applied biosystems

Axiom™ Analysis Suite 3.1

USER GUIDE

Publication Number 703307
Revision 5

ThermoFisher SCIENTIFIC

For Research Use Only. Not for use in diagnostic procedures.

The information in this guide is subject to change without notice.

DISCLAIMER

TO THE EXTENT ALLOWED BY LAW, LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) WILL NOT BE LIABLE FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING YOUR USE OF IT.

Important Licensing Information

This product may be covered by one or more Limited Use Label Licenses. By use of this product, you accept the terms and conditions of all applicable Limited Use Label Licenses.

Legal entity

Life Technologies | Carlsbad, CA 92008 USA | Toll Free in USA 1 800 955 6288

TRADEMARKS

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

Contents

1 Introduction

Overview

Axiom Analysis Suite (AxAS) enables you to perform the following functions:

- Run QC and Genotyping Algorithms.
- View QC Data within tables and graphs at a Sample and/or SNP level.
- View Cluster Graphs with the ability to change calls and/or highlight by attribute.
- Export your Data.

Software and Hardware Requirements

64-bit Operating System	Speed	Memory (RAM)	Available Disk Space[1]	Web Browser
Microsoft Windows® 7 (64 bit) Professional with Service Pack 1	2.83 GHz Intel Pentium Quad Core Processor	16 GB	150 GB HD + data storage Refer to table below.	Internet Explorer 8.0 and above
Microsoft Windows 10 (64 bit) Professional	2.83 GHz Intel Pentium Quad Core Processor	16 GB	150 GB HD + data storage Refer to table below.	Internet Explorer 8.0 and above

[1] Minimum storage requirements are for a single run. Total storage space should include additional space for data storage of input and output files from current and previously completed analyses. In addition, you must have a minimum of 5GB of free space on your C: drive to run an analysis.

Sample Data Size Estimates and Required Disk Space

Before using AxAS, make sure you have enough disk space. See the table below for size estimates. Note: The estimates shown include the contents of the batch name folder.[2]

# of Markers	Storage Type[3]	50 samples	100 samples	500 samples	1000 samples	5000 samples
50K	Input	1.33 GB	2.66 GB	13.3 GB	26.6 GB	133 GB
	Output	158 MB	286 MB	1.27 GB	2.51 GB	12.4 GB
	Total	1.49 GB	2.95 GB	14.57 GB	29.11 GB	145.4 GB
500K	Input	1.33 GB	2.66 GB	13.3 GB	26.6 GB	133 GB
	Output	1.53 GB	2.77 GB	12.6 GB	25.0 GB	124 GB
	Total	2.86 GB	5.43 GB	25.9 GB	51.6 GB	257 GB
850K	Input	1.33 GB	2.66 GB	13.3 GB	26.6 GB	133 GB
	Output	2.59 GB	4.69 GB	21.4 GB	42.4 GB	209 GB
	Total	3.92 GB	7.35 GB	34.7 GB	69.0 GB	342 GB

[2] A batch name folder is auto-generated during the analysis process. This folder includes all the necessary files needed to view your analysis results in the Viewer.

[3] Input is the storage size required for CEL files to be analyzed. Output is the storage size required for analysis results files.

Installation Instructions

1. Go to thermofisher.com, then navigate to the following location:

   Applications and Techniques > **Life Sciences** > **Microarray Analysis** > **Software** > **Axiom Analysis Suite**

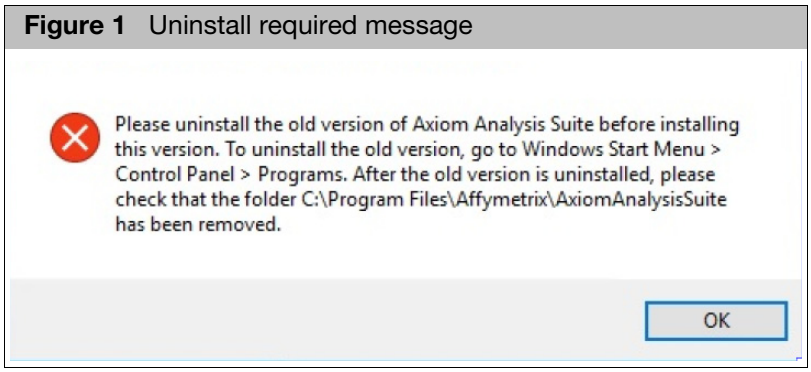
   Or

   Click on this link:

   https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/axiom-analysis-suite.html
2. Locate and download the zipped Axiom Analysis Suite software package.
3. Unzip the file, then double-click **AxiomAnalysisSuiteSetup.exe**.
4. Follow the on-screen instructions to complete the installation.

   If your system has a previous version installed, the following message appears: (Figure 1)

Figure 1 Uninstall required message



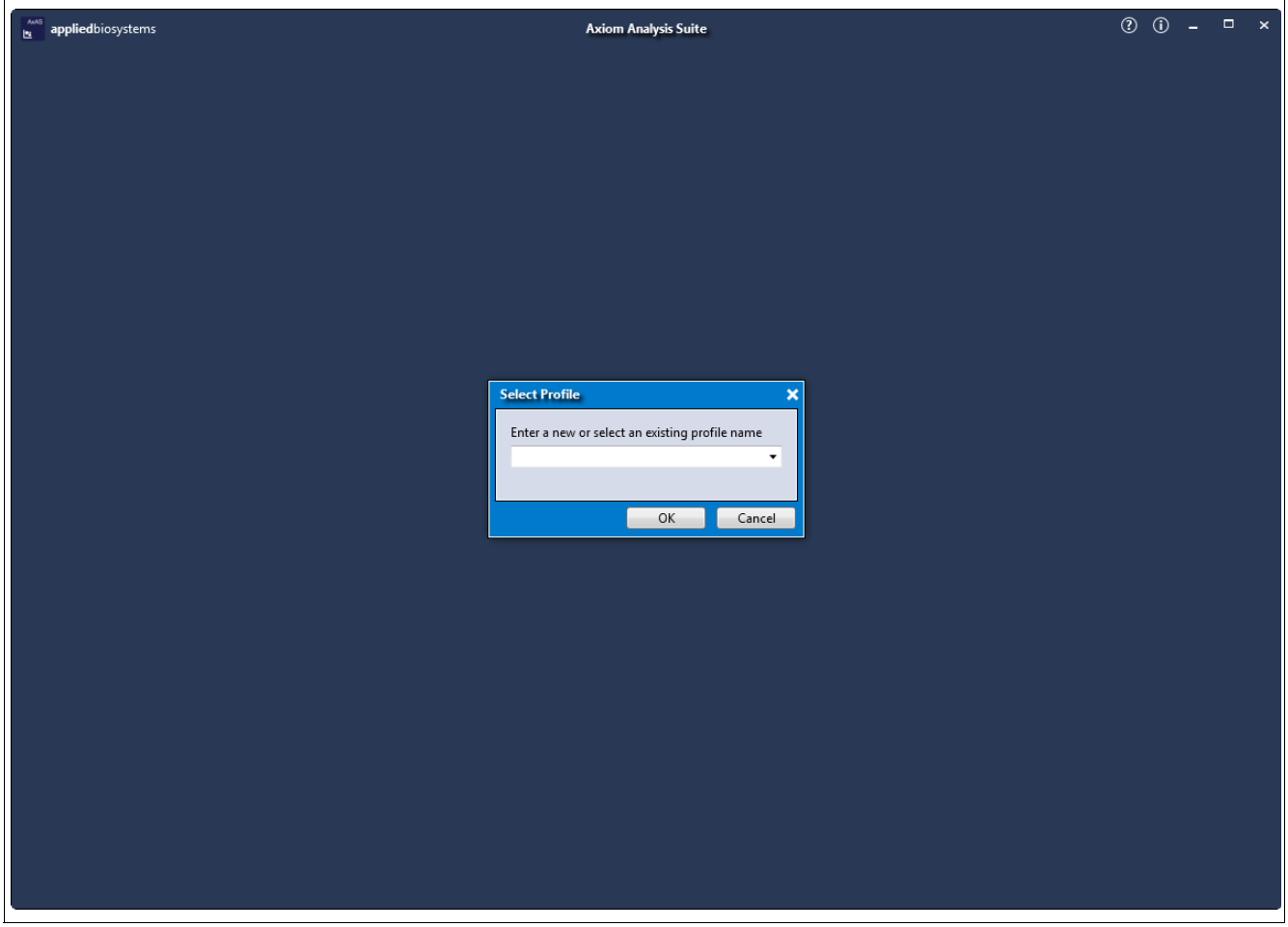
Acknowledge the message, click **OK**, then go to "Uninstalling" on page 18.

Starting Axiom Analysis Suite

1. Double-click on the Axiom Analysis Suite Desktop shortcut or click **Start** > **All Programs** > **Thermo Fisher Scientific** > **Axiom Analysis Suite**.

   The following window appears: (Figure 2)

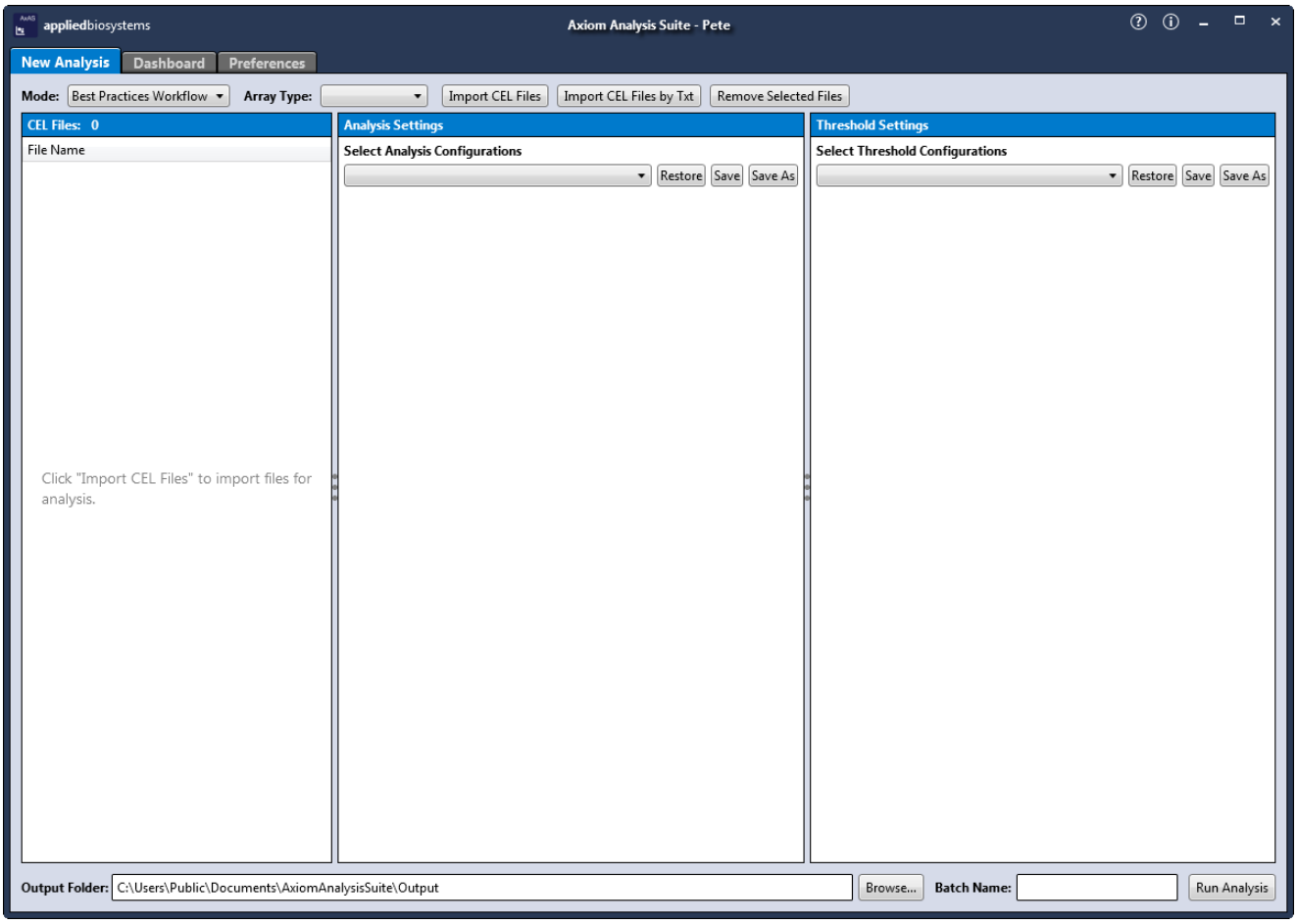
Figure 2 Opening window



2. Enter a new profile name or click the down-arrow to select an existing profile name.
3. Click **OK**.

   The following window appears: (Figure 3)

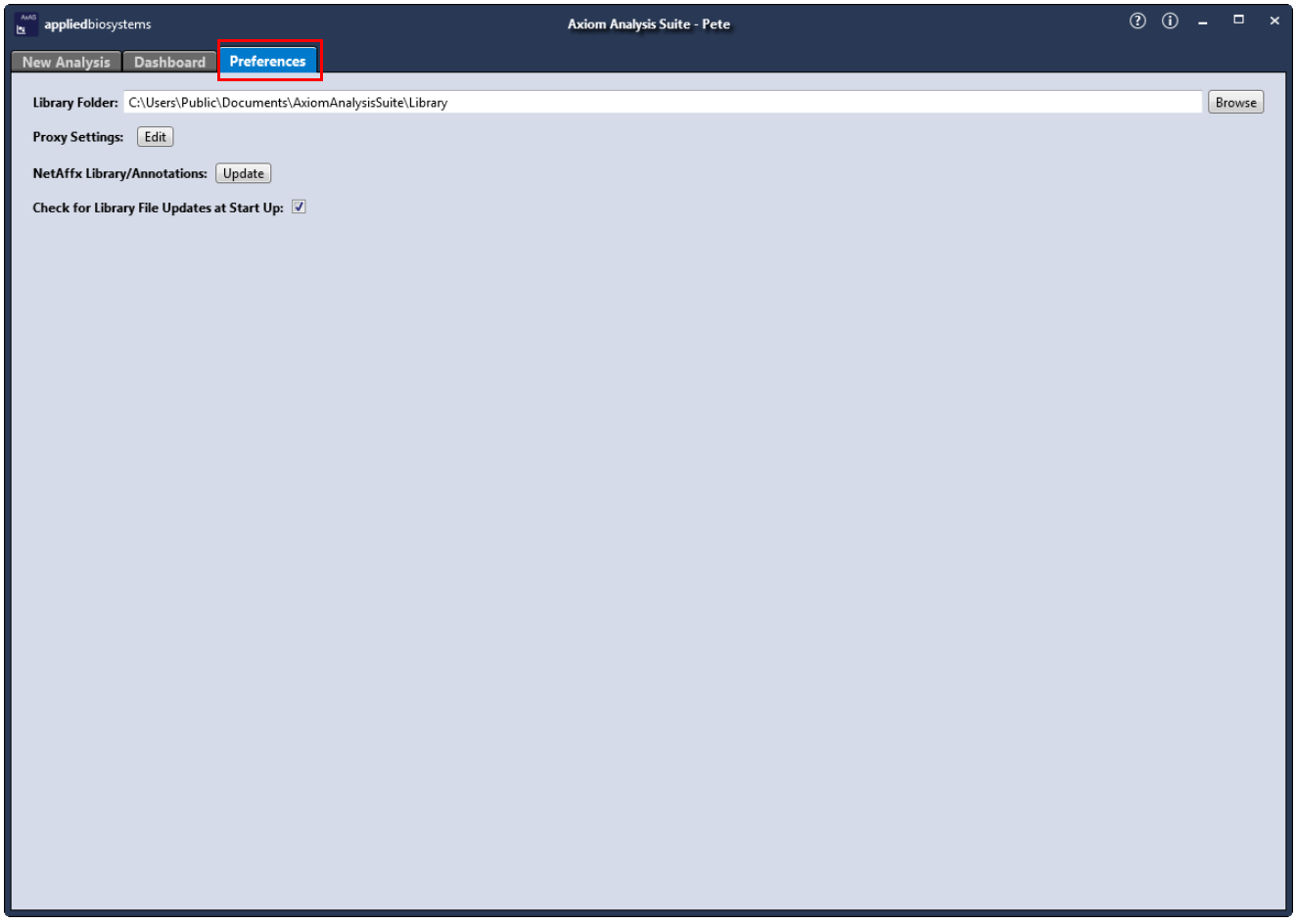
Figure 3 Main window



Using the Preferences Window Tab

Click the **Preferences** window tab (Figure 4) to setup or change a library path, edit Proxy settings, download or update Library/Annotation files.

Figure 4 Main Preferences window



Changing the Default Library Folder/Path

IMPORTANT! The library folder contains the library and annotation files required to run the Axiom Analysis Suite software.

Do the following to change the default Library folder/path:

1. Click **Browse** (right of library path field).

   The **Select Library Folder** window appears.
2. Navigate to the new location you want the library folder to reside.
3. Click **New Folder**.
4. Rename the New Folder (as you normally would), then click **Select Folder**.

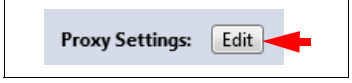
   Your newly assigned Library folder is set and reflected in the Library Folder directory/path field, as shown in Figure 5.

Figure 5 Populated Library Path example

Setting Up Proxy Server Access

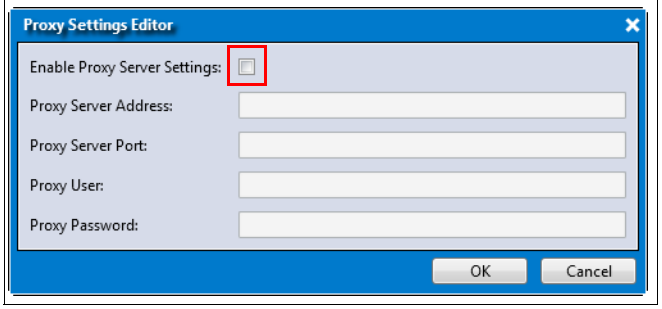
If your system has to pass through a Proxy Server before it can access the NetAffx server (Internet), click the **Edit** button. (Figure 6)

Figure 6 Proxy Settings



The following window appears: (Figure 7)

Figure 7 Proxy Settings Editor window



5. Click the **Enable Proxy Server Settings** check box (Figure 7), then contact your IT department for help with completing the required text fields.
6. Click **OK**.

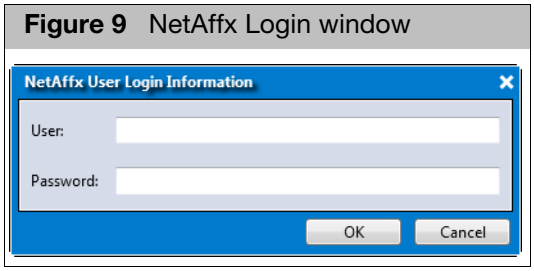
Updating NetAffx Library/Annotations

1. Click on the **Update** button. (Figure 8)

Figure 8 Update button

The following window appears: (Figure 9)

Figure 9 NetAffx Login window

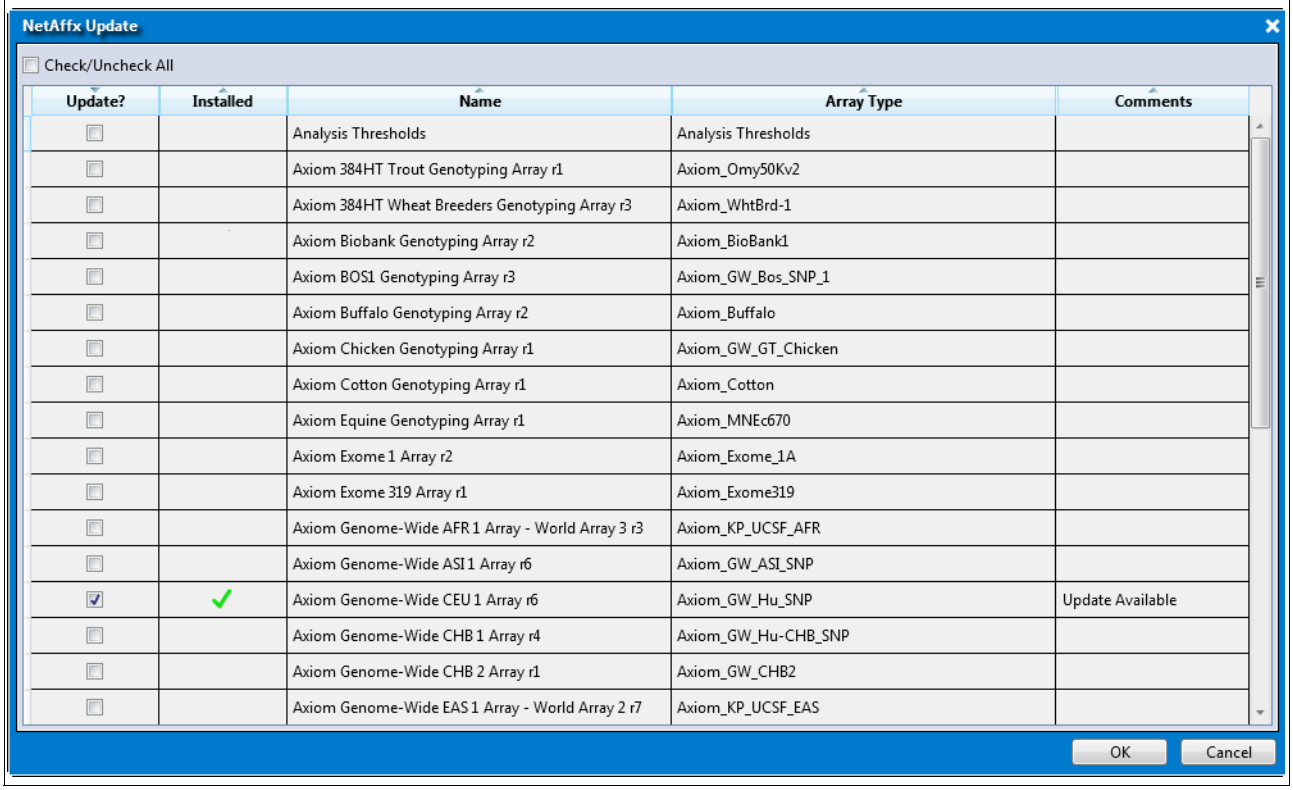


2. Enter your NetAffx account email and password, then click **OK** or go to www.netaffx.com and click **Register** to sign up.

Note: If you are unable to connect to NetAffx, make sure you have an active Internet connection, and/or correct Proxy Server settings.

The NetAffx Update window appears. (Figure 10)

Figure 10 NetAffx Update window



3. You must click the check box(es) that correspond with the type of CEL files you want to analyze.

   Click the **Check/Uncheck All** check box to select/deselect all the listed check boxes.

4. Click **OK**.

   An Installing Updates progress bar appears.

Enabling/Disabling Check for Library File Updates at Start Up

1. This check box (Figure 11) is checked by default to enable automatic Library File update alerts each time you launch the Axiom Analysis Suite application. (Recommended)

Figure 11 Auto-update notifications check box

Installing Custom Array Library Files

IMPORTANT! Library files for custom designs must be manually installed.

1. Download the zip package provided to you by Affymetrix Bioinformatics Services.
2. Unzip the contents of the analysis library files into a single sub-folder within the library file folder.

   For multi-species designs, each species should be in its own sub-folder. There should be no other folders within each sub-folder and all annotation information must be in the same location as the .CDF file.

Uninstalling

IMPORTANT! Upgrade installations are NOT supported. You must uninstall the existing version of Axiom Analysis Suite BEFORE installing this new version. Administrative rights to the computer are required before you can uninstall the Axiom Analysis Suite software. For your convenience, no existing library files or user settings are removed during the uninstall process.

Windows 7

1. Click **Start > Control Panel**.

   The Control Panel window appears.
2. Click the **View by** drop-down menu (upper-right), then click to select **Category**.
3. In the **Programs** category, click **Uninstall a program**.

   The Programs and Features window appears.
4. Click to select **Axiom Analysis Suite**, then click **Uninstall**.
5. Follow the on-screen instructions.
6. After the uninstall process is complete, close the Programs and Features window.
7. Use Windows Explorer as you normally would to navigate to the directory: **C:\Program Files\Affymetrix**
8. Verify that the **Axiom Analysis Suite** folder has been removed.
9. If the folder is present, double-click on it to open it.
10. Search for any files you want to keep, then move them to different (easily accessible) location.
11. Delete the **Axiom Analysis Suite** folder.
12. Close all open windows, then install the new version, as described in the "Installation Instructions" on page 11.

Windows 10

1. Click the Windows icon (bottom left corner).
2. Click **All apps > Windows System > Control Panel**.

   The Control Panel window appears.
3. In the **Programs** category, click **Uninstall a program**.

   The Programs and Features window appears.
4. Click to select **Axiom Analysis Suite**, then click **Uninstall**.
5. Follow the on-screen instructions.
6. After the uninstall process is complete, close all open windows.
7. Use Windows Explorer as you normally would to navigate to the directory: **C:\Program Files\Affymetrix**
8. Verify that the **Axiom Analysis Suite** folder has been removed.
9. If the folder is present, double-click on it to open it.
10. Search for any files you want to keep, then move them to different (easily accessible) location.
11. Delete the **Axiom Analysis Suite** folder.
12. Close all open windows, then install the new version, as described in the "Installation Instructions" on page 11.

2 Performing an Analysis

After downloading the library and annotation files that match the array type of the CEL files you want to analyze, click the **New Analysis** tab.

The **New Analysis** window and its three individual panes appear, as shown in Figure 12.

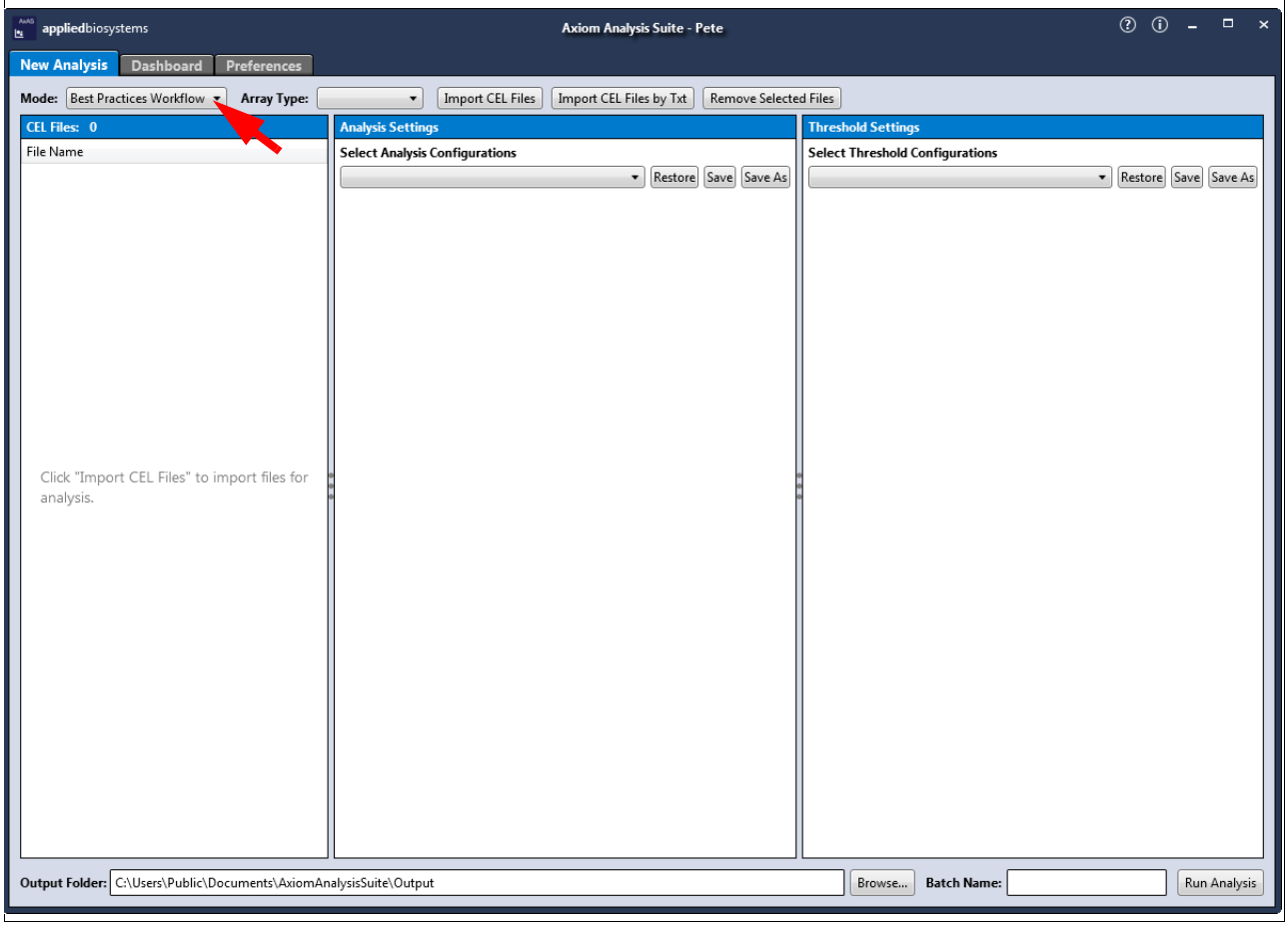
Setting Up an Analysis

If you want to run and view CN-aware genotypes, go to page 122. If not, continue to "Selecting a Mode (Workflow)".

Selecting a Mode (Workflow)

1. From the main Axiom Analysis Suite window tab, click the **Mode** drop-down. (Figure 12)

Figure 12 Main window



2. Click to select the workflow you want to use.

- **Best Practices Workflow (Default):** This workflow performs quality control analysis for samples and plates, genotypes those samples which pass the defined QC thresholds, and then categorizes the probe sets to identify those whose genotypes are recommended for statistical tests in downstream study. Details are available in the Axiom Genotyping Solution Data Analysis Guide (P/N 702961)
- **Sample QC:** This workflow performs the quality control analysis for samples and plates. Note this workflow does not produce genotype calls for the passing samples.
- **Genotyping:** This performs genotyping on the imported CEL files, regardless of the sample and plate QC metrics. Note: Including samples that do not pass defined QC thresholds may reduce the quality of the results for passing samples.
- **Summary Only:** This workflow produces a summary of the intensities for the probe sets for use in copy number analysis tools. Note: Summary Only does not perform sample QC nor genotyping.

Selecting an Array Type

1. Click the **Array Type** drop-down to select the array type you want to use.

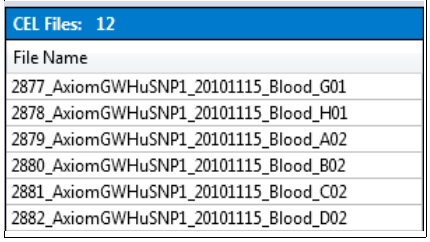
Importing CEL Files

1. Click **Import CEL Files**.

   The **Add CEL Files** window appears.
2. Navigate to your CEL file location. Make sure the CEL Files you select coincide with the array type you selected earlier, otherwise a warning message appears.
3. Single-click on a CEL file or Ctrl click, Shift click, or press Ctrl A (to select multiple files).
4. Click **Open**.

   The CEL Files pane populates and displays your selected CEL files. (Figure 13)

Figure 13 Populated CEL File pane example



Importing CEL Files by Text

IMPORTANT! The CEL files paths in a TXT file are not recognized (and will not import) if they begin with back slashes (e.g: **path\to\my\file.cel**).

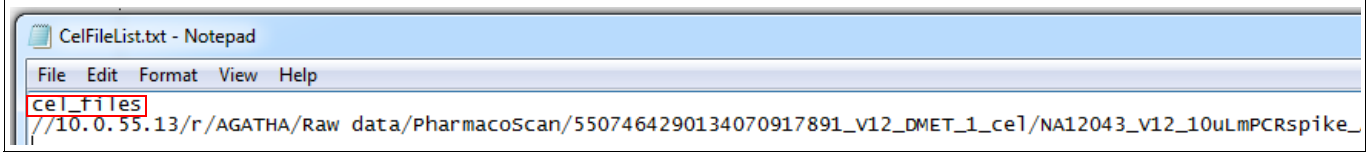
Before importing, the two starting back slashes must be replaced with forward slashes (e.g: //**path\to\my\file.cel**).

Also, your ***.txt** list file must start with the header **cel_files**, as shown in Figure 14.

1. Click **Import CEL Files by Txt**.

   The **Import CEL Files by Txt** window appears.
2. Navigate to the .txt file that contains the list of CEL files you want to process.

Figure 14 Text CEL file list example shown in Notepad



Make sure the CEL Files you select coincide with the array type you selected earlier, otherwise a warning message appears.

3. Single-click on a CEL file or Ctrl click, Shift click, or press Ctrl A (to select multiple files).
4. Click **Open**.

   Your CEL Files pane populates and displays each CEL file extracted from your selected text file.

Removing Selected CEL Files

Use this option to remove unwanted CEL files.

1. Single-click on a CEL file or Ctrl click, Shift click, or press Ctrl A (to select multiple files), then click **Remove Selected Files**.

Setting Up an Analysis Configuration

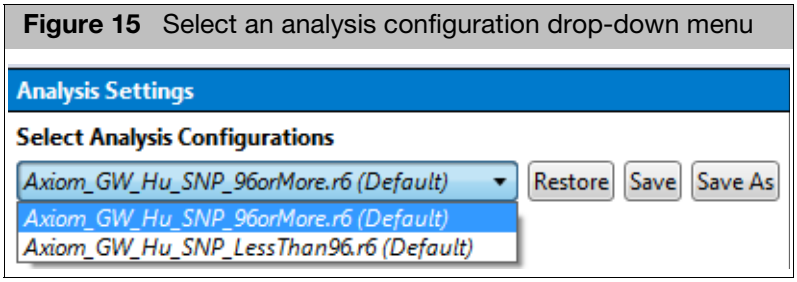
The Analysis Settings are populated based on the Mode (Workflow) chosen. For example, if Genotyping mode is selected, the Sample QC section of the Analysis Settings is hidden and only the Genotyping section is visible.

Selecting an Analysis Configuration

1. It is highly recommended you click the drop-down menu (Figure 15) and select the option that best matches the number of samples you want to analyze.

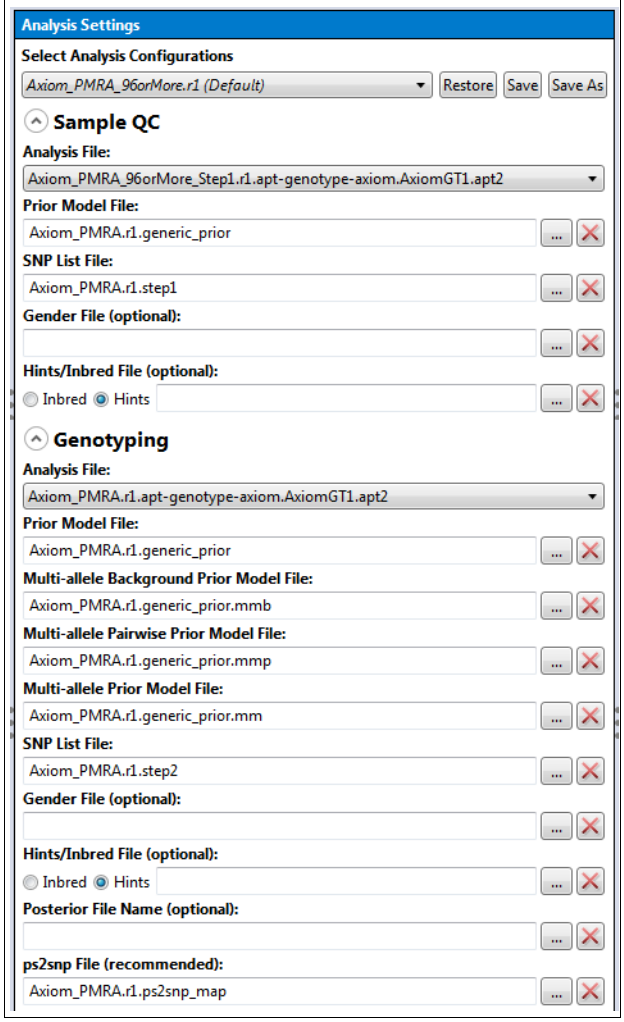
   Note: The default configuration options displayed in the drop-down menu are based on your array type.

Figure 15 Select an analysis configuration drop-down menu



After selecting the appropriate default for the number of your samples, the Analysis Setting pane auto-populates, as shown in Figure 16.

Figure 16 Auto-populated Analysis Setting pane example



Using the Analysis Settings Fields

Follow the instructions below to create a new analysis configuration or edit a pre-populated field(s).

Sample QC Fields

1. Click the **Analysis File** drop-down button to select the appropriate XML file.
2. Click the **Prior Model File** Browse ... button.
   The Prior Model File window appears.
3. Navigate and select the appropriate file, then click **Open**.
   Your newly assigned filename is displayed.
4. (Optional) Click the **SNP List File** Browse ... button.
   The SNP List File window appears.
5. Navigate and select the appropriate file, then click **Open**.
   Your newly assigned filename is displayed.
6. (Optional) Click the **Gender File** Browse ... button.
   The Gender File window appears.
7. Navigate and select the appropriate file, then click **Open**.
   Your assigned filename is displayed.
8. (Optional) Click the **Hints/Inbred File** Browse ... button.
   The Hints/Inbred File window appears.
9. Navigate and select the appropriate file, then click **Open**.
   Your newly assigned path is displayed.
10. Click the either the **Inbred** or **Hints** radio button.

Genotyping Fields

1. Click the **Analysis File** drop-down button to select the appropriate XML file.
2. Click the **Prior Model File** Browse ... button.
   The Prior Model File window appears.
3. If multi-allelic probesets are available, the following fields appear:
   - Multi-allele Background Prior Model File
   - Multi-allele Pairwise Prior Model File
   - Multi-allele Prior Model File
   - If you want to change the currently displayed Model file, click the appropriate Multi-allele Browse ... button to select a different Model file.
4. Navigate and select the appropriate file, then click **Open**.
   Your newly assigned filename is displayed.
5. (Optional) Click the **SNP List File** Browse ... button.
   The SNP List File window appears.
6. Navigate and select the appropriate file, then click **Open**.
   Your newly assigned filename is displayed.
7. (Optional) Click the **Gender File** Browse ... button.
   The Gender File window appears.

8. Navigate and select the appropriate file, then click **Open**.

   Your assigned filename is displayed.
9. (Optional) Click the **Hints/Inbred File** Browse … button.

   The Hints/Inbred File window appears.
10. Navigate and select the appropriate file, then click **Open**.

    Your assigned filename is displayed.
11. Click the either the **Inbred** or **Hints** radio button.
12. (Optional) Click the **Posterior File Name** Browse … button.

    The Posterior File Name window appears.
13. Navigate to a location where you want to save your posterior file, enter a name, then click **Open**.

    Your assigned filename is displayed.
14. Click the **ps2snp File** Browse … button.

    The ps2snp File window appears.
15. Navigate to your ps2snp-file location, then click **Open**.

    Your newly assigned filename is displayed.
16. (Optional) If the library package supports it, click the **Genotype Frequency** Browse button.

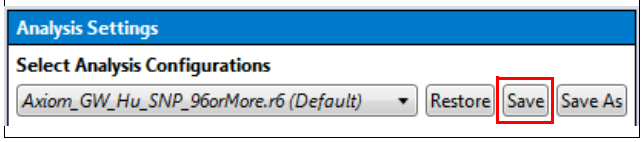
    The Genotype Frequency window appears.
17. Navigate to a location for your genotype frequency file, enter a name, then click **Open**.

    Your assigned filename is displayed.

Saving your Analysis Configuration

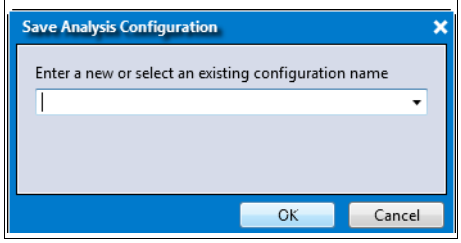
1. After editing your Analysis Configuration settings, click Save (top of Analysis Setting pane). (Figure 17)

Figure 17 Analysis Configuration Save window



If the Analysis Configuration that came with the library package already exists, then a **Save Analysis Configuration** window appears. (Figure 18)

Figure 18 Analysis Configuration Save window



2. Enter a different configuration name, then click **OK**.

   Your new Analysis Configuration name is saved and available for use in the **Select Analysis Configuration** drop-down menu.

Modifying an Existing Analysis Configuration

1. Click the **Select Analysis Configuration** drop-down, then click to select the saved analysis configuration you want to modify.

Do one or more of the following to modify an existing analysis configuration:

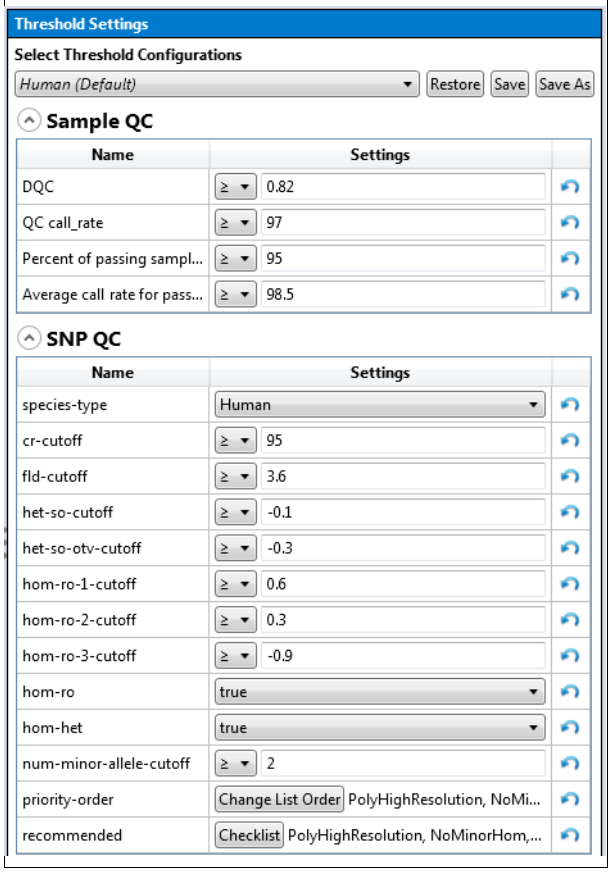
- Click the applicable File field's **Browse** button to navigate to a different location, then click **Open** to reassign its path.
- If needed, click a File field's  button to delete a displayed path setting.
- Click the **Restore** button to return to the last saved values of the analysis configuration file.
- Click the **Save** button to overwrite your previously saved configuration
- Click the **Save As** button to save your modified configuration with a different name. [Recommended]

Setting Up Threshold Settings

The settings shown in the **Threshold Setting** pane (Figure 19) are based on the Mode (Workflow) you selected.

For Sample QC and SNP QC name definitions, see "Threshold Names" on page 149.

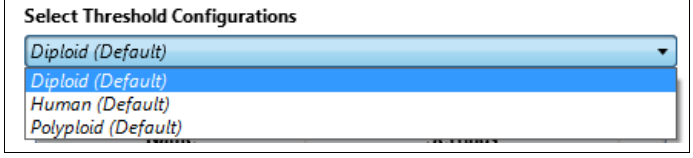
Figure 19 Automated QC Mode Threshold Settings pane example



Customizing Thresholds

Click the **Select Threshold Configuration** drop-down (Figure 20) to select an appropriate Default Threshold for your starting point. Note: The comparison signs/ operators are preset and cannot be changed.

Figure 20 Select Threshold Configuration



Sample QC

All the Sample QC Threshold Settings are populated with default values.

1. Click inside each text field to enter a different value, as shown in Figure 21.

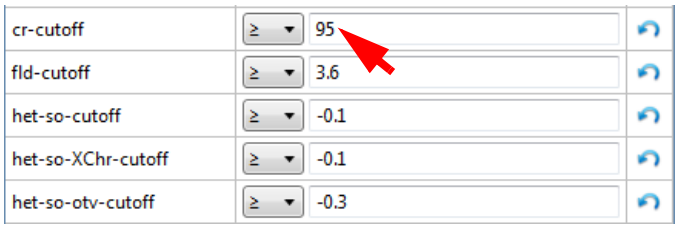
   Click the text field's button to return its value back to its last saved value within the threshold configuration file.

Figure 21 Threshold Name text field example

SNP QC

1. Click the **species-type** drop-down menu to select a different species type.
2. Click inside each text field to enter a different value, as shown in Figure 22. Note: General Rule: The het-so-otv-cutoff should be less or equal to het-so-cutoff and het-so-XChr-cutoff.

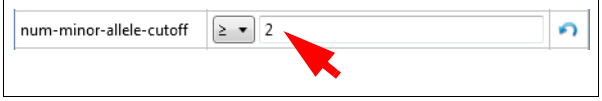
Figure 22 SNP QC text fields



   Note: To return a value back to its last saved value, click the text field's button.

3. Use the **hom-ro and hom-het** drop-down menus to change their True or False values.
4. Click inside the **num-minor-allele-cutoff** text field to enter a different value, as shown in Figure 23.

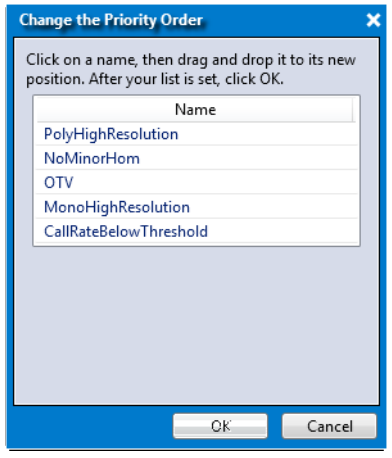
Figure 23 SNP QC text fields



5. The priority-order option enables you to change the order of categories when determining which probesets are selected as the best probeset for a SNP. To change the priority-order of your SNP QC Metric, click Change List Order.

The following window appears: (Figure 24)

Figure 24 Change the Priority Order window



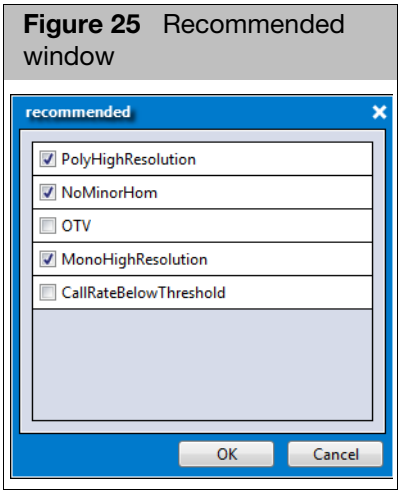
6. Click and hold onto the selection you want to move, then drag and drop it into its new position. After you get the order of priority you want, click **OK**.

   Note: To return the list back to its default priority, click the priority-order field's button.

7. Use the **recommended** checklist to choose the PS_Classification conversion types for your analysis. To change the recommended options, click Checklist.

   The recommended window appears. (Figure 25)

Figure 25 Recommended window



8. Click to check/uncheck the available recommended options, then click **OK**.

Note: If all recommended options are unchecked, the software uses the following default values:

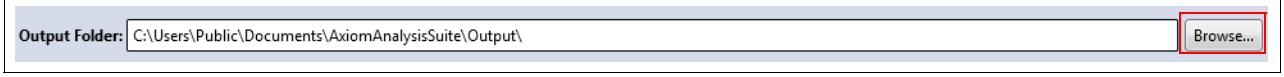
- **Human**: PolyHighResolution, NoMinorHom, MonoHighResolution
- **Diploid**: PolyHighResolution, NoMinorHom, MonoHighResolution
- **Polyploid**: PolyHighResolution

Assigning an Output Folder Path

Assigning a New Output Folder Path

1. Click the **Output Folder** path's **Browse** button. (Figure 26)

Figure 26 Output Folder field



An Explorer window appears.

2. Navigate to the recommended path C:\Users\Public\Documents\AxiomAnalysisSuite\Output, then click **Select Folder**.

Your selected output folder path is now displayed.

Adding Sub-Folders

Note: To better organize your output results, you can add sub-folders to your newly assigned output result path's folder.

To add sub-folders to your newly assigned result path's folder:

1. Click the Output Folder's **Browse** button to return to your assigned output path and/or folder.
2. In the Explorer window, click **New Folder**.
3. Enter a sub-folder name.
4. Click **Select Folder**.

   The newly created sub-folder now appears in the output result information window.
5. Repeat the above steps 1-4 to add more sub-folders, then click **Select Folder**.

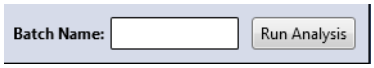
Assigning a Batch Name

The batch file is produced while your analysis is running and includes all the necessary files needed to view your analysis in the Axiom Analysis Suite Viewer.

1. Enter a name in the **Batch Name** field. (Figure 27)

IMPORTANT! Each Batch Name you create must be unique for the set of batches listed in the Dashboard window tab and unique within the same destination folder.

Figure 27 Enter a Batch Name



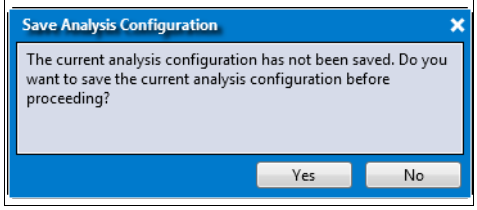
Note: A folder (with the same name as your entered batch name) is auto-generated during the analysis process. This folder includes all the necessary files needed to view your analysis results in the Viewer.

Running your Analysis

1. Click **Run Analysis**.

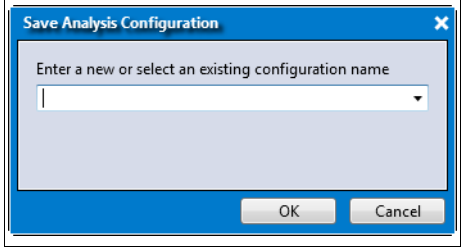
- If you have not saved any changes to your configured Analysis Settings, a Save Analysis Configuration window appears. (Figure 28) Click **Yes**.

Figure 28 Save Analysis Configuration prompt window



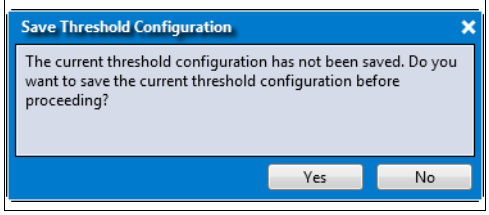
- Enter a new analysis name or use the drop-down to select a previously saved name, then click **OK.** (Figure 29)

Figure 29 Save Analysis Configuration window



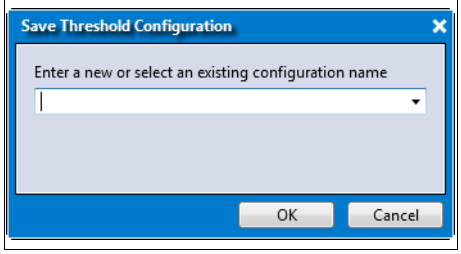
- If you have not saved any changes to your configured Threshold Settings, a Save Threshold Configuration window appears. (Figure 30) Click **Yes**.

Figure 30 Save Threshold Settings prompt window



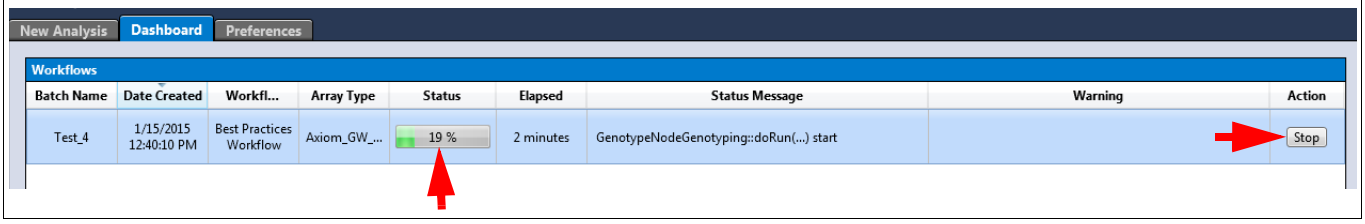
- Enter a new threshold name or use the drop-down to select a previously saved name, then click **OK.**

Figure 31 Save Threshold Settings window



The Dashboard window/tab appears and shows the status of your running analysis. (Figure 32) Click Stop to cancel an analysis in progress.

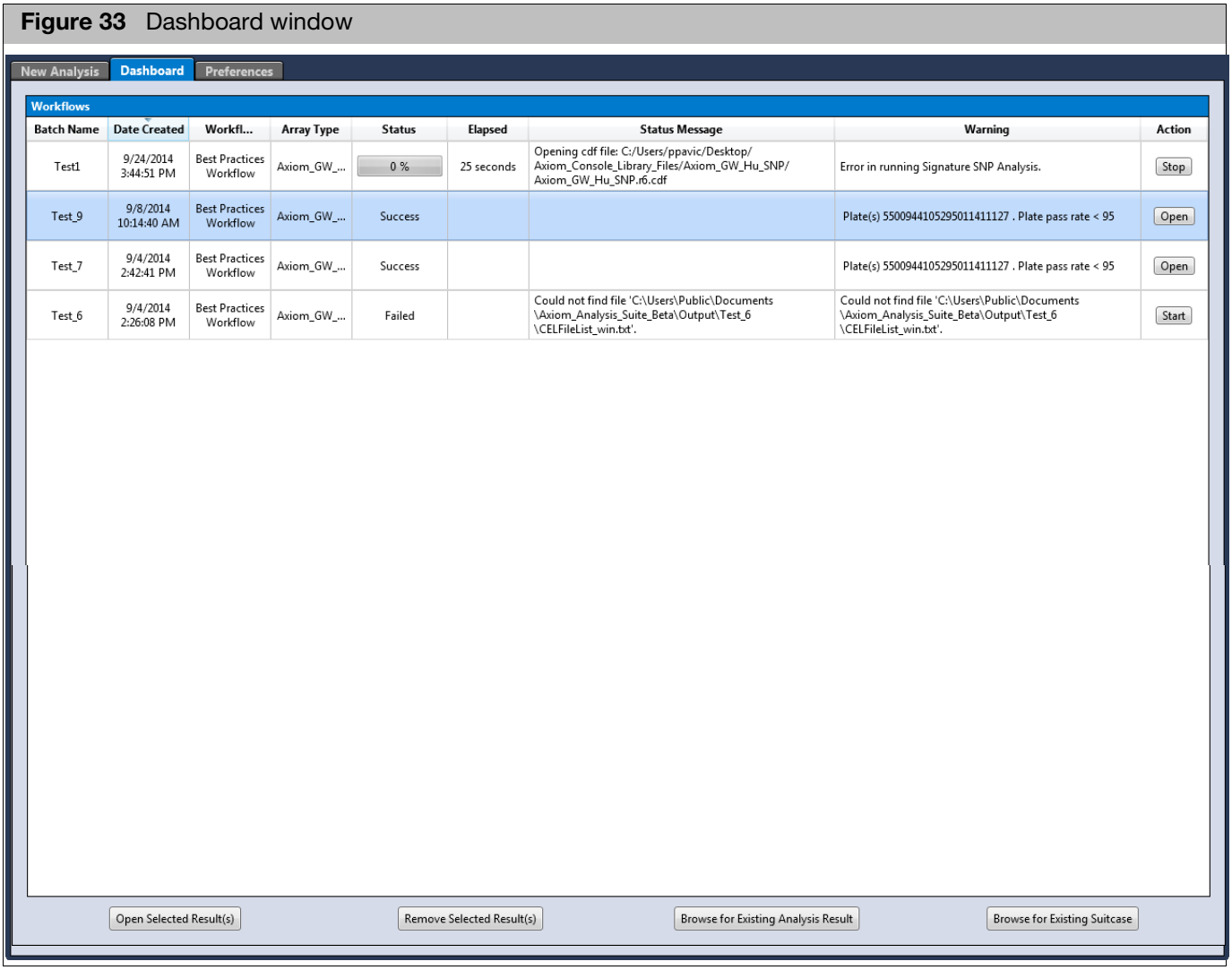
Figure 32 Dashboard window/tab - Status bar and Stop button example



Using the Dashboard Window Tab

The Dashboard tab window displays existing results. (Figure 33)

Figure 33 Dashboard window



Open Selected Result(s)

Do one of the following to open a selected result:

- Click on the Action column's **Open** button to open a study.

  Note: The **Open** button is relabeled **Start** if the Dashboard's Status column displays Failed. A workflow can fail if there is a processing error (for example, needed files are unavailable) or if no samples pass QC. If the Warning column indicates that no samples passed QC, yet you still want to open the study to review its QC results, use one of the alternate methods below to open a selected study.
- Single-click on a study, then click **Open Selected Result(s)**.
- Double-click on a study to open it.
- Right-click on a study, then click **Open**. Note: Click on any of the Workflow header columns to sort your listed results.

  After a few moments, the Viewer opens and displays your study (as you last left it).

Remove Selected Result(s)

Do one of the following to remove a selected result:

- Single-click to highlight the analysis you want to remove, then click **Remove Selected Result(s)**.
- Right-click on the highlighted analysis, then click **Remove from List**.

Viewing the Results Folder in Windows Explorer

1. Right-click on the highlighted analysis, then click **Open in Windows Explorer**.

   Your Analysis Results folder now appears in the Explorer window.

Browsing For Existing Analysis Results

The Axiom Analysis Suite Dashboard displays/stores your results for each user profile. Note: Navigate to the Workflows sub-folder to locate other User Profile folders (that may contain more recent results).

Do the following if a study is not listed on the Dashboard:

1. Click **Browse for Existing Analysis Result**.

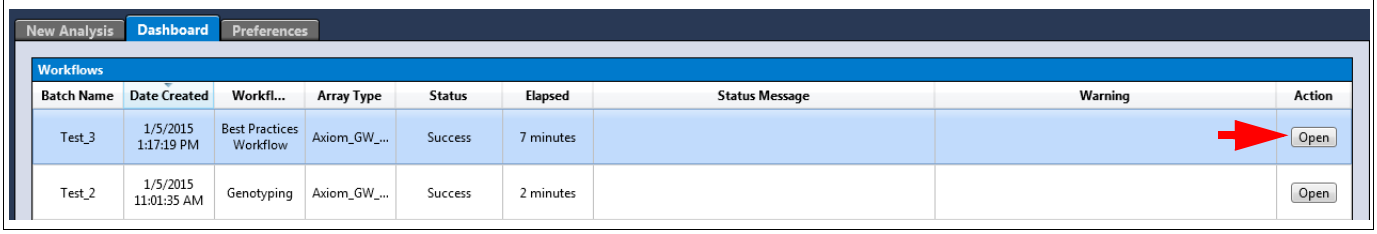
   A Select Analysis Result File window appears.

2. Click on a recent analysis, then click **Select Folder**.

   After a few moments, your analysis opens as you last left it.

3. After your analysis has successfully completed. click Open (Figure 34). Alternatively, click to highlight the completed analysis, then click **Open Selected Result(s)**.

Figure 34 Dashboard window/tab



The Axiom Analysis Suite Viewer appears.

For instructions on how to use the Viewer, continue to Chapter 3.

Browsing for Existing Suitcases

In AxAS v1.0, a suitcase file was auto-generated during an analysis and included all the necessary files needed to view your analysis in the Viewer.

For AxAs v1.1 and newer, the suitcase has been replaced with a higher performing batch name folder. Your suitcase file must be converted to this new folder format before it can be opened in the newly installed Viewer.

Do the following to convert your suitcase file to a batch name folder:

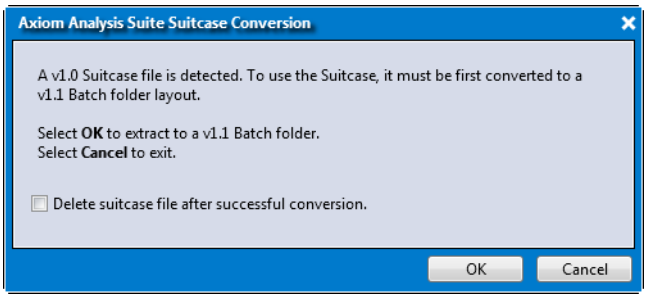
1. Click **Browse for Existing Suitcase**.

   A Select Analysis Result File window appears.

2. Click to highlight a suitcase file, then click **Open**.

   An **Axiom Analysis Suite Suitcase Conversion** message window appears. (Figure 35)

Figure 35 Convert suitcase file to batch name folder message



3. If you want to retain your v1.0 suitcase file for archiving purposes, leave the **Delete suitcase file after successful conversion** check box unchecked. Click on this check box if you want your suitcase file to be auto-deleted after it is converted.
4. Click **OK**.

   Allow a few moments for your suitcase file to convert to the v1.1 batch name folder format.

   The Axiom Analysis Suite Viewer appears.

   For instructions on how to use the Viewer, continue to Chapter 3.

Opening an Analysis from Windows Explorer

You can open an existing analysis directly from Windows Explorer, instead of through the application's Dashboard widow tab. To do this:

1. Right-click on an analysis batch folder and select **Open in Axiom Analysis Suite Viewer**.

   A Select Profile window appears.

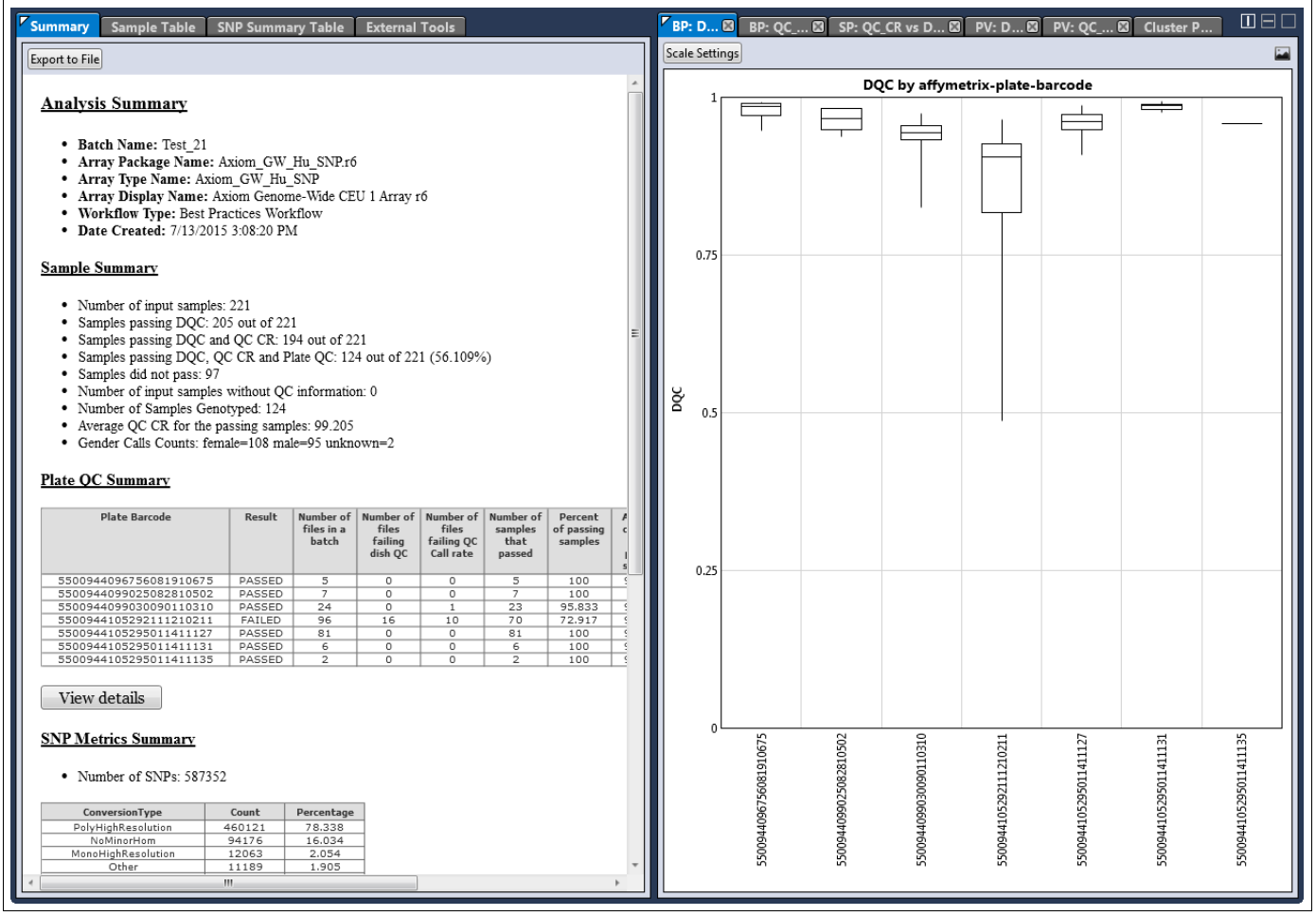
2. At the Select Profile window, enter a new or select an existing profile name.

   If your selected folder is a valid analysis folder, the application's Viewer opens as it normally would.

3 The Viewer: Summary Window and Sample Table

After setting up and successfully running an analysis, as described in Chapter 2, the Axiom Analysis Suite Viewer opens. (Figure 36)

Figure 36 Main Viewer window



Viewing Options

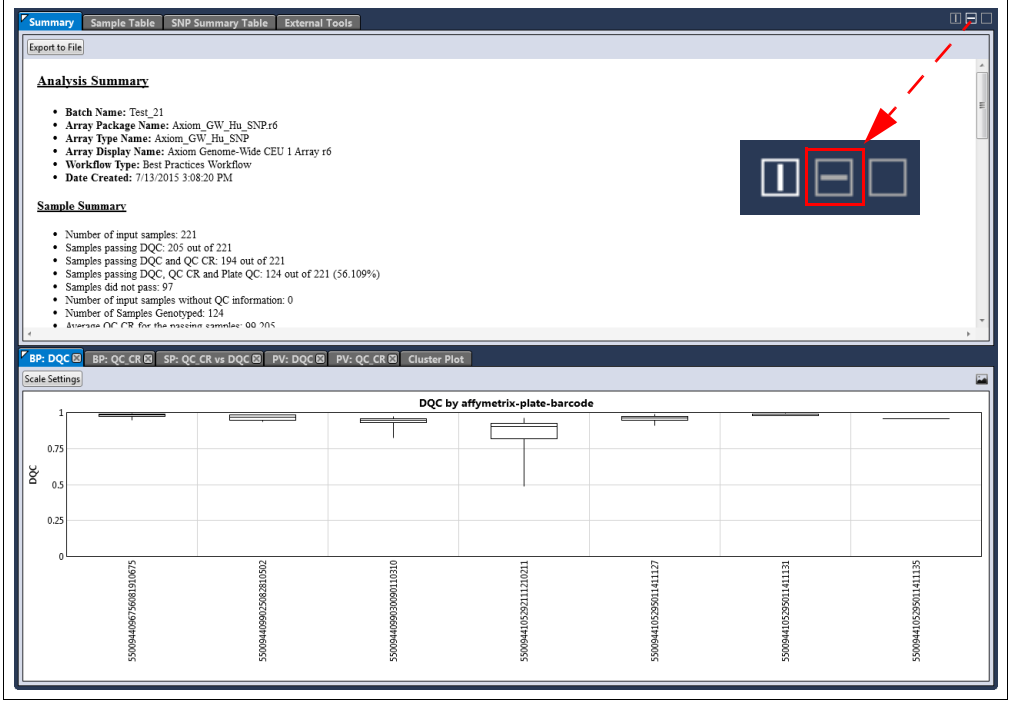
As shown in Figure 36 on page 38, the Viewer (by default) displays a **side-by-side** split-screen configuration.

Split-Screen Options

To change side by side split-screen to a top and bottom configuration:

1. Click the **Horizontal Split** icon. (Figure 37)

Figure 37 Split Horizontal View icon and window layout example



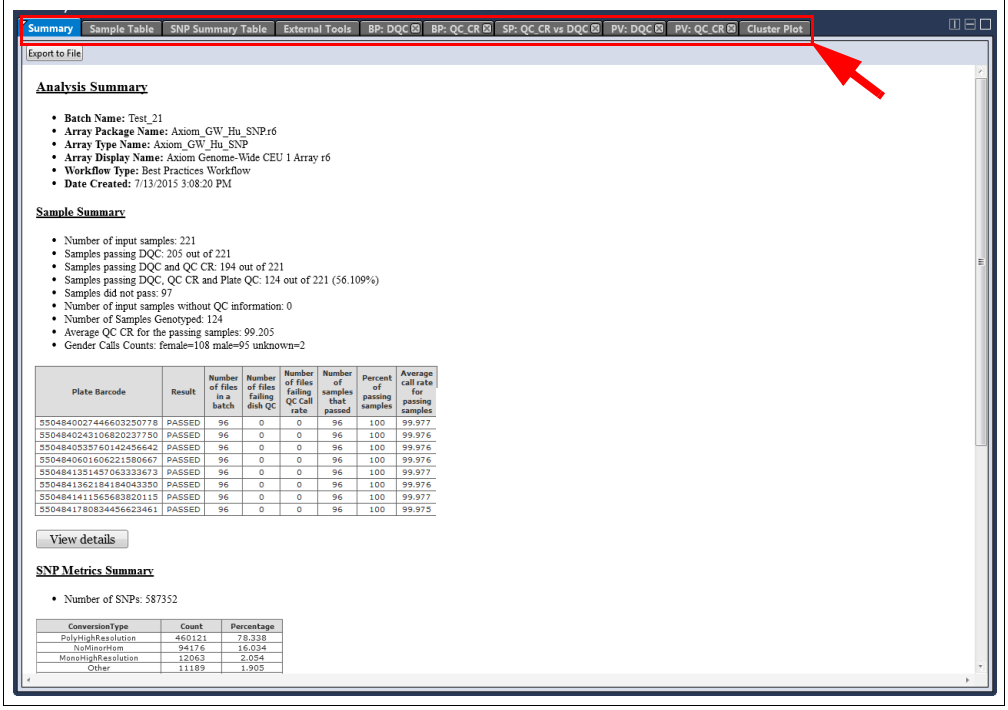
To disable the split-screen:

1. Click the **Disable Split-Screen** icon. (Figure 38)

Figure 38 Disable Split-Screen icon

The split-screen becomes 1 window. (Figure 39)

Figure 39 Full window view example



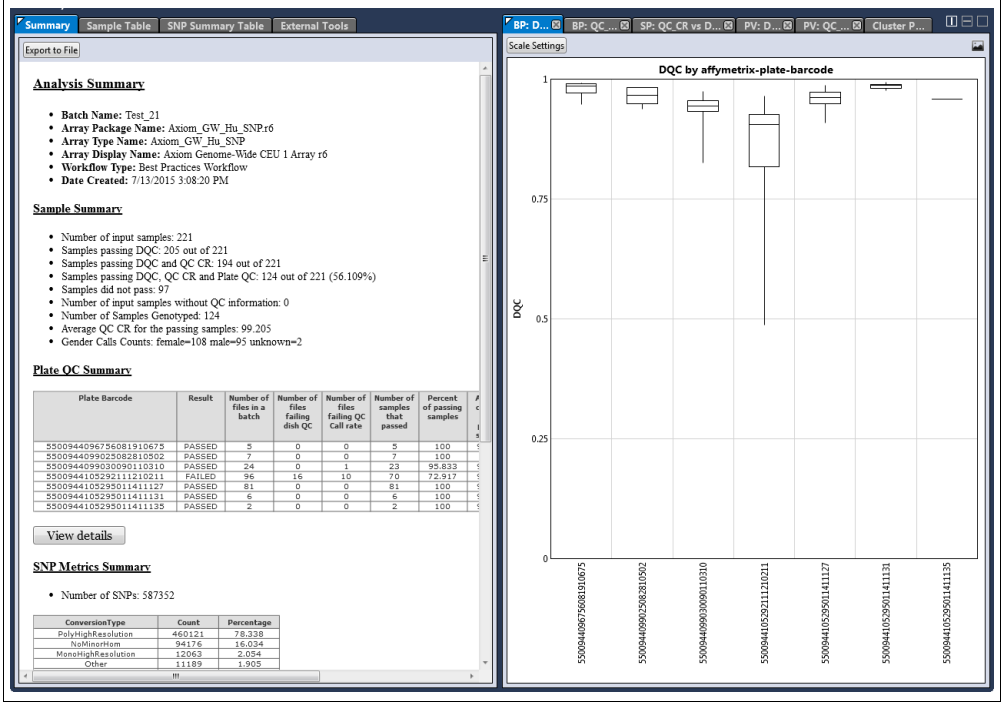
2. Click on any window tab (Figure 39) to view it in full window mode.

To return to the default side by side split-screen configuration:

1. Click the **Vertical Split** icon. (Figure 40)

Figure 40 Vertical Split icon

Figure 41 Split Vertical View example



Changing a Tab Window to a Full Screen Windows

To toggle a tab window to full screen:

1. Locate the tab you want to make full screen.
2. Click on a tab's white triangle graphic. (Figure 42)

Figure 42 Tab to window toggle icon

The window tab is now a window.

3. Double-click anywhere along the top of the window to change it to full screen.

To toggle a full screen window to its default tab window:

1. Double-click anywhere along the top of the window.
2. Click the **X** button (top right) to close the window.

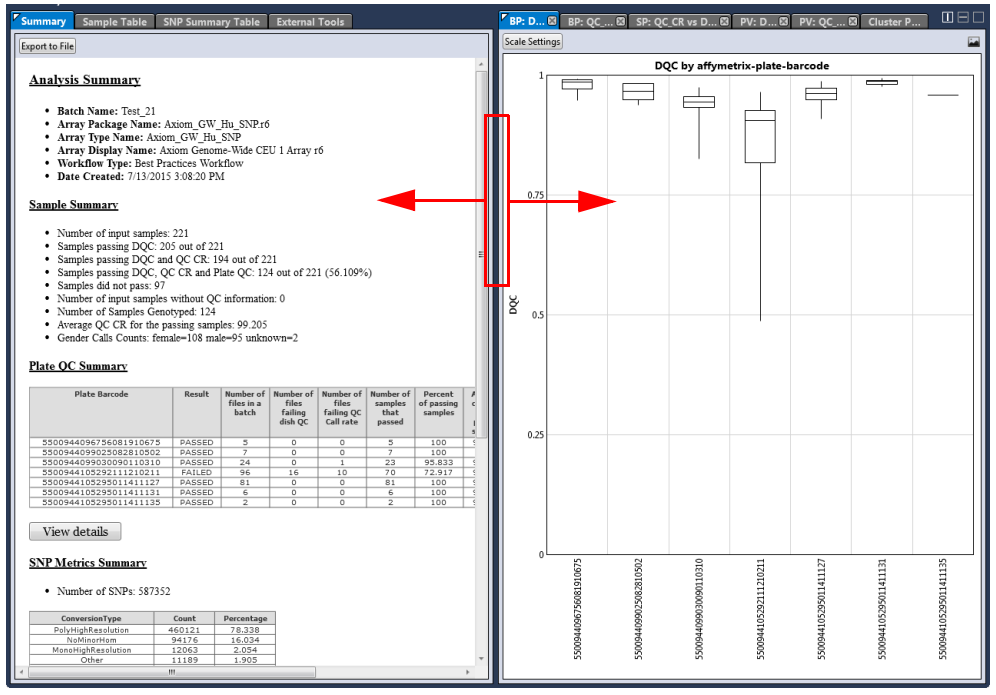
The window returns to its default tab window and position.

Adjusting the Window Size

To change the size of a window pane:

1. Click, hold, then drag the edge of the window pane (Figure 43) to resize it.

Figure 43 Split Vertical View example



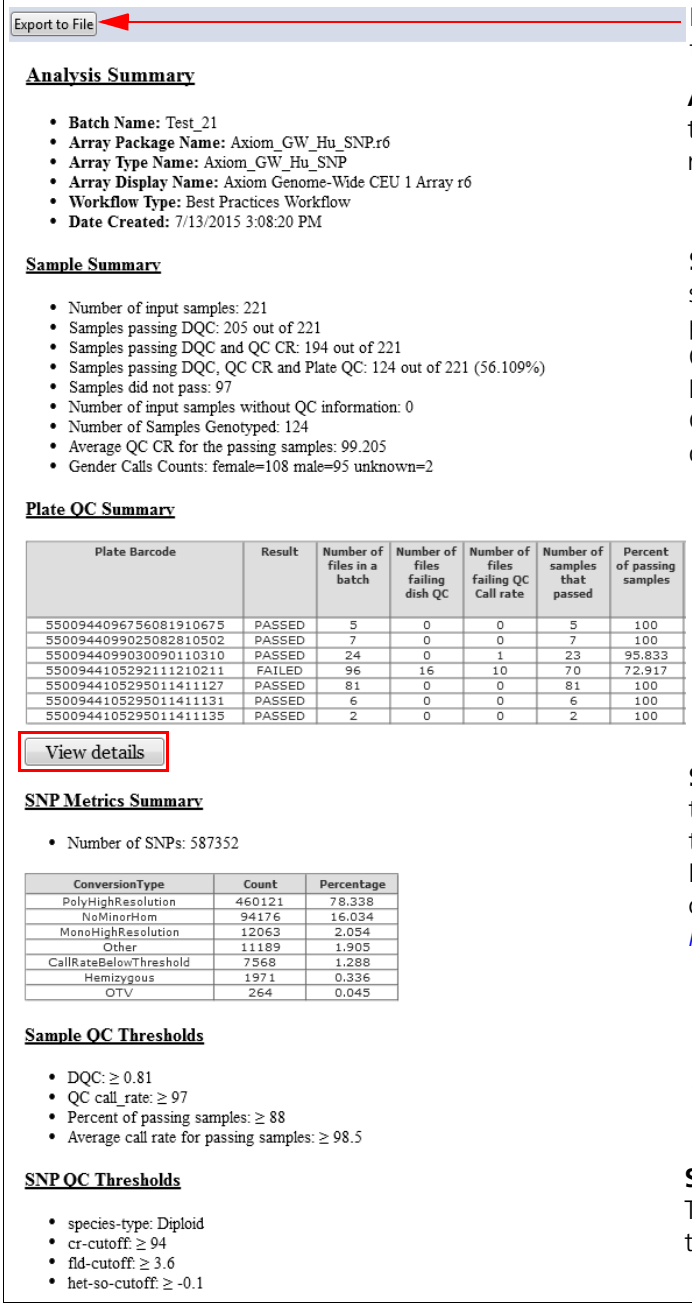
Summary Window/Tab

The Summary window/tab (Figure 44) displays a summary snapshot of your analysis, including detailed threshold values, and tables based on your analysis.

Data Analysis Summary

Note: Each workflow type reports different information within the Analysis Summary window. Figure 44 is an example of a Best Practices workflow.

Figure 44 Summary window tab



Export to File: Click this button to export the Summary report as a PDF file.

Analysis Summary: Contains information about the array type, the workflow run and the date processed

Sample Summary: Breaks down the sample QC for your analysis run and displays the number that pass each of your QC Thresholds.
In addition, it provides the average QC Call Rate (CR) and breakdown of the genders found within your batch of samples.

Plate QC Summary: Contains sample QC information for each plate including the number samples failing DQC, QC Call Rate, the Percent of passing samples. and the average Call Rate for your passing samples.

SNP Metrics Summary: This section contains a summary of the categorization of the SNPs in the analysis by PS_Classification. For more information on these categories *see "Regenerate SNP Metrics" on page 82*.

Sample QC Thresholds: Displays the Sample QC Thresholds used for your analysis run and their associated SNP QC Metrics.

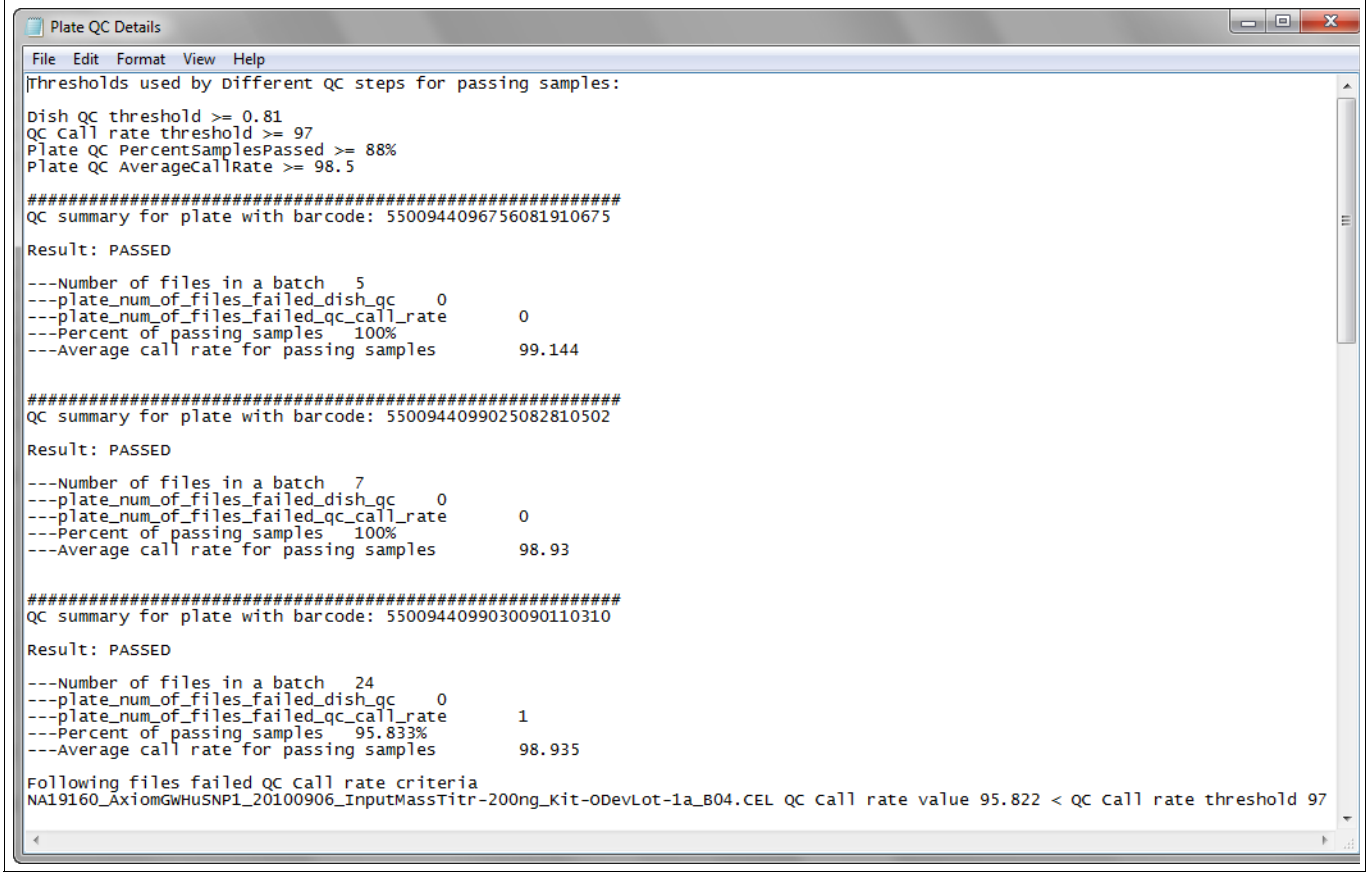
SNP QC Thresholds: Displays the Thresholds used for your analysis run and their associated SNP QC Metrics.

Viewing the Plate Barcode Table Details

1. In the Summary window tab, click View Details. (Figure 44)

   A window opens and displays a text file version of your Sample QC information (by plate). (Figure 45)

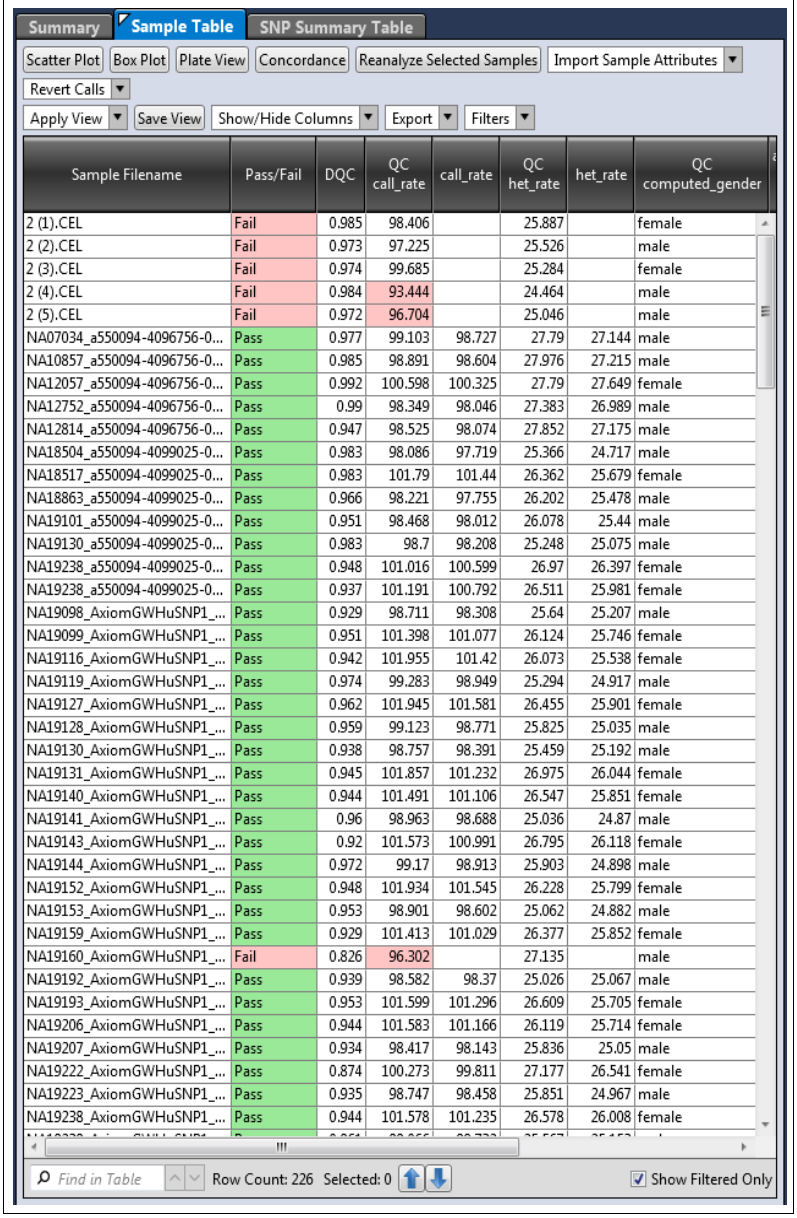
Figure 45 Notepad window



Sample Table

Note: Depending on the Threshold values you set (prior to running your analysis), color-coded Pass or Fail cells may appear in the table, as shown in Figure 46.

Figure 46 Sample Table window tab



Sample Filename	Pass/Fail	DQC	QC call_rate	call_rate	QC het_rate	het_rate	QC computed_gender
2 (1).CEL	Fail	0.985	98.406		25.887		female
2 (2).CEL	Fail	0.973	97.225		25.526		male
2 (3).CEL	Fail	0.974	99.685		25.284		female
2 (4).CEL	Fail	0.984	93.444		24.464		male
2 (5).CEL	Fail	0.972	96.704		25.046		male
NA07034_a550094-4096756-0...	Pass	0.977	99.103	98.727	27.79	27.144	male
NA10857_a550094-4096756-0...	Pass	0.985	98.891	98.604	27.976	27.215	male
NA12057_a550094-4096756-0...	Pass	0.992	100.598	100.325	27.79	27.649	female
NA12752_a550094-4096756-0...	Pass	0.99	98.349	98.046	27.383	26.989	male
NA12814_a550094-4096756-0...	Pass	0.947	98.525	98.074	27.852	27.175	male
NA18504_a550094-4099025-0...	Pass	0.983	98.086	97.719	25.366	24.717	male
NA18517_a550094-4099025-0...	Pass	0.983	101.79	101.44	26.362	25.679	female
NA18863_a550094-4099025-0...	Pass	0.966	98.221	97.755	26.202	25.478	male
NA19101_a550094-4099025-0...	Pass	0.951	98.468	98.012	26.078	25.44	male
NA19130_a550094-4099025-0...	Pass	0.983	98.7	98.208	25.248	25.075	male
NA19238_a550094-4099025-0...	Pass	0.948	101.016	100.599	26.97	26.397	female
NA19238_a550094-4099025-0...	Pass	0.937	101.191	100.792	26.511	25.981	female
NA19098_AxiomGWHuSNP1_...	Pass	0.929	98.711	98.308	25.64	25.207	male
NA19099_AxiomGWHuSNP1_...	Pass	0.951	101.398	101.077	26.124	25.746	female
NA19116_AxiomGWHuSNP1_...	Pass	0.942	101.955	101.42	26.073	25.538	female
NA19119_AxiomGWHuSNP1_...	Pass	0.974	99.283	98.949	25.294	24.917	male
NA19127_AxiomGWHuSNP1_...	Pass	0.962	101.945	101.581	26.455	25.901	female
NA19128_AxiomGWHuSNP1_...	Pass	0.959	99.123	98.771	25.825	25.035	male
NA19130_AxiomGWHuSNP1_...	Pass	0.938	98.757	98.391	25.459	25.192	male
NA19131_AxiomGWHuSNP1_...	Pass	0.945	101.857	101.232	26.975	26.044	female
NA19140_AxiomGWHuSNP1_...	Pass	0.944	101.491	101.106	26.547	25.851	female
NA19141_AxiomGWHuSNP1_...	Pass	0.96	98.963	98.688	25.036	24.87	male
NA19143_AxiomGWHuSNP1_...	Pass	0.92	101.573	100.991	26.795	26.118	female
NA19144_AxiomGWHuSNP1_...	Pass	0.972	99.17	98.913	25.903	24.898	male
NA19152_AxiomGWHuSNP1_...	Pass	0.948	101.934	101.545	26.228	25.799	female
NA19153_AxiomGWHuSNP1_...	Pass	0.953	98.901	98.602	25.062	24.882	male
NA19159_AxiomGWHuSNP1_...	Pass	0.929	101.413	101.029	26.377	25.852	female
NA19160_AxiomGWHuSNP1_...	Fail	0.826	96.302		27.135		male
NA19192_AxiomGWHuSNP1_...	Pass	0.939	98.582	98.37	25.026	25.067	male
NA19193_AxiomGWHuSNP1_...	Pass	0.953	101.599	101.296	26.609	25.705	female
NA19206_AxiomGWHuSNP1_...	Pass	0.944	101.583	101.166	26.119	25.714	female
NA19207_AxiomGWHuSNP1_...	Pass	0.934	98.417	98.143	25.836	25.05	male
NA19222_AxiomGWHuSNP1_...	Pass	0.874	100.273	99.811	27.177	26.541	female
NA19223_AxiomGWHuSNP1_...	Pass	0.935	98.747	98.458	25.851	24.967	male
NA19238_AxiomGWHuSNP1_...	Pass	0.944	101.578	101.235	26.578	26.008	female

Importing Sample Attributes

To import sample attributes into your Sample Table:

1. Click the **Import Sample Attributes** drop-down.
2. Click to select either **Import from ARR Files** or **Import from CSV/Tab-Delimited Text File**.

