Agilent GeneSpring GX Software

Overview and Biological Significance Analysis Quick Start Guide

This quick start guide gives you an overview of the basic concepts of data analysis and terminology in the GeneSpring GX module of the GeneSpring suite of products. It also introduces you to the main elements of the GeneSpring GX user interface.

The second section of this quick start guide illustrates data analysis in GeneSpring with a sample exercise. This sample exercise takes you through the **Analysis: Biological Significance** guided workflow for microarray data in GeneSpring GX. The data used for this exercise is an Agilent Single Color Expression data set that is included in your software installation. However, the steps of this guided workflow are common to many of the other experiment and analysis types available in GeneSpring.

This quick start guide covers:

1 Overview of GeneSpring GX

2 The Analysis: Biological Significance guided workflow

3 Where to find additional information

Agilent Technologies

What is Agilent GeneSpring?

The Agilent GeneSpring suite of products provides powerful, accessible tools for statistical analysis and effective visualization of genomics, transcriptomics, proteomics, and metabolomics data. The GeneSpring platform is specifically designed for scientists, offering an interactive desktop computing environment to analyze and visualize data within a biological context. GeneSpring helps scientists identify biological entities such as genes, proteins, and cellular pathways that are significant to a biological hypothesis.

The following four modules are available for GeneSpring:

- **GeneSpring GX** for data from microarray technologies.
- **GeneSpring NGS** for data from next-generation sequencing platforms.
- **Mass Profiler Professional** for mass spectrometry data.
- **Pathway Architect** to perform pathway and network analysis across data types and organisms.

With GeneSpring you can:
- Normalize experimental data using various methods, like percentile shift normalization, quantile normalization, or Lowess normalization.
- Easily create experiments that link trends in data to various test parameters. Therefore, GeneSpring allows you to test complex hypotheses by running previously defined analyses against different combinations of samples.
- Assess data quality across samples using correlation analysis and Principal Components Analysis (PCA) and filter out bad quality data.
- Identify biologically significant entities using powerful statistical algorithms, such as t-test, ANOVA, multiple testing corrections, false discovery rate prediction, Tukey and Student-Newman-Keuls post hoc tests.

- Find similarities across entities and conditions with clustering algorithms and visualization techniques such as hierarchical clustering, K-means clustering, Self-Organizing Map (SOM), and PCA.

- Visualize and compare data and analysis results with a variety of visualization options, such as Genome Browser, Profile Plots, Scatter Plots, MvA plots, Box-and-Whisker plots, Venn Diagrams, and more.

- Interpret your results in a biological context using Gene Ontology (GO) Analysis, Gene Set Enrichment Analysis (GSEA), Gene Set Analysis (GSA), or Pathway Analysis.

- Discover new interactions for your entities of interest from the latest published literature using a powerful natural language processing (NLP) algorithm.

- Import pathway information from many sources and in different formats, including native import from the WikiPathways[1] online portal.

- Perform multi-omic pathway analysis to integrate your results from genomics, transcriptomics, proteomics, and metabolomics experiments for one or multiple organisms.

- Automate complex analysis tasks using scripts.

- Export data, images, and analysis reports in standard file formats.

Overview of GeneSpring GX

How do I get started with Agilent GeneSpring GX?

After installation of GeneSpring GX, you can get started immediately using the preloaded demonstration experiment. The demonstration experiment allows you to familiarize yourself with the software functionality and workflow. The project called "Demo Project" contains an experiment called "HeLa cells treated with compound X". This experiment consists of six samples of HeLa cells treated with an unknown compound in a cancer screening program. Three samples are treated biological replicates (control) and the other three are untreated biological replicates. You are encouraged to explore this demonstration project to familiarize yourself with GeneSpring.

Basic concepts of data analysis and terminology

This part of the quick start guide explains the basic concepts of data analysis and terminology in GeneSpring. It also describes the main elements of the GeneSpring user interface.

Basic concepts

In GeneSpring, data is organized in terms of projects, which can hold multiple experiments of different experiment and analysis types and differing organisms. In addition to naming the projects and experiments, you can also provide a short description of your project or experiment in the form of notes (see "Setting up a project" on page 15). This combination of a name and descriptive notes allows you to get a quick view into your project or experiment when you return to it at a later date. GeneSpring also uses the name and notes to find a particular project or experiment when you use the Search functionality.

For creating experiments, GeneSpring offers a choice of two workflow types:

1 A guided workflow (**Analysis: Biological Significance**) that provides data import and biological significance analysis of your data based on default parameter settings.

For data import, GeneSpring preselects the most appropriate parameters for your experiment type. These parameters are predefined and you cannot modify them. You can review these settings by inspecting the experiment after it is created. During the guided analysis steps you are able to modify some of the preselected parameters, for example p-value cut-off or flag filters. On completion of the workflow, GeneSpring creates an experiment with all the workflow results. You can now continue your analysis by choosing any of the advanced quality control, filtering, analysis, and biological interpretation steps from the Workflow Browser. This guided workflow option is described in detail in this quick start guide from onwards.

2 A data import only option (**Data Import Wizard**). This option is only recommended if you have advanced knowledge of the data import options in GeneSpring. It allows you to adjust flag settings, as well as data alignment and normalization parameters for importing your data. No further analysis steps are performed automatically (like in the guided workflow). GeneSpring immediately creates the experiment using your inputs. You can then select the appropriate experiment grouping, quality control, filtering, analysis, and biological interpretation steps from the Workflow Browser.

Terminology

Some terms commonly used in the general biological research community have a more specialized use in GeneSpring. The following table explains the most important GeneSpring - specific terms:

Table 1 GeneSpring-specific terminology

Term	GeneSpring Definition
• Project	A **project** is the primary workspace which allows analysis between experiments, across different technologies, and different organisms.
• Technology	A **technology** refers to the microarray type that was used to generate the data. For example, Agilent Whole Genome and Affymetrix HG-U133 PLUS 2 are two different array types. Technologies in GeneSpring contain biological information about all the genes on a specific array type. Install the appropriate technology for each new array type before you start analyzing data from that array. Select the Analysis type depending on the technology used when creating an experiment (see "Setting up a project" on page 15).
• Sample	A **sample** contains data from an array for a single biological source.
• Experiment	An **experiment** contains a group of samples used for a particular research study. An experiment consists of multiple interpretations that group these samples by user-defined conditions.
• Condition	A **condition** consists of one or more samples that represent a common biological state. For example, if you have tumor tissues from three different patients, these tissues describe the tumor condition. A different set of healthy patient tissue samples accordingly represents the normal condition.
• Interpretation	Multiple **interpretations** can be made from the same experiment data. Interpretations group samples into different conditions, if applicable to the study, and therefore allow alternative analysis approaches.

