Agilent
Genomic Workbench 6.0

SureSelect Quality Analyzer

User Guide

Agilent Technologies

Notices

Revision

G3800-90003

Revision A, February 2010

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

Trademarks

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

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

Software Revision

This guide is valid for 6.0 and later revisions of the Agilent Genomic Workbench software, until superseded.

Warranty

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

Technology Licenses

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

Restricted Rights Legend

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

Safety Notices

CAUTION

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

WARNING

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

In This Guide...

This guide describes how to use the SureSelect Quality Analyzer program that is part of the Agilent Genomic Workbench Standard Edition suite.

1 Getting Started

This chapter describes SureSelect Quality Analyzer and how it fits into the typical SureSelect Target Enrichment research workflow. It gives instructions on how to start the program, and describes the main program window. It describes the files that you must have before you can use the program, and includes an example exercise that leads you through a typical SureSelect QC analysis experiment. This chapter also tells you how you can get additional help.

2 Using SureSelect Quality Analyzer

This chapter describes how to import sequence read and target interval files, and how to organize these files into analysis experiments. It describes how to display and analyze the reads and target intervals, and how to create and display QC reports of several types, including summary QC metrics, and reports on read depth and enrichment analyses.

3 SureSelect Quality Analyzer Reference

This chapter describes the tabs, panes, commands, menus, shortcut menus, and dialog boxes that can appear when you use SureSelect Quality Analyzer. It also contains reference information on the reports that the program creates.

Contents

Contents

1
Getting Started

This chapter describes SureSelect Quality Analyzer and how it fits into the typical SureSelect Target Enrichment research workflow. It gives instructions on how to start the program, and describes the main program window. It describes the files that you must have before you can use the program, and includes an example exercise that leads you through a typical SureSelect QC analysis experiment. This chapter also tells you how you can get additional help.

SureSelect Quality Analyzer

SureSelect Quality Analyzer is a program that lets you assess the effectiveness of the pull-down of targeted genomic fragments when you use the Agilent SureSelect Target Enrichment system. This system uses libraries of biotinylated RNA oligonucleotide “baits” to harvest genomic DNA fragments of interest for sequencing. This forms the basis of a powerful selection method that lets you focus your sequencing efforts.

The selective nature of the process makes it ideal for targeted resequencing that uses next-generation sequencing technology. After you sequence the harvested fragments, you can use SureSelect Quality Analyzer to calculate statistical metrics and analyze read depth. You can also do an enrichment analysis, and display the results in the UCSC Genome Browser next to the annotation tracks of your choice.

SureSelect Quality Analyzer is part of the Agilent Genomic Workbench Standard Edition suite, a collection of essential programs that supplies a single solution for all of your genomics data processing needs. Figure 1 shows a typical sequencing workflow that uses the Agilent SureSelect Target Enrichment system.

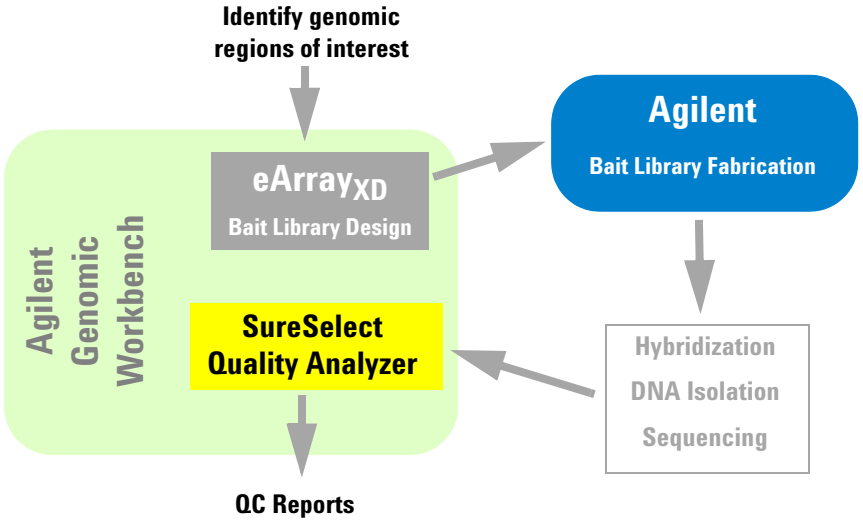


Figure 1 Typical sequencing workflow with the SureSelect Target Enrichment system

After you identify the genomic regions of interest, you use the eArray$_{XD}$ program within Agilent Genomic Workbench to design a library of oligonucleotide baits to the desired regions. You submit the library to Agilent, who supplies the biotinylated RNA oligonucleotides in the library to you in any desired quantity. You use the bait library to enrich your sample DNA for the desired target genomic regions, and sequence the resulting DNA fragments.

You import the sequence reads into Agilent Genomic Workbench, and use SureSelect Quality Analyzer to do several crucial quality-related analyses:

- **QC metrics** – Includes overall statistics, such as the percentage of reads in targeted regions, the overall fold enrichment, and the average read depth.
- **Read depth analysis** – Shows the overall distribution of reads over the targeted regions and targeted bases. Distributions are available both as tables and as graphs.

- **Enrichment analysis** – Shows the degree of enrichment for target intervals in the genome of interest. The program creates a WIG format file, and uploads it to the UCSC Genome Browser, where it appears graphically in the User Track area. For information on the WIG file format, go to genome.ucsc.edu/FAQ/FAQformat.

  You can also create a text version of the enrichment analysis results as a *.xls file.

Agilent Genomic Workbench

To use SureSelect Quality Analyzer, you must set up all parts of Agilent Genomic Workbench Standard Edition. The system has three main components, illustrated in Figure 2.

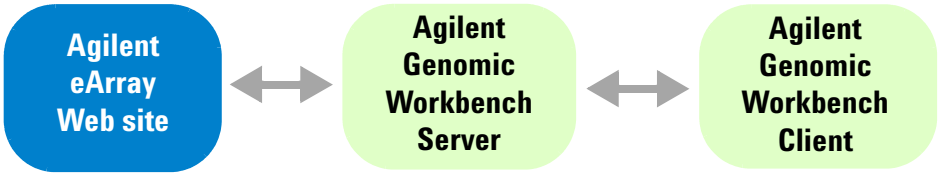


Figure 2 Agilent Genomic Workbench Standard Edition – Main components

- **Agilent Genomic Workbench Client** – You install this program on your computer. It contains all of the parts of Agilent Genomic Workbench with which you have direct interaction, including SureSelect Quality Analyzer. It communicates as necessary with your Agilent Genomic Workbench server.
- **Agilent Genomic Workbench Server** – You install this software on one machine in your workgroup. It contains a shared database of workgroup content, as well as utilities that coordinate the flow of data between the components of the system. Many users, who each run the client program, can connect to this server.

- **Agilent eArray Web site** – A Web site that contains a large database of microarray-related content (probes, probe groups, and microarrays), as well as Target Enrichment-related content (baits, bait groups, and libraries). The site, located at https://earray.chem.agilent.com, includes both Agilent Catalog content as well as custom content that your workgroup may have created. It also contains a full set of tools that can be used for the design of custom microarrays and bait libraries.

  Your Agilent Genomic Workbench server communicates with the eArray Web site to transfer data and to submit and retrieve jobs and results. Agilent Genomic Workbench Standard Edition includes eArray$_{XD}$, a program that lets you use most eArray functionality on your desktop, and lets you access Agilent content and certain tools on the eArray Web site from within the program.

  Before you can use the other components of the Agilent Genomic Workbench suite, your workgroup must be registered on the eArray Web site, and you must be a registered user. In addition, when you first install the Agilent Genomic Workbench server software, you must wait for the eArray Web site to transfer the custom content of your workgroup to your server.

To use Agilent Genomic Workbench Standard Edition, you must set up all parts of the system. However, SureSelect Quality Analyzer runs almost entirely within the client program on your computer. By default, sequence read data, as well as analysis results are stored locally, and are not accessible to other users.

Setting up SureSelect Quality Analyzer

This section describes what you must do before you can use SureSelect Quality Analyzer, and explains how to start the program. It also includes a description of the main program window as it appears for the SureSelect Target Enrichment application type.

Before you use SureSelect Quality Analyzer

- You must be a registered user on the eArray Web site. For details, go to https://earray.chem.agilent.com, and click **Help.** Also, the Agilent Genomic Workbench client software that is installed on your computer must be linked to this user account. See the *eArray$_{XD}$ User Guide*.
- You must have several types of files available before you can use SureSelect Quality Analyzer:
  - **Sequence read file** – A *.zip archive that contains the nucleotide sequences and locations of the fragments in your DNA sample, along with related header information. The program supports these read file types:

File Type	Comments
Illumina GA Export	Revised version of the ELAND file format that contains additional fields, including quality scores for all bases.
AB SOLiD ma	Color space *.ma file produced by the Applied Biosystems SOLiD system. The AB SOLiD alignment software produces this CSFASTA format file, and places sequence alignment information in the FASTA headers. Color calls are converted to bases when the file is imported.
MAQ aln.txt	The aln.txt output file from the MAQ (Mapping and Assembly with Qualities) program. This program is specifically designed for Illumina-Solexa/AB-SOLiD reads. For more information, go to maq.sourceforge.net

NOTE Import sequence read files as *.zip archives. Do not use the gzip format.

- **Chromosome mapping file** – A *.txt file that contains the names of the chromosomes that are referenced in the read file, correlated with the names of the chromosomes in the desired genome build of your species of interest. For details, see "To import sequence read files" on page 30.
- **Target intervals file** – A *.bed format track file that contains a list of the target genomic intervals. Typically, these are the target genomic intervals to which you designed your SureSelect Target Enrichment baits. If you used eArray to create the bait library, you can download this file from eArray when you download the library. You can also create the target intervals file manually.

To start SureSelect Quality Analyzer

SureSelect Quality Analyzer is part of Agilent Genomic Workbench. All programs in the suite run within a single user interface.

1 Start Agilent Genomic Workbench Standard Edition.

The Home tab appears.

2 Click **Switch Application,** then select **SureSelect Target Enrichment.** For details on the other options that are available, see "To set the application type" on page 18.

The tabs and content that are relevant to this application type become available.

3 Click the **Quality Analyzer** tab.

The Quality Analyzer tab appears. See Figure 3.

Figure 3 Agilent Genomic Workbench main window – Quality Analyzer tab

The Quality Analyzer tab contains several main components:

Component	Description
Tab bar	Contains the tabs that let you use the main functions of the program. See "Agilent Genomic Workbench tab bar" on page 54. All commands for SureSelect Quality Analyzer can be found in the Quality Analyzer tab.
Switch Application menu	Lets you select the experimental application type. This is one of the main settings for Agilent Genomic Workbench, since functionality as well as design, sample, and sequence data are partitioned by application type. See "Switch Application menu" on page 55 and "To set the application type" on page 18. To use SureSelect Quality Analyzer, you must set the application type to **SureSelect Target Enrichment.** See "To set the application type" on page 18.
Command ribbon	Contains the commands that you use to access the main functionality of Agilent Genomic Workbench. A different command ribbon appears for each main tab. The available commands also vary by application type. For specific information on the commands that appear in the Quality Analyzer tab, see "Command ribbon – Quality Analyzer tab" on page 58.
Navigator	Contains lists of the designs, sample data, sequence data, analysis experiments, tasks, and other related items that are available to you. A search feature is also available. See "Navigator overview" on page 63. When you use SureSelect Quality Analyzer, you mainly use the Experiment pane of the Navigator. See "Navigator – Experiment pane" on page 64.
Genomic Viewer	Shows the genome of the selected organism at several levels of detail. For the different application types and programs in Agilent Genomic Workbench, genes, tracks, probes, baits, and data of many types can appear in the Genomic Viewer in the context of the genome. For general information on how to use the Genomic Viewer, see the *Data Viewing User Guide.* SureSelect Quality Analyzer uses the Genomic Viewer to display reads and target intervals. Several different panes appear. See "Genomic Viewer – overview" on page 67.

To set the application type

In Agilent Genomic Workbench, content as well as function are partitioned by experimental application type. When you select an application type, only the tools and content that are relevant to the selected type are accessible. The content and functionality for SureSelect Quality Analyzer are only available when you select **SureSelect Target Enrichment** as the application type.

1 Start Agilent Genomic Workbench.

2 Click  **Switch Application.**

3 Select the desired application type.

The selected experimental application type appears in Application Type. The tools and content relevant to this application type become available.

You can use the tools in Agilent Genomic Workbench to work with microarray and library design content, experimental data, and other related files for many different application types. The program supports these application types:

Application Type	Description
CGH	Design of microarrays and analysis of data for comparative genomic hybridization studies
ChIP-on-chip	Design of microarrays and analysis of data for chromatin immunoprecipitation studies
CH3	Design of microarrays and analysis of data for methylation studies
Expression	Design of microarrays for gene expression studies
microRNA	Design of microarrays for microRNA studies
SureSelect Target Enrichment	Design of bait libraries for the retrieval of specific DNA fragments for sequencing, and QC analysis of target fragment pull-down

To select where data and results are stored

By default, sequence data and analysis results are stored locally on your computer in the **data** folder, within the folder that contains the Agilent Genomic Workbench client program. You can change this location.

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

The User Preferences dialog box appears.

2 Click the **Miscellaneous** tab.

The Miscellaneous tab appears. See "User Preferences – Miscellaneous tab" on page 93.

3 In Data Location, click **Browse.**

An Open dialog box appears.

4 Select the desired location, then click **Open.**

The selected location appears in Data Location.

5 Click **OK.**

Example exercise

In this example exercise, you import a sequence read file and other supporting files. You create and view a PDF format QC report that contains summary QC statistics and the results of read depth analysis, and you also use the UCSC Genome Browser to view the results of enrichment analysis. In addition, you view sequence reads and target intervals in the Genomic Viewer.

Before you start the example exercise

- Your workgroup must be registered on the eArray Web site, and you must be a registered user on the site. For more information, see the online help at earray.chem.agilent.com. Also, the Agilent Genomic Workbench client software that is installed on your computer must be linked to this user account. See the *eArray$_{XD}$ User Guide*.
- Download the example files for this exercise. Follow these steps:
  - **a** Go to earray.chem.agilent.com.

    The login page of the eArray Web site appears. You do not need to log in.
  - **b** Under Additional Information, click **Download Agilent Genomic Workbench Standard Edition.**

    A page of available Agilent Genomic Workbench files appears.
  - **c** Under Download the Agilent Genomic Workbench Standard Edition 6.0 Client Software, next to Sample files for SureSelect Target Enrichment Quality Analyzer, right-click **Download,** then click **Save Target As.**
  - **d** Save the file to your desktop.

    The eArray Web site transfers a zip format file to your computer.
  - **e** Extract the zip file to your desktop.

    The folder of example files appears on your desktop. Do not extract the zip format files within this folder.
- Familiarize yourself with the parts of the Agilent Genomic Workbench interface. See "To start SureSelect Quality Analyzer" on page 15 and "Genomic Viewer – overview" on page 67.

To import read data and create QC reports

The main steps in this example exercise appear in the first column of the table below. More detailed instructions appear in the second column. Additional notes and comments appear in the third column.

