

GeneSpring GX 10

Gold standard software for: Micro Array data analysis, Detect alternative splicing events, Identify microRNAs, Analyze real-time PCR data, and Build relevant biological interaction networks & pathway analysis

Guided workflow analysis - Step through typical analysis with limited decision points. Advanced workflows allow expert user to access all GeneSpring GX tools and define parameters and more

Suite of statistical tools - Tools include paired and unpaired t-tests, 1-way and multi-way ANOVA, Repeated Measures ANOVA, and permutative method of p-value calculation

Tool set for uncovering unique patterns - Use Principal Component Analysis or clustering methods such as Hierarchical, k-means, and SOM and quickly graphically visualize your data

Algorithms for pattern correlation - Use training set samples to identify entities whose patterns of expression correlate with disease phenotypes and use them to classify unknown samples

Results viewable in biological context - Place statistically significant results in biological context using GO analysis, Gene Set Enrichment Analysis (GSEA), Gene Set Analysis (GSA), and pathway analysis

Extend functionality through programming - Use **JYTHON Application Programming Interface** (API) and fully integrated R programming language to extend GeneSpring GX functionality

Software Features

GeneSpring GX 10 takes a big step beyond providing quick, easy, and robust gene expression analysis. As systems-level studies become more prevalent in functional genomics research, GeneSpring GX adds tools for integrative data analysis critical to the discovery of linkages and concordance between different data types.

GeneSpring GX also expands its pathway analysis capabilities by providing a database of gene product interactions from which scientists can build biological interaction networks from their genes of interest. Frequent product updates and a persistent emphasis on making powerful statistical tools accessible through guided analysis and intuitive interface have placed GeneSpring GX at the top of the list of preferred analysis applications within the community of gene expression biologists. GeneSpring GX has 4,400 references in Google Scholar, including over 1,600 in peer reviewed publications.

GeneSpring GX Key Features

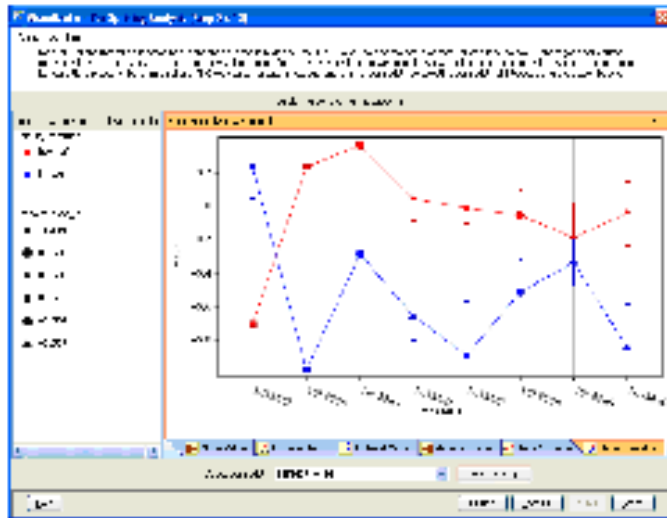
Support for new applications

GeneSpring GX adds visualization and analysis tools to allow identification of interesting biological events such as alternative splicing or changes in miRNA and gene expression. As much as 75 percent of human mRNAs undergo alternative splicing, generating splice variants that can produce proteins with distinct biological functions. GeneSpring GX provides multivariate splicing ANOVA to detect statistically significant differences in splicing events between experimental conditions. These splicing events can then be visualized in various intuitive plots.

MicroRNAs have been shown to be important regulators of gene expression. GeneSpring GX not only provides statistical tools to detect significant changes in miRNA abundance, but integrates TargetScan miRNA gene target information to determine the biological consequences of the detected changes in miRNA abundance. Results from microarray experiments are necessarily validated by more quantitative methods such as real-time PCR.