Table 1   GeneSpring-specific terminology

Term	GeneSpring Definition
• Entity	An **entity** is a discrete feature measured by microarray analysis such as a probe, probeset, gene, or protein. The term entity used in this guide appears infrequently in GeneSpring itself. GeneSpring features context-sensitive terminologies for lists, pathways, and trees. Therefore the term "entity" in this guide represents any one of eight possible entity types displayed in GeneSpring: **1** Exon **2** Feature **3** Gene **4** Probe **5** Probe Set **6** Protein **7** Splice Event **8** Transcript.
• Probe	A **probe** is a discrete feature on a microarray used to capture biological measurements. A probe can be an oligonucleotide or cDNA complimentary to a region of genomic DNA or a protein which interacts with other proteins.

User Interface

Main functional areas

The main functional areas of GeneSpring are shown in Figure 1.
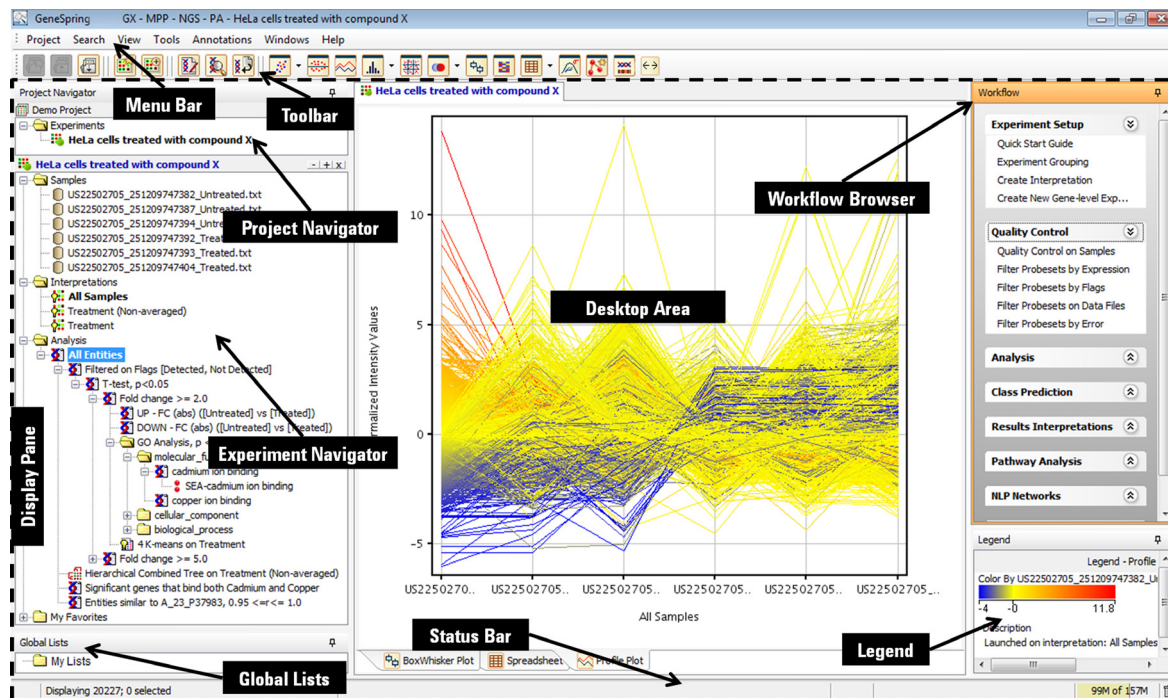

Figure 1   The main functional areas of the GeneSpring software

The main GeneSpring window consists of four parts:

• Menu Bar

• Toolbar

• Display Pane

• Status Bar.

Menu Bar   The menu bar shown in Figure 2 provides actions that are used for managing your projects, experiments, pathways, and display pane views.
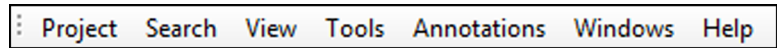

| ⋮ | Project | Search | View | Tools | Annotations | Windows | Help |

Figure 2   Menu Bar

Toolbar   The toolbar shown in Figure 3 on page 10 is located below the menu bar and contains four sections of icons providing fast access the following commonly used tasks:

- **Project section:** New project, Open project, and Close project.

- **Experiment section:** New experiment and Add experiment.

- **Entity List section:** Create entity list from selection, Inspect selected entity, and Import entity list from file.

- **Views section:**
  - **Scatter Plot** - shows a 2-D scatter of all entities in a selected entity list along the first two conditions of the active interpretation.
  - **MvA Plot** - plots the difference of probe measurements between two samples against the average as a scatter plot.
  - **Profile Plot** - plots normalized expression values against the selected interpretation.
  - **Histogram** - shows the frequency or number of entities across equal intervals of the normalized signal values for the first condition in the selected interpretation.
  - **Matrix Plot** - shows a matrix of pairwise 2-D scatter plots for conditions in the selected interpretation.
  - **Venn Diagram** - illustrates commonalities between entity lists or pathway lists in or across experiments.

- **Box-and-Whisker** - represents the distribution of the conditions in the selected interpretation with respect to the selected entity list in the experiment.
- **Heatmap** - visualizes the normalized signal values for all the entities in the selected entity list by the conditions in the active interpretation using a color range.
- **Spreadsheet** - presents a tabular view of your data.
- **Summary Statistics** - presents a summary of common statistical measurements (for example, mean, standard deviation, etc.) for the selected entity list grouped by the conditions in the selected interpretation.
- **Create new Pathway** - opens a new view with options to create your own interaction network or pathway.
- **Launch Genome Browser** - visualizes entities in a selected entity list against the reference genome for the experiment organism.
- **Hide Sidebars** - hides the Project and Experiment Navigator, Global Lists, Workflow Browser, and Legend and extends the Desktop Area to occupy the entire GeneSpring window.