Example Exercise

Step	Detailed Instructions	Notes/Comments
1 Start SureSelect Quality Analyzer	**a** Start **Agilent Genomic Workbench 6.0.** The main window of Agilent Genomic Workbench appears. **b** Click **Switch Application > SureSelect Target Enrichment.** SureSelect Target Enrichment appears in Application Type. The appropriate tools and panes appear in the main window of the program. **c** Click **Quality Analyzer.** The Quality Analyzer tab appears.	• One or more dialog boxes can appear when you first start the program. You must select the appropriate options and/or type the requested information before you can use Agilent Genomic Workbench. For more information, see "To start SureSelect Quality Analyzer" on page 15. • You can use SureSelect Quality Analyzer free of charge. You do not need to purchase or enter a license.
2 Import a read file and a chromosome mapping file	**a** Click **Import > Reads.** A dialog box appears. **b** In the folder of sample files that you extracted to your desktop, select the file **s_1_1_export_top.zip.** **c** Click **Open.** The Specify Input Information dialog box appears. **d** Under Read Info, in **File Type,** select **Illumina GA Export.** **e** Under Map Chromosomes, in Chr Mapping File, click **Browse.** An Open dialog box appears. **f** In the folder of sample files that you extracted to your desktop, select the file **chrMapping.txt,** then click **Open.** The name of the selected file appears in Chr Mapping File. **g** Click **Import.** A progress bar appears.	• In this step, you import these files: ◦ **Read file** – Contains header information, and the sequence data to be analyzed. The program supports read files in several formats. See "To import sequence read files" on page 30. ◦ **Chromosome mapping file** – A *.txt file that associates the names of chromosomes in the read file with the names of chromosomes as they are specified in the selected genome build. • Under Species Info, keep Select Species set to **H. sapiens,** and Select Genome Build set to **hg18.**

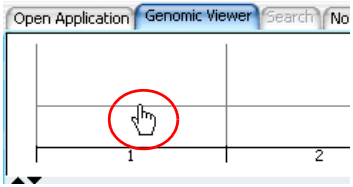
Example Exercise

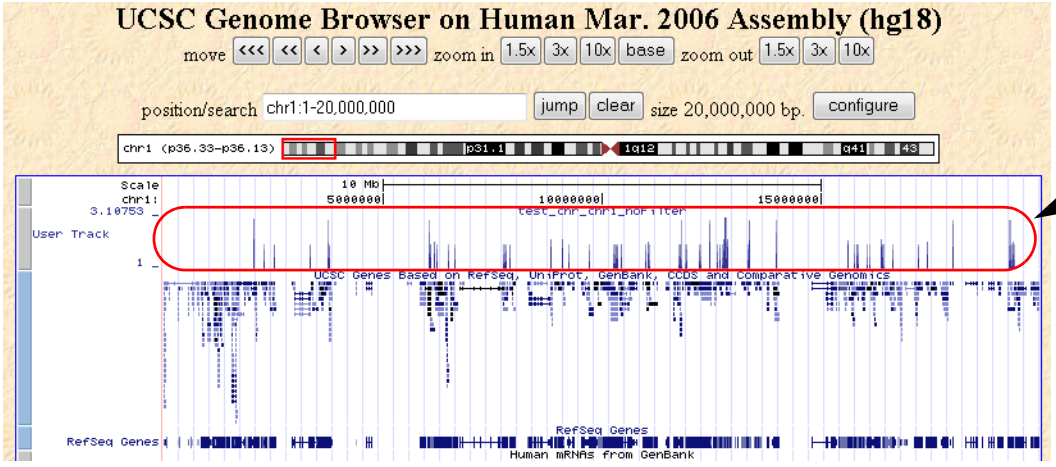
Step	Detailed Instructions	Notes/Comments
3 Import a target intervals file	**a** Click **Import > Track.** The Import Track dialog box appears. **b** In **Track Name,** type `Example Targets`. **c** In **Track File,** click **Browse.** An Open dialog box appears. **d** In the folder of sample files that you extracted to your desktop, select **SureSelect_All_Exon_G3362_with_names.bed,** then click **Open.** The location of the file appears in Track File. **e** Click **OK.** A dialog box tells you that the track was successfully imported. **f** Click **OK.**	• In this step, you import a BED format track file that contains the desired target genomic intervals. • When you select a track file to import, the program only displays *.bed format files.
4 Create a QC analysis experiment	**a** In the Experiment pane of the Navigator, right-click the **Experiments** node, then click **New Experiment.** The Create Experiment dialog box appears. **b** Enter the following information: • **Name** – Type `Example Expt` • **Read Length** – Delete the existing value, then type `76` • **Targets** – Select **Example Targets** • **SEQ Sample** – Select **s_1_1_export_top.** **c** Click **OK.** A progress bar appears. In the Experiment pane of the Navigator, a new experiment with the name **Example Expt** appears.	• In SureSelect Quality Analyzer, an experiment is a required organizational level that relates a specific read file to a specific set of target intervals.

Example Exercise

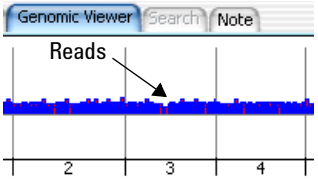
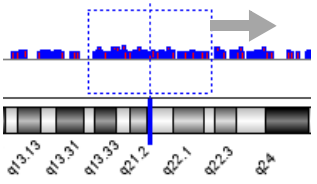
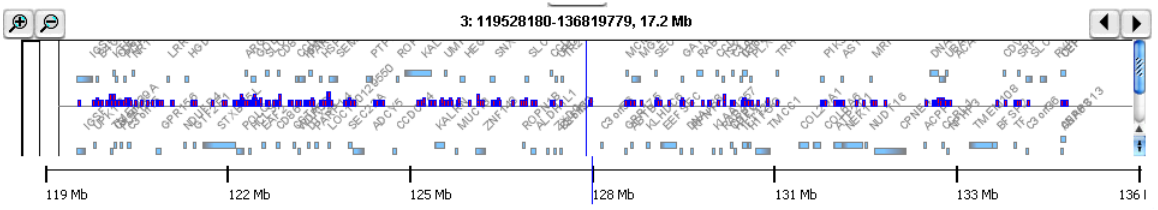
Step	Detailed Instructions	Notes/Comments
5 Select the experiment for analysis	**a** In the Experiment pane of the Navigator, in the Experiments folder, right-click **Example Expt,** then click **Select Experiment.** A dialog box asks if you want to select the experiment. **b** Click **Yes.** A progress bar appears. A list of reads appears in Tab View of the Genomic Viewer.	• Later in this example exercise, you view the reads graphically in the context of the human genome. • In the Experiment pane of the Navigator, the name of the selected experiment appears in blue. • The reads that appear in Tab View are the ones that are found on the chromosome that appears in Chromosome View.
6 Calculate QC metrics and analyze read-depth and enrichment	**a** Click **Create Report.** A progress bar appears. A dialog box tells you that reports were successfully created. **b** Click **OK.**	• In this step, the program analyzes the sequence data against the intervals in your selected target intervals file. You view the results in subsequent steps.
7 Create and view a PDF format QC Report	**a** Click **Summary Report > Save as PDF.** A Save dialog box appears **b** Select a location, and type a name for the report file, then click **Save.** A dialog box tells you that the file was successfully created, and asks if you would like to view it. **c** Click **Yes.** The QC Report opens in Adobe Reader.	• In this step, you create a PDF format report that contains calculated QC metrics and the results of read depth analysis. For details, see: ◦ "To view and save reports" on page 49 ◦ "QC metrics" on page 101 ◦ "Read depth distribution" on page 103 • You can also save the report as a text file. See "To view and save reports" on page 49.

Example Exercise

Step	Detailed Instructions	Notes/Comments
8 Display enrichment analysis results for chromosome 1 in the UCSC Genome Browser	**a** In the Genomic Viewer, in Genome View, click the center of the Chromosome 1 area.  **b** Click **Enrichment > View in UCSC.** Your Web browser opens. **c** If a security message tells you that the browser has restricted scripts or ActiveX controls, follow the on-screen instructions to allow the blocked content. The UCSC Genome Browser opens in your Web browser. **d** In **position/search,** delete the existing chromosomal location, then type `chr1:1-20,000,000` **e** Click **jump.** Enriched peaks appear in the User Track area.	• In this step, you transfer enrichment analysis results for chromosome 1 to the UCSC Genome Browser, and view the enriched peaks in the first 20 Mb of that chromosome. • You can display the enriched peaks next to your choice of annotation tracks. For instructions, see the online help that is available in the UCSC Genome Browser. • You can also save the enrichment analysis results as a *.xls file. See "To view and save reports" on page 49.

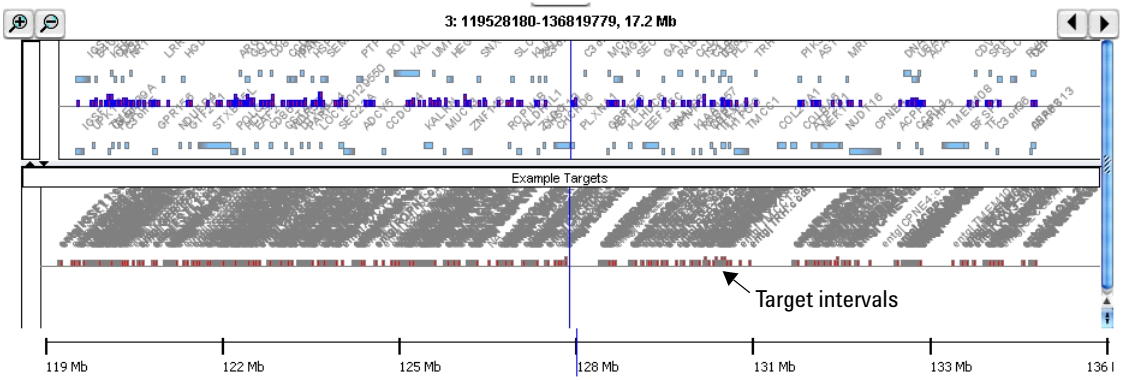


Example Exercise

Step	Detailed Instructions	Notes/Comments
9 View the reads in the selected experiment.	**a** In the **View** tab, under Custom Data, mark **Show.** The reads in the experiment appear in the Genomic Viewer, in Genome View.  **b** In Genome View, click the reads on chromosome 3. Chromosome 3 appears in Chromosome View. **c** Drag the pointer across some of the reads in Chromosome View.  The reads appear in Gene View.  	In this step, you use the Genomic Viewer to view the reads on chromosome 3 from the experiment. For more information on the features of the Genomic Viewer, see: • "To view the data in an experiment" on page 39 • "To view the intervals in a target interval track" on page 42 • "Genomic Viewer – overview" on page 67 • "Genomic Viewer – Genome View" on page 69 • "Genomic Viewer – Chromosome View" on page 71 • "Genomic Viewer – Gene View" on page 73 • "Genomic Viewer – Tab View" on page 75 • The *Data Viewing User Guide* A list of the reads on chromosome 3 appears in Tab View.

Example Exercise

Step	Detailed Instructions	Notes/Comments
10 View the target intervals that have been defined for the experiment	**a** In Chromosome View, drag the pointer across several reads. The reads appear in Gene View. **b** Drag the lower border of Gene View downward until Gene View is approximately twice its original size.	• The experiment for this exercise contains many target intervals. To zoom Gene View to see the boundaries of specific intervals in greater detail, click .



Getting help

To view other Agilent Genomic Workbench user guides

You can view all of the user guides that are available for Agilent Genomic Workbench from within the program. The user guides that are available vary by application type.

1 Select the desired application type. See "To set the application type" on page 18. To view the user guides that are relevant to the Quality Analyzer tab and the other tabs that are available for the SureSelect Target Enrichment application type, set the application type to **SureSelect Target Enrichment.**

2 Click **Help.**

Buttons for the available user guides for the selected application type appear in the command ribbon. See "Command ribbon – Help tab" on page 60.

3 Click the button for the desired user guide.

The selected user guide opens in Adobe Reader.

To contact Agilent Technical Support

If you have an issue or question that is not addressed in one of the user guides, you can contact Agilent Technical Support. Contact information for each region appears in the table below.

Region	Technical support contact information
North America	Telephone: (800) 227 9770 E-mail: pdl-earraysupport_afo@agilent.com
Europe	E-mail: pdl-earraysupport_efo@agilent.com
Asia Pacific	E-mail: pdl-earraysupport_apfo@agilent.com
Africa	E-mail: pdl-earraysupport_efo@agilent.com
Middle East	E-mail: pdl-earraysupport_efo@agilent.com

2
Using SureSelect Quality Analyzer

This chapter gives detailed instructions on how to assess the effectiveness of the pull-down of DNA fragments when you use the SureSelect Target Enrichment system. It describes how to import the necessary files into the program, and how to organize the files into *experiments* for analysis. It also describes how to create and export SureSelect QC reports of several different types. In addition, it describes how to use the Genomic Viewer in Agilent Genomic Workbench to display reads and target intervals in the context of your genome of interest.

Importing files

You must import a sequence read file, a chromosome mapping file, and a target intervals file before the program can calculate QC metrics and analyze read depth and enrichment. If the applicable genome build for the species of interest is not available in the program, you must also import genome build files. Refer to these topics in this section:

Type of file	See this topic
Sequence read	"To import sequence read files" on page 30
Chromosome mapping	"To import sequence read files" on page 30
Target intervals	"To import target genomic regions as a track" on page 33
Genome build	"To import a genome build" on page 35

To import sequence read files

Sequence read files are *.zip archives that contain the base sequences of the fragments in your enriched DNA sample, along with related header information. The program supports these read file formats:

File Type	Comments
Illumina GA Export	Revised version of the ELAND file format that contains additional fields, including quality scores for all bases.
AB SOLiD ma	Color space *.ma file produced by the Applied Biosystems SOLiD system. The AB SOLiD alignment software produces this CSFASTA format file, and places sequence alignment information in the FASTA headers. Color calls are converted to bases when the file is imported.
MAQ aln.txt	The aln.txt output file from the MAQ (Mapping and Assembly with Qualities) program. This program is specifically designed for Illumina-Solexa/AB-SOLiD reads. For more information, go to maq.sourceforge.net

NOTE Import sequence read files as *.zip archives. Do not use the gzip format.

In addition, you must have a *.txt file that maps the names of chromosomes in your sequence read file to the names of the chromosomes in the desired genome build of the species of interest.

1 Set the application type to **SureSelect Target Enrichment.** See "To set the application type" on page 18.

2 In the Quality Analyzer tab, click **Import > Reads.**

A dialog box appears.

3 Select the desired read file, then click **Open.**

The Specify Input Information dialog box appears. See "Specify Input Information" on page 87.

4 Enter the following information:

Group	Instructions/Details
Read Info	The program displays the name of your read file, and the Seq (sequence read file) input type. • In **File Type,** select the specific file format that best represents the data in the file.
Species Info	Select the most appropriate species and genome build. The sequence data in your imported read file must be linked to a specific species and genome build. If the desired genome build does not appear, you must import it. See "To import a genome build" on page 35.

Group	Instructions/Details

Map Chromosomes	You must import a file that maps the names of chromosomes in the read file to the names of chromosomes that are defined in the selected genome build.
	a Create a tab-delimited text file that contains the desired mapping information. Enter one chromosome per line. Give the name of the chromosome in the read file followed by the name of the matching chromosome in the genome build, separated by a tab character. Do not include a header row.
	Example file content:
	`chrA.fa chrA`
	`chrB.fa chrB`
	`chrC.fa chrC`
	b In Chr Mapping File, click **Browse.** An Open dialog box appears.
	c Select the desired chromosome mapping file, then click **Open.** The name of the selected file appears in Chr Mapping File.

5 Click **Import.**

The program imports the read file and the chromosome mapping file. A progress bar appears.

Before you can analyze your sequence data, you must also import a BED format track file that defines the target genomic regions. See "To import target genomic regions as a track" on page 33.

To import target genomic regions as a track

To analyze the sequence data in a read file, the program must correlate reads with target genomic intervals. Typically, these intervals are those to which you have designed you SureSelect Target Enrichment baits. To define the target intervals, you import them as a BED format track file. If you used eArray to create baits, you can download the library as a BED format file, and use this file directly in SureSelect Quality Analyzer. You can also create the file manually.

1 Set the application type to **SureSelect Target Enrichment.** See "To set the application type" on page 18.

2 In the Quality Analyzer tab, click **Import > Track.**

The Import Track dialog box appears. See "Import Track" on page 86.

3 Enter the following information:

Parameter	Instructions/Details
Species	Select the species that is associated with the genomic intervals in your BED format track file.
Build Name	Select the genome build that is associated with the genomic intervals in your BED format track file.
Color	(Optional) To customize the display color of the intervals in the target intervals track, follow these steps: **a** Click **Change.** The Choose Track Color Dialog box appears. **b** In the Swatches tab, select the desired color, then click **OK.** You can also use the HSB and RGB tabs to define or adjust the color using the HSB and RGB color scales, respectively. See "Choose Track Color" on page 77.
Track Name	Type a name for the track as you want it to appear in reports and lists, and in Gene View of the Genomic Viewer.
Track File	The track file must be a BED format track file that contains the target genomic intervals. **a** Click **Browse.** An Open dialog box appears. **b** Select the desired *.bed track file, then click **Open.** The name of the selected file appears in Track File.

4 Click **OK.**

A dialog box tells you that the track was imported successfully.

5 Click **OK.**

NOTE

- The name, species, genome build, description, and display color of the track can only be entered as you import the track. You cannot edit this information later.
- You can combine tracks to create a single merged track. Import the individual tracks, then export all of the desired tracks as a single *.zip file. See "To export tracks" on page 42. You can then import this merged file as the target intervals file.

To import a genome build

If you want to import a read file, but the genome build that applies to the reads is not available in the program, you must import the relevant Agilent-supplied genome build files before you import the read file.

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

The Import Genome Build dialog box appears. See "Import Genome Build" on page 85.

2 Enter the following information:

Item	Instructions/Details
Species	Type the name of the species to which the imported genome build applies.
Build Name	Type the name of the genome build, as you would like it to appear in lists, dialog boxes, and the like.
Refseq file	This file contains the names and locations of genes in the genome of the selected species. **a** Click **Browse.** An Open dialog box appears. **b** Select the desired Agilent-supplied Refseq file, then click **Open.**
Cytoband file	This file contains the names and locations of cytobands in the genome of the selected species. **a** Click **Browse.** An Open dialog box appears. **b** Select the desired Agilent-supplied cytoband file, then click **Open.**

A dialog box tells you that the genome build has been successfully imported.