To facilitate validation of microarray experiments, GeneSpring GX supports analysis of real-time PCR data. Results from different data types can also be directly compared. For example, intensity values from real-time PCR experiment can be plotted against their corresponding intensity values from the microarray experiment. Support for new applications in GeneSpring GX includes:

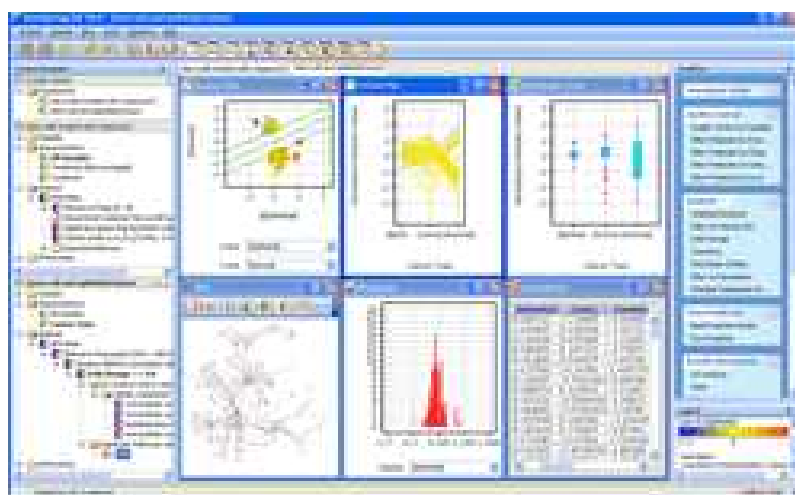
- Agilent miRNA array
- Affymetrix GeneChip Exon array
- Applied Biosystems 7900HT Real-time PCR system



Transcript clusters with significant differences in alternative splicing events can be detected using multivariate splicing ANOVA. Each transcript cluster can then be visualized by plotting the intensity values for its corresponding probesets in each experimental condition. Probesets are ordered according to its chromosomal location.

Project-based hierarchical data organization

GeneSpring GX data are organized in projects. Each project can contain one or more experiments of different data types, array platforms, or organisms. This organization enables users to easily compare results from related experiments contained in the same project. For each experiment, the results of the analysis are stored in a hierarchy that maintains the order and dependency of how the results were created. Any data objects in this hierarchy can be copied into a My Favorites folder, allowing you to segregate your most important data objects and organize it in anyway you desire.



Project-based data organization enables you to store data from related experiments in a single workspace, regardless of array type or technology used. This organization facilitates comparison of results from related experiments. Multiple visualization tools allow you to interrogate different aspects of the data. Views are linked such that selection of entities in one view automatically highlights the same entities in all other views.

Intuitive graphical displays

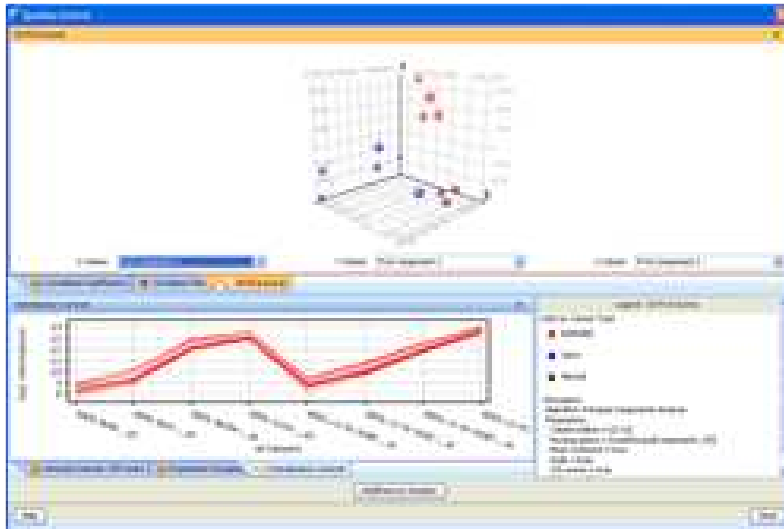
GeneSpring GX displays expression data in ways that help you conceptualize the information in your data and convey it to your peers. The various types of plots, graphs and diagrams highlight different aspects of the data, allowing you to extract visual information in different ways. Virtually any graphical image can be exported as HTML or as image in .tiff, .jpeg, .png or .bmp format compatible with publishing software applications. Visualization tools in GeneSpring GX include:

- Profile plot
- 3D scatter plot
- Histogram plot
- Volcano plots
- Box-Whisker plot
- Matrix plot
- MvA plot
- Venn diagram
- Spreadsheet
- 2D dendrogram
- Heatmap
- Chromosome map

