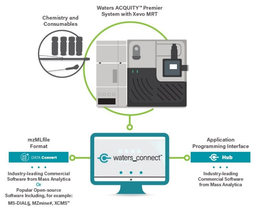
CONVERSION AND INTEGRATION OF OMICS DATA FROM A PROTOTYPE, BENCHTOP MULTI-REFLECTING TIME-OF-FLIGHT (MRT) PLATFORM WITH THIRD-PARTY INFORMATIC WORKFLOWS

Lee A. Gethings, Ian Morns, Pete Reay, Simon Jones, Nyasha Munjoma, Jayne Kirk, Richard Lock
Waters Corp., Wilmslow, Cheshire, United Kingdom

INTRODUCTION

OMIC studies typically generate large and complex data sets, derived from various sample types such as biofluids. As mass spectrometry (MS) technology rapidly advances and with data acquisition methods such as data independent analysis (DIA) progressing, the ability to delve deeper into the metabolome and lipidome is significantly improved, whilst at the same time generating high-dimensional datasets. In this study, we describe a data processing pipeline which converts MS data (DIA and DDA based) that were collected using a novel, multi-reflecting ToF mass spectrometry into a generic file format, that can be readily processed by a variety of third-party informatic tools. Data were primarily processed via Skyline[1] and MS-Dial[2] for lipidomic datasets, whilst metabolomic datasets were processed using XCMS[3].



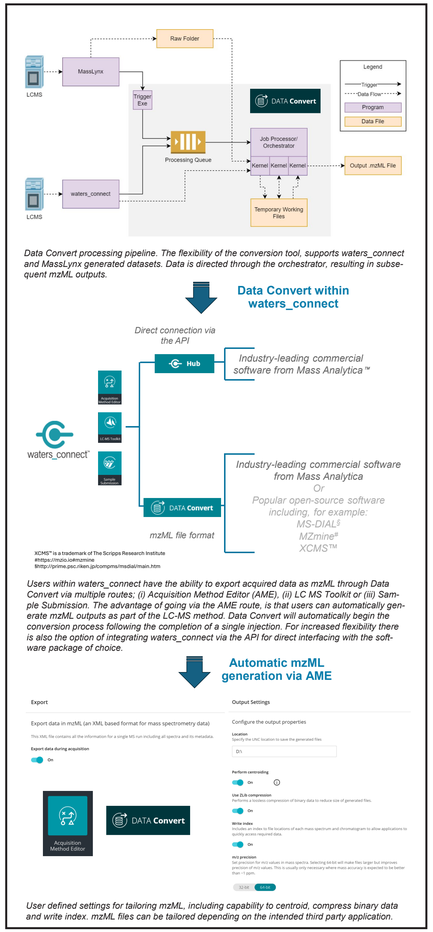
* HUPO Proteomics Standards Initiative, #https://mzio.io/#mzmine, @http://prime.psc.riken.jp/compms/msdial/main.html

Figure 1. waters_connect™ and routes of connectivity to third party informatic packages.

mzML output options available within the LC-MS method created within AME. Users have the option of tailoring the mzML format for compatibility to the various third party informatic packages available:

- *Automatic generation of mzML during the course of the acquisition to ensure*
- *Centroiding function for those applications requiring centroided data. Peak detection is performed with lockmass correction applied.*
- *Compression to minimise the size of mzML's (binary data), particularly suitable for larger sample sets and samples of high complexity.*
- *Writing index out to file, whereby each mass spectrum and chromatogram are indexed (referenced). An option which is dependent on the third party application used for processing.*
- *m/z precision available as 32 or 64-bit versions (dependent on the instrument used for data acquisition).*

Once created, the mzML is available in the defined subdirectory folder specified by the user.



Data Convert processing pipeline. The flexibility of the conversion tool, supports waters_connect and MassLynx generated datasets. Data is directed through the orchestrator, resulting in subsequent mzML outputs.

Users within waters_connect have the ability to export acquired data as mzML through Data Convert via multiple routes; (i) Acquisition Method Editor (AME), (ii) LC MS Toolkit or (iii) Sample Submission. The advantage of going via the AME route, is that users can automatically generate mzML outputs as part of the LC-MS method. Data Convert will automatically begin the conversion process following the completion of a single injection. For increased flexibility there is also the option of integrating waters_connect via the API for direct interfacing with the software package of choice.