[Project   ][Experiment][Entity List][Views - Plots, Statistics, Pathways   ]

Figure 3   Toolbar

Display Pane   The display pane, see Figure 1 on page 8, is further divided into six areas:

- **Project Navigator:** Displays the current project and lists all the experiments within the project.
- **Experiment Navigator:** Displays information related to Samples, Interpretation, Analysis, and My Favorites in respective folders related to the selected experiment in the Project Navigator. Each experiment within a project has a separate experiment navigator window.

- **Desktop Area:** Displays one or more interactive views associated with the experiments. You can configure each view in the desktop area separately. Window views can be arranged using Tile, Cascade, or Tabbed from the **Window** menu. Right-clicking anywhere in the active view shows you a menu of options (Figure 4) to customize the view, copy the view to the system clipboard, or export the view in popular image, html, or text file formats as specified in Table 2 on page 12.
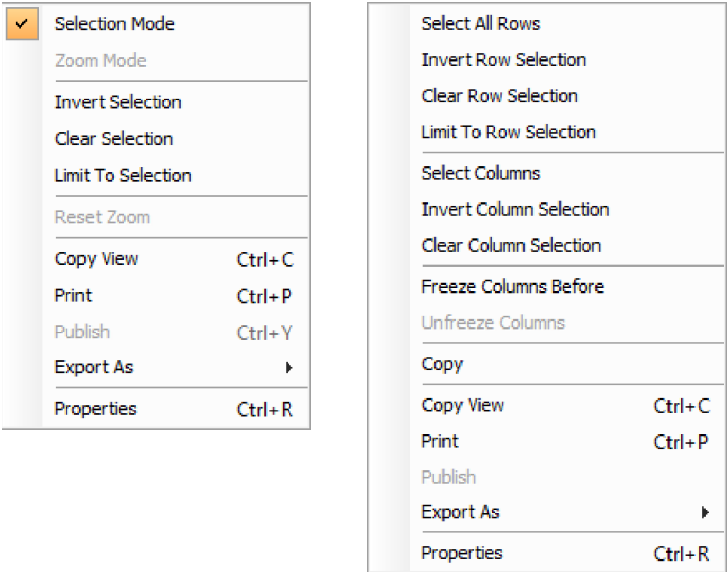

Figure 4   The right-click options available on plots or spreadsheets

Table 2   File export options for active views in the desktop area

Image	HTML	Text
*.jpg	*.html	*.txt
*.png		*.tsv
*.jpeg		
*.bmp		
*.tiff		

- **Legend:** The Legend (Figure 5) shows the key (scale) to the use of color in the active desktop view. Right-click the window title to copy and export the legend as described in the desktop area description.
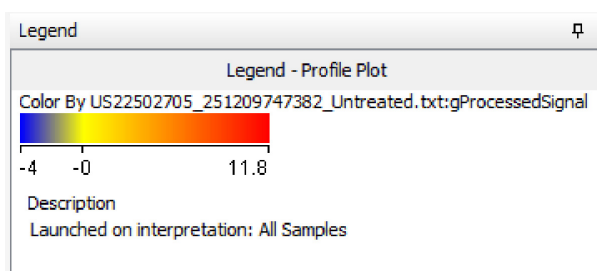

Figure 5   Legend

- **Workflow Browser:** The Workflow Browser (Figure 6 on page 13) is organized into sequential groups of operations for the analysis of your data:
  - Experiment Setup
  - Quality Control
  - Analysis
  - Class Prediction
  - Results Interpretations
  - Pathway Analysis *(optional)*
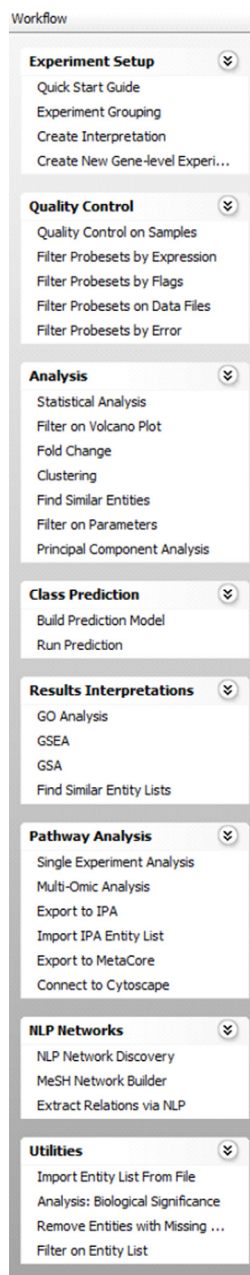 - NLP Networks *(optional)*
  - Utilities

Figure 6 Workflow Browser

- **Global Lists:** The Global Lists folder holds entity lists that you want to be able to use across different projects, rather than just in their original project.

Status Bar   The status bar (Figure 7) has three informative areas:

- **Status Area:** Displays high-level information of the current view. For example, the number of rows and columns in table views and the number of entities or conditions selected in plot views.
- **Ticker Area:** Displays the coordinates of the cursor in active plot views or the entity identification and value in table views.
- **Memory Monitor:** Displays the total memory being used and the total memory allocated by GeneSpring. You can click the garbage can icon at any time to reduce memory usage.
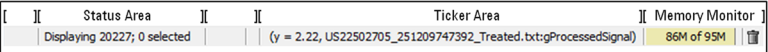

| [ ][| Status Area |][ ][| Ticker Area |][Memory Monitor] |
| | Displaying 20227; 0 selected | | (y = 2.22, US22502705_251209747392_Treated.txt:gProcessedSignal) | 86M of 95M |

Figure 7   Status Bar
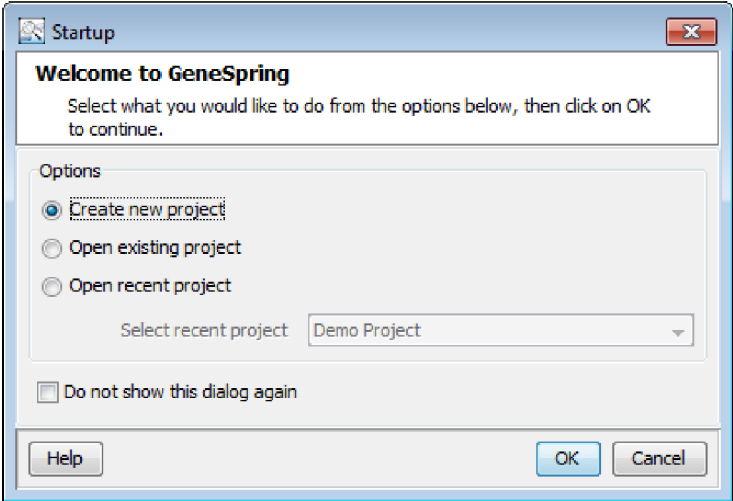
Setting up a project

GeneSpring guides you through four steps to create a new project and experiment to receive imported data:

1 **Startup:** Select creation of a new project.

