For Research Use Only. Not for Use In Diagnostic Procedures.

Biomarkers and Omics

SCIEX

Advanced Data Processing Software enables Autonomous Metabolite Identification on the X500R QTOF Platform
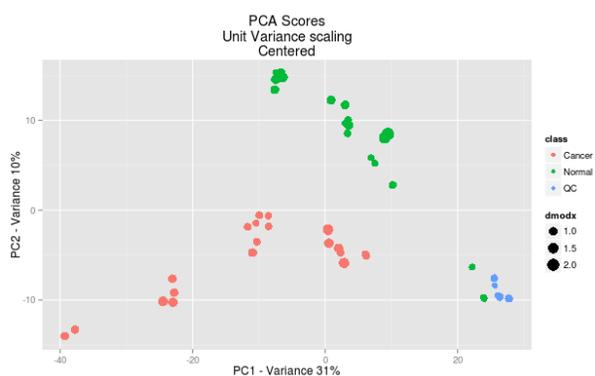
Using XCMS[plus] Software

Baljit K. Ubhi[1] and Cyrus Papan[2]
[1]SCIEX, CA, USA and [2]SCIEX, Germany

Metabolomics is the scientific study of the chemical fingerprint which is left behind by cellular processes[1]. Untargeted metabolomics studies are valuable to biomarker researchers for studying disease effects over time and to identify novel biomarkers to track disease progression. Mass Spectrometry tends to be the analytical tool of choice given its sensitivity to other techniques such as ^{1}H NMR Spectroscopy. The new X500R QTOF System was developed for routine, robust workflows and requires minimal MS expertise. Data from a previously acquired from a prostate cancer study was processed as proof of concept, through the XCMS[plus] software pipeline to highlight advanced data processing features required of untargeted metabolomics studies. XCMS software is the "World's most cited metabolomics software" as it has over 1000+ citations in literature and very trusted in the metabolomics community The software allows you to load, process and analyze their data in one interactive workspace. Peaks are annotated and matched against the METLIN database for metabolite identification.


Figure 1: Metabolomic Profiles can Clearly Differentiate Diseased (Cancer) from Normal Samples. Using the untargeted metabolomics strategy described here, samples from a prostate cancer study were analyzed, as well as QC samples (spiked with known compounds). Shown in the above Scores plot, PCA analysis clearly differentiates healthy from disease and samples, confirming the original sample classification. Samples within the same group cluster together highlighting the reproducibility of the technique. The related loadings plot is used to determine which peaks are responsible for this differentiation.

Key Benefits of the Untargeted Metabolomics Workflow on the X500R QTOF System with XCMSPlus Software Data Processing

- Simplified operation with new interface and workflow setup powered by SCIEX OS Software with minimal MS expertise.

- Get to actionable results quickly with user-friendly processing software and libraries

- Maximize your lab's efficiency using a robust and integrated system with automated calibration and a benchtop footprint

- Data can be loaded, processed and reviewed in a single interactive workspace

- Submit and monitor the status of multiple jobs simultaneously

- Statistical analysis including paired/unpaired t-tests, parametric and non-parametric testing (including FDR), and multivariate data analysis techniques.

- Simplified metabolite identification by linking your data to the METLIN database

SCIEX

Methods

Sample Preparation: Urine samples were obtained with disease classifications that were previously determined using accepted clinical techniques. The specific gravity of the samples was measured by testing 15 µL of urine samples with a refractometer prism. A 50 µL volume of the thawed urine samples was then transferred to a clean, labeled microcentrfuge-filter tube. An isotopically labeled internal standard mixture (20 µL) was added. The urine sample was then diluted with 400 µL 98:2 acetonitrile/water with 0.1% sodium azide, then vortexed. The sample was then centrifuged, and the supernatant was isolated. This was then dried down and reconstituted in 50 µL of 0.1% formic acid in water. The samples were transferred to glass vial and loaded into the autosampler.

