Agilent Genomic Workbench 7.0

ChIP Interactive Analysis

User Guide

Agilent Technologies

Notices

Manual Part Number

G3800-90043

Edition

Revision A, December 2015

Printed in USA

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

Trademarks

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

Adobe®, Acrobat® and 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This guide is valid for 7.0 and later revisions of the Agilent Genomic Workbench software, until superseded.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 contains information to use the interactive analysis portion of the Agilent Genomic Workbench 7.0 ChIP program. See the *Workflow User Guide* to learn how to set up ChIP analysis methods for running workflows.

1 Getting Started

This chapter gives an overview of how to use the ChIP application to import, analyze, and create reports for chromatin immunoprecipitation (ChIP) data.

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

This chapter describes how to import, organize, manage, and export ChIP data and other content within the user interface of Agilent Genomic Workbench 7.0.

3 Displaying ChIP 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 modify the display to visualize the data and content the way you prefer.

4 Setting Up ChIP Interactive Analysis

This chapter gives instructions on how to activate and change options to prepare extracted data for ChIP analysis, analyze the data and set up probe, gene and QC reports.

5 ChIP Interactive Analysis Reference

This chapter describes the commands, shortcut menus, dialog boxes, and tabs of the interactive analysis portion of the ChIP application. A second section describes report output formats.

6 ChIP Statistical Algorithms

This chapter describes implementation details of the algorithms used in the ChIP application of Agilent Genomic Workbench. The ChIP algorithms facilitate the statistical analysis of ChIP-on-chip experiments, for example, detected genomic regions with bound protein.

Contents

1
Getting Started

This chapter gives an overview of how to use Agilent Genomic Workbench 7.0 to import, display, and analyze chromatin immunoprecipitation (ChIP) data.

Before or after you import Feature Extraction 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 ChIP data, you organize the data files into logical units called *experiments*. Experiments are used to select the data you want to display or analyze using Agilent Genomic Workbench. After you create an experiment, and add array data, you can then use the Preprocessing, Analysis and Reports tabs of the program to analyze the data in the experiment for binding events.

Experiments can also serve as the basis for automated, unattended ChIP analyses in the Workflow tab. The commands under this tab also let you set up image files for automated, unattended Feature Extraction before array analysis. See the *Workflow User Guide* for information.

What is the ChIP Application?

The ChIP application contains a complete set of features that lets you analyze chromatin immunoprecipitation (ChIP) microarray data. ChIP microarray analysis can identify the genomic loci that contain proteins that bind to DNA, including individual transcription factors, chromatin modifiers, and components of the general transcription machinery.

With ChIP interactive analysis, you set up an experiment and apply analysis algorithms to your data one experiment at a time. You also set parameters for some report templates for use in Workflow analysis.

You do most of your operations with the Preprocessing, Analysis, and Reports tabs.

See the sequence by which you analyze data, do calculations on results and make reports, "General Instructions for Setting Options for ChIP Interactive Analysis" on page 36.

With the interactive portion of the ChIP application, you can:

- Import data from the Agilent Feature Extraction programs.
- Use an intuitive graphical interface to display data and annotations in the context of an organism's genome, at several simultaneous levels of detail.
- Normalize your data with several statistical methods.
- Use error modeling and event detection algorithms to identify probes, genes, and genomic loci that have significant binding.
- Display significant binding events graphically within the program, and export report files that you can analyze further with other programs.

With workflow analysis, you configure an analysis method ahead of time, and then run it in a workflow. This method is useful for unattended operation and for consistent analyses of multiple data sets. To review workflow results, you use the Genomic Viewer. For more information and quick–start instructions to set up and run a ChIP workflow, see the *Workflow User Guide*.

Using Agilent Genomic Workbench on a Mac

The content of this User Guide applies to both the Windows and Mac versions of Agilent Genomic Workbench. 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 the License and Starting the ChIP Application

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.

NOTE

If a message appears about performance when analyzing 1M feature arrays, read the message, mark the **Please do not show this warning again** box, and click **OK**. Instructions for how to increase heap memory are located in the *Agilent Genomic Workbench 7.0 Installation Guide*.

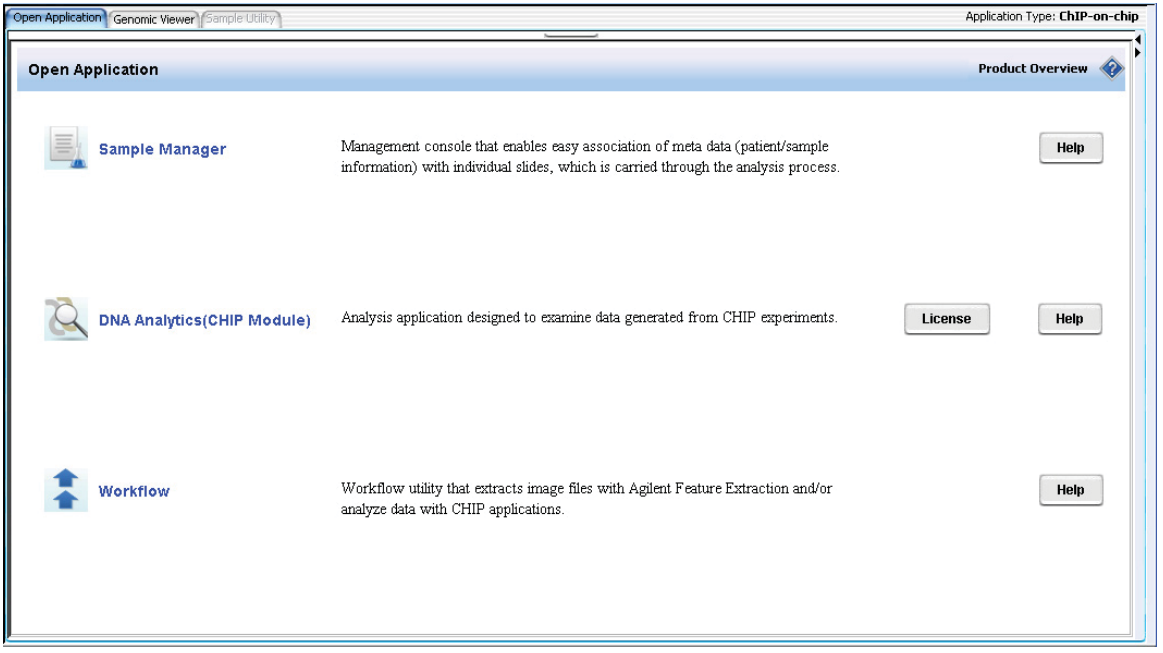

Figure 1 Open Application tab for Agilent Genomic Workbench ChIP Application

To enter your license for analyzing ChIP data interactively

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

2 Click **License** next to the description of DNA Analytics in the Open Application pane.

The License tab of the User Preferences dialog box appears.

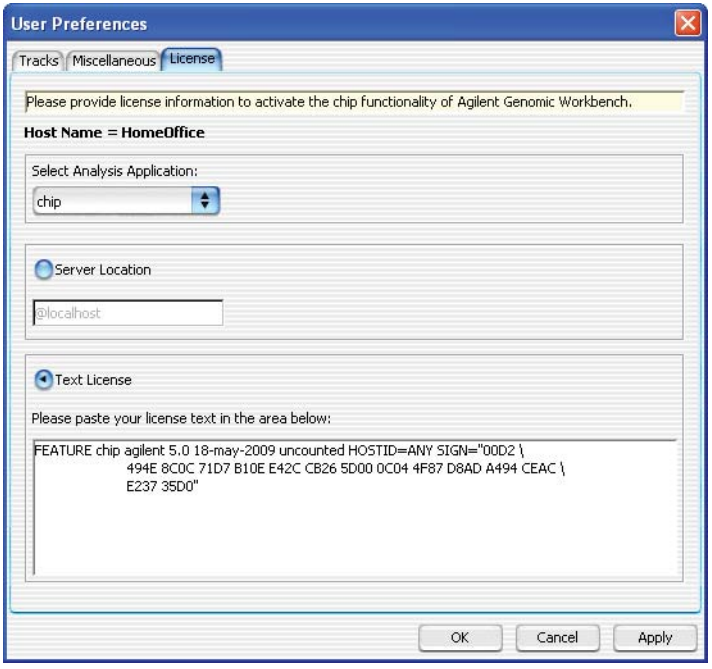

Figure 2 License tab for ChIP application

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 box 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 ChIP application license .txt file.

2 Double-click the license name to open the file in Notepad, and copy the text in the Notepad window.

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 ChIP application

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

  The ChIP 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 with or without a license to display data for many types of arrays, including CGH, ChIP, and Methylation (CH3). You can use the data *analysis* capability in Agilent Genomic Workbench 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 7.0 main window showing major components for ChIP

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), or SureSelect Target Enrichment	**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
Show properties, rename, or delete designs and data in the program	**Data pane of the Navigator:** Double-click to expand data node, then right-click to select option.
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 ChIP 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 ChIP 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 ChIP Interactive Analysis" for information 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 ChIP 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 ChIP 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 ChIP 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 reports. For more information on what you can do in these tabs, see "Tabs" on page 25.

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.

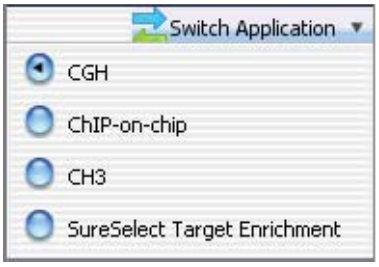

Figure 4 Switch application menu

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 ChIP interactive analysis

The following table summarizes what you can do from the ChIP interactive analysis tabs of Agilent Genomic Workbench.

Table 2 ChIP Interactive Analysis capabilities

Tabs	ChIP capabilities
Preprocessing	Select/edit normalization calculation
	Select/edit error model
	Combine designs and replicates
	Display QC metrics
Analysis	Select/edit/apply event detection model
Reports	Probe report
	Gene report
	QC report
View	Set up preferences for display of data
	Copy displayed data to the Clipboard
	Turn on or off display of views and Navigator
	Turn on or off tabular display of signal intensity and annotations
	Turn on or off display of Line Connector in Gene View
	Turn on or off display of custom data

Commands

The area where commands appear is called a *command ribbon*. The command ribbon that appears when you click the Home tab for ChIP 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 ChIP 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 or Workflow tabs. See the User Guides for those applications for information on the Navigator contents.

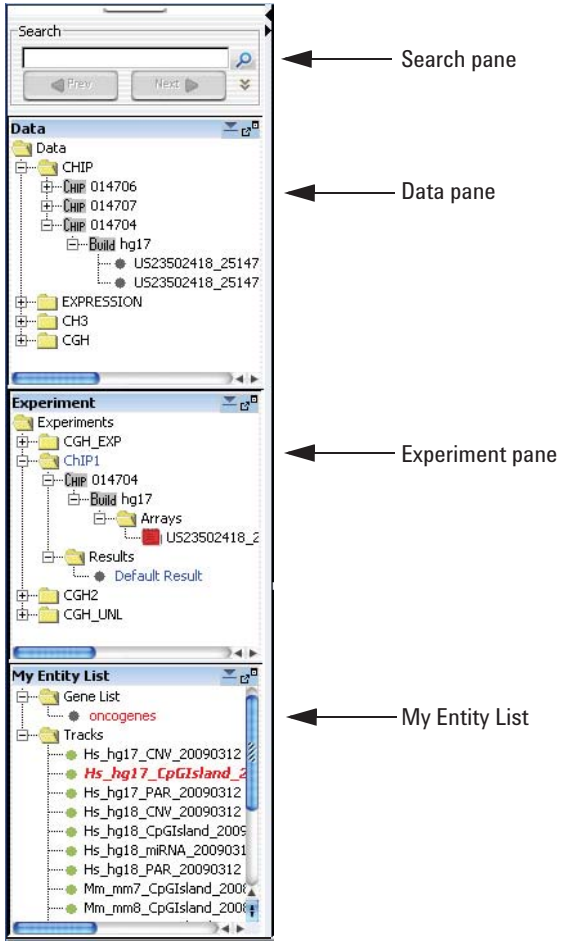

Figure 7 Navigator panes

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 wildcards for unspecified strings. For example, type *1234* to find any item that contains "1234".
Data	Displays the data content available to you in the program. Contains microarray data files, organized by application type, then by design and genome build. 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 141 and "Data pane – actions and shortcut menus" on page 142.
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. See "Experiment pane – icons, special text, and buttons" on page 144.
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 151.

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.

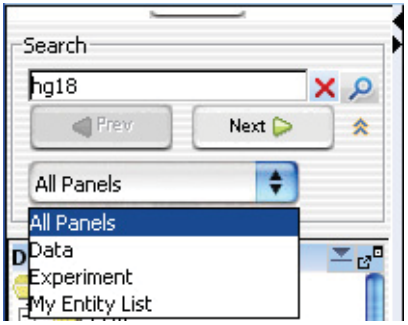

Figure 8 Navigator search pane

1. At the top of the Navigator, in the Panels list, select the Navigator pane to search. To search in all panes, select **All Panels.** If the panel 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 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 can be 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, “ChIP Interactive Analysis Reference”.

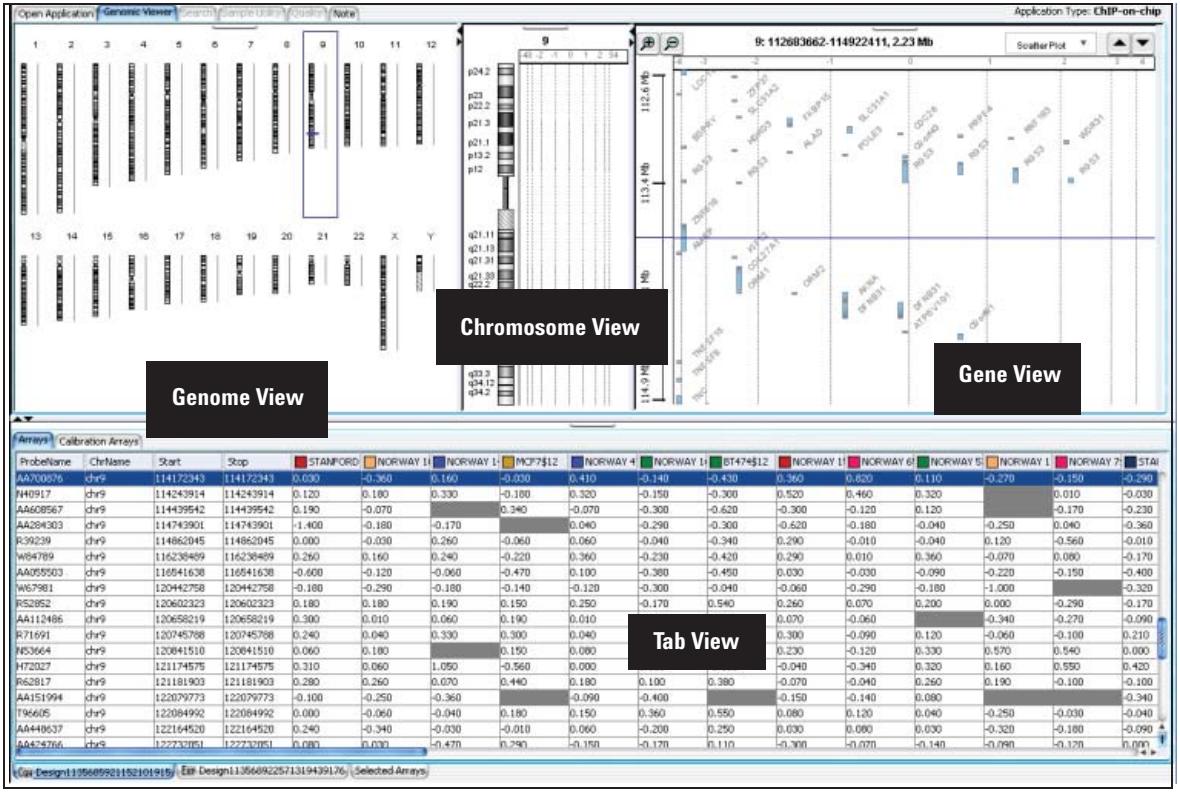

Figure 9 Genomic Viewer in vertical orientation

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 .

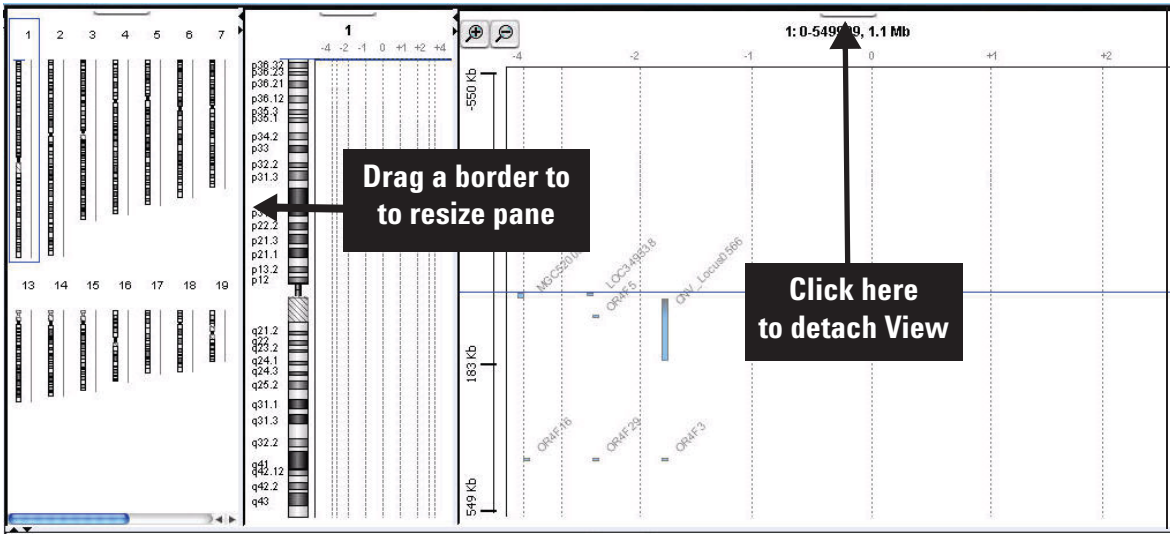

Figure 10 Changing the size of and detaching panes

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.

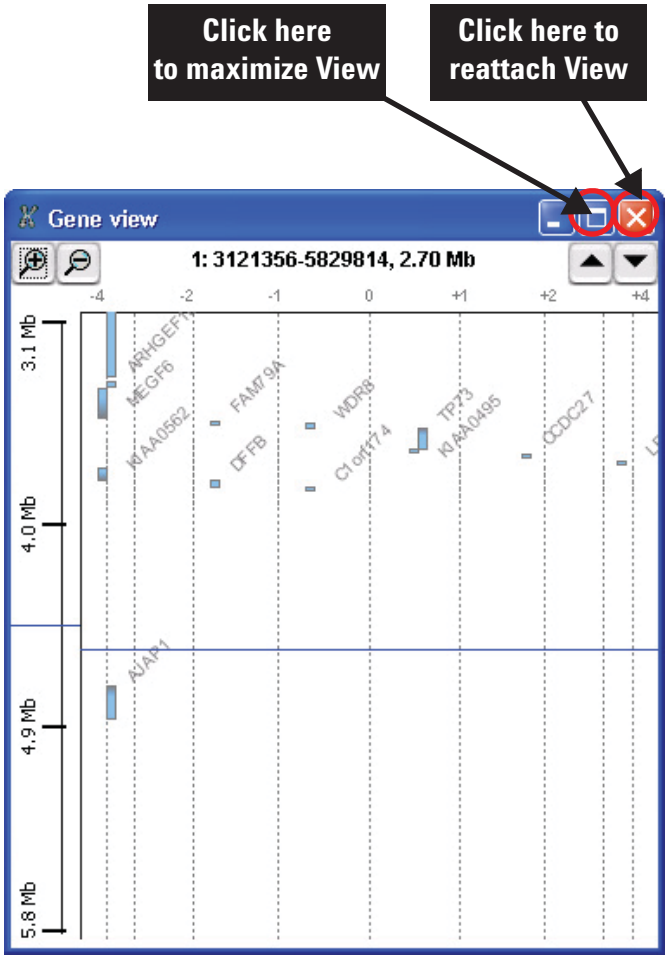

Figure 11 Maximizing and reattaching panes

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*.

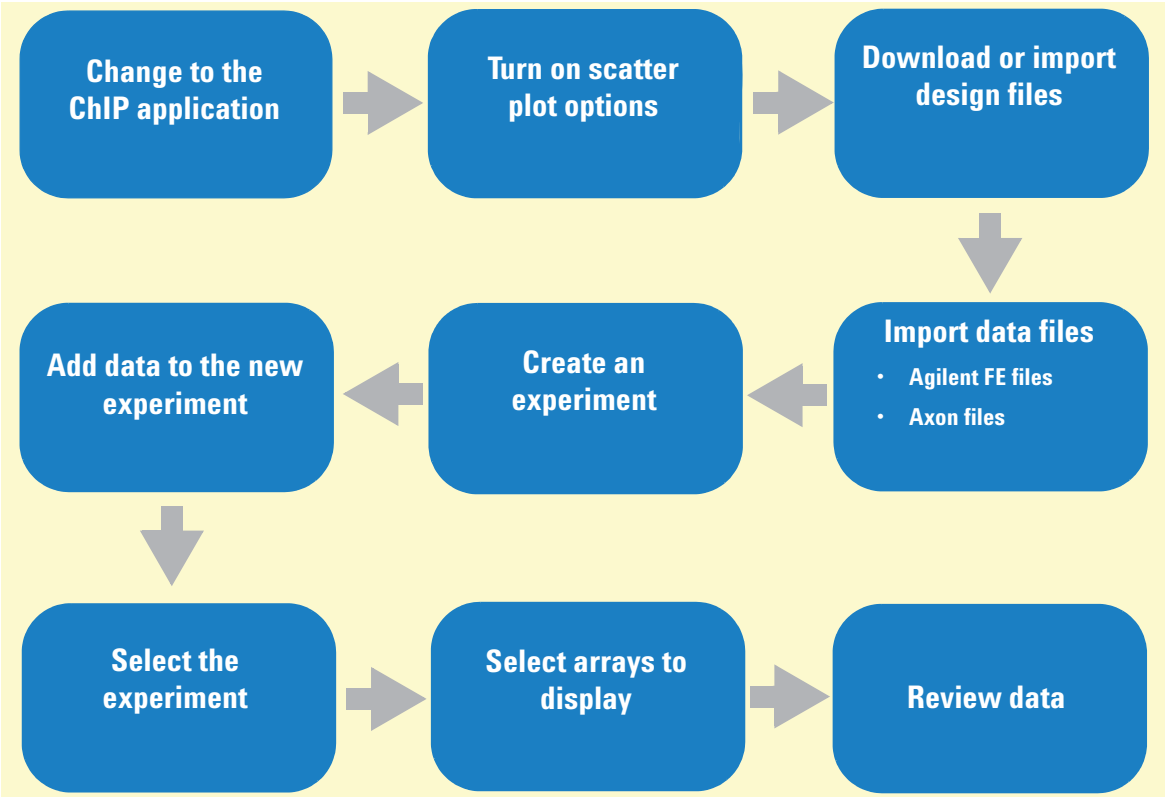

Figure 12 Typical pathway for displaying microarray data/results

General Instructions for Setting Options for ChIP Interactive Analysis

After you import data and set up experiments, you can set up preprocessing and analysis calculations before you make the experiment active. You can also change and apply them after an experiment has been activated. You can combine designs and replicate probes, apply a previously modified algorithm to show binding effects, and produce probe, gene and QC reports. Once you activate an experiment, the program recalculates the results after you change each option.

For more information on how to change analysis options, see Chapter 4, “Setting Up ChIP Interactive Analysis”.

For information on how to analyze ChIP data as part of a Workflow, see the *Workflow User Guide*.

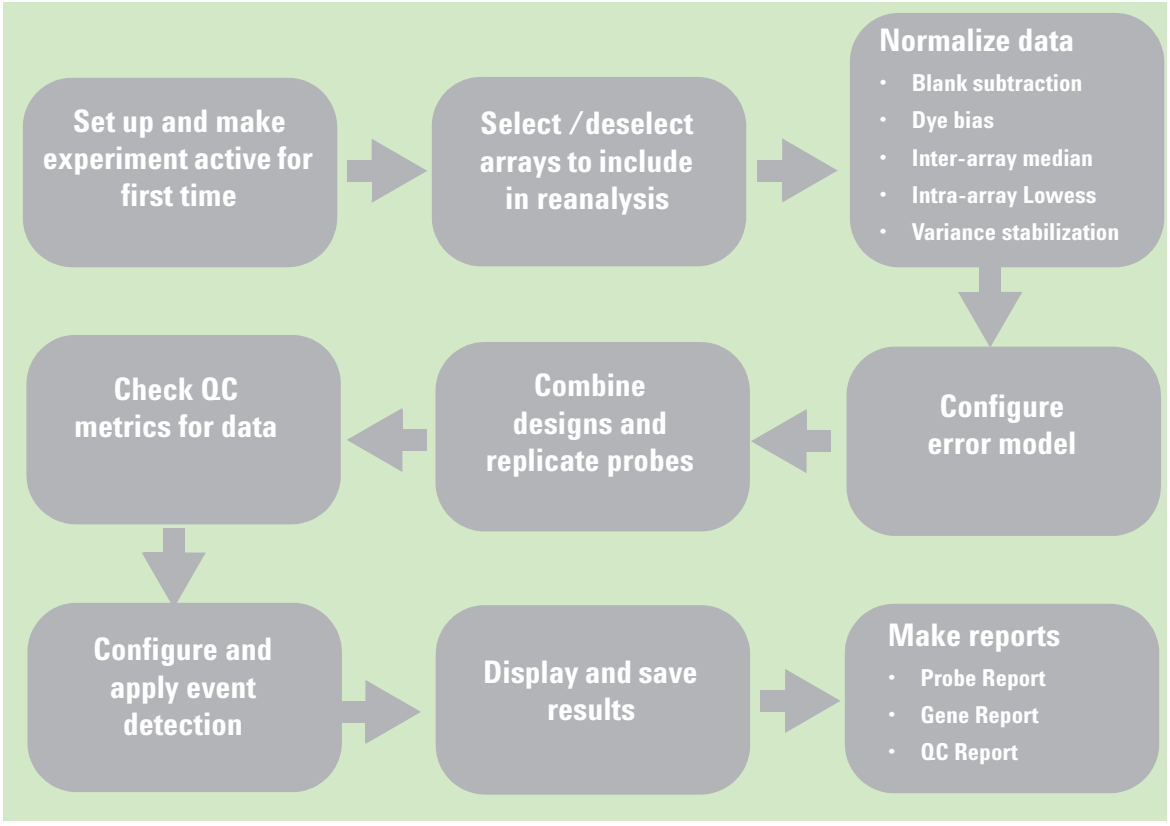

Figure 13 Typical analysis pathway – Interactive analysis for ChIP application

You can also use the output from the Agilent Feature Extraction program, instead of normalization and error modeling within the ChIP application.

Getting Help

To get help within Agilent Genomic Workbench

Agilent Genomic Workbench has several help resources. Help guides open with Adobe® Reader®.

Help Resource	Description/Instructions
ChIP Interactive Analysis User Guide	This user guide, which you are now reading, supplies comprehensive help on all available ChIP 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.** ChIP Interactive 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 from the Switch Application menu. **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 Assistance, 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 ChIP Data and Other Content

This chapter describes how to import, organize, manage, export and display ChIP data and other content within the user interface of the Agilent Genomic Workbench.

Importing Files

From the Home tab, you can import many kinds of files into the Agilent Genomic Workbench. The table below summarizes 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 in the program. See "Navigator" on page 137 for more information on the Navigator panes and how to use them.

Type of file	Comments	See these topics
Microarray data files	• Agilent Feature Extraction (*.txt) data files • Axon (*.gpr) data files	"To import Agilent FE or Axon data files" on page 47
Microarray design files	• Agilent GEML (*.xml) design files • Axon (*.gal) design files	"To import Agilent GEML design files" on page 43 "To import Axon design files" on page 45
Genome builds	Agilent-supplied genome information for human, mouse and rat genomes	"To import a genome build" on page 46
Tracks	BED format annotation track files	"To import tracks" on page 49
Array attributes	.txt files that you have created yourself or previously exported from the Agilent Genomic Workbench	"To import array attributes" on page 50
Experiments	ZIP format file of experiments exported from the Agilent Genomic Workbench applications	"To import an experiment file" on page 50
Filters	.xml file that contains filters created using the Agilent Genomic Workbench	"To import filters" on page 51

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 <version>\data**. If you like, you can select a different location.

CAUTION Do not select a location that contains a backup data folder. Data in the selected location will be overwritten by the current data.

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

The User Preferences dialog box appears. See "User Preferences" on page 245.

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

An Open dialog box appears.

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

NOTE Make sure you have full permissions in the data location.

The selected location appears in the 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 can be 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.

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

The Import Design Files dialog box appears. See "Import" on page 204. 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 and 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 207.

- If a design file passes validation, the Status column shows **Valid** in green.
- If the design is an Agilent Catalog design, and is not yet downloaded from the eArray Web site, the Status shows **Not Allowed** in red. You must download the file from the eArray Web site.
- If a design and build is already in the database, the Status shows **Overwrite** in yellow. If you continue, the imported design replaces the design in the database.
- If a design is already in the database, but has a different build, the Status shows **Update** in green. If you continue, this build of the design will be added to the database. The existing design build will not be overwritten.
- 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. 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 the 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 204. 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 and 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 236.

If a design file passes validation, the Status column will show **Valid** in green. If a design file fails validation, **Corrupt** appears in the Status column beside it, and the program does 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 use eArray to update 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. For more information on User Preferences, see "User Preferences" on page 245.

3 Click OK.

4 Expand the Navigator to display the design of interest.

5 Right-click the design in the Navigator, and click **Update from eArray.**

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 209.

2 Set the following. All are required.

Setting	Instructions
Species	• Type the genome's species of origin, as you would like it to appear within the program.
Build Name	• Type the name of the genome build you want to import, as you would like it to appear within the program.

Setting	Instructions
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 cytobands 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 Agilent FE or Axon data files

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

- Agilent Feature Extraction (FE) *.txt data files
- Axon (*.gpr) data files

To import Agilent Feature Extraction files, the representative GEML array design files must imported first. In order to import Axon data files, the representative Axon.gal design files must be imported first. See "To import Agilent GEML design files" on page 43 or "To import Axon design files" on page 45.

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 204.

2 To select a file for import, click its name. To select additional files, hold down the **Ctrl** key and 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 170.

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 allows you to work while the program imports the files.

To import tracks

You can import BED format track files into the 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 210.

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 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 combine tracks" on page 69 and "To show tracks in Gene View" on page 93.

To import array attributes

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

To import an array attributes file

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

The Import AttributeFiles dialog box appears. See "Import" on page 204.

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 the 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 the Agilent Genomic Workbench on another computer
- Agilent Genomic Workbench 5.x and 6.x experiment files

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

The Import Experiments dialog box appears. See "Import" on page 204.

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, with the appropriate arrays, in the Experiment pane.

NOTE

The 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 73.

To import filters

Filters are used in the Agilent Genomic Workbench to include or exclude data from an analysis, based on filter criteria. Filters are created in the interactive CGH and ChIP applications, or in workflow setup.

1 In the Home tab, on the Command Ribbon, click **Import > Filters**.

The Import dialog box appears. See "Import" on page 204 for more information.

2 Select the file that contains the exported filter(s) for import. and then click **Import**.

3 In the filters Import dialog box, mark the **Import** box next to each filter you want to import, and then click **OK.**

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 data analysis applications (CGH, ChIP, or methylation, for example), 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.

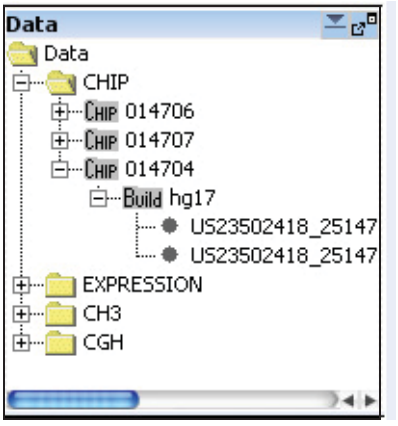

Figure 14 Data pane of the Navigator

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 information, see "Data pane – actions and shortcut menus" on page 142.

Many icons can appear in the Data pane. See "Data pane – icons, special text, and buttons" on page 141 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 63.

To create a new experiment

In the 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 180.

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 55.

- To create an experiment and add data to it now, follow these steps: (You can add or remove data from the experiment later, as well.)
  - **a** Click **Properties.**
    The Experiment Properties dialog box appears. See "Experiment Properties" on page 189.
  - **b** Under **Select Design,** select the design and genome build for 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 and click the names of additional arrays.
  - **d** Click  .
    The program transfers the selected arrays to the Selected Array List.
    The dialog box also gives you other options for adding arrays. See "Experiment Properties" on page 189 for information.
  - **e** Click **OK.**
    The program creates the new experiment, and adds 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.

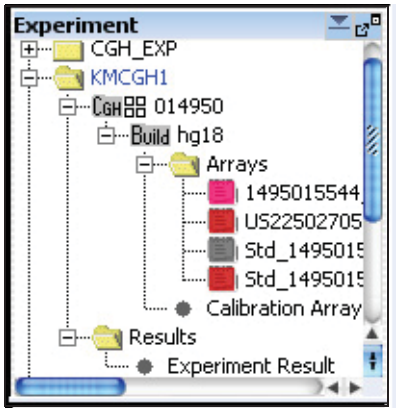

Figure 15 Experiment pane of the Navigator

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 desired experiment, then click **Show Properties.**

The Experiment Properties dialog box appears. See "Experiment Properties" on page 189.

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 and click their names.

5 Click > .

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

The dialog box also gives you other options for adding arrays. See "Experiment Properties" on page 189 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.

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.

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 and, if selected, signal intensities for arrays in the experiment. See "Tab View" on page 164. You can change the order in which the arrays appear in the table. If you display separate (stacked) 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 188.

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 the 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 remains 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.

NOTE To see the original name of the array, move the mouse pointer over the array name. A ToolTip appears that displays the original array name.

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 remain 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 65.

1 In the **Experiment** pane, expand folders until you can see the desired experiment, and the array(s) to remove from it.

2 In the **Arrays** or **Calibration Arrays** folder of the desired experiment, click the name of an array to select if for removal. Hold down the **Ctrl** key and 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 shows 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 to select as a calibration array.

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

To deselect multiple calibration arrays at once, select all of the 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 array-specific information, such as the amount of label used 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 228.

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

1. Right-click the experiment whose attributes you want to show or hide, or to change.
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*.
3. Click **Show/Hide Attributes**.

   The Show/Hide Columns dialog box appears. See "Show/Hide Columns" on page 242.
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 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 228.
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 array-specific information, such as the amount of label used or hybridization temperature. You can display a list of attributes for each array that is available in the program.

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 211. 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 display header and feature information passed through 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* for more information.

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 the Agilent Genomic Workbench. To display the data, gene list and track content, see Chapter 3, “Displaying ChIP 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 the Agilent Genomic Workbench.

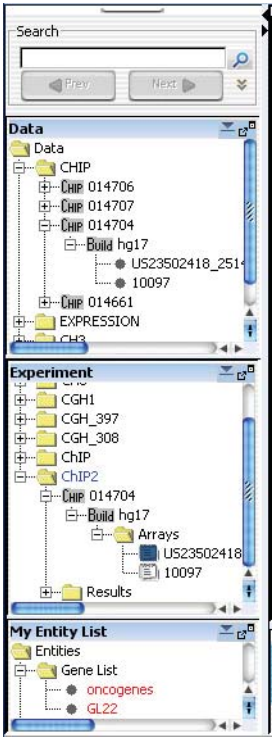

Figure 16 The Navigator

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 52.

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 138.

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 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 highlights them in red. You may need to scroll down to see retrieved items.
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 184.

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 the Agilent Genomic Workbench, and you are a registered eArray user, you can download the updated design files from within the Agilent Genomic Workbench. For more information about eArray, go to https://earray.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 245.

3 Click **OK.**

4 Expand the folders of the Data pane until you can see the design 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

1 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 57.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 to delete is associated with an experiment, first delete it from the experiment. See "To remove arrays from an experiment" on page 58.

2 In the Data pane, expand folders until you can see the design folder or array to delete. Do one of the following:

- For array data files, click the name of the first array, then hold down the **Ctrl** key and 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 and click the names of additional ones. This selects the designs and all array data files within them for deletion.

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

A confirmation dialog box appears.

4 Click **Yes.**

The program deletes the selected files.

CAUTION When you delete files, you permanently remove them from the 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 select. 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 68.

1. Follow these steps to select 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 157.
   b. In Chromosome View, in the plotting area to the right of the chromosome, drag the pointer over the approximate desired chromosomal region.
   The program encloses the region in a blue box, 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 159.
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 181.
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 My Entity List of the Navigator in the Gene List folder.

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 the 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 204.
3. Select the desired gene list file. To select additional gene list files, hold down the **Ctrl** key and click their names.
4. 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 200.

