

Supporting Information

Phenotypic mapping of metabolic profiles using self-organizing maps of high-dimensional mass spectrometry data

Cody R. Goodwin,^{†‡} Stacy D. Sherrod,^{‡||} Christina C. Marasco,^{‡§} Brian O. Bachmann,[†] Nicole Schramm-Sapyta,[○] John P. Wikswo,^{‡||§*} John A. McLean^{†‡*}

[†] Dept. of Chemistry and Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, TN

[‡] Vanderbilt Institute for Integrative Biosystems Research and Education, Vanderbilt University, Nashville, TN

^{||} Physics and Astronomy, Vanderbilt University, Nashville, TN

[§] Dept. of Biomedical Engineering, Vanderbilt University, Nashville, TN

[○]Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN

^{*}Dept. of Psychiatry and Behavioral Sciences, Duke University, Durham, NC

Abstract

A metabolic system is comprised of inherently interconnected metabolic precursor, intermediates, and products. The analysis of untargeted metabolomics data has conventionally been performed through the use comparative statistics or multivariate statistical analysis-based approaches; however, each falls short in representing the related nature of metabolic perturbations. Herein, we describe a complementary method for the analysis of large metabolite inventories using a data-driven approach based upon a self-organizing map algorithm. This workflow allows for the unsupervised clustering, and subsequent prioritization, of correlated features through Gestalt comparisons of metabolic heat maps. We describe this methodology in detail, including a comparison to conventional metabolomics approaches, and demonstrate the application of this method to the analysis of the metabolic repercussions of prolonged cocaine exposure upon rat sera profiles.

Table of Contents

- Figure S1.** *Density map for Figure 2(b)*
- Figure S2.** *Orthogonal partial least square-discriminant analysis comparing cocaine-experienced (-1) to cocaine-naïve (+1) rat sera profiles.*
- Table S1.** *Statistical summary for region “a”.*
- Table S2.** *Statistical summary for region “b”.*
- Table S3.** *Statistical summary for region “c”.*
- Table S4.** *Statistical summary for region “d”.*
- Table S5.** *Statistical summary for region “e”.*
- Table S6.** *Statistical summary for region “f”.*
- Table S7.** *Features within region extracted from the S-plot in Figure 4(b).*
- Table S8.** *Feature list with corresponding mass accuracies.*
- Figure S3-22.** *Spectra corresponding to putative metabolite identifications.*

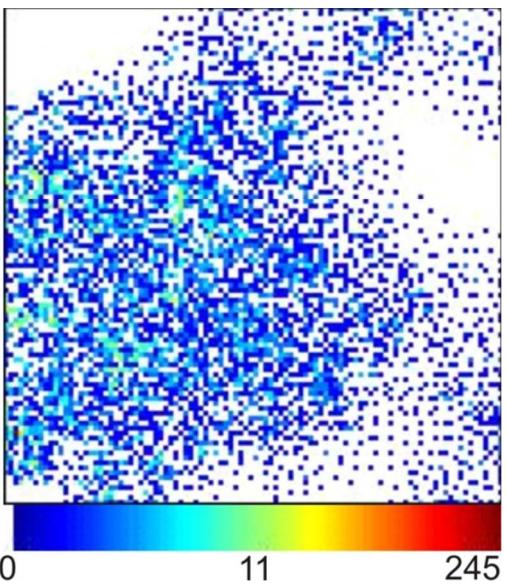


Figure S1. Density map for Figure 2(b). Color indicates the density of features per node. These features are then summed to determine the node intensity.

