.9 SQL Server Management Studio

5. In the Name box, type the name of the operator: AFFXUser.
6. In the Net send address box, enter the name of the computer (Figure 7.10).

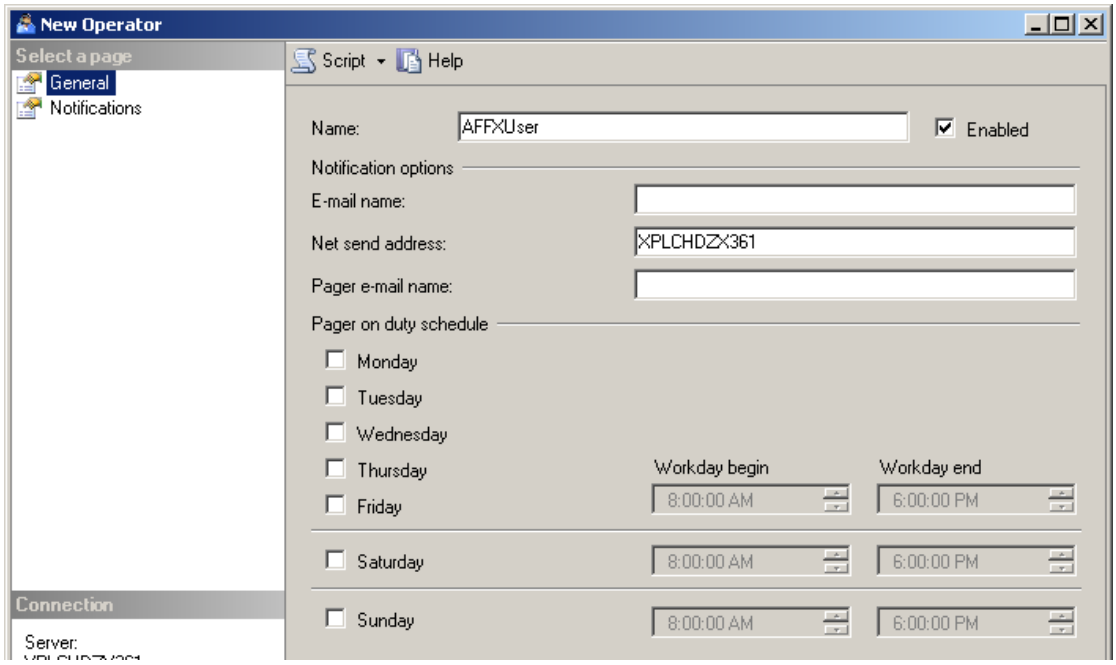


Figure 7.10 Creating an Operator in SQL Server

7. Click OK.

Enable Message Popup for Database Jobs in SQL Server 2005

1. Click **Start → Programs → Microsoft SQL Server 2005 → SQL Server Management Studio**
2. In the left pane, expand the local machine name.

3. Expand **SQL Server Agent**.
4. Expand **Jobs**.
5. Right-click the first job in the list and select **Properties**.
6. In the left pane, select the **Notifications** page.
7. Check **Net** send operator: and select **AFFXUser** from the drop down box next to it **(**Figure 7.11).

NOTE: Be sure that you select When the job fails.

