Agilent CytoGenomics 1.0

Reference Guide

Research Use Only. Not for Diagnostic Procedures

Agilent Technologies

Notices

Manual Part Number

G2602-90005

Edition

Revision A0, November 2010

Agilent Technologies, Inc.
5301 Stevens Creek Blvd.
Santa Clara, CA 95051 USA

Trademarks

Microsoft ® is a registered trademark of Microsoft Corporation in the United States and other countries.

Adobe®, Adobe®Acrobat® and Adobe®Reader® are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Software Revision

This guide is valid for the Agilent CytoGenomics 1.0 software and later revisions, until superseded.

Warranty

The material contained in this document is provided "as is," and is subject to being changed, without notice, in future editions. Further, to the maximum extent permitted by applicable law, Agilent disclaims all warranties, either express or implied, with regard to this manual and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Agilent shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein. Should Agilent and the user have a separate written agreement with warranty terms covering the material in this document that conflict with these terms, the warranty terms in the separate agreement shall control.

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

Restricted Rights Legend

U.S. Government Restricted Rights. Software and technical data rights granted to the federal government include only those rights customarily provided to end user customers. Agilent provides this customary commercial license in Software and technical data pursuant to FAR 12.211 (Technical Data) and 12.212 (Computer Software) and, for the Department of Defense, DFARS 252.227-7015 (Technical Data - Commercial Items) and DFARS 227.7202-3 (Rights in Commercial Computer Software or Computer Software Documentation).

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

In This Guide...

This guide describes the tabs, windows, parameter panels, dialog boxes, and reports you see in the Agilent CytoGenomics 1.0 software. For information on how to use the software to set up and run workflows and examine results, see the *Setup and Data Review User Guide* and the *Running CytoGenomics Analyses User Guide*. For information on installing and administering the software, see the *Installation and Administration Guide*.

1 Window and Command Ribbons Reference

This chapter gives descriptions of the tabs, windows, and command ribbons that appear within Agilent CytoGenomics 1.0.

2 Parameter Panels and Dialog Boxes

This chapter describes the parameter panels and dialog boxes that appear. They are listed in alphabetical order.

3 Reports

This chapter describes the default reports provided in the program. The Cyto Report can be customized by the user and saved as a new report template to use for reporting results of a workflow. For more information, see the *Setup and Data Review User Guide*.

4 Statistical Algorithms

This chapter describes the statistical algorithms used to generate results in Agilent CytoGenomics 1.0.

Contents

1 Window and Command Ribbons Reference

This chapter describes the tabs, commands, and views for all user roles in Agilent CytoGenomics 1.0.

Windows and Tabs

In CytoGenomics, functional tasks that are typically performed together are located together, under tabs at the top of the program window. You only see the tabs that are allowed for your assigned user role (Technician, Scientist, Administrator.) The task window changes depending on the tab and command currently selected. Figure 1 shows the window that appears when the program is opened by an Administrator. This window contains the elements listed in Table 1 below. More detailed descriptions of each element appear later in this section.

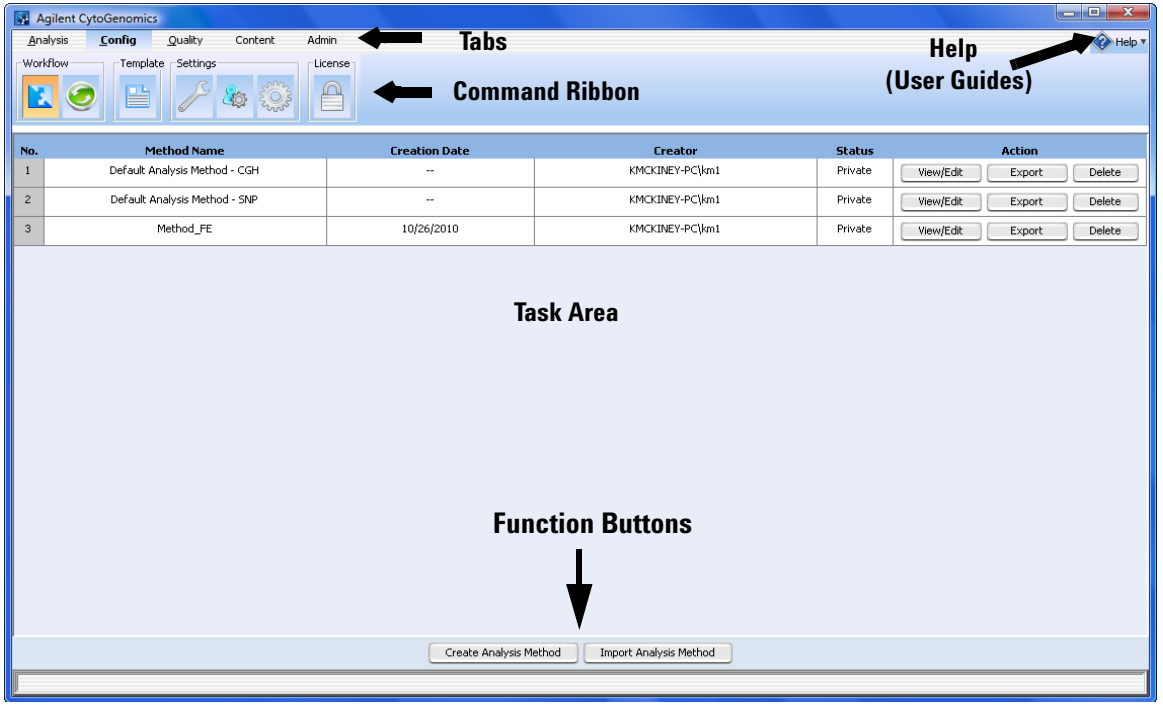


Figure 1 Window with Config tab selected (Administrator role)

Table 1 Window Elements

Element	Purpose
Tabs	Used to change functional areas of Agilent CytoGenomics 1.0. The tabs that appear depend on the role of the current user.
Command Ribbon	Contains the command buttons for the tasks available in the selected tab. To see the command button name, move the mouse cursor over the command button. A tooltip appears with the command name.
Help	Opens a menu of user guides available for the program. The user guides include: • Product Overview • Installation and Administration (Administrator tasks) • Setup and Data Review (Scientist tasks) • Running CytoGenomics Analyses (Technician tasks) • Reference Guide (this document) • About - shows the program version and license agreement
Task Area	The contents of the task area change depending on the command selected in the Command Ribbon.
Function Buttons	Function buttons appear at the bottom of the task area. Appearance of task buttons, and their purpose, change depending on the command selected in the Command Ribbon.

CytoGenomics Tabs

Table 1 shows the tabs that appear in Agilent CytoGenomics 1.0 and describes the tasks they contain.

Table 2 CytoGenomics Tabs

Tab	Tasks
Analysis (Available to Technician, Scientist, Administrator)	• Select and run workflows • View workflow jobs • Search for and display results and reports • Sign off results (Scientist and Administrator only)
Config (Available to Scientist and Administrator)	• Configure analysis methods and workflows • Create and save report templates • Set up preferences for viewing tracks and genomic boundaries • Set up view preferences for genomic viewer • Set up default data folders • Display or set program license
Quality (Available to Scientist and Administrator)	• Import QC data • Create and run search queries • Create metrics • Display, create, and edit metric sets • Display, create, and edit quality charts
Content (Available to Scientist and Administrator)	• Set up sample lists and assign sample attributes • Import designs • Import genotype references (required for SNP analysis)
Admin (Available only to Administrator)	• Add users and set their roles • Change database configuration

Command Ribbons

The commands are divided into groups, located at the top of the window directly below the tabs. The following sections describe the commands that are available for the selected tab.

NOTE Move the mouse cursor over a command icon in the command ribbon to see its name.

Command ribbon for Analysis

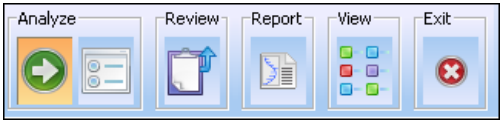


Figure 2 Command ribbon for the Analysis tab

The commands in this ribbon are described in the following table, in order from left to right. The name of the command appears when you move the mouse over the command icon in the command ribbon.

Table 3 Commands for Analysis command ribbon

Command	Purpose
Analyze	Select and run workflows
Job Monitor	Display status of workflows. Cancel and delete workflows. Review and sign off workflow results.
Review	Display sample list. Open samples in Triage View to review and sign off results.
Report	Search for samples and display cyto reports.
Multi Sample View	Search for samples and display results in the View Aberrations window.
Exit	Closes the Agilent CytoGenomics 1.0 program.

Command ribbon for Config

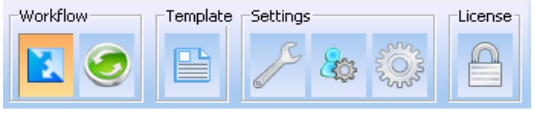


Figure 3 Command ribbon for the Config tab

The commands in this ribbon are described in the following table, in order from left to right. These commands are not available to users with the role of "Technician." The name of the command appears when you move the mouse over the command icon in the command ribbon.

Table 4 Commands for Config command ribbon

Command	Purpose
Analysis Method	Create, edit, and publish analysis methods.
Workflow	Create, edit, and publish workflows.
Report Template	Create, edit, and display Cyto Report templates.

Command	Purpose
Preferences	Select and configure tracks and genomic boundaries.
View Settings	Set up default viewing preferences and data display for the Genomic Viewer.
Default Settings	Set up default locations for reports, designs, and arrays.
Licenses	Display or enter license information.

Command ribbon for Quality

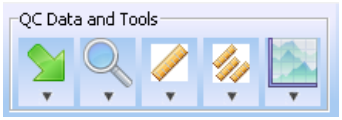


Figure 4 Command ribbon for the Quality tab

The commands in this ribbon are described in the following table, in order from left to right. These commands are not available to users with the role of "Technician." The name of the command appears when you move the mouse over the command icon in the command ribbon.

Table 5 Commands for Quality command ribbon

Command	Purpose
Import File	• Import Feature Extraction stats and parameters. • Import query results.
Queries	Create, change, delete, rename, export, and run queries.
Metrics	Create and delete metrics.
Metric Sets	Display, create, change, rename, import, and export metric sets.
Chart	Display, create, change, delete, and rename QC charts.

Command ribbon for Content

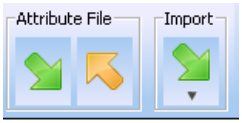


Figure 5 Command ribbon for the Content tab – Sample Manager

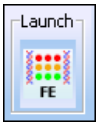


Figure 6 Command ribbon for the Content tab – Feature Extraction

The commands in this ribbon are described in the following table. These commands are not available to users with the role of "Technician." The name of the command appears when you move the mouse over the command icon in the command ribbon.

Table 6 Commands for Content command ribbon

Command	Purpose
Import Attribute File	Import an attribute file.
Export Attribute File	Export an attribute file.
Import	Import designs and genotype references.
Launch Feature Extraction	Starts the Feature Extraction program.

Command ribbon for Admin

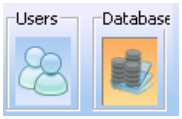


Figure 7 Command ribbon for the Admin tab

The commands in this ribbon are described in the following table. These commands are not available to users with the role of “Technician” or “Scientist.” The name of the command appears when you move the mouse over the command icon in the command ribbon.

Table 7 Commands for Admin command ribbon

Command	Purpose
Manage Users	Add users and set their roles.
Database	Display and change database configuration.

Analysis Tab Windows

This section describes the windows that appear when you select commands in the Analysis tab. These tasks are typically performed by users with the role of Technician.

Analyze window

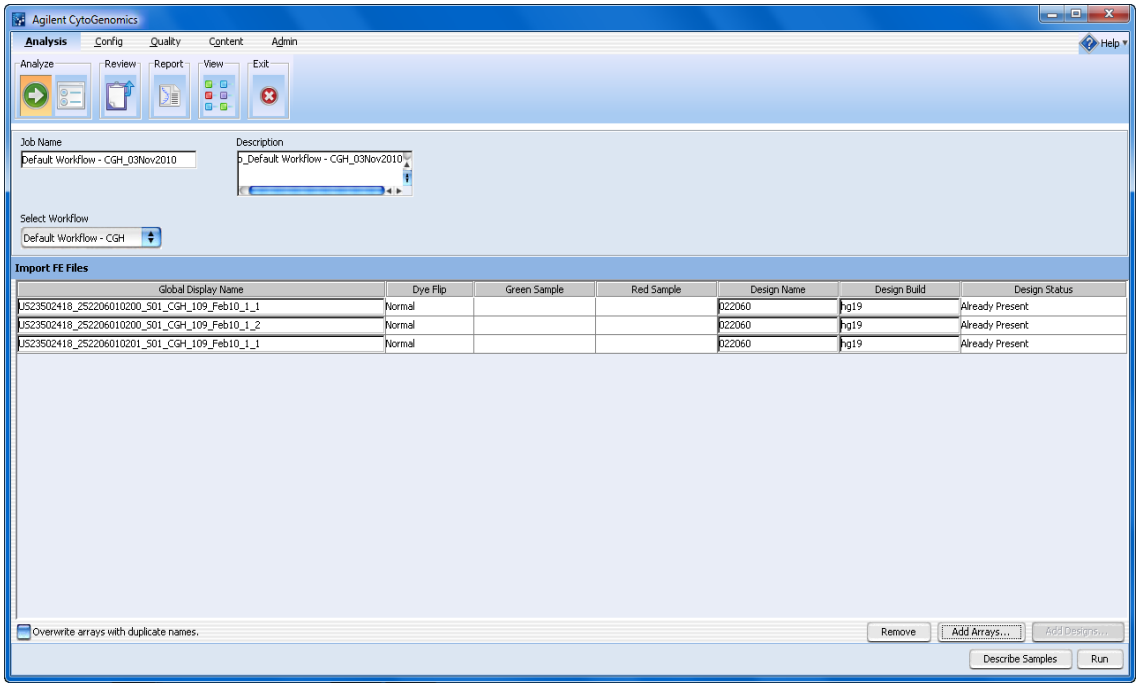


Figure 8 Analyze window

Purpose: To select a workflow and select input files.

To open: In the Analysis tab, under Analyze, click **Analyze**.

General window buttons

The type of files you select in the Analyze window and the required parameters change, depending on the input required for the selected workflow. The items in this section are available for all workflow types.

Job Name Type the name you want to use for the workflow job. The job name appears as an identifier for the workflow run throughout Agilent CytoGenomics 1.0

Description Type a description for the workflow.

Select Workflow Click the arrow and select from the available workflows. You only see workflows that you created or that were published by another user. The task window and function buttons change depending on the type of input required by the workflow.

Describe Samples Opens the Describe Samples table, where you can show or hide attributes, or change attributes for the samples in the workflow. From this view, you can click **<< Back** to go back to the workflow, or click **Run** to run the workflow. See "Describe Samples window" on page 28.

Run Starts the workflow.

Import FE Files

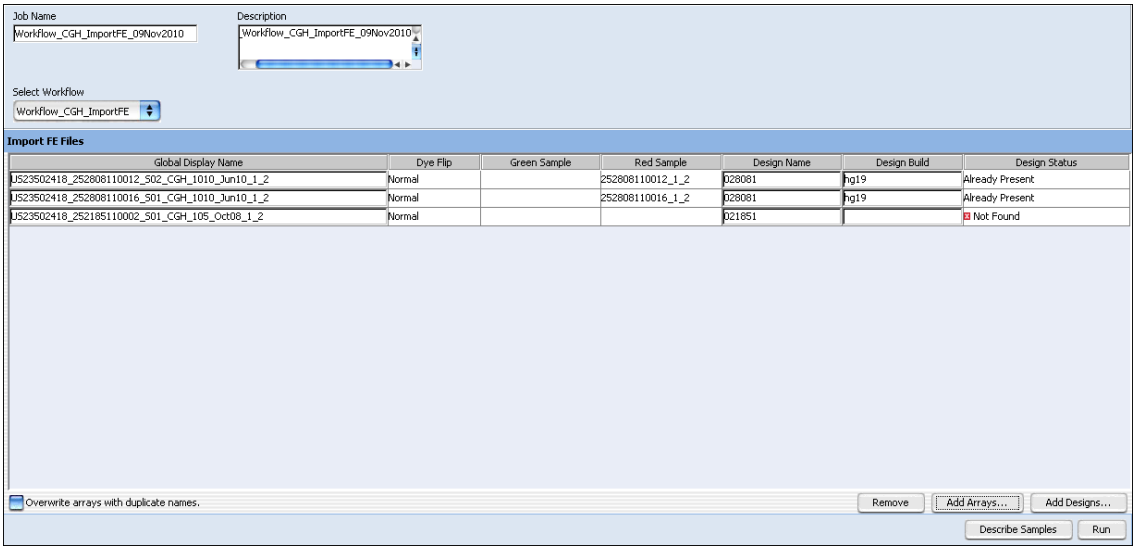


Figure 9 Analyze Window – Import FE Files

Purpose: Used to select extracted Feature Extraction files to use for workflow that uses imported FE files as the input.

To open: In the Analysis Window, select a workflow that requires FE files as input.

Global Display Name Lists the global display names of array data files selected for import. To add files, click **Add.** To edit the global display name of a file, double-click the name, type the name, then press **Enter.**

Dye Flip Lets you annotate an array as a dye-flip array (an array in which the Cy3 and Cy5 fluorochrome-labeled samples have been reversed). If you select **Flipped,** the program inverts ratios so you can make proper comparisons. Also, the program does not combine dye-flip pairs.

Select one of these options for each array:

Select this option	If:
Normal	• The test samples were labeled with cyanine 5 (red). • The control samples were labeled with cyanine 3 (green). • The imported ratio (test/control) should be reported directly.
Flipped	• The test samples were labeled with cyanine 3 (green). • The control samples were labeled with cyanine 5 (red). • The imported ratio (control/test) should be reported with the ratio inverted (test/control)

Green Sample For normal (non-flipped) CGH+SNP samples, click the arrow and select a genotype reference.

Red Sample For flipped CGH+SNP samples, click the arrow and select a genotype reference.

Design Name The design name for the array is displayed in this field.

Design Build The design build for the array is displayed in this field. If this field is blank, you need to add the design before you can run a workflow that imports data files.

Design Status This field displays the status of the design file.

- Not Found - The design file is not currently in the program or a path has not been selected.
- Already Present - The design file is in the program database.
- Path Provided - The design file was not imported, but a location was selected for the file.

Overwrite arrays with duplicate names If you select this option, the workflow deletes an existing array if it has the same name as one you import.

Remove Highlight one or more samples and then click this button to remove them from the list. To select more than one sample, hold down the **Ctrl** key and then click the files to remove. To select a contiguous series of samples, hold down the **Shift** key and click the first sample and then click the last sample.

Add Arrays Opens a dialog box where you select the extracted feature extraction (.txt) files to use for the workflow.

Add Designs If the design that matches a sample is not in the database, the Design Status shows "Not Found" for the sample, and this button becomes active. Click to open the Open dialog box where you select the design for the sample. After you select the design, the Design Status shows "Path provided." When you run the workflow, the design is imported from this location.

Select Imported Data

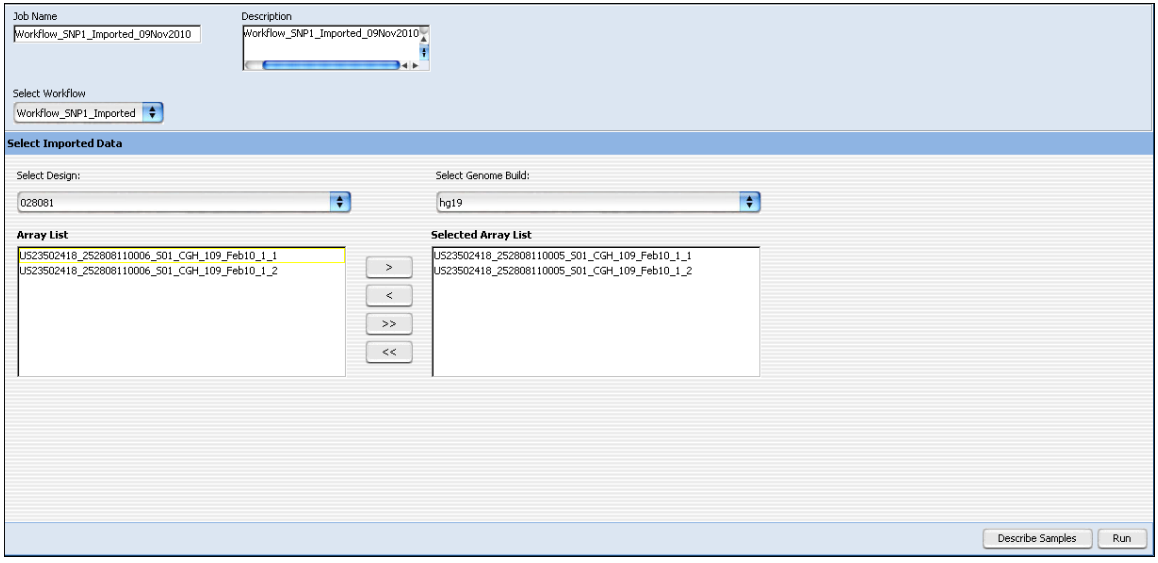


Figure 10 Analysis Window – Select Imported Data

Purpose: Used to select data already present in the database to use as input for the workflow.

To open: In the Analysis Window, select a workflow that requires imported data for the input.

Select Design Displays the array designs available in the database. Select the design from the list. The arrays for the design appear in the Array List. Although the program displays the arrays from one design at a time, you can add arrays from more than one design to the Selected Array List. The genome build must be the same for all selected arrays.

Select Genome Build If the design you select in Select Design has arrays from more than one genome build, select the desired genome build.

Array List Displays the arrays for the selected design and genome build.

Selected Array List Displays the arrays the program uses as input for the workflow.

> Moves selected arrays from the Array List to the Selected Array List. To select an array, click its name. To select additional arrays, hold down the **Ctrl** key and click their names. To select a block of arrays, click the name of the first array in the block, then hold down the **Shift** key and click the last one.

< Moves selected arrays from the Selected Array List to the Array List.

>> Moves all of the arrays in the Array List to the Selected Array List.

<< Clears all of the arrays from the Selected Array List, and restores them to their original location(s).

Import Image Files

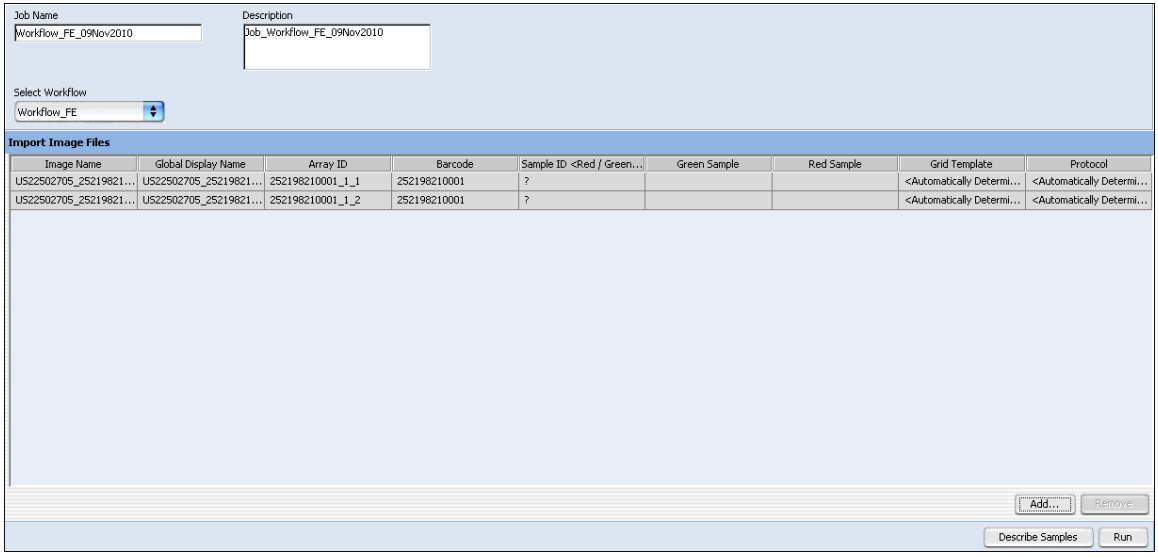


Figure 11 Analysis Window – Import Image Files

Image Name (Read-only) The file name of the image.

Global Display Name The Global Display Name for the array. To change the Global Display Name, double-click the name and type the new name.

Array ID (Read-only) The unique identifier for the microarray.

Barcode (Read-only) This is the barcode identifier for the slide. A microarray slide contains one array, or for multiplex slides, multiple arrays on one slide.

Sample ID <Red/Green (ArrayID)> If the Red/Green attributes for the array are assigned in Sample Manager, they appear here. Otherwise, a ? appears in the field. Click to select the genotype reference (for CGH+SNP arrays.)

Grid Template By default, the Feature Extraction program automatically determines the Grid Template for the microarray. Or, you can select a grid file (.grd) from your hard disk.

Protocol By default, the Feature Extraction program automatically determines the Protocol for the microarray. Or, you can select a Protocol from the list of available protocols. The program automatically determines the protocol, based on the following:

- If the grid template has a protocol associated with it, that protocol will be assigned to the workflow.
- If the grid template has no protocol associated with it, the protocol will be assigned based on the application type (CGH, for example).
- If no proper application type is specified in grid template, the default type is assumed to be Expression, and the corresponding protocol is associated, depending on whether the image is one-color or two-color.

NOTE If the user selects a protocol or grid template, and there is a newer version of the protocol or grid template in the database, the program will use the protocol or grid template selected by the user, and not the newer version. To use the newer version, the user must select it or let the program automatically determine the protocol.

Add Click this button to open the Open dialog box, where you can select the image file(s) to be added. See "Open" on page 226.

Remove Highlight one or more microarrays and then click this button to remove them from the list. To select more than one microarray, hold down the **Ctrl** key and then click the microarrays to remove. To select a contiguous series of microarrays, hold down the **Shift** key and click the first microarray and then click the last microarray.

Describe Samples window

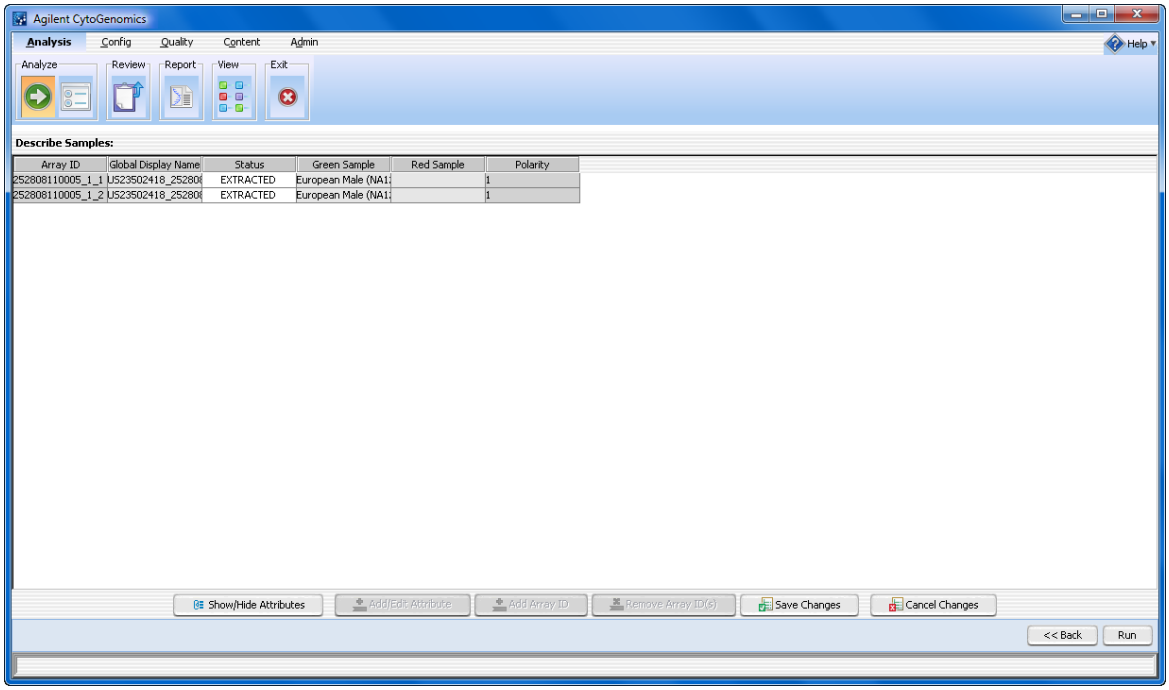


Figure 12 Describe Samples window

Purpose: Used to display and change sample attributes or Global Display Name for samples in the current workflow.

To open: In the Analysis tab, click **Analyze**. Select a workflow and the data files to use, then click **Describe Samples**.

Show/Hide Attributes Opens the Show/Hide Columns dialog box, where you can select attributes to show in the sample table. Clear the check box for attributes you do not want to show. See "Show/Hide Columns" on page 242.

Save Changes If you make a change in the table, this saves the changes.

Cancel Changes If you make a change in the table, this cancels the change.

<<Back Go back to the workflow Analyze window.

Run Starts the workflow run.

Job Monitor window

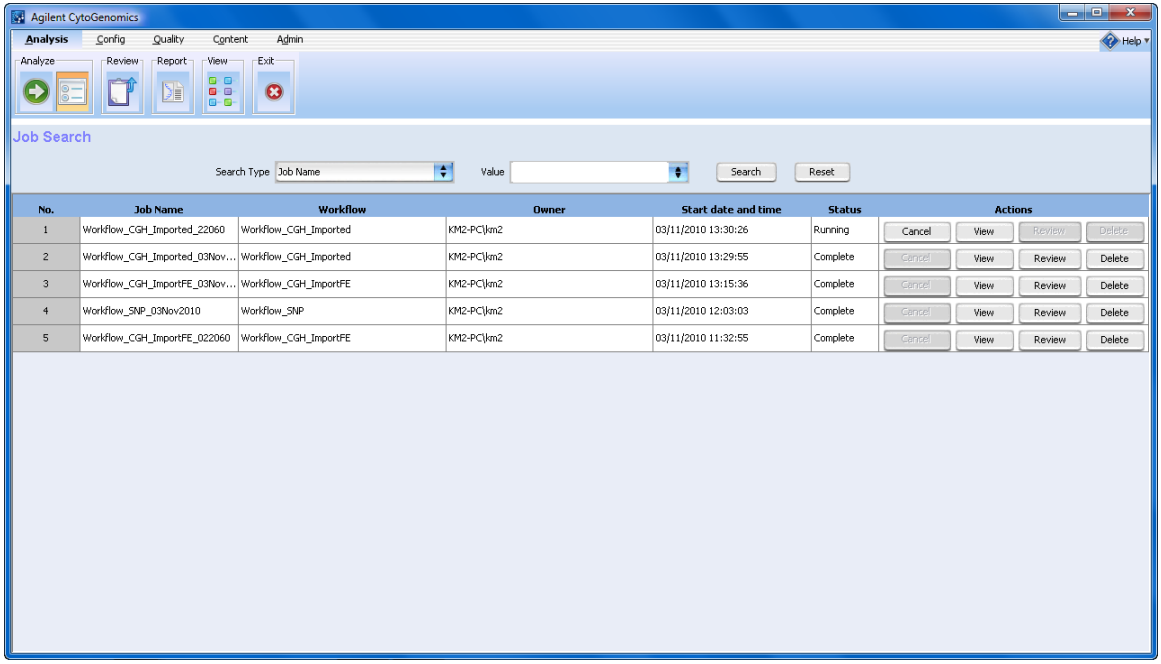


Figure 13 Job Monitor window

Purpose: To show and manage workflow runs.

To open: In the Analysis tab, under Analyze, click **Job Monitor**.

Job Search

Use this area to search for specific workflow jobs.

Search Type Click to select what field you want to search:

- Job Name
- User
- Job Status

Value Type the Job Name or select a User or workflow Job Status from the list.

Search Click to search the available jobs and display jobs that match the search conditions.

Reset Click to show all jobs.

Cancel This button is active for jobs that are not finished. Click to cancel a workflow in progress.

View Displays a summary of the selected workflow job.

Review Opens the Analysis Review window, where you can see a list of analyzed microarray samples for the selected job. From here, you can click a sample's Status to open the Triage View to review and sign off results. See "Triage View" on page 59.

Delete Deletes the selected workflow job.

Review window

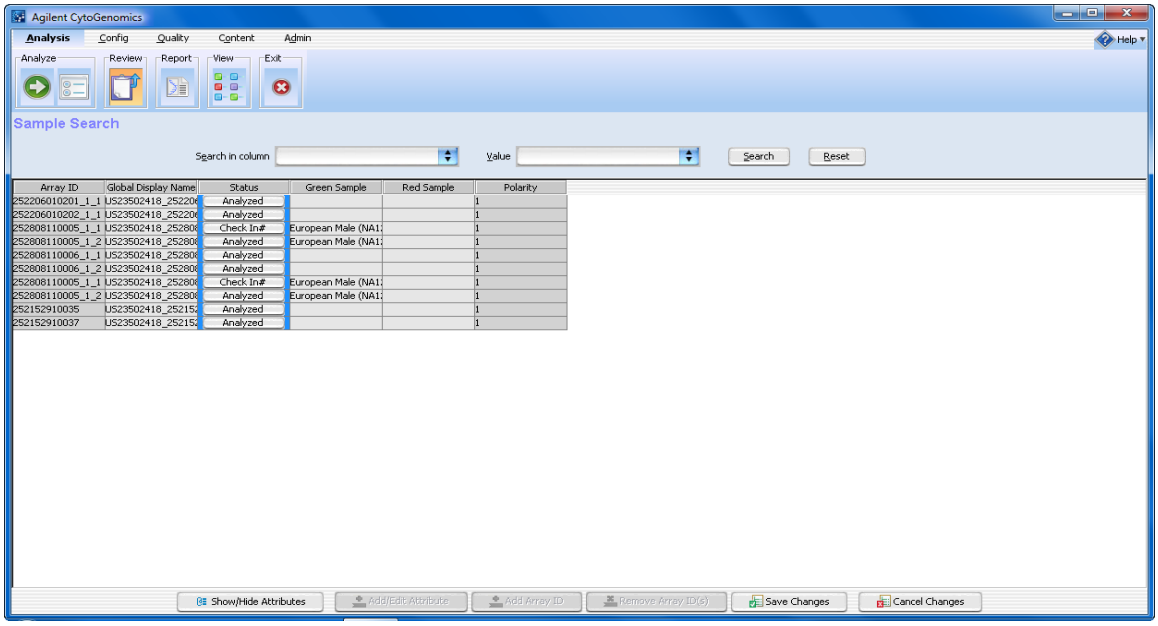


Figure 14 Review window

Purpose: To show microarray samples and their status, and open Triage View to display and sign off on results. See "Triage View" on page 59.

To open: In the Analysis tab, click **Review**.

Sample table

Samples that were analyzed in the selected job are displayed in the sample table. You can change global display names and attribute values, and save the changes. You can also show or hide attributes in the sample table. The Status column shows the status of each sample. To open a sample in Triage View, click its status. The possible sample status and meanings are shown in the following table.

Status	Meaning
Analyzed	Workflow was completed and the sample was analyzed. The sample was not checked in or checked out yet.
Check Out	The sample is checked out. You cannot open a sample that is checked out unless you are the user who checked it out.
Check In	The sample was checked out and then checked in. You can open the sample in Triage View.
Reviewed	The sample was signed off. You can open the sample in Triage View, but you cannot check out, make changes, or check in the sample.
"Status#"	The sample status has # next to it for samples that have more than one record with different status. For example, a sample was analyzed twice but only one record was reviewed and the other has the status of 'Analyzed'. In this case the status field in the content tab will display 'Reviewed#'.

Show/Hide Attributes Opens Show/Hide Columns dialog box where you can select which attributes to display in the sample table. See "Show/Hide Columns" on page 242.

Save Changes Saves changes made to the array list.

Cancel Changes Cancels changes made to the array list.

Sample Search

Search in column Click to select a column in the table to use for the search.

Value Click to select a value for the selected search column.

Search Click to search the sample table for the array(s) that match the search conditions.

Reset Click to show all samples in the table.

Report window

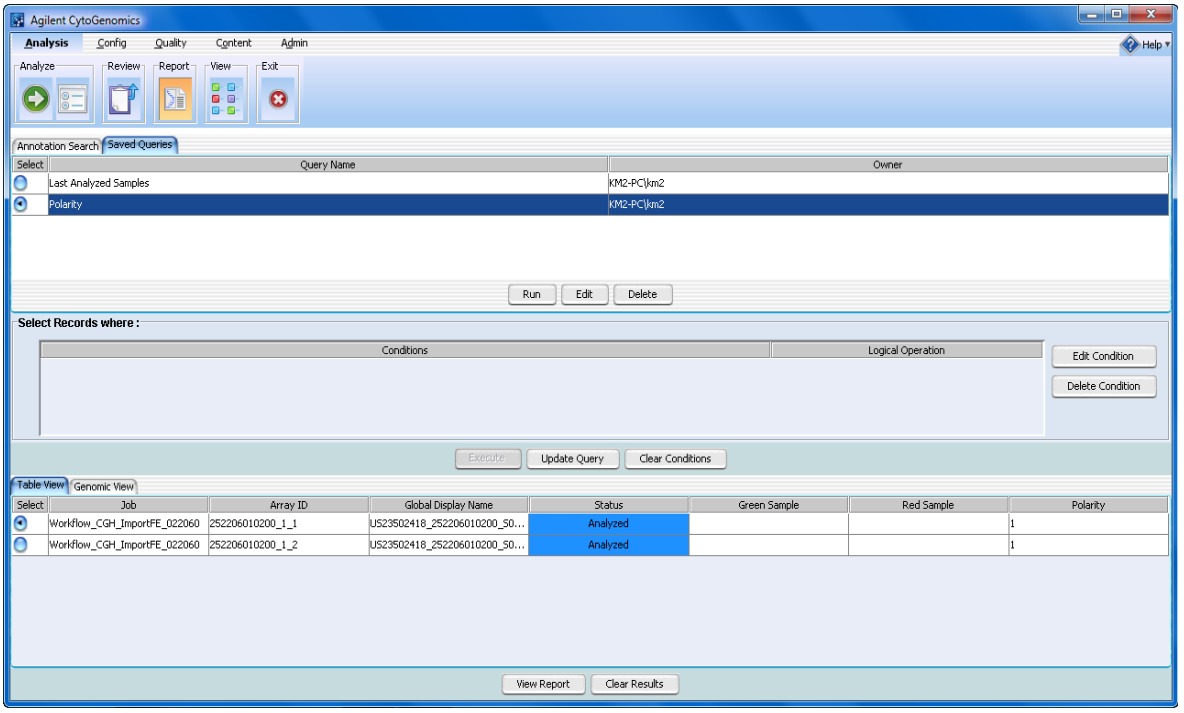


Figure 15 Report window

Purpose: Used to search for analyzed samples and display their cyto reports.

To open: In the Analysis tab, click **Report**.

This window contains two tabs that enable you to search for samples based on search parameters you enter or using saved queries.

NOTE Only samples with cyto reports are shown for search results.

General window commands

View Report Lets you select a report for the selected sample, then opens it in the Manage Cyto Report dialog box, where you can examine the report page-by-page. See "Manage Cyto Report" on page 213.

Clear Results Clears all search results from the Table View and Genomic View.

Annotation Search tab

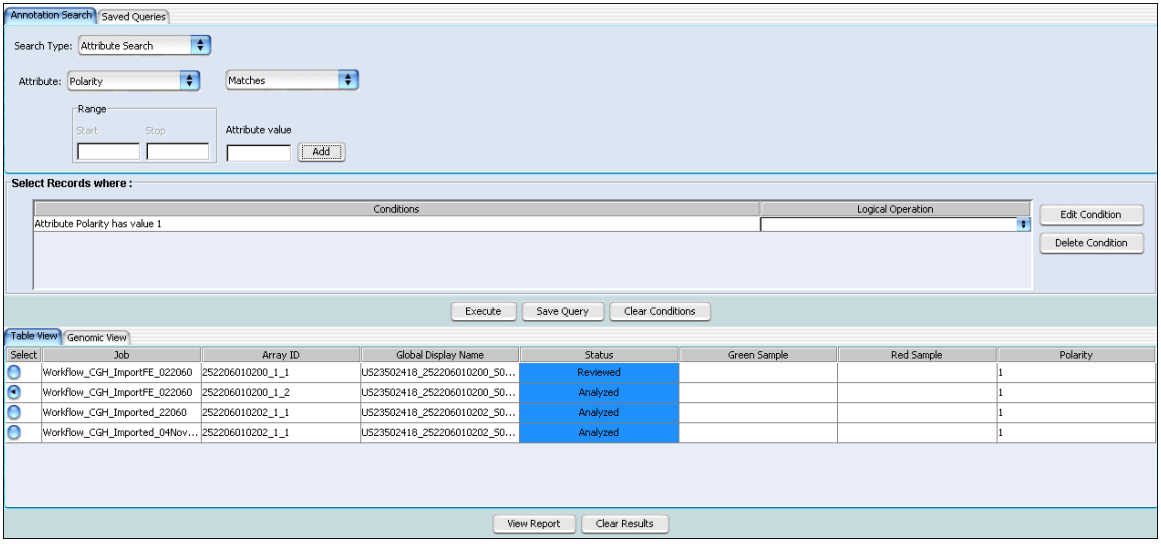


Figure 16 Annotation Search tab

Use this tab to search for samples based on search conditions you specify, and to create and save queries.

Search Type Click to select the search type. Select from the following options:

Option	Purpose
Simple Search	Lets you search for samples based on a string that you type.
Attribute Search	Lets you search for samples based on a selected attribute.

Simple Search

String Search Type a string that you want to use for the search. The search finds and shows all samples that contain the string.

Attribute Search

Attribute Click to select the attribute you want to use for the search. Next to the selected attribute, select how to match the attribute:

- Matches – Finds all samples with attributes that match the Attribute value exactly.
- Is in range – Finds all samples with attributes that fall within the Range.

Range If you select Is in range, type a start and stop range for the attribute.

Attribute value If you select Matches, type a value for the search attribute.

Add Click to add the search condition to the list of conditions.

Saved Queries Tab

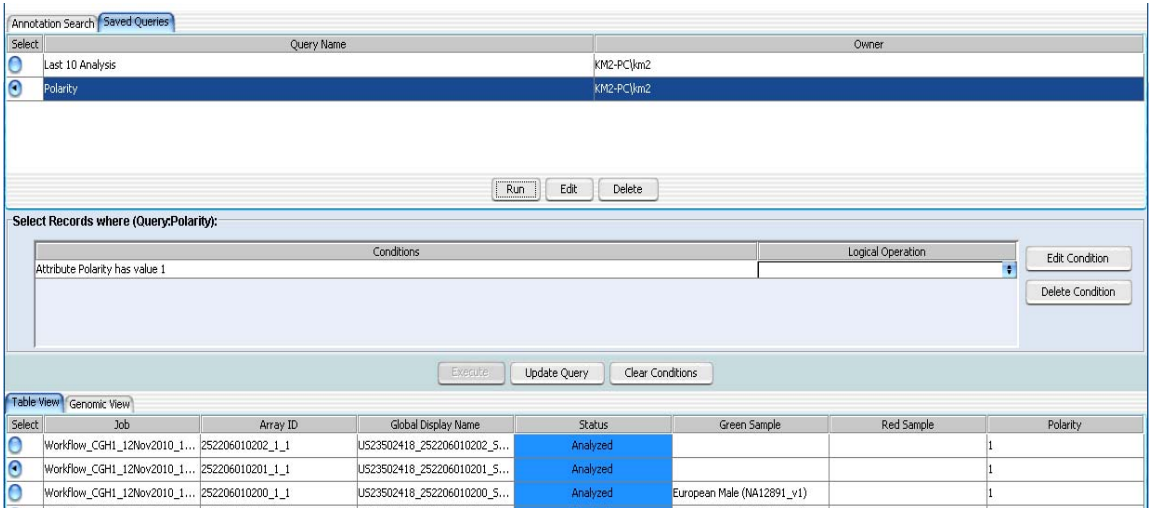


Figure 17 Saved Queries tab

Lets you select and run a sample search using a saved query.

Run Runs the selected query.

Edit Opens the query search conditions so that you can change them.

Delete Deletes the selected query.

Select Records where

This table contains the list of conditions for the search or query.

Logical Operation Select a logical operation if you want to add another condition for the search. For example, if you select **AND** in **Logical Operation** for the first condition, the search includes a sample if it matches both the first condition and the next condition. If you select a logical operation, you must add another condition to the list.

Edit Condition Click to show the selected condition and make changes to it. When you finish making changes, click Add to save the changes.

Delete Condition Deletes the selected condition.

Execute Runs the current search.

Save Query Saves the current search parameters as a query.

Clear Conditions Clears the conditions in the conditions list.

Table View

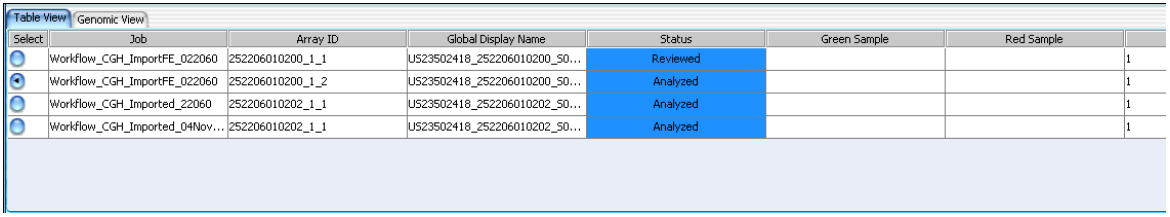


Figure 18 Table View

Shows results of the search in table form. Mark **Select** next to samples for which you want to show the report.

Genomic View

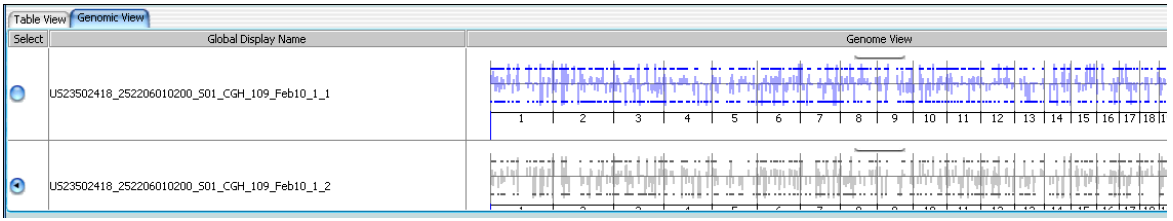


Figure 19 Genomic View tab

Shows the results of the search in a graphical view. Each sample is displayed as a graphical depiction that indicates aberrations. Mark Select next to samples for which you want to show the report.

Multi Sample View window

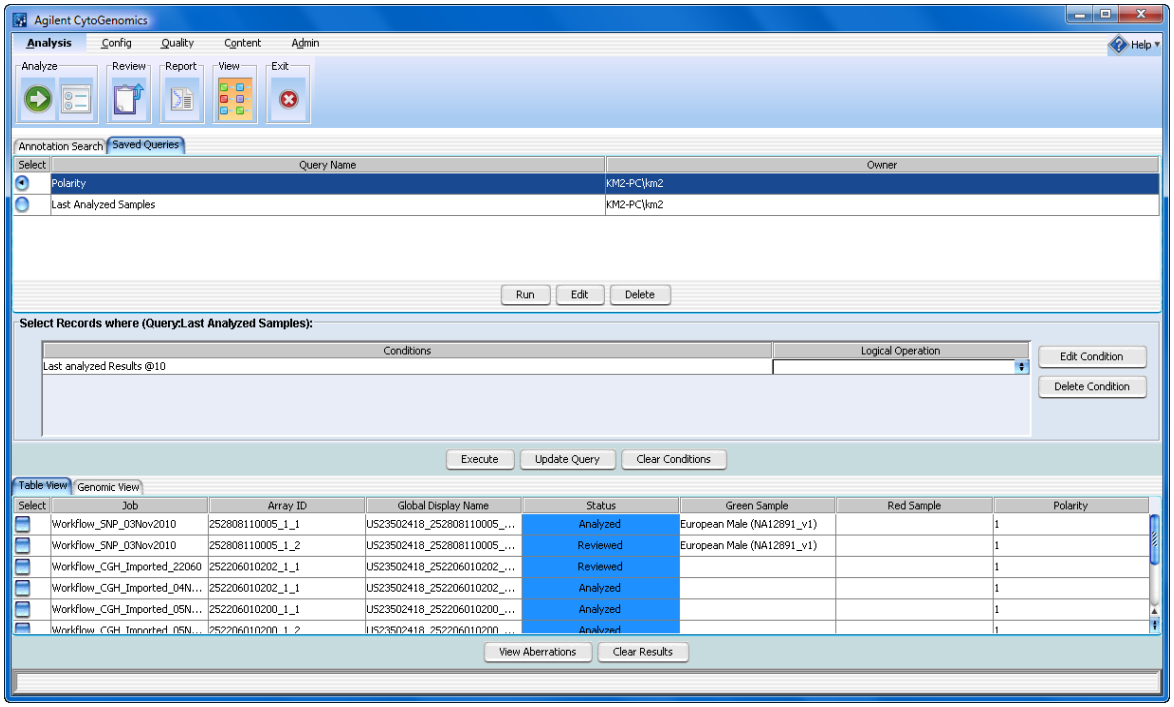


Figure 20 Multi Sample View window

Purpose: Used to search for analyzed samples and show the aberration results.

To open: In the Analysis tab, click **Multi Sample View**.

This window contains two tabs that enable you to search for samples based on search parameters you enter or using saved queries. For information on how to set up searches and queries, see "Annotation Search tab" on page 34, "Simple Search" on page 35, "Attribute Search" on page 35, "Saved Queries Tab" on page 35, and "Select Records where" on page 36.

NOTE Only samples from completed workflows are shown for search results.

View Aberrations Opens the View Aberrations window, with the aberration results for the selected samples. See "View Aberrations Window" on page 54.

Clear Results Clears the results of the sample search from the Table View and Genomic View.

Genomic Viewer

Genomic Viewer is the display of genomic results within Agilent CytoGenomics 1.0. It includes the three views (Genome, Chromosome and Gene) along with a Tab View that contains tables of data and results. The Genomic Viewer is displayed in both the Triage View and the View Aberrations windows. You can also include views from the Genomic Viewer in Cyto Reports.

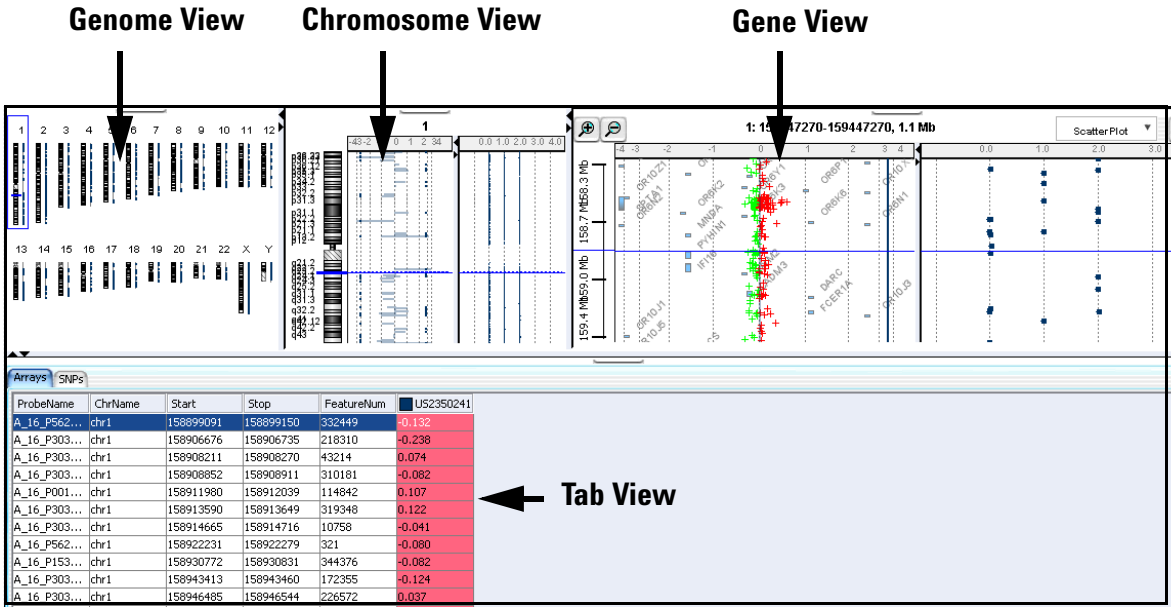


Figure 21 Components of the Genomic Viewer

Genome View

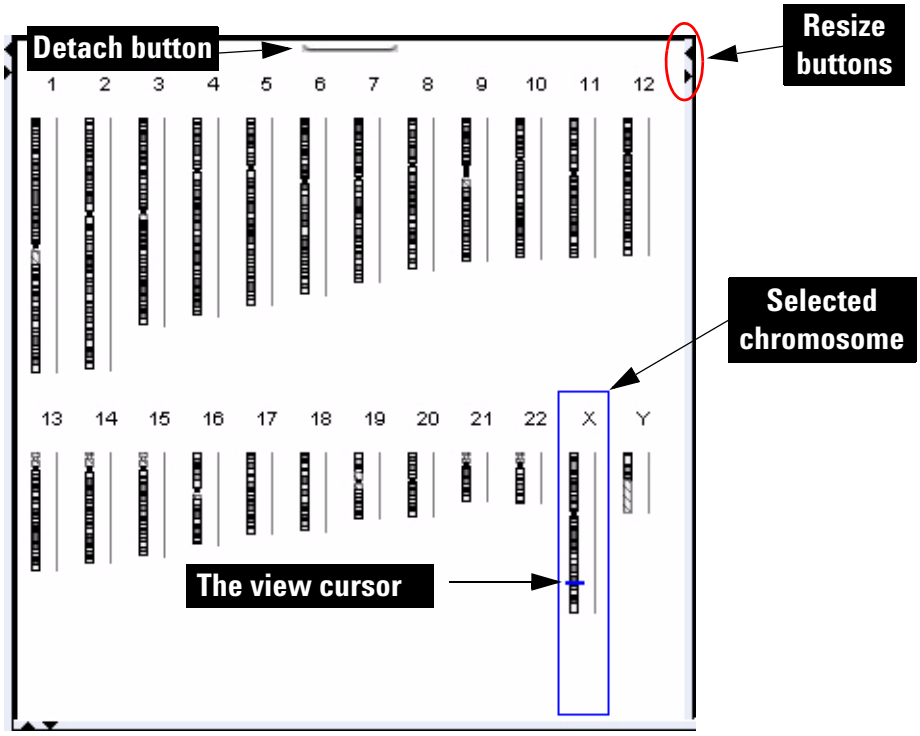


Figure 22 Genome View (vertical orientation), with human chromosomes. The X chromosome is selected.

Genome View shows pictures of each of the distinct types of chromosomes in the selected genome. A blue box encloses the selected chromosome, and the cursor appears as a blue line across the chromosome.

Genome View actions and shortcut menus

- Click a chromosome within the Genome View to select it. When you select a chromosome, Chromosome, Gene, and Tab Views show only genomic regions, genes, and data associated with it. The specific location in which you click the chromosome sets the position of the cursor. See "The View Cursor" on page 49.

- On the selected chromosome, click anywhere to reposition the cursor. See "The View Cursor" on page 49. This also repositions the cursor in Chromosome, Gene, and Tab Views.
- Right-click anywhere within Genome Views to display a menu. If you click **View Preferences,** the View Preferences dialog box opens, where you can set preferences for the display. See "View Preferences" on page 251.
- Click the **Detach** button (located at the top center of the view) to remove Genome View from the main window and open it in a separate window. To reattach the view, click its **Close** button  .
- Drag the side or bottom borders of the view to resize it.
- On a border of the view, click a resize button (for example,  or ) that points away from the view to move that border all the way to the edge of the main window. To move the border back to its previous location, click the other resize button.

Chromosome View

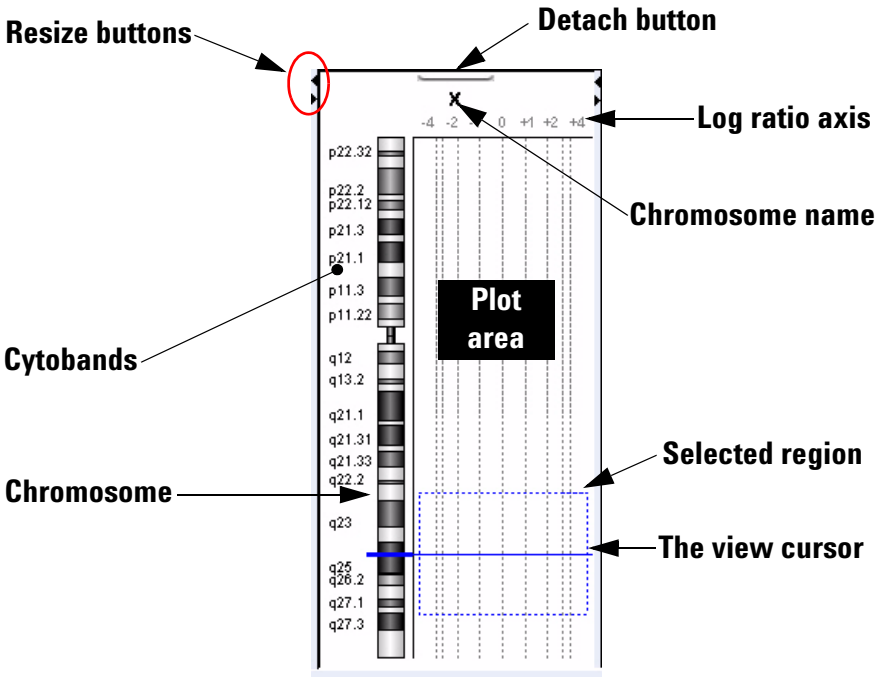


Figure 23 Chromosome View (vertical orientation), human X chromosome shown

Chromosome View shows a more detailed diagram of the chromosome you select in Genome View.

- Cytobands and a plot area appear next to the chromosome.
- When you select arrays for display, their data appear in the plot area.
- The cursor appears as a solid blue line across the chromosome and the plot area.
- The selected region of the chromosome (if any) appears as a dotted blue box in the plot area.

Chromosome View actions and shortcut menus

- Click a cytoband, any part of the chromosome, or anywhere in the plot area to reposition the cursor at that location. See "The View Cursor" on page 49.
- Drag the pointer over any part of the plot area to select a chromosomal region for display in Gene View. Drag parallel to the chromosome. This also repositions the cursor to the center of the selected region. See "The View Cursor" on page 49.
- Right-click anywhere within Chromosome View to display a menu. If you click **View Preferences,** the View Preferences dialog box opens, where you can set preferences for the display. See "View Preferences" on page 251.
- Click the **Detach** button (located at the top center of the view) to remove Chromosome View from the main window and open it in a separate window. To reattach the view, click its **Close** button  .
- Drag the side or bottom borders of the view to resize it.
- On a border of the view, click a resize button (for example,  or ) that points away from the view to move that border all the way to the edge of the main window. To move the border back to its previous location, click the other resize button.

Gene View

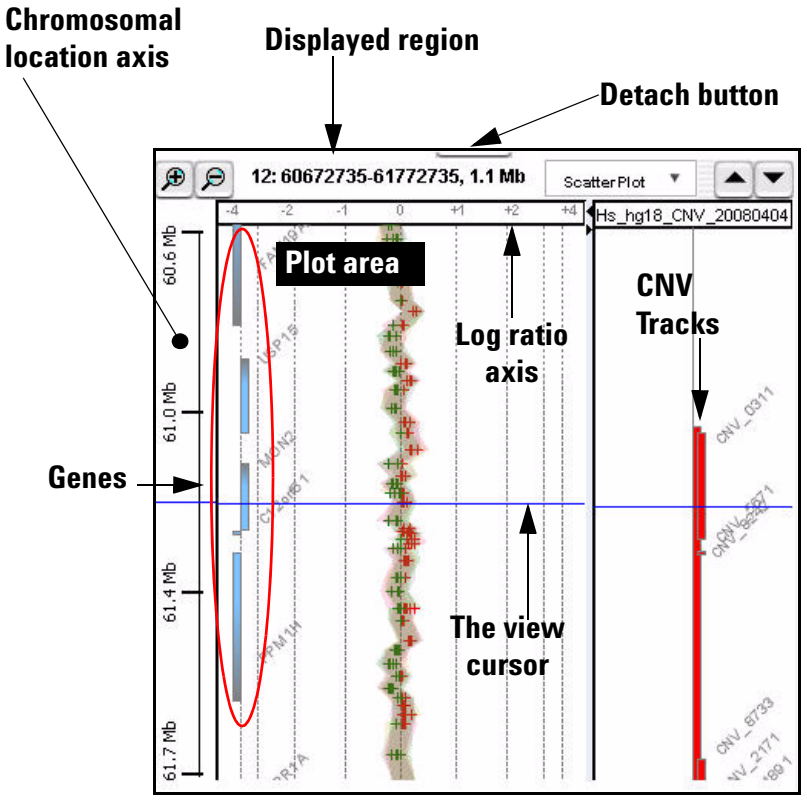


Figure 24 Gene View (vertical orientation), with log ratio data and CNV tracks

Gene View shows a more detailed view of the chromosomal region you select in Chromosome View. See "Chromosome View" on page 43.

- Regions occupied by genes appear as small blue boxes. Gene names appear nearby. You can customize the appearance of gene names. Also, you can use a gene list to highlight genes of interest, or to limit the genes that appear to those in the list.
- Log ratio data from selected arrays appear as a scatter plot. You can also customize the scatter plot.

- The location of the cursor matches the location of the cursors in other views. See "The View Cursor" on page 49.
- The name of the chromosome, and the coordinates and size of the displayed chromosomal region appear at the top of the view.
- Cytobands can also be displayed in Gene View.
- Imported tracks can also be displayed in Gene View.

Scatter Plot

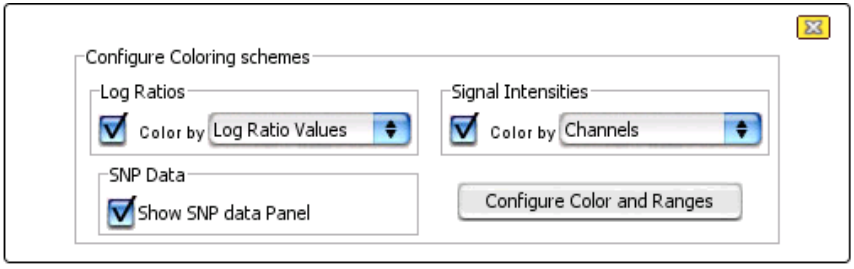


Figure 25 Scatter Plot command group in Gene View

You access the scatter plot command group in Gene View or View Preferences from the View tab. Scatter plot data appear in the Chromosome and Gene Views, but only if they have been selected under Data Visibility in the View Preferences dialog box. See "View Preferences" on page 251.

Log Ratios Select the check box to enable the Log Ratios scatter plot. Choices for the plot are Log Ratio Values or Probe Score Values.

Signal Intensities Select the check box to enable the Signal Intensities scatter plot. Selections for the plot are Channels, Probe Score Values, or Intensity Values.

SNP Data Select the check box to enable the SNP data panel that shows copy number and LOH data for CGH+SNP microarrays.

Configure Color and Ranges Opens the Configure Coloring Ranges and Shades dialog box, where you can set up the colors and ranges for Primary and Secondary scatter plots. For more information, see "Configure Coloring Ranges and Shades" on page 152.

Gene View buttons


Zooms in to see a smaller region in more detail.


Zooms out to see a larger region in less detail.


When in vertical orientation, scrolls up through the genes and data to lower-numbered chromosomal coordinates.


When in vertical orientation, scrolls down through the genes and data to higher-numbered chromosomal coordinates.


When in horizontal orientation, scrolls left through the genes and data to lower-numbered chromosomal coordinates.


When in horizontal orientation, scrolls right through the genes and data to higher-numbered chromosomal coordinates.


(**Resize** buttons) The button that points away from Gene View expands the view. The other button restores the view to its original size.(These buttons will appear horizontal if the view orientation is horizontal.)

(**Detach** button) Removes Gene View from the main window, and opens it in a separate window.

Gene View shortcut menu and other actions

- Click anywhere in the plot area of Gene View to move the cursor to that location. See "The View Cursor" on page 49.
- Drag an inside border of Gene View to resize the view.

Right-click anywhere in the plot area of Gene View to display these options:

Table 8 Gene View shortcut menu tasks

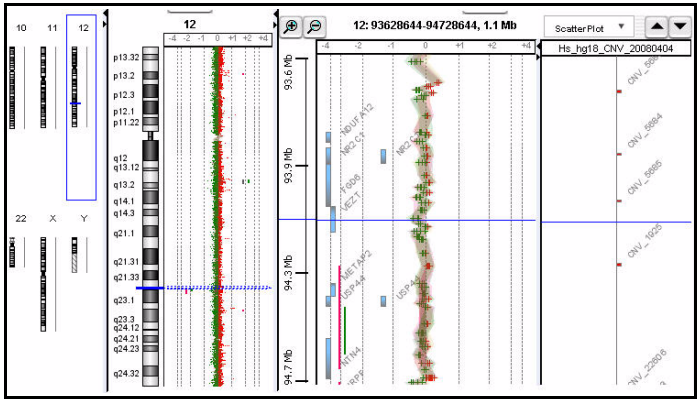
Command	Purpose
Gene View tasks	
Create Track	Opens the Create Track dialog box, where you select a name and the chromosome locations for a track. See "Create Track" on page 168.

Command	Purpose
Show Intensity Bar Charts	Opens the Create Signal Bar Chart dialog box, where you select parameters to create a signal intensity chart for the data. See "Create Signal Bar Chart" on page 167.
Show in UCSC	Opens the View Coordinates in UCSC Browser dialog box where you select track information for display in the UCSC (University of California at Santa Cruz) Genome Browser. You can then view the track. See "View coordinates in UCSC browser" on page 249.
User Preferences	Opens the User Preferences dialog box, where you import and set up the appearance of tracks next to the Gene View. Tracks are additional graphic displays of genomic information loaded from an external file. They align with genomic coordinates in Gene View. See "User Preferences" on page 247.
View Preferences	Opens the View Preferences dialog box, where you set preferences for the Genomic Viewer. See "View Preferences" on page 251.

The View Cursor

The View cursor reflects the center of the current chromosomal location of interest. It appears in several Views:

- In Genome View, it appears as a blue bar across the selected chromosome.
- In Chromosome View, it is a blue bar that appears across the chromosome and across the plot area of the view.
- In Gene View, it is a blue bar that appears across the plot area and tracks of the view.



The position of the cursor in one view is also the position of the cursor in all views. The exact chromosomal location of the cursor appears in the first cell of the Status bar. Several actions affect the position of the cursor:

- In Genome View, click anywhere on a chromosome to move the cursor to that location.
- In Chromosome View, click a cytoband name, part of the chromosome, or anywhere in the plot area to move the cursor to that location.
- In Gene View, click anywhere in the plot area to move the cursor to that location.

  The cursor used in Gene View is the same cursor used for the tracks.
- In Tab View, click a row of a data table to move the cursor to the chromosomal location associated with that row.

Tab View for Genomic Viewer

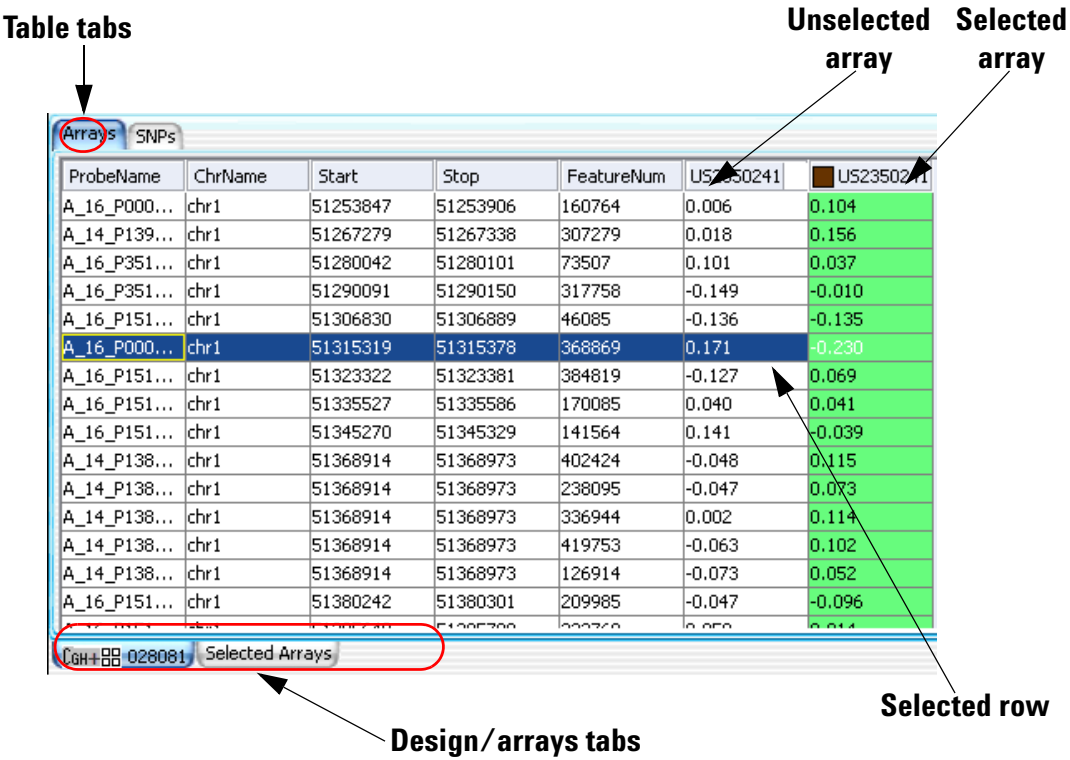


Figure 26 Tab View

Tab View displays design annotation and log ratio data related to the chromosome you select in Chromosome View. CGH probes are displayed in the Arrays tab. For CGH+SNP arrays, an additional SNPs tab is available with SNP data for the selected chromosome. This tab does not show results unless the SNP algorithms are selected and data was analyzed.