   An Explorer window appears.

IMPORTANT! Your text-based CEL file must start with the header *Sample Filename* and include the full CEL file name, as shown in Figure 47.

Figure 47 Tab-delimited text CEL file example shown in Excel

	A	B	C
1	Sample Filename	String Name	Number Name
2	NA18961_AxiomGWHuSNP1_20101007_TrainingMix_Rtstab-DL1a-Omo_E09.CEL	A1	1
3	NA12760_AxiomGWHuSNP1_20101007_TrainingMix_Rtstab-DL1a-Omo_B06.CEL	A2	1
4	NA19160_AxiomGWHuSNP1_20100906_InputMassTitr-200ng_Kit-ODevLot-1a_B04.CEL	A3	1
5	NA18960_AxiomGWHuSNP1_20101007_TrainingMix_Rtstab-DL1a-Omo_E08.CEL	A4	1
6	NA12814_AxiomGWHuSNP1_20101007_TrainingMix_Rtstab-DL1a-Omo_B09.CEL	A5	1
7	NA19210_AxiomGWHuSNP1_20101007_TrainingMix_Rtstab-DL1a-Omo_F12.CEL	A6	1

3. Navigate to the applicable file location, then click **Open**.

Column Headers

The default Sample Table column view is as shown. (Figure 48)

Figure 48 Default Sample Table Columns

To show or hide table columns:

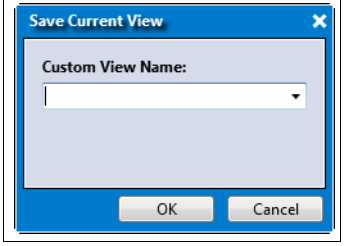
1. Click the **Show/Hide Columns** drop-down menu.
2. Click each available column name's check box to show it or remove it from the table. See "Annotations and Columns" on page 159 for their definitions.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

To save your customized Sample Table column view:

1. Click **Save View.**

   The Save Current View window appears. (Figure 49)

Figure 49 Save Current View



2. Enter a name for your custom table view, then click **OK**.
   Your newly saved name is now added to the **Apply View** drop-down menu.

To show ALL available columns within the Sample Table:

1. Click the **Apply View** drop-down menu, then select **All Columns View**.

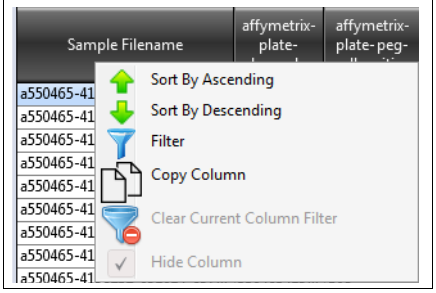
Rearranging Columns

1. Click on a column you want to move.
2. Drag it (left or right) to its new location.
3. Release the mouse button.
   The column is now in its new position.

Sorting Columns

1. Select a column, then right-click on it.
   A right-click menu appears. (Figure 50)

Figure 50 Right-click Column Menu



2. Click to select either **Sort By Ascending** (A-Z) or **Sort By Descending** (Z-A).

Single-Click Sorting Method

1. Single-click on a column header to sort its data in an ascending order. Single-click on the same column header to sort its data in a descending order

Hiding the Column

1. Select the column you want to hide from the table, then right-click on it.
   A right-click menu appears. (Figure 50)
2. Click the **Hide Column** check box to remove it from the table.

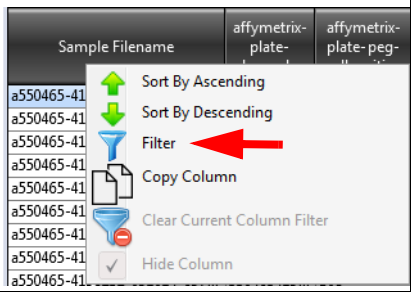
Filtering Column Data

Note: All Sample Table columns are filterable.

Adding Filters (Method 1)

1. Select a column, then right-click on it.
   The following window appears: (Figure 51)

Figure 51 Right-click Column Menu

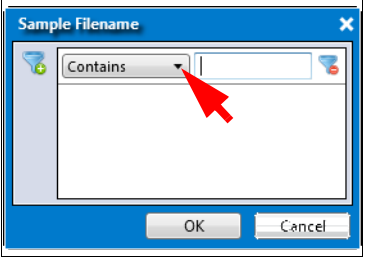


2. Click **Filter**.

Text-based Columns

If the column you want to filter contains text-based data, the Contains drop-down menu appears. (Figure 52)

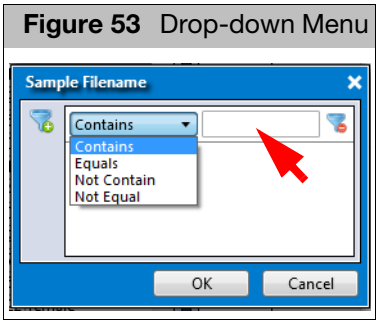
Figure 52 Filter Properties



To apply a filter to a text-based column:

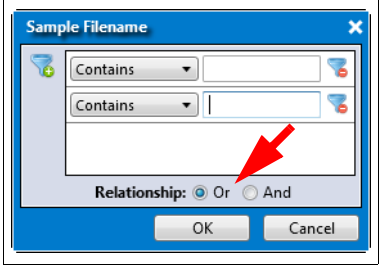
1. Click the **Contains** drop-down menu to select a filtering property. (Figure 53)

Figure 53 Drop-down Menu



2. Click inside the text entry box [ ] to enter a value. (Figure 53)
3. OPTIONAL: Click to add additional filters.

Figure 54 Or or And Relationship Logic

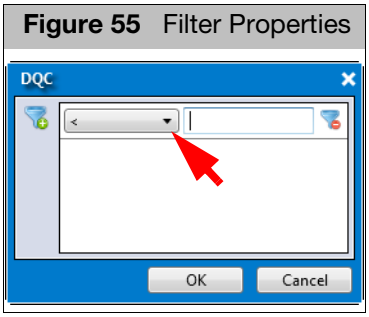


4. Click the **Or** or **And** radio button to choose **Or** or **AND** relationship logic. (Figure 54)
5. Repeat steps 1-4 as needed.
6. To remove a filter(s), click .

Numeric Data Columns

If the column you want to filter contains numeric data, a symbol drop-down menu appears. (Figure 55)

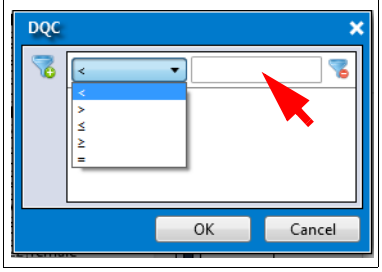
Figure 55 Filter Properties



To apply a filter to a value-based column:

1. Click the **Symbol Value** drop-down menu to select the filtering symbol you want. (Figure 56)

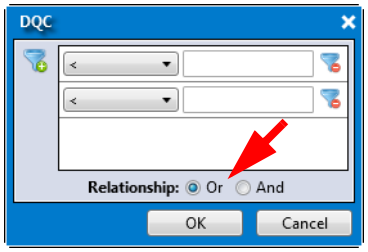
Figure 56 Drop-down Menu



2. Click inside the text entry box [ ] to enter the value(s). (Figure 56)

3. OPTIONAL: Click to add filter(s).

Figure 57 Or or And Relationship Logic



4. Click the **Or** or **And** radio button to choose **Or** or **AND** relationship logic. (Figure 57)
5. If needed, repeat steps 1-4.
6. Click **OK**.

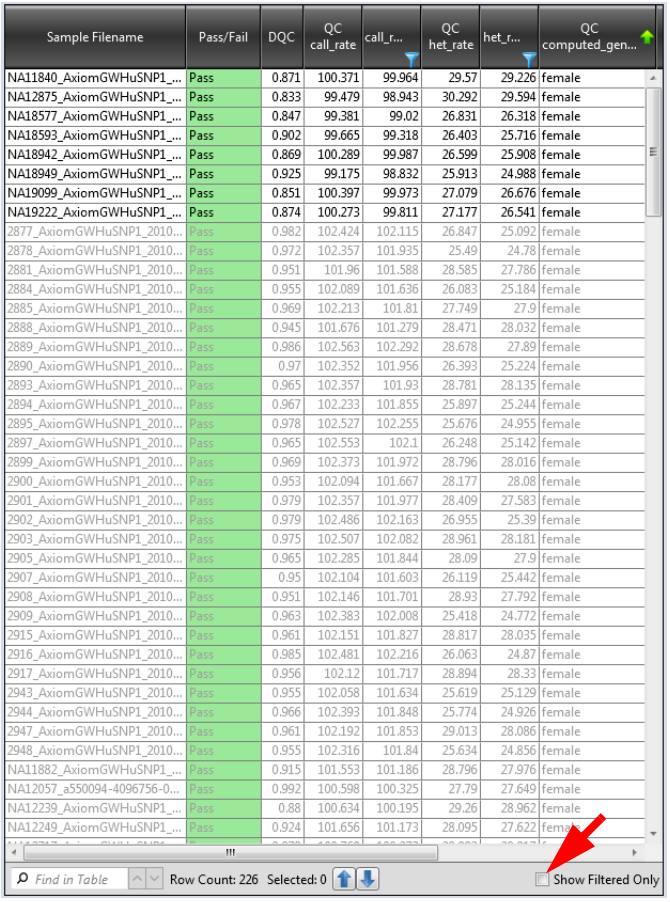
   To remove a filter(s), click .

Showing Filtered Data Only

- Click the **Show Filtered Only** check box Show Filtered Only to show only the data that passes the filters.

  Uncheck this box to show all data, including data that did not pass your filter criteria setting(s). In this mode, data that passes the filter appears in light gray, as shown in Figure 58 on page 51.

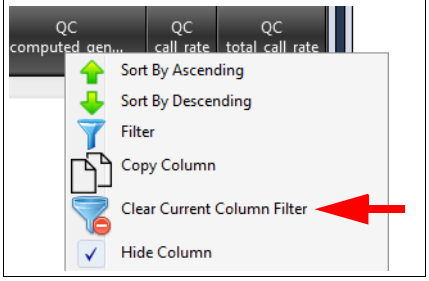
Figure 58 Sample Table window tab - Show Filter Only unchecked example



Clearing an Individual Filter

1. Right-click on the filtered column you want to clear.
   The following window appears: (Figure 59)

Figure 59 Right-click Column Menu

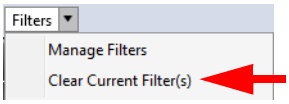


2. Click **Clear Current Column Filter**.
   The filter is removed.

Clearing All Current Filters

- Click the Filters drop-down, then select **Clear Current Filters**. (Figure 60)

Figure 60 Filters Menu



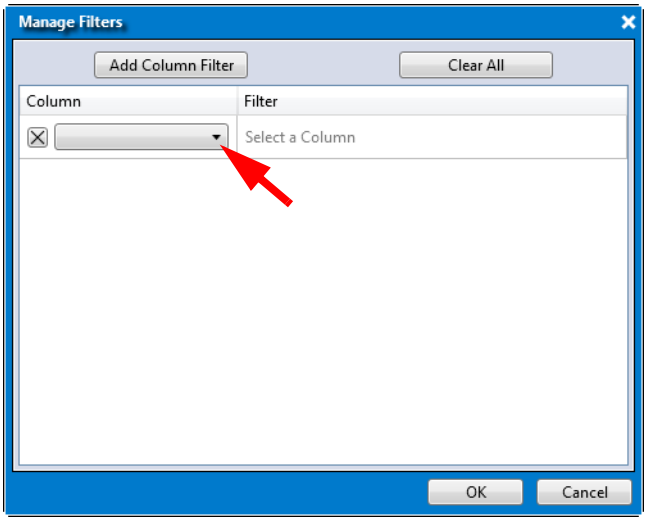
Adding Filters (Method 2)

Use this method if you want to change more than one of your Sample Table column filters at the same time.

1. Click the **Filters** drop-down menu, then click **Manage Filters**.

   The Manage Filters window appears. (Figure 61)

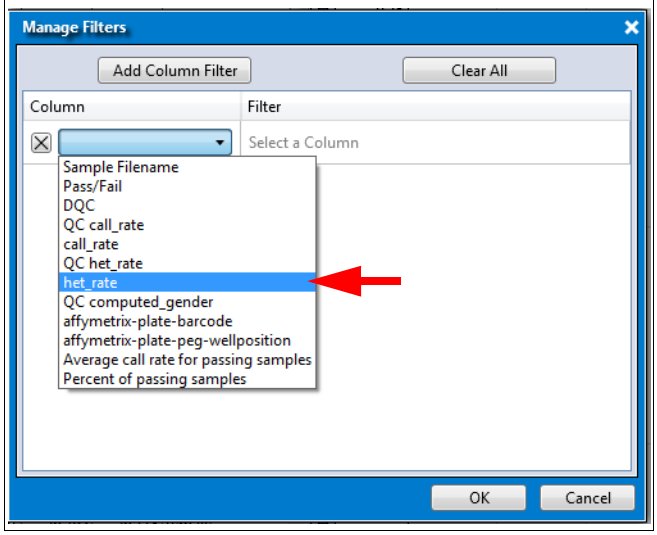
Figure 61 Manage Filters window



Note: If the column you want to filter contains text-based data, the Contains drop-down menu appears. If the column you want to filter contains numeric data, a symbol drop-down menu appears.

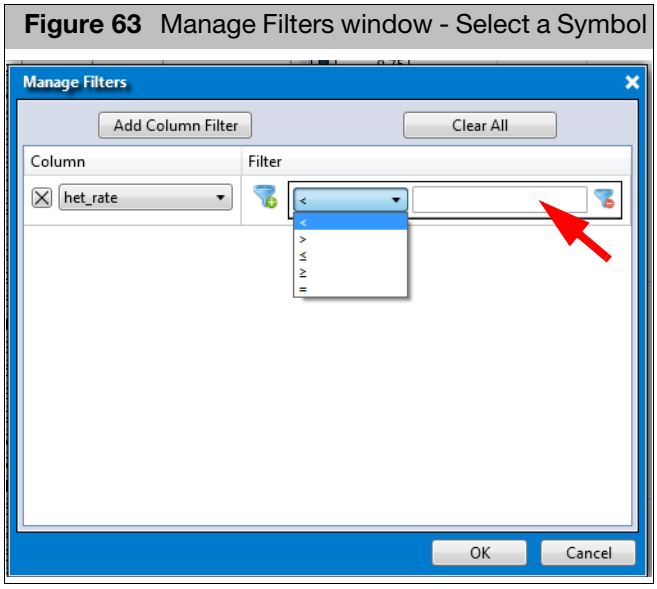
2. Click the **Column** drop-down, then click to select the Column name you want to create a filter for. (Figure 62)

Figure 62 Manage Filters window - Select a Column



3. Click the **Symbol Value** drop-down menu to select the filtering symbol you want. (Figure 63)

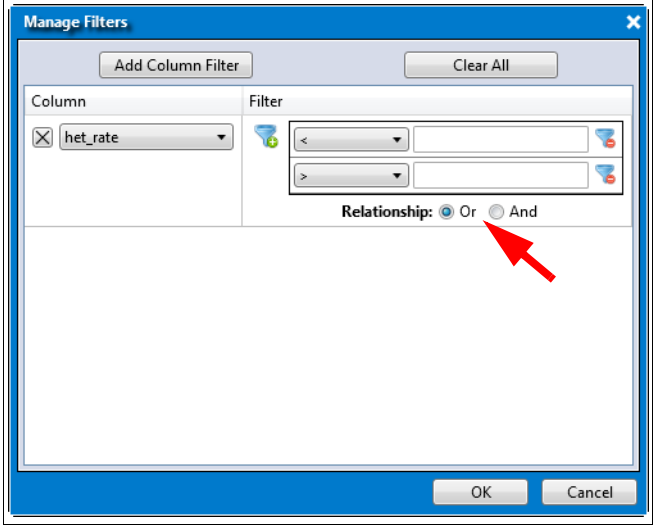
Figure 63 Manage Filters window - Select a Symbol



4. Click inside the text entry box [ ] to enter new value(s). (Figure 63)
5. OPTIONAL: If you want to add an additional filter to a column, click .

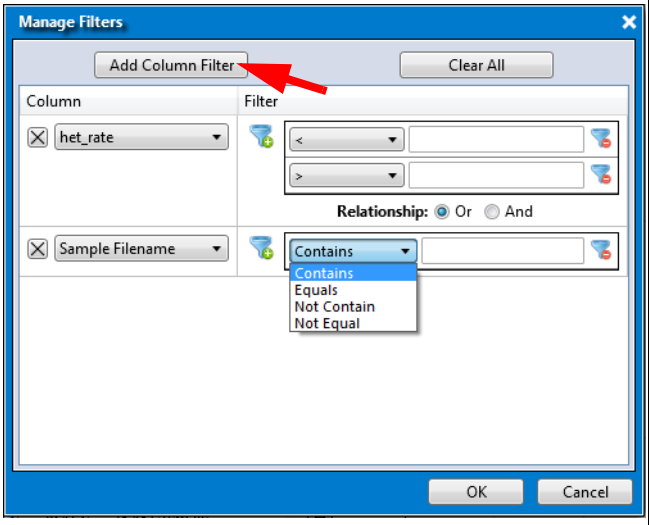
6. Click the **Or** or **And** radio button to choose **Or** or **AND** relationship logic. (Figure 64)

Figure 64 Manage Filters window - OR or AND Relationship



7. If needed, click **Add Column Filter**, then repeat the above steps. (Figure 65)

Figure 65 Manage Filters window - Adding another Column Filter



8. Click **OK**.

   To remove a filter(s), click .

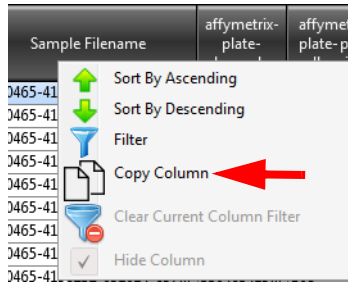
   Click **Clear All** to remove ALL filters in the Manage Filters window.

Copying Column Data

To copy column data to your clipboard:

1. Click to select a column you want to copy to a clipboard, then right-click on it.
   The following window appears: (Figure 66)

Figure 66 Right-click Column Menu



2. Click **Copy Column**.
   The column data is now ready for pasting (Ctrl v).

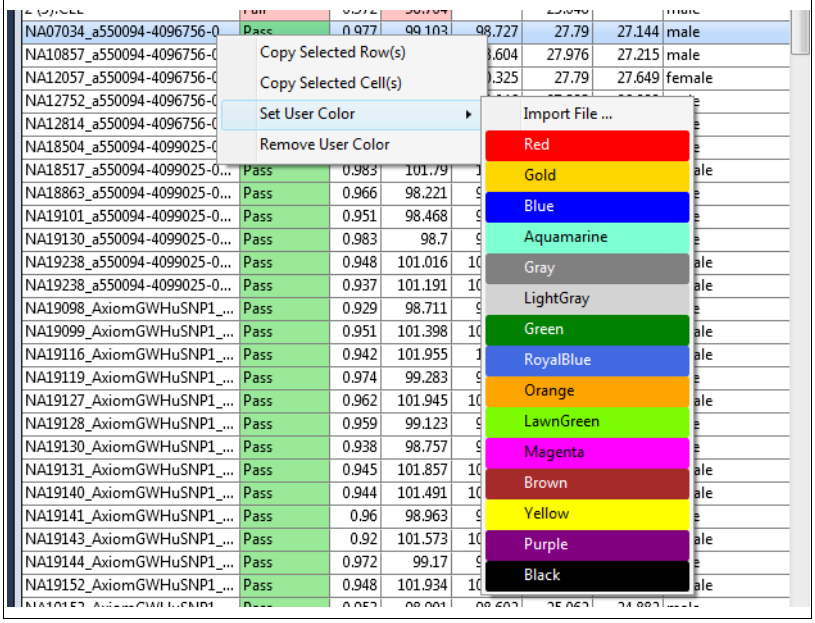
Setting User Colors

Use this feature to more easily identify different sets between the Sample Table and Cluster Graph.

Assigning a Color to a Sample

1. Right-click on the sample you want to assign a color to.
   A menu appears. (Figure 67)

Figure 67 Right-click menu - Set User Color

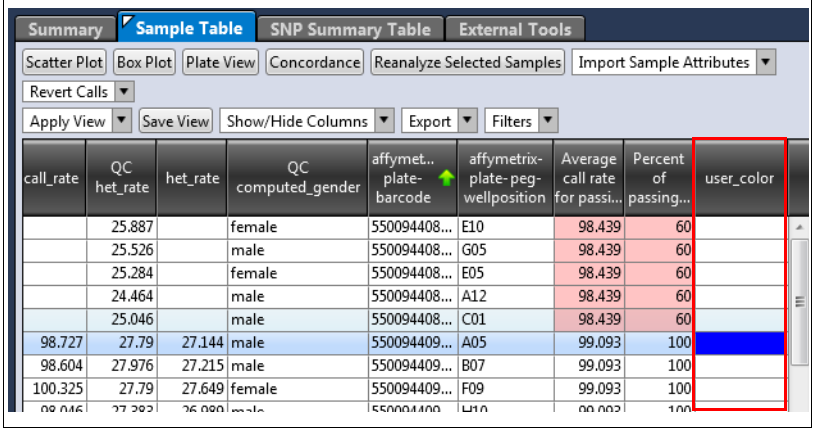


2. Mouse over **Set User Color**.
   A color pallet appears.

3. Click on the color you want.

   A **user_color** column is automatically added to your Sample Table and contains your sample's newly assigned color, as shown in Figure 68.

Figure 68 Right-click menu - Set User Color

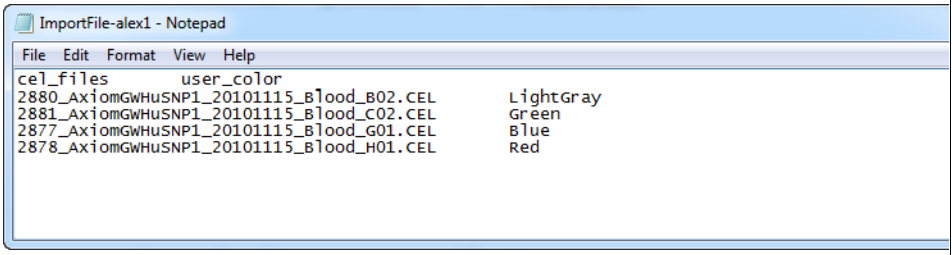


Importing Assigned Colors

Use this feature if you want to assign colors to a large number of samples or if your Sample Table contains a vast amount of samples and you want to assign a color to only a few samples.

1. Use MS Excel or MS Notepad (as you normally would) to create a two column table. (Figure 69)

Figure 69 Two column example in MS Notepad

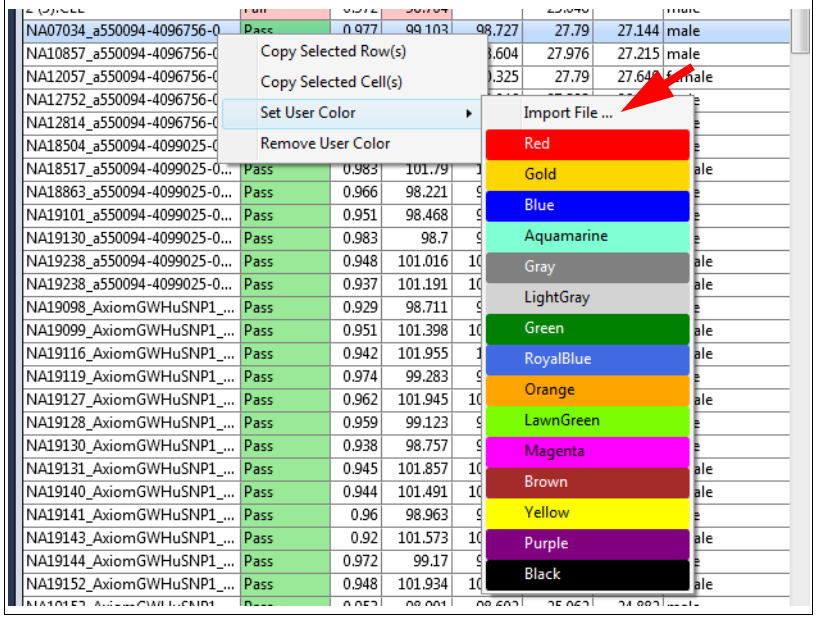


IMPORTANT! Your **user_color** entries must match the color pallet naming conventions shown in Figure 70. Example: **RoyalBlue** not **Royal Blue**.

2. Save your two column table as a tab-delimited text file to an easily accessible location.
3. At the Sample Table, right-click on any sample.

   A menu appears. (Figure 70)

Figure 70 Right-click menu - Set User Color - Import File



4. Mouse over **Set User Color**.
5. Click on **Import File...**

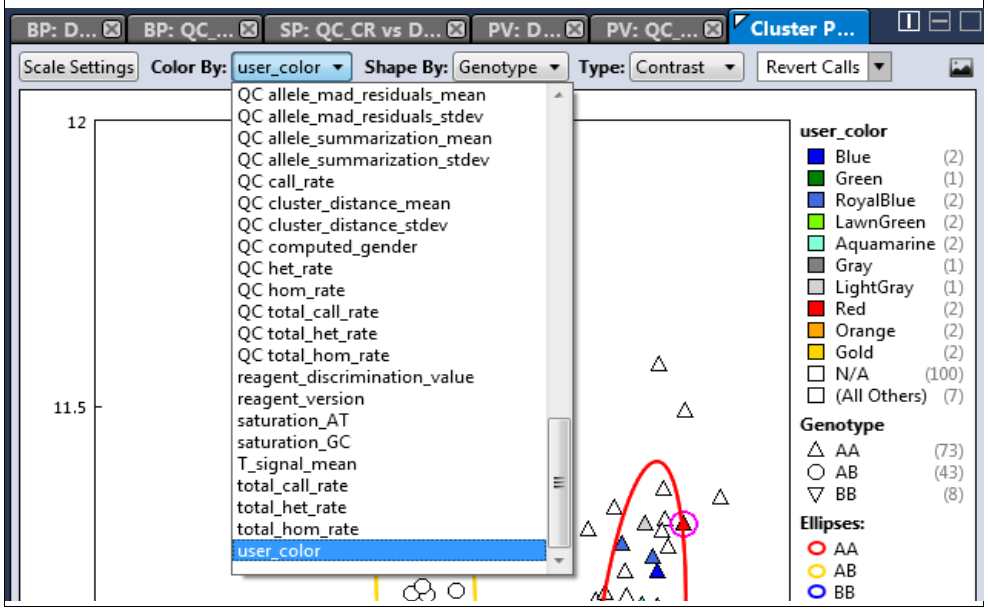
   An **Import User Colors** Explorer window appears.
6. Locate your saved TXT file, click to highlight it, then click **Open**.

   Your TXT file is now incorporated into the Sample Table.
7. Scroll the Sample Table right to see the added **user_color** column and assigned sample colors.

Viewing User Colors in the Cluster Graph

1. From the Cluster Graph, click the **Color By** drop-down menu. (Figure 71)

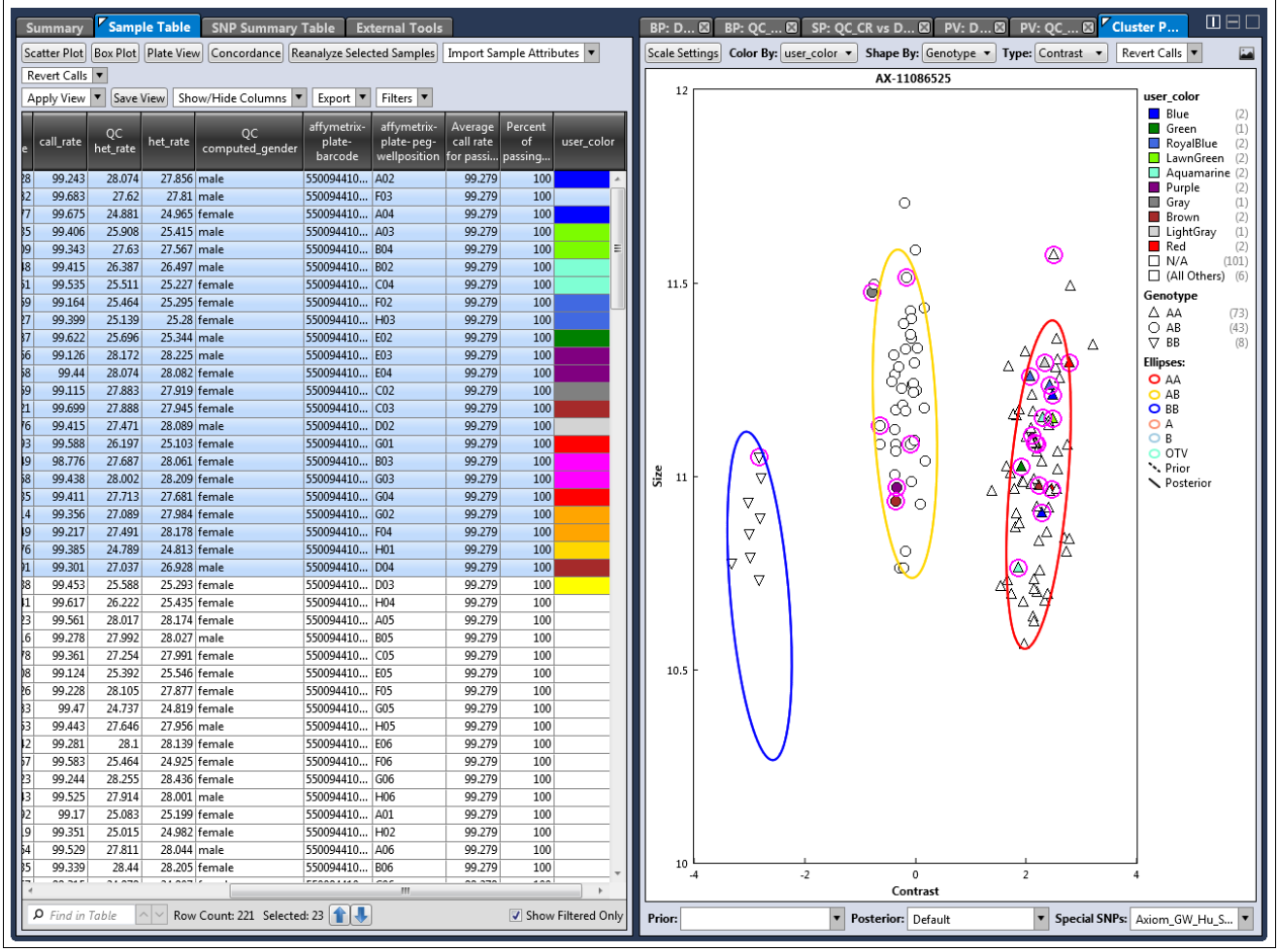
Figure 71 Color By menu - user_color



2. From the Sample Table, single-click on a color-coded sample file or Ctrl click, Shift click, or press Ctrl A (to select multiple color-coded sample files).

Each highlighted sample and its assigned color are now displayed in the Cluster Graph, as show in Figure 72.

Figure 72 user_color shown in Sample Table and Cluster Graph



Removing an Assigned User Color

1. From the Sample Table, right-click on the sample containing the color you want to remove.

   A menu appears.

2. Click **Remove User Color**.

   Your previously assigned sample color is now removed.

Searching Keywords

Note: The Find in Table tool can locate exact (case insensitive) matches. It also accepts wild-card (*) characters to aid in your search. Example: ABC*

1. Click inside the 

text field (bottom left corner of table).
2. Enter a keyword or number.
3. Click the **Up** or **Down** button.

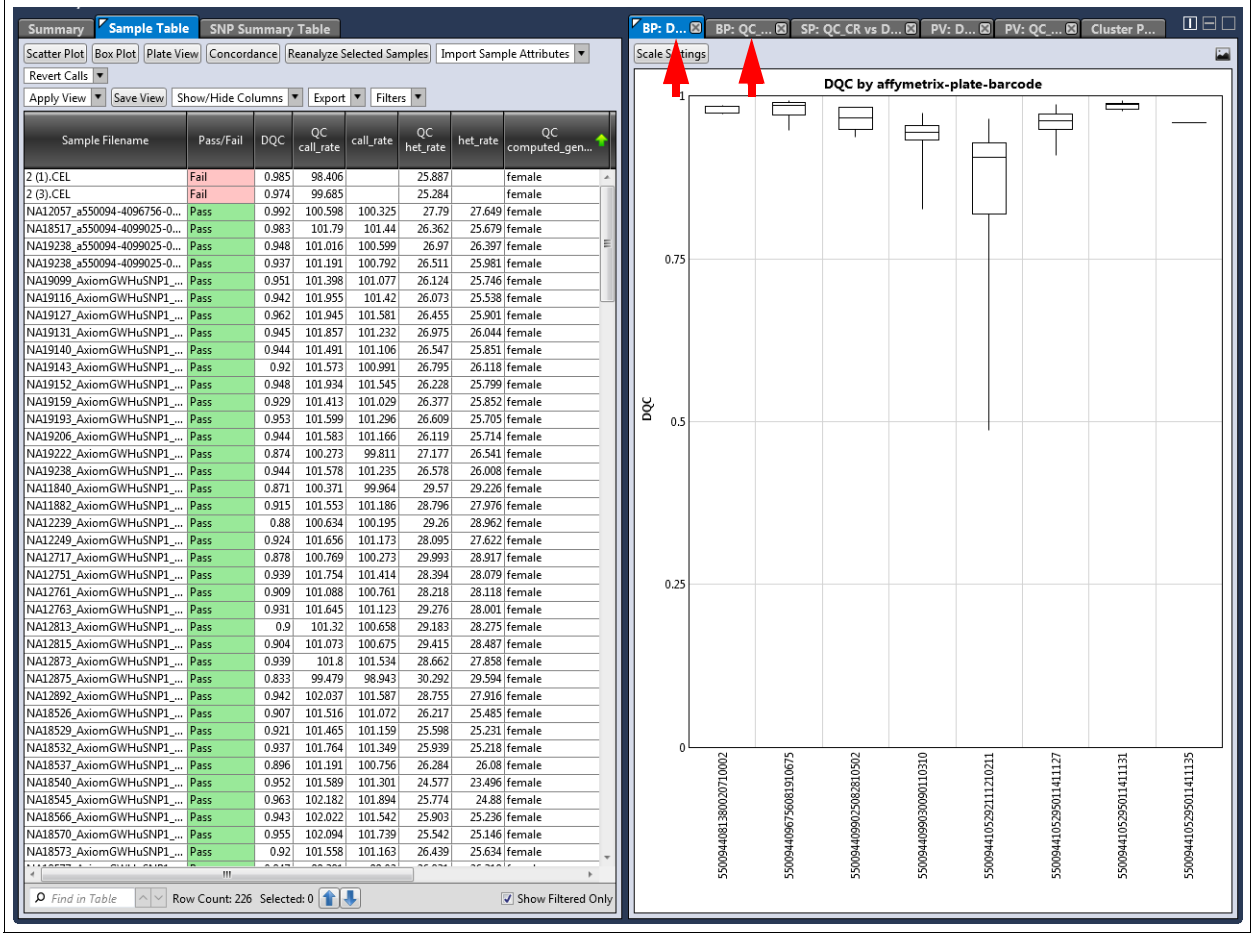
   When a match is found, the appropriate table entry is highlighted. If a graph is displayed, the appropriate graph point is also highlighted.

Box Plots

Note: By default, the Viewer generates two Box Plots. (Figure 73)

Viewing the Default Box Plots

Figure 73 Table and Box Plot 1



IMPORTANT! You cannot change a plot's axis values after it has been created. However, you can change its scale and coloring properties. See "Changing the Box Plot's Scale Setting Ranges" on page 62.

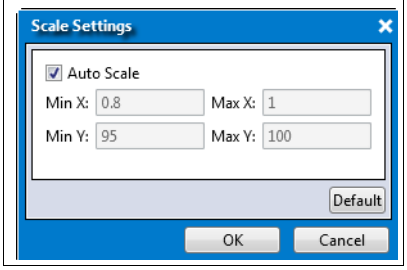
To change a Box Plot's axis properties, you must create a new Box Plot. See "Adding a New Box Plot" on page 62.

Changing the Box Plot's Scale Setting Ranges

1. Click Scale Settings.

   The following window appears. (Figure 74)

Figure 74 Scale Settings window



   By default, the window displays your current range values.

2. Uncheck the **Auto Scale** check box to enter different ranges in the provided text fields.
3. Click **OK**.

   Your new settings are now reflected within the Box Plot. Modified Set Scale values are auto-saved.

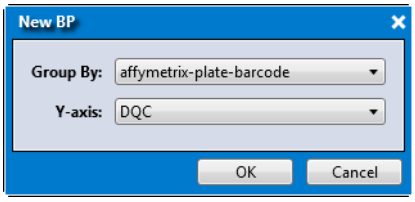
   If needed, click Default to return all values back to their factory settings.

Adding a New Box Plot

1. Click the **Box Plot** button.

   The following window appears: (Figure 75)

Figure 75 Box Plot Versus menus



2. Click the **Group By** drop-down menu to select the X-axis for your new Box Plot.

   Your X-axis selection determines your new Box Plot's boxes and whiskers, based on the data group of values that are compiled.
3. Click the **Y-axis** drop-down menu to select the Y-axis you want.

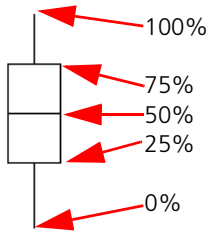
   For Group By and Y-axis definitions, see "Sample Table" on page 154.
4. Click **OK**.

   A new Box Plot window tab is created.

Reading Box Plot Percentiles

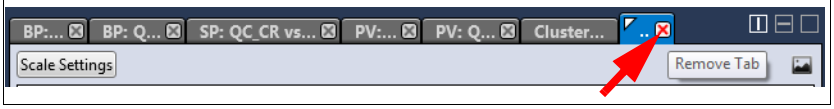
See Figure 76.

Figure 76 Box Plot percentiles



At any time, click **X** to remove a window/tab, as shown in Figure 77.

Figure 77 New Window/Tab



Saving the Current Box Plot View

1. Click the **Save Image** button.
   An Explorer window appears.
2. Navigate to where you want to save the .PNG file, enter a filename, then click **Save**.

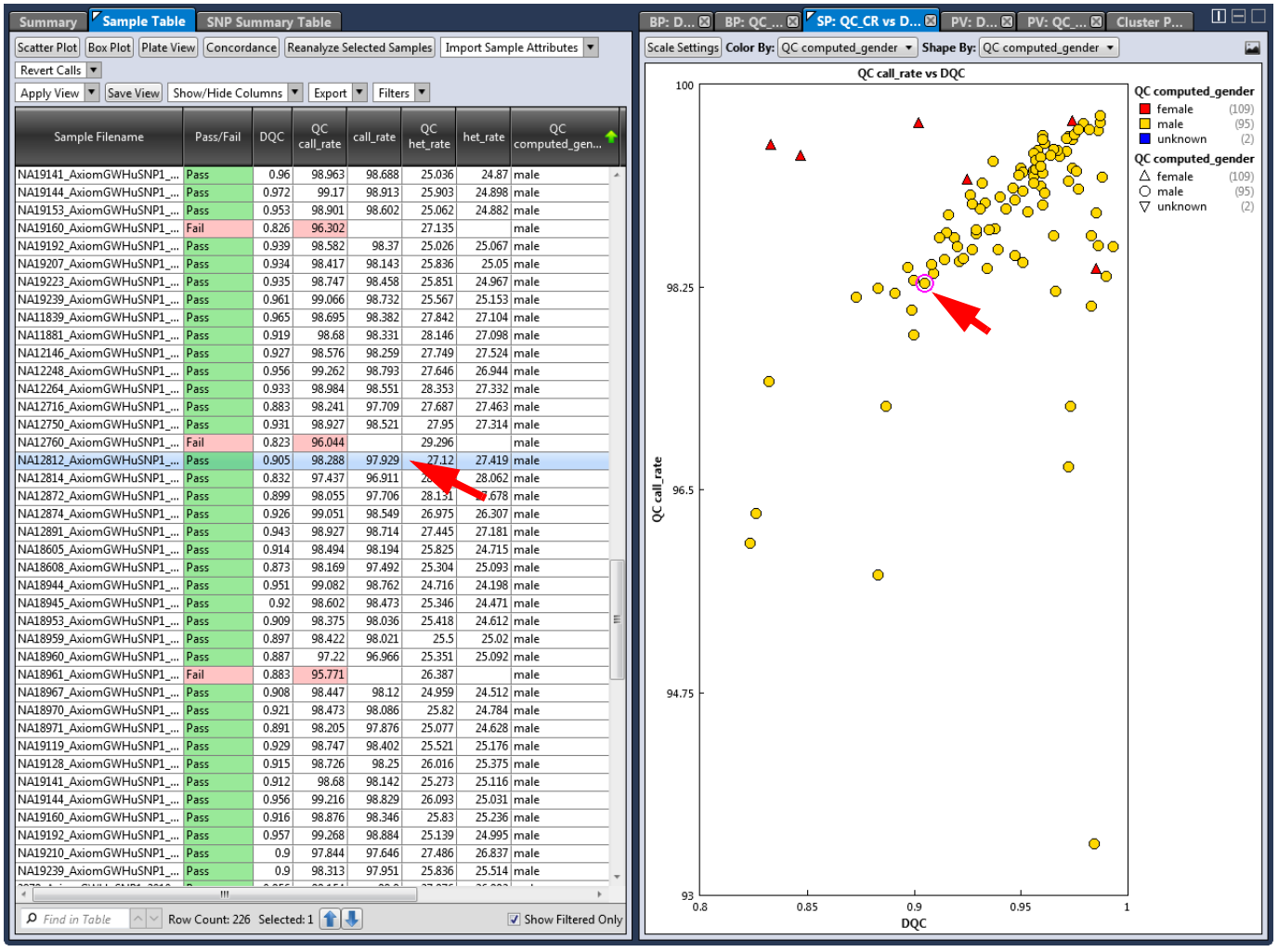
Scatter Plot

By default, the Viewer generates 1 Scatter Plot of **QC call_rate** vs. **DQC**. The data displayed in the plot are colored and shaped by **QC computed_gender**, as shown in Figure 78.

Viewing the Default Scatter Plot

1. Click to highlight a table entry to view its location within the Scatter Plot or click on a data point to highlight its corresponding entry in the Sample Table. (Figure 78)

Figure 78 Table and Scatter Plot



IMPORTANT! You cannot change the default Scatter Plot's pre-defined X and Y definitions, however you can change its Scale Settings and Color By and Shape By configuration.

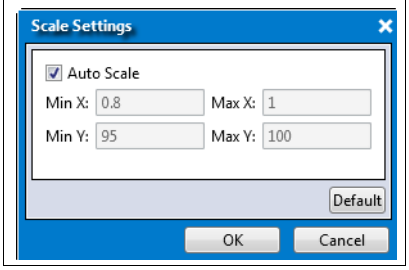
To change a Scatter Plot's axis properties, you must create a new Scatter Plot. See "Adding a New Scatter Plot and Selecting its X and Y Properties" on page 65.

Changing the Scatter Plot's Setting Ranges

1. Click Scale Settings.

   The following window appears. (Figure 79)

Figure 79 Scale Settings window



   By default, the window displays your current range values.

2. Enter your new ranges in the appropriate text fields.
3. Click **OK**.

   Your new settings are now reflected within the Scatter Plot. Modified Set scale values are auto-saved.

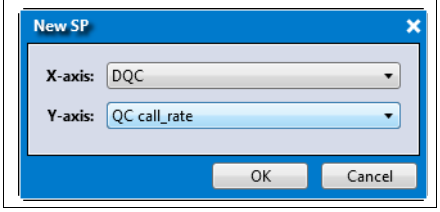
   If needed, click Default to return all values back to their factory settings.

Adding a New Scatter Plot and Selecting its X and Y Properties

1. Click the **Scatter Plot** button.

   The following window appears: (Figure 80)

Figure 80 Scatter Plot Versus menus

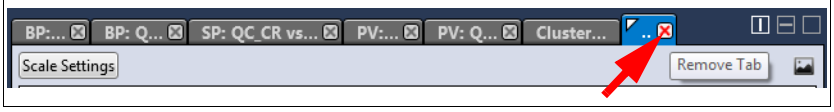


2. Use the drop-down menus to select your Plot's versus scenario (X and Y axis). For definitions, see "Sample Table" on page 154.
3. Click **OK**.

   A new Scatter Plot window tab is created.

   At any time, click **X** to remove a window/tab, as shown in Figure 81.