3 Click **OK.**

CAUTION Use only Agilent-supplied genome build files. You can download these files from the eArray Web site. In your individual user workspace in the CGH or ChIP application type, in the Probe tab, click **DNA Analytics Download.** For more information, see the online help on the eArray Web site.

Setting up QC experiments

To create QC reports on your sequence read data, you must set up an experiment. An experiment is a required level of organization that links a specific read file with a specific set of genomic targets.

To create a new experiment

An experiment links a specific sequence read file with the desired target genomic intervals. Target genomic intervals are defined in a BED format track file. Before you can create an experiment, you must import the sequence read file and the track file into the program . See "To import sequence read files" on page 30 and "To import target genomic regions as a track" on page 33.

1 Set the application type to **SureSelect Target Enrichment.** See "To set the application type" on page 18.

2 In the Experiment pane of the Navigator, right-click the **Experiments** node (or folder), then click **New Experiment.**

The Create Experiment dialog box appears. See "Create Experiment" on page 81.

3 Enter the following information:

Parameter	Instructions/Details
Name	Type a name for the experiment. The program uses this name to identify the experiment in the Experiment pane of the Navigator.
Description	(Optional) Type a description for the experiment.
Read Length	Type the number of bases in each read in your sequence read file. The program assumes all reads in the file have the same length.
Targets	Select the track that contains the definitions of the desired genomic targets. If the desired track does not appear in the list, you must import it. Close the dialog box, then see "To import target genomic regions as a track" on page 33.

Parameter	Instructions/Details
SEQ Sample	Select the desired read file. If the desired read file does not appear in the list, you must import it. Close the dialog box, then see "To import sequence read files" on page 30.
Paired	If you have two output files in which the reads are paired, mark this option, then select the paired file. The program imports the two files separately, then merges them before analysis.

4 Click **OK.**

A progress bar appears. The program adds the experiment to the Experiment pane of the Navigator. To view the experiment, and to create QC reports, you must select the experiment. See "To select an experiment for analysis" on page 38.

To select an experiment for analysis

To view and analyze the reads in an experiment, and to create QC reports, you must first select the desired experiment to make it active. Only one experiment can be active at a time.

1 Set the application type to **SureSelect Target Enrichment.** See "To set the application type" on page 18.

2 Click **Quality Analyzer.**

SureSelect Quality Analyzer tab appears.

3 In the Experiments pane of the Navigator, in the Experiments folder, right-click the name of the desired experiment, then click **Select Experiment.**

A Confirm dialog box asks if you want to select the experiment.

4 Click **Yes.**

A progress bar appears. The program activates the experiment. In the Experiment pane of the Navigator, the name of the selected experiment appears in blue. You can now also create and view reports based on the reads in the experiment. See "Creating Target Enrichment QC reports" on page 48. To display the reads and intervals from the experiment in the Genomic Viewer, see "To view the data in an experiment" on page 39.

To deselect an experiment

When you deselect an experiment, its reads no longer appear in the Genomic Viewer, and it is no longer activated for analysis and report creation. However, no files are deleted.

Do one of the following:

- In the Experiment pane of the Navigator, right-click the experiment that you want to deselect, then click **Deselect.**
- Select another experiment. See "To select an experiment for analysis" on page 38.

To view the data in an experiment

After you select an experiment, the reads in the experiment appear in all views of the Genomic Viewer.

1. Set the application type to **SureSelect Target Enrichment.** See "To set the application type" on page 18.
2. In the View tab, under Custom Data, mark **Show.**
3. Select the desired experiment. See "To select an experiment for analysis" on page 38.

   The sequence reads from the experiment appear in all views of the Genomic Viewer. Figure 4 shows the Genomic Viewer with reads on human chromosomes 3, 6, and 11.

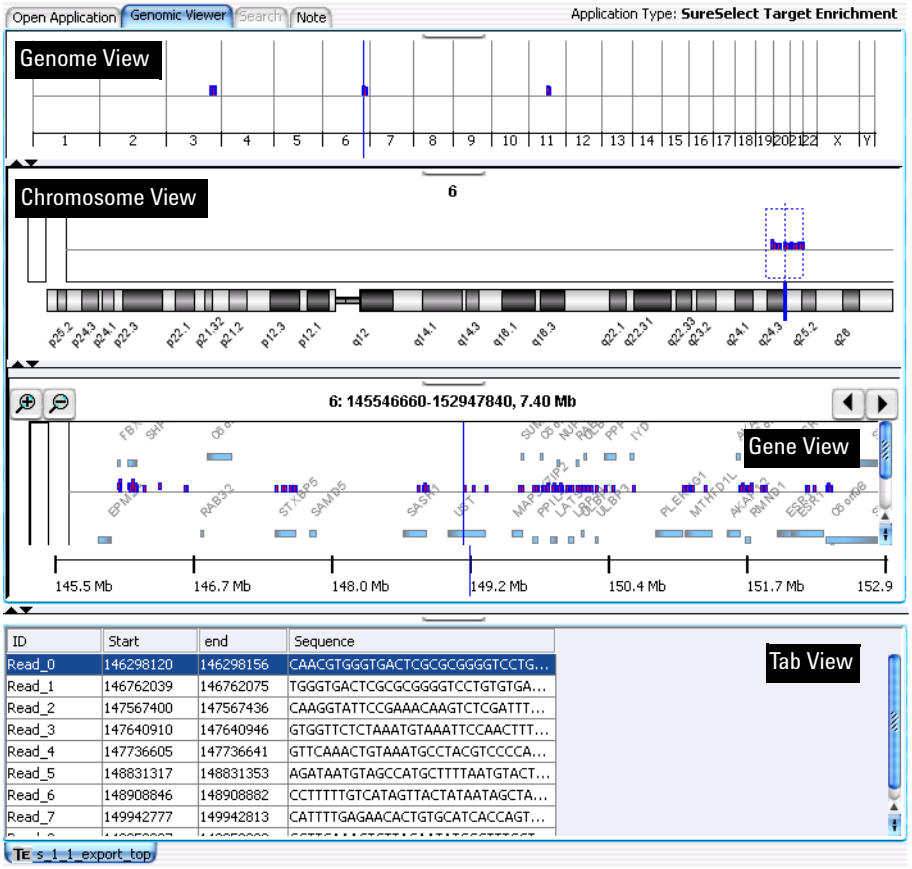


Figure 4 Genomic Viewer. with sequence reads from an experiment

The following appears in each pane of the Genomic Viewer:

View	Description/Instructions
Genome View	Shows all of the chromosomes of the organism. For chromosomes that have reads associated with them, red/blue bars appear at the approximate loci of reads. When you select a chromosome, its reads appear in all of the other views. Also, to view the results of enrichment analysis in the UCSC Genome Browser, you must first select a chromosome in Genome View. The program subsequently transfers results only for the selected chromosome. • To select a chromosome, click anywhere within its boundaries. See "Genomic Viewer – Genome View" on page 69.
Chromosome View	Shows the chromosome that was selected in Genome View. A diagram of the selected chromosome appears. Red/Blue bars appear above this diagram that show the approximate locations of reads. • In the area where the reads appear, drag the pointer through a region, or click a specific location. The region appears in Gene View in greater detail. See "Genomic Viewer – Chromosome View" on page 71.
Gene View	Shows the selected region of the chromosome. You can adjust Gene View in several ways: • To re-center Gene View on a specific read or gene, click the desired read or gene. • To zoom in, click  . If you zoom in far enough on a specific read, its nucleotide sequence appears. • To zoom out, click  . • To scroll the view, click  or  . See "Genomic Viewer – Gene View" on page 73.
Tab View	Displays a list of the reads that appear in Gene View. See "Genomic Viewer – Tab View" on page 75.

For further instructions on how to use the Genomic Viewer, see the *Data Viewing User Guide*. To open this guide, click **Help > Data Viewing.**

To view the intervals in a target interval track

After you import a set of target intervals as a track, and link the track and a read file in an experiment, you can view the list of intervals in it.

1 In the Experiment pane of the Navigator, in the Experiments folder, double-click the name of an experiment that contains the desired target intervals track.

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

The Track dialog box appears, with header information and a list of intervals. See "Track" on page 90.

NOTE You can also use the Genomic Viewer to see the intervals in the genomic region that appears in Gene View. Drag the bottom border of Gene View downward until Gene View is approximately twice its original size. The target intervals track appears at the bottom of gene view.

To export tracks

You can export some or all of the tracks that are available in the program. The program exports all of the intervals the selected tracks in a single BED format file. This makes it possible to combine tracks. You can then import the combined track and use it in an analysis experiment.

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

The Export Tracks dialog box appears. See "Export Tracks" on page 83.

2 In **Select Tracks,** mark the tracks that you want to export.

3 Click **OK.**

An Export dialog box appears.

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

The program exports all selected tracks in a single BED format file. You can open the file in a compatible genome browser, and you can also import the BED file into Agilent Genomic Workbench as a new track. See "To import target genomic regions as a track" on page 33.

To remove a track from the program

If you import a track in error, or if you no longer need one, you can permanently remove it from the program.

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

The Tracks tab of the User Preferences dialog box appears. See "User Preferences – Tracks tab" on page 95.

2 In the list of tracks, in the Delete column, mark the track that you want to delete. You can mark as many tracks as you want.

3 Click the **Delete** button.

A dialog box asks if you are sure you want to delete the selected track(s).

CAUTION When you delete a track, you permanently remove it from the program. To restore a track, you must import it again.

4 Click **Yes.**

The program deletes the track.

5 Click **OK.**

To customize the appearance of an experiment in the Genomic Viewer

The table below describes options that can be useful when you view sequence reads in the Genomic Viewer:

Customization option	Instructions/Details
To view a different chromosome in Chromosome, Gene, and Tab views.	• In Genome View, click the desired chromosome.
To select a specific chromosomal region to view in Gene View	• In Chromosome View, in the plotting area where reads appear, drag the pointer across the desired region.
To move the cursor to a specific genomic location	The cursor is a thin blue line that appears in Genome, Chromosome, and Gene views, and indicates the center of the currently selected region. **a** In the View tab, click **Go to Gene/Genomic Location.** The Go to Gene/Genomic Location dialog box appears. See "Go to Gene/Genomic Location" on page 84. **b** Do one of the following: ◦ In RefSeq by Symbol, type a valid RefSeq gene symbol (such as `BRCA1` or `CTSB`), then click **Go.** ◦ In Genomic Location, select a chromosome, type a base location, then click **Go.**
To scroll Gene View	• Click the scroll buttons at the top of Gene View, as desired.  – Scrolls Gene View left.  – Scrolls Gene View right. **Note:** You can also click anywhere within the gene display area of Gene View to reposition the cursor to that location.
To show only certain chromosomes in Genome View	• Right-click anywhere within Genome View, then mark or clear the check boxes next to the names of the desired chromosomes. To close the chromosome selection menu, click anywhere outside of the menu.
To show or hide reads in Genome, Chromosome, and Gene Views.	• In the View tab, under Custom Data, mark or clear **Show,** as desired.

Customization option	Instructions/Details
To show or hide genes in Gene View	**a** In the Home tab, click **User Preferences.** The User Preferences dialog box appears. See "User Preferences – Tracks tab" on page 95. **b** In the Tracks tab, under Visualization Parameters, in Genes, mark or clear **Show Gene Symbols in Gene View,** as desired.
To show or hide the intervals in the target intervals track	**a** In the Home tab, click **User Preferences.** The User Preferences dialog box appears. See "User Preferences – Tracks tab" on page 95. **b** In the Tracks tab, under Visualization Parameters, in Tracks, mark or clear **Show Annotations,** as desired. **c** Click **OK.**
To change the width of a column in Tab View.	You can change the width of any column, but it can be especially useful to expand the Sequence column to view the full sequence of reads. • In the column heading row of Tab View, drag the right edge of the desired column to the left or to the right, as desired.
To increase the size of Gene View	You can expand Gene View to see the target interval track, which can be hidden. Do one of the following: • If Gene View appears horizontally, drag its bottom border down until you can see the target interval track. • If Gene View appears vertically, drag the right border of the main program window to the right, or the left border of Gene View to the left.

Customization option	Instructions/Details
To change the zoom level in Gene View	You can increase the zoom level in Gene View to see the boundaries of individual reads, and how they overlap other reads. • To zoom in, click  . • To zoom out, click  .
To switch between horizontal and vertical orientation	Because of the shape of most monitors, it can be more convenient to use the Genomic Viewier in its default (horizontal) orientation. However, you can use the Genomic Viewer in vertical orientation. **a** In the View tab, click **View Preferences.** The View Preferences dialog box appears. See "View Preferences" on page 98. **b** Under View Alignment, in Orientation, select **Horizontal** or **Vertical,** as desired. **c** Click **OK.**

For additional information about the Genomic Viewer, see the topics referenced in the table above, and the *Data Viewing User Guide*.

To delete an experiment

1 In the Experiment pane of the Navigator, right-click the name of the desired experiment, then click **Delete.**

A dialog box asks if you want to delete the experiment.

CAUTION When you delete an experiment, you permanently remove the link between the particular sequence read file and the target intervals file. To restore the experiment, you must create a new one. However, the program does not delete any data files that you have imported, or any reports that you have saved.

2 Click **OK.**

To save an experiment

The program automatically saves experiments as you create them. No additional action is required on your part.

To copy the what appears on your screen to the clipboard

You can copy one or all parts of the main program window to the clipboard. You can then paste the copied image into a document in a word processing, graphics, or presentation program.

1 In the View tab, click **Copy,** then select the pane that you want to copy. To copy all panes, including the navigator, select **All.**

2 In a word processing, graphics, or presentation program, paste the contents of the clipboard into the desired document.

NOTE This procedure transfers the selected item as a bitmapped image. This image contains only what appears on your screen. It does not contain any hidden panes, or any areas to which you must scroll.

Creating Target Enrichment QC reports

SureSelect Quality Analyzer can calculate QC metrics and analyze read depth and enrichment, based on the sequence reads in the currently selected experiment.

This section contains the following topics:

To analyze read quality

After you import read data and target intervals files, and set up and select an experiment, the program can calculate overall QC metrics, and can also analyze read depth and enrichment.

1. Set the application type to **SureSelect Target Enrichment.**
2. Import files and set up and activate an experiment as described in "Importing files" on page 30 and "Setting up QC experiments" on page 36.
3. In the Quality Analyzer tab, click **Create Report.**

   The program calculates QC metrics and analyzes read depth and enrichment. A dialog box tells you that reports were successfully created.
4. Click **OK.**

   At this point, the program has created all reports internally. To view a report, you must request a specific report. See "To view and save reports" on page 49.

To view and save reports

After the program calculates QC metrics and analyzes read depth and enrichment, you can view and save PDF and text format reports. You can also view the results of enrichment analysis in the UCSC Genome Browser.

1 Analyze read quality as described in "To analyze read quality" on page 48.

2 View and/or save the desired report as described in the table below:

Report	Description/Instructions
PDF summary report	*.pdf file that contains overall QC metrics. Also contains results of read depth analysis both as tables and as a graph. **a** In the Quality Analyzer tab, click **Summary Report > Save as PDF.** A Save dialog box appears. **b** Select a location and type a name for the report file, then click **Save.** A dialog box tells you that the report was successfully saved, and asks if you would like to view it. **c** Click **Yes.** The report opens in Adobe Reader.
Text summary report	*.txt file that contains overall QC metrics. This report also contains a table of results of read depth analysis. **a** In the Quality Analyzer tab, click **Summary Report > Save as Text.** A Save As dialog box appears. **b** Select a location and type a name for the report file, then click **Save As.** A dialog box tells you that the report was successfully saved, and asks if you would like to view it. **c** Click **Yes.** The report opens.

Report	Description/Instructions
Enrichment analysis results – UCSC	Opens the UCSC Genome Browser and lets you view the results of enrichment analysis for a selected chromosome. See "Enrichment analysis results – UCSC" on page 104. **a** In the Genomic Viewer, in Genome View, click anywhere within the boundaries of the desired chromosome. **b** In the command ribbon, click **Enrichment > View in UCSC.** Your Web browser opens. A message can appear that tells you the browser has blocked scripts and ActiveX controls. If this occurs, allow the blocked content. The UCSC Genome Browser opens. **c** In **position/search,** type the genomic coordinates of a region that contains reads, then click **jump.** The enrichment analysis results appear as a bar graph in the User Track area. For additional information on how to use the UCSC Genome Browser, see the online help on the site.
Enrichment analysis results – XLS report	*.xls format file that contains enrichment analysis results for a selected chromosome. See "Enrichment analysis results – XLS report" on page 105. **a** In the Genomic Viewer, in Genome View, click anywhere within the boundaries of the desired chromosome. **b** In the command ribbon, click **Enrichment > Save as XLS.** A Save As dialog box opens. **c** Select a location and type an name for the report file, then click **Save As.** A dialog box tells you that the report was saved, and asks if you want to view it. **d** Click **Yes.** The report opens in Microsoft Excel. If Microsoft Excel is not installed on your computer, a dialog box asks you to select a program to open the file. If this dialog box appears, select a spreadsheet program that can open *.xls files.