- The exact column content of the tables depends on the tab and design, but it always includes chromosomal locations of probes.
- The selected row of data is highlighted in blue. This row represents data that corresponds approximately with the location of the cursor.

- Columns of log ratio data appear below the names of the arrays to which they correspond. If an array is selected for display in Chromosome and Gene Views, a colored square appears next to its name.
- Signal intensity (raw signals) and/or annotations appear if selected from the View command ribbon.
- SNPs display two probes, separated by a comma. This corresponds to a probe for each strand. Corresponding feature numbers are displayed, also separated by a comma.

Tab View tabs and buttons

You can see the following tabs and buttons in Tab View. See Figure 26 for a diagram that identifies some of these elements.

Design tabs A separate tab appears for each microarray design included. The name of the design appears on each tab, along with an icon:

CGH – An aCGH array design.

CGH+ – A CGH+SNP array design.

When you click a design tab, the data and annotation for the arrays in the design appear in Tab View.

Arrays tab (Available when you click a specific design tab.) Contains a table of data and annotation for all arrays in a design that contain biological data.

SNPs tab (Available when CGH+SNP array is selected.) Contains a table of SNP data for the selected chromosome. Includes genotype information on a per-array basis.

Selected Arrays tab Contains a table of data and annotation for the selected arrays from all designs in the window.

(**Resize** buttons) The button that points away from Tab View expands the view. The other button restores the view to its original size.

(**Detach** button) Removes Tab View from the main window, and opens it in a separate window.

Tab View actions and shortcut menus

- Right-click a *heading of a column other than an array data column* to display a Scroll To Column option. If you click this option, the Scroll to Column dialog box appears, where you can select a column in the current tab. The program then scrolls the data table in the tab so you can see the column.
- Click an *entry in a data table* to select the row in which it appears. This also moves the cursor to the location of the data point that corresponds to the selected row.
- Right-click a *data table entry* to display these options:

Option	Description
Find in column	Opens the Find in column dialog box, where you can search for a specific text string within the column you clicked. "Find in column" on page 190.
Google LocusLink PubMed UCSC HG15(April '03) UCSC HG16(July'03) UCSC HG17(May'04) UCSC HG18(March'06) UCSC mm8(Feb'06) UCSC mm9(July'07) DGV(hg18) GO KEGG(HUMAN)	Opens your Web browser, and passes the column entry you clicked as a search string to the selected site. The UCSC links search the indicated University of California, Santa Cruz database related to the indicated genome build.
Customize Link	Opens the Customize Search link dialog box, where you can create or edit a custom Web link that appears in this shortcut menu. When you click a custom link, the program opens your Web browser, and passes the column entry you clicked as a search string to the site. See "Customize Search Link" on page 170.

Status Bar

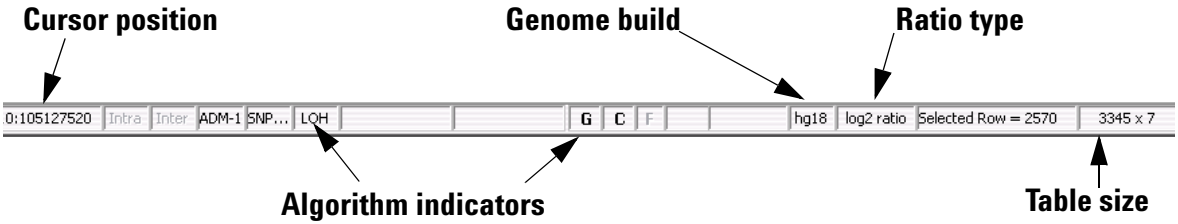


Figure 27 Status bar

The Status Bar displays information related to the displayed data.

Cursor position The chromosomal location of the cursor. See "The View Cursor" on page 49.

Algorithm indicators Calculations that are currently selected in the Preprocessing and Analysis tabs are shown in **bold**.

Genome build The genome build associated with the displayed data.

Ratio type The mathematical type of the array data. The possible types are:

- **ratio**
- **$\log_2$ ratio**
- **$\log_{10}$ ratio**
- **ln (natural log) ratio**

Selected Row The row in the displayed data table that is selected. The location of the cursor is approximately the chromosomal location associated with this row.

Table size The number of rows and columns in the displayed tab. The size appears as
<# of rows> x <# of columns>.

View Aberrations Window

The View Aberrations window is used to examine the results from one or more samples simultaneously. The results are displayed graphically using the Genomic Viewer. For information on how the Genomic Viewer displays results, see "Genomic Viewer" on page 40.

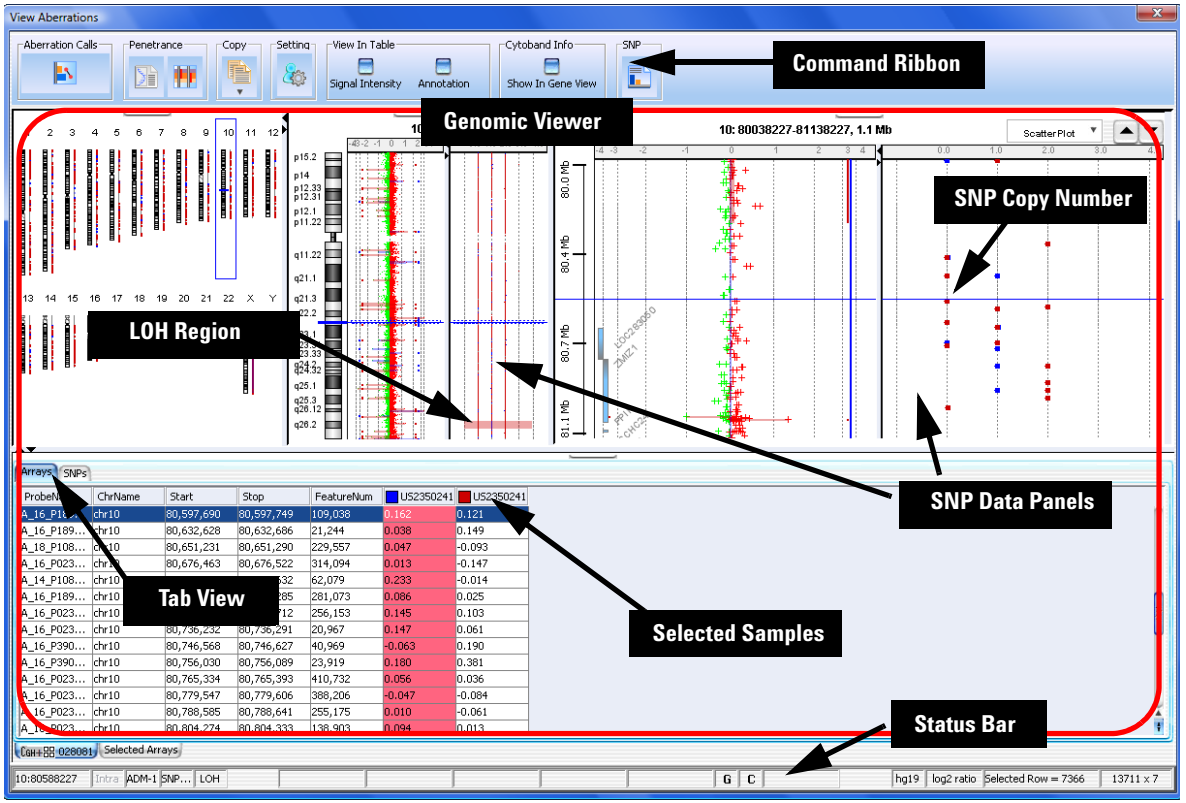


Figure 28 View Aberrations window showing CGH+SNP results

Purpose: Used to display results from one or more analyzed samples, display differrential aberrations, and create graphical Probe and Interval summaries.

To open: In the Analysis tab, under View, click **Multi Sample View**. Search for and select one or more analyzed samples, and then click **View Aberrations**. For information on how to set up annotation searches or queries, see the *Running CytoGenomics Analyses User Guide*.

Aberration View is used only for display of results. To display and sign off on results (for users with Scientist role or higher), you must use the Review function to open the Triage View. See "Triage View" on page 59.

The sections below describe the capabilities you have within the View Aberrations window.

View Aberrations command ribbon

Figure 29 Command ribbon for View Aberrations

The commands in this ribbon are described in the following table.

Table 9 Commands for View Aberrations command ribbon

Command	Purpose
Aberration Calls	
Differential	Set up and display a differential aberration summary for two or more samples.
Penetrance	
Probe	Creates a graphical penetrance summary that shows penetrance scores for each probe position across selected chromosomes.
Interval	Creates a graphical interval penetrance summary that lets you display and save interval penetrance results for the selected arrays in the current view.

Command	Purpose
Copy	Opens a menu where you can select one of the following items in the View Aberrations window, and copy them to the Clipboard. • All • Tab view • Genome view • Chromosome view • Gene view
Setting	
View Settings	Lets you select what content to display in the View Aberrations window, and how it appears.
View In Table	
Signal Intensity	Select to add green and (if applicable) red signal intensity columns to the tab view for each sample.
Annotation	Select to add annotation columns to the tab view.
Cytoband Info	
Show In Gene View	Select to show cytobands in the Gene View.
SNP	
Generate Genotype Reference	Opens the Generating Genotype Reference Files dialog box, where you set up and create a genotype reference file from selected CGH+SNP reference microarray(s). This genotype reference can be imported later into the database, and used to analyze CGH+SNP data. See "Generating Genotype Reference File" on page 191.

View Aberrations window actions and shortcut menus

When you right-click the mouse in the various views with the View Aberrations window, menus appear that let you perform specific tasks. The table below describes the tasks that are available from the right-click menus within the View Aberrations window. For more information about the different data views within the View Aberrations window, see "View Aberrations Window" on page 54.

Gene View shortcut menu and other actions

For a description of the shortcut menu items and other actions within the Gene View, see "Gene View shortcut menu and other actions" on page 47. These items are the same for both the View Aberrations window and the Triage View window.

Tab View actions and shortcut menus

- Right-click the *name of an array in a column heading* to display these options:

Option	Description
Select Array	(Available if the array is not selected.) Selects the array for display. A colored square appears next to the name of the array.
Deselect Array	(Available if the array is selected.) Removes the array data from scatter plots, and removes the column of the array from the Selected Arrays tab.
Edit Array Color	Opens the Select Color dialog box, where you can change the display color of the array. See "Select Color" on page 237.
Edit Array Order	Opens the Edit Array Order dialog box, where you can change the order in which the names of the arrays in a given design of appear in Tab View and in the Navigator. In Gene View, when you view separate scatter plots for each array, the plots also appear in this order. See "Edit Array Order" on page 181.
Select All Arrays	Selects all arrays in all designs for display. All arrays appear in the Selected Arrays tab.

Option	Description
Deselect All Arrays	Removes all arrays from display, and from the Selected Arrays tab.
Scroll To Column	Opens the Scroll to Column dialog box, where you can select a column in the current tab. The program then scrolls the data table in the tab so you can see the selected column. "Scroll to Column" on page 236.

- Right-click a *heading of a column other than an array data column* to display a Scroll To Column option. If you click this option, the Scroll to Column dialog box appears, where you can select a column in the current tab. The program then scrolls the data table in the tab so you can see the column.
- Click an *entry in a data table* to select the row in which it appears. This also moves the cursor to the location of the data point that corresponds to the selected row.
- Right-click a *data table entry* to display a shortcut menu for the table. For information on the shortcut menu items and tasks, see "Tab View actions and shortcut menus" on page 52.

Triage View

In Agilent CytoGenomics 1.0, Triage View is where a user examines analysis results, changes or adds aberration calls, adds notes, generates reports, and then (for users with Scientist or Administrator role) signs off the results.

The diagram below shows the Triage View for a CGH+SNP sample. Results are depicted graphically in the *Genomic Viewer*. For more information on the Genomic Viewer, see "Genomic Viewer" on page 40.

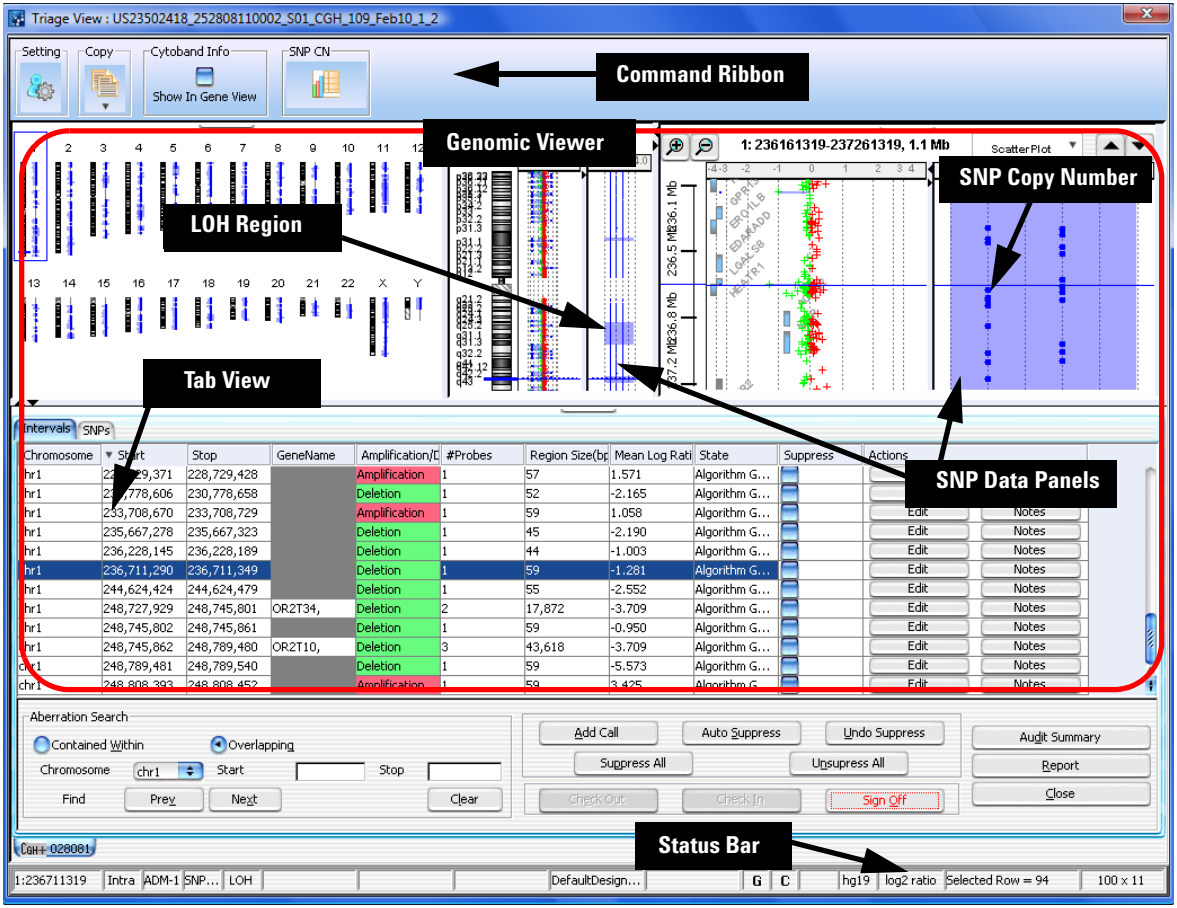


Figure 30 Triage View for CGH+SNP sample (Scientist or Administrator user role)

Purpose: Used to review and (for users with role of Scientist or Administrator) sign off results for completed samples.

To open: In the Analysis tab, under Review, click **Review**. In the list of analyzed samples, click the Status for a sample. In the Analyzed Sample dialog box, select the sample results you want to display, and click **OK**.

OR

In the Analysis tab, under Analyze, click **Job Monitor**. Locate the completed workflow job you wish to review, and click **Review**. In the sample list for the workflow job, click the Status of the sample results you want to review.

OR

In the Content tab, in Sample Manager, click the Status of any sample with a Status of Analyzed, Check In, or Reviewed.

Samples can have the following Status:

- Analyzed – Sample was successfully analyzed, but was not checked out or in, or reviewed.
- Check Out – Sample is checked out. Only the user who checked out this sample can currently open it in Triage View.
- Check In – Sample is checked in. Changes were made to this sample and were checked in. This sample is available to open in Triage View.
- Reviewed – Sample was signed off and the results are now locked. You can open the sample in Triage View, but you cannot make any changes.
- Samples that have more than one analysis result are indicated with a "#".

Changes to the call State

In the Intervals or SNPs table, the State of each call indicates how it was determined.

- Algorithm Generated – Calls that are based on the results of the algorithms chosen in the analysis method used to analyze the sample.
- Added – Calls that were added by the user.
- Edited – Calls that were changed by the user.

The following actions and commands appear in the Triage View window. Some actions and commands are not available until you *Check Out* the sample.

Command Ribbon

View Settings Opens the View Preferences dialog box, where you can change how you want results to appear in the Genomic Viewer. See "View Preferences" on page 251.

Copy Opens a menu where you select part or all of the view to copy to the Clipboard. You can then paste the contents of the Clipboard to a document of your choice.

Cytoband Info Select the check box to show cytobands in the Gene View of the display.

SNP CN QC Metrics Opens the SNP CN QC Metrics Table dialog box, that displays QC metrics for the selected sample. See "SNP CN QC Metrics Table" on page 245. The sample must be analyzed using one or more SNP algorithms in order to show this table.

Aberration Search

Contained Within Select to search for aberrations that fall within the selected chromosome and location range.

Overlapping Select to search for aberrations that overlap with the selected chromosome and location range.

Chromosome Select the chromosome you want to search.

Start Type the start location of the range to search.

Stop Type the end of the range to search.

Prev Goes to the previous match for the search.

Next Goes to the next match for the search.

Clear Clears the search parameters.

Sample review commands

Check Out Checks out the sample so that you can make changes.

Check In Saves your changes and checks in the sample so that other users can review and make changes to the sample.

Sign Off Saves the changes and signs off the sample. Changes cannot be made by any user after a sample is signed off. (Not available for users with the role of Technician.)

Changes to aberration calls

Add Call Opens the Add Aberration Call dialog box, where you can create a new aberration call for the table. See "Add Aberration Call" on page 133.

Auto Suppress Opens the Auto Suppress dialog box, where you select a track to use for aberration suppression. Calls within the selected track are suppressed. See "Auto Suppress Dialog" on page 145 and "Create Track" on page 168.

Undo Suppress Removes suppress from all intervals.

Suppress All Suppresses all aberrations for the currently selected chromosome.

Unsuppress All Removes suppress from all intervals.

Interval and SNPs table actions

Suppress Select to suppress an individual aberration call in the table.

Edit Opens the Edit Aberration dialog box, where you can change the selected aberration call. See "Edit Aberration" on page 180.

Notes Opens the Notes dialog box, where you can type information about the selected aberration call. See "Notes" on page 225.

General window actions

Audit Summary Shows a complete list of who checked the sample in and out, and what changes were made.

Report Generates a Cyto Report and saves it in a location you specify.

NOTE This option is not available for samples that did not have a cyto report selected in the workflow.

Close Closes the Triage View. Changes that were not checked in are not saved.

Config Tab Windows

This section contains descriptions of the windows that appear when you select the commands in the Config tab. These commands are tasks that are performed by users with the role of Scientist.

Analysis Method window

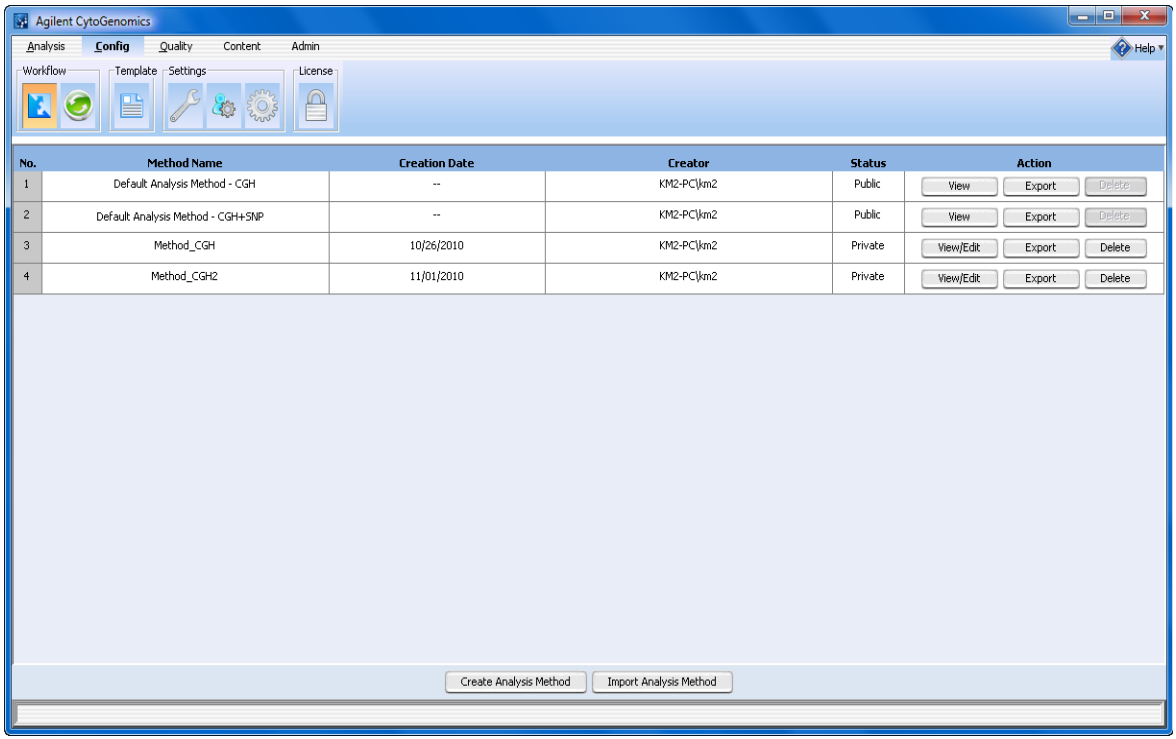


Figure 31 Analysis Method window

Purpose: Create and manage analysis methods.

To open: In the Config tab, under Workflow, click **Analysis Method**.

View/Edit Opens the Analysis method setup window, where you can change the selected analysis method. For published (public) analysis methods or methods created by other users (private), you cannot save changes.

Export Opens the Export dialog box, where you type a name and save the analysis method to a network or local location.

Delete For analysis methods you created that are not published, this deletes the selected analysis method.

Create Analysis Method Opens the Analysis Method setup window, where you set up the parameters for a new analysis method.

Import Analysis Method Opens the Import Analysis Method dialog box, where you can search for and select an analysis method file that was previously exported, to import to the program.

Workflow window

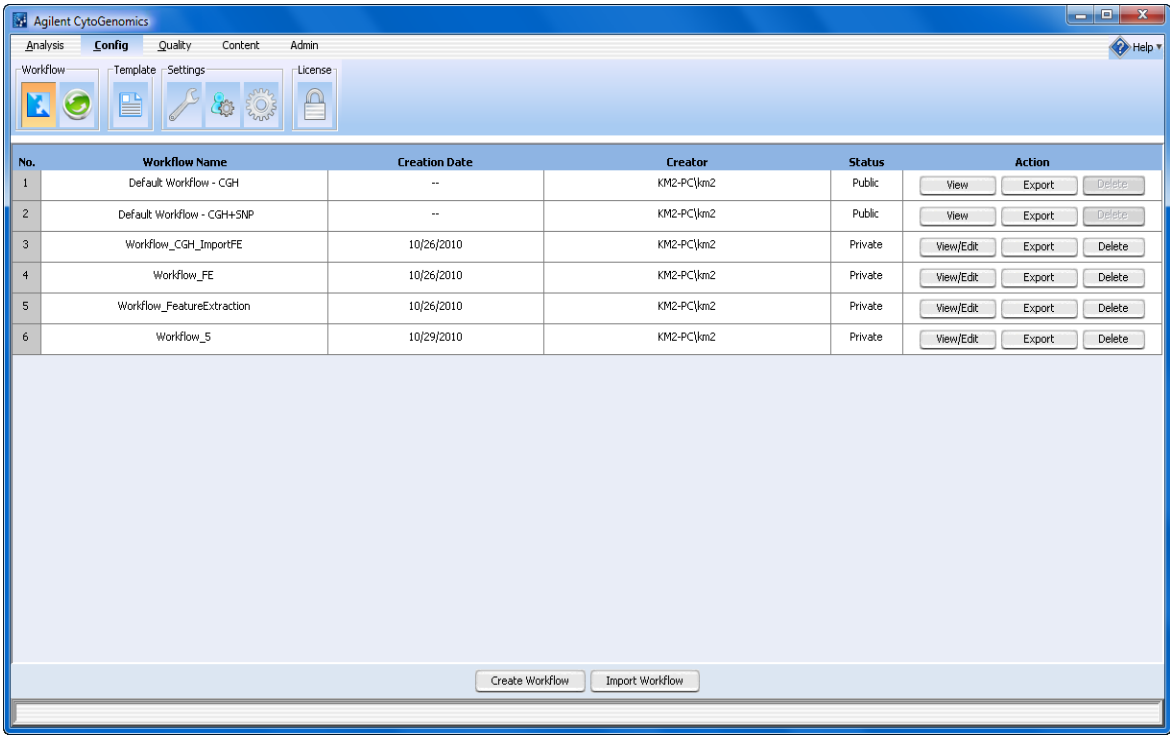


Figure 32 Workflow window

Purpose: Create and manage workflows.

To open: In the Config tab, under Workflow, click **Workflow**.

View/Edit Opens the Workflow setup window, where you can change the selected workflow. For published (public) workflows or for workflows created by other users, you cannot save changes.

Export Opens the Export dialog box, where you type a name and save the workflow to a network or local location.

Delete For workflows you created that are not published, this deletes the selected workflow.

Create Workflow Opens the Workflow setup window, where you set up the parameters for a new workflow.

Import Workflow Opens the Import Workflow dialog box, where you can search for and select a workflow file that was previously exported, to import to the program.

Report Template window

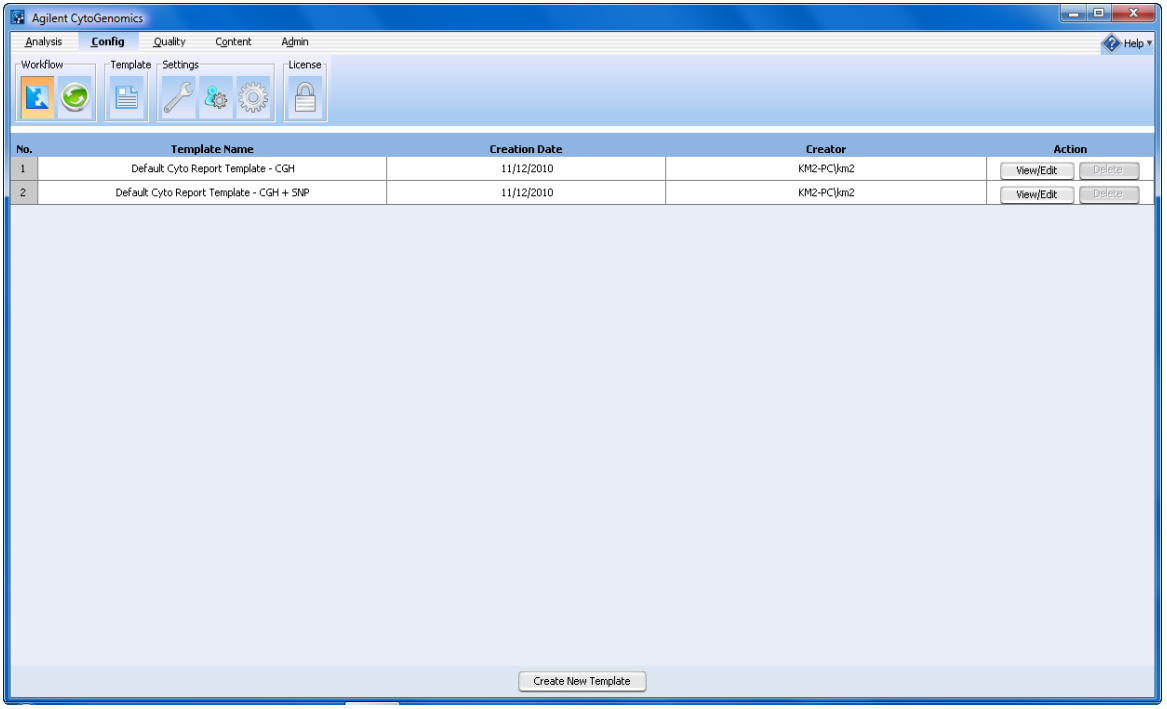


Figure 33 Report Template window

Purpose: Create, display, and change cyto report templates.

To open: In the Config tab, under Template, click **Report Template**.

View/Edit Opens the Edit Cyto Report Template dialog box where you can display or change the selected report template. You cannot change and save report templates that were created by another user or the default report templates. See "Edit Cyto Report Template" on page 182.

Delete Deletes the selected report template. You cannot delete default report templates, or templates that were created by another user.

Create New Template Opens the Create Cyto Report Template dialog box, where you select the items you want to include in a Cyto Report. See "Create Cyto Report Template" on page 158.

Preferences window

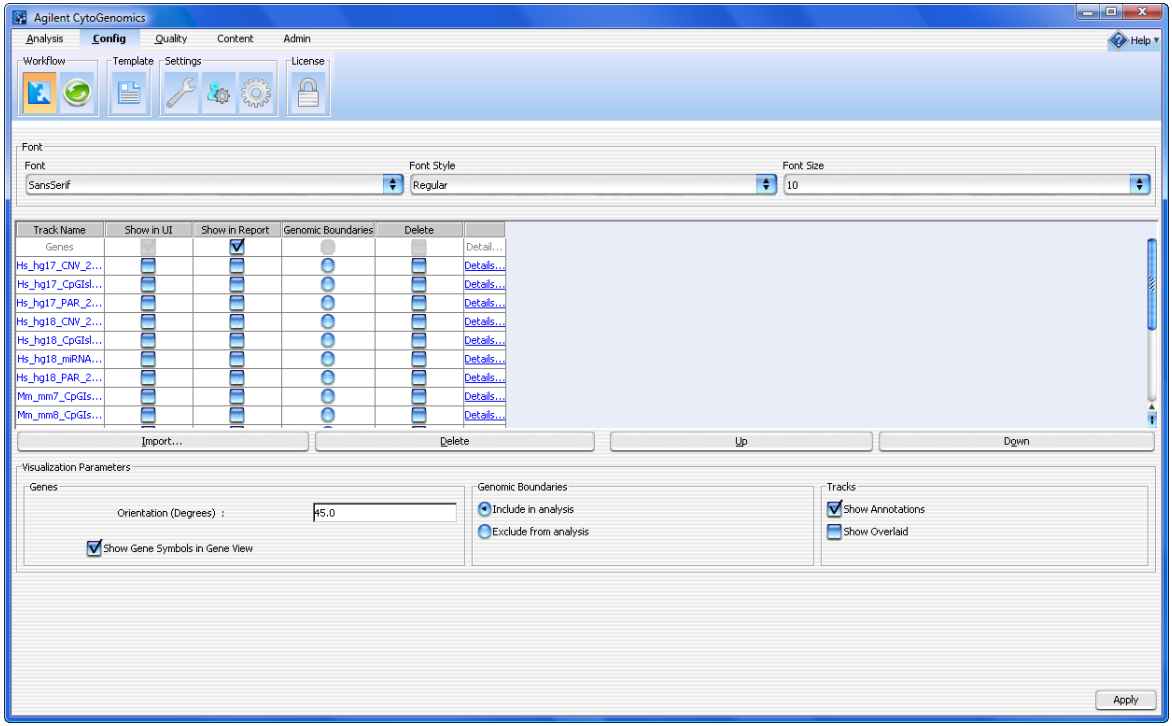


Figure 34 Preferences window

Purpose: To import and set up the appearance of tracks next to the Gene View. Tracks are additional graphic displays of genomic information loaded from an external file. They align with genomic coordinates in Gene View.

To open: In the Config tab, under Settings, click **Preferences**.

Apply Accepts the current settings and applies them to the Genomic Viewer.

Font Options

Select the font type, style and size for the gene annotations that appear in the selected tracks.

Tracks List

Track Name Name of the track already loaded or imported

Show in UI Select the check box to display the track next to Gene View.

Show in Report Select the check box to display the track information in all the reports.

Genomic Boundaries Click to use the track to define only the regions that aberration detection algorithms will run. You can choose to do this for only one track.

Delete Select the check box to delete the track from the list. Then, click **Delete** to delete the track from the list.

Details Click to display all the chromosome locations defined in the track.

Import Click to import new tracks.

Delete Click to delete the tracks selected in the Delete column.

Up Click to move a track up the list.

Down Click to move a track down the list.

Visualization Parameters

Genes These options affect the appearance of the Track and Gene View.

- Orientation – Type a number to set the angle at which the Gene Symbols appear in Gene View and the Track Annotations appear in the tracks.
- Show Gene Symbols – Select to show gene symbols in Gene View, and clear the check box to hide them.

Genomic Boundaries These options let you include or exclude the Genomic Boundaries from the analysis.

Tracks These options affect the appearance of the Track Views.

- Show Annotations – Select to show the names of the gene regions for the tracks, and clear to hide them.
- Show Overlaid – Select to overlay all the tracks that appear next to Gene View, and clear the check box to display the information in separate tracks.

View Settings window

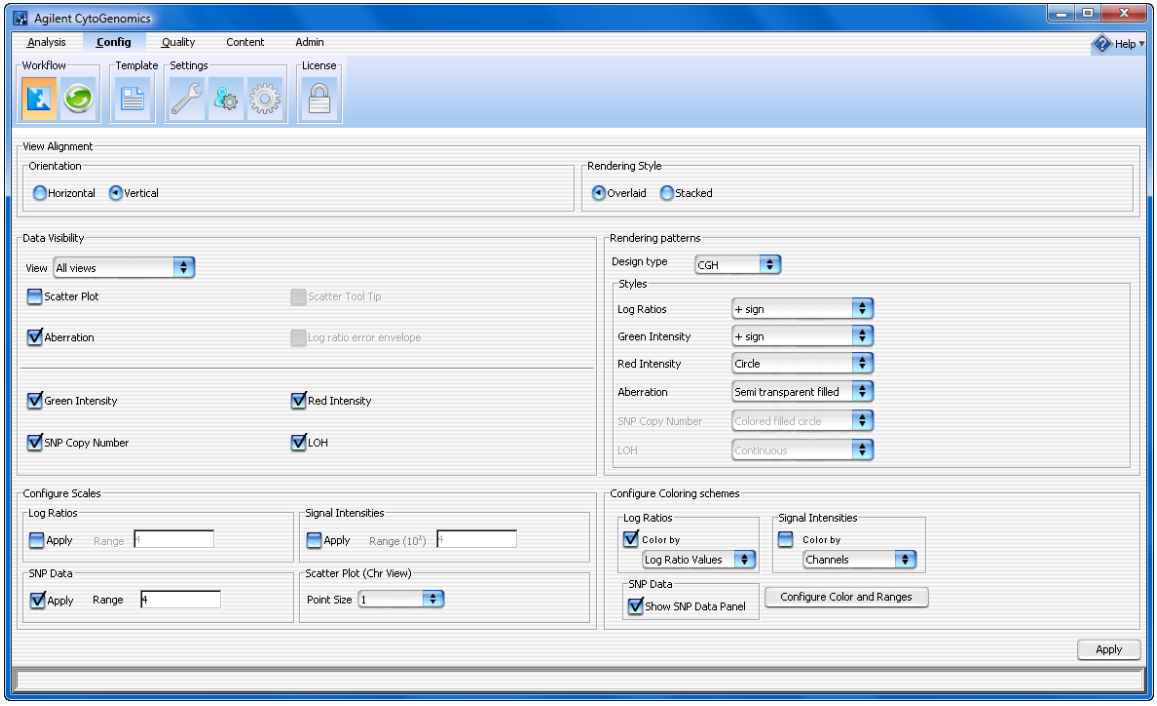


Figure 35 View Settings window

Purpose: This dialog box allows you to configure how data and results appear in Genome, Chromosome, and Gene Views of the Genomic Viewer. See "Genomic Viewer" on page 40.

To open: In the Config tab, under Settings, click **View Settings**. Or, right-click in any of the views in the Genomic Viewer and select **View Preferences**.

View Alignment Selects the orientation and rendering style (described below).

Option	Description
Orientation	
Horizontal	Stacks Genome, Chromosome, and Gene Views horizontally in the main program window. Genomic locations appear across the bottom of each view.
Vertical	Displays Genome, Chromosome, and Gene Views from left to right as side-by-side panes in the main program window.
Rendering Style	
Overlaid	In Chromosome View and in Gene View, displays data and results as a single, combined pane for all arrays. (Default)
Stacked	In Chromosome View and in Gene View, displays a separate pane for each array.

Data Visibility For each view, or all views, selects the kind(s) of data and results to display.

In **View,** select the view you want to configure. To set availability of display items for all views, select **All views.** Some display items are only available for certain views and modules. When you select a display item, it enables the item for display; for some items, you must take additional steps to display them. For example, you may need to configure a specific algorithm in the toolbar.

Select any of the following options, as available:

Option	Description/Comments
Scatter Plot	The plot(s) of individual log ratio, intensity, or probe score data points.
Scatter Tool Tip	The ToolTips that appear when you place the pointer over specific data points on the scatter plot(s) in Gene View. The tool tip shows the array of origin and the numerical log ratio value for the data point.
Aberration	The result of the selected aberration detection algorithm.
Log ratio error envelope	The log ratio error envelope is a visual representation of the log ratio error calculated by Feature Extraction.
Green Intensity	Select the check box to display green raw signal intensity.

Option	Description/Comments
Red Intensity	Select the check box to display red raw signal intensity.
SNP Copy Number	Select the check box to display the SNP copy number data for CGH+SNP samples. Note: Data is only displayed for samples where SNP copy number results exist.
LOH	Select the check box to display the LOH (Loss/Lack of Heterogyzosity) data for CGH+SNP samples. Note: Data is only displayed for samples where LOH results exist.

Rendering Patterns

These options control the specific appearance of data and results in Genome, Chromosome, and Gene Views. You configure these options separately for each type of application design.

- **Design Type** – Select the application design type for which you want to define rendering patterns.
- **Styles** – Select the display style for each of these elements:

Option	Description/Comments
Log Ratios	Select the symbol used for log ratio data points in the scatter plots in Chromosome and Gene Views.
Green Intensity	Select the symbol to use for display of the green raw signal intensity.
Red Intensity	Select the symbol to use for display of the red raw signal intensity.
Aberration	Select the rendering style for detected aberrations. • **Semi transparent filled** – Solid, colored regions (in the display colors defined for each array, if applicable). • **Hatched** – Cross-hatched colored lines (in the display colors defined for each array, if applicable). • **Do not show area** – Aberrations do not appear.
SNP Copy Number	Select the symbol to use for showing SNP Copy Number.
LOH	The only selection for showing regions of LOH is “continuous”.

NOTE Rendering scatter plots for more than 10 high density arrays in the Chromosome View may take significant time. Selecting filled circles as the rendering style for CGH scatter plots can also decrease performance. For faster performance, change the rendering style for CGH data from the filled circle to the plus (+) or cross hair sign.

Configure Scales For Log Ratios, Signal Intensities, SNP Data plots, select **Apply** to enable the custom scale. In Range, type the value to use as the range for the scatter plot.

Scatter Plot (Chr View) Point Size Select a point size to use for display of scatter plot data points in the Chromosome View.

Configure Coloring schemes Use these options to change the display of the scatter plot in the Gene View. These options are the same as those displayed in the Scatter Plot box in the Gene View.

To do this	Follow these steps
Show or hide the log ratio values in the Log Ratios plot	• To show the data points - Select the **Log Ratios** check box and select **Log Ratio Values** from the list. • To hide all data points - Clear the **Log Ratios** check box.
Show or hide LogRatios color-coded by Probe Score Values in the Log Ratios plot	• To show the data points - Select the **Log Ratios** check box and select **Probe Score Values** from the list. • To hide the data points - Clear the **Log Ratios** check box.
Show or hide Intensity values in the Signal Intensities plot	• To show the data points - Select the **Signal Intensities** check box and select **Intensity Values** from the list. • To hide all data points - Clear the **Signal Intensities** check box.
Show or hide Signal Intensities color-coded by Channels in the Signal Intensities plot	• To show the data points - Select the **Signal Intensities** check box and select **Channels** from the list. • To hide the data points- Clear the **Signal Intensities** check box.
Show or hide Signal Intensities color-coded by Probe Score values in the Signal Intensities plot	• To show the data points - Select the **Signal Intensities** check box and select **Probe Score Values** from the list. • To hide the data points- Clear the **Signal Intensities** check box.

To do this	Follow these steps
Show or hide SNP data panel	• To show the SNP data panel - Select the **Show SNP Data Panel** check box. • To hide the SNP data panel - Clear the **Show SNP Data Panel** check box.
Change the ranges and colors for scatter plot and signal intensities panels	• Click **Configure Color and Ranges** to enter ranges and change colors. See "Configure Coloring Ranges and Shades" on page 152 for more information.

Apply Applies changes without closing the dialog box.

Default Settings window

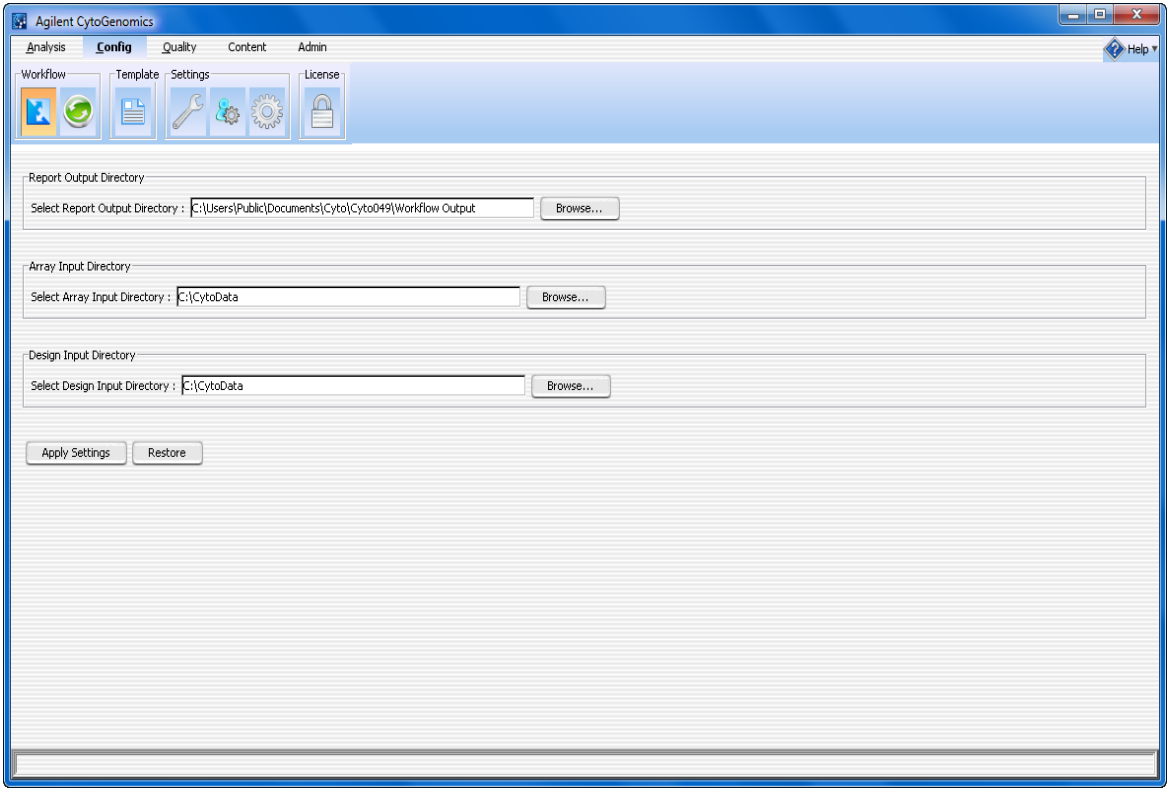


Figure 36 Default Settings window

Purpose: To set the default folders to use for data.

To open: In the Config tab, under Settings, click **Default Settings**.

Select Report Output Directory Type the name of the path to use for storing reports, or click **Browse** to browse to and select the folder.

Select Array Input Directory Type the name of the path to use for importing arrays, or click **Browse** to browse to and select the folder.

Select Design Input Directory Type the name of the path to use for design files, or click **Browse** to browse to and select the folder.

Apply Settings Accepts the displayed folders for default locations.

Restore Restores previous folder locations after you made changes but did not click Apply Settings.

License window

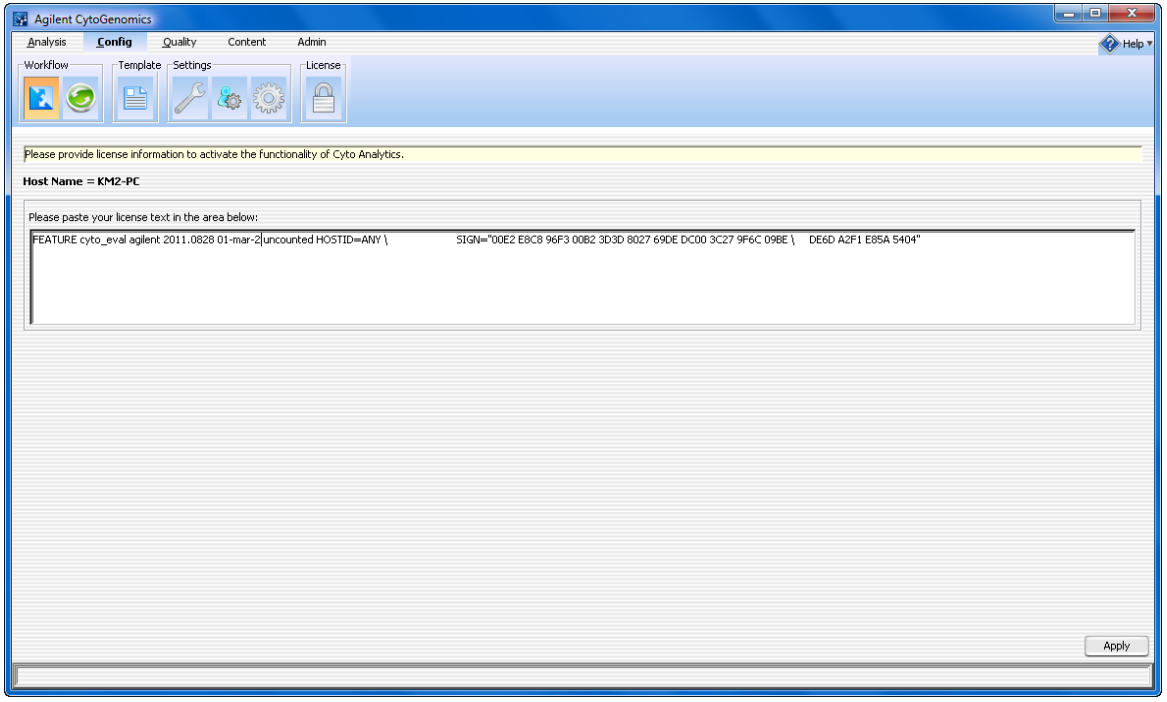


Figure 37 License window

Purpose: Show or enter a license for Agilent CytoGenomics 1.0.

To open: In the Config tab, click **License**.

Please paste your license in the area below Copy and paste your license.

Apply Click to apply the license to your program.

Analysis Method Window

The Analysis Method window appears when you create an analysis method or click **View** or **View/Edit** in the analysis method list. This window lets you set the parameters for an analysis method, save the analysis method, and publish it so that it is available for all users.

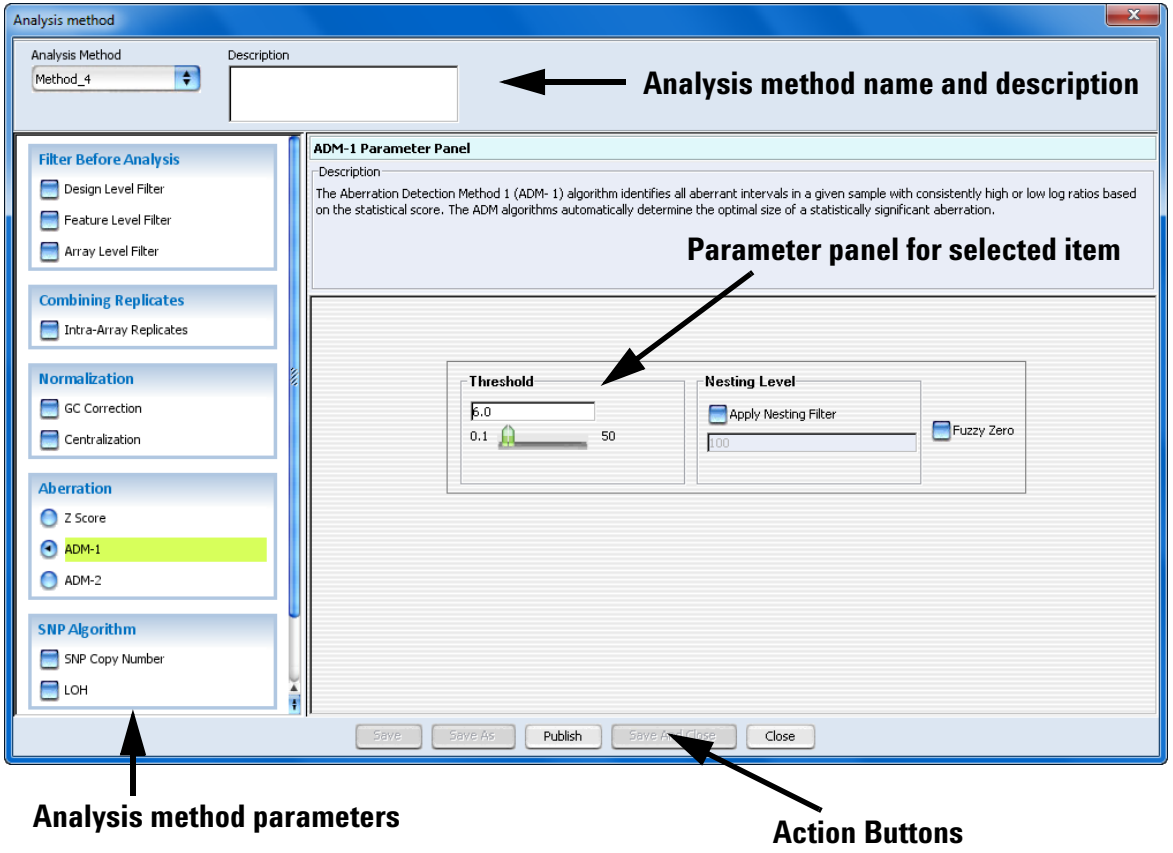


Figure 38 Analysis method setup

Purpose: Used to set parameters for an analysis method.

To open: In the Config tab, under Workflow, click **Analysis Method**. Then click **Create Analysis Method** to create a new analysis method, or click **View/Edit** or **View** for an existing analysis method.

Filter Before Analysis

The options available in the Filter Before Analysis area let you select filters that remove or include data based on specific criteria. You can also create and edit these filters.

Select any of these options:

Option	Description
Design Level Filter	Opens the Design Level Filter Parameter Panel. Design level filters let you include or exclude probes, based on criteria set in the filter. For example, a design filter can be used to filter out probes that fail the homology filter or have a low probe score. See "Design Level Filter Parameter Panel" on page 120.
Feature Level Filter	Opens the Feature Level Filter Parameter Panel. See "Feature Level Filter Parameter Panel" on page 122. Feature level filters let you include or exclude data from specific microarray features, based on information from the imported Feature Extraction output files.
Array Level Filter	Opens the Array Level Filter Parameter Panel. See "Array Level Filter Parameter Panel" on page 113. Array level filters let you include or exclude arrays in the current workflow based on their attributes.

Combining Replicates

For CGH arrays, intra-array replicates are features within the same array that contain the same probe. When you combine replicates, you define how the program handles replicate probes. The program can combine multiple biological and technical replicates within arrays.

Select this option:

Option	Description
Intra-Array Replicates	Combines replicate probes within each array. If you select this option, the Intra-Array Replicates Parameter Panel appears. However, no input parameters are required.

Normalization

Select any of these options:

Option	Description
GC Correction	Corrects for artifacts by performing a regression fit to GC content in a specified region flanking the probes. GC Correction is recommended for SNP Copy Number and LOH analyses. See "GC Correction Parameter Panel" on page 124.
Centralization	Centralization recenters log ratio values. It finds a constant value to subtract from or add to all values, and ensures that the zero-point reflects the most-common-ploidy state. For a description of the centralization algorithm, see "Centralization Algorithm" on page 275. All of the aberration algorithms can use the Centralization calculation, used to normalize data. Centralization is recommended for SNP Copy Number and LOH analyses. See "Centralization Parameter Panel" on page 115.

Aberration

The options in the Aberration folder let you select the aberration detection algorithm for the workflow. For a detailed discussion of all aberration detection algorithms, see "Aberration detection algorithms" on page 267. Select one of these options:

Option	Description
Z Score	The Z-Score algorithm is a quick method of detecting aberrant regions. It assesses intervals using a sliding window of fixed size, and is especially useful in the exploratory phase of CGH data analysis. When you select this option, the Z Score Parameter Panel appears, where you can set the parameters of the algorithm. See "Z Score Parameter Panel" on page 131. For a discussion of the Z-Score algorithm, see "Z-Scoring for Aberrant Regions" on page 287.

Option	Description
ADM-1	The ADM-1 algorithm searches for intervals in which the average log ratio of the sample and reference channels exceeds a threshold that you specify. When you select this option, the ADM-1 Parameter Panel appears, where you can set the parameters of the algorithm. See "ADM-1 Parameter Panel" on page 111. For a discussion of the ADM-1 algorithm, see "ADM-1" on page 290.
ADM-2	The ADM-2 algorithm is similar to the ADM-1 algorithm, except that it is more accurate and also takes into account probe quality. It is especially useful for the detection of small aberrant intervals. When you select this option, the ADM-2 Parameter Panel appears, where you can set the parameters of the algorithm. See "ADM-2 Parameter Panel" on page 112. For a discussion of the ADM-2 algorithm, see "ADM-2" on page 293.

SNP Algorithm

These options let you set up the workflow to calculate results for CGH arrays that contain SNP probes. Select one or both of the following options:

Option	Description
SNP Copy Number	For each SNP site that is represented on the array, SNP Copy Number (ASCN algorithm) calculates an expectation value for the copy number of the uncut SNP allele. See "SNP Copy Number Parameter Panel" on page 129. For more information on the SNP Copy Number algorithm, see "SNP analysis algorithms" on page 270. **Note**: In order to select SNP Copy Number, you must first select an Aberration (except z-score). GC Correction and Centralization are recommended.
LOH	LOH algorithm identifies copy-neutral genomic regions with a statistically significant scarcity of heterozygous SNP calls. The algorithm reports the regions where the LOH score exceeds a definable threshold. See "LOH Parameter Panel" on page 125. For more information on the LOH algorithm, see Chapter 4, "LOH (Loss or lack of heterozygosity) algorithm". **Note:** In order to select LOH, you must first select SNP Copy Number.

Filter After Analysis

Aberration filters exclude certain detected aberrations from the output of the workflow, based on specific criteria. To apply an aberration filter to the results of the workflow, select Aberration Filter. The Aberration Filter Parameter Panel appears, where you can create and select a filter. See "Aberration Filter Parameter Panel" on page 109.

Workflow Window

The Workflow window appears when you create a workflow or click View or View/Edit in the workflow list. This window lets you set the parameters for a workflow, save the workflow, and publish it so that it is available for all users.

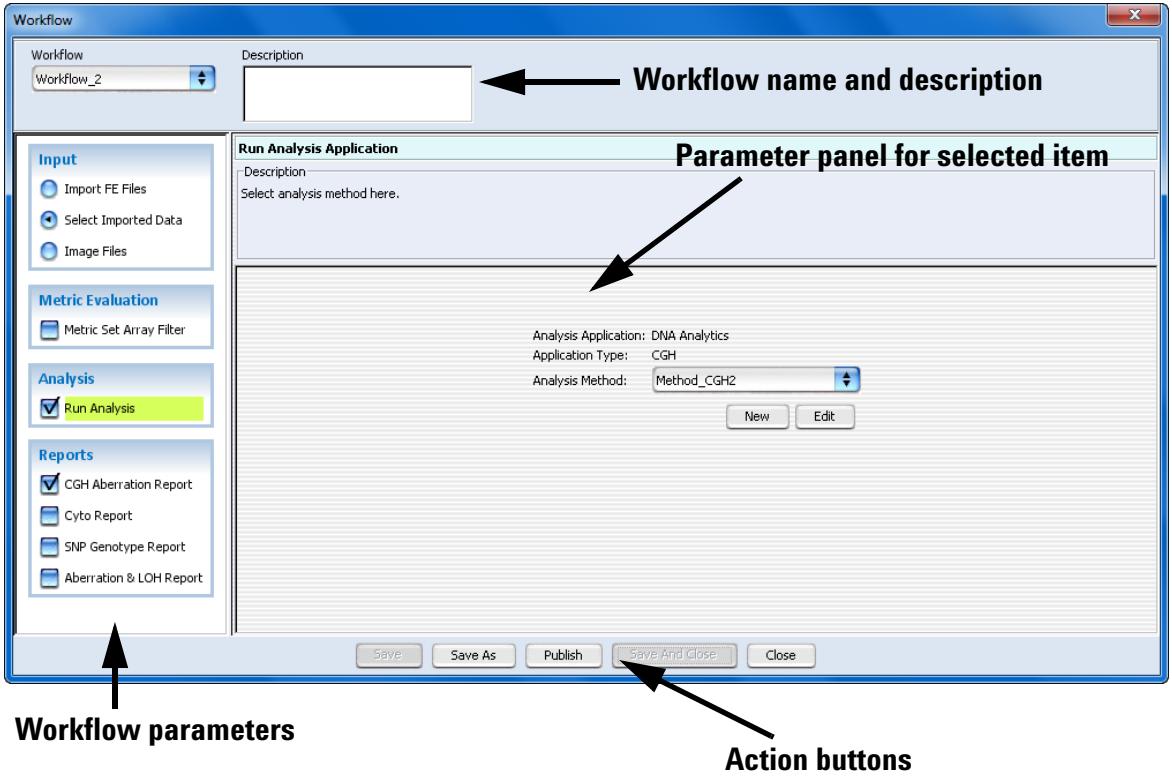


Figure 39 Workflow setup

Purpose: Used to create or change a workflow.

To open: In the Config tab, under Workflow, click **Workflow**. Then click **Create Workflow** to create a new workflow, or click **View/Edit** or **View** for an existing workflow to display or change it.

The parameters are described below. For more information on the individual parameter panels that appear when you select the parameters, see "Parameter Panels" on page 108.

Input

In Input, you select the source of data for the workflow. Select one of the input options:

Table 10 Input for workflow

Option	Description
Import FE Files	Configures the workflow to import Agilent Feature Extraction microarray data files. Users select the files when they run the workflow.
Select Imported Data	Configures the workflow to use microarray data that you previously imported into the program. Users select the microarrays when they run the workflow.
Image Files	This option lets you select the image files (.tif) to extract during the workflow. When image files are selected, the workflow will run the Feature Extraction program on these files first, then analyze the extraction results using the selected analysis method. Users select the image files when they run the workflow.

Metric Evaluation

The Metric Set Filter option lets you include or exclude data from the workflow, based on QC metric set filters.

Analysis

In this area, you select the analysis method to use for the workflow. You must select an analysis method for the workflow before you can save or publish it.

Reports

The options in the Reports folder let you select the reports that are created by the workflow. The reports contain one or more files that you can open in other programs such as Microsoft® Excel or Adobe® Reader®.

Select any of these options:

Table 11 Report selections for CGH

Report	Description
CGH Aberration Report	The CGH Aberration Report describes regions that have detected aberrations. You can report these regions by genomic interval, by probe, or both. The program reports aberrations separately for each array in the workflow, and creates one or more *.xls files that you can work with in Microsoft Excel. When you select this option, the CGH Aberration Report Parameter Panel appears, where you can configure the report, and type a name and select a location for it. See "CGH Aberration Report Parameter Panel" on page 117.
Cyto Report	Cyto reports summarize analysis settings and detected aberrations by array. The workflow creates a separate PDF and .XML file for each array. When you select this option, the Cyto Report Parameter Panel appears, where you select the desired cyto report template to use, and a location for the report. See "Cyto Report Parameter Panel" on page 119. Note – You create Cyto Report templates using the Report Templates option in the Config tab.
SNP Genotype Report	The SNP Genotype Report contains genotype and *p*-values for SNP probes. It generates reports in .xls format for the entire genome, or for each chromosome. When you select this box, the SNP Genotype Report Parameter Panel appears, where you select the format and a location for the report. See "SNP Genotype Report Parameter Panel" on page 130.
Aberration & LOH Report	The Aberration & LOH Report contains aberration and log ratio information for significant intervals. When you select this box, the Aberration & LOH Report Parameter Panel appears, where you select the location for the report. See "Aberration & LOH Report Parameter Panel" on page 108.

Quality Tab Window

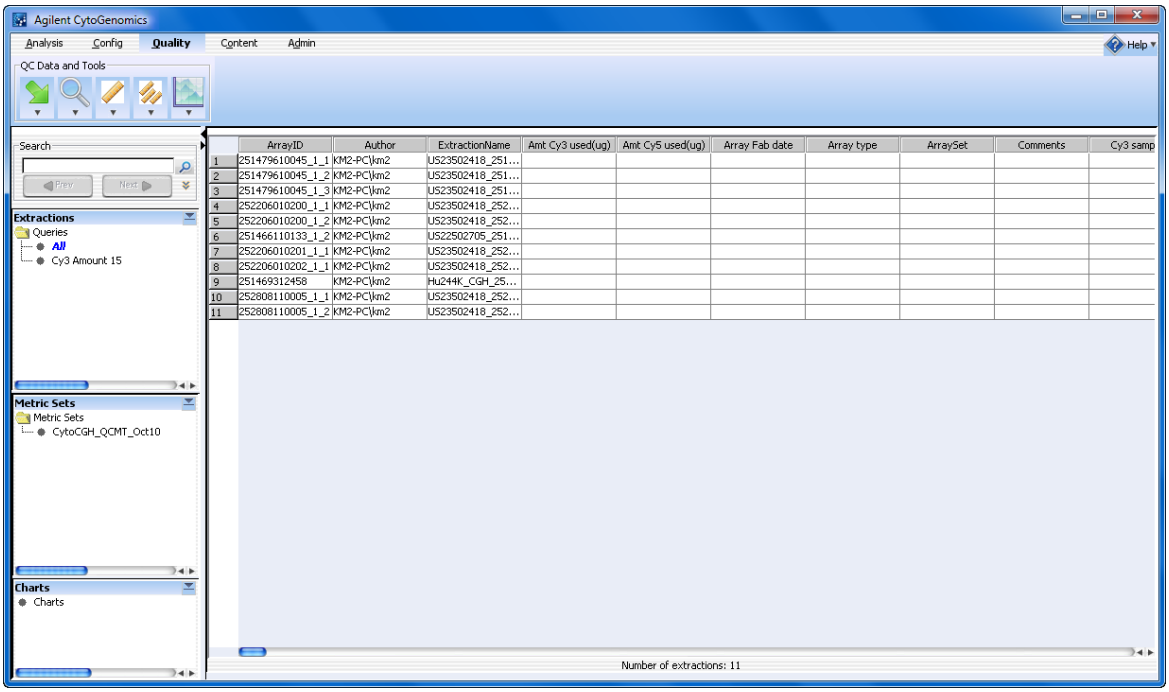


Figure 40 Quality tab

Purpose: To set up QC metrics and metric sets, and to create and display quality charts for extracted data in the program.

To open: On the tab bar, click **Quality**.

Navigator

The Navigator consists of the **Extractions** pane, the **Metric Sets** pane, and the **Charts** panes. These panes are described below.

Extractions Pane

This pane displays a list of query views of extractions in the database. The query named “All” is a permanent, unchangeable query and consists of no query parameters; it displays all extractions in the database.

Right-click to see the following options in the Extractions pane:

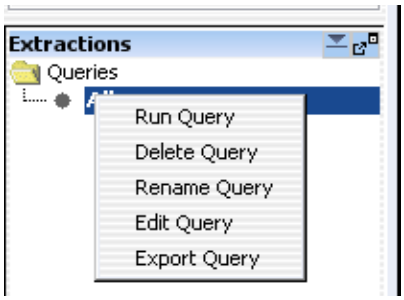


Figure 41 Extractions Navigator Options

Menu item	Description
Run Query	Runs the selected query on the database and displays the extractions from that query in the Tab View.
Delete Query	Permanently removes the selected query from Agilent CytoGenomics 1.0 Quality tools.
Rename Query	Opens the Enter New Name dialog box to change the name of the selected query.
Edit Query	Opens the Query Builder dialog box, identical to clicking Query in the Quality tab and then selecting Edit.
Export Query	Opens the Export Query dialog box, identical to clicking Query in the Quality tab and then selecting Export.

Metric Sets Pane

This pane displays a list of metric sets that were created in or imported into Quality tools.

Right-click to see the following options in the Metric Sets pane:

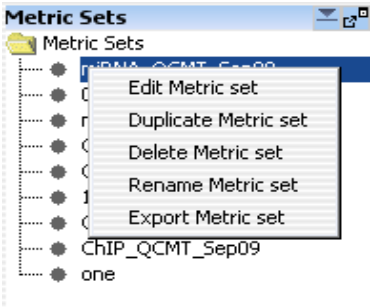


Figure 42 Metric Sets Navigator Options

Menu item	Description
Edit Metric set	Opens the Metric Set Configuration dialog box, identical to when you click **Metric Sets** in the Quality tab and then select **Edit**. See"Metric Set Configuration Dialog Box: Add Metrics to Metric Set Tab" on page 215.
Duplicate Metric set	Opens the Duplicate Metric Set dialog box, which lets you enter a name for the metric set to be copied.
Delete Metric set	Permanently removes the selected metric set from Agilent CytoGenomics 1.0 Quality tools.
Rename Metric set	Opens the Enter New Name dialog box to change the name of the selected metric set.
Export Metric set	Opens the Export Metric Set dialog box, identical to when you click **Metric Sets** in the Quality tab and then select **Export**.

Charts Pane

This pane displays a list of charts that were created in or imported into Quality tools.

Right-click to see the following options in the Metric Sets pane:

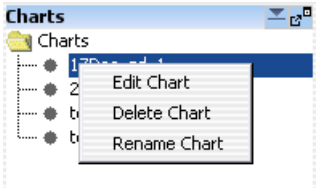


Figure 43 Metric Sets Navigator Options

Menu item	Description
Edit Chart	Opens the Chart Configuration dialog box, identical to clicking Chart in the Quality tab and then selecting Edit.
Delete Chart	Permanently removes the selected chart from Agilent CytoGenomics 1.0 Quality tools.
Rename Chart	Opens the Enter New Name dialog box to change the name of the selected chart.

Table View

The Table View displays the results of a selection from any pane in the Navigator (Query, Metric set, or Chart) in the Navigator. Examples of each pane are shown below.

Tip When you click on a row heading in the table, the results are sorted by the values in that column.

Table View - Extractions

The extractions loaded into the Agilent CytoGenomics 1.0 database are displayed in a table. See the figure below:

	ArrayID	Author	ExtractionName	Amt Cy3 used(ug)	Amt Cy5 used(ug)	Array Fab date	Array type	ArraySet	Comment
1	252136510460_1_2	Ann	O1-NA18517_25...						
2	252136510460_1_2		O1-NA18517_25...					1	
3	252136510317_1_2		O2-NA18517_25...					1	
4	252136511006_1_2		S1-NA18517_252...					1	
5	252136511015_1_2		S2-NA18517_252...					1	
6	252136510356_1_2	Jay	alsc_NA19000-TA...						
7	252136510357_1_1	Jay	alsc_NA19000-TA...						

Figure 44 Table View - Extractions

The default order of the column is:

- **ArrayID** (the barcode and pack number, if the image is a multi-pack array)
- **Author** (the user who is logged in to Agilent CytoGenomics 1.0 when the file is imported)
- **Extraction Name** (from FE parameters)
- **Array attributes**
- **FE parameters**
- **FE Stats**

To change the order of the columns in the table, drag the column headings to the desired positions.

Table View - Metric Sets

The Metric Sets created in or imported into the Quality tab are displayed in the table. See the figure below:

	Metric Name	Expression	Upper Limit	Upper Warning Limit	Lower Warning Limit	Lower Limit	Calculation Type	Defined by	Date Create
1	IsGoodGrid	IsGoodGrid	NA	NA	1.0	1.0	Manual		07-Jul-2009 18:
2	AddErrorEstimate...	AddErrorEstimate...	12.0	5.0	NA	NA	Manual		11-Aug-2009 18
3	AnyColorPrcntFe...	AnyColorPrcntFe...	15.0	8.0	NA	NA	Manual		11-Aug-2009 18
4	gNonCtrlMedPrcn...	gNonCtrlMedPrcn...	15.0	10.0	NA	0.0	Manual		11-Aug-2009 18
5	gTotalSignal75pctile	gTotalSignal75pctile	NA	NA	NA	NA	Manual		12-Aug-2009 14
6	LabelingSpike-InS...	0.5 * (gdmr285G...	NA	NA	NA	2.5	Manual		12-Aug-2009 14
7	HybSpike-InSignal	0.5 * (gdmr3Gen...	NA	NA	NA	2.5	Manual		12-Aug-2009 14
8	StringencySpike-I...	gdmr3ProbeRatio	NA	NA	NA	NA	Manual		12-Aug-2009 14

Figure 45 Table View - Metric Sets

Disassociate Metric A button on the far right that lets you quickly remove a metric from a custom Metric Set.

Charts View

To open the Chart View, in the Charts Navigator, double-click the name of a chart.

This pane has both a table and graphical view of the data. In the graphical view, you can zoom in on a chart by using the mouse to drag and release on the area to zoom. Double-click the chart to return to normal view.

The Charts View supports the operations described below.