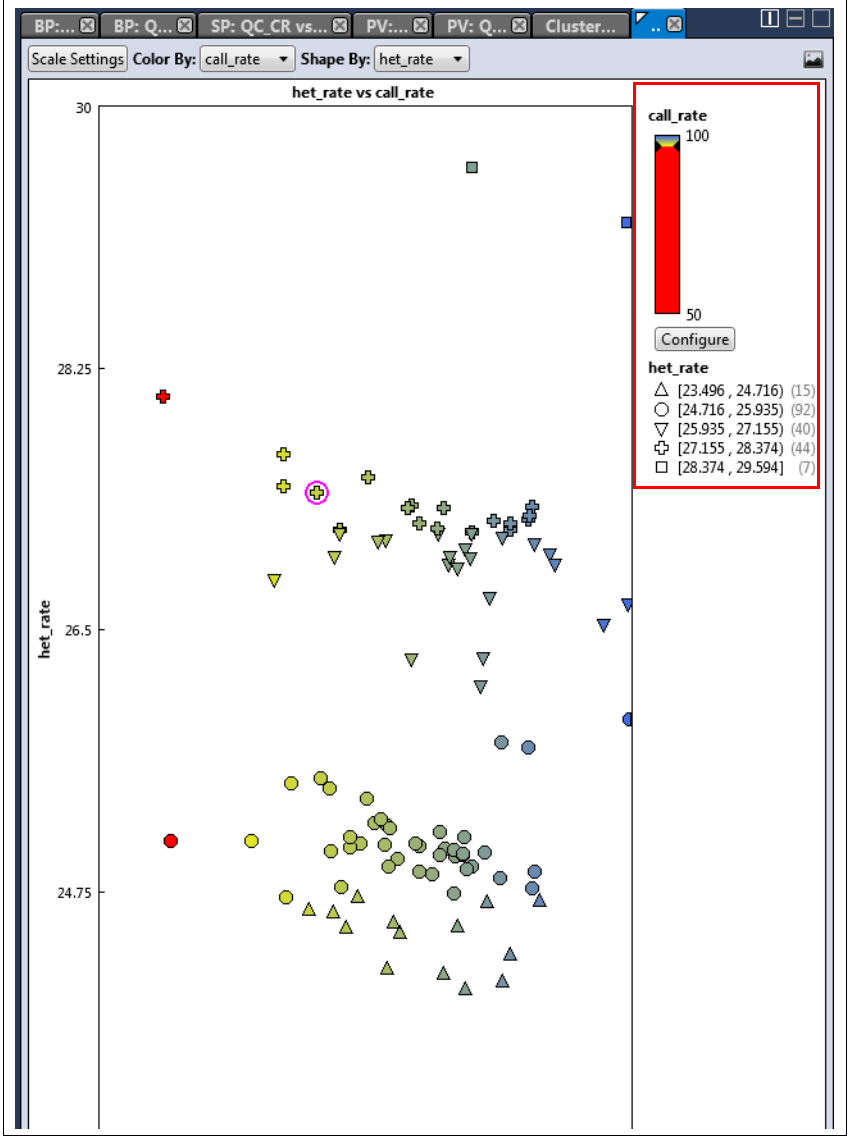
Figure 81 New Window/Tab



4. Click the **Color By** and **Shape By** drop-down menus to select the combination view you want. See "Sample Table" on page 154 for Color By and Shape By definitions. Note: Your imported sample attributes are also available for use within the Color By and Shape By drop-down menus. For information on importing sample attributes, see "Importing Sample Attributes" on page 46.

   A legend appears within the plot. (Figure 82)

Figure 82 Color By and Shape By Legend example



The graph can display up to 10 different colors and up to 10 different shapes. If the attributes selected for display have more than 10 categories, categories 1 through 9 are displayed normally, but categories 10 and higher get grouped together.

If your study has more than 10 values:

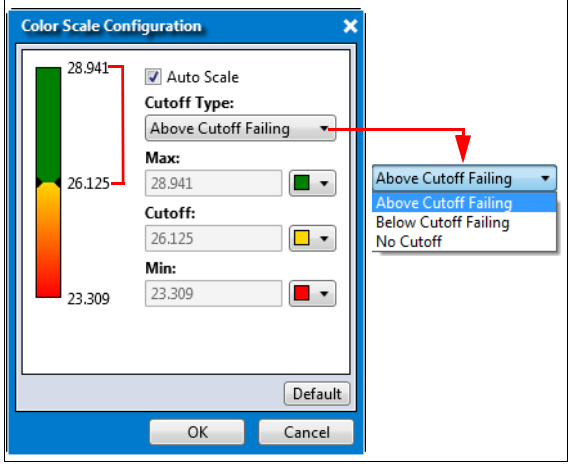
- If the value is text, the software takes the first nine values and assigns each a color or shape. The remaining values are put into a bin labeled **Other**. All values in the Other bin have the same color or shape.
- If the value is a date or number, the software divides the range of data into 10 equal bins and assigns a color or shape to each bin. If the data includes one or more outliers, it is possible to have one value in a particular bin and all other values in another bin.

Customizing Color By Settings

1. Click Configure.

   The Color Scale Configuration window appears. (Figure 83)

Figure 83 Color By options

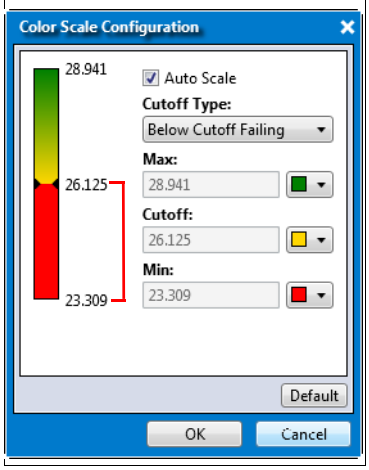


2. Use the provided text fields and color drop-down menus to customize your Color By selection.

- **Auto Scale** check box (when checked) uses the actual minimum (lower bound) and maximum (upper bound) as your min/max scale. Uncheck the Auto Scale check box to enter your min and max number scales in the provided fields.
- Click the **Cutoff Type** drop-down menu to select your cutoff preference.
  - **Above Cutoff Failing** - This presents a hard visual cutoff graph of all values that fail ABOVE the Cutoff value entered. The Above Cutoff data is represented by the color defined for Max. (**Green** in Figure 83)

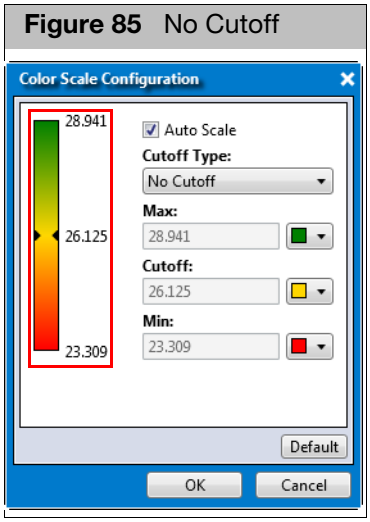
– **Below Cutoff Failing -** This presents a hard visual cutoff graph of all values that fail BELOW the Cutoff value entered. The Below Cutoff data is represented by the color defined for Min. (**Red** in Figure 84)

Figure 84 Below Cutoff



– **No Cutoff -** This presents a smooth 3-point gradient of your defined Max, Min, and colors. (Figure 85)

Figure 85 No Cutoff



3. Click **OK**.

   Your Cutoff preference, entered values, and color selections are now displayed on the graph and saved for future use. If needed, click Default to revert all values back to their factory settings.

Saving the Current Scatter Plot View

1. Click the **Save Image** button.

   An Explorer window appears.

2. Navigate to where you want to save the .PNG file, enter a filename, then click **OK**.

Plate Views

By default, the Viewer generates two Plate Views. (Figure 86) Each have a factory set metric. Plate View 1 is **DQC by Plate** and Plate View 2 is **QC call_rate by Plate**.

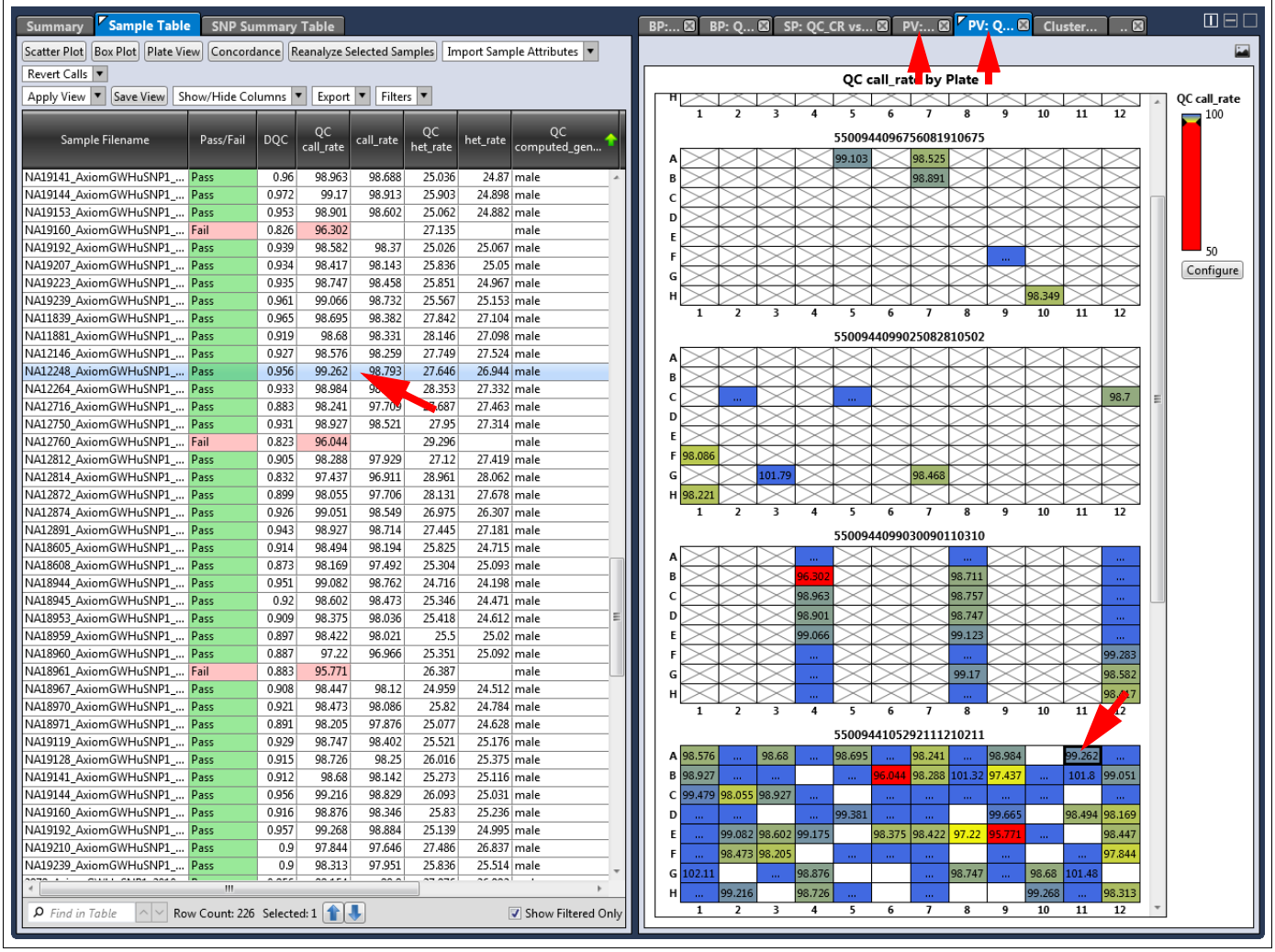
To display a different metric you must create a new Plate View. For more details, see "Adding a New Plate View Metric" on page 70.

The Plate Views display the currently selected (highlighted) metric from the Sample Table and are a graphic representation of the plate used. For example, 96 count plate layouts are shown in Figure 86.

Viewing the Default Plate Views

1. Click to highlight a table entry to view its location within the Plate View or click on a plate position to highlight its corresponding table entry, as shown in Figure 86.

Figure 86 Table and Plate View 1



IMPORTANT! You cannot change a default Plate View, however you can change its Scale Settings, gradient, and coloring. See "Customizing your Plate View Settings" on page 71.

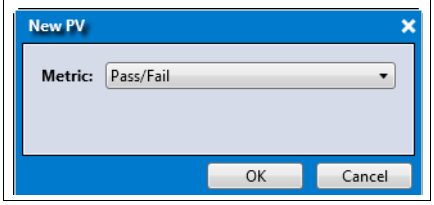
Adding a New Plate View Metric

The default Plate Views cannot be altered, therefore you must click the **Plate View** button to create a new Plate View to reflect your Metric change.

1. Click the **Plate View** button.

   The following window appears: (Figure 87)

Figure 87 Plate View Metric setting

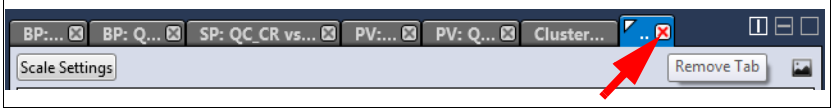


2. Use the drop-down menus to select your Plate View's Metric setting. See the tables in "Sample Table" on page 154 for Metric definitions.
3. Click **OK**.

   The new Plate View window tab appears.

   At any time, click **X** to remove a window/tab, as shown in Figure 88.

Figure 88 New Window/Tab

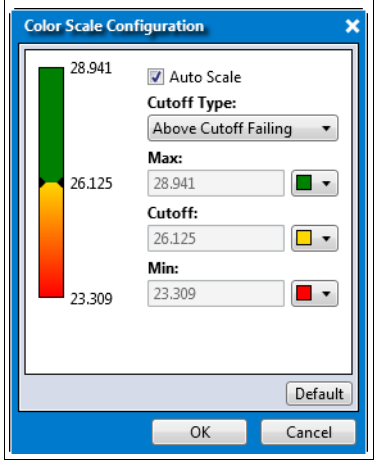


Customizing your Plate View Settings

1. Click Configure .

   The Color Scale Configuration window appears. (Figure 89)

Figure 89 Color Scale options



2. Use the provided text fields and color drop-down menus to customize your Color By selection.

- **Auto Scale** check box (when checked) uses the actual minimum (lower bound) and maximum (upper bound) as your min/max scale. Uncheck the Auto Scale check box to enter your min and max number scales in the provided fields. Note: If the Auto Scale check box remains unchecked, you must enter new scale limits for each subsequent analysis.
- Click the **Cutoff Type** drop-down menu to select the appropriate cutoff (based on the custom Cutoff value you entered).

3. Click **OK**.

   Your new preferences are now displayed and saved for future use.

   At any time, click the **Default** button to revert all the Color Scale Configuration window values back to their factory setting.

Saving the Current Plate View

1. Click the **Save Image** button.

   An Explorer window appears.

2. Navigate to where you want to save the .PNG file, enter a filename, then click **OK**.

Concordance Checks

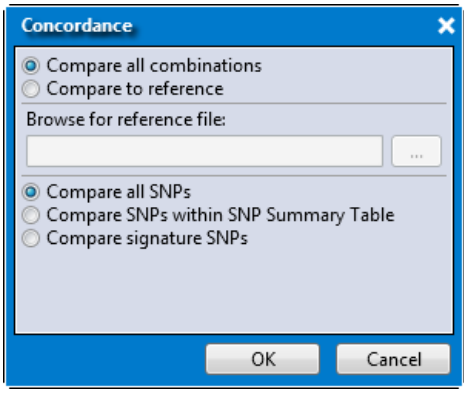
- **Compare all combinations** enables you to compare the SNP calls for all samples. The concordance between all pairwise comparisons for the samples in the dataset/ suitcase are reported.
- **Compare to reference** enables you to compare every sample to a single reference file.

Running a Concordance Check

1. Click the **Concordance** button.

   The following window appears: (Figure 90)

Figure 90 Concordance window



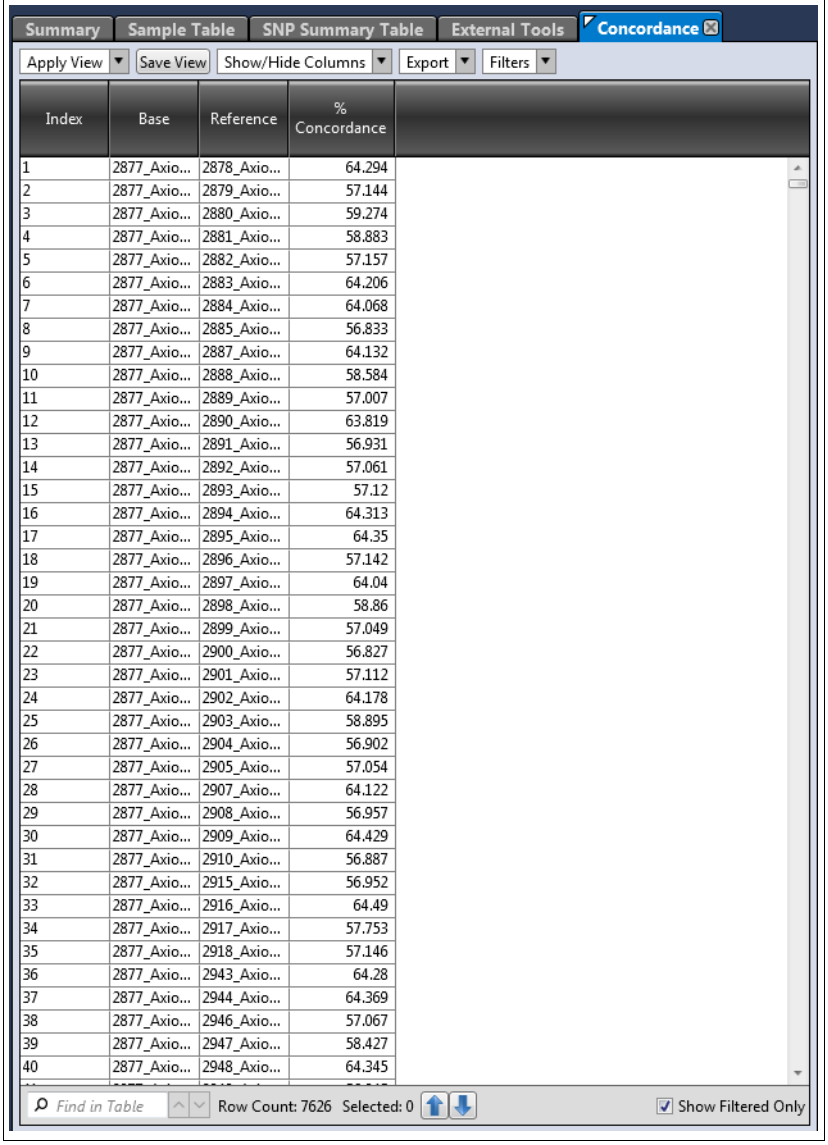
Comparing All Combinations

1. Make sure the **Compare all combinations** radio button is selected.
2. By default, the **Compare all SNPs** button is selected. If needed, click the **Compare signature SNPs within the SNP Summary Table**, or **Compare signature SNPs** radio button.
3. Click **OK**.

   After a few moments, the following Concordance window tab appears: (Figure 91)

IMPORTANT! The amount of time to calculate concordance is proportional to the number of samples squared and the number of SNPs. It is highly recommended you use <1000 SNPs for an All versus All concordance check.

Figure 91 Default Concordance Table example



For definitions of the Concordance columns, see Table 13 on page 158.

Comparing to Reference

1. Click the **Compare to reference** button.

   The **Browse for the reference file** field is now activated.
2. Click the **Browse** button.

   A Windows Explorer window appears.
3. Navigate to the appropriate reference file location, then click **Open**.

   Your Reference file is displayed.

IMPORTANT! The reference file you select can have more than two columns. However, only the first two columns are used during the *Compare to Reference* concordance check. Also, your reference genotypes must be reported using letter call codes (e.g. AA, AB, BB, NoCall).

4. By default, the **Compare all SNPs** button is selected. If needed, click the **Compare signature SNPs within the SNP Summary Table**, or **Compare signature SNPs** radio button.
5. Click **OK**.

   After a few moments, a Concordance window tab appears.

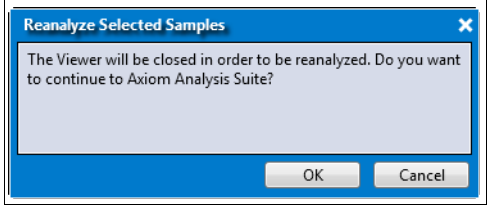
Reanalyzing Samples

To reanalyze sample(s) displayed in the Sample Table:

1. Single-click on a CEL file or Ctrl click, Shift click, or press Ctrl A (to select multiple files).
2. Click the **Reanalyze Selected Samples** button.

   The following message appears. (Figure 92)

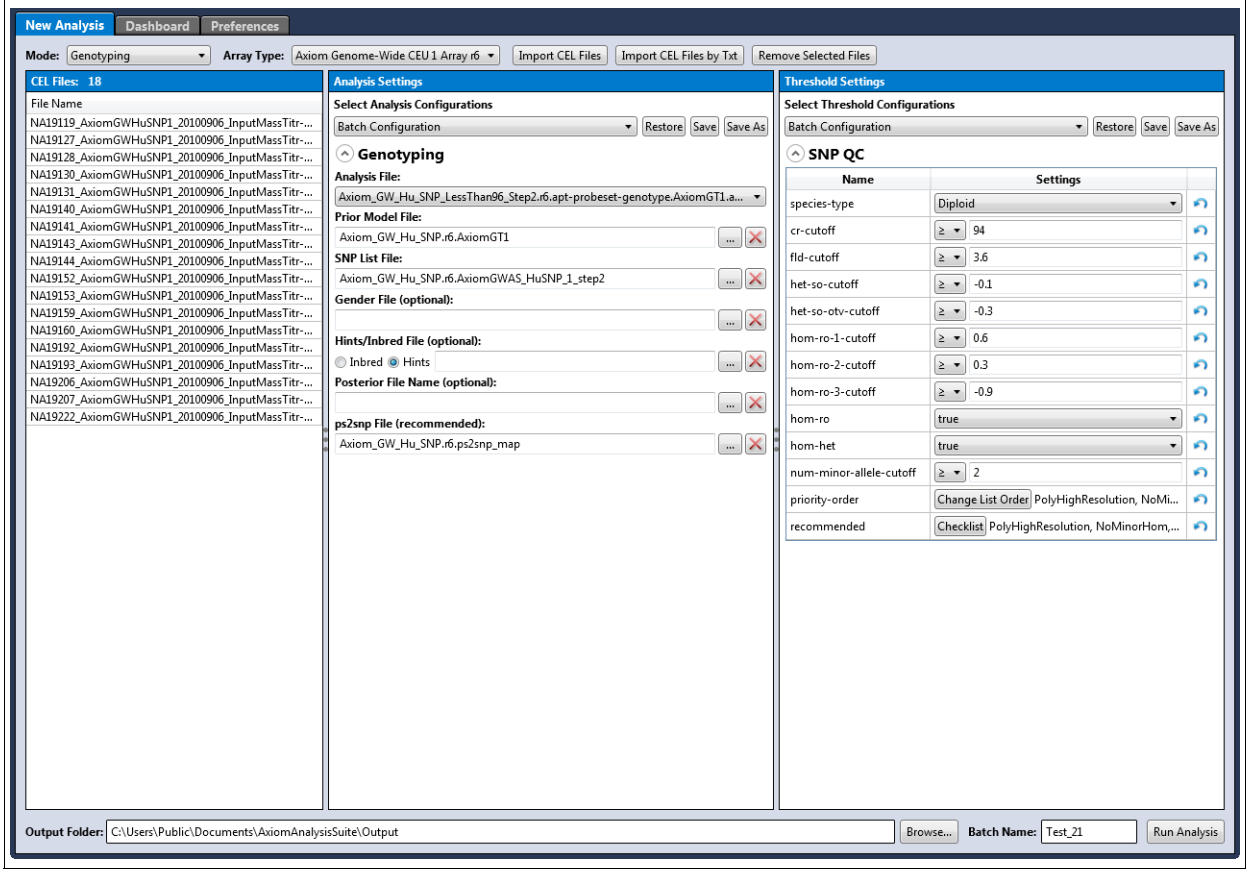
Figure 92 Viewer will be closed message



3. Click **OK**.

The Viewer closes. Your selected samples are displayed and ready for re-analysis in a New Analysis window tab. (Figure 93)

Figure 93 New Analysis window tab



4. See Chapter 2, "Performing an Analysis" on page 19 for instructions on setting up an analysis.

4 The Viewer: SNP Summary Table and Cluster Plot

SNP Summary Table

Figure 94 is an example of a standard SNP Summary Table

Figure 94 SNP Summary Table window

SNP Summary Table

Select Annotation | Manage SNP List | Change/Revert Calls | Reanalyze

Apply View | Save View | Show/Hide Columns | Export | Filters

probeset_id	ConversionType	CR	MinorAlleleFrequency	H.W.p-Value	FLD	HomFLD	HetSO	HomRO	Nclus	n_AA	n_AB	n_BB	n_NC	hemizygous	HomHet	gender_metrics	Call Modified
AFFX-KIT-000001	PolyHighResol...	100	0.464	0.306	20.691	44.561	0.325	2.754	3	39	76	51	0	0	0	all	False
AFFX-KIT-000002	PolyHighResol...	99.398	0.33	0.482	8.127	18.035	0.191	0.846	3	76	69	20	1	0	0	all	False
AFFX-KIT-000003	PolyHighResol...	100	0.22	0.007	6.492	13.807	0.146	0.738	3	107	45	14	0	0	0	all	False
AFFX-KIT-000004	PolyHighResol...	100	0.319	0.145	15.227	33.186	0.318	2.124	3	81	64	21	0	0	0	all	False
AFFX-KIT-000005	PolyHighResol...	100	0.081	1	5.139	11.779	0.094	1.069	3	140	25	1	0	0	0	all	False
AFFX-KIT-000008	PolyHighResol...	100	0.322	0.532	5.412	11.248	0.133	1.127	3	19	69	78	0	0	0	all	False
AFFX-KIT-000009	NoMinorHom	100	0.075	0.602	21.081		0.247	4.47	2	0	25	141	0	0	1	all	False
AFFX-KIT-000012	PolyHighResol...	100	0.368	0.388	20.114	42.061	0.238	2.784	3	69	72	25	0	0	0	all	False
AFFX-KIT-000013	PolyHighResol...	99.398	0.136	0.19	5.134	10.814	0.169	0.813	3	125	35	5	1	0	0	all	False
AFFX-KIT-000014	PolyHighResol...	100	0.099	0.663	18.947	38.956	0.567	2.803	3	2	29	135	0	0	0	all	False
AFFX-KIT-000015	PolyHighResol...	100	0.181	1	30.972	67.77	0.439	4.658	3	5	50	111	0	0	0	all	False
AFFX-KIT-000016	PolyHighResol...	100	0.135	1	23.086	48.374	0.579	2.409	3	3	39	124	0	0	0	all	False
AFFX-KIT-000017	PolyHighResol...	100	0.078	1	20.903	44.214	0.42	3.888	3	141	24	1	0	0	0	all	False
AFFX-KIT-000018	PolyHighResol...	99.398	0.23	1	16.629	34.399	0.254	2.451	3	9	58	98	1	0	0	all	False
AFFX-KIT-000019	PolyHighResol...	99.398	0.373	0.043	6.281	15.899	0.235	1.536	3	29	65	71	1	0	0	all	False
AFFX-KIT-000021	PolyHighResol...	99.398	0.179	0.79	8.169	17.256	0.331	1.384	3	112	47	6	1	0	0	all	False
AFFX-KIT-000022	PolyHighResol...	100	0.434	0.381	18.57	38.557	0.397	2.523	3	56	76	34	0	0	0	all	False
AFFX-KIT-000023	PolyHighResol...	99.398	0.309	0.238	6.384	14.553	0.193	1.015	3	82	64	19	1	0	0	all	False
AFFX-KIT-000025	PolyHighResol...	100	0.494	0.439	17.222	36.978	0.348	2.333	3	43	78	45	0	0	0	all	False
AFFX-KIT-000026	PolyHighResol...	99.398	0.436	0.147	13.408	33.126	0.509	1.785	3	36	72	57	1	0	0	all	False
AFFX-KIT-000027	PolyHighResol...	100	0.41	0	23.078	47.688	0.238	2.894	3	41	54	71	0	0	0	all	False
AFFX-KIT-000029	PolyHighResol...	99.398	0.397	0	11.659	24.178	0.135	1.097	3	75	49	41	1	0	0	all	False
AFFX-KIT-000031	PolyHighResol...	100	0.238	0.671	15.932	33.638	0.379	2.353	3	95	63	8	0	0	0	all	False
AFFX-KIT-000032	PolyHighResol...	100	0.377	0.071	20.768	43.819	0.607	2.467	3	29	67	70	0	0	0	all	False
AFFX-KIT-000033	PolyHighResol...	99.398	0.258	0	13.454	27.084	0.505	2.629	3	102	41	22	1	0	0	all	False
AFFX-KIT-000049	PolyHighResol...	100	0.386	0.583	12.018	25.089	0.511	2.074	3	61	82	23	0	0	0	all	False
AFFX-KIT-000050	PolyHighResol...	100	0.473	0.227	6.032	12.798	0.147	0.776	3	50	75	41	0	0	0	all	False
AFFX-KIT-000121	PolyHighResol...	100	0.283	0.15	13.623	28.574	0.309	2.419	3	10	31	49	0	0	0	female	False
AFFX-KIT-000198	PolyHighResol...	100	0.356	0.228	17.117	36.866	0.365	2.704	3	40	36	14	0	0	0	female	False
AFFX-KIT-000209	PolyHighResol...	100	0.333	0.018	4.766	10.896	0.129	0.697	3	15	30	45	0	0	0	female	False
AFFX-KIT-000226	PolyHighResol...	98.889	0.371	0.002	12.302	27.039	0.556	2.075	3	42	28	19	1	0	0	female	False
AFFX-KIT-000408	PolyHighResol...	100	0.217	1	19.59	41.485	0.12	3.567	3	4	31	55	0	0	0	female	False
AX-162003418	MonoHighRes...	100							1							NA	False
AX-162005580	PolyHighResol...	100							2							NA	False
AX-162007660	CallRateBelow...	77.711							3							NA	False
AX-162011899	PolyHighResol...	100							3							NA	False
AX-162020167	Other	100							1							NA	False

NA | Row Count: 37 Selected: 0 | Show Filtered Only

Multi-allele probesets are probesets that can report calls from more than two alleles. Multi allele probesets can be identified by selecting the annotation field **Allele_Count**. Additional columns can be shown that report metrics for multi-allele probesets.

Multi-allele probesets are reported differently than the more common biallele probesets.

- Multi-allele probesets will not report NoMinorHom or OTV ConversionTypes.
- Multi-allele probesets do not report values for many of the default SNP Summary Table columns. Click **Show/Hide Columns** to select additional Columns with multi-allele probeset metrics.

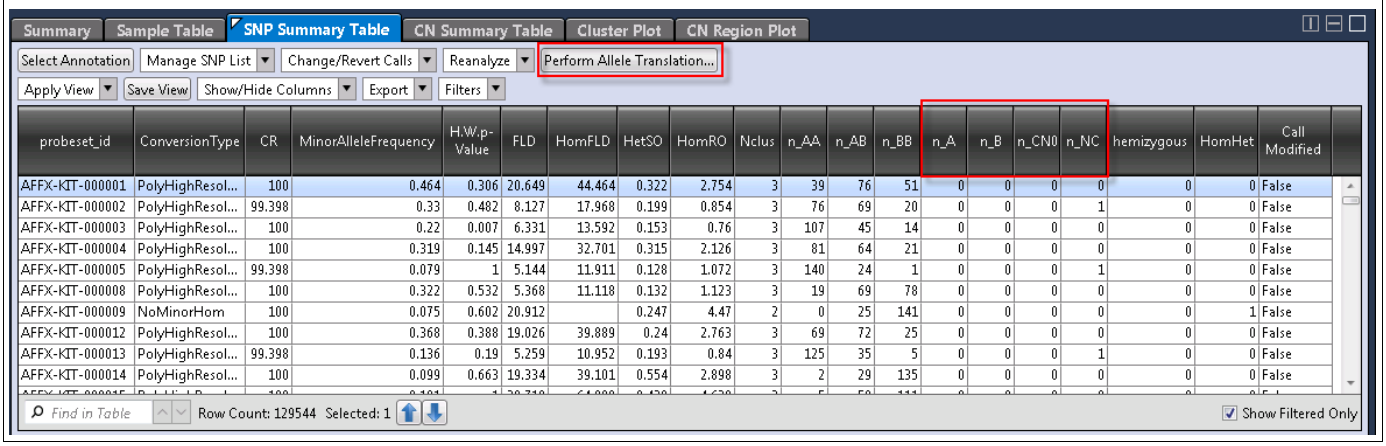
Note: If an analysis batch does not include samples needed to evaluate a SNP's performance, then no metrics (including ConversionType) will be reported. For example, if an analysis batch does not contain male samples, no metrics are reported for Y chromosome SNPs and none of these SNPs are added to the Recommended SNP list.

Figure 95 represents a SNP Summary Table with CN-aware genotyping and Allele Translation.

Note: The Perform Allele Translation... button only appears if supported arrays are available. See Chapter 5, "Allele Translation" on page 97 for more information.

A SNP Summary Table with CN-aware genotyping also displays a count of calls for hemizygous genotypes (n_A, n_B) and a count of ZeroCN calls (n_CN0), as shown in Figure 95. These additional columns appear for arrays that support copy number-aware genotyping. For more information, see Appendix A, "Predefined Region CN Analysis Variations" on page 122.

Figure 95 SNP Summary Table with CN-aware genotyping and Allele Translation



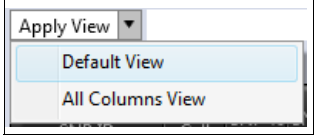
probeset_id	ConversionType	CR	MinorAlleleFrequency	H.W.p-Value	FLD	HomFLD	HetSO	HomRO	Nclus	n_AA	n_AB	n_BB	n_A	n_B	n_CN0	n_NC	hemizygous	HomHet	Call Modified
AFFX-KIT-000001	PolyHighResol...	100	0.464	0.306	20.649	44.464	0.322	2.754	3	39	76	51	0	0	0	0	0	0	False
AFFX-KIT-000002	PolyHighResol...	99.398	0.33	0.482	8.127	17.968	0.199	0.854	3	76	69	20	0	0	0	1	0	0	False
AFFX-KIT-000003	PolyHighResol...	100	0.22	0.007	6.331	13.592	0.153	0.76	3	107	45	14	0	0	0	0	0	0	False
AFFX-KIT-000004	PolyHighResol...	100	0.319	0.145	14.997	32.701	0.315	2.126	3	81	64	21	0	0	0	0	0	0	False
AFFX-KIT-000005	PolyHighResol...	99.398	0.079	1	5.144	11.911	0.128	1.072	3	140	24	1	0	0	0	1	0	0	False
AFFX-KIT-000008	PolyHighResol...	100	0.322	0.532	5.368	11.118	0.132	1.123	3	19	69	78	0	0	0	0	0	0	False
AFFX-KIT-000009	NoMinorHom	100	0.075	0.602	20.912		0.247	4.47	2	0	25	141	0	0	0	0	0	1	False
AFFX-KIT-000012	PolyHighResol...	100	0.368	0.388	19.026	39.889	0.24	2.763	3	69	72	25	0	0	0	0	0	0	False
AFFX-KIT-000013	PolyHighResol...	99.398	0.136	0.19	5.259	10.952	0.193	0.84	3	125	35	5	0	0	0	1	0	0	False
AFFX-KIT-000014	PolyHighResol...	100	0.099	0.663	19.334	39.101	0.554	2.898	3	2	29	135	0	0	0	0	0	0	False

Using the SNP Summary Table

Setting your SNP Summary Table View

1. Click the **Apply View** drop-down. (Figure 96)

Figure 96 Apply View drop-down

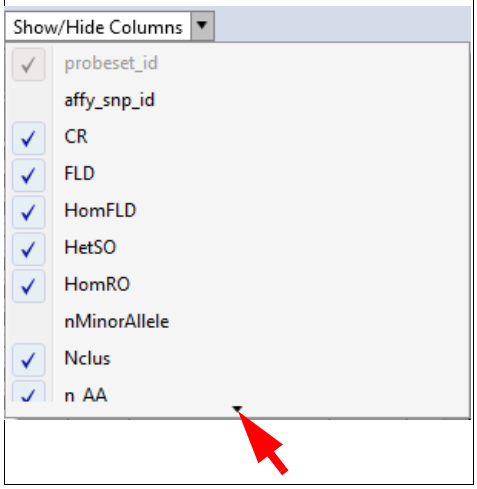


- The **Default View** is the initial table view and includes a preset number of columns.
- The **All Columns View** displays the maximum available columns.

Adding and Removing Table Columns

1. Click the **Show/Hide Columns** drop-down. (Figure 97)
   A list of available columns appear.

Figure 97 Show/Hide Columns drop-down



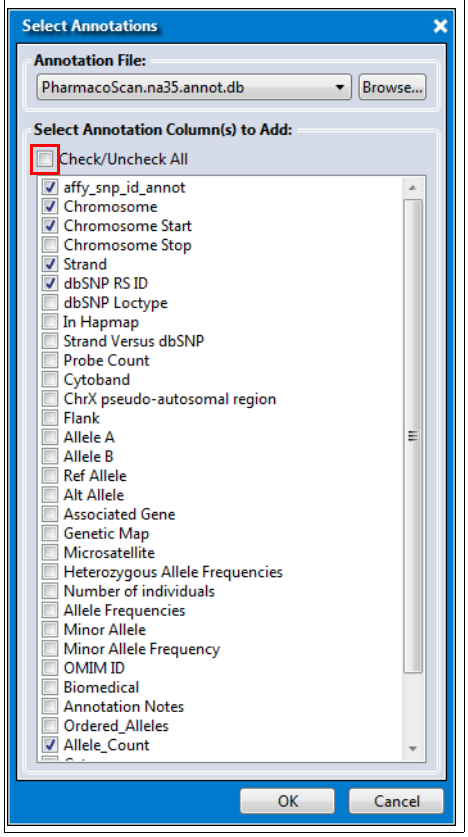
2. Click the check box next to the column(s) you want to add (Show) on the table. Click to uncheck a column you want to remove (Hide) from the table. Mouse over the menu's down arrow to reveal more available column choices. See Table 15 on page 162 for column definitions.

Selecting Annotations

1. Click the **Select Annotation** button.

   The following window appears. (Figure 98)

Figure 98 Available Annotations



2. Select the appropriate Annotation File from the drop-down menu list.
3. Click the check box next to the Annotation Column(s) you want to add to the table or click the **Check/Uncheck All** check box (Figure 98) to add or remove ALL available annotations. See "Annotations and Columns" on page 159 for each Annotation's definition.
4. Click **OK**.

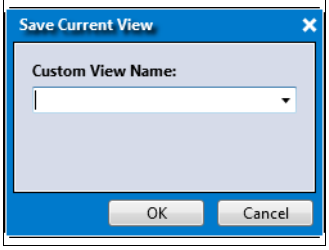
   Your selected annotation columns are now added to the right side of the SNP Summary Table.

Saving your Table Column View

1. After you have your preferred SNP Summary Table columns set, click **Save View**.

   The following window appears: (Figure 99)

Figure 99 Save New Custom View



2. Enter a name, then click **OK**.

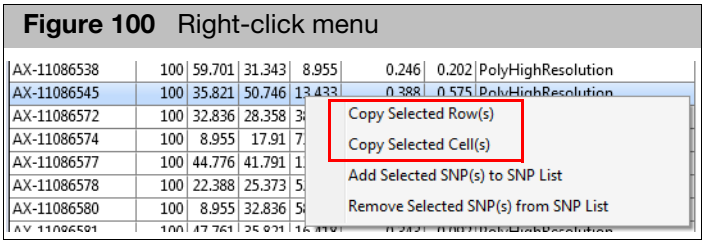
   Your custom table view is now saved and stored inside the Apply View menu for future use.

Copying Selected Row(s)

1. Right-click on a row you want to copy. Optional: Single-click on a row or Ctrl click, Shift click, or press Ctrl A (to select multiple rows).

   A menu appears. (Figure 100)

Figure 100 Right-click menu



2. Click **Copy Selected Row(s)**.

   The row data is now ready for pasting (Ctrl v).

Copying Selected Cell(s)

1. Right-click on a cell you want to copy. Optional: Single-click on a row or Ctrl click, Shift click, or press Ctrl A (to select multiple cells).

   A menu appears. (Figure 100)

2. Click **Copy Selected Cell(s)**.

   The cell data is now ready for pasting (Ctrl v).

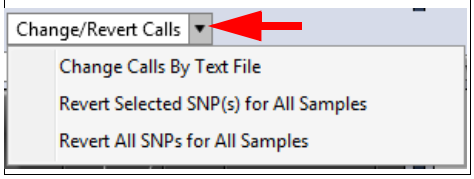
Changing or Reverting Genotype Calls

Reverting Calls changes them back to what they were originally called by the algorithm. No other history is saved, only its current and original values.

1. Click the **Change/Revert Calls** drop-down.

   The following menu appears: (Figure 101)

Figure 101 Revert Calls drop-down menu



Changing Genotype Calls by Text File

1. Click **Change Calls By Text File**.

   An Explorer window appears.
2. Navigate to the text file's location.

IMPORTANT! Your Change Call by Text file must start with the header *probeset_id* and use numeric or alphabetic codes for the genotype, as shown in Figure 102. It also cannot contain any annotation columns.

Figure 102 Change Call by Text file example shown in Excel (as a tab-delimited text file)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	probeset_id	2877_Axic	2878_Axic	2879_Axic	2880_Axic	2881_Axic	2882_Axic	2883_Axic	2884_Axic	2885_Axic	2887_Axic	2888_Axic	2889_Axic	2890_Axic	2891_Axic	2892_A
2	AX-11699985	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
3	AX-11700086	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
4	AX-11700405	BB	BB	AB	BB	AB	AA	BB	BB	BB	AB	BB	AA	BB	AA	AA
5	AX-11700418	AB	AB	BB	BB	AB	AB	AB	AB	BB	BB	AB	BB	AB	BB	BB
6	AX-11700675	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
7	AX-11700679	AB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
8	AX-11700939	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA

3. Click **Open**.

   Your Genotype Calls are now changed and reflected in the Cluster Plot.

Reverting Selected SNPs for all Samples

1. Single-click on a SNP file or Ctrl click, Shift click, or press Ctrl A (to select multiple files) within the SNP Sample Table.
2. Click **Revert Selected SNP(s) for All Samples**.

Reverting all SNPs for all Samples

Use this feature to perform a master SNPs reset.

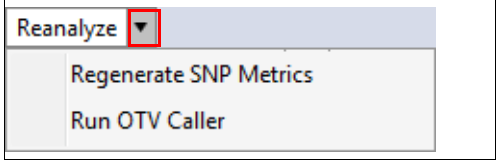
1. Click **Revert All SNPs for All Samples**.

IMPORTANT! Once *Revert Calls* is performed, the selected calls will be reverted to original calls. This cannot be undone.

Reanalyzing your SNP Summary Table Data

The **Reanalyze** drop-down menu (Figure 103) offers optional steps for post-processing SNP data. These functions utilize the output files from the previous genotyping and classification steps as input.

Figure 103 Reanalyze drop-down menu



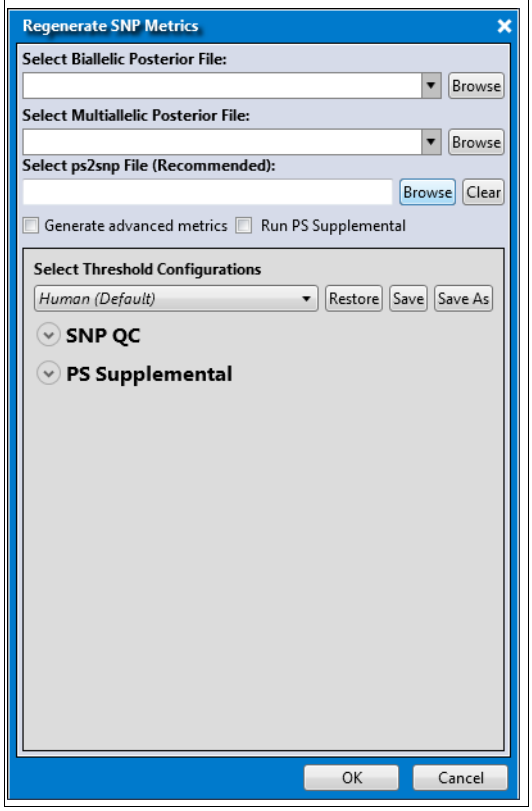
Regenerate SNP Metrics

The **Regenerate SNP Metrics** operation allows you to recalculate SNP Metrics and ConversionType classifications, which you may want to do if you have edited any genotype calls, or if you want to change SNP QC thresholds, or if you want to generate additional metrics and classifications.

1. Click **Regenerate SNP Metrics**.

   The following window appears: (Figure 104)

Figure 104 Regenerate SNP Metrics window

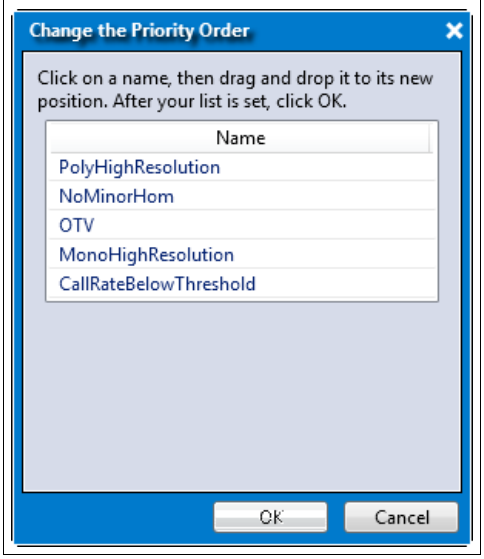


1. Use the drop-down to select the appropriate Posterior File or click its **Browse** button.

   An Explorer window appears.
2. Navigate to Posterior File you want to use, then click **Open**.
3. Choose an appropriate ps2snp File (recommended), as described in "Using the Analysis Settings Fields" on page 24.
4. (Optional) Click the **Generate advanced metrics** check box to calculate additional SNP-specific posterior metrics.
5. (Optional) Click the **Run PS Supplemental** check box to generate additional metrics and probeset ConversionTypes. Note: PS Supplemental performs further classification that may be needed for polyploid organisms, complex genomes, or inbred populations.
6. Select the Threshold Configuration you want to use, as described in"Customizing Thresholds" on page 28.
7. Click to expand and use the **SNP QC** drop-down menu selections and text fields to setup the regeneration of your SNP Metric.
8. To change the priority-order of the various assigned ConversionTypes, go to the SNP QC setting's priority-order column and click the Change List Order button.