Chromatography: The reverse phase HPLC separation was performed using a Shimadzu LC System, operating at a flow rate of 350 µL/min. The column used was an Ace Excel C18-PFP column (100 x 1mm, 2 µm) from ACE, maintained at 30 ºC. A standard reverse phase gradient was used employing mobile phase A as 0.1% formic acid in water and mobile phase B as acetonitrile. The injection volume was 3 µL in positive ion mode and 5 µL in negative ion mode.

Mass Spectrometry: The data was collected using information dependent acquisition (IDA) on the X500R QTOF System (SCIEX). Using optimized source conditions, the MS mass range analyzed was 50-600 m/z and the MS/MS was acquired with a mass range of 40-600 m/z with a 25 msec accumulation time.
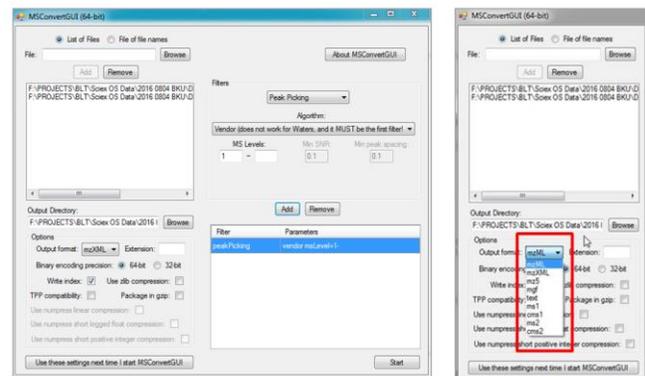
Figure 2: Data Conversion Parameters. Using Proteowizard wiff2 files can be converted using the above parameters to mzxml which can be loaded into XCMS[plus].

The collision energy was set to 35 V with a 15 V collision energy spread to ensure high quality MS/MS on most metabolites.

Data Processing: The data was processed in XCMS[plus] software, a desktop version of XCMS Online. The wiff2 data were converted using the msconvert tool of Proteowizard (3.0.9992) to mzXML format and pushed through XCMS[plus] to pick peaks, align features, normalize data and subsequently aid metabolite identification of the significantly extracted features.
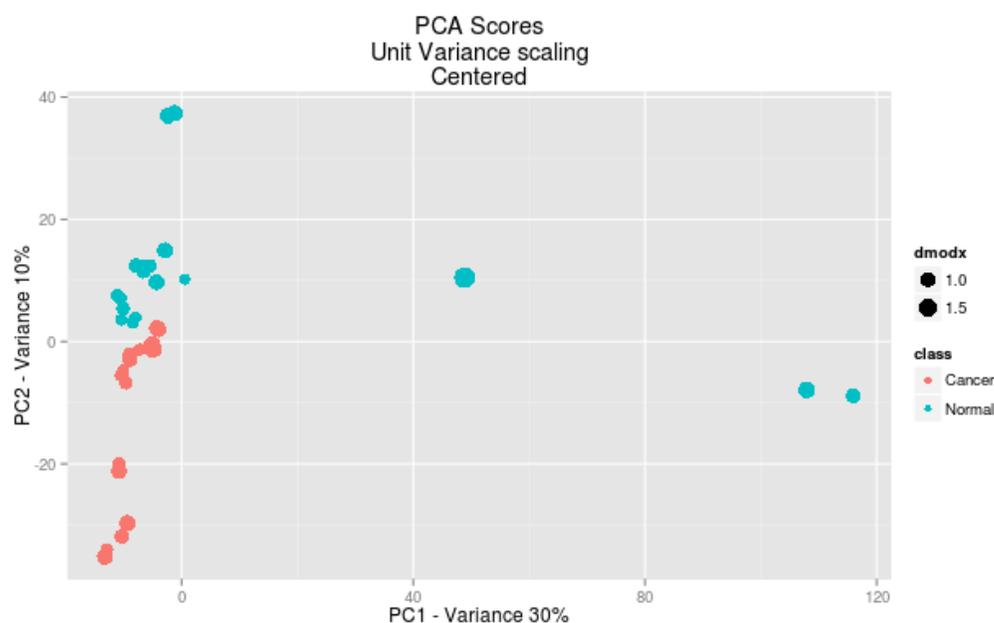
Figure 3: A Scores Plot from XCMS[plus] Software. The QC samples were removed from the same dataset as in Figure 1, leaving only a comparison of the healthy and diseased samples. There is a differentiation between the known sample groups highlighting that the metabolic profile is quite different in the diseased samples.

