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. 703307

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>January 2020</td>
<td>5.0.1 release. Updated multiallelic SNP QC options in Threshold Settings window. Added filter for multiallelic markers.</td>
</tr>
<tr>
<td>9</td>
<td>November 2019</td>
<td>5.0 release. Added Whole Genome View feature and updated multiallelic SNP metrics.</td>
</tr>
<tr>
<td>8</td>
<td>June 2019</td>
<td>4.1 release. Copy Number QC Metrics have been updated to include two additional metrics.</td>
</tr>
<tr>
<td>7</td>
<td>September 2018</td>
<td>4.0.1 release. VCF export updates. Added non-human copy number variation analysis.</td>
</tr>
<tr>
<td>6</td>
<td>July 2018</td>
<td>4.0 release. Added CN Analysis capabilities.</td>
</tr>
<tr>
<td>5</td>
<td>August 2017</td>
<td>3.1 release. Added SNP QC call frequency check.</td>
</tr>
<tr>
<td>3</td>
<td>August 2016</td>
<td>2.0 release. Added PharmacoScan array and star allele translation.</td>
</tr>
<tr>
<td>2</td>
<td>July 2015</td>
<td>1.1 release. Added dynamic probeset selection genotyping if new v5 CDF have wobble probesets.</td>
</tr>
<tr>
<td>1</td>
<td>March 2015</td>
<td>Initial 1.0 release.</td>
</tr>
</tbody>
</table>

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.

NOTICE TO PURCHASER: DISCLAIMER OF LICENSE

Purchase of this software product alone does not imply any license under any process, instrument or other apparatus, system, composition, reagent or kit rights under patent claims owned or otherwise controlled by Life Technologies Corporation, either expressly, or by estoppel.

Legal entity

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

TRADEMARKS

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

©2019 Thermo Fisher Scientific Inc. All rights reserved.
Contents

CHAPTER 1 Introduction .............................................. 11

Overview .............................................................. 11
Software and hardware requirements ............................... 11
Sample data size estimates and required disk space ............... 12
Validated space requirements and run times for CN batches ....... 12
Installation ............................................................. 13
Starting ................................................................. 14
Preferences .............................................................. 16
General preferences .................................................. 16
Changing the default library folder/path ........................... 16
Setting up proxy server access ....................................... 17
Updating NetAffx library/annotations .............................. 17
Enabling/disabling check for library file updates .................. 18
Viewer preferences ..................................................... 19
Hide SNP Classification View ......................................... 19
Toggling the linking out to view RSID alert ......................... 21
Installing custom array library files .................................. 22
Uninstalling .............................................................. 22
Windows 7 .............................................................. 22
Windows 10 ............................................................ 22

CHAPTER 2 Performing an analysis .................................... 24

Setting up an analysis .................................................. 24
Selecting an array type .................................................. 24
Selecting a workflow .................................................... 24
Importing CEL files ..................................................... 25
Importing CEL files by text ............................................. 26
Import CEL by ARR ...................................................... 26
Removing selected CEL files .......................................... 26
Setting up an analysis configuration ................................. 27
Selecting an analysis configuration ................................. 27
Using the analysis settings fields ..................................... 29
General Analysis fields ................................................. 29
Sample QC fields ....................................................... 29
Genotyping fields ....................................................... 29
Setting up a multi-allelic enabled analysis ......................... 30
Saving your analysis configuration ................................. 31
CHAPTER 3 Summary window and sample table

Viewing options ......................................................... 44
Split-screen options .................................................. 44
Changing a tab window to a full screen ...................... 46
Sizing a window ......................................................... 47
Summary window tab .................................................. 48
Data analysis summary ................................................. 48
Viewing the plate barcode table details ......................... 49
Sample table ............................................................. 50
Importing sample attributes ....................................... 51
Column headers ........................................................ 51
Rearranging columns ................................................ 52
Sorting columns ......................................................... 52
Single-click sorting ................................................... 52
Hiding the column ..................................................... 52
Filtering column data ................................................ 53
Adding filters [method 1] ............................................. 53
Text-based columns ................................................. 53
Numeric data columns .............................................. 54
Showing filtered data only .......................................... 55
Clearing an individual filter ....................................... 56
Clearing all current filters ......................................... 57
Adding filters [method 2] ............................................. 57
Copying column data ................................................ 60
Setting user colors .................................................... 60
Assigning a color to a sample .................................... 60
Importing assigned colors ........................................ 61

Modifying an existing analysis configuration ..................... 32
Setting up threshold settings ....................................... 33
Customizing thresholds ............................................. 35
  Sample QC ............................................................... 35
  SNP QC ................................................................. 35
Assigning an output folder path .................................... 37
  Assigning a new output folder path ......................... 37
  Adding sub-folders ............................................... 37
Assigning a batch name ............................................ 38
Running your analysis ............................................... 39
Dashboard window tab .............................................. 41
  Open selected result(s) .......................................... 41
  Remove selected result(s) .................................... 42
  Viewing the results folder in Windows Explorer .......... 42
  Browsing for existing analysis results .................... 42
  Opening an analysis from Windows Explorer ............ 42

Adding filters (method 1) ............................................ 53
Setting user colors .................................................... 60
Copying column data ................................................ 60
Assigning a color to a sample .................................... 60
Importing assigned colors ........................................ 61
CHAPTER 4 Probeset summary table and cluster plot.

Probeset summary table ....................................................... 85
Multiallele probsets .......................................................... 86
Probeset summary table ....................................................... 88
Setting your Probeset summary table view ................................. 88
Adding and Removing Table Columns ...................................... 88
Selecting annotations .......................................................... 88
PCA tables and plot ......................................................... 123
PCA sample table ....................................................... 125
CN discovery segments table ........................................... 125
LOH segments table ..................................................... 126
Whole Genome View (WGV) ............................................. 127
Overview ................................................................. 127
Tool bar and viewing features ......................................... 128
WGV settings and options .............................................. 129
Graph settings ............................................................ 129
Saving a WGV ............................................................ 131
Printing a WGV .......................................................... 131
Showing ToolTips ........................................................ 131
Setting up a WGV ......................................................... 132
Frequency Histogram/Annotation (upper) pane ......................... 136
Sample data plots (lower) pane ......................................... 137
Customizing the sample data plots pane ............................... 138
Filtering Samples ......................................................... 139
Highlighting regions of interest ......................................... 139
Zooming into a highlighted region ..................................... 142
Viewing sample data plot segment information ........................ 143
Synchronizing the CN segment table and whole genome plot ........ 143
CN fixed regions batches ................................................ 145
Summary ................................................................. 145
Sample table .............................................................. 146
CN segments table ....................................................... 147
LOH segments table ..................................................... 148
CN summary table and CN region plot ................................ 150
Exporting a CN summary table ........................................ 151
Mega region CN plot ...................................................... 151
Mega region CN plot overview ......................................... 152
Exporting ................................................................. 153
Exporting to Integrative Genome Viewer (IGV) ....................... 153
Export Nexus format ..................................................... 154
Export VCF format ...................................................... 154

CHAPTER 6 Allele translation ............................................ 156

About translations ........................................................ 156
Performing allele translation ............................................. 156
Allele Translation Options .............................................. 158
CHAPTER 7 Exporting ......................................................... 170

Using the sample table export options .................................. 170
Using the ProbeSet summary table export options .................. 171
Exporting the current table ................................................. 171
Exporting all data .......................................................... 171
Exporting signature SNPs .................................................. 171
Exporting genotyping data .................................................. 171
  Result Output Formats .................................................. 172
  Filters ................................................................. 173
  Exporting genotyping data filters .................................. 173
  Call Output Formats .................................................. 173
  Exported Data Selections ............................................. 174
  ProbeSet List Filter .................................................... 174
  Output Location (Required) ......................................... 174
  Output Name (Required) ............................................. 175
  Changing the SNP Identifier ....................................... 175
  Changing the Current Annotation File (Optional) ............... 175
  Adding and Removing Annotation Columns ....................... 176
  Including ProbeSet Summary table columns in your genotyping report ........................................... 177
Exporting cluster plots to PDF. ......................................... 178

CHAPTER 8 External tools ..................................................... 180

Council on Dairy Cattle Breeding (CDCB) Export Tool .......... 180
Axiom Long Format Export Tool ....................................... 181
APPENDIX A  Predefined region CN analysis variations  . . . . . . . . . . . . 182

Predefined copy number analyses ............................................................................. 182
  With predefined CN genotyping .................................................................................. 183
  With CN-aware genotyping ......................................................................................... 183
Genotyping with CN analysis ....................................................................................... 184
  Setting up a genotyping with CN analysis ................................................................. 184
  Selecting a mode (Workflow) .................................................................................... 184
Importing CEL files ........................................................................................................ 184
Analysis settings with copy number options ............................................................... 187
  Sample QC .................................................................................................................. 187
  Genotyping ................................................................................................................. 187
Threshold configurations with CN options ................................................................. 189
  Sample QC .................................................................................................................. 189
  CN QC ......................................................................................................................... 189
  SNP QC ........................................................................................................................ 189
Assigning an output folder path .................................................................................. 189
Assigning a batch name ............................................................................................... 189
Running your genotyping with CN analysis ............................................................... 190
Viewing your genotyping and CN results ................................................................. 190
  Summary report ......................................................................................................... 191
  Sample table .............................................................................................................. 193
  ProbeSet summary table ............................................................................................ 194
  CN summary table and CN region plot ..................................................................... 195
  CN summary table [Overview] .................................................................................... 195
  CN region plot [Overview] ....................................................................................... 195
Overview and use of the best practices workflow ....................................................... 196

APPENDIX B  About allele translation ................................................................. 198

Overview ....................................................................................................................... 198
Gene table layout for haplotyping .............................................................................. 198
Biological annotations in translation reports ............................................................. 200
Impact of phase ambiguity in haplotyping ................................................................. 201
Diplootype to phenotype translation ............................................................................ 203
Creating a custom metabolizer library file ............................................................... 204
Metabolizer library file format .................................................................................... 205
Sources used in translation data curation .................................................................... 207
APPENDIX C Definitions ......................................................... 208
  Threshold names ..............................................................208
  Sample table .................................................................214
  Concordance columns .....................................................218
  Annotations and columns ................................................219
  Probeset summary table definitions ....................................222

APPENDIX D Gender calling ................................................... 229
  Overview .................................................................229
  Gender calling and scatter plots ......................................229
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>
<thead>
<tr>
<th>64-bit Operating System</th>
<th>Speed</th>
<th>Memory (RAM)</th>
<th>Available Disk Space(^1)</th>
<th>Web Browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Windows® 7 (64 bit) Professional with Service Pack 1</td>
<td>2.83 GHz Intel Pentium Quad Core Processor</td>
<td>32 GB</td>
<td>150 GB HD + data storage Refer to table below.</td>
<td>Internet Explorer 8.0 and higher</td>
</tr>
<tr>
<td>Microsoft Windows 10 (64 bit) Professional</td>
<td>2.83 GHz Intel Pentium Quad Core Processor</td>
<td>32 GB</td>
<td>150 GB HD + data storage Refer to table below.</td>
<td>Internet Explorer 8.0 and higher</td>
</tr>
</tbody>
</table>

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

Before using AxAS, make sure you have enough disk space. See the table below for size estimates.

**Note:** The estimates shown include the contents of the batch name folder.\(^2\)

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

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

\(^3\) Input is the storage size required for CEL files to be analyzed. Output is the storage size required for analysis results files.

Validated space requirements and run times for CN batches

**Table 1** 5000 UKBioBankv2 sample batch on a 32GB system

<table>
<thead>
<tr>
<th>Workflow</th>
<th>Space Required</th>
<th>Run Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy Number Fixed Regions</td>
<td>124GB</td>
<td>23 hours, 31 minutes</td>
</tr>
<tr>
<td>Copy Number Discovery</td>
<td>124GB</td>
<td>30 hours</td>
</tr>
<tr>
<td>Copy Number Reference Creation</td>
<td>260GB</td>
<td>19 hours</td>
</tr>
</tbody>
</table>

**Table 2** 5000 UKBioBankv2 sample batch on a 16GB system

<table>
<thead>
<tr>
<th>Workflow</th>
<th>Space Required</th>
<th>Run Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy Number Fixed Regions</td>
<td>124GB</td>
<td>55 hours</td>
</tr>
<tr>
<td>Copy Number Discovery</td>
<td>124GB</td>
<td>68 hours</td>
</tr>
<tr>
<td>Copy Number Reference Creation</td>
<td>260GB</td>
<td>42 hours</td>
</tr>
</tbody>
</table>
Installation

1. Go to thermofisher.com, then navigate to:
   Applications and Techniques → Life Sciences → Microarray Analysis → Software → Axiom Analysis Suite
2. Locate and download the zipped Axiom Analysis Suite software package.
3. Unzip the file, then double-click AxiomAnalysisSuiteSetup.exe.
4. Follow the on-screen instructions to complete the installation.
   **Note:** If your system has a previous version installed, an uninstall warning message appears. (Figure 1)

![Figure 1 Uninstall required message](image)

Acknowledge the message by clicking OK, then go to "Uninstalling" on page 22.
Starting

1. Double-click on the Desktop shortcut or click Start → All Programs → Thermo Fisher Scientific → Axiom Analysis Suite.
   The Select Profile window appears. (Figure 2)

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. For more information, see “Enabling/disabling check for library file updates” on page 18.
Figure 3  Main window

Click 'Import CEL Files' to import files for analysis.
Preferences

1. Click the Preferences tab.
The Preferences window tab appears. (Figure 4)

![Preferences window](image)

**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.

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

1. Click Browse (right of library path field).
The Select Library Folder window appears.
2. Navigate to 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 5**.

![Populated Library Path example](image)

**Figure 5** Populated Library Path example

**Library Folder: C:\Users\Public\Documents\AxiomAnalysisSuite\Library**

**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 6**)

![Proxy Settings Editor window](image)

**Figure 6** Proxy Settings Editor window

5. Click the **Enable Proxy Server Settings** check box (**Figure 6**), 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 User Login Information window appears. (**Figure 7**)

![NetAffx Login window](image)

**Figure 7** NetAffx Login window

2. Enter your NetAffx account email and password, then click **OK** or go to www.netaffx.com and click **Register** to sign up.
Note: If you are unable to connect to NetAffx, make sure you have an active Internet connection, and/or correct Proxy Server settings. The NetAffx Update window appears. (Figure 8)

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

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

4. Click OK.

   An Installing Updates progress bar appears.

Enabling/disabling check for library file updates

Each time you launch AxAS, the automatic Library File update alerts are enabled. To disable this feature, click the check box.
Viewer preferences  Hide SNP Classification View

- To hide the SNP classification view, click its check box. A Warning message appears.

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

To view 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.
The ProbeSet Summary Table will not show the Conversion Type column, as shown in Figure 11.
**Best ProbeSet** is the default filter when exporting Genotyping data, as shown in Figure 12.

![Figure 12 Export Genotyping Data window](image)

**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**

1. **Turn off Alert when linking out to view RSID**

   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.

**Windows 7**

1. Click **Start** → **Control Panel**.
   
   The Control Panel window appears.
2. Click the **View by** drop-down menu (upper-right), then click to select **Category**.
3. In the **Programs** category, click **Uninstall a program**.
   
   The Programs and Features window appears.
4. Click to select **Axiom Analysis Suite**, then click **Uninstall**.
5. Follow the on-screen instructions.
6. After the uninstall process is complete, close the Programs and Features window.
7. Use Windows Explorer as you normally would to navigate to the directory: `C:\Program Files\Affymetrix`
8. Verify that the **Axiom Analysis Suite** folder has been removed.
9. If the folder is present, double-click on it to open it.
10. Search for any files you want to keep, then move them to different (easily accessible) location.
11. Delete the **Axiom Analysis Suite** folder.
12. Close all open windows, then install the new version, as described in the "Installation" on page 13.

**Windows 10**

1. Click the Windows icon (bottom left corner).
2. Click **All apps** → **Windows System** → **Control Panel**.
   
   The Control Panel window appears.
3. In the **Programs** category, click **Uninstall a program**.
The Programs and Features window appears.

4. Click to select Axiom Analysis Suite, then click Uninstall.
5. Follow the on-screen instructions.
6. After the uninstall process is complete, close all open windows.
7. Use Windows Explorer as you normally would to navigate to the directory: C:\Program Files\Affymetrix
8. Verify that the Axiom Analysis Suite folder has been removed.
9. If the folder is present, double-click on it to open it.
10. Search for any files you want to keep, then move them to different (easily accessible) location.
11. Delete the Axiom Analysis Suite folder.
12. Close all open windows, then install the new version, as described in the "Installation" on page 13.
Performing an analysis

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

The New Analysis window and its three individual panes appear, as shown in Figure 13.

Setting up an analysis

For Array Types that are CN enabled, go to "Copy number analysis" on page 110 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 190.

Selecting an array type

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

Selecting a workflow

Note: 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.

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 (P/N 702961)

- Sample QC: This workflow performs the quality control analysis for samples and plates. Note this workflow does not produce genotype calls for the passing samples.

- Genotyping: This performs genotyping on the imported CEL files, regardless of the sample and plate QC metrics.
Note: Including samples that do not pass defined QC thresholds may reduce the quality of the results for passing samples.

- **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.

- **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 Gender check is performed at the beginning of the run.

- **Copy Number Fixed Regions**: Copy Number analysis is performed on targeted regions.

### Importing CEL files

Files can be imported in three ways:

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

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](image-url)
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.

3. Click **Open**.

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

### Figure 15  Text CEL file list example shown in Notepad

<table>
<thead>
<tr>
<th>Sample Filename</th>
<th>Alternate Sample Name</th>
<th>Gender</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>D:\AxAS Data\Axiom_Buffalo\LG392_D01.CEL</td>
<td>LGEV023724</td>
<td>female</td>
<td>Blood</td>
</tr>
<tr>
<td>D:\AxAS Data\Axiom_Buffalo\LG394_B06.CEL</td>
<td>LGEV023783</td>
<td>female</td>
<td>Blood</td>
</tr>
</tbody>
</table>

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 text 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. It is highly recommended you click the drop-down menu (Figure 16) and select the option that best matches the number of samples you want to analyze.

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

![Figure 16 Select an analysis configuration drop-down menu](image-url)
After selecting the appropriate default for the number of your samples, the Analysis Setting pane auto-populates. By default, the sections are collapsed. Click on each section’s down arrow button to view its content, as shown in Figure 17.

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

### General Analysis
- **Includ File**: Value (optional):  
  - Use value for all samples
- **Hints File (optional)**:  
  - Axiom_PM16_r6.generic_prior
- **Gender File (optional)**:  
  - Axiom_PM16_r6.lmm_regions

### Sample QC
- **Analysis File**:  
  - Axiom_PM16_r6SeqMore_Step1.r6.aapt-genotype-axiom.SnpSpecificPrior...
- **Prior Model File**:  
  - Axiom_PM16_r6.AxiomGT1
- **SNF List File (recommended)**:  
  - Axiom_PM16_r6.step1

### Genotyping
- **CN Analysis File**:  
  - Axiom_PM16_r6.aapt-copynumber-axiom.cnmnix.AxiomCNmix.aapt2
- **CN Reference Model File**:  
  - Axiom_PM16
- **CN Priors File**:  
  - Axiom_PM16_r6
- **Analysis File**:  
  - Axiom_PM16_r6SeqMore_Step2.r6.aapt-genotype-axiom.mmm.SnpSpecificPrior...
- **Prior Model File**:  
  - Axiom_PM16_r6.AxiomGT1
- **Multi-allelic Background Prior Model File**:  
  - Axiom_PM16_r6.AxiomGT1.mmb
- **Multi-allelic Pairwise Prior Model File**:  
  - Axiom_PM16_r6.AxiomGT1.mmp
- **Multi-allelic Prior Model File**:  
  - Axiom_PM16_r6.AxiomGT1.mm
- **SNF List File (recommended)**:  
  - Axiom_PM16_r6.step2
- **Posterior File Name (optional)**:  
  - ps2snp File (recommended):
    - Axiom_PM16_r6.ps2snp_map

**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.

**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.

- Gender file (Optional): Click the button, navigate to your Gender file, then click Open. The name of the file appears in the Gender File (optional) field. To remove the file, click the button.