   The following window appears: (Figure 105)

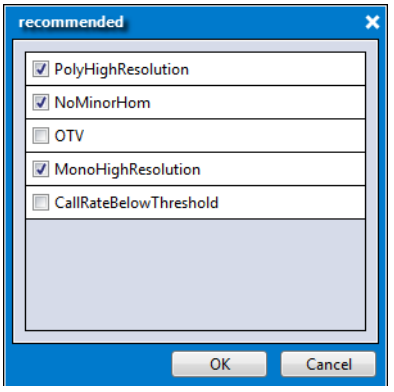
Figure 105 Change the Priority Order window



9. Click and hold onto the selection you want to move, then drag and drop it into its new position. After you get the order of priority you want, click **OK**.
10. To change the recommended options, click Checklist.

The following window appears: (Figure 106)

Figure 106 Recommended window



11. Click to check/uncheck the available recommended options. Click a field's button to return its value back to its default setting.

Note: If you have checked **Run PS Supplemental** check box, then the associated threshold configurations are relevant. Edit them as needed.

12. Click **OK**.

    After the process is complete, your SNP Summary Table is updated.

Running OTV Caller

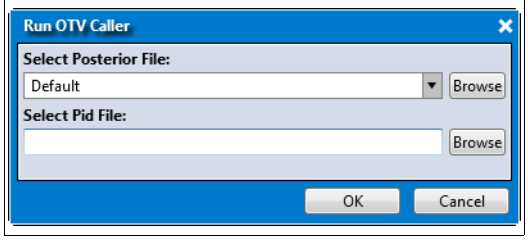
The OTV Caller is intended for SNPs that have been classified as likely having off-target variants (OTV), or for SNPs with unusually large Y-dimension variance (as identified by PS Supplemental option of the Regenerate SNP Metrics).

OTV Caller function performs post-processing analysis to identify miscalled clustering and identify which samples should be in the OTV cluster and which samples should remain in the AA, AB, or BB clusters. Samples in the OTV cluster are re-labeled as OTV.

1. Click **Run OTV Caller**.

   The following window appears: (Figure 107)

Figure 107 Run OTV Caller window



1. Use the drop-down to select the appropriate Posterior File or click its **Browse** button.

   An Explorer window appears.
2. Navigate to Posterior File you want to use, then click **Open**.
3. Click the Select Pid File's **Browse** button.

   An Explorer window appears.
4. Navigate to Pid File you want to use, then click **Open**.
5. Click **OK**.

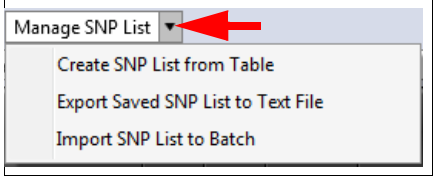
Managing your SNP List

Axiom Analysis Suite enables lists of SNPs to be saved within the application.

1. Click the Manage SNP List drop-down menu (Figure 108), then select the option you want.

Note: The Recommended SNP List is auto-generated and updated whenever SNP metrics are calculated. Avoid using this reserved (default) SNP List name for your custom lists.

Figure 108 Manage SNP List drop-down menu

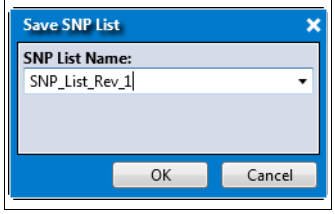


Saving your Current SNP List

1. To save all SNPs currently displayed in the SNP Summary Table, click **Create SNP List from Table**.

   The following window appears: (Figure 109)

Figure 109 Save SNP List



2. Enter a name, then click **OK**.

Exporting your SNP List

Before exporting a SNP List you must first create one. If no SNP Lists are detected, a message box appears. Click **OK** to acknowledge the message, then go to "Saving your Current SNP List" to create a SNP List.

1. Click **Export Saved SNP List to Text File**.

   An Explorer window appears.

2. Navigate to an export location, enter a name, then click **Save**.

Importing a SNP List

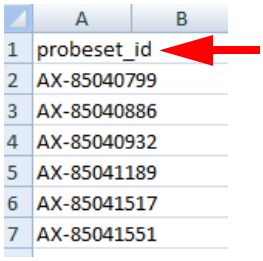
1. Click **Import SNP List to Batch**.

   An Explorer window appears.

2. Navigate to your SNP List location containing your tab-delimited text file.

   Your first row/column header must be labeled **probeset_id**, as shown in Figure 110, otherwise an error message appears.

Figure 110 Required probeset_id column header



3. Click **Open**.

   Your imported SNP List now appears in the SNP Summary Table.

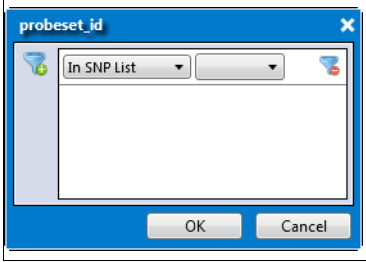
Using your Saved SNP List

Displaying SNPs in a SNP list

1. Click to select the **probeset_id** column, then right-click on it.
2. Click **Filter**.

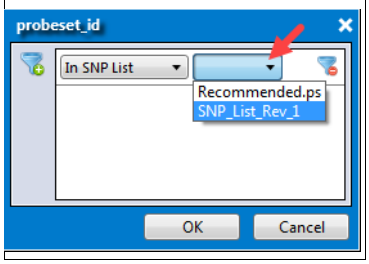
   The Filter window appears. (Figure 111)

Figure 111 Filter window



3. Add your previously saved SNP List by selecting it from the drop-down list. (Figure 112)

Figure 112 Saved SNP List drop-down



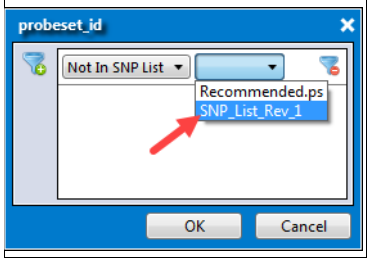
4. Click **OK**.

   Only the SNPs in your SNP List are displayed in the SNP Summary Table.

Displaying SNPs that are not in your SNP List

1. Click to select the **probeset_id** column, then right-click on it.
2. Click **Filter**.
3. Click the **In SNP List** drop-down, then click to select **Not in SNP List**. (Figure 113)
4. Click the Saved SNP List drop-down, select your saved SNP List, then click **OK**.

Figure 113 Saved SNP List drop-down



Your SNPs from the SNP List are no longer displayed in the SNP Summary Table.

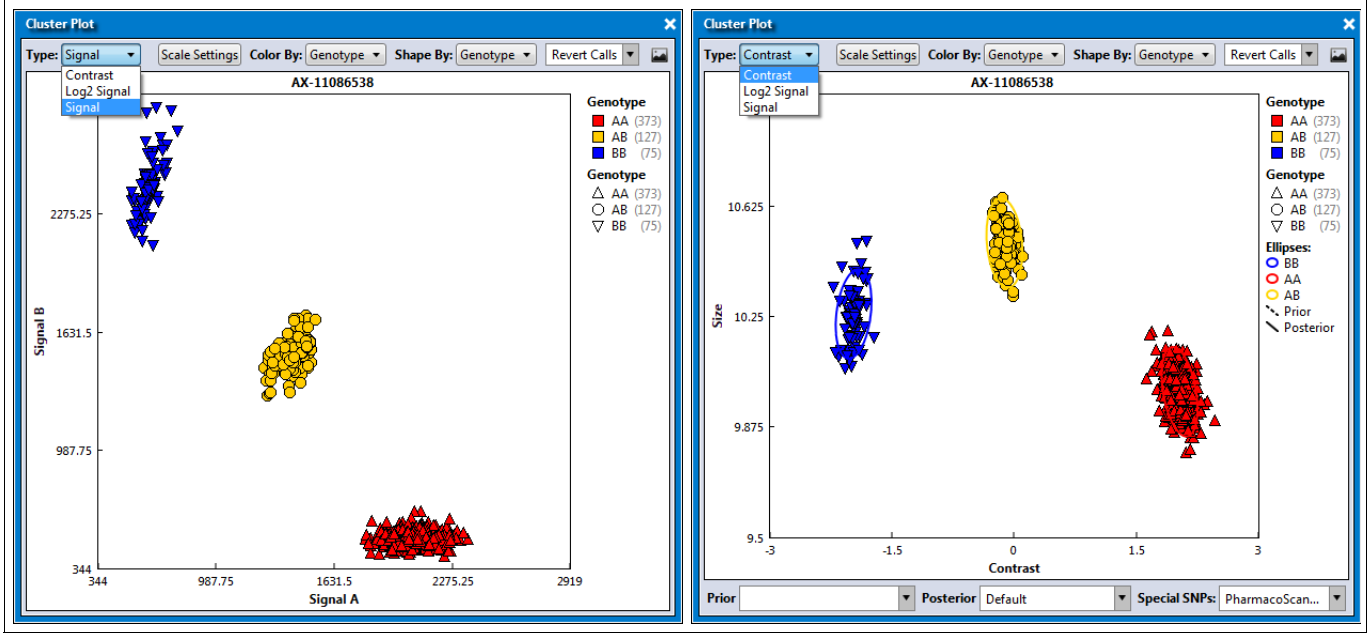
Cluster Plot

The Cluster Plot displays the SNP calls for selected samples as a set of points in the clustering space used for making the calls. A visual inspection of select Cluster Plots aids in identifying problematic SNPs and enables you to manually change calls.

As shown in Figure 114, use the **Type** option above the plot to switch between **Signal** view (left plot), **Log2 Signal** view (not shown), and **Contrast** view (right plot).

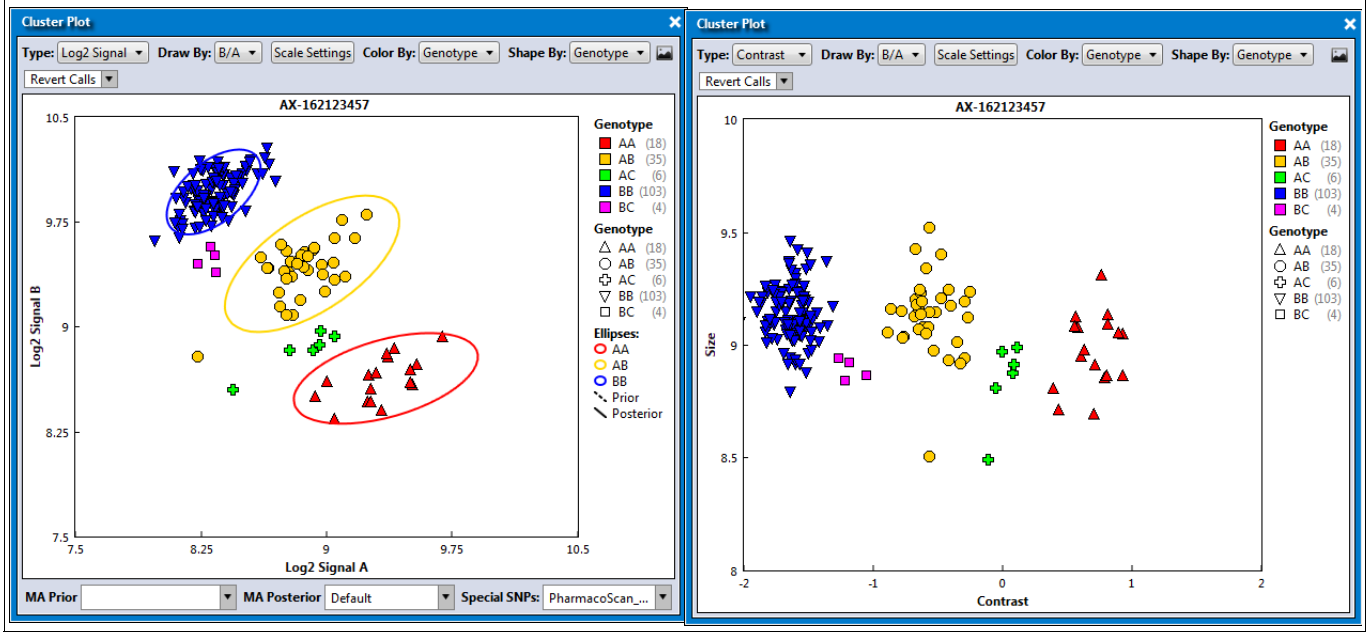
Note: For bi-allele probesets (those that measure two alleles), only the **Contrast** view displays the ellipses that illustrate the prior and posterior knowledge of genotype cluster positions.

Figure 114 Cluster window tab for bi-allelic probesets



Multi-allele probesets measure more than two alleles. Unlike bi-allele probesets, multi-allele probesets are genotyped in the Log2 Signal space. For multi-allele probesets, the ellipses that illustrate the prior and posterior knowledge of genotype cluster positions are only displayed when selecting plot Type **Log2 Signal**. See Figure 115 on page 90 for plot examples of a multi-allele probeset.

Figure 115 Cluster window tab for a multi-allele probeset

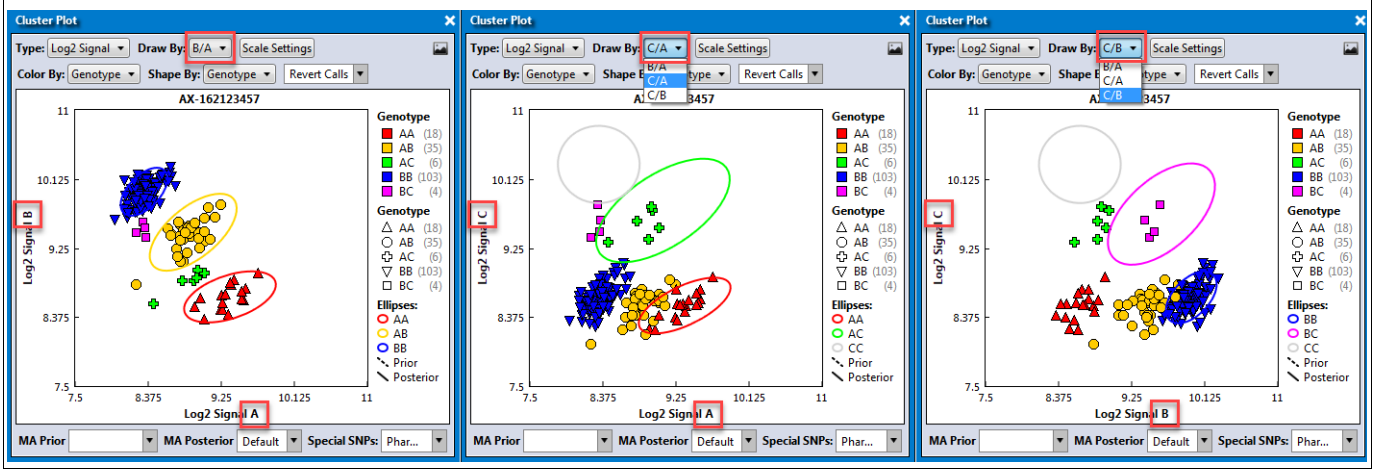


Note: Multi-allele probesets in the SNP Summary Table can be more easily identified by adding the annotation field Allele_Count and/or Ordered_Alleles. If a probeset has Allele_Count of 3 or greater, it is a multi-allele probeset.

Multi-allele probesets measure more than two alleles, and so more than two allele signal channels are present. The cluster plot only shows data for two allele signal channels. If multi-allele probesets are present in the current results, then the top of the cluster plot will also display the Draw By menu.

See Figure 116 for an example of using Draw By to select different pairs of allele signal channels for plotting. The default Draw By option is selected based on the probeset's most common genotype allele calls in the data set.

Figure 116 Cluster window tab - Using Draw By to select different pairs of allele signal channels for plotting.



Using the Cluster Plot

Displaying a SNP Cluster Plot that Corresponds with a SNP

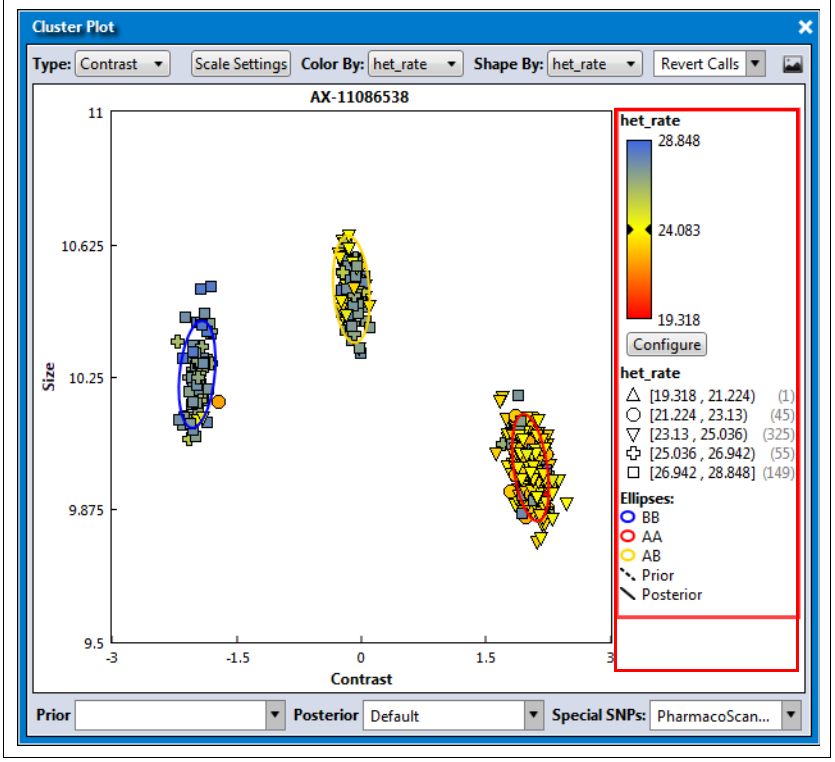
1. In the SNP Summary Table, click on row (SNP) of interest.

Note: Use the arrow keys on the keyboard to toggle through the list. As you toggle through the list, the Cluster Plot auto-updates to match your selected SNP.

2. Click the **Color By** and **Shape By** drop-down menus to select the combination (X and Y axis) view you want. See Table 12 on page 154 for Color By and Shape By definitions. Note: Your imported sample attributes are also available for use within the Color By and Shape By drop-down menus.

   The appropriate legend appears within the plot. (Figure 117)

Figure 117 Color By and Shape By Legend example

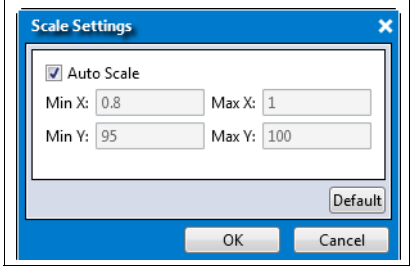


Setting New Scale Setting Ranges

1. Click Scale Settings.

   The following window appears. (Figure 118)

Figure 118 Scale Settings window



   By default, the window displays your current range values.

2. Enter your new ranges in the appropriate text fields.
3. Click **OK**.

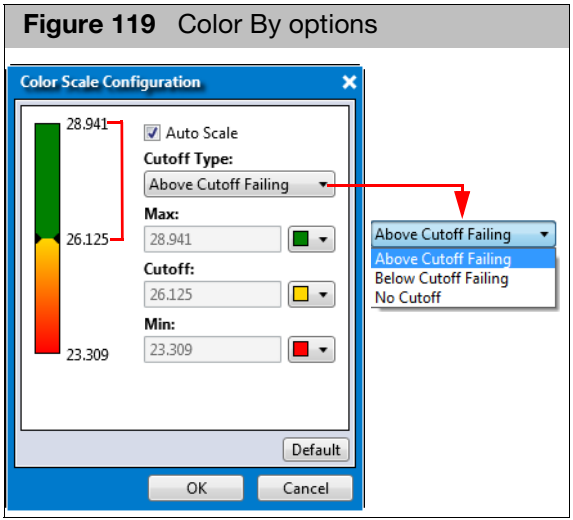
   Click Default to return all values back to their factory settings.

Customizing Color By Settings

1. Click Configure.

   The Color Scale Configuration window appears. (Figure 119)

Figure 119 Color By options

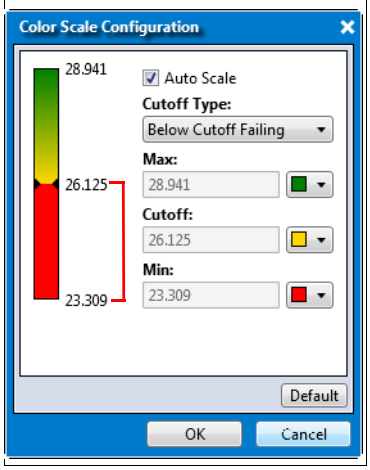


2. Use the provided text fields and color drop-down menus to customize your Color By selection.

- **Auto Scale** check box (when checked) uses the actual minimum (lower bound) and maximum (upper bound) as your min/max scale. Uncheck the Auto Scale check box to enter your min and max number scales in the provided fields. Note: If the Auto Scale check box remains unchecked, you must enter new scale limits for each subsequent analysis.
- Click the **Cutoff Type** drop-down menu to select your cutoff preference.
  - **Above Cutoff Failing** - This presents a hard visual cutoff graph of all values that fail ABOVE the Cutoff value entered. The Above Cutoff data is represented by the color defined for Max. (**Green** in Figure 119)

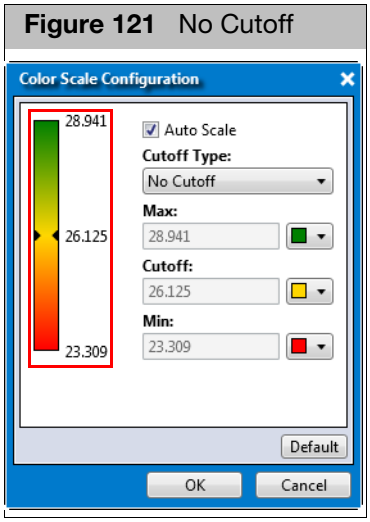
– **Below Cutoff Failing -** This presents a hard visual cutoff graph of all values that fail BELOW the Cutoff value entered. The Below Cutoff data is represented by the color defined for Min. (**Red** in Figure 120)

Figure 120 Below Cutoff



– **No Cutoff -** This presents a smooth 3-point gradient of your defined Max, Min, and colors. (Figure 121).

Figure 121 No Cutoff



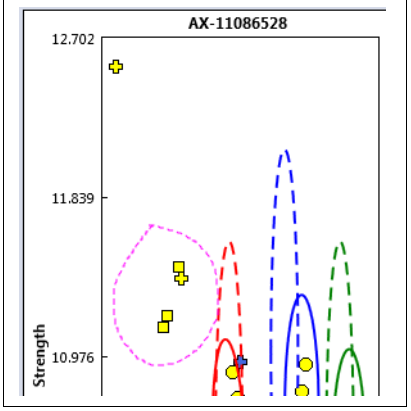
3. Click **OK**.

   Your Cutoff preference, entered values, and color selections are now displayed on the graph and saved for future use. If needed, click the **Default** button to revert ALL values back to their factory setting.

Selecting Multiple Samples in a Cluster Plot

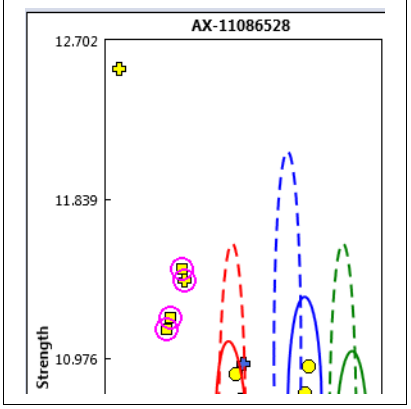
1. Drag the cursor around a group of samples to draw a pink-dotted closed loop around them, as shown in Figure 122.

Figure 122 Selecting multiple samples



2. Release the mouse button to select the group of lassoed samples. (Figure 123)

Figure 123 Selecting multiple samples

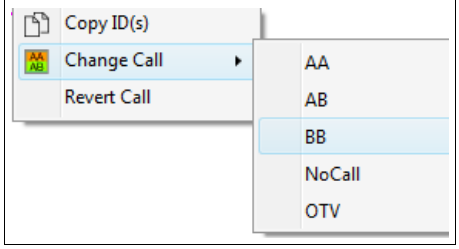


Changing a Sample's Call for a Single SNP

1. Highlight the sample or samples you want to modify, then right-click on them.

   A menu appears.
2. Click **Change Call**, then move your cursor to the right, then click to select a different call, an OTV (Off Target Variant), or No Call. (Figure 124) Note: If the array supports CN-aware Genotyping, additional Call Code choices appear.

Figure 124 Change Call menu



The Call is now changed, but not the position. The image may or may not change, as it depends on the Color By and Shape By options you selected.

Reverting a Single Call

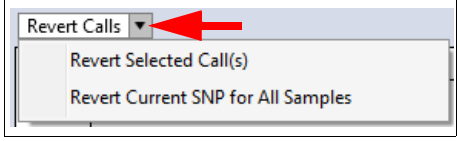
1. Single-click to highlight the Call you want to revert back, then right-click on it.

   A menu appears.
2. Click **Revert Call**.

Reverting Multiple Calls

1. Drag the cursor around a group of samples to draw a pink-dotted closed lasso shape around them.
2. Release the mouse button to highlight your selected samples.
3. Click the **Revert Call** drop-down menu (Figure 125), then click to select **Revert Selected Call(s)**.

Figure 125 Revert Calls drop-down menu

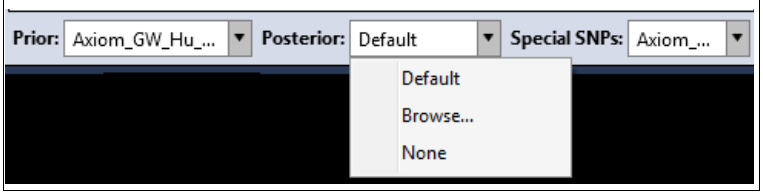


Displaying Cluster Model Data

By default, the **Prior**, **Posterior**, and **Special SNPs** drop-down selections are preset to best suit the currently displayed Cluster Plot.

1. Click the appropriate drop-down menu (Figure 126), then click to select a new setting.

Figure 126 Bi-Allele Cluster Model Data drop-down menus



- If you select **Browse**, an Explorer window appears. Navigate to your folder location as you normally would, then click **Open** to display your data within the graph.
- Selecting **None**, conceals (hides) the selected graph data.

Saving the Current Cluster Plot View

1. Click the **Save Image** button.
   An Explorer window appears.
2. Navigate to where you want to save the .PNG file, enter a filename, then click **OK**.

5 Allele Translation

About Translations

For supported array types (e.g. PharmacoScan), Axiom Analysis Suite will provide the option to convert (translate) the genotype calls of an important subset of SNPs to functional allele calls using standardized nomenclature wherever possible. The software enables you to:

- Quickly identify possible rare alleles or missing data.
- Identify haplotype and SNP-level sequence variation in the test samples.
- Annotate the reported genotypes across translated SNPs to indicate genomic, mRNA, or peptide changes resulting from any observed variation.
- Predict general gene activity based on detected diplotypes.

See Appendix B, "About Allele Translation" on page 139 for more information.

Performing Allele Translation

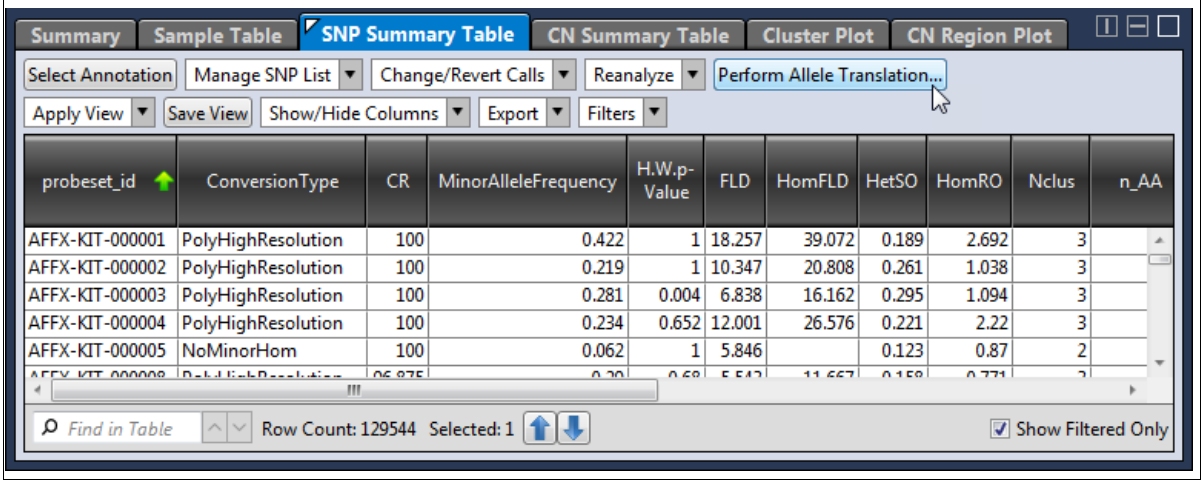
The Perform Allele Translation operation is available only if the following conditions are met:

- The library folder must include the same library package (Array Type with revision number) used to generate the batch results.
- Inside the matched library package folder, the files with the extensions ***.dc_annot.csv** and ***.translation** must exist.

To perform an allele translation:

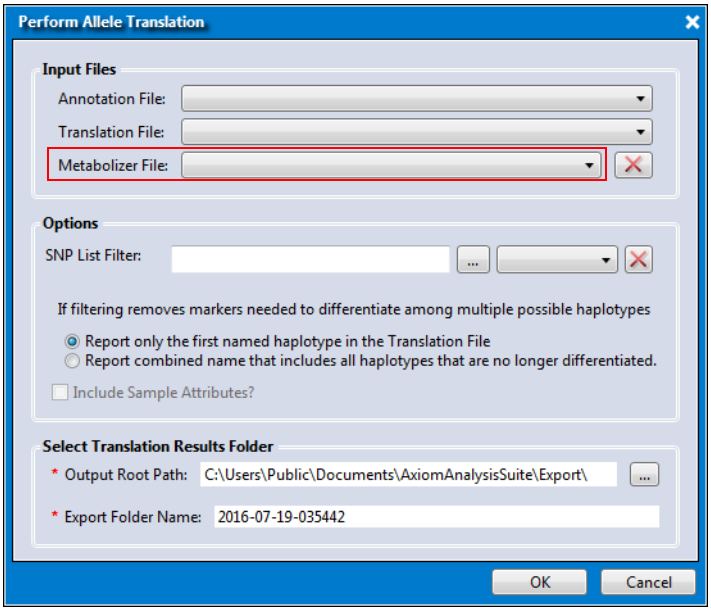
1. From an open batch in the Viewer, navigate to the SNP Summary Table, then click the **Perform Allele Translation** button, as shown in Figure 127.

Figure 127 SNP Summary Table window



The Perform Allele Translation window appears. (Figure 128)

Figure 128 Perform Allele Translation window



2. At the Perform Allele Translation window, use the provided drop-down menus to select an **Annotation File**, **Translation File**, and an optional **Metabolizer File**.
   Note: If you disagree with the phenotype interpretations, you may want to leave the Metabolizer File option blank. (Figure 128) A Phenotype report will not be created if this option is not used.

Allele Translation Options

Refer to Table 1 and Table 2 for descriptions of the available **Allele Translation** options.

Table 1 Allele Translation options

Select Options	Description
SNP List Filter	Choose this option to translate only the genotypes of SNPs in a user-specified probeset list. Click the Browse button to select the marker list, or select from SNP lists you're already imported to the batch results. WARNING: If you supplied a custom optional SNP List File for Genotyping in Analysis Settings when setting up your analysis, you should either filter on the same marker list for Allele Translation, or filter on an even more restrictive list. Otherwise, you may see SNPs with NotAvailable calls in your translation reports. The presence of NotAvailable calls may increase the number of reported haplotype possibilities.
Report only the first named haplotype in the translation file (Default)	This option is only relevant IF you filter by a SNP List AND if the SNP List contains some (but not all) of the available probesets in a gene that is allele translated. If both conditions are met, then it is possible that your SNP List excludes a probeset needed to differentiate among two or more named haplotypes in the translation library file. Selecting this option means that only the first haplotype will be reported from the set of possible haplotypes that are non-distinguishable due to probeset exclusion. The haplotypes are ordered by name from left to right in the translation library file. For example, in gene CYP1A1, the *2C haplotype may be differentiated from the *1 haplotype by a variation in a single probeset. If this probeset is omitted, and the data indicates that both *1 and *2C are possible (due to a NoCall at that probeset), then only *1 is reported as a possibility (since *1 is listed before *2C in the translation library file). The disadvantage of selecting this option is that you may be excluding the actual haplotype for a tested sample. The advantage of selecting this option is that you may want to exclude haplotypes that are differentiable only by probesets you have decided not to translate, AND you agree that the selection of which haplotype to report is correct.

Table 1 Allele Translation options

Select Options	Description
Report combined name that includes all haplotypes that are no longer differentiated	This option is only relevant IF you filter by a SNP List AND if the SNP List contains some (but not all) of the available probesets in a gene that is allele translated. If both conditions are met, then it is possible that your SNP List excludes a probeset needed to differentiate among two or more named haplotypes in the translation library file. Selecting this option means that a combined haplotype name will be reported using the set of possible haplotypes that are non-distinguishable due to probeset exclusion. For example, in gene CYP1A1, the *2C haplotype may be differentiated from the *1 haplotype by a variation in a single probeset. If this probeset is omitted, and the data indicates that both *1 and *2C are possible (due to a NoCall at that probeset), then "*1_or_*2C" is reported as a possibility. The advantage of selecting this option is that you are not excluding possible haplotypes. The disadvantage of selecting this option is that the report will include haplotypes that require a variant allele of a probeset you have decided to exclude for translation. Note: This option is only available if you do not need a phenotype report, as phenotyping requires haplotype names to not change (depending on the set of probesets used for translation). To enable this option that excludes the generation of a phenotype report, you must deselect the usage of the metabolizer library file. To do this, click  (right of the Metabolizer File option).
Include Sample Attributes	Click this check box to include sample attributes in the translation reports.This option is enabled if you have imported sample attributes to your analysis results.

Table 2 More Allele Translation options

Select Translation Results Folder	Description
Output Root Path	The path to the output folder for the translation reports and log. Click its Browse button to set the path.
Export Folder Name	The new folder name for the translation results, whose default name has the date-time format YYYY-MM-DD_HHMMSS_translations. Confirm the default or enter a new folder name.

3. After completing the Allele Translation selections, click **OK**.

   When the translation is finished, an Explorer window appears displaying the folder containing the translation reports and corresponding run log.

Translation Reports

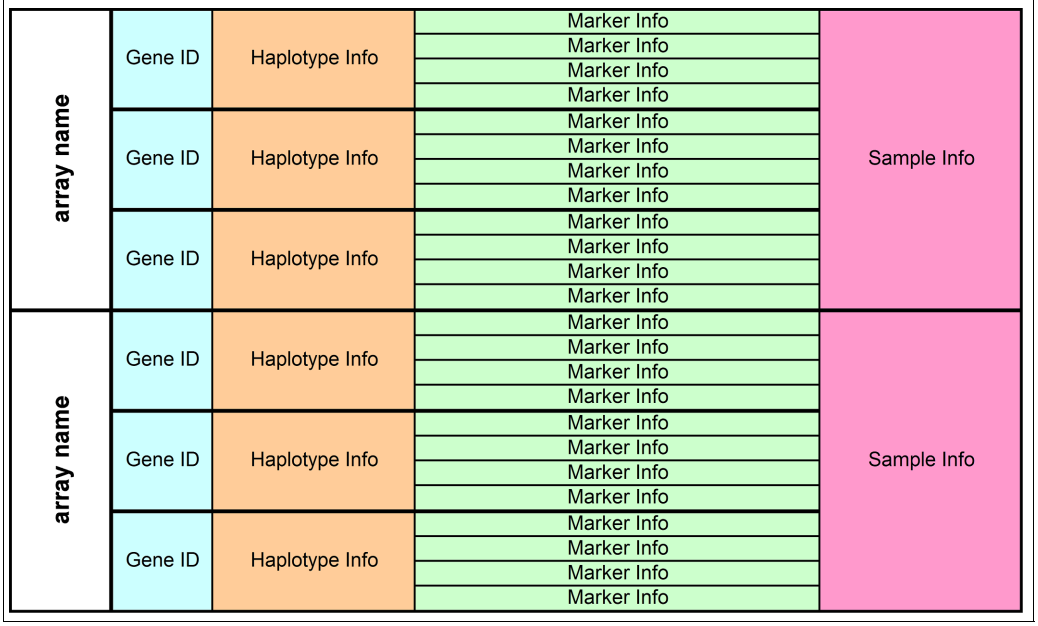
The allele translation operation creates the following reports and supporting files:

- **Comprehensive Translation** - Displays one row per translated SNP for each sample. Provides information on each SNP in addition to haplotype calls.
- **Summary Translation** - An abbreviated version of the Comprehensive report, which displays at least one row for every translated gene for each sample. It also includes rows for every genotype where the translation identifies a variant call. It also includes rows listing SNPs with missing data. In the Summary report only, if no SNPs responsible for functional changes report a variant allele, then information for those SNPs is replaced with a comment to this effect. If a copy number state of zero is indicated, then information for SNPs in that gene is replaced with a comment to this effect, and the copy number haplotype code is reported in the Known Call field.
- **Phenotype Translation** - Displays one row per phenotyped gene for each sample, based on the diplotypes from the source Comprehensive Translation report
- **Uncalled probeset list** - A list of probesets with NoCall genotype calls from SNPs used for translation. This probeset list can then be importing into the Batch, so that you can filter the SNP Summary Table by the uncalled.ps SNP list. This provides a quick way to review the cluster plots of probesets with missing data, and possibly edit the calls directly to "fill in" the missing data in preparation for another round of allele translation reports.
- **Uncalled** - A list of NoCall genotype calls from SNPs used for translation. This report is useful for identifying samples and SNPs for follow up genotyping.
- **MD5** - An electronic signature that can be used to verify that the comprehensive and phenotype reports have not been modified. Interested users can contact devnet@affymetrix.com for information on accessing tools to verify the integrity of the translation results files.
- **Log file** - A list of messages generated by the software as the data is processed. This file is useful for troubleshooting errors.

Comprehensive and Summary Translation Report

The basic layout of this report is shown in Figure 129.

Figure 129 Example: Basic layout of a Comprehensive and Summary Translation report



Summary Translation Report

To make this report easier to read, bold and regular fonts are used. For example, genotype calls are represented in bold, as shown in Figure 130.

Figure 130 Example: Summary Translation report

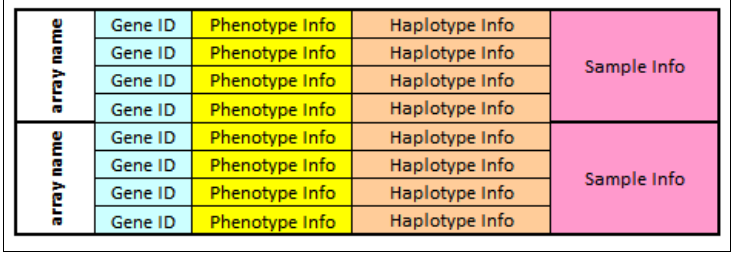
2008-11-05_122550_translations _summary.rpt

Index	CHP Filename	Gene	Known Call	Unknown Call	Interpretation Code	Summary Flag	Relevant Alleles	Common Name	Probe Set ID	Basecall	Reference Base	Variant Base	Call	Haplotype Marker
0001-	TEST_0001.	SLC15A2	*4A/*4A		UNIQ	R503K	*2,*4A,*4B	SLC15A2_34845A>G(R503K	AM_13307	A/A	G	A	Var/Var	Y
0001-	TEST_0001.	UGT1A3	*22/WT		UNIQ	*22-	*22	UGT1A3*22_insT-118	AM_12947	-/T	-	T	Ref/Var	Y
0001-	TEST_0001.	UGT1A3	*22/WT		UNIQ	ATS	ATS	UGT1A3_(rs6753892)	AM_12963	T/G	T	G	Ref/Var	N
0001-	TEST_0001.	UGT2B7	*1/*2E,*2/*5	*2B/UNK,'	MULT+UNK	*2B	*2B,*2E,*5	UGT2B7*2B_-327>(rs7662(	AM_13458	G/A	G	A	Ref/Var	Y
0001-	TEST_0001.	UGT2B7	*1/*2E,*2/*5	*2B/UNK,'	MULT+UNK	*2A	*2A,*2C,*2	UGT2B7*2A_-161>(rs76682	AM_13459	C/T	C	T	Ref/Var	Y
0001-	TEST_0001.	UGT2B7	*1/*2E,*2/*5	*2B/UNK,'	MULT+UNK	*2	*2,*2B,*2D	UGT2B7*2_2100C>T(Y268I	AM_13465	C/T	C	T	Ref/Var	Y
0001-	TEST_0001.	VKORC1	B2/B2		UNIQ	G30R	G30R	VKORC1_1120A>G(G30R)	AM_11032	A/A	G	A	Var/Var	N
0001-	TEST_0001.	VKORC1	B2/B2		UNIQ	B-Hap	B1,B2,B3,I	VKORC1_2255(rs2359612)	AM_11040	C/C	T	C	Var/Var	Y
0001-	TEST_0001.	VKORC1	B2/B2		UNIQ	B-Hap	B1,B2,B3,I	VKORC1_1173(rs9934438)	AM_11045	C/C	T	C	Var/Var	Y
0001-	TEST_0001.	VKORC1	B2/B2		UNIQ	B-Hap	B1,B2,B3,I	VKORC1_(rs9923231)	AM_11054	G/G	A	G	Var/Var	Y
0001-	TEST_0001.	DCK			NoHAP	P122S	P122S	DCK_(P122S)	AM_13645	C/T	C	T	Ref/Var	N
0001-	TEST_0001.	DCK			NoHAP	35708A>G	35708A>G	DCK_35708>(rs4643786)	AM_13646	T/C	T	C	Ref/Var	N
0001-	TEST_0001.	SLCO1A2			NoHAP	T277N	T277N	SLCO1A2_>(T277N)	AM_10528	C/C	A	C	Var/Var	N
0001-	TEST_0001.	UGT1A1		UNK/UNK	UNDH	*60-	*28var,*60	UGT1A1*60_(rs4124874)	AM_13018	G/G	T	G	Var/Var	Y
0001-	TEST_0001.	UGT1A1		UNK/UNK	UNDH	*28-	*28,*28var	UGT1A1*28_(rs34815109)	AM_13024	(TA)7or8	(TA)5or6	(TA)7or	Var/Var	Y
0001-	TEST_0001.	UGT2B15			NoHAP	*2	*2	UGT2B15*2_>(rs1902023)	AM_13439	G/T	G	T	Ref/Var	N
0001-	TEST_0001.	ABCG2	*1/*1		UNIQ	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	CDA	*1C/*3	*1/UNK	UNIQ+UNK	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	CES2	*1/*1		UNIQ	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	CYP2A13	*1A/*1A		UNIQ	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	CYP2A6	*1G/*1G		UNIQ	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	CYP2J2	*1/*1		UNIQ	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	FAAH			NoHAP	All markers responsible for functional changes are Ref/Ref								
0001-	TEST_0001.	FMO3	H1/H2A,H2B/H7		MULT	All markers responsible for functional changes are Ref/Ref								

Phenotype Translation Report

The basic layout of this report is shown in Figure 131.

Figure 131 Example: Phenotype Translation report



Phenotype Report

The basic layout of this report is shown in Figure 132.