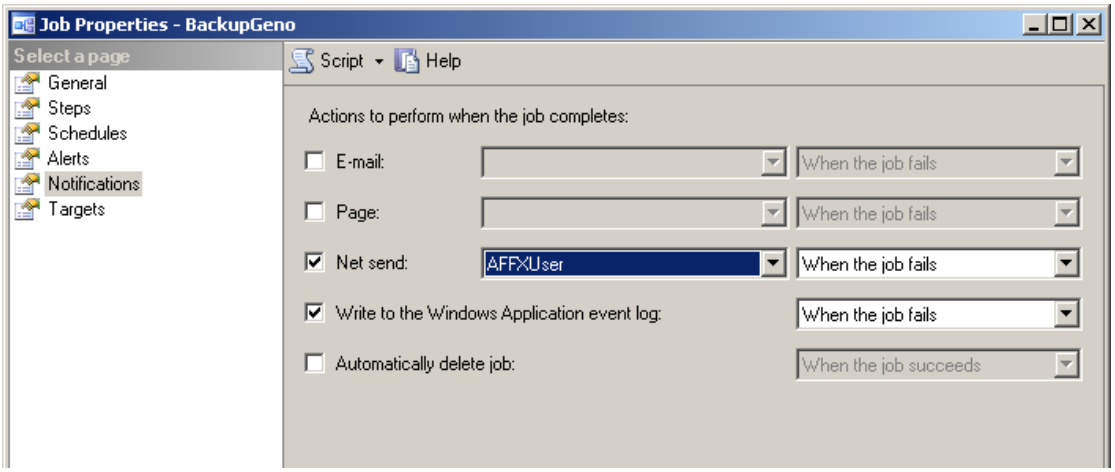


Figure 7.11 BackupGeno Properties Window

8. Click **OK**.
9. Repeat steps 5 through 8 for each job in the list of jobs.

SQL Server Batch Jobs

Scheduled Jobs

- **BackupGeno**

  Backs up the Geno database onto [drive letter]:\mssql\backup\Genodb.bkup. This job runs daily at 1:00 AM.

- **Trunc_logs**

  Truncates the Geno and Tempdb database transaction logs. This job runs every Sunday at 2:00 AM.

- **Check HostName**

  If the server name changes, this job will make the necessary changes within SQL Server to recognize the new server name. This job runs each time SQL Server is started.

- **DBLoginAudit (not available for systems using SQL Server 2005)**

  Tracks user login and logout of the application and the database. This job runs every 30 minutes on server startup. It writes the audit information into the database table DBLoginAudit.

- **Clean_DBLoginAudit (not available for systems using SQL Server 2005)**

  Deletes data that is older than 60 days from the DBLoginAudit table. This job runs daily at 3 AM. See *Login/Logout Audit Information* on page 115 for more information.

- **DefragDB**

  Defragments the Geno database. This job runs once per week on Friday at 11:30 PM.

- **UpdateStats**

  Update the Geno database statistics on a nightly basis. The update will help the database return query results faster. This job runs nightly at 1 AM.

Terminating Processes Prior to a Database Restore

To terminate all processes prior to a database restore:

1. Go to each of the computers that may be running GeneChip® Scanner 3000 Targeted Genotyping System applications, and quit these applications.

   If a user is in the middle of an operation, first ask the user whether or not the application can be quit.
2. Terminate any remaining processes that may be connected to the Geno database as follows from the Windows Desktop.

For SQL Server 2000 users

A. Open **Start** → **Settings** → **Control Panel**.

B. Open **Administrative Tools**; then open **Computer Management**.

C. In the left pane, expand **Services and Applications** → **Microsoft SQL Servers** → **(local)(Windows NT)** → **Management** → **Current Activity**.

D. Select **Process Info**.

E. In the right-hand pane, right-click all of the processes that are connected to the Geno database and select **Kill Process**.

For SQL Server 2005 users

A. Open **Start → Programs → Microsoft SQL Server 2005** → **SQL Server Management Studio.**

B. In the left pane, expand the computer name hosting the database.

C. Expand **Management**

D. Right-click on **Activity Monitor** and select **View Processes**.

E. In the window that opens, right-click all of the processes that are connected to the Geno database and select **Kill Process**.

Manual Restore Jobs

NOTE: The following Restore jobs should be used only to recover from database corruption or unintended deletion of information. They do not run on a schedule. Before the jobs can be run, no processes can be connected to the Geno database. See *Terminating Processes Prior to a Database Restore* on page 114.

Restore OLD Db

Archives the current Geno database to:

[drive letter]:\mssql\backup\Genodbcurrent.bkup

and restores the old Geno database backup from:

[drive letter]:\mssql\backup\Genodb.bkup

This job should only be run when a restore of the Geno database is necessary.

To run Restore OLD Db using SQL 2000:

1. From the Windows Desktop, open **Start** → **Settings** → **Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.

3. In the left pane, expand **Services and Applications → Microsoft SQL Servers → (local)(Windows NT) → Management → SQL Server Agent**.
4. Select **Jobs**.
5. In the right-hand pane, right-click **Restore OLD Db** and select **Start Job**.
6. Accept the default Start execution at step 1 and click **Start**.
7. To check on the status of the job, right-click in the job list window and select **Refresh**.

To run Restore OLD Db using SQL Server 2005:

1. From the Windows Desktop, open **Start → Programs → Microsoft SQL Server 2005 → SQL Server Management Studio**.
2. In the left pane, expand the computer name hosting the database.
3. Expand **SQL Server Agent**.
4. Select **Jobs**.
5. In the right-hand pane, right-click **Restore OLD Db** and select **Start Job at Step....**
6. To check the status of the job, right-click in the job list window and select **Refresh**.

