applied biosystems

Axiom™ Analysis Suite v5.4

USER GUIDE

Publication Number MAN0027928
Revision M00

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

ThermoFisher SCIENTIFIC

Affymetrix, Inc.
3450 Central Expressway
Santa Clara, CA 95051

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.

Revision history: Pub No. MAN0027928

Revision	Date	Description
M00	July 2024	v5.4 release. Enabled Remote Copy Number Genotyping for arrays with library packages that support it. Changed Gender Calling to Sex Calling. Revision 14 is now M00 to comply with new numbering guidelines.
13	February 2023	v5.3 release. Initial release in Thermo Fisher Scientific document control system. Supersedes legacy Affymetrix publication number 703307. Documented new Multiplate Normalization and Dual Workflow features.
12	April 2022	v5.2 release. Updated to include best practices copy number reference creation for fixed region copy number library packages.
11	June 2020	v5.1 release. Enabled Rare het adjustment. Fixed genomic position for indel in VCF export.
10	January 2020	v5.0.1 release. Updated multiallelic SNP QC options in Threshold Settings window. Added filter for multiallelic markers.
9	November 2019	v5.0 release. Added Whole Genome View feature and updated multiallelic SNP metrics.
8	June 2019	v4.1 release. Copy Number QC Metrics have been updated to include two additional metrics.
7	September 2018	v4.0.1 release. VCF export updates. Added non-human copy number variation analysis.
6	July 2018	v4.0 release. Added CN Analysis capabilities.
5	August 2017	v3.1 release. Added SNP QC call frequency check.
4	January 2017	v3.0 release. Added predefined region CN calls using females as plate controls. Multiallele probeset genotyping, visualization and Y,MT probeset classification. Rebranded user guide.
3	August 2016	v2.0 release. Added PharmacoScan array and star allele translation.
2	July 2015	v1.1 release. Added dynamic probeset selection genotyping if new v5 CDF have wobble probesets.
1	March 2015	Initial v1.0 release.

Important Software Licensing Information

Your installation and/or use of this Axiom Analysis Suite software is subject to the terms and conditions contained in the End User License Agreement (EULA) which is incorporated within the Axiom Analysis Suite software, and you will be bound by the EULA terms and conditions if you install and/or use the software.

Legal entity

Affymetrix, Inc.| Santa Clara, CA 95051 USA | Toll Free in USA 1 800 955 6288

TRADEMARKS

Contents

Contents

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 Probeset level.
- Run Copy Number algorithms on select array types.
- View Cluster Plots with the ability to change calls and/or highlight by attribute.
- Export your Data in various formats for use in 3rd party software.

Software and hardware requirements

Table 1 Requirements

64-bit Operating System	Speed	RAM	Available Disk Space[1]	Web Browser
Microsoft Windows 11 (64 bit) Professional	3.8 GHz Intel Pentium Quad Core Processor	32 GB	150 GB HD + data storage Refer to table below.	Internet Explorer 8.0 and higher

[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:** Batches were run with the default setting **Reduce the size of the genotyping batch** =**True**[2]

Table 2 Sample Data size estimates and required disk space

# 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	235 MB	290 MB	729 MB	1.24 GB	5.52 GB
	Total	**1.56 GB**	**2.94 GB**	**14 GB**	**27.84 GB**	**138.52 GB**
500K	Input	1.33 GB	2.66 GB	13.3 GB	26.6 GB	133 GB
	Output	1.05 GB	1.57 GB	5.69 GB	10.7 GB	52 GB
	Total	**2.38 GB**	**4.23 GB**	**18.99 GB**	**37.3 GB**	**185 GB**
850K	Input	1.33 GB	2.66 GB	13.3 GB	26.6 GB	133 GB
	Output	1.97 GB	2.88 GB	10 GB	19.1 GB	90.2 GB
	Total	**3.3 GB**	**5.54 GB**	**23.3 GB**	**45.7 GB**	**223.2 GB**

[2] A batch 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.

Validated space requirements and run times for CN batches

Table 3 Testing performed on a 96f 885k marker array using Windows 11, 32 GB RAM system

Workflow	Space Required	Run Time
Initial Copy Number Reference Creation	199 MB	1h 14min
Best Practices Copy Number Reference Creation Step 1 (6 plates)	3.93 GB	1h 31min
Best Practices Copy Number Reference Creation Step 2 (6 plates)	38.4 MB	2h 30min

Table 4 5000 UKBioBankv2 sample batch on a 32 GB system

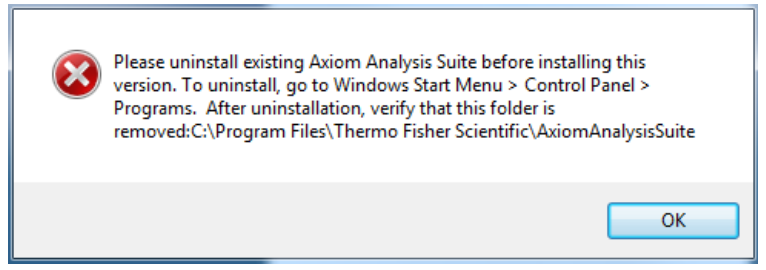
Workflow	Space Required	Run Time
Copy Number Fixed Regions	79.3 MB	18h 35min
Copy Number Discovery	60.2 GB	28 hours
Copy Number Reference Creation	106 GB	23 hours

Installation

1. Go to https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/applications/predictive-genomics/population-genomics/software.html
2. Locate the Axiom Analysis Suite software package, then click **Download now**.
3. Unzip the file, then double-click **AxiomAnalysisSuiteSetup.exe**.
4. Follow the on-screen instructions to complete the installation.

   Note: If your system has a previous version installed, an uninstall warning message appears. (Figure 1)

Figure 1 Uninstall required message



Acknowledge the message by clicking **OK**, then go to "Uninstalling" on page 19.

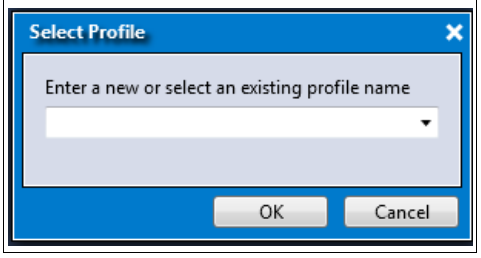
Starting

1. Double-click on the 

Desktop shortcut or click **Start** → **All apps** → **Thermo Fisher Scientific** → **Axiom Analysis Suite**.

   The Select Profile dialog window appears. (Figure 2)

Figure 2 Select profile window

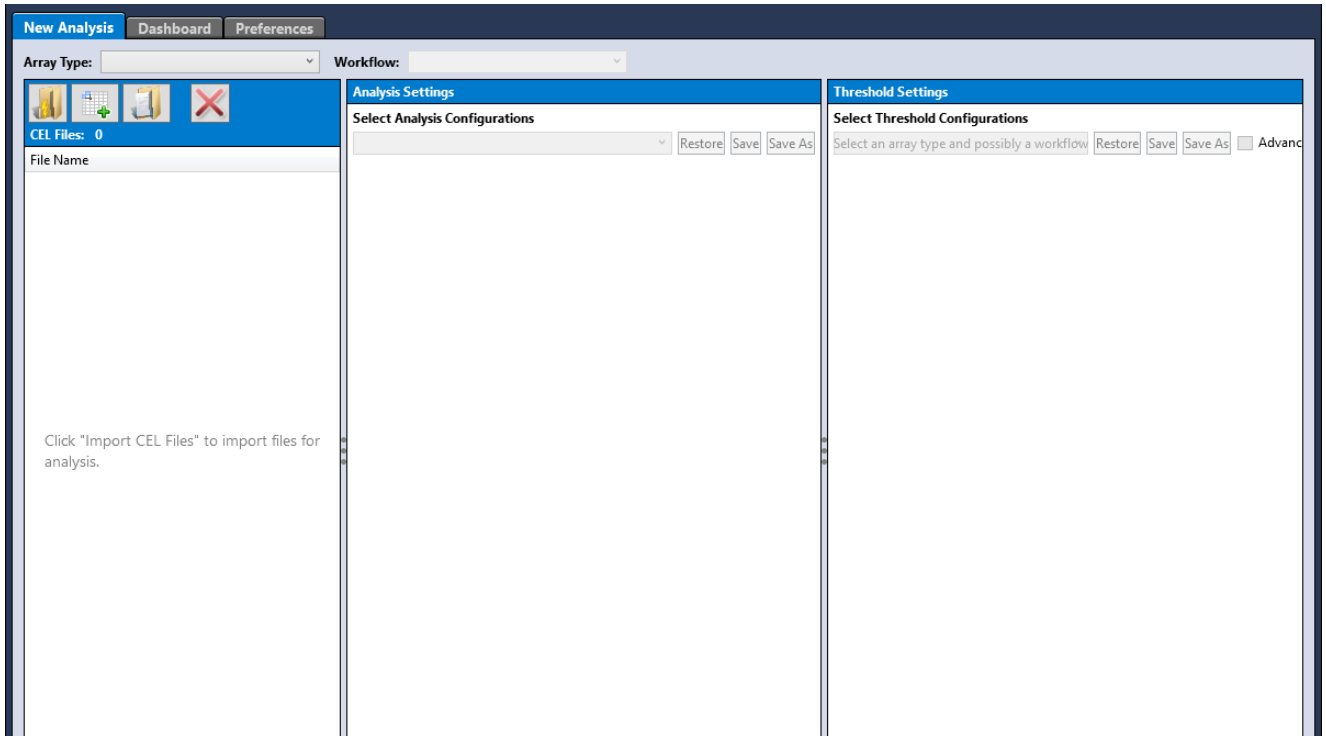


2. Enter a new profile name or click the drop-down to select an existing profile name.
3. Click **OK**.

   The main window appears. (Figure 3)

   Note: A Library Folder Updates dialog may appear. If it does, acknowledge the available updates by clicking **OK**. To disable this feature, see "Turning off library file update alerts" on page 16.

Figure 3 Main window

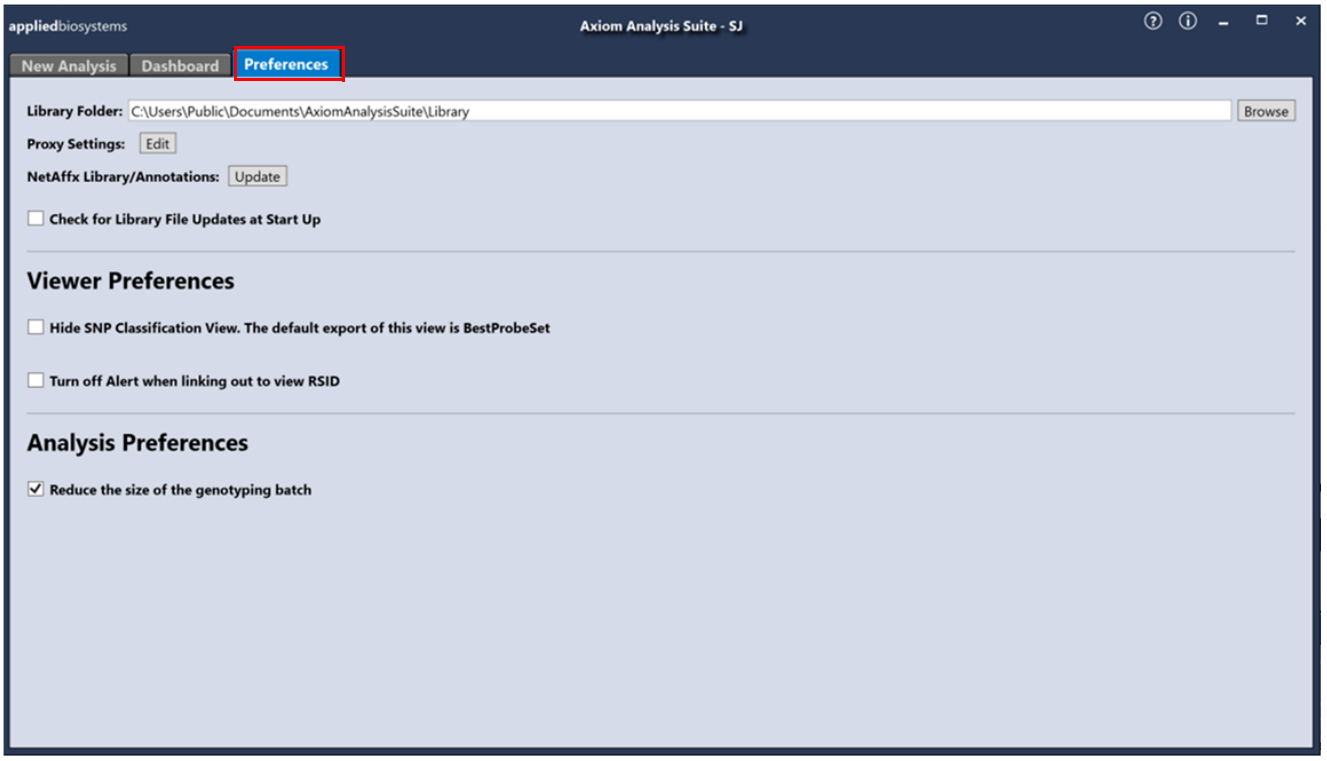


Preferences

1. Click the **Preferences** tab.

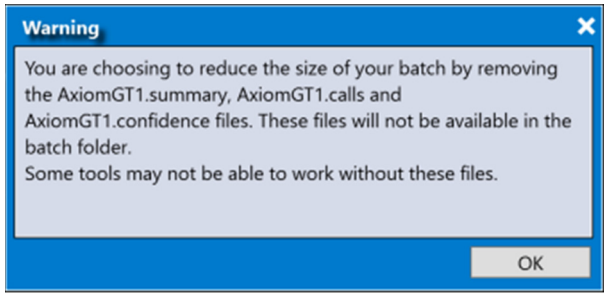
   The Preferences window tab appears. (Figure 4)

Figure 4 Preferences window



Note: The **Reduce the size of the genotyping batch** check box is checked by default. If it is unchecked, then checked again, the following Warning appears. (Figure 5)

Figure 5 Warning



Read the warning, then click **OK**.

General preferences

Changing the default library folder/path

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

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

   The **Select Library Folder** window appears.
2. Navigate to a new library file folder location.
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 6.

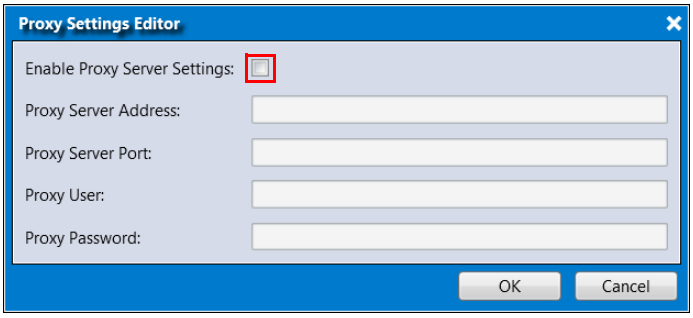
Figure 6 Populated Library Path example

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.

The Proxy Settings Editor window appears. (Figure 7)

Figure 7 Proxy Settings Editor window



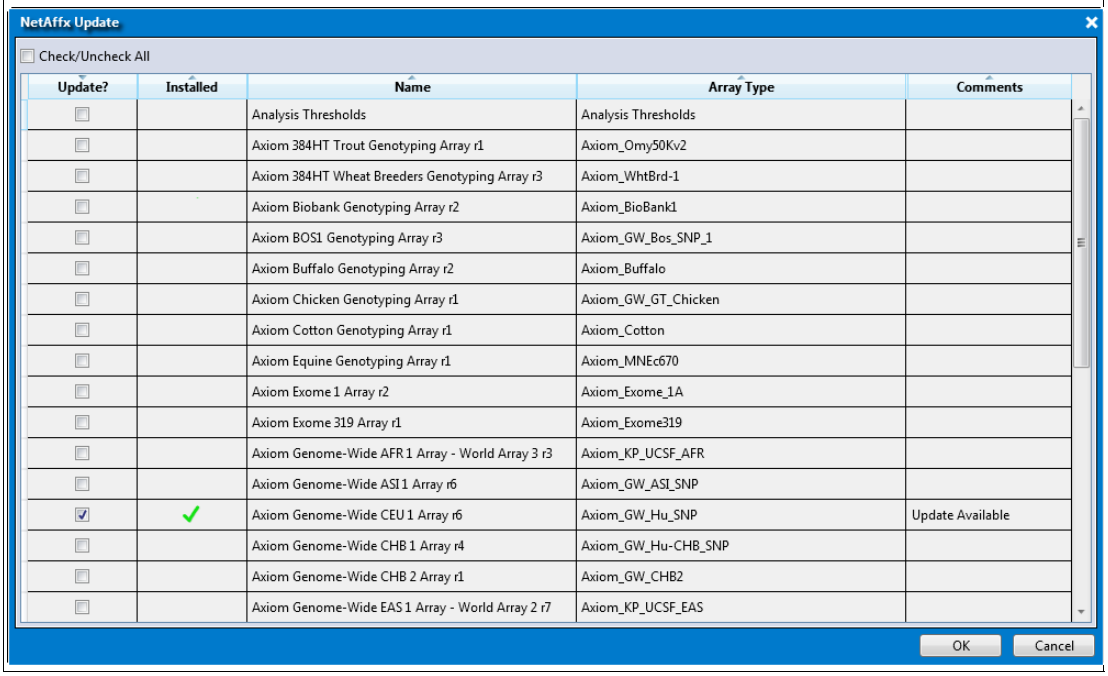
5. Click the **Enable Proxy Server Settings** check box (Figure 7), then contact your IT department for help with completing the required text fields.
6. Click **OK**.

Updating NetAffx library/annotations

1. Click the **Update** button.

   The NetAffx Update window appears. (Figure 8)

Figure 8 NetAffx Update window



2. 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.

3. Click **OK**.

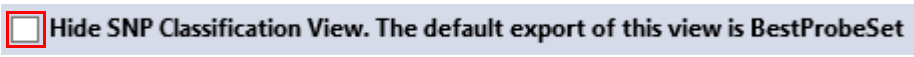
   An Installing Updates progress bar appears.

Turning off library file update alerts

This feature is on (checked) by default. Uncheck to turn it off.

Viewer preferences

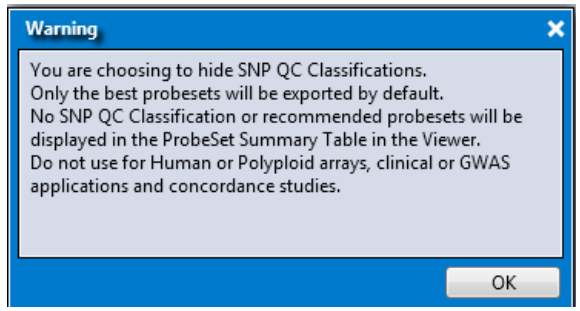
Hide SNP Classification View



1. To hide the SNP classification view, click its check box.

   A Warning message appears. (Figure 9)

Figure 9 Warning message



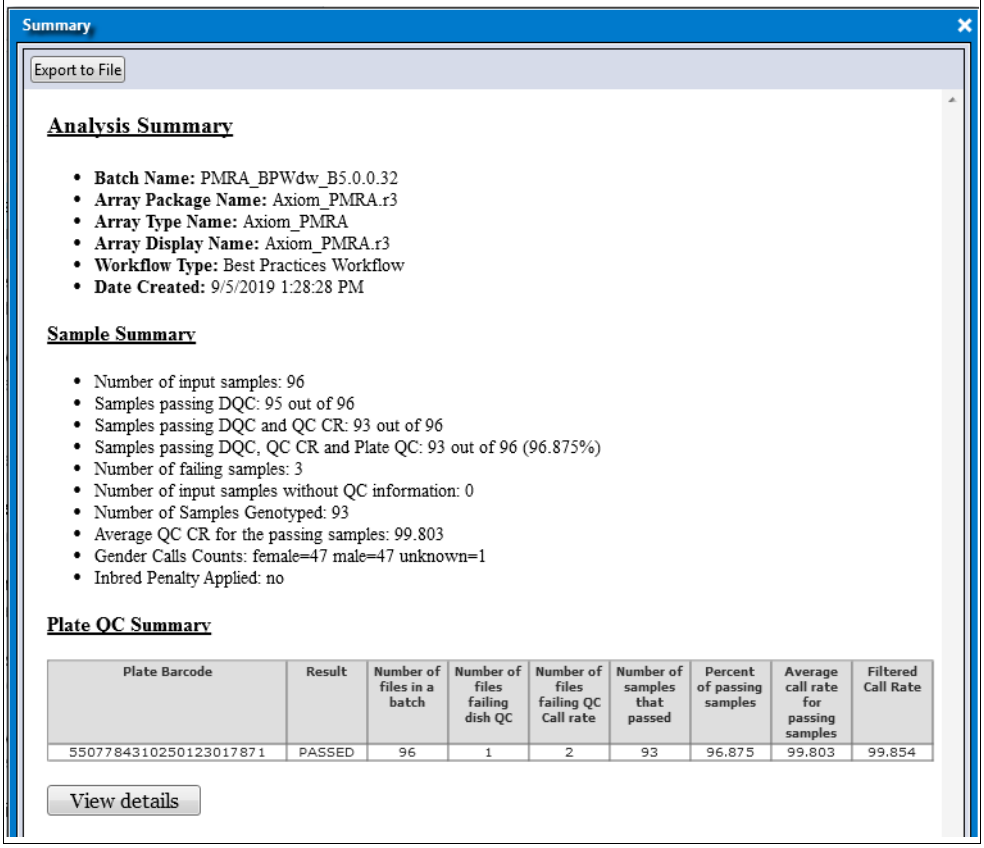
2. Acknowledge the message, then click **OK** to enable this hide feature or close the message window to cancel.

Viewing a batch with Hide SNP Classification View enabled

1. Click on the Dashboard widow tab.
2. Click **Actions** → **Open**.

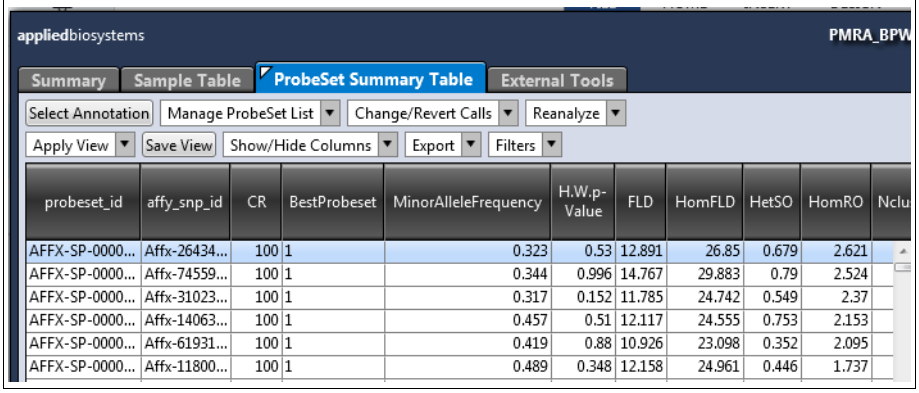
   The batch loads and appears in a Summary window. The Summary tab will not contain any Probeset or Marker Metric information, as shown in Figure 10.

Figure 10 Summary window



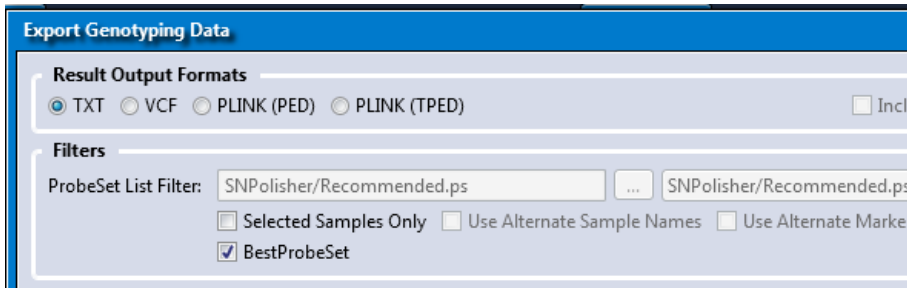
The Probeset Summary Table will not show the Conversion Type column, as shown in Figure 11.

Figure 11 Probeset Summary Table



Best Probeset is the default filter when exporting Genotyping data, as shown in Figure 12.

Figure 12 Export Genotyping Data window



Note: Hide SNP Classification View does not change how workflow is performed. Standard Best Practices is still run, however the SNP conversion types or count is not shown. To view full batch information, simply close out the batch and uncheck Hide SNP Classification, then reopen the batch.

Toggling the linking out to view RSID alert

By default, the alert when linking out to view RSID is set to **ON**.

1. To turn OFF this linking out alert, click its check box.

   Note: For select genotyping probesets, AxAS enables you to view corresponding TaqMan Assays by linking out via RSID.

IMPORTANT! Make sure the most current annotation file is loaded in your Probeset Summary tab, before linking out.

Installing custom array library files

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

1. Download the zip package provided to you by Thermo Fisher 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.

1. Click **Windows Start** → **All apps** → **Settings** → **Apps** → **Installed apps**.
2. Click on **Axiom Analysis Suite**, then click the **Uninstall** button.
3. Follow the on-screen instructions.
4. After the uninstall process is complete, close all open windows.
5. Use Windows Explorer as you normally would to navigate to the directory: **C:\Program Files\ThermoFisherScientific**
6. Verify that the **Axiom Analysis Suite** folder has been removed.
7. If the folder is present, double-click on it to open it.
8. Search for any files you want to keep, then move them to different (easily accessible) location.
9. Delete the **AxiomAnalysisSuite** folder.
10. Close all open windows, then install the new version, as described in the "Installation" on page 12.

2 Performing an analysis

Setting up 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 13.

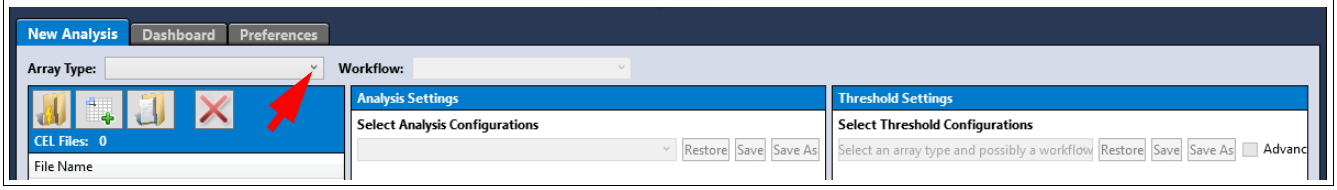
For Array Types that are CN enabled, go to "Copy number analysis" on page 98 for analysis set up. If not, continue to "Selecting an array type".

Note: If you want to run and view CN-aware genotypes, go to page 191.

Selecting an array type

1. From the main AxAS window tab, click the **Array Type** drop-down (Figure 13), then select the array type you want to use.

Figure 13 Main window



Selecting a workflow

Workflows are driven by array type, therefore the arrays residing in your library folder is what determines what type of workflow is available in the Workflow drop-down.

Note: For cases with rare heterozygous calls, the Best Practices and Genotyping workflows include an updated genotyping algorithm to distinguish true calls from false. For more details, see the Rare Heterozygous Adjusted Genotyping Tech Note.

Note: Remote Copy Number is enabled for arrays with library packages that support it. For more information, see the *Axiom Genotyping Solution Data Analysis User Guide* (P/N MAN0018363).

1. Click to select the workflow you want to use.
   - **Best Practices Workflow:** 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 (#MAN0018363)*.
   - **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.

- **Summarized Signal Intensity:** This workflow produces a summary of the intensities for the probe sets. **Note:** Summarized Signal Intensity does not perform sample QC nor genotyping.
- **Initial Copy Number Reference Creation:** This workflow generates a Reference file using the Axiom Training Plate as input samples.
- **Best Practices Copy Number Reference Creation:** This workflow generates a Reference file using customer samples, the Initial Copy Number Reference file and performs quality control to ensure the file generated is robust.
- **Copy Number Reference Creation:** This workflow generates a Reference file using all samples imported into the cel file pane.
- **Copy Number Discovery:** This workflow performs whole genome copy number analysis on imported CEL files. A sex check is performed at the beginning of the run.
- **Copy Number Fixed Regions:** Copy Number analysis is performed on targeted regions.

Importing files

Files can be imported in three ways:

- Import CEL Files
- Import Samples and Attributes by text
- Import by ARR

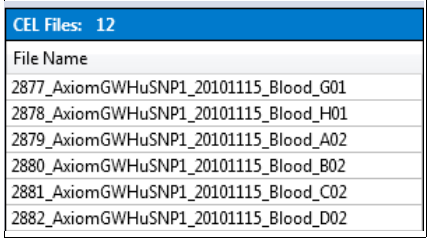
Importing CEL files

1. Click the **Import CEL Files**  icon.

   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 files. (Figure 14)

Figure 14 Populated CEL File pane example



Importing CEL files by text

IMPORTANT! To include your own sample names and plate names, make sure you include a column title, alternate sample name, and plate name, as shown in Figure 15.

1. Click **Import Samples and Attributes by TXT**  icon.

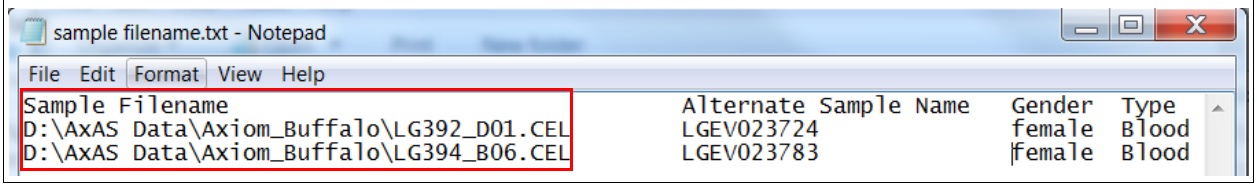
   The Import Samples and Attributes by TXT window appears.

2. Navigate to the TXT file that contains the list of CEL files you want to process.

   Note: If the text file contains sample attributes, they will be visible in the Viewer, after the batch has been analyzed. The very first column of the text file header must be either **cel_files** or **Sample Filename**. The CEL files must be entered exactly as they are named. Example: LG392_D01.CEL

IMPORTANT! If the text file is not located in the same folder as the CEL files, you must include the full path of the text file (under the Sample Filename column) before the CEL filename, as shown in Figure 15.

Figure 15 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. Click **Open**.

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

Import CEL by ARR

1. Click **Import Sample and Attributes by ARR**  icon.

   An Import Samples and Attributes by ARR window appears.

2. Navigate to your ARR files folder.

3. Select the appropriate ARR files, then click **Open**.

   Your CEL Files pane populates and displays each ARR file extracted from your selected ARR file. After the batch has been successfully analyzed, the Sample attribute information in the ARR will be displayed in the viewer.

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 contents of the **Analysis Settings** pane is based on the array type and workflow selected. For example, if Genotyping mode is selected, Sample QC is hidden and the Genotyping section is displayed.

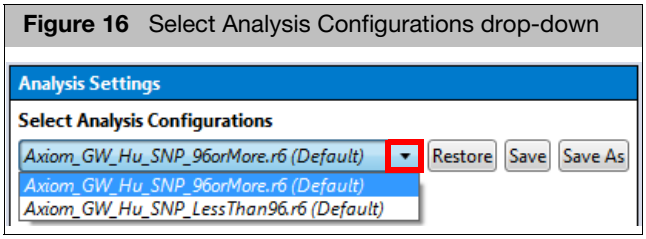
Note: Workflows and analysis configurations are only selectable after an Array Type has been selected.

Selecting an analysis configuration

1. Click the drop-down menu (if available) (Figure 16), then select the option that best matches the number of samples that you want to analyze.

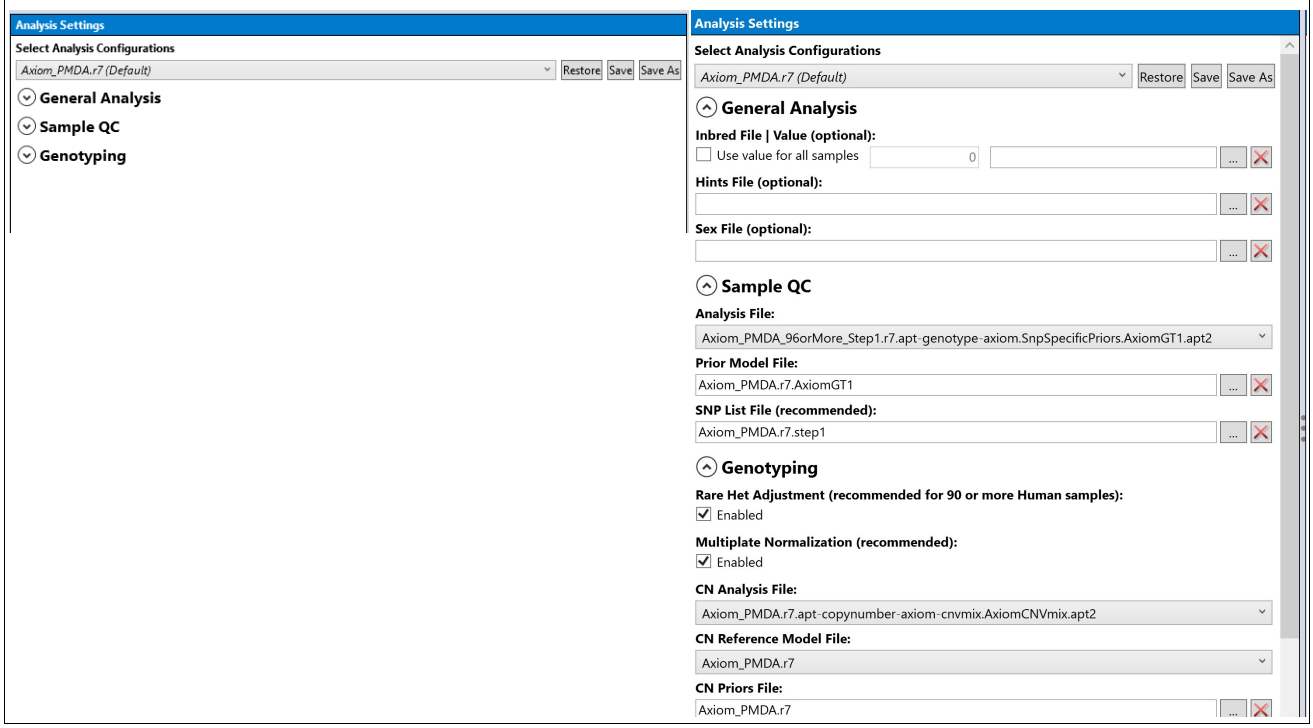
   Note: The default configuration options displayed in the drop-down menu are based on your array type.

Figure 16 Select Analysis Configurations drop-down



After selecting the appropriate default for the number of your samples, the Analysis Setting pane auto-populates. Click on each section's down-arrow to expand its content, as shown in Figure 17.

Figure 17 Analysis Settings pane - Collapsed (left) Expanded (right)



IMPORTANT! By default, the last workflow appears.

Note: For newer library packages, Analysis Settings will contain a General Analysis section.

Using the analysis settings fields

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

General Analysis fields

- To use an inbred penalty file, click the ... button, navigate to your file, then click **Open**. The name of the file appears next to value text box. To remove the file, click the button. **Note**: For Inbred Penalty, click the **Use value for all samples** check box, then enter a value (from 0-16) in the adjacent text box to apply the value to all samples in the batch.
- **Hints file** (Optional): Click the ... button, navigate to your Hints file, then click **Open**. The name of the file appears in the Hints File (optional) field. To remove the file, click the button.
- **Sex file** (Optional): Click the ... button, navigate to your Sex file, then click **Open**. The name of the file appears in the Sex File (optional) field. To remove the file, click the button.

Note: The term *Gender* has been replaced with *Sex* in the software user interface. The algorithm to determine sex still uses gender, therefore the input file remains the same.

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**.

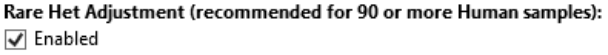
   Your newly assigned filename is displayed.
4. (Optional) Click the **SNP List File** Browse ... button.

   The SNP List File window appears.
5. Navigate and select the appropriate file, then click **Open**.

   Your newly assigned filename is displayed.

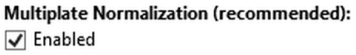
Genotyping fields

Note: Rare Het Adjustment is checked (enabled) by default for human arrays. It can reduce false positive heterozygous calls and increase sensitivity when genotyping rare variants.



If using less than 90 samples, uncheck this check box. It is not recommended to use with nonhuman arrays.

Note: Multiplate Normalization is checked (enabled) by default. For batches with multiple plates, it can increase the number of probesets that pass SNP QC.



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.

   Note: If you want to change the currently displayed Model file, click the appropriate Multiallele **Browse** button to select a different Model file.
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 **Posterior File Name** Browse button.

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

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

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

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

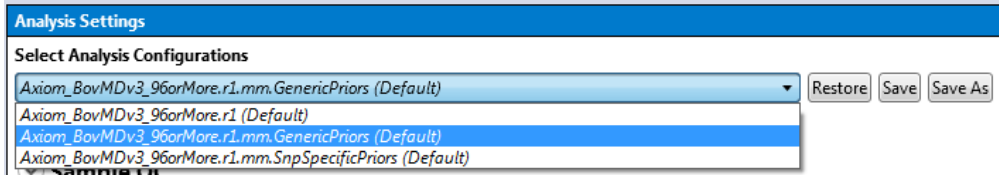
    The Genotype Frequency window appears.
11. Navigate to a location for your genotype frequency file, enter a name, then click **Open**.

    Your assigned filename is displayed.

Setting up a multiallelic enabled analysis

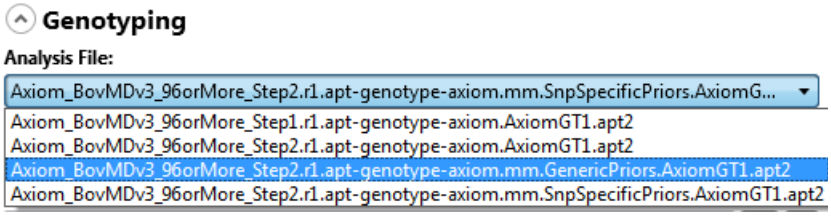
1. At the Analysis Settings pane, click the **Select Analysis Configurations** drop-down to select the appropriate multiallele configuration file, as shown in Figure 18.

Figure 18 Select Analysis Configuration drop-down



2. Expand the Genotyping section, then in the Analysis File drop down, select the appropriate **Step2** file, as shown in Figure 19.

Figure 19 Genotyping Analysis File drop-down



Note: The ps2snp file must be a multiallelic file.

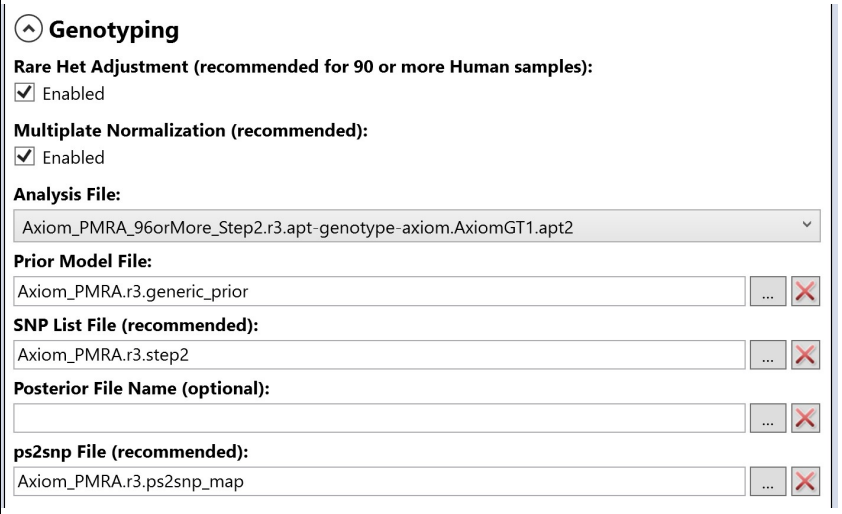
3. (Optional) If not already displayed by default, click the ... button.

   An Explorer window appears.

4. Navigate to the library package folder and manually select the appropriate ps2multisnp_map file, the click **Open**.

   The file is now displayed. (Figure 20)

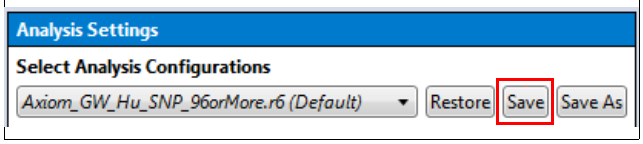
Figure 20 ps2snp File



Saving your analysis configuration

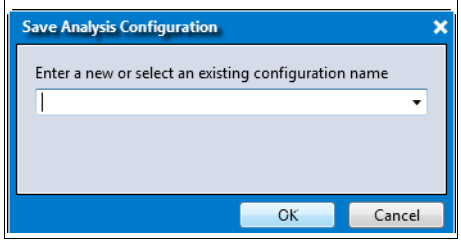
1. After editing your Analysis Configuration settings, click Save (top of Analysis Setting pane). (Figure 21)

Figure 21 Analysis Configuration Save window



If the Analysis Configuration that came with the library package already exists, then a **Save Analysis Configuration** window appears. (Figure 22)

Figure 22 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.

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

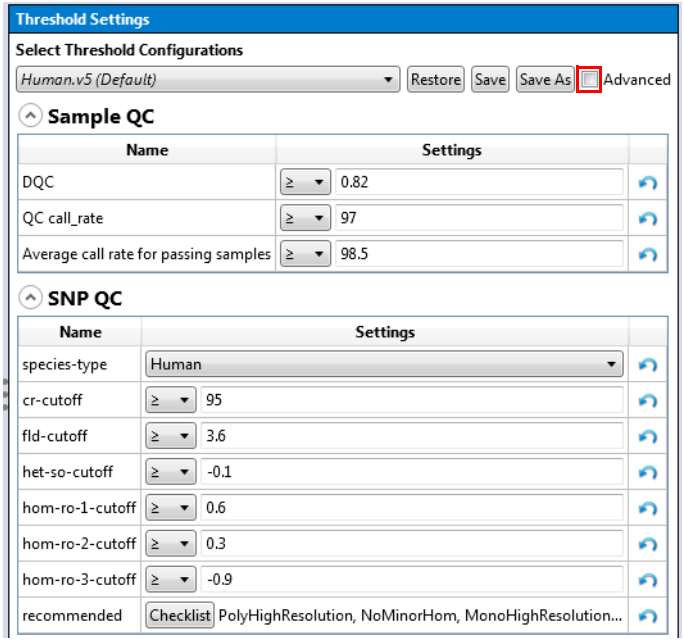
- Click the applicable File field's **Browse** button to navigate to a different location, then click **Open** to reassign its path.
- If needed, click a File field's button to delete a displayed path setting.
- Click the **Restore** button to return to the last saved values of the analysis configuration 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

IMPORTANT! The Threshold Setting Configuration will be blank when an Array Type is selected that has not yet been run. If an Array Type has been run, the Threshold Configuration displays the settings last used.

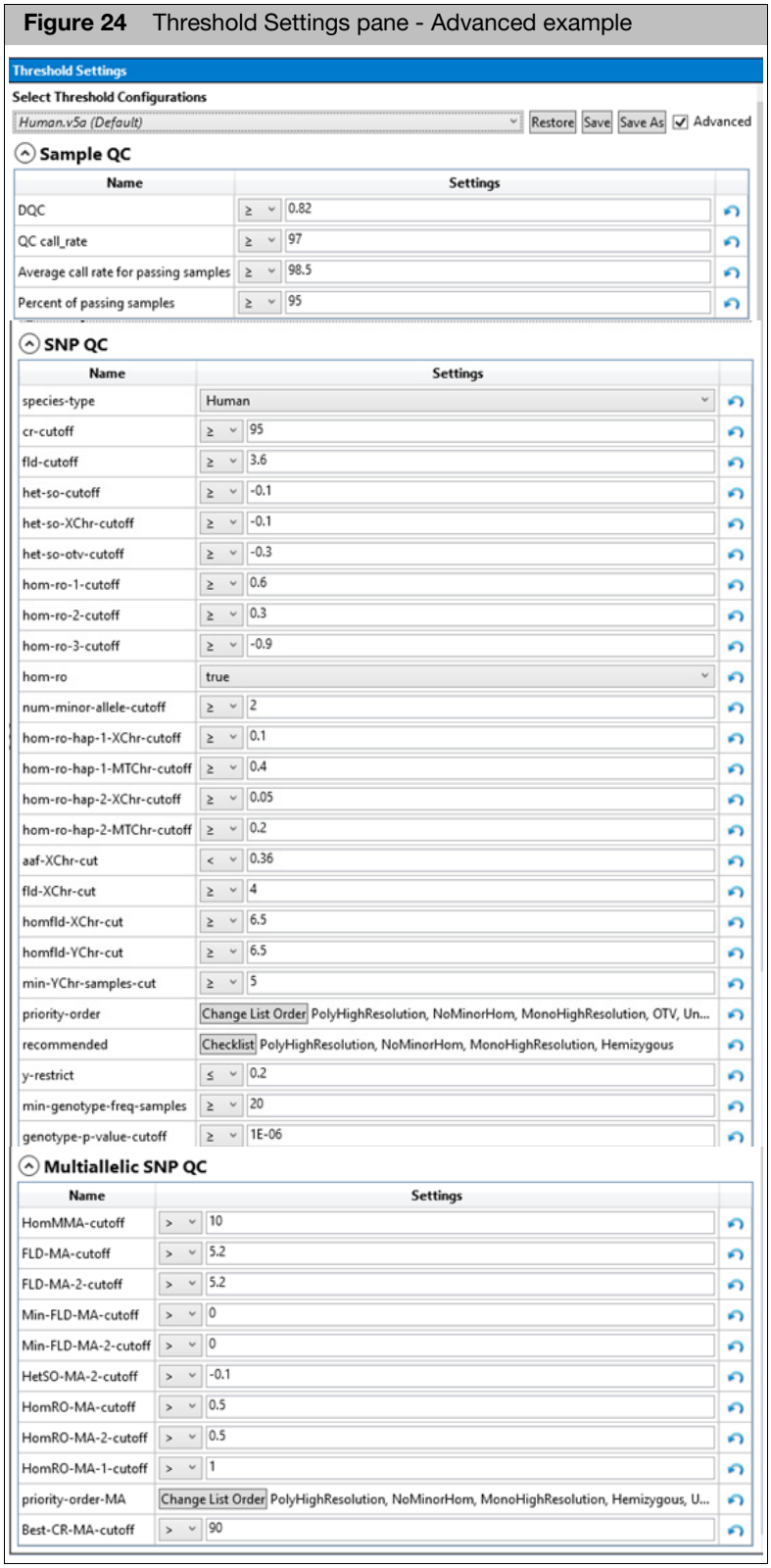
- The settings shown in the **Threshold Setting** pane (Figure 23) are based on the selected array type and workflow.
- Expanding the Sample QC and SNP QC drop-down arrows displays commonly edited threshold parameters.
- For Sample QC, SNP QC, and Multiallelic SNP QC name definitions, see "Threshold names" on page 216.

Figure 23 Threshold Settings pane example



- Check the **Advanced** check box for a complete listing of thresholds. (Figure 24)
  Note: Advanced mode may display thresholds that are not applicable to your analysis.

Figure 24 Threshold Settings pane - Advanced example



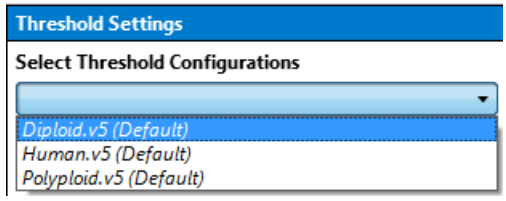
Customizing thresholds

Click the **Select Threshold Configuration** drop-down (Figure 25) to select an appropriate Default Threshold for your starting point.

Note: The comparison signs/operators are preset and cannot be changed.

Note: Newer library file packages will contain an array specific **ax_thresholds** file.

Figure 25 Select Threshold Configurations



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 26.

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

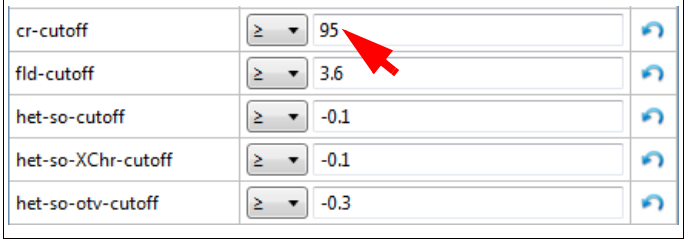
Figure 26 Threshold Name text field example

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 27.

   Note: General Rule: The het-so-otv-cutoff should be less or equal to het-so-cutoff and het-so-XChr-cutoff.

Figure 27 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** 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 28.

Figure 28 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 Change the Priority Order window appears. (Figure 29)

Figure 29 Change the Priority Order window

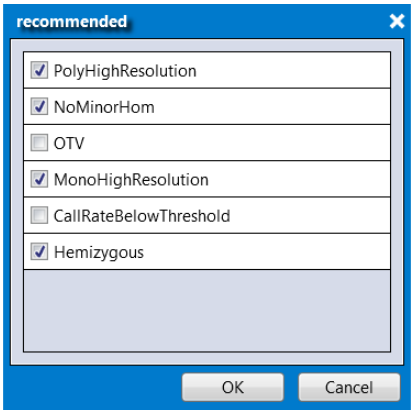
6. Click and hold onto the selection you want to move, then drag and drop it into its new position. After you get the order of priority you want, click **OK**.

   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 30)

Figure 30 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, Hemizygous
   - **Diploid**: PolyHighResolution, NoMinorHom, MonoHighResolution, Hemizygous
   - **Polyploid**: PolyHighResolution

Assigning an output folder path

1. Click the **Output Folder** path's **Browse** button. (Figure 31)

Figure 31 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 to a newly assigned results path folder

Note: To better organize your output results, you can add sub-folders to your newly assigned output 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 folder 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 **Name** field. (Figure 32)

IMPORTANT! Each a name you enter must be unique for the set of batches listed in the Dashboard window tab and unique within the same destination folder.

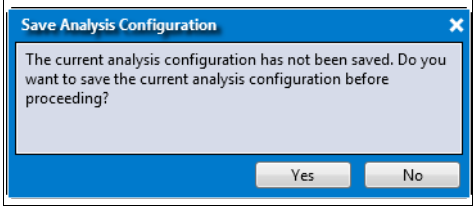
Figure 32 Enter a Name

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

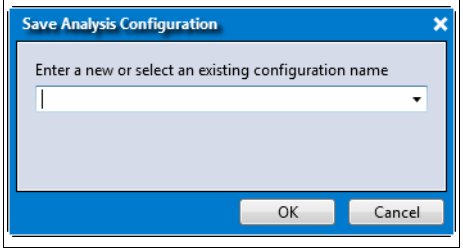
1. Click **Run**.
2. If you have not saved any changes to your configured Analysis Settings, a Save Analysis Configuration window appears. (Figure 33) Click **Yes**.

Figure 33 Save Analysis Configuration prompt window



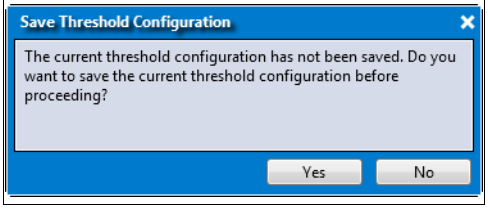
3. Enter a new analysis name or use the drop-down to select a previously saved name, then click **OK.** (Figure 34)

Figure 34 Save Analysis Configuration window



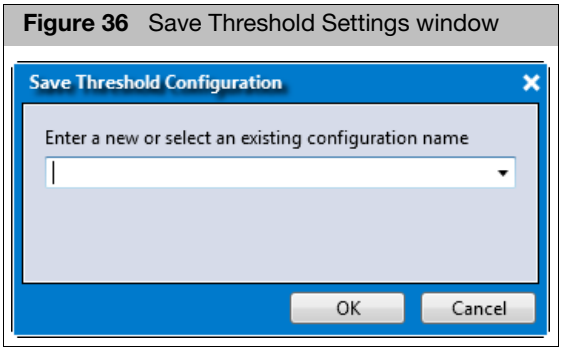
4. If you have not saved any changes to your configured Threshold Settings, a Save Threshold Configuration window appears. (Figure 35) Click **Yes**.

Figure 35 Save Threshold Settings prompt window



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

Figure 36 Save Threshold Settings window



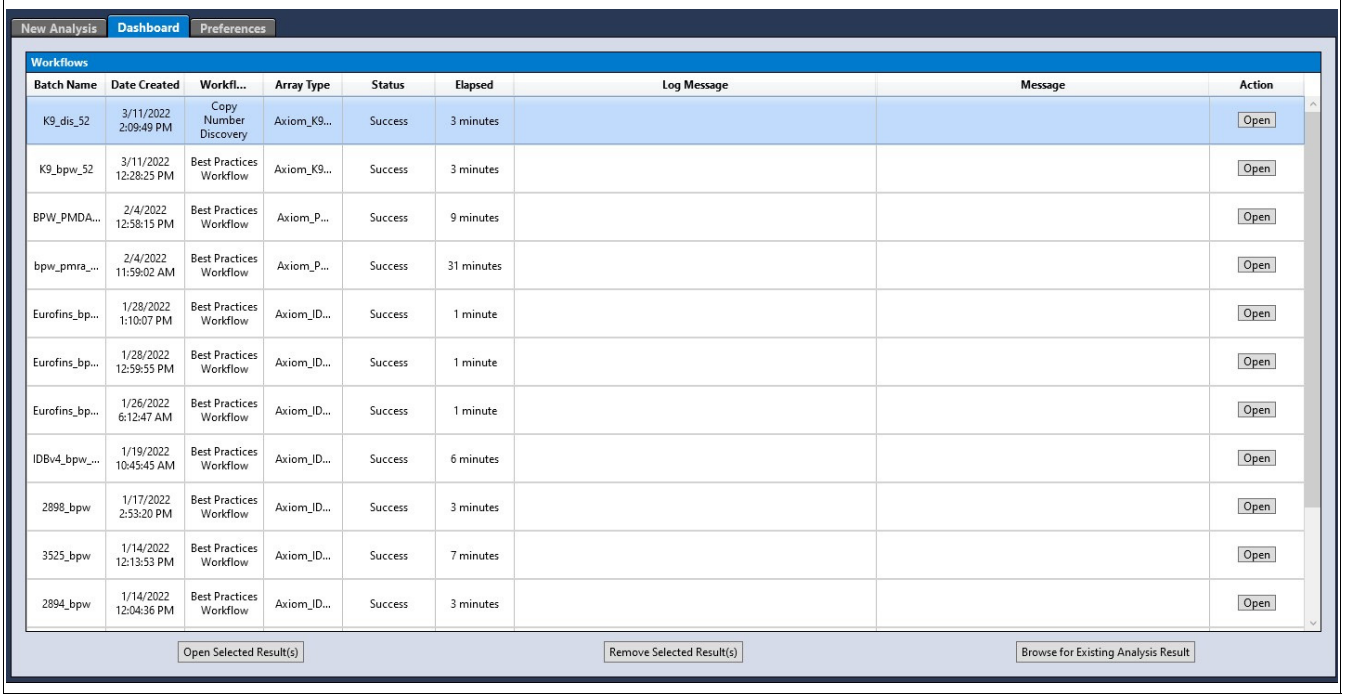
The Dashboard window tab appears. (Figure 37) Its **Status** column displays a green progress bar that monitors the status of your running analysis.

To cancel an analysis in progress, click Stop (far right).

Dashboard window tab

The Dashboard window tab displays existing results, as shown in Figure 37.

Figure 37 Dashboard window



Open selected result(s)

Perform 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)

Perform 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.

   Note: Copy number reference creation cannot be opened.

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).

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

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

   A Select Existing Analysis Result Folder window appears.

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

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

   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, select **Show more options,** then select **Open in Axiom Analysis Suite**.

   A Select Profile window appears.

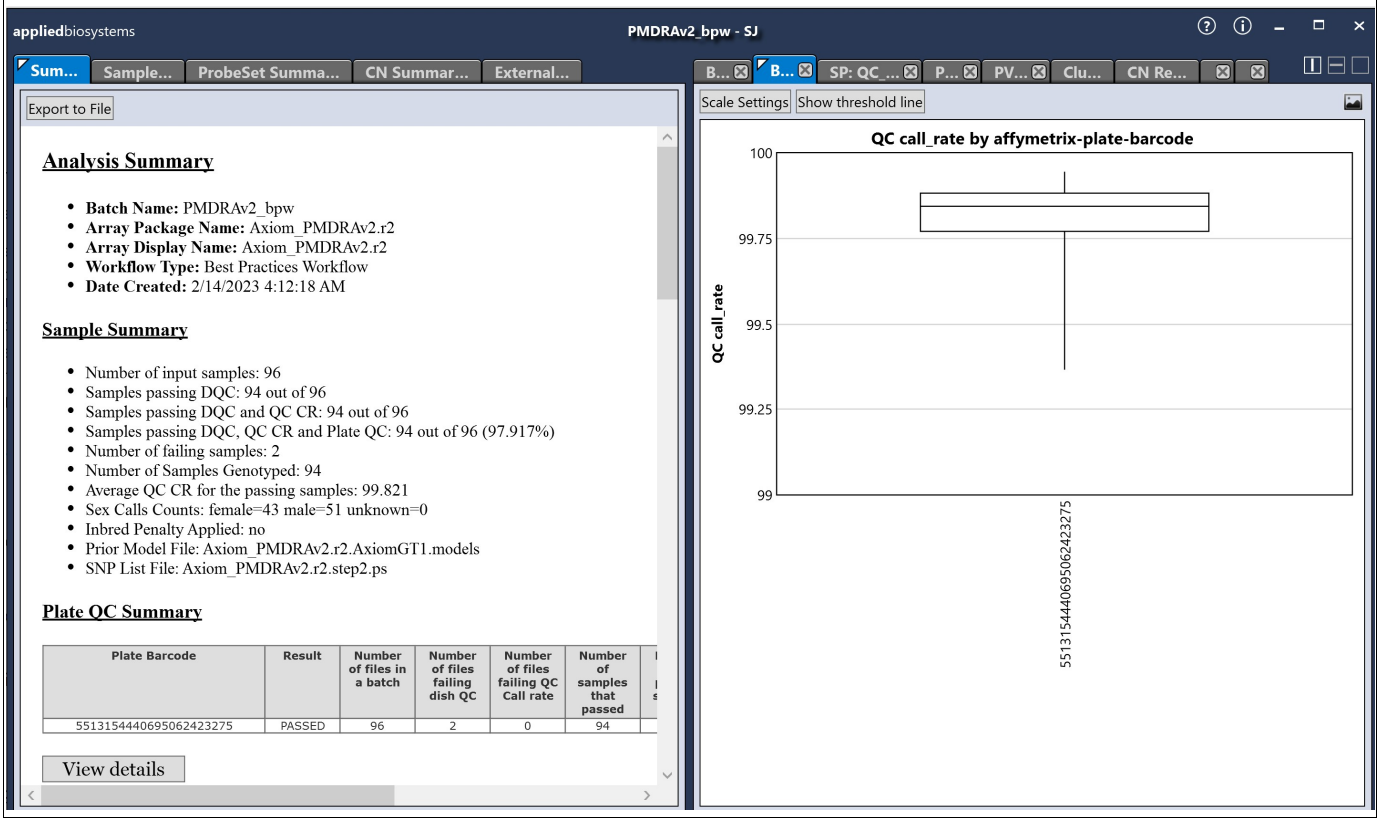
2. At the Select Profile window, enter a new or select an existing profile name.

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

3 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 38)

Figure 38 Main Viewer window



Viewing options

As shown in Figure 38 on page 37, the Viewer (by default) displays a **side-by-side** split-screen configuration.

Split-screen options

Changing side by side split-screen to a top and bottom configuration

1. Click the **Horizontal Split** icon (far right) 

   Click on any window tab to view it in full/single window mode.

Returning to the default side by side split-screen

1. Click the **Vertical Split** icon (far right) 

Changing a tab window to a full screen

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

The window tab is now a window.

