Agilent Genomic Workbench Lite Edition 6.0

Methylation (CH3) Analysis

User Guide

Agilent Technologies

Notices

Manual Part Number

G3800-90017

Edition

Revision A, May 2010

Printed in USA

Agilent Technologies, Inc.
3501 Stevens Creek Blvd.
Santa Clara, CA 95051

Trademarks

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

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This guide is valid for 6.0 and later revisions of the Agilent Genomic Workbench Lite Edition 6.0 software, until superseded.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 how to use the Methylation (CH3) application of Agilent Genomic Workbench Lite Edition 6.0 to apply algorithms that help identify methylated regions.

1 Getting Started

This chapter gives instructions on how to start the application and enter license information. It also gives an overview of how to analyze CH3 data.

2 Importing, Managing, and Exporting CH3 Data and Other Content

This chapter describes how to import, organize, manage, and export CH3 data and other content in Agilent Genomic Workbench Lite Edition.

3 Displaying CH3 Data and Other Content

This chapter shows you how to display log ratio data from imported feature extraction data files and analysis results, as well as gene list and track content, in the Genomic Viewer. It also gives you instructions on how to change parameters to display the data and content the way you prefer.

4 Setting Up Methylation (CH3) Analysis

This chapter gives instructions on how to set up the analysis functions for Methylation (CH3) experiments. These include the Preprocessing, Analysis and Reports tabs.

5 Methylation (CH3) Analysis Reference

This chapter describes the tabs, commands, shortcut menus and dialog boxes specific to Agilent Genomic Workbench methylation (CH3) data analysis. The chapter also includes information on the format of the reports created by the program.

6 Statistical Algorithms

This chapter provides implementation details of the algorithms used in the CH3 application of Agilent Genomic Workbench Lite Edition. The CH3 application algorithms facilitate the statistical analysis of methylated genomic regions. The first section presents the main methylation detection algorithm. The next section describes the display options available for methylation analysis.

Contents

1 Getting Started

This chapter gives an overview of the Methylation (CH3) application of Agilent Genomic Workbench Lite Edition. It shows you how to start the application and find help, and gives general instructions on how to get started with analyzing CH3 data.

Before or after you import extracted data into the program, you can assign identification information and attributes to the samples through the Sample Manager tab. See the *Sample Manager User Guide*.

To display or analyze imported CGH data, you organize the data files into logical units called *experiments*. Experiments are used to define the data you want to display or analyze using Agilent Genomic Workbench. You can then use the Preprocessing, Analysis, Discovery and Reports tabs of the program to interactively analyze the data in the experiment for aberrations. For detailed steps on how to create an experiment, see the *Data Viewing User Guide*.

What is the CH3 Application?

The CH3 (methylation) application is a framework used to identify methylation events in your samples. CH3 microarray analysis can help identify methylated regions isolated using affinity based methods such as methylated DNA immunoprecipitation. The software helps you:

- Identify molecular events associated with DNA methylation
- Find and validate gene regulation and regulatory networks by creating high-resolution, genome-wide methylation profiles
- Show modes of action and potential therapeutic activities of compounds and target genes by understanding the relation of DNA methylation to transcriptional control

With the CH3 application, you can:

- Import data from the Agilent Feature Extraction and Axon programs, and import UDF files
- Use an intuitive graphical interface to display data and annotations in the context of an organism's genome, at several simultaneous levels of detail
- Use a Z-score or BATMAN (Bayesian Tool for Methylation Analysis) algorithm to calculate the probability that probes are methylated or unmethylated
- Compare the moving average of your log ratio and Z-score data within the boundaries of CpG Island tracks

You cannot use workflow mode for methylation event detection.

Using Agilent Genomic Workbench Lite Edition on a Mac

The content of this User Guide applies to both the Windows and Mac versions of Agilent Genomic Workbench Lite Edition. Both of these versions have the same features. However, when you use the Mac version of the program, please note the following:

Windows command	Equivalent Mac command
Right-click	• Command-click (⌘ -click) • On Macs with trackpads, other options are available. On certain machines, you place two fingers on the trackpad while you press the button below the trackpad. See the user guide for your specific machine. • If you have a third-party mouse that has more than one button, you may be able to use one of the buttons as a right mouse button.
Control-click	Control-click (Same as the Windows command)
Shift-click	Shift-click (Same as the Windows command)
 (Close button)	 (Close button)

Entering a License and Starting the CH3 Application

This section describes how to enter your license for the CH3 application to begin analyzing CH3 data.

When you start Agilent Genomic Workbench for the first time, the program opens in the **Home** tab, with the **Open Application** tab displayed. From this tab, you can click any of the application areas, enter license information, or click **Help** to open the User Guide for that application.

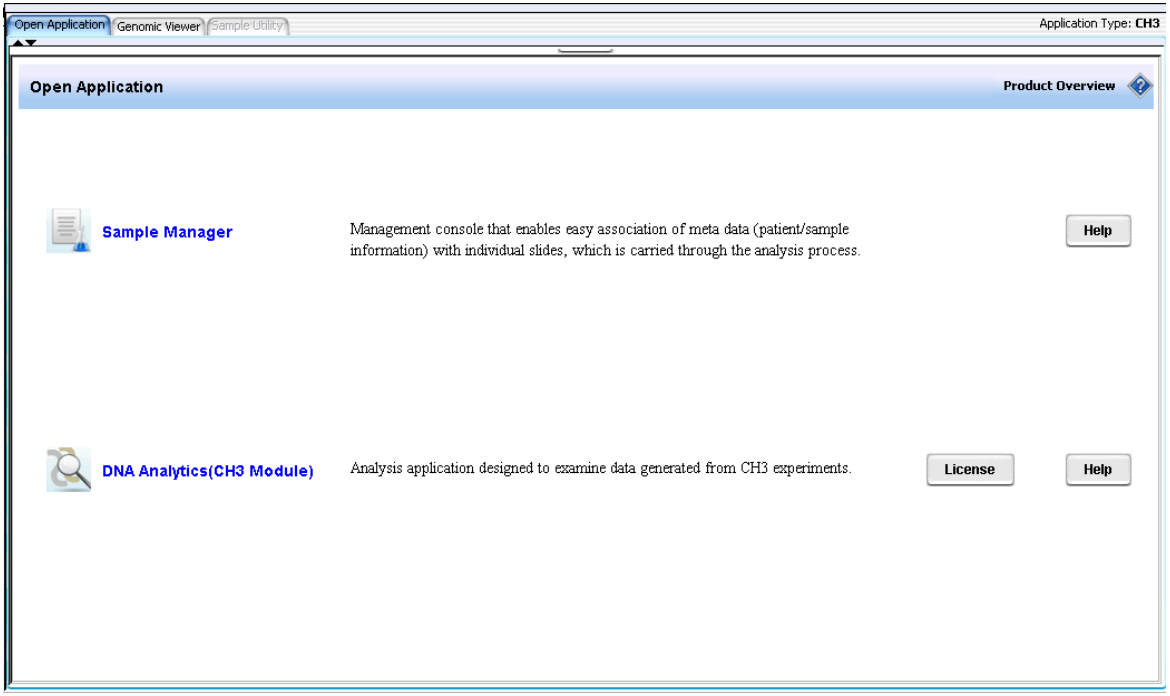


Figure 1 Open Application tab for CH3

To enter your license for analyzing Methylation (CH3) data interactively

1 Click the **Open Application** tab if it is not already displayed.

2 Click **License** next to the description of DNA Analytics (CH3 Module). The License tab of the User Preferences dialog box appears.

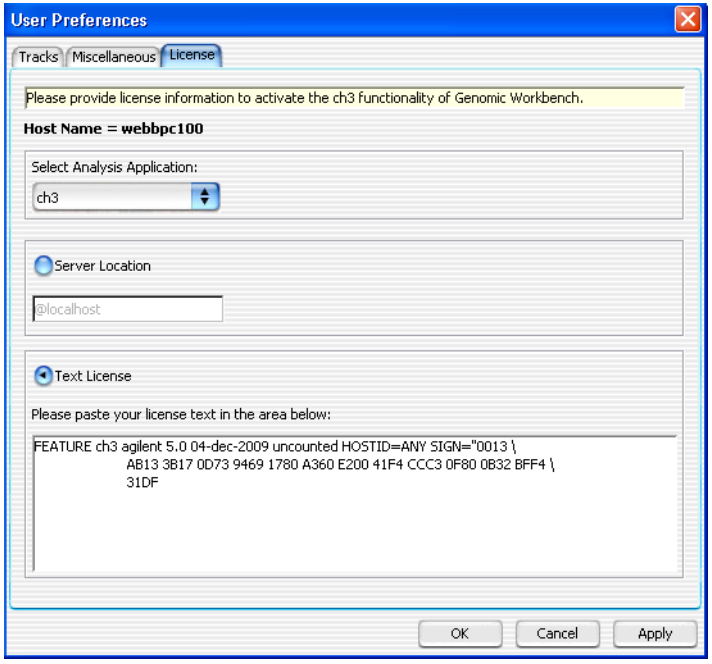


Figure 2 License tab of the User Preferences dialog box

There are two ways to provide the license information:

Use a Server Location

1 Unzip the license .txt file into a folder on your server, to which the program has access.

2 Copy the path for that folder to the Clipboard.

3 In the User Preferences License tab, click **Server Location**.

4 Paste the license folder path into the field below Server Location. (To paste the license for both Windows and Mac computers, hold down the **ctrl** key and press **V**.)

5 Click **Apply**, or click **OK** to apply the license and close the dialog box.

Enter a Text License

1 Find the folder that contains the CH3 application license .txt file.

2 Double-click the license name to open the file in Notepad (or open the file in another text editor), and copy the text displayed to the Clipboard.

3 In the User Preferences License tab, click **Text License**.

4 Paste the license information into the License text box. (To paste the license for both Windows and Mac computers, hold down the **ctrl** key and press **V**.)

5 Click **Apply**.

6 If you have no other licenses, click **OK**.

OR

If you have another license, click the arrow from the Select Analysis Application list, select the application and repeat steps 1-5.

To start the CH3 application

- In the Open Application tab, click the icon next to **DNA Analytics (CH3 Module)**  .

  The CH3 application starts and the Genomic Viewer is displayed.

Using Main Window Components to Display/Analyze Data

You can use the data *viewing* capability in Agilent Genomic Workbench Lite Edition without a license to view data for many types of arrays, including CGH, ChIP, and Methylation (CH3). You can use the data *analysis* capability in Agilent Genomic Workbench Lite Edition only if you have a license for one or more of the DNA Analytics programs (CGH, ChIP, or Methylation).

What are the main window components?

You use four primary components of the Agilent Genomic Workbench main window to import, manage, export, display, and analyze extracted data.

- Home tab commands – import, manage and export data
- Navigator – create and fill new experiments with array data

  When you make the experiment active, the data appear in the display, called Genomic Viewer.
- Genomic Viewer – display data and content in four Views: Genomic View, Chromosome View, Gene View, and Tab View

  You use commands in the interactive analysis tabs to perform preprocessing, analysis, and reporting of data. You can view the results of data analysis in the Genomic Viewer.
- View tab commands – change appearance of Genomic Viewer display

Figure 3 shows the main window of Agilent Genomic Workbench when the Genomic Viewer tab is selected, and identifies the names of its components.

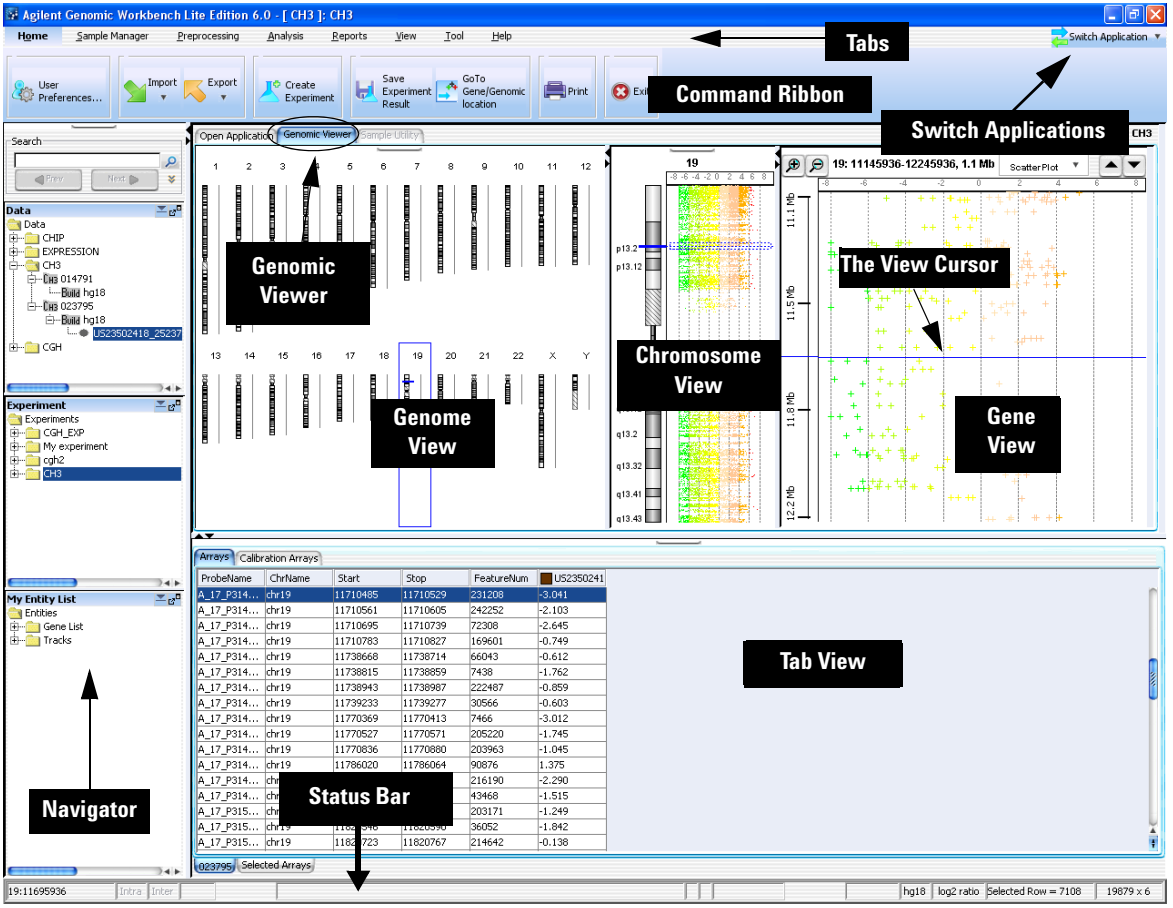


Figure 3 Agilent Genomic Workbench Lite Edition showing main components for CH3

What can you do with the main components for display of data and results?

See the table below for the parts of the main window you use to display log ratio data and results.

Table 1 Components of Agilent Genomic Workbench main window for display of data and results

To do this	Use this part of the main window
Change program to CGH, ChIP, Methylation (CH3)	**Switch Application button**: Click the button and click the program you want to open. The scatter plot options are different for the different program types
Import or export data	**Home tab**: Click the **Import** or **Export** button to select the data you want to import or export. See Chapter 2, "Importing, Managing, and Exporting CH3 Data and Other Content" for more information.
Select array data to display in the three graphical views or in the Tab View as a table	**Experiment pane of the Navigator**: Create an experiment with the imported data, select the experiment, and then select the data within the experiment to display or analyze. See Chapter 3, "Displaying CH3 Data and Other Content" for more information.
Display array data/results for only a certain portion of a chromosome	**Genome View**: Select a chromosome to display in Chromosome View. You cannot view log ratio data points here. **Chromosome View**: Select a gene region to display in Gene View. You can display log ratio data points here if you select **Scatter Plot** in the View Preferences dialog box. **Gene View**: See the log ratio data next to a selected region of a chromosome, with associated genes and track-based annotation. See Chapter 4, "Setting Up Methylation (CH3) Analysis" for details about these Views.

Table 1 Components of Agilent Genomic Workbench main window for display of data and results (continued)

To do this	Use this part of the main window
Show/Hide or customize the data points for the scatter plots	**Gene View**: Move the mouse pointer over **Scatter Plot** to display the options. Or, right-click and then click **View Preferences**. **Chromosome View**: Right-click and then click **View Preferences**. **View tab**: Click **View Preferences**. See Chapter 3, "Displaying CH3 Data and Other Content" for information on how to do this.
Display array data next to tracks or gene lists	**My Entity List pane of Navigator**: Add or select a track or gene list to have it appear in Gene View. See Chapter 3, "Displaying CH3 Data and Other Content" for information on how to do this.
Change the appearance of the display	**View Tab**: Click **View Preferences**. From the View Preferences dialog box, you can change the orientation, select what type of data to view, and configure scatter plot options. **Genomic Viewer**: Right-click any View except the Tab View and select **View Preferences**. In the View Preferences dialog box, you can select to show or hide the scatter plots and how to display them, including results. See Chapter 3, "Displaying CH3 Data and Other Content" for more information.
Analyze or reanalyze displayed data	**Preprocessing Tab**: Click this tab to display commands you use to manipulate the data before you apply the algorithms. **Analysis Tab:** Click this tab to display commands you use to analyze the data. **Reports Tab:** Click this tab top display commands you use to generate and manage reports. For more information on what you can do in these tabs, see "Tabs" on page 24.

Switching Applications

You can use the Agilent Genomic Workbench to work with a variety of different data types. Because the requirements for the display of data (and calculation of results, if using a license) are different for different data types, you must switch the application for the type of data you want to display.

The Switch Applications menu, located at the upper right corner of the Agilent Genomic Workbench window, is used to change the application. The selected application is marked . The selected application is also displayed in the title bar of the Agilent Genomic Workbench main window.

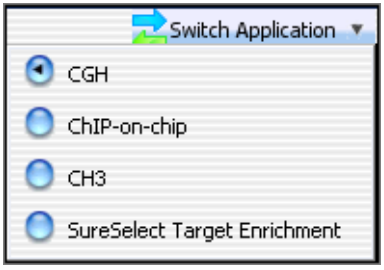


Figure 4 Switch Application menu

Using Tabs and Command Ribbons

Tabs

When you click a *tab*, groups of commands or single commands appear that are specific for that tab. The tabs that are displayed change depending on what licenses you have, and what application is selected (such as CGH, ChIP, CH3).

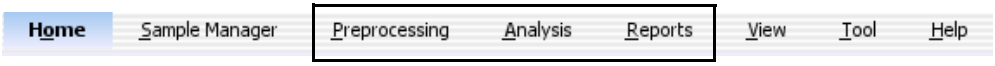


Figure 5 Tabs for CH3 interactive analysis

The following table summarizes the capabilities in the interactive tabs of the CH3 application of Agilent Genomic Workbench:

Tabs	CH3 Capabilities
Preprocessing	Combine (fuse) array designs
Analysis	Calculate a moving average on log ratio data Calculate a moving average on ZScores generated by probe methylation algorithm Apply probe methylation algorithm Apply BATMAN algorithm
Reports	Generate probe report Generate BATMAN report

Commands

The area where commands appear is called a *command ribbon*. The command ribbon that appears when you click the Home tab for CH3 is shown below. The commands that appear in the command ribbon change depending on what application module is selected, and which tab in that application module is selected.

Figure 6 Home command ribbon for CH3 interactive analysis

For a complete description of all of the command ribbons and commands you see in Agilent Genomic Workbench, see "Command Ribbons" on page 124.

Using the Navigator to Search for Data

This section gives you instructions on how to search for design files, extracted FE data, experiments and other information in the Navigator of Agilent Genomic Workbench. The Navigator contains different panes when you select the Sample Manager tab. See the *Sample Manager User Guide* for information on the Navigator contents.

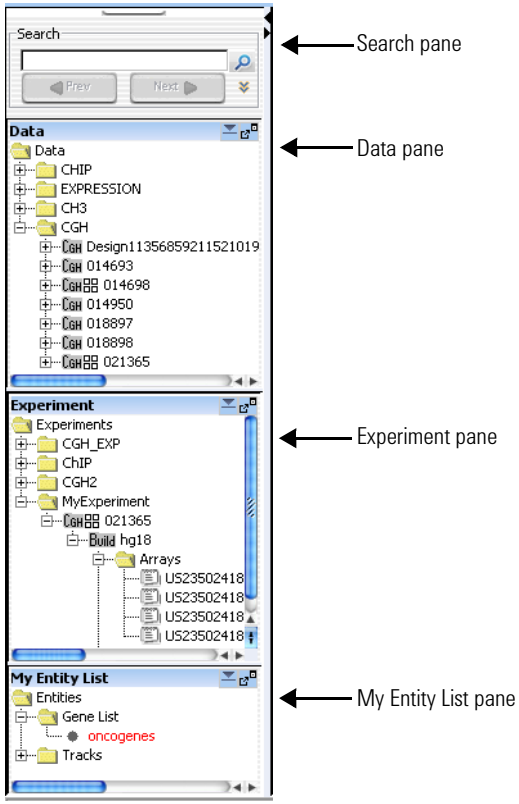


Figure 7 Navigator panes

The Navigator shows the array data, experiments, and other content stored in Agilent Genomic Workbench that is available to the user. It contains the following panes:

Pane	Comments
Search	Lets you search within any pane of the Navigator for a specific item (array or build, for example). You must type the entire array name or term; otherwise, use asterisks (*) as wild cards for unspecified strings. For example, type *1234* to find any item that contains "1234".
Data	Contains microarray data files, organized by application, then by design and genome build. Shows all probe groups and microarray designs that are available to you, organized by folders. In general, you can: • Expand or collapse folders to show or hide content. • Right-click the name of a folder or item to open a shortcut menu that lets you take action on the item. See "Data pane – icons, special text, and buttons" on page 139 and "Data pane – actions and shortcut menus" on page 139.
Experiment	Contains Agilent Genomic Workbench experiments. Experiments are organizational units that contain links to microarray data and design files. In data analysis modules, experiments also contain saved results.
My Entity List	Contains gene lists and tracks: • **Gene Lists** are collections of genes of interest. You can create them within the program, import and export them, and apply them to Gene View and Chromosome View. • **Tracks** are collections of annotation or other information that map to specific genomic locations. You can import, export, and combine tracks, and display them in Gene View with your array data and analysis results.

To search the Navigator

You can search one or all of the panes of the Navigator for items that match a specific search term. Figure 8 shows the search pane of the Navigator, and identifies a couple of its elements.

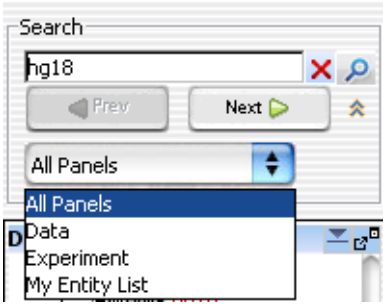


Figure 8 Search pane of the Navigator

1. At the top of the Navigator, in the Pane list, select the pane to search. To search in all panes, select **All Panels.** If the pane list is not visible, click to show it.
2. In the search term box, type the desired search term. The search term is not case sensitive, but it must contain the complete entry that you want to find. You can use asterisks (*) to represent one or more unspecified characters. For example, type *12345* to find any item that contains "12345".
3. Click .

   The program searches the selected pane(s) for items that match your search term. If it finds matching items, the program expands the appropriate folders, and displays the names of the matching items in red. The first matching item is highlighted in yellow.
4. Do any of the following:
   - To highlight the next matching item, if one is available, click Next.
   - To highlight the previous matching item, click Prev.
5. After you complete the search, click to clear the results of the search, as well as your search term.

Using the Genomic Viewer to Display Data

What is the Genomic Viewer?

Genomic Viewer is the graphics and tabular display section of the Agilent Genomic Workbench main window. In the Genomic Viewer, extracted data and analysis results are tabulated and displayed next to depictions of the genome, selected chromosome, and selected genes of the species whose array data you are analyzing.

There are four main views in the Genomic Viewer, as shown in Figure 9.

- **Genome View** – A graphical representation of the entire genome for the selected species. Use this view to select the chromosome to show in the other views.
- **Chromosome View** – A graphical representation of the selected chromosome, displayed with cytobands and a plot area. Click or drag the mouse to select a region to display in the Gene View.
- **Gene View** – A more detailed view of the chromosomal region selected in the Chromosome View.
- **Tab View** – Displays design annotation and log ratio data related to the chromosome you select in Chromosome View

For more information on the Genomic Viewer and its views, see Chapter 5, “Methylation (CH3) Analysis Reference”.

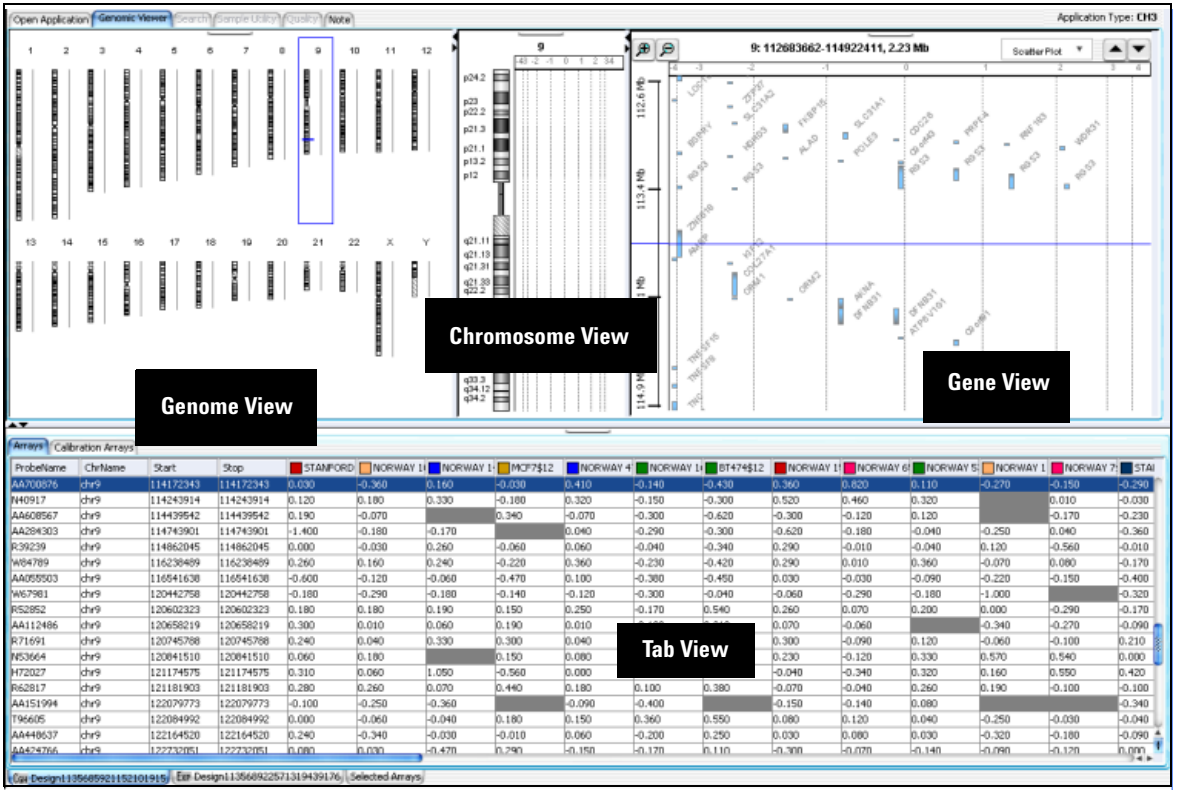


Figure 9 Genomic Viewer in vertical orientation

To change the size of and detach panes from the Agilent Genomic Workbench main window

- To change the size of a pane in the main window, drag one of its inside borders.
- To detach a pane from the main window and open it in a separate window, click its **Detach** button .

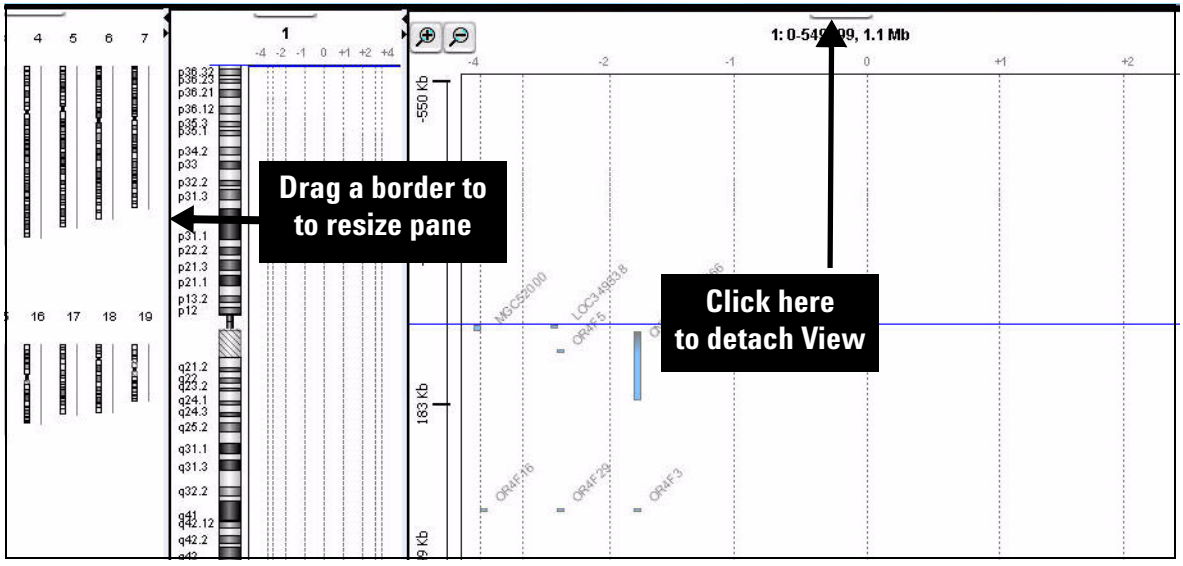


Figure 10 Changing the size of and detaching panes

To maximize and reattach panes to the Agilent Genomic Workbench main window

- To display a view full-screen in a separate window, click its **Maximize** button.
- To reattach a view in a separate window to the main window, click its **Close** button.

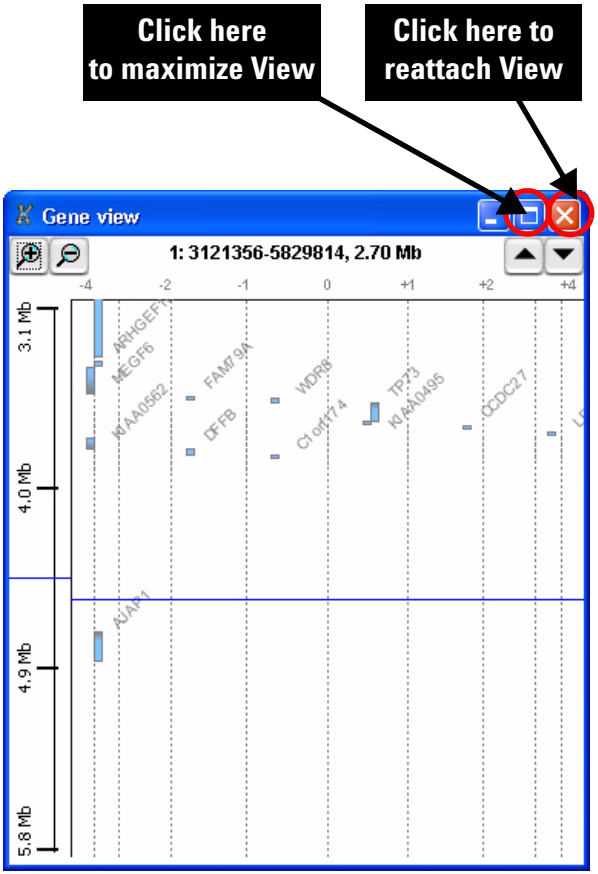


Figure 11 Maximizing and reattaching panes

General Instructions for Displaying Microarray Data/Results

An *experiment* is the folder that holds data from any array set you select for the experiment. The folder also holds analysis results.

You set up experiments to display all data and results in the Genomic Viewer. To set up an experiment you:

- Import data
- Create a new experiment
- Add the imported data to the experiment
- Select the experiment to display data

For step-by-step instructions on how to display data, see the *Data Viewing User Guide*.

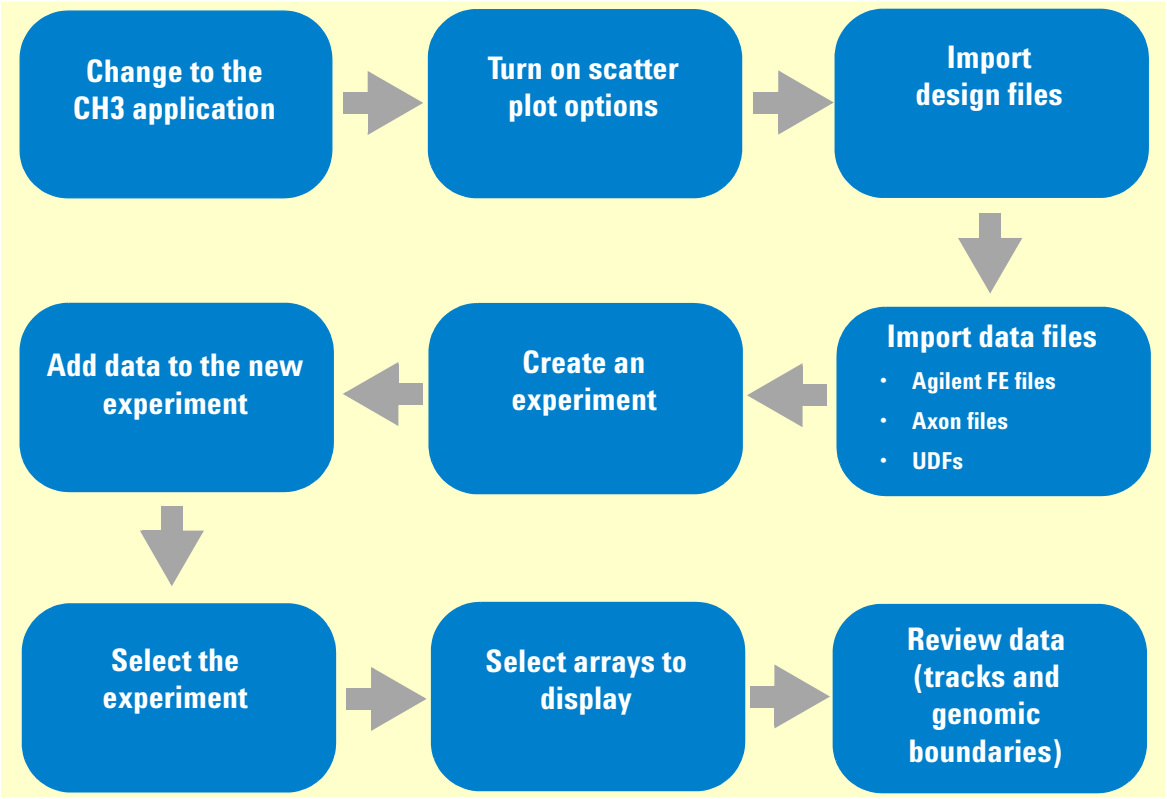


Figure 12 Typical pathway for displaying microarray data/results

General Instructions for Setting Options for CH3 Interactive Analysis

Figure 3 shows the pathway for setting up an experiment and analyzing data interactively with the CH3 application. After you import data and set up experiments, you can set up preprocessing and analysis calculations before you make the experiment active, or apply them afterwards. You can combine designs, apply an algorithm to show probabilities that genomic regions are methylated or not and report the results. When you change each option after experiment activation, the program recalculates the results. For more information on how to change analysis options, see Chapter 4, “Setting Up Methylation (CH3) Analysis”.

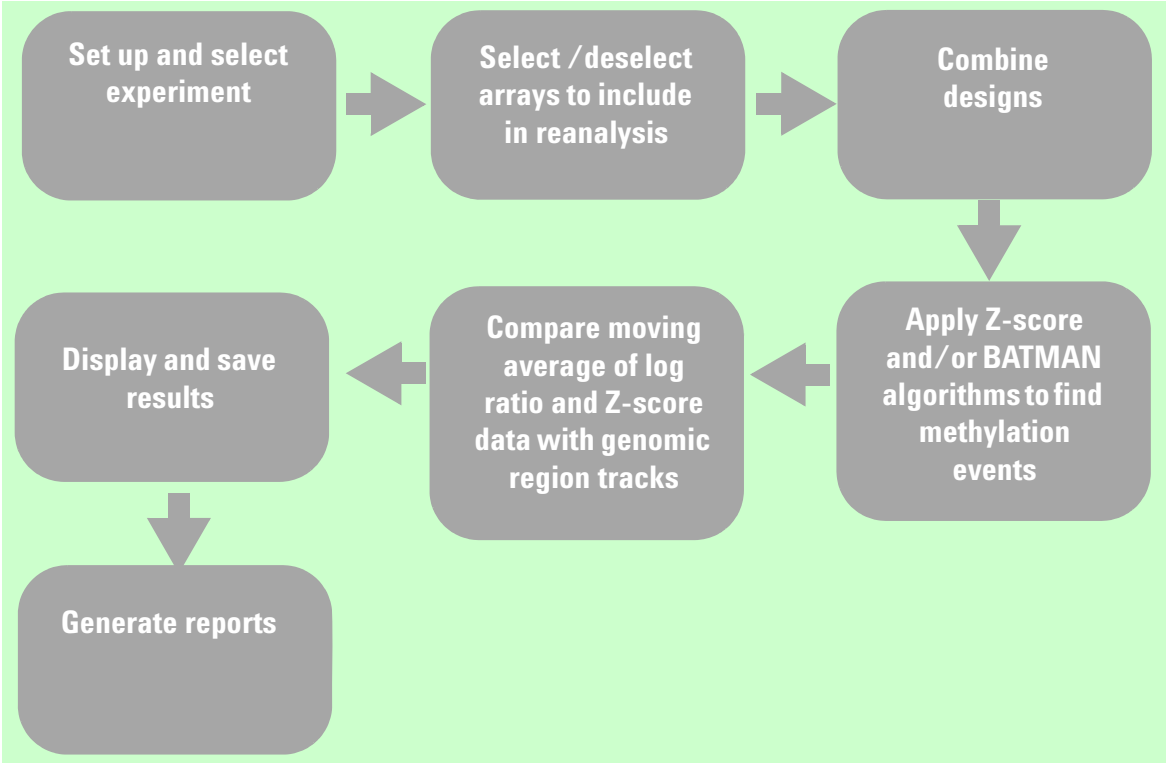


Figure 13 Typical analysis pathway – Interactive mode for CH3 application

Getting Help

To get help within Agilent Genomic Workbench

Agilent Genomic Workbench has several help resources. All help guides open using Adobe® Acrobat®.

Help Resource	Description/Instructions
Methylation (CH3) Analysis User Guide	This user guide, which you are now reading, supplies comprehensive help on all available CH3 tasks. You can access it easily from anywhere within the program. **1** In any tab of Agilent Genomic Workbench, click the **Help** tab. **2** On the Help Ribbon, click **Application Guide.** Methylation (CH3) Analysis User Guide opens.
Other User Guides	The Help tab in Agilent Genomic Workbench lets you view any of the available user guides that apply to the currently selected application type. **1** Set the desired application type. See "Switching Applications" on page 23. **2** In the Agilent Genomic Workbench tab bar, click **Help.** The names of the available user guides appear in the command ribbon. **3** Click the desired user guide. The selected user guide opens.
Product Overview Guide	An additional guide gives an overview of the capabilities within Agilent Genomic Workbench and describes how to start and find help for all of the programs. In addition, it helps you with system administration and troubleshooting. **1** In any interactive analysis tab of Agilent Genomic Workbench, click the **Open Application** tab. **2** At the upper right corner of the Open Application tab, click **Product Overview**.

To contact Agilent Technical Support

Technical support is available by phone and/or e-mail. A variety of useful information is also available on the Agilent Technical Support Web site.

Resource	To find technical support contact information
Agilent Technical Support Web site	**1** Go to http://chem.agilent.com. **2** Select a country or area. **3** Under Quick Links, select **Technical Support**. **4** Select from the available links to display support information.
Contact Agilent Technical Support by telephone or e-mail (United States and Canada)	Telephone: (800-227-9770) E-mail: informatics_support@agilent.com
Contact Agilent Technical Support by telephone or e-mail (for your country)	**1** Go to http://chem.agilent.com. **2** Select **Contact Us**. **3** Under Worldwide Sales and Support Phone Assisstance, click to select a country, and then click **Go**. Complete e-mail and telephone contact information for your country is displayed.

To learn about Agilent products and services

To view information about the Life Sciences and Chemical Analysis products and services that are available from Agilent, go to www.chem.agilent.com.

2 Importing, Managing, and Exporting CH3 Data and Other Content

This chapter describes how to import, organize, manage, and export methylation (CH3) data and other content within the user interface of Agilent Genomic Workbench Lite Edition.

Importing Files

You can use the Home tab to import many kinds of files into Agilent Genomic Workbench. The table below describes the kinds of files you can import, and the topics in this section that describe how to import them.

The Data pane of the Navigator displays all of the content available for the user who is logged in. See "Navigator" on page 135 for more information on the Navigator panes and how to use them.

Type of file	Description	See these topics
Microarray data files	• Agilent Feature Extraction (*.txt) data files • GenePix/Axon (*.gpr) data files • Universal Data Files (UDFs) (*.txt files)	"To import Agilent FE or Axon data files" on page 41 "To import a UDF file" on page 42
Microarray design files	• Agilent GEML (*.xml) design files • GenePix/Axon (*.gal) design files	"To import Agilent GEML design files" on page 39 "To import Axon design files" on page 40
Genome builds	Agilent-supplied genome information for human, mouse and rat genomes	"To import a genome build" on page 45
Tracks	BED format annotation track files	"To import tracks" on page 46
Array attributes	.txt files that you have created yourself or previously exported from Agilent Genomic Workbench	"To import array attributes" on page 48
Experiments	ZIP format file of experiments exported from Agilent Genomic Workbench	"To import an experiment file" on page 48

To select a different location for data files

By default, the program stores microarray and experimental data files in **C:\Program Files\Agilent\Agilent Genomic Workbench Lite Edition<version>\data**. If you want, you can select a different location.

CAUTION Do not select a location that contains a backup data folder; the current data overwrites the data in the folder you select.

1 In the Home tab, click **User Preferences.**

The User Preferences dialog box appears. See "User Preferences" on page 230.

2 In the **Miscellaneous** tab, under **Data Location,** click **Browse.**

An Open dialog box appears.

3 Select a location, then click **Open.**

The selected location appears in the User Preferences dialog box, in Data Location.

