

A complete workflow for improved untargeted metabolome annotation and identification using ultra high-resolution accurate mass and LC-MSⁿ Orbitrap-based mass spectrometry

David A. Peake, Reiko Kiyonami, Ioanna Ntai, Amanda Souza and Ralf Tautenhahn, Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose, CA, USA, 95134

ABSTRACT

Purpose: Demonstrate confidence in metabolite annotation from ultra-high resolution LC-MS and data dependent MS² data using a combined *global semi-targeted* (GST) metabolomics workflow.

Methods: NIST SRM1950 human plasma was analyzed using data dependent LC-MS/MS analysis and the AcquireX data acquisition strategy. A list of highly-confident annotations was prepared and 58 reference compounds were targeted for confirmation of metabolite annotations.

Results: Comparing the results from standard mixtures, SRM1950 plasma and spiked plasma extracts provided unequivocal evidence that 56 out of the 58 compounds were identified.

INTRODUCTION

Annotation of hundreds of unknown metabolites from human plasma is one of the most difficult challenges faced by metabolomics scientists. Automated annotation must incorporate accurate mass, isotope pattern and isotopic fine structure to confirm accurate elemental formula for database searching. Unknown MS² and MS³ spectra should be searched for identity or similarity matches against a high-quality MSⁿ mass spectral library and the identification level reported for each metabolite (Table 1) based on the consensus of the available analytical measures¹. Unequivocal identification of annotated metabolites requires confirmation by spiking authentic standards into the plasma and matching the retention time and fragmentation patterns of the unknown MS/MS spectra.

Preliminary analysis of SRM1950 human plasma using the AcquireX acquisition strategy showed that more than 5000 compounds were detected in the positive ion reversed-phase analysis with 82% of MS² spectra being acquired on the preferred protonated ions. Using the elemental composition of molecular species determined from ultra-high resolution MS data 2699 ChemSpider database hits were found. MS² product ion spectra were searched against the mzCloud MSⁿ library (>17,000 compounds and 4.75 million spectra) providing and 349 identity matches to known metabolites. Similarity results (1272 molecules) were prioritized using the mzLogic algorithm (rank orders annotations by mapping potential structures to known fragment ions in mzCloud).

An automated GST workflow for combined untargeted and semi-targeted sample analysis and data processing pipeline (Figure 1) results in improved metabolome coverage and highly confident annotations as confirmed here using 58 metabolite standards spiked into SRM1950 plasma extracts.

MATERIALS AND METHODS

Sample Preparation

SRM1950 human plasma was purchased from NIST (<https://srm1950.nist.gov>). Human metabolite reference standards were purchased from MetaSci (<https://www.metasci.ca>). Lipid reference standards were obtained from Avanti Polar Lipids, Inc. (<https://www.avantilipids.com>), Naproxen from Sigma-Aldrich (<https://www.sigmaaldrich.com>) and unlabeled carnitine and acyl carnitines standards (NSK-B-US) were obtained from Cambridge Isotope Laboratories, Inc. (<https://www.isotope.com>).

Stock solutions (1.00 mg/mL) were prepared in water, methanol, 1:1 methanol-water or 1:1 methanol-chloroform. Some lipid reference standards were supplied as 10mg/mL solutions in chloroform. Ten different mixtures containing 5-10 compounds each were prepared at 10µg/mL in 1:1 methanol-water or methanol. Mixtures (Mix01-Mix10) were diluted 1:10 in water to give 1.0 µg/mL per compound for LC-MS² analysis. Aliquots of SRM1950 human plasma (100µL) were precipitated (21,000g at 4°C for 20min) using a 3:1 (v/v) methanol to sample ratio, and 300µL of the supernatant was evaporated under vacuum at 7°C overnight. Each plasma extract was reconstituted in 90µL of 95% water and 5% methanol, vortexed and the extracts were combined and centrifuged. Aliquots of 90µL plasma extract were spiked with 10µL of standard mixtures (M01-M10). Four aliquots of unspiked SRM1950 plasma were prepared as QC samples (P1-P4).

LC-MS and data dependent MS² analysis

SRM1950 plasma extracts, standard mixtures and spiked plasma samples were injected (2 µL) and separated on a Thermo Scientific™ Hypersil GOLD™ C18 column (2.1x150mm, 1.9µm, 45°C) using the gradient shown in Table 2. LC-MS analyses were performed using a Thermo Scientific™ Vanquish™ UHPLC system coupled with a Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer. Mass spectral data were acquired separately in positive and negative ion modes at 120,000 MS resolution (FWHM @m/z 200) using internal mass calibration and data dependent HCD MS² (30,000 res.) using stepped collision energy for metabolite characterization (Tables 2, 3). An accurate mass inclusion list was employed to ensure that targeted reference compounds (Table 4) were prioritized for MS² analysis; untargeted metabolites were acquired during the remaining time.

RESULTS

Data Analysis

Analysis of SRM1950 plasma spiked with metabolite standards was conducted to validate confidence in unknown annotations. Ultra-high resolution MS data were analyzed with Thermo Scientific™ Compound Discoverer™ version 3.1 software by first annotating unknown metabolites based on a formula search of selected ChemSpider databases (BioCyc, DrugBank, HMDB, KEGG, LipidMaps, PubMed) and Metabolika pathways. From the same dataset a targeted mass list generated from the reference standards in Table 3 was used to perform targeted analysis (Figure 1) for high confidence annotations determined in a preliminary analysis of SRM1950 plasma. All 58 reference compounds were identified by a formula search, with an average mass measurement error of < 0.1 ppm, and all but one compound (an isomer of Levulinic acid) matched the retention time of the standards (Table 5).

A fragmentation search was conducted of unknown MS² spectra against the MSⁿ mzCloud library (<https://mzcloud.org>) and a targeted search using the mzVault library (constructed from MS² spectra obtained from the reference compounds). The combination of mzCloud and mzVault searches (MS² match reported in Table 5 is the highest match score from mzCloud or mzVault for plasma/spiked plasma) unequivocally identified 56 out of 58 compounds in the SRM1950 and spiked plasma extracts. Stearoyl ceramide was not identified due to the very low abundance of the M+H ion.

Figure 1. Data analysis workflow for untargeted and semi-targeted metabolite identification.