- For information about QC metrics, see "QC metrics" on page 101.
- For information about the graph of read depth distribution, see "Read depth distribution" on page 103.

3
SureSelect Quality Analyzer Reference

This chapter describes the parts of the Agilent Genomic Workbench main window that are relevant to SureSelect Quality Analyzer. It also describes the menus, shortcut menus, panes, and dialog boxes that can appear. In addition, it contains details about the reports that the program can produce.

For general help with the user interface, see the *Data Viewing User Guide*. In addition, separate user guides cover each of the main programs in Agilent Genomic Workbench. To view these other guides, select the relevant application type, then click the **Help** tab.

Quality Analyzer tab

Figure 5 Agilent Genomic Workbench – Quality Analyzer tab

The Agilent Genomic Workbench main window has these main elements:

Element	Description
Title bar	Displays the program name and the currently selected application type.
Tab bar	Contains the tabs that let you use the main functions of the program. See "Agilent Genomic Workbench tab bar" on page 54.
Switch Application menu	Lets you select the experimental application type. This is one of the main settings for Agilent Genomic Workbench, since functionality as well as design, sample, and sequence data are partitioned by application type. See "Switch Application menu" on page 55 and "To set the application type" on page 18.
Command ribbon	Contains the commands that you use to access the main functionality of Agilent Genomic Workbench. A different command ribbon appears for each main tab. The available commands also vary by application type. See these topics: • "Command ribbon – Home tab" on page 56 • "Command ribbon – Quality Analyzer tab" on page 58 • "Command ribbon – Help tab" on page 60
Navigator	Contains lists of the designs, sample data, sequence data, analysis experiments, pending jobs, and other related items that are available to you. A search feature is also available. See these topics: • "Navigator overview" on page 63 • "Navigator – Experiment pane" on page 64
Genomic Viewer	Shows the genome of the selected organism at several levels of detail. Genes, tracks, probes, baits, and data of many types can appear in the genomic viewer in the context of the genome. For general information on how to use the Genomic Viewer, see the *Data Viewing User Guide*. For information on the specific features of the Genomic Viewer that are relevant to the Quality Analyzer tab, see "Genomic Viewer – overview" on page 67.

Agilent Genomic Workbench tab bar

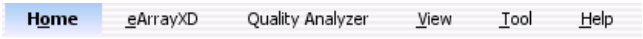


Figure 6 Tab bar (SureSelect Target Enrichment application type)

The tabs that are available in the tab bar vary by application type. To create and view Target Enrichment QC reports, you set the application type to SureSelect Target Enrichment. For this application type, these tabs appear in the tab bar:

Tab	Description
Home	Lets you set user preferences, import and export files, create analysis experiments, view data, genes, and tracks, print certain panes, and exit the program. See "Command ribbon – Home tab" on page 56. **Note:** To calculate Target Enrichment QC metrics, and to produce reports, you do not need to use this tab. All necessary commands are available in the Quality Analyzer tab.
eArrayXD	Lets you create custom bait libraries for Target Enrichment experiments. For other application types, $eArray_{XD}$ lets you create custom microarrays. For details, see the *$eArray_{XD}$ User Guide*.
Quality Analyzer	Lets you calculate QC metrics based on selected sequence read data, and analyze read depth and enrichment. You can also create PDF and text QC reports. See "Command ribbon – Quality Analyzer tab" on page 58.
View	Lets you set view preferences, and customize the appearance of the main program window. For details, see the *Data Viewing User Guide*.
Tool	Lets you create new accounts on the eArray Web site for users in your workgroup. (Workgroup administrator access is required.) Also lets you send error logs to Agilent and create text notes.
Help	Lets you view the user guides that are relevant to the SureSelect Target Enrichment application type. See "Command ribbon – Help tab" on page 60.

Switch Application menu

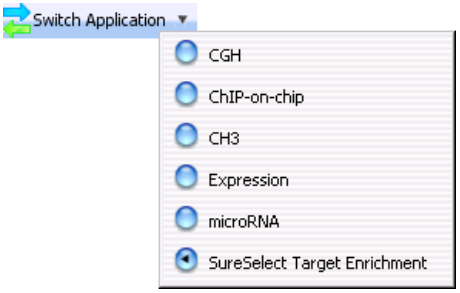


Figure 7 Switch Application menu

Purpose: Lets you select the main experimental application type for the program. Design, sample, and sequence data, as well as program functionality, are partitioned by application type. See "To set the application type" on page 18. To use the Target Enrichment Quality Analyzer, you select **SureSelect Target Enrichment** as the application type.

To open: Click **Switch Application.**

These application types are available:

Application Type	Description
CGH	Microarray design and data analysis for comparative genomic hybridization studies.
ChIP-on-chip	Microarray design and data analysis for chromatin immunoprecipitation studies.
CH3	Microarray design and data analysis for methylation studies.
Expression	Microarray design for gene expression studies.
microRNA	Microarray design for microRNA studies.
SureSelect Target Enrichment	Design of bait libraries for the retrieval of specific DNA fragments for sequencing. Also lets you do a QC analysis of target fragment pull-down.

Command ribbon – Home tab

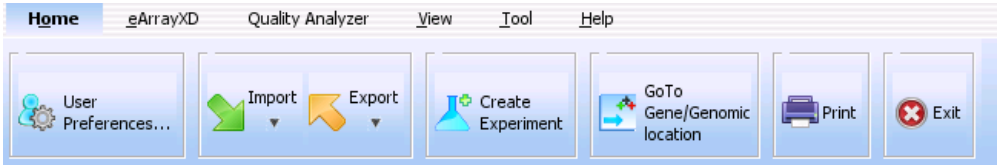


Figure 8 Command ribbon – Home tab (SureSelect Target Enrichment application type)

Purpose: Lets you set user preferences, import and export files, create analysis experiments, view data, genes, and tracks, print certain program panes, and exit the program.

To open: Set the application type to **SureSelect Target Enrichment,** then click **Home.**

NOTE You can use the Home tab to import files, create analysis experiments, and view read data and target intervals. However, you can also use the Quality Analyzer tab for this purpose, and conveniently do the desired analyses and create reports in one location.

User Preferences Opens the User Preferences dialog box, where you can customize the appearance of genes and tracks in the Genomic Viewer, select a default data location, and select other program options. See "User Preferences – Miscellaneous tab" on page 93 and "User Preferences – Tracks tab" on page 95.

Import Opens a menu with these options:

Option	Description
Reads	Opens a dialog box where you can import a sequence read file. See "To import sequence read files" on page 30.
Track	Opens the Import Track dialog box, where you can set up and import a target intervals track file. See "To import target genomic regions as a track" on page 33 and "Import Track" on page 86.
Genome Build	Opens the Import Genome Build dialog box, where you can set up and import a new Agilent-supplied genome build. See "To import a genome build" on page 35 and "Import Genome Build" on page 85.

Option	Description
Bait Upload	This option is used by eArray$_{XD}$, and is not applicable to SureSelect Quality Analyzer. See the *eArray$_{XD}$ User Guide*.
Custom Genome for Tiling	This option is used by eArray$_{XD}$, and is not applicable to SureSelect Quality Analyzer. See the *eArray$_{XD}$ User Guide*.

Export Opens a menu with a Tracks option. This option opens the Export Tracks dialog box, where you can select one or more tracks to export as a single BED format file. See "To export tracks" on page 42 and "Export Tracks" on page 83.

Create Experiment Opens the Create Experiment dialog box, where you can begin the creation of a Target Enrichment QC analysis experiment. Experiments are required organizational units that link a read file with related files and information. See "Create Experiment" on page 81 and "To create a new experiment" on page 36.

Go to Gene/Genomic Location Opens the Go to Gene/Genomic Location dialog box, where you can enter a gene symbol or genomic location. This lets you move the cursor of the Genomic Viewer to a specific locus. See "To customize the appearance of an experiment in the Genomic Viewer" on page 44 and "Go to Gene/Genomic Location" on page 84.

Print Opens a Print dialog box, where you can print the contents of several program panes. See the *Data Viewing User Guide*.

Exit Exits the program.

Command ribbon – Quality Analyzer tab

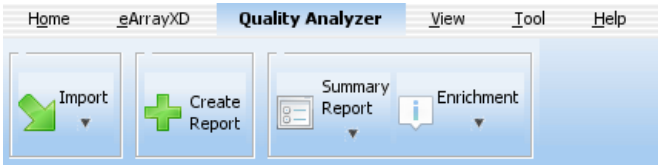


Figure 9 Command ribbon – Quality Analyzer tab (SureSelect Target Enrichment application type)

Purpose: Lets you import sequence read files and target interval tracks, do QC analyses, and create QC reports.

To open: Set the application type to **SureSelect Target Enrichment,** then click **Quality Analyzer.**

Import Opens a menu with these options:

Option	Description
Reads	Opens a dialog box where you can import a sequence read file. You must import a *.zip file that contains a read file in one of these formats: • Illumina GA Export • AB SOLiD ma • MAQ aln.txt See "To import sequence read files" on page 30.
Track	Opens the Import Track dialog box, where you can set up and import a target intervals track. The track must be a *.txt file in BED format. See "To import target genomic regions as a track" on page 33.

Create Report Analyzes the data in the currently selected experiment. To view the results, you use the **Summary Report** and **Enrichment** commands (see below).

Summary Report Opens a menu with these options:

Option	Description
Save as PDF	Opens a Save dialog box, where you can select a location for a PDF QC report file. This report contains calculated QC metrics. It also contains the results of read depth analysis both as tables and as a graph. See these topics: • "To view and save reports" on page 49 • "QC metrics" on page 101 • "Read depth distribution" on page 103.
Save as Text	Opens a Save As dialog box where you can select a location for a *.txt format report. This report contains calculated QC metrics. It also contains tables of the results of read depth analysis. See "To view and save reports" on page 49 and "QC metrics" on page 101.

Enrichment Opens a menu with these options:

Option	Description
View in UCSC	Opens the UCSC Genome Browser in your Web browser, and displays the results of enrichment analysis in the User Track area. This command lets you view these results one chromosome at a time. You must select the desired chromosome in Genome View of the Genomic Viewer before you click this command. See "To view and save reports" on page 49 and "Enrichment analysis results – UCSC" on page 104.
Save as XLS	Opens a Save As dialog box, where you can select a location for a *.xls format report. This report contains the results of enrichment analysis for each target interval in the selected chromosome. You select the desired chromosome in Genome View of the Genomic View before you click this command. You can use a spreadsheet program to view this report.

NOTE To view or save summary or enrichment reports, you must first import a read file and a target intervals file, create and select an analysis experiment, and analyze the data in the selected experiment.

Command ribbon – Help tab

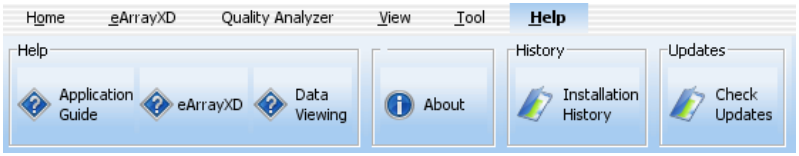


Figure 10 Help tab (SureSelect Target Enrichment application type)

Purpose: Lets you view the user guides for Agilent Genomic Workbench that are relevant to the selected application type. This tab also lets you view information about the client and server patches that you have installed for Agilent Genomic Workbench, and check for software updates.

To open: Set the desired application type to then click **Help.**

Help

All guides open in Adobe Reader. These buttons can appear:

Button	Description
Application Guide	(Available for all application types except Expression and microRNA) For each of these application types, this button opens the indicated user guide: • **CGH** – Opens the *CGH Interactive Analysis User Guide.* This guide describes how to use the CGH application of Agilent Genomic Workbench to analyze comparative genomic hybridization data and create reports. • **ChIP-on-chip** – Opens the *ChIP Interactive Analysis User Guide.* This guide describes how to use the ChIP application of Agilent Genomic Workbench to analyze chromatin immunoprecipitation data and create reports. • **CH3** – Opens the *Methylation (CH3) Analysis User Guide.* This guide describes how the use the Methylation (CH3) application of Agilent Genomic Workbench to apply algorithms that help identify methylated regions. • **SureSelect Target Enrichment** – Opens the *SureSelect Quality Analyzer User Guide.* This guide describes how to use the SureSelect Quality Analyzer application of Agilent Genomic Workbench to assess the effectiveness of fragment pull-down for target enrichment experiments

Button	Description
$eArray_{XD}$	(Available for all application types) Opens the *$eArray_{XD}$ User Guide.* This guide describes how to design and submit custom microarray designs and SureSelect Target Enrichment bait libraries.
Sample Manager	(Available for all application types except SureSelect Target Enrichment) Opens the *Sample Manager User Guide.* This guide describes how to use the Sample Manager application of Agilent Genomic Workbench to organize microarrays and edit their attributes. You can use Sample Manager for Feature Extraction without any of the analysis applications.
Feature Extraction	(Available for all application types except SureSelect Target Enrichment) Opens a menu with these options: • **Quick Start** – Opens the *Feature Extraction Quick Start Guide.* This guide gives an overview of how to use the Feature Extraction software to extract and generate QC reports for Agilent microarrays. • **User Guide** – Opens the *Feature Extraction User Guide.* This guide shows you how to set up and run Feature Extraction to automatically extract a batch of image files. It also describes how to extract image files in real time. • **Reference Guide** – Opens the *Feature Extraction Reference Guide.* This guide contains tables that contain lists of default parameter values and results for Feature Extraction analyses, and explanations of how Feature Extraction uses its algorithms to calculate results.
Quality Tools	(Available for the CGH, ChIP-on-chip, and CH3 application types) Opens the *Quality Tools User Guide.* This guide describes how to query, filter, and evaluate microarray extractions within Agilent Genomic Workbench. It also describes how to visualize current and historical batch microarray extraction processes.

Button	Description
Workflow	(Available for the CGH and ChIP-on-chip application types) Opens the *Workflow User Guide.* This guide describes how to use the workflow module of Agilent Genomic Workbench to extract image files with Agilent Feature Extraction software and/or use the CGH and ChIP analysis applications to analyze data.
Data Viewing	(Available for all application types except Expression and microRNA) Opens the *Data Viewing User Guide.* This guide describes how to import, organize, manage, export, and display data and other content (experiments, gene lists, tracks) within Agilent Genomic Workbench. It is targeted for users who have no DNA Analytics application license(s).

An additional guide is available in the Open Application tab of the program. The *Agilent Genomic Workbench Product Overview Guide* gives an overview of the capabilities within the Standard Edition of Agilent Genomic Workbench. It also describes how to start each of the component programs and find help, and how to enter your license information. In addition, it helps you with system administration and troubleshooting. To open this guide, click the **Open Application** tab, then click **Product Overview**.

About Opens a dialog box that displays title, copyright, and version information for the program. It also lets you view the License Agreement.

History

Installation History Opens a dialog box that displays the software updates that you have installed, and their installation histories. Both client and server updates appear in this dialog box.

Updates

Check Updates Connects the program to Agilent, and lets you download updates to the Agilent Genomic Workbench software.

Navigator overview

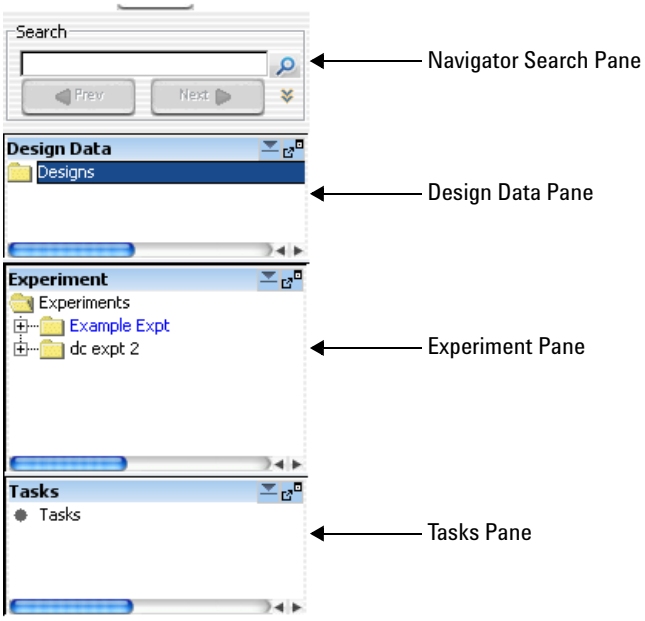


Figure 11 Navigator (SureSelect Target Enrichment application type)

For the SureSelect Target Enrichment application type, the Navigator contains several panes that show the content that is available to you:

Pane of Navigator	Description
Search	Lets you search all available panes of the Navigator for items that match a search term. For information about this pane, see the *Data Viewing User Guide.*
Design Data	Lets you work with bait groups and bait libraries in the SureSelect Target Enrichment application type in the eArray$_{XD}$ tab. You do not use this pane of the Navigator when you use the Quality Analyzer tab.