   You can also create gene lists. For more information, see "To create a gene list" on page 66.

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 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 169.

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 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 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 and click the names of additional gene lists.

This selects the 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

The table that you display 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 243.

To combine tracks

You can create a track that contains elements from two or more existing tracks. The existing tracks must be available in the 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 173.

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 consisting of elements that appear in both tracks.
OR	Creates a combined track out of 2 tracks consisting of 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 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 of the Navigator, expand the Tracks folder.
2. Click the name of a track to delete. Hold down the **Ctrl** key and click the names of additional tracks.

   This selects the 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 that you choose. You first select the arrays and then the attributes that you want exported for your array selection. You can access this capability from the Home tab or the shortcut 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 192.

If you opened this dialog box by right-clicking an experiment, only those arrays associated with 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 and click their names.

4 Click > .

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

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

The Export Array Attributes dialog box appears with the Attribute tab displayed. See "Attribute tab" on page 194.

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

6 In the Selected Attribute List, highlight those attributes you do not intend to export.

7 Click  .

8 Click **OK**.

The Export dialog box appears. See "Export" on page 191.

9 Select the folder in which to locate 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 196.

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

3 Click **OK.**

An Export dialog box appears. See "Export" on page 191.

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 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 success message appears.

4 Click **OK.**

To export tracks

You can export selected tracks as a BED format track file. You can then import this file into the 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 198.

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 export filters

You can export selected array, feature, and design filters that are available in some data analysis applications in the Agilent Genomic Workbench. The program exports all selected filters as a single *.xml file that you can import at a later time.

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

The Export Filters dialog box appears. See "Export Filters" on page 197.

2 Under Export, mark the check boxes beside the filter(s) to export. To select all filters for export, click **Select All.**

3 Click **OK.**

An Export dialog box appears.

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

The program exports all selected filters as a single *.xml 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 Microsoft® 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 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.

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 187.

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 saved file, then click **Save.**

3
Displaying ChIP 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 modify the display to show the data and content in different ways.

Selecting an Experiment for Data Display

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 137 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 selected experiment. See "To locate and display data (or results) within the Views" on page 87.

When you select an experiment and Preprocessing and Analysis options have been turned on, the program automatically begins the analysis of the selected array data with selected 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 to make it active, 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 a selected experiment, the program recalculates results.

1 If necessary, do one of the following to add an 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 53.
- Import a saved DNA Analytics 5.x, or 6.x ChIP experiment. See "To import an experiment file" on page 50.

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 both before and after you select it. Every time you select or deselect an array in an *active* 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 haven't changed. If they have changed, the program reanalyzes all of the arrays selected 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 will be 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 change color after activation.

In the Navigator, an array's icon has two appearances after experiment selection:


Array not selected.

Array selected. The specific color matches the color of the column headings for the array in Tab View. In addition, the program displays plot data 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 analysis.

   To include additional arrays, hold down the **Ctrl** key and 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 and click the name of the last one.
3. Right-click the name of one of the highlighted arrays, then click **Select.**

   After you select the arrays, the program reanalyzes the data set within the experiment and posts 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 view results of analysis" on page 115.

   To customize the appearance of the results in Genome, Chromosome, and Gene Views, see "To customize scatter plot ranges and colors" on page 84.

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 and click a column heading to select or deselect an array without affecting the status of other arrays.
- Right-click a column heading to open a shortcut menu with options that allow you to select or deselect that array, or all arrays.

For more information on the Tab View, see "Tab View" on page 164.

To change the display color of an array

The color assigned to an array changes 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. "Select Color" on page 230.

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 experiment, then click **Edit Array Color.** For more information, see "Edit Array Color" on page 187.

To analyze an experiment one sample at a time

You can use Sample-By-Sample View to analyze or reanalyze the *active* experiment one array at a time. When you work in this special mode, the program displays only one array at a time, and it also adds a new Sample-By-Sample View to the main window, where you can create or edit the Karyotype and Research Notes attributes of each array as you go.

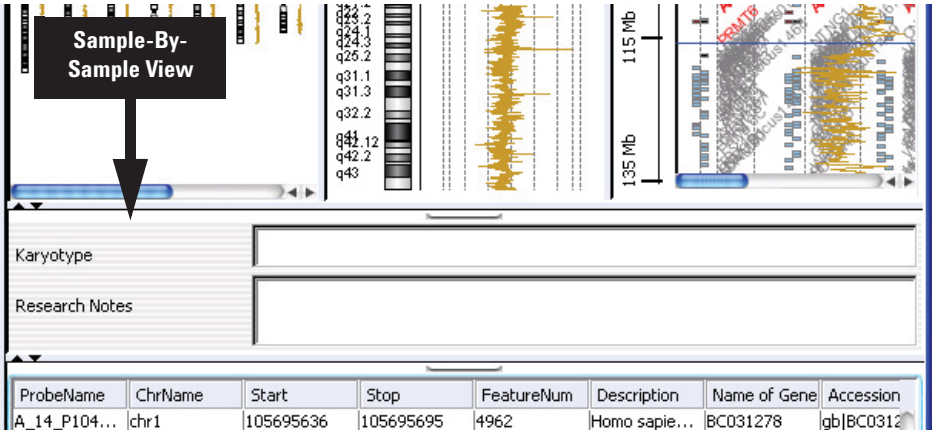

Figure 17 Sample-By-Sample View

1 Select an experiment. See "To select an experiment" on page 78.

2 Click **View > Show,** then mark **SampleBySample View.**

Sample-By-Sample View appears. The program also displays the first array in the experiment in Genome, Chromosome, Gene, and Tab Views.

3 Analyze the first array, as desired. You can also type comments in **Karyotype** and **Research Notes.**

The program saves what you type in the Karyotype and Research Notes attributes of the array.

4 Do any of the following:

- To show and analyze another array, right-click its name within the selected experiment in the Navigator, and click **Select**.
- To leave Sample-By-Sample mode, click **View > Show,** then clear **SampleBySample View.**

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 252 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 252 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, click **Scatter Plot,** and do any of the following:

To do this	Follow these steps
Show or hide ratio values in the Ratios plot	• To show the data points - Mark the **Ratios** check box and select **Ratio Values** from the list. • To hide all data points - Clear the **Ratios** 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 next to Color by, select **Log Ratio Values** from the list. • To hide all data points - Clear the **Log Ratios** check box.

To do this	Follow these steps
To color-code Log Ratios by Probe Score values in the Log Ratios plot	• To show the log ratios and color-code by Probe Score values- 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 175 for more information.

2 Click **X** to close the Scatter Plot window.

To customize scatter plot ranges and colors

To make it easier to see significant results, you can customize the display of scatter plot data. For each data type (ratio, log ratio) you can set custom ranges and colors for the display. For channels, you can set custom colors only.

To add and customize a range

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

2 Under Configure Coloring schemes, mark the check box under **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 175.

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

6 Type 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 230 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.

To edit or remove a range

1 In the Configure Coloring Ranges and Shades dialog box, click the **Ratios** 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 Preferences dialog box to change the appearance of the scatter plots in Chromosome and Gene Views.

1 In the Genomic Viewer, right-click in the Gene View or Chromosome View, and 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 252.

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.**

To do this	Follow these steps
Change the symbol that appears for data points	You can select the symbol separately for each design type. **a** 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.**

3 You can also configure ranges and colors for the scatter plot from the **View Preferences** dialog box. For more information, see "To customize scatter plot ranges and colors" on page 84 and "View Preferences" on page 252**.**

4 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 a separate page. Chromosomes and genes appear on the printed pages, but tracks do not.

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

2 Set print options, as desired, 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 the Ratios and/or Log Ratios plot.

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 move 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 desired chromosome. All other views switch to the selected chromosome.
Display data in a region of the selected chromosome	• In Chromosome View, in the scatter plot, drag the pointer over the desired region. Gene View expands (or shrinks) to show only the selected region. Tab View scrolls to the new cursor location.
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.

To do this	Follow these steps
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.
Center all views on the location of a specific gene	**a** Click **Home > Go To Gene/Genomic location** A dialog box appears. **b** Under **RefSeq by Symbol,** either select the desired gene (if available) or type the name of the gene. **c** Click **Go.** All views move to the location of the selected gene.
Center Chromosome and Gene views based on data in Tab View	• In Tab View, click any entry in any table, except a column heading. Chromosome and Gene Views become centered on the genomic location that corresponds to the selected entry.
Scroll to a specific column in Tab View	**a** In Tab View, right-click any column heading, then click **Scroll To Column** A dialog box appears **b** In **Select Column,** select the desired column. **c** Click **OK.**
View 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 168.

To display QC metrics of arrays and set array QC status

QC Metrics are indicators of the quality of data from an array. QC Metrics, which are available only for Agilent arrays, are either passed through from the FE program, or are calculated by the ChIP application

itself. You can view QC Metrics for an individual array, the arrays in an experiment, or the arrays associated with a specific design and genome build.

1 In the Preprocessing tab, do one of the following:

- To display the QC metrics of Agilent arrays in the active experiment – In the ribbon, click **QC Metric.**
- To display the QC metrics of a specific Agilent array – In the **Data** or **Experiment** panes of the Navigator, right-click the name of the array. In the shortcut menu that appears, click **QC Metrics.**
- To display the QC metrics of all Agilent arrays associated with a genome build within a specific design folder – In the **Data** or **Experiment** panes of the Navigator, right-click the name of the genome build folder within the desired design folder. In the shortcut menu that appears, click **QC Metrics.**

In each case, the QC Metrics Table appears. For each array, the QC Metrics Table color codes each metric as Evaluate (pink), Good (turquoise), Excellent (yellow) or NA (white). This dialog box also offers several other options for display of the QC metrics. See "QC Metrics Table" on page 219.

2 In the QC Status column, select the desired overall status for each array.

This sets the QCMetricStatus attribute of each array. You can use the value of this attribute as the basis for an array filter.

3 Click **Close.**

After you view the QC metrics of an array, you can set the QC status of the array. Later, you can use an array filter to include or exclude arrays from an analysis based on their QC metric status.

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 the Agilent Genomic Workbench, and you can also create them in the program and export them. See "To import a gene list" on page 67, and "To export a gene list" on page 74.

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 selected 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 desired gene list does not appear, create or import it. See "To create a gene list" on page 66, or "To import a gene list" on page 67.

2 Right-click the desired 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 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 selected 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 choose the desired color. See "Select Color" on page 230.

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 200. 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 74.

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 "Track" on page 243.

3 Do any 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 245.

2 Click **Tracks**.

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

4 Click **OK.**

The program displays the selected tracks in Gene View.

To change the appearance of tracks

Within the Tracks tab of the User Preferences dialog box, you can modify 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.
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.**

To do this	Follow these steps
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.**
View the parameters and the list of annotations of a track	• In the list of tracks, next to the desired track, click **Details...**
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 display tracks in UCSC Browser

1 Right-click Gene View, and click **Show in UCSC**.

The View coordinates in UCSC browser dialog box appears. See "View coordinates in UCSC browser" on page 250.

2 Complete the dialog box with the track parameters, and click **OK**.

The UCSC Browser appears, if you are connected to the Internet. It may be necessary to enable pop-ups or set other preferences on the UCSC Web site; see the browser help for information.

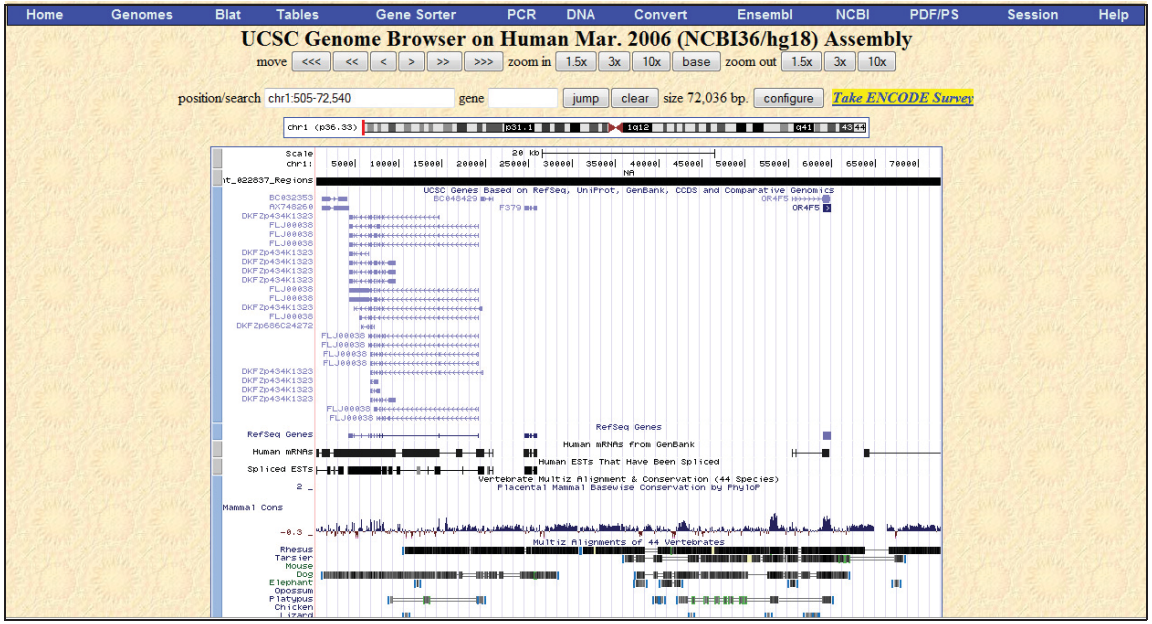

Figure 18 Track displayed in UCSC browser

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, activate an experiment whose data are based on a design file of a different genome build.

The display automatically changes when you activate 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 46.

The program does 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 53, "To add arrays to an experiment" on page 55, 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 164.

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 column to be searched 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** – Return entries that match upper and lower case characters in the search term.
	  ◦ **Match whole word** – Return 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 like, and the program

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

4 When you finish 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 be a registered eArray user. For more information, go to https://eArray.chem.agilent.com. You must also provide your eArray user name and password in the Miscellaneous tab of the User Preferences dialog box. See "User Preferences" on page 245.

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 229.

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 you want 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 passes 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, the Agilent Genomic Workbench opens the site in your Web browser and passes 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 183.

2 Click **New.**

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

This name appears 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 pass a search string to the site. Use `<target>` as the query string value.

For example, this URL transfers the 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 you want 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 ChIP Interactive Analysis

This chapter gives instructions on how to set up the interactive analysis functions for ChIP Interactive Analysis experiments. These include the Preprocessing, Analysis and Reports tabs.

Working with Interactive Analysis Options

For a detailed description of the interactive analysis tabs – Preprocessing, Analysis, Reports – and their commands see "Command Ribbons" on page 124. For information on the **View** command ribbon, see "View command ribbon" on page 133 and Chapter 3, "Displaying ChIP Data and Other Content".

Figure 19 ChIP Interactive Analysis tabs

Table 3 shows the tasks in this chapter that you can use to change preprocessing and analysis options, display results, and create reports.

Table 3 Interactive analysis topics

Subject	See these topics
"Changing Preprocessing and Analysis Options"	"To apply normalization" on page 103 "To configure the error model" on page 106 "To combine (fuse) arrays" on page 108 "To combine intra-array replicates" on page 110 "To combine interarray replicates" on page 110 "To apply event detection" on page 111
"Displaying Results"	"To view results of analysis" on page 115 "To save a result" on page 116 "To restore a saved result" on page 117
"Creating Reports"	"To create a probe report" on page 118 "To create a gene report" on page 119 "To create a QC report" on page 120

Changing Preprocessing and Analysis Options

To apply normalization

You normalize microarray data to correct for three known factors that cause the reported signal intensities to differ from the "true" signal:

- Non-specific binding (noise)
- Variations from one array to another
- Dye bias (the tendency of a specific fluorescent dye to alter binding)

See Table 4 on page 104 for a description of the normalization methods. You can apply any combination of normalization methods to your data, or you can use the normalized output from the Agilent Feature Extraction program instead.

To apply normalization

1 In the Preprocessing tab, under Normalization, select **User Defined.**

The Edit button becomes available.

2 Click the **Edit** button  .

The Setting Normalization Parameters dialog box appears. The available normalization methods appear under Select Normalization.

3 Under Select Normalization, mark the desired normalization methods.

4 To set the specific parameters of a normalization method, click the name of the method.

The name of the method and its parameters appear in the right pane of the dialog box. Select the desired parameters from the list(s) in this pane. See "Setting Normalization Parameters" on page 238.

5 Click **OK.**

To use processed Feature Extraction output

- In the Preprocessing tab, under Normalization, select **Use FE Output.**

The Edit button becomes unavailable.

By default, the program uses raw feature intensity data from the imported data files. When you use this option with files from the Agilent Feature Extraction program, the Agilent Genomic Workbench uses the normalized feature intensities from the FE software instead.

Table 4 Normalization methods available in interactive mode

Normalization method	Comments
Blank subtraction normalization	This kind of normalization corrects for nonspecific binding. It first calculates the central tendency of the negative controls on the array for both the immunoprecipitated (IP) and whole cell extract (WCE) channels. It then subtracts these central tendencies from the raw signal intensities of each feature on the array.
Interarray median normalization	This kind of normalization corrects for variations from one replicate array to another. The Agilent Genomic Workbench calculates and applies it separately for each channel. It first calculates the median signal intensity over the common probes in each replicate array. It then finds the average of these median intensities over all replicates of all arrays. For each array, it computes the ratio of its median signal intensity to the average of the median signal intensities of all arrays. Finally, it normalizes data by multiplying each signal intensity by the applicable ratio.
Dye-bias (intra array) median normalization	This kind of normalization corrects for dye bias within each array in an experiment, and it normalizes the intensities of the IP channel only. It can calculate the dye bias in two ways: • **By equalizing central tendencies of IP and WCE channels** – This method first calculates the ratio of the median IP signal intensity to the median WCE signal intensity. Then, it multiplies the signal intensities of the data probes by this ratio. • **By normalizing central tendency of log ratios to 1** – This method multiplies the signal intensities of all data probes on the array by a correction factor. This correction factor adjusts the central tendency of log ratios of the data probes on the array to 1. **Note:** If Dye-bias (intra-array) Median normalization is selected, Variance Stabilization normalization and Intra-array Lowess normalization cannot be selected.

Table 4 Normalization methods available in interactive mode

Normalization method	Comments
Intra-array Lowess (intensity dependent) normalization	Intra-array normalization attempts to correct for artifacts caused by nonlinear rates of dye incorporation, as well as inconsistencies in the relative fluorescence intensity between some red and green dyes. The Lowess normalization algorithm normalizes the channels within each array using a nonlinear polynomial fit to the data, and effectively normalizes by probes and by arrays. **Note:** If you are using feature extraction data that have been normalized by the Lowess approach, you do not need to apply the intra-array Lowess normalization here. **Note:** If Intra-array Lowess normalization is selected, blank subtraction and Interarray Median normalization are not available.
Variance stabilization	This normalization is useful for data that is either "blank-subtracted" or "spatially detrended" but it may have utility for data processed by other means as well. Variance stabilization is an alternative to Lowess normalization that fits a regression curve to signal intensities after applying an "arcsinh(x)" transform to each channel. This approach uses a two-parameter error model to compress the reported ratios of probes with weak signals after blank-subtraction. After the transform is applied, the variance of the reported log ratios should be independent of the signal strength. **Note:** If you are using feature extraction data that have been normalized by the Lowess approach, you do not need to apply Variance Stabilization here. **Note:** If Variance Stabilization normalization is selected, Dye-bias (intra-array) Median normalization and Intra-array Lowess normalization are not available.

For a detailed description of the statistical algorithms the Agilent Genomic Workbench uses to normalize data, see "Normalization Algorithms" on page 274.

To configure the error model

The main goal of ChIP data analysis is to find the chromosomal locations where protein binding (or other events) occur. The error model calculates the likelihood that a given probe signal represents a binding event, and assigns p-values to each probe as a measure of this likelihood. A p-value close to 1 indicates that a probe is unlikely to represent a significant binding event. A very small p-value (for example, < 0.001) indicates that the probe is very likely to represent one.

After the ChIP application uses the error model to assign p-values, it then evaluates the p-values of groups of neighboring probes to make binding calls. See "To apply event detection" on page 111. For a detailed discussion of the statistical algorithms used in the error model, see "Error Models" on page 284.

To select an error model

- In the Preprocessing tab, under Error Model, select one of these options:

Option	Description
Whitehead Error Model	The program uses this error model by default. Select this model unless you have tried both models and know that the other one provides a better match to biological truths and/or positive controls that you have available for your specific experiment. When you select this model, the Edit button becomes available, and you can set additional advanced parameters. For a description of the statistical algorithm used in this model, see "Error Models" on page 284.
Use FE Error Model	Uses data produced by the error model from the Agilent Feature Extraction (FE) program. These data are available in the imported FE file(s). Select this model if you have tried both models, and know this one provides a better match to biological truths and/or positive controls you have available for your specific experiment. If you select this model, the Edit button becomes unavailable; you do not need to set any additional parameters. For a description of the statistical algorithm used in this model, see the *Feature Extraction 10.7 Reference Guide*.

To set parameters for the Whitehead error model

The parameters for the Whitehead error model are optional, advanced settings. You can use them to optimize the statistical calculations of the error model using training data specific to your particular assay.

1 Select the Whitehead error model. See "To select an error model" on page 106.

2 Under Error Model, click the **Edit** button  .

The Whitehead Error Model Parameter Settings dialog box appears.

3 Set any of these parameters, as desired:

Parameter	Comments/Instructions
Source of additive (intensity-dependent) error of each channel is	The options for this parameter affect the additive (intensity-dependent) component of the estimate of the error in IP – WCE. Select one of these sources: • Standard deviation of background pixels • Additive error as computed by Agilent Feature Extractor • Observed spread of negative controls
Custom defined f-value	The f-value of one replicate of an array is the rate at which the multiplicative error increases with signal intensity. Normally, the ChIP application calculates f-values automatically, but you can type a custom value. **1** Mark **Custom defined f-value.** **2** In the box, type the desired f-value.

4 Click **OK.**

To combine (fuse) arrays

If you have two arrays that use different design files, you can combine them into one larger virtual array. This can increase the coverage of the genome in your design. For example, if 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 of the probe data in the display at once.

The program cannot combine arrays from more than two different design files; see the requirements for fusing arrays at the end of this topic.

1 Create and activate a new experiment. See "To create a new experiment" on page 53 and "Selecting an Experiment for Data Display" on page 78.

2 To the new, active experiment, add the arrays that you want to fuse. See "To add arrays to an experiment" on page 55.

3 Assign the same value to the **ArraySet** attribute of every array you want to fuse. 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  .
   The field becomes active.

   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 fused.

NOTE Double-check the values in the ArraySet Attribute column of the dialog box. The Agilent Genomic Workbench fuses 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.

- **Remove arrays from experiment after fuse** – To delete the initial unfused arrays from the experiment, mark this option. This reduces the duplication of data within the experiment.

6 Click **Continue.**

The program fuses the arrays. The fused 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.

Requirements for fusing arrays:

- Each array must be associated with a different design file.
- All of the arrays must be of the same application type (for example, ChIP).
- None of the arrays can be fused arrays.
- The samples you hybridize to the arrays must all be aliquots from the same preparation.
- (Preferred) Hybridization and labeling occur for all samples together under the same conditions.

NOTE

- If the arrays that you fuse have probes in common, these probes appear as replicates in the fused array. You can combine these replicates. See "To combine intra-array replicates" on page 110.
- If the original arrays that you fused change, you can manually update the fused array. Fuse the same set of arrays again.
- If you want to fuse many arrays, consider importing a sample attributes file. See "To import array attributes" on page 50, or the *Sample Manager User Guide*.

To combine intra-array replicates

Intra-array replicates are features within the same array that contain the same probe. The probe information is contained in the design file for each type of array. You can combine these replicates to increase the statistical power of your analyses. For a discussion of the statistical model that the Agilent Genomic Workbench uses to combine replicates, see "Intra-array (dye-bias) median normalization" on page 277.

To combine intra-array replicates for the selected arrays of the active experiment:

1 Click **Preprocessing.**

2 Under Combine, in Replicates, mark **Intra Array.**

The Agilent Genomic Workbench combines the replicates in the selected arrays within the experiment.

If you set this option before you activate an experiment, it is applied when you activate the experiment and select the arrays.

To restore the replicates to their uncombined state, clear the check box.

To combine interarray replicates

When you combine interarray replicates, you select an array attribute. The Agilent Genomic Workbench combines replicates from the arrays with the same value for the selected attribute. For a discussion of the statistical model the Agilent Genomic Workbench uses to combine replicates, see "Interarray median normalization" on page 276.

To combine interarray replicates for all arrays of the active experiment:

1 Click **Preprocessing.**

2 Under Combine, in Replicates, mark **Inter Array.**

3 In Group By, select the desired array attribute.

For the attribute that you select, each array must have the same value. You can assign values in one of two ways:

- In the **Sample Manager** tab, assign the values in the column of the selected attribute for the arrays whose replicates you intend to combine.

- In the Experiment pane, right-click the name of one of the arrays to be fused, then click **Show Properties.** Assign the value for the selected attribute. Repeat for each array.

4 Click **Go.**

The Agilent Genomic Workbench combines replicates from the arrays with the same value for the selected attribute in the active experiment.

If you set this option before you activate an experiment, the program applies it when you activate the experiment and select the arrays.

To restore the replicates to their uncombined state, clear **Inter Array,** then click **Go.**

Microarray experiments can include biological or technical replicate arrays, which are often combined to create one virtual array for downstream analysis. Biological or technical replicates of the physical arrays may be used to mitigate systematic variation arising from sample preparation or array processing.

Because the replicates are hybridized to microarrays with the same design, inter-array replicates are simply features on different arrays that contain the same probe. The combined signal values give an increase in statistical power for measurement of log ratios.

To apply event detection

Whitehead per-array neighbourhood detection model

The Whitehead per-array neighbourhood model (and the modified version) used to detect binding events, considers the *p*-values of both the probe in question and its neighbors. You can customize the parameters of the model, including the maximum distance between neighbor probes and the stringency of the detection process. The model considers probes in groups of three, shown in Figure 20. Two neighbor probes (blue) flank a central probe (red).

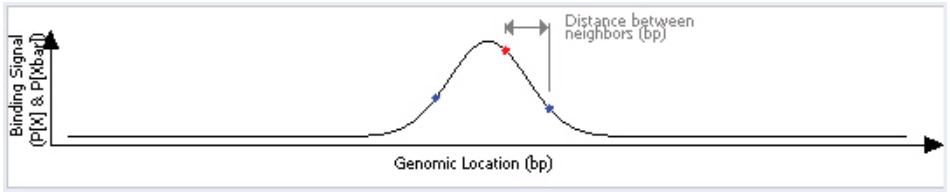

Figure 20 Central probe flanked by two neighbor probes

The program considers the probe "bound" if the p-value of the composite error-corrected ratio ("X") of all three probes ("X_{bar}") is less than a set cut-off value, and if either of the following is true:

- The p-values for the central probe and at least one of its neighbors are less than set cut-off values.
- The p-value of one (or optionally, another number) of the neighbors of the central probe is less than a set cut-off value.

For a detailed description of the statistical calculations involved in event detection, see "Peak Detection Algorithms" on page 289.

Predefined peak shape detection algorithm

The predefined peak shape detection algorithm produces a peak shape that is universal to each microarray/microarray set. In this way, detailed per-probe information can be used to detect peaks:

- If a probe is not inside any of the "peak" objects, then it is "not bound." Nothing special is drawn for this probe, except a baseline at exactly a ratio of 1 (log ratio of zero).
- If a probe is inside a peak, then it gets the significance value and score value of the associated peak. If it is inside two overlapping peaks, it gets the values for the peak with the better score. Peaks are drawn by computing ratios from the collection of detected peaks.

To apply Whitehead per-array neighbourhood event detection to the current experiment

1 Configure the error model. See "To configure the error model" on page 106.

2 Click **Analysis.**

3 Under Event Detection, select either **Whitehead Per-Array Neighbourhood Model** or **Whitehead Per-Array Neighbourhood Model (Modified).**

4 Click the **Edit** button  .

The Whitehead Per-Array Neighbourhood Model (or Modified) Parameter Settings dialog box appears. "Whitehead Per-Array Neighbourhood Model Parameter Settings" on page 257 or "Whitehead Per-Array Neighbourhood Model (Modified) Parameter Settings" on page 259.

5 In **A probe is considered "bound" if,** change the parameters of the event detection algorithm as desired. The parameters are based on the three-probe model shown in Figure 20. Table 5 describes the parameters.

6 Click **OK.**

7 Under Event Detection, mark **Apply.**

The program applies event detection to the experiment. By default, binding events appear as shaded regions in Gene View.

Table 5 Whitehead per-array neighbourhood model parameters

Parameter	Comments
Maximum distance (in bp) for two probes to be considered as neighbors.	The program only considers probes to be neighbors if their genomic locations are within this threshold distance. The default value for this parameter is 1000 base pairs. • To change the value, delete the old value and type a new one in the box.
$P(X_{bar}) <$	• This parameter refers to the *p*-value of the average error-corrected ratio ("X") of the central probe and its left and right neighbors. (The "bar" indicates the average.) • The default value is 0.001. • To make detection more stringent, decrease the value.

Table 5 Whitehead per-array neighbourhood model parameters

Parameter	Comments
Central probe has P(X) <	• The central probe is the red probe in Figure 20. • The default value is 0.001. • To make detection more stringent, decrease the value.
At least one neighboring probe has P(X) <	• Neighboring probes are probes to either side of the central probe. The blue probes in Figure 20 are the neighbors of the central (red) probe. • The default value is 0.1. • To make detection more stringent, decrease the value.
At least **n** of the neighbors has P(X) <	• The default value for **n** is 1 or 2 for the Modified Model. • The default cut-off value for P(X) is 0.005. • To make detection more stringent, decrease the value.

To apply predefined peak shape v2.1 event detection to the current experiment

1 Configure the error model. See "To configure the error model" on page 106.

2 Click **Analysis.**

3 Under Event Detection, select **Pre-defined Peak Shape detection v2.1.**

4 Click the **Edit** button  .

The Pre-defined Peak Shape detection v2.1 dialog box appears. See "Pre-defined Peak Shape detection v2.1" on page 214.

5 Complete the dialog box. Descriptions of the parameters for the dialog box are shown in "Pre-defined peak shape detection parameters v2.1" on page 215.

6 Click **OK.**

7 Under Event Detection, mark **Apply.**

The program applies event detection to the experiment. By default, binding events appear as shaded regions in Gene View.

Displaying Results

This section explains how to display, save, and restore the analysis results of the active experiment.

To view results of analysis

After you set up an experiment and analyze it, the program displays the results automatically in Chromosome and Gene Views. Figure 21 shows an example of a result displayed in Gene View. To customize the way the results appear, see "To change scatter plot appearance" on page 85.

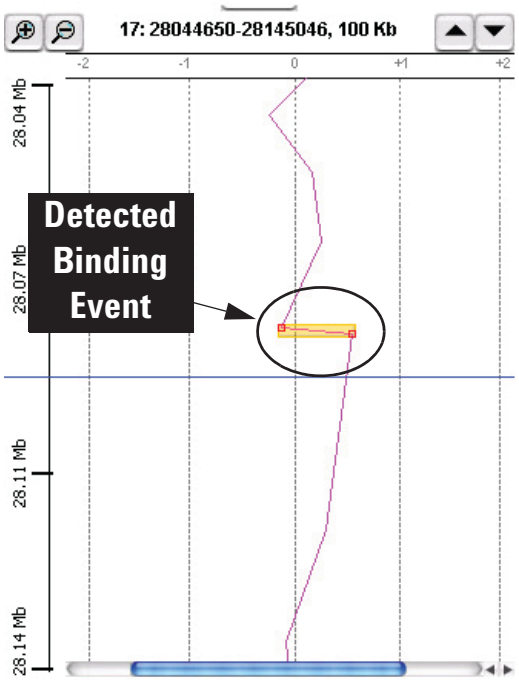

Figure 21 Gene view, showing detected binding event

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

Check these three potential problem areas in order:

- You may not have applied the event detection algorithm. In the **Analysis** tab, under Event Detection, mark **Apply.**
- You may not have selected the experiment or selected the arrays of interest. See "To select an experiment" on page 78.
- You may not be looking in the right region of Chromosome View.

See "To show or hide data in the scatter plots" on page 83 to find out how to display the Scatter Plot, if it is not visible, and how to navigate the View displays to observe results.

To save a result

The program allows you to save the 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 name of the experiment, then 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, then click **Save Experiment Result.**

A dialog box appears.

3 Select one of these options:

- 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

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.

Creating Reports

To create a probe report

The ChIP application makes information about the probes in the current experimental result available in tab-separated value (*.tsv) format. A probe report contains one row for each probe in the array (or array set). The program generates a separate file for each array. See "Probe Report format" on page 260 for a description of the columns in the report. You can view probe reports and perform further analysis on them with a spreadsheet program.

To create a probe report based on the current experimental result:

1 Click **Reports > Probe Report.**

The Probe Report Settings Parameter Settings dialog box appears.

2 Click **Browse.**

The Select report folder dialog box appears. See "Select report folder" on page 234.

3 Select a folder for the report.

4 In File name**,** type a name for the report.

If you type a name that matches an existing report in the selected folder, the program overwrites the existing report when it creates the new one.

5 Click **Open.**

The location of the report appears in the Probe Report Settings Parameter Settings dialog box in Report Location.

6 Click **OK.**

The program creates the report. If the report contains more than one file, the program creates a folder that contains the report files.

A dialog box appears.

7 To view the report file(s) in Windows Explorer (in the Finder in Mac OSX), click **Yes.** Otherwise, click **No.**

To create a gene report

The ChIP application makes information about the genes in the current experimental result available in tab-separated value (*.tsv) format. A gene report contains one row for each probe in the array (or array set), grouped by the genes to which the probes bind. The program generates a separate file for each array. It also includes loci represented by probes on the array that are not associated with genes. The program creates gene reports in several formats. See "Gene Report formats" on page 262 for a description of these formats, and the columns in each. You can view gene reports and perform further analysis on them with a spreadsheet program.

1 Click **Reports > Gene Report.**

The Gene Report Settings Parameter Settings dialog box appears. See "Gene Report Settings Parameter Settings" on page 202.

2 Mark either of these check boxes, if desired.

- **Show only gene names** – The resulting gene report contains only accession numbers of genes (or chromosomal locations for probe loci not associated with genes). This option overrides the next one.
- **Show probe information** – The resulting gene report contains additional information about the probes in the array associated with the bound genes.

3 Click **Browse.**

The Select report folder dialog box appears. See "Select report folder" on page 234.

4 Open a folder for the report.

5 In File name, type a name for the report.

If you type a name that matches an existing report in the selected folder, the program overwrites the existing report.

6 Click **Open.**

The path name of the report appears in the Gene Report Settings Parameter Settings dialog box in Report Location.

7 Click **OK.**

The program creates the report. If the report contains more than one file, the program creates a folder that contains the report files.

8 To display the report file(s) in Windows Explorer (in the Finder in Mac OSX), click **Yes.** Otherwise, click **No.**

To create a QC report

The QC report summarizes the settings of the current analysis, and the overall statistics of each array. In addition to summary tables, it includes plots that summarize the data graphically. The program creates the QC Report in HTML format, and generates a separate folder for each array. For more information about the contents of the report, see "QC report format" on page 263.

1 Click **Reports > ChIP QC Report.**

The QC Report Settings Parameter Settings dialog box appears. See "QC Report Settings Parameter Settings" on page 227.

2 Click **Browse.**

The Select report folder dialog box appears. See "Select report folder" on page 234.

3 Open a folder for the report.

4 In File name**,** type a name for the report.

If you type a name that matches an existing report in the selected folder, the program overwrites the existing report when it creates the new one.

5 Click **Open.**

The location of the report appears in the QC Report Settings Parameter Settings dialog box in Report Location.

6 Click **OK.**

The program creates a folder that contains the report files. A dialog box appears.

7 To display a report, follow these steps:

a Click **Yes.**
The QC report folder for each microarray appears in Windows Explorer (In a new Finder window in Mac OSX).

b Open the desired folder, then double-click the **QCReport** HTML file. Your Internet browser opens, displaying the QC report.

5
ChIP Interactive Analysis Reference

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

ChIP Application Main Window

The main window of the ChIP Application contains the main components illustrated in Figure 22.

Figure 22 Agilent Genomic Workbench 7.0 – ChIP application main window

Switch Application Menu

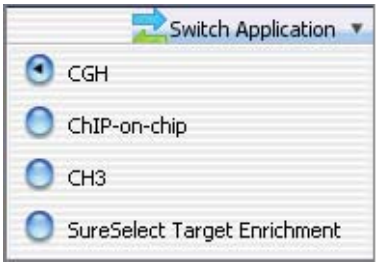

Figure 23 Switch Application menu

The Switch Application menu lets you switch to the other data display and analysis application type in Agilent Genomic Workbench. 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-on-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. Import, export, and view data. See the *Target Enrichment User Guide* for more information.

Command Ribbons

When you click a tab, groups of commands or single commands appear. The entire string of commands is called a command ribbon. The tabs that are displayed (Figure 24) change depending on what application is selected (such as CGH, ChIP, CH3). This section describes the tab commands used to import, manage, export, and display data/results for ChIP interactive analysis.

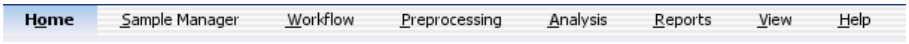

Figure 24 Agilent Genomic Workbench 7.0 - ChIP Application tabs

Home command ribbon

The Home command ribbon lets you import, manage, export and display ChIP data and other content (gene lists, tracks) for further analysis.

Figure 25 Command ribbon for the Home tab for ChIP

User Preferences Opens the User Preferences dialog box with three 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 245.
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 245.

Tab	Description
License	Opens a dialog box where you can add an Agilent Genomic Workbench application license, if you choose to purchase one after using the unlicensed version. See "User Preferences" on page 245.

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 204 and "To import Agilent FE or Axon data files" on page 47. • **Axon File** – Opens the Import Axon Files dialog box, where you can select Axon (*.gpr) files for import. See "Import" on page 204 and "To import Agilent FE or Axon data files" on page 47.
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 204 and "To import Agilent GEML design files" on page 43. • **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 204 and "To import Axon design files" on page 45.
Genome Build	Opens the Import Genome Build dialog box, where you can import Agilent-provided genome build files. See "Import Genome Build" on page 209 and "To import a genome build" on page 46.
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 210 and "To import tracks" on page 49.
ArrayAttributes	Opens the Import Array Attributes dialog box, where you can select a .txt file for import. See "Import" on page 204 and "To import array attributes" on page 50.

Option	Description
Experiments	Opens the Import Experiments dialog box, where you can select a ZIP format experiment file for import. See "Import" on page 204 and "To import an experiment file" on page 50.
Filters	Opens the Import dialog box, where you select a filter file to import. For more information, see "Import" on page 204 and "To import filters" on page 51.

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 196 and "To export experiments" on page 73.
Filters	Opens the Export Filters dialog box, where you can select one or more filters for export as a single *.xml file. See "Export Filters" on page 197 and "To export filters" on page 75.
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 198 and "To export tracks" on page 74.
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 192.

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 180 and "To create a new experiment" on page 53.

Save Experiment Result Saves ChIP results for the selected experiment.

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 203.

Print Opens the Print window to print the display.

Exit Closes the program.

Preprocessing command ribbon

The Preprocessing command ribbon contains several sets of commands:

- **Normalization** – Configure and apply normalization method(s).
- **Error Model** – Select and configure the error model.
- **Combine** – Fuse array designs and configure how the program combines inter- and intra-array replicate probes.

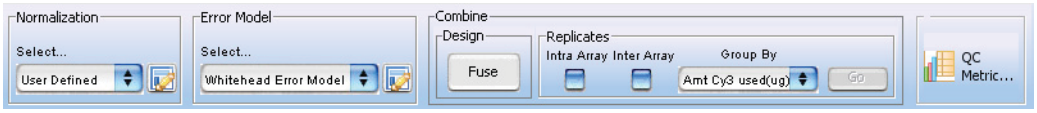

Figure 26 ChIP Preprocessing command ribbon

Normalization

The settings in Normalization control how the program normalizes data before it applies event detection. You normalize data to correct it for known factors such as nonspecific binding, overall variations from one array to another, and dye bias (the tendency for a fluorescent dye molecule to influence binding). See "To apply normalization" on page 103.

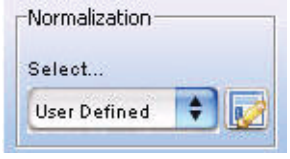

Figure 27 Normalization group box

Select Lets you select one of two choices for the normalization:

User Defined – The program applies user-selected normalization steps and parameters. These steps normalize the selected array data in the current experiment before the program applies the event-detection algorithm. This option also makes the other normalization settings available.

Use FE Output – The program uses the processed feature intensity values available in the output files of the Agilent Feature Extraction program. The program uses these values instead of applying its own normalization algorithms. See "To use processed Feature Extraction output" on page 103. Unless you select this option, the program uses the raw, unprocessed feature intensities.

(Edit button) Opens the Setting Normalization Parameters dialog box, where you can select and configure the program's normalization algorithms. This button is available only if you select **User Defined** as the normalization option. See "Setting Normalization Parameters" on page 238.

Error Model

The error model calculates the likelihood that probes represent binding events, and assigns *p*-values to probes. A *p*-value close to 1 indicates that a probe is unlikely to represent a significant binding event. A very small *p*-value (for example, P < 0.001) indicates that the probe is very likely to represent one.

After the ChIP application uses the error model to assign *p*-values, it then evaluates the *p*-values to make binding calls. See "To apply event detection" on page 111. For a detailed discussion of the statistical algorithms used for error modeling and event detection, see "Error Models" on page 284.

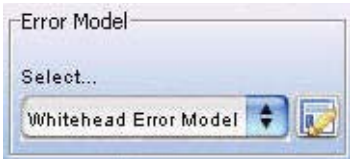

Figure 28 Error Model group box

Select Lets you select one of two options for the error model:

- **Whitehead Error Model** – The program uses this error model by default. Select this model unless you have tried both models and know the other one provides a better match to biological truths and/or positive controls you have available for your specific experiment. For a description of the statistical algorithm used in this model, see "Error Models" on page 284. If you choose this error model, click  to set additional parameters.
- **Use FE Error Model** – Uses the error model from the Agilent Feature Extraction program. Select this model if you have tried both models, and know that this one provides a better match to biological truths and/or positive controls you have available for your specific experiment. If you select this model, the edit button becomes unavailable; you do not need to set any additional parameters.

 (Edit button) Opens the Whitehead Error Model Parameter Settings dialog box, where you can customize the configuration of Whitehead error model. See "Whitehead Error Model Parameter Settings" on page 256.

Combine

The settings and commands in Combine enable you to combine multiple array designs into a larger virtual combined design. They also allow you to combine both intra- and inter-array replicate probes. See "To combine (fuse) arrays" on page 108 and "To combine intra-array replicates" on page 110. For a discussion of the algorithms used to combine replicates, see "Normalization Algorithms" on page 274.

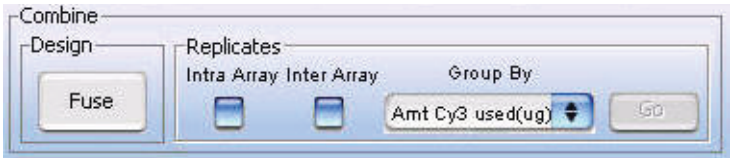

Figure 29 Combine group box

Fuse Opens the Array Set dialog box, where you can view the names of arrays to be fused, set options, and fuse designs together to form a larger virtual design. See "Array Set" on page 172, and "To combine (fuse) arrays" on page 108.

Intra Array Mark this check box to combine intra-array replicates (probes within an array that have the same name). The program combines the replicates when you click **Go**. See "To combine intra-array replicates" on page 110.

Inter Array Mark this check box to combine replicates across multiple arrays. The program uses the array attribute you select in **Group By** and combines replicates from arrays that have the same value selected for that attribute. The program combines the replicates when you click **Go**.

Group By Select an array attribute. The program combines replicates from arrays that have the same value assigned to the attribute you select.

Go Combines replicates according to the other settings under Replicates.

QC Metric

Figure 30 QC Metrics command

Opens the QC Metrics Table for Agilent arrays in the selected experiment, where you can select QC Metrics and display QC Plots or Frequency Distributions. See "QC Metrics Table" on page 219.

Analysis command ribbon

Event Detection

Event detection is the algorithm the program uses to determine binding events. See "To apply event detection" on page 111, and "Peak Detection Algorithms" on page 289. The settings in Event Detection allow you to select and customize the model.

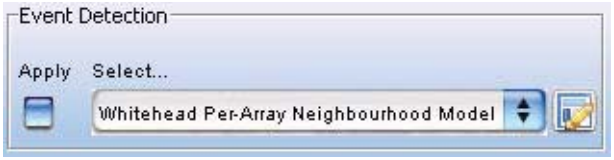

Figure 31 Event Detection group box

Apply Mark this check box to apply the selected algorithm to the selected array data in the active experiment.

Select Select the Event Detection algorithm to use for the analysis.

Opens the Parameter Settings dialog box for the selected model, where you can customize the model. See "Whitehead Per-Array Neighbourhood Model Parameter Settings" on page 257, "Pre-defined Peak Shape detection v2.1" on page 214, and "To apply event detection" on page 111.

Reports command ribbon

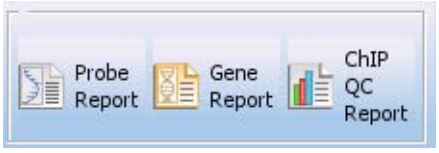

Figure 32 ChIP Reports command ribbon

The Reports command ribbon lets you save detailed information about the probes, genes, and overall statistics of the current experiment result. You use a spreadsheet program or your Internet browser to open the report files that the program generates. This lets you further view, analyze, and organize the results.

Probe Report Opens the Probe Report Settings Parameter Settings dialog box, where you can select a location and a name for the probe report file. "Probe Report Settings Parameter Settings" on page 216.

The probe report makes information about the probes in the current experiment result available in tab-separated value (*.tsv) format. You can use a spreadsheet program to open this file. A probe report contains one row for each probe in the array (or array set). See "Probe Report format" on page 260 for a description of the columns in the report.

Gene Report Opens the Gene Report Settings Parameter Settings dialog box, where you can customize the content of the report, and select a location and a name for the gene report file. See "Gene Report Settings Parameter Settings" on page 202.

The gene report makes information about the genes in the current experiment result available in tab-separated value (*.tsv) format. The gene report contains one row for each probe in an array, grouped by the genes to which the probes bind. It also includes loci represented by probes on the array that are not associated with genes. The program creates gene reports in several formats. See "Gene Report formats" on page 262 for a description of these formats, and the columns in each one. You can view gene reports and perform further analysis on them with a spreadsheet program.

ChIP QC Report Opens the QC Report Settings Parameter Settings dialog box, where you can select a location and a name for the QC report file. See "QC Report Settings Parameter Settings" on page 227.

The QC report summarizes the settings of the current analysis, and the overall statistics of each array. In addition to summary tables, it includes four plots:

- Scatter plot of IP versus WCE
- Scatter plot of enrichment ratio versus intensity
- Histogram of normalized intensity distribution, by channel
- Histogram of normalized $\log_2$ ratio distribution

The program creates the QC report in HTML format, and you can view the report with an Internet browser. For more information about the contents of the report, see "QC report format" on page 263.

View command ribbon

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

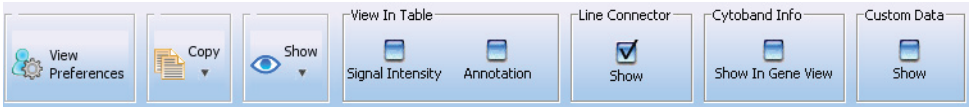

Figure 33 Command ribbon of the View tab

View Preferences Opens the **View Preferences** dialog box, where you can customize the display of your data in the Genomic Viewer. See "View Preferences" on page 252.

Copy 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 75.

Option	Description
All	Copies all panes of the main window to the Clipboard as an image.
Navigator	Isolates and copies the Navigator to the Clipboard as an image.
Tab view	Isolates and copies Tab View to the Clipboard as an image.
Genome view	Isolates and copies Genome View to the Clipboard as an image.
Chromosome view	Isolates and copies Chromosome View to the Clipboard as an image.
Gene view	Isolates and copies Gene View to the Clipboard as an image.

Show Opens a menu with all available elements of the main window. Mark the check box(es) next to the element(s) that you want to display in the 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.

Line Connector Mark **Show** to draw lines between the log ratio data points on the scatter plot in Gene View. Clear the check box to hide the lines.

Cyto band info

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

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 34 Help command ribbon for ChIP

Table 6 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.
Workflow	Opens the *Workflow User Guide*, that describes how to use the Workflow module of Agilent Genomic Workbench to extract image files with Agilent Feature Extraction software and/or analyze data using CGH and ChIP analysis software.
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.

An additional guide is available in the Open Application tab of the program. The *Agilent Genomic Workbench Product Overview Guide* gives an overview of the capabilities within Agilent Genomic Workbench. It also

describes how to start each of the component programs and find help, and how to enter your license information. To open this guide, click the **Open Application** tab, then click **Product Overview**.

Navigator

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

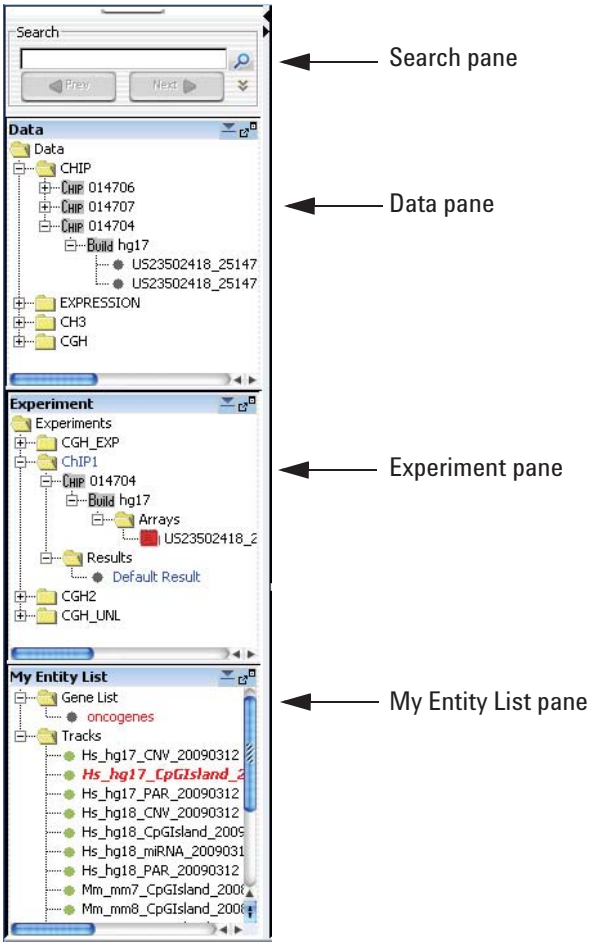

Figure 35 Navigator panes

The Navigator catalogs 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".
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. For the SureSelect Target Enrichment application type, the program shows all bait groups and libraries. 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 141 and "Data pane – actions and shortcut menus" on page 142.
Experiment	Contains the Agilent Genomic Workbench experiments. Experiments are organizational units within the program that contain links to microarray data and design files. In data analysis applications, 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 is correlated with specific genomic locations. You can import, export, and combine tracks, and display them in Gene View alongside your array data and analysis results.

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 63. It also contains several buttons that you can use to detach, hide, show or resize the Navigator.

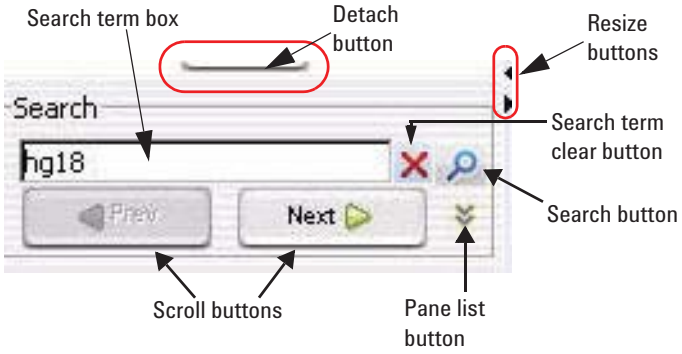

Figure 36 Navigator – Search pane

Detach button Detaches the Navigator from the main window of the program and opens it in a new, separate window.

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

Search term box Provides a box for you to 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".

Panel 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.

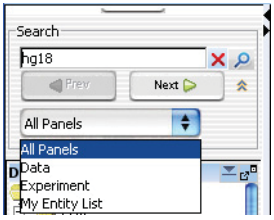

Figure 37 Open pane list

(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 reflect the entire name of the desired items. You can use asterisks (*) as wildcards to represent groups of unspecified characters.

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
	A collapsed 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 associated with the design, organized by genome build.
	A CGH array design. This folder contains array data associated with the design, organized by genome build.
	A gene expression array design. This folder contains array data associated with the design, organized by genome build.
	A ChIP array design. This folder contains array data associated with the design, organized by genome build.
	A genome build folder within a specific design folder. This folder contains arrays associated with 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.
highlighted text	The current result when you search the Navigator. Click **Next** to highlight the next result, and **Previous** to highlight the previous one.
	(Dock out button) Detaches 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 in the program for the selected application type, and any external content that you imported.

Data Folder

- Double-click any folder to expand or collapse it.
- Double-click a data folder (ChIP, Expression, CGH, CH3) to display the designs for that data type.
- Double-click the name of a design folder to display a list of genome builds.
- Double-click the name of a genome build folder to display 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 64. **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 245.
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 184.
QC Metrics	Opens the QC Metrics Table dialog box with the arrays from the selected genome. See "QC Metrics Table" on page 219.
Delete	Opens a Confirm dialog box. If you click **Yes,** the program permanently deletes all of the arrays in this genome build folder. (Not available for read-only builds.)

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 211 and "To display or edit the attribute values of a specific array" on page 61.
QC Metrics	Opens the QC Metrics Table dialog box with the arrays from the selected genome. See "QC Metrics Table" on page 219.
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 associated with the design, organized by genome build.
	A CGH array design. This folder contains array data associated with the design, organized by genome build.
	A gene expression array design. This folder contains array data associated with the design, organized by genome build.
	A ChIP array design. This folder contains array data associated with the design, organized by genome build.
	A genome build folder within a specific design folder. This folder contains arrays associated with 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.
	An array that has been designated as a calibration array.
	An empty folder.
	Data created from a multi-pack array.
blue text	The currently active experiment. All data and results that appear in Chromosome, Gene, and Tab Views are derived from this experiment.
red text	An item that matches the search term in a search.
	(Dock out button) Detaches 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

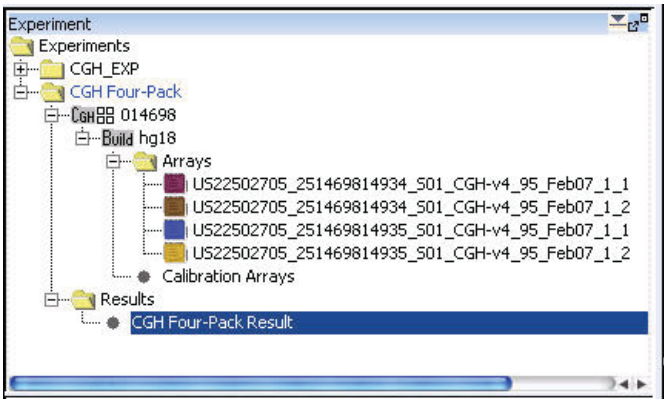

Figure 38 Expanded Experiment Pane

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 180), where you can name the new experiment, and open another dialog box that lets you add microarrays to the experiment. See "To create a new experiment" on page 53.
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 196 and "To export experiments" on page 73.

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. 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 currently selected experiment result in the 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 116. • **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 there are unsaved results, a Confirm dialog box opens with these options: • **Overwrite Current Result** – Replaces the currently selected experiment result in the 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
Save Experiment Result	Opens a dialog box that asks if you want to save the results of the current experiment. When you click OK, one of these dialog boxes opens: • If you have not yet saved a result for the experiment, the Save experiment result dialog box opens, where you can type a name for the experiment result. • If you have previously saved the result, the confirm dialog box appears, which lets you overwrite the current result, create a new result, or discard any changes. See "Confirm" on page 177. • If you have previously saved the result, and no changes have occurred, a message informs you that the current experimental condition is the same as an existing one.
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 189.
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 228. 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 196, and "To export experiments" on page 73.
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 192 and "To export array attributes" on page 72.
QC Metrics	Opens the QC Metrics table for all arrays in the experiment. The QC Metrics Table is available only for Agilent arrays. See"QC Metrics Table" on page 219, and "To display QC metrics of arrays and set array QC status" on page 88.
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 187.
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 188.

Option	Description
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.
Expand Node	Expands the selected node to display all folders and their contents.
Collapse Node	Closes all folders for the selected node.

Design Folder

- In the folder of a specific experiment, right-click the name of a design to display 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	This is not used.
QC Metrics	Opens the QC Metrics table for all arrays in the genome build. The QC Metrics Table is available only for Agilent arrays. See"QC Metrics Table" on page 219, and "To display QC metrics of arrays and set array QC status" on page 88.

Option	Description
Save As Text File	Opens the Save Design dialog box, where you can save all of the data associated with 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 remain 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 from the Selected Arrays tab in Tab View.
Select for Calibration	(Available only in the Arrays folder.) Designates the array as a calibration array. Within the specific experiment, the program moves the name of the array to the Calibration Arrays folder. Within the specific design tab in Tab View, the program also moves the data for the array from the Arrays tab to the Calibration Arrays tab.
Deselect from Calibration	(Available only within the **Calibration Arrays** folder.) Designates the array as a "regular" non-calibration array. Within the specific experiment, the program moves the name of the array to the Arrays folder. Within the specific design tab in Tab View, the program also moves the data for the array from the Calibration Arrays tab to the Arrays tab.

Option	Description
Rename	Opens an Input dialog box, where you can type a new name for the experiment. Click **OK** to accept the new name for the 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 211. 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 61.
QC Metrics	Opens the QC Metrics table for the array(s) you have selected. The QC Metrics Table is available only for Agilent arrays. See"QC Metrics Table" on page 219, and "To display QC metrics of arrays and set array QC status" on page 88.
Edit Array Color	Opens the Select Color dialog box, where you can select a display color for the array. See "Select Color" on page 230.
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 188.

Results Folder

- In the **Results** folder of an experiment, right-click the name of a saved result to display these options:

Option	Description
Restore result	Replaces the result that appears in Genome, Chromosome, Gene, and Tab Views with the saved result. See "To restore a saved result" on page 117. The experiment associated with the saved result must be the selected experiment.
Rename	Opens an Input dialog box. Type a new name for the result, then click **OK**.
Delete	Opens a Confirm dialog box that asks if you want to delete the result. Click **Yes** to delete the result from the experiment.
Show Properties	Opens a Properties dialog box that you can use to view or edit a description of the result, and to view other attributes of the result. See "Properties (of an experiment result)" on page 217.

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 an unapplied gene list 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.

Item	Comments
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) Detaches 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 display 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 "To import a gene list" on page 67 and "Import" on page 204.
- 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 200 and "To select gene list display color" on page 91.
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 74.

Option	Description
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 169 and "To add one gene list to another" on page 68.
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) Limits 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 210 and "To import tracks" on page 49.
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 198 and "To export tracks" on page 74.
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 173 and "To combine tracks" on page 69.

- 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 alongside the data and results of the selected experiment.
Show In UCSC	Opens the UCSC Genome Browser in your Web browser and uploads the track.
View Details	Opens a table listing all the chromosome locations selected in the track. See "Track" on page 243.
Edit Color	Opens the Select Color dialog box, where you can select a color for display of the selected track. See "Select Color" on page 230.
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

The Genomic Viewer is the display for the Agilent Genomic Workbench CGH, ChIP and CH3 applications. It includes three graphical views: Genome, Chromosome and Gene Views. It also contains Tab View and the Cursor.

Genome View

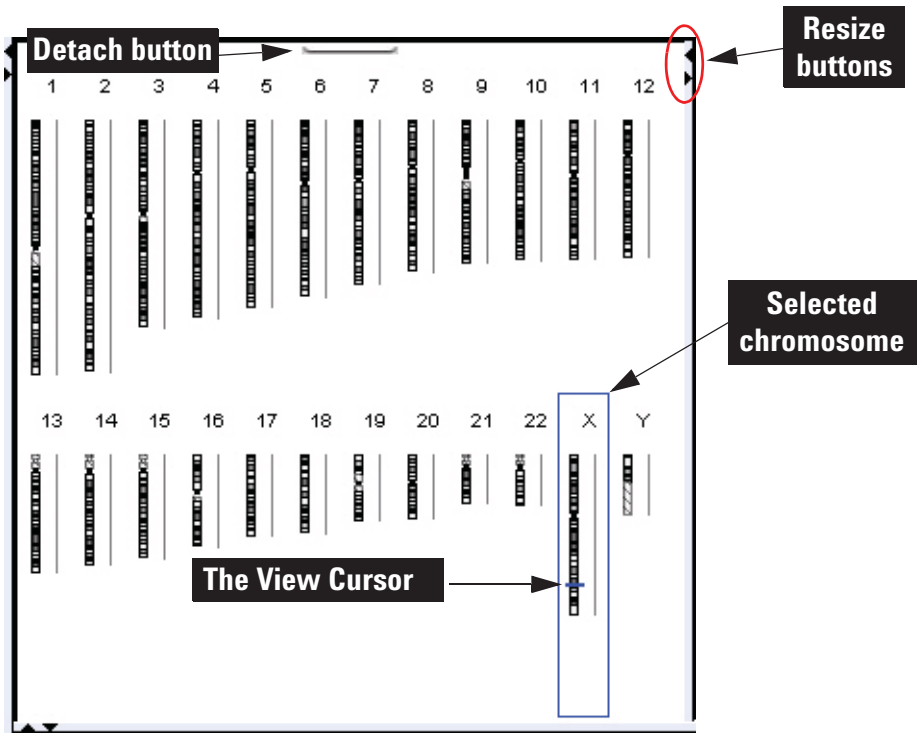

Figure 39 Genome View (vertical orientation), with human chromosomes. The X chromosome is selected.

Genome View shows pictures of each of the distinct types of chromosomes in the selected genome. A blue box is drawn around the currently selected chromosome, and the 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 associated with it. The specific location in which you click the chromosome sets the position of the cursor. See "The View Cursor" on page 163.
- On the selected chromosome, click anywhere to reposition the cursor. See "The View Cursor" on page 163. This also repositions 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 252.
- 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

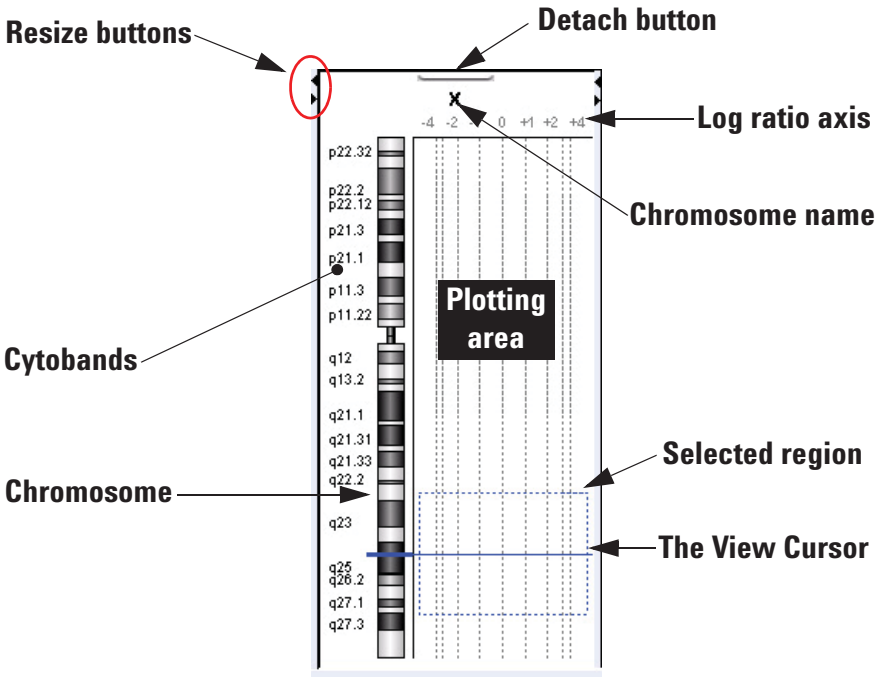

Figure 40 Chromosome View, human X chromosome shown

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

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

Chromosome View actions and shortcut menus

- Click a cytoband, any part of the chromosome, or anywhere in the plotting area to reposition the cursor at that location. See "The View Cursor" on page 163.
- Drag the pointer over any part of the plotting area to select a chromosomal region for display in Gene View. Drag parallel to the chromosome. This also repositions the cursor to the center of the selected region. See "The View Cursor" on page 163.
- 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 252.
- 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 an inside border of Chromosome View to resize the view.
- 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

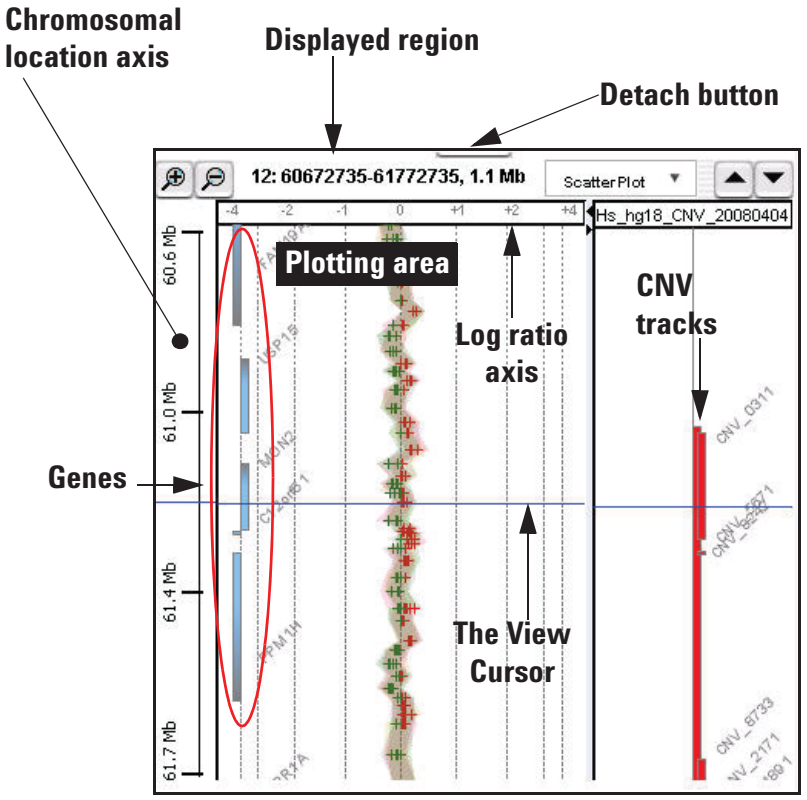

Figure 41 Gene View, with log ratio data from an experiment and CNV tracks

Gene View shows a more detailed view of the chromosomal region you select in Chromosome View. See "Chromosome View" on page 157.

- Regions occupied by genes appear as small blue boxes. Gene names appear nearby. You can customize the appearance of gene names. Also, you can use a gene list to highlight genes of interest, or to limit the genes that appear to those in the list. 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. Points appear in colors that you can customize. 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 163.
- 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 93.

Scatter Plot

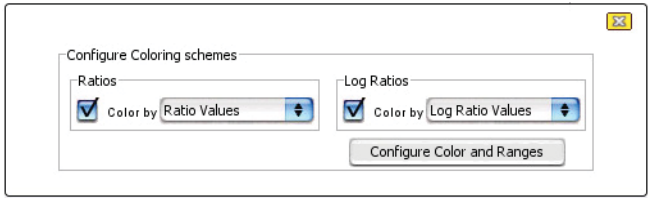

Figure 42 Scatter Plot command group in ChIP Gene View

You access the scatter plot command group through Gene View or through the View Preferences dialog box from the View tab. The commands differ for each Agilent Genomic Workbench application. Scatter plots appear in the Chromosome and Gene Views, but only if they have been activated in the View Preferences dialog box.

Ratios Mark the box to enable the Ratios scatter plot. The only choice of data for the plot is Ratio Values.

Log Ratios Mark the box to enable the Log Ratios scatter plot. Selections let you color-code the Log Ratios by Log Ratio Values or Probe Score Values.

Configure Color and Ranges Opens the Configure Coloring Ranges and Shades dialog box, where you can set up the colors and ranges for the Ratios and Log Ratios scatter plots. For more information, see "Configure Coloring Ranges and Shades" on page 175.

Gene View buttons


Zooms in to see a smaller region in more detail.


Zooms out to see a larger region in less detail.


(Available when Gene View is in vertical orientation.) Scrolls up through the genes and data to lower-numbered chromosomal coordinates.


(Available when Gene View is in vertical orientation.) Scrolls down through the genes and data to higher-numbered chromosomal coordinates.


(Available when Gene View is in horizontal orientation.) Scrolls left through the genes and data to lower-numbered chromosomal coordinates.


(Available when Gene View is in horizontal orientation.) Scrolls right through the genes and data to higher-numbered chromosomal coordinates.


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

Gene View shortcut menu and other actions

- Click anywhere in the plotting area of Gene View to move the cursor to that location. See "The View Cursor" on page 163.
- Drag an inside border of Gene View to resize the view.
- Right-click anywhere in the plotting 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 currently selected (or another) chromosomal region. See "Create Gene List" on page 181 and "To create a gene list" on page 66.
Show in UCSC	Opens the View Coordinates in UCSC Browser dialog box where you select track information for display in the UCSC (University of California at Santa Cruz) Genome Browser. You can then view the track. See "View coordinates in UCSC browser" on page 250.
Search probes in eArray	Opens the Search probes in eArray dialog box, where you can start a search of the Agilent eArray web site for probes in the selected (or another) chromosomal region. See "Sample Attributes" on page 228.
User Preferences	Opens the User Preferences dialog box, where you can set user preferences on three separate tabs. See "User Preferences" on page 245.
View Preferences	Opens the View Preferences dialog box, where you set preferences for the Genomic Viewer. See "View Preferences" on page 252.

The View Cursor

The View cursor reflects the center of the current chromosomal region 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 plotting area of the view.
- In Gene View, it is a blue bar that appears across the plotting area and tracks of the view.

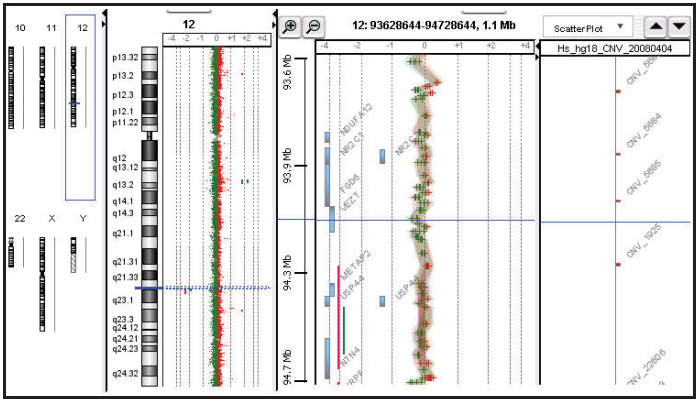

Figure 43 Genomic viewer showing cursors

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

- In Genome View, click anywhere on a chromosome to move the cursor to that location.
- In Chromosome View, click a cytoband name, part of the chromosome, or anywhere in the 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 in Gene View is the same cursor used for tracks.

- In Tab View, click a row of a data table to move the cursor to the chromosomal location associated with that row.

Tab View

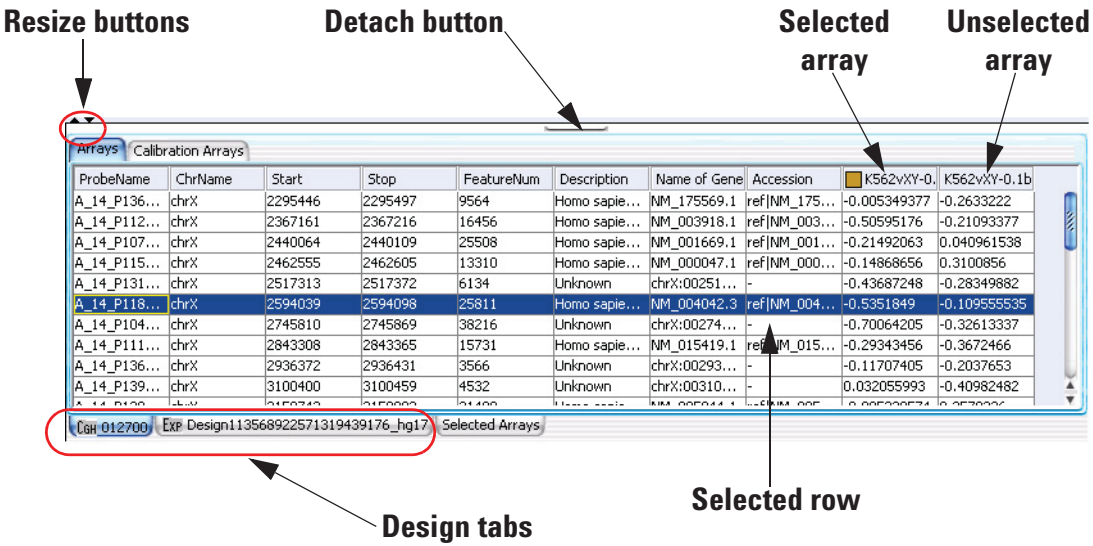

Figure 44 Tab View

- 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 44 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.
- CGH+ – A CGH+SNP 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 Tab View expands the view. The other button restores the view to its original size. See the *Product Overview Guide* for more information.

(**Detach** button) Removes Tab View from the main window, and opens it in a separate window. See the *Product Overview Guide* for more information.

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 65.
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 187 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 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 188 and "To change the order of arrays in an experiment" on page 56.
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 230.

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

Option	Description
Find in Column	Opens the Find in column dialog box, where you can search for a specific text string within the column you clicked. See "Find in column" on page 199.
Google LocusLink PubMed UCSC HG15(April '03) UCSC HG16(July'03) UCSC HG17(May'04) UCSC HG18(March'06) UCSC mm8(Feb'06) UCSC mm9(July'07) DGV(hg18) GO KEGG(HUMAN)	Opens your Web browser, and passes the column entry you clicked as a search string to the selected site. The UCSC links search the indicated University of California, Santa Cruz database related to the indicated genome build. See "To search the Web for information on probes in Tab View" on page 99.
Customize Link	Opens the Customize Search link dialog box, where you can create or edit a custom Web link that appears in this shortcut menu. When you click a custom link, the program opens your Web browser, and passes the column entry you clicked as a search string to the site. See "Customize Search Link" on page 183 and "To create a custom Web search link" on page 99.
(other options)	If other options appear in this shortcut menu, they are custom Web search links. Click them to open your Web browser, and pass the column entry you clicked as a search string to the site.

Status Bar

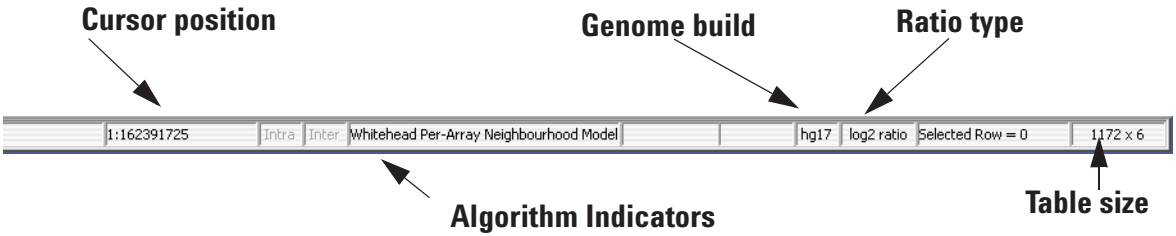

Figure 45 Status bar

The Status Bar (Figure 45) displays information related to the currently displayed data.

Cursor position The chromosomal location of the cursor. See "The View Cursor" on page 163.

Genome build The genome build associated with the currently 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 currently displayed data table that is selected. The location of the cursor is approximately the chromosomal location associated with this row.

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

Dialog Boxes

This section describes the dialog boxes specific to the interactive mode of the ChIP program. They are listed in alphabetical order by title. Many dialog boxes in interactive mode also appear when you use the Genomic Viewer.

Add Gene List <name> to

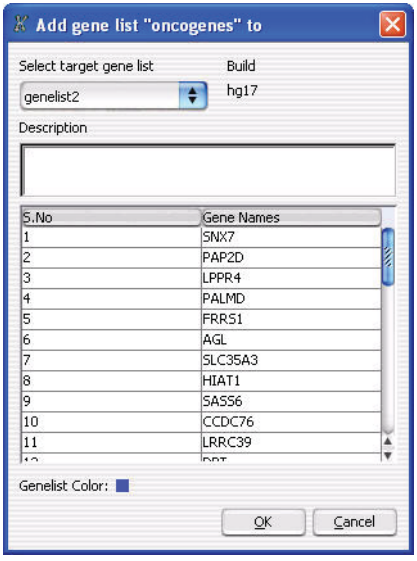

Figure 46 Add Gene List <name> to

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

To open: In the My Entity List 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 associated with 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.

Genelist Color (Read-only) The current 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

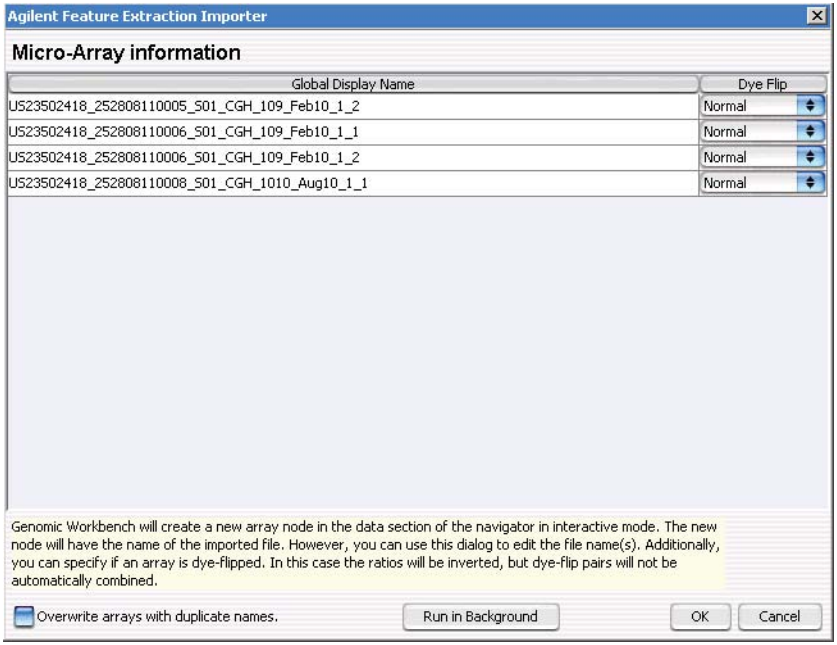

Figure 47 Agilent Feature Extraction Importer

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

To open: In the 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 you are more likely to recognize or remember.

Dye Flip For each array:

Select **Normal** if:

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

Select **Flipped** if:

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

The program does not combine dye-flip pairs.

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

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

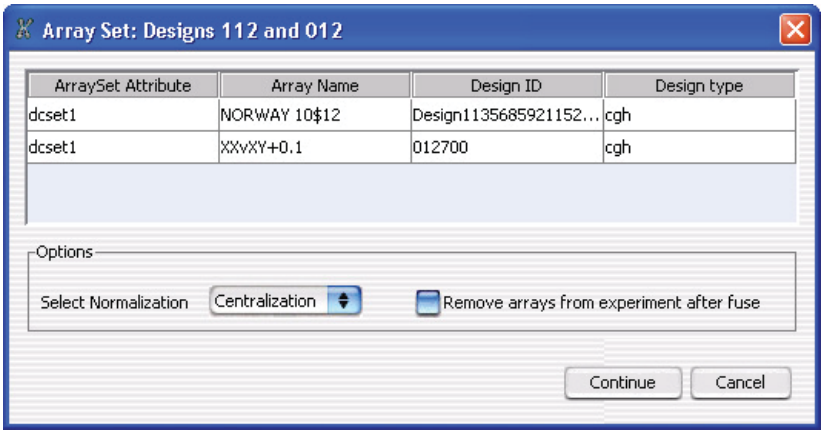

Figure 48 Array Set dialog box

Purpose: Displays the names of arrays to be fused in the analysis.

To open: The array set dialog box opens when you fuse designs. Click the **Fuse** button in the Preprocessing command ribbon. See "To combine (fuse) arrays" on page 108.

Table Displays the arrays to be fused, arranged by their values for the ArraySet attribute. The program creates a separate fused array for each group of arrays with a given value for ArraySet. The table also lists the design ID associated with each array, and the design type.

Select Normalization Currently, the program always applies the Centralization normalization algorithm to the arrays in each fused design. See "Normalization Algorithms" on page 274.

Remove arrays from experiment after fuse Deletes the original unfused arrays after creating fused arrays. This reduces the duplication of data within the experiment.

Continue Click to create fused designs using the selected options.

Cancel Cancels any selections, and closes the dialog box.

Combine Tracks

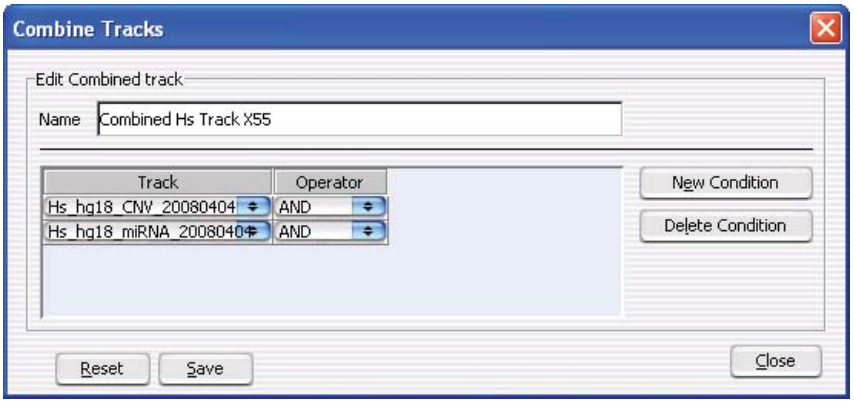

Figure 49 Combine Tracks dialog box

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 69.

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

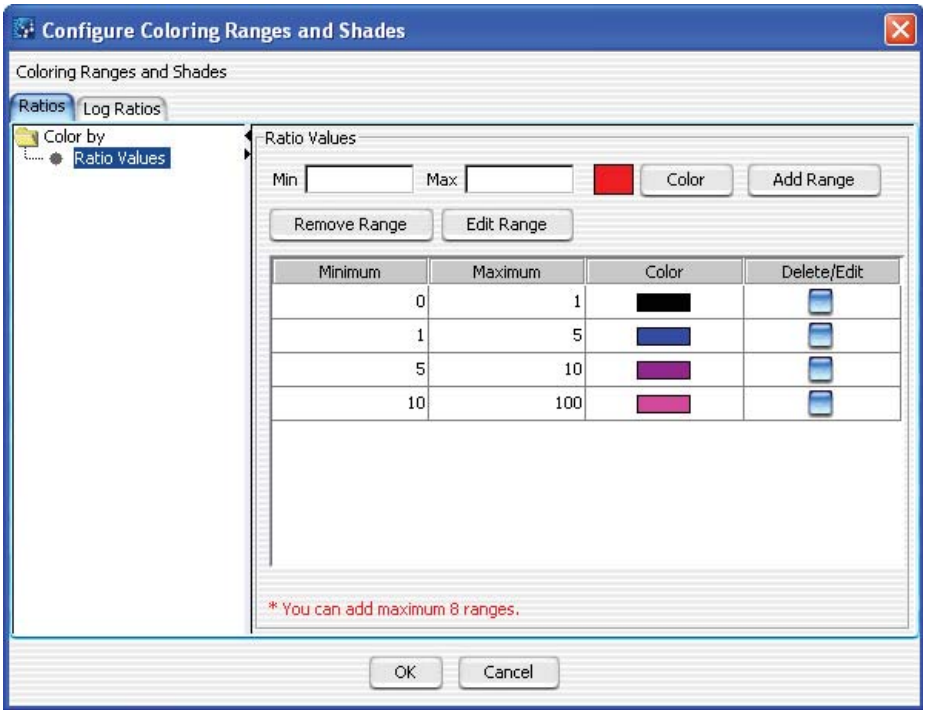

Figure 50 Set Colors and Ranges dialog box

Purpose: This dialog box is used to type ranges and select colors for scatter plot options. Tabs show selections for Ratios and Log Ratio plot 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 Scatter Plot, click **Configure.**

Table 7 Ratios Plot Parameters

Scale Parameter	Description
Ratio	
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 230 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 8 Log Ratios Plot Parameters

Scale 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 230 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.

Table 8 Log Ratios Plot Parameters (continued)

Scale Parameter	Description
Range table	This table displays the defined ranges, including minimum and maximum values, color for each range, and Edit/Delete selection.
Probe Score Values	
Min	Type a minimum value for the range.
Max	Type a maximum value for the range.
Color	Click to open the Select Color dialog box, where you can select the color you want to display for this range. See "Select Color" on page 230 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

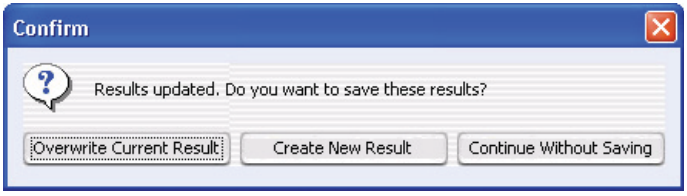

Figure 51 Confirm dialog box

Purpose: Selects how the state of the current analysis should be retained when switching between applications.

To open: Click the **Switch Application** button when an experiment is currently selected.

Overwrite Current Result Replaces the currently selected experiment result in the 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. See "To save a result" on page 116.

Continue Without Saving The program does not save the result that appears on your screen.

Confirm overwrite

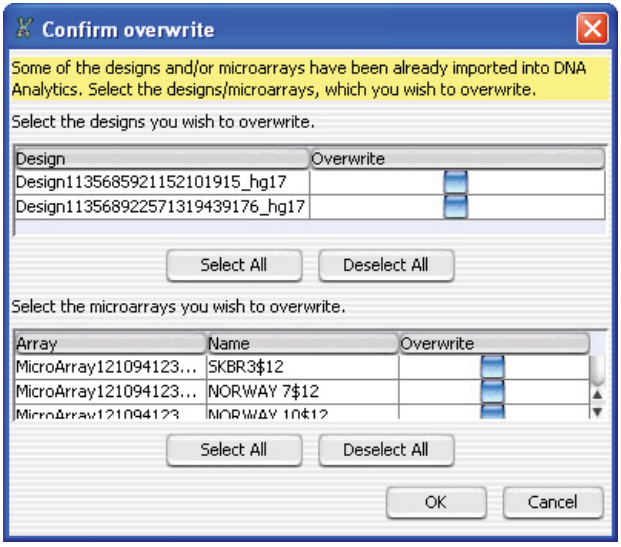

Figure 52 Confirm overwrite dialog box

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 50.

Select the designs you wish 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 next to each existing 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 you wish to overwrite

Array Identification number for the array.

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 206.

Create Experiment

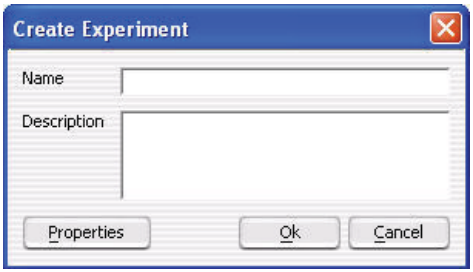

Figure 53 Create Experiment dialog box

Purpose: To create an organizational unit in the Experiment pane of the Navigator to link to array data for viewing and analysis.

To open: In the Experiment pane of the Navigator, right-click the **Experiments** folder, then click **New Experiment,** or click **Home > Create Experiment.**

Name Type a name for your new experiment.

Description Brief description of your experiment with information that helps you identify it.

Properties Click to access the Experiment Properties dialog box, where you can select microarrays to add to your new experiment. See "Experiment Properties" on page 189.

NOTE If you do not add data to your new experiment from arrays in the Experiment Properties dialog box, the program creates an empty experiment.

Create Gene List

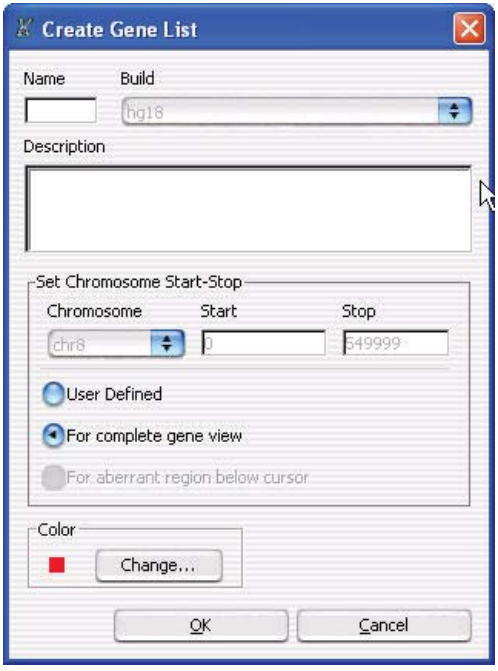

Figure 54 Create Gene List

Purpose: Creates a list that you can use to highlight or limit the genes that appear in Gene View.

To open: Right-click Gene View, then click **Create Gene List**.

Name Type a name for the gene list.

Build Select the genome build for the genes to be selected for list.

Description Describe the type or nature of the genes in the list.

Set Chromosome Start-Stop

You can select a chromosomal region in Chromosome View before you open the Create Gene List dialog box. If you do, the program presets the chromosomal coordinates of the region.

User Defined Select this option to use the Chromosome, Start, and Stop options to limit the gene list. This option overrides any previously selected region.

For complete gene view Selects all of the genes that currently appear in Gene View.

Chromosome (Available if you select User Defined.) Lets you select the chromosome for the gene list.

Start (Available if you select User Defined.) Lets you set the beginning of the genomic region for the gene list.

Stop (Available if you select User Defined.) Lets you set the end of the genomic region for the gene list.

Color

Change Click to change the display color of the gene list in Gene View and in the Navigator.

Customize Search Link

Figure 55 Customize Search Link dialog box

Purpose: Lets you create a custom Web search link in the shortcut menu that appears when you right-click an entry in Tab View. The link opens the URL of your choice, and passes the entry to it as a search string. See "To create a custom Web search link" on page 99.

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

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

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

For example, this URL passes the selected tab view entry to google.com:

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

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

Update Saves the settings in the dialog box.

Delete Deletes the currently selected custom Web search link.

Close Closes the dialog box.

Design Properties

Purpose: Provides general and detailed information about a given microarray design. See "To display the properties of a specific design" on page 64.

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.

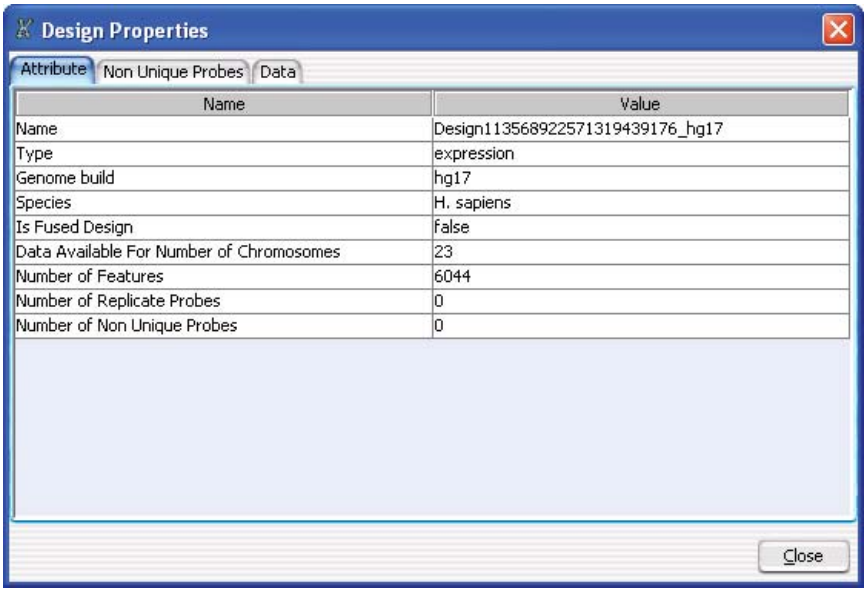

Figure 56 Design Properties dialog box – Attribute tab

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.

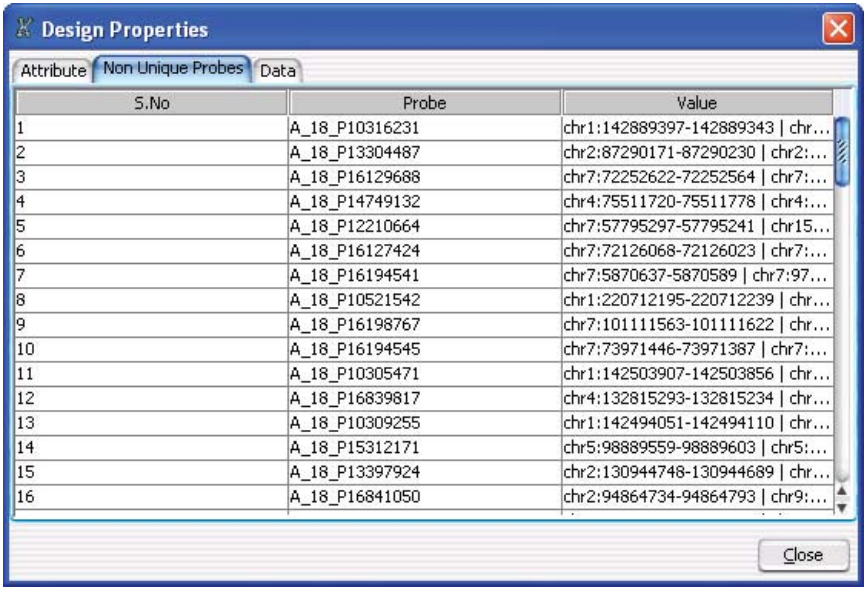

Figure 57 Design Properties dialog box – Non Unique Probes tab

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

Probe The name of each nonunique probe.

Value The chromosomal locations to which each of the probes binds. Because these are nonunique probes, at least two locations appear for each probe.

Data tab

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

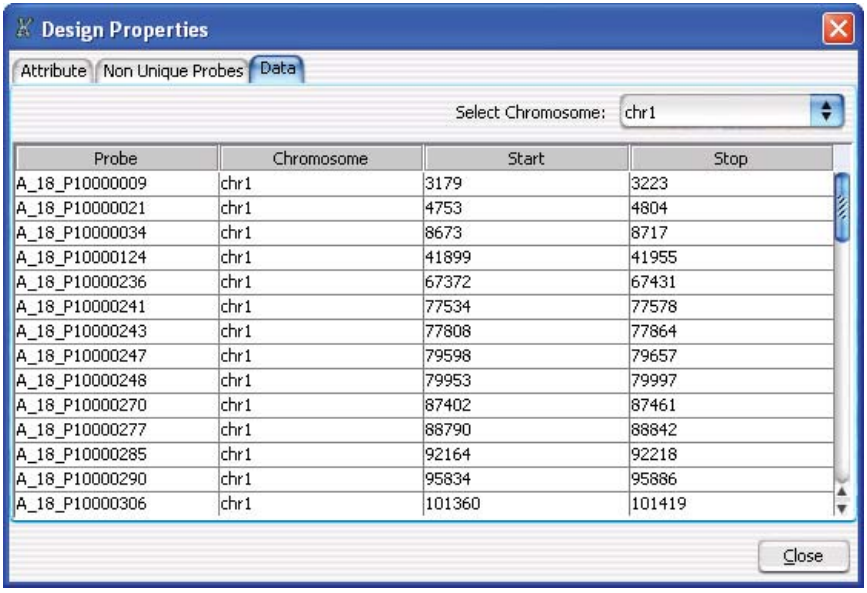

Figure 58 Design Properties dialog box – Data tab

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

Probe The name (Probe ID) of each probe.

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

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

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

Edit Array Color

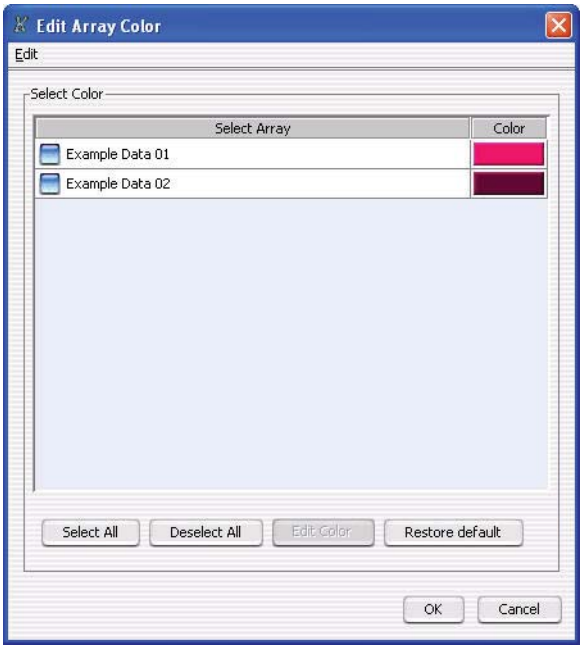

Figure 59 Edit Array Color dialog box

Purpose: To change the display color of the arrays in an experiment

To open: Right-click an experiment name, then click **Edit Array Color.**

Select Array Mark the check box for the array(s) whose color you want to change.

Color Each colored rectangle opens the Select Color dialog box, where you can change the display color for the specific array.

Select All Marks all of the check boxes.

Deselect All Clears all of the check boxes.

Edit Color Opens the Select Color dialog box, where you can change the color assigned to all of the selected array(s). This assigns the same color to all selected arrays.

Restore default Click to restore the original color(s) to the array(s) in the experiment.

Edit Array Order

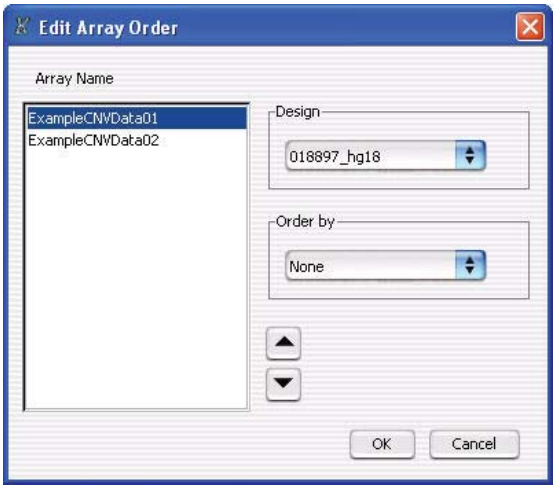

Figure 60 Edit Array Order dialog box

Purpose: Lets you change the order of the arrays in an experiment

To open: Right-click an experiment name, then click **Edit Array Order.**

Array Name The arrays in the selected design, listed in the order that they appear in the Navigator. To move an array up or down in the list, you click the name of the array, then click the Move Up and Move Down buttons. The order of arrays in the list selects the left-to-right display order of the arrays in Tab View. The array order also selects the order in which individual scatter plots appear in Gene View.

Design Select the desired design from the list.

Order by Displays the attributes associated with the arrays. Select an attribute to use for ordering the list. The arrays are reordered based on their respective values for that attribute.

(Move Up button) Moves the selected array up in the list.

(Move Down button) Moves the selected array down in the list.

OK Applies the array order to the active experiment.

Cancel Closes the dialog box without changing the array order.

Experiment Properties

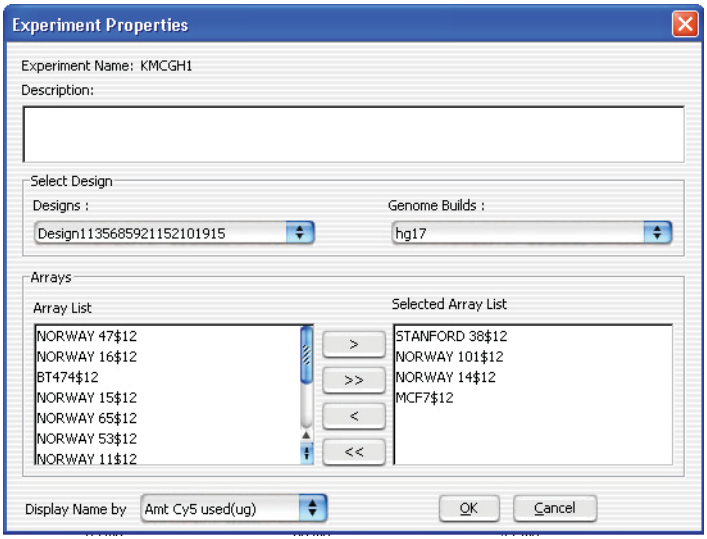

Figure 61 Experiment Properties dialog box

Purpose: To select the arrays to be linked 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, then click **Show Properties.**

Experiment Name The name of the selected experiment appears automatically.

Description Displays the description of the experiment you typed when you created the experiment.

Select Design

Designs From the list, select the design whose arrays you want to add to the experiment.

Genome Builds From the list, 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 this 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 this experiment.

Moves the selected arrays in Array List to the Selected Arrays List. You can move arrays from as many designs as you like, as long as they are all associated with the same genome build.

Moves all of the arrays in Array List to the Selected Arrays 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 to display the array names 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 selected arrays in the Selected Array List to the experiment, and closes the dialog box.

Cancel Closes the dialog box without adding any arrays to the experiment.

Export

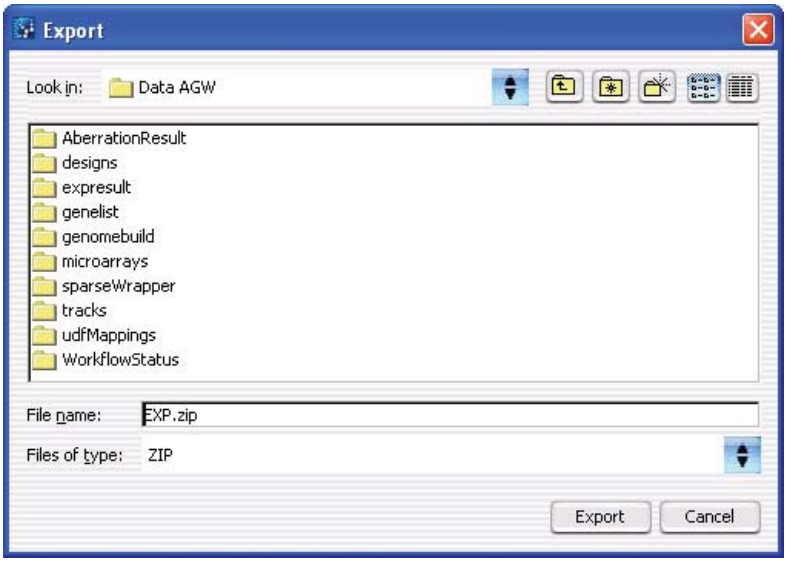

Figure 62 Export dialog box – Several types of file exports use this dialog box. This specific example exports selected experiment(s) as a ZIP format file.

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 73, "To export tracks" on page 74, or "To export array attributes" on page 72.

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.

Moves to 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 details of 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*. The program limits listed files to the type selected in *Files of type*. To select a file, click its name. To open a folder or other location, double-click its name.

File name Displays the name of the file where 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 Limits the files listed in the main pane to those of the appropriate type for your specific kind of export. To show all files, click  , then select **All Files.**

Export Saves the selected content to the location specified 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 72.

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

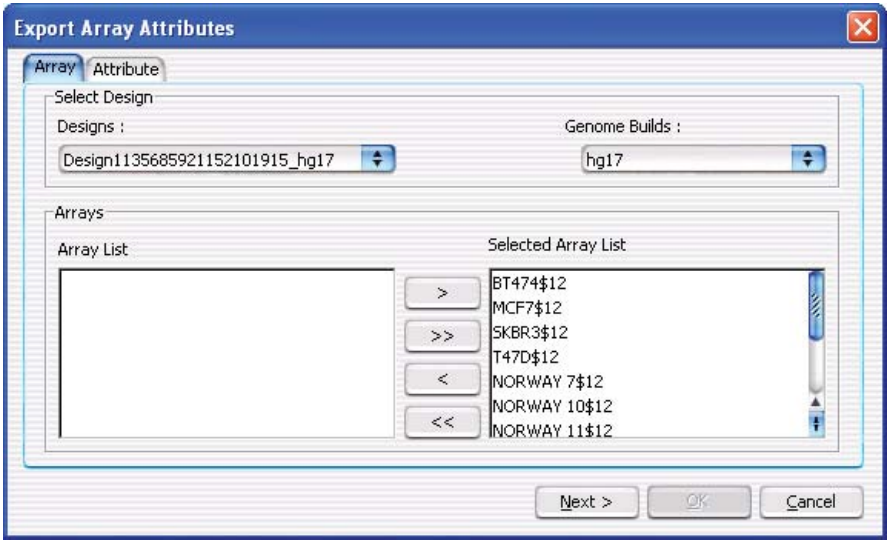

Figure 63 Export Array Attributes dialog box – Array tab

Select Design

Designs Displays all of the designs available in the program. Select the design associated with arrays whose attributes you want to export.

Genome Builds Displays the genome build(s) associated with the design. Select the desired genome build to display the arrays.

Arrays

Array List Displays the arrays in the selected design.

- To select an array for subsequent transfer 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 this experiment.

> Moves the selected arrays in Array List to the Selected Array List. You can move arrays from as many designs as you like, as long as they are all associated with 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.

Next Moves to the Attribute tab for attribute removal. See "Attribute tab" on page 194.

Cancel Closes the dialog box without selecting any array attributes for export.

Attribute tab

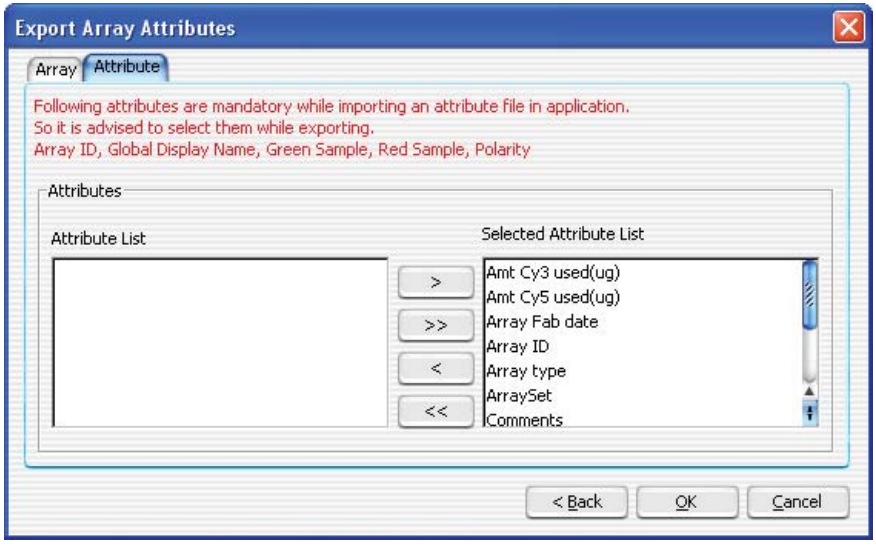

Figure 64 Export Array Attributes dialog box – Attribute tab

Attributes

Selected Attribute List For the selected arrays, lists the attributes to be exported when you click **OK**.

- To select an array for subsequent removal to the Attribute 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.

NOTE You must select the following mandatory attributes, or else the attribute file cannot be imported at a later time: Array ID, Global Display Name, Green Sample, Red Sample, Polarity.

Attribute List Displays the attributes that are not exported for the selected arrays.

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

<< 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 191.

Cancel Closes the dialog box without adding any attributes to the list to be exported.

Export Experiments

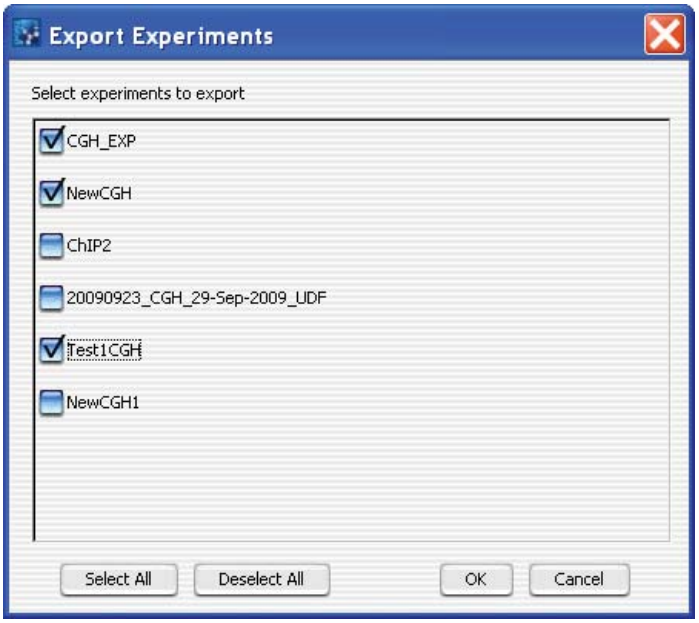

Figure 65 Export Experiments dialog box

Purpose: Lets you select experiments for export. The program exports all array designs and data associated with 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 73.

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 191.

Cancel Cancels the export and closes the dialog box.

Export Filters

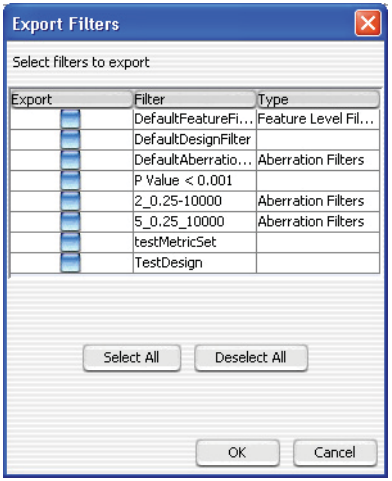

Figure 66 Export Filters dialog box

Purpose: Lets you select feature-level, array-level, and/or design filters to export as a single *.xml file. You can create and use filters only if you have a DNA Analytics application license. See "To export filters" on page 75.

To open: In the **Home** tab, click **Export > Filters.**

Select filters to export Displays all of the filters available in the program. The table has these columns:

- Export – Mark the check box for each filter to export.
- Filter – The name of each filter.
- Type – The type of content to which the program applies each filter.

Select All Selects all available filters for export.

Deselect All Clears all of the check boxes under Select filters to export.

OK Opens the Export dialog box, where you can select a location for the exported *.xml file of filters. See "Export" on page 191.

Cancel Cancels the export and closes the dialog box.

Export Tracks

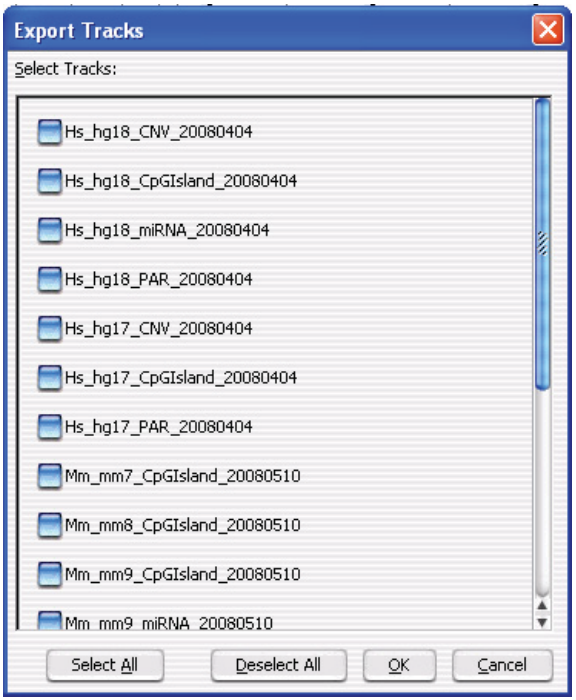

Figure 67 Export Tracks dialog box

Purpose: Lets you select tracks to export as a single BED format file. See "To export tracks" on page 74.

To open: In the **Home** tab, click **Export > Tracks.**

Select Tracks Displays all of the tracks 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 191.

Cancel Cancels the export and closes the dialog box.

Find in column

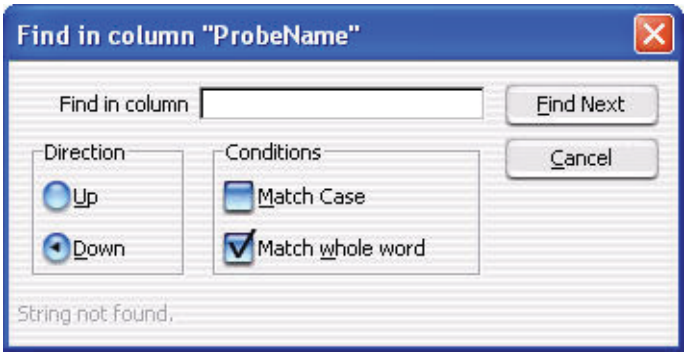

Figure 68 Find in column dialog box

Purpose: Lets you set search parameters for a specific column entry for the selected chromosome. Based on these parameters, the program highlights the row of the first entry that matches. The cursor then moves to the location selected 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 that you want to find.

Direction Select a search direction:

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

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 selected in the row that contains the entry.

Cancel Closes the dialog box.

Gene List

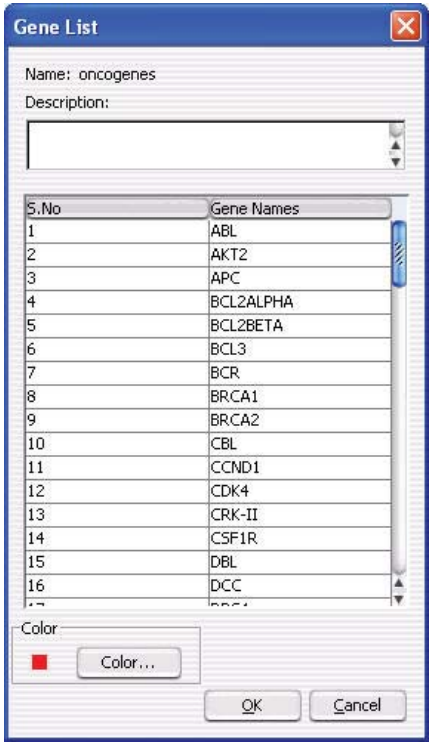

Figure 69 GeneList dialog box

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 67.

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.

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.

Gene Report Settings Parameter Settings

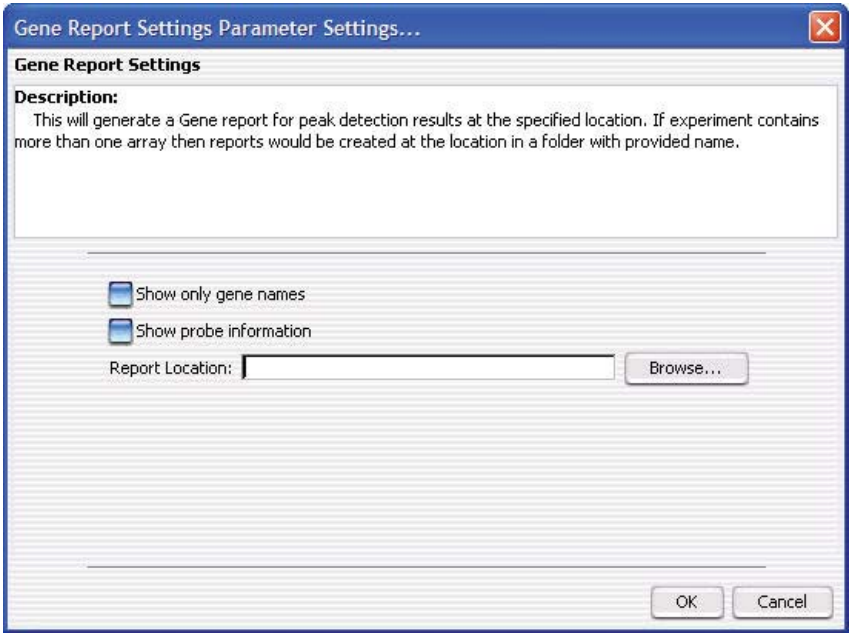

Figure 70 Gene Report Settings Parameter Settings dialog box

Purpose: The Gene Report Settings Parameter Settings dialog box lets you set options for the content and location of the gene report. The gene report contains one row for each probe in an array, grouped by the genes to which the probes bind. It is a tab-separated value (*.tsv) file that you can open and analyze further with a spreadsheet program. For multiple arrays, the program creates a separate *.tsv file for each array.

To open: Click **Reports > Gene Report.**

The program creates gene reports in several formats. See "Gene Report formats" on page 262 for descriptions of the column content in each.

Show only gene names If you mark this check box, the resulting gene report contains only accession numbers of genes (or chromosomal locations for probe loci not associated with genes). A mark in this check box overrides a mark in Show probe information.

Show probe information If you mark this check box, the resulting gene report contains additional information about the probes in the array.

Report Location The location where the program saves the probe report. You can type a location in the text box, or you can click **Browse** to select a location.

Browse Opens the Select report folder dialog box, where you can select a location for the gene report. See "Select report folder" on page 234.

OK Creates the gene report.

Cancel Closes the dialog box without creating a gene report.

Go To Gene/Genomic Location

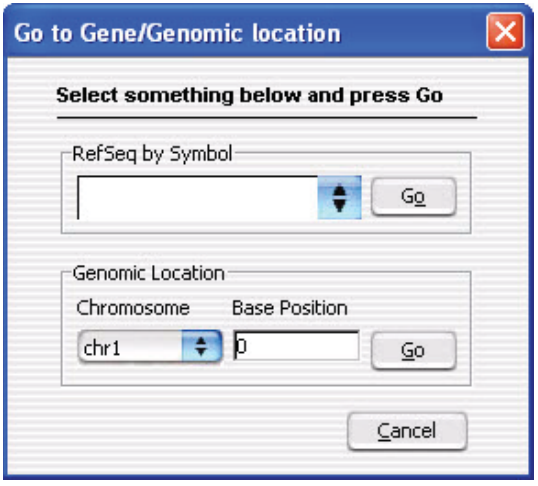

Figure 71 Go To Gene/Genomic location dialog box

Purpose: To find a specific gene location in Gene View based on either its associated RefSeq Symbol or its specific genomic location.

To open: Click **Home > Go to Gene/Genomic location.**

RefSeq by Symbol Select the Reference Sequence accession symbol from NCBI, then 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

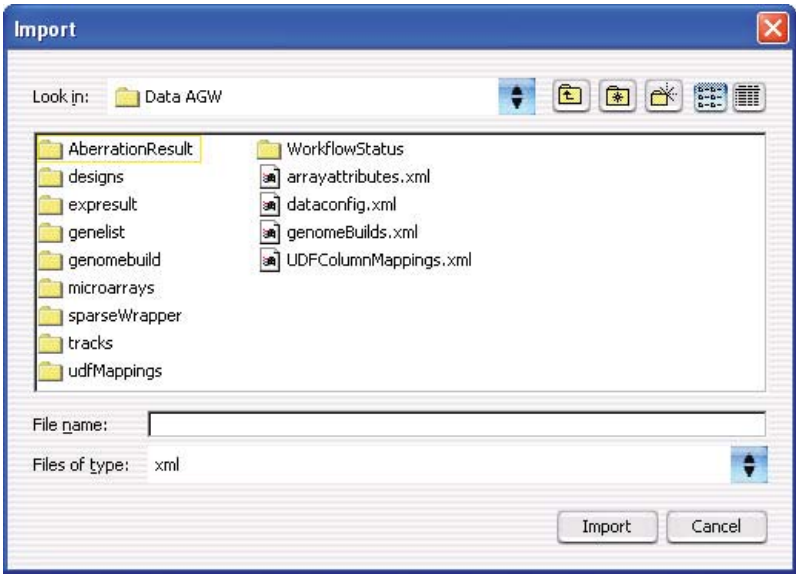

Figure 72 Import dialog box

Purpose: Lets you select files for import 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 to be imported 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 that you select for import.

Files of type The program limits the list of files to the specific types expected for the import. 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
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 select the content for import. See the instructions for each specific type of import in Chapter 2.

Cancel Cancels the import and closes the dialog box.

Import (experiments)

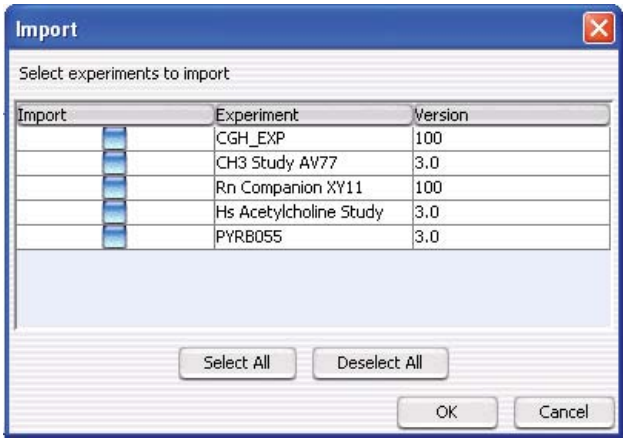

Figure 73 Import dialog box (for experiments)

Purpose: Lets you select the specific experiments within a ZIP format experiment file to load into the program. See "To import an experiment file" on page 50.

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 next to the experiment(s) that you want to import.
- **Experiment** – The names of the experiments available for import in the ZIP format experiment file.
- **Version** – Version of software in which the experiment was created.

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 178.

Cancel Cancels the import and closes the dialog box.

Import GEML design files

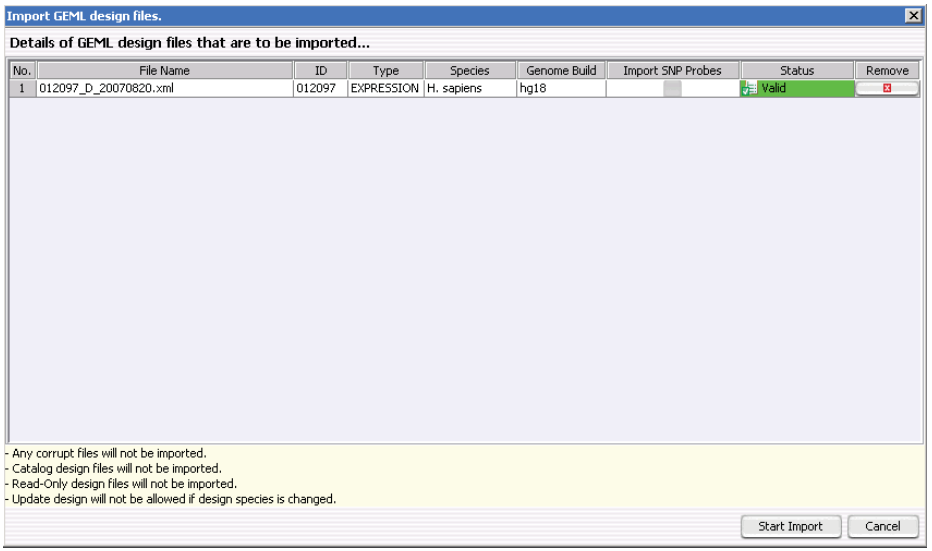

Figure 74 Import GEML design files dialog box

Purpose: Lets you view general identifying information about the design and to remove any files that you do not 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) to be imported.

ID The Agilent ID number for the design file.

Type The application type, which can be 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

- If a design file passes validation, the Status column shows **Valid** in green.
- If the design is an Agilent Catalog design, and is not yet downloaded from the eArray Web site, the Status shows **Not Allowed** in red. You must download the file from the eArray Web site.
- If a design and build is already in the database, the Status shows **Overwrite** in yellow. If you continue, the imported design replaces the design in the database.
- If a design is already in the database, but has a different build, the Status shows **Update** in green. If you continue, this build of the design will be added to the database. The existing design build will not be overwritten.
- If a design file fails validation, **Corrupt** appears in the Status column beside it, and the program will not import the file.

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

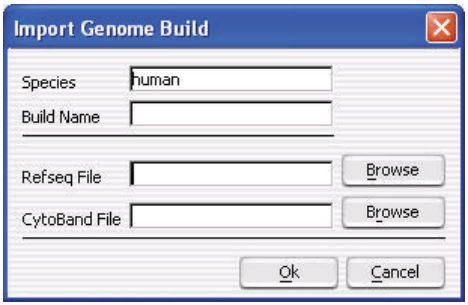

Figure 75 Import Genome Build dialog box

Purpose: To import a new set of genome build files into Agilent Genomic Workbench. See "To import a genome build" on page 46.

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 be imported.

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-provided genome build files.

Import Track

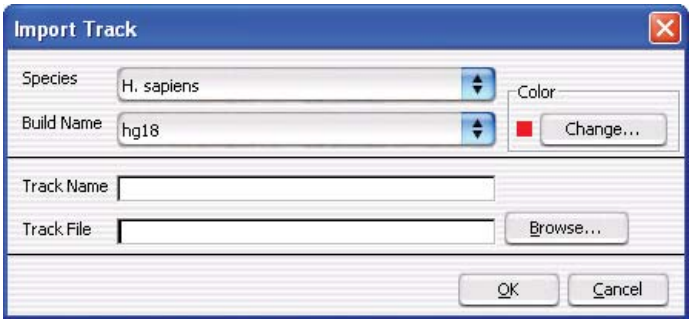

Figure 76 Import Track dialog box

Purpose: Lets you import a BED format track file. See "To import tracks" on page 49. Track information can appear in Gene view.

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 Shows the currently 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 that you want 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

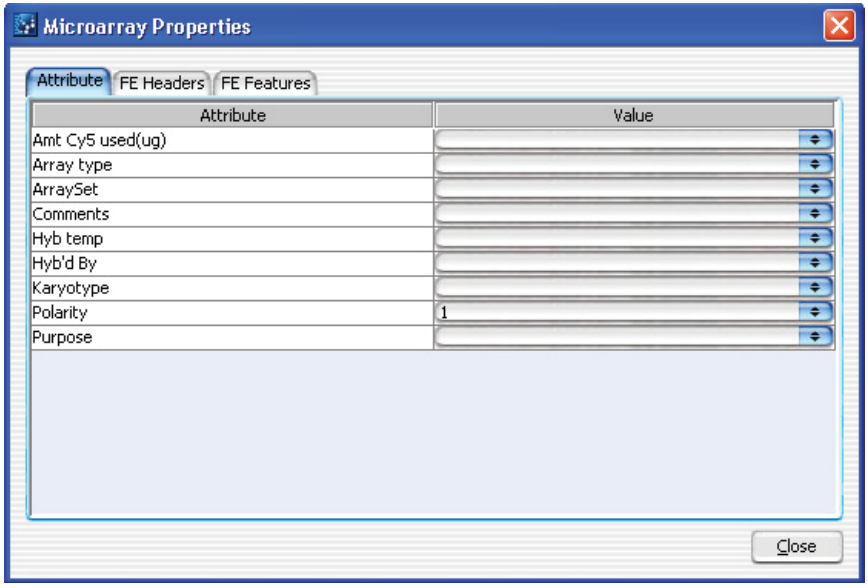

Figure 77 Microarray Properties dialog box with list of Attributes and their values

- **Attribute** — Displays the attributes associated with the array.
- **Value** – Indicates the value, if any, for each attribute.

Close Closes the dialog box.

FE Headers

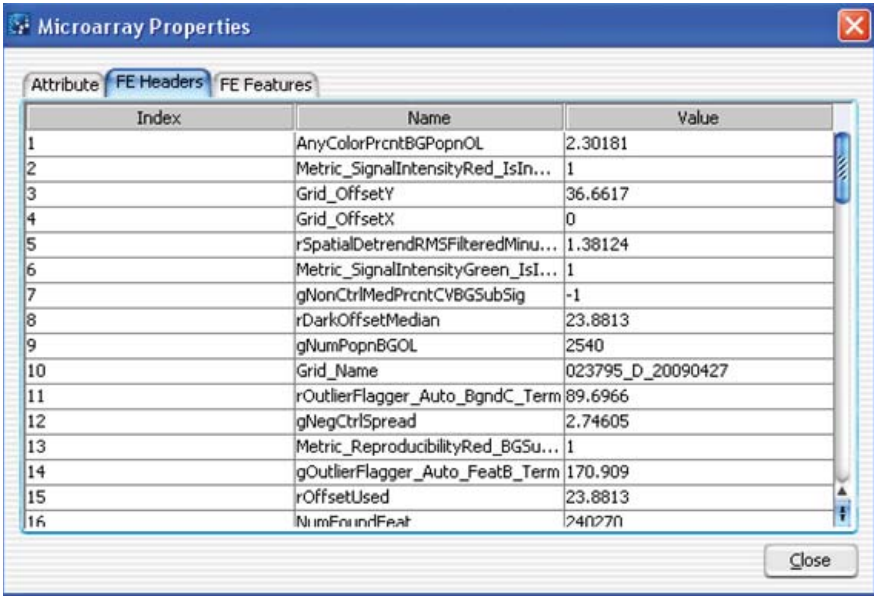

Figure 78 Microarray Properties dialog box listing FE Headers their values

Index Displays a sequential index to help identify FE properties.

Name Displays feature parameters, statistics and constants for the array.

Value Displays the value for each parameter, statistic and constant.

Close Closes the dialog box.

FE Features Tab

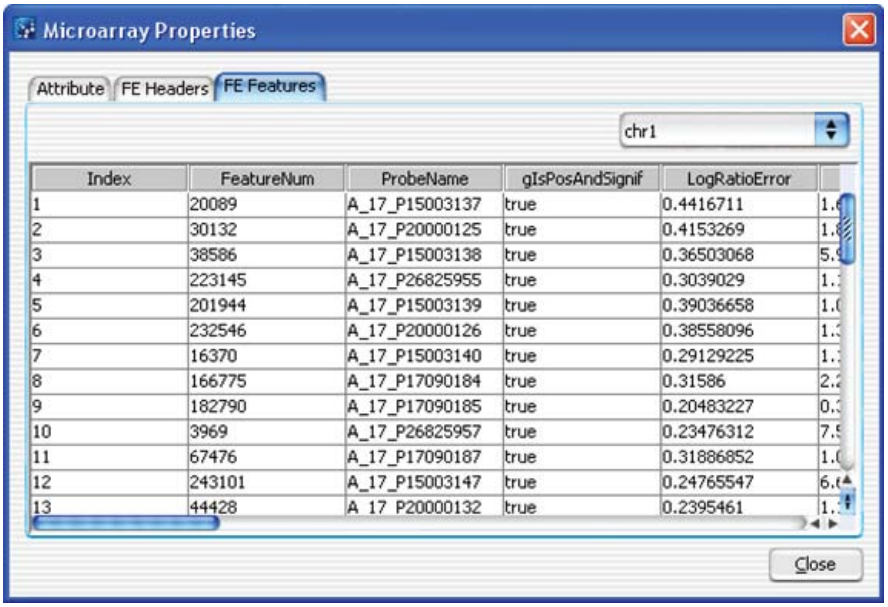

Figure 79 Microarray Properties dialog box listing FE Features and associated data

List of chromosomes Select the chromosome whose feature information you want to view.

Table of features Displays FE features and their associated data.

Pre-defined Peak Shape detection v2.1

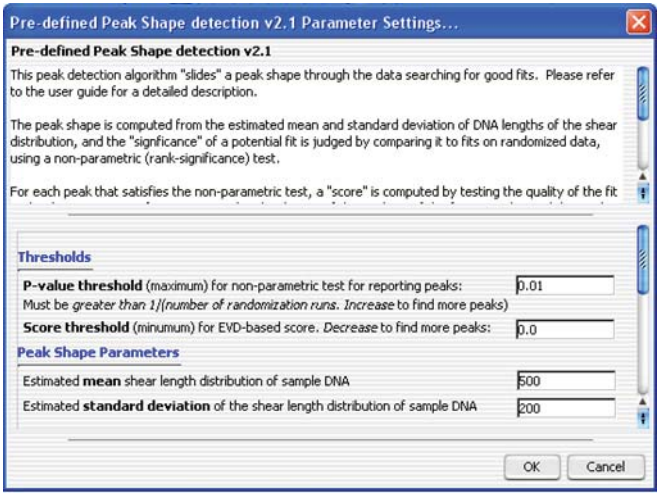

Figure 80 Pre-defined Peak Shape detection v2.1 dialog box

Purpose: This peak detection algorithm **"**slides**"** a peak shape through the data searching for good fits. The peak shape is computed from the estimated mean and standard deviation of DNA lengths of the shear distribution, and the "significance" of a potential fit is judged by comparing it to fits on randomized data, using a nonparametric (rank-significance) test.

For each peak that satisfies the non-parametric test, a "score" is computed by testing the quality of the fit under the assumption of an extreme value distribution of the qualities of the fits to randomized data. The significance derived from this test is converted to a score by computing -log10(significance) for the peak fit. For more information, see "Predefined peak-shape detection" on page 272.

To open: In the **Analysis** tab, under **Event Detection**, select **Pre-defined Peak Shape detection v2.1** and then click the edit icon .

Table 9 Pre-defined peak shape detection parameters v2.1

Parameter	Comments
Thresholds	
p-value threshold	Maximum threshold for the nonparametric test for reporting peaks. Value must be greater than 1/number of randomized runs. (Increase the value to find more peaks.)
Score threshold	Minimum threshold for extreme value distribution (EVD)-based score. (Decrease value to find more peaks.)
Peak Shape Parameters	
Estimated mean shear length distribution of sample DNA	Type a mean to be used in the gamma distribution calculation for approximation of the distribution of sheared DNA fragments.
Estimated standard deviation of the shear length distribution of sample DNA	Type a standard deviation to be used in the gamma distribution calculation of the distribution of sheared DNA fragments.
Other Algorithmic Parameters	
Precision of peak placement on the chromosome (in base pairs)	This is the window within which the algorithm searches for potential positions for the peak center. When you decrease this window, the time it takes for the algorithm to run increases.
Number of randomizations for determination of peak significance (via nonparametric test) and score	The program computes *p*-value and peak score through a number of random samplings. Increase the number of samples to increase the accuracy of the prediction; however this also increases the time to do the calculation.
Window size (in bp) for computing local baseline.	Use smaller number for smaller genomes.
Desired spacing of interpolated data points between probe.	Must be less than or equal to probe spacing on the array.

Table 9 Pre-defined peak shape detection parameters v2.1

Parameter	Comments
Automatically re-run calculation after learning peak shape?	Selection increases accuracy, but will double the runtime.
Use errors estimated by Error model?	Select to use the estimated error for each probe to weight its contribution to the peak fit measurement.

Probe Report Settings Parameter Settings

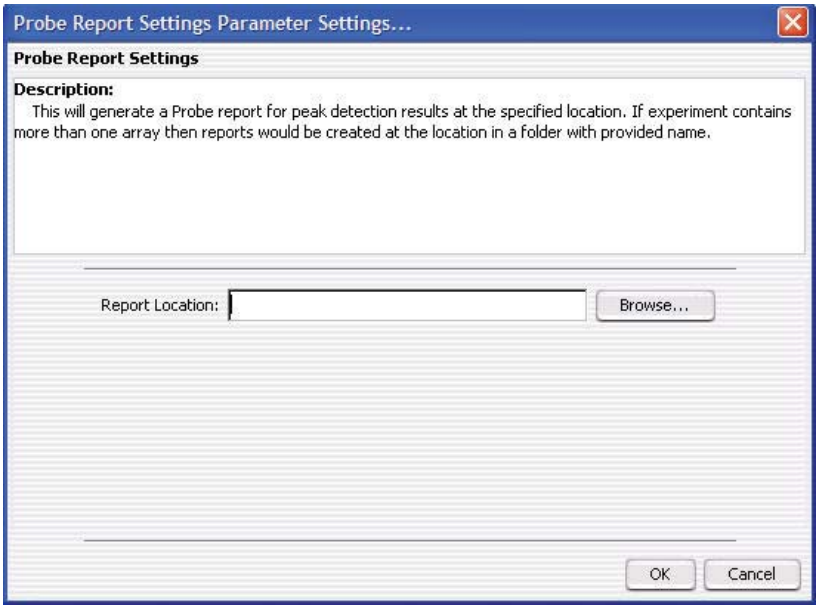

Figure 81 Probe Report Settings Parameter Settings dialog box

Purpose: The Probe Report Settings Parameter Settings dialog box lets you set the location where the program saves the probe report. A probe report contains information about the probes in the current experimental result in tab-separated value (*.tsv) format. It contains one row for each probe

in the array. See "Probe Report format" on page 260 for a description of the columns in the report. You can view probe reports and perform further analysis on them with a spreadsheet program.

To open: Click **Reports > Probe Report.**

Report Location Shows the location where the program saves the probe report. You can type a location in the text box or click **Browse.** See "Select report folder" on page 234.

OK Generates the probe report in the selected location.

Cancel Closes the dialog box without generating a report.

Properties (of an experiment result)

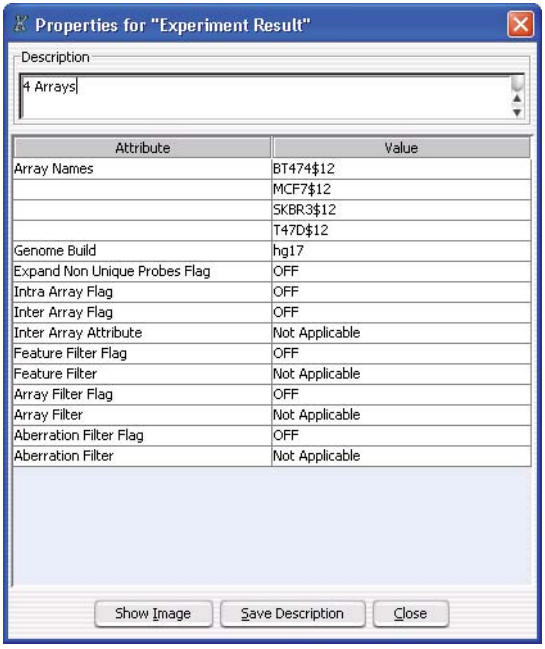

Figure 82 Properties dialog box (for experiment results)

Purpose: This dialog box lists the attributes of a saved experiment result.

To open: Right-click the name of the result in the Results folder of an experiment.

Description Shows the description associated with the result. You can add a description, or change an existing one. Click **Save Description** to save the description with the experiment result.

Table Shows the attributes of the experiment result and their corresponding values.

Save Description (Available after you add or change a description) Saves the current description with the experiment result.

Close Closes the dialog box.

QC Metrics Table

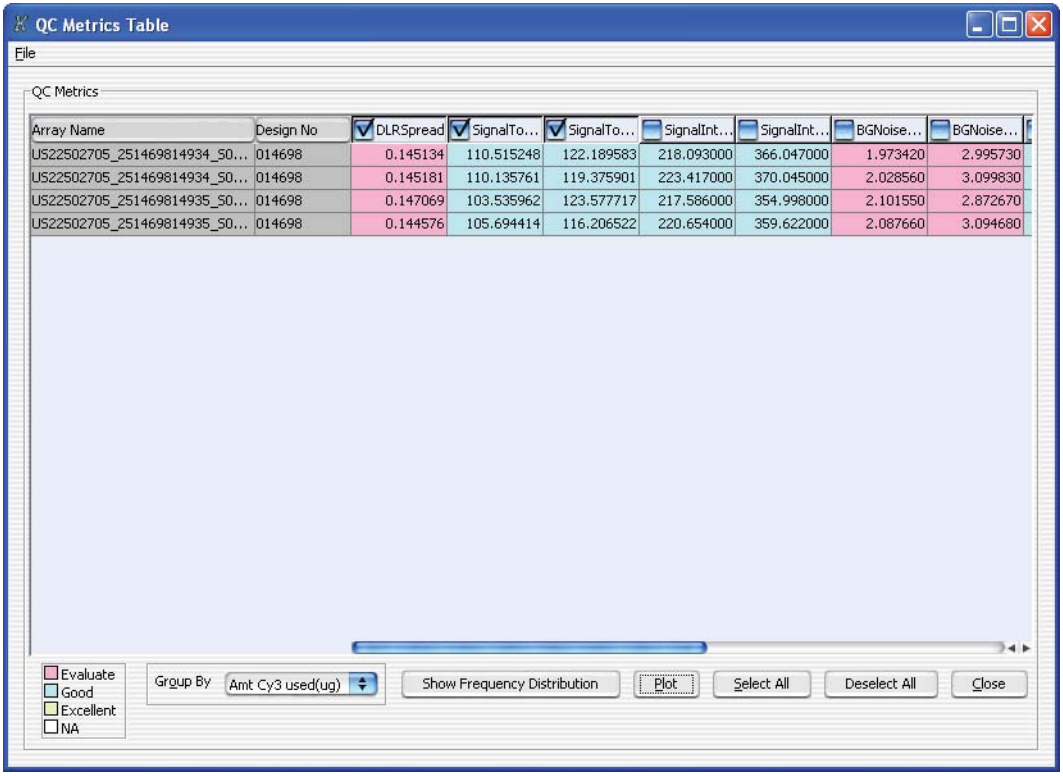

Figure 83 QC Metrics table

Purpose: The QC Metrics Table shows the available metrics for one or more arrays. With this table and the available plots, you can evaluate the quality of your microarray results, and assign a manual QC status to each microarray. Some metrics come from the Agilent Feature Extraction program, while others are calculated by the ChIP application itself. These metrics are available only for Agilent microarrays. See "To display QC metrics of arrays and set array QC status" on page 88.

To open: Click **Preprocessing > QC Metric.** Alternatively, in the **Experiment** pane of the Navigator, right-click the name of a genome build, experiment, or individual array, then click **QC Metrics.**

File Lets you save the QC Metrics Table as a Microsoft Excel (*.xls) format file. When you click **File,** a menu opens with an Export command. This command opens a Save dialog box, where you can select a location and type a name for the exported file.

Table The values of the QC metrics for arrays appear under QC Metrics, one array per row. The table has many columns:

- **Array Name** – Displays the names of microarrays. Because you can open the QC Metrics Table in several ways, the list can contain an individually-selected microarray, or those associated with an experiment or with a design.
- **Design No.** – Identifies the Agilent design ID for each microarray.
- **Metrics** – The program evaluates each metric, and assigns it a rating of Excellent (yellow), Good (turquoise), Evaluate (pink), or NA (white). The name of each metric appears as a column heading. Mark the check box next to the name of the metric to include it in the available plots. Drag the column heading of a metric horizontally to change its position in the table.
- **ManualQC Flag** – Lets you set the QCMetricStatus attribute of the array. Status can be Pass, Fail, Marginal, or NA. Later, you can filter arrays based on this attribute.

  The following metrics are included in the table.

Metric	Comments
SignalToNoiseGreen (and Red)	For each channel, this metric is the Signal Intensity divided by BGNoise. If this ratio is low, fail the array. A ratio over 100 indicates that the DNA quantity is sufficient and that no significant error was introduced during hybridization, washing, or scanning. To display the cut-offs the program uses to score this metric, open the current ChIP_QCMT_(date) default metric set in Feature Extraction.

Metric	Comments
SignalIntensityGreen (and Red)	For each channel, this metric is the median background-subtracted signal after rejecting nonuniform outliers and saturated features. If the signals are too low, fail the array. If the signals are marginal, expect noisy results. Low signals can result from poor quality input DNA or from losses during labeling and clean-up. To display the cut-offs the program uses to score this metric, open the current ChIP_QCMT_(date) default metric set in Feature Extraction.
BGNoiseGreen (and Red)	For each channel, this metric is the standard deviation of negative control probes after rejecting feature nonuniform outliers, saturated features, and feature population outliers. If the noise is high, examine the array image for visible non-uniformities. High background noise is often introduced during slide handling or from contaminated buffers. To display the cut-offs the program uses to score this metric, open the current ChIP_QCMT_(date) default metric set in Feature Extraction.
ReproducibilityGreen (and Red)	For each channel, this metric is the Median %CV (coefficient of variation) of background-subtracted signal for replicate non-control probes after outlier rejection. The calculation excludes any probe whose average signal is below the additive noise of that channel (i.e. Average (BGSubSignal) * Multiplicative error < Additive error / Dye Norm Factor). After rejecting non-uniform outliers and saturated features, the program requires at least three probes to calculate the CV for that sequence. It then calculates the median of the CVs of the remaining sequences. If the number of sequences that pass the filter is less than 10, the program does not calculate this metric. High scores on this metric may signal catastrophic failures (for example, that the slide leaked or fell out of the rotisserie). Large bubbles cause moderate values on this metric, but do not compromise the results significantly. To display the cut-offs the program uses to score this metric, open the current ChIP_QCMT_(date) default metric set in Feature Extraction.

Group By The program displays the arrays in the table, and in the QC Metrics Graph, grouped by the array attribute you select here.

Show Frequency Distribution Opens the Frequency Distribution dialog box. This dialog box contains line plots of the distribution of each selected metric over the all of the arrays in the QC Metrics Table. See "QC Metrics – Frequency Distribution" on page 223.

Plot Opens the QC Metrics Graph dialog box. This dialog box contains plots of each selected metric for each array. See "QC Metrics Graph" on page 224.

Select All Marks the check boxes of all metrics.

Deselect All Clears the check boxes of all metrics.

Close Closes the QC Metrics Table.

QC Metrics – Frequency Distribution

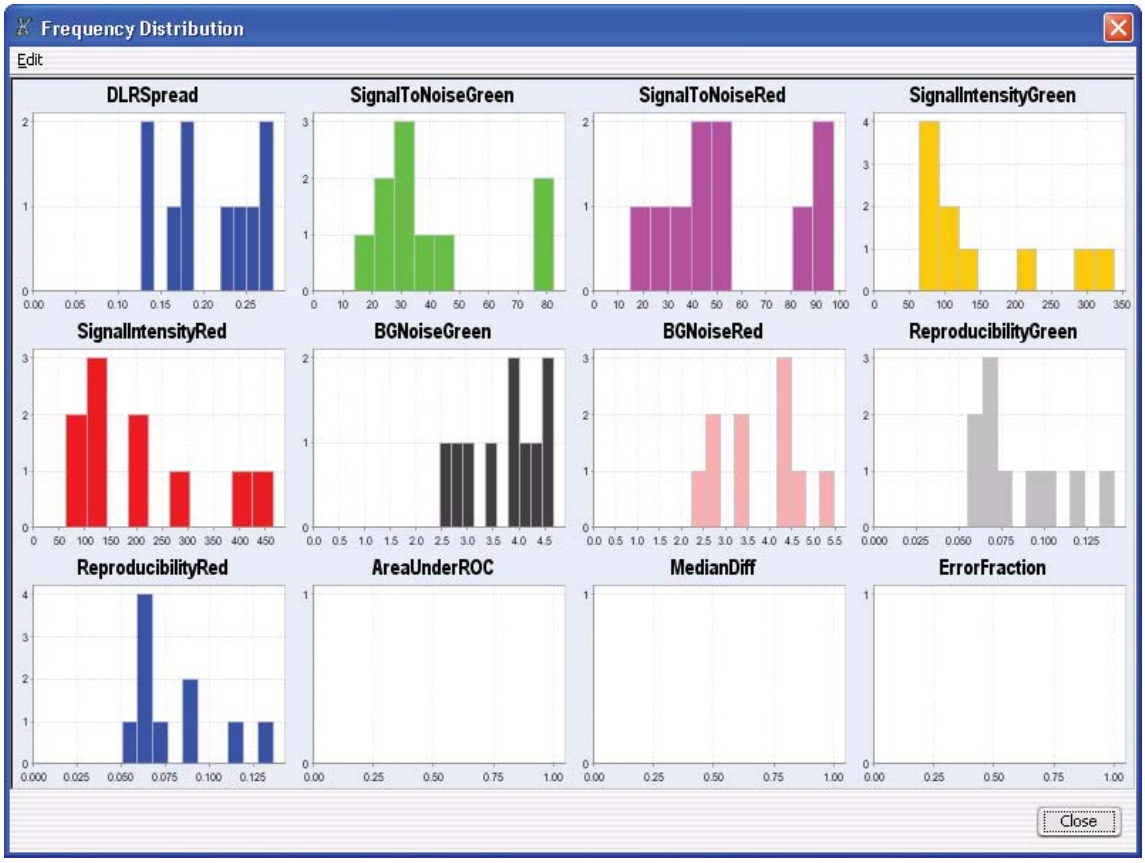

Figure 84 QC Metrics Frequency Distribution Plot

Purpose: The plots in this dialog box represent the selected columns in the QC Metrics Table. Each plot shows the number of arrays within each value range for a metric. See "QC Metrics Table" **on page 219**.

To open: Click **Show Frequency Distribution** in the QC Metrics Table.

Edit Opens a menu with a Copy command that copies the plots in the dialog box to the Clipboard. You can then paste the image into a document.

Close Closes the dialog box.

QC Metrics Graph

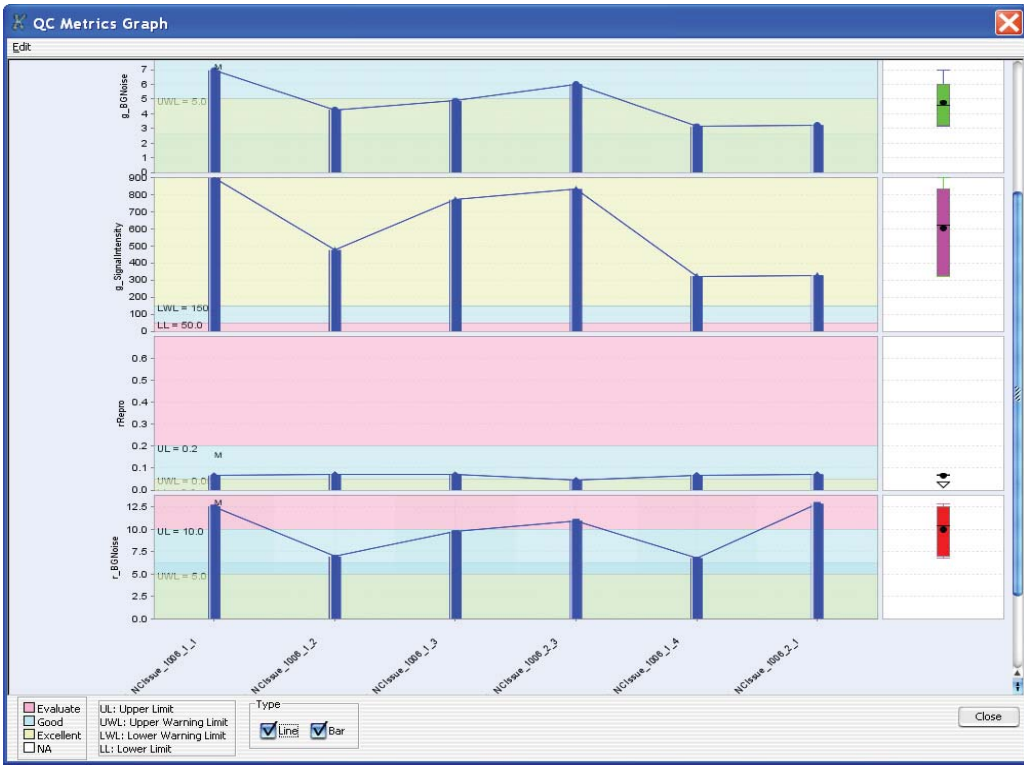

Figure 85 QC Metrics Graph

Purpose: The plots in this dialog box represent the selected columns in the QC Metrics Table. Each plot shows the value of a given metric for all arrays.

To open: Click **Plot** in the QC Metrics Table (see "QC Metrics Table" on page 219).

Main Plots These plots have several features:

- The background colors in each plot correspond to quality guidelines developed by Agilent, based on normal ranges observed for analyses of well-established cell lines using standard Agilent protocols. See the descriptions of each of these metrics in "QC Metrics Table" on page 219.
- A "Box & Whisker" plot appears to the right of the main plot for each metric. See "'Box & Whisker' Plot," below.
- The program can plot the data as a line graph, a bar graph, or both. See "Line" and "Bar," below.
- If you set your user preferences to show ToolTips (see "View Preferences" on page 252), a ToolTip appears when you place the pointer over any bar. The ToolTip lists the value of each bar and the name of the corresponding array. See Figure 86.
- You can right-click any bar to open a shortcut menu for the corresponding array. The options in the shortcut menu allow you to set the QCMetricStatus attribute for the array. See Figure 86.
- The QCMetricStatus attribute for each array appears over all of the corresponding bars of the main plot. The four possible values for QCMetricStatus are: (P)ass, (F)ail, (M)arginal, and (N)A.

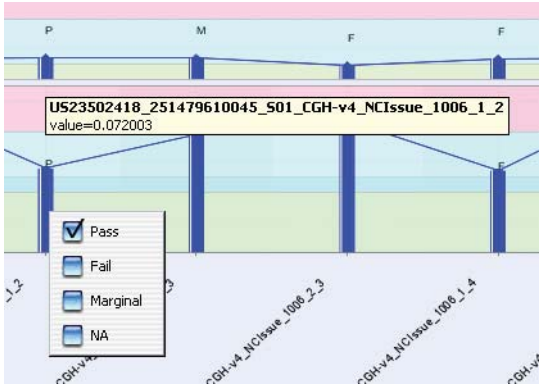

Figure 86 Portion of the QC Metrics Graph, showing a ToolTip (values in white box), and a shortcut menu. You use the shortcut menu to set the QCMetricStatus attribute for the array. Right-click any bar to open the shortcut menu for the corresponding array.

"Box & Whisker" Plots

A small plot appears to the right of each of the main plots. It represents the overall distribution of values for the metric. Two examples appear in Figure 87. The symbols have the following meanings:

- The lower and upper edges of the box represent the 25th and 75th percentiles, respectively.
- The black horizontal line in the box is the median.
- The black dot is the mean.
- The "whiskers" represent the range of values that are not outliers. An outlier is a point that is out of the 25th to 75th percentile range by more than 150%.
- Open circles represent outliers, and an open triangle represents outliers that plot beyond the available space on the graph.

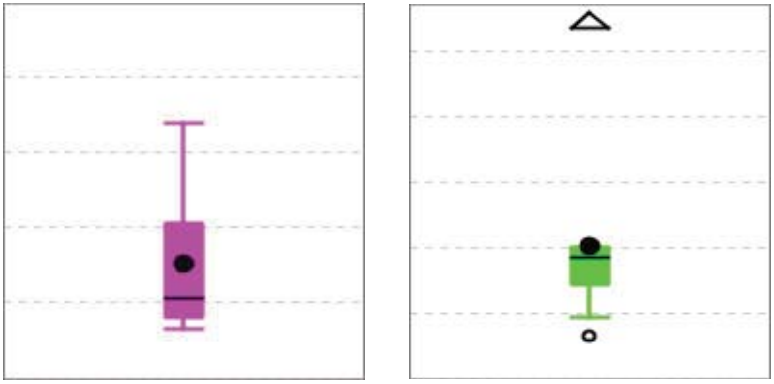

Figure 87 Two "Box & Whisker" plots

Line Mark this option to display each metric as a line graph.

Bar Mark this option to display each metric as a bar graph.

Edit Opens a menu with a Copy command. This command copies the plots in the dialog box to the Clipboard as an image. You can then paste the image into a document in another program.

Close Closes the dialog box.

QC Report Settings Parameter Settings

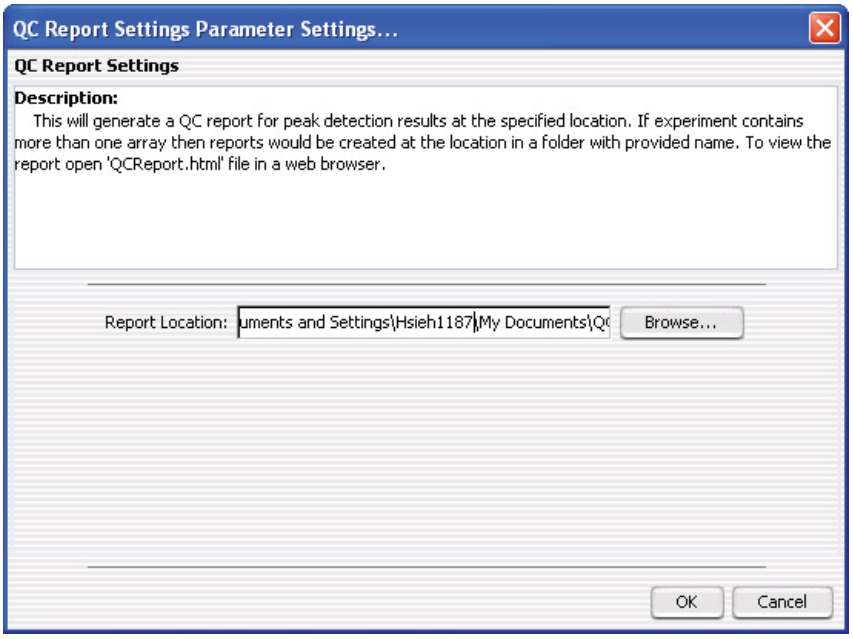

Figure 88 QC Report Settings Parameter Settings dialog box

Purpose: This dialog box lets you set the location where the program saves the QC report. The QC report summarizes the settings of the current analysis, and the overall statistics of each array. In addition to summary tables, it includes four plots. The program creates the QC report in HTML format for viewing in your Internet browser. For more information about the contents of this report, see "QC report format" on page 263.

To open: Click **Reports > QC Report.**

Report Location Shows the location where the program saves the QC report. You can type a location in the box or click **Browse.** See "Select report folder" on page 234.

OK Generates the QC report in the selected location.

Cancel Closes the dialog box without generating a report.

Sample Attributes

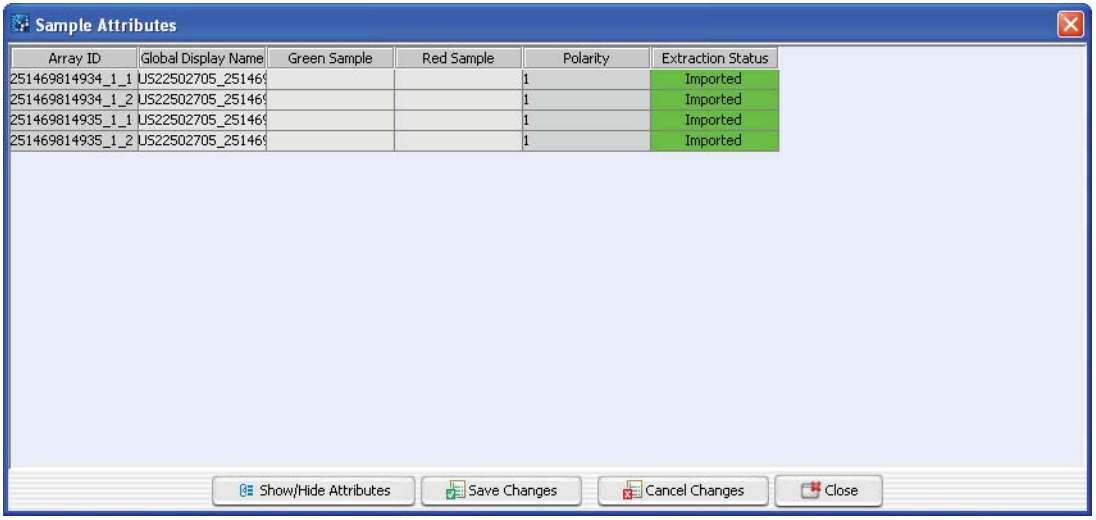

Figure 89 Sample Attributes dialog box

Purpose: To show, hide, or edit the attributes of experiment arrays

To open: In the Experiment Pane of the Navigator, right-click an experiment, then click **Sample Attributes.**

Table of Attributes Lets you select or change the values for the attributes for the arrays in the experiment. The columns that appear initially are the default columns (Array ID, Green Sample, Red Sample, Polarity and Extraction Status) plus any that have been made visible in the Sample Manager table. See the *Sample Manager User Guide*. Changes you make are applied globally.

Show/Hide Attributes Opens the Show/Hide Columns dialog box, where you can add or remove attribute columns from the table of attributes. See "Show/Hide Columns" on page 242.

Save Changes Saves any attribute changes you have made.

Cancel Changes Restores attributes to their prior values, and leaves the dialog box open.

Close Closes the dialog box.

Search Probes in eArray

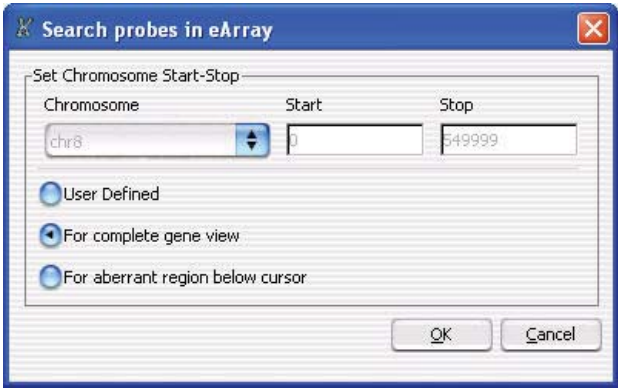

Figure 90 Search probes in eArray dialog box

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 from which the probes to be searched in eArray will be selected. 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 will be searched.

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, you can type in Start and Stop positions for defining the region contained the genes to be in the list.

Scroll to Column

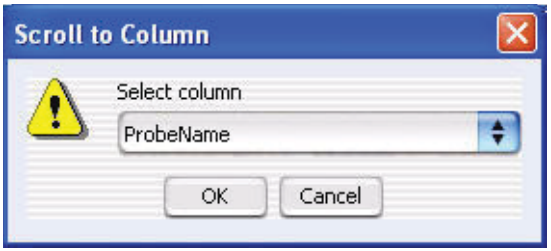

Figure 91 Scroll to Column dialog box

Purpose: Lets you select a "Scroll to" column. The program then scrolls the tab so that you can see the selected column.

To open: Right-click a column heading in Tab View, then click **Scroll To Column** in the shortcut menu.

Select column Displays the columns available in the currently selected tab. Select the one you want to display.

OK Scrolls the current tab so that you can see the selected column.

Cancel Closes the dialog box.

Select Color

Purpose: To select a color. Three tabs are available for selecting colors:

- Swatches tab - select colors based on samples (swatches)
- HSB tab - select colors based on an HSB schema (Hue, Saturation, and Brightness)
- RGB tab - select colors based on an RGB schema (Red-Green-Blue)

To open: This dialog box opens when a function allows you to change a color. For example, right-click on an array in an experiment, click **Edit Array Color** and click the **Swatches, HSB,** or **RGB** tab.

Swatches tab

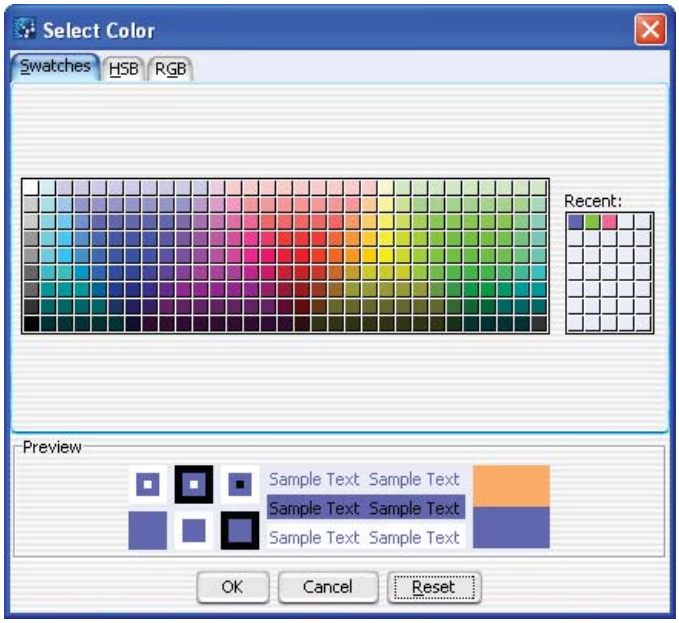

Figure 92 Select Color - Swatches tab

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

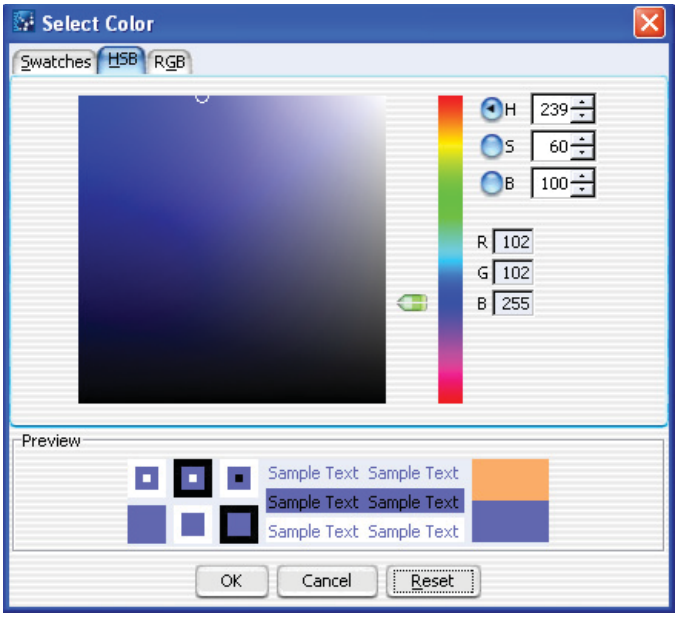

Figure 93 Select Color - HSB tab

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

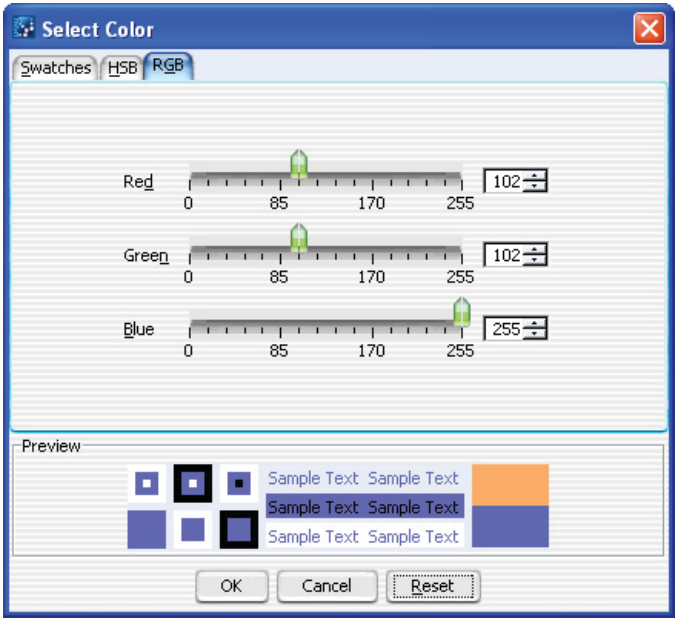

Figure 94 Select Color - RGB tab

This tab is used to select a color based on an RGB schema.

Red Move the slider to change the amount of red in the color. Or, click the up or down arrow to select a number.

Green Move the slider to change the amount of green in the color. Or, click the up or down arrow to select a number.

Blue Move the slider to change the amount of blue in the color. Or, click the up or down arrow to select a number.

Preview The Preview area shows how the selected color appears. When you change the color, the original color appears at the top of the color box on the right.

OK Click to select the color and close the dialog box.

Cancel Click to close the dialog box without changing the color.

Reset Click to return the swatches, HSB, and RGB colors back to default values.

Select report folder

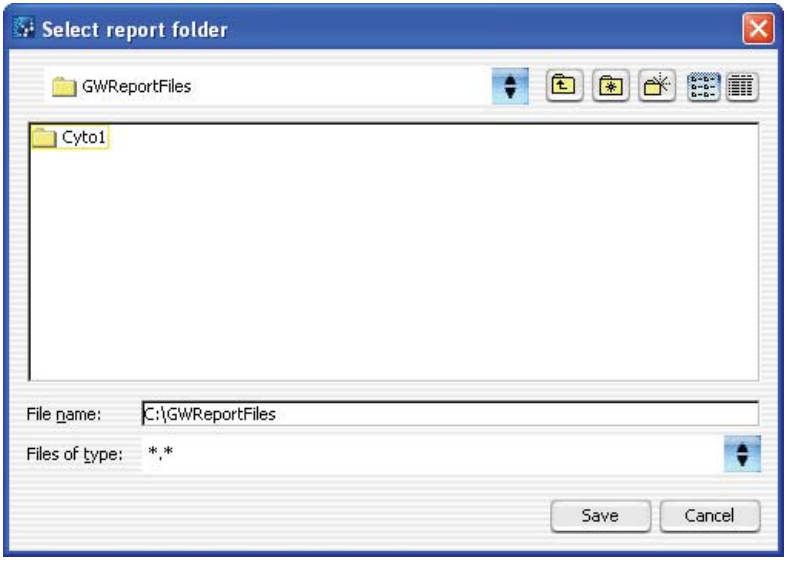

Figure 95 Select report folder dialog box

Purpose: This dialog box lets you select the location for a gene, probe, or QC report.

To open: Select a report from the Reports menu, then click **Browse** in the dialog box that appears. See "Reports command ribbon" on page 131.

Use the standard Windows Explorer commands available in the Select report folder dialog box to select a location for a report. Only the files that have the same file type as the selected report appear in the list of flies.

File name Shows the file name of the report to be created. Type a name in the text box.

Files of type Shows the file type of the report to be created. The program creates gene and probe reports in tab-separated value (*.tsv) format, and creates QC reports in HTML (*.htm) format.

Open Opens the selected location.

Cancel Closes the dialog box without opening a location.

Set genome build and species for Axon design files

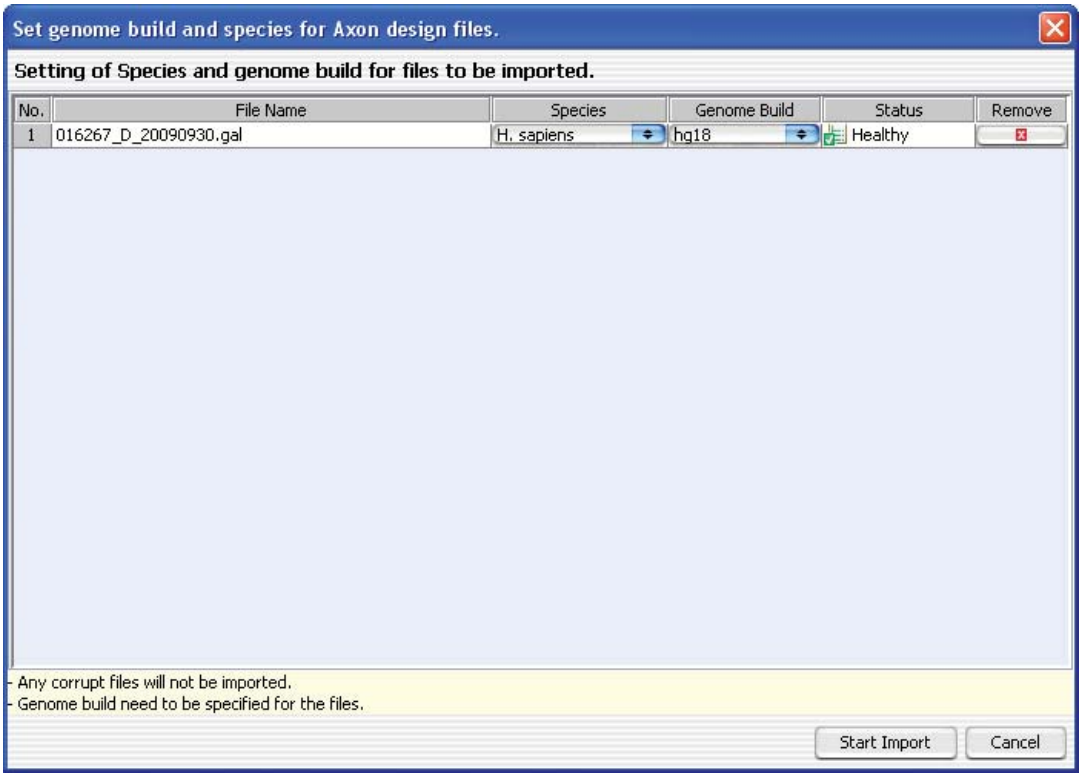

Figure 96 Set genome build and species for Axon design files dialog box

Purpose: Lets you set the species and genome builds associated with imported Axon design file(s), and to remove specific designs files from the import, if necessary. See "To import Axon design files" on page 45.

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 associated with each design file. If a species is incorrect, select the correct one from the appropriate list.

Genome Build The genome build associated with 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 has passed 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.
- **Overwrite** – The file is a valid design file, but when you import it, it replaces 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.

Setting Normalization Parameters

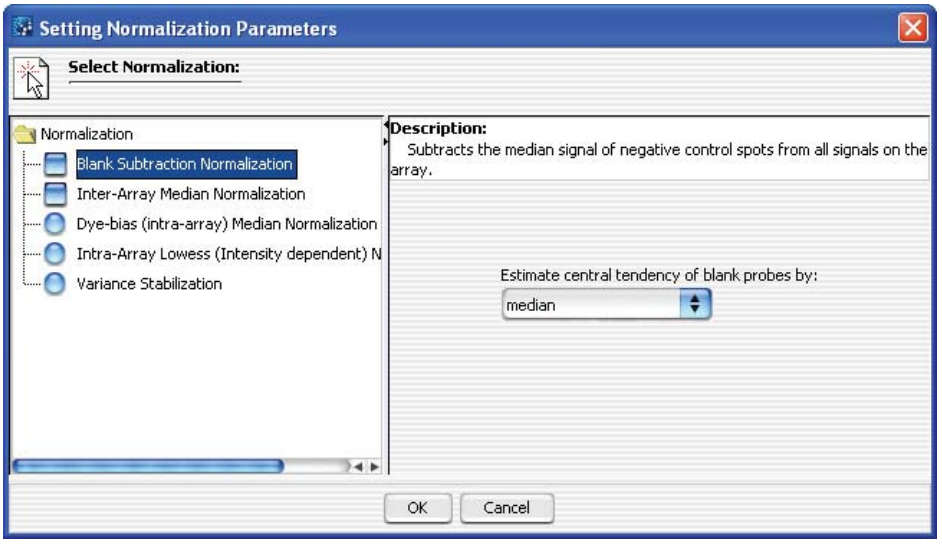

Figure 97 Setting Normalization Parameters dialog box

Purpose: The Setting Normalization Parameters dialog box lets you include and configure the normalization methods that the program applies during analysis. See "To apply normalization" on page 103 and "Normalization Algorithms" on page 274.

To open: This dialog box opens when you select a Normalization method in the Preprocessing ribbon, and click .

Blank Subtraction Normalization Mark this check box to include blank subtraction normalization in your analysis. This kind of normalization corrects for non-specific binding. It first calculates the central tendency of the negative controls on the array for both the immunoprecipitated (IP) and whole cell extract (WCE) channels. It then subtracts these central tendencies from the raw signal intensities of each feature on the array. As with all of the normalization methods, if the method causes a probe to have a negative value for intensity, it flags the probe as *excluded*. For information on this algorithm, see "Blank subtraction normalization" on page 274.

Click the words **Blank Subtraction Normalization** to display the settings for this kind of normalization in the right pane of the dialog box. You can configure the following setting:

Setting	Description
Estimate central tendency of blank probes by	Selects how the program calculates the central tendency that it subtracts from the raw signal intensity values. There is currently only one selection: • median

Dye-bias (intra-array) Median Normalization

Mark this check box to include intra-array median normalization in your analysis. This kind of normalization corrects for dye bias within each array in an experiment, and it normalizes the intensities of the IP channel only. For information on this algorithm, see "Intra-array (dye-bias) median normalization" on page 277.

Click the words **Dye-bias (intra-array) Median Normalization** to display the settings for this kind of normalization in the right pane of the dialog box. You can configure the following settings:

Setting	Description
Normalize by	Selects how the program computes the dye bias when it applies this kind of normalization. • **By equalizing central tendencies of IP and WCE channels** – This method first calculates the ratio of the median IP signal intensity to the median WCE signal intensity. Then, it multiplies the signal intensities of the data probes by this ratio. • **By normalizing central tendency of log ratios to 1** – This method multiplies the signal intensities of all data probes on the array by a correction factor. This correction factor adjusts the central tendency of log ratios of the data probes on the array to 1.

Inter-Array Median Normalization

Mark this check box to include interarray median normalization in your analysis. This kind of normalization corrects for variations from one replicate array to another. The program calculates and applies the normalization separately for each channel. It first calculates the median

signal intensity over the common probes in each replicate array. It then finds the average of these median intensities over all replicates of all arrays. For each array, it computes the ratio of its median signal intensity to the average of the median signal intensities of all arrays. Finally, it normalizes data by multiplying each signal intensity by the applicable ratio. For information on this algorithm, see "Interarray median normalization" on page 276.

Click the words **Inter-Array Median Normalization** to display a description of this kind of normalization. No settings are required.

Intra-array Lowess (Intensity dependent) Normalization

Mark this check box to include intra-array Lowess (intensity dependent) normalization in your analysis. This kind of normalization attempts to correct for artifacts caused by nonlinear rates of dye incorporation, as well as inconsistencies in the relative fluorescence intensity between some red and green dyes. The Lowess normalization algorithm normalizes the channels within each array using a nonlinear polynomial fit to the data, and effectively normalizes by probes and by arrays. For information on this algorithm, see "Intra-array (intensity-dependent) Lowess normalization" on page 280.

NOTE If you are using feature extraction data that have been normalized by the Lowess approach, you do not need to apply the intra-array Lowess normalization here.

Click the words **Intra-array Lowess (Intensity dependent) Normalization** to display the settings for this kind of normalization. You can configure the following settings:

Setting	Description
Regression curve is fitted to	Select the data for the regression curve. • **All data probes** - Includes all of the data probes in the regression curve. • **All common probes** - Includes probes whose names start with "LACC". • **Gene desert probes** - Includes data for probes whose names start with "LACC:GD".

Variance Stabilization

Mark this check box to include intra-array (intensity dependent) variance stabilization normalization in your analysis. This normalization is useful for data that is either "blank-subtracted" or "spatially detrended" but it may have utility for data processed by other means as well. Variance stabilization is an alternative to Lowess normalization that fits a regression curve to signal intensities after applying an "asinh(x)" transform to each channel. This approach uses a two-parameter error model to compress the reported ratios of probes with weak signals after blank-subtraction. After the transform is applied, the variance of the reported log ratios should be independent of the signal strength. For information on this algorithm, see "Intra-array (intensity-dependent) variance stabilization" on page 282.

Click the words **Variance Stabilization** to display a description of this normalization and the settings. The following settings are available:

Setting	Description
Regression curve is fitted to	Select the data for the regression curve. • **All data probes** - Includes all of the data probes in the regression curve. • **All common probes** - Includes probes whose names start with "LACC". • **Gene desert probes** - Includes data for probes whose names start with "LACC:GD".

OK Accepts the selections and closes the dialog box.

Cancel Closes the dialog box without making any changes to the settings.

Show/Hide Columns

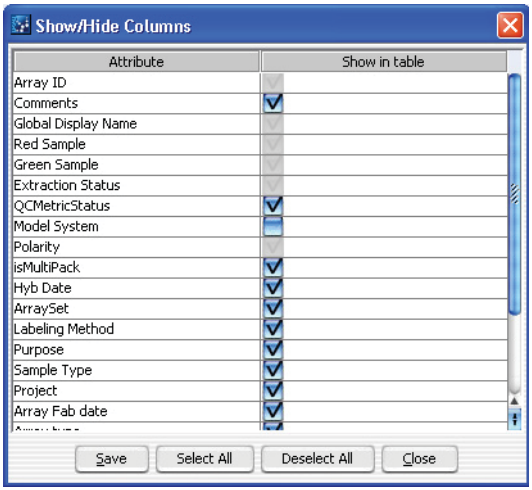

Figure 98 Show/Hide Columns dialog box

Purpose: Used to select the attributes to be displayed 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 will be displayed in the Sample Attributes and Sample Utilities. 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 again.

Save Saves the current list of selected attributes and updates the Sample Utility table based on the selections.

Select All Selects all the attributes in the list.

Deselect All Removes 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

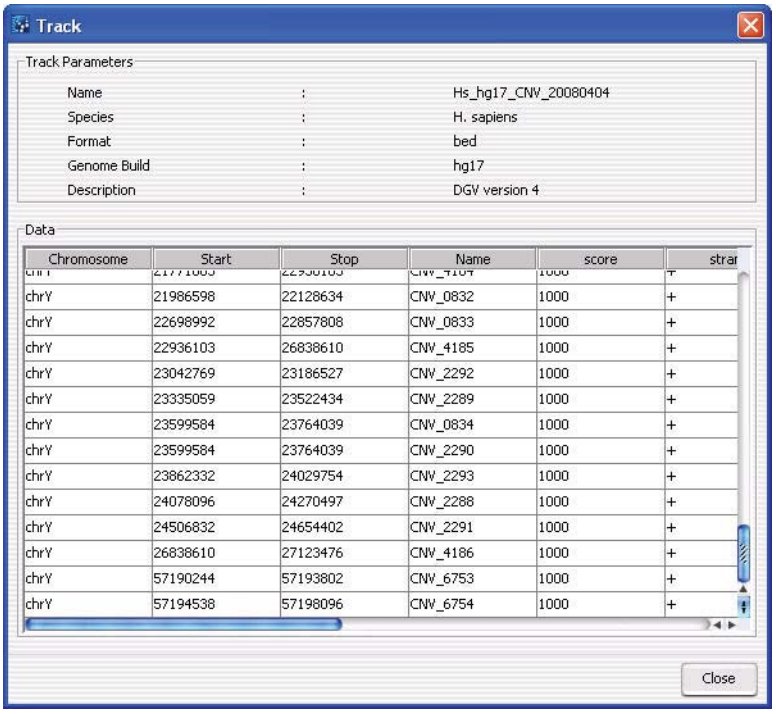

Figure 99 Track details

Purpose: This dialog box lets you view the parameters and data of the track.

To open: Click the **Details** link next to the desired track in the **Tracks** tab of the User Preferences dialog box. See "User Preferences" on page 245.

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 alongside the selected region for the feature.

The other columns are additional BED track file columns that can appear for some tracks. Agilent Genomic Workbench does not support these.

Close Closes the Track dialog box.

User Preferences

Purpose: This dialog box is used to set up preferences for display of tracks, eArray information and 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

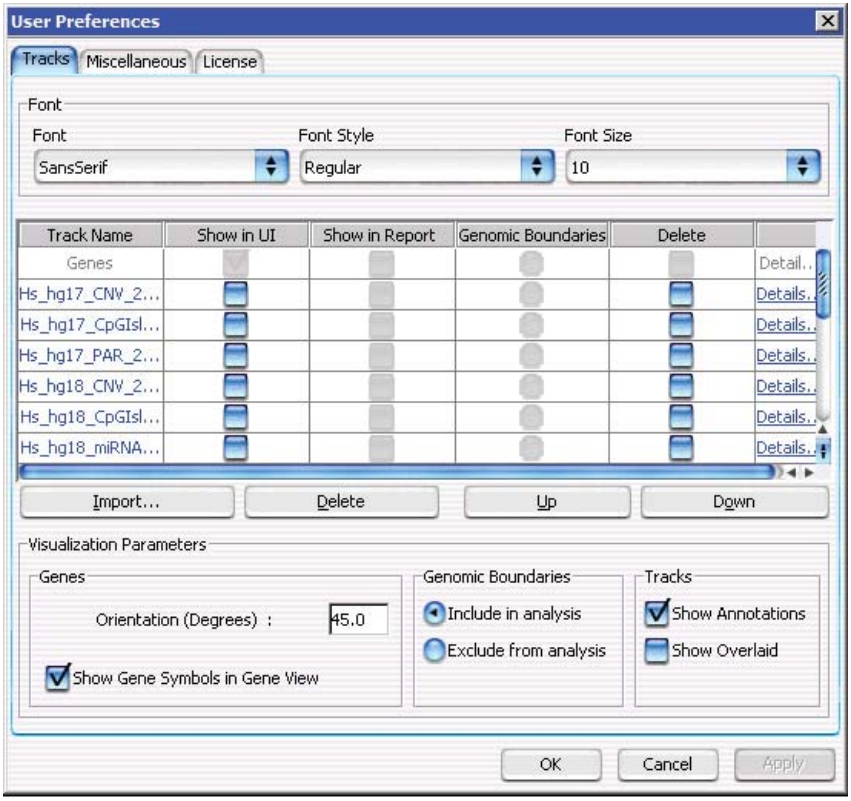

Figure 100 User Preferences dialog box - Tracks tab

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 (Not available in ChIP) Mark the check box to display the track information in all the reports.