Results

From the scores plot (Figure 3) one can see a difference along the y-axis between the healthy and diseased classified samples. From here the user can "view results table" and view a table of extracted features. The table can be sorted based on p-value significance or q-value significance (Figure 4). The q-value is the false discovery rate adjusted p-value (not featured in this figure). The feature table identifies the retention time at which a particular feature was found as well as the intensity. The fold changes between any sample groups of that particular feature are also listed in the column labeled "fold" (Figure 4). Other observations (as columns) of interest can be added by the user and used for sorting (such as m/z value or retention time (RT), Corr Var = correlation variation, Max Int = maximum intensity, Feature Gp = feature group). Any isotopes and adducts are also identified as well as feature grouping which groups together any related ions.

From the feature table, a feature can be highlighted and the XIC can be visualized as an overlay from all the samples processed (Figure 4 top right). One can then review the raw data and see if this is an actual feature or just noise or an irrelevant peak eluting in the solvent front. Retention time information as well as the accurate mass is displayed in this XIC plot. The mass spectrum for that XIC can be viewed as well as a simple box and whisker plot highlighting the differences in intensities between the groups (including the average and upper and lower limits of the intensity per group) (Figure 4, middle right). Finally each feature is matched for identification to the METLIN database. If MS/MS spectra are available then the column labelled "MS/MS" is populated with "y" meaning MS/MS is available (not shown

Figure 4: Table of Extracted Features from XCMSplus Software. Showing all features extracted from this dataset (a total of 2532 features). These can be sorted by p value, fold change or any other parameter in the feature table. The highlighted feature (#1 in the table) then allows the user to view that particular extracted feature from the total ion chromatogram (TIC, top right). Here one can see the huge concentrations of this feature in the cancer group vs the normal group (where there is virtually no signal). The related box and whisker plot can be visualized (middle, right). Lastly any matches to the METLIN database can be highlighted in the table (bottom, right). Identifications are ordered from highest match to the accurate mass of the MS1 parent ion. One can click directly on the composite ID number to be directed to the METLIN database webpage regarding that particular compound (Figure 5).

here). The database identifications are listed in a table (Figure 4 bottom, right)) and ranked by m/z error (ppm) and then alpha/numerically. Then using composite ID column one can click the METLIN ID and be referred to the METLIN website (https://metlin.scripps.edu) where any MS/MS confirmation could be compared to the experimental spectra (Figure 5). A range of MSMS spectra are available at varying collision energies.

The most compelling visualization tool in XCMS^{plus} software is the interactive cloud plot (Figure 6). Key visualization features include:

- P-value is represented by how dark or light the color is.
- Fold change is represented by the radius of each feature.
- Retention time is represented by position on the x-axis.
- Mass-to-Charge ratio is represented by position on y-axis.
- Sliders for p-value and fold change are in the Main Panel.
- Sliders for intensity, retention time, and mass-to-charge are in the Advanced Tab. A link to the table representation of the graph is displayed at the bottom along with the settings used to generate the graph.