Pane of Navigator	Description
Experiment	Lets you create, select, and view experiments in the Target Enrichment Quality Analyzer. See "Navigator – Experiment pane" on page 64.
Tasks	Lets you view and take action on the jobs that you have submitted to your Agilent Genomic Workbench server, or to the eArray Web site. The Quality Analyzer tab does not create any jobs of this type.

Navigator – Experiment pane

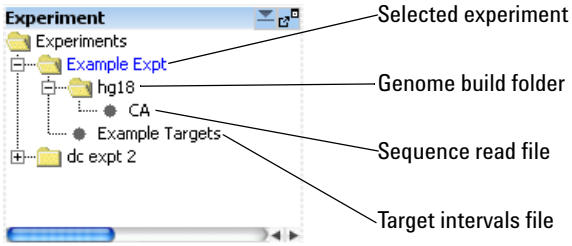


Figure 12 Navigator – Experiment pane

Purpose: In the SureSelect Target Enrichment application type, the Experiment pane of the Navigator lets you create, view, and analyze Target Enrichment QC analysis experiments, which are organizational units that link sequence read data with target intervals and other information. You create an experiment as a prerequisite to QC analysis.

Experiment pane – Icons and special text

These icons and special text items can appear in the Experiment pane of the Navigator:

Item	Description
	(Available if the Experiment pane is not minimized) Minimizes the experiment pane to the bottom of the Navigator.
	(Available if the Experiment pane is minimized) Restores the Experiment pane to its position in the middle of the Navigator.

Item	Description
	If the Experiment pane is docked in the Navigator, this button detaches the Experiment pane from the Navigator and opens it in a new, separate window. If the experiment pane is in its own separate window, this button re-attaches the Experiment pane to the Navigator.
	A collapsed folder whose contents are hidden. Folders can contain data items and/or other folders.
	An expanded folder. Its contents appear below the name of the folder in a hierarchical "tree" format.
	Expands a folder to show its contents.
	Collapses a folder to hide its contents.
	A data node that represents a specific imported read file or target interval track. It also represents the Experiments node (folder) when no experiments are present.
blue text	The name of an experiment that is selected for analysis.
red text	An item that matches the search term in a search of the Navigator.
highlighted text	The currently selected search result from a search of the Navigator.

Experiment pane – Actions and shortcut menus

- Double-click the **Experiments** folder to expand it. Double-click it again to collapse it.
- Right-click the **Experiments** folder to open a shortcut menu with a New Experiment option. This option opens the Create Experiment dialog box, where you can create a new QC analysis experiment. See "Create Experiment" on page 81 and "To create a new experiment" on page 36.

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

Option	Description
Select Experiment	(Available for experiments other than the currently active one) Lets you view and analyze the data in the experiment. The reads and target intervals that are defined for the selected experiment can appear in the Genomic Viewer, and are used for QC analysis. When you select an experiment, the program displays the name of the experiment in blue. You can only select one experiment at a time. See "To select an experiment for analysis" on page 38.
Deselect Experiment	(Available for the currently selected experiment, if any) Removes the reads and target intervals that are defined in the experiment from the Genomic Viewer, and makes them unavailable for QC analysis. See "To deselect an experiment" on page 39. If you deselect an experiment, you can select it again as needed.
Delete Experiment	Removes the experiment from the Experiment pane of the Navigator. To restore a deleted experiment, you must create a new experiment. See "To delete an experiment" on page 46.

- Double-click the name of an experiment to view its contents. Double-click the name of the experiment again to hide its contents. Each experiment contains a genome build folder that contains a link to the uploaded read file that has been selected for the experiment. It also contains a link to the target intervals track that has been selected for the experiment.
- Right-click the name of a target intervals file to open a shortcut menu with a View Details option. This option opens the Track dialog box, where you can view a list of the target intervals in the track, and header information. See "To view the intervals in a target interval track" on page 42 and "Track" on page 90.

Genomic Viewer – overview

This section describes the panes that appear in the Genomic Viewer, and the commands, options, and shortcut menus that are available. These descriptions are specific to the SureSelect Target Enrichment application type. Other applications, such as the CGH, ChIP-on-chip, and CH3 data analysis applications, display data and results differently.

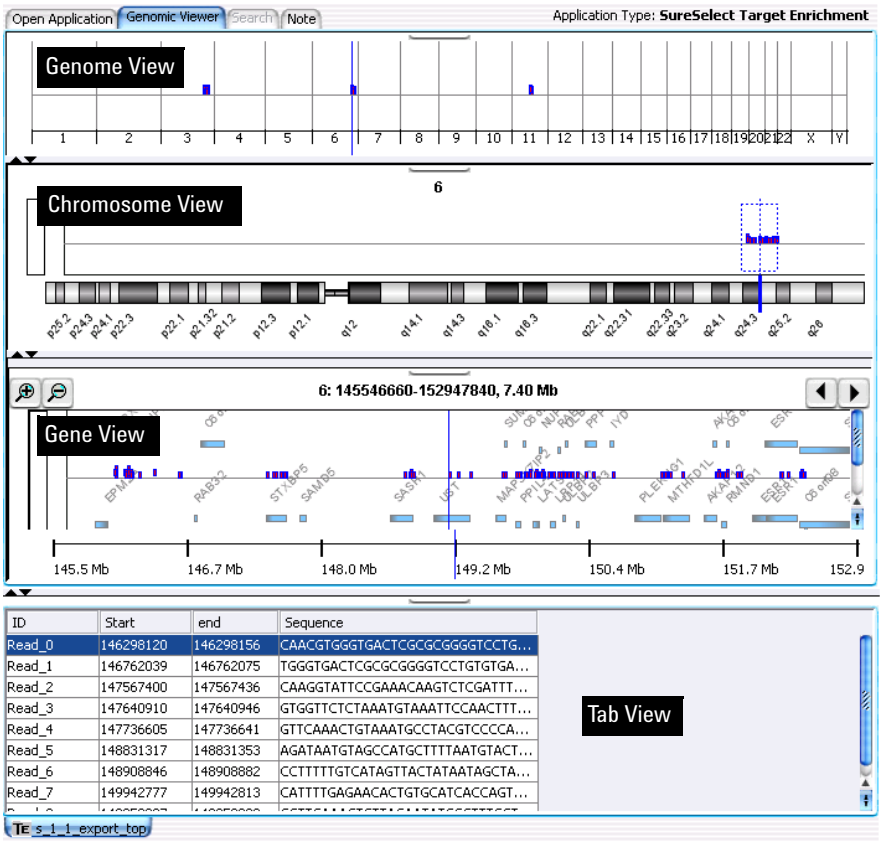


Figure 13 Genomic Viewer – SureSelect Target Enrichment application type

By default, for the SureSelect Target Enrichment application type, the Genomic Viewer appears in horizontal orientation. However, if desired, you can change the orientation to vertical. See "To customize the appearance of an experiment in the Genomic Viewer" on page 44.

The Genomic Viewer contains four panes:

Pane	Description
Genome View	Shows reads in the context of the genome of the species of interest. See "Genomic Viewer – Genome View" on page 69.
Chromosome View	Shows the reads that are mapped to a selected chromosome. See "Genomic Viewer – Chromosome View" on page 71.
Gene View	Shows an expanded view of a selected chromosomal region. The reads, target intervals, and genes in the selected region appear. See "Genomic Viewer – Gene View" on page 73.
Tab View	Shows a list of the reads on the selected chromosome. See "Genomic Viewer – Tab View" on page 75.

The four views are linked. When you select a location in one view, the other views change to display that location. The selected location appears as a blue line ("the cursor") in Genome, Chromosome, and Gene Views.

Buttons available in all panes

Detach button. Detaches the view and opens it in a separate window. To re-attach the detached view to the main window, click its close box ✕.

Resize buttons. Increases or decreases the size of the given pane. You can also drag the border between panes to resize them.

Genomic Viewer – Genome View

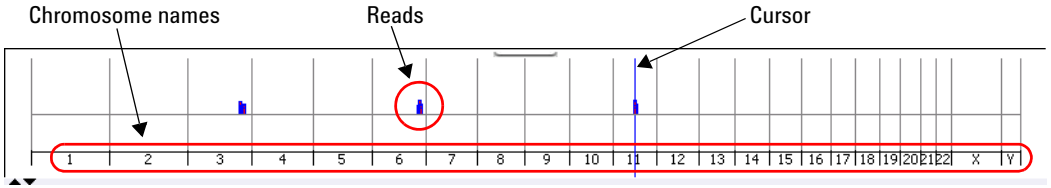


Figure 14 Genome View, with reads on chromosomes 3, 6, and 11. Chromosome 11 is selected.

When an experiment is selected, this view shows reads as blue and red bars in the context of the genome of the organism of interest. The chromosome that you select in this view sets the chromosome for the other views. Also, before you can view the results of enrichment analysis, you must first select the desired chromosome in Genome View. The program lets you view these results for one chromosome at a time. See "To view and save reports" on page 49.

Chromosome Names The names of the chromosomes of the organism, in order. The width of the space that each chromosome occupies is proportional to its size.

Reads The reads from the read file appear as blue and red bars. Their placement within the boundary of a chromosome reflects their approximate location within that chromosome.

Cursor The cursor appears as a thin blue line at a specific genomic location. To move the cursor, click the desired new location within Genome View. The other views adjust to reflect the new position of the cursor.

Shortcut menu

- Right-click anywhere within Genome View to open a menu with these options:

Option	Description
User Preferences	Opens the User Preference dialog box, where you can customize the appearance of genes and tracks in Gene View and enter the default data storage location. See "User Preferences – Tracks tab" on page 95 and "User Preferences – Miscellaneous tab" on page 93. You can also enter database configuration parameters and proxy server settings, and your login credentials for the eArray Web site.
View Preferences	Opens the View Preferences dialog box, where you can customize the orientation of the main program window. See "To customize the appearance of an experiment in the Genomic Viewer" on page 44 and "View Preferences" on page 98.
Select All	Marks the check boxes next to the names of all chromosomes. All chromosomes of the organism appear in Genome View.
Deselect All	Clears the check boxes next to the names of all chromosomes, except for the one in which the cursor currently appears.
Chromosomes	The names of the all of the chromosomes of the current species appear. If the check box next to the name of a chromosome is marked, that chromosome appears in Genome View.

Genomic Viewer – Chromosome View

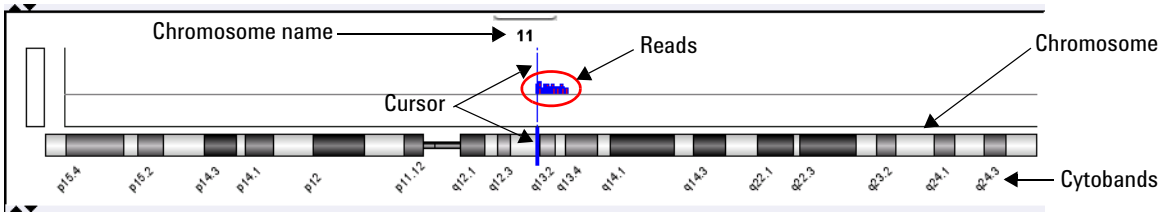


Figure 15 Chromosome View

This view shows an expanded view of the chromosome that is selected in Genome View. When an experiment is selected, reads appear as blue and red bars above the chromosomal locations to which they map. A region that you select in this view becomes the region that you can view in greater detail in Gene View.

Actions and shortcut menus

- Click a location within Chromosome View to move the cursor to the new location. The cursor in all views adjusts to the new location.
- Drag the pointer through part Chromosome View where reads appear to select a region to view in greater detail. The selected region appears in Gene View.

- Right-click anywhere within Chromosome View to open shortcut menu with these options:

Option	Description
User Preferences	Opens the User Preference dialog box, where you can customize the appearance of genes and tracks in Gene View and enter the default data storage location. See "User Preferences – Tracks tab" on page 95 and "User Preferences – Miscellaneous tab" on page 93. You can also enter database configuration parameters and proxy server settings, and your login credentials for the eArray Web site.
View Preferences	Opens the View Preferences dialog box, where you can customize the orientation of the main program window. See "To customize the appearance of an experiment in the Genomic Viewer" on page 44 and "View Preferences" on page 98.

Genomic Viewer – Gene View

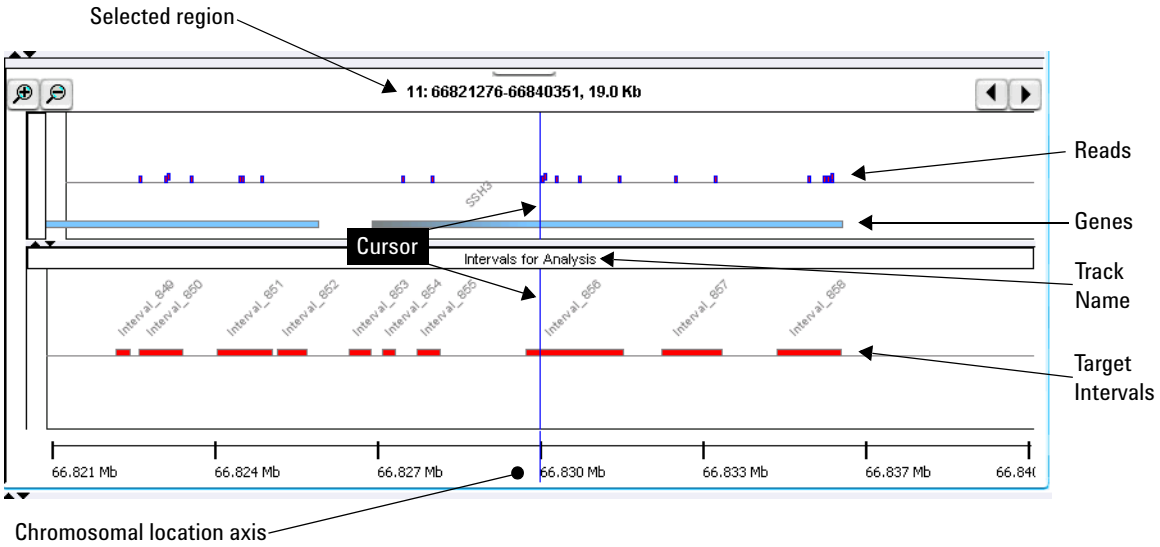


Figure 16 Gene View

Gene View displays the reads, target intervals, and genes that are associated with a specific region of a chromosome. Many zoom levels are available. See "To view the data in an experiment" on page 39.

Selected region The genomic coordinates of the chromosomal region that appears in Gene View.

Example: In Figure 16, the notation **11:66821276-66840351, 19.0 Kb** tells you that Gene View currently shows chromosome 11, base pairs 66,821,276 to 66,840,351, which is a region that is approximately 19.0 Kb in size.

Zooms in to see a smaller region in greater detail.

Zooms out to see a larger region.

Scrolls left to lower-numbered base pairs on the chromosome.

Scrolls right to higher-numbered base pairs on the chromosome.

Reads (Appear if an experiment is selected) The reads in the read file from the selected experiment appear as blue and red bars that are aligned with the genomic locations to which they map.

Genes The genes that map to the currently displayed chromosomal region. A light blue bar indicates the genomic region that is covered by each gene. By default, the name of each gene also appears.

Cursor The cursor is a thin blue line that indicates the center of the currently displayed chromosomal region. To change the location of the cursor, click one of the scroll buttons, or click anywhere within the gene display area of Gene View.

Track name (Appears if an experiment is selected) The name of the target intervals track from the currently selected experiment. To see the target intervals track, you must expand Gene View.

Target intervals (Appear if an experiment is selected) The target intervals from the selected experiment. By default, each target interval appears as a red bar that extends over the applicable genomic region. The name of each interval also appears.

Chromosomal location axis This axis applies to the reads, genes, and tracks that appear in Gene View. The values on the axis change as you zoom in or out, or go to different chromosomal locations.

Actions and shortcut menus

- Click anywhere within the gene display area of Gene View to move the cursor to a new location.
- Right-click anywhere within the gene display area of the Gene View to open a shortcut menu with these options:

Option	Description
Create Gene List	This option does not apply to SureSelect Quality Analyzer.
Show in UCSC	Opens the View Coordinates in UCSC Browser dialog box, where you can create a user track based on the region that appears in Gene View. From this dialog box, the program can open the UCSC Genome Browser in your Web browser to view the region(s) that you defined. See the *Data Viewing User Guide*.