User defined settings for tailoring mzML, including capability to centroid, compress binary data and write index. mzML files can be tailored depending on the intended third party application.

Figure 2. Data convert integration into waters_connect. Flexibility is demonstrated with the option of wither generating mzML automatically during data acquisition.

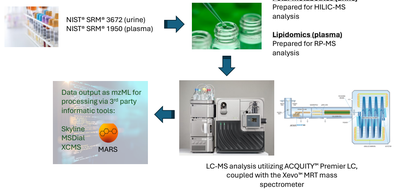


Figure 3. LC-MS analysis of biological samples (urine and plasma) for polar metabolomics and lipidomics. Analytes were chromatographically separated using HILIC (metabolites) or reversed-phase (RP) (lipids)and data collected using the benchtop Xevo MRT mass spectrometer. mzML files were automatically generated via AME during acquisition and processed with a variety of third party informatic tools.

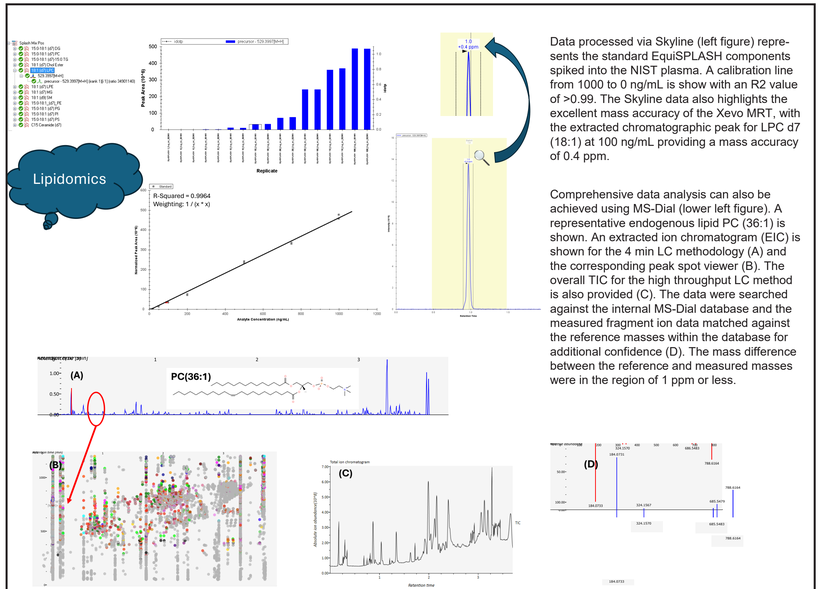


Figure 4. Lipidomic data processing via a variety of third party informatic tools, including Skyline and MS-Dial. Data represented by the Skyline example is based on the EquiSPLASH spiked components, whereas an endogenous lipid (PC(36:1)) is shown when processing with MS-Dial.

RESULTS

Lipidomic data resulting from the plasma extracts were subsequently processed using Skyline and MS-Dial, whilst polar metabolite data relating to the urine samples were processed via XCMS. Exemplar data for each software package is illustrated for the lipidomics (below) and metabolomics (right).

Lipidomics

NIST plasma extracts were spiked with EquiSPLASH™ (Avanti® Polar Lipids, Alabama, USA). Data were imported as mzML format into Skyline and screened against a library consisting of the EquiSPLASH components, highlighting consistency, linear dynamic range and high mass accuracy. The MS-Dial workflow provides a discovery approach with identification of endogenous components, indicating high mass accuracy at the precursor and fragment ion level.

Metabolomics

NIST urine extracts from each of the three levels were pooled to form a QC, which was periodically acquired throughout the LC-MS analysis. Data as mzML format were then uploaded to XCMS for peak picking, statistical interrogation and compound identification.

Data processed via Skyline (left figure) represents the standard EquiSPLASH components spiked into the NIST plasma. A calibration line from 1000 to 0 ng/mL is show with an R2 value of >0.99. The Skyline data also highlights the excellent mass accuracy of the Xevo MRT, with the extracted chromatographic peak for LPC d7 (18:1) at 100 ng/mL providing a mass accuracy of 0.4 ppm.

Comprehensive data analysis can also be achieved using MS-Dial (lower left figure). A representative endogenous lipid PC (36:1) is shown. An extracted ion chromatogram (EIC) is shown for the 4 min LC methodology (A) and the corresponding peak spot viewer (B). The overall TIC for the high throughput LC method is also provided (C). The data were searched against the internal MS-Dial database and the measured fragment ion data matched against the reference masses within the database for additional confidence (D). The mass difference between the reference and measured masses were in the region of 1 ppm or less.

