Multi-omic Analysis with Agilent's GeneSpring 11.5 Analysis Platform

Product Note

Analyze many types of biological data simultaneously

Compare data from many experiments using a single application

New workflow for Agilent's Exon microarray data

Biological research involves producing and integrating many types of large, 'omics-level datasets. This heterogeneous data can measure everything from gene expression changes to proteomic and metabolomic abundance levels. The ability to compare different data types is important in multi-omic research. With GeneSpring 11.5, researchers can easily identify key similarities among data that directly contributes to a more complete understanding of the underlying mechanism of disease.

Integrative platform for multi-omic data analysis

GeneSpring 11.5 provides comprehensive analytical and visualization tools for multiple data types within a single data analysis application. With this platform, heterogeneous data such as gene expression, miRNA, exon splicing, genomic copy number, genotyping, proteomic, and metabolomic abundance can be combined into one project, allowing researchers to analyze and view results from different experiments in a single user interface (Figure 1). Additionally, GeneSpring 11.5 offers capabilities that link probes across data types, array platforms, and organisms that map to the same biological entity. By supporting analysis and translation of multiple data types in a single application, GeneSpring 11.5 avoids common issues of software interoperability and faulty semantic mapping to increase a researcher's ability to leverage concordances among data types.

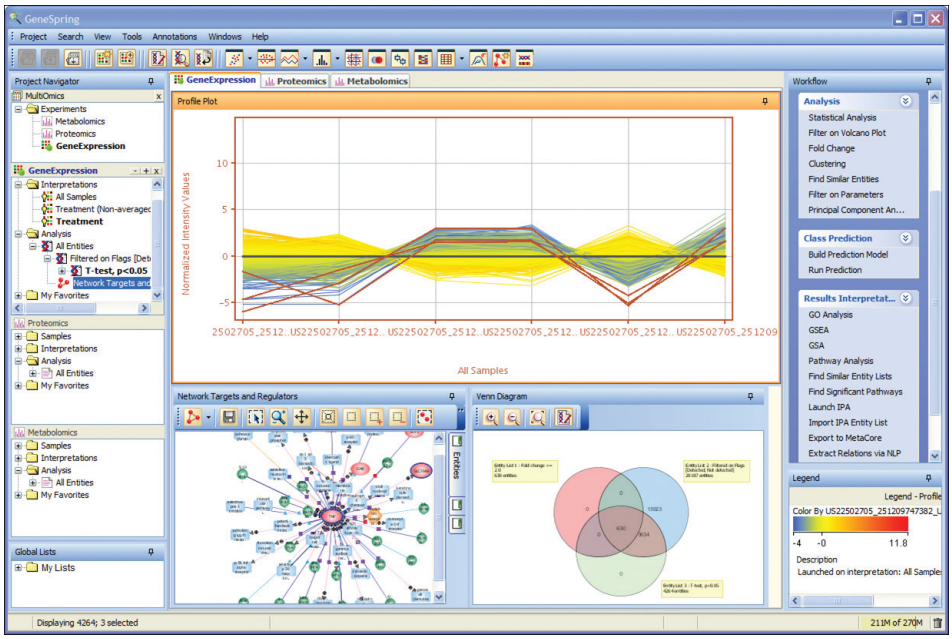

Figure 1. *Project-based data organization enables researchers to store data from related experiments in a single workspace, regardless of data type, array platform, and organism.*

Transcriptomic analysis

GeneSpring 11.5 provides flexible and comprehensive workflows for a variety of transcriptomic applications. Several new features that leverage the content on Agilent Exon Microarrays have been implemented, including a new workflow, statistical routines, and a host of enhanced visualizations (Figure 2). A broad spectrum of data pre-processing and normalization methods are available for both one- and two-color gene expression data. Depending on the researcher's level of expertise, data analysis can be performed using either the Guided Workflow or Advanced Analysis mode. Quality control can be performed using platform-specific metrics, enabling researchers to optimize pre-processing steps before statistical analysis.