Option	Description
Chromosomal Location Search	This option opens the Bait Search pane of eArray$_{XD}$, and transfers the species and the currently selected region in Gene View as a search term for a Chromosomal Location search. See the *eArray$_{XD}$ User Guide.*
User Preferences	Opens the User Preference dialog box, where you can customize the appearance of genes and tracks in Gene View and set the default data storage location. See "User Preferences – Tracks tab" on page 95 and "User Preferences – Miscellaneous tab" on page 93. You can also set database configuration parameters and your eArray.com login credentials.
View Preferences	Opens the View Preferences dialog box, where you can customize the orientation of the main program window. See "To customize the appearance of an experiment in the Genomic Viewer" on page 44 and "View Preferences" on page 98.

Genomic Viewer – Tab View

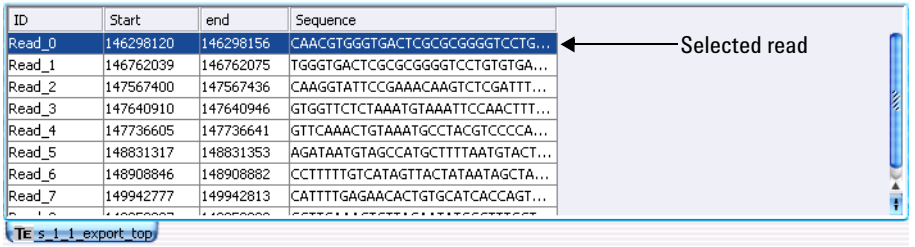


Figure 17 Tab View

When an experiment is selected, Tab View shows a list of the reads that are found on the currently selected chromosome.

ID The name of each read, as defined in the read file.

Start The first base pair on the chromosome to which each read maps.

End The last base pair on the chromosome to which each read maps.

Sequence The base sequence of each read.

Selected read To select a read, and move the cursor to its location, click anywhere in the row of the desired read. If you select a read, the program highlights it in blue.

Actions and shortcut menus

- Click anywhere within the row of a specific read to select it. When you select a read, the cursor location changes to the location of that read, and the other views adjust accordingly.
- Right-click anywhere in the column heading row to open a shortcut menu with a Scroll to Column option. This option opens the Scroll to Column dialog box, which lets you select a column, and scrolls Tab View so that the selected column is visible.
- In the column heading row, drag the right border of a column to increase or decrease the width of the column. This can be especially useful when you want to inspect the full base sequence of reads.
- Right-click any data item in Tab View to open a shortcut menu with these options:

Option	Description
Find in Column	Opens the Find in Column dialog box, where you can start a search for a text string in the column. See the *Data Viewing User Guide*.
Search Links	Many links to databases such as LocusLink, PubMed, and the UCSC Genome Browser appear. Each link passes the data item that you right-clicked to the selected database as a search string.
Customize Link	Opens the Customize Search Link dialog box, where you can create a new search link to include in the shortcut menu. See the *Data Viewing User Guide*.

Dialog boxes

Choose Track Color

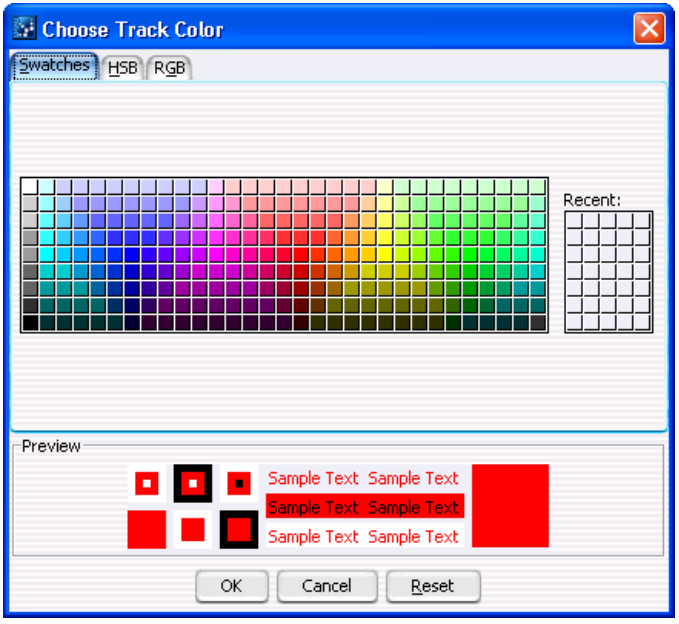


Figure 18 Choose Track Color dialog box – Swatches tab

Purpose: For the SureSelect Target Enrichment application type, lets you select a display color for the bars that represent genomic regions in the target intervals track in Gene View. See "To import target genomic regions as a track" on page 33.

To open: In the Import Track dialog box (see "Import Track" on page 86), under Color, click **Change.**

Swatches tab

Swatches To select a color, click a swatch. The new color appears in the Preview pane and also under Recent. The color that you select also becomes the selected color in the HSB and RGB tabs, where you can further refine the color.

Recent Displays the colors that you have recently selected in the Swatches tab. To select one of these recent colors, click it.

HSB tab

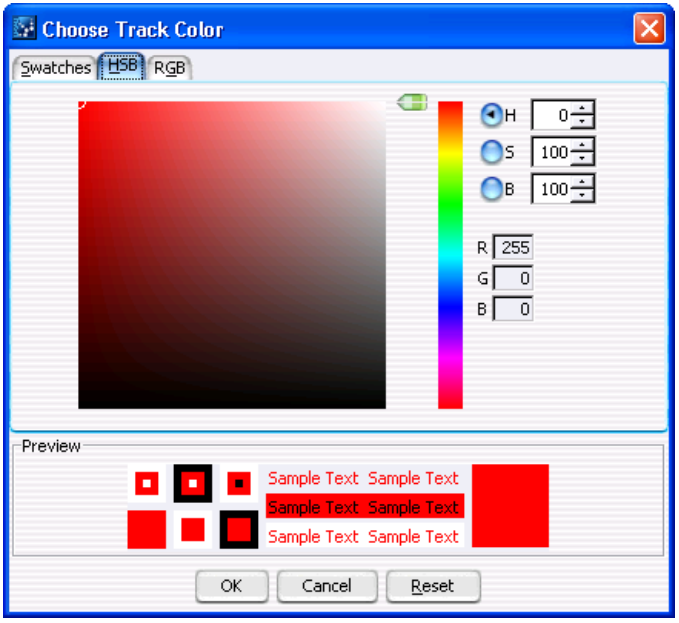


Figure 19 Choose Track Color dialog box – HSB tab

This tab lets you define the hue (H), saturation (S), and brightness (B) levels of for a color. These three values uniquely define a color. The initial color settings in this tab reflect any changes that you have made to the currently selected color in the other tabs. You can set these values in several ways:

- Directly edit the numbers in **H, S,** and **B.** You can also click the up or down button to the right of each value to increase or decrease it. In addition, to use the up and down arrow keys on your keyboard to change values, click the number that you want to change, then press the up or down arrow key, as desired.
- Use the green slider to change values. Select **H, S,** or **B,** as desired, then drag the slider up or down to change the selected value.
- Use the green slider and the large, square color selection box. Select **H, S,** or **B,** then drag the slider up or down to set the desired value. The color selection box shows all of the available colors given the particular setting of the slider. To select a color, click anywhere within the color selection box. This sets the remaining two HSB values.

H **Hue** – A number from 0 to 359 that represents the basic color. The color spectrum is a 360 degree color circle.

S **Saturation** – A number from 0 to 100 that represents the intensity of the color. A setting of 100 gives maximum color intensity. A setting of 0 gives no color, and reduces the available color spectrum to grayscale, only.

B **Brightness** – A number from 0 to 100 that represents the amount of black that is mixed in with the color. A setting of 100 gives maximum brightness, with no black added, and a setting of 0 results in the color black, without regard to the other settings.

R, G, B Each of these values shows the color settings of the selected color using the RGB (Red, Green, Blue) color model. You cannot directly edit the values from HSB tab, but they change when you select a different color. You can set specific RGB values in the RGB tab.

RGB tab

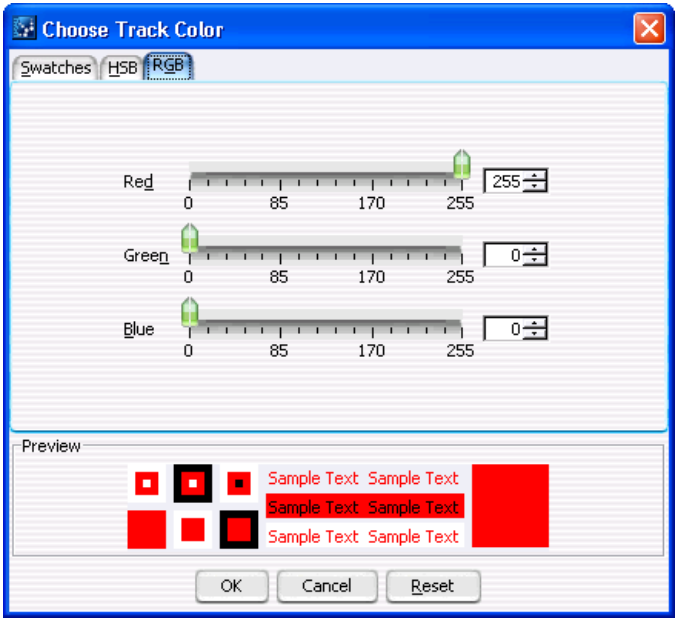


Figure 20 Choose Track Color dialog box – RGB tab

This tab lets you use the RGB (red-green-blue) color model to define a color. In this model, you select the amounts of red, green, and blue to combine to form the desired color. All colors can be defined.

You can change the R, G, or B value of the selected color in two ways:

- In **Red, Green,** or **Blue,** drag the green slider to the desired value.
- Directly edit the number in **Red, Green,** or **Blue.** You can also click the up and down buttons to the right of the value to increase or decrease it. In addition, to use the up and down arrow keys on your keyboard to change a value, click the number that you want to change, then press the up or down arrow key, as desired.

Red A number from 0 to 255 that represents the amount of red. 0 is the minimum and 255 is the maximum.

Green A number from 0 to 255 that represents the amount of green. 0 is the minimum and 255 is the maximum.

Blue A number from 0 to 255 that represents the amount of blue. 0 is the minimum and 255 is the maximum.

Items that appear in all tabs

Preview Shows the selected color in a number of contexts. The right-most diagram in this pane can show two colors:

- **Top color** – The color that is currently selected for the track.
- **Bottom color** – The color that is currently selected to replace it.

OK Accepts any changes that you made to the color, and closes the dialog box.

Cancel Closes the dialog box, and discards any changes that you made to the color.

Reset Restores the color settings in the dialog box to what they were before you made any changes. The dialog box remains open.

Create Experiment

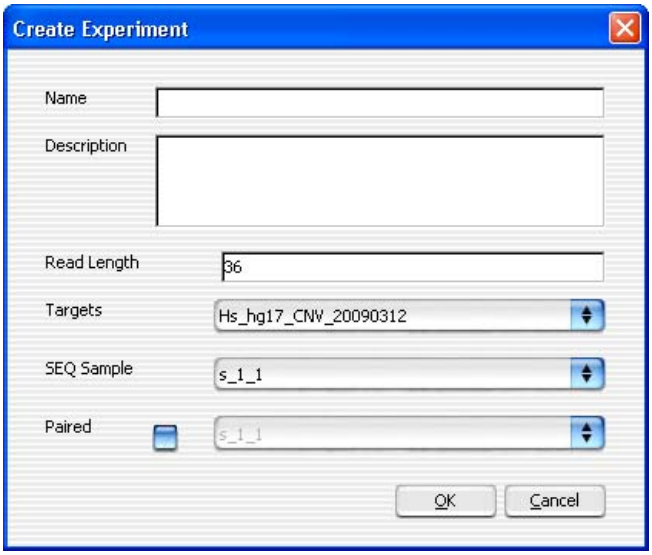


Figure 21 Create Experiment dialog box

Purpose: Lets you set up a QC analysis experiment. You use this dialog box to name the experiment, and to select a sequence read file and a set of target intervals for the analysis. See "To create a new experiment" on page 36.

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

Name Type a name for the experiment. This name identifies the experiment in the Experiment pane of the Navigator.

Description (Optional) Type a brief description of the experiment.

Read Length Type the number of bases in each read in the selected read file. The program assumes all reads in the file have the same length.

Targets Shows a list of all of the annotation tracks that are available in the program. Select the track that contains the desired target intervals for analysis.

SEQ Sample Shows a list of the names of all imported sequence read files that are available for analysis. Select the desired sequence read file.

Paired If you have two output files in which the reads are paired, mark this option, then select the paired file. The program imports the two files separately, then merges them before analysis.

OK Creates a new experiment with the given name that contains links to the selected sequence read file and target interval track. Creates a new folder in the Experiment pane of the Navigator that bears the name of the experiment.

Cancel Closes the dialog box, but does not create a new experiment.

NOTE Before you create an experiment, you must import the desired sequence read file, target file, and chromosome mapping file into the program. See "Importing files" on page 30.

Export Tracks

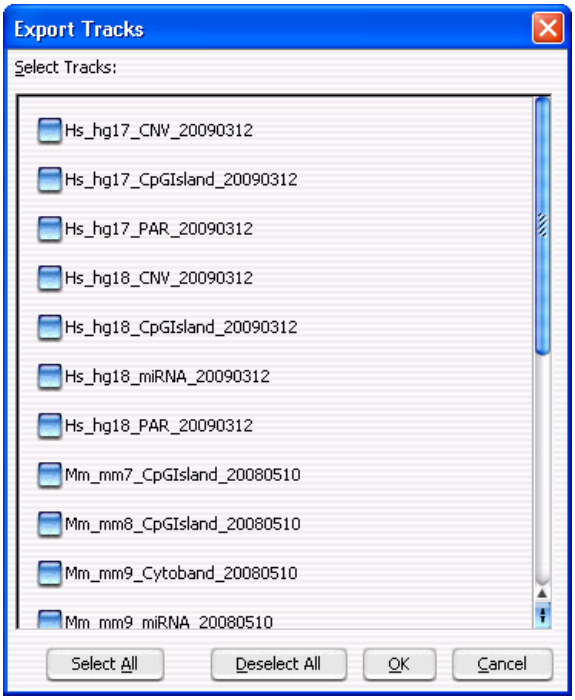


Figure 22 Export Tracks dialog box

Purpose: Lets you select one or more tracks for export as a single BED format track file. You can use this functionality combine multiple tracks, and also to transfer a track to another computer. In addition, you can view the track in any compatible genome browser. See "To export tracks" on page 42.

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

Select Tracks Shows a list of all of the tracks that are available in the program. Mark the check box next to a track to select it for export.

Select All Marks all tracks for export.

Deselect All Clears the check boxes next to all tracks.

OK Opens an Export dialog box, where you can select a location for the exported BED format track file.

Cancel Closes the dialog box, but does not export any tracks.

Go to Gene/Genomic Location

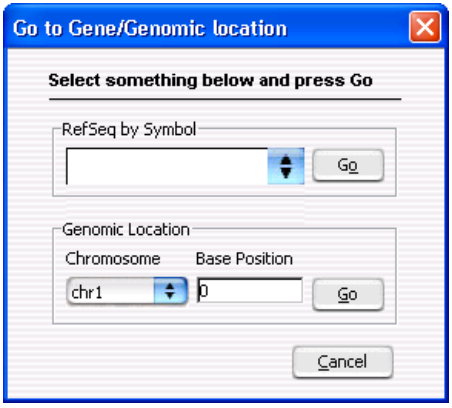


Figure 23 Go to Gene/Genomic Location dialog box

Purpose: Moves the cursor in the Genomic Viewer to a specific location. You can enter either a gene symbol or a genomic location. See "To customize the appearance of an experiment in the Genomic Viewer" on page 44.

To open: In the **Home** tab, click **Go to Gene/Genomic Location.**

RefSeq by Symbol Type a valid RefSeq gene symbol.

Go – Moves the cursor to the location of the gene that you entered.

Genomic Location **Chromosome** – Select the desired chromosome.

Base Position – Type the desired base position on the selected chromosome.

Go – Moves the cursor to the specified genomic location.

Cancel Closes the dialog box, but does not move the cursor.

Import Genome Build

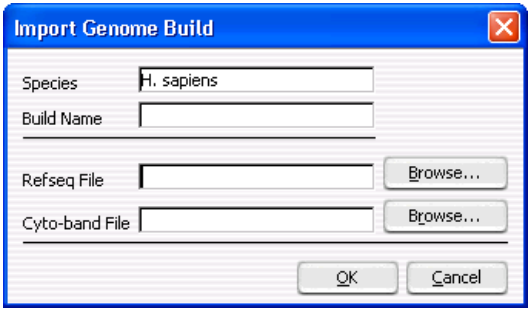


Figure 24 Import Genome Build dialog box

Purpose: Lets you import a new genome build into the program. See "To import a genome build" on page 35. Genome builds contain the names and locations of genes and cytobands. However, they do not contain any nucleotide sequence data.