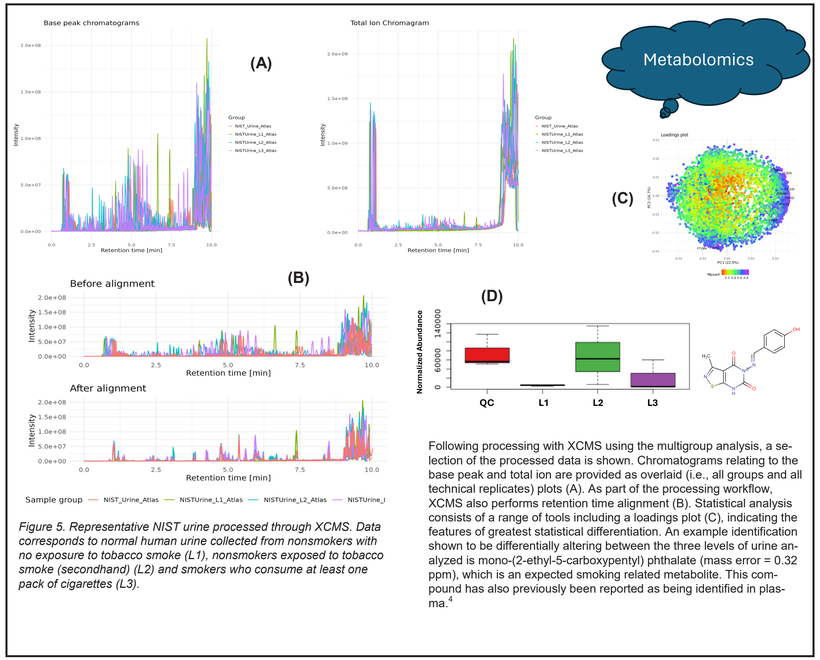


Figure 5. Representative NIST urine processed through XCMS. Data corresponds to normal human urine collected from nonsmokers with no exposure to tobacco smoke (L1), nonsmokers exposed to tobacco smoke (secondhand) (L2) and smokers who consume at least one pack of cigarettes (L3).

Following processing with XCMS using the multigroup analysis, a selection of the processed data is shown. Chromatograms relating to the base peak and total ion are provided as overlaid (i.e., all groups and all technical replicates) plots (A). As part of the processing workflow, XCMS also performs retention time alignment (B). Statistical analysis consists of a range of tools including a loadings plot (C), indicating the features of greatest statistical differentiation. An example identification shown to be differentially altering between the three levels of urine analyzed is mono-(2-ethyl-5-carboxypentyl) phthalate (mass error = 0.32 ppm), which is an expected smoking related metabolite. This compound has also previously been reported as being identified in plasma.[4]

CONCLUSION

- **Data Convert provides a flexible and easy approach to generating mzML files from DDA and DIA acquisitions.**
- **Integration of Data Convert into waters_connect allows a seamless workflow of automatically generating mzML during data acquisition, in addition to having flexibility for users to export mzML from projects at any point in time.**
- **mzML files can be tailored for the intended third party informatics package (i.e., centroided, compressed etc).**
- **Data from the benchtop Xevo MRT has demonstrated the workflow from data acquisition through to data processing with third party tools, such as Skyline, MS-Dial and XCMS using the waters_connect/Data Convert pipeline.**
- **Metabolomic and lipidomic datasets highlight the high quality data that can be generated from the Xevo MRT mass spectrometer:**

⇒ **Superior mass accuracy (sub 1 ppm)**
⇒ **High mass resolution which can be maintained with fast, high throughput LC gradients**
⇒ **High levels of sensitivity/dynamic range**

References

1. Tsugawa, H *et al.*, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, 523-526 (2015).
2. Adams, K.J., *et al.*, Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *J Proteome Res.* **19**(4), 1447-1458 (2020).
3. Tautenhahn R. *et al.*, XCMS Online: A Web-Based Platform to Process Untargeted Metabolomics Data. *Anal. Chem.* **84**(11), 5035-5039 (2012).
4. Barupal *et al.*, Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. *Environ Health Perspect.* **127**(9). Doi: 10.1289/EHP4713.

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS