ChIP Interactive Analysis

Agilent Genomic Workbench 5.0

User Guide

Agilent Technologies

Notices

Edition

First Edition, April 2009

Printed in USA

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

Software Revision

This guide is valid for 5.0 and later revisions of the Agilent Genomic Workbench 5.0 software, until superseded.

Warranty

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

Technology Licenses

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

Restricted Rights Legend

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

Safety Notices

CAUTION

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

WARNING

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

In This Guide...

This guide contains information to use the interactive analysis portion of the Agilent Genomic Workbench 5.0 ChIP application program. See the *Workflow User Guide* to learn how to set up ChIP analysis methods for running workflows.

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

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

2 Visualizing 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.

3 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.

4 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.

5 ChIP Statistical Algorithms

This chapter describes implementation details of the algorithms used in the ChIP application of Genomic Workbench 5.0. The ChIP application algorithms facilitate the statistical analysis of ChIP-on-chip experiments, i.e. detected genomic regions with bound protein.

Contents

1 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 Agilent Genomic Workbench.

The program lets you import many different kinds of files, including array data and design files from Agilent products and other sources, experiment files from DNA Analytics 5.0 and its predecessor DNA Analytics 4.0, genome build files, and other content such as annotation tracks and gene lists.

Once you have imported Feature Extraction files into the program, you can assign identification information and attributes to the array files through the Sample Manager tab. See the *Sample Manager Guide*.

To view or analyze the ChIP data you have imported, you organize the data files into logical units called *experiments*. Experiments are central to the array viewing capabilities of Genomic Workbench. After you create them, and populate them with array data, you can then use the Preprocessing, Analysis and Report tabs of the program to analyze the data in the experiment for binding events.

For an overview of the Preprocessing, Analysis and Reports options and what they do, see the *Agilent Genomic Workbench 5.0 Quick Start Guide*.

For details see Chapters 3 and 4 in this guide.

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 Guide*.

Importing Files

You can use the Home tab to import many kinds of files into 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.

Type of file	Comments	See these topics
Microarray data files	• Agilent Feature Extraction (*.txt) data files • GenePix/Axon (*.gpr) data files	"To import Agilent FE or GenePix/Axon data files" on page 14
Microarray design files	• Agilent GEML (*.xml) design files • GenePix/Axon (*.gal) design files	"To import Agilent GEML design files" on page 16 "To import GenePix/Axon design files" on page 17
Genome builds	Agilent-supplied genome information for human, mouse and rat genomes	"To import a genome build" on page 18
Tracks	BED format annotation track files	"To import tracks" on page 19
Array attributes	.txt files that you have created yourself or previously exported from Agilent Genomic Workbench	"To import array attributes" on page 20
Experiments	ZIP format file of experiments exported from DNA Analytics	"To import an experiment file" on page 21
Gene list	*.txt file – Creates a new gene list in the program with the list of imported genes	"To import a gene list" on page 21

NOTE To import sample attributes files (SAFs), you use the **Sample Manager** tab. See the *Sample Manager Guide*.

To select a new location for data files

By default, the program stores design, data and experiment files in C:\Program Files\Agilent\Genomic Workbench Standard (or Enterprise) Edition 5.0\data. If you like, you can select a different location.

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

The Preferences dialog box appears. See "Preferences – Miscellaneous Tab" on page 178.

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

An Open dialog box appears.

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

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

4 Click **OK.**

To import Agilent FE or GenePix/Axon data files

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

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

If you import Agilent Feature Extraction files, the program requires the representative GEML array design files. If you import GenePix/Axon data files, the program requires the representative GenePix/Axon *.gal design files. See "To import Agilent GEML design files" on page 16 or "To import GenePix/Axon design files" on page 17.

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

- To import Agilent FE data files, click **Import > Array Files > FE File...**
- To import GenePix/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 166.

2 To select a file for import, click its name. To select additional files, control-click their names.

3 Do one of the following:

- For Agilent FE files, click **Open.**
- For GenePix/Axon files, click **Import.**

In either case, the Agilent Feature Extraction Importer dialog box appears. "Agilent Feature Extraction Importer" on page 133.

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 (Return** on a Mac).
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 ration (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 Agilent GEML design files

You must import Agilent GEML (*.xml) microarray design files into Genomic Workbench that match the Agilent Feature Extraction data files. 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 166. The dialog box shows only *.xml files.

2 To select a file for import, click its name. To select additional files, control-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 169.

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 array 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 GenePix/Axon design files

You can import GenePix/Axon (*.gal) microarray design files into Genomic Workbench. The program requires the GenePix/Axon design files that match all GenePix/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 166. The dialog box shows only *.gal files.

2 To select a file for import, click its name. To select additional files, control-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 201.

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 array from the list, click its **Remove** button  .

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

5 Click **Start Import.**

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

To import 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 170.

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.
RefSeq File	This file contains information on gene locations for Gene View. **a** Click **Browse...** A dialog box appears. **b** Select the desired file, then click **Open.**
Cyto-band File	This file contains the graphic information on the cyto-bands for Genome and Chromosome Views. **a** Click **Browse...** A dialog box appears. **b** Select the desired file, then click **Open.**

3 Click **OK.**

To import tracks

You can import BED format track files into DNA Analytics. 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 171.

2 Set the following. All are required.

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

3 Click **OK.**

The program imports the track. To view the track in Gene View, and to manage tracks, see "To show tracks in Gene View" on page 56.

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 much more easily. See the *Sample Manager User Guide*. This menu item will be eliminated from the next version of the program, and the only way to import array attributes will be through the Sample Manager application.

To import an array attributes file

1 Click **File > Import > ArrayAttributes.**

The Import microarray attributes dialog box appears. See "Import" on page 166.

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

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

Array attributes files must follow these guidelines:

- The first line of the file contains the names of array attributes, separated by tabs.
- The first attribute must be `Chip Barcode`.
- The rest of the lines of the file list the values of each attribute, one line per array. The order of attribute values must follow the order of attributes in the first line of the file.

Here is an example.

Chip Barcode	ArraySet	Hyb'd By
251270010402	E986	Maurice R.
251270010423	E986	Maurice R.
251270019455	E986	Maurice R.

To import an experiment file

In Genomic Workbench, an experiment is a set of links to microarray data and design files, and any associated results. A Genomic Workbench experiment file is a single ZIP file that contains the design and data files associated with one or many experiments. You can import experiment files created in Genomic Workbench on another computer, as well as DNAx 4.0 experiments.

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

The Import Experiments dialog box appears. See "Import" on page 166.

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, populated with the appropriate arrays, in the Experiment pane.

NOTE

Genomic Workbench experiment files contain all of the design and array data files associated with 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 39.

To import a gene list

(Available only with DNA Analytics application license) A gene list file is a plain text (*.txt) file that contains one gene name per line. When you import a gene list into 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 restrict the appearance of genes and data, in Gene and Chromosome Views. See "Gene List" on page 162.

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 166.

3 Select the desired gene list file. To select additional gene list files, control-click their names.

4 Click **OK.**

Working with Experiments to Organize Imported Data

This section describes how to arrange imported array data and designs into organizational units called *experiments*. Experiments, found 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 contain saved experiment results.

To view the array designs and data in the program

- To view the directory of data in the program, use the Data pane (Figure 1). Double-click a folder to expand or collapse it, or click the ⊞ and ⊟ buttons.

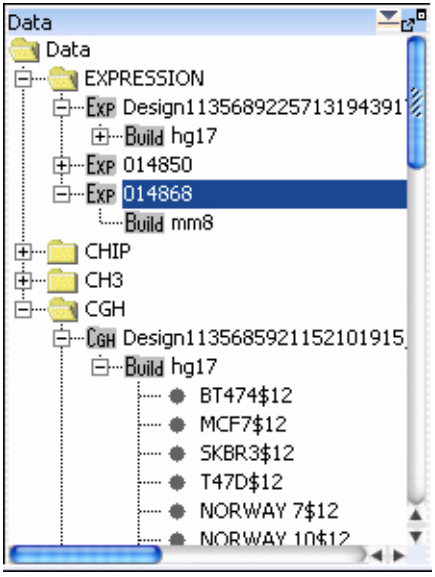


Figure 1 Data pane of the Navigator

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 details, see "Data pane – actions and shortcut menus" on page 104.

Many icons can appear in the Data pane. See "Data pane – icons, special text, and buttons" on page 103 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 30.

To create a new experiment

In Genomic Workbench, *experiments* are organizational units that contain links to data and design files. To view 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 141.

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 25.

- To create an experiment and populate it with data 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 151.

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

  c In **Array List,** click the name of an array that you want in your experiment. Control-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 151 for details.

  e Click **OK.**
  The program creates the new experiment, and populates it with the selected arrays.

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

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, any number of 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 151.

3 Under **Select Design,** select the design file and genome build associated with the arrays you wish to add.

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

4 In **Array List,** select the arrays you wish to add to the experiment. To select a single array, click its name. To select additional arrays, control-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 151 for details.

6 Click **OK.**

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

You can change the order in which arrays appear in an experiment in tables in Tab View. If you choose to display separate scatter plots in Gene View and Chromosome View for each array, the array order also determines the order in which these plots appear. You can use this feature to organize your arrays more logically, or to make it more convenient to view 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 149.

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 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 given 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 you wish 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.**

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 32.

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

2 In the **Arrays** or **Calibration Arrays** folder of the desired experiment, click the name of an array to select if for removal. Control-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 edit the attributes of an experiment

(Available only with DNA Analytics application license) Array attributes are pieces of array-specific information, such as chip barcode or hybridization temperature. You can show or hide attributes for the arrays in the experiment with the Experiment Attributes dialog box. See "Export Array Attributes – Array" on page 154.

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

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

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

   The Show/Hide Columns dialog box appears. See "Show/Hide Columns" on page 206.

4. Mark the check boxes whose attribute columns you want to show, or clear the check boxes whose attribute columns you want to remove.
5. Double-click the cell whose array attribute value you want to change.
6. Click **Save Changes**.

   If you do not want to save the changes, click **Cancel Changes**.

7. Click **Close**.

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

To edit the attribute values of a specific array

Array attributes are pieces of array-specific information such as chip barcode or hybridization temperature. You can view 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 - Attribute Tab" on page 173. 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 view 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

Managing Content

This section describes how to create, find, rename, update, combine, and/or remove several types of content stored in Genomic Workbench. To display the data, gene list and track content, see Chapter 2, "Visualizing ChIP Data and Other Content".

To view a list of the content stored in the program

The Data and My Entity List panes of the Navigator show the content stored in Genomic Workbench.

Data pane – Shows all of the array data files stored in the program, organized by application (CGH, ChIP, or methylation, for example), then by array design, then by applicable genome build. For more information, see "To view the array designs and data in the program" on page 23.

My Entity List pane – Shows the gene lists and tracks stored in the program. To view 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.

NOTE Content that is available exclusively in the Design tab, such as the probes and probe groups that you use to create custom microarray designs, is covered in a separate section. See

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 101.

1. Type a search term in the box at the top of the Navigator. The search term is not case-sensitive, but it must reflect the entire name of the content item that you want to find. You can use asterisks (*) as wildcards to represent a group of unspecified characters.
2. By default, the program searches all panes of the Navigator. To restrict 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 view 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 145.

To update probe annotation in design files

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

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

The Preferences dialog box appears.

2 In the Miscellaneous tab, under **eArray User Details,** type your eArray **Username** and **Password.** See "Preferences – Miscellaneous Tab" on page 178.

3 Click **OK.**

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

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

A confirmation dialog box appears.

6 Click **Yes.**

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

To rename an array in the Data pane

This topic describes how to rename an array in the Data pane. 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. However, the name of the array is unaffected 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 27.

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

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

An Input dialog box appears.

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

The program renames the array.

To remove data or design files from the program

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

1 If an array that you want to delete is associated with an experiment, first delete it from the experiment. See "To remove arrays from an experiment" on page 27.

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

3 Do one of the following:

- For array data files, click the name of the first array, then control-click the names of additional arrays within the same design.
- For array design folders, click the name of the first design folder, then control-click the names of additional ones. This selects the designs and all array data files within them for deletion.

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

A confirmation dialog box appears.

5 Click **Yes.**

The program deletes the selected files.

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

To create a gene list

When you create a gene list, you create a list of the genes in a contiguous chromosomal region that you define. To create a list of genes in multiple regions, create multiple gene lists, and combine them. See "To add one gene list to another" on page 34.

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

a In Genome View, select the desired chromosome.
The selected chromosome appears in Chromosome View.

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 122.

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 Experiment" on page 141.

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

4 In the dialog box select the desired 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 view the genes in a gene list

You can view 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 Report Settings Parameter Settings" on page 163.

You can also create gene lists. For details, see "To create a gene list" on page 33.

To add one gene list to another

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

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

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

A dialog box appears. For details, see "Add Gene List <name> to" on page 132.

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

4 Click **OK.**

To rename a gene list

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

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

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

An Input dialog box appears.

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

To delete gene list(s)

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

2 Click the name of a gene list that you want to delete. Control-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 combine tracks

You can create a track that contains elements from two or more existing tracks. The existing tracks must be available in 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 137.

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 that you want to combine.

6 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.

7 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.

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.**

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 view the details of a track

The table that you bring up 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.**

Data describing the track appear in a Track table. See "Track" on page 207.

To rename a track

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

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

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

An Input dialog box appears.

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

To delete tracks

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

2 Click the name of a track that you want to delete. Control-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. You can also export many additional kinds of files from the separate data analysis applications and other parts of the program. See the appropriate User Guides.

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 short-cut menu for an experiment.

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

OR

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

The Export Array Attributes dialog box appears with the Array tab displayed. See "Export Array Attributes – Array" on page 154.

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 associated with the arrays you wish to add.

The arrays associated with 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, control-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 "Export Array Attributes – Attribute" on page 156.

All of the attributes for the arrays are already located in the Selected Attribute List. If you don't want all the attributes exported, then you must transfer those to the Available List.

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

7 Click  .

8 Click **OK**.

The Export dialog box appears. See "Export" on page 152.

9 Select the folder in which to locate the attributes, and click **Export**.

The attributes will be 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 158.

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

3 In **Export Format,** select one of these options:

- **5.0 Format** – Exports the experiment(s) in a format that you can import into Genomic Workbench. This is the most current experiment format, but it is not compatible with previous versions of the program.
- **3.0 Format** – Exports the experiment(s) in a format that you can import into Agilent CGH Analytics 3.0 or later. This is a "legacy" format that you can use to maintain compatibility with earlier versions of the program.

4 Click **OK.**

An Export dialog box appears. See "Export" on page 152.

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

6 Click **Export.**

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

To export a gene list

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

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

A Save As dialog box appears.

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

3 Click **Save.**

A 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 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 159.

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

3 Click **OK.**

An Export dialog box appears.

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

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

To copy what you see in the main window

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

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

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

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

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 148.

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) associated with 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.**

2
Visualizing 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.

To find out how to import, organize, manage and export data and other content, Chapter 1, “Importing, Managing, and Exporting ChIP Data and Other Content”.

To learn about the options for the main window and the dialog boxes for visualizing data, see Chapter 4, “ChIP Interactive Analysis Reference”.

See the *Agilent Genomic Workbench 5.0 Quick Start Guide* for an overview of all the applications you can use with this software, both those that require a license and those that are free capabilities.

Activating an Experiment for Visualizing Data

An experiment is a set of links to microarray data and design files, and any associated results.

When you activate an experiment and the Preprocessing and Analysis options have not been turned on or set to apply, the program shows the log ratio data of selected arrays in the active experiment, if certain options have been set.

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

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

To activate an experiment

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

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

- Create a new experiment and populate it with data. See "To create a new experiment" on page 24.
- Import a saved DNA Analytics 4.0 ChIP experiment. See "To import an experiment file" on page 21.

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 chosen to show the results of an algorithm calculation, then moving average or aberration results appear for the first array when you activate the experiment, if you have not selected any other arrays.

You can select or deselect arrays in the experiment both before and after you activate it. Every time you select or deselect an array in an *active* experiment or change a setting, the program re-analyzes the new data set with the changed settings. See "To select or deselect arrays in the experiment" on page 45.

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

To select or deselect arrays in the experiment

To include arrays for view and analysis, you select them from among the arrays available either in an inactive experiment or the active one. When you first create an experiment, the program automatically sets the first array in the experiment for analysis. If you do not select any more arrays for analysis, only the first one will be analyzed when the experiment is activated.

To select the arrays for analysis before experiment activation:

1. Hold down the **Shift** key to highlight contiguous arrays or hold down the **Ctrl** key to highlight noncontiguous arrays.
2. Right-click the highlighted arrays, and click **Select**.

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

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

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 aberration results and moving averages related to this array in this color. To configure a custom color for the array, see "To change the display color of an array" on page 47.

To select or deselect arrays in an *active* experiment:

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

   To include additional arrays, control-click their names. To include a contiguous block of arrays, click the name of the first array in the block, then shift-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 re-analyzes 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 78.

   To customize the appearance of the results in Genome, Chromosome, and Gene Views, see the *Data Viewing Guide*.

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.
- Control-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 details on Tab View, see the *Data Viewing Guide*.

To change the display color of an array

The color assigned to an array affects the color of its icon when you select the array within an experiment. It also affects 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 (Edit Array Color) — Swatches Tab" on page 197.

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 right-most color sample shows the original color for comparison.

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

You can also manage all of the colors for all of the arrays in an experiment. Right-click the desired experiment, then click **Edit Array Color...** For details on the dialog box that appears, see "Edit Array Color" on page 148.

Visualizing Array Data

After you activate 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

1 In the Gene View, click **Scatter Plot,** and do any of the following:

To do this	Follow these steps
Show or hide the log ratio data point	• To show the data points – Mark the **Log Ratio** check boxes • To hide all data points – Clear the **Log Ratio** check boxes.
Show or hide the intensity ratio (ChIP) data point	• To show all data points – Mark the check box for **Intensity Ratio.** • To hide all data points – Clear the check box for **Intensity Ratio**.
Change the size of data points	• In **Point Size,** select a size for the data points.

2 Click X to close the Scatter Plot window.

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

The Preferences dialog box appears. "Preferences – View Tab" on page 181.

2 Do any of the following:

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

3 Click **OK.**

To print the scatter plot

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

1 Click **File > Print.**

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

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

- To navigate 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 view	• In Genome View, click the desired chromosome. All other views switch to the selected chromosome.
View 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.
Re-center Gene View or Chromosome view	Click anywhere in Chromosome View, or anywhere within the scatter plot in Gene View. The location you click becomes the new cursor location.
Move all views to a specific genomic location	**a** Click **Home > Go To Gene/Genomic location...** A dialog box appears. **b** Under **Genomic Location,** select a **Chromosome,** and type a **Base Position.** **c** Click **Go.** All views move to the selected location.
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 corresponding to the selected entry.

To do this	Follow these steps
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 131.

Visualizing Content (Gene Lists/Tracks)

To show gene lists in Gene View

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

You can import gene lists into DNA Analytics, and you can also create them in the program and export them. See "To import a gene list" on page 21, and "To export a gene list" on page 40.

In Gene View, the names of all genes normally appear in gray. When you apply a gene list, the program highlights the listed genes in their defined display color. You can also restrict 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 33, or "To import a gene list" on page 21.

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 associated with 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.**

Select gene list display color

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

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

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

The Gene List dialog box appears.

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

A dialog box appears.

4 Select the desired color.

The dialog box offers three different ways to choose the desired color. See "Select gene list display color" on page 54.

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

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

View a gene list as a table

You can view 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 Report Settings Parameter Settings" on page 163. 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 "Select gene list display color" on page 54.

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

You can also export a gene list. See "Choose Gene List Color" on page 136.

To change the appearance of genes in Gene View

You use the 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 **Preferences.**

The Preferences dialog box appears.

2 Click **Tracks**.

"Preferences – Tracks Tab" on page 179.

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

1 Select and show microarray data. See "To activate an experiment" on page 44.

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 **Preferences.**

The Preferences dialog box appears. See "Preferences – Tracks Tab" on page 179.

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.

Within the Preferences – Tracks dialog box, you can modify the appearance of tracks in several additional ways. See the table below.

To do this	Follow these steps
Include track information in reports	**a** In the list of tracks, in the **Show in Report** column, mark the check boxes of the desired tracks. **b** Click **Apply.** Doing this adds a column with the hits from the track file. For each aberrant interval, it reports the entries from the track file that falls under that interval in that separate column.
Show or hide annotations in all tracks	• To show annotations in all tracks: under **Tracks,** mark **Show Annotations.** • To hide annotations in all tracks: under **Tracks,** clear **Show Annotations.**

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.

Show track information in reports

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

2 Click **Apply.**

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

Restrict data to the genomic boundaries of the track

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

3 Mark **Genomic Boundaries**.

You can remove the boundaries by clearing the check box.

Display tracks in UCSC Browser

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

The UCSC Browser appears if you are connected to the Internet.

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

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 containing 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 18.

The program will not let you add arrays belonging to one genome build to an experiment containing arrays of a different genome build.

See also "To create a new experiment" on page 24, "To add arrays to an experiment" on page 25, and "To activate an experiment" on page 44.

To copy Views

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

1 Click **View > Copy.**

A menu of Views appears.

2 Click the View you want to copy to the clipboard. To copy all available views as a single image, click **All.**

3 Open a document in another program that accepts images, such as a word processor or graphics program.

4 In the other program, click **Edit > Paste.**

The copied image appears.

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 127.

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.

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 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 eArray.chem.agilent.com. You must also provide your eArray user name and password in the Miscellaneous tab of the Preferences dialog box. See "Preferences – Miscellaneous Tab" on page 178.

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. "Scroll to Column" on page 195.

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, DNA Analytics 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 144.

2 Click **New.**

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

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

4 Click **OK.**

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

For example, this URL passes 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.**

3
Setting Up ChIP Interactive Analysis

This chapter gives instructions on how to set up the interactive analysis functions for ChIP experiments. These include the Preprocessing, Analysis and Reports tabs. To find instructions for importing, organizing, managing, exporting and displaying data and content, see Chapter 1 and Chapter 2.

Chapter 4 presents the descriptions of the tabs, commands, shortcut menus, window options, dialog box options and icons that you use to set up interactive analysis for ChIP data.

Chapter 5 explains how the analysis algorithms work.

For an overview of ChIP Interactive Analysis and a quick run-through of how you can use it to view and analyze ChIP data, from importing data to generating reports, see the *Agilent Genomic Workbench Quick Start Guide*.

If you are reading a hard copy of this guide and have not installed the software yet or do not know how to activate the license, please see the installation and licensing instructions online or in the *Agilent Genomic Workbench Quick Start Guide*.

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 88.

Figure 2 ChIP Interactive Analysis tabs

Table 1 lists the tasks in this chapter that you can use to change preprocessing and analysis options, visualize results, and create reports.

Table 1 Interactive analysis topics

Subject	See these topics
"Changing Preprocessing and Analysis options"	"To apply normalization" on page 67 "To configure the error model" on page 69 "To combine (fuse) arrays" on page 71 "To combine intra-array replicates" on page 73 "To combine inter-array replicates" on page 73 "To view QC metrics of arrays and set array QC status" on page 74 "To apply event detection" on page 75
"Visualizing results"	"To view results of analysis" on page 78 "To save a result" on page 79 "To restore a saved result" on page 80
"Creating Reports"	"To create a probe report" on page 81 "To create a gene report" on page 82 "To create a QC report" on page 83

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 2 on page 68 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 Order and 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 Order and Parameters” on page 203.

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, DNA Analytics uses the normalized feature intensities from the FE software instead.

Table 2 Normalization methods available in Interactive mode

Normalization method	Comments
Blanks subtraction normalization	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. The algorithm can use either medians or a one-step Tukey biweight function to estimate the central tendencies of blank probes.
Inter-array median normalization	This kind of normalization corrects for variations from one replicate array to another. DNA Analytics 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.
Intra-array (dye-bias) 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.

For a detailed description of the statistical algorithms DNA Analytics uses to normalize data, see "Normalization Algorithms" on page 224.

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 assesses 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 75. For a detailed discussion of the statistical algorithms used in the error model, see "Error Models" on page 232.

