

# GeneSpring GX

# **Quick Start Guide**

The information in this guide is for use with GeneSpring GX.

# What is GeneSpring GX?

GeneSpring GX is a powerful visualization and analysis solution designed for use with genomic expression data. It allows you to display and analyze large data sets on your personal computer. GeneSpring GX can analyze data from many sources, including academic, biotechnology, and pharmaceutical organizations.

The GeneSpring GX platform is designed to break through bottlenecks in the analysis process and to help identify genes/pathways that are truly relevant to your biological question by comparing analysis results from expression, genotyping, protein, metabolite and other data types. The program allows you to:

- Identify targets quickly and reliably
- Uncover statistically meaningful results, using techniques such as 1-way and 2-way ANOVA, multiple testing corrections, false discovery rate prediction, and Tukey and Student-Newman-Keuls post hoc tests
- Display expression data with customizable visualization tools
- Create complex experiments that link trends in expression data to a variety of test parameters. This allows you to test complex hypotheses by running previously defined analyses against different combinations of samples
- Ensure data quality with powerful statistical tools and normalization methods
- Automate complex analysis tasks using scripts
- Import pathways from many sources, including native import from the Kyoto Encyclopedia of Genes and Genomes (KEGG)



- Import and compare expression data from multiple platforms
- Export data and images in standard formats (PICT and PNG)

# What's New in GeneSpring GX

The major new features in this release are:

- Ontology Browser for the GO ontology
- Filter by volcano plot
- Box plots

Other new features include:

- FE import
- GCOS integration package
- Tree improvements
- Import of custom pathways and GenMAPP pathways
- Pathway Inspector improvements

# **GeneSpring GX Workflow**

The following diagram shows the steps that can occur in a typical data analysis session using GeneSpring GX. Use this as a guideline for analyzing your data.



GeneSpring GX has additional features that are not contained in this diagram.

# **Terminology**

The following terms are used in GeneSpring GX. Some terms have slightly different meanings than you are used to.

#### Genome

A **genome** in GeneSpring GX contains information about all the genes in your chip or microarray setup. Note that a GeneSpring GX genome *doesn't* correspond exactly to the biological definition of a genome. A genome in GeneSpring GX is composed of discrete genes rather than the full nucleotide sequence. Therefore, it can contain two genes that represent alternately spliced variants of a single gene, whereas a true genome would include the DNA sequences for only one.

#### **Parameters**

**Parameters** in GeneSpring GX are the variables in experiments such as age, weight, LDL level, blood pH, or IQ. Parameter values are values assigned to experiment parameters. For example, Embryonic, Postnatal or Adult could represent parameter values of the experiment parameter "Stage", while .01 ppm could represent a parameter value of the experiment parameter "Concentration".

#### Replicates

**Replicates** in GeneSpring GX are experiments repeated with the same sample. They provide a measure of the experimental variation, such as:

- multiple spots on the same array representing the same gene (i.e. a copy)
- the same sample on more than one array
- a biological replicate (equivalent samples taken from more than one organism)

Graphically, a parameter that is defined as a replicate is a *hidden variable*; such that no visual distinction is made based on this parameter or its parameter values.

#### Regions

**Regions** in GeneSpring GX divide data into specific sections. This allows you to normalize sections of an array separately, rather than normalizing across the entire data set.

#### **Raw Data**

The analysis process begins by obtaining data in the form of flat files that were generated by your scanning software or other expression analysis technology. GeneSpring GX is capable of recognizing most commercially-available formats and can be customized to work with other formats, as necessary. Typically, the gene, spot, or probe-set intensity values in these files are referred to as **raw data**.

#### **Normalized Data**

If GeneSpring GX recognizes your file format, it applies a set of default normalizations appropriate for your expression analysis technology. The denominator used to normalize each measurement is referred to as the control strength.

#### **Interpreted Data**

GeneSpring GX can interpret normalized data in many different ways. Multiple samples can be treated as replicates and averaged. You can set assumptions to determine the precision of averaged values. Normalized data can be displayed and analyzed using ratio, log of ratio, or fold change modes. The mode you select affects the graphical display of normalized values and the numbers used for all analyses such as clustering. Numbers that are displayed as text, such as in the Gene Inspector window, and those entered as parameters for analyses, such as in the Filter Genes tools, are always in ratio mode.

#### Flags

**Flags** are additional measurement markers in your data set that can be used in subsequent analyses. They can be assigned as present, marginal, unknown, or absent.

# **Starting GeneSpring GX**

- **1** Install GeneSpring GX according to the instructions given in the *GeneSpring GX Installation Manual.*
- **2** Double-click the GeneSpring GX icon on the desktop.



**3** A Helpful Hint is displayed that tells you how to load data. Click **OK** to close the Hint window.

The GeneSpring GX main window appears, with all the genes in your experiment displayed. Genome



**Navigator Pane** 

- **4** To see your genome, select **Open Genome or Array** from the **File** menu and select your genome from the submenu.
- **5** Use the following features to view and manipulate your data in GeneSpring GX:
  - "Menu Commands in the GeneSpring GX Main Window" on page 7
  - "Navigator Pane" on page 20
  - "Genome Browser" on page 25
  - Genome Inspector window, see "To inspect the genome" on page 32

# Menu Commands in the GeneSpring GX Main Window

File	Edit	View	Experiments	Filtering	Tools	
Import Data Import from Database Open Genome or Array View Projects Genome Man-	Copy >Gene List >Experiment >Annotated Gene List Paste >Gene List >Classification	Graph >Line Graph >Bar Graph >Graph by Genes >Box Plot Physical Position Blocks Tree Array Layout	Experiment Inspector Experiment Nor- malizations Experiment Parameters	Filter on: >Expression Level >Fold Change Error >Confidence >Parameter >Volcano Plot	Statistical Analy- sis (ANOVA) Clustering Class Prediction Find Potential Regulatory Sequences Principal Compo-	
ager Import Genome	>Experiment	Pathway Ordered List Scatter Plot	Cross-Gene Error Model	Filter on: >Flags >Data File	nents Analysis Find Similar Samples	
New Window New Linked Win-	Undo Edit Gene List	3D Scatter Plot Compare Genes	Experiment Inter- pretation	>Arbitrary File	Find Significant Parameters	
dow Login to Work- group Server Bulk Upload to	Find Gene Find Next Gene Advanced Find	to Genes Spreadsheet Condition Scatter- plot	Create New Experiment Duplicate Experi-	Filter on Gene List Numbers Advanced Fil- tering	Find Similar Genes Draw Expres- sion Profile	
Workgroup Server Copy Genome from WorkGroup Server	Gene Inspect Selected Gene	Zoom In Zoom Out Zoom Fully Out Unsplit Window	ment Sample Manager	Annotations	Check Remote Execution Queue Script Editor	
New Pathway New Program Import Gene-	Standard	Animate Animate Second- ary Visible Remove Second- ary Gene List Display Options Show Average of	Animate Animate Second- ary Visible Remove Second- ary Gene List Display Options Show Average of	Colorbar Color by:	Annotations from Agilent Build Homol- ogy Tables	
Spring GX Zip Load Bookmark File Save Bookmark	Preferences			<ul> <li>&gt;Significance</li> <li>&gt;Venn Diagram</li> <li>&gt;Parameter</li> <li>&gt;Classification</li> </ul>	GO Ontology Browser Make Gene Lists from	
Print Image Save Image		Genes	No Color	Annotations Build Ontology Edit Genes and		
Close Quit				Annotations Genome Inspector		

These menu commands are described in the following sections.

## **File Menu Commands**

These commands are available from the **File** menu in the GeneSpring GX main window.