Restore Current Db

Run this job only if you need to undo the Restore OLD Db job. The job restores the current Geno database backup from:

[drive letter]:\mssql\backup\Genodbcurrent.bkup

This job sets the database back to the state prior to running Restore OLD Db. This job will not function if Restore OLD Db has never been run.

To run Restore Current Db using SQL Server 2000:

1. From the Windows Desktop, open **Start → Settings → Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.
3. In the left pane, expand **Services and Applications → Microsoft SQL Servers → (local)(Windows NT) → Management → SQL Server Agent**.
4. Select **Jobs**.
5. In the right-hand pane, right-click **Restore Current Db** and select **Start Job**.
6. To check on the status of the job, right-click in the job list window and select **Refresh**.

To run Restore Current Db using SQL Server 2005:

1. From the Windows Desktop, open **Start → Programs → Microsoft SQL Server 2005 → SQL Server Management Studio**.
2. In the left pane, expand the computer name hosting the database.
3. Expand **SQL Server Agent**.
4. Select **Jobs**.
5. In the right-hand pane, right-click **Restore Current Db** and select **Start Job at Step....**
6. To check the status of the job, right-click in the job list window and select Refresh.

Login/Logout Audit Information

The database audits user login and logout of the application and of the database. This auditing is enabled at all time while the database server is running.

NOTE: **The audit jobs are only provided for SQL Server 2000 installs.**

The audit information is first written to a binary file. Then every 30 minutes, a batch job called DBLoginAudit writes the information from the binary file to the DBLoginAudit table in the Geno database. (See ***Scheduled Jobs*** on page 113.)

Another batch job called Clean_DBLoginAudit runs daily to delete data from the DBLoginAudit table that is older than 60 days. The detailed audit information can be viewed by following the instructions below.

To view detailed audit information:

1. From the Windows Desktop, open **Start** → **Settings** → **Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.
3. In the left pane, expand **Services and Applications** → **Microsoft SQL Servers** → (**local**)(**Windows NT**) → **Databases** → **geno**.
4. Select **Tables**.
5. In the right pane of the window, right-click the table **DBLoginAudit** and select **Open Table** → **Return all rows**.

A description of the fields in the DBLoginAudit table is presented in Table 7.1.

Table 7.1 Field Descriptions for the DBLoginAudit Table

Event Class	The event action. Value is usually *Login* or *Logout*.
NTUserName	NT user name of the event. The value will be <*NULL*> if it is an application login or logout event. The user name is specified if it is a database login or logout event.
HostName	The hostname where the event is initiated.
LoginName	The database user name of the event.
DBName	The database name access for the specified event.
Application	The application that is accessing the database.
StartTime	Start time of the event.
Success	*1* if the event was successful; *0* if the event was unsuccessful.
ServerName	The server name of the database.

Computer Maintenance

We recommend performing the following tasks on a regular basis to ensure proper system operation:

Check the Logs

To check the logs:

1. From the Windows Desktop, open **Start** → **Settings** → **Control Panel**.
2. Open **Administrative Tools**; then open **Computer Management**.
3. Expand **System Tools** → **Event Viewer**.
4. Check the Warning and Error logs by clicking **Application**, **Security** and **System**.
5. Determine the level of severity of the problems and fix if necessary.

Copy the Database Backup to Another Location

The Geno database is regularly backed up to the same computer in a file called [drive letter]:\mssql\backup\Genodb.bkup. It is recommended that you copy this file to a separate computer on a nightly basis or at the very least on a weekly basis. The frequency depends on how much data you are willing to lose, if for some reason the hard drives containing the Geno.mdf and Genodb.bkup files crash or the computer becomes unavailable.

Defragment the hard drive(s)

Once every few months, we recommend that you defragment the hard drive. Perform this task when the computer is not in use.

To defragment the hard drives:

1. From the Windows Desktop, navigate to **Start** → **Programs** → **Accessories** → **System Tools** → **Disk Defragmenter**.
2. Click **Analyze**. This step may take an hour to run.
3. Click **Defragment** if the results from the Analyze indicates serious fragmentation.

Chapter 8

GLOSSARY

Many of the terms that appear in this glossary are the column headings in the various tables available in Affymetrix GeneChip® Targeted Genotyping Analysis Software (GTGS). Some of these terms appear in multiple tables.