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

Returning to the tab window view

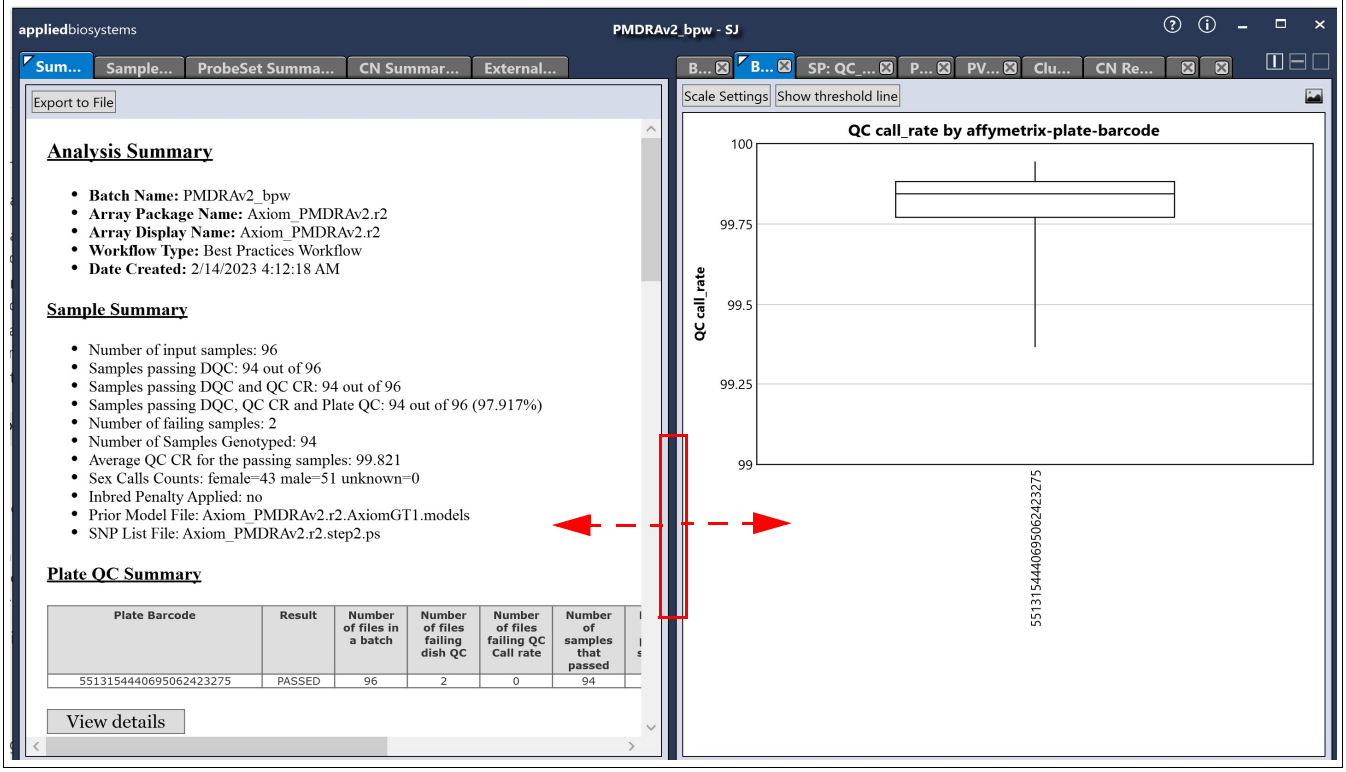
1. Click the **X** button (top right) to close the window.

   The window returns to its default tab window and position.

Sizing a window pane

1. Click, hold, then drag the edge of the window pane (Figure 39) to resize it.

Figure 39 Sizing side by side panes



Summary window tab

The Summary window/tab (Figure 40 and Figure 41) displays a summary snapshot of your analysis, including detailed threshold values, and tables based on your analysis.

Figure 40 Summary window tab

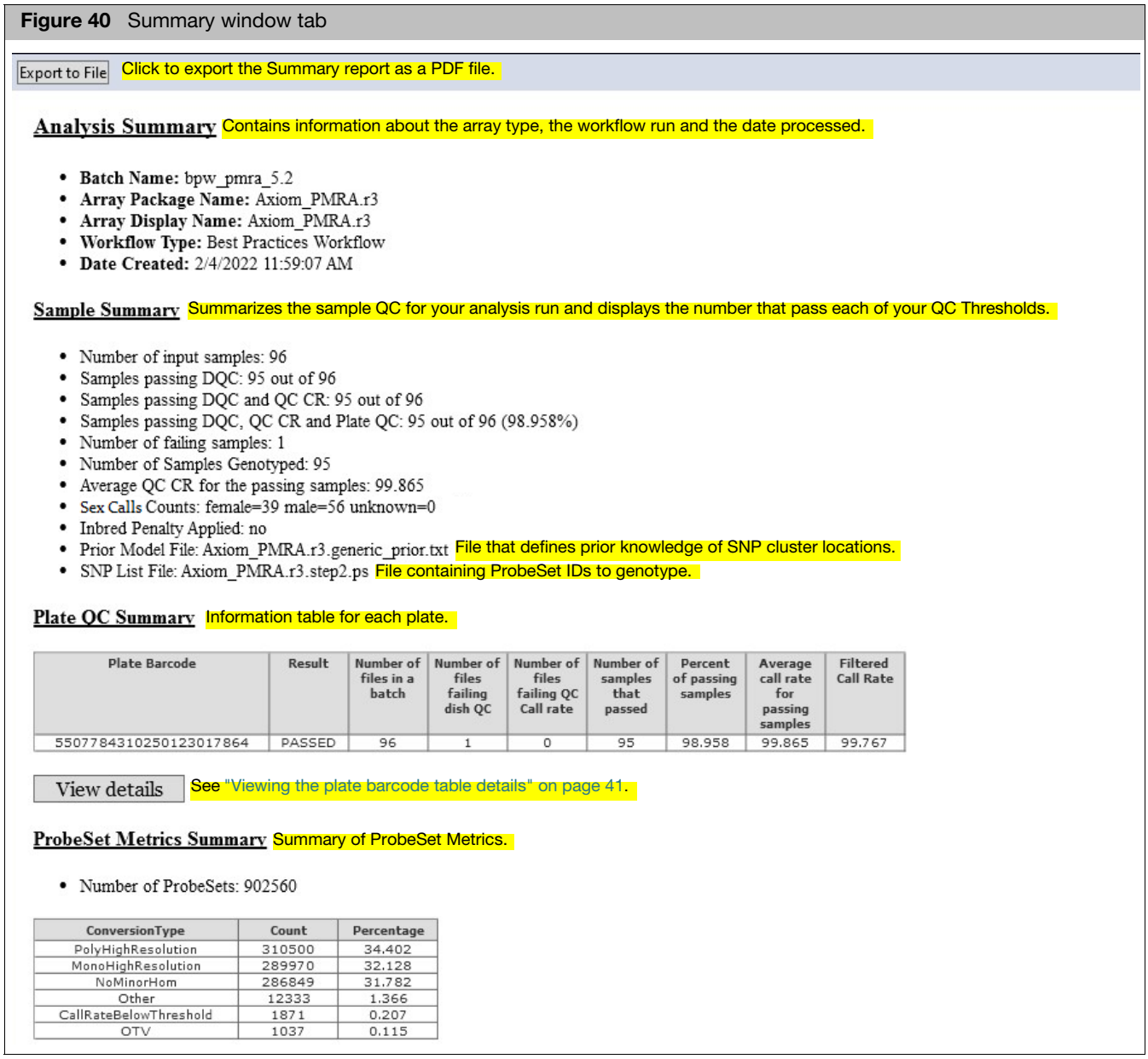


Figure 41 Summary window tab - Continued

Marker Metrics Summary

This summary details the markers by conversion type.
Note: The *Number of BestandRecommended* is the count of probesets in the **Recommended.ps** filter file. This count can be smaller than the sum of recommended ConversionType best probesets because no more than one probeset from a multiallelic marker is selected for the filter file.

- Number of Markers: 888799
- Number of BestandRecommended: 874048
- Percent BestandRecommended: 98.34

ConversionType	Count	Percentage
PolyHighResolution	310177	34.898
NoMinorHom	286708	32.258
MonoHighResolution	277163	31.184
Other	11914	1.34
CallRateBelowThreshold	1829	0.206
OTV	1008	0.113

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

- DQC: ≥ 0.82
- QC call_rate: ≥ 97
- Average call rate for passing samples: ≥ 98.5
- Percent of passing samples: ≥ 95

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

- species-type: Human
- cr-cutoff: ≥ 95
- fld-cutoff: ≥ 3.6
- het-so-cutoff: ≥ -0.1
- het-so-XChr-cutoff: ≥ -0.1
- het-so-otv-cutoff: ≥ -0.3
- hom-ro-1-cutoff: ≥ 0.6
- hom-ro-2-cutoff: ≥ 0.3
- hom-ro-3-cutoff: ≥ -0.9
- hom-ro: true
- num-minor-allele-cutoff: ≥ 2
- hom-ro-hap-1-XChr-cutoff: ≥ 0.1
- hom-ro-hap-1-MTChr-cutoff: ≥ 0.4
- hom-ro-hap-2-XChr-cutoff: ≥ 0.05
- hom-ro-hap-2-MTChr-cutoff: ≥ 0.2
- aaf-XChr-cut: < 0.36
- fld-XChr-cut: ≥ 4
- homfld-XChr-cut: ≥ 6.5
- homfld-YChr-cut: ≥ 6.5
- min-YChr-samples-cut: ≥ 5
- priority-order: PolyHighResolution, NoMinorHom, MonoHighResolution, OTV, UnexpectedGenotypeFreq, CallRateBelowThreshold, Other, OtherMA
- recommended: PolyHighResolution, NoMinorHom, MonoHighResolution, Hemizygous
- y-restrict: ≤ 0.2
- min-genotype-freq-samples: ≥ 20
- genotype-p-value-cutoff: ≥ 1E-06

Multi-Allelic SNP QC Thresholds Displays Multi-Allelic SNP QC Thresholds used in your analysis run (if applicable).

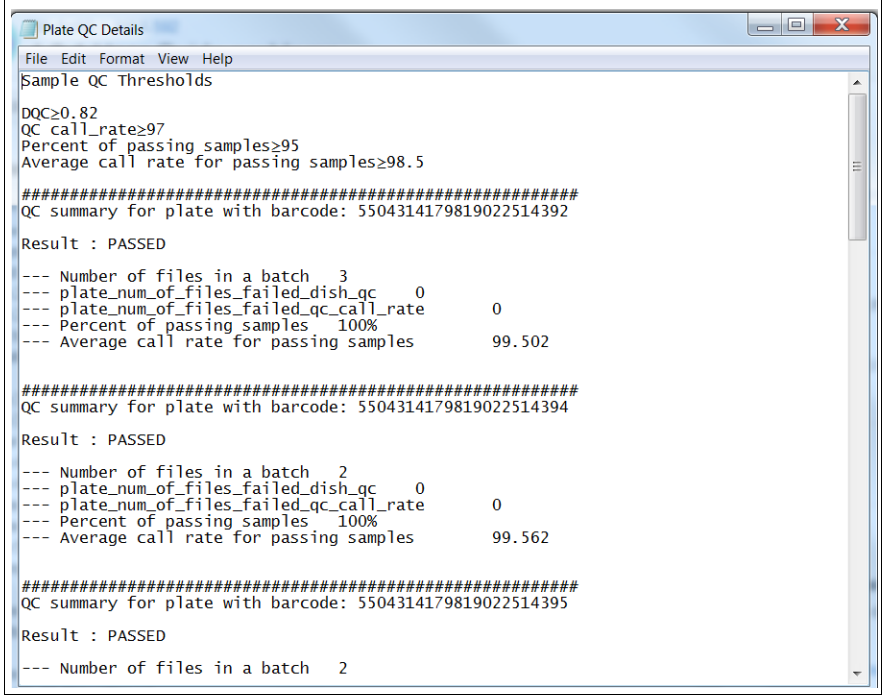
- HomMMA-cutoff: > 10
- FLD-MA-cutoff: > 5.2
- FLD-MA-2-cutoff: > 5.2
- Min-FLD-MA-cutoff: > 0
- Min-FLD-MA-2-cutoff: > 0
- HetSO-MA-2-cutoff: > -0.1
- HomRO-MA-cutoff: > 0.5
- HomRO-MA-2-cutoff: > 0.5
- HomRO-MA-1-cutoff: > 1
- priority-order-MA: PolyHighResolution, NoMinorHom, MonoHighResolution, Hemizygous, UnexpectedGenotypeFreq, CallRateBelowThreshold,
- Best-CR-MA-cutoff: > 90

Viewing the plate barcode table details

1. In the Summary window tab (Figure 40 on page 39), click View Details.

   A window opens and displays a text file version of your Sample QC information (by plate). (Figure 42)

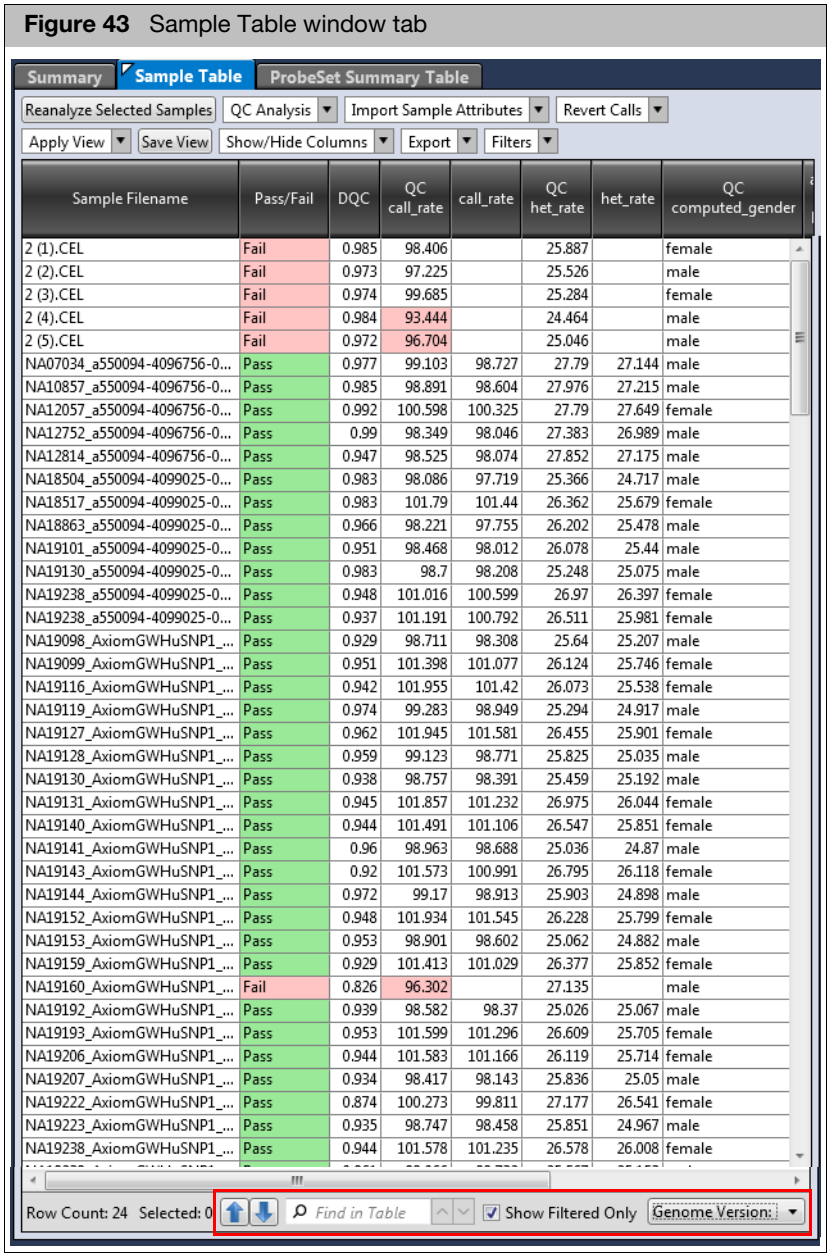
Figure 42 Notepad window



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 43.

Figure 43 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

- Click the **Genome Version** drop-down (bottom right) to view annotation build information.
- Use the **Find in Table** field to locate keywords. See "Searching keywords" on page 56.
- Click the **Show Filtered Only** check box 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).

Importing sample attributes into the 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 or cel_files and include the full CEL file name, as shown in Figure 44. Sample attribute column names must be different than standard column names.

Figure 44 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**.

IMPORTANT! You can use other names for your samples and plates. To do this, add two columns to your text file. Label one column header Alternate Sample Name and the other Plate Name.

Column headers

The default Sample Table column view is as shown. (Figure 45)

Figure 45 Default Sample Table Columns

Showing or hiding table columns

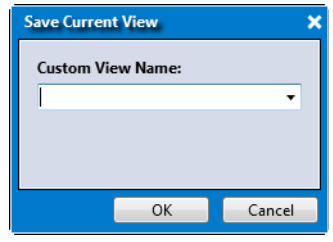
1. Click the **Show/Hide Columns** drop-down menu.
2. Click each available column name's check box to show it or remove it from the table. See "Annotations and columns" on page 226 for their definitions.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

Saving your customized Sample Table column view

1. Click **Save View.**

   The Save Current View window appears. (Figure 46)

Figure 46 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.

Showing 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.

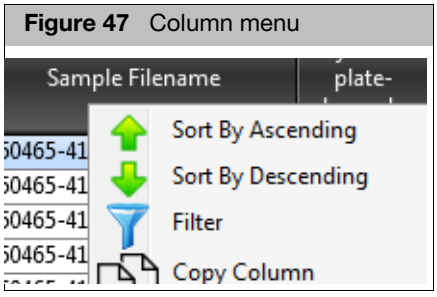
   The column is now in its new table position.

Sorting columns

1. Select a column, then right-click on it.

   The column menu appears. (Figure 47)

Figure 47 Column menu



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

Single-click sorting

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.
2. Click the **Hide Column** check box to remove it from the table.

Filtering column data

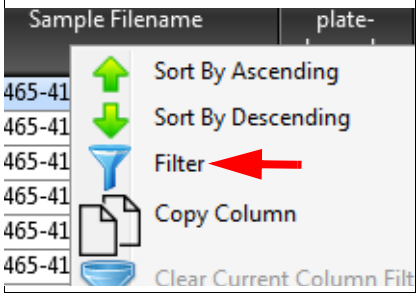
Note: All Sample Table columns are filterable.

Adding filters (method 1)

1. Select a column, then right-click on it.

   The column menu appears. (Figure 48)

Figure 48 Column Menu

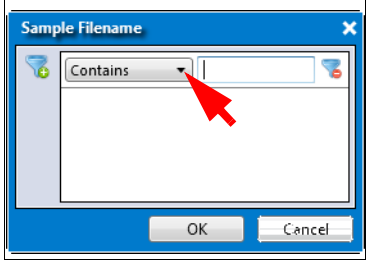


2. Click **Filter**.

Text-based columns

If the column you want to filter contains text-based data, the Contains drop-down menu appears, as shown in Figure 49.

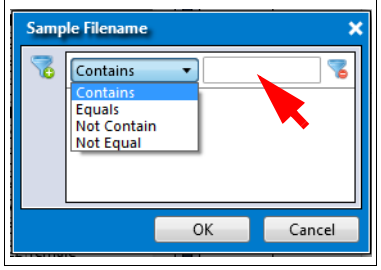
Figure 49 Contains drop-down



Applying a filter to a text-based column

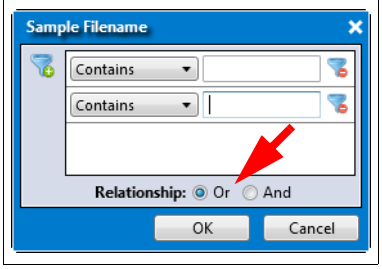
1. Click the **Contains** drop-down menu to select a filtering property.
2. Click inside the text entry box to enter a value, as shown in Figure 50.

Figure 50 Drop-down Menu



3. (Optional) Click to add additional filters.

Figure 51 Or or And Relationship Logic

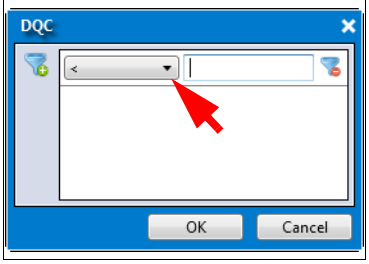


4. Click the **Or** or **And** radio button to choose **Or** or **AND** relationship logic. (Figure 51)
5. Repeat steps 1-4 as needed.
6. To remove a filter(s), click .

Numeric data columns

If the column you want to filter contains numeric data, a symbol drop-down menu appears. (Figure 52)

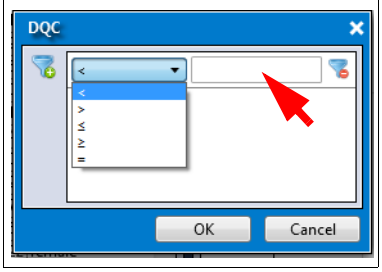
Figure 52 Filter Properties



Applying a filter to a value-based column

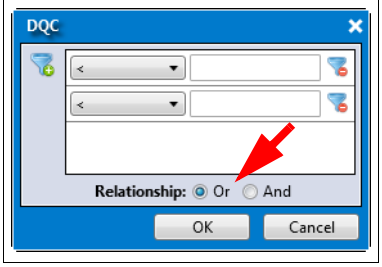
1. Click the **Symbol Value** drop-down menu to select the filtering symbol you want.
2. Click inside the text entry box to enter the value(s), as shown in Figure 53.

Figure 53 Drop-down Menu



3. (Optional) Click to add filter(s). (Figure 54)

Figure 54 Or or And Relationship Logic



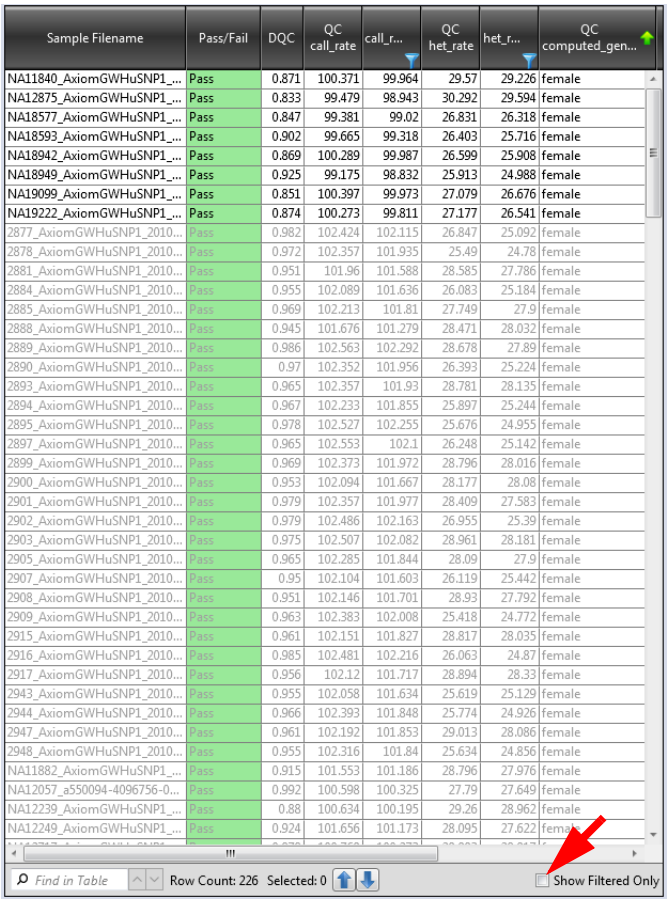
4. Click the **Or** or **And** radio button to choose **Or** or **AND** relationship logic. (Figure 54)
5. If needed, repeat steps 1-4.
6. Click **OK**.

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 55 on page 48.

Figure 55 Sample Table window tab - Show Filter Only unchecked example

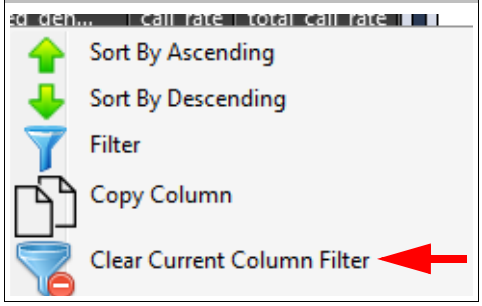


Clearing an individual filter

1. Right-click on the filtered column you want to clear.

   The column menu appears. (Figure 56)

Figure 56 Column Menu



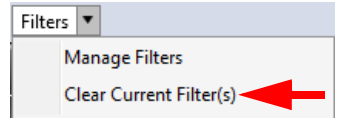
2. Click **Clear Current Column Filter**.

   The filter is removed.

Clearing all current filters

- Click the Filters drop-down, then select **Clear Current Filters**. (Figure 57)

Figure 57 Filters Menu



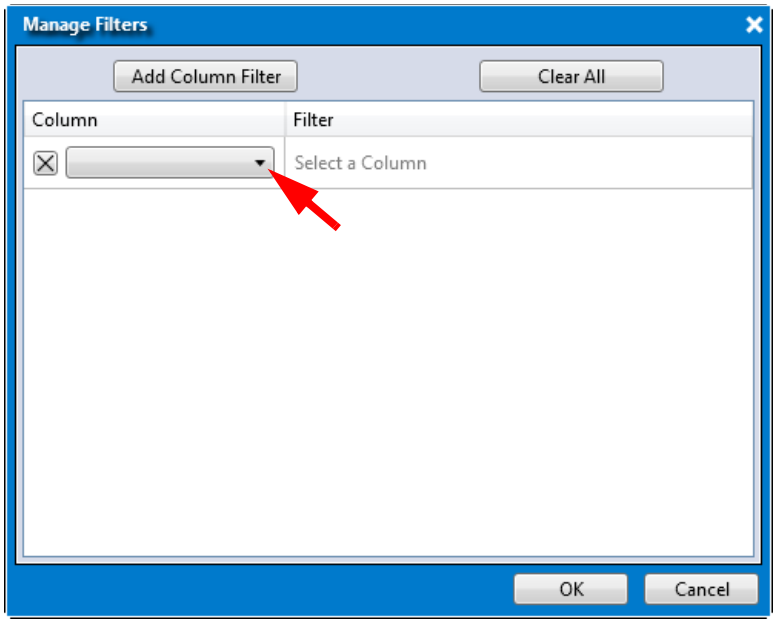
Adding filters (method 2)

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 58)

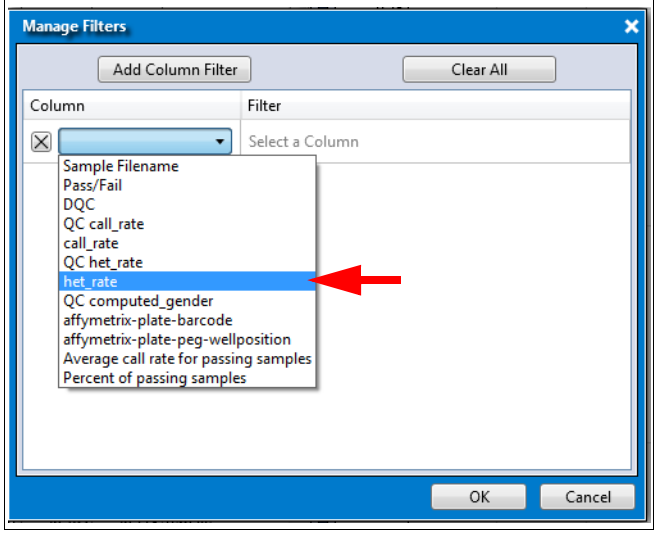
Figure 58 Manage Filters window



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.

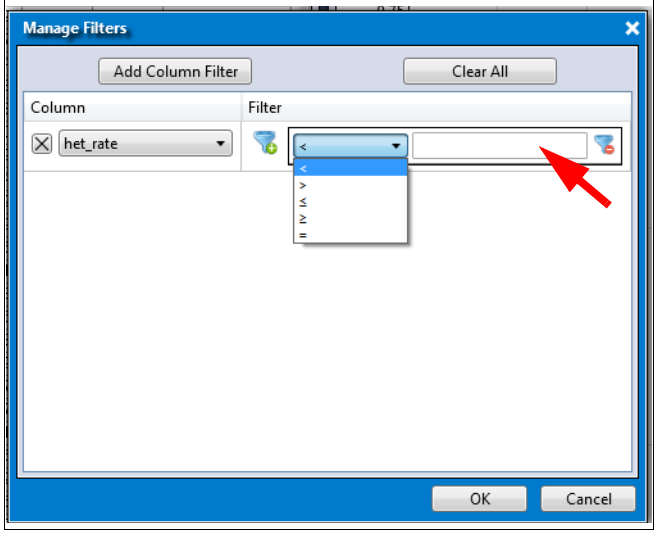
2. Click the **Column** drop-down, then click to select the Column name you want to create a filter for. (Figure 59)

Figure 59 Manage Filters window - Select a Column



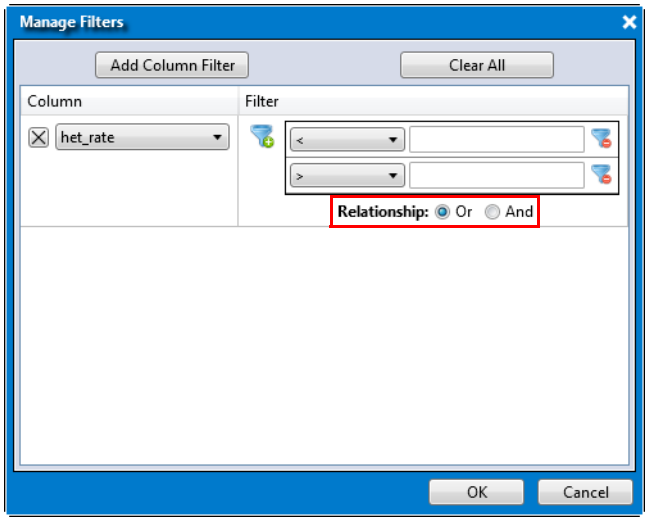
3. Click the **Symbol Value** drop-down menu to select the filtering symbol you want. (Figure 60)

Figure 60 Manage Filters window - Select a Symbol



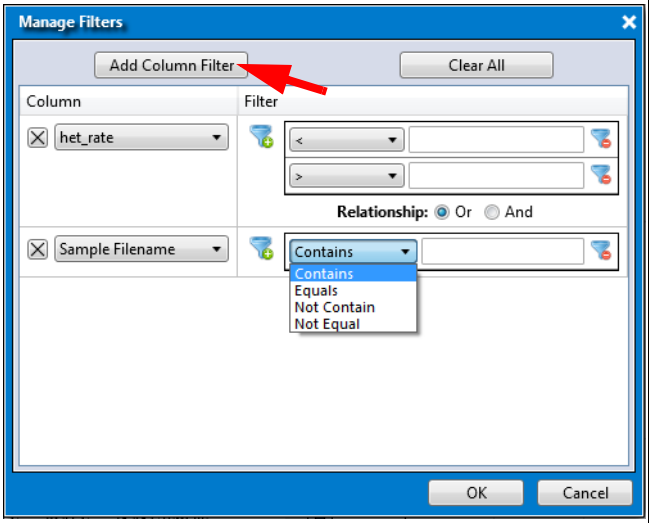
4. Click inside the text entry box to enter new value(s), as shown in Figure 60.
5. (Optional) If you want to add an additional filter to a column, click .
6. Click the **Or** or **And** radio button to choose **Or** or **AND** relationship logic. (Figure 61)

Figure 61 Manage Filters window - OR or AND Relationship



7. If needed, click **Add Column Filter**, then repeat the above steps. (Figure 62)

Figure 62 Manage Filters window - Adding another Column Filter



8. Click **OK**.

   To remove a filter(s), click .

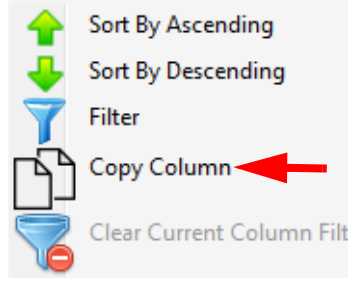
   Click **Clear All** to remove ALL filters in the Manage Filters window.

Copying column data to a clipboard

1. Click to select a column you want to copy to a clipboard, then right-click on it.

   The column menu appears. (Figure 63)

Figure 63 Right-click Column Menu



2. Click **Copy Column**.

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

Setting user colors

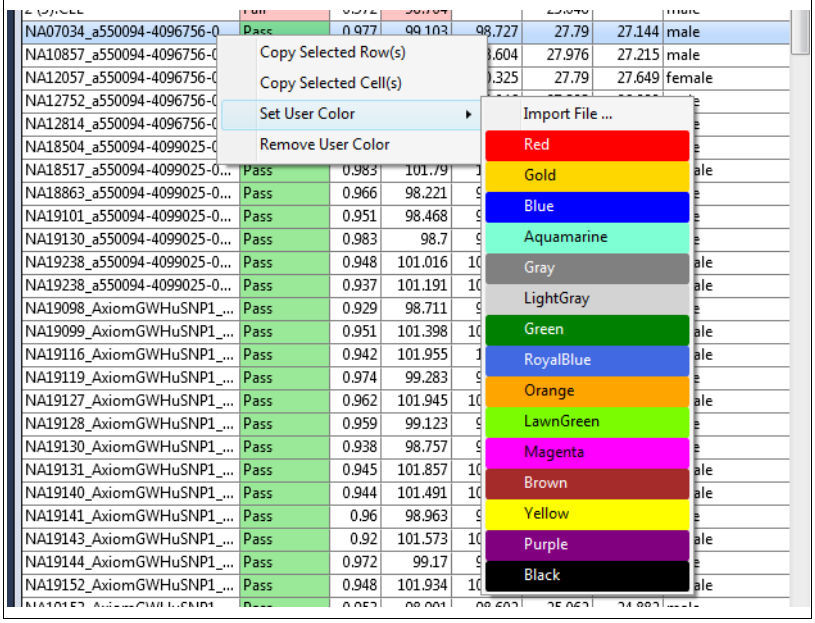
Use this feature to more easily identify different sets between the Sample Table and Cluster Plot.

Assigning a color to a sample

1. Right-click on the sample you want to assign a color to.

   A menu appears. (Figure 64)

Figure 64 Right-click menu - Set User Color



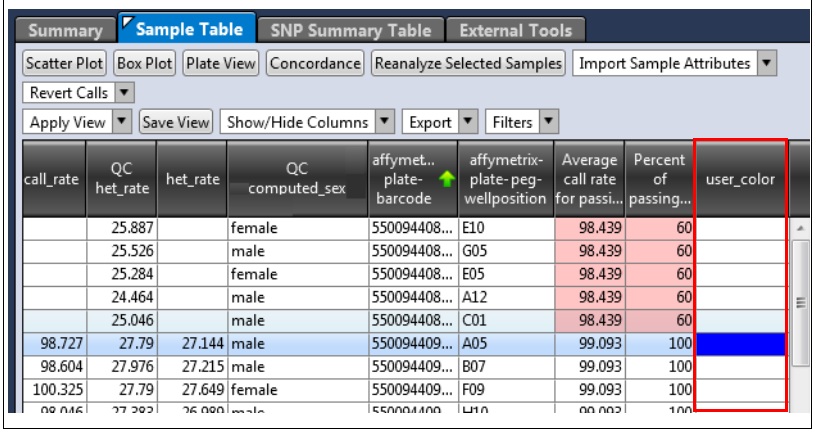
2. Mouse over **Set User Color**.

   A color palette appears, as shown in Figure 64.

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 65.

Figure 65 Right-click menu - Set User Color

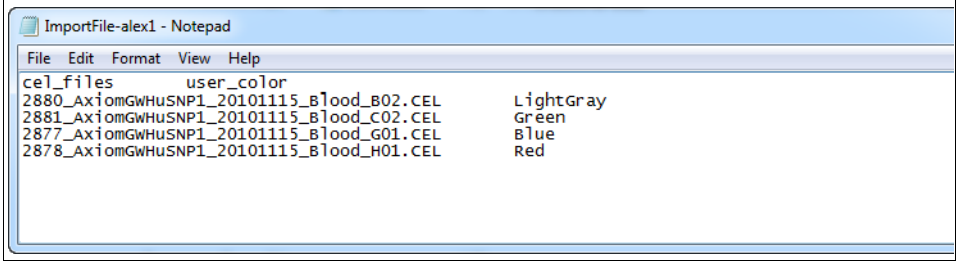


Importing assigned colors

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 66)

Figure 66 Two column example in MS Notepad

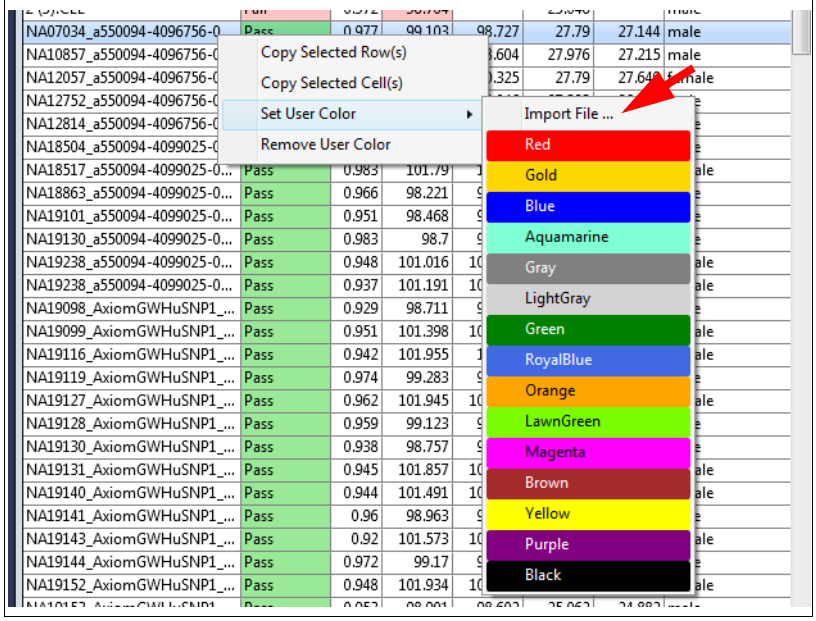


IMPORTANT! Your user_color entries must match the color palette naming conventions shown in Figure 67. Example: RoyalBlue not Royal Blue.

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 67)

Figure 67 Right-click menu - Set User Color - Import File



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

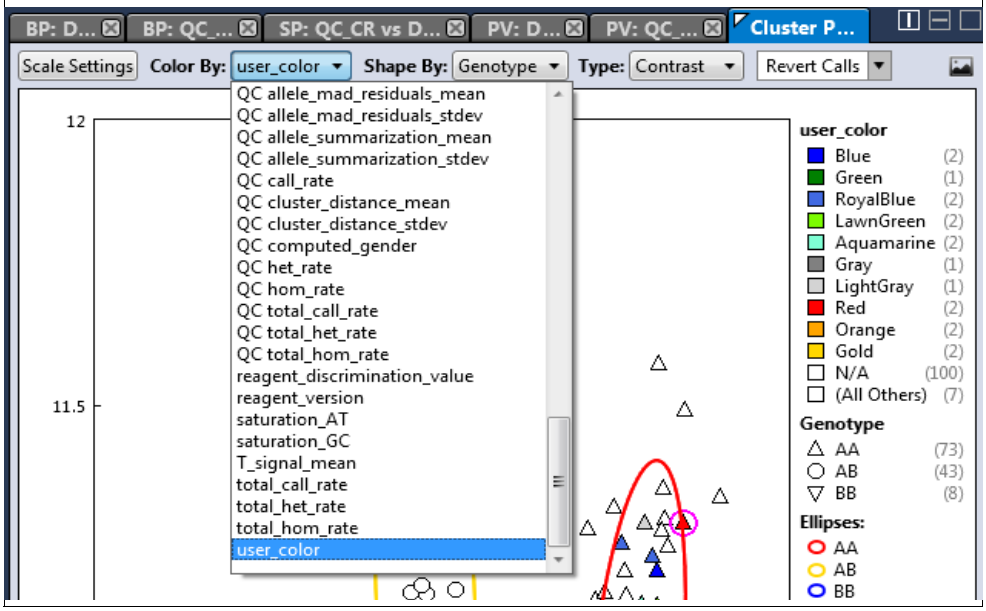
   An **Import User Colors** Explorer window appears.
6. Locate your saved TXT file, click to highlight it, then click **Open**.

   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 plot

1. From the Cluster Plot, click the **Color By** drop-down menu. (Figure 68)

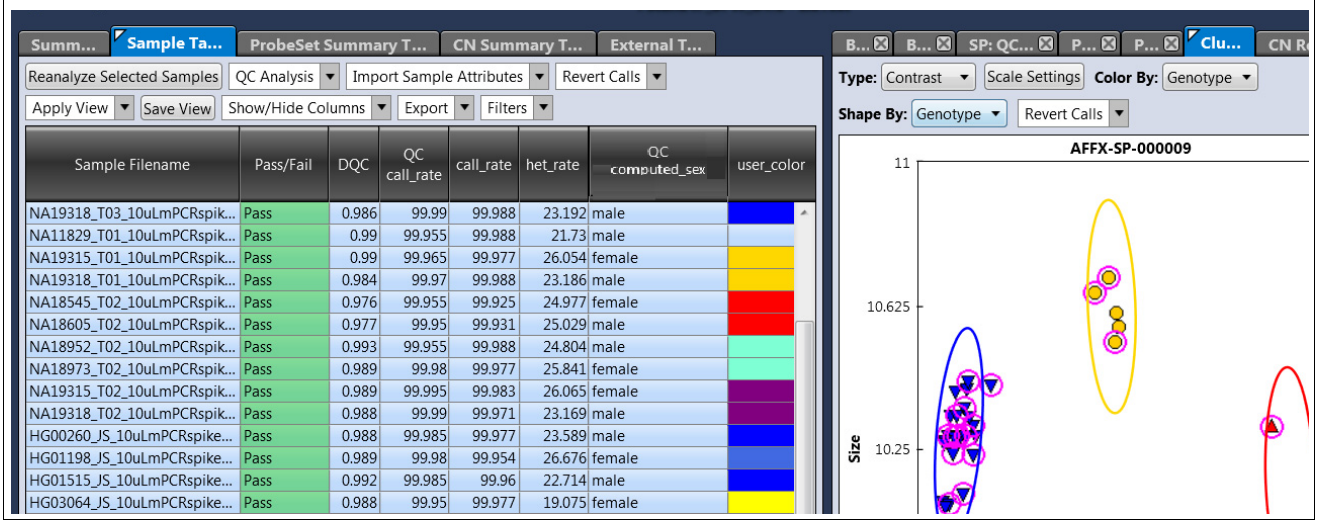
Figure 68 Color By menu - user_color



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 Plot, as show in Figure 69.

Figure 69 user_color shown in Sample Table and Cluster Plot



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

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

1. Click inside the 

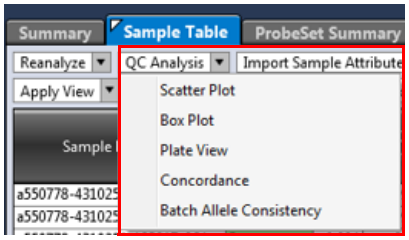
text field (bottom left corner of table).
2. Enter a keyword or number.
3. Click the **Up** or **Down** button.

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

QC Analysis

From the Sample Table, click on the **QC Analysis** drop-down to select the type of plot or view you want, as shown in Figure 70.

Figure 70 QC Analysis drop-down menu

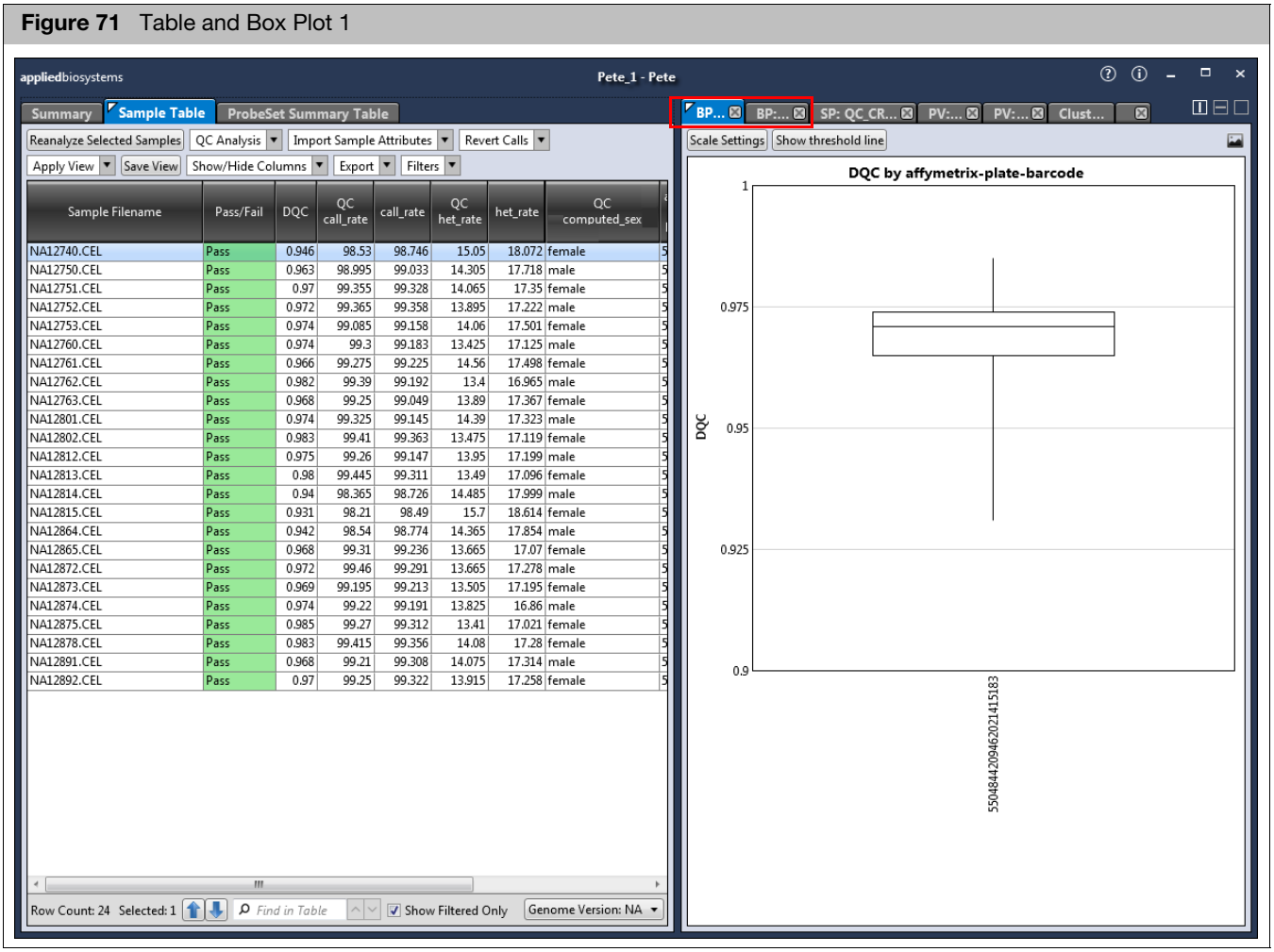


Box Plots

Note: By default, the Viewer generates two Box Plot window tabs. (Figure 71)

Viewing the default box plots

Figure 71 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 58.

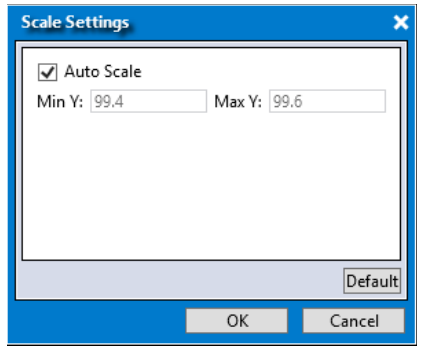
To change a Box Plot's axis properties, you must create a new Box Plot. See "Adding a new box plot" on page 58.

Changing the box plot's scale setting ranges

1. Click Scale Settings.

   The Scale Settings window appears. (Figure 72)

Figure 72 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 to their factory settings.

Creating a threshold line

1. Click the Box Plot's **Show threshold line** button.

   A thick dotted line appears at the bottom of the plot.

2. There are two ways to set a threshold line:
   a. Click and hold onto the thick dotted line, then drag your mouse cursor to move it up or down.
   b. Enter a value in the text field box (right of thick dotted line).

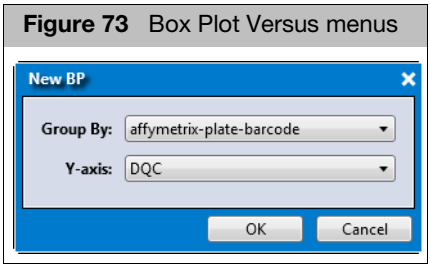
   To remove the threshold, click the Box Plot's **Hide threshold line** button.

Adding a new box plot

1. From the Sample Table's QC Analysis drop-down menu, click **Box Plot**.

   The New BP window appears. (Figure 73)

Figure 73 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 222.

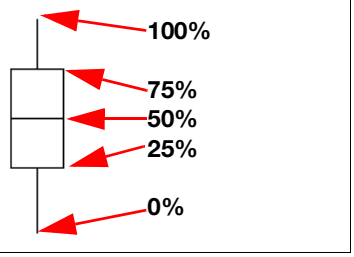
4. Click **OK**.

   A new Box Plot window tab is created.

Reading box plot percentiles

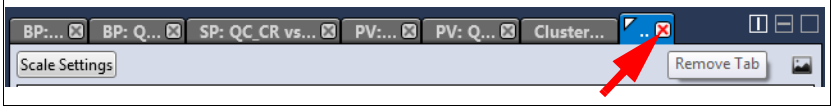
See Figure 74.

Figure 74 Box Plot percentiles



At any time, click **X** to remove a window/tab, as shown in Figure 75.

Figure 75 New Window/Tab



Saving the current box plot view

1. Click the  button (upper right).

   An Explorer window appears.

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

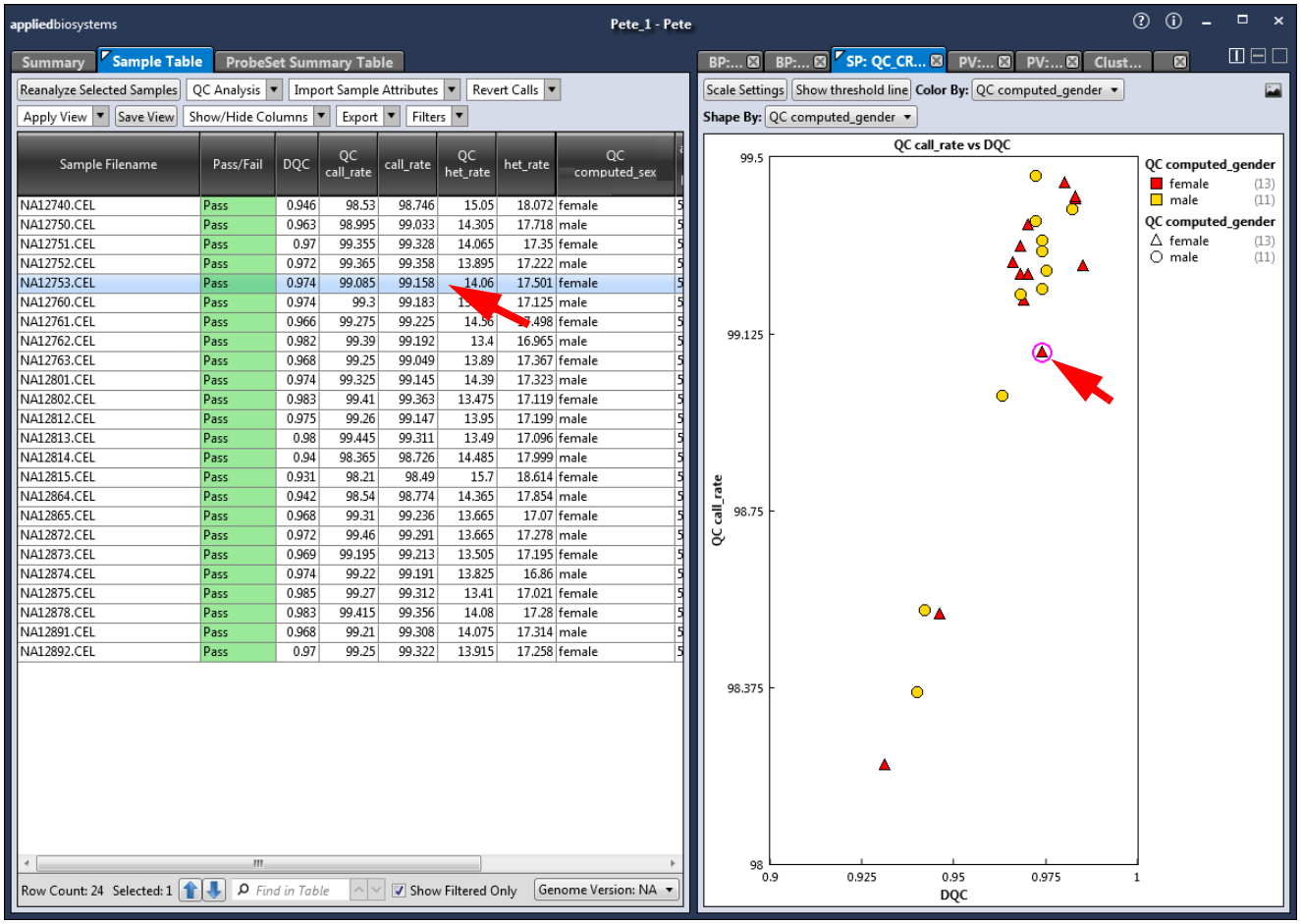
Scatter Plot

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_sex**, as shown in Figure 76.

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 76)

Figure 76 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.

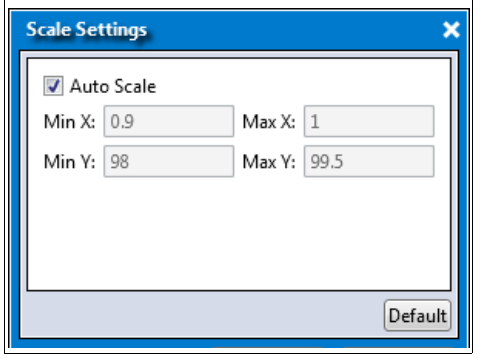
To change a Scatter Plot's axis properties, you must create a new Scatter Plot. See "Adding a new scatter plot and selecting its X and Y properties" on page 61.

Changing the scatter plot's setting ranges

1. Click Scale Settings.

   The Scale Settings window appears. (Figure 77)

Figure 77 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 to their factory settings.

Creating a threshold line

1. Click the Scatter Plot's **Show threshold line** button.

   A thick dotted line appears at the bottom of the plot.

2. There are two ways to set a threshold line:
   a. Click and hold onto the thick dotted line, then drag your mouse cursor to move it up or down.
   b. Enter a value in the text field box (right of thick dotted line).

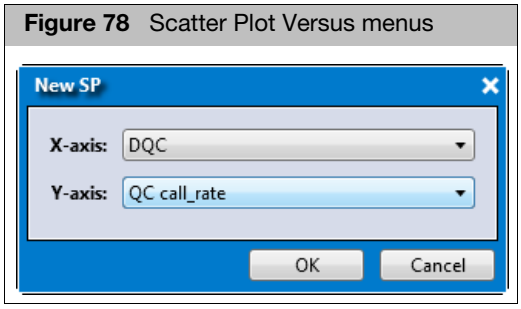
   To remove the threshold, click the Scatter Plot's **Hide threshold line** button.

Adding a new scatter plot and selecting its X and Y properties

1. From the Sample Table's QC Analysis drop-down menu, click **Scatter Plot**.

   The New SP window appears. (Figure 78)

Figure 78 Scatter Plot Versus menus



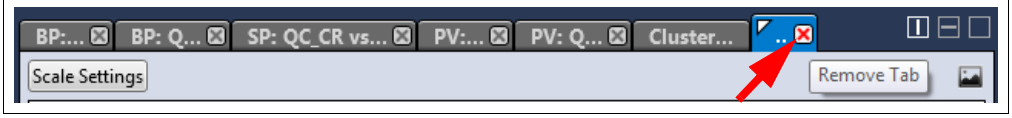
2. Use the drop-down menus to select your Plot's versus scenario (X and Y axis). For definitions, see "Sample table" on page 222.

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 79.

Figure 79 New Window/Tab



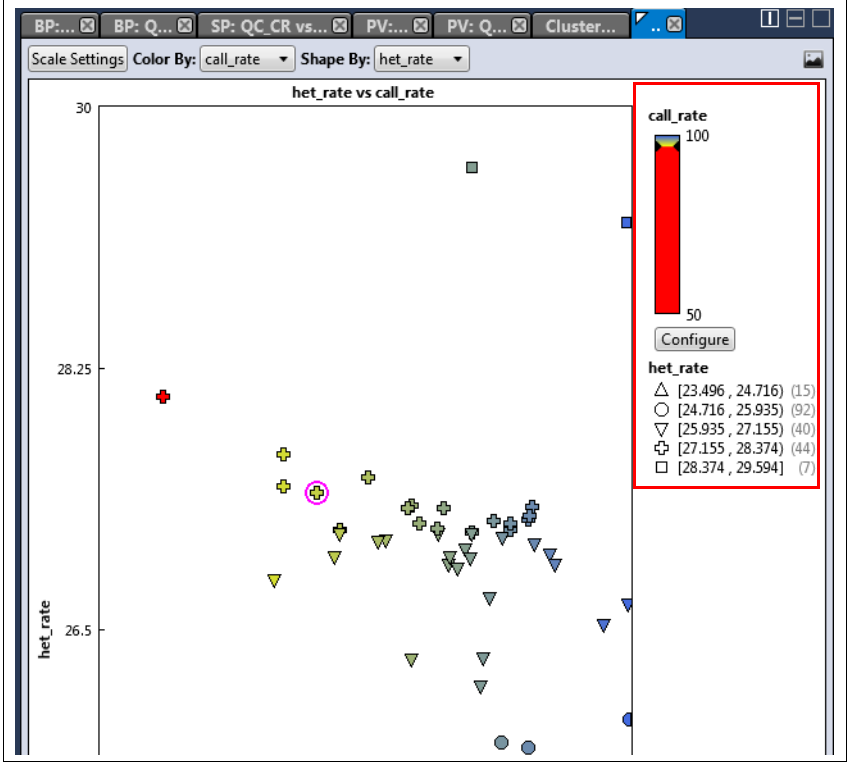
4. Click the **Color By** and **Shape By** drop-down menus to select the combination view you want. See "Sample table" on page 222 for Color By and Shape By definitions.

   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 into the sample table" on page 43.

   A legend appears within the plot. (Figure 80)

   Note: If you hover your mouse cursor over a point on the plot, the name of the sample is revealed, however. if you have imported Sample Attributes and included the Alternate Sample Name, your alternate name appears in place of the original sample filename.

Figure 80 Color By and Shape By Legend example



The graph can display up to 10 different colors and up to 11 different shapes. If the attributes selected for display have more than 11 categories, categories 1 through 10 are displayed normally, but categories 11 and higher get grouped together.