File Menu Command	Description
Import Data	Selects one data file for import
Import Data from Database	Selects and imports data from a database
Open Genome or Array > Affymetrix	Opens an array from Affymetrix
Open Genome or Array > Demo Chip	Opens a genome from Silicon Genetics
View Projects	Lists projects available for display. Not Assigned - Lists objects, such as gene lists or experiments, that are not assigned to a project
Genome Manager	Opens the Genome Manager window, which allows you to inspect, delete, and rename genomes or arrays. It also allows you to log in to Workgroup Server, upload data; and import and export zip files to and from GeneSpring GX.
Import Genome	Imports a standard genome or creates a cus- tom genome
New Window	Creates a separate copy of the current window which allows you to work independently in that window. Changes that you make in the new window are not implemented in the original window.
New Linked Window	Creates a copy of the current window that allows you to select one gene or gene list in two windows simultaneously. Changes you make in the linked window are reflected in the original window too.
Login to Workgroup Server	Allows you to log in to Workgroup Server
Bulk Upload to Workgroup Server	Uploads entire genomes or large amounts of data to Workgroup Server
Copy Genome from Workgroup Server	Copies a genome from Workgroup Server and downloads it to GeneSpring GX

File Menu Command	Description
New Pathway > Import New Pathway	Imports an image file from the desktop or another server
New Pathway > Import KEGG Pathways	Imports a folder containing KEGG pathways available from the KEGG FTP site
New Program > New External Program	Opens the New External Program window, which allows you to introduce an external pro- gram for use in GeneSpring GX
New Program > New Interactive Plugin*	Opens the New Interactive Plugin window, which allows you to introduce an interactive plugin for use in GeneSpring GX
New Program > New Script Plugin*	Opens the New Script Plugin window, which allows you to create or view a custom-written script
New Program > New Database Preprocessor Plugin*	Opens the New Data File Preprocessor win- dow, which allows you to introduce a custom program that can be run while importing raw data files in order to perform a special function (like GC-RMA) on the raw data values

\*For information on how to build a plugin program, see the following resources:

•The Plugin API tutorial, which can be accessed by opening the GeneSpring GX API Tutorial.pdf file in the GeneSpring\docs directory

•The GeneSpring API JAVADOC, which can be accessed by opening the index.html file in the API folder in the GeneSpring\docs directory

Import GeneSpring GX Zip	Imports a zip file created previously in Gene- Spring GX, using the Export As Zip command. See "Navigator Shortcut Menu Commands" on page 23 for more information.
Load Bookmark File	Displays the settings in the selected bookmark saved previously in GeneSpring GX. See Save Bookmark below.
Save Bookmark	Saves the current display settings, including experiments, gene list, coloration, and selected genes in a bookmark
Print Image	Prints an image of the Genome Browser, Col- orbar, or Browser and Colorbar together

File Menu Command	Description
Save Image	Saves the image of the Genome Browser or Colorbar in either a PNG or PICT file format to export to other programs
Close	Closes the current GeneSpring GX window
Quit	Ends the current GeneSpring GX session and closes <i>all</i> windows

## **Edit Menu Commands**

These commands are available from the **Edit** menu in the GeneSpring GX main window.

Edit Menu Command	Description
Сору	Copies the selected gene list, experiment, or annotated gene list (see the <i>Paste</i> command)
Paste > Gene List or Experiment	Pastes the gene list or experiment most recently copied (see <i>Copy</i> command)
Paste > Classification	Pastes the classification selected in the Classifications folder in Navigator
Undo	Reverses the last Edit command
Edit Gene List	Opens the Gene List Editor window, which allows you to filter on a gene list, create a gene list, filter on an anno- tation, and show all genes in a filtered list
Find Gene	Opens the Find Gene in View window, which allows you to find a gene based on its systematic name, common name, or synonym
Find Next	Finds the next gene that meets the defined search crite- ria
Advanced Find Gene	Opens the Advanced Find Genes window, which allows you to find a gene using advanced search criteria
Inspect Selected Gene	Opens the Gene Inspector window, with the gene selected in the Browser displayed

Edit Menu Command	Description
Search	Opens the Search window, which allows you to find data by keyword, data by project assignment, data by advanced search criteria, or genes stored locally or on Workgroup Server
Filter Navigator	Opens the Filter Navigator window, which allows you to filter projects by keyword or advanced search criteria
Standard Attributes	Opens the Standard Attributes window, which allows you to modify the standard xml files for sample attributes and experiment parameters
Preferences	Opens the GeneSpring GX Preferences window, which allows you to specify global preferences for data files, database, colors, Browser, firewall, system, Workgroup Server use, computation, and miscellaneous options.

## **View Menu Commands**

These commands are available from the **View** menu in the GeneSpring GX main window.

View Menu Command	Description
Line Graph	Displays one experiment or a set of experiments by plotting the relative expression of each gene against experimental parameters, such as time or drug concentration. Each gene is represented as a line.
Bar Graph	Displays one experiment or a set of experiments by plotting the relative expression of each gene against experimental parameters, such as time or drug concentration. Each gene is represented as a vertical bar.
Graph by Genes	Displays an experiment as one line, where each point on the line represents the relative expression of one gene
Box Plot	Displays data in box plot, which allows you to see the fol- lowing statistical information about a condition or sample: median values, outliers, the quantiles and the distribution for a particular sample or condition
Physical Position	Displays genes according to their physical position (when the gene loci are known and loaded into GeneSpring GX) within the DNA sequence of the organism.

View Menu Command	Description
Blocks	Displays a rectangle for every gene in the active genome, ordered by trust
Tree	Displays the results of hierarchical clustering in the form of a mock phylogenetic tree or dendrogram. In such a tree, genes having similar expression patterns are clustered together.
Array Layout	Produces a synthetic picture of the arrays used in the cur- rent experiment. This view is useful in identifying arrays that display local shifts in intensity due to problems in probe deposition, hybridization, washing, or blocking.
Pathway	Displays an imported gif or jpeg pathway image
Ordered List	Displays a gene list in the order of its associated values
Scatter Plot	Displays the expression levels of genes in two distinct con- ditions, samples, or normalization schemes
3D Scatter Plot	Displays genes in a 3D scatter plot in which each dot represents a gene
Compare Genes to Genes	Displays the similarity between the expression profiles of two genes in one list or in two separate lists
Spreadsheet	Displays gene data in a spreadsheet
Condition Scatter Plot	Displays the results of Principal Components Analysis per- formed on conditions
Zoom In	Expands the view of data in the Browser
Zoom Out	Decreases the detail shown in the view of data in the Browser
Zoom Fully Out	Returns the view to the original data display
Unsplit Window	Rejoins windows that have been split into several small windows
Animate	Shows changes in classification assignments in real time
Animate Secondary	Shows changes in classification assignments in real time when viewing two gene lists or experiments simultaneously
Visible > Picture	Shows or hides the optional picture at the lower right cor- ner of the main GeneSpring GX window

View Menu Command	Description
Visible > Animation Controls	Shows or hides the slider and the Animate box at the bot- tom of the window. Hiding this check box does not disable the Animation feature.
Visible > Magnification	Shows or hides the Magnification feature and the Zoom Out button at the bottom of the GeneSpring GX window. Zoom Out is still available from the menu.
Visible > Secondary Picture	Shows or hides a secondary picture when viewing two gene lists or experiments simultaneously in the Genome Browser
Visible > Secondary Animation Controls	Shows or hides the secondary Animation Controls check box and slider when viewing two gene lists or experiments simultaneously
Visible > Navigator	Shows or hides the Navigator in the GeneSpring GX win- dow
Visible > Hide All	Hides everything in the GeneSpring GX window, except the Genome Browser
Visible > Show All	Shows all elements in the GeneSpring GX window
Visible > Hide All in All Windows	Hides everything in all windows, except the Genome Browser
Visible > Show All in All Windows	Shows all elements in all windows
Remove Secondary Gene List	Removes a secondary gene list from the Navigator
Display Options	Changes display options for views
Show Average of Genes	Displays the centroid of all genes in the view

## **Experiments Menu Commands**

These commands are available from the **Experiments** menu in the GeneSpring GX main window.

Experiments Menu Command	Description
Experiment Inspector	Opens the Experiment Inspector window, which allows you to view more information about the selected experiment
Experiment Normalizations	Opens the Experiment Normalizations window, which allows you to apply a transformation scenario to normalize data
Experiment Parameters	Opens the Experiment Parameters window, which allows you to specify the parameters to use in the experiment
Cross-Gene Error Model	Opens the Cross-Gene Error Model window, which allows you to apply a statistical model to estimate error in control and repeat measurements
Experiment Interpretation	Opens the Change Interpretation window, which allows you to specify the type of analysis to apply, inspect experi- ment details, and display the results
Create New Experiment	Opens a series of windows that allows you to select samples to include in a new experiment
Duplicate Experiment	Opens the Save New Experiment window, which allows you to save a copy of an experiment under a new name
Sample Manager	Opens the Sample Manager window, which allows you to apply different filters to find samples for creating experi- ments, editing attributes, assigning projects, and publish- ing to Workgroup Server

## **Colorbar Menu Commands**

These commands are available from the **Colorbar** menu in the GeneSpring GX main window.