Other key features for transcriptomic applications include:

- Probe-or gene-level expression analysis on all major microarray platforms, including Agilent, Affymetrix, and Illumina
- microRNA analysis and identification of gene targets using integrated TargetScan information
- Exon splicing analysis using t-tests or multivariate splicing ANOVA and filtering for transcripts on splicing index
- Real-time PCR data analysis
- NCBI Gene Expression Omnibus Importer tool for expression datasets
- MAGE-TAB Export Tool

Genomic copy number analysis

GeneSpring 11.5 supports the interrogation of genomic structural variations and their implications in disease susceptibility and progression. Providing workflows for paired and unpaired analysis of Affymetrix and Illumina genotyping data, GeneSpring 11.5 enables scientists to quickly detect regions of genomic copy number variation (CNV) and loss of heterozygosity (LOH). Once regions of interest are discovered, genes overlapping those regions can be identified and the biological impact of copy number variation can be assessed in downstream GO analysis or pathway analysis.

Features of the genomic copy number workflow include:

- Ability to create and use a custom reference in addition to pre-packaged HapMap reference
- Batch effect correction method
- Circular Binary Segmentation
- Filters to identify copy-neutral LOH events and regions of allelic imbalance
- Ability to identify common variations across a set of samples

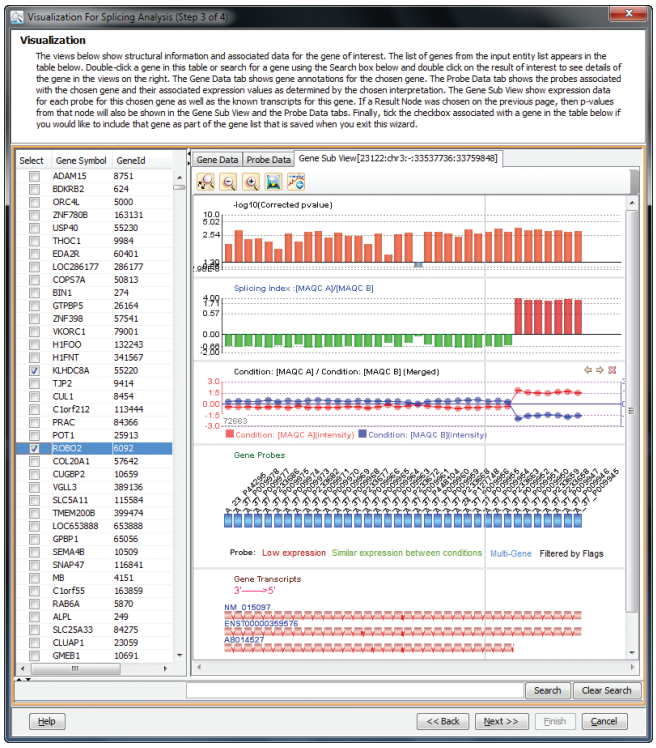

Figure 2. *GeneSpring 11.5 incorporates a new workflow, statistical routines, and a host of enhanced visualizations specifically created for the analysis of Agilent Exon microarray data.*

Genome-wide association analysis

Genome-wide association studies (GWAS) utilize high-density genotyping microarrays to identify SNPs associated with qualitative or quantitative traits. GeneSpring 11.5 provides a comprehensive workflow for case-control GWAS using Affymetrix and Illumina genotyping microarray platforms. The flexible workflow offers a suite of statistical tests applied under various genetic models as well as multiple testing correction and correction methods for population stratification. GeneSpring can also perform filtering by HWE, known SNPs, and the Database of Genomic Variants. After identifying genes harboring SNPs or haplotype blocks associated with trait, researchers can perform GO analysis and pathway analysis to determine what biological process and pathways may be involved in the disease under study.