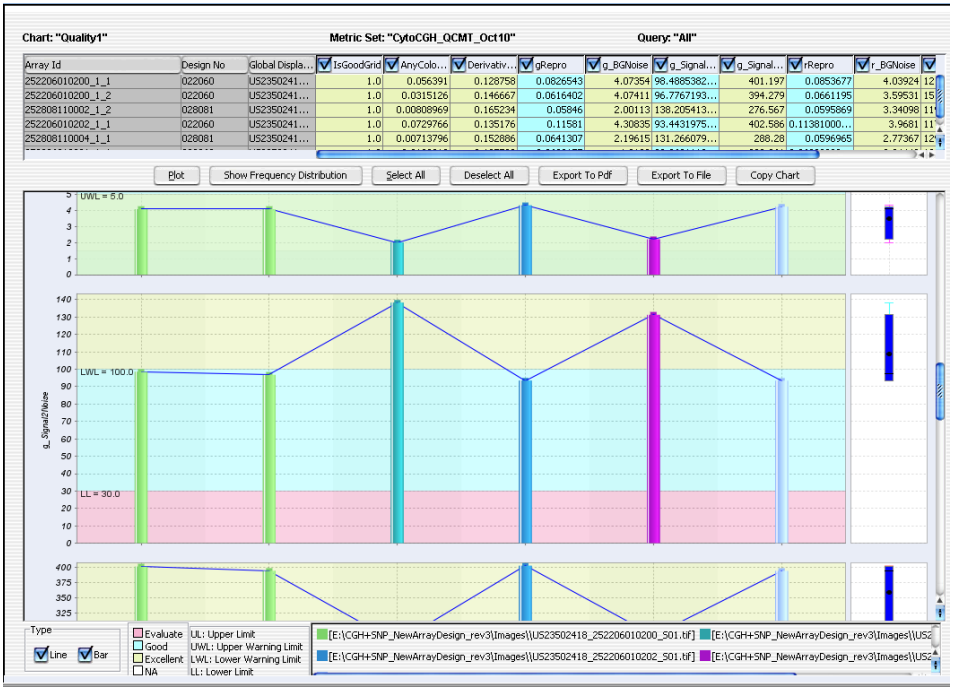


Figure 46 Charts View

Plot Draws the chart according to the data selected in the view.

Show Frequency Distribution Displays a binned vertical bar chart (a histogram) of each metric selected in the table view.

Select All/Deselect All Selects all or none of the available metrics to include in the chart.

Export to PDF Saves the chart in PDF format.

Export to File Exports the data from the chart to a tab-delimited text file.

Copy Chart Copies the chart to the Clipboard as a bitmap that can be pasted in MS Word and MS Paint, or in any other appropriate software.

Type Select Line, Bar, or both for the type of chart.

At the bottom of the Chart View, legends appear for coloring of thresholds, (Evaluate, Good, Excellent, NA) limits, and plotted data.

Content Tab Window

You use the Content tab to display and manage samples and their attributes. The Content tab has two different views – the Sample Manager View and the Feature Extraction View.

Content tab – Sample Manager window

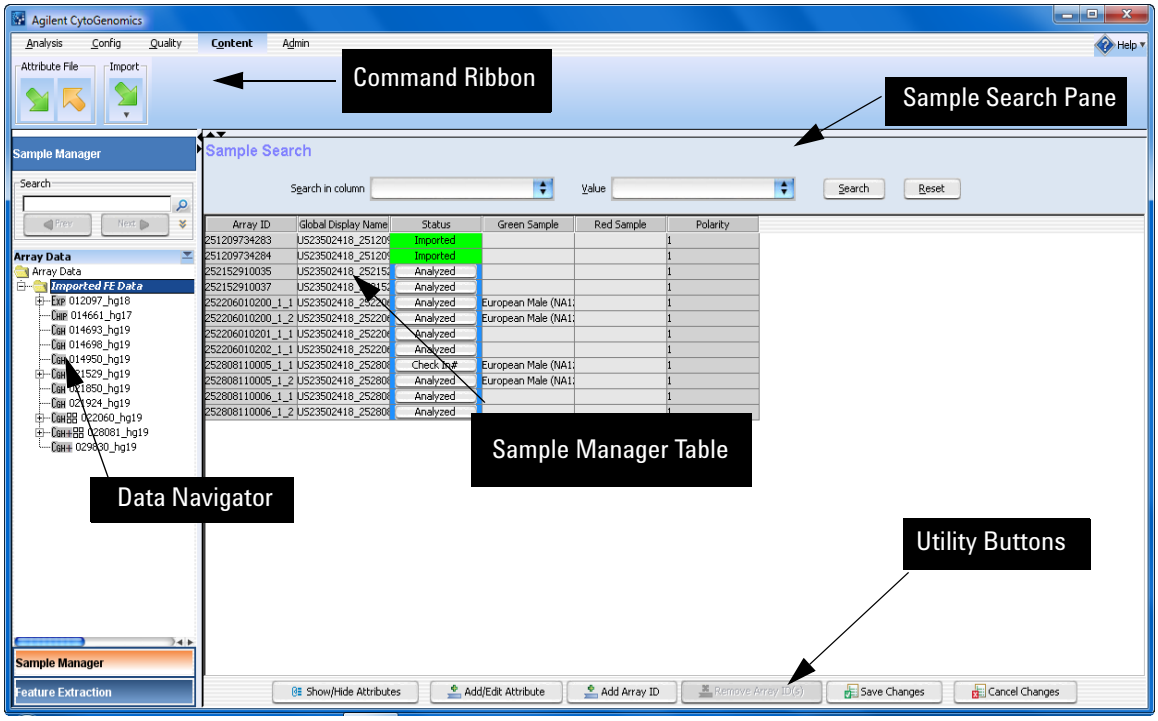


Figure 47 Content tab – Sample Manager View

Purpose: Used to manage samples and their attributes.

To open: In the Content tab, click **Sample Manager**.

Table 12 Content tab – Sample Manager View elements

Part	Purpose
Command Ribbon	Contains the commands used to perform Sample Manager functions, such as import and export Attribute Files.
Data Navigator	Displays list of microarrays, organized into folders that represent imported files or user-added attribute records.
Sample Search Pane	Lets you find an array ID or any part of a Global Display name in the list.
Sample Manager Table	Contains a list of sample microarrays, organized by Array ID. You can select which attributes to display, but at a minimum Array ID, Green Sample, Red Sample, Polarity, and Extraction Status are displayed. To display only the microarrays for an Array Data node, double-click the name of the data node in the Data Navigator.
Utility Buttons	Buttons that let you edit and add to the sample microarray list, and for saving or canceling changes.

The Sample Manager Command Ribbon contains the commands used for importing and exporting attribute files.

Table 13 Sample Manager Command Ribbon

Select	Purpose
Import Attribute File	To import contents of an Attribute File.
Export Attribute File	To export the selected Sample Manager data to an Attribute File.
Import Design	To import a design file. A design that matches the microarray must be present in the database before you can run a workflow for that microarray.
Import Genotype Reference	To import a genotype reference file to use for CGH+SNP analysis.

Sample Search Pane

Figure 48 Sample Manager – Sample Search pane

Search in column Select the column in the Sample Manager table to use to search for the data.

Value Based on your selection for Search Type, choose a value from the list to match for the search.

Search Searches the Sample Manager table and displays all samples that match the search condition.

Reset Clears the search condition.

Array Data Navigator

The Array Data Navigator for Sample Manager contains an Array Data folder, which contains one or more data folders. Each of these folders contains a list of microarrays in the program database. The microarrays are listed under the appropriate design for the array.

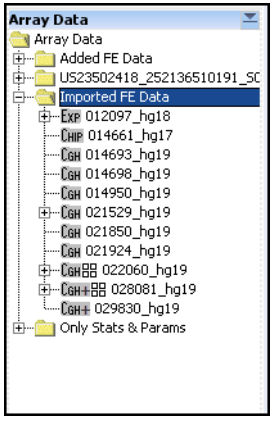


Figure 49 Array Data Navigator

You can use the Array Data Navigator to:

- display the microarrays for each imported attribute file
- delete a single unimported microarray
- delete a user-created folder of unimported microarrays
- search for Array IDs or Global Display Names

Table 14 Data Navigator icons, special text, and buttons

Icon	Comments
	Click to expand a folder and display its contents.
	Click to collapse a folder and hide its contents.
	A folder that contains data. In the case of Array Data, each folder, or "Data folder" contains a list of the microarrays in that imported or user-created attribute file.
	A microarray that you imported.
	A read-only microarray that was imported by another user.
	A CGH design.
	A CGH+SNP design.
	A multipack design.
	The **Detach** button, located at the top of the Data Navigator pane, removes the Data Navigator pane from the main window, and opens it in a separate window.
	Click to find an Array ID or Global Display Name entered in the Search box.

The Array Data Navigator has the following shortcuts and actions.

- Single-click a data folder to expand or collapse it.
- Double-click the name of a folder to display a list of the Array IDs and their attributes in the Sample Manager table.
- Right-click the name of a microarray to open a menu with the following choices.

  Show Properties – Opens the Microarray Properties dialog box, that displays the properties of the selected microarray. See "Microarray Properties" on page 221.

QC Metrics – Opens the QC Metrics Table, that shows the QC Metrics for the selected microarray, and lets you show plot them and show the frequency distribution. See "QC Metrics Table" on page 231, "QC Metrics Graph" on page 228, and "QC Metrics – Frequency Distribution" on page 227.

- Right-click the name of a design build to open a menu with the following choices:

  Show Properties – Opens the Design Properties dialog box, that displays information about the selected design. See "Design Properties" on page 173.

  QC Metrics – Opens the QC Metrics Table, that shows the QC Metrics for the selected microarray, and lets you show plot them and show the frequency distribution. See "QC Metrics Table" on page 231, "QC Metrics Graph" on page 228, and "QC Metrics – Frequency Distribution" on page 227.

Array Data Navigator Search pane The Search pane lets you find all occurrences of an entire name or specific search string in the Array Data Navigator. It also contains several buttons that you can use to move, hide, show or resize the Navigator.

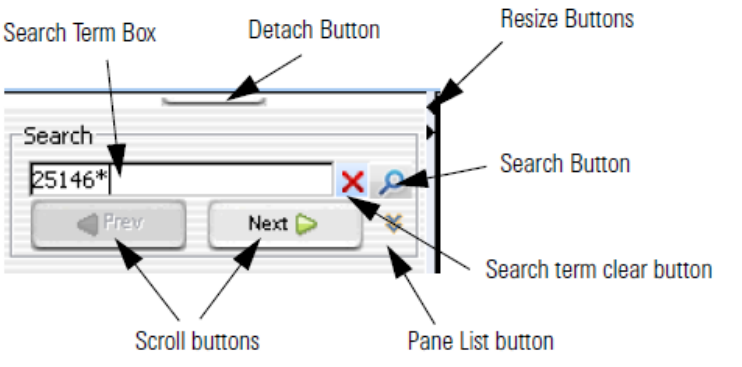


Figure 50 Navigator Search pane

Detach button Click to move the Navigator from the main window of the program and open it in a new, separate window.

Resize buttons Click to hide, show, or expand the Navigator.

Search term box The place where you type your desired search term. Search terms are not case-sensitive, but they must reflect the entire name of an array or other content item that you want to find. You can use asterisks (*) as wildcards to represent groups of unspecified characters. For example, a search term *25887* searches for any content that contains the string "25887".

Pane list Lets you limit a search to a specific pane. Select the name of the desired pane from the list. To select all panes, select **All Panels.** By default, the program searches all panes.

Sample Manager Table

The Sample Manager table displays the list of microarrays that are selected in the Data Navigator. See "Array Data Navigator" on page 97. Each microarray is displayed along with its sample attributes and analysis status. Buttons at the bottom of the tab let you display or edit sample attributes and do other tasks in the list, as described in the table below.

The columns (attributes) in the Sample Manager table can change, but the following required attributes are always displayed, in addition to extraction status:

Table 15 Default columns in the Sample Manager table

Attribute	Description
Array ID	The unique identifier for each array on a microarray slide. This cannot be changed by the user.
Global Display Name	By default, the name of the extraction. This name is used to identify the microarray throughout the program, and can be changed by the user.

Table 15 Default columns in the Sample Manager table

Attribute	Description

Status	• Shows the status of the microarray sample. Samples can have the following Status:
	  ◦ Not Extracted – An array ID that was added, but was not extracted or analyzed using a workflow.
	  ◦ EXTRACTED – Extracted Array ID is in the database, but the sample was not analyzed (possibly due to a workflow failure).
	  ◦ Imported – Sample was imported, but not analyzed. Samples for which you imported the Stats and Params from the Quality tab have this status.
	  ◦ Analyzed – Sample was successfully analyzed, but was not checked out or in, or reviewed.
	  ◦ Check Out – Sample is checked out.
	  ◦ Check In – Sample is checked in.
	  ◦ Reviewed – Sample was signed off.
	• For samples with Status of *Analyzed, Check In, Reviewed, or Check Out (if the sample is checked out by you)* click the status to open the sample in Triage View.
	• Samples that have more than one analysis result are indicated with a "#".

Green Sample	Names of Cy3- labeled samples; can be changed by the user. Also used to associate or display the genotype reference sample for CGH+SNP arrays.
Red Sample	Names of Cy5- labeled samples; can be changed by the user. Also used to associate or display the genotype reference sample for dye-flipped CGH+SNP arrays.
Polarity	Sample polarity (1 or -1); can be changed by the user before import of the extraction, not after import
Extraction status	For arrays extracted using Feature Extraction, shows the results status from the FE results file. See the *Feature Extraction Reference Guide* for more information. For new rows added by the user and not yet imported (white rows), displays *Not Extracted.* For arrays that are imported, not extracted within the program, displays *Imported*.

Table 16 Sample Manager table buttons

Button	Function
Show/Hide Attributes	Opens Show/Hide Columns dialog box where you can select which attributes, FE Statistics, and FE Parameters to display. See "Show/Hide Columns" on page 242.
Add/Edit Attribute	Opens Attributes dialog box where you can add a new attribute or edit an existing attribute. "Attributes" on page 143.
Add Array ID	Lets you add a single Array ID to a folder. See "Add Array ID to Data Folder" on page 134.
Remove Array ID	Removes selected Array ID from the list.
Save Changes	Saves changes made to the array list.
Cancel Changes	Cancels changes made to the array list.

Content tab – Feature Extraction View

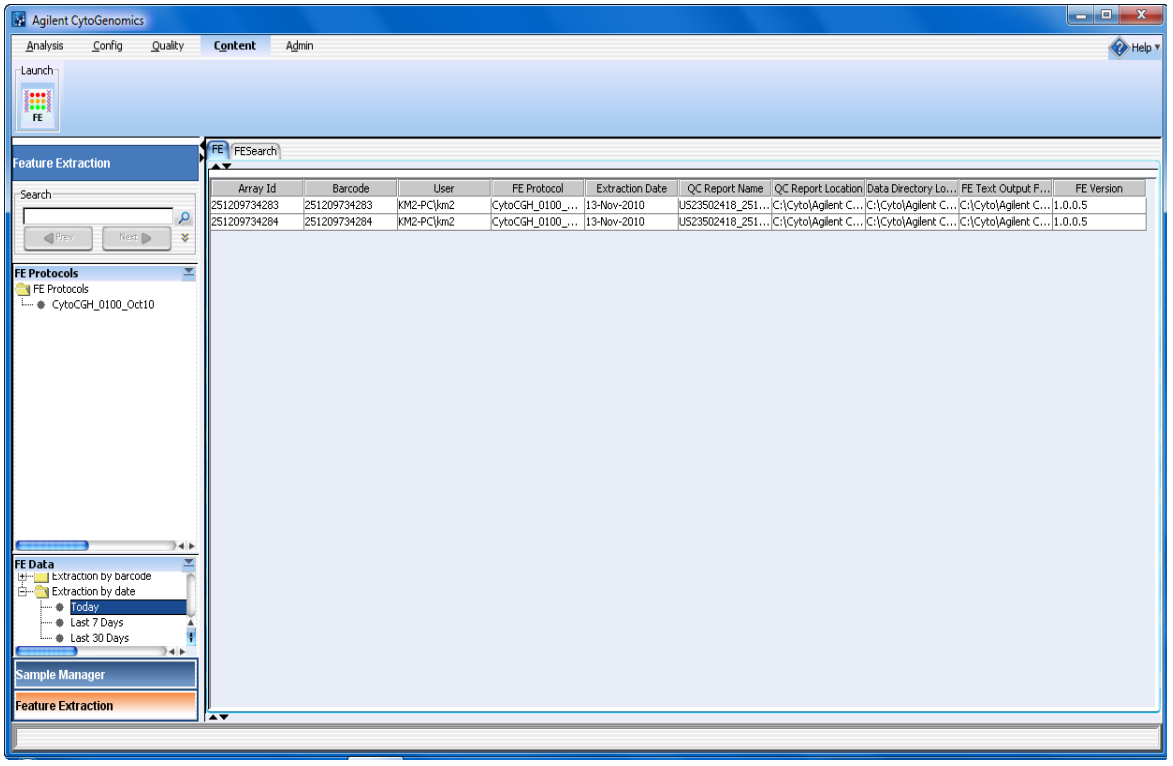


Figure 51 Content tab – Feature Extraction View

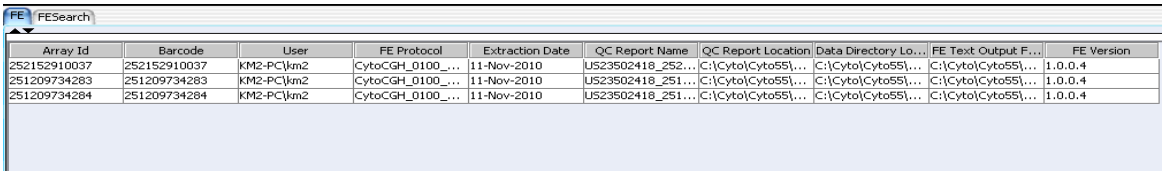
Purpose: To display Feature Extraction protocols and extracted data within the program.

To open: In the Content tab, in the Navigator, click **Feature Extraction**.

The Feature Extraction View in the Content tab lets you show extractions, FE protocol details, and search extractions based on Array ID, User, Barcode, or Extraction Date. You can also open the QC Report .pdf file for an extraction.

Starts the Feature Extraction program, where you can change FE parameters and create protocols. For more information, see the *Feature Extraction User Guide*.

FE tab



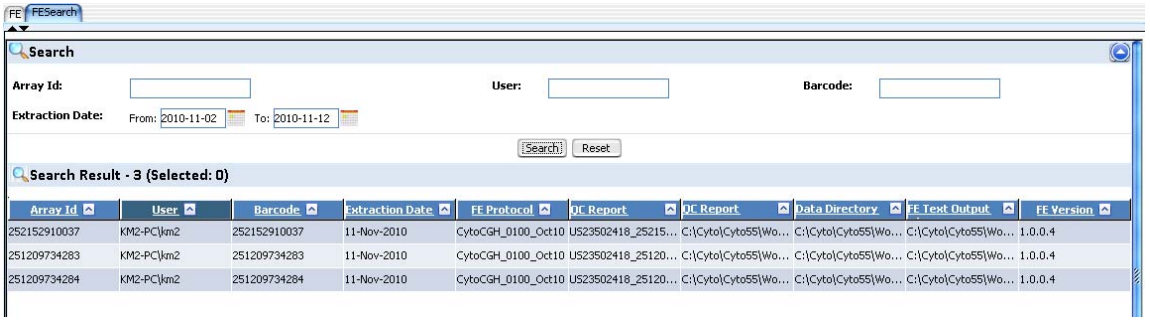
Array Id	Barcode	User	FE Protocol	Extraction Date	QC Report Name	QC Report Location	Data Directory Lo...	FE Text Output F...	FE Version
252152910037	252152910037	KM2-PC\km2	CytoCGH_0100_...	11-Nov-2010	US23502418_252...	C:\Cyto\Cyto55\...	C:\Cyto\Cyto55\...	C:\Cyto\Cyto55\...	1.0.0.4
251209734283	251209734283	KM2-PC\km2	CytoCGH_0100_...	11-Nov-2010	US23502418_251...	C:\Cyto\Cyto55\...	C:\Cyto\Cyto55\...	C:\Cyto\Cyto55\...	1.0.0.4
251209734284	251209734284	KM2-PC\km2	CytoCGH_0100_...	11-Nov-2010	US23502418_251...	C:\Cyto\Cyto55\...	C:\Cyto\Cyto55\...	C:\Cyto\Cyto55\...	1.0.0.4

Figure 52 Sample Manager – FE tab

Purpose: This tab displays the extractions you selected in the Navigator (Extraction by barcode or Extraction by date.) All columns are read-only except for the QC Report Name. Double-click **QC Report Name** to open the QC Report for the selected extraction.

To open: In the Content tab, in the Navigator, click **Sample Manager**, then click the **FE** tab.

FESearch tab



Array Id	User	Barcode	Extraction Date	FE Protocol	QC Report	QC Report	Data Directory	FE Text Output	FE Version
252152910037	KM2-PC\km2	252152910037	11-Nov-2010	CytoCGH_0100_Oct10	US23502418_25215...	C:\Cyto\Cyto55\Wo...	C:\Cyto\Cyto55\Wo...	C:\Cyto\Cyto55\Wo...	1.0.0.4
251209734283	KM2-PC\km2	251209734283	11-Nov-2010	CytoCGH_0100_Oct10	US23502418_25120...	C:\Cyto\Cyto55\Wo...	C:\Cyto\Cyto55\Wo...	C:\Cyto\Cyto55\Wo...	1.0.0.4
251209734284	KM2-PC\km2	251209734284	11-Nov-2010	CytoCGH_0100_Oct10	US23502418_25120...	C:\Cyto\Cyto55\Wo...	C:\Cyto\Cyto55\Wo...	C:\Cyto\Cyto55\Wo...	1.0.0.4

Figure 53 Sample Manager – FESearch tab

Purpose: This tab lets you search for extractions, based on parameters that you select. To sort the list, click the arrow at the top of the column for a given parameter.

To open: In the Content tab, in the Navigator, click **Sample Manager**, then click the **FESearch** tab.

Array Id Type an Array ID to use for the search. Search finds all extractions with the designated Array ID.

User Type the name of a user to use for the search. User name must include the domain name and the user. For example: domain\user. Search returns all extractions performed by the designated user.

Barcode Type a barcode to use for the search. Search returns all extractions that match the designated barcode.

Extraction Date Type a date range to use for the search. Or, click the calendar icon  to select a date. Search returns all extractions performed within the designated date range.

Search Performs a search using the given search conditions.

Reset Clears the search parameters and search results.

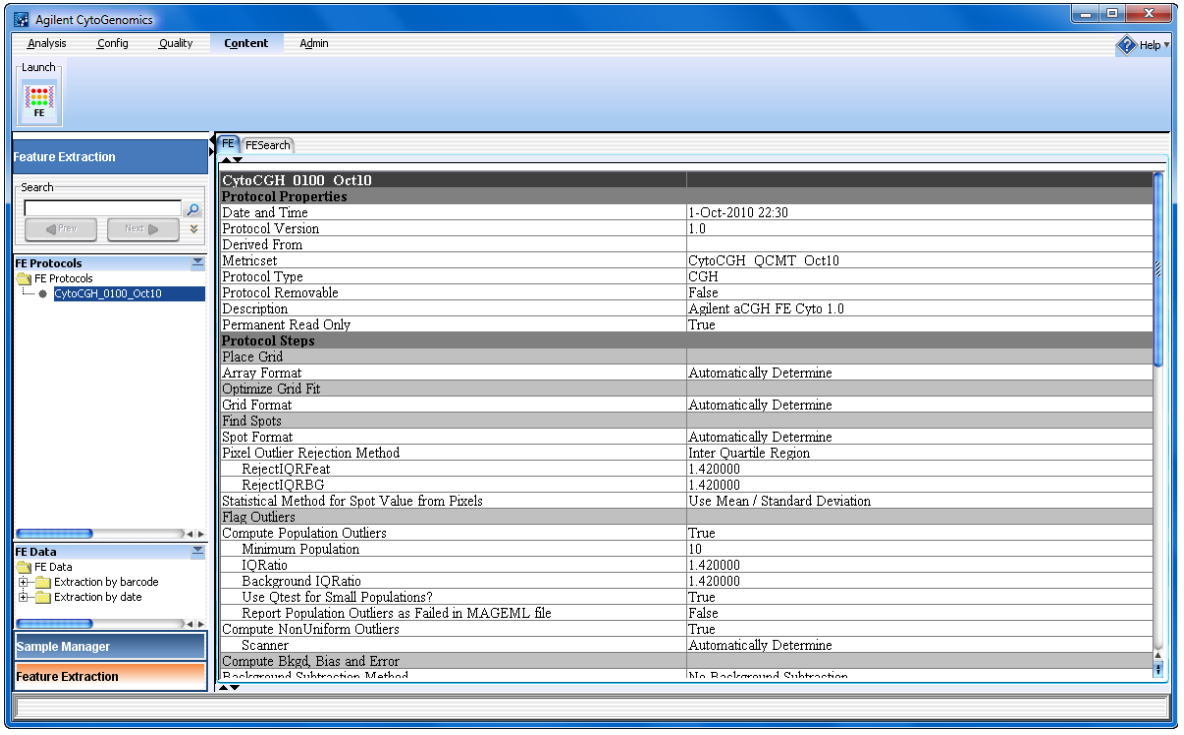


Figure 54 Content tab – Feature Extraction View with protocol selected

Purpose: This view lets you display an FE protocol.

To open: In the Content tab, in the Navigator, click **Feature Extraction**. n the Navigator, under FE Protocols, double-click a protocol to display.

2 Parameter Panels and Dialog Boxes

This chapter describes the parameter panels and dialog boxes for all user roles in Agilent CytoGenomics 1.0.

Parameter Panels

As you select options for an analysis method or workflow, the parameters available for each option appear in specific parameter panels. If you do not need to set any parameters for a specific option, **No input parameters are required** appears.

This section describes the parameter panels that appear in Agilent CytoGenomics 1.0, organized alphabetically by name.

Aberration & LOH Report Parameter Panel

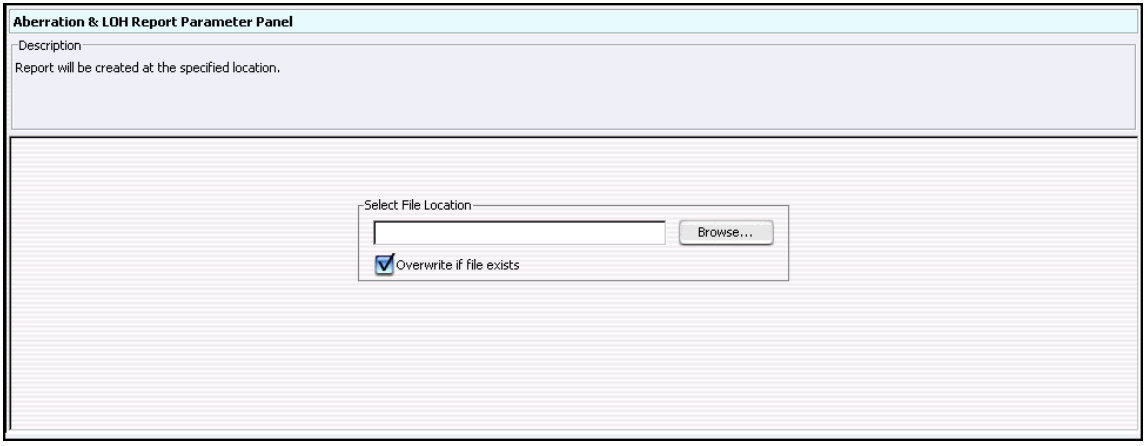


Figure 55 Aberration & LOH Report Parameter Panel

Purpose: To set location for Aberration & LOH Report created by Workflow.

To open: In the Workflow window, under Reports, select **Aberration & LOH Report**

Select File Location Displays the location where the workflow saves the files. To select a location for the report, click **Browse.** An Open dialog box appears. Type a name and select a location for the report, then click **Open.**

Overwrite if file exists If you select this option, the workflow deletes an existing file if it has the same name and location as a generated report.

Aberration Filter Parameter Panel

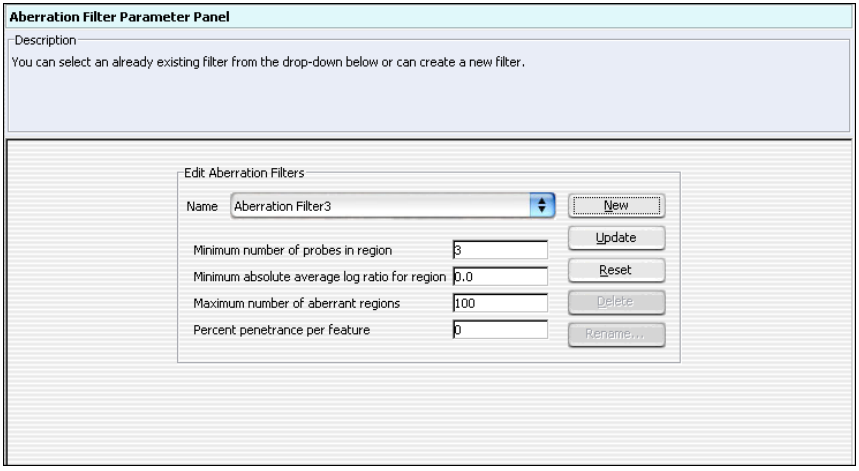


Figure 56 Aberration Filter Parameter Panel

Purpose: This parameter panel is used to create or edit aberration filters. Aberration filters exclude detected aberrations from the output of the workflow, based on selected conditions.

To open: The Aberration Filter Parameter Panel appears when you select **Aberration Filter** under Filter After Analysis in the Analysis Method window.

Name Select the name of the filter you want to use. To create a new aberration filter, click **New**.

New Opens an Input dialog box, where you can type a name for the new aberration filter. To accept the name, click **OK**. The program creates the filter, and adds the new name to the Name list.

Update Saves any changes you make to the filter conditions.

Reset Restores the values of the filter conditions to what they were before you made any changes to them.

Delete Opens a Confirm dialog box that asks you if you want to delete the currently selected filter. To delete the filter, click **Yes**.

Rename Opens an Input dialog box where you can type a new name for the filter. To accept the name, click **OK**.

Minimum number of probes in region Type a whole number. The filter excludes putative aberrant regions that contain fewer probes than the number you type.

Minimum absolute average log ratio for region Type a value. The filter excludes putative aberrant regions if the average log2 ratio within the region is less than the value you type.

Maximum number of aberrant regions Type a whole number. For each microarray, the filter includes up to this number of aberrant regions that have the highest statistical significance.

Percent penetrance per feature Type a value. The filter excludes putative aberrations that have less than the specified minimum percent penetrance across the set of selected arrays.

ADM-1 Parameter Panel

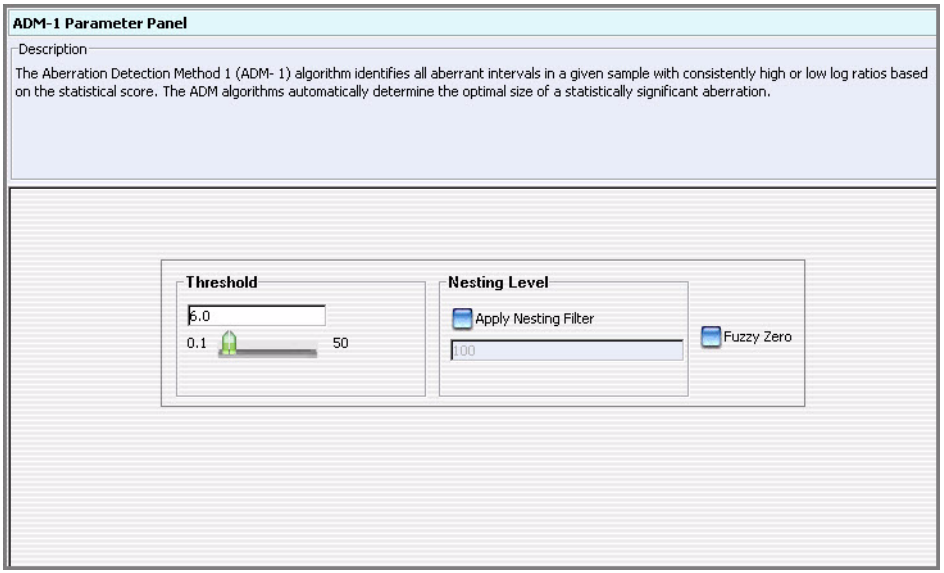


Figure 57 ADM-1 Parameter Panel

Purpose: This panel lets you set the parameters for the ADM-1 aberration detection algorithm. For more information on ADM-1, see Chapter 4, "Statistical Algorithms".

To open: The ADM-1 Parameter Panel appears when you select **ADM-1** under Aberration in the Analysis Method window.

Threshold Type a numerical value from 0.1 to 50, or use the slider to set a value. The threshold is the minimum ADM-1 score for the detection algorithm to consider a given genomic interval significant. In general, increase this value to make the detection process more stringent.

Nesting Level To apply a nesting-level filter, select **Apply Nesting Filter,** then type a whole number from 0 to 2147483647 in the box.

The ADM-1 aberration detection algorithm reports nested aberrations; that is, aberrations within other aberrations. If you set the nesting level to 0, the program reports only the parent aberration, without any child (nested) aberrations. If you set it to 1, the program reports the first level of child

aberrations. By default, the program sets the nesting-level filter to its maximum value, which essentially applies no filter. To make the filter more stringent, decrease the value.

Fuzzy Zero Select **Fuzzy Zero** to apply Fuzzy Zero correction to detected aberrant intervals. This correction applies a "global error model" to the intervals, and can result in fewer errors in aberration calls. For more information on the fuzzy zero algorithm, see "Fuzzy Zero" on page 281.

ADM-2 Parameter Panel

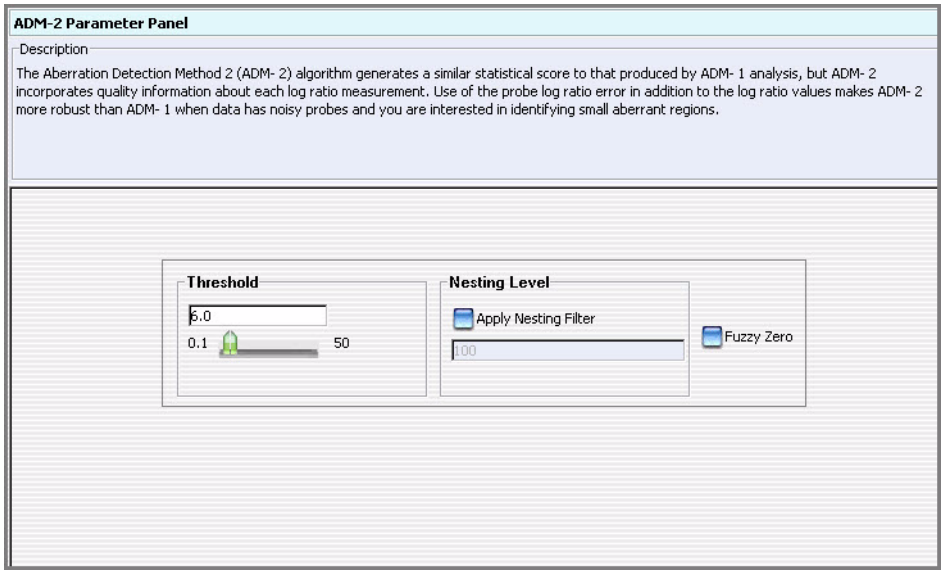


Figure 58 ADM-2 Parameter Panel

Purpose: This parameter panel lets you set the parameters for the ADM-2 aberration detection algorithm. For more information on ADM-2, see Chapter 4, "Statistical Algorithms".

To open: The ADM-2 Parameter Panel appears when you select **ADM-2** under Aberration in the Analysis Method window.

Threshold Type a numerical value from 0.1 to 50, or use the slider to set a value. The threshold is the minimum ADM-2 score for the detection algorithm to consider a given genomic interval significant. In general, increase this value to make the detection process more stringent.

Nesting Level To apply a nesting-level filter, select **Apply Nesting Filter,** then type a whole number from 0 to 2147483647 in the box.

The ADM-2 aberration detection algorithm reports nested aberrations; that is, aberrations within other aberrations. If you set the nesting level to 0, the program reports only the parent aberration, without any child (nested) aberrations. If you set it to 1, the programs reports the first level of child aberrations. By default, the program sets the nesting-level filter to its maximum value, which applies no filter. To make the filter more stringent, decrease the value.

Fuzzy Zero Select **Fuzzy Zero** to apply Fuzzy Zero correction to detected aberrant intervals. This correction applies a "global error model" to the intervals, and can result in fewer aberration call errors. For more information on the fuzzy zero algorithm, see Chapter 4, "Statistical Algorithms".

Array Level Filter Parameter Panel

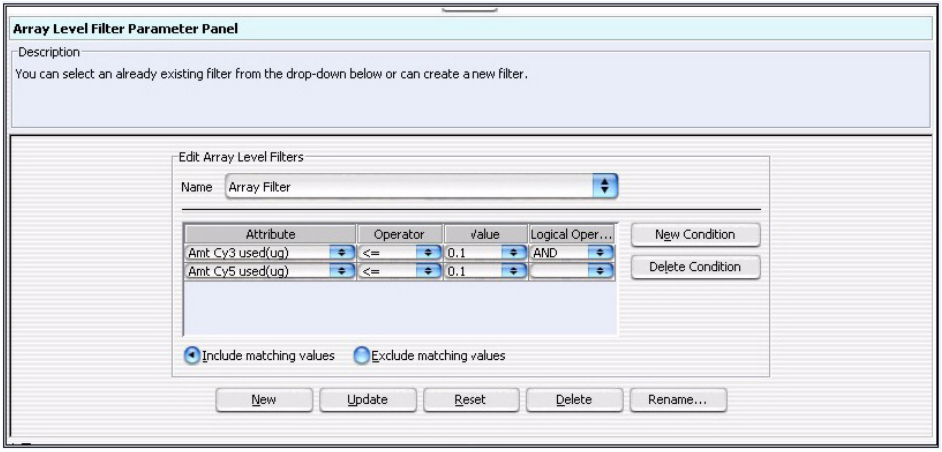


Figure 59 Array Level Filter Parameter Panel

Purpose: This parameter panel lets you create or edit array level filters. An array filter excludes arrays from the workflow based on selected conditions.

To open: The Array Filter Parameter Panel appears when you select **Array Level Filter** in Filter Before Analysis in the Analysis Method window.

Name Select the name of the array filter you want to edit. To create a new filter, and add its name to the list, click **New**.

Filter conditions table Displays the conditions defined for the selected array level filter. When you create or edit the filter, for each condition (row), select options from the lists. In Value, select an option from the list, if available. Otherwise, type a value, then press **Enter**. To add another row to the table, click **New Condition**. Each condition has these elements:

- Attribute – The array attribute evaluated by the filter
- Operator – How the filter uses the entry in Value to evaluate arrays. For example, the >= operator configures the filter to include or exclude arrays where the selected attribute is greater than or equal to the entry in Value.
- Value – The value the filter uses as the basis to evaluate an attribute. For example, if you select the **AmtCy3used(ug)** attribute, the = operator, and a value of **5**, the filter includes or excludes an array if its AmtCy3used(ug) attribute is equal to 5.
- Logical Operator – (Available only if you configure more than one condition) The relationship between the condition and the next one in the list. For example, if you select **AND** in Logical Operator for the first condition, the filter includes or excludes an array if it passes both the first condition and the next condition.

New Condition Adds a new, blank condition (row) to the table.

Delete Condition Removes a condition from the list. To remove a condition, click anywhere within the condition row, then click **Delete Condition**.

Include/Exclude matching values Select one of these options:

- Include matching values – If an attribute passes the filter conditions, the program includes the array in the analysis.
- Exclude matching values – If an attribute passes the filter conditions, the program excludes the array from the analysis.

New Opens an Input dialog box, where you can type a name for the new filter. To accept the name, click **OK**. The program creates the filter, and adds the new name to the Name list.

Update Saves any changes you make to the filter conditions without closing the dialog box.

Reset Restores the values of the filter conditions to what they were before you made any changes to them.

Delete Opens a Confirm dialog box that asks you if you want to delete the selected filter. To delete the filter, click **Yes**.

Rename Opens an Input dialog box where you can type a new name for the filter. To accept the name, click **OK**.

Centralization Parameter Panel

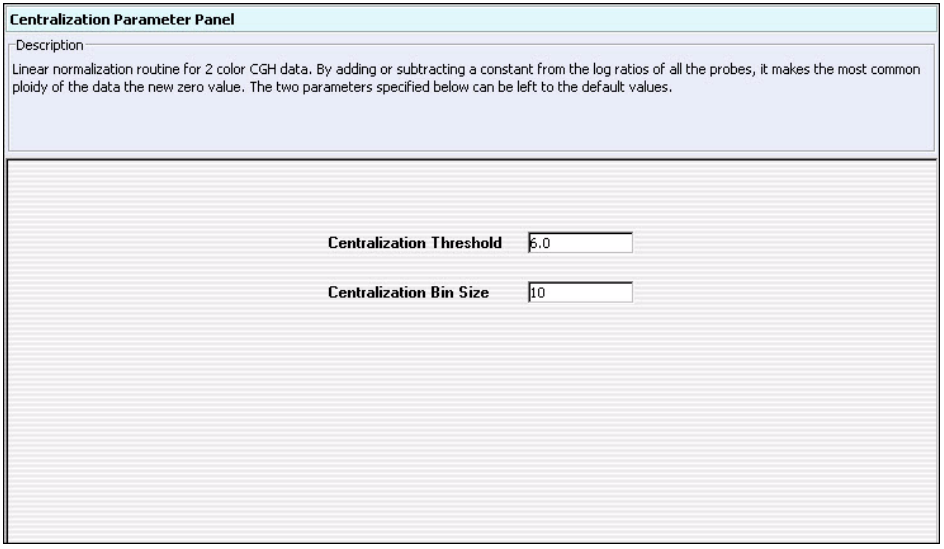


Figure 60 Centralization Parameter Panel

Purpose: This parameter panel lets you set up centralization parameters. Centralization recenters log ratio values. It finds a constant value to subtract from or add to all values, and makes sure that the zero-point reflects the most-common-ploidy state. For a description of the centralization algorithm, see Chapter 4, **"Statistical Algorithms"**.

To open: The Centralization Parameter Panel appears when you select **Centralization** in Normalization in the Analysis Method window.

Centralization Threshold

This value is the ADM-1 threshold used to call aberrations for the centralization algorithm. Ideally, set this value to the ADM-1 threshold defined in the current analysis. However, because the centralization value is fairly robust over a wide range of threshold values, Agilent does not recommend that you change the default settings for this algorithm.

Centralization Bin Size

Type a whole number. To improve performance, the algorithm finds the average of groups (or "bins") of contiguous probes and runs the centralization process on these averages, rather than on individual probes. The centralization bin size is the number of probes that the program groups together for each of these averages. The default value is 10 probes.

NOTE

Centralization is recommended for SNP Copy Number and LOH analyses.

CGH Aberration Report Parameter Panel

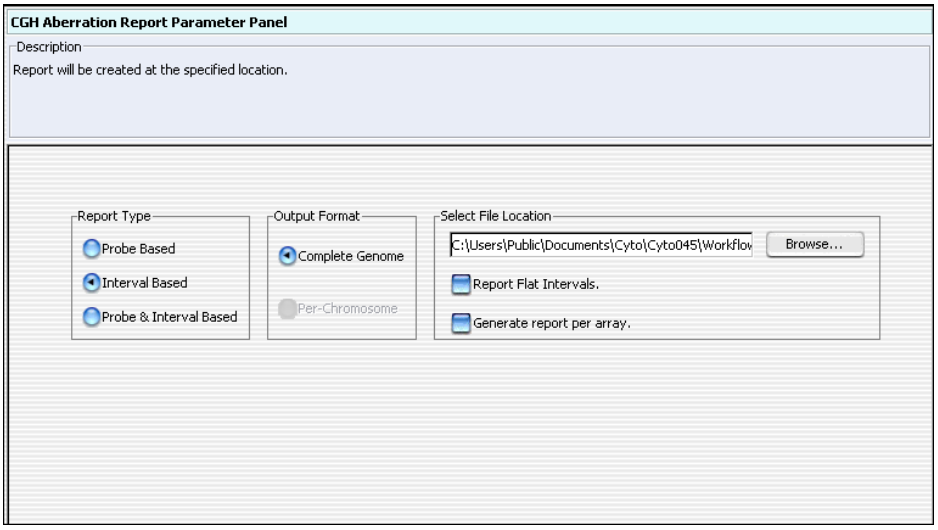


Figure 61 CGH Aberration Report Parameter Panel

Purpose: This parameter panel lets you configure the CGH Aberration Report, and select a location for it. This report describes regions that have detected aberrations. You can report these regions by genomic interval, by probe, or both. The program reports aberrations separately for each array in the workflow, and creates one or more *.xls files you can work with in Microsoft Excel. See "CGH Aberration Reports" on page 258.

To open: The CGH Aberration Report Parameter Panel appears when you select **CGH Aberration Report** under Reports in the Workflow window.

Report Type Configures the organization of reported aberrations. Select one of these options:

Option	Description
Probe Based	Creates a report that contains one line for each probe showing an aberration.
Interval Based	Creates a report that contains one line for each aberrant genomic interval.
Probe & Interval Based	Creates both a probe-based report and an interval-based report.

Output Format Select one of these options:

Option	Description
Complete Genome	Creates a single report file for each requested report type.
Per-Chromosome	(Available only for probe-based reports) Creates a separate report file for each chromosome.

Select File Location Displays the location to which the workflow saves the CGH Aberration Report. To select a location, click **Browse.** An Open dialog box appears. Select a location and type a name for the report, then click **Open.**

Report Flat Intervals Select this to have aberration intervals reported without any nested structure.

Generate report per array Select this to generate a report after each microarray is analyzed. This lets you look at the results for individual microarrays before the workflow is completed.

Cyto Report Parameter Panel

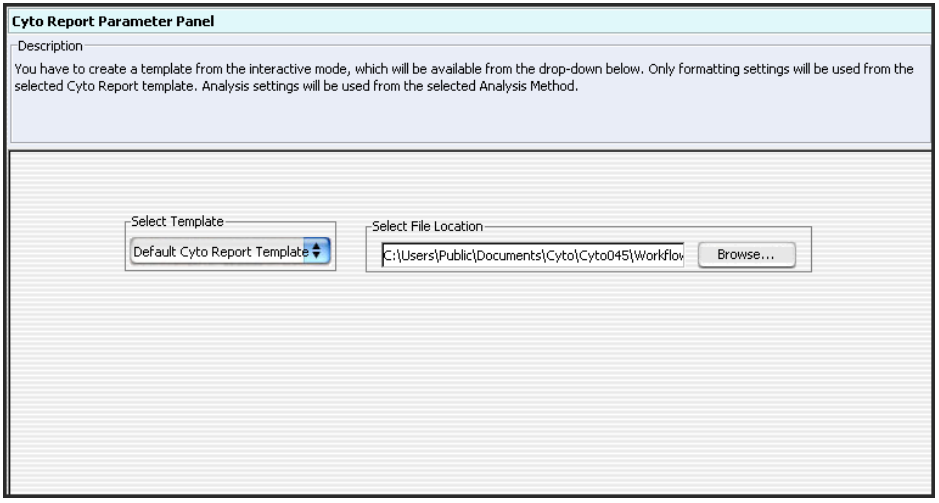


Figure 62 Cyto Report Parameter Panel

Purpose: This panel lets you select the Cyto Report template to use, and a location for the report. Cyto Reports summarize analysis settings and detected aberrations by array. The workflow creates separate PDF and .XML files for each array.

To open: The Cyto Report Parameter Panel appears when you select **Cyto Report** in Reports in the Workflow window.

Select Template Select the desired Cyto Report template. You create and edit Cyto Report templates in the Config tab. Also, when you generate a Cyto Report in a workflow, the program uses the data analysis settings from the workflow, not from the selected Cyto Report template. The program uses only the format settings from the template.

Select File Location Click **Browse.** The Select report folder dialog box appears, where you can type a name and select a location for the Cyto Report file(s). The workflow creates a new folder that contains all of the files.

Design Level Filter Parameter Panel

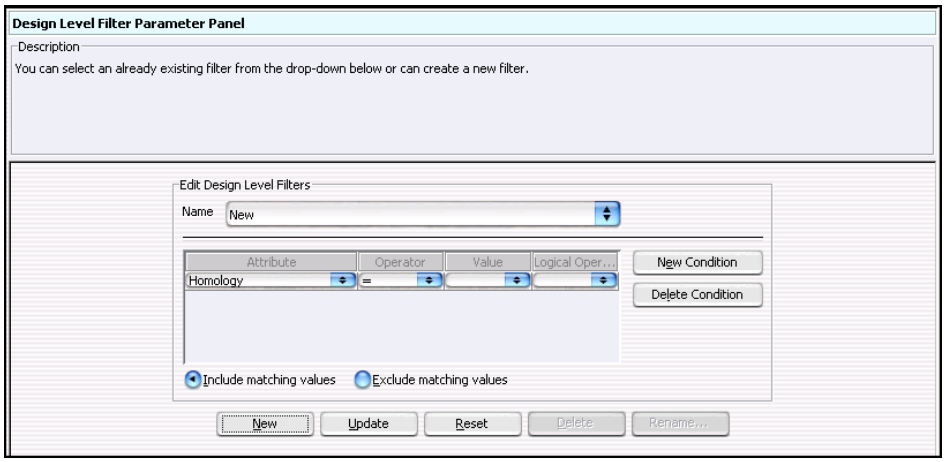


Figure 63 Design Level Filter Parameter Panel

Purpose: This parameter panel is used to create or edit design filters that let you include or exclude probes from analysis based on selected design attributes.

To open: In the Analysis Method window, under Filter Before Analysis, select the box next to **Design Level Filter.**

Name Select the name of the design filter you want to edit. To create a new filter, and add its name to the list, click **New.**

Filter Conditions table For each condition (row), select options from the list or type a value, then press **Enter.** To add another row to the table, click **New Condition.**

Each condition has these elements:

- Attribute – The design attribute evaluated by the filter.
- Operator – How the filter uses the entry in Value to evaluate data. For example, the **>=** operator configures the filter to include or exclude probes where the selected attribute is greater than or equal to the entry in Value.

- Value – The value the filter uses as the basis to evaluate an attribute. You either select a value or, for some attributes, type a value. For example, if you select the **Homology** attribute with an operator of =, the filter will include or exclude probes that exhibit homology if the value is set to **1.** (The homology attribute is assigned to probes that have more than one mapping in the genome or probes that have secondary hits that are not perfect matches.)
- Logical Operator – (Available only if you configure more than one condition) The relationship between the condition and the next one in the list. For example, if you select **AND** in Logical Operator for the first condition, the filter includes or excludes an interval if it passes both the first condition and the next condition.

New Condition Adds a new, blank condition (row) to the table.

Delete Condition Removes a condition from the list. To remove a specific condition, click anywhere within the condition, then click **Delete Condition.**

Include/Exclude matching values Select one of these options:

- **Include matching values** – If a probe passes the filter condition, the program *includes* it in the analysis.
- **Exclude matching values** – If a probe passes the filter condition, the program *excludes* it from the analysis.

New Opens an Input dialog box, where you can type a name for the new filter. To accept the name, click **OK.** The program creates the filter, and adds the new name to the Name list.

Update Saves any changes you make to the filter conditions without closing the dialog box.

Reset Restores the values of the filter conditions to what they were before you made any changes to them.

Delete Opens a Confirm dialog box that asks you if you want to delete the selected filter. To delete the filter, click **Yes.**

Rename Opens an Input dialog box where you can type a new name for the filter. To accept the name, click **OK.**

Feature Level Filter Parameter Panel

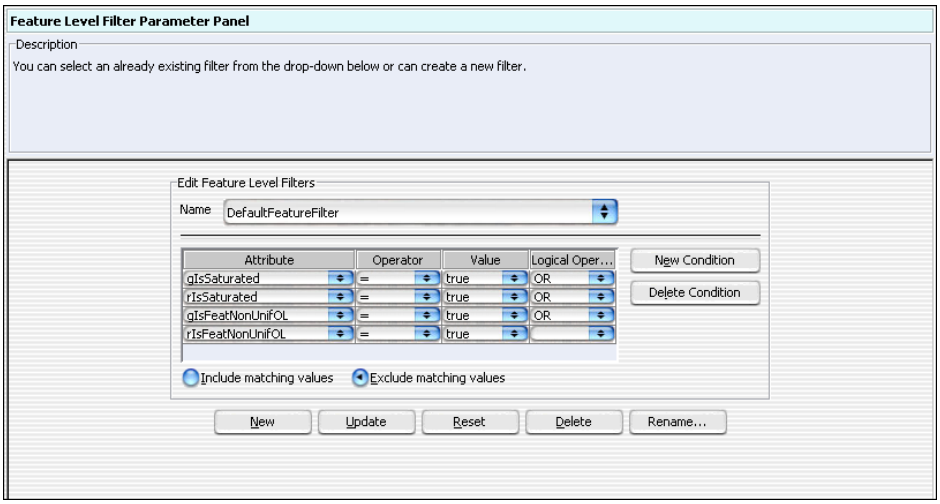


Figure 64 Feature Level Filter Parameter Panel

Purpose: This parameter panel lets you select a feature level filter, display a description of its filtering conditions, or create a new filter. Feature level filters include or exclude data from the workflow, based on selected feature conditions.

To open: The Feature Filter Parameter Panel appears when you select **Feature Level Filter** under Filter Before Analysis in the Analysis Method window.

Name Select the name of the feature filter you want to edit. To create a new filter, and add its name to the list, click **New**.

Filter Conditions table Below the Name is a list that displays the conditions defined for the selected feature level filter. For each condition (row), select options from the list or type a value, then press **Enter.** To add another row to the table, click **New Condition**. Each condition has these elements:

- Attribute – The feature attribute evaluated by the filter.

- Operator – How the filter uses the entry in Value to evaluate arrays. For example, the >= operator configures the filter to include or exclude features where the selected attribute is greater than or equal to the entry in Value.
- Value – The value the filter uses as the basis to evaluate a feature. For example, if you select the **gIsSaturated** attribute, the = operator, and a value of **true**, the filter includes or excludes a feature if its gIsSaturated attribute is true.
- Logical Operator – (Available only if you configure more than one condition) The relationship between the condition and the next one in the list. For example, if you select **AND** in Logical Operator for the first condition, the filter includes or excludes an array if it passes both the first condition and the next condition.

New Condition Adds a new, blank condition (row) to the table.

Delete Condition Removes a condition from the list. To remove a specific condition, click anywhere within the condition row, then click **Delete Condition**.

Include/Exclude matching values Select one of these options:

- Include matching values – If a feature passes the filter condition, the program includes it in the analysis.
- Exclude matching values – If a feature passes the filter condition, the program excludes it from the analysis.

New Opens an Input dialog box, where you can type a name for the new filter. To accept the name, click **OK**. The program creates the filter, and adds the new name to the Name list.

Update Saves any changes you make to the filter conditions without closing the dialog box.

Reset Restores the values of the filter conditions to what they were before you made any changes to them.

Delete Opens a Confirm dialog box that asks you if you want to delete the currently selected filter. To delete the filter, click **Yes**.

Rename Opens an Input dialog box where you can type a new name for the filter. To accept the name, click **OK**.

GC Correction Parameter Panel

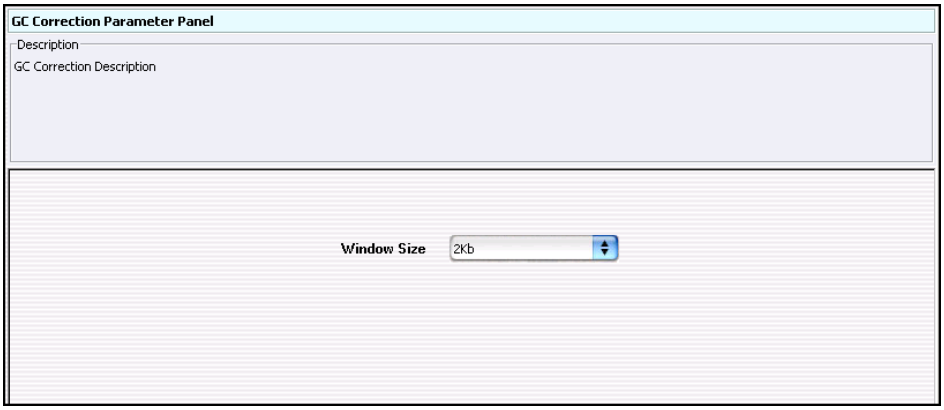


Figure 65 GC Correction parameter panel

Purpose: To select the window size to use for GC content correction.

To open: This panel appears when you click **GC Correction** under Normalization in the Analysis Method window.

Window Size Select a window size from the drop-down menu of choices.

LOH Parameter Panel

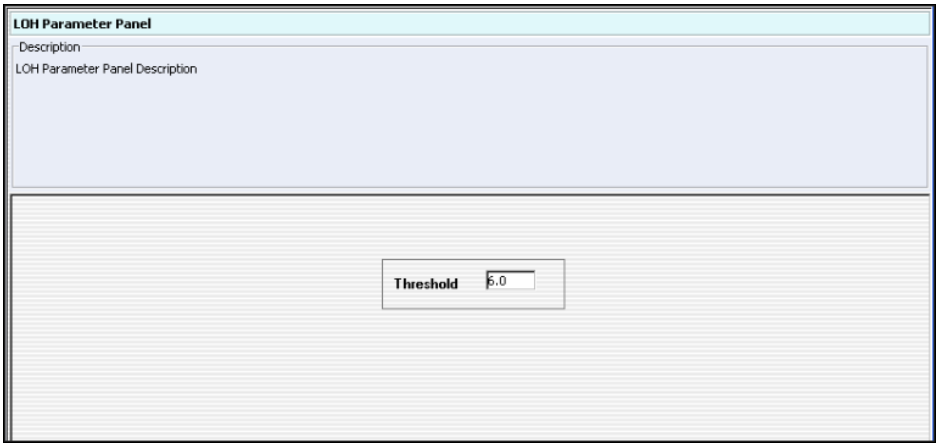


Figure 66 LOH Parameter Panel

Purpose: To set the threshold level for calculation of LOH (Loss or lack of heterozygosity) regions.

To open: In the Analysis Method window, under SNP Algorithm, select **LOH**.

Threshold Type the threshold to use for the LOH calculation. For more information on this algorithm, see "LOH (Loss or lack of heterozygosity) algorithm" on page 322.

Metric Set Filter Parameter Panel

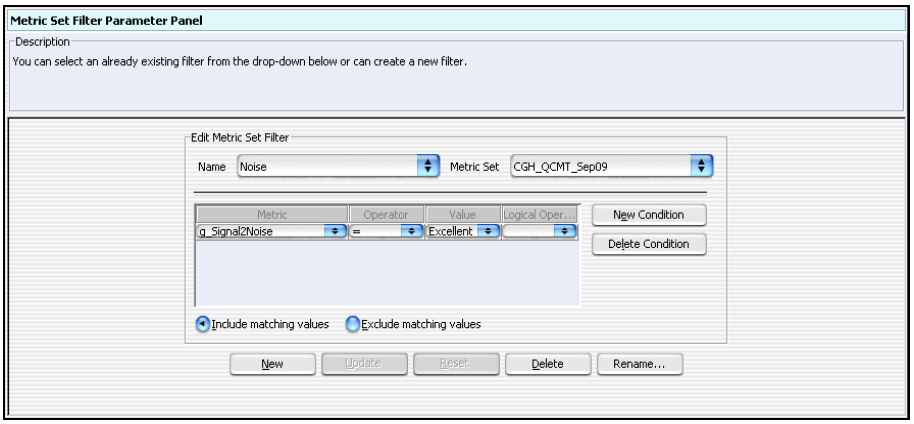


Figure 67 Workflow - Metric Set Level Filter Parameter Panel

Purpose: In this panel, you create or edit a metric set filter that is used to include or exclude arrays from the workflow, based on their QC metrics. For each metric set selected, you select conditions for how to apply the filter.

To open: This panel is displayed when you select **Metric Set Filter** under Metric Evaluation in the Workflow Navigator.

Edit Metric Set Filter

Name Select the name of the metric set filter you want to edit. To create a new filter, and add its name to the list, click **New.**

Metric Set The default metric set for the selected application is shown. Click the arrow to select a different metric set.

Filter Conditions table For each condition (row), select an option from the list, if available. Otherwise, type a value, then press **Enter.** To add another row to the table, click **New Condition.**

Each condition has these elements:

- Metric – The QC metric evaluated by the filter.

- Operator – How the filter uses the entry in Value to evaluate arrays. For example, the **Is** operator configures the filter to include or exclude arrays where the calculated extraction result of the selected metric is the same as the threshold selected in Value.
- Value – The value the filter uses to evaluate an array. For example, if you select the **g_BGNoise** metric and the Operator **Is**, when you select the **Evaluate** Value, the array will pass the filter if its extraction result is in the Evaluate threshold range.
- Logical Operator – (Available only if you configure more than one condition) The relationship between the condition and the next one in the list. For example, if you select **AND** in **Logical Operator** for the first condition, the filter includes or excludes an array if it passes both the first condition and the next condition.

New Condition Adds a new, blank condition (row) to the table.

Delete Condition Removes a condition from the list. To remove a specific condition, click anywhere within the condition, then click **Delete Condition.**

Include/Exclude matching values Select one of these options:

- **Include matching values** – If an array meets the filter conditions, the program *includes* it in the analysis.
- **Exclude matching values** – If an array meets the filter conditions, the program *excludes* it from the analysis.

New Opens an Input dialog box, where you can type a name for the new filter. To accept the name, click **OK.** The program creates the filter, and adds the new name to the Name list.

Update Saves any changes you make to the filter conditions.

Reset Restores the values of the filter conditions to what they were before you made any changes to them.

Delete Opens a Confirm dialog box that asks you if you want to delete the selected filter. To delete the filter, click **Yes.**

Rename Opens an Input dialog box where you can type a new name for the filter. To accept the name, click **OK.**

Run Analysis Application Panel

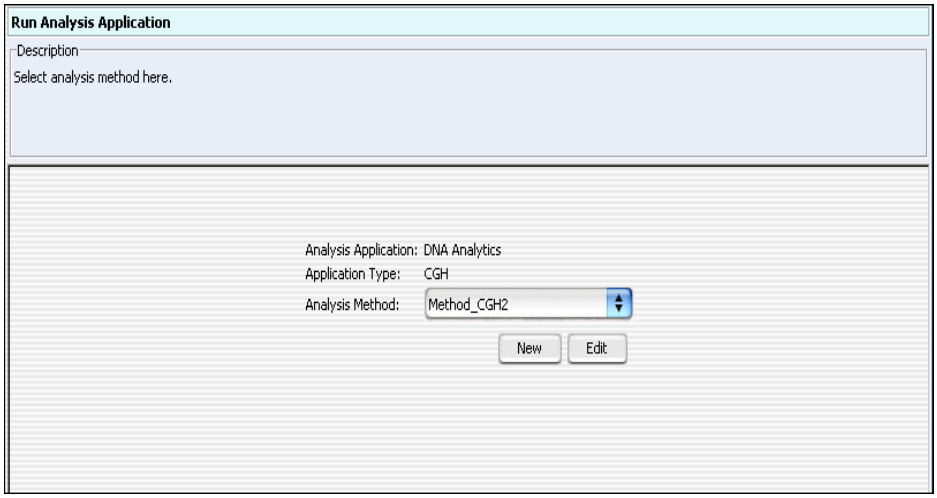


Figure 68 Workflow – Run Analysis Application panel

Purpose: This panel is used to display the analysis application and application type, and to select, create, or edit the analysis method to use for the workflow.

To open: This panel is displayed when Run Analysis is selected under Analysis in the Workflow Navigator.

Analysis Method Click the arrows to select the Analysis Method to run in the workflow.

New Opens the Create Analysis Method dialog box, where you type a name for the new analysis method. The Analysis Method window then opens, where you can create and save the new analysis method for the workflow.

Edit Opens the Analysis Method window, where you can edit the selected analysis method.

SNP Copy Number Parameter Panel

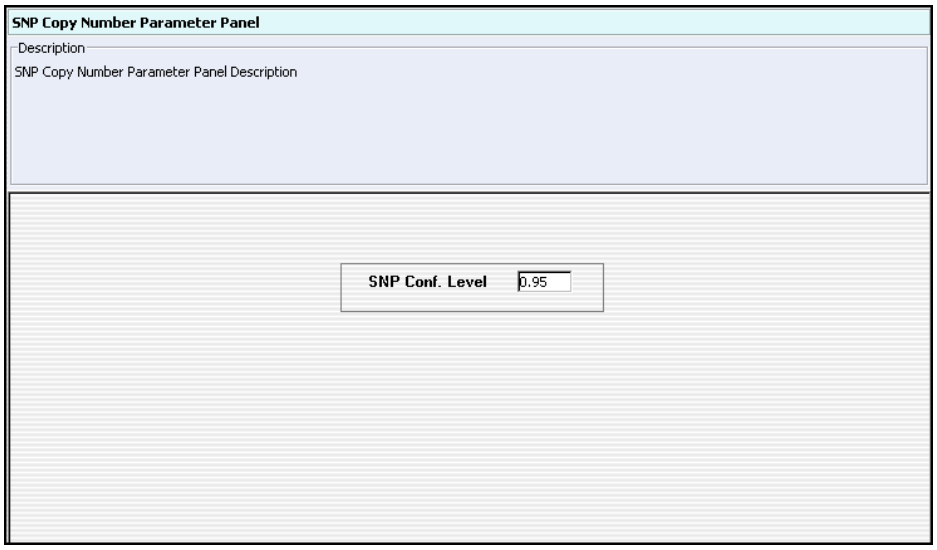


Figure 69 SNP Copy Number Parameter Panel

Purpose: This panel is used to type a confidence level value to use in the SNP Copy Number calculation.

To open: In the Analysis Method window, under SNP Algorithm, select **SNP Copy Number**.

SNP Conf. Level Type a value for the confidence level to use in the SNP copy number calculation.

SNP Genotype Report Parameter Panel

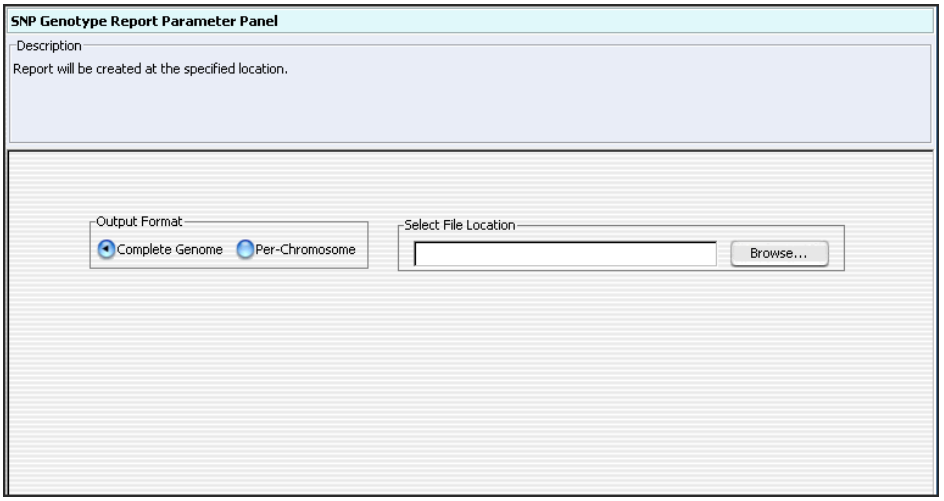


Figure 70 SNP Genotype Report Parameter Panel

Purpose: To set the format and storage location for SNP Genotype Reports generated in the Workflow.

To open: In the Analysis Method window, under Reports, select **SNP Genotype Report**

Output Format Select one of these options:

- **Complete Genome –** Creates a single report file.
- **Per- Chromosome –** Creates a separate report file for each chromosome.

Select File Location Displays the location where the workflow saves the files. To select a location for the report, click **Browse.** An Open dialog box appears. Type a name and select a location for the report, then click **Open.**

Overwrite if file exists If you select this option, the workflow deletes an existing file if it has the same name and location as a generated report.

Z Score Parameter Panel

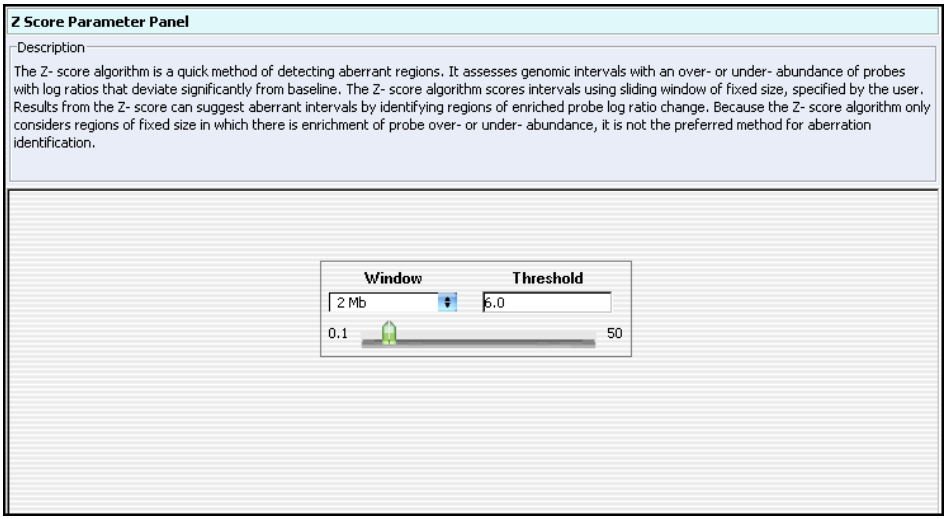


Figure 71 Z Score Parameter Panel

Purpose: This parameter panel lets you configure the Z-Score aberration detection algorithm. The Z-Score algorithm is a quick method of detecting aberrant regions. It calculates intervals using a sliding window of fixed size, and is especially useful in the exploratory phase of CGH data analysis. For more information on the Z-Score algorithm, see "Z-Scoring for Aberrant Regions" on page 287.

To open: The Z Score Parameter Panel appears when you Select **Z Score** under Aberration in the Analysis Method window.

Window The size of the sliding window. You can type a window size in Kb or Mb – for example, `2 Mb`. You can also select a value from the list.

Alternatively, you can specify the window size in terms of data points (pt). This sets the window size to the specified number of contiguous data points. Type the desired number of data points, for example `30 pt`, or select a value from the list.