The cloud plot is completely interactive and can be filtered on any function listed above. So you can be more stringent for the p-value and also the mean fold change allowing for only the more highly significant features to be displayed.
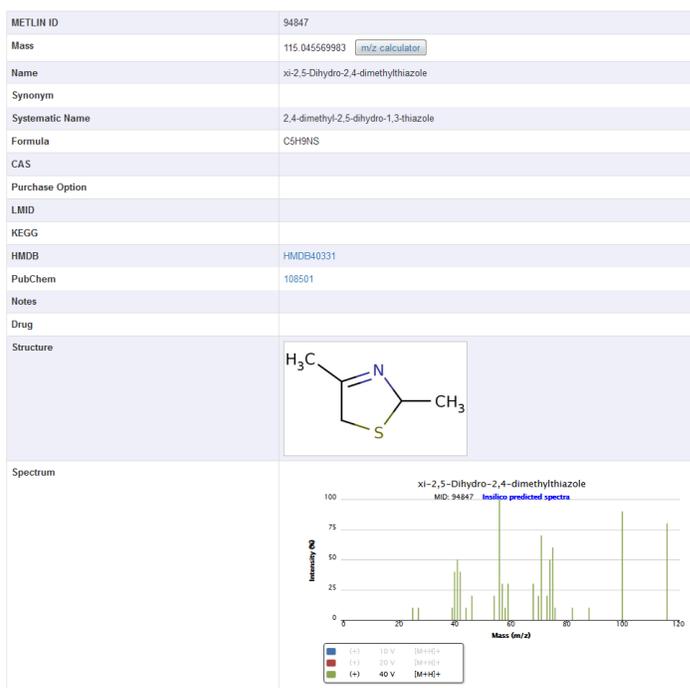
Conclusions

The X500R QTOF system is a robust, easy to operate, benchtop system that requires minimal MS expertise to perform untargeted metabolomics analyses. As metabolomics continues to expand in disease research, robust easy-to-use solutions that provide quality answers will be increasingly important.

In this study, samples from a pilot prostate cancer study was analyzed and a clear difference between healthy and disease urine samples were detected using this untargeted metabolomics approach, confirming the original disease classifications.

XCMS^{Plus} Software was used to extract the features from the dataset and simultaneously identify them against a large well-known database. All of the data review and any statistical data analysis could be completed within a single software package.

The software allows advanced data processing and data analysis which are often required by researchers for untargeted metabolomics studies.

This pilot study provided confidence in the approach and the next phase of the study analyzing a much larger set of samples is underway.

METLIN ID	94847
Mass	115.045569983 m/z calculator
Name	xi-2,5-Dihydro-2,4-dimethylthiazole
Synonym	
Systematic Name	2,4-dimethyl-2,5-dihydro-1,3-thiazole
Formula	C5H9NS
CAS	
Purchase Option	
LMID	
KEGG	
HMDB	HMDB40331
PubChem	108501
Notes	
Drug	
Structure	
Spectrum	