Genomic Boundaries (Not available in ChIP) Click to use the track to define only the regions that event detection algorithms 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 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 (Does not apply to ChIP) 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

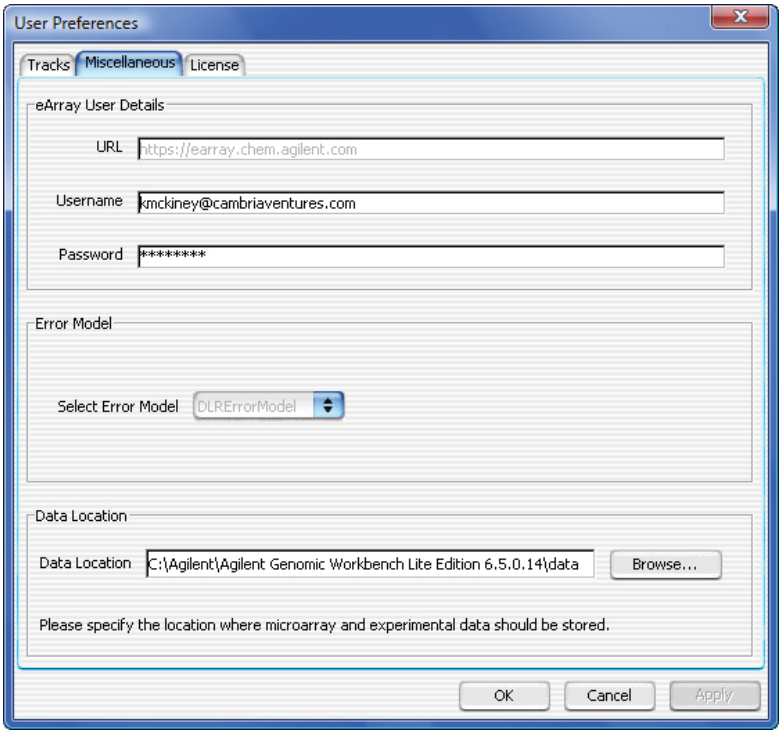

Figure 101 User Preferences dialog box - Miscellaneous tab

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 ChIP) The DLR Error Model (Derivative Log Ratio). This 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

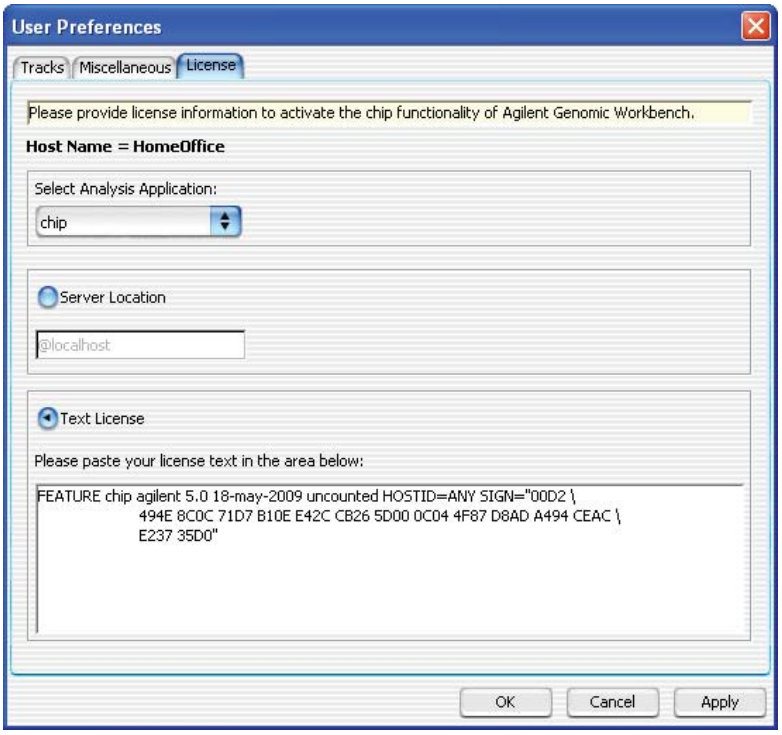

Figure 102 User Preferences dialog box - License tab

Purpose: The License tab allows you to display and update your ChIP application license. This license enables the ChIP application, and allows you to use it to analyze array data.

To open: 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, or want to use a server location for your user license(s). To edit this name, select **Server Location,** then type or copy the name of the computer used as your

license server, or the fully qualified path name for the folder where your license text files 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 coordinates in UCSC browser

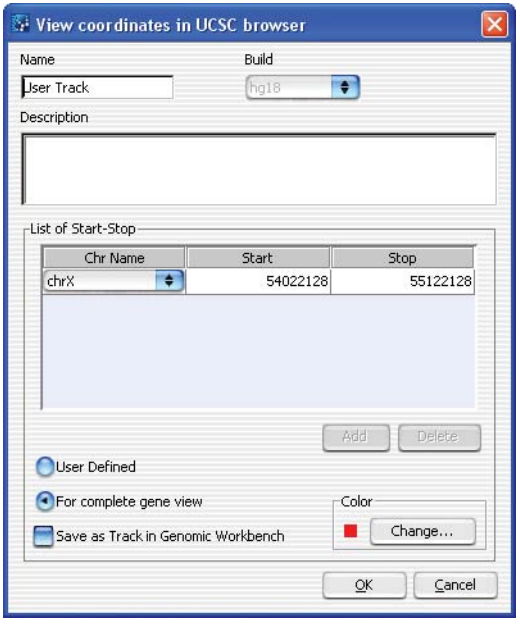

Figure 103 View coordinates in UCSC browser

Purpose: Defines a track to upload to the UCSC Web site so that you can see the information in the UCSC Genome Browser.

To open: Right-click in the Gene View, and select **Show in UCSC**.

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

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

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

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

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

Save as Track in Genomic Workbench Mark the check box to save this track in the Tracks folder in the My Entity List pane of the Navigator.

Change Click to open the Choose Track Color dialog box to select the color to use for display of the track in the Tracks folder. See "Select Color" on page 230.

OK Creates the track and opens the UCSC Web site, where you can display the track and associated information. For information on using the UCSC Web site, see the help and information provided there.

Cancel Closes the dialog box without creating a track.

View Preferences

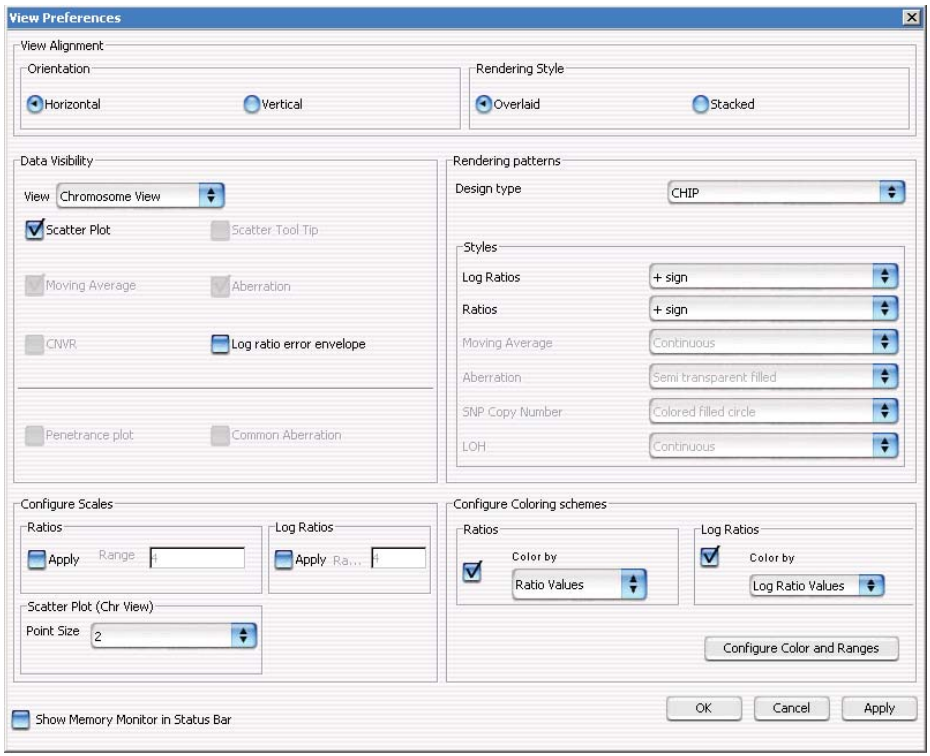

Figure 104 View Preferences dialog box for ChIP

Purpose: This dialog box lets you configure how data and results appear in Genome, Chromosome, and Gene Views.

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

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

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

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

In **View,** select the view you want to configure. To set availability of display items for all views, select **All views.** Some display items are only available for certain views and applications. When you select a display item, it enables the item for display; for some items, you must also 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.
Log ratio error envelope	The log ratio error envelope is a visual representation of the log ratio error calculated by Feature Extraction.

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 array design type for which you want to define rendering patterns.

- **Styles** – Select the display style for each of the elements in the following table. (Elements that appear grayed-out and do not apply to ChIP are not described here.)

Display element	Details
Log Ratios	Select the symbol used for log ratio data points in the scatter plots in Chromosome and Gene Views.
Ratios	Select the symbol used for ratios or probe score values data points in the scatter plots in Chromosome and Gene Views.

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

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

Configure Scales For Log Ratios or Signal Intensities 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.

To do this	Follow these steps
Show or hide all ratio data points	• To show the data points – Mark the **Ratios** check box and select **Ratio Values** from the list. • To hide the ratio data points – Clear the **Ratios** check box.
Show or hide all log ratio data points	• To show the data points – Mark the **Log Ratios** check box and select **Log RatioValues** from the list. • To hide the log ratio data points – Clear the **Log Ratios** check box.

To do this	Follow these steps
Show or hide all probe score data points	• To show the data points – Mark the **Log Ratios** check box and select **Probe Score Values** from the list. • To hide the probe score data points – Clear the **Log Ratios** check box.
Configure colors and ranges for the Scatter Plots	• Click **Configure Color and Ranges**. In the Set Colors and Ranges dialog box, type the ranges and colors you want to use for the scatter plots. See "Configure Coloring Ranges and Shades" on page 175 for more information.

Show Memory Monitor in Status Bar Displays a memory usage monitor in the eighth cell of the status bar. For information about the Status Bar, see "Status Bar" on page 168.

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.

Whitehead Error Model Parameter Settings

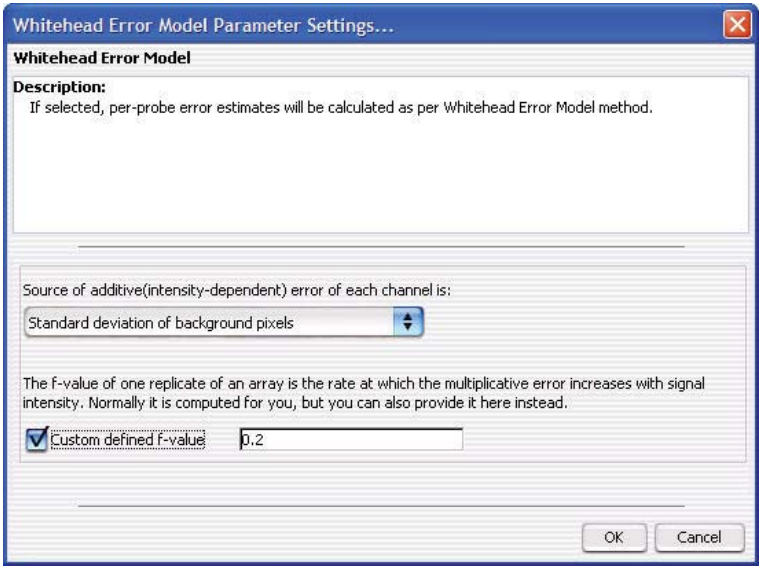

Figure 105 Whitehead Error Model Parameter Settings dialog box

Purpose: This dialog box lets you customize the advanced parameters of the Whitehead error model. Set these parameters to optimize the statistical calculations of the error model using training data specific to your particular assay.

To open: Select **Whitehead Error Model** under Error Model in the Preprocessing command ribbon, then click  . See "To configure the error model" on page 106.

Source of additive (intensity-dependent) error in each channel is:

The options for this parameter affect the additive (intensity-dependent) component of the estimate of the error in IP – WCE. Select one of these sources:

- Standard deviation of background pixels
- Additive error as computed by Agilent Feature Extractor
- Observed spread of negative controls

Custom defined f-value Mark this check box to type a custom f-value. In the box to the right, type an f-value. The f-value of one replicate of an array is the rate at which the multiplicative error increases with signal intensity. By default, the ChIP application calculates f-values automatically. For information on the statistical algorithm, see "Whitehead error model" on page 284.

OK Configures the Whitehead Error Model with the settings in the dialog box.

Cancel Closes the dialog box without changing any settings.

Whitehead Per-Array Neighbourhood Model Parameter Settings

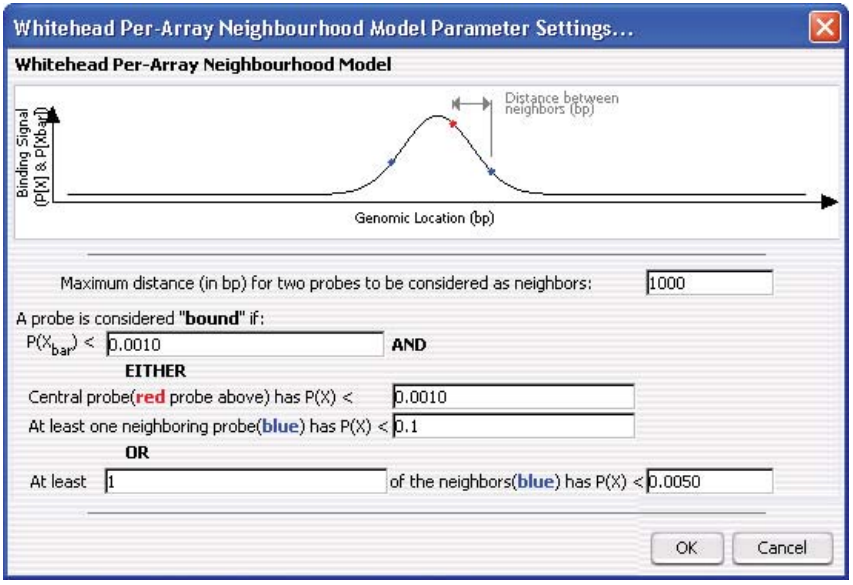

Figure 106 Whitehead Per-Array Neighbourhood Model Parameter Settings dialog box

Purpose: This dialog box lets you customize the parameters of the Whitehead per-array neighbourhood model. The ChIP application uses this model to make binding calls based on the *p*-values of each probe and its neighbors.

To open: Select **Whitehead Per-Array Neighbourhood Model** under Event Detection in the Analysis command ribbon, then click  . See "To apply event detection" on page 111.

Graph Illustrates the probes considered in the model. Two neighbor probes (blue) flank a central one (red).

Parameters The program considers the probe "bound" if the average *p*-values for all three probes is less than a set cut-off value, and if either of the following is true:

- The *p*-values for the central probe and at least one of its neighbors are less than set cut-off values.
- The *p*-value of one (or optionally, another number) of the neighbors of the central probe is less than a set cut-off value.

For a detailed description of the statistical algorithms, see "Whitehead per-array neighbourhood models" on page 271.

You can set the following parameters:

Parameter	Comments
Maximum distance (in bp) for two probes to be considered as neighbors	The program only considers probes to be neighbors if their genomic locations are within this threshold distance. The default value for this parameter is 1000 base pairs.
$P(X_{bar})$ <	• This parameter refers to the average *p*-value for the central probe and its neighbors. • The default cut-off value is 0.001. • Decreasing the cutoff value makes the detection more stringent.
Central probe has P(X) <	• The central probe is the red probe in Figure 106. • The default cut-off value is 0.001. • Decreasing the cut-off value makes the detection more stringent.

Parameter	Comments
At least one neighboring probe has P(X) <	• Neighboring probes are probes to either side of the central probe. The blue probes in Figure 106 are the neighbors of the central (red) probe. • The default cut-off value is 0.1. • Decreasing the cut-off value makes the detection more stringent.
At least **n** of the neighbors has P(X) <	• The default value for **n** is 1. • The default cut-off value for P(X) is 0.005 • Decreasing the cut-off value makes the detection more stringent.

OK Configures the event detection model with the parameters selected in the dialog box.

Cancel Closes the dialog box without changing any parameters.

Whitehead Per-Array Neighbourhood Model (Modified) Parameter Settings

This model is exactly like the unmodified model, except that in the modified model, the number of neighbors includes the probe itself, so the default value for **n** is 2 instead of 1. For information on the statistical algorithm, see "Whitehead per-array neighbourhood models" on page 271 and "Modified model" on page 295.

Purpose: This dialog box lets you customize the parameters of the Whitehead per-array neighbourhood model. The ChIP application uses this model to make binding calls based on the *p*-values of each probe and its neighbors.

To open: Select **Whitehead Per-Array Neighbourhood Model (Modified)** under Event Detection in the Analysis command ribbon, then click  . See "To apply event detection" on page 111.

Report Formats

Probe Report format

The probe report contains information about the probes in the current experiment result in tab-separated value (*.tsv) format. It contains one row for each probe in the array (or array set). The program generates a separate file for each array, and you can use a spreadsheet program to open them.

Probe report files contain the following columns, in this order:

Column	Description
Name	Passed through from the feature extraction program
ID	Passed through from the feature extraction program
Chromosome	Name of the chromosome (for example, chr17)
Start Location	The first base pair of the chromosomal location to which the probe binds
Control Type	The type of probe. Possible values are: blank, common, data, pos, negative, or FALSE.
IP Foreground	Passed through from the feature extraction program. This is the median pixel intensity for the foreground of the IP channel.
IP Background	Passed through from the feature extraction program. This is the median pixel intensity for the background of the IP channel.
WCE Foreground	Passed through from the feature extraction program. This is the median pixel intensity for the foreground of the WCE channel.
WCE Background	Passed through from the feature extraction program. This is the median pixel intensity for the background of the WCE channel.
IP Background SD	Passed through from the feature extraction program. This is the standard deviation of the intensities of the background pixels in the IP channel.

Column	Description
WCE Background SD	Passed through from the feature extraction program. This is the standard deviation of the intensities of the background pixels in the WCE channel.
Norm IP Signal	The IP channel normalized signal (foreground minus background), after both inter- and intra-array normalizations.
Norm WCE Signal	The WCE channel normalized signal (foreground minus background), after both inter- and intra-array normalizations.
Normalized Log Ratio	The natural log (base e) of the ratio of **Norm IP Signal** to **Norm WCE Signal.**
X	The difference between **Norm IP Signal** and **Norm WCE Signal,** divided by estimated error.
p-value	The probability that the **X** value is due to non-biological causes.
Z	The same as X (above), except that the distribution of Z values has a mean of zero, and a standard deviation of 1.
Xbar	Averaged neighborhood X, equal to the average of X for this probe and for any adjacent neighboring probes (within 1000 base pairs, user configurable)
P[Xbar]	The probability that X_{bar} is due to non-biological causes.
Is In BoundRegion	Set to 1 if this probe and its immediate neighbors within 1000 base pairs (user configurable) have sufficiently high X_{bar} values.
Exclude	Set to 1 if the feature extraction program or the analysis algorithm flagged the probe to be excluded (that is, the normalized signal surpassed the legal range).
Primary Annotation	Passed through from the design file. It identifies the transcript or gene ID nearest to this probe using simple proximity heuristics.
Secondary Annotation	Passed through from the design file. It typically represents features that overlap a window centered on the probe. The specific format and window size can vary based on the genome and array design.

Column	Description
Accession Numbers	Passed through from the design file. These accession numbers identify the gene(s) most likely to be regulated by the transcription factor(s) to which the probe binds. These identifications are made using a heuristic that looks for nearby genes, and checks their direction relative to the genomic location covered by the probe.
Primary Annotation Type	Passed through from the design file. It represents the relationship between the probe and the nearby Primary Annotation, and typically assumes a 10 kilobase pair window. The possible values are: • **PROMOTER** – This probe is upstream from the transcription start site. • **DIVERGENT** – This probe is upstream from two genes that are transcribed in opposite directions. • **INSIDE** – This probe is inside a gene. • **NONE** – There are no genomic features upstream or downstream of this probe.

Gene Report formats

The gene report contains one row for each probe in an array (or array set), grouped by the genes to which the probes bind. It is a tab-separated value (*.tsv) format file that you can open and analyze further with a spreadsheet program. For multiple arrays, the program creates a separate *.tsv file for each array. The program can create a gene report in one of three formats:

- **Show only gene names**– A single-column report that contains only the names of genes (and other loci) associated with the probes in the array. The program generates this report if you mark **Show only gene names** when you configure the report.
- **Show probe information** – A report that contains information about each probe, in addition to all of the information in the Standard Report. See the table, below. The program generates this report if you mark **Show probe information** when you configure the report.

The table below describes the column content of each of these report formats.

Column	Show only gene names	Standard report	Show probe information	Description
Gene	included	included	included	The primary name of this gene as specified in the design file. If more than one probe binds to the gene, the gene name appears only for the first probe.
Chromosome		included	included	The chromosome on which this gene is located.
Start		included	included	The first base pair of the locus to which this probe binds.
End		included	included	The last base pair of the locus to which this probe binds.
Min P[Xbar]		included	included	The best (minimum) of the P[Xbar] values of the probes for this gene. For a description of P[Xbar] see the probe report format.
Probe Name			included	The name of this probe.
Type			included	See "Primary Annotation Type" in the probe report format.
Normalized Log Ratio			included	The natural log (base e) of the ratio of **Norm IP Signal** to **Norm WCE Signal.** See the probe report format.

QC report format

The QC report summarizes the settings of the current analysis, and the overall statistics of each array. It also contains scatter plots and histograms that allow you to graphically display the overall results of the analysis. The program creates a separate QC report for each array in your experiment. Each report is a folder that contains an HTML file that you can open in an Internet browser, along with the graphics (*.png files) linked to the HTML file.

Summary

The summary table describes the overall statistics of the array, and the settings of the analysis. It contains the following specific information:

Item	Description
Total number of replicate sets	The number of replicate arrays (or array sets) that the program grouped together for this analysis.
Total number of Probes	The total number of probes on the array.
Significant Probes ($p < 0.001$)	The number of probes with $p < 0.001$, as calculated by the analysis.
Bound Probes	The number of probes that satisfy the requirements of the event detection model.
Segments	A segment is a contiguous sequence of bound and unbound probes that satisfy certain conditions. This item reports the number of bound segments, and the number of genes that correspond to these segments.
Normalizations	The kind(s) of normalization that the program applied during analysis.
Error Models	The error model the program used to calculate binding probabilities.

Details for replicate sets

Item	Description
Slide Name	The name of the microarray slide detailed in this section.
Number of Probes	Number of probes on the slide.
Bound Probes	The number of probes on the slide that satisfy the requirements of the event detection model.
Derivative Log ratio spread	An indicator of the noise in the experiment, and thus the minimum log ratio difference required to make reliable binding calls. For ChIP experiments, this value typically can be up to 0.5 (or even higher). For information on how the program calculates this value, see "Normalization Algorithms" on page 274.

Item	Description
Normalized Red/Green Correlation (R^2)	Similarity in the normalized intensities from both channels in a two-color experiment. The correlation is based upon negative control probes and gives an indication of the amount of dye bias present.
IP vs WCE	Scatter plot of normalized intensities of the IP and WCE channels. See Figure 107. Well-matched distributions indicate efficacy of normalization. The distributions may not match if a large proportion of genomic loci represented on the array are enriched, as might be the case for the characterization of a general transcription factor.
Enrichment (log scale) ratio vs Intensity	Scatter plot that shows the enrichment ratio as a function of a probe's signal intensity. See Figure 108. The distributions may not match if a large proportion of genomic loci represented on the array are enriched, as might be the case for the characterization of a general transcription factor.
Normalized Intensity Distribution	Histogram that shows the distribution of normalized signal intensities, by channel (red and green). See Figure 109. In general, a symmetric distribution that is free from skew indicates effective normalization. However, you can expect probes that are genuinely enriched to create additional density on the positive side of the distribution.
Normalized Ratio ($\log_2$) Distribution	Histogram of the distribution of normalized $\log_2$ ratios. See Figure 110. In general, a symmetric distribution that is free from skew indicates effective normalization. However, you can expect probes that are genuinely enriched to create additional density on the positive side of the distribution.

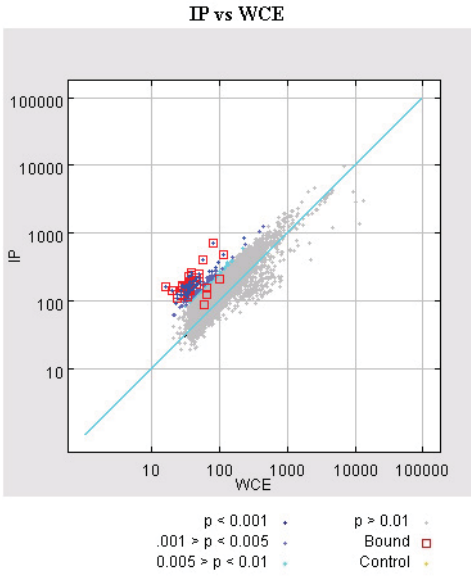

Figure 107 Example scatter plot of IP vs WCE intensities

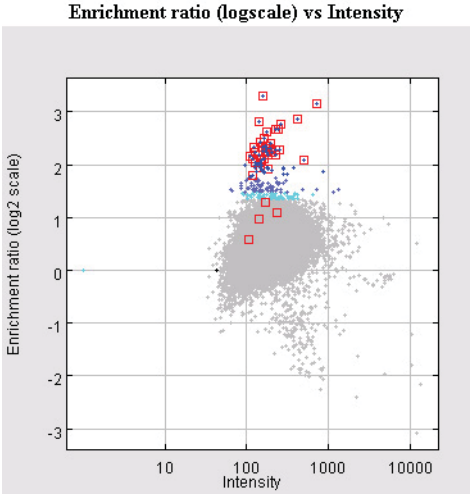

Figure 108 Example scatter plot of enrichment ratio versus intensity. This plot uses the same probe color legend as in Figure 107.

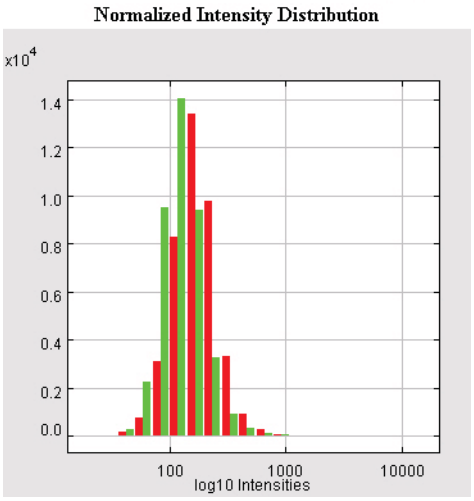

Figure 109 Example histogram of normalized intensity distribution. The distributions appear in a separate color for each channel.

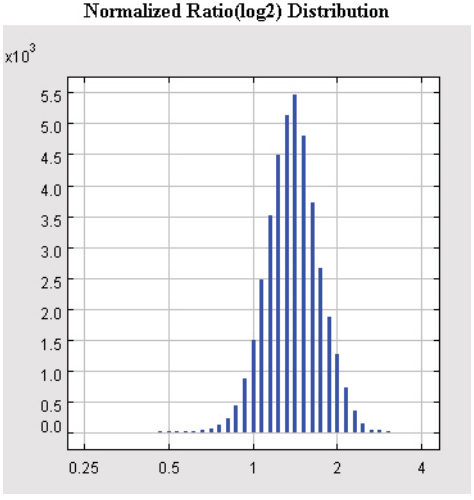

Figure 110 Example histogram of normalized $\log_2$ enrichment ratio distribution

6
ChIP Statistical Algorithms

This chapter provides implementation details for the algorithms used in the ChIP application of the Agilent Genomic Workbench. The ChIP algorithms allow for microarray normalization, and facilitate the statistical analysis of regions with an overabundance of probe signal.

If you plan a chromatin immunoprecipitation on chip (ChIP-on-chip) analysis, this chapter provides an overview of identification of protein-bound regions. Further contents follow the logical order of an analysis workflow after data import. The first set of topics compares and contrasts the normalization approaches, based on the algorithmic

procedures. The next section presents methods to arrive at error estimates. The final section explains the peak detection algorithms used in the Agilent Genomic Workbench.

Peak Detection and Evaluation Overview

ChIP-on-chip is a powerful technology to explore gene regulation by determination of precise genomic locations where a protein is bound. Typical ChIP microarrays have a high number of probes that cover both coding and noncoding DNA sequences. The Agilent Genomic Workbench detects robust peaks of probe signal that correspond to the chromatin immunoprecipitation. The Agilent Genomic Workbench also provides genomic visualization of the statistical analysis for identifying and evaluating regions of bound DNA.

Whitehead per-array neighbourhood models

One set of algorithms used in the Agilent Genomic Workbench to detect protein-DNA binding events is comprised of the two Whitehead neighbourhood models. These methods analyze the distribution of all probes on each array, to identify robust regions of increased probe signal ("peaks").

These algorithms examine groups of probe triplets that are significantly enriched, for robust detection of binding events. For example, in a probe-rich region, the area enriched by bound protein may span several probes. (See Figure 111 below.) In this case, probe signal attenuation is often observed at the genomic boundaries of the ChIP fragment. The Whitehead model samples every probe and its immediate upstream or downstream neighboring probe, to identify a robust estimate of the location of bound protein.

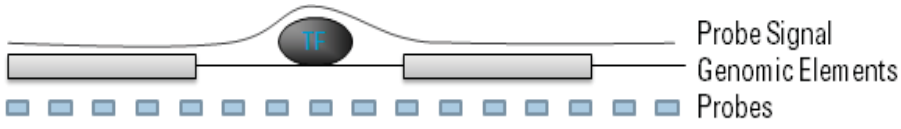

Figure 111 A graphical representation of the probe signal perturbation caused by a protein-DNA binding event, such as a transcription factor (TF)

Predefined peak-shape detection

A second algorithm used in the Agilent Genomic Workbench to detect protein-DNA binding events is the predefined peak-shape detection algorithm (PDPD). This method uses the expected distribution of sheared DNA fragments to model the shape of increased probe signal ("peaks").

The algorithm assumes that probes will show signal enrichment in a manner that is proportional to two factors — the enrichment at the peak center and the distance of the probe from the peak center. Probes that are farther away from the peak center (that is, the chromosomal position of the antibody target) will show enrichment only for longer fragments that span both the probe location and the peak center.

Definitions used in the statistical algorithms sections

This chapter uses these abbreviations and mathematical symbols:

Abbreviation or symbol	Definition
IP, *IP*	In a ChIP-on-chip experiment, the signal channel is called the IP or *immunoprecipitated* channel.
WCE, *WCE*	In a ChIP-on-chip experiment, the background channel is called the WCE or *whole cell extract* channel.
I	The probe intensity from either the IP or WCE
D	Any distribution of probe signals
A	A vector of values, also known as an array, is a one-dimensional matrix and is identified by the symbol *A*.
MAD	The median absolute deviation
UI	User interface for the Agilent Genomic Workbench – for entry of user-defined values into algorithms
left	Refers to probes closer to the front of a chromosome

Abbreviation or symbol	Definition
right	Refers to probes closer to the end of a chromosome
M versus A	Refers to the *M* vs. *A* plot, or a plot (usually log - transformed) which relates the ratio of intensities between channels to the average of intensities between channels. Examination of such a plot can reveal trends due to systematic errors.

Normalization Algorithms

Changes in probe signals within a microarray and across samples may be attributed to protein-DNA binding events or to systematic variation. To answer the biological questions that the ChIP experiment was designed to address, you need to measure only the true variation from a binding event. Application of data normalization allows you to limit the systematic variation in the data, such that true biological variations are more readily detected.

The Agilent Genomic Workbench can import data that has already been normalized by Agilent Feature Extraction software, or it can optionally apply user-defined procedures for both intra- and inter-array normalization. For example, if you use the Lowess normalization that Feature Extraction performs on Agilent two-color data, you do not have to normalize the data files again in the Agilent Genomic Workbench. For more information about Lowess normalization, refer to the *Agilent Feature Extraction User Guide*.

The normalization scheme provided by the Agilent Genomic Workbench contains both intra- and inter-array procedures. If you do not use the normalizations from Feature Extraction, each normalization step may be combined with any of the other normalization steps. It is important to make sure that the normalization steps and options are appropriate for your particular experimental design. The following sections describe each normalization procedure available in the Agilent Genomic Workbench.

Blank subtraction normalization

The blank subtraction normalization subtracts from all signals on the array a measure of central tendency of the signal from negative control probes. This gives a more robust estimate of the true signal ratio, independent of nonspecific binding.

Blank subtraction is a method to transform the data prior to the application of normalization algorithms. Data transformation is the process of applying mathematical modifications to the values of a variable. When you use a data transformation method such as blank subtraction, the Agilent Genomic Workbench recalculates the signal data values, and uses them in any subsequent analyses that you perform.

NOTE A more nuanced version of blank subtraction is available in the *Agilent Feature Extraction* software.

The Agilent Genomic Workbench uses median subtraction to estimate the central tendency of control (blank) probes:

Median subtraction

Purpose This normalization procedure corrects for nonspecific binding. It uses all control probes — the ones that do not overlap with protein binding events — as a baseline measurement for array probe signals.

Use Median blank subtraction uses the median of the negative control probes as a baseline. You apply this normalization if you know that some probes do not hybridize during an experiment. If the data are noisy or contain nonspecific binding on the control probe set (yielding outliers), then median blank subtraction may yield unwanted results. For example, subtraction of a large median estimate from the experimental probes, can produce negative probe signal values across the array.

Algorithm The program computes corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels, by subtraction of the background intensities from the foreground intensities. It then applies the normalization steps to these corrected values:

$$I_{IP} = normalized\ (I_{IP,foreground} - I_{IP,background})$$

$$I_{WCE} = normalized(I_{WCE,foreground} - I_{WCE,background})$$

Given a data vector for each probe on each array:

1. Compute the median of the negative control probes on the array for the IP and WCE channels.
2. Subtract the median from the signal intensities of every probe on the chip for both channels.
3. Flag any probes with a negative normalized signal for exclusion from further analysis.

Interarray median normalization

The interarray median normalization algorithm normalizes the median signal in both channels across all of the arrays in an experiment.

Purpose To increase the statistical power of the experiment, you often do microarray experiments with replicates. Furthermore, many ChIP-on-chip microarrays come in sets of multiple arrays. Because each array is a separate physical entity, variations between these arrays may exist due to differences in sample preparation, instrumentation, and hybridization. To facilitate the comparison and combination of values from these different arrays, the Agilent Genomic Workbench offers interarray normalization as a linear transformation of the signal data values on each array, for each channel.

Use If you have replicate arrays, or more than one array for a given genomic ChIP- on-chip experiment, use the interarray median normalization algorithm to ensure that comparisons between probe values are robust, and differences are not due to systematic variation.

Algorithm Before application of the normalization, the program computes corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels, by subtraction of the background intensities from the foreground intensities. It then applies the normalization steps to these corrected values:

$$I_{IP} = normalized\ (I_{IP,foreground} - I_{IP,background})$$

$$I_{WCE} = normalized(I_{WCE,foreground} - I_{WCE,background})$$

1 Create an array A as the intersection of probes between all arrays.

For each array:

2 Compute the median of the common control probes on the array (denoted “LACC” inside the design file). If there are no common control probes in the design, the Agilent Genomic Workbench uses all the probes on the array.

3 Compute the median intensity of each channel of the common probes in A.

For each channel:

4 Calculate the target median tm as the mean of the medians of blanks on each array.

5 Calculate the normalized signal S as:

$$S = (signal - median(blanks)) * tm / median(A)$$

6 Flag any probes with a negative normalized signal for exclusion from further analysis.

Intra-array (dye-bias) median normalization

Interarray normalizations control for chip-wide variations in intensity, due to inconsistent washing, inconsistent sample preparation, or other microarray technology imperfections. Two-color array experiments contain signal ratios determined by the concurrent use of contrasting dyes. The normalized values for two-color data are determined by the ratio between the signal and control channels. Dye-related artifacts can affect the intensities of these ratios. The intra-array (dye-bias) median normalization attempts to correct for artifacts caused by nonlinear rates of dye incorporation, as well as inconsistencies in the relative fluorescence intensity between some red and green dyes.