2 **Create New Project:** Type descriptive information about the project.

3 **Experiment Selection:** Select **Create new experiment** as part of the project.

4 **New Experiment:** Type and select custom information to store with the experiment and import your data files.

Follow the steps described in this section to set up your new project. The Agilent Expression Single Color demo data set is used to illustrate each step.

The sample files you need are located in the **samples** folder of the main installation folder. If you did not specify another folder during installation, the location is: **C:\Program Files\Agilent\ GeneSpring\samples\Agilent Expression Single Color Demo**. You can substitute the demo information with information for your data.
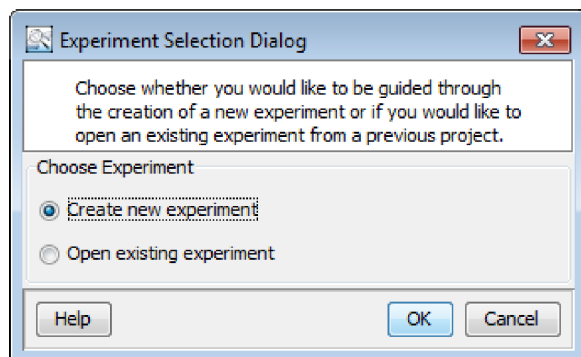
Steps	Detailed Instructions	Comments
1 Start GeneSpring.	a Double-click the GeneSpring shortcut icon on your desktop, or click **Start > All Programs > Agilent > GeneSpring > GeneSpring**.	• This opens the Startup dialog box.

Steps	Detailed Instructions	Comments
2 Create a new project in the **Startup** dialog box.	**a** Select **Create new project**. **b** Click **OK**.	• **Create new project** allows you to create a new project and new experiments or import existing experiments into the new project.

3 Alternatively, if you have already been working in GeneSpring, create a new project via the main menu, or the toolbar.	**a** Close the open project.
	b Create a new project from the
	  • main menu by selecting **Project > New Project**, or
	  • Toolbar by clicking the **New project** icon.
		• In GeneSpring, only one project can be open at any given time.

4 Type the descriptive information in the **New Project Details** area of the **Create New Project** dialog box.	**a** Type `Agilent Demo Project` in Name. **b** Type `Project containing the Agilent HeLa cells demo experiment.` in **Notes**. **c** Click **OK**.	• You can view and edit the project name and notes in the **Project Inspector** at any time. Click **Project > Inspect Project** from the menu bar to open the inspector.

Steps	Detailed Instructions	Comments
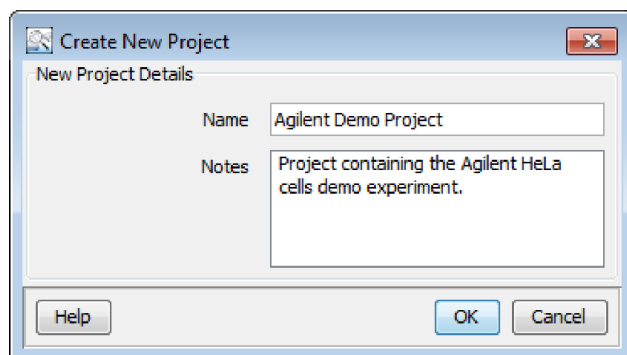
| **5** Select the option to create a new experiment in the **Choose Experiment** area of the **Experiment Selection Dialog** dialog box. | **a** Select **Create new experiment**.
b Click **OK**. | • If you select **Open existing experiment**, you can select an existing experiment to be added to the new project.
• Alternatively, you can create a new experiment to add to an existing, open project as follows:
From the main menu, select **Project > New Experiment**, or in the toolbar, click the **New experiment** icon. |

Steps	Detailed Instructions	Comments
6 Type and select information that guides the experiment creation in the **New Experiment** dialog.	**a** Type the descriptive name `HeLa cells treatment` for the experiment in **Experiment name**. This name can be different from the name previously entered for the project. **b** Select **Expression** for **Analysis type**. **c** Select **Agilent Expression Single Color** for **Experiment type**. **d** Select **Analysis: Biological Significance** for **Workflow type.** **e** Type `Agilent demonstration data` in **Experiment notes**.	• You can also access the **Analysis: Biological Significance** workflow from the Utilities section of the Workflow Browser in any open experiment that supports the guided workflow. • The Workflow type **Data Import Wizard** is only recommended for those with advanced knowledge of data import settings in GeneSpring.

	f Click **OK**.	• Your new project is now set up. • You are **immediately** guided through importing your data files in the **Load Data** part of the **New Experiment** dialog.

Steps	Detailed Instructions	Comments
7 Select your data files and reorder them if necessary in the **Load Data** part of the **New Experiment** dialog.	**a** Click **Choose Files**.	• This process to load data samples is common to all Workflow types.
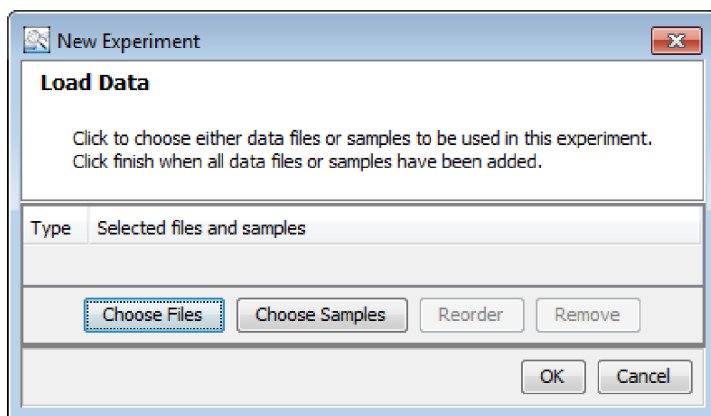
|		
	b The **Open** dialog box will most likely already point to the samples folder in the main GeneSpring installation folder. If not, browse to **C:\Program Files\ Agilent\GeneSpring\samples** and then select the **Agilent Expression Single Color Demo** folder.	
	c Browse for the proper data file types based on your data source selection.	
	d Click the sample expression data files to import into the experiment. The example HeLa cells data files are:	

