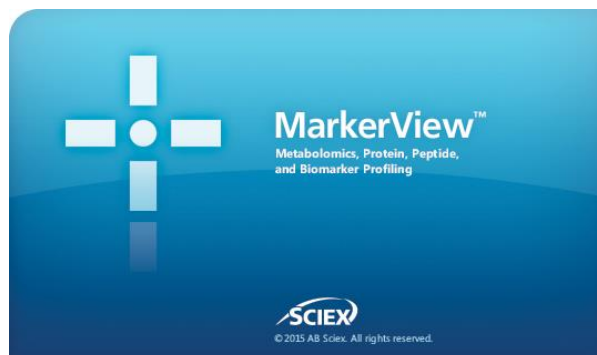


## MarkerView™ Software 1.3.1 for Statistical Analysis of MS Datasets

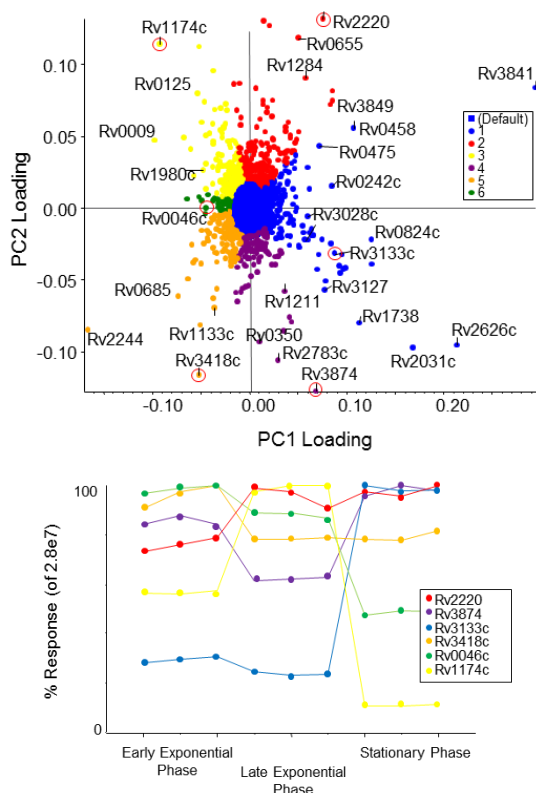
Today's mass spectrometers can readily generate large amounts of quantitative data, moving the bottleneck to processing the data efficiently, making sense of the complexity, and generating trustworthy results. Therefore, proper statistical analysis is increasingly important. MarkerView™ software is powerful, easy to use software designed for the statistical analysis of mass spectrometry data, from Omics experiments to food authenticity, water testing experiments and many others. Using sophisticated statistical analyses and graphical tools, MarkerView software allows rapid review of data acquired on all SCIEX mass spectrometers to determine what components are different or changing between complex samples. With both multivariate and univariate statistical capabilities including principal component analysis (PCA), principal component



variable grouping (PCVG), and t-tests, MarkerView software enable features of interest can be easily discovered. Normalization algorithms, peak detection algorithms and extensive reporting capabilities round out the functionality to make this software extremely versatile for downstream analysis.

### Key Features of MarkerView™ Software

- Powerful tools to perform statistical analysis to accelerate the discovery of components responsible for interesting changes between complex samples
- Visualize data using supervised and/or unsupervised techniques such as PCA-DA (discriminant analysis) or PCA as well as t-tests to accelerate the discovery of new biomarkers
- Principal Component Variable Grouping (PCVG) to assist data interpretation by locating and highlighting correlated variables
- Intuitive interactive plots can be easily generated to assist in interpretation of complex datasets
- Broad flexibility to analyze many different types of datasets using a broad range of statistical tools
  - Import multiple types of raw data acquired on any SCIEX mass spectrometer, including both ESI- and MALDI-based platforms
  - Import processed data from SCIEX software such as SWATH® Software, MasterView™ Software, SCIEX OS and others
- Automatically create reports on potential biomarkers and export data to third-party statistical packages for additional data mining



**Figure 1. MarkerView™ Software Visualization Plots.** MarkerView software uses PCA to reveal similarities and/or differences in a set of samples. Here, the Loadings plot (top) indicates which variables are responsible for the observed separation and PCVG has been used to determine correlated variables (colored together). The bottom plot shows select variables from each PCVG group, in this case protein expression differences<sup>1</sup>.