Colorbar Menu Command	Description
Color by Expression	Colors genes according to their normalized expression values and trustworthiness
Color by Significance	Colors genes according to the p-value from the t-test of dif- ferential expression
Color by Venn Diagram	Colors genes based on their membership in one or more gene lists in a Venn diagram
Color by Parameter	Colors genes based on the value of parameters. This scheme is best suited for use with Graph view and Bar Graph view when different conditions are indicated with discrete symbols.
Color by Classification	Colors genes based on a previously-defined classification scheme. You can use a folder of lists to color by classifica- tion or a classification method such as k-means or self-organizing maps
No Color	Displays all genes in gray

## **Filtering Menu Commands**

These commands are available from the **Filtering** menu in the GeneSpring GX main window.

Filtering Menu Command	Description
Filter on Expression Level	Filters genes that have certain values present in some of the conditions or samples in an experiment or interpreta- tion
Filter on Fold Change	Filters genes based on a comparison of two samples or conditions
Filter on Error	Filters data based on the standard deviation, standard error, or range of replicates
Filter on Confidence	Filters data based on a measure of confidence, such as of confidence from the t-test p-value or number of replicates

Filtering Menu Command	Description
Filter on Parameter	Filters genes based on similar parameters
Filter on Volcano Plot	Filters genes on volcano plots, which are constructed using fold change values and p-values and are used to visually represent differential expression between two dif- ferent conditions
Filter on Flags	Filters genes based on quality flags in the original data files
Filter on Data File	Filters genes based on values in a specific column of the experiment data file
Filter on Arbitrary File	Filters genes based on values in a specific column of any data file
Filter on Gene List Numbers	Filters genes according to the numbers associated with them in a gene list, such as correlation coefficients, p-val- ues, or fold change ratios
Advanced Filtering	Provides advanced filters for more fine-grained filtering

## **Tools Menu Commands**

These commands are available from the **Tools** menu in the GeneSpring GX main window.

Tools Menu Command	Description
Statistical Analysis (ANOVA)	Compares mean expression levels between two or more groups of samples
Clustering > K-Means	Produces groups of genes with a high degree of similarity within each group, and a low degree of similarity between groups
Clustering > Gene Tree	Produces dendrograms that show the relationships among the different expression levels of genes over a series of conditions.
Clustering > Condition Tree	Divides genes or conditions into groups that have expres- sion patterns similar to nearby genes in a tree
Clustering >Self-Organizing Map	Divides genes into groups based on expression patterns and illustrates the relationship between groups by arrang- ing them in a two-dimensional map

Tools Menu Command	Description
Clustering > QT Clustering	Identifies clusters of genes, such that each gene in the cluster is within a specified distance (based on a user-defined distance metric) of every other gene in the cluster
Class Prediction	Predicts the value (or "class") of an individual parameter in an uncharacteristic sample or set of samples using k-near- est neighbors or Support Vector Machines techniques
Find Potential Regulatory Sequences	Finds common regulatory sequences in a gene list, or searches for a known regulatory sequence
Principal Component Analysis	Reduces a multi-dimensional data set by transforming a number of correlated variables into a smaller number of uncorrelated variables (principal components)
Find Similar Samples	Finds similar samples based on a gene list, target sample, or sample pool
Find Significant Parameters	Finds parameters and attributes that are correlated to an expression
Find Similar Genes	Finds similar genes based on a gene list and a similarity measure, such as a Pearson or Spearman correlation
Draw Expression Profile	Draws a pseudo-gene to represent a hypothetical expres- sion pattern
Check Remote Execution Queue	Shows the status of all jobs: running, paused, pending, and completed, on Workgroup Server
Script Editor	Creates re-usable scripts that can be applied to any data set

## **Annotations Menu Commands**

These commands are available from the **Annotations** menu in the GeneSpring GX main window.

Annotations Menu Command	Description
Update Annotations from Agilent	Updates annotations from the Agilent database
Build Homology Tables	Builds homologies between any pair of organisms that are included in both HomoloGene and UniGene, or within-genome homologies based on the UniGene Cluster ID or the LocusLink Locus ID
GO Ontology Browser	Opens the GO Ontology Browser window, allows you to explore the overlap between an ontology and a gene list
Make Gene Lists from Annotations	Creates custom gene lists based on annotations
Build Ontology	Hierarchically groups genes into meaningful biological cat- egories (gene lists) based on the Gene Ontology Consor- tium Classifications (GO)
Edit Genes and Annotations	Displays the Master Table of Genes Editor which lets you modify genes and annotations. Also provides access to the Genome Inspector
Genome Inspector	Opens the Genome Inspector window

## Window Menu Commands

These commands are available from the **Window** menu in the GeneSpring GX main window.

Window Menu Command	Description
List of open windows	Allows you to switch to the selected window
Tile Windows	Arranges all open GeneSpring GX windows in a tile pattern for easy access to all open windows

# **Help Menu Commands**

These commands are available from the **Help** menu in the GeneSpring GX main window.

Help Menu Command	Description
Tutorial	Opens your default browser and takes you to the Gene- Spring GX Tutorial in PDF format. You can save this file to your local machine and print it. The tutorial provides step-by-step procedures on how to use many basic feature and functions of GeneSpring GX.
User Manual	Opens the manual installed on your hard drive during installation or updating. The GeneSpring GX User Manual is a PDF document you can save or print.
Quick Start Guide	Opens this Quick Start Guide in PDF format for online viewing or printing.
Version Notes	Displays notes for your version of GeneSpring GX. The versionnotes.html are located in the\silicongenet- ics\genespring\docs directory.
Keyboard Shortcuts	Opens a dialog box that shows the keyboard shortcuts for menus, tools, Views, 2D graphs, 3D graphics, and trees.
Update GeneSpring GX	Downloads the latest version of GeneSpring GX
Support and Training Resources	Displays the Agilent training page on the web site, where you can view available training options
Email Technical Support	Opens an email to request Agilent technical support
Agilent Online	Opens the Agilent web site. This site contains a wealth of information including manuals and information on work- shops designed to help you use GeneSpring GX more effectively.
Workgroup Server Database	Opens the Workgroup Server web browser. From here you can download a demo version of Workgroup Server and upload or download additional information. See the Work-group Server User Manual for more information.
System Monitor	Displays the Java system monitor to track free memory and view the processes running on your computer
Test Database Connectivity	Verifies that GeneSpring GX and the database are com- municating with each other

Help Menu Command	Description
Show Helpful Hints	Displays a new helpful hint each time you start Gene- Spring GX
About	Displays information about GeneSpring GX, such as the version number and demo expiration date. You will need this information if you contact Technical Support.

## **Navigator Pane**

The Navigator pane provides access to all data elements relating to genomes. A Navigator pane appears on the left side of the GeneSpring GX main window and in many other GeneSpring GX windows as well. In some windows, only selected folders are displayed in the Navigator, but the basic operation is the same.

Information is organized into files and folders, which store information about different data objects in GeneSpring GX. Each folder contains a specific type of information. By default, folders in the Navigator are closed, although an "all genes" or "all genomic elements" gene list is displayed on start-up. The folders and objects are color-coded as follows:

- A yellow folder icon indicates a folder that can contain data objects.
- A bright-colored icon indicates an object that is loaded into memory.
- A pale blue icon indicates an object that is *not* loaded into memory.

Clicking on an object in the Navigator displays that data object in the Genome Browser pane. Double-clicking on an object opens an Inspector window, which displays more information for that object.

The Navigator pane in the GeneSpring GX main window contains the following folders.

### **Gene Lists**

This folder contains collections of genes called as gene lists, which are created and used during analysis in GeneSpring GX. By default, GeneSpring GX makes and displays an "all genes" list containing all genes in the genome.

### **Experiments**

This folder contains experiment information. Experiments are divided into interpretations. Experiment Interpretations determine how to experiment variables, called experiment parameters, are treated and displayed. Experiment interpretations have conditions. Conditions are groupings of one or more samples. Experiment folders also contain lists of samples. The following is an example of an Experiments folder and the various types of objects contained in it:



### **Gene Trees**

This folder contains gene trees created in GeneSpring GX. Gene trees are dendrograms that show relationships among the expression levels of genes over a series of conditions.

### **Condition Trees**

This folder contains condition trees created in GeneSpring GX. Condition trees are dendrograms that show relationships among the expression levels of samples.

#### Classifications

This folder contains genes that have been grouped or classified into groups as defined by k-means or SOM clustering.

### **Pathways**

This folder contains pathway objects, which are images of regulatory or metabolic pathways imported into GeneSpring GX. Genes are overlaid on these images, which allows you to observe their changing expression levels across experimental conditions. The item **Find Genes Which Could Fit Here** on the menu is a tool to predict new pathway elements.

#### **Array Layouts**

The Array Layouts folder contains information about the arrangement of the spots on your array. These can be used to recreate an image of your arrays to check for regional abnormalities.

### **Expression Profiles**

Expression profiles are lines representing gene profiles that you draw in the Genome Browser. You can then search for genes matching that profile. Any expression profiles you create are stored in the Expression Profiles folder.

### **External Programs**

External programs are analysis programs outside GeneSpring GX that can be launched from within GeneSpring GX. Data from GeneSpring GX is sent to the program and output from the program is recognized by GeneSpring GX. These programs are kept in the External Programs folder.

#### **Bookmarks**

Bookmarks are saved display settings such as an experiment, gene list, color scheme, or selected genes. You can always save your current display and return to it later by opening the Bookmarks folder and selecting a particular bookmark.

#### **Scripts**

Scripts are tools that save time by allowing a long series of data analysis steps to be performed at once. Scripts are re-usable and can be applied to any data set. You can create your own scripts using the Script Editor window. All scripts, including complimentary scripts shipped with GeneSpring GX, are stored in the Scripts folder.

### **Show options**

The Show box in the bottom of the Navigator pane allows you to access the following commands for filtering data in the Genome Browser:

All Data - Shows all data in the Genomic browser

Not Assigned - Shows data that have not been assigned to a project

**Filter** - Opens the Filter Navigator window, which allows you to search for data by keyword or other search criteria

#### **Navigator Shortcut Menu Commands**

The following menu commands appear when you right-click on a folder or object in the Navigator pane. The commands can differ depending on which data object you select. Some items apply to folders only, as indicated below.

Command	Description
Display	Displays the contents of an object in the Genome Browser
Attachments	Adds an attachment to the selected object
Inspect	Opens an Inspector to display the selected object
Upload to Workgroup Server	Uploads data from GeneSpring GX to Work- group Server
Copy from Workgroup Server	Copies data from Workgroup Server to Gene- Spring GX
Assign Project	Assigns a data object to a project
Change Owner	Changes the owner of a project
Rename	Renames a folder or object
Delete	Deletes the selected folder or object
Export as Zip	Exports the selected data as a zip file
Clear (folders only)	Clears the image in the Genome Browser
Import Zip (folders only)	Imports the selected zip file
Add New Folder (folders only)	Creates a new folder

## **Other Navigator Shortcut Menu Commands**

The following menu commands appear when you right-click on certain objects or folders in the Navigator pane.

Command	Object/Folder	Description
Set as Default	Experiment	Automatically loads this experiment when GeneSpring GX starts
Use as Coloring	Classification/ Gene List folder	Colors genes based on some previ- ously-defined knowledge about them
Split Window	Classification/ Gene List folder	Splits the Genome Browser into several small windows
Display as Second Gene List	Gene List	Displays a second gene list in the Genome Browser for comparison
Set Secondary Exper- iment	Experiment	Shows an additional experiment in the Naviga- tor (only enabled for certain views)
Translate	Gene List/ Experiment	Creates a homology gene list after logging in to Workgroup Server
Re-Annotate	Gene Tree	Annotates genes using only standard lists
Venn Diagram	Gene Lists	Creates a Venn diagram and colors a specific gene according to your specification
Make Gene Lists	Classification/ Pathways	Creates a list of genes
Run	Script/External Pro- grams	Runs a script or an external program
Use as Coloring	Classification	Colors chromosome numbers by classification
Use to Color Branches	Classification	Colors tree branches by classification
Edit	Script	Allows you to edit a script
Help	Script	Displays the help topic for the selected script
Duplicate	Experiment	Copies an experiment and saves it with a new name
Save DB Data Locally	Experiment	Saves experiment information from a database to your local computer

Command	Object/Folder	Description
Export as MAGE-ML	Experiment	Exports data in MAGE-ML format
Draw Gene Profile	Expression Profile	Draws lines representing gene profiles in the Genome Browser

# **Genome Browser**

The Genome Browser is the main area of the GeneSpring GX window that displays a visual representation of genes. You can zoom in on data using the mouse, pan around the data using arrow keys, or display the data in different views as described below. See "Viewing and Selecting Data" on page 40 for more information about using the Browser.

### **Genome Browser Mouse and Key Strokes**

Mouse Action or Key Stroke	Result	Menu Command
<b>Drag</b> a rectangle across a region that you want to expand. Repeat the process until you reach the desired magnification level.	Zooms in (expands selected region)	View > Zoom In
Click the <b>Zoom Out</b> button at the bottom of the <b>Genome Browser or right-click</b> and select <b>Zoom Out</b> .	Zooms out one level	View > Zoom Out
Click the <b>Zoom Fully Out</b> button at the bottom of the <b>Genome Browser or right-click</b> and select <b>Zoom Fully Out</b> .	Zooms out com- pletely	View > Zoom Fully Out
Left arrow, Right arrow	Pans horizontally	None
Up arrow, Down arrow	Pans vertically	None
PageUp, PageDown	Pans vertically by a whole screen	None

### **Genome Browser Views**

The following items on the **View** menu in the GeneSpring GX window allow you to change the display of data in the Genome Browser. Note that the Browser view changes automatically when you select certain data objects in the Navigator. For example, when you select a pathway object, the Browser view automatically changes to Pathway view.

• Line Graph	Array Layout
• Bar Graph	• Pathway
• Graph by Genes	Ordered List
• Box Plot	• Scatter Plot
Physical Position	• 3D Scatter Plot
• Blocks	Compare Genes to Genes
<b>M</b>	

• Tree

## **Browser Shortcut Menu Commands**

The following menu commands appear when you right-click in the Browser pane. Some commands appear only when certain data objects are selected.

Menu Command	Action
Zoom In	Increases the detail shown in the view
Zoom Out	Decreases the detail shown in the view
Zoom Fully Out	Returns the view to the original data display
Unsplit Windows	Returns to an unsplit Genome Browser window
Remove Secondary Gene List	Removes the assignment of a secondary gene list
Make List from Selected Genes	Makes a list of all the genes you have selected in the Genome Browser
Save Expression Profile	Saves a new expression profile in the Expression Profiles folder in the Navigator
Save Image	Allows you to save the image of the Genome Browser in either a PNG or PICT file format to export to other programs
Save Bookmark	Opens the Save New Bookmark Window. Bookmarks allow you save your current display settings, including experiments, gene list, coloration, and selected genes. The saved bookmark appears in the Bookmarks folder in the Navigator pane, so that you can return to it later.
Load Sequence	
Delete Sequence Region Element	
Display Options	Opens the Display Options window, which allows you to modify the appearance of the current view.

## **Other Browser Shortcut Menu Commands**

The following menu commands appear when you right-click in certain views in the Browser pane.