4 Click **OK.**

To import Agilent GEML design files

The Agilent Genomic Workbench database must contain designs that match the Agilent Feature Extraction data files you want to import. The design file must be present before any extraction data files are imported. Your imported GEML files contain array-specific information such as probe names, annotations, and chromosomal locations, and are associated with a specific genome build. To import an Agilent GEML file, use the following procedure:

1 In the Home tab, click **Import > Design Files > GEML File.**

The Import Design Files dialog box appears. See "Import" on page 200. The dialog box shows only *.xml files.

2 To select a file for import, click its name. To select additional files, hold down the **ctrl** key while you click their names.

3 Click **Open.**

The program validates the selected file(s), and the Import GEML Design Files dialog box appears. See "Import GEML design files" on page 203.

If a design file passes validation, the Status column will show **Update** in green. If a design file fails validation, **Corrupt** appears in the Status column beside it, and the program will not import the file. To remove the corrupt design from the list, click its **Remove** button  .

4 Click **Start Import.**

The program imports the file(s). The files appear as new design folders in the Data folder of the Navigator with the genome build as a node within the folder.

You can import two design files with the same name, but associated with different genome builds; for example, Hg17 or Hg18. If you do, the program creates a single design folder with two nodes, one for each genome build.

To import Axon design files

You can import Axon (*.gal) microarray design files into Agilent Genomic Workbench. The program requires the Axon design files that match all Axon array data files you import.

1 In the Home tab, click **Import > Design Files > Axon File.**

The Import Axon Design Files dialog box appears. See "Import" on page 200. The dialog box shows only *.gal files.

2 To select a file for import, click its name. To select additional files, hold down the **ctrl** key while you click their names.

3 Click **Import.**

The program validates the selected file(s), and the Set genome build and species for Axon design files dialog box appears. See "Set genome build and species for Axon design files" on page 219.

If a design file passes validation, the Status column will show **Update** in green. If a design file fails validation, **Corrupt** appears in the Status column beside it, and the program will not import the file. To remove the corrupt design from the list, click its **Remove** button  .

4 For each design file, select the appropriate **Species** and **Genome Build.**

5 Click **Start Import.**

The program imports the file(s). The files appear as new design folders in the Data folder of the Data pane, organized by application (CGH, ChIP, or methylation, for example).

To import Agilent FE or Axon data files

You can import several types of microarray data files into Agilent Genomic Workbench:

- Agilent Feature Extraction (FE) *.txt data files
- Axon (*.gpr) data files
- Universal Data Files (UDFs) (*.txt files) See "To import a UDF file" on page 42 for instructions on how to import this file type.

In order to import Agilent Feature Extraction files, you must import the representative GEML array design files first. In order to import Axon data files, you must import the representative Axon.gal design files first. See "To import Agilent GEML design files" on page 39 or "To import Axon design files" on page 40.

1 In the Home tab, do one of the following:
   - To import Agilent FE data files, click **Import > Array Files > FE File.**
   - To import Axon data files, click **Import > Array Files > Axon File.**

   A dialog box appears. Only data files of the appropriate type appear. See "Import" on page 200.
2 To select a file for import, click its name. To select additional files, hold down the ctrl key while you click their names.
3 Do one of the following:
   - For Agilent FE files, click **Open.**
   - For Axon files, click **Import.**

   In either case, the Agilent Feature Extraction Importer dialog box appears. "Agilent Feature Extraction Importer" on page 166.
4 Set the following, as needed:

Setting	Comments
Name	The names of imported arrays are often cryptic. You can give any array a more meaningful label. **a** Double-click the name of the array. **b** Edit the name, as desired. **c** Press **Enter.**
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	If you mark this option, the program deletes an existing array data file if it has the same name as one you import.

5 Do one of the following:

- To import the file(s) while you wait, click **OK.**
- To import the file(s) in the background, click **Run in Background.** This lets you continue while the program imports the files.

To import a UDF file

UDF files are plain text files that contain array data in tab-delimited format. Files must contain the following six columns of information, in any order. Each column must contain the following column names, as column headers, or you must "map" the names from the file to these columns in Agilent Genomic Workbench:

- Probe name
- Chromosome name
- Start position

- Stop position
- Description
- Signal intensity data (The file can contain additional columns, each with data from an additional array.)

When you import a UDF file, the program creates a new design based on the information you enter during import, and the information in the file itself. This design contains all of the arrays represented in the file. The program also creates a new experiment that contains the arrays.

1 In the Home tab, click **Import > Array Files > UDF File.**

The UDF Files dialog box appears. See "Import" on page 200. Only *.txt files appear in the dialog box.

2 Select the desired UDF file, then click **Open.**

The Select data type for experiments dialog box appears. "Select data type for experiments" on page 218.

3 For each array, set the following, as needed:

Setting	Comments
Experiment Name	By default, the program creates an experiment with the same name as the imported file. To change the name: **a** Double-click the name. **b** Edit the name as desired. **c** Press **Enter.**
Data type	• Select the mathematical form of the signal intensity data for the array. The options are **ratio, $\log_2$ ratio, $\log_{10}$ ratio,** and **ln ratio.**
Design type	• Select **CH3.**

4 Click **Continue.**

The Universal Data Importer – Map column headers dialog box appears. The main table in the dialog box contains the first few rows of data from the file. Column headings that are contained in the first line of the file appear at the top of the table as a guide. See "Universal Data Importer - Map Column Headers" on page 225 for more information.

5 Below each column heading, select the label that identifies the content of the column. Use each label exactly once, except for LogRatio, which

you can use multiple times. Alternatively, in **Select Mapping,** select a saved column map.

These options are available:

Label	Description
ProbeName	The column contains names of probes.
ChrName	The column contains names of chromosomes.
Start	First chromosomal location for the probe.
Stop	Last chromosomal location for the probe.
Description	Text annotation related to the probe.
LogRatio	The column contains array data values that correspond to each probe. You can use this label more than once.

6 Under Species Info, select the **species** and **Genome Build** appropriate to the data in the file.

7 If you expect to import many similar UDFs in the future, follow these steps to save the column map:

   a Under Mapping Info, click **Save Mapping As.**
   An Input dialog box appears.

   b Type a name for the column map, then click **OK.**
   The name of the saved map appears in Select Mapping.

   In the future, you can select this mapping and apply it to any UDF file that you import.

8 By default, the program creates a "Virtual Array ID" that becomes the Array ID attribute for the array(s) in the UDF. To create your own virtual Array ID, follow these steps:

   a Under ArrayID Info**,** clear **Use System Generated ArrayID.**

   b Double-click the number in Virtual ArrayID, then type the desired new Virtual Array ID.

   For more information on Array IDs, see the *Sample Manager User Guide*.

9 Click **Import**.

The program validates your column mapping. A dialog box appears. If you need to fix the column map, the dialog box has a list of the missing column label(s). If the column map is complete, a message asks if you want to import additional files with the same mapping.

10 Do one of the following:

- If you want to import additional files with the same column mapping, follow these steps to include these files in the import:
  - **a** Click **Yes.**
    The UDF Files dialog box appears.
  - **b** Click the name of a file to select it for import. Hold down the **ctrl** key while you click the names of additional files.
  - **c** Click **Open.**
- If you do not want to include additional file(s) in the import, click **No.**

The Program imports all requested files, and the UDF Import Summary dialog box appears. This dialog box shows the imported files, the number of lines of data that were imported for each file, and the number of lines that were skipped, if any. If a file name appears in red, the program may not have imported the file. See "UDF Import Summary" on page 224.

11 Click **OK.**

In the Data pane, in the appropriate application type folder, a new design folder appears. The design folder contains the imported array data.

A new experiment appears in the Experiments folder in the Experiment pane, that contains the array data. This experiment has the name of the imported UDF file, unless you changed it during import.

To import a genome build

In general, the program uses the genome build specified in the array design file, and protects it from changes. If a genome build is not available in the program, you can import one.

NOTE Use arrays from a single genome build in an experiment.

1 In the Home tab, click **Import > Genome Build.**

The Import Genome Build dialog box appears. See "Import Genome Build" on page 204.

2 Set the following. All are required.

Setting	Instructions
Species	• Type the genome's species of origin, as you would want it to appear within the program.
Build Name	• Type the name of the genome build you want to import, as you would want it to appear within the program.
Refseq File	This file contains information on gene locations for Gene View. **a** Click **Browse.** A dialog box appears. **b** Select the desired file, then click **Open.**
Cyto-band File	This file contains the graphic information on the cyto-bands for Genome and Chromosome Views. **a** Click **Browse.** A dialog box appears. **b** Select the desired file, then click **Open.**

3 Click **OK.**

To import tracks

You can import BED format track files into Agilent Genomic Workbench. Track files contain specific features correlated with chromosomal locations, and apply to a specific genome build of a given species.

1 In the Home tab, click **Import > Track.**

The Import Track dialog box appears. See "Import Track" on page 205.

2 Set the following. All are required.

Setting	Instructions
Species	• Select the species to which the track applies.
Build Name	• Select the specific genome build of the species to which the track applies.
Track Name	• Type a name for the track. This name identifies the track within the program, including the name that appears if you include the track in Gene View.
Track File	**a** Click **Browse.** A dialog box appears. **b** Select the name of the track (*.bed) file that you want to import. **c** Click **Open.** The location of the file appears in Track File.

3 Click **OK.**

The program imports the track. To display the track in Gene View, and to manage tracks, see "To show tracks in Gene View" on page 92.

To import array attributes

An array attributes file is a tab-delimited *.txt file that contains a list of arrays by Array ID, and values for array attributes. Attributes are pieces of array-specific information, such as the hybridization temperature or the name of an array set that contains the array.

Although you can import array attributes with this function, the Sample Manager application lets you import and assign array attributes more easily. See the *Sample Manager User Guide* for more information. To import an array attributes file

1 From the Home tab, click **Import** and then select **ArrayAttributes**.

The Import microarray attributes dialog box appears. See "Import" on page 200.

2 Select the microarray attributes file, then click **Import.**

The program imports the file. If the ArrayIDs in the file do not match the ArrayIDs of arrays in the program, a dialog box appears. The dialog box has a list of the ArrayIDs in the file that do not match. Click **No** to stop the import process, or click **Yes** to continue anyway.

To import an experiment file

In Agilent Genomic Workbench, an experiment is a set of links to microarray data and design files, and any associated results. An Agilent Genomic Workbench experiment file is a single ZIP file that contains the design and data files for one or more experiments. You can import

- Experiment files created in Agilent Genomic Workbench on another computer
- Agilent Genomic Workbench 5.0 and 6.0 experiment files

1 In the Home tab, click **Import > Experiments.**

The Import Experiments dialog box appears. See "Import" on page 200.

2 Select the ZIP file that contains the experiment(s) you want to import, then click **OK.**

The program imports the experiment file. Designs appear as new folders in the Data pane, in the applicable design type folder. Array data appears within the applicable design folder, organized by genome build. In addition, the experiment(s) appear in the Experiment pane, containing the appropriate arrays.

Agilent Genomic Workbench experiment files contain all of the design and array data files for an experiment, but do not include any analysis parameter settings, array selections, or analysis results. To export the data and design files from one or more experiments, see "To export experiments" on page 72.

Working with Experiments to Organize Imported Data

This section describes how to organize imported array data and designs into *experiments*. Experiments, shown in the Experiment pane of the Navigator, contain links to specific array data and design files in the Data pane. After you set up an experiment, you can then analyze selected array data within the experiment.

Because experiments only contain *links* to the actual data and design files, any number of experiments can use a given set of files. In the CH3 application, experiments also can contain saved experiment results.

To display the array designs and data in the program

- To display the directory of data in the program, use the Data pane (Figure 14). Double-click a folder to expand or collapse it, or click the ⊞ and ⊟ buttons.

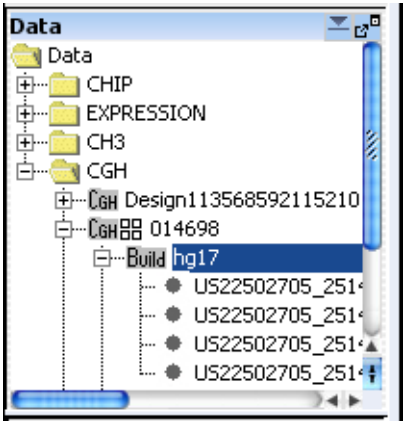


Figure 14 Data pane of the Navigator

In the Data pane, the program organizes design files by the application (CGH, ChIP, or methylation, for example) to which they apply. It organizes array data files by genome build under the design with which they are associated.

You can right-click many elements of the Data pane to open shortcut menus. For more information, see "Data pane – actions and shortcut menus" on page 139.

Many icons can appear in the Data pane. See "Data pane – icons, special text, and buttons" on page 139 for a complete list.

The Search pane can help you find specific data files or other content. See "To find specific content items in the Navigator" on page 62.

To create a new experiment

In Agilent Genomic Workbench, *experiments* are organizational units that contain links to data and design files. To display or analyze data, you must first create an experiment and associate the desired data files with it. Because experiments only contain *links* to the actual data and design files, any number of experiments can use a given set of files. In data analysis applications (CGH, ChIP, or methylation, for example), experiments can also contain saved experiment results.

1 In the Home tab, click **Create Experiment.**

  The Create Experiment dialog box appears. See "Create Experiment" on page 177.

2 Type a **Name** and an optional **Description** for the experiment.

3 Do one of the following:

  - To create an empty experiment, and add data to it later, click **OK.** The program creates the experiment. To add arrays to the experiment later, see "To add arrays to an experiment" on page 53.

- To create an experiment and add data to it now, follow these steps: (You can add or remove data from the experiment later.)

  a Click **Properties.**
  The Experiment Properties dialog box appears. See "Experiment Properties" on page 187.

  b Under **Select Design,** select the design and genome build associated with the desired array data.
  The applicable arrays appear in Array List.

  c In **Array List,** click the name of an array that you want in your experiment. Hold down the **ctrl** key while you click the names of additional arrays.

  d Click  .
  The program transfers the selected arrays to the Selected Array List.
  The dialog box also shown other options for adding arrays. See "Experiment Properties" on page 187 for more information.

  e Click **OK.**
  The program creates the new experiment, and adds data to it from the selected arrays.

- To create an experiment and add data to it using the "drag and drop" method, follow these steps:

  a To create an empty experiment, click **OK.**
  The program creates the experiment.

  b From the Data pane, expand a design to see the build and array data.

  c Drag an array from the Data pane and drop it onto the experiment folder in the Experiment pane.

In all cases, a folder with the name of the new experiment appears in the Experiment pane of the Navigator.

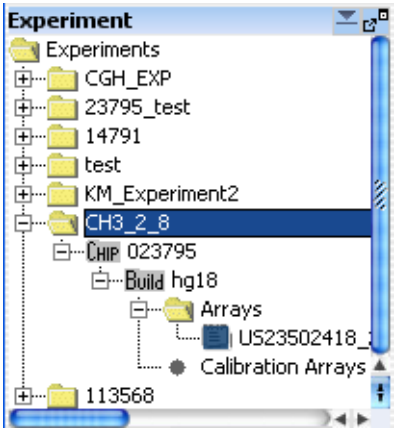


Figure 15 Experiment pane of the Navigator

To add arrays to an experiment

After you create an experiment, or import one, you can add arrays to it. When you add arrays to an experiment, you create links between the experiment and the array data and design files. Because the program does not move the actual files, multiple experiments can share the same arrays.

1 In the **Experiment** pane, double-click the **Experiments** folder to expand it.

2 Right-click the name of the experiment, then click **Show Properties.**

The Experiment Properties dialog box appears. See "Experiment Properties" on page 187.

3 Under **Select Design,** select the design file and genome build for the arrays to add.

The arrays for the selected design file and genome build appear in Array List.

4 In **Array List,** select the arrays to add to the experiment. To select a single array, click its name. To select additional arrays, hold down the **ctrl** key while you click their names.

5 Click > .

The program transfers the selected arrays to the Selected Array List.

The dialog box also has other options for adding arrays. See "Experiment Properties" on page 187 for more information.

6 Click **OK.**

Or, to add array data to an experiment using the "drag and drop" method,

1 From the Data pane, expand a design to see the build and array data.

2 Drag an array from the Data pane and drop it onto the experiment folder in the Experiment pane.

If needed, the program adds appropriate design and genome build folders to your experiment folder in the Experiment pane. It places the arrays you selected in the appropriate genome build folder.

NOTE If you add arrays to a selected experiment, and analysis options are marked (for example, Probe Methylation or Batman), you need to clear and then mark these options again to apply the analysis to the arrays.

To change the order of arrays in an experiment

When you select an experiment, a table appears in the Tab View of Genomic Viewer that contains log ratio values for arrays in the experiment. See "Tab View" on page 160. You can change the order in which the arrays appear the table. If you display separate scatter plots in Gene View and Chromosome View for each array, the array order also determines the order in which these plots appear. You can use this feature to organize your arrays more logically, or to make it more convenient to display certain arrays. It is especially useful if you have many arrays.

1 In the Experiment pane, right-click the name of the desired experiment, then click **Edit Array Order.**

The Edit Array Order dialog box appears. See "Edit Array Order" on page 186.

2 In **Design,** select the design that contains the arrays whose order you want to change.

The arrays from the selected design appear in Array Name.

3 Do any of the following:

- To move an array up in the list, click its name, then click  .
- To move an array down in the list, click its name, then click  .
- To sort the list based on a specific microarray attribute, select the desired attribute in **Order by.**

4 Click **OK.**

To change the display names for arrays in an experiment

You can change the name displayed for arrays in an experiment, based on array attributes. When you change the display names for arrays in an experiment, the array names are changed only for the selected experiment. The display names are unchanged in the Data pane and in the other experiments.

1 Expand the folders in the Experiment pane until you see the experiment you want to change.

2 Right-click the experiment name, and select **Show Properties**.

3 In the Experiment Properties dialog box, click **Display Name by** and select an attribute to use for display of array names.

Click **OK**. The names of the arrays in the experiment are changed to the selected attribute. If the attribute does not exist for an array, the Global Display Name is displayed.

NOTE

To change the name of an array that is displayed throughout Agilent Genomic Workbench, change its Global Display Name using Sample Manager. For more information, see the *Sample Manager User Guide*.

To rename an array in an experiment

When you rename an array in an experiment, you change the array's name only within the context of the selected experiment. The name of the array is unchanged in the Data pane, and in other experiments.

1. Expand the folders in the **Experiment** pane until you can see the array to rename.
2. Right-click the name of the desired array, then click **Rename.**

   An Input dialog box appears.
3. Type the new name for the array, then click **OK.**

   The name of the array in the tab view of the selected experiment is renamed. The global display name of the array is not changed.

To remove arrays from an experiment

When you remove arrays from an experiment, you only remove the links between the experiment and the data files. The files are still available in the program for use in other experiments. To completely remove files from the program, see "To remove data or design files from the program" on page 64.

1. In the **Experiment** pane, expand folders until you can see the desired experiment, and the array(s) that you want to remove from it.
2. In the **Arrays** or **Calibration Arrays** folder of the desired experiment, click the name of an array to select it for removal. Hold down the **ctrl** key while you click the names of additional arrays.
3. Right-click one of the selected array names, then click **Delete.**

   A Confirm dialog box appears.
4. Click **Yes.**

   The program removes the links between the experiment and the selected array data files. If the removal of arrays leaves a design folder in the experiment empty, the program removes this folder as well.

To select or remove calibration array(s)

After you add an array to an experiment, you can select it as a calibration array. The program lists calibration arrays within the Calibration Arrays folder of the experiment with a special icon  . You can also remove the calibration designation from an array.

To select an array as a calibration array

1 Expand the folders of the **Experiment** pane until you can see the array that you want to select as a calibration array.

2 Right-click the name of the desired array, then click **Select for Calibration.** To select all of the arrays of a given design in the experiment as calibration arrays, right-click the genome build folder of the desired design, then click **Set for Calibration**.

The program selects the array as a calibration array. In the Calibration Arrays folder of the applicable genome build and design within the experiment, the array appears with a special icon  .

To deselect an array from calibration

1 Expand the folders of the **Experiment** pane until you can see the desired array. The program lists calibration arrays in the Calibration Arrays folder(s) of the applicable genome build(s) and design(s) within each experiment.

2 Right-click the name of the desired array, then click **Deselect from Calibration.**

To deselect multiple calibration arrays at once, select all of the desired arrays. Right-click one of the arrays, then click **Deselect from Calibration.** (To select multiple arrays, click the name of one array, then hold down the **ctrl** key and click the names of additional arrays. 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.)

The program removes the array(s) from calibration, and moves the arrays to the Arrays folder of the applicable genome build and design within the experiment. The icons of the arrays change to the standard (non-calibration) array icon .

To show or hide array attributes in an experiment

Sample attributes are pieces of information specific to an array, such as chip barcode or hybridization temperature. You can show or hide attributes for the arrays in the experiment with the Sample Attributes dialog box. See "Sample Attributes" on page 212.

1. Right-click the experiment whose attributes you want to show or hide, or to change.

   You see the array attributes and their values that were set up in the Sample Manager table. See the *Sample Manager User Guide*.

2. Click **Sample Attributes**.
3. Click **Show/Hide Attributes**.

   The Show/Hide Columns dialog box appears. See "Show/Hide Columns" on page 221.

NOTE You cannot hide the required attributes. These include Array ID, Global Display Name, Green Sample, Red Sample (for 2-color arrays), and Polarity.

4. Mark the check boxes for the attributes you want to show, or clear the check boxes for the attributes you want to hide. These changes are applied globally for the arrays.
5. Click **Save**.
6. In the Show/Hide Columns dialog box, click **Close**.
7. Click **Close**.

NOTE You cannot create new attributes using this dialog box. To do this, you must use the Sample Manager tab. See the *Sample Manager User Guide*.

To display or edit array attributes in an experiment

1. Right-click the experiment whose attributes you want to display or edit.
2. Click **Sample Attributes**.

You see the array attributes and their values that were set up in the Sample Manager table. See the *Sample Manager Guide*. See "Sample Attributes" on page 212.

3 Double-click the cell whose array attribute value you want to change.

NOTE You cannot change Array ID, Polarity, Extraction Status, or IsMultiPack attributes for extracted or UDF arrays.

4 Click **Save Changes**.

5 Click **Close**.

To display or edit the attribute values of a specific array

Array attributes are pieces of information specific to an array, such as chip barcode or hybridization temperature. In the Navigator, you can display or change attributes for each array.

1 Expand the folders of the Data pane or the **Experiment** pane until you can see the array of interest.

2 Right-click the name of the array, then click **Show Properties.**

The Microarray Properties dialog box appears, with a list of array attributes. See "Microarray Properties" on page 207. You can also edit the attributes of a specific array from this dialog box. In addition, if the array is an Agilent array, you can see header and feature information sent from the Agilent Feature Extraction program.

3 When you are finished, click **Close.**

NOTE You use the Sample Manager tab to organize, create, import, and export array attributes. See the *Sample Manager User Guide*.

Managing Content

This section describes how to create, find, rename, update, combine, and/or remove content such as data, gene lists, and tracks, stored in Agilent Genomic Workbench. To display the data, gene list and track content, see Chapter 3, “Displaying CH3 Data and Other Content”.

To display a list of the content stored in the program

The Data and My Entity List panes of the Navigator show the content stored in Agilent Genomic Workbench. For more information on the Navigator and its contents, see “Navigator” on page 135.

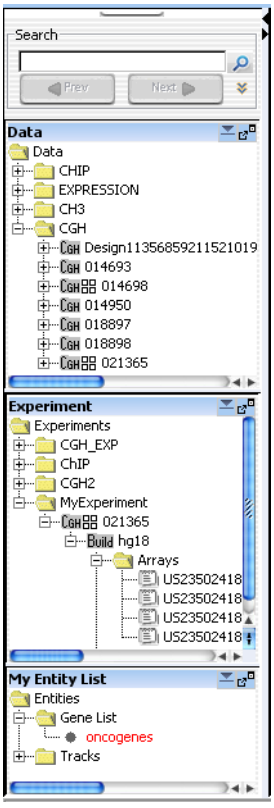


Figure 16 Navigator panes

Data pane – Shows all of the design and array data files stored in the database. For more information, see "To display the array designs and data in the program" on page 50.

My Entity List pane – Shows the gene lists and tracks stored in the program. To display the names of gene lists or tracks available in the program, double-click the names of folders to expand or collapse them, or click the ⊞ or ⊟ buttons.

To find specific content items in the Navigator

At the top of the Navigator is a search pane that can help you find specific content items. See "Search pane" on page 137.

1. Type a search term in the box at the top of the Navigator. The search term is not case-sensitive, but it must reflect the entire name of the content item that you want to find. You can use asterisks (*) as wildcards to represent a group of unspecified characters. For example, if you type *1234*, the search will find all items that contain "1234" in the name.
2. By default, the program searches all panes of the Navigator. To limit your search to a specific pane, click . In the list that appears, select the desired pane.
3. Click  .

   The program searches the selected pane(s). If it finds item(s) that match your search term, it expands folders so that the items are visible, and selects them in red. You may need to scroll down to see all the search results.
4. To clear the results of a search, click  .

To display the properties of a specific design

Design properties include general information about a design, such as its name, application type, and associated species. They also include a list of the names and chromosomal locations of probes.

1. Expand the folders of the Data pane until you can see the genome build folder(s) within the desired design folder.
2. Right-click the desired genome build folder, then click **Show Properties.**

   The Design Properties dialog box appears. See "Design Properties" on page 182.

To update probe annotation in design files

Agilent regularly makes updates to probe annotations on its eArray Web portal. If you have imported Agilent array designs into Agilent Genomic Workbench, and you are a registered eArray user, you can download the updated design files from within Agilent Genomic Workbench. For more information about eArray, go to https://earray.chem.agilent.com and click **Help.**

1 In the Home tab, click **User Preferences.**

   The User Preferences dialog box appears.

2 In the Miscellaneous tab, under **eArray User Details,** type your eArray **Username** and **Password.** See "User Preferences" on page 230.

3 Click **OK.**

4 Expand the folders of the Data pane until you can see the design that you want to update.

5 Right-click the desired design, then click **Update from eArray.** This option appears only for Agilent designs.

   A confirmation dialog box appears.

6 Click **Yes.**

   The program downloads an updated design, if one is available.

To rename an array in the Data pane

This topic describes how to rename an array in the Data pane, which changes the Global Display Name for the array. If you rename an array in this way, and subsequently add the array to an experiment, the array appears in the experiment with the new name. It also changes the array name in any experiment to which it is already linked. To rename an array only within the context of a specific experiment, see "To rename an array in an experiment" on page 56.

1 Expand the folders of the Data pane until you can see the array you want to rename.

2 Right-click the name of the array, then click **Rename.**

   An Input dialog box appears.

3 Type a new name for the array, then click **OK.**

The program renames the array.

To remove data or design files from the program

You can delete array design and data files from the program when you are finished with them.

1 If an array that you want to delete is associated with an experiment, first delete it from the experiment. See "To remove arrays from an experiment" on page 56.

2 In the Data pane, expand folders until you can see the design folder or array that you want to delete.

3 Do one of the following:

- For array data files, click the name of the first array, then hold down the **ctrl** key while you click the names of additional arrays within the same design.
- For array design folders, click the name of the first design folder, then hold down the **ctrl** key while you click the names of additional ones. This selects the designs and all array data files within them for deletion.

4 Right-click the name of a selected design folder or array data file, then click **Delete.**

A confirmation dialog box appears.

5 Click **Yes.**

The program deletes the selected files.

CAUTION When you delete files, you permanently remove them from Agilent Genomic Workbench. To restore deleted files, you must import them again.

To create a gene list

When you create a gene list, you create a list of the genes in a contiguous chromosomal region that you define. To create a list of genes in multiple regions, create multiple gene lists, and combine them. See "To add one gene list to another" on page 67.

1 Follow these steps to define a chromosomal region for your gene list. If you know the exact start and end locations of the desired chromosomal region, skip to step 2.

   a In Genome View, select the desired chromosome.
   The selected chromosome appears in Chromosome View. See "Chromosome View" on page 153.

   b In Chromosome View, in the plotting area to the right of the chromosome, drag the pointer over the chromosomal region of interest.
   The program draws a blue box around the region, and displays the region in greater detail in Gene View.

   c In Gene View, adjust the view so only the genes of interest appear. For a description of the adjustment commands available in Gene View, see "Gene View" on page 155.

2 Right-click anywhere within the log ratio plotting area in Gene View, then click **Create Gene List**.

   The Create Gene List dialog box appears. See "Create Gene List" on page 179.

3 In the dialog box set the Name, Description and Color.

4 In the dialog box select the chromosomal region for the new gene list.

5 Click **OK.**

   The new gene list appears in the Gene List folder in the My Entity List pane of the Navigator.

To import a gene list

A gene list file is a plain text (*.txt) file that contains one gene name per line. When you import a gene list into Agilent Genomic Workbench, it appears in the Gene List folder in the My Entities List pane. You can use the gene list to highlight specific genes, or to show or hide the appearance of genes and data, in Gene and Chromosome Views. See "To show gene lists in Gene View" on page 90.

1. In the **My Entities List** pane, double click the **Entities** folder to expand it.
2. Right-click the **Gene List** folder, then click **Import Gene List.**

   An Import dialog box appears. See "Import" on page 200.
3. Select the desired gene list file. To select additional gene list files, hold down the **ctrl** key and click their names.

   Click **OK.**

To display the genes in a gene list

You can display the genes in a gene list as a table.

1. Expand the folders in the **My Entity List** pane until you can see the desired gene list.
2. Right-click the gene list, then click **View In Table.**

   The Gene List dialog box appears, with a table that contains the names of the genes in the gene list. You can also use this dialog box to edit the description of the gene list and its display color. See "Gene List" on page 198.

   You can also create gene lists. For more information, see "To create a gene list" on page 65.

To add one gene list to another

You can add one gene list (a source gene list) to another (the target gene list). The program appends the source gene list to the end of the target gene list, and leaves the source gene list unchanged.

1 Expand the folders in the **My Entity List** pane until you can see the gene lists that you want to combine.

2 Right-click the desired source gene list, then click **Add to Gene List.**

A dialog box appears. For more information, see "Add Gene List <name> to" on page 165.

3 In **Select target gene list,** select the desired target gene list.

4 Click **OK.**

To rename a gene list

The name of a gene list identifies it within the Gene List folder of the My Entity List pane. You can rename gene lists.

1 Expand the folders of the **My Entity List** pane until you can see the gene list that you want to rename.

2 Right-click the desired gene list, then click **Rename.**

An Input dialog box appears.

3 Type a new name for the gene list, then click **OK.**

To delete gene list(s)

1 In the **My Entity List** pane of the Navigator, click to expand the **Gene List** folder.

2 Click the name of a gene list to delete. Hold down the **ctrl** key while you click the names of additional gene lists.

3 Right-click one of the selected gene lists, then click **Delete.**

A confirmation dialog box appears.

4 Click **Yes.**

To display the details of a track

You can display a table that contains the values for a list of track attributes.

1 In **My Entity List** pane, expand the Tracks folder to see the track.

2 Right-click the name of the track, then click **View Details.**

3 Data describing the track appear in a Track table. See "Track" on page 222.

To combine tracks

You can create a track that contains elements from two or more existing tracks. The existing tracks must be available in Agilent Genomic Workbench, and they must be associated with the same genome build.

1 In the **My Entities List** pane, double-click the **Entities** folder to expand it, if necessary.

2 Right-click the **Tracks** folder, then click **Combine Tracks.**

The Combine Tracks dialog box appears. See "Combine Tracks" on page 171.

3 In **Name,** type a name for the combined track. The program uses this name to identify the track in the Tracks folder, and to label the track if it appears in Gene View.

4 Click **New Condition.**

A new row appears in the Track/Operator list.

5 Under **Track,** select the first track to combine.

6 Click **New Condition,** then select another Track/Operator pair. You can set up as many Track/Operator pairs as you like, but you must set up at least two. When you add a track, the program automatically assigns the AND operator to the previous track.

To remove the bottom row from the list, click **Delete Condition.** To delete all rows from the list, and erase any entry in Name, click **Reset.**

7 Under **Operator,** select one of the following:

Operator	Comments
AND	Creates a combined track out of 2 tracks that will contain elements that appear in both tracks
OR	Creates a combined track out of 2 tracks that will contain elements that appear in either of the tracks
MINUS	Removes the elements of the second track from the first track.

8 Click **Save.**

Your combined track appears in the Tracks folder of the My Entities List pane. The Combine Tracks dialog box remains open for you to create another combined track.

9 Click **Close.**

To rename a track

The name of a track identifies it both within the Tracks folder of the My Entity List pane, and in Gene View when you select **Show In UI** for the track. You can rename tracks.

1 Expand the folders of the My Entity List pane until you can see the track that you want to rename.

2 Right-click the desired track, then click **Rename.**

An Input dialog box appears.

3 Type a new name for the track, then click **OK.**

To delete tracks

1. In the My Entity List pane of the Navigator, expand the Tracks folder.
2. Click the name of a track to delete. Hold down the **ctrl** key while you click the names of additional tracks.
3. Right-click one of the selected tracks, then click **Delete.**

   A confirmation dialog box appears.
4. Click **Yes.**

Exporting and Saving Content

This section describes how to export several kinds of files from the program.

To export array attributes

You can export selected array attributes for any imported arrays. You first select the arrays and then the attributes for the selected arrays. You can export array attributes from the Home tab or from the short-cut menu for an experiment.

1 Click **Home > Export > Array Attributes.**

OR

In the Experiment pane of the Navigator, right-click an experiment of interest, and click **Export Attributes**.

The Export Array Attributes dialog box appears with the Array tab displayed. See "Export Array Attributes" on page 190.

If you opened this dialog box by right-clicking an experiment, only those arrays selected for the experiment appear in the Selected Array List. You can add or subtract from the list.

2 Under Select Design, select the design file and genome build for the arrays to add.

The arrays for the selected design file and genome build appear in Array List.

3 In Array List, select the arrays whose attributes you intend to export. To select a single array, click its name. To select additional arrays, hold down the **ctrl** key while you click their names.

4 Click  .

The program moves the selected arrays to the Selected Array List.

5 Click **Next** to select attributes for the selected arrays.

The Export Array Attributes dialog box appears with the Attribute tab displayed. See "Attribute Tab" on page 192.

All of the attributes for the arrays are already located in the Selected Attribute List.

6 Move any attributes you don't want to export to the Available List.

   a In the Selected Attributes List, select those attributes you do not intend to export.

   b Click  .

7 Click **OK**.

   The Export dialog box appears. See "Export" on page 189.

8 Select the folder in which to save the attributes, and click **Export**.

   The attributes are saved to the selected folder as a .txt file.

To export experiments

You can export experiments as a ZIP file to transfer them to another computer. Exported experiments contain the associated design and array data files, only. The program does not export information about array selections, or any analysis parameters or results.

1 In the Home tab, click **Export > Experiments.**

   The Export Experiments dialog box appears. See "Export Experiments" on page 194.

2 Mark the experiments that you want to export. To export all experiments, click **Select All.**

3 Click **OK.**

   An Export dialog box appears. See "Export" on page 189.

4 Select a location and type a name for the exported ZIP file.

5 Click **Export.**

   The program exports all selected experiment(s) together as a single ZIP file.

To export a gene list

You can export a gene list as a text file that contains one gene per line.

1 In the My Entity List pane, in the Gene List folder, right-click the gene list that you want to export, then click **Save As.**

   A Save As dialog box appears.

2 Select a location and type a name for the file.

3 Click **Save.**

   A message appears when the operation is complete.

4 Click **OK.**

To export tracks

You can export selected tracks as a BED format track file. You can then import this file into Agilent Genomic Workbench on another computer, or into a genome browser that accepts BED format files.

1 In the Home tab, click **Export > Tracks**.

   The Export Tracks dialog box appears. See "Export Tracks" on page 195.

2 Mark the tracks to export. To select all tracks for export, click **Select All.**

3 Click **OK.**

   An Export dialog box appears.

4 Select a location and type a name for the exported track file, then click **Export.**

   The program exports the track(s) as a single BED format track file.

To copy what you see in the main window

You can copy panes of the main window to the clipboard as images, and then paste them into a new document in another program (such as Word, or PowerPoint). The images contain only what actually appears on your screen — regions to which you must scroll are not included.

1 In the View tab, click **Copy.**

2 In the shortcut menu that appears, click the name of the pane that you want to copy. You can copy any view, or the Navigator. To copy all of the panes, click **All.**

The program copies the selected pane(s) to the clipboard.

NOTE To adjust how data is displayed in the panes, use the View Preferences dialog box. See "View Preferences" on page 235 for more information.

3 Open a document in a program that accepts images. In that program, click **Edit > Paste,** or the appropriate paste command.

To copy the list of array colors for an experiment

You can copy the list of arrays in an experiment, and the colors assigned to them, to the clipboard as an image. You then paste the image into a document in another program such as Word or PowerPoint.

1 In the Experiment pane, expand the **Experiments** folder.

2 Right-click the name of the desired experiment, then click **Edit Array Color.**

The Edit Array Color dialog box appears. See "Edit Array Color" on page 185.

3 In the dialog box, click **Edit > Copy.**

The program copies the names of the arrays and their colors to the clipboard as an image.

4 Open a program that accepts images. Click **Edit > Paste,** or the appropriate paste command for the specific program.

To save data and design information from an experiment

You can save the data and design information from a single design in an experiment as a tab-delimited text file.

1. In the Experiment pane, expand the **Experiments** folder until you see the genome build(s) for the design you want to export.
2. Right-click the name of the desired genome build, then click **Save As Text File**.

   A dialog box appears.
3. Select a location and type a name for the file, then click **Save.**

3 Displaying CH3 Data and Other Content

This chapter shows you how to display log ratio data from imported feature extraction data files, as well as gene and track content, in the Genomic Viewer. It also has instructions on how to customize the display of data and content to meet your needs.

To learn about the options for the main window and the dialog boxes for displaying data, see Chapter 5, “Methylation (CH3) Analysis Reference”.

Selecting an Experiment

An experiment is a set of links to microarray data and design files, and any associated results. You can see a list of the experiments in the Experiments pane of the Navigator. See "Navigator" on page 135 for more information.

When you select an experiment and the Preprocessing and Analysis options have not been turned on or set to apply, the program shows the log ratio data of selected arrays in the active experiment, if certain options have been set. See "To locate and display data (or results) within the Views" on page 87 for more information.

When you select an experiment and Preprocessing and Analysis options have been turned on or set to apply, the program automatically begins the analysis of the selected array data with current settings and displays its results.

This section describes how to select an experiment, select or deselect arrays for further analysis, and analyze arrays one at a time.

To select an experiment

When you select an experiment, the program begins the analysis with the current settings. You can either set the Preprocessing and Analysis parameters before you select the experiment or change the settings one at a time after the first analysis and reanalyze. Every time you change a Preprocessing or Analysis setting for an active experiment, the program recalculates results.

1 If necessary, do one of the following to add the desired experiment to the Experiment Pane in the Navigator:

- Create a new experiment and add data to it. See "To create a new experiment" on page 51.
- Import a saved experiment file. See "To import an experiment file" on page 48.

2 In the Navigator, double-click the name of the experiment.

The Experiment Selection dialog box appears.

3 Click **Yes.**

In the Experiment pane of the Navigator, the name of the experiment turns blue. The name also appears in the title bar of the main window. Tables of data and design information appear in Tab View.

If you have selected to show the results of an algorithm calculation, then results appear for the first array when you select the experiment, if you have not selected any other arrays.

You can select or deselect arrays in the experiment before or after you select the experiment. Every time you select or deselect an array in a *selected* experiment or change a setting, the program reanalyzes the new data set with the changed settings. See "To select or deselect arrays in the experiment" on page 79.

When you select the experiment after deselecting it or selecting another one, the experiment is simply restored if the settings in the UI haven't changed. If they have changed, the program reanalyzes all of the arrays assigned when the experiment was last selected.

To select or deselect arrays in the experiment

To include arrays for display and analysis, you select them from the arrays available, either in an inactive experiment or the selected one. When you first create an experiment, the program automatically sets the first array in the experiment for analysis. If you do not select additional arrays for analysis, only the first one is analyzed when the experiment is selected.

To select the arrays for analysis before experiment selection:

1 Hold down the **Shift** key to highlight contiguous arrays or hold down the **ctrl** key to highlight noncontiguous arrays.

2 Right-click the highlighted arrays, and click **Select**.

Even though the selected arrays do not change color, they will change color after activation.

In the Navigator, the color or an array's icon has the following meaning after experiment selection:

Array is not selected.

Array is selected. The specific color matches the color of the column headings for the array in Tab View. In addition, the program displays aberration results and moving averages related to this array in this color. To configure a custom color for the array, see "To change the display color of an array" on page 81.

To select or deselect arrays in a *selected* experiment:

1. In the Navigator, expand the folders of the selected experiment.
2. Click the name of an array you want to include in the display.

   To include additional arrays, hold down the **ctrl** key while you click their names. To include a contiguous block of arrays, click the name of the first array in the block, then hold down the **Shift** key while you click the name of the last one.
3. Right-click the name of one of the highlighted arrays, then click **Select.**
4. After you select the arrays, if you have Analysis options selected (Methylation or Batman, for example), clear the options and then mark them again. The program re-analyzes the data set within the experiment and displays the data in Genome, Chromosome, and Gene Views. You can see the data and results for just the selected arrays in the Selected Arrays tab in Tab View.

   To show analysis results if they do not appear, see "To display results of analysis" on page 111.

   To customize the appearance of the results in Genome, Chromosome, and Gene Views, see "To change scatter plot appearance" on page 85.

You can also use the headings of columns in Tab View that contain array data to select and deselect arrays.

- Click a column heading to select that array only.
- Hold down the **ctrl** key while you click a column heading to select or deselect an array without changing the status of other arrays.
- Right-click a column heading to open a shortcut menu with options that let you select or deselect that array, or all arrays.

For more information on Tab View, see "Tab View" on page 160.

To change the display color of an array

The color assigned to an array sets the color of its icon when you select the array within an experiment. It also changes the colored square in the array's column heading in Tab View.

1 In the Experiment pane, in the Experiments folder, expand the folder of an experiment until you can see the array whose color you want to edit.

2 Right-click the desired array, then click **Edit Array Color.**

The Select Color dialog box appears. The dialog box offers three different ways to choose the desired color. See "Select Color" on page 215.