Threshold The minimum Z-Score for the Z-Score algorithm to consider a region aberrant. Type a value from 0.1 to 50, or use the slider to set the value. In general, to increase the stringency of aberration detection, increase the threshold value.

Dialog Boxes

Add Aberration Call

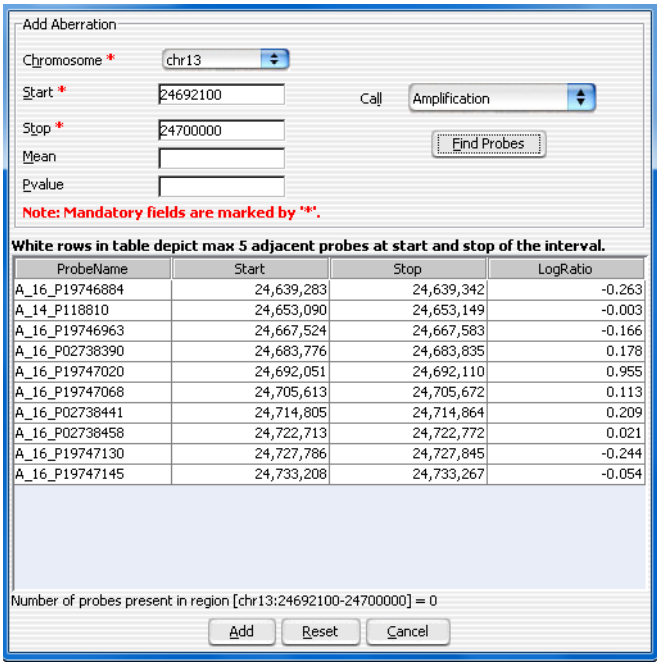


Figure 72 Add Aberration Call dialog box

Purpose: To add an aberration to the results in Triage View.

To open: In Triage View, click **Add Call**. See "Triage View" on page 59.

Chromosome Click arrow and select the chromosome for which you want to add the aberration.

Call Select the type of aberration for the call. Choices are Amplification or Deletion.

Start Type the starting location for the call.

Stop Type the end location for the call.

Find Probes Locates probes within the Start/Stop range, and displays them in the table.

Mean If desired, type the mean value for the call.

Pvalue If desired, type the pValue for the call.

Add Adds the aberration call to the Intervals or SNPs table, and closes the dialog box.

Reset After you make changes, this clears the boxes.

Cancel Cancels the changes and closes the dialog box without adding the call.

Add Array ID to Data Folder

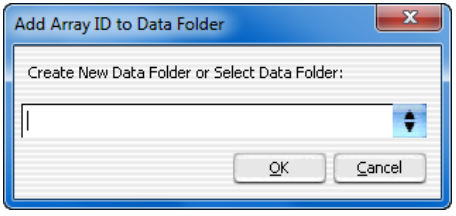


Figure 73 Add Array ID to Data Folder dialog box

Purpose: Adds an Array ID row to the list in the Sample Manager table. Also used to add a new Array folder to the Array Data.

To open: This dialog box appears when you click **Add Array ID** at the bottom of the Sample Manager table.

Select Node Click the arrow and select one of the current data folders, or type the name of a new data folder.

OK Click **OK** to add a blank array row to the Sample Manager table.

Cancel Click **Cancel** to cancel the operation without adding an array ID.

NOTE When you use the Add Array ID function to add an Array ID, you must enter the required attributes for the new array and then save the changes in order for it to appear in the Array Data list. If you don't add all of the required attributes, when you save the row, an error message appears and the row is not saved.

Add Image Pack Information for FE Extraction

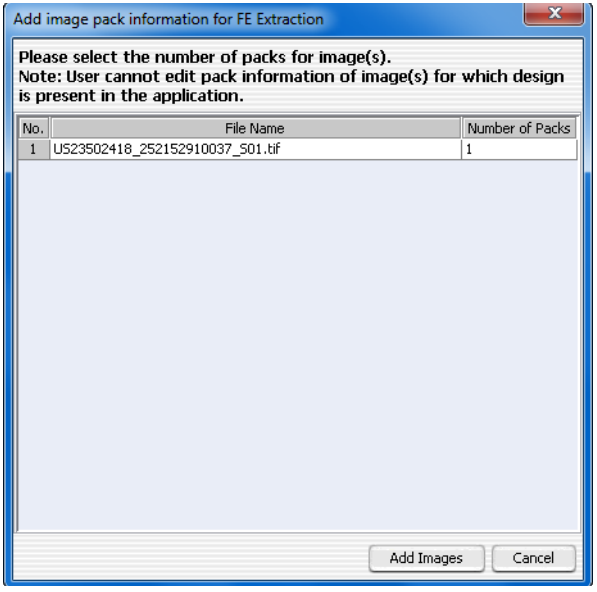


Figure 74 Add image pack information for FE Extraction dialog box

Purpose: To display or select the number of image packs for multi-pack image files to be opened.

To open: This dialog box appears when you click **Open** after you select an image file for a workflow in the Open dialog box. See "Import Image Files" on page 26.

Number of Packs For each image file, the number of packs is displayed. For images with no design in the database, click the **Number of Packs** and select the number of packs included in the image file to be imported. For example, for a 2-pack array, select **2** for Number of Packs. The Number of Packs is set to 1 by default.

Add Images Click to add the images to the workflow.

Cancel Closes the dialog box. No images are added.

Add User Dialog

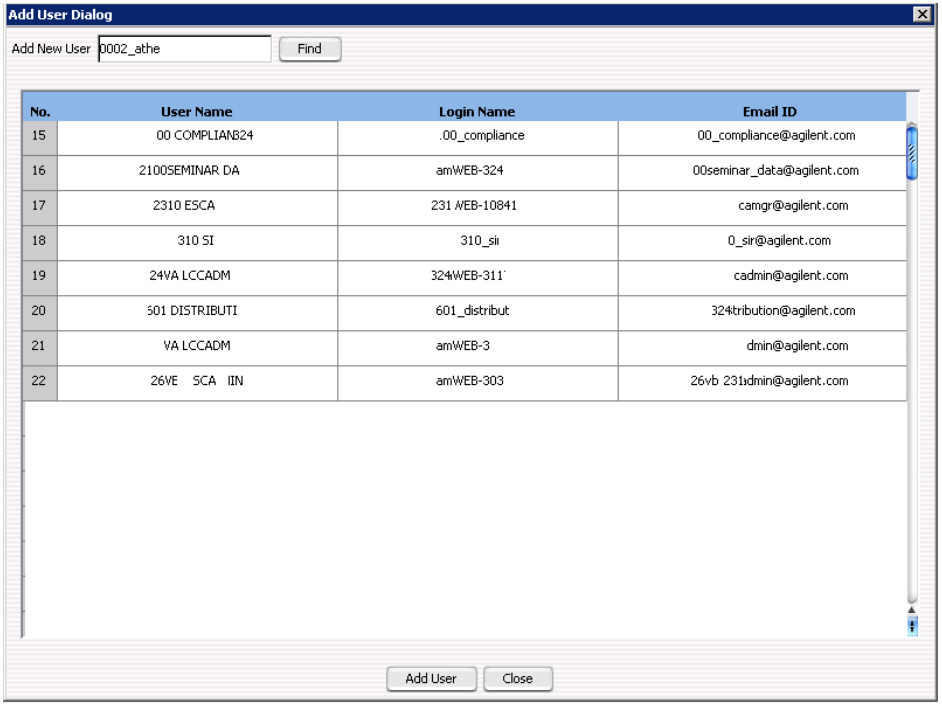


Figure 75 Add User Dialog box

Purpose: This dialog lets you select a user to add to the Agilent CytoGenomics 1.0 database. After you add a user, you should click Edit to verify or change the user role for that user. See "Edit User Roles" on page 183.

To open: In the Admin tab, under Users, click **Manage Users** and then at the bottom of the window, click **Add New User**.

Add New User Type a name or part of a name to look for in the list of users.

Find Click this to find a user in the list that matches the user you typed in the Add New User box.

Add User Select the user you want to add in the list of available users, and click this to add that user to the database.

Close Closes the dialog box.

Agilent Feature Extraction Importer

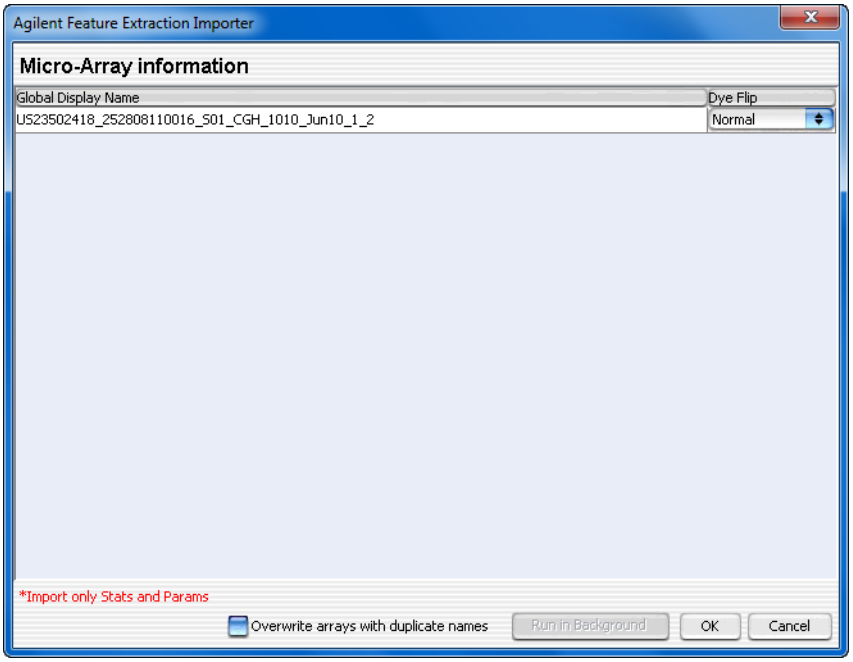


Figure 76 Agilent Feature Extraction Importer dialog box

Purpose: Lets you edit the name of the FE data file you intend to import, and to select if you want to flip the red/green ratio for the data.

To open: In the Quality tab, under QC Data and Tools, click **Import File.** In the menu, select **FE Stats and Parameters.** In the Import FE Files dialog box, select the desired FE data file(s), then click **Open.**

Global Display Name Shows the global display name of each microarray to import. You can change the names of the files to names that you are more likely to recognize or remember.

Dye Flip For each array:

Select **Normal** if:

- The test samples were labeled with cyanine-5 (red).

- The control samples were labeled with cyanine-3 (green).
- The imported ratio (test/control) should be reported directly.

Select **Flipped** if:

- The test samples were labeled with cyanine-3 (green).
- The control samples were labeled with cyanine-5 (red).
- The imported ratio (control/test) should be reported with the ratio inverted (test/control).

The program does not combine dye-flip pairs.

Overwrite arrays with duplicate names Mark this option to replace existing file(s) in the program with the imported one(s), if they have the same name(s).

OK Imports the files in the foreground. You cannot use your computer for other purposes while the import occurs.

Cancel Cancels the entire import process without importing anything.

Attribute File Importer-Map Column Headers

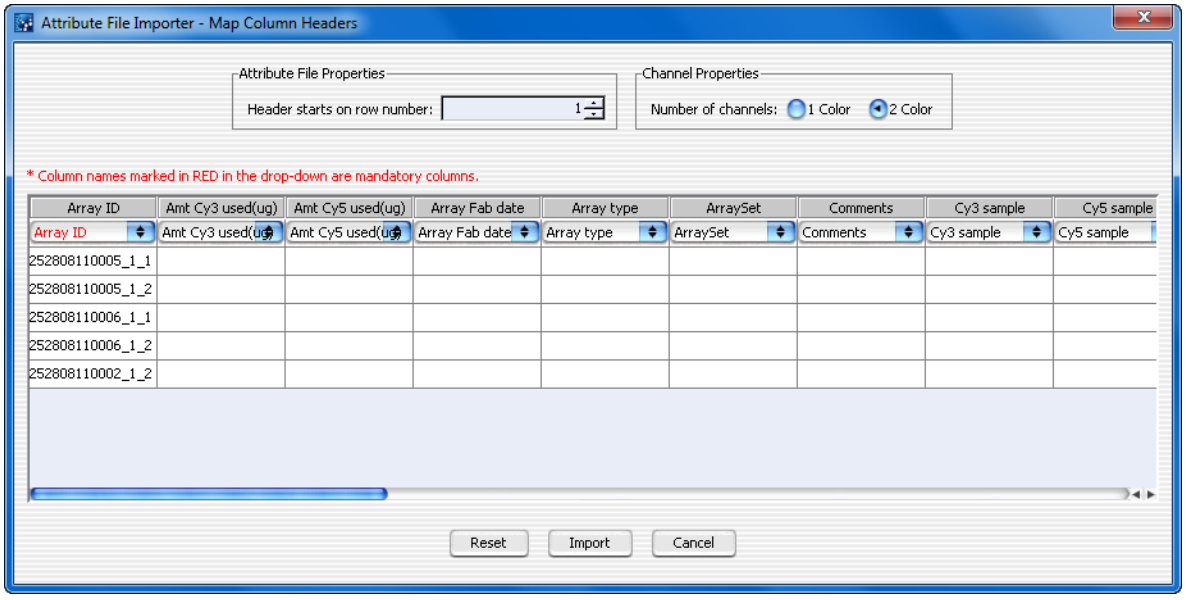


Figure 77 Attribute File Importer - Map Column Headers dialog box

Purpose: Use this dialog box to assign column headers from an attribute file created with a spreadsheet program to the headers expected by the Sample Manager (if necessary). This dialog box is also used to import the attribute file.

To open: This dialog box opens when you click **Open** in the Import Attribute Files dialog box.

To map a column header

1. Click the arrow and select the attribute header to be assigned from the list of currently-defined attributes.
2. When all columns are assigned, click **Import** to import the file. See Table 17 for a list of required columns.

Header starts on row number Select the number of the row that contains the header in the imported file.

Channel Properties Click to select the number of channels for the file. For a 1-color experiment, select **1 Color**. For a 2-color experiment, select **2 Color**.

Reset Resets the column headers to the defaults when the file was opened.

Import Imports the file. A list of exceptions and conditions that may occur when you import a file are shown in Table 18 on page 142.

Cancel Cancels the file import operation and closes the dialog box.

Table 17 Required Columns in an Attribute File

Column	Requirements
For one-channel arrays	
Array ID	Unique for each sample
Global Display Name	By default, this is the file name. Must be unique for each microarray and cannot be blank.
Green Sample	Green sample ID
For two-channel arrays	
Array ID	Unique for each sample
Global Display Name	By default, this is the file name. Must be unique for each microarray and cannot be blank.
Green Sample	Green Sample ID
Red Sample	Red Sample ID
Polarity	Value of 1 or -1 only. Blank value is set to "1". Values other than 1, -1 or blank will cause the row to be ignored and not imported. If a polarity is assigned as -1, then for later analysis, it is processed as a dye-flip array. For more information on dye-flipped arrays, see "Import FE Files" on page 23.

Table 18 Exceptions and Conditions for Importing Attribute Files

Condition	Outcome
The attribute file being imported has more than 1 entry for the same Array ID.	Error message appears. User is asked whether the last entry should overwrite the previous entry.
The attribute file to be imported has rows with missing required column values.	The row is skipped and an error message appears that shows which required columns are missing.
The attribute file to be imported has rows with values that do not match the data type for the particular column.	Those rows are skipped and a warning message appears.
The attribute file to be imported has rows with Array IDs already present in the table.	Error message appears "Some Array ID's are already present - Do you want to over-write these rows?" Click **Yes** to overwrite the existing rows with the duplicate rows from the attribute file. Click **No** to cancel the import.
The attribute file to be imported has rows with ArrayID already present in the table as FE data (gray row).	Polarity and isMultiPack value will not be overwritten.

Attributes

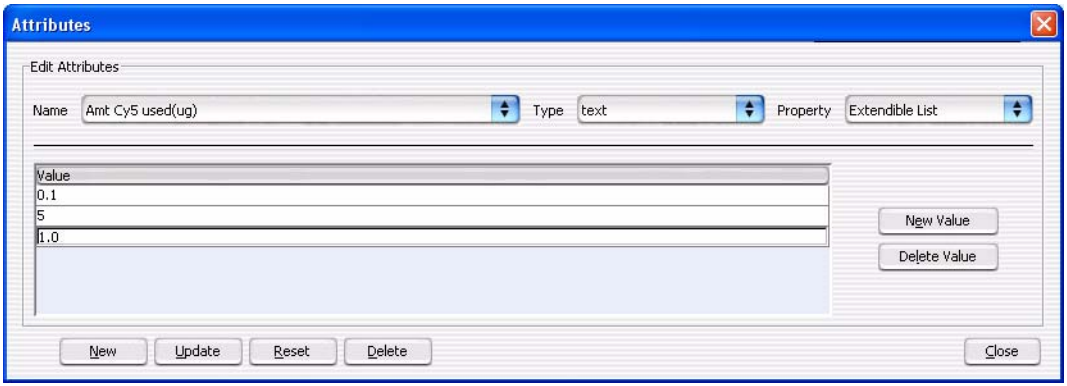


Figure 78 Attributes dialog box

Purpose: Use this dialog box to create, change, or delete available sample attributes. Attributes are pieces of sample-specific information, such as hybridization temperature or sample name.

To open: From the Sample Manager table, click **Add/Edit Attribute**.

NOTE When you add or change an attribute from this dialog box, that attribute/value is then available for you to associate with a microarray.

Name The name of the attribute whose details appear in the dialog box. Click the arrow to select the desired attribute from the list.

Two lists appear to the right of the attribute name. These lists define the type of value the program will use for the attribute. You select one value from each list.

The Type list has these options:

Option	The attribute is:
int	an integer
double	a double-precision floating point number (a high-precision number that can include decimals)

Option	The attribute is:
boolean	true or false
text	a string of plain text

NOTE If you select an option (other than "text"), and type a value that does not match the selected type, you will get the following error: "Invalid Value for column".

The Property list has these options:

Option	Description
Extendible List	This lets you define a choice of values that are displayed when the attribute is selected. It also displays a text box where you can type a custom value.
Single valued	The program displays a single text box to edit this attribute for a sample.
Dynamic valued	The program does not allow you to edit the value of this attribute for a sample. Instead, it defines the value of the attribute, based on information in the array data file.
Pre-defined list	This lets you define a choice of values that are displayed when the attribute is selected.

Value For extendible and predefined lists, the program displays these values as choices when you edit the selected attribute for a sample. Double-click a value to change it. Click **Update** to accept the edited value. Click **Reset** to return the values to their previous state.

New Value For extendible and predefined lists, adds another value to the Value list. Click **New Value** to make it available for editing. Click **Update** to accept the edited value.

Delete Value For extendible and predefined lists, deletes a value from the Value list. Click the value you want to delete, then click **Delete Value**.

New Use this command to create a new attribute. It opens an Input dialog box, where you type a name for the new attribute.

Update (Enabled only if there are unsaved changes to the attribute.) Saves the changes you made to the attribute without closing the Attributes dialog box.

Reset (Enabled only if there are unsaved changes to the attribute.) Restores the attribute it to its state before you made any changes.

Delete Deletes the attribute from the program.

CAUTION If you delete an attribute, the program also deletes all values assigned to it for all samples. To restore the attribute, you must create the attribute again, and reassign the desired value to each array.

Close Closes the Attributes dialog box. If you made changes to the attribute, but did not save them, a dialog box asks if you want to save the changes to the attribute. Click **Yes** to save the changes, or click **No** to close the dialog box without saving the changes, or click **Cancel** to go back to the Attributes dialog box without saving the changes.

Auto Suppress Dialog

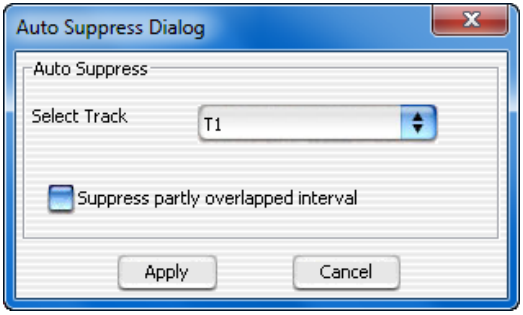


Figure 79 Auto Suppress dialog box

Purpose: To automatically suppress aberrations within a selected track.

To open: In Triage View, click **Auto Suppress**. See "Triage View" on page 59.

Select Track Click the arrow and select an available track. If no tracks are listed, you must create one before you can use this feature. See "Create Track" on page 168.

Suppress partly overlapped interval Select this box to suppress aberrations that occur in intervals that partly overlap each other.

Chart Configuration Dialog Box

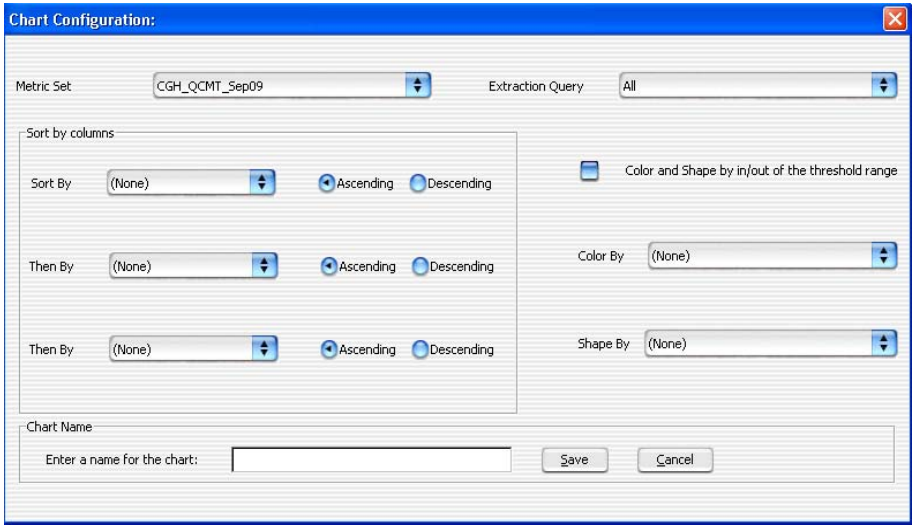


Figure 80 Chart Configuration dialog box

Purpose: To configure the appearance of a Quality chart.

To open: On the Quality ribbon, click **Chart** and then select **New**. See "Quality Tab Window" on page 87.

The Chart Configuration dialog box has the following components and functionality:

Metric Set Select the metric set to be used with the chart.

Extraction Query Select the extraction query that is to be evaluated to produce a chart.

Sort by Columns Select how you want to order the extractions in the chart. You can select three conditions, and select whether to sort them in Ascending or Descending order.

Color and shape by in/out of the threshold range When this check box is selected, the color and shape of all points are set by whether the value is in or out of range. The Color By and Shape By settings are ignored.

Color By Indicates whether to color-code extraction data points depending on whether they fall inside or outside the threshold level. Points that are outside the limits are color-coded in red and the ones within the limits are color-coded in blue. All the points are connected by a *light-gray* line. The ShapeBy and ColorBy columns are disabled if this check box is selected.

Shape By Indicates what groups are used to differentiate the data points by shape. For example, selecting Username causes all the extractions that are from a particular user to have the same shape.

Chart Name Lets you enter a name to be associated with the chart.

Save Saves the chart using the name you entered.

Cancel Closes the Chart Configuration dialog box without saving any changes.

Chart Properties

Purpose: This dialog box is used to create or edit settings for plots.

To open: This dialog box opens when you right-click within the line plot in the Graphical Differential Aberration Summary dialog box. See "Graphical Differential Aberration Summary" on page 198.

This dialog box has three tabs. At any point, click **OK** to accept the settings in all three tabs, or click **Cancel** to close the dialog box without making any changes to the settings.

Title tab

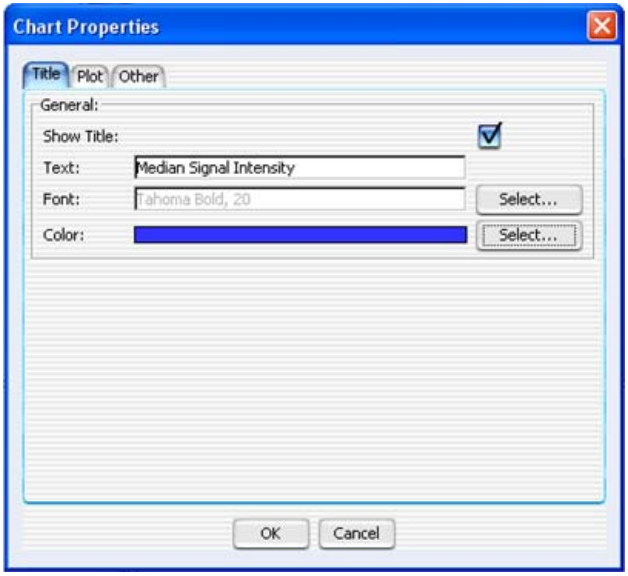


Figure 81 Chart Properties dialog box – Title tab

- **Show Title** – Select this option to display a title across the top of the chart.
- **Text** – Type a title for the chart.
- **Font** – (Available if you select **Show Title**) Click **Select** to open the Font Selection dialog box. Select the desired font attributes, then click **OK.**
- **Color** – (Available if you select **Show Title**) Click **Select** to open the Title Color dialog box. Select or configure a color for the title, then click **OK.** This dialog box is identical to the Select Color dialog box. See "Select Color" on page 237.

Plot tab

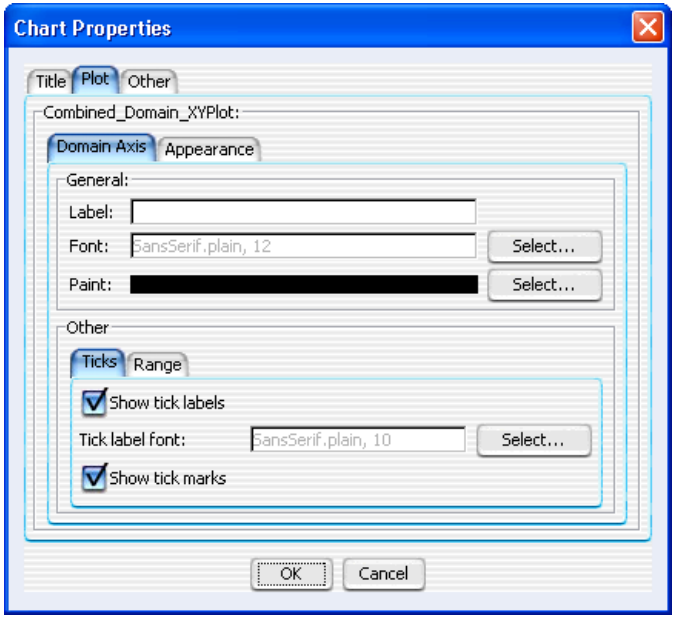


Figure 82 Chart Properties dialog box – Plot tab

- In the Plot tab, you can set these properties in the Domain Axis tab ("X" axis):

Property	Details
General	
Label	A custom label for the Domain (X) axis of the chart. Type the desired label.
Font	The font for the custom label on the Domain (X) axis. Click **Select** to open the Font Selection dialog box. Select the desired font attributes, then click **OK.**
Paint	The color of the custom label on the Domain (X) axis. Click **Select** to open the Label Color dialog box. Select the desired color, then click **OK.** This dialog box is identical to the Select Color dialog box. See "Select Color" on page 237.

Property	Details
Other – Ticks tab	
Show tick labels	Select this option to show, or clear it to hide, the numerical values on the Domain (X) axis.
Tick label font	The font for the numerical values on the Domain (X) axis. Click **Select** to open the Font Selection dialog box. Select the desired font attributes, then click **OK.**
Show tick marks	Select this option to show, or clear it to hide, tick marks on the Domain (X) axis.
Other – Range tab	
Auto-adjust range	Select this option to automatically set the range of values on the X-axis to include all data.
Minimum range value	(Available if you do not select Auto-adjust range) The lowest value represented on the X-axis.
Maximum range value	(Available if you do not select Auto-adjust range) The highest value represented on the X-axis. The program automatically converts large numbers to scientific "E" notation – for example, **1.22E8.**

- In the Plot tab, you can set the following properties in the Appearance tab:

Property	Details
Outline stroke	The thickness of the lines that enclose each plot. Click **Select** to open the Stroke Selection dialog box. Select the desired line thickness, then click **OK**.
Outline paint	The color of the lines that enclose each plot. Click **Select** to open the Outline Color dialog box. Select the desired color, then click **OK.** This dialog box is identical to the Select Color dialog box. See "Select Color" on page 237.
Background paint	The color of the background within each plot area. Click **Select** to open the Background Color dialog box. Select the desired color, then click **OK.** This dialog box is identical to the Select Color dialog box. See "Select Color" on page 237.
Orientation	Select either Vertical (domain-axis on the bottom of the chart) or Horizontal (domain-axis on the left side of the chart).

Other tab

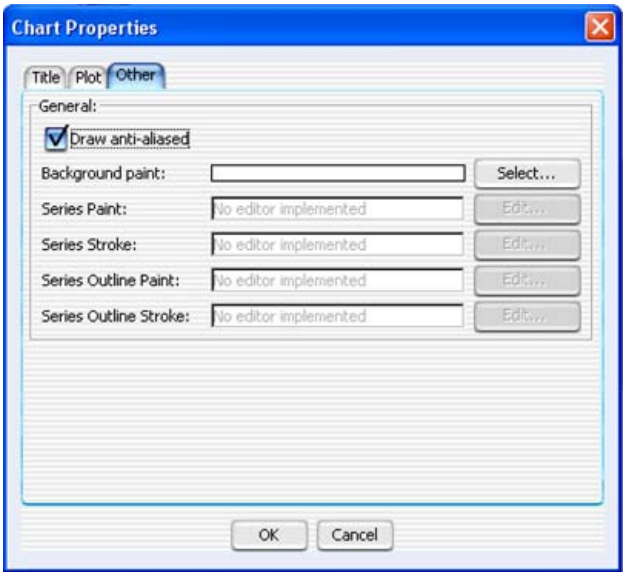


Figure 83 Chart Properties dialog box – Other tab

The Other tab offers these options:

- **Draw anti-aliased** – Select this option to minimize distortion and visual artifacts in the plot image. This produces a smoother image, but it can be less sharp than the original one.
- **Background paint** – The color of the chart outside of the plot area and legend. Click **Select** to open the Background Color dialog box. Select the desired color, then click **OK.** This dialog box is identical to the Select Color dialog box. See "Select Color" on page 237.

The other options are for future expansion, and are not available in the current release of Agilent CytoGenomics 1.0.

Configure Coloring Ranges and Shades

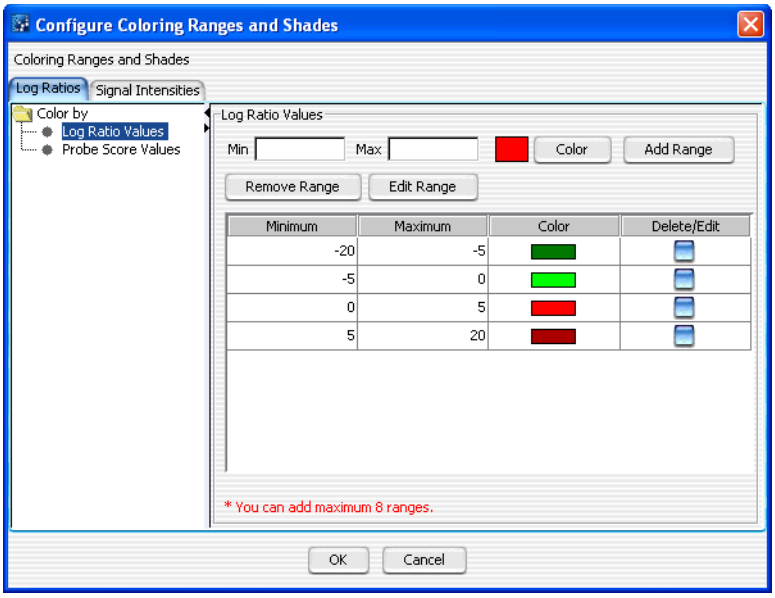


Figure 84 Configure Coloring Ranges and Shades dialog box

Purpose: This dialog box is used to enter ranges and select colors for scatter plot options. Tabs show selections for Log Ratios and Signal Intensities plots.

To open: In Gene View, move the mouse pointer over **Scatter Plot** to display the scatter plot options and then click **Configure Color and Ranges**. Or, right-click the mouse in any of the Genomic Viewer Views, and click **View Preferences**. Then, under Configure Coloring schemes, click **Configure Colors and Ranges.** See "Genomic Viewer" on page 40.

Table 19 Log Ratios

Color by	Description
Log Ratio Values	
Min	Type a minimum value for the range.
Max	Type a maximum value for the range.
Color	Click to open the Select Color dialog box, where you can select the color you want to display for this range. See "Select Color" on page 237 for more information.
Add Range	Click to add a row to the range table, using the values displayed in Min and Max, and the selected Color.
Remove Range	Click to remove the ranges with Edit/Delete box selected.
Edit Range	Click to edit range(s) with Edit/Delete box selected.
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.
Probe Score Values	
Min	Type a minimum value for the range.
Max	Type a maximum value for the range.
Color	Click to open the Select Color dialog box, where you can select the color you want to display for this range. See "Select Color" on page 237 for more information.
Add Range	Click to add a row to the range table, using the values displayed in Min and Max, and the selected Color.
Remove Range	Click to remove the ranges with Edit/Delete box selected.
Edit Range	Click to edit range(s) with Edit/Delete box selected.
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.

Table 20 Signal Intensities

Color by	Description
Channels	
Green Intensity	Click to open the Select Color dialog box, where you can select the color you want to display for this channel. See "Select Color" on page 237 for more information.
Red Intensity	Click to open the Select Color dialog box, where you can select the color you want to display for this channel. See "Select Color" on page 237 for more information.
Probe Score Values	
Min	Type a minimum value for the range.
Max	Type a maximum value for the range.
Color	Click to open the Select Color dialog box, where you can select the color you want to display for this range. See "Select Color" on page 237 for more information.
Add Range	Click to add a row to the range table, using the values displayed in Min and Max, and the selected Color.
Remove Range	Click to remove the ranges with Edit/Delete box selected.
Edit Range	Click to edit range(s) with Edit/Delete box selected.
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.
Intensity Values	
Min	Type a minimum value for the range.
Max	Type a maximum value for the range.
Color	Click to open the Select Color dialog box, where you can select the color you want to display for this range. See "Select Color" on page 237 for more information.
Add Range	Click to add a row to the range table, using the values displayed in Min and Max, and the selected Color.
Remove Range	Click to remove the ranges with Edit/Delete box selected.

Table 20 Signal Intensities

Color by	Description
Edit Range	Click to edit range(s) with Edit/Delete box selected.
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.

Create a new Metric

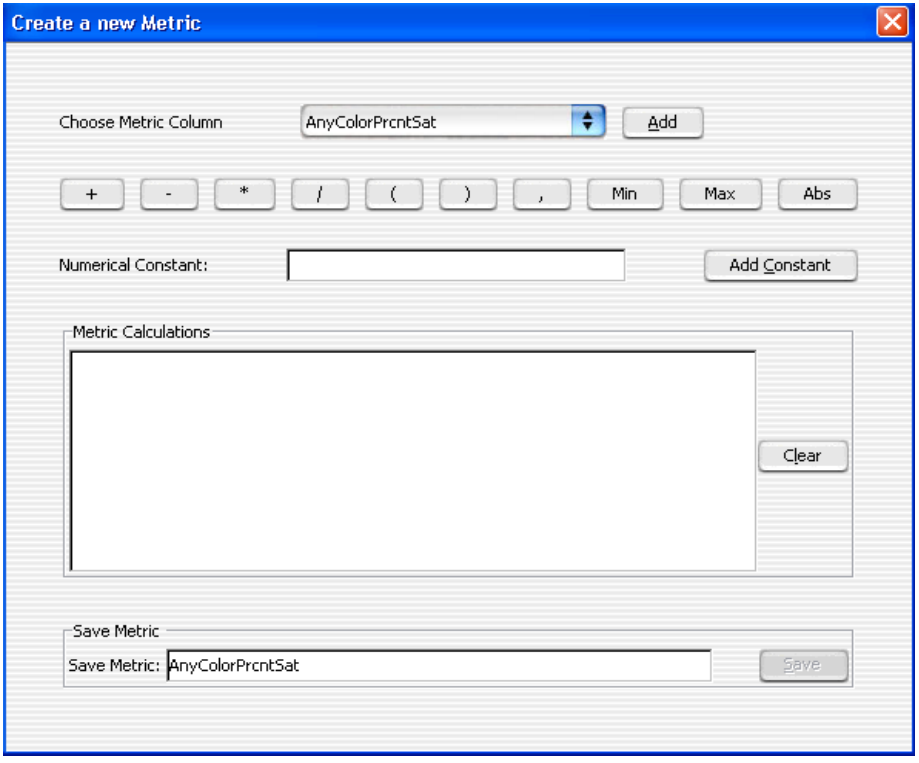


Figure 85 Create a new Metric dialog box

Purpose: Lets you create a metric in the Quality tab to track selected statistical values for a set of extractions.

To open: On the Quality ribbon, click Metrics, and then select New. See "Quality Tab Window" on page 87.

The Create a new Metric dialog box has the following components and functionality:

Choose Metric Column A list of metrics that can be used to create a calculation as a new metric.

Add Selects the chosen metric and adds it to the Metric Calculations text area for review.

Operations The following mathematical operators are available for any metric or collection of metrics that are selected using the Choose Metric Column.

- Precedence of operations is left to right, except when interrupted by parentheses.
- Operations proceed left to right inside any set of parentheses, and inside out in terms of stacked parentheses.

+ Adds any two metrics or collection of metrics grouped by matched parentheses.

- Subtracts any two metrics or collection of metrics grouped by matched parentheses.

***** Multiplies any two metrics or collection of metrics grouped by matched parentheses.

/ Divides any two metrics or collection of metrics grouped by matched parentheses.

( ) Let you subset and prioritize the mathematical operations.

, Lets you list any two metrics or collection of metrics grouped by parenthetical operators for the evaluative operations listed below.

Min Returns the smallest value from a list of metrics (or collection of metrics grouped by matching parenthesis). The list elements are separated by the ',' operator.

Max Returns the largest value from a list of metrics (or collection of metrics grouped by matching parenthesis). The list elements are separated by the ',' operator.

Abs Returns the absolute value of a metric or a collection of metrics grouped by matching parenthesis.

Numerical Constant Lets you enter a value to be added to the metric calculation formula.

Add Constant button Accepts the value entered in the Numerical Constant field and adds it to the metric calculation formula.

Metric Calculations

The area in which the metric calculation formula is displayed for review.

Clear Removes all metrics, mathematical operators, and constants from the formula in the Metric Calculations area.

Save Metric Lets you type a name to be associated with the metric.

Save Saves the metric using the name you entered.

Create Analysis Method

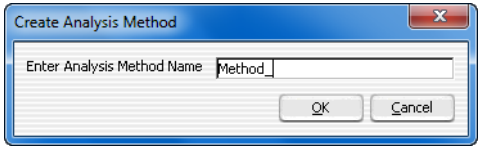


Figure 86 Create Analysis Method dialog box

Purpose: Used to create and name an analysis method.

To open: In the Config tab, under Workflow, click **Analysis Method**. Then at the bottom of the window, click **Create Analysis Method**.

Enter Analysis Method Name Type the name for the analysis method you are creating.

OK Click this button to create the new analysis method with the designated name.

Cancel Click to cancel the operation.

Create Cyto Report Template

Purpose: To create or change a cyto report template that can be used to generate a report at the end of a workflow analysis.

To open: In the Config tab, under Template, click **Report Template**. At the bottom of the window, click **Create New Template**. Or, to change an existing template, under Action, click **View/Edit** for a report template in the list.

The Create Cyto Report Template dialog has four Steps, described below. The following commands appear at the bottom of all step dialogs, and are used to move forward and backward through the steps.

Back Click to move to the previous step.

Next Click to move to the next step.

Cancel Click to cancel the changes and exit the dialog box.

Create Cyto Report Template Step 1 of 4

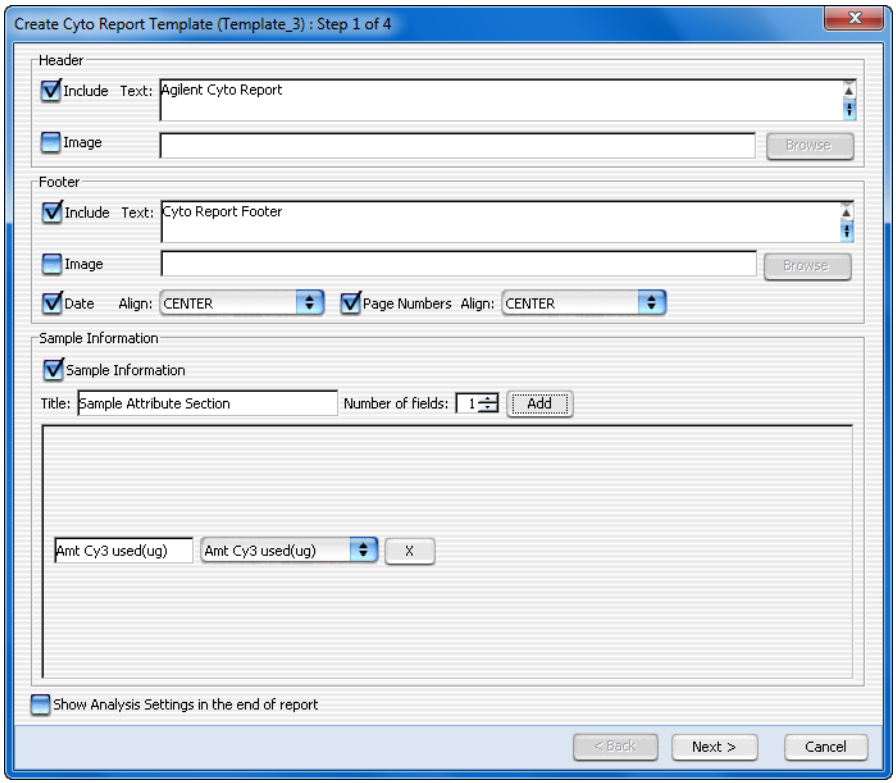


Figure 87 Create Cyto Report Template (template name): Step 1 of 4 dialog box

In this step, you can select and configure the following items for the report:

Item	Details
Header	
Include	Select to include a header in the report. You must select this to include any of the header items in the report.
Text	Type the text that appears in the header of the report.

Item	Details
Image	Select to include an image in the header of the report. (This can be a company logo, for example.)
Browse	Click to browse to the location and select an image file to include in the report. Make sure the image is in a location where users have read permissions.
Footer	
Include	Select to include a footer in the report. You must select this to include any of the footer items in the report.
Text	Type the text that appears in the footer of the report.
Image	Select to include an image in the footer of the report. (This can be a company logo, for example.)
Browse	Click to browse to the location and select an image file to include in the report. Make sure the image is in a location where users have read permissions.
Date	Select to include the date in the footer of the report.
Align	Click the arrow and select the alignment of the date in the footer.
Page Numbers	Select to include page numbers in the footer of the report.
Align	Click the arrow and select the alignment of the date in the footer.
Sample Information	
Sample Information	Select to include sample information in the report.
Title	Type the title to appear above the sample information. You must include a title.
Number of fields	Click and select the number of sample information fields to include in the report.

Item	Details
Add	Click to add the selected number of fields to the report. The number of fields selected are added to the box below. For each field, the name for the field is shown. Select the sample information from the list. You can change the text title for the field, if desired. Select X to remove a field from the report.
Show Analysis Settings in the end of report	Select this to list the analysis settings used when the report was generated to the end of the report.

Create Cyto Report Template Step 2 of 4

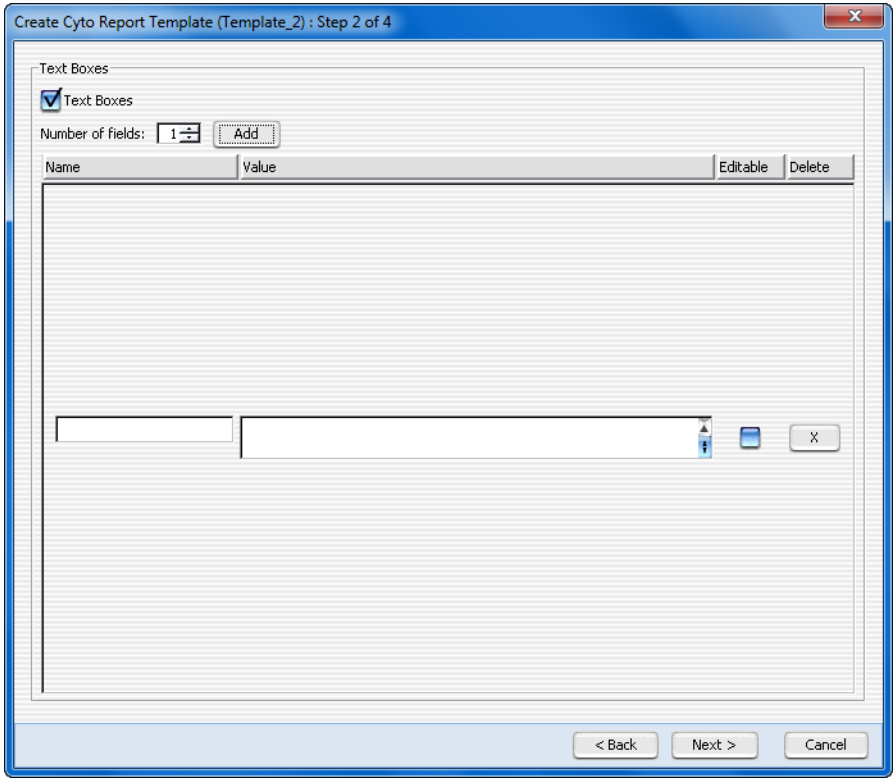


Figure 88 Create Cyto Report Template (template name): Step 2 of 4 dialog box

In this step, you can select and configure the following items for the report:

Item	Details
Text Boxes	
Text Boxes	Select to include text boxes in the report.
Number of fields	Click the arrow and select the number of text fields you want to include in the report. Text fields are used to let you include customizable information in the report.
Add	Click to add the selected number of text boxes. The selected number of text boxes are added to the table.
Name	Type a name for the text box.
Value	Type the text to include for the field.
Editable	Select to make the text box customizable. Text in editable text boxes can be changed in the Manage Cyto Report dialog box when you use the View Report function in the Report window of the Analysis tab. See "Manage Cyto Report" on page 213.
Delete	Click to delete the selected text box from the report.

Create Cyto Report Template Step 3 of 4

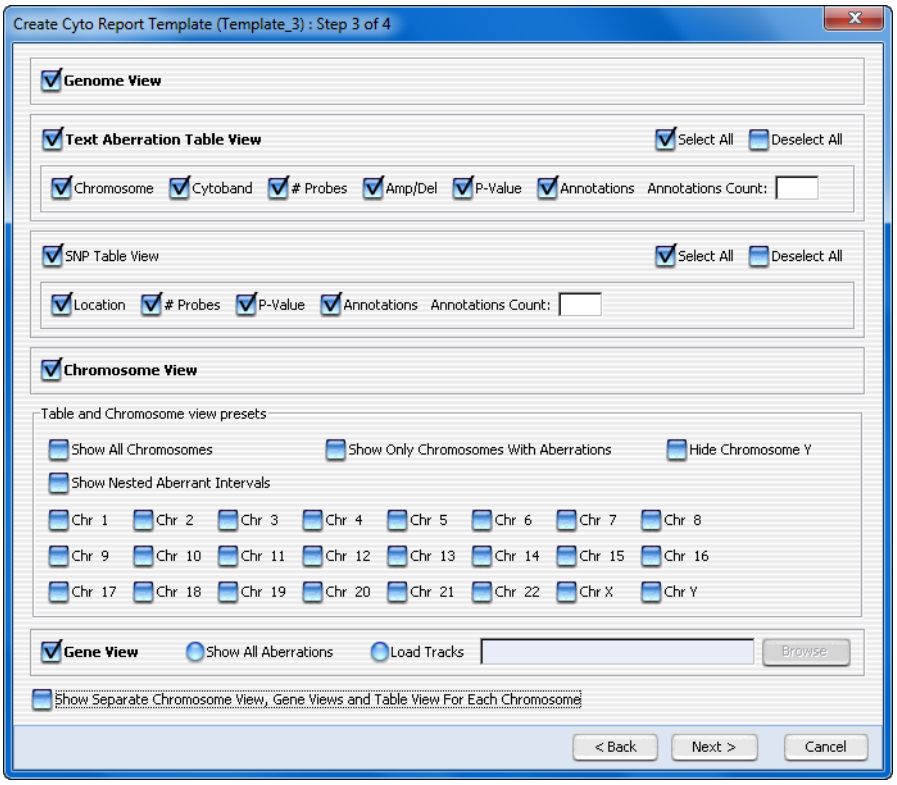


Figure 89 Create Cyto Report Template (template name) Step 3 of 4 dialog box

In this step, you can select and configure the following items for the report.

Item	Details
Genome View	Select to include a graphical representation of results across the genome in the report. See "Genome View" on page 41.
Text Aberration Table View	Select to include a table of aberration results in the report.
Select All	Select to select all columns for the table.

Item	Details
Deselect All	Select to clear all columns for the table.
Table view columns	• Select one or more columns to include in the table: Chromosome, Cytoband, #Probes, Amp/Del, P-Value, Annotations. • If you select Annotations, you must type a count number for the number of annotations to show.
Annotations Count	Type a number that controls the number of genes displayed in the Annotations column of the cyto report. To display all genes, set this value to a high number (for example, 10000.)
SNP Table View	Select to include a table of SNP results in the report.
Select All	Select to select all columns for the table.
Deselect All	Select to clear all columns for the table.
Table view columns	• Select one or more SNP columns to include in the table: Location, #Probes, P-Value, Annotations.
Annotations Count	Type a number that controls the number of genes displayed in the Annotations column of the cyto report. To display all genes, set this value to a high number (for example, 10000.)
Chromosome View	Select to show a graphical depiction of results along selected chromosomes in the report. See "Chromosome View" on page 43.
Table and Chromosome view presets	
Show All Chromosomes	Select to include all chromosomes in the Table View and Chromosome View of the report.
Show Only Chromosomes With Aberrations	Select to include only chromosomes with aberrations in the Table View and Chromosome View of the report.
Hide Chromosome Y	Select to never include chromosome Y in the Table View and Chromosome View of the report.

Item	Details
Show Nested Aberrant Intervals	Select to show nested aberrant intervals in the Table View and Chromosome View of the report. • ADM-1 and ADM- 2 scores iteratively identify all aberrations that differ significantly from log ratios that show no change between sample channels. In the iterations, the algorithms identify nested, or "child", aberrations that are contained within other "parent" aberrations but show a significant difference from the parent aberration log ratio.
Chromosome list	Select the individual chromosomes you want to include in the Table View and Chromosome View of the report. (Not available if Show Only chromosomes With Aberrations is selected.)
Gene View	Select to include the Gene View in the report. See "Gene View" on page 45.
Show All Aberrations	Select to show all aberrations in the Gene View of the report.
Load Tracks	Select to show only selected tracks in the Gene View of the report.
Browse	Click to browse to a location and select tracks to include in the report. See "Create Track" on page 168,
Show Separate Chromosome View, Gene View, and Table View for each chromosome	This is enabled if you selected Text Aberration Table View, Chromosome View, and Gene View.

Create Cyto Report Template Step 3 of 4

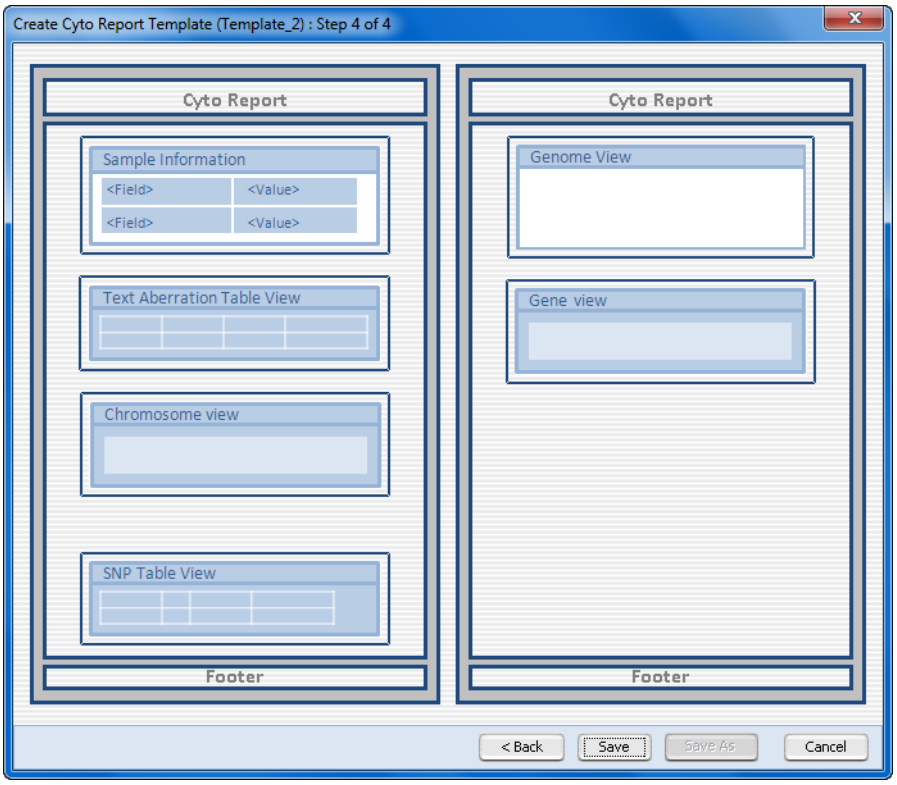


Figure 90 Create Cyto Report Template (template name) Step 3 of 4 dialog box

In this step, you see a graphical depiction of the report template you created in Steps 1 - 3. You can change the order in which the report sections appear. To move a report section, point to the section with your mouse, click and hold the mouse button while you drag the section to the location where you want it to appear on the report, then let go of the mouse button.

Save Click to save the report template.

Save As Click to save the report template and give it a new name.

Create Signal Bar Chart

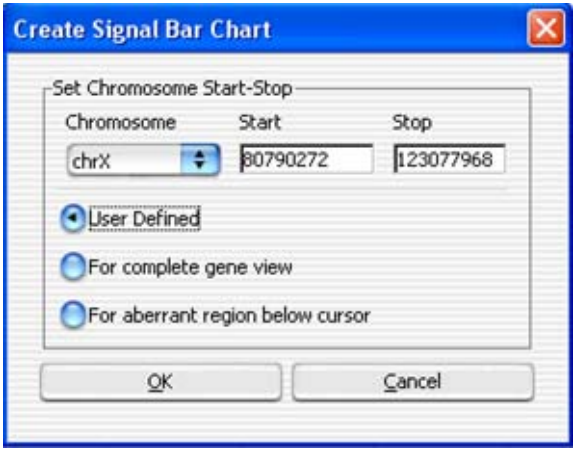


Figure 91 Create Signal Bar Chart dialog box

Purpose: This dialog box lets you set parameters to create a histogram of signal intensities. You can customize the region you want to display by selections in Set Chromosome Start-Stop.

To open: Right-click in the Gene View and select **Show Intensity Bar Charts**.

Set Chromosome Start-Stop Defines the region of the chromosome for which the bar chart will be defined. Select one of these options:

- **User Defined** – Lets you define an arbitrary region of any chromosome. If you select this option, select the desired chromosome in **Chromosome,** then type the beginning **(Start)** and end **(Stop)** locations of the desired interval.
- **For complete gene view** – The chromosomal region that appears in Gene View.
- **For aberrant region below cursor** – All of the intervals that begin before the cursor position and end after the cursor position. (This option is available if an aberration algorithm is selected.)

OK Creates the histogram using the selected region.

Cancel Closes the dialog box without creating the histogram.

Create Track

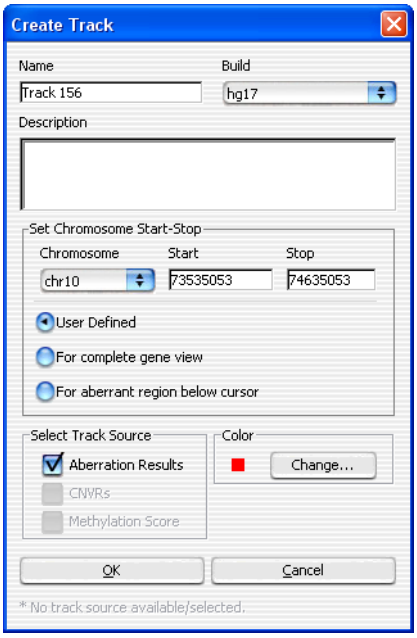


Figure 92 Create Track dialog box

Purpose: This dialog box lets you create a track for a chromosomal region, based on certain results or other parameters. You can display one or more tracks next to the genes, data, and results in Gene View. For information on setting how tracks are displayed, see "User Preferences" on page 247.

To open: Right-click in the plot area of Gene View for the CGH module, then click **Create Track** in the shortcut menu. See "Genomic Viewer" on page 40.

Name Type a name for the track. This name identifies the track when it appears in lists and displays.

Build (Available if you select **User Defined** in **Set Chromosome Start-Stop.**) Select the genome build with which to associate the track.

Description Type descriptive text to attach to the track for reference.

Set Chromosome Start-Stop This parameter defines the region of the chromosome for which the track will be defined. Select one of these options:

- **User Defined** – Lets you define an arbitrary region of any chromosome. If you select this option, select the desired chromosome in **Chromosome,** then type the beginning **(Start)** and end **(Stop)** locations of the desired interval.
- **For complete gene view** – The chromosomal region that appears in Gene View.
- **For aberrant region below cursor** – All of the intervals that begin before the cursor position and end after the cursor position.

Select Track Source The type of analysis result the program uses to construct the regions defined in the track. Select one or both of these options:

- **Aberration Results** – The current aberration analysis results in the defined chromosomal region.
- **CNVRs** – The detected CNVRs in the defined chromosomal region.

Change Click to open the Choose Track Color dialog box to select the color to use for display of the track in the Tracks folder. See "Select Color" on page 237.

OK Creates the track. To configure how tracks are displayed in Gene View, see "User Preferences" on page 247.

Cancel Closes the dialog box without creating a track.

Create Workflow

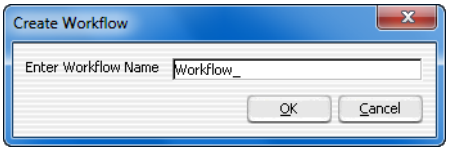


Figure 93 Create Workflow dialog box

Purpose: To create and name a new workflow.

To open: In the Config tab, under Workflow, click **Workflow** and then click **Create Workflow**.

Enter Workflow Name Type the name of the new workflow here.

Customize Search Link

Figure 94 Customize Search Link dialog box

Purpose: This dialog box lets you create a custom Web search link in the shortcut menu that appears when you right-click a table entry.

To open: Right-click any entry in a tab in Tab View, other than a column heading, then click **Customize Link.**

URL Name The name of the custom Web search link that appears in the shortcut menu (see above). To edit an existing custom Web search link, select it from the list.

URL The full uniform resource locator (URL) of the desired search page. For the query string value, type `<target>`

For example, this URL passes the selected Tab View entry to google.com:

```
http://www.google.com/search?hl=eng&q=<target>
```

New Opens an Input dialog box, where you can type a name for a new custom Web search link. Click **OK** to accept the name and add it to the URL name list.

Update Saves the settings in the dialog box.

Delete Deletes the selected custom Web search link.

Close Closes the dialog box.

Define Sets

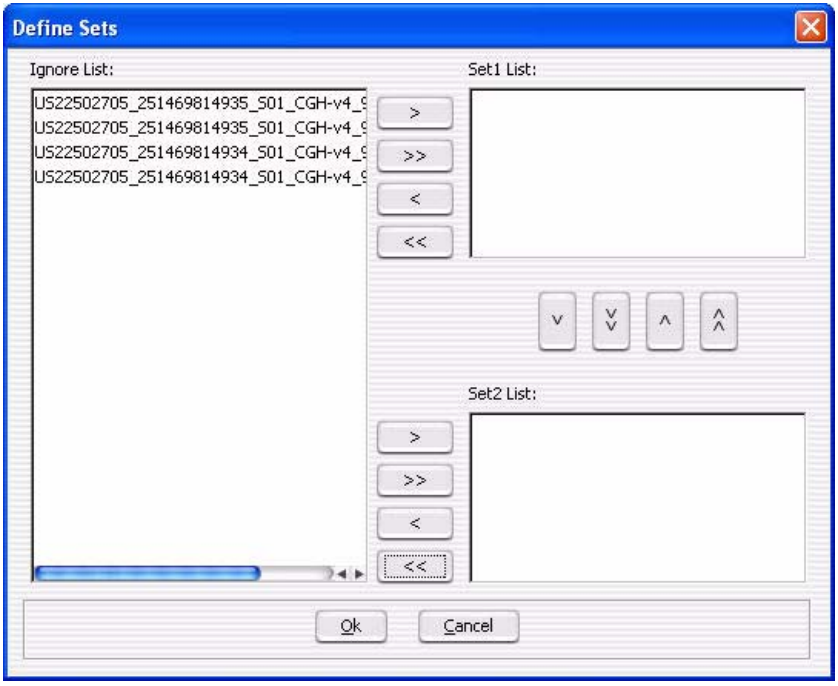


Figure 95 Define Sets dialog box

Purpose: This dialog box lets you assign arrays to comparison sets for a differential aberration analysis. See "Differential Aberration Setup" on page 178.

To open: The Define Sets dialog box opens when you click **Define Sets** in the Differential Aberration Setup dialog box as you set up a differential aberration analysis within the View Aberrations window. See "View Aberrations Window" on page 54.

The lists in this dialog box reflect the assignments you make with the Set1/Set2/Ignore options in the Differential Aberration Setup dialog box. Similarly, the selected Set1/Set2/Ignore options in the Differential Aberration Setup dialog box reflect the assignments you make in this Define Sets dialog box.

Set1 List Displays the arrays assigned to comparison set 1.

Set2 List Displays the arrays assigned to comparison set 2.

Ignore List Displays the arrays that are not assigned to a comparison set.

Moves selected arrays from one list to another, in the direction indicated by the arrowhead. To select arrays:

- To select an array, click its name.
- To select additional arrays, hold down the **ctrl** key and click their names.
- To select a contiguous block of arrays, click the name of the first array, then hold down the **shift** key and click the name of the last one.

Moves all of the arrays in one list to another, in the direction indicated by the arrowheads.

Ok Accepts your array assignments. These assignments will be reflected in the settings of the Set1/Set1/Ignore options for each array in the Differential Aberration Setup dialog box.

Cancel Discards any changes to array assignments, and closes the dialog box.

Design Properties

Purpose: Gives general and detailed information about a given microarray design.

To open: In the Content tab, in the Sample Manager Data Navigator, right-click a design build, then click **Show Properties.** Several tabs are available.

Attribute tab

Displays general identifying attributes of the array design, and statistics such as the total number of features in the design.

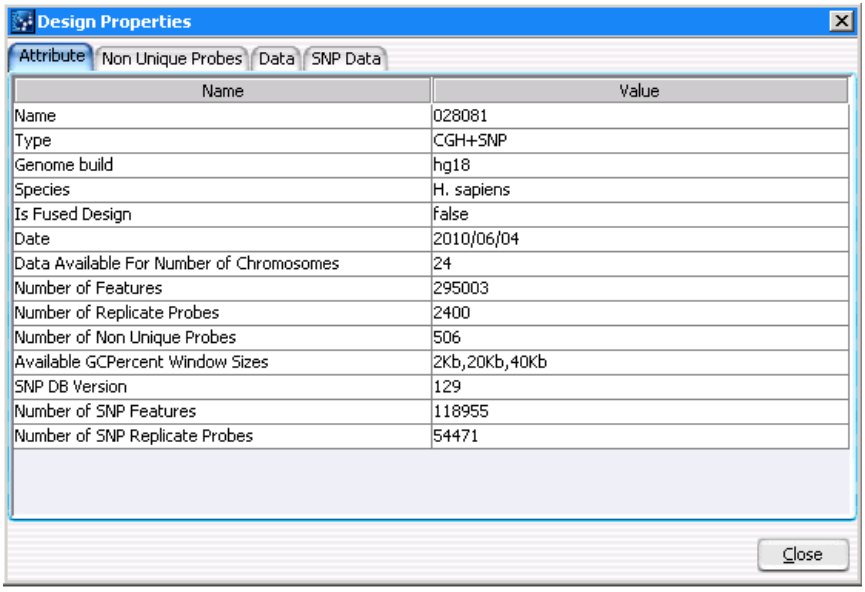


Figure 96 Design Properties dialog box – Attribute tab

Non Unique Probes tab

Displays the nonunique probes in the design. Nonunique probes have more than one mapping in the genome that is a perfect match.

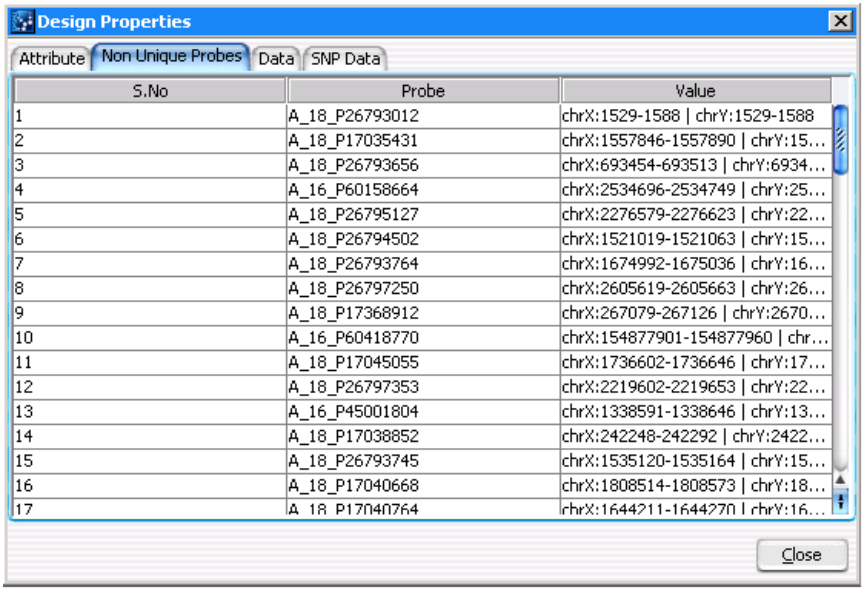


Figure 97 Design Properties dialog box – Non Unique Probes tab

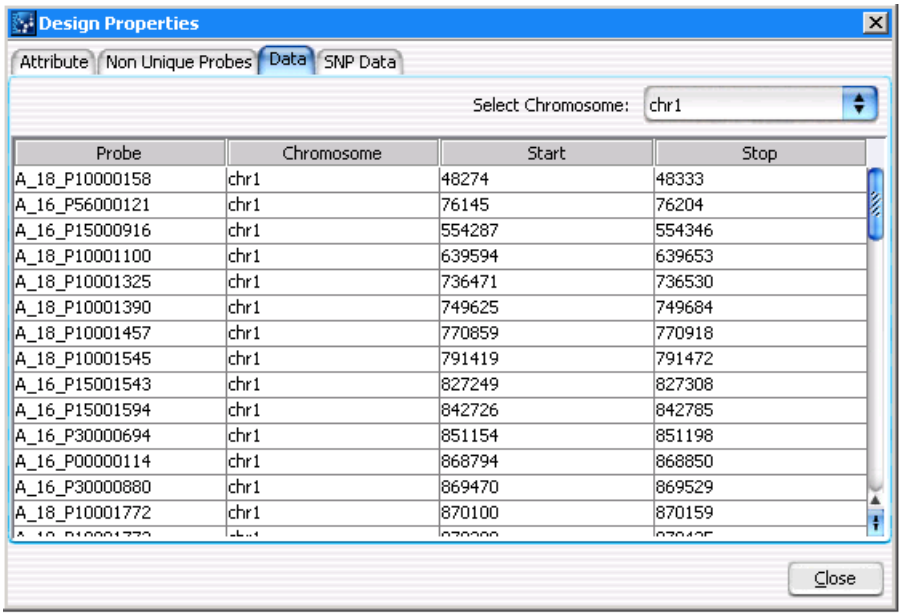
S. No The sequence order of the probes within the tab.

Probe The name of the each nonunique probe.

Value The chromosomal locations to which each of the probes maps. Because these are nonunique probes, two or more locations appear for each probe.

Data tab

Displays the names of the probes in the design and the genomic locations to which they are designed. The tab displays the probes for one chromosome at a time.



Probe	Chromosome	Start	Stop
A_18_P10000158	chr1	48274	48333
A_16_P56000121	chr1	76145	76204
A_16_P15000916	chr1	554287	554346
A_18_P10001100	chr1	639594	639653
A_18_P10001325	chr1	736471	736530
A_18_P10001390	chr1	749625	749684
A_18_P10001457	chr1	770859	770918
A_18_P10001545	chr1	791419	791472
A_16_P15001543	chr1	827249	827308
A_16_P15001594	chr1	842726	842785
A_16_P30000694	chr1	851154	851198
A_16_P00000114	chr1	868794	868850
A_16_P30000880	chr1	869470	869529
A_18_P10001772	chr1	870100	870159

Figure 98 Design Properties dialog box – Data tab

Select Chromosome The chromosome whose probes appear in the list. To display the probes designed to a different chromosome, select one from this list.

Probe The name (Probe ID) of each probe.

Chromosome The name of the chromosome to which the probe is designed.

Start The location on the selected chromosome of the first base pair to which each probe is designed.

Stop The location on the selected chromosome of the last base pair to which each probe is designed.

SNP Data tab

This tab shows design information for SNPs in the design.