If your study has more than 10 values:

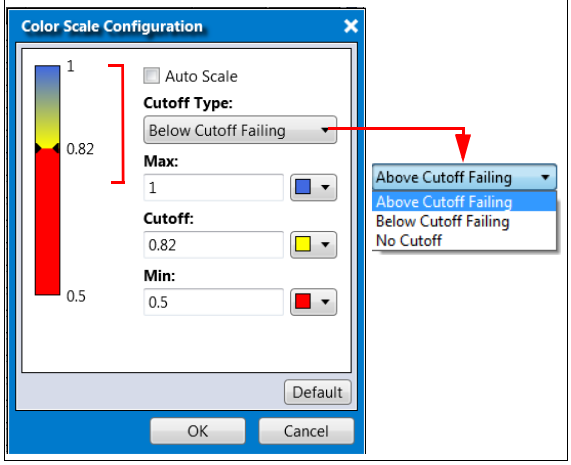
- If the value is text, the software takes the first nine values and assigns each a color or shape. The remaining values are put into a bin labeled **All Others**. All values in the Other bin have the same color or shape.
- If the value is a date or number, the software divides the range of data into 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 your scatter plot view settings

1. Click Configure.

   The Color Scale Configuration window appears. (Figure 81)

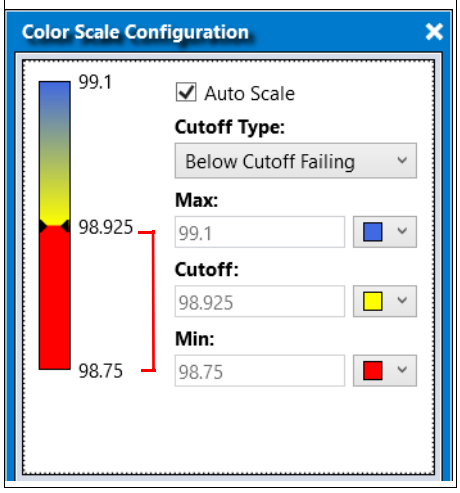
Figure 81 Color By options



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. (**Blue** in Figure 81)

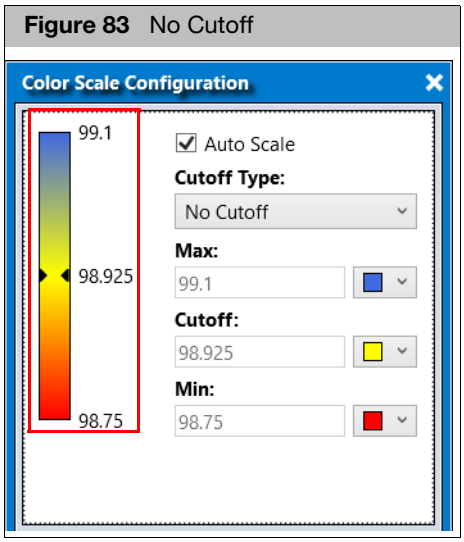
- **Below Cutoff Failing -** This presents a hard visual cutoff graph of all values that fail BELOW the Cutoff value entered. The Below Cutoff data is represented by the color defined for Min. (**Red** in Figure 82)

Figure 82 Below Cutoff



- **No Cutoff -** This presents a smooth 3-point gradient of your defined Max, Min, and colors. (Figure 83)

Figure 83 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  button (upper right).

   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 84) Each have a factory set metric. Plate View 1 is **DQC by Plate** and Plate View 2 is **QC call_rate by Plate**.

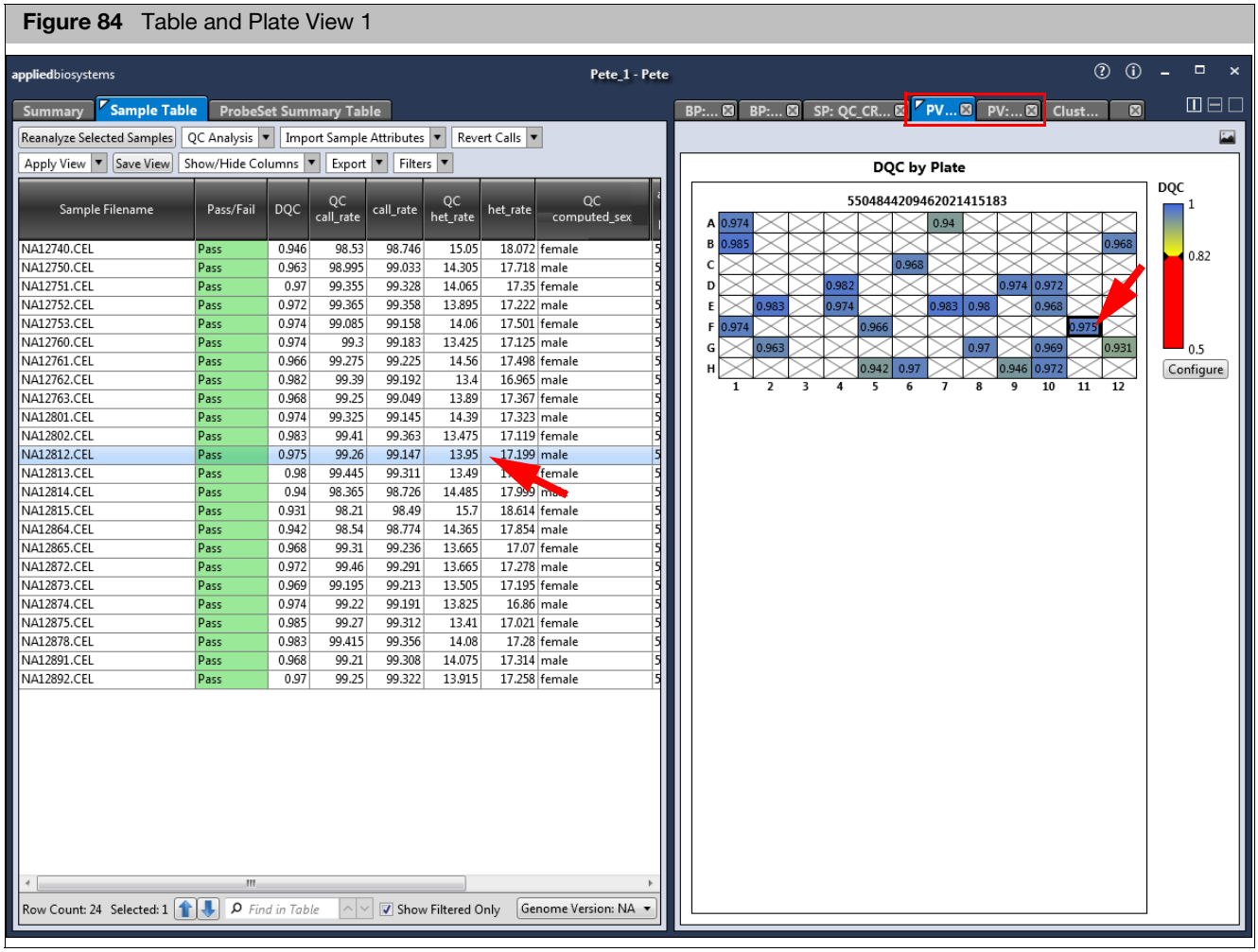
To display a different metric you must create a new Plate View. For more details, see "Adding a new plate view metric" on page 66.

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 84.

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 84.

Figure 84 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 67.

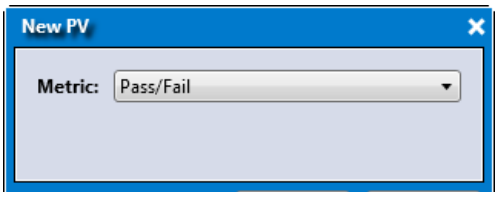
Adding a new plate view metric

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. From the Sample Table's QC Analysis drop-down menu, click **Plate View**.

   The New PV window appears. (Figure 85)

Figure 85 Plate View Metric setting



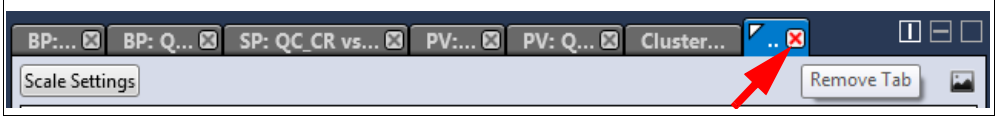
IMPORTANT! If your sample attributes file contains Plate Name a check box appears enabling you to title your plate view by Plate Name and not by the plate's barcode.

2. Use the drop-down menus to select your Plate View's Metric setting. See the tables in "Sample table" on page 222 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 86.

Figure 86 New Window/Tab

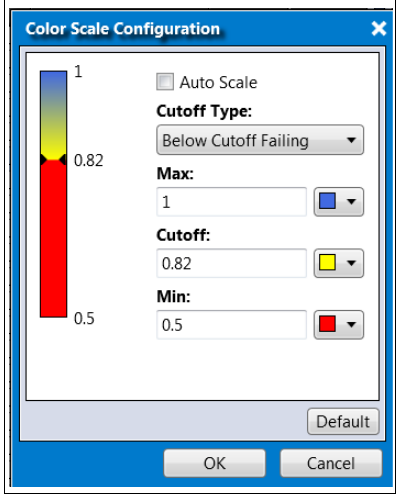


Customizing your plate view settings

1. Click Configure .

   The Color Scale Configuration window appears. (Figure 87)

Figure 87 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  button (upper right).

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

Concordance checks

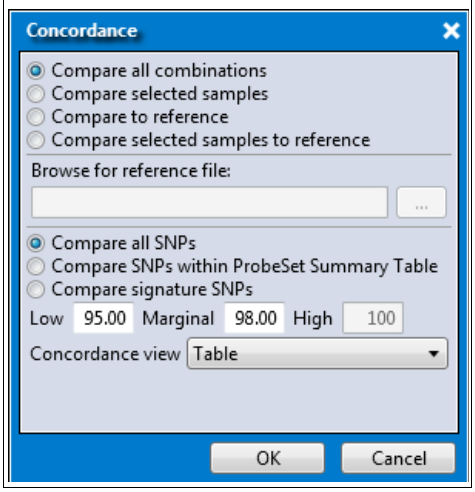
- **Compare all combinations** enables you to compare the SNP calls for all samples. The concordance between all pairwise comparisons for the samples in the dataset/ suitcase are reported.
- Compare selected samples
- **Compare to reference** enables you to compare every sample to a multi-sample reference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

Running a concordance check

1. From the Sample Table's QC Analysis drop-down menu, click **Concordance**.

   The Concordance window appears. (Figure 88)

Figure 88 Concordance window



Comparing all combinations

1. Make sure the **Compare all combinations** radio button is selected.

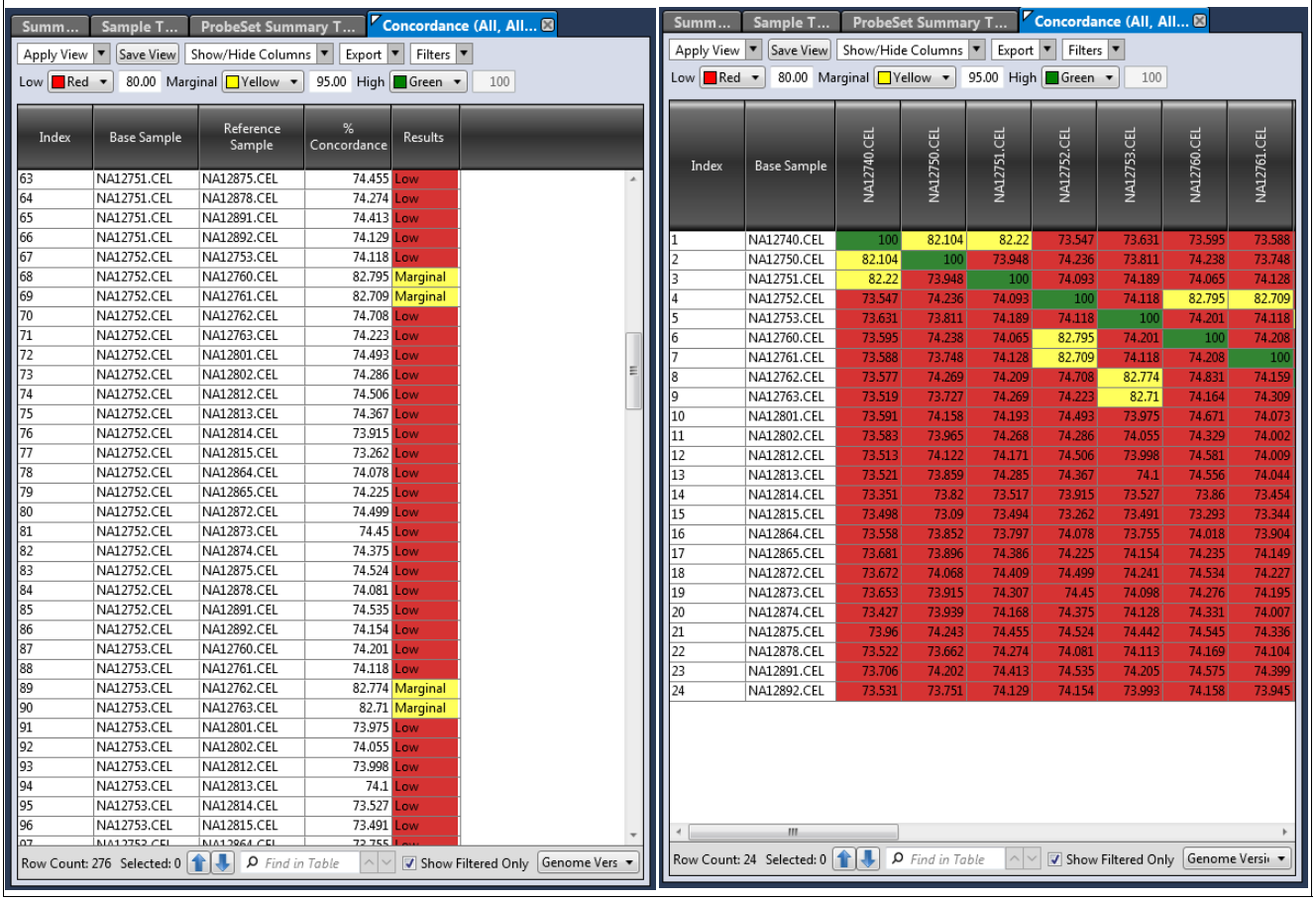
   Note: By default, the **Compare all SNPs** radio button is selected.

2. (Optional) Click the **Compare SNPs within the Probeset Summary Table**, or **Compare signature SNPs** radio button.
3. (Optional) Enter a different **Low** and **Marginal** threshold value.
4. The default Concordance view is a table. Click the Table drop-down to select **Matrix**.
5. Click **OK**.

   After a few moments, the Concordance window tab appears in your selected Table or Matrix view, as shown in Figure 89.

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 89 Example concordance table view (left) Matrix view (right)



For definitions of the Concordance columns, see Table 24 on page 226.

To change the default threshold colors, use the color drop-down menus.

Comparing to selected samples

1. Within the Sample Table, click to highlight the samples you want to run concordance on.
2. Click the **Compare selected samples** radio button.

   Note: By default, the **Compare all SNPs** radio button is selected.
3. (Optional) Enter a different **Low** and **Marginal** threshold value.

   Note: The lowest setting allowed for the Low threshold is 80.
4. The default Concordance view is a table. Click the Table drop-down to select **Matrix**.
5. Click **OK**.

   After a few moments, the Concordance window tab appears in your selected Table or Matrix view.

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.

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

5. Click **OK**.

   After a few moments, a Concordance window tab appears.

Comparing selected samples to reference

1. Within the Sample Table, click to highlight the samples you want to run concordance on.

2. Click the **Compare selected samples to reference** radio button.

3. Click the Browse for reference file Browse button.

   An Explorer window opens.

4. Navigate to your reference file location, then click **Open**.

   Your reference file is displayed.

IMPORTANT! The first column of your reference file must have the probeset ID, the following columns are the reference samples. In addition, your reference genotypes must be reported using letter call codes (e.g. AA, AB, BB, NoCall).

   Note: By default, the **Compare all SNPs** radio button is selected.

5. (Optional) Enter a different **Low** and **Marginal** threshold value.

6. Click **OK**.

   After a few moments, the Concordance window tab appears in the Table view.

   Note: The Matrix view is not available when comparing selected samples to reference(s).

Reverting calls

If you need to revert call(s), see "Changing or reverting calls" on page 80.

Batch allele consistency

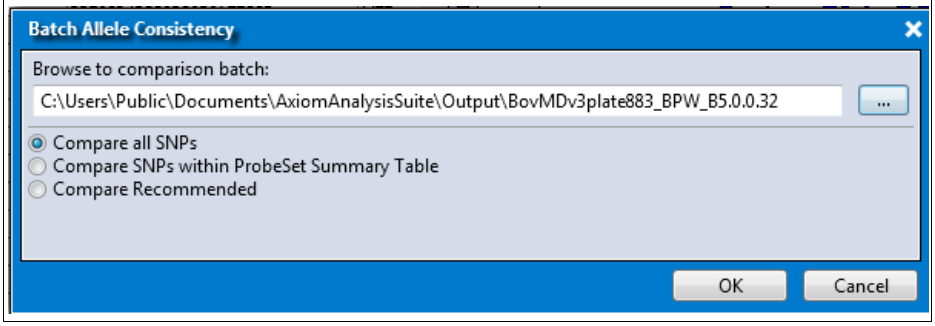
- Batch Allele Consistency (BAC) QC analysis compares the frequency of the B allele between two batches. This can be helpful in identifying batch effects and whether plates should be genotyped separately.
- To perform BAC, two Best Practices batches should be run, with one batch containing a subset of samples of the other.
- The output is a B allele frequency value per probeset per batch and pass or fail for each probeset.
- If the difference between the B allele frequency value is less than the threshold, the probeset passing. If the difference is greater than the threshold, the probeset fails.

  Note: See the Axiom Data Analysis Guide for details on when to run.

1. From the Sample Table tab, click the **QC Analysis** drop-down, select **Batch Allele Consistency**.

   The BAC window appears. (Figure 90)

Figure 90 BAC window

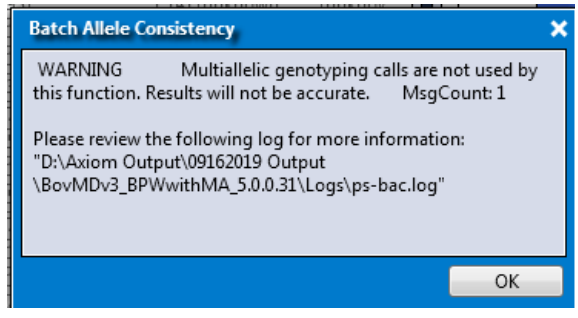


2. Browse to the folder that contains the subset batch, then click **OK**.
   - Click the **Compare all SNPs** radio button to compare all SNPs.
   - Click the **Compare SNPs with Probeset Summary Table** radio button to compare probeset that are listed in the probeset summary table.
   - Click the **Compare Recommended** radio button to select recommended probesets from the probeset summary table
3. Click **OK**.

   The analysis continues.

   If the batch contains any multi allelic genotyping, a warning message appears. (Figure 91)

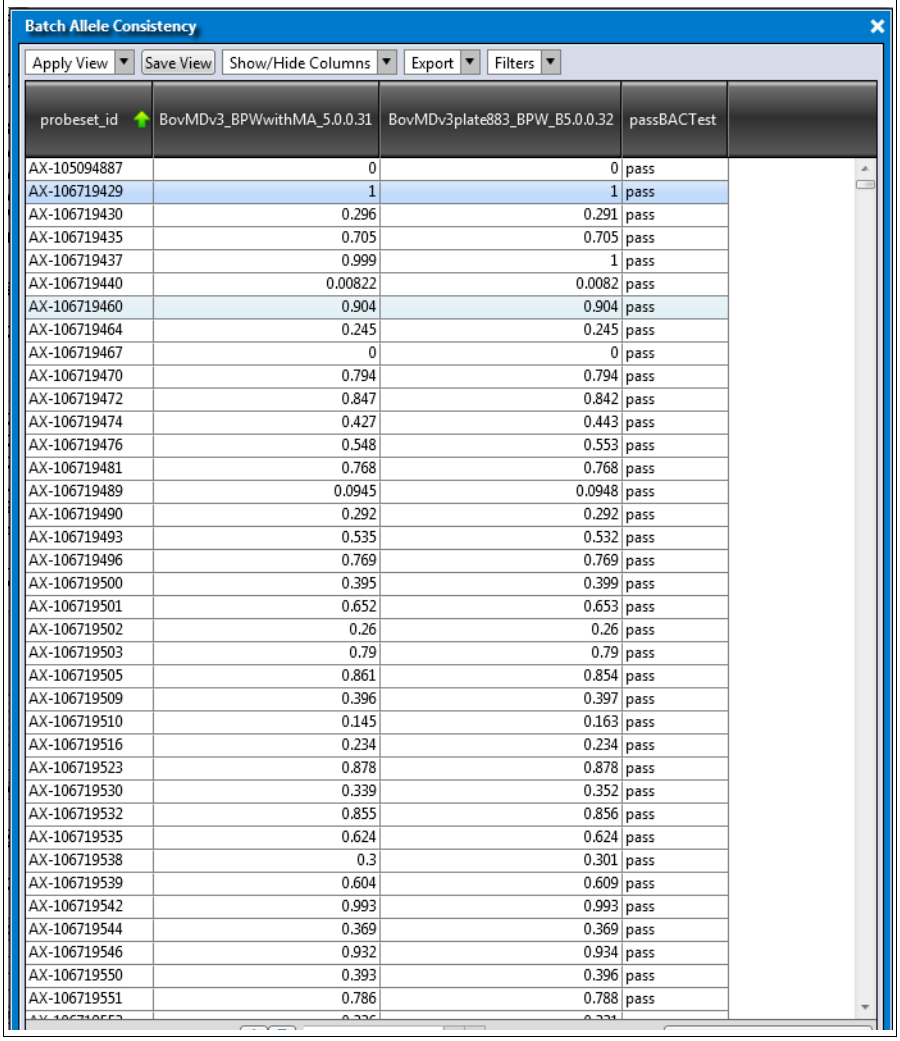
Figure 91 BAC Warning window



Acknowledge the warning, then click **OK**.

After the BAC analysis is complete, a BAC table appears (Figure 92) and becomes an additional tab within the Viewer.

Figure 92 BAC table



probeset_id	BovMDv3_BPWwithMA_5.0.0.31	BovMDv3plate883_BPW_B5.0.0.32	passBACTest
AX-105094887	0	0	pass
AX-106719429	1	1	pass
AX-106719430	0.296	0.291	pass
AX-106719435	0.705	0.705	pass
AX-106719437	0.999	1	pass
AX-106719440	0.00822	0.0082	pass
AX-106719460	0.904	0.904	pass
AX-106719464	0.245	0.245	pass
AX-106719467	0	0	pass
AX-106719470	0.794	0.794	pass
AX-106719472	0.847	0.842	pass
AX-106719474	0.427	0.443	pass
AX-106719476	0.548	0.553	pass
AX-106719481	0.768	0.768	pass
AX-106719489	0.0945	0.0948	pass
AX-106719490	0.292	0.292	pass
AX-106719493	0.535	0.532	pass
AX-106719496	0.769	0.769	pass
AX-106719500	0.395	0.399	pass
AX-106719501	0.652	0.653	pass
AX-106719502	0.26	0.26	pass
AX-106719503	0.79	0.79	pass
AX-106719505	0.861	0.854	pass
AX-106719509	0.396	0.397	pass
AX-106719510	0.145	0.163	pass
AX-106719516	0.234	0.234	pass
AX-106719523	0.878	0.878	pass
AX-106719530	0.339	0.352	pass
AX-106719532	0.855	0.856	pass
AX-106719535	0.624	0.624	pass
AX-106719538	0.3	0.301	pass
AX-106719539	0.604	0.609	pass
AX-106719542	0.993	0.993	pass
AX-106719544	0.369	0.369	pass
AX-106719546	0.932	0.934	pass
AX-106719550	0.393	0.396	pass
AX-106719551	0.786	0.788	pass

Column	Description
probeset_id	The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP probe sets only).
Case Control Batch Name	BAC value for probeset in batch that is used as the case control for the comparison. Calculated **(2*n_BB+n_AB)/2*(n_AA+n_AB+n_BB)**
Case Batch Name	BAC value for probeset in batch that is used as the case for the comparison. Calculated **(2*n_BB+n_AB)/2*(n_AA+n_AB+n_BB)**
passBACTest	Pass or Fail depending on: **abs(BAF_casecontrol - BAF_case) < intercept where intercept = 0.02**

A BAC Summary table is also added to the Summary window. This table details the number of probesets, percent passing and failing BAC QC Analysis, as shown in the example below.

Batch Allele Consistency Summary

Category	Number	Percent Passing	Percent Failing
Total Probesets	65008	95.828	4.172
BestAndRecommended	58151	98.932	1.068

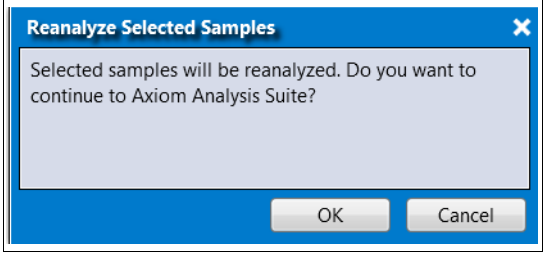
For more details, see "Summary window tab" on page 39.

Reanalyzing samples 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 93)

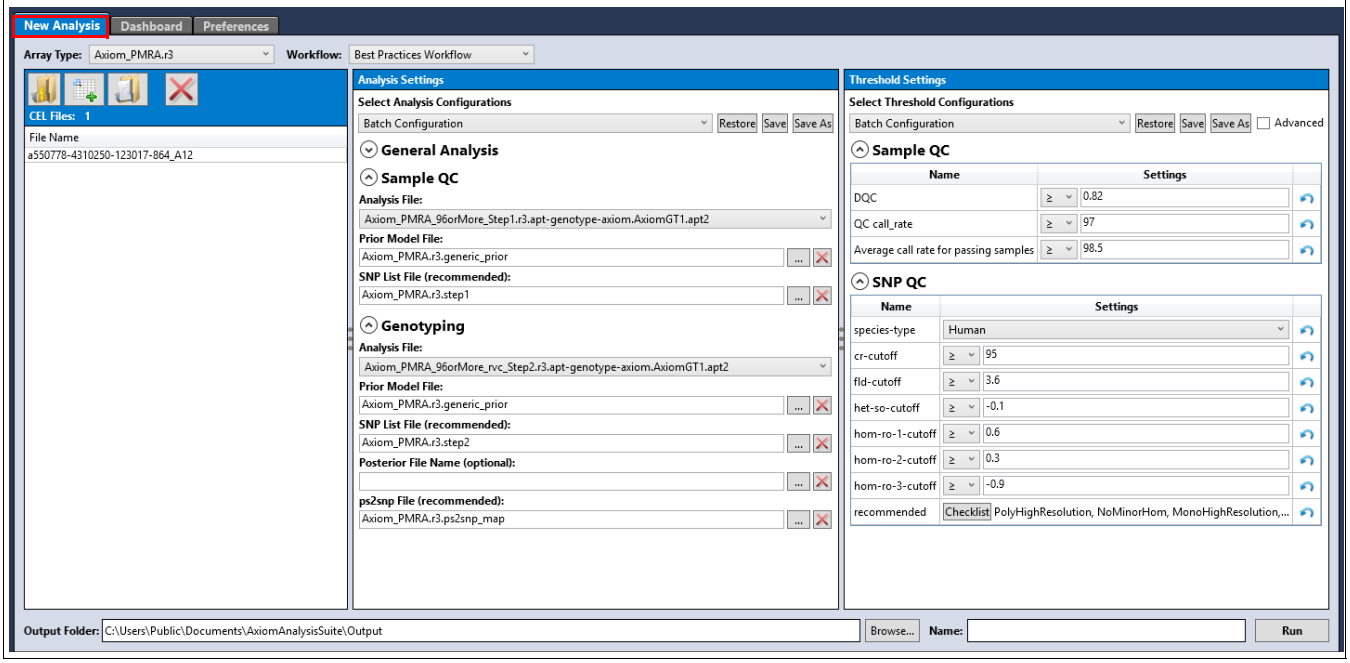
Figure 93 Return to AxAS message



3. Click **OK**.

   Your selected samples are displayed and ready for re-analysis in a New Analysis window tab. (Figure 94)

Figure 94 New Analysis window tab



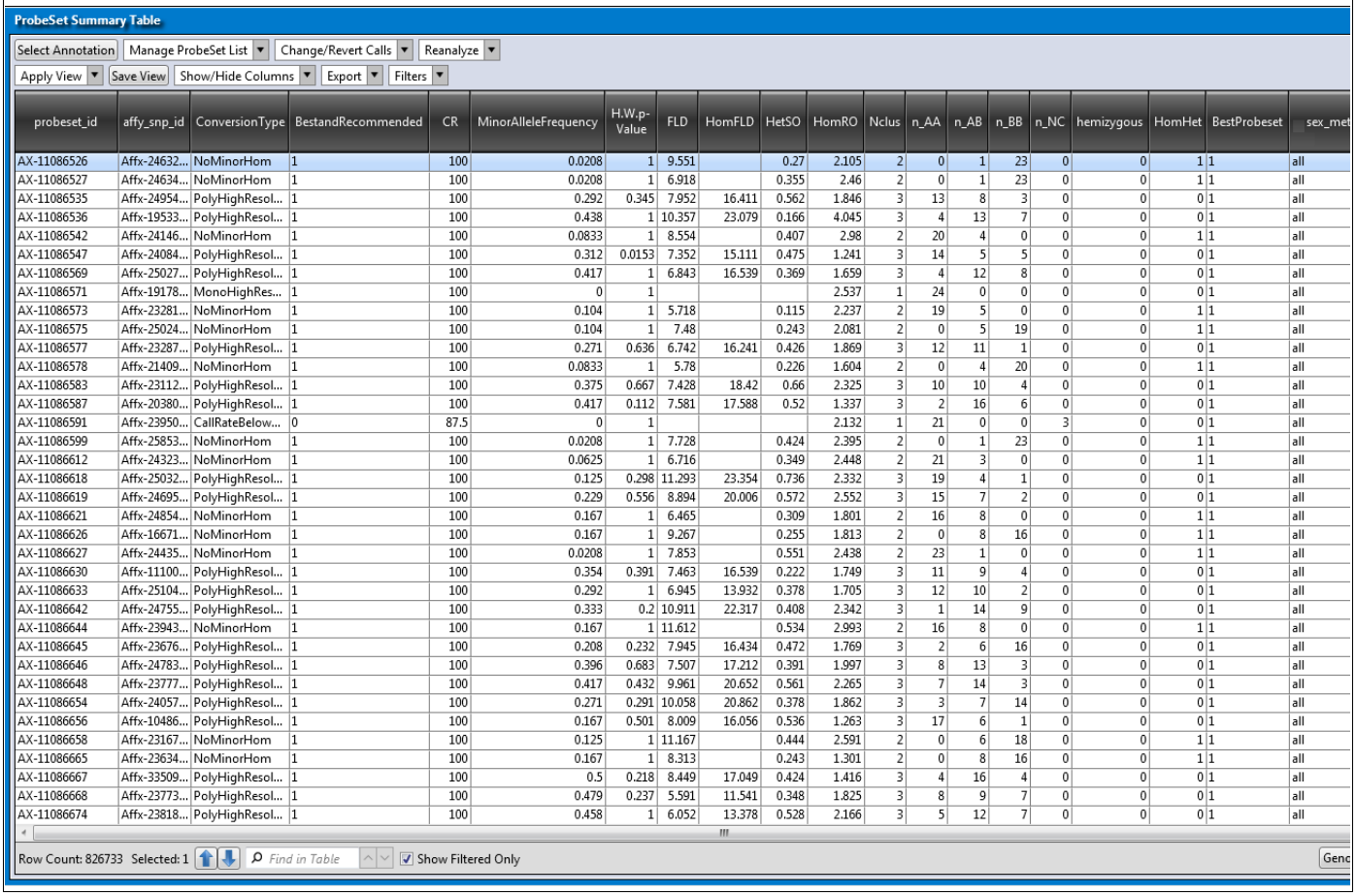
4. See Chapter 2, "Performing an analysis" on page 20 for instructions on setting up an analysis.

4 Probeset summary table and cluster plot

Probeset summary table

Figure 95 is an example of a standard Probeset Summary Table.

Figure 95 ProbeSet summary table window tab



probeset_id	affy_snp_id	ConversionType	BestandRecommended	CR	MinorAlleleFrequency	H.W.p-Value	FLD	HomFLD	HetSO	HomRO	Nclus	n_AA	n_AB	n_BB	n_NC	hemizygous	HomHet	BestProbeset	sex_met
AX-11086526	Affx-24632...	NoMinorHom	1	100	0.0208	1	9.551		0.27	2.105	2	0	1	23	0	0	1	1	all
AX-11086527	Affx-24634...	NoMinorHom	1	100	0.0208	1	6.918		0.355	2.46	2	0	1	23	0	0	1	1	all
AX-11086535	Affx-24954...	PolyHighResol...	1	100	0.292	0.345	7.952	16.411	0.562	1.846	3	13	8	3	0	0	0	1	all
AX-11086536	Affx-19533...	PolyHighResol...	1	100	0.438	1	10.357	23.079	0.166	4.045	3	4	13	7	0	0	0	1	all
AX-11086542	Affx-24146...	NoMinorHom	1	100	0.0833	1	8.554		0.407	2.98	2	20	4	0	0	0	1	1	all
AX-11086547	Affx-24084...	PolyHighResol...	1	100	0.312	0.0153	7.352	15.111	0.475	1.241	3	14	5	5	0	0	0	1	all
AX-11086569	Affx-25027...	PolyHighResol...	1	100	0.417	1	6.843	16.539	0.369	1.659	3	4	12	8	0	0	0	1	all
AX-11086571	Affx-19178...	MonoHighRes...	1	100	0	1				2.537	1	24	0	0	0	0	0	1	all
AX-11086573	Affx-23281...	NoMinorHom	1	100	0.104	1	5.718		0.115	2.237	2	19	5	0	0	0	1	1	all
AX-11086575	Affx-25024...	NoMinorHom	1	100	0.104	1	7.48		0.243	2.081	2	0	5	19	0	0	1	1	all
AX-11086577	Affx-23287...	PolyHighResol...	1	100	0.271	0.636	6.742	16.241	0.426	1.869	3	12	11	1	0	0	0	1	all
AX-11086578	Affx-21409...	NoMinorHom	1	100	0.0833	1	5.78		0.226	1.604	2	0	4	20	0	0	1	1	all
AX-11086583	Affx-23112...	PolyHighResol...	1	100	0.375	0.667	7.428	18.42	0.66	2.325	3	10	10	4	0	0	0	1	all
AX-11086587	Affx-20380...	PolyHighResol...	1	100	0.417	0.112	7.581	17.588	0.52	1.337	3	2	16	6	0	0	0	1	all
AX-11086591	Affx-23950...	CallRateBelow...	0	87.5	0	1				2.132	1	21	0	0	3	0	0	1	all
AX-11086599	Affx-25853...	NoMinorHom	1	100	0.0208	1	7.728		0.424	2.395	2	0	1	23	0	0	1	1	all
AX-11086612	Affx-24323...	NoMinorHom	1	100	0.0625	1	6.716		0.349	2.448	2	21	3	0	0	0	1	1	all
AX-11086618	Affx-25032...	PolyHighResol...	1	100	0.125	0.298	11.293	23.354	0.736	2.332	3	19	4	1	0	0	0	1	all
AX-11086619	Affx-24695...	PolyHighResol...	1	100	0.229	0.556	8.894	20.006	0.572	2.552	3	15	7	2	0	0	0	1	all
AX-11086621	Affx-24854...	NoMinorHom	1	100	0.167	1	6.465		0.309	1.801	2	16	8	0	0	0	1	1	all
AX-11086626	Affx-16671...	NoMinorHom	1	100	0.167	1	9.267		0.255	1.813	2	0	8	16	0	0	1	1	all
AX-11086627	Affx-24435...	NoMinorHom	1	100	0.0208	1	7.853		0.551	2.438	2	23	1	0	0	0	1	1	all
AX-11086630	Affx-11100...	PolyHighResol...	1	100	0.354	0.391	7.463	16.539	0.222	1.749	3	11	9	4	0	0	0	1	all
AX-11086633	Affx-25104...	PolyHighResol...	1	100	0.292	1	6.945	13.932	0.378	1.705	3	12	10	2	0	0	0	1	all
AX-11086642	Affx-24755...	PolyHighResol...	1	100	0.333	0.2	10.911	22.317	0.408	2.342	3	1	14	9	0	0	0	1	all
AX-11086644	Affx-23943...	NoMinorHom	1	100	0.167	1	11.612		0.534	2.993	2	16	8	0	0	0	1	1	all
AX-11086645	Affx-23676...	PolyHighResol...	1	100	0.208	0.232	7.945	16.434	0.472	1.769	3	2	6	16	0	0	0	1	all
AX-11086646	Affx-24783...	PolyHighResol...	1	100	0.396	0.683	7.507	17.212	0.391	1.997	3	8	13	3	0	0	0	1	all
AX-11086648	Affx-23777...	PolyHighResol...	1	100	0.417	0.432	9.961	20.652	0.561	2.265	3	7	14	3	0	0	0	1	all
AX-11086654	Affx-24057...	PolyHighResol...	1	100	0.271	0.291	10.058	20.862	0.378	1.862	3	3	7	14	0	0	0	1	all
AX-11086656	Affx-10486...	PolyHighResol...	1	100	0.167	0.501	8.009	16.056	0.536	1.263	3	17	6	1	0	0	0	1	all
AX-11086658	Affx-23167...	NoMinorHom	1	100	0.125	1	11.167		0.444	2.591	2	0	6	18	0	0	1	1	all
AX-11086665	Affx-23634...	NoMinorHom	1	100	0.167	1	8.313		0.243	1.301	2	0	8	16	0	0	1	1	all
AX-11086667	Affx-33509...	PolyHighResol...	1	100	0.5	0.218	8.449	17.049	0.424	1.416	3	4	16	4	0	0	0	1	all
AX-11086668	Affx-23773...	PolyHighResol...	1	100	0.479	0.237	5.591	11.541	0.348	1.825	3	8	9	7	0	0	0	1	all
AX-11086674	Affx-23818...	PolyHighResol...	1	100	0.458	1	6.052	13.378	0.528	2.166	3	5	12	7	0	0	0	1	all

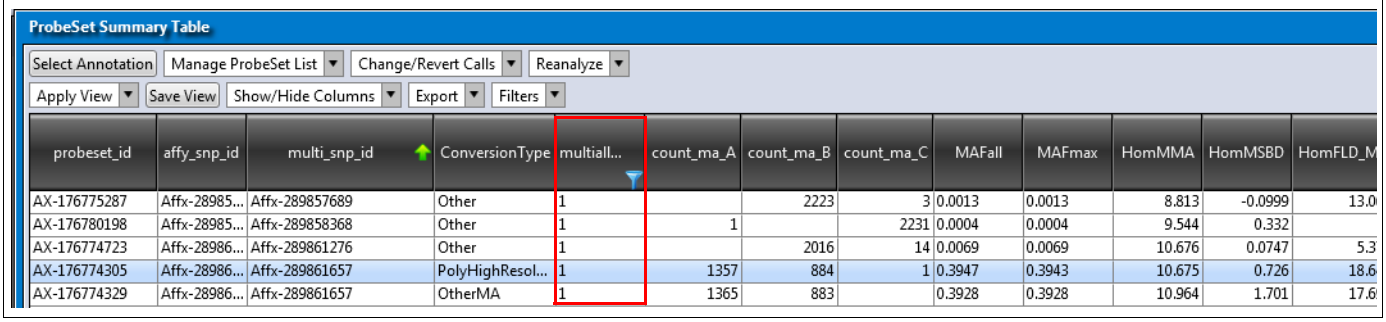
Multiallele probsets

Multiallelic probesets are probesets that can report calls from probe sets with more than one alternate allele. Multiallele probesets can be identified by selecting the **multiallelic** column. (Figure 96) A value of **1** indicates a multiallelic probeset. Additional columns can be shown that report metrics for multiallele probesets.

Multiallele probesets are classified differently than the more common biallelic probesets.

- Multiallele probesets cannot be classified as an OTV probeset.
- Multiallele probesets report PolyHighRes, NoMinorHom, MonoHighRes, Call Rate Below Threshold (CRBT), UnexpectedGenotypeFreq, and OtherMA. **Note**: OtherMA is different from Other. OtherMA means that the probeset failed a multiallelic-specific check.
- Multiallele probesets do not display values for many of the default Probeset Summary Table columns. Click **Show/Hide Columns** to select additional columns with multiallelic probeset metrics.

Figure 96 ProbeSet Summary Table - multiallelic column



probeset_id	affy_snp_id	multi_snp_id	ConversionType	multiall...	count_ma_A	count_ma_B	count_ma_C	MAFall	MAFmax	HomMMA	HomMSBD	HomFLD_M
AX-176775287	Affx-28985...	Affx-289857689	Other	1		2223	3	0.0013	0.0013	8.813	-0.0999	13.0
AX-176780198	Affx-28985...	Affx-289858368	Other	1	1		2231	0.0004	0.0004	9.544	0.332	
AX-176774723	Affx-28986...	Affx-289861276	Other	1		2016	14	0.0069	0.0069	10.676	0.0747	5.3
AX-176774305	Affx-28986...	Affx-289861657	PolyHighResol...	1	1357	884	1	0.3947	0.3943	10.675	0.726	18.6
AX-176774329	Affx-28986...	Affx-289861657	OtherMA	1	1365	883		0.3928	0.3928	10.964	1.701	17.6

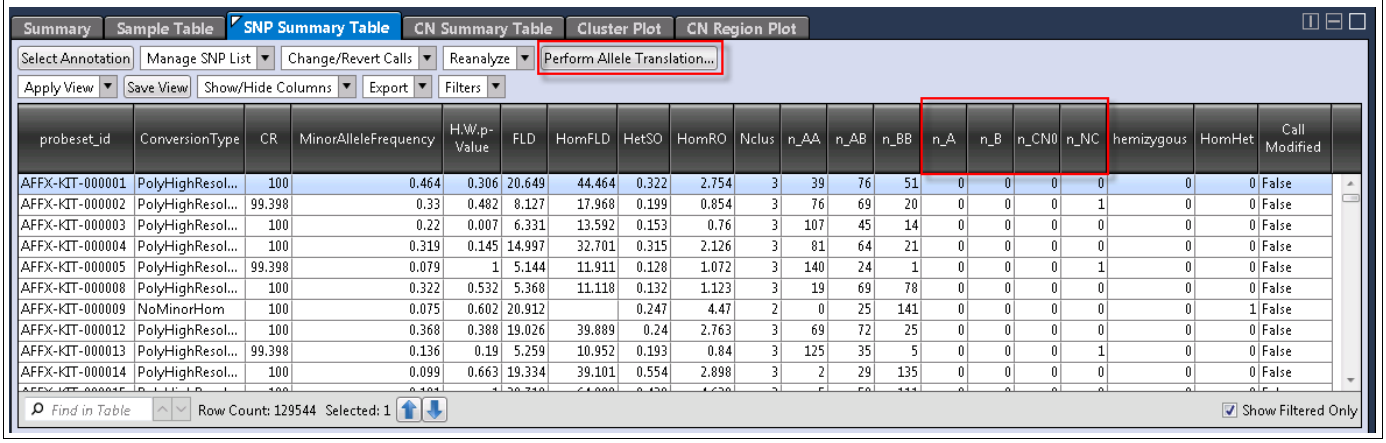
Note: If an analysis batch does not include samples needed to evaluate a probeset's performance, then no metrics (including ConversionType) will be reported. For example, if sex information for all samples is missing, AxAS outputs a message in the Dashboard explaining that sex-separated metrics can't be calculated and Y and X probesets will be treated as autosomal probesets and Y probesets will end up classified as Hemizygous.

Figure 97 represents a Probeset Summary Table with CN-aware genotyping and Allele Translation.

Note: The Perform Allele Translation... button only appears if supported arrays are available. See Chapter 6, "Allele translation" on page 158 for more information.

A Probeset 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 97. 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 183.

Figure 97 ProbeSet Summary Table with CN-aware genotyping and Allele Translation

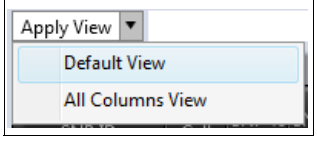


Using the probeset summary table

Setting the table view

1. Click the **Apply View** drop-down. (Figure 98)

Figure 98 Apply View drop-down



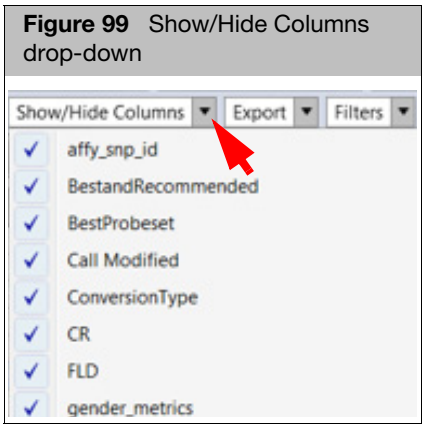
- The **Default View** is the initial table view and includes a preset number of columns.
- The **All Columns View** displays the maximum available columns.

Adding and Removing Table Columns

1. Click the **Show/Hide Columns** drop-down. (Figure 99)

   A list of available columns appear.

Figure 99 Show/Hide Columns drop-down



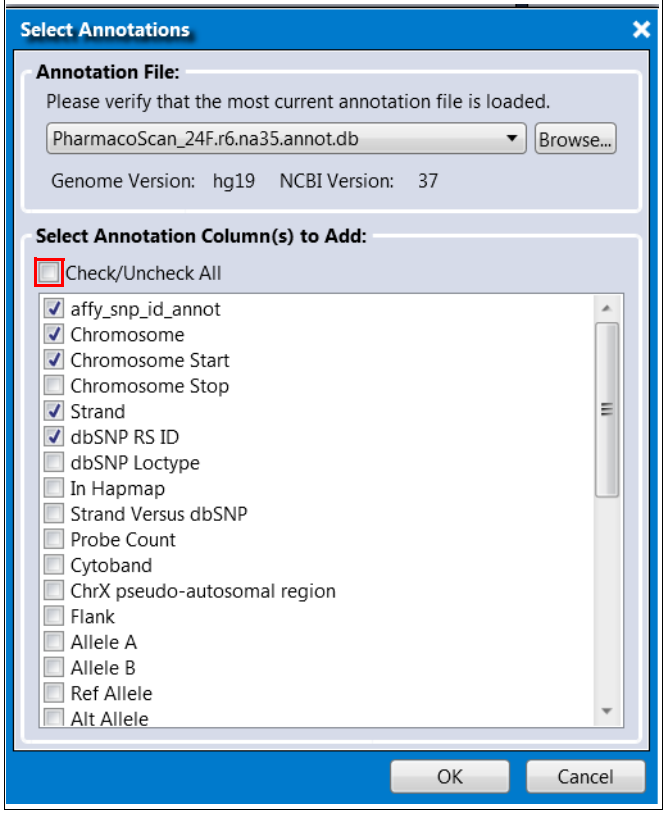
2. Click the check box next to the column(s) you want to add (Show) on the table. Click to uncheck a column you want to remove (Hide) from the table. Mouse over the menu's down arrow to reveal more available column choices. See Table 26 on page 229 for column definitions.

Selecting annotations

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

   The Select Annotations window appears. (Figure 100)

Figure 100 Available Annotations

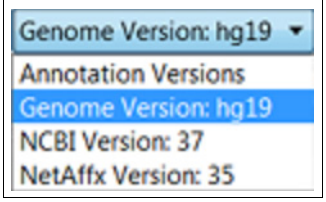


2. Select the appropriate Annotation File from the drop-down menu list.
3. Click the check box next to the Annotation Column(s) you want to add to the table or click the **Check/Uncheck All** check box (Figure 100) to add or remove ALL available annotations. See "Annotations and columns" on page 226 for each Annotation's definition.
4. Click **OK**.

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

   Note: After the annotations have been loaded into the batch, the Genome, NCBI and NetAffx versions are populated in the Summary Table's **Annotation Versions** (lower right) drop-down menu. (Figure 101)

Figure 101 Annotation versions

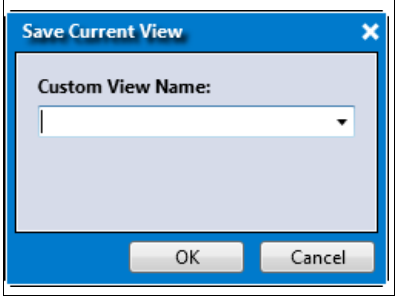


Saving your table column view

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

   The Save Current View window appears. (Figure 102)

Figure 102 Save Current View window



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

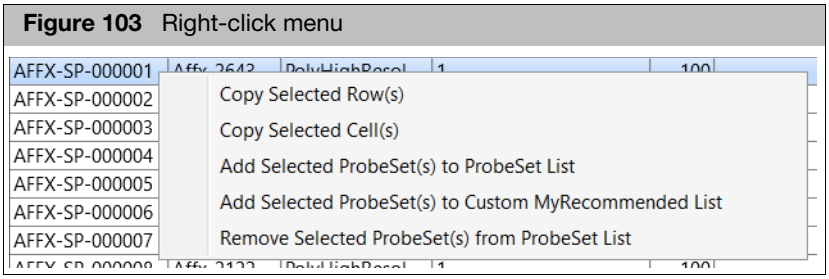
   Your custom table view is now saved and stored inside the Apply View menu for future use.

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 103)

Figure 103 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 103)

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

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

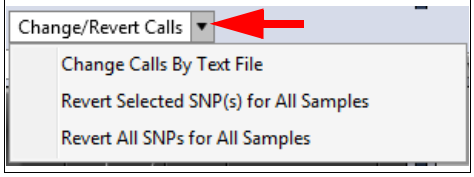
Changing or reverting calls

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 104)

Figure 104 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 105. It also cannot contain any annotation columns.

Figure 105 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	
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	28
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 probesets for all samples

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

Reverting all probesets for all samples

Use this feature to perform a master probesets reset.

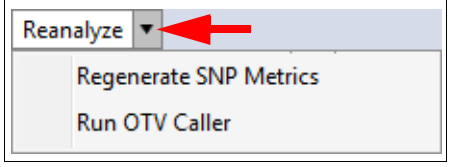
1. Click **Revert All SNP(s) for All Samples**.

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

Reanalyzing your probeset summary table data

The **Reanalyze** drop-down menu (Figure 106) offers optional steps for post-processing Probeset data. These functions utilize the output files from the previous genotyping and classification steps as input.

Figure 106 Reanalyze drop-down menu



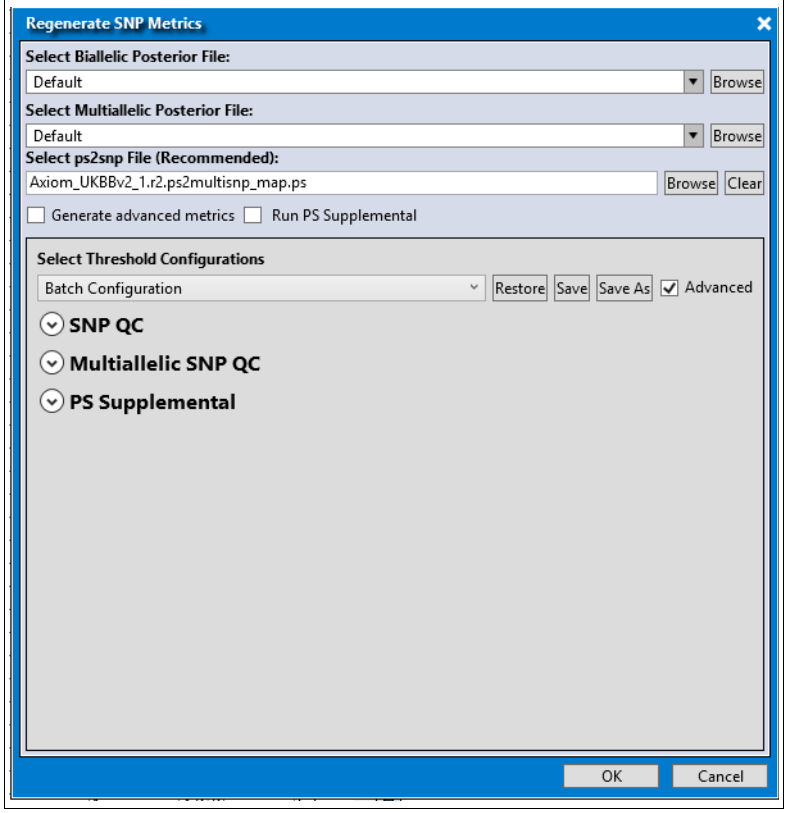
Regenerate probeset metrics

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

1. Click **Regenerate SNP Metrics**.

   The Regenerate SNP Metrics window appears. (Figure 107)

Figure 107 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 Probeset-specific posterior metrics.
5. (Optional) Click the **Run PS Supplemental** check box to generate additional metrics and probeset ConversionTypes.

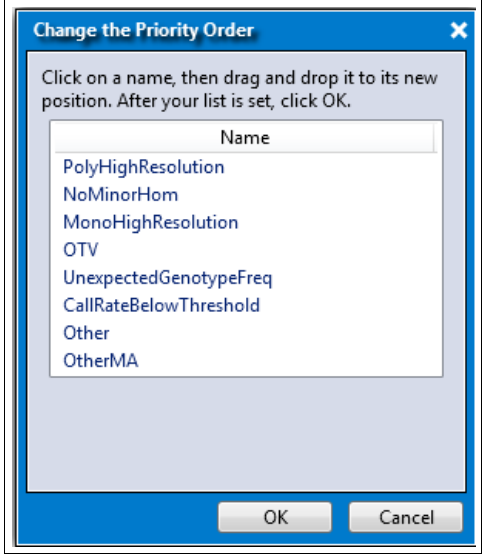
   Note: PS Supplemental performs further classification that may be needed for polyploid organisms, complex genomes, or inbred populations.

6. Select the Threshold Configuration you want to use, as described in"Customizing thresholds" on page 30.
7. Click to expand and use the **SNP QC** drop-down menu selections and text fields to setup the regeneration of your Probeset Metric.
8. To change the priority-order of the various assigned ConversionTypes, go to the Probeset QC setting's priority-order column and click the Change List Order button.

The Change the Priority Order window appears. (Figure 108)

Note: If your analysis contains multiallele probesets, additional conversion types are displayed.

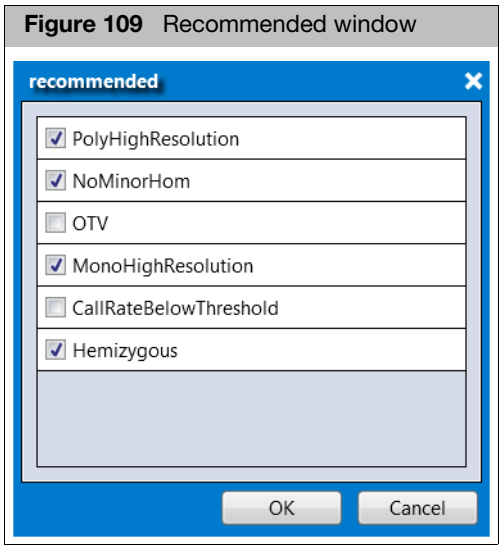
Figure 108 Change the Priority Order window



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 recommended window appears. (Figure 109)

Figure 109 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 ProbeSet Summary Table is updated.

Running OTV caller

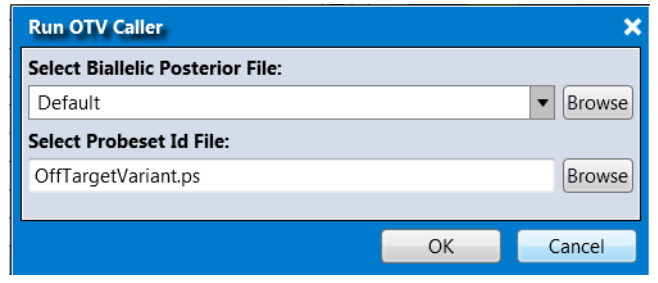
OTV Caller is designed for Probesets that have been classified as likely having off-target variants (OTV), or for Probesets with unusually large Y-dimension variance (as identified by PS Supplemental option of the Regenerate Probeset Metrics).

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 Run OTV Caller window appears. (Figure 110)

Figure 110 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 Probeset Id File's **Browse** button.

   An Explorer window appears.

4. Navigate to Probeset Id file you want to use, then click **Open**.
5. Click **OK**.

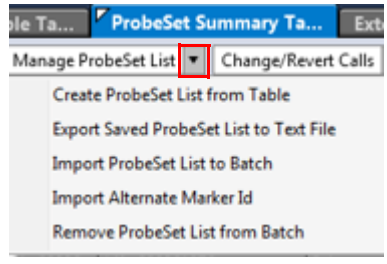
Managing your probeset list

AxAS enables lists of probesets to be saved within the application.

1. Click the **Manage ProbeSet List** drop-down menu (Figure 111), then select the option you want.

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

Figure 111 Manage ProbeSet List drop-down menu

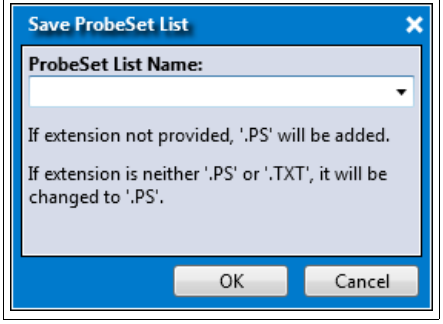


Saving your current probeset list

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

   The Save ProbeSet List window appears. (Figure 112)

Figure 112 Save ProbeSet List



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

Exporting your probeset list

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

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

   An Explorer window appears.

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

Importing a probeset list

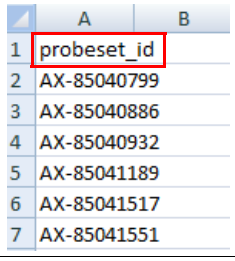
1. Click **Import ProbeSet List to Batch**.

   An Explorer window appears.

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

   Your first row/column header must be labeled **probeset_id**, as shown in Figure 113, otherwise an error message appears.

Figure 113 Required probeset_id column header



3. Click **Open**.

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

Import Alternate Marker ID

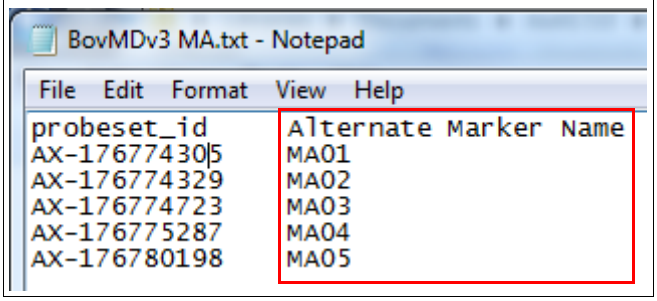
1. Click Import Alternate Marker ID.

   An Explorer window appears.

2. Navigate to your Alternate Marker ID text file location.

   The first column must be labeled **probeset_id** and the second column **Alternate Marker Name**, as shown in Figure 114.

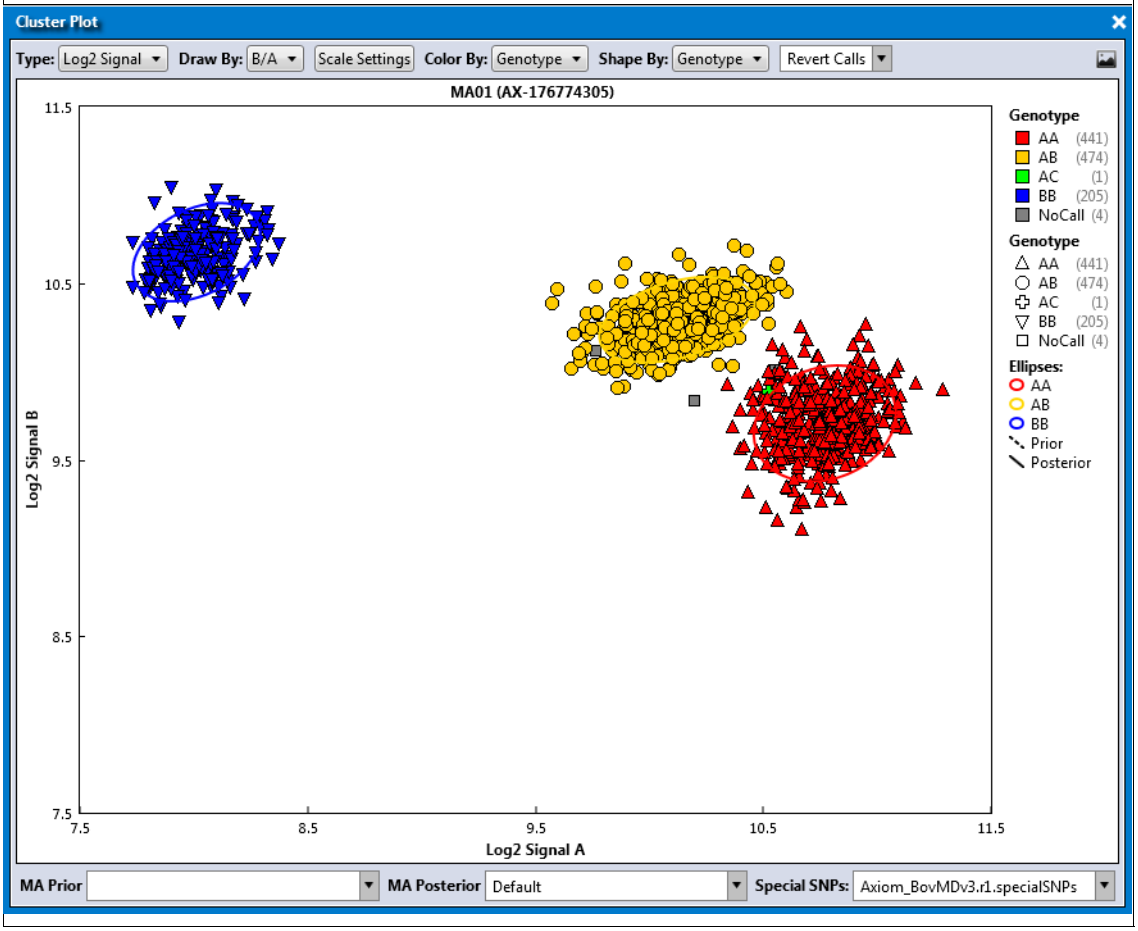
Figure 114 Alternate marker name



3. Click **Open**.

   Your Alternate Marker ID is now displayed in the Probeset Summary Table and Cluster Plot. (Figure 115)

Figure 115 Alternate Marker ID in Cluster Plot



Removing a probeset list from a batch

1. From the **Manage ProbeSet List** drop-down, select **Remove ProbeSet List from batch**,

   A Remove ProbeSet list window appears.