Figure 132 Example: Phenotype report

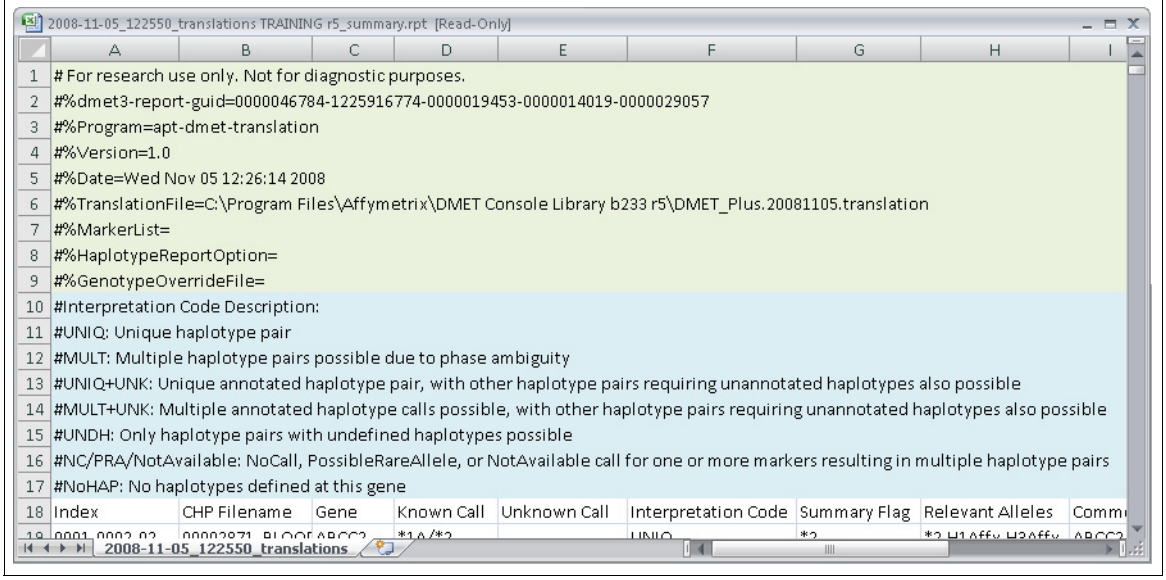
Index	CHP File	Gene	Phenotype Call	Gene Activity	Known Call	Unknown Call	Interpretation Code
0001-0020	test_01.c	CYP1A2	EM	normal/normal	*1F/*1F		UNIQ
0001-0022	test_01.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0001-0024	test_01.c	CYP2B6	EM	normal/normal	*1/*1		UNIQ
0001-0029	test_01.c	CYP2D6	PM	none/none	*5/*5		UNIQ
0002-0020	test_02.c	CYP1A2	EM_or_IM	normal/reduced	*1A/*1L,*1C/*1F		MULT
0002-0022	test_02.c	CYP2A6	EM_or_IM	normal/reduced	*1/*17		UNIQ
0002-0024	test_02.c	CYP2B6	EM_or_IM	normal/reduced	*1/*6	*4/UNK	UNIQ+UNK
0002-0029	test_02.c	CYP2D6	EM_or_IM	normal/reduced	*2/*29	*2/UNK,*29/UNK	NC/PRA/NA
0004-0020	test_04.c	CYP1A2	EM	normal/normal	*1A/*1A		UNIQ
0004-0022	test_04.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0004-0024	test_04.c	CYP2B6	IM	normal/none	*1/*18		UNIQ
0004-0029	test_04.c	CYP2D6	EM	normal/normal	*1/*1		UNIQ
0005-0020	test_05.c	CYP1A2	EM	normal/normal	*1A/*1F		UNIQ
0005-0022	test_05.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0005-0024	test_05.c	CYP2B6	IM	reduced/reduced	*6/*6		UNIQ
0005-0029	test_05.c	CYP2D6	IM	normal/none	*2/*4	*1/UNK,*2/UNK,*4/I	NC/PRA/NA
0009-0020	test_09.c	CYP1A2	EM	normal/normal	*1F/*1F		UNIQ
0009-0022	test_09.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0009-0024	test_09.c	CYP2B6	IM	reduced/reduced	*6/*6		UNIQ
0009-0029	test_09.c	CYP2D6	IM	normal/none	*2/*4	*1/UNK,*10/UNK,UI	UNIQ+UNK

Opening Translation Report in MS Excel

1. Use Windows Explorer as you normally would to navigate to the export folder with the translation results.
2. Double-click the report (.rpt) to be viewed. You may be asked choose an application to open the report. Select Microsoft Excel, then step through the Text Import Wizard (use the tab-delimited default options).

   The report header includes basic information that helps track study data and definitions of interpretation codes, as shown in Figure 133.

Figure 133 Example: Header for Comprehensive and Summary report



Available Report Fields and Descriptions

Array Tracking

Refer to Table 3 for descriptions of the available **Array Tracking** fields.

Table 3 Array Tracking fields

Array Tracking fields	Description
Index	A row index in the format: [filename index]-[gene index within filename]-[Probe Set ID index within gene]. This field can be parsed for sorting or row filtering. For the phenotype report, the index is shortened to [filename index]-[gene index within filename]
Filename	Name of the sample file.

Gene-specific

Refer to Table 4 for descriptions of the available **Gene-specific** fields.

IMPORTANT! Haplotypes are not reported for genes whose Interpretation Code is **NoHap**. The fields described in the table below will therefore be empty for these genes. The exception is if the gene reports a gene deletion, in which case the associated haplotype names are reported.

Table 4 Gene-specific fields

Gene-specific fields	Description
Associated Gene	Gene symbol
Phenotype Call	In the Phenotype report, the predicted phenotype given the supplied Known Call diplotypes. Multiple comma-separated phenotypes are reported when multiple Known Call diplotypes are associated with different phenotypes. Most genes use the following terminology when the default metabolizer library file is selected: • UM = ultra-rapid metabolizer • RM = rapid metabolizer • NM = normal metabolizer • IM = intermediate metabolizer • PM= poor metabolizer Variations on these terms also exist to describe some level of uncertainty: • NM_or_IM = normal or intermediate metabolizer • IM_or_PM =intermediate or poor metabolizer • Not_PM = not a poor metabolizer • unknown = unknown metabolizer state Some genes use different phenotype terms to be consistent with literature usage. Refer to the header of the phenotype report for additional information. Users are responsible for reviewing the *.metabolizer library file for accuracy! Users may modify the *.metabolizer file as needed, and are not restricted to this terminology. Refer to "Diplotype to Phenotype Translation" on page 144 for more information.

Table 4 Gene-specific fields

Gene-specific fields	Description
Gene Activity	In the Phenotype report, the predicted pair of gene activities given the supplied Known Call diplotypes. Multiple comma-separated activity pairs are reported when multiple Known Call diplotypes are associated with different activity pairs. Most genes use the following terminology when the default metabolizer library file is selected: • increased = increased gene function • normal = normal gene function • reduced: reduced gene function • no = no gene function • unknown = unknown or uncertain gene function Some genes use different phenotype terms to be consistent with literature usage. Users are responsible for reviewing the *.metabolizer library file for accuracy! Users may modify the *.metabolizer file as needed, and are not restricted to this terminology. Refer to "Diplotype to Phenotype Translation" on page 144 for more information.
Known Call	Haplotype pairs (diplotypes) identified in the gene of interest. When more than one pair of haplotypes is implicated (due to phase ambiguity in compound heterozygous samples), the reported diplotypes are separated by a comma.

Table 4 Gene-specific fields

Gene-specific fields	Description
Unknown Call	When the gene table includes haplotyping SNPs and a complete diplotype pair cannot be identified in a sample, one or more unknown haplotypes is assumed. This is designated as UNK in the report. Multiple haplotype pairs (diplotypes) that have unknown alleles are separated with a comma in this field. An example record might be in the format: *2/UNK,*13/UNK,*24/UNK,*32/UNK to indicate that there are at least 4 defined alleles consistent with the data, but each would require matching to a haplotype pattern that does not exist in the translation library file.
Interpretation Code	This diplotype interpretation code indicates whether one and only one unique haplotype pair is consistent with the data (UNIQ), whether there are multiple haplotype pairs consistent with the observed genotypes (MULT) and whether these are observed in conjunction with other unknown haplotypes (UNIQ+UNK or MULT+UNK). Additional codes indicate that no known haplotype pairs have been identified (UNDH) or if there is missing data leading to additional haplotype possibilities. The missing data could be NoCall or NotAvailable (NC/PRA/NA). The PossibleRareAllele call is not used by Axiom Analysis Suite. For genes for which copy number state is available, the following interpretation codes may also appear: • CN_HybridLoss = Partial gene deletion is detected, so haplotype pair calling is not available. • CN_HybridGain = Partial gene duplication is detected. If a non-wild-type allele is detected, the software can't determine whether the variant allele is on the partial copy of the gene. • CN_Gain = Gene duplication is detected, but the software cannot determine which haplotype(s) are duplicated. For this reason you won't see xN nomenclature in the reported haplotype pairs. • CN_NoCall = Copy Number state not reported, so there is less confidence in the reported haplotype pairs. • CN_Error = Genotypes from multiple Copy Number states are detected, so haplotype pair calling is not available. This can happen if you manually edit a genotype call in such a way as to change the SNP's copy number state. For example, if you change a "NoCall_1" genotype (a NoCall for a CN=1 sample) to a "BB" CN=2 genotype, its copy number has changed. This would cause allele translation problems if other SNPs for this sample for the same gene have genotypes of a different copy number state.

Marker-specific

Refer to Table 5 for descriptions of the available **Marker-specific** fields.

IMPORTANT! Marker annotations will always be the same throughout the reports for a given SNP. Fields in the Marker Information section of the Comprehensive and Summary reports include the biological information at the SNP level, along with the interpreted genotypes identified in each sample

Table 5 Marker-specific fields

Marker-specific fields	Description
Summary Flag (marker annotation)	This annotation field contains an abbreviated name when structural or functional differences are known to result with mutations at the SNP locus. For triallelic SNPs, there may be more than one flag. For example, ABCB1_c.2677G>T>A(A893SorT) marker is triallelic and can result in different function changes in the protein. The two summary flags for the marker are thus reported "A893S,A893T". This flag is N (No) if the marker is not defining for a haplotype, and its genotype doesn't affect the structure or expression of the gene product. If the flag is N, the marker will not appear in the Summary report unless it reports NoCall.
Relevant Alleles (marker annotation)	This annotation field is the full listing of haplotype-based alleles defined in the gene table that contain the variant version of the marker. For non-haplotype-based SNPs, this is an abbreviated name indicating the protein change that results when the variant base is present.
Common Name (marker annotation)	A SNP identifier describing either the gene location, coding change or dbSNP rsID for the SNP. The Common Name is retrieved from the translation library file, and may not be the same as the Common Name seen within Axiom Analysis Suite (which instead uses the Common Name in the *.annot.db library file)
Probe Set ID	Unique identifier for the SNP.
Basecall	The observed bases, also known as the "raw" genotypes.
Reference Base	This field generally indicates the more common allele in biallelic SNPs. Certain genes use a particular GenBank entry as the "Reference genome" and the observed allele at each marker across the gene is then reported as Reference.
Variant Base	These are the alternate alleles for each SNP. When there is more than one variant allele (e.g. triallelic SNPs) the alternate alleles are reported together and separated by a comma (e.g. A,T).
Call	The first level of translation of the Basecall field, replacing the individual nucleotide calls with the associated reference (Ref) or variant (Var) allele state. For Basecalls associated with copy numbers less than 2, this field will show haplotype names as needed. This field will be empty if the Basecall value has an unclear call.
Haplotype Marker (marker annotation)	Differentiates SNPs used to make haplotype calls or single-marker variant calls. • Y = A flag to indicate that the Allele translation algorithm will match allele variants in blocks of SNPs defining haplotypes in the gene tables. Called haplotypes are reported in the "Known Calls" and Unknown Calls of the report. • N = A flag to indicate that the haplotype background of a variant is not known. Genotyping results for these SNPs are only reported in the "Call" column.
Change for Variant (marker annotation)	Amino acid substitution or other structural change (such as splicing variant, promoter mutation, Frame shift mutation, etc.) caused by the presence of the variant allele.

Table 5 Marker-specific fields

Marker-specific fields	Description
cDNA Change (marker annotation)	Location of the mutation on a reference mRNA sequence.
Genome Position (marker annotation)	This is the chromosomal position of the mutation.
dbSNP RS ID (marker annotation)	The dbSNP identifier for the marker.

Tracking Edited Genotype Calls

Refer to Table 6 for descriptions of the available fields for tracking edited genotype calls.

IMPORTANT! Fields for tracking genotyping changes are recorded in the Change Tracking portion of the translation reports (located immediately before the User Defined Sample Information section of the report).

Table 6 Fields for tracking edited genotype calls

Tracking Edited Genotype Calls	Description
Original Basecall	This field reports the original genotype for the sample. The field is not empty when the user has edited the call within Axiom Analysis Suite.
Override Comment	This field reports edited in AxAS if the call has been edited within Axiom Analysis Suite.

Uncalled Report

Refer to Table 7 for descriptions of the available fields for uncalled reports.

IMPORTANT! The Reference Allele and Variant Allele values of the Uncalled report are the same as the Reference Base and Variant Base values of the Comprehensive and Summary reports.

Table 7 Uncalled Report

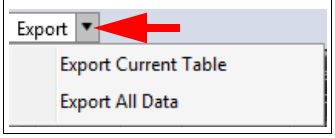
Uncalled Report	Description
Filename	Name of the sample file.
Gene	Gene symbol.
Common Name	The Common Name defines the gene and positional information about the genetic change tested with the probeset.
Basecall	The Uncalled report contains all the NoCall genotypes from SNPs used for allele translation.
Override Comment	User-specified annotation field that enables an audit trail of the source of genotyping results done outside of Axiom Analysis Suite. In the Uncalled report, this field is primarily useful when creating a copy of this file to use as an Override file. The Override file may be useful for Affymetrix Power Tool users.
Reference Allele	Reference base indicates the allele in a reference sequence known to be present at this genetic location. Generally this is the more common allele at SNPs with low minor allele frequency (<1%).
Variant Allele	The variant base(s) defined by the marker are alternative known genotypes known to be present at this genetic location. For triallelic SNPs, the reporting format is: A,T for ABCB1_68883G>T(S893A) because two specific mutations are known to occur at this genomic location (G>A and G>T).

6 Exporting

Using the Sample Table Export Options

1. Click the **Export** drop-down.
   Your Export options appear. (Figure 134)

Figure 134 Sample Table Export Menu

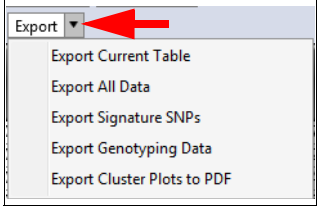


2. Click **Export Current Table** or **Export All Data**.
   A **Save As** window appears.
3. Click on an existing folder or click **New Folder** to choose a new save location.
4. Type a filename for the table, then click **Save**.
   The table data is now saved as a tab-delimited .txt file.

Using the SNP Summary Table Export Options

1. Click the **Export** drop-down.
   The following window appears: (Figure 135)

Figure 135 SNP Summary Table Export menu



Exporting the Current Table

To export the columns and rows currently displayed in the table:

1. Click **Export Current Table**.

   An Explorer window appears.
2. Enter a filename, then click **Save**.

   The current table data is now saved and exported as a tab-delimited text file.

Exporting All Data

To export all columns and rows, including hidden and filtered data:

1. Click **Export All Data**.

   An Explorer window appears.
2. Enter a filename, then click **Save**.

   All data in the table (displayed or not) is saved and exported as a tab-delimited text file.

Exporting Signature SNPs

Use this option to export only the signature SNPs in your data.

1. Click **Export Signature SNPs**.

   An Explorer window appears.
2. Enter a filename, then click **Save**.

   All data is now saved a tab-delimited text file.

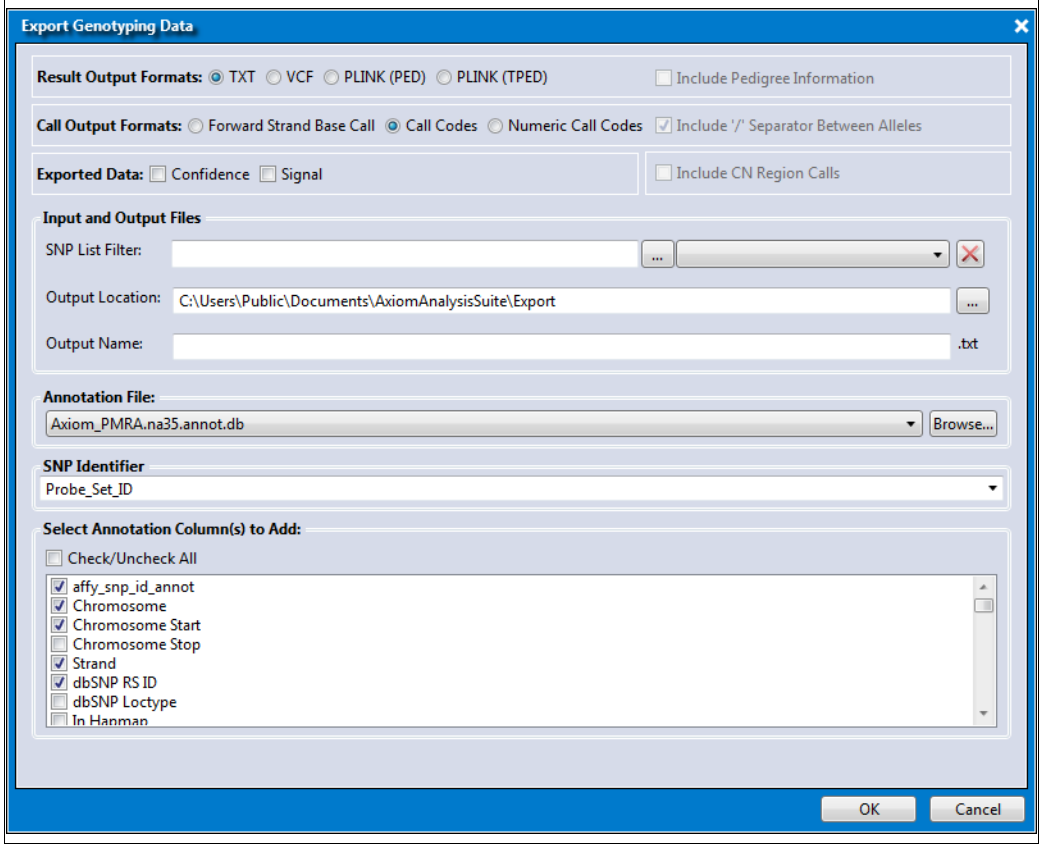
Exporting Genotyping Data

IMPORTANT! Not all options are available and are dependent on the export format you select and its applicable format restrictions.

1. Click **Export Genotyping Data**.

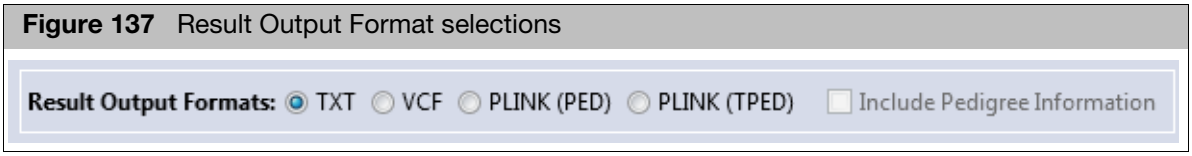
   The following window appears: (Figure 136)

Figure 136 Export Genotype Data window



Result Output Formats

Figure 137 Result Output Format selections



1. Click the radio button to select the Result Output Format you want use.
2. Optional: If you selected a PLINK format, make sure you click the **Include Pedigree Information** check box. Not checking this box may require special handling (within PLINK) to make your exported output work properly.

Note: PLINK format specifies that all markers be biallelic, therefore multi-allele probesets will not be exported to PLINK files.

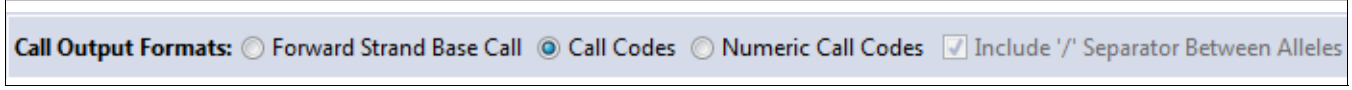
IMPORTANT! If you click the Include Pedigree Information check box with your PLINK format, make sure your Sample Attributes include an Index ID and Pedigree Information (Family ID, Individual ID, Father ID, Mother ID, Sex, and Affection Status).

3. If you have used OTV Caller, some of the calls may be OTV. To export these OTV calls, click the **TXT** and **Call Codes** radio buttons or **TXT** and **Numeric Call Codes** radio buttons. Note: For all other formats, OTV calls are treated as No Calls.

Call Output Formats

1. Click the radio button to select the appropriate Call Output Format (Figure 138) you want to use.

Figure 138 Call Output Format selections



If you select Forward Strand Base Call, you have the option of not including the '/' separator between alleles. Note: Excluding separators is NOT advised when exporting probesets with multi-base alleles.

IMPORTANT! If exporting Numeric Call Codes, the exported data file's header rows includes a table mapping numeric call code to call code. The format of these header rows is:

#%%call-code-N=call_code:numeric_call_code:code_ploidy

The assignment of numeric call codes to call codes is NOT guaranteed to be the same across all supported array types.

Exported Data Selections

1. Click inside the check box(es) to check the additional type(s) of Exported Data (Figure 139) you want to include.

Figure 139 Exported Data selections

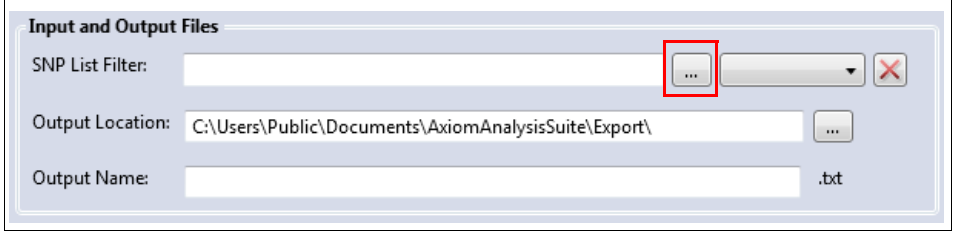
Note: Multi-allele probesets will not export signal values. Normalized signal values for multi-allele and bi-allele probesets are available in the AxiomGT1.summary.txt file inside the batch results folder.

If you selected the VCF output format and copy number analysis was performed, you have the option to not include CN Region Calls from the VCF file.

Input and Output Files

(Figure 140)

Figure 140 Input and Output Files selections



SNP List Filter (Optional)

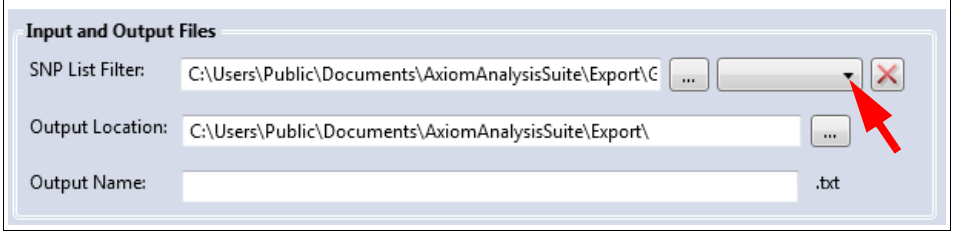
1. To restrict the output to a list of SNPs (probeset_IDs) contained in a file, click the SNP List Filter field's **Browse** button (Figure 140).

   An Explorer window appears.

2. Navigate to the SNP Filter List location, then click **Open**.

   The SNP Filter List field is now populated. (Figure 141)

Figure 141 SNP List Filter and Output path



3. Click the SNP List Filter's drop-down menu to view and select a previously saved SNP List. (Figure 141)

Output Location (Required)

1. Click the Output Location field's **Browse** button.

   An Explorer window appears.

2. Navigate to an output location, create a new folder if needed, then click **Select Folder** button.

   The Output Location path is displayed. (Figure 141)

Output Name (Required)

1. Use the output name already in the Output Name field, or click inside the field to enter a new name. Note: Your output name's file extension reflects the Results Output Format you selected in Step 1.

Changing the SNP Identifier

IMPORTANT! Only SNPs that have a value for the selected annotation are exported. For example, markers that have a dbSNP RSID are exported, while markers without a dbSNP RSID are not exported.

If the selected SNP Identifier has more than 1 probeset mapped to it, it will have multiple entries in the exported file.

1. Click the drop-down arrow, then click to select the SNP Identifier you want to use. (Figure 142)

Figure 142 SNP Identifier

Changing the Current Annotation File (Optional)

1. To change the currently displayed Annotation File, click the Annotation File field's **Browse** button (Figure 143).

   An Explorer window appears.

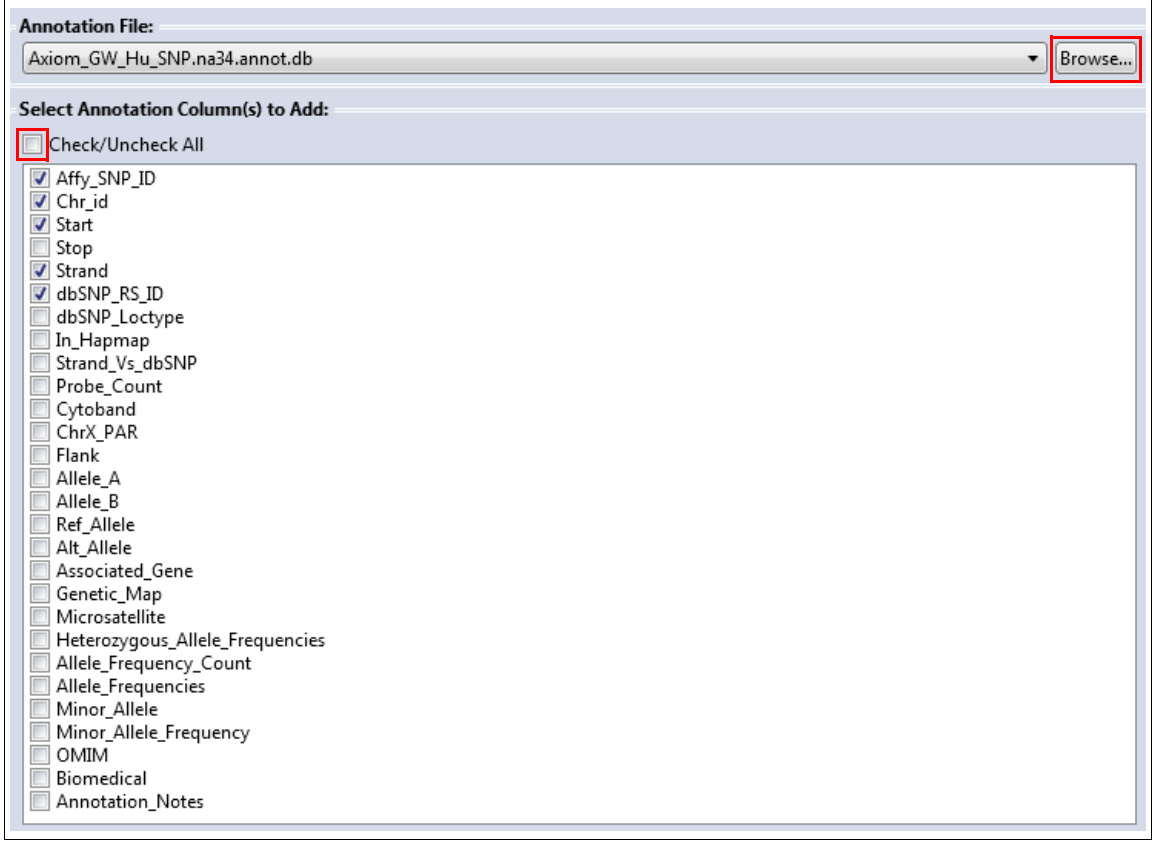
2. Navigate to the appropriate Annotation File location, then click **Open**.

   Your newly selected Annotation file is displayed.

Adding and Removing Annotation Columns

1. Click the check box next to the Annotation Column(s) you want to add to your format results or click to uncheck/remove a column. If you want to add or remove ALL available annotations, click the **Check/Uncheck All** check box. (Figure 143)

Figure 143 Select Annotation Column(s) pane



2. After the Export Genotype Data form is complete, click **OK**.
3. Your newly exported data now reside in the output location you defined in Step 1 on page 115.

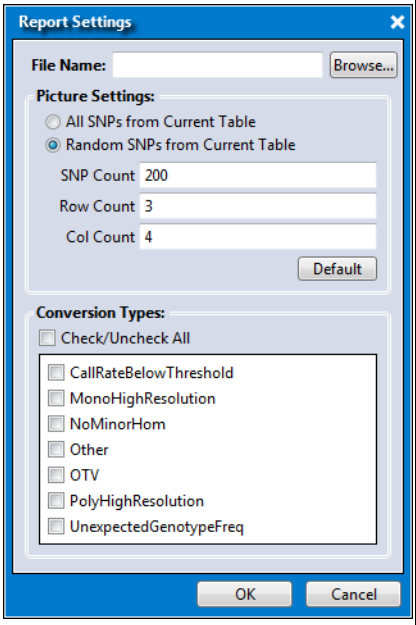
Exporting Cluster Plots to PDF

This exporting option generates a specific number of cluster plots from the selected Conversion Type choices.

1. Click **Export Cluster Plots to PDF**.

   The following window appears: (Figure 144)

Figure 144 Report Settings window



2. Click **Browse**.

   An Explorer window appears.

3. Navigate to a desired location, then enter a name for your PDF report.
4. Click **Save**.

   You are returned to the **Report Settings** window.

5. In the Picture Settings section, click either:
   - All SNPs from Current Table
   - Random SNPs from Current Table
6. Click inside the applicable **Count** field(s) to enter how many cluster pots you want to export.

   Click the **Default** button to revert your modified counts back to their original states.

7. Click to check a Conversion Type(s) you want to add to the report or click the **Check/Uncheck All** check box, then click **OK**.

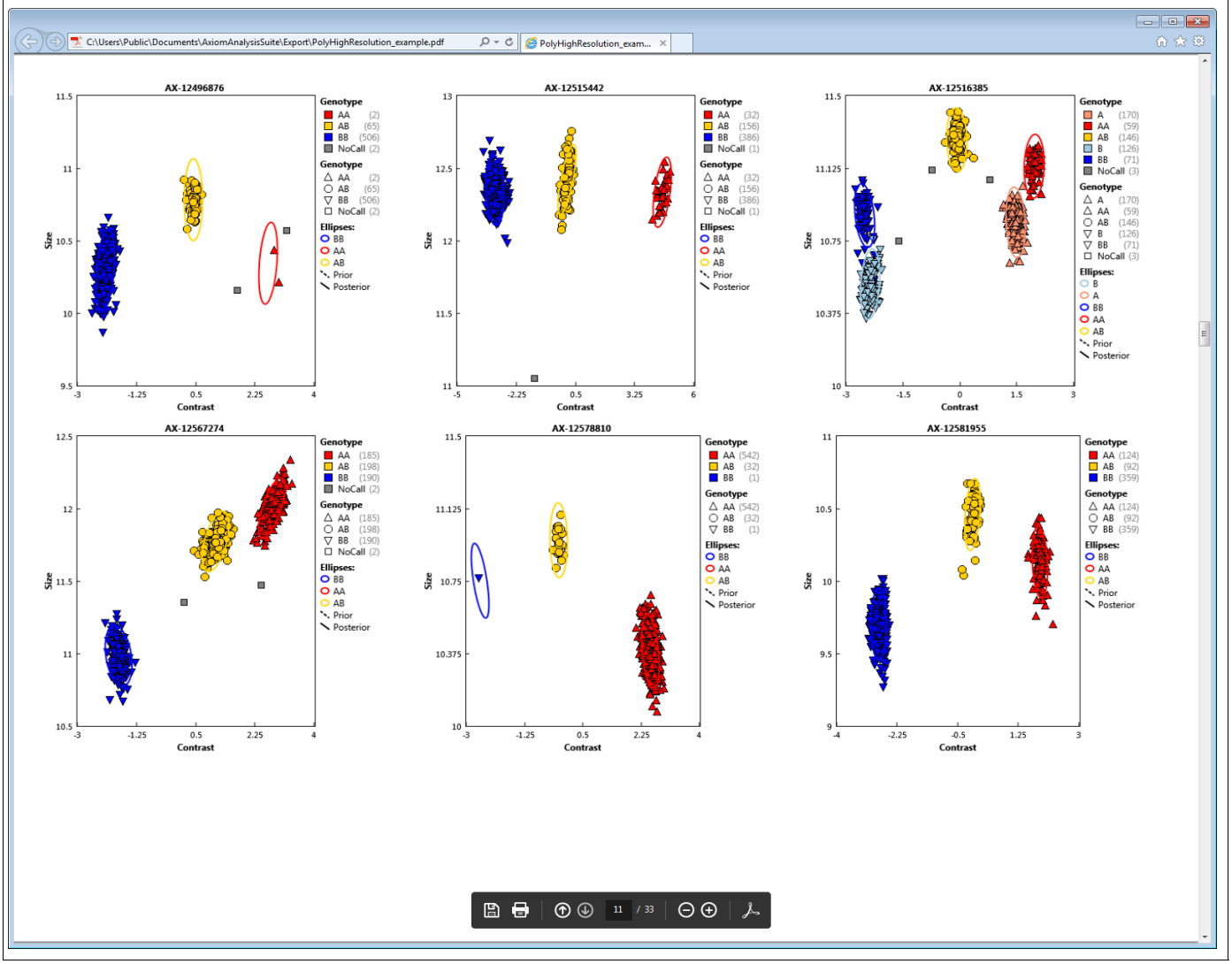
   A Please Wait message and progress bar appear. Allow several minutes if multiple Conversion Type PDF report(s) were selected.

   An Explorer window (where you saved the PDF Report location in Step 3) appears.

8. Double-click on the PDF Report you want to view.

   The PDF Report opens. (Figure 145)

Figure 145 PDF Report example

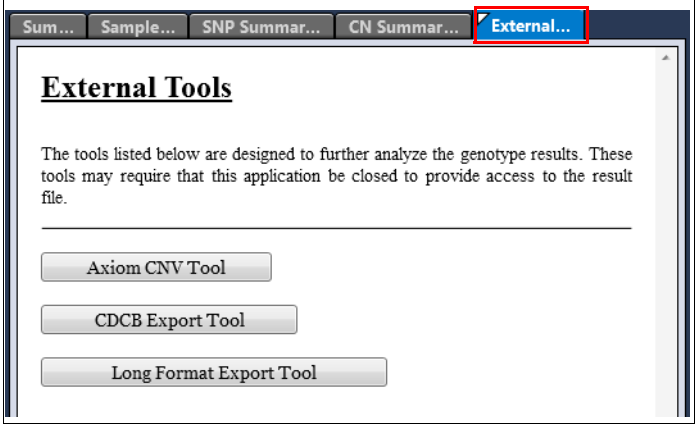


Note: One plot is made for each probeset. For multi-allele probesets, the most informative pair of allele signal channels in the data set is used for plotting, based on the distribution of genotype calls.

7 External Tools

Axiom Analysis Suite auto-detects any previously installed application that may further analyze your genotype results. When a compatible application is detected, an **External Tools** window tab is generated, as shown in Figure 146.

Figure 146 External Tools window tab

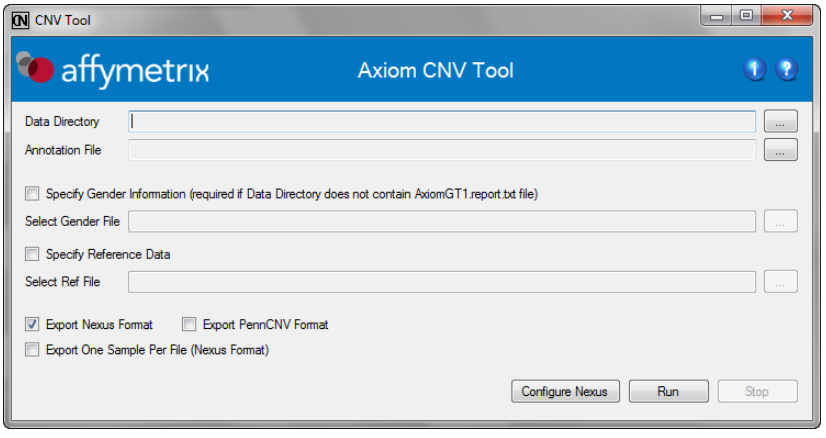


IMPORTANT! The Axiom Analysis Suite can be left open/active if you are using an external application to further view your data. However, you must exit and close the Axiom Analysis Suite before editing its data with an external application.

Axiom CNV Tool 1.1

1. Click the **Axiom CNV Tool** button.
   The application opens. (Figure 147)

Figure 147 CNV Tool - Main window

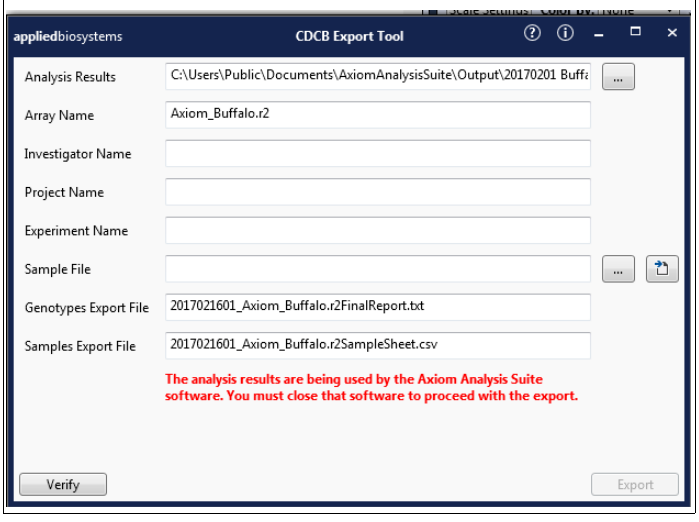


For details on how to use the Axiom CNV Tool 1.1 application, refer to its User Guide (**P/N 703216**).

Council on Dairy Cattle Breeding (CDCB) Export Tool

1. Click the **CDCB Export Tool** button.
   The application opens. (Figure 148)

Figure 148 CDCB Export Tool - Main window

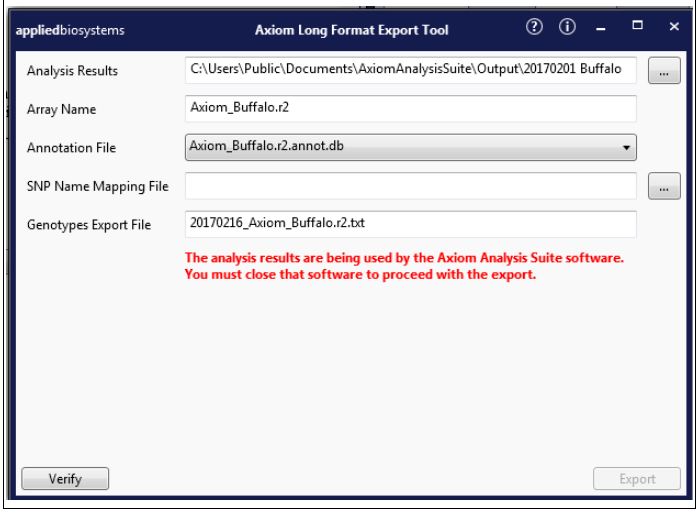


For details on how to use the CDCB Export Tool application, refer to its User Guide (**P/N 703465**).

Axiom Long Format Export Tool

1. Click the **Long Format Export Tool** button.
   The application opens. (Figure 149)

Figure 149 Long Format Export Tool - Main window



For details on how to use the Long Format Export Tool application, refer to its User Guide (**P/N 703455**).

A Predefined Region CN Analysis Variations

Predefined Copy Number Analyses

If the array content and library package supports it, the Genotyping and Best Practices workflows perform predefined region copy number analysis. Copy number analysis is restricted in that the copy number regions are pre-specified, with a fixed set of probesets used to estimate a single copy number state for each region. This form of copy number analysis is done within the Genotyping and Best Practices workflows.

Axiom Analysis Suite supports two variations of restricted copy number analysis. These analysis variations are determined by the array library package and are compared in Figure 150.

Figure 150 Table showing workflow analysis variations that support copy number

Workflow Stage		Analysis Description	Analysis Variation: Typical Axiom	Analysis Variation: Predefined CN Genotyping †	Analysis Variation: CN-Aware Genotyping ‡
Best Practices	Sample QC	Sample QC	yes	yes	yes
		Identify Copy Number (CN) control samples that pass Sample QC	not done	female samples used as CN plate controls	CN controls identified by comparing measured with reference genotype calls
	Genotyping	Normalize plate signals using CN control samples that pass CN QC. Compute CN state in pre-defined regions. Samples that fail CN QC will report a CN state of NoCall.	not done	yes	yes
		Supply CN results to genotyping engine	not done	no	yes
		Final genotyping	yes	yes, including OffTargetVariant calls	yes, using additional call codes for haploid and zero CN calls

† library package specifies analysis_category = "CN_GT_2"

‡ library package specifies analysis_category = "CN_GT"

With Predefined CN Genotyping

- CN states will be reported for the predefined regions specified by the library package
- Unless overridden by the user, the control samples used for per-plate signal correction for CN analysis are selected from the set of samples determined to be female and passing sample QC

- Unless overridden by custom library file settings, genotyping probesets will report diploid genotype codes (such as "AA", "AB", "BB", "NoCall"), whether or not they fall within CN regions reporting deletions.
- OTV Caller is automatically run if any SNPs are assigned to the Conversion Type "OTV". Some of the resulting genotype calls may be "OTV", indicating a detected Off Target Variant.

With CN-aware Genotyping

- CN states will be reported for the predefined regions specified by the library package
- Unless overridden by the user, the control samples used for per-plate signal correction for CN analysis are identified by high concordance between measured and reference genotypes for expected control samples, and should also pass sample QC
- SNPs in a CN=0 region will be called as "ZeroCN"
- SNPs in a CN=1 region will be called with haploid genotype codes such as "A", "B", "NoCall_1"
- SNPs in a CN=2 or higher region will continue to be called with diploid genotype codes such as "AA", "AB", "BB", "NoCall"
- Special SNPs (those on chromosomes Y, MT, and non-PAR X) will also be assigned hemizygous and ZeroCN genotype codes when expected copy number is less than two
- CN-aware genotyping can be applied to probesets between two measured CN regions. For example, if two adjacent regions predict the same CN state, genotyping probesets between those two regions can be assigned genotype call codes consistent with that CN state.

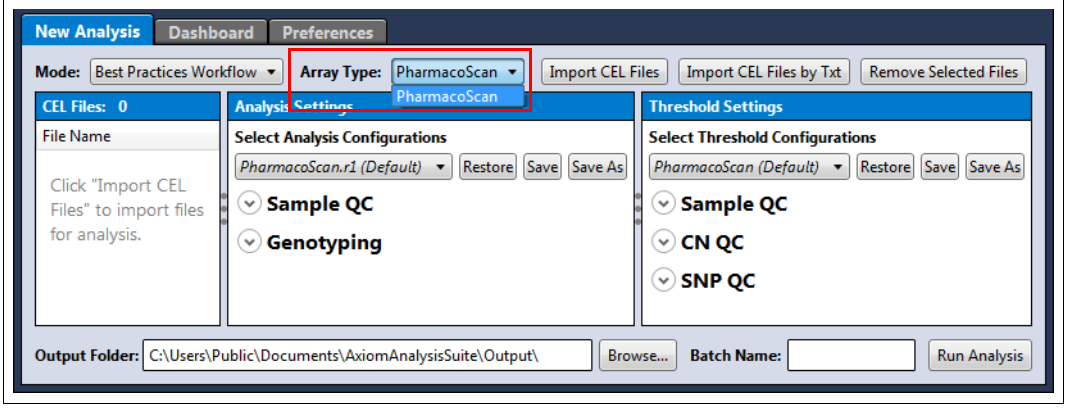
Note: Examples of library packages that support these copy number analyses include; CarrierScan for Predefined CN genotyping, and PharmacoScan for CN-aware genotyping.

Genotyping with CN Analysis

Setting Up a Genotyping with CN Analysis

Before setting up a genotyping with predefined CN analysis, click the New Analysis window tab, then click the Array Type drop-down menu to confirm an array such as PharmacoScan or CarrierScan is available, as shown in Figure 151. If it is, continue to "Selecting a Mode (Workflow)".

Figure 151 Preferences window tab



Selecting a Mode (Workflow)

From the main Axiom Analysis Suite window tab, click the **Mode** drop-down.

1. Select **Best Practices Workflow**.

   Best Practices Workflow performs and combines the Sample QC Workflow and Genotyping Workflow. For more information, see "Overview and Use of the Best Practices Workflow" on page 137.

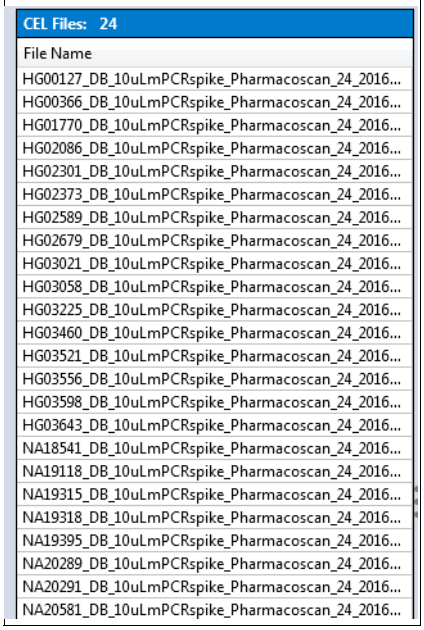
Importing CEL Files

1. Click **Import CEL Files**.

   The **Add CEL Files** window appears.
2. Navigate to your CEL file location.
3. Single-click on a CEL file or Ctrl click, Shift click, or press Ctrl A (to select multiple files).
4. Click **Open**.