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

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 Open.
   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 multi-allelic 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 an analysis configuration drop-down menu](image)

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

   ![Figure 19 Select an analysis configuration drop-down menu](image)

   **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.ps file, the click Open.
   The file is now displayed. (Figure 20)
Chapter 2 Performing an analysis

Saving your analysis configuration

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

2. Enter a different configuration name, then click \(\text{OK}\).
   Your new Analysis Configuration name is saved and available for use in the Select Analysis Configuration drop-down menu.
Modifying an existing analysis configuration

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

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

- Click the applicable File field’s **Browse** button to navigate to a different location, then click **Open** to reassign its path.
- If needed, click a File field’s **X** 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 will display the appropriate settings.

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 208.

![Figure 23 Automated QC Mode Threshold Settings pane example](image)

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.
### Chapter 2 Performing an analysis

#### Setting up threshold settings

**Figure 24  Threshold Settings pane - Advanced example**

<table>
<thead>
<tr>
<th>Threshold Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select Threshold Configurations</td>
</tr>
<tr>
<td>HumansSa (Default)</td>
</tr>
</tbody>
</table>

**Sample QC**

<table>
<thead>
<tr>
<th>Name</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOC</td>
<td>$\geq 0.82$</td>
</tr>
<tr>
<td>QC_callRate</td>
<td>$\geq 0.97$</td>
</tr>
<tr>
<td>Average call rate for passing samples</td>
<td>$\geq 0.95$</td>
</tr>
<tr>
<td>Percent of passing samples</td>
<td>$\geq 0.85$</td>
</tr>
</tbody>
</table>

**SNP QC**

<table>
<thead>
<tr>
<th>Name</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>species-type</td>
<td>Humans</td>
</tr>
<tr>
<td>cr-cutoff</td>
<td>$\geq 0.95$</td>
</tr>
<tr>
<td>f1d-cutoff</td>
<td>$\geq 3.6$</td>
</tr>
<tr>
<td>het-iso-cutoff</td>
<td>$\geq 0.1$</td>
</tr>
<tr>
<td>het-iso-XChr-cutoff</td>
<td>$\geq 0.1$</td>
</tr>
<tr>
<td>het-iso-aut-cutoff</td>
<td>$\geq 0.1$</td>
</tr>
<tr>
<td>hom-ro-1-cutoff</td>
<td>$\geq 0.65$</td>
</tr>
<tr>
<td>hom-ro-2-cutoff</td>
<td>$\geq 0.3$</td>
</tr>
<tr>
<td>hom-ro-3-cutoff</td>
<td>$\geq 0.9$</td>
</tr>
<tr>
<td>hom-ro</td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td>num-minor-allele-cutoff</td>
<td>$\geq 2$</td>
</tr>
<tr>
<td>hom-ro-hap-1-XChr-cutoff</td>
<td>$\geq 0.1$</td>
</tr>
<tr>
<td>hom-ro-hap-1-MChr-cutoff</td>
<td>$\geq 0.4$</td>
</tr>
<tr>
<td>hom-ro-hap-2-XChr-cutoff</td>
<td>$\geq 0.05$</td>
</tr>
<tr>
<td>hom-ro-hap-2-MChr-cutoff</td>
<td>$\geq 0.2$</td>
</tr>
<tr>
<td>asf-XChr-cut</td>
<td>$\leq 0.38$</td>
</tr>
<tr>
<td>f1d-XChr-cut</td>
<td>$\leq 4$</td>
</tr>
<tr>
<td>hom-fid-XChr-cut</td>
<td>$\leq 0.55$</td>
</tr>
<tr>
<td>hom-fid-YChr-cut</td>
<td>$\leq 0.55$</td>
</tr>
<tr>
<td>min-YChr-samples-cut</td>
<td>$\geq 3$</td>
</tr>
<tr>
<td>priority-order</td>
<td>Change List Order: PolyHighResolution, NoMinorHem, MonoHighResolution, QTV, Un...</td>
</tr>
<tr>
<td>recommended</td>
<td>Change List: PolyHighResolution, NoMinorHem, MonoHighResolution, Hemizygous</td>
</tr>
<tr>
<td>p-restrict</td>
<td>$\leq 0.2$</td>
</tr>
<tr>
<td>min-genotype-freq-samples</td>
<td>$\leq 20$</td>
</tr>
<tr>
<td>genotype-p-value-cutoff</td>
<td>$\leq 1E-06$</td>
</tr>
</tbody>
</table>

**Multiallelic SNP QC**

<table>
<thead>
<tr>
<th>Name</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>HomMA-cutoff</td>
<td>$\geq 10$</td>
</tr>
<tr>
<td>FLD-MA-cutoff</td>
<td>$\geq 5.2$</td>
</tr>
<tr>
<td>FLD-MA-2-cutoff</td>
<td>$\geq 3.2$</td>
</tr>
<tr>
<td>Min-FLD-MA-cutoff</td>
<td>$\geq 0$</td>
</tr>
<tr>
<td>Min-FLD-MA-2-cutoff</td>
<td>$\geq 0$</td>
</tr>
<tr>
<td>HetSD-MA-2-cutoff</td>
<td>$\geq 0.1$</td>
</tr>
<tr>
<td>HomRO-MA-cutoff</td>
<td>$\geq 0.5$</td>
</tr>
<tr>
<td>HomRO-MA-2-cutoff</td>
<td>$\geq 0.5$</td>
</tr>
<tr>
<td>HomRO-MA-1-cutoff</td>
<td>$\geq 1$</td>
</tr>
<tr>
<td>priority-order-MA</td>
<td>Change List Order: PolyHighResolution, NoMinorHem, MonoHighResolution, Hemizygous, U...</td>
</tr>
<tr>
<td>Best-CR-MA-cutoff</td>
<td>$\geq 90$</td>
</tr>
</tbody>
</table>
Chapter 2  Performing an analysis

Setting up threshold settings

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.

![Figure 25 Select Threshold Configurations](image)

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](image)

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](image)

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

3. Use the hom-ro and hom-het drop-down menus to change their True or False values.
4. Click inside the `num-minor-allele-cutoff` text field to enter a different value, as shown in Figure 28.

![Figure 28 SNP QC text fields](image)

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](image)

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`.
Chapter 2 Performing an analysis
Setting up threshold settings

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

Assigning a new output folder path

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

   ![Figure 31 Output Folder field](image)

   An Explorer window appears.

2. Navigate to the recommended path `C:\Users\Public\Documents\AxiomAnalysisSuite\Output`, then click **Select Folder**.
   Your selected output folder path is now displayed.

Adding sub-folders

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

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

1. Click the Output Folder’s **Browse** button to return to your assigned output path and/or folder.
2. In the Explorer window, click New Folder.
3. Enter a sub-folder name.
4. Click Select Folder.
   The newly created sub-folder now appears in the output result information window.
5. Repeat the above steps 1-4 to add more sub-folders, then click Select Folder.

Assigning a batch name

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

1. Enter a name in the 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](Image)

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

   ![Save Analysis Configuration prompt window](Figure 33)

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

   ![Save Analysis Configuration window](Figure 34)

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

   ![Save Threshold Settings prompt window](Figure 35)
• 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](image)

The Dashboard window tab appears (**Figure 37**) and shows the status of your running analysis. To cancel an analysis in progress, click **Stop**.

![Figure 37 Dashboard window/tab - Status bar and Stop button example](image)
Dashboard window tab

The Dashboard window tab displays existing results, as shown in Figure 38.

**Figure 38  Dashboard window**

<table>
<thead>
<tr>
<th>Workflows</th>
<th>Batch Name</th>
<th>Date Created</th>
<th>Workflow Type</th>
<th>Array Type</th>
<th>Status</th>
<th>Elapsed Time</th>
<th>Status Message</th>
<th>Warning</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyping_1</td>
<td>5/31/2018 09:54:51 AM</td>
<td>Genotyping</td>
<td>Axiom_UKB,</td>
<td>Success</td>
<td>24 minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample_QC_1</td>
<td>5/31/2018 09:54:31 AM</td>
<td>Sample QC</td>
<td>Axiom_UKB,</td>
<td>Success</td>
<td>48 minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SII</td>
<td>5/31/2018 08:15:27 AM</td>
<td>Summarized Signal Intensity</td>
<td>Axiom_UKB,</td>
<td>Success</td>
<td>26 minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN_RC</td>
<td>5/31/2018 08:34:51 AM</td>
<td>Copy, Number Reference Regions</td>
<td>Axiom_UKB,</td>
<td>Failed</td>
<td>4 minutes Failed to open 'L:\Axiom 544 Library and Calf Files 'CEL' file Axiom UKB Bank N42175S.CEL' for reading.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN_FR</td>
<td>5/31/2018 08:33:53 AM</td>
<td>Copy, Number Fixed Regions</td>
<td>Axiom_UKB,</td>
<td>Success</td>
<td>26 minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN_L</td>
<td>5/31/2018 08:20:00 AM</td>
<td>Copy, Number Discovery</td>
<td>Axiom_UKB,</td>
<td>Success</td>
<td>25 minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>File_3</td>
<td>5/30/2018 00:12:05 PM</td>
<td>Best Practices Workflow</td>
<td>Axiom_UKB,</td>
<td>Success</td>
<td>20 minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Open selected result(s)

*Do one of the following to open a selected result:*

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

  **Note:** The **Open** button is relabeled **Start** if the Dashboard’s Status column displays Failed. A workflow can fail if there is a processing error (for example, needed files are unavailable) or if no samples pass QC. If the Warning column indicates that no samples passed QC, yet you still want to open the study to review its QC results, use one of the alternate methods below to open a selected study.

- Single-click on a study, then click **Open Selected Result(s).**
- Double-click on a study to open it.
- Right-click on a study, then click **Open.**

  **Note:** Click on any of the Workflow header columns to sort your listed results. After a few moments, the Viewer opens and displays your study (as you last left it).
Remove selected result(s)  

Do one of the following to remove a selected result:
- Single-click to highlight the analysis you want to remove, then click **Remove Selected Result(s)**.
- Right-click on the highlighted analysis, then click **Remove from List**.

Viewing the results folder in Windows Explorer

1. Right-click on the highlighted analysis, then click **Open in Windows Explorer**. Your Analysis Results folder now appears in the Explorer window.
   **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).

**Do 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 window tab. To do this:

1. Right-click on an analysis batch folder and select **Open in Axiom Analysis Suite Viewer**.
   A Select Profile window appears.
2. At the Select Profile window, enter a new or select an existing profile name.
   If your selected folder is a valid analysis folder, the application’s Viewer opens as it normally would.
After setting up and successfully running an analysis, as described in Chapter 2, the Axiom Analysis Suite Viewer opens. (Figure 39)

**Summary window and sample table**

**Figure 39   Main Viewer window**

- **Analysis Summary**
  - Batch Name: Pete_1
  - Array Package Name: Affymetrix_GT_Biotype
  - Array Type Name: Affymetrix_GT
  - Workflow Name: Best Practices Workflow
  - Date Created: 5/30/2018 2:12:41 PM

- **Sample Summary**
  - Number of input samples: 24
  - Arrays passing QC: 24 out of 24
  - Samples passing QC and QC CR: 24 out of 24
  - Samples passing QC, QC CR and Plate QC: 24 out of 24 (100%)
  - Number of failing samples: 0
  - Number of input samples without QC information: 0
  - Number of Samples Genotyped: 24
  - Average QC CR for the passing samples: 99.34
  - Gender Call Coverage: females=13, males=11, unknown=0
  - Hombre Penalty Applied: no

- **Plate QC Summary**

- **ProbeSet Metrics Summary**
  - Number of ProbeSets: 826733
Viewing options

As shown in Figure 39 on page 43, the Viewer (by default) displays a side-by-side split-screen configuration.

Split-screen options

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

1. Click the Horizontal Split icon. (Figure 40)

To disable the split-screen:

1. Click the Disable Split-Screen icon. (Figure 41)

The split-screen becomes a single window, as shown in Figure 41.

Click on any window tab to view it in full/single window mode.
Chapter 3 Summary window and sample table

Viewing options

Figure 41 Full window view example

Analysis Summary

- Batch Name: Pro_1
- Array Package Name: Axiom_UKB_WCSG.r3
- Array Type Name: Axiom_UKB_WCSG
- Array Display Name: Axiom_UKB_WCSG.r3
- Workflow Type: Best Practices Workflow
- Date Created: 5/30/2018 2:12:41 PM

Sample Summary

- Number of input samples: 24
- Samples passing QC: 24 out of 24
- Samples passing QC and QC CR: 24 out of 24
- Samples passing QC, QC CR and Plate QC: 24 out of 24 (100%)
- Number of calling samples: 0
- Number of input samples without QC information: 0
- Number of Samples Genotyped: 24
- Average QC CR for the passing samples: 0.143
- Gender Call Counts: female=12 male=12 unknown=0
- Inbred Penalty Applied: no

Plate QC Summary

<table>
<thead>
<tr>
<th>Plate barcode</th>
<th>Result</th>
<th>Number of files in a batch</th>
<th>Number of files failing QC</th>
<th>Number of files failing QC Call rate</th>
<th>Number of samples passing</th>
<th>Percent of passing samples</th>
<th>Average QC Call rate for passing samples</th>
<th>Filtered Call Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>plate1</td>
<td>PASS</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>100</td>
<td>0.8832</td>
<td>100</td>
</tr>
</tbody>
</table>

View details

Probeset Metrics Summary

- Number of Probesets: 256733

<table>
<thead>
<tr>
<th>Probeset Type</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhouse</td>
<td>256733</td>
<td>100</td>
</tr>
<tr>
<td>millennium</td>
<td>256733</td>
<td>100</td>
</tr>
<tr>
<td>Inhouse millennium</td>
<td>256733</td>
<td>100</td>
</tr>
</tbody>
</table>
Chapter 3 Summary window and sample table

Viewing options

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

1. Click the Vertical Split icon. (Figure 42)

Changing a tab window to a full screen

To toggle a tab window to full screen:

1. Locate the tab you want to make full screen.
2. Click on a tab’s white triangle graphic. (Figure 43)

The window tab is now a window.

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

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

1. Double-click anywhere along the top of the window.
2. Click the X button (top right) to close the window.
   The window returns to its default tab window and position.

Sizing a window

To change the size of a window pane:

1. Click, hold, then drag the edge of the window pane (Figure 44) to resize it.
Summary window tab

The Summary window/tab (Figure 45) displays a summary snapshot of your analysis, including detailed threshold values, and tables based on your analysis.

### Data analysis summary

**Note:** Each workflow type reports different information within the Analysis Summary window. Figure 45 is an example of a Best Practices workflow.

<table>
<thead>
<tr>
<th>Analysis Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch Name: Pete_1</td>
</tr>
<tr>
<td>Array Package Name: Axiom_UKB_WCSG.13</td>
</tr>
<tr>
<td>Array Type Name: Axiom_UKB_WCSG</td>
</tr>
<tr>
<td>Array Display Name: Axiom_UKB_WCSG.13</td>
</tr>
<tr>
<td>Workflow Type: Best Practices Workflow</td>
</tr>
<tr>
<td>Date Created: 2:30-3:00 PM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of input samples: 24</td>
</tr>
<tr>
<td>Samples passing DQC: 24 out of 24</td>
</tr>
<tr>
<td>Samples passing DQC and QC CR: 24 out of 24</td>
</tr>
<tr>
<td>Samples passing DQC, QC CR and Plate QC: 24 out of 24 (100%)</td>
</tr>
<tr>
<td>Number of failing samples: 0</td>
</tr>
<tr>
<td>Number of input samples without QC information: 0</td>
</tr>
<tr>
<td>Number of Samples Genotyped: 24</td>
</tr>
<tr>
<td>Average QC CR for the passing samples: 98.143</td>
</tr>
<tr>
<td>Gender Calls Counts: female=13 male=11 unknown=0</td>
</tr>
<tr>
<td>Inbred Penalty Applied: no</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plate QC Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plate Barcode</strong></td>
</tr>
<tr>
<td><strong>Result</strong></td>
</tr>
<tr>
<td><strong>Number of samples in batch</strong></td>
</tr>
<tr>
<td><strong>Number of samples failing DQC</strong></td>
</tr>
<tr>
<td><strong>Number of samples failing QC CR</strong></td>
</tr>
<tr>
<td><strong>Number of samples that passed</strong></td>
</tr>
<tr>
<td><strong>Percent of samples that passed</strong></td>
</tr>
<tr>
<td>S00424102426224103153</td>
</tr>
</tbody>
</table>

| View Details |

<table>
<thead>
<tr>
<th>Probeset Metrics Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Probesets: 82,073</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification Type</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>31,860</td>
<td>39.645</td>
</tr>
<tr>
<td>Deviation</td>
<td>21,232</td>
<td>26.525</td>
</tr>
<tr>
<td>MeanHighScore</td>
<td>21,797</td>
<td>26.354</td>
</tr>
<tr>
<td>GenotypeHighScore</td>
<td>4,432</td>
<td>5.364</td>
</tr>
<tr>
<td>Other</td>
<td>5075</td>
<td>6.226</td>
</tr>
<tr>
<td>GTN</td>
<td>1325</td>
<td>1.618</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker Metrics Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Markers: 81,883</td>
</tr>
<tr>
<td>Number of BestandRecommended: 74,063</td>
</tr>
<tr>
<td>Percent BestandRecommended: 90.998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification Type</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>31,860</td>
<td>39.158</td>
</tr>
<tr>
<td>Deviation</td>
<td>21,232</td>
<td>26.540</td>
</tr>
<tr>
<td>MeanHighScore</td>
<td>21,797</td>
<td>26.050</td>
</tr>
<tr>
<td>GenotypeHighScore</td>
<td>4,432</td>
<td>5.393</td>
</tr>
<tr>
<td>Other</td>
<td>5075</td>
<td>6.186</td>
</tr>
<tr>
<td>GTN</td>
<td>1325</td>
<td>1.618</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample QC Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQC ≥ 6.82</td>
</tr>
</tbody>
</table>

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

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

**Sample Summary:** Breaks down the sample QC for your analysis run and displays the number that pass each of your QC Thresholds. In addition, it provides the average QC Call Rate (CR) and breakdown of the genders found within your batch of samples, as well as whether an Inbred Penalty Score was applied.

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

**View Details:** See “Viewing the plate barcode table details” on page 49.

**Probeset Metrics Summary:** This section contains a summary of the categorization of the probesets in the analysis by PS_Classification. For more information on these categories see “Regenerate probeset metrics” on page 92.

**Marker Metrics Summary:** This summary has a breakdown of the number of markers that are in BestandRecommended category, percentage, and conversion type.