2. Select the ProbeSet list name you want to remove from the drop-down, then click **OK**.

   Your selected ProbeSet List name is removed.

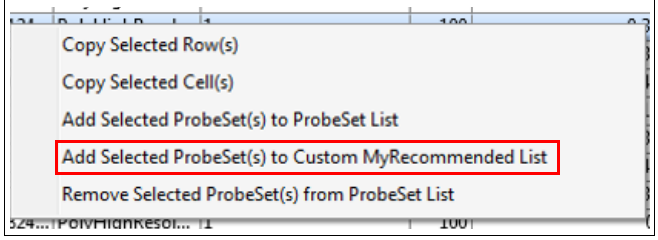
Creating a custom recommended probeset list

You have the ability to add probesets of interest to recommended list for downstream exporting.

1. Click to select the probeset(s) in the ProbeSet Summary Table.
2. Right-click on the table.

   A pop-up menu appears. (Figure 116)

Figure 116 ProbeSet Table - pop-up window



3. Click **Add Selected ProbeSet(s) to Custom MyRecommended List**.

The probeset(s) is now be added to your recommended probeset list. The resulting recommended list can be found in your batch snplist folder. and is named **MyRecommended.ps**.

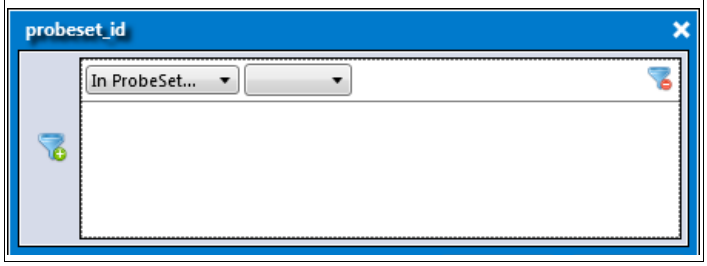
Using your saved probeset list

Displaying Probesets in a Probeset list

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

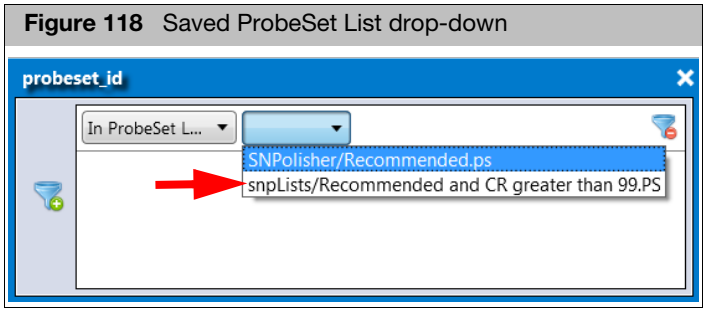
   The Filter window appears. (Figure 117)

Figure 117 Filter window



3. Add your previously saved Probeset List by selecting it from the drop-down list. (Figure 118)

Figure 118 Saved ProbeSet List drop-down



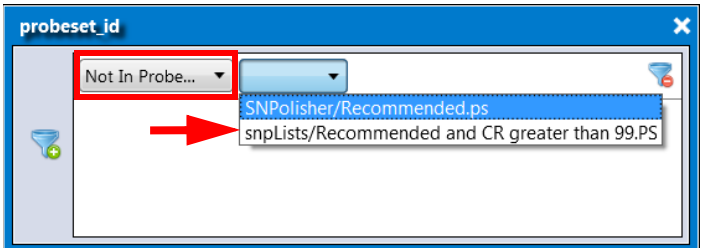
4. Click **OK**.

   Only the Probesets in your ProbeSet List are displayed in the ProbeSet Summary Table.

Displaying Probesets that are not in your Probeset list

1. Click to select the **probeset_id** column, then right-click on it.
2. Click **Filter**.
3. Click the **In ProbeSet List** drop-down, then click to select **Not in ProbeSet List**, as shown in Figure 119.
4. Click the **Saved ProbeSet List** drop-down, select your saved ProbeSet List, as also shown in Figure 119, then click **OK**.

Figure 119 Saved ProbeSet List drop-down



Your Probesets from the ProbeSet List are no longer displayed in the ProbeSet Summary Table.

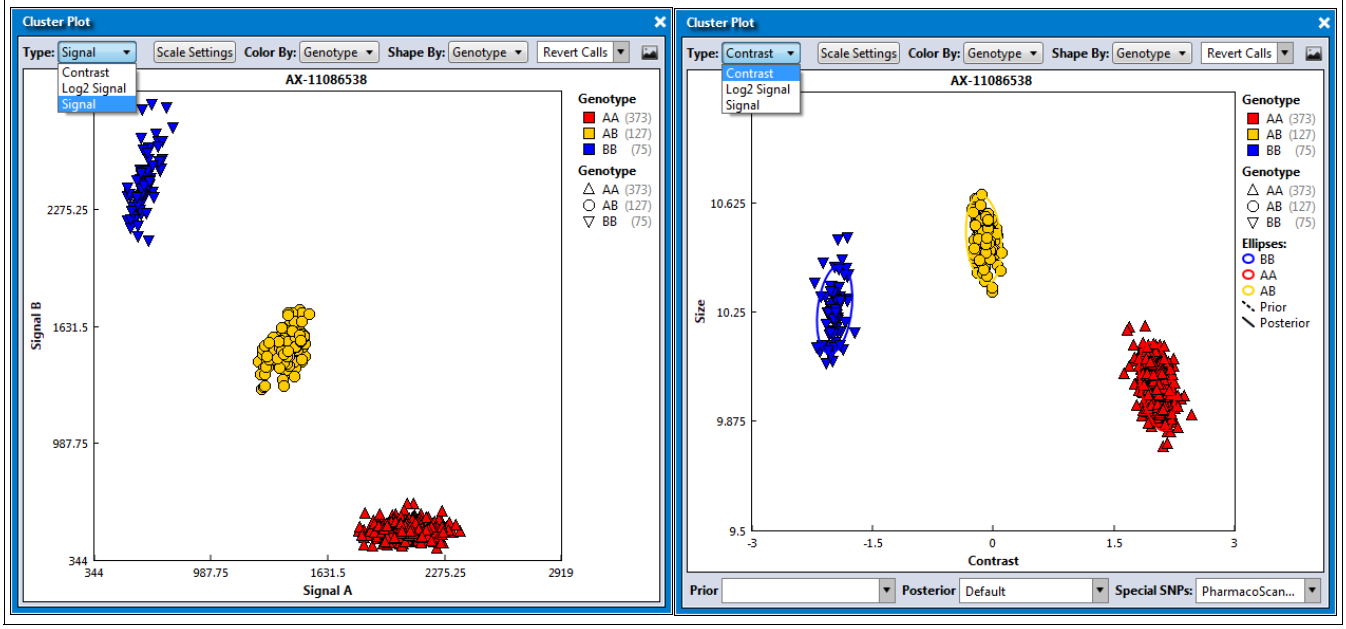
Cluster plot

The Cluster Plot displays the Probeset 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 Probesets and enables you to manually change calls.

As shown in Figure 120, 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 biallelic probesets (those that measure two alleles), only the **Contrast** view displays the ellipses that illustrate the prior and posterior knowledge of genotype cluster positions.

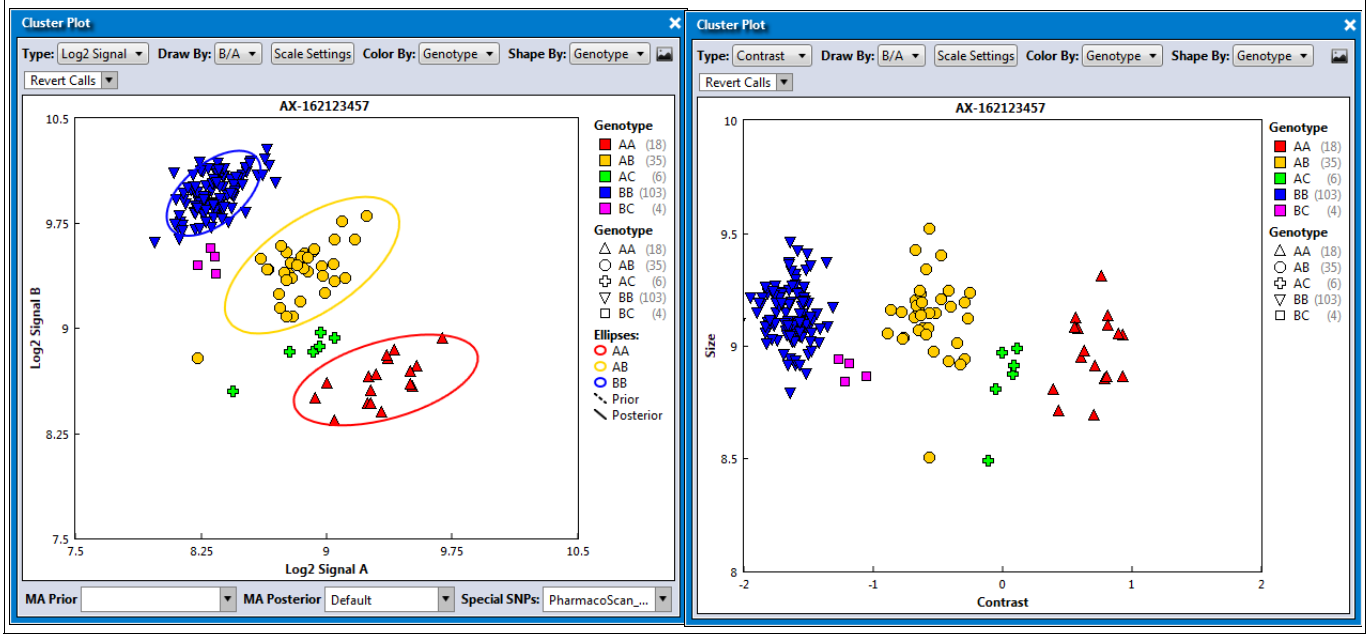
Figure 120 Cluster window tab for biallelic probesets



Multiallelic probeset visualization

Multiallele probesets measure more than two alleles. Unlike biallelic probesets, multiallele probesets are genotyped in the Log2 Signal space. For multiallele 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 121 on page 91 for examples of a multiallele probeset.

Figure 121 Cluster window tab for a multiallele probeset

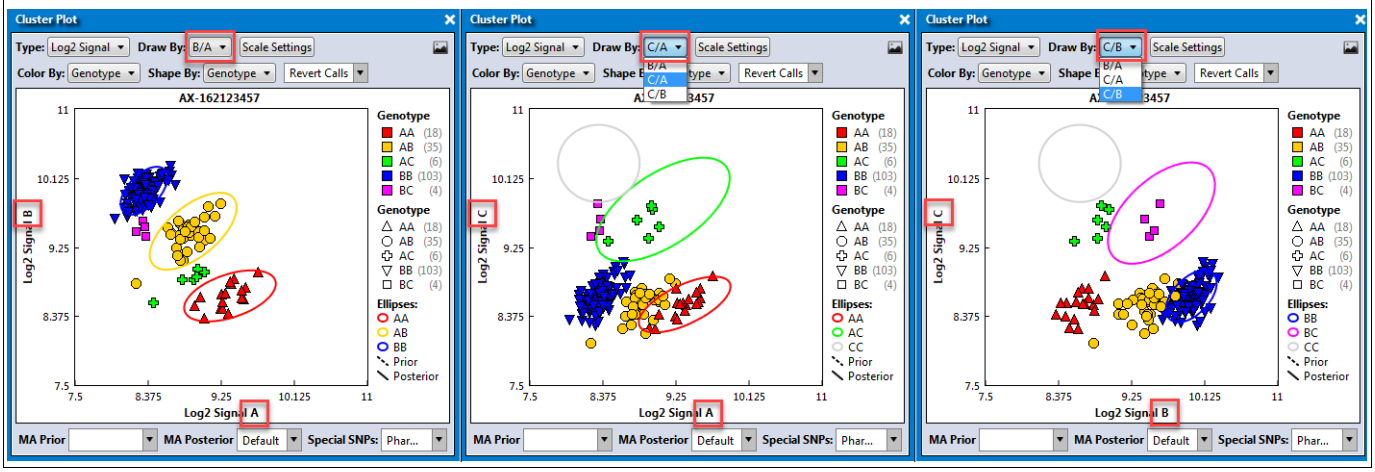


Note: Multiallele probesets in the ProbeSet Summary Table can be more easily identified by displaying the multiallelic column. If a probeset has a multiallelic count of one, it is deemed a multiallele probeset.

Multiallelic probesets have more than one alternate allele, and so more than two allele signal channels are present. The cluster plot only shows data for two allele signal channels. If multiallele probesets are present in the current results, then the top of the cluster plot will also display the Draw By menu.

See Figure 122 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 122 Cluster window tab - Using Draw By to select different pairs of allele signal channels for plotting.



Using the cluster plot

Displaying a Probeset cluster plot that corresponds with a Probeset

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

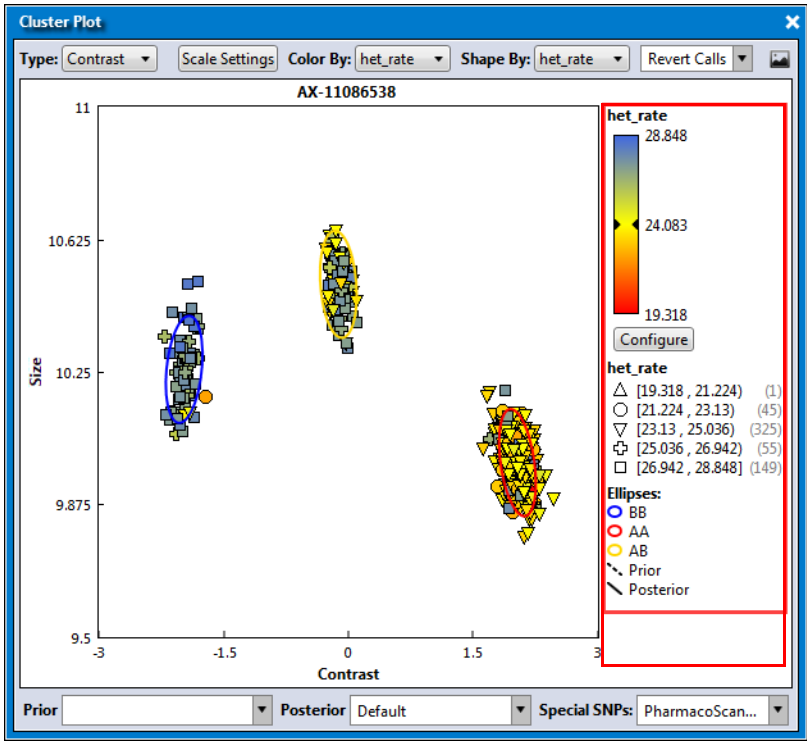
   Note: Use the arrow keys on the keyboard to toggle through the list. As you toggle through the list, the Cluster Plot auto-updates to match your selected Probeset.

2. Click the **Color By** and **Shape By** drop-down menus to select the combination (X and Y axis) view you want. See Table 23 on page 222 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.

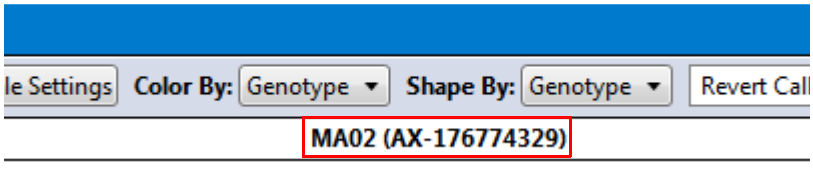
   The appropriate legend appears within the plot. (Figure 123)

Figure 123 Color By and Shape By Legend example



Note: If an Alternate Marker Name has been imported into your ProbeSet Summary Table, the cluster plot title displays your Alternate marker name and probeset id, as shown in Figure 124.

Figure 124 Probeset and Alternate Marker ID in Cluster Plot name

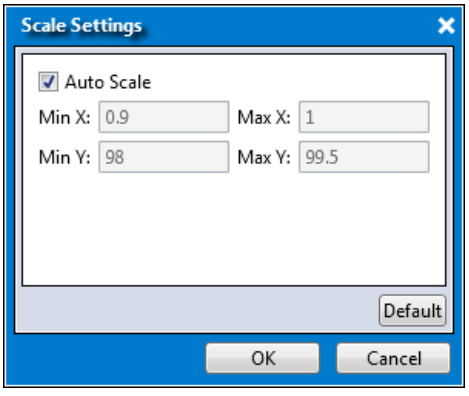


Setting new scale setting ranges

1. Click Scale Settings

   The Scale Settings window appears. (Figure 125)

Figure 125 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**.

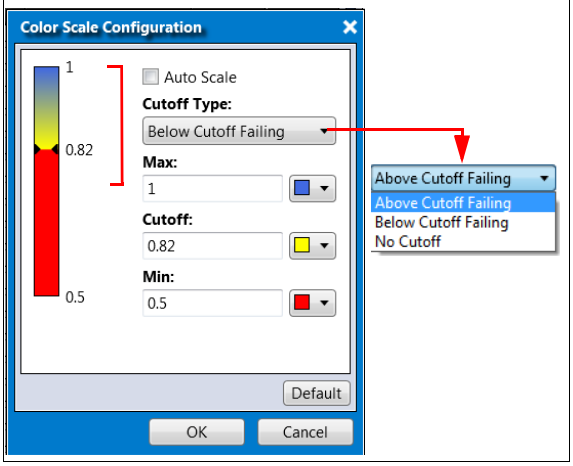
   Click Default to return all values to their factory settings.

Customizing your cluster plot view settings

1. Click Configure

   The Color Scale Configuration window appears. (Figure 126)

Figure 126 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.

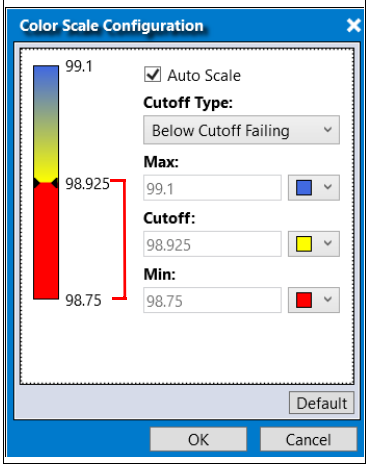
Note: If the Auto Scale check box remains unchecked, you must enter new scale limits for each subsequent analysis.

- 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. (**Blue** in Figure 126)

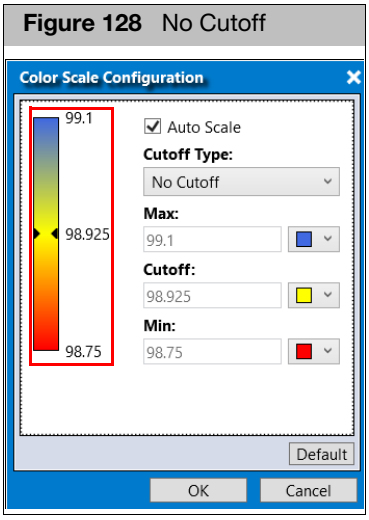
– **Below Cutoff Failing -** This presents a hard visual cutoff graph of all values that fail BELOW the Cutoff value entered. The Below Cutoff data is represented by the color defined for Min. (**Red** in Figure 127)

Figure 127 Below Cutoff



– **No Cutoff -** This presents a smooth 3-point gradient of your defined Max, Min, and colors. (Figure 128).

Figure 128 No Cutoff



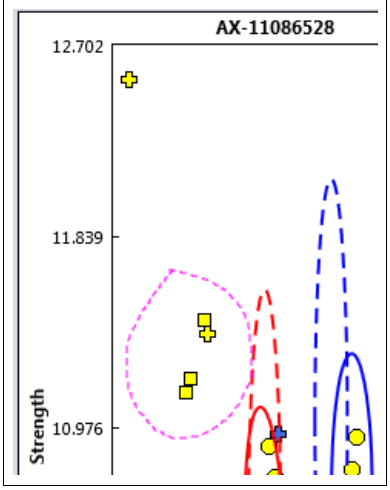
3. Click **OK**.

   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

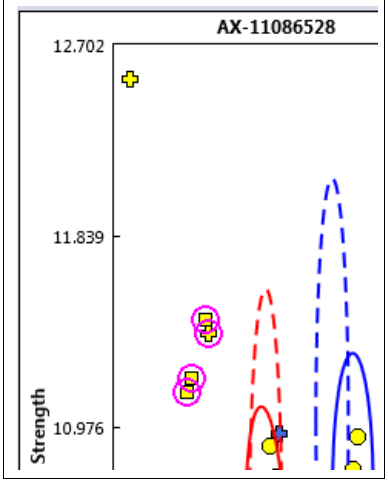
1. Drag the cursor around a group of samples to draw a pink-dotted closed loop around them, as shown in Figure 129.

Figure 129 Selecting multiple samples



2. Release the mouse button to select (lock in) the group of lassoed samples, as shown in Figure 130.

Figure 130 Selecting multiple samples



Changing a sample's call for a single probeset

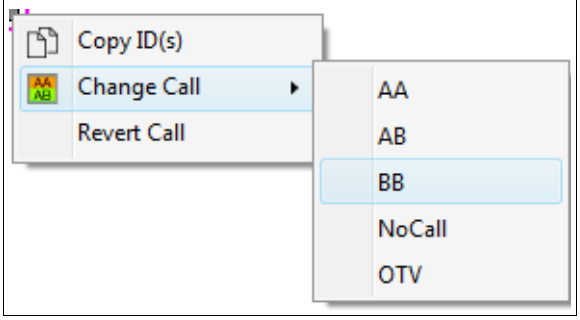
1. Highlight the sample or samples you want to modify, then right-click on them.

   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 131)

   Note: If the array supports CN-aware Genotyping, additional Call Code choices appear.

Figure 131 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.

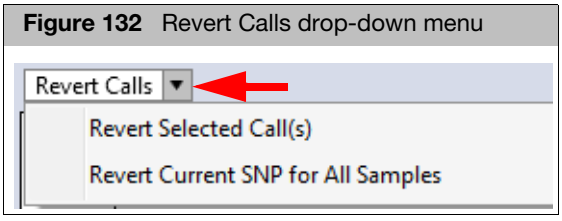
   A menu appears.

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 132), then click to select **Revert Selected Call(s)**.

Figure 132 Revert Calls drop-down menu

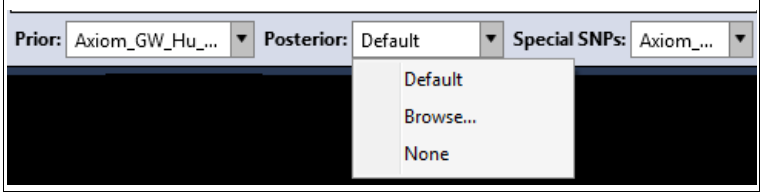


Displaying cluster model data

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

1. Click the appropriate drop-down menu (Figure 133), then click to select a new setting.

Figure 133 Biallele Cluster Model Data drop-down menus



- If you select **Browse**, an Explorer window appears. Navigate to your folder location, 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  button (upper right).

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

5 Copy number analysis

Overview

Copy Number (CN) analysis capabilities are enabled for newer library file versions of some array types.

The Copy Number Discovery workflow supports genome-wide copy number variation (CNV) breakpoint detection in targeted regions (*de novo* discovery). The Copy Number Fixed Regions workflow enables CNV calling in defined regions within the genome.

Note: Loss of Heterozygosity (LOH) detection is enabled for each of these workflows.

IMPORTANT! Batch sizes containing 2000 samples have been validated on a Windows 11 64-bit system. See "Software and hardware requirements" on page 11 and "Validated space requirements and run times for CN batches" on page 12 for more information.

Some library files do not include a CN reference file, which is needed for CN analysis. Library files that include a CN reference file can benefit from making a new CN reference file that is more appropriate for the samples run in the lab.

To make the first CN reference file, the library files may support either the **Copy Number Reference Creation** workflow, or the **Initial Copy Number Reference Creation** workflow.

The newer Initial workflow requires a training plate of samples available from Thermo Fisher. The library files may also support the **Best Practices Copy Number Reference Creation** workflow. This workflow has additional checks to improve the quality of the CN reference.

Note: For more information on copy number workflows, see the *Axiom Copy Number Data Analysis Guide (P/N 703518)*.

Copy number reference creation

The Copy Number Reference Creation workflow generates a CN Reference File that manages the expected probeset signal levels for copy neutral samples. If the CN Reference File does not exist for your Array Type, you must run the Copy Number Reference Creation workflow before you can run other copy number workflows.

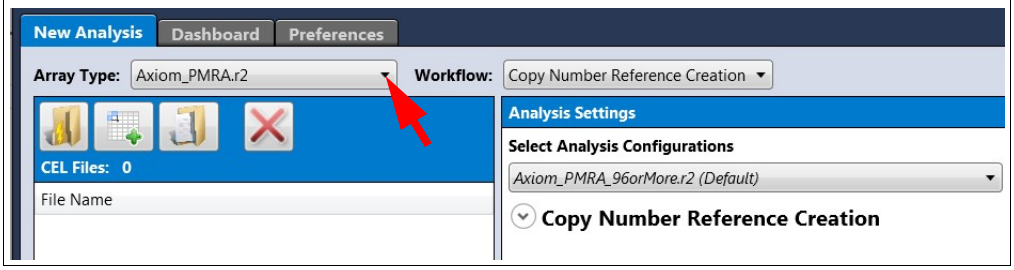
This workflow generates a reference file using all samples imported into the CEL file pane.

IMPORTANT! All samples used must have passed Sample QC and be CN neutral. Also, it is recommended that a CN Reference be generated using at least 80 samples; 40 of each sex.

If you have an array that supports fixed region CN analysis as part of the Genotyping workflow, go to page 183. If not, go to Step 1 on page 99.

1. From the New Analysis window tab, click the **Array Type** drop-down (Figure 134) to select the array type you want to use.

Figure 134 Main window

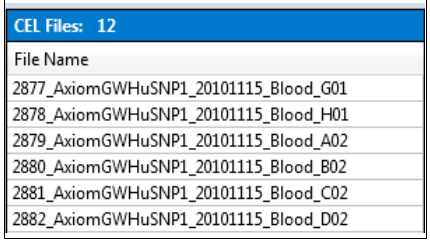


2. Click on the Workflow drop-down, then select **Copy Number Reference Creation**.
3. Click **Import CEL Files**.

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

   The CEL Files pane populates and displays your selected files. (Figure 135)

Figure 135 Populated CEL File pane example



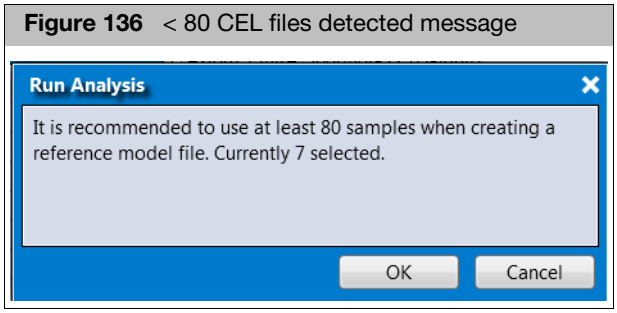
7. (Optional) To change the default annotation file, click the **Copy Number Reference Creation** down arrow to expand the section, then click to highlight the annotation file you want to use to create your reference file.

IMPORTANT! The Output folder path is set by default and cannot be changed. C:\Users\Public\Documents\AxiomAnalysisSuite\Library\Array\custom

8. Enter a name in the **Name** field, then click **Run**.

   AxAS performs an initial count to verify there is at least 80 CEL files loaded. If less than 80 are detected, a message appears. (Figure 136)

Figure 136 < 80 CEL files detected message



9. Click **Cancel** to import more CEL files or click **OK** to continue.

   The Dashboard window tab appears and displays the progress. **Note:** A Warning appears if the analysis completes with less than 40 male and female or unknown sex samples.

   Note: There is no **Open** button in the Dashboard's Action column, because CN Reference workflows create additional library files and therefore cannot be opened in the AxAS Viewer.

CN reference file creation with CN region calls file option

Note: This is an advanced user workflow.

When there are regions that are not normally CN neutral for a population, you have the option of creating a reference in which you designate the CN neutral samples for those regions in question.

Prior to running this advanced workflow, you must first create a custom CN Region Calls file. This is a recommended file that is imported into the reference file creation workflow.

Note: A standard reference file must have been created or be part of the array library file package.

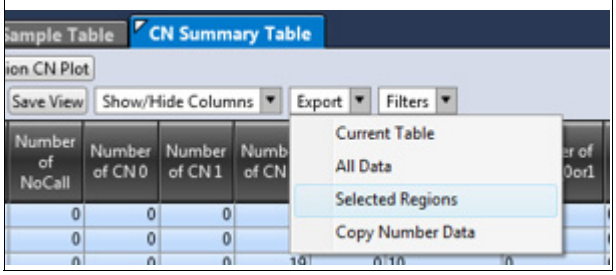
Creating a custom CN Region Calls file

1. Open a successfully completed Copy Number Fixed Regions batch created using the desired reference CEL files in the Viewer.

   Note: For PGx-enabled arrays (like Axiom_PMDA_Plus, PharmacoScan) there is no "Copy Number Fixed Regions batch". Instead open a "Genotyping" or "Best Practices Workflow" batch.

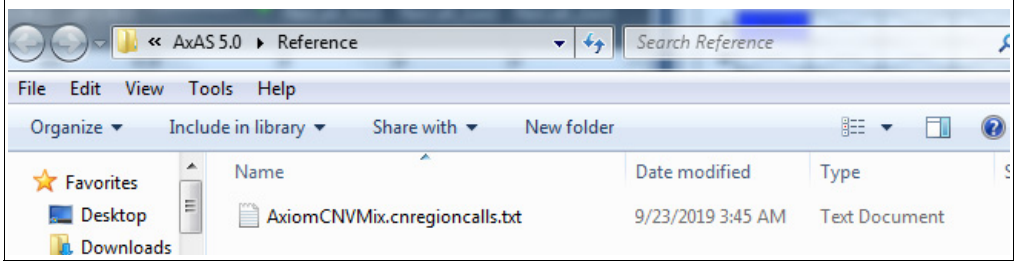
2. In the CN Summary Table, select the CN Regions for which you want to select specific CN neutral samples for Copy Number Reference Creation.
3. At the CN Summary Table, click **Export** → **Selected Regions** (Figure 137), then create a new or locate an empty folder to save results.

Figure 137 Export drop-down menu



4. (Optional) To edit, locate the exported **.cnregionscalls.txt** file (Figure 138), then open it in a text editor (e.g. Excel).

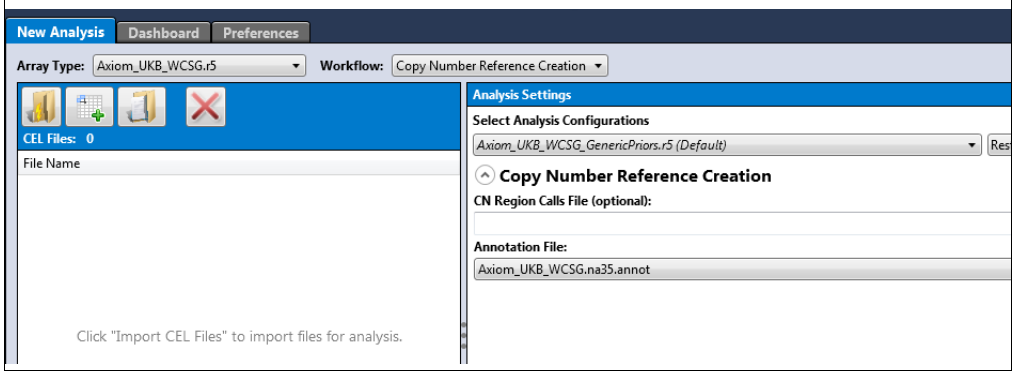
Figure 138 Exported cnregionscalls.txt file



The CEL files that are reporting **CN_State = 2** in a region are used to determine the normal signal for that region. If you want to remove a CEL file, delete the entire row. After completing your edits, save your ***.cnregioncalls.txt** file.

5. Open AxAS, then click on the **New Analysis** window tab.
6. Click the Workflow drop-down to select **Copy Number Reference Creation**. (Figure 139)

Figure 139 Analysis Settings pane - Copy Number Reference Creation



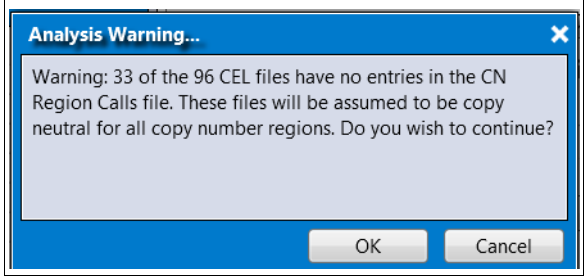
7. In the Analysis Setting pane, expand Copy Number Reference Creation section.
8. Click the ... button, then navigate to the (optionally-edited) **.cnregioncalls.txt** file that you exported earlier.
9. (Optional) To change the default annotation file, click the **Annotation File** down arrow, then click to highlight the annotation file you want to use.

   AxAS performs an initial count to verify there is at least 80 CEL files loaded. If less than 80 are detected, a message appears.

   In addition to the checks listed in the standard CN Reference workflow procedure, AxAS checks and displays a warning message (Figure 140) if the CEL files listed in CEL file pane are not listed in the CN Region Calls file. If you choose not to proceed and click **Cancel**, the CEL files not listed in the CN Region Calls file are displayed.
10. Enter a name for the new reference model in the Name field, then click **Run**.

    Note: Reference file creation may take several hours to complete.

Figure 140 Analysis Warning message



11. Click **OK** to continue.

    The Dashboard window tab appears and displays the progress. **Note:** A Warning appears if the analysis completes with less than 40 male and female or unknown sex samples.

    Note: There is no **Open** button in the Dashboard's Action column, because CN Reference workflows only create additional library files that cannot be opened in the AxAS Viewer.

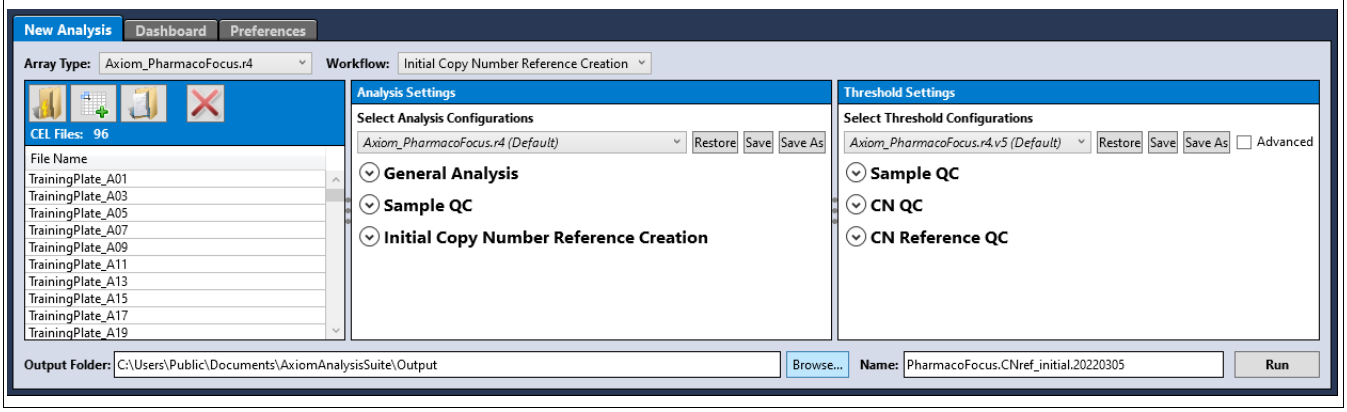
Initial copy number reference creation

Library files that do not include a CN Reference file may support the Initial Copy Number Reference Creation workflow. This workflow requires CEL files from supported plates like the **Axiom DNA Training Plate, 96F**.

Note: The CN Reference file created by this workflow should only be used until enough plates are available to run the Best Practices Copy Number Reference Creation workflow.

1. From the **Array Type** drop-down, select your array type.
2. In Analysis Settings, select your Analysis Configuration.
3. From the **Workflow** drop-down select your workflow, in this case *Initial Copy Number Reference Creation*.

Figure 141 Copy Number Discovery option



   The **Name** field (lower right) auto-populates, but can be edited in its text field.

4. In **Threshold Settings**, select your Threshold Configuration.
5. Optional: If you need to add more CEL files, click the **Add CEL files** button.
6. Click **Run**.

   A **Run Analysis** pop-up window appears.

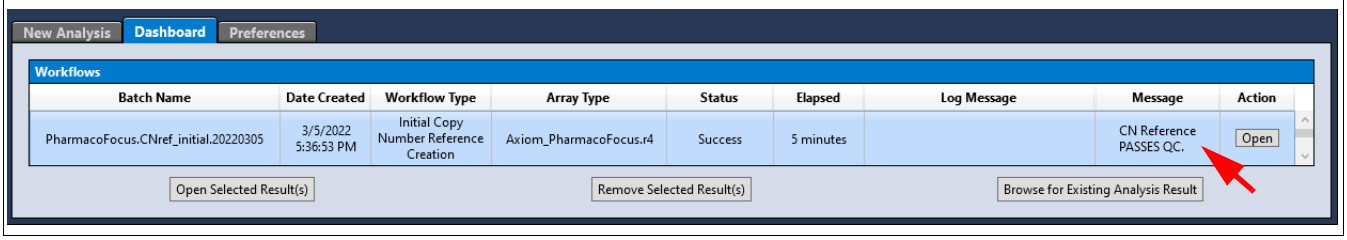
7. Note where the CN reference will be saved, then click **OK**.

   The Dashboard appears and displays the progress of your workflow.

Reviewing the initial copy number reference creation batch

When the workflow successfully completes, the CN reference is saved to the library folder. Confirm the **Message** in the Dashboard states **CN Reference PASSES QC**, as shown in Figure 142. If not, the CN Reference File should not be used for other CN workflows.

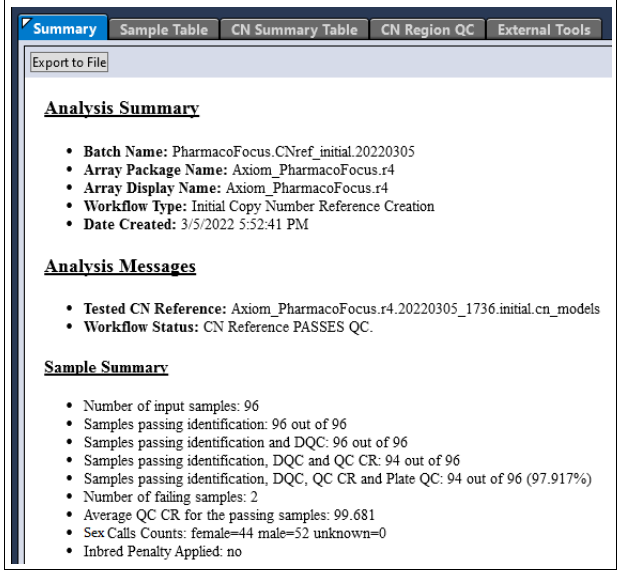
Figure 142 Dashboard tab window



1. Click the Open button (far right) to review CN reference file test results.

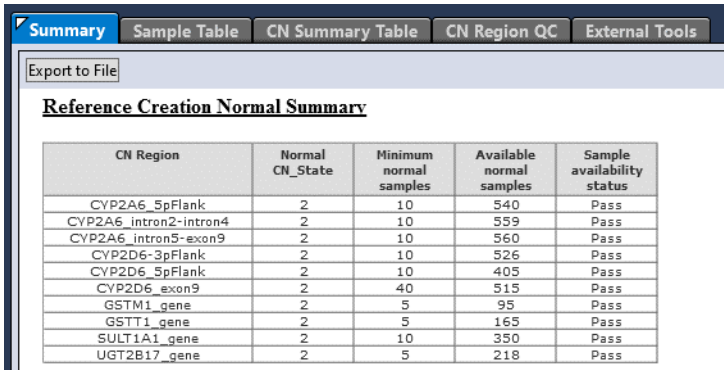
   The batch opens in the Viewer's Summary tab window. (Figure 143)

Figure 143 Summary tab window



- The Sample Summary section includes a "Samples passing identification" metric. CEL files are matched to known samples before DishQC is calculated. Samples pass the identification check if they match to known samples AND the samples are arranged on the plate in one of a few allowed arrangements. These checks ensure that expected samples are used when making the important initial reference.
- In the Summary tab, the Reference Creation Normal Summary table (Figure 144) reports how many normal copy number samples pass QC, and so are available to make a reference for each CN Region. A passing CN reference needs to measure signals from a minimum number of normal copy number samples for each region.

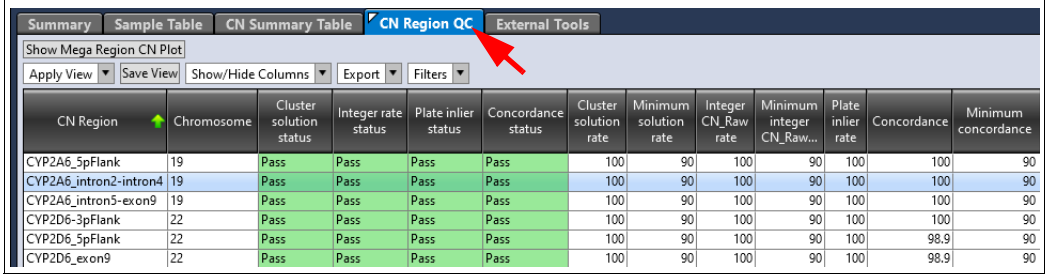
Figure 144 Summary tab window (Reference Creation Normal Summary



CN Region	Normal CN_State	Minimum normal samples	Available normal samples	Sample availability status
CYP2A6_5pFlank	2	10	540	Pass
CYP2A6_intron2-intron4	2	10	559	Pass
CYP2A6_intron5-exon9	2	10	560	Pass
CYP2D6-3pFlank	2	10	526	Pass
CYP2D6_5pFlank	2	10	405	Pass
CYP2D6_exon9	2	40	515	Pass
GSTM1_gene	2	5	95	Pass
GSTT1_gene	2	5	165	Pass
SULT1A1_gene	2	10	350	Pass
UGT2B17_gene	2	5	218	Pass

2. Click on the CN Region QC tab. Confirm that the status column values are Pass for all CN Regions (green background), as shown in Figure 145.

Figure 145 Dashboard tab window



CN Region	Chromosome	Cluster solution status	Integer rate status	Plate inlier status	Concordance status	Cluster solution rate	Minimum solution rate	Integer CN_Raw rate	Minimum integer CN_Raw...	Plate inlier rate	Concordance	Minimum concordance
CYP2A6_5pFlank	19	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2A6_intron2-intron4	19	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2A6_intron5-exon9	19	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2D6-3pFlank	22	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2D6_5pFlank	22	Pass	Pass	Pass	Pass	100	90	100	90	100	98.9	90
CYP2D6_exon9	22	Pass	Pass	Pass	Pass	100	90	100	90	100	98.9	90

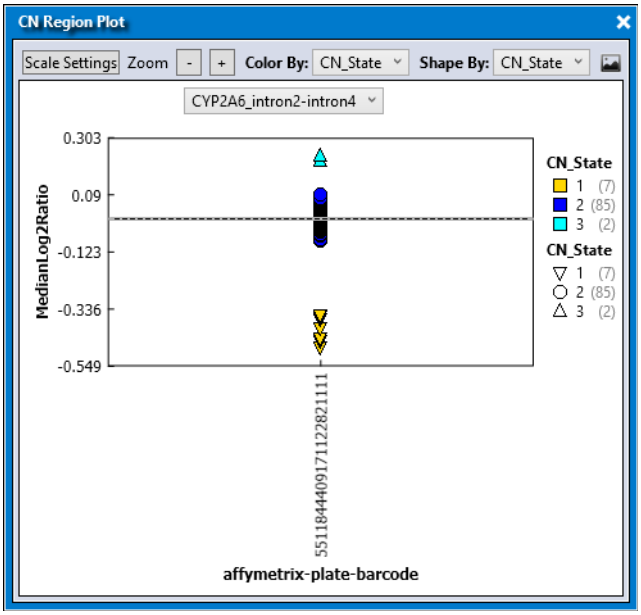
CN Regions table definitions

Column	Description
Cluster solution status	'Pass' if 'Cluster solution rate' is at least 'Minimum solution rate'; 'Fail' otherwise.
Integer rate status	'Pass' if 'Integer CN_raw' rate is at least 'Minimum integer CN_raw rate'; 'Fail' otherwise.
Plate inlier status	'Pass' if 'Plate inlier rate' is above a threshold; 'Review' otherwise.
Concordance status	'Pass' if 'Concordance' is at least 'Minimum concordance';'Fail' otherwise.
Minimum concordance	The required minimum concordance for the region, expressed as a percentage.
Cluster solution rate	Percentage of plates for which at least some samples are assigned a call other than NoCall.
Minimum solution rate	The required minimum percentage of fits with a valid solution for the region
Integer CN_Raw rat	Percentage of calls with integer CN_Raw values (i.e. 1,2,3), excluding samples that fail CN QC.
Minimum integer call rate	The required minimum percentage of integer (non-intermediate) calls for the region, after excluding NoCalls from CN QC failed samples and fits with no valid solution.

Column	Description
Plate inlier rate	Percentage of plates whose distribution of CN_State calls is consistent with at least one population's expected distribution of CN_State calls or is consistent with the CN_State calls of the other plates in the batch.
Concordance	Percentage of CN_State calls that agree with expected calls, for samples matched to known samples. The library files include reference CN data for some 1000Genomes samples and the Ref103 control sample.

3. Select a CN Region in the CN Region QC table. The Region is displayed in the CN Region Plot, as shown in Figure 146.

Figure 146 CN Region Plot window



4. Scroll through all the CN Regions to see if the CN2 samples are centered on MedianLog2Ratio=0 (the dashed line), as they should be. You can also select the region from the title drop-down menu, and then up/down arrow key to scroll through the plots.

Best practices copy number reference creation

Best Practices Copy Number Reference Creation is a 2-step procedure to make and test an updated CN Reference File. The Step 1 workflow determines which CEL files pass all QC tests. It then uses an existing CN Reference File to identify and select the normal copy number samples that should be used to make a new reference for each CN Region.

To start Step 2 of the procedure, you will need to 1) open Step 1 analysis results in the viewer, 2) review and possibly edit the selected **Reference Normal** samples for each

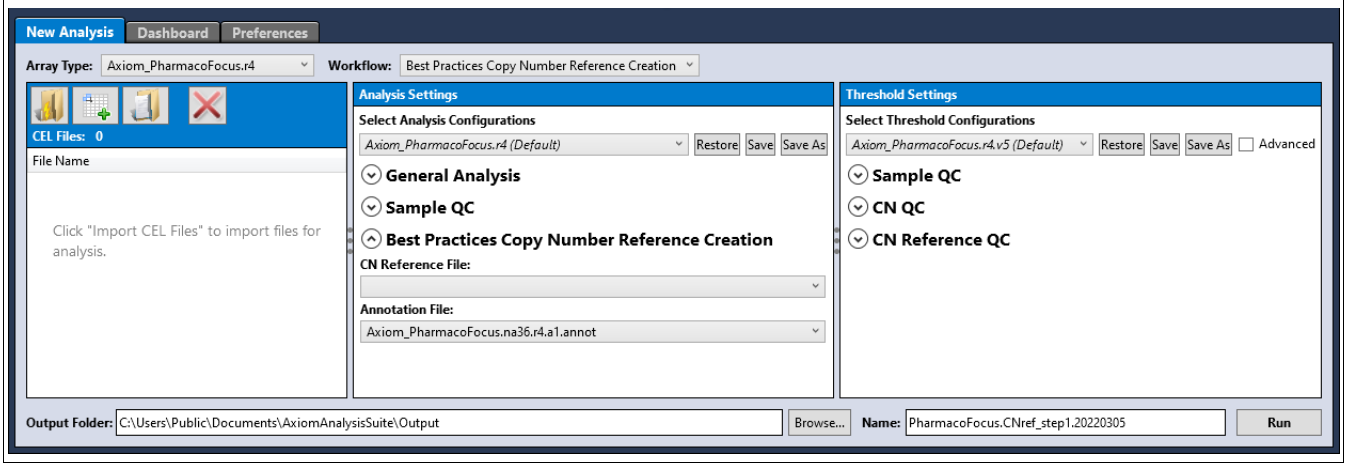
CN Region, and 3) click **Create CN Reference** from the Sample Table tab. Step 2 of the workflow will then create and test the new CN Reference File.

The new CN Reference File can be used for all workflows requiring a CN Reference File.

Note: Although it is possible to run this workflow with only one plate of CEL files, it is recommended that at least six plates be used. If the training plate from Thermo Fisher is available, it should also be included. For more information on copy number workflows, see the *Axiom Copy Number Data Analysis Guide*.

1. From the **Array Type** drop-down, select your array type.
2. In Analysis Settings, select your Analysis Configuration.
3. From the **Workflow** drop-down select your workflow, in this case *Best Practices Copy Number Reference Creation*.

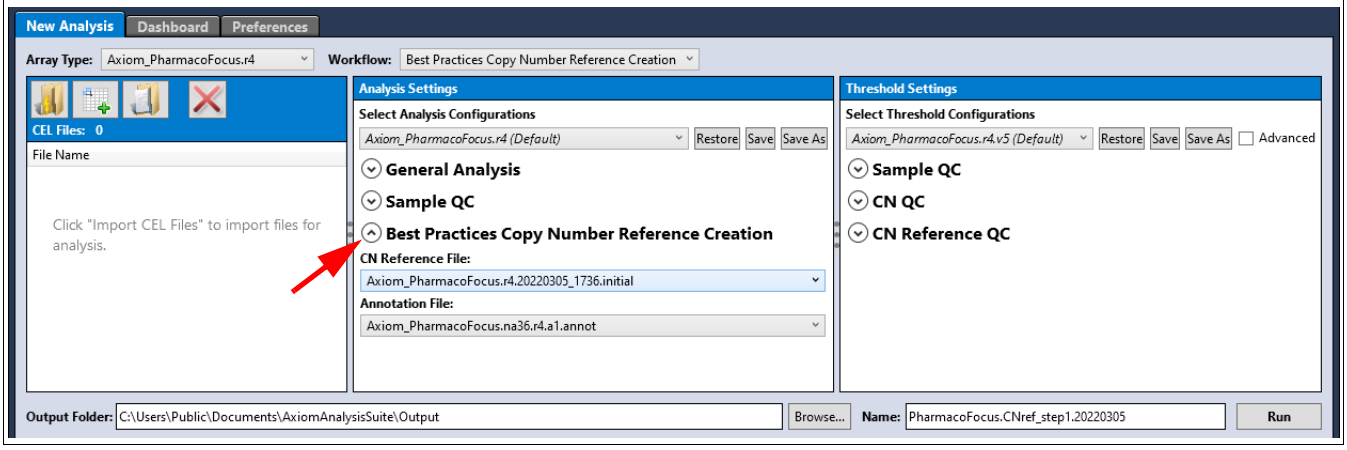
Figure 147 Copy Number Discovery option



The **Name** field auto-populates, but can be edited in its text field.

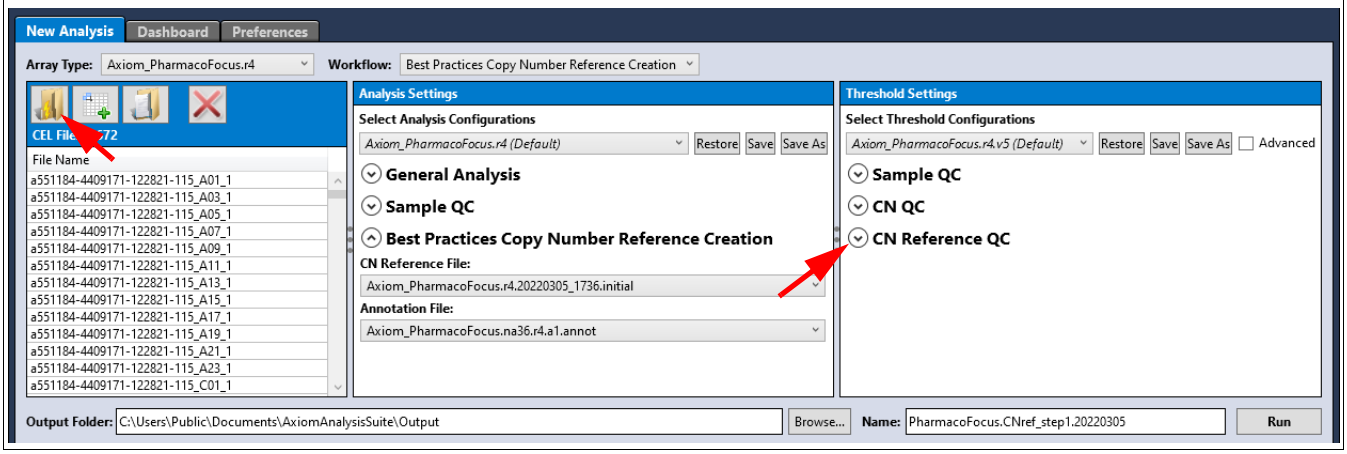
4. In **Threshold Settings**, select your Threshold Configuration.
5. In the Analysis Settings pane, click the **Best Practices Copy Number Reference Creation** arrow to expand it. (Figure 148)

Figure 148 Best Practices Copy Number Reference Creation arrow



6. Click the **CN Reference QC** drop-down arrow button to select an existing copy number reference file. If one doesn't exist, you must create one using a different copy number reference creation workflow.
7. Click the **Add CEL files** button to add CEL files.

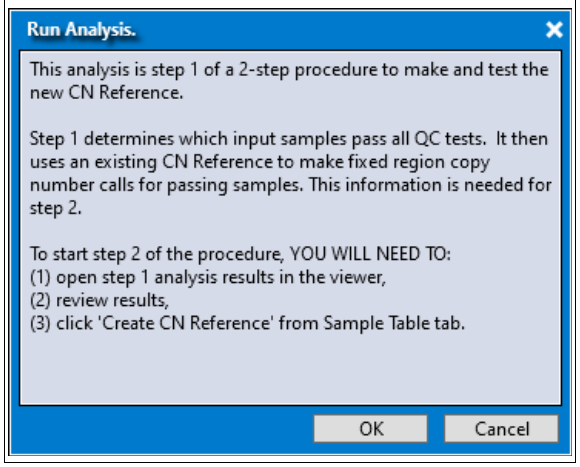
Figure 149 Add CEL files button and CN Reference QC drop-down arrow button



8. Click **Run**.

   A **Run Analysis** pop-up window appears. (Figure 150)

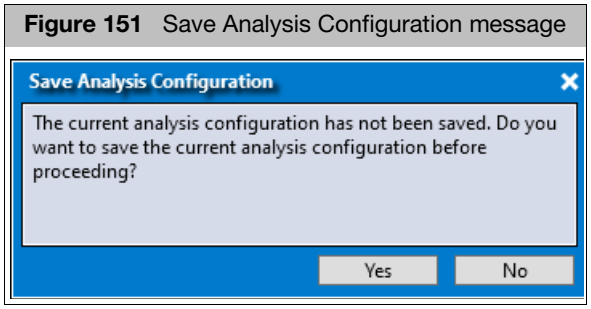
Figure 150 Run Analysis message



9. Read this message carefully to understand what will happen, then click **OK** to proceed.

   A Save Analysis Configuration pop-up window appears. (Figure 151) This message appears because the CN Reference File setting was changed from [empty] to [the initial CN Reference].

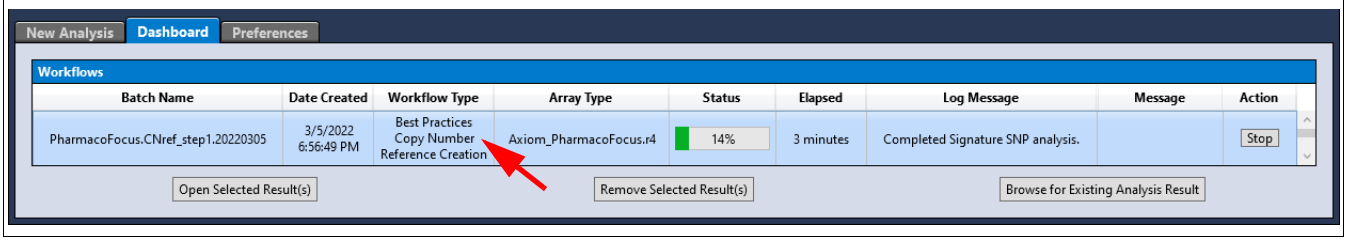
Figure 151 Save Analysis Configuration message



10. Saving the Configuration is not required at this time, therefore you may click **No**.

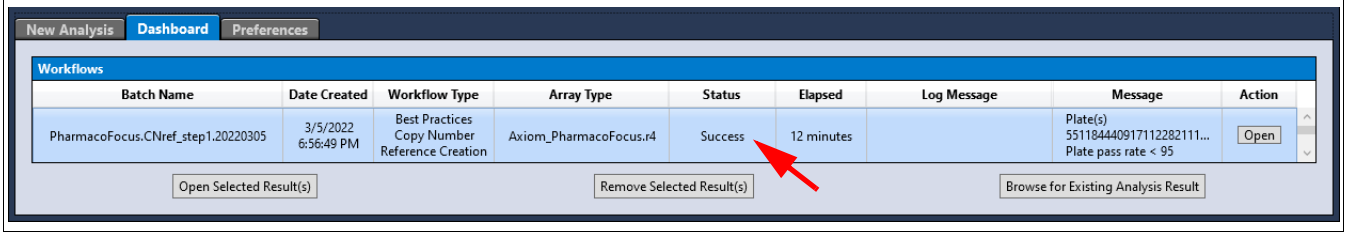
    The Dashboard appears and displays Best Practices Copy Number Reference Creation under the Workflow Type. (Figure 152)

Figure 152 Dashboard: Best Practices Copy Number Reference Creation



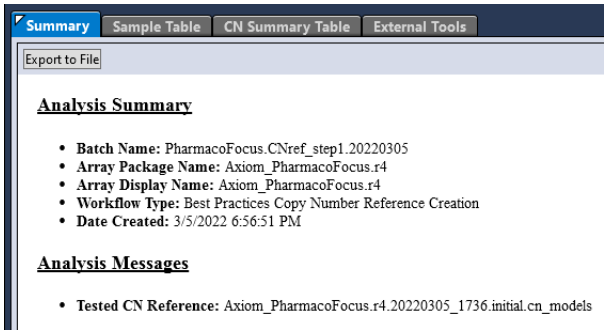
After the run successfully completes, confirm that the Status for Step 1 states Success (Figure 153), then click **Open**.

