COMPARISON OF FAST SCANNING DATA DEPENDENT AND DATA INDEPENDENT ACQUISITION METHODS FOR A MULTI-OMIC CANCER STUDY USING HIGH-SPEED CHROMATOGRAPHY

Lee A. Gethings[1], Matthew Daly[1], Martin Palmer[1], Richard Lock[1], Jason Wildgoose[1], James I. Langridge[1]
[1]Waters Corp., Wilmslow, Cheshire, United Kingdom

CONNECTING SCIENCE
To Solve Problems That Matter

Waters™

INTRODUCTION

Metabolomics and lipidomics allows changes within biological systems resulting from disease, treatment, lifestyle, etc to be investigated. Typically, OMICS analyses are performed using a combination of liquid chromatography (LC) and mass spectrometry (MS). Despite developments in analytical technologies the detection and identification of compounds and subsequent biological interpretation remains a significant challenge. Here we evaluate data dependent (DDA) and data independent acquisition (DIA) modes to demonstrate their utilization for multi-OMIC studies performed using fast chromatographic gradients, and thereby addressing challenges such as spectral complexity, datafile size and quantitation. The benefits of these approach are shown using plasma samples from a colorectal cancer (CRC) (colon/rectum cancer) pilot study, using the Xevo™ MRT mass spectrometer with vendor neutral third-party informatic solutions for data processing.

EXPERIMENTAL

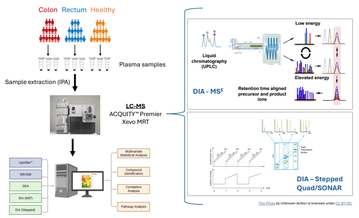

Figure 1. Experimental workflow - Plasma samples collected from individuals diagnosed with colon cancer (n=2), rectal cancer (n=4) and healthy controls (n=6). Samples were protein crashed (chilled IPA) and analyzed via LC-MS. Data were acquired using DDA and DIA techniques. The subsequent data were processed using a variety of software packages for data alignment, peak picking and compound identification. Additional data interrogation using multivariate statistics allowed for potential differential markers of interest between subject groups to be identified and pathway analyzed (Metacore).

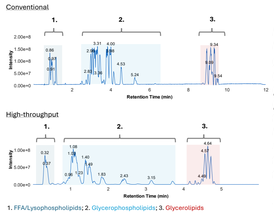

Figure 2. Example lipidomic chromatographic conditions (reversed-phase) for plasma-based samples. The conventional separation is based on a 12-min separation, however, the analyses undertaken within this study comprised of 5 -min separations.

RESULTS

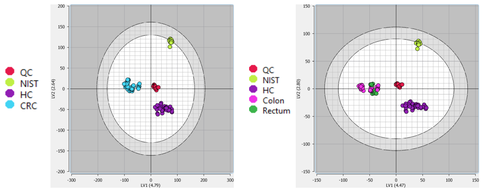

Figure 3. PCA analysis of the sample groups analyzed using DIA. Tight clustering of the replicate injections signifies the high level of reproducibility and confidence in the measurements. Clear separation between groups is also shown with differentiation between the healthy control and CRC groups.

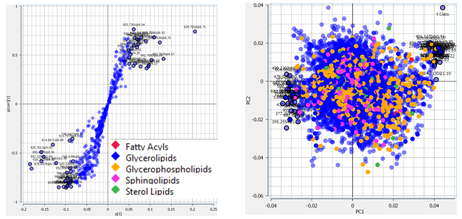

Figure 4. Additional multivariate statistical analysis based on DIA datasets, which include the S-plot (left) and associated loadings plot (right) for healthy controls vs. CRC. Differential markers identified from the S-plot were subjected to identification using a variety of databases including in-house (comprised of standards) and LipidMaps. Features are colour coded on the loadings plot based on lipid class.

Samples were analyzed using LC-MS which comprised of 5-min chromatographic separations and MS data collected using DDA (20 Hz MS/50 Hz MS/MS) and DIA (20 Hz).

Following data processing, the data outputs were further investigated using multi-variate statistical analysis to identify potential markers of interest within the CRC group and to provide potential differentiation between colon and rectum cancers. Example principal component analyses (PCA) from the DIA data are provided in figure 3.

Additional multivariate statistical analysis were performed to establish differential markers of interest (initially curated based on the study QC, ensuring only those with a CV ≤20% were considered) between healthy controls and CRC subjects. These features were subsequently database searched to provide tentative identifications (figure 4). Database searching was conducted using mass tolerances of +/- 1 ppm for both precursor and product ions.

Comparing the fragmentation spectra of all the acquisition methods utilized, resulted in comprehensive structural elucidation and therefore highly confident identifications (figure 5). Incorporating the stepped quadrupole DIA acquisitions resulted in 'DDA-like' spectra, resulting in greater specificity and cleaner spectra, ultimately providing higher identification scores and thus greater confidence in the ID's returned. Data were exported as generic mzML files and investigated using Skyline (figure 6). The stepped quadrupole DIA approach, highlights the high level of specificity, combined with high mass resolution and high mass accuracy.

All data were processed and searched using a variety of software packages, providing a comprehensive list of lipids in return (figure 7). Utilizing the most highly confident lipid identifications, box-whisker profiles indicated the lipids (lipid classes) which were significantly over/under expressed when comparing healthy controls vs. CRC (figure 8). Additional analysis to map significant lipids to their biological role was conducted using Metacore (figure 9), highlighting a number of highly scored pathways, including the WNT pathway.