**Sample QC Thresholds:** Displays the Sample QC Thresholds used for your analysis run and their associated SNP QC Metrics.
Chapter 3  Summary window and sample table

Summary window tab

Viewing the plate barcode table details

1. In the Summary window tab, click [View Details]. (Figure 45)
   A window opens and displays a text file version of your Sample QC information (by plate). (Figure 46)

Figure 46  Notepad window

```
Plate QC Details

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

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%
QC summary for plate with barcode: 5504314179819022514392
Result: PASSED
--- Number of files in a batch 3
--- plate_num_of_files_failed_dishqc 0
--- plate_num_of_files_failed_qc_call_rate 0
--- Percent of passing samples 100%
--- Average call rate for passing samples 99.502

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%
QC summary for plate with barcode: 5504314179819022514394
Result: PASSED
--- Number of files in a batch 2
--- plate_num_of_files_failed_dishqc 0
--- plate_num_of_files_failed_qc_call_rate 0
--- Percent of passing samples 100%
--- Average call rate for passing samples 99.562

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%
QC summary for plate with barcode: 5504314179819022514395
Result: PASSED
--- Number of files in a batch 2
```
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 47.

- Click the Genome Version drop-down (bottom right) to select a different genome version.
- Use the Find in Table field to locate keywords. See "Searching keywords" on page 64.
- 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

To import sample attributes into your Sample Table:

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

   An Explorer window appears.

**IMPORTANT!** Your text-based CEL file must start with the header Sample Filename or cel_files and include the full CEL file name, as shown in Figure 48.

Figure 48  Tab-delimited text CEL file example shown in Excel

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample Filename</td>
<td>String Name</td>
</tr>
<tr>
<td>2</td>
<td>NA18901_AxiomGWhuSNP1_20101007_TrainingMix_R3stab-Di1a-Omo_509.CEL</td>
<td>A1</td>
</tr>
<tr>
<td>3</td>
<td>NA12701_AxiomGWhuSNP1_20101007_TrainingMix_R3stab-Di1a-Omo_506.CEL</td>
<td>A2</td>
</tr>
<tr>
<td>4</td>
<td>NA19160_AxiomGWhuSNP1_20100908_ExampleMix_CEL</td>
<td>A3</td>
</tr>
<tr>
<td>5</td>
<td>NA19560_AxiomGWhuSNP1_20100107_TrainingMix_R3stab-Di1a-Omo_502.CEL</td>
<td>A4</td>
</tr>
<tr>
<td>6</td>
<td>NA12801_AxiomGWhuSNP1_20101007_TrainingMix_R3stab-Di1a-Omo_505.CEL</td>
<td>A5</td>
</tr>
<tr>
<td>7</td>
<td>NA19210_AxiomGWhuSNP1_20101007_TrainingMix_R3stab-Di1a-Omo_512.CEL</td>
<td>A6</td>
</tr>
</tbody>
</table>

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 49)

Figure 49  Default Sample Table Columns

<table>
<thead>
<tr>
<th>Sample Filename</th>
<th>Pass/Fail</th>
<th>QC call rate</th>
<th>cell_rate</th>
<th>QC net rate</th>
<th>het rate</th>
<th>QC computed_gender</th>
<th>affymetrix plate barcode</th>
<th>affymetrix plate prep wash position</th>
<th>Average call rate for passing</th>
<th>Percent of passing</th>
</tr>
</thead>
</table>

To show or hide table columns:

1. Click the **Show/Hide Columns** drop-down menu.
2. Click each available column name’s check box to show it or remove it from the table. See "Annotations and columns" on page 219 for their definitions.
3. Click outside the **Show/Hide Columns** drop-down menu to close it.

To save your customized Sample Table column view:

1. Click **Save View**.
Chapter 3  Summary window and sample table

Sample table

The Save Current View window appears. (Figure 50)

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

To show ALL available columns within the Sample Table:
1. Click the Apply View drop-down menu, then select All Columns View.

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

Sorting columns
1. Select a column, then right-click on it.
   The column menu appears. (Figure 51)
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 52)*

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 53.*

*Figure 52  Column Menu*

*Figure 53  Contains drop-down*

**To apply 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 54.*
3. (Optional) Click \( \text{filter symbol} \) to add additional filters.

4. Click the \textbf{Or} or \textbf{And} radio button to choose \textbf{Or} or \textbf{AND} relationship logic. (Figure 55)

5. Repeat steps 1-4 as needed.

6. To remove a filter(s), click \( \text{filter symbol} \).

\textbf{Numeric data columns}

If the column you want to filter contains numeric data, a symbol drop-down menu appears. (Figure 56)

\textbf{To apply a filter to a value-based column:}

1. Click the \textbf{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 57.
3. (Optional) Click to add filter(s).

4. Click the Or or And radio button to choose Or or AND relationship logic. (Figure 58)
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 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 59 on page 56.
Chapter 3 Summary window and sample table

Filtering column data

Clearing an individual filter

1. Right-click on the filtered column you want to clear.
   The column menu appears. (Figure 60)

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 61)

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 62)

   **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 63)

![Figure 63 Manage Filters window - Select a Column](image)

3. Click the **Symbol Value** drop-down menu to select the filtering symbol you want. (Figure 64)

![Figure 64 Manage Filters window - Select a Symbol](image)

4. Click inside the text entry box to enter new value(s), as shown in Figure 64.

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 65)

![Figure 65 Manage Filters window - OR or AND Relationship](image)

7. If needed, click **Add Column Filter**, then repeat the above steps. (Figure 66)

![Figure 66 Manage Filters window - Adding another Column Filter](image)

8. Click **OK**.

   To remove a filter(s), click 🗑.

   Click **Clear All** to remove ALL filters in the Manage Filters window.
### Copying column data

To copy column data to your clipboard:

1. Click to select a column you want to copy to a clipboard, then right-click on it. The column menu appears. (Figure 67)

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 68)

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

   A color pallet appears, as shown in Figure 68.
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 69.

![Figure 69 Right-click menu - Set User Color](image)

**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 70)

![Figure 70 Two column example in MS Notepad](image)

**IMPORTANT!** Your user_color entries must match the color pallet naming conventions shown in Figure 71. 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 71)
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 72)

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 shown in Figure 73.

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 74**.
Box Plots

Note: By default, the Viewer generates two Box Plot window tabs. (Figure 75)

Viewing the default box plots

Figure 75  Table and Box Plot 1

<table>
<thead>
<tr>
<th>Sample Filename</th>
<th>Pass/Tail</th>
<th>QC</th>
<th>cell state</th>
<th>QC</th>
<th>cell state</th>
<th>DQC</th>
<th>QC</th>
<th>computed gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAI2794.CEL</td>
<td>Pass</td>
<td>0.964</td>
<td>0.732</td>
<td>0.246</td>
<td>17.015</td>
<td>18.472</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>HAI2755.CEL</td>
<td>Pass</td>
<td>0.963</td>
<td>0.889</td>
<td>0.093</td>
<td>14.205</td>
<td>17.758</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>HAI2773.CEL</td>
<td>Pass</td>
<td>0.960</td>
<td>0.855</td>
<td>0.076</td>
<td>14.076</td>
<td>17.401</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>HAI2769.CEL</td>
<td>Pass</td>
<td>0.960</td>
<td>0.855</td>
<td>0.076</td>
<td>14.076</td>
<td>17.401</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>HAI2783.CEL</td>
<td>Pass</td>
<td>0.960</td>
<td>0.855</td>
<td>0.076</td>
<td>14.076</td>
<td>17.401</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>HAI2801.CEL</td>
<td>Pass</td>
<td>0.960</td>
<td>0.855</td>
<td>0.076</td>
<td>14.076</td>
<td>17.401</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>HAI2862.CEL</td>
<td>Pass</td>
<td>0.960</td>
<td>0.855</td>
<td>0.076</td>
<td>14.076</td>
<td>17.401</td>
<td>female</td>
<td></td>
</tr>
</tbody>
</table>

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 66.

To change a Box Plot’s axis properties, you must create a new Box Plot. See “Adding a new box plot” on page 66.
Changing the box plot’s scale setting ranges

1. Click **Scale Settings**.
   The Scale Settings window appears. (Figure 76)

   ![Figure 76 Scale Settings window](image)

   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 77)

   ![Figure 77 Box Plot Versus menus](image)

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 214.

4. Click **OK**.
   A new Box Plot window tab is created.

---

**Reading box plot percentiles**

See Figure 78.

![Box Plot Percentiles](image)

At any time, click X to remove a window/tab, as shown in Figure 79.

**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_gender, as shown in Figure 80.

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 80)

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 69.
Chapter 3 Summary window and sample table

Scatter Plot

Changing the scatter plot’s setting ranges

1. Click [Scale Settings].
   The Scale Settings window appears. (Figure 81)

   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 82)

2. Use the drop-down menus to select your Plot’s versus scenario (X and Y axis). For definitions, see “Sample table” on page 214.
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 83.

   ![Figure 83 New Window/Tab](image)

4. Click the **Color By** and **Shape By** drop-down menus to select the combination view you want. See "Sample table" on page 214 for Color By and Shape By definitions.

   **Note:** Your imported sample attributes are also available for use within the Color By and Shape By drop-down menus. For information on importing sample attributes, see "Importing sample attributes" on page 51.

   A legend appears within the plot. (Figure 84)
**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.

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

**Customizing your scatter plot view settings**

1. Click the Configure button.
   
The Color Scale Configuration window appears. (Figure 85)

![Figure 85 Color By options](image)

2. Use the provided text fields and color drop-down menus to customize your Color By selection.
   - **Auto Scale** check box (when checked) uses the actual minimum (lower bound) and maximum (upper bound) as your min/max scale. Uncheck the Auto Scale check box to enter your min and max number scales in the provided fields.
   - Click the Cutoff Type drop-down menu to select your cutoff preference.
     - **Above Cutoff Failing** - This presents a hard visual cutoff graph of all values that fail ABOVE the Cutoff value entered. The Above Cutoff data is represented by the color defined for Max. (Green in Figure 85)
– **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 86)*

![Figure 86: Below Cutoff](image)

– **No Cutoff** - This presents a smooth 3-point gradient of your defined Max, Min, and colors. *(Figure 87)*

![Figure 87: No Cutoff](image)

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 88) 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 75.

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 88.

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 88.

![Figure 88 Table and Plate View 1](image)

**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 76.
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 89)*

![Figure 89 Plate View Metric setting](image)

2. Use the drop-down menus to select your Plate View’s Metric setting. See the tables in *"Sample table" on page 214* 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 90*.

![Figure 90 New Window/Tab](image)

**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 214* 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 90*. 

![Figure 90 New Window/Tab](image)
Customizing your plate view settings

1. Click [Configure].

   The Color Scale Configuration window appears. (Figure 91)

   ![Figure 91 Color scale options](image)

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 92)

![Figure 92 Concordance window](image)

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 signature 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 93.

**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.
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.

For definitions of the Concordance columns, see Table 22 on page 218.

To change the default threshold colors, use the color drop-down menus.
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 signature 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 91.
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 passes. 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 94)

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 95)
Acknowledge the warning, then click **OK**.

After the BAC analysis is complete, a BAC table appears (Figure 96) and becomes an additional tab within the Viewer.
### Chapter 3 Summary window and sample table

#### Batch allele consistency

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.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>probeset_id</td>
<td>The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP probe sets only).</td>
</tr>
<tr>
<td>Case Control Batch Name</td>
<td>BAC value for probeset in batch that is used as the case control for the comparison. Calculated ( \frac{2 \times n_{BB} + n_{AB}}{2 \times (n_{AA} + n_{AB} + n_{BB})} )</td>
</tr>
<tr>
<td>Case Batch Name</td>
<td>BAC value for probeset in batch that is used as the case for the comparison. Calculated ( \frac{2 \times n_{BB} + n_{AB}}{2 \times (n_{AA} + n_{AB} + n_{BB})} )</td>
</tr>
<tr>
<td>passBACTest</td>
<td>Pass or Fail depending on: (</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percent Passing</th>
<th>Percent Failing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Probesets</td>
<td>65008</td>
<td>95.828</td>
<td>4.172</td>
</tr>
<tr>
<td>BestAndRecommended</td>
<td>58151</td>
<td>98.932</td>
<td>1.068</td>
</tr>
</tbody>
</table>

For more details on the Summary window, see "Data analysis summary" on page 48.
Reanalyzing samples

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

1. Single-click on a CEL file or Ctrl click, Shift click, or press Ctrl A (to select multiple files).
2. Click the **Reanalyze Selected Samples** button.
   The following message appears. *(Figure 97)*

![Figure 97](image)

3. Click **OK**.
Chapter 3 Summary window and sample table
Reanalyzing samples

Your selected samples are displayed and ready for re-analysis in a New Analysis window tab. (Figure 98)

4. See Chapter 2, "Performing an analysis" on page 24 for instructions on setting up an analysis.
Probeset summary table

Figure 99 is an example of a standard Probeset Summary Table.

<table>
<thead>
<tr>
<th>probeset_id</th>
<th>Affy探针ID</th>
<th>ConversionType</th>
<th>Banding/Recommended</th>
<th>CR</th>
<th>MinClusterFrequency</th>
<th>H韦尔-Value</th>
<th>FLD</th>
<th>HomFLD</th>
<th>HetSO</th>
<th>HetRO</th>
<th>Ncis</th>
<th>n_BA</th>
<th>n_AB</th>
<th>n_BN</th>
<th>n_NC</th>
<th>HomHet</th>
<th>BestProbeSet</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>AX-1108626</td>
<td>Aff-2631...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.0008</td>
<td>1</td>
<td>9.553</td>
<td>-0.27</td>
<td>2.055</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108627</td>
<td>Aff-2834...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.0090</td>
<td>1</td>
<td>6.315</td>
<td>0.305</td>
<td>2.46</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108659</td>
<td>Aff-3492...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.263</td>
<td>0.359</td>
<td>7.905</td>
<td>16.453</td>
<td>0.960</td>
<td>1.848</td>
<td>3</td>
<td>13</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108663</td>
<td>Aff-3533...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.438</td>
<td>1</td>
<td>16.357</td>
<td>23.079</td>
<td>0.666</td>
<td>4.045</td>
<td>3</td>
<td>4</td>
<td>13</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108672</td>
<td>Aff-3446...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.813</td>
<td>1</td>
<td>6.554</td>
<td>0.407</td>
<td>2.98</td>
<td>2</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>AX-1108691</td>
<td>Aff-3848...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.312</td>
<td>0.013</td>
<td>7.352</td>
<td>15.131</td>
<td>0.473</td>
<td>1.241</td>
<td>3</td>
<td>14</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108690</td>
<td>Aff-3027...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.417</td>
<td>1</td>
<td>6.831</td>
<td>16.538</td>
<td>0.369</td>
<td>1.610</td>
<td>3</td>
<td>4</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108671</td>
<td>Aff-3178...</td>
<td>MonotHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.164</td>
<td>1</td>
<td>5.753</td>
<td>0.115</td>
<td>2.237</td>
<td>2</td>
<td>19</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>AX-1108673</td>
<td>Aff-3181...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.164</td>
<td>1</td>
<td>7.48</td>
<td>0.243</td>
<td>2.081</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108675</td>
<td>Aff-3204...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.271</td>
<td>0.656</td>
<td>7.842</td>
<td>16.241</td>
<td>0.426</td>
<td>1.809</td>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108677</td>
<td>Aff-3287...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.831</td>
<td>1</td>
<td>5.78</td>
<td>0.228</td>
<td>1.694</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108678</td>
<td>Aff-3249...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.0098</td>
<td>1</td>
<td>7.778</td>
<td>0.424</td>
<td>2.395</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108680</td>
<td>Aff-3238...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.417</td>
<td>0.112</td>
<td>7.581</td>
<td>17.588</td>
<td>0.52</td>
<td>1.377</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108690</td>
<td>Aff-3299...</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>0.163</td>
<td>1</td>
<td>5.78</td>
<td>0.228</td>
<td>1.694</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108695</td>
<td>Aff-3552...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.312</td>
<td>0.013</td>
<td>7.352</td>
<td>15.131</td>
<td>0.473</td>
<td>1.241</td>
<td>3</td>
<td>14</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108691</td>
<td>Aff-3848...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.271</td>
<td>0.656</td>
<td>7.842</td>
<td>16.241</td>
<td>0.426</td>
<td>1.809</td>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108694</td>
<td>Aff-3848...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.271</td>
<td>0.656</td>
<td>7.842</td>
<td>16.241</td>
<td>0.426</td>
<td>1.809</td>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108695</td>
<td>Aff-3584...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.271</td>
<td>0.656</td>
<td>7.842</td>
<td>16.241</td>
<td>0.426</td>
<td>1.809</td>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
<tr>
<td>AX-1108696</td>
<td>Aff-3590...</td>
<td>PolyHighRecal</td>
<td>1</td>
<td>0</td>
<td>0.271</td>
<td>0.656</td>
<td>7.842</td>
<td>16.241</td>
<td>0.426</td>
<td>1.809</td>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>all</td>
</tr>
</tbody>
</table>
Multiallele probsets

Multiallele probsets are probsets that can report calls from probe sets with more than one alternate allele. Multiallele probsets can be identified by selecting the multiallelic column. (Figure 100) A value of 1 indicates a multi-allelic probset. Additional columns can be shown that report metrics for multiallele probsets.

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

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

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

Figure 101 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 156 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 101. 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 182.
### Probeset Summary Table with CN-aware genotyping and Allele Translation

**Figure 101**

<table>
<thead>
<tr>
<th>Probeset ID</th>
<th>Conversion Type</th>
<th>CR</th>
<th>Minor Allele Frequency</th>
<th>HLG Value</th>
<th>FLD</th>
<th>HomFLD</th>
<th>HetISO</th>
<th>HomRD</th>
<th>n_A</th>
<th>n_A</th>
<th>n_A</th>
<th>n_A</th>
<th>n_B</th>
<th>n_B</th>
<th>n_CN</th>
<th>n_CN</th>
<th>Heterozygous</th>
<th>Homozygous</th>
<th>Cell Altered</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0101001</td>
<td>PolyHighResal</td>
<td>180</td>
<td>0.404</td>
<td>0.310</td>
<td>20.638</td>
<td>44.654</td>
<td>8.302</td>
<td>2.774</td>
<td>3</td>
<td>19</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>A0101002</td>
<td>PolyHighResal</td>
<td>19.360</td>
<td>0.312</td>
<td>0.127</td>
<td>17.090</td>
<td>1.199</td>
<td>0.054</td>
<td>76</td>
<td>63</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>A0101003</td>
<td>PolyHighResal</td>
<td>180</td>
<td>0.22</td>
<td>0.097</td>
<td>0.302</td>
<td>1.153</td>
<td>0.168</td>
<td>71</td>
<td>43</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>A0101004</td>
<td>PolyHighResal</td>
<td>180</td>
<td>0.112</td>
<td>0.047</td>
<td>0.164</td>
<td>0.313</td>
<td>0.224</td>
<td>0</td>
<td>16</td>
<td>41</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>A0101005</td>
<td>PolyHighResal</td>
<td>180</td>
<td>0.275</td>
<td>0.149</td>
<td>1.097</td>
<td>12.612</td>
<td>8.313</td>
<td>2.224</td>
<td>0</td>
<td>81</td>
<td>64</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>A0101006</td>
<td>PolyHighResal</td>
<td>180</td>
<td>0.122</td>
<td>0.052</td>
<td>0.344</td>
<td>0.341</td>
<td>0.124</td>
<td>1.672</td>
<td>3</td>
<td>140</td>
<td>24</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
</tbody>
</table>
Probeset summary table

Setting your Probeset summary table view

1. Click the Apply View drop-down. *(Figure 102)*

- 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 103)*
   A list of available columns appear.