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

Species The species represented by the new genome build.

Build Name A name for the genome build, as you want it to appear in the program.

RefSeq file The location of the file that contains gene symbol information for the genome.

Browse – Opens an Open dialog box, where you can select the desired file.

Cytoband file The location of the file that contains cytoband information for the genome.

Browse – Opens an Open dialog box, where you can select the desired file.

OK Imports the genome build.

Cancel Closes the dialog box, and does not import a genome build.

CAUTION Use only Agilent-supplied genome build files. You can download these files from the eArray Web site. In your individual user workspace in the CGH or ChIP application types, in the Probe tab, click **DNA Analytics Download.** For more information, see the online help on the eArray Web site.

Import Track

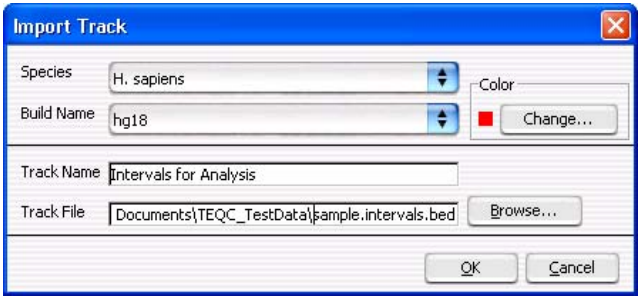


Figure 25 Import Track dialog box

Purpose: Lets you import a BED format track file into the program. For SureSelect Quality Analyzer, this is how you import target genomic intervals for analysis experiments. See "To import target genomic regions as a track" on page 33.

To open: In the **Quality Analyzer** tab, click **Import > Track.**

Species Shows a list of the species for which genome builds are available in the program. Select the desired species.

Build Name Shows a list of the genome build(s) that are available for the selected species. Select the desired genome build. If the desired genome build does not appear in the list, you must import it. See "To import a genome build" on page 35.

Color The color swatch shows the currently selected display color for the bars that represent the target intervals in Gene View.

Change – Opens the Choose Track Color dialog box, where you can select a new color for the track. See "Choose Track Color" on page 77.

Track Name The program uses this name to identify the track in Gene View, and also in lists. Type a name for the new track.

Track File The location of the BED format track file to be imported.

Browse – Opens a dialog box, where you can import the file.

OK Imports the track and closes the dialog box.

Cancel Closes the dialog box, and does not import a track.

Specify Input Information

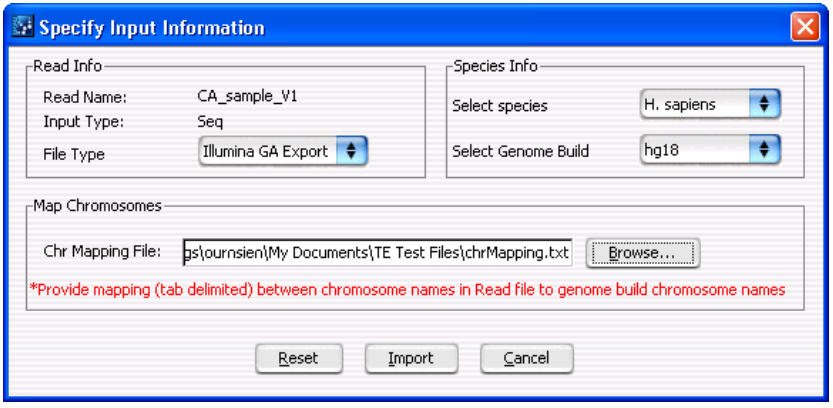


Figure 26 Specify Input Information dialog box

Purpose: Lets you define the attributes of a sequence read file as you import it. See "To import sequence read files" on page 30.

To open: In the **Quality Analyzer** tab, click **Import > Reads.** In the dialog box that appears, select a *.zip file that contains the desired sequence read file, then click **Open.**

Read Info

Read Name (Read-only) The name of the selected sequence read file.

Input Type (Read-only) **Seq** appears, which denotes a file that contains sequence reads.

File Type Select the specific data format that applies to the imported file. The program can import *.zip archives that contain these types of files:

File Type	Comments
Illumina GA Export	Revised version of the ELAND file format that contains additional fields, including quality scores for all bases.
AB SOLiD ma	Color space *.ma file produced by the Applied Biosystems SOLiD system. The AB SOLiD alignment software produces this CSFASTA format file, and places sequence alignment information in the FASTA headers. Color calls are converted to bases when the file is imported.
MAQ aln.txt	The aln.txt output file from the MAQ (Mapping and Assembly with Qualities) program. This program is specifically designed for Illumina-Solexa/AB-SOLiD reads. For more information, go to maq.sourceforge.net

NOTE Import sequence read files as *.zip archives. Do not use the gzip format.

Species Info

Select Species Select the species that is represented in the imported sequence read file.

Select Genome Build The available genome builds for the selected species appear in the list. Select the build that applies to the reads in the imported file.

Map Chromosomes

Because the names of chromosomes in the imported read file can differ from the names of chromosomes in the selected genome build, you must import a chromosome mapping file that relates the chromosome names in these two locations.

Chr Mapping File The name and location of the applicable chromosome mapping file.

Browse – Opens an Open dialog box, where you can select the desired chromosome mapping file.

File format information: Create a tab-delimited text file (*.txt) that contains the desired mapping information. Enter one chromosome per line. Give the name of the chromosome in the read file followed by the name of the matching chromosome in the genome build, separated by a tab character. Do not include a header row.

Example file content:

```
chrA.fa     chrA
chrB.fa     chrB
chrC.fa     chrC
```

Other commands

Reset Opens a dialog box that asks if you are sure you want to reset the mapping. If you click Yes, the program restores the settings in the Specify Input Information dialog box to what they were before you made any changes.

Import Imports the sequence read file into the program, and makes it available for analysis.

Cancel Closes the dialog box, but does not import the read file.

Track

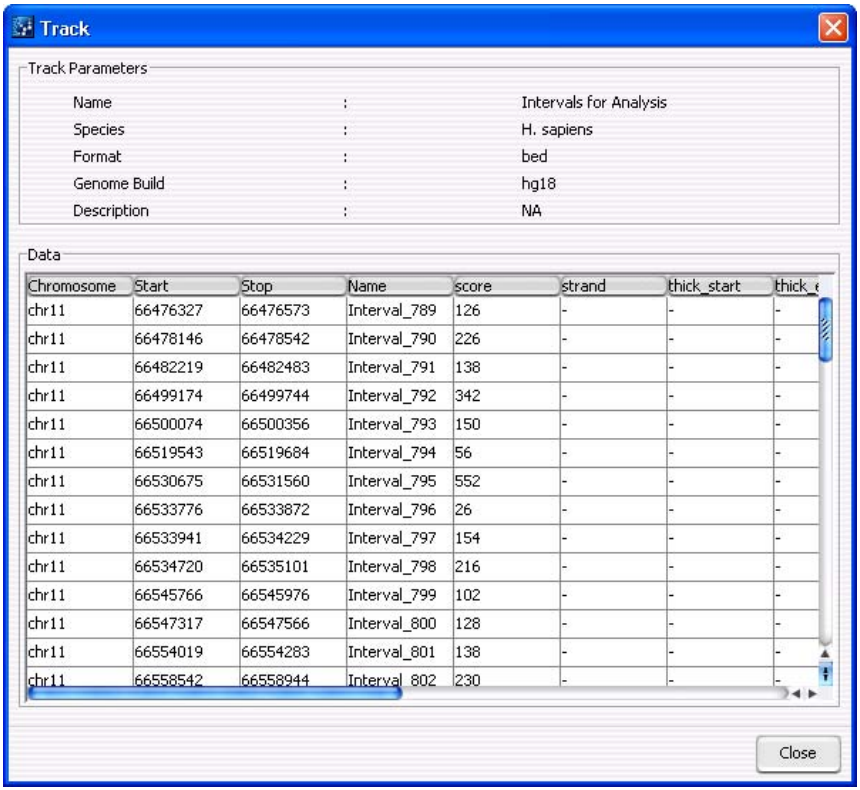


Figure 27 Track dialog box

Purpose: Lets you view the attributes of a specific track that is available in the program. You can view header information, and a list of all of the intervals in the track. All information is read-only. See "To view the intervals in a target interval track" on page 42.

To open: In the Experiment pane of the Navigator, right-click the name of the target intervals track within a specific experiment, then click **View Details.**

Alternatively, in the **Home** tab, click **User Preferences.** In the list of available tracks, click **Details** in the row of the desired track.

Track Parameters

This header information appears:

Item	Description
Name	The name of the track
Species	The name of the species to which the intervals in the track apply
Format	The format of the imported track file. The program supports BED format track files
Genome Build	The genome build of the applicable species upon which the track is defined
Description	A description of the track, if one is available.

Data

The Data pane of the dialog box shows a list of the intervals in the track, and all of the available BED file columns that are available for them. These columns can appear:

Column	Description
Chromosome	The chromosome on which the given interval is located.
Start	The location of the first base pair in the interval.
Stop	The location of the last base pair in the interval.
Name	The name of the interval that appears when the program displays the track.
Score	A number that represents how darkly the feature appears (higher numbers give darker results). The score can be a number from 0 to 1000.
Strand	The strand to which the annotation applies, which can be + (sense) or – (antisense)
thick_start	Each genomic feature can contain a thickly drawn region. This value is the starting position of this region.
thick_end	The ending position of the thickly drawn region of the genomic feature.

Column	Description
item_rgb	The display color of the data for the interval. The color is defined in the RGB color space. **Example:** An RGB value of 255, 0, 255 produces a bright fuchsia color.
block_count	The number of separate blocks for the given region. Typically, blocks refer to exons.
block_size	The size of each block. Multiple values are separated by commas.
block_starts	The starting position of each block relative to the location defined in **Start.** Multiple values are separated by commas.
exp_count	These columns use experiment scores and IDs to set the colors of a track. They are not relevant to SureSelect Quality Analyzer.
exp_Ids	
exp_scores	

Close Closes the dialog box.

User Preferences – Miscellaneous tab

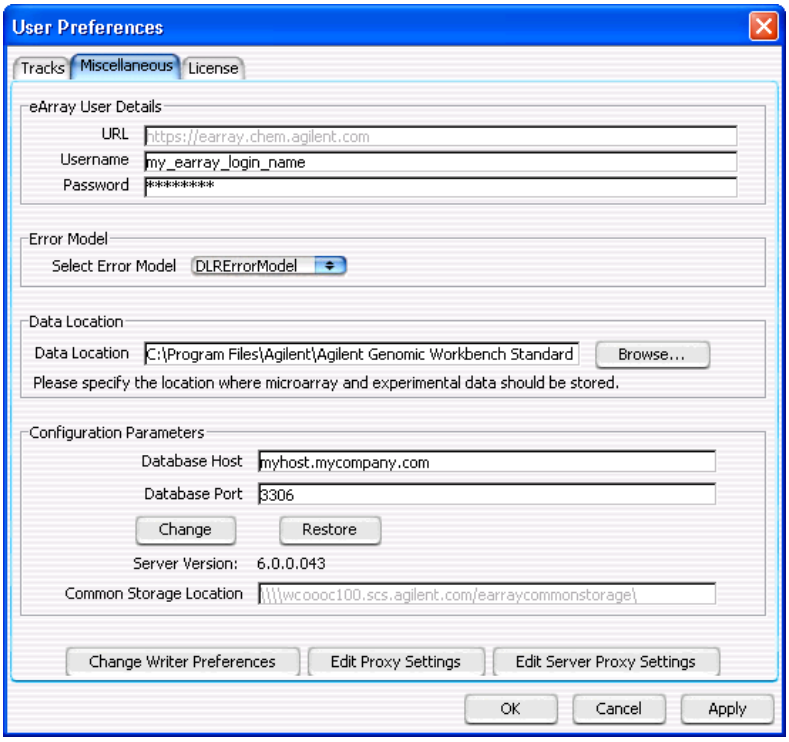


Figure 28 User Preferences dialog box – Miscellaneous tab

Purpose: For SureSelect Quality Analyzer, this dialog box lets you select a default location for analysis results and imported files. See "To select where data and results are stored" on page 19. This dialog box also lets you enter certain user preferences for other programs within Agilent Genomic Workbench.

To open: In the **Home** tab, click **User Preferences,** then click the **Miscellaneous** tab.

Alternatively, within the Genomic Viewer, right-click anywhere in Genome or Chromosome Views, or within the gene display area in Gene View, then click **User Preferences.** In the User Preferences dialog box, click the **Miscellaneous** tab.

Data Location Shows the default location for imported sequence read data and related files. The program also stores microarray sample data and analysis results for the CGH, ChIP-on-chip, and CH3 applications in this location.

Browse – Opens an Open dialog box, where you can select a default location.

Other Options All other options in this dialog box let you enter installation and configuration options for Agilent Genomic Workbench as a whole, or options that are specific for parts of the program other than the Target Enrichment Quality Analyzer.

OK Accepts any changes, and closes the dialog box.

Cancel Closes the dialog box, but does not save any changes.

Apply (Available after you make change(s) to any tab of the Preferences dialog box) Applies the changes that you made in any tab of the dialog box, but keeps the dialog box open.

User Preferences – Tracks tab

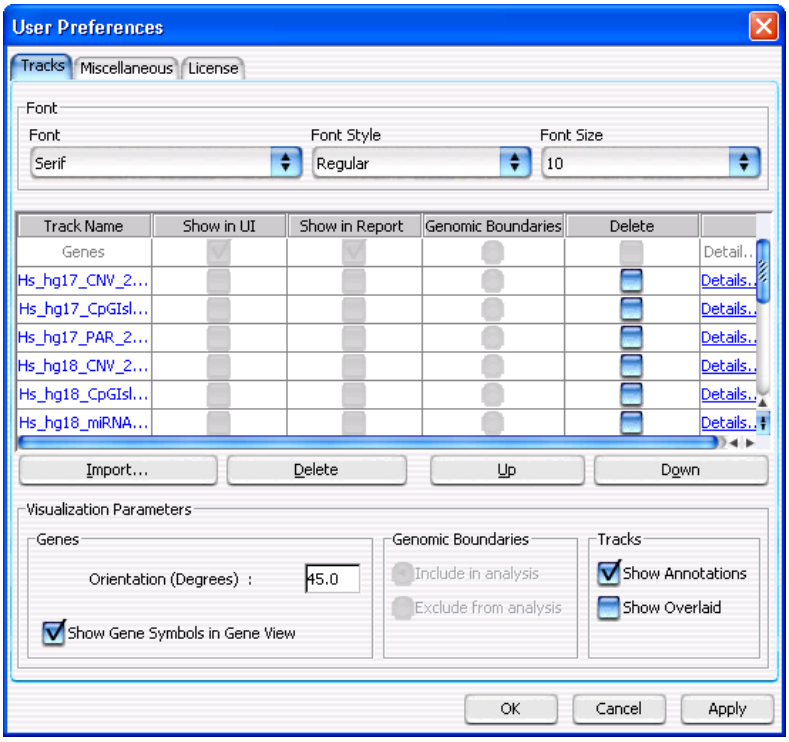


Figure 29 User Preferences dialog box – Tracks tab

Purpose: Lets you import and delete tracks from the program. Also lets you view the details of tracks and customize how genes and tracks appear in Gene View.

To open: In the **Home** tab, click **User Preferences.**

Alternatively, right-click anywhere within Genome or Chromosome Views, or within the gene display area in Gene View, then click **User Preferences.**

Font

Sets the font, style, and size of gene names and interval names in Gene View. These settings are available:

Setting	Description
Font	Select the desired font family.
Style	Select **Regular, Bold, Italic,** or **Bold Italic.**
Size	The size (in points) of gene and interval names. This is an absolute measure—the size of names does not change as you zoom in or out in Gene View.

NOTE These settings do not affect the appearance of the nucleotide sequences of reads, which can vary in size as you change the zoom level.

List of Tracks – Columns

Shows a list of the tracks that are available in the program.

Track Name The name of each track.

Show in UI For the SureSelect Target Enrichment application type, the check boxes in this column are unavailable. The program automatically marks only Genes and the target interval track that has been defined for the selected experiment.

Show in Report This column is not available for the SureSelect Target Enrichment application type.

Genomic Boundaries The program automatically selects the target interval track that has been defined for the selected experiment.

Delete Mark a check box to select a given track for deletion. To delete the selected track(s), click **Delete.**

Details Opens the Track dialog box for the given track, where you can view a list of intervals in the track, as well as the header and other data that is associated with the track. See "Track" on page 90.

List of Tracks – Buttons

Import Opens the Import Track dialog box, where you can select a BED format track file, and configure the track before you import it. See "Import Track" on page 86.