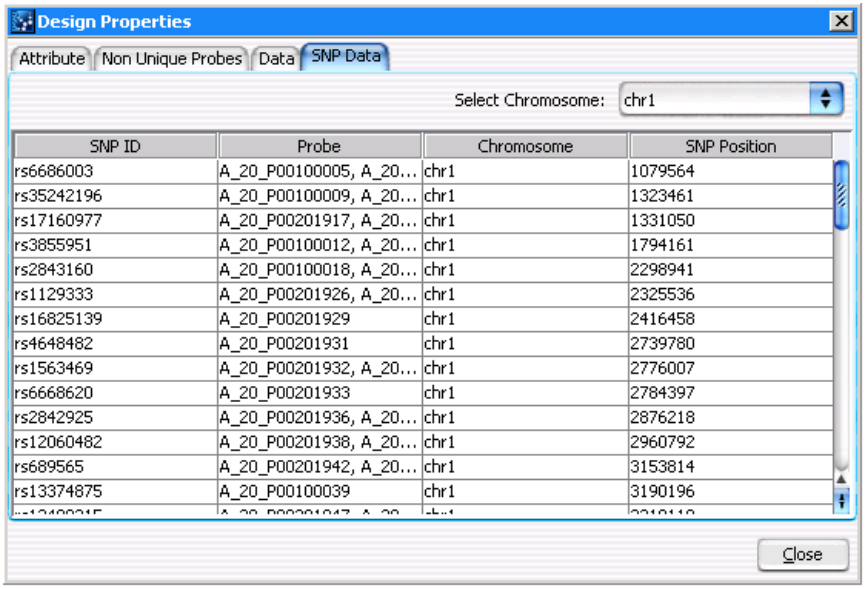


Figure 99 Design Properties dialog box – SNP Data tab

SNP ID The SNP identification.

Probe The name (Probe ID) of the probe. The probe names are separated with a comma.

Chromosome The chromosome on which the probe is located.

SNP Position The position of the SNP on the chromosome.

Differential Aberration Setup

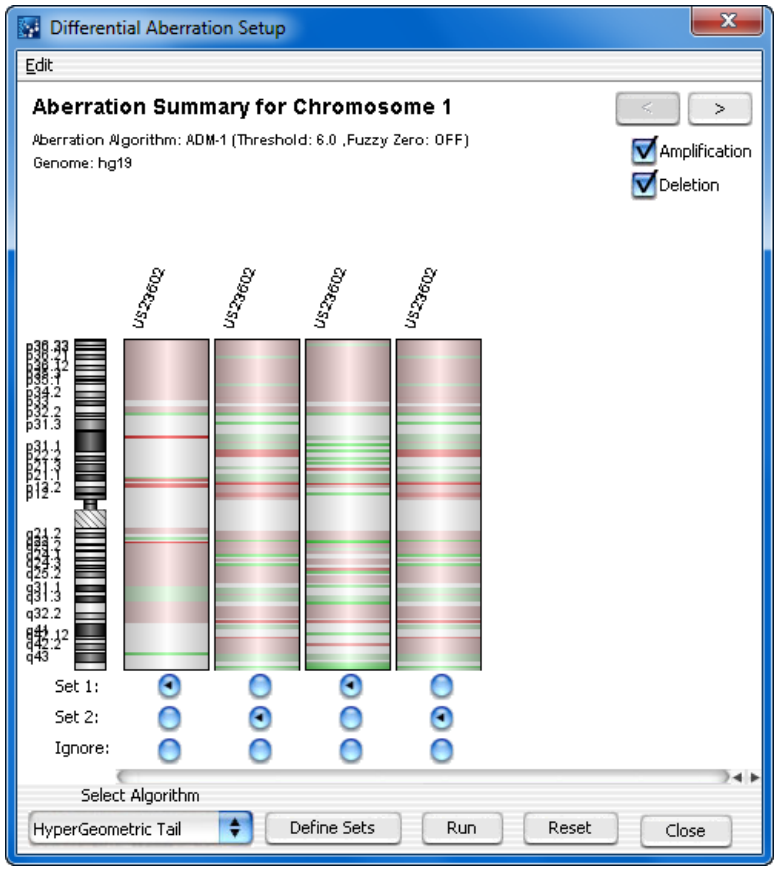


Figure 100 Graphical Aberration Summary dialog box – Special version used to set up a differential aberration analysis

Purpose: This dialog box lets you configure and start a differential aberration analysis.

To open: In the Analysis tab, open more than one sample in the View Aberrations tab. The samples must contain results of an aberration detection algorithm. In the View Aberrations command ribbon, under Aberration, click **Differential**. See "View Aberrations Window" on page 54.

Edit Opens a menu with a **Copy summary to clipboard** option. This option lets you move an image of the dialog box to a document in another program.

Displays aberrations for the previous chromosome in the genome.

Displays aberrations for the next chromosome in the genome.

Amplification Select this option to display the aberrant regions that represent amplifications as red bars.

Deletion Select this option to display the aberrant regions that represent deletions as green bars.

Graphical aberration plots Display the selected types of aberrant regions for the selected chromosome for all arrays.

Set 1 Assigns the aberrations in the given array to comparison set 1.

Set 2 Assigns the aberrations in the given array to comparison set 2.

Ignore Does not assign the aberrations in the given array to a comparison set.

Select Algorithm Sets the algorithm used to score differential aberrations. This algorithm is available:

- **HyperGeometric Tail** – The Hypergeometric distribution tests for enrichment, or the likelihood of an increase in the number of aberrations present in one group, given the number of aberrations present in both groups. The HyperGeometric Tail sums the likelihoods as extreme or greater than the original number of aberrations from one group, to calculate an enrichment *p*-value.

For further information on this algorithm, see "Differential Aberration Analysis" on page 304.

Define Sets Opens the Define sets dialog box, where you can assign arrays to comparison groups. This is an alternative to the use of the Set1/Set2/Ignore options that appear below each array.

Run Runs the differential aberration analysis with the options you selected. The results appear in the Graphical Differential Aberration Summary dialog box. See "Graphical Differential Aberration Summary" on page 198.

Reset Resets the assignment of all arrays to comparison set 1.

Close Closes the dialog box.

Edit Aberration

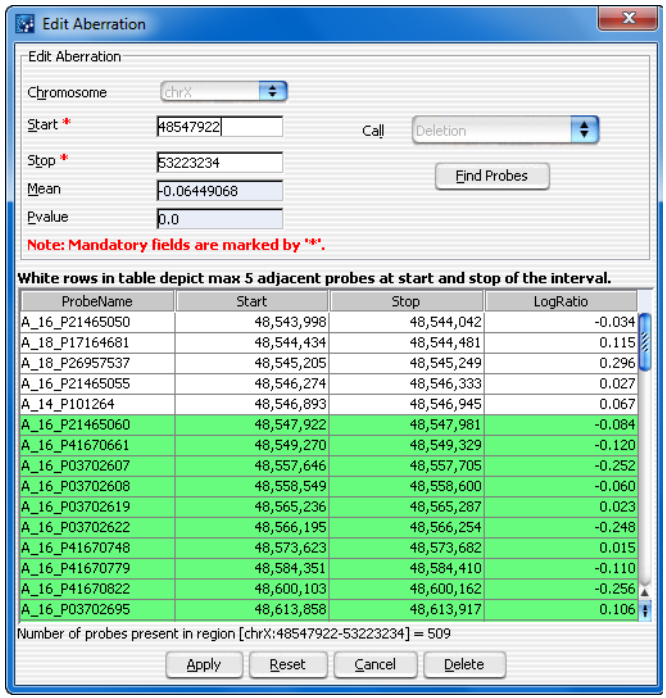


Figure 101 Edit Aberration dialog box

Purpose: Lets you make changes to an aberration result.

To open: In the Triage View, under Actions, click **Edit** for the aberration you want to change. You cannot change the type of aberration – only its location range, mean, and *p*Value. See "Triage View" on page 59.

Start Type the start location for the aberration call.

Stop Type the stop location for the aberration call.

Mean Type a new mean log ratio.

PValue Type a new *p*Value.

Find Probes Click to find and highlight the probes found in the selected range.

Apply Applies the changes to the selected aberration.

Reset Resets the changes made to the aberration.

Cancel Cancels the changes and closes the dialog box.

Delete Deletes the aberration interval from the table and closes the dialog box.

NOTE Changes made to aberrations are not saved unless you Check In or Sign Off the sample in Triage View.

Edit Array Order

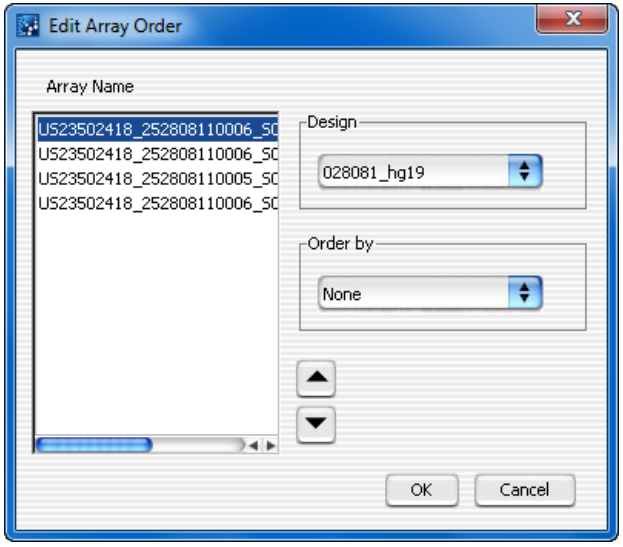


Figure 102 Edit Array Order dialog box

Purpose: This dialog box lets you change the order for display of arrays you open in the View Aberrations window.

To open: In the View Aberrations window, in the Arrays or SNP tab, right-click an array name, then click Edit Array Order in the shortcut menu. "View Aberrations Window" on page 54.

Array Name The arrays in the selected design, listed in their current order.

Design Select the name of a design. In Array Name, the program displays the arrays associated with the selected design.

Order by Select an attribute to use as a basis for ordering the list. For example, if you select Barcode, the program reorders that Array Name list based on Barcode.

 Moves a selected array up in the Array Name list. To select an array in this list, click the name of the array.

 Moves a selected array down in the Array Name list. To select an array in this list, click the name of the array.

OK Applies the new array order.

Cancel Closes the dialog box without making any changes to the array order.

Edit Cyto Report Template

Purpose: To change a cyto report template that can be used to generate a report at the end of a workflow analysis.

To open: In the Config tab, under Template, click **Report Template**. In the list of templates, click **View/Edit** for the report template you want to change.

The Edit Cyto Report Template dialog has four Steps. For details on these dialogs, see "Create Cyto Report Template" on page 158. The following commands appear at the bottom of all step dialogs, and are used to move forward and backward through the steps.

Back Click to move to the previous step.

Next Click to move to the next step.

Cancel Click to cancel the changes and exit the dialog box.

Edit User Roles

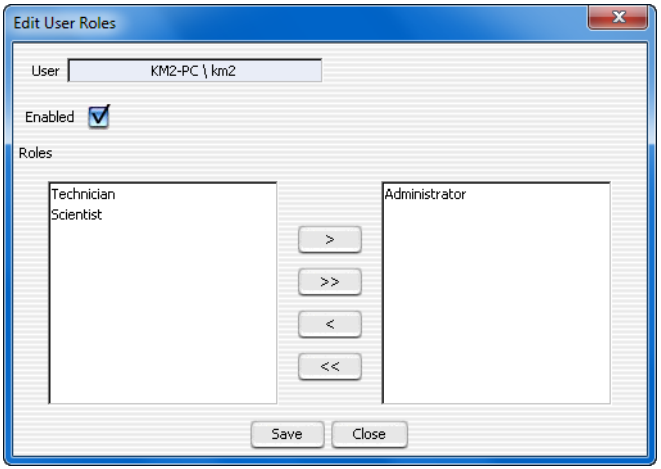


Figure 103 Edit User Roles

Purpose: To display or change the role associated with a user.

To open: In the Admin tab, under Actions, click **Edit** for the user for whom you want to change a role.

User (Read-only) The selected user is displayed here.

Enabled Select to enable the user for Agilent CytoGenomics 1.0. To prevent a user from using the program, clear this box. You can select the box at a later time, if desired.

Roles Available roles are displayed in the left pane. The assigned role for this user is displayed in the right pane. Move the roles from the left to the right and back using the arrows.

Assigns a role to the user, and moves the selected role from the left pane to the right pane.

Assigns all user roles, and moves all of the roles in the left pane to the right pane.

Removes a user role from the user and moves it to the left pane.

<<	Clears all selected user roles.
Save	Saves the assigned user roles for the user.
Close	Closes the dialog box.

Export

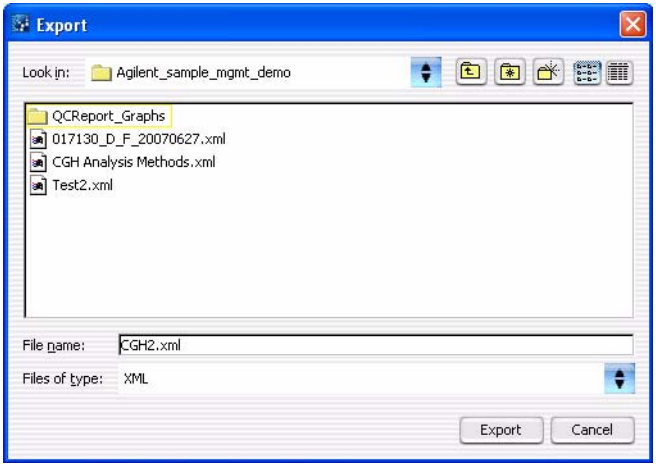


Figure 104 Export dialog box

Purpose: Used to designate a location and file name to export analysis methods or workflows.

To open: This dialog box opens when you click **Export** from the Workflow or Analysis Method window in the Config tab, or whenever the program saves database information to a hard disk location.

Use the buttons at the top of the dialog box to change the display.

Look in	Click the arrow and select the folder where you wish to export the file.
File name	Type the name you wish to use for the exported file.
Files of type	Shows the type of files displayed in the window. Click the arrow to change the type of files displayed.
Export	Click to export the selected item to the file.

Cancel Click to cancel the operation.

Export Attribute Files

Purpose: This dialog box contains two tabs used to define what Array IDs and attributes are exported to an exported attribute file.

To open: In the Content tab, in the Navigator, click **Sample Manager**. On the Sample Manager command ribbon, under Attribute File click **Export Attribute File**.

Array tab

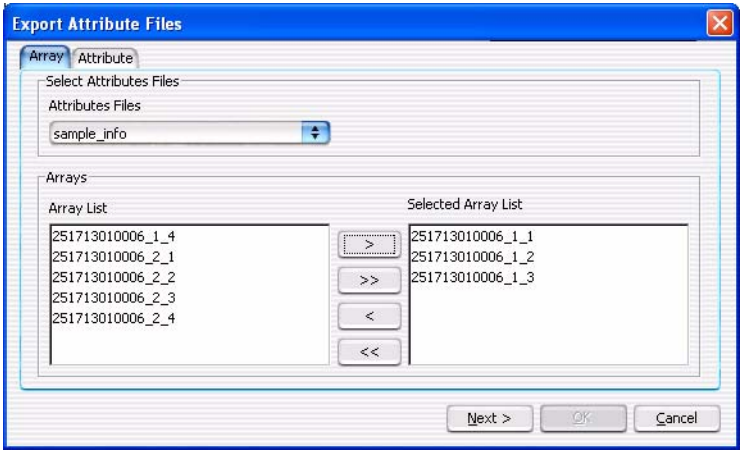


Figure 105 Export Attribute Files dialog box - Array tab

Purpose: This tab lets you select the arrays to export. Once you have selected the arrays you want to export, click **Next**. This opens the Export Attribute Files - Attribute tab.

To open: In the Export Attribute Files dialog box, click the **Array** tab.

Attribute Files Select an attribute file from the list. The sample arrays from this attribute file are displayed in the Arrays list. After you select arrays for export from this file, you can select another attribute file and then select more arrays. In this way, you create a custom attribute file that contains some or all of the arrays from different attribute files.

Array list Displays a list of available arrays for export. Move desired arrays for export to the Selected Array List using one of the following methods.

- To select an array to move to the Selected Array List, click its name.
- To select additional arrays, hold down the **Ctrl** key and click their names.
- To select a contiguous block of arrays, click the name of the first array, then hold down the **Shift** key and click the name of the last one.

Selected Array List Displays the arrays that are currently selected for export.

Moves the selected arrays in Array List to the Selected Array List.

Moves all of the arrays in Array List to the Selected Array List.

Removes an array from the Selected Array List. To select an array to be removed, click its name. If desired, you can add the array again.

Clears the Selected Array List.

Next Displays the Attribute tab that allows you to select attributes.

Cancel Closes the dialog box without selecting any array attributes to be exported.

Attribute Tab

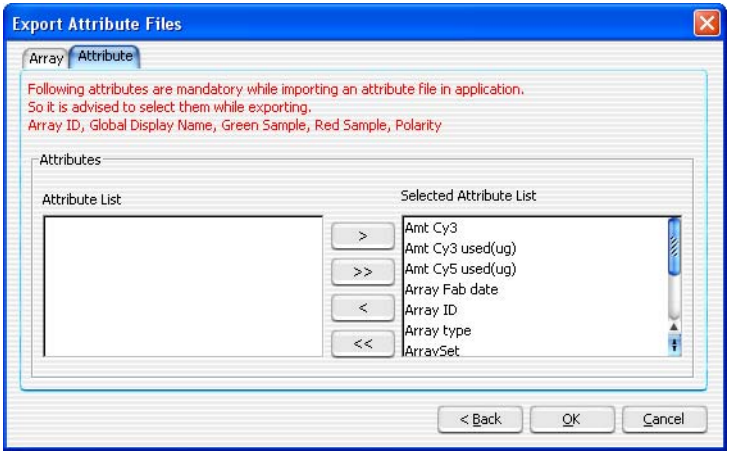


Figure 106 Export Attribute Files dialog box - Attribute tab

Purpose: This tab lets you select the attributes to include in the sample attribute file. When you have finished selecting the attributes for export, click **OK** and the Export dialog box will appear where you type the file name. See "Export" on page 184.

To open: In the Array tab of the Export Attribute Files dialog box, click **Next>** after you add one or more arrays to the Selected Array List.

NOTE

Because certain attributes are required for importing an attributes file, it is important that you select these attributes when you export an attributes file. Required attributes are: Array ID, Global Display Name, Green Sample, Red Sample, and Polarity.

Attribute list Displays a list of available attributes for export. Move desired attributes for export to the Selected Attribute List, using one of the following methods.

- To select an attribute for subsequent transfer to the Selected Attribute List, click its name.
- To select additional attributes, hold down the **Ctrl** key and click their names.

- To select a contiguous block of attributes, click the name of the first attribute, then hold down the **Shift** key and click the name of the last one.

Selected Attribute List Displays the attributes currently selected for export.

Moves the selected attributes in the Attribute List to the Selected Attribute List.

Moves all of the attributes in the Attribute List to the Selected Attribute List.

Removes an attribute from the Selected Attribute List. To select an attribute for removal, click its name. If desired, you can add the attribute again.

Clears the Selected Attribute List.

Back Moves to Array tab for selection of arrays.

Cancel Closes the dialog box without selecting any array attributes to be exported.

OK Opens the **Export** dialog box, where you find a location, type a file name, and export the data. See "Export" on page 184. You must select one or more arrays and one or more attributes before you can export an attribute file.

Feature Extraction Properties

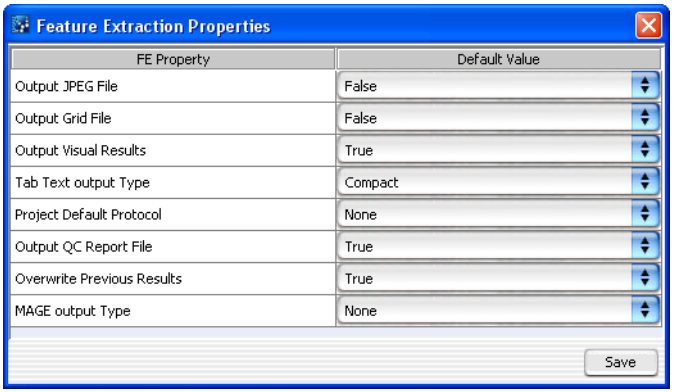


Figure 107 Feature Extraction Properties dialog box

Purpose: To use Workflow to display and change the parameters for Feature Extraction.

To open: In the Workflow window, under Input, select **Image Files**, then click **FE default parameters being used.** See "Workflow Window" on page 84.

This dialog box displays the current FE parameters. To change a parameter, click the arrow in the field next to the parameter and select the new value. Click **Save** to save any changes and close the dialog box. Click .

Find in column

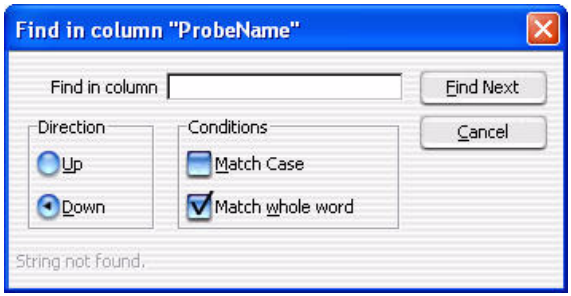


Figure 108 Find in column dialog box

Purpose: This dialog box allows you to set search parameters for a specific column entry for the selected chromosome. Based on these parameters, the program can highlight the row of the first entry that matches. The cursor then moves to the location defined in the row.

To open: Right-click any entry in a tab in Tab View other than a column heading, then click **Find in column** in the shortcut menu. See "Tab View for Genomic Viewer" on page 50.

Find in column Type all or part of the entry you want to find.

Direction Select a search direction:

- **Up** – Sets the search to scan the column you clicked in an upward direction from the highlighted row.
- **Down** – Sets the search to scan the column you clicked in an downward direction from the highlighted row.

Conditions Select any of these search options:

- **Match Case** – Select this option to take case into account. For example, if you select Match Case, and you type `aa351` in Find in column, the search finds the next entry in the column that contains **aa351.** It does *not* find entries that contain **AA351** or **Aa351.**

- **Match whole word** – Select this option to only find entries in which the complete entry matches what you type in Find in column. For example, if you type `AA351` in Find in column, and select **Match whole word,** the program finds the next **AA351** entry. It does not find entries such as **AA3512** or **AA351992.**

Find Next Finds the next matching entry in the selected column, and moves the cursor to the location defined in the row that contains the entry. The search is performed only for the chromosome selected in the Genome View.

Cancel Closes the dialog box.

Generating Genotype Reference File

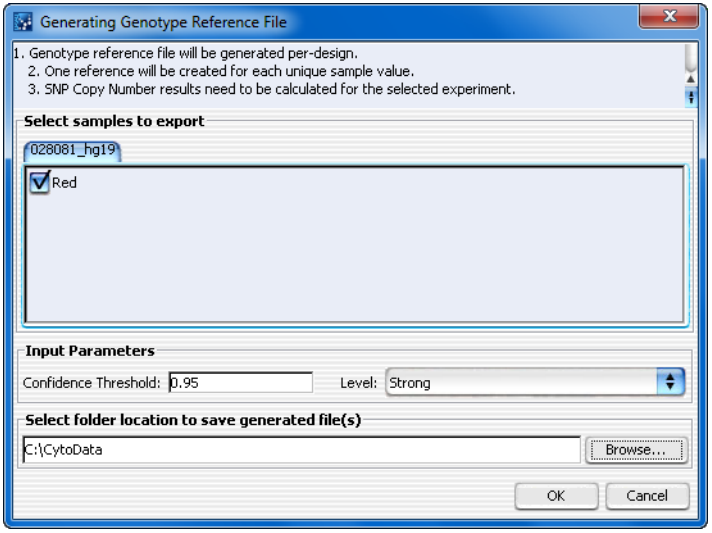


Figure 109 Generating Genotype Reference File dialog box

Purpose: Creates a genotype reference file using selected samples and parameters. You can import the new genotype reference file later to the database, and use it to analyze CGH+SNP data.

To open: In the View Aberrations window, under SNP, click **Generate Genotype Reference.** See "View Aberrations Window" on page 54.

Select samples to export

For samples you have set up for generating genotype references, select the check box next to the sample(s) you want to genotype. Only the arrays containing the selected sample(s) are used.

NOTE

In order to use a sample to generate a genotype reference, the Red Sample field for the microarray (Green for dye-flipped,) must contain the unknown reference sample name. A validated genotype reference must be selected in the other sample channel.

Input parameters

Select the confidence parameters that must be met to include the genotype information.

Confidence threshold

Type the confidence threshold to apply. A higher value makes the selection more stringent. For the default threshold of 95%, the following confidence assignments are made:

Confident: The software reports an AsCN within 0.05 of an integer. It is considered to have called the AsCN as the nearest integer, with 95% confidence.

Tentative: The software reports an AsCN which is *not* within 0.05 of an integer. It is considered to have called the AsCN as the nearest integer, but with low confidence.

No call: The software makes no call for the SNP in that sample.

Confidence level

Select a level of confidence for the selection.

- **Strong** – The SNP is called as Confident in at least one sample, and it is never called as Tentative. Also, all samples in which it is called report the same (integral) AsCN.
- **Weak** – All samples in which the SNP is called report the same (integral) AsCN. Some samples are called as Tentative.
- **Majority** – The SNP is called as Confident in at least one sample, but it is sometimes called with a different (integral) AsCN. All such calls are Tentative.
- **Contradictory** – The SNP is called with different (integral) AsCNs in different samples, but none of the calls are Confident.

Select folder location to save the generated file(s) Type the path to the folder where you want to save the generated genotype reference file.

Browse Click this to browse to and select the location where you want to save the generated genotype reference file.

Genotype Reference Importer

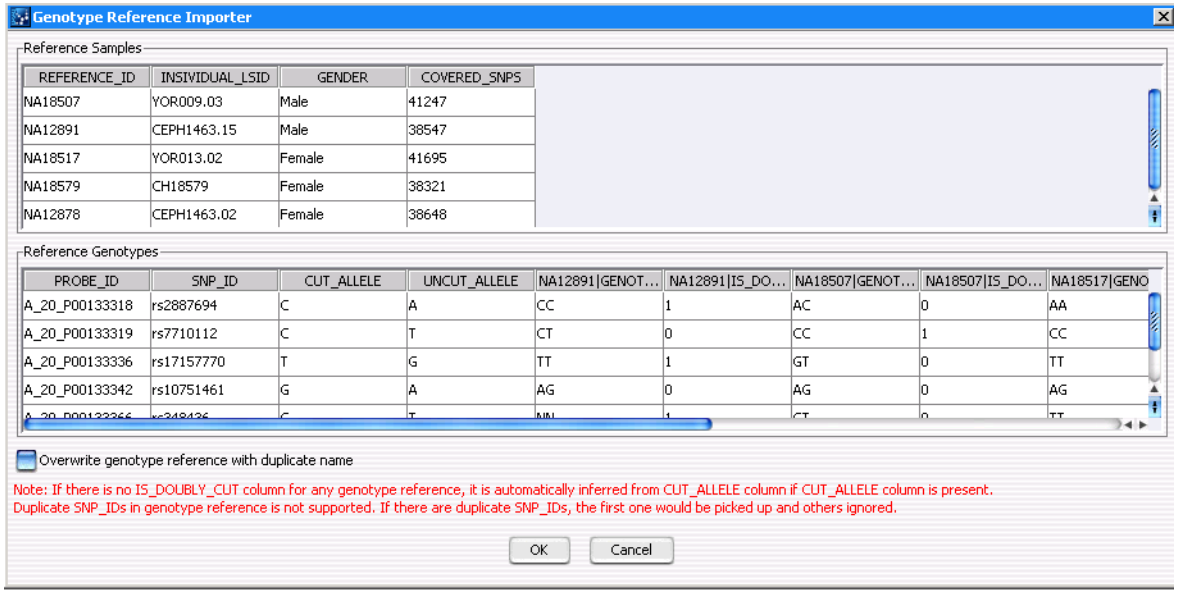


Figure 110 Genotype Reference Importer dialog box

Purpose: Displays the contents of a genotype reference file you want to import, and lets you choose to overwrite existing genotype references in the database when you import the file. For the requirements for this text file, see "Required format for genotype reference files" on page 194.

To open: From the Content tab, click **Import > Genotype References**. See "Content Tab Window" on page 95.

Reference Samples Displays a table of the samples in the file, including number of SNP probes covered by the sample.

Reference Genotypes Displays a table of the genotypes in the file. Duplicate SNP_IDs are not allowed. If there are duplicate SNP_IDs in the file, only the first SNP_ID is imported.

NOTE If the CUT_ALLELE column is present for a genotype reference, and there is no IS_DOUBLY_CUT column, the IS_DOUBLY_CUT column will be automatically inferred from the CUT_ALLELE column.

Required format for genotype reference files

A genotype reference file must be a tab delimited text file. It has the following format and requirements.

Header section – Contains description and information about the references.

- **Reference_id** – An identifier for the genotype reference
- **Individual_LSID** – ID for the individual
- **Gender** of the reference
- **Covered_SNPs** – how many SNPs in reference file
- **dbSNP_version** – the SNP version

Reference genotype table – each row corresponds to a Probe, with the following column headings

- **Probe_Id**
- **Species**
- **Snp_id** – The ID of SNP associated with probe. There may be more than one probe associated with a SNP. Duplicate SNP IDs are not allowed. If there are duplicates, only the first one is used.
- **Cut Allele**
- **Uncut Allele**
- **Reference ID|genotype** – genotype for the SNP in that reference.
- **Reference ID|is_doubly_cut** – a flag denoting whether the SNP is doubly cut in the reference (1 doubly cut, 0 not doubly cut)

#reference samples

reference_id	individual_LSID	Gender	covered_SNPs	dbSNP_version
NA18507	YOR009.03	Male	41247	130
NA12891	CEPH1463.15	Male	38547	130
NA18517	YOR013.02	Female	41695	130
NA18579	CH18579	Female	38321	130
NA12878	CEPH1463.02	Female	38648	130

#

#reference genotypes

probe_id	species	snp_id	cut_allele	uncut_allele	NA12891|genotype	NA12891|is_doubly_cut	NA18507|genotype	NA18507|is_doubly_cut
A_20_P00122122	Hs	rs2405741	G	C	CC	0	CC	0
A_20_P00224028	Hs	rs2405741	G	C	CC	0	CC	0
A_20_P00122159	Hs	rs1492169	G	A	AA	0	AG	0

Figure 111 Example of part of a genotype reference file (including five references)

Graphical Penetrance Summary

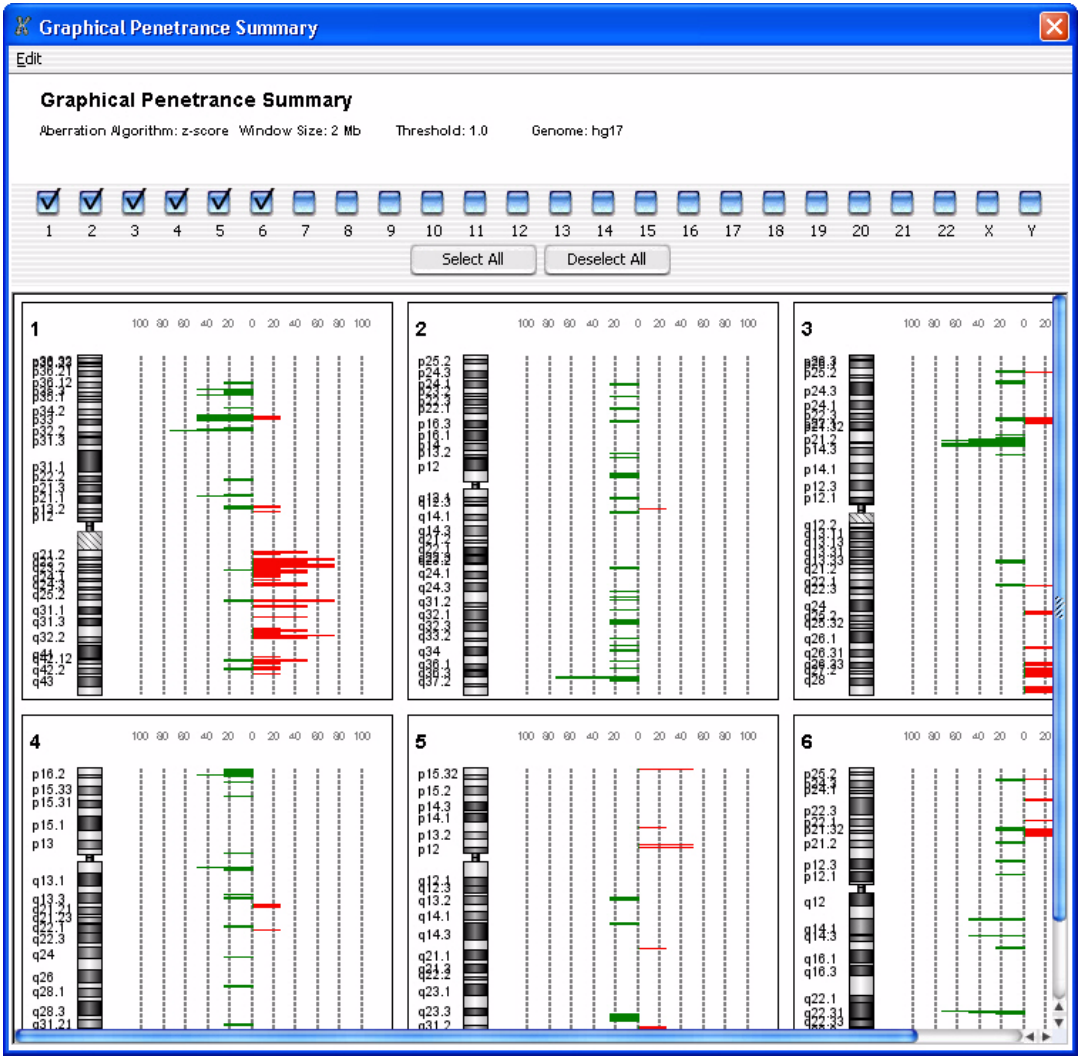


Figure 112 Graphical Penetrance Summary

Purpose: This window allows you to display the results of probe penetrance analysis for each chromosome in the genome.

To open: In the command ribbon of the View Aberrations window, under Penetrance, click **Probe.** See "View Aberrations Window" on page 54.

Probe penetrance analysis produces penetrance scores for each probe position across multiple arrays. A probe penetrance score reflects the percentage of the selected arrays in the experiment that show a significant amplification or deletion at the position. For a more detailed description of penetrance analysis, see "Penetrance" on page 302.

Plots Penetrance plots for selected chromosomes appear in the main pane. For each probe position on a given chromosome, the percentage of arrays that show a significant deletion appear in green. The percentage of arrays that show a significant amplification appear in red. Amplifications and deletions reflect the results of the experimental analysis as configured in the main window of the program.

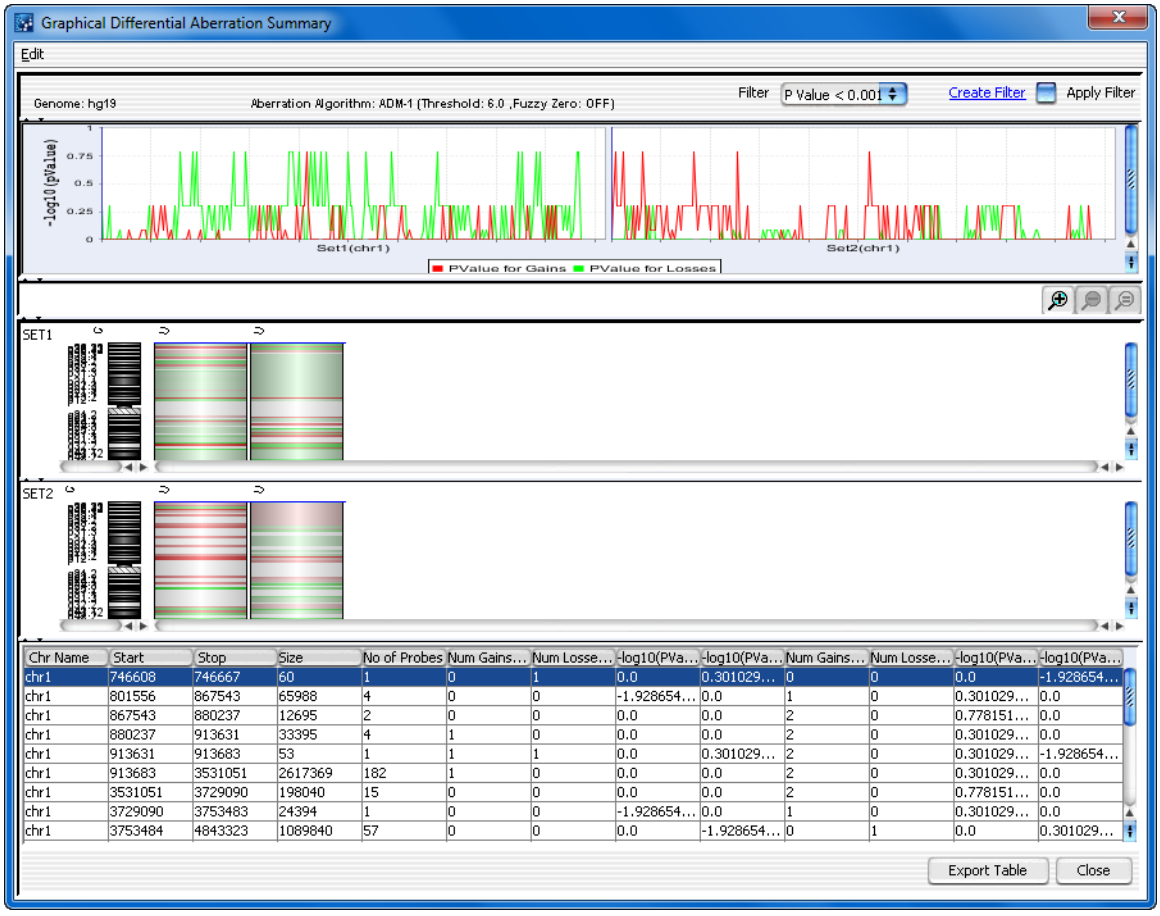
When you double-click within the plot area of any penetrance plot or within any chromosome, the cursor in Genome, Chromosome, Gene, and Tab Views in the main window moves to the new location.

Select Chromosomes The check boxes at the top of the window correspond to the chromosomes of the applicable genome. Select as many as you like. The program displays a probe penetrance plot for each selected chromosome.

Select All Produces probe penetrance plots for all chromosomes in the genome.

Deselect All Clears all of the check boxes, and removes all of the probe penetrance plots from the window.

Graphical Differential Aberration Summary



Chr Name	Start	Stop	Size	No of Probes	Num Gains...	Num Losse...	-log10(PVa...	-log10(PVa...	Num Gains...	Num Losse...	-log10(PVa...	-log10(PVa...
chr1	746608	746667	60	1	0	1	0.0	0.301029...	0	0	0.0	-1.928654...
chr1	801556	867543	65988	4	0	0	-1.928654...	0.0	1	0	0.301029...	0.0
chr1	867543	880237	12695	2	0	0	0.0	0.0	2	0	0.778151...	0.0
chr1	880237	913631	33395	4	1	0	0.0	0.0	2	0	0.301029...	0.0
chr1	913631	913683	53	1	1	1	0.0	0.301029...	2	0	0.301029...	-1.928654...
chr1	913683	3531051	2617369	182	1	0	0.0	0.0	2	0	0.301029...	0.0
chr1	3531051	3729090	198040	15	0	0	0.0	0.0	2	0	0.778151...	0.0
chr1	3729090	3753483	24394	1	0	0	-1.928654...	0.0	1	0	0.301029...	0.0
chr1	3753484	4843323	1089840	57	0	0	0.0	-1.928654...	0	1	0.0	0.301029...

Figure 113 Graphical Differential Aberration Summary dialog box

Purpose: This shows the detected differentially aberrant regions between two sets of microarrays, both graphically and as a table.

To open: The Graphical Differential Aberration Summary dialog box opens when you click **Run** in the Differential Aberration Setup dialog box that is part of the differential aberration analysis process. See "Differential Aberration Setup" on page 178.

NOTE At least two samples are required in order to run a differential aberration analysis.

Analysis Summary The top of the dialog box shows information about the original aberration analysis, including the aberration detection algorithm that was used, and other associated analysis settings.

Line Plot(s) The line plots in the dialog box represent the p-values for gains and losses in both comparison sets as a function of genomic location. When you select a row in the table of differentially aberrant regions at the bottom of the dialog box, the line plots for the corresponding chromosome appear. Also, the blue line in the line plots reflects the genomic location of the selected row.

In addition, when you right-click anywhere within the line plots, a shortcut menu opens with these options:

- **Properties** – Opens the Chart Properties dialog box, where you can customize the plots, and add and format a title and a legend. See "Chart Properties" on page 147.
- **Save as** – Opens a Save dialog box, where you can type a name and select a location for an exported *.png image file of the plots.
- **Print** – Opens a Page Setup dialog box, which allows you to set page and printer options for the printing of the plots. After you click **OK** in this dialog box, the Print dialog box appears, where you can set additional options, and send the plots to your printer.
- **Zoom In** – Opens a menu that allows you to zoom in the scale of all of the line plots. These options are available:
  - **Both Axes** – Zooms in both axes of all plots
  - **Domain Axis** – Zooms in only the chromosomal location axis
  - **Range Axis** – Zooms in only the p-value axis
- **Zoom Out** – Opens a menu that allows you to zoom out the scale of all of the line plots. These options are available:

  - **Both Axes** – Zooms out both axes of all plots
  - **Domain Axis** – Zooms out only the chromosomal location axis
  - **Range Axis** – Zooms out only the *p*-value axis
- **Auto Range** – Adjusts the scale of both axes for optimal viewing of all of the data associated with a given chromosome.

Cylinder plots

The main pane of the summary shows heat-map style plots of the aberrant regions of each individual array next to an ideogram of the applicable chromosome. Detected amplifications appear in red, and detected deletions appear in green.

This pane has several additional features:

- A *differential aberration,* as detected by the analysis, appears as a blue box around the given region across all of the plots. To display each of these regions, and to move among the chromosomes, click a row of the results table that appears at the bottom of the dialog box.
- To identify the array that a particular plot represents, place the pointer over the plot. A ToolTip shows the name of the array.
- To zoom the view, click any of these zoom buttons, located at the top right of the pane:

 – Zooms in the view.

 – Zooms out the view.

 – Resets the zoom to its initial setting (zoomed all the way out to see the entire chromosome).

Table of regions

At the bottom of the window, the detected differentially aberrant regions appear in a table. Click any row of the table to shift the plots to the appropriate chromosome. In the *p*-value line plots, a blue line shows the specific chromosomal location of the selected region. In the aberration plots, a blue box identifies the differentially aberrant regions.

These columns appear in the table:

Column	Description
Chr Name	The name of the chromosome on which the differential aberration is found.
Start	The location of the first base pair in the differentially aberrant region.
Stop	The location of the last base pair in the differentially aberrant region.

Column	Description
Size	The number of base pairs in the differentially aberrant region.
No of Probes	The number of probes upon which the differential aberration call is based.
Num Gains in Set 1	Number of aberration gains for the differentially aberrant region in first group.
Num Losses in Set 1	Number of aberration losses for the differentially aberrant region in the first group.
-log10(PVal) Gain in Set 1	The likelihood that the region represents a differential gain for set 1. (Lower *p*-values indicate a greater likelihood.)
-log10(PVal) Loss in Set 1	The likelihood that the region represents a differential loss for set 1. (Lower *p*-values indicate a greater likelihood.)
Num Gains in Set 2	Number of aberration gains for the differentially aberrant region in second group.
Num Losses in Set 2	Number of aberration losses for the differentially aberrant region in the second group.
-log10(PVal) Gain in Set 2	The likelihood that the region represents a differential gain for set 2. (Lower *p*-values indicate a greater likelihood.)
-log10(PVal) Loss in Set 2	The likelihood that the region represents a differential loss for set 2. (Lower *p*-values indicate a greater likelihood.)

Other commands

The dialog box also offers these additional commands:

Edit Opens a menu with a Copy summary to clipboard command. This command copies the entire summary to the clipboard as an image. You can then paste the image into a document in another program.

Filter Displays the interval filters available for differential aberration results. Interval filters include or exclude certain intervals based on specific conditions. Select an interval filter. To apply the filter, select **Apply Filter.** To create a new interval filter, click **Create Filter** (see below).

Create Filter Click this link to open the Interval Filter dialog box, where you can create, change, or delete interval filters. See "Interval Filter" on page 210.

Apply Filter Select this option to apply the selected interval filter to the results.

Export Table Opens an Open dialog box, where you can type a name and select a location for a *.xls file that contains the list of aberration analysis settings, and the complete table of detected differentially aberrant regions.

Close Closes the dialog box.

Graphical Interval Penetrance Summary

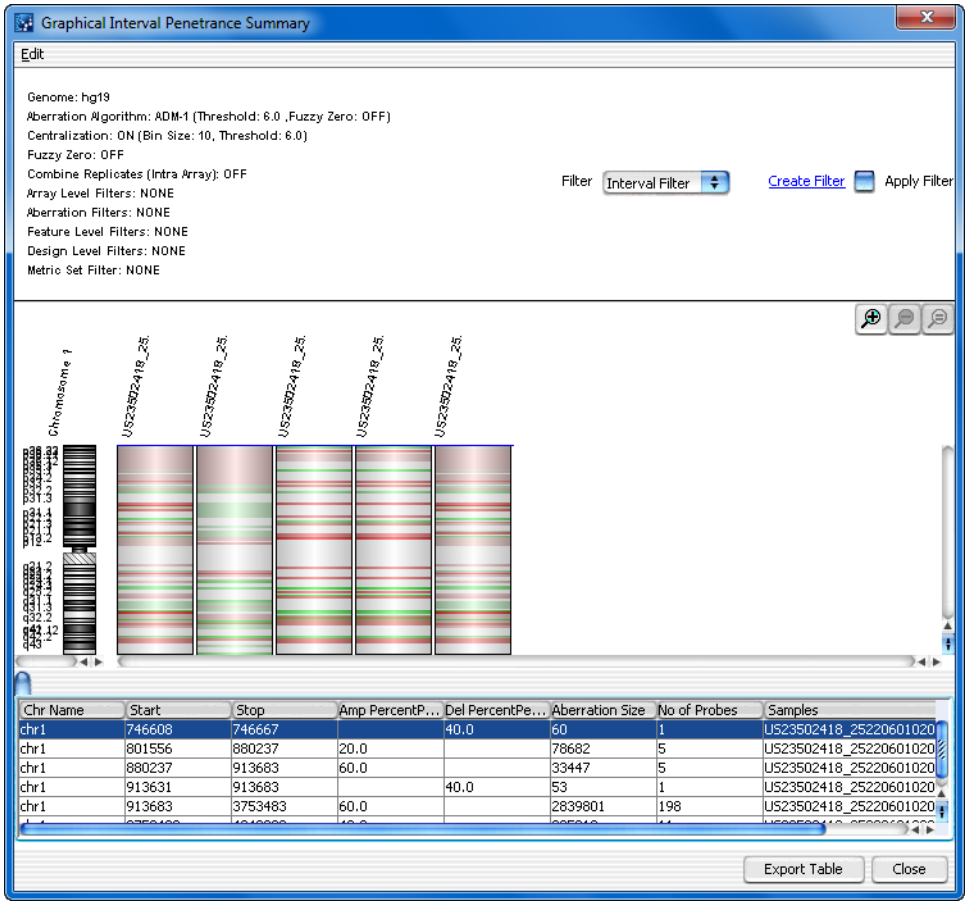


Figure 114 Graphical Interval Penetrance Summary

Purpose: This window allows you to display and save interval penetrance results for the selected arrays in the View Aberrations window. Interval penetrance analysis produces scores for detected aberrant intervals across multiple samples. These scores reflect the percentage of samples that show each region to be aberrant. For a more detailed discussion of penetrance analysis, see "Penetrance" on page 302.

To open: In the View Aberrations window, under Penetrance, click **Interval.** See "View Aberrations Window" on page 54.

Main Pane

The main pane of the summary shows heat-map style plots of the aberrant regions of each individual array next to an ideogram of the applicable chromosome. Detected amplifications appear in red, and detected deletions appear in green.

Each interval scored by the analysis appears as a blue box around the given region across all of the plots. To display each of these regions, and to move among the chromosomes, click a row of the results table that appears at the bottom of the window.

To zoom the view, click any of these zoom buttons, located at the top right of the main pane:

 – Zooms in the view.

 – Zooms out the view.

 – Resets the zoom to its initial setting (zoomed all the way out to see the entire chromosome).

Results table

At the bottom of the window, the scored aberrant intervals appear in a table. Click any row of the table to move to the appropriate chromosome, and center the graphical plots in the main pane on the particular aberrant interval.

These columns appear in the table:

Column	Description
Chr Name	The name of the chromosome on which the aberrant interval is found.
Start	The location of the first base pair in the aberrant interval.
Stop	The location of the last base pair in the aberrant interval.
Amp Percent Penetrance	The percentage of arrays that show a significant amplification for the interval.
Del Percent Penetrance	The percentage of arrays that show a significant deletion for the interval.
Aberration Size	The number of base pairs in the aberrant interval.

Column	Description
No of Probes	The number of probes used to make the aberrant interval call.
Samples	The arrays on which an aberration was found in this interval.

Other commands The window also offers these additional commands:

- **Edit** – Opens a menu with a Copy summary to clipboard command. If you click this command, the program copies the entire summary to the clipboard as an image. You can then paste the image into a document in another program.
- **Filter** – Displays the interval filters available for the interval penetrance results. Interval filters include or exclude certain aberrant intervals based on specific conditions. Select an interval filter. To apply the selected filter, select **Apply Filter.** To create a new interval filter, click **Create Filter** (see below).
- **Create Filter** – Click this link to open the Interval Filter dialog box, where you can create, change, or delete interval filters. See "Interval Filter" on page 210.
- **Apply Filter** – Select this option to apply the selected interval filter to the results.
- **Export Table** – Opens an Open dialog box, where you can type a name and select a location for the exported table. The program saves the results table that appears at the bottom of the window as a *.xls file you can view in Microsoft Excel.
- **Close** – Closes the Graphical Interval Penetrance Summary.

Import Analysis Method(s)

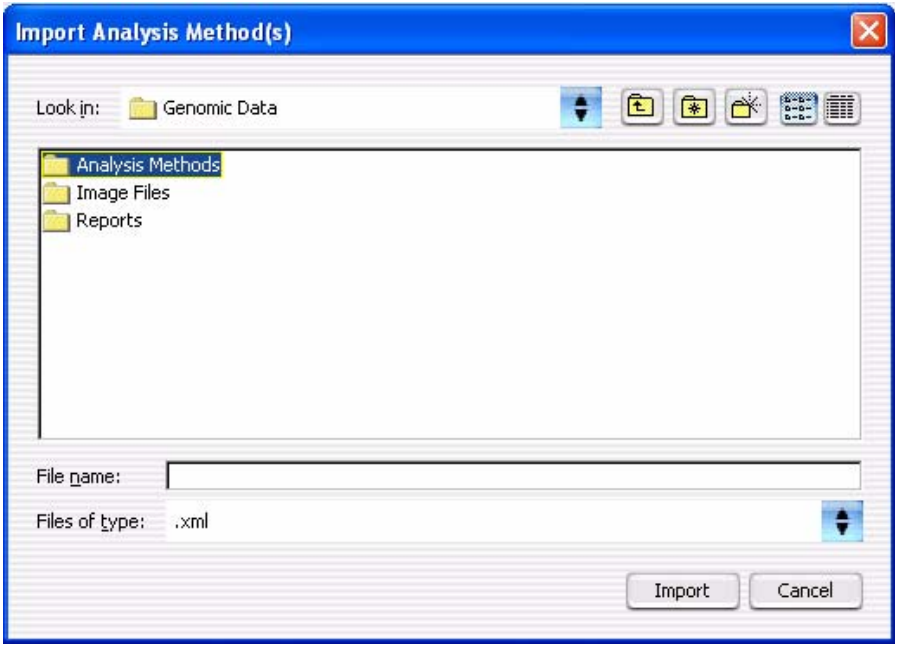


Figure 115 Import Analysis Method(s) dialog box

Purpose: Used to select an analysis method file to be imported into the program.

To open: In the command ribbon of the Config tab, under Workflow, click **Analysis Method**, then at the bottom of the window, click **Import Analysis Method**. See "Analysis Method Window" on page 79.

Look in Click the arrow and select the folder from which you want to import a file.

File name Type the name of the file you wish to import. Or, click to select the file from the displayed files.

Files of type Shows the type of files displayed in the window. Click the arrow to change the type of files displayed.

Import Click to import the currently-selected analysis method file into the program.

Cancel Click to cancel the operation.

Import GEML design files

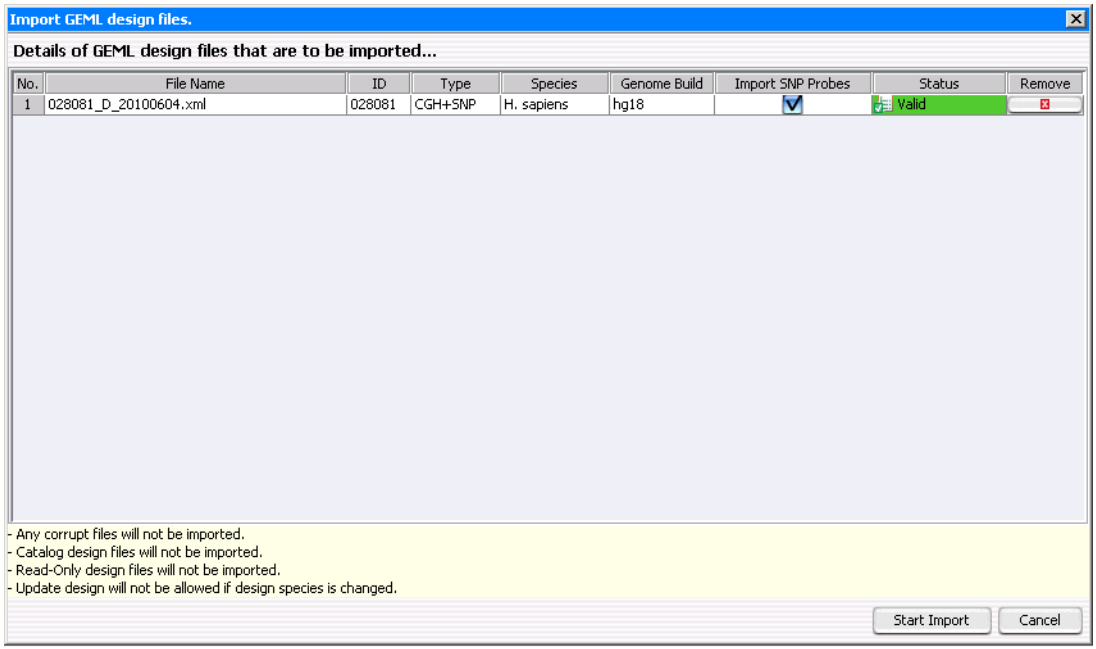


Figure 116 Import GEML design files dialog box

Purpose: To display information in the design file and to remove any files you don't want to import.

To open: In the Content tab Navigator, click **Sample Manager**. On the Sample Manager command ribbon, click **Import > Design.** Select the desired *.xml design files, then click **Open.** See "Content tab – Sample Manager window" on page 95.

File Name The name(s) of the design file(s) to be imported.

ID The Agilent ID number for the design file.

Type The module type, which can be CGH or CGH+SNP.

Species The species for the genome build. This appears automatically when the Genome Build is selected.

Genome Build The genome build for the design. If the genome build is not read automatically, a "?" appears. Click **Genome Build** and select the correct value from the list.

Import SNP Probes Available for CGH+SNP designs. Select the box to include SNP probes with the design data.

Status

- **Not Set** – Appears if Genome Build and Species information is not shown.
- **Not Allowed** – Appears if a Genome Build is selected that does not match the design.
- **Overwrite** – Appears when the design file has been updated and will overwrite any existing one of the same name.
- **Update** – Appears when the design build is different than the design build in the database.
- **Valid** – Appears when the file is new.
- **Corrupt** – Appears when the file is corrupt.

Remove Click  to remove a specific design file from the list.

Start Import Starts the import of the design files in the list.

Cancel Cancels the transfer and closes the dialog box.

Import Workflow(s)

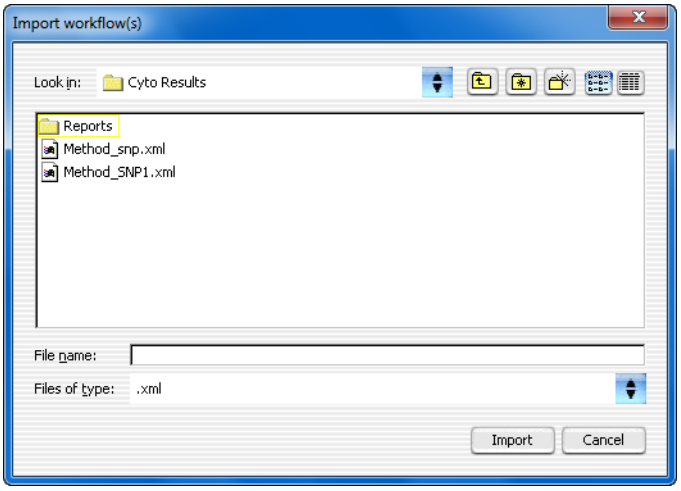


Figure 117 Import Workflow(s) dialog box

Purpose: Used to select an workflow file to be imported into the program.

To open: In the Config tab, under Workflow, click **Workflow**. At the bottom of the window, click **Import Workflow**. See "Workflow Window" on page 84.

Look in Click the arrow and select the folder from which you want to import a file.

File name Type the name of the file you wish to import. Or, click to select the file from the displayed files.

Files of type Shows the type of files displayed in the window. Click the arrow to change the type of files displayed.

Import Click to import the currently-selected workflow file into the program.

Cancel Click to cancel the operation.

Interval Filter

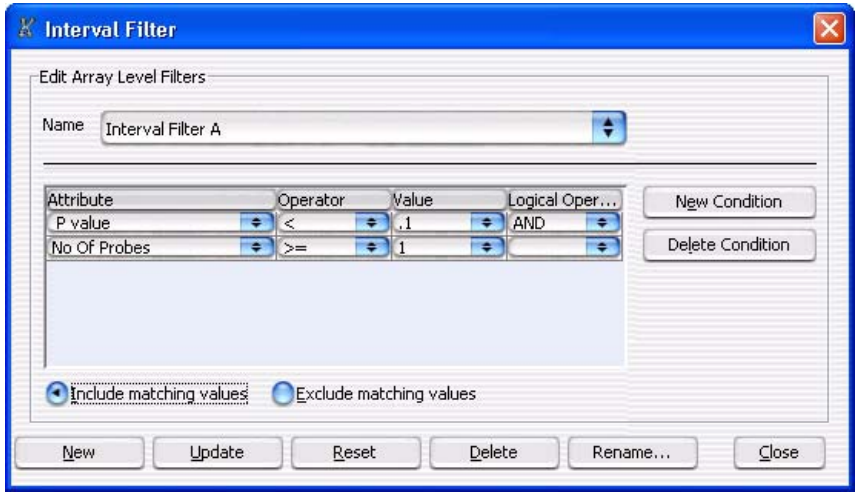


Figure 118 Interval Filter dialog box

Purpose: Use the Interval Filter dialog box to create, change, or delete interval filters. An interval filter includes or excludes certain detected intervals from the results, based on specific conditions.

To open: The Interval Filter dialog box opens when you click **Create Filter** in any of these dialog boxes:

- Graphical Differential Aberration Summary – See "Graphical Differential Aberration Summary" on page 198.
- Graphical Interval Penetrance Summary – See "Graphical Interval Penetrance Summary" on page 203.

Name Select the name of the interval filter you that want to edit. To create a new filter, and add its name to the list, click **New.** You can apply interval filters to both common aberration results and interval penetrance results. However, because the available attributes for interval filters in these two situations are not the same, the filters you create in one context are not available in the other.

Filter conditions For each condition (row), select options from the lists. Specifically in **Value,** select an option from the list, if available, or type a value, then press **Enter.** To add another row to the table, click **New Condition.**

Each condition has these elements:

- **Attribute** – The attribute of the intervals evaluated by the filter. The attributes available in this list vary with the specific context of the filter.
- **Operator** – How the filter uses the entry in Value to evaluate common aberrant intervals. For example, the **>=** operator configures the filter to include or exclude features where the selected attribute is greater than or equal to the entry in Value.
- **Value** – The value the filter uses as the basis to evaluate an interval. For example, if you select the **Aberration Size** attribute, the **>** operator, and a value of **1000,** the filter includes or excludes an interval if it is greater than 1000 bp in size.
- **Logical Operator** – (Available only if you configure more than one condition) The relationship between the condition and the next one in the list. For example, if you select **AND** in **Logical Operator** for the first condition, the filter includes or excludes an interval if it passes both the first condition and the next condition.

Include/Exclude matching values Select one of these options:

- **Include matching values** – If an interval passes the filter conditions, the program *includes* it in the final result.
- **Exclude matching values** – If an interval passes the filter conditions, the program *excludes* it from the final result.

New Condition Adds a new, blank condition (row) to the table.

Delete Condition Removes a condition from the list. To remove a specific condition, click anywhere within the condition, then click **Delete Condition.**

New Opens an Input dialog box, where you can type a name for a new filter. To accept the name, click **OK.** The program creates a new filter, and adds its name to the Name list.

Update Saves any changes you make to the filter conditions.

Reset Restores the values of the filter conditions to what they were before you made any changes to them.

Delete Opens a Confirm dialog box that asks you if you want to delete the selected filter. To delete the filter, click **Yes.**

Rename Opens an Input dialog box where you can type a new name for the filter. To accept the name, click **OK.**

Close Closes the dialog box. If you created or changed a filter, but did not update it, a Confirm dialog box opens. Click **Yes** to accept the changes, **No** to reject the changes, or **Cancel** to return to the dialog box.

Manage Cyto Report

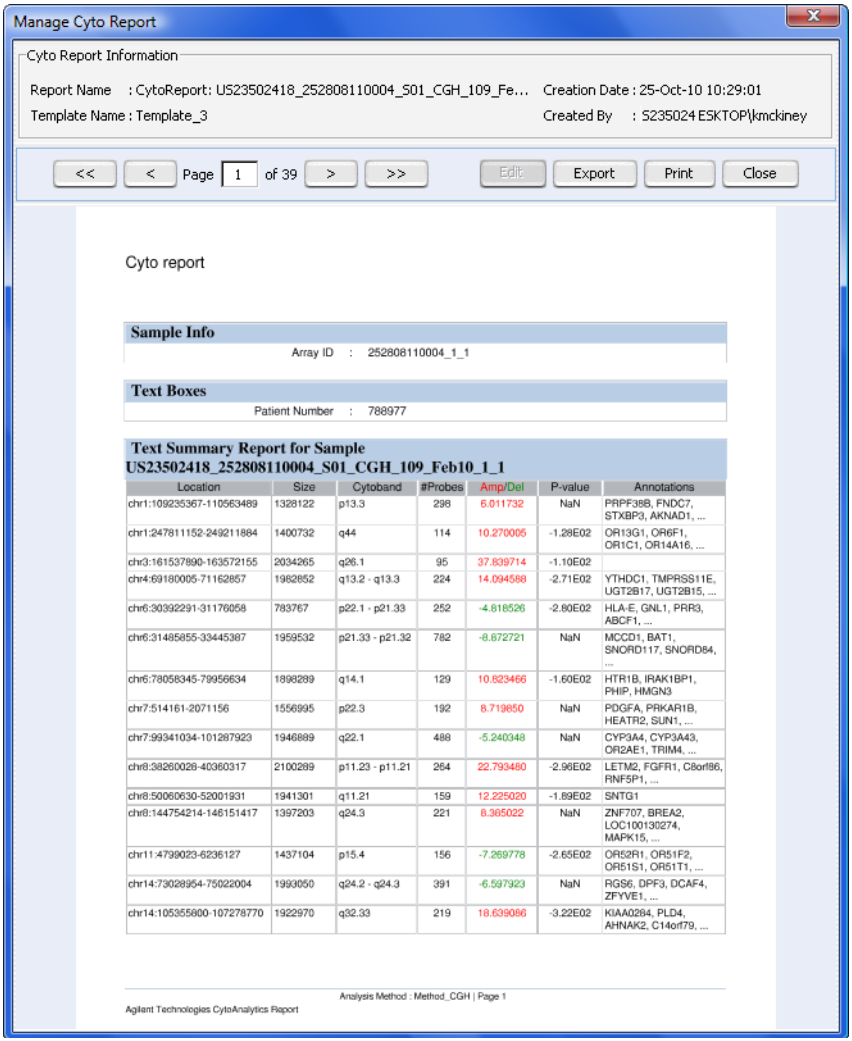


Figure 119 Manage Cyto Report dialog box

Purpose: To display, export, and print a workflow report. Also used to change editable text boxes in the report.

To open: In the Analysis tab, under Report, click **Report**. Search for and select an analyzed sample, and then click **View Report** at the bottom of the window. See "Report window" on page 33.

>	Displays the next page of the report.
>>	Displays the last page of the report.
Page x **of** y	Displays the current page number (x) and the total number of page numbers (y).
<	Displays the previous page of the report.
<<	Displays the first page of the report.
Edit	Opens the Edit Cyto Report panel, where you can change or type information in the editable text boxes.
Save	Saves the report with changes you made to the text boxes.
Save As	Lets you type a new report name and saves it with the changes you made to the text boxes.
Cancel	Cancels the text box changes and returns to the main view.
Export	Opens the Select report name dialog box, where you can browse to a location and type a name to save the report. See "Select Report Name" on page 241.
Print	Prints the report to the default printer for your computer.
Close	Closes the dialog box.

Metric Set Configuration Dialog Box: Add Metrics to Metric Set Tab

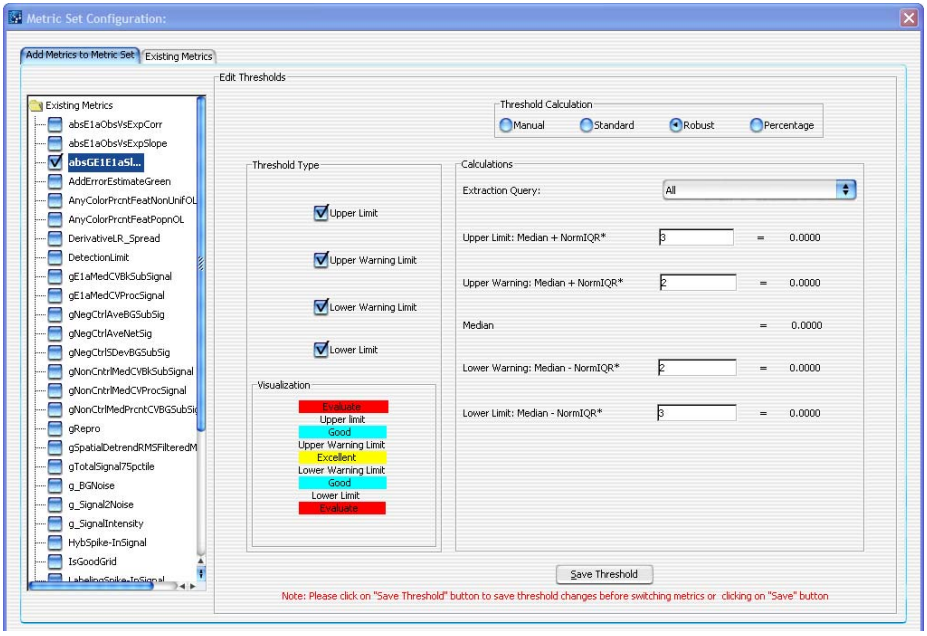


Figure 120 Metric Set Configuration dialog box - Add Metrics to Metric Set tab

Purpose: Used to define a metric set that consists of existing metrics and user-defined thresholds.

To open: On the Quality ribbon, click Metric Sets, and then click New. See "Quality Tab Window" on page 87.

The Add Metrics to Metric Set tab of the Metric Set Configuration dialog box has the following components and functionality:

Existing Metrics Displays a list of the metric(s) that can be used in the metric set.

Threshold Type

The following fields are threshold types, which are used to select which threshold(s) to apply and display.

Upper Limit Sets a limit where extraction values for the appropriate metric that are greater than the limit calculation are displayed in the color red and flagged as “Evaluate”. Extractions with metric values lower than this limit are displayed in the color blue and flagged as “Good”, unless there are additional limits selected that may further separate the extractions.

Upper Warning Limit Sets a limit where extraction values for the appropriate metric that are greater than the limit calculation are displayed in the color blue and flagged as “Good”. Extractions with metric values lower than this limit are displayed in the color yellow and flagged as “Excellent”, unless there are additional limits selected that may further separate the extractions.

Lower Warning Limit Sets a limit where extraction values for the appropriate metric that are less than the limit calculation are displayed in the color blue and flagged as “Good”. Extractions with metric values higher than this limit are displayed in the color yellow and flagged as “Excellent”, unless there are additional limits selected that may further separate the extractions.

LowerLimit Sets a limit where extraction values for the appropriate metric that are less than the limit calculation are displayed in the color red and flagged as “Evaluate”. Extractions with metric values higher than this limit are displayed in the color blue and flagged as “Good”, unless there are additional limits selected that may further separate the extractions.

Threshold Calculation

Manual Applies a constant value for Upper Limit, Upper Warning Limit, Lower Warning Limit, and LowerLimit.

Standard Takes a constant value as the number of standard deviations of the data beyond the mean to calculate the limit. For detailed information, see “Standard Threshold Calculations” on page 217.

Robust Takes a constant value as the number of inter-quartile ranges (IQR) of the data beyond the mean to calculate the limit. For detailed information, see “Robust Threshold Calculations” on page 217.

Percentage Takes a percentage range of the data to calculate the limit. For detailed information, see “Percentage Threshold Calculations” on page 218.

Extraction Query Optional: Lets you assign an Extraction Query to the metric set for Standard, Robust, and Percentage Threshold Calculation modes. This query will filter the appropriate extractions from the database, so that just the data from the queried extractions is used in the calculation of the statistical summary values used to set the thresholds.

Standard Threshold Calculations

Take a constant value as the number of standard deviations of the data beyond the mean to calculate the limits.

Upper Limit Lets you define a multiplier for the number of standard deviations to be added to the mean to create the Upper Limit. For example, to apply an upper limit of 2 standard deviations, enter the number 2 in the text field. The Upper Limit is defined as the mean + constant*SD.