Figure 5: METLIN Database Entry. Highlighting information relevant to the metabolite selected from the table of extracted features. One can view the METLIN ID number and other details such as other names for the compound as well as molecular formula. Most importantly the structural and MSMS spectra are available to compare with any experimental data collected. Fragmentation spectra are collected at multiple collision energies and in both positive and negative ionization modes (only if applicable in both modes).
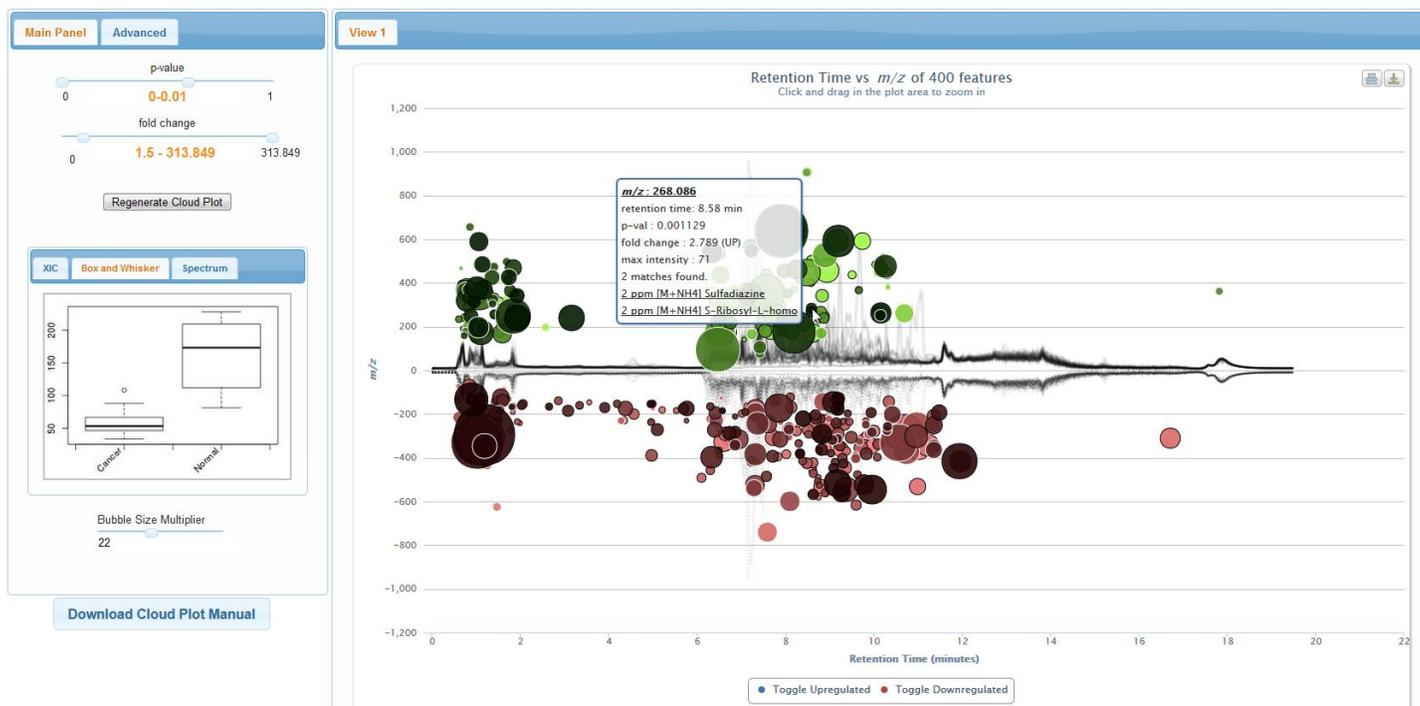
Figure 6: The Interactive Cloud Plot. The cloud plot displays features whose intensities are altered between sample groups. In this case 400 features where highlighted as having a p-value of less than 0.01. Up-regulated features are represented as circles on the top of the plot and down-regulated features are represented as circles on the bottom of the plot, where the size and the degree of color saturation corresponds to the (log) fold change of the feature. Circles with black outlines indicate hits in the database. The lighter or darker the color of the circle relates directly to the significance of the p-value. The cloud plot is completely interactive, where you can filter p value, m/z value, as well as retention time to allow the viewer to see more/less features based on the filtering criteria.

References

1. Bennett Daviss April 2005). Growing Pain In metabolomics. The Scientist. 19 (8): 25–28

2. Baljit K. Ubhi, H. Paul Benton, Duane Rinehart and Gary Siuzdak. A Streamlined Workflow for Untargeted Metabolomics Application Note. RUO-MKT-02-2326-A

3. Benton H.P. and Ivanisevic J. *et al.* Autonomous metabolomics for rapid metabolite identification in global profiling. Analytical Chemistry, 2015 87(2):884-91

4. METLIN database (https://metlin.scripps.edu)

Acknowledgments

Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at
sciex.com/offices