    • US22502705_251209747404_Treated
    • US22502705_251209747393_Treated
    • US22502705_251209747392_Treated
    • US22502705_251209747394_Untreated
    • US22502705_251209747387_Untreated
    • US22502705_251209747382_Untreated

Steps	Detailed Instructions	Comments
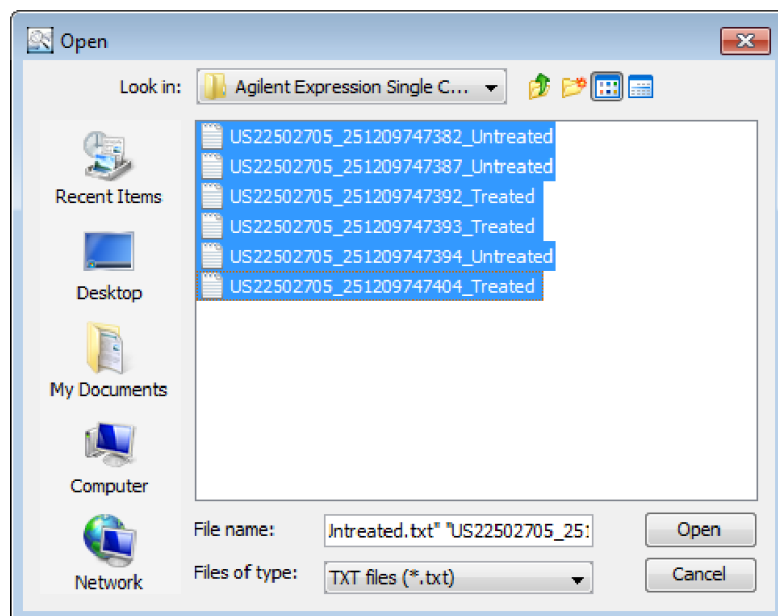
| | e Click **Open** to load the selected files for further preparation. | • A progress indicator is displayed while the files are being imported into GeneSpring. |

Steps	Detailed Instructions	Comments
	f If the samples do not appear in the required order, click **Reorder**.	• **NOTE:** This step represents the only opportunity to reorder your samples. If you want your samples to appear in a different order after you completed this step, you have to create a new experiment.

New Experiment

Load Data

Click to choose either data files or samples to be used in this experiment.
Click finish when all data files or samples have been added.

Type	Selected files and samples
📋	US22502705_251209747382_Untreated.txt
📋	US22502705_251209747387_Untreated.txt
📋	US22502705_251209747392_Treated.txt
📋	US22502705_251209747393_Treated.txt
📋	US22502705_251209747394_Untreated.txt
📋	US22502705_251209747404_Treated.txt

Choose Files   Choose Samples   Reorder   Remove

OK   Cancel

Steps	Detailed Instructions	Comments
	g Click one or more samples that you want to reorder in the **Reorder Samples** dialog box. **h** Click the **Up** [⬆] or **Down** [⬇] buttons to reorder the selected sample or samples. **i** Repeat the reordering steps as often as necessary to obtain your order. **j** Click **OK**.	• Select a continuous range of files with a click on a first file and a Shift-click on a last file that includes the range of files you want to select. • To select multiple samples that do not appear in sequence, Ctrl-click any sample name.

Reorder Samples

Samples
```
US22502705_251209747382_Untreated.txt
US22502705_251209747387_Untreated.txt
US22502705_251209747392_Treated.txt
US22502705_251209747393_Treated.txt
US22502705_251209747394_Untreated.txt
US22502705_251209747404_Treated.txt
```

[⬆] [⬇]

[OK] [Cancel]

Steps	Detailed Instructions	Comments
	k Click **OK** in the **New Experiment - Load Data** dialog box in step f) on page 21.	• What happens next depends on the workflow type that you chose in the New Experiment dialog box on page 18: • **Analysis: Biological Significance**: GeneSpring **immediately** launches the wizard-driven guided workflow. GeneSpring imports, thresholds, aligns, and normalizes your data in the background based on a number of default settings. The steps of the workflow are described in "Analysis: Biological Significance guided workflow" on page 24. • **Data Import Wizard**: The New Experiment dialog continues after sample import is complete. In the remaining steps, you can determine the flag settings, alignment, and normalization parameters for your data.

Analysis: Biological Significance guided workflow

The **Analysis: Biological Significance** guided workflow helps you identify differential expression in your data. The sample exercise in this section of the quick start guide illustrates the steps in the workflow with the same Agilent demonstration data set that was used in "Setting up a project" on page 15.

Following the import of your samples, the **Analysis: Biological Significance** workflow guides you through differential analyses between different conditions based on fold change and significance in eight steps. Some steps are automatically skipped for your experiment depending on the experiment type, experiment grouping, and conditions you specify during experiment creation and **Step 2** of this workflow.

1 Summary Report

Displays a summary view of your experiment. A box-and-whisker plot representing the distribution of data in each of the imported samples is displayed. If the number of samples is more than 30, GeneSpring displays your data in a spreadsheet view instead of the plot.

2 Experiment Grouping

In GeneSpring, independent variables are referred to as parameter names. Attribute values within an independent variable are referred to as parameter values.

Specify independent variables and the attribute values of the independent variables to define grouping of the samples. An independent variable is referred to as a parameter name. The attribute values within an independent variable are referred to as parameter values. Samples with the same parameter values within a parameter name are treated as replicates.

3 Quality Control on Samples

Presents samples by grouping and the current Principal Component Analysis (PCA). PCA calculates all the possible principal components and visually represents them in a 3D scatter plot. The scores shown by the scales of the axes are used to check data quality. The scatter plot shows one point

per sample color-coded by the experiment grouping. Additional quality metrics present in the imported feature extraction data files are displayed as a table and line plot.

4 Filter Probesets

Filtered (removes) entities from further analysis based on the presence of certain flag values across samples and parameter values (now referred to as a condition). In the absence of flags, entities are filtered based on signal intensity.

5 Significance Analysis

The entities are filtered based on their p-values calculated from a statistical analysis. The statistical analysis performed depends on the samples and experiment grouping.

6 Fold Change

Filters entities based on their abundance ratios or differences between a treatment and a control that are greater than a specified cut-off or threshold value.

7 GO Analysis

Finds enriched Gene Ontology terms for entities in the entity list resulting from **Step 6**.