3 Select the desired color in one of the following ways:

Dialog box tab	Instructions
Swatches	• Click the desired color swatch.
HSB (Hue/Saturation/Brightness)	Type or adjust the values in H (Hue), S (Saturation), and B (Brightness), or alternately, follow these steps: **a** Select **H**, then drag the slider to select a hue based on the color strip to its right. **b** Click an appropriate location in the large color box to the left of the slider to set the saturation and brightness levels of the color. Both the HSB and equivalent RGB values of the color appear in the dialog box. Note these values; they will be useful if you need to duplicate this color in the future.
RGB (Red/Green/Blue)	Do any of the following. Note the final RGB Values; they will be useful if you need to duplicate this color in the future. • Drag the Red, Green, and Blue sliders. • Type or adjust values in the boxes to the right of the sliders.

Samples of the color in different contexts appear under Preview. The upper half of the color sample on the right shows the original color for comparison.

4 Adjust the color as desired, then click **OK.**

You can also manage all of the colors for all of the arrays in an experiment. Right-click the desired experiment, then click **Edit Array Color.** For more information, see "Edit Array Color" on page 185.

Displaying Array Data

After you select an experiment, you can change how data appear within the Views or change the appearance of the Views that contain the data (or results).

To display the scatter plots

By default, display of scatter plots is turned On. If you do not see the scatter plot(s), do one of the following:

1 From the View tab, click **View Preferences**. See "View Preferences" on page 235 for more information.

2 In the View Preferences dialog box, under Data Visibility, select **All views** and then mark the box next to **Scatter Plot**.

OR

1 Right-click in any of the views, and select **View Preferences**. See "View Preferences" on page 235 for more information.

2 In the View Preferences dialog box, under Data Visibility, select **All views** and then mark the box next to **Scatter Plot**.

To show or hide data in the scatter plots

1 In the Gene View, move the mouse cursor over the arrow next to **Scatter Plot,** and do any of the following:

To do this	Follow these steps
Show or hide Z-Score values in the Methylation Results plot	• To show the data points – Mark the **Methylation Results** check box and select **Z-Score Values** from the list. • To hide all data points – Clear the **Methylation Results** check box.
Show or hide Log Ratio values in the Log Ratios plot	• To show the data points – Mark 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 Probe Score values in the Log Ratios plot	• To show the data points – Mark the **Log Ratios** check box and select **Probe Score Values** from the list. • To hide the data points, Clear the **Log Ratios** check box.
Change the ranges and colors for all scatter plots	• Click **Configure Color and Ranges** to enter ranges and change colors. See "Configure Coloring Ranges and Shades" on page 173 for more information.

2 Click X to close the Scatter Plot window.

To customize scatter plot ranges and colors

In order to make it easier to see significant results, you can customize the display of scatter plot data. For each data type (Z-Score, Log Ratio, probe score) you can set custom ranges and colors for the display. For channels, you can set custom colors only.

Add and customize a range

1 In Gene View, move the mouse pointer over **Scatter Plot** to display the options.

2 Mark the check box under **Methylation Results** or **Log Ratios**.

3 Select a data type from the **Color by** list.

4 Click **Configure Color and Ranges**.

The Configure Coloring Ranges and Shades dialog box appears where you set ranges and colors for any of the data types. For more information, see "Configure Coloring Ranges and Shades" on page 173.

5 In the Configure Coloring ranges and Shades dialog box, click the **Methylation Results** or **Log Ratios** tab and then select the data type to configure.

6 Enter minimum and maximum numbers to define a range for the data type.

7 Click **Color** to open the **Select Color** dialog box. Use the tabs to select a color for the range. See "Select Color" on page 215 for more information.

8 Click **OK** to close the **Select Color** dialog box and return to the Configure Coloring ranges and Shades dialog box.

9 Click **Add Range** to add the custom range to the range list.

10 When you are done, click **OK** to close the dialog box.

Edit or remove a range

1 In the Configure Coloring ranges and Shades dialog box, click the **Methylation Results** or **Log Ratios** tab and then select the data type to configure.

2 In the range list, mark the **Edit/Delete** box to select the range. You can mark more than one range.

3 Click **Edit Range** to change the minimum and maximum values, or to change the color for the selected range.

4 Click **Delete Range** to delete the selected range.

5 Click **OK** to close the dialog box.

To change scatter plot appearance

You use the View Preferences dialog box to change the appearance of the scatter plots in Chromosome and Gene views.

1 Right-click any part of a scatter plot, then click **View Preferences.**

Or, click the View tab, and then click **View Preferences**.

The View Preferences dialog box appears. See "View Preferences" on page 235.

2 Do any of the following:

To do this	Follow these steps
Show or hide the scatter plot	**a** In the View tab under **Data Visibility,** in **View,** select **All Views.** **b** Do one of the following: To show the scatter plot, mark **Scatter Plot.** To hide the scatter plot, clear **Scatter Plot.** **c** Click **OK.**
Change the symbol that appears for data points	You can select the symbol separately for each design type. **a** In the View tab, under **Rendering Patterns,** select the desired **Design type.** **b** Under **Styles,** for each data type, select the desired symbol. **c** Click **Apply.**
Show a separate scatter plot in Gene and Chromosome Views for each selected array	**a** In the View tab, under **View Alignment,** under **Rendering Style,** select **Stacked.** **b** Click **Apply.**
Show one scatter plot that contains data for selected arrays	**a** In the View tab, under **View Alignment,** under **Rendering Style,** select **Overlaid.** **b** Click **Apply.**
Enable ToolTips for the scatter plot in Gene View	ToolTips show information about an individual data point when you place the pointer over it. **a** Click the **View** tab. **b** Under **Data Visibility,** in **View,** select **Gene View.** **c** Mark **Scatter Tool Tip.** **d** Click **Apply.**

3 Click **OK.**

To print the scatter plot

You can print the scatter plot as it appears in Genome, Chromosome, and Gene views. Each view selected in the analysis is printed on separate pages. Chromosomes and genes appear on the printed pages, but tracks do not.

1 In the Home tab, click **Print**.

The Print dialog box appears.

2 Set print options, then click **OK.**

To create custom scales for Views

You can customize the scale used for display in the Chromosome View and Gene View. Custom scales are applied to both views.

1 Click the View tab and then click **View Preferences.**

2 In the View Preferences dialog box, under Configure Scales, mark the box next to **Apply** for Methylation Results and/or Log Ratios.

3 In Range, type a value to use for the range. The range you type changes the scale for the display of the selected data.

To locate and display data (or results) within the Views

To look through the data of the selected arrays, do any of the following. In general, all views are synchronized; if you select a location or region in one view, the other views move there as well.

To do this	Follow these steps
Select a specific chromosome to display	• In Genome View, click the chromosome. All other views switch to the selected chromosome.
Display data in a region of the selected chromosome	• In Chromosome View, drag the pointer over the region. Gene View expands (or shrinks) to show only the selected region. Tab View scrolls to the new cursor location.

To do this	Follow these steps
Zoom in and out in Gene View	• Click  to zoom in. • Click  to zoom out.
Scroll through the selected chromosome	• Click  to scroll up. • Click  to scroll down. **Note:** These arrows will appear side by side for horizontal orientation.
Re-center Gene View or Chromosome view	Click anywhere in Chromosome View, or anywhere within the scatter plot in Gene View. The location you click becomes the new cursor location.
Move all views to a specific genomic location	**a** Click **Home > Go To Gene/Genomic location.** A dialog box appears. **b** Under **Genomic Location,** select a **Chromosome,** and type a **Base Position.** **c** Click **Go.** All views move to the selected location.
Display the location of a specific gene in the center of all Views	**a** Click **Home > Go To Gene/Genomic location.** A dialog box appears. **b** Under **RefSeq by Symbol,** either select the gene (if available) or type the name of the gene. **c** Click **Go.** All views move to the location of the selected gene.
Display the data selected in Tab View in the center of Chromosome and Gene Views	• In Tab View, click any entry in any table, except a column heading. **Chromosome and Gene views:** The genetic location of the selected data appears in the center of Chromosome and Gene Views.
Scroll to a specific column in Tab View (for the selected chromosome)	**a** In Tab View, right-click any column heading, then click **Scroll To Column.** The Scroll to Column dialog box appears. See "Scroll to Column" on page 213. **b** In **Select Column,** select the column. **c** Click **OK.**

To do this	Follow these steps
Search for a specific column entry in Tab View, and move the cursor there	**a** In Tab View, right-click any entry except a column heading, then click **Find in column.** The Find in column dialog box appears. See "Find in column" on page 196. **b** Set the search parameters, then click **Find Next.** The program searches the column using your search parameters, and highlights the row of the first entry that matches. The cursor moves to the location defined in the highlighted row.
Display the exact chromosomal location of the cursor	At the bottom of the main window, look at the first cell of the Status bar. The location appears as the chromosome followed by the base position. For more information on the status bar, see "Status Bar" on page 164.

Displaying Content (Gene Lists/Tracks)

To show gene lists in Gene View

A gene list is a set of genes of interest. Within the program, you can highlight the genes in the gene list in Gene View, or limit the display of data, genes, and tracks to the regions selected by a gene list.

You can import gene lists into Agilent Genomic Workbench, and you can also create them in the program and export them. See "To import a gene list" on page 66, and "To export a gene list" on page 73.

In Gene View, the names of all genes normally appear in gray. When you apply a gene list, the program highlights the listed genes in their defined display color. You can also limit the genes and/or data that appear in Gene View and Chromosome View to only the listed genes.

1 In the My Entity List pane of the Navigator, expand the **Gene List** folder. If the gene list does not appear, create or import it. See "To create a gene list" on page 65, or "To import a gene list" on page 66.

2 Right-click the gene list, then do one of the following to apply it:

- To show all genes and all data, and highlight the listed genes in their display color, click **Highlight.**
- To show only the listed genes and only the data for those genes, click **Show only.**

Gene and Chromosome views change accordingly. In the My Entity List pane of the Navigator, the name of the gene list appears in italics.

To remove the effects of a gene list, right-click the active gene list in the Navigator, then click **Show All.**

To select gene list display color

In Gene View, the names of all genes normally appear in gray. When you apply a gene list, the program highlights the listed genes in their defined display color. You can customize this color.

1 In the My Entity List pane of the Navigator, expand the **Gene List** folder.

2 Right-click the name of the gene list whose color you want to change, then click **View in Table.**

The Gene List dialog box appears.

3 Under **Color,** click **Color.**

A dialog box appears.

4 Select the desired color.

The dialog box offers three different ways to select the color. See "To select gene list display color" on page 91.

5 Adjust the color as desired, then click **OK.**

6 In the Gene List dialog box, click **OK.**

To display a gene list as a table

You can display the description of a gene list and the names of the genes in it.

1 In the My Entity List pane of the Navigator, in the **Gene List** folder, right-click the desired gene list, then click **View in Table.**

The Gene List dialog box appears. See "Gene List" on page 198. The names of the genes appear in Gene Names. You can also use this dialog box to edit the description of the gene list, or to change its display color. To change the display color, see "To select gene list display color" on page 91.

2 When you are finished displaying the list, click **OK.**

You can also export a gene list. See "To export a gene list" on page 73.

To change the appearance of genes in Gene View

You use the User Preferences dialog box to change the appearance of the genes in Chromosome and Gene views.

1 Right-click any part of the Gene View, then click **User Preferences.**

The User Preferences dialog box appears.

2 Click **Tracks**.

See "Tracks tab" on page 230.

3 Do one or more of the following:

To do this	Follow these steps
Show or hide genes in Gene View	**a** Under **Visualization Parameters**: To show genes – Under **Genes,** mark **Show Gene Symbols.** To hide genes – Under **Genes,** clear **Show Gene Symbols.** **b** Click **Apply.**
Change the display font for genes (and track annotations) in Gene View	**a** In the Gene Symbols tab, under **Font,** select a new **Font, Font Style,** and **Font Size.** **b** Click **Apply**
Change the display angle for genes (and track annotations) in Gene View	**a** Under **Visualization Parameters,** under **Genes,** in **Orientation (Degrees),** type a new orientation in degrees. 0° is horizontal. **b** Click **Apply.**

4 Click **OK.**

To show tracks in Gene View

Tracks contain information for specific genomic locations. A multitude of tracks from diverse sources is available for many species. You can display tracks next to genes and microarray data in Gene View.

1 Select and show microarray data. See "To select an experiment" on page 78.

2 In the My Entity List pane, open the Tracks folder.

3 Right-click the track you want to display, and click **Show In UI**.

Or, you can do this:

1 In Gene View, right-click anywhere within the scatter plot, then click **User Preferences.**

The User Preferences dialog box appears. See "User Preferences" on page 230.

2 Click **Tracks**.

3 Mark the **Show In UI** check box of each track.

4 Click **OK.**

The program displays the selected tracks in Gene View.

To change the appearance of tracks

In the Tracks tab of the User Preferences dialog box, you can change the appearance of tracks as described in the table below.

To do this	Follow these steps
Include track information in reports	**a** In the list of tracks, in the **Show in Report** column, mark the check boxes of the desired tracks. **b** Click **Apply.** Doing this adds a column with the hits from the track file. For each aberrant interval, it reports the entries from the track file for that interval in that separate column.
Show or hide annotations in all tracks	• To show annotations in all tracks: under **Tracks,** mark **Show Annotations.** • To hide annotations in all tracks: under **Tracks,** clear **Show Annotations.**
Display all selected tracks as a single track	• Under **Tracks,** mark **Show Overlaid.** The program combines the annotations of all selected tracks into a single track named **Overlaid Track.** • To show tracks individually again, clear **Show Overlaid.**
Display the parameters and the list of annotations of a track	• In the list of tracks, for the desired track, click **Details.**

To do this	Follow these steps
Change the display font for track annotations (and genes)	**a** Under **Font,** select a new **Font, Font Style,** and **Font Size** for track annotations. **b** Click **Apply.** The program changes the display font of track annotations and genes in Gene View.
Change the order in which tracks appear in Gene View.	The order of tracks in the Gene Symbols tab controls the left-to-right order of tracks in Gene View. **a** Click the name of the track you want to move. **b** Do any of the following: • To move the track up in the list of tracks (and farther left in Gene View), click its name, then click **Up.** • To move the track down in the list of tracks (and farther right in Gene View), click its name, then click **Down.** **c** Click **Apply.**
Change the display angle of track annotations (and genes)	• Under **Genes,** in **Orientation,** type a new orientation (in degrees). 0° is horizontal. The program changes the display angle of track annotations and genes in Gene View.

To show track information in reports

1 In the list of tracks, in the **Show in Report** column, mark the check boxes of the desired tracks.

2 Click **Apply.**

Doing this adds a column with the hits from the track file. For each aberrant interval, it reports the entries from the track file for that interval in that separate column.

NOTE The annotations that are displayed in tab view are taken from the design file. This may include names of genes and other annotations. The positions of these genes may also be supplied as a separate Gene track and may contain unnecessary information.

To limit data to the genomic boundaries of the track

1 In the list of tracks in My Entity List, right-click the track whose boundaries you want to use to limit the display of the data.

2 Mark **Genomic Boundaries**.

You can remove the boundaries by clearing the check box.

To display a track in UCSC Browser

1 In the My Entity List pane, expand the Tracks folder and find the track you want to view in the UCSC browser.

2 Right-click the track name, and click **Show in UCSC**.

The UCSC Browser appears, if you are connected to the Internet (Figure 17). You may have to enable pop ups or set your other browser preferences on the UCSC browser Web site.

3 Follow the instructions on the web site for what you want to do.

To upload a track to UCSC Browser

1 Right-click Gene View, and click **Upload Track to UCSC**.

2 The Upload Track to UCSC dialog box appears.

Complete the dialog to define the track you wish to upload. See "Upload Track to UCSC" on page 228.

The UCSC Browser appears, if you are connected to the Internet (Figure 17). You may have to enable pop ups or set your other browser preferences on the UCSC browser Web site.

3 Follow the instructions on the web site for what you want to do.

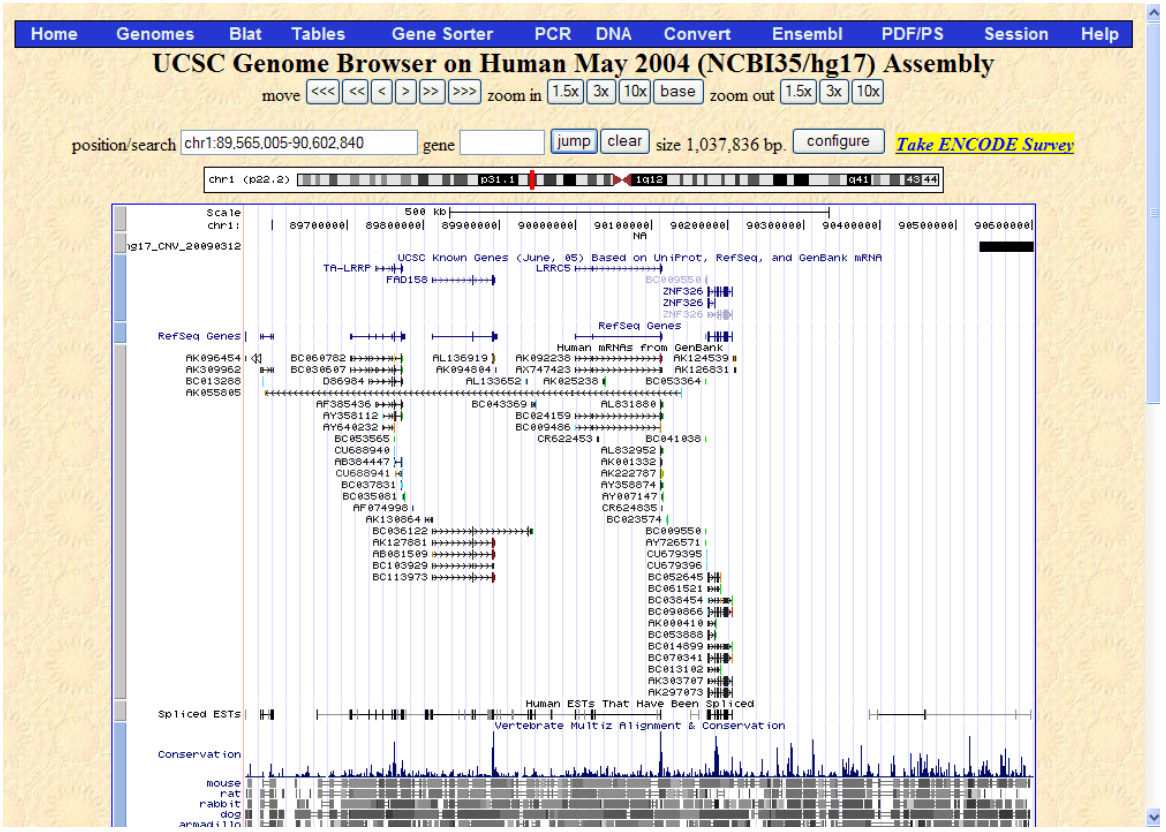


Figure 17 Track displayed in UCSC browser

To change the graphical display to a different genome build

The default graphical display for Genome, Chromosome and Gene Views represents human genome build 18.

- To change the graphical display to a different genome build, select an experiment whose data are based on a design file of a different genome build.

The display automatically changes when you select an experiment that contains a design file with a different genome build, such as human genome build 17, or a mouse or rat genome build.

If a genome build is not available for the design file you import, you must import the genome build first. See "To import a genome build" on page 45.

The program will not let you add arrays that belong to one genome build to an experiment that contains arrays of a different genome build.

See also "To create a new experiment" on page 51, "To add arrays to an experiment" on page 53, and "To select an experiment" on page 78.

Searching for Probe and Gene Information

To search Tab View for specific probe information

You can find a specific entry in a column of a data table in Tab View. For more information on Tab View, see "Tab View" on page 160.

1 In Tab View, right-click anywhere in the column you want to search, then click **Find in column.**

The Find in column dialog box appears. The search column also appears in the title bar of the dialog box.

NOTE The Find in column function works within the selected chromosome.

2 Set the search parameters, as described below.

Parameter	Comments/Instructions
Find in column	• Type the text you want to find (the *search term*). This can be an entire entry, or part of one.

Direction	• Select one of these options:
	  ◦ **Up** – Search the column upwards from the current cursor location (the highlighted row of the table).
	  ◦ **Down** – Search the column downwards from the current cursor location (the highlighted row of the table).
	Tip: Click a row in Tab View to highlight it.

Conditions	• Mark any of these, as desired:
	  ◦ **Match Case** – Find entries that match upper and lower case characters in the search term.
	  ◦ **Match whole word** – Find an entry only if the entire entry matches the search term.

3 Click **Find Next.**

If the program finds a match, it highlights the row that contains the matching entry, and resets the cursor to the corresponding position. You can click **Find Next** as many times as you want, and the program

continues to search for additional matching entries in the column. If it finds no match, the message: **String not found** appears in black in the lower part of the dialog box.

4 When you complete your search, click **Cancel** to close the dialog box.

To search Agilent eArray for probe information

You can use the chromosomal region that appears in Gene View, or another chromosomal region as the basis for a probe search on the Agilent eArray Web site. eArray is a powerful microarray design system for CGH, ChIP and gene expression applications. It contains a massive database of validated, annotated probes, and a full complement of tools for custom microarray design.

Before you can search for probes in eArray, you must register for an eArray account if you do not already have one. For more information, go to https://eArray.chem.agilent.com. Provide your eArray user name and password in the Miscellaneous tab of the User Preferences dialog box. See "User Preferences" on page 385.

1 In Gene View, right-click anywhere in the plotting area, then click **Search probes in eArray**.

The Search probes in eArray dialog box appears. See "Search Probes in eArray" on page 214.

2 Do one of the following to define the chromosomal region for your search:

- To set the region to the one that currently appears in Gene View, select **For complete gene view.**
- To set the region numerically, select **User Defined,** then select a **Chromosome** and type **Start** and **Stop** locations for the desired region.

3 Click **OK.**

The eArray Web portal opens in your internet browser.

To search the Web for information on probes in Tab View

You can use any entry in a table in Tab View as the basis for a Web search.

1 In Tab View, right-click any data table entry other than a column heading.

2 Click one of the available sites.

If the site of interest does not appear in the shortcut menu, you can create a custom search link. See "To create a custom Web search link" below.

The selected site opens in your Internet browser. The program sends the table entry to the site as a search string.

To create a custom Web search link

If you need to search a different database or site based on data table entries, you can create your own custom search link. When you right-click a table entry in Tab View, a shortcut menu opens, and your custom link appears in it. If you select this link, Agilent Genomic Workbench opens the site in your Web browser and sends the table entry to the site as a search string.

1 Right-click any data table entry in Tab View, except a column heading, then click **Customize Link.**

The Customize Search link dialog box appears. See "Customize Search Link" on page 181.

2 Click **New.**

3 In the Input dialog box, in **URL name,** type a name for the link.

This name will appear in the shortcut menu that opens when you right-click a data table entry.

4 Click **OK.**

5 In **URL,** type the complete URL needed to send a search string to the site. Use `<target>` as the query string value.

For example, this URL sends selected table entries to Google.com:
`http://www.google.com/search?hl=eng&q=<target>`

6 Click **Update**, then click **Yes.**

To update or delete a custom Web search link

1 Right-click any data table entry in Tab View other than a column heading, then click **Customize Link.**

The Customize Search link dialog box appears.

2 In **URL Name,** select the custom search link to update or delete.

3 Do one of the following:

To do this	Follow these steps
Update a Web search link	**a** Edit the **URL name** and the **URL** as needed. **b** Click **Update.** A Confirm dialog box appears. **c** Click **Yes.**
Delete a Web search link	• Click **Delete.**

4 Click **Close.**

4 Setting Up Methylation (CH3) Analysis

This chapter gives instructions on how to set up the analysis functions for Methylation (CH3) Analysis experiments. These include the Preprocessing, Analysis and Reports tabs.

Working with Methylation Options

For a detailed description of the Methylation (CH3) analysis tabs and their commands – Preprocessing, Analysis, Reports – see Chapter 5, "Methylation (CH3) Analysis Reference".

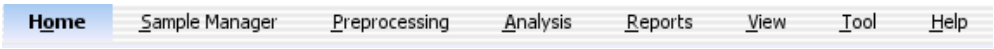


Figure 18 Tabs available in the CH3 application

Table 1 describes the sections in this chapter and the tasks you do during interactive analysis.

Table 2 Methylation analysis topics

Subject	See these topics
Changing Preprocessing and Analysis options	"To combine (fuse) arrays" on page 105 "To set up a moving average (Log Ratio) calculation to smooth the data" on page 107 "To set up a moving average (ZScore) calculation to smooth the data" on page 108 "To apply methylation (CH3) event detection" on page 109 "To apply BATMAN (Bayesian Tool for Methylation Analysis)" on page 110
Displaying results and generating reports	"To display results of analysis" on page 111 "To upload a track to UCSC" on page 113 "To save a result" on page 114 "To restore a saved result to the display" on page 115 "To generate a probe report" on page 115 "To generate a Batman report" on page 118

Changing Preprocessing and Analysis options

To combine (fuse) arrays

If you have two arrays that use different design files, you can combine (fuse) them into one larger virtual array. You do this to increase the coverage of the genome in your design. If for example, you have a catalog array, you can design another array to add probes between the catalog probes to increase the density of coverage. With the Fuse function, you can combine the array data to see all the probe data in the display at once. The program cannot combine arrays from more than two different design files.

Requirements for fusing arrays:

- Each array has a different design file.
- All of the arrays are of the application type. (For this application, CH3.)
- None of the arrays are fused arrays.
- The samples you hybridize to the arrays are all aliquots from the same preparation.
- (Preferred) Hybridization and labeling occur for all samples together under the same conditions.

1 Create and select a new experiment. See "To create a new experiment" on page 51 and "To select an experiment" on page 78.

2 Add the arrays to be combined to the experiment. See "To add arrays to an experiment" on page 53.

3 Assign the same value to the **ArraySet** attribute of every array you want to combine. Follow these steps for each array:

   a In the **Experiment** pane of the Navigator, right-click the name of the array, then click **Show Properties.**
   The Microarray Properties dialog box appears.

   b Next to the ArraySet attribute, under **Value,** click  .
   A text box appears.

   c Type a value in the text box. Type the same value for the ArraySet attribute of every array.

   d Click **Close.**

4 In the **Preprocessing** tab, under **Combine,** click **Fuse.**

   The Array Set dialog box lists the arrays to be combined. See "Array Set" on page 168.

NOTE Double-check the values in the ArraySet Attribute column of the dialog box. Agilent Genomic Workbench combines all of the array pairs that have the same value for this attribute.

5 Set any of these options, as desired:

   - **Select Normalization** – Select **None** or **Centralization.** Centralization adds or subtracts a constant value from each log ratio measurement. This recenters the log ratio values, and makes sure that the zero-point reflects the most common ploidy state.
   - **Remove arrays from experiment after fuse** – To delete the initial individual arrays from the experiment, mark this option so that the experiment will not contain duplicate data.

6 Click **Continue.**

   The program combines the arrays. The new array appears in the Experiment pane of the Navigator in a new design folder within the active experiment. The folder name contains the names of both designs.

NOTE

- If the original arrays change, you can manually update the combined array. Fuse the same set of arrays again.
- If you want to combine many arrays, consider importing an array attributes file. See "To import array attributes" on page 48.
- When you combine (fuse) arrays, the centralization algorithm (if selected) is applied to the individual arrays before they are combined.

To set up a moving average (Log Ratio) calculation to smooth the data

With the Moving Average tool, a moving average is computed for each point in the data set using a sliding window of fixed size, centered on that point. This algorithm uses the pre-algorithm probe log ratios to calculate a moving average line plot. See "Visualization algorithms" on page 243.

1 In the **Analysis** tab, under **Moving Average,** mark or clear **Show** to show or hide the Moving Average Line Plot.

2 In **Algorithm,** select **Linear** or **Triangular.**

- **Linear** – The linear algorithm calculates a standard, unweighted average using every probe Log Ratio score within a specified window size (specified by either a number of adjacent measurements or a positional range). This average is applied sequentially to every point by moving the window along the chromosome. See "Moving average (linear smoothing)" on page 259.
- **Triangular** – The triangular algorithm calculates a weighted average using every probe Log Ratio within a specified window size. The triangular algorithm is more sensitive to localized variations in the data. See "Triangular smoothing" on page 260.

3 In **Line width,** select the thickness (in pixels) of the Moving Average Line Plot.

The range is 1-5 pixels.

4 In **Window,** select the size of the moving average window.

Moving averages are computed with windows of sizes based on either a specific length of base-pairs (5 Kb to 50 Kb or .1 Mb to 50 Mb) or a fixed number of data points (1 pt. to 60 pt.).

To set up a moving average (ZScore) calculation to smooth the data

The Moving Average (ZScore) applies a moving average to each point in the combined methylation Z-score data set using a sliding window of fixed size, centered on that point. This algorithm uses the post algorithm combined Z-scores to calculate a moving average line plot. See "Overview of Methylation Detection and Visualization Algorithms" on page 242.

1 In the **Analysis** tab, in the **Moving Average (ZScore)** group, mark or clear **Show** to show or hide the Moving Average Line Plot.

2 In **Algorithm,** select **Linear** or **Triangular.** See "Visualization algorithms" on page 243**.**

Linear – The linear algorithm calculates a standard, unweighted average using every combined Z-score within a window size (that is defined by either a number of adjacent measurements or a positional range). This average is applied sequentially to every point by moving the window along the chromosome. See "Visualization algorithms" on page 243.

Triangular – The triangular algorithm calculates a weighted average using every combined Z-score within a specified window size. The triangular algorithm is more sensitive to localized variations in the data. See "Triangular smoothing" on page 260.

3 In **Line width,** select the thickness (in pixels) for the Moving Average Line Plot.

The range is 1-5 pixels.

4 In **Window,** select the size of the moving average window, or type the size of the window.

Moving averages are computed with windows of sizes based on either a specific length of base-pairs (5 Kb to 50 Kb or .1 Mb to 50 Mb) or a fixed number of data points (1 pt. to 60 pt.).

The window can be set manually to as low as .5 Kb (type .5 Kb).

To apply methylation (CH3) event detection

The purpose of a methylation detection algorithm is to identify methylation sites located within genomic regions such as promoters, CpG islands, etc. It is used in conjunction with CpG Island tracks that either already exist in the program's database or that you import.

1 Set up genomic region tracks (such as CpG islands) in Gene View.

  See "To import tracks" on page 46 for instructions on how to import and display these tracks in Gene View and in the report.

2 Select the experiment now, or after step 3.

  See "To select an experiment" on page 78.

3 Click **Analysis**.

4 Under **Probe Methylation Select ZScore**, select the algorithm, and then mark **Apply**.

  The Probe Methylation Status Setup dialog box appears.

  See "Probe Methylation Status Setup" on page 210.

5 To select a Tm (melting temperature) mapping file, click **Browse** and find the folder containing the file.

6 Select the file, and click **Open**.

7 Click **Continue**.

The algorithm is automatically applied if you've already selected the experiment, or is applied when you select it. To learn how the algorithms work see "Methylation Detection and Measurement Algorithms" on page 244.

The Z-score or BATMAN score results appear in a second channel of Gene View next to the log ratio data. You can now calculate a moving average for the Z-scores. See the next topic.

To apply BATMAN (Bayesian Tool for Methylation Analysis)

When BATMAN is selected, a Bayesian tool is used to calculate methylation results (BATMAN). The BATMAN algorithm reflects the underlying distribution of CpG dinucleotides by calculating a Gaussian distribution of methylated sites to give the observed probe signal. The signal used in BATMAN is either the probe log ratio or the methylation detection algorithm Z- score. For more information on the BATMAN algorithm, see "Bayesian tool for methylation analysis (BATMAN)" on page 254.

1 In the **Analysis** tab, under **Batman**, mark the check box to enable the Batman algorithm.

   The **Batman Parameter Setup** dialog box is displayed. See "Batman Parameter Setup" on page 169**.**

2 Complete the Batman Parameter Setup and click **Continue** to apply the calculations and show results in the views.

Displaying results and generating reports

The section shows you how to make sure the analysis results are displayed automatically after you select an experiment. To learn more about what the results mean, see "Visualization algorithms" on page 243.

To display results of analysis

After you set up an experiment and analyze it, the program displays the results automatically in Genome, Chromosome, and Gene Views if the display is turned on. Figure 19 and Figure 20 show examples of results displayed in the Gene View. To customize the way the results appear, see Chapter 3, "Displaying CH3 Data and Other Content".

Results for probe methylation and log ratios compared to a CpG Island track

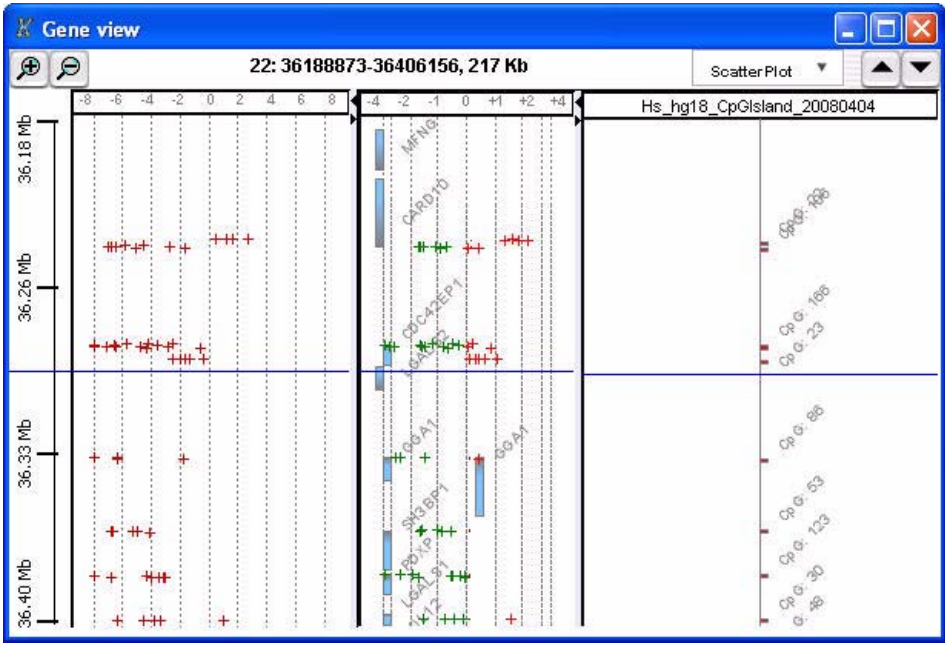


Figure 19 Gene View, showing Z-score data next to log ratios and a CpG Island track

Results for moving averages within genomic boundaries of CpG Island track

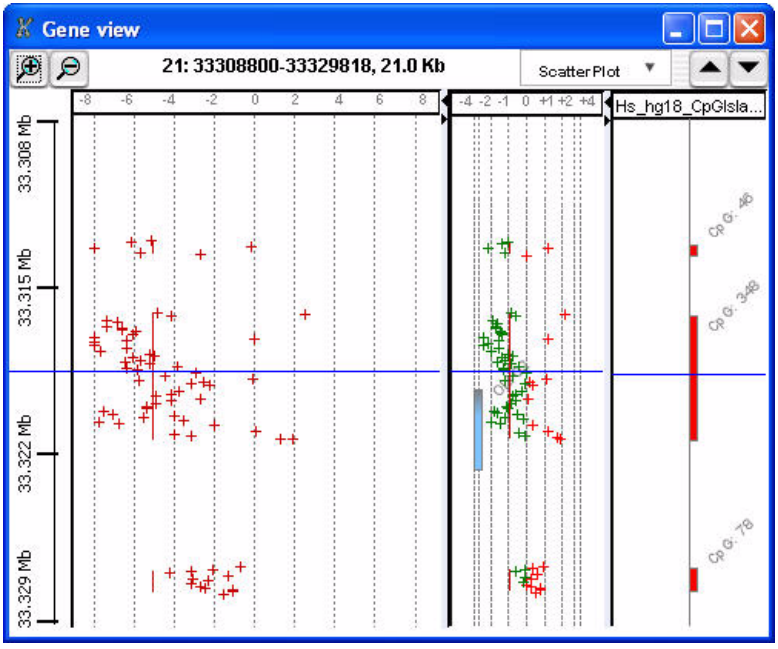


Figure 20 Gene View, showing moving averages within genomic boundaries of tracks

If the analysis results do not appear in any of the Views

Check these four potential problem areas, in order:

- Make sure you have marked the Apply check box. Under Analysis, mark **Apply** for Probe Methylation and/or Batman.
- Make sure you have selected the experiment or selected the arrays of interest. See "To select an experiment" on page 78.
- Make sure you have turned on the display for the moving averages or the Z-score algorithm. Do the following:
  1. Right-click any of the Views, and click **View Preferences**.
  2. Click the **View** tab.
  3. Under **Data Visibility,** in **View,** select **All views,** then mark any of the check boxes for the information to display.

You can also select a single View.

4 Click **OK.**

- Make sure you are looking in the correct region of Chromosome View.

See Chapter 3, "Displaying CH3 Data and Other Content" to find out how to display the Scatter Plot and how to search the View displays to see results.

To upload a track to UCSC

The program lets you upload methylation results to the UCSC Web site for further review.

1 Right-click in Gene View and select **Upload track to UCSC**.

The Upload Track to UCSC dialog box appears. See "Upload Track to UCSC" on page 228.

2 Complete the dialog box, (mark **Methylation Score** under Select Track Source) and click **OK**.

The track is uploaded to the UCSC Web site (Figure 21). Enable pop ups or set preferences on the UCSC Web site, if necessary.

3 Follow the instructions on the UCSC Web site to review your data.

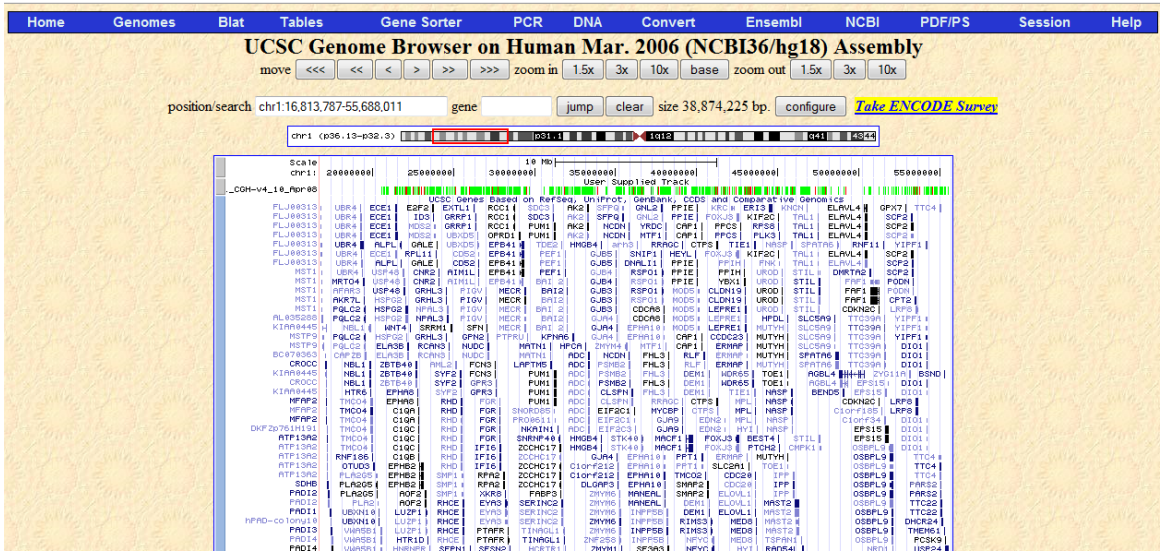


Figure 21 Uploaded track in UCSC browser

To save a result

The program lets you save the current result of the active experiment. You can run many different analyses in the same experiment, and save each one. Later, you can restore any of your saved results.

If you are saving a result for the first time for the experiment:

1 In the Experiment pane of the Navigator, right-click the experiment, and click **Save Experiment Result**, or

Click **Home > Save Experiment Result.**

A dialog box asks if you want to save the results of the current experiment.

2 Click **Yes**.

The Save experiment result dialog box appears.

3 Type a name for the result, then click **OK.**

If you have already saved at least one result for the experiment:

1 In the Experiment pane of the Navigator, expand the folders of the current experiment.

The currently selected result, if any, appears in blue in the Results folder.

2 Click **Home > Save Experiment Result** or

Right-click the experiment, and click **Save Experiment Result**.

A dialog box asks if you want to select one of the following actions:

- To replace the current result with another saved result, click **Overwrite Current Result.**
- To add the current results to the list of experimental results, click **Create New Result.**
- To change views to another result without changing the current result, click **Continue Without Saving.**

To restore a saved result to the display

1 If necessary, select the experiment that contains the result that you want to see. See "To select an experiment" on page 78.

2 In the Experiment pane of the Navigator, expand the folder of the active experiment, then expand its **Results** folder.

3 Right-click the desired result, then click **Restore result.**

The restored result appears in Genome, Chromosome, and Gene Views.

To generate a probe report

The probe report includes the following information either for single chromosomes or the complete genome:

- Log ratio for each event
- Number of probes
- Start and stop of each event

- Whether the probe is in a CpG island, a promoter, or inside a gene boundary
- Probable gene for the probe

1 Click **Reports**, then click **Generate Report**.

The Probe Methylation Report Setup dialog box appears.

2 To select if you want a report to include the results for the whole genome or for individual chromosomes, click either **Complete Genome** or **Per- Chromosome**.

