Introducing DiscoverM

Statistical analysis and reporting for Metabolomics

www.chenomx.com

INTRODUCTION

DiscoverM is a software tool capable of analyzing metabolomics data using a simple pipeline of preprocessing, visualization and statistical analysis methods. Based on standard and optimized statistical techniques, **DiscoverM** can be used for the complete process of data analysis ranging from data imputation and extensive preprocessing, to interactive visualization, to flexible but comprehensive steps of data interpretation using various univariate and multivariate statistical analysis methods. **DiscoverM** allows the user to extract and save data and intermediate results at any step in the entire pipeline. **DiscoverM** is a user friendly tool which can be used to analyze metabolomics data.

INPUT DATA AND PREPROCESSING

Data files in CSV (comma separated value) format can be exported from Chenomx NMR Suite for use in **DiscoverM**. The concentration data file contains metabolite concentration, whereas the metadata file contains sample information such as gender, age, sample dilution details, dates of data collection, or anything else that can be used as a factor in experimental analysis.

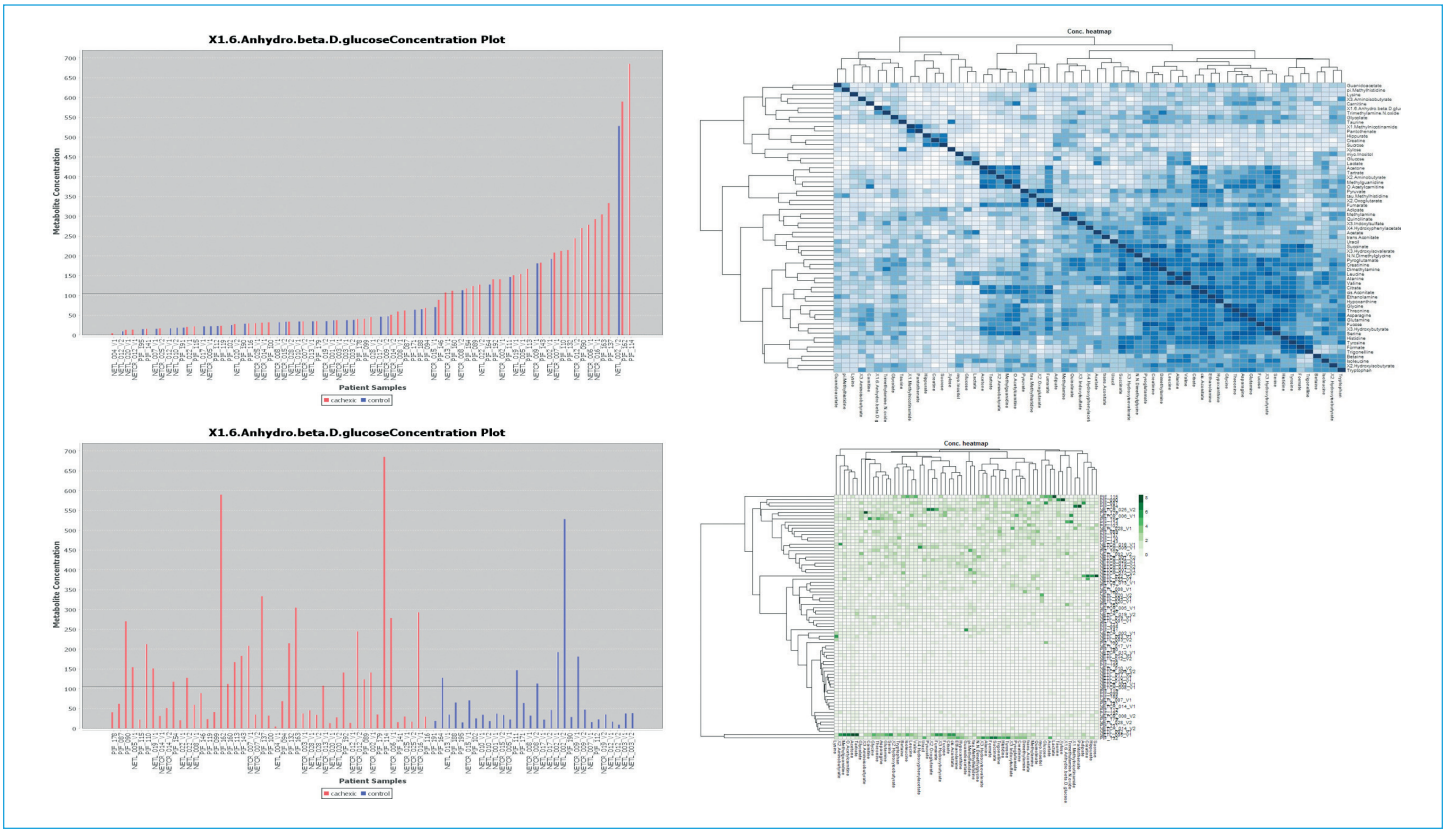
1	Profiled Data Type	Concentrations (uM)			
2	Export Date	Sat Apr 09 16:28:13 MDT 2016			
3		1,6-Anhydro-beta-D-glucose	1-MethylNicotinamide	2-Aminobutyrate	2-Hydroxyisobutyrate
4	compound id	999	999	999	999
5	PIF_178	40.85	65.37	18.73	26.05
6	PIF_087	62.18	340.36	24.29	41.68
7	PIF_090	270.43	64.72	12.18	65.37
8	NETL_005_V1	154.47	52.98	172.43	74.44
9	PIF_115	22.2	73.7	15.64	83.93
10	PIF_110	212.72	31.82	18.36	80.64

Profiled Data Type	Concentrations (uM)
Export Date	Sat Apr 09 16:28:13 MDT 2016
Sample ID	Class
PIF_178	control
PIF_087	cachexic
PIF_090	cachexic
NETL_005_V1	cachexic
PIF_115	cachexic
PIF_110	cachexic

A number of preprocessing steps are available within **DiscoverM**. These include: normalization by sum, median, pooled samples, specific sample and more; data transformation by cube root, or logarithmic; and data scaling (auto or Pareto).