Other key features for the genetic association workflow include:

- EIGENSTRAT and Genomic Control population stratification correction
- Tag SNP identification
- Haplotype inference and Haplotype Trend Regression
- Pearson's Chi-Square, Fisher's Exact, Cochran Armitage and Chi-Square correlation
- Logistic and linear regression
- LD plot (Figure 3)

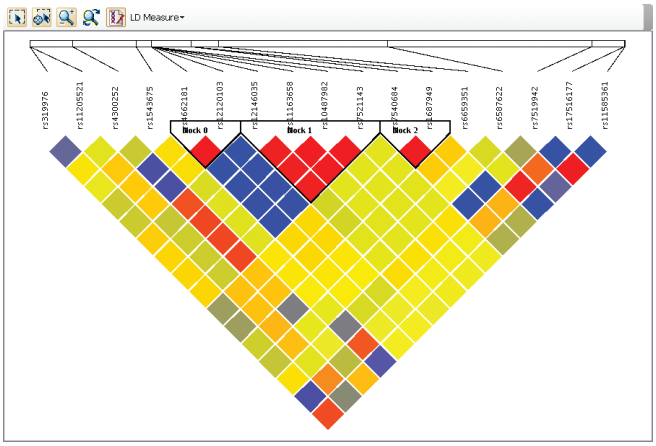

Figure 3. *LD plot provides fine-grained representation of the LD blocks, which are colored according to the pair-wise correlation provided as D′ or r^2 values. SNPs within regions of strong LD can be saved as entity lists for further analysis.*

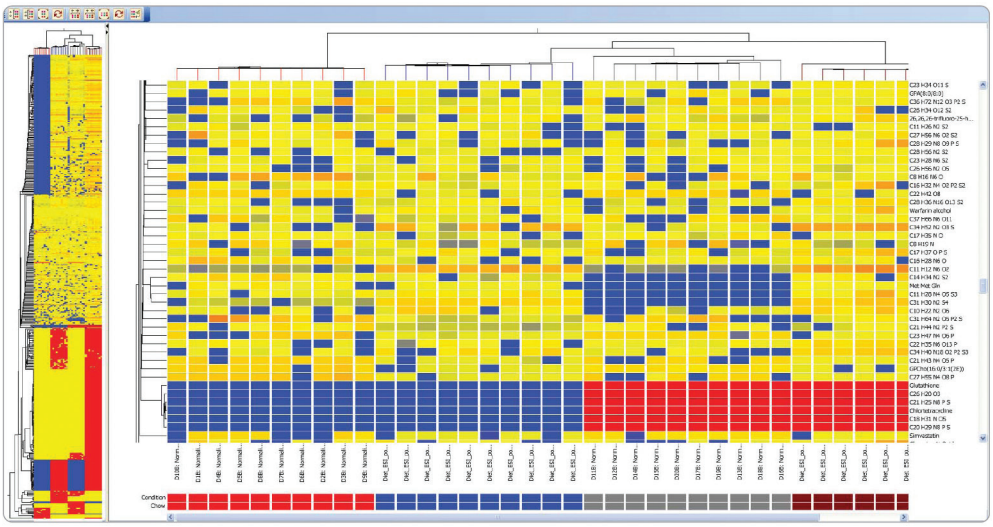

Figure 4. *Hierarchical clustering connects similar abundance profiles together in a group within a tree structure. The dendrogram view reveals the relationships between mass entities in one dimension and between samples in the other dimension. In this example, rat chow diets with differing bean supplements show similarities in abundance profiles for the replicate samples.*

Proteomic and metabolomic analysis

GeneSpring 11.5 can be used to analyze many different types of mass spectral data through integration with Agilent's Mass Profiler Professional, including GC/MS, LC/MS, CE/MS, and ICP-MS. An optional mzXML package expands these capabilities even further for other instruments. Recursive analysis—a unique feature of Mass Profiler Professional—lets users easily re-mine data sets based on preliminary data to improve the quality of statistical analysis results. It's also easy to export an inclusion list for Q-TOF MS/MS analysis and re-import the results into Mass Profiler Professional. A single, consistent user interface supports all data types, and advanced visualization tools let users inspect and annotate their results in new ways (Figure 4).

With functionality for unsupervised (classification without prior group assignments) or supervised (using pre-classified groups) analysis, the software lets users:

- Quickly and easily discover differences between sample groups
- Plot changing patterns of compound abundances over time
- Develop useful multivariate models for class prediction

Built-in ID browser automates database and spectral library searches

Converting entities into actual chemical compounds using Mass Profiler Professional is made easy, using comparative techniques against public or private databases. The software automatically annotates the entity list and projects the compound names onto any of the various visualization and pathway analysis tools.

