INFORMATICS GENOMICS PROTEOMICS METABOLOMICS

0 1 1 0 1 0 0 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 1 1 0 1 1 0 1 0 0 0 1
0 0 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 1 0 1 1 0 1 0 1 1 1 0 1 1 0 0 1 1

GeneSpring MS from Agilent

Comprehensive Software Solution for Protein and Metabolite Biomarker Discovery

Mass spectrometry in high-throughput protein and metabolite profiling has led to an exponential accumulation of biological data. A key goal is to identify biomarkers—molecular species that correlate to a particular phenotype and may play an important role in the biological process or condition under study. GeneSpring MS was developed to meet the challenge of discovering biomarkers and distilling other biologically interesting and relevant information from complex, multi-dimensional GC/MS and LC/MS data. It allows researchers to easily import, analyze and visualize GC/MS and LC/MS data from large sample sets and complex experimental designs. Using a comprehensive array of powerful statistical analyses, GeneSpring MS can profile proteins or small molecules that are associated with changes in cellular function, thereby enabling the rapid discovery of biomarkers that can detect disease or drug toxicity.

GeneSpring MS is part of the GeneSpring Analysis Platform, an expanding suite of integrated software applications designed to enable systems-level investigation of biological research questions. Based on the widely popular GeneSpring GX expression analysis software, GeneSpring MS software offers a similar straightforward interface, knowledgeable worldwide support, and powerful statistical tools.

Intuitive and informative display of data

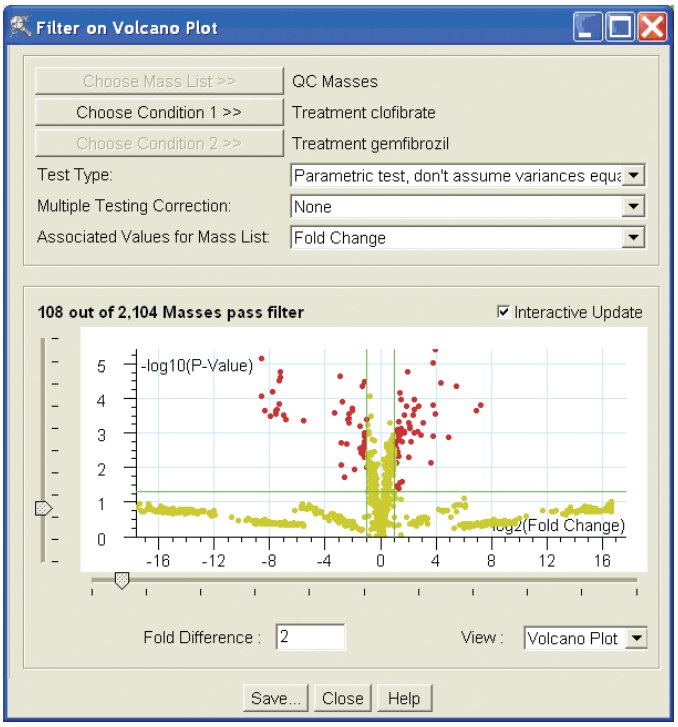
GeneSpring MS contains a variety of graphical displays to clearly and intuitively visualize data. These displays are useful for assessing data quality and finding mass entities of interest. The Spectrum Viewer allows access to the acquired total ion chromatogram (TIC) and associated mass spectra for each LC/MS sample. The Mass Details window provides access to LC/MS raw data information, such as the reconstructed mass spectrum and the associated abundance values of each ion species for the mass entity of interest. The Volcano Plot view facilitates rapid identification of mass entities with a statistically significant fold-change between two experimental conditions. Other types of graphical displays in GeneSpring MS include:

- 2D and 3D Scatter plots
- Mass vs Ratio of Two Conditions Plot
- 2D dendrograms
- RT vs Mass Plot
- Compound Viewer

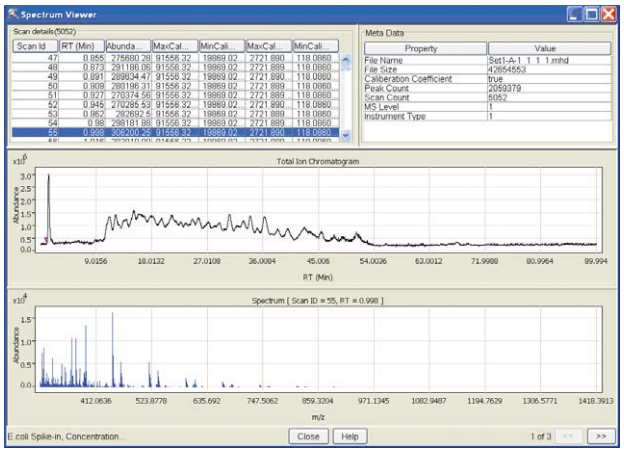
Detecting differential abundance of mass entities