Figure 153 Dashboard: Status Success



The Summary tab Analysis Messages section describes the initial CN Reference that this workflow created and tested. (Figure 154)

Figure 154 Summary tab window Workflow created



In the Summary tab, check if any plate has an unusually low percent of passing samples. If it appears that a plate suffers from a significant processing issue, no samples from it are appropriate for building a CN reference, and these samples should be excluded later. (Figure 155)

Figure 155 Summary tab window Plate QC Summary

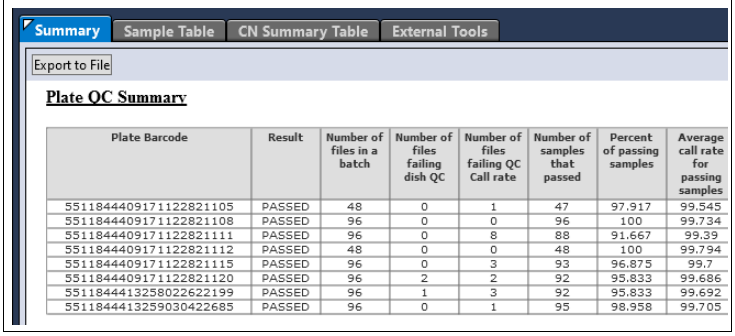
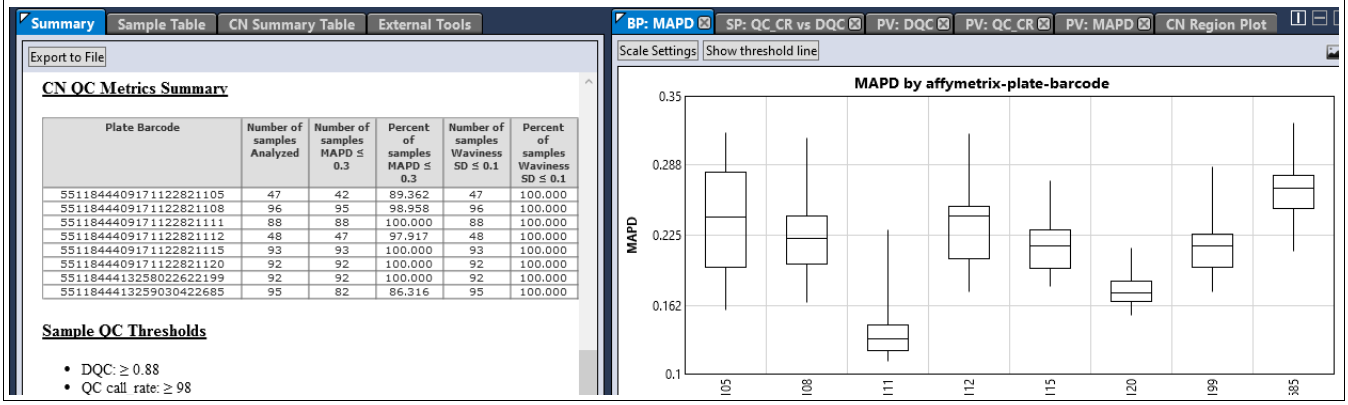


Plate Barcode	Result	Number of files in a batch	Number of files failing dish QC	Number of files failing QC Call rate	Number of samples that passed	Percent of passing samples	Average call rate for passing samples
551184440917112282105	PASSED	48	0	1	47	97.917	99.545
551184440917112282108	PASSED	96	0	0	96	100	99.734
551184440917112282111	PASSED	96	0	8	88	91.667	99.39
551184440917112282112	PASSED	48	0	0	48	100	99.794
551184440917112282115	PASSED	96	0	3	93	96.875	99.7
551184440917112282120	PASSED	96	2	2	92	95.833	99.686
551184441325802262199	PASSED	96	1	3	92	95.833	99.692
551184441325903042268	PASSED	96	0	1	95	98.958	99.705

In the Summary tab, review the percent of samples below the MAPD threshold for each plate. It should be a high percentage for most plates. (Figure 156)

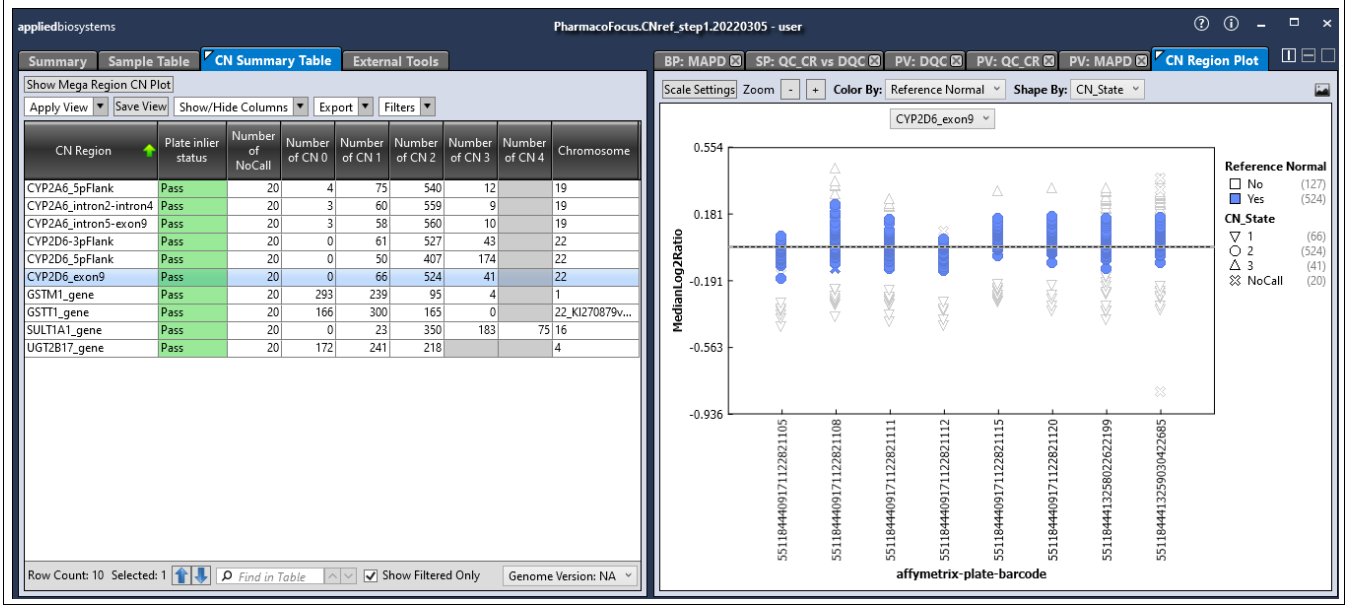
Figure 156 Summary tab window Percentages



The **MAPD Box Plot** (Figure 156) shows how appropriate it is for each plate to be used for the tested CN reference, when doing CN analysis. Samples with higher MAPD values behave more differently from the samples used to make the tested CN reference. One reason to make a new CN reference using this workflow is to lower the MAPD values for most samples. Lowering the MAPD values should reduce the number of false segment calls from the Copy Number Discovery workflow, and increase the call rate and accuracy of fixed regions from the Copy Number Fixed Regions workflow.

11. To review the results for each fixed region, navigate to the CN Summary Table (Figure 157) and in the right section, click on the **CN Region Plot**. Click on each region in the CN Summary Table and view the CN Region Plot.

Figure 157 CN Summary Table tab window



The purpose of step 1 of Best Practices Copy Number Reference Creation is to select the normal CN samples for each CN region, in order to make a good multiplate CN

reference. The selected normal CN samples are labeled 'Reference Normal' in the CN Region Plot. The CN Summary Table, the CN Region Plot, and Mega Region CN Plot are the primary visualizations to help you decide if the software made good Reference Normal selections.

In the CN Summary Table, **Plate inlier status** column has values of Pass or Review for each region. 'Pass' means that, for a given region, AxAS does not detect more than 10% of the plates in the batch to be outliers (threshold set in library files). An outlier plate is one that has unusually few measured normal CN samples, or has BOTH:

a. An unusual distribution of CN calls compared to multiple known populations AND
b. An unusual plate-averaged CN_State compared to the other plates in the batch

When a plate is labeled as an outlier, NO samples from that plate are typically selected as Reference Normal. If a sample is not labeled as Reference Normal, it will not be used when making a CN reference for this CN Region.

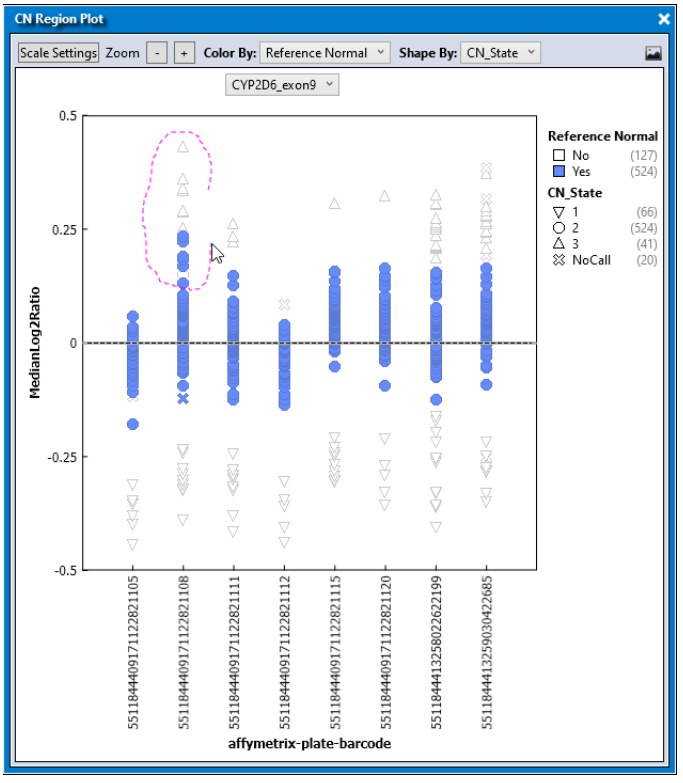
IMPORTANT! If a CN Region has 'Plate inlier status' of 'Review', it should be reviewed to confirm or edit the handling of outlier plates. It is possible that the CN calling algorithm assigned the wrong CN State to the samples.

How to evaluate the CN Region Plot for a step1 Best Practices CN Reference Creation batch:

- The blue data points are the samples AxAS selected as appropriate to use as normal samples when building the CN reference for that region. Different regions will have different 'Reference Normal' samples selected.
- The normal CN samples for each plate should be near the horizontal dashed line, which represents MedianLog2ratio=0. The default shape of the samples is by 'CN_State'. Since there are between-plate effects, it is not unusual for the normal CN sample cluster (usually CN2) to shift a little above and below the line. If the labeled normal CN cluster has shifted too far away from the dashed line, miscalls become more likely.
- Since the normal CN state for most regions is 2, most or all of the CN2-shaped data points should be blue. It is possible for a sample to be labeled as Reference Normal even if the labeled CN_State is not 2, and vice versa. This can happen if AxAS has matched the CEL file to a recognized sample and has knowledge of the expected CN_State.
- If any plate for a region has NO blue data points (Reference Normal samples), that plate is considered an outlier. A plate with no Reference Normal samples will not be used to create a reference for the region (even if it has some samples labeled by shape as CN2).

12. To change which samples are selected as Reference Normal for a region, drag-select (lasso) the data points to edit. (Figure 158)

Figure 158 CN Region Plot window Lasso

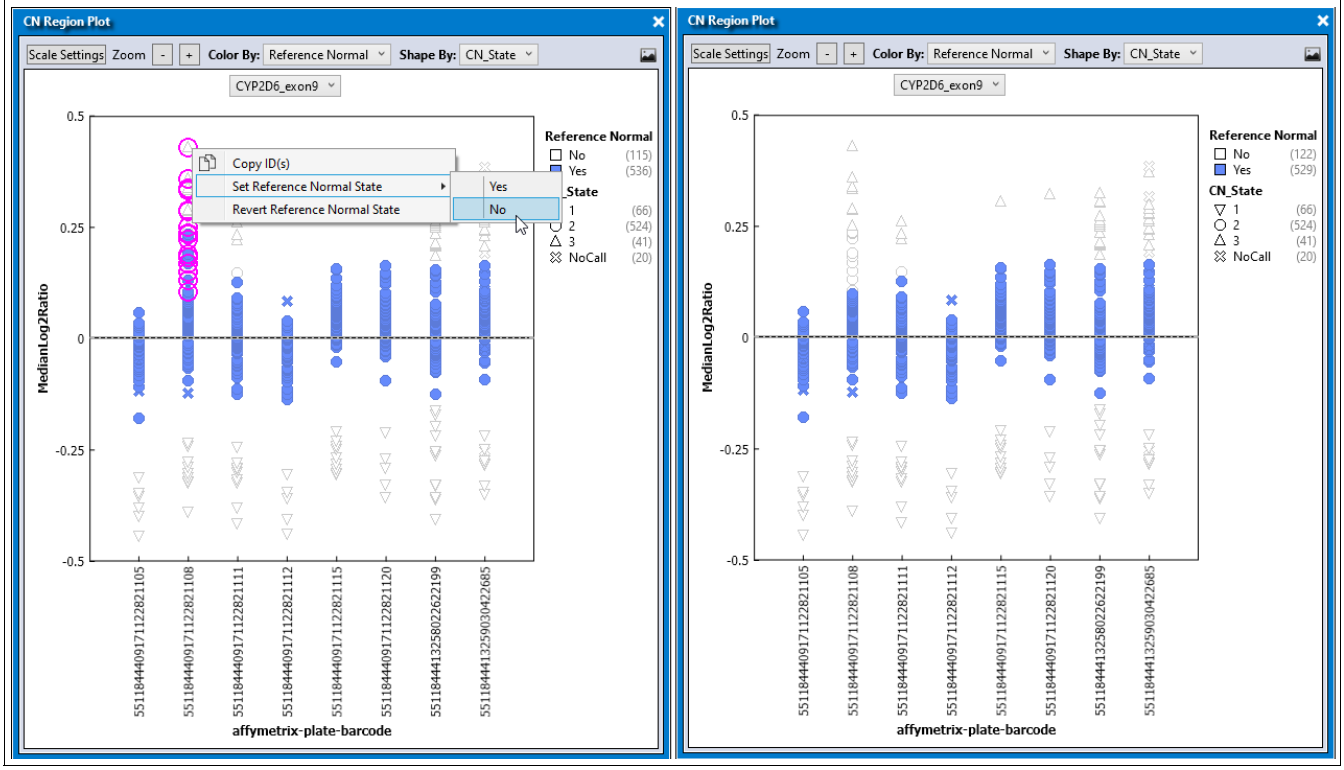


13. Use the secondary mouse button to click on the selected samples. Then select **Set Reference Normal State**, and in the sub-menu, select the desired option. (Figure 159)

    a. To remove samples from the Reference Normal set, select **Set Reference Normal State > No**.

    b. To add samples to the Reference Normal set, select **Set Reference Normal State > Yes**.

Figure 159 CN Region Plot window Change Reference

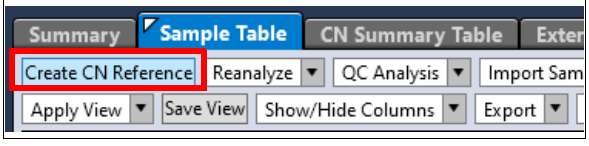


To undo all changes to the Reference Normal set, select some or all samples, then secondary mouse click on them, and select **Revert Reference Normal**. Reverting changes must be done separately for each region.

Note: To make a good CN Reference, it helps if the Reference Normal set includes ONLY normal samples for each region, not samples that appear to have CN gains or losses. It is OK if some of the normal sample are not included, as long as most plates include a high percentage of their normal samples labeled as Reference Normal.

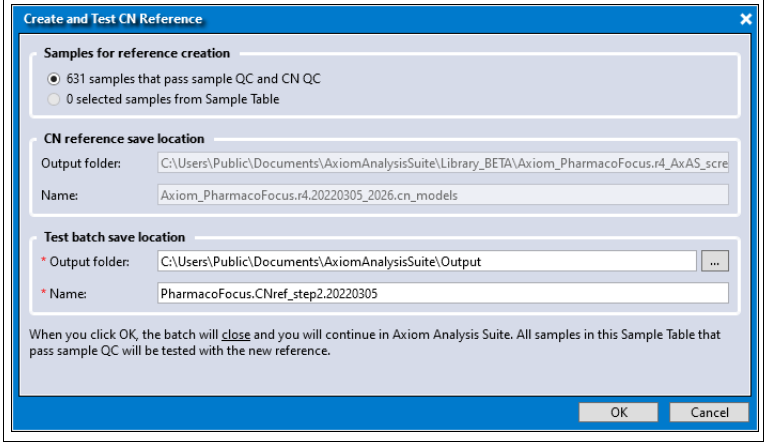
14. After reviewing and optionally editing the samples selected as Reference Normal, it is time to start the second half of this workflow. Click **Create CN Reference** button in the Sample Table tab. (Figure 160)

Figure 160 Create CN Reference button



The Create and Test CN Reference window appears. (Figure 161)

Figure 161 Create and Test CN References window



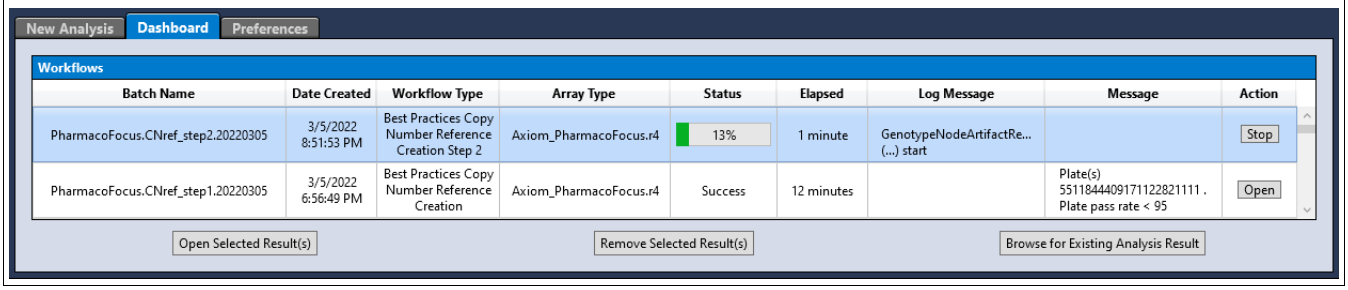
15. In the "Samples for reference creation" section, select the set of samples used to make the new CN Reference. The default option is to use all samples that pass both sample QC and CN QC (Sample Table "Pass/Fail" column = Pass, and "CN passes QC" = yes). If you want to select a different set from the Sample table, then 1) select the desired sample rows from the Sample Table, 2) click Create and Test CN Reference, and 3) select the "selected samples from Sample Table" option.

    Note: Regardless of which samples are selected to make the new CN reference, all samples in this batch that pass sample QC will be tested using the new CN reference.

    - The CN reference save location is shown and cannot be changed
    - The test batch save location is pre-filled, but can be changed.

16. Click **OK** to start the second half of this workflow. The Viewer closes and the Dashboard displays the progress of the new workflow 'Best Practices Copy Number Reference Creation Step 2'. (Figure 162)

Figure 162 Dashboard Best Practices Copy Number Reference Creation Step 2



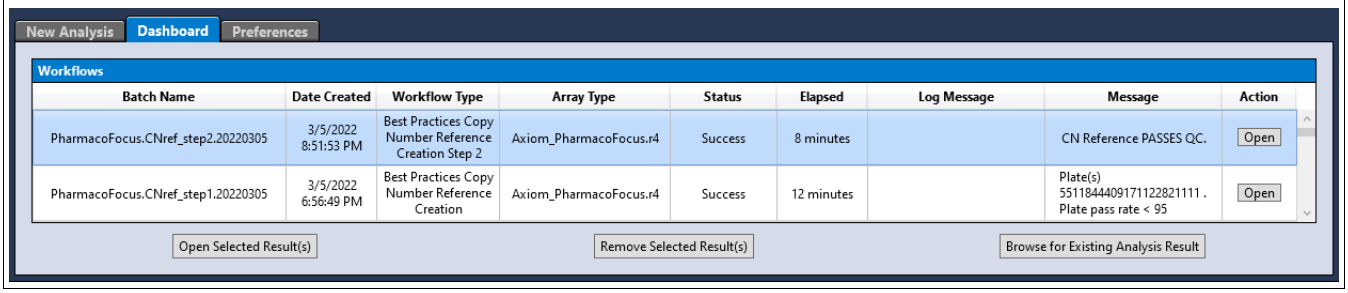
Note: The Step 2 workflow will not complete successfully if the Step 1 batch is moved or renamed while the Step 2 workflow is running.

Note: The only way to create and start Step 2 of this workflow is from an open Step1 batch, not from the New Analysis tab.

Reviewing the Best Practices CN reference

1. From the Dashboard, select the batch that tests the best practices CN reference you just created. Its Workflow Type is **Best Practices Copy Number Reference Creation Step 2**. (Figure 163)

Figure 163 Dashboard Reviewing Best Practices Copy Number Reference Creation Step 2



Test results appear in the Message column of the Dashboard. A good message is 'CN Reference PASSES QC'.

IMPORTANT! At this point the new CN reference exists, and you can use either the previous CN reference or the new CN reference for another workflow. However, it is essential to first review the test results of this step 2 batch. Since the CN results are from samples processed using a CN reference trained on the same data, future CN results on new plates may not look better than these test results. If the current results are not satisfactory, now is the best time to try to improve the CN reference, before using it routinely.

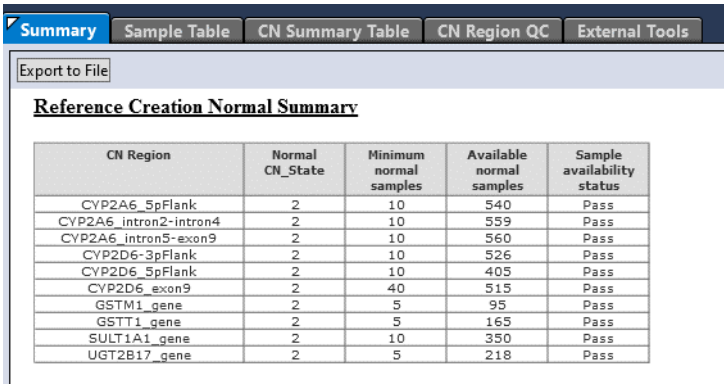
2. To open this batch, click **Open**.
3. From the Summary tab window, review the Analysis Messages section. (Figure 164)

Figure 164 Summary tab window Analysis Messages

The Sample Summary and Plate QC Summary are the same as the Step1 batch from which this Step2 batch was created.

4. Review the number of normal samples used to create the reference for each CN Region in the **Reference Creation Normal Summary** table. (Figure 165) A good CN Reference requires Sample availability status to be a Pass for all regions.

Figure 165 Reference Creation Normal Summary table

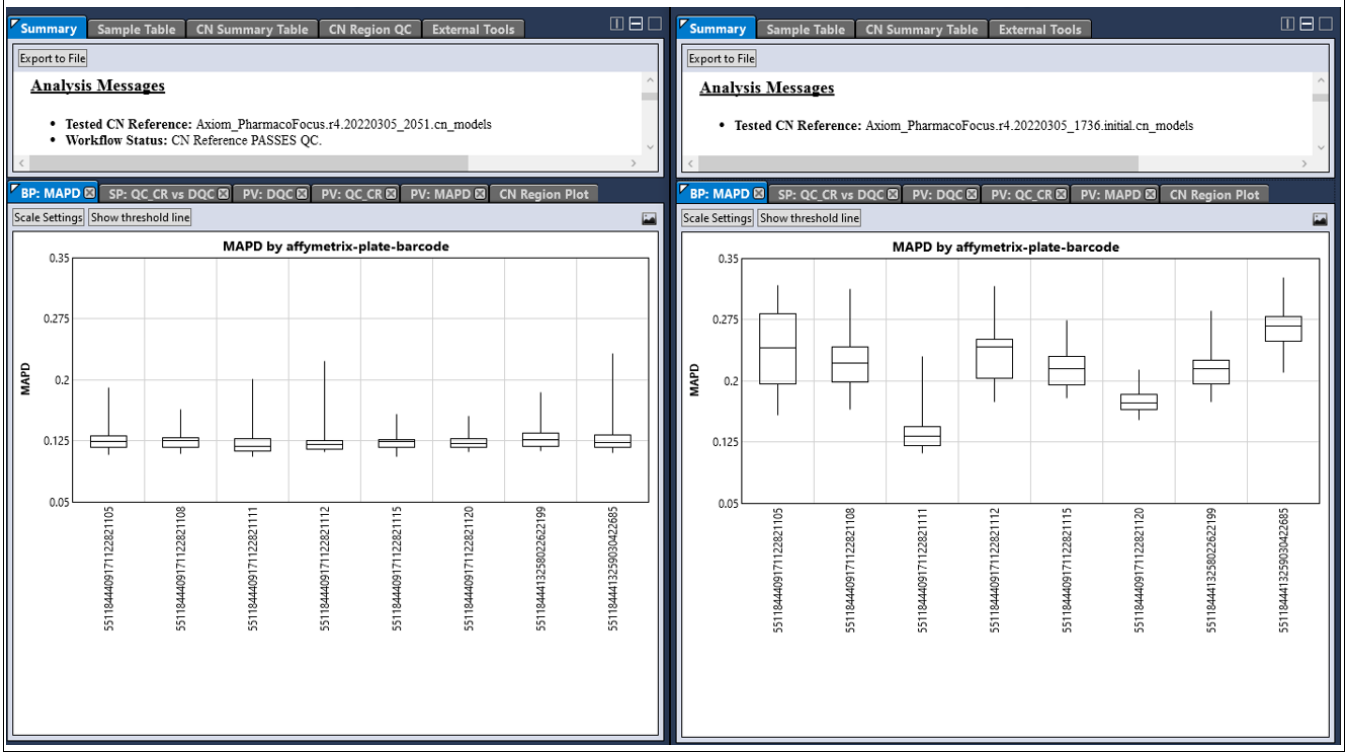


CN Region	Normal CN_State	Minimum normal samples	Available normal samples	Sample availability status
CYP2A6_5pFlank	2	10	540	Pass
CYP2A6_intron2-intron4	2	10	559	Pass
CYP2A6_intron5-exon9	2	10	560	Pass
CYP2D6-3pFlank	2	10	526	Pass
CYP2D6_5pFlank	2	10	405	Pass
CYP2D6_exon9	2	40	515	Pass
GSTM1_gene	2	5	95	Pass
GSTT1_gene	2	5	165	Pass
SULT1A1_gene	2	10	350	Pass
UGT2B17_gene	2	5	218	Pass

5. Scroll down to the CN QC Metrics Summary table. Confirm that a high percentage of samples pass both MAPD and WavinessSD thresholds. A good CN Reference requires that the percentage of samples passing CN QC also meets the CNQC Pass Rate threshold.

   At this point it helps to see how the best practices CN reference tested in step 2 compares to the previous CN reference used in step 1, when tested on the same data. Figure 166 captures compares the MAPD box plot of the current step 2 batch, to the MAPD box plot from a re-opened step 1 batch.

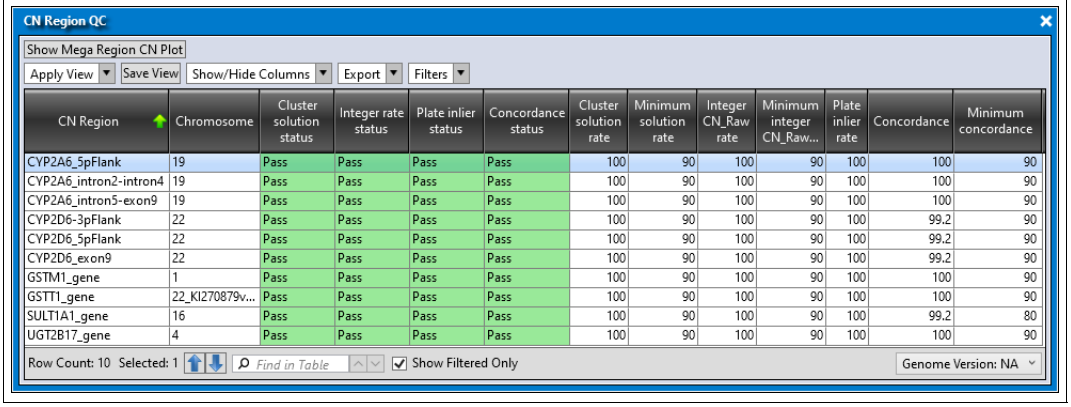
Figure 166 MAPD Box Plot - Step 2 vs. Step 1 Comparison



It is better for MAPD values to be low and more consistent across plates. The Best Practices CN Reference built on these plates (step 2 batch) is a better reference for CN analysis for this data than the previous reference used in the step 1 batch.

6. For the step 2 batch that evaluates the multiplate reference, switch to the CN Region QC tab, and check whether all regions pass all tests. (Figure 167)

Figure 167 CN Region QC tab window

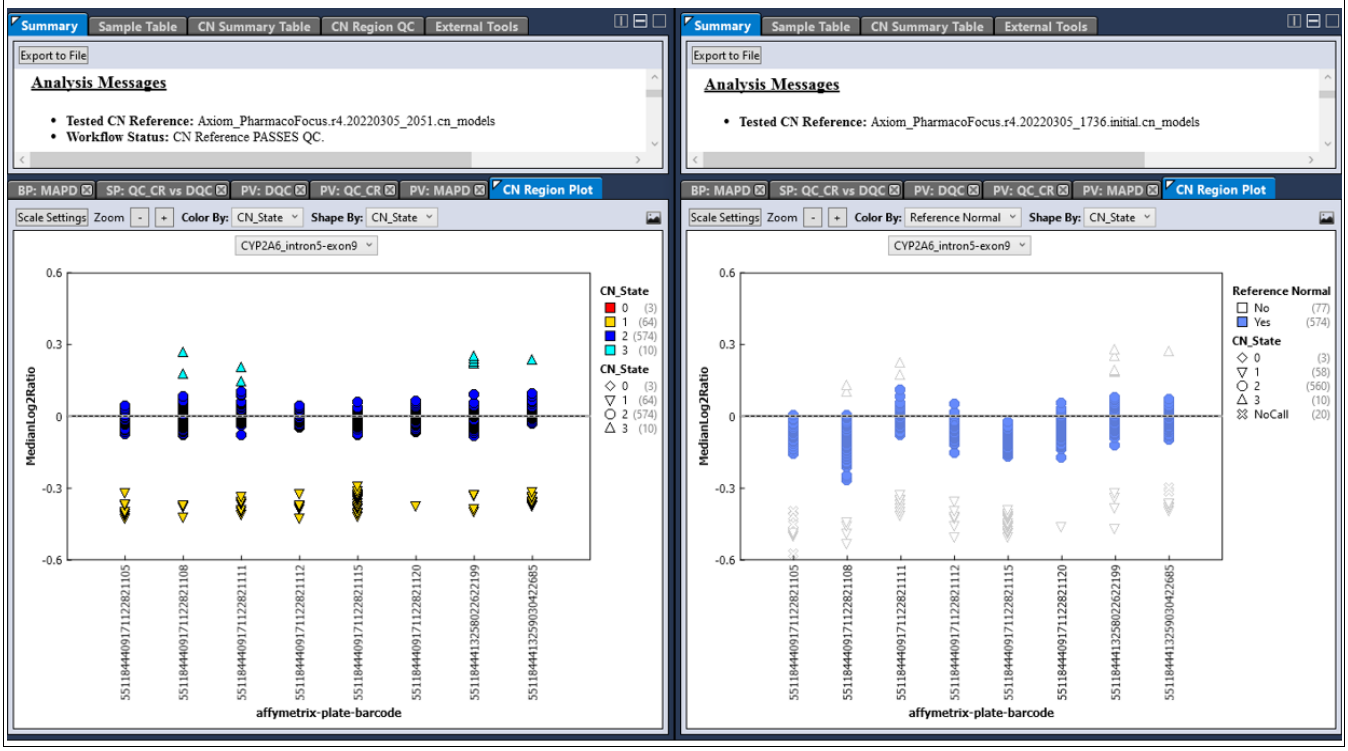


CN Region	Chromosome	Cluster solution status	Integer rate status	Plate inlier status	Concordance status	Cluster solution rate	Minimum solution rate	Integer CN_Raw rate	Minimum integer CN_Raw...	Plate inlier rate	Concordance	Minimum concordance
CYP2A6_5pFlank	19	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2A6_intron2-intron4	19	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2A6_intron5-exon9	19	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
CYP2D6-3pFlank	22	Pass	Pass	Pass	Pass	100	90	100	90	100	99.2	90
CYP2D6_5pFlank	22	Pass	Pass	Pass	Pass	100	90	100	90	100	99.2	90
CYP2D6_exon9	22	Pass	Pass	Pass	Pass	100	90	100	90	100	99.2	90
GSTM1_gene	1	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
GSTT1_gene	22_KI270879v...	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90
SULT1A1_gene	16	Pass	Pass	Pass	Pass	100	90	100	90	100	99.2	80
UGT2B17_gene	4	Pass	Pass	Pass	Pass	100	90	100	90	100	100	90

Note: Concordance columns appear if at least one sample is matched to samples for which the library files have expected CN calls. Known samples include Ref103 and many 1000Genomes project samples.

Review the CN Region Plot. See if all plates for all regions appear to have the CN2 samples near the horizontal dashed line. To see the benefit of the new CN reference vs previous one, it helps to have the CN Region Plot from BOTH the step 2 and previous step 1 batches open side by side, as shown in Figure 168.

Figure 168 CN Region Plot - Step 2 vs Step 1 comparison



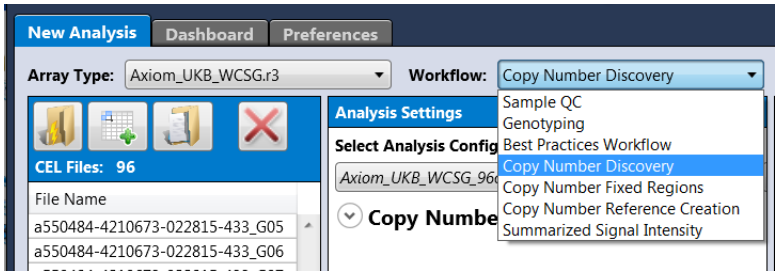
In the Step 2 vs Step1 comparison above, you can see the impact that the new reference has on the results. The right plot shows the blue samples selected for building a new multiplate CN reference. The left plot shows how the multiplate reference is better able to correct for between-plate variability. Having the CN2 samples more centered at MedianLog2Ratio makes correct CN calling easier. Also, the clusters are sometimes smaller, which makes it easier to resolve the different copy number states within each plate.

If the CN Reference PASSES QC and a visual inspection of the CN Region Plot across all regions reveals all plates made CN2 calls near the MedianLog2Ratio=0 line, then this CN Reference is ready for routine use. If the CN Reference File passes QC, it is OK to use it for other CN analyses.

Copy number discovery workflow

1. From the New Analysis window tab, click the **Array Type** drop-down, then select the array type you want to use.
2. Click on the Workflow drop-down, then select **Copy Number Discovery**. (Figure 169)

Figure 169 Copy Number Discovery option

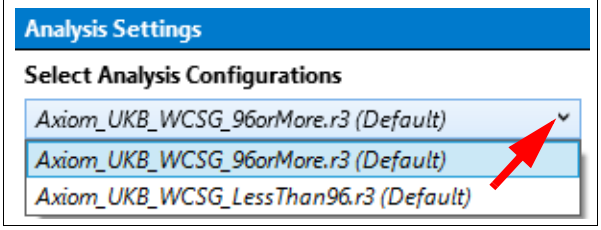


3. Click either **Import CEL Files**, **Import Samples and Attributes by Txt**, or **Import Samples and Attributes by ARR** icon.

   The appropriately named **Add Files** window appears.
4. Navigate to your files, as you normally would, then click **Open**.

   The CEL Files pane populates and displays your selected files.
5. (Optional) Click the **Select Analysis Configurations** drop-down (Figure 170) to choose a different configuration.

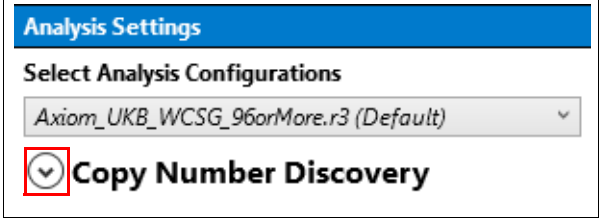
Figure 170 Select Analysis Configurations drop-down



IMPORTANT! The Analysis Configuration in the Analysis Settings pane is dynamically linked to the Workflow selected. If the analysis configuration is changed, the workflow resets to the last used for that analysis configuration.

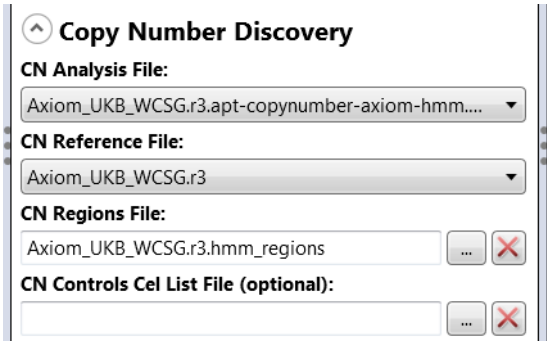
6. In the Analysis Settings pane, click on the **Copy Number Discovery** drop-down arrow to expand it. (Figure 171)

Figure 171 Copy Number Discovery drop-down



7. (Optional) The **CN Reference File** defaults to the file supplied in the library package, as shown in Figure 172. To select a different reference file, click the down arrow to expand the window, then click to select reference file you want.

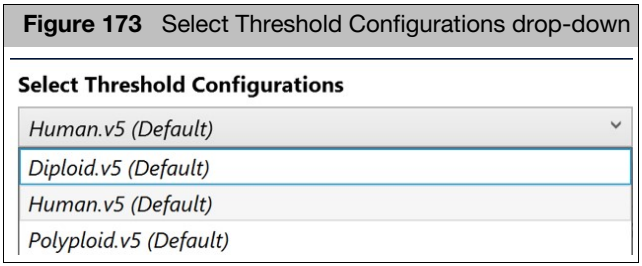
Figure 172 CN Reference File drop-down



8. In the Threshold Settings pane, click the Select Threshold Configuration drop-down, then click to select the appropriate threshold, as shown in Figure 173.

   Note: Newer array packages will have their own ax threshold file.

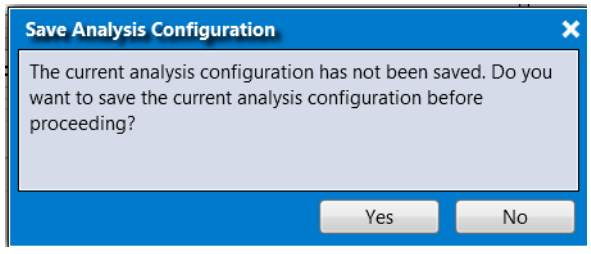
Figure 173 Select Threshold Configurations drop-down



9. Enter a name for the Copy Number Discovery Batch in the **Name** field, then click **Run**.

A Save Analysis Configuration window appears. (Figure 174)

Figure 174 Save message



10. If this is configuration will be used for future analyses, click **Yes**, then enter a save name. If not, click **No**.

    The Dashboard window tab appears and displays the progress. After the analysis successfully completes, click the **Open** button to view the results in the AxAS Viewer.

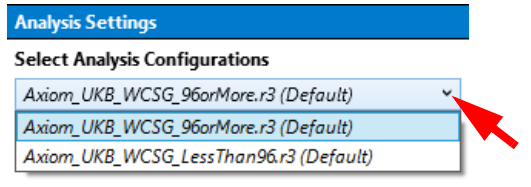
Copy number fixed regions workflow

1. From the New Analysis window tab, click the **Array Type** drop-down, then select the array type you want to use.
2. Click on the Workflow drop-down, then select **Copy Number Fixed Regions**.
3. Click the **Import CEL Files**, **Import Samples and Attributes by Txt**, or **Import Samples and Attributes by ARR** icon.

   The **Add Files** window appears.
4. Navigate to your files, as you normally would, then click **Open**.

   The CEL Files pane populates and displays your selected files.
5. (Optional) Click the **Select Analysis Configurations** drop-down (Figure 175) to choose a different configuration.

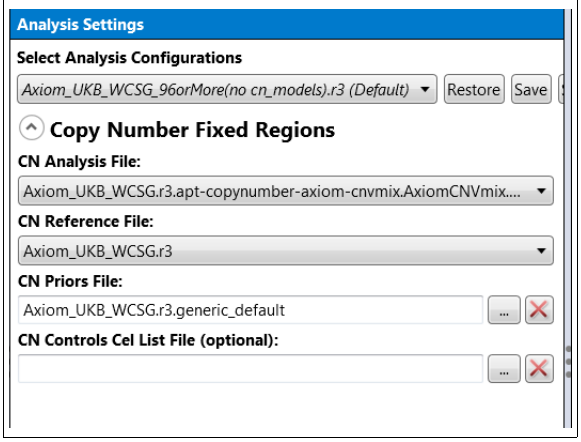
Figure 175 Select Analysis Configurations drop-down



IMPORTANT! The Analysis Configuration in the Analysis Settings pane is dynamically linked to the Workflow selected. If the analysis configuration is changed, the workflow resets to the last used for that analysis configuration.

6. In the Analysis Settings pane, click on the **Copy Number Fixed Regions** drop-down arrow to expand it, as shown in Figure 176.

Figure 176 Copy Number Fixed Regions drop-down

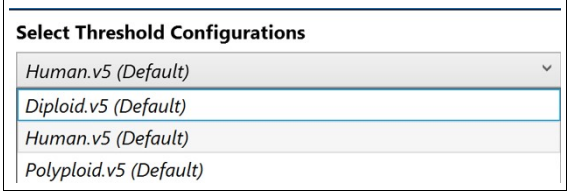


7. (Optional) The **CN Reference File** defaults to the file supplied in the library package, as shown in Figure 176. To select a different reference file, click the down arrow to expand the window, then click to select reference file you want.

   Note: A CN Reference file must be selected. If using a CN Controls CEL List File, "Threshold configurations with CN options" on page 190 for plate control information.

8. In the Threshold Settings pane, click the Select Threshold Configuration drop-down, then click to select the appropriate threshold, as shown in Figure 177.

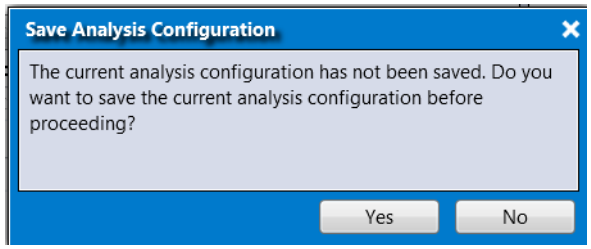
Figure 177 Select Threshold Configurations drop-down



9. Enter a name for your CN Fixed Region Batch in the **Name** field, then click **Run**.

A Save Analysis Configuration window appears. (Figure 178)

Figure 178 Save message



10. If this configuration will be used for future analyses, click **Yes**, then enter a save name. If not, click **No**.

    The Dashboard window tab appears and displays the progress. After the analysis successfully completes, click the **Open** button to view the results in the AxAS Viewer.

Visualization for CN batches

- CN Discovery analysis batches have different content in the Viewer than CN Fixed Region batches.
- The Summary contains CN specific thresholds and metrics. A MAPD plate view plot is the default graph in CN Result batches. CN Segment and LOH Segment tables can be also be generated from the Sample Table.
- For CN Discovery results, in addition to standard sample, and plate information, the Sample Table contains MAPD, Waviness as well as count of CN 0, 1, 2 and 3 Segments and Autosome %LOH. CN Discovery batches also include PCA Plot generation for selected or all samples.
- Whole Genome View plots can be created from Discovery batches. Whole Genome View allow users to view frequency of copy number events in the batch as well as individual sample's copy number segments, LOH, Log2Ratio and smooth signal tracks.
- Results from CN Fixed Regions batches include a CN Summary Table and a CN Region Plot.
- The Sample Table for CN Fixed Region batches includes CN specific QC metrics.
- Export to IGV, Nexus and VCF format for further visualization is also available.

Viewing CN discovery batches

1. From the Dashboard window tab, click on the appropriate batch's **Open** button.

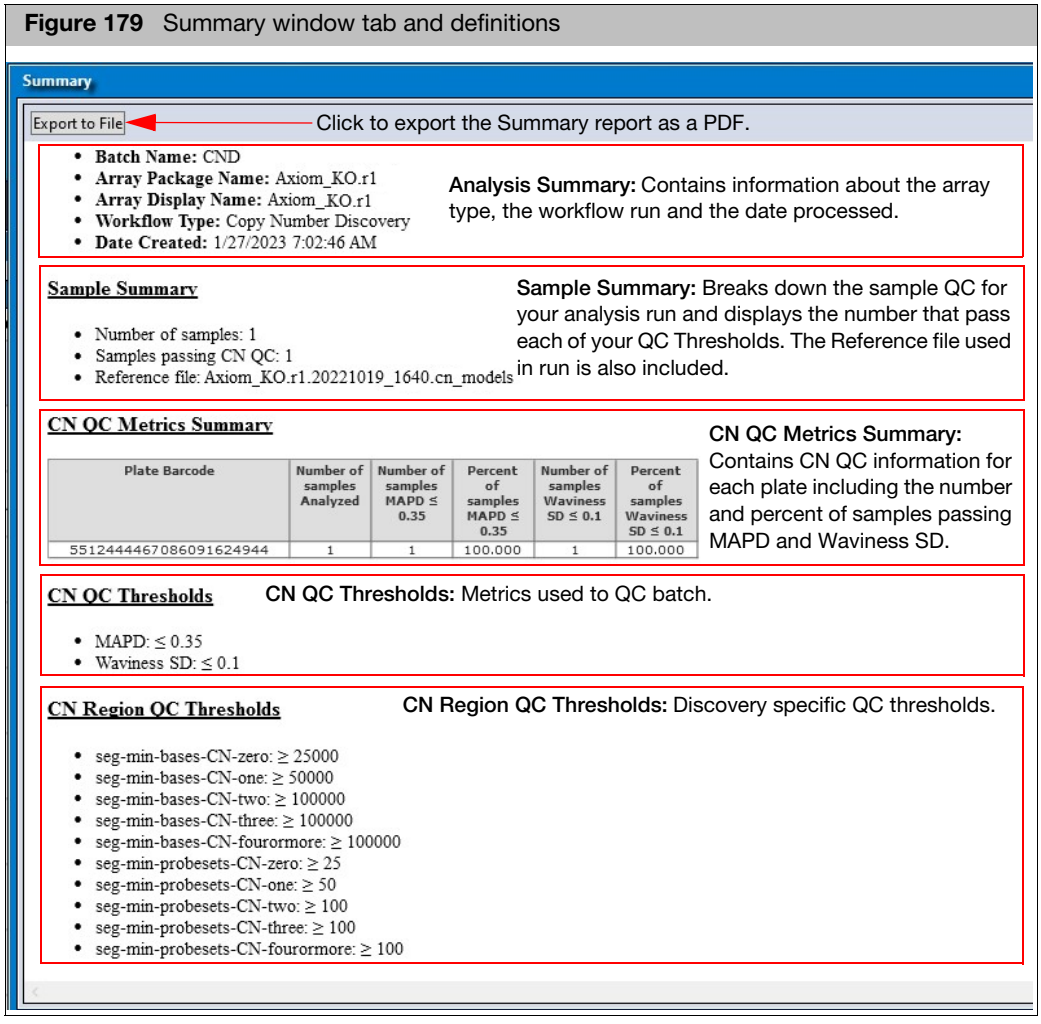
   Three default window tabs appear:

   - Summary
   - Sample Table
   - Plate View (PV): MAPD

   Note: By default, the Viewer auto-generates a MAPD Plate View plot for CN analyzed batches.

Summary window

Figure 179 Summary window tab and definitions



Sample table

Figure 180 Sample Table window tab

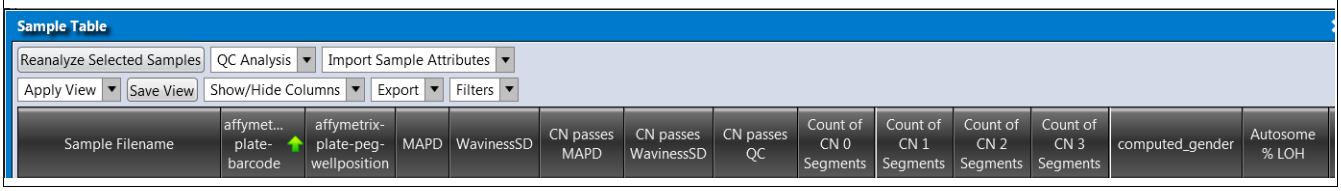
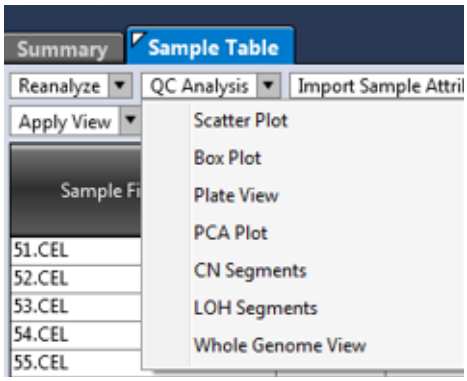


Table 5 Default CN Discovery Table

Column	Description
Sample Filename	CEL file name.
MAPD	A global measure of the variation of all microarray probes across the genome. It represents the median of the distribution of changes in Log2 Ratio between adjacent probes. Since it measures differences between adjacent probes, it is a measure of shortrange noise in the microarray data.
Waviness SD	A global measure of variation of microarray probes that is insensitive to shortrange variation and focuses on long-range variation.
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.
Count of CN 0 Segments	Number of segments that are CN zero.
Count of CN 1 Segments	Number of segments that are CN 1.
Count of CN 2 Segments	Number of segments that are CN 2.
Count of CN 3 Segments	Number of segments that are CN 3 or more.
Autosome % LOH	Percentage of the autosomal regions of the genome reporting loss of heterozygosity.

(Optional) For more CN Discovery Batch options, click the QC Analysis drop-down menu, as shown in Figure 181.

Figure 181 CN Discovery Batches; Additional options



PCA tables and plot

AxAS enables you to perform Principle Component Analysis (PCA) on CN Discovery data, which will visually cluster samples by similarity of CN segment calls.

PCA constructs a new set of variables called loadings from combinations of the original variables. The first three of the new loadings (PCA1) captures the largest variability in the data from all of the loadings. The second loading (PCA2) captures the second largest amount of variability in the data, and the third loading (PCA3) captures the third largest amount of variability.

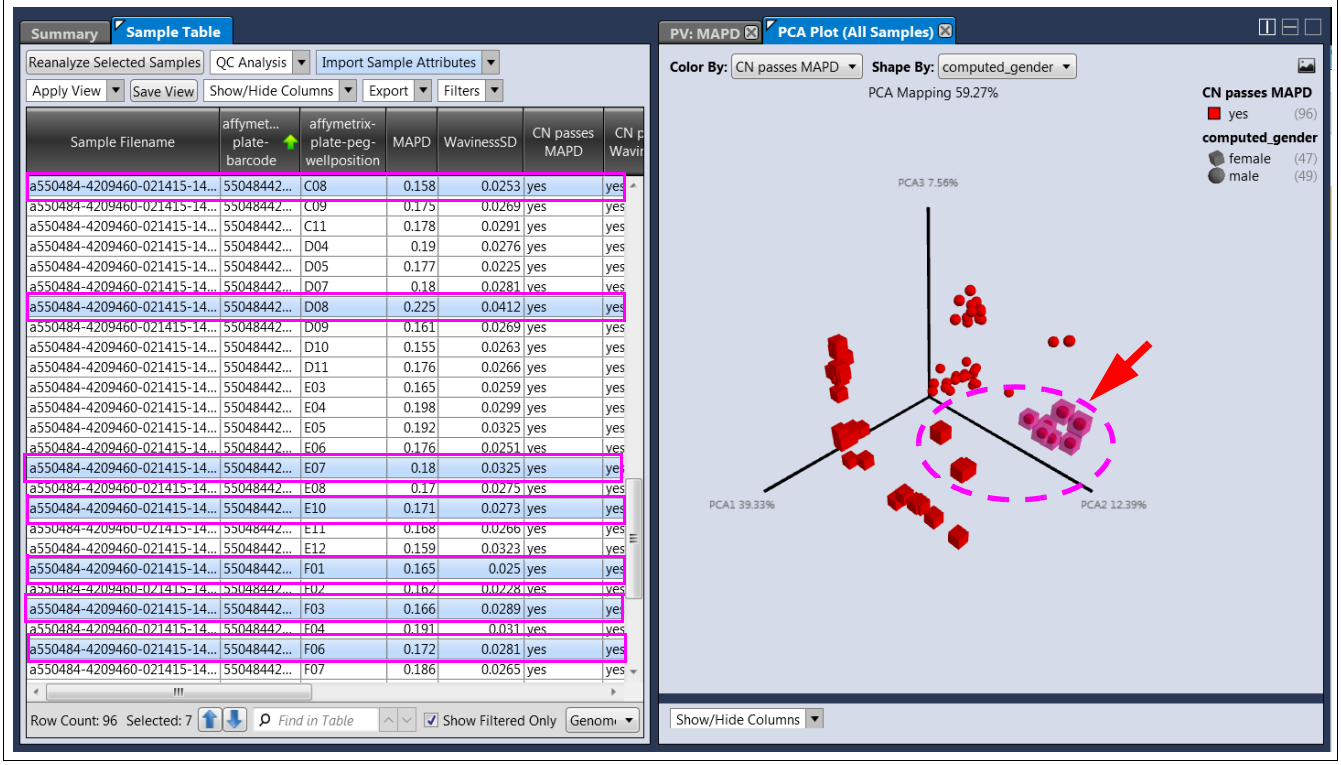
1. Click to highlight the samples in Sample Table that you would like to run PCA.
2. From the QC Analysis drop-down menu, select **PCA Plot**.

   A PCA window appears displaying two running options.
3. Select one by clicking on the appropriate radio button, then click **OK**.

   A new PCA Plot window tab and selected samples table appears. (Figure 182)

   The PCA graph's axes represent the top three loadings (PCA1, PCA2, and PCA3) that account for the part of the variability among the samples.

Figure 182 PCA tables and plot



- To expose the subset of samples used, use your mouse to drag the bottom border of the PCA Plot, as shown in Figure 182.

  Note: By default, the PCA Plot is generated with all samples highlighted. To remove the highlights from the graph and table(s), right-click on the graph and select **Clear Highlight**.

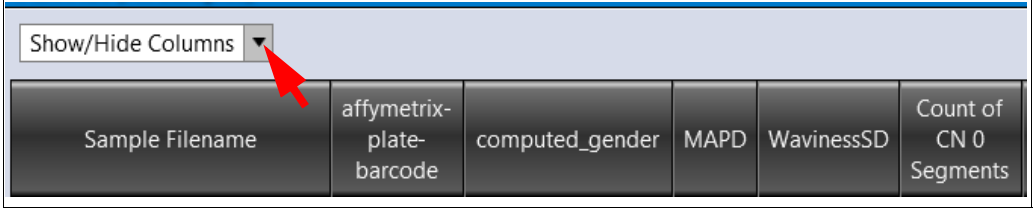
- To lasso a group of samples of interest, hold the left mouse down while drawing a circle, as shown in Figure 182. This action also highlights the corresponding files in the table(s).

- To hide a selected set of samples or all samples, right-click.
  - Click to select one of the following:
  - **Hide Selected:** Hides the selected samples that your mouse is hovering on.
  - **Hide All Selected:** Hides all highlighted samples.

  Note: After samples are hidden from PCA Plot, they cannot be unhidden. You must generate a new PCA Plot to reveal the previously hidden samples.
  - To display **Color By** and/or **Shape By**, click the appropriate drop-down menu.
  - To rotate the graph, right-click, then drag the graph to change its view perspective.
  - Click the  icon (upper right) to export the plot as a PNG file.

PCA sample table

Figure 183 Default sample table columns



- Click the **Show/Hide Columns** drop-down menu (Figure 183) to remove or add columns.

CN discovery segments table

Generating a CN Discovery Segments Table

1. At the Sample Table, highlight your samples of interest.
2. Click the QC Analysis drop-down, then select **CN Segments**.

   A CN Discovery Segments window tab appears. (Figure 184)

Figure 184 CN Discovery Segments Table window tab

CN Discovery Segments Table

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

Sample Filename	Chromosome	Region	Start Position	Stop Position	Size	Marker Count	State	MedianLog2Ratio
51.CEL	1	1	101148487	101358060	209574	92	3	0.158
51.CEL	5	5	40255573	40419789	164217	92	3	0.161
51.CEL	Z	Z	62285490	62548985	263496	69	3	0.145
52.CEL	13	13	7175634	7252935	77302	67	3	0.19
53.CEL	4	4	36156550	36269573	113024	54	3	0.195
53.CEL	4	4	78802442	78969955	167514	71	3	0.173
54.CEL	1	1	113961340	114280951	319612	162	3	0.176
55.CEL	3	3	62260209	62408766	148558	80	3	0.181
55.CEL	Z	Z	62222376	62411253	188878	57	3	0.199
56.CEL	3	3	5971134	6354538	383405	205	3	0.171
56.CEL	8	8	8943	275503	266561	89	3	0.155
56.CEL	Z	Z	62285490	62548985	263496	69	3	0.158
57.CEL	4	4	62104448	62217446	112999	60	3	0.186
58.CEL	1	1	113954625	114282864	328240	168	3	0.177
58.CEL	2	2	81319557	81439515	119959	57	3	0.189
58.CEL	3	3	6694011	6839706	145696	64	3	0.183
58.CEL	3	3	36088184	36183089	94906	54	3	0.206
58.CEL	4	4	36155216	36257756	102541	51	3	0.21
58.CEL	4	4	62111427	62217446	106020	57	3	0.161
58.CEL	8	8	8943	413758	404816	129	3	0.151
58.CEL	11	11	5725605	5857407	131803	83	3	0.156
58.CEL	Z	Z	18472649	18741003	268355	85	3	0.17
58.CEL	Z	Z	20574483	20716587	142105	55	3	0.21
58.CEL	Z	Z	37627724	37894245	266522	95	3	0.162
58.CEL	Z	Z	62285490	62548985	263496	69	3	0.216
59.CEL	8	8	8943	619520	610578	208	3	0.166
59.CEL	Z	Z	62284803	62548985	264183	70	3	0.211

Note: Copy number neutral segments are filtered out of the CN Discovery Segments Table.

Table 6 Default CN Discovery Segments Table

Column	Description
Sample Filename	CEL file name.
Chromosome	The chromosome on which the CN event is located.
Region	Specified region within chromosome, as defined in the library file package, on which copy number analysis is performed.
Start Position	The nucleotide base start position of the segment. The genomic coordinates given are in relation to the current genome version and may shift as subsequent genome builds are released.
Stop Position	The nucleotide base stop position of the segment. The genomic coordinates given are in relation to the current genome version and may shift as subsequent genome builds are released.
Size	Length of segment in base pairs.
Marker Count	Number of ProbeSets that make up segment.
medianlog2ratio	The median Log2Ratio of CN ProbeSets within that segment.
State	CN state of segment.

LOH segments table

Loss of Heterozygosity (LOH) is also determined during CN analysis. For more information see *Affymetrix® White Paper: The Loss of Heterozygosity Algorithm in Genotyping Console 2.0*.

Generating a LOH Segments Table

1. At the Sample Table, highlight your samples of interest.
2. Click the QC Analysis drop-down, then select **LOH Segments**.