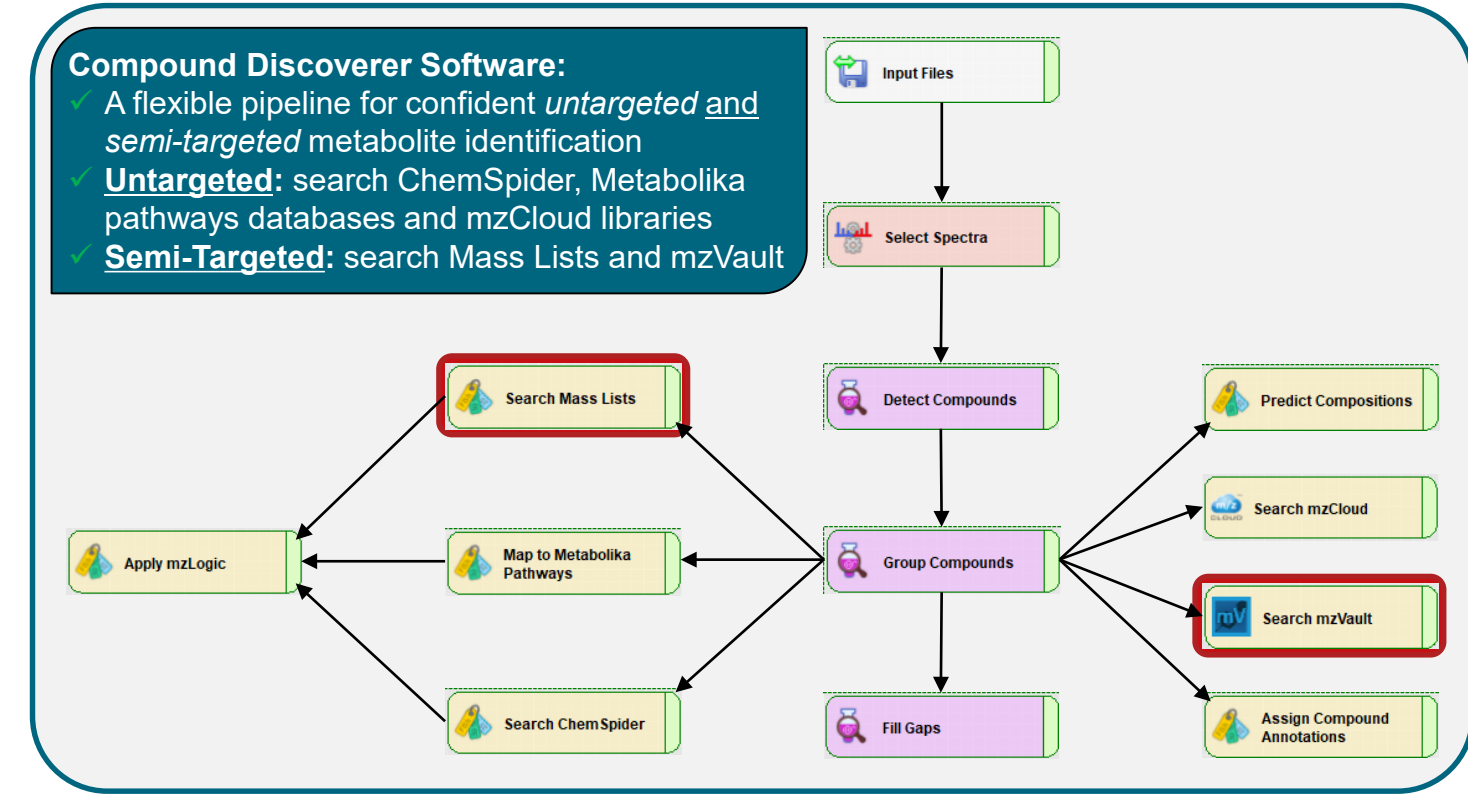


Table 1. Unknown Annotation Level Definitions

Level	Database m/z match	HR-MS Composition	Library MS ² match	Standard Rt match	Standard MS ² match
0	Structure confirmed by 2D-NMR, HRMS and other spectroscopic methods				
1	✓	✓	✓	✓	✓
2	✓	✓	✓		
3	✓	✓			
4	Feature; retention time and m/z, with several possible elemental compositions				

Table 2. UHPLC Gradient Method

Time min	Flow mL/min	%A = H ₂ O	%B = MeOH
-4.00	0.300	100	0
0.00	0.300	100	0
8.00	0.300	50	50
9.00	0.300	2	98
14.0	0.300	2	98
14.1	0.300	100	0
15.0	0.300	100	0

Table 3. Orbitrap ID-X MS Conditions

Ion Source	Orbitrap MS
ESI pos. = 3500V	Method 1 - LC-MS
ESI neg. = 3000V	MS R = 120K
Sheath gas = 40	Max IT = 50msec
Aux gas = 8	AGC = 1E5
Sweep = 1	Method 2 - LC-MS ²
Vaporizer = 320°C	MS R = 60K, MS ² R = 30K
IT temp = 275°C	IW = 1.5 Da
RF lens = 35%	AGC = 5E4
Int. Cal. = Easy-IC	Max IT = 54msec
Method t = 15 min	CE = 20, 35, 50%

Table 4. Reference standards for confirming metabolite annotation

#	Compound name	HMDB ID	Location Catalog #	Formula	µg/mL	Std Mix
1	Choline	HMDB0000097	MetaSci 3J3	C ₅ H ₁₃ N O	1.00	M10
2	Serine	HMDB0000187	MetaSci 3E5	C ₃ H ₇ N O ₃	1.00	M09
3	Creatinine	HMDB0000562	MetaSci 3C3	C ₄ H ₇ N ₃ O	1.00	M08
4	Proline	HMDB0000162	MetaSci 3F6	C ₅ H ₉ N O ₂	1.00	M09
5	Levulinic acid	HMDB0000720	MetaSci 3H2	C ₆ H ₈ O ₃	1.00	M06
6	Betaine	HMDB0000043	MetaSci 3A1	C ₅ H ₁₁ N O ₂	1.00	M07
7	Butyric acid, 2-OH-3-Me	HMDB0000407	MetaSci 4G3	C ₆ H ₁₀ O ₃	1.00	M02,06
8	Pyroglutamic acid	HMDB0000267	MetaSci 5J10	C ₅ H ₇ N O ₃	1.00	M03
9	Isocaproic acid, α-keto	HMDB0000695	MetaSci 10D7	C ₆ H ₁₀ O ₃	1.00	M06
10	Valeric acid, 3-methyl-2-oxo	HMDB0000491	MetaSci 5E5	C ₆ H ₁₀ O ₃	1.00	M03
11	Proline, 4-hydroxy	HMDB0000725	MetaSci 6G9	C ₅ H ₉ N O ₃	1.00	M04
12	Salicylic acid	HMDB0015470	MetaSci 5I5	C ₇ H ₆ O ₃	1.00	M03
13	Glutamine	HMDB0000641	MetaSci 3A2	C ₆ H ₁₂ N O ₃	1.00	M08
14	Lysine	HMDB0000182	MetaSci 4D1	C ₆ H ₁₄ N ₂ O ₂	1.00	M02
15	Glutamic acid	HMDB0000148	MetaSci 3B10	C ₅ H ₉ N O ₃	1.00	M08
16	Methionine	HMDB0000696	MetaSci 6F9	C ₅ H ₁₁ N O ₂ S	1.00	M04
17	Acetaminophen	HMDB0001859	MetaSci 5I6	C ₈ H ₉ N O ₂	1.00	M03
18	Genetic acid	HMDB0000152	MetaSci 3G7	C ₇ H ₆ O ₄	1.00	M06
19	Histidine	HMDB0000177	MetaSci 3G3	C ₆ H ₉ N ₃ O ₂	1.00	M09
20	Carnitine	HMDB0000062	CIL NSK-B-US	C ₇ H ₁₅ N O ₃	24.50	M01,05
21	Phenylalanine	HMDB0000159	MetaSci 3E6	C ₉ H ₉ N O ₂	1.00	M09
22	Uric acid	HMDB0000289	MetaSci 3D7	C ₅ H ₄ N ₄ O ₃	0.50	M06
23	Arginine	HMDB0000517	MetaSci 4C10	C ₆ H ₁₄ N ₄ O ₃	1.00	M02
24	Indole-3-acetic acid	HMDB0000197	MetaSci 3A6	C ₁₀ H ₉ N O ₂	1.00	M08
25	Hippuric acid	HMDB0000714	MetaSci 4A9	C ₉ H ₉ N O ₃	1.00	M02,06
26	Paraxanthine	HMDB0001860	MetaSci 4B2	C ₇ H ₈ N ₄ O ₂	1.00	M02
27	Theobromine	HMDB0002825	MetaSci 7C9	C ₇ H ₈ N ₂ O ₂	0.50	M05
28	Theophylline	HMDB0001889	MetaSci 1B9	C ₇ H ₈ N ₄ O ₂	1.00	M07
29	Tyrosine	HMDB0000158	MetaSci 3A7	C ₉ H ₉ N O ₃	1.00	M08
30	Azelaic acid	HMDB0000784	MetaSci 3G10	C ₉ H ₁₆ O ₄	1.00	M09
31	3-Indolepropionic acid	HMDB0002302	MetaSci 10A7	C ₁₁ H ₁₁ N O ₂	1.00	M07
32	Caffeine	HMDB0001847	MetaSci 4H9	C ₈ H ₁₀ N ₄ O ₂	1.00	M03
33	Hippuric acid, 2-hydroxy	HMDB0000840	MetaSci 9F4	C ₉ H ₉ N O ₄	1.00	M06
34	Carnitine, 2:0	HMDB0000201	CIL NSK-B-US	C ₉ H ₁₇ N O ₄	7.72	M01,05
35	Tryptophan	HMDB0000929	MetaSci 3H8	C ₁₁ H ₁₂ N ₂ O ₂	1.00	M04
36	Carnitine, 3:0	HMDB0062514	CIL NSK-B-US	C ₁₀ H ₁₉ N O ₄	1.65	M01
37	Naproxen	HMDB0001923	Sigma	C ₁₄ H ₁₄ O ₃	1.00	M05
38	Carnitine, 4:0	HMDB0002013	CIL NSK-B-US	C ₁₁ H ₂₁ N O ₄	1.76	M01
39	Carnitine, 5:0 (Isovaleryl)	HMDB0000688	CIL NSK-B-US	C ₁₂ H ₂₃ N O ₄	1.86	M01
40	Atenolol	HMDB0001924	MetaSci 7C4	C ₁₄ H ₂₂ N ₂ O ₃	1.00	M05
41	Piperine	HMDB0029377	MetaSci 9G5	C ₁₇ H ₁₉ N O ₃	1.00	M07
42	Carnitine, 8:0	HMDB0000791	CIL NSK-B-US	C ₁₅ H ₂₉ N O ₄	2.18	M01
43	Sphingosine, d18:1	HMDB0000252	Avanti 860490	C ₁₈ H ₃₇ N O ₂	1.00	M10
44	Cortisone	HMDB0002802	MetaSci 9E5	C ₂₁ H ₂₈ O ₅	1.00	M05
45	Cortisol (Hydrocortisone)	HMDB0014879	MetaSci 6E6	C ₂₁ H ₃₀ O ₅	1.00	M04
46	Carnitine, 14:0	HMDB0005066	CIL NSK-B-US	C ₂₁ H ₄₁ N O ₄	2.82	M01
47	Cholesterol	HMDB0000067	Avanti 700100	C ₂₇ H ₄₆ O	1.00	M10
48	Carnitine, 16:0	HMDB0000222	CIL NSK-B-US	C ₂₃ H ₄₅ N O ₄	6.07	M01
49	Carnitine, 18:1(9Z)	HMDB0005065	Avanti 870852	C ₂₅ H ₄₇ N O ₄	10.00	M10
50	Glycoursodeoxycholic acid	HMDB0000708	MetaSci 9B1	C ₂₆ H ₄₃ N O ₅	1.00	M07
51	Glycochenodeoxycholic acid	HMDB0000637	MetaSci 6D2	C ₂₆ H ₄₃ N O ₅	1.00	M04
52	Glycocholic acid	HMDB0000138	MetaSci 3I10	C ₂₆ H ₄₃ N O ₅	1.00	M06
53	Lyso PC, 1-18:1(9Z)	HMDB0002815	Avanti 845875	C ₂₆ H ₅₂ N O ₇ P	10.00	M06
54	Lyso PC, 1-18:0	HMDB0010384	Avanti 855775	C ₂₆ H ₅₄ N O ₇ P	10.00	M10
55	Ceramide, d18:1/16:0	HMDB0000790	Avanti 860516	C ₃₄ H ₆₇ N O ₃	1.00	M10
56	Ceramide, d18:1/18:0	HMDB0000829	Avanti 860518	C ₃₆ H ₇₁ N O ₃	1.00	M10
57	Bilirubin	HMDB0000054	MetaSci 4A8	C ₃₃ H ₃₆ N ₄ O ₆	1.00	M10
58	Sphingomyelin, d18:1/16:0	HMDB0061712	Avanti 860584	C ₃₉ H ₇₈ N ₂ O ₆ P	1.00	M10

Table 5. Summary of compounds identified in SRM1950 human plasma

ID level	Compound name	Rt, min	Rt Δ, min	Formula confirmed	Ion	m/z, plasma	Δ, ppm	MS ² match	ID conf
1	Choline	1.13	0.01	C ₅ H ₁₃ N O	Pos	104.1069	-0.9	90.9	✓
1	Serine	1.09	0.02	C ₃ H ₇ N O ₃	Pos	106.0499	0.3	84.3	✓
1	Creatinine	1.18	0.02	C ₄ H ₇ N ₃ O	Pos	114.0661	-0.8	90.3	✓
1	Proline	1.27	0.00	C ₅ H ₉ N O ₂	Pos	116.0705	-0.9	89.0	✓
3	Levulinic acid	3.76	0.30	C ₆ H ₈ O ₃	Neg	115.0401	0.3	84.6	✗
1	Betaine	1.17	0.00	C ₅ H ₁₁ N O ₂	Pos	118.0862	-0.5	90.4	✓
1	Butyric acid 2-OH 3-Me	5.13	0.03	C ₆ H ₁₀ O ₃	Neg	117.0557	-0.1	92.9	✓
1	Pyroglutamic acid	2.40	0.02	C ₅ H ₇ N O ₃	Pos	130.0498	-0.5	89.2	✓
1	Valeric acid 3-Me-2-Oxo	5.93	-0.02	C ₆ H ₁₀ O ₃	Neg	129.0557	-0.1	85.1	✓
1	α-Ketoisocaproic acid	6.35	-0.03	C ₆ H ₁₀ O ₃	Neg	129.0557	-0.1	88.0	✓
1	Proline, 4-OH	1.12	0.01	C ₅ H ₉ N O ₃	Pos	132.0655	-0.1	90.2	✓
1	Salicylic acid	9.15	0.00	C ₇ H ₆ O ₃	Neg	137.0244	-0.1	95.9	✓
1	Glutamine	1.11	0.02	C ₆ H ₁₂ N ₂ O ₃	Pos	147.0763	-0.8	87.4	✓
1	Lysine	0.97	-0.01	C ₆ H ₁₄ N ₂ O ₂	Pos	147.1128	0.0	89.7	✓
1	Glutamic acid	1.13	0.01	C ₅ H ₉ N O ₃	Neg	146.0459	0.1	88.5	✓
1	Methionine	1.90	0.01	C ₅ H ₁₁ N O ₂ S	Pos	150.0583	-0.2	91.8	✓
1	Acetaminophen	4.26	0.00	C ₈ H ₉ N O ₂	Pos	152.0706	0.0	94.9	✓
1	Genetic acid	5.33	-0.02	C ₇ H ₆ O ₄	Neg	153.0193	-0.2	90.0	✓
1	Histidine	1.09	0.09	C ₆ H ₉ N ₃ O ₂	Pos	156.0768	0.3	94.1	✓
1	Carnitine	1.15	0.01	C ₇ H ₁₅ N O ₃	Pos	162.1123	-1.0	98.1	✓
1	Phenylalanine	3.91	-0.01	C ₉ H ₉ N O ₂	Pos	166.0862	-0.3	95.4	✓
1	Uric acid	1.80	0.00	C ₅ H ₄ N ₄ O ₃	Neg	167.0210	-0.4	91.4	✓
1	Arginine	1.09	0.07	C ₆ H ₁₄ N ₄ O ₃	Pos	175.1190	0.3	88.6	✓
1	Indole-3-acetic acid	8.53	0.00	C ₁₀ H ₉ N O ₂	Pos	176.0706	0.0	92.5	✓
1	Hippuric acid	5.85	-0.03	C ₉ H ₉ N O ₃	Neg	178.0510	-0.1	89.2	✓
1	Paraxanthine	5.03	0.00	C ₇ H ₈ N ₄ O ₂	Pos	181.0720	0.0	93.4	✓
1	Theobromine	4.35	0.01	C ₇ H ₈ N ₂ O ₂	Pos	181.0720	0.0	95.7	✓
1	Theophylline	5.21	0.01	C ₇ H ₈ N ₄ O ₂	Pos				