Command	View	Description	
Load Sequence	Physical Position	Loads nucleic acid sequences with genome data	
Make Subtree	Tree	Creates a new tree from a node of another tree	
Make List from Sub- tree	Tree	Makes a gene list from a subtree	
Display Subtree	Tree	Displays a subtree by clicking on a node	
Display Parent of Sub- tree	Tree	Displays the tree directly above the one selected	
Display Entire Tree	Tree	Displays the entire tree by clicking anywhere on the tree	
Find Genes Which Could Fit Here	Pathway	Predicts new pathway elements	
Delete Entire Tree	Tree	Deletes the selected tree	
Reset Orientation	3D Scatter Plot/Condition Scatter Plot	Resets the scatter plot to the default orienta- tion	
Select Genes Within Lines	2D Scatter Plot	Selects and highlights genes within the fold lines	
Show /Hide Associ- ated Values	Compare Genes to Genes	Shows or hides the values from the selected gene list	
Unsplit Window	Any view split by classification	Rejoins windows that have been split into several smaller windows	
Remove Secondary Gene List	When a secondary gene list is selected	Removes a secondary gene list from the Navi- gator	
Make list of genes in this list only	Physical Position and Color by Venn Diagram	Makes a Venn diagram based on a gene list	

Command	View	Description
Clear this selection	Physical Position and Color by Venn Diagram	Clears the gene list in the Venn diagram
Set Coloring	Physical Position and Color by Venn Diagram	Opens the Display Options window with the Coloring tab selected, which allows you to set the colorbar range and change colors

# **Basic Operations**

The following basic operations are described in the following sections:

- "Setting Up the Genome" on page 30, which includes the following substeps:
  - "To import genomes from Agilent" on page 30
  - "To update annotations with GeneSpider" on page 31
  - "To inspect the genome" on page 32
- "Importing Data and Creating Experiments" on page 33, which includes the following substeps:
  - "To import data from text files in recognized formats" on page 33
  - "To set sample attributes" on page 34
  - "To assign normalizations" on page 35
  - "To assign parameter values" on page 37
  - "To assign interpretations" on page 38
  - "To filter samples in an experiment" on page 39
- "Viewing and Selecting Data" on page 40, which includes the following substeps:
  - "To select genes" on page 40
  - "To view data objects" on page 40
  - "To find genes" on page 42
  - "To perform an advanced gene search" on page 42
  - "To make gene lists from selected genes" on page 43
- "Filtering Data" on page 43
  - "To filter on expression level" on page 44
- "Performing Analyses" on page 46

- "Managing Your Genome" on page 47, which includes the following substeps:
  - "To build homology tables" on page 47
  - "To build ontologies" on page 48
  - "To create gene lists from annotations" on page 48
  - "To import KEGG pathways" on page 48
  - "To compare a gene list with the GO ontologies" on page 49

# **Setting Up the Genome**

You can import genomes into GeneSpring GX in any of the following ways:

- Import genomes from Agilent. See "To import genomes from Agilent" on page 30
- Import genomes from GenBank/EMBL files. See "To import genomes from GenBank/EMBL files" in the *GeneSpring GX User's Guide/Online help*.
- Import tab-delimited genome files. See "To import tab-delimited genome files" in the *GeneSpring GX User's Guide/Online help*.

Annotate the genome as described in "To update annotations with GeneSpider" on page 31.

Inspect the genome as described in "To inspect the genome" on page 32.

#### To import genomes from Agilent

Use the following procedure to download genomes from the Agilent web site for the Affymetrix, Agilent, and CodeLink platforms.

- 1 Select **Import Genome** from the **File** menu to open the Import Genome window.
- 2 Select the **Download a Standard Genome** option.

The list of pre-made genomes is organized into folders and files. The folders list the name of the manufacturer and the folders contain the genomes from that manufacturer. Each genome is named with the exact name provided by the manufacturer. Different versions of a chip are treated as different genomes.

- **3** Click the **Genomes and Arrays** folder and find the genome you want to import.
- 4 Click on the genome of interest to select it.

- **5** Click the **Next** button. A progress bar opens and the genome begins to download. When the downloading has finished, the genome appears in a new Genome Browser.
- **6** Name and save the genome.
- 7 The New Genome Checklist window opens, which allows you to perform the following activities. Most of these are optional and can be performed at a later time by clicking the appropriate command on the Annotations menu.
  - Click the **GeneSpider** button to update annotations from public databases. See "To update annotations with GeneSpider" on page 31.
  - Click the Make Gene Lists button to make gene lists from annotations.
  - Click the **Build Ontology** button to build the GO SLIMS gene ontology.
  - Click the **Pathways** button to import a KEGG pathway folder.
  - Click the Homology Tables button to build homology tables from multiple genomes.
  - Click the **Resources** button to download supplemental information from the Agilent web site, such as new pathway diagrams, chromosome maps, or sequence information that can be used in the Find Regulatory Sequence search function.
  - Click the Import Data button to import a data file. This is the same as selecting Import Data from the File menu. See "Importing Data and Creating Experiments" on page 33.
- 8 Click the **Close** button to exit the New Genome Checklist window. You can edit the Genome later by selecting the appropriate menu item from the Annotations menu.

#### To update annotations with GeneSpider

Use the procedure below to retrieve annotations from the NCBI web site or the Agilent mirror site. The Agilent mirror server returns the same annotations as the GeneSpiders that access GenBank, LocusLink, and UniGene, and also fills in GO biological process, GO molecular function, GO cellular component, and RefSeq identifier fields from the LocusLink database and the UniGene cluster ID from the UniGene database.

**Note:** Depending on the number of genes in the genome, this process can take a number of hours to complete. You may want to let the process run over night.

- **1** Open the Update Genome window in either of the following ways:
  - Click the GeneSpider button in the New Genome Checklist window.
  - Select Update Annotations from Agilent from the Annotations menu.

- **2** Select the annotation source and source priority options you want:
  - **a** Select the boxes next to the annotation sources from which to retrieve data.
  - **b** Select **Concatenate annotations from different sources** to retrieve annotations from all sources as a semicolon-delimited list. Exact duplicates are not retrieved. The order is fixed: GenBank, then LocusLink, and finally, UniGene.
  - **c** Select **Keep the highest priority annotation** to retrieve only the annotation from the highest priority source available for each gene. Use the **Move Up** and **Move Down** buttons to reorder the priority.
  - **d** Select the **Overwrite Existing Annotations** check box to update information in places where data already exists. Otherwise new information is added only to blank fields.
  - e Select the **Retrieve Sequence Data** option to update sequence data from Agilent or GenBank.
- **3** Select the column in your Edit Genes and Annotations table that contains GenBank accession numbers.
- **4** Click the **Start** button. Status information is displayed while the GeneSpider runs.
- **5** When GeneSpider is finished, click the **Save and Close** button to update the Edit Genes and Annotations table.

## To inspect the genome

- **1** Open the Genome Inspector window in one of the following ways:
  - Select Genome Inspector from the Annotations menu.
  - Select **Genome Manager** from the **File** menu, select the genome you want to inspect, and then click the **Inspect** button.
  - Select **Edit Genes and Annotations** from the **Annotations** menu, select a genome you want to inspect, and then click the **Inspect Genome** button.
- **2** Review the information shown on the following tabs:
  - **Genome Properties** tab displays a summary of the contents in the Edit Genes and Annotations table
  - History tab displays a history of changes made to the genome
  - **GenBank/EMBL Files** tab displays the GenBank/EMBL files associated with the genome

- **Sequence** tab allows you to import the full nucleotide sequence (if available)
- **Homology Tables** tab allows you to view, add, edit, and delete homology tables associated with the genome
- **Web Links** tab allows you to view, create, edit, or delete gene and experiment web links that are associated with the genome

#### See Also "Managing Your Genome" on page 47

The "Genome Inspector Window" topic in the "Genome Windows" chapter of the *GeneSpring GX User's Guide/Online help* 

## Importing Data and Creating Experiments

The process of importing data to create experiments is made up of the following activities. A series of windows are displayed to guide you through the process.

- Importing data in one of the following ways:
  - from tab-delimited text files in either a recognized or custom format
  - from Affy CEL or CHP files. See the *GeneSpring GX User's Guide/Online* help.
  - by setting up a database and importing data from the database. See *GeneSpring GX User's Guide/Online help*.
- Setting sample attributes. See "To set sample attributes" on page 34.
- Creating an experiment, including the following activities:
  - Setting normalizations. See "To assign normalizations" on page 35.
  - Setting experiment parameters. See "To assign parameter values" on page 37.
  - Defining experiment interpretations. See "To assign interpretations" on page 38.

#### To import data from text files in recognized formats

- **1** Select **Import Data** from the **File** menu.
- **2** Select the file that contains the data of interest.
- **3** Click the **Open** button.
- **4** When the Import Data: Define File Format and Genome window opens, select the format of your file from the list of file formats. See "To load expression data from files with custom format" in the *GeneSpring GX*

*User's Guide/Online help* if the format of your file does not appear in the list.