Mass Profiler Professional includes an integrated ID browser (Figure 5) that mirrors MassHunter's qualitative analysis functionality to allow identification using:

- LC/MS Personal Compound Database (METLIN, pesticides, forensics)
- GC/MS libraries (NIST and Fiehn library)
- Empirical Formula Calculation using Agilent's Molecular Formula Generator (MFG) algorithm

Comprehensive analytical and visualization toolkit

At the core of GeneSpring 11.5 is a set of statistical tests, algorithms, and visualization tools that allow researchers to analyze a broad range of experimental designs.

Statistical tools for testing differential expression

GeneSpring 11.5 provides a comprehensive suite of statistical tests so that differential analysis can be applied robustly on a variety of experimental designs. GeneSpring 11.5 statistical toolkit includes parametric and non-parametric tests that can be applied to paired and unpaired experimental designs, permutative and asymptotic p-value computation, and multiple testing correction methods, including permutation-based options. Post-hoc statistical tests can also be applied to pinpoint pairs of experimental conditions where differential expression is detected. Multivariate analysis is available to test the effects of each factor, and their interaction, on changes in expression across experimental conditions.

Pattern discovery

GeneSpring 11.5 provides a broad choice of tools to identify unique patterns in data. Clustering algorithms can be employed to group entities and/or samples based on the similarity of their expression profiles, revealing information regarding the biological function or the co-regulation of genes. GeneSpring 11.5 also provides robust classification algorithms that use training datasets to find clinically predictive expression patterns. By offering multiple classifiers including Decision Tree, Support Vector Machine, Naive Bayesian, Neural Network, and Partial least squares discrimination, GeneSpring 11.5 enables biomarker discovery for a variety of experimental designs.

Extensible functionality with Jython and R

GeneSpring 11.5 provides scientists the ability to write, execute, and save their own scripts to combine operations in GeneSpring 11.5 with a more general Jython (Python with Java class import capabilities) programming framework. Users can develop their own data transformation operations, automatically pull up data views, and run external algorithms within GeneSpring 11.5. An embedded R scripting editor allows R scripts to be written and run from within GeneSpring 11.5. Any R function can be given access to GeneSpring 11.5 data, with results being automatically incorporated back into the GeneSpring 11.5 environment.

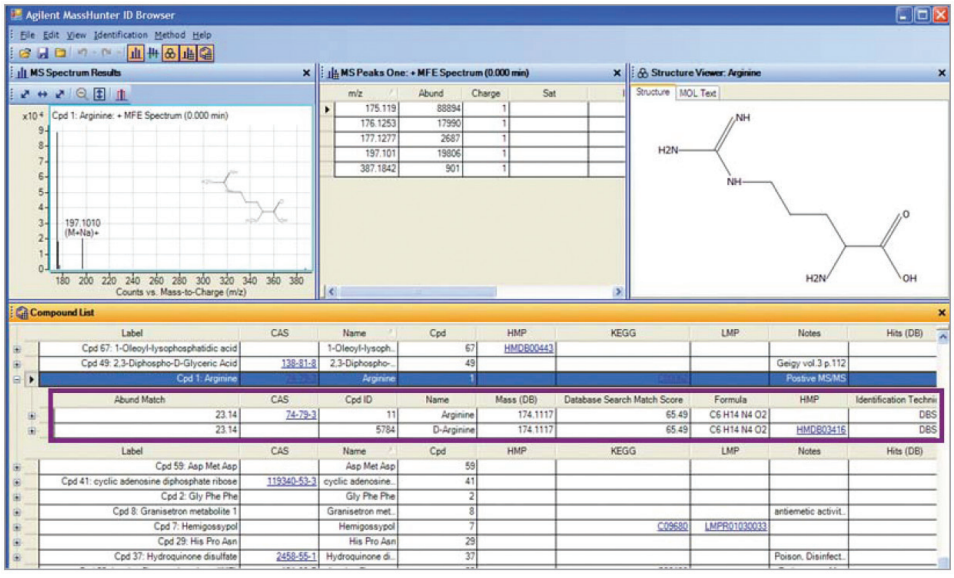

Figure 5. *MassHunter ID Browser identifies compounds from an entity list generated in Mass Profiler Professional. In this example, results for mass 174.1117 matched against the METLIN database reveals it to be arginine; molecular formula, database match score, name, KEGG, and CAS ID are also shown.*