Upper Warning Limit Lets you define a multiplier for the number of standard deviations to be added to the mean to create the Upper Warning Limit. The Upper Warning Limit is defined as the mean + constant*SD.

Lower Warning Limit Lets you define a multiplier for the number of standard deviations to be subtracted from the mean to create the Lower Warning Limit. The Lower Warning Limit is defined as the mean - constant*SD.

LowerLimit Lets you define a multiplier for the number of standard deviations to be subtracted from the mean to create the Lower Limit. The Lower Limit is defined as the mean - constant*SD.

Robust Threshold Calculations

Take a constant value as the number of inter-quartile ranges (IQR) of the data beyond the mean to calculate the limits.

Upper Limit Lets you define a multiplier for the number of IQRs to be added to the mean to create the Upper Limit. For example, to apply an upper limit of 2 IQR, enter the number 2 in the text field. The Upper Limit is defined as the Median + constant*IQR.

Upper Warning Limit Lets you define a multiplier for the number of IQRs to be added to the mean to create the Upper Warning Limit. The Upper Warning Limit is defined as the Median + constant*IQR.

Lower Warning Limit Lets you define a multiplier for the number of IQRs to be subtracted from the mean to create the Lower Warning Limit. The Lower Warning Limit is defined as the Median - constant*IQR.

LowerLimit Lets you define a multiplier for the number of IQRs to be subtracted from the mean to create the Lower Limit. The Lower Limit is defined as the Median - constant*IQR.

Percentage Threshold Calculations

Take a percentage range of the data to calculate the limits.

Upper Limit Lets you define percentage from the uppermost values of the data to calculate the Upper Limit. For example, to set an Upper Limit that is equal to the best 1% of the extractions for a particular metric, set the Upper Limit text box to 99%.

Upper Warning Limit Lets you define a multiplier for the percentage from the uppermost values of the data to calculate the Upper Warning Limit. For example, to set an Upper Warning Limit that is equal to the best 25% of the extractions for a particular metric, set the Upper Warning Limit text box to 75%.

Lower Warning Limit Lets you define percentage from the bottommost values of the data to calculate the Lower Warning Limit. For example, to set an Lower Warning Limit that is equal to the lower 25% of the extractions for a particular metric, set the Lower Warning Limit text box to 25%.

Lower Limit Lets you define percentage from the bottommost values of the data to calculate the Lower Limit. For example, to set an Lower Limit that is equal to the lower 1% of the extractions for a particular metric, set the Lower Limit text box to 1%.

Save Threshold Saves the threshold settings for the selected metric(s). These metric(s) with their associated thresholds become available in the "Metric Set Configuration Dialog Box: Existing Metrics Tab" on page 219.

Metric Set Name Lets you enter a name for the new metric set.

Save Saves the metric set.

Cancel Closes the Metric Set Configuration dialog box without saving any changes.

Metric Set Configuration Dialog Box: Existing Metrics Tab

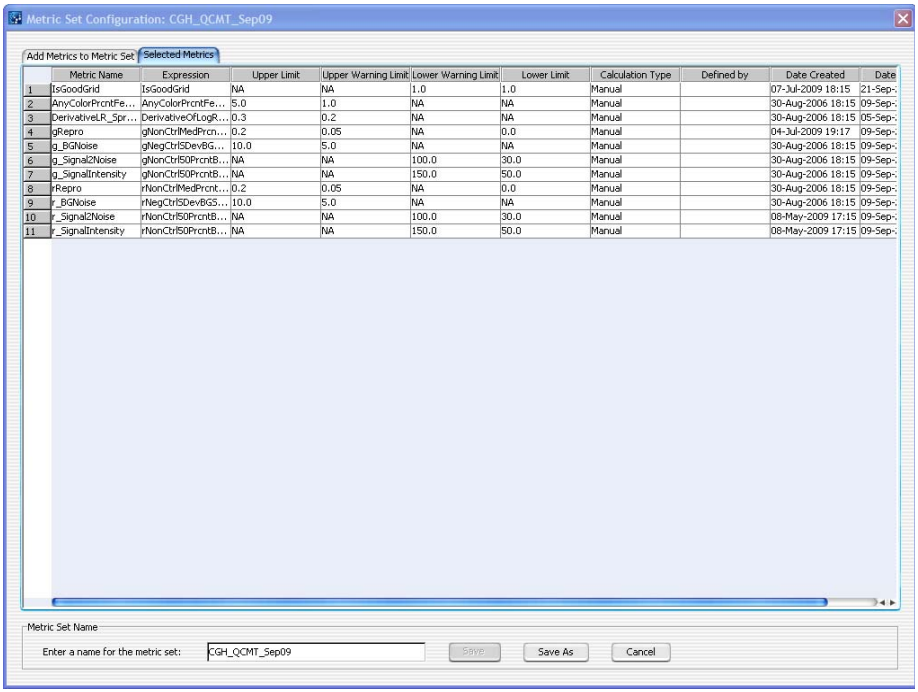


Figure 121 Metric Set Configuration dialog box - Existing Metrics tab.

Purpose: Used to display existing metrics in a metric set.

To open: In the Metric Set Configuration dialog box, click the **Existing Metrics** or **Selected Metrics** tab. See "Quality Tab Window" on page 87.

The Existing Metrics tab of the Metric Set Configuration dialog box has the following components and functionality:

Table

The Metric Sets created in or imported into Quality tools are displayed in the form of a table, or grid-view.

The column headers appear in this order:

- Metric Name
- Expression
- Upper Limit
- Upper Warning Limit
- Lower Warning Limit
- Lower Limit
- Calculation Type
- Defined By
- Date Created
- Date Modified

Metric Set Name Lets you enter a name to save the metric set.

Save Saves the existing metric set with the previously defined name.

Save As Saves the metric set with a newly defined name.

Cancel Closes the Metric Set Configuration dialog box without saving any changes.

Microarray Properties

Purpose: Displays the properties associated with an array. You can also edit the values of specific attributes. For information on how to add attributes to the list, see the *Setup and Data Review User Guide*.

To open: For any array in the Array Data folder in Sample Manager, right-click the array name, then click **Show Properties.** See "Content tab – Sample Manager window" on page 95.

Attribute tab

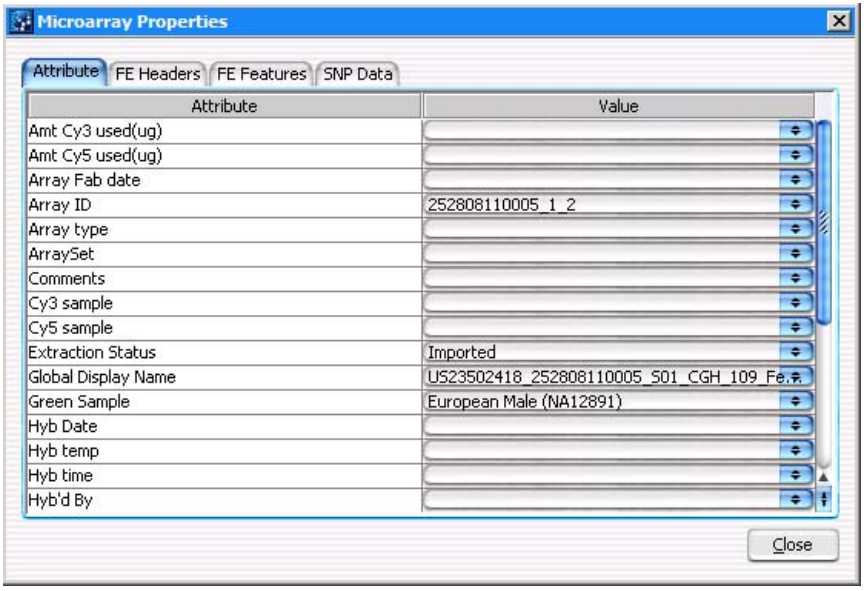


Figure 122 Microarray Properties dialog box with list of Attributes and their values

- **Attribute** – Displays the attributes in the array by name.
- **Value** – Indicates the values, if any, for each array. To edit the value of an attribute, select a new value for it under Value. Alternatively, click , then type or edit the value.

NOTE You cannot edit values for read-only arrays.

Close Closes the dialog box.

FE Headers tab

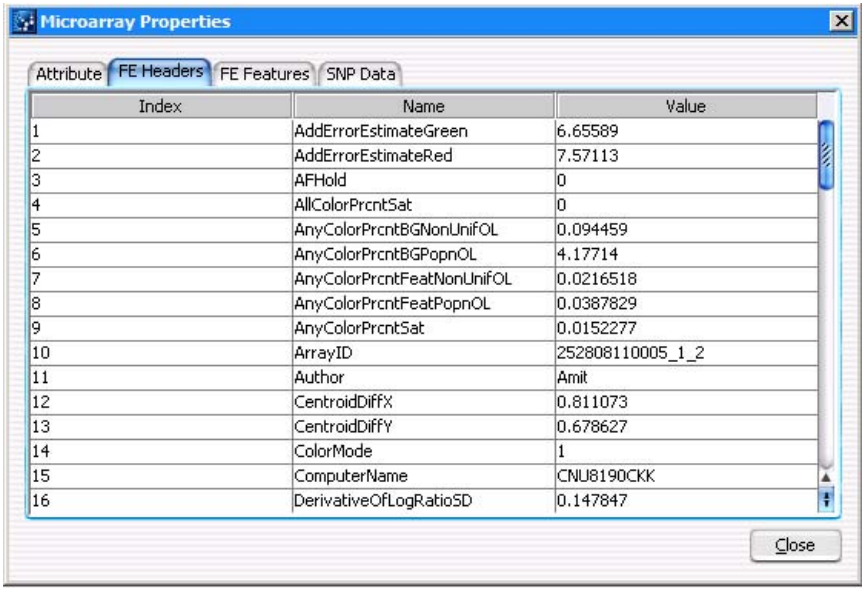


Figure 123 Microarray Properties dialog box with list of FE Headers their values

Index Displays a sequential index to help identify FE properties.

Name Displays feature parameters, statistics, and constants for the whole array.

Value Displays the value for each parameter, statistic, and constant.

Close Closes the dialog box.

FE Features tab

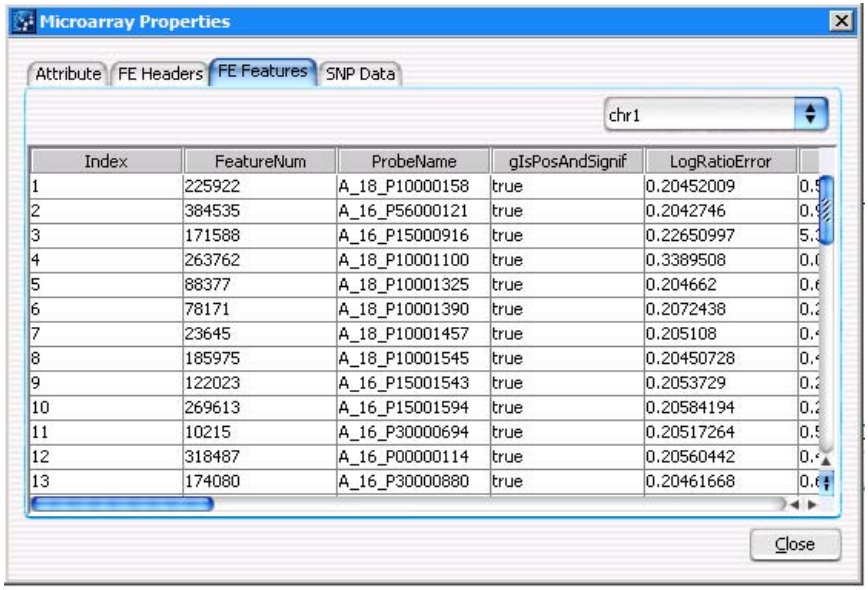


Figure 124 Microarray Properties dialog box with list of FE Features and associated data

Selection List Select the chromosome whose feature information you want to display.

List Box Displays FE features and the associated data.

SNP Data tab

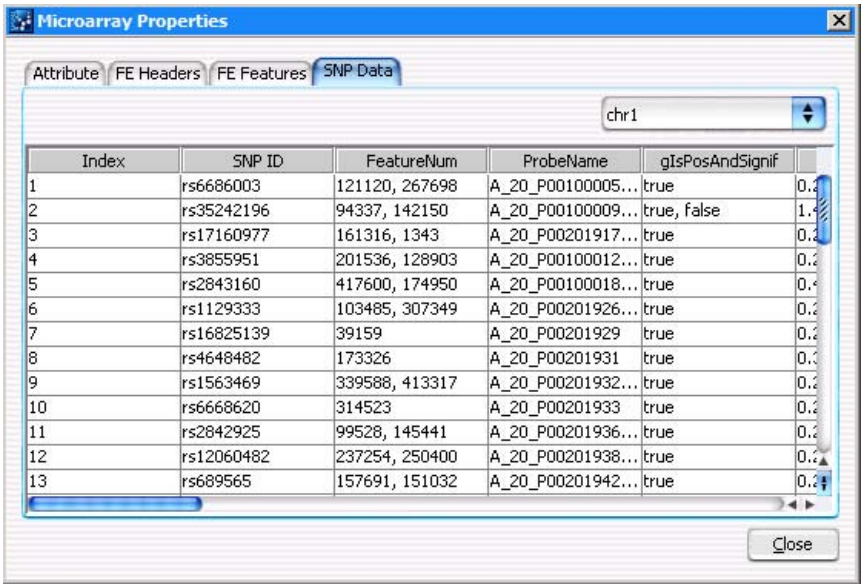


Figure 125 Microarray Properties dialog box with list of SNP data and values

Selection List Select the chromosome whose SNP information you want to display.

List Box Displays SNP data.

Notes

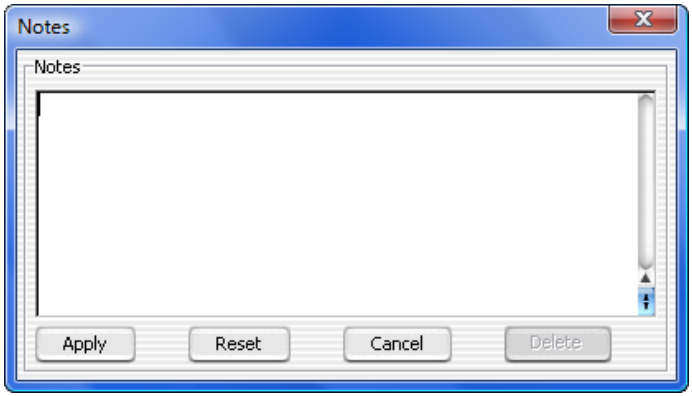


Figure 126 Notes dialog box

Purpose: Lets you type information about a selected aberration call.

To open: In Triage View, under Actions, click **Notes** for the aberration call for which you want to create notes. See "Triage View" on page 59.

Apply Saves the note and closes the dialog box.

Reset Clears the Notes text box.

Cancel Closes the dialog box without saving the note.

Delete Deletes a note once it has been saved.

Open

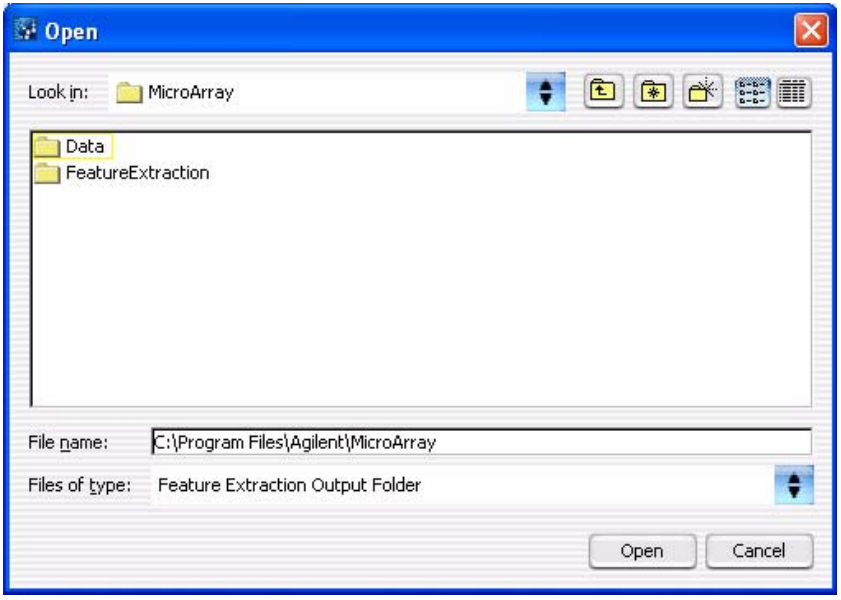


Figure 127 Open dialog box

Purpose: To select the files you want to open, or to identify the location where you want files to be stored.

To open: This dialog box appears at various places in the program where you need to select a file or identify a file output location.

Look in Select folder where the files are located. To browse for a location, click the arrow and browse to the desired folder.

File name Type the name for the file you want to open, or click the file to select it. To select multiple files to open, hold down the **Ctrl** key and click the files to open.

Files of type Expected file type is displayed. Click to change displayed file types.

Open Opens the selected file, or selects the output location.

Cancel Click this to cancel the operation.

QC Metrics – Frequency Distribution

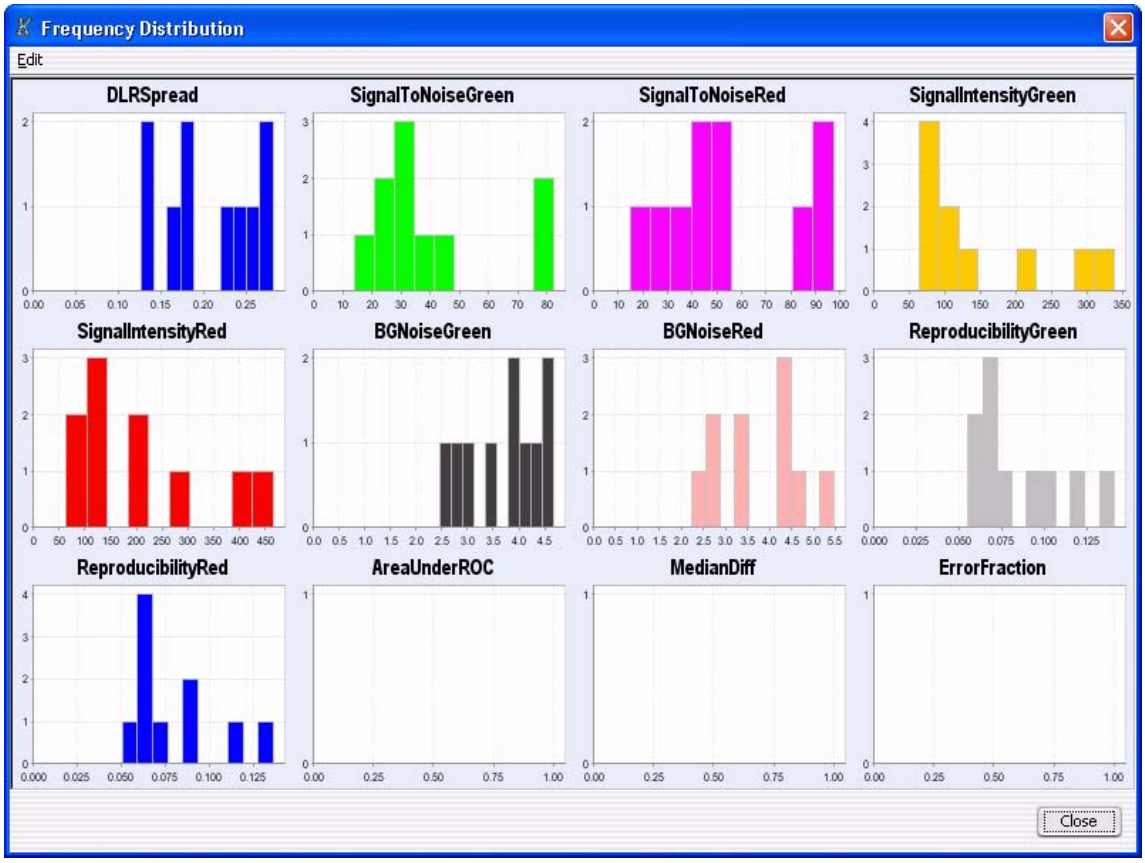


Figure 128 QC Metrics Frequency Distribution Plot

Purpose: The plots in this dialog box represent the selected columns in the QC Metrics Table. Each plot shows the number of arrays within each value range for a metric. See "QC Metrics Table" **on page 231**.

To open: Click **Show Frequency Distribution** in the QC Metrics Table. See "Quality Tab Window" on page 87.

Edit Opens a menu with a Copy command that copies the plots in the dialog box to the Clipboard. You can then paste the image into a document.

Close Closes the dialog box.

QC Metrics Graph

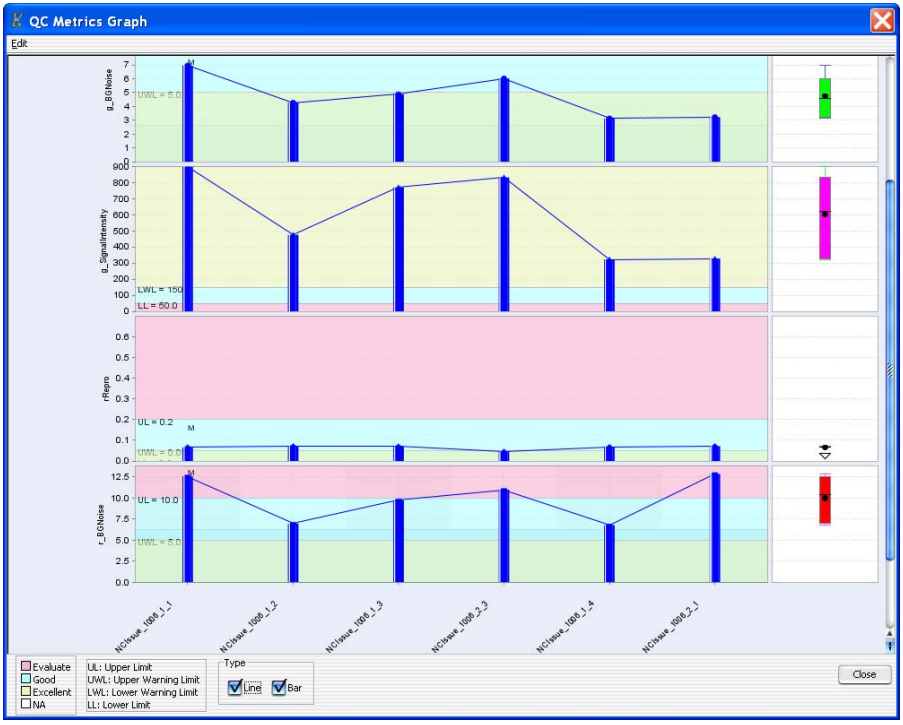


Figure 129 QC Metrics Graph

Purpose: To display plots of selected metric(s) for all arrays.

To open: Click **Plot** in the QC Metrics Table See "QC Metrics Table" on page 231.

Main Plots The plots in this dialog box represent the selected columns in the QC Metrics Table. Each plot shows the value of a given metric for all arrays.

These plots have several additional features:

- The background colors in each plot correspond to quality guidelines developed by Agilent, based on normal ranges observed for analyses of well-established cell lines using standard Agilent protocols. See the descriptions of each of these metrics in "QC Metrics Table" on page 231.
- A "Box & Whisker" plot appears to the right of the main plot for each metric. See "'Box & Whisker' Plot," below.
- The program can plot the data as a line graph, a bar graph, or both. See "Line" and "Bar," below.
- If you set your view preferences to show ToolTips (see "View Preferences" on page 251), a tool tip appears when you place the pointer over any bar. The ToolTip shows the value of each bar and the name of the corresponding array. See Figure 130.
- You can right-click any bar to open a shortcut menu for the corresponding array. The options in the shortcut menu let you set the QCMetricStatus attribute for the array. See Figure 130.
- The QCMetricStatus attribute for each array appears over all of the corresponding bars of the main plot. The four possible values for QCMetricStatus are: (P)ass, (F)ail, (M)arginal, and (N)A.

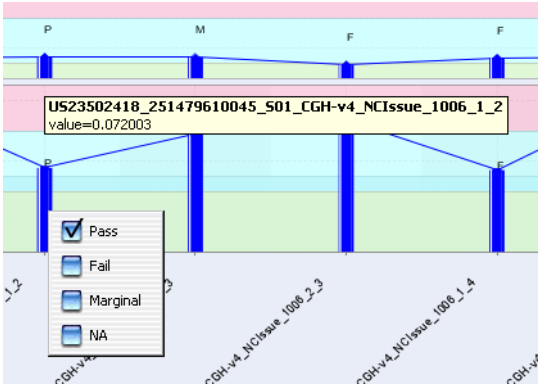


Figure 130 Portion of the QC Metrics Graph, showing a ToolTip (values in white box), and a shortcut menu. Use the shortcut menu to set the QCMetricStatus attribute for the array. Right-click any bar to open the shortcut menu for the corresponding array.

"Box & Whisker" Plots

A small plot appears to the right of each of the main plots. It represents the overall distribution of values for the metric. Two examples appear in Figure 131. The symbols have the following meanings:

- The lower and upper edges of the box represent the 25th and 75th percentiles, respectively.
- The black horizontal line in the box is the median.
- The black dot is the mean.
- The "whiskers" represent the range of values that are not outliers. An outlier is a point that is out of the 25th to 75th percentile range by more than 150%.
- Open circles represent outliers, and an open triangle represents outliers that plot beyond the available space on the graph.

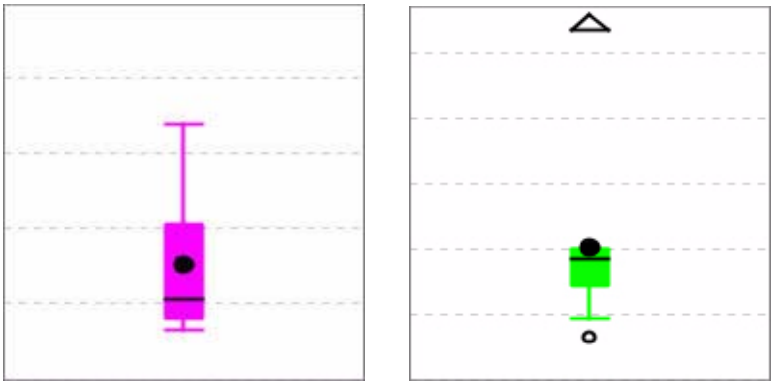


Figure 131 Two "Box & Whisker" plots

Line Select this option to display each metric as a line graph.

Bar Select this option to display each metric as a bar graph.

Edit Opens a menu with a Copy command. This command copies the plots in the dialog box to the Clipboard as an image. You can then paste the image into a document in another program.

Close Closes the dialog box.

QC Metrics Table

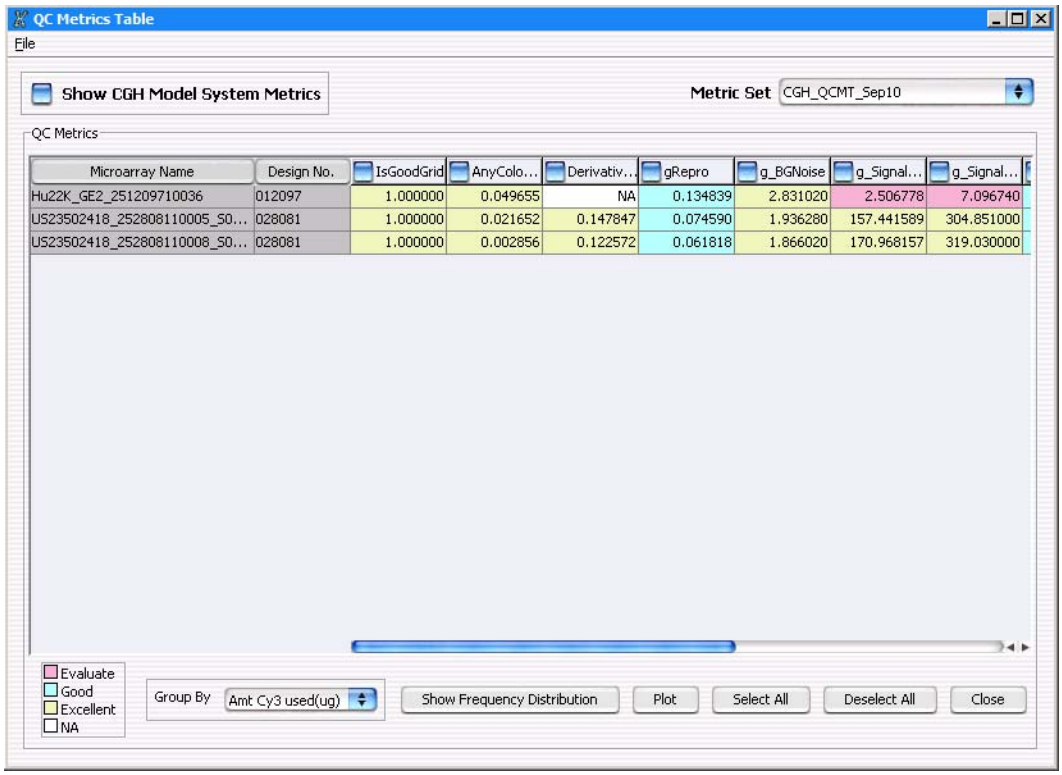


Figure 132 QC Metrics Table

Purpose: The QC Metrics Table shows the available metrics for one or more arrays. With this table and the available plots, you can evaluate the quality of your microarray results, and assign a manual QC status to each microarray. Some metrics come from the Agilent FE program, while others are calculated by the CGH module itself. These metrics are available only for Agilent microarrays.

To open: In the Content tab, in the Navigator, click **Sample Manager**. In the Array Data folder, right-click a design build or a microarray, and click **QC Metrics**. See "Content tab – Sample Manager window" on page 95.

File Lets you save the QC Metrics Table as a Microsoft Excel (*.xls) format file. When you click **File,** a menu opens with an Export command. This command opens a Save dialog box, where you can select a location and type a name for the exported file.

Show CGH Model System Metrics Click to include display of the CGH Model System Metrics.

Metric Set The default metric set for the selected module is displayed. Click the arrow to select a different metric set. Metric sets are created or edited from the Quality tab.

Table The values of the QC metrics for arrays appear under QC Metrics, one array per row. The table has many columns:

- **Microarray Name** – Displays the names of microarrays. Because you can open the QC Metrics Table in several ways, the list can contain an individually-selected microarray, or those associated with an experiment or with a design.
- **Design No.** – Identifies the Agilent design ID for each microarray.
- **Metrics** – The program evaluates each metric, and assigns it a rating of Excellent (yellow), Good (turquoise), Evaluate (pink), or NA (white). The name of each metric appears as a column heading. Select the check box next to the name of the metric to include it in the available plots. Drag the column heading of a metric horizontally to change its position in the table.
- **ManualQCFlag** – Lets you set the QCMetricStatus attribute of the array. Status can be Pass, Fail, Marginal, or NA. Later, you can filter arrays based on this attribute.

NOTE To see a list of the metrics included in any metric set, including the threshold limits, click the **Quality** tab. In the Metric Sets Navigator, double-click the name of the metric set. The details for the metric set appear in the Quality table, along with thresholds for each.

When you select **Show CGH Model System Metrics** the following metrics appear in the table, in addition to the metrics for the selected Metric Set:

Metric	Comments
Model System Metrics	The metrics below apply only to CGH model systems. Currently the only model system supported is a male (XY) CY5-labeled vs. female (XX) CY3-labeled comparison.
AreaUnderROC	Method: Sort the log ratios in ascending order for the entire array. Each log ratio in the data set comes from an X-probe or an autosome. If it is an X-probe, it contributes to the number of true positives (TP). If it is an autosome, it contributes to the number of false positives (FP). So for each log ratio, start from the lowest and continue incrementing either TP (if an X-probe) or FP (if an autosome). Then for each log ratio, plot FP/(total number of autosomes) vs. TP/(total number of X-probes). Use the trapezoidal rule to estimate the area under this curve. The result is this metric: AreaUnderROC. Excellent >0.95 Good 0.85 – 0.95 Evaluate < 0.85
MedianDiff	This metric is the difference between the medians of the histograms of X-probes and autosomes. Excellent > 0.9 Good 0.8 – 0.9 Evaluate < 0.8
ErrorFraction	The minimum value of all error fractions calculated. The error fraction is: ((FP/(total number of autosomes) + (1 – TP/(total number of X-probes))/2. FP is the number of false positives, and TP is the number of true positives. Excellent < 0.05 Good 0.05 – 0.1 Evaluate > 0.1

Group By The program displays the arrays in the table, and in the QC Metrics Graph, grouped by the array attribute you select here.

Show Frequency Distribution Opens the Frequency Distribution dialog box. This dialog box contains line plots of the distribution of each selected metric over the all of the arrays in the QC Metrics Table. See "QC Metrics – Frequency Distribution" on page 227.

Plot Opens the QC Metrics Graph dialog box. This dialog box contains plots of each selected metric for each array. See "QC Metrics Graph" on page 228.

Select All Selects the check boxes of all metrics.

Deselect All Clears the check boxes of all metrics.

Close Closes the QC Metrics Table.

Query Builder Dialog Box

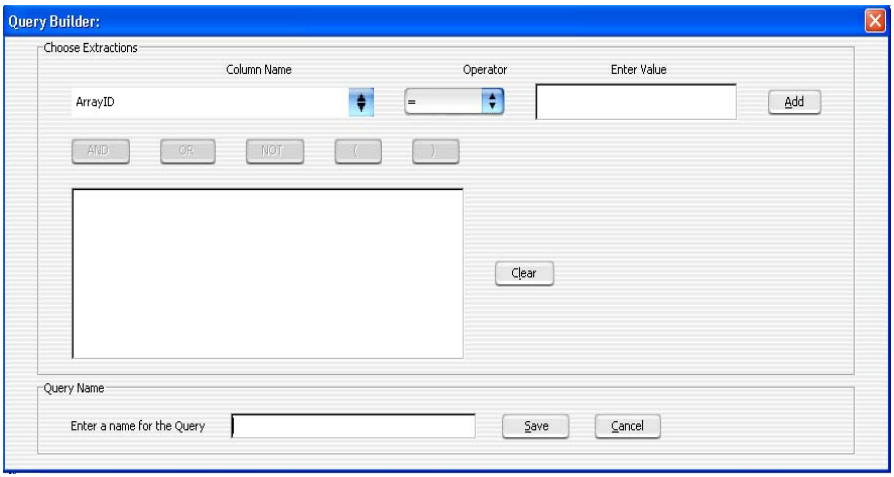


Figure 133 The Query Builder dialog box

Purpose: Used to define a subset of extractions for a representative data set, for use in metric and threshold development, and in producing Charts.

To open: In the Quality tab, click **Queries** and then click **New**. See "Quality Tab Window" on page 87.

The Query Builder dialog box has the following components and functionality:

Column Name Displays a list of the metrics that can be used to create an expression with which to query the database.

Operator Displays a list of the relational operators that can be used with the selected metric.

Enter Value A text area where a value can be compared to the metric for each extraction. If the chosen relation between the metric and value is valid, the query will produce extractions that pass that condition (for which the relation between the metric and value is true).

Operations The following buttons are logical operations used to link two or more metric-value relations built using the functions listed above. Each logical operation can link two metric-value relations at a time.

AND Produces a complex query which is true only if *both* metric-value relations are true.

OR Produces a complex query which is true if *either* metric-value relations are true.

NOT Produces a complex query which is true only if both metric-value relations are *not* true.

( ) Lets you create a subset for and prioritize the complex query.

Text area The area in which complex relations using the logical operations are listed for review.

Query Name A text area for entering a name under which to save the query.

Save Saves the query using the name specified in the Query Name field.

Cancel Cancels all query operations and closes the Query Builder dialog box.

Scroll to Column

Figure 134 Scroll to Column dialog box

Purpose: Lets you select a “Scroll to” column. The program then scrolls the tab so that you can see the selected column.

To open: Right-click a column heading in Tab View, then click **Scroll To Column** in the shortcut menu. See "Tab View for Genomic Viewer" on page 50.

Select column Displays the columns available in the selected tab. Select the one you want to display.

OK Scrolls the current tab so that you can see the selected column.

Cancel Closes the dialog box.

Select Color

Purpose: To select a color. Three tabs are available for selecting colors:

- Swatches tab - select colors based on samples (swatches)
- HSB tab - select colors based on an HSB schema (Hue, Saturation, and Brightness)
- RGB tab - select colors based on an RGB schema (Red-Green-Blue)

To open: This dialog box opens when a function allows you to change a color. For example, right-click an array in the Genomic Viewer, click **Edit Array Color** and click the **Swatches, HSB,** or **RGB** tab.

Swatches tab

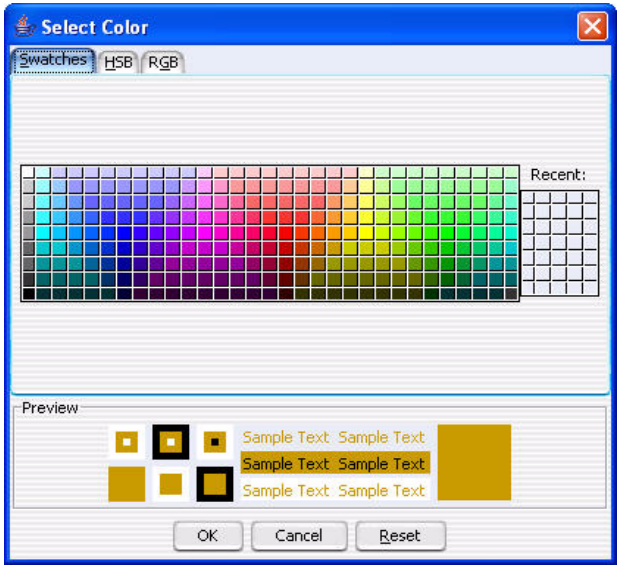


Figure 135 Select Color - Swatches tab

This tab is used to select a color based on color samples (swatches).

Preview The Preview area shows how the selected color appears. When you change the color, the original color appears at the top of the color box on the right.

Recent: Click to choose a recent color selection.

OK Click to select the color and close the dialog box.

Cancel Click to close the dialog box without changing the color.

Reset Click to change swatches, HSB, and RGB colors back to the default colors.

HSB tab

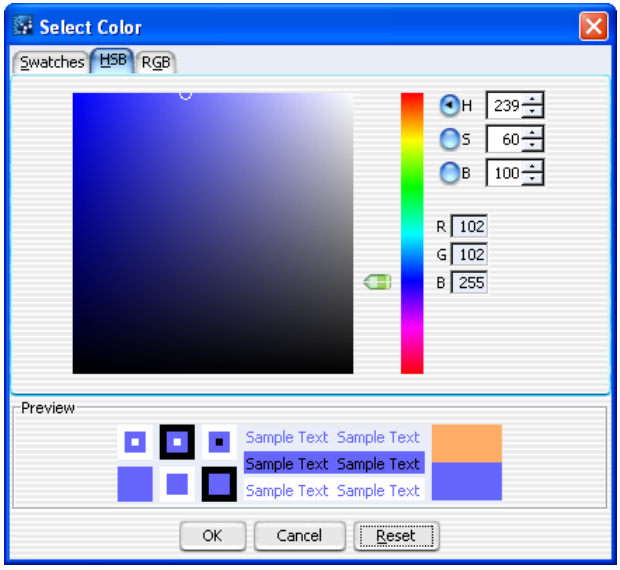


Figure 136 Select Color - HSB tab

In this tab, you can select a color based on an HSB schema (Hue, Saturation, and Brightness).

Hue Click the **H** button, and move the slider up and down, or go up and down the list of numbers, to select the hue or color of the array.

Saturation Click the **S** button, and move the slider up and down, or go up and down the list of numbers, to select the saturation level for the color.

Brightness Click the **B** button and move the slider up and down, or go up and down the list of numbers, to select the brightness level for the color.

RGB Numbers Reflect the amount of red, green and blue in the resulting color.

Preview The Preview area shows how the selected color appears. When you change the color, the original color appears at the top of the color box on the right.

OK Click to select the color and close the dialog box.

Cancel Click to close the dialog box without changing the color.

Reset Click to change the swatches, HSB, and RGB colors back to default values.

RGB tab

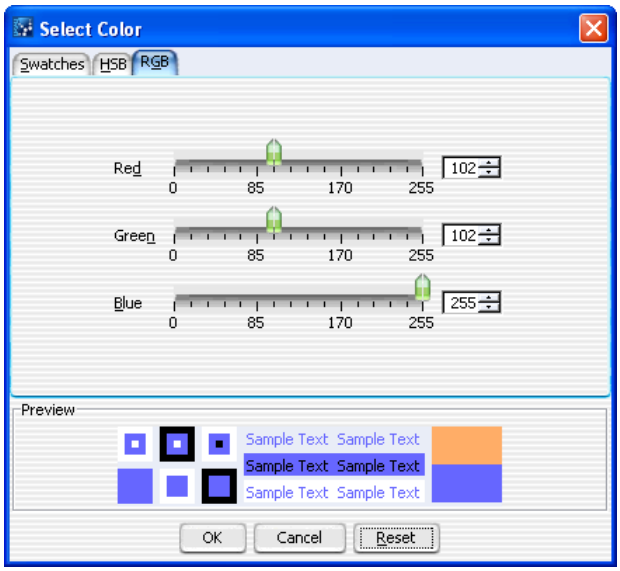


Figure 137 Select Color - RGB tab

This tab is used to select a color based on an RGB schema.

Red Move the slider to change the amount of red in the color. Or, click the up or down arrow to select a number.

Green Move the slider to change the amount of green in the color. Or, click the up or down arrow to select a number.

Blue Move the slider to change the amount of blue in the color. Or, click the up or down arrow to select a number.

Preview The Preview area shows how the selected color appears. When you change the color, the original color appears at the top of the color box on the right.

OK Click to select the color and close the dialog box.

Cancel Click to close the dialog box without changing the color.

Reset Click to return the swatches, HSB, and RGB colors back to default values

Select Report Name

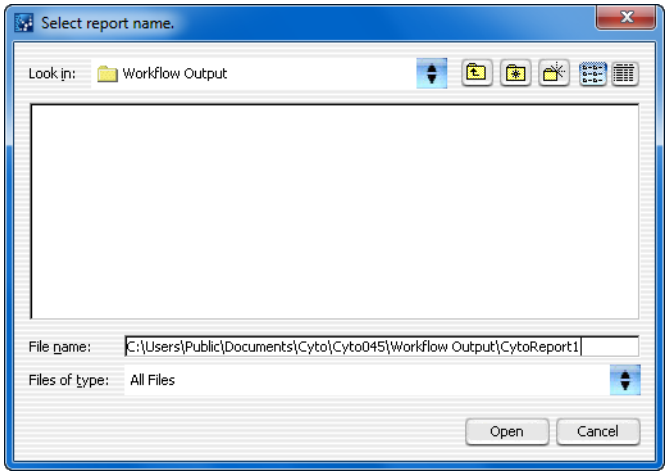


Figure 138 Select report name dialog box

Purpose: Used to select the folder location and name to store a report.

To open: From Cyto Report Parameter Panel in the Workflow window, click **Browse**. See "Cyto Report Parameter Panel" on page 119.

Click the arrow and select the folder, or browse to the location where you want to save the report file.

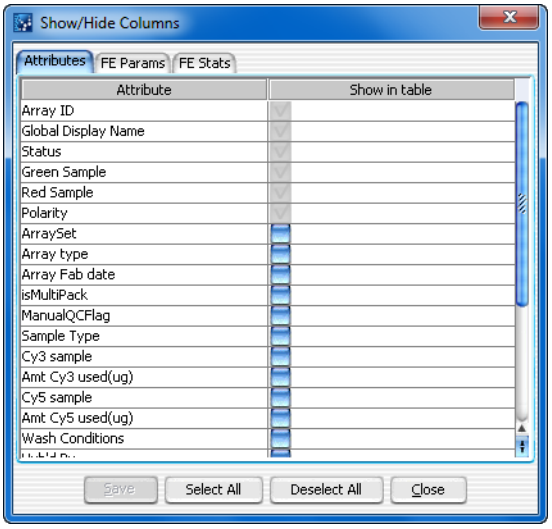
File name Type the name of the file you wish to save. Or, click to select the file from the displayed files.

Files of type Shows the type of files displayed in the window. Click the arrow to change the type of files displayed.

Open Click to save the report file.

Cancel Click to cancel the operation.

Show/Hide Columns



Show/Hide Columns dialog box

Purpose: Used to select the attributes to be displayed in the Sample Manager table and other views in the program. The Sample Manager table is available in the Content tab, and also appears in the Analysis tab when you click **Review**.

To open: This dialog box appears when you click **Show/Hide Attributes** at the bottom of the Sample Manager table in the Content tab. See "Content Tab Window" on page 95.

Attributes tab All available attributes are shown in the Attribute column. Attributes with a check-mark next to them will be displayed in the Sample Attributes and Sample Manager tab. To select an attribute for display, select the **Show in Table** box next to it. To hide an attribute, clear the **Show in Table** box again.

FE Params tab All available Feature Extraction parameters are shown in the FE Param column. Click to select the FE parameters you want to show in the Sample Manager table. Clear the check mark to hide the parameter in the Sample Manager table.

FE Stats tab All available Feature Extraction statistics are shown in the FE Stat column. Click to select the FE statistics you want to show in the Sample Manager table. Clear the check mark to hide the statistic in the Sample Manager table.

Save Saves the current list of selected attributes and updates the Sample Manager table, based on the selections.

Select All Selects all the attributes in the list.

Deselect All Clears all check selects from attributes in the list.

Close Closes the dialog box. If changes were made, the program asks if you want to save your changes before closing.

Show Intensity Bar Charts

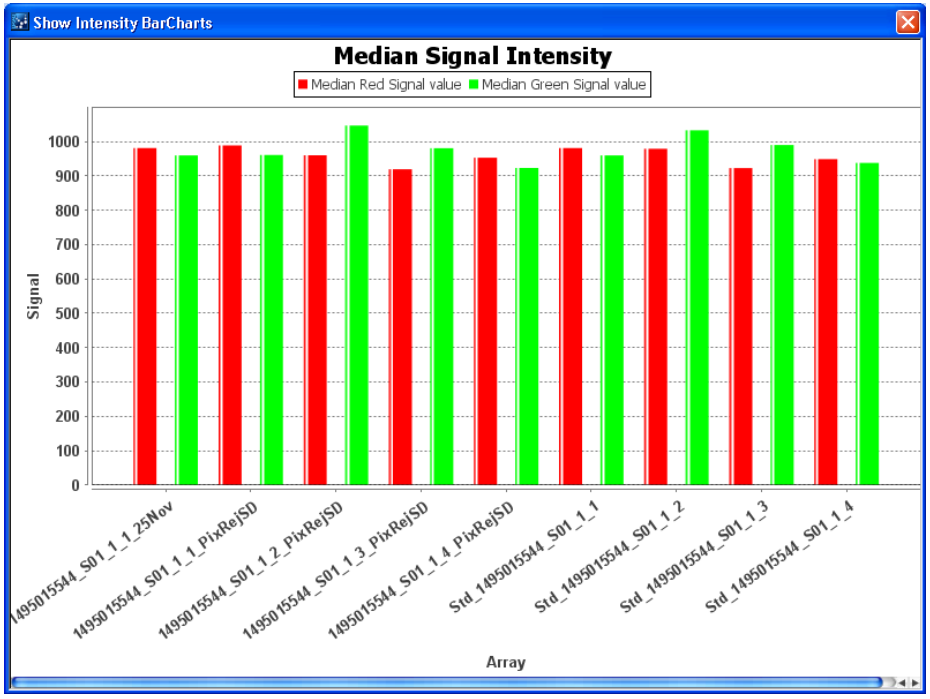


Figure 139 Show Intensity Bar Charts display

Purpose: This window displays a signal intensity bar chart.

To open: This display appears when you create a signal intensity bar chart from Gene View. See "Gene View" on page 45.

Right-click the chart and select from the following options.

Copy Click to copy the bar chart to the Clipboard. You can then paste the bar chart image into another application that supports the Clipboard.

Save as Click to open the **Save as** dialog box that lets you save the bar chart in a file on your hard disk.

SNP CN QC Metrics Table

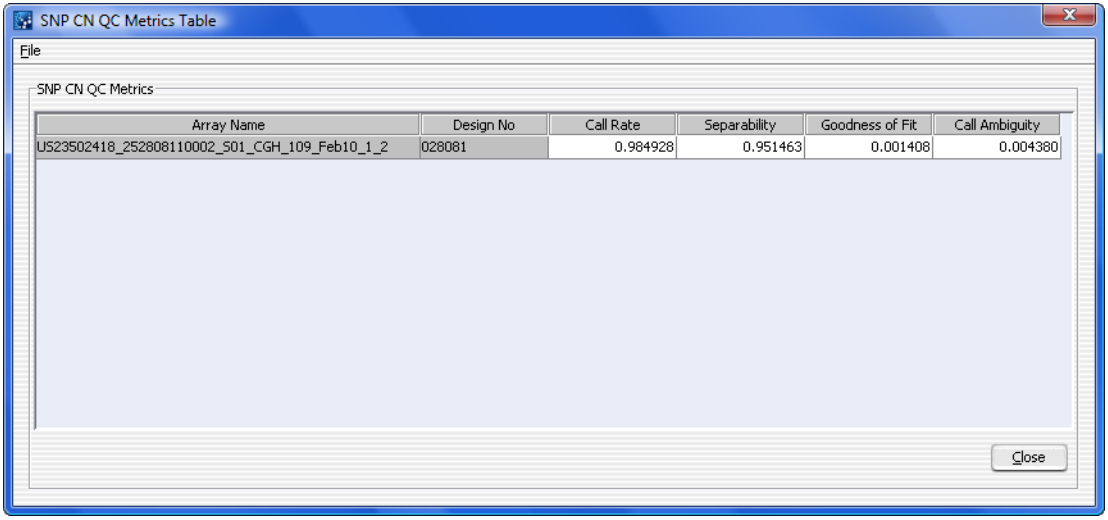


Figure 140 SNP CN QC Metrics Table

Purpose: To display calculated metrics for a CGH+SNP sample in Triage View.

To open: In the command ribbon for Triage View, under SNP CN, click **QC Metrics**. See "Triage View" on page 59.

File Opens the Export command, where you can choose to export the displayed metrics table to a file on your hard disk.

Array Name (Read only) The name of the microarray.

Design No (Read only) The microarray design number.

Call Rate Call rate is the number of SNPs that receive a genotype call at a 95% threshold divided by the total number of SNPs for which there is signal in the reference. For high-quality DNA samples hybridized against a genotyped and qualified reference, this value should exceed 90%.

Separability The separation between the CN peaks. It is defined as the distance between the average log ratio of SNPs with CN 1 and 2, which is a measure of the log ratio compression. For high-quality DNA samples this value should exceed 0.85. Lower values can indicate that the hybridization or wash stringency was too low, or the hybridization time was too short.

Goodness of Fit The error in the Gaussian fit for the peaks. It is obtained from the difference between the observed log ratio distributions and the modeled Gaussian distributions. It is a measure of the quality of the curve fitting model. For high-quality DNA samples this value should not exceed 2%.

Call Ambiguity This measures the overlap between the CN 1 and 2 Gaussian peaks. If a probe falls into this region, the CN call for that probe will not be made with high confidence. In other words, it is a measure of the degree of ambiguity of copy number calls. For high-quality DNA samples this value should not exceed 2%.

Close Closes the dialog box.

Summary

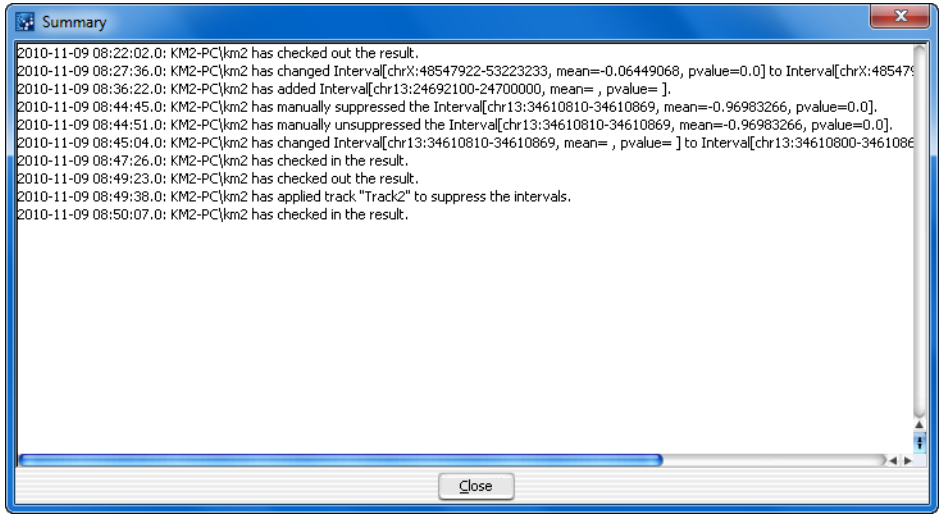


Figure 141 Summary dialog box

Purpose: Used to show audit trail for the sample.

To open: In Triage View, click **Audit Summary**. See "Triage View" on page 59.

Close Closes the Summary dialog box.

User Preferences

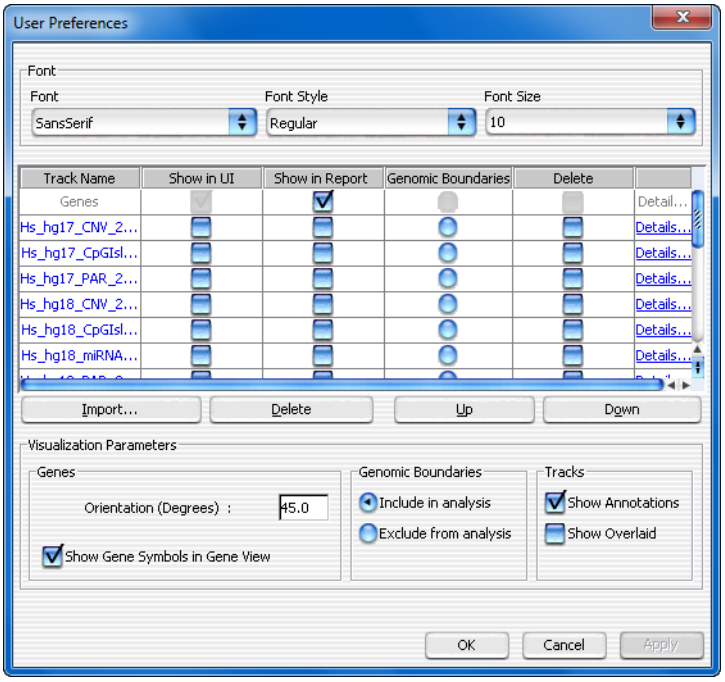


Figure 142 User Preferences dialog box

Purpose: To import and set up the appearance of tracks next to the Gene View. Tracks are additional graphic displays of genomic information loaded from an external file. They align with genomic coordinates in Gene View.

To open: In the Genomic Viewer, right click in the Genome, Chromosome, or Gene View, and click **User Preferences**. Or, in the Config tab, under Settings, click **Preferences**. See "Genomic Viewer" on page 40 or "Config Tab Windows" on page 64.

Font Options

Select the font type, style and size for the gene annotations that appear in the selected tracks.

Tracks List

Track Name Name of the track already loaded or imported

Show in UI Select the check box to display the track next to Gene View.

Show in Report Select the check box to display the track information in all the reports.

Genomic Boundaries Click to use the track to define only the regions that aberration detection algorithms will run. You can choose to do this for only one track.

Delete Select the check box to delete the track from the list. Then, click **Delete** to delete the track from the list.

Details Click to display all the chromosome locations defined in the track.

Import Click to import new tracks.

Delete Click to delete the tracks selected in the Delete column.

Up Click to move a track up the list.

Down Click to move a track down the list.

Visualization Parameters

Genes These options affect the appearance of the Track and Gene View.

- Orientation – Type a number to set the angle at which the Gene Symbols appear in Gene View and the Track Annotations appear in the tracks.
- Show Gene Symbols – Select to show gene symbols in Gene View, and clear the check box to hide them.

Genomic Boundaries These options let you include or exclude the Genomic Boundaries from the analysis.

Tracks These options affect the appearance of the Track Views.

- Show Annotations – Select to show the names of the gene regions for the tracks, and clear to hide them.
- Show Overlaid – Select to overlay all the tracks that appear next to Gene View, and clear the check box to display the information in separate tracks.

OK Accepts any changes and closes the dialog box.

Cancel Cancels all changes and closes the dialog box.

Apply Applies any changes to the preferences.

View coordinates in UCSC browser

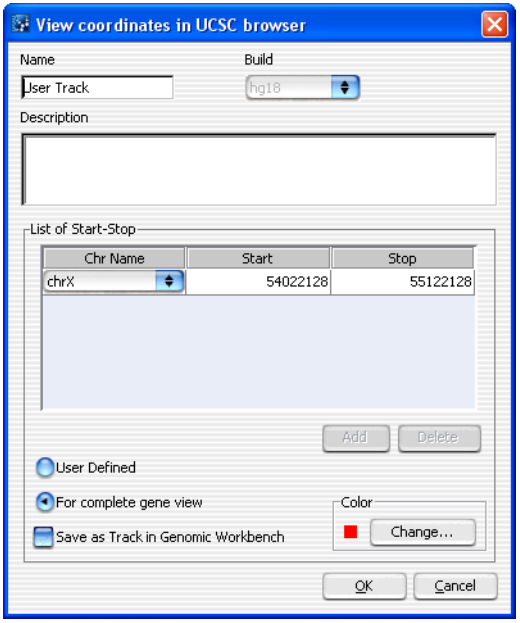


Figure 143 View coordinates in UCSC browser

Purpose: Defines a track to upload to the UCSC Web site so that you can see the information in the UCSC Genome Browser.

To open: Right-click in the Gene View, and select **Show in UCSC**. See "Gene View" on page 45.

Name Type a name for the track. This name identifies the track when it appears in lists and displays.

Build (Available if you select **User Defined** in **Set Chromosome Start-Stop.**) Select the genome build with which to associate the track.

Description Type descriptive text to attach to the track for reference.

Set Chromosome Start-Stop This parameter defines the region of the chromosome for which the track will be defined. Select one of these options:

- **User Defined** – Lets you define an arbitrary region of any chromosome. If you select this option, select the desired chromosome in **Chromosome,** then type the beginning **(Start)** and end **(Stop)** locations of the desired interval.
- **For complete gene view** – The chromosomal region that appears in Gene View.

Save as Track Select the check box to save this track. The track appears in the tracks list in the User Preferences. See "User Preferences" on page 247.

Change Click to open the Choose Track Color dialog box to select the color to use for display of the track in the Tracks folder. See "Select Color" on page 237.

OK Creates the track and opens the UCSC Web site, where you can display the track and associated information. For information on using the UCSC Web site, see the help and information provided there.

Cancel Closes the dialog box without creating a track.

View Preferences

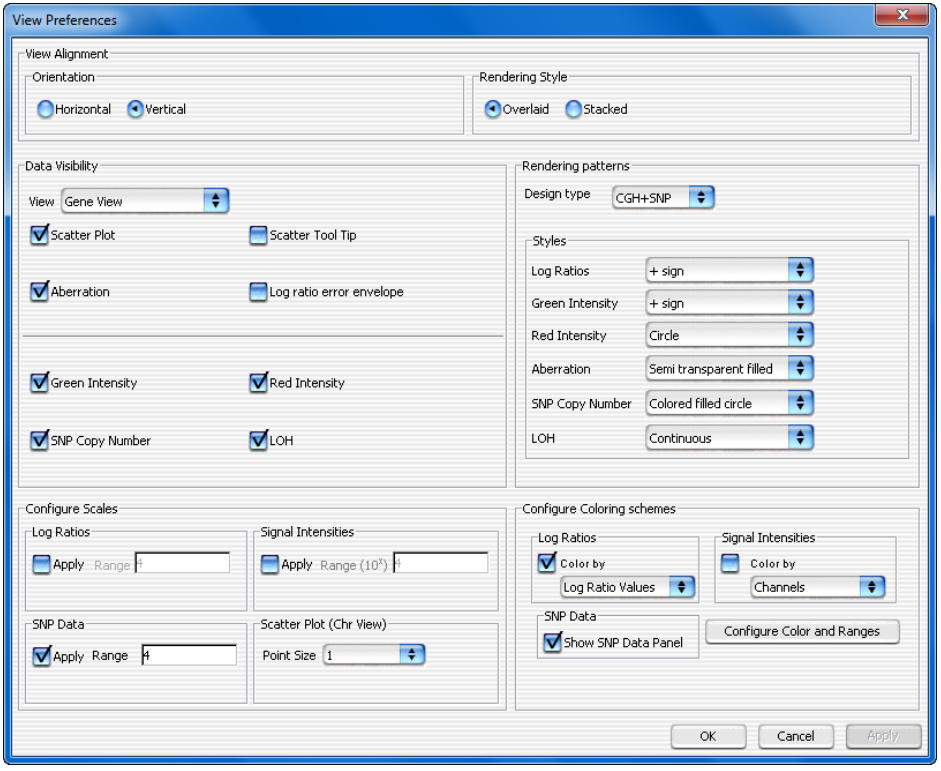


Figure 144 View Preferences dialog box

Purpose: This dialog box allows you to configure how data and results appear in Genome, Chromosome, and Gene Views of the View Aberrations or Triage windows.

To open: In the View Aberrations or Triage window, right-click in any of the views and select **View Preferences**. Or, in the Config tab, under Settings, click **View Settings**. See "Triage View" on page 59 or "Config Tab Windows" on page 64.

View Alignment Selects the orientation and rendering style (described below).

Option	Description
Orientation	
Horizontal	Stacks Genome, Chromosome, and Gene Views horizontally in the main program window. Genomic locations appear across the bottom of each view.
Vertical	Displays Genome, Chromosome, and Gene Views from left to right as side-by-side panes in the main program window.
Rendering Style	
Overlaid	In Chromosome View and in Gene View, displays data and results as a single, combined pane for all arrays. (Default)
Stacked	In Chromosome View and in Gene View, displays a separate pane for each array.

Data Visibility For each view, or all views, selects the kind(s) of data and results to display.

In **View,** select the view you want to configure. To set availability of display items for all views, select **All views.** Some display items are only available for certain views and modules. When you select a display item, it enables the item for display; for some items, you must take additional steps to display them. For example, you may need to configure a specific algorithm in the toolbar.

Select any of the following options, as available:

Option	Description/Comments
Scatter Plot	The plot(s) of individual log ratio, intensity, or probe score data points.
Scatter Tool Tip	The ToolTips that appear when you place the pointer over specific data points on the scatter plot(s) in Gene View. The tool tip shows the array of origin and the numerical log ratio value for the data point.
Moving Average	The result of the Moving Average algorithm.
Aberration	The result of the selected aberration detection algorithm.
Green Intensity	Select the check box to display green raw signal intensity.
Red Intensity	Select the check box to display red raw signal intensity.

Rendering Patterns

These options control the specific appearance of data and results in Genome, Chromosome, and Gene Views. You configure these options separately for each type of application design.

- **Design Type** – Select the application design type for which you want to define rendering patterns.
- **Styles** – Select the display style for each of these elements:

Option	Description/Comments
Log Ratios	Select the symbol used for log ratio data points in the scatter plots in Chromosome and Gene Views.
Green Intensity	Select the symbol to use for display of the green raw signal intensity.
Red Intensity	Select the symbol to use for display of the red raw signal intensity.
Aberration	Select the rendering style for detected aberrations. • **Semi transparent filled** – Solid, colored regions (in the display colors defined for each array, if applicable). • **Hatched** – Cross-hatched colored lines (in the display colors defined for each array, if applicable). • **Do not show area** – Aberrations do not appear.
SNP Copy Number	Select the symbol to use for showing SNP Copy Number.
LOH	The only selection for showing regions of LOH is "continuous".

Scatter Plot (Chr View) Point Size

Select a point size to use for display of scatter plot data points in the Chromosome View.

NOTE

Rendering scatter plots for more than 10 high density arrays in the Chromosome View may take significant time. Selecting filled circles as the rendering style for CGH scatter plots can also decrease performance. For faster performance, change the rendering style for CGH data from the filled circle to the plus (+) or cross hair sign.

Configure Scales

For Log Ratios or Signal Intensities plots, select **Apply** to enable the custom scale. In Range, type the value to use as the range for the scatter plot.

Configure Coloring schemes

Use these options to change the display of the scatter plot in the Gene View. These options are the same as those displayed in the Scatter Plot box in the Gene View.

To do this	Follow these steps
Show or hide the log ratio values in the Log Ratios plot	• To show the data points - Select the **Log Ratios** check box and select **Log Ratio Values** from the list. • To hide all data points - Clear the **Log Ratios** check box.
Show or hide LogRatios color-coded by Probe Score Values in the Log Ratios plot	• To show the data points - Select the **Log Ratios** check box and select **Probe Score Values** from the list. • To hide the data points - Clear the **Log Ratios** check box.
Show or hide Intensity values in the Signal Intensities plot	• To show the data points - Select the **Signal Intensities** check box and select **Intensity Values** from the list. • To hide all data points - Clear the **Signal Intensities** check box.
Show or hide Signal Intensities color-coded by Channels in the Signal Intensities plot	• To show the data points - Select the **Signal Intensities** check box and select **Channels** from the list. • To hide the data points- Clear the **Signal Intensities** check box.
Show or hide Signal Intensities color-coded by Probe Score values in the Signal Intensities plot	• To show the data points - Select the **Signal Intensities** check box and select **Probe Score Values** from the list. • To hide the data points- Clear the **Signal Intensities** check box.
Show or hide SNP data panel	• To show the SNP data panel - Select the **Show SNP Data Panel** check box. • To hide the SNP data panel - Clear the **Show SNP Data Panel** check box.
Change the ranges and colors for scatter plot and signal intensities panels	• Click **Configure Color and Ranges** to enter ranges and change colors. See "Configure Coloring Ranges and Shades" on page 152 for more information.

Show Memory Monitor in Status Bar

Displays a memory usage monitor in the eighth cell of the status bar. For information about the Status Bar, see "Status Bar" on page 53.

OK

Applies the changes you made to all preferences and closes the dialog box.

Cancel Closes the dialog box without applying changes.

Apply Applies changes without closing the dialog box.

View Report

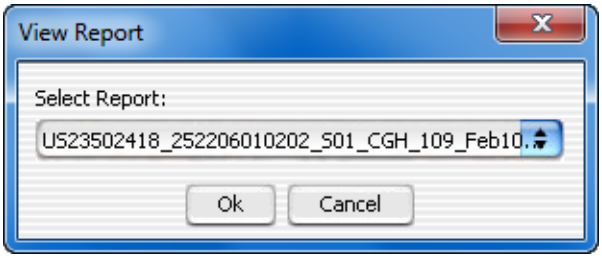


Figure 145 View Report dialog box

Purpose: To select a cyto report to view for the selected sample.

To open: In the Analysis tab, select **Report**. Search for and select a sample for which you want to show the report. Click **View Report**. See "Report window" on page 33.

Select Report Click the arrow and select the report you want to show for the sample.

OK Opens the Manage Cyto Report dialog box with the selected cyto report. See "Manage Cyto Report" on page 213.

Cancel Closes the dialog box without opening the report.

3 Reports

This chapter contains examples of the standard reports available in the Agilent CytoGenomics 1.0 program.

Examples of Agilent CytoGenomics Reports

The type of report generated during a workflow is selected when you create or edit a workflow. See "Workflow Window" on page 84. Available reports include:

- CGH Aberration Report
- Cyto Report (customizable)
- SNP Genotype Report (for samples analyzed with SNP algorithms)
- Aberration & LOH Report (for samples analyzed with SNP and LOH algorithms)

CGH Aberration Reports

Each report shown below also includes header information that contains the parameters used for data preprocessing and analysis. The following reports were generated using the same microarray and analysis parameters, with common intervals or regions highlighted for comparison, if applicable.

NOTE The *Stop* position listed in the report is the start of the last probe in that interval.