 # Assays
As displayed under the Fit Results tab – the number of assays performed for each experiment.

 # Experiments
As displayed under the Fit Results tab – the number of experiments available in the Array Data section at the time cluster genotyping was performed.

 # Genotyped Experiments
As displayed under the Fit Results tab – the number of experiments genotyped in the current cluster.

 # Genotyped Unique Samples
As displayed under the Fit Results tab – the number of unique samples among the experiments genotyped in the current cluster.

 # Passed Assays
As displayed under the Fit Results tab – the number of assays in a cluster set that pass quality control criteria.

 # Rejected Outliers
Number of rejected features in a channel, based on scanner signal saturation, blemishes, or high pixel CV.

 # Repeat Disc.
Number of discordant calls in an assay or an experiment (depending on which table the term is in), compared to consensus calls from repeated experiments for the same sample.

 # Repeated Sample Concordances
As displayed under the Fit Results tab – the number of concordant calls in the cluster set, compared to consensus calls from repeated experiments for the same samples.

 # Repeated Sample Discordances
As displayed under the Fit Results tab – the number of disconcordant calls in the cluster set, compared to consensus calls from repeated experiments for the same samples.

 # Trio Concordances

(Fit Results tab)

As displayed under the Fit Results tab – the number of Mendel tests in the entire cluster fit whose genotypes are consistent, based on a comparison of the genotypes of the mother/father/child trios. This is the sum of # Trio Tests minus the sum of # Trio Disc. in the Trios tab.

 # Trio Disc.

(Assays tab)

Number of Mendel errors for a given assay, based on a comparison of the genotypes of the mother/father/trios in the cluster set.

 # Trio Disc.

(Trios table of Genotype Results)

Number of Mendel errors for a given trio, based on a comparison of the genotypes of the mother/father/child trio. ***See also*** # Trio Tests.

 # Trio Disc.

(Experiments tab)

Number of Mendel errors for a given experiment, based on a comparison of the genotypes of the selected experiment with the best experiment (based on QC Call Rate) of the other members of the trio.

 # Trio Discordances

(Fit Results tab)

As displayed under the Fit Results tab – the number of Mendel errors in the entire cluster fit, based on a comparison of the genotypes of the mother/father/child trios. This is the sum of the listed discordances in the Trios tab.

 # Trio Tests

(Trios tab)

Number of tests for Mendelian inheritance analysis. Only the experiment with the highest QC Call Rate for each trio member is used.

 # Trio Tests

(Assays tab)

Number of tests for Mendelian inheritance analysis for the assay. Only the experiment with the highest QC Call Rate for each trio member is used.

 # Trio Tests

(Experiments tab)

Number of tests for Mendelian inheritance analysis for the experiment, based on a comparison of the genotypes of the selected experiment with the best experiment (based on QC Call Rate) of the other members of the trio.

 # Trios

As displayed under the Fit Results tab – the number of trios loaded from the Trio Names file at the time of cluster genotyping. A trio is one mother/father/child set of sample names. A trio is loaded if at least one trio member has an experiment in the Array Data section at the time of clustering. Note that trio comparison cannot be done if experiment data for all three members of the trio are not available.

 # Unique Samples

Number of unique samples among the experiments in the Array Data folder at the time of cluster genotyping.

Allele

(Assay and Target)

Target allele is the allele of the strand submitted for assay design.

Assay allele is the allele of the strand to be genotyped by the assay (displayed under the Assays tab when viewing cluster genotyping results).

The target and assay alleles may or may not be on the same strand of DNA. If not, they are complements.

Allele 1, Allele 2

In the Genotypes Short format, the genotype call for a marker on one of the chromosomes. Calls are alphabetically sorted into these two columns, so that for a CT call, "C" will appear in the Allele 1 column, and "T" in the Allele 2 column.

Allele Freq.

Minor Allele Frequency percentage.

Allele Freq. = 100%*minimum(p,q), where p = (2*numHom1Calls + numHetCalls)/2*numCalls, q = (2*numHom2Calls + numHetCalls)/ 2*numCalls, and only the best repeat of each experiment is considered. See Glossary for definition of best repeat experiment. If Trio information is supplied, child experiments are excluded. The reported allele frequency is reset to 0 for markers on the X and Y chromosome..

Array

GeneChip® Universal Tag Array

Array Definition

Specifies the function of each feature on the array.

Assay

A Targeted Genotyping Protocol test that utilizes a Molecular Inversion Probe (MIP) and reagents to genotype a locus.

Assay Allele

Possible alleles of strand for which the assay is designed.

Assay Call Rate %

Percentage of assays for which the genotype for both alleles can be determined. This rate is based on the number of successful genotypes divided by the number of attempted genotypes for an assay. It applies to all passed experiments only.

Assay Column

Column of the assay plate assigned to the experiment.

Assay Date

Date the assay plate barcode is entered into the database.

Assay Id

Assay identifier assigned by Affymetrix.

Assay Panel

A group of assays. Can be standard or customized.

Assay Panel File

A file that contains information about the assays grouped together to create an assay panel. This information includes assay IDs, target alleles, and assay alleles.

Assay Protocol

Barcode of the assay protocol followed to make the assay plate.

Assay Repeatability %

Percentage of calls in an assay that agree with the consensus calls from repeated experiments for the same sample.

Assay Status

The status of an assay in a genotype result. "Pass" if it passes all the user-adjustable Assay Cuts displayed when the user initiates a genotyping operation, or "Fail" if it fails at least one of the Assay Cuts.

Assay Trio Conc. %

Number of Mendel tests for the assay whose genotypes are consistent, divided by the total number of tests for the assay.

Auto

(and Auto Status)

In all tables, charts and windows where displayed – the experiment status automatically assigned by GTGS. The status can be Pass or Fail.

Average Signal (Assays table of Genotype Results)
Average sum of the allele 1 + allele 2 signal (Sum Signal) for the assay across all experiments in the cluster set. Used to sort assays by relative signal strength.

Best Repeat
As displayed under the Experiments tab when viewing cluster genotyping results – a *Y* indicates which experiment for a particular sample yielded the best call rate. If a sample was run only once, that experiment will be marked with a Y. If a sample was used for more than one experiment in a particular cluster, the experiment with the best call rate for that sample will be marked with a Y.

Call Confidence
A quality score for each genotype call, derived from the number of standard deviations away from the cluster center the data point lies, and the relative probability of the data point belonging to the most likely versus second most likely clusters. Call Confidence ranges from 0 to 1, with 1 being the most confident. Reported in the Genotypes Long file.

CEL File Location
As displayed in the Experiment Details window under the Summary tab – the directory where GTGS expects to access the .cel files. If the .cel are moved, the directory displayed in this window must be updated using the Change Location button.

Channel A, Channel C, Channel G, Channel T
The four channels on each array that are scanned. The data collected from each scan is saved in a file that includes a suffix designating the channel (A, B, C and D correlate to channel A, channel C, channel G and channel T respectively).

Channel Median S/N
The typical channel signal for assay features divided by the standard deviation of dimmest features.

Child
Sample name of the child of a trio set.

Chip Type
Type of Universal Tag Array used for an experiment, for example Universal 3K Tag Array.

Chrom. Position
Base pair position of the SNP on the chromosome.

Chromosome
Name of chromosome where SNP is located.

Cluster Fit Repeatability %
As displayed under the Fit Results tab – the percentage of calls in a cluster set that agree with the consensus calls from repeated experiments for the same sample.

Cluster Fit Trio Concordance %
Number of parent/child trio genotypes that are consistent with Mendelian inheritance divided by the total number of trio genotypes tested and expressed as a percentage.

Color Format
Specifies the number of color channels to scan for each array.

Column
Column of the relevant plate assigned to the experiment, to which a unique sample is assigned.

Common Primers
PCR amplification primers.

Complementary Tag The complement of a tag.

Completeness % As displayed under the Fit Results tab – the percentage of calls in a cluster set, in which:

Only passed assays are considered.

Only the best genotyped experiment replicate for each sample is considered.

Concentration User-reported sample concentration, in ng/µL.

Contrast Contrast = sinh(2y)/sinh(2), where y = (2*S2/Signal Sum) - 1 .

See Signal Sum for more information. If the array uses replicate features, the value reported in the Assay Allele Signals file is the average of the individual feature contrasts whose genotypes agree with the consensus genotype. If no consensus genotype exists, the average of all replicate feature contrasts is reported.

Controls CV % As displayed in the Experiment QC Summary table – the percent coefficient of variation of control feature signals across a array.

Converted Assay An assay that has passed genotyping quality criteria.

Current Barcode of the currently-selected item.

Current Location Region or well of the current plate.

Date In all tables and charts where displayed – the date the operation occurred.

Destination Barcode of the plate or array to which the current well contents are transferred.

Destination Location Region or well of the destination plate or array to which the current well(s) are transferred.

Exp. Call Rate % As displayed under the Experiments tab when viewing cluster genotyping results – the percentage of passed assays in the experiment that report full genotypes.

Exp. Count Experiment Count: number of experiments performed on a sample.

Exp. Repeatability % As displayed under the Experiments tab when viewing cluster genotyping results – the percentage of calls in an experiment that agree with the consensus calls from repeated experiments for the same sample.

Exp. Status The status of an experiment at the time it was genotyped. "Auto Pass" if the experiment passed QC and wasn't manually failed, "Manual Pass" if the experiment failed QC and was manually passed, and "Fail" if the experiment failed QC and wasn't manually passed.

Exp. Trio Conc. % Number of Mendel tests for the experiment whose genotypes are consistent with the experiments of the other members of the trio, divided by the total number of tests.

Experiment
A test that utilizes one assay panel to genotype one sample.

(and Experiment Name)
In all tables, charts and windows where displayed – the barcode for the array on which a particular sample was hybridized. Typically, the barcode is annotated.

External Id
User-supplied locus identifier.

Failed by Low Call Rate %
Percentage of assays failed during cluster genotyping due to low call rate.

Failed by Low H.W. P-Value %
Percentage of assays failed during cluster genotyping due to low Hardy Weinberg p-value.

Failed by Low Repeatability %
Percentage of assays failed during cluster genotyping due to low repeatability. This metric is relevant only if multiple experiments exist for the same sample within the same cluster.

Failed by Low Trio Concordance %
Percentage of assays failed during cluster genotyping due to low trio concordance. This metric is not relevant if Mendelian inheritance analysis is performed at the time of genotyping.

Failed?
User-specified status of the relevant plate.

Father
Sample name of the father of a trio set.

Feature ID
Identifier of each feature on the array.

Feature Type
The function of the array feature.

Fit By
User who created the specific Genotype Result.

Fit Status
Cluster fit status for an assay or an experiment (depends upon which table the term is in.)

Gender
As displayed in the Experiment QC Summary and other tables – the known gender of a sample. This information is imported into GTGS via Sample Info files. ***See also*** Inferred Gender.

Gene
Name of gene where SNP is located.

Genome Map
Source of the annotation information for the SNP.

Half %

(or Half Call Rate)
As displayed under the Assays and Experiments tabs when viewing cluster genotyping results – the half call rate. Only one of the two chromosomes can be called. When genotypes are exported, half calls are exported as no calls.

Hardy-Weinberg P-value

Statistical likelihood that this assay of the SNP, given the observed genotypes, obeys the Hardy Weinberg Equilibrium model. The Hardy Weinberg equilibrium model says that the frequency for a SNP of the genotype AA is pp, the frequency of AB is 2pq, and frequency of BB is qq, where p represents the frequency of allele A, q represents the frequency of allele B, and p + q = 1. This model is relevant for a large set of samples belonging to the same population. GTGS does not calculate the H.W.P-Value for X and Y chromosome SNPs. GTGS does not use the population annotation information associated with each sample when calculating the p-value. See Allele Freq. for more information on how p and q are calculated.

Het. %

As displayed under the Assays and Experiments tabs when viewing cluster genotyping results – the heterozygous call rate.

High Background Rate

Estimated percentage of assays in an experiment that will be uncalled due to high signal in the non-allele channels.

High Pixel CV Rate

Percent of uncalled assays due to array image artifacts or unsuccessful gridding.

Hom Tail Cut

If the ratio of Half to Het calls is above a threshold, it is assumed that a diffuse homozygous cluster is being incorrectly assigned both a hom and a het cluster. In this situation, both the het and minor hom clusters are turned off. Genotypes that would have been called from the het and minor hom cluster are now assigned "No Call" due to this cut. Typically, "MinHetToHalfRatio = 2".

Hom. %

As displayed under the Assays and Experiments tabs when viewing cluster genotyping results – the homozygous call rate.

Hyb Date

In all tables, charts and windows where displayed – the date the Hyb Plate barcode was entered into GTGS.

InActive

User state: inactive users do not appear in the list of available users during application login.

Index

Table index.

Inferred Gender

As displayed in the Experiment QC Summary table – the gender as called by GTGS when a sample has been processed using the Targeted Genotyping Protocol and an assay panel that includes gender control assays. ***See also*** Gender.

Label Date

Date the Label plate barcode is entered into the database.

Label Protocol

Barcode of the genotyping protocol followed to make the Label plate.

Locus

A genomic sequence of interest targeted by a MIP.

Low Call Confidence Cut %

As displayed under the Fit Results tab – the percentage of genotypes that were failed because they did not clearly belong to a particular cluster.

Low S/BG Cut

May be displayed for a data point in a cluster plot that is identified as "?/?" (No Call). A data point will fail the signal/background ratio test if the sum of its allele channel signals isn't sufficiently larger than the sum of its two non-allele channel signals. The setting MinSignalToBgnd defines this threshold, and is reported in the exported Genotype Settings file. See Normalized Signal for more information.

Low Signal Rate

Estimated percentage of assays in an experiment that will be uncalled due to weak signal in the allele channels.

Manual (Status)

As displayed in the Experiment QC Summary table – if an experiment status has been manually changed, the new value (Pass or Fail) is displayed in the manual column See Auto.

Map

See Genome Map.

Max Controls CV %

Percent coefficient of variation of control feature signals across a array, reported as the maximum value of all arrays in the experiment. Must be less than 30% for an experiment to be considered Pass.

MinAssayCallRate

As displayed under the Genotype Settings tab – a cluster genotype setting. Assays below this call rate are failed and no genotypes are reported.

MinAssayH.W.P-Value

As displayed under the Genotype Settings tab. *See also* Hardy-Weinberg P-value. Assays with an H.W.P-Value below this minimum will be failed, if this setting is used during genotyping. X and Y chromosome assays are not affected by this cut.

MinAssayRepeatability

As displayed under the Genotype Settings tab – a cluster genotyping setting. When multiple experiments of the same sample exist, the genotype of each experiment is compared against the consensus genotype for that sample. This comparison determines the repeatability of the assay. Assay repeatability scores that fall below the MinAssayRepeatability value are failed and no genotypes are reported.

MinAssayTrioConcordance

As displayed under the Genotype Settings tab – a cluster genotyping setting. If a Trio Names file is loaded during a clustering operation, Mendelian inheritance analysis will be performed for each assay, by comparing the genotypes of the mother, father, and child trio. A trio concordance rate (or consistency) of the genotypes is calculated for each assay. Assays below this minimum % trio concordance are failed, and no genotypes are reported.

MinCallConfidence

As displayed under the Genotype Settings tab – the cluster genotype setting that defines how clearly an assay must belong to a particular cluster to be genotyped. If the call confidence score falls below the MinCallConfidence value, the assay will not be genotyped. See Call Confidence.

MIP Allele

The allele targeted by the MIP. It can be on either strand of DNA.

Molecular Inversion Probe (MIP)

The name given to the DNA oligonucleotide that functions as a probe in a Targeted Genotyping Protocol test.

Mother Sample name of the mother of a trio set.

Normalized Signal (A,C,G, or T) The signals used to derive Contrast and Signal Sum. Intensity file (.cel) signals are normalized before QC metrics are determined, and may be further normalized at the point of genotyping to scale all arrays to the same average signal. If the array uses replicate features, the normalized signals reported in the Assay Allele Signals file are the average of the individual feature signals whose genotypes agree with the consensus genotype. If no consensus genotype exists, the average of all replicate features signals is reported.

Outlier Rate % In all windows, tables and charts where displayed such as the Experiment QC Summary table – see Rejected Outliers %.

Panel Conversion Rate Percentage of loci converted to an assay panel.

Passed Assays % As displayed under the Fit Results tab – the percentage of assays in a cluster set of experiments for which genotypes are reported.

Pixel CV In .dat files, the percent coefficient of variation of the pixels for each feature (standard deviation / raw signal.)

Plate Id Database plate identifier for plates that have been created in the database but don not yet have a barcode.

Plate Name File name of the file loaded by user that describes the plate contents.

Plate Type Plate can be of type Assay, Label, or Hyb.

Position (bp) Base pair position of the SNP on the chromosome.

Processed Processed data view, after background subtraction, color separation, and signal balancing between channels.

Project A collection of experiments tracked and/or genotyped as a set.

Protocol The barcode designation for the Targeted Genotyping Protocol being performed.