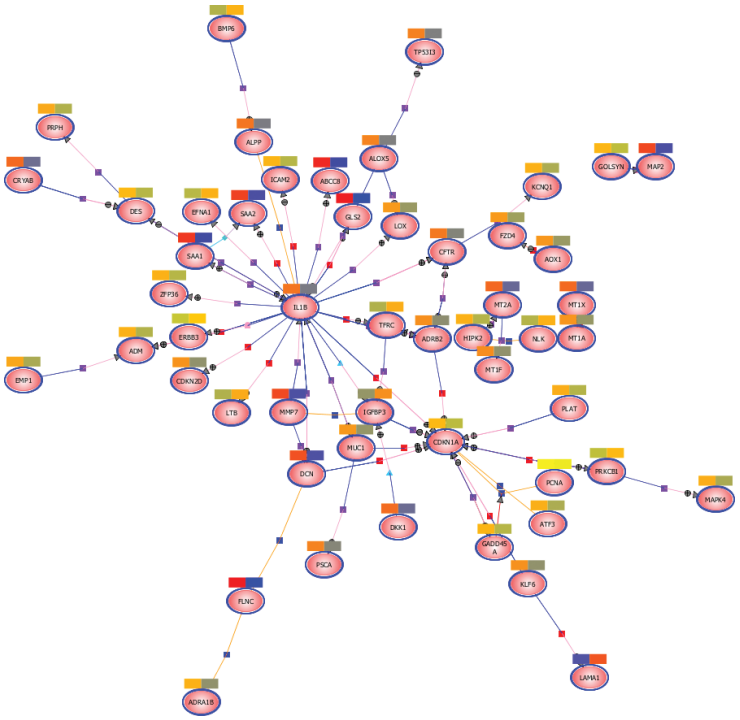

Figure 6. *Pathway and network diagrams help place statistical results in a biological context. Direct navigation between biological pathways and their associated genes provides systems-level insight.*

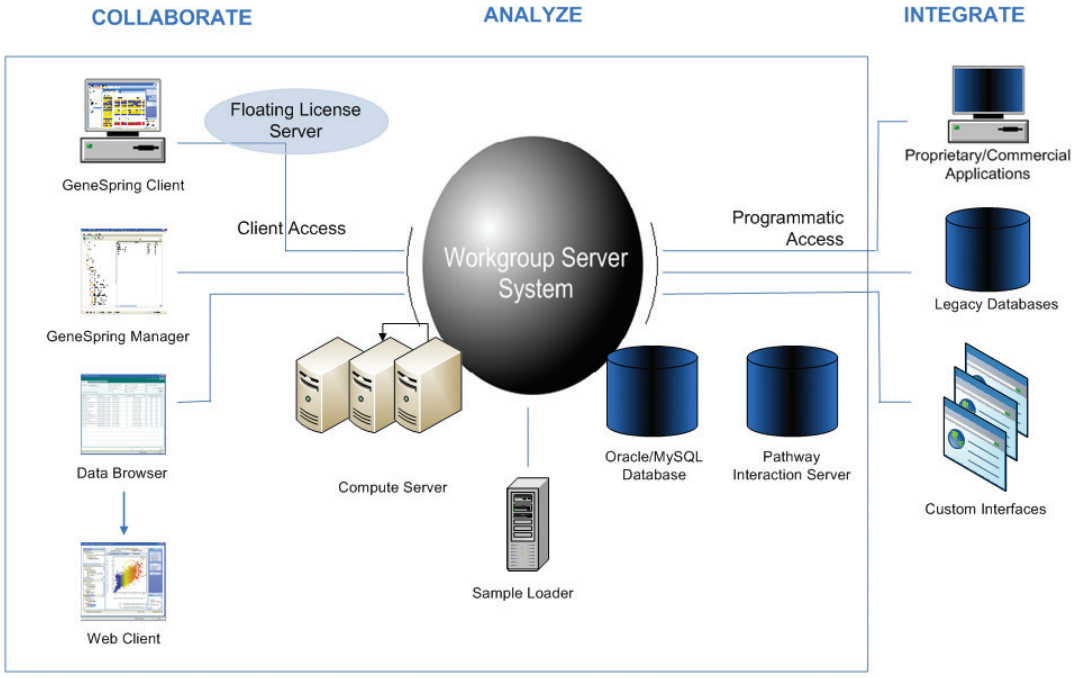

Figure 7. *GeneSpring Workgroup provides enterprise-level central storage and analysis capabilities for multi-omic data.*