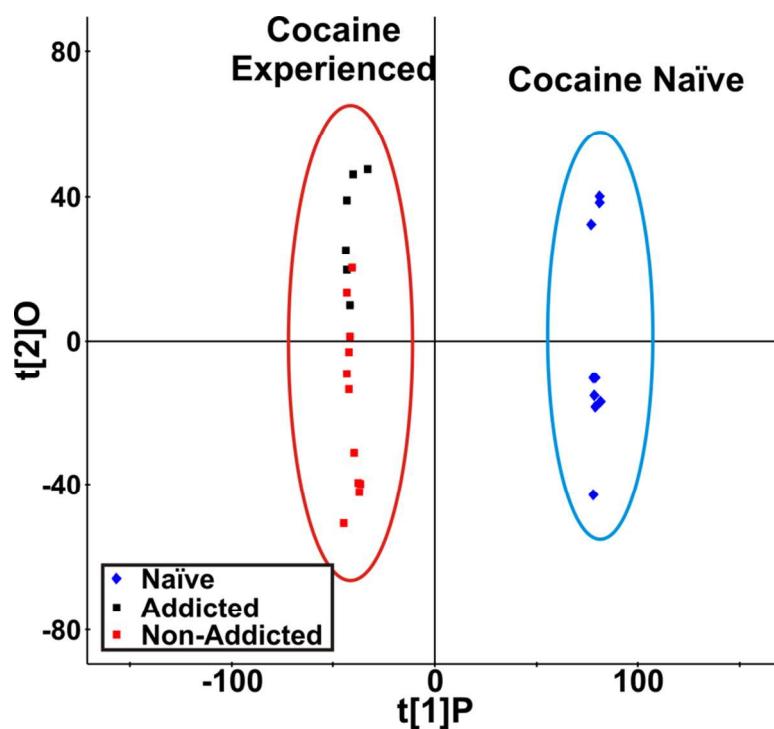


Figure S2. Orthogonal partial least square-discriminant analysis comparing cocaine-experienced (-1) to cocaine-naïve (+1) rat sera profiles.

Supplemental Information

Table S1. Region “a” statistical summary.

Blue	Non-addicted averages	Addicted averages	Naïve averages	Experienced averages	
	6200.54 10163.6 8724.59 7838.07 9355.55 5919.37 8781.63 5323.97 6343.24 6202.28	8959.59 7979.64 11042.4 8700.31 8140.3 8945.93	3002.5 2892.6 2606.9 2218.3 2057.7 2298.2 2461 2551.5 2704.5	6200.54 10163.6 8724.59 7838.07 9355.55 5919.37 8781.63 5323.97 6343.24 6202.28 8959.59 7979.64 11042.4 8700.31 8140.3 8945.93	
Kruskal-Wallace One-way ANOVA					
SUMMARY					
Groups	Count	Sum	Average	Variance	
Column 1	10	74852.9	7485.286	2859614	
Column 2	6	53768.1	8961.355	1207373	
Column 3	9	22793.3	2532.586	96271.43	
Column 4	16	128621	8038.812	2662921	
ANOVA					
H=22.02	df=3	Non-addicted	Addicted	Naïve	Experienced
p<0.0001	Mean Ranks for Sample	22.7	30.2	5	25.5
Post hoc test: Bonferroni-Holm					
Group 1	Group 2	Critical	P	Significant?	
Addicted average	Naïve average	0.008333	3.25E-10	Yes	
Naïve average	Experienced average	0.01	8.69E-10	Yes	
Non-addicted average	Naïve average	0.0125	1.28E-07	Yes	
Non-addicted average	Addicted average	0.016667	0.078596	No	
Addicted average	Experienced average	0.025	0.218321	No	
Non-addicted average	Experienced average	0.05	0.414697	No	
Test for Equal Variance					
F	DFn	DFd	P		
6.716522861	3	37	0.000987	FAIL-equal variance cannot be assumed	

Supplemental Information

Table S2. Region “b” statistical summary.