3 Click **Browse** to find the folder for containing the report(s).

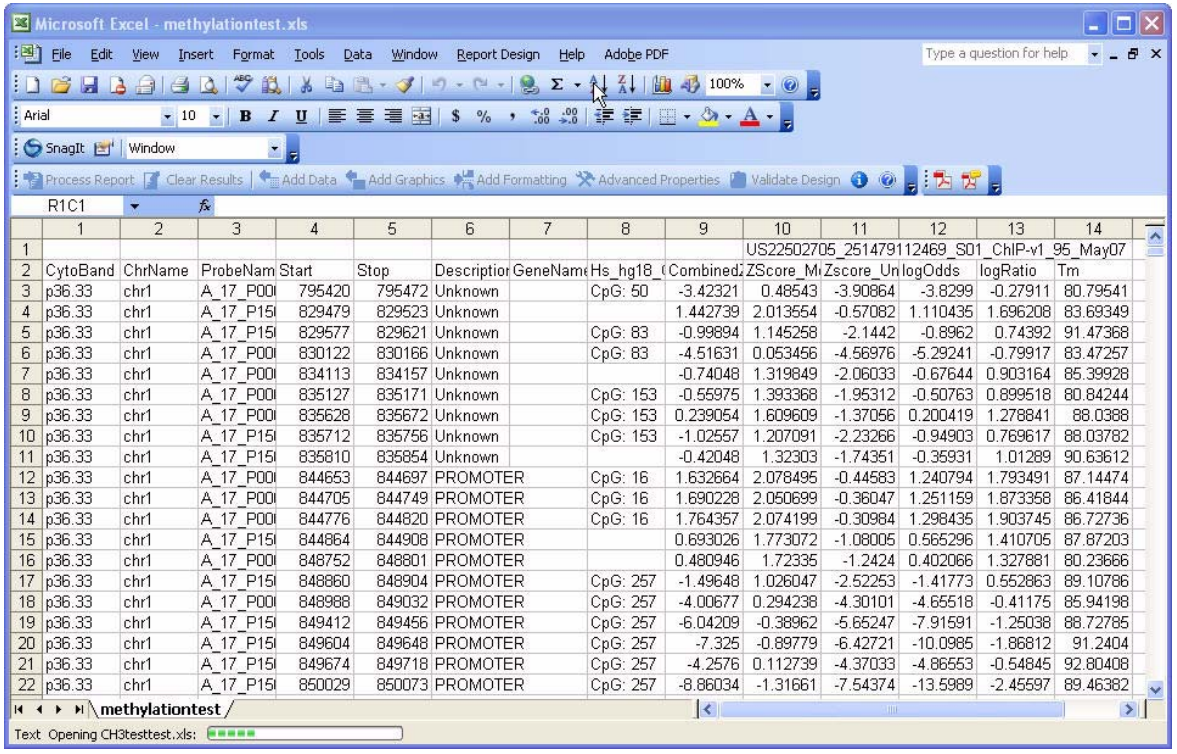
4 Type the report name, and click **Open**.

The report file name has the extension .xls.

5 Click **OK**.

6 When asked if you want to display the report, click **Yes**.

If you selected Complete Genome, an Excel spreadsheet appears containing all the data:



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1										US22502705_251479112469_S01_ChIP-v1_95_May07				
2	CytoBand	ChrName	ProbeNam	Start	Stop	Description	GeneName	Hs_hg18_(	Combined	ZScore_Mi	Zscore_Un	logOdds	logRatio	Tm
3	p36.33	chr1	A_17_P00	795420	795472	Unknown		CpG: 50	-3.42321	0.48543	-3.90864	-3.8299	-0.27911	80.79541
4	p36.33	chr1	A_17_P15	829479	829523	Unknown			1.442739	2.013554	-0.57082	1.110435	1.696208	83.69349
5	p36.33	chr1	A_17_P15	829577	829621	Unknown		CpG: 83	-0.99894	1.145258	-2.1442	-0.8962	0.74392	91.47368
6	p36.33	chr1	A_17_P00	830122	830166	Unknown		CpG: 83	-4.51631	0.053456	-4.56976	-5.29241	-0.79917	83.47257
7	p36.33	chr1	A_17_P00	834113	834157	Unknown			-0.74048	1.319849	-2.06033	-0.67644	0.903164	85.39928
8	p36.33	chr1	A_17_P00	835127	835171	Unknown		CpG: 153	-0.55975	1.393368	-1.95312	-0.50763	0.899518	80.84244
9	p36.33	chr1	A_17_P00	835628	835672	Unknown		CpG: 153	0.239054	1.609609	-1.37056	0.200419	1.278841	88.0388
10	p36.33	chr1	A_17_P15	835712	835756	Unknown		CpG: 153	-1.02557	1.207091	-2.23266	-0.94903	0.769617	88.03782
11	p36.33	chr1	A_17_P15	835810	835854	Unknown			-0.42048	1.32303	-1.74351	-0.35931	1.01289	90.63612
12	p36.33	chr1	A_17_P00	844653	844697	PROMOTER		CpG: 16	1.632664	2.078495	-0.44583	1.240794	1.793491	87.14474
13	p36.33	chr1	A_17_P00	844705	844749	PROMOTER		CpG: 16	1.690228	2.050699	-0.36047	1.251159	1.873358	86.41844
14	p36.33	chr1	A_17_P00	844776	844820	PROMOTER		CpG: 16	1.764357	2.074199	-0.30984	1.298435	1.903745	86.72736
15	p36.33	chr1	A_17_P15	844864	844908	PROMOTER			0.693026	1.773072	-1.08005	0.565296	1.410705	87.87203
16	p36.33	chr1	A_17_P00	848752	848801	PROMOTER			0.480946	1.72335	-1.2424	0.402066	1.327881	80.23666
17	p36.33	chr1	A_17_P15	848860	848904	PROMOTER		CpG: 257	-1.49648	1.026047	-2.52253	-1.41773	0.552863	89.10786
18	p36.33	chr1	A_17_P00	848988	849032	PROMOTER		CpG: 257	-4.00677	0.294238	-4.30101	-4.65518	-0.41175	85.94198
19	p36.33	chr1	A_17_P15	849412	849456	PROMOTER		CpG: 257	-6.04209	-0.38962	-5.65247	-7.91591	-1.25038	88.72785
20	p36.33	chr1	A_17_P15	849604	849648	PROMOTER		CpG: 257	-7.325	-0.89779	-6.42721	-10.0985	-1.86812	91.2404
21	p36.33	chr1	A_17_P15	849674	849718	PROMOTER		CpG: 257	-4.2576	0.112739	-4.37033	-4.86553	-0.54845	92.80408
22	p36.33	chr1	A_17_P15	850029	850073	PROMOTER		CpG: 257	-8.86034	-1.31661	-7.54374	-13.5989	-2.45597	89.46382

Figure 22 Excel methylation report for Complete Genome

If you selected Per-Chromosome, the text files for the individual chromosomes appear in the folder you selected.

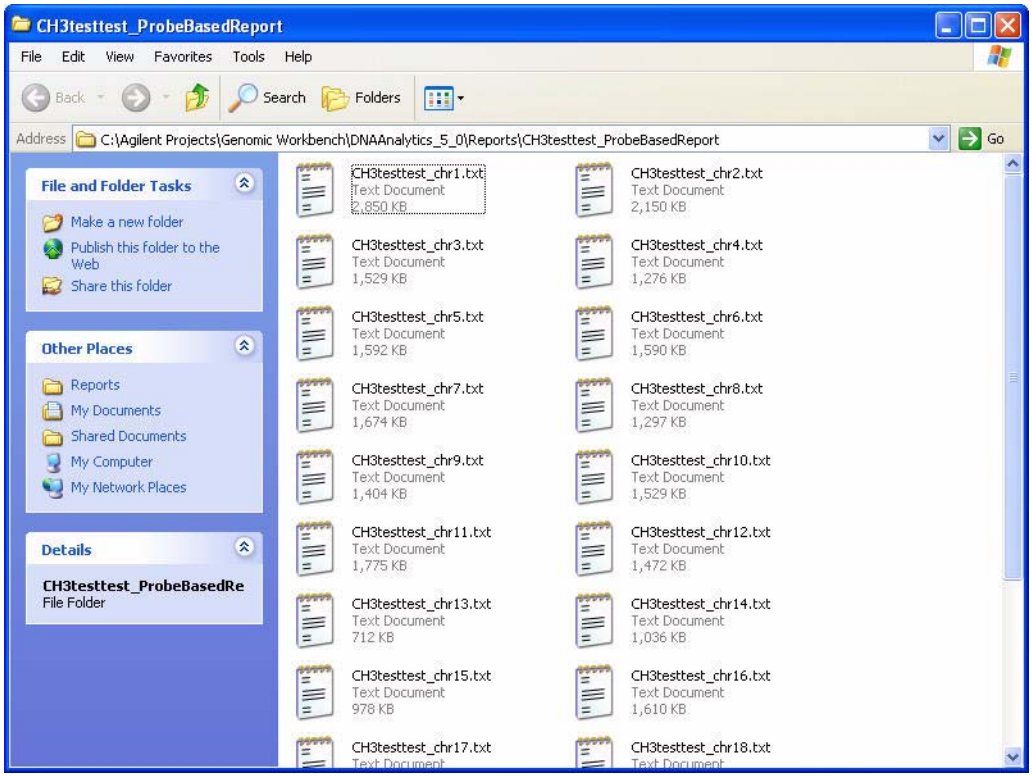


Figure 23 List of text methylation reports for each chromosome

To generate a Batman report

The Batman report includes the following information either for single chromosomes or for the entire genome:

- Cytoband name
- Chromosome name
- Probe name

- Start and Stop of each event

1 Click **Reports** and then select **Generate Batman Report.**

The Batman Report Dialog dialog box opens. See "Batman Report Dialog" on page 170.

2 To select if you want a report to include the results for the whole genome or for individual chromosomes, click either **Complete Genome** or **Per- Chromosome**.

3 Click **Browse** to find the folder for containing the report(s).

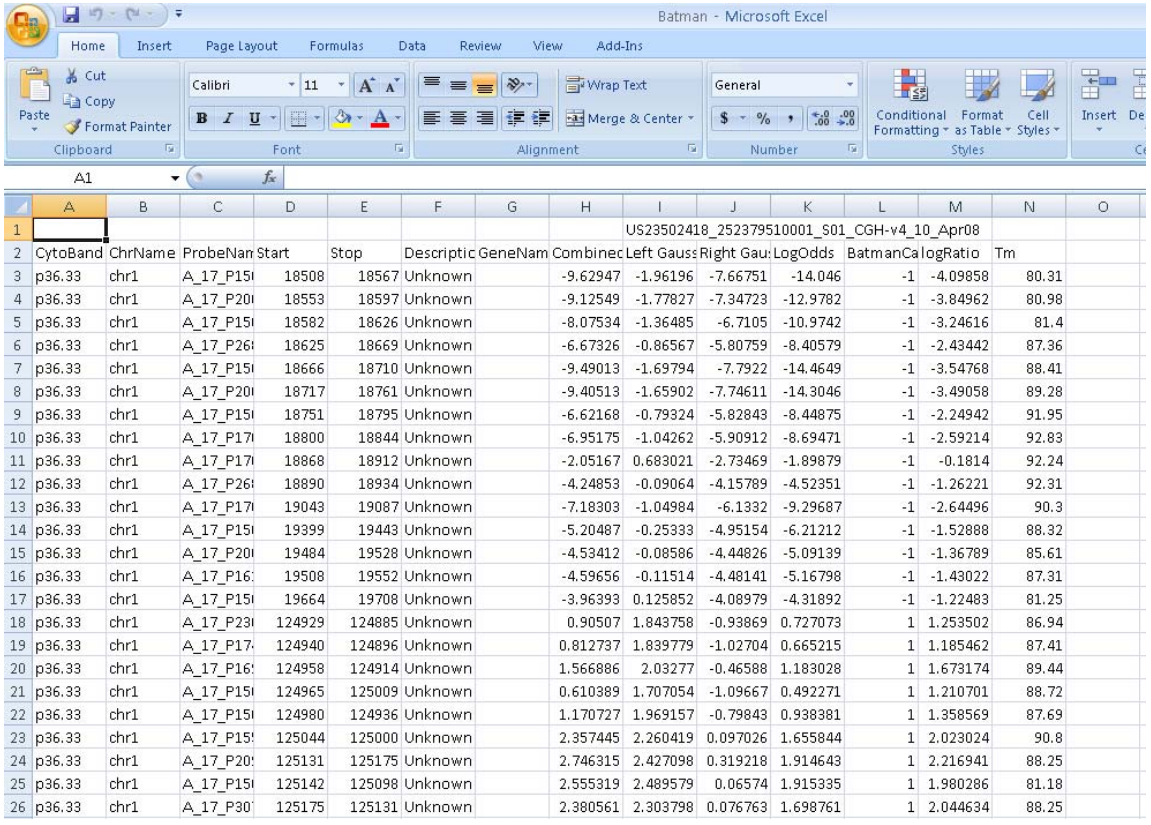
4 Type the report name, and click **Open**.

The report file name has the extension .xls.

5 Click **OK**.

6 When asked if you want to display the report, click **Yes**.

If you selected Complete Genome, an Excel spreadsheet appears containing all the data:



Batman - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1									US23502418_252379510001_S01_CGH-v4_10_Apr08						
2	CytoBand	ChrName	ProbeNam	Start	Stop	Descriptic	GeneNam	Combinec	Left Gauss	Right Gau	LogOdds	BatmanCa	logRatio	Tm	
3	p36.33	chr1	A_17_P15	18508	18567	Unknown		-9.62947	-1.96196	-7.66751	-14.046	-1	-4.09858	80.31	
4	p36.33	chr1	A_17_P20	18553	18597	Unknown		-9.12549	-1.77827	-7.34723	-12.9782	-1	-3.84962	80.98	
5	p36.33	chr1	A_17_P15	18582	18626	Unknown		-8.07534	-1.36485	-6.7105	-10.9742	-1	-3.24616	81.4	
6	p36.33	chr1	A_17_P26	18625	18669	Unknown		-6.67326	-0.86567	-5.80759	-8.40579	-1	-2.43442	87.36	
7	p36.33	chr1	A_17_P15	18666	18710	Unknown		-9.49013	-1.69794	-7.7922	-14.4649	-1	-3.54768	88.41	
8	p36.33	chr1	A_17_P20	18717	18761	Unknown		-9.40513	-1.65902	-7.74611	-14.3046	-1	-3.49058	89.28	
9	p36.33	chr1	A_17_P15	18751	18795	Unknown		-6.62168	-0.79324	-5.82843	-8.44875	-1	-2.24942	91.95	
10	p36.33	chr1	A_17_P17	18800	18844	Unknown		-6.95175	-1.04262	-5.90912	-8.69471	-1	-2.59214	92.83	
11	p36.33	chr1	A_17_P17	18868	18912	Unknown		-2.05167	0.683021	-2.73469	-1.89879	-1	-0.1814	92.24	
12	p36.33	chr1	A_17_P26	18890	18934	Unknown		-4.24853	-0.09064	-4.15789	-4.52351	-1	-1.26221	92.31	
13	p36.33	chr1	A_17_P17	19043	19087	Unknown		-7.18303	-1.04984	-6.1332	-9.29687	-1	-2.64496	90.3	
14	p36.33	chr1	A_17_P15	19399	19443	Unknown		-5.20487	-0.25333	-4.95154	-6.21212	-1	-1.52888	88.32	
15	p36.33	chr1	A_17_P20	19484	19528	Unknown		-4.53412	-0.08586	-4.44826	-5.09139	-1	-1.36789	85.61	
16	p36.33	chr1	A_17_P16	19508	19552	Unknown		-4.59656	-0.11514	-4.48141	-5.16798	-1	-1.43022	87.31	
17	p36.33	chr1	A_17_P15	19664	19708	Unknown		-3.96393	0.125852	-4.08979	-4.31892	-1	-1.22483	81.25	
18	p36.33	chr1	A_17_P23	124929	124885	Unknown		0.90507	1.843758	-0.93869	0.727073	1	1.253502	86.94	
19	p36.33	chr1	A_17_P17	124940	124896	Unknown		0.812737	1.839779	-1.02704	0.665215	1	1.185462	87.41	
20	p36.33	chr1	A_17_P16	124958	124914	Unknown		1.566886	2.03277	-0.46588	1.183028	1	1.673174	89.44	
21	p36.33	chr1	A_17_P15	124965	125009	Unknown		0.610389	1.707054	-1.09667	0.492271	1	1.210701	88.72	
22	p36.33	chr1	A_17_P15	124980	124936	Unknown		1.170727	1.969157	-0.79843	0.938381	1	1.358569	87.69	
23	p36.33	chr1	A_17_P15	125044	125000	Unknown		2.357445	2.260419	0.097026	1.655844	1	2.023024	90.8	
24	p36.33	chr1	A_17_P20	125131	125175	Unknown		2.746315	2.427098	0.319218	1.914643	1	2.216941	88.25	
25	p36.33	chr1	A_17_P15	125142	125098	Unknown		2.555319	2.489579	0.06574	1.915335	1	1.980286	81.18	
26	p36.33	chr1	A_17_P30	125175	125131	Unknown		2.380561	2.303798	0.076763	1.698761	1	2.044634	88.25	

Figure 24 Excel Batman report for complete genome

If you selected per-chromosome for the report output, the text files for the individual chromosomes appear in the folder you selected.

5 Methylation (CH3) Analysis Reference

This chapter describes the parts of the Agilent Genomic Workbench main window that you use to import, organize, manage, export and display array data and other content. It also describes the tab commands, shortcut menus, and dialog boxes that can appear.

Agilent Genomic Workbench CH3 Application Main Window

The main window of the CH3 Application contains the major components illustrated in Figure 25.

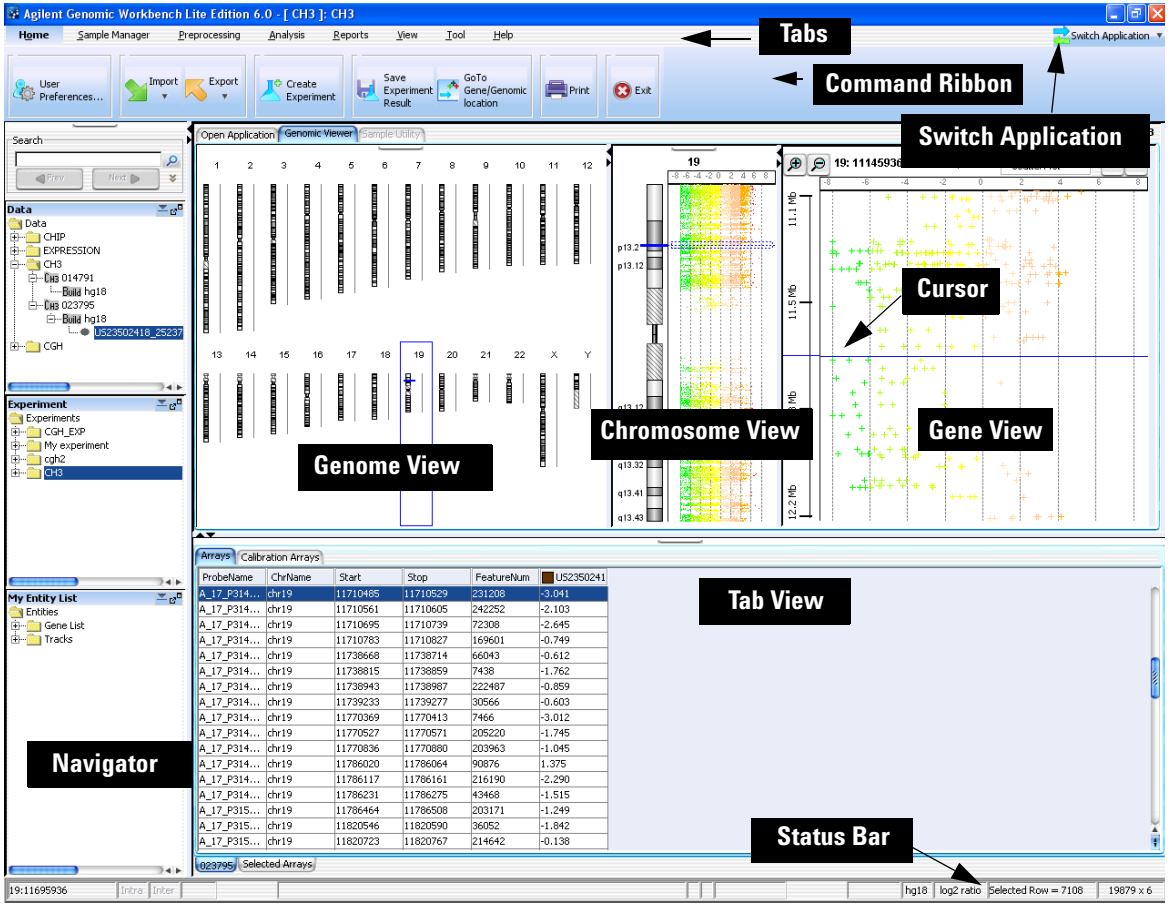


Figure 25 Agilent Genomic Workbench Lite Edition – CH3 application main window

Switch Application Menu

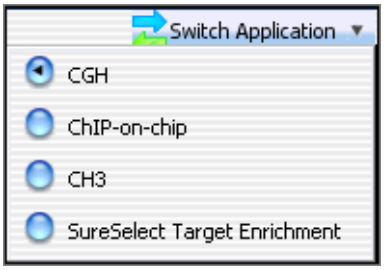


Figure 26 Switch Application menu

The Switch Application menu lets you change to the other data display and analysis application type in DNA Analytics. Mark the desired application type.

CGH (Separate license required) Imports, displays, and analyzes array-based comparative genomics hybridization (aCGH) data in both an interactive "analyze as you go" mode, and an automated workflow mode.

ChIP (Separate license required) Imports, displays, and analyzes ChIP-on-Chip microarray data in both an interactive "analyze as you go" mode, and an automated workflow mode.

CH3 (Separate license required) Imports and displays data from microarray-based studies of genomic methylation patterns.

SureSelect Target Enrichment Use the Quality Analyzer function for SureSelect Target Enrichment. See the *SureSelect Quality Analyzer User Guide* for more information.

Command Ribbons

When you click a tab, groups of commands or single commands appear at the top of the tab. This group of commands is called a command ribbon. The interactive tabs of the CH3 Application include Home, Preprocessing, Analysis, Reports, View, and Help (Figure 27). This section provides descriptions of the commands for each of these tabs, as well as for the Help tab. Descriptions of the remaining tabs are located in the following guides:

- Sample Manager — *Sample Manager User Guide*

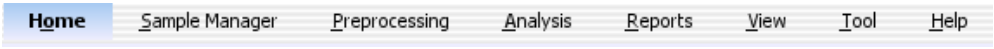


Figure 27 CH3 Application tabs

Home command ribbon

The Home command ribbon displays the functions that let you import, manage, export and display CH3 data and content for further CH3 analysis.

Figure 28 Command ribbon in the Home tab of Agilent Genomic Workbench

User Preferences Opens the User Preferences dialog box with the following tabs:

Tab	Description
Tracks	Opens a dialog box that lets you manage which tracks to display in Genomic Viewer and how they appear. See "User Preferences" on page 230.

Tab	Description
Miscellaneous	Opens a dialog box where you can select a new location for your data files and set up access to the eArray web site. See "User Preferences" on page 230.
License	Opens a dialog box where you can enter an Agilent Genomic Workbench application license, if you purchase one after using the unlicensed version. See "User Preferences" on page 230.

Import Opens a menu of file types that you can import:

Option	Description
Array Files	Opens a menu with these options: • **FE File** – Opens the Import FE Files dialog box, where you can select an Agilent Feature Extraction array data file to import. See "Import" on page 200 and "To import Agilent FE or Axon data files" on page 41. • **Axon File** – Opens the Import Axon Files dialog box, where you can select Axon (*.gpr) files for import. See "Import" on page 200 and "To import Agilent FE or Axon data files" on page 41. • **UDF File** – Opens the UDF Files dialog box, where you can select a Universal Data File (UDF) to import. See "Import" on page 200 and "To import a UDF file" on page 42.
Design Files	Opens a menu with these options: • **GEML File** – Opens the Import Design Files dialog box, where you can select Agilent GEML-based (*.xml) array design files for import. See "Import" on page 200 and "To import Agilent GEML design files" on page 39. • **Axon Design File** – Opens the Import Axon Design Files dialog box, where you can select Axon (*.gal) array design files for import. See "Import" on page 200 and "To import Axon design files" on page 40.
Genome Build	Opens the Import Genome Build dialog box, where you can import Agilent-supplied genome build files. See "Import Genome Build" on page 204 and "To import a genome build" on page 45.
Array Attributes	Opens the Import microarray attributes dialog box, where you can select a microarray attributes file. See "Import" on page 200.

Option	Description
Track	Opens the Import Track dialog box, where you can select a BED format track file for import, and create a display name for the track. See "Import Track" on page 205 and "To import tracks" on page 46.
Experiments	Opens the Import Experiments dialog box, where you can select a ZIP format experiment file for import. See "Import (experiments)" on page 202 and "To import an experiment file" on page 48.

Export Opens a menu that lets you export several kinds of files.

Option	Description
Experiments	Opens the Export Experiments dialog box, where you can select one or more experiments for export as a single ZIP file. See "Export Experiments" on page 194 and "To export experiments" on page 72.
Tracks	Opens the Export Tracks dialog box, where you can select one or more tracks to export as a single BED format file. See "Export Tracks" on page 195 and "To export tracks" on page 73.
Array Attributes	Opens the Export Array Attributes dialog box, where you can select arrays and their attributes for export. See "Export Array Attributes" on page 190.

Create Experiment Opens the Create Experiment dialog box, where you can create a new, empty experiment and add data to it. See "Create Experiment" on page 177 and "To create a new experiment" on page 51.

Save Experiment Results Opens a confirm dialog box. Click **Yes** and the Save experiment result dialog box opens where you enter a name and description for the results to save.

Go to Gene/Genomic Location Moves the cursor to the location in Chromosome and Gene Views that you select. See "Go To Gene/Genomic Location" on page 199.

Print Opens the Print window to print the display.

Exit Closes the program.

Preprocessing command ribbon

The Preprocessing ribbon has a single command you use to prepare for CH3 analysis: Fuse.

Figure 29 CH3 Preprocessing command

The Fuse command lets you combine multiple array designs into a larger virtual combined design. It also lets you combine both intra- and inter-array replicate probes.

Click to open the Array Set dialog box, where you can select the names of arrays that are combined, set options, and combine designs together to form a larger virtual design. See "Array Set" on page 168, and "To combine (fuse) arrays" on page 105.

Analysis command ribbon

The Analysis ribbon contains the commands you use for CH3 analysis.

Figure 30 CH3 Analysis ribbon

Moving Average (Log Ratio)

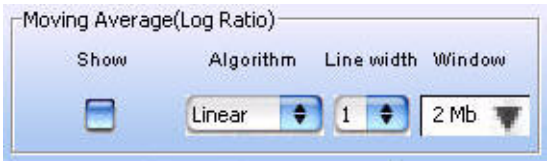


Figure 31 Moving Average (Log Ratio) command

The settings under Moving Average(Log Ratio) control the calculation and display of line plots that represent smoothed log ratio data for each selected array. These plots can appear in Genome, Chromosome, and Gene views.

Show Mark this check box to show moving average line plots, or clear the check box to hide them. You turn the display of moving average line plots for specific views on or off from the **View** tab by selecting **View Preferences**.

In View Preferences, under Data Visibility, select the desired view, then mark or clear **Moving Average.** See "View Preferences" on page 235.

Algorithm Select one of these options:

- **Linear** – The linear algorithm calculates a standard, unweighted average using every Log Ratio score within a defined window size (specified by either a number of adjacent measurements or a positional range). This average is applied sequentially to every point by moving the window along the chromosome.

- **Triangular** – The triangular algorithm calculates a weighted average using every Log Ratio within a specified window size. The triangular algorithm is more sensitive to localized variations in the data. See "Overview of Methylation Detection and Visualization Algorithms" on page 242.

Line width Select the desired thickness (in pixels) for the moving average line plots. You can select a thickness from 1 to 5 pixels.

Window Select the desired size of the moving average window. You can select either a specific number of base pairs (5 Kb to 50 Mb), or a specific number of data points (1 pt to 60 pt). You can also type a value. The program calculates a moving average for each selected array based on a window of the given size centered on each point in the array.

Moving Average (ZScore)

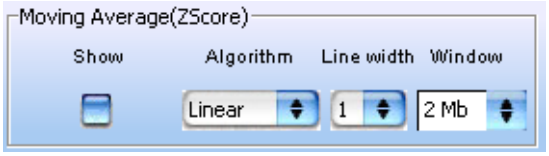


Figure 32 Moving Average (ZScore) command

The settings under Moving Average (Z-Score) control the calculation and display of line plots that represent smoothed Z-score data for each selected array. These plots can appear in Genome, Chromosome, and Gene views. You must apply the Probe Methylation algorithm first.

Show Mark this check box to show moving average line plots for the Z-score data, or clear the check box to hide them. You enable or disable the display of moving average line plots for selected views in View Preferences. See "View Preferences" on page 235.

Algorithm Select one of these options:

- **Linear** – The linear algorithm calculates a standard, unweighted average using every combined Z-score within a selected window size (specified by either a number of adjacent measurements or a positional range). This average is applied sequentially to every point by moving the window along the chromosome.

- **Triangular** – The triangular algorithm calculates a weighted average using every Z-score within a selected window size. The triangular algorithm is more sensitive to localized variations in the data. See "Triangular smoothing" on page 260.

Line width Select the thickness (in pixels) for the moving average line plots. You can select a thickness from 1 to 5 pixels.

Window Select the size of the moving average window. You can select either a number of base pairs (5 Kb to 50 Mb), or a number of data points (1 pt to 60 pt). You can also type a value. The program calculates a moving average for each defined array based on a window of the given size, centered on each point in the array.

Probe Methylation

You apply this command to make Z-score calculations to find out whether a Methylation detection array probe is methylated or unmethylated. See "Overview of Methylation Detection and Visualization Algorithms" on page 242.

Figure 33 Probe Methylation command

Select ZScore Select the algorithm to use for calculation of probe methylation score.

Apply Click Apply to display the Probe Methylation Status Setup dialog box to assign a melting temperature (Tm) mapping file, if necessary, to the experiment. See "Probe Methylation Status Setup" on page 210.

Batman

Figure 34 Batman command

Click **Apply** to display the Batman Parameters Setup dialog box where you set the parameters for the Batman tool for calculation and display of methylation data. See "Batman Parameter Setup" on page 169 for more information.

Reports command ribbon

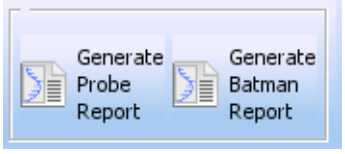


Figure 35 CH3 Report command

The Reports command ribbon lets you save all the important information about the Z-score statistics and CpG Island tracks that can help you find out whether regions of the genome are methylated or not. You use a spreadsheet program or your Internet browser to open the report files the program creates. This lets you further display, analyze, and organize the result.

Generate Probe Report Opens the Probe Methylation Report Dialog dialog box, where you can enter a location and a name for the probe report file and select whether you want individual reports for each chromosome or one report on the complete genome. See "Probe Methylation Report Dialog" on page 211.

The Methylation Report makes information about the results of the probe methylation algorithm available in .xml format (complete genome) or in .txt format (individual chromosomes). You can use a spreadsheet program to open all these files. See "Report Format" on page 239 for a description of the columns in the report.

Generate Batman Report Opens the Batman Report Dialog box where you set the parameters and select a location to save the report.

View command ribbon

The View command ribbon lets you change the display of data/results in Genomic Viewer.

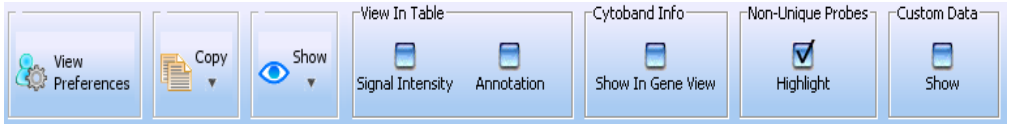


Figure 36 View command ribbon

View Preferences Opens the View Preferences dialog box where you customize the display of data and results in the Genomic Viewer. For more information, see "View Preferences" on page 235.

Copy This command opens a menu with the options listed below. In general, the Copy command copies pane(s) of the main window to the Clipboard as an image. You can then paste the image into a document in another program. See "To copy what you see in the main window" on page 74.

Option	Description
All	Copies all panes of the main window to the Clipboard as an image.
Navigator	Copies only the Navigator to the Clipboard as an image.
Tab View	Copies only Tab View to the Clipboard as an image.
Genome view	Copies only Genome View to the Clipboard as an image.

Option	Description
Chromosome view	Copies only Chromosome View to the Clipboard as an image.
Gene view	Copies only Gene View to the Clipboard as an image.

Show Opens a menu with all available elements of the main window. Mark the check box for the ones to display in Genomic Viewer.

View In Table

Signal Intensity Mark the check box to see the red and green raw signal intensities of the log ratio data in the Tab View.

Annotation Mark the check box to show annotations in the Tab View.

Cyto band info

View In Gene View Mark the check box to display cytobands in the Gene View.

NonUnique Probes

Highlight Mark the check box to display non unique probes in a different color.

Custom Data

Show Mark the check box to display custom data in the Genomic Viewer.

Help command ribbon

The Help command ribbon lets you display the available Agilent Genomic Workbench help guides, and get information about software version, installation history, and check for software updates. Help guides are opened in Adobe Reader.

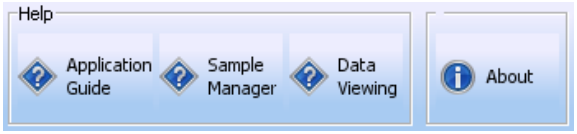


Figure 37 Help ribbon for CH3 application

Table 3 Table of Agilent Genomic Workbench Help

Help Command	Action
Application Guide	Opens the Agilent Genomic Workbench application user guide for the selected application.
Sample Manager	Opens the *Sample Manager User Guide,* that shows how to use the Sample Manager module of Agilent Genomic Workbench to organize microarrays and edit their attributes.
Data Viewing	Opens the *Data Viewing User Guide* that describes how to import, organize, manage, export and display data and other content (experiments, gene lists, tracks) within Agilent Genomic Workbench. It is targeted for users who have no DNA Analytics application license(s).
About	Opens a message with information about the version number and copyright of the program.

Navigator

This section describes the parts of the Navigator, and the shortcut menus and other functionality available within it.

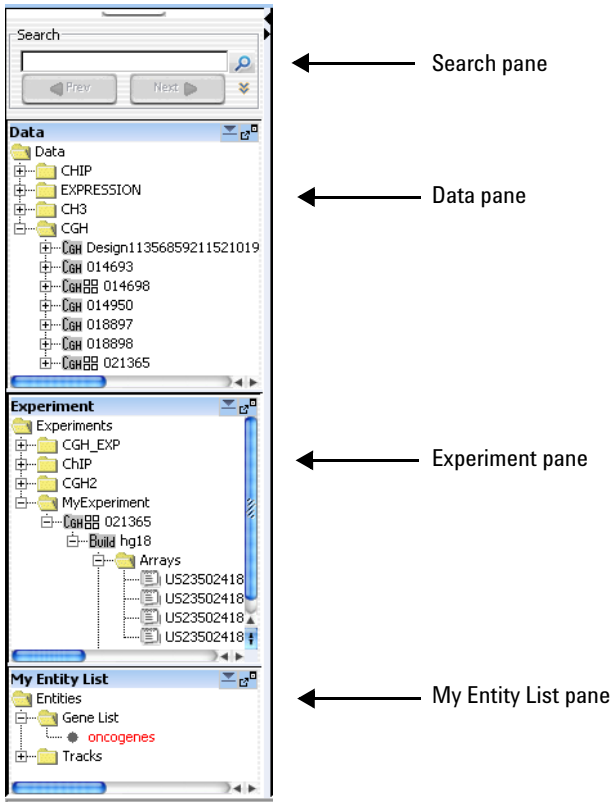


Figure 38 Navigator panes

The Navigator (Figure 38) shows the array data, experiments, and other content stored in Agilent Genomic Workbench. It contains the following panes:

Pane	Comments
Search	Lets you search within any pane of the Navigator for a specific item (array or build, for example). You must type the entire array name or term; otherwise, use asterisks (*) as wildcards for unspecified strings. For example, type *1234* to find any item that contains "1234". See "Search pane" on page 137.
Data	Contains microarray data files, organized by application type, then by design and genome build. Shows all probe groups and microarray designs that are available to you, organized by folders.In general, you can: • Expand or collapse folders to show or hide content • Right-click the name of a folder or item to open a shortcut menu that lets you take action on the item. See "Data pane – icons, special text, and buttons" on page 139 and "Data pane – actions and shortcut menus" on page 139.
Experiment	Contains Agilent Genomic Workbench experiments. Experiments are organizational units that contain links to microarray data and design files. In data analysis modules, experiments also contain saved results.
My Entity List	Contains gene lists and tracks: • **Gene Lists** are collections of genes of interest. You can create them within the program, import and export them, and apply them to Gene View and Chromosome View. • **Tracks** are collections of annotation or other information that map to specific genomic locations. You can import, export, and combine tracks, and display them in Gene View with your array data and analysis results. • See "My Entity List pane – Icons, buttons, and special text" on page 148.

Search pane

The Search pane lets you find all occurrences of a specific search term in the Data, Experiment, and/or My Entity List panes. See "To find specific content items in the Navigator" on page 62. It also contains several buttons that you can use to move, hide, show or resize the Navigator.

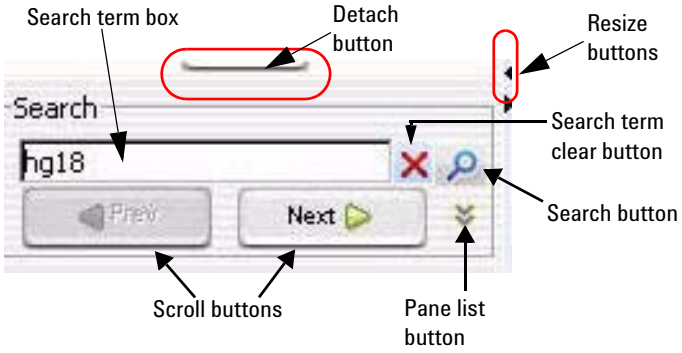


Figure 39 Navigator – Search pane

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 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, type *12345* to search for any content that contains the string "12345".

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.

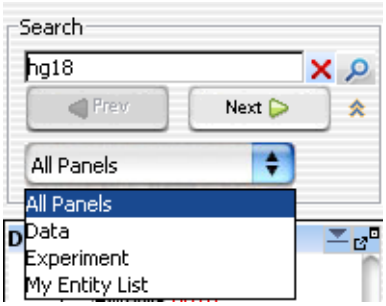


Figure 40 Search pane selection list

(Show Pane List button, available only if the Pane list is not visible) Makes the Pane list visible.

(Hide Pane List button, available only if the Pane list is visible) Hides the Pane list.

(Search button) Searches the pane(s) selected in the Pane list for all occurrences of the term you typed in the Search term box. If the program finds a matching item, it expands the folder structure to make the matching item(s) visible, makes the lettering of each item red and highlights the item in yellow. Note: The search term is not case-sensitive, but it must contain the entire name of the desired items.

Scroll buttons (Available only after a search) Lets you scroll up and down the lists of highlighted search items after a search.

(Clear button, available only after a search) Clears the search term from the Search term box, and resets the color of any matching item to its original color.

Data pane – icons, special text, and buttons

Item	Comments
	An unexpanded folder (domain) that contains subfolders or other items.
	An expanded folder. The items that it contains are visible in the Navigator.
	Expands a folder to show its contents.
	Collapses a folder to hide its contents.
	A methylation array design. This folder contains array data for the design, organized by genome build.
	A CGH array design. This folder contains array data for the design, organized by genome build.
	A gene expression array design. This folder contains array data for the design, organized by genome build.
	A ChIP array design. This folder contains array data for the design, organized by genome build.
	A genome build folder within a specific design folder. This folder contains arrays for the specific genome build and design.
	A single array data file.
	Data created from a multi-pack array.
red text	An item that matches the search term in a search.
	(Dock out button) Moves the Data pane from the Navigator, and opens it in a, separate window.
	(Collapse button, available only if the Data pane is not collapsed) Collapses the Data pane, and shows its title bar at the bottom of the Navigator.
	(Expand button, available only if the Data pane is collapsed) Expands the Data pane.

Data pane – actions and shortcut menus

The Data pane of the Navigator shows available content items that are stored on your server for the selected application type, and any external content that you imported.

- Double-click any folder to expand or collapse it.

Data Folder

- Double-click a data type folder (CGH, ChIP, Expression, or CH3) to expand it and display its contents.
- Double-click a design folder to display the genome builds for the design.
- Double-click a build folder to display the imported arrays for that build.

Design Folder

- Right-click the name of a design folder to display these options:

Option	Description
Update from eArray	Updates the annotations for your array design from the eArray Web site. Agilent regularly updates annotations in eArray as new ones become available. See "To update probe annotation in design files" on page 63. **Note:** In order to use this function, you must enter your eArray Username and Password in the Miscellaneous tab of the User Preferences dialog box. See "User Preferences" on page 230.
Delete	Opens a Confirm dialog box. If you click **Yes,** the program permanently deletes the design and all arrays associated with it.

Genome Build Folder

- Right-click the name of a genome build folder to display these options:

Option	Description
Show Properties	Opens the Design Properties dialog box. See "Design Properties" on page 182.
Delete	Opens a Confirm dialog box. If you click **Yes,** the program permanently deletes all of the arrays in this genome build folder.

Individual Arrays

- Right-click the name of an array to display these options:

Option	Description
Show Properties	Opens the Microarray Properties dialog box. See"Microarray Properties" on page 207 and "To display or edit the attribute values of a specific array" on page 59.
Rename	Opens an Input dialog box, where you can type a new name for the array. Click **OK** to rename the array.
Delete	Opens a Confirm dialog box. If you click **Yes,** the program permanently deletes the array.

- Drag an array from the Data pane to an experiment folder in the Experiment pane to associate it with an experiment. You can drag multiple arrays at once from one genome build in a design. Hold down the **ctrl** key and click the additional arrays to select them. You can also select a contiguous block of arrays; click the first array in the block, then hold down the **Shift** key and click the last one.

Experiment pane – icons, special text, and buttons

Item	Comments
	Click to expand a folder and display its contents.
	Click to collapse a folder and hide its contents.
	A folder that contains files or other folders.
	A methylation array design. This folder contains array data for the design, organized by genome build.
	A CGH array design. This folder contains array data for the design, organized by genome build.
	A gene expression array design. This folder contains array data for the design, organized by genome build.
	A ChIP array design. This folder contains array data for the design, organized by genome build.
	A genome build folder within a specific design folder. This folder contains arrays for the specific genome build and design.
	An array that is not selected for view and analysis.
	An array that is selected for view and analysis. The specific color of this icon can vary.
	A calibration array.
	An empty folder.
	Data created from a multi-pack array.
blue text	The active experiment. All data and results that appear in Chromosome, Gene, and Tab Views are from this experiment.
red text	An item that matches the search term in a search.
	(Dock out button) Moves the Experiment pane from the main window, and opens it in a, separate window.
	(Collapse button, available only if the Experiment pane is not collapsed) Collapses the Experiment pane, and shows its title bar at the bottom of the Navigator.
	(Expand button, available only if the Experiment pane is collapsed) Expands the Experiment pane.