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 24 on page 222 for column definitions.

Selecting annotations

1. Click the Select Annotation button.
The Select Annotations window appears. (Figure 104)

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 104) to add or remove ALL available annotations. See "Annotations and columns" on page 219 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 105)
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 106)

   ![Figure 106 Save Current View window](image)

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 107)

   ![Figure 107 Right-click menu](image)

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 107)

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 108)

   ![Revert Calls drop-down menu](image)

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

**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 109. It also cannot contain any annotation columns.

![Change Call by Text file example shown in Excel (as a tab-delimited text file)](image)

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 110) offers optional steps for post-processing Probeset data. These functions utilize the output files from the previous genotyping and classification steps as input.

Reanalyze drop-down menu

| Reanalyze | Regenerate SNP Metrics | Run OTV Caller |

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 ProbeSet Metrics.

   The Regenerate SNP Metrics window appears. (Figure 111)
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 29.
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 35.
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 112)
   
   **Note**: If your analysis contains multiallele probesets, additional conversion types are displayed.
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 113)

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 intended 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 114)

   ![Figure 114 Run OTV Caller window](image)

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 115), 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.

2. To save all Probesets currently displayed in the ProbeSet Summary Table, click Create ProbeSet List from Table. The Save ProbeSet List window appears. (Figure 116)

   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 117, otherwise an error message appears.
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, as shown in Figure 118, and the second column Alternate Marker Name.
3. Click Open.
   Your Alternate Marker ID is now displayed in the Probeset Summary Table and Cluster Plot. (Figure 119)
Chapter 4 Probeset summary table and cluster plot
Managing your probeset list

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 120)
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 121)*

3. Add your previously saved Probeset List by selecting it from the drop-down list. *(Figure 122)*

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 123.
4. Click the **Saved ProbeSet List** drop-down, select your saved ProbeSet List, as also shown in Figure 123, then click **OK**.

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 124, 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 124 Cluster window tab for biallelic probesets](image)

**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 125 on page 102 for examples of a multiallele probeset.
**Chapter 4 Probeset summary table and cluster plot**

**Cluster plot**

**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 126 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.
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 21 on page 214 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 127)*

   ![Figure 127 Color By and Shape By Legend example](image)

   **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 128.

   ![Figure 128 Probeset and Alternate Marker ID in Cluster Plot name](image)
Setting new scale setting ranges

1. Click `Scale Settings`.
   The Scale Settings window appears. (Figure 129)

   ![Figure 129 Scale Settings window](image)

   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 130)

   ![Figure 130 Color By options](image)

2. Use the provided text fields and color drop-down menus to customize your Color By selection.
   - **Auto Scale** check box (when checked) uses the actual minimum (lower bound) and maximum (upper bound) as your min/max scale. Uncheck the Auto Scale check box to enter your min and max number scales in the provided fields.
Note: If the Auto Scale check box remains unchecked, you must enter new scale limits for each subsequent analysis.

- Click the Cutoff Type drop-down menu to select your cutoff preference.
  - **Above Cutoff Failing** - This presents a hard visual cutoff graph of all values that fail ABOVE the Cutoff value entered. The Above Cutoff data is represented by the color defined for Max. ([Green in Figure 130](#))
  - **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 131](#))
  - **No Cutoff** - This presents a smooth 3-point gradient of your defined Max, Min, and colors. ([Figure 132](#)).

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 133.

2. Release the mouse button to select (lock in) the group of lassoed samples, as shown in Figure 134.
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 135)

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

   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 136), then click to select Revert Selected Call(s).
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 137**), then click to select a new setting.

   ![Figure 137 Biallele Cluster Model Data drop-down menus](image)

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

Saving the current cluster plot view

1. Click the button (upper right).

   An Explorer window appears.

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

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 5000 samples have been validated on a Windows 7 system with 32GB RAM and on a Windows 10 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.

**Note:** For more information on copy number workflows, see the Axiom CN Data Analysis Guide.

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 gender.

If you have an array that supports fixed region CN analysis as part of the Genotyping workflow, go to page 182. If not, go to Step 1 on page 111.
1. From the New Analysis window tab, click the **Array Type** drop-down (Figure 138) to select the array type you want to use.

![Figure 138 Main window](image)

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 139)

![Figure 139 Populated CEL File pane example](image)

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 140)*

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

The Dashboard window tab appears and displays the progress. When the analysis is complete and there are less than 40 male and female or any unknown gender samples a Warning note appears.

**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 required 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.

**To create 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 141)*, then create a new or locate an empty folder to save results.
4. (Optional) To edit, locate, the exported .cnregioncalls.txt file (Figure 142), then open it in a text editor (e.g. Excel).

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 143)

7. In the Analysis Setting pane, expand Copy Number Reference Creation section.

8. Click the **Select CN Region Calls File (optional)** 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 144) 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.

11. Click OK to continue.

   The Dashboard window tab appears and displays the progress. When the analysis is complete and there are less than 40 male and female or any unknown gender samples a Warning note appears.

   **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.

**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 145)
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 146) to choose a different configuration.

![Figure 146 Select Analysis Configurations drop-down](image)

**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 147)

![Figure 147 Copy Number Discovery drop-down](image)

7. (Optional) The **CN Reference File** defaults to the file supplied in the library package, as shown in Figure 148. To select a different reference file, click the down arrow to expand the window, then click to select reference file you want.
8. In the Threshold Settings pane, click the Select Threshold Configuration dropdown, then click to select the appropriate threshold, as shown in Figure 149.

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

A Save Analysis Configuration window appears. (Figure 150)

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.

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 151) to choose a different configuration.

   ![Figure 151 Select Analysis Configurations drop-down](image)

   **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 152.
7. (Optional) The CN Reference File defaults to the file supplied in the library package, as shown in Figure 152. 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 189 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 153.

9. Enter a name for your CN Discovery Batch in the **Name** field, then click **Run**.
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 155  Summary window tab and definitions

* **Batch Name**: Discovery import via txt
* **Array Package Name**: Axiom_UKB_W
* **Array Type Name**: Axiom_UKB_WCS
* **Array Display Name**: Axiom_UKB_WC
* **Workflow Type**: Copy Number Discovery
* **Date Created**: 5/22/2018 7:42:17 AM

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

**Sample Summary**: Breaks down the sample QC for your analysis run and displays the number that pass each of your QC Thresholds. The Reference file used in run is also included.

**CN QC Metrics Summary**: Contains CN QC information for each plate including the number and percent of samples passing MAPD and Waviness SD.

**CN QC Thresholds**: Metrics used to QC batch.

**CN Region QC Thresholds**: Discovery specific QC thresholds.

---

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

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

**Sample Summary**: Breaks down the sample QC for your analysis run and displays the number that pass each of your QC Thresholds. The Reference file used in run is also included.

**CN QC Metrics Summary**: Contains CN QC information for each plate including the number and percent of samples passing MAPD and Waviness SD.

**CN QC Thresholds**: Metrics used to QC batch.

**CN Region QC Thresholds**: Discovery specific QC thresholds.

---

<table>
<thead>
<tr>
<th>Plate Barcode</th>
<th>Number of samples Analyzed</th>
<th>% of samples Passing</th>
</tr>
</thead>
<tbody>
<tr>
<td>529404420946032143248</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>
Sample table

Figure 156  Sample Table window tab

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Filename</td>
<td>CEL file name.</td>
</tr>
<tr>
<td>MAPD</td>
<td>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.</td>
</tr>
<tr>
<td>Waviness SD</td>
<td>A global measure of variation of microarray probes that is insensitive to shortrange variation and focuses on long-range variation.</td>
</tr>
<tr>
<td>CN passes MAPD</td>
<td>Yes, if the sample’s MAPD value is not greater than the MAPD threshold used by CN QC.</td>
</tr>
<tr>
<td>CN passes WavinessSD</td>
<td>Yes, if the sample’s WavinessSD value is not greater than the WavinessSD threshold used by CN QC.</td>
</tr>
<tr>
<td>CN passes QC</td>
<td>Yes, if the sample passes both MAPD and WavinessSD threshold tests.</td>
</tr>
<tr>
<td>Count of CN 0 Segments</td>
<td>Number of segments that are CN zero.</td>
</tr>
<tr>
<td>Count of CN 1 Segments</td>
<td>Number of segments that are CN 1.</td>
</tr>
<tr>
<td>Count of CN 2 Segments</td>
<td>Number of segments that are CN 2.</td>
</tr>
<tr>
<td>Count of CN 3 Segments</td>
<td>Number of segments that are CN 3 or more.</td>
</tr>
<tr>
<td>Autosome % LOH</td>
<td>Percentage of the autosomal regions of the genome reporting loss of heterozygosity.</td>
</tr>
</tbody>
</table>

(Optional) For more CN Discovery Batch options, click the QC Analysis drop-down menu, as shown in Figure 157.
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 158)
   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.
Chapter 5 Copy number analysis

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 158.

**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 158. 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

To generate 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 160)

---

**Table**

<table>
<thead>
<tr>
<th>Sample Filename</th>
<th>Chromosome</th>
<th>Region</th>
<th>Start Position</th>
<th>Stop Position</th>
<th>Size</th>
<th>Median Count</th>
<th>State</th>
<th>Median LogRatio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5L.CML</td>
<td>1</td>
<td>1</td>
<td>10114847</td>
<td>10115000</td>
<td>1533</td>
<td>52</td>
<td>3</td>
<td>0.15</td>
</tr>
<tr>
<td>5L.CML</td>
<td>5</td>
<td>5</td>
<td>40235931</td>
<td>40410920</td>
<td>15627</td>
<td>52</td>
<td>3</td>
<td>0.16</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>6385406</td>
<td>63481895</td>
<td>646</td>
<td>3</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>5L.CML</td>
<td>13</td>
<td>13</td>
<td>7178157</td>
<td>7253857</td>
<td>7774</td>
<td>68</td>
<td>3</td>
<td>0.19</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>38135906</td>
<td>38035923</td>
<td>1023</td>
<td>54</td>
<td>3</td>
<td>0.19</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>78962442</td>
<td>78963505</td>
<td>10753</td>
<td>71</td>
<td>3</td>
<td>0.17</td>
</tr>
<tr>
<td>5L.CML</td>
<td>1</td>
<td>1</td>
<td>31361236</td>
<td>31360051</td>
<td>2615</td>
<td>12</td>
<td>3</td>
<td>0.17</td>
</tr>
<tr>
<td>5L.CML</td>
<td>3</td>
<td>3</td>
<td>62066209</td>
<td>64066766</td>
<td>14055</td>
<td>85</td>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>63122896</td>
<td>63112151</td>
<td>188878</td>
<td>57</td>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>5L.CML</td>
<td>3</td>
<td>3</td>
<td>5911134</td>
<td>635493</td>
<td>33939</td>
<td>205</td>
<td>3</td>
<td>0.17</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>6315906</td>
<td>6315907</td>
<td>1</td>
<td>51</td>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>6315906</td>
<td>6315907</td>
<td>1</td>
<td>51</td>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>5L.CML</td>
<td>1</td>
<td>1</td>
<td>11804621</td>
<td>11428264</td>
<td>32824</td>
<td>168</td>
<td>3</td>
<td>0.17</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>81359055</td>
<td>81359015</td>
<td>10959</td>
<td>52</td>
<td>3</td>
<td>0.19</td>
</tr>
<tr>
<td>5L.CML</td>
<td>3</td>
<td>3</td>
<td>6090001</td>
<td>6090086</td>
<td>15686</td>
<td>64</td>
<td>3</td>
<td>0.18</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>36802184</td>
<td>36802408</td>
<td>10224</td>
<td>51</td>
<td>3</td>
<td>0.21</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>38112816</td>
<td>38112896</td>
<td>10280</td>
<td>51</td>
<td>3</td>
<td>0.21</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>6311447</td>
<td>63117446</td>
<td>10026</td>
<td>57</td>
<td>3</td>
<td>0.18</td>
</tr>
<tr>
<td>5L.CML</td>
<td>5</td>
<td>5</td>
<td>8943</td>
<td>415798</td>
<td>40480</td>
<td>123</td>
<td>3</td>
<td>0.15</td>
</tr>
<tr>
<td>5L.CML</td>
<td>11</td>
<td>11</td>
<td>14670100</td>
<td>14670100</td>
<td>0</td>
<td>83</td>
<td>3</td>
<td>0.16</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>2984483</td>
<td>2981681</td>
<td>18002</td>
<td>65</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>37627724</td>
<td>37629645</td>
<td>19122</td>
<td>95</td>
<td>3</td>
<td>0.16</td>
</tr>
<tr>
<td>5L.CML</td>
<td>5</td>
<td>5</td>
<td>62884906</td>
<td>62884906</td>
<td>0</td>
<td>210</td>
<td>3</td>
<td>0.21</td>
</tr>
<tr>
<td>5L.CML</td>
<td>8</td>
<td>8</td>
<td>8943</td>
<td>615920</td>
<td>63078</td>
<td>201</td>
<td>3</td>
<td>0.16</td>
</tr>
<tr>
<td>5L.CML</td>
<td>2</td>
<td>2</td>
<td>62884906</td>
<td>62884906</td>
<td>0</td>
<td>210</td>
<td>3</td>
<td>0.21</td>
</tr>
<tr>
<td>5L.CML</td>
<td>1</td>
<td>1</td>
<td>53440901</td>
<td>53440901</td>
<td>0</td>
<td>73</td>
<td>3</td>
<td>0.15</td>
</tr>
<tr>
<td>5L.CML</td>
<td>4</td>
<td>4</td>
<td>6474966</td>
<td>6474966</td>
<td>0</td>
<td>114</td>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>5L.CML</td>
<td>7</td>
<td>7</td>
<td>1505906</td>
<td>1704961</td>
<td>21835</td>
<td>120</td>
<td>3</td>
<td>0.24</td>
</tr>
</tbody>
</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.

To generate 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 161)

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Filename</td>
<td>CEL file name.</td>
</tr>
<tr>
<td>Chromosome</td>
<td>The chromosome on which the CN event is located.</td>
</tr>
<tr>
<td>Region</td>
<td>Specified region within chromosome, as defined in the library file package, on which copy number analysis is performed.</td>
</tr>
<tr>
<td>Start Position</td>
<td>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.</td>
</tr>
<tr>
<td>Stop Position</td>
<td>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.</td>
</tr>
<tr>
<td>Size</td>
<td>Length of segment in base pairs.</td>
</tr>
<tr>
<td>Marker Count</td>
<td>Number of ProbeSets that make up segment.</td>
</tr>
<tr>
<td>medianlog2ratio</td>
<td>The median Log2Ratio of CN ProbeSets within that segment.</td>
</tr>
<tr>
<td>State</td>
<td>CN state of segment.</td>
</tr>
</tbody>
</table>

**LOH segments table**

Table 4 Default CN Discovery Segments Table

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Filename</td>
<td>CEL file name.</td>
</tr>
<tr>
<td>Chromosome</td>
<td>The chromosome on which the CN event is located.</td>
</tr>
<tr>
<td>Region</td>
<td>Specified region within chromosome, as defined in the library file package, on which copy number analysis is performed.</td>
</tr>
<tr>
<td>Start Position</td>
<td>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.</td>
</tr>
<tr>
<td>Stop Position</td>
<td>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.</td>
</tr>
<tr>
<td>Size</td>
<td>Length of segment in base pairs.</td>
</tr>
<tr>
<td>Marker Count</td>
<td>Number of ProbeSets that make up segment.</td>
</tr>
<tr>
<td>medianlog2ratio</td>
<td>The median Log2Ratio of CN ProbeSets within that segment.</td>
</tr>
<tr>
<td>State</td>
<td>CN state of segment.</td>
</tr>
</tbody>
</table>
Chapter 5  Copy number analysis

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!** 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.

**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.

### Table 5  Default LOH Segments Table

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Filename</td>
<td>CEL file name.</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Chromosome number where LOH is located.</td>
</tr>
<tr>
<td>Start Position</td>
<td>The nucleotide base start position of the LOH segment.</td>
</tr>
<tr>
<td>Stop Position</td>
<td>The nucleotide stop position of the LOH segment.</td>
</tr>
<tr>
<td>Size</td>
<td>Length of LOH in base pairs.</td>
</tr>
<tr>
<td>Marker Count</td>
<td>Number of ProbeSets that make up segment.</td>
</tr>
<tr>
<td>LOH</td>
<td>A flag of 1 indicates the presence of LOH.</td>
</tr>
</tbody>
</table>
Tool bar and viewing features

The WGV window tab features a tool bar. (Figure 162)

- 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 163) to zoom 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 164)*

**Graph settings**

1. Click **Graph Settings**.

   A Whole Genome View Configuration window appears. *(Figure 165)*
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 166)*
Chapter 5 Copy number analysis

Whole Genome View (WGV)

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 167.
Chapter 5 Copy number analysis
Whole Genome View (WGV)

Setting up a WGV

1. At the Sample Table, right-click on the **CNpassesQC** header, then select **Filter**. The CNpassesQC window appears. (Figure 168)

![Figure 168 CNpassesQC window](image)

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 169.

![Figure 167 WGV ToolTip example](image)
5. Verify the annotation that was used in your analysis is displayed, then click OK.

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 171)
Chapter 5 Copy number analysis
Whole Genome View (WGV)

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 172.

![Select Annotations window](image-url)
Chapter 5 Copy number analysis
Whole Genome View (WGV)

The WGV window tab (Figure 172) 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 173). The samples contained in your highlighted rectangle are now highlighted in the Sample Table. (Figure 174)
Chapter 5 Copy number analysis
Whole Genome View (WGV)

Sample data plots (lower) pane

Samples highlighted in the Sample Table appear in the Sample Data Plot pane, as shown in Figure 175. To filter the displayed samples, see "Filtering Samples" on page 139.
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 176)
• 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.

To filter 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 177.
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 178.
3. (Optional) Right-click, then select **Clear Highlighted Region** to remove the highlight.
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 179.

Figure 179  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 180.

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 181.
Figure 181  Synchronized CN Segment table and Whole Genome Plot example
CN fixed regions batches

To view 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

![Summary window tab and definitions](image)

**Analysis Summary**

- **Batch Name**: UKBB Fixed B4.0.0.14
- **Array Package Name**: Axiom_UKB_WCGS.r3
- **Array Type Name**: Axiom_UKB_WCGS
- **Array Display Name**: Axiom_UKB_WCGS.r3
- **Workflow Type**: Copy Number Fixed Regions
- **Date Created**: 5/8/2018 3:35:45 AM

**Sample Summary**

- Number of samples: 96
- Samples passing CN QC: 61
- Reference file: Reference UKBB 100 samples B1058.cn_models

**CN QC Metrics Summary**

<table>
<thead>
<tr>
<th>Plate Barcode</th>
<th>Number of samples Analyzed</th>
<th>Number of samples MAPD ≤ 0.05</th>
<th>Percent of samples passing MAPD ≤ 0.05</th>
<th>Waviness SD ≤ 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50484420946021451540</td>
<td>95</td>
<td>95</td>
<td>62.158</td>
<td></td>
</tr>
<tr>
<td>50484424715127259718</td>
<td>1</td>
<td>1</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

**CN QC Thresholds**

- MAPD: ≤ 0.35
- Waviness SD: ≤ 0.1

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

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

**Sample Summary**: Breaks down the sample QC for your analysis run and displays the number that pass each of your QC Thresholds. The Reference file that was used in run is also included.

**CN QC Metrics Summary**: Contains CN QC information for each plate including the number and percent of samples passing MAPD and Waviness SD.