Guidance for most common tasks

Guided Workflows step a user through a defined series of the most common analyses performed on expression data. Each Guided Workflow is specific for the technology type of the data. For example, the Guided Workflow for Agilent data will use the QC metrics produced by Feature Extraction software to help assess the processing quality of your samples in GeneSpring GX. The Guided Workflow for Affymetrix data will use Affymetrix specific metrics, such as hybridization control plots and 3'/5' ratios. Once you have finished the Guided Workflow, all data objects will be saved to the Navigator, where they will be available for further analysis in the Advanced Analysis mode. The Advanced Analysis mode gives you full control over all GeneSpring GX tools and functionalities, allowing you to specify all analysis parameters and cut-offs. Guided Workflows in GeneSpring GX include finding differentially expressed genes for:

- Agilent One-color Expression Array
- Agilent Two-color Expression Array
- miRNAs for Agilent miRNA Array
- Affymetrix GeneChip® Expression Array
- Affymetrix GeneChip® Exon 1.0ST Array
- Illumina Beadchips Expression Array



GeneSpring GX incorporates vendor specific sample quality control metrics within an intuitive interface to facilitate the identification of poor quality samples. These samples can then be removed from the experiment, allowing you to proceed with analysis using only data from high-quality samples.

Data normalization and transformation methods

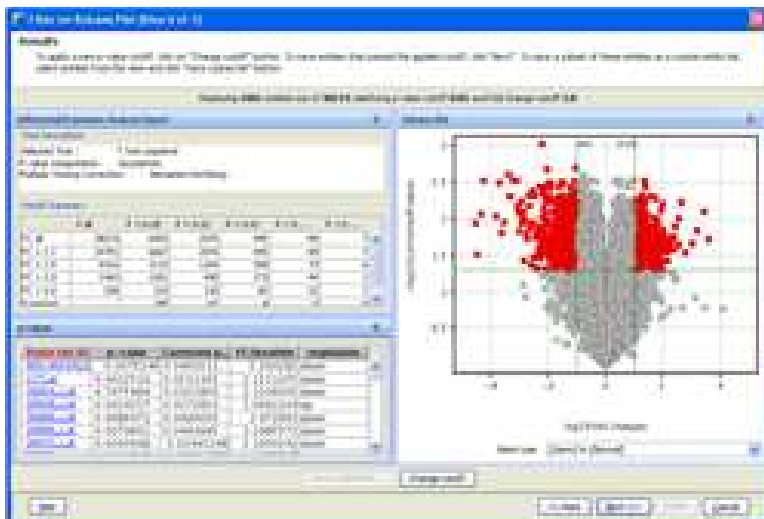
Despite tightly controlled experimental conditions, technical bias invariably is introduced into your dataset. GeneSpring GX provides a suite of normalization methods that will help limit the technical variation in the data so that true biological variations are revealed. These normalization methods include:

- Probe-level and probeset level summarization for Affymetrix data:
 - RMA
 - GC-RMA
 - PLIER 16
 - MAS 5
 - Li Wong
- Percentile shift
- Quantile normalization
- Scaling
- Normalize to control genes
- Intensity dependent (LOWESS) normalization
- Dye swap
- Baseline transformation to median of all samples or median of control samples

Statistical tools for testing differential expression

GeneSpring GX provides a large variety of statistical tools that can be applied to test for differential expression in experiments of various designs. Powerful statistical tools accessible through GeneSpring GX's flexible and easy-to-use interface include:

- T-tests
 - Parametric and non-parametric
 - Paired and unpaired
- One-way ANOVA
 - Parametric and non-parametric
- Multi-way ANOVA
 - Balanced and unbalanced two-way ANOVA
 - Three-way ANOVA
- Repeated Measures ANOVA
- Tukey and Student-Newman-Keuls post-hoc tests
- Multiple testing correction methods
- Permutative and asymptotic p-value computation



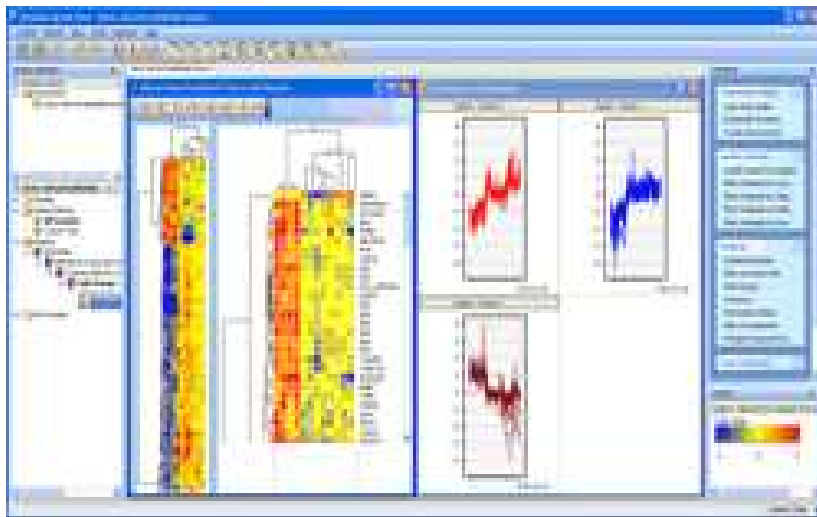
Statistical analysis result window allows you to save a list of entities that were called to be differentially expressed at the specified level of significance. Relevant information such as the calculated p-values and corrected p-values are saved along with the significant entities.

Pattern discovery

GeneSpring GX gives you a broad choice of tools for uncovering unique patterns in your expression data. Principal Component Analysis (PCA) reduces the dimensionality of complex datasets by identifying principal components that represent the most dominant profiles in the dataset. Clustering algorithms can be employed to group entities and/or samples based on the similarity of their expression profiles. Results from such an analysis may reveal information regarding the biological function or co-regulation of the genes

belonging to the same cluster. In addition, GeneSpring GX allows you take an entity of interest, perhaps a gene that you know plays an important role in the disease that you are studying, and identify other genes with similar expression profiles. Results from such an analysis may allow you to discover other genes that play an important role in the disease process under study. Pattern discovery tools available in GeneSpring GX include:

- PCA on Entities or Conditions
- Clustering Algorithms
 - Hierarchical
 - Self-organizing maps
 - k-means
- Find Similar Entities tool



GeneSpring GX offers three different clustering algorithms to help identify unique expression patterns in the dataset. Entities and samples can be grouped according to the similarity of their expression profiles. An entity list for each cluster identified can be created, allowing closer interrogation of the genes found within a cluster.

Class prediction

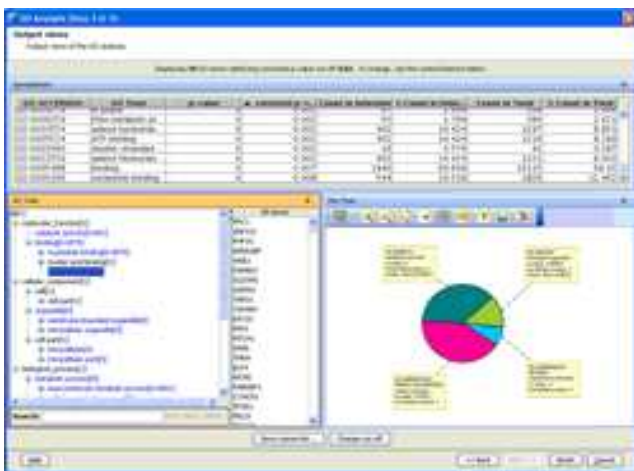
Classification methods applied to high throughput expression data have proved useful in identifying genes whose patterns of expression correlate with disease phenotypes or other qualitative class labels. Using the expression data for these “predictor genes” from samples of known classification as input allows you to predict the classification of unknown samples. GeneSpring GX provides several powerful classification algorithms for class prediction analysis:

- Support Vector Machines
- Neural Network
- Naive Bayesian
- Decision Tree

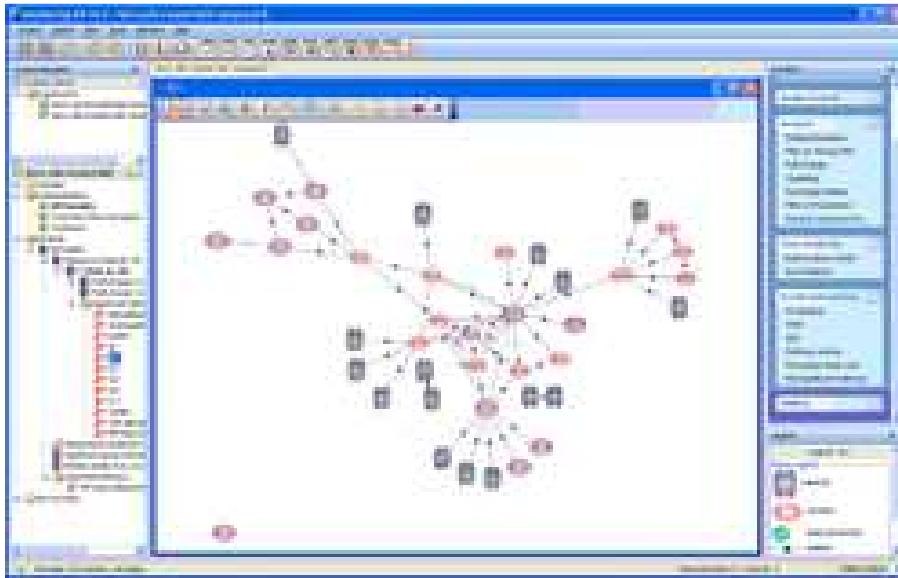
Biological contextualization

One of the most critical steps of data analysis is to put statistically significant findings into a biological context. Instead of looking at your results at the individual gene level, you can explore what biological process, molecular function, and biological pathways are involved or are affected in your study. GeneSpring GX provides many methods of updating technology gene annotations to help you put results in a biological context. Determining the direct or indirect molecular interactions between identified genes of interest can also shed light into the underlying mechanism of the disease or biological process you are studying. Biological contextualization tools provided in GeneSpring GX include:

- Update technology annotation
 - from Agilent eArray
 - from Agilent server
 - from NCBI
 - from text file
- GO enrichment analysis
- Gene Set Enrichment Analysis (GSEA) and Gene Set Analysis (GSA)
 - Use of Gene Sets from the Broad Institute
 - Use of any list of genes as gene sets
- Pathway Analysis
 - Import and view BioPAX pathway exchange format (OWL)
 - Find Similar Pathways tool
 - Build biological interaction networks
 - Relevance Interaction Network analysis
 - Create networks using interactions mined by next-generation natural language
- Integration with Ingenuity Pathway Analysis (IPA)



The GO Analysis tool helps identify biological processes and functions that may be impacted in your experiment by calculating the enrichment of gene lists with genes from GO categories. Pathway and network diagrams help place statistical results in a biological context. Direct navigation between biological pathways and their associated genes provides a rich user experience and systems-level insight.



Multi-platform compatibility

GeneSpring GX lets you analyze data from all modern array platforms, such as Agilent, Affymetrix and Illumina expression analysis systems. GeneSpring GX guides users to the proper normalization algorithms to enable valid combinations and comparisons using expression profile data derived from different hardware platforms. Supported array data files include:

- Agilent Feature Extraction files
 - FE version 8.5 or newer
- Affymetrix CEL, CHP and ARR files
 - Text and binary (XDA) formats (version 3 and 4)
 - New AGCC data format (3'-IVT and Exon expression arrays)
- Illumina BeadStudio files
 - Sample Probe Profile GeneSpring export format
 - Group Profile GeneSpring export format files can be imported, but each group will be loaded as one sample
 - Sample Gene Profile GeneSpring export format can be import as custom format in custom technology
- Applied Biosystems SDS RQ files
 - SDS 2.1, 2.2, 2.3 RQ and RQ Manager 1.2 files
- Molecular Devices GenePix Pro files
 - GenePix Results GPR files version 1.4 or newer
- Tab-delimited text files

Extensible functionality with JYTHON and R

Programmers can make use of GeneSpring GX documented JYTHON (Python for JAVA) Application Programming Interface (API) to extend GeneSpring GX functionality. The API enables incorporation of third-party applications such as SAS and MATLAB®. In addition to the JYTHON programming language, the R programming language is fully integrated into GeneSpring GX and can be used to implement novel algorithms.