DATA VISUALIZATION

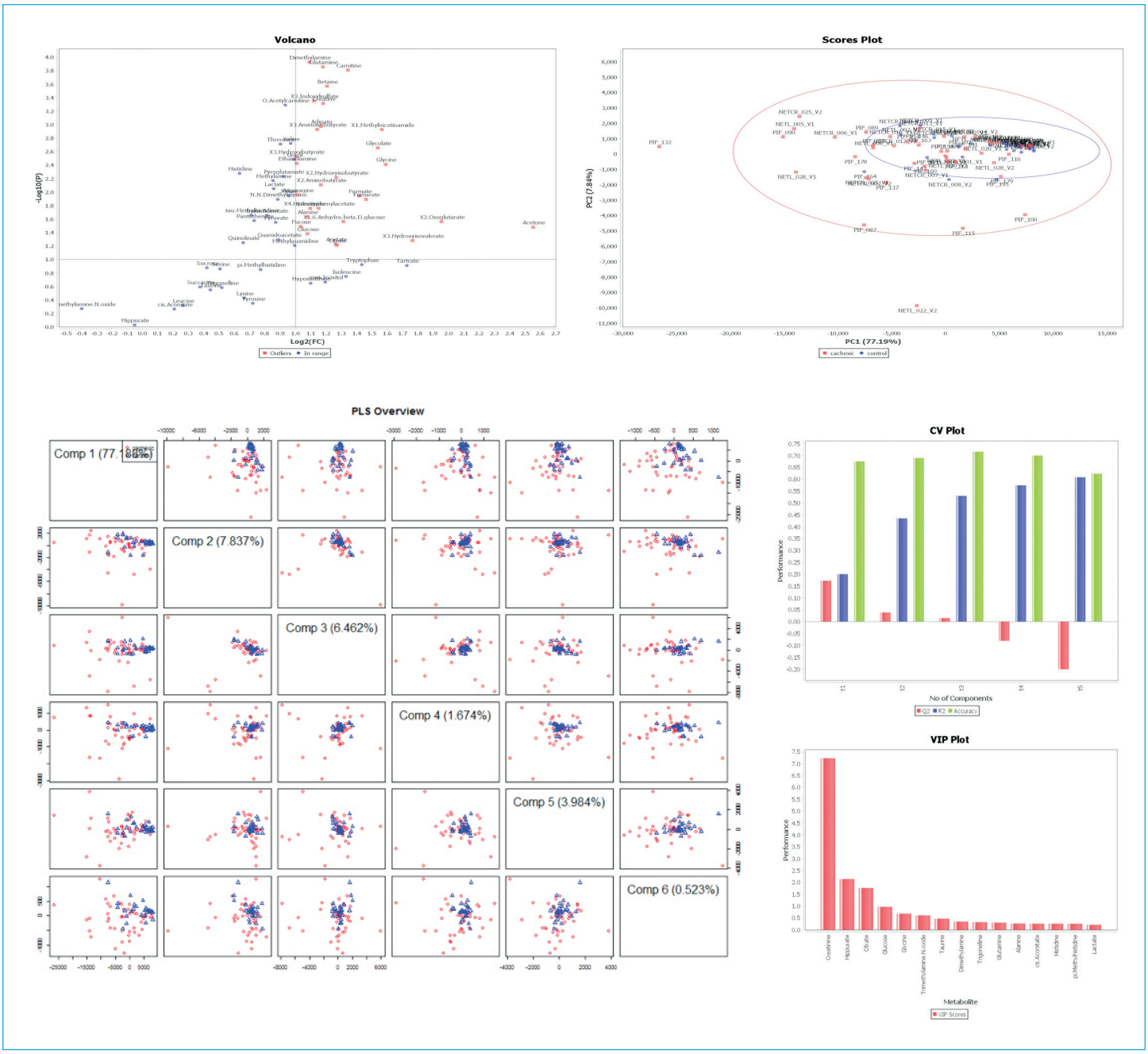
This reports the value of a particular metabolite within all the samples and the various metabolites within a particular sample. Additional features include sorting, heat maps, and dendrograms from clustering. Various distance calculation methods can be used: complete, average, single, and ward.



UNIVARIATE AND MULTIVARIATE ANALYSIS

The statistical analysis methods present in **DiscoverM** include t- test, Fold Change Analysis and Volcano Plot in univariate category and Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) in multivariate category.

In addition to the wide range of analytical methods, **DiscoverM** provides clear and ready to use graphical outcomes.



DiscoverM has been developed to meet the user demand for an interactive and easy-to-use tool for statistical metabolomic data analysis. **DiscoverM** is robust and well-adapted software for this type of analysis. With the provided preprocessing, univariate and multivariate analysis methods, **DiscoverM** can be used to draw useful biological/clinical inference from information-rich metabolomics data.

UNIVERSITY OF ALBERTA | CHENOMX

LICENSING AND USE: **DiscoverM** has been developed together with an acacemic collaboration and utilizes open source software (R) as part of the application. Chenomx does not charge a license fee for this software but does require the latest version of Chenomx NMR Suite (professional version) in order to use the data export features that support discoverM.

www.chenomx.com

Printed in Canada | May 2018