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

Biomarkers and Omics

SCIEX

MS-DIAL Data Processing for Untargeted Metabolomics

With SWATH® Acquisition on the TripleTOF® 6600 System

Cyrus Papan, SCIEX, Germany

The major bottle neck in untargeted metabolomics analysis since its conception has been the accurate identification of the metabolites in a complex biological sample. A confident approach for unknown metabolite identification is to match the product ion fragments to a reference MS/MS spectrum. However, data dependent techniques often do not collect MS/MS on all precursors to allow for the identification of metabolites.

The data independent approach using SWATH® Acquisition ensures that product ion spectra is acquired on all detectable compounds in a sample, effectively generating a digitized record of all detectable metabolites in a sample.[1] The fast MS/MS acquisition speed of the TripleTOF® 6600 System (up to 100 MS/MS per second) is key for acquiring high quality SWATH acquisition data on metabolomic samples. In addition, the use of variable sized Q1 windows[2] improves the specificity in the fragment assignment. Furthermore, the SWATH acquisition MS/MS data can be used for quantitation, allowing an MRM-style quantification approach at the MS/MS level.

MS-DIAL (**M**ass **S**pectrometry – **D**ata **I**ndependent **A**na**L**ysis) is an open-source software for the identification and quantification of small molecules and lipids from DIA and DDA-based untargeted LC-MS/MS analysis[3]. It leverages the power of the

SWATH acquisition method for untargeted metabolomics analysis using a two-step process: data is deconvoluted by MS²Dec algorithm, then the precursor and fragment ions are re-associated to obtain purified specific product ion spectra of each precursor ion (Figure 1). The 'purified MS/MS spectra' provides high accuracy for metabolite identification and better identification coverage of low abundant metabolites. Here, the use of MS-DIAL for processing SWATH acquisition data from the TripleTOF 6600 system is demonstrated. A comparison of metabolites from two different strains of *Arabidopsis* was performed.

Key Features of SWATH Acquisition with MS-DIAL for Untargeted Metabolomics

- Single injection workflow for metabolite quantification and identification using SWATH acquisition
  - Variable window Q1 isolation provides improved specificity while maintaining sample coverage
- Complete fragmentation of all detectable analytes within the selected mass range
- Purified product ion spectra are generated from peak detection and deconvolution for better library matching
- Quantitative data obtained from MS/MS provides higher specificity of quantitation
- Large and open compound libraries can be used for spectral matching and compound identification
- Generation of a comprehensive digital record of each metabolomics sample enables retrospective data mining
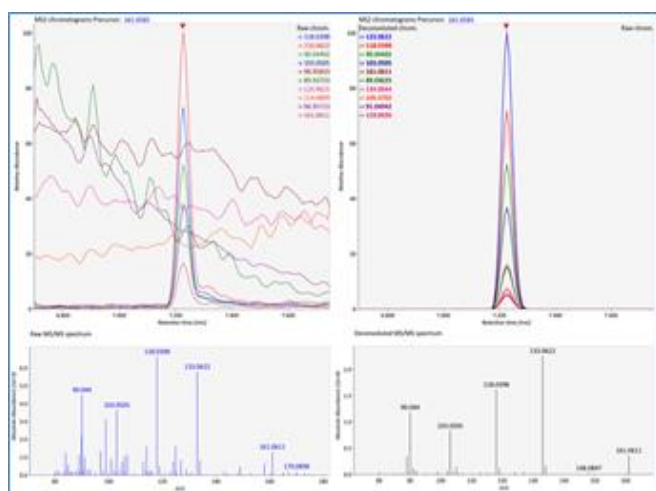

Figure 1. Purification of Convoluted SWATH Acquisition Spectra by MS-DIAL. The left shows the raw MS/MS XIC and spectrum of methoxycinnamic acid. After deconvolution, the noise at both the XIC and spectrum level is substantially reduced, providing higher quality quantitation and more identification confidence. Note the small red arrows above the XICs, indicating the retention time of the aligned chromatographic peaks.

Methods

Sample Preparation: The upper aqueous phase of a chloroform: methanol extraction of plant material from two different strains of the widely used model organism *Arabidopsis thaliana* (mouse ear cress) were prepared.

LC-MS/MS Analysis: The samples were analyzed using a TripleTOF® 6600 system with a Shimadzu Nexera HPLC system using variable window SWATH acquisition method in positive ion mode. Q1 mass range from 80 to 600 Da was covered and MS/MS was acquired in high resolution mode (30000 resolution).

Data Processing: Data was converted using ABF converter interface, Reifycs Analysis Base File Convertor to convert the SCIEX data file (*.wiff) to ABF format[3]. The ABF converter interface with experimental data files loaded for conversion is shown in Figure 2. ABF files were then processed in MS-DIAL software pipeline, with multiple files loaded in one work session. Chromatographic peaks are integrated and aligned for quantitative comparison between sample groups. For the library matching, MS-DIAL utilizes the open source NIST MSP text file format library for the fragment spectral library matching to MSP-libraries from public compound data bases such as MassBank[3] or LipidBlast[4] for compound identification.
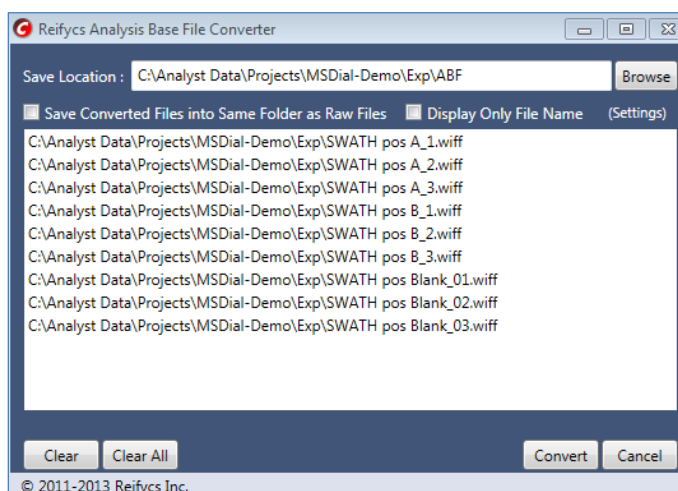

Figure 2. The ABF Converter Interface. Data files loaded for conversion by dragging into the window.

Metabolomic Comparison of Two Strains of *A. thaliana*

Three data files of each cell line group (A1-A3 and B1-B3) were loaded into the MS-DIAL software version 2.94 and alignment was done. The main results window is shown in Figure 3.
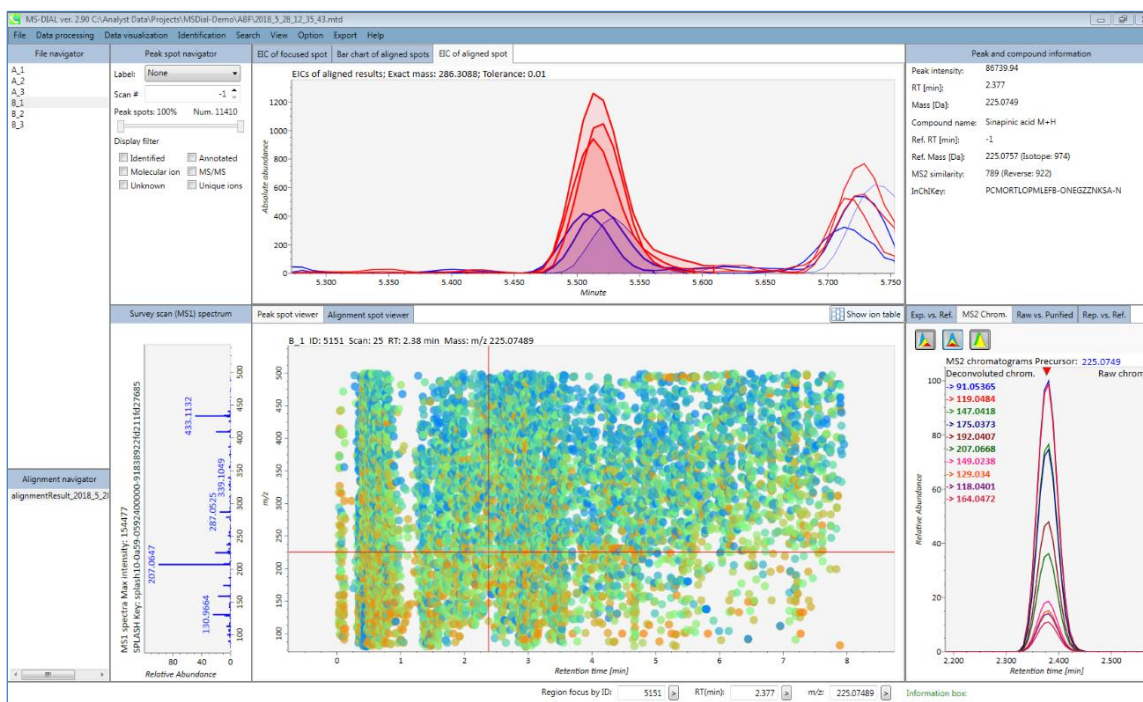

Figure 3. The MS-DIAL Results Window for Easy Data Visualization. The graphical interface includes the *File Navigator* which shows the loaded files, while the *Alignment Navigator* shows the alignment results of the loaded files. The *Peak Spot Navigator* allows filtering of the results in several ways (identified peaks, annotated peaks, etc). The *EIC Panel* shows the extracted ion chromatograms of the spots highlighted in the *Peak Spot Viewer* where the color corresponds to the intensity. Lastly, the top panel shows *Peak and Compound Information* for an identification, and below are the aligned raw MS/MS chromatograms or the raw and purified product ion spectra.

Figure 1 shows the raw (upper left panel) and purified (upper right panel) spectra at the retention time indicated by the red arrow. It is evident that some of the co-eluting fragment ions are low abundant compared to contaminating but not co-eluting fragments. Many peaks present in the un-purified product ion spectrum have been eliminated by the purification process. This results in a much cleaner product ion spectrum, which is then better suitable for matching with a spectral library.

Matching of SWATH Acquisition Data to Spectral Libraries in MS-DIAL

MS-DIAL uses spectral libraries in the open NIST msp text format, converted from MassBank databases, a public mass spectral database system. Currently, MassBank contains around 220 000 experimental and *in silico* spectral records of over 73,000 unique compounds. The library is loaded into MS-DIAL, then the deconvolved SWATH acquisition or DDA product ion spectra can be matched to the database reference spectra.

In Figure 4, three examples of identifications from the SWATH acquisition data by MS/MS spectral matching are shown. The dot-plot of m/z vs. retention time is filtered to show only identified peaks. Note that the software also assigns possible adducts, as shown in the middle and bottom left dot plots.

Table 1 shows a list of identified compounds exported from the MS-DIAL software. Thirty-nine metabolites were identified with confidence from both of the plant strains. The average RT and average m/z are reported for the measured precursor ion.
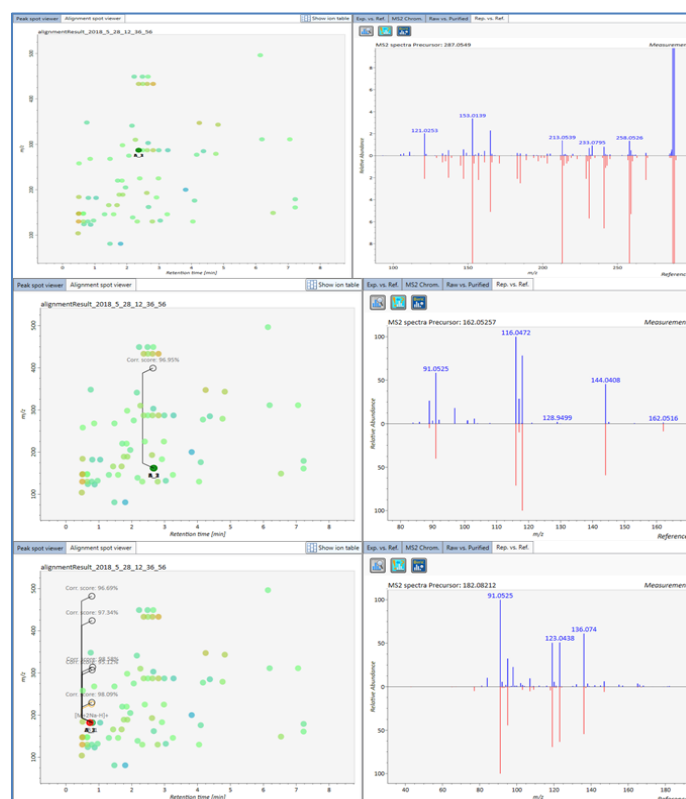

Figure 4. Three Examples of Matched Product Ion Spectra with Database Spectra from SWATH Acquisition Data. The left panels show the m/z vs. retention time plot, with each dot representing one feature. Adducts are represented with lines on the dot plot. The right panels show the spectra comparison with the purified SWATH acquisition product ion spectra in blue and the reference library spectra in red. Top: Kaempferol; Middle: Indole Carboxylic Acid; Bottom: Tyrosine.
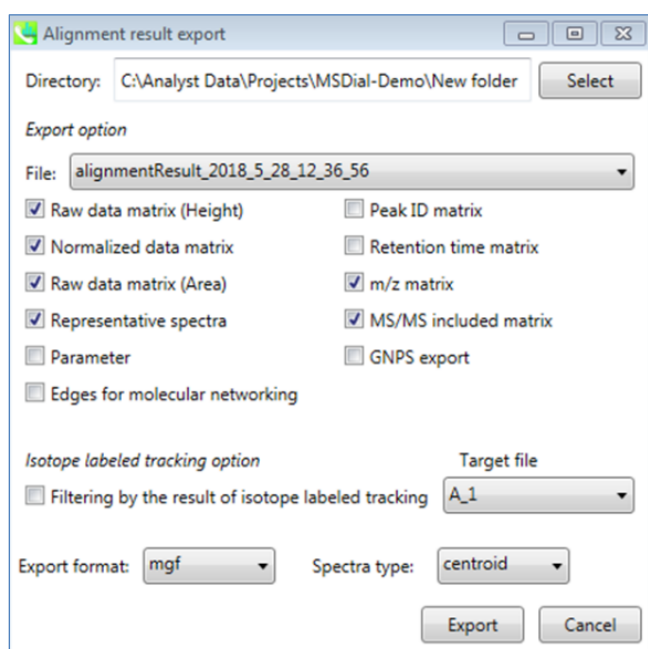
Post-Processing Data Analysis Options

MS-DIAL data output can be exported in several file formats for post processing data analysis as shown in Figure 5. Various metrics are selectable in the data export window. Data exported in individual text file formats can be used for further data processing in other tools, such as Excel, or other statistical software packages.

Data can also be exported as a generic text file for import into MarkerView™ software. Here, the aligned chromatographic peak areas were exported for features with significant changes for further analysis.

Figure 5. The MS-DIAL Data Export Window.

Table 1: List of Identified Compounds with MS-DIAL. Metabolite identification was done based on MS/MS spectral matching to the selected spectral database with high confidence.

RT (min)	m/z	Ion Type	Metabolite Name	Area
0.50	130.0525	[M+H]+	L-5-Oxoproline	5754681
0.50	147.0805	[M+H]+	Glutamine	2271359
0.51	148.0640	[M+H]+	Glutamic acid	2896310
0.51	184.0781	[M]+	Phosphocholine	1475994
0.59	104.1062	[M]+	Choline	661963.4
0.75	348.0695	[M+H]+	Adenosine 3'-monophosphate	112507
0.79	130.0845	[M+H]+	Pipecolate	252351.1
1.03	182.0795	[M+H]+	Tyrosine	136850.9
1.47	268.1052	[M+H]+	Adenosine	582340.4
1.70	166.0866	[M+H]+	Phenylalanine	940387.6
1.82	220.1175	[M+H]+	Pantothenate	240060.6
1.87	298.0965	[M+H]+	5'-S-Methylthioadenosine	315190.7
1.95	205.0985	[M+H]+	Tryptophan	109640.9
2.19	139.0386	[M+H]+	Salicylic acid	154445.5
2.27	310.1643	[M]+	Sinapine	56305.83
2.33	130.0652	[M+H-H2O]+	Indole-3-carbinol	423719
2.40	225.0745	[M+H]+	Sinapic acid	44123.27
2.63	193.0509	[M+H]+	Scopoletin	674717.8
2.63	287.0542	[M+H]+	Luteolin	104320
2.67	162.0536	[M+H]+	1H-indole-3-carboxylic acid	63939.38
2.82	287.0544	[M+H]+	Kaempferol	321494.4
2.49	287.0542	[M+H]+	Luteolin	221075.1
2.82	433.1125	[M+H]+	Kaempferol-7-O-alpha-L-rhamnoside	1427938
2.99	225.0721	[M+H]+	Sinapinic acid	240669.5
3.25	146.0591	[M+H]+	3-Formylindole	50114.78
4.25	347.2222	[M+H]+	Reichsteins substance S	2921677
4.34	220.0822	[M+H]+	1-Isothiocyanato-7-(methylsulfinyl)-heptane	140871.7
4.72	234.0993	[M+H]+	1-Isothiocyanato-8-(methylsulfinyl)-octane	708197.6
4.83	343.2955	[M+H]+	CocamidopropylBetaine	1096473
6.15	496.3396	[M+H]+	LysoPhosphatidylcholine16:0	93726.45
6.54	149.0228	[M+H]+	Phthalic anhydride	868146.5
7.06	311.1637	[M+H]+	Avobenzone (Parsol 1789)	225197.2
7.22	161.0585	[M+H-H2O]+	3-Methoxycinnamic acid	249725.4
7.22	179.0685	[M+H]+	3-Methoxycinnamic acid	123098.6

Quantitative Metabolomics

Principal component analysis (PCA) can then be performed on the identified metabolites within MS-DIAL. A clear separation between the two plant groups was observed in the scores plot (Figure 6). The loadings plot highlights the individual components responsible for the separation. The labels in the loadings plot are filtered to show only the identified components. The contribution plot below shows that over 70% of the variability in the detected features are correlated with the type of sample.

Conclusions

The fast acquisition speed and wide dynamic range of the TripleTOF 6600 system combined with SWATH acquisition generates structural and quantitative information of all detectable compounds across a wide concentration range from biological samples. The MS-DIAL software leverages the power of the SWATH acquisition data for untargeted metabolomics acquisition by extracting the product ion spectra for MS/MS fragment matching. It then utilizes the accurate mass, isotope ratios information, and retention time prediction for identification which exceeds the two orthogonal parameters guideline by the Metabolomics Standards Initiative[7].

The MS-DIAL generated product ion spectra are purified from co-eluting chemical noise and are often cleaner compared to DDA-derived spectra, resulting in better matching with spectral libraries. MS-DIAL also supports normalization methods for MS/MS quantitation analysis. The complete workflow has been demonstrated here for the quantitative comparison of two strains of *A. Thaliana*.
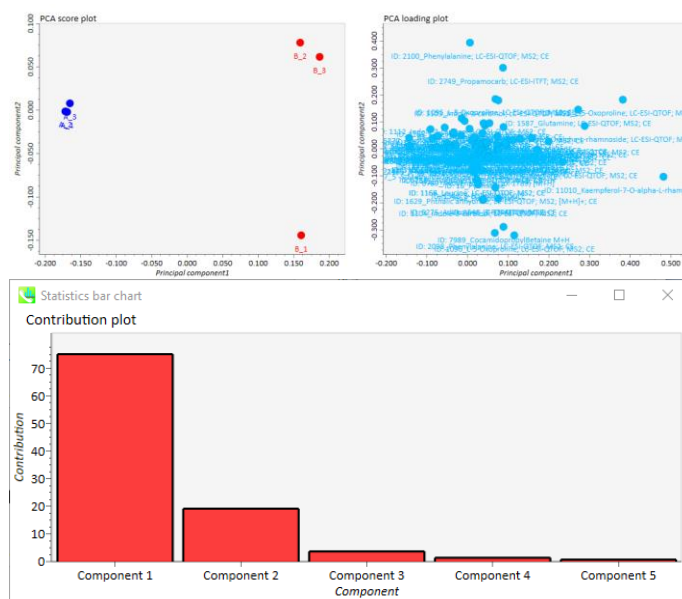

Figure 6. Principal Component Analysis. The Scores plot of the two samples, A and B on the left shows the differentiation between the two samples and the loadings plot on the right shows the aligned features. The contributions plot below shows that over 70% of the variation in the data is correlated with the sample groups.

References

1. Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal Chem.* (2014) **86(2)**: 1202-9.

2. Improved Data Quality Using Variable Q1 Window Widths in SWATH® Acquisition. SCIEX technical note RUO-MKT-02-2879-B.

3. MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat Methods.* (2015) **12(6)**: 523–526.

4. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spec.* (2010) **45(7)**: 703-14.

5. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat Methods.* (2013) **10(8):** 755-8.

6. Download ABF converter. http://www.reifycs.com/AbfConverter/

7. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics.* (2007) **3(3)**: 211-221.