Experiment pane — actions and shortcut menus

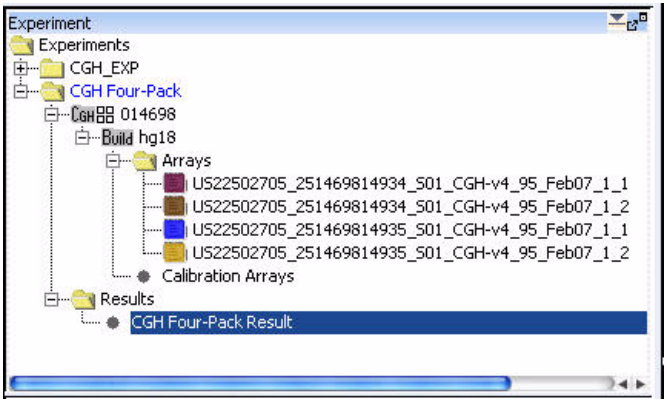


Figure 41 Expanded Experiment Pane

You can use many items in the Experiment Pane of the Navigator to open shortcut menus or take other actions.

- In general, double-click the **Experiments** folder within the Experiment Pane, and the folders within it, to expand and collapse them. Exception: Double-click the name of an unselected experiment to select it for analysis. This action opens the Experiment Selection dialog box. To select the experiment for analysis, click **Yes.**

Experiments Folder

- Right-click the **Experiments** folder to display these options:

Option	Description
New Experiment	Opens the Create Experiment dialog box (see "Create Experiment" on page 177), where you can name the new experiment, and open another dialog box that lets you add microarray data to the experiment. See "To create a new experiment" on page 51.
Export	Opens the Export Experiments dialog box, where you can export one or more experiments as a single ZIP file. See "Export Experiments" on page 194 and "To export experiments" on page 72.

Specific Experiment Folder

- In the **Experiments** folder, right-click the name of a specific experiment folder to display these options:

Option	Description
Select Experiment	(Appears only if the experiment is not selected.) Opens the Experiment Selection dialog box, which asks if you want to select the experiment. Click **Yes** to select the experiment for display and analysis. Or, in the Experiments folder, double-click the name of an experiment that is not selected to open the Experiment Selection dialog box. To select the experiment for analysis, click **Yes.** If you switch experiments, a Confirm dialog box asks if you want to save the current result. Select one of these options: • **Overwrite Current Result** – Replaces the selected experiment result in the Data Navigator with the result that appears on your screen. • **Create New Result** – Opens the Save experiment result dialog box, where you can save the result that appears in the main window as a new experiment result. See "To save a result" on page 114. • **Continue Without Saving** – The program does not save the result that appears on your screen. In some cases, the Confirm dialog box offers only Yes and No choices. If you click **Yes,** the Save experiment result dialog box appears, where you can save the result that appears on your screen with the name of your choice.
Deselect Experiment	(Appears only if the experiment is selected.) If the results are unsaved, a Confirm dialog box opens with these options: • **Overwrite Current Result** – Replaces the selected experiment result in the Data Navigator with the result that appears on your screen. • **Create New Result** – Opens the Save Experiment result dialog box, where you can save the result that appears on your screen as a new experiment result. • **Continue Without Saving** – The program does not save the result that appears on your screen. In all three cases, the program then removes the experiment data and results from all views.

Option	Description
Show Properties	Opens the Experiment Properties dialog box. Use this dialog box to see the names of the arrays in the experiment, and also to add or remove arrays from the experiment. See "Experiment Properties" on page 187.
Sample Attributes	Opens the Sample Attributes dialog box, where you can change the values for the attributes assigned to the arrays in the experiment. See "Sample Attributes" on page 212. To add new attributes you must use the Sample Manager. See *Sample Manager User Guide*.
Export	Opens the Export Experiments dialog box, where you can export this and other experiments as a single ZIP file. See "Export Experiments" on page 194, and "To export experiments" on page 72.
Export Attributes	Opens the Export Array Attributes dialog box, where you can save a file that contains selected attributes of the arrays in your experiment. See "Export Array Attributes" on page 190.
Edit Array Color	Opens the Edit Array Color dialog box, where you can select a display color for each of the arrays in the experiment. See "Edit Array Color" on page 185.
Edit Array Order	Opens the Edit Array Order dialog box, where you can change the order of the arrays in the experiment pane of the Navigator, and in Chromosome, Gene, and Tab Views. See "Edit Array Order" on page 186.
Rename	Opens an Input dialog box, where you can type a new name for the experiment. Click **OK** to rename the experiment.
Delete	Opens a Confirm dialog box that asks if you want to delete the Experiment. Click **Yes** to delete it. Note: You can delete any experiment except the selected one.

Design Folder

- In the folder of a specific experiment, right-click the name of a design to open a shortcut menu with a Delete command. If you click **Delete,** a Confirm dialog box opens. Click **Yes** to disassociate all of the arrays under the design from the experiment.

Genome Build Folder

- In the folder of a specific experiment, in a specific design folder, right-click the name of a genome build to display these options:

Option	Description
Set for Calibration	Use of calibration arrays is not recommended by Agilent. This feature has been deprecated.
Save As Text File	Opens the Save Design dialog box, where you can save all of the data for the genome build and design within the experiment as a tab-delimited text file.
Delete	Opens a Confirm dialog box that asks if you want to disassociate all arrays under the design from the experiment. Click **Yes** to remove the links between the arrays and the experiment. Note: • If you delete a design from an experiment, the program removes the links between the experiment and the design and its arrays. The actual design and array data stay in the Data folder. • Saved results become unavailable if they involve arrays you delete with this command.

Individual Arrays

- In a specific experiment, right-click the name of an individual array to display these options:

Option	Description
Select	(Available only if the array is not already selected) Selects the array for display and analysis.
Deselect	(Available only if the array is selected) Removes the array data from Genome, Chromosome, and Gene views, and excludes it from the analysis. Also removes the array data from the Selected Arrays tab in Tab View.
Select for Calibration	Use of calibration arrays is not recommended by Agilent. This feature has been deprecated.
Deselect from Calibration	Use of calibration arrays is not recommended by Agilent. This feature has been deprecated.

Option	Description
Rename	Opens an Input dialog box, where you can type a new name for the array in this experiment. Click **OK** to accept the new name for the array. The array name is changed only for the selected experiment.
Delete	Opens a Confirm dialog box that asks if you want to disassociate the array from the experiment. Click **Yes** to remove the link between the array and the experiment. **Note:** • If you delete an array from an experiment, the program removes the link between the experiment and the array. The actual array data remains in the Data folder. • You cannot restore an experiment result that includes a deleted array.
Show Properties	Opens the Microarray Properties dialog box, where you can view and edit microarray attributes. See "Microarray Properties" on page 207. For array files from the Agilent Feature Extraction program, you can also view the headers and feature data from the file. See "To display or edit the attribute values of a specific array" on page 59.
Edit Array Color	Opens the Select Color dialog box, where you can select a display color for the array. See "Select Color" on page 215.
Edit Array Order	Opens the Array Order dialog box, where you can change the order of the arrays in the Experiment pane of the Navigator, and in Chromosome, Gene, and Tab Views. See "Edit Array Order" on page 186.

My Entity List pane – Icons, buttons, and special text

Item	Comments
	Click to expand a folder and display its contents.
	Click to collapse a folder and hide its contents.
	A folder that contains files or other folders.
	An individual gene list or track.
red regular text	An item that is an exact match with the search term in a search, or a gene list that has not been applied or that has red chosen as its custom color.
colored italics	A gene list that has been applied.
red bold italics	A track that is selected for display in Gene View.
black bold italics	A "combined" track that is selected for display in Gene View. A combined track contains information from two or more individual tracks associated by logical criteria.
	(Dock out button) Moves the My Entity List pane from the main window, and opens it in a, separate window.
	(Collapse button, available only if the My Entity List pane is not collapsed) Collapses the My Entity List pane, and shows its title bar at the bottom of the Navigator.
	(Expand button, available only if the My Entity List pane is collapsed) Expands the My Entity List pane.

My Entity List pane – Actions and shortcut menus

Gene List folder

- Right-click the **Gene List** folder to open a shortcut menu with an **Import Gene List** option. This command opens an Import dialog box that you can use to import a gene list into the program. See "Import" on page 200.
- Double-click the **Gene List** folder to show or hide its gene lists.

- In the **Gene List** folder, right-click the name of a gene list to display these options:

Option	Description
View in Table	Opens the Gene List dialog box, where you can view the list of genes. You can also edit the description of the gene list, and change the display color of the genes. See "Gene List" on page 198 and "To display the genes in a gene list" on page 66.
Rename	Opens an Input dialog box, where you can type a new name for the gene list. Click **OK** to accept the new name.
Delete	Opens a confirm dialog box that asks if you are sure you want to delete the gene list. Click **Yes** to confirm.
Save As	Opens a Save As dialog box, where you can save the gene list as a text (*.txt) file. See "To export a gene list" on page 73.
Add to gene list	Opens the Add gene list dialog box, where you can add the gene list to any other one in the Gene List folder. See "Add Gene List <name> to" on page 165 and "To add one gene list to another" on page 67.
Highlight	(Available if the gene list is not selected.) Displays all genes in Gene View, and highlights the genes from the gene list in their display color. See "To show gene lists in Gene View" on page 90.
Show only	(Available only if all genes appear in Gene View, or if the gene list is not selected) Restricts the genes in Gene View to those on the gene list. No other genes appear. The program displays the genes in their display color. See "To show gene lists in Gene View" on page 90.
Show All	(Available only for the selected gene list.) In Gene View, displays all genes, without highlighting. See "To show gene lists in Gene View" on page 90.

Tracks folder

- Right-click the Tracks folder to display these options:

Option	Comments
Import Tracks	Opens the Import Track dialog box, where you can import a BED format track file into the program. See "Import Track" on page 205 and "To import tracks" on page 46.
Export Tracks	Opens the Export Tracks dialog box, where you can select tracks for export as a single BED format track file. See "Export Tracks" on page 195 and "To export tracks" on page 73.
Combine Tracks	Opens the Combine Tracks dialog box, where you can associate two or more individual tracks by logical criteria to create a new combined track. See "Combine Tracks" on page 171 and "To combine tracks" on page 68.

Track name

- Right-click the name of a track to display these options:

Option	Comments
Show in UI	Mark this option to display the track in Gene View next to the data and results of the selected experiment. See "To show tracks in Gene View" on page 92.
Show in Report	Mark this option to display the track in the reports.
Genomic Boundaries	Click to analyze data invoking the genomic boundaries for only that track. You can choose to do this for only one track.
Show in UCSC	Opens the UCSC Genome Browser in your Web browser and uploads the track. You can then view the track.
View Details	Opens a dialog box that displays information about the track. See "Track" on page 222.
Rename	Opens an Input dialog box, where you can type a new name for the track. Click **OK** to rename the track.
Delete	Opens a Delete Track dialog box that asks if you are sure you want to delete the track. Click **Yes** to delete the track.

Genomic Viewer

Genomic Viewer is the display for the DNA Analytics applications (CGH, ChIP and CH3). It includes the three Views – Genome, Chromosome and Gene Views – the Tab View and the View Cursor.

Genome View

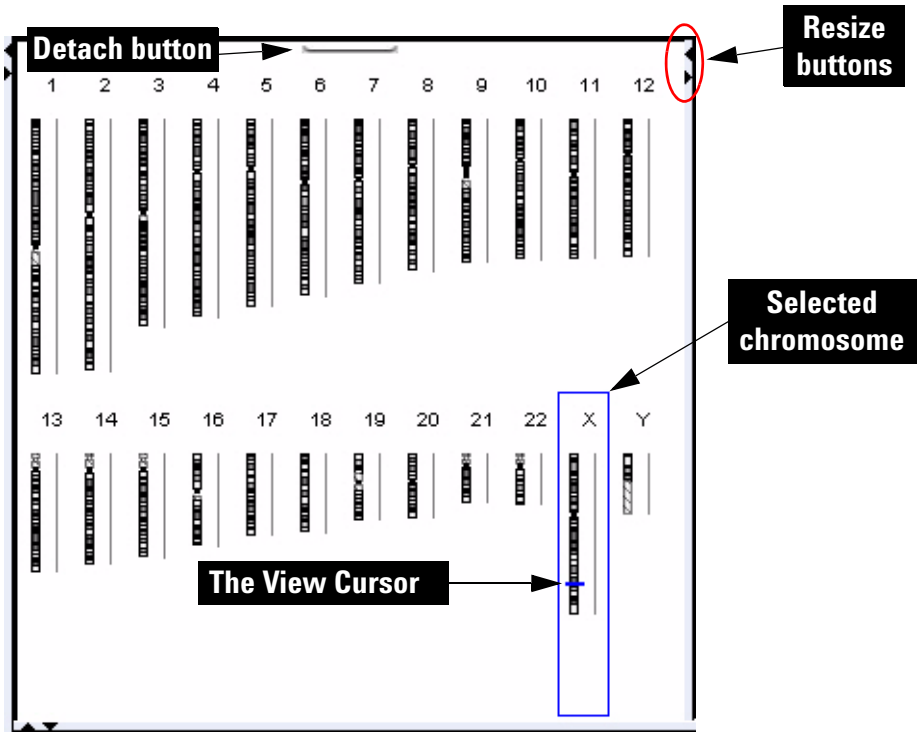


Figure 42 Genome View (vertical orientation), with human chromosomes. The X chromosome is selected.

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

Genome View actions and shortcut menus

- Click a chromosome to select it. When you select a chromosome, Chromosome, Gene, and Tab Views show only genomic regions, genes, and data for it. The specific location in which you click the chromosome sets the position of the cursor. See "The View Cursor" on page 159.
- On the selected chromosome, click anywhere to move the cursor. See "The View Cursor" on page 159. This also moves the cursor in Chromosome, Gene, and Tab Views.
- Right-click anywhere within Genome 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 235.
- Click the **Detach** button (located at the top center of the pane) 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 pane to resize it.
- On a border of the pane, click a resize button (for example,  or ) that points away from the pane 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

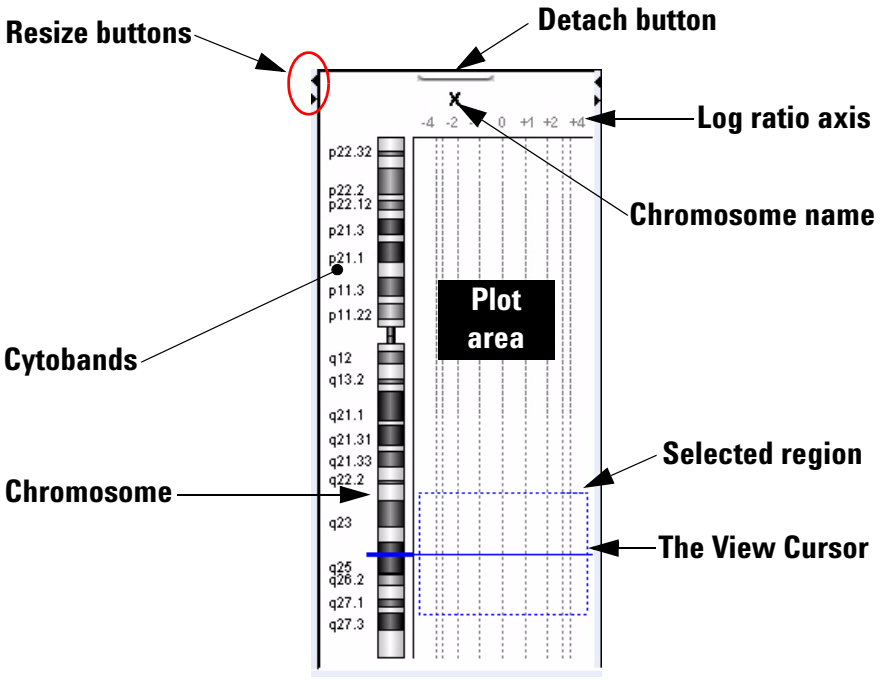


Figure 43 Chromosome View, human X chromosome shown

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 move the View cursor to that location. See "The View Cursor" on page 159.
- 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 moves the cursor to the center of the selected region. See "The View Cursor" on page 159.
- 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 235.
- Click the **Detach** button (located at the top center of the pane) 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 pane to resize it.
- On a border of the pane, click a resize button (for example,  or ) that points away from the pane 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

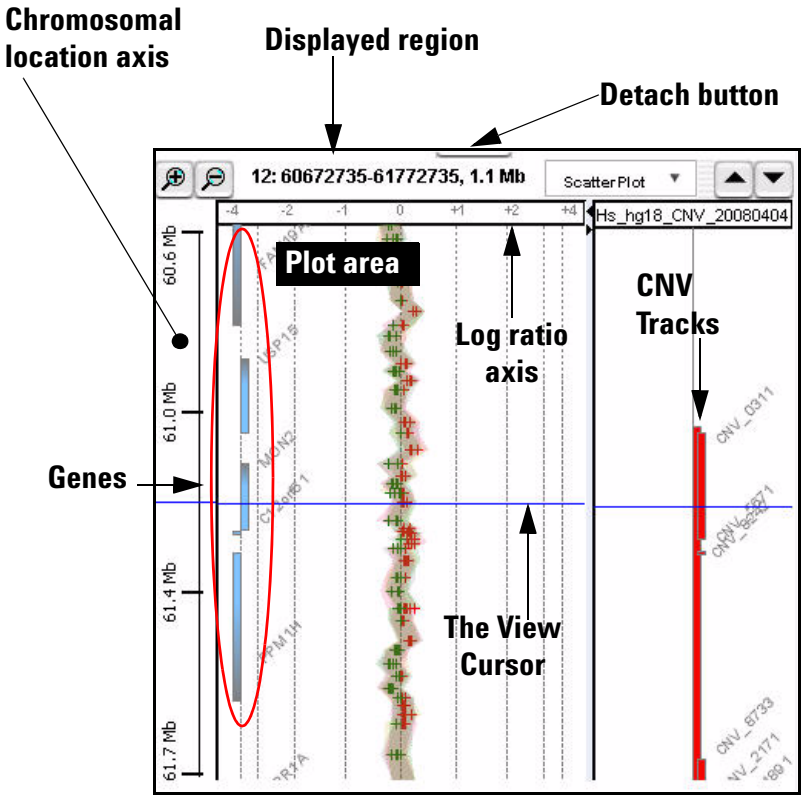


Figure 44 Gene View, with log ratio data from an experiment and CNV tracks.

Gene View shows a more detailed view of the chromosomal region you select in Chromosome View. See "Chromosome View" on page 153.

- Regions that contain 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 display only the genes in the list. See "To change the appearance of genes in Gene View" on page 92, and "To show gene lists in Gene View" on page 90.

- Log ratio data from selected arrays in the active experiment appear as a scatter plot. You can also customize the scatter plot. See "To change scatter plot appearance" on page 85.
- The location of the cursor matches the location of the cursors in other views. See "The View Cursor" on page 159.
- The name of the chromosome, and the coordinates and size of the displayed chromosomal region appear at the top of the pane.
- Imported tracks can also appear in Gene View. See "To show tracks in Gene View" on page 92.

Scatter Plot

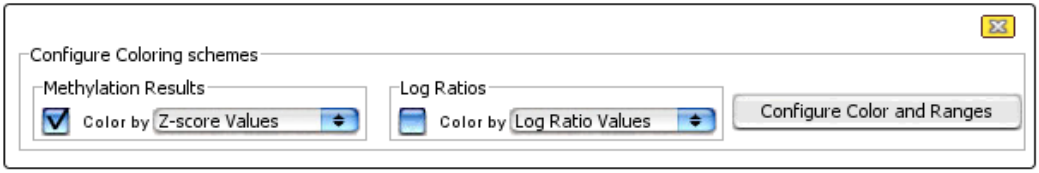


Figure 45 Scatter Plot command group in CH3 Gene View

You access the scatter plot command group in Gene View or through View Preferences from the **View** tab. The commands are different for CGH, ChIP and CH3 applications. Scatter plots appear in the Chromosome and Gene Views but only if they have been selected in the View Preferences dialog box.

Methylation Results Mark the box to enable the Methylation Results scatter plot. Selections for the CH3 Methylation Results scatter plot are Z-score Values or Probe Score Values.

Log Ratios Mark the box to enable the Log Ratios scatter plot. Selections for the CH3 Log Ratios scatter plot is Log Ratio Values.

Configure Colors and Ranges Opens the Configure Coloring Ranges and Shades dialog box, where you can set up the colors and ranges for the Methylation Results and Log Ratios scatter plots. For more information, see "Configure Coloring Ranges and Shades" on page 173.

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


When in vertical 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

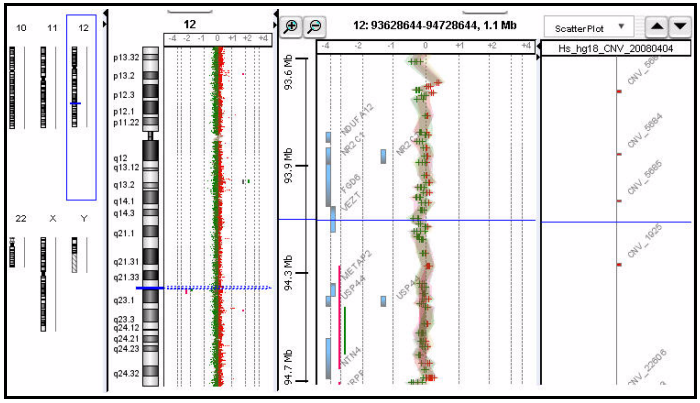
- Click anywhere in the plot area of Gene View to move the cursor to that location. See "The View Cursor" on page 159.
- Drag an inside border of Gene View to resize the view.
- Right-click anywhere in the plot area of Gene View to display these options:

Option	Description
Create Gene List	Opens the Create Gene List dialog box, where you can create a new gene list based on the selected (or another) chromosomal region. See "Create Gene List" on page 179 and "To create a gene list" on page 65.
Upload Track to UCSC	Opens the Upload Track to UCSC dialog box, where you can set parameters to upload the track to the UCSC (University of California at Santa Cruz) Genome Browser in your Web browser. You can then display the track and use the tools available in the UCSC Web site to examine the data. See "Upload Track to UCSC" on page 228.
Search Probes in eArray	Opens the Search Probes in eArray dialog box, where you can set the region to search for probes in eArray.
User Preferences	Opens the User Preferences dialog box, where you can set user preferences on three separate tabs. See "User Preferences" on page 230 for more information.
View Preferences	Opens the View Preferences dialog box, where you set preferences for the Genomic Viewer. See "View Preferences" on page 235.

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 change 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 plotting area to move the cursor to that location.
- In Gene View, click anywhere in the plotting 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 for that row.

Tab View

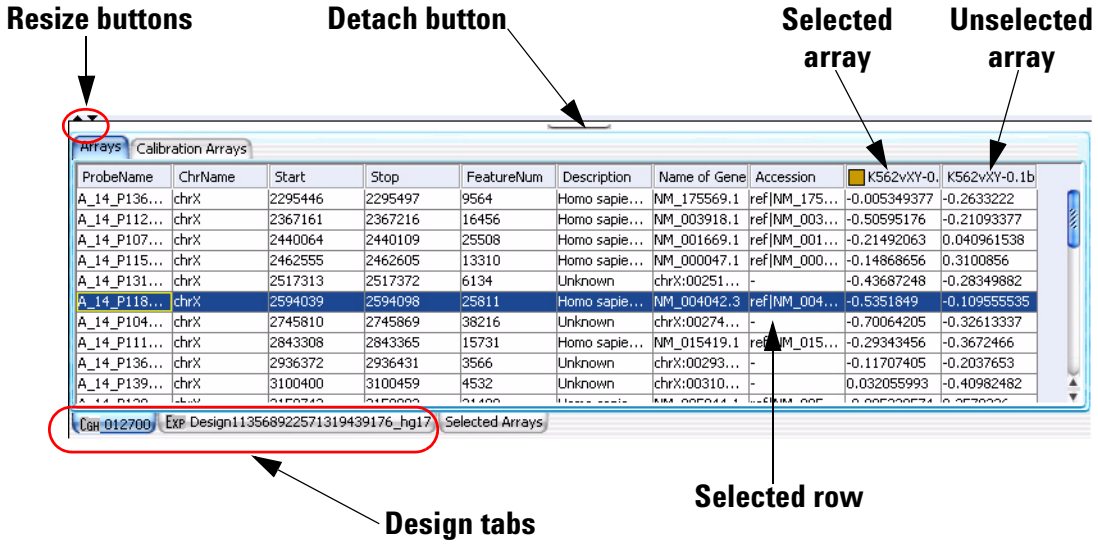


Figure 46 Tab View

Tab View displays design annotation and log ratio data related to the chromosome you select in Chromosome View.

- The exact column content of the tables depends on the specific tab and design, but it always includes chromosomal locations of probes.
- The selected row of data appears 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 specific 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.

Tab View tabs and buttons

You can see the following tabs and buttons in Tab View. See Figure 46 for a diagram that identifies some of these elements.

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

- CH3 – A methylation array design
- CGH – An aCGH array design.
- EXP – A gene expression array design.
- CHIP – A ChIP-on-Chip array design.

When you click a design tab, the data and annotation for the arrays in the design appear in Tab View. The program separates the arrays of the design into the Arrays tab and the Calibration Arrays tab (see below).

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.

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

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

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

Tab View actions and shortcut menus

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

Option	Description
Rename Array	Opens an Input dialog box, where you can type a new name for the array. This only changes the name of the array within the active experiment.
Remove Array From Experiment	Opens a confirmation dialog box. Click **Yes** to remove the link between the array and the active experiment. This command does not delete the data file from the program. To do this, see "To remove data or design files from the program" on page 64.
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 "Edit Array Color" on page 185 and "To change the display color of an array" on page 81.
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 the active experiment appear in Tab View and in the Data 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 186 and "To change the order of arrays in an experiment" on page 54.
Select All Arrays	Selects all arrays in all designs in the active experiment for display. All arrays appear in the Selected Arrays tab.
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. See "Scroll to Column" on page 213.

- Right-click a *heading of a column other than an array data column* to open a shortcut menu with 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 corresponding 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. See "Find in column" on page 196.
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 sends 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. See "To search the Web for information on probes in Tab View" on page 100.
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 sends the column entry you clicked as a search string to the site. See "Customize Search Link" on page 181 and "To create a custom Web search link" on page 100.
(other options)	If other options appear in this shortcut menu, they are custom Web search links. Click them to open your Web browser, and send the column entry you clicked as a search string to the site.

Status Bar

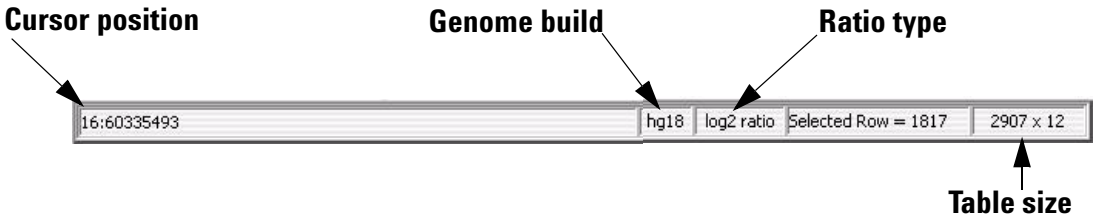


Figure 47 Status bar

The Status Bar displays information related to the displayed data.

Cursor position The chromosomal location of the cursor. See "The View Cursor" on page 159.

Genome build The genome build for 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 for this row.

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

Dialog Boxes

This section describes the dialog boxes specific to the interactive analysis of the CH3 Application. They are listed in alphabetical order by title.

Add Gene List <name> to

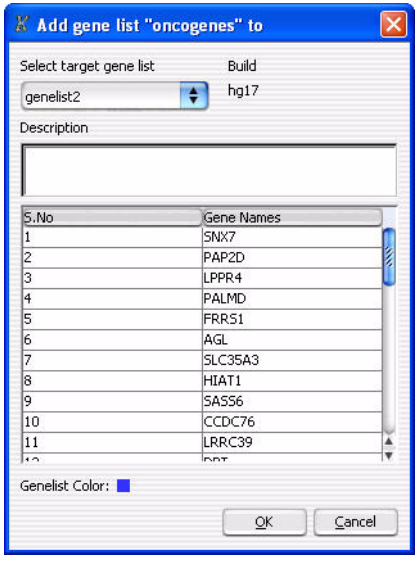


Figure 48 Add Gene List <name> to

Purpose: Adds genes from one gene list (the source gene list) to another (the target gene list).

To open: In the Data pane, right-click the name of a gene list, then click **Add to Gene List.**

Select target gene list The gene list to which genes are added. Select one from the list.

Build (Read-only) The genome build for the genes in the list. The builds of the two gene lists must match.

Description (Optional) Description of the combined gene list.

List of genes A list of the genes in the target gene list.

Gene List Color (Read-only) The display color of the target gene list.

OK Adds the genes from the source gene list to the target gene list.

Cancel Closes the dialog box without adding any genes to the target gene list.

Agilent Feature Extraction Importer

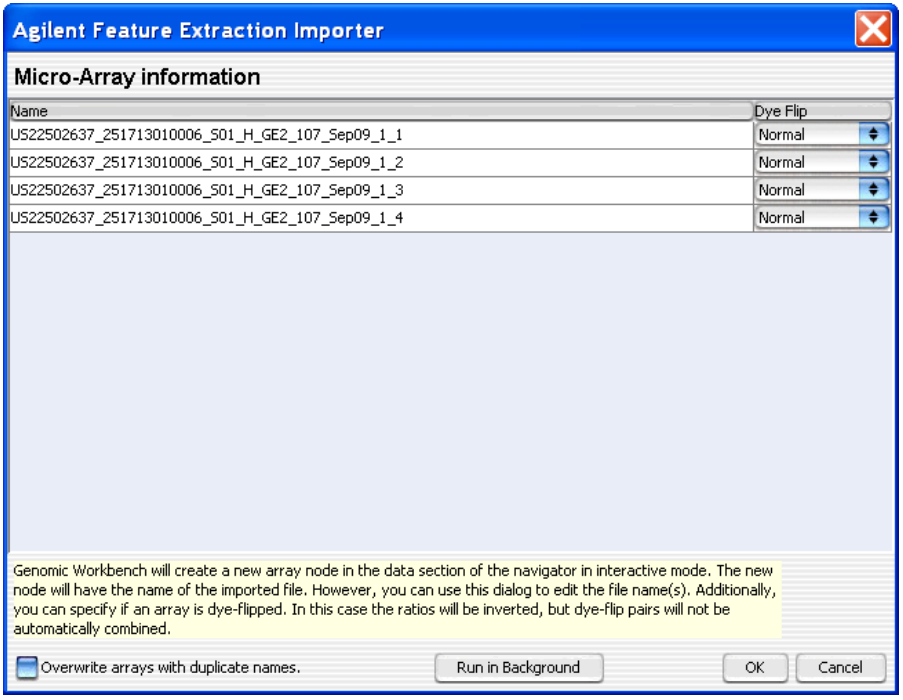


Figure 49 Agilent Feature Extraction Importer

Purpose: Lets you edit the name of the FE data file you will import and to indicate whether you want to flip the red/green ratio for the data.

To open: In the Home tab, click **Import > Array Files > FE File,** select the desired FE data file(s), then click **Open.**

Name Lets you edit the names of the FE files. You can change the names of the files to names that are easier 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 as is.

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).

Run in Background Imports the files, and lets you use your computer for other purposes while the import occurs. This is especially useful if you have many files to import.

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.

Array Set

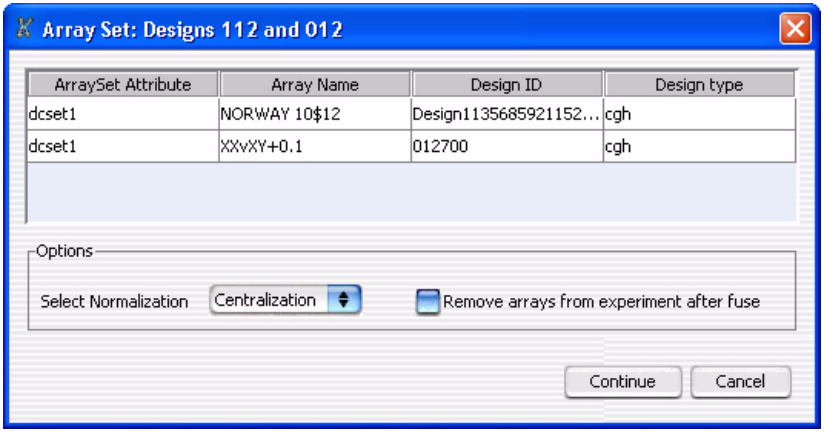


Figure 50 Array Set dialog box.

Purpose: To select replicate arrays to combine in the analysis.

To open: Click **Fuse** in the Preprocessing ribbon.

The array set dialog box opens when you combine designs. See "To combine (fuse) arrays" on page 105.

Table Displays the arrays that are combined, in order of their values for the ArraySet attribute. The program creates a separate combined array for each group of arrays with a given value for ArraySet. The table also shows the design ID for each array, and the design type.

Select Normalization Select Centralization normalization or none. See "Methylation Detection and Measurement Algorithms" on page 244.

Remove arrays from experiment after fuse Deletes the original individual arrays after creating combined arrays so that the experiment won't contain duplicate data.

Continue Click to create combined designs using the selected options.

Cancel Cancels any selections, and closes the dialog box.

Batman Parameter Setup

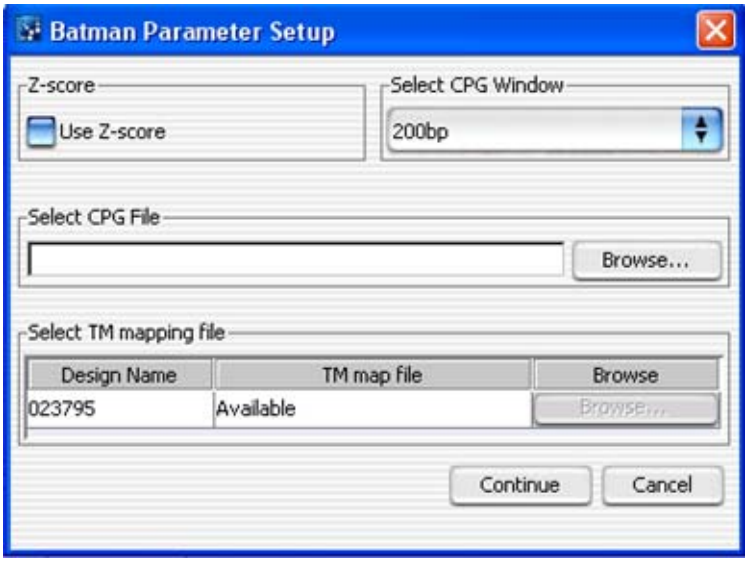


Figure 51 Batman Parameter Setup dialog box

Purpose: This dialog box is used to set the parameters for the Batman tool of methylation analysis. For details on how this algorithm works, see "Bayesian tool for methylation analysis (BATMAN)" on page 254.

To open: In the **Analysis** tab, under **Batman**, mark the **Apply** box.

Z-score Mark Use Z-score to use the Z-score in place of the log ratio values.

Select CPG Window Select the local CPG window to use.

Select CPG File Click **Browse** and select a FASTA file containing sequences with 400 bases on either side of the center of the probe. The file must be in standard FASTA format.

Select TM mapping file If the Tm information is present in the design, the TM map file will display *Available*. If the Tm information is not in the design, you must select a TM file to use. The Tm file must be a tab delimited text file with probe name and Tm as two columns.

Continue Click to apply the Batman parameters and display them in the views.

Cancel Click to close the dialog box without applying the Batman algorithm.

Batman Report Dialog

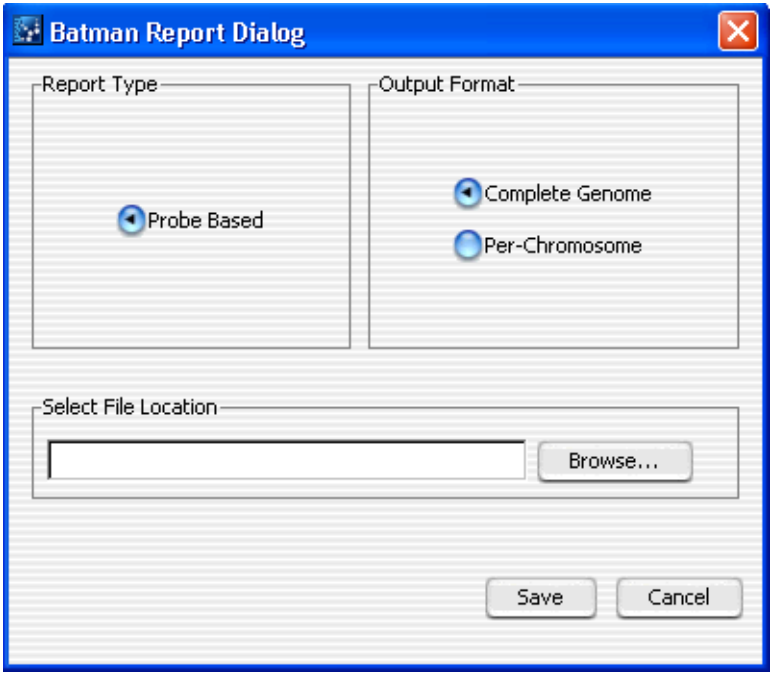


Figure 52 Batman Report Dialog

Purpose: This dialog box is used to select the output format and location of the Batman report.

To open: In Reports, click **Generate Batman Report**.

Report Type All methylation reports are probe-based.

Output Format **Complete Genome** – Creates report as .xls file for entire genome

Per-Chromosome – Creates report for each chromosome as .txt file. Open in spreadsheet program to see headers properly aligned.

Select File Location Click **Browse** to select a location for the file, and type a file name.

Combine Tracks

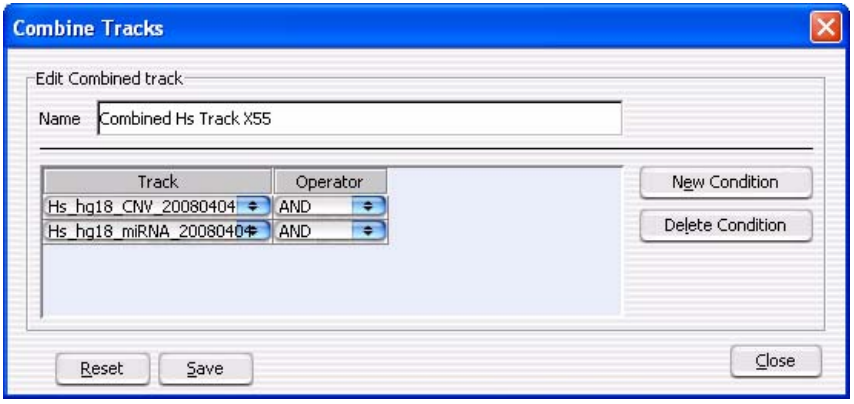


Figure 53 Combine Tracks dialog box

Purpose: Lets you create a combined track that contains elements of two or more source tracks, associated by logical criteria. See "To combine tracks" on page 68.

To open: In the **My Entity List** pane, right-click the **Tracks** folder, then click **Combine Tracks.**

Name The name of the combined track

New Condition Adds a new, empty row to the Track/Operator table in the dialog box.

Delete Condition Removes the bottom row from the Track/Operator table in the dialog box.

Track In each row, select a track to include in the combined track.

Operator In each row, select the desired logical operator. This operator controls the manner in which the program combines the track in this row with the others. Select one of these options:

Operator	Comments
AND	Places an element in the combined track if it appears in both this track and any of the others.
OR	Places an element in the combined track if it appears in either this track or any of the others. If you set this operator for all tracks in the list, the result is a nonredundant set of elements from all tracks.
MINUS	Removes the elements that appear in this track from the combined track, if they otherwise appear there.

Reset Removes all Track/Operator pairs from the table in the dialog box, and clears the Name of the combined track.

Save Creates the combined track, but does not close the dialog box.

Close Closes the dialog box. Opens the Confirm track save dialog box if you created a combined track, but did not save it.

Configure Coloring Ranges and Shades

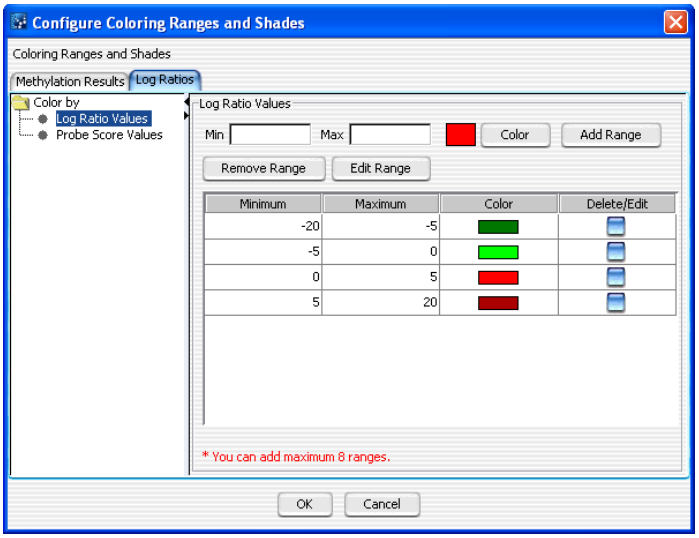


Figure 54 Configure Coloring Ranges and Shades dialog box

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

To open: In Gene View, move the mouse pointer over **Scatter Plot** to display the scatter plot options and then click **Configure**. Or, click the **View** tab and click **View Preferences**. Under Configure Coloring schemes, click **Configure Colors and Ranges.**

Table 4 Methylation Results Parameters

Parameter	Description
Z-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 215 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 marked.
Edit Range	Click to edit range(s) with Edit/Delete box marked.
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.

Table 5 Log Ratios Parameters

Parameter	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 215 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 marked.
Edit Range	Click to edit range(s) with Edit/Delete box marked.
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.

Table 5 Log Ratios Parameters (continued)

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

Confirm Overwrite

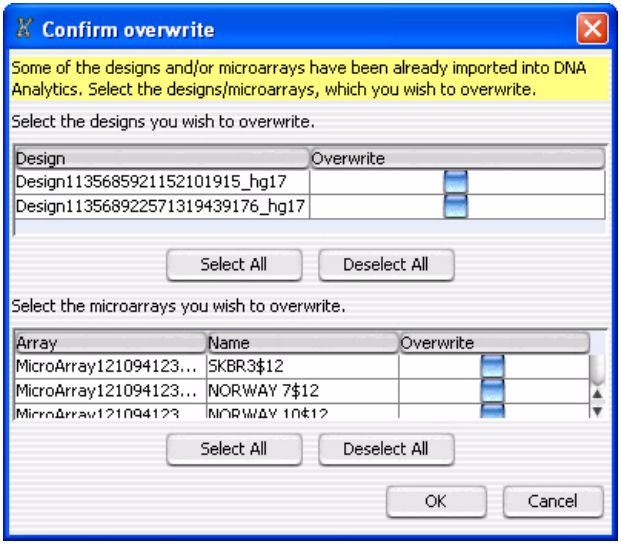


Figure 55 Confirm overwrite dialog box

Purpose: When you import an experiment, it can contain designs and/or arrays that have the same names as those already available in Agilent Genomic Workbench. This dialog box lets you select which designs and/or arrays to overwrite.