**CN QC Thresholds**: Metrics used to QC batch.

*Note*: If plate correction has been applied, MAPDc and WavinessSDc thresholds are also shown here.
Sample table

Table 6  Default CN Fixed Regions columns

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPD</td>
<td>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.</td>
</tr>
<tr>
<td>Waviness SD</td>
<td>A global measure of variation of microarray probes that is insensitive to shortrange variation and focuses on long-range variation.</td>
</tr>
<tr>
<td>CN passes MAPD</td>
<td>Yes, if the sample’s MAPD value is not greater than the MAPD threshold used by CN QC.</td>
</tr>
<tr>
<td>CN passes WavinessSD</td>
<td>Yes, if the sample’s WavinessSD value is not greater than the WavinessSD threshold used by CN QC.</td>
</tr>
<tr>
<td>CN passes QC</td>
<td>Yes, if the sample passes both MAPD and WavinessSD threshold tests.</td>
</tr>
<tr>
<td>Control Chosen</td>
<td>“Yes” if the control was listed in the CN Control CEL List file supplied for CN analysis.</td>
</tr>
<tr>
<td>Control Used</td>
<td>“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.</td>
</tr>
<tr>
<td>Number of Controls</td>
<td>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.</td>
</tr>
<tr>
<td>CN tuned using controls</td>
<td>If Number of controls for CN tuning &gt; 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.</td>
</tr>
<tr>
<td>Autosome % LOH</td>
<td>Percentage of the autosomal regions of the genome reporting loss of heterozygosity.</td>
</tr>
</tbody>
</table>

See Table 17 on page 192 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 184.

**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 185)

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 186.
Note: The asterisk on the window tab (Figure 186) indicates the table is synchronized with the Sample Table.

![Figure 186  CN Segments Table window tab](image)

Table 7  CN Segment Table columns

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Filename</td>
<td>CEL file name</td>
</tr>
<tr>
<td>CN Region</td>
<td>Chromosomal region specified in library file package that copy number analysis will be performed.</td>
</tr>
<tr>
<td>CN State</td>
<td>Copy number state of sample in region.</td>
</tr>
<tr>
<td>MedianLog2Ratio</td>
<td>Median log2ratio of selected probesets for each CN Region for each sample.</td>
</tr>
<tr>
<td>Confidence</td>
<td>A calculation to determine likelihood of CN_raw. Higher confidence values indicated more confident CN calls.</td>
</tr>
<tr>
<td>CN_raw</td>
<td>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.</td>
</tr>
<tr>
<td>affymetrix-plate-barcode</td>
<td>Plate barcode number</td>
</tr>
</tbody>
</table>

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.

To generate 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 187)
Figure 187  LOH Segments Table window

<table>
<thead>
<tr>
<th>Sample Filename</th>
<th>Chromosome</th>
<th>Start Position</th>
<th>Stop Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>752721</td>
<td>35</td>
</tr>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>35256386</td>
<td>36</td>
</tr>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>36818066</td>
<td>72</td>
</tr>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>72224718</td>
<td>73</td>
</tr>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>73533529</td>
<td>121</td>
</tr>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>143761875</td>
<td>147</td>
</tr>
<tr>
<td>NA12750.CEL</td>
<td>1</td>
<td>147382550</td>
<td>150</td>
</tr>
</tbody>
</table>

Table 8  Default LOH Segments Table

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome</td>
<td>Chromosome number where LOH is located.</td>
</tr>
<tr>
<td>Start Position</td>
<td>The nucleotide base start position of the LOH segment.</td>
</tr>
<tr>
<td>Stop Position</td>
<td>The nucleotide stop position of the LOH segment.</td>
</tr>
<tr>
<td>Size</td>
<td>Length of LOH in base pairs.</td>
</tr>
<tr>
<td>Marker Count</td>
<td>Number of ProbeSets that make up segment.</td>
</tr>
<tr>
<td>LOH</td>
<td>A flag of 1 indicates the presence of LOH.</td>
</tr>
</tbody>
</table>

**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 188.

**Figure 188  CN Summary table and CN Region Plot**

![CN Summary Table and CN Region Plot](image)

**Table 9  CN Summary Table**

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN Region</td>
<td>Predefined region in chromosome on which copy number is called.</td>
</tr>
<tr>
<td>Number of NoCall</td>
<td>Count of NoCalls per region. If a sample fails CN QC, it is reported as 'NoCall'.</td>
</tr>
<tr>
<td>Number of CN0</td>
<td>Count of CN0 per region</td>
</tr>
<tr>
<td>Number of CN1</td>
<td>Count of CN1 per region</td>
</tr>
<tr>
<td>Number of CN2</td>
<td>Count of CN2 per region</td>
</tr>
<tr>
<td>Number of CN3</td>
<td>Count of CN3 per region</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Chromosome number which CN Region is located.</td>
</tr>
<tr>
<td>Number of NoCall_*</td>
<td>Count of copy number calls for the specific bin described by * . These copy number calls are calls that are between identified copy number states.</td>
</tr>
</tbody>
</table>
Exporting a CN summary table

To export a CN Summary Table:
1. Click the Export drop-down menu.
   Four export options appear. (Figure 189)

<table>
<thead>
<tr>
<th>Export</th>
<th>Filters</th>
<th>Current Table</th>
<th>All Data</th>
<th>Selected Regions</th>
<th>Copy Number Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selected Regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copy Number Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 190)
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 190.
- 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.
   An Export IGV window appears. (Figure 191)

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 192)*

   ![Figure 192 Export Nexus window](image)

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 192)*

   ![Figure 193 Export VCF window](image)

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.
Allele translation

About translations

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

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

See Appendix B, "About allele translation" on page 198 for more information.

Performing allele translation

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

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

1. From an open batch in the Viewer, navigate to the ProbeSet Summary Table, then click the **Perform Allele Translation** button, as shown in Figure 194.

**Figure 194  ProbeSet Summary Table window**

The Perform Allele Translation window appears. (Figure 195)

**Figure 195  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 195) A Phenotype report will not be created if this option is not used.
Refer to Table 10 and Table 11 for descriptions of the available Allele Translation options.

Table 10 Allele Translation options

<table>
<thead>
<tr>
<th>Select Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProbeSet List Filter</td>
<td>This option translates only the genotypes of SNPs in a user-specified probe set list. Only translate SNPs in a probe set list selected by the user.</td>
</tr>
<tr>
<td></td>
<td>Click Browse to select a different probe set list, or select from existing probe set lists. You can also click X to deselect probe set list filtering, which enables allele translation for all available probe sets, including those that may not have passed SNP QC.</td>
</tr>
</tbody>
</table>
| Report only the first named haplotype in the translation file (Default) | This option is relevant if you filter by a ProbeSet List AND if the ProbeSet List contains some (but not all) of the available probe sets in a gene that is allele translated.  
If both conditions are met, then it is possible that your ProbeSet List excludes a probe set 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 probe set 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 probe set. If this probe set is omitted, and the data indicates that both *1 and *2C are possible (due to a NoCall at that probe set), 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 probe sets you have decided not to translate, AND you agree that the selection of which haplotype to report is correct. |
Performing allele translation

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.

<table>
<thead>
<tr>
<th>Select Translation Results Folder</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Folder</td>
<td>The path to the output folder for the translation reports and log. Click its Browse button to set the path.</td>
</tr>
<tr>
<td>Name</td>
<td>The new folder name for the translation results, whose default name has the date-time format YYYY-MM-DD_HHMSS_translations. Confirm the default or enter a new folder name.</td>
</tr>
</tbody>
</table>
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 196.

**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 197.

**Figure 196** Example: Basic layout of a Comprehensive and Summary Translation report

**Figure 197** Example: Summary Translation report
Phenotype translation report

The basic layout of this report is shown in Figure 198.

![Figure 198 Example: Phenotype Translation report](image)

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Phenotype Info</th>
<th>Haplotype Info</th>
<th>Sample Info</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Phenotype report

The basic layout of this report is shown in Figure 199.

![Figure 199 Example: Phenotype report](image)

<table>
<thead>
<tr>
<th>Index</th>
<th>CIP File</th>
<th>Gene</th>
<th>Phenotype Cell</th>
<th>Gene Activity</th>
<th>Known Cell</th>
<th>Unknown Cell</th>
<th>Interpretation Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001-0002</td>
<td>test_01.c</td>
<td>CYP1A2</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1F/1F</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0001-0004</td>
<td>test_01.c</td>
<td>CYP2A6</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1/1</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0001-0009</td>
<td>test_01.c</td>
<td>CYP2D6</td>
<td>EM</td>
<td>none/none</td>
<td>*5/5</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0002-0020</td>
<td>test_02.c</td>
<td>CYP1A2</td>
<td>EM_or IM</td>
<td>normal/reduced</td>
<td>*1A/1C/1F</td>
<td></td>
<td>MULT</td>
</tr>
<tr>
<td>0002-0022</td>
<td>test_02.c</td>
<td>CYP2A6</td>
<td>EM_or IM</td>
<td>normal/reduced</td>
<td>*1/17</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0002-0024</td>
<td>test_02.c</td>
<td>CYP2D6</td>
<td>EM_or IM</td>
<td>normal/reduced</td>
<td>*1/5</td>
<td>*4/UNK</td>
<td>UNIQ+UNK</td>
</tr>
<tr>
<td>0002-0029</td>
<td>test_02.c</td>
<td>CYP2D6</td>
<td>EM_or IM</td>
<td>normal/reduced</td>
<td>*2/29</td>
<td>*2/UNK,*29/UNK</td>
<td>NC/PRA/NA</td>
</tr>
<tr>
<td>0004-0020</td>
<td>test_04.c</td>
<td>CYP1A2</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1A/1A</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0004-0022</td>
<td>test_04.c</td>
<td>CYP2A6</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1/1</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0004-0024</td>
<td>test_04.c</td>
<td>CYP2D6</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1/1</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0005-0020</td>
<td>test_05.c</td>
<td>CYP1A2</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1A/1F</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0005-0022</td>
<td>test_05.c</td>
<td>CYP2A6</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1/1</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0005-0024</td>
<td>test_05.c</td>
<td>CYP2D6</td>
<td>EM</td>
<td>reduced/reduced</td>
<td>*6/5</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0005-0029</td>
<td>test_05.c</td>
<td>CYP2D6</td>
<td>IM</td>
<td>normal/reduced</td>
<td>*2/4</td>
<td>*1/UNK,*2/UNK,*4/1NC/PRA/NA</td>
<td>UNIQ</td>
</tr>
<tr>
<td>0009-0020</td>
<td>test_09.c</td>
<td>CYP1A2</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1F/1F</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0009-0022</td>
<td>test_09.c</td>
<td>CYP2A6</td>
<td>EM</td>
<td>normal/normal</td>
<td>*1/1</td>
<td></td>
<td>UNIQ</td>
</tr>
<tr>
<td>0009-0024</td>
<td>test_09.c</td>
<td>CYP2D6</td>
<td>IM</td>
<td>reduced/reduced</td>
<td>*6/5</td>
<td></td>
<td>UNIQ</td>
</tr>
</tbody>
</table>
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 200.

Figure 200  Example: Header for Comprehensive and Summary report
Available report fields and descriptions

Array tracking  Refer to Table 12 for descriptions of the available Array Tracking fields.

Table 12  Array Tracking fields

<table>
<thead>
<tr>
<th>Array Tracking fields</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>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]</td>
</tr>
<tr>
<td>Filename</td>
<td>Name of the sample file.</td>
</tr>
</tbody>
</table>

Gene-specific  Refer to Table 13 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 13  Gene-specific fields

<table>
<thead>
<tr>
<th>Gene-specific fields</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associated Gene</td>
<td>Gene symbol</td>
</tr>
<tr>
<td>Phenotype Call</td>
<td>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 203 for more information.</td>
</tr>
</tbody>
</table>
### Table 13  Gene-specific fields

<table>
<thead>
<tr>
<th>Gene-specific fields</th>
<th>Description</th>
</tr>
</thead>
</table>
| **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 203 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. |
Available report fields and descriptions

### 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.

### Table 13 Gene-specific fields

<table>
<thead>
<tr>
<th>Gene-specific fields</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unknown Call</strong></td>
<td>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.</td>
</tr>
</tbody>
</table>
| **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. |
Chapter 6 Allele translation

Available report fields and descriptions

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 14 Marker-specific fields

<table>
<thead>
<tr>
<th>Marker-specific fields</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary Flag (marker annotation)</td>
<td>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&gt;T[A</td>
</tr>
<tr>
<td>Relevant Alleles (marker annotation)</td>
<td>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.</td>
</tr>
<tr>
<td>Common Name (marker annotation)</td>
<td>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)</td>
</tr>
<tr>
<td>Probeset ID</td>
<td>Unique identifier for the SNP.</td>
</tr>
<tr>
<td>Basecall</td>
<td>The observed bases, also known as the &quot;raw&quot; genotypes.</td>
</tr>
<tr>
<td>Reference Base</td>
<td>This field generally indicates the more common allele in biallelic SNPs. Certain genes use a particular GenBank entry as the &quot;Reference genome&quot; and the observed allele at each marker across the gene is then reported as Reference.</td>
</tr>
<tr>
<td>Variant Base</td>
<td>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).</td>
</tr>
<tr>
<td>Call</td>
<td>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.</td>
</tr>
<tr>
<td>Haplotype Marker (marker annotation)</td>
<td>Differentiates SNPs used to make haplotype calls or single-marker variant calls.</td>
</tr>
<tr>
<td></td>
<td>• 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 &quot;Known Calls&quot; and Unknown Calls of the report.</td>
</tr>
<tr>
<td></td>
<td>• 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 &quot;Call&quot; column.</td>
</tr>
<tr>
<td>Change for Variant (marker annotation)</td>
<td>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.</td>
</tr>
</tbody>
</table>
### Table 14  Marker-specific fields

<table>
<thead>
<tr>
<th>Marker-specific fields</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA Change (marker annotation)</td>
<td>Location of the mutation on a reference mRNA sequence.</td>
</tr>
<tr>
<td>Genome Position (marker annotation)</td>
<td>This is the chromosomal position of the mutation.</td>
</tr>
<tr>
<td>dbSNP RS ID (marker annotation)</td>
<td>The dbSNP identifier for the marker.</td>
</tr>
</tbody>
</table>

### Tracking Edited Genotype Calls

Refer to Table 15 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 15  Fields for tracking edited genotype calls

<table>
<thead>
<tr>
<th>Tracking Edited Genotype Calls</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Basecall</td>
<td>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.</td>
</tr>
<tr>
<td>Override Comment</td>
<td>This field reports edited in AxAS if the call has been edited within Axiom Analysis Suite.</td>
</tr>
</tbody>
</table>
Uncalled report

Refer to Table 16 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>
<thead>
<tr>
<th>Uncalled Report</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filename</td>
<td>Name of the sample file.</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene symbol.</td>
</tr>
<tr>
<td>Common Name</td>
<td>The Common Name defines the gene and positional information about the genetic change tested with the probeset.</td>
</tr>
<tr>
<td>Basecall</td>
<td>The Uncalled report contains all the NoCall genotypes from SNPs used for allele translation.</td>
</tr>
<tr>
<td>Override Comment</td>
<td>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.</td>
</tr>
<tr>
<td>Reference Allele</td>
<td>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 (&lt;1%).</td>
</tr>
<tr>
<td>Variant Allele</td>
<td>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&gt;T&gt;A[A893SorT], because two specific mutations are known to occur at this genomic location (G&gt;A and G&gt;T).</td>
</tr>
</tbody>
</table>
Using the sample table export options

1. From the Sample Table, click the Export drop-down.  
   ![Figure 201 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.

To export 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 202)

   ![Figure 202 Export menu]

   **Figure 202 Export menu**

   - Export
     - Filters
       - Current Table
       - All Data
       - Signature SNPs
       - Genotyping Data
       - Cluster Plots to PDF

   **Exporting the current table**

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

   1. Click Export Current Table.
      An Explorer window appears.
   2. Enter a filename, then click Save.
      The current table data is now saved and exported as a tab-delimited text file.

   **Exporting all data**

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

   1. Click Export All Data.
      An Explorer window appears.
   2. Enter a filename, then click Save.
      All data in the table (displayed or not) is saved and exported as a tab-delimited text file.

   **Exporting signature SNPs**

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

   1. Click Export Signature SNPs.
      An Explorer window appears.
   2. Enter a filename, then click Save.
      All data is now saved a tab-delimited text file.

   **Exporting genotyping data**

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

   1. Click Export Genotyping Data.
      The Export Genotyping Data window appears. (Figure 203)
Result Output Formats

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**

- **Alternate Marker Names:** Click this check box (Figure 205) 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.

---

**Call Output Formats**

1. Click the radio button to select the appropriate Call Output Format (Figure 206) you want to use.

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.
Chapter 7  Exporting

Using the ProbeSet summary table export options

**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 207) you want to include.

   ![Figure 207 Exported Data selections](image)

   **Exported Data:**
   - Confidence
   - Signal
   - Include CN Region Calls

   **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 208 ProbeSet List Filter and Output path](image)

2. Click the ProbeSet List Filter’s drop-down menu to view and select a previously saved list. (Figure 208)

**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 208)
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 209)

![Figure 209 SNP Identifier](image)

Changing the Current Annotation File (Optional)

1. To change the currently displayed Annotation File, click the Annotation File field’s **Browse** button. (Figure 210)
   
   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 210)

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 174.
Including ProbeSet Summary table columns in your genotyping report

1. Check ProbeSet data columns to include in your export, as shown in Figure 211. The included ProbeSet data columns will be appended at the end of the files.

![Figure 211 Select ProbeSet Data Columns to Include pane](image)
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 212)

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 213)
Figure 213   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).
External tools

AxAS auto-detects any previously installed application that may further analyze your genotype results. When a compatible application is detected, an External Tools window tab is generated, as shown in Figure 214.

Figure 214   External Tools window tab

The tools listed below are designed to further analyze the genotype results. These tools may require that this application be closed to provide access to the result file.

- CDCB Export Tool
- Long Format Export Tool

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 215)*
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 216)

For details on how to use the Long Format Export Tool application, refer to its User Guide (P/N 703455).
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 217.

<table>
<thead>
<tr>
<th>Workflow Stage</th>
<th>Analysis Description</th>
<th>Typical Axiom</th>
<th>Predefined CN Genotyping †</th>
<th>CN-Aware Genotyping ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample QC</td>
<td>Sample QC</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Best Practices</td>
<td>Identify Copy Number (CN) control samples that pass Sample QC</td>
<td>not done</td>
<td>female samples used as CN plate controls</td>
<td>CN controls identified by comparing measured with reference genotype calls</td>
</tr>
<tr>
<td>Genotyping</td>
<td>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.</td>
<td>not done</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Supply CN results to genotyping engine</td>
<td>not done</td>
<td>No</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Final genotyping</td>
<td>yes</td>
<td>yes, including OffTargetVariant calls</td>
<td>yes, using additional call codes for haploid and zero CN calls</td>
</tr>
</tbody>
</table>

† library package specifies analysis_category = "cn_gt_2" or "cnvmix_gt_2"

‡ library package specifies analysis_category = "cn_gt" or "cnvmix_gt"
Appendix A Predefined region CN analysis variations
Predefined copy number analyses

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

![Figure 218   New Analysis window tab](Image)

Selecting a mode (Workflow)

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


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 196.

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.
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 220.
For information on typical Axiom array analysis and threshold fields, see "Setting up an analysis" on page 24 and "Setting up threshold settings" on page 33.

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 221.
Analysis settings
with copy number
options

IMPORTANT! Only experienced users should modify default analysis settings.

Sample QC

- **GT Analysis File**: Parameters file for the genotyping step that calculates QC Call Rate.

- **Prior Model File**: Defines prior knowledge of SNP cluster locations. This file has the same format as a posteriors file, which is generated by the genotyping step. This means that you can "train" on a custom data set, and use the updated knowledge of cluster locations as a "seed" to possibly improve future genotyping batches. This file must contain two row entries for the GENERIC and GENERIC:1 probesets (if there are any probesets to be genotyped that are not listed in this file).

- **SNP List File**: A file of probeset IDs to genotype. For Sample QC it defines the probesets used to calculate QC Call Rate.

- **Gender File**: A file specifying the desired gender of every sample. If supplied, software will use values in this file instead of the computed gender. Gender impacts genotyping of chromosome X and Y SNPs.

- **Hints/Inbred File**: If a hints file, a file of expected genotype calls. This is used to influence the predicted cluster locations, which influences the final calls. In the Sample QC section, it influences the QC Call Rate.

- **Control Reference Calls File**: For CN-aware genotyping only, a recommended file containing the expected signature SNP calls of the CN control samples and used by Sample QC to identify the control samples among the supplied CEL files.

Genotyping

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

<table>
<thead>
<tr>
<th>Workflow</th>
<th>CN Control CEL List File Input</th>
<th>Chosen Controls for CN Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best Practices</strong></td>
<td>None</td>
<td>Passing controls identified by Sample QC step</td>
</tr>
<tr>
<td><strong>Genotyping</strong></td>
<td>None</td>
<td>No controls, therefore no plate signal tuning is done.</td>
</tr>
<tr>
<td><strong>Best Practices</strong></td>
<td>User-supplied</td>
<td>User-supplied controls</td>
</tr>
<tr>
<td><strong>Genotyping</strong></td>
<td>User-supplied</td>
<td>User-supplied controls</td>
</tr>
</tbody>
</table>

- **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).

• 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.

• **Gender File**: A file specifying the known gender of every sample. If supplied, software will use values in this file instead of the computed gender. Gender impacts genotyping of chromosome X and Y SNPs.

• **Hints/Inbred File**: If a hints file, a file of expected genotype calls. This is used to influence the predicted cluster locations, which influences the final calls.

• **Posterior File Name**: The desired file output of a genotyping analysis, specifying identified SNP probeset cluster locations, variance, and relative weight. One use of this file is as a prior model file for future genotype analyses.

• **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
  - snpid (snpid = affy_snp_id)
  - multi.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 226.

### SNP QC

Refer to Table 20 on page 208 for SNP QC Threshold name definitions.

### Assigning an output folder path

1. Click the **Output Folder path’s Browse** button. (Figure 222)

   ![Figure 222 Output Folder field](image)

   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 223)*

   ![Figure 223](image)

   **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 224)*

   ![Figure 224](image)

   After **Success** is displayed in the Status column *(Figure 224)*, 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 225)* 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 225.

- **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 225 Viewer window tabs](image)

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 225.
Refer to Table 17 for descriptions of each CN Summary Report column.

<table>
<thead>
<tr>
<th>CN Summary Report</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of files analyzed</td>
<td>Count of CEL files supplied to the Genotyping step (which includes CN analysis).</td>
</tr>
<tr>
<td>Number of samples failing CNQC</td>
<td>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.</td>
</tr>
<tr>
<td>Number of samples passing CNQC</td>
<td>Count of CEL files that pass copy number quality control checks.</td>
</tr>
<tr>
<td>Number of controls analyzed</td>
<td>Count of CEL files supplied to the Genotyping step that are identified as controls.</td>
</tr>
<tr>
<td>Number of controls failing CNQC</td>
<td>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.</td>
</tr>
<tr>
<td>Note:</td>
<td>For the CN-aware genotyping analysis variation, Non-PAR X, Y, and mitochondrial SNPs use gender and other information to report non-diploid calls, regardless of CN QC status.</td>
</tr>
<tr>
<td>Number of controls for CN tuning</td>
<td>Count of control CEL files that pass copy number quality control checks.</td>
</tr>
<tr>
<td>CN tuned using controls</td>
<td>If Number of controls for CN tuning &gt; 0, then this value is &quot;Yes&quot;. 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.</td>
</tr>
</tbody>
</table>
After successfully running a CN analysis, additional columns appear within the Sample Table, as shown in Figure 226.

Table 18  Added Sample Table columns after running a CN Analysis

<table>
<thead>
<tr>
<th>Sample Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPD</td>
<td>Median Absolute Pairwise Difference of log2ratio signals of adjacent copy number (CN) probesets.</td>
</tr>
<tr>
<td>MAPDc</td>
<td>This metric is used to determine CN passes MAPD and CN passes QC and is calculated after plate based signal correction.</td>
</tr>
<tr>
<td>WavinessSD</td>
<td>Waviness Standard Deviation of log2ratio signals of adjacent copy number (CN) probesets.</td>
</tr>
<tr>
<td>WavinessSDc</td>
<td>This metric is used to determine CN passes WavinessSD and CN passes QC and is calculated after plate based signal correction.</td>
</tr>
<tr>
<td>CN passes MAPD</td>
<td>“Yes” if the sample’s MAPDc value is not greater than the MAPDc threshold used by CN QC.</td>
</tr>
<tr>
<td>CN passes WavinessSD</td>
<td>“Yes” if the sample’s WavinessSDc value is not greater than the WavinessSDc threshold used by CN QC.</td>
</tr>
</tbody>
</table>

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 gender and other information to report non-diploid calls, regardless of CN QC status.

Control chosen "Yes" if the control was listed in the CN Control CEL List file supplied for CN analysis.

Control used “1” if the control was used to adjust CN signals of all samples in the same plate. The value will be 1 if Control chosen = 1 and CN passes QC = Yes.
Appendix A Predefined region CN analysis variations

Viewing your genotyping and CN results

Table 18  Added Sample Table columns after running a CN Analysis

<table>
<thead>
<tr>
<th>Sample Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of controls for CN tuning</td>
<td>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.</td>
</tr>
<tr>
<td>CN tuned using controls</td>
<td>If Number of controls for CN tuning &gt; 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.</td>
</tr>
</tbody>
</table>

Probeset summary table

After successfully running a CN-aware genotyping analysis, additional columns appear within the SNP Summary Table, as shown in Figure 227.

Figure 227  ProbeSet Summary Table after a CN Analysis run

Refer to Table 19 for descriptions of each added SNP Summary Table column.

Table 19  Added ProbeSet Summary Table columns after CN-aware genotyping is performed

<table>
<thead>
<tr>
<th>ProbeSet Summary Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_A</td>
<td>Count of hemizygous “A” calls.</td>
</tr>
<tr>
<td>n_B</td>
<td>Count of hemizygous “B” calls.</td>
</tr>
<tr>
<td>n_CN0</td>
<td>Count of ZeroCN calls.</td>
</tr>
<tr>
<td>CopyNumIssue (hidden by default)</td>
<td>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.</td>
</tr>
</tbody>
</table>
Appendix A Predefined region CN analysis variations

Viewing your genotyping and CN results

**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 228.

### 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 Median Log2Ratio 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.
Appendix A Predefined region CN analysis variations

Overview and use of the best practices workflow

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

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

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

**Step 1 Sample QC:** Identify and exclude the poor quality sample CEL files.

**Step 2 Genotyping:** Genotype only the sample CEL files that pass Step 1.

**Step 1: The Sample QC Workflow performs the following steps:**

- **Genotype a small set of SNPs used to uniquely identify each sample (“Signature SNPs”).** This method uses static calling boundaries, so a given sample's calls are not influenced by other samples.
- **DishQC:** Calculate DishQC metric and exclude CEL files with too small a DishQC from next step. For this method, a given sample's metrics are not influenced by other samples.
- **QC Call Rate:** Initial cluster genotyping on remaining samples for selected QC SNPs, to identify and remove any additional CEL files with low QC call rates. For this method, batch information from other samples influences a given samples QC call rates. Genotypes are not stored.
– **Plate QC**: Compare the average QC Call Rate of passing samples within a plate against the threshold "Average call rate for passing samples". If the metric is below this threshold, all the samples on the plate will fail Plate QC, and will not be genotyped.

– **Identify Copy Number plate controls**:
  - For CN-aware genotyping arrays like PharmacoScan, auto-identify reagent control samples by comparing measured SignatureSNP genotypes from all CEL files against reference genotypes. A control is identified if [number of Signature SNPs with a call is >= "Control comparisons"] AND [Concordance % of signature SNP calls compared with reference genotypes is >= "Control concordance"]. Control samples that pass all sample QC checks will be used in the subsequent Genotyping workflow, unless overridden by the user.
  - For predefined CN region arrays like CarrierScan, select as CN controls all female samples passing sample QC.
  - **Step 2**: Genotyping can be run by itself or as part of the Best Practices Workflow. When run as part of the Best Practices Workflow, only CEL files passing Sample QC are genotyped. When run by itself, the user has full control of which samples to genotype. Genotyping does the following steps:
    – For arrays supporting predefined region CN analysis or CN-aware genotyping, collate normalized signals for CN probesets of interest and perform predefined-region CN analysis. This step also calculates CN-specific QC metrics MAPD and wavinessSD. Samples that fail CN QC have their CN states reset to NoCall. If control samples are supplied (by user or by Sample QC step), then plate-based signal correction is performed. Control samples that fail CN QC are not used for plate-based signal correction.
    – 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.

---

**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.
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 clinically-recognized star nomenclature via Allele Translation. This section explains the organization of the translation reports to help you interpret the translation data. Key concepts such as phase ambiguity and the impact of missing data on haplotype-based allele calling are described.

Gene table layout for haplotyping

To appreciate how haplotyping operates, it is essential to describe the organization of the gene tables (Figure 229 and Figure 230) in the translation library file (*.translation).

<table>
<thead>
<tr>
<th>Reference Link</th>
<th>Probe Set ID</th>
<th>Strand to Report</th>
<th>dbsNP ID</th>
<th>rsSNP ID</th>
<th>Defining</th>
<th>cDNA Nucleotide Position</th>
<th>Genome Position</th>
<th>Change</th>
<th>Common Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMID: 11295847</td>
<td>AX:112162174</td>
<td>N</td>
<td>113</td>
<td>134G/A</td>
<td>Ch15:75015305 G45D</td>
<td>CYP1A1*13_134G=A(G45D)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 11295847</td>
<td>AX:112262501</td>
<td>N</td>
<td>113</td>
<td>993G&gt;T</td>
<td>Ch15:75013804 M331I</td>
<td>CYP1A1*6_1635G=T(M331I)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 15618738</td>
<td>AX:112279125</td>
<td>N</td>
<td>113</td>
<td>127G_1276insT</td>
<td>Ch15:75013093 E426frameshift</td>
<td>CYP1A1*7_2345_2346insT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 8895751</td>
<td>AX:112162073</td>
<td>N</td>
<td>113</td>
<td>1382C&gt;A</td>
<td>Ch15:75012987 T461N</td>
<td>CYP1A1*4_2452C=A(T461N)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 9070254</td>
<td>AX:173402723</td>
<td>N</td>
<td>113</td>
<td>1384A&gt;G</td>
<td>Ch15:75012985 G462V</td>
<td>CYP1A1*2C_2454A=G(G462V)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 11295847</td>
<td>AX:165878418</td>
<td>N</td>
<td>113</td>
<td>1390C&gt;A</td>
<td>Ch15:75012979 R464S</td>
<td>CYP1A1<em>3Ror</em>9_2460C=A(R464ScR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 11295847</td>
<td>AX:165878418</td>
<td>N</td>
<td>113</td>
<td>1390C&gt;T</td>
<td>Ch15:75012979 R464C</td>
<td>CYP1A1<em>3Ror</em>9_2460C=A(R464ScR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 15618738</td>
<td>AX:112262005</td>
<td>N</td>
<td>113</td>
<td>1429C&gt;T</td>
<td>Ch15:75012940 R477W</td>
<td>CYP1A1*10_24960C=A(R477W)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 15618738</td>
<td>AX:112161841</td>
<td>N</td>
<td>113</td>
<td>*595T&gt;C</td>
<td>Ch15:75012225 3UTR</td>
<td>CYP1A1*3_3204T=C(3UTR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 21490707</td>
<td>AX:11839795</td>
<td>N</td>
<td>113</td>
<td>1694G&gt;A</td>
<td>Ch15:75019449 Promoter</td>
<td>CYP1A1*-1_1694G=A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 25348619</td>
<td>AX:86570574</td>
<td>N</td>
<td>113</td>
<td>*2760G&gt;T</td>
<td>Ch15:75017176 Intron</td>
<td>CYP1A1*-1_2760G=T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 11295847</td>
<td>AX:165751549</td>
<td>N</td>
<td>113</td>
<td>835C&gt;T</td>
<td>Ch15:75014049 R279G</td>
<td>CYP1A1*13_835C=G(R279G)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 9351182</td>
<td>AX:83250267</td>
<td>N</td>
<td>113</td>
<td>857T&gt;C</td>
<td>Ch15:75014027 286T</td>
<td>CYP1A1*14_286T=C(286T)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMID: 15618738</td>
<td>AX:11419619</td>
<td>N</td>
<td>113</td>
<td>1143C&gt;A</td>
<td>Ch15:75013563 F381L</td>
<td>CYP1A1*1876C=A(F381L)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B About allele translation

#### Gene table layout for haplotyping

- **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*1, 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).

#### 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 |   |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*6_1635G>T(M331I) | Y | C | A | A |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*7_2345_2346insT | Y | - | A |    |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*8_2413T>A(I448N) | Y | A | T |    |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*4_2452C>A(T461N) | Y | G | T | T |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*2C_2454A>G(I462V) | Y | T | C | C |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*5or9_9_2460C>A>T(R4645orC) | Y | G | T | T |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1*5or9_9_2460C>A>T(R4645orC) | 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-274606G>T | N | C | A |   |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1_1390C>GorT(R279GorW) | N | G | C | C |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1_1390C>GorC(R279GorW) | N | G | A | A |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1_1412T>C(I286T) | N | A | G | G |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1_1875C>A(F381L) | N | G | T | T |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |
| CYP1A1_2458C>G(A463G) | N | G | C | C |    |    |    |    |    |    |    |    |    |    |    |    |       |       |       |       |       |       |

**Figure 230** Haplotype descriptions [Example gene table data for markers in CYP1A1]

**Figure 229** 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 231.

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 232 on page 202 lists the multiple possible calls due to phase ambiguity that were observed in six HapMap populations, and how often they occurred.
It is worth pointing out that the predicted phenotypes of some of these alternative diplotype calls are identical, and in these cases the Phenotype report will then report a single phenotype. When phase ambiguity is encountered and the Phenotype report does not resolve the multiple calls to a single Phenotype Call, follow-up metabolic screening may be merited to differentiate the actual genetic configuration of the test samples.
Diplotype to phenotype translation

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

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

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

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

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

1. Import a custom SNP List into the 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.

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 233 on page 206 for field descriptions.
Figure 233 Field descriptions for the metabolizer library file

<table>
<thead>
<tr>
<th>Metabolizer field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<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</td>
<td>The haplotype name of an allele for a gene as reported in the comprehensive report, e.g. &quot;*2&quot;. A Known Call in the comprehensive report is usually a single pair of alleles, e.g. &quot;*1/*2&quot;. To have this call be matched to a specific row in the metabolizer table, <strong>only one of the rows is needed</strong> in the following table:</td>
</tr>
<tr>
<td>allele_2</td>
<td></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:</td>
</tr>
<tr>
<td>function_1</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. &quot;normal/decreased&quot;. If you leave these fields empty, the phenotype report will display '?' for the Gene Function.</td>
</tr>
<tr>
<td>function_2</td>
<td></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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>gene</th>
<th>allele_1</th>
<th>allele_2</th>
<th>phenotype</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</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.
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
## Threshold names

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

<table>
<thead>
<tr>
<th>Threshold Group</th>
<th>Threshold Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample QC</td>
<td>DQC</td>
<td>A sample’s dish QC value must be greater than or equal to ‘DQC’ to pass sample QC.</td>
</tr>
<tr>
<td>Sample QC</td>
<td>QC call_rate</td>
<td>A sample’s call rate value must be greater than or equal to ‘QC call_rate’ to pass genotyping QC.</td>
</tr>
<tr>
<td>Sample QC</td>
<td>Percent of passing samples</td>
<td>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.</td>
</tr>
<tr>
<td>Sample QC</td>
<td>Average call rate for passing samples</td>
<td>A plate’s average QC call rate of passing samples must be greater than or equal to this number to pass plate QC.</td>
</tr>
<tr>
<td>Sample QC</td>
<td>Control comparisons</td>
<td>CN-aware genotyping parameter. For identifying control samples, this is the minimum number of SignatureSNP probesets compared to a reference.</td>
</tr>
<tr>
<td>Sample QC</td>
<td>Control concordance</td>
<td>CN-aware genotyping parameter. For identifying control samples, this is the minimum percent concordance of SignatureSNP calls to a reference.</td>
</tr>
<tr>
<td>CN QC</td>
<td>MAPD</td>
<td>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.</td>
</tr>
<tr>
<td>CN QC</td>
<td>Waviness SD</td>
<td>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.</td>
</tr>
<tr>
<td>CN Region QC</td>
<td>seg-min-bases-CN-oneormore</td>
<td>Minimum size in bases for segments that report copy number greater than 0.</td>
</tr>
<tr>
<td>CN Region QC</td>
<td>seg-min-bases-CN-zero</td>
<td>Minimum size in bases for segments that report copy number 0.</td>
</tr>
<tr>
<td>CN Region QC</td>
<td>seg-min-probesets-CN-oneormore</td>
<td>Minimum size in probesets for segments that report copy number greater than 0.</td>
</tr>
<tr>
<td>CN QC</td>
<td>MAPDc</td>
<td>MAPD calculated after plate based signal correction.</td>
</tr>
<tr>
<td>Threshold Group</td>
<td>Threshold Name</td>
<td>Description</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CN QC</td>
<td>WavinessSDc</td>
<td>WavinessSD calculated after plate based signal correction.</td>
</tr>
<tr>
<td>CN Region QC</td>
<td>seg-min-probesets-CN-zero</td>
<td>Minimum size in probesets for segments that report copy number 0.</td>
</tr>
</tbody>
</table>
| 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'. |
<p>| 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. |
| 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. |</p>
<table>
<thead>
<tr>
<th>Threshold Group</th>
<th>Threshold Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP QC</td>
<td>hom-ro-hap-1-ZChr-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-ro-hap-2-ZChr-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HomMSBD-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HomMSBD-1-cutoff</td>
<td>For multiallelic probesets, the minimum difference between copy number 2 signals and background signals for probesets with 1 allele.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HomMMA-cutoff</td>
<td>The minimum of the means for each multiallelic probeset allele’s homozygous cluster.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>FLD-MA-cutoff</td>
<td>For multiallelic probesets, the minimum acceptable average FLD value between all populated homozygous and heterozygous clusters.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>FLD-MA-2-cutoff</td>
<td>For multiallelic probesets, minimum acceptable average FLD values between all populated homozygous and heterozygous clusters for probesets with only 2 alleles.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>Min-FLD-MA-cutoff</td>
<td>For multiallelic probesets, the smallest acceptable minimum FLD value between all populated homozygous and heterozygous clusters.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>Min-FLD-MA-2-cutoff</td>
<td>For multiallelic probesets, the smallest acceptable minimum FLD value between homozygous and heterozygous clusters for probesets with only 2 alleles.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HetSO-MA-2-cutoff</td>
<td>For multiallelic probesets, the minimum acceptable value for the Size Offset of heterozygous clusters for probesets with only 2 alleles.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HomRO-MA-cutoff</td>
<td>For multiallelic probesets, the minimum acceptable value for the Contrast (X position) of the homozygous clusters (Ratio Offset).</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HomRO-MA-2-cutoff</td>
<td>For multiallelic probesets, the minimum acceptable value for the Contrast (X position) of the homozygous clusters (Ratio Offset) for probesets with 2 genotype clusters.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>HomRO-MA-1-cutoff</td>
<td>For multiallelic probesets, the minimum acceptable value for the Contrast (X position) of the homozygous clusters (Ratio Offset) when a probeset has 1 genotype cluster.</td>
</tr>
<tr>
<td>Multiallelic SNP QC</td>
<td>priority-order-MA</td>
<td>Priority order of probeset conversion types when performing probeset selection.</td>
</tr>
</tbody>
</table>
### Table 20  Threshold names