Intuitive graphical displays

GeneSpring 11.5 displays data in ways that help users conceptualize and interpret the information in their data. The various types of plots, graphs, and diagrams highlight different aspects, allowing visual information to be extracted in multiple 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. An interactive and powerful Genome Browser is a key component to visualizing genomic structural data and integrating results from multi-omic studies. Results from different experiments can be dragged and dropped into individual data tracks and viewed simultaneously. The image overlay feature permits the user to overlay any data or annotation tracks, thus allowing researchers to qualitatively assess correlation between different data types. Other visualization tools in GeneSpring 11.5 include scatter plot, MvA, profile plot, histogram, and many more.

Integrated toolbox for pathway analysis and biological contextualization

Genes and proteins interact in a biochemical network to orchestrate the biological processes involved in disease. GeneSpring 11.5 provides capabilities to allow researchers to quickly generate and dynamically explore these networks. Instead of looking at results at the individual gene level, scientists can explore what biological process, molecular function, and biological pathways are involved in a disease process using provided GO analysis, GSEA, GSA and pathway and network analysis tools. Using a set of algorithms and provided organism-specific interaction databases, researchers can build a range of network types, including targets and regulators, transcription regulators, biological processes, and shortest connect. Networks can also be created based on user-specified MeSH terms. Networks generated in GeneSpring 11.5 are dynamic and interactive, allowing overlay of expression values and analysis results onto the network, interrogation of nodes and edges, and expansion of network from selected nodes (Figure 6). Natural language processing-based (NLP) algorithms can be applied to a body of text, HTML, a PDF, or Medline XML to extract and add interactions to an existing interaction database. GeneSpring 11.5 facilitates the process of extracting relevant interactions by allowing scientists to use MeSH terms to retrieve Medline abstracts that the NLP will run on. This allows researchers to customize network analysis to their experimental models. In addition to network building, GeneSpring 11.5 allows researchers to import and view any pathways in the BioPAX exchange format. These pathways can be used in the Find Similar Pathway tool, where researchers could quickly determine if there is an enrichment of the genes of interest in any pathways. In addition to providing built-in pathway and network analysis tools, GeneSpring 11.5 extends biological contextualization capabilities. Through its integration with Ingenuity Pathway Analysis (IPA) and GeneGo's MetaCore pathway analysis, data can be seamlessly transferred between the two applications for iterative analysis.

Scalable architecture

In addition to a desktop computing environment, the GeneSpring platform offers GeneSpring Workgroup, a scalable client-server product that serves the need for higher computation and throughput (Figure 7). GeneSpring Workgroup consists of a central server for secure content management of multi-omic data that is connected to various components and clients. The Data Browser and Web Client interfaces allow scientists to search, visualize, and share analysis results. Automatic batch sample loading and experiment creation provides efficient means of importing data into the Workgroup server for scientists to access and analyze. Computationally intensive analyses can be offloaded to a remote server farm. Web services of the Workgroup server provide secure programmatic access to any data in the system.

GeneSpring GX Products

Description	Academic Part Number	Commercial Part Number
GeneSpring GX Standalone 1 year license	G3784AA	G3778AA
GeneSpring GX Concurrent 1 year license	G3783AA	G3777AA
GeneSpring GX Standalone 2 year license	G3782AA	G3776AA
GeneSpirng GX Concurrent 2 year license	G3781AA	G3775AA
GeneSpring GX Standalone 3 year license	G3780AA	G3774AA
GeneSpring GX Concurrent 3 year license	G3779AA	G3773AA
GeneSpring Workgroup Server 1 year license	G1754AA	G1771AA

www.genespring.com

Published in USA, March 25, 2011
Publication Number 5990-7505EN

Agilent Technologies