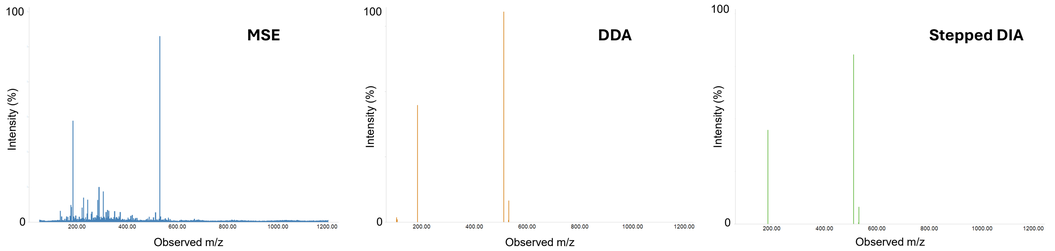

Figure 5. Exemplar spectra for MSE, DDA and stepped quadrupole DIA of the EquiSPLASH LPC (d7) component spiked into human plasma. The quality of the fragmentation data for DDA and the stepped quadrupole DIA (based on 100 Hz MS/MS acquisition rate) are highly comparable and of high specificity, thereby providing increased identification matching scores and confidence in the identifications returned.

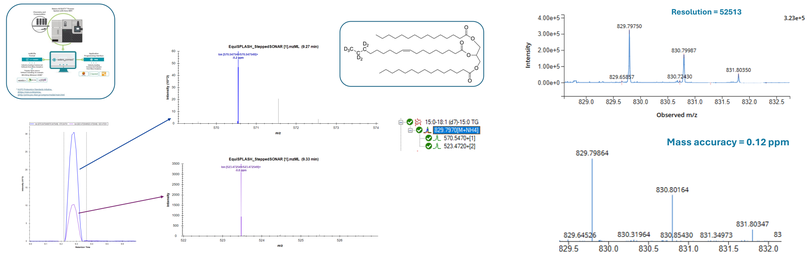

Figure 6. Maintaining instrument performance with increased throughput. The example shown here represents the spiked EquiSPLASH component, 15:0-18:1(d7)-15:0 TAG), acquired using the quadrupole isolated DIA acquisition combined with the high throughput method (6 min). The data was processed using Skyline, demonstrating high mass resolution (top right) and excellent mass accuracy (bottom right) and thereby demonstrating that instrument performance is not compromised with the speed of analysis. Compatibility with 3rd party applications is also highlighted with the flexibility provided via Data Convert with the generation of generic mzML files. The Skyline data (lower left) highlights the representative fragment ions, perfectly aligned with equally excellent mass accuracies at <1ppm.

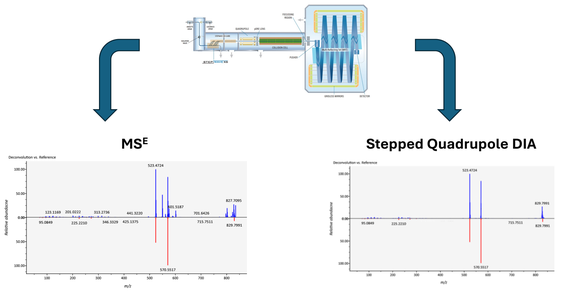

Figure 7. Example lipid identifications returned following database searching. Data acquired with MSE and stepped quadrupole DIA were processed and database searched using MS-Dial[1]. The data here represents the identification of the 15:0-18:1(d7)-15:0 TAG. The fragmentation spectrum for the stepped quadrupole DIA provides a cleaner background (greater specificity) and therefore higher identification score.

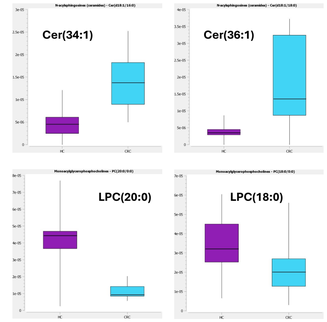

Figure 8. Representative, statistically highly significant lipid identifications which differentiate between healthy controls and CRC. Box-whisker plots based on the relative (normalized) abundances indicate significant under expression for LPC's with CRC (blue), whilst ceramides are up regulated in the CRC group when compared with healthy controls (purple).

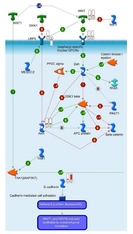

Figure 9. Pathway analysis (Metacore) signifying the implication on the WNT signalling cascade with lipid dysregulation. The WNT pathway has previously been associated with the CRC with a particular role in the regulation of intestinal epithelial cell proliferation.[2]

CONCLUSIONS

- **Xevo MRT MS demonstrates flexible acquisition strategies with an integrated informatic workflow for generic file formats with integration into third party application programs.**
- **Quadrupole based DIA acquisitions demonstrate DDA-like spectra with increased specificity over full scan DIA approaches.**
- **Data collected with DDA or DIA, highlight the performance attributes of the Xevo MRT MS with high levels of sensitivity, high mass accuracy (sub 1 ppm precursor and fragment ions) and the capability of acquiring high quality data with high throughput chromatography.**
- **Based on the analyses of this CRC study using these DDA and DIA strategies, the following were identified:**
  - **Clear separation (based on MVA) of the healthy controls from the CRC group.**
  - **Highly confident lipid identifications (i.e., stringent mass accuracy thresholds, quality of fragmentation data) were determined following database searching.**
  - **Lipid classes identified which differentiated healthy controls from CRC, included LPC's and ceramides.**
  - **Pathway analysis revealed a variety of significant pathways, including the WNT signalling pathway.**

References

1. Tsugawa, H *et al.*, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, 523-526 (2015).
2. Nguyen L *et al.*, Pathways of Colorectal Carcinogenesis. Gastroenterology **158**, 291-302 (2019).

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