QC Call Rate (%) In all tables, charts and windows where displayed – the estimated percentage of assays clearly genotyped. Must be at least 80% for an experiment to be automatically assigned the status Pass.

QC Half Rate (%) In all windows, tables and charts where displayed such as the Experiment QC Summary table – the estimated percentage of assays marginally genotyped. Must be no more than 10% for an experiment to be automatically assigned the status Pass.

Raw Raw data view, with signal values as loaded into database.

Rejected Assays % As displayed under the Experiments tab when viewing cluster genotyping results –the percentage of genotypes cut because of poor data quality.

Rejected Outlier % As displayed in various tables and charts, the percent of assay features on an array that will not be genotyped due to rejected features in the allele channels. For arrays where there is only one feature per assay, this corresponds to the percent of assays that will not be genotyped.

Rejected Outliers As displayed in the Experiment Details window under the Array Views tab – a view that highlights the rejected features in a channel based on scanner signal saturation, blemishes, or high pixel CV.

Repeat Count As displayed under the Experiments tab when viewing cluster genotyping results – the number of experiments run on the same DNA sample.

Repeatability Number of concordant genotypes divided by the total number of compared genotypes in two or more repeated samples.

Sample Some quantity of DNA used in an experiment.

Sample Call Rate Number of successful genotypes divided by the number of genotypes attempted for one sample. Assays that are failed for all experiments in a cluster set are not included in this calculation.

Sample Info General information for each sample, which can be imported via a Sample Info file. Includes Sample Name, and optional Gender, Population, and Species fields.

Sample Location In all tables, charts and windows where displayed – the sample plate barcode and the well location of a particular sample.

Sample Name In all tables, charts and windows where displayed – the name assigned to a particular DNA sample.

Sample ng/ul User-reported sample DNA concentration in ng/µL.

Sample Repeat Count See Repeat Count.

Sample Type User-specified sample type. Sample types can be water, control, or sample.

Sample Well Well location of sample in the sample plate.

Scanner Type As displayed in the Experiment Details window under the Summary tab – the type of scanner used to scan the array.

Sig./Bkgd. In all windows, tables and charts where displayed such as the Experiment QC Summary table – signal/background is the median ratio of assay allele/non-allele channel signals. Must be ≥ 20 for an experiment to be automatically assigned the status Pass. Calculated as (S1 + S2) / (B1 + B2), where S = allele signal and B = background (non-allele signal)

Signal Sum Signal 1 + Signal 2, where Signal 1 is the first assay allele, e.g. "C Normalized Signal" for a C/T SNP, and Signal 2 is the second assay allele, e.g "T Normalized Signal". If the array uses replicate features, the value reported in the Assay Allele Signals file is the average of the individual feature Signal Sums whose genotypes agree with the consensus genotype. If no consensus genotype exists, the average of all replicate feature contrasts is reported.

Signal/Noise As displayed in the Channel Metrics charts – the typical channel signal for assay features / standard deviation of dimmest features.

Software Version Version of the current software at the time of the creation of the Genotype Result.

Source Previous plate in the series SMP > ANN > ASY > LBL > HYB > ARRAY that contained the same samples during the genotyping procedure.

Source Location Region or well of the source plate.

Status Status of the current well transfer. Can be ready or complete.

Strand Switched? Has the assay been designed to interrogate the strand complementary to the target strand (from which the target allele genotypes can be derived)? A check indicates this is the case.

Tag Features Number of potential array features that can be used to genotype SNPs.

Tag Id Identifier of the unique oligo sequence present on the array.

Target Allele Allele of strand user wants genotyped.

Thermocycler Numerical identifier of the specific thermocycler used to process this plate.

Trio Conc. % (Trios table of Genotype Results) Number of Mendel tests in a trio whose genotypes are consistent, divided by the total number of tests.

Trio Status If the current cluster does not have data for all members of the Trio, no Mendelian inheritance tests can be performed, and the status is Not Set. If data is available for all members, the status is Valid.

Type Type of samples loaded onto a particular plate. Sample types can be external samples or internal samples.

User User name of person accessing the software.

User Comments In all windows where displayed such as the Experiment Details window – a field in which users can enter general comments.

Well Well location of the sample in the current plate.

X Array column coordinate; feature location on the array of the tag oligo, X position.

X Max	Number of feature columns in the array.
Y	Array row coordinate; feature location on the array of the tag oligo, Y position
Y Max	Number of feature rows in the array.

INDEX

T

W