Delete Permanently removes selected tracks from the program. To select a track for deletion, mark its check box in the Delete column of the list of tracks.

Up, Down Changes the display order of tracks in Gene View. These commands are not relevant to the SureSelect Target Enrichment application type.

Visualization Parameters

Genes **Orientation (Degrees)** – The angle at which gene and interval names appear in Gene View. A value of 0° orients the names horizontally, and higher values rotate the name in a counterclockwise direction, up to a maximum of 360°.

Show Gene Symbols in Gene View – If you mark this option, the names of genes, as well as blue bars that show their genomic locations, appear in Gene View. If you clear this option, neither of these appears.

Genomic Boundaries These options are not available for the SureSelect Target Enrichment application type.

Tracks Mark any of these options:

- **Show Annotations** – In the target intervals track, displays the names of target intervals along with bars that represent their genomic locations.
- **Show Overlaid** – This option is only relevant to other programs in Agilent Genomic Workbench that can display more than one track at a time.

Commands that appear in all tabs

OK Accepts the change(s) that you made in any tab of the dialog box, and closes the dialog box.

Cancel Closes the dialog box, but does not make any changes.

Apply (Available if you have made changes to one or more preferences) Applies any change(s) that you have made in any tab of the dialog box, and leaves the dialog box open.

View Preferences

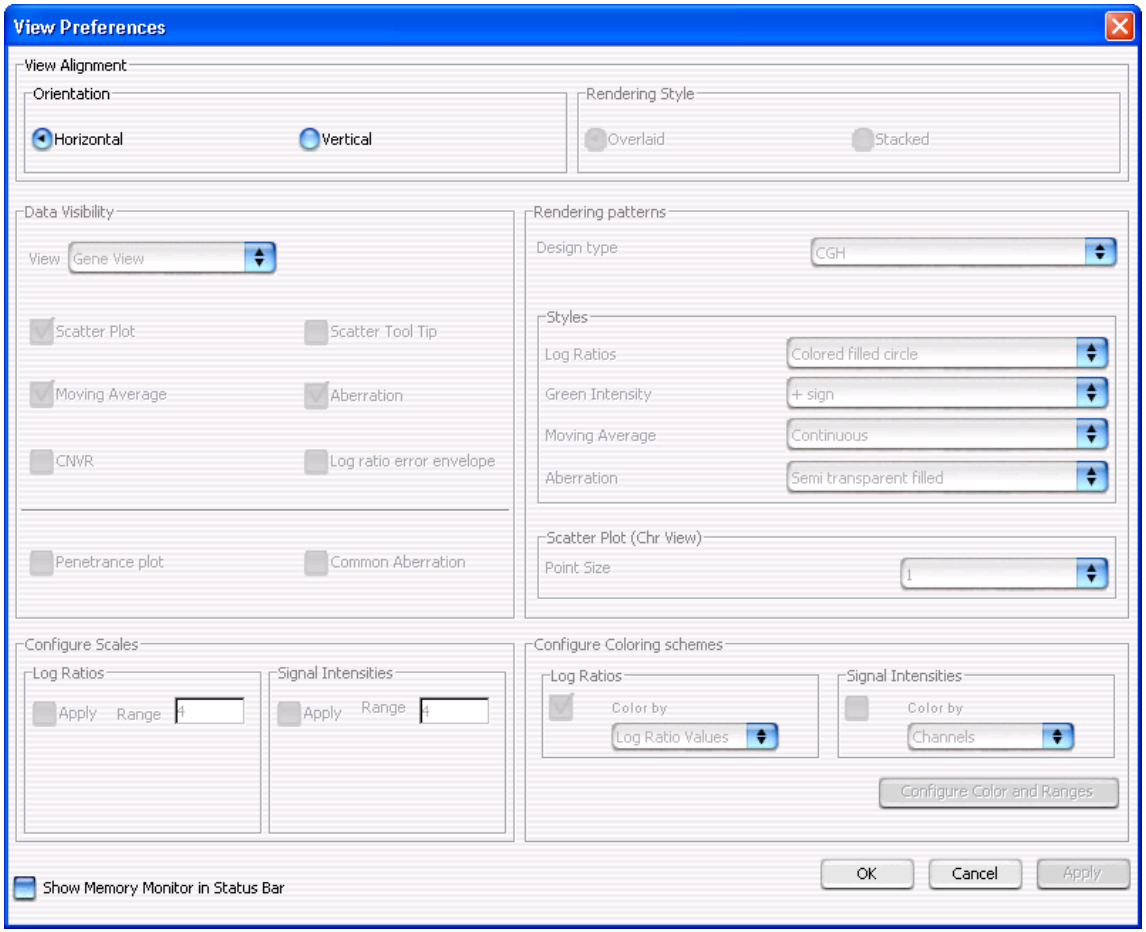


Figure 30 View Preferences dialog box

Purpose: For the SureSelect Target Enrichment application type, this dialog box lets you switch the Genomic Viewer between horizontal and vertical orientation.

To open: Click **View > View Preferences.**

View Alignment

Orientation The layout of Genome, Chromosome, and Gene Views. Select one of these options:

- **Horizontal** – Stacks the three views horizontally, from top to bottom. Because of the oblong shape of a typical monitor, Horizontal orientation can be somewhat more useful when you use the program to view sequence reads and the target interval track.
- **Vertical** – Stacks the three views vertically, from left to right.

In both cases, Tab view appears horizontally across the bottom of the Genomic Viewer.

Other commands

OK Accepts any changes you made, and closes the dialog box.

Cancel Closes the dialog box, but does not accept any changes.

Apply (Available only if you changed the orientation) Changes the orientation of the Genomic Viewer, and keeps the dialog box open.

No other options in this dialog box are applicable to the SureSelect Target Enrichment application type.

Target Enrichment reports

This section describes the analysis results that are produced by SureSelect Quality Analyzer. These results/reports are available:

Result/Report	Description/Comments
Summary – PDF	For this report, the program creates a PDF format report that contains tables of QC metrics. It also contains the results of read depth analysis shown both as tables and as a graph. See these topics: • "To view and save reports" on page 49 • "QC metrics" on page 101 • "Read depth distribution" on page 103
Summary – Text	For this report, the program creates a *.txt file that contains a table of QC metrics and the results of read depth analysis. You can open this file with a text editor or word processor. See these topics: • "To view and save reports" on page 49 • "QC metrics" on page 101
Enrichment Analysis– UCSC	For this report, the program creates a track file that contains enrichment analysis results for the chromosome that is currently selected in Genome View, and uploads it to the UCSC Genome Browser. See these topics: • "To view and save reports" on page 49 • "Enrichment analysis results – UCSC" on page 104
Enrichment Analysis – XLS	For this report, the program creates a *.xls file that contains enrichment analysis results for the chromosome that is currently selected in Genome View. You can view this file in Microsoft Excel. See these topics: • "To view and save reports" on page 49 • "Enrichment analysis results – XLS report" on page 105

QC metrics

The table below describes the QC metrics that appear in both the PDF and *.txt format reports that are produced by Target Enrichment Quality Analyzer.

Metric	Description/comments
Total reads (good and bad)	Total number of reads contained in the read file in the selected experiment.
Total HQ uniquely mapped reads	The total number of high quality (HQ) input reads that map to a single genomic location. HQ reads are those that meet the quality criteria in SureSelect Quality Analyzer.
Read length	The number of bases in each of the input reads. SureSelect Quality Analyzer assumes all reads in the input file have the same length.
Total number of bases mapped	The total number of bases to which the reads in the selected experiment map. Typically, not all bases map to genomic regions.
Number of reads in target regions	The number of reads that overlap the regions that are defined in the target intervals track in the selected experiment. The program considers a read to be in a targeted region if the midpoint of the read falls within the targeted region.
Percent of reads in target regions	The fraction of the total reads that overlap one of the regions that are defined in the input target intervals track. The program considers a read to be in a targeted region if the midpoint of the read falls within the targeted region.
Percent of genome targeted	The fraction of the genome that is represented by the regions that are defined in the target intervals track in the selected experiment.
Enrichment in target regions	The overall fold-increase in representation of the target intervals over that expected for a completely random process.
Average read depth	The average number of reads that cover each targeted region. The program considers a read to be in a targeted region if the midpoint of the targeted region falls within the read.

Metric	Description/comments
Uniformity (3/4 mean with upper tail)	A measure of how evenly reads are distributed. It is the percentage of all reads that fall in regions where the read depth is within a window defined by ±75% of the mean, around the mean, plus the entire upper tail of the distribution. Effectively, this is the percentage of all reads that fall in regions where the read depth is at least 25% of the mean read depth. **Example:** If the mean read depth is 38.42 reads, this metric gives the percentage of reads that fall in regions where the read depth is at least 9.61 reads. As the distribution of reads becomes broader, the Uniformity metric decreases. This metric can be useful in applications where you need a certain minimum level of coverage—for example, to make SNP calls. It lets you know the fraction of reads that fall in regions where the coverage is close to the mean, or above.
Number of target regions	The number of intervals that were defined in the target intervals track in the selected experiment.
Number of target regions covered by at least **n** read(s)	The number of regions in the target intervals track in the selected experiment that are covered by at least the indicated number of reads. Totals appear for n = 1, 5, 10, and 20.
Percentage of target regions covered by at least **n** read(s)	The percentage of regions in the target intervals track in the selected experiment that are covered by at least the indicated number of reads. Percentages appear for n = 1, 5, 10, and 20.
Number of target bases covered by at least **n** read(s)	The number of bases in targeted regions that are covered by at least the indicated number of reads. Totals appear for n = 1, 5, 10, 20, and 1000.
Percentage of target bases covered by at least **n** read(s)	The percentage of bases in targeted regions that are covered by at least the indicated number of reads. Totals appear for n = 1, 5, 10, 20, and 1000.
Percent target regions with zero coverage	The percentage of target regions that have no reads that overlap them by at least 1 bp.

Read depth distribution

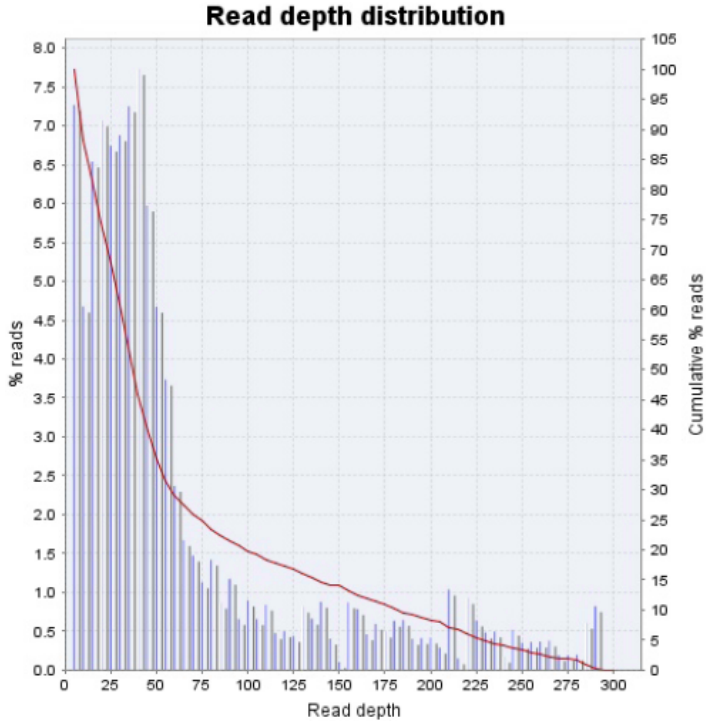


Figure 31 Distribution of read depths

The PDF format SureSelect QC report contains a graph that shows the distribution of read depths for the selected experiment. The program partitions the range of read depths into a fixed number of bins.

Bars (% Reads) – Shows the percentage of reads in each read depth bin. In general, a desirable distribution is gaussian, skewed toward higher values, with the peak at the desired level of coverage.

Red line (Cumulative % reads) – Shows the cumulative distribution function for read depths for the selected experiment. This function shows the probability that the read depth will be greater than or equal to a given depth.

Enrichment analysis results – UCSC

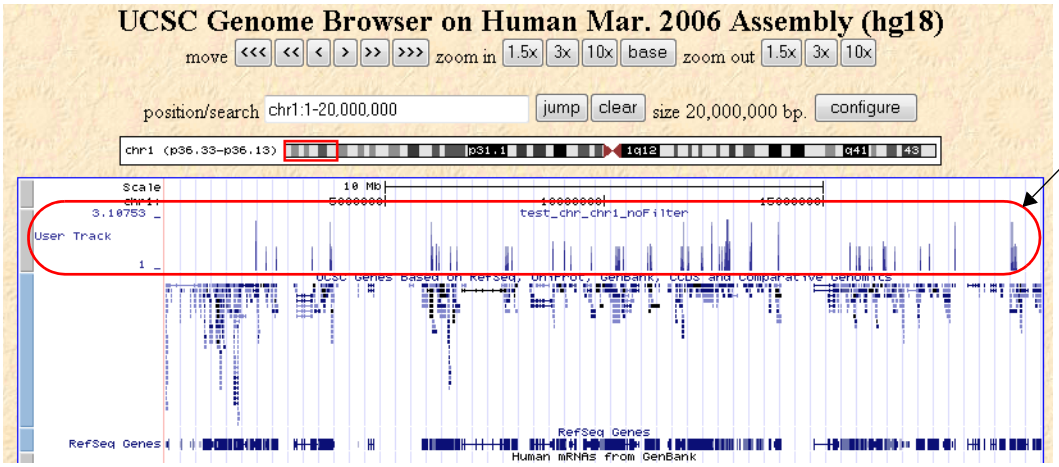


Figure 32 UCSC Genome Browser – Enriched peaks appear in the User Track area

The program displays enriched peaks for the selected chromosome as a histogram in the User Track area of the UCSC Genome Browser. It lets you see the fold-increase in the representation of each interval over that expected for a random process. See "To view and save reports" on page 49.

When you ask the program to display enrichment analysis results, it creates a WIG format file for the currently selected chromosome and transfers it to the UCSC genome browser. The UCSC Genome Browser opens in your Web browser.

You can also create an *.xls format file that contains a list of all of the enriched peaks for the selected chromosome. See "Enrichment analysis results – XLS report" on page 105.

Enrichment analysis results – XLS report

	A	B	C	D	E	F
1	Read File: s_1_1_export_top					
2	Target Regions: Example Targets					
3	Chromosome Definition: chrMapping.txt					
4	Chromosome: chr1					
5	Report Generated By: SSQA User					
6						
7	Interval Number	Chromosome	Start	Stop	Mid Point	Enrichment
8	1	chr1	879221	879520	879371	1
9	2	chr1	881276	881575	881426	1
10	3	chr1	947568	947867	947718	1
11	4	chr1	970915	971214	971065	1
12	5	chr1	971222	971521	971372	1
13	6	chr1	1129703	1130002	1129853	1
14	7	chr1	1142911	1143210	1143061	1
15	8	chr1	1169203	1169502	1169353	1
16	9	chr1	1182149	1182448	1182299	1
17	10	chr1	1221278	1221577	1221428	1
18	11	chr1	1223817	1224116	1223967	1

Figure 33 Partial view of an enrichment analysis report (XLS format)

This report contains the results of enrichment analysis for the selected chromosome, organized by interval. You view this report in Microsoft Excel.
To create this report, see "To view and save reports" on page 49.

You can also view enriched intervals graphically in the UCSC Genome Browser. See "Enrichment analysis results – UCSC" on page 104.

Header information

Item	Description
Read File	The name of the sequence read file that was analyzed.
Target Regions	The name of the target intervals track that was used to analyze the read file. You supply this name when you import the track. See "To import target genomic regions as a track" on page 33.
Chromosome Definition	The name of the chromosome definitions file. This is a *.txt file that contains the names of the chromosomes that are referenced in the read file, correlated with the names of the chromosomes in the desired genome build of the species of interest. For details, see "To import sequence read files" on page 30.
Chromosome	The chromosome that was selected for enrichment analysis. Before you request an enrichment analysis report, you select the desired chromosome in Genome View of the Genomic Viewer.
Report generated by	The name and login ID of the person who requested the enrichment analysis report.

Data columns

Below the header information, these columns of data appear:

Column	Description
Interval Number	Intervals are numbered sequentially.
Chromosome	The name of the chromosome that contains the enriched interval
Start	The location of the first base of the enriched interval in the given chromosome
Stop	The location of the last base of the enriched interval in the given chromosome
Mid Point	The location of the midpoint of the enriched interval in the given chromosome
Enrichment	The fold-increase in representation of each interval over that expected for a purely random process.

www.agilent.com

In this book

This User Guide describes how to use the SureSelect Quality Analyzer application of Agilent Genomic Workbench to assess the quality of fragment pull-down for target enrichment experiments.

Revision A, February 2010

G3800-90003

Agilent Technologies