Gray	Non-addicted averages	Addicted averages	Naïve averages	Experienced averages
	2170.11	1016.38	245.32	2170.11
	2101.32	692.3	327.03	2101.32
	1542.41	782.36	331.68	1542.41
	843.41	879.43	400.34	843.41
	748.87	879.85	338.84	748.87
	814.03	1082.14	439.14	814.03
	737.19		401.79	737.19
	704.62		476.22	704.62
	593.51		515.46	593.51
	620.41			620.41
				1016.38
				692.3
				782.36
				879.43
				879.85
				1082.14
ANOVA: Single Factor				
SUMMARY				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Column 1	10	10875.9	1087.588	376374.2
Column 2	6	5332.46	888.7433	20753.32
Column 3	9	3475.82	386.2022	7077.63
Column 4	16	16208.3	1013.021	242627.1
ANOVA				
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Between Groups	2908170	3	969389.9	4.990485
Within Groups	7187162	37	194247.6	
Total	10095332	40		
Post hoc test: Bonferroni-Holm				
Group 1	Group 2	Critical	P	Significant?
Addicted average	Naïve average	0.008333	1.02E-06	Yes
Naïve average	Experienced average	0.01	0.001037	Yes
Non-addicted average	Naïve average	0.0125	0.003471	Yes
Non-addicted average	Addicted average	0.016667	0.453466	No
Addicted average	Experienced average	0.025	0.555188	No
Non-addicted average	Experienced average	0.05	0.735433	No
Test for Equal Variance				
F	DFn	DFd	P	
2.31283416	3	37	0.091936	PASS - equal variance may be assumed (p > 0.05).

Supplemental Information

Table S3. Region “c” statistical summary.

Green	Non-addicted averages	Addicted averages	Naïve averages	Experienced averages
	1878.46	2836.05	504.97	1878.46
	1469.38	3042.55	448.33	1469.38
	1993.71	2200.57	345.58	1993.71
	2543.64	1483.77	251.39	2543.64
	1045.84	1382.98	293.38	1045.84
	8985.16	726.63	296.93	8985.16
	6632.13		263.09	6632.13
	7836.09		405.93	7836.09
	7118.23		333.42	7118.23
	6014.89			6014.89
				2836.05
				3042.55
				2200.57
				1483.77
				1382.98
				726.63
Kruskal-Wallace one-way ANOVA				
SUMMARY				
Groups	Count	Sum	Average	Variance
Column 1	10	45517.5	4551.753	9223617
Column 2	6	11672.6	1945.425	815385.2
Column 3	9	3143.02	349.2244	7550.15
Column 4	16	57190.1	3574.38	7504202
ANOVA				
H=21.83	df=3	Non-addicted	Addicted	Naïve
p<0.0001	Mean Ranks for Sample	28.1	21.2	5
Post hoc test: Bonferroni-Holm				
Group 1	Group 2	Critical	P	Significant?
Addicted average	Naïve average	0.008333	0.000128	Yes
Naïve average	Experienced average	0.01	0.000688	Yes
Non-addicted average	Naïve average	0.0125	0.001938	Yes
Non-addicted average	Addicted average	0.016667	0.062533	No
Addicted average	Experienced average	0.025	0.174181	No
Non-addicted average	Experienced average	0.05	0.404085	No
Test for Equal Variance				
F	DFn	DFd	P	
11.32572354	3	37	2.05E-05	FAIL-equal variance cannot be assumed

Supplemental Information

Table S4. Region “d” statistical summary.

Purple	Non-addicted averages	Addicted averages	Naïve averages	Experienced averages
	1961.77 1714.38 1701.73 1803.42 3796.06 964.43 841.67 1642.09 1295.13 1657.07	1068.04 1373.17 1044.29 1004.43 1516.42 1350.24	564.75 459.78 582.72 388.91 345.29 536.04 772.68 675.7 673.84	1961.77 1714.38 1701.73 1803.42 3796.06 964.43 841.67 1642.09 1295.13 1657.07 1068.04 1373.17 1044.29 1004.43 1516.42 1350.24
ANOVA: Single Factor				
SUMMARY				
Groups	Count	Sum	Average	Variance
Column 1	10	17377.8	1737.775	656170.8
Column 2	6	7356.59	1226.098	45700.29
Column 3	9	4999.71	555.5233	19740.94
Column 4	16	24734.3	1545.896	474389.1
ANOVA				
Source of Variation	SS	df	MS	F
Between Groups	7871591	3	2623864	7.240781
Within Groups	13407803	37	362373.1	
Total	21279394	40		
Post hoc test: Bonferroni-Holm				
Group 1		Group 2	Critical	P
Addicted average		Naïve average	0.008333	5.35E-06
Naïve average		Experienced average	0.01	0.00032
Non-addicted average		Naïve average	0.0125	0.000476
Non-addicted average		Addicted average	0.016667	0.156613
Addicted average		Experienced average	0.025	0.283378
Non-addicted average		Experienced average	0.05	0.524272
Test for Equal Variance				
F	DFn	DFd	P	
2.801616821	3	37	0.053266	PASS - equal variance may be assumed (p > 0.05).

Supplemental Information

Table S5. Region “e” statistical summary.

Red	Non-addicted averages	Addicted averages	Naïve averages	Experienced averages
	411.94	655.04	174.65	411.94
	386.85	238.99	176.65	386.85
	1046.58	487.88	209.55	1046.58
	889.8	897.8	155.11	889.8
	1555.52	652.19	151.64	1555.52
	494.63	1134.66	158.86	494.63
	581.41		384.33	581.41
	878.57		344.2	878.57
	417.64		345.88	417.64
	723.63			723.63
				655.04
				238.99
				487.88
				897.8
				652.19
				1134.66
ANOVA: Single Factor				
SUMMARY	Count	Sum	Average	Variance
Groups				
Column 1	10	7386.57	738.657	135840.8
Column 2	6	4066.56	677.76	97383.75
Column 3	9	2100.87	233.43	9164.929
Column 4	16	11453.1	715.8206	114892.8
ANOVA				
H=18.9	df=3	Non-addicted	Addicted	Naïve
p<0.0001	Mean Ranks for Sample	25.5	25	5.7
Post hoc test: Bonferroni-Holm				
Group 1	Group 2	Critical	P	Significant?
Addicted average	Naïve average	0.008333	0.000395	Yes
Naïve average	Experienced average	0.01	0.000963	Yes
Non-addicted average	Naïve average	0.0125	0.001348	Yes
Non-addicted average	Addicted average	0.016667	0.740765	No
Addicted average	Experienced average	0.025	0.813415	No
Non-addicted average	Experienced average	0.05	0.872901	No
Test for Equal Variance				
F	DFn	DFd	P	
5.963301904	3	37	0.002013	FAIL - equal variance cannot be assumed.

Supplemental Information

Table S6. Region “f” statistical summary.

White	Non-addicted averages	Addicted averages	Naïve averages	Experienced averages
	1950.42 786.35 915.45 1068.64 2538.34 1199.5 694.17 1263.33 1087.39 728.13	1520.41 2827.55 2014.75 2372.6 3169.89 2626.97	470.8 511.69 446.62 359.06 414.26 294.02 408.2 300.77 318.73	1950.42 786.35 915.45 1068.64 2538.34 1199.5 694.17 1263.33 1087.39 728.13 1520.41 2827.55 2014.75 2372.6 3169.89 2626.97
Kruskal-Wallis One-way ANOVA				
SUMMARY				
Groups	Count	Sum	Average	Variance
Column 1	10	12231.7	1223.172	345940.3
Column 2	6	14532.2	2422.028	349396.1
Column 3	9	3524.15	391.5722	6081.414
Column 4	16	26763.9	1672.743	683343.6
ANOVA				
H=25.6	df=3	Non-addicted	Addicted	Naïve
p<0.0001	Mean Ranks for Sample	20.3	34.2	5
Post hoc test: Bonferroni-Holm				
Group 1	Group 2	Critical	P	Significant?
Addicted average	Naïve average	0.008333	1.18E-07	Yes
Naïve average	Experienced average	0.01	0.000128	Yes
Non-addicted average	Naïve average	0.0125	0.000606	Yes
Non-addicted average	Addicted average	0.016667	0.00148	Yes
Addicted average	Experienced average	0.025	0.05688	No
Non-addicted average	Experienced average	0.05	0.148061	No
Test for Equal Variance				
F	DFn	DFd	P	
10.23623293	3	37	4.79E-05	FAIL - equal variance cannot be assumed.

Supplemental Information

Table S7. Features within region extracted from the S-plot in Figure 4(b). Included are the retention time, mass, loadings and correlation coefficients, averages for each group, factor of change with uncertainty, in addition to the regional location in the MEDI heat map in Figure 4(c).