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 232.
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.

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 69.

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 define 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 24 and "To activate an experiment" on page 44.

2 To the new, active experiment, add the arrays that you want to fuse. See "To add arrays to an experiment" on page 25.

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  .
   A text box appears.

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

   d Click **Close.**

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

   The Array Set dialog box lists the arrays to be fused.

NOTE Double-check the values in the ArraySet Attribute column of the dialog box. DNA Analytics 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 re-centers the log ratio values, and ensures that the zero-point reflects the most-common-ploidy state.

- **Remove arrays from experiment after fuse** – To delete the initial un-fused 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, CGH).
- 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 73.
- 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 an array attributes file. See "To import array attributes" on page 20, or the *Sample Manager User Guide*.
- When you combine (fuse) arrays, the program applies the centralization algorithm to the individual arrays before it fuses them.

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 DNA Analytics uses to combine replicates, see "Intra-Array (Dye-Bias) Median Normalization" on page 228.

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

1 Click **Preprocessing.**

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

DNA Analytics combines the replicates in the selected arrays within the experiment.

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

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

To combine inter-array replicates

When you combine inter-array replicates, you select an array attribute. DNA Analytics combines replicates from the arrays with the same value for the selected attribute. For a discussion of the statistical model DNA Analytics uses to combine replicates, see "Inter-Array Median Normalization" on page 227.

To combine inter-array 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.**

DNA Analytics 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 measuring log ratios.

To view 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 only available for Agilent arrays, are either passed through from the FE program, or are calculated by DNA Analytics 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 Do one of the following:

- To view the QC metrics of Agilent arrays in the active experiment – In the Preprocessing command ribbon, click **QC Metrics.**
- To view 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 view 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 Poor (pink), Good (turquoise), or Excellent (yellow). This dialog box also offers several other options for visualizing the QC metrics of the array(s). See "QC Metrics Table" on page 185.

After you view the QC metrics of an array, you can set the QC status of the array.

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.**

To apply event detection

The ChIP application uses the Whitehead Per-Array Neighbourhood Model to make binding calls. This model 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 3. Two neighbor probes (blue) flank a central probe (red).

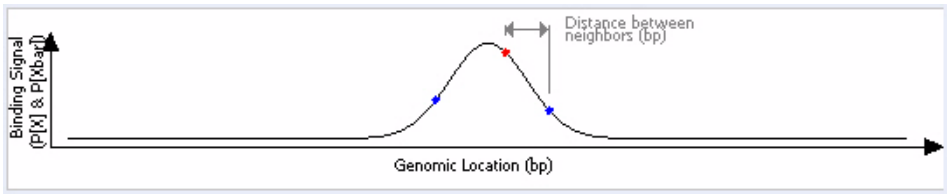


Figure 3 Central probe flanked by two neighbor probes

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 Algorithm" on page 237.

To apply event detection to the current experiment

1. Configure the error model. See "To configure the error model" on page 69.
2. Click **Analysis.**
3. Under **Event Detection,** click the **Edit** button  .

   The Whitehead Per-Array Neighbourhood Model Parameter Settings dialog box appears. "Whitehead Per-Array Neighbourhood Model Parameter Settings" on page 210.
4. 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 3. Table 3 describes the parameters.
5. Click **OK.**
6. 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 3 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.
Central probe has P(X) <	• The central probe is the red probe in Figure 3. • 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 3 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 • The default cut-off value for P(X) is 0.005 • To make detection more stringent, decrease the value.

Visualizing 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 Genome, Chromosome, and Gene Views. Figure 4 shows an example of a result displayed in Gene View. To customize the way the results appear, see the *Data Viewing Guide*.

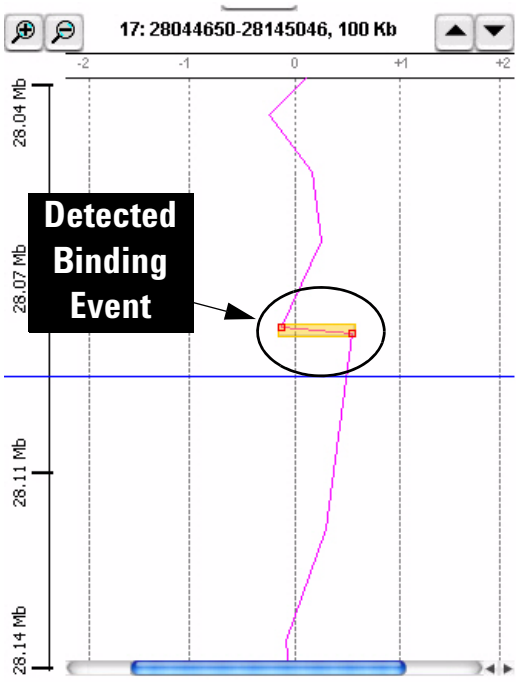


Figure 4 Gene view, showing detected binding event

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 activated the experiment or selected the arrays of interest. See "To activate an experiment" on page 44.
- You may not be looking in the right region of Chromosome View.

See the *Data Viewing Guide* 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 current result of the active experiment. You can run many different analyses in the same experiment, and save each one. Later, you can restore any of your saved results.

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

1 In the **Experiment** pane of the Navigator, right-click the 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, activate the experiment that contains the result that you want to see. See "To activate an experiment" on page 44.

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 213 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 name dialog box appears. See "Select report name" on page 200.

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 215 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 163.

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 name dialog box appears. See "Select report name" on page 200.

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 view 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 details about the contents of the report, see "QC Report format" on page 216.

1 Click **Reports > QC Report.**

The QC Report Settings Parameter Settings dialog box appears.

2 Click **Browse.**

The Select report name dialog box appears. See "Select report name" on page 200.

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.

4
ChIP Interactive Analysis Reference

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

For specific instructions on how to use Genomic Workbench to manage data and accomplish related tasks, see Chapter 1, “Importing, Managing, and Exporting ChIP Data and Other Content.” To learn how to visualize data and content, see Chapter 2, “Visualizing ChIP Data and Other Content”.

This chapter also describes the tabs, commands, shortcut menus, and dialog boxes specific to ChIP data analysis. For specific instructions on analyzing and reporting on ChIP data, see Chapter 3, “Setting Up ChIP Interactive Analysis.”

To learn how to set up and run analysis workflows with ChIP analysis methods, see the *Workflow User Guide*.

To understand how all the guides work together to help you use Agilent Genomic Workbench, see the *Agilent Genomic Workbench 5.0 Quick Start Guide*.

Genomic Workbench ChIP Application Main Window

The main window of the ChIP Application contains the eight main components illustrated in Figure 1.

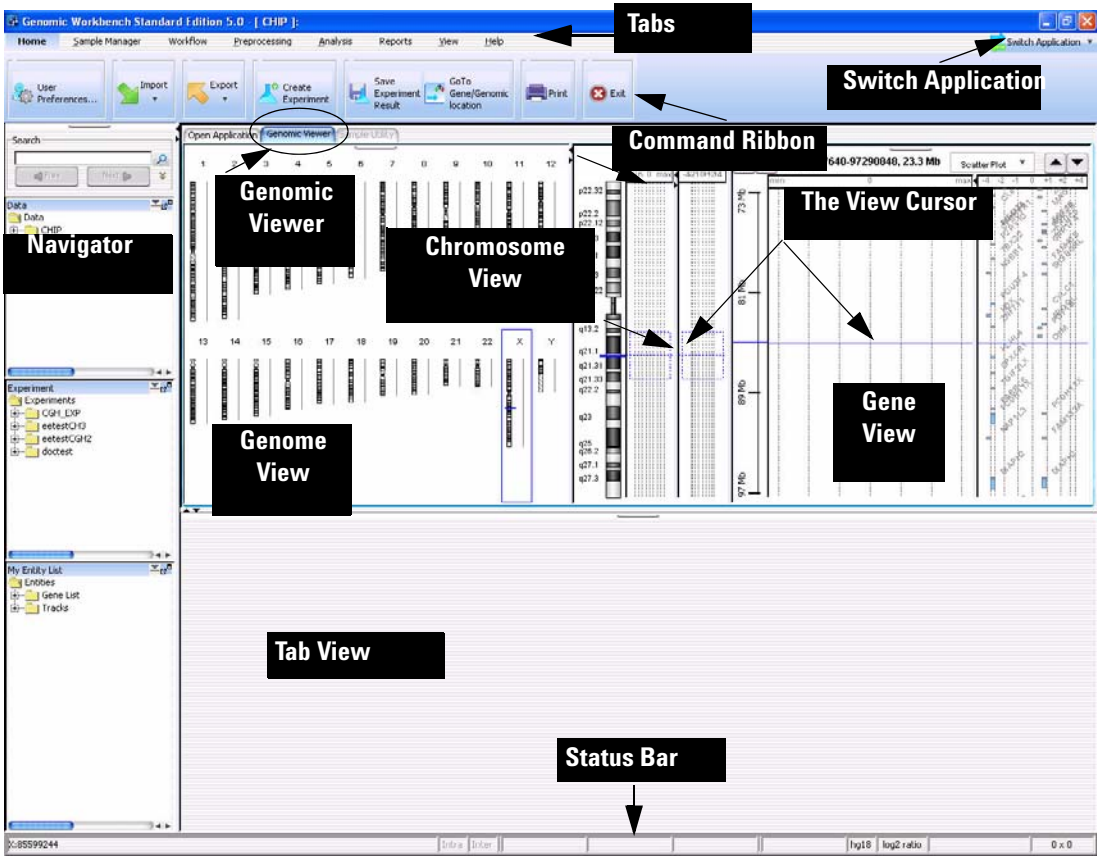


Figure 1 Genomic Workbench Standard Edition – ChIP Application main window

Switch Application Menu

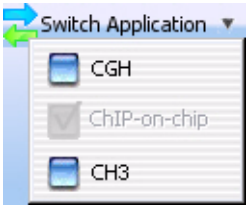


Figure 2 Switch Application menu

The Switch Application menu allows you to switch to the other data display and analysis applications in DNA Analytics. Mark the desired application.

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.

Command Ribbons

When you click a tab, groups of commands or single commands appear. The entire string of commands is called a command ribbon. This section describes the tab commands used to import, manage, export, and display data/results and those specific to ChIP interactive analysis.

Figure 3 ChIP Application tabs

Home command ribbon

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

Figure 4 Command ribbon in the Home tab of Genomic Workbench

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

Tab	Description
View	Opens a dialog box that lets you change in what form the data will appear in Genomic Viewer. See "Preferences – View Tab" on page 181.
Tracks	Opens a dialog box that lets you manage which tracks to display in Genomic Viewer and how they appear. See "Preferences – Tracks Tab" on page 179.

Tab	Description
Miscellaneous	Opens a dialog box where you can select a new location for your data files and set up access to the eArray web site. See "Preferences – Miscellaneous Tab" on page 178.
License	Opens a dialog box where you can add a DNA Analytics application license, if you choose to purchase one after using the unlicensed version. Preferences – License 176.

Import/Export

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 166 and "To import Agilent FE or GenePix/Axon data files" on page 14. • **Axon File** – Opens the Import Axon Files dialog box, where you can select Axon (*.gpr) files for import. See "Import" on page 166 and "To import Agilent FE or GenePix/Axon data files" on page 14.
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 166 and "To import Agilent GEML design files" on page 16. • **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 166 and "To import GenePix/Axon design files" on page 17.
Genome Build	Opens the Import Genome Build dialog box, where you can import Agilent-provided genome build files. See "Import Genome Build" on page 170 and "To import a genome build" on page 18.
Probe Upload	Allows you to import a file of probe sequences and annotation. For details, see the *eArray$_{XD}$ User Guide.* You see this only if you have installed Genomic Workbench Enterprise Edition.

Option	Description
Track	Opens the Import Track dialog box, where you can select a BED format track file for import, and create a display name for the track. See "Import Track" on page 171 and "To import tracks" on page 19.
Array Attributes	Opens the Import Array Attributes dialog box, where you can select a .txt file for import. See "Import" on page 166 and "To import array attributes" on page 20.
Experiments	Opens the Import Experiments dialog box, where you can select a ZIP format experiment file for import. See "Import" on page 166 and "To import an experiment file" on page 21.

Export Opens a menu that allows you to 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 158 and "To export experiments" on page 39.
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 159 and "To export tracks" on page 40.
Array Attributes	Opens the Export Array Attributes dialog box, where you can select arrays and their attributes for export. See "Export Array Attributes – Array" on page 154.

Create Experiment Opens the Create Experiment dialog box, where you can create a new, empty experiment and populate it with data. See "Create Experiment" on page 141 and "To create a new experiment" on page 24.

Save Experiment Result Saves ChIP results after analysis.

Go to Gene/Genomic Location Moves the cursor to the location in Chromosome and Gene Views that you specify. See "Go To Gene/Genomic Location" on page 165.

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.
- **QC Metric** – View quality control metrics, and set their status.

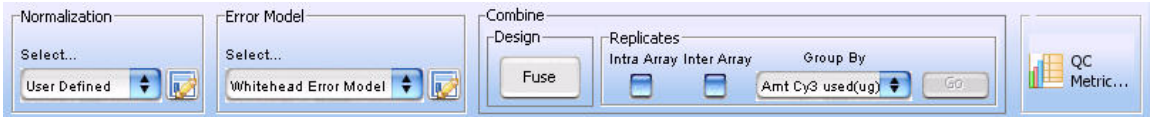


Figure 5 ChIP Preprocessing command ribbon

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 non-specific 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 67.

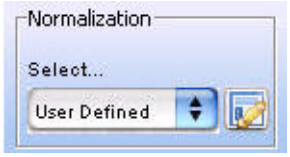


Figure 6 Normalization group box

Select Allows you to select one of two choices for the normalization:

User Defined – The program applies user-defined 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 67. Unless you select this option, the program uses the raw, unprocessed feature intensities.

(Edit button) Opens the Setting Normalization Order and 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 Order and Parameters" on page 203.

Error Model

The error model assesses 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 75. For a detailed discussion of the statistical algorithms used for error modeling and event detection, see "Error Models" on page 232.

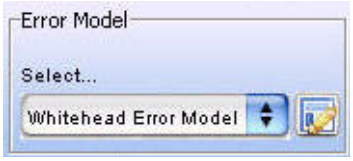


Figure 7 Error Model group box

Select Allows you to 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 232. 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 209.

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 71 and "To combine intra-array replicates" on page 73. For a discussion of the replicate support statistical algorithm, see "Normalization Algorithms" on page 224.

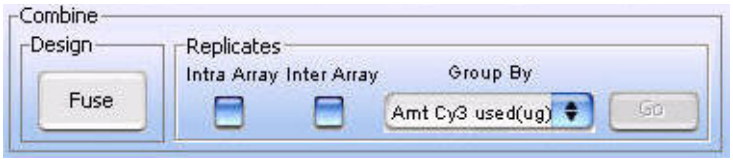


Figure 8 Combine group box

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 135, and "To combine (fuse) arrays" on page 71.

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 73.

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 defined 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

The dialog box that opens when you click this command allows you to open the QC Metrics Table, where you can view the calculated QC metrics for the arrays in the active experiment. You use QC metrics to evaluate the quality of microarray data. They are available only for Agilent microarrays. See "QC Metrics Table" on page 185.

Figure 9 QC Metrics command

You use other commands to view the QC Metrics of individual arrays or designs. See "Data pane – actions and shortcut menus" on page 104.

Analysis command

Event Detection

The program uses the Whitehead Per-Array Neighbourhood Model to make binding calls. See "To apply event detection" on page 75, and "Peak Detection Algorithm" on page 237. The settings in Event Detection allow you to customize the model.

Figure 10 Event Detection group box

Apply Mark this check box to apply the selected algorithm (Whitehead Per-Array Neighbourhood Model) to the selected array data in the active experiment.

Select Currently, the Whitehead Per-Array Neighbourhood Model is the only event detection model available for ChIP experiments.

Opens the Whitehead Per-Array Neighbourhood Model Parameter Settings dialog box, where you can customize the model. See "Whitehead Per-Array Neighbourhood Model Parameter Settings" on page 210, and "To apply event detection" on page 75.

Reports command ribbon

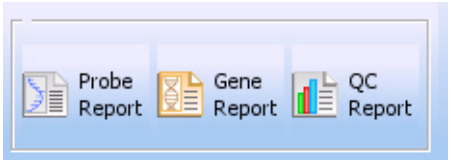


Figure 11 ChIP Reports command ribbon

The Reports command ribbon allows you to 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 allows you to further view, analyze, and organize the result.

Probe Report Opens the Probe Report Settings Parameter Settings dialog box, where you can define a location and a name for the probe report file. "Probe Report Settings Parameter Settings" on page 183.

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 213 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 define a location and a name for the gene report file. See "Gene Report Settings Parameter Settings" on page 163.

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 215 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.

QC Report Opens the QC Report Setting Parameter Settings dialog box, where you can define a location and a name for the QC Report file. See "QC Report Settings Parameter Settings" on page 194.

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 vs. WCE
- Scatter plot of enrichment ratio vs. 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 details about the contents of the report, see "QC Report format" on page 216.

View command ribbon

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

Figure 12 Command Ribbon of the View tab

Orientation Select one of these options:

- **Horizontal** – Stacks Genome, Chromosome, and Gene views horizontally. Chromosomes and chromosomal locations appear in left to right orientation.
- **Vertical** – Stacks Genome, Chromosome, and Gene views vertically. Chromosomes and chromosomal locations appear in top to bottom orientation.

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 41.

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.

Signal Intensity Mark **View in Table** to see the red and green raw signal intensities of the log ratio data in 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.

Scatter Plot See "Scatter Plot" on page 123.

Help command ribbon

The Help command ribbon provides access to the Agilent Genomic Workbench Quick Start Guide and all the other user guides.

Figure 13 Help command ribbon for Genomic Workbench Standard Edition

Table 1 Table of Agilent Genomic Workbench Help

Help Command	Action
Help	Opens the DNA Analytics application user guide for which you have the associated license. ChIP Interactive Analysis – (ChIP license required) Shows how to set up all preprocessing, analysis, and reporting options for analyzing ChIP data interactively. Includes details on the algorithms used.
Quick Start	Opens the *Genomic Workbench Quick Start Guide* in Adobe Reader. This guide provides brief instructions on how to install and start the program, and how to use the basic features of the program to create custom microarray designs, and to analyze microarray data.
Data View	Shows you how to import, manage, export and display log ratio data from Agilent and other sources.
Workflow	Gives instructions on how to set up a workflow for automated feature extraction and/or analysis. Also shows you how to set up the CGH and ChIP analysis methods to be used in a workflow.
Sample Manager	Shows you how to assign identification and attribute information to image files, imported feature extraction (FE) files, or UDF files.
eArray	(Genomic Workbench Enterprise Edition only) Gives instructions on how to use $eArray_{XD}$ to design custom microarrays on your desktop. **Note:** For instructions on how to use the eArray Web site, see the online help on earray.chem.agilent.com.
About	Opens a message with details about the version number and copyright of the program.

Navigator

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

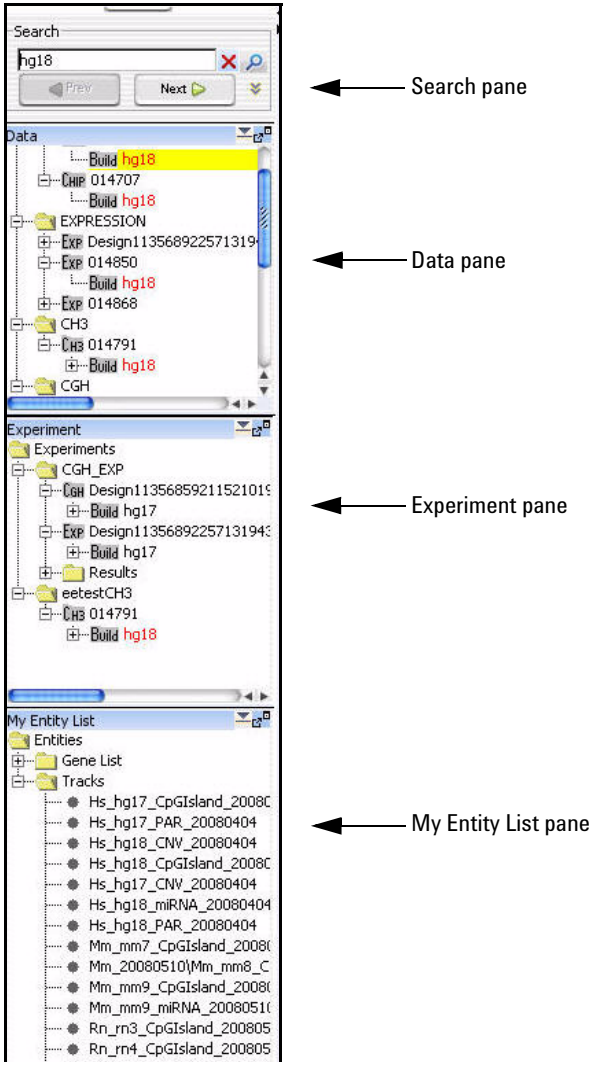


Figure 14 Navigator – note the four panes within the Navigator

The Navigator (Figure 14) catalogs the array data, experiments, and other content stored in Genomic Workbench. It contains four panes:

Pane	Comments
Search	Allows you to find all occurrences of a specific search term in the Data, Experiment, and/or My Entity List panes.
Data	Contains microarray data files, organized by application type and design, and then by genome build.
Experiment	Contains 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 allows you to 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 30. It also contains several buttons that you can use to detach, hide, show or resize the Navigator.

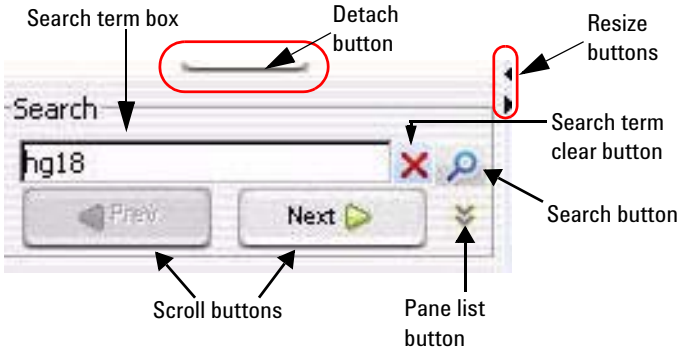


Figure 15 Navigator – Search pane

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

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

Search term box 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.

Pane list Allows you to restrict 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.

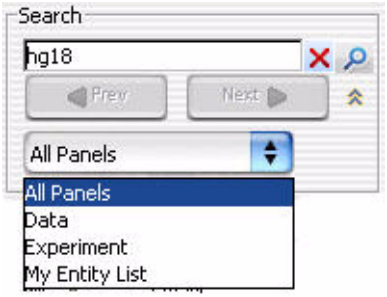


Figure 16 Open Pane List

(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.

Scroll buttons (Available only after a search) Allows you to 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
	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.

Item	Comments
	A single array data file.
	Data created from a multi-pack array.
red text	An item that matches the search term in a search.
	(Dock out button) Detaches the Data pane from the Navigator, and opens it in its own, 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

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

Data Folder

- Right-click the **Data** folder to open a shortcut menu with an Import option. When you select this option, a menu appears with these options for file import:

Option	Description
Design File	Opens the Import Design Files dialog box, where you can select an Agilent GEML-based (*,xml) file for import. See "Import" on page 166 and "To import Agilent GEML design files" on page 16.
Axon Design File	Opens the Import Axon Files dialog box, where you can select GenePix/Axon design (*.gal) files for import. See "Import" on page 166 and "To import GenePix/Axon design files" on page 17.
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 166 and "To import Agilent FE or GenePix/Axon data files" on page 14.
Axon File	Opens the Import Axon Files dialog box, where you can select Axon (*.gpr) files for import. See "Import" on page 166 and "To import Agilent FE or GenePix/Axon data files" on page 14.