8 Single Experiment Pathway Analysis

This is the last step in the guided workflow. It identifies matching entities between the entity list resulting from **Step 5** and previously downloaded pathways for the experiment organism. Before creating your experiment with the guided workflow, select **Tools > Import Pathways from WikiPathways** from the main menu to download pathways for the experiment organism from WikiPathways.

<table>
<tr><td>**NOTE**</td><td>You require a license for the Pathway Architect module to perform this last step. If you do not have a license, or do not want to download the required pathways, follow the workflow until **Step 7** and click **Finish** to complete the guided workflow.</td></tr>
</table>

The **Analysis: Biological Significance** workflow allows you to proceed through each step by clicking **Next >>**. A summary of your analysis is presented in each subsequent step. After

reviewing your analysis progress, you can return to a previous step and change the analysis parameters by clicking **<<Back**. Frequently use the **<<Back** and **Next >>** options to become more familiar with the analysis parameters and how the parameters affect your data.

To exit the wizard and skip the later steps in the wizard, click **Finish** at any step. When you click **Finish**, the All Entities entity list and all the entity lists resulting from the analysis steps performed so far are saved in the **Experiment Navigator.** You can now continue your analysis using the advanced operations available in the **Workflow Browser**.

Sample exercise

Steps	Detailed Instructions	Comments
1 Start GeneSpring and create a new project and experiment.	**a** Follow the steps described in "Setting up a project" on page 15 to create a new project and experiment with the Agilent demonstration data set.	
2 **Summary Report (Step 1 of 8)**: Review your data in the **Summary Report** workflow step.	**a** Review the summary report. Review the data, export selected data, or export the plot to a file using clicking and right-clicking features available on your summary report view. **b** Click **Next >>**.	• This view depends on the number of samples in your data. A spreadsheet is displayed if you have more than 30 samples. A profile plot is displayed if you have 30 or fewer samples.

Steps	Detailed Instructions	Comments
3 **Experiment Grouping (Step 2 of 8)**: Enter the sample grouping with respect to the independent variables and the replicate structure of your experiment in the **Experiment Grouping** workflow step.	**a** Click **Add Parameter...**. **b** Type a name for your Parameter name in the **Add/Edit Experiment Parameter** dialog box (Figure 8 on page 28). **c** Click your replicate **Samples** that share the first parameter value in your data. **d** Select the **Parameter type** for your grouping. **e** Click **Assign Value...**. **f** Type the value for your first grouping in the **Assign Value** dialog box. **g** Click **OK**. **h** Click your replicate **Samples** that share the second parameter value in your data. **i** Select the **Parameter type** for your grouping. **j** Click **Assign Value...**. **k** Type the value for your second grouping in the **Assign Value** dialog box. **l** Click **OK**. **m** Repeat the value assignment steps until you have assigned a parameter name, type, and value to all of your samples. **n** Review your entries and grouping assignment accuracy in the **Add/Edit Experiment Parameter** dialog box. **o** Repeat the value assignments for individual or multiple samples as necessary to make corrections or changes. **p** Click **OK** when the grouping for this parameter is complete.	• To proceed to the next step, assign at least one parameter with two values. • An independent variable is an essential element, constituent, attribute, or quality in a data set that is deliberately controlled in an experiment. An independent variable is referred to as a parameter and is assigned a parameter name. • The attribute values within an independent variable are referred to as parameter values. Samples with the same parameter values within a parameter name are treated as replicates. • Assignment of parameter names and values to the parameters is a process called experiment grouping. • **Parameter type**: Select **Non-numeric** if the grouping is not based on a quantitative value. Select **Numeric** if the grouping value is quantitative with respect to your parameter.

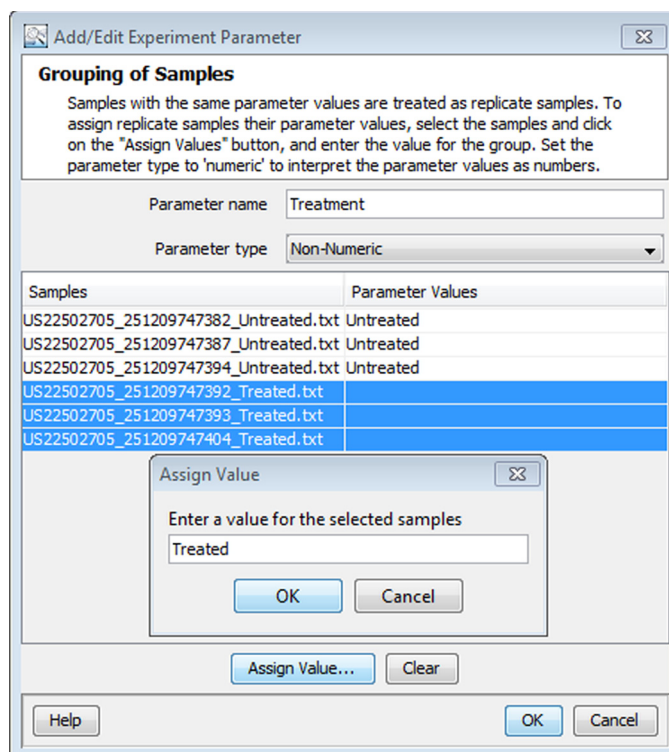| Steps	Detailed Instructions	Comments

Figure 8   Add/Edit Experiment Parameter and Assign Value dialog box

q If your data has more than one independent variable, repeat the **Add Parameter...** steps.

 • The experiment grouping for the Agilent demonstration data is shown in Figure 9 on page 29.

r Click **Next >>** when you have completed the experiment grouping. Click **<<Back** if you want to adjust settings in a prior step in the workflow.
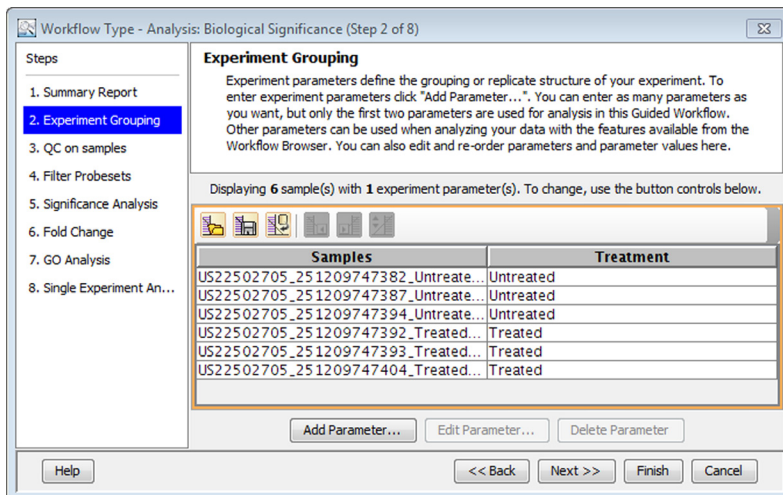
Steps	Detailed Instructions	Comments

Figure 9   Agilent demonstration data experiment grouping with one parameter and two parameter values per parameter.