Probe Based

AberrationNo	CytoBand	ChrName	ProbeName	Start	Stop	Description	Genes		Logratio	Amplification	Deletion
1.188	p34.1	chr1	A_16_P30147946	45753224	45753275	Homo sapiens peroxi	PRDX1		1.1566998	1.0346224	
1.189	p33	chr1	A_16_P30155942	47678098	47678147	Homo sapiens forkhe	FOXD2		-1.7148536		-1.71485
1.19	p33	chr1	A_16_P30158058	51198005	51198049	Homo sapiens Fas (TN	FAF1		-1.890943		-1.89094
1.191	p32.3	chr1	A_16_P00064465	51583516	51583560	Unknown			-1.2975296		-1.29753
1.192	p32.3	chr1	A_16_P00068800	54861780	54861824	Homo sapiens acyl-Cc	ACOT11 ,FAM151A		1.1017393	1.1017393	
1.193	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown			-0.6178095		-0.67304
1.194	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown			-0.7367969		-0.67304
1.195	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown			-0.6244842		-0.67304
1.196	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown			-0.7957025		-0.67304
1.197	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown			-0.5904218		-0.67304
1.198	**p32.2**	**chr1**	**A_16_P00071447**	**56817487**	**56817534**	**Homo sapiens phospl**	**PPAP2B**		**-1.91642**		**-1.91642**
1.199	p31.1	chr1	A_16_P15174860	75698956	75699015	Homo sapiens solute	SLC44A5		-1.050432		-1.05043
1.2	p31.1	chr1	A_16_P00101977	78284071	78284130	Homo sapiens cDNA, FLJ97370.			-3.43538		-3.43538
1.201	p31.1	chr1	A_16_P00102728	78879706	78879765	Homo sapiens interfe	IFI44L		1.202869	1.202869	
1.202	p31.1	chr1	A_16_P30203405	82194201	82194260	Homo sapiens latroph	LPHN2		-1.1051049		-1.1051

Figure 146 CGH Aberration Report - Probe Based

Interval Based

AberrationNo	Chr	Cytoband	Start	Stop	#Probes	Amplification	Deletion	pval	Gene Names
36	chr1	p34.1	45045016	45045016	1	0	-2.29112	6.11E-44	TCTEX1D4
37	chr1	p34.1	45752746	45753224	2	1.034622	0	4.55E-19	PRDX1
38	chr1	p33	47678098	47678098	1	0	-1.71485	1.73E-25	FOXD2
39	chr1	p33	51198005	51198005	1	0	-1.89094	1.43E-30	FAF1
40	chr1	p32.3	51583516	51583516	1	0	-1.29753	2.19E-15	
41	chr1	p32.3	54861780	54861780	1	1.101739	0	1.38E-11	ACOT11, FAM151A
42	chr1	p32.3	55469748	55469748	5	0	-0.67304	4.58E-20	
43	**chr1**	**p32.2**	**56817487**	**56817487**	**1**	**0**	**-1.91642**	**2.40E-31**	**PPAP2B**
44	chr1	p31.1	75698956	75698956	1	0	-1.05043	1.09E-10	SLC44A5
45	chr1	p31.1	78284071	78284071	1	0	-3.43538	3.89E-96	
46	chr1	p31.1	78879706	78879706	1	1.202869	0	1.79E-13	IFI44L
47	chr1	p31.1	82194201	82194201	1	0	-1.10511	1.20E-11	LPHN2
48	chr1	p31.1	83664313	83712963	3	0.575142	0	8.77E-10	
49	chr1	p22.3	85235056	85235056	1	0	-0.99296	9.88E-10	MCOLN2

Figure 147 CGH Aberration Report - Interval Based with comparison interval highlighted

Probe and Interval Based

AberrationNo	CytoBand	ChrName	ProbeName	Start	Stop	Description	Genes	Logratio	Amplification	Deletion
1.192	p32.3	chr1	A_16_P00068800	54861780	54861824	Homo sapiens a	ACOT11 ,FAM151A	1.101739	1.1017393	
1.193	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown		-0.61781		-0.67304
1.194	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown		-0.7368		-0.67304
1.195	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown		-0.62448		-0.67304
1.196	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown		-0.7957		-0.67304
1.197	p32.3	chr1	A_16_P15126656	55469748	55469807	Unknown		-0.59042		-0.67304
1.198	**p32.2**	**chr1**	**A_16_P00071447**	**56817487**	**56817534**	**Homo sapiens p**	**PPAP2B**	**-1.91642**		**-1.91642**
1.199	p31.1	chr1	A_16_P15174860	75698956	75699015	Homo sapiens s	SLC44A5	-1.05043		-1.05043
1.2	p31.1	chr1	A_16_P00101977	78284071	78284130	Homo sapiens c	DNA, FLJ97370.	-3.43538		-3.43538
1.201	p31.1	chr1	A_16_P00102728	78879706	78879765	Homo sapiens i	IFI44L	1.202869	1.202869	
1.202	p31.1	chr1	A_16_P30203405	82194201	82194260	Homo sapiens l	LPHN2	-1.1051		-1.1051
1.203	p31.1	chr1	A_16_P15194313	83664313	83664372	Unknown		0.664638	0.5751422	
1.204	p31.1	chr1	A_16_P00109785	83695596	83695655	Unknown		0.403801	0.5751422	
1.205	p31.1	chr1	A_18_P10204499	83712963	83713022	Unknown		0.656988	0.5751422	
1.206	p22.3	chr1	A_16_P30206844	85235056	85235100	Homo sapiens r	MCOLN2	-0.99296		-0.99296
1.207	p22.3	chr1	A_16_P35209254	86173546	86173605	Homo sapiens c	COL24A1	1.004279	1.004279	

Figure 148 CGH Aberration Report - Probe and Interval Based with comparison region highlighted

Default Cyto Report

There are two default cyto report templates provided with the program. You can display the default cyto report template in the Config tab when you select **Report Template** and then **View/Edit** in the Default Cyto Report Template - CGH or Default Cyto Report Template - CGH+SNP row. If you make changes to a default report template, you must save the report template with a new name.

NOTE You must have a user role of Scientist or Administrator to display and edit report templates.

The default cyto report template creates .PDF and .XML files that include the following items.

- Header “Cyto Report Header”
- Footer “Cyto Report Footer” including right-aligned date and left-aligned page number.

- Sample information for the required sample attributes.
- Analysis settings at the end of the report
- One editable text box “Comments”
- Genome View
- Text Aberration Table View (all items selected, Annotations count of 3)
- SNP selections (for SNP template only) with all items selected and Annotations count of 3
- Chromosome View, showing all chromosomes and nested aberrant intervals
- No Gene View

SNP Reports

SNP Genotype Report

Index	ArrayName	ProbeID	SNP ID	Chr	SNP Position	Genotype	p-Val	Log Ratio
4	US23502418_25	A_20_P00100012, A_20_P00201	rs3855951	chr1	1794161	TC	0.001024	-0.82333
5	US23502418_25	A_20_P00100018, A_20_P00201	rs2843160	chr1	2298941	TT	0.001024	-4.57129
6	US23502418_25	A_20_P00201926, A_20_P00100	rs1129333	chr1	2325536	GG	0.001024	0.027611
7	US23502418_25	A_20_P00201929	rs16825139	chr1	2416458	NN	0.001024	-0.21117
8	US23502418_25	A_20_P00201931	rs4648482	chr1	2739780	TTT	0.001024	3.259384
9	US23502418_25	A_20_P00201932, A_20_P00100	rs1563469	chr1	2776007	GA	0.001024	-0.72524
10	US23502418_25	A_20_P00201933	rs6668620	chr1	2784397	NN	0.001024	-0.67233
11	US23502418_25	A_20_P00201936, A_20_P00100	rs2842925	chr1	2876218	NN	0.001024	0.107138
12	US23502418_25	A_20_P00201938, A_20_P00100	rs12060482	chr1	2960792	NN	0.001024	0.092238
13	US23502418_25	A_20_P00201942, A_20_P00100	rs689565	chr1	3153814	GGG	0.001024	0.799873
14	US23502418_25	A_20_P00100039	rs13374875	chr1	3190196	NN	0.001024	0.276925
15	US23502418_25	A_20_P00201947, A_20_P00100	rs12409315	chr1	3219119	NN	0.001024	0.17788
16	US23502418_25	A_20_P00201948	rs7516150	chr1	3243749	CCC	0.001024	2.675796
17	US23502418_25	A_20_P00201949	rs882430	chr1	3248975	NN	0.001024	1.891996
18	US23502418_25	A_20_P00100046, A_20_P00201	rs12085231	chr1	3287500	CA	0.001024	-0.96495
19	US23502418_25	A_20_P00100047	rs2483225	chr1	3314896	NN	0.001024	0.134196
20	US23502418_25	A_20_P00201954, A_20_P00100	rs6670123	chr1	3316656	CCC	0.001024	2.4159

Figure 149 SNP Genotype Report

SNP Aberration & LOH Report

This report is similar to the Text Aberration Probe Based Summary Report, except that cytobands are not shown, and the Amplifications and Deletions columns are combined into one column “Type” that indicates AMP or DEL. Log ratio values are listed in the AveCGHLR column for both aberration types.

Index	ArrayName	Class	Chr	Start	Stop	Type	#Probes	p-Val	AvgCGHLR	Gene Names
37	US23502418	CGH	1	45752746	45753224	AMP	2	4.55E-19	1.034622	PRDX1
38	US23502418	CGH	1	47678098	47678098	DEL	1	1.73E-25	-1.714854	FOXD2
39	US23502418	CGH	1	51198005	51198005	DEL	1	1.43E-30	-1.890943	FAF1
40	US23502418	CGH	1	51583516	51583516	DEL	1	2.19E-15	-1.29753	
41	US23502418	CGH	1	54861780	54861780	AMP	1	1.38E-11	1.101739	ACOT11, FAM151A
42	US23502418	CGH	1	55469748	55469748	DEL	5	4.58E-20	-0.673043	
43	**US23502418**	**CGH**	**1**	**56817487**	**56817487**	**DEL**	**1**	**2.40E-31**	**-1.91642**	**PPAP2B**
44	US23502418	CGH	1	75698956	75698956	DEL	1	1.09E-10	-1.050432	SLC44A5
45	US23502418	SNP	1	75999074	79472657	LOH	95	8.351616		ACADM, RABGGTB, SNORD45C,
46	US23502418	CGH	1	78284071	78284071	DEL	1	3.89E-96	-3.43538	
47	US23502418	CGH	1	78879706	78879706	AMP	1	1.79E-13	1.202869	IFI44L
48	US23502418	CGH	1	82194201	82194201	DEL	1	1.20E-11	-1.105105	LPHN2
49	US23502418	CGH	1	83664313	83712963	AMP	3	8.77E-10	0.575142	

Figure 150 SNP Aberration & LOH Report with comparison interval highlighted

4 Statistical Algorithms

This chapter describes the algorithms in Agilent CytoGenomics. The program uses many different algorithms to perform the statistical analyses that are needed for comparative genomic hybridization (CGH), copy number variation (CNV), and single nucleotide polymorphism (SNP) studies. Additional algorithms let you find common aberrant regions.

For an overview of the available algorithms, see "Overview of CGH Algorithms" on page 265. More detailed information on each algorithm follows later in the chapter. Algorithms are organized by general purpose. An appendix provides necessary background information for some of the more complex algorithms.

Overview of CGH Algorithms

Figure 151 summarizes the algorithms that are available in Agilent CytoGenomics, and their relationship to each other.

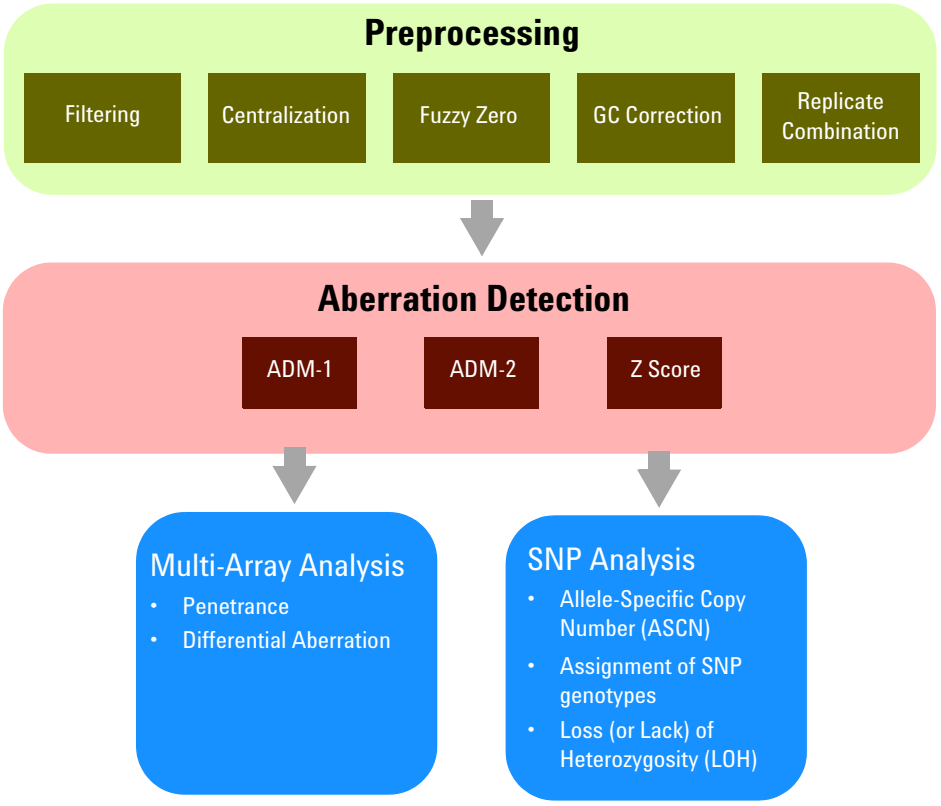


Figure 151 Overview of Agilent CytoGenomics algorithms. The relationship among the algorithms is shown.

The algorithms in Agilent CytoGenomics fall into several general categories. The list below describes these categories, and tells you where you can go in this section for brief descriptions of the algorithm(s) in each category. More detailed descriptions of the algorithms appear later in the chapter.

- **Preprocessing algorithms** prepare array data for downstream analyses. They include data correction and data centering algorithms, as well as algorithms that combine replicates and establish error models. For a brief description of these algorithms, see "Preprocessing Algorithms" on page 275.
- **Aberration detection algorithms** define the boundaries and magnitudes of regions of DNA loss or gain. For a brief description of these algorithms, see "Aberration detection algorithms" on page 267.
- **Algorithms for multi-array analysis** combine and evaluate aberrations across multiple samples. For a brief description of these algorithms, see "Algorithms for multi-array analysis" on page 270.
- **Algorithms for SNP analysis** analyze data from Agilent CGH+SNP microarrays, which combine CGH and SNP probes on the same array. For a brief description of these algorithms, see "SNP analysis algorithms" on page 270.

Preprocessing algorithms

Centralization Many statistical algorithms for aberration detection assume that log ratio values are centered around zero if no aberration occurs. This is a reasonable assumption if there is no difference between the reference and sample channels. However, for samples with a high aberration percentage, this assumption can lead to erroneous results as the measured center of the data can deviate from a log ratio value of zero. To re-center the data, the centralization algorithm finds a constant value to add to or subtract from all log ratio measurements. This ensures that the zero-point reflects the most-common-ploidy state. See "Centralization Algorithm" on page 275.

GC Correction A frequent observation in aCGH profiles is a "wavy" technical artifact that correlates with the local GC content of genomic regions. The exact cause of this artifact has not been ascertained, but its presence can interfere with detection algorithms and lead to inaccurate aberration, SNP, and LOH

calls. The GC correction algorithm corrects the log ratio value for each CGH and SNP probe on the array for GC-correlated "wavy" artifacts. The algorithm bases its corrections on values in the array design file that give the GC content of genomic segments that are centered on the locations of CGH probes. See "GC Correction Algorithm" on page 278.

Fuzzy Zero

ADM-1 and ADM-2 scores can identify extended aberrant segments with low absolute mean ratios, especially in data sets with a varying baseline. Often such aberrations represent noise, and are detected because of a high number of probes in the region. If long, low aberrations are detected in an analysis, you can apply the fuzzy zero algorithm to correct for the reliance on segment probe number. See "Fuzzy Zero" on page 281.

Replicate Combination

To combine replicate probes within an array (intra-array replicates) or among multiple arrays (inter-array replicates), the program calculates a weighted average of the log ratio values for the replicated probes. The weight given to each log ratio value is proportional to quality, which in turn is based on quality measures from the Agilent Feature Extraction (FE) program. Agilent CytoGenomics uses an error model that is based on the log ratio error values in the FE files. See "Error Model and Combining Replicates" on page 285.

Aberration detection algorithms

The aberration detection algorithms in Agilent CytoGenomics form the basis of CGH analysis. These algorithms detect DNA gain and loss events, show them in the Genomic Viewer, and make them available in reports. You can select from among many different aberration detection algorithms in Agilent CytoGenomics.

You can refine the analysis of CGH data with several options that are available in the program. Some functions estimate the global or local noise in the data. Others centralize or filter the data before the aberration regions are identified. Additional options let you filter out aberrations and do further analysis after aberrations are detected. The algorithms can be limited to known genomic regions.

A threshold can be applied to each algorithm. The optimal threshold depends upon several factors. These factors can include the DNA sample type, the DNA sample quality, or other measures, and can be explored interactively in the program.

NOTE A recommended threshold starting point for both the ADM-1 and ADM-2 algorithms is a value of 6.

The aberration detection algorithms cover a range of computational complexity and are best suited to different stages of analysis. For example, the Z-score algorithm is a straightforward test that can quickly identify regions with aberrant probes. The ADM-1 and ADM-2 algorithms are more sophisticated ways to identify aberrant regions.

Further, the output of the aberration detection algorithms can form the basis for further exploration. For example, the aberration calling algorithms ADM-1 and ADM-2 supply the input for the multi-array analysis algorithms.

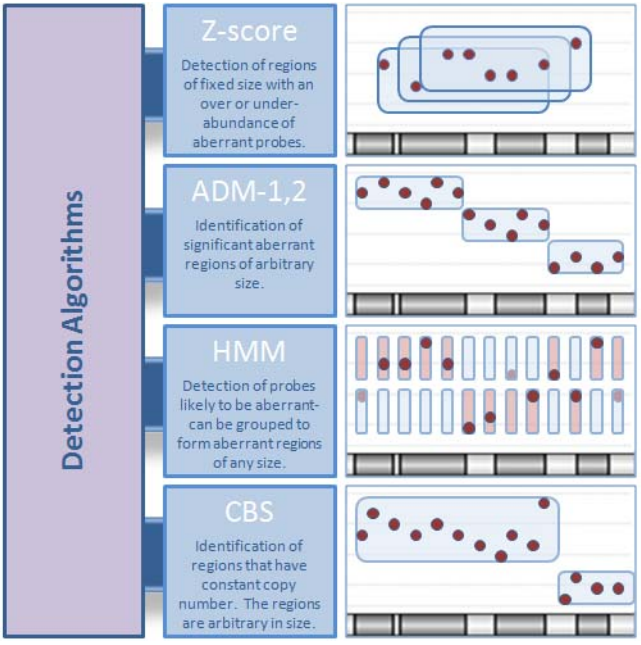


Figure 152 Overview of the aberration detection algorithms – HMM and CBS algorithms are not used in CytoGenomics.

Z-score The Z-score algorithm is a quick method of detecting aberrant regions. It finds genomic intervals with an over- or under-abundance of probes with log ratios that deviate significantly from baseline. To score intervals, the Z-score algorithm uses a sliding window of a fixed size that you set. Results from the Z-score identify regions of enriched probe log ratio change, and can suggest aberrant intervals. See "Z-Scoring for Aberrant Regions" on page 287. Because the Z-score algorithm only considers regions of fixed size in which there is enrichment of probe over- or under-abundance, it is not the preferred method for aberration identification. The ADM-1 or ADM-2 algorithms, described below, are recommended for such identification.

ADM-1 The Aberration Detection Method 1 (ADM-1) algorithm identifies all aberrant intervals in a given sample with consistently high or low log ratios based on a statistical score. The ADM algorithms automatically determine the optimal size of a statistically significant aberration. See "ADM-1" on page 290.

ADM-2 The Aberration Detection Method 2 (ADM-2) algorithm generates a similar statistical score to that produced by ADM-1 analysis, but ADM-2 incorporates quality information about each log ratio measurement. Use of the probe log ratio error in addition to the log ratio values makes ADM-2 more robust than ADM-1 when the data has noisy probes and you want to identify small aberrant regions. See "ADM-2" on page 293.

Derivative Log Ratio Spread (*dLRsd*) To make aberration calls, a detection algorithm needs a measure of probe-to-probe noise. A measure of the minimum log ratio difference is needed to make reliable amplification or deletion calls. The *dLRsd* algorithm is a robust method that estimates noise from the sample array alone. It calculates the spread of the log ratio differences between consecutive probes along all chromosomes. See "Noise Estimation – the Derivative Log Ratio Spread" on page 294.

Algorithms for multi-array analysis

Agilent CytoGenomics has algorithms that can combine the results of the aberration algorithms in biologically and statistically meaningful ways. These algorithms identify and summarize commonly aberrant regions within multiple samples, which can supply a robust report of aberration coverage.

Penetrance The probe-based and interval-based penetrance algorithms give a count score for probes and regions, respectively, which show aberration across multiple samples. The count is the percentage of samples which show an aberration in the same direction (over- or under-abundance). The penetrance algorithms use as input the aberrant regions found using any of the detection algorithms. See "Penetrance" on page 302.

Differential Aberration The algorithms to find genomic regions sharing common aberrations or having significant difference in aberrations across multiple samples use as input the list of per-sample aberrant regions as determined by either the ADM-1 or ADM-2 algorithm. The aberrant regions considered in the differential aberration algorithm are therefore subject to the user-defined ADM threshold. The probability that a given genomic interval has far more (or less) aberrations between two groups of samples is calculated in the differential aberration analysis. See "Differential Aberration Analysis" on page 304 for more information.

SNP analysis algorithms

Agilent CytoGenomics supports CGH+SNP arrays, which combine CGH probes and SNP probes on the same array. The inclusion of SNP probes allows the determination of allele-specific copy number for specific SNP sites (ASCN), the assignment of genotypes for specific SNP sites, and the detection of regions of loss or lack of heterozygosity (LOH).

The Agilent CGH+SNP platform uses restriction enzyme cleavage to differentiate between alleles at a given SNP site. At the SNP sites that can be detected by this method, one allele is cut by the enzymes, while the other is not. The method also uses a known genotyped reference. Since the reference genotype is known, the raw $\log_2$ ratios are "reference adjusted" to values that reflect a hypothetical ASCN of 2 for the reference at all SNP sites.

Allele-specific Copy Number (ASCN)

For each SNP site targeted on the array, the ASCN algorithm calculates an expectation value for the copy number of the uncut SNP allele. In Agilent CytoGenomics, this algorithm is known as the **SNP CN** (SNP Copy Number) algorithm.

A normal, diploid region of the genome typically contains a distribution of SNP genotypes with SNP copy numbers of 0, 1 and 2 uncut alleles. Three distinct ASCN states will appear in the SNP CN panel. However, aberrations can affect this distribution:

- In a diploid genomic region that comprises a copy-neutral LOH or UPD aberration, the SNP probes only report alleles that are homozygously cut and uncut (0 and 2 uncut copies). Only two states appear.
- For a region of the genome that is affected by a hemizygous deletion, two states also appear. These states represent SNPs that have only one copy of an allele, either cut or uncut (0 or 1 copy of the uncut allele).
- The amplification of a region or an entire chromosome adds one or more states. In this case, the SNP probes report four or more states that correspond to 0, 1, 2, 3 (or more) copies of the uncut allele.

Table 21 summarizes the relationship of genotype to SNP status.

Table 21 Relationship of genotype to SNP status (number of uncut alleles)

Genomic Status	Genotype	Number of uncut alleles
Normal diploid genome	AA, AB, BB	0, 1, 2
Diploid Genome with copy-neutral LOH or UPD	AA, BB	0, 2
Hemizygous deletion	A, B	0, 1
Amplification (*e.g.* trisomy)	AAA, AAB, ABB, BBB	0, 1, 2, 3

To calculate ASCN expectation values, the algorithm first adjusts the measured log ratios to account for the (known) ASCN of the reference sample. It finds the peaks in the probability density distribution of log ratio values, and assigns the most likely copy number to each peak. It then fits a separate Gaussian distribution to each peak. This creates a model that gives the likelihood that a given log ratio value corresponds to each possible copy number. See Figure 169 on page 315. The expectation

value that is reported for the ASCN of each SNP is the most likely copy number, given the observed log ratios. See "ASCN (SNP CN) – Allele-specific copy number detection algorithm" on page 312.

Assignment of SNP genotypes

The program reports the genotypes of the targeted SNP sites. To do this, it uses SNP allele information from the known genotyped reference, and expectation values for CGH copy number and the allele-specific SNP copy number (ASCN). See "Assignment of SNP genotypes" on page 319.

Loss or Lack of Heterozygosity (LOH)

The LOH algorithm identifies genomic regions that report a statistically significant scarcity of heterozygous SNP calls. Identification of such regions can inform constitutional cytogenetic studies. The algorithm reports LOH for the regions that are also detected as deletions in standard CGH analysis. However, it can also detect LOH in amplified and copy-neutral regions.

The algorithm uses total and allele-specific copy numbers to label each SNP site as homozygous or heterozygous. It then uses a binomial probability distribution to report regions that contain an unusually high fraction of homozygous SNPs. See "LOH (Loss or lack of heterozygosity) algorithm" on page 322.

Definitions Used in the Statistical Algorithms Sections

The following abbreviations and mathematical symbols are utilized in this chapter, listed in order of appearance:

Abbreviation or symbol	Definition
L	The log ratio of signal obtained from comparing each channel for each probe. The log ratio is obtained from the Feature Extraction file by converting the ratio to log base 2.
LE	The error associated with the log ratio value of a feature. The Log Ratio error (LE) is obtained from Feature Extraction files after converting it to log base 2.
Z	The *Z*-normalized value of the log ratio.
w	The moving average window. This is a user-defined window width. The average (or weighted average) is calculated from every probe within that region. The sliding window is then advanced across the genome and a new average is calculated for the new range covered by that window.
h	The height of a measured aberrant region. h is the unsigned magnitude, or the absolute value of the average log ratio for the aberrant region.
I	A genomic interval used for calculation of the magnitude of a aberrant region.
$S(I)$	The aberration score for interval *I*.
t	A user defined threshold for $S(I)$. Intervals with scores greater than t are marked significant and retained as aberrant regions.
q	The weight of a probe in further calculations, equal to the inverse of *LE* squared.
G	An entire genome, chromosome, or defined genomic boundary.
e	The probe calculated error, defined as either *LE* or *dLRsd*, whichever is greater.

Abbreviation or symbol	Definition
$\|\|A\|\|$	For a vector *A*, this is the magnitude of the vector.
$A.B$	For two vectors *A* and *B*, this is the product of the vectors: $A \cdot B = \sum_{i=1}^{n} A_i B_i$

Preprocessing Algorithms

Centralization Algorithm

Given a data vector for a single sample or entire genome, this algorithm attempts to find the best way to center the data by adding or subtracting the same constant to or from all log ratio measurements. Doing so will make the most-common-ploidy the new zero-point.

Purpose Many algorithmic approaches to aberration detection on aCGH data assume that the data points are distributed around some zero value if no aberration occurs. Typically, aCGH data fluorescence ratios for each array are normalized by setting the average log fluorescence ratio for all array elements to zero. This may lead to erroneous aberration calls for highly-aberrant genomes such as those found in tumor samples.

Use The centralization algorithm is a normalization algorithm which is used as a parameter for detecting aberrant regions or regions of constant copy number using ADM-1, ADM-2.

Algorithm Define a score S for a possible centralization value c where $S(c)$ equals the number of probes that are not included in all aberrations as called by aberration finding routines applied to the original log-ratios, shifted by c.

Try to find the value of c that minimizes score $S(c)$. That is, a value that minimizes the number of probes that are called aberrant. The search for the optimal value of c can be time consuming because you must run ADM-1 on each possible value.

In order to speed up the computation, without affecting the performance, contiguous probes are binned across the genome. In the user interface, you can choose a bin size for this algorithm (the default bin size is 10). In the default case each ten contiguous probes are averaged to reduce the number of probes used in the centralization procedure.

In the following example for the given array, the log ratio values are plotted in Figure 153. The plot of score $S(c)$ for different values of c is generated, and the plot is shown in Figure 154. The centralization algorithm defines the new zero where the center of the highest peak lies in Figure 154.

Interpretation The centralization algorithm affects the output from the aberration algorithms. See "ADM-1" on page 290.

Visualization See "ADM-1" on page 290 for more information about displaying the ADM algorithms.

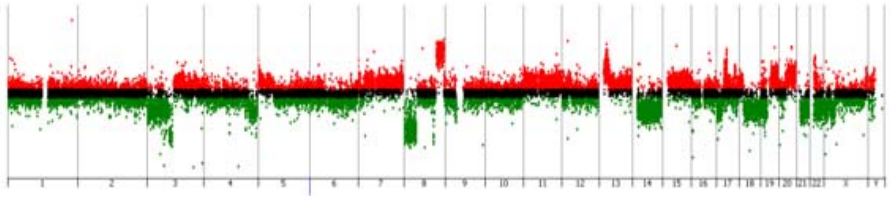


Figure 153 Log ratio values of an HT29 cell line in Genome view

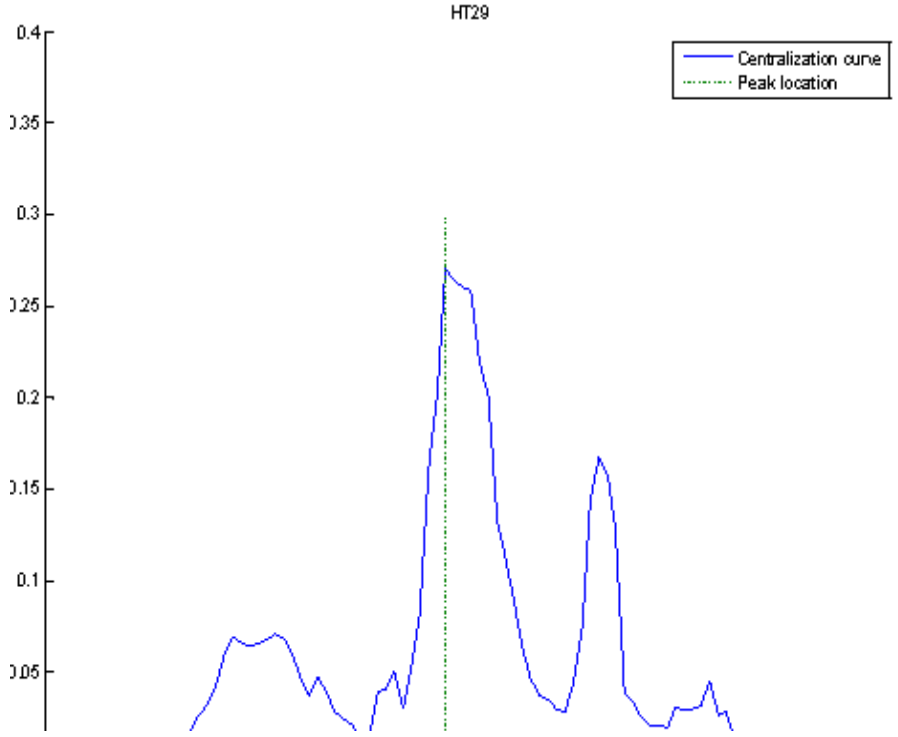


Figure 154 The plot of the score S(c) and the location of the peak for this centralization curve. For this example the log ratios will be shifted after centralization by 0.06, the absolute value of the x-coordinate of the peak.

GC Correction Algorithm

The GC Correction algorithm corrects aCGH log ratio data for the presence of “wavy” artifacts. This specific type of technical artifact correlates with the local GC content of genomic regions. The exact cause of these artifacts has not been ascertained, but their presence can interfere with detection algorithms and lead to inaccurate aberration, copy number, and LOH calls.[1]

The algorithm only corrects for GC-correlated “wavy” artifacts. The effects of the algorithm are often negligible, but it can help considerably in many cases when the baseline shows these artifacts.

Agilent recommends that you apply GC Correction when you perform CGH+SNP analysis. For the algorithm to work properly, the array design file must contain GC content values for the genomic regions that flank each CGH probe. The design files for Agilent Catalog arrays, for example, currently contain GC content values for window sizes of 2 kb, 10 kb, and 40 kb surrounding each CGH probe. In Agilent CytoGenomics, a window size of 2 kb is selected by default. If GC content data for the selected window size is not present in the design file, the program prompts you to select another value. If no GC content data is present in the design file, the algorithm cannot run.

Algorithm

The GC correction algorithm first removes significant aberrations from the log ratio data for the CGH probes on the array. The residual log ratio values are then correlated with local GC content, and the result is used to correct the log ratio values for all CGH and SNP probes on the array.

Steps

Step 1 Find correlation with GC content

In this step, the algorithm first calls aberrations in the log ratio data from the CGH probes on the array. It removes the aberrations from the data, which leaves a set of residual log ratio data.

It then assumes a linear relationship between the residual log ratio data and local GC content. The array design file supplies values for local GC content. GC content values are typically available for several window sizes around each CGH probe, and you select the desired window size when you set up GC correction. Although the algorithm is not particularly sensitive to window size, Agilent recommends that you select the smallest window size of 2 kb.

The algorithm computes a robust linear regression of log ratios on GC content. See Figure 155.

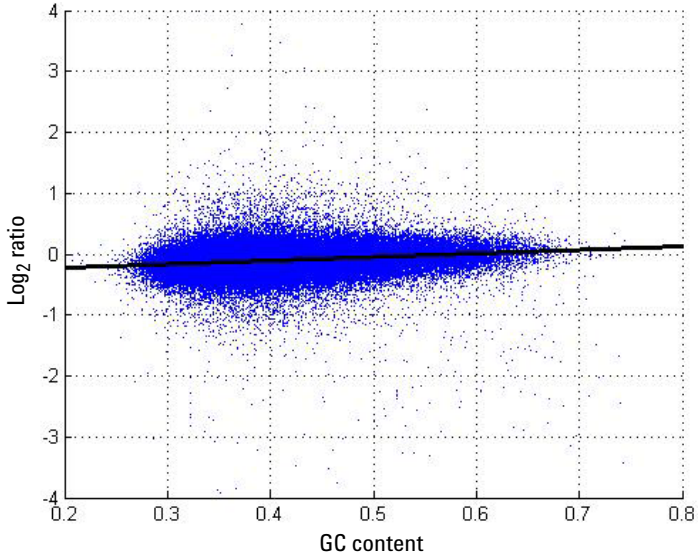


Figure 155 Linear relationship between local GC content and log ratio signal intensity from residual artifacts. This relation supplies correction factors for all probes on the array. For each probe, the GC content value reflects a region ±1 kb around the location of the probe.

Step 2 Subtract correction factors

From the log ratio value of each CGH and SNP probe on the array, the algorithm subtracts the linear trend determined by the regression. If GC Correction is selected, the corrected log ratio values are used in all downstream analysis and visualization routines for the array.

Visualization

Agilent CytoGenomics lets you view a plot of the probability density for the probes on an array both before and after GC Correction is applied. Figure 156 shows an example of this output. When you apply GC Correction, check the final distributions to assure that the correction is reasonable.

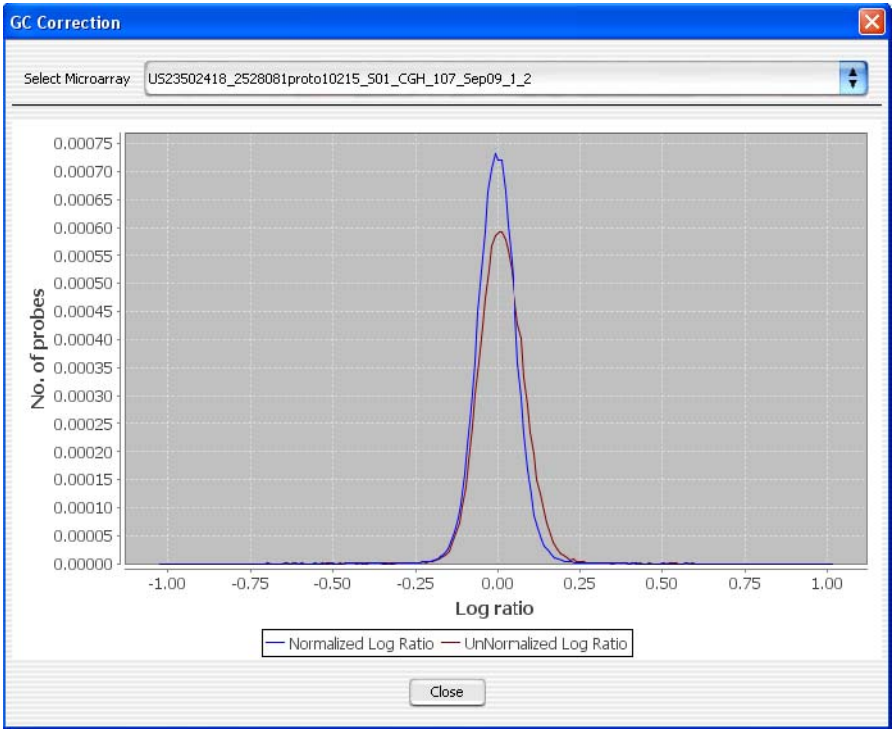


Figure 156 Distribution of $\log_2$ ratio values for an array before (red) and after (blue) the application of GC Correction. To view these distributions, in the Preprocessing tab, under Normalization, click **Plot Distribution.**

Fuzzy Zero

The Fuzzy Zero algorithm is an improved error model that explicitly includes the effects of long-range log ratio correlations. In this model, we assume that there are two independent sources of noise contributing to the total log ratio variation: the local uncorrelated probe-to-probe noise:

σ^1_{Local}

and a global noise describing baseline variation:

σ_{Global}

Purpose The ADM-1 and ADM-2 algorithms estimate the standard deviation of the mean log ratio of an interval using a statistical error model that treats probe to probe errors as independent. In many samples, the assumption that the log ratio errors of successive probes are independent is not in fact valid. The errors of the probes are often correlated over wide genomic intervals, and the ADM algorithms therefore underestimate the error for long intervals. Long aberrations with low average log ratios are thus often incorrectly deemed significant.

Use Fuzzy zero correction applies a "Global error model" to all aberrant intervals identified in ADM-1or ADM-2 analysis. The global error model uses a more realistic error model to avoid erroneous aberration calls when the errors are correlated.

Algorithm For the global error model, we assume that there are two independent sources of noise contributing to the total noise of the intervals. A local probe-to-probe noise, σ^1_{Local}, which is not correlated among different probes along the interval as described above, and a global noise, σ_{Global}, which is correlated among probes in an interval. The global noise component, σ_{Global}, is calculated as the variation of the average log ratios in large genomic intervals. As local probe-to-probe noise, σ^1_{Local}, is not correlated between different probes, when k probes are averaged, we assume that the local noise is reduced by a factor of $1/\sqrt{k}$. Thus,

$$\sigma^k_{Local} = \sigma^1_{Local} / \sqrt{k} \quad \textbf{(1)}$$

The score of interval I under the global error model, $S_g(I)$, is

$$S_g(I) = \frac{h}{\sqrt{(\sigma^k_{Local})^2 + \sigma^2_{Global}}} \quad (2)$$

here h is the average log ratio of all probes in the interval I. If the ADM-2 algorithm is used, h is the quality weighted average log ratio of all probes in the interval I.

Using α to denote $\left(\frac{\sigma_{Global}}{\sigma^1_{Local}}\right)^2$ we derive

$$\sigma^k = \sigma^1_{Local}\sqrt{\frac{1}{k} + \alpha} \quad (3)$$

and

$$S_g(I) = \frac{h}{\sigma^1_{Local}\sqrt{\frac{1}{k} + \alpha}} \quad (4)$$

Fitting the model parameters

For a given log ratio vector v of length N (for a particular sample), σ^1_{Local} equals the Derivative Log Ratio Spread (*dLRsd*):

$\sigma^1_{Local} = dLRsd(v_0)$, where v_0 is the vector of the individual probe log ratios L.

See "Noise Estimation – the Derivative Log Ratio Spread" on page 294 for a description of *dLRsd*.

Then α is estimated using the following iterative procedure:

1 Start with an initial estimate of $\alpha_0 = 0.01$

Start with an initial value of the vector v as v_0.

2 At each iteration i, new values of α and v are calculated as α_i and v_i:

a Find all aberrant intervals I in v_i with the score $S_g(I)$ (**4**) above the user defined threshold, T (the default recommended starting threshold is 6).

Note that the score $S_g(I)$ (**4**) depends on the current value of α_i. This set of aberrant intervals is considered as the signal component of the data.

b Compute the residual vector v^r.

To compute v^r, we subtract from v_i the heights of each aberrant interval I. Namely, we subtract from each probe in I the height h of the aberration containing the probe. The resulting vector v^r represents the current estimate of the noise in the data.

c Estimate the combined noise σ_i^k from the residual vector v^r.

To estimate the combined noise σ_i^k, we bin consecutive probes into bins of size $k = \sqrt{N}$. Then we derive a binned vector u_k, where each element of u_k is the average log ratio of all probes in one bin. We estimate σ_i^k by computing $dLRsd(u_k)$. To make this estimation more robust we repeat the binning using 10 different starting positions of the first bin. The final estimation of σ_i^k is the median of these 10 different estimations.

d Compute the new α_{i+1} based on the current estimate of σ_i^k.

e Set $i+1 = v^r$.

3 Continue the iterations until the process converges, *i.e.* $|\alpha_i - \alpha_{i-1}| < 0.001$, or 10 iterations were made.

How the data centering algorithms can affect aberration calls

The data centering algorithms are used in the aberration detection algorithms. Output from application of the detection algorithms is visible in Genome, Chromosome, and Gene views. Figure 157 and Figure 158 show how the application of the data centering algorithms can affect aberration detection.

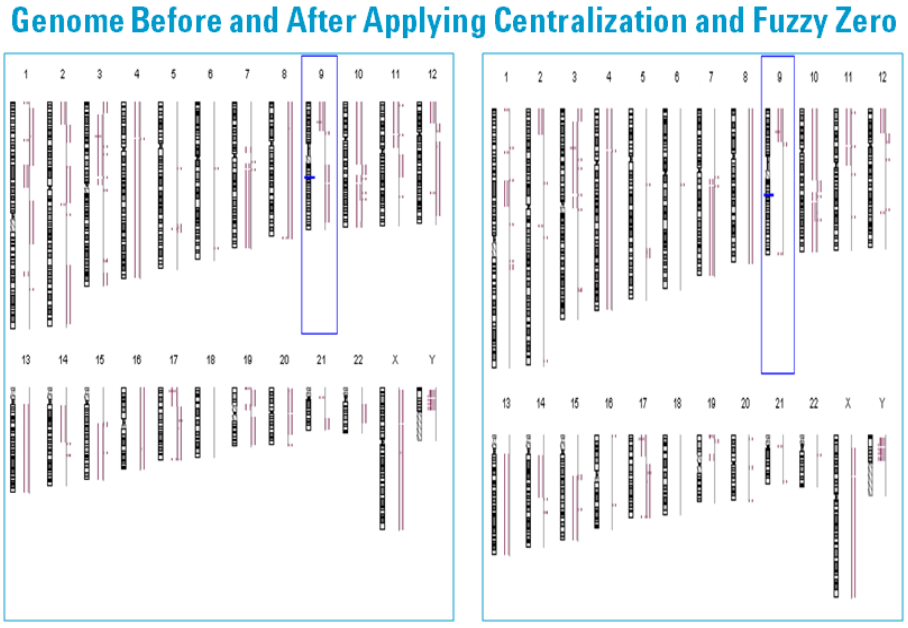


Figure 157 Aberration results before (left panel) and after (right panel) the application of data centering algorithms. Before application of these algorithms, aberrant regions (dark red lines) are present in regions without a clear over- or under-abundance of log ratios (light red blocks). In the right panel, most spurious aberrant regions have been removed, while the robust aberrant regions remain. A detailed view of chromosome 9 appears in Figure 158.

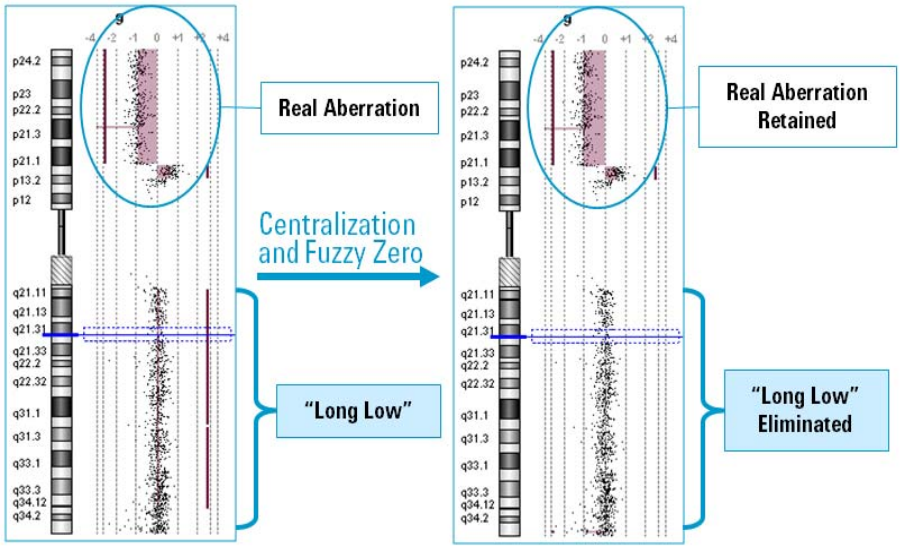


Figure 158 Application of the data centering (centralization and fuzzy zero) algorithms, illustrated. Re-centering the data allows identification of robust aberrant regions while removing spurious aberrations that are reported when the log ratios of individual probes are not completely independent.

Error Model and Combining Replicates

Error modelling is a way to separate events measured from biological signals such as protein-DNA binding from signals measured from systematic variations in the technology. The parameters for error calculation use probe log ratio and quality measures from Agilent Feature Extraction (FE).

Error Model

Purpose The purpose of the error model is to identify which probes are most reliable in the calculation of aberrant regions. The confidence of such events is calculated based on the log ratio error of each probe. For more information regarding the FE log ratio error calculation, see the *Agilent Feature Extraction User Guide*.

Use The error model works by selecting robust probe signals for inclusion in the detection algorithms and is therefore on by default. The error model is especially important whenever there are replicate probes within an array or between samples.

Algorithm To combine log ratios of replicated features, a weight is applied first. The weight is proportional to quality, and quality is defined as the inverse of square of log ratio error.

To combine replicated probes in an array (intra array) or within replicated arrays (inter array), combine the log ratio and the log ratio error as follows:

1 Define a weight q_i for *each* probe to be $q_i = 1/\ e_i^2$ That is, the noisier a given probe is, the smaller is its weight. The error, e_i, is defined as the maximum between the log ratio error, LE_i, of that probe and the spread of derivative of log ratio, *dLRsd*, for that array, i.e. $e_i = max(LE_i, dLRsd)$. The *dLRsd* is described under "Noise Estimation – the Derivative Log Ratio Spread" on page 294.

2 Define the quality-weighted average log ratio for replicated probes as:

$$L_{ave} = \frac{\sum_{i \in I} q_i L_i}{\sum q_i} \quad (5)$$

3 Estimate the Log Ratio Error of the above mean:

$$LE_{ave} = \frac{1}{\sqrt{\sum_{i \in I} \frac{1}{e_i^2}}} = \sqrt{\frac{1}{\sum q_i}} \quad (6)$$

Combining Dye Swap

Algorithm When combining dye-swapped arrays, Agilent CytoGenomics separately combines any replicates using the original dye channels and the dye-swapped chips according to the above equations. This yields both the average weighted sample and the combined error measurement for each polarity.

The error model for the dye-swapped arrays is the sum of the Log Ratio Error of the two polarities in quadrature:

$$\sigma_{combined}(dyeSwap) = sqrt((\sigma^2_{polarity1} + \sigma^2_{polarity2})/\ 2) \quad (7)$$

Aberration Detection Algorithms for CGH Analysis

This section describes in detail the purpose, comparative use, and step-by-step methods used by the aberration detection algorithms for Agilent CytoGenomics. These algorithms interpret log ratios and probe quality measures from the microarray to identify contiguous genomic regions that correspond to chromosomal aberrations or copy number variations. At the end of this section is a guide to the interpretation and visualization of the detection algorithm output. Table 22 lists the topics available in this section.

Table 22 Detection algorithms topics

Subject	See these topics
Z-score algorithm	"Z-Scoring for Aberrant Regions" on page 287
Aberration Detection Method (ADM) algorithms	"ADM-1" on page 290 "ADM-2" on page 293
Noise estimation	"Noise Estimation – the Derivative Log Ratio Spread" on page 294
Interpretation and Visualization	"Interpretation and Visualization of the Detection Algorithms" on page 297 "How the data centering algorithms can affect aberration calls" on page 284

NOTE For descriptions of the algorithms for the analysis of CGH+SNP arrays, see "Algorithms for CGH+SNP Analysis" on page 311.

Z-Scoring for Aberrant Regions

This method identifies all aberrant regions in a given sample using statistical analysis based on hypergeometric Z-scores.

Purpose The Z-score algorithm is a quick method of detecting aberrant regions. It finds intervals using a sliding window of fixed size, specified by the user. Enrichment in the number of probes with high log ratios for any given window size yields an aberrant region.

Use The exploratory phase of an analysis in Agilent CytoGenomics may include visual inspection of regions with an over abundance of probes with log ratios which pass a scoring threshold. This abundance is compared to the number of probes expected to pass the same threshold at random, yielding a Z-score for each window.

Algorithm The scoring method has essentially two steps. In the first step it identifies the total number of probes with log ratios significantly different from zero in the sample array. These probes are referred to as outlier probes.

NOTE Ideally, the probe statistics would be computed for samples that contain no genetic anomalies, so that μ and σrepresent the distribution of a non-diseased sample.

In the second step, the method determines if the actual sample of interest has a significantly higher proportion of aberrant probes in any given genomic region than the proportion of total aberrant probes in the sample used in the first step. If it identifies any such region, then this region gets a higher score and it is called as an aberrant region. The two steps are explained in more detail in the following sections:

Step 1: Calibration

For each probe, the log ratio is *Z*-normalized by computing the usual formula:

$$Z(L) = \frac{L - \mu}{\sigma} \quad \textbf{(8)}$$

where L is the signal log ratio, μ is the mean and σ is the noise level of the population of such log ratios. Chromosomes X and Y are not included in the calculation of μ and σ since gender differences between arrays can offset the statistics.

Each *Z*-value can be classified as significantly above or below the mean by using a Z cutoff, Z_C. This cutoff can be supplied as a user-specified value. In essence, you are simply stating that you consider log ratios greater than Z_C to be outliers from the normal population of log ratios.

NOTE

Z_C is not a cutoff used to *filter* data. It is a cutoff for *classifying* data as being significantly above or below the mean. To avoid reinforcing the idea that this value filters *Z*-scores, the Z Score Parameter Panel refers to Z_C as threshold.

As part of the computation, the number of entries in each of three classes are counted using the threshold setting Z_C from the UI:

- R = number above positive cutoff (*i.e.* the number of outliers above Z_C in a normal array after proper normalization)
- R' = number below negative cutoff (*i.e.* the number of outliers below $-Z_C$ in a normal array after proper normalization)
- N = the total number of measurements

These Z values and counts are pre-computed and reserved for calculations in **step 2**. The values computed in **step 1** would only need to be recomputed if a different Z_C were desired. Even so, μ and σ can still be reused without computation.

If you want to determine the pre-computed statistics more accurately, you can select the specific arrays to be used in this step.

Step 2: Computation

To compute a moving average, log ratios are averaged over a small subset of points in the genome. This moving average window, w, may be simply a number of adjacent measurements or it may be over a positional window (such as every megabase). The objective is to analyze the over- or under-abundance of log ratios within the window that deviate significantly from the mean from step 1. For this smaller subset, the same three counts as in step 1 are computed using exactly the same cutoff values, but in this case, only for the points within the averaging window, w:

- r = the number above the positive cutoff (Z_C) in w
- r' = the number below the negative cutoff ($-Z_C$) in w
- n = the total number of measurements in w

Now, compute an exact Z-score that measures the significance of this over abundance or under abundance in w of significant positive deviations as:

$$Z(w) = \frac{\left(r - n\frac{R}{N}\right)}{\sqrt{n\left(\frac{R}{N}\right)\left(1 - \frac{R}{N}\right)\left(1 - \frac{n-1}{N-1}\right)}} \quad \textbf{(9)}$$

You can compute the same formula for r' to obtain a score for negative deviations. This score can be plotted in a manner analogous to a moving average. This would identify statistically significant groups of probes that appear to deviate from the typical distribution of values for the given microarrays. In this way, it provides some predictive power to call amplification or deletion events in CGH studies.

ADM-1

Aberration Detection Method 1 (ADM-1 or "adam-one") is a detection algorithm that identifies all aberrant intervals in a given sample with consistently high or low log ratios based on the statistical score.

Purpose The ADM-1 algorithm searches for intervals in which the statistical score exceeds a user specified threshold. The statistical score is calculated based on the average log ratios of the probes and the number of probes in the interval. ADM-1 reports contiguous genomic regions of arbitrary size as aberrant regions.

Use In contrast to the *Z*-score algorithm, the ADM algorithms do not rely upon a set window size, instead sampling adjacent probes to arrive at a robust estimation of the true range of each aberrant segment. The ADM-1 algorithm may be limited to a subset of genomic locations which have annotated features. This restriction, termed 'genomic boundaries', is available in the Agilent CytoGenomics Settings UI.

Algorithm The ADM-1 statistical score is computed as the average normalized log ratios of all probes in the genomic interval multiplied by the square root of the number of these probes. It represents the deviation of the average of the normalized log ratios from its expected value of zero.

The ADM-1 score is proportional to the height h (absolute average log ratio) of the genomic interval, and to the square root of the number of probes in the interval. Roughly, for an interval to have a high ADM-1 score, it should have high height or/and include large number of probes.

Before calling the ADM-1 routine, the log ratios are normalized in the following way:

Step 1: Normalization

In the normalization step, the expected average μ is subtracted from all log ratios L, and then these modified log ratios are divided by the estimated variance σ. This transforms the log ratio scores into a normal Z distribution with a mean of 0 under the null model assumption:

$$Z(L) = \frac{L - \mu}{\sigma} \quad \textbf{(10)}$$

where μ is the mean and σ is the noise level of the population of such log ratios.

Once the data are transformed the following score is assigned to each interval I:

$$S(I) = \frac{\sum_{i \in I} Z_i}{\sqrt{N_I}}, \quad \textbf{(11)}$$

where N_I is the number of probes in the interval, and $S(I)$ represents the number of standard deviations that the sum of values in I deviates from its expected value of 0 under the null model.

Step 2: Iteration

A call to the ADM-1 function starts a recursive process. The first step is to identify the interval I for which $S(I)$ is maximal and exceeds a predefined threshold parameter, t, specified in the user interface. Then the process is called on the interior of this interval, using the interval median as a mean for re-centering the values, as well as on the two intervals, one to the left and the other to the right flanking I, towards the two ends of the chromosome.

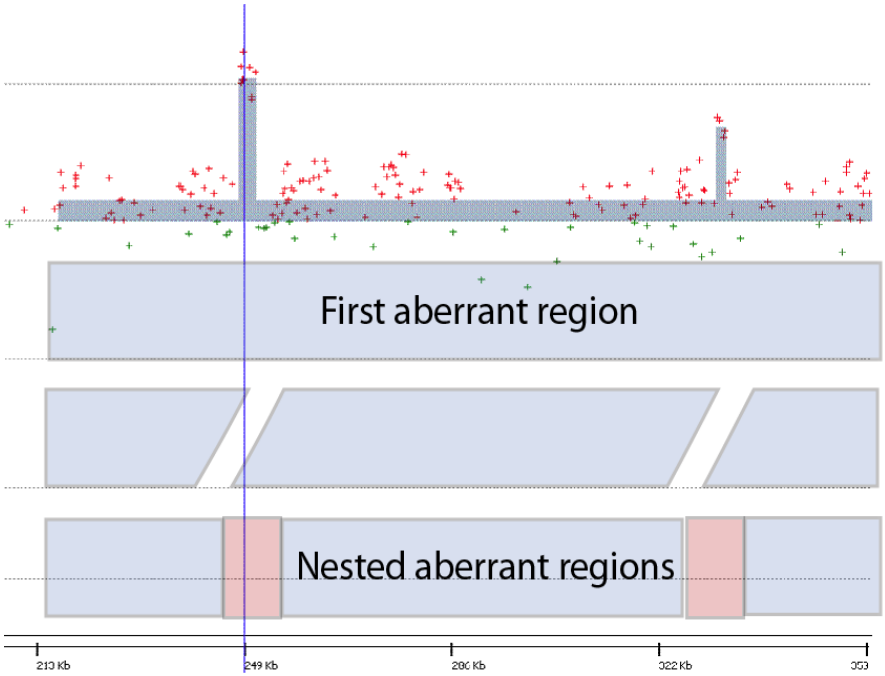


Figure 159 Aberration region identification by recursive interval scoring. For each genomic interval *I*, Agilent CytoGenomics computes an interval score, *S(I)*.

The overall recursive structure of the algorithm is:

Given a data vector for a single sample, single chromosome, and a statistical threshold value:

1. Find the most significant interval *I* in the chromosome.
2. If *S(I)* >= *t*, mark *I* a significant interval.
3. Add *I* to the list of intervals.

Search recursively for more intervals (a) to the left of *I*, (b) to the right of *I*, and after normalizing, (c) inside *I*.

NOTE A text report can also be generated which reports the p-value corresponding to each interval. The p-value is calculated using the normal probability distribution function and the score of that interval.

ADM-2

The Aberration Detection Method 2 (ADM-2 or "adam-two") algorithm identifies all aberrant intervals in a given sample with consistently high or low log ratios based on a statistical score.

Purpose The ADM-2 algorithm searches for intervals in which a statistical score based on the average quality weighted log ratio of the sample and reference channels exceeds a user specified threshold. ADM-2 reports contiguous genomic regions of arbitrary size as aberrant regions.

Use In contrast to the *Z*-score algorithm, the ADM algorithms do not rely upon a set window size, instead sampling adjacent probes to arrive at a robust estimation of the true range of the aberrant segment. ADM-2 differs from ADM-1 by using probe quality information to weight the log ratios before calculating the score for the interval. The ADM algorithms may be restricted to a subset of genomic locations which have annotated features. This restriction, termed 'genomic boundaries', is available in the Agilent CytoGenomics Settings UI.

Algorithm ADM-2 uses the same iterative procedure as ADM-1 to find all genomic intervals with the score above a user specified threshold. In ADM-2, the score represents the deviation of the weighted average of the log ratios from its expected value of zero. This score is similar to the statistical score used in ADM-1 analysis, but ADM-2 incorporates quality information about each probe measurement.

The Quality-Weighted Interval Score algorithm (ADM-2) computes a set of aberrations for a given sample. The overall recursive structure of the algorithm is the same as it is in ADM-1.

Step 1: Log ratio error model The only difference between ADM-1 and ADM-2 is in the definition of the score of the interval. ADM-1 considers only the log-ratios, while in ADM-2 you also consider the log-ratio error information, hence the name Quality-Weighted Interval Score.

The following describes the ADM-2 score:

1 Input is a vector of pairs $(L_1,LE_1),(L_2,LE_2),...,(L_n,LE_n)$, where

- L_i is the log-ratio signal for the i-th probe
- LE_i is the log-ratio error for the i-th probe ordered
- If *dLRsd* is greater than LE_i, then LE_i is set to the *dLRsd* value.

Define $q_i = 1/(LE_i)^2$ **(12)**

Assume that under the null model, $L_i \sim N(0, 1/q_i)$ and the different L_i are independent of each other.

2 Consider the weighted sum, for an interval *I*:

$$\sum_{i \in I} q_i L_i \quad \textbf{(13)}$$

3 Compute the variance of *S(I)*:

$$var[S(I)] = var(\sum q_i L_i) = \sum q_i^2 var L_i) = \sum q_i \quad \textbf{(14)}$$

4 Compute the ADM-2 interval score:

$$S(I) = \frac{\sum q_i L_i}{\sqrt{\sum q_i}} \quad \textbf{(15)}$$

This score reflects the deviation of the weighted sum from its expected value (0) in units of standard deviation. If the quality weight of each probe is the same as the probe to probe noise of the array, then the score will be same as ADM-1.

NOTE A text report can also be generated which reports the p-value corresponding to each interval. The p-value is calculated using the normal probability distribution function and the score of that interval.

Noise Estimation – the Derivative Log Ratio Spread

To make aberration calls, Agilent CytoGenomics needs a measure of the log ratio noise for each array. A good estimate of noise is attained by calculating the robust standard deviation (spread) of the log ratio differences between consecutive probes (*dLRsd*) along all chromosomes, divided by $\sqrt{2}$ to compensate for noise averaging. The *dLRsd* is an estimate of noise that is robust even in the presence of highly aberrant samples.

Purpose Even highly aneuploid samples have chromosomes with extensive stretches along which the genomic copy number is constant, or nearly so. In such constant-copy-number regions, the true log ratios are constant, although

not necessarily zero. Estimations of log ratio error, and therefore the minimum log ratio difference required to make reliable amplification or deletion calls, is based on observations of the variation in such constant-copy-number regions. The *dLRsd* metric is an attempt to quantitate such "eyeball" estimates.

For normal samples, the *dLRsd* is the width of a self-self distribution, and should be below 0.2 log units. It will be somewhat greater for abnormal chromosomes because (a) the width of regions of constant copy number different from two will include both noise and the variable log ratio compression observed for many probes, (b) in deleted regions, the noise tends to be slightly higher, as expected from the error model.

To make *dLRsd* more robust and a true measure of noise, outliers are removed from the constant-copy-number regions. To remove outliers, IQR (Inter Quartile Range) statistics are used with appropriate scaling to calculate the spread of the distribution instead of calculating the standard deviation of the derivative of log ratio directly. See Figure 160.

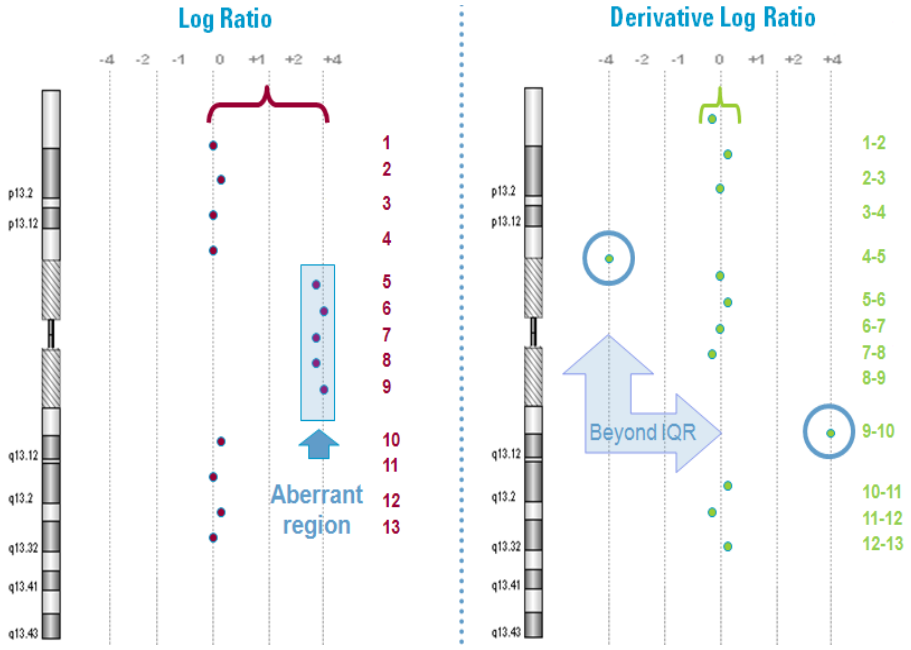


Figure 160 The difference between log ratios and derivative log ratios, illustrated. When using the log ratio spread as an estimate of array variance, aberrant regions can artificially inflate the noise of the array. The derivative log ratio calculates the difference between adjacent log ratios (the derivative log ratio, or DLR) as noise. Agilent CytoGenomics uses the inter-quartile range (IQR) as 50% of the data around the mean of the DLR to estimate the variance.

Interpretation and Visualization of the Detection Algorithms

Output from application of the detection algorithms is visible in the genome, chromosome, and gene view panels. This section describes the interpretation of the visualization capabilities in Agilent CytoGenomics.

Z-score algorithm:

Interpretation

For each moving window, the *Z*-score algorithm takes the points within that window as a sample and computes a hypergeometric *Z*-score that measures the significant number of outliers — points that lie above (positive values) or below (negative values) the threshold *Z*-score. Note that a point that is slightly beyond the threshold is counted the same as a point that is considerably beyond the threshold.

Z-scores are plotted to indicate statistically significant groups of probes that appear to deviate from the typical distribution of values for the given microarrays. They provide some predictive power to call amplification or deletion events in CGH studies. You should adjust the cut-off, *Z*, appropriately based on your visual analysis of amplified and deleted regions in the chromosomes.

Also note that the score that is plotted has nothing to do with the log ratio, and you should not expect the values to necessarily line up with the log ratios. It simply represents a statistical measure of aberration that can be used to track the distribution of outliers, which you can usually see by comparing the scatter plot to the *Z*-score plot.

One final interesting point is that if you set the threshold too low (i.e. cut-off is too small), most of the data points on an array are outliers (i.e. very high values of *R* and *R'*), and you will probably get a *Z*-score of zero. If on the other hand, you set the threshold too high, none of the points are outliers and again, the *Z*-score will be zero. Usually a threshold of 2-3 is the best setting. You can go slightly higher if you want to look for very deviant aberrations, but going too high will show no aberration.

Visualization

When the final *Z*-scores are computed, they can be plotted as a line graph similar to the moving average. To enhance the visibility of the plots and to distinguish them from the moving average, the graphs are filled from the origin. As a further refinement, the filling is alpha-blended for transparency. When plotting multiple microarrays simultaneously, this

minimizes obscuring of the data and allows you to detect overlaps. For two or three simultaneous plots, it is often possible to distinguish the various intersections based on the color blending.

The *Z*-score is also reduced by a factor of 10, thus allowing you to read the actual underlying value by interpolating the location on the graph scale (2, 4) and simply multiplying that value by 10. All *Z*-scores are positive, and those shown in the negative direction are actually positive *Z*-scores for decreased copy number.

It can still be difficult to read small segments of significant *Z*-scores, particularly in the overview. So as a further aid, *side-bars* are provided. Usually, these side-bars are not overlapped and provide a quick means for locating interesting anomalies in single microarrays. They also provide better separation when looking at multiple microarrays. Normally, the side-bars are stacked, but if there is insufficient room they may overlap. This allows you to see clear areas of interest. You can always manipulate the display to enlarge the available space in order to separate the side-bars as needed.

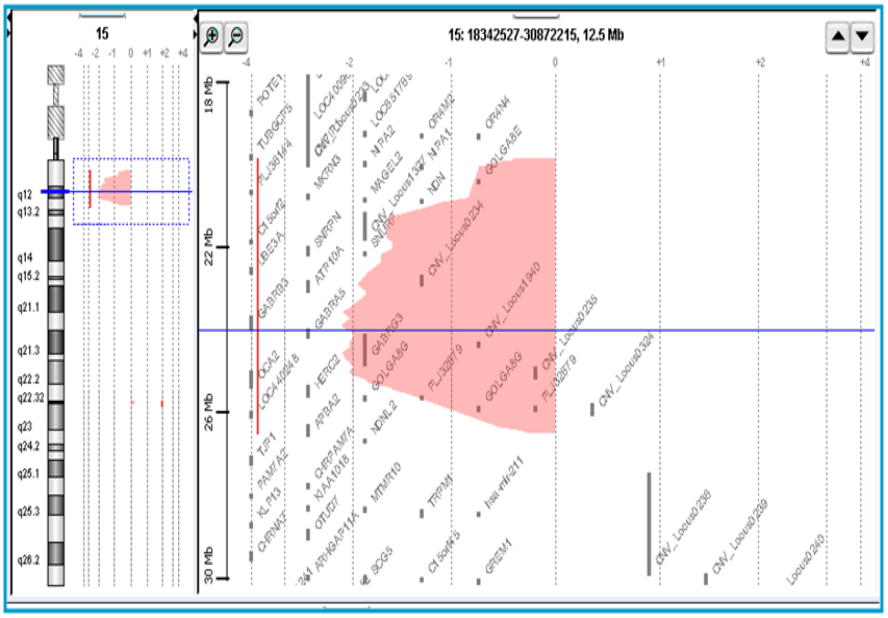


Figure 161 Regions with enriched over- or under-abundance probes as visualized with the Z-score algorithm output.

Interpretation

ADM algorithms:

The Aberration Detection Algorithms (ADM-1 and ADM-2) use a recursive process to identify contiguous stretches of aberrant probes. The recursion stops when no interval with *S(I)* exceeding the user-specified threshold *t* is found. All intervals found in this process are reported, and a plot is generated as output. The intervals are rendered as steps in the visualization panel. The height of each step is equal to the average log ratio of that interval. Steps are also extended on each side of an interval by a fixed distance (other than at the end of the chromosome and centromere). See Figure 162.

NOTE The fixed interval is set to a default distance of halfway to a neighboring probe, or 200 base pairs (bp), whichever is less.

A threshold can be applied to the ADM algorithms. The optimal threshold depends upon several factors. These factors may be the DNA sample type, the DNA sample quality, or other measures, and can be explored interactively in Agilent CytoGenomics. A recommended starting point for finding the optimum threshold is a value of 6.

Visualization

When the aberration regions are computed, they are plotted as a bar graph, colored by sample. To enhance the visibility of the plots and to distinguish them from the moving average, the graphs are filled from the origin. As a further refinement, the filling is alpha-blended for transparency. When plotting multiple microarrays simultaneously, this minimizes obscuring of the data and allows you to detect overlaps. For two or three simultaneous plots, it is often possible to distinguish the various regions based on the color blending.

It can still be difficult to read small segments of aberrant regions. By clicking on a chromosomal region of interest or by dragging a rectangular selection around a genomic region you can easily zoom in on that region.

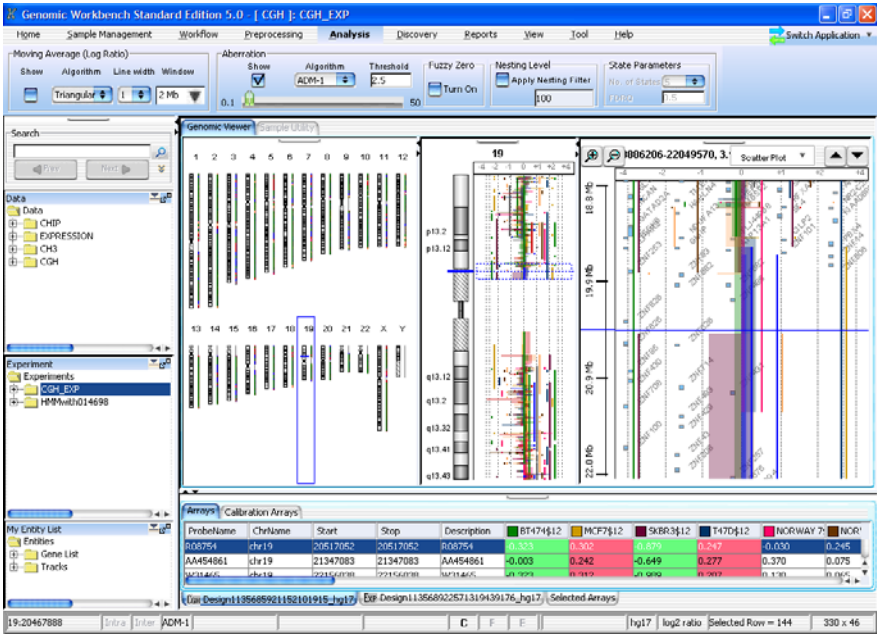


Figure 162 An example of ADM output. All intervals are rendered as steps where the height of each step is equal to the average log ratio for that interval.

Interpretation **CBS algorithm:**

The recursion stops when no additional change points can be found. All intervals with constant copy number bounded by these change points found in this process are reported, and a plot is generated as output. The intervals are rendered as bars in the visualization panel. Identification of these intervals can be used for further exploration of aberration regions or for copy change analysis.