   A LOH Segments Table window tab appears. (Figure 185)

Figure 185 LOH Segments Table window

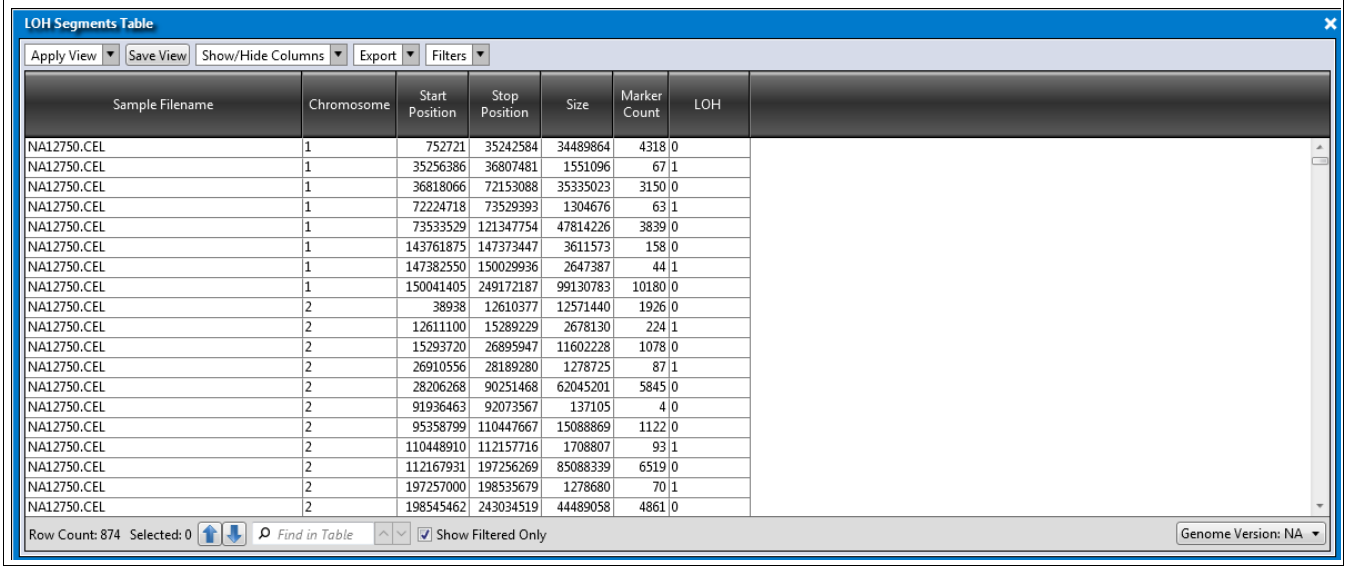


Table 7 Default LOH Segments Table

Column	Description
Sample Filename	CEL file name.
Chromosome	Chromosome number where LOH is located.
Start Position	The nucleotide base start position of the LOH segment.
Stop Position	The nucleotide stop position of the LOH segment.
Size	Length of LOH in base pairs.
Marker Count	Number of ProbeSets that make up segment.
LOH	A flag of 1 indicates the presence of LOH.

IMPORTANT! The default view of the LOH Segments table does not have any filters on Size or Marker Count. As part of reviewing your LOH data, create a Filter on Marker Count >25. To view only LOH segments, click on the Table's Filter drop-down menu and select contains 1.

Whole Genome View (WGV)

Overview

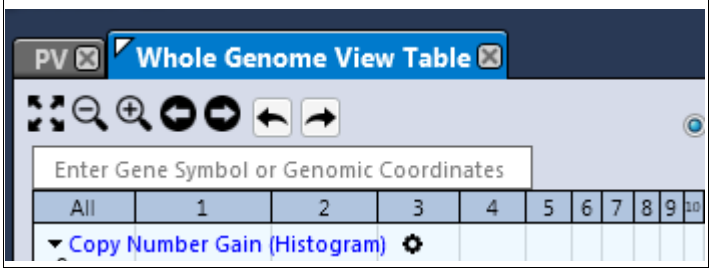
- A genome-wide plot can be created from a Discovery batch, allowing viewing of 400 samples simultaneously.
- A WGV requires two annotation files; annotation database file that was used to analyze samples and a NetAffxGenomic Annotation file. The annotation database file is part of the array library file package and should be located within the array library package folder. NetAffxGenomic Annotation files are used by the WGV to display recent snapshots of genomic annotations downloaded from public databases. The NetAffxGenomic Annotation file should be located in the library folder data root.
- Frequency plots of batch copy number gains, losses and LOH are generated.
- Individual sample data plots of CN segment, including CN state, log2 ratio, smooth signal, BAF and LOH are also generated.
- Users have the ability to drill down to per chromosome segment as well as highlight selected regions of interest.
- WGV and CN Segment table are synchronized to allow users to see in tabular form start, stop and size of copy number events.

IMPORTANT! Samples must pass CN QC to be visualized via WGV. It is recommended to filter passing samples in the CN passes QC column of the Sample Table prior to generating WGV plot.

Tool bar and viewing features

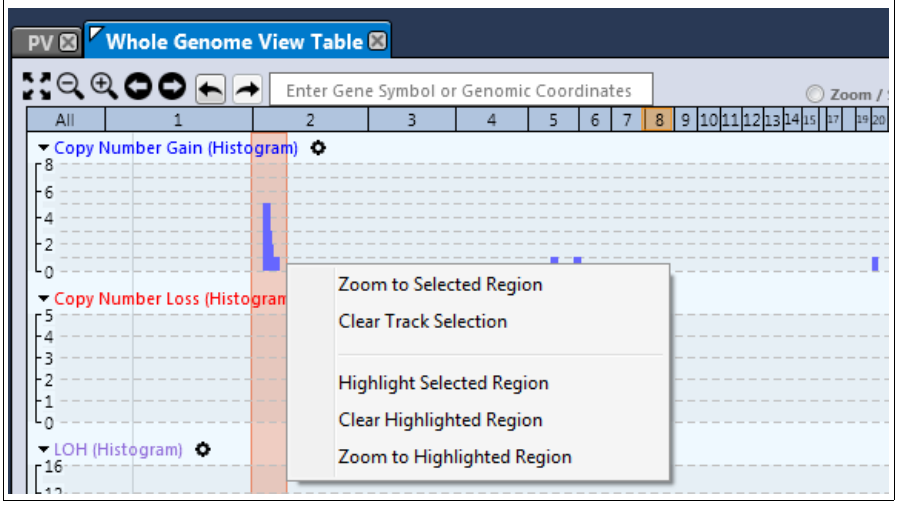
The WGV window tab features a tool bar. (Figure 186)

Figure 186 WGV tool bar



- Click  to display the entire whole genome view of the Whole Genome View tab.
- Click  to zoom in or zoom out.
- Click on a Chromosome number to zoom in on that chromosome.
- Right-click and drag to draw a rectangle in the Whole Genome View, then release the mouse button to zoom into the area within the rectangle.
- To return to the previous view, click the undo  button. Click the redo  button to go back to your last view.
- Click  to move the current view left or right.
- Click the  radio button (far right) to enable a left/right pan action of the current view. **Note**: While in Panning mode, you must right-click to draw a rectangle, then select **Zoom to Selected Region** (Figure 187) to zoom in on a region.

Figure 187 Panning mode and zooming in on a region



- Click inside the  text box, then enter the gene name, accession number, or genome coordinates you want to locate.

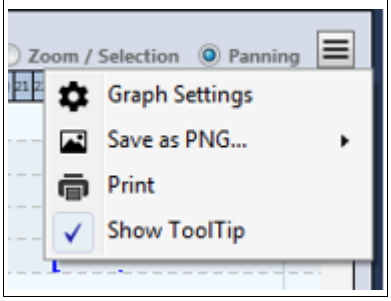
- Click to select the appropriate finding from the list, then press **Enter** to display the selection. To return to the previous view, click the undo button. Click the redo to go back to your last view.

WGV settings and options

1. Click (upper right corner).

   A menu appears. (Figure 188)

Figure 188 WGV Settings menu

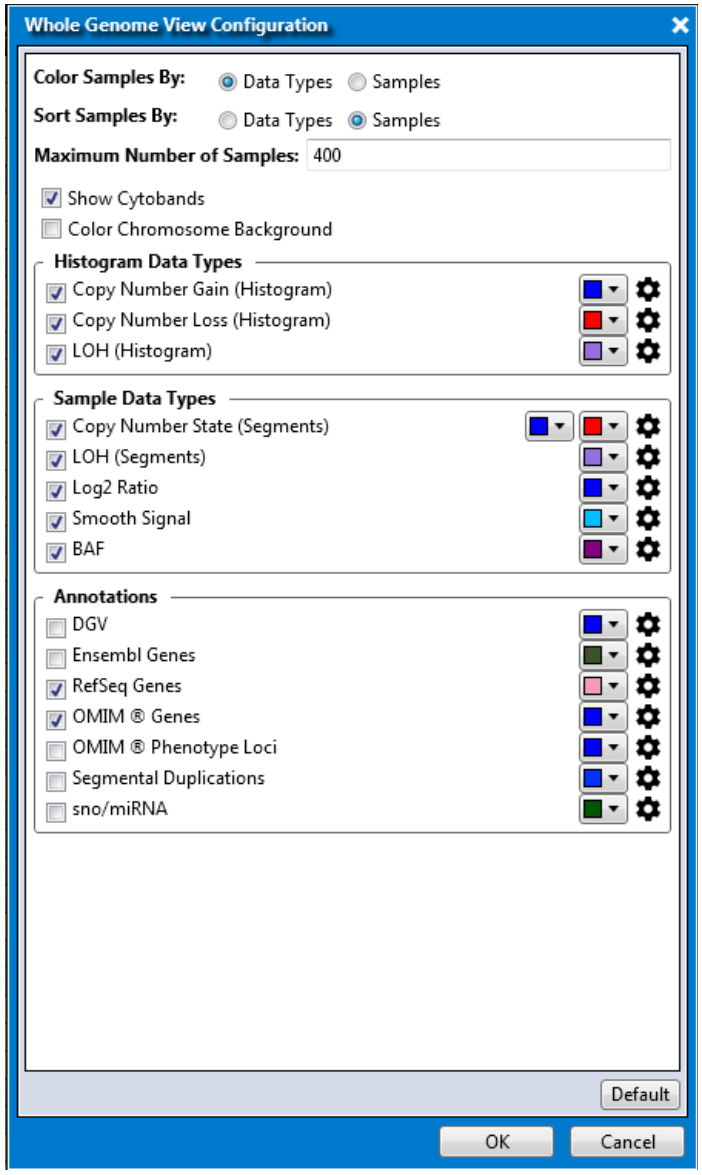


Graph settings

1. Click **Graph Settings**.

   A Whole Genome View Configuration window appears. (Figure 189)

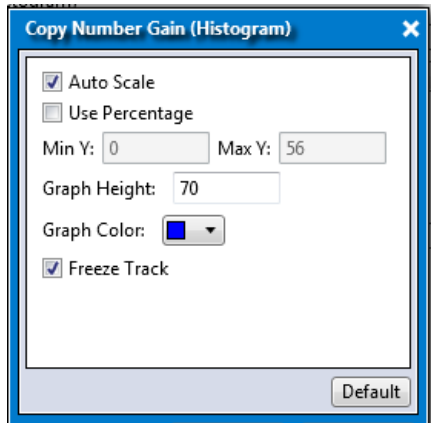
Figure 189 Whole Genome View Configuration window



2. To personalize your Whole Genome View, click the appropriate radio buttons, check boxes, and color drop-downs.
3. Click **OK** to save your configurations or click **Default** to return all settings back to their factory defaults.
4. (Optional) For additional options, click the appropriate ⚙.

   A settings window appears. (Figure 190)

Figure 190 Copy Number Gain (Histogram) window



- By default, the **Auto Scale** check box is checked and each axis is given a value (based on the data). Uncheck the Auto Scale check box to activate both the Min and Max value fields, then enter the new values.
- Click the **Use Percentage** check box to see percent of samples with copy number events instead of number.
- Click inside the Graph Height value field to enter a new value.
- Click the Graph Color drop-down to assign a different color.
- The Freeze Track check box is checked by default and displays the track in the upper WGV window. Uncheck this box to turn off the track. This option is checked by default because it can be useful for annotation tracks. Frozen/Static tracks can make for easier viewing while scrolling through your probe level data. Uncheck this box to turn off this feature off.
- Click **OK** to save your configurations or click **Default** to return all settings back to their factory defaults.

Saving a WGV

1. Click ☰ (upper right corner), then click **Save as PNG**.
2. Navigate to a save location, the enter a filename.
3. Click **OK**.

Printing a WGV

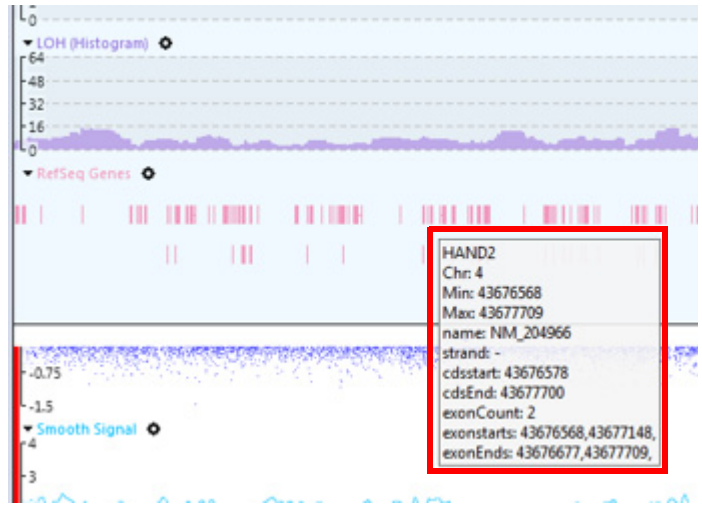
1. Click ☰ (upper right corner), then click **Print**.
2. Select a pre-configured printer as you normally would, then click **Print**.

   The Frequency Histogram and Sample Data Plots are printed.

Showing ToolTips

The **Show ToolTip** is checked by default. It enables you to hover over annotations to see gene information in the Frequency Histogram plot and view specific CN state information in the Sample Data plots, as shown in Figure 191.

Figure 191 WGV ToolTip example

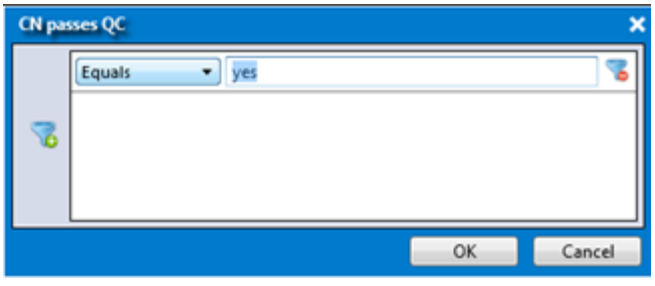


Setting up a WGV

1. At the Sample Table, right-click on the **CNpassesQC** header, then select **Filter**.

   The CNpassesQC window appears. (Figure 192)

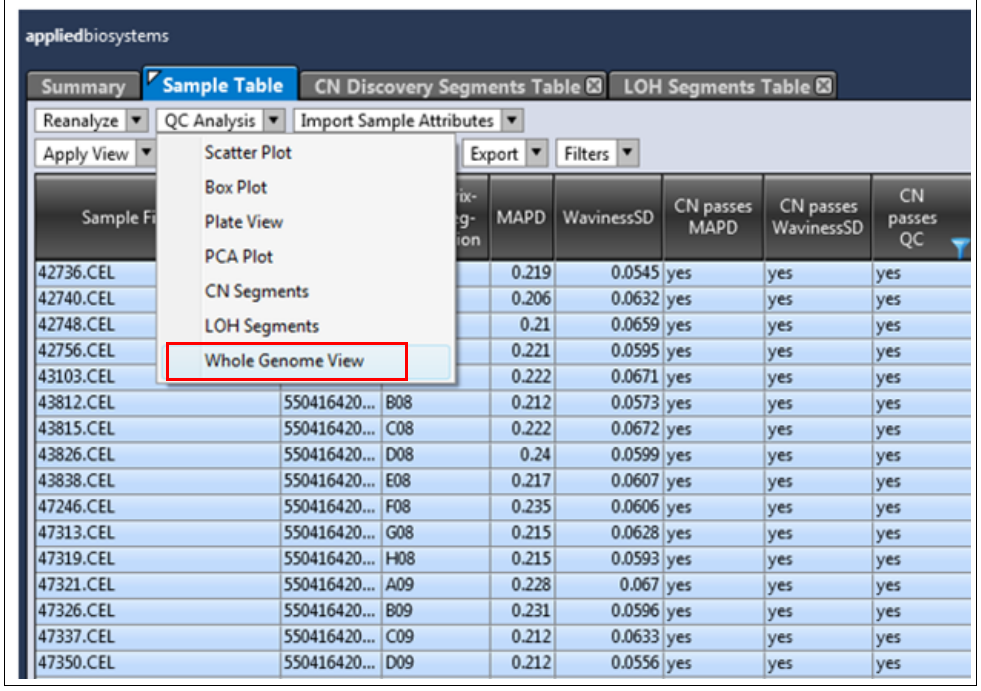
Figure 192 CNpassesQC window



2. Click the drop-down to select **Equals**.
3. In text field, type **yes**, then click **OK**.

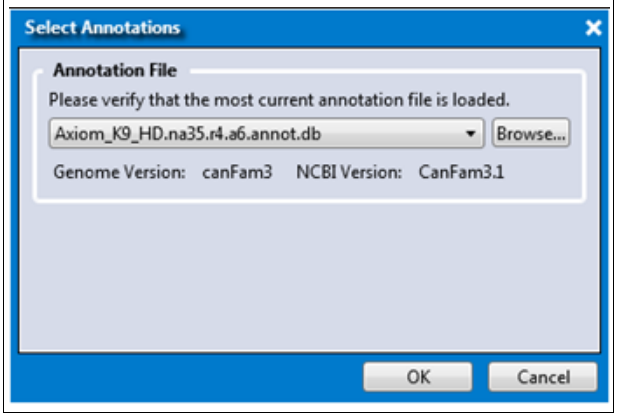
   Only samples that have passed QC now appear in the Sample Table.
4. At the Sample Table, click or Ctrl click on the samples you want to view in WGV, then click the QC Analysis drop-down and select **Whole Genome View**, as shown in Figure 193.

Figure 193 Sample Table QC Analysis drop-down



5. Verify the annotation that was used in your analysis is displayed, then click **OK**.

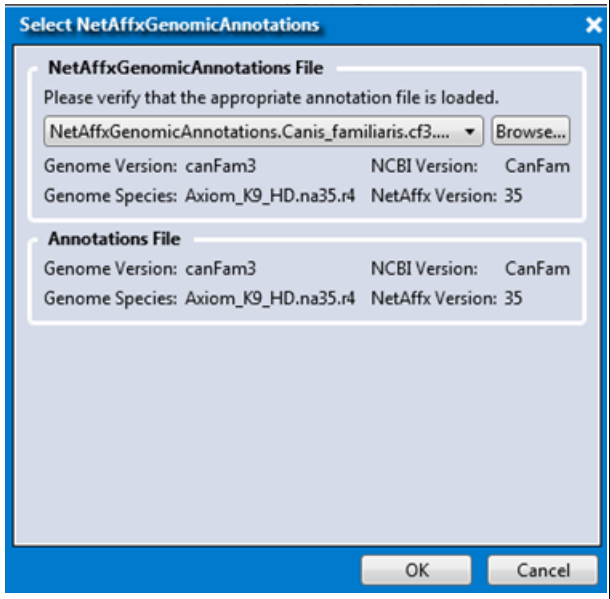
Figure 194 Select Annotations window



6. If the annotation file is not correct, click the **Browse** button, then navigate to the folder containing the correct annotation.
7. Click **Open**.
8. At the Select Annotation window, click **OK**.

   A Select NetAffxGenomicAnnotations window appears. (Figure 195)

Figure 195 Select Annotations window

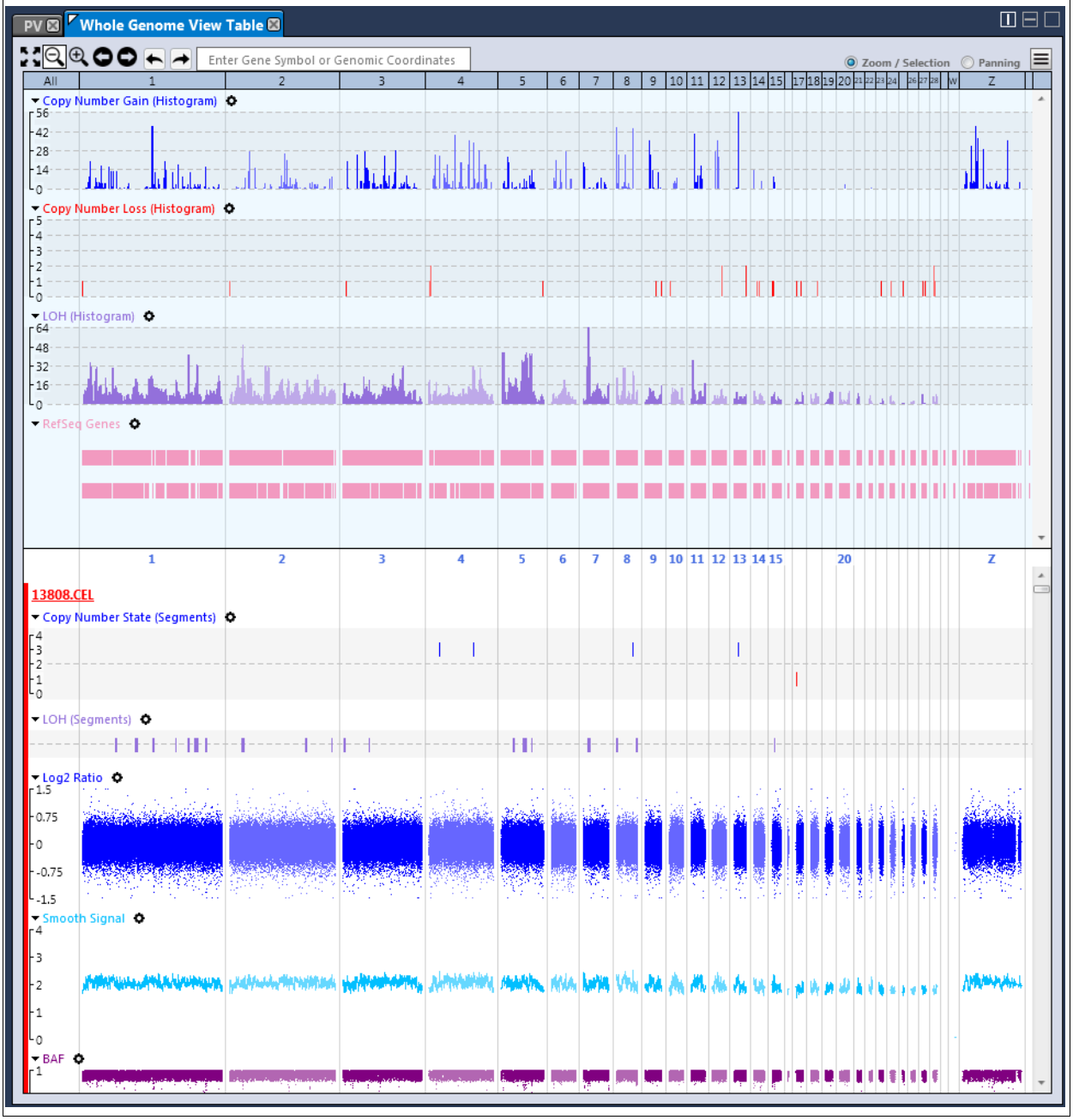


Note: A red warning message appears if the NetAffx versions are not the same. AxAS will error out if Genome species versions are different from the annotation files.

9. Confirm Genome Versions are the same between the annotation files, then click **OK**.

   A WGV table is created and is available as an addition window tab, as shown in Figure 196.

Figure 196 Whole Genome View Table window tab



The WGV window tab (Figure 196) is divided into two viewing sections:

- (Upper pane) Frequency Histograms and Annotations
- (Lower pane) Sample Data Types

Frequency Histogram/ Annotation (upper) pane

- Top banner is defaulted on 'All' chromosomes, with breakdown of CN aberrations on each chromosome.
- CN gains, CN losses, LOH, Ensembl Genes and RefSeq (OMIM if human)
- The upper portion of the WGV tab offers frequency histograms for Segment Data (Copy Number and LOH). These graphs are useful for visually identifying common regions of copy number gain/loss and/or LOH across a set of samples.
  - For gains/losses: Y axis is number of samples with changes, X axis is segments across genome
- Click on a Chromosome number to zoom in on that chromosome.
- Left-click, then drag to draw a rectangle in the Frequency histogram plot (Figure 197). The samples contained in your highlighted rectangle are now highlighted in the Sample Table. (Figure 198)

Figure 197 Highlighting a region

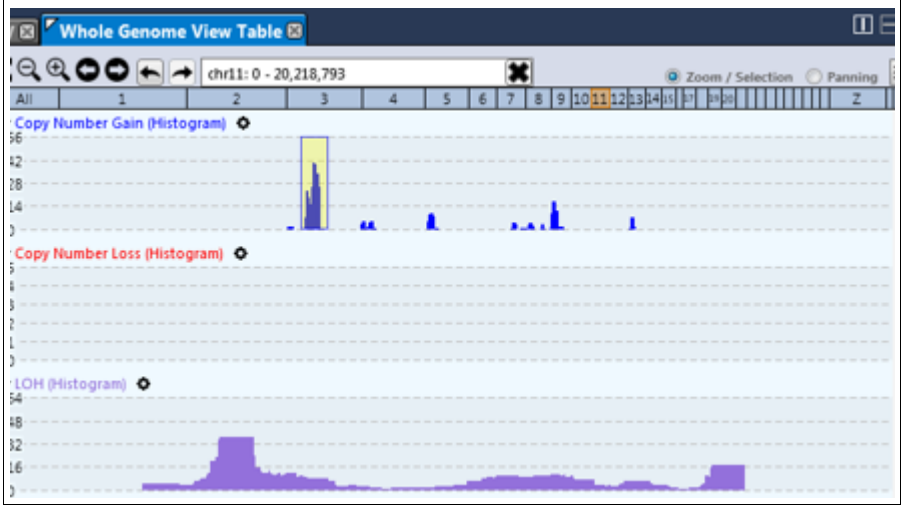
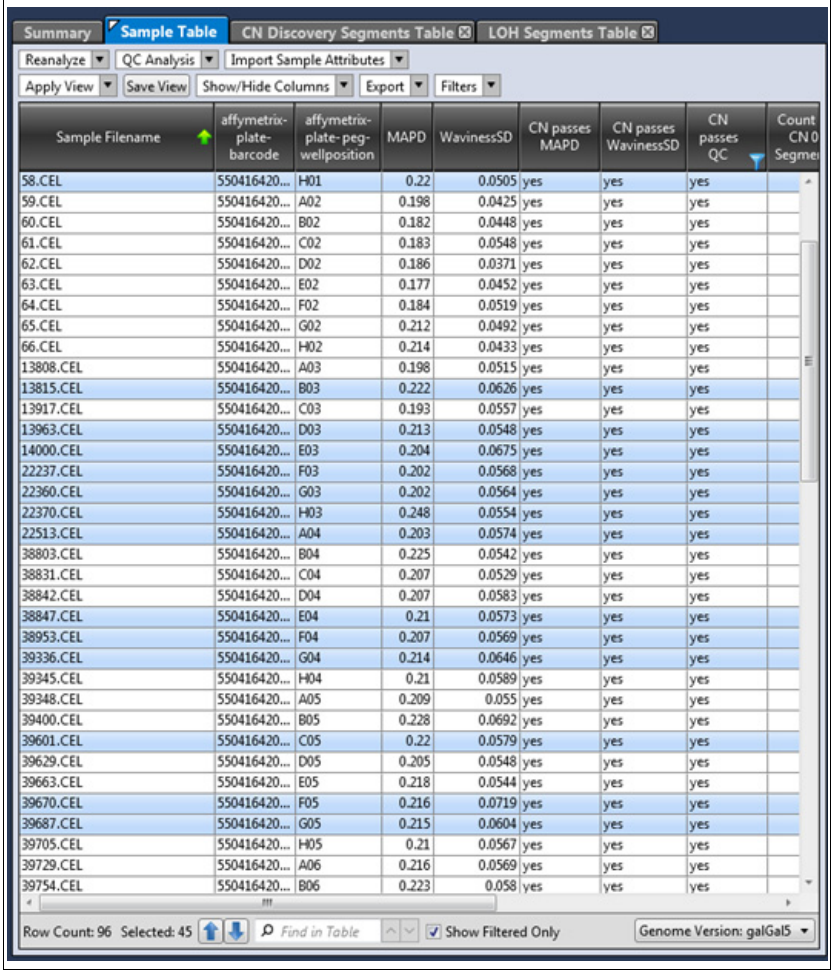


Figure 198 Samples with the selected or chosen copy number event are highlighted in the Sample Table

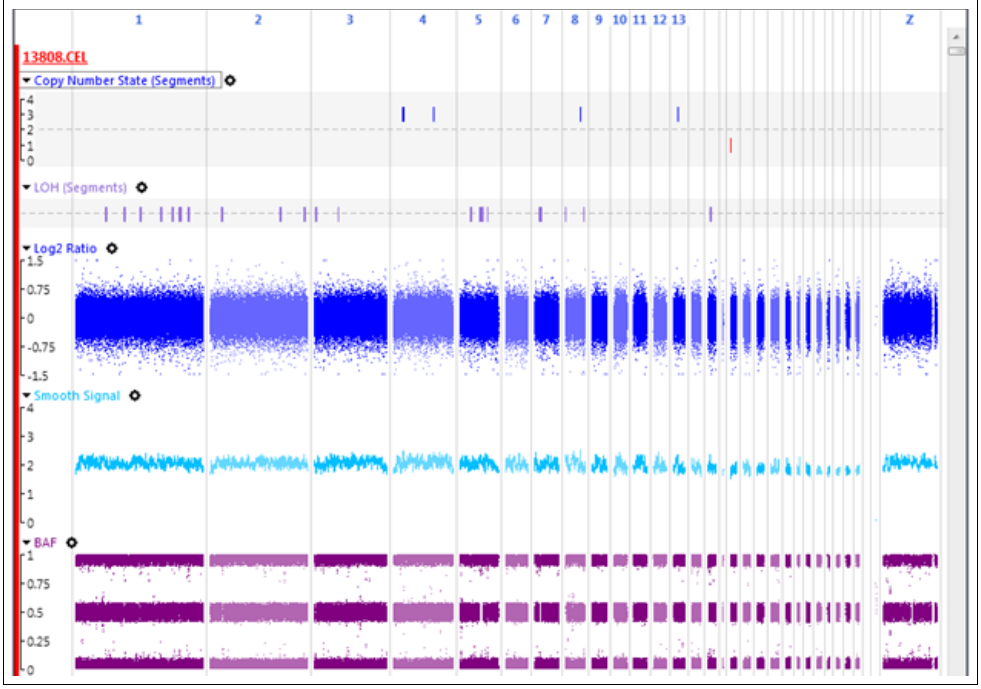


Sample Filename	affymetrix-plate-barcode	affymetrix-plate-peg-wellposition	MAPD	WavinessSD	CN passes MAPD	CN passes WavinessSD	CN passes QC	Count CN 0 Segme
58.CEL	550416420...	H01	0.22	0.0505	yes	yes	yes	
59.CEL	550416420...	A02	0.198	0.0425	yes	yes	yes	
60.CEL	550416420...	B02	0.182	0.0448	yes	yes	yes	
61.CEL	550416420...	C02	0.183	0.0548	yes	yes	yes	
62.CEL	550416420...	D02	0.186	0.0371	yes	yes	yes	
63.CEL	550416420...	E02	0.177	0.0452	yes	yes	yes	
64.CEL	550416420...	F02	0.184	0.0519	yes	yes	yes	
65.CEL	550416420...	G02	0.212	0.0492	yes	yes	yes	
66.CEL	550416420...	H02	0.214	0.0433	yes	yes	yes	
13808.CEL	550416420...	A03	0.198	0.0515	yes	yes	yes	
13815.CEL	550416420...	B03	0.222	0.0626	yes	yes	yes	
13917.CEL	550416420...	C03	0.193	0.0557	yes	yes	yes	
13963.CEL	550416420...	D03	0.213	0.0548	yes	yes	yes	
14000.CEL	550416420...	E03	0.204	0.0675	yes	yes	yes	
22237.CEL	550416420...	F03	0.202	0.0568	yes	yes	yes	
22360.CEL	550416420...	G03	0.202	0.0564	yes	yes	yes	
22370.CEL	550416420...	H03	0.248	0.0554	yes	yes	yes	
22513.CEL	550416420...	A04	0.203	0.0574	yes	yes	yes	
38803.CEL	550416420...	B04	0.225	0.0542	yes	yes	yes	
38831.CEL	550416420...	C04	0.207	0.0529	yes	yes	yes	
38842.CEL	550416420...	D04	0.207	0.0583	yes	yes	yes	
38847.CEL	550416420...	E04	0.21	0.0573	yes	yes	yes	
38953.CEL	550416420...	F04	0.207	0.0569	yes	yes	yes	
39336.CEL	550416420...	G04	0.214	0.0646	yes	yes	yes	
39345.CEL	550416420...	H04	0.21	0.0589	yes	yes	yes	
39348.CEL	550416420...	A05	0.209	0.055	yes	yes	yes	
39400.CEL	550416420...	B05	0.228	0.0692	yes	yes	yes	
39601.CEL	550416420...	C05	0.22	0.0579	yes	yes	yes	
39629.CEL	550416420...	D05	0.205	0.0548	yes	yes	yes	
39663.CEL	550416420...	E05	0.218	0.0544	yes	yes	yes	
39670.CEL	550416420...	F05	0.216	0.0719	yes	yes	yes	
39687.CEL	550416420...	G05	0.215	0.0604	yes	yes	yes	
39705.CEL	550416420...	H05	0.21	0.0567	yes	yes	yes	
39729.CEL	550416420...	A06	0.216	0.0569	yes	yes	yes	
39754.CEL	550416420...	B06	0.223	0.058	yes	yes	yes	

Sample data plots (lower) pane

Samples highlighted in the Sample Table appear in the Sample Data Plot pane, as shown in Figure 199. To filter the displayed samples, see "Filtering Samples" on page 143.

Figure 199 Samples Data Plots pane



Data Tracks that are generated in the Sample Data Plots pane are as follows:

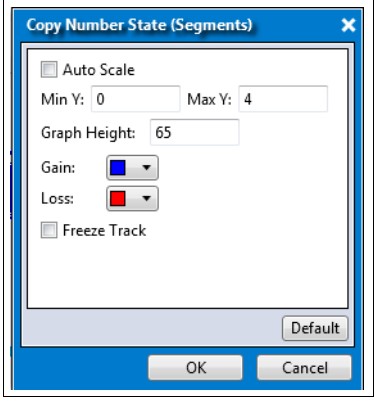
- Copy Number State Segment: shows copy number non neutral states
- Log2ratio track
- Smooth Signal track
- BAF track

Customizing the sample data plots pane

From the Sample Data Plots pane, click ⚙ button.

A Copy Number State (Segments) window appears. (Figure 200)

Figure 200 Copy Number State (Segments) window



- By default, the **Auto Scale** check box is checked and each axis is given a value (based on the data). Uncheck the Auto Scale check box to activate both the Min and Max value fields, then enter the new values.
- Click inside the Graph Height value field to enter a new value.
- Click the Gain/Loss Color drop-down to assign a different color.
- The Freeze Track check box is checked by default and displays the track in the upper WGV window. Uncheck this box to turn off the track. This option is checked by default because it can be useful for annotation tracks. Frozen/Static tracks can make for easier viewing while scrolling through your probe level data. Uncheck this box to turn off this feature off.
- Click **OK** to save your configurations or click **Default** to return all settings back to their factory defaults.

Filtering Samples

1. In the Frequency Histogram pane, highlight a region of interest by right-clicking on it, then dragging the pointer.

   The Samples contained in the segment are highlighted in the Sample Table.

Filtering your samples without specifying a region(s)

1. From the Sample Table, click or Ctrl click to highlight the samples you want to filter.
1. Click ☰ (upper right corner).

   A menu appears.
2. Click **Graph Settings**.

   A Whole Genome View Configuration window appears.
3. Click **OK**.

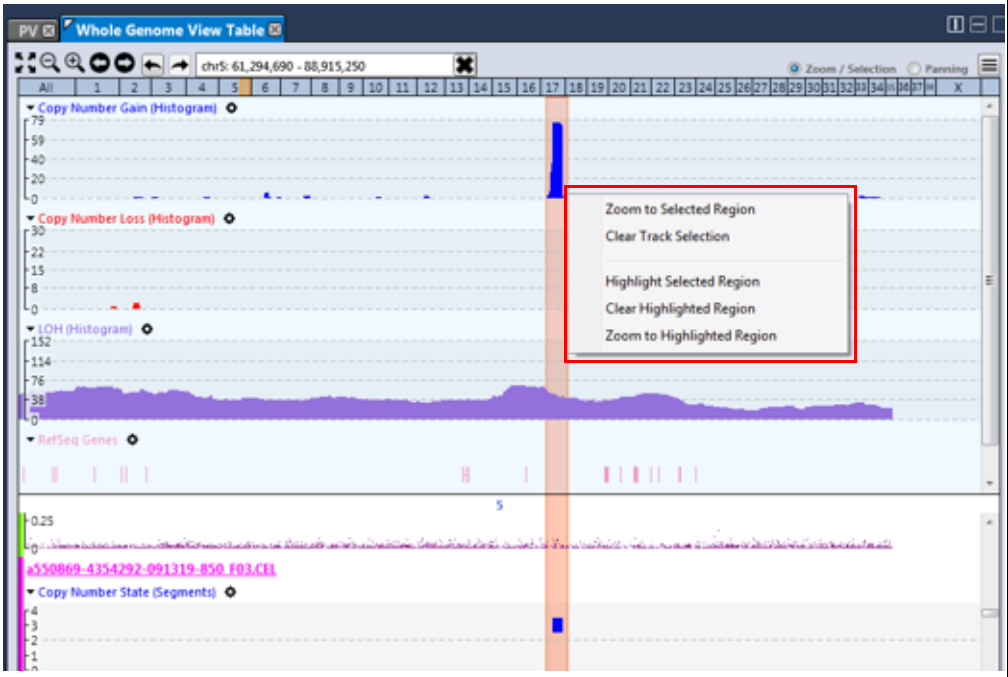
   The WGV plot is now filtered for samples displayed in the Sample Table.
4. Optional: To reset the WGV plot filter, highlight all the samples (Ctrl a) in the Sample Table, click ☰, click **Graph Settings**. At the Whole Genome View Configuration window, click **OK**.

Highlighting regions of interest

1. To highlight a region of interest in the Frequency Histogram or Sample Data Plot, right-click on the region, then drag your mouse to highlight it.

   A pop up appears, as shown in Figure 201.

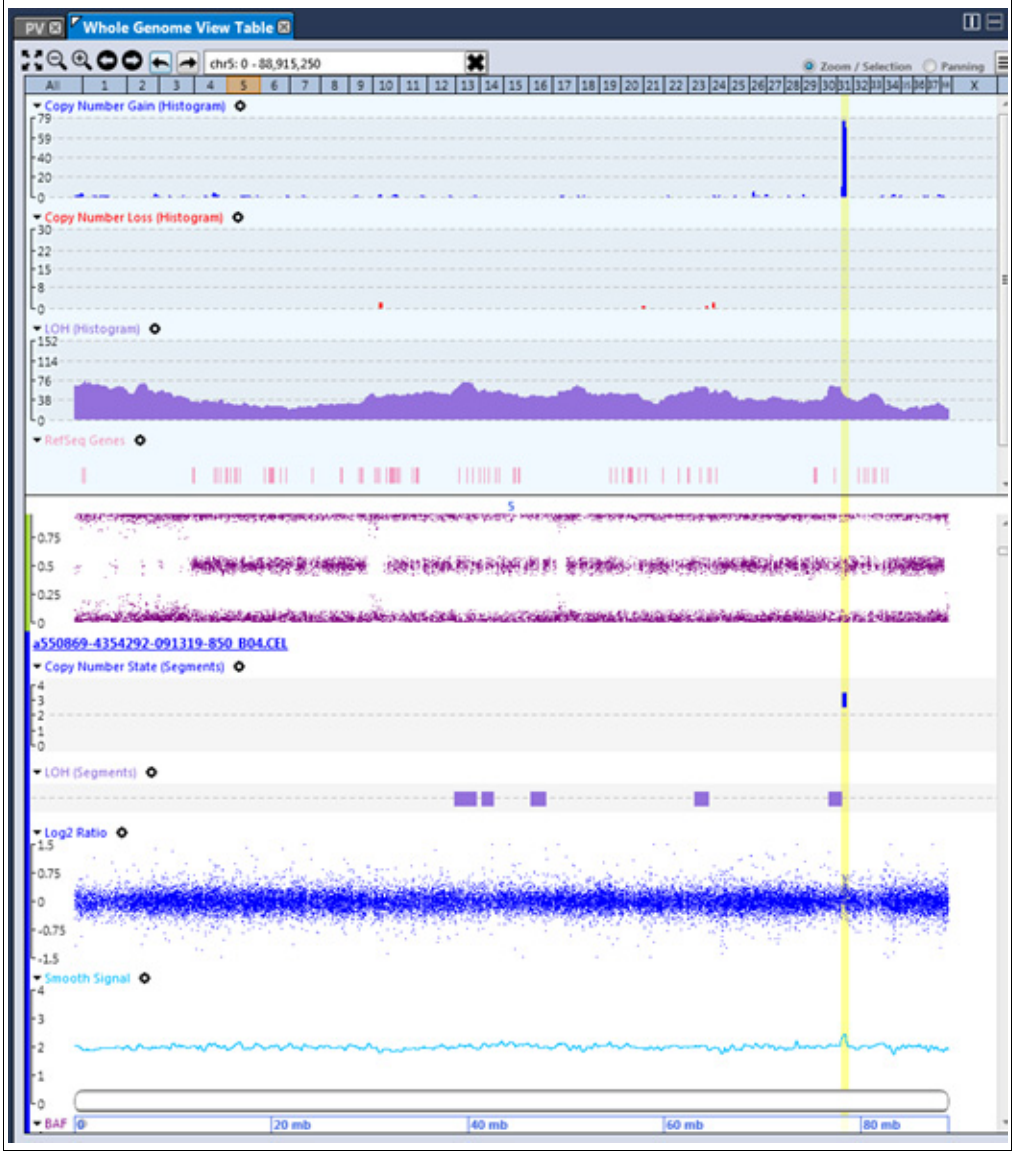
Figure 201 Region of interest



2. To highlight the region of interest in the Frequency Histogram and Sample Data Plots, click **Highlight Selected Region**.

   The region is now highlighted, as shown in Figure 202.

Figure 202 Highlighted region of interest

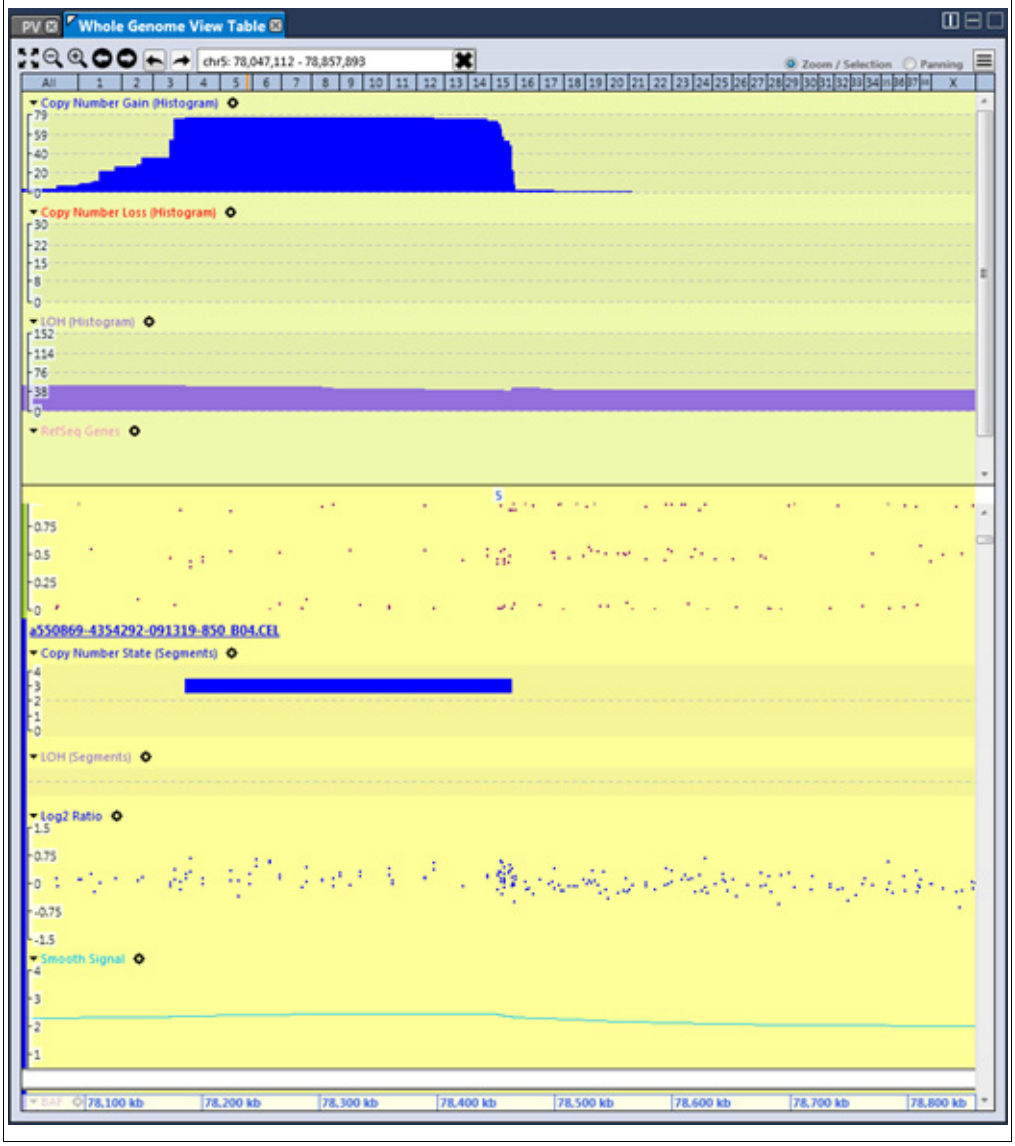


3. (Optional) Right-click, then select **Clear Highlighted Region** to remove the highlight.

Zooming into a highlighted region

To zoom in on your highlighted region in the Frequency Histogram and Sample Data plots, click **Zoom to Highlighted Region**. The highlighted region is now magnified, as shown in Figure 203.

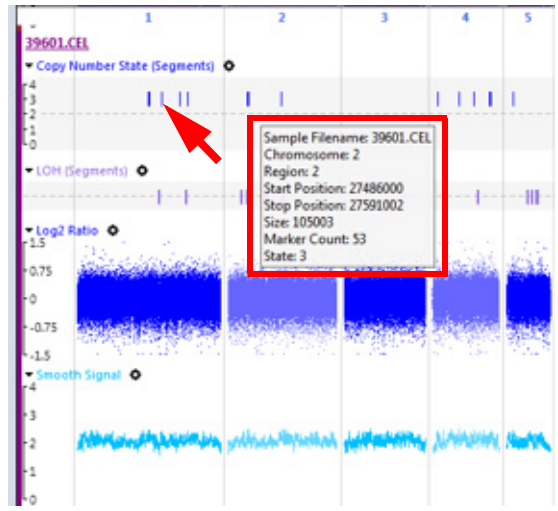
Figure 203 Highlighted region of interest - Magnified (Zoomed in) example



Viewing sample data plot segment information

At the Sample Data plot, use your mouse to hover over a highlighted gain to see its detailed information for the segment, as shown in Figure 204.

Figure 204 Hover mouse to see details example



Synchronizing the CN segment table and whole genome plot

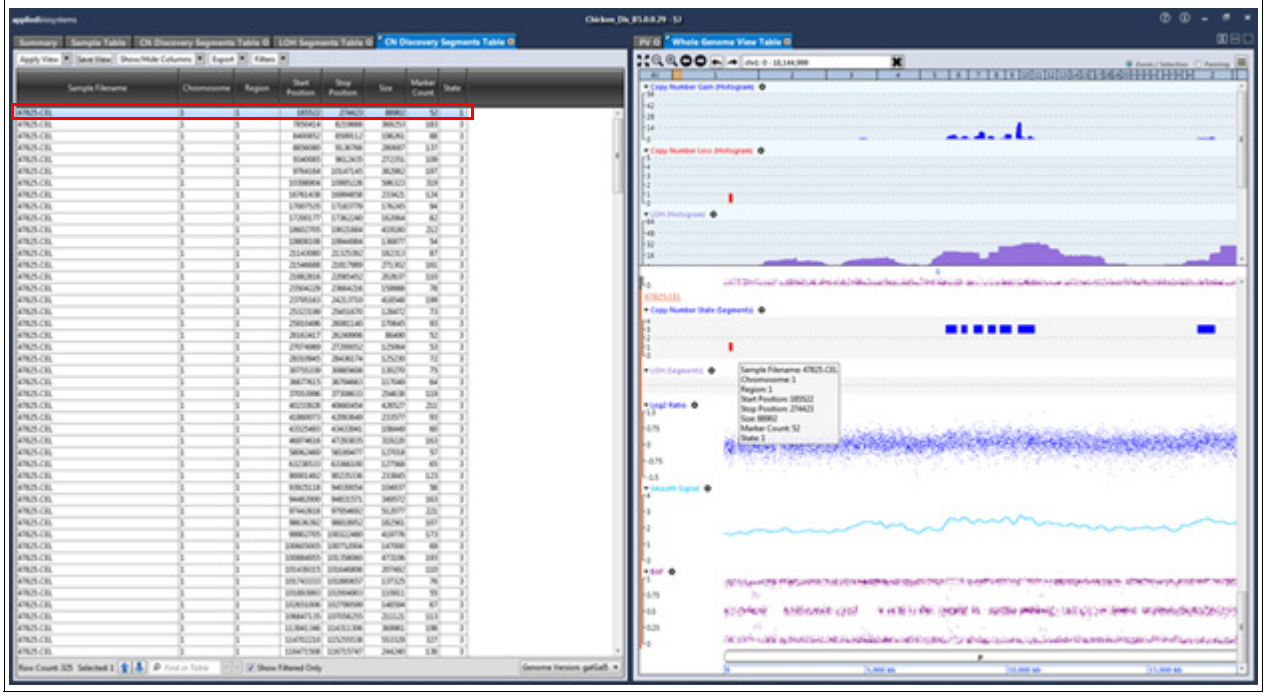
1. From the Sample Table, click the **QC Analysis** drop-down, then select **CN Segments**.

   A CN Discovery Segments Table tab is created. The table contains all the segments for the samples selected in the sample table.

2. Left-click on a gain in the Sample Data Plot of the WGV.

   The segment is highlighted in the CN Discovery Segment Table, as shown in Figure 205.

Figure 205 Synchronized CN Segment table and Whole Genome Plot example



CN fixed regions batches

Viewing a CN Fixed Regions Batch

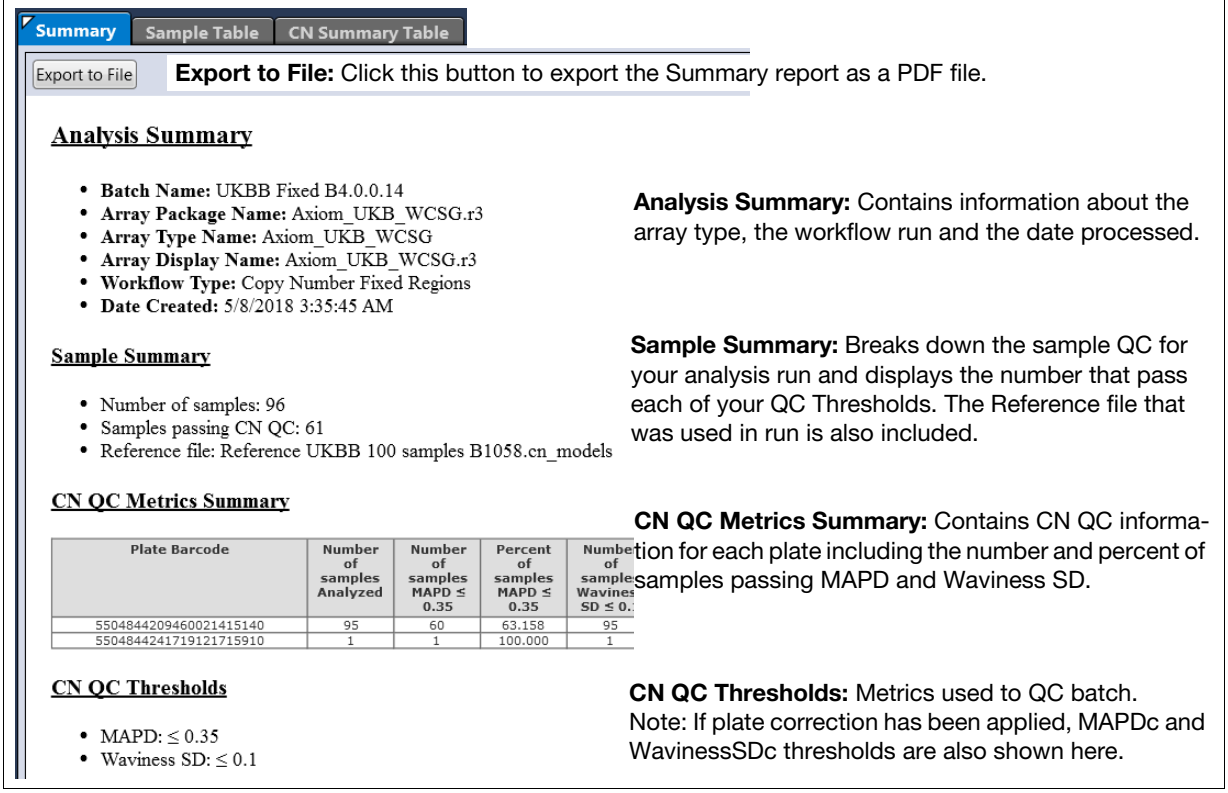
1. At the Dashboard window tab, click the **Open** button of the batch you want to open in the AxAS Viewer.

   The Viewer opens with five default window tabs.

   - Summary Table
   - Sample Table
   - CN Summary Table
   - Plate View (PV) MAPD
   - CN Region Plot

Summary

Figure 206 Summary window tab and definitions



Sample table

Figure 207 Sample Table window tab

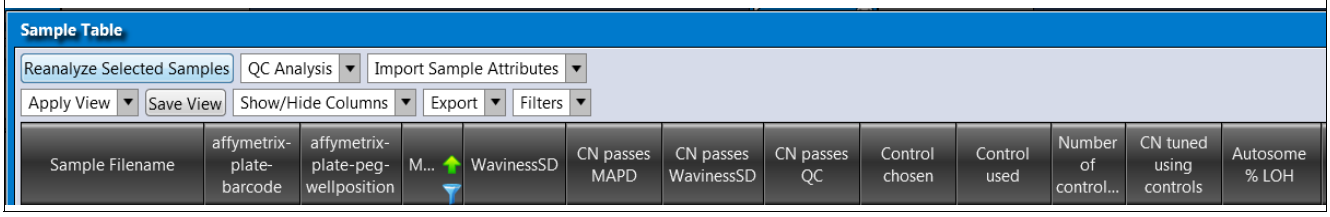


Table 8 Default CN Fixed Regions columns

Column	Description
MAPD	A global measure of the variation of all microarray probes across the genome. It represents the median of the distribution of changes in Log2 Ratio between adjacent probes. Since it measures differences between adjacent probes, it is a measure of shortrange noise in the microarray data.
Waviness SD	A global measure of variation of microarray probes that is insensitive to shortrange variation and focuses on long-range variation.

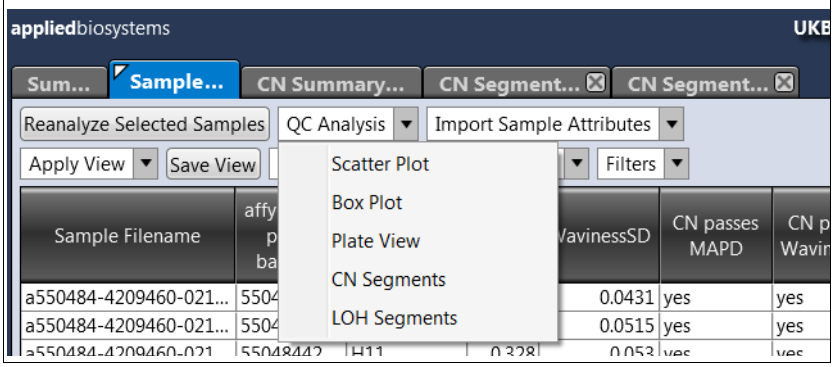
Table 8 Default CN Fixed Regions columns

Column	Description
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.
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.
Number of Controls	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.

See Table 19 on page 193 for definitions of the columns that get added after a plate correction is applied to a batch.

(Optional) For more CN Fixed Regions options, click the QC Analysis drop-down menu, as shown in Figure 208.

Figure 208 CN Fixed Regions; Additional options



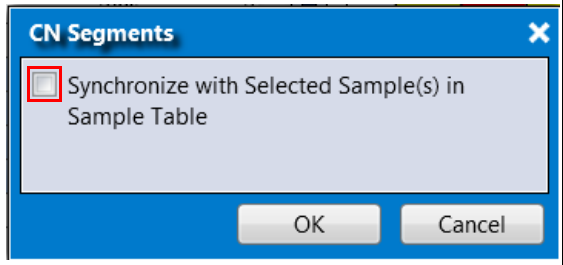
CN segments table

You can generate a CN Segments Table for a sample list and to show which regions show a CN state.

1. At the Sample Table, highlight your samples of interest.
2. Click the QC Analysis drop-down, then select **CN Segments**.

   A CN Segments window appears. (Figure 209)

Figure 209 CN Segments window



3. Click to check the check box, then click **OK**.

   A CN Segments Table window tab appears that is based on the samples you previously highlighted in the Sample Table, as shown in Figure 210.

   Note: The asterisk on the window tab (Figure 210) indicates the table is synchronized with the Sample Table.

Figure 210 CN Segments Table window tab

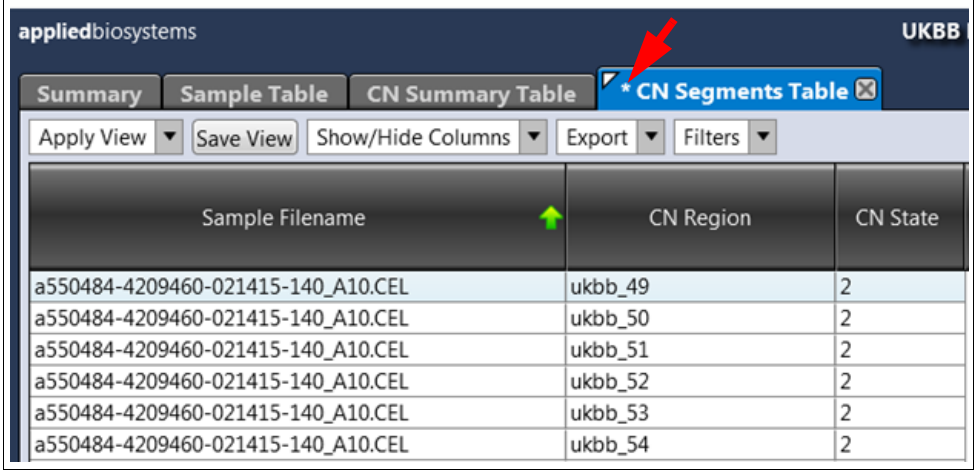


Table 9 CN Segment Table columns

Column	Description
Sample Filename	CEL file name
CN Region	Chromosomal region specified in library file package that copy number analysis will be performed.
CN State	Copy number state of sample in region.
MedianLog2Ratio	Median log2ratio of selected probesets for each CN Region for each sample.
Confidence	A calculation to determine likelihood of CN_raw. Higher confidence values indicated more confident CN calls.

Table 9 CN Segment Table columns

Column	Description
CN_raw	The initial assignment of CN calls by the copy number algorithm. Each raw CN call is then mapped to the final CN State based on library file package.
affymetrix-plate-barcode	Plate barcode number

LOH segments table

Loss of Heterozygosity (LOH) is also determined during CN Fixed Regions analysis. For more information see *Affymetrix® White Paper: The Loss of Heterozygosity Algorithm in Genotyping Console 2.0*.

Generating a LOH Segments Table

1. At the Sample Table, highlight your samples of interest.
2. Click the QC Analysis drop-down, then select **LOH Segments**.

