Untargeted Metabolomics Workflow Using UHPLC/Quadrupole Orbitrap Mass Spectrometer and SIEVE 2.1 Software

Junhua Wang, David A. Peake, Mark Sanders, Michael Athanas, and Yingying Huang
Thermo Fisher Scientific, San Jose, CA

Thermo SCIENTIFIC

Overview

Purpose: Demonstrate a generic, integrated workflow for untargeted metabolomics study using a UHPLC/benchtop Thermo Scientific™ Orbitrap™ mass spectrometer and informatics software.

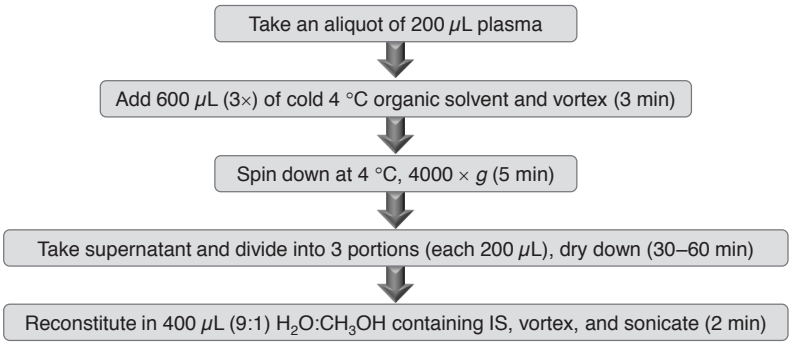
Introduction

Metabolomics is a rapidly growing field of post-genomic biology, aiming to comprehensively characterize the small molecules in biological systems. Nonbiological systematic biases from instrument calibration or the order of sample injection account for the most significant errors in LC/TOF-MS data [1]. Here we present a workflow using a UHPLC/benchtop quardrupole Orbitrap platform and automated data analysis software for untargeted metabolomic profiling of plasma samples for biomarker discovery from the Zucker diabetes fatty (ZDF) rat model. The optimal conditions for sample preparation, liquid chromatography (LC), column, mass spectrometry (MS), and data processing parameters are explored.

Methods

Sample Preparation

Plasma samples were deproteinized with organic solvent. Four extraction solvent systems including methanol (MeOH), acetonitrile (ACN), acetone, and 1:1:1 of the above were tested in this work. Endogenous metabolites were reconstituted in methanol/water (1:9) containing isotopically labeled internal standard (IS), d5-hippuric acid for LC-MS analysis. Solvent blank, pooled QC, and biological samples were analyzed in a randomized injection order.



Liquid Chromatography

UHPLC separation was implemented on a Thermo Scientific™ Dionex™ UltiMate™ 3000 HPG (high-pressure gradient) pump using Thermo Scientific™ Hypersil GOLD™ RP C18 column at 450 µL/min, column temperature at 55 °C. LC solvents were 0.1% FA (A) and 0.1% FA in MeOH (B). Apply linear gradient from 0.5–50% B for 5.5 min, followed by increasing to 98% at 6 min, hold 98% B for 6 min, then decrease to 0.5% at 13 min, then equilibrate for another 2 min.

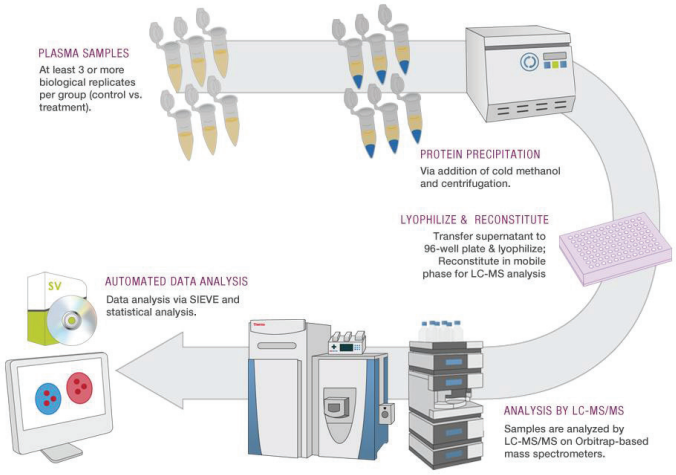
Mass Spectrometry

The Thermo Scientific™ Q Exactive™ mass spectrometer was operated under electrospray ionization (ESI) positive, negative, and polarity switching modes. Full scan (*m/z* 67–1000) used resolution 70,000 with automatic gain control (AGC) target of 1×10^6 ions and a maximum ion injection time (IT) of 35 ms. Data-dependent MS/MS were acquired on a "Top5" data-dependent mode using the following parameters: resolution 17,500; AGC 1×10^5 ions; maximum IT 80 ms; 2.0 amu isolation window; normalized collision energy 35% ± 50%; underfill ratio 1.0%; Apex trigger 2–4 s, and dynamic exclusion 6 s. Source ionization parameters were: spray voltage, 3.8 kV; capillary temperature, 300 °C; and S-Lens level, 45.

Data Analysis

Differential analyses of the obese versus lean plasma were performed using Thermo Scientific™ SIEVE™ 2.1 software, which executes background subtraction, component detection, peak alignment, differential analysis (Figure 1). Statistical results, putative IDs, and pathways were generated through searching ChemSpider and KEGG™. Metabolites of interest were identified via MS/MS mass spectral database matching. The *.raw files were converted to mzXML format using ProteoWizard and also analyzed by XCMS Online [2] to compare the results.

FIGURE 1. Untargeted metabolomics workflow



Results

Challenges in Untargeted Metabolomics Study

- Complexity of biological samples
- Diversity of small molecule metabolites: polar and non-polar analytes
- Ionization requires both positive and negative ion
- Wide range of concentrations
- No universal method for chromatographic separation
- Multiple sources of variability
- Structure elucidation of unknowns is expensive: lack of synthetic standards

Preparing for the UHPLC-MS Data Acquisition

Prior to the real samples analysis, a solvent blank with internal standard (IS) is injected at the beginning to check the solvent and the LC-MS status. The injections of the real samples should be randomized in order to eliminate systematic bias. Triplicate injections of the pooled plasma are intermittently repeated throughout the whole batch to validate consistent performance of the overall system. The experimental design and run sequence are shown in Figure 2.

FIGURE 2. UHPLC/MS Experimental Design and Run Sequence. Left, schematic showing the vials and sample names. Right, detailed content and overall time for each step.

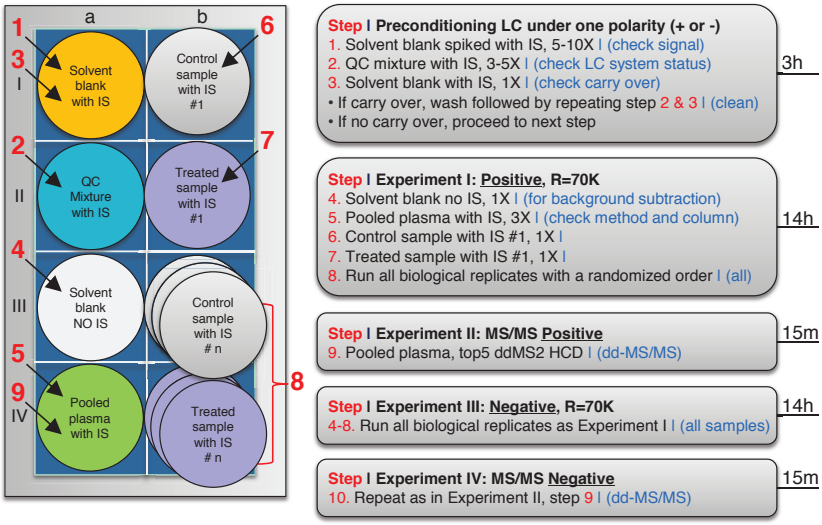
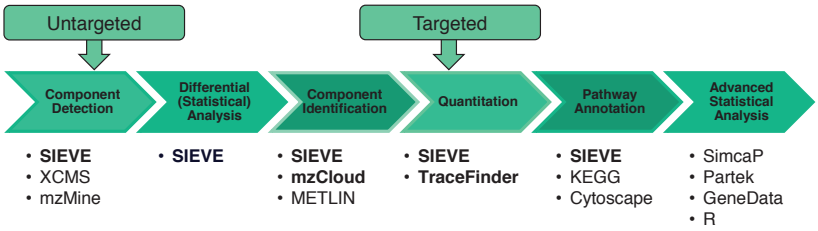


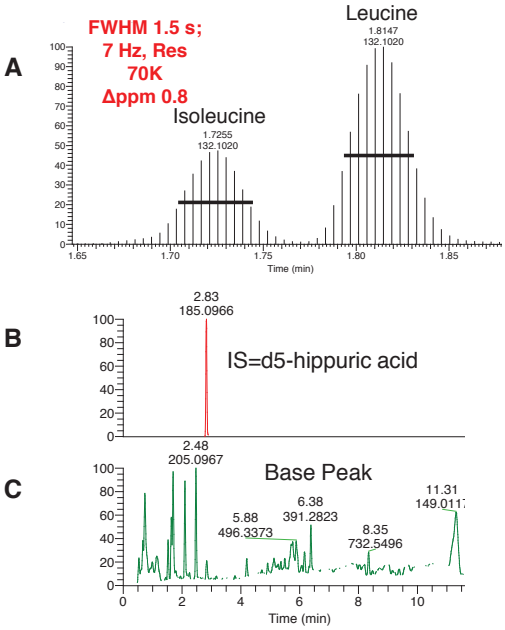
FIGURE 3. Metabolomics data analysis. SIEVE 2.1 software, Thermo Scientific™ TraceFinder™ 3.1 software, and mzCloud support untargeted and targeted metabolomics workflows.



UHPLC provides fast chromatography for high throughput analysis, the typical peak width is 4–6 seconds. Our method can baseline resolve Isoleucine and Leucine, generating peak width 1.2 s at FWHM. Refer to Figure 4. For such narrow peaks, Q Exactive mass spectrometer operating at 70,000 resolution acquires >15 point across the peak without losing sensitivity (A).

The QC of each run was performed by monitoring the intensity of IS, d5-hippuric acid (B), and the overall base peak (C). When all samples were finished, the selected ion chromatograms can be quickly viewed with RawMeat 2.1 (a free app with SIEVE 2.1 software). By inspecting the RT and intensity of the IS, runs with large retention time drift and bad injections can be excluded from the following data analysis.

FIGURE 4. Method validation, data quality control in metabolomics application



Comparing Differential Analysis Results from SIEVE 2.1 software and XCMS Online

The results from SIEVE 2.1 software are compared to XCMS Online, an academia-developed open-source software for metabolomics data analysis (Table 1) (2). As shown in Figure 5, the significant components (*p*-value <0.05, ratio > 2.0) identified by SIEVE 2.1 software and XCMS Online are similar, while the SIEVE software started with a much smaller (~20× less) number of components because of its capability to automatically remove solvent background, thus saving users time and labor in data review (Figure 6).

Metabolite Identification Guided by mzCloud

The mzCloud is the first cloud-based MS^n spectral trees library built for small molecule structural elucidation (3). It contains a standard reference database, an elucidated compound database with putative structures, and a database of virtual spectra. Q Exactive HCD MS/MS spectra of interested metabolites can be searched against m/zCloud. The matching compounds are scored and the fragment ions are annotated accordingly (Figure 7).

Table 1. Statics of component *m/z* 170.081 from SIEVE 2.1 software and XCMS Online

SIEVE 2.1 Software	XCMS Online
Ratio 5.0	Fold change 4.8
p-value 3.44e-4	p-value 3.4e-4
RT 2.03 min	RT 2.03 min
IDs: Norepinephrine Pyridoxine Oxidopiamine, etc.	**IDs:** Norepinephrine Pyridoxine (VB6) Hydroxydopamin, etc.

FIGURE 5. PCA by SIEVE 2.1 software (A), Numbers in SIEVE 2.1 software (B), Numbers in XCMS Online (C)

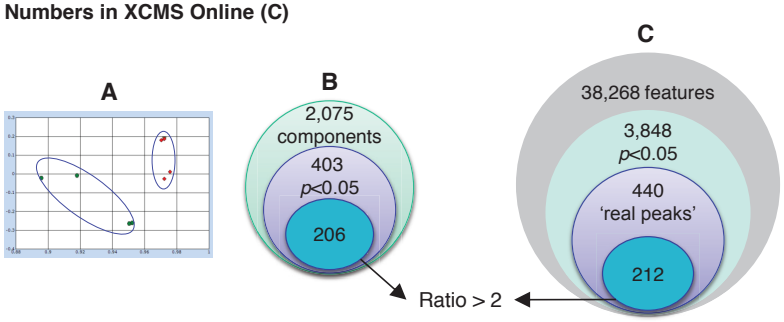


FIGURE 6. SIEVE 2.1 software output of component *m/z* 170.081, showing alignment across different samples (A), adducts grouping (B), peak integration (C), and trend intensity view (D)

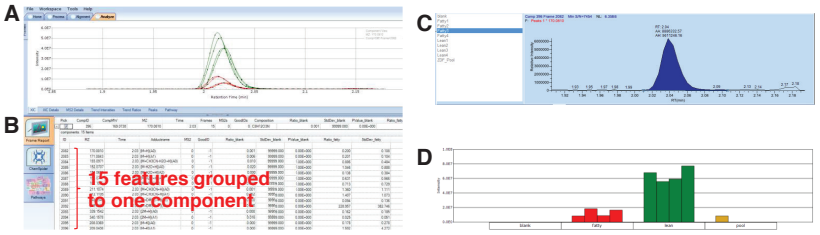
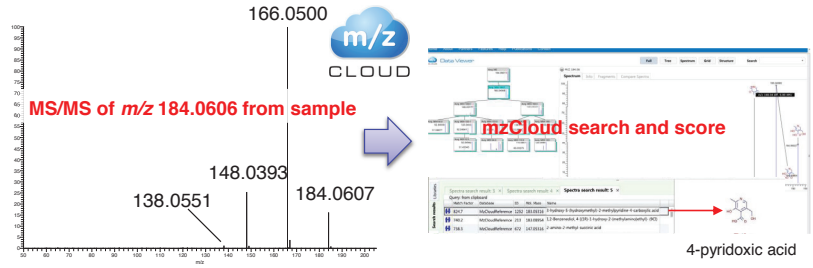


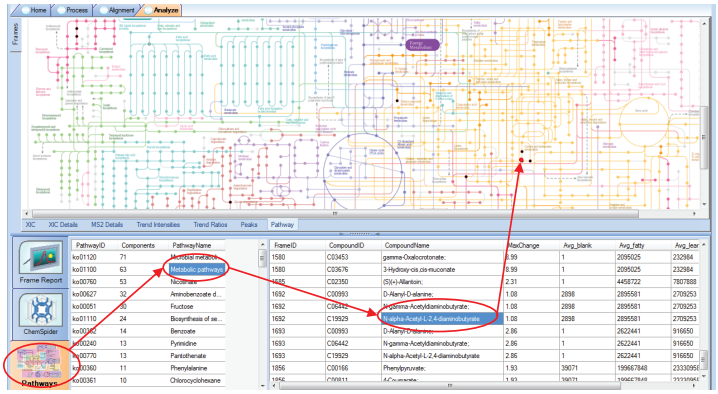
FIGURE 7. Searching MS/MS spectrum of component *m/z* 184.0607 in mzCloud. The entry of 4-pyridoxic acid matches the spectrum with the highest score.



An example of SIEVE 2.1 software with KEGG pathway visualization demonstrated with component *m/z* 161.0918 is shown in Figure 8. All matching metabolites are labeled with black dots.

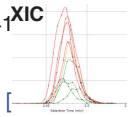
Table 2 shows some of the representative metabolites that are significantly changed. Up-regulated metabolites in fatty vs. lean rat plasma are shown in green while down-regulated metabolites are shown in blue (*p*-value <0.05, ratio >2).

FIGURE 8. SIEVE 2.1 software with KEGG pathway visualization



p-value: 3.76e-3
Ratio: 2.86
Composition: $C_6H_{12}N_2O_3$

ChemSpider ID: 389841
(2S)-4-Acetamido-2-aminobutanoic acid



Branched amino acids are known hallmarks of Type II Diabetes [

TABLE 2. Representative metabolites that are significantly changed

[M+H]+	Time	Fatty/Lean	P-value	Δppm	Formula	Name
854.5657	7.25	2.81	3.65E-02	4.0	C50H80NO8P	PC44:10
838.6292	8.84	2.48	6.45E-03	3.0	C48H88NO8P	PC 40:4
836.6133	8.55	4.54	2.22E-02	0.0	C46H88NO8P	PC38:2
812.6095	8.58	2.79	3.08E-02	4.0	C46H86NO8P	PC 38:3
788.6127	9.04	2.13	3.56E-02	3.0	C44H86NO8P	PC36:1
778.5351	7.20	4.15	3.38E-03	0.0	C42H78NO8P	PC34:3
768.5498	8.01	2.14	3.17E-04	4.0	C43H78NO8P	PC35:4
764.5196	7.44	3.71	6.37E-04	3.0	C43H74NO8P	PC35:6
754.5347	7.29	2.67	1.35E-02	3.0	C42H76NO8P	PC34:4
834.5936	8.10	2.83	7.13E-04	0.0	C44H83NO13	LacCer(d18:1/14:0)
836.6041	8.01	2.42	1.10E-03	4.0	C44H85NO13	LacCer(d18:0/14:0)
813.6790	9.72	2.40	3.84E-02	4.0	C47H93N2O6P	SM(d18:2/24:0)
675.5412	7.20	2.29	3.34E-02	3.0	C37H75N2O6P	SM(d16:1/16:0)
546.3525	5.84	2.33	9.23E-03	4.0	C28H52NO7P	LysoPC(20:3)
526.2910	5.69	2.42	1.67E-04	3.0	C27H44NO7P	LysoPE(22:6)
153.0521	1.79	2.46	1.56E-03	3.0	C4H12N2S2	Cystamine
230.0954	0.80	2.53	2.80E-02	1.0	C9H15N3O2S	Ergothioneine
245.0917	3.56	3.93	2.16E-03	1.0	C13H12N2O3	Haematopodin
172.1693	5.10	3.18	4.19E-02	1.0	C10H21NO	decanamide
302.3043	5.48	2.28	2.50E-02	3.0	C18H39NO2	Sphinganine
204.1228	3.75	3.10	1.83E-02	1.0	C9H17NO4	Acetylcarnitine
232.1534	2.26	4.49	1.99E-03	2.0	C11H21NO4	Butyryl-L-carnitine
344.2786	5.31	2.13	1.57E-02	2.0	C19H37NO4	Lauroylcarnitine
400.3409	5.53	2.10	3.49E-04	2.0	C23H45NO4	Palmitoyl-L-carnitine
466.3160	4.97	0.07	3.52E-03	0.0	C26H43NO6	Glycocholic Acid
357.2780	5.39	0.15	3.03E-03	2.0	C24H36O2	THA
355.2626	5.13	0.17	8.89E-03	1.0	C24H34O2	delta2-THA
170.0810	2.03	0.20	3.44E-04	0.0	C8H11NO3	Norepinephrine, 5-Hydroxydopamine
184.0968	2.69	0.23	2.70E-02	1.0	C9H13NO3	Normetanephrine; Methylnoradrenaline
212.1279	3.74	0.34	4.04E-03	1.0	C11H17NO3	Methoxamine
226.1433	4.20	0.29	8.10E-03	1.0	C12H19NO3	N-Methylmescaline; Terbutaline
161.0918	2.45	0.35	3.76E-03	2.0	C6H12N2O3	(2S)-4-Acetamido-2-aminobutanoic acid
224.0915	2.59	0.34	3.56E-02	1.0	C11H13NO4	Acetyl-L-tyrosine
146.1173	1.44	0.42	3.54E-02	1.0	C7H15NO2	DL-Aminoheptanoic acid

Conclusion

An efficient and robust workflow for untargeted metabolomics is presented here. The reliable high-resolution, accurate-mass (HR/AM) performance of the Q Exactive LC-MS system eliminates the need for technical replicates on biological samples. The superior S/N in Orbitrap data allows efficient data reduction in SIEVE 2.1 software, resulting in much reduced data analysis effort. KEGG pathway visualization allows quick access to biological pathway mapping. The MS^n spectral library mzCloud facilitates accurate compound identification.

References

1. Wang S.Y., Kuo, C.H., Tseng, Y. J. *Anal. Chem.*, **2013,** 85 (2), 1037–1046.
2. Scripps Center For Metabolomics Home Page. https://xcmsonline.scripps.edu/index.php
3. m/zCloud Home Page. https://www.mzcloud.org/home.aspx
4. Muoio, D.M.; Newgard, C.B. Nature Rev. *Mol. Cell. Biol.* **2008,** 9, 193–205.

www.thermoscientific.com

Thermo Fisher Scientific, San Jose, CA USA is ISO 9001:2008 Certified.

Africa-Other +27 11 570 1840
Australia +61 3 9757 4300
Austria +43 1 333 50 34 0
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China +86 10 8419 3588
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland/Norway/Sweden +46 8 556 468 00
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9434
Italy +39 02 950 591
Japan +81 45 453 9100
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Russia/CIS +43 1 333 50 34 0
South Africa +27 11 570 1840
Spain +34 914 845 965
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC

Part of Thermo Fisher Scientific

ASMS13_T382_JWang_E 07/13S