The CEL Files pane populates and displays your selected CEL files. (Figure 152)

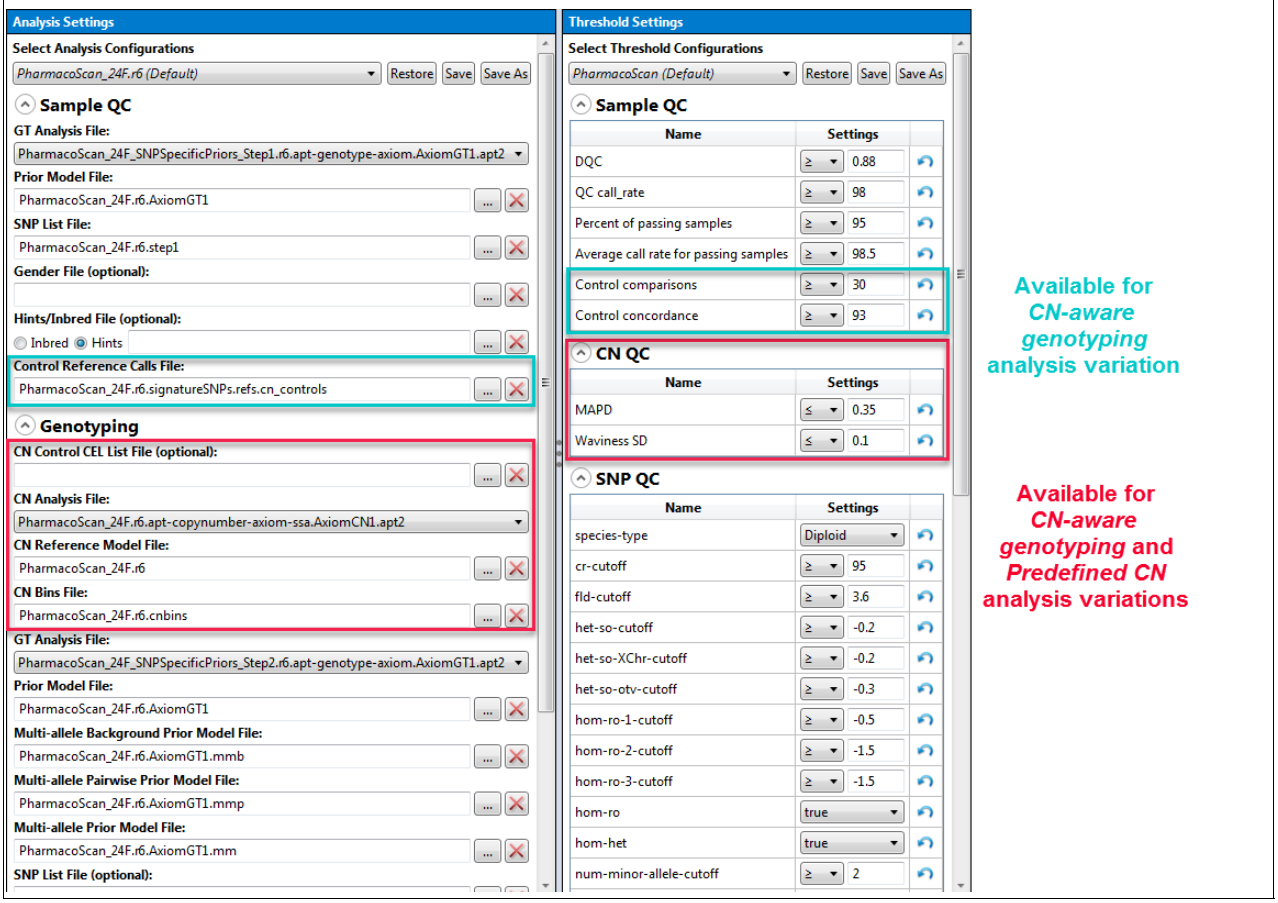
Figure 152 Cell File pane



After importing your CEL files, the Analysis and Threshold Setting panes auto-populate with default settings and additional user-configurable fields and settings specific to the library package, as shown in Figure 153.

For information on typical Axiom array analysis and threshold fields, see "Setting Up an Analysis" on page 19 and "Setting Up Threshold Settings" on page 28.

Figure 153 Example: Auto-populated Analysis and Threshold Setting panes



Analysis Settings with Copy Number Options

IMPORTANT! Only experienced users should modify default analysis settings.

Sample QC

- **GT Analysis File**: Parameters file for the genotyping step that calculates QC Call Rate.
- **Prior Model File**: Defines prior knowledge of SNP cluster locations. This file has the same format as a posteriors file, which is generated by the genotyping step. This means that you can "train" on a custom data set, and use the updated knowledge of cluster locations as a "seed" to possibly improve future genotyping batches. This file must contain two row entries for the GENERIC and GENERIC:1 probesets (if there are any probesets to be genotyped that are not listed in this file).
- **SNP List File**: A file of probeset IDs to genotype. For Sample QC it defines the probesets used to calculate QC Call Rate.
- **Gender File**: A file specifying the desired gender of every sample. If supplied, software will use values in this file instead of the computed gender. Gender impacts genotyping of chromosome X and Y SNPs.
- **Hints/Inbred File**: If a hints file, a file of expected genotype calls. This is used to influence the predicted cluster locations, which influences the final calls. In the Sample QC section, it influences the QC Call Rate.
- **Control Reference Calls File**: For CN-aware genotyping only, a recommended file containing the expected signature SNP calls of the CN control samples and used by Sample QC to identify the control samples among the supplied CEL files.

Genotyping

- **CN Control CEL List File**: This optional file identifies the CEL files that are to be used for per-plate tuning of CN signals. The file has the same format as file **[Batch Name]\CNData\CNcontrolSamples.pass.txt**, which is generated by a Sample QC step. The following table describes how the Workflow and this input option interact to select the controls used for CN analysis.

Workflow	CN Control CEL List File Input	Chosen Controls for CN Analysis
Best Practices	None	Passing controls identified by Sample QC step
Genotyping	None	No controls, therefore no plate signal tuning is done.
Best Practices	User-supplied	User-supplied controls
Genotyping	User-supplied	User-supplied controls

- **CN Analysis File**: Parameters file for the CN Analysis step.
- **CN Reference Model File**: Reference information for CN Analysis step.

- **CN Bins File**: Specifies for each predefined CN Region the callable CN states, and the MedianLog2Ratio boundaries defining each state.
- **GT Analysis File**: Parameters file for the final genotyping step.
- **Prior Model File**: Defines prior knowledge of bi-allelic SNP cluster locations. This file has the same format as a posteriors file, which is generated by the genotyping step. This means that you can "train" on a custom data set, and use the updated knowledge of cluster locations as a "seed" to possibly improve future genotyping batches. This file must contain two row entries for the GENERIC and GENERIC:1 probesets (if there are any probesets to be genotyped that are not listed in this file).
- If the library package supports multi-allele SNPs, then there are three additional model files for the three stages of multi-allele genotyping: Multi-allele Background Prior Model File, Multi-allele Pairwise Prior Model File, and Multi-allele Prior Model File.
- **SNP List File**: A file of probeset IDs to genotype. For Genotyping it defines the probesets for which genotypes will be reported.
- **Gender File**: A file specifying the known gender of every sample. If supplied, software will use values in this file instead of the computed gender. Gender impacts genotyping of chromosome X and Y SNPs.
- **Hints/Inbred File**: If a hints file, a file of expected genotype calls. This is used to influence the predicted cluster locations, which influences the final calls.
- **Posterior File Name**: The desired file output of a genotyping analysis, specifying identified SNP probeset cluster locations, variance, and relative weight. One use of this file is as a prior model file for future genotype analyses.
- **ps2snp File**: If multiple probeset designs exist on the array for a given SNP (for example, one forward and one reverse strand design), then the ps2snp file is used by the SNP classification step to identify the best performing probeset for the SNP, using the priority-order setting in the SNP QC section in the New Analysis tab. This text file has two tab delimited columns with the headers probeset_id and snpid (**snpid = affy_snp_id**).
- **Genotype Frequency File**: If the library package supports a check for unexpectedly high call frequency for specific genotypes, this optional file specifies the maximum expected frequency for reviewed genotypes.

Threshold Configurations with CN Options

IMPORTANT! Only experienced users should modify default threshold settings.

Sample QC

Control Comparisons (CN-aware genotyping only): For identifying control samples, this is the minimum number of SignatureSNP probesets compared to a reference.

Control Concordance (CN-aware genotyping only): For identifying control samples, this is the minimum percent concordance of SignatureSNP calls to a reference.

CN QC

MAPD: Median Absolute Pairwise Difference of log2ratio signals of adjacent copy number (CN) probesets must be below this value to make CN calls.

Waviness SD: Waviness Standard Deviation of log2 ratio signals of copy number (CN) probesets must be below this value to make CN calls.

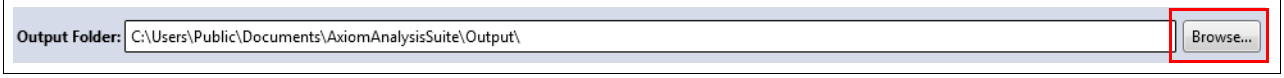
SNP QC

Refer to Table 11 on page 149 for SNP QC Threshold name definitions.

Assigning an Output Folder Path

1. Click the **Output Folder** path's **Browse** button. (Figure 154)

Figure 154 Output Folder field



An Explorer window appears.

2. Navigate to the recommended path **C:\Users\Public\Documents\AxiomAnalysisSuite\Output**, then click **Select Folder**.

Your selected output folder path is now displayed.

Assigning a Batch Name

The batch file is produced while your analysis is running and includes all the necessary files needed to view your analysis in the Axiom Analysis Suite Viewer.

1. Enter a name in the **Batch Name** field. (Figure 155)

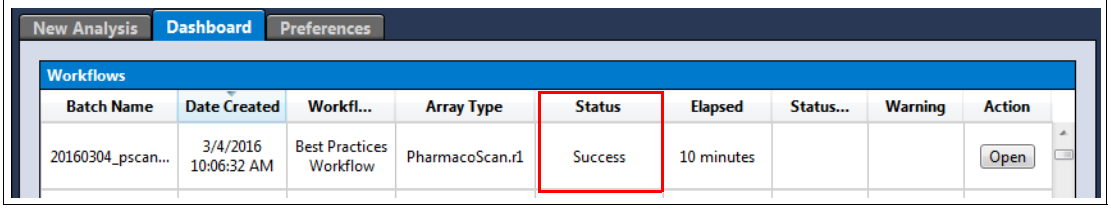
Figure 155 Enter a Batch Name

Note: A folder (with the same name as your entered batch name) is auto-generated during the analysis process. This folder includes all the necessary files needed to view your analysis results in the Viewer.

Running your Genotyping with CN Analysis

1. Click Run Analysis.
   The Dashboard window tab appears. (Figure 156)

Figure 156 Dashboard window tab



After **Success** is displayed in the Status column (Figure 156), click Open.

The Open button will not appear if a processing error occurs. If a message appears stating that no samples passed QC, click the **Open Selected Result(s)** button to review the analysis result that may need troubleshooting.

The Axiom Analysis Suite Viewer opens in a new window (Figure 157) and displays your completed analysis results.

Note: To make comparisons between your completed data analyses easier, open additional Viewer windows. To do this, click the **Dashboard** window tab, then click on the **Open** button again.

Viewing your Genotyping and CN Results

After processing arrays that support predefined CN or CN-aware genotyping, the following is added to the Axiom Analysis Suite Viewer:

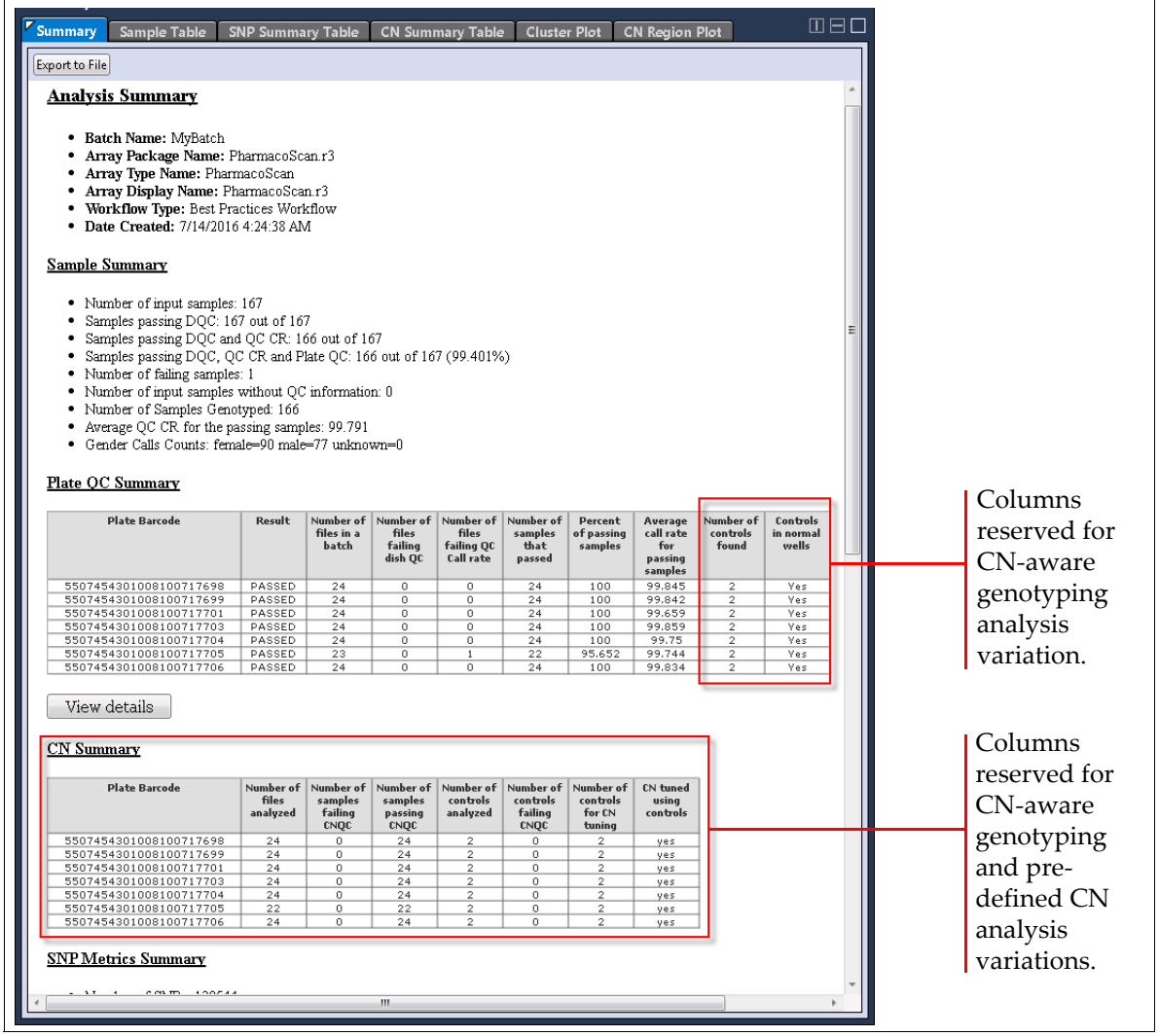
- New entries in the Summary Report, Sample Table, and SNP Summary Table.
- For CN-aware genotyping, the Cluster Plot also displays hemizygous and ZeroCN clusters for SNPs in CN regions.
- CN Summary Table, with Export CN Data option.
- A CN Region Plot window tab.
- Exported VCF formats can now include CN and SNP calls.

Summary Report

After successfully running a Sample QC or Best Practices workflow, the Summary report's **Plate QC Summary** section features two additional fields for the CN-aware genotyping analysis variation, as shown in Figure 157.

- **Number of controls found** - This metric is a count of samples identified as controls based on high signature SNP concordance to expected calls in the Control Reference Calls library file.
- **Controls in normal wells** - The Control Reference Calls library file lists the expected plate wells for the control samples. This metric will report "Yes" if all the detected control samples are in the expected plate wells. It is OK to put the control samples in any wells you choose.

Figure 157 Viewer window tabs



After successfully running a Genotyping or Best Practices workflow, the Summary report's **CN Summary** section features a table of information for each plate, as shown in Figure 157.

Refer to Table 8 for descriptions of each CN Summary Report column.

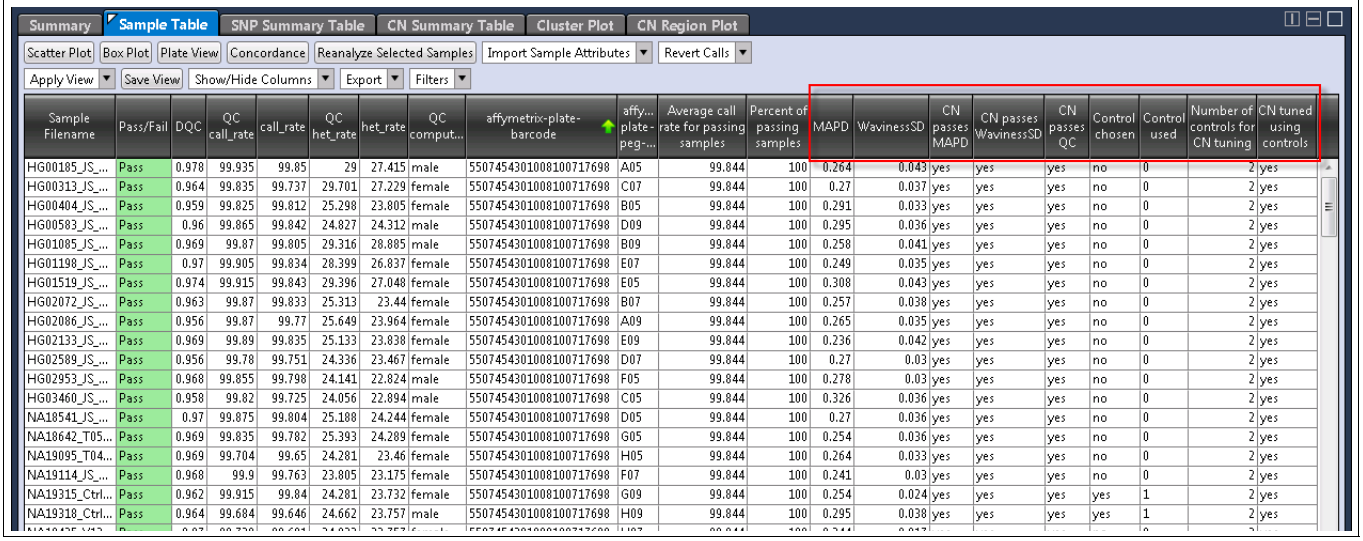
Table 8 CN Summary Report

CN Summary Report	Description
Number of files analyzed	Count of CEL files supplied to the Genotyping step (which includes CN analysis).
Number of samples failing CNQC	Count of CEL files that fail copy number quality control checks, which is the result of a comparison of measured MAPD and waviness SD metrics against thresholds. Samples that fail CN QC report a CN state of NoCall. For the CN-aware genotyping analysis variation, SNPs from samples that fail CN QC are called without the benefit of CN-aware genotyping.
Number of samples passing CNQC	Count of CEL files that pass copy number quality control checks.
Number of controls analyzed	Count of CEL files supplied to the Genotyping step that are identified as controls.
Number of controls failing CNQC	Count of control CEL files that fail copy number quality control checks, which is the result of a comparison of measured MAPD and waviness SD metrics against thresholds. Controls that fail CN QC report a CN state of NoCall. For the CN-aware genotyping analysis variation, SNPs from samples that fail CN QC are called without the benefit of CN-aware genotyping. Note: For the CN-aware genotyping analysis variation, Non-PAR X, Y, and mitochondrial SNPs use gender and other information to report non-diploid calls, regardless of CN QC status.
Number of controls for CN tuning	Count of control CEL files that pass copy number quality control checks.
CN tuned using controls	If Number of controls for CN tuning > 0, then this value is "Yes". This means that CN signals are adjusted based on the signal measured in the control samples from the same plate. If the controls are appropriate for the samples being processed, this adjustment can improve CN calling accuracy.

Sample Table

After successfully running a CN analysis, additional columns appear within the Sample Table, as shown in Figure 158.

Figure 158 Sample Table after a CN Analysis run



Sample Filename	Pass/Fail	DQC	QC call_rate	call_rate	QC het_rate	het_rate	QC comput...	affymetrix-plate-barcode	affy... plate-peg-...	Average call rate for passing samples	Percent of passing samples	MAPD	WavinessSD	CN passes MAPD	CN passes WavinessSD	CN passes QC	Control chosen	Control used	Number of controls for CN tuning	CN tuned using controls
HG00185_JS_...	Pass	0.978	99.935	99.85	29	27.415	male	550745430100810071769 8	A05	99.844	100	0.264	0.043	yes	yes	yes	no	0	2	yes
HG00313_JS_...	Pass	0.964	99.835	99.737	29.701	27.229	female	550745430100810071769 8	C07	99.844	100	0.27	0.037	yes	yes	yes	no	0	2	yes
HG00404_JS_...	Pass	0.959	99.825	99.812	25.298	23.805	female	550745430100810071769 8	B05	99.844	100	0.291	0.033	yes	yes	yes	no	0	2	yes
HG00583_JS_...	Pass	0.96	99.865	99.842	24.827	24.312	male	550745430100810071769 8	D09	99.844	100	0.295	0.036	yes	yes	yes	no	0	2	yes
HG01085_JS_...	Pass	0.969	99.87	99.805	29.316	28.885	male	550745430100810071769 8	B09	99.844	100	0.258	0.041	yes	yes	yes	no	0	2	yes
HG01198_JS_...	Pass	0.97	99.905	99.834	28.399	26.837	female	550745430100810071769 8	E07	99.844	100	0.249	0.035	yes	yes	yes	no	0	2	yes
HG01519_JS_...	Pass	0.974	99.915	99.843	29.396	27.048	female	550745430100810071769 8	E05	99.844	100	0.308	0.043	yes	yes	yes	no	0	2	yes
HG02072_JS_...	Pass	0.963	99.87	99.833	25.313	23.44	female	550745430100810071769 8	B07	99.844	100	0.257	0.038	yes	yes	yes	no	0	2	yes
HG02086_JS_...	Pass	0.956	99.87	99.77	25.649	23.964	female	550745430100810071769 8	A09	99.844	100	0.265	0.035	yes	yes	yes	no	0	2	yes
HG02133_JS_...	Pass	0.969	99.89	99.835	25.133	23.838	female	550745430100810071769 8	E09	99.844	100	0.236	0.042	yes	yes	yes	no	0	2	yes
HG02589_JS_...	Pass	0.956	99.78	99.751	24.336	23.467	female	550745430100810071769 8	D07	99.844	100	0.27	0.03	yes	yes	yes	no	0	2	yes
HG02953_JS_...	Pass	0.968	99.855	99.798	24.141	22.824	male	550745430100810071769 8	F05	99.844	100	0.278	0.03	yes	yes	yes	no	0	2	yes
HG03460_JS_...	Pass	0.958	99.82	99.725	24.056	22.894	male	550745430100810071769 8	C05	99.844	100	0.326	0.036	yes	yes	yes	no	0	2	yes
NA18541_JS_...	Pass	0.97	99.875	99.804	25.188	24.244	female	550745430100810071769 8	D05	99.844	100	0.27	0.036	yes	yes	yes	no	0	2	yes
NA18642_T05...	Pass	0.969	99.835	99.782	25.393	24.289	female	550745430100810071769 8	G05	99.844	100	0.254	0.036	yes	yes	yes	no	0	2	yes
NA19095_T04...	Pass	0.969	99.704	99.65	24.281	23.46	female	550745430100810071769 8	H05	99.844	100	0.264	0.033	yes	yes	yes	no	0	2	yes
NA19114_JS_...	Pass	0.968	99.9	99.763	23.805	23.175	female	550745430100810071769 8	F07	99.844	100	0.241	0.03	yes	yes	yes	no	0	2	yes
NA19315_Ctrl...	Pass	0.962	99.915	99.84	24.281	23.732	female	550745430100810071769 8	G09	99.844	100	0.254	0.024	yes	yes	yes	yes	1	2	yes
NA19318_Ctrl...	Pass	0.964	99.684	99.646	24.662	23.757	male	550745430100810071769 8	H09	99.844	100	0.295	0.038	yes	yes	yes	yes	1	2	yes

Refer to Table 9 for descriptions of each added Sample Table column.

Table 9 Added Sample Table columns after running a CN Analysis

Sample Table	Description
MAPD	Median Absolute Pairwise Difference of log2ratio signals of adjacent copy number (CN) probesets.
WavinessSD	Waviness Standard Deviation of log2ratio signals of adjacent copy number (CN) probesets
CN passes MAPD	"Yes" if the sample's MAPD value is not greater than the MAPD threshold used by CN QC.
CN passes WavinessSD	"Yes" if the sample's WavinessSD value is not greater than the WavinessSD threshold used by CN QC
CN passes QC	"Yes" if the sample passes both MAPD and WavinessSD threshold tests. If "no", the sample reports a CN state of NoCall. For the CN-aware genotyping analysis variation, SNPs from samples that fail CN QC are called without the benefit of CN-aware genotyping. Note: For the CN-aware genotyping analysis variation, Non-PAR X, Y, and mitochondrial SNPs use gender and other information to report non-diploid calls, regardless of CN QC status.
Control chosen	"Yes" if the control was listed in the CN Control CEL List file supplied for CN analysis.
Control used	"1" if the control was used to adjust CN signals of all samples in the same plate. The value will be 1 if Control chosen = 1 and CN passes QC = Yes.

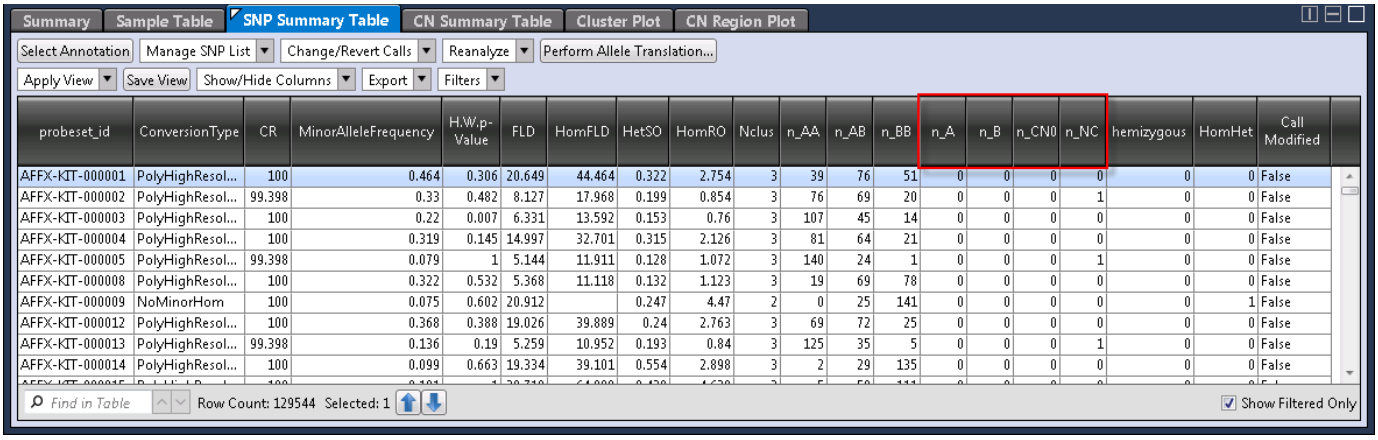
Table 9 Added Sample Table columns after running a CN Analysis

Sample Table	Description
Number of controls for CN tuning	A count of control samples used to adjust the CN signals for the given sample. This value will be the same for all samples on the same plate.
CN tuned using controls	If Number of controls for CN tuning > 0, then this value is "Yes". This means that CN signals are adjusted based on the signal measured in the control samples from the same plate. If the controls are appropriate for the samples being processed, this adjustment can improve CN calling accuracy.

SNP Summary Table

After successfully running a CN-aware genotyping analysis, additional columns appear within the SNP Summary Table, as shown in Figure 159.

Figure 159 SNP Summary Table after a CN Analysis run



probeset_id	ConversionType	CR	MinorAlleleFrequency	H.W.p-Value	FLD	HomFLD	HetSO	HomRO	Nclus	n_AA	n_AB	n_BB	n_A	n_B	n_CN0	n_NC	hemizygous	HomHet	Call Modified
AFFX-KIT-000001	PolyHighResol...	100	0.464	0.306	20.649	44.464	0.322	2.754	3	39	76	51	0	0	0	0	0	0	False
AFFX-KIT-000002	PolyHighResol...	99.398	0.33	0.482	8.127	17.968	0.199	0.854	3	76	69	20	0	0	0	1	0	0	False
AFFX-KIT-000003	PolyHighResol...	100	0.22	0.007	6.331	13.592	0.153	0.76	3	107	45	14	0	0	0	0	0	0	False
AFFX-KIT-000004	PolyHighResol...	100	0.319	0.145	14.997	32.701	0.315	2.126	3	81	64	21	0	0	0	0	0	0	False
AFFX-KIT-000005	PolyHighResol...	99.398	0.079	1	5.144	11.911	0.128	1.072	3	140	24	1	0	0	0	1	0	0	False
AFFX-KIT-000008	PolyHighResol...	100	0.322	0.532	5.368	11.118	0.132	1.123	3	19	69	78	0	0	0	0	0	0	False
AFFX-KIT-000009	NoMinorHom	100	0.075	0.602	20.912		0.247	4.47	2	0	25	141	0	0	0	0	0	1	False
AFFX-KIT-000012	PolyHighResol...	100	0.368	0.388	19.026	39.889	0.24	2.763	3	69	72	25	0	0	0	0	0	0	False
AFFX-KIT-000013	PolyHighResol...	99.398	0.136	0.19	5.259	10.952	0.193	0.84	3	125	35	5	0	0	0	1	0	0	False
AFFX-KIT-000014	PolyHighResol...	100	0.099	0.663	19.334	39.101	0.554	2.898	3	2	29	135	0	0	0	0	0	0	False

Refer to Table 10 for descriptions of each added SNP Summary Table column.

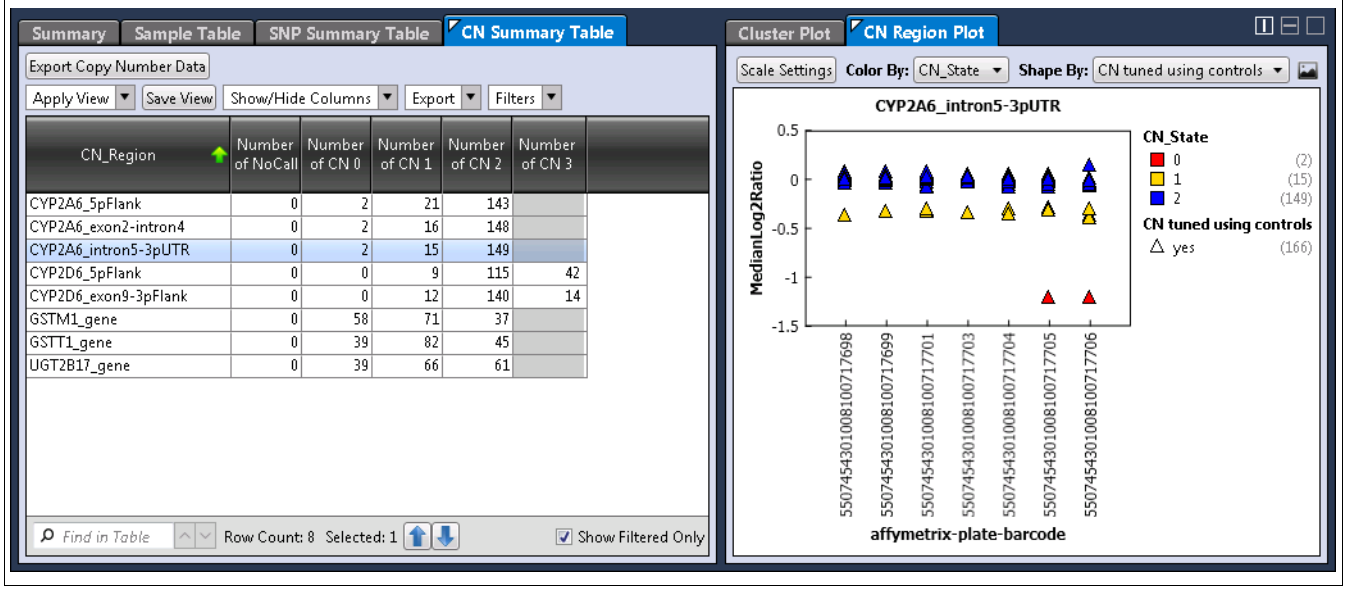
Table 10 Added SNP Summary Table columns after CN-aware genotyping is performed

SNP Summary Table	Description
n_A	Count of hemizygous "A" calls.
n_B	Count of hemizygous "B" calls.
n_CN0	Count of ZeroCN calls.
CopyNumIssue (hidden by default)	Copy Number Issue is "1" if the probeset is assigned to the "Other" ConversionType because the haploid or ZeroCN clusters are in an unusual location compared to each other or to the diploid clusters.

CN Summary Table and CN Region Plot

The Genotyping and Best Practices workflows perform predefined region copy number analysis (if the array content and library package support it). If this is the case, a **CN Summary Table** and **CN Region Plot** window tab appear in the Axiom Analysis Suite Viewer, as shown in Figure 160.

Figure 160 CN Summary Table and CN Plot window tabs



CN Summary Table (Overview)

- The CN Summary Table gives a count of samples called with a given copy number state (CN_State) for a given copy number region (CN_Region).
- CN_States that cannot be reported are displayed as empty with a gray background. If the CN_State can be reported but no samples in the batch report that CN State, then the count is 0. As long as the default CN Bins library file is used during analysis setup, the only samples that will report a CN_State of NoCall are those that fail Copy Number QC.
- Selecting a specific CN_Region row in the table updates the associated CN Region Plot.
- The button **Export Copy Number Data** copies several CN reports to another location specified by the user. Note that CN_States for each CN_Region for each sample are also available in a VCF format export of genotyping data, available from the SNP Summary Table tab.

CN Region Plot (Overview)

- The CN Region Plot displays the MedianLog2Ratio of each genotyped sample for the copy number region selected in the CN Summary Table, grouped by affymetrix-plate-barcode.
- Selecting a sample or samples in the CN Region Plot also selects the same samples in the Sample Table and the Cluster Plot, and vice versa.
- Genes with a CN of 2 typically have a MedianLog2Ratio centered at 0, which indicates that the signals of probesets used for CN estimation are similar to the expected signals for a CN=2 sample.
- CN States are called by comparing the MedianLog2Ratio to the thresholds defined in the CN Bins library file selected during analysis setup. CN States cannot be edited in the Viewer.
- Sometimes between-plate assay variation can shift the observed MedianLog2Ratio values away from 0 for CN=2 samples. Fortunately, if the assay includes appropriate control samples on the assay plate, then the copy number signals can be tuned using the control samples. If CN plate correction was done, then the default "Shape By" metric "CN tuned using controls" identifies which samples had their CN measurements adjusted using the plate controls.

Overview and Use of the Best Practices Workflow

The recommended genotyping method is to use information from a batch of samples to improve the calling of individual samples. To this end, the cluster locations in signal space for each probeset (see the Cluster Plot) adapt in a Bayesian fashion to the supplied data. Prior knowledge of cluster locations influences the final calls. The more samples that are supplied in a batch, the more the final cluster locations will be influenced by the supplied data.

The dynamic nature of this genotyping algorithm means that if a given sample's CEL file is genotyped in a group of 24 CEL files, or in a group of 2400 CEL files, you can expect that some SNPs may experience a changed call (call <-> NoCall, or sometimes call 1 <-> call 2). This is more likely to happen if the SNP's data quality is not great (close clusters), if there are strong between-plate effects in cluster positions, or for very low Minor Allele Frequency SNPs where good information on rare genotype cluster positions may not be available before genotyping. Calling accuracy of both rare and common SNPs is improved even for small batch sizes if enough prior knowledge of cluster locations is available.

The batch nature of the genotyping means that if there are some samples of poor data quality in a group of good quality samples, then sometimes the poor quality samples will harm the calling accuracy of the good samples. For this reason, Axiom Analysis Suite's “Best Practices Workflow” can be thought of as a two-step process:

Step 1 Sample QC: Identify and exclude the poor quality sample CEL files.
Step 2 Genotyping: Genotype only the sample CEL files that pass Step 1.

- **Step 1:** The Sample QC Workflow performs the following steps:
  - Genotype a small set of SNPs used to uniquely identify each sample ("Signature SNPs"). This method uses static calling boundaries, so a given sample's calls are not influenced by other samples.
  - **DishQC**: Calculate DishQC metric and exclude CEL files with too small a DishQC from next step. For this method, a given sample's metrics are not influenced by other samples.
  - **QC Call Rate**: Initial cluster genotyping on remaining samples for selected QC SNPs, to identify and remove any additional CEL files with low QC call rates. For this method, batch information from other samples influences a given samples QC call rates. Genotypes are not stored.
  - **Plate QC**: Compare the average QC Call Rate of passing samples within a plate against the threshold "Average call rate for passing samples". If the metric is below this threshold, all the samples on the plate will fail Plate QC, and will not be genotyped.
  - **Identify Copy Number plate controls**:
    - For CN-aware genotyping arrays like PharmacoScan, auto-identify reagent control samples by comparing measured SignatureSNP genotypes from all CEL files against reference genotypes. A control is identified if [number of Signature SNPs with a call is >= "Control comparisons"] AND [Concordance % of signature SNP calls compared with reference genotypes is >= "Control concordance"]. Control samples that pass all sample QC checks will be used in the subsequent Genotyping workflow, unless overridden by the user.
    - For predefined CN region arrays like CarrierScan, select as CN controls all female samples passing sample QC.

- **Step 2:** Genotyping can be run by itself or as part of the Best Practices Workflow. When run as part of the Best Practices Workflow, only CEL files passing Sample QC are genotyped. When run by itself, the user has full control of which samples to genotype. Genotyping does the following steps:
  - For arrays supporting predefined region CN analysis or CN-aware genotyping, collate normalized signals for CN probesets of interest and perform predefined-region CN analysis. This step also calculates CN-specific QC metrics MAPD and wavinessSD. Samples that fail CN QC have their CN states reset to NoCall. If control samples are supplied (by user or by Sample QC step), then plate-based signal correction is performed. Control samples that fail CN QC are not used for plate-based signal correction.

IMPORTANT! If you plan to run only the Genotyping workflow, and if the CN Control CEL List File is an Analysis Settings input option, this file should be supplied. If it is not, then plate signal correction is not performed and CN results may suffer.

The file **[Batch Name]\CNData\CNcontrolSamples.pass.txt** (from the Sample QC step) is supplied for CN analysis during a Best Practices Workflow.

  - For CN-aware genotyping arrays like PharmacoScan, CN results are supplied as an input to the next genotyping step.
  - Perform genotyping, using either default or user-supplied input options. Genotypes are saved.
  - Generate summary statistics on each SNP.

B

About Allele Translation

Overview

Human genome sequence variation, which includes both single nucleotide polymorphisms (SNPs) as well as more complex structural variation in the form of insertions, duplications and deletions, underlies each individual's response to drugs. Products like PharmacoScan and DMET Plus are designed to enable comprehensive and accurate genotyping of specific polymorphisms involved in drug-metabolizing enzymes and transporters. Axiom Analysis Suite enables conversion of genotype calls to clinically-recognized star nomenclature via Allele Translation. This section explains the organization of the translation reports to help you interpret the translation data. Key concepts such as phase ambiguity and the impact of missing data on haplotype-based allele calling are described.

Gene Table Layout for Haplotyping

To appreciate how haplotyping operates, it is essential to describe the organization of the gene tables (Figure 161 and Figure 162) in the translation library file (***.translation**).

Figure 161 Biological annotations [Example: Gene table data for markers in CYP1A1]

CYP1A1	Reference Link	Probe Set ID	Switch Design Strand to Report	dbSNP RS ID	Defining	cDNA Nucleotide Position	Genome Position	Change	Common Name
CYP1A1	PMID: 11295847	AM_10774	Y	rs56313657	*6	993G>T	Ch15:75013804	M331I	CYP1A1*6_1635G>T(M331I)
CYP1A1	PMID: 15618738	AM_10771	Y	rs72547510	*7	1275_1276ir	Ch15:75013093	E426Fra	CYP1A1*7_2345insT
CYP1A1	PMID: 15618738	AM_10770	Y	rs72547509	*8	1343T>A	Ch15:75013026	I448N	CYP1A1*8_2413T>A(I448N)
CYP1A1	PMID: 8895751	AM_10769	Y	rs1799814	*4	1382C>A	Ch15:75012987	T461N	CYP1A1*4_2452C>A(T461N)
CYP1A1	PMID: 9070254	AM_10768	Y	rs1048943	*2C	1384A>G	Ch15:75012985	I462V	CYP1A1*2C_2454A>G(I462V)
CYP1A1	PMID: 11295847	AM_10766	N	rs41279188	*5	1390C>A	Ch15:75012979	R464S	CYP1A1*5or*9_2460C>A>T(R464SorC)
CYP1A1	PMID: 11295847	AM_10766	N	rs41279188	*9	1390C>T	Ch15:75012979	R464C	CYP1A1*5or*9_2460C>A>T(R464SorC)
CYP1A1	PMID: 15618738	AM_10765	Y	rs56240201	*10	1429C>T	Ch15:75012940	R477W	CYP1A1*10_2499C>T(R477W)
CYP1A1	PMID: 15618738	AM_10762	N	rs1800031	*3	*595T>C	Ch15:75012235	3'UTR	CYP1A1*3_3204T>C(3'UTR)
CYP1A1	PMID: 11295847	AM_10778	N	rs4646422	G45D	134G>A	Ch15:75015305	G45D	CYP1A1_134G>A(G45D)
CYP1A1	PMID: 11295847	AM_10776	N	rs34260157	R279W	835C>T	Ch15:75014049	R279W	CYP1A1_1390C>T(R279W)
CYP1A1	PMID: 9353182	AM_10775	Y	rs4987133	I286T	857T>C	Ch15:75014027	I286T	CYP1A1_1412T>C(I286T)
CYP1A1	PMID: 15618738	AM_10772	Y	rs2856833	F381L	1143C>A	Ch15:75013563	F381L	CYP1A1_1876C>A(F381L)
CYP1A1	GBID: NM_00049	AM_10767	Y	rs2278970	A463G	1388C>G	Ch15:75012981	A463G	CYP1A1_2458C>G(A463G)

Figure 162 Haplotype descriptions [Example gene table data for markers in CYP1A1]

Common Name	Haplotype	Reference	Variant	*1	*2C	*3	*4	*5	*6	*7	*8	*9	*10	G45D	R279	I286T	F3
CYP1A1*6_1635G>T(M331I)	Y	G	T						T								
CYP1A1*7_2345insT	Y	-	T							T							
CYP1A1*8_2413T>A(I448N)	Y	T	A								A						
CYP1A1*4_2452C>A(T461N)	Y	C	A				A										
CYP1A1*2C_2454A>G(I462V)	Y	A	G		G												
CYP1A1*5or*9_2460C>A>T(R464SorC)	Y	C	A					A									
CYP1A1*5or*9_2460C>A>T(R464SorC)	Y	C	T									T					
CYP1A1*10_2499C>T(R477W)	Y	C	T										T				
CYP1A1*3_3204T>C(3'UTR)	Y	T	C			C											
CYP1A1_134G>A(G45D)	N	G	A											A			
CYP1A1_1390C>T(R279W)	N	C	T												T		
CYP1A1_1412T>C(I286T)	N	T	C													C	
CYP1A1_1876C>A(F381L)	N	C	A														A
CYP1A1_2458C>G(A463G)	N	C	G														

Biological annotations: The first set of columns in the table are annotations for the markers in each of the translated genes. Haplotype descriptions: The columns beyond the Common Name field contain information used for interpretation and translation of the gene file.

Following the columns enumerating the Reference and Variant alleles, haplotypes and markers in the gene are listed. In this example, CYP1A1 has 10 haplotypes described and they are named in the column headers. In addition, there are five additional markers for rare variants that can also be identified in this gene. Notice that the first haplotype described in this table is CYP1A1*1A, and that all markers except the last five are haplotyping markers (see the Haplotype field). Markers are characterized as "non-haplotyping" if their state in every reportable haplotype is not known. For example, a variant has been identified in this gene resulting in a non-synonymous change in the protein (CYP1A1_134G>A(G45D)), but the haplotype background of that variant is not available in the literature references used (for example, the Karolinska reference database; http://www.cypalleles.ki.se/).

Figure 161 illustrates the way that haplotypes are called. Notice that the haplotype names appear in the header of each gene table (for CYP1A1, when all markers are tested, this is *1, *2C, *3, *4, etc). Only differences from the reference haplotype appear in the haplotype columns, and the change for the altered base is indicated. For example, CYP1A1*2C contains one difference relative to CYP1A1*1, specifically a mutation 5'-prime to the gene, defined by the probe "CYP1A1*2C_2454A>G(I462V)".

The Defining field in the translation file lists the effect that a variant allele of this marker has. For example, because a variant at the final probe in the table, "CYP1A1_2458C>G(A463G)", results in a structural change in the protein, this marker is flagged with the name of the amino acid change that results: alanine at position 463 of the protein is a glycine in this variant (A463G). Although not shown in this gene, if a marker contains a variant allele in multiple haplotypes, then the marker does not uniquely define a single haplotype. That marker would then have an "N" (No) in the Defining field.

Haplotype field names with a # prefix, and rows with a # prefix in the Probe Set ID field are "commented out", and not used for translation.

Biological Annotations in Translation Reports

The primary function of the translation reporting is to summarize genotypes into commonly recognized variant names. In the case of the CYP450 core gene set, this translates to the Star-nomenclature followed by standardized nomenclature committee direction. Similar names are used by other steering committees such as the two Phase II enzyme genes, N-acetylase genes (NAT1 and NAT2) or the UGT-transferase gene families. Wherever possible, we have attempted to use a standard naming convention for the markers. To facilitate interpretation of the genotyping results, the translation reports provide:

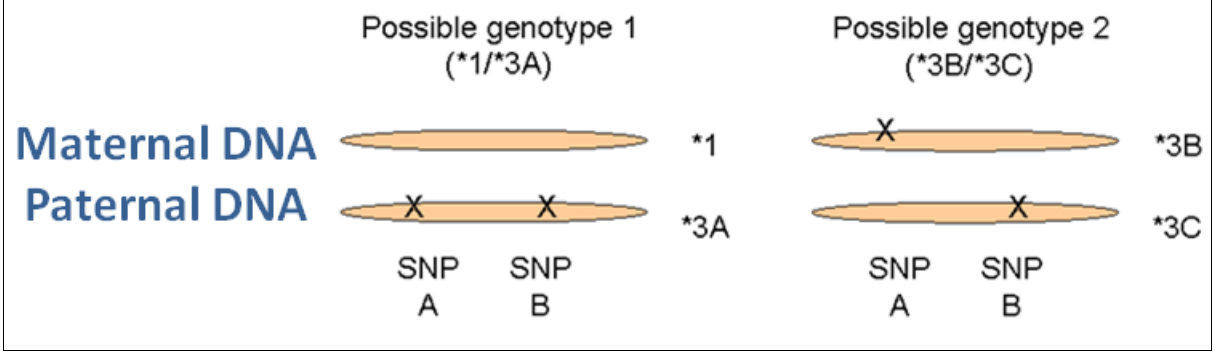
- Reference publications, sequences or dbSNP identifiers for following previously published information about the variant site
- Precise genomic location in a recent genome build for identifying confirmatory genotyping assays
- Notation of protein changes that may result from the mutations in the panel. This field may also indicate whether the variant allele is strategically positioned in the promoter region or causes changes in splice junction sequences in the gene.
- Description of the initial star-allele which the variant was identified. Generally, this corresponds to the Summary flag entry
- Alternative alleles at each marker and whether the defined allele is the Reference base or Variant (corresponding to the altered gene form)

Along with the identified genotypes of the sample, this information provides biological evidence supporting haplotype calls.