   A LOH Segments Table window tab appears. (Figure 211)

Figure 211 LOH Segments Table window

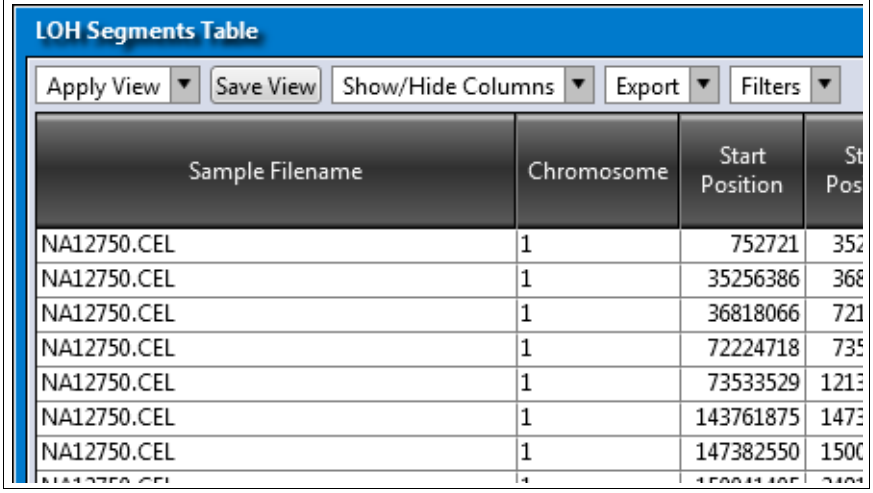


Table 10 Default LOH Segments Table

Column	Description
Chromosome	Chromosome number where LOH is located.
Start Position	The nucleotide base start position of the LOH segment.
Stop Position	He nucleotide stop position of the LOH segment.
Size	Length of LOH in base pairs.
Marker Count	Number of ProbeSets that make up segment.
LOH	A flag of 1 indicates the presence of LOH.

IMPORTANT! The default view of the LOH Segments table does not have any filters on Size or Marker Count. As part of reviewing your LOH data, create a Filter on Marker Count >25. To view only LOH segments, click on the Table's Filter drop-down menu and select contains 1.

CN summary table and CN region plot

CN Fixed Regions Batches auto-generate a CN Summary Table and CN Region Plot window tabs, as shown in Figure 212.

Figure 212 CN Summary table and CN Region Plot

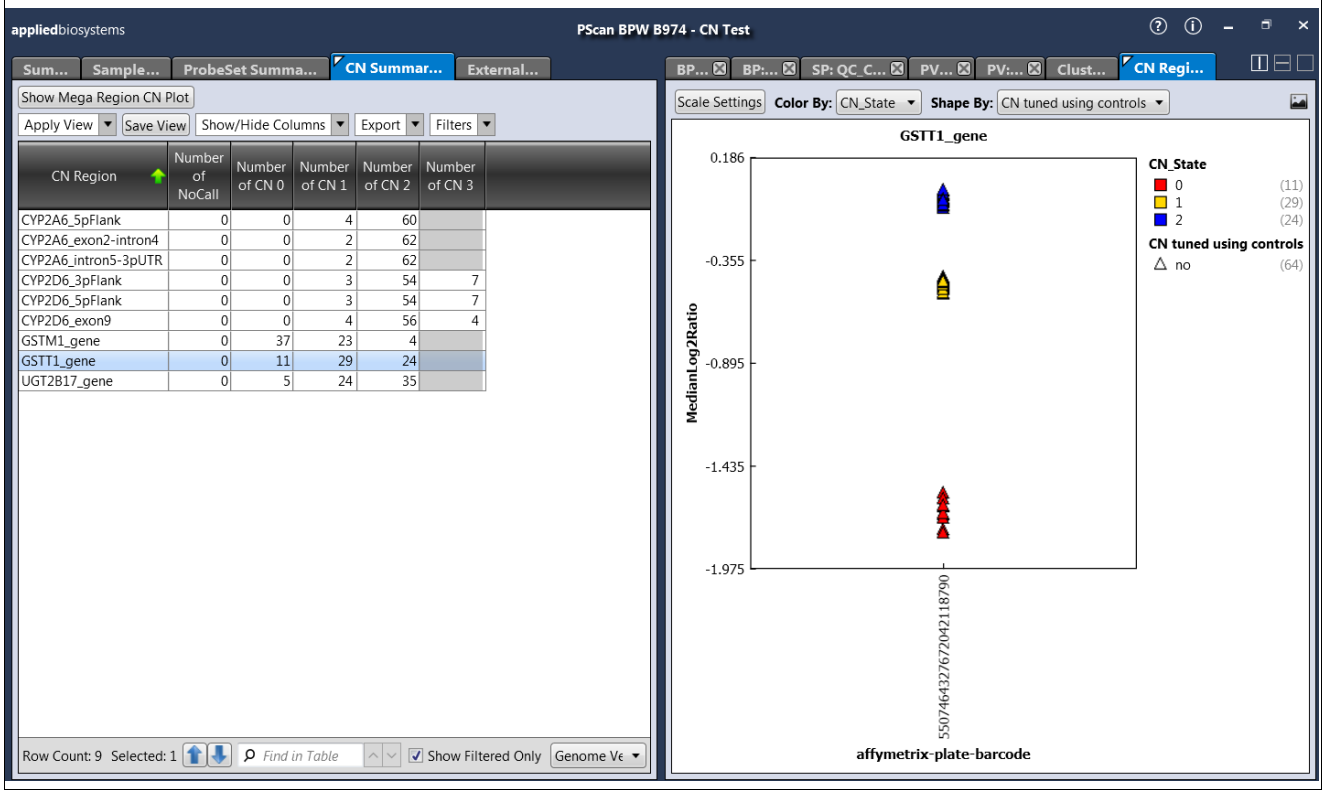


Table 11 CN Summary Table

Column	Description
CN Region	Predefined region in chromosome on which copy number is called.
Number of NoCall	Count of NoCalls per region. If a sample fails CN QC, it is reported as 'NoCall'.
Number of CN0	Count of CN0 per region
Number of CN1	Count of CN1 per region
Number of CN2	Count of CN2 per region
Number of CN3	Count of CN3 per region

Table 11 CN Summary Table

Column	Description
Chromosome	Chromosome number which CN Region is located.
Number of NoCall_*	Count of copy number calls for the specific bin described by *. These copy number calls are calls that are between identified copy number states.

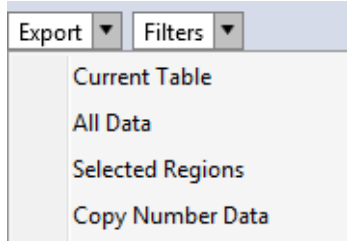
Exporting a CN summary table

Exporting a CN Summary Table

1. Click the Export drop-down menu.

   Four export options appear. (Figure 213)

Figure 213 Export drop-down



2. Click to select the option you want to use.
   - **Current Table:** Exports data currently shown in table.
   - **All Data:** Export all columns and data in the CN Summary Table
   - **Selected Regions:** Exports only highlighted regions.

   Note: The resulting **AxiomCNVMix.cnregioncalls.txt** file is sorted by regions.

   - **Copy Number Data:** Exports all associated CN files.

Mega region CN plot

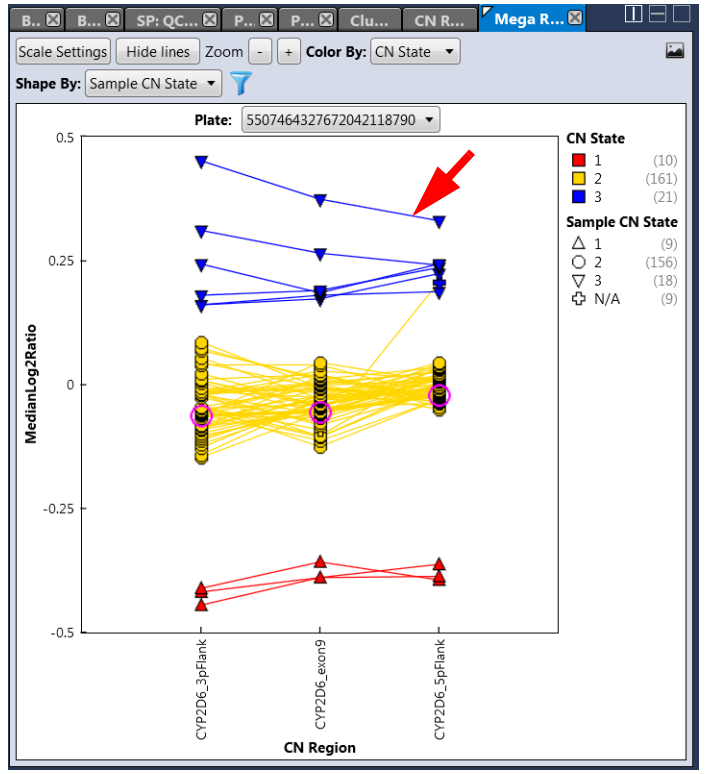
A single plot of adjacent regions can be created from the CN Summary Table.

IMPORTANT! A Mega Region Plot can only be created with CN Regions from the same chromosome.

1. From the CN Summary Table, highlight the regions of interest.
2. Click the **Show Mega Region CN Plot** button.

   A Mega Region Plot (window tab) is generated. (Figure 214)

Figure 214 Mega Region CN Plot window tab



Mega region CN plot overview

- Y axis is median Log2Ratio, the CN Regions in genomic order are shown on the X axis
- Samples in the plot are represented by a shape that are connected via a line, as shown in Figure 214.
- Click the **Hide lines** button to remove the lines.
- The default Color By is CN State of each of the samples regions.
- The Shape By is the Sample CN State. If all regions do not agree, then the Sample will be given a CN State of N/A.
- Click the  icon (upper right) to export the plot as a PNG file.
- Other plot features include, a zoom feature and a filtering option to view specific samples.

Exporting

CN data files in various formats can be exported from the Sample Table for both CN Discovery and CN Fixed Regions analyzed batches.

Exporting to Integrative Genome Viewer (IGV)

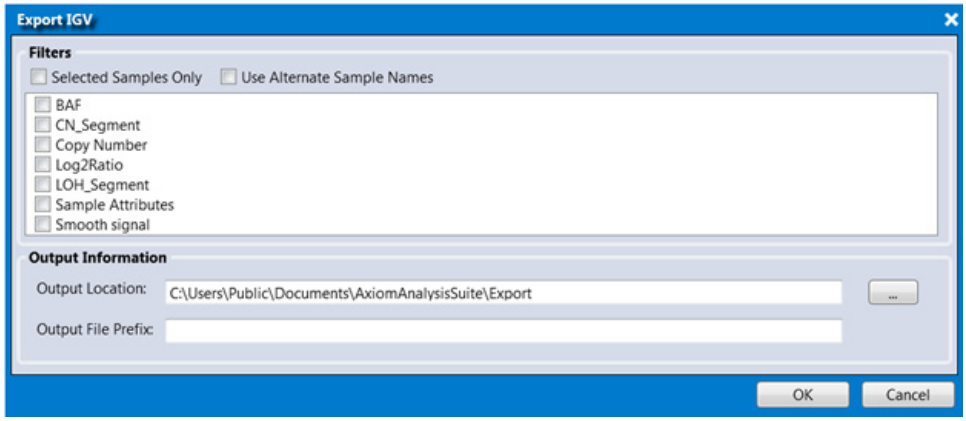
Before you can export to IGV, you must first download and install the IGV software. To download IGV, click on this link:

http://software.broadinstitute.org/software/igv/home

1. After IGV has been successfully installed, go to the Sample Table, then click the **Export** drop-down menu and select **IGV**.

   The Export IGV window appears. (Figure 215)

Figure 215 Export IGV window



2. Click the data file check box(es) to select files you want to export.
3. (Optional) Click to check the **Selected Samples Only** check box to export only samples that are highlighted in Sample Table. If sample attributes with alternate sample name columns are present in the Sample Table, click to check the **Use Alternate Sample Names** check box.
4. (Optional) To change the output location, click the browse button, then navigate to a new output location.
5. (Optional) Add an optional Output File Prefix to your files. If this field is left blank, the file is exported with a date and time prefix.
6. Click **OK** to start the export.

   Note: Depending on number of samples, size of array and number of data files selected for export, the exporting might take several minutes to complete.

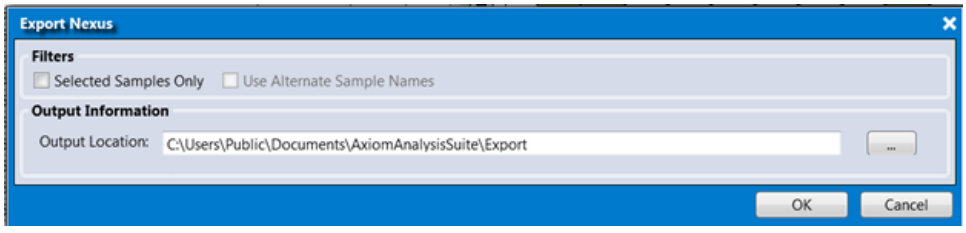
   Once export is complete, the output folder with the exported files opens automatically and your data is now ready to be imported into the IGV browser.

Export Nexus format

1. From the Sample Table, click the **Export** drop-down menu, then click to select **Nexus**.

   A Export Nexus window appears. (Figure 216)

Figure 216 Export Nexus window



2. Click to check the **Selected Samples Only** check box to export only samples that are highlighted in Sample Table. If sample attributes with alternate sample name columns are present in the Sample Table, click to check the **Use Alternate Sample Names** check box.
3. (Optional) To change the output location, click the browse button, then navigate to a new output location.
4. Click **OK** to start the export.

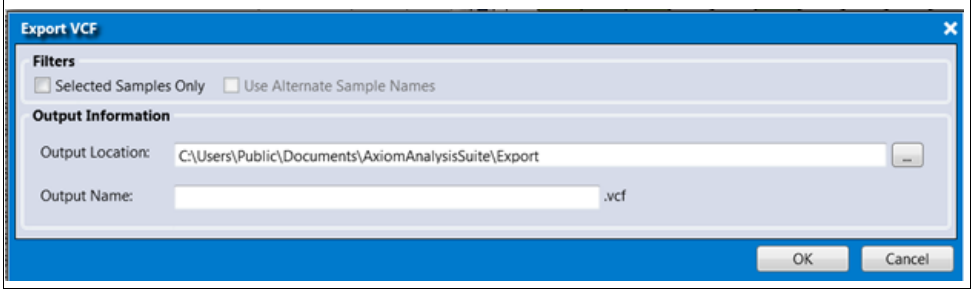
   Once export is complete, the output folder with the exported files opens automatically.

Export VCF format

1. From the Sample Table, click the **Export** drop-down menu, then click to select **VCF**.

   A Export VCF window appears. (Figure 216)

Figure 217 Export VCF window



2. Click to check the **Selected Samples Only** check box to export only samples that are highlighted in Sample Table. If sample attributes with alternate sample name columns are present in the Sample Table, click to check the **Use Alternate Sample Names** check box.
3. (Optional) To change the output location, click the browse button, then navigate to a new output location.
4. Click **OK** to start the export.

   Once export is complete, the output folder with the exported files opens automatically.

6 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 C, "About allele translation" on page 202 for more information.

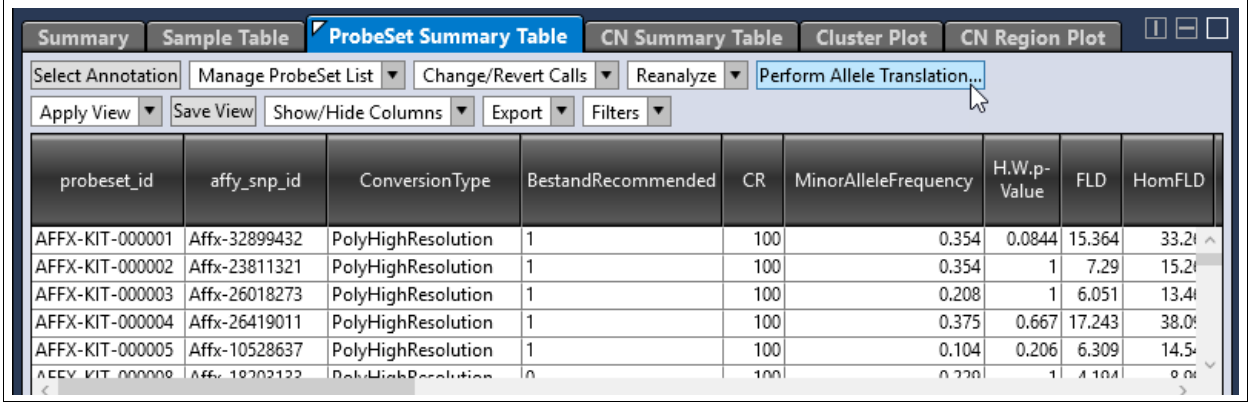
Performing translations

Perform Allele Translation 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.

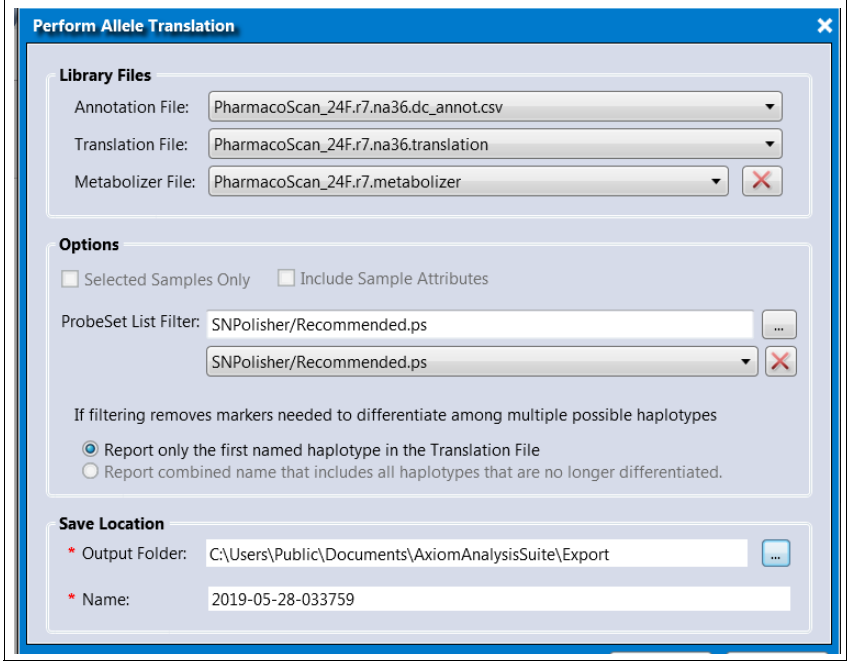
1. From an open batch in the Viewer, navigate to the ProbeSet Summary Table, then click the **Perform Allele Translation** button, as shown in Figure 218.

Figure 218 ProbeSet Summary Table window



The Perform Allele Translation window appears. (Figure 219)

Figure 219 Perform Allele Translation window



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 219) A Phenotype report will not be created if this option is not used.

Allele Translation Options

Refer to Table 12 and Table 13 for descriptions of the available **Allele Translation** options.

Table 12 Allele Translation options

Select Options	Description
ProbeSet List Filter	This option to translates only the genotypes of SNPs in a user-specified probeset list. Click Browse to select a different probeset list, or select from existing probeset lists. You can also click X to deselect probeset list filtering, which enables allele translation for all available probesets, including those that may not have passed SNP QC.
Report only the first named haplotype in the translation file (Default)	This option is only relevant IF you filter by a ProbeSet List AND if the ProbeSet 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 ProbeSet 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.
Report combined name that includes all haplotypes that are no longer differentiated	This option is only relevant IF you filter by a ProbeSet List AND if the ProbeSet 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 ProbeSet 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).

Table 12 Allele Translation options

Select Options	Description
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.
Selected Samples Only	If this check box is checked, only samples highlighted in the Sample Table will be used.

Table 13 More Allele Translation options

Select Translation Results Folder	Description
Output Folder	The path to the output folder for the translation reports and log. Click its Browse button to set the path.
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.

2. 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:

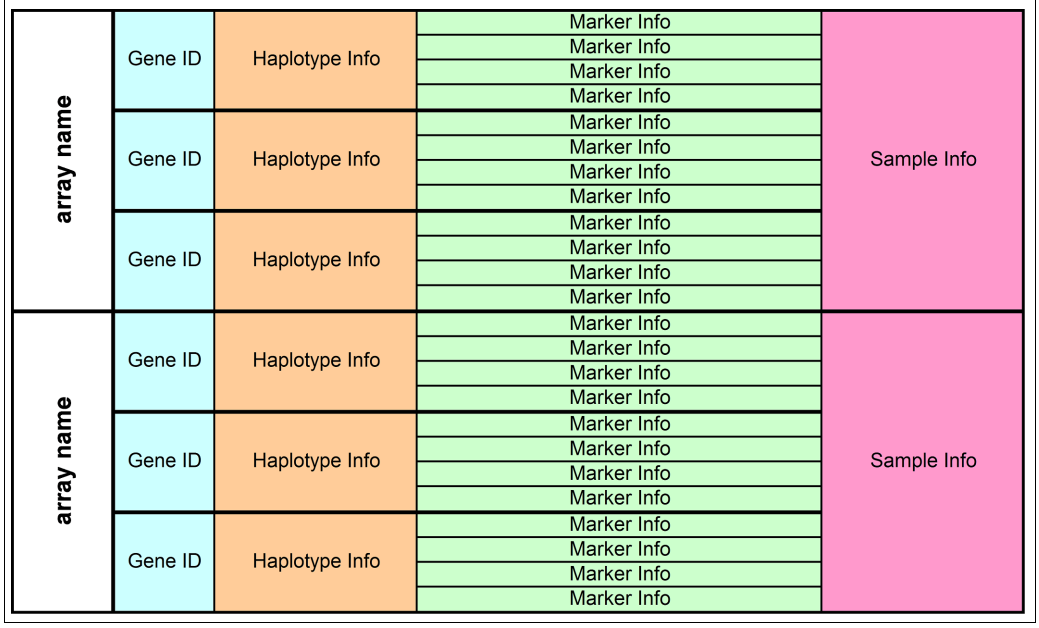
- **Alleles Report:** Displays the sets of haplotypes that can and cannot be called in the current batch. The full set of reportable haplotypes is managed by the translation library file. If some probeset data needed to call a haplotype is not available (either not genotyped, or not in the ProbeSet List File), then all non-reportable haplotypes will be listed for each gene.
- **Comprehensive Report**: Displays one row per translated SNP for each sample. Provides information on each SNP in addition to haplotype calls.
- **Summary Report**: 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 Report:** Displays one row per phenotyped gene for each sample, based on the diplotypes from the source Comprehensive 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 ProbeSet 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 220.

Figure 220 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 221.

Figure 221 Example: Summary Translation report

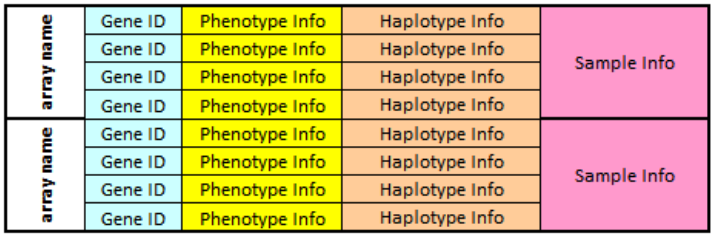
2008-11-05_122550_translations _summary.rpt

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	Haplotyp Marker
0001-	TEST_0001.	SLC15A2	"4A/"4A		UNIQ	R509K	"2,"4A,"4B	SLC15A2_34845A>G(R509I	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_(rs6759892)	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	Q30R	Q30R	VKORC1_1120A>G(Q30R)	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_(rs34815103)	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								

Phenotype translation report

The basic layout of this report is shown in Figure 222.

Figure 222 Example: Phenotype Translation report



Phenotype report

The basic layout of this report is shown in Figure 223.

Figure 223 Example: Phenotype report

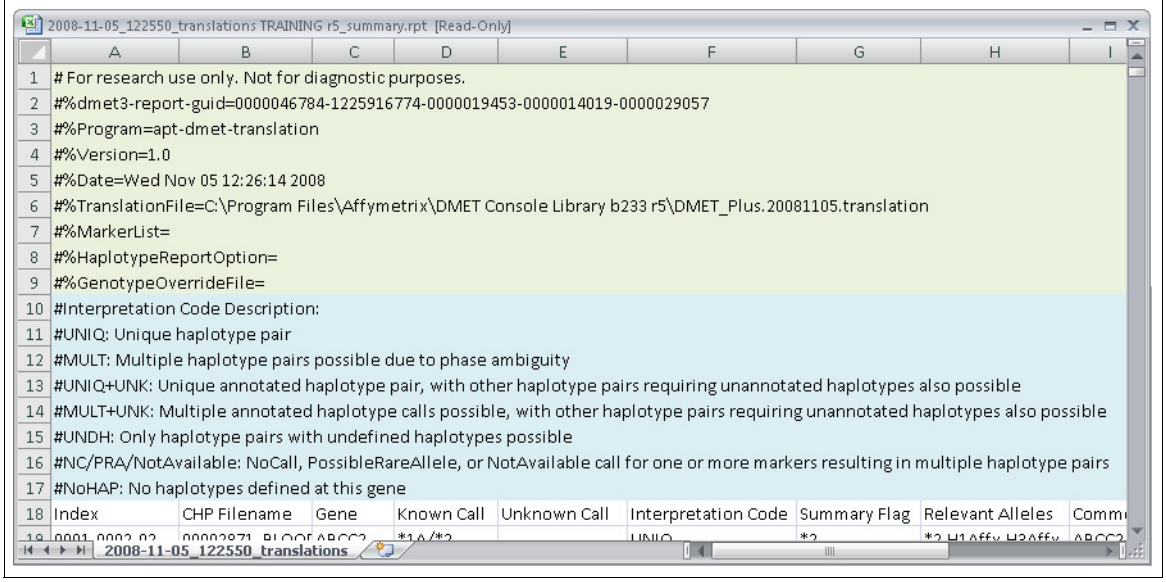
Index	CHP File	Gene	Phenotype Call	Gene Activity	Known Call	Unknown Call	Interpretation Code
0001-0020	test_01.c	CYP1A2	EM	normal/normal	*1F/*1F		UNIQ
0001-0022	test_01.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0001-0024	test_01.c	CYP2B6	EM	normal/normal	*1/*1		UNIQ
0001-0029	test_01.c	CYP2D6	PM	none/none	*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,UN	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 224.

Figure 224 Example: Header for Comprehensive and Summary report



Available report fields and descriptions

Array tracking

Refer to Table 14 for descriptions of the available **Array Tracking** fields.

Table 14 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 15 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 15 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 207 for more information.

Table 15 Gene-specific fields

Gene-specific fields	Description
Gene Function	In the Phenotype report, the predicted pair of gene functions given the supplied Known Call diplotypes. Multiple comma-separated function pairs are reported when multiple Known Call diplotypes are associated with different function 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 207 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. If a copy number gain is detected, the xN string is added to indicated an unknown number of extra copies are present. For example, *1/*1xN means at least 3 copies of the *1 haplotype are reported. If there are two different haplotypes reported and a copy number gain is reported, the software cannot determine which haplotype(s) were duplicated. In that case, xN indicates at least one of the haplotypes was duplicated, and the diplotype is bracketed by parentheses. For example, (*1/*2)xN indicates at least one extra copy of *1 and/or *2 haplotypes.

Table 15 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, undefined haplotype present 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 16 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 16 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)
Probeset 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.
cDNA Change (marker annotation)	Location of the mutation on a reference mRNA sequence.

Table 16 Marker-specific fields

Marker-specific fields	Description
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 17 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 17 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 18 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 18 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.

Table 18 Uncalled Report

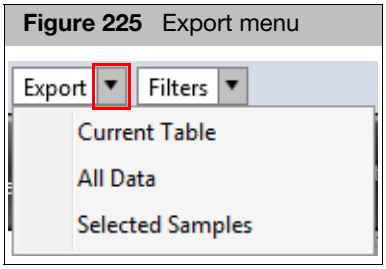
Uncalled Report	Description
Reference Allele	Reference base indicates the allele in a reference sequence known to be present at this genetic location. Generally this is the more common allele at SNPs with low minor allele frequency (<1%).
Variant Allele	The variant base(s) defined by the marker are alternative known genotypes known to be present at this genetic location. For triallelic SNPs, the reporting format is: A,T for ABCB1_c.2677G>T>A(A893SorT), because two specific mutations are known to occur at this genomic location (G>A and G>T).

7 Exporting

Using the sample table export options

1. From the Sample Table, click the **Export** drop-down. (Figure 225)

Figure 225 Export menu



2. Click **Current Table** or **All Data**.

   An Export 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.

Exporting specific samples from the table

1. Click to highlight the samples you want to export, then click the Export drop-down and click **Selected Samples**.

   An Export window appears.

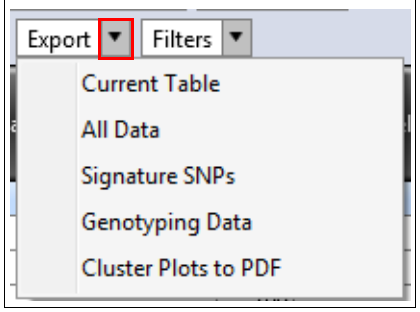
2. Click on an existing folder or click **New Folder** to choose a new save location.
3. Click **Save**.

   Your selected sample data is now saved as a tab-delimited TXT file.

Using the ProbeSet summary table export options

1. From the ProbeSet Summary Table, click the **Export** drop-down. (Figure 226)

Figure 226 Export menu



Exporting the currently displayed 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

Including 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 only 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.

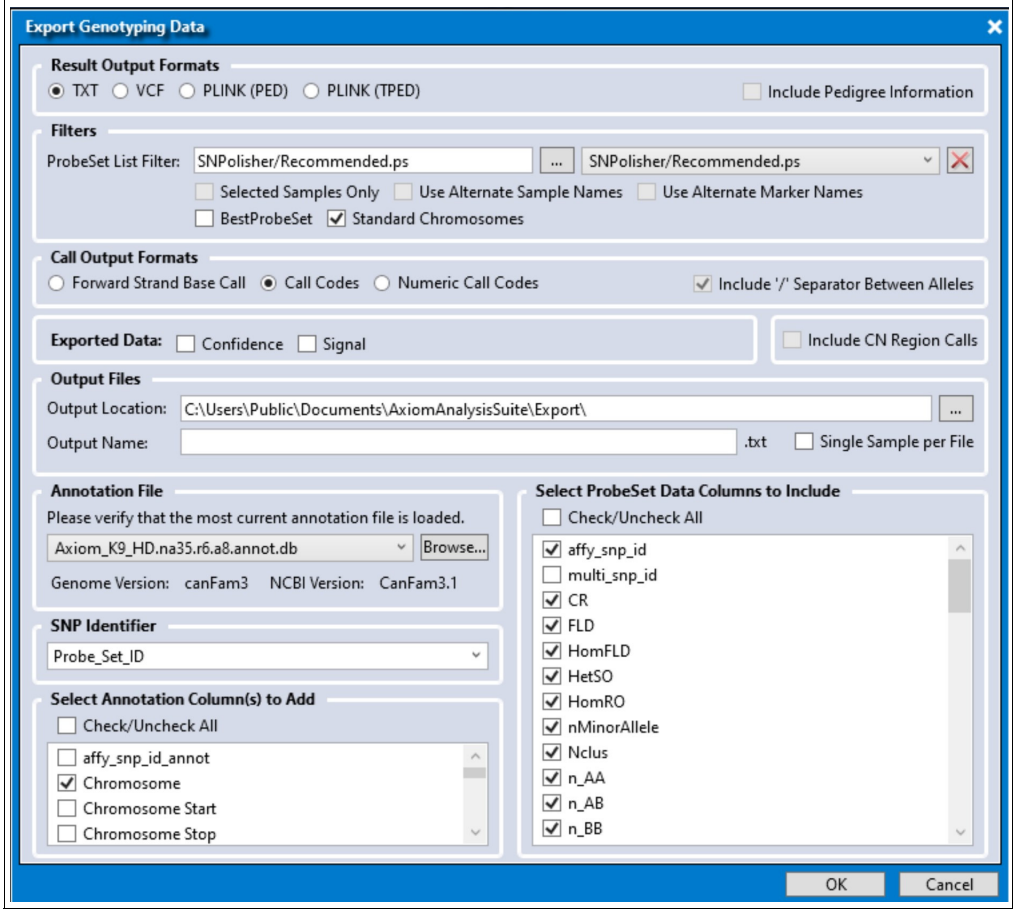
Exporting genotyping data

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**.

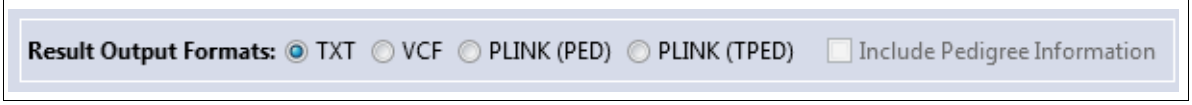
   The Export Genotyping Data window appears. (Figure 227)

Figure 227 Export Genotype Data window



Result Output Formats

Figure 228 Result Output Format selections



1. Click the radio button to select the Result Output Format you want use.
2. 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 multiallele 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.

Filters

- **Selected Samples Only:** Click the check box to export selected (highlighted) samples. **Note:** If no samples were previously highlighted in the Sample Table, this filter option is grayed out (unavailable).
- **Use Alternate Sample Names**: If your sample attributes include an Alternate Sample Name column, click the check box to enable its export.

Exporting genotyping data filters

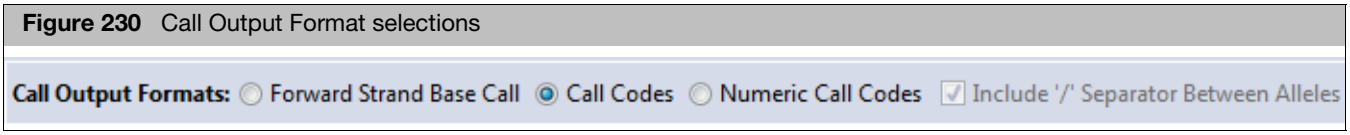
- **Use Alternate Marker Names:** Click this check box (Figure 229) to use your alternate marker name in place of probeset id.
- **BestProbeSet:** Default selection for Hide SNP Classification View export. Click the BestProbeSet check box to exports only probesets that are in the BestProbeSet category.
- **Standard Chromosomes:** Click this check box to include only probesets that are assigned to standard chromosomes or other annotated contigs.

Figure 229 Exporting Genotyping data filters

Call Output Formats

1. Click the radio button to select the appropriate Call Output Format (Figure 230) you want to use.

Figure 230 Call Output Format selections



Note: 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 231) you want to include.

Figure 231 Exported Data selections

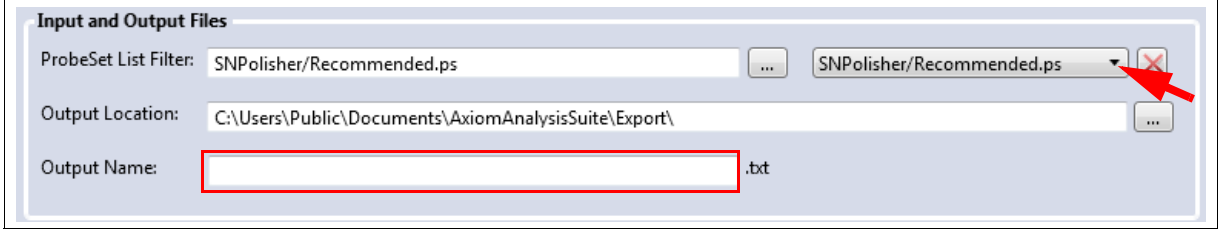
Note: Multiallelic probesets will not export signal values. Normalized signal values for multiallelic and biallelic probesets are available in the AxiomGT1.summary.txt file inside the batch results folder.

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.

ProbeSet List Filter

The ProbeSet List Filter defaults to the recommended list. To change, click on the drop-down, then select from the available list. If you have previously created a ProbeSet list, it will be displayed here.

Figure 232 ProbeSet List Filter and Output path



2. Click the ProbeSet List Filter's drop-down menu to view and select a previously saved list. (Figure 232)

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 232)

Output Name (Required)

1. Click inside the field to enter a new name.

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.

In a text format, 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 233)

Figure 233 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 234)

   An Explorer window appears.

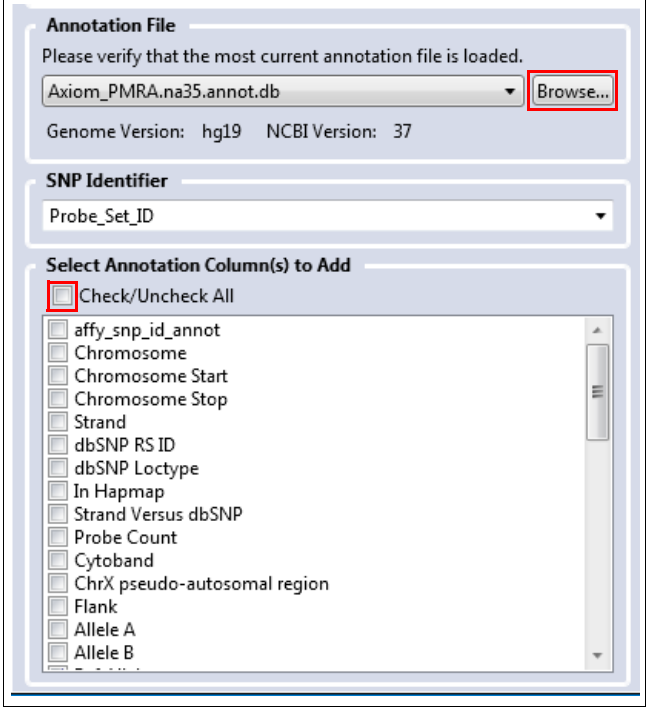
2. Navigate to the appropriate Annotation File location, then click **Open**.

   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 234)

Figure 234 Select Annotation Column(s) pane



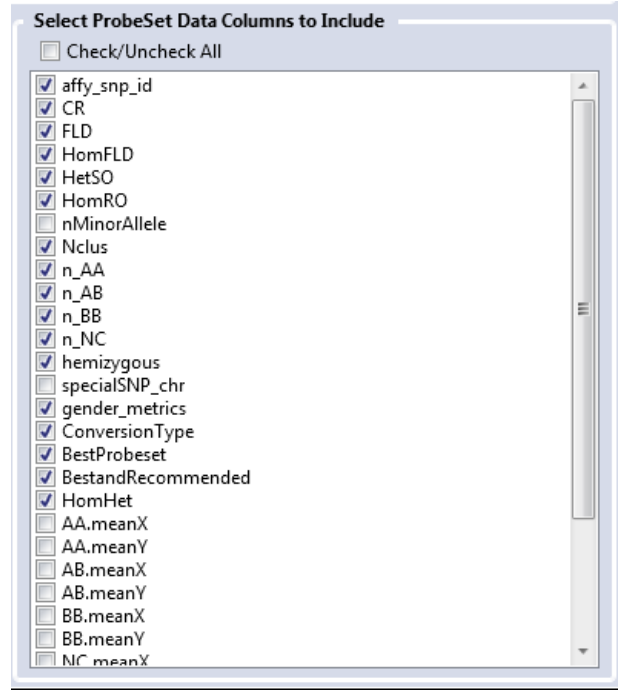
2. After the Export Genotype Data form is complete, click **OK**.
3. Your newly exported data now reside in the output location you defined in Step 1 on page 175.

Including ProbeSet Summary table columns in your genotyping report

1. Check ProbeSet data columns to include in your export, as shown in Figure 235.

   Note: The included ProbeSet data columns will be appended at the end of the files.

Figure 235 Select ProbeSet Data Columns to Include pane



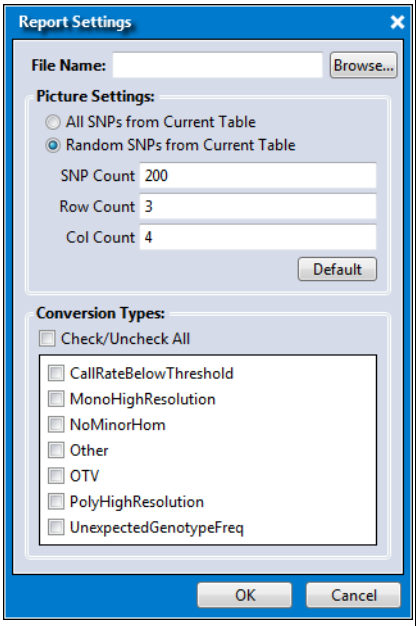
Exporting cluster plots to PDF

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

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

   The Report Settings window appears. (Figure 236)

Figure 236 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**.

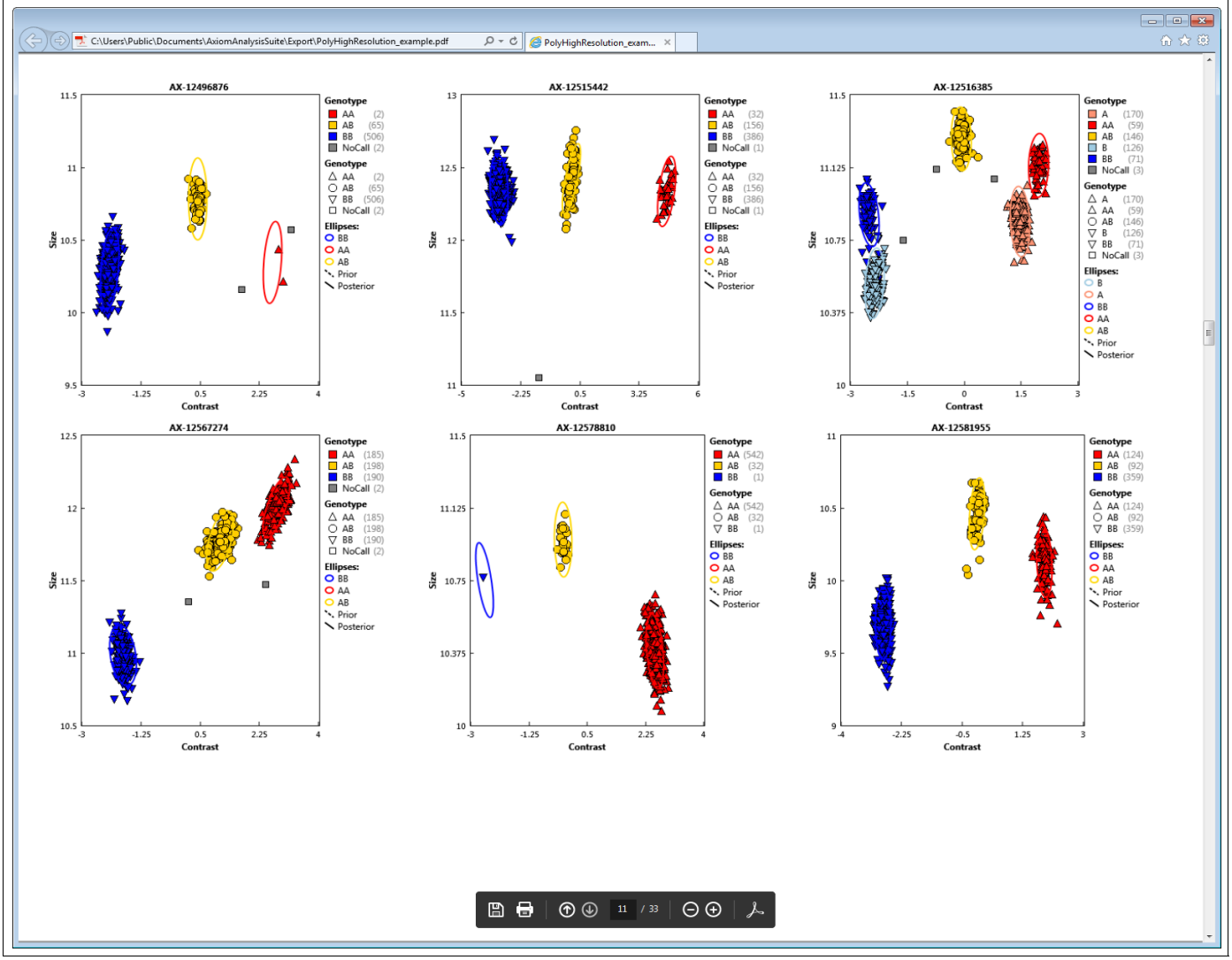
   A Please Wait message and progress bar appear. Allow several minutes if multiple Conversion Type PDF report(s) were selected.

   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 237)

Figure 237 PDF Report example



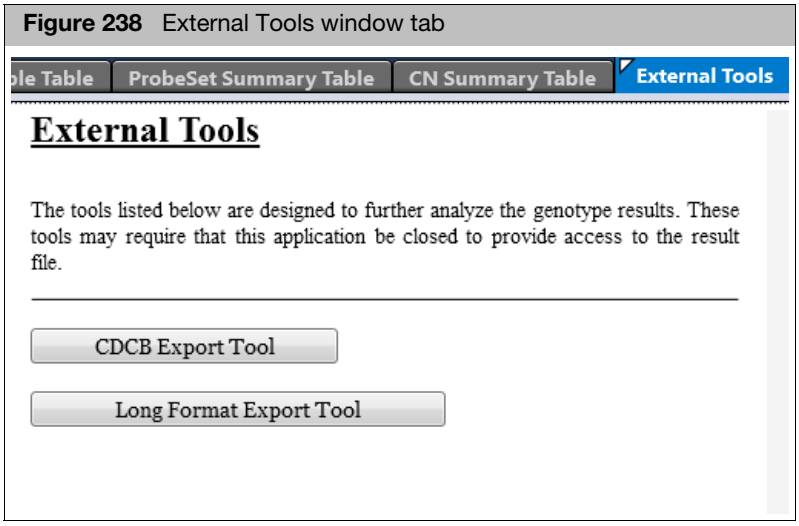
Note: One plot is made for each probeset. For multiallele probesets, the most informative pair of allele signal channels in the data set is used for plotting (based on the distribution of genotype calls).

8 External tools

Overview

AxAS auto-detects any previously installed application(s) that may be used to further analyze your genotype results. If a compatible application is detected, an **External Tools** window tab appears displaying the available tools, as shown in Figure 238.

Figure 238 External Tools window tab



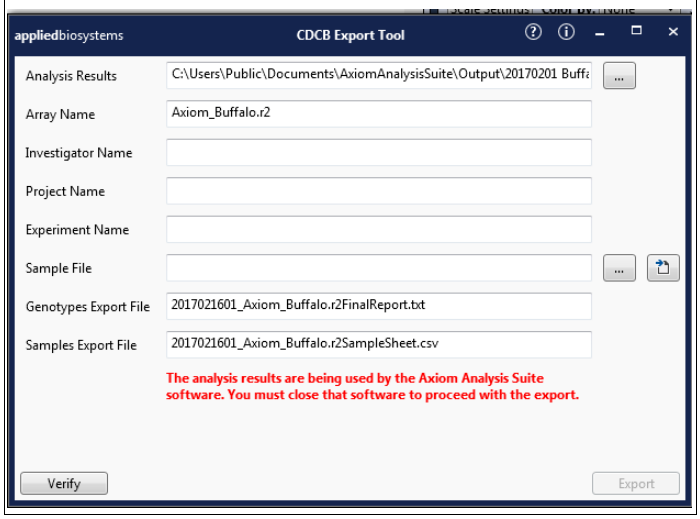
IMPORTANT! You must exit and close AxAS before editing its data with an external application.

Council on Dairy Cattle Breeding (CDCB) Export Tool

1. Click the **CDCB Export Tool** button.

   The Tool opens. (Figure 239)

Figure 239 CDCB Export Tool - Main window



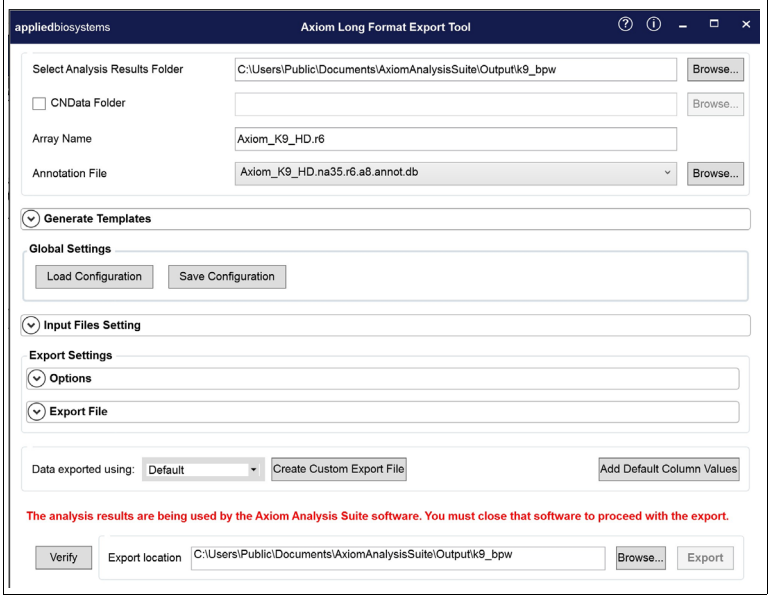
For details on how to use the CDCB Export Tool application, refer to its User Guide *(P/N 703465)*.

Axiom Long Format Export Tool

1. Click the **Long Format Export Tool** button.

   The Tool opens. (Figure 240)

Figure 240 Long Format Export Tool - Main window



For details on how to use the Long Format Export Tool application, refer to the *Axiom Long Format Export Tool 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.

AxAs supports two variations of restricted copy number analysis that also include genotyping. These analysis variations are determined by the array library package and are compared in Figure 241.

Figure 241 Table showing workflow analysis variations that support copy number

Workflow Stage		Analysis Description	Analysis Variation		
			Typical Axiom	Predefined CN Genotyping †	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" or "cnvmix_gt_2"

‡ library package specifies analysis_category = "cn_gt" or "cnvmix_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, CP, W, non-PAR X, and Z) 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.

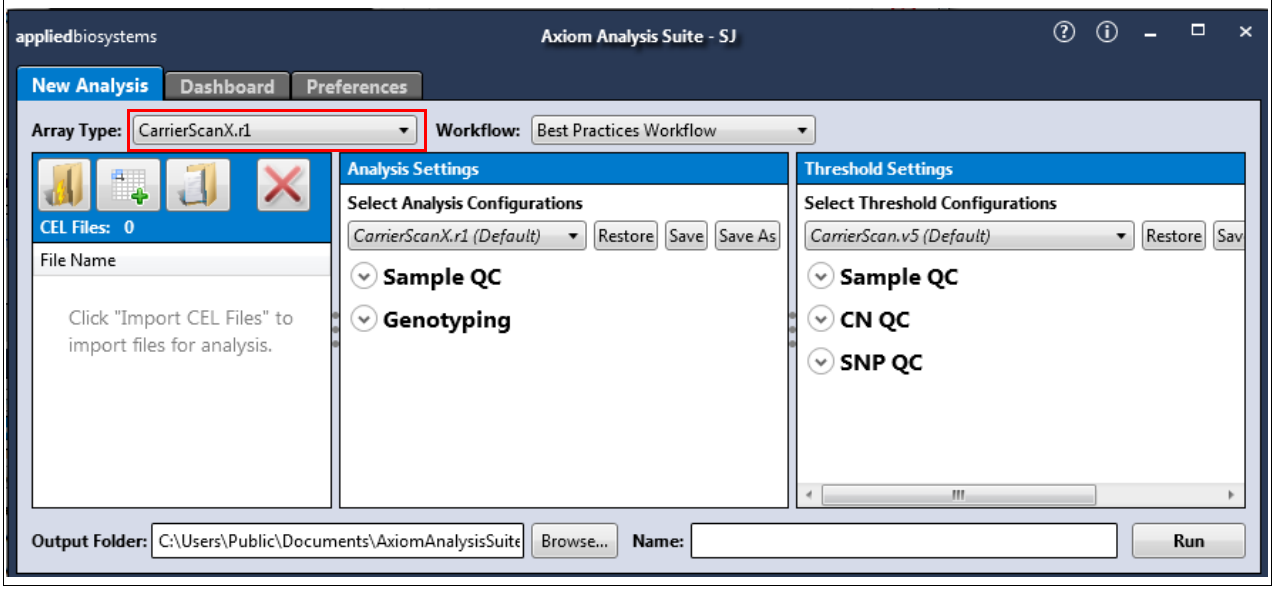
  Note: Examples of library packages that support these copy number analyses include; CarrierScan for Predefined CN genotyping, and PharmacoScan for CN-aware genotyping.

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 (Figure 242) to confirm the array you want to use (example PharmacoScan) is available. If it is, continue to "Selecting a mode (Workflow)".

Figure 242 New Analysis window tab



Selecting a mode (Workflow)

From the main Axiom Analysis Suite window tab, click the **Workflow** 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 198.

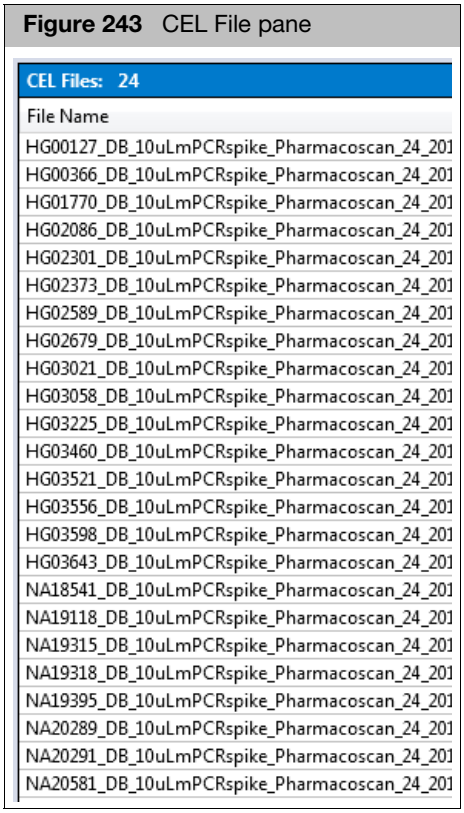
Importing CEL files

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 files. (Figure 243)

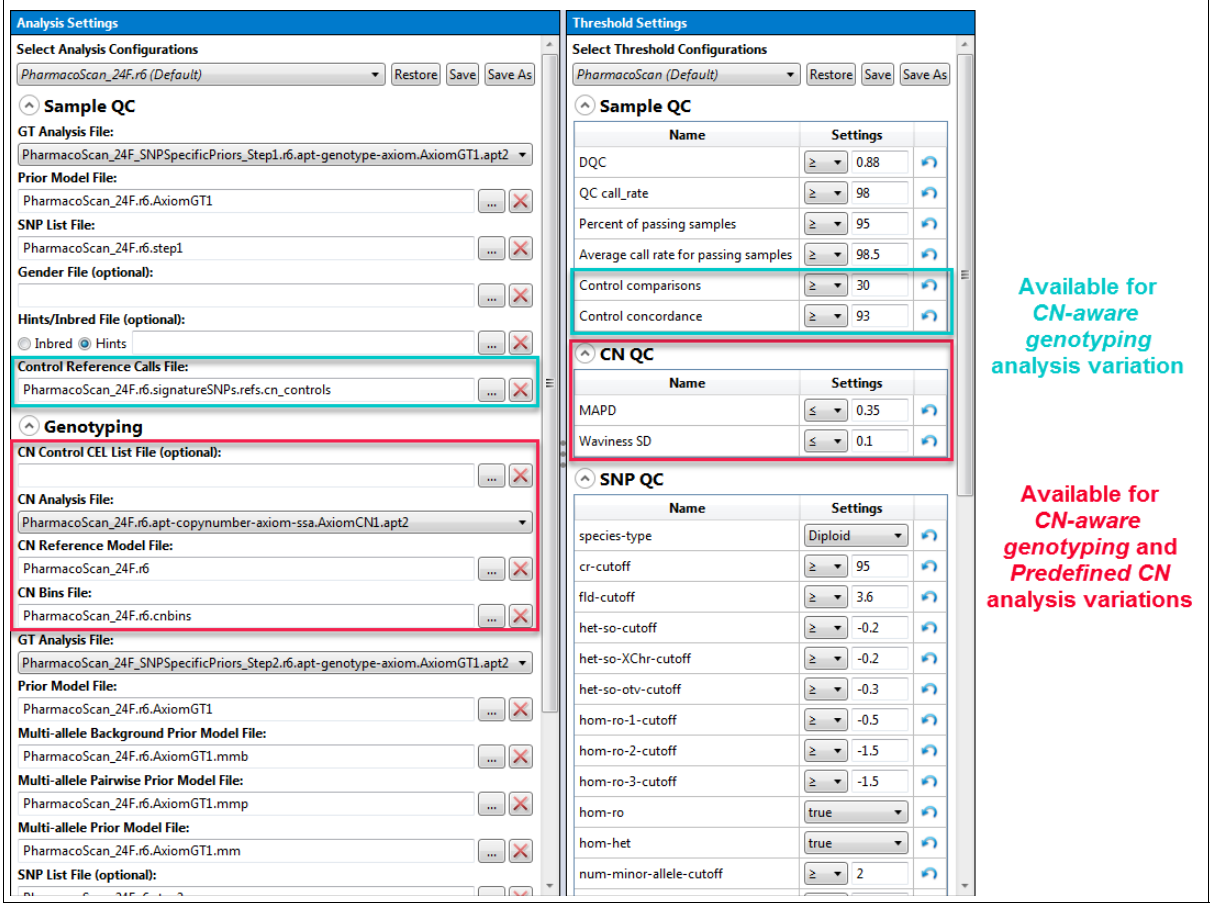
Figure 243 CEL File pane



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 244.

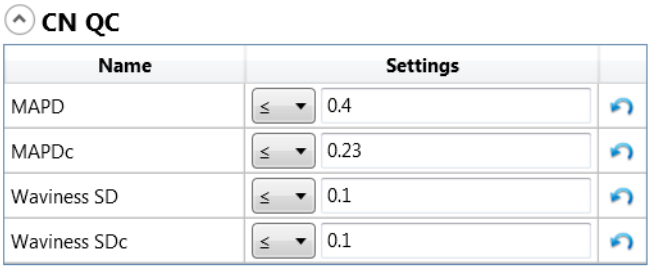
For information on typical Axiom array analysis and threshold fields, see "Setting up an analysis" on page 20 and "Setting up threshold settings" on page 28.

Figure 244 Example: Auto-populated Analysis and Threshold Setting panes



Note: For arrays that have post-plate based signal correction metrics enabled, MAPDc and WavinessSDc entries are added to the Threshold Settings pane's CN QC pane, as shown in Figure 245.

Figure 245 Added MAPDc and WavinessSDc entries



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.
- **Sex File**: A file specifying the desired Sex of every sample. If supplied, software will use values in this file instead of the computed sex. Sex impacts genotyping of chromosome X and Y SNPs. The Sex File is a two column tab-delimited text file, where the first row has the column names 'cel_files' and 'gender', and each row lists one filename and its sex (female, male, or unknown).
- **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 biallelic 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). **Note:** If the library package supports multiallele SNPs, then there are three additional model files for the three stages of multiallele genotyping: Multiallele Background Prior Model File, Multiallele Pairwise Prior Model File, and Multiallele 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.
- **Sex File**: A file specifying the known sex of every sample. If supplied, software will use values in this file instead of the computed sex. Sex 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.
- **ps2multisnp_map.ps**: 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 four tab delimited columns with the headers:
  - **probeset_id**
  - **multisnpid (snpid = affy_snp_id)**
  - **_snp_id (multiallelic probeset id)**
  - **ordered_allele**
- **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.
- **MAPDc**: MAPD calculated after plate based signal correction. This metric is used to determine whether the samples pass CN QC.
- **Waviness SD**: Waviness Standard Deviation of log2 ratio signals of copy number (CN) probesets must be below this value to make CN calls.
- **Waviness SDc:** WavinessSD calculated after plate based signal correction. This metric is used to determine whether the samples pass CN QC.

Note: For plates with control samples, MAPD and WavinessSD is calculated for the whole plate. Only controls that pass MAPD and WavinessSD are used for plate correction. Plate correction is then performed. After plate correction, MAPDc and WavinessSDc are calculated and used to determine CN passes QC. MAPD, MAPDc, WavinessSD and WavinessSDc are all shown in the Sample Table. See Figure 250.

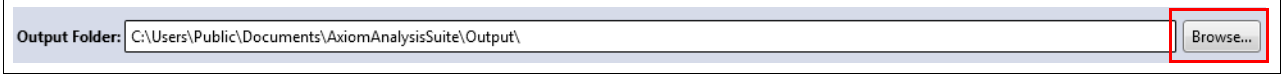
SNP QC

- Refer to Table 22 on page 216 for SNP QC Threshold name definitions.