Isolating mass entities with differential abundance between a set of experimental conditions is a key step in identifying biomarkers, proteins or metabolites that mediate the underlying mechanism of a biological process, and new potential therapeutic drug targets. To enable identification of these mass entities of interest in a statistically rigorous fashion, GeneSpring MS provides a broad set of tests that include:

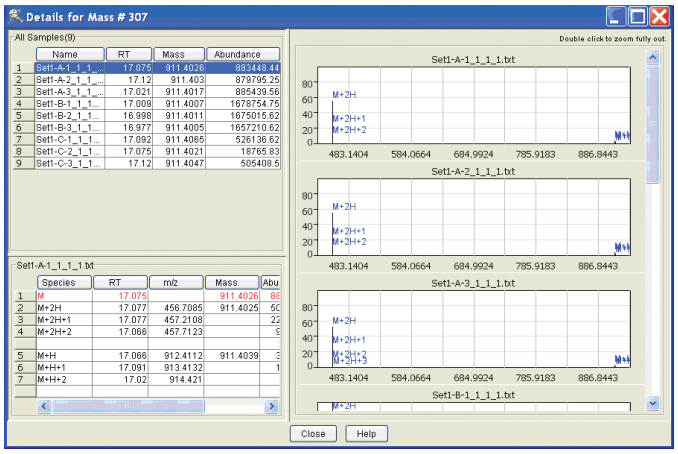
- Parametric and non-parametric one-way tests
- Parametric and non-parametric two-way ANOVA
- Family-wise error rate and false discovery rate multiple testing corrections
- Tukey and Student-Newman-Keuls post-hoc tests



The Volcano Plot simultaneously calculates the fold-change in abundance for each mass entity between two experimental conditions and determines the statistical significance of that change.



The Spectrum Viewer displays the TIC and spectra associated with each LC/MS sample within the experiment.



The Mass Details view displays the reconstructed mass spectrum for a selected mass entity and reports the abundance values for each ion species.

Grouping mass entities with similar abundance profiles

Clustering analysis groups mass entities based on the similarity of their abundance profiles to uncover the most prominent patterns in the data. In addition, clustering analysis may reveal interesting biological relationships, as mass entities that exhibit similar behavior across a set of experimental conditions may share similar biological functions. GeneSpring MS provides a broad choice of clustering algorithms that include:

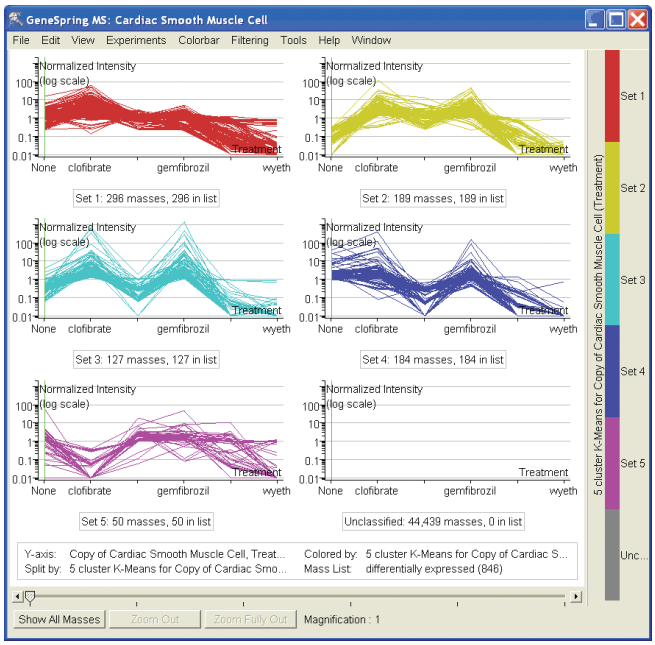
- Hierarchical clustering
- k-means clustering
- Quality Threshold (QT) clustering
- Self-organizing maps (SOM)

Comparing sample abundance profiles

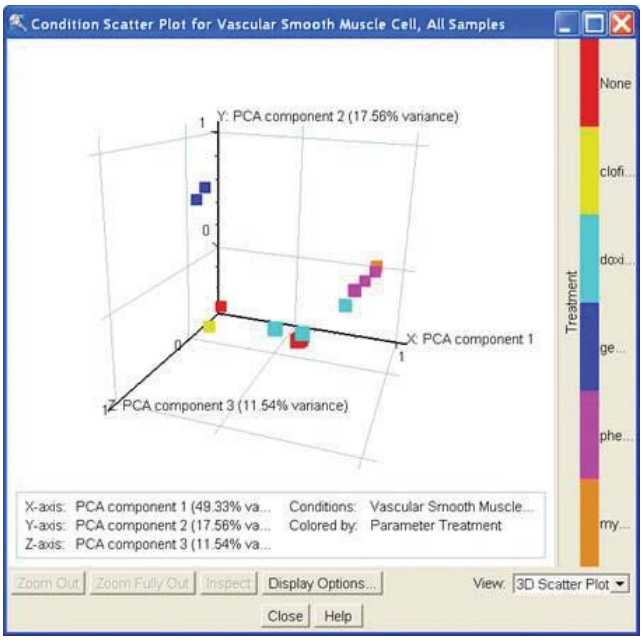
Comparing sample abundance profiles can provide information on the quality of replicate samples within an experimental condition and give insight into similarities in biological responses. For example, if a compound is known to induce a mass abundance signature associated with liver toxicity, compounds that induce a similar profile may also possess toxic effects. Principal Component Analysis (PCA) and hierarchical clustering are two powerful methods for comparing sample abundance in GeneSpring MS.

Supervised classification of abundance profiles

Class prediction analysis using metabolomic and proteomic data is becoming an increasingly valuable tool in toxicometabolomics, toxicoproteomics, and for early diagnosis, prognosis, and prediction of clinical treatment outcomes. Compound classification and predicting toxic effects of potential therapeutics is essential to the process of prioritizing compound pipelines and eliminating costly failures in drug development. In class prediction analysis, altered metabolite and protein expression patterns reflecting biological responses to well characterized toxicants can be used to predict toxicological classification of unknown compounds. For such supervised classification analyses, GeneSpring MS provides two class prediction algorithms: k-nearest neighbors and Support Vector Machines (SVM).



K-means clustering results are presented in a graph view to clearly illustrate the different mass abundance patterns in the data. Mass entities in each cluster can be isolated into separate mass lists, allowing a closer interrogation of mass entities with similar abundance profiles.

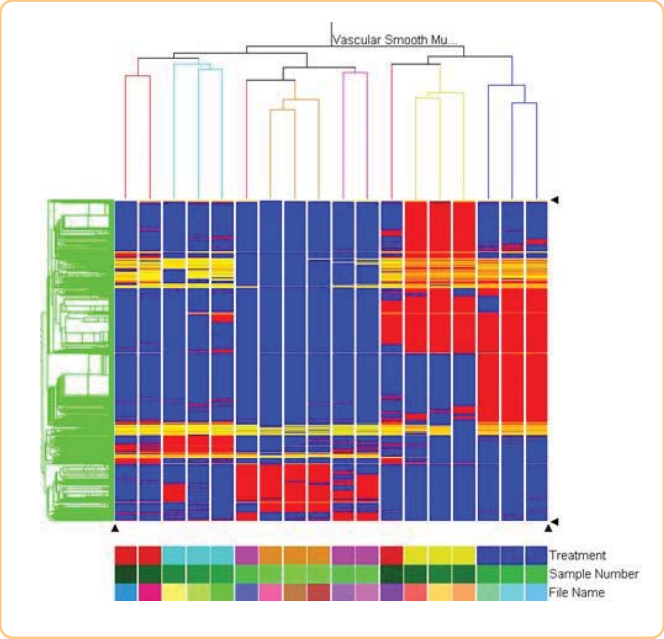


Results from PCA on Condition delineate the similarity in abundance profiles between samples. The 3D scatter plot can be rotated and samples can be colored according to any experimental parameters, thus facilitating the comparison of biological responses to each experimental condition.

Protein and metabolite identification of mass entities of interest

Frequently, the identity of mass entities of interest from mass spectrometry data is unknown. A key step in placing statistically significant findings into a biological context is to identify the proteins or metabolites represented by these mass entities of interest. GeneSpring MS provides several tools to address common methodologies for protein or metabolite identification, protein and metabolite database searching, and tandem mass spectrometry analysis. Specifically, these tools include:

- A direct link to perform metabolite identification through Metlin Metabolite Database search.
- A direct link to perform protein identification using Protein Mass Fingerprinting in Agilent Spectrum Mill database.
- Direct export of the MS/MS inclusion list containing mass entities to be targeted during MS/MS acquisition for peptide/protein identification.



Tree diagrams produced by hierarchical clustering in GeneSpring MS reveal the relationships between mass entities in one dimension and between samples in the other dimension.

About Agilent's Integrated Biology Solutions

Agilent Technologies is a leading supplier of life science research systems that enable scientists to understand complex biological processes, determine disease mechanisms, and speed drug discovery. Engineered for sensitivity, reproducibility, and workflow productivity, Agilent's integrated biology solutions include instrumentation, microfluidics, software, microarrays, consumables, and services for genomics, proteomics, and metabolomics applications.

Find an Agilent customer center in your country:
www.agilent.com/chem/contactus

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Asia Pacific
adinquiry_aplsca@agilent.com

Europe
info_agilent@agilent.com

Printed in the U.S.A. August 24, 2006

5989-5553EN