Visualization When the change points are identified using CBS, each region of nearly constant copy number is reported as a bar graph colored by sample and plotted by magnitude. Change points are then simply identified as breaks in the bar series for any given color. To enhance the visibility of the plots, the bars are alpha-blended for transparency. When plotting multiple microarrays simultaneously, this minimizes obscuring of the data and allows you to detect overlaps. For two or three simultaneous plots, it is often possible to distinguish the various regions based on the color blending.

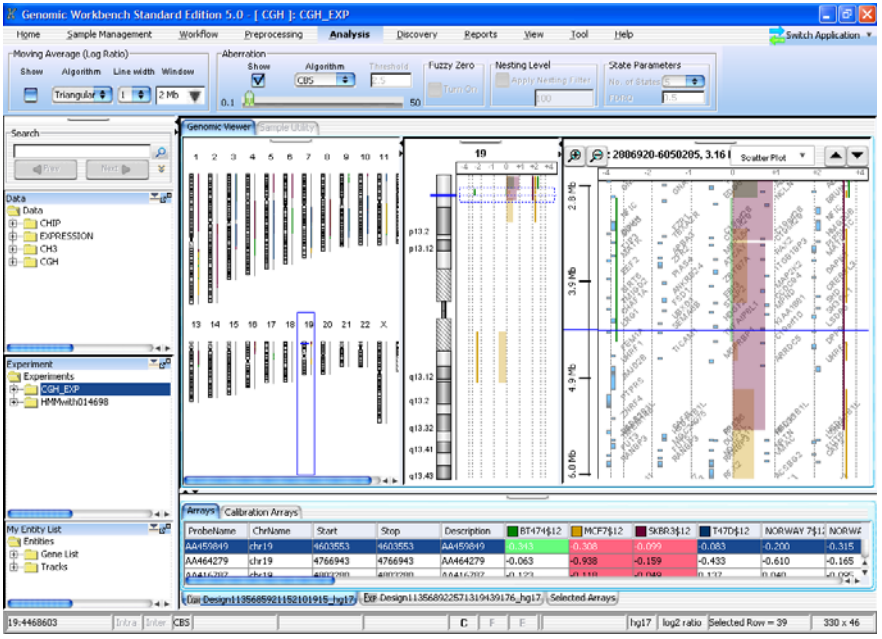


Figure 163 An example of CBS output. All intervals are rendered as bars indicating regions of constant copy number.

Algorithms for Multi-array Analysis

Penetrance

Penetrance analysis is used to find the percentage of samples that share aberrations in a particular genomic region among multiple samples. Amplification and deletions are considered separately.

In probe based penetrance, for each aberrant probe, the percentage of samples that shows that amplification or deletion is calculated. The height of aberration is not considered for this calculation.

In interval based penetrance, for each interval, the percentage of samples that share that amplification or deletion is counted.

Purpose This algorithm is one of the algorithms that can identify aberrant regions that are common among multiple samples that use output from ADM-1, ADM-2, or CBS.

Use To generate a report showing the overlap of identified regions from ADM-1, ADM-2, or CBS across many samples. The algorithm differs from the common aberration analysis in that it does not recalculate the score of each candidate interval using the ADM algorithms, nor does it attempt to identify statistically significant common aberrations. Instead, it reports how many samples agree with a given aberrant probe or region prediction.

Interval Based Penetrance Algorithm Interval-based penetrance is the ability to calculate the percentage of samples that share regions of amplification or deletion. The analysis has the following steps:

1. Collect all start and end points from all the aberrant intervals in all the samples determined from any of the detection algorithms.
2. Sort by genomic position the resulting break points.
3. Construct a set of candidate genomic intervals using all break points listed in **Step 2**, such that there is no break point inside any candidate interval.
4. For each candidate interval A, count the number of agreements in amplifications and deletions separately, representing the penetrance of aberration A in sample set S.

5 Report candidate intervals with penetrance scores above a user-defined threshold (the default threshold is 6) and the names of the corresponding supporting samples.

NOTE For counting penetrance, amplification and deletion are considered separately. The aberration state is set by the aberration height of the deepest nested interval.

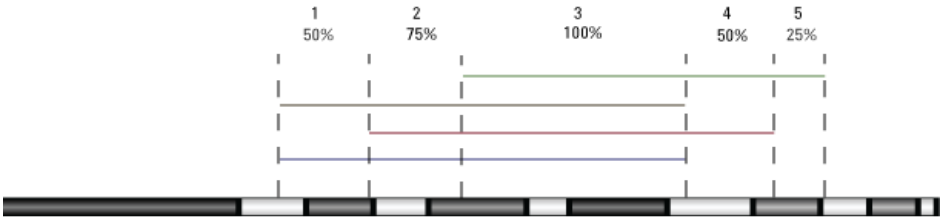


Figure 164 Scoring detected aberrant regions from different samples. Here each sample is color coded. Note that interval based penetrance does not take into account the height of the aberrant region. The vertical axis in the illustration serves only to differentiate the different samples, and not a difference in average log ratios.

Probe Based Penetrance Algorithm

Probe based penetrance calculates the percentage of samples that share a probe amplification or deletion. The analysis has the following steps:

1 Collect all aberrant probes from all samples.

2 For each aberrant probe, calculate the percentage of sample which also show that probe to be aberrant (in the same direction, amplification or deletion), without regard to the magnitude of the log ratio.

Differential Aberration Analysis

Differential aberration is used with multiple samples which have been partitioned into two comparative groups. It is used to identify genomic intervals which have an overabundance of gains or losses in one group of samples compared to the other group in a given genomic region. This algorithm uses output from the aberration detection algorithms such as ADM-1 and ADM-2 and reports only statistically significant differences in aberration distributions between the two groups of samples.

Purpose The differential aberration method identifies genomic regions with statistically significant enrichment of gains or losses in one of the group of samples. The tool then allows one to select the regions of interest and save them for further investigation.

Use Differential aberration is used to identify genomic intervals that have statistically significant aberration differences across multiple samples. After calculating aberration profiles using one of the aberration detection methods, groups of interest are defined and the differential aberration is performed.

For samples which are naturally grouped into two groups by some characteristic (e.g. disease, tissue type, etc.) then this tool can detect the aberrant frequencies in this partition and assign statistical significance to the differences. For example, assume you have two groups named *Group 1*and *Group 2*, both consisting of 100 samples. If *Group 1* has 40 aberration amplifications (called gains) in a specific genomic region and *Group 2* has 5 gains in that same region, the statistical significance of observing the 40 gains in *Group 1* given observing only 5 gains in *Group 2* is calculated by using *p-values* from the Hypergeometric distribution. In this example, the *p*-value of observing 40 gains in *Group 1* is 1.3×10^{-9}, meaning that *Group 1* has a significant enrichment of gains compared to *Group 2*.

Algorithm The report analysis first finds all intervals defined by breakpoints for each sample in the study. Enrichment from the perspective of gains or losses from each of the two groups is calculated using the Hypergeometric distribution. It then selects significant results.

NOTE A minimum of three samples in each group is recommended for Differential Aberration analysis. The analysis method becomes statistically more powerful with increasing numbers of samples in each group.

Step 1: Enumerate aberrations by sample groups

The algorithm identifies genomic regions that share aberrations across sample members in each group.

1 Apply one of aberration detection algorithms (ADM-1or ADM-2) to a set of samples, S_1, S_2, ... S_n, to identify a set of aberrant genomic intervals in each sample that have a score above a user-specified threshold.

2 Construct a set of candidate genomic intervals for common aberration analysis using all intervals identified in **Step 1**.

3 Count the number of amplifications (gains) and deletions (losses) in each of the groups *1* and *2* for each candidate interval identified in **Step 2**. Total the number of gains and losses from the identical genomic intervals from the two groups.

NOTE

The candidate intervals can be filtered based on the number of probes in the interval either before or after Differential Aberration analysis. Prior to Differential Aberration analysis an aberration filter can be applied before generating candidate genomic intervals. Following the Differential Aberration analysis you can select the number of probes as an interval filter condition. See "Interval Filter" on page 210

Step 2: Calculate probabilities using the Hypergeometric distribution

The enrichment of gains or losses for each group is calculated for each candidate interval. Enrichment of gains or losses is expressed as the area under the Hypergeometric tail (HGT). The Hypergeometric tail probabilities reflect the probability of observing y or more number of gains or losses in a group - *i.e.* it is a measure of enrichment of gains (losses) in a group. This is also referred to as a *p*-value, which is reported in Agilent CytoGenomics as $-log_{10}(p)$; thus larger values are more significant. See "Enrichment Analysis - the Hypergeometric distribution" on page 333.

1 For each candidate interval compute the enrichment of gains in *Group 1*. This is the probability of observing a or more gains in *Group 1* where *Group 1* consists of A out of a total of N samples and there are n total aberrations. This cumulative probability is expressed as a *p*-value calculated from the Hypergeometric Tail (*HGT*) in the following way:

$$'(X \geq a) = HGT(a,\ N,\ A,\ n) = \sum_{i=a}^{min(n,\ A)} \frac{\binom{n}{i}\binom{N-n}{A-i}}{\binom{N}{A}} \quad \textbf{(16)}$$

2 Compute the enrichment of losses in *Group 1*, and the enrichment of gains and losses in *Group 2* using the formula in **Step 1.**

As stated, the enrichment analysis is directional. This means each calculation considers only whether or not there is an overabundance of aberrations within an interval as compared to the other comparison group. Four total calculations must then be completed in order to assess whether there is enrichment of either gains or losses of *Group 1* and of *Group 2*. The following four *p*-values are therefore calculated for every interval:

- $p\text{-}value(G_1)$ = Enrichment of Gains in group *1* given the observed number of gains in group *1* and *2*.
- $p\text{-}value(G_2)$ = Enrichment of Gains in group *2* given the observed number of gains in group *1* and *2*.
- $p\text{-}value(L_1)$ = Enrichment of Losses in group *1* given the observed number of losses in group *1* and *2*.
- $p\text{-}value(L_2)$ = Enrichment of Losses in group *2* given the observed number of losses in group *1* and *2*.

Step 3: Select Significant Results

The four *p*-values for every candidate interval are available as results for further analysis and filtering. The filter applied is either the default filter described below, or a customized filter defined by the user. See "Interval Filter" on page 210.

The default filter selects intervals with a $-log_{10}(p\text{-value})$ greater than the selected threshold of $-log_{10}(p\text{-value})$. The default threshold is a *p*-value less than 0.001. Because Agilent CytoGenomics reports the *p*-values in $-log_{10}$ values, the default threshold for enrichment is a value greater than *3*.

Interpretation and visualization of the algorithms for the comparison of commonly aberrant regions

Output from the comparison of common aberrant regions is generally visible in the algorithm-specific result windows. This section describes the interpretation of the visualization capabilities in Agilent CytoGenomics for these analysis algorithms.

Differential Aberration analysis

Interpretation

For each candidate genomic interval, the differential aberration method identifies those regions with statistically significant enrichment of gains or losses in one of the group of samples. The tool then allows one to select the regions of interest and save them for further investigation.

For samples which are naturally grouped into two groups by some characteristic (e.g. disease, tissue type, etc.) then this tool can detect the aberrant frequencies in this partition and assign statistical significance to the differences. For example, assume you have two groups named *A* and *B*, both consisting of 100 samples. If group *A* has 40 aberration amplifications (called gains) in a specific genomic region and group *B* has 5 gains in that same region, the statistical significance of observing the 40 gains in group *A* given observing only 5 gains in group *B* is calculated by using *p-values* from the Hypergeometric distribution. In this example, the *p*-value of observing 40 gains in group *A* is 1.3×10^{-9}, meaning that group *A* has a significant enrichment of gains compared to group *B*.

Visualization

The output of the differential aberration analysis is a graphical and tabular output which allows visual inspection of the aberration amplification (colored red) and deletion (colored green) $-log_{10}$(*p*-values). The topmost pane of the Graphical Differential Aberration Summary window is a line plot of chromosomal location by $-log_{10}$(*p*-value). The enrichment values comparing the first group to the two combined groups are on the left and the enrichment values from the second group is on the right. Below the line chart is a graphical representation of the chromosomes under consideration from the comparative groups. The top row of the graphical representation of chromosomes belongs to first group and the bottom row belongs to the second group. A scroll bar at the bottom allows aberration profiles of that chromosome from all samples to be displayed. The bottom-most pane of the Summary window contains the table lists the four comparative enrichment *p*-value scores of the candidate intervals chosen from one of the aberration detection

algorithms. In addition, the chromosome, the chromosomal position, and the interval size of the candidate intervals is displayed. Clicking on an interval in the interval table switches the graphical view to display the chromosome to which the selected interval belongs. On that chromosome, a blue line is superimposed to show the location (and relative size) of the selected interval. See Figure 165.

Figure 165 Per-chromosome results of applying the Differential Aberration algorithm.

Filters can be created and applied to broaden or narrow the range of differential intervals displayed based upon *p*-value or other thresholds. For example, if you are not interested in detecting short differentially aberrant intervals, you can apply a filter (like one shown below in Figure 166) to remove all intervals that contain less than a certain number of probes. In this example, a custom filter has been created by using the attribute drop-down menu to select “No of Probes” as a filter criteria. The greater-than inequality is used as an operator on the target value 10. Additional attributes, or filtering criteria, can be added as new

conditions, and the results can either display those results passing the filter, or those specifically not passing the filter. See Figure 166. See "Interval Filter" on page 210.

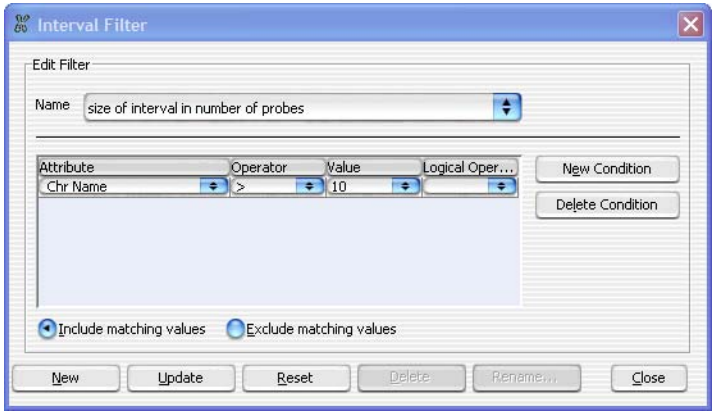


Figure 166 Creating a custom filter based on the number of probes in an interval for Differential Aberration analysis.

Applying either the default or a custom filter will change both the p-value plot for gains and losses and restrict the interval table to those intervals which pass the filter. The graphical representation of gains and losses in the chromosome diagrams will, however, remain unfiltered. This is useful for identifying regions that did not pass a filtering threshold. See Figure 167 for an example output of Differential Aberration analysis after applying the filter shown in Figure 166.

Figure 167 Per-chromosome results of applying the Differential Aberration algorithm with a custom filter to remove intervals containing less than 10 probes.

Algorithms for CGH+SNP Analysis

Agilent CytoGenomics supports the analysis of Agilent CGH+SNP microarrays, which combine both CGH probes and SNP probes on the same array. Considered together, the data from these probes allow for the determination of SNP allele-specific copy number (ASCN), the assignment of SNP genotypes, and the identification of regions of loss or lack of heterozygosity (LOH). These results can be useful in constitutional cytogenetic studies.

ASCN (SNP CN) – Allele-specific copy number detection algorithm

The Agilent CGH+SNP platform distinguishes the two alleles of a SNP by whether or not the SNP site is cleaved by the AluI/RsaI restriction enzyme mixture that is used during the sample labeling process. One allele (the "cut" allele) is cleaved by the enzymes, binds poorly to the corresponding SNP probe(s), and produces very little signal. The other allele (the "uncut" allele) is not cleaved by the enzymes, binds strongly to the corresponding SNP probe(s), and produces a high signal level. See Figure 168.

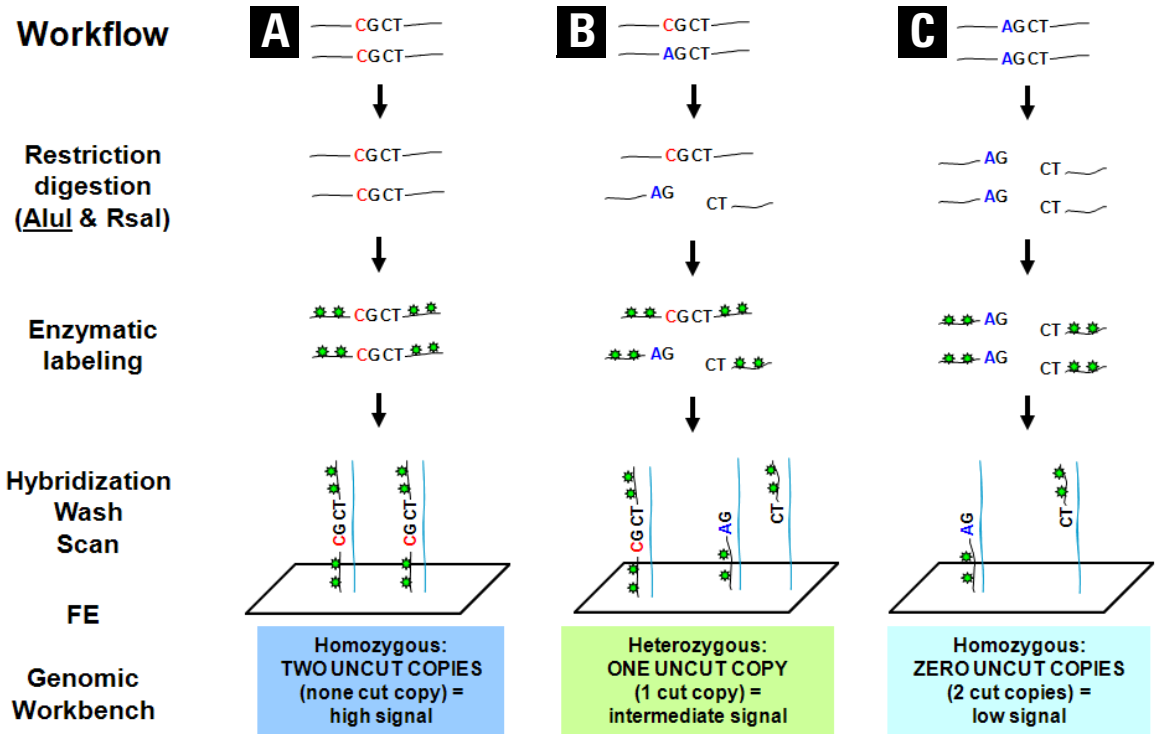


Figure 168 The Agilent CGH+SNP microarray workflow, with three possible cases for SNPs shown. In each of the three possibilities, homologous regions of two chromosomes are shown, with the bases of the SNPs in red and/or blue. **A** – Neither of the SNP sites is cut by AluI or RsaI, which lets the greatest amount of labeled material hybridize with the probes on the microarray. This yields the highest signal level. **B** – One of the SNP sites is cut by AluI or RsaI, which yields an intermediate (half) signal level. **C** – Both of the SNP sites are cut by AluI or RsaI, which yields the lowest signal level.

The ASCN detection algorithm finds the most likely copy number of the "uncut" SNP allele for each SNP that is interrogated on an Agilent CGH+SNP array. SNP calls are made from the $\log_2$ ratios of the sample signal versus the signal from a genotyped internal reference, which compensates for labeling and hybridization bias. Since the reference genotype is known, the raw $\log_2$ ratios are "reference adjusted" to values that reflect a hypothetical ASCN of 2 for the reference at all SNP sites. The details of these steps are described later in this section.

Allele-specific copy number values are used by the LOH detection algorithm. See "LOH (Loss or lack of heterozygosity) algorithm" on page 322. The output of the algorithm is also used to deduce the genotypes of SNP sites. See "Assignment of SNP genotypes" on page 319.

When you use the ASCN algorithm, you must also use Centralization and GC Correction, both of which are applied to the CGH probes on the array. See "Centralization Algorithm" on page 275 and "GC Correction Algorithm" on page 278.

Algorithm

The ASCN algorithm uses several main inputs:

- The $\log_2$ ratio values for all SNP probes on the array.
- CGH copy numbers, which are internally calculated expectation values for the copy numbers of the genomic regions including each CGH probe on the array.
- A file that contains information about the individual genotyped reference. For each SNP, this file includes the SNP ID, the SNP genotype at the site (*i.e.* the number of copies of each variant allele - see note below), and information about which SNP allele is cut by the AluI/RsaI enzyme mixture that is used in the genomic DNA labeling process. Agilent CytoGenomics supports five standard HapMap references, as well as custom references. If you have a custom reference file, you can import it. You can also use Agilent CytoGenomics to construct a custom reference file based on a prior experiment where the supported HapMap reference samples have been hybridized with the custom reference on the CGH+SNP microarray.

NOTE

Traditionally, the genotypes of SNPs are represented as two-character strings (*e.g.* "AA" or "AG"). However, when amplifications are identified, SNP genotypes will have additional characters that correspond to the extra alleles (e.g. "AAG" or "AGG").

The algorithm constructs the distribution of $\log_2$ ratio values for all SNP probes on the array, and finds the peaks of the distribution. It fits a separate Gaussian distribution to each peak, and uses a Bayesian model to calculate an expectation value for the uncut SNP allele copy number at each SNP.

Steps

Step 1 Reference correct $\log_2$ ratios

Since the reference genotype is known, the raw $\log_2$ ratios are "reference adjusted" to values that reflect a hypothetical ASCN of 2 for the reference at all SNP sites. SNPs with zero copies of the uncut allele in the reference sample are ignored in this step, and the algorithm subtracts 1 from the $\log_2$ ratios of SNPs with one copy in the reference sample.

NOTE

The correction that is applied to the $\log_2$ ratios of SNPs with one copy in the reference sample is not precisely 1.0. It is modified by the measured log ratio compression, namely the Separability QC metric.

Step 2 Find peaks in distribution

In this step, the algorithm constructs the distribution of $\log_2$ ratio values for all SNP probes on the array, and identifies the peaks in the distribution. It then assigns an uncut allele copy number to each peak. Typically, the assigned copy numbers range from 0 to the measured total copy number in the general region that includes the SNP site.

Step 3 Fit Gaussians to peaks

In this step, the algorithm uses an iterative method to fit separate Gaussian distributions to each of the peaks in the overall distribution.

Typical results from the first three steps of the algorithm appear in Figure 169.

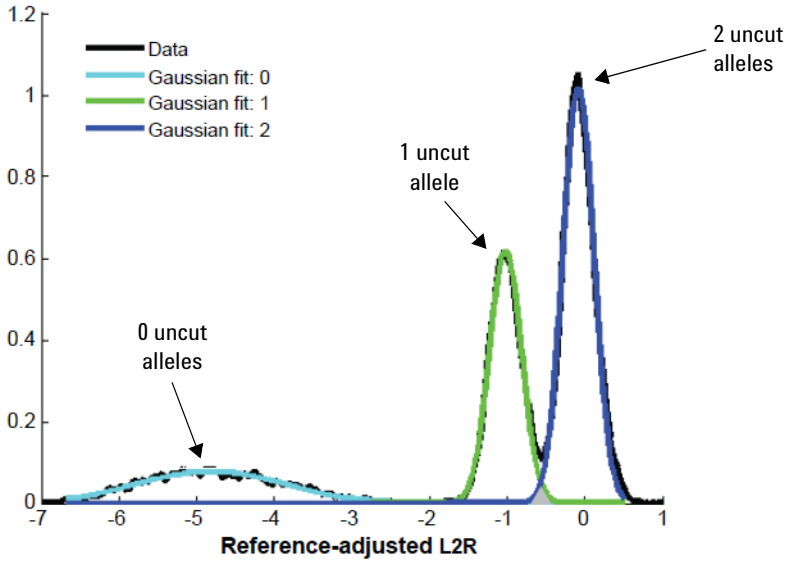


Figure 169 Example distribution of reference-adjusted $\log_2$ ratio values for SNP probes on an array (black). Gaussian distributions (blue, green, cyan) have been fit to each of the three peaks. Each peak, and thus each Gaussian distribution, has been assigned an uncut SNP allele copy number.

Step 4 Compute Bayesian likelihoods

In this step, the algorithm uses a Bayesian method to compute the likelihood that each SNP site has each possible number of uncut copies. The algorithm defines three terms for every possible combination of copy number (CN) and distinct $\log_2$ ratio value (LR) in the distribution:

- The prior probability of observing each $\log_2$ ratio value. This value, called P(LR), is obtained directly from the distribution of $\log_2$ ratio values.
- The probability that the specific $\log_2$ ratio value under consideration would be obtained given the specific copy number. This value, called P(LR|CN), is computed from the fitted Gaussian distribution that corresponds to each copy number.
- The prior probability that a given SNP has the copy number under consideration. This value, P(CN), is the area under the Gaussian curve that corresponds to each copy number.

Based on these terms, the algorithm uses Bayes's rule to calculate the likelihood that the uncut alleles would exist in each given copy number:

$$P(CN|LR) = \frac{P(CN) \cdot P(LR|CN)}{P(LR)} \quad \textbf{(17)}$$

Step 5 Adjust for outliers

In this step, the algorithm filters out outliers. Sometimes the measured log ratios of the probes violate the assumptions of the model. In particular, some probes very far from the fitted peaks have P(LR) very close to zero, which causes the model to produce a very high, spurious result (a division by zero in Equation 17). Also, since the distribution for CN=0 is much wider than the other distributions, probes reporting high $\log_2$ ratio values sometimes have a P(CN=0) that is higher than their P(CN=2). Both of these types of outliers are assigned an ASCN by linear interpolation, rather than by the Bayesian Equation 17.

Step 6 Rescue probes

Previously, in Step 1, the algorithm excluded data from probes targeting doubly-cut SNP sites of the known genotyped reference. In this step, the algorithm rescues the genotypes of some of the probes that were previously excluded.

SNPs can often be assigned an ASCN of zero based on the probe intensity alone, even when (as is the case when the ASCN of the reference sample is zero) the log ratio cannot be interpreted. Gaussian curves are fitted to the sample $\log_2$(signal) levels of SNP probes already assigned to ASCN=0 or ASCN<>0. Unassigned SNP probes which can be confidently assigned to the ASCN=0 distribution, based on their signal alone, are assigned a likelihood of having a copy number of zero. The confidence value of the probe is then reevaluated based on the likelihood that it represents a copy number of zero.

Step 7 Calculate ASCN expectation values

In the final step, the algorithm uses the likelihood values to calculate an expectation value, $<C_p>$, for the uncut allele copy number for each SNP. SNPs whose maximum likelihood exceeds a user-specified threshold (0.95 recommended) are assigned a genotype based on the maximum likelihood ASCN. See "Assignment of SNP genotypes" on page 319. If no ASCN exceeds the likelihood threshold, the SNP genotype is not reported.

Visualization

When you apply the ASCN algorithm, ASCN assignments appear in the SNP CN panel in Gene View and in Chromosome View. The calculated expectation values for the ASCN of the SNPs are plotted against chromosomal location. A typical SNP array represents many different SNPs that cover all states of uncut SNP allele copy numbers. Because of this, separate bands of points are seen in the SNP CN panel for each possible uncut SNP allele copy number. See Figure 170.

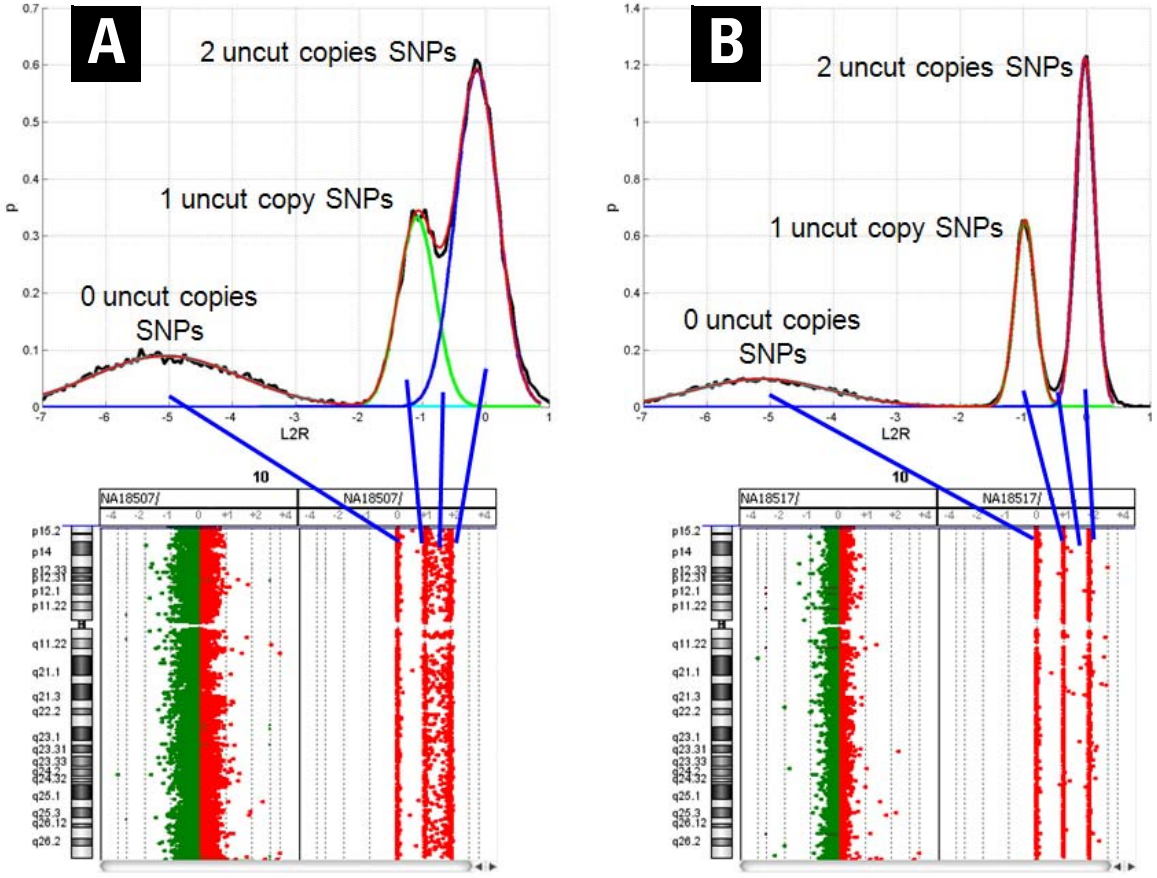


Figure 170 Two examples of SNP copy number calls. **A** – Upper plot: Copy number distribution plot for a sample with a poor derivative log ratio spread (DLRS = 0.33). Note how the peaks for copy numbers of 1 and 2 are not well-differentiated. Lower plot: Gene View. The right panel is the SNP CN panel. ASCN expectation values for probes are plotted against chromosomal location. Note how allele-specific copy number calls are not well defined, especially for copy numbers of 1 and 2.
B – Upper plot: Copy number distribution plot for a sample with a good derivative log ratio spread (DLRS = 0.15). Note the well-separated peaks for all copy numbers. Lower plot: Gene View. The right panel is the SNP CN panel. ASCN expectation values for probes are plotted against chromosomal location. Well-separated bands of points are seen for each uncut SNP allele copy number.

Assignment of SNP genotypes

The program reports the genotypes of the SNP sites targeted by SNP probes. To do this, it uses SNP allele information from the reference file, and expectation values for CGH copy number and ASCN. Examples of the logic that the program uses appear in "Example 1" on page 320 and "Example 2" on page 321.

To deduce the genotypes of SNP sites, the algorithm requires integer values for CGH copy numbers and ASCNs. To obtain these values, the algorithm rounds the measured expectation values for CGH copy number and ASCN (both of which can contain fractional values) to the nearest integer. Copy numbers are excluded from the analysis if the amount of rounding that is required is too great.

To assess the amount of rounding, and to decide whether or not to include a given copy number value in the SNP genotype analysis, the algorithm uses a threshold value that is entered by the user. It includes a given copy number value in the analysis if the following relation is true:

$$|\text{ASCN} - \text{rounded ASCN}| < (1 - \text{threshold value})$$

If the CGH copy number value and/or the ASCN value for a given region are not reported, the algorithm reports both alleles as "N" (not determined).

For genotype assignment and for visualization, you can set a threshold value for the confidence level for ASCN calls (0.95 recommended).

NOTE For the LOH algorithm, all SNP probes are considered without regard to the confidence threshold that you set. This threshold applies only to the visualization of copy number calls in the SNP CN panel, and to genotype assignment.

Example 1 In the example that appears in Figure 171, a given reference SNP site has a genotype of CT (a C base as one allele, and a T base as the other). Further, it is known that the C allele is the one that remains uncut by the AluI/RsaI enzyme mixture during sample preparation. If, for the sample, the CGH copy number for the region is 2, and the uncut allele copy number (ASCN) for the site is 2, the SNP genotype for the sample is CC.

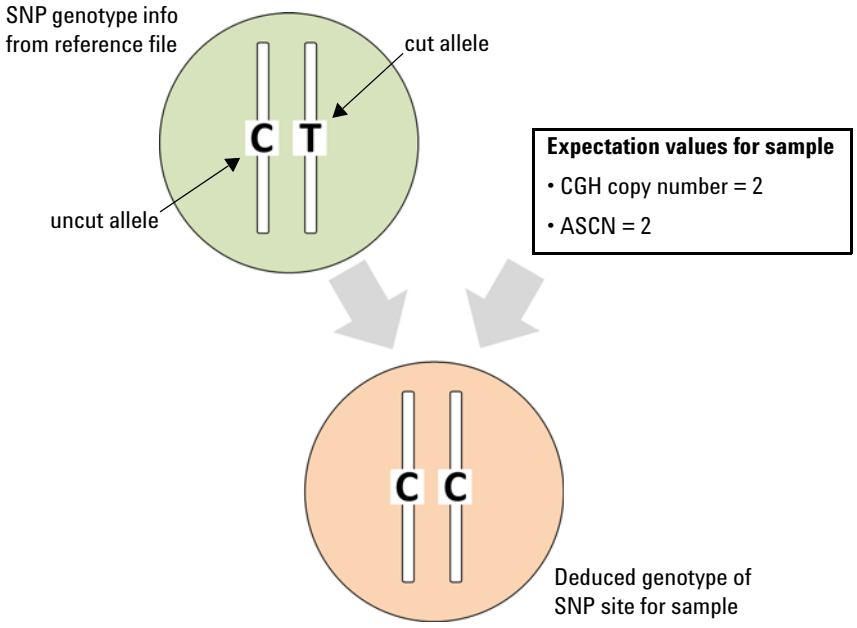


Figure 171 Reference genotype and deduced sample genotype at a specific SNP site. The CGH copy number of 2 indicates a total of two alleles in the sample at the SNP site, and the ASCN of 2 indicates that two alleles are the "uncut" C allele.

Example 2 As in the previous example, the example that appears in Figure 172 shows a genotype of CT for the reference SNP site, and the allele that remains uncut is C. For the sample, the CGH copy number is 4, and the uncut allele copy number (ASCN) is 2. The sample thus has two C alleles. Because SNPs are typically bi-allelic, the other two alleles must be the T allele.

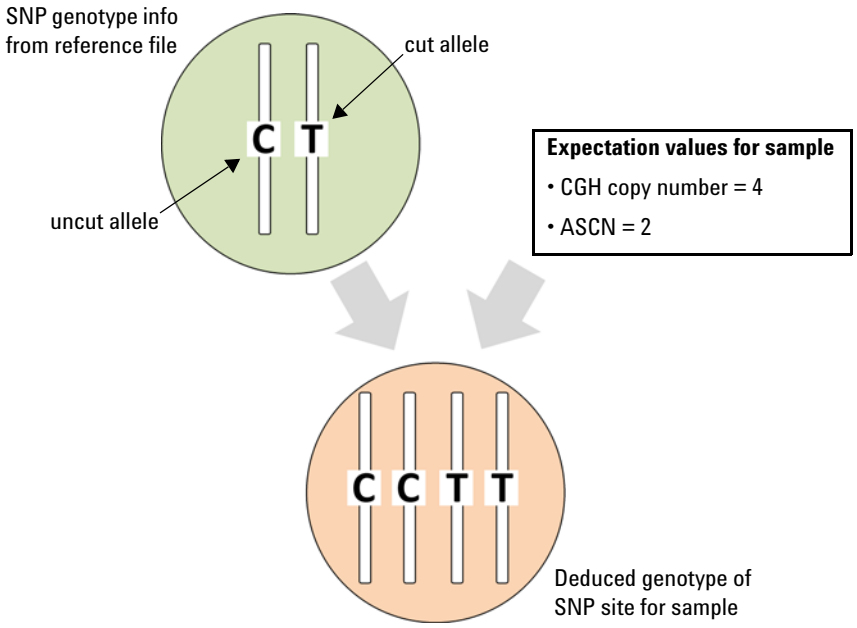


Figure 172 Reference genotype and deduced sample genotype at another specific SNP site. The CGH copy number of 4 indicates a total of four alleles in the sample at the SNP site, and the ASCN of 2 indicates that two of the alleles in the sample are the "uncut" C allele.

NOTE The new dbSNP databases record a small number of SNPs with more than two different alleles. The Agilent assay still works for these sites, with some reinterpretation.

You can inspect or use the reported genotypes in several ways:

- The reported genotypes of SNP sites for the selected experiment appear in the Genotype column in the SNPs tab of Tab View. See "Tab View tabs and buttons" on page 51.
- You can create a SNP Genotype Report that contains a summary of analysis parameters and list(s) of reported SNP genotypes.
- You can create a custom genotype reference file that you can import into Agilent CytoGenomics to use as a reference for future SNP analyses.

LOH (Loss or lack of heterozygosity) algorithm

For the SNP sites that are interrogated by Agilent SNP probes on a CGH+SNP array, the LOH algorithm identifies genomic regions with a statistically significant scarcity of heterozygous SNP calls to discover regions of copy-neutral LOH and uniparental disomy (UPD). Identification of such regions can inform constitutional cytogenetic studies.

The algorithm uses total and allele-specific copy numbers to label each SNP site as homozygous or heterozygous. It then uses a binomial probability distribution to report regions that contain an unusually high fraction of homozygous SNPs.

When you use the LOH algorithm, you must apply both Centralization and GC Correction. See "Centralization Algorithm" on page 275 and "GC Correction Algorithm" on page 278. You must also apply the ASCN algorithm, since the LOH algorithm requires ASCN calls as input. See "ASCN (SNP CN) – Allele-specific copy number detection algorithm" on page 312.

The algorithm reports LOH for the regions that are also detected as deletions in standard CGH analysis (see Figure 173). Amplified regions, while not detected as LOH, are visible in the results of CGH aberration analysis, and can also be seen in the SNP CN pane (see Figure 174). Regions of copy-neutral LOH, such as those that arise from UPD and parental consanguinity, are detected by the LOH algorithm, but are not detected by CGH analysis. See Figure 175 and Figure 176.

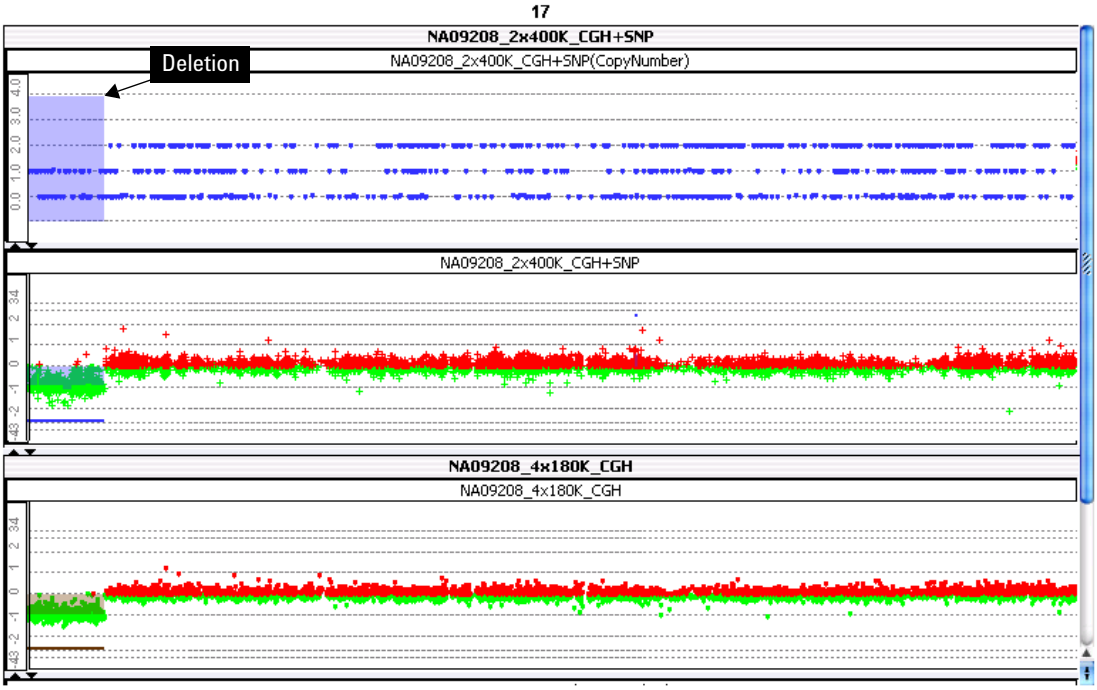


Figure 173 Hemizygous deletion in chromosome 17. **Upper pane** – Deletion identified by the LOH algorithm (purple shaded area) and ASCN calls for SNP probes (blue dots). In the region of the deletion, note the significant absence of probes that correspond to a copy number of 2. **Middle pane** – Log_2 ratio data for the CGH+SNP array. The same deletion is detected by aberration analysis of the CGH probes on the array. **Lower pane** – Log_2 ratio data for the same sample on a standard CGH array. Again, the same deletion is detected by CGH aberration analysis.

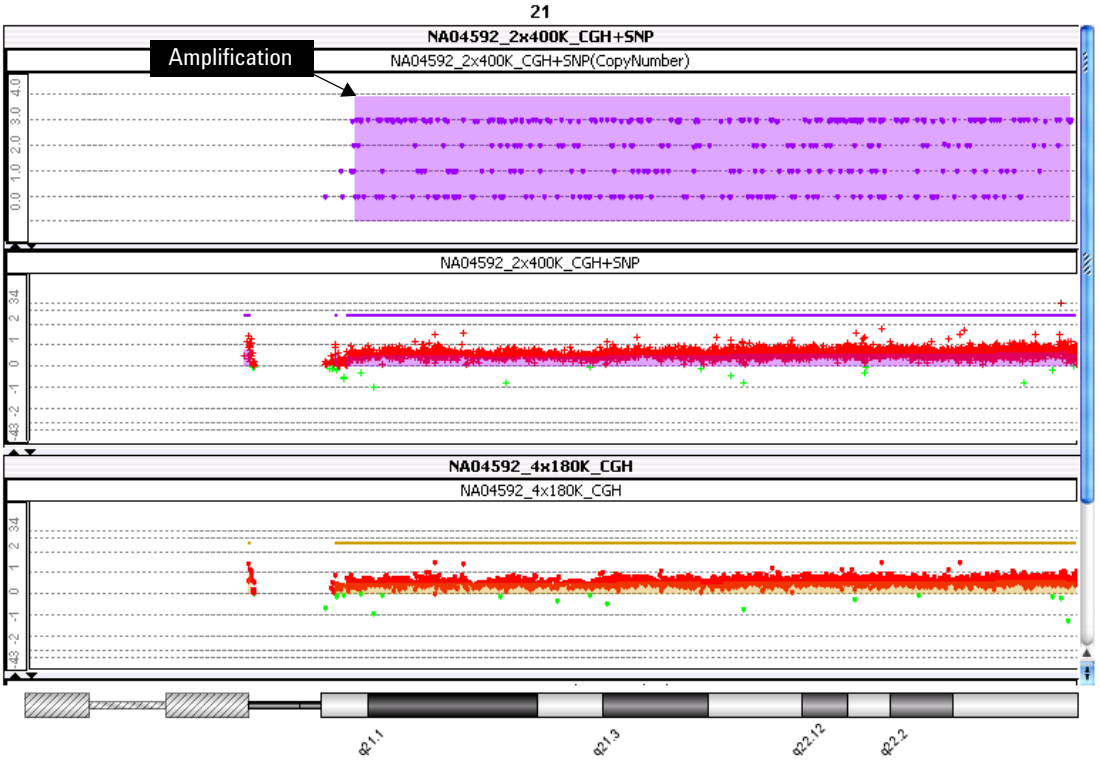


Figure 174 Amplification (trisomy) on human chromosome 21 (purple shaded area). **Upper pane** – ASCN calls for SNP probes (purple dots). Note the significant band of probes that corresponds to a copy number of 3. **Middle pane** – Log_2 ratio data from the CGH+SNP array. The same amplification is detected by aberration analysis of the CGH probes on the array. **Lower pane** – Log_2 ratio data for the same sample on a standard CGH array. Again, the same amplification is detected by CGH aberration analysis.

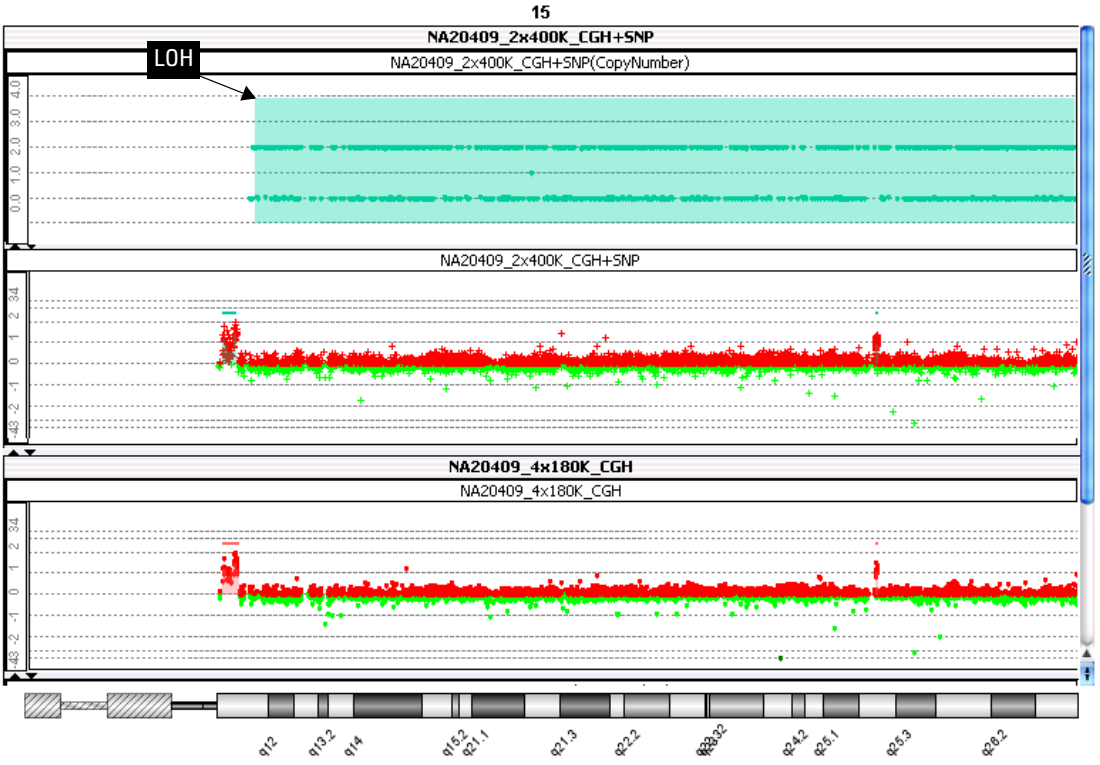


Figure 175 LOH caused by UPD of the entire chromosome 15. **Upper pane** – ASCN calls for SNP probes (green dots). In the region of the detected LOH (green shaded area), note the lack of probes that correspond to a copy number of 1. **Middle pane** – Log_2 ratio data from the CGH+SNP array, with an aberration analysis of the CGH probes on the array. The CGH aberration analysis does not detect the region of copy-neutral LOH caused by the UPD. **Lower pane** – Log_2 ratio data for the same sample from a standard CGH array. Again, the copy-neutral LOH is not detected by standard CGH analysis.

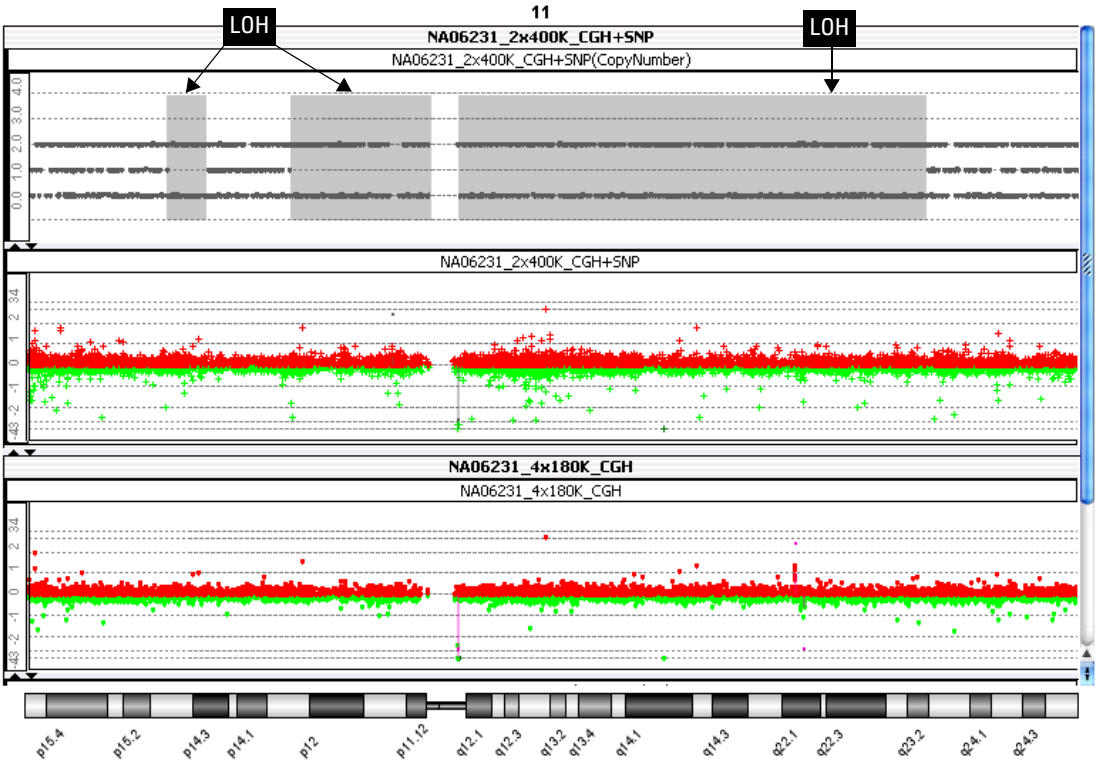


Figure 176 LOH in chromosome 11 caused by parental consanguinity. **Upper pane** – ASCN calls for SNP probes. In the regions of LOH (gray shaded regions), note the lack of probes that correspond to a copy number of 1. **Middle pane** – Log_2 ratio data from the CGH+SNP array, with an aberration analysis of the CGH probes on the array. The CGH aberration analysis does not detect the copy-neutral LOH regions. **Lower pane** – Log_2 ratio data for the same sample from a standard CGH array. Again, the copy-neutral LOH regions are not detected by standard CGH analysis.

Algorithm

The LOH algorithm uses two main inputs:

- The expectation values for CGH copy number. This value is calculated for every CGH aberration interval and non-aberrant interval on the array.
- The expectation value for the uncut SNP allele copy number (ASCN) for each SNP site under study in the sample genome. These values are the output of the ASCN algorithm. See "ASCN (SNP CN) – Allele-specific copy number detection algorithm" on page 312.

For each SNP site that is interrogated on the array, the algorithm deduces, to the extent possible, whether the given SNP is homozygous or heterozygous. The algorithm scans the SNP sites in the genome and uses an iterative, stepwise expansion and comparison method to develop putative LOH regions. This part of the algorithm is similar to the ADM-1 aberration detection algorithm. See "ADM-1" on page 290. As it develops these regions, the algorithm calculates an LOH score for each region after each expansion. The algorithm reports a final LOH call for genomic regions whose scores exceed a user-defined threshold (default 6). The highest scoring of overlapping LOH regions is reported.

Step 1 Calculate LOH scores

To calculate an LOH score for a given region, the algorithm first assigns each SNP as either homozygous or heterozygous. It assumes a binomial distribution, and calculates the probability that the given set of SNP sites forms an LOH region that is not just a result of random chance. For a region that contains n total SNP sites, x of which are homozygous, the algorithm first calculates the probability that x or fewer of n sites would be homozygous as a result of random chance:

$$F(x, p, n) = \sum_{1}^{x} P(x, p, n) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{(n-x)}, \quad \textbf{(18)}$$

where F is the cumulative distribution function, P is the probability distribution function for the binomial distribution, and p is the prior probability that a given SNP is heterozygous. p is estimated from the genome-wide fraction:

$$p_{(\text{heterozygous})} = \frac{n_{\text{heterozygous}}}{n_{\text{heterozygous}} + n_{\text{homozygous}}}, \quad \textbf{(19)}$$

where n is the total number of SNP sites in each given state. If the value of p is less than 0.22, the algorithm sets p to 0.22.

Smaller values of F signify a lower likelihood that this event happens as a result of random chance. The LOH score is the negative common logarithm of this probability:

$$\text{LOH score} = -\log_{10}[F(x, p, n)] \quad \textbf{(20)}$$

Step 2 Separate regions if needed

In this step, the algorithm examines each putative LOH region. If a region contains a statistically significant number of consecutive heterozygous SNPs, that region is separated into two smaller regions. Each of the resulting smaller regions is then scored again.

Step 3 Make final LOH calls

As the algorithm develops putative LOH regions, it makes a final LOH call when the score for a given region falls below a user-specified threshold value. You enter the threshold value when you set up LOH analysis. Agilent recommends that you use a value of 6.0 as a starting point for this threshold value for the LOH score. A higher value makes the algorithm more stringent (*i.e.* it will report fewer false positives, but also fewer true positives).

Visualization

When you apply the LOH algorithm in Agilent CytoGenomics, LOH calls can appear in the Genomic Viewer in Genome View (Figure 177), Chromosome View (Figure 178), and Gene View (Figure 179 and Figure 180). In addition, you can create a SNP Aberration & LOH Report that contains a summary of analysis parameters and a list of all aberrant and LOH intervals.

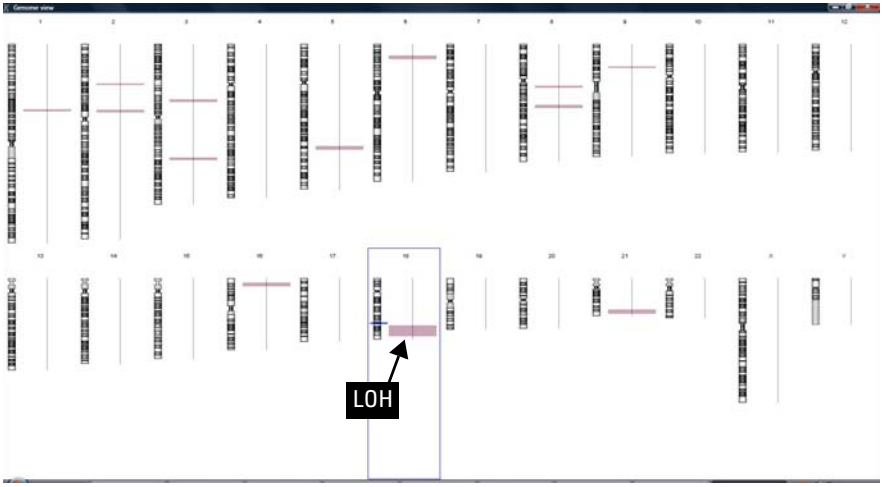


Figure 177 Visualization of LOH regions in Genome View in the Genomic Viewer of Agilent CytoGenomics. LOH calls appear as shaded areas next to the relevant chromosomes.

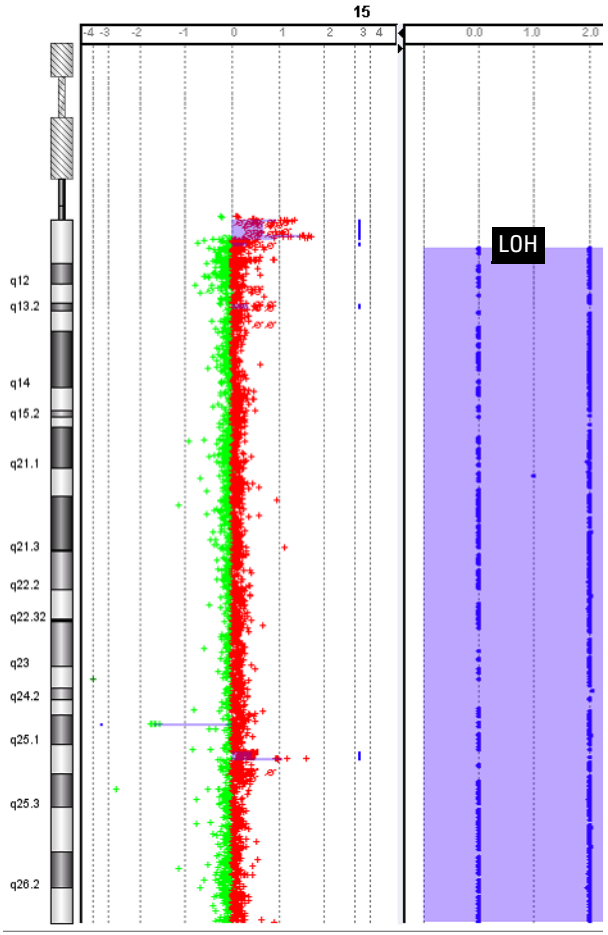


Figure 178 Chromosome View of the Genomic Viewer in Agilent CytoGenomics. **Left pane** – Log_2 ratio data and aberration calls on chromosome 15 based on data from the CGH probes on a CGH+SNP array. **Right pane** – SNP CN panel, with ASCN calls (blue dots) for SNP sites. An LOH call appears as a large purple shaded area. Note the absence of sites with a copy number of 1.

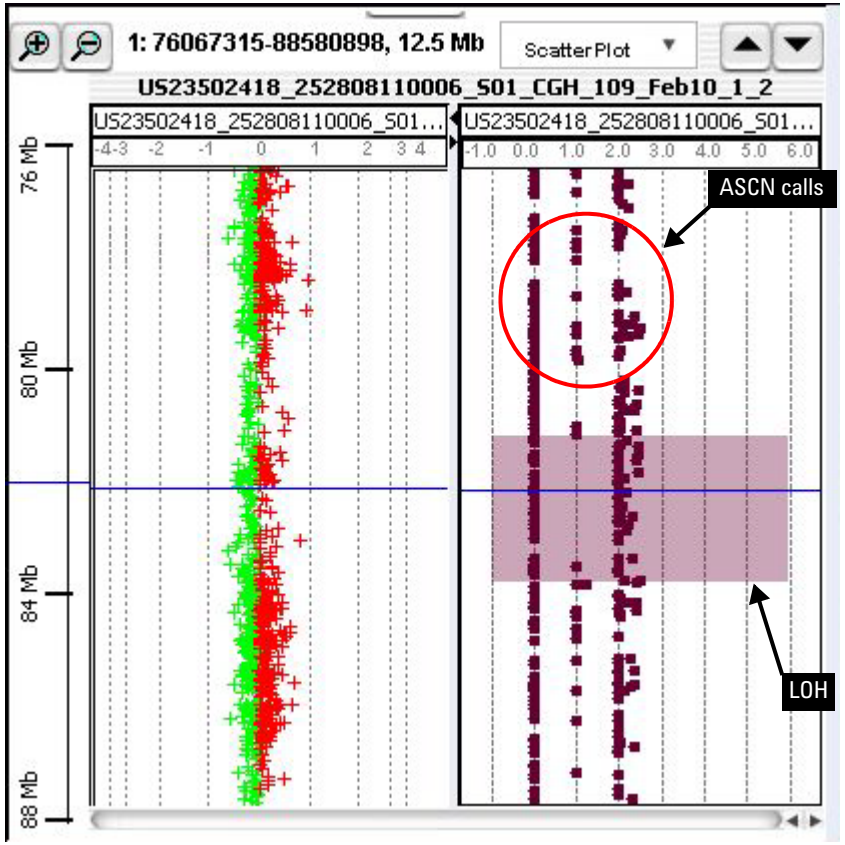


Figure 179 Gene View of the Genomic Viewer in Agilent CytoGenomics, showing a 12.5 Mb region of chromosome 1. **Left pane** – Log_2 ratio data from a CGH+SNP array. **Right pane** – ASCN calls for SNP sites. LOH call appears as a mauve shaded region. In the LOH region, note the absence of sites with a copy number of 1.

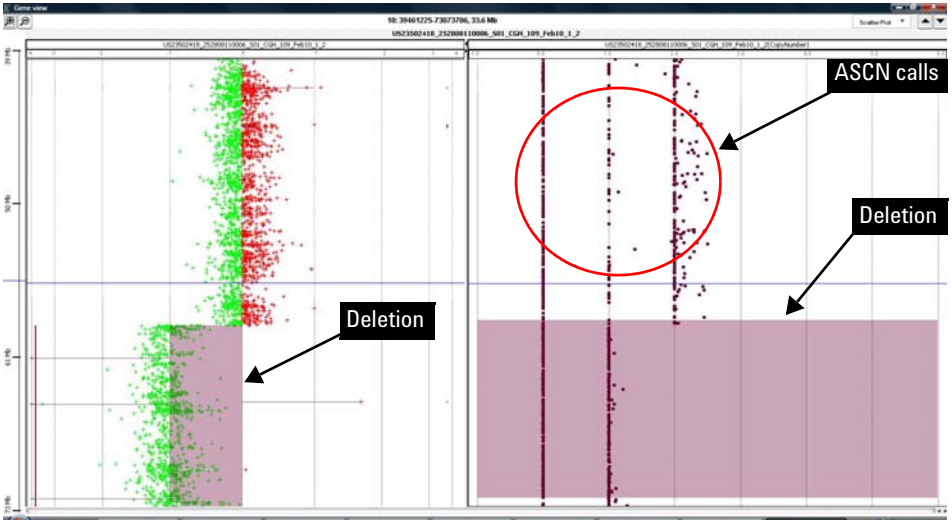


Figure 180 Gene View of the Genomic Viewer in Agilent CytoGenomics, showing a 33.6 Mb region of chromosome 18. A hemizygous deletion affects part of chromosome 18. **Left pane** – Log_2 ratio data from the CGH+SNP array. Deletion call, based on data from the CGH probes on the array. **Right pane** – SNP CN panel, with ASCN calls for the same region, with the analogous deletion call made by the LOH algorithm. In the aberrant region, note the lack of SNP sites with a copy number of 2.

Appendix

This section contains additional information about statistical algorithms and is useful for understanding algorithm steps in detail.

Enrichment Analysis - the Hypergeometric distribution

The Hypergeometric distribution is a generalization of the binomial distribution. The key difference between the two distributions is that the Hypergeometric distribution samples without replacement and is used to calculate the overabundance, or *enrichment* of a certain event occurring, given the constraints of the total number of matching events possible. In this way, the Hypergeometric distribution addresses a question of conditional probability - the probability that one event occurred, given that another event has occurred within the groups under consideration.

Because of the similarity to the binomial distribution, it is familiar to use terms used to describe binomial events - mainly, the concept of the success or failure of an event to occur. The first step to enrichment analysis using the Hypergeometric distribution is to partition trials into successes and failures based upon a given condition. For example, one might wish to measure the probability of an aberration event occurring (a success), where the probability itself (usually called p) is a measure of the likelihood of observing the number of events under a certain condition compared to observing the same number of events in a random model. Another experimental parameter might then be considered, such as the disease state of the subject, in order to ascertain whether or not the probability of the number of observed mutations is increased with a certain disease. This is called an enrichment analysis, and such an analysis asks whether or not the observed number of mutations is enriched in a certain group. Other factors might be the sex of the subject, the age of the subject, the medical history, and so forth.

To demonstrate the derivation of the Hypergeometric distribution, an enrichment analysis such as that used in the Differential Aberration algorithm (see "Differential Aberration Analysis") begins by partitioning a group of samples based upon whether or not an event is observed in that group. For N total subjects, assume that n aberrations are observed within a genomic interval in disease-free subjects (preselected by some criteria),

and therefore N-n aberrations are observed within the same interval in diseased subjects. From these, assume that the total number of all aberrations within a given genomic interval is m, and that therefore are N-m absence of detected aberrations in the same interval. An aberration can be either an amplification (gain) or deletion (loss), so the calculations will have to be repeated to assess enrichment for both scenarios.

The probability of observing y aberrations within healthy subjects for a given interval can therefore be described using the binomial distribution:

$$P_Y(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad \textbf{(21)}$$

The probability of observing z aberrations within diseased subjects is also described using the binomial distribution:

$$P_Z(z) = \binom{N-n}{m-y} p^{m-y} (1-p)^{N-m-(m-y)} \quad \textbf{(22)}$$

And the probability of observing m total aberrations is derived from the binomial distribution:

$$P_M(m) = \binom{N}{m} p^m (1-p)^{N-m} \quad \textbf{(23)}$$

The probability of observing the number of aberrations from healthy subjects within the interval, given the probability of observing the number of aberrations in the total sample for that same interval is given by the conditional probability property:

$$P(Y_y | M_m) = \frac{P(Y_y M_m)}{P(M_m)} \quad \textbf{(24)}$$

Equation 24 allows use of the probability of observing the number of successes and failures in each of the partitions from the samples (the healthy and diseased groups), normalized by the probability of observing this aberration in the population:

$$P(y|m) = \frac{\binom{n}{y} p^y (1-p)^{n-y} \bullet \binom{N-n}{m-y} p^{m-y} (1-p)^{N-n-(m-y)}}{\binom{N}{m} p^m (1-p)^{N-m}} \quad \textbf{(25)}$$

Rearranging Equation 25 yields:

$$P(y|m) = \frac{\binom{n}{y}\binom{N-n}{m-y}p^m(1-p)^{N-m}}{\binom{N}{m}p^m(1-p)^{N-m}} \quad (26)$$

Equation 26 cancels to yield the Hypergeometric distribution, which is a quick way to calculate enrichment independently of the binomial probabilities, and instead entirely upon the binomial coefficients:

$$P(y|m) = \frac{\binom{n}{y}\binom{N-n}{m-y}}{\binom{N}{m}} \quad (27)$$

Because this calculation is a distinct probability, the calculation is usually summed across all number of observations equal or greater than that of the number of aberrations in the interval from the partitioned sample to yield a p-value, which is known as the Hypergeometric tail score (H.G.T.), or simply the probability under the extreme tail of the Hypergeometric distribution:

$$HGT = p-value = P(Y \geq y|m) = \sum_{y}^{m} \frac{\binom{n}{y}\binom{N-n}{m-y}}{\binom{N}{m}} \quad (28)$$

The hypergeometric tail probabilities reflect the probability of observing y or more number of gains or losses in a group in this example - *i.e.* it is a measure of the enrichment of the gains or losses in a group.

References

1 Diskin, Sharon J., Mingyao Li, Cuiping Hou, Shuzhang Yang, Joseph Glessner, Hakon Hakonarson, Maja Bucan, John M. Maris, and Kai Wang 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* 36:e126. http://nar.oxfordjournals.org. doi:10.1093/nar/gkn556.

2 Hedenfalk, I. et. al. (2003) Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc. Natl. Acad. Sci. USA,* 2003 Mar 4;100(5):2532-7.

3 Kincaid, R., A. Ben-Dor, and Z. Yakhini. Exploratory visualization of array-based comparative genomic hybridization. *Information Visualization*, 2005, 4:176-190. doi:10.1057/palgrave.ivs.9500095.

4 Lipson, Doron, Yonatan Aumann, Amir Ben-Dor, Nathan Linial, and Zohar Yakhini. Efficient Calculation of Interval Scores for DNA Copy Number Data Analysis. Proceedings of Recomb, 2005.

5 Olshen, A. B., Venkatraman, E. S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.*" **5**, 4, 557-572.

6 E. S. Venkatraman and Adam B. Olshen. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*, **23(6)**: 657-663.

7 JV Braun, HG Muller (1998), Statistical methods for DNA sequence segmentation. *Statistical Science,* 13(2):142-162.

8 E. S. Venkatraman and Adam B. Olshen. (2005). DNAcopy: A Package for Analyzing DNA Copy Data. *The Bioconductor Project*: http://www.bioconductor.org.

9 Stanton A. Glantz. Correlation and Correlation Coefficients. *Primer of Biostatistics, 5th Edition*. McGraw-Hill, 2002. 262-265.

10 Pierre Baldi and G. Wesley Hatfield. Statistical analysis of array data: Dimensional reduction, clustering, and regulatory regions. *DNA Microarrays and Gene Expression.* Cambridge University Press, 2002. 78-84.

11 Lipson D, Aumann Y, Ben-Dor A, Linial N, Yakhini Z. (2006) Efficient calculation of interval scores for DNA copy number data analysis. *J. Comput Biol.* 2006 Mar;13(2):215-28.

12 Ben-Dor, A., Lipson, D., Tsalenko, A., Reimers, M., Baumbusch, L., Barrett, M., Weinstein, J., Børresen-Dale, A.-L. and Yakhini, Z. (2007). Framework for identifying common aberrations in DNA copy number data. Research in Computational Molecular Biology, Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, Vol. 4453.

13 Rice, John A. (1995) The Hypergeometric Distribution. *Mathematical Statistics and Data Analysis, 2nd Edition.* Duxbury Press, 1995, 39-40.

14 Chaya Ben-Zaken Zilberstein et al., *Lecture Notes in Computer Science: Regulatory Genomics: RECOMB 2004 International Workshop, Revised Selected Papers* in Eskin E. and Workman, C. (eds.), Springer Berlin / Heidelberg 2005(3318). ISBN: 3-540-24456-5

www.agilent.com

In this book

This guide describes the tabs, windows, parameter panels, dialog boxes, and reports you see in the Agilent CytoGenomics 1.0 software. For information on how to use the software to set up and run workflows and examine results, see the Setup and Data Review User Guide and the Running CytoGenomics Analyses User Guide. For information on installing and administering the software, see the Installation and Administration Guide.

Revision A0, November 2010

G2602-90005

Agilent Technologies