To open: This dialog box appears when you import a ZIP format experiment file, and it contains designs and/or arrays that are already available in Agilent Genomic Workbench. See "To import an experiment file" on page 48.

Select the designs to overwrite

Design The names of the designs in the imported file that have the same names as designs that are already available in Agilent Genomic Workbench.

Overwrite Mark the check box for each design that you want to overwrite.

Select All Marks all of the check boxes under Overwrite.

Deselect All Clears all of the check boxes under Overwrite.

Select the microarrays to overwrite

Array The microarray identification, usually a barcode.

Name The names of the arrays in the imported file that have the same names as arrays that are already available in Agilent Genomic Workbench.

Overwrite Mark the check box next to each existing array that you want to overwrite.

Select All Marks all of the check boxes under Overwrite.

Deselect All Clears all of the check boxes under Overwrite.

OK Overwrites the selected files (both designs and arrays) and closes the dialog box.

Cancel Closes the dialog box, and returns you to the Import (experiments) dialog box. See "Import (experiments)" on page 202.

Create Experiment

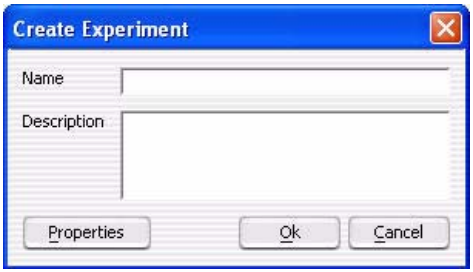


Figure 56 Create Experiment dialog box

Purpose: To create an organizational unit in the Experiment pane of the Navigator to link to array data for display and analysis and to create the links to the data for the experiment (see "Experiment Properties" on page 187).

To open: In the Experiment pane of the Navigator, right-click the Experiments folder, and click **New Experiment**, or click **File > New Experiment**.

Name Type a name for your new experiment.

Description Briefly, describe your experiment with information that will help you identify it.

Properties Click to display the Experiment Properties dialog box where you can select microarrays to add to your new experiment. See "Experiment Properties" on page 187.

NOTE

Do not click **OK** until you have added arrays to your experiment in the Experiment Properties dialog box or you will have an empty experiment. You can also add arrays to the experiment later, by dragging and dropping the arrays from the Data pane of the navigator. See "To add arrays to an experiment" on page 53.

Create Gene List

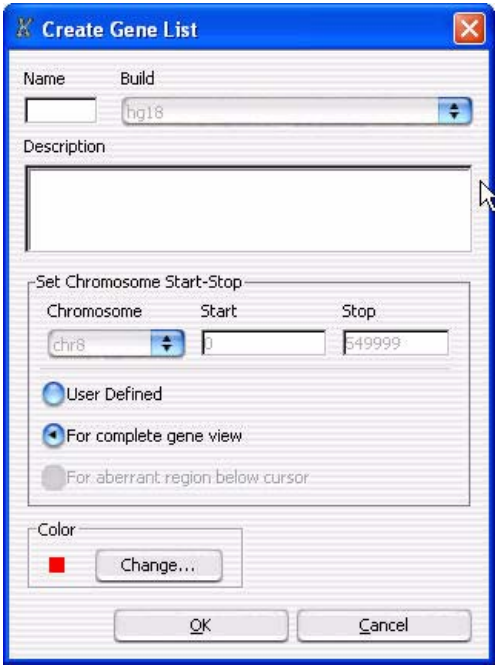


Figure 57 Create Gene List dialog box

Purpose: To limit the genes presented in Gene View to a preselected number valuable for interpreting data

To open: Right-click Gene View, and click **Create Gene List**.

Name Type in name of gene list.

Build Select the genome build for the genes.

Description Describe the type or nature of the genes in the list.

Set Chromosome Start-Stop

Select a chromosome and a region in Chromosome View for selecting the genes in the list before you open the Create Gene List dialog box.

User Defined Select to choose region from which the genes in Gene View are selected. The chromosome selection list and the Start and Stop positions on the Y axis are enabled when this option is selected. With this option, you can override the selections you made before opening Create Gene List.

For complete gene view Select all the genes in Gene View.

For aberrant region below cursor Select those genes that appear in the aberrant region just below the cursor position in Gene View.

Chromosome If you select User Defined, you can select a different chromosome than had been selected before opening the Create Gene List dialog box.

Start If you select User Defined, you type in a Start position to define the region that contains the genes for the list.

Stop If you select User Defined, you type in a Stop position to define the region that contains the genes for the list.

Color

Change Click to change the color of the gene list name in Data Navigator.

Customize Search Link

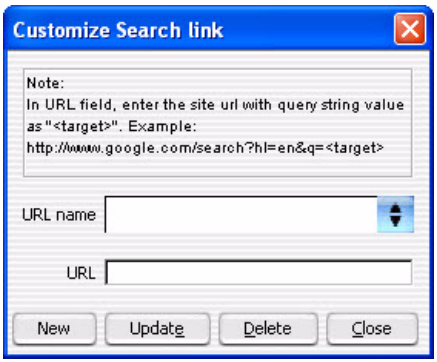


Figure 58 Customize Search Link dialog box

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. The link opens the URL of your choice, and sends the table entry to it as a search string. See "To create a custom Web search link" on page 100.

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 sends 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 add the name 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.

Design Properties

Purpose: Gives general and detailed information about a given microarray design. See "To display the properties of a specific design" on page 62.

To open: In the Data pane of the Navigator, right-click the name of a genome build within a design folder, 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. For Catalog designs, the "Date" field is the date of the downloaded catalog design file.

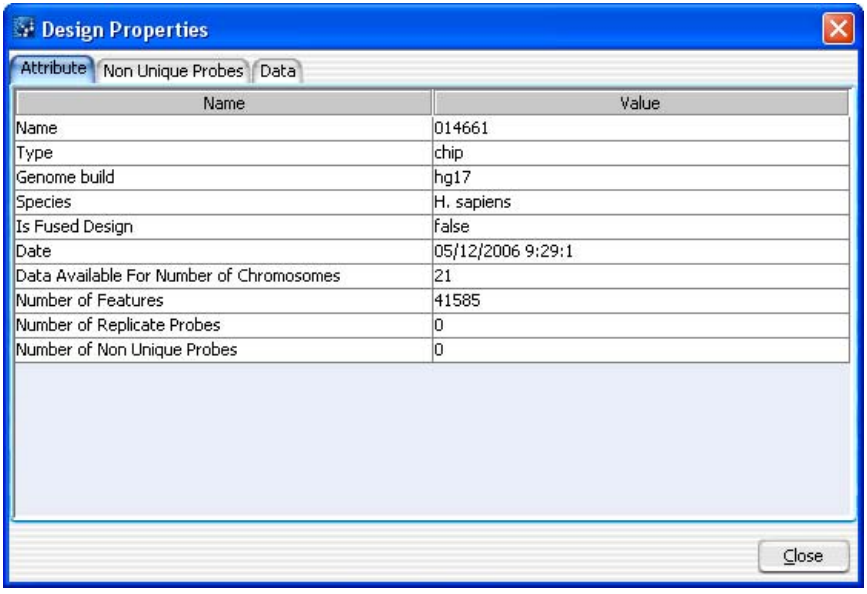


Figure 59 Design Properties dialog box – Attribute tab

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.

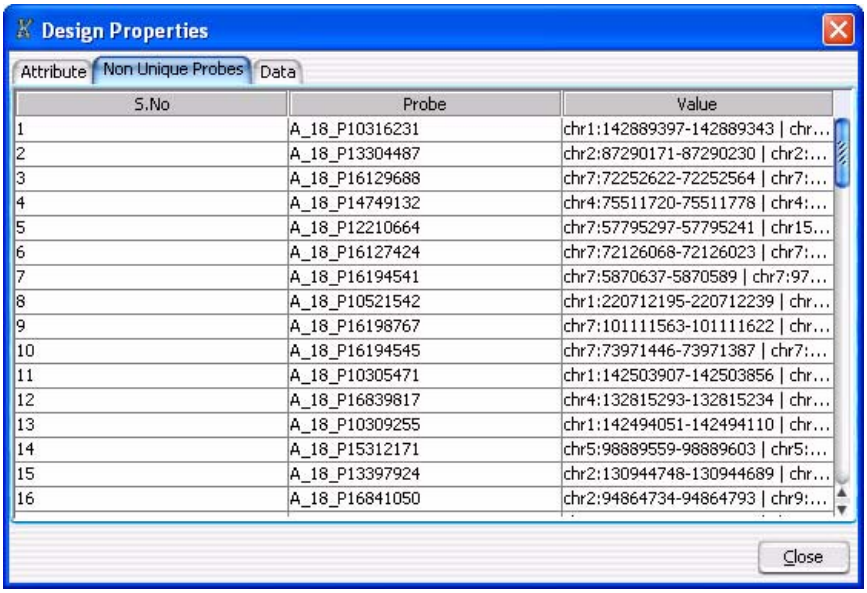


Figure 60 Design Properties dialog box – Non Unique Probes tab

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 binds. Because these are nonunique probes, two locations appear for each probe.

Data tab

Displays the names of the probes in the design and their target genomic locations. The tab displays the probes for one chromosome at a time.

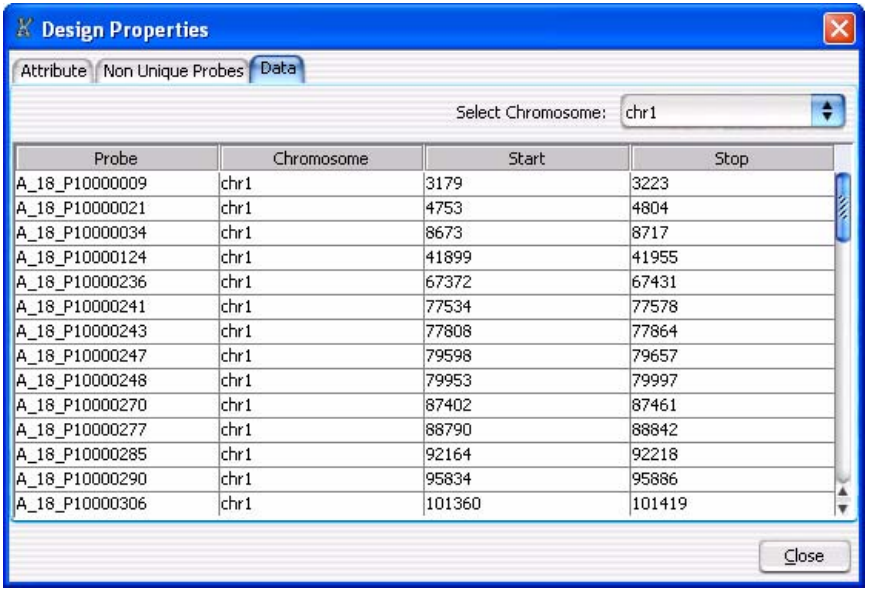


Figure 61 Design Properties dialog box – Data tab

Select Chromosome The chromosome whose probes appear in the list. To display the probes for another target 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 for the probe.

Stop The location on the selected chromosome of the last base pair for the probe.

Edit Array Color

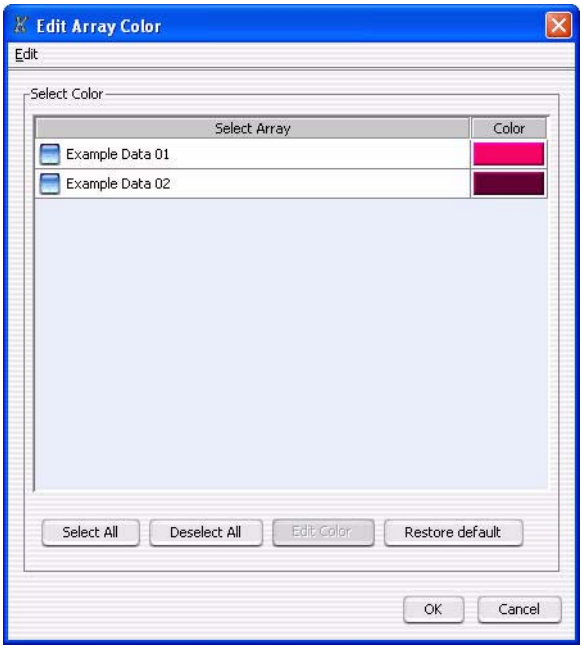


Figure 62 Edit Array Color dialog box

Purpose: To change the color of the arrays in an experiment

To open: Right-click the experiment name, and click **Edit Array Color**.

Select Array Mark the check box for the array(s) whose color you want to change.

Color Click to change the color for the selected array(s). If you selected more than one, all the selected arrays will change to the same color.

Select All Click to mark all the check boxes.

Deselect All Click to clear all the check boxes.

Edit Color Click to change the color for the selected array(s). Same as Color button.

Restore default Click to restore the original color(s) to the selected array(s).

Edit Array Order

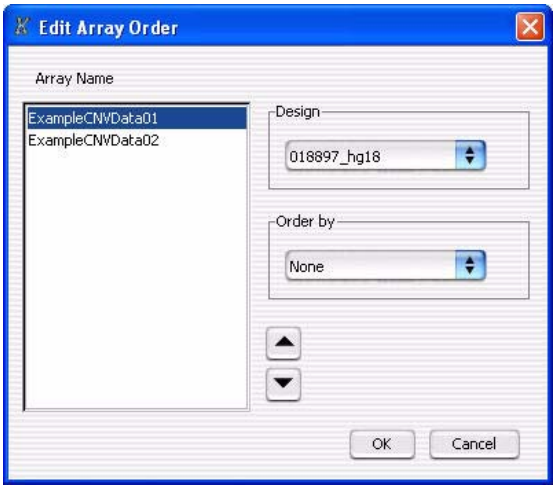


Figure 63 Edit Array Order dialog box

Purpose: This dialog box lets you change the order of arrays in an experiment, which defines the order in which the program displays arrays and array data in the Experiment pane of the Navigator and in Tab View. If you select the **Stacked** rendering style for scatter plots, the array order also determines the order in which the scatter plots for the arrays appear.

To open: Right-click the name of an experiment, then click **Edit Array Order** in the shortcut menu.

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 reordered 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.

Experiment Properties

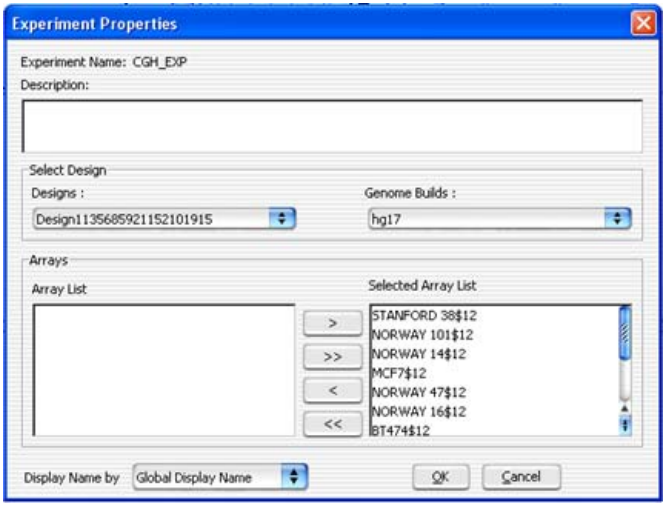


Figure 64 Experiment Properties dialog box

Purpose: To select the arrays to link to the experiment

To open: In the Create Experiment dialog box, click **Properties**, or in the Experiment pane of the Navigator, right-click the experiment name, and click **Show Properties**.

Experiment Name: Displays the name of the selected experiment.

Description Text Box: Displays the description of the experiment that was entered when the experiment was created.

Select Design

Designs Select the design whose arrays you want to add to the experiment.

Genome Builds Select the genome build for the design you selected, if the design has more than one genome build.

Arrays

Array List Displays the arrays in the selected design that are available for the experiment.

- 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 you have selected for the experiment.

Moves the selected arrays in Array List to the Selected Array List. You can move arrays from as many designs as you want, if they are all for the same genome build.

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 for removal, click its name. If desired, you can re-add an array.

Clears the Selected Array List.

Display name by Click to select how the array names are displayed in the experiment. The Global Display name is the name assigned in Sample Manager for the array. See the *Sample Manager User Guide* for more information.

OK Adds the selected arrays to the experiment and closes the dialog box.

Cancel Closes the dialog box without making any changes.

Export

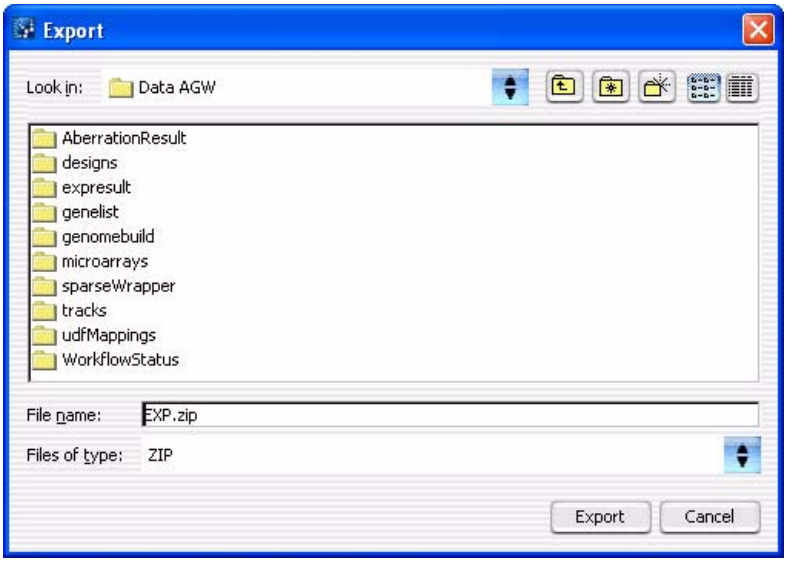


Figure 65 Export dialog box – Several types of file exports use this dialog box. This specific example exports selected experiment(s) as a ZIP format file.

Purpose: Lets you select a location for an exported file.

To open: This dialog box appears after you select specific experiment(s), track(s), or array attribute(s) to export. See "To export experiments" on page 72, "To export tracks" on page 73, or "To export array attributes" on page 71.

Look in Displays the folder or other location whose contents appear in the main pane of the dialog box. To select another folder or other location, click .

Moves to the next higher folder level.

Displays the Desktop.

Creates a new folder in the selected location in *Look in*.

Displays the names, only, of folders, files, and other locations in the main pane of the dialog box.

Displays both the names and information about folders, files, and other locations in the main pane of the dialog box.

Main pane Displays the folders, files, and other locations in the selected location in *Look in*. Only files of the selected type are displayed. To select file, click its name. To open a folder or other location, double-click its name.

File name Displays the name of the file to which the exported content is saved. To change the name, you can either select a file in the main pane of the dialog box, or type a new name.

Files of type Sets the type of files that are displayed. To show all files, click  , then select **All Files.**

Export Saves the selected content to the location given in the dialog box.

Cancel Cancels your selections and closes the dialog box.

Export Array Attributes

Purpose: This dialog box lets you select arrays whose attributes you want to export. It contains two tabs: an Array tab where you select the arrays, and an Attribute tab where you select the attributes of the selected arrays to export. See "To export array attributes" on page 71.

To open: In the Home command ribbon, click **Export > Array Attributes**, or in the Experiment pane of the Navigator, right-click the name of an experiment, then click **Export Attributes.**

Array tab

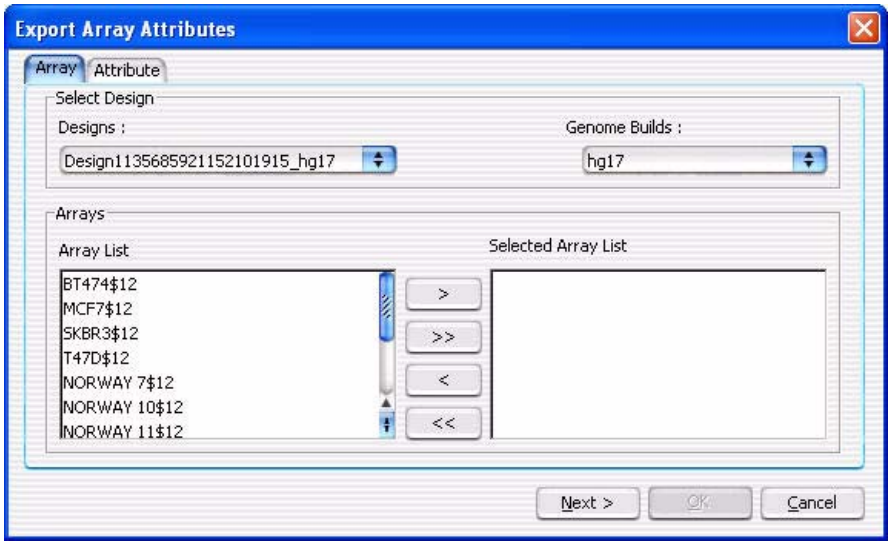


Figure 66 Export Array Attributes dialog box – Array tab

Select Design

Designs Displays all of the designs available in the program. Select the design for arrays that you want to export.

Genome Builds Displays the genome build(s) for the design. Select the desired genome build to display the arrays for a single genome build.

Arrays

Array List Displays the arrays in the selected design that are available for export.

- 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 you have selected for export.

> Moves the selected arrays in Array List to the Selected Array List. You can move arrays from as many designs as you want, as long as they are all for the same genome build.

>> 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 for removal, click its name. You can later add a removed array.

<< Clears the Selected Array List.

Next Moves to Attribute tab to select attributes for export.

Cancel Closes the dialog box without exporting any array attributes.

Attribute Tab

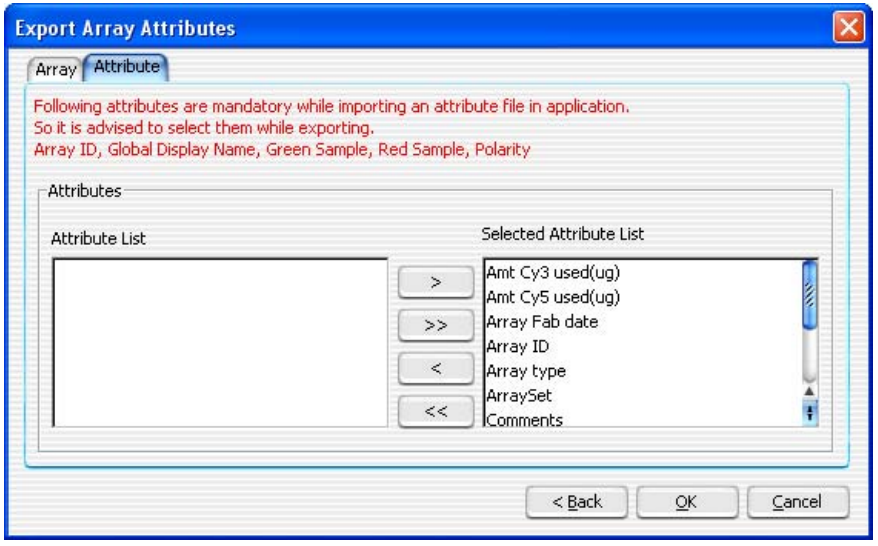


Figure 67 Export Array Attributes dialog box – Attribute tab

Attributes

Attribute List Displays the attributes that are not exported for the selected arrays.

Selected Attribute List Displays the exported attributes for the selected arrays.

- To select an attribute to move to the 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.

NOTE You must select the following mandatory attributes, or else you cannot import the attribute file at a later time: Array ID, Global Display Name, Green Sample, Red Sample, Polarity.

< Removes an attribute from the Selected Attribute List. To select an attribute for removal, click its name. You can add a removed attribute later.

<< Clears the Selected Attribute List.

> 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.

Back Moves back to the Array tab for array selection or removal.

OK Opens the Export dialog box. See "Export" on page 189.

Cancel Closes the dialog box without exporting any array attributes.

Export Experiments

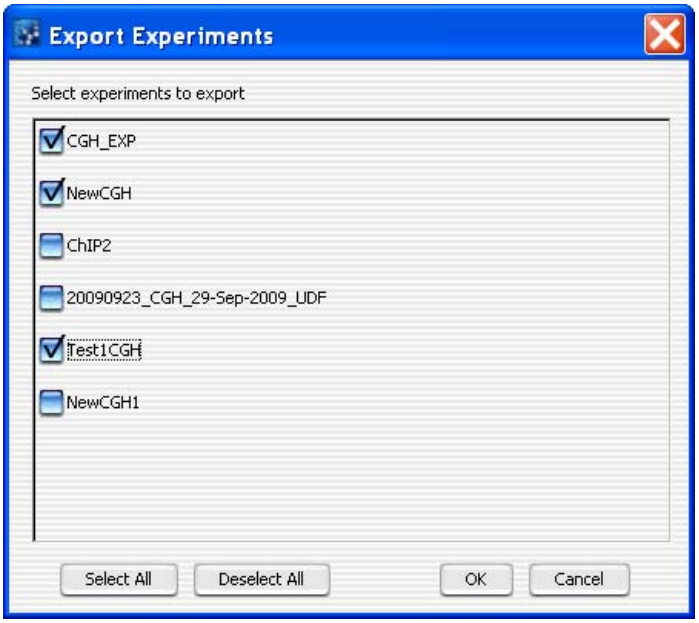


Figure 68 Export Experiments dialog box

Purpose: Lets you select experiments for export. The program exports all array designs and data for the experiments as a single ZIP file. This file does not include any parameter settings, array selections, or results. See "To export experiments" on page 72.

To open: In the Home tab, click **Export > Experiments.**

Select experiments to export Displays all experiments available for export. Mark each experiment you want to export.

Select All Selects all experiments for export.

Deselect All Clears all check boxes under Select experiments to export.

OK Opens an Export dialog box. See "Export" on page 189.

Cancel Cancels the export and closes the dialog box.

Export Tracks

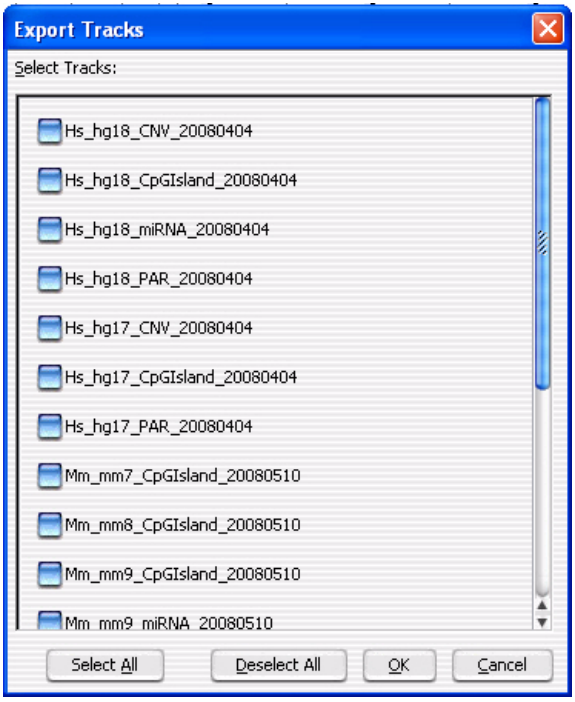


Figure 69 Export Tracks dialog box

Purpose: Lets you select tracks to export as a single BED format file. See "To export tracks" on page 73.

To open: In the **Home** tab, click **Export > Tracks.**

Select tracks Displays all of the filters available in the program. Mark the check box next to each track that you want to export.

Select All Selects all available tracks for export.

Deselect All Clears all of the check boxes under Select Tracks.

OK Opens the Export dialog box, where you can select a location for the exported BED format file. See "Export" on page 189.

Cancel Cancels the export and closes the dialog box.

Find in column

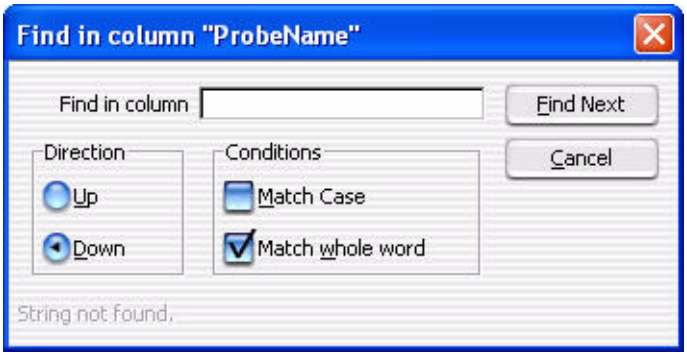


Figure 70 Find in column dialog box

Purpose: This dialog box lets you 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.

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

Direction Select a search direction:

- **Up** – Sets the search to move up the column from the selected location.
- **Down** – Sets the search move down the column from the selected location.

Conditions Mark any of these search options:

- **Match Case** – Mark this option to take case into account. For example, if you mark 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** – Mark 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 mark **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.

Gene List

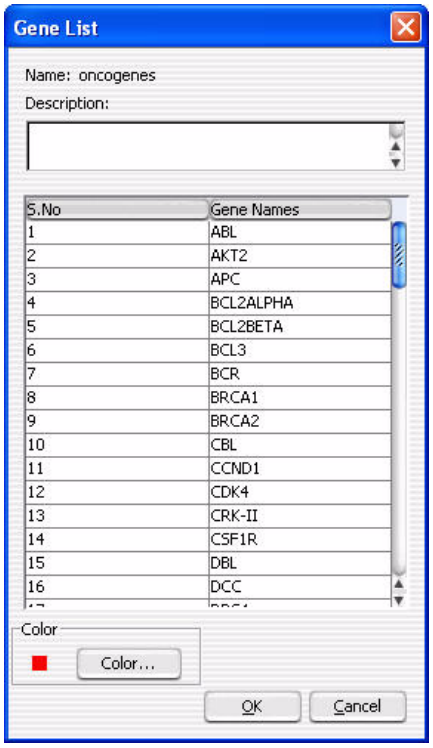


Figure 71 GeneList dialog box

Purpose: Lets you view the names of the genes in a specific gene list and to change the display color of the gene list. See "To display the genes in a gene list" on page 66.

To open: In the **My Entity List** pane of the Navigator, right-click the name of a gene list, then click **View in Table.**

Name (Read-only) The name of the gene list.

Description (Optional) Brief descriptive comments about the gene list, such as how it was created or the nature of the genes in the list. You can edit the description.

S. No The sort order number. This is the index number of each gene within the gene list.

Gene Names The names of the genes in the gene list.

Color Opens the Choose Gene List Color dialog box, where you can change the display color for the gene list. See "Combine Tracks" on page 171.

OK Saves the gene list with any new description or display color, and closes the dialog box.

Cancel Closes the dialog box without making any changes to the gene list.

Go To Gene/Genomic Location

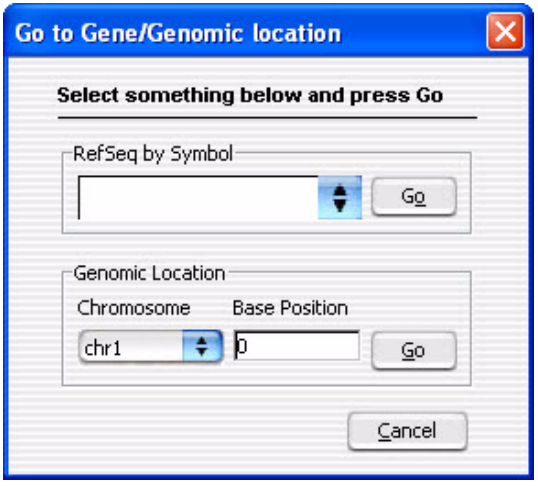


Figure 72 Go To Gene/Genomic location dialog box

Purpose: To find a specific gene location in Gene View by either selecting the RefSeq by Symbol or by selecting the Genomic Location.

To open: Click **Home > Go to Gene/Genomic location**.

RefSeq by Symbol Select the Reference Sequence accession symbol from NCBI, and click **Go.**

Genomic Location

- **Chromosome** – The chromosome number.

- **Base Position** – The position on the chromosome.

Click **Go** after selecting the chromosome number and the position of the gene on the chromosome.

Cancel Closes the dialog box.

Import

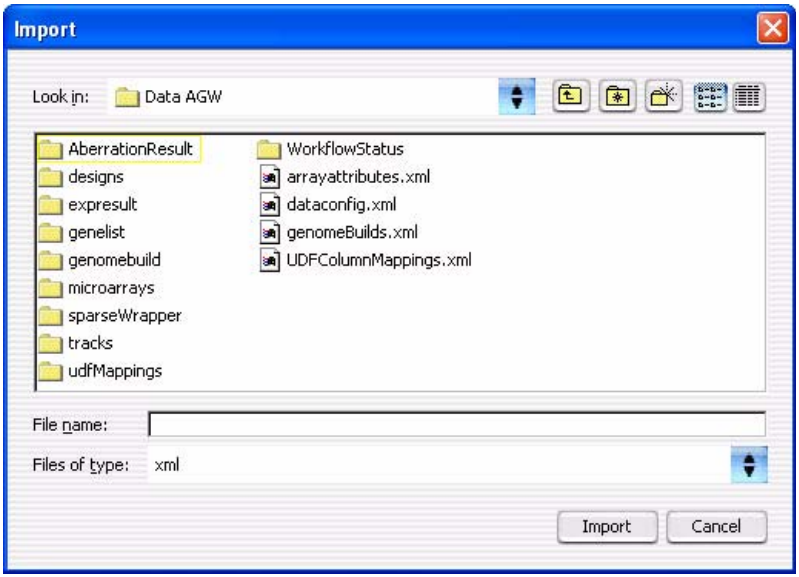


Figure 73 Import dialog box

Purpose: Lets you select files and import them into Agilent Genomic Workbench.

To open: In the Home tab, click **Import,** then select any kind of import except Genome Build or Track. The type of file for import appears in the title of the dialog box. To import a gene list, right-click the **Gene List** folder in the My Entities List pane of the Navigator, then click **Import Gene List.**

Use the standard Windows Explorer commands in the dialog box to select a file for import.

For some imports, you can select multiple files. Click the name of the first file, then hold down the **ctrl** key and click the names of additional files. To select a contiguous block of files, click the name of the first file in the block, then hold down the **Shift** key and click the name of the last one.

File name Displays the name of a file you select for import.

Files of type Lets you select the types of files to display from the types shown in the table below. To display all files, click  ,then select **All Files.**

File type	Extension
FE array File	*.txt
Axon array file	*.gpr
UDF file	*.txt
Design file (GEML)	*.xml
Axon design file	*.gal
Array attributes	*.txt
Experiments	*.zip
Filters	*.xml
Gene list	*.txt

Import or Open Imports the file into the program. In some cases, the name of this button is *Open*, rather than *Import*. Also, when you click **Import,** in many cases one or a series of additional dialog box(es) lets you further define the content for import. See the instructions for each type of import in Chapter 2.

Cancel Cancels the import and closes the dialog box.

Import (experiments)

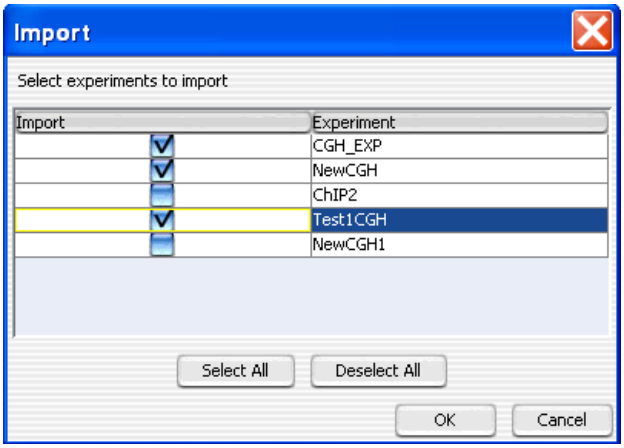


Figure 74 Import dialog box (for experiments)

Purpose: Lets you select the specific experiments within a ZIP format experiment file to import into the program. See "To import an experiment file" on page 48.

To open: In the **Home** tab, click **Import > Experiments.** In the dialog box that appears, select the desired ZIP format experiment file, then click **Import.**

Select experiments to import

These columns appear:

- **Import** – Mark the check box for the experiment(s) to import.
- **Experiment** – The names of the experiments available for import in the ZIP format experiment file.

Select All

Selects all of the experiments in the ZIP file for import.

Deselect All

Clears all of the check boxes under Import.

OK

Imports the selected experiments into the program. If the name of an imported array design or data file matches one that is already available in the program, the Confirm overwrite dialog box appears, where you can select the data and/or design files that you want to overwrite. See "Confirm Overwrite" on page 176.

Cancel Cancels the upload and closes the dialog box.

Import GEML design files

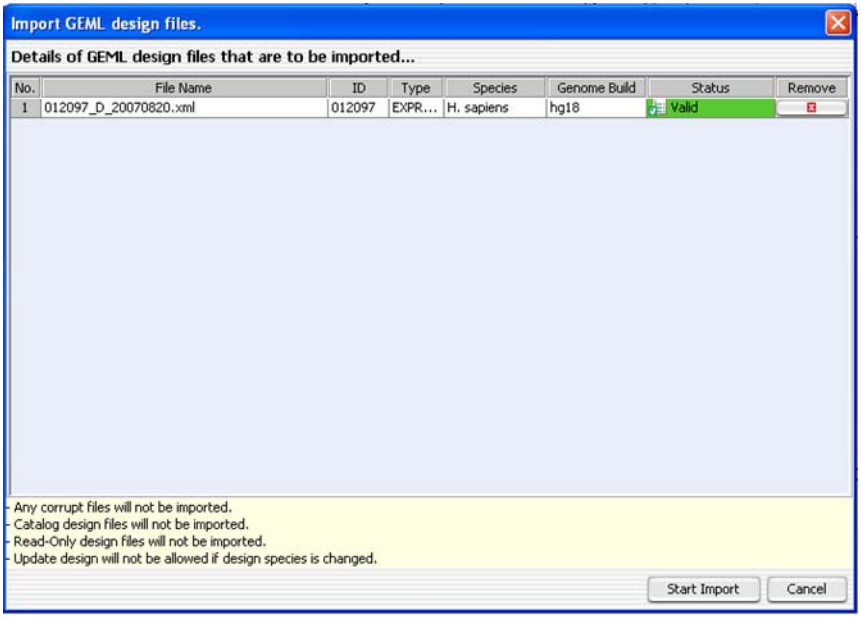


Figure 75 Import GEML design files dialog box

Purpose: To display information in the design file and to remove any files that you don't want to import.

To open: In the Home tab, click **Import > Design Files > GEML File.** Select the desired *.xml design files, then click **Open.**

File Name The name(s) of the design file(s) for import.

ID The Agilent ID number for the design file

Type The application type; CGH, ChIP, miRNA, or gene expression.

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.

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.
- **Valid** – Appears when the file is new.

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 upload and closes the dialog box.

Import Genome Build

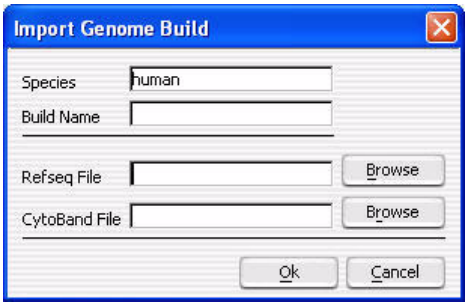


Figure 76 Import Genome Build dialog box

Purpose: To import a new set of genome build files into Agilent Genomic Workbench. See "To import a genome build" on page 45.

To open: In the Home tab, click **Import > Genome Build.**

Species The genome's species of origin.

Build Name The name of the build to import.

Refseq File The location of the RefSeq database file. This file contains chromosomal locations of genes. To select a Refseq file, click **Browse.**

CytoBand File The location of the applicable cytoband file. This file contains graphical cytoband information for Gene View and Chromosome View. To select a cytoband file, click **Browse.**

OK Imports the genome build and closes the dialog box.

Cancel Cancels the import and closes the dialog box.

CAUTION Import only Agilent-supplied genome build files.

Import Track

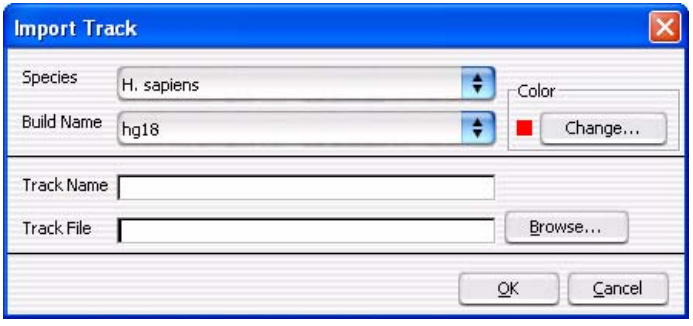


Figure 77 Import Track dialog box

Purpose: Lets you import a BED format track file. See "To import tracks" on page 46. Track information can appear in Gene view. See "Gene View" on page 155.

To open: In the **Home** tab, click **Import > Track.**

Species Select the species to which the track relates.

Build Name This list contains the available genome builds for the selected species. Select the desired genome build.

Color The assigned display color for the track. To change this color, click **Change.**

Track Name Type a name to identify the imported track.

Track File Type the location of the BED track file that you want to import, or click **Browse** to select a file.

Browse Opens an Open dialog box, where you can select the BED track file to import.

OK Imports the track into the program.

Cancel Cancels the import and closes the dialog box.

Microarray Properties

Purpose: Displays the properties associated with an array. You can also edit the values of specific attributes. To add attributes to the list, see the *Sample Manager User Guide*.

To open: For any array in the Data folder or Experiments folder, right-click the array name, then click **Show Properties.** For non-Agilent arrays, only the Attribute tab appears.