The intra-array median normalization algorithm normalizes the channels within each array so that medians are equivalent. This algorithm is a linear transformation. However, another normalization procedure, termed Lowess, applies a nonlinear normalization, and effectively normalizes by probes and by arrays. The Lowess approach is described in the next section.

NOTE If you import the data from the Feature Extraction software using the Lowess approach, then you need not apply the intra-array median normalization.

You have two options for intra-array normalization:

- Equalizing the IP and WCE channels
- Normalizing central tendency of log ratios to 1

For details, see the two subsections below.

Equalizing the IP and WCE channels

Purpose This normalization procedure corrects for dye-bias artifacts by transforming the WCE signal values. The algorithm uses the median values from both channels to calculate a transform, then applies the transform to correct for intensity differences between the dyes used in each channel.

Use Equalization of the IP and WCE channels uses the median of the nonexcluded probes in each channel. You apply this normalization if you have a two-color experiment without previous normalization steps (such as those available in the Feature Extraction software). Normally, this approach works well if you have a large number of regions in which you expect the ratio between signals to deviate from the value 1 (which indicates no change). However, if the majority of the genome is expected not to change probe value intensities between each channel, it may be appropriate to normalize the median of log ratios to 1, to avoid variance that can result from this algorithm.

Algorithm Before application of the normalization, the program computes corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels, by subtraction of the background intensities from the foreground intensities. It then applies the normalization steps to these corrected values:

$$I_{IP} = normalized\ (I_{IP,foreground} - I_{IP,background})$$

$$I_{WCE} = normalized(I_{WCE,foreground} - I_{WCE,background})$$

For each array:

1. Compute the median of the nonexcluded probes on the array for the IP and WCE channels.
2. Calculate the dye bias as the ratio between the medians of the IP and WCE channels.

For each WCE signal:

3. Transform the signal by multiplying the signal by the dye bias factor obtained in **step 2**.
4. Flag any probes with a negative normalized signal for exclusion from further analysis.

Normalizing central tendency of log ratios to 1

Purpose This normalization procedure corrects for dye-bias artifacts by transforming the WCE signal values. The algorithm uses the median values from both channels to calculate a transform, then applies the transform to correct for intensity differences between the dyes used in each channel.

Use This normalization should be applied if a two-color experiment is used without previous normalization steps (such as those available in the Feature Extraction software). Normally, this approach works well if the majority of the genome is expected not to change probe value intensities between each channel. If there are a large number of regions in which the ratio between signals is expected to deviate from the value 1 (indicating no change), it may be appropriate to normalize the median of log ratios to the central tendency of the IP and WCE channels. Equalizing the IP and WCE channels uses the median of the nonexcluded probes in each channel.

Algorithm Before application of the normalization, the program computes corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels, by subtraction of the background intensities from the foreground intensities. It then applies the normalization steps to these corrected values:

$$I_{IP} = normalized\ (I_{IP,foreground} - I_{IP,background})$$

$$I_{WCE} = normalized(I_{WCE,foreground} - I_{WCE,background})$$

For each array:

1. Compute the median of the nonexcluded probes on the array for the IP and WCE channels.
2. Calculate the dye bias as the ratio between the medians of the IP and WCE channels.

For each WCE signal:

3. Transform the signal by multiplying the signal by the reciprocal of the dye bias factor obtained in **step 2**.
4. Flag any probes with a negative normalized signal for exclusion from further analysis.

Intra-array (intensity-dependent) Lowess normalization

Intra-array normalization attempts to correct for artifacts caused by nonlinear rates of dye incorporation, as well as inconsistencies in the relative fluorescence intensity between some red and green dyes. The Lowess (LOcally WEighted polynomial regreSSion) algorithm normalizes the channels within each array using a nonlinear polynomial fit to the data, and effectively normalizes by probes and by arrays. The Lowess approach is also available in the Agilent Feature Extraction software. You can find more information about the application of Lowess in the Feature Extraction software in the *Agilent Feature Extraction User Guide*.

NOTE If you import the data from the Feature Extraction software using the Lowess approach, then you need not apply the intra-array Lowess normalization here.

Purpose This normalization procedure corrects for dye-related artifacts in two-color experiments that cause the Cy5/Cy3 ratio to be affected by the total intensity of the spot. A transform yields the log of the ratio of the signals versus the log of the intensities of the spot. A nonlinear fit is applied to a moving window across the data to obtain a good fit. This fit is used to normalize the data to a more linear response.

Use The algorithm attempts to correct for artifacts caused by nonlinear rates of dye incorporation, as well as inconsistencies in the relative fluorescent intensity between some red and green dyes. In the absence of bias, the IP signal should be independent of the WCE signal, and thus the data points would be scattered symmetrically around the 45° line.

In many cases, these artifacts can cause the graph of the IP versus WCE signal to be curved. The Lowess transform fits a polynomial curve through the data and uses this curve to adjust the WCE values for each measurement. When the resulting normalized data are graphed versus the adjusted control values, the points are distributed more symmetrically around the 45° line.

The regression curve can be computed using the signals in either *all data probes*, *all common probes* (probes whose names starts with “LACC”), or an even smaller set using the *all gene-desert probes* (probes whose names start with “LACC:GD”).

NOTE This option does not determine which probes are normalized; *all data probes* are always normalized.

If Lowess is applied, blank subtraction and interarray median normalization are turned off.

Algorithm This is a straight implementation of the Lowess method.[1] Before application of the normalization, the program computes corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels, by subtraction of the background intensities from the foreground intensities. It then applies the normalization steps to these corrected values:

$$I_{IP} = normalized\ (I_{IP,foreground} - I_{IP,background})$$

$$I_{WCE} = normalized(I_{WCE,foreground} - I_{WCE,background})$$

The Lowess method corrects for dye-bias artifacts by log transforming the IP and WCE signal values. The transform yields the log of the ratio of the I_{IP}/I_{WCE} signals against the log of the intensities (where intensity is the square root of the product $I_{IP}{}^{*}I_{WCE}$). A nonlinear fit is applied to a moving window across the data to obtain a good fit. This fit is used to normalize the data to a more linear response.

For each array:

1. Add a constant value to every I_{IP} and I_{WCE} value to avoid the problem of taking logarithms of negative values in subsequent steps.
2. Compute the signal ratio of the I_{IP}/I_{WCE} channels for the nonexcluded probes on the array.
3. Calculate the signal intensity of the square root of the product of the I_{IP} and I_{WCE} channels for the nonexcluded probes on the array.
4. Log_{10} transform the ratio and intensity values.

The resulting data are partitioned into a set of overlapping intervals, and a polynomial function is fit to each partition. For curved data trends, lower-degree polynomials (such as 1°, which yields a straight line) tend to

approximate the data very well within a small neighborhood, but deteriorate quickly. Higher-degree polynomials tend to overfit the data. The algorithm therefore uses a lower-degree polynomial fit across each partition created by the sliding window. A weighting function is applied to the data points to fit the polynomial function such that points farther away contribute less to the fit.

5 Partition the data obtained in **step 4** into n overlapping regions. The smoothness of the fit curve is inversely proportional to the size of the regions.

6 Starting with the first partition, apply the weighting function to the new data neighborhood.

7 Fit a polynomial of degree d to the weighted data using least-squares regression for that region.

8 Move forward to the next partition and apply **step 6** and **step 7**. Stop at the n^{th} partition.

Interpretation Lowess uses well-known techniques such as least-squares regression, and does not rely upon a set function to be fit to the data. However, it should be emphasized that the Lowess transform must be run on each data set, every time, and cannot be transferred to different data sets, as a slightly different result will occur.

Intra-array (intensity-dependent) variance stabilization

Intra-array normalization attempts to correct for artifacts caused by nonlinear rates of dye incorporation, as well as inconsistencies in the relative fluorescence intensity between some red and green dyes.

The Lowess normalization algorithm normalizes the channels within each array using a nonlinear polynomial fit to the data, and effectively normalizes by probes and by arrays. The variance stabilization method[2] is an alternative to Lowess normalization that fits a regression curve to signal intensities after applying an arcsinh(x) transform to each channel rather than a log(x) transform. The arcsinh(x) transform attempts to adjust the variance so that it is more consistent across the range of the data set.

NOTE If you import the data from the Feature Extraction software using the Lowess approach, then you need not apply Variance Stabilization here.

Purpose The variance stabilization procedure corrects for dye-related artifacts in two-color experiments that cause the variance of the intensities between the dye channels to be affected by the total intensity of the spot. This variance can be described as a relationship between the standard deviation of a given spot and the mean intensity of the spot. Log ratios are often skewed because of this trend. The variance stabilization approach uses a two-parameter error model to compress the reported ratios of probes with weak signals after blank-subtraction. After the transform is applied, the variance of the reported log ratios should be independent of the signal strength.

Use This normalization technique is based on the observation that in two-channel microarray experiments, increases in raw signal intensities tend to be accompanied by increases in variance between the channels. Log transformation yields a better trend with values that are more normally distributed, but suffer from intensity-dependent variation. The transformation therefore attempts to find the parameters of a transform that yields constant variance across the range of the data set.

Interpretation Variance stabilization allows quantitative analysis of log ratios between channels in two-dye microarray experiments by ensuring that the variance is uniform between all levels of signal intensity.

Error Models

Error modeling aids in the separation of true biological events, such as protein-DNA binding interactions, from signals that arise from artifacts in the measurement technology. You use error modeling to adjust the ratios (or signal intensities) according to an estimate of confidence in the ratio (or intensity) reported by a probe. The parameters for error calculation may include probe quality measures, or may be calculated from the distribution of probe values on each array. Likewise, the error model may be generated prior to data import, in the Agilent Feature Extraction software, or may be determined by the Agilent Genomic Workbench software.

Whitehead error model

Purpose

The purpose of the error model is to estimate how values measured on the array should be adjusted to reflect the confidence in their measurements. The confidence of such events is calculated based on the confidence of a gain in probe signal intensity.

For example, an intensity ratio with the value "3" can correspond to any intensity measurements in one channel that are three times higher than in the other channel. If the ratio of "3" corresponds to an intensity of 30 in one channel and an intensity of 10 in the other channel, this ratio is considered to be low in confidence, because the measurements are so close to the noise level. On the other hand, a ratio of "3" that corresponds to an intensity of 30,000 in one channel and an intensity of 10,000 in the other channel is much more reliable.

The probe signal confidence is reported as a "p-value", a parametric measure of how likely it is to observe a probe signal intensity of equal or greater magnitude, according to the error model. The significance of a p-value is scored over a range of 0 to 1. Significant events have p-values close to 0 ($p < 0.001$); events that likely are due to systematic variations have p-values close to 1.

The Agilent Genomic Workbench computes a p-value from the distribution of probes on each array:

- P(X) – An intensity-based p-value for each probe, derived from a variant on the X_{dev} quantity, as described by the Rosetta compendium.

Use The error model selects robust probe signals to include in the Whitehead per-array neighbourhood model, and therefore is "on" by default. The Whitehead error model is generally used for error modeling in the Agilent Genomic Workbench. However, if your data include probe quality scores from the Feature Extraction software, you may choose to use the Feature Extraction error model to see if the biological results are closer to those expected. For more information on the Feature Extraction error model, please refer to the *Agilent Feature Extraction User Guide*.

Algorithm The algorithm obtains a p-value, termed P(X), by first computing an "X" value for each probe. This quantity "X" correlates with the log ratio, but imparts a correction for low intensities.

The program computes intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels, according to the normalization options described in the previous sections:

$$I_{IP} = normalized\ (I_{IP,foreground} - I_{IP,background})$$

$$I_{WCE} = normalized(I_{WCE,foreground} - I_{WCE,background})$$

If no normalization steps are applied, then the intensities I for the IP and WCE channels are used directly.

1. Calculate the standard deviations of the IP and WCE channels. You may calculate this measure of variance from one of the following:
   - The standard deviation of background pixels. This is the default option. In such cases, the error model is based on the overall performance of the microarray.
   - The observed spread of the negative controls. An estimate of the variance of negative control probes gives a good indication of the error inherited by every probe on the array, due to technological limitations. However, it has not been tested for all situations.
   - The additive error as computed by the Agilent Feature Extraction software. For more information, refer to the *Agilent Feature Extraction User Guide*.
2. Calculate the value f as a factor to adjust the symmetry of the distribution of X. You may define this value, or the Agilent Genomic

Workbench may calculate it automatically. To calculate the value of *f*, the Agilent Genomic Workbench performs an iterative search of all values of *f* such that the difference between the values for $X < 0$ are minimized for probes in the highest and lowest quartile:

a Create an array *A* that includes the values from all nonexcluded, noncontrol probes in the bottom 25th percentile by WCE, with an *X* value less than 0.

b Create an array *A'* that includes the values from all nonexcluded, noncontrol probes in the top 25th percentile by WCE, with an *X* value less than 0.

Beginning with an initial estimate of $f = 1.0$, iterate over the following steps until convergence is met, or 50 iterations have occurred:

c Using the current value of *f*, assemble a distribution *D* by computing *X* for every member of *A*, using the equation below. Append both the *X*-value and -*X* to the distribution.

d Using the current value of *f*, assemble a distribution *D'* by computing *X* for every member of *A'*, using the equation below. Append both the *X*-value and -*X* to the distribution.

e Compute the difference between the standard deviations of *D* and *D'*.

f Report convergence if the difference computed in **step e** is less than 0.0001, or 50 iterations have occurred. Otherwise, update *f* as the previous value of *f*, multiplied by the ratio of the standard deviation of *D'* to the standard deviation of *D*.

3 Compute *X* for each probe, using the value of *f* computed in **step 2**:

$$X = \frac{I_{IP} - I_{WCE}}{\sqrt{f^2(I_{IP}^2 + I_{WCE}^2) + (\sigma_{IP}^2 + \sigma_{WCE}^2)}}$$

4 For each probe, compute P(*X*) as the probability of observing a value equal to or greater than *X*, assuming a normal distribution with the computed mean and standard deviation.

5 For each data probe *i*, compute a *Z*-score based on the *X*-values.

Replicate error model

The Whitehead error model in the Agilent Genomic Workbench supports the use of multiple biological or technical replicates. The program extends the error model for replicates by first defining X, and then allowing for multiple arrays in the calculations:

$$X = \frac{IP - WCE}{\sigma_{IP-WCE}}$$

$$= \frac{IP - WCE}{\sqrt{\sigma_{IP}^2 + \sigma_{WCE}^2}}$$

$$= \frac{IP - WCE}{\sqrt{f^2(\sigma_{IP,\,mult}^2 + \sigma_{WCE,\,mult}^2) + \sigma_{IP,\,add}^2 + \sigma_{WCE,\,add}^2}}$$

$$= \frac{IP - WCE}{\sqrt{f^2(IP^2 + WCE^2) + \sigma_{IP,\,add}^2 + \sigma_{WCE,\,add}^2}}$$

The denominator can be broken into its constituents:

$$\sigma_{IP} = \sqrt{(f \cdot IP)^2 + \sigma^2_{IP,\,add}}$$

and

$$\sigma_{WCE} = \sqrt{(f \cdot WCE)^2 + \sigma^2_{WCE,\,add}}$$

Then define:

$$\sigma_L = \left|\frac{L}{X}\right|$$

where $L = \log(IP/WCE)$.

The average log ratio is a weighted average of the replicate log ratios where each weight is:

$$w_i = \frac{1}{\sigma^2_{L_i}}$$

where L_i is the log ratio of the i-th replicate, so the log ratio of the probe is:

$$L_{avg} = \frac{1}{\sum w_i} \sum w_i L_i$$

Then, working backwards from this averaged log ratio, define the weighted average of Xs:

$$X_{avg} = \frac{L_{avg}}{\sigma_{L_{avg}}}$$

where the error in the averaged log ratio can be shown to be:

$$\sigma_{L_{avg}} = \sqrt{\frac{1}{\sum w_i}}$$

Peak Detection Algorithms

The Agilent Genomic Workbench detects robust peaks of probe signal that correspond to binding events. You typically normalize these probe signals, and apply an error model, as described in previous sections of this chapter. The normalization allows for direct probe-to-probe intensity ratio comparisons, and comparisons of biological or technical replicates. Error modeling estimates how values measured on the array should be adjusted to reflect the confidence in their measurements. The Agilent Genomic Workbench also provides genomic visualization and evaluation of the statistical analysis. Agilent offers two peak detection algorithms in the Agilent Genomic Workbench to identify and evaluate regions of bound DNA in a genomic context.

Whitehead per-array neighbourhood models

One set of algorithms used in the Agilent Genomic Workbench to detect protein-DNA binding events is comprised of the Whitehead per-array neighbourhood models. These methods use the distribution of all probes on each array to compute robust regions of increased probe signal ("peaks"). The algorithms do this by examining groups of probe triplets that are significantly enriched. The Whitehead models sample every probe and its immediate upstream or downstream neighboring probe, to identify a robust estimate of the location of bound protein. See Figure 112.

Figure 112 A graphical representation of the probe signal "neighbourhood" used in the Whitehead model. The diagrams show the binding event as a transcription factor (TF), and the resultant signals from discrete probes in the binding region. The Whitehead per-array neighbourhood models use probes in the immediate vicinity of the probe signal (here represented as a triplet) to evaluate whether the local increase in signal intensity is likely to represent a binding event.

Original model

The following outlines the original Whitehead per-array neighbourhood model in the Agilent Genomic Workbench. Many of the methods described below are also used in the modified Whitehead per-array neighbourhood model, which is detailed under "Modified model" on page 295.

Purpose Typical ChIP-on-chip microarrays use discrete probes that cover coding and noncoding genomic DNA. Because a protein-DNA binding event may span regions described by several distinct probes, you need a strategy to aggregate and evaluate such probe signals into a robust region of increased signal intensity. Such regions are termed "peaks". Peak detection is a method to identify binding events from complex genomic data.

Use After normalization of the data and application of an error model, the data analysis typically shifts from "low-level" analysis to "high-level" analysis. In low-level analysis, the program applies statistics at the probe level to account for systematic variations, and correlates quality information about probes with signal intensities. In high-level analysis, biological events are catalogued, statistically examined, and collected into information about biological annotation, pathway, and network. The peak detection algorithms in the Agilent Genomic Workbench provide a bridge between low-level and high-level analysis, enabling a complete data analysis workflow for ChIP-on-chip experiments.

Algorithm: $\mathbf{P}(X_{\text{neighb}})$ The Whitehead model uses two types of p-values, and then employs a heuristic that uses both values to predict whether a probe represents a binding event rather than noise in the data. These p-values are:

- P(X) – The intensity-based p-value for each probe, derived from the X_{dev} quantity. P(X) represents the error- corrected significance of the probe ratio. The Agilent Genomic Workbench may calculate this value directly, from the distribution of probe signals on each array, or you may import the value from the Agilent Feature Extraction software, by selecting the Feature Extraction error model.
- P(X_{neighb}) – A related p-value that uses the intensities of probes that represent regions of the genome that are close to each other. The program calculates this value using the adjacent probe triplets described in the introduction to this section.

Although the binding events (or epigenetic marks) assayed by ChIP-on-chip occur at genomic loci with base-pair resolution, you expect signals to span several hundreds of bases, due to the size of the DNA fragments produced by shearing (see Figure 112 on page 290). Therefore, if a peak is biologically genuine, you can expect that positive signals in a probe are corroborated by the signals of probes that are its genomic neighbors, provided the probes are close enough to each other.

The program computes this corroboration in terms of X_{neighb}, an extension of the p-value calculated as part of the Whitehead or Feature Extraction error model. The value X_{neighb} is a simple average of the X value of a given probe with its genomic neighbors. The program applies a distance threshold to determine whether neighboring probes are close enough to be

considered in the average of the probe triplet described in the introduction to this section. If not, it uses a value of $X = 0$ instead. X_{neighb} is calculated as follows:

1 Compute the mean X value for each probe on all arrays through the following simple function:

$$\overline{X} = \begin{cases} X & \text{X has no neighbor within range} \\ \frac{(X + X_{left})}{2} & \text{X has neighbor on the left only within range} \\ \frac{(X + X_{right})}{2} & \text{X has neighbor on the right only within range} \\ \frac{(X_{left} + X + X_{right})}{3} & \text{X has neighbors on the left and right within range} \end{cases}$$

2 Compute the distribution D as all mean X values of all probes on all arrays.

For each probe on all arrays:

3 For each data probe i, compute a Z-score based on the mean X value.

4 Compute the p-value of the mean X-value by evaluation of the area under the normal Gaussian curve from Z to infinity.

NOTE P(X) is the area under the normal curve from Z to infinity. The distribution of values for X are presumed to be Gaussian.

Whether a set of enriched probes identifies a genomic location at which a protein is bound is subject to several user-defined parameters. The UI allows specification of:

- The maximum distance (in bp) for two probes to be considered as neighbors (default: 1000 bp)
- P(X) and P(Xbar) criteria for a probe to be considered "bound" to another probe, extending a binding event area

A probe identifies the location of a "bound" protein if either of the following user conditions are met:

- The calculated *p*-value, P(*X*bar), is less than a threshold defined in the UI (default: 0.001) *and*:
  - The error model *p*-value, P(*X*), for the central probe is less than a threshold defined in the UI (default: 0.001) *or*:
  - The error model *p*-value, P(*X*), for at least one neighboring probe is less than a threshold defined in the UI (default: 0.1).
- The calculated *p*-value, P(*X*bar), is less than a threshold defined in the UI (default: 0.001) *and*:
  - The number of neighbors defined in the UI (default: 1) has an error model *p*-value, P(*X*), less than a threshold defined in the UI (default: 0.005).

The decision tree to call a probe "bound" is summarized in the following table and diagram:

Table 10 Parameters for Whitehead per-array neighbourhood model

Parameter	Comments
Maximum distance (in bp) for two probes to be considered as neighbors	The program only considers probes to be neighbors if their genomic locations are within this threshold distance. The default value for this parameter is 1000 base pairs. • To change the value, delete the old value and type a new one in the box.
$P(X_{bar})$ <	• This parameter refers to the average P-value for the central probe and its neighbors. • The default value is 0.001. • To make detection more stringent, decrease the value.
Central probe has $P(X)$ <	• The central probe is the red probe in Figure 113. • The default value is 0.001. • To make detection more stringent, decrease the value.
At least one neighboring probe has $P(X)$ <	• Neighboring probes are probes to either side of the central probe. The blue probes in Figure 113 are neighbors of the central (red) probe. • The default value is 0.1. • To make detection more stringent, decrease the value.
At least **n** of the neighbors has $P(X)$ <	• The default value for **n** is 1. • The default cut-off value for $P(X)$ is 0.005. • To make detection more stringent, decrease the value.

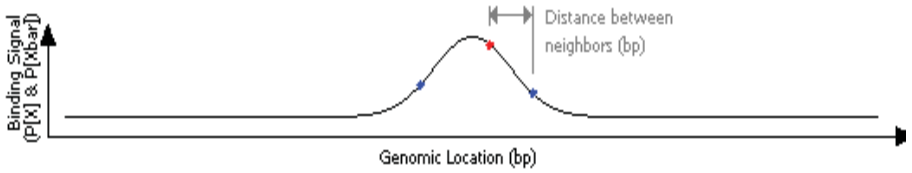

Figure 113 A graphical representation of the probe signal "neighbourhood" used in the Whitehead model

Interpretation Probe binding ends when the program finds all neighboring probes that pass the condition tree determined from the user-defined parameters. The program reports the average signal intensity found from bound probes,

and generates a plot as output. It renders the binding events as points in the visualization panel. The height of each point is equal to the average log ratio of that region.

Visualization When the program computes the binding regions, it plots them as a graph, colored by sample. For two or three simultaneous plots, it is often possible to distinguish the various regions based on the color blending.

It can still be difficult to read small regions. To zoom in on a region, click a chromosomal region of interest, or drag a rectangular selection around a genomic region.

Modified model

The modified Whitehead per-array neighbourhood model is very similar to the original model, with one exception. You can see this difference if you compare the dialog boxes for the parameter settings for the two models. Note the setting at the bottom of the dialog boxes.

- The original version says at least one of the neighbors has a p(*X*) less than 0.005.
- The modified version says at least two of the neighbors, including the center probe, have a P(*X*) less than 0.005.

Therefore, in the modified algorithm, the *center* probe, shown in red in Figure 113 on page 294, must pass the same P(*X*) value as the neighbors.

With the original Whitehead algorithm, the average of the probes P(*X*bar) could pass the threshold when the *neighboring* probes that flank the center probe were high, while the center probe was low. Because the ChIP binding event is centered on the center probe, the modified Whitehead algorithm ensures that the center probe always passes the P(*X*) requirement as well.

The overall decision tree for the modified Whitehead per-array neighbourhood model is similar to that of the original model. A probe identifies the location of a “bound” protein if either of the following user conditions are met:

- The calculated *p*-value, P(*X*bar), is less than a threshold defined in the UI (default: 0.001) *and*:
  - The error model *p*-value, P(*X*), for the central probe is less than a threshold defined in the UI (default: 0.001) *or*:

  - The error model p-value, P(X), for at least one neighboring probe is less than a threshold defined in the UI (default: 0.1).
- The calculated p-value, P(Xbar), is less than a threshold defined in the UI (default: 0.001) *and*:
  - The number of neighbors defined in the UI (default: 2, including the center probe) has an error model p-value, P(X), less than a threshold defined in the UI (default: 0.005). In the modified model, the center probe must pass this P(X) requirement, in addition to the neighbors.

The modified Whitehead per-array neighbourhood model is more reliable than the original model, but the software still includes the original model in case you have used it in the past and wish to compare data using the same algorithm.

Predefined peak-shape detection algorithm

A second protein-DNA binding event detection algorithm used in the Agilent Genomic Workbench is the predefined peak-shape detection algorithm (called "PDPD" hereafter). This method uses the expected distribution of sheared DNA fragments to model the shape of increased probe signal ("peaks").

The algorithm assumes that probes will show signal enrichment in a manner proportional to two factors — the enrichment at the peak center and the distance of the probe from the peak center. Probes that are farther away from the peak center (that is, the chromosomal position of the antibody target) will show enrichment only for longer fragments that span both the probe location and the peak center.

The model implies that the peak shape is derived from the shear-length distribution by computing the expected number of fragments that span the distance between the peak center and the probe location. See Figure 114.

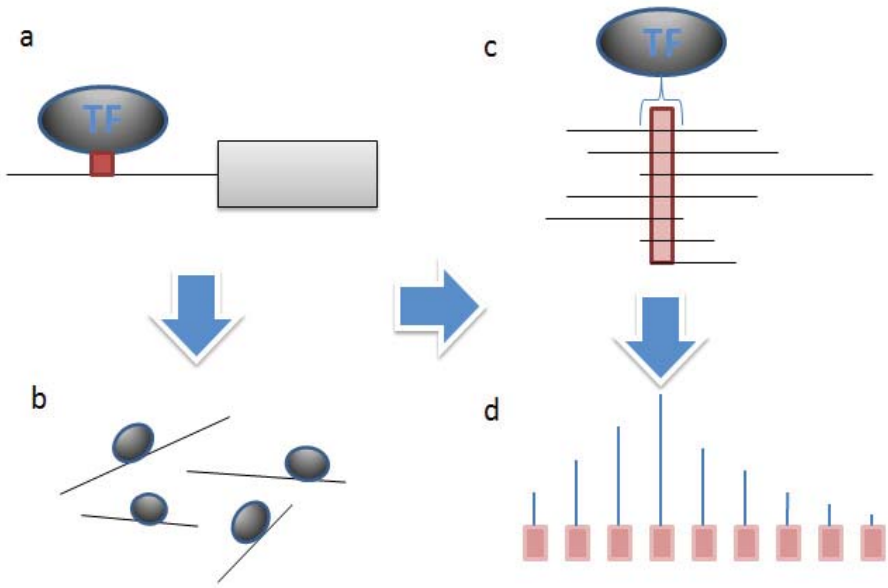

Figure 114 A graphical representation of the probe signal from sheared DNA used in the PDPD model. The diagrams show (a) the binding event as a transcription factor (TF), (b) shearing of the DNA enriched by bound TF, (c) a schematic of the way in which a portion of sheared DNA may overlap the TF site, and (d) the resultant signals from different sheared overlaps.

Purpose Typical ChIP-on-chip microarrays use discrete probes that cover coding and noncoding genomic DNA. Because a protein-DNA binding event may span regions described by several distinct probes, you need a strategy to aggregate and evaluate such probe signals into a robust region of increased signal intensity. Such regions are termed “peaks”. Peak detection is a method to identify binding events from complex genomic data.

Use After normalization of the data and application of an error model, the data analysis typically shifts from “low-level” analysis to “high-level” analysis. In low-level analysis, the program applies statistics at the probe level to account for systematic variations, and correlates quality information about probes with signal intensities. In high-level analysis, biological events are catalogued, statistically examined, and collected into information about biological annotation, pathway, and network. The peak

detection algorithms in the Agilent Genomic Workbench provide a bridge between low-level and high-level analysis, enabling a complete data analysis workflow for ChIP-on-chip experiments.

Algorithm

PDPD uses two methods to score peaks — a *p*-value, and an extreme value distribution (EVD) of log-likelihood values. It then employs a filter that uses both methods to predict whether a probe represents a binding event rather than noise in the data. The algorithm must first model the expected peak shape, followed by normalization and peak detection. Lastly, the peak scoring and score filtering are applied.

Modeling the expected peak shape

To model the expected peak shape, the peak-detection algorithm first approximates the distribution of sheared DNA fragments by a gamma distribution that uses a user-specified mean and standard deviation.

The mathematical model assumes that probes will show enrichment in a manner proportional to two factors — the enrichment at the peak center and the distance of the probe from the peak center. Probes that are farther away from the peak center (that is, the chromosomal position of the antibody target) would show enrichment only for longer fragments that span both the probe location and the peak center.

The model implies that the peak shape is derived from this shear-length distribution by computing the expected number of fragments that would span the distance between the peak center and the probe location. As a result, the peak shape is generated as the difference of two gamma distribution functions:

$$height(d) = G(d, \alpha + 1, \beta) - G(d, \alpha, \beta) \cdot \frac{d\beta}{\alpha} \quad \textbf{(1)}$$

Where d is the distance from the peak center, α and β are computed from the user-specified mean and standard deviation, and $G(d,\alpha,b)$ is the Gamma distribution function:

$$G(d, \alpha, \beta) = 1 - \frac{\Gamma(\alpha, d\beta)}{\Gamma(\alpha)} \quad \textbf{(2)}$$

Since peak-shape should be constant across an experiment, the Agilent Genomic Workbench can optionally try to learn the peak shape by analyzing the highest scoring peaks in the data set. The program reports the mean and standard deviation of the shear distribution that underlies

the idealized peak shape, and you can enter these as parameters if you choose to repeat the calculation. Alternatively, you can configure the software to automatically rerun the calculation on each array after learning its peak shape. (Note that this doubles the time that it takes to run the PDPD algorithm.)

Peak detection

The next steps identify peaks that match the predefined peak shape discovered in the steps described above. The PDPD algorithm first partitions the input data into mutually exclusive windows. The algorithm then performs local normalization of the data. Finally, it iteratively identifies peaks in each window and scores those discovered peaks.

Partitioning data into windows

For the initial partitioning of the data into windows, the algorithm identifies all contiguous chromosomal regions of the input data, such that

1. The distance between all neighboring probes within the window is less than a predefined width. (The width is based on the peak shape, and it corresponds to the distance at which two probes would not have any influence on the same peak), and
2. The number of probes in the window is at least 3.

Local normalization

For each potential position, the algorithm first normalizes the data to the local environment of the window by subtracting a baseline value from each data point. The baseline is computed from the average of all probes that are within 1 million base pairs of the data point. The size of this window is user-configurable and should be made smaller for smaller genomes. This baseline-subtraction normalizes for biases that exist in the environment of the window.

Also, because the density of probes will differ slightly across windows, the algorithm makes the density uniform by linearly interpolating the data at points that are “missing” between neighboring probes. By default, the interpolated points are spaced at 25 bp, but this spacing is user-configurable.

Iterative peak detection

Within a window, the algorithm searches in steps of 50 bp for potential positions for the peak center. This spacing is user-configurable, but decreasing it will increase the time it takes for the algorithm to run.

For each candidate position for the peak center, the algorithm computes the cross-correlation between the vector of the peak shape and the normalized data within its span. The data has been normalized to be of equal density across any window that might be examined. Therefore, the correlation is comparable across windows and is used as a measure of quality for placing a peak center at the current putative position. The peak position with the best cross-correlation is considered to be the predicted peak center.

After finding the peak center, the algorithm determines the peak height through an analytical procedure. The peak height is computed as the weighted average of the data around the peak center divided by the weighted average of the peak shape around the peak center, where both weighted averages are computed at the locations relative to the peak center at which there are measurements.

When computing the weighted average of the data, the algorithm uses the residual data, which is defined here as the original data after subtracting any previous peaks found in the region. When computing the weighted average of the peak shape, the algorithm takes the height at each position for which there are measurements of the peak shape. The weighting is done by the height of the peak shape at each position.

The peak is evaluated according to is log-likelihood. The log-likelihood of an entire peak is computed by summing the likelihood of each data point. The log of the ratio values are assumed to be normally distributed either (a) with a fixed standard deviation, or (b) with a standard deviation provided by an external error model around each data point. (This is user-configurable.)

In the case of the assumption of a fixed standard deviation, the data log-likelihood is simply proportional to

$$[-\log(predicted) - \log(observed)]^2 \quad \textbf{(3)}$$

In this former case, all other terms of the normal distribution are constant (since the standard deviation is fixed for all data points) and thus do not need to be computed. In the latter case, of varying standard deviation based on an error model, the term is

$$-\log(error) - \left[\frac{\log(predicted) - \log(observed)}{error}\right]^2 \quad \textbf{(4)}$$

Where *error* is the standard deviation of the log ratio of the measurements, as predicted by the error model.

Once the algorithm identifies a peak, it subtracts its values from the original data, leaving only the residuals of the data. The algorithm then attempts to find subsequent peaks in the window as described above, except it now uses the residual data.

Peak scoring

There are two ways that the algorithm scores peaks:

- By using a *p*-value, and
- By using a "score" based on an assumption of an extreme value distribution of log-likelihoods.

The algorithm computes these quantities through random sampling. By default, the algorithm performs 100 samples, but this is user-configurable. Increasing the number of samples (for example, to 1000) improves the accuracy of the predictions, but it also increases the time it takes to run the algorithm in a linear fashion (by 10 times, in this example.)

To perform each random sampling, the algorithm considers the current tentative peak position, but it randomly assigns values to the locations of the original data measurements instead of using the original data. These values are taken from randomly selected measurements on the array.

Randomly selected values may come from locations that have different baselines. Therefore, when picking a random value, the algorithm first subtracts the local baseline of the original genomic location of the value and then adds the baseline of the location into which the probe is inserted — the baseline at the region in which the tentative peak is being considered. Both of these baselines are computed as described above.

For each set of randomly assigned values in the window, the algorithm computes the log-likelihood of the peak centered at the putative position, and stores the result. To evaluate the log-likelihood of a peak found in the real data, the algorithm compares its log-likelihood to the distribution of log-likelihoods of the peaks found when fitting randomized data.

During this comparison, the algorithm estimates the p-value according to the number of times that peaks found in the randomized data exhibit better log-likelihoods than the peak discovered in the real data. For example, if 95 of the 100 randomized values have log-likelihoods that are better than the number associated with the putative peak, then $p < 0.05$.

The algorithm also assigns a "score" by testing the significance of the log-likelihood under the assumption of an extreme value distribution (EVD) of the log-likelihoods from the randomized data. The significance derived from the test is converted to a score by computing the negative $\log_{10}$ of the significance.

Filtering results

The Agilent Genomic Workbench allows you to specify both an initial p-value threshold and a score threshold. The default settings are:

p-value threshold: 0.01 Score threshold: 0.0

With these settings, any peak with a p-value < 0.01 will be retained (and reported) by the Agilent Genomic Workbench, regardless of its EVD-based score. Alternate filtering schemes may also be useful, for example:

p-value threshold: 1.0 Score threshold: 3.0

In this example, any peak with an EVD-based score above 3.0 will be retained, regardless of its p-value.

Interpretation The program creates a score by comparing the fit of the peak shape of a binding region to a predefined peak shape from randomized data. A p-value and score threshold are produced, which allows for fine-tuning of those binding regions that are more likely to be reproducible in a number of experiments with similar noise envelopes.

Visualization When the program computes the binding regions, it plots them as a graph, colored by sample. For two or three simultaneous plots, it is often possible to distinguish the various regions based on the color blending.

It can still be difficult to read small regions. To zoom in on a region, click a chromosomal region of interest, or drag a rectangular selection around a genomic region.

References

1 Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992). "Local regression models." Chapter 8 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

2 Huber, W., von Heydebreck, A., *et al.* (2002). "Variance stabilization applied to microarray data calibration and to the quantification of differential expression." *Bioinformatics* 18 Suppl 1: S96-104.

3 Bolstad, B. M., Irizarry, R. A., *et al.* (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." *Bioinformatics* 19(2): 185-193.

www.agilent.com

In this book

The User Guide presents instructions on how to analyze your ChIP microarray data with Agilent Genomic Workbench 7.0.

PR7000-0045

Revision A, December 3, 2015

G3800-90043

Agilent Technologies