<table>
<thead>
<tr>
<th>Threshold Group</th>
<th>Threshold Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiallelic SNP QC</td>
<td>Best-CR-MA-cutoff</td>
<td>The minimum Call Rate for selecting a best and recommended probeset from multiple probesets for a single multi.snp.id.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-ro-hap-1-MTChr-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-ro-hap-2-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-ro-hap-2-XChr-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-ro-hap-2-MTChr-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-hap-X-cutoff</td>
<td>For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the haploid clusters relative to the homozygous clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-hap-Y-lower-cutoff</td>
<td>For autosomal probesets, maximum acceptable value for the correctness of the Size (Y position) of the haploid clusters relative to homozygous clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>hom-hap-Y-upper-cutoff</td>
<td>For autosomal probesets, maximum acceptable value for the correctness of the Size (Y position) of the haploid clusters relative to homozygous clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>CN0-hap-X-cutoff</td>
<td>For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the ZeroCN cluster relative to the haploid clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>CN0-hap-X-YChr-cutoff</td>
<td>For Y probesets, minimum acceptable value for the correctness of the Contrast (X position) of the ZeroCN cluster relative to the haploid clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>CN0-hap-Y-cutoff</td>
<td>For autosomal probesets, minimum acceptable value for the correctness of the Size (Y position) of the ZeroCN cluster relative to the haploid clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>CN0-dip-X-cutoff</td>
<td>For autosomal probesets, minimum acceptable value for the correctness of the Contrast (X position) of the ZeroCN cluster relative to the diploid clusters.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>CN0-dip-Y-cutoff</td>
<td>For autosomal probesets, minimum acceptable value for the correctness of the Size (Y position) of the ZeroCN cluster relative to the diploid clusters.</td>
</tr>
</tbody>
</table>
### Table 20 Threshold names

<table>
<thead>
<tr>
<th>Threshold Group</th>
<th>Threshold Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP QC</td>
<td>aaf-XChr-cut</td>
<td>For non-PAR X probesets, maximum acceptable difference in male versus female A-allele frequency. Any probeset that fails is categorized as Other.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>fld-XChr-cut</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>homfld-XChr-cut</td>
<td>For non-PAR X probesets, minimum acceptable FLD value for male haploid cluster separation.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>homfld-YChr-cut</td>
<td>For Y probesets, minimum acceptable FLD value for male haploid cluster separation.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>min-genotype-freq-samples</td>
<td>Minimum count of samples to support the genotype frequency p-value calculation. This parameter is used if a genotype frequency file is supplied.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>genotype-p-value-cutoff</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>priority-order</td>
<td>Priority order of probeset conversion types when performing probeset selection.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>recommended</td>
<td>Probesets having ConversionTypes in this recommended set will be included in the recommended probeset list.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>y-restrict</td>
<td>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.</td>
</tr>
<tr>
<td>SNP QC</td>
<td>min-YChr-samples-cut</td>
<td>For Y probesets, minimum count of called samples of each gender needed for the Euclidean distance test between male and female genotype clusters.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>homfld-cut</td>
<td>Minimum acceptable FLD value for homozygous cluster separation.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>bb-varx-cut</td>
<td>Maximum acceptable Contrast (X position) variance for BB cluster. Not used if bb-varx-z-cut-enabled is true.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>bb-vary-cut</td>
<td>Maximum acceptable Size (Y position) variance for BB cluster. Not used if bb-vary-z-cut-enabled is true.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>ab-varx-cut</td>
<td>Maximum acceptable Contrast (X position) variance for AB cluster. Not used if ab-varx-z-cut-enabled is true.</td>
</tr>
</tbody>
</table>
### Table 20  Threshold names

<table>
<thead>
<tr>
<th>Threshold Group</th>
<th>Threshold Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS Supplemental</td>
<td>ab-vary-cut</td>
<td>Maximum acceptable Size (Y position) variance for AB cluster. Not used if ab-vary-z-cut-enabled is true.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>aa-varx-cut</td>
<td>Maximum acceptable Contrast (X position) variance for AA cluster. Not used if aa-varx-z-cut-enabled is true.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>aa-vary-cut</td>
<td>Maximum acceptable Size (Y position) variance for AA cluster. Not used if aa-vary-z-cut-enabled is true.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>bb-varx-z-cut-enabled</td>
<td>If true, bb-varx-z-cut is used.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>bb-vary-z-cut-enabled</td>
<td>If true, bb-vary-z-cut is used.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>ab-varx-z-cut-enabled</td>
<td>If true, ab-varx-z-cut is used.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>ab-vary-z-cut-enabled</td>
<td>If true, ab-vary-z-cut is used.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>aa-varx-z-cut-enabled</td>
<td>If true, aa-varx-z-cut is used.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>aa-vary-z-cut-enabled</td>
<td>If true, aa-vary-z-cut is used.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>bb-varx-z-cut</td>
<td>Maximum acceptable Contrast (X position) Z-score for BB cluster. Not used if bb-varx-z-cut-enabled is false.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>bb-vary-z-cut</td>
<td>Maximum acceptable Size (Y position) Z-score for BB cluster. Not used if bb-vary-z-cut-enabled is false.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>ab-varx-z-cut</td>
<td>Maximum acceptable Contrast (X position) Z-score for AB cluster. Not used if ab-varx-z-cut-enabled is false.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>ab-vary-z-cut</td>
<td>Maximum acceptable Size (Y position) Z-score for AB cluster. Not used if ab-varx-z-cut-enabled is false.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>aa-varx-z-cut</td>
<td>Maximum acceptable Contrast (X position) Z-score for AA cluster. Not used if aa-varx-z-cut-enabled is false.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>aa-vary-z-cut</td>
<td>Maximum acceptable Size (Y position) Z-score for AA cluster. Not used if aa-varx-z-cut-enabled is false.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>clustermin</td>
<td>Minimum number of samples in a cluster for the new variance or Z-score value to be calculated.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>n-minor-hom-cut</td>
<td>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.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>hetv-maf-cut</td>
<td>Maximum acceptable heterozygosity parameter. PolyHighResolution probesets with het rate &gt; MAF*hetv-maf-cut are categorized as UnexpectedHeterozygosity.</td>
</tr>
<tr>
<td>PS Supplemental</td>
<td>variance-class</td>
<td>List of categories that supplemental filters are used on.</td>
</tr>
</tbody>
</table>
## Sample table

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

<table>
<thead>
<tr>
<th>Selection</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Filename</td>
<td>CEL file name.</td>
</tr>
<tr>
<td>Pass/Fail</td>
<td>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.</td>
</tr>
<tr>
<td>DQC</td>
<td>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.</td>
</tr>
<tr>
<td>call_rate</td>
<td>Percentage of autosomal SNPs with a call other than NoCall. ‘SpecialSNP’ probesets are excluded, as they are mostly non-autosomal.</td>
</tr>
<tr>
<td>Filtered call rate</td>
<td>Call Rate metric that is restricted to genotyped probesets that are in the batch-specific Recommended probeset list.</td>
</tr>
<tr>
<td>QC call_rate</td>
<td>Percentage of autosomal SNPs with a call other than NoCall (measured at the Sample QC step).</td>
</tr>
<tr>
<td>het_rate</td>
<td>Percentage of SNPs called AB (i.e. the heterozygosity) for autosomal SNPs.</td>
</tr>
<tr>
<td>QC het_rate</td>
<td>Percentage of SNPs called AB (i.e. the heterozygosity) for autosomal SNPs (measured at the Sample QC step).</td>
</tr>
<tr>
<td>computed_gender</td>
<td>Computed gender for the sample.</td>
</tr>
<tr>
<td>QC computed_gender</td>
<td>Computed gender for the sample (measured at the Sample QC step).</td>
</tr>
<tr>
<td>affymetrix-plate-barcode</td>
<td>Plate barcode number.</td>
</tr>
<tr>
<td>QC affymetrix-plate-barcode</td>
<td>Plate barcode number (measured at the Sample QC step).</td>
</tr>
<tr>
<td>affymetrix-plate-peg-wellposition</td>
<td>Well position of the plate’s peg.</td>
</tr>
<tr>
<td>QC affymetrix-plate-peg-wellposition</td>
<td>Well position of the plate’s peg (measured at the Sample QC step).</td>
</tr>
<tr>
<td>Average call rate for passing samples</td>
<td>Average QC Call Rate of passing samples within the plate to which this sample belongs.</td>
</tr>
<tr>
<td>Percent of passing samples</td>
<td>Percentage of samples passing sample QC within the plate to which this sample belongs.</td>
</tr>
<tr>
<td>cel_filepath</td>
<td>CEL file path.</td>
</tr>
<tr>
<td>cel_file_identifier</td>
<td>CEL file identifier.</td>
</tr>
<tr>
<td>affymetrix-array-id</td>
<td>Command Console array identifier.</td>
</tr>
<tr>
<td>total_call_rate</td>
<td>Call rate at the default or user-specified threshold for all SNPs.</td>
</tr>
<tr>
<td>QC total_call_rate</td>
<td>Call rate at the default or user-specified threshold for all SNPs (measured at the Sample QC step).</td>
</tr>
<tr>
<td>total_het_rate</td>
<td>Percentage of SNPs called AB (i.e., the heterozygosity) for all SNPs.</td>
</tr>
</tbody>
</table>
## Appendix C Definitions

### Sample table

<table>
<thead>
<tr>
<th>Selection</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC total_het_rate</td>
<td>Percentage of SNPs called AB [i.e., the heterozygosity] for all SNPs (measured at the Sample QC step).</td>
</tr>
<tr>
<td>hom_rate</td>
<td>Percentage of SNPs called AA or BB [i.e. the homozygosity] for autosomal SNPs.</td>
</tr>
<tr>
<td>QC hom_rate</td>
<td>Percentage of SNPs called AA or BB [i.e. the homozygosity] for autosomal SNPs (measured at the Sample QC step).</td>
</tr>
<tr>
<td>total_hom_rate</td>
<td>Percentage of SNPs called AA or BB [i.e. the homozygosity] for all SNPs.</td>
</tr>
<tr>
<td>QC total_hom_rate</td>
<td>Percentage of SNPs called AA or BB [i.e. the homozygosity] for all SNPs (measured at the Sample QC step).</td>
</tr>
<tr>
<td>cluster_distance_mean</td>
<td>Average distance to the cluster center for the called genotype.</td>
</tr>
<tr>
<td>QC cluster_distance_mean</td>
<td>Average distance to the cluster center for the called genotype (measured at the Sample QC step).</td>
</tr>
<tr>
<td>cluster_distance_stdev</td>
<td>Standard deviation of the distance to the cluster center for the called genotype.</td>
</tr>
<tr>
<td>QC cluster_distance_stdev</td>
<td>Standard deviation of the distance to the cluster center for the called genotype (measured at the Sample QC step).</td>
</tr>
<tr>
<td>allele_summarization_mean</td>
<td>Average of the allele signal estimates (log2 scale).</td>
</tr>
<tr>
<td>QC allele_summarization_mean</td>
<td>Average of the allele signal estimates (log2 scale) (measured at the Sample QC step).</td>
</tr>
<tr>
<td>allele_summarization_stdev</td>
<td>Standard deviation of the allele signal estimates (log2 scale).</td>
</tr>
<tr>
<td>QC allele_summarization_stdev</td>
<td>Standard deviation of the allele signal estimates (log2 scale) (measured at the Sample QC step).</td>
</tr>
<tr>
<td>allele_deviation_mean</td>
<td>Average of the absolute difference between the log2 allele signal estimate and its median across all arrays.</td>
</tr>
<tr>
<td>QC allele_deviation_mean</td>
<td>Average of the absolute difference between the log2 allele signal estimate and its median across all arrays (measured at the Sample QC step).</td>
</tr>
<tr>
<td>allele_deviation_stdev</td>
<td>Standard deviation of the absolute difference between the log2 allele signal estimate and its median across all arrays.</td>
</tr>
<tr>
<td>QC allele_deviation_stdev</td>
<td>Standard deviation of the absolute difference between the log2 allele signal estimate and its median across all arrays (measured at the Sample QC step).</td>
</tr>
<tr>
<td>allele_mad_residuals_mean</td>
<td>Average of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model.</td>
</tr>
<tr>
<td>QC allele_mad_residuals_mean</td>
<td>Average of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model (measured at the Sample QC step).</td>
</tr>
<tr>
<td>allele_mad_residuals_stdev</td>
<td>Standard deviation of the median absolute deviation (MAD) between observed probe intensities and probe intensities fitted by the model.</td>
</tr>
</tbody>
</table>
### Table 21 Sample table

<table>
<thead>
<tr>
<th>Selection</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC allele_mad_residuals_stdev</td>
<td>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).</td>
</tr>
<tr>
<td>A_signal_mean</td>
<td>Average of control A probe raw intensities in the AT channel.</td>
</tr>
<tr>
<td>T_signal_mean</td>
<td>Average of control T probe raw intensities in the AT channel.</td>
</tr>
<tr>
<td>G_signal_mean</td>
<td>Average of control G probe raw intensities in the GC channel.</td>
</tr>
<tr>
<td>C_signal_mean</td>
<td>Average of control C probe raw intensities in the GC channel.</td>
</tr>
<tr>
<td>AT_B</td>
<td>AT channel background, which is the average signal of the GC control probes in AT channel.</td>
</tr>
<tr>
<td>GC_B</td>
<td>GC channel background, which is the average signal of the AT control probes in GC channel.</td>
</tr>
<tr>
<td>AT_S</td>
<td>AT channel signal, which is the average signal of the AT control probes in AT channel.</td>
</tr>
<tr>
<td>GC_S</td>
<td>GC channel signal, which is the average signal of the GC control probes in GC channel.</td>
</tr>
<tr>
<td>AT_SBR</td>
<td>AT channel signal to background ratio, defined as AT_S/AT_B.</td>
</tr>
<tr>
<td>GC_SBR</td>
<td>GC channel signal to background ratio, defined as GC_S/GC_B.</td>
</tr>
<tr>
<td>AT_B_IQR</td>
<td>AT channel interquartile range (middle 50%) of background intensities, measured using GC control probes.</td>
</tr>
<tr>
<td>GC_B_IQR</td>
<td>GC channel interquartile range (middle 50%) of background intensities, measured using AT control probes.</td>
</tr>
<tr>
<td>AT_S_IQR</td>
<td>AT channel interquartile range (middle 50%) of signal intensities, measured using AT control probes.</td>
</tr>
<tr>
<td>GC_S_IQR</td>
<td>GC channel interquartile range (middle 50%) of signal intensities, measured using GC control probes.</td>
</tr>
<tr>
<td>CV_AT</td>
<td>AT channel median coefficient of variation of replicate probe signals for control probesets.</td>
</tr>
<tr>
<td>CV_GC</td>
<td>GC channel median coefficient of variation of replicate probe signals for control probesets.</td>
</tr>
<tr>
<td>AT_FLD</td>
<td>Fisher’s Linear Discriminant between signal and background in the AT channel, defined as [median_of_AT_probe_intensities - median_of_GC_probe_intensities]^2 / [0.5*(AT_S_IQR^2 + AT_B_IQR^2)].</td>
</tr>
<tr>
<td>GC_FLD</td>
<td>Fisher’s Linear Discriminant between signal and background in the GC channel, defined as [median_of_GC_probe_intensities - median_of_AT_probe_intensities]^2 / [0.5*(GC_S_IQR^2 + GC_B_IQR^2)].</td>
</tr>
<tr>
<td>log_diff_qc</td>
<td>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.</td>
</tr>
</tbody>
</table>
### Table 21  Sample table

<table>
<thead>
<tr>
<th>Selection</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>saturation_AT</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>saturation_GC</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>cn-probe-chrXY-ratio_gender_meanX</strong></td>
<td>Average probe intensity [raw, untransformed] of X chromosome nonpolymorphic probes.</td>
</tr>
<tr>
<td><strong>cn-probe-chrXY-ratio_gender_ratio</strong></td>
<td>Gender ratio $Y/X = \text{cn-probe-chrXY-ratio_gender_meanY} / \text{cn-probe-chrXY-ratio_gender_meanX}$.</td>
</tr>
<tr>
<td><strong>cn-probe-chrXY-ratio_gender</strong></td>
<td>Predicted gender, based on the value of \text{cn-probe-chrXY-ratio_gender_ratio}.</td>
</tr>
<tr>
<td><strong>reagent_version</strong></td>
<td>Reagent version used for processing the arrays, based on data intensity values. <strong>Note:</strong> You can only perform batch genotyping analysis on CEL files processed using the same reagent version.</td>
</tr>
<tr>
<td><strong>reagent_discrimination_value</strong></td>
<td>Value assigned to the reagent, and used to determine \text{reagent_version}.</td>
</tr>
<tr>
<td><strong>user_color</strong></td>
<td>User-supplied color to associate with the sample. For more information, see “Setting user colors” on page 60.</td>
</tr>
</tbody>
</table>
Concordance columns

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

Table 22 Concordance columns

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>The first sample in the comparison.</td>
</tr>
<tr>
<td>Reference</td>
<td>The second sample in the comparison.</td>
</tr>
<tr>
<td>#SNPs Called</td>
<td>Number of SNPs common to both sample and reference files with genotype calls.</td>
</tr>
<tr>
<td>#Concordant SNP’s</td>
<td>Number of called SNPs that have the same genotype call.</td>
</tr>
<tr>
<td>%Concordance</td>
<td>Percentage of called SNPs that have the same genotype call.</td>
</tr>
</tbody>
</table>
## Annotations and columns

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

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probeset ID</td>
<td>The Affymetrix unique identifier for the set of probes used to detect a particular Single Nucleotide Polymorphism (SNP probesets only).</td>
</tr>
<tr>
<td>Affx SNP ID</td>
<td>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).</td>
</tr>
<tr>
<td>dbSNP RS ID</td>
<td>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: <a href="http://www.ncbi.nlm.nih.gov/SNP/">http://www.ncbi.nlm.nih.gov/SNP/</a> (SNP probesets only).</td>
</tr>
<tr>
<td>Chromosome</td>
<td>The chromosome on which the SNP is located on the current Genome Version.</td>
</tr>
<tr>
<td>Chromosome Start</td>
<td>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.</td>
</tr>
<tr>
<td>Chromosome Stop</td>
<td>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.</td>
</tr>
<tr>
<td>Strand</td>
<td>Genomic strand that the SNP resides on.</td>
</tr>
<tr>
<td>Cytoband</td>
<td>Cytoband location of the SNP derived from the SNP physical map and the chromosome band data provided by UCSC.</td>
</tr>
<tr>
<td>Strand Vs dbSNP</td>
<td>Indicates whether the SNP is on the same or reverse strand as compared to dbSNP (SNP probesets only).</td>
</tr>
<tr>
<td>ChrX pseudo-autosomal region</td>
<td>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.</td>
</tr>
<tr>
<td>Probe Count</td>
<td>The total number of probes in the Probeset.</td>
</tr>
<tr>
<td>Flank</td>
<td>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).</td>
</tr>
</tbody>
</table>
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.