Impact of Phase Ambiguity in Haplotyping

It is not unusual for individuals to be heterozygous at more than one defining marker (compound heterozygote genotype) in a gene. When this happens, multiple haplotype pairs may be consistent with resulting profiles, as shown in Figure 163.

Figure 163 Two phase ambiguity in the TPMT gene



In this example, it is clear that the child could have inherited both variant alleles from one parent (in that case their diplotype would be *1/*3A), or could have inherited one variant allele from each of the two parents (in that case the diplotype would be *3B/*3C). Although the *3A haplotype is less common than either *3B or *3C, the translation reports list both potential haplotype pairs in the output reports. One reason for this is that the phenotypes may differ between the two alternative genetic configurations. In this case, three of the four alleles of TPMT have reduced activity: *3A, *3B and *3C, whereas the reference allele, *1 is a normally functioning allele. Phase ambiguity is relatively common in genes with common polymorphisms. Figure 164 on page 143 lists the multiple possible calls due to phase ambiguity that were observed in six HapMap populations, and how often they occurred.

Figure 164 Example: Observed phase ambiguities in DMET Plus in a data set of six HapMap populations with 597 individuals and no children.

Gene	Call 1	Call 2	Call 3	Call Rate
CYP1A2	*1A/*1L	*1C/*1F		22.6%
CYP2B6	*1/*7	*5/*6		2.0%
CYP2D6	*2/*64	*10/*17		0.3%
FMO2	*2A/*3	*1/*2C		0.5%
NAT1	*4/*11	*11C/*30		1.8%
NAT2	*4/*5E	*5/*6		9.2%
NAT2	*4/*6J	*6/*7		3.7%
NAT2	*4/*14D	*6/*14		2.0%
NAT2	*4/*14F	*5/*14		1.3%
NAT2	*4/*7D	*7/*14		0.2%
SLC22A2	*1/*3D	*3A/*6		7.4%
SLC22A2	*1/*2B	*2A/*3A		6.2%
SLC22A2	*1/*3E	*2A/*3D	*2B/*6	5.5%
SLC22A2	*2B/*3D	*3A/*3E		1.3%
SLCO1B1	*1b/*17	*15/*21		5.9%
SLCO1B1	*1a/*14	*1b/*4		3.2%
SLCO1B1	*1a/*15	*1b/*5		3.2%
SLCO1B1	*1a/*17	*5/*21		2.5%
TPMT	*1/*3A	*3B/*3C		1.3%
UGT1A1	*1/*28+60	*28/*60		1.5%
UGT1A1	*1/*27+28+60+93	*27/*28+60+93		0.5%
UGT1ACOMMON	*76+79/*IA	*76/*79		8.0%
UGT2B15	*1/*5	*2/*4		12.9%

It is worth pointing out that the predicted phenotypes of some of these alternative diplotype calls are identical, and in these cases the Phenotype report will then report a single phenotype. When phase ambiguity is encountered and the Phenotype report does not resolve the multiple calls to a single Phenotype Call, follow-up metabolic screening may be merited to differentiate the actual genetic configuration of the test samples.

Diplotype to Phenotype Translation

Allele translations include a Phenotype report if the required metabolizer library file has been selected. The Phenotype report further translates the reported diplotypes (star allele pairs) from a subset of genes in the Comprehensive report into one of several phenotypes (e.g. "Poor Metabolizer"). As the software reads the comprehensive.rpt file, it will try to match Known Call diplotype values for each gene of each sample to one row of the metabolizer library file table. If a match is found, the associated phenotype and allele activities are written to the phenotype.rpt. If a match is not found, a Phenotype Call of "unknown" is reported. More information on this software feature is available in the DMET™ Plus Allele Translation white paper.

IMPORTANT! Users are responsible for reviewing the metabolizer library file for accuracy!

Phenotype Call and Gene Activity interpretations for a Known Call are supported by differing levels of evidence from in vivo and/or in vitro research studies. Refer to metabolizer library file for a list of references. The actual phenotype and gene activities may be dependent on the substrate and dose.

If you do not want to generate a phenotype report, leave the Metabolizer File option blank in the Perform Allele Translation dialog.

If you want to report phenotypes for only a subset of genes, there are two ways to accomplish this:

1. Import a custom SNP List into the workspace containing probesets from only the genes of interest. At the point when you normally perform allele translation, select the option to filter to just probesets in this list.

   Or

2. Create a copy of the metabolizer file that only contains the genes of interest.

If you want to change what phenotypes are reported for a particular combination of diplotypes, or you would like to change what is written to the header of the Phenotype Translation report, you will need to create and use a custom version of the .metabolizer library file. Instructions for doing this follow.

Creating a Custom Metabolizer Library File

If you choose to create a custom metabolizer table, start with the default metabolizer file as a template, then save a copy of this file using a new name.

IMPORTANT! Use caution if editing the metabolizer file with Microsoft Excel. For example, Excel inserts quotation marks around text containing commas, which may make the file unreadable by Axiom Analysis Suite.

Before using the file with Axiom Analysis Suite, open it in another text editor and remove any unexpected text such as quotation marks.

To be recognized by Axiom Analysis Suite, the file:

- Must have the file extension ***.metabolizer**, where the * indicates your custom text.
- Must be encoded in ANSI, not Unicode or other encoding.
- Must exist in the library folder used by Axiom Analysis Suite.
- Must be selected for use from the **Configuration** -> **Options** menu, in the Translations tab.
- Must be properly formatted. It is recommended that you use a file comparison utility to verify that the only changes between the original and modified files are expected changes.

If you want to add phenotype reporting (for genes not currently in the metabolizer library file) the gene names and star allele names you want to add must exist in the ***.translation** library file used to generate the ***_comprehensive.rpt** file.

Metabolizer Library File Format

The **.metabolizer** library file is a tab-delimited text file that can be edited in any text editor. This file consists of a header section followed by a single table. Any rows from the start of the file until the beginning of the main table are considered header rows, and must begin with a pound or hash sign (#). Header rows are optional. Header rows beginning with **#%Info=** is added to the header of the ***_phenotype.rpt** file, so you can put custom text into your reports. The first row that does not begin with # must use the names shown below, be tab separated, and contain only lowercase letters.

```
gene    allele_1    allele_2    phenotype    activity_1    activity_2
```

Additional field names can be added to the first table row, but they will not be used. After the first table row, all rows require a value for the following fields:

```
gene    allele_1    allele_2    phenotype
```

The **activity_1** and **activity_2** fields can be left blank. See Figure 165 on page 147 for field descriptions.

Figure 165 field descriptions for the Metabolizer library file

Metabolizer field	Description
gene	The gene name as reported in the comprehensive report. These values can also be found in the *.translation library file needed to generate the comprehensive report.
allele_1 allele_2	The haplotype name of an allele for a gene as reported in the comprehensive report, e.g. '*2'. A Known Call in the comprehensive report is usually a single pair of alleles, e.g.'*1/*2'. To have this call be matched to a specific row in the metabolizer table, **only one of the rows is needed** in the following table:

gene	allele_1	allele_2	phenotype
CYP2D6	*1	*2	NM
CYP2D6	*2	*1	NM

Axiom Analysis Suite will report an error IF it detects duplicate rows (as in the above example) AND IF the duplicate rows report disagreeing phenotypes.

phenotype	The value that should be reported for the associated 'gene allele_1/allele_2' call. The string should be short with no commas, quotes, or whitespace characters. Common phenotype names are:

phenotype	definition
UM	Ultra-rapid metabolizer
RM	Rapid metabolizer
NM	Normal metabolizer
IM	Intermediate metabolizer
PM	Poor metabolizer
unknown	Unknown metabolizer

activity_1 activity_2	The reported gene activity or function level for an allele, e.g. 'normal' or 'decreased'. Activity_1 is for allele_1, and activity_2 is for allele_2. The values in these fields are used to populate the Gene Activity field in the phenotype report, e.g. 'normal/decreased'. If you leave these fields empty, the phenotype report will display '/' for the Gene Activity.
Optional fields	Axiom Analysis Suite will ignore additional fields in the metabolizer file. Additional fields may be used to annotate each row.

Reference Databases Used in Translation Data Curation

The databases used to curate the allele translation gene tables include:

- PharmGKB - Stanford University Pharmacogenomics reference database
  http://www.pharmgkb.org
- Karolinska cytochrome P450 gene standard nomenclature
  http://www.cypalleles.ki.se
- Database of NAT genes (Democritus University of Thrace)
  http://nat.mbg.duth.gr
- Database of UGT genes
  https://www.pharmacogenomics.pha.ulaval.ca/ugt-alleles-nomenclature
- Drug interaction database (University of Indiana)
  http://medicine.iupui.edu/clinpharm/ddis
- PubMed - On-line National Library of Medicine publication database
  http://www.ncbi.nlm.nih.gov/pubmed

C Definitions

Threshold Names

Use the table of definitions below to help select thresholds (when setting up an analysis or regenerating SNP metrics).

Table 11 Threshold Names

Threshold Group	Threshold Name	Description
Sample QC	**DQC**	A sample's dish QC value must be greater than or equal to 'DQC' to pass sample QC.
Sample QC	**QC call_rate**	A sample's call rate value must be greater than or equal to 'QC call_rate' to pass genotyping QC.
Sample QC	**Percent of passing samples**	If a plate's percent of passing samples is smaller than this number, all samples on the plate will show a warning in the Sample Table.
Sample QC	**Average call rate for passing samples**	A plate's average QC call rate of passing samples must be greater than or equal to this number to pass plate QC.
Sample QC	**Control comparisons**	CN-aware genotyping parameter. For identifying control samples, this is the minimum number of SignatureSNP probesets compared to a reference.
Sample QC	**Control concordance**	CN-aware genotyping parameter. For identifying control samples, this is the minimum percent concordance of SignatureSNP calls to a reference.
CN QC	**MAPD**	For arrays that support copy number (CN) analysis, the Median Absolute Pairwise Difference of log2ratio signals of adjacent copy number (CN) probesets must be less than this value to make CN calls.
CN QC	**Waviness SD**	For arrays that support copy number (CN) analysis, the Waviness Standard Deviation of log2 ratio signals of copy number (CN) probesets must be less than this value to make CN calls.
SNP QC	**species-type**	Species type for the array, which affects some SNP QC checks. Species types include: Diploid, Human, and Polyploid. Note: Selecting 'Human' will assign a probeset with two clusters that are both homozygous to a non-recommended ConversionType like 'Other'.
SNP QC	**cr-cutoff**	Minimum acceptable call rate.

Table 11 Threshold Names

Threshold Group	Threshold Name	Description
SNP QC	**fld-cutoff**	For autosomal probesets, minimum acceptable FLD value for cluster separation.
SNP QC	**het-so-cutoff**	Minimum acceptable value for the correctness of the Size (Y position) offset of the heterozygous cluster.
SNP QC	**het-so-XChr-cutoff**	For probesets on the non-pseudoautosomal regions of chromosome X, the minimum acceptable value for the correctness of the Size (Y position) offset of the female heterozygous cluster.
SNP QC	**het-so-otv-cutoff**	Minimum acceptable value for the correctness of the Size (Y position) offset of the heterozygous cluster, possibly indicating a fourth cluster below the heterozygous cluster (OTV).
SNP QC	**hom-ro-1-cutoff**	Minimum acceptable value for the correctness of the Contrast (X position) of the homozygous clusters (Ratio Offset) when a probeset has 1 genotype cluster.
SNP QC	**hom-ro-2-cutoff**	Minimum acceptable value for the correctness of the Contrast (X position) of the homozygous clusters (Ratio Offset) when a probeset has 2 genotype clusters.
SNP QC	**hom-ro-3-cutoff**	Minimum acceptable value for the correctness of the Contrast (X position) of the homozygous clusters (Ratio Offset) when a probeset has 3 genotype clusters.
SNP QC	**hom-ro**	Flag indicating whether the metric HomRO is used in classification.
SNP QC	**hom-het**	If hom-het flag is FALSE, a probeset with two clusters that are homozygous and heterozygous is assigned a non-recommended ConversionType like 'Other'. Hom-het flag should be TRUE for human and diploid species for a large batch of samples, and should be FALSE for highly inbred samples such as polyploid species.
SNP QC	**num-minor-allele-cutoff**	Minimum minor allele count for categorizing a probeset as PolyHighResolution.
SNP QC	**hom-ro-hap-1-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 1 haploid genotype cluster.
SNP QC	**hom-ro-hap-1-XChr-cutoff**	For non-PAR X probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 1 haploid genotype cluster.
SNP QC	**hom-ro-hap-1-YChr-cutoff**	For Y probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 1 haploid genotype cluster.

Table 11 Threshold Names

Threshold Group	Threshold Name	Description
SNP QC	**hom-ro-hap-1-MTChr-cutoff**	For mitochondrial probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 1 haploid genotype cluster.
SNP QC	**hom-ro-hap-2-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 2 haploid genotype clusters.
SNP QC	**hom-ro-hap-2-XChr-cutoff**	For non-PAR X probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 2 haploid genotype clusters.
SNP QC	**hom-ro-hap-2-YChr-cutoff**	For Y probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 2 haploid genotype clusters.
SNP QC	**hom-ro-hap-2-MTChr-cutoff**	For mitochondrial probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters from 0 (Ratio Offset) when a probeset has 2 haploid genotype clusters.
SNP QC	**hom-hap-X-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters relative to the homozygous clusters.
SNP QC	**hom-hap-Y-lower-cutoff**	For autosomal probesets, maximum acceptable value for the correctness of the Size (Y position) of the haploid clusters relative to homozygous clusters.
SNP QC	**hom-hap-Y-upper-cutoff**	For autosomal probesets, maximum acceptable value for the correctness of the Size (Y position) of the haploid clusters relative to homozygous clusters.
SNP QC	**CN0-hap-X-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the ZeroCN cluster relative to the haploid clusters.
SNP QC	**CN0-hap-X-YChr-cutoff**	For Y probesets, minimum acceptable value for the correctness of the Contrast (X position) of the ZeroCN cluster relative to the haploid clusters.
SNP QC	**CN0-hap-Y-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Size (Y position) of the ZeroCN cluster relative to the haploid clusters.
SNP QC	**CN0-hap-Y-YChr-cutoff**	For Y probesets, minimum acceptable value for the correctness of the Size (Y position) of the ZeroCN cluster relative to the haploid clusters.
SNP QC	**CN0-dip-X-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the ZeroCN cluster relative to the diploid clusters.

Table 11 Threshold Names

Threshold Group	Threshold Name	Description
SNP QC	**CN0-dip-Y-cutoff**	For autosomal probesets, minimum acceptable value for the correctness of the Size (Y position) of the ZeroCN cluster relative to the diploid clusters.
SNP QC	**aaf-XChr-cut**	For non-PAR X probesets, maximum acceptable difference in male versus female A-allele frequency. Any probeset that fails is categorized as Other.
SNP QC	**fld-XChr-cut**	For non-PAR X probesets, minimum acceptable FLD value for diploid cluster separation. Note: FLD is computed differently for non-PAR X probesets than for autosomal probesets.
SNP QC	**homfld-XChr-cut**	For non-PAR X probesets, minimum acceptable FLD value for male haploid cluster separation.
SNP QC	**homfld-YChr-cut**	For Y probesets, minimum acceptable FLD value for male haploid cluster separation.
SNP QC	**sign-diff-hom-1-cutoff**	For multi-allele probesets, minimum acceptable difference between mean log2 signal and background when there is 1 homozygous genotype cluster.
SNP QC	**sign-diff-hom-2-cutoff**	For multi-allele probesets, minimum acceptable difference between mean log2 signal and background when there are 2 homozygous genotype clusters.
SNP QC	**min-mean-cp2-cutoff**	For multi-allele probesets, minimum acceptable mean log2 allele signal for samples that appear to have at least 2 copies of that allele.
SNP QC	**max-mean-cp2-cutoff**	For multi-allele probesets, maximum acceptable mean log2 allele signal for samples that appear to have at least 2 copies of that allele.
SNP QC	**min-genotype-freq-samples**	Minimum count of samples to support the genotype frequency p-value calculation. This parameter is used if a genotype frequency file is supplied.
SNP QC	**genotype-p-value-cutoff**	Minimum acceptable value for the genotype frequency p-value calculation. Probesets not meeting this threshold may be categorized as 'UnexpectedGenotypeFreq'. This parameter is used if a genotype frequency file is supplied, and if the count of genotyped samples is at least min-genotype-freq-samples.
SNP QC	**priority-order**	Priority order of probeset conversion types when performing probeset selection.
SNP QC	**recommended**	Probesets having ConversionTypes in this recommended set will be included in the recommended probeset list.
SNP QC	**y-restrict**	Y-restrict is the maximum vertical distance 2 points in 2 clusters can be when matching up points for calculating the edge metric values BB_dis_x_adj and AA_dis_x_adj . Y-restrict is not used to set probeset ConversionType. It is only used during 'Regenerate SNP Metrics' when the 'Generate advanced metrics' option is selected.

Table 11 Threshold Names

Threshold Group	Threshold Name	Description
PS Supplemental	**homfld-cut**	Minimum acceptable FLD value for homozygous cluster separation.
PS Supplemental	**bb-varx-cut**	Maximum acceptable Contrast (X position) variance for BB cluster. Not used if bb-varx-z-cut-enabled is true.
PS Supplemental	**bb-vary-cut**	Maximum acceptable Size (Y position) variance for BB cluster. Not used if bb-vary-z-cut-enabled is true.
PS Supplemental	**ab-varx-cut**	Maximum acceptable Contrast (X position) variance for AB cluster. Not used if ab-varx-z-cut-enabled is true.
PS Supplemental	**ab-vary-cut**	Maximum acceptable Size (Y position) variance for AB cluster. Not used if ab-vary-z-cut-enabled is true.
PS Supplemental	**aa-varx-cut**	Maximum acceptable Contrast (X position) variance for AA cluster. Not used if aa-varx-z-cut-enabled is true.
PS Supplemental	**aa-vary-cut**	Maximum acceptable Size (Y position) variance for AA cluster. Not used if aa-vary-z-cut-enabled is true.
PS Supplemental	**bb-varx-z-cut-enabled**	If true, bb-varx-z-cut is used.
PS Supplemental	**bb-vary-z-cut-enabled**	If true, bb-vary-z-cut is used.
PS Supplemental	**ab-varx-z-cut-enabled**	If true, ab-varx-z-cut is used.
PS Supplemental	**ab-vary-z-cut-enabled**	If true, ab-vary-z-cut is used.
PS Supplemental	**aa-varx-z-cut-enabled**	If true, aa-varx-z-cut is used.
PS Supplemental	**aa-vary-z-cut-enabled**	If true, aa-vary-z-cut is used.
PS Supplemental	**bb-varx-z-cut**	Maximum acceptable Contrast (X position) Z-score for BB cluster. Not used if bb-varx-z-cut-enabled is false.
PS Supplemental	**bb-vary-z-cut**	Maximum acceptable Size (Y position) Z-score for BB cluster. Not used if bb-vary-z-cut-enabled is false.
PS Supplemental	**ab-varx-z-cut**	Maximum acceptable Contrast (X position) Z-score for AB cluster. Not used if ab-varx-z-cut-enabled is false.
PS Supplemental	**ab-vary-z-cut**	Maximum acceptable Size (Y position) Z-score for AB cluster. Not used if ab-vary-z-cut-enabled is false.
PS Supplemental	**aa-varx-z-cut**	Maximum acceptable Contrast (X position) Z-score for AA cluster. Not used if aa-varx-z-cut-enabled is false.
PS Supplemental	**aa-vary-z-cut**	Maximum acceptable Size (Y position) Z-score for AA cluster. Not used if aa-vary-z-cut-enabled is false.
PS Supplemental	**clustermin**	Minimum number of samples in a cluster for the new variance or Z-score value to be calculated.
PS Supplemental	**n-minor-hom-cut**	Minimum number of minor hom samples for PolyHighResolution classification. PolyHighResolution probesets where the number of minor hom samples is less than n-minor-hom-cut are categorized as nMinorHom.

Table 11 Threshold Names

Threshold Group	Threshold Name	Description
PS Supplemental	**hetv-maf-cut**	Maximum acceptable heterozygosity parameter. PolyHighResolution probesets with het rate > MAF*hetv-maf-cut are categorized as UnexpectedHeterozygosity.
PS Supplemental	**variance-class**	List of categories that supplemental filters are used on.

Sample Table

Use the table of definitions below for the fields in the Sample Table.

Table 12 Sample Table

Selection	Definition
Sample Filename	CEL file name.
Pass/Fail	Sample quality control status. A sample is called 'Pass' by the Sample QC step if it meets the minimum thresholds for DQC, QC call_rate, and average call rate for passing samples.
DQC	DishQC measures the amount of overlap between two homozygous peaks created by non-polymorphic probes. DQC of 1 is no overlap, which is good. DQC of 0 is complete overlap, which is bad.
call_rate	Percentage of autosomal SNPs with a call other than NoCall. 'SpecialSNP' probesets are excluded, as they are mostly non-autosomal.
QC call_rate	Percentage of autosomal SNPs with a call other than NoCall (measured at the Sample QC step).
het_rate	Percentage of SNPs called AB (i.e. the heterozygosity) for autosomal SNPs.
QC het_rate	Percentage of SNPs called AB (i.e. the heterozygosity) for autosomal SNPs (measured at the Sample QC step).
computed_gender	Computed gender for the sample.
QC computed_gender	Computed gender for the sample (measured at the Sample QC step).
affymetrix-plate-barcode	Plate barcode number.
QC affymetrix-plate-barcode	Plate barcode number (measured at the Sample QC step).
affymetrix-plate-peg-wellposition	Well position of the plate's peg.
QC affymetrix-plate-peg-wellposition	Well position of the plate's peg (measured at the Sample QC step).
Average call rate for passing samples	Average QC Call Rate of passing samples within the plate to which this sample belongs.
Percent of passing samples	Percentage of samples passing sample QC within the plate to which this sample belongs.
cel_filepath	CEL file path.
cel_file_identifier	CEL file identifier.

Table 12 Sample Table

Selection	Definition
affymetrix-array-id	Command Console array identifier.
total_call_rate	Call rate at the default or user-specified threshold for all SNPs.
QC total_call_rate	Call rate at the default or user-specified threshold for all SNPs (measured at the Sample QC step).
total_het_rate	Percentage of SNPs called AB (i.e., the heterozygosity) for all SNPs.
QC total_het_rate	Percentage of SNPs called AB (i.e., the heterozygosity) for all SNPs (measured at the Sample QC step).
hom_rate	Percentage of SNPs called AA or BB (i.e. the homozygosity) for autosomal SNPs.
QC hom_rate	Percentage of SNPs called AA or BB (i.e. the homozygosity) for autosomal SNPs (measured at the Sample QC step).
total_hom_rate	Percentage of SNPs called AA or BB (i.e. the homozygosity) for all SNPs.
QC total_hom_rate	Percentage of SNPs called AA or BB (i.e. the homozygosity) for all SNPs (measured at the Sample QC step).
cluster_distance_mean	Average distance to the cluster center for the called genotype.
QC cluster_distance_mean	Average distance to the cluster center for the called genotype (measured at the Sample QC step).
cluster_distance_stdev	Standard deviation of the distance to the cluster center for the called genotype.
QC cluster_distance_stdev	Standard deviation of the distance to the cluster center for the called genotype (measured at the Sample QC step).
allele_summarization_mean	Average of the allele signal estimates (log2 scale).
QC allele_summarization_mean	Average of the allele signal estimates (log2 scale) (measured at the Sample QC step).
allele_summarization_stdev	Standard deviation of the allele signal estimates (log2 scale).
QC allele_summarization_stdev	Standard deviation of the allele signal estimates (log2 scale) (measured at the Sample QC step).
allele_deviation_mean	Average of the absolute difference between the log2 allele signal estimate and its median across all arrays.
QC allele_deviation_mean	Average of the absolute difference between the log2 allele signal estimate and its median across all arrays (measured at the Sample QC step).
allele_deviation_stdev	Standard deviation of the absolute difference between the log2 allele signal estimate and its median across all arrays.
QC allele_deviation_stdev	Standard deviation of the absolute difference between the log2 allele signal estimate and its median across all arrays (measured at the Sample QC step).
allele_mad_residuals_mean	Average of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model.

Table 12 Sample Table

Selection	Definition
QC allele_mad_residuals_mean	Average of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model (measured at the Sample QC step).
allele_mad_residuals_stdev	Standard deviation of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model.
QC allele_mad_residuals_stdev	Standard deviation of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model (measured at the Sample QC step).
A_signal_mean	Average of control A probe raw intensities in the AT channel.
T_signal_mean	Average of control T probe raw intensities in the AT channel.
G_signal_mean	Average of control G probe raw intensities in the GC channel.
C_signal_mean	Average of control C probe raw intensities in the GC channel.
AT_B	AT channel background, which is the average signal of the GC control probes in AT channel.
GC_B	GC channel background, which is the average signal of the AT control probes in GC channel.
AT_S	AT channel signal, which is the average signal of the AT control probes in AT channel.
GC_S	GC channel signal, which is the average signal of the GC control probes in GC channel.
AT_SBR	AT channel signal to background ratio, defined as AT_S/AT_B.
GC_SBR	GC channel signal to background ratio, defined as GC_S/GC_B.
AT_B_IQR	AT channel interquartile range (middle 50%) of background intensities, measured using GC control probes.
GC_B_IQR	GC channel interquartile range (middle 50%) of background intensities, measured using AT control probes.
AT_S_IQR	AT channel interquartile range (middle 50%) of signal intensities, measured using AT control probes.
GC_S_IQR	GC channel interquartile range (middle 50%) of signal intensities, measured using GC control probes.
CV_AT	AT channel median coefficient of variation of replicate probe signals for control probesets.
CV_GC	GC channel median coefficient of variation of replicate probe signals for control probesets.
AT_FLD	Fisher's Linear Discriminant between signal and background in the AT channel, defined as [median_of_AT_probe_intensities – median_of_GC_probe_intensities]^2 / [0.5*(AT_S_IQR^2 + AT_B_IQR^2)].

Table 12 Sample Table

Selection	Definition
GC_FLD	Fisher's Linear Discriminant between signal and background in the GC channel, defined as [median_of_GC_probe_intensities – median_of_AT_probe_intensities]^2 / [0.5*(GC_S_IQR^2 + GC_B_IQR^2)].
log_diff_qc	A cross channel QC metric, defined as mean(log(AT_SBR))/ std(log(AT_SBR))+mean(log(GC_SBR))/std(log(GC_SBR)), where signal and background are calculated for control non-polymorphic probes after intensity normalization.
saturation_AT	Fraction of features in the AT channel with intensity greater than or equal to 3800. Features likely to saturate the scanner in the long exposure image will instead be measured in the short exposure image, and all feature signals are scaled using a high dynamic range exposure merging technique.
saturation_GC	Fraction of features in the GC channel with intensity greater than or equal to 3800. Features likely to saturate the scanner in the long exposure image will instead be measured in the short exposure image, and all feature signals are scaled using a high dynamic range exposure merging technique.
cn-probe-chrXY-ratio_gender_meanX	Average probe intensity (raw, untransformed) of X chromosome nonpolymorphic probes.
cn-probe-chrXY-ratio_gender_meanY	Average probe intensity (raw, untransformed) of Y chromosome nonpolymorphic probes.
cn-probe-chrXY-ratio_gender_ratio	Gender ratio Y/X = cn-probe-chrXY-ratio_gender_meanY / cn-probe-chrXY-ratio_gender_meanX.
cn-probe-chrXY-ratio_gender	Predicted gender, based on the value of cn-probe-chrXY-ratio_gender_ratio.
reagent_version	Reagent version used for processing the arrays, based on data intensity values. Note: You can only perform batch genotyping analysis on CEL files processed using the same reagent version.
reagent_discrimination_value	Value assigned to the reagent, and used to determine reagent_version.
user_color	User-supplied color to associate with the sample. For more information, see "Setting User Colors" on page 55.

Concordance Columns

Use the table of definitions below for the columns in the Concordance table.

Table 13 Concordance Columns

Column Name	Description
Base	The first sample in the comparison.
Reference	The second sample in the comparison.
#SNPs Called	Number of SNPs common to both sample and reference files with genotype calls.
#Concordant SNP's	Number of called SNPs that have the same genotype call.
%Concordance	Percentage of called SNPs that have the same genotype call.

Annotations and Columns

Use the table of definitions below to help select the annotations you can add to the SNP Summary Table.

Table 14 Annotations and Columns

Column Name	Description
Probe Set ID	The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP probe sets only).
Affx SNP ID	The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP). (SNP probe sets only, not available for Axiom™ Genome-Wide Human Array).
dbSNP RS ID	The dbSNP ID that corresponds to this probe set or SNP. The dbSNP at the National Center for Biotechnology Information (NCBI) attempts to maintain a unified and comprehensive view of known single nucleotide polymorphisms (SNPs), small scale insertions/deletions, polymorphic repetitive elements, and microsatellites from TSC and other sources. The dbSNP is updated periodically, and the dbSNP version used for mapping is given in the dbSNP version field. For more information, see: http://www.ncbi.nlm.nih.gov/SNP/ (SNP probe sets only).
Chromosome	The chromosome on which the SNP is located on the current Genome Version.
Chromosome Start	The nucleotide base start position where the SNP is found. The genomic coordinates given are in relation to the current genome version and may shift as subsequent genome builds are released.
Chromosome Stop	The nucleotide base stop position where the SNP is found. The genomic coordinates given are in relation to the current genome version and may shift as subsequent genome builds are released.
Strand	Genomic strand that the SNP resides on.
Cytoband	Cytoband location of the SNP derived from the SNP physical map and the chromosome band data provided by UCSC.
Strand Vs dbSNP	Indicates whether the SNP is on the same or reverse strand as compared to dbSNP (SNP probe sets only).
ChrX pseudo-autosomal region	SNPs on the X Chromosome which are mapped to the two pseudo-autosomal region have a value of 1 or 2 in this field. All other SNPs are indicated by 0. A value of “1” indicates that the marker maps to the PAR-1 region and a value of “2” indicates that the marker maps to the PAR-2 region. A value of “0” indicates that the marker does not map to either of the two PAR regions.
Probe Count	The total number of probes in the probe set.
Flank	The nucleotide sequence surrounding the SNP. This is a 33-mer sequence with 16 nucleotides on either end of the SNP position. The alleles at the SNP position are provided in the brackets (SNP probe sets only).

Table 14 Annotations and Columns

Column Name	Description
Allele A, Allele B	At array (or underlying database) design time, the following naming convention is used to assign allele nucleotide bases to the "Abstract" allele codes "A" and "B": 1. SNPs are fixed on the forward strand of the design-time reference genome. 2. For AT or CG SNPs (SNP alleles are A/T or C/G), the alleles are named in alphabetical order (A and C are the "A" alleles, in these cases); 3. For non-AT and non-CG SNPs, allele A is A or T, allele B is C or G; 4. For indels, allele A is -, allele B is the insertion. 5. For multi-base alleles, the alleles are named in alphabetical order. (For [AGT/TTA], AGT would be "Allele A". For [GGT/TTA], GGT would be "Allele A".)
Ref Allele, Alt Allele	The reference allele and alternative alleles are specified according to the current reference genome build. The value of Ref Allele could be "-", which indicates an insertion after the specified position. Otherwise it is the sequence of the allele of the marker which matches the current reference genome. The value of Alt Allele could be "-", which indicates that the variant is a deletion with respect to the current genome build. Otherwise it is the sequence(s) of the allele(s) of the marker which does not match the current reference genome. If neither allele of the marker matches the current genome build sequence, then the value of the Ref Allele is set to "." and the value of the Alt Allele(s) is set to <allele_1>/<allele_2>, where <allele_1> and <allele_2> are the alleles of the marker. If the current genome build position of the marker is unknown then the value of the Ref Allele and Alt Allele are set to "---" to denote missing information.
Associated Gene	SNPs were associated with human genes by comparing the genomic locations of the SNPs to genomic alignments of human mRNA sequences. In cases where the SNP is within a known gene, NetAffx reports the association. Additionally, for genes with exon or CDS annotations, NetAffx reports whether or not the SNP is in an exon, and in the coding region. If the SNP is not within a known gene, NetAffx reports the closest genes in the genomic sequence, and the distance and relationship of the SNP relative to the genes. A SNP is upstream of a gene if it is located closer to the 5' end of the gene and is downstream of a gene if it is located closer to the 3' end of the gene.
Genetic Map	Describes the genetic location of the SNP derived from three separate linkage maps (deCODE, Marshfield, or SLM). The physical distance between the markers is assumed to be linear with their genetic distance. The genetic location is computed using the linkage maps from the latest physical location of the SNP and the neighboring microsatellite markers (SNP probe sets only).
Microsatellite	Describes the nearest microsatellite markers (upstream, downstream and overlapping) for the SNP.
Enzyme Fragment	Lists the enzyme, the restriction fragment containing the SNP and the fragment length. The Whole Genome Assay protocol detects SNPs that are contained within the genomic restriction fragments to simplify the sequence background for genotyping arrays (not available for Axiom Genome-Wide Human Array).
Copy Number Variation	When available, a description of Copy Number Variation Region (CN) probe sets as described by the Database of Genomic Variants (not available for Axiom Genome-Wide Human Array).

Table 14 Annotations and Columns

Column Name	Description
SNP Interference	This column is for Copy Number probe sets. It indicates whether or not a known SNP overlaps a copy number probe (CN probe sets only, not available for Axiom Genome-Wide Human Array).
In Final List	This column annotates extended content for genotyping arrays. A value of “1” indicates that the marker is included in the final version of the library file and a value of “0” indicates that the marker is not included in the final version of the library file (SNP probe sets only, not available for Axiom Genome-Wide Human Array).
% GC	The fraction of bases that are G or C in a window of 250,000 bases to each side of the SNP or CN position. All positions that are nearer to the end than 250,001 are set to the value of the position at 250,001 from that end. Position and chromosome values for SNPs and CN probes were mapped to the position of bases in the FASTA files for the build of the genome used in this release of NetAffx, and these bases were then used for all calculations (not available for Axiom Genome-Wide Human Array).
Heterozygous Allele Frequencies	Describes the heterozygous frequency of the allele from Yoruba, Japanese, Han Chinese and CEPH studies using the Affymetrix genotyping arrays. (SNP probe sets only)
Allele Sample Size	Sample size used for Allele Frequency estimates (SNP probe sets only).
Allele Frequencies	Describes the major and minor frequency of the allele from Yoruba, Japanese, Han Chinese and CEPH studies using the Affymetrix genotyping arrays (SNP probe sets only).
Minor Allele	Indicates the Minor Allele of a SNP (SNP probe sets only).
Minor Allele Frequency	The Minor Allele Frequency of a SNP (SNP probe sets only).
OMIM ID	Furnishes OMIM and Morbid Map IDs and their respective gene titles. This database contains information from the Online Mendelian Inheritance in Man® (OMIM®) database, which has been obtained under a license from the Johns Hopkins University. This database/product does not represent the entire, unmodified OMIM® database, which is available in its entirety at www.ncbi.nlm.nih.gov/omim/.
Ordered Alleles	A list of alleles alphabetically ordered by abstract allele code. For bi-allele probesets, the order is "Allele A // Allele B" For multi-allele probesets, the order is "Allele A // Allele B // Allele C // Allele D", with the number of alleles equal to Allele Count.
Allele Count	The number of measured alleles. For bi-allele probesets the count is 2. For multi-allele probesets, the count is 3 or more.

SNP Summary Table Definitions

Use the table of definitions below to help select your SNP Summary Table columns.

Note: Most SNP Summary Table metrics are calculated only for the subset of samples reported by the column 'gender_metrics'. Refer to the gender_metrics description for more information.

Note: Some of the metrics defined in the following table are not displayed by default in the SNP Summary Table. Select **Show/Hide Columns** or **Apply View** > **All Columns View** to access more metrics. Some metrics are reported only after selecting the option(s) **Generate advanced metrics** or **Run PS Supplemental**, when choosing to **Regenerate SNP Metrics**. Some metrics are reported only if multi-allele probesets are present, or if CN-aware genotyping was performed.

Note: A few metrics are calculated only for multi-allele probesets. Most metrics are calculated only for biallele probesets.

Table 15 SNP Summary Table Metrics

Column Name	Description
ConversionType	Probeset classification
CR	Call rate (CR) is the percentage of samples with a genotype call other than "No Call" for the SNP. Note: Call Rate on non-pseudoautosomal regions of chromosome X is reported only for female samples. Call Rate on chromosome Y is reported only for male samples.
MinorAlleleFrequency	The allele frequency for the A allele is calculated as: $P_A = \frac{(\#\,AA\ Calls + 0.5 * AB\ Calls)}{Total\ \#\ Calls}$ Where the Total # Calls does not include the No Calls. The B allele frequency is . $P_B = 1 - P_A$ The minor allele frequency is the Min(P_A, P_B).

Table 15 SNP Summary Table Metrics

Column Name	Description
H.W.p-Value	Hardy Weinberg p-value is a measure of the significance of the discrepancy between the observed ratio or heterozygote calls in a population and the ratio expected if the population was in Hardy Weinberg equilibrium. There are two statistical tests used for HWE. When AA, AB, and BB counts are all >=10, a Chi-squared test is used. When one or more of the AA, AB, and BB counts are <10, an Exact test is used. An Exact test means that the p-value is calculated exactly and not approximated from a population distribution. $$x^2 = \frac{(f^2 aa - fa)^2}{f^2 aa} + \frac{(2faafbb - fab)^2}{2faafbb} + \frac{(f^2 bb - fb)^2}{f^2 bb}$$ Where: $$fa = \frac{(\#AA\ Calls)}{Total\ \#\ Calls}$$ $$fb = \frac{(\#BB\ Calls)}{Total\ \#\ Calls}$$ $$faa = \frac{(\#AA\ Calls + 0.5 * \#\ AB\ Calls)}{Total\ \#\ Calls}$$ $$fbb = \frac{(\#BB\ Calls + 0.5 * \#\ AB\ Calls)}{Total\ \#\ Calls}$$ $$fab = \frac{(\#AB\ Calls)}{Total\ \#\ Calls}$$ $$P_{HW} = CDF(x^2)$$ Where CDF is the Cumulative Distributive Function for the chi-squared distribution. The Exact test used is the one implemented in R package "HardyWeinberg" for more information see: [2] Haldane, J., 1954. An exact test for randomness of mating. J. Genet. 52 631-635. [3] Levene, H., 1949. On a matching problem arising in genetics. Ann. Math. Stat. 20 91-94.
H.W.statistic	H.W.statistic is 1 if H.W.p-Value is calculated using an exact test. H.W.statistic is 0 if the chi-squared test is used. Refer to H.W.p-Value definition (above) for more information.
minGenotypeFreqPval	The minimum probability that any of the clusters present for a probeset have more observations than the supplied expected values. If the ConversionType is 'UnexpectedGenotypeFrequency' then minGenotypeFreqPval is smaller than the SNP QC parameter genotype-p-value-cutoff. The metric minGenotypeFreqPval is available for supported arrays if a genotype frequency file is supplied for SNP QC, and if the genotyped sample batch size is at least as large as the SNP QC parameter ‘min-genotype-freq-samples’.

Table 15 SNP Summary Table Metrics

Column Name	Description
minGenotypeFreqCluster	The genotype cluster for which minGenotypeFreqPval is reported.
FLD	Fisher's Linear Discriminant (FLD) is a measure of the cluster quality of a probeset. High-quality probeset clusters have well-separated centers, and the clusters are narrow. FLD is measured from the SNP posteriors produced during genotyping. FLD is undefined if either the heterozygous or one of the homozygous clusters is empty.
HomFLD	HomFLD is a version of FLD computed for the homozygous genotype clusters. HomFLD is undefined for probesets without two homozygous clusters.
HomFLD_hap	HomFLD_hap is a version of HomFLD computed for the haploid genotype clusters, but only for samples specified by gender_metrics. HomFLD_hap is undefined for probesets without two haploid clusters.
HetSO	Heterozygous Strength Offset measures how far the heterozygous cluster center sits above the homozygous cluster centers in the Size dimension (Y position). Low HetSO values are produced either by misclustering events or by the inclusion of samples that contain variations from the reference genome. Most well-clustered diploid SNPs have positive HetSO values.
HomRO	Homozygote Ratio Offset is the distance to zero in the Contrast dimension (X position) from the center of the homozygous cluster that is closest to zero. If there is only one homozygous cluster, HomRO is the distance from that cluster center to zero on the Contrast dimension.
HomRO_hap	HomRO_hap is a version of HomRO computed for haploid clusters, but only for samples specified by gender_metrics. HomRO_hap will report -10 otherwise.
nMinorAllele	The count of minor alleles, which is one for each heterozygous call, one for each minor haploid call, and two for each minor homozygous call.
Nclus	The number of genotype clusters.
n_AA	The number of AA calls.
n_AB	The number of AB calls.
n_BB	The number of BB calls.
n_A	The number of A calls (haploid).
n_B	The number of B calls (haploid).
n_CN0	The number of ZeroCN calls (zero copy number).
n_NC	The number of NoCall calls, including NoCall_1 (haploid).
AA.meanX	Average Contrast (X position) for AA cluster.
AA.meanY	Average Size (Y position) for AA cluster.
AB.meanX	Average Contrast (X position) for AB cluster.
AB.meanY	Average Size (Y position) for AB cluster.
BB.meanX	Average Contrast (X position) for BB cluster.

Table 15 SNP Summary Table Metrics

Column Name	Description
BB.meanY	Average Size (Y position) for BB cluster.
hemizygous	Hemizygous flag is 1 if the probeset measures chromsome Y or mitochondrial DNA, indicating that diploid genotypes are not possible. Otherwise the flag is 0.
BB_dis_x_adj	Smallest Contrast (X position) distance between any BB cluster sample and any neighbor cluster sample, adjusted to only comparing the sample pairs that have a smaller Size (Y position) distance than y_restrict parameter.
AA_dis_x_adj	Smallest Contrast (X position) distance between any AA cluster sample and any neighbor cluster sample, adjusted to only comparing the sample pairs that have a smaller Size (Y position) distance than y_restrict parameter.
freq_diff	A-allele frequency difference between male and female samples.
HomHet	HomHet flag is 1 if, when two diploid genotype clusters are present, one cluster is homozygous and the other is heterozygous. Otherwise the flag is 0.
BB.varX	Contrast (X position) variance for BB cluster.
BB.varY	Size (Y position) variance for BB cluster.
AB.varX	Contrast (X position) variance for AB cluster.
AB.varY	Size (Y position) variance for AB cluster.
AA.varX	Contrast (X position) variance for AA cluster.
AA.varY	Size (Y position) variance for AA cluster.
BB.varX.Z	Contrast (X position) variance Z-score for BB cluster.
BB.varY.Z	Size (Y position) variance Z-score for BB cluster.
AB.varX.Z	Contrast (X position) variance Z-score for AB cluster.
AB.varY.Z	Size (Y position) variance Z-score for AB cluster.
AA.var.X.Z	Contrast (X position) variance Z-score for AA cluster.
AA.var.Y.Z	Size (Y position) variance Z-score for AA cluster.
CopyNumIssue	CopyNumIssue flag is 1 if the probeset is categorized as 'Other' because of unusual locations of the haploid or ZeroCN clusters. Otherwise the flag is 0.
BestProbeset	BestProbeset flag is available when multiple probesets are mapped to the same SNP (Affy-SNP-ID) by a ps2snp file. A probeset is selected based on the priority order of the conversion types. BestProbeset flag is 1 when it is the best or only probeset for a SNP. Otherwise the flag is 0.
BestandRecommended	BestandRecommended flag is 1 if BestProbeset is 1 and the ConversionType belongs to the Recommended set of conversion types. Otherwise the flag is 0.

Table 15 SNP Summary Table Metrics

Column Name	Description
gender_metrics	Not all probeset metrics are reported using all samples. Number of clusters (Nclus) is calculated on all samples. See the Call Rate (CR) definition for samples used. The remaining metrics are calculated using the sample set described in gender_metrics. If gender_metrics is 'all', all samples are used. If gender_metrics is 'female', only female samples are used. If gender_metrics is 'male', only male samples are used. If gender_metrics is 'diploid', then haploid and ZeroCN samples are excluded before calculation.
MinMean_cp_2	For multi-allele probesets, the minimum mean log2 allele signal for samples that appear to have at least two copies of that allele. If this value is not between the thresholds for the SNP QC parameters 'min-mean-cp2-cutoff' and 'max-mean-cp2-cutoff', then the probeset fails this QC test.
minSigBgndDiffHom	For multi-allele probesets, minimum difference between mean log2 signal and background for the homozygous genotype clusters. If this value is less than the threshold for the appropriate SNP QC parameter 'sign-diff-hom-1-cutoff' or 'sign-diff-hom-2-cutoff', then the probeset fails this QC test.
count_ma_A, count_ma_B, count_ma_C, …	For multi-allele probesets, the count of the associated allele. For example, if there are only two samples with the calls AA and AC, then count_ma_A is 3 and count_ma_C is 1.
Call Modified	Call Modified flag is True if any calls for this probeset are changed since the batch results were first created. If no calls are changed the flag is False.