Rapid analysis and interpretation of metabolomics SWATH acquisition data using a cloud-based processing pipeline

Using the OneOmics suite with the SCIEX ZenoTOF 7600 system

Alexandra Antonoplis[1], Jason Causon[2], Christie Hunter[1]
[1]SCIEX, USA, [2]SCIEX, Canada

SWATH acquisition is a data independent acquisition (DIA) workflow that has been demonstrated to improve metabolite coverage over traditional data dependent techniques for untargeted metabolomics.[1,2] The workflow enables creation of a digitized record of the metabolome present in a sample, with full (MS1) and MS/MS scans capturing every detectable analyte in a single injection. The richness of SWATH acquisition data offers numerous data analysis opportunities, one of which is to identify differential metabolites across sample groups. While instrumentation and appropriate methods for collecting metabolomics SWATH acquisition data are well-established, the lack of software tools for processing large-scale metabolomics SWATH acquisition studies remains a challenge for wide-spread adoption of the workflow in laboratories.[3]

In this work, the OneOmics suite, a cloud-based solution for SWATH acquisition data processing, was used to investigate analytes present in the urine of Zucker Diabetic Fatty (ZDF) vs. control Sprague Dawley (SD) rats. ZDF rats are widely used as an animal model of genetic type 2 diabetes. The OneOmics suite facilitated complete end-to-end processing of the metabolomics data set, starting with ion-library driven extraction of analyte peak groups from the SWATH acquisition data. The platform features built-in false discovery rate (FDR) analysis and normalization algorithms to facilitate accurate identification of differentially expressed analytes. Data sets can be further interrogated using multivariate statistical analysis tools and viewed in a biological context with pathway mapping (Figure 1).

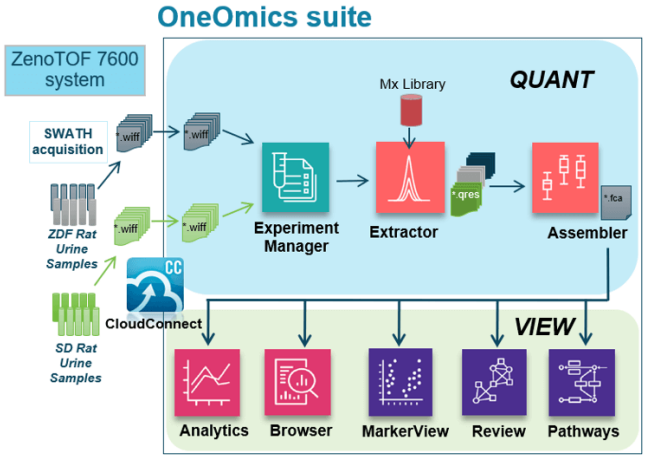

Figure 1. Overview of the metabolomics workflow for SWATH acquisition data processing in the OneOmics suite. Following collection of variable window SWATH acquisition data using the SCIEX ZenoTOF 7600 system, data files were uploaded to the OneOmics suite for processing. Experimental meta data was entered, and then the Extractor and Assembler Apps were used to extract and quantify metabolites across the samples. Additional apps were used to visualize the results and perform statistical analysis to determine the metabolite differences between ZDF vs. SD rat urine.

Key features of the OneOmics suite for metabolomics data processing

- OneOmics suite enables fast and confident identification of differential metabolites across experiment groups in large-scale metabolomics SWATH acquisition studies
- The entire processing workflow, from meta data management to examining biological significance of results, is supported in the platform
- OneOmics suite features specific algorithms for metabolomics SWATH acquisition peak extraction and scoring, FDR analysis, and normalization, along with data dashboards for rapid assessment of data quality and results
- Enriched analytes can be mapped to biological pathways using the Bioreviews App

Methods

Sample preparation: Urine samples were collected from four distinct rat groups: male ZDF rats, female ZDF rats, male SD rats, and female SD rats. Urine was collected from N=5 rats per group and diluted 10-fold in mobile phase A prior to LC-MS/MS analysis.

Chromatography: An ExionLC AD HPLC system (SCIEX) with a Phenomenex Luna Omega Polar C18, 3 µm 150 x 2.1 mm (00F-4760-AN) at 40 ºC was used with a flow rate of 300 µL/min. The reversed-phase gradient method is listed in Table 1. The injection volume was 2 µL.

Table 1. LC Gradient for Urine Metabolomics Analysis

Time	%A	%B
0.0	*99*	*1*
0.5	*99*	*1*
3.0	*90*	*10*
5.75	*75*	*25*
7.5	*72*	*28*
10.0	*20*	*80*
11.0	*5*	*95*
11.5	*5*	*95*
11.6	*99*	*1*
13.1	*99*	*1*

Mobile phase A – Water with 0.1% formic acid
Mobile phase B – Acetonitrile with 01.% formic acid

Mass spectrometry: SWATH acquisition data was collected using a SCIEX ZenoTOF 7600 system and 80 variable Q1 windows. For the TOF MS survey scan, an accumulation time of 50 ms was used, while an accumulation time of 10 ms was used for MS/MS scans. The ion source conditions were as follows: CUR 30, GS1 50, GS2 50, ISVF 5000, TEM 450.

Data processing: CloudConnect in PeakView software 2.2 was used to upload data files to the OneOmics suite in the cloud. The meta data was assigned for the study using the Experiment Manager App. Analyte peak groups were then extracted in the Metabolomics App using an extracted ion chromatogram (XIC) extraction width of 40 ppm with 6 transitions required per analyte. Upon normalization and computation of log2 signed fold change values, data were examined in the Bioreviews App for identification of enriched analytes and pathway mapping.

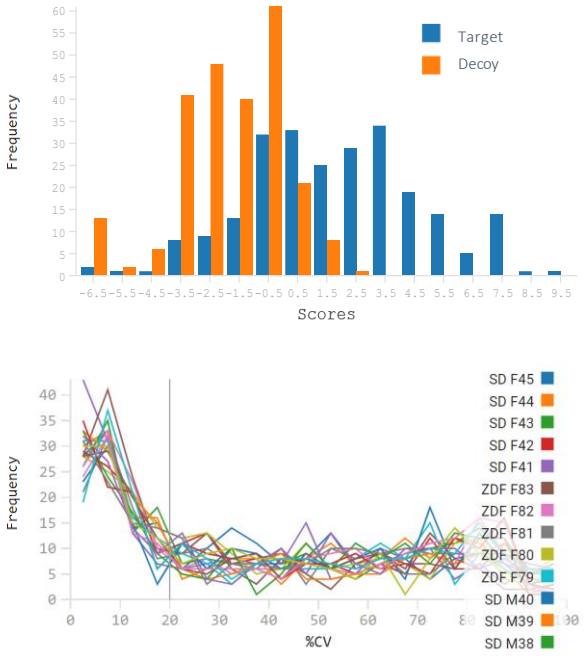

Figure 2. Assessing analytical reproducibility using the Analytics App. Top: for each data file, an FDR analysis is performed to identify analytes that pass a 1% FDR cutoff filter. Within the Analytics App, the distribution of scores for both target and decoy analytes can be viewed for each SWATH acquisition data file, along with the number of analytes passing the 1% FDR cutoff threshold. Bottom: the peak area reproducibility of analytes with at least one confident detection (<1% FDR) across the technical replicates was assessed. The vertical line marks the number of analytes passing a 20% CV filter.

Examining data quality in the Analytics App

To investigate different abundance levels of metabolites in the urine of diabetic (ZDF) rats as compared to healthy control (SD) rats, SWATH acquisition data files were analyzed in the OneOmics suite. Prior to exploration of the quantitative results, the OneOmics suite was first used to evaluate the quality of the MS data. Using the Experiment Manager App, the data files were organized into groups of technical replicates to assess the analytical reproducibility of the LC-MS/MS workflow. The technical replicate groups were then processed using the Extractor App, the first in a series of applications for processing of SWATH metabolomics data.

The Extractor App extracts and integrates analyte precursor and fragment XIC areas using an ion library. The resulting peak groups are then scored according to chromatographic attributes, such as peak width and peak intensity ratio, and spectral attributes, including mass error of the monoisotopic peak and MS/MS m/z error. These target analyte scores are then incorporated into an FDR analysis within the application to assess the statistical accuracy of the detections. The FDR

analysis follows a target-decoy approach, in which decoy analytes are generated *in silico* by selecting precursor and decoy fragment ions from the ion library used for data processing according to predetermined criteria.[4] The resulting distribution of target vs. decoy is easily visualized in the Analytics App (Figure 2, top).

Upon analyte extraction, the technical replicate data files were then processed using the Assembler App, which normalizes the extracted results using a most likely ratio (MLR) approach and then calculates areas, metabolite fold changes, and confidences across the samples. Finally, the results could then be visualized to assess the reproducibility of analyte detections across the technical replicates (Figure 2, bottom). Across the technical replicate groups, analytes with <20% CV exhibited the highest total frequencies, with maximal total frequencies occurring at 7-8% CV for the majority of the replicate groups.

Figure 3. Analytes identified in urine of ZDF rats as compared to SD rats. (Top) Heatmap of analytes quantified can be generated in Browser for rapid visualization of the changes observed for the 236 detected analytes as compared to a control sample group. For visualization, a 75% fold change confidence filter and MLR reproducibility filter of 0.15 were applied, 84 metabolites passed these filters and thus had significant change. The intensity of the color (orange for up-regulation, blue for down-regulation) indicates the magnitude of the log2 signed fold change, and a subset of the heatmap is shown, as visualized in the plot on the right. (Bottom) Selecting an individual analyte in the heat map (indoxyl) enables viewing of its associated normalized transition peak areas across all samples. Each color represents a different transition, and overall normalized analyte areas can also be viewed for each sample.

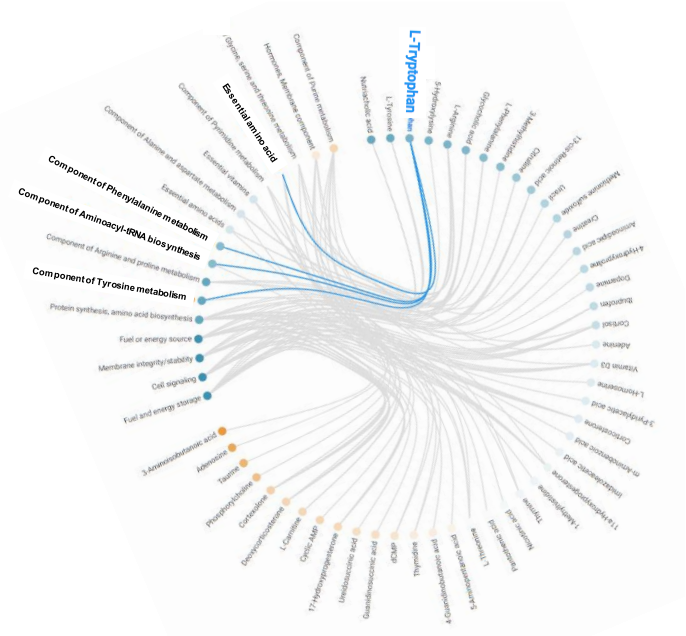

Figure 4. Exploring individual analytes in Browser. Shown at the top are biofunction ontology terms for L-tryptophan, a differential analyte in ZDF vs. SD rats. Hovering over an analyte of interest in the ontology wheel reveals its associated ontology terms, which are provided by the HMDB. Hovering over an ontology will highlight all the metabolites found within that ontology. The color of the dot highlights the direction of the fold change (positive fold change = orange, negative fold change = blue, no fold change = white).

Exploring differential analytes in the Browser App

Next, the data files were grouped according to rat type (diabetic vs. healthy) and sex (female vs. male) to explore up-regulated and down-regulated metabolites. The files were processed again using the Extractor and Assembler App, and then results were visualized using the Browser App. The Browser App features a dashboard of visuals to showcase the direction, extent, and confidence of detected metabolite fold changes, as well as ontology information from the Human Metabolome Database (HMDB) for each differentially expressed analyte. Metabolite fold changes in the Browser App can further be interrogated in heatmap form, with processing filters to visualize detections with high confidence and reproducibility (Figure 3).

As compared to control rats, several highly differential analytes were identified in the ZDF rats across both sexes. The analytes exhibiting the greatest log2 signed fold change values were indoxyl (6.521), xanthosine (4.536), and L-tryptophan (-3.02).

Indoxyl sulfate is a known uremic toxin that has been detected in the urine of type 2 diabetes patients with low renal function, and xanthosine has been identified as a urinary metabolite biomarker of type 2 diabetic nephropathy.[5,6] In a metabolomics analysis of the urine of ZDF rats, tryptophan was also found to be down-regulated.[7] The identified analytes can also be correlated to their biofunction in the Ontology section of Browser (Figure 4). The analyte L-tryptophan was linked to its role in amino acid metabolism using the ontologies provided by HMDB, highlighting processes that might be affected by its down-regulation in ZDF rat urine as compared to SD rat urine.

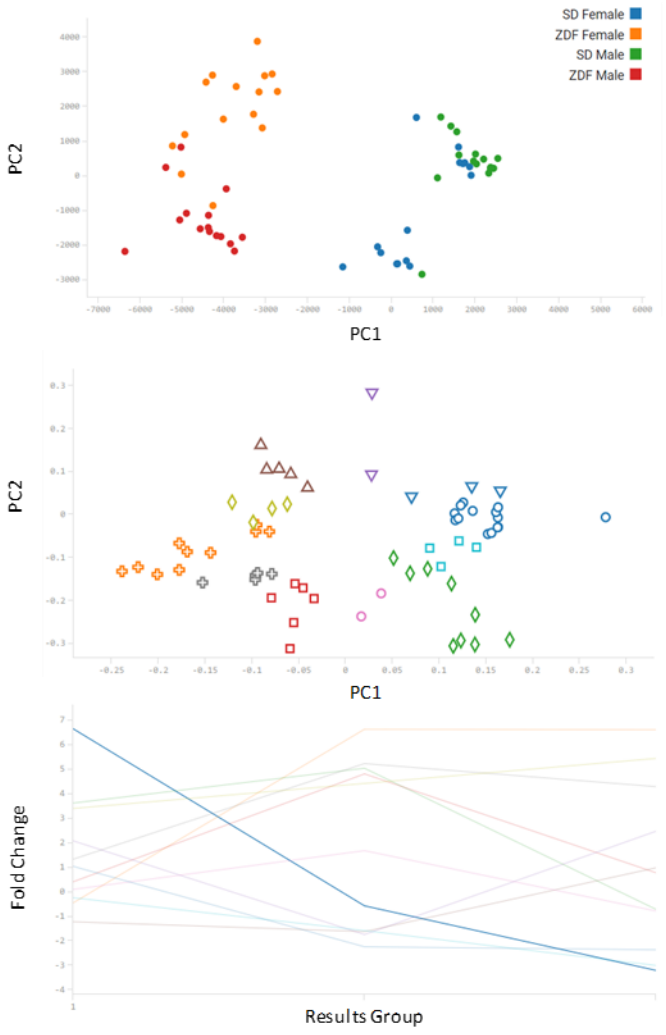

Figure 5. Using the Bioreviews App for multivariate statistical analysis and pathway mapping. (Top) PCA analysis revealed distinct clustering of the sample groups, with healthy SD rats forming one distinct subset and ZDF rats forming a second subset. (Middle) The loadings plot shows the distinct PCVG groups of analytes, clustered according to similar patterns of change across samples. (Bottom) Selecting a particular PCVG cluster group reveals quantitative differences across the sample groups.

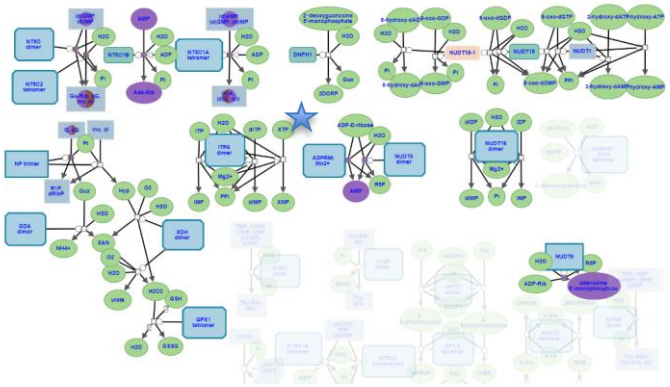

Figure 6. Pathway mapping with Bioreviews App. Mapping of PCA-PCVG cluster groups that contained metabolite differences to biological pathways revealed enrichment of the purine catabolism pathway (p =.0049). Green ovals represent small molecules and outlined blue boxes represent proteins and protein complexes. Analytes identified from the data set, including adenosine, guanosine, and deoxyguanosine, are shown in purple and brown. Xanthosine triphosphate (XTP – marked with star) is also an intermediate in this pathway. Xanthosine was found to be significantly enriched in the urine of ZDF rats in the SWATH acquisition data set.

Using the Bioreviews App for multivariate data analysis and pathway mapping

In the Bioreviews App, the MarkerView App can be used to perform multivariate statistical analysis of quantitative metabolomics results, including principal components analysis-principal component variable grouping (PCA-PCVG) analysis. For these analyses, a PCA Scores plot of the sample groups indicated significant differentiation between the ZDF rat and SD rat sample groups (Figure 5, top). The Loadings plot shows the PCVG groups, clustering metabolites with similar patterns of expression (Figure 5, middle). Selecting a group will highlight the analytes within that group and show the quantitative differences across the sample groups (Figure 5, bottom).

After performing multivariate statistical analysis, analyte cluster groups can be mapped to biological pathways in the Pathways App (Figure 6). Pathways analysis is conducted using Reactome, and the application enables visualization of the enriched biological and chemical pathways from the results. The Pathway Browser revealed significant enrichment of the purine catabolism pathway in the data set, in which purine bases are converted to xanthine. This pathway was previously identified as enriched in a study of the urinary metabolome of Zucker diabetic rats as compared to healthy control rats.[7] Increases in the analyte xanthosine in urine have been linked to increased purine catabolism, and xanthosine has been observed to be significantly increased in the urine of diabetic rats.[8] Increased

purine catabolism has been proposed as a homeostatic response of mitochondria to oxidative stress.[9]

Conclusions

In this study, the OneOmics suite was used to process a metabolomics data set acquired using SWATH acquisition on the SCIEX ZenoTOF 7600 system and to detect differentially regulated metabolites in the urine of ZDF versus healthy rats. Results were compared with previous metabolomic studies of diabetic rats to interpret the findings. The complete integration of data processing applications in the cloud-based platform enabled rapid processing of the study data. Built-in FDR analysis enabled library-driven peak identifications to be assessed for statistical accuracy. The platform also features tools for exploring the biological relevance of detected analytes, including ontology enrichments and pathway mapping.

References

1. SWATH acquisition improves metabolite coverage over traditional data dependent techniques for untargeted metabolomics. SCIEX technical note, RUO-MKT-02-7128-A.
2. Improved metabolite identification using data independent analysis for metabolomics. SCIEX technical note, RUO-MKT-02-10617-A.
3. Raetz M, Bonner R, Hopfgartner G. (2020) SWATH-MS for metabolomics and lipidomics: critical aspects of qualitative and quantitative analysis. *Metabolomics* **16**, 71.
4. How does FDR scoring work in the OneOmics suite for metabolomics? SCIEX community post RUO-MKT-18-13219-A.
5. Ng DPK et al. (2012) A metabolomic study of low estimated GFR in non-proteinuric type 2 diabetes mellitus. *Diabetologia* **55**, 499–508.
6. Chen CJ et al. (2018) Identification of Urinary Metabolite Biomarkers of Type 2 Diabetes Nephropathy Using an Untargeted Metabolomic Approach. *J Proteome Res.* **Nov 2;17(11)**: 3997-4007.
7. Salek RM et al. (2007) A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. *Physiol Genomics.* **Apr 24;29(2):** 99-108.
8. Men L et al. (2017) Metabolomics insights into diabetes nephropathy and protective effects of Radix Scutellariae on rats using ultra-high performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry. *RSC Adv.*, **7**, 16494-16504.
9. Kristal BS et al. (1999) Purine catabolism: links to mitochondrial respiration and antioxidant defenses? *Arch Biochem Biophys.* **Oct 1;370(1):** 22-33.

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