- **5** Specify the genome for the data in one of the following ways:
  - Select the **Select Genome** option, and then select the genome from the folder or file you want if the genome is already loaded.
  - Otherwise, select the **Create a New Genome** option to create a genome based on the data in the file you are loading.
- 6 Click the Next button. The Import Data: Selected Files window opens.
- 7 Select the files to import from the Files list.
  - Click on a column heading if you want to sort the table by the values in that column.
  - Click on a file to select it.
  - <Shift>+click or <Ctrl>+click to select multiple files.
  - <Ctrl>+click on a selected file to deselect it.
  - Click the **Add**>> button to add the selected files, or click the **Add All**>> button to add all files. The selected files appear in the Selected Files column.
- 8 Click the Next button. The Import Data: Sample Attributes window opens. Continue as described in "To set sample attributes" on page 34.

#### To set sample attributes

The Sample Attributes window opens automatically when you click the **Next** button on the Import Data: Selected Files window, if no additional clarification of the import file is needed. See "To import data from text files in recognized formats" on page 33.

- 1 Click the **New Attribute** button on the Sample Attributes window. The New Attribute window opens. You must fill in the required attributes, which are colored yellow.
- **2** Do *one* of the following:
  - To add a standard attribute, select it from the list and click the **OK** button.
  - To add a custom attribute, select the **Custom** option, create the new attribute, then click **OK**.

A new column appears in the Sample Attributes table.

**3** Click the **Edit Attribute Value** button to enter the value of an attribute on the Edit Attribute Value window.

- **4** Do *one* of the following:
  - To add a standard attribute value, select it from the list then click OK.
  - To add a custom attribute value, select the **Custom** option, enter the value, and then click **OK**.

The updated value appears in the Sample Attributes table.

- **5** When have finished adding attributes, click the **Next** button. After the samples have been imported, the Import Data: Create Experiment window opens.
- **6** Do *one* of the following:
  - Click the **Yes** button to create a new experiment from the samples. Continue with Step 7.
  - Click the **No** button on the Create Experiment window. In this case, the imported data are saved as samples, but a new experiment file is not created. You can view the samples in the Sample Inspector window and can create experiments from them later by selecting **Create New Experiment** from the **Experiments** menu.
- 7 If the Normalization Errors window opens, review the information in the window and click the **OK** button to continue.You can renormalize your data at another time.
- 8 Name and save the experiment.
- **9** The New Experiment Checklist window opens. Follow all steps in the checklist in order to complete the creation of the experiment.
  - Click the Normalizations button to define or edit normalizations. See "To assign normalizations" on page 35. This is equivalent to selecting Experiment Normalizations from the Experiments menu.
  - Click the **Parameters** button to define or edit experiment parameters. See "To assign parameter values" on page 37. This is equivalent to selecting **Experiment Parameters** from the Experiments menu.
  - Click the **Experiment Interpretation** button to define or edit default interpretations. "To assign interpretations" on page 38. This is equivalent to selecting **Experiment Interpretation** from the Experiments menu.

10 Click the Close button to exit the New Experiment Checklist window.

#### To assign normalizations

You can use the default normalization settings, or you can use this procedure to assign different normalization steps to your experiment.

- **1** Open the Experiment Normalizations window in *one* of the following ways:
  - Click the **Next** button on the Create New Experiment: Experiment Parameters window
  - Click the **Normalizations** button on the New Experiment Checklist window
  - Select **Experiment Normalizations** from the **Experiments** menu
- **2** Add a normalization step in one of the following ways:
  - Click to select the normalization step from the list on the left side of the window, then click the **Normalization Step** button.
  - Double-click on the normalization step in the list on the left side of the window.
- **3** Enter the information on the window displayed for the selected normalization step.
- **4** (optional) To select the samples in the current experiment that you want the normalization step applied to:
  - **a** Click the **Apply Only to Specific Samples** option. A list of samples in the current experiment appears.
  - **b** Select the box of any samples to which you want to apply this step.

Otherwise the normalization step is applied to *all* samples in the experiment.

- **5** Click **OK**. The step is added to the list of normalization steps to perform on the right side of the Experiment Normalizations window.
- 6 Repeat Steps 2 5 to add additional normalization steps if desired.
- **7** (optional) To view or modify the normalization settings for a step in the list of normalizations to perform:
  - **a** Click to select a step from the list.
  - **b** Click the **Inspect** button. A configuration window appears. View or modify the data, then click **OK**.
- **8** (optional) Click the **Get Text Description** button to get a text description of the normalization steps to perform.
- **9** Order the normalization steps using *one* of the following:
  - Click the **Move Up** or **Move Down** buttons to position a selected step in the list.
  - Select the **Use Recommended Order** checkbox to order the steps in the order they appear in the list on the left side of the window, which is the generally the best order to perform the normalization steps.

**10** (optional) Click the **Save As Scenario** button to save this normalization scenario for future use.

#### To assign parameter values

Use the procedure below to assign parameter values for each of the samples in your experiment. Examples of experiment parameters and values could be:

Parameter: Drug Concentration, Value: Drug Concentration in ppm

Parameter: Yeast,	Value: Strain of Yeast
Parameter: Cancer Outcome,	Value: Malignant, Benign, or Normal
Parameter: Replicate,	Value: Replicate Number

Experiment parameters are used to group samples with the same parameter values into groups called "Conditions", which can be used to display data in graphs and to analyze the variability of genes.

- **1** Open the Experiment Parameters window in any of the following ways:
  - Select Experiment Parameters from the Experiments menu
  - Click the **Next** button on the **Create New Experiment: Select Samples** window
  - Click the Parameters button on the New Experiment Checklist window
  - Click the **Experiment Parameters** button on the Change Interpretation window
- **2** For each parameter value you want to edit, click the appropriate sample cell in the table and enter the value of the parameter for that sample.
- **3** To copy a value from one cell to other cells below it in the same column:
  - **a** Click on the cell that has the value you want to copy.
  - **b** <Shift>+click to define the range of cells to replace.
  - **c** Click the **Fill Down** button. The text from the top cell of the range is copied to all the other cells in the range.

- **4** To fill down alphabetic or numeric sequences:
  - **a** <Shift>-click on the cells underneath those values you would like replaced with the original sequence.
  - **b** To fill down a sequence of numbers, select three or more cells, where the first two cells contain the sequence you want to expand.
  - **c** Click the **Fill Sequence Down** button. If the spacing between the numbers can be determined from the first two sequences, it will be continued to the selected cells.
- **5** To replace text:
  - **a** Select the entries you want to change.
  - **b** Click the **Replace Text** button to open the Replace window.
  - c Enter the replacement text.
  - d Click OK to close the Replace window.
- **6** Save the changes to the parameters in one of the following ways:
  - Click the **Save** button to change the parameters in the current experiment.
  - Click the **Save As** button and enter a new name to save these parameters in a new experiment.

#### To assign interpretations

Use the procedure below to set up or view interpretations for your experiments. Experiment interpretations determine how your experiment parameters are displayed and how normalized values are handled.

- **1** Open the Experiment Interpretation window in one of the following ways:
  - Select **Experiment Interpretation** from the **Experiments** menu.
  - Right-click on an interpretation in the Navigator and select **Inspect** from the shortcut menu.
  - Double-click on an interpretation in the Navigator. See "Experiments" on page 21 for a diagram that shows the icons used for experiments, conditions, interpretations, and samples in the Navigator.
- **2** Select the following:
  - An analysis mode for the vertical axis.
  - Lower and upper bounds to limit the boundaries of the vertical axis in use in the Graph views.

- (optional) A flag if you want to limit the data that is used for the interpretation.
- (optional) Whether or not to use the Cross-Gene Error Model in this interpretation.
- A display mode for each parameter.
- **3** (optional) Click the **Error Model Structure** button to edit the cross-gene error model at this time.
- **4** Name and save the interpretation in one of the following ways:
  - Click the **Save** button to update the current interpretation.
  - Enter a new name and click the **Save As** button to create a new interpretation.

#### To filter samples in an experiment

Use this procedure to view only selected samples in your experiments.

- 1 Select **Sample Manager** from the **Experiments** menu.
- **2** To filter on experiments:
  - Click the **Filter on Experiment** tab.
  - In the **Use samples stored** list, select the location of the samples you want to search.
  - Select the experiment you want in Navigator. All samples associated with that experiment appear in the Filter Results list.
- **3** Click the appropriate tab to filter on the following:
  - **Filter on Parameter** tab filters samples based on a parameter and value range.
  - **Filter on Attributes** tab filters on sample attributes such as patient biography, lab technician, date, and ambient temperature.
  - **Filter on Keywords** tab filters samples based on whether a particular keyword appears in any of the specified fields. This method is useful in cases where a string such as "cancer" is sometimes the parameter, sometimes the parameter value, and sometimes is part of the experiment name. Select the conditions, then click the **Search** button.
  - Filter on Projects tab filters samples based on project assignments.