Retention Time	Mass	p[1]P	p(corr)[1]P	Experienced Average	Naïve Average	Factor of Change	Uncertainty	Region
11.76	130.066	-0.077	-0.730	60.693	4.294	14.1	5.121	blue
12.87	614.174	-0.046	-0.584	25.551	0.000	-3.4E+07	10000	green
12.83	616.179	-0.186	-0.689	354.740	0.000	10000	10000	blue
13.22	158.154	-0.056	-0.576	51.840	13.964	3.7	0.709	blue
10.19	162.056	-0.051	-0.739	26.495	1.509	17.6	12.02	purple
9.92	182.192	-0.198	-0.811	475.820	139.393	3.4	0.274	blue
10.82	185.080	-0.066	-0.633	64.801	17.044	3.8	0.653	blue
10.46	186.223	-0.031	-0.577	16.783	5.390	3.1	0.819	red
11.78	190.048	-0.059	-0.720	35.808	1.642	21.8	14.47	blue
9.09	201.068	-0.057	-0.670	47.493	13.496	3.5	0.548	grey
1.2	203.054	-0.049	-0.636	31.450	4.884	6.4	1.842	white
9.98	211.143	-0.056	-0.583	39.752	2.159	18.4	8.956	blue
9.6	215.125	-0.078	-0.615	96.993	29.218	3.3	0.497	blue
11.94	223.066	-0.042	-0.628	25.111	5.240	4.8	1.275	white
12.62	225.088	-0.065	-0.591	54.029	3.078	17.6	7.917	green
9.85	229.091	-0.084	-0.549	139.021	49.404	2.8	0.44	blue
10.53	229.142	-0.182	-0.709	418.809	97.261	4.3	0.527	blue
10.77	243.095	-0.065	-0.643	46.404	0.571	81.3	89.034	blue
10.85	257.136	-0.055	-0.602	48.442	14.323	3.4	0.556	blue
12.63	261.109	-0.100	-0.575	165.129	44.117	3.7	0.613	blue
9.55	269.132	-0.050	-0.704	35.318	9.933	3.6	0.595	purple
10.34	273.175	-0.069	-0.577	97.670	39.432	2.5	0.327	blue
9.09	279.688	-0.044	-0.571	33.130	10.547	3.1	0.667	blue
9.22	305.159	-0.032	-0.555	16.740	3.643	4.6	1.467	green
10.19	308.091	-0.049	-0.659	27.579	2.347	11.8	5.586	blue
9.49	338.087	-0.053	-0.598	32.848	0.000	-2.4E+07	10000	blue
9.74	340.104	-0.049	-0.630	29.279	2.385	12.3	6.83	blue
10.55	361.138	-0.049	-0.609	54.713	27.654	2	0.225	blue
9.48	371.227	-0.052	-0.606	41.785	10.601	3.9	0.875	purple
10.71	383.113	-0.042	-0.556	21.828	0.000	10000	10000	white
10.58	392.138	-0.046	-0.594	30.563	5.803	5.3	1.534	green
9.54	399.625	-0.037	-0.610	15.955	0.000	10000	10000	red
9.6	406.164	-0.046	-0.600	27.037	1.929	14	7.712	green
12.18	430.296	-0.067	-0.712	44.363	0.000	-4.2E+07	10000	green
9.96	431.090	-0.046	-0.728	19.933	0.000	-4E+07	10000	green
9.58	432.280	-0.048	-0.680	33.715	10.099	3.3	0.601	gray
12.18	448.306	-0.050	-0.626	29.513	1.137	26	20.871	green
12.19	466.318	-0.067	-0.556	58.968	2.487	23.7	11.808	green

Supplemental Information

14.09	500.277	-0.063	-0.577	48.380	1.157	41.8	33.792	blue
9.72	520.336	-0.069	-0.854	51.818	12.795	4	0.548	blue
14.11	542.327	-0.045	-0.582	24.440	0.000	-3.2E+07	10000	white
9.86	547.334	-0.031	-0.558	18.339	6.459	2.8	0.559	purple
9.93	608.387	-0.034	-0.680	18.797	7.236	2.6	0.481	purple

Supplemental Information

Table S8. Putative metabolite identification. Metabolites were given preliminary identifications based upon accurate mass, ion type, and database searching.