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).

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele A, Allele B</td>
<td>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”:</td>
</tr>
<tr>
<td></td>
<td>1. SNPs are fixed on the forward strand of the design-time reference genome.</td>
</tr>
<tr>
<td></td>
<td>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);</td>
</tr>
<tr>
<td></td>
<td>3. For non-AT and non-CG SNPs, allele A is A or T, allele B is C or G;</td>
</tr>
<tr>
<td></td>
<td>4. For indels, allele A is -, allele B is the insertion.</td>
</tr>
<tr>
<td></td>
<td>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”).</td>
</tr>
<tr>
<td>Ref Allele, Alt Allele</td>
<td>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 &lt;allele_1&gt;/&lt;allele_2&gt;, where &lt;allele_1&gt; and &lt;allele_2&gt; 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.</td>
</tr>
<tr>
<td>Associated Gene</td>
<td>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.</td>
</tr>
<tr>
<td>Genetic Map</td>
<td>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).</td>
</tr>
<tr>
<td>Microsatellite</td>
<td>Describes the nearest microsatellite markers (upstream, downstream and overlapping) for the SNP.</td>
</tr>
<tr>
<td>Enzyme Fragment</td>
<td>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).</td>
</tr>
<tr>
<td>Copy Number Variation</td>
<td>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).</td>
</tr>
</tbody>
</table>
### Table 23  Annotations and columns

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP Interference</td>
<td>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).</td>
</tr>
<tr>
<td>In Final List</td>
<td>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).</td>
</tr>
<tr>
<td>% GC</td>
<td>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).</td>
</tr>
<tr>
<td>Heterozygous Allele Frequencies</td>
<td>Describes the heterozygous frequency of the allele from Yoruba, Japanese, Han Chinese and CEPH studies using the Affymetrix genotyping arrays. (SNP probe sets only)</td>
</tr>
<tr>
<td>Allele Sample Size</td>
<td>Sample size used for Allele Frequency estimates (SNP probe sets only).</td>
</tr>
<tr>
<td>Allele Frequencies</td>
<td>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).</td>
</tr>
<tr>
<td>Minor Allele</td>
<td>Indicates the Minor Allele of a SNP (SNP probe sets only).</td>
</tr>
<tr>
<td>Minor Allele Frequency</td>
<td>The Minor Allele Frequency of a SNP (SNP probe sets only).</td>
</tr>
<tr>
<td>OMIM ID</td>
<td>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 <a href="http://www.ncbi.nlm.nih.gov/omim/">www.ncbi.nlm.nih.gov/omim/</a>.</td>
</tr>
<tr>
<td>Ordered Alleles</td>
<td>A list of alleles alphabetically ordered by abstract allele code. For biallele probesets, the order is &quot;Allele A // Allele B&quot; For multiallele probesets, the order is &quot;Allele A // Allele B // Allele C // Allele D&quot;, with the number of alleles equal to Allele Count.</td>
</tr>
<tr>
<td>Allele Count</td>
<td>The number of measured alleles. For biallele probesets the count is 2. For multiallele probesets, the count is 3 or more.</td>
</tr>
</tbody>
</table>
Probeset summary table definitions

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

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

A few metrics are calculated only for multi-allele 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 24 Probeset summary table metrics

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConversionType</td>
<td>Probeset classification</td>
</tr>
<tr>
<td>CR</td>
<td>Call rate (CR) is the percentage of samples with a genotype call other than &quot;No Call&quot; for the SNP. <strong>Note:</strong> Call Rate on non-pseudoautosomal regions of chromosome X is reported only for female samples. Call Rate on chromosome Y is reported only for male samples.</td>
</tr>
</tbody>
</table>
| MinorAlleleFrequency | The allele frequency for the A allele is calculated as:  

\[
P_A = \frac{(\# AA \text{ Calls} + 0.5 \times AB \text{ Calls})}{Total \# \text{ Calls}}
\]

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\}. |
Appendix C Definitions
Probeset summary table definitions

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.W.p-Value</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>There are two statistical tests used for HWE. When AA, AB, and BB counts are all &gt;=10, a Chi-squared test is used. When one or more of the AA, AB, and BB counts are &lt;10, an Exact test is used. An Exact test means that the p-value is calculated exactly and not approximated from a population distribution.</td>
</tr>
</tbody>
</table>
|                         |  \[
|                         | \quad x^2 = \frac{(f^2_{aa} - f_{a})^2}{f^2_{aa}} + \frac{(2f_{aa}f_{bb} - f_{ab})^2}{2f_{aa}f_{bb}} + \frac{(f^2_{bb} - f_{b})^2}{f^2_{bb}} 
|                         | \]                                                                                                                                                                                                            |
|                         | Where:                                                                                                                                                                                                       |
|                         | \[
|                         | \quad f_{a} = \frac{(#AA \text{ Calls})}{\text{Total # Calls}} \quad f_{b} = \frac{(#BB \text{ Calls})}{\text{Total # Calls}} \quad f_{ab} = \frac{(#AB \text{ Calls})}{\text{Total # Calls}} \quad f_{aa} = \frac{(#AA \text{ Calls} + 0.5 \times #AB \text{ Calls})}{\text{Total # Calls}} \quad f_{bb} = \frac{(#BB \text{ Calls} + 0.5 \times #AB \text{ Calls})}{\text{Total # Calls}} 
|                         | \]                                                                                                                                                                                                            |
|                         |  \[
|                         | \quad P_{HW} = CDF(x^2) 
|                         |  \]                                                                                                                                                                                                            |
|                         | 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:                                                                                                           |
| 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.

HomFLD
HomFLD is a version of FLD computed for the homozygous genotype clusters. HomFLD is undefined for probesets without two homozygous clusters.

HomFLD_hap
HomFLD_hap is a version of HomFLD computed for the haploid genotype clusters, but only for samples specified by gender_metrics. HomFLD_hap is undefined for probesets without two haploid clusters.

HetSO
Heterozygous Strength Offset measures how far the heterozygous cluster center sits above the homozygous cluster centers in the Size dimension (Y position). Low HetSO values are produced either by misclustering events or by the inclusion of samples that contain variations from the reference genome. Most well-clustered diploid SNPs have positive HetSO values.

HomRO
Homozygote Ratio Offset is the distance to zero in the Contrast dimension (X position) from the center of the homozygous cluster that is closest to zero. If there is only one homozygous cluster, HomRO is the distance from that cluster center to zero on the Contrast dimension.

HomRO_hap
HomRO_hap is a version of HomRO computed for haploid clusters, but only for samples specified by gender_metrics.

nMinor Allele
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.

---

Table 24 Probesets summary table metrics

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLD</td>
<td>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.</td>
</tr>
<tr>
<td>HomFLD</td>
<td>HomFLD is a version of FLD computed for the homozygous genotype clusters. HomFLD is undefined for probesets without two homozygous clusters.</td>
</tr>
<tr>
<td>HomFLD_hap</td>
<td>HomFLD_hap is a version of HomFLD computed for the haploid genotype clusters, but only for samples specified by gender_metrics. HomFLD_hap is undefined for probesets without two haploid clusters.</td>
</tr>
<tr>
<td>HetSO</td>
<td>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.</td>
</tr>
<tr>
<td>HomRO</td>
<td>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.</td>
</tr>
<tr>
<td>HomRO_hap</td>
<td>HomRO_hap is a version of HomRO computed for haploid clusters, but only for samples specified by gender_metrics.</td>
</tr>
<tr>
<td>nMinorAllele</td>
<td>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.</td>
</tr>
<tr>
<td>Nclus</td>
<td>The number of genotype clusters.</td>
</tr>
<tr>
<td>n_AA</td>
<td>The number of AA calls.</td>
</tr>
<tr>
<td>n_AB</td>
<td>The number of AB calls.</td>
</tr>
<tr>
<td>n_BB</td>
<td>The number of BB calls.</td>
</tr>
<tr>
<td>n_A</td>
<td>The number of A calls (haploid).</td>
</tr>
<tr>
<td>n_B</td>
<td>The number of B calls (haploid).</td>
</tr>
<tr>
<td>n_CN0</td>
<td>The number of ZeroCN calls (zero copy number).</td>
</tr>
<tr>
<td>n_NC</td>
<td>The number of NoCall calls, including NoCall_1 (haploid).</td>
</tr>
<tr>
<td>AA.meanX</td>
<td>Average Contrast (X position) for AA cluster.</td>
</tr>
<tr>
<td>AA.meanY</td>
<td>Average Size (Y position) for AA cluster.</td>
</tr>
<tr>
<td>AB.meanX</td>
<td>Average Contrast (X position) for AB cluster.</td>
</tr>
<tr>
<td>AB.meanY</td>
<td>Average Size (Y position) for AB cluster.</td>
</tr>
<tr>
<td>BB.meanX</td>
<td>Average Contrast (X position) for BB cluster.</td>
</tr>
<tr>
<td>BB.meanY</td>
<td>Average Size (Y position) for BB cluster.</td>
</tr>
</tbody>
</table>
### Table 24 Probesets summary table metrics

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hemizygous</td>
<td>Hemizygous flag is 1 if the probeset measures chromosome Y or mitochondrial DNA, indicating that diploid genotypes are not possible. Otherwise the flag is 0.</td>
</tr>
<tr>
<td>BB_dis_x_adj</td>
<td>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.</td>
</tr>
<tr>
<td>AA_dis_x_adj</td>
<td>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.</td>
</tr>
<tr>
<td>freq_diff</td>
<td>A-allele frequency difference between male and female samples.</td>
</tr>
<tr>
<td>HomHet</td>
<td>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.</td>
</tr>
<tr>
<td>BB.varX</td>
<td>Contrast (X position) variance for BB cluster.</td>
</tr>
<tr>
<td>BB.varY</td>
<td>Size (Y position) variance for BB cluster.</td>
</tr>
<tr>
<td>AB.varX</td>
<td>Contrast (X position) variance for AB cluster.</td>
</tr>
<tr>
<td>AB.varY</td>
<td>Size (Y position) variance for AB cluster.</td>
</tr>
<tr>
<td>AA.varX</td>
<td>Contrast (X position) variance for AA cluster.</td>
</tr>
<tr>
<td>AA.varY</td>
<td>Size (Y position) variance for AA cluster.</td>
</tr>
<tr>
<td>BB.varX.Z</td>
<td>Contrast (X position) variance Z-score for BB cluster.</td>
</tr>
<tr>
<td>BB.varY.Z</td>
<td>Size (Y position) variance Z-score for BB cluster.</td>
</tr>
<tr>
<td>AB.varX.Z</td>
<td>Contrast (X position) variance Z-score for AB cluster.</td>
</tr>
<tr>
<td>AB.varY.Z</td>
<td>Size (Y position) variance Z-score for AB cluster.</td>
</tr>
<tr>
<td>AA.var.X.Z</td>
<td>Contrast (X position) variance Z-score for AA cluster.</td>
</tr>
<tr>
<td>AA.var.Y.Z</td>
<td>Size (Y position) variance Z-score for AA cluster.</td>
</tr>
<tr>
<td>CopyNumIssue</td>
<td>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.</td>
</tr>
<tr>
<td>BestProbeset</td>
<td>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.</td>
</tr>
<tr>
<td>BestandRecommended</td>
<td>BestandRecommended flag is 1 if BestProbeset is 1 and the ConversionType belongs to the Recommended set of conversion types. Otherwise the flag is 0.</td>
</tr>
</tbody>
</table>
### Appendix C Definitions

#### Probeset summary table definitions

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender_metrics</td>
<td>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 gender_metrics. If gender_metrics is 'all', all samples are used. The metrics are calculated by gender depending on the chromosomes. What is displayed is restricted to the gender in the gender_metrics column. For chromosomes MT and CP: all genders, no splitting by gender. 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.</td>
</tr>
<tr>
<td>MinMean_cp_2</td>
<td>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.</td>
</tr>
<tr>
<td>minSigBgdndDiffHom</td>
<td>For multiallelic probesets, minimum difference between mean log2 signal and background for the homozgyous 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.</td>
</tr>
<tr>
<td>count_ma_A, count_ma_B, count_ma_C, ...</td>
<td>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.</td>
</tr>
<tr>
<td>Call Modified</td>
<td>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.</td>
</tr>
<tr>
<td>nSamples</td>
<td>Total number of samples [where non-diploid calls have been removed].</td>
</tr>
<tr>
<td>nCalls</td>
<td>The number of assigned calls.</td>
</tr>
<tr>
<td>count_ma_A</td>
<td>The number of alleles that appear per probeset for A allele.</td>
</tr>
<tr>
<td>count_ma_B</td>
<td>The number of alleles that appear per probeset for B allele.</td>
</tr>
<tr>
<td>count_ma_C</td>
<td>The number of alleles that appear per probeset for C allele.</td>
</tr>
<tr>
<td>nAllelesTested</td>
<td>The number of total possible alleles that a probeset can have.</td>
</tr>
<tr>
<td>nAllelesDetected</td>
<td>How many of the nallelesTested actually appeared in the assigned calls for that probeset. nAllelesDetected is always lower than nAllelesTested.</td>
</tr>
<tr>
<td>NHetClus</td>
<td>The number of heterzygous clusters.</td>
</tr>
<tr>
<td>nMajorAlleles</td>
<td>nMajorAlleles is the maximum of all allele counts [count_ma].</td>
</tr>
<tr>
<td>maxMinorAllele</td>
<td>maxMinorAllele is the second largest allele count. nMinorAlleles is the sum of all allele counts without nMajorAlleles.</td>
</tr>
<tr>
<td>nMinorAlles</td>
<td>nMinorAlleles is the sum of all allele counts without nMajorAlleles.</td>
</tr>
<tr>
<td>MAFall</td>
<td>MAFall is the minor allele frequency across all non-major alleles [the number of minor alleles is nMinorAlleles].</td>
</tr>
</tbody>
</table>
### Probeset summary table definitions

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAFmax</td>
<td>MAFmax is the minor allele frequency using the second largest allele count (the number of minor alleles is $\text{maxMinorAllele}$).</td>
</tr>
<tr>
<td>HomCount</td>
<td>HomCount is the number of samples in all homozygous clusters.</td>
</tr>
<tr>
<td>MajorHomCount</td>
<td>MajorHomCount is the number of samples in the largest homozygous cluster.</td>
</tr>
<tr>
<td>MinorHomCount</td>
<td>MinorHomCount is the number of samples in all homozygous clusters without MajorHomCount.</td>
</tr>
<tr>
<td>HetCount</td>
<td>HetCount is the number of samples in all heterozygous clusters.</td>
</tr>
<tr>
<td>HomMMA</td>
<td>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 A{\text{meanA}, BB{\text{meanB}, and CC{\text{meanC}. The higher the HomMMA value, the stronger the allele signal is and the more defined the cluster is.</td>
</tr>
<tr>
<td>FLD_MA</td>
<td>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.</td>
</tr>
<tr>
<td>MinFLD_MA</td>
<td>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.</td>
</tr>
<tr>
<td>HomFLD_MA</td>
<td>HomFLD_MA is similar to FLD_MA except it uses all pairwise combinations of populated homozygous clusters instead of homozygous-heterozygous pairs.</td>
</tr>
<tr>
<td>HetSO_MA</td>
<td>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.</td>
</tr>
<tr>
<td>HomRO_MA</td>
<td>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.</td>
</tr>
<tr>
<td>multiallelic</td>
<td>A ‘1’ in this column indicates the probeset is multiallelic.</td>
</tr>
<tr>
<td>OriginalCT</td>
<td>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.</td>
</tr>
</tbody>
</table>
Table 24  Probesets summary table metrics

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternate Marker Name</td>
<td>This column lists Alternate marker name, if one has been supplied.</td>
</tr>
<tr>
<td>passBACTest</td>
<td>This column appears if a BAC test has been run and indicates Pass or Fail per probeset_id.</td>
</tr>
</tbody>
</table>
Gender calling

Overview

In Axiom Analysis Suite, the gender calling algorithm used to populate the QC Computed Gender call in the Sample Table and the Geno_QCResults.txt is cn-probe-chrXY-ratio_gender 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 gender call is male, while diploid models are used for X chromosome SNPs, when the gender call is female. A No Call is made for Y chromosome SNPs when the gender call is female.

The cn-probe-chrXY-ratio_gender method determines gender based on the ratio of the average probe intensity of nonpolymorphic probes on the Y chromosome (cn-probe-chrXY-ratio_gender_meanY) to the average probe intensity of nonpolymorphic probes on the X chromosome (cn-probe-chrXY-ratio_gender_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 gender call is female, and if it is greater than 0.95, the gender call is male. If the ratio is between these values, the gender call is unknown.

Gender calling and scatter plots

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

The Y axis is the signal of the AT channel, the X axis is cn-probe-chrXY-ratio_gender_ratio, as shown in Figure 234.
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 gender sample (blue) is in between the two threshold values.

The four gender 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_gender_meanX**: The average probe intensity (raw, untransformed) of X chromosome non polymorphic probes.

- **cn-probe-chrXY-ratio_gender_meanY**: The average probe intensity (raw, untransformed) of Y chromosome non polymorphic probes.

- **cn-probe-chrXY-ratio_gender_ratio**: Gender ratio $Y/X = \frac{cn\text{-}probe\text{-}chrXY\text{-}ratio\_gender\_meanY}{cn\text{-}probe\text{-}chrXY\text{-}ratio\_gender\_meanX}$

- **cn-probe-chrXY-ratio_gender**: Designated gender based on cn-probe-chrXY-ratio_gender_ratio

A Sample QC computed gender call will not be made if the sample fails DQC. The QC computed gender column will be blank. If a sample passes DQC but fails QC Call Rate, QC computed gender will be assigned. A sample must pass Sample QC to have a computed gender assigned.
### Scatter plot of Signal vs. \texttt{cn-probe-\textit{chr}XY-ratio\textunderscore gender\_ratio}

The Y axis is the signal of the AT channel, the X axis is \texttt{cn-probe-\textit{chr}XY-ratio\textunderscore gender\_ratio}. Samples are colored by Computed Gender and shaped by \texttt{cn-probe-\textit{chr}XY-ratio\textunderscore gender}, as shown in Figure 235.

Figure 235 is an example gender calling of PMRA batch that contains samples that failed Sample QC and thus given N/A gender assignment. Samples displayed below red line have failed Sample QC and are not given Computed Gender call.

### Scatter plot of signal vs. \texttt{cn-probe-\textit{chr}XY-ratio\textunderscore gender\_ratio} for 1152 BovMDv3 samples

Female samples, in red have \texttt{cn-probe-\textit{chr}XY-ratio\textunderscore gender\_ratio} of 1.1 or less. Male samples, in yellow, have \texttt{cn-probe-\textit{chr}XY-ratio\textunderscore gender\_ratio} of 1.2 or more. Unknown samples are in blue as shown in Figure 236.
Figure 236   Scatter plot of signal vs. cn-probe-chrXY-ratio_gender_ratio for 1152 BovMDv3 samples example