Samples that match your search criteria appear in the Filter Results list.

# **Viewing and Selecting Data**

Some common procedure for viewing and selecting data are described below. For more information, see the "Finding, Inspecting, and Viewing Data" chapter of the *GeneSpring GX User's Guide/Online help*.

#### To select genes

- Single gene<br/>selection• Click on any line or rectangle representing a gene in the Genome Browser.• Click on any line or rectangle representing a gene in the Genome Browser.
- Multiple 1 Click on any line or rectangle representing a gene in the Genome Browser.
- genes selection 2 <Shift>-click to select additional genes.
- Alternate method <Shift>-click-and-drag across the genes you want to select. A box appears as you drag. When you release the mouse, the selected genes are highlighted (selected).
  - To deselect a gene, <Shift>-click on it.

#### To view data objects

In the Browser area of the Physical Position view, the horizontal lines are chromosomes and the vertical lines are genes.

**View genes** In the Navigator, open the **Gene Lists** folder and click on the **All Genes** list. In the Physical Position view, genes appear as many gray or colored boxes on the chromosomes as shown in the example below.



To zoom in on a gene, drag a rectangle to select the area to expand. If the genome is sequenced, and you continue to expand the region, the sequence within the gene becomes visible as shown in the example below:



**View sequence** Drag a rectangle to select the area to expand as shown below:



The expanded area is displayed when you release the mouse button, as shown below. Each block represents a gene. Which side of the chromosome the block is drawn on represents the coding strand for the gene.



Continue to zoom in on a chromosome until the sequence is visible as shown in the example below. If sequence data is not already loaded for your genome, right-click and select **Load Sequence** to load it.



## To find genes

Use this procedure to find one or more genes in the current Browser view.

- 1 Select **Find Gene** from the **Edit** menu. The Find Gene in View window opens.
- **2** To find a gene in the current view, do the following:
  - a In the Search For field, enter the name of the gene you want to find.
  - **b** Click the **Find** button. Results are highlighted in the table. In some views, the Genome Browser zooms in on the gene and automatically selects it.
- 3 Click the Find Next button to view additional matches.
- **4** When finished, click the **Close** button.

## To perform an advanced gene search

Use this procedure to find one or more genes in the current Browser view, using more advanced criteria to expand or narrow your search.

- **1** Do one of the following to open the Advanced Find Genes window.
  - Click the Advanced button in the Find Gene in View window.
  - Select Advanced Find Gene from the Edit menu.
- 2 Enter the search terms. Separate multiple terms with the keywords "and", "or", or "not". If the Use \* as Wildcard option is selected in Step 5, you can enter an \* in this field to match any characters. Spaces between search terms is treated as an "and" keyword in the search.
- **3** Select the annotations to search.
- **4** (optional) Use the **Map Location** fields to restrict your search to specific regions within the genome. You can specify the chromosome number, cytogenic band, or sequence. Sequences can include the IUPAC-IUB ambiguity codes, A, K, Y, W etc. The symbol, X, is not allowed; use N instead to specify a single wildcard-base.
- 5 (optional) Specify additional search options: Case Sensitive, Whole Words Only, Use \* as a Wildcard, or Search Only Current Gene List.
- **6** Click the **Find** button. A list of genes that match the search criteria is displayed in a table at the bottom of the window, with the field that matched the search criteria colored red.
- 7 To learn more about one or more genes, do the following:
  - Click to select a gene of interest in the appropriate row or press <Ctrl>-click to select multiple rows (genes).
  - Click the **Select** button to highlight the genes in the Genome Browser.

- Click the **Select All** button to select all of the genes identified by the search.
- Click the **Zoom & Select** button to select and zoom-in on a gene in the Genome Browser.
- Click the **Inspect** button to display the Gene Inspector window for the selected gene.
- Click the **Make Gene List** button to make a gene list from all of the genes identified by the search.
- 8 When you are done, click the **Close** button.

#### To make gene lists from selected genes

**1** Select the genes in the Genome Browser in one of the following ways:

**Single gene** Click on any line or rectangle representing a gene in the Genome Browser. The name of the selected gene appears in the legend at the bottom of the Genome Browser.

**Multiple genes** Click on any line or rectangle representing a gene in the Genome Browser, then <Shift>-click to select additional genes.

**Alternate method** <Shift>-click-and-drag across the genes you want to select. A box appears as you drag. When you release the mouse, the selected genes are highlighted (selected).

- **2** Right-click and select **Make gene list from selected genes** from the shortcut menu.
- 3 (optional) Select Save Image > Browser from the File menu. This opens the Export Options window, which allows you to save the image as a PICT or PNG file for use with other applications.

## **Filtering Data**

You can filter your data based on the following criteria by selecting the appropriate command from the **Filtering** menu:

- Filter on Expression Level
- Filter on Fold Change
- Filter on Error
- Filter on Confidence
- Filter on Parameters

- Filter on Volcano Plot
- Filter on Flags
- Filter on Data Files
- Filter on Arbitrary Files
- Filter on Gene List Numbers

The procedure to filter on expression level is described below. For the other filtering methods, see "Filtering Data" chapter of the *GeneSpring GX User's Guide/Online help*.

#### To filter on expression level

Use the procedure below to find genes with certain values present in some of the conditions or samples in an experiment or interpretation. You can set what proportion of conditions must meet a certain threshold.

- **1** Select **Filter on Expression Level** from the **Filtering** menu to open the Filter on Expression Level window.
- **2** To select a gene list other than the current selection, do *one* of the following. Only genes that are members of the selected gene list are subjected to the filter.
  - Click on the gene list to use in Navigator, then click the **Choose Gene** List button.
  - Double-click on the gene list to use in Navigator.
- **3** To select an experiment other than the current selection, do one of the following:
  - Click on the data (experiment, condition, or subset of conditions) to use in Navigator, then click the **Choose Experiment** button.
  - Double-click on the data (experiment, condition, or subset of conditions) to use in Navigator.
- **4** (optional) Click the **Exclude Conditions** button to exclude certain samples or conditions from a filter. Otherwise, all conditions and samples are used in the filter.
- 5 Select the type of data to use: Normalized, Raw, or Control Signal.
- **6** Specify the following values for the filter: **Minimum**, **Maximum**, and **Values must appear in at least [ ] out of [ ] conditions**. The latter field allows genes that pass the filtering settings in some but not all conditions to be included in the filter results.
- 7 (optional) Clear the **Interactive Update** check box if you don't want the graphical display to be dynamically updated with genes that pass the

current filter settings. In this case, you won't be given any indication of how many genes pass the filter until you click the **Save** button. Clearing this option may help if the system becomes sluggish when working with large experiments.

- 8 (optional) By default, the best view for the current filter is used to display data in the preview pane. You can select the **Main Window** view option if you want lock the preview pane view to that of the main window.
- **9** Click the **Save** button to save the filter results as a gene list, for use in other filter steps or analyses.

# **Performing Analyses**

You can perform the following types of analyses on your data with GeneSpring GX

Analysis	Access via	Description
Statistical Analysis ANOVA	Tools Menu	Compares mean expression levels between two or more groups of samples
Clustering > K-Means	Tools menu	Produces groups of genes with a high degree of similarity within each group, and a low degree of similarity between groups
Clustering > Gene Tree	Tools menu	Produces dendrograms that show the relationships among the differ- ent expression levels of genes over a series of conditions
Clustering > Condition Tree	Tools menu	Divides genes or conditions into groups that have expression patterns similar to nearby genes in a tree
Clustering > Self-Organizing Map	Tools menu	Divides genes into groups based on expression patterns and illus- trates the relationship between groups by arranging them in a two-dimensional map
Clustering > QT Clustering	Tools menu	Identifies clusters of genes, such that each gene in the cluster is within a specified distance (based on a user-defined distance metric) of every other gene in the cluster
Class Prediction	Tools menu	Predicts the value (or "class") of an individual parameter in an unchar- acteristic sample or set of samples using k-nearest neighbors or Sup- port Vector Machines techniques
Find Potential Regulatory- Sequences	Tools menu	Finds common regulatory sequences in a gene list, or searches for a known regulatory sequence
Principal Component Analy- sis	Tools menu	Reduces a multi-dimensional data set by transforming a number of correlated variables into a smaller number of uncorrelated variables (principal components)
Filter on Fold Change	Filtering menu	Filters genes based on a comparison of two samples or conditions
Filter on Parameter	Filtering menu	Filters genes based on similar parameters
Filter on Volcano Plot	Filtering menu	Filters genes on volcano plots, which are constructed using fold change values and p-values. Volcano plots allow you to visualize the relationship between fold-change (magnitude of change) and statisti- cal significance (which takes both magnitude of change and variability into consideration). Volcano plots are used to visually represent differ- ential expression between two different conditions.