4 QC on samples (Step 3 of 8): Assess the sample quality of your experiment using the **QC on samples** workflow step.	**a** Review your data, change the 3D PCA Scores plot view, export selected data, or export the plots to a file through clicking and right-clicking features available. • Good quality samples form discrete groups in the 3D PCA Scores view based on their parameter assignments. **b** *Highly recommended:* click **<<Back** to adjust settings in a prior step in the workflow to improve the QC on samples results. **c** Click **Next>>** when you have completed the QC on samples.	• QC on samples provides you with the first view of the data using a Principal Component Analysis (PCA). PCA allows you to assess the data by viewing a 3D scatter plot of the calculated principle components. • The Quality Control Metrics plot and report produced for this demo experiment are specific to Agilent Feature Extraction data files. If you used a different technology to generate your data, this plot and spreadsheet tab vary according to the QC-related information present in your data files.

Steps	Detailed Instructions	Comments
5 **Filter Probesets (Step 4 of 8)**: Select values that filter entities in your samples based on the quality of their presence in specified samples and conditions in the **Filter Probesets** workflow step.	**a** Review your data, change the plot view, export selected data, or export the plot to a file by clicking and right-clicking features available on the plots. **b** Click **Re-run Filter** to select acceptable flags in the **Filter Parameters** dialog box (Figure 10).	• If flags are detected in your data, Filter Probesets filters your data based on flags. If at least one of the imported samples has acceptable values, GeneSpring retains entities that have either a Detected or Not Detected flag value, by default. • A flag is a term used to denote the quality of an entity within a sample.

Figure 10   Filter Parameters dialog box

c Select the **Detected** and **Not Detected** options under **Acceptable Flags**.

d Unselect **Compromised** under **Acceptable Flags**. This flag is useful when you want to identify entities that are missing in the samples.

e Click **OK**.

f The number of entities displayed above the profile plot is expected to decrease in value as you progress through the workflow. Review the profile plot. You are encouraged to repeat these **Re-run Filter** steps until you obtain the best results for your experiment.

g Click **Next>>** when you have completed filtering the probeset. Click **<<Back** if you want to adjust settings in a prior step in the workflow.

• For Agilent microarray technologies, a flag indicates whether the entity was detected in each sample as follows: Detected - the entity was detected, Compromised - the entity was not detected, and Not Detected - the signal for the entity was saturated.

• For non-Agilent technologies, a flag indicates whether the entity was detected in each sample as follows: Present - the entity was detected, Absent - the entity was not detected, and Marginal - the signal for the entity was saturated.

• In the absence of flags, your data is filtered based on signal intensity values. By default, the filter removes entities in the lowest 20 percentile and displays a profile plot of the remaining entities.

Steps	Detailed Instructions	Comments

6 Significance Analysis (Step 5 of 8): Assess the differential significance of your samples from the **Significance Analysis** workflow step.	**a** Review your data, change the plot view, export selected data, or export the plot to a file by clicking and right-clicking the features available.
	b Customize the window panes within the view.
	c Move the **corrected p-value cut-off** slider or type a value to change the **p-value cut-off**. The default value is 0.05.
	d Adjust the **p-value cut-off** until the results displayed are satisfactory to your experiment.
	  • It is recommended that you adjust the **p-value cut-off** several times to develop an understanding of how it affects your results.
	  • A larger p-value passes a larger number of entities through to the final results.
	  • The last row of data in the Result Summary shows the number of entities that is expected to pass significance analysis by chance for each p-value specified in the column header. If that number is much smaller than the number of entities expected to pass based on the corrected p-value, your selection of entities shows significance among the parameter values.
	e *Highly recommended:* Click **<<Back** to adjust the settings in a prior step in the workflow to improve the Significance Analysis results.
	f Click **Next>>** when you have completed the Significance Analysis.
		• Significance Analysis filters the entities based on their p-values calculated from statistical analysis.
		• The statistical analysis is either a T-test or an Analysis of Variance (ANOVA) based on the samples and experiment grouping.
		• If any parameter value is associated with only one sample, p-values cannot be calculated and the Significance Analysis step is not displayed.
		• **Volcano Plot**: Display of the Volcano Plot or other plot depends on the samples and experiment grouping for analysis. Entities that satisfy the p-value cut-off appear in red color and the remaining entities appear in gray color. The Volcano Plot draws negative log10 of the p-value vs log of the fold change. Probesets with large fold change and low p-value can easily be identified on this view.

Steps	Detailed Instructions	Comments
7 Fold Change (Step 6 of 8): Enter values that filter the remaining entities in your samples based on their relative abundance ratios among the samples and conditions.	**a** Review your data, change the plot view, export selected data, or export the plot to a file by clicking and right-clicking features available. The fold change values for each entity are sign color coded: red for positive and blue for negative fold change by default. **b** Move the **Fold change cut-off** slider or type a value to change the **Fold change cut-off**. The default value is 2.0. **c** Adjust the **Fold change cut-off** until the results displayed are satisfactory to your experiment. 　• It is recommended that you adjust the **Fold change cut-off** several times to develop an understanding of how the **Fold change cut-off** affects your results. 　• A larger **Fold change cut-off** passes a smaller number of entities through to the final results. **d** *Highly recommended:* click **<<Back** to adjust the settings in a prior step in the workflow to improve the Fold Change results. **e** Click **Next>>** when you have completed the Fold Change step.	• Fold change is a signed value that describes how much an entity changes from its initial to its final value. For example, when an entity changes from a value of 60 to a value of 15, the fold change is -4. The quantity experienced a four-fold decrease. Fold change is the ratio of the final value to the initial value. • Fold change analysis is used to identify entities with abundance ratios that are in excess of a specified cut-off or threshold value. Fold change is calculated between two conditions where Condition 1 and another condition, Condition 2, are treated as a single group.