Attribute tab

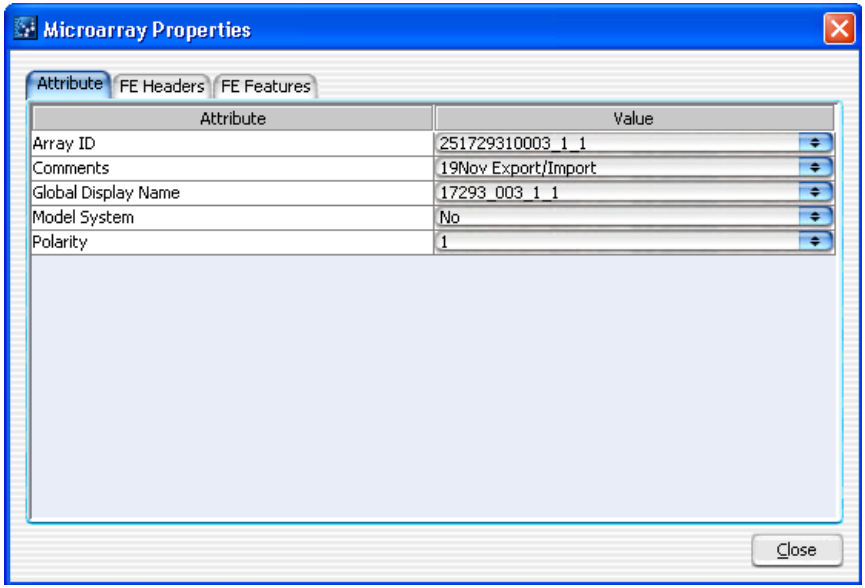


Figure 78 Microarray Properties dialog box with list of Attributes and their values

- **Attribute** — Displays the attributes in an array by name. You can load these from an Excel spreadsheet.
- **Value** – Indicates the values, if any, for each array.

NOTE You cannot edit values for read-only arrays.

Close Closes the dialog box.

FE Headers Tab

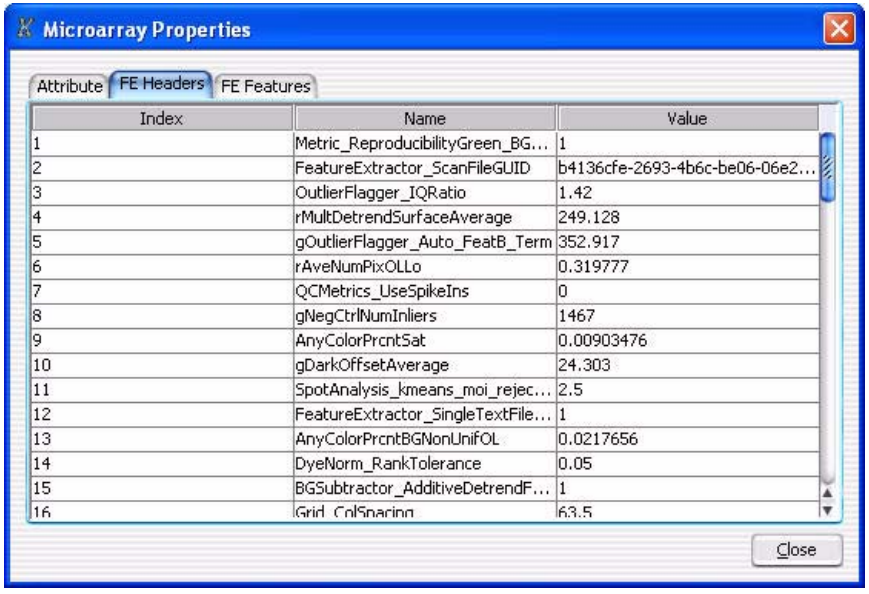


Figure 79 Microarray Properties dialog box with list of FE Headers and their values

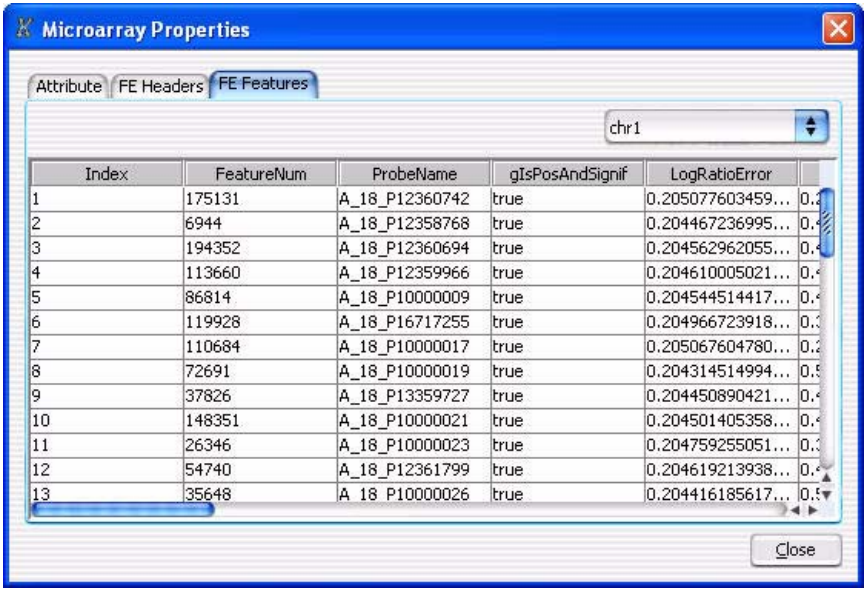
Index Displays a sequential index to help identify FE properties.

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



Index	FeatureNum	ProbeName	gIsPosAndSignif	LogRatioError
1	175131	A_18_P12360742	true	0.205077603459...
2	6944	A_18_P12358768	true	0.204467236995...
3	194352	A_18_P12360694	true	0.204562962055...
4	113660	A_18_P12359966	true	0.204610005021...
5	86814	A_18_P10000009	true	0.204544514417...
6	119928	A_18_P16717255	true	0.204966723918...
7	110684	A_18_P10000017	true	0.205067604780...
8	72691	A_18_P10000019	true	0.204314514994...
9	37826	A_18_P13359727	true	0.204450890421...
10	148351	A_18_P10000021	true	0.204501405358...
11	26346	A_18_P10000023	true	0.204759255051...
12	54740	A_18_P12361799	true	0.204619213938...
13	35648	A_18_P10000026	true	0.204416185617...

Figure 80 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. The fields are:

Index	FeatureNum	ProbeName
gIsPosAndSignif	LogRatioError	PValueLogRatio
gProcessedSignal	rProcessedSignal	gMedianSignal
rMedianSignal	gBGSubSignal	rBGSubSignal
gIsSaturated	rIsSaturated	gIsFeatNonUnifOL
rIsFeatNonUnifOL	gIsBGNonUnifOL	rIsBGNonUnifOL
rIsPosAndSignif	gIsWellAboveBG	rIsWellAboveBG

Probe Methylation Status Setup

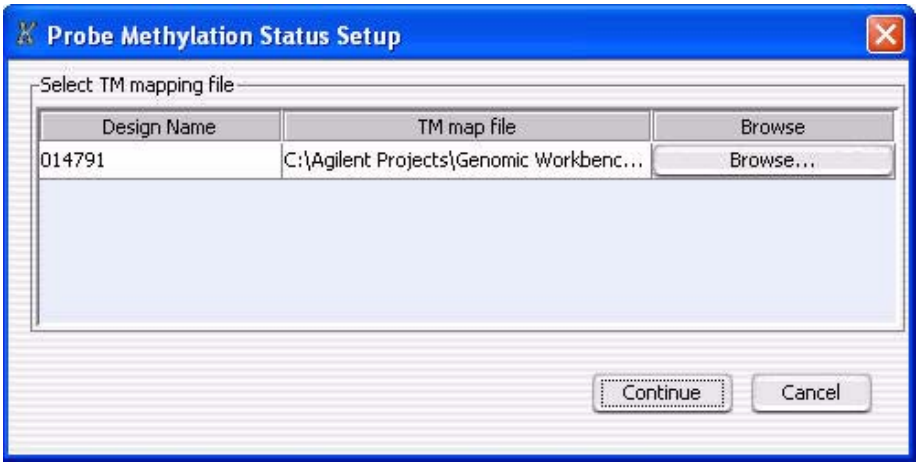


Figure 81 Probe Methylation Status Setup dialog box

Purpose: To associate Tm (melting temperature) map files with the design files for the arrays in the experiment or to make sure the Tm information is already in the design file.

To open: Click **Analysis**, then mark **Apply** for the **Probe Methylation** command

Design Name Shows the names of the design files contained in the experiment

Tm Map File Click **Browse** to find the Tm map file for the design file. This is necessary only if the design is a custom design or if Tm is still not available after an attempted update.

If the design is an Agilent Catalog array, the Tm information is available within the file as long as it has been updated. If the design file has not been updated, the Browse button is active. If it has been updated, the Tm Map File option says Available. If not, update it before continuing.

Continue The methylation algorithm calculates and displays the Z-score results.

Cancel Closes the dialog box without generating a report.

Probe Methylation Report Dialog

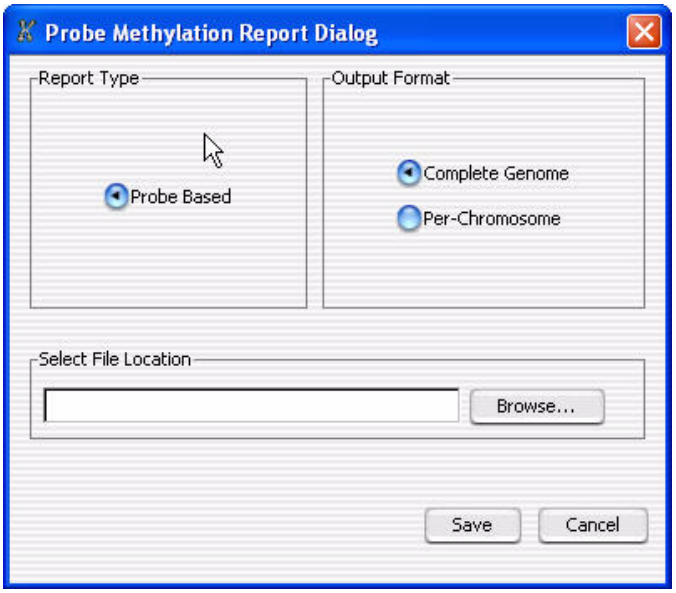


Figure 82 Probe Methylation Report dialog box

Purpose: To select whether the results are reported for the complete genome or for individual chromosomes. See "Report Format" on page 239.

To open: Click **Reports**, then click **Generate Probe Report**.

Report Type All methylation reports are probe-based.

Output Format **Complete Genome** – Creates report as .xls file for entire genome

Per-Chromosome – Creates report for each chromosome as .txt file. Open in spreadsheet program to see headers properly aligned.

Select File Location Click **Browse** to select a location for the file, that you must name.

Sample Attributes

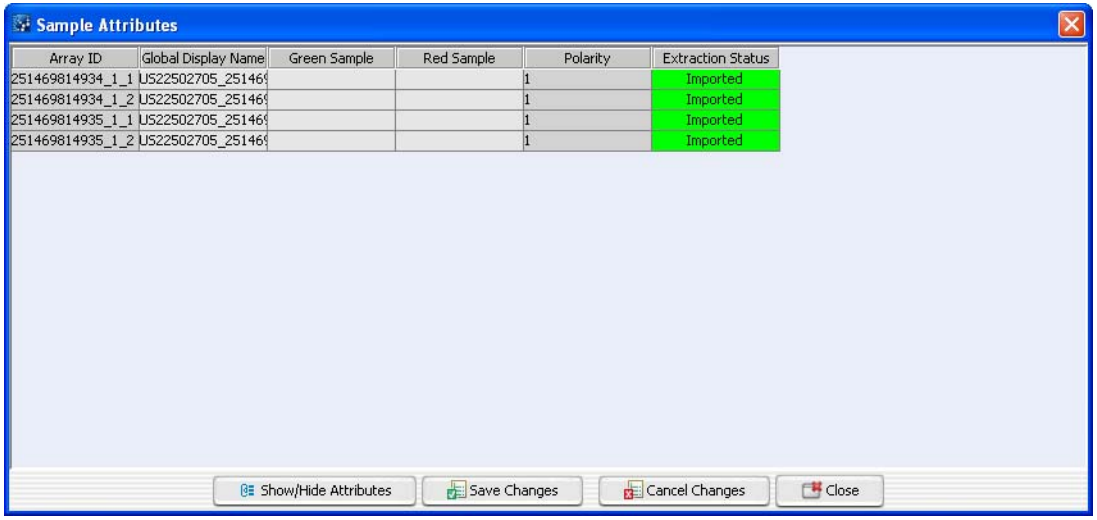


Figure 83 Sample Attributes dialog box

Purpose: To show, hide, or edit array attributes

To open: In the Experiment Pane of the Navigator, right-click the experiment, and click **Sample Attributes**.

This dialog box lets you enter or change the existing values for the attributes listed for the arrays in the experiment. You can also show or hide attribute columns. The columns that appear initially are the default columns (Array ID, Global Display Name, Green Sample, Red Sample, Polarity and Extraction Status) plus any that you selected to show. Changes you make are applied globally. See the *Sample Manager User Guide*.

NOTE Changes to array attributes you make in this table appear also in the Sample Manager table.

Scroll to Column

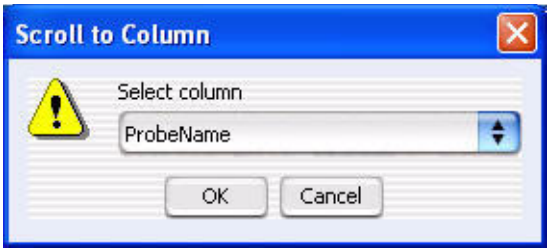


Figure 84 Scroll to Column dialog box

Purpose: This dialog box lets you select a 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.

Select column Displays the columns available in the selected tab. Select the one you want to display.

OK Scrolls the current tab to show the selected column.

Cancel Closes the dialog box.

Search Probes in eArray

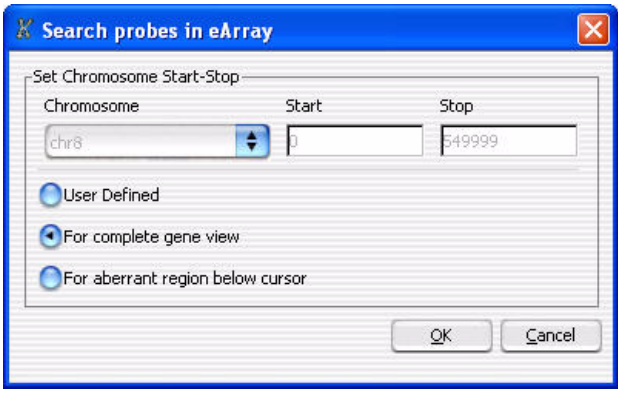


Figure 85 Search probes in eArray dialog box

Purpose: To select the probes you want to update in eArray

To open: Right-click Gene View, and click **Search probes in eArray**.

Select a chromosome and a region in Chromosome View for selecting the probes related to the genes in this region.

User Defined Select to choose the region to search for probes in eArray. The chromosome selection list and the Start and Stop positions on the Y axis are activated when this option is selected.

For complete gene view All the probes related to the genes in Gene View are searched.

For aberrant region below cursor Selects those probes for the genes that appear just below where the cursor sits in Gene View.

Chromosome If you select User Defined, you can select a different chromosome than had been selected before opening this dialog box.

Start/Stop If you select User Defined, type Start and Stop positions for defining the region for the genes in the list.

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 on an array in an experiment, click **Edit Array Color** and click the **Swatches, HSB,** or **RGB** tab.

Swatches tab

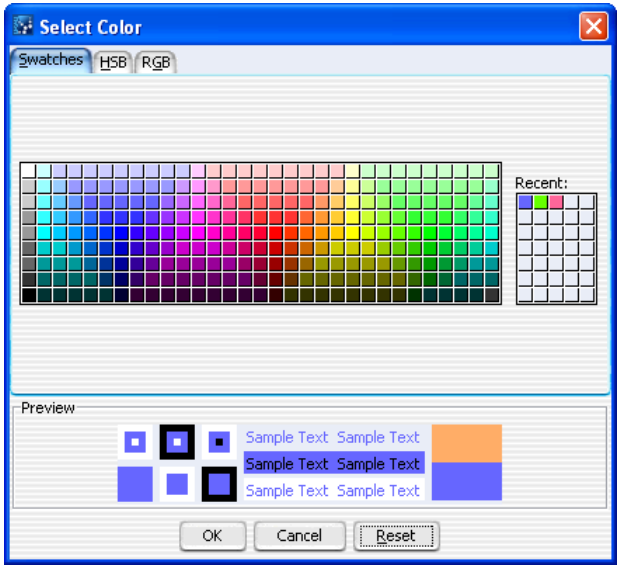


Figure 86 Select Color - Swatches tab

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: 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

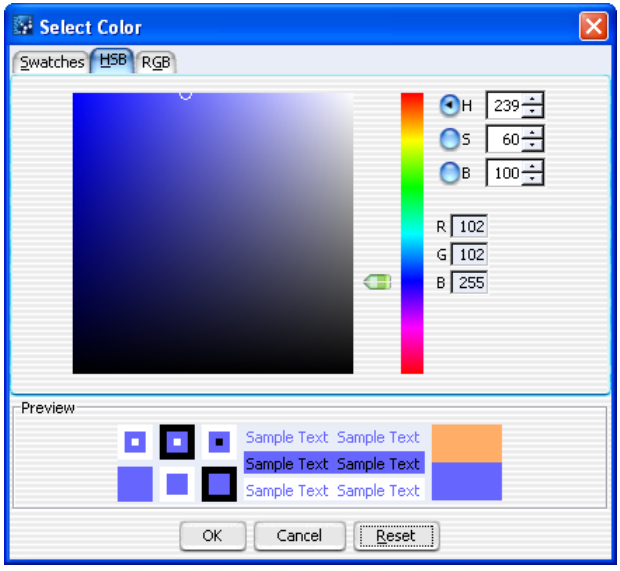


Figure 87 Select Color - HSB Tab

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

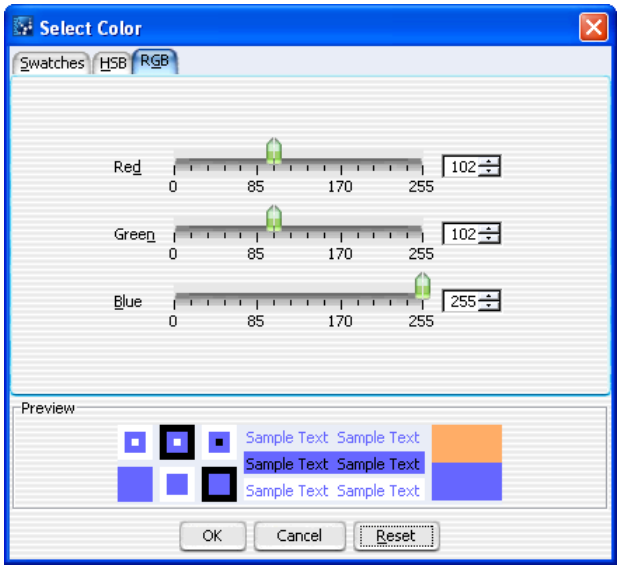


Figure 88 Select Color - RGB Tab

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 data type for experiments

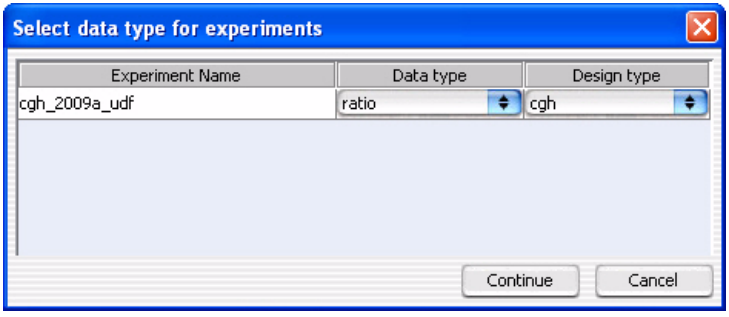


Figure 89 Select data type for experiments dialog box

Purpose: Lets you select the mathematical form of the data in an imported UDF file, and its associated application type. See "To import a UDF file" on page 42.

To open: In the **Home** tab, click **Import > Array Files > UDF File.** In the dialog box that appears, select the desired UDF file, then click **Open.**

Experiment Name By default, the experiment name is the name of the imported UDF file. To change the name, double-click it, then edit it as desired.

Data Type Select the mathematical form of the array data in the UDF file. The options are:

- **ratio**

- **$\log_2$ ratio**
- **$\log_{10}$ ratio**
- **ln ratio** (base e)

Design type Select the application type (CH3, CGH, or expression, for example) for the array data in the UDF file.

Continue Accepts your selections, and goes to the next step in the UDF import process.

Cancel Cancels the UDF import.

Set genome build and species for Axon design files

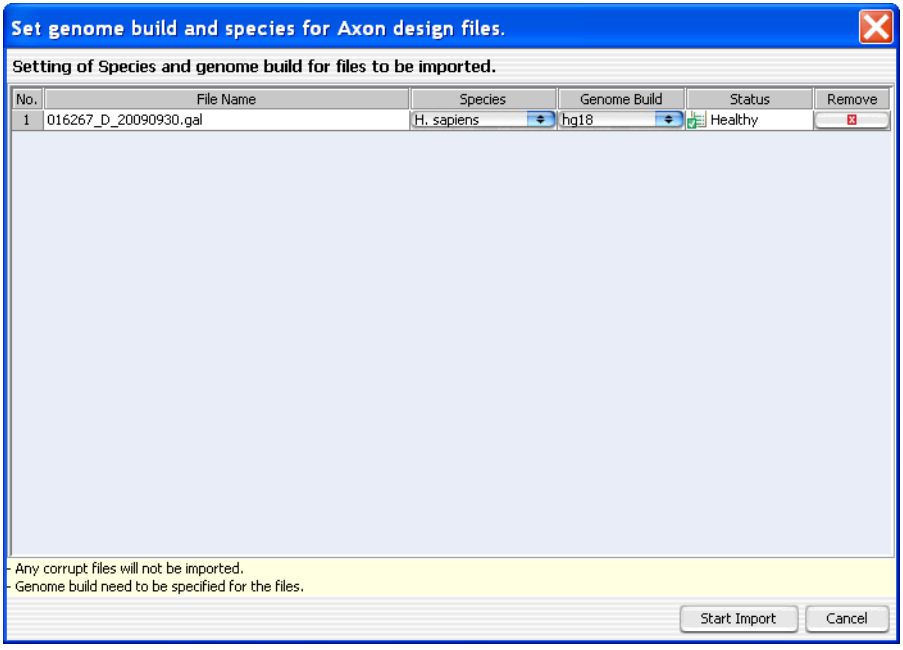


Figure 90 Set genome build and species for Axon design files dialog box

Purpose: Lets you set the species and genome builds for imported Axon design file(s), and to remove specific designs files from the import, if necessary. See "To import Axon design files" on page 40.

To open: In the Home tab, click **Import > Design Files > Axon File.** In the dialog box that appears, select at least one Axon design file, then click **Import.**

No. An index number within the dialog box for each Axon file.

File Name The names of each Axon design file selected for import.

Species The species for each design file. If a species is incorrect, select the correct one from the appropriate list.

Genome Build The genome build for each of the design files. If a genome build is incorrect, select the correct one from the appropriate list.

Status The status of the file is one of the following:

- **Valid** – The file is a new file that can be imported.
- **Healthy** – The file passes validation and can be imported.
- **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, or if the design is an Agilent Catalog design.
- **Overwrite** – The file is a valid design file, but when you import it, it will replace an existing design that has the same name.
- **Corrupt** – The file failed validation. When you start the import process, the program ignores the file.

Remove Click  to remove a specific design file from the list. This is useful if you select a design file in error, or if you do not want to overwrite an existing one.

Start Import Imports the file(s) and closes the dialog box.

Cancel Cancels the import and closes the dialog box.

Show/Hide Columns

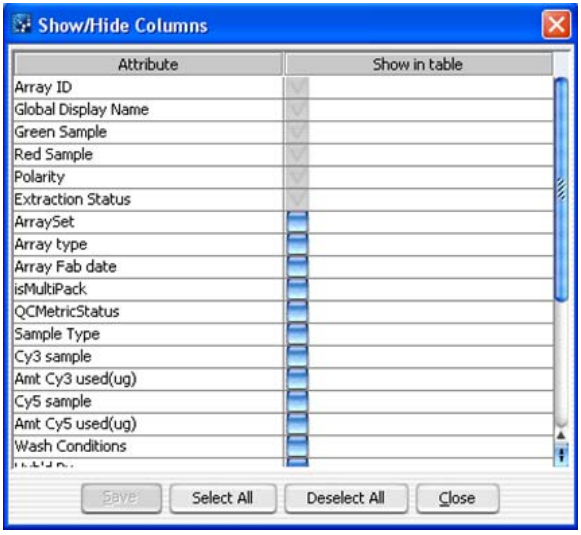


Figure 91 Show/Hide Columns dialog box

Purpose: Used to select the attributes for display in the Sample Attributes dialog box and the Sample Utility tab. The Sample Utility tab is available when you go to Sample Manager. See the *Sample Manager User Guide* for information about Sample Manager.

To open: This dialog box appears when you click **Show/Hide Attributes** at the bottom of the Sample Attributes dialog box.

All available attributes are shown in the Attributes column. Attributes with a check-mark next to them are displayed in the Sample Attributes and Sample Utilities tab for each sample. To select an attribute for display, mark the **Show in Table** box next to it. To hide an attribute, clear the **Show in Table** box.

Save Saves the current list of selected attributes and updates the Sample Utilities table based on the selections.

Select All Selects all the attributes in the list.

Deselect All Clears all check marks from attributes in the list.

Close Closes the dialog box. If changes have been made, the program asks if you want to save your changes before closing.

Track

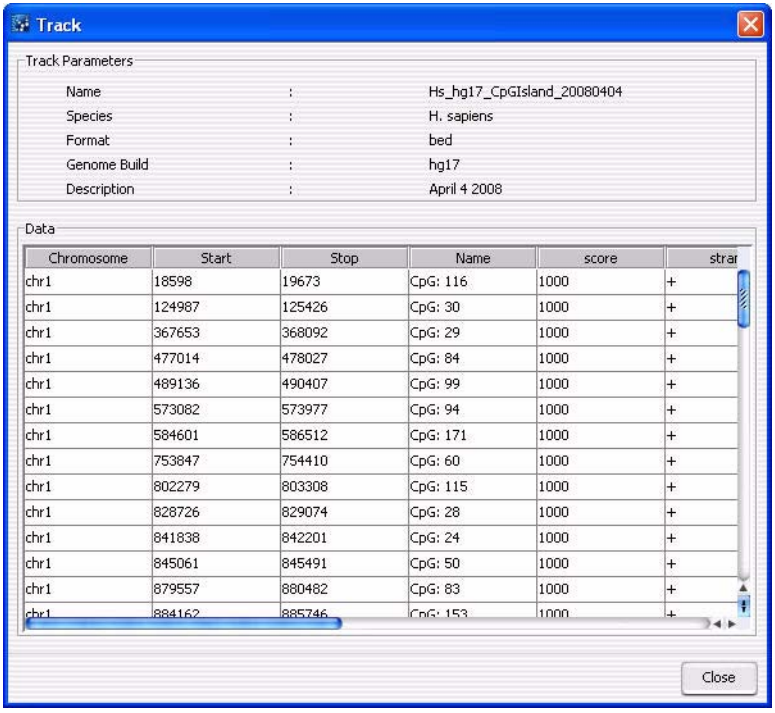


Figure 92 Track details dialog box

Purpose: This dialog box lets you view the chromosome locations in the track.

To open: Click the **Details** link for the desired track in the Tracks tab of the Preferences dialog box. See "Tracks tab" on page 230.

Track Parameters These parameters appear:

Parameter	Description
Name	The name of the track.
Species	The species to which the track applies.
Format	The format of the track data. Agilent Genomic Workbench supports the BED format.
Genome Build	The specific genome build of the species to which the track applies.
Description	Descriptive text saved with the track.

Data Tracks must contain entries for at least these four columns in the table:

Column	Description
Chromosome	The name of the chromosome
Start	The first base pair of the particular feature in the chromosome.
Stop	The last base pair of the particular feature in the chromosome.
Name	The name of the feature. This name appears next to the defined region for the feature.

The other columns are additional BED track file columns that can appear for some tracks. Agilent Genomic Workbench does not display these.

Close Closes the Track dialog box.

UDF Import Summary

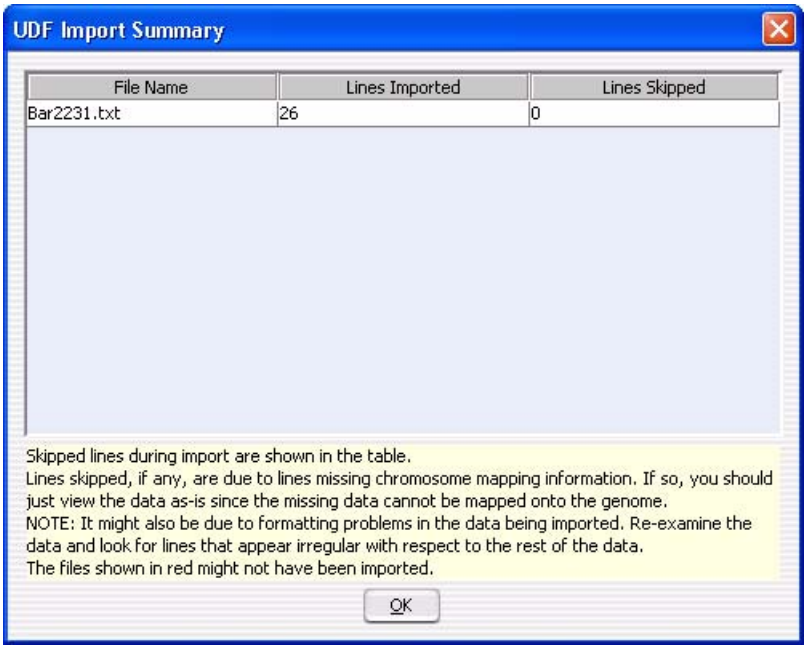


Figure 93 UDF Import Summary dialog box

Purpose: Reports how many lines of data were successfully imported from a UDF file, and how many lines were skipped. Skipped lines are caused by missing chromosome mapping information, or improper formatting of the UDF file.

To open: Import a UDF file (see "To import a UDF file" on page 42). This dialog box appears after you map the columns of the UDF file.

Table Displays the file name of the imported UDF file, the number of lines that were successfully imported, and the number of lines, if any, that were skipped during import. If many lines were skipped, review the data for improper formatting or missing chromosome mapping information.

OK Closes the dialog box.

Universal Data Importer - Map Column Headers

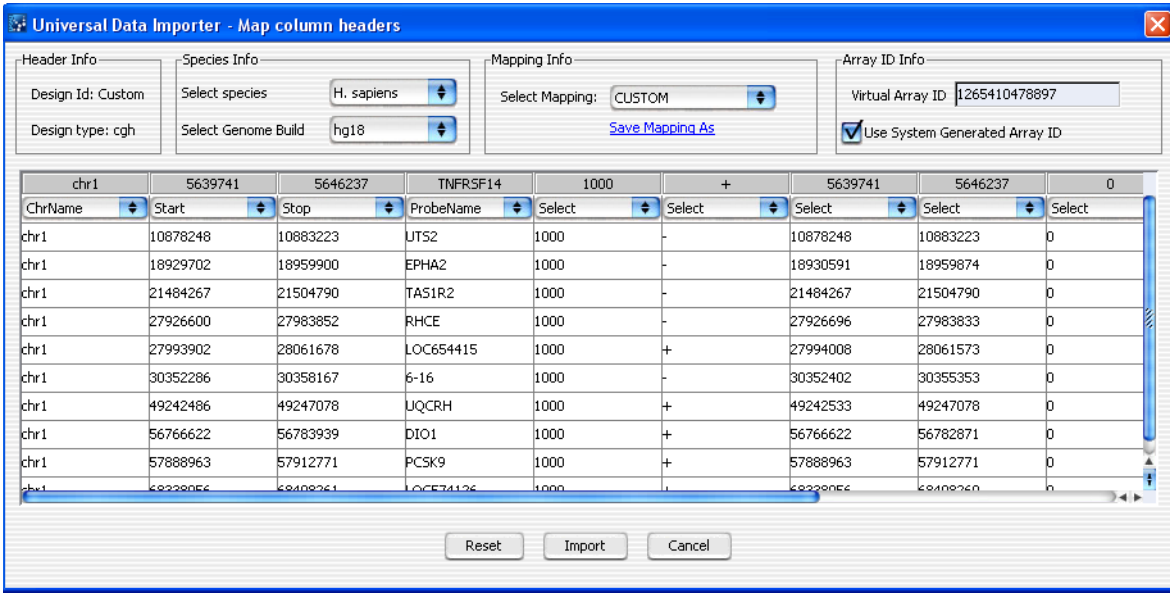


Figure 94 Universal Data Importer - Map Column Headers dialog box

Purpose: Lets you set up a universal data file (UDF) for import. You select several properties for the UDF, and identify the contents of each column of data in the file. You can also save column mappings for re-use.

To open: As you go through the UDF import process (see "To import a UDF file" on page 42), in the Select data type for experiments dialog box, click **Continue.** See "Select data type for experiments" on page 218.

Species Info

Select Species Select the species for the array data in the UDF.

Select Genome Build Sets the species-specific build to use.

Mapping Info

Select Mapping Applies an existing column map to the current UDF. A column map identifies the contents of each column of data. To create a new column map for the current UDF, select **CUSTOM.**

Save Mapping As Saves the column map under a new name. Opens an Input dialog box, where you can type a name for the new map.

Array ID Info

Virtual Array ID A number that uniquely identifies the data in the UDF. Typically, an Agilent microarray slide has a physical barcode that Agilent Genomic Workbench uses to generate an Array ID. The Array ID is used to track the data from the slide as it goes through the steps of an analysis workflow. A "virtual" Array ID is, by default, a system-generated ID that serves the same purpose for data from UDFs. You can also create your own virtual Array ID.

Use System Generated Array ID By default, the virtual barcode assigned to the array data in a UDF is a number that is created by the program. To create your own barcode, clear **Use System Generated Array ID,** then type a new number in **Virtual Array ID.**

Table

This table lets you identify the contents of the columns of data in the UDF. The first row of the table displays the column heading information from the UDF. The second row contains labels that you apply to each column, and the rest of the table displays lines of data from the UDF. If the UDF contains data from Agilent arrays, the column headings will exactly match the labels in the lists.

In the list below each column heading, select the applicable label. You must use each of the labels exactly once, except LogRatio, which you can use more than once. These labels are available:

Label	Description
ProbeName	Names of probes.
ChrName	Names of chromosomes.
Start	First chromosomal location for each probe.

Label	Description
Stop	Last chromosomal location for each probe.
Description	Text annotation for the probe.
LogRatio	Array data values that correspond to each probe. You can use this label more than once.

NOTE If you select an existing column mapping, then change or reset the column labels in the table, the program changes or resets the saved column map as well.

Reset Clears all the column labels in the second row of the table. If you have selected an existing column mapping, this command also clears the labels in the saved map.

Import Imports the UDF file with the specified parameters, and opens the UDF Import Summary dialog box (see "UDF Import Summary" on page 224).

Cancel Cancels the import and closes the dialog box.

Upload Track to UCSC

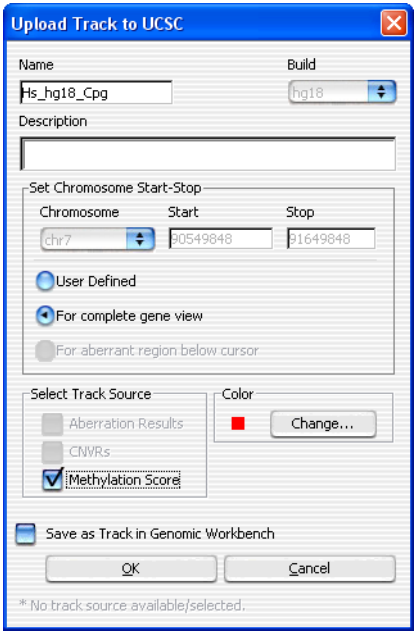


Figure 95 Upload Track to UCSC dialog box

Purpose: Lets you select a track to upload to the UCSC Web site, where you can view it in the UCSC genomic browser.

To open: Right-click in Gene View, and select **Upload Track to UCSC**.

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 used for the track. 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:

- **Methylation Score** – Uses the current methylation score as the source for the track.

Save as Track in Genomic Workbench Saves the selected track in the Tracks folder in the My Entity List pane of the Navigator.

OK Creates the track. To display the track in Gene View, use the **Tracks** tab of the User Preferences dialog box to enable it. See "Tracks tab" on page 230. To export the track, see "To export tracks" on page 73.

Cancel Closes the dialog box without creating a track.

User Preferences

Purpose: This dialog box is used to set up preferences for display of tracks, data storage locations, and licenses.

To open: From the Home tab, click **User Preferences**. Or, right-click in the Gene View, Chromosome View, or Genome View, and click **User Preferences**.

Tracks tab

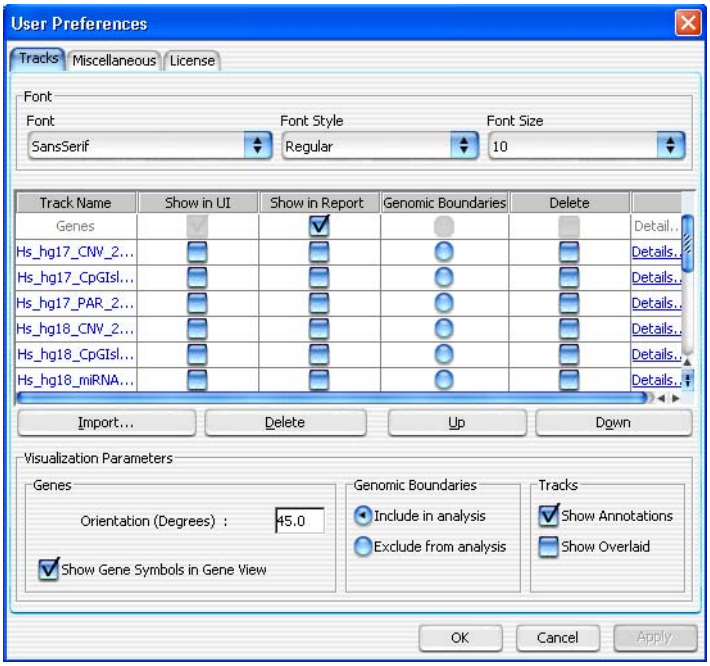


Figure 96 User Preferences dialog box - Tracks tab

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 User Preferences dialog box, click the **Tracks** tab.

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 Mark the check box to display the track next to Gene View.

Show in Report Mark 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 Mark 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 will appear in Gene View and the Track Annotations appear in the tracks.
- Show Gene Symbols – Mark 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 – Mark to show the names of the gene regions for the tracks, and clear to hide them.

- Show Overlaid – Mark to overlay all the tracks that appear next to Gene View, and clear the check box to display the information in separate tracks.

Miscellaneous tab

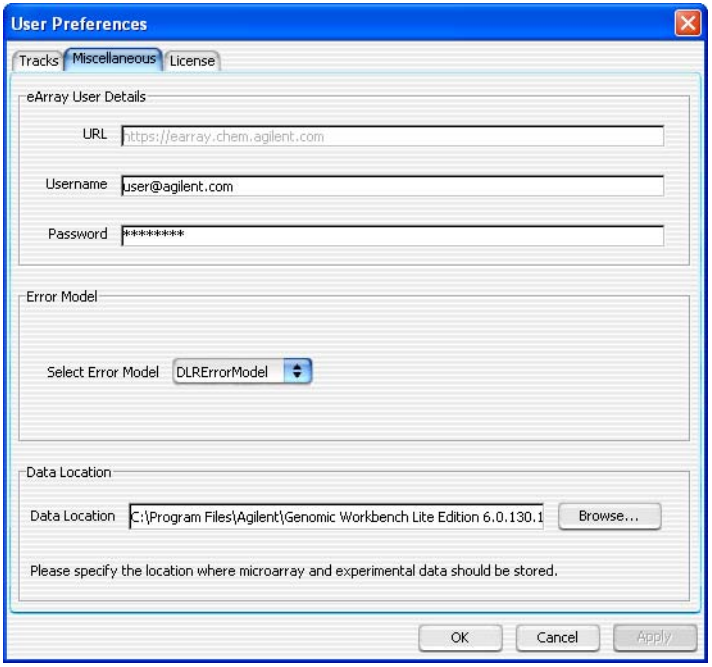


Figure 97 User Preferences dialog box – Miscellaneous tab

Purpose: For data/content set-up, this dialog box allows you to set up eArray access and to change the location for data.

To open: In the User Preferences dialog box, click the **Miscellaneous** tab.

eArray User Details

Sets login details for the Agilent eArray Web site.

- **URL** – At present, https://earray.chem.agilent.com
- **Username** – The name registered on the eArray site.
- **Password** – The password registered on the eArray site.

Error Model (Not available for CH3) The DLRErrorModel (Derivative Log Ratio) measures noise in the data for CGH analyses.

Data Location The folder where the program stores array data and design files. To select a location, click **Browse.**

Apply Applies any changes to the preferences.

OK Accepts any changes and closes the dialog box.

Cancel Cancels all changes and closes the dialog box.

License tab

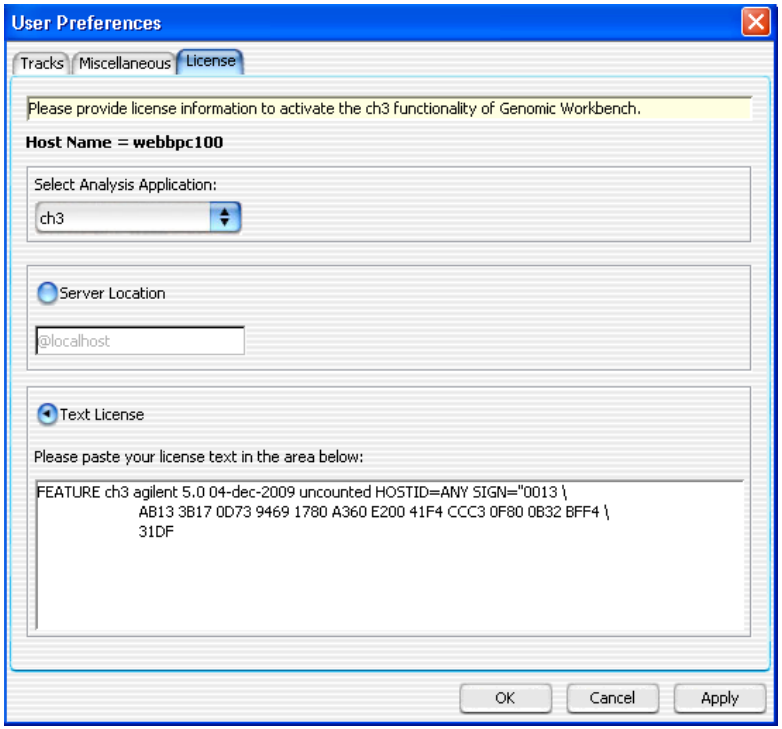


Figure 98 User Preferences dialog box – License tab

Purpose: The License tab allows you to display and update your CH3 application license. This license enables the CH3 application, and allows you to use it to analyze array data.

To open: the In the User Preferences dialog box, click the **License** tab.

Host Name Displays the host computer name automatically.

Select Analysis Application Select the Agilent Genomic Workbench application for which you have a license.

Server Location Select this option if you have a concurrent user license. To edit this name, select **Server Location,** then type the path where your license(s) are located. If you select this option, the Text License option is unavailable.

Text License Select this option if you have an application license (CGH, ChIP, CH3). To change the license, delete the old license text, and paste the new license text in the box.

OK Accepts any changes you have made, and closes the dialog box.

Cancel Closes the dialog box without changing any license information.

Apply Accepts any changes you have made, but does not close the dialog box.

View Preferences

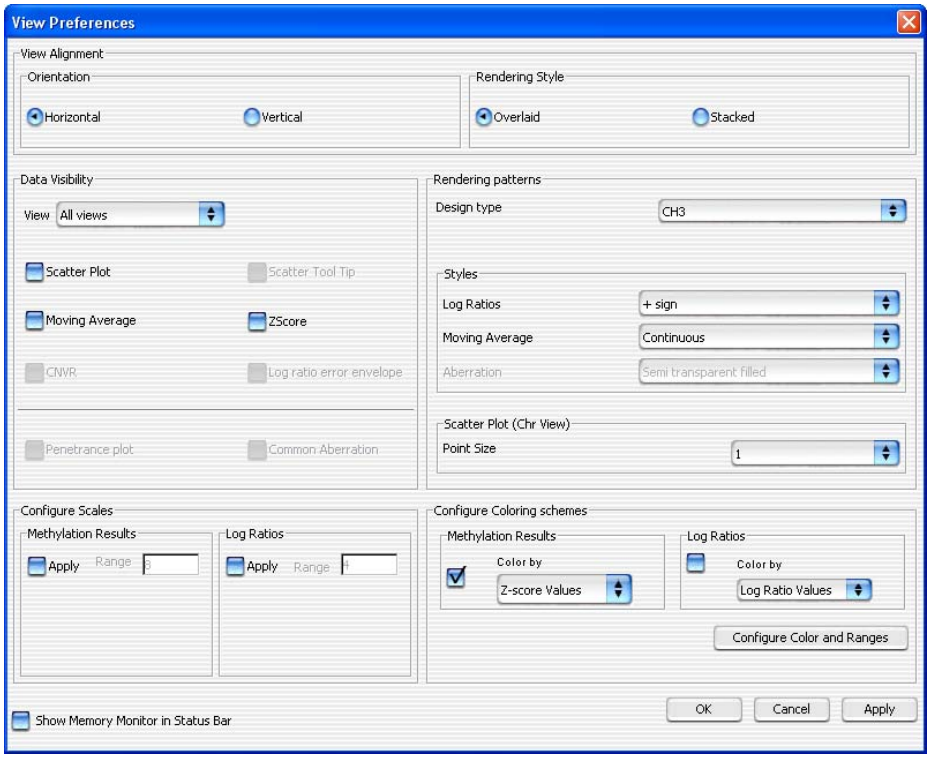


Figure 99 View Preferences dialog box

Purpose: This dialog box allows you to configure how data and results appear in Genome, Chromosome, and Gene views.

To open: In the View tab, click **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. 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.

Mark any of the following options, as available:

Display item	Description/Comments
Scatter Plot	The plot(s) of individual log ratio 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. See "To set up a moving average calculation to smooth the data" on page 126.

Display item	Description/Comments
ZScore	Marking this check box displays the results from the methylation algorithm to produce the statistical likelihood of methylation or no methylation in CpG Island regions of the genome.
Log Ratio Error Envelope	Mark this check box to display the error envelope for log ratio values.

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 array design.

- **Design Type** – Select the type of design to which the patterns are applied: **CH3, Expression, or Other.**
- **Styles** – Select the display style for each of these elements:

Display element	Details
Log Ratios	Select how to display individual data points as: **Color filled circles (ellipses), circles, rectangles, filled rectangles, + signs,** or **x signs.** The latter two choices increase performance.
Moving Average	Select the line style for the moving average display. Lines appear in the display color defined for each array. See "To set up a moving average calculation to smooth the data" on page 126. • **Continuous** – A solid line. • **Dashed** – A dashed line. • **Dotted** – A dotted line. • **Do not show area** – No line.

Scatter Plot (Chr View) Point Size

Select a point size to use for display of scatter plot data points.

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 CH3 scatter plots can also decrease performance. For faster performance, change the rendering style for CH3 data from the filled circle to the plus (+) or cross hair sign.

Configure Scales For Methylation Results or Log Ratios plots, mark **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 ToolTip in the Gene View.

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 164.

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.

Report Format

The Methylation Report contains all the statistical results to help you determine if there are methylated or unmethylated regions in the genome. The report is created in .xls format for the complete genome or in .txt format for individual chromosomes. Use a spreadsheet program to open the report(s).

To learn the meaning of the values in the ZScore columns see "Overview of Methylation Detection and Visualization Algorithms" on page 242.

Methylation Reports contain the following columns, in this order:

Column	Description
CytoBand	Accepted name of the cytoband for each section of chromosome
ChrName	Name of the chromosome (for example, chr17)
ProbeName	Name of the probe (for example, P)
Start	The first base pair of the chromosomal location to which the probe binds
Stop	The last base pair of the chromosomal location to which the probe binds
Description	Name or phrase for the type of probe (for example, promoter)
GeneNames	Names of each gene in whose region probes are located
"Name of CpG Island Track"	Name of each CpG Island in whose region probes are located
Combined ZScore	Combination of both the methylated and unmethylated ZScores. The higher the positive combined ZScore, the more likely the probe is methylated and vice versa.
ZScore_Methylated	Probe ZScore that are contained in right Gaussian. All of these values are positive or small negative values.
ZScore_Unmethylated	Probe ZScore that are contained in left Gaussian. All of these values are negative or small positive values.
logOdds	This number reflects how likely it is that a probe is methylated rather than unmethylated. The higher the positive value, the more likely it is methylated.

Column	Description
logRatio	Log ratio from the extracted FE image file for the probe
Tm	Melting temperature for the probe region

6 Statistical Algorithms

This chapter provides implementation details for the algorithms used in the Methylation (CH3) application of Agilent Genomic Workbench 6.0. The methylation algorithms facilitate the statistical analysis of methylated genomic regions. The first section presents the methylation detection and measurement algorithms. This is followed by a section that describes the visualization options available for methylation analysis.

Overview of Methylation Detection and Visualization Algorithms

The Methylation module of Agilent Genomic Workbench provides algorithms for genomic methylation detection and visualization. The main methylation detection algorithms are described below. Information on algorithms that visualize the data follows.

Methylation detection and measurement algorithms

Methylation of cytosines in DNA is an epigenetic modification that can play a role in the regulation of gene expression. Generally, methylation is associated with repression of gene expression. Much of the aberrant DNA methylation associated with disease is found in CG-rich regions, termed CpG islands.

Agilent catalog methylation (CH3) arrays are specifically designed to assess the methylation pattern within CpG islands and gene promoter regions. The methylation detection algorithm is designed to be compatible only with the Agilent protocol for DNA methylation analysis and the Agilent methods for affinity-based methylated DNA enrichment.[1]

Methylation status detection algorithm

The methylation status detection algorithm is designed for two-color assays, where the green (Cy3) channel is comprised of input DNA, and the red (Cy5) channel is comprised of the affinity-enriched DNA. As a result of constraints imposed on the array design, the probes that target genomic regions with varying CG content do not share a uniform melting temperature (T_m), which can result in compression of the log-ratios.

The methylation detection algorithm allows you to normalize the log-ratios for each probe, based on its T_m, and returns the methylation status of a probe. This sharpens the bimodal distribution observed with log-ratios alone. For more information, see "Methylation status detection algorithm" on page 244.

BATMAN algorithm

The BATMAN (BAyesian Tool for Methylation ANalysis) algorithm[2] has been adapted here for analysis of Agilent methylation microarrays. This algorithm takes into account the local density of CpG dinucleotides in the

genomic vicinity of the interrogating probes, and estimates methylation events using this information. The algorithm is used with log ratio observations and can additionally incorporate the effects of probe melting temperature corrections based on the output *Z*-scores from the methylation status detection algorithm. For more information, see "Bayesian tool for methylation analysis (BATMAN)" on page 254.

Visualization algorithms

Moving average – linear

The moving average algorithm sets a fixed window size around every point of interest. The points are either the probe log-ratio scores determined prior to the methylation detection algorithm, or the combined scores that result from application of the methylation detection algorithm. The program reports the value for each point as the average of that point and neighboring points within the window boundaries. For more information, see "Moving average (linear smoothing)" on page 259.

Moving average – triangular smoothing

The triangular smoothing algorithm is a shaped algorithm based on a moving average. As the moving average centers on a point of interest, the program applies a maximum weight, and that weight falls off with increasing distance. The program adjusts the point of interest as the weighted mean of itself and neighboring points. For more information, see "Triangular smoothing" on page 260.

Methylation Detection and Measurement Algorithms

Agilent Genomic Workbench provides two algorithms for methylation detection and measurement of the methylation levels detected by Methylated DNA Immunoprecipitation (MeDIP). The methylation status detection algorithm uses a melting temperature differential to discriminate between methylated CpG regions and unmethylated regions. The Bayesian tool for methylation analysis (BATMAN) algorithm uses conditional probabilities from a calibration step to estimate methylation levels in CpG regions. The BATMAN algorithm may be used directly on the log-ratios, or it may be used on the resultant *Z*-scores from running the methylation status detection algorithm.

Methylation status detection algorithm

You find epigenetic methylation predominantly in genomic areas of increased GC content, known as CpG islands. Agilent human methylation arrays specifically query approximately 25,000 CpG islands in the human genome, in addition to gene promoter regions. The Agilent methylation protocol uses antibody enrichment for 5-methylcytosine, which results in a log-ratio of the relative abundance of genomic fragments with methylation to those without.

Although the log-ratio scores give information about the relative abundance of the genomic fragments, individual specificities of the probes in each log-ratio score vary by melting temperature. Differences in melting temperature can therefore decrease probe specificity. Normally, the probes on Agilent arrays are chosen in such a way as to minimize probe-to-probe differences in melting temperature.

The algorithm first bins the probes by their melting temperature. Within these bins, you can make direct probe comparisons. When you compare probes by log-ratio scores within each bin, you observe a bimodal distribution — a mode comprised of those probes that were not methylated, and a mode comprised of methylated probes. You also observe a bimodal distribution in per-T_m *Z*-normalized scores.

The algorithm determines the methylation status of a probe in a T_m-dependent manner. To determine which probes are methylated, the binning process compares log-ratios among probes of similar melting temperatures.

To apply parametric statistical techniques, the algorithm fits the bimodal distribution to a set of Gaussian curves, each of which requires only a limited number of parameters to accurately fit the data. The algorithm uses normalized *Z*-scores from the Gaussian distributions to effectively judge the methylation status of a given probe on the array.

Purpose Detection of methylation status is a visualization and report analysis that determines the confidence score for probes that query known CpG islands and gene promoter regions. To determine the probe methylation status, the analysis uses a combination of log-ratio scores from competitive hybridization methods, and the range in melting temperatures among probes located in CpG islands.

Use To establish a link between genomic methylation and gene expression, detection of methylation status is used to identify probes likely to query methylated genomic regions. CpG islands, measured on the Agilent catalog methylation (CH3) design array, generally are either fully methylated or fully unmethylated. You use the algorithm to identify methylation status for probes. You can then visualize the probes in a genomic context — for example, CpG islands within the genome.

Algorithm The algorithm first bins the probes by their melting temperature. For each bin, it applies Gaussian fits using one of three models. It fits the probe log-ratios to Gaussians, using a local searching algorithm called *random hill climbing*.[3] *Z*-scores and *p*-values derived from the Gaussian data give probabilities and confidence values for methylated and unmethylated probe populations. The algorithm then calculates a methylation logOdds, which gives the relative probability that a probe is more likely methylated than unmethylated.

NOTE The logOdds score indicates how likely a probe is to be methylated. It compares *p*-values from the methylated to unmethylated populations. If they are the same, this value is zero.

Step 1: Create a binned distribution of probes

To compare log-ratio scores between genomic fragments enriched by methylation and those without methylation, the algorithm orders probes into discrete melting temperatures.

7 Probes are binned according to their melting temperature. Each bin is 1ºC wide.

8 The algorithm fits the Gaussian distributions to the probe log-ratios within each T_m bin. It does this by further dividing the probe log-ratios into sub-bins (with default size of 0.1 ºC), to generate the observed distribution of signals, as shown in Figure 100.

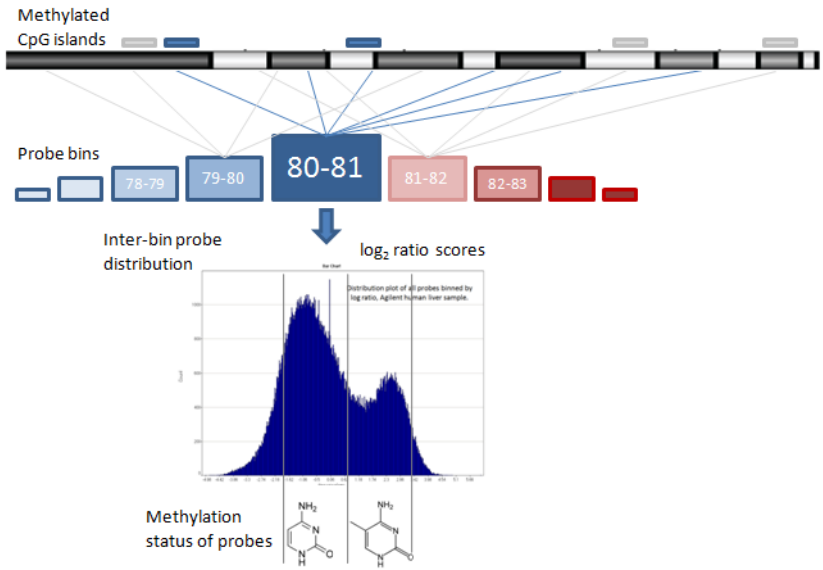


Figure 100 Overview of procedures in the methylation status algorithm

Step 2: Fit Gaussian curves to the binned data

For each bin of 1ºC, the algorithm applies Gaussian fits in one of three models, depending on whether the data points are unimodal or multimodal:

- If the data are unimodal, then it fits one Gaussian distribution to the data.
- If the data are bimodal (the most common fit for methylation data), then it fits two overlapping Gaussian distributions to the data.
- If the data are trimodal, then it fits three overlapping Gaussian distributions to the data.

NOTE

One Gaussian goodness of fit can exceed two Gaussians if, for example, no methylated probes were pulled down during the antibody enrichment process, or if the number of methylation events is exceedingly small. If this occurs, Agilent Genomic Workbench automatically fits one Gaussian to the data, and the message "SINGLE GAUSS" appears in the report. The *Z*-score in the visualization and report is then calculated from the single Gaussian (that is, how likely the probe is to be methylated), instead of a combined *Z*-score (which is reported from models using either two or three Gaussians).

The process to fit a Gaussian curve to the data has two steps:

1 Initial parameter approximation

2 Improvement of the model parameters using a local search procedure called random hill climbing.[3]

Initial parameter approximation

For each Gaussian curve, the parameters to be fit are the mean (μ), standard deviation (σ), and mixture coefficient (α, the total Gauss integral). For each of the three possible models, the algorithm approximates the initial parameters as follows:

- First Gaussian

  For any of the three models, the parameter estimations are done in the following ways:

1 The mean is estimated by the maximum bin.

2 The standard deviation is estimated using the IQR (Inter-Quartile Range) of the distribution, as inferred from the data left of the mean.

3 The mixture coefficient is estimated from the maximum bin height.

- Second Gaussian

If the data points are bimodal or trimodal, the first Gaussian is subtracted from the distribution. The mean, standard deviation, and mixture coefficient are then estimated in the same way as the first Gaussian.

- Third Gaussian

  If the data are not well explained by either a unimodal or bimodal distribution, then a third Gaussian curve can be fit to allow a better estimate of the first and second Gaussian parameters. In this case, the parameter estimates are:

1. The mean is estimated from the mean from the entire bin.
2. The standard deviation is estimated from the entire bin.
3. The mixture coefficient is estimated from the height of the bin used to estimate the third Gaussian mean.

Parameter improvement procedure

Because many probes are generally present in each T_m bin (on the order of thousands), the algorithm uses a straightforward local search procedure called random hill climbing[3] for parameter optimization. This procedure is highly accurate and extremely fast when fitting the Gaussian models to the data.

In each step, the algorithm alters a random parameter. It accepts this alteration if it improves the target fitness function. To converge to a (locally) optimal result, each random step gets shorter with each iteration. This reduction in step is made logarithmically by n, the number of iterations, and a default starting value of 100 (termed LB):

$$\frac{\log(LB+n)}{\log(LB)} \quad (1)$$

The algorithm achieves convergence when a new step does not yield an improvement in the target fitness function. The procedure converges if no change occurs to the fitness value for 1000 iterations, or after the algorithm has performed 50000 iterations.

Target fitness function

The default target function is the Chi-squared goodness of fit. This function is described using an observed distribution $OD_{i=1...N}$, and an estimated distribution function EF (for example, the convex combination of two Gaussians), as follows:

$$\chi^2 = \sum_{i=1}^{N} \frac{OD_i - EF_i{}^2}{EF_i} \quad (2)$$

Step 3: Calculate *Z*-normalization scores

After the algorithm fits the Gaussian curve(s) to the data, it performs parametric statistical analysis. The log-ratios from the T_m-binned probes may differ in value. Agilent Genomic Workbench therefore uses a measure from the Gaussian distributions themselves to normalize the scores. This measure is called a *Z*-score:

$$Z = \frac{log - ratio_{probe} - \mu_{bin}}{\sigma_{bin}} \quad (3)$$

where μ_{bin} is the mean and σ_{bin} is the standard deviation of the Gaussian distribution. The *Z*-score is a measure of the distance of a given probe log-ratio score from the mean of any of the Gaussian curves, given the standard deviation of that curve. The *Z*-score normalizes comparisons of probes from different bins, by taking into account the Gaussian-fit curves of the different bins.

Because the methylation model typically contains two Gaussians, the algorithm calculates *Z*-scores for each one. The left Gaussian represents those probes with a lower log-ratio score, and consists predominantly of unmethylated probes. The left Gaussian is also the major mode, because the majority of probes are generally unmethylated. The right Gaussian represents probes with a high log-ratio score, and consists predominantly of probes enriched for methylation. The algorithm calculates the following *Z*-scores as statistical measures of whether or not a probe is methylated:

1. The *Z*-score derived from the left Gaussian. A negative or small positive value means that the probe is likely to be unmethylated.
2. The *Z*-score derived from the right Gaussian. A positive or a small negative value means that the probe is likely to be methylated.
3. The combined *Z*-score. This is the summation of the left and right Gaussian *Z*-scores. It reflects the location of a probe log-ratio value in relation to the Gaussian distribution(s) of probes with similar T_m. A strong positive value of the combined score means it is methylated, while a strong negative value indicates the probe is unmethylated. See Figure 101 for more information.

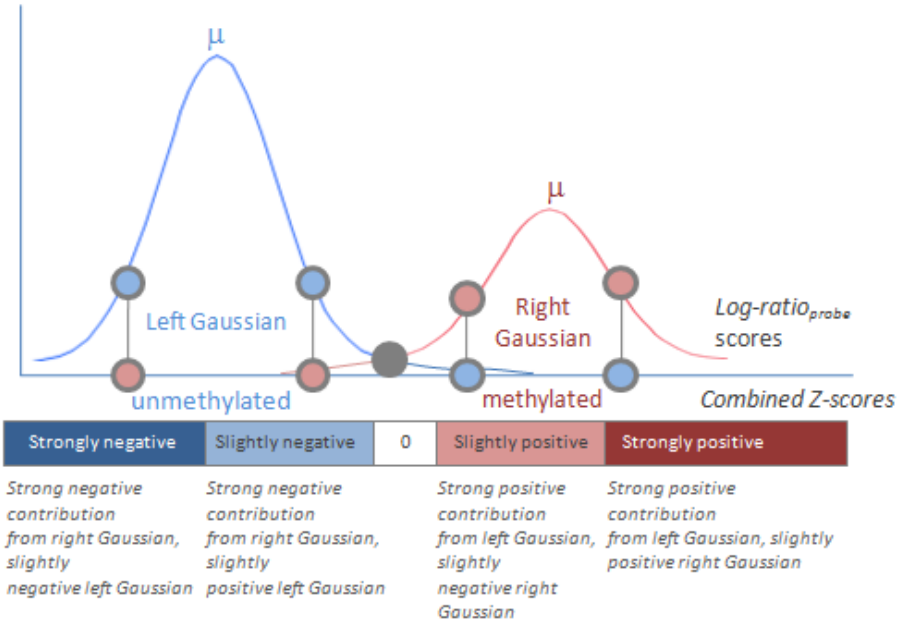


Figure 101 Effect of probe log-ratio score placement on a mixture of Gaussians – result ranges for *Z*-normalization score

In addition to the *Z*-score calculation, Agilent Genomic Workbench calculates a *p*-value that considers how much of a tail continues past the log-ratio point on the Gaussian curve. Such a calculation is important because the tails of the Gaussian distributions overlap; hence it is often difficult to clearly assign a given probe log-ratio score to the left or the right Gaussian if a pronounced valley exists between the two distributions. See Figure 102 for more information.

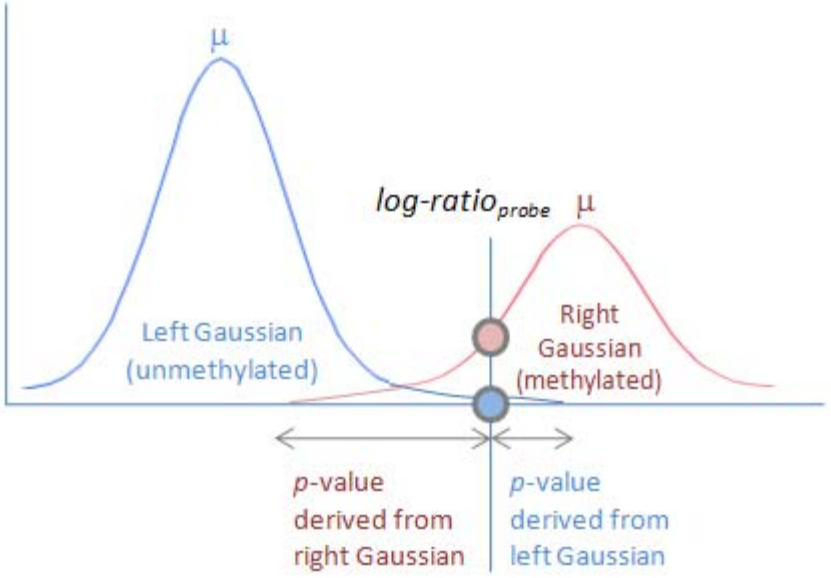


Figure 102 Calculating *Z*-scores and *p*-values from two overlapping Gaussian distributions

The following *p*-values are calculated for the usual case of two Gaussians:

1. The *p*-value derived from the left Gaussian. This is denoted p_M in the report, and is the confidence at which the non-methylation null hypothesis is rejected (that is, the confidence at which you can call the probe methylated).
2. The *p*-value derived from the right Gaussian. This is denoted p_U in the report, and is the confidence at which the methylation null hypothesis is rejected (that is, the confidence at which you can call the probe unmethylated).
3. Finally, the algorithm uses the bimodal distribution of log-ratio scores among probes of similar T_m to calculate a logOdds score of the probe. This final logOdds score reflects the likelihood that a probe is methylated, and is calculated as ω

$$\omega = -\log\left(\frac{p_M}{p_U}\right) \quad (4)$$

Interpreting the results

As shown in Figure 102 above, the boundary to decide whether a probe is methylated often comes from log-ratio scores that fall in a valley between two Gaussians. Although the program can establish a ratio of the contribution of each Gaussian by calculation of the logOdds score ω, a region exists where ω approaches a value of 1, and a methylation status decision cannot be made. In other words, a threshold on the logOdds score decides the methylation status call. This threshold is not fixed, and can be determined after you inspect the output in the context of a project.

Visualization

Agilent Genomic Workbench can display a moving average for both the probe log-ratio scores and the *Z*-normalization scores from the methylation status algorithm. The *Z*-scores used in the moving average visualization are the combined *Z*-scores, and reflect the position of the probe log-ratio score on both Gaussian curves, as shown in Figure 101.

Figure 103 displays the Gene View results of the methylation status algorithm for Chromosome 21. The left pane displays the *Z*-score moving average. There is no defined cut-off value to determine whether a probe or probe region is methylated. The middle pane shows the moving average for the probe logOdds ratios, with genomic tracks annotated. The right pane displays the CpG island track. Note that the probes fall specifically within CpG islands, and upstream of genomic tracks such as coding regions.

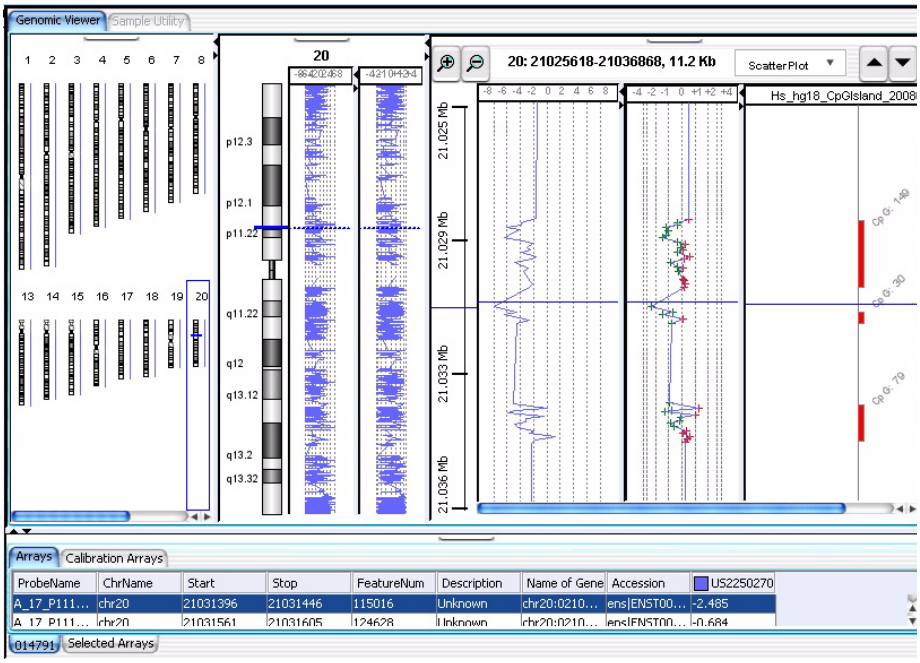


Figure 103 Output of methylation status detection algorithm. On the right side of Agilent Genomic Workbench, Gene View displays the moving average for both the *Z*-score (left pane of Gene View) and the probe log-ratios (middle pane of Gene View). Human CpG islands are displayed as a track along the right side of the moving averages. Note that the probe log-ratios, which are only slightly positive in this region (colored red), may take on negative combined *Z*-scores (colored by sample).

Output report The program generates an output report for each chromosome. The report contains the following statistical measures:

Table 6 Contents of methylation output report

Reported parameter	Meaning
Z-score unmethylated	*Z*-score derived from the left Gaussian. This is the probability that the observed value for a probe comes from the unmethylated population.
Z-score methylated	*Z*-score derived from the right Gaussian. This is the probability that the observed value for a probe comes from the methylated population.
Combined *Z*-score	Summation of the methylated and unmethylated *Z*-scores
p_M	The *p*-value derived from the left Gaussian. This is the confidence at which the non-methylated null hypothesis is rejected (that is, the confidence at which you can call the probe methylated).
p_U	The *p*-value derived from the right Gaussian. This is the confidence at which the methylated null hypothesis is rejected (that is, the confidence at which you can call the probe unmethylated).
logOdds	The logOdds score ω the likelihood that a probe is methylated rather than unmethylated.

Bayesian tool for methylation analysis (BATMAN)

Epigenetic methylation is found predominantly in genomic areas of increased GC content, known as CpG islands. Agilent human methylation arrays specifically query approximately 25,000 CpG islands in the human genome, in addition to gene promoter regions. Like many MeDIP protocols, the Agilent methylation protocol uses antibody enrichment for 5-methylcytosine, which results in a log-ratio of the relative abundance of genomic fragments with methylation to those without.

Although the log-ratio scores give information about the relative abundance of the genomic fragments, the scores do not reflect the underlying distribution of CpG dinucleotides, the abundance of which can increase fragment enrichment. Differences in CpG dinucleotide frequencies

can therefore confound methylation state quantitation because the signal log-ratios themselves depend upon the density of the methylated CpG sites.

To obtain an absolute measure of methylation, the BATMAN algorithm first calibrates a set of conditional probabilities based upon the distance of a probe from any CpG sites. Based on expected fragment sizes during DNA shearing, the total number of CpG sites per probe can be estimated, yielding a Gaussian distribution of methylated sites to give the observed probe signal. This conditional probability is then used as a likelihood to generate a Bayesian posterior probability of the methylation status of any CpG site given the probe signal. The signal used in BATMAN can be either the probe log ratio or the *Z*-score from the methylation detection algorithm.

In general, probes chosen for MeDIP- microarray experiments are limited by their genomic location and therefore have a wider range of melting temperatures. The methylation detection algorithm accounts for this by binning the probes according to their melting temperature and standardizing the probe scores. Agilent Genomic Workbench therefore allows you to combine the output from the methylation detection algorithm with the BATMAN algorithm, to account for the differences in probe melting temperatures and the effects of inhomogeneous CpG densities.

To increase the speed of the overall calculation, BATMAN uses the assumption that the methylation state of neighboring CpG dinucleotides is not independent, but rather homogeneous due to the effect of methylation spreading and maintenance. A Monte-Carlo simulation is then used to sample windows of fixed genomic size to establish methylation levels for all CpG sites.

Purpose BATMAN is a visualization and report analysis that determines the absolute methylation level for CpG dinucleotides that are contained within DNA fragments queried by probes across the microarray. These probes typically capture known CpG islands and gene promoter regions. To determine the absolute methylation level of individual CpG sites, the analysis uses a calibration method to arrive at a distribution that describes the conditional probability of the probe signal given a set of methylation states. Bayesian inference is then used to obtain the methylation states given the probe signal.

Use To establish a link between genomic methylation and gene expression, or to observe epigenetic changes, Agilent Genomic Workbench uses the BATMAN algorithm to quantify the methylation level of CpG dinucleotides across probes that query methylated genomic regions. You can then visualize the probes in a genomic context – for example, CpG islands within the genome, or individual CpG dinucleotides within CG-sparse regions.

Algorithm The algorithm begins with a calibration step, which uses a regression to fit a trend between local CpG density and either probe log-ratios, or *Z*-scores if the methylation status detection algorithm is used. This is followed by a modeling step, which uses the calibration results to construct the conditional probability of the log-ratios (or *Z*-scores) given the methylation state. A sampling step occurs to generate a large number of possible methylation states. Bayesian methodology is used to invert the conditional probability and summarize the methylation state per probe. Finally, a methylation call step fits a distribution to the scores on a per-chromosome basis. Additionally, for each CpG island, genomic intervals are found with the same methylation call.

Step 1: Calibration step

As described in Down *et al.*,[2] a coupling factor C_{cp} is defined as:

$$C_{cp} = \frac{m}{n}$$

where c is any given CpG dinucleotide, n is the number of fragments containing c, and m is the total number of DNA fragments hybridizing to any probe p. A linear regression is then fit to the plot of the log-ratios (or *Z*-scores) versus C_{tot}, where:

$$C_{tot} = \sum_{p} C_{cp}$$

Since most CpG-rich areas are hypomethylated while most CpG-poor regions are hypermethylated, the regression fit is increased by limiting the least-squares method to the low-CpG regions. The least square estimates are given by:

$$\hat{\beta}_1 = \frac{\sum(x - \bar{x})(y_i - \bar{y})}{\sum(xi - \bar{x})^2}$$

and:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

The variance of the predicted response (that is, the expected range of values of y at 95% confidence) is given by:

$$Var(y_d - [\hat{\alpha} + \hat{\beta} x_d]) = \sigma^2 \left(1 + \frac{1}{m} + \frac{(x_d - \bar{x})^2}{\sum(x_d - \bar{x})^2} \right)$$

This value is then used in the modeling step along with the regression line slope and intercept.

Step 2: Modeling and sampling

This step uses the data from the calibration results to create a likelihood model. The MeDIP model is given by:

$$f(A|m) = \prod_p \mathcal{F}\left(A_p | A_{base} + r \sum_c C_{cp} M_c,\ v^{-1} \right)$$

Where A_p is the log ratio (or Z-score) for the probe, A_{base} is the baseline value given by the linear regression, and M_c, the methylation state at c, is regarded as a continuous variable with a Gaussian distribution.

Possible combinations of methylation steps are sampled from the MeDIP model using nested sampling.[4]

Step 3: Summarization (Bayesian inference)

The sampling step results are then summarized to create a score for each probe. Each of the sampling states are weighted by the likelihood estimate from the model. The IQR and median values are calculated across samples and for every probe in every array.

The responses are then trimmed to the regions of interested (for example, track boundaries).

Step 4: Methylation calling

A beta distribution is fit to the scores per chromosome obtained in the summarization step. (See "Methylation calling – the beta distribution" on page 263 for more information). Agilent Genomic Workbench then calculates the modes of the distribution and compares the score to the mode, assigning methylated / status calls (+1 for methylated, -1 for unmethylated). If the modes are not unique, then zero is returned as a call.

The beta distribution is fit using a local searching algorithm called *random hill climbing*.[3]

Interpreting the results

For each CpG island, Agilent Genomic Workbench displays consecutive regions that share the same call across each chromosome.

Visualization

The consecutive regions appear as intervals for inspection.

Visualization Algorithms

After you apply the methylation algorithm, it is useful to visualize the general trend of the pre-algorithm probe log-ratios and post-algorithm combined *Z*-scores, using a moving average line plot. To facilitate the visualization of large datasets, Agilent Genomic Workbench employs a sliding window across the genome, to smooth the data points for detailed examination. The smoothing functions are available independently from the main methylation algorithm. A change in window size in the visualization algorithms does not affect the methylation calls.

To reduce the complexity of the data from probe-to-probe noise, you can apply one of two smoothing algorithms. One algorithm applies linear smoothing (the moving average algorithm), and the other applies a triangular smoothing function. These algorithms use a sliding window of fixed size, set in the user interface (UI). You can apply them to the log ratios, the *Z*-score output from the methylation algorithm, or both.

The moving average visualization algorithm sets a fixed window size around every point of interest. The algorithm reports the value for that point as the average of that point and neighboring points within the window boundaries.

Moving average (linear smoothing)

To compute a moving average, the program averages probe log-ratios or combined *Z*-scores 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).

Triangular smoothing

The triangular smoothing algorithm is a shaped smoothing algorithm based on a moving average. As the moving average centers on a point of interest, the program applies a maximum weight, and that weight falls off with increasing distance. Then the program adjusts the point of interest as the weighted mean of itself and neighboring points.

Purpose When you visualize or analyze array data, it is common to smooth the data using a moving average. However, the moving average approach is not the optimal means to reduce noise associated with each independent point, because it can minimize log-ratio changes or methylation *Z*-scores and obscure individual points. Triangular smoothing is a good compromise that reduces noise of individual points, while remaining sensitive to true localized or small-scale variations in the data.

Use The smoothing functions are used for visualization purposes only, and do not affect the methylation algorithm. However, the program applies the moving average (*Z*-score) after the methylation algorithm, to smooth the methylation *Z*-scores that you visualize in the user interface.

Algorithm The number of neighboring points used for smoothing depends on the type of moving average. If you select point input (pt input), the program keeps constant the number of points that are averaged. It may use, for example, 3, 5, 7, 9, or 11 points, and it gives each point equal weight. Alternatively, you can select a window of constant width (in Mb or Kb). The window moves across the data and centers on the point of interest. All points within its range are averaged to yield the moving average value for the point.

NOTE In linear smoothing, the user-defined point input (pt input) uses the same number of probes as in triangular smoothing. However, an equal number of probes is taken in each direction (to the left and right of the center probe) to compute the moving average. This might not be ideal if the probes are placed at varying distances.

Two potential problems exist in linear smoothing:

1 Use of a fixed window width (by choice of a base pair from the UI) causes a variable number of probes to be averaged at each smoothed point. Therefore, the degree of averaging varies from probe to probe, depending on how many probes are in the fixed-width window. Because

varying numbers of measurements contribute to each smoothed point, the degree of statistical noise reduction also varies for each point. This can complicate the error analysis.

2 For so-called “zoom-in” arrays, where some genomic regions of interest are covered more densely than neighboring regions, the appropriate window size can vary greatly between the different genomic regions. Smoothing windows that are appropriate for sparsely tiled regions obliterate all structure in densely tiled regions. Windows appropriate for densely tiled regions perform practically no averaging at all in sparsely tiled regions.

Triangular smoothing avoids these problems, by use of smoothing windows that contain a fixed number of probes, regardless of the total range of sequences those probes span. This respects the fact that nearby probes are more relevant than distant probes to the average at any point.

The concept is illustrated in Figure 104. Fixed-count smoothing includes the same number of points in each average, but weights probes far from the averaged point as much as points near it. In triangular smoothing with pt input, the program enlarges a symmetric window around the averaged point until it contains the number of points chosen for the fixed-size window. These points may be on one or both sides of the averaged point, depending on the probe density around the averaged point. These points are then weighted appropriately for the triangular smoothing function, depending on their distance from the averaged point.

When you enter a number of points from the user interface, the program uses a variable window width, chosen to be the smallest window that is symmetrical about the averaged point and includes the specified number of points. Figure 104 illustrates the application of this method for pt input.

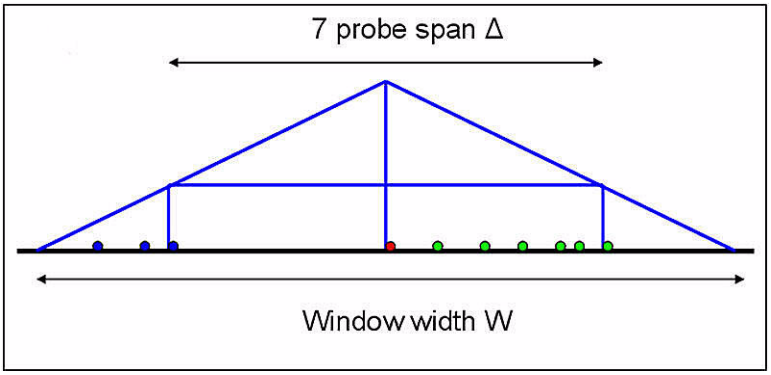


Figure 104 Triangular smoothing with pt input

The program applies smoothing to a region of varying probe density. The effective width of the smoothing window, *W*, depends on the length of the smallest symmetrical region (Δ) that includes the specified number of points. The weight given to each point is proportional to the height of the triangle at that point.

Weights assigned to the log-ratio or combined *Z*-score values of these probes are given by the equation:

$$w(x) = (W - |x|) / W^2 \quad \textbf{(1)}$$

where the effective window width, *W*, is determined by the length, Δ, of the symmetrical region that includes the specified number of points:

$$W = \Delta / (2 - \sqrt{2}) \quad \textbf{(2)}$$

Interpretation The smoothing algorithms affect the scatter plot in the Chromosome and Gene Views. See Figure 103 for more information.

Visualization Agilent Genomic Workbench can display scores from the methylation algorithms in the UCSC genome browser for visual interpretation.

Appendix

This section contains additional information about statistical algorithms and is useful to understand algorithm steps in detail.

Methylation calling – the beta distribution

The beta distribution is given by:

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\ \beta)}$$

Where B in the denominator is an incomplete beta function given by:

$$B(x,\ y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

And the gamma function is:

$$\Gamma(z) = \int_0^\infty (t^{z-1}e^{-1})dt$$

The beta distribution is shown in Figure 105.

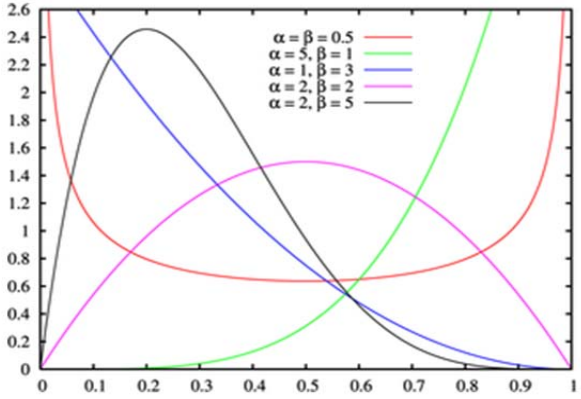


Figure 105 The beta distribution, illustrated with various parameter values

References

1 "Agilent Microarray Analysis of Methylated DNA Immunoprecipitation - Protocol, version 1.0" *Agilent Life Sciences and Chemical Analysis Protocols*. © 2008. <http://www.chem.agilent.com/Library/usermanuals/Public/G4170-90012_Methylation_Protocol.pdf>.

2 Down, T. A., Rakyan, V. K., *et al.* (2008). "A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis." *Nat Biotechnol* 26(7):779-785.

3 Russell, Stuart J., Norvig, Peter (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.), Upper Saddle River, NJ: Prentice Hall, pp. 111-114, ISBN 0-13-790395-2.

4 http://www.inference.phy.cam.ac.uk/bayesys/nest.pdf

www.agilent.com

In this book

This guide describes how to use the Methylation (CH3) application of Agilent Genomic Workbench Lite Edition 6.0 to apply algorithms that help identify methylated regions.

Revision A, May 2010

G3800-90017

Agilent Technologies