These analyses are described in more detail in the GeneSpring GX User's Guide.

# **Managing Your Genome**

Once you have imported and annotated your genome as described earlier in this book, you can perform the following optional activities to update and manage your genome.

- Build homology tables (necessary for multi-chip or multi-organism studies)
- Build ontologies
- Create gene lists from annotations
- Import pathways from KEGG and/or GenMapp. See *GeneSpring GX* User's Guide/Online help to learn how to import custom pathways.
- Compare a gene list with the GO ontologies.

#### To build homology tables

Use the procedure below to build a homology table. Homologies can be made between any pair of organisms that are included in both HomoloGene and UniGene. Within-genome homologies are based solely on UniGene Cluster ID or LocusLink Locus ID.

- **1** Open the Build Homology Tables window in one of the following ways:
  - Click the **Homology Tables** button in the **New Genome Checklist** window.
  - Select **Build Homology Tables** from the **Annotations** menu.
  - Click the **Add/Update** button on the **Homology Tables** tab in the **Genome Inspector** window.
- **2** In the upper table, select the appropriate column that contains the GenBank (or EMBL) accession number for the selected genome.
- **3** Select the next genome you want in the Navigator and click the **Add>>** button. The genome is added to the lower table on the right side of the window.
- **4** In the lower table, select the name of the column that contains the GenBank Accession Number for the selected genome.
- **5** Repeat Steps 2 through 4 for each genome you want to add.
- 6 Click the Start button. The process can take several hours to complete.

## To build ontologies

Use the procedure below to create a gene list for each of the ontological classifications. The Gene Ontology (GO SLIMS) classification system is used to build ontologies. All the genes with the same GO classification are grouped into the same gene list.

- 1 Click the **Build Ontology** button in the **New Genome Checklist** window or select **Build Homology Tables** from the **Annotations** menu.
- **2** Enter the name of the folder in which to store the GO gene lists.
- **3** Click **OK**. A message is displayed if the ontology is successfully built. The gene lists that were generated are stored in the Gene Lists folder specified in Step 2.
- **4** (optional) Use the **Similar Lists** function in the Gene List Inspector to find overlap between selected gene lists and the GO gene lists. This information can be used to determine whether a gene list contains an over-representation of genes from one or more GO classifications.

## To create gene lists from annotations

Use the procedure below to create a gene list for each of the values of a selected annotation. This allows you to group genes that have similar values for a particular annotation.

- 1 Click the **Make Gene Lists** button in the **New Genome Checklist** window or select **Make Gene Lists from Annotations** from the **Annotations** menu to open the Make Gene List from Annotations window.
- **2** Select the annotation column to use to generate the gene list.
- **3** Select the **Divide by Semicolons** check box.
- **4** Click **OK**. The gene lists that were generated are stored in the Gene Lists folder named for the annotation selected in Step 2.

## To import KEGG pathways

Use the procedure below to import Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway diagrams into GeneSpring GX. Pathway diagrams from KEGG depict many different types of biological interactions, including metabolic and regulatory pathways. Superimposing the expression values of pathway genes on the diagrams may provide you with insight about the behavior of the genes in the expression experiment and their role in the pathways.

- 1 Click the **Pathways** button in the New Genome Checklist window to open the Import KEGG Pathways window.
- **2** If you don't already have a pathway downloaded, download it from the KEGG FTP site as described in the *GeneSpring GX User's Guide/Online help*.

**Note**: Organism-specific folders are named by the first letter of the genus and first two letters of the species (for example, 'hsa' stands for Homo sapiens).

- **3** Import downloaded pathways into GeneSpring GX as follows:
  - **a** Select the pathway folder you downloaded from the KEGG site in Step 2.
- **4** Decide if you want to group the pathways into subfolders. This option can be useful if the folder contains many pathways, e.g. more than 100.
- **5** Decide whether you want to save pathways as gene lists as well as pathway diagrams. Each gene list will have the same name as the pathway it was created from, and will contain all the genes that are represented in that pathway.

**Note**: Use the "Lists and Pathways Containing Gene" feature in the Gene Inspector to determine if a gene is a member of a pathway or if a gene list you created contains an over-representation of genes that belong to specific pathways.

6 Click OK.

#### To compare a gene list with the GO ontologies

Use this procedure to examine the overlap between a GO ontology and a gene list.

- **1** Open a genome that has entries in the GO Ontology columns. This is usually the result of running the GeneSpider.
- **2** Select a gene list of interest.
- **3** Select **Ontology Browser** from the **Annotations** menu.
- **4** Review the information in the Navigator on the right side of the Ontology Browser window in Pie Chart view, including the ontology category and the number of genes from the selected gene list that are within that category. The default ontology is Biological Process. Hover the mouse over an area of the pie chart to get more information on that area.
- **5** To select a different gene list, double-click on the gene list in the Navigator *or* select a gene list in the Navigator and click the **Choose Gene List** button.

- **6** Double-click on a pie segment to drill down and see the categories that make up that segment. Click the **Move Up** or **Move to Top** buttons to reverse the direction of drilling.
- 7 Click to select a particular ontology or ontology category from the Navigator on the left side of the window.
- 8 Select a subset in the **Restrict to subset** list to limit the ontology defined to selected subset of GO.
- **9** To make a gene list of the *overlap* between ontology and selected list in Pie Chart view:
  - **a** Select the categories of interest by clicking on them. Use <Ctrl> + Click to select multiple categories.
  - **b** Right-click and select **Make List from Selected Categories** or click the **Make List** button.

**Note:** If multiple categories are selected, the gene list created is the result of the following Boolean expression: (gene list AND category 1) OR (gene list AND category 2) OR (gene list AND category 3)...

- **10** Right-click in the pie chart and select **Save Image** to save the image in a file for use with another application.
- **11** Click the **Table** tab to filter ontologies for the ones that have the highest overlap with the selected gene list:
  - **a** Move the slider on the left side of the table down to see only the gene lists with low p-values or change the **Minimum p-Value** setting.
  - **b** See the "Ontology Browser" topic in the *GeneSpring GX User's Guide/Online help* for descriptions of the types of information displayed in the table.
  - **c** Click a column heading to sort the table according to the entries in that column. The sorting of the Category column is based on the hierarchy of the ontologies, rather than alphabetical.
- **12** To make a gene list of the *overlap* between ontology and selected list in Table view:
  - **a** Select the categories of interest by clicking on them. Use <Ctrl> + Click to select multiple categories.
  - **b** Right-click and select **Make List from Selected Categories** or click the **Make List** button.
  - **c** The resulting gene list contains genes that are in the gene list *and* in the selected categories.

- **13** To restrict the entries in the table:
  - **a** Select the **Suppress empty categories** option to display only those lists that have any genes in the category at all. Some genomes won't have any genes that map to those defined as part of the GO ontology.
  - **b** Select the **Suppress non-overlapping categories** option to display only those lists that have genes in both the selected gene list and the listed category.
- **14** Right-click and select **Display Options** to change the way the data is displayed.

# Where to Find More Information

You can access more information about GeneSpring GX as follows.

## **Agilent Web Site**

To view support information for GeneSpring GX and other Agilent products, see: http://www.agilent.com/chem/genespring.

# **GeneSpring GX User's Guide**

Consult the *GeneSpring GX User's Guide* either as online help or in printable PDF format for in-depth information about using GeneSpring GX. To access online help, click the **Help** button in the GeneSpring GX windows or select **User Manual** from the **Help** menu in the GeneSpring GX main window.

# www.agilent.com

# In this Book

The *Quick Start Guide* presents steps to analyze data with GeneSpring GX.

© Agilent Technologies, Inc. 2005

Printed in USA First Edition, December 2005

