**1. Summary**

All worked was performed during a two-week trip of Dr William Nash and Dr Elliott Palmer (University of Birmingham) to visit Dr Ioanna Ntai (Thermo Fisher Scientific, San Jose, USA) in February 2019. 875 chemical standards purchased from MetaSci (<https://www.metasci.ca/>) were prepared in 50 groups of varying numbers of metabolites. All solutions were analysed applying UHPLC-MS to acquire HCD MS2 data and UHPLC retention time data in both positive and negative ion mode using an aqueous C18 reversed-phase assay and a HILIC assay.

## 2. Sample preparation

### Assigning Standard Groups

The MetaSci COMPLETE human standards library (1027 standards at the date of purchase) was kindly provided by Thermo Fisher Scientific. It was suitable for use as it represents a variety of metabolites normally present in human biofluids. To analyse them all individually was not feasible with the given time constraints. As a result, they were divided into groups of approximately 20 metabolites with the goal to analyse as many as possible. Groups were created manually within Microsoft Excel using the information in the reference file provided by MetaSci and the following rules. Groups were approximately sorted by hydrophobicity of the standards within a self-imposed *m/z* difference restriction of at least 4 *m/z* where possible between all metabolites within a single group. This was maintained in the majority of cases for all metabolites and groups. Any metabolites with a molecular mass of less than 100 were not included, all other metabolites were included and considered of equal importance. 50 groups were created in total.

### Sample Preparation

All standards in a single group were weighed out and placed into a 20 mL glass vial. The masses added for all solid-state standards ranged between 0.6 mg and 2.8 mg. Once added they were resuspended in 2 mL of either 50:50 ACN:H2O (for groups assigned as hydrophilic or mixtures) or 50:40:10 MeOH:H2O:IPA (for groups assigned as hydrophobic). The solution was then vortexed for 1 minute and sonicated for 10 minutes. The solution was vortexed again for 1 minute and then diluted 1:5 into either of the following 6:2:2 ACN:MeOH:H2O for the HILIC assay or 50:50 MeOH:H2O for the RP assay. The solution was then centrifuged at 14000 g for 20 mins at 4⁰C. 200 µL of the supernatant was transferred into a LC vial and stored at 4⁰C ready for analysis. Some standards were in liquid form, for these standards 1 µL was utilised, some were also viscous, 1 µL was attempted to be added for these but the true volume added was unknown. Liquid groups were treated the same as the hydrophilic and mixture groups.

**3. Data acquisition**

#### **UHPLC and MS system**

#### Vanquish UHPLC (binary pump) coupled to Thermo Scientific Orbitrap ID-X Tribrid Mass Spectrometer

#### **HILIC UHPLC-MS**

The mobile phase was different for positive and negative ion modes otherwise all parameters were the same.

Column: Accucore Amide-HILIC (100mm x 2.1 mm, 2.6µm)

Scan Range: 70 – 1050 *m/z*

Injection volume: 2µL

Flow Rate: 0.5 mL/min

Mobile phase - positive mode:

1. 10 mM Ammonium formate 95% ACN + 0.1% Formic acid
2. 10 mM Ammonium formate 50% ACN/Water + 0.1% Formic acid

Mobile phase - negative mode:

1. 10 mM Ammonium acetate 95% ACN + 0.1% Acetic Acid
2. 10 mM Ammonium acetate 50% ACN/Water + 0.1% Acetic Acid

The gradient for the HILIC method is outlined in the Figure 1 below.

Graphical user interface

Description automatically generated

**Figure 1**: The gradient utilised for the HILIC method.

#### **Aqueous C18 reversed phase UHPLC-MS**

Column: Hypersil GOLD aQ (100 x 2.1 mm, 1.9 µm)

Scan Range: 100 – 1500 *m/z*

Injection volume: 2µL

Flow rate: 0.3 mL/min

Mobile Phase:

1. Water + 0.1% Formic Acid
2. Methanol + 0.1% Formic Acid

The gradient utilised for the RP method is displayed in Figure 2 below.

Graphical user interface, application

Description automatically generated

**Figure 2**: The gradient utilised for the RP method.

### 

### Acquisition of Data for Metabolite Standards

Data were acquired for the 50 groups in both positive and negative ion mode for HILIC and RP. HCD MS2 data were collected.

**Raw data processing**

All data were processed within Compound Discoverer 3.0 (CD3.0) (Thermo Fisher Scientific, USA). The following nodes were used in the workflow, Select Spectra (Figure 3), Detect Compounds (Figure 4), Group Compounds (Figure 5), Search MassList (Figure 6), Search MzCloud (Figure 7) and Assign Compound Annotations (Figure 8). A mass list containing mol structure files for all standards analysed was created and set as the first priority for annotation assignments. This allowed easy identification of the relevant features in each file if they had been detected. Each standards group/method/polarity combination was processed separately. The resulting compound lists were then browsed manually. When a standard had been annotated the check box adjacent to it was ticked. If MS2 spectra had been acquired they were inspected manually for quality. The mzCloud match score if the standard was already present in mzCloud was checked and quality was further assessed by applying the fISH scoring algorithm. This information along with the intensity and EIC shape informed the manual decision about which spectra to export to the mzVault library. Selected spectra were exported by checking the tickbox adjacent to them and selecting export to mzVault. More than one spectrum may have been exported depending on the number, quality and variance present. A single RP library was created, separate HILC libraries were made for positive and negative ion mode. For each group a manual record was kept in Excel of whether the spectra had been added to the library, if this was for the molecular ion or an alternative adduct form, if spectra were present but not added to the library, if just MS1 data were recorded or if it was not detected at all. All standards detected in a group whether MS2 data were acquired or not had the checkbox ticked, this allowed the export of just these compounds into an excel file. After all groups had been processed the excel files were combined to generate a csv file of all compounds detected. This was a combined file for RP but separate lists were created for HILIC in positive and negative ion mode due to the different mobile phases utilised. Where there was more than one entry in the list for a single standard the feature with the greatest intensity had the RT selected. The resulting .csv files could then be used as a mass list with RT information to provide extra confidence in the identification of features in the biological samples.

Graphical user interface

Description automatically generated

**Figure 3:** CD3.0 Parameters in the Select Spectra node for standards processing.

Graphical user interface, text, application

Description automatically generated

**Figure 4**: CD3.0 Parameters in the Detect Compounds node for standards processing. All ions available were selected.

Graphical user interface, text, application

Description automatically generated

**Figure 5**: CD3.0 Parameters in the Group Compounds node for standards processing.

Graphical user interface, application

Description automatically generated

**Figure 6**: CD3.0 Parameters in the Search Mass Lists node for standards processing.

Graphical user interface, text, application

Description automatically generated

**Figure 7**: CD3.0 Parameters in the Search MzCloud node for standards processing.

Graphical user interface, text, application

Description automatically generated

**Figure 8**: CD3.0 Parameters in the Assign Compound Annotations node for standards processing