Design Folder

- Right-click the name of design folder to open a shortcut menu with these options:

Option	Description
Update from eArray	(Available only for Agilent microarrays) 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 31.
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 open a shortcut menu with these options:

Option	Description
Show Properties	Opens the Design Properties dialog box. See "Design Properties" on page 145.
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 185, and "To view QC metrics of arrays and set array QC status" on page 74.
Delete	Opens a Confirm dialog box. If you click **Yes,** the program permanently deletes all of the arrays in this genome build folder.

Individual Arrays

- Right-click the name of an array to open a shortcut menu with these options:

Option	Description
Show Properties	Opens the Microarray Properties dialog box. See"Microarray Properties - Attribute Tab" on page 173 and "To edit the attribute values of a specific array" on page 28.
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 185, and "To view QC metrics of arrays and set array QC status" on page 74.
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. Control-click the additional arrays to select them. You can also select a contiguous block of arrays—click the first array in the block, then shift-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.

Item	Comments
	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 its own, 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

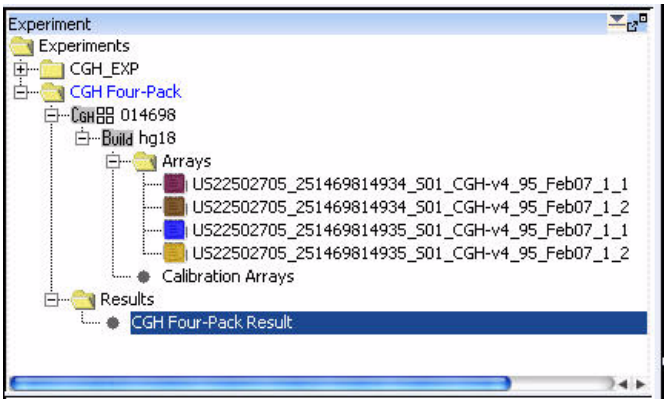


Figure 17 Expanded Experiment Pane

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 open a shortcut menu with these options:

Option	Description
New Experiment	Opens the Create Experiment dialog box (see "Create Experiment" on page 141), where you can name the new experiment, and open another dialog box that allows you to populate the experiment with microarrays. See "To create a new experiment" on page 24.
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 158 and "To export experiments" on page 39.

Specific Experiment Folder

- In the **Experiments** folder, right-click the name of a specific experiment folder to open a shortcut menu with 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 79. • **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 allows you to overwrite the current result, create a new result, or discard any changes. See "Confirm" on page 138. • 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 151.
Edit Attributes	Opens the Experiment Attributes dialog box, where you can change the values for the attributes assigned to the arrays in the experiment. See "Experiment Attributes" on page 150. 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 158, and "To export experiments" on page 39.
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 – Array" on page 154 and "To export array attributes" on page 38.
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 185, and "To view QC metrics of arrays and set array QC status" on page 74.
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 148.
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 149.

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.

Design Folder

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

Genome Build Folder

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

Option	Description
Set for Calibration	Designates all arrays associated with the specific genome build and design as calibration arrays. The program transfers their names to the Calibration Arrays folder of the experiment, and moves the array data in Tab view for those arrays to the Calibration Arrays tab.
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 185, and "To view QC metrics of arrays and set array QC status" on page 74.

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 open a shortcut menu with 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.
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.

Option	Description
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 - Attribute Tab" on page 173. For array files from the Agilent Feature Extraction program, you can also view the headers and feature data from the file. See "To edit the attribute values of a specific array" on page 28.
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 185, and "To view QC metrics of arrays and set array QC status" on page 74.
Edit Array Color	Opens the Select Color dialog box, where you can select a display color for the array. See "Scroll to Column" on page 195.
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 149.

Results Folder

- In the **Results** folder of an experiment, right-click the name of a saved result to open a shortcut menu with 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 80. 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 184.

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 its own, 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

- Right-click the **Gene List** folder to open a shortcut menu with an **Import Gene List** option. This command opens an Import dialog box that you can use to import a gene list into the program. See "To import a gene list" on page 21 and "Import" on page 166.
- 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 open a shortcut menu with 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 Report Settings Parameter Settings" on page 163 and "To show gene lists in Gene View" on page 53.
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 40.

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 132 and "To add one gene list to another" on page 34.
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 53.
Show only	(Available only if all genes appear in Gene View, or if the gene list is not selected) Restricts the genes in Gene View to those on the gene list. No other genes appear. The program displays the genes in their display color. See "To show gene lists in Gene View" on page 53.
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 53.

- Right-click the Tracks folder to open a shortcut menu with 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 171 and "To import tracks" on page 19.
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 159 and "To export tracks" on page 40.
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 137 and "To combine tracks" on page 35.

- Right-click the name of a track to open a shortcut menu with 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. See "Gene List" on page 162.
Show in Report	Mark this option to display the track in the reports.
Genomic Boundaries	Click to use the genome track to define only the regions that aberration detection algorithms will run. You can choose to do this for only one track.
Show in UCSC	Opens the UCSC Genome Browser in your Web browser and uploads the track. You can then view the track. See
View Details	Opens a table listing all the chromosome locations defined in the track. See "Track" on page 207.
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 DNA Analytics applications. It includes three graphical views: Genome, Chromosome and Gene Views. It also contains Tab View and the Cursor.

Genome View

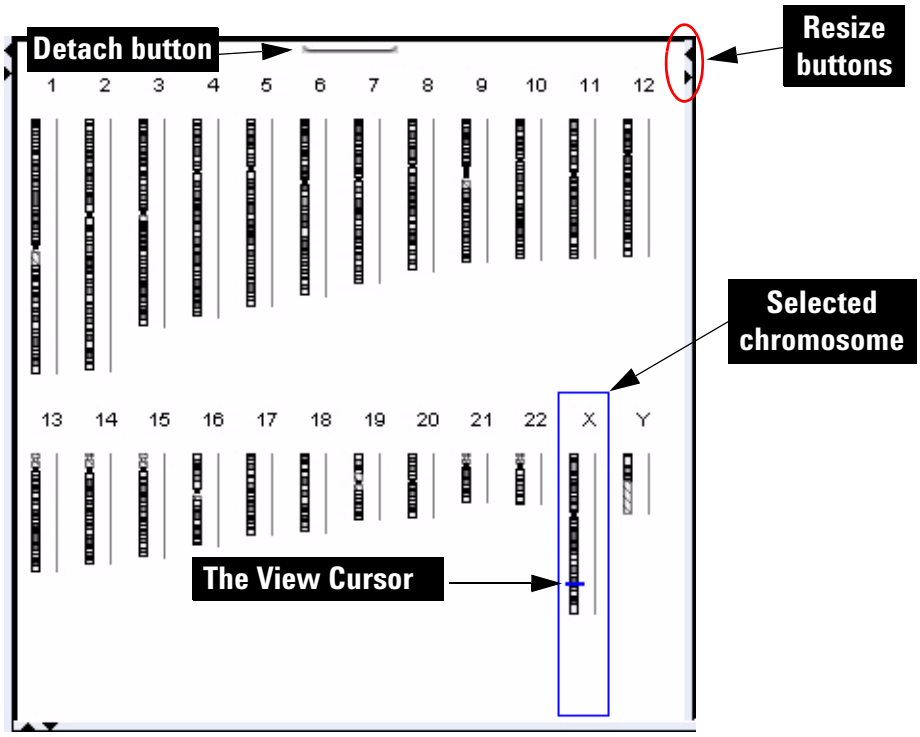


Figure 18 Genome View, with human chromosomes. The X chromosome is selected.

Genome View shows pictures of each of the distinct types of chromosomes in the selected genome. A blue box encloses the 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 126.
- On the selected chromosome, click anywhere to reposition the cursor. See "The View Cursor" on page 126. This also repositions the cursor in Chromosome, Gene, and Tab Views.
- Right-click anywhere within Genome View to open a shortcut menu with a Preferences option. If you click **Preferences,** the Preferences dialog box opens, where you can set user preferences on four separate tabs. See "Preferences – License" on page 176 and the topics that follow.
- Click the **Detach** button (located at the top center of the pane) to remove Genome View from the main window and open in its own separate window. To reattach the view, click its **Close** button  . See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 and "Maximizing and reattaching panes to the Genomic Workbench main window" on page 20 of the *Quick Start Guide*.
- 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. See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 of the *Quick Start Guide*.

Chromosome View

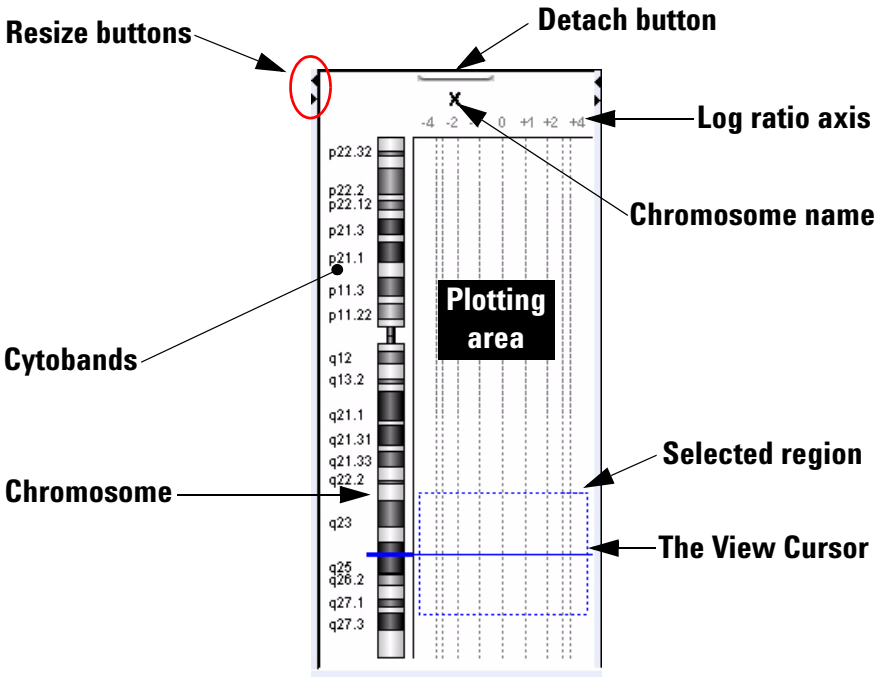


Figure 19 Chromosome View, human X chromosome shown

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 126.
- 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 126.
- Right-click anywhere in the plotting area of Chromosome View to open a shortcut menu with a Preferences option. If you click **Preferences,** the Preferences dialog box opens, where you can set user preferences on four separate tabs. See "Preferences – License" on page 176 and the topics that follow.
- Click the **Detach** button (located at the top center of the pane) to remove Chromosome View from the main window and open in its own separate window. To reattach the view, click its **Close** button  . See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 and "Maximizing and reattaching panes to the Genomic Workbench main window" on page 20 of the *Quick Start Guide*.
- 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. See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 of the *Quick Start Guide*.

Gene View

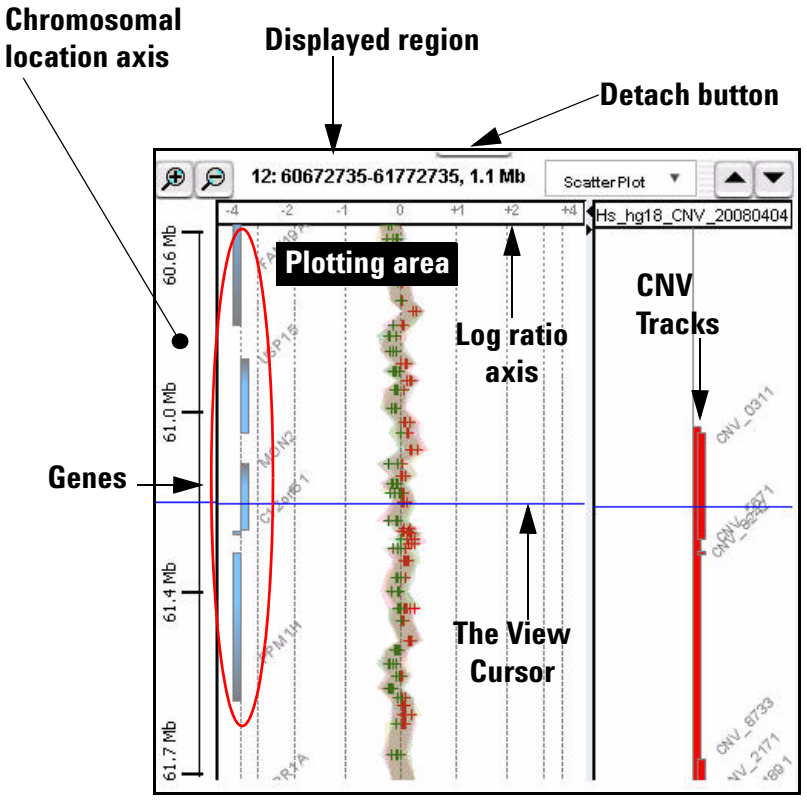


Figure 20 Gene View, with log ratio data from an experiment and CNV tracks.

Gene View shows a more detailed view of the chromosomal region you select in Chromosome View. See "Chromosome View" on page 120.

- 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 restrict the genes that appear to those in the list. See "To change the appearance of genes in Gene View" on page 55, and "To show gene lists in Gene View" on page 53.

- Log ratio data from selected arrays in the active experiment appear as a scatter plot. Points appear in up to three different colors. You can also customize the scatter plot. See "To change scatter plot appearance" on page 49.
- The location of the cursor matches the location of the cursors in other views. See "The View Cursor" on page 126.
- 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 56.

Scatter Plot

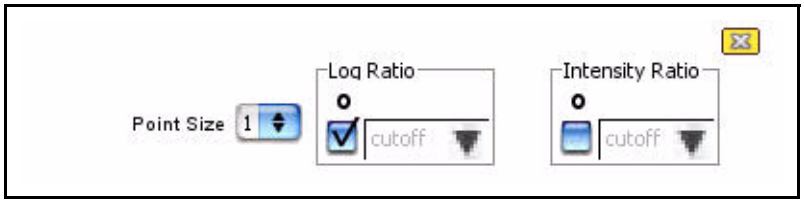


Figure 21 Scatter Plot command group in ChIP Gene View

You access the scatter plot command group through Gene View or through the View tab. The commands differ depending on the DNA Analytics application you are working with. All the scatter plot command groups contain the commands for log ratio data. The ChIP application also contains these same choices for the Intensity Ratio. Scatter plots appear in the Chromosome and Gene Views but only if they have been activated in the Preferences dialog box.

Mark this black check box to display, or clear it to hide, points that reflect insignificant changes. You select the cutoff for significance in **Cutoff.** These points appear in black.

Point size Sets the size (in pixels) of the points in the scatter plots.

In addition to options for log ratio plotting, the command group includes options for raw intensity plotting.

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 "Resizing and detaching panes from the Genomic Workbench main window" on page 19 of the *Quick Start Guide*.

(**Detach** button) Removes Gene View from the main window, and opens it in its own separate window. See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 of the *Quick Start Guide*.

Gene View shortcut menu and other actions

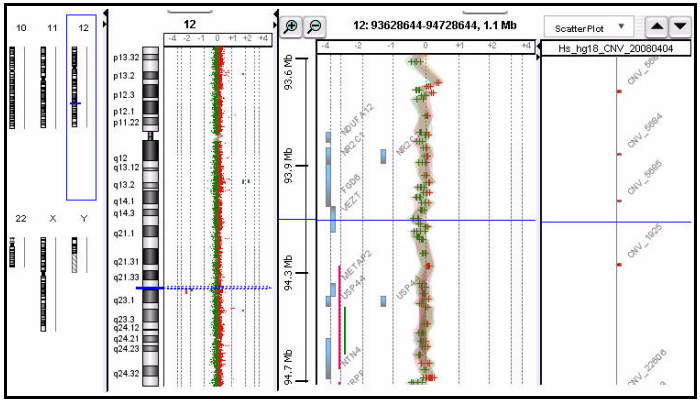
- Click anywhere in the plotting area of Gene View to move the cursor to that location. See "The View Cursor" on page 126.
- Drag an inside border of Gene View to resize the view.
- Right-click anywhere in the plotting area of Gene View to open a shortcut menu with these options:

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 142 and "To create a gene list" on page 33.
Show in UCSC	Opens the UCSC (University of California at Santa Cruz) Genome Browser in your Web browser and uploads the track. You can then view the track.
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 "Scroll to Column" on page 195 and "To search Agilent eArray for probe information" on page 61.
Preferences	Opens the Preferences dialog box, where you can set user preferences on four separate tabs. See "Preferences – License" on page 176 and the related pages that follow.

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.



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

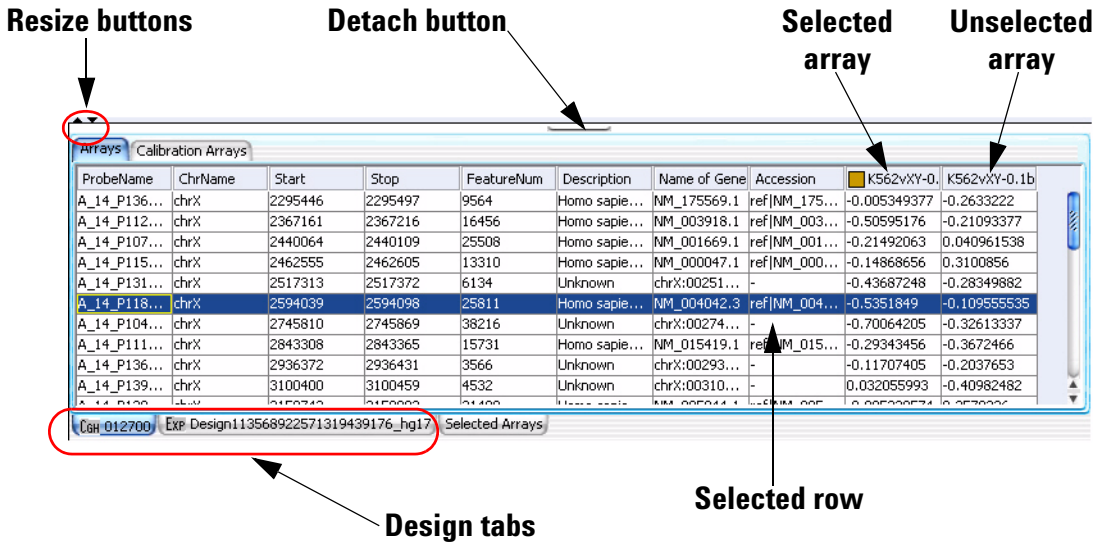


Figure 22 Tab View

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.

Tab View tabs and buttons

You can see the following tabs and buttons in Tab View. See Figure 22 for a diagram that identifies some of these elements.

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

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

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

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

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

(**Resize** buttons) The button that points away from Tab View expands the view. The other button restores the view to its original size. See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 of the *Quick Start Guide*.

(**Detach** button) Removes Tab View from the main window, and opens it in its own separate window. See "Resizing and detaching panes from the Genomic Workbench main window" on page 19 of the *Quick Start Guide*.

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 open a shortcut menu with 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 32.
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 148 and "To change the display color of an array" on page 47.
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 149 and "To change the order of arrays in an experiment" on page 26.
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 195.

- Right-click a *heading of a column other than an array data column* to open a shortcut menu with a Scroll To Column option. If you click this option, the Scroll to Column dialog box appears, where you can select a column in the current tab. The program then scrolls the data table in the tab so you can see the column.
- Click an *entry in a data table* to select the row in which it appears. This also moves the cursor to the location of the data point corresponding to the selected row.
- Right-click a *data table entry* to open a shortcut menu with 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 160.
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)	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 62.
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 144 and "To create a custom Web search link" on page 62.
(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

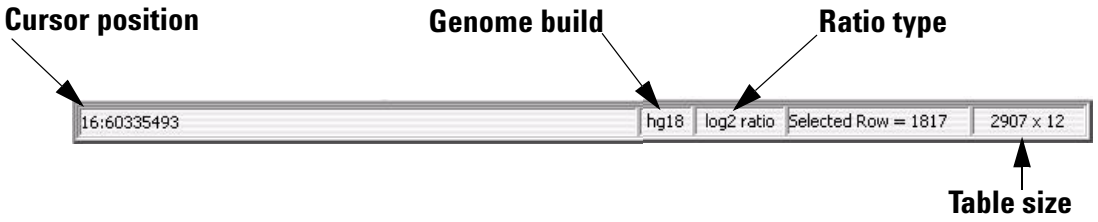


Figure 23 Status bar

The Status Bar (Figure 23) displays information related to the currently displayed data.

Cursor position The chromosomal location of the cursor. See "The View Cursor" on page 126.

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 columns in the currently 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 Application. They are listed in alphabetical order by title. Many dialog boxes in Interactive mode can also appear when you use the Genomic Viewer.

Add Gene List <name> to

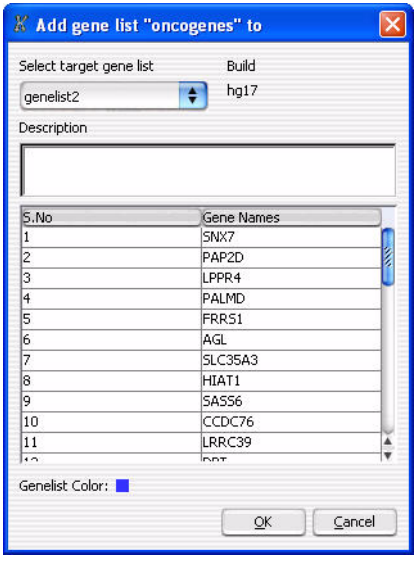


Figure 24 Add Gene List <name> to

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

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

Select target gene list The gene list to which genes will be 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.

Gene List 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

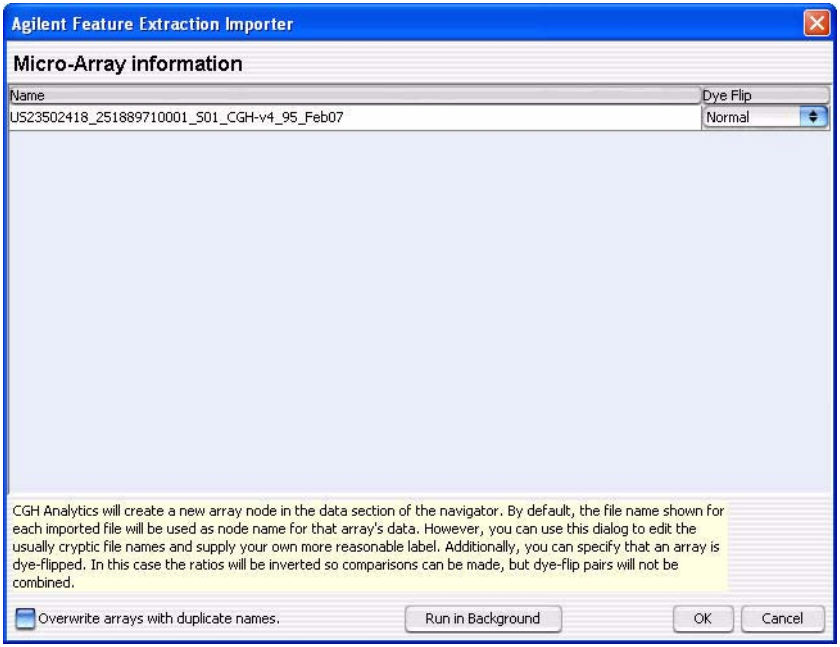


Figure 25 Agilent Feature Extraction Importer

Purpose: Allows you to edit the name of the FE data file you intend to import and to specify 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 Allows you to 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 ration (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 allows you to 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

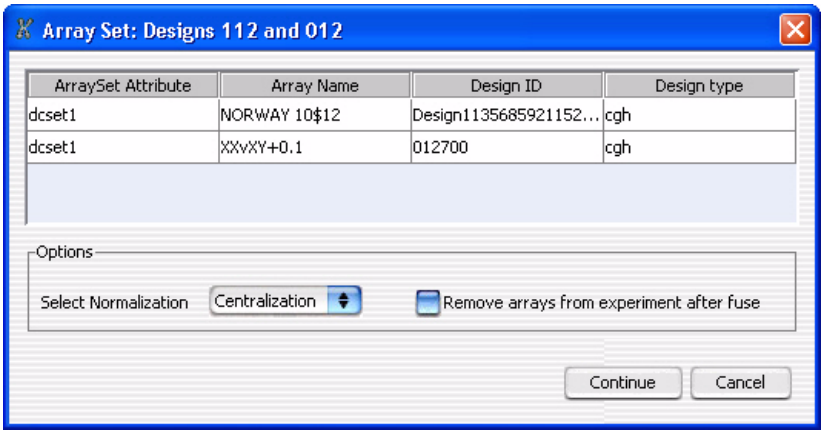


Figure 26 Array Set dialog box.

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 71.

Table Lists 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. "Normalization Algorithms" on page 224.

Remove arrays from experiment after fuse Deletes the original un-fused arrays after creating fused arrays. This can help reduce 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.

Choose Gene List Color

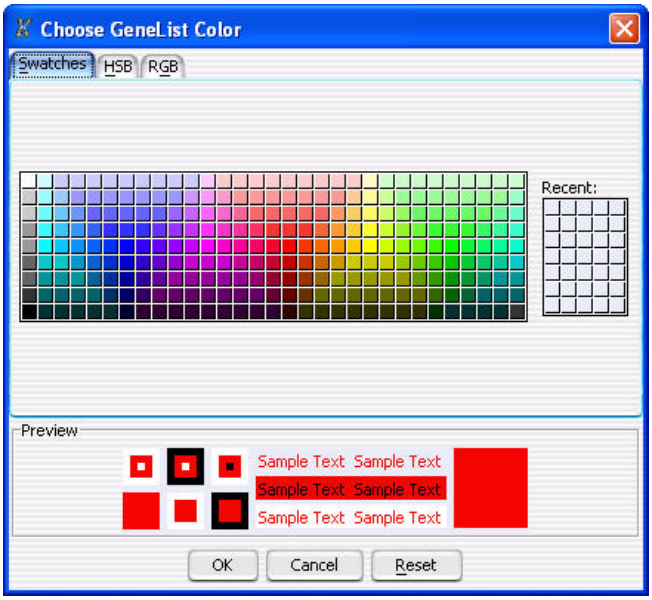


Figure 27 Choose GeneList Color dialog box

Purpose: Allows you to select a display color for a specific gene list.

To open: Right-click Gene View, click **Create Gene List > Change**, or in the Navigator, right-click a specific gene list, click **View in Table > Color**.

Swatches Tab Choose color based on color samples (Swatches)

HSB Tab Choose colors based on an HSB schema (Hue, Saturation, and Brightness or Value). See "Select Color (Edit Array Color) — HSB Tab" on page 198.

RGB Tab Choose colors based on an RGB schema (Red-Green-Blue). See "Select Color (Edit Array Color) — RGB Tab" on page 199.

Recent Choose a recent color selection.

Reset Click to return HSB or RGB values back to default values.

Combine Tracks

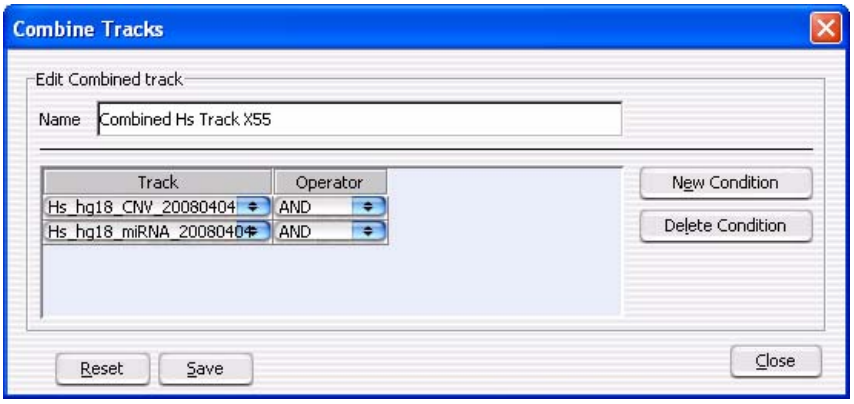


Figure 28 Combine Tracks dialog box

Purpose: Allows you to create a combined track that contains elements of two or more source tracks, associated by logical criteria. See "To combine tracks" on page 35.

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 non-redundant 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.

Confirm

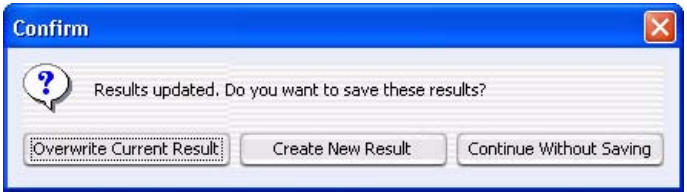


Figure 29 Confirm dialog box

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 79.

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

Confirm overwrite

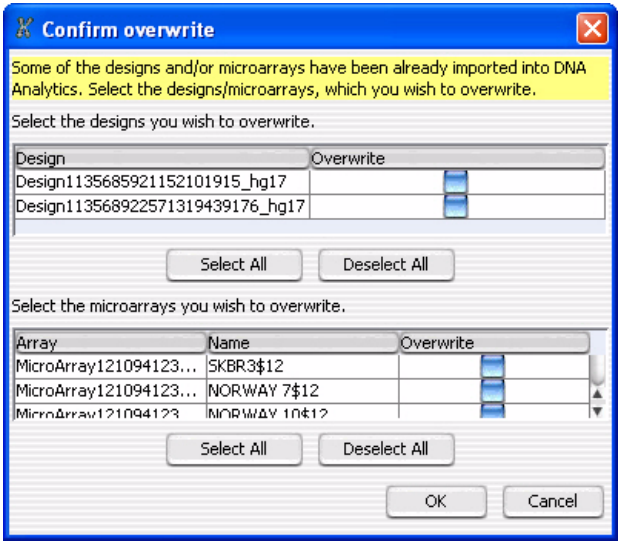


Figure 30 Confirm overwrite dialog box

Purpose: When you import an experiment, it can contain designs and/or arrays that have the same names as those already available in Genomic Workbench. This dialog box allows you to 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 Genomic Workbench. See "To import an experiment file" on page 21.

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 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

Name The names of the arrays in the imported file that have the same names as arrays that are already available in 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 168.

Create Experiment

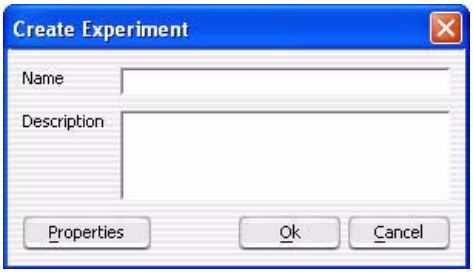


Figure 31 Create Experiment dialog box

Purpose: To create an organizational unit in the Experiment pane of the Navigator to link to array data for viewing and analysis and to create the links to the data (populate) the experiment (in Experiment Properties dialog box)

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 Briefly, describe your experiment with information that will help you identify it.

Properties Click to access the Experiment Properties dialog box where you can select microarrays to populate your new experiment. See "Experiment Properties" on page 151.

NOTE If you do not populate your new experiment with arrays in the Experiment Properties dialog box, the program creates an empty experiment.

Create Gene List

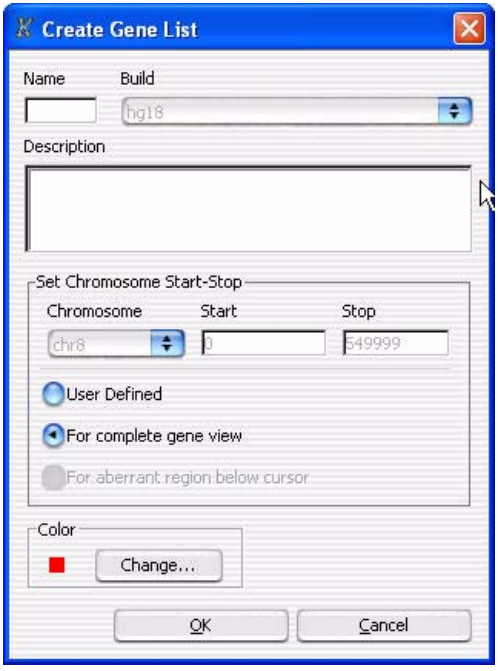


Figure 32 Create Gene List

Purpose: Creates a list that you can use to highlight or restrict the genes data 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 manually define the region from which the genes in Gene View will be selected. The Chromosome, Start, and Stop options become available. This option overrides any previously defined region.

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

For aberrant region below cursor Selects the genes that appear in the aberrant region just below where the cursor sits in Gene View.

Chromosome (Available if you select User Defined) Allows you to select the chromosome associated with the genomic region defined for the gene list.

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

Stop (Available if you select User Defined) Allows you to set the end of the genomic region defined 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

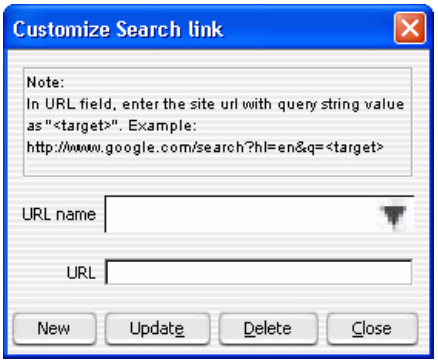


Figure 33 Customize Search Link dialog box

Purpose: Allows you to 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 62.

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 view the properties of a specific design" on page 31.

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

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

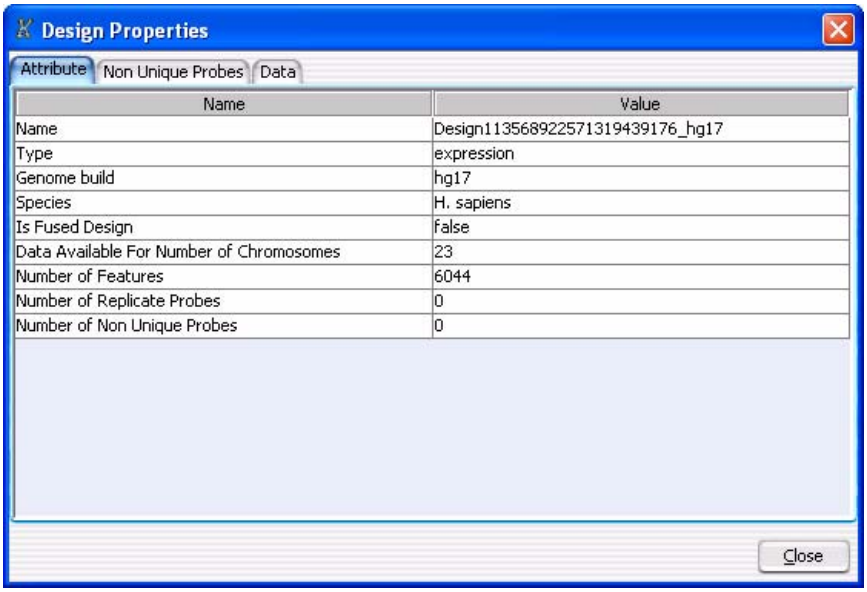


Figure 34 Design Properties dialog box – Attribute tab

Non-Unique Probes tab

Lists the non-unique probes in the design. Non-unique probes bind to more than one location in a target genome.

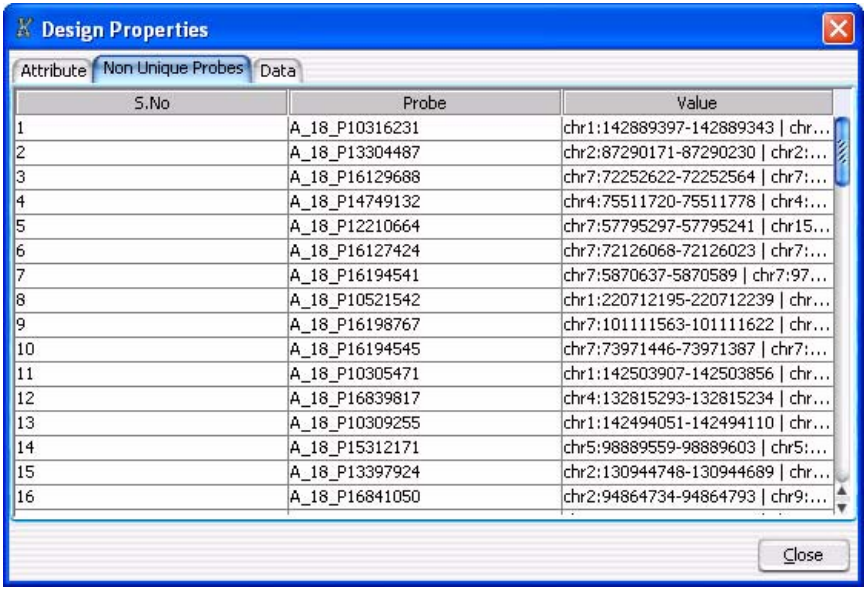


Figure 35 Design Properties dialog box – Non-Unique Probes tab

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

Probe The name of each non-unique probe.

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

Data tab

Lists 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.

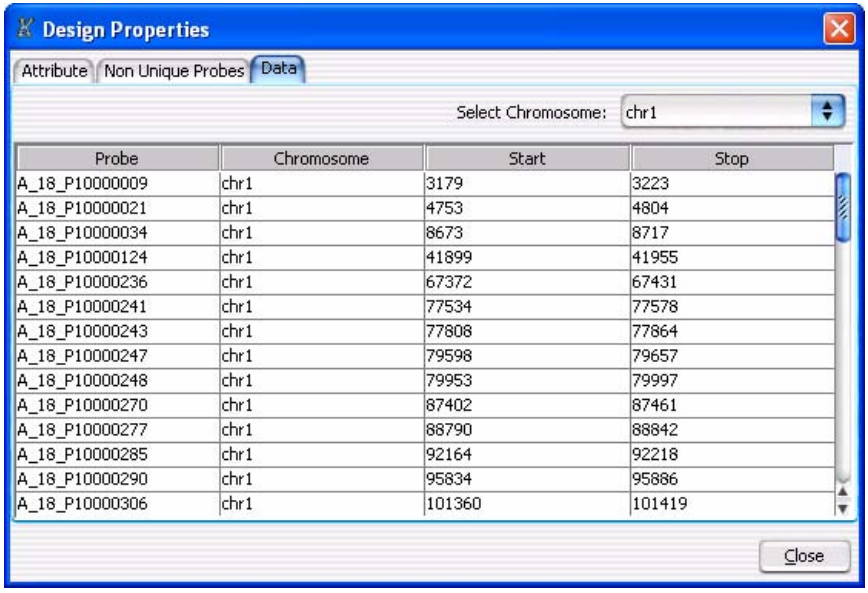


Figure 36 Design Properties dialog box – Data tab

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

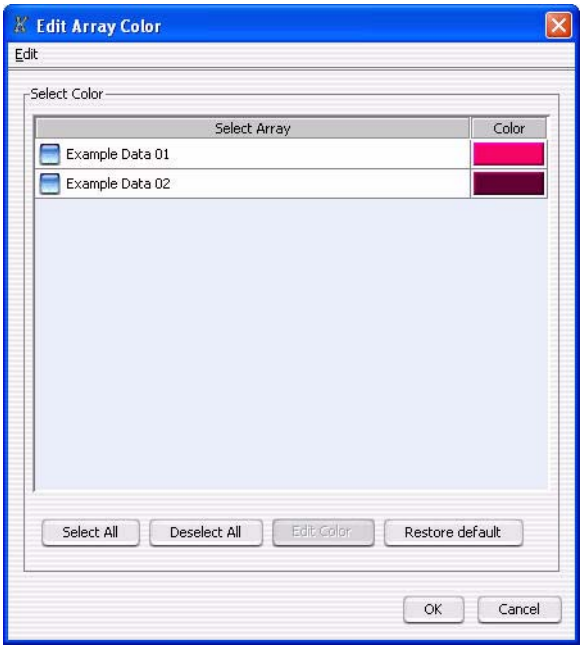


Figure 37 Edit Array Color dialog box

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

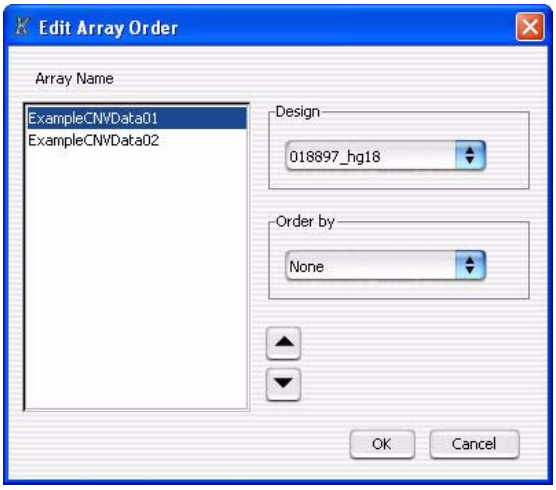


Figure 38 Edit Array Order dialog box

Purpose: Allows you to 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 defines the left-to-right display order of the arrays in Tab View. The array order also defines the order in which individual scatter plots appear in Gene View.

Design Select the desired design from the list.

Order by Lists the attributes associated with the arrays. Select an attribute to use for ordering the list. The arrays are re-ordered 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 Attributes

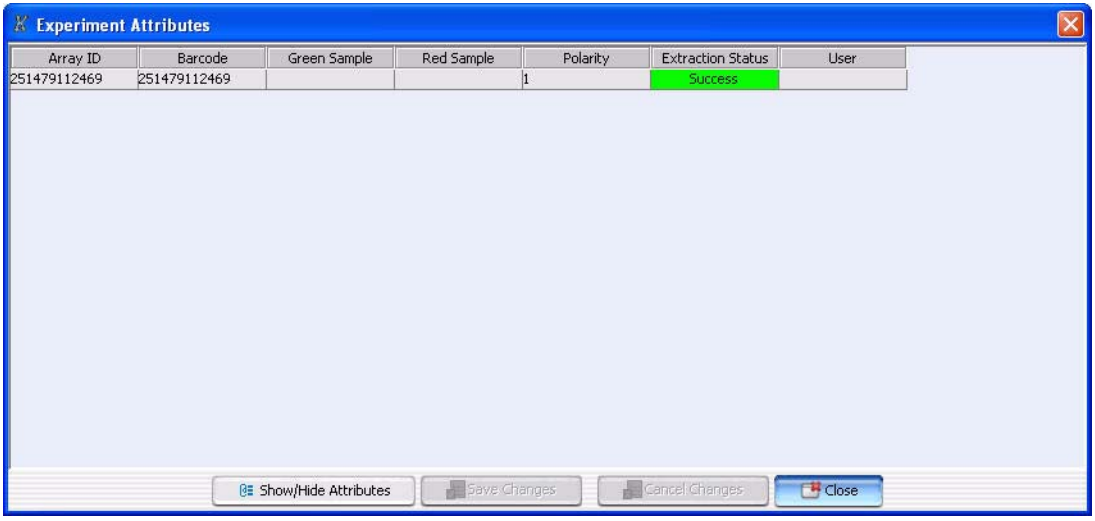


Figure 39 Experiment Attributes dialog box

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 **Edit Attributes.**

Table of Attributes Allows you to define 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*.

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 206.

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.

Experiment Properties

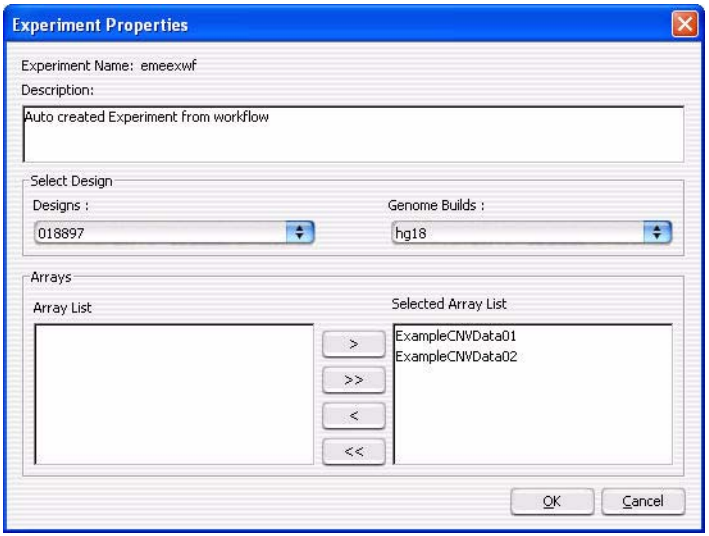


Figure 40 Experiment Properties dialog box

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 Text Box: Displays the description of the experiment that was entered when the experiment was created.

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 – A list of arrays that are available for this experiment.
- Selected Array List – A list of the arrays that you have selected for this experiment.
- Operators used to move files between Array List and Selected Array List list boxes:
  - > – Move selected file(s) from left to right list box.
  - >> – Move all files from left to right list box.
  - < – Move selected file(s) from right to left list box.
  - << – Move all files from right to left list box.

Export

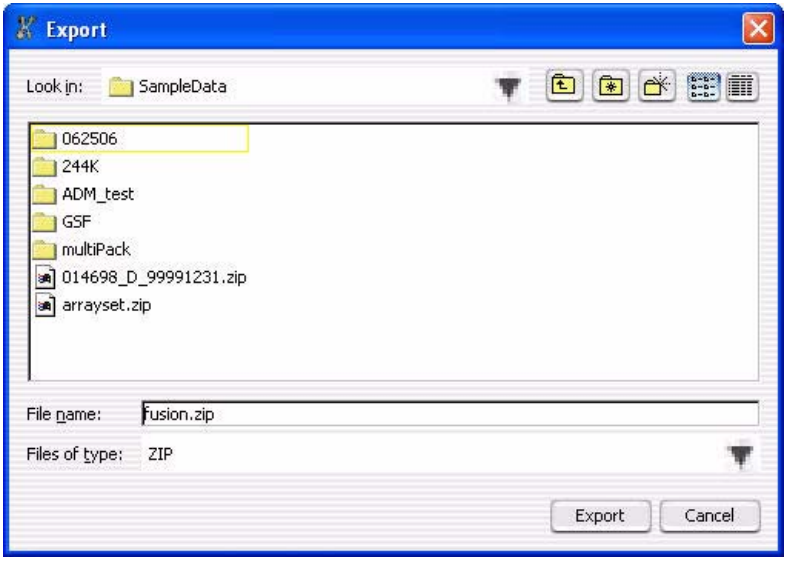


Figure 41 Export dialog box – Several types of file exports use this dialog box. This specific example exports selected experiment(s) as a ZIP format file.

Purpose: Allows you to 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 39, "To export tracks" on page 40, or "To export array attributes" on page 38.

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 ▼.


Navigates up one level.


Navigates 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 pain of the dialog box.

Main pane Displays the folders, files, and other locations in the selected location in *Look in*. The program restricts listed files to the type selected in *Files of type*. To select file, click its name. To open a folder or other location, double-click its name.

File name Displays the name of the file to which the exported content will be 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 Restricts 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 – Array

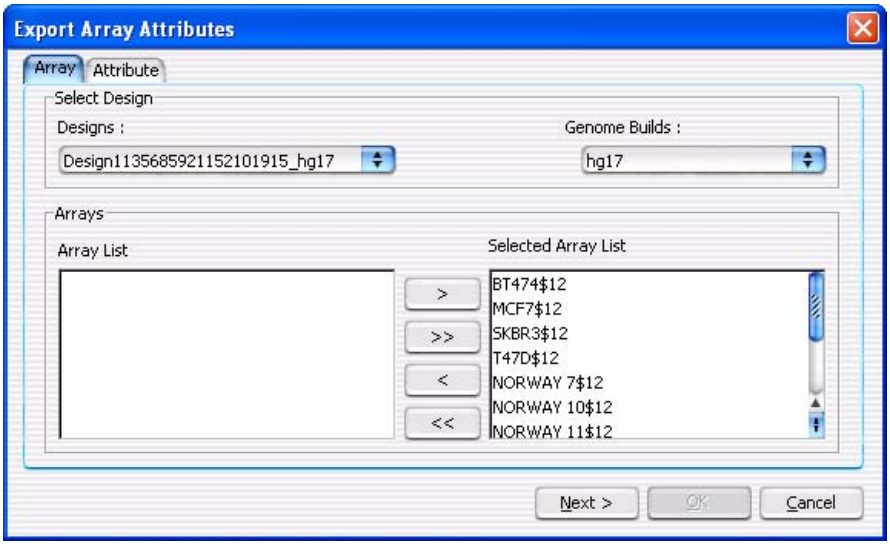


Figure 42 Export Array Attributes dialog box – Array tab

Purpose: Allows you to select arrays whose selected attributes you will then export. See "To export array attributes" on page 38.

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.**

Select Design

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

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

Arrays

Array List Lists 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, control-click their names.
- To select a contiguous block of arrays, click the name of the first array, then shift-click the name of the last one.

Selected Array List Lists 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.

Next Moves to the Attribute tab for attribute removal. See "Export Array Attributes – Attribute" on page 156.

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

Export Array Attributes – Attribute

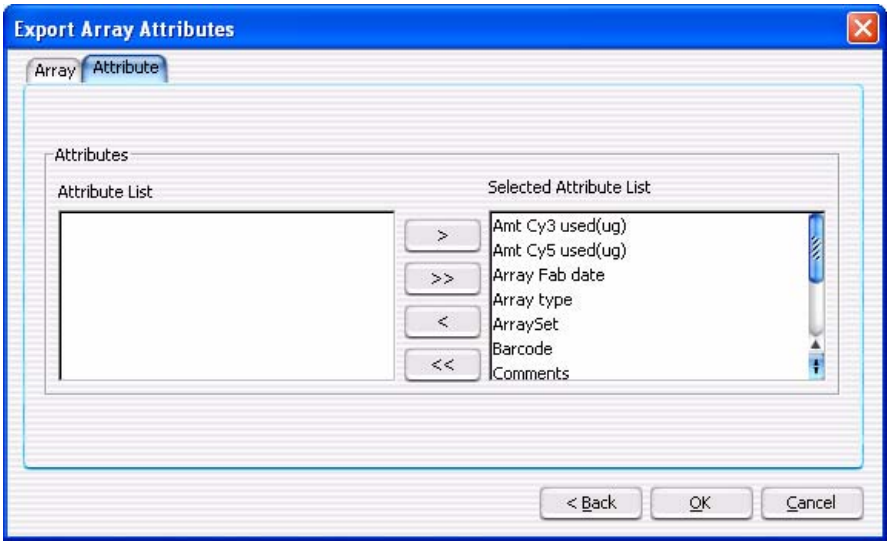


Figure 43 Export Array Attributes dialog box – Attribute tab

Purpose: Allows you to remove selected array attributes from the list you will then export. See "To export array attributes" on page 38.

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.** After selecting the arrays whose attributes you intend to export, click **Next.**

Attributes

Selected Attribute List

Lists the attributes for the selected arrays that will be exported when you click OK.

- To select an array for subsequent removal to the Attribute List, click its name.
- To select additional arrays, control-click their names.
- To select a contiguous block of arrays, click the name of the first array, then shift-click the name of the last one.

Attribute List Lists the attributes that will not be 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 152.

Cancel Closes the dialog box without adding the attributes in the Selected Attribute List to be exported.

Export Experiments

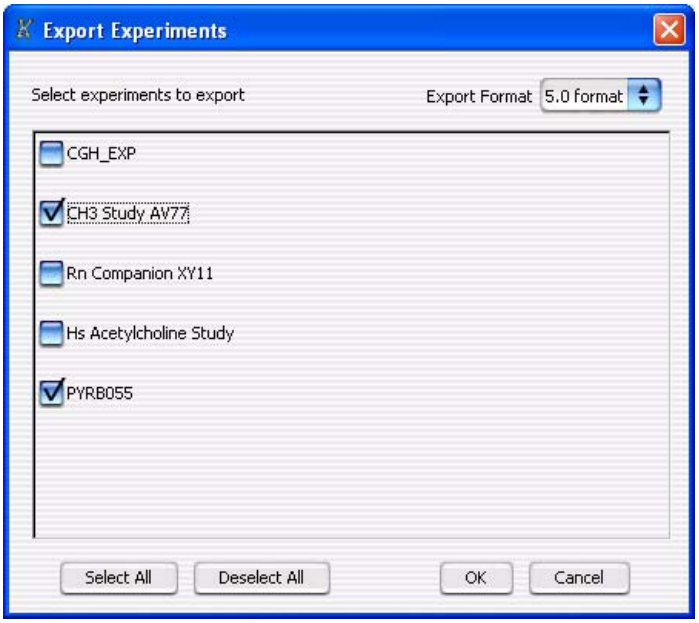


Figure 44 Export Experiments dialog box

Purpose: Allows you to 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 39.

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

Export Format The file format for the exported experiment file. Select one of these options:

- **5.0 Format** – Exports the experiment(s) in a format that you can import into Genomic Workbench. This is the most current experiment format, but it is not compatible with previous versions of the program.
- **3.0 Format** – Exports the experiment(s) in a format that you can import into Agilent CGH Analytics 3.0 or later. This is a "legacy" format that you can use to maintain compatibility with earlier versions of the program.

Select experiments to export Lists 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 152.

Cancel Cancels the export and closes the dialog box.

Export Tracks

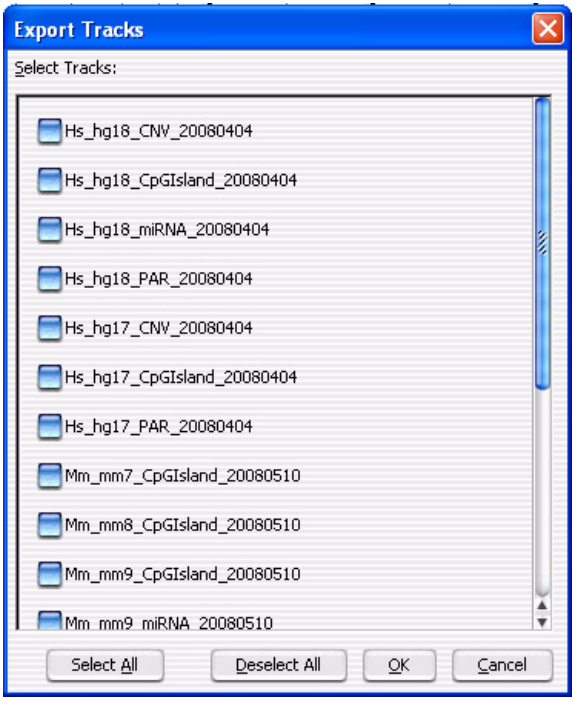


Figure 45 Export Tracks dialog box

Purpose: Allows you to select tracks to export as a single BED format file. See "To export tracks" on page 40.

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

Select tracks Lists 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 152.

Cancel Cancels the export and closes the dialog box.

Find in column

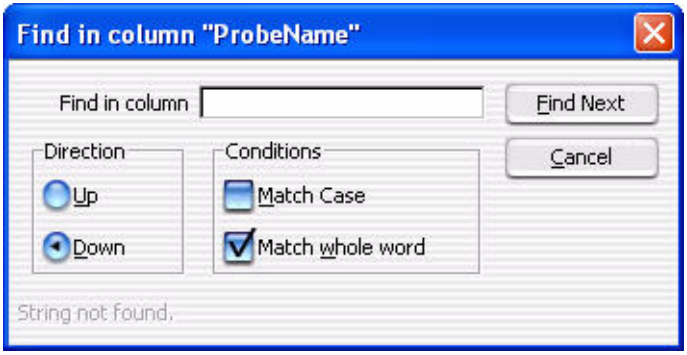


Figure 46 Find in column dialog box

Purpose: Allows you to set search parameters for a specific column entry. Based on these parameters, the program highlights the row of the first entry that matches. The cursor then moves to the location defined in the row.

To open: Right-click any entry in a tab in Tab View other than a column heading, then click **Find in column** in the shortcut menu.

Find in column Type all or part of the entry 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 defined in the row that contains the entry.

Cancel Closes the dialog box.

Gene List

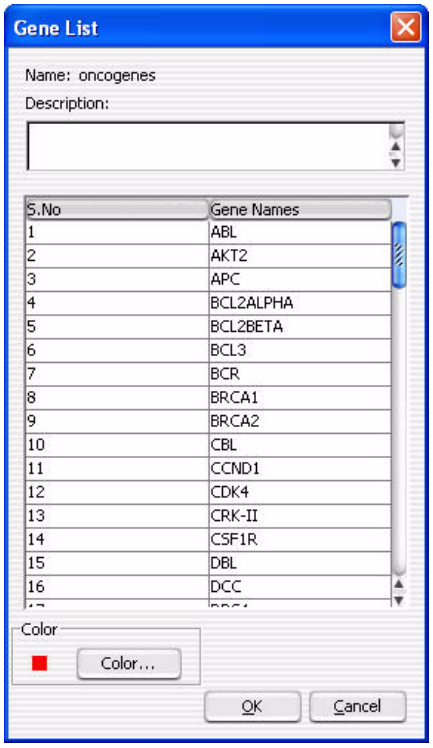


Figure 47 GeneList dialog box

Purpose: Allows you to view the names of the genes in a specific gene list and to change the display color of the gene list. See "To view the genes in a gene list" on page 34.

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 Name 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

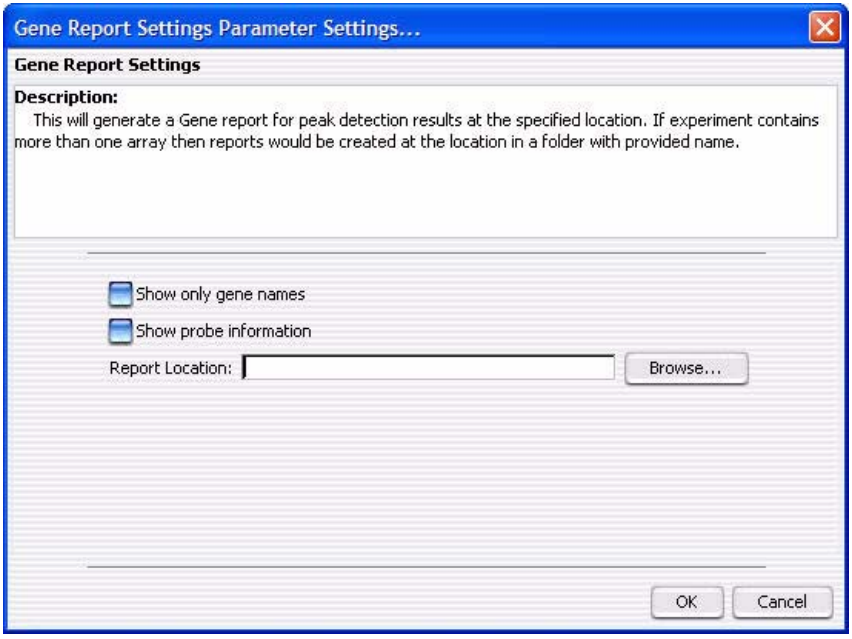


Figure 48 Gene Report Settings Parameter Settings dialog box

Purpose: The Gene Report Settings dialog box allows you to 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) 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.

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

The program creates gene reports in several formats. See "Gene Report formats" on page 215 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 Name dialog box, where you can select a location for the gene report. See "Select report name" on page 200.

OK Creates the gene report.

Cancel Closes the dialog box without creating a gene report.

Go To Gene/Genomic Location

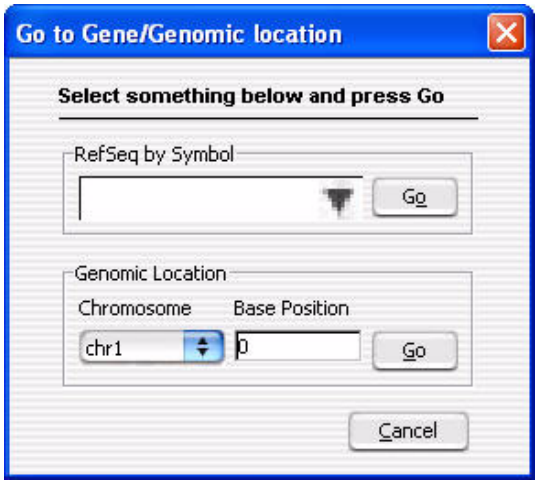


Figure 49 Go To Gene/Genomic location dialog box

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

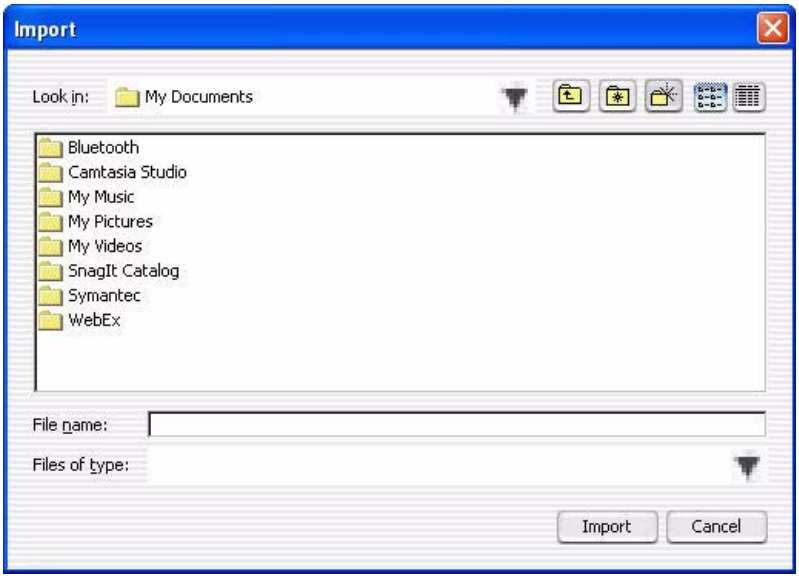


Figure 50 Import dialog box

Purpose: Allows you to select files for import into 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 control-click the names of additional files. To select a contiguous block of files, click the name of the first file in the block, then shift-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 restricts the list of files to the specific types expected for the import. To display all files, click ▼,then select **All Files.**

Type of import	File type
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 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) allows you to further define the content for import. See the instructions for each specific type of import in Chapter 1.

Cancel Cancels the import and closes the dialog box.

Import (experiments)

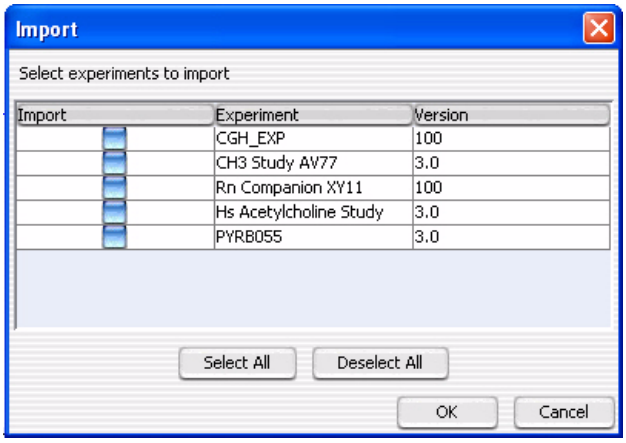


Figure 51 Import dialog box (for experiments)

Purpose: Allows you to select the specific experiments within a ZIP format experiment file to load into the program. See "To import an experiment file" on page 21.

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**

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 139

Cancel Cancels the upload and closes the dialog box.

Import GEML design files

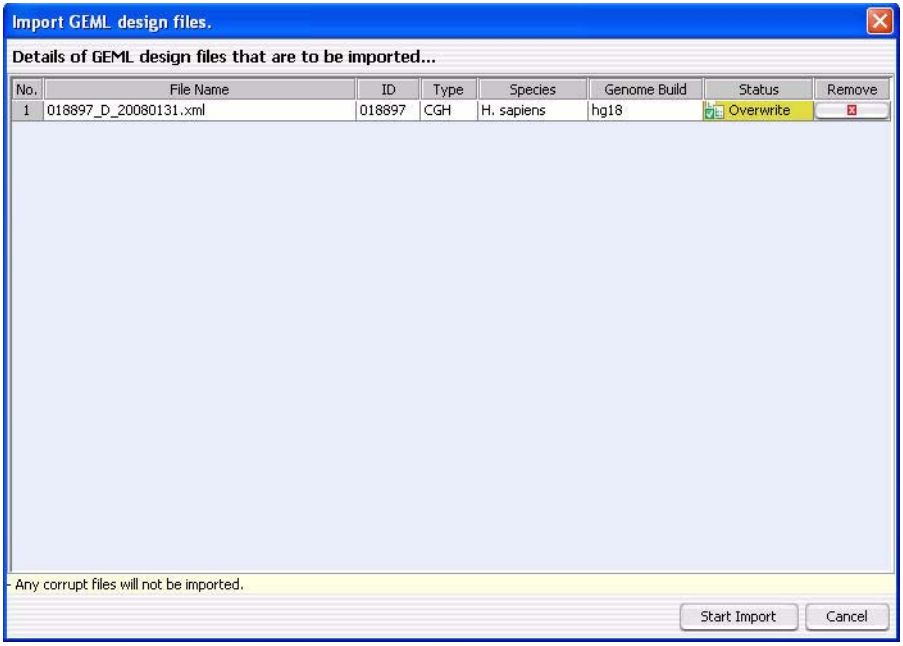


Figure 52 Import GEML design files dialog box

Purpose: Allows you to 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 At present, Genomic Workbench supports these species:

- *H. sapiens*
- *M. musculus*
- *R. norvegicus*

Genome Build The genome build with which this design is associated.

Status

- **Overwrite** – Appears when the design file has been updated and will overwrite any existing one of the same name.
- **Valid** – Appears when the file is new.
- **Corrupt** – Appears when the file is corrupt.

Remove Click  to remove a specific design file from the list.

Start Import Starts the import of the design files in the list.

Cancel Cancels the upload and closes the dialog box.

Import Genome Build

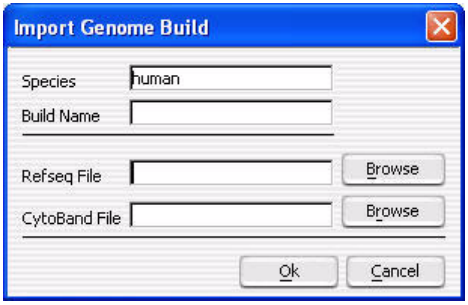


Figure 53 Import Genome Build dialog box

Purpose: To import a new set of genome build files into Genomic Workbench. See "To import a genome build" on page 18.

To open: In the Home tab, click **Import > Genome Build.**

Species The genome's species of origin. The program supports these species:

- *H. sapiens*
- *M. musculus*
- *R. norvegicus*

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

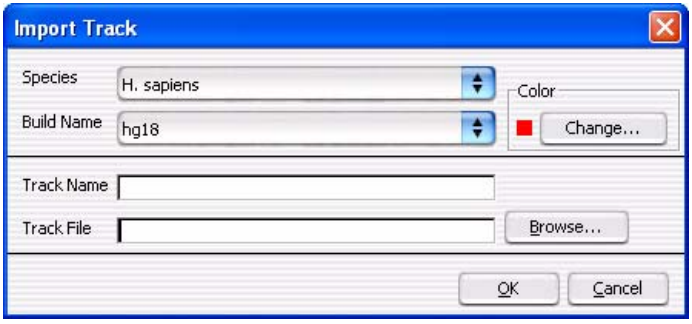


Figure 54 Import Track dialog box

Purpose: Allows you to import a BED format track file. See "To import tracks" on page 19. 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. The program supports these species:

- *H. sapiens*
- *M. musculus*
- *R. norvegicus*

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 - Attribute Tab

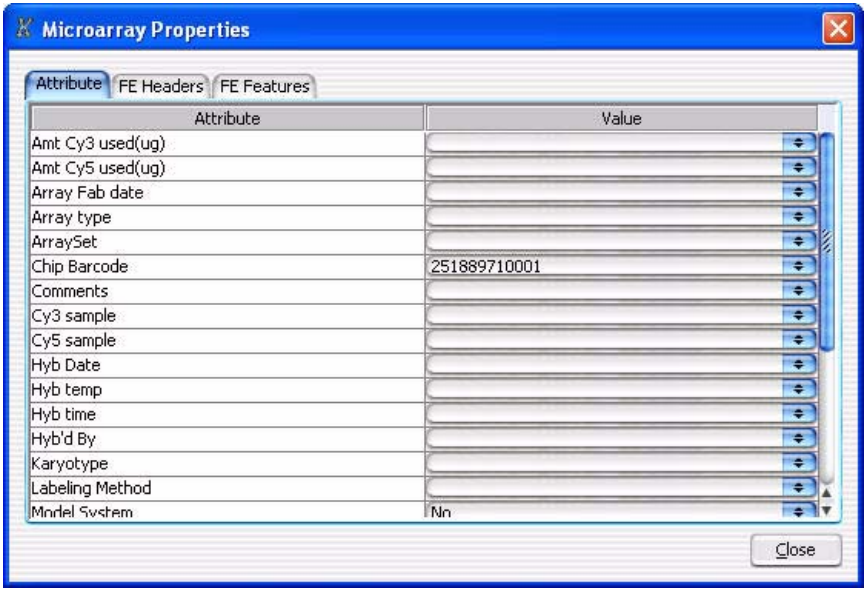


Figure 55 Microarray Properties dialog box listing Attributes and their values

Purpose: To view attributes, and their associated values, that are associated with an array.

To open: For any array in the Data folder or Experiments folder, right-click the array name, then click **Show Properties.** For third-party or non-Agilent arrays, only the Attribute Tab appears.

Attribute Tab

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

Close Closes the dialog box.

For the FE Headers tab options, see "Microarray Properties - FE Headers" on page 175.

For the FE Features tab options, see "Microarray Properties - FE Features Tab" on page 174.

Microarray Properties - FE Features Tab

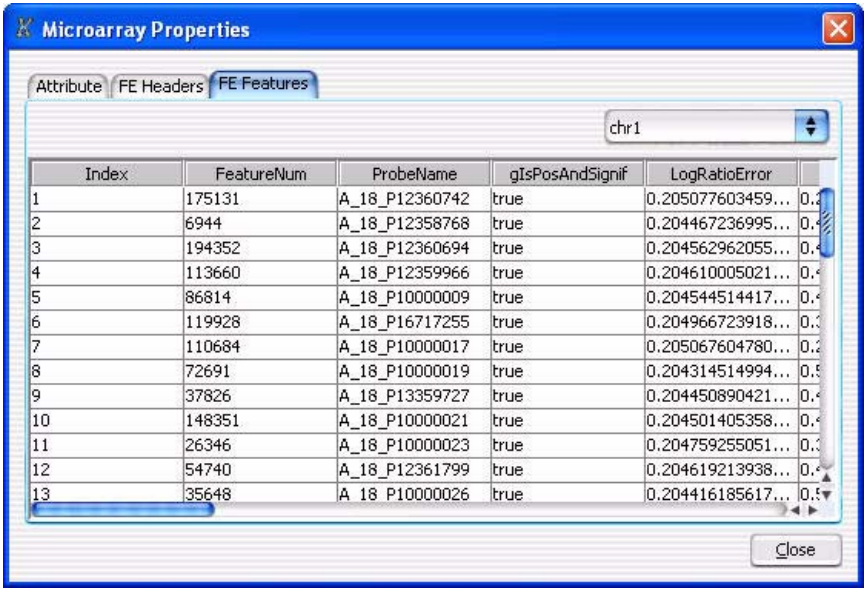


Figure 56 Microarray Properties dialog box listing FE Features and associated data

Purpose: To view feature information for arrays extracted with Agilent Feature Extraction software.

To open: Right-click the array name and click Show Properties.

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

Table of features Displays FE features and their associated data.

Microarray Properties - FE Headers

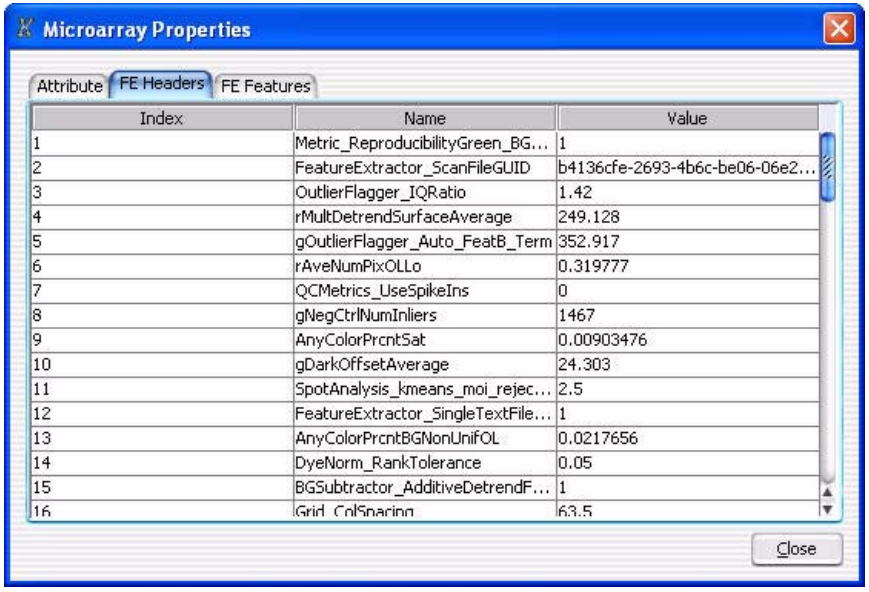


Figure 57 Microarray Properties dialog box listing FE Headers their values

Purpose: To view feature parameters, statistics and constants for the whole array extracted with Agilent Feature Extraction software.

To open: For Agilent arrays in the Data folder or Experiments folder, right-click the array name, then click **Show Properties.**

Index Displays a sequential index to help identify FE properties.

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

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

Close Closes the dialog box.

Preferences – License

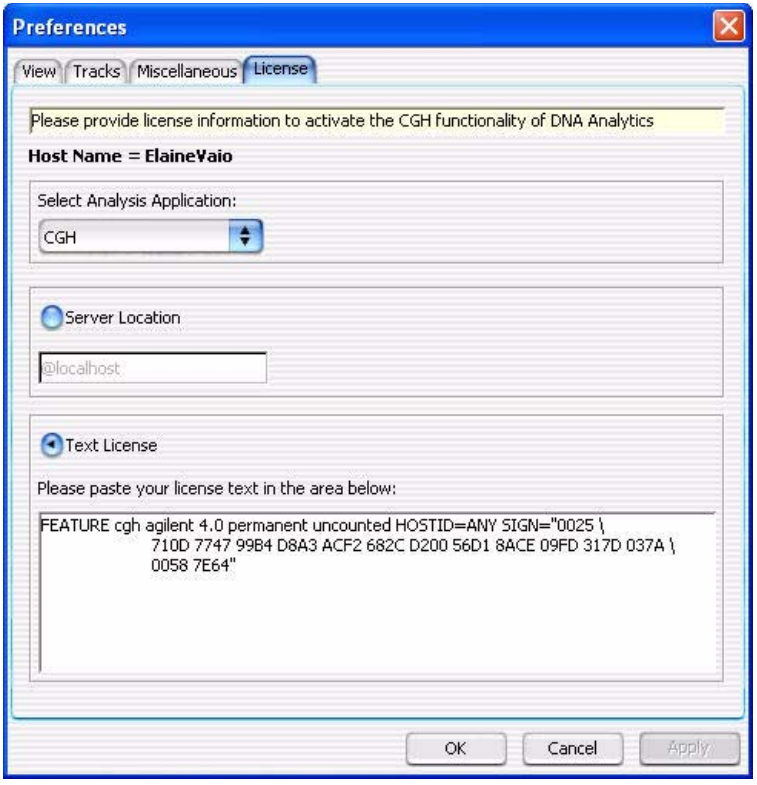


Figure 58 Preferences dialog box displaying License tab options

Purpose: To view and update the license for each application that you have installed with the DNA Analytics software, if necessary.

To open: Right-click anywhere in the graphical interface, either in Genome View, Chromosome View or Gene View, click Preferences, then click the License tab.

Host Name Displays the host name, automatically.

Select Analysis Application Select the DNA Analytics application for which you have a license.

Server Location Server location should be selected if you have a concurrent user license. If appropriate, click to enable and type in the name of your license server. The default is @localhost. Replace localhost with the name of the computer used as the license server. **Text License** is unavailable if **Server Location** is selected.

Text License Text licenses are used if you have a workstation license. If you have a workstation license, paste your license in the box. If you have entered a license previously, it appears in the box. **Server Location** is unavailable if **Text License** is selected.

Apply Applies your changes to the parameters.

OK/Cancel Accepts your changes and exit, or cancel all changes and return to the previously selected parameters.

Preferences – Miscellaneous Tab

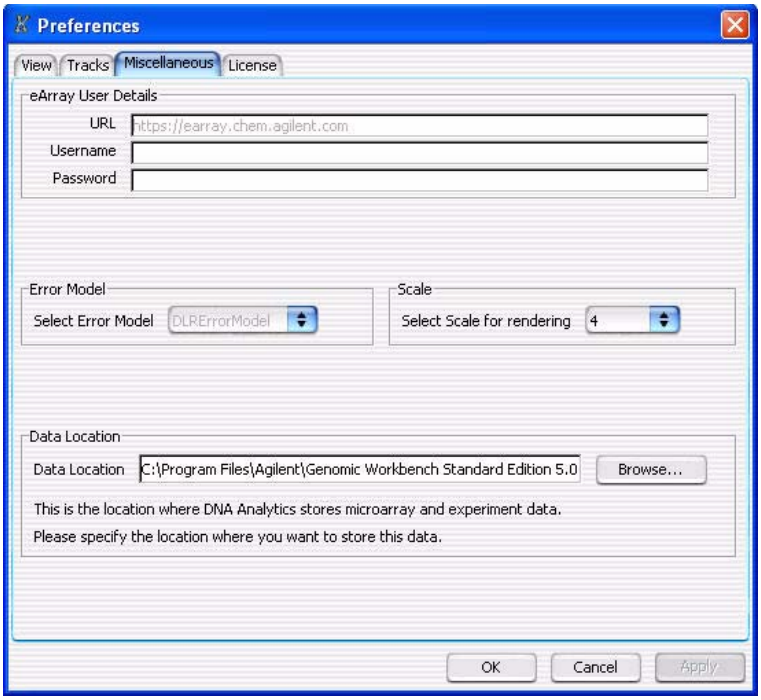


Figure 59 Preferences dialog box displaying Miscellaneous tab options

Purpose: In the ChIP Interactive Analysis application, allows you to change the location for data.

To open: Right-click anywhere in the graphical interface, either in Genome View, Chromosome View or Gene View, click Preferences and 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.

Data Location

Type in the path to the folder where you want the ChIP application to store the microarray and experiment data.

Apply Applies your changes to the parameters.

OK Accepts your changes and closes the dialog box.

Cancel Cancels all changes and closes the dialog box.

Preferences – Tracks Tab

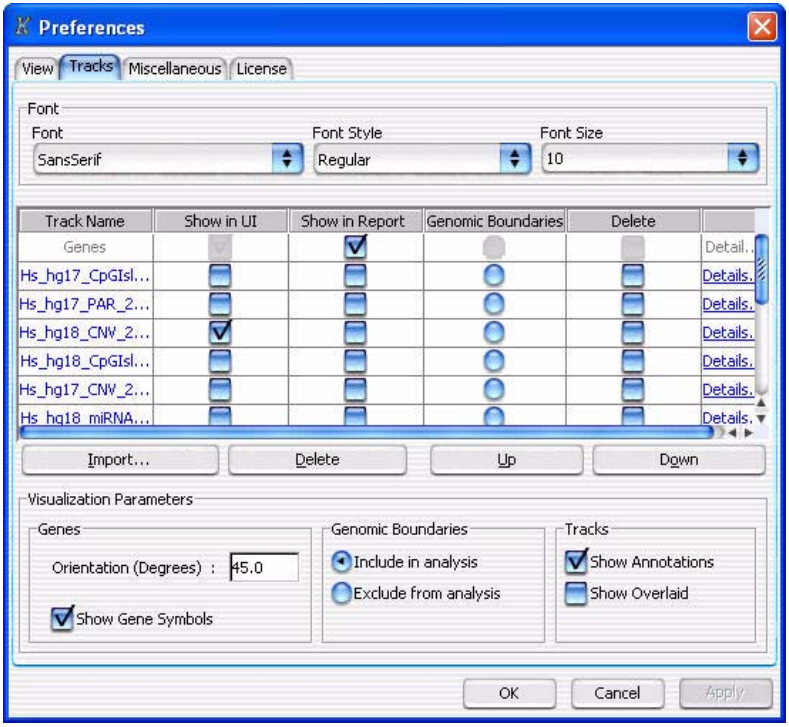


Figure 60 Preferences dialog box displaying Tracks Tab options

Purpose: To import and set up the appearance of tracks next to Gene View. Tracks contain annotation or other information aligned with specific genomic coordinates.

To open: Right-click anywhere in the graphical interface, either in Genome View, Chromosome View or Gene View, click **Preferences** and 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 view the track next to Gene View.

Show in Report Mark the check box to view the track information in all the reports.

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

Delete Click to delete the track from the list.

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

Import Opens the Import Track dialog box, where you can import a new track. See

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 Gene View.

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

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

Tracks These options affect the appearance of the Track Views.

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

Preferences – View Tab

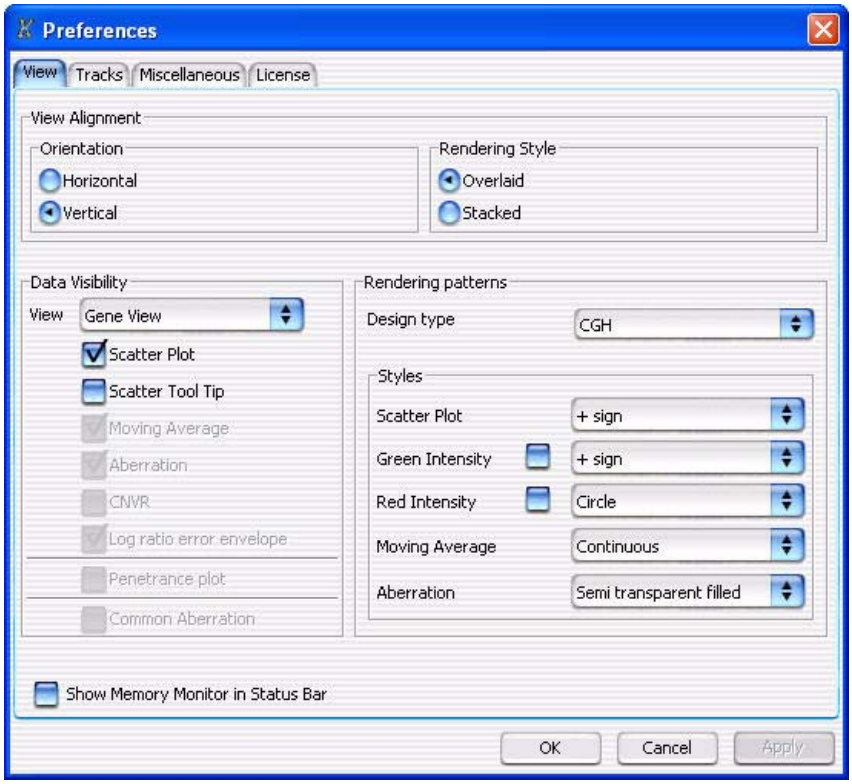


Figure 61 Preferences dialog box displaying View tab options for ChIP Interactive Analysis

Purpose: To set up how the Views are aligned and settings for the scatter plot in the Genome Viewer. The View tab options control the general characteristics of how data are displayed on your monitor.

To open: Right-click anywhere in Genome View, Chromosome View or Gene View, click **Preferences,** then click the **View** tab.

View Alignment

Orientation Selects the orientation of three graphical views in the main window:

- Horizontal – Reorients the views horizontally in the order of Gene, Chromosome, and Genome views, top to bottom. The Navigator and Tab view orientation remains unchanged.
- Vertical – Displays all views vertically, left to right: Navigator, Genome, Chromosome, and Gene views. See "Genomic Workbench ChIP Application Main Window" on page 86.

Rendering Style Selects the way Chromosome and Gene data are rendered on your screen.

- Overlaid – Displays data from multiple arrays superimposed one on top of another (default).
- Stacked – Displays data from each array in a separate plot.

Data Visibility

View Choose what features you want to display for the Genome, Chromosome, and Gene views, either individually or together. Select one or more check boxes:

- Scatter Plot
- Scatter Tool Tip (Gene View only)

Rendering patterns

Design type Specify the type of design to which you are applying these patterns: **ChIP, Expression, or Other.**

Styles Set the parameters for displaying your data.

- Scatter Plot – Specify how to display individual data points as: **Color filled circles (ellipses), + signs,** or **x signs.**

NOTE Rendering scatter plots for more then 10 high density arrays in the Chromosome View may take significant time. Selecting ellipses as the rendering style for ChIP scatter plots can also decrease performance. Please change the rendering style for ChIP data from ellipse to the plus (+) or cross hair sign.

Apply Applies your changes to the parameters.

Probe Report Settings Parameter Settings

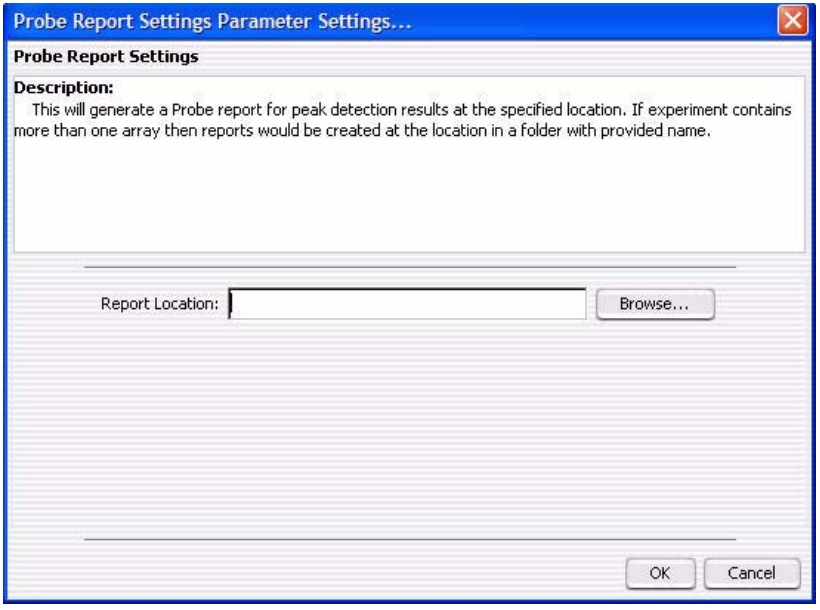


Figure 62 Probe Report Settings Parameter Settings dialog box

Purpose: The Probe Report Settings dialog box allows you to 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 213 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 name" on page 200.

OK Generates the probe report in the selected location.

Cancel Closes the dialog box without generating a report.

Properties (of an experiment result)

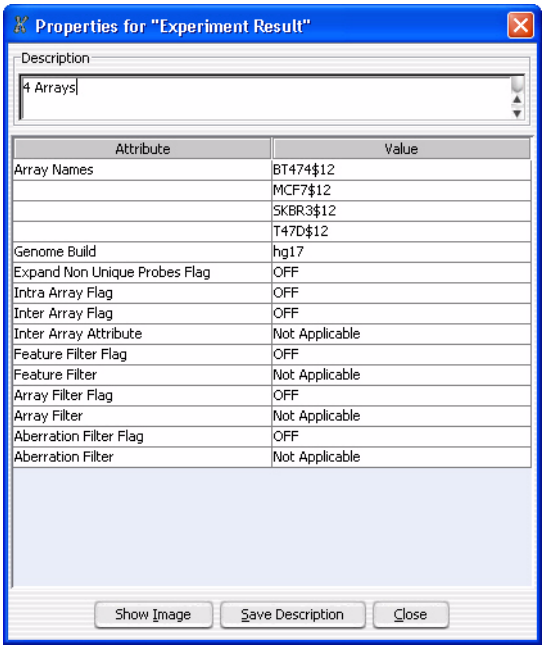


Figure 63 Properties dialog box (for experiment results)

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

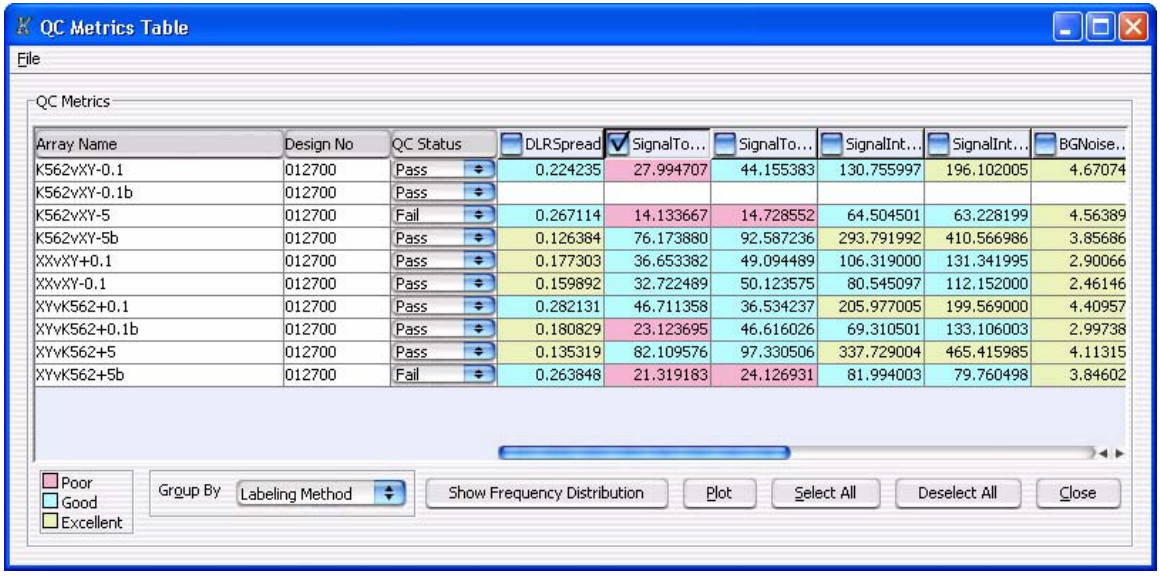


Figure 64 QC Metrics Table

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 QC status to each microarray. Some metrics are passed through from the Agilent FE program, while others are calculated by the CGH application itself. These metrics are available only for Agilent microarrays. See "To view QC metrics of arrays and set array QC status" on page 74.

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.**

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

- **Array Name** – Lists 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 design that contains each microarray.
- **QC Status** – Allows you to set the QCMetricStatus attribute of the array. Status can be Pass, Fail, Marginal, or NA. Later, you can filter arrays based on this attribute.
- **Metrics** – The program evaluates each metric, and assigns it a rating of Excellent (yellow), Good (turquoise), or Poor (pink). 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.

The following metrics appear in the table:

Metric	Comments
DLRSpread	(CGH only) This metric is an indicator of the noise in the experiment, and thus the minimum log ratio difference required to make reliable aberration calls. For details on how the program calculates this value, The cut-offs the program uses to score this metric are: • Excellent < 0.2 • Good 0.2 – 0.3 • Poor > 0.3
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. The cut-offs the program uses to score this metric are: • Excellent: > 100 • Good: 30 – 100 • Poor: < 30
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. The cut-offs the program uses to score this metric are: • Excellent: > 150 • Good: 50 – 150 • Poor: < 50
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. The cut-offs the program uses to score this metric are: • Excellent: < 5 • Good: 5 – 10 • Poor: > 10

Metric	Comments
ReproducibilityGreen (and Red)	For each channel, this metric is the Median %CV 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, at least three probes are required 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 (e.g. 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. The cut-offs the program uses to score this metric are: • Excellent: < 0.05 • Good: 0.05 – 0.2 • Poor: > 0.2
Model System Metrics	The metrics below apply only to CGH model systems. Currently the only model system supported is a male (XY) CY5-labeled vs. female (XX) CY3-labeled comparison.
AreaUnderROC	Method: Sort the log ratios in ascending order for the entire array. Each log ratio in the data set comes from an X-probe or an autosome. If it is an X-probe, it contributes to the number of true positives (TP). If it is an autosome, it contributes to the number of false positives (FP). So for each log ratio, start from the lowest and continue incrementing either TP (if an X-probe) or FP (if an autosome). Then for each log ratio, plot FP/(total number of autosomes) vs. TP/(total number of X-probes). Use the trapezoidal rule to estimate the area under this curve. The result is this metric: AreaUnderROC. Excellent < 0.95 Good 0.85 – 0.95 Poor > 0.85

Metric	Comments
MedianDiff	This metric is the difference between the medians of the histograms of X-probes and autosomes. Excellent > 0.9 Good 0.8 – 0.9 Poor < 0.8
ErrorFraction	The minimum value of all error fractions calculated. The error fraction is: (FP/(total number of autosomes) + (1 – TP/(total number of X-probes)))/2. FP is the number of false positives, and TP is the number of true positives. Excellent > 0.05 Good 0.05 – 0.1 Poor < 0.1

Group By The program displays the arrays in the table, and also 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 190.

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 191.

File Allows you to 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.

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

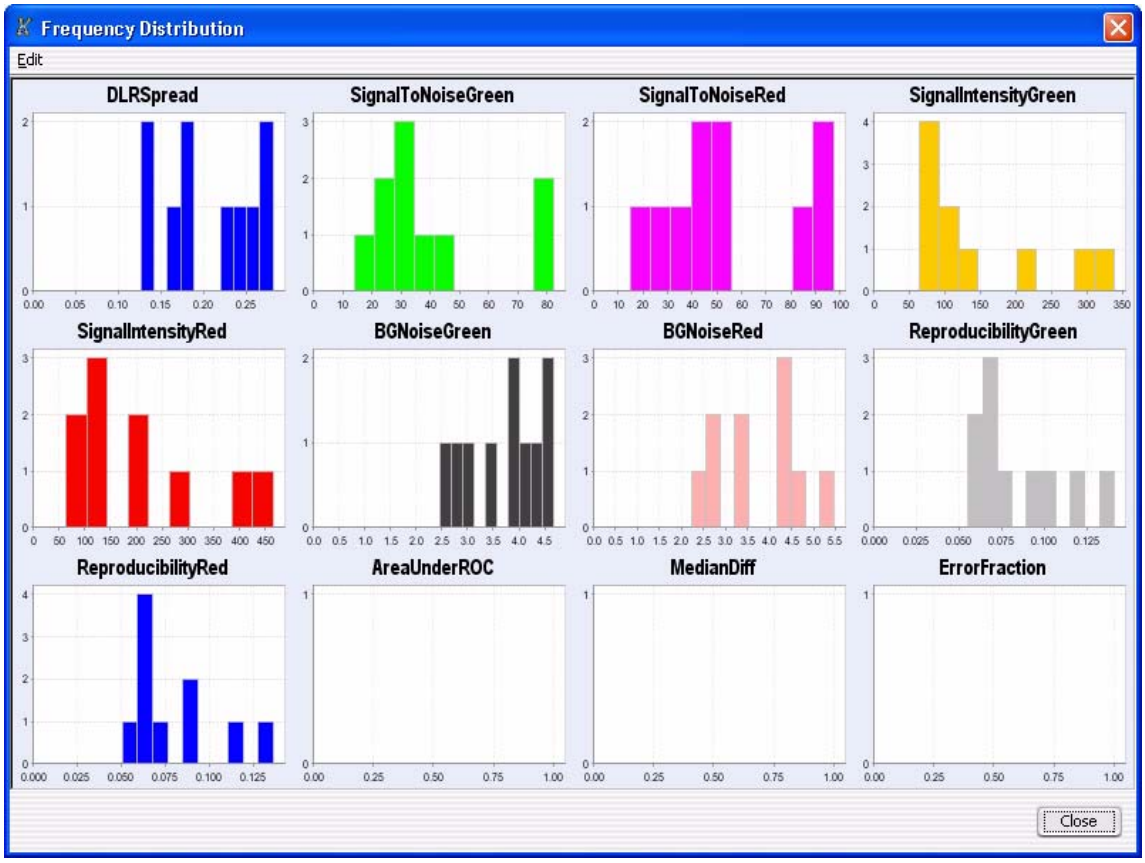


Figure 65 QC Metrics Frequency Distribution Plot

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 185**.

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, whose image you can then paste into a document.

Close Closes the dialog box.

QC Metrics Graph

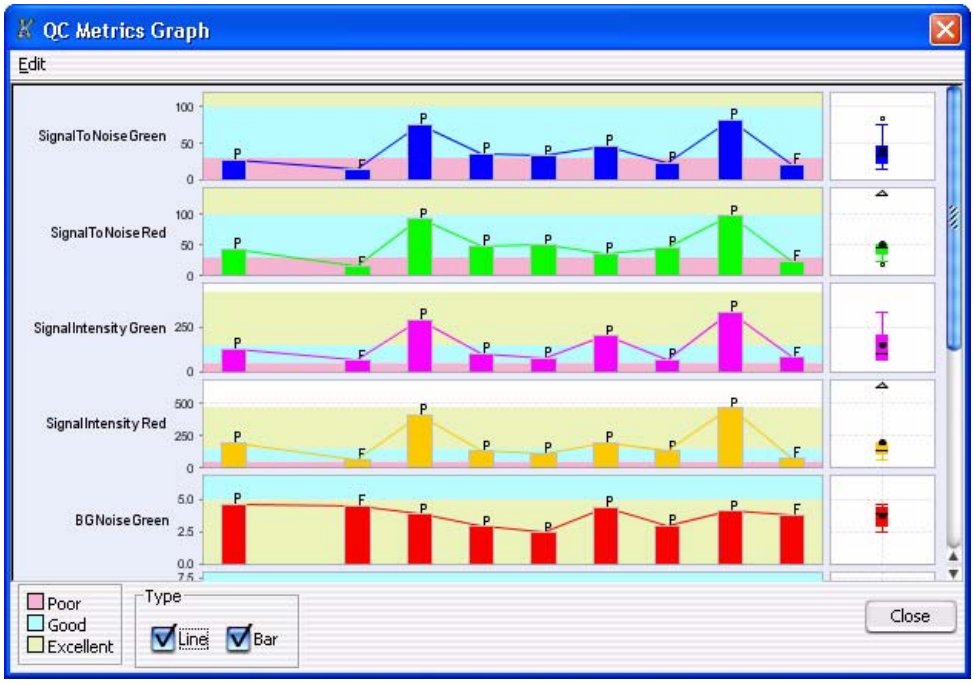


Figure 66 QC Metrics Graph

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 185).

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 185.
- 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 tool tips (see "Preferences – View Tab" on page 181), a tool tip appears when you place the pointer over any bar. The tool tip lists the value of each bar and the name of the corresponding array. See Figure 67.
- 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 67.
- 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.

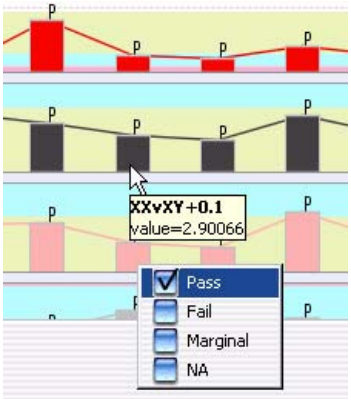


Figure 67 Portion of the QC Metrics Graph, showing a ToolTip (values in yellow box), and a shortcut menu. You use the shortcut menu to set the QCMetricStatus attribute for the array. Right-click any bar to open the shortcut menu for the corresponding array.

"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 68. The symbols carry 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.

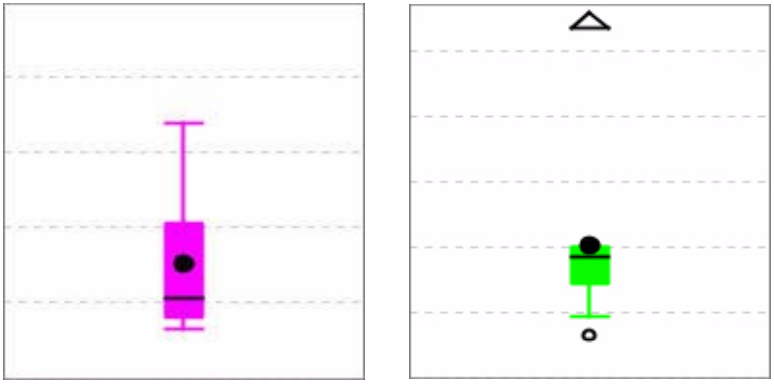


Figure 68 Two "Box & Whisker" plots

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

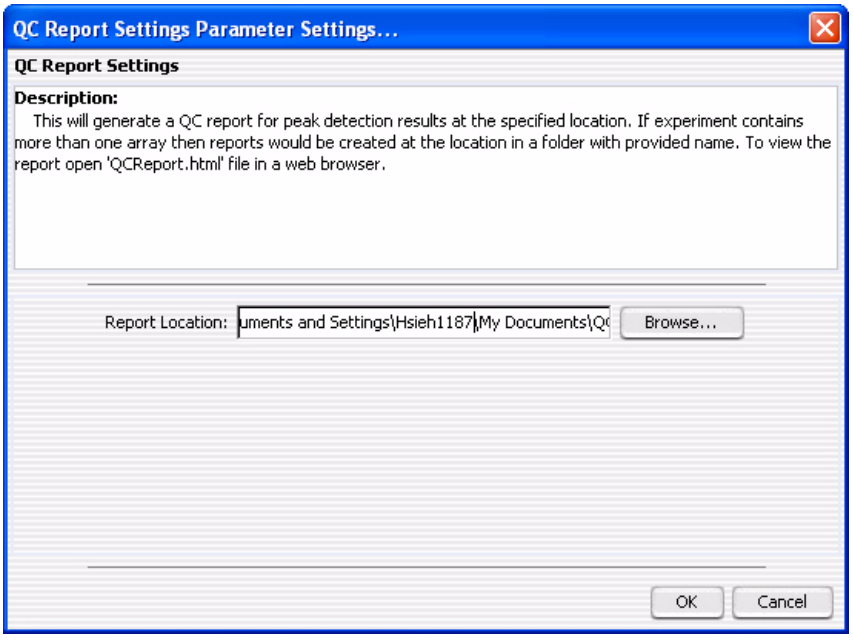


Figure 69 QC Report Settings Parameter Settings dialog box

Purpose: This dialog box allows you to 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 details about the contents of this report, see "QC Report format" on page 216

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 name" on page 200.

OK Generates the QC report in the selected location.

Cancel Closes the dialog box without generating a report.

Scroll to Column

Figure 70 Scroll to Column dialog box

To open this dialog box, right-click a column heading in Tab View, then click Scroll To Column... in the shortcut menu. It allows you to select a column. The program then scrolls the tab so that you can see the selected column.

Select column Lists the columns available in the currently selected tab. Select the one you want to view.

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

Cancel Closes the dialog box.

Search probes in eArray

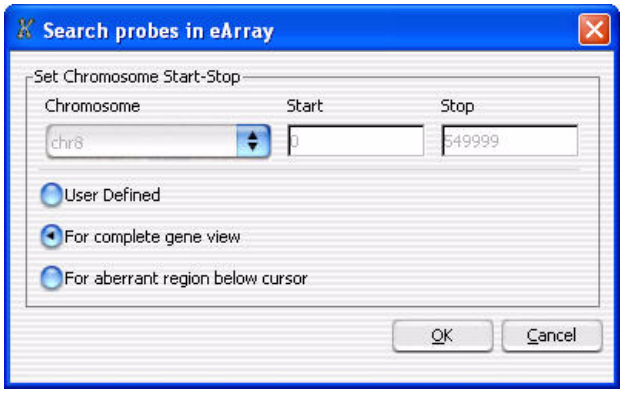


Figure 71 Search probes in eArray

Purpose: To select the probes you want to update in eArray

To open: Right-click Gene View, then click **Search probes in eArray**.

By default, the program pre-sets Chromosome Start-Stop values based on the genomic region that is appears in Gene View.

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.

For aberrant region below cursor Selects those probes for the genes that appear just below where the cursor sits in Gene View. Not operational in ChIP Interactive Analysis.

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 Start and Stop positions for defining the region contained the genes to be in the list.

Select Color (Edit Array Color) — Swatches Tab

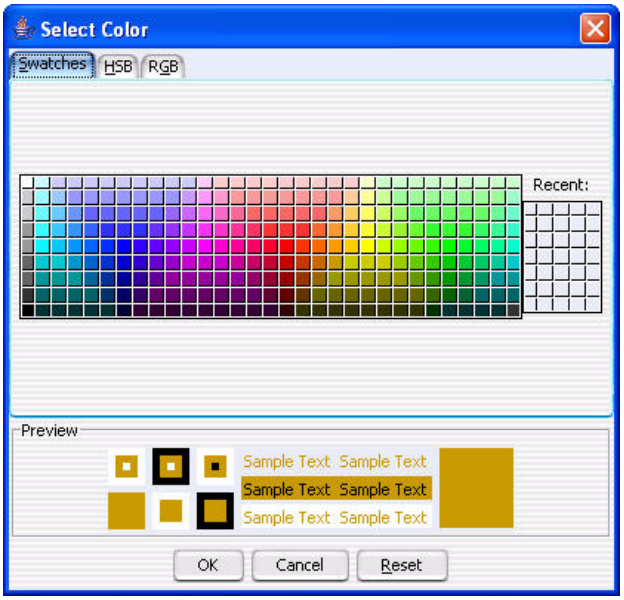


Figure 72 Select Color — Swatches Tab

Purpose: To select a color for each array based on color samples (swatches)

To open: Right-click on an array in an experiment, click Edit Array Color and click Swatches tab.

HSB Tab Choose colors based on an HSB schema (Hue, Saturation, and Brightness or Value). See

RGB Tab Choose colors based on an RGB schema (Red-Green-Blue).

Recent: Choose a recent color selection.

Reset Click to return HSB or RGB values back to default values.

Select Color (Edit Array Color) — HSB Tab

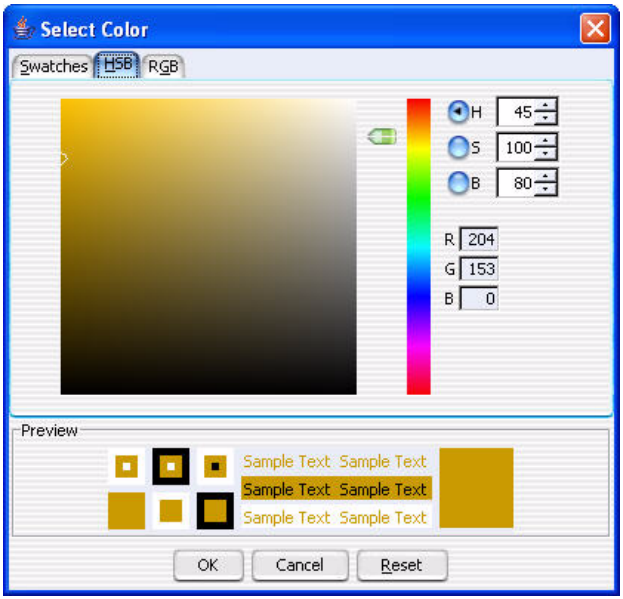


Figure 73 Select Color — HSB Tab

Purpose: To select a color for each array based on an HSB schema (Hue, Saturation, and Brightness)

To open: Right-click on an array in an experiment, click Edit Array Color and click HSB tab.

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.

Reset Click to return HSB values back to default values.

Select Color (Edit Array Color) — RGB Tab

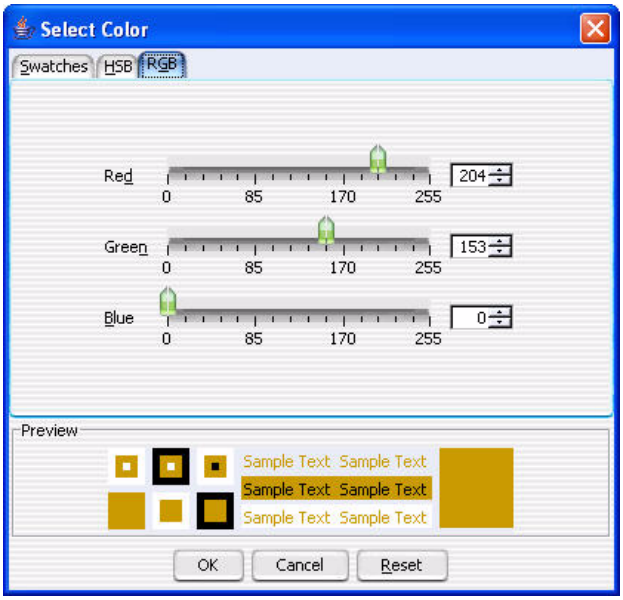


Figure 74 Select Color — RGB Tab

Purpose: To select a color for each array based on an RGB schema (Red- Green- Blue)

To open: Right- click on an array in an experiment, click Edit Array Color and click RGB tab.

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.

Reset Click to reset the RGB values back to default values.

Select report name

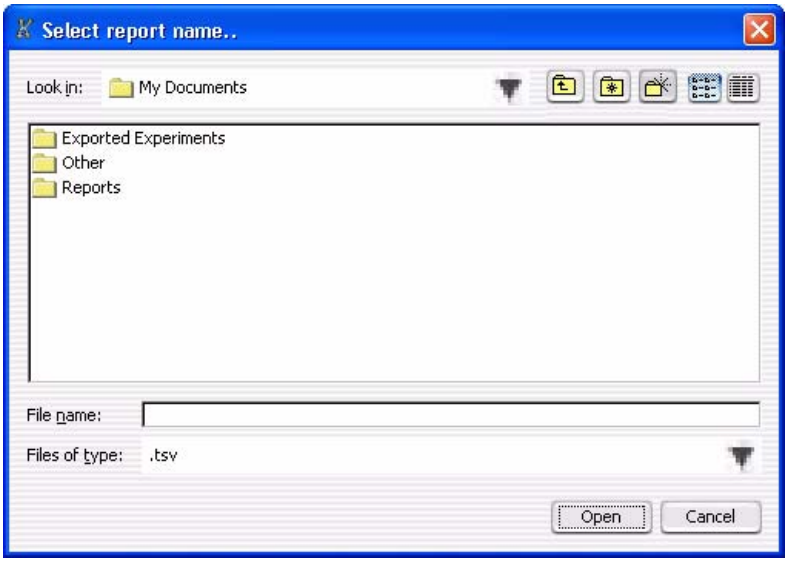


Figure 75 Select report name dialog box

Purpose: This dialog box allows you to 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 96.

Use the standard Windows Explorer commands available in the Select report name 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

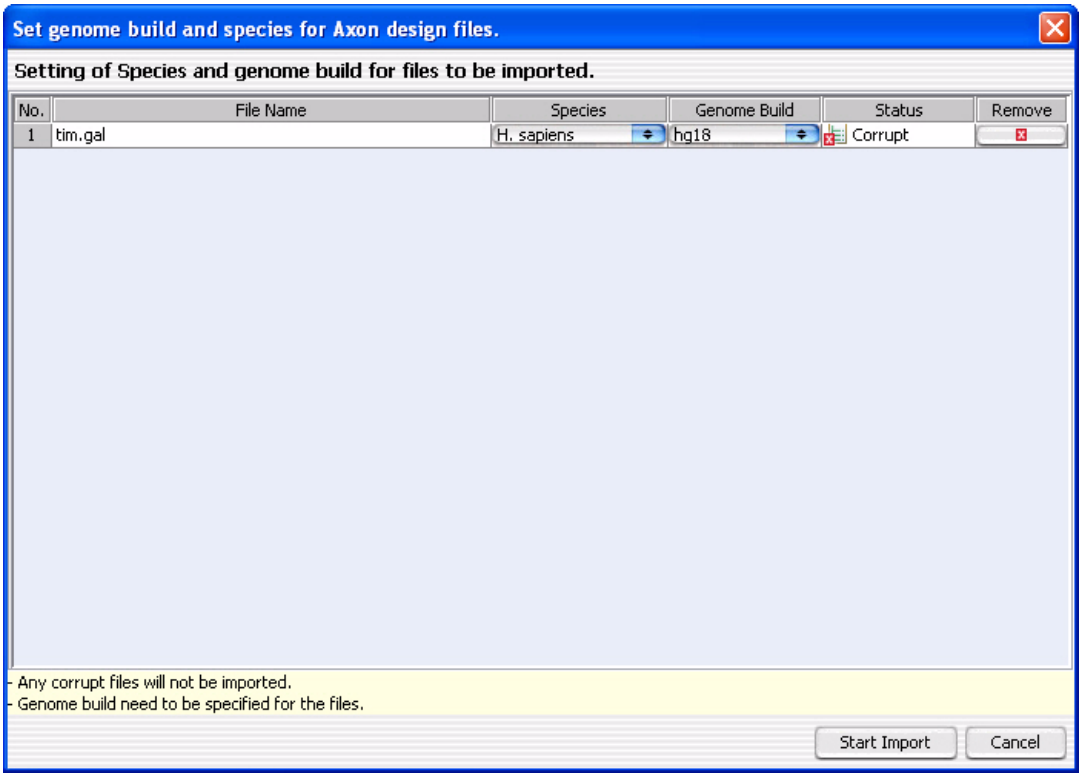


Figure 76 Set genome build and species for Axon design files dialog box

Purpose: Allows you to 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 GenePix/Axon design files" on page 17.

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.
- **Overwrite** – The file is a valid design file, but when you import it, it will replace an existing design that has the same name.
- **Corrupt** – The file failed validation. When you start the import process, the program ignores the file.

Remove Click  to remove a specific design file from the list. This can be 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 Order and Parameters

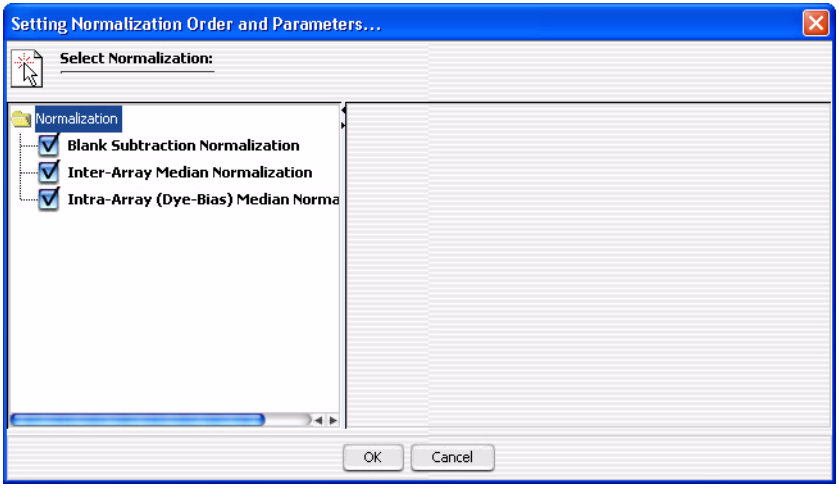


Figure 77 Setting Normalization Order and Parameters dialog box

Purpose: The Setting Normalization Order and Parameters dialog box allows you to include and configure the normalization methods that the program applies during analysis. See "To apply normalization" on page 67 and "Normalization Algorithms" on page 224.

To open: This dialog box opens when you mark Apply under Normalization 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*.

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	Defines how the program calculates the central tendency that it subtracts from the raw signal intensity values. You can select either of the following options: • median • One-step Tukey biweight

Inter-Array Median Normalization

Mark this check box to include inter-array median normalization in your analysis. This kind of normalization corrects for variations from one replicate array to another. The program 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.

Click the words Inter-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 setting:

Setting	Description
Estimate central tendency of common probes by	Defines how the program computes a "median" when it applies this kind of normalization. You can select either of the following options: • median • One-step Tukey biweight

Intra-Array (Dye Bias) 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.

Click the words Intra-Array (Dye-Bias) 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	Defines 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.
Estimate central tendency of data probes by	Defines how the program computes a "median" when it applies this kind of normalization. You can select either of the following options: • median • One-step Tukey biweight

OK Configures normalization according to the settings in the dialog box.

Cancel Closes the dialog box without changing any normalization parameters.

Show/Hide Columns

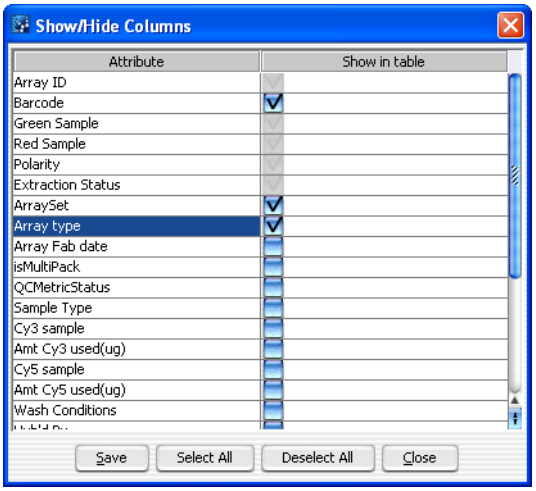


Figure 78 Show/Hide Columns dialog box

Purpose: To select the attributes that will be displayed in the Sample Utilities tab.

To open: This dialog box appears when you press the Show/Hide Attributes button at the bottom of the Sample Utilities tab.

All available attributes are shown in the Attributes column. Attributes with a check-mark next to them will be displayed in the Sample Utilities tab for each sample. To select an attribute for display, click the Show in Table box next to it. To de-select a currently-selected attribute, click the Show in Table box again.

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

Select All Selects all the attributes in the list.

Deselect All 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

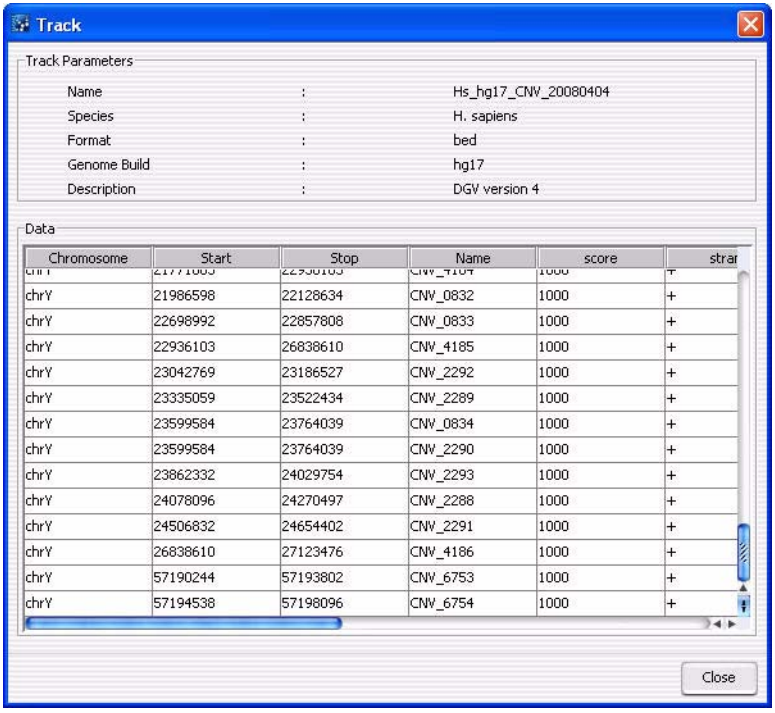


Figure 79 Track details

Purpose: This dialog box allows you to 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 Preferences dialog box. See "Preferences – Tracks Tab" on page 179.

Track Parameters These parameters appear:

Parameter	Description
Name	The name of the track.
Species	The species to which the track applies.

Parameter	Description
Format	The format of the track data. DNA Analytics 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 defined region for the feature.

The other columns are additional BED track file columns that can appear for some tracks. DNA Analytics does not render these.

Close Closes the Track dialog box.

Whitehead Error Model Parameter Settings

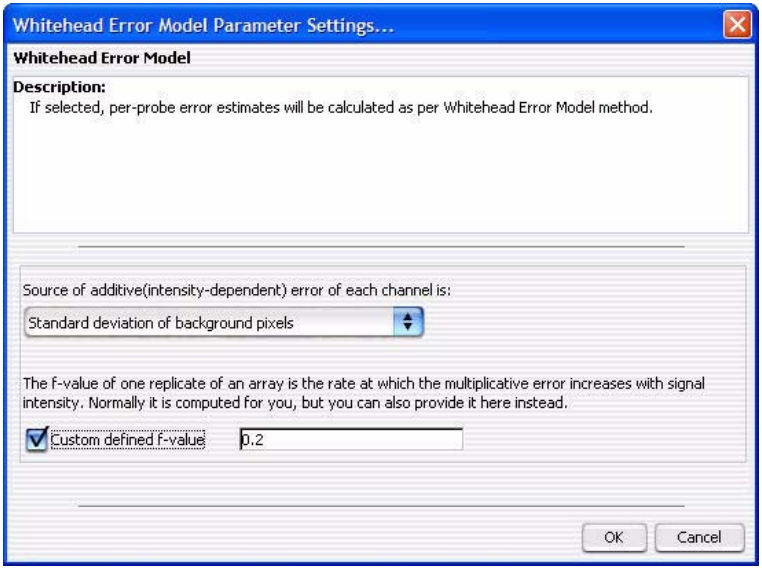


Figure 80 Whitehead Error Model Parameter Settings dialog box

Purpose: This dialog box allows you to 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 69.

Subtract background from signals in each channel

Mark the check box to select this option.

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 define 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.

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

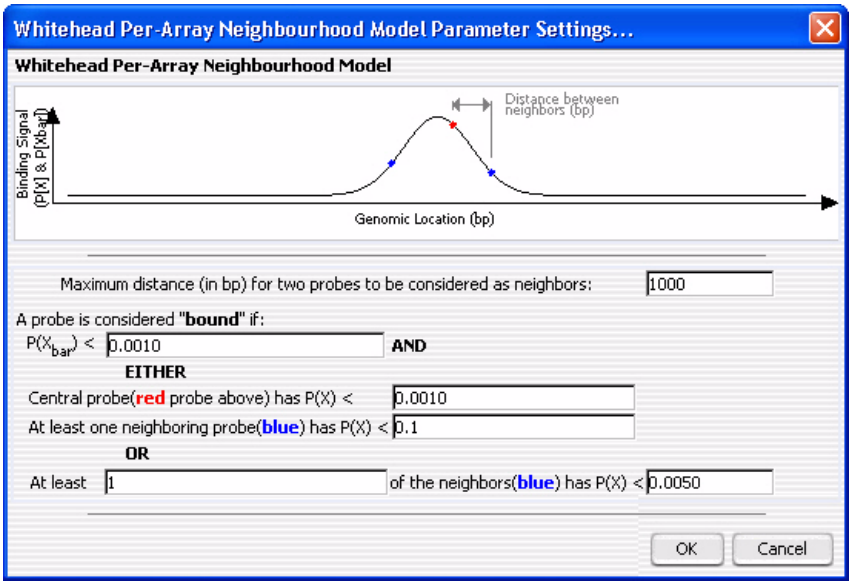


Figure 81 Whitehead Per-Array Neighbourhood Model Parameter Settings dialog box.

Purpose: This dialog box enables you to 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 75.

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 involved in event detection, see "Peak Detection Algorithm" on page 237.

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 81. • 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 81 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 defined in the dialog box

Cancel Closes the dialog box without changing any parameters.

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 (i.e. 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 accessions 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 Gene Names Only** – 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 Gene Names Only** when you configure the report.
- **Standard Report** – A report that contains the columns indicated in the table, below. This is the “default” report that the program generates if you do not mark any additional options.
- **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 Gene Names Only	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 visually asses 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 format 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.

Item	Description
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 details on how the program calculates this value, see "Normalization Algorithms" on page 224.
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 82. 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 83. 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 84. 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 85. 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.

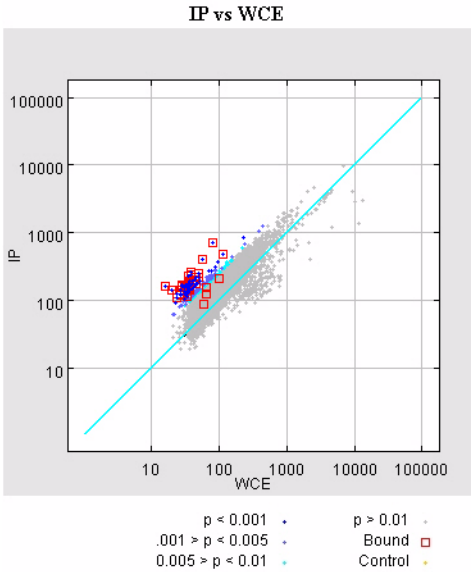


Figure 82 Example scatter plot of IP vs WCE intensities.

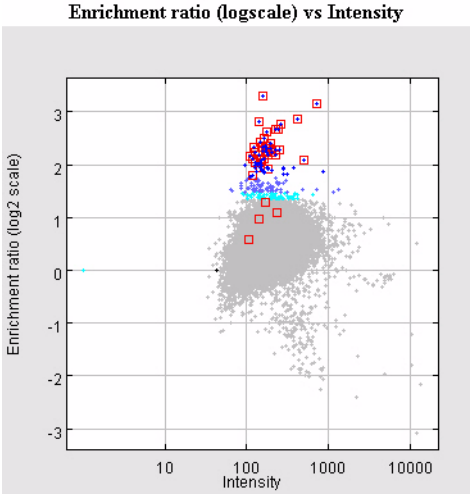


Figure 83 Example scatter plot of enrichment ratio vs intensity. This plot uses the same probe color legend as in Figure 82.

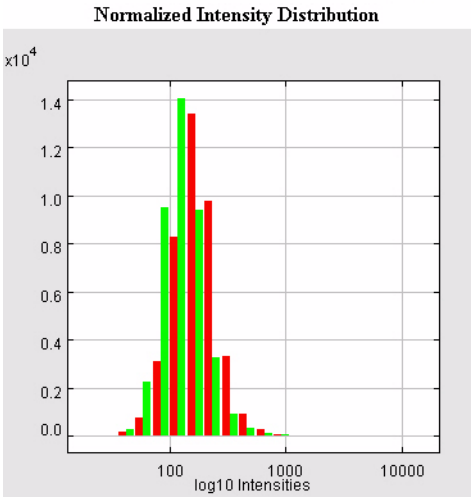


Figure 84 Example histogram of normalized intensity distribution. The distributions appear in a separate color for each channel.

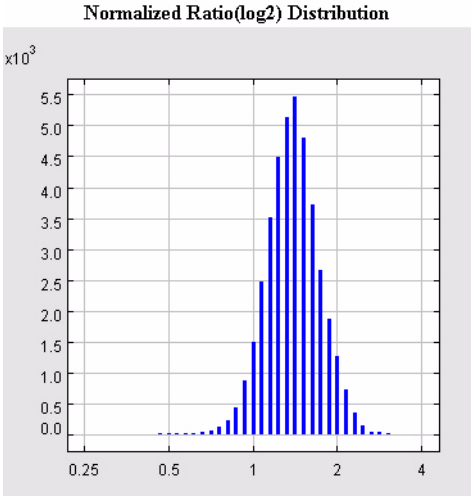


Figure 85 Example histogram of normalized $\log_2$ enrichment ratio distribution.

5 ChIP Statistical Algorithms

This section provides implementation details of the algorithms used in the ChIP application of DNA Analytics. The ChIP application algorithms allow for microarray normalization and facilitate the statistical analysis of regions with an overabundance of probe signal.

If you are planning a Chromatin ImmunoPrecipitation on chip (ChIP-on-chip) analysis, the following pages in this section provide an overview of identification of protein-bound regions in this context. Further contents follow the logical order of an analysis workflow following data import.

The first section compares and contrasts the normalization approaches based on the algorithmic procedures. The second section presents methods to arrive at error estimates. The third section details the peak region detection algorithm used in DNA Analytics.

Peak Detection and Evaluation Overview

ChIP-on-chip is a powerful technology to explore gene regulation activity by determining precise genomic locations where a protein is bound. Phenotypic analysis of ChIP data allows promoter and regulatory network analysis, assessment of toxicogenomic biomarkers, and therapeutic screening. Typical ChIP microarrays have a high number of probes covering both coding and noncoding DNA sequences. DNA Analytics detects robust peaks of probe signal corresponding to the chromatin immunuprecipitation. DNA Analytics also provides genomic visualization and evaluation of the statistical analysis for identifying and evaluating regions of bound DNA in a genomic context.

Whitehead Per-Array Neighbourhood Model

The protein-DNA binding event detection algorithm used in DNA Analytics is the Whitehead Neighbourhood Model. This method analyzes the distribution of all probes on each array to identify robust regions of increased probe signal (termed 'peaks'). The algorithm does this by examining groups of probe triplets which are significantly enriched, yielding robust binding event detection. For example, in a probe-rich region, the area enriched by bound protein may span several probes (see Figure 1 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.

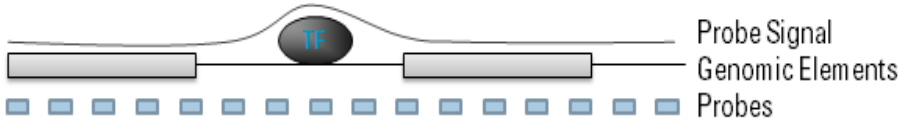


Figure 1 A graphical representation of the probe signal pertebation caused by a protein-DNA binding event such as a transcription factor (TF).

Definitions Used in the Statistical Algorithms Sections

The following abbreviations and mathematical symbols will be utilized in this chapter:

Abbreviation or symbol	Definition
IP, *IP*	In a ChIP on Chip experiment, the signal channel is referred to as the IP or *immunoprecipitated* channel
WCE, *WCE*	In a ChIP on Chip experiment, the background channel is referred to as 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	The DNA Analytics User Interface - refers to algorithms which allow user-defined values to be supplied.
left	Refers to probes closer to the front of a/the chromosome.
right	Refers to probes closer to the end of a/the chromosome.

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, only the true variation from a binding event should be measured. Applying data normalization allows you to limit the systematic variation in the data such that true biological variations are revealed and more readily detected.

DNA Analytics can import data that has already been normalized by Agilent Feature Extraction (FE) software, or it can optionally apply user-defined procedures for both intra- and inter-array normalization. For example, if you wish to utilize the Lowess normalization that FE performs on Agilent two-color data, any data files that have been pre-processed by the FE software do not have to be normalized again in DNA Analytics. For more information about Lowess normalization, refer to the *Agilent Feature Extraction User Guide*.

The normalization scheme provided by DNA Analytics contains both intra- and inter-array procedures. If you are not using the normalizations from FE, 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 present details of each normalization procedure available in DNA Analytics.

Blank Subtraction Normalization

The blank subtraction normalization procedure subtracts a measure of central tendency of the signal from negative control probes from all signals on the array. Doing so gives a more robust estimate of the true signal ratio independent of non-specific binding.

Blank subtraction is a method for transforming 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 choose to use a data transformation method such as blank subtraction, DNA Analytics 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 (version 8).

Options to estimate the central tendency of control (blank) probes:

Median Subtraction

Purpose This normalization procedure corrects for non-specific binding by using all control probes which are known not to 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. This normalization should be applied if there are probes known not to hybridize during an experiment. If the data is noisy or contains non-specific 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 result in negative probe signal values across the array. In such a situation a more robust measure of central tendency should be used, such as the One-Step Tukey Biweight estimate.

Algorithm Corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels are computed by subtracting the background intensities from the foreground intensities. The normalization steps are then applied 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.

One-Step Tukey Biweight Subtraction

Purpose This normalization procedure corrects for non-specific binding by using a weighted set of negative control probes as a baseline measurement for array probe signals.

Use One-Step Tukey Biweight Subtraction uses the Tukey Biweight, also known as the Tukey Bisquare, of the negative control probes as a baseline. The Tukey Biweight is a type of robust measure of central tendency. This normalization should be applied if there is a set of probes known not to hybridize during an experiment. If the data is noisy or contains non-specific binding on the control probe set (yielding outliers) then the Tukey Biweight may be used to avoid negative signal values and outliers. The Tukey Biweight will exclude some of the negative control probes from the blank subtraction. If the set of control probes is small then it may be advantageous to avoid such exclusion. In such a case the median subtraction option may be used.

Algorithm The Tukey Biweight excludes probes from the calculation of the central tendency by their distance from the estimate. For each value, the deviation from the median is normalized by the Median Absolute Deviation (MAD), yielding in effect the median version of a Z-score. This score is used for a measurement of central tendency if it is close to the estimate of central tendency; otherwise it is not used. In such a way outlying probe signals do not affect the blank subtraction and negative normalized probe values are avoided.

Corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels are computed by subtracting the background intensities from the foreground intensities. The normalization steps are then applied 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. Compute the Median Average Deviation (MAD) from the negative control probes on the array.

3 Subtract the median from the signal intensities of every probe on the chip for both channels.

4 For each probe, divide, or normalize, the modified MAD value obtained in **step 4** by the value in **step 3**. Call this value *t*.

5 Obtain the biweight for every value obtained in **step 5** as follows:

$$biweight = \begin{cases} |t| < 1 & (1-t^2)^2 \\ |t| \geq 1 & 0 \end{cases}$$

6 Sum the products of each biweight by each probe signal value.

7 Calculate the measure of central tendency by dividing the sum in **step 7** by the sum of each biweight obtained in **step 6**.

Inter-Array Median Normalization

The Inter-Array Median Normalization algorithm normalizes the median signal across all of the arrays in an experiment, in both channels.

Purpose Microarray experiments are often done with replicates to increase the statistical power of the experiment. 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 preparations, instrumentation, and hybridization. To facilitate the comparison and combination of values from these different arrays, DNA Analytics offers inter-array 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, you should use the Inter-Array Median Normalization algorithm to ensure that comparison between probe values are robust and not due to systematic variation.

Algorithm Before applying the normalization, corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels are computed by subtracting the background intensities from the foreground intensities. The normalization steps are then applied 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, DNA Analytics will use 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

Inter-Array normalizations control for chip-wide variations in intensity. Such variations may be 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 non-linear 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, there is another normalization procedure termed Lowess that applies a non-linear normalization and effectively normalizes by probes and by arrays. The Lowess approach is available in the Agilent Feature Extraction (FE) software and more information can be found in the *Agilent Feature Extraction User Guide*. Importantly, if the data to be analyzed are imported from the FE software using the Lowess approach, then the Intra-Array Median Normalization step need not be applied.

Options to normalize by:

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 and then applies the transform to correct for intensity differences between the dyes used in each channel.

Use Equalizing the IP and WCE channels uses the median of the non-excluded probes in each channel. This normalization should be applied if a two-color experiment is used without previous normalization steps (such as those available in the FE software). Normally this approach works well 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). However, in the case that 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 applying the normalization, corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels are computed by subtracting the background intensities from the foreground intensities. The normalization steps are then applied 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 non-excluded 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 and then applies the transform to correct for intensity differences between the dyes used in each channel.

Use Equalizing the IP and WCE channels uses the median of the non-excluded probes in each channel. This normalization should be applied if a two-color experiment is used without previous normalization steps (such as those available in the FE 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.

Algorithm Before applying the normalization, corrected intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels are computed by subtracting the background intensities from the foreground intensities. The normalization steps are then applied 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 non-excluded 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.

Error Models

Error modelling aids in the separation of true biological events, such as protein-DNA binding interactions, from signals arising from artifacts in the measurement technology. To do this, error modelling is used 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 (FE) software or may be determined by the DNA Analytics 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 over 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' which corresponds to an intensity of 30,000 in one channel over an intensity of 10,000 in the other channel is much more reliable.

The probe signal confidence is reported as a 'p-value'. P-values are a parametric measure of how likely it is to observe a probe signal intensity of equal or greater magnitude, according to the error model. A range from 0 to 1 is used for the significance scoring of a p-value. A general guide to interpreting p-values is that significant events have p-values close to the value 0 (e.g. $p < 0.001$) and events which are likely due to systematic variations have p-values close to the value 1.

DNA Analytics 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 works by selecting robust probe signals for inclusion in the peak detection algorithm and is therefore on by default. The Whitehead error model is generally used as for error modelling in DNA Analytics. However, if your data includes probe quality scores from the FE software, you may choose the option to use the FE error model to see if the biological results are closer to those expected. For more information regarding the FE error model, please refer to the *Agilent Feature Extraction User Guide*.

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

Intensities for both the immunoprecipitated (IP) and whole cell extract (WCE) channels are computed according to the normalization options described in the previous sections:

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

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

The normalized values are those obtained after applying any normalization steps (as described in the preceding section). 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 both the IP and WCE channels. This measure of variance may be calculated from one of the following:
   - **a** The standard deviation of background pixels. This is the default option. In such cases the error model will be based on the overall performance of the microarray.
   - **b** The observed spread of the negative controls. Estimating the variance of negative control probes gives a good indication of the error inherited by every probe on the array due to technological limitations, but it has not been tested for all situations.
   - **c** The additive error as computed by the Agilent Feature Extraction (FE) software. For more information, refer to the *Agilent Feature Extraction User Guide*.
2. Calculate the value f to as a factor to adjust the symmetry of the distribution of X. This value may be user defined or may be calculated automatically by DNA Analytics. To calculate the value of f, DNA

Analytics performs an iterative search of all values of *f* such that the difference between the values for X<0 for probes in the highest and lowest quartile are minimized:

a Create an array *A* that includes the values from all non-excluded, non-control probes in the bottom 25 percentile by WCE with an X value less than 0.

b Create an array *A'* that includes the values from all non-excluded, non-control probes in the top 25 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'* by 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 an X value as high as its own or higher, 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 DNA Analytics supports the use of multiple biological or technical replicates. The error model is extended 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 ration, 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 Algorithm

DNA Analytics detects robust peaks of probe signal corresponding to binding events. These probe signals are typically normalized and an error model is applied, 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. DNA Analytics also provides genomic visualization and evaluation of the statistical analysis. Agilent has partnered with the Whitehead Institute to provide a technology solution in DNA Analytics for identifying and evaluating regions of bound DNA in a genomic context.

Whitehead Per-Array Neighbourhood Model

The protein-DNA binding event detection algorithm used in DNA Analytics is the Whitehead Per-Array Neighbourhood Model. This method uses the distribution of all probes on each array to compute robust regions of increased probe signal (termed 'peaks'). The algorithm does this by examining groups of probe triplets which are significantly enriched, yielding robust binding event detection. 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. See Figure 2 on page 237.

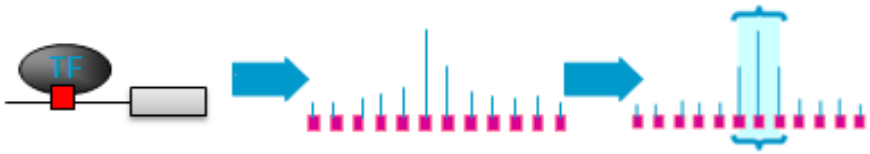


Figure 2 A graphical representation of the probe signal 'neighborhood' used in the Whitehead model. The diagrams shows the binding event as a transcription factor (TF) and the resultant signals from discrete probes in the binding area region. The peak detection algorithm uses probes in the immediate vicinity of the probe signal (here represented as a triplet) to evaluate whether or not the local increase in signal intensity is likely to represent a binding event.

Purpose Typical ChIP-on-chip microarrays use discrete probes covering both coding and non-coding genomic DNA. Because a protein-DNA binding event may span regions described by several distinct probes, a strategy to aggregate and evaluate such probe signals into a robust region of increased signal intensity is required. Such regions are term “peaks”. Peak detection is a method to identify binding events from complex genomic data.

Use Following normalization of the data and the application of an error model, the data analysis typically shifts from “low-level” analysis to “high-level” analysis. In the low-level analysis, statistics are applied at the probe level in order to ensure that systematic variations are accounted for and quality information about probes is correlated with signal intensities. In the high-level analysis, biological events are catalogued, statistically examined, and collected into information about biological annotation, pathway, and network. The peak detection algorithm in DNA Analytics provides a bridge between low-level and high-level analysis, thereby enabling complete data analysis workflow for ChIP-on-chip experiments.

Algorithm: $P(X_{neighb})$ DNA Analytics utilizes two types of p-values and then employs a heuristic that uses both values to predict whether or not 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 described earlier in this chapter. P(X) represents the error-corrected significance of the probe ratio. This value may be calculated directly by DNA Analytics using the distribution of probe signals on each array or may be imported from the Agilent Feature Extraction (FE) software by choosing the FE 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. This value is computed 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, signals are expected to span several hundreds of bases, due to the size of the DNA fragments produced by shearing (see Figure 2 on page 237). 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.

This corroboration is computed in terms of X_{neighb}, an extension of the p-value calculated as part of the Whitehead or FE error model. The value X_{neighb} is a simple average of the X value of a given probe with its genomic neighbors. A distance threshold is applied 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, a value of X=0 is used 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 evaluating the area under the normal Gaussian curve from *Z* to infinity.

NOTE P(X) is the area under the normal curve from Z to infinite. The distribution of values for X are presumed to be Gaussian.

Whether a set of enriched probes should be considered as identifying a genomic location at which a protein is bound is subject to several user-defined parameters available in the user interface (UI).

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 is considered as identifying the location of a “bound” protein if either of the following conditions user 1 or 2 are met:

1. The calculated p-value termed P(Xbar) is less than a threshold defined in the UI (default: 0.001) AND:
   a. The error model p-value termed P(X) for the central probe is less than a threshold defined in the UI (default: 0.001) OR:
   b. The error model p-value termed P(X) for at least one neighboring probe is less than a threshold defined in the UI (default: 0.1).
2. The calculated p-value termed P(Xbar) is less than a threshold defined in the UI (default: 0.001) AND:
   a. A number of the neighbors defined in the UI (default: 1) has an error model p-value termed P(X) less than a threshold defined in the UI (default: 0.005).

The decision tree for calling a probe “bound” can be summed in the following table and diagram:

Table 1 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 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 3. • 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 3 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 • The default cut-off value for P(X) is 0.005 • To make detection more stringent, decrease the value.

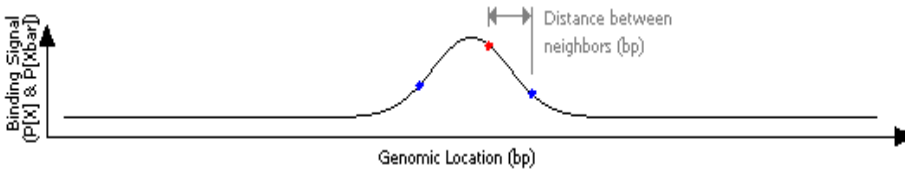


Figure 3 A graphical representation of the probe signal 'neighborhood' used in the Whitehead model.

Interpretation Probe binding ends when all neighboring probes that pass the condition tree determined from the user-defined parameters are found. The average signal intensity found from bound probes in this process are reported, and

a plot is generated as output. The binding events are rendered as points in the visualization panel. The height of each point is equal to the average log ratio of that region.

Visualization When the binding regions are computed, they are plotted 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. By clicking on a chromosomal region of interest or by dragging a rectangular selection around a genomic region you can easily zoom in on that region.

NOTE In version 4.0 of the software, graphs are drawn in log2-space. This has the effects of exaggerating the baseline noise and compressing strong signals. Moreover, ratios above 4-fold (or 8-fold, depending on your settings) will be rendered as 4-fold (or 8-fold). This can make it difficult to observe strong peaks visually in the software. This is a known issue and will be resolved in future releases.

www.agilent.com

In This Book

The User Guide presents instructions on how to analyze your ChIP microarray data with the Agilent Genomic Workbench ChIP application software.

First Edition, April 2009

Agilent Technologies