## Processing of Raw Data

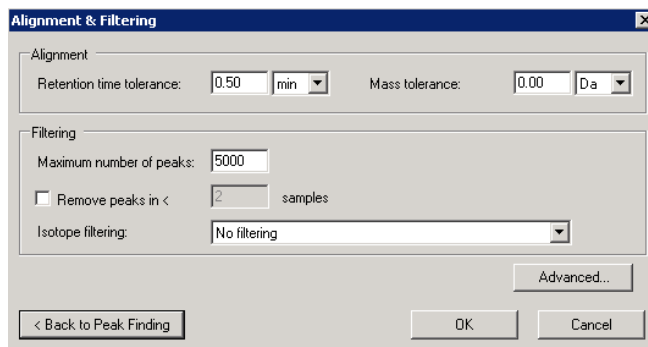
MarkerView™ software uses sophisticated processing algorithms that accurately find chromatographic and spectral peaks direct from raw data in complex data sets. Data alignment by the software compensates for minor variations in both mass and retention time values, ensuring that identical compounds in different samples are accurately compared to one another (Figure 2). Subsequent normalization of the data accounts for sample to sample variation, such as sample size or injection volume, and provides more precise results. MarkerView software can process data acquired by LC-MS, MALDI-MS, flow injection, and/or direct infusion from all SCIEX mass spectrometers. After trends are found, it can link back to raw mass spectra and extracted ion chromatograms to support identification of features.

## Direct Import of Processed Data

Processed data can also be exported and imported directly into MarkerView software for statistical analysis which significantly streamlines the processing pipeline. Here the primary data processing software enables the identification and quantitation of the known analytes. Data analysis can be then undertaken in MarkerView software.

Software packages that can export data to MarkerView Software:

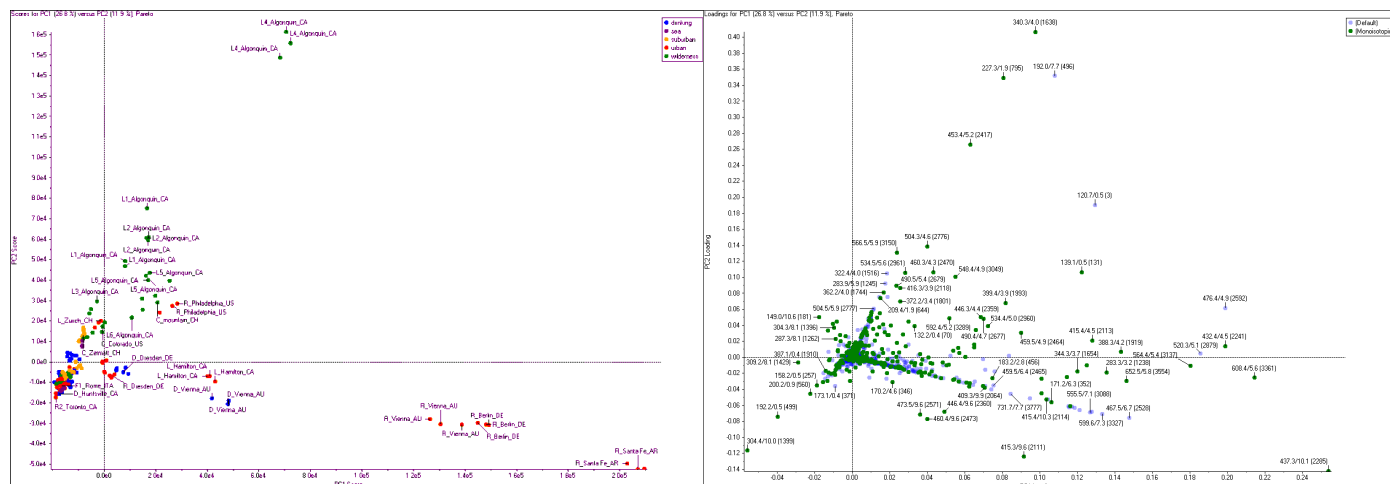
- SWATH® Acquisition MicroApp in PeakView® Software
- LipidView™ Software
- MasterView™ Software
- SCIEX OS



**Figure 2. Feature Alignment.** The Alignment & Filtering dialog box compensates for minor variations in mass and retention time, automatically ensuring that identical compounds in different samples are accurately compared. Retention time correction can be based on standards if available.

## Principal Component Analysis

MarkerView software allows visualization of many types of data with principal component analysis (PCA). PCA is an unsupervised multivariate statistical analysis approach that allows trends to be recognized across groups of samples within a dataset. This is graphically represented in a Scores plot (Figure 3, left). The scores plot highlights the largest variation in the dataset in principal component 1 (PC1, x-axis) and the second largest variation in the dataset as PC2 (y-axis). Reviewing the Loadings plot provides insight into variables that lead to any sample trend in the scores plot (Figure 3, right). The Loadings plot would then illustrate which compounds are being up- or down-regulated. Discriminant analysis (PCA-DA) can also be employed, a supervised version of PCA wherein prior knowledge of sample groups is used to determine the variables that maximize the variation between groups and those which minimize the variation within a group.



**Figure 3. Principal Component Analysis.** Powerful statistical tools such as Principal Component Analysis (PCA) enable grouping of samples based on common features. The Scores plot (left) shows groups and differences among the samples, for example the quality of technical replicates can quickly be ascertained as well as the categorization of samples based on differentially observed components. The Loadings Plot (right) reflects the variables that are causing the separation and those with higher loadings values are generally more significant for the separation.

Additionally, MarkerView Software offers the flexibility to use different scaling algorithms, such as Mean Centering, Autoscaling, and Pareto scaling, providing more flexibility in data analysis.

## Principal Component Variable Grouping

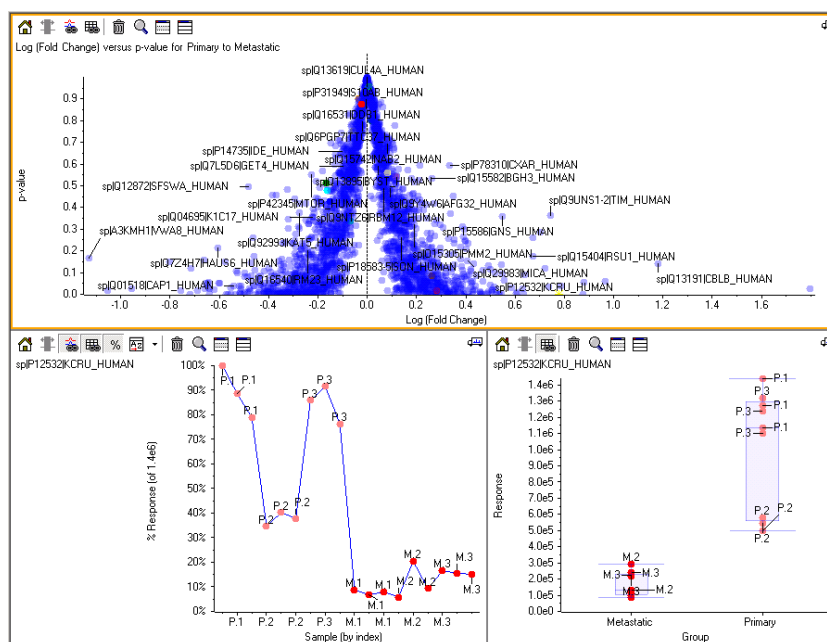
Principal Component Variable Grouping (PCVG) is a tool that analyzes the PCA loadings values to find correlated variables, i.e. those that share a common expression pattern across the samples<sup>2</sup>. Essentially it uses the samples to group the variables, which simplifies interpretation since there are usually far fewer groups than variables and the behavior of any one represents the rest of the group.

MarkerView™ Software user interface allows all members of the group to be selected together for display and also allows them to be removed if the behavior is not relevant (contamination for example). The top-right graph in Figure 1 shows an example of PCVG. As shown, the variables are color-coded with respect to the sets of variables that are responsible for separation of the samples into groups. At the bottom, Group 3 and Group 6 variables are displayed in Profile Plots facilitating the identification of those variables responsible for distinguishing the groups.

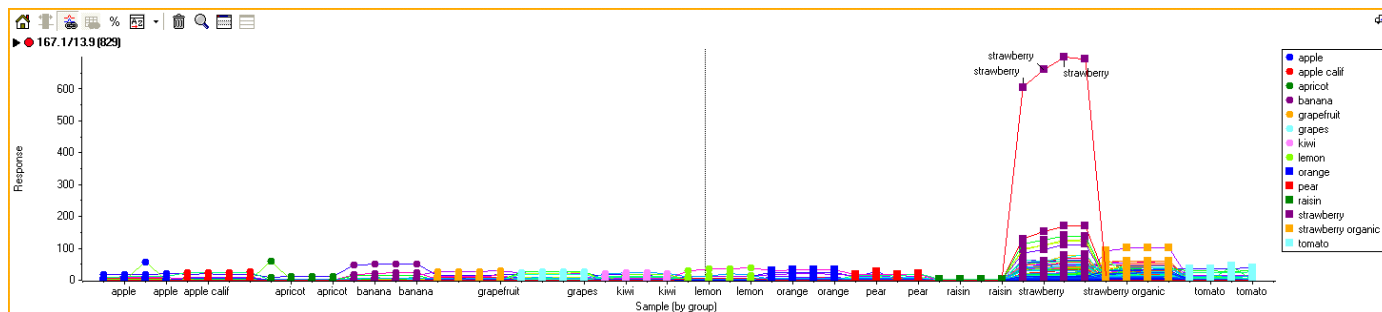
## T-Test

The t-test is a supervised analysis technique and is useful when two or more predetermined classes of samples are present. MarkerView Software will perform a pair-wise comparison of all the classes or alternatively, compare one class to all others. The results of the t-test indicate how well each variable distinguishes the two groups. This is reported as a p-value, i.e. the probability that the observed difference occurs by chance.

The profiles for individual variables can be generated or the overall behavior can be summarized in displays such as the “volcano plot” (Figure 5). For any two groups, the volcano plot shows for each variable the log of the fold change (the ratio of the average response in each group) as a function of the p-value (the probability that the observed difference occurs by chance). Especially significant variables are those that have a low p-value (small probability of occurring by chance) and a large fold change. At the extremes of the horizontal axis are groups of variables that are completely absent in one sample group, that is, they have an infinite fold change. The changes for significant variables can be easily visualized, with response plots for the variable across the different samples (Figure 5, bottom left) and box and whisker plots (Figure 5, bottom right).



**Figure 5. T-Tests.** The volcano plot of p-value vs. log (fold change) combines the fold change in the response of a variable between groups with the probability (p-value) that the variable distinguishes the groups to quickly highlight important variables (top). In this proteomics example using SWATH® Acquisition, significant differential proteins between two biological groups were found. Example of a differential protein is shown (bottom) with the response profile across the 18 samples and the computed Box and Whisker plot.



**Figure 6. Simple Data Visualization.** Additional visualization graphics and tools aid in the detailed investigation of trends across samples. By plotting the abundance of one or more variables across multiple samples, features that behave similarly can be quickly spotted as in this example, where selected variables distinguish strawberries and organic strawberries from other fruit samples.

## Data Normalization

For more precise quantification, sometimes data normalization is needed to account for systematic variance across a dataset. There are a number of different ways to perform normalization and the right choice depends on the experiment and data type. In MarkerView Software, there are numerous normalization strategies that can be applied.

- Using Internal Standards or Selected Peaks – specific features are used to normalize across the dataset
- Using Total Area Sums – sum of all peak areas used to compute the scaling factor
- Using Most Likely Ratio (MLR)<sup>3</sup> – used when a large number of features are variable, ratio between features used to compute a ratio histogram which is aligned across samples
- Using Manual Scale Factors – using scale factors computed by other means