Steps	Detailed Instructions	Comments

8 GO Analysis (Step 7 of 8): Assess which of the entities retained following the Fold Change step show significance for a particular Gene Ontology (GO) category.	**a** Review your data, change the view, or export selected data from the spreadsheet or the visible area of the GO Tree through the available right-click options.
	b Move the **corrected p-value cut-off** slider or type a value to change the **p-value cut-off**. The default value is 0.1.
	c Adjust the **p-value cut-off** until the results displayed are satisfactory to your experiment.
	  • It is recommended that you adjust the **p-value cut-off** several times to develop an understanding of how it affects your results.
	  • A larger p-value passes a larger number of entities through to the final results.
	d Click **Next>>** when you have completed the GO Analysis.
		• The GO classification scheme allows you to categorize genes by biological process, molecular function, and cellular component.
		• Genespring provides two display options for your results: Spreadsheet and GO Tree:
		  • The Spreadsheet displays all entities satisfying the p-value cut-off sorted by corrected p-value.
		  • The GO Tree highlights all the GO terms with entities satisfying the p-value cut-off within the GO Tree hierarchical view.

Steps	Detailed Instructions	Comments
9 Single Experiment Pathway Analysis (Step 8 of 8): Review which of the entities retained following the Fold Change step have matching entities in pathways for your experiment organism.	**a** Review your data or export selected data by clicking and right-clicking features available. **b** To save a selection of the listed pathways, select the desired pathways (use Ctrl-Click to select multiple pathways) and click **Custom Save**. **c** Click **Finish** when you have completed the Single Experiment Pathway Analysis. **d** The result of Single Experiment Pathway Analysis is saved in the Experiment Navigator as a pathway list. You can filter this list of pathways by p-value and number of matching entities after completing this workflow.	• The Single Experiment Pathway Analysis workflow step is only available if 　• you have a license for the GeneSpring Pathway Architect module, and 　• you have imported pathways from the WikiPathways portal (via **Tools > Import pathways from WikiPathways**) for the experiment organism before starting the **Analysis: Biological Significance** workflow. • Single Experiment Pathway Analysis identifies matching entities between the entity list created as a result of Fold Change Analysis (**Step 5**) and pathways for the experiment organism. In this step, GeneSpring also computes p-values to establish the significance of the matching entities in a pathway.

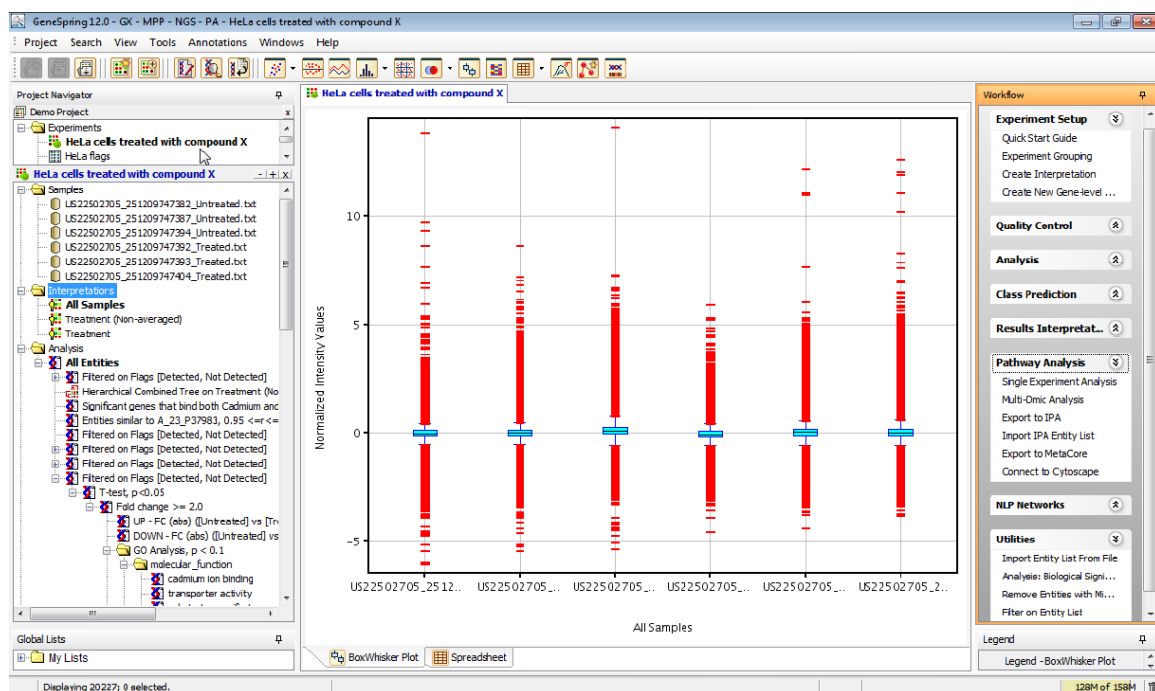| Steps	Detailed Instructions	Comments
10 Review your results in the experiment that GeneSpring creates on completion of the guided workflow.	**a** Review your results in the box-and-whisker plot (Figure 11).	• The **Analysis: Biological Significance** workflow is now complete.

Figure 11 The GeneSpring screen after completing the Analysis: Biological Significance workflow using the demonstration data set supplied by Agilent.

11 Continue your analysis.	• Perform custom analyses on your data using the operations available in the Workflow Browser.	

Where to find additional information

Online help

Press F1   To get more information about a menu, toolbar, window, or dialog box, place the cursor on the part of the menu, toolbar, window, or dialog box of interest and press the **F1** key.

Help menu   Click **Help > Documentation Index** to access the release notes, quick start guides, and HTML and PDF versions of the GeneSpring manual.

Online support portal   Visit www.genespring-support.com to contact the GeneSpring support team and access additional demo data sets, user guides, and video tutorials.

E-seminars   Visit our e-seminar calendar to register for live seminars on available features in GeneSpring.

Documents

GeneSpring manual   Agilent GeneSpring User Manual (Agilent publication n/a, August 2012).

References

1 Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. (2008) WikiPathways: Pathway Editing for the People. *PLoS Biol* 6(7)

www.agilent.com

In this guide

This guide gives you an overview of the GeneSpring GX module of the GeneSpring suite of products. It also describes the Analysis: Biological Significance guided workflow for microarray data.

G9220-90002

Agilent Technologies