Retention Time (min)	m/z	MEDI Region	Putative Identification	Mass Accuracy (ppm)	Database
11.76	130.066	a	3-Methylene-indolene	6.9	HMDB
9.92	182.192	a	No Database Match		
10.82	185.080	a	(3-Methoxy-4-hydroxyphenyl) ethylene glycol	4.3	HMDB
11.78	190.048	a	Kynurenic acid	10.0	METLIN
9.85	229.091	a	Deoxyuridine	28.8	METLIN
10.53	229.142	a	No Database Match	N/A	
10.77	243.102	a	Thymidine	14.4	METLIN
10.34	273.175	a	Estradiol	37.7	LIPID MAPS
10.19	308.091	a	Glutathione	0.3	METLIN
9.49	338.087	a	3-Indole carboxylic acid glucuronide	0.9	HMDB
9.74	340.104	a	5-Hydroxy-6-methoxyindole glucuronide	3.8	HMDB
10.55	361.138	a	Dityrosine	3.9	HMDB
14.09	500.277	a	LPE(20:5)	0.4	LIPID MAPS
9.72	520.336	a	LPC(18:2)	7.3	LIPID MAPS
9.09	201.068	b	Bilirubin oxidation product	21.3	HMDB
12.62	225.088	c	3-Hydroxykynurenine	5.3	HMDB
9.22	305.159	c	Sodiated steroid-like molecule	N/A	
12.18	430.296	c	Glycocholic acid -2H ₂ O	1.9	METLIN
12.18	448.306	c	Glycocholic acid-H ₂ O	0.7	METLIN
12.19	466.318	c	Glycocholic acid	3.6	METLIN
10.19	162.056	d	4,6-Dihydroxyquinoline	6.1	HMDB
9.55	269.132	d	3-Carboxy-4-methyl-5-pentyl-2-furanpropionic acid	24.2	HMDB
9.48	371.227	d	6-Keto-prostaglandin F1a	21.5	LIPID MAPS
9.54	399.625	e	[M+2H] ²⁺	N/A	
1.20	203.054	f	Succinylacetacetate	4.9	HMDB
11.94	223.066	f	No Database Match	N/A	
14.11	542.327	f	LPC(20:5)	5.3	LIPID MAPS

Supplemental Information

Figure S3. Retention time selected isotope envelope for 3-methylene-indolenine.

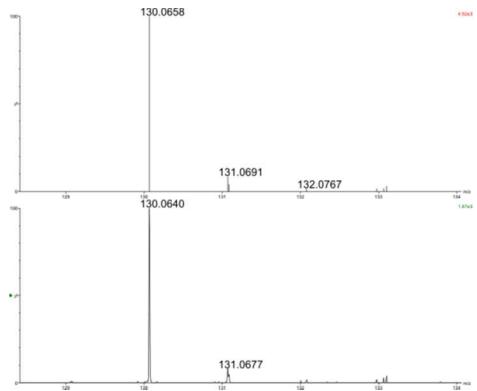


Figure S4. Retention time selected isotope envelope for (3-methoxy-4-hydroxyphenyl) ethylene glycol.

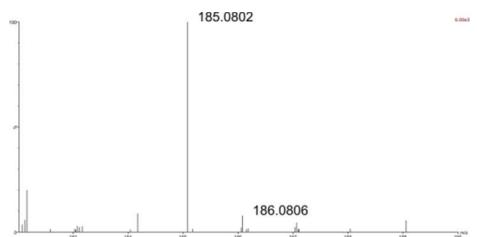


Figure S5. Retention time selected isotope envelope for kynurenic acid.

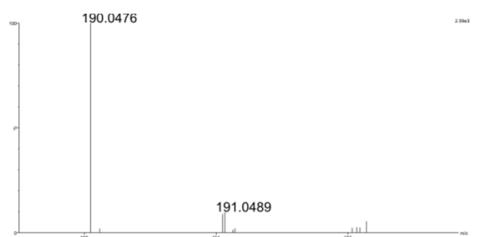


Figure S6. Retention time and drift time selected isotope envelope for deoxyuridine.

Supplemental Information