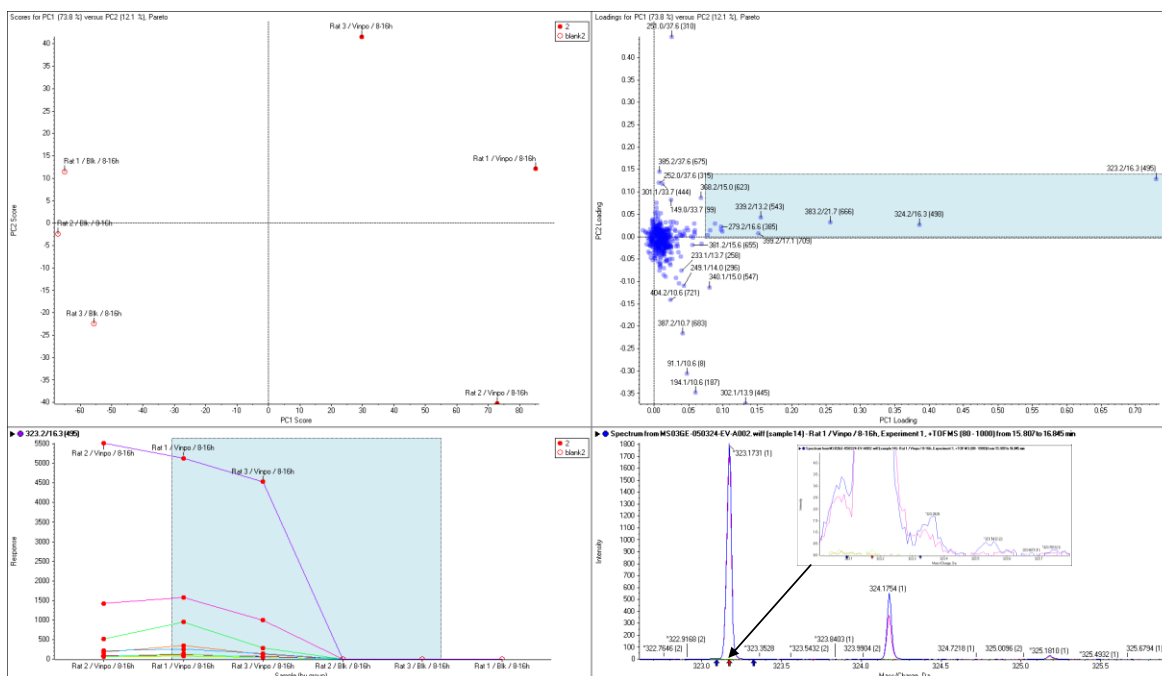
## Other Data Visualization Plots

When analyzing complex data sets, tools that allow visualization and manipulation of the data to discover true patterns and correlations within the data are critical. In MarkerView™ software, not only are the traditional Scores and Loadings plots generated from PCA valuable, but direct links to raw mass spectra and chromatographic data provides a simple way to go through multiple data files simultaneously to confirm findings (Figure 7).

Profile plots can be easily generated to show the intensity of selected compounds across multiple samples to assist in finding features of interest (Figure 6). This view can also highlight expression patterns resulting from systematic errors or contamination that can then be excluded.

## Create Reports

During data review, MarkerView™ software allows you to create a list of compounds of interest. This list can be annotated and is used to automatically generate a report in Microsoft Word containing user specified information. The reporting tools can easily be customized in terms of both the Word template and layout. Information that can be reported includes the Scores and Loadings plots, Profile plots, Mass Spectra as well as Extracted Ion Chromatograms for all compounds of interest such as suspected biomarkers.



**Figure 7. Automatic Links to Raw Data.** The data shown here were obtained by analyzing the urine from three rats before (open circles, top left) and after (closed circles) administration of a single dose of vinpocetin. The upper two panels show the Scores (left) and Loadings (right) plots. The variables within the blue rectangle were selected for display and generated the profile plots shown in the lower left panel which shows the response for the six samples. Four data points were selected and the raw spectra displayed (bottom right). It is clear that the most intense variable ( $m/z$  323.2 at 16.3 min) is absent from the pre-dose samples (orange and green traces).

## Summary

Easily go from raw data to results - MarkerView software provides the essential tools to aid in statistical analysis and visualization of large numbers of samples and prevents data overload. Originally designed for metabolomics and biomarker profiling researchers, the powerful and easy-to-use features of MarkerView software help to accelerate your discovery of differentially expressed compounds and new biomarkers. MarkerView software also applies to the food or forensic drug chemist, helping to identify the minor component differences between samples to aid in determining the source of substances and/or authenticity.

## References

1. Enabling Systems Biology Driven Proteome Wide Quantitation of Mycobacterium Tuberculosis, SCIEX technical note RUO-MKT-02-4292-B.
2. Ivosev G, et. al. (2008) Dimensionality Reduction and Visualization in Principal Component Analysis. *Anal. Chem.* **80**, 4933-4944.
3. Lambert et al. (2013) Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nature Methods*, **10**, 1239-1245

AB Sciex is doing business as SCIEX.

© 2017 AB Sciex. For Research Use Only. Not for use in diagnostic procedures. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Document number: RUO-MKT-02-2775-B