Assigning an output folder path

1. Click the **Output Folder** path's **Browse** button. (Figure 246)

Figure 246 Output Folder field



An Explorer window appears.

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

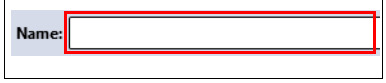
Your selected output folder path is now displayed.

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 247)

Figure 247 Enter a Batch Name



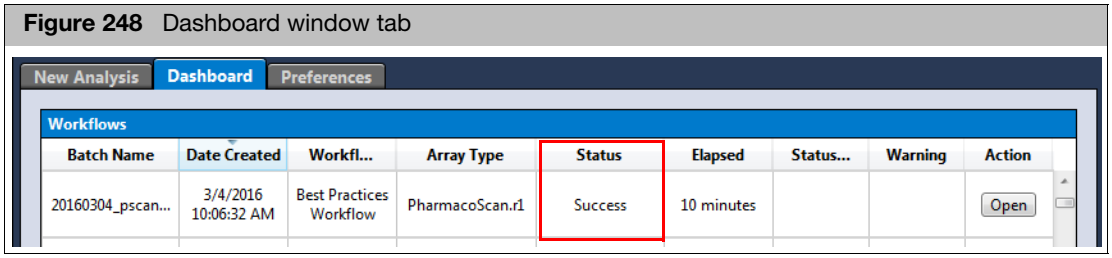
Note: A folder (with the same name as your entered batch name) is auto-generated during the analysis process. This folder includes all the necessary files needed to view your analysis results in the Viewer.

Running your genotyping with CN analysis

1. Click Run.

   The Dashboard window tab appears. (Figure 248)

Figure 248 Dashboard window tab



After **Success** is displayed in the Status column (Figure 248), 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 249) 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:

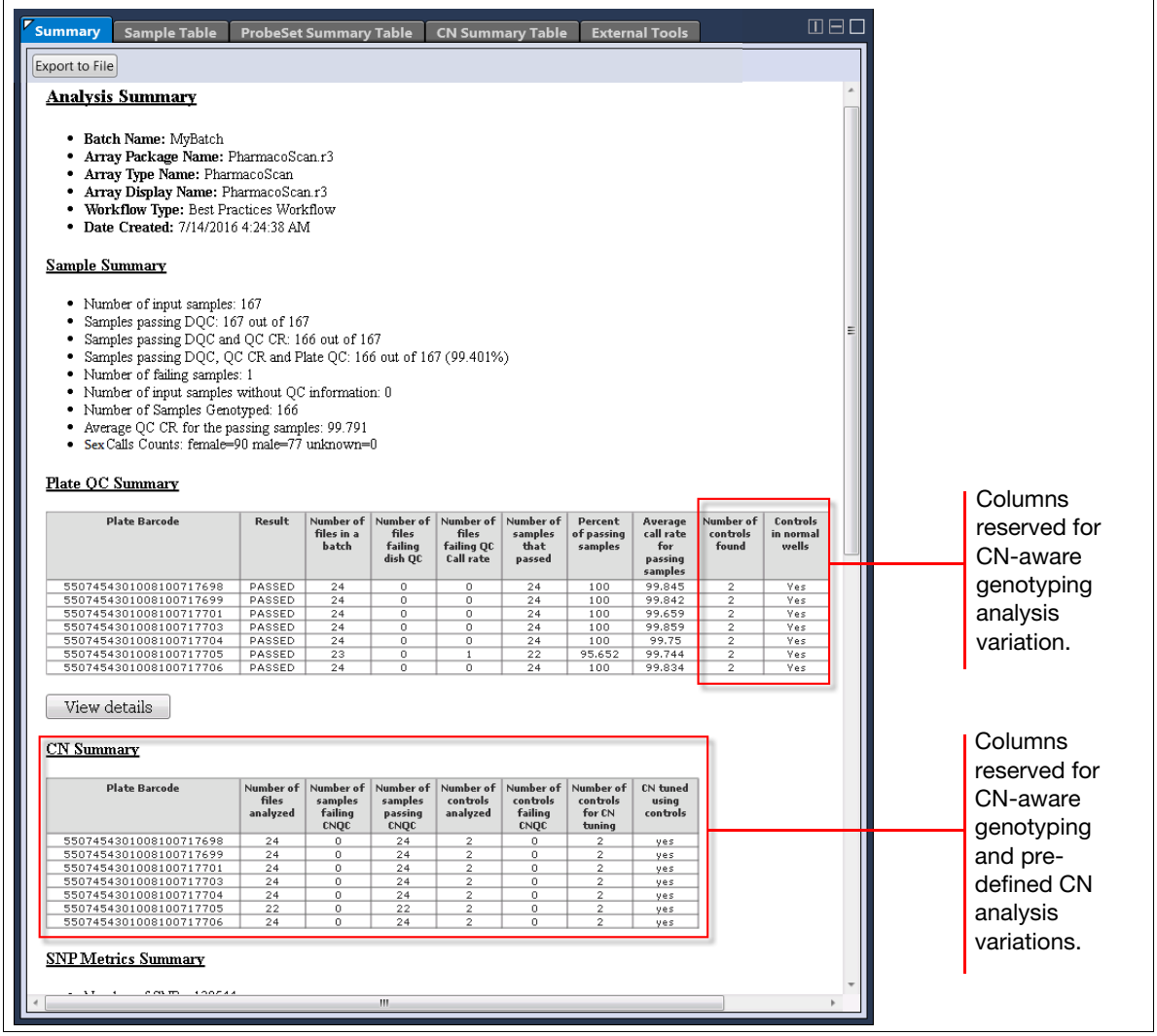
- New entries in the Summary Report, Sample Table, and ProbeSet Summary Table.
- For CN-aware genotyping, the Cluster Plot also displays hemizygous and ZeroCN clusters for SNPs in CN regions.
- CN Summary Table, with Export CN Data option.
- A CN Region Plot window tab.
- Exported VCF formats can now 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 249.

- **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 249 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 249.

Refer to Table 19 for descriptions of each CN Summary Report column.

Table 19 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.

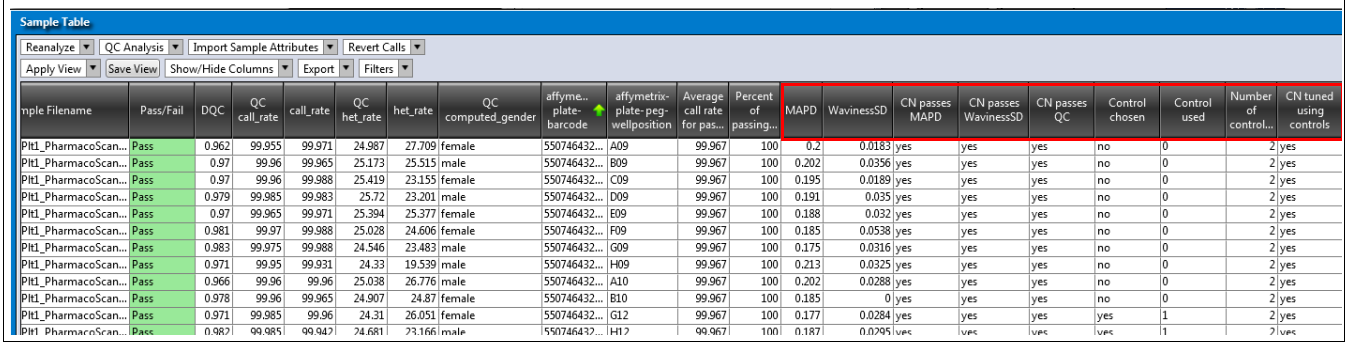
Table 19 CN Summary Report

CN Summary Report	Description
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 sex 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 250.

Figure 250 Sample Table after a CN Analysis run



nple Filename	Pass/Fail	DQC	QC call_rate	call_rate	QC het_rate	het_rate	QC computed_gender	affyme... plate-barcode	affymetrix-plate-peg-wellposition	Average call rate for pas...	Percent of passing...	MAPD	WavinessSD	CN passes MAPD	CN passes WavinessSD	CN passes QC	Control chosen	Control used	Number of control...	CN tuned using controls
Plt1_PharmacoScan...	Pass	0.962	99.955	99.971	24.987	27.709	female	550746432...	A09	99.967	100	0.2	0.0183	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.97	99.96	99.965	25.173	25.515	male	550746432...	B09	99.967	100	0.202	0.0356	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.97	99.96	99.988	25.419	23.155	female	550746432...	C09	99.967	100	0.195	0.0189	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.979	99.985	99.983	25.72	23.201	male	550746432...	D09	99.967	100	0.191	0.035	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.97	99.965	99.971	25.394	25.377	female	550746432...	E09	99.967	100	0.188	0.032	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.981	99.97	99.988	25.028	24.606	female	550746432...	F09	99.967	100	0.185	0.0538	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.983	99.975	99.988	24.546	23.483	male	550746432...	G09	99.967	100	0.175	0.0316	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.971	99.95	99.931	24.33	19.539	male	550746432...	H09	99.967	100	0.213	0.0325	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.966	99.96	99.96	25.038	26.776	male	550746432...	A10	99.967	100	0.202	0.0288	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.978	99.96	99.965	24.907	24.87	female	550746432...	B10	99.967	100	0.185	0	yes	yes	yes	no	0	2	yes
Plt1_PharmacoScan...	Pass	0.971	99.985	99.96	24.31	26.051	female	550746432...	G12	99.967	100	0.177	0.0284	yes	yes	yes	yes	1	2	yes
Plt1_PharmacoScan...	Pass	0.982	99.985	99.942	24.681	23.166	male	550746432...	H12	99.967	100	0.187	0.0295	yes	yes	yes	yes	1	2	yes

Refer to Table 20 for descriptions of each added Sample Table column.

Table 20 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.
MAPDc	This metric is used to determine CN passes MAPD and CN passes QC and is calculated after plate based signal correction.
WavinessSD	Waviness Standard Deviation of log2ratio signals of adjacent copy number (CN) probesets.

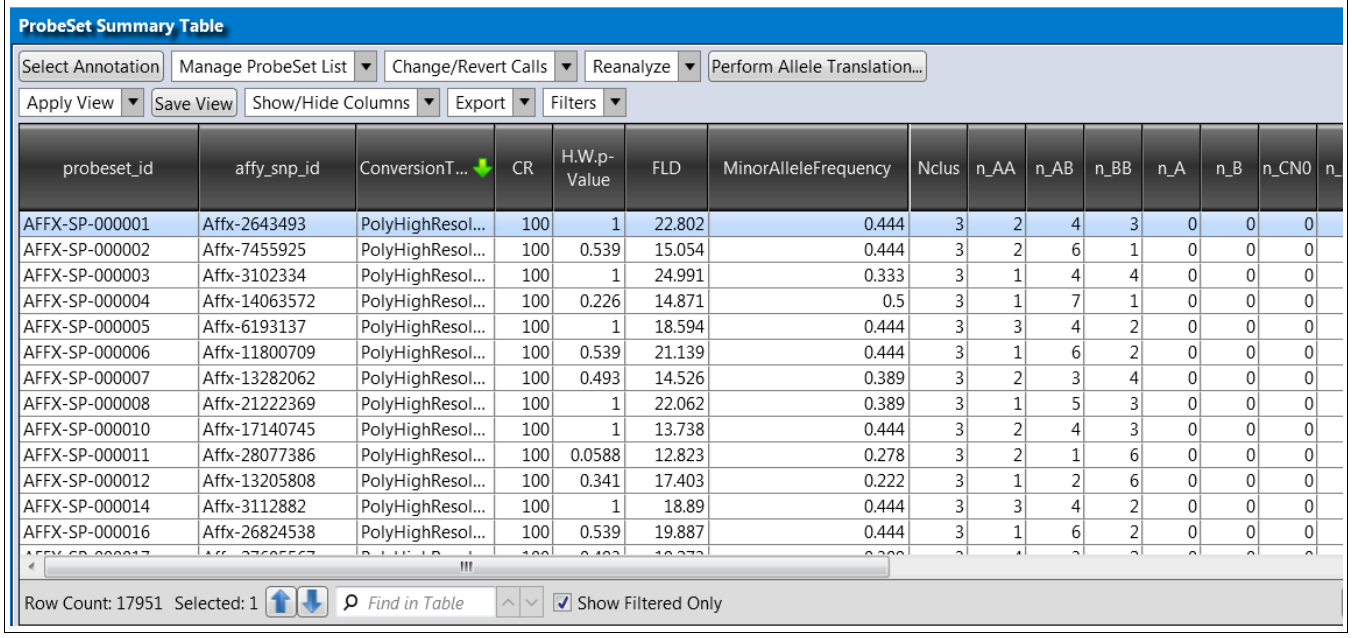
Table 20 Added Sample Table columns after running a CN Analysis

Sample Table	Description
WavinessSDc	This metric is used to determine CN passes WavinessSD and CN passes QC and is calculated after plate based signal correction.
CN passes MAPD	"Yes" if the sample's MAPDc value is not greater than the MAPDc threshold used by CN QC.
CN passes WavinessSD	"Yes" if the sample's WavinessSDc value is not greater than the WavinessSDc threshold used by CN QC.
CN passes QC	"Yes" if the sample passes both MAPDc and WavinessSDc 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:** MAPDc and WavinessSDc are used to determine whether the samples pass this metric. **Note:** For the CN-aware genotyping analysis variation, Non-PAR X, Y, and mitochondrial SNPs use sex 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.
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.

ProbeSet summary table

After successfully running a CN-aware genotyping analysis, additional columns appear within the ProbeSet Summary Table, as shown in Figure 251.

Figure 251 ProbeSet Summary Table after a CN Analysis run



probeset_id	affy_snp_id	ConversionT...	CR	H.W.p-Value	FLD	MinorAlleleFrequency	Nclus	n_AA	n_AB	n_BB	n_A	n_B	n_CN0	n_
AFFX-SP-000001	Affx-2643493	PolyHighResol...	100	1	22.802	0.444	3	2	4	3	0	0	0	
AFFX-SP-000002	Affx-7455925	PolyHighResol...	100	0.539	15.054	0.444	3	2	6	1	0	0	0	
AFFX-SP-000003	Affx-3102334	PolyHighResol...	100	1	24.991	0.333	3	1	4	4	0	0	0	
AFFX-SP-000004	Affx-14063572	PolyHighResol...	100	0.226	14.871	0.5	3	1	7	1	0	0	0	
AFFX-SP-000005	Affx-6193137	PolyHighResol...	100	1	18.594	0.444	3	3	4	2	0	0	0	
AFFX-SP-000006	Affx-11800709	PolyHighResol...	100	0.539	21.139	0.444	3	1	6	2	0	0	0	
AFFX-SP-000007	Affx-13282062	PolyHighResol...	100	0.493	14.526	0.389	3	2	3	4	0	0	0	
AFFX-SP-000008	Affx-21222369	PolyHighResol...	100	1	22.062	0.389	3	1	5	3	0	0	0	
AFFX-SP-000010	Affx-17140745	PolyHighResol...	100	1	13.738	0.444	3	2	4	3	0	0	0	
AFFX-SP-000011	Affx-28077386	PolyHighResol...	100	0.0588	12.823	0.278	3	2	1	6	0	0	0	
AFFX-SP-000012	Affx-13205808	PolyHighResol...	100	0.341	17.403	0.222	3	1	2	6	0	0	0	
AFFX-SP-000014	Affx-3112882	PolyHighResol...	100	1	18.89	0.444	3	3	4	2	0	0	0	
AFFX-SP-000016	Affx-26824538	PolyHighResol...	100	0.539	19.887	0.444	3	1	6	2	0	0	0	

Refer to Table 21 for descriptions of each added SNP Summary Table column.

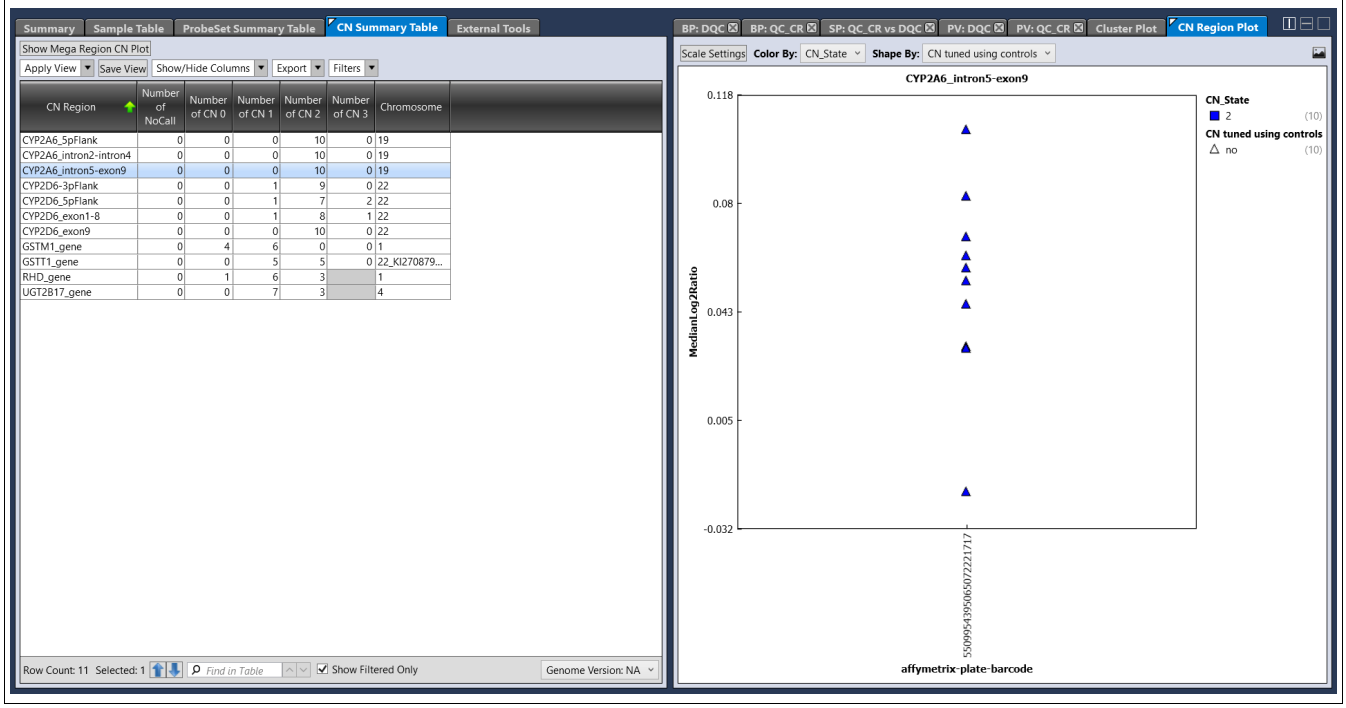
Table 21 Added ProbeSet Summary Table columns after CN-aware genotyping is performed

ProbeSet 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 252.

Figure 252 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 these 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, you have full control of which samples to genotype. Genotyping performs these 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

Dual workflow

IMPORTANT! The Dual Workflow feature requires an updated library file package. The Best Practices Workflow must be run using the updated library file package prior to the Dual Workflow analysis. To determine if a user has an updated library package or to request an updated library package, please contact the local support team.

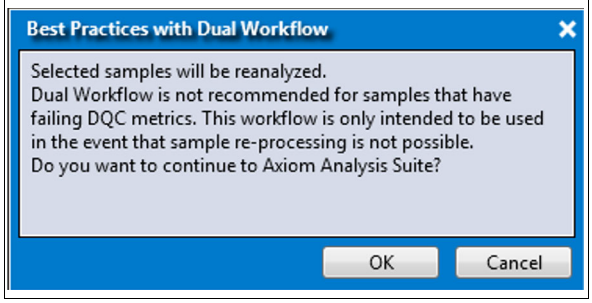
Dual workflow steps

Note: Make sure you execute the Best Practices Workflow. The Dual Workflow is intended for use with single-plate batches of samples but may also be applied to larger batches.

1. Open batch in Viewer.
2. In the Sample Table tab, select samples for the Dual Workflow, according to the following rules and procedures.
   - The expected use must be appropriate for the reduced accuracy of recovered sample genotypes.
   - The sample(s) were assayed on a plate that passes plate QC and has a minimum of 75% passing samples.
   - The samples have a passing DQC single-sample QC metric (usually $\geq$ 0.82).
   - The samples have a QC Call Rate below passing (usually 97%) but $\geq$ 90%
   - Highlight to select samples that pass cut offs listed above.
3. From the **Reanalyze** drop-down, select **Best Practices with Dual Workflow**.

   A Warning message appears. (Figure 253)

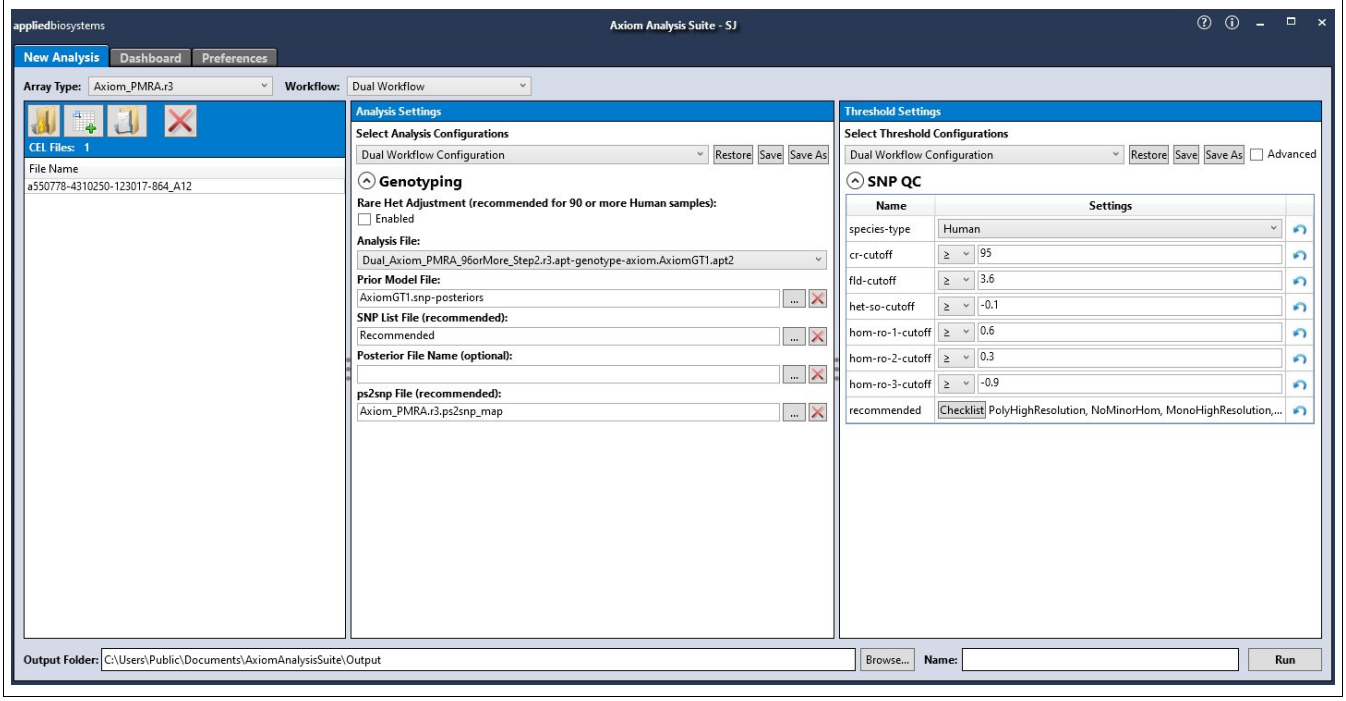
Figure 253 Warning message



4. Acknowledge the message, then click **OK**.

   A New Analysis window appears. (Figure 254)

Figure 254 New Analysis window tab



The Dual Workflow Analysis settings are now modified.

- The Dual Workflow Configuration is auto-selected and appears in the Analysis Setting's Genotyping section (Figure 254) with the following changes:
  - The Analysis File for Dual Workflow has changed settings: (a) single sample genotyping mode, (b) more stringent confidence threshold of 0.01.
  - The Prior Model File section is automatically populated to **AxiomGT1.snp-posteriors.txt** (the posteriors from the Best Practices workflow.
  - The SNP List File section is automatically populated with the **Recommended.ps** file from the Best Practices Workflow.

5. Name the batch, then click **Run**.

   The selected samples generate a separate batch that can be opened in the Viewer.

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.

AxAS enables conversion of genotype calls to relevant 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 255 and Figure 256) in the translation library file (***.translation**).

Figure 255 Biological annotations [Example: Gene table data for markers in CYP1A1]

Reference Link	Probe Set ID	Switch Design Strand to Report	dbSNP RS ID	Defining	cDNA Nucleotide Position	Genome Position	Change	Common Name
PMID: 11295847	AX-112162174	N	rs4646422	*13	134G>A	Ch15:75015305	G45D	CYP1A1*13_134G>A(G45D)
PMID: 11295847	AX-112262501	N	rs56313657	*6	993G>T	Ch15:75013804	M331I	CYP1A1*6_1635G>T(M331I)
PMID: 15618738	AX-112279125	N	rs72547510	*7	1275_1276insT	Ch15:75013093	E426Frameshift	CYP1A1*7_2345_2346insT
PMID: 15618738	AX-112162121	N	rs72547509	*8	1343T>A	Ch15:75013026	I448N	CYP1A1*8_2413T>A(I448N)
PMID: 8895751	AX-112162073	N	rs1799814	*4	1382C>A	Ch15:75012987	T461N	CYP1A1*4_2452C>A(T461N)
PMID: 9070254	AX-173402723	N	rs1048943	*2C	1384A>G	Ch15:75012985	I462V	CYP1A1*2C_2454A>G(I462V)
PMID: 11295847	AX-165878418	N	rs41279188	*5	1390C>A	Ch15:75012979	R464S	CYP1A1*5or*9_2460C>A>T(R464SorC)
PMID: 11295847	AX-165878418	N	rs41279188	*9	1390C>T	Ch15:75012979	R464C	CYP1A1*5or*9_2460C>A>T(R464SorC)
PMID: 15618738	AX-112262005	N	rs56240201	*10	1429C>T	Ch15:75012940	R477W	CYP1A1*10_2499C>T(R477W)
PMID: 15618738	AX-112161877	N	rs28399430	*11	1475C>G	Ch15:75012894	P492R	CYP1A1*11_2545C>G(P492R)
PMID: 15618738	AX-112161841	N	rs1800031	*3	*595T>C	Ch15:75012235	3'UTR	CYP1A1*3_3204T>C(3'UTR)
PMID: 21490707	AX-11393795	N	rs2470893	N	-1694G>A	Ch15:75019449	Promoter	CYP1A1_c.-1694G>A
PMID: 25348619	AX-31705745	N	rs2606345	N	-27+606G>T	Ch15:75017176	Intron	CYP1A1_c.-27+606G>T
PMID: 11295847	AX-165751549	N	rs34260157	R279G	835C>G	Ch15:75014049	R279G	CYP1A1_1390C>GorT(R279GorW)
PMID: 11295847	AX-165751549	N	rs34260157	R279W	835C>T	Ch15:75014049	R279W	CYP1A1_1390C>GorT(R279GorW)
PMID: 9353182	AX-83250267	N	rs4987133	I286T	857T>C	Ch15:75014027	I286T	CYP1A1_1412T>C(I286T)
PMID: 15618738	AX-11419619	N	rs2856833	F381L	1143C>A	Ch15:75013563	F381L	CYP1A1_1876C>A(F381L)
PMID: 18779756	AX-165874668	N	rs2278970	A463G	1388C>G	Ch15:75012981	A463G	CYP1A1_2458C>G(A463G)

Figure 256 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	*11	*13	R279G	R279W	I286T	F381L	A463G
CYP1A1*13_134G>A(G45D)	Y	C	T												T					
CYP1A1*6_1635G>T(M331I)	Y	C	A						A											
CYP1A1*7_2345_2346insT	Y	-	A							A										
CYP1A1*8_2413T>A(I448N)	Y	A	T								T									
CYP1A1*4_2452C>A(T461N)	Y	G	T				T													
CYP1A1*2C_2454A>G(I462V)	Y	T	C		C															
CYP1A1*5or*9_2460C>A>T(R464SorC)	Y	G	T					T												
CYP1A1*5or*9_2460C>A>T(R464SorC)	Y	G	A									A								
CYP1A1*10_2499C>T(R477W)	Y	G	A										A							
CYP1A1*11_2545C>G(P492R)	Y	G	C											C						
CYP1A1*3_3204T>C(3'UTR)	Y	A	G			G														
CYP1A1_c.-1694G>A	N	C	T																	
CYP1A1_c.-27+606G>T	N	C	A																	
CYP1A1_1390C>GorT(R279GorW)	N	G	C													C				
CYP1A1_1390C>GorT(R279GorW)	N	G	A														A			
CYP1A1_1412T>C(I286T)	N	A	G															G		
CYP1A1_1876C>A(F381L)	N	G	T																T	
CYP1A1_2458C>G(A463G)	N	G	C																	C

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_1412T>C(I286T), but the haplotype background of that variant is not available in the literature references used (for example, the Pharmacogene Variation Consortium; https://www.pharmvar.org).

Figure 255 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:

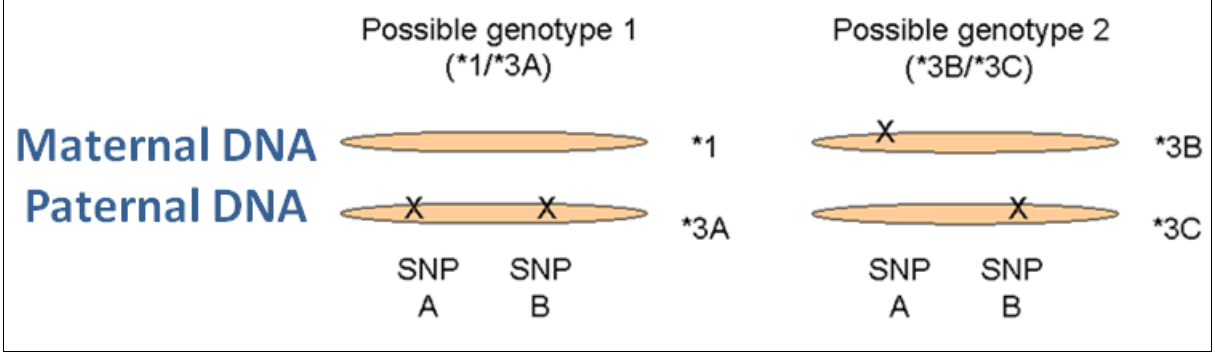
- Reference publications, sequences or dbSNP identifiers for following previously published information about the variant site
- Precise genomic location in a recent genome build for identifying confirmatory genotyping assays
- Notation of protein changes that may result from the mutations in the panel. 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)

Note: 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 257.

Figure 257 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 258 on page 206 lists the multiple possible calls due to phase ambiguity that were observed in six HapMap populations, and how often they occurred.

Figure 258 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%

Note: 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.

Note: 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 analysis batch 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.

Note: 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 AxAS, 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 function_1 function_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

Note: The **function_1** and **function_2** fields can be left blank. See Figure 259 on page 210 for field descriptions.

Figure 259 Field descriptions for the metabolizer library file

<table>
<tr><th>Metabolizer field</th><th>Description</th></tr>
<tr><td>gene</td><td>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.</td></tr>
<tr><td>allele_1

allele_2</td><td>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:
<table>
<tr><th>gene</th><th>allele_1</th><th>allele_2</th><th>phenotype</th></tr>
<tr><td>CYP2D6</td><td>*1</td><td>*2</td><td>EM</td></tr>
<tr><td>CYP2D6</td><td>*2</td><td>*1</td><td>EM</td></tr>
</table>
DMET Console will report an error IF it detects duplicate rows (as in the above example) AND IF the duplicate rows report disagreeing phenotypes.</td></tr>
<tr><td>phenotype</td><td>The value that should be reported for the associated 'gene allele_1/allele_2' call. It is recommended that the string be short with no commas, quotes, or whitespace characters. Common phenotype names are:
<table>
<tr><th>phenotype</th><th>definition</th></tr>
<tr><td>UM</td><td>Ultra-rapid metabolizer</td></tr>
<tr><td>EM</td><td>Extensive metabolizer</td></tr>
<tr><td>IM</td><td>Intermediate metabolizer</td></tr>
<tr><td>PM</td><td>Poor metabolizer</td></tr>
<tr><td>unknown</td><td>Unknown metabolizer</td></tr>
</table></td></tr>
<tr><td>function_1

function_2</td><td>The reported gene function level for an allele. Function_1 is for allele_1, and function_2 is for allele_2. The values in these fields are used to populate the Gene Function field in the phenotype report, e.g. 'normal/decreased'. If you leave these fields empty, the phenotype report will display '/' for the Gene Function.</td></tr>
<tr><td>Optional fields</td><td>DMET Console will ignore additional fields in the metabolizer file. Additional fields may be used to annotate each row.</td></tr>
</table>

Sources 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**
- Pharmacogene Variation Consortium:
  - **https://www.pharmvar.org**
- 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**

D

Sex calling

Overview

In Axiom Analysis Suite, the Sex calling algorithm used to populate the QC Computed Sex call in the Sample Table and the Geno_QCResults.txt is cn-probe-chrXY-ratio_sex method. Optimal genotyping of sex chromosome SNPs requires use of the correct model type, haploid or diploid. Haploid models are used for X and Y chromosome SNPs, when the Sex call is **male**, while diploid models are used for X chromosome SNPs, when the Sex call is **female**. A **No Call** is made for Y chromosome SNPs when the Sex call is female.

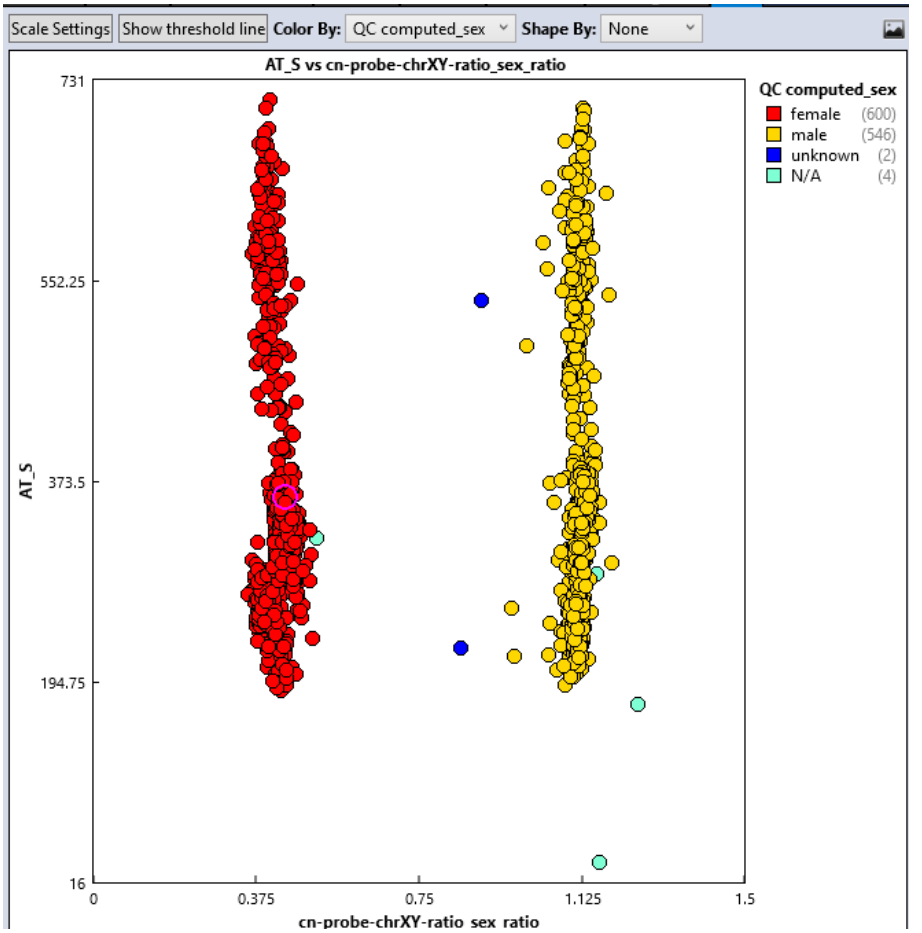
The cn-probe-chrXY-ratio_sex method determines sex based on the ratio of the average probe intensity of nonpolymorphic probes on the Y chromosome (cn-probe-chrXY-ratio_sex_meanY) to the average probe intensity of nonpolymorphic probes on the X chromosome (cn-probe-chrXY-ratio_sex_meanX). The probe intensities are raw and untransformed for these calculations and copy number probes within the pseudoautosomal regions (PAR region) of the X and Y chromosomes are excluded. For Axiom Human arrays, if the ratio is less than 0.65, the Sex call is female, and if it is greater than 0.95, the Sex call is male. If the ratio is between these values, the Sex call is unknown.

Sex calling and scatter plots

Scatter Plot of Signal vs. cn-probe-chrXY-ratio_sex_ratio

The Y axis is the signal of the AT channel, the X axis is **cn-probe-chrXY-ratio_sex_ratio**, as shown in Figure 260.

Figure 260 Scatter Plot of Signal vs. cn-probe-chrXY-ratio_sex_ratio example



Example of separation of gender in PMRA Best Practices Workflow batch. The female samples (red) are less than 0.65, while the male samples are above 1.0 (yellow). The unknown sex (blue) is in between the two threshold values.

The four sex columns can be displayed in the Sample Table. To do this, click the **Apply View** drop-down, then select **All Columns View**. The columns are:

- **cn-probe-chrXY-ratio_sex_meanX**: The average probe intensity (raw, untransformed) of X chromosome non polymorphic probes.
- **cn-probe-chrXY-ratio_sex_meanY**: The average probe intensity (raw, untransformed) of Y chromosome non polymorphic probes.
- **cn-probe-chrXY-ratio_sex_ratio**: Sex ratio Y/X = cn-probe-chrXY-ratio_sex_meanY/cn-probe-chrXY-ratio_sex_meanX
- **cn-probe-chrXY-ratio_sex**: Designated sex based on cn-probe-chrXY-ratio_sex_ratio

A Sample QC computed sex call will not be made if the sample fails DQC. The QC computed sex column will be blank. If a sample passes DQC but fails QC Call Rate, QC computed sex will be assigned. A sample must pass Sample QC to have a computed sex assigned.

Scatter plot of Signal vs. cn-probe-chrXY-ratio_sex_ratio

The Y axis is the signal of the AT channel, the X axis is cn-probe-chrXY-ratio_sex_ratio. Samples are colored by computed sex and shaped by cn-probe-chrXY-ratio-sex, as shown in Figure 261.

Figure 261 Scatter plot of Signal vs. cn-probe-chrXY-ratio_sex_ratio

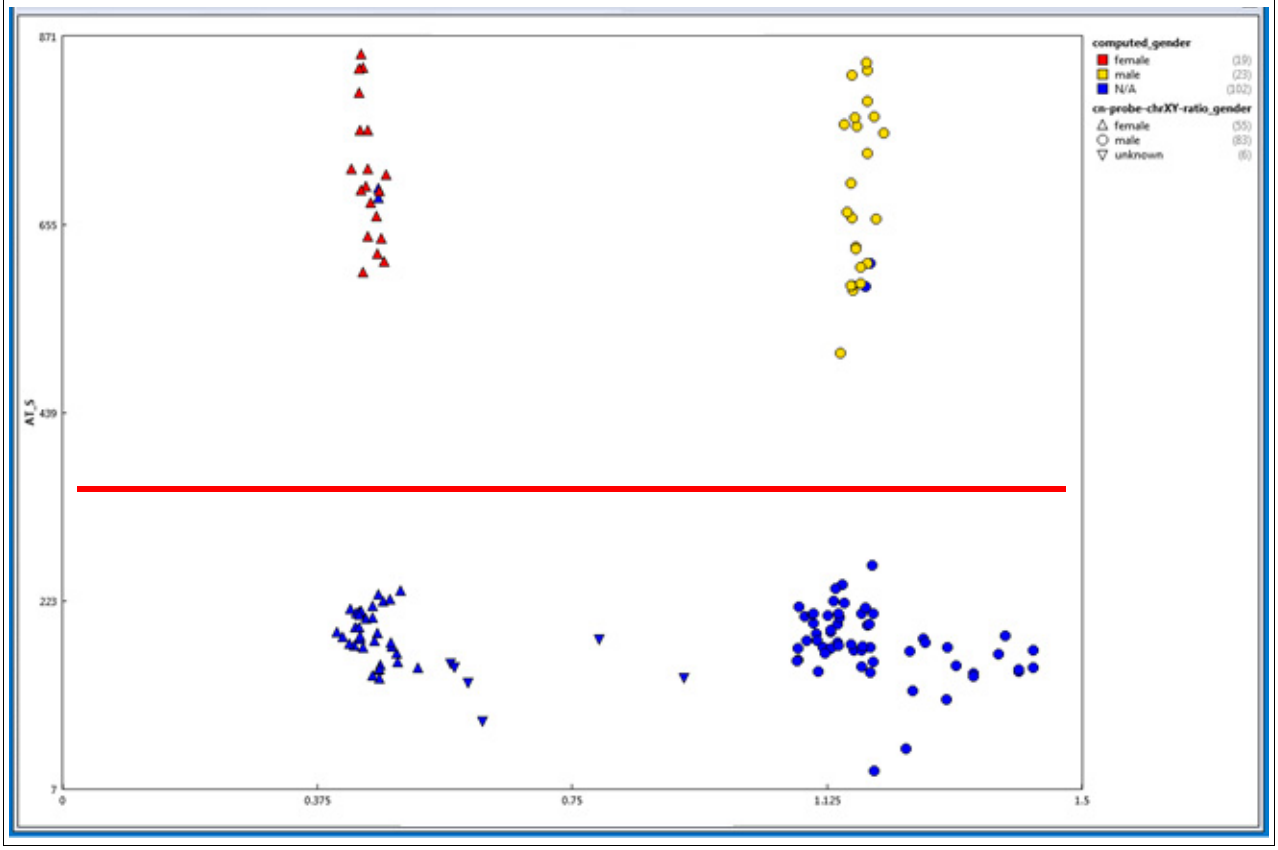
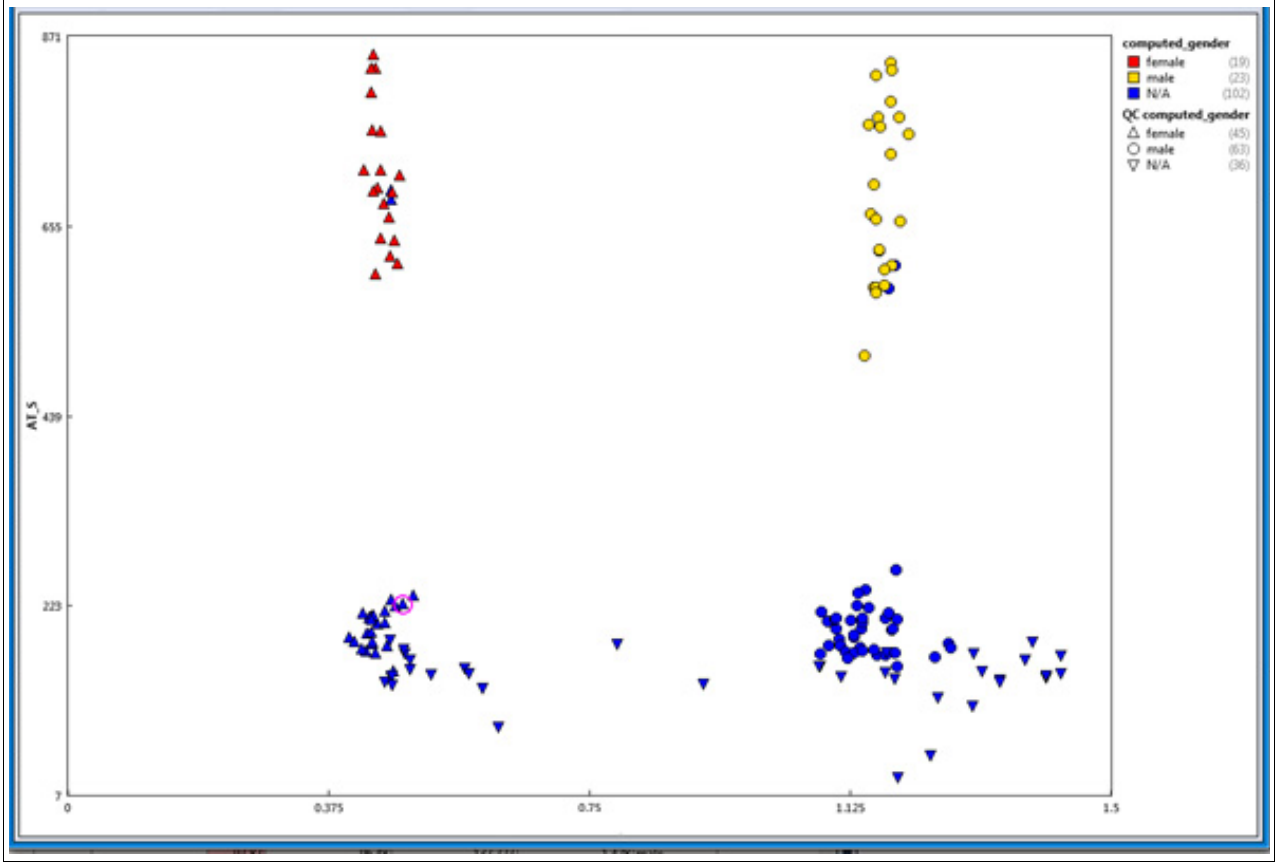


Figure 261 is an example sex calling of PMRA batch that contains samples that failed Sample QC and thus given N/A sex assignment. Samples displayed below red line have failed Sample QC and are not given computed sex call.

Scatter plot of signal vs. cn-probe-chrXY-ratio_sex_ratio for 1152 BovMDv3 samples

Female samples, in red have cn-probe-chrXY-ratio_sex_ratio of 1.1 or less. Male samples, in yellow, have cn-probe-chrXY-ratio_sex_ratio of 1.2 or more. Unknown samples are in blue as shown in Figure 262.

Figure 262 Scatter plot of signal vs. cn-probe-chrXY-ratio_sex_ratio for 1152 BovMDv3 samples example



E Definitions

Threshold names

Use the definitions in Table 22 to help select thresholds (when setting up an analysis or regenerating SNP metrics).

Table 22 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.
CN Region QC	seg-min-bases-CN-oneormore (*Not available in newer library packages*)	Minimum size in bases for segments that report copy number greater than 0.
CN Region QC	seg-min-bases-CN-zero	Minimum size in bases for segments that report copy number 0.
CN Region QC	seg-min-bases-CN-one	Minimum size in bases for segments that report copy number 1.
CN Region QC	seg-min-bases-CN-two	Minimum size in bases for segments that report copy number 2.
CN Region QC	seg-min-bases-CN-three	Minimum size in bases for segments that report copy number 3.

Table 22 Threshold names

Threshold Group	Threshold Name	Description
CN Region QC	seg-min-bases-CN-four or more	Minimum size in bases for segments that report copy number 4 or more.
CN Region QC	seg-min-probesets-CN-oneormore (*Not available in newer library packages*)	Minimum size in probesets for segments that report copy number greater than 0.
CN Region QC	seg-min-probesets-CN-zero	Minimum size in probesets for segments that report copy number 0.
CN Region QC	seg-min-probesets-CN-one	Minimum size in probesets for segments that report copy number 1.
CN Region QC	seg-min-probesets-CN-two	Minimum size in probesets for segments that report copy number 2.
CN Region QC	seg-min-probesets-CN-three	Minimum size in probesets for segments that report copy number 3.
CN Region QC	seg-min-probesets-CN-four or more	Minimum size in probesets for segments that report copy number 4 or more.
CN QC	MAPDc	MAPD calculated after plate based signal correction.
CN QC	WavinessSDc	WavinessSD calculated after plate based signal correction.
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.
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.

Table 22 Threshold names

Threshold Group	Threshold Name	Description
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	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-ZChr-cutoff	For avian Z 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-ZChr-cutoff	For avian Z 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.
Multiallelic SNP QC	HomMSBD-cutoff	For multiallelic probesets, the minimum acceptable difference between copy number 2 signals and background signals across all alleles for homozygous clusters. This metric measures how well the homozygous clusters separate from the background.
Multiallelic SNP QC	HomMSBD-1-cutoff	For multiallelic probesets, the minimum difference between copy number 2 signals and background signals for probesets with 1 allele.
Multiallelic SNP QC	HomMMA-cutoff	The minimum of the means for each multiallelic probeset allele's homozygous cluster.
Multiallelic SNP QC	FLD-MA-cutoff	For multiallelic probesets, the minimum acceptable average FLD value between all populated homozygous and heterozygous clusters.
Multiallelic SNP QC	FLD-MA-2-cutoff	For multiallelic probesets, minimum acceptable average FLD values between all populated homozygous and heterozygous clusters for probesets with only 2 alleles.
Multiallelic SNP QC	Min-FLD-MA-cutoff	For multiallelic probesets, the smallest acceptable minimum FLD value between all populated homozygous and heterozygous clusters.
Multiallelic SNP QC	Min-FLD-MA-2-cutoff	For multiallelic probesets, the smallest acceptable minimum FLD value between homozygous and heterozygous clusters for probesets with only 2 alleles.
Multiallelic SNP QC	HetSO-MA-2-cutoff	For multiallelic probesets, the minimum acceptable value for the Size Offset of heterozygous clusters for probesets with only 2 alleles.
Multiallelic SNP QC	HomRO-MA-cutoff	For multiallelic probesets,the minimum acceptable value for the Contrast (X position) of the homozygous clusters (Ratio Offset).

Table 22 Threshold names

Threshold Group	Threshold Name	Description
Multiallelic SNP QC	HomRO-MA-2-cutoff	For multiallelic probesets,the minimum acceptable value for the Contrast (X position) of the homozygous clusters (Ratio Offset) for probesets with 2 genotype clusters.
Multiallelic SNP QC	HomRO-MA-1-cutoff	For multiallele probesets, the minimum acceptable value for the Contrast (X position) of the homozygous clusters (Ratio Offset) when a probeset has 1 genotype cluster.
Multiallelic SNP QC	priority-order-MA	Priority order of probeset conversion types when performing probeset selection.
Multiallelic SNP QC	Best-CR-MA-cutoff	The minimum Call Rate for selecting a best and recommended probeset from multiple probesets for a single multi_snp_id.
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-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-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 22 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. Specifically, it only uses the female samples and ignores the male samples for this calculation.
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	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.
SNP QC	min-YChr-samples-cut	For Y probesets, minimum count of called samples of each sex needed for the Euclidean distance test between male and female genotype clusters.
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.

Table 22 Threshold names

Threshold Group	Threshold Name	Description
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.
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.
CN Reference QC	Sample Count	Minimum number of CEL files passing QC, for making a copy number reference.
CN Reference QC	CNQC Pass Rate	Minimum percentage of CEL files passing copy number QC thresholds, for the new copy number reference to pass QC.

Table 22 Threshold names

Threshold Group	Threshold Name	Description
CN Reference QC	Initial Matched Samples Rate	For Initial Copy Number Reference Creation, the minimum percentage CEL files matched to known samples.
CN Reference QC	Initial Matched Wells Rate	For Initial Copy Number Reference Creation, the minimum percentage of identified samples matched to expected plate location.
CN Reference QC	Initial Plate Count	For Initial Copy Number Reference Creation, the maximum number of plate barcodes.
CN Reference QC	Initial Scan Count	For Initial Copy Number Reference Creation, the maximum number of CEL files with the same barcode and well.

Sample table

Use the table of definitions in Table 23 for the fields used in the Sample Table.

Table 23 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.
Filtered call rate	Call Rate metric that is restricted to genotyped probesets that are in the batch-specific Recommended probeset list.
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_sex	Computed sex for the sample.
QC computed_sex	Computed sex for the sample (measured at the Sample QC step).
affymetrix-plate-barcode or 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).

Table 23 Sample table

Selection	Definition
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.
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.

Table 23 Sample table

Selection	Definition
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.
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_MQR	AT channel interquartile range (middle 50%) of background intensities, measured using GC control probes.
GC_B_MQR	GC channel interquartile range (middle 50%) of background intensities, measured using AT control probes.
AT_S_MQR	AT channel interquartile range (middle 50%) of signal intensities, measured using AT control probes.
GC_S_MQR	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.

Table 23 Sample table

Selection	Definition
AT_MLD	Median 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_MQR^2 + AT_B_MQR^2)].
GC_MLD	Median 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_MQR^2 + GC_B_MQR^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_sex_meanX	Average probe intensity (raw, untransformed) of X chromosome nonpolymorphic probes.
cn-probe-chrXY-ratio_sex_meanY	Average probe intensity (raw, untransformed) of Y chromosome nonpolymorphic probes.
cn-probe-chrXY-ratio_sex_ratio	Sex ratio Y/X = cn-probe-chrXY-ratio_sex_meanY / cn-probe-chrXY-ratio_sex_meanX.
cn-probe-chrXY-ratio_sex	Predicted sex based on the value of cn-probe-chrXY-ratio_sex_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 52.
Result	Sample quality control status.

Concordance columns

See the definitions in Table 24 for the columns used in the Concordance table.

Table 24 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 in Table 25 to help select the annotations that you can add to the Probeset Summary Table.

Table 25 Annotations and columns

Column Name	Description
Probeset ID	The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP probesets only).
Affx SNP ID	The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP). (SNP probesets 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 probesets 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.

Table 25 Annotations and columns

Column Name	Description
Strand Vs dbSNP	Indicates whether the SNP is on the same or reverse strand as compared to dbSNP (SNP probesets 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 Probeset.
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 probesets only).
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 probesets only).
Microsatellite	Describes the nearest microsatellite markers (upstream, downstream and overlapping) for the SNP.

Table 25 Annotations and columns

Column Name	Description
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) probesets as described by the Database of Genomic Variants (not available for Axiom Genome-Wide Human Array).
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 biallele probesets, the order is "Allele A // Allele B" For multiallele 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 biallele probesets the count is 2. For multiallele probesets, the count is 3 or more.

Probeset summary table definitions

Use the definitions in Table 26 to help select your Probeset Summary Table columns.

Most Probeset Summary Table metrics are calculated by sex depending on what the chromosomes are. What is displayed is restricted to the sex in the **sex_metrics** column. Refer to the sex_metrics description for more information.

A few metrics are calculated only for multiallele probesets. Most metrics are calculated only for biallele probesets.

Some of the metrics defined in the following table are not displayed by default in the Probeset 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 multiallelic probesets are present, or if CN-aware genotyping was performed.

Note: A few metrics are calculated only for multiallelic probesets. Most metrics are calculated only for biallele probesets.

Table 26 ProbeSet 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.
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 26 ProbeSet 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 chisquared.statistic	H.W.chisquared-statistic is NA if the HW p-value is calculated using the exact test, and it's the chi-squared test statistic value if the chi-squared test was used. Refer to H.W.p-Value definition (above) for more information.
minGenotypeFreqPval	If the ConversionType is 'UnexpectedGenotypeFreq' (unexpectedly-high genotype frequency) 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’.
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.

Table 26 ProbeSet summary table metrics

Column Name	Description
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 sex_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 sex_metrics.
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.
BB.meanY	Average Size (Y position) for BB cluster.
NC.meanX	Average contrast(X position) for No Calls.
NC.meanY	Average size for No Calls.
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 heterozygous cluster AB sample, adjusted to only comparing the sample pairs that have a smaller Size (Y position) distance than y_restrict parameter.

Table 26 ProbeSet summary table metrics

Column Name	Description
AA_dis_x_adj	Smallest Contrast (X position) distance between any AA cluster sample and heterozygous cluster AB 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.
sex_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 displayed using the sample set described in sex_metrics. If sex_metrics is 'all', all samples are used. The metrics are calculated by sex depending on the chromosomes. What is displayed is restricted to the sex in the sex_metrics column. For chromosomes MT and CP: all sexes, no splitting by sex. Chromosome X: all metrics on females, some metrics on males. Chromosome Y: all metrics on males, a small number of metrics on females. For avian species, chromosome Z: all metrics on males, some metrics on females and chromosome W: all metrics on females, a small number of metrics on males.
MinMean_cp_2	For multiallelic 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.

Table 26 ProbeSet summary table metrics

Column Name	Description
minSigBgndDiffHom	For multiallelic 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 multiallelic 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.
nSamples	Total number of samples (where non-diploid calls have been removed).
nCalls	The number of assigned calls.
count_ma_A	The number of alleles that appear per probeset for A allele.
count_ma_B	The number of alleles that appear per probeset for B allele.
count_ma_C	The number of alleles that appear per probeset for C allele.
nAllelesTested	The number of total possible alleles that a probeset can have.
nAllelesDetected	How many of the nallelesTesteed actually appeared in the assigned calls for that probeset. nAllelesDetected is always lower than nAllelesTested.
NHetClus	The number of heterzygous clusters.
nMajorAlleles	nMajorAlleles is the maximum of all allele counts (count_ma).
maxMinorAllele	maxMinorAllele is the second largest allele count. nMinorAlleles is the sum of all allele counts without nMajorAlleles.
nMinorAlles	nMinorAlleles is the sum of all allele counts without nMajorAlleles.
MAFall	MAFall is the minor allele frequency across all non-major alleles (the number of minor alleles is nMinorAlleles).
MAFmax	MAFmax is the minor allele frequency using the second largest allele count (the number of minor alleles is maxMinorAllele).
HomCount	HomCount is the number of samples in all homozygous clusters.
MajorHomCount	MajorHomCount is the number of samples in the largest homozygous cluster.
MinorHomCount	MinorHomCount is the number of samples in all homozygous clusters without MajorHomCount.
HetCount	HetCount is the number of samples in all heterozygous clusters.
HomMMA	HomMMA involve signal to background strength and use copy number to measure signal strength. The copy numbers are measured with respect to each allele: AA is copy number 2 for allele A (signal) but copy number 0 for allele D (background). HomMMA is the minimum of the means for each allele's homozygous cluster. For allele A, the mean value is mean A from the row for cluster AA. If a probeset has alleles A, B, and C, HomMMA is the minimum of AAmeanA, BBmeanB, and CCmeanC. The higher the HomMMA value, the stronger the allele signal is and the more defined the cluster is.

Table 26 ProbeSet summary table metrics

Column Name	Description
FLD_MA	Three multiallelic metrics are based on the biallelic metric FLD and use transformed variance values in their calculations: FLD_MA, MinFLD_MA, and HomFLD_MA. The posterior values are transformed from log2 space to size vs contrast space, and then the new FLD metrics are calculated using the new weighted variances. FLD_MA is the average FLD between all populated homozygous and heterozygous clusters across all biallelic pairs. Low FLD_MA values indicate low resolution clusters with little separation.
MinFLD_MA	MinFLD_MA is the same calculation as FLD_MA except the minimum value is taken instead of the average value. This is similar to the biallelic FLD calculation, which also uses the minimum.
HomFLD_MA	HomFLD_MA is similar to FLD_MA except it uses all pairwise combinations of populated homozygous clusters instead of homozygous-heterozygous pairs.
HetSO_MA	HetSO_MA is the minimum HetSO across all the biallelic pairs. A high HetSO_MA value indicates that the heterozygous clusters are well separated from the homozygous clusters and their strong signal is visible by their placement above the homozygous clusters.
HomRO_MA	HomRO_MA is the minimum of the absolute distance between the mean of a populated homozygous cluster and 0 on the X axis (contrast) across all the biallelic pairs. This is similar to taking the minimum of all biallelic HomRO values for every pairwise biallelic combination. HomRO_MA values should be much larger or smaller than 0, indicating that the location of the homozygous clusters are far from the origin and have stronger non-zero signal.
multiallelic	A '1' in this column indicates the probeset is multiallelic.
OriginalCT	An extra column called OriginalCT is included in the performance file when a 4-column ps2snp file is provided with biallelic and multiallelic content. If any problematic probesets are identified for a multiallelic SNP, then all probesets have their categories updated to Other and the original classification types are stored in OriginalCT.
Alternate Marker Name	This column lists Alternate marker name, if one has been supplied.
passBACTest	This column appears if a BAC test has been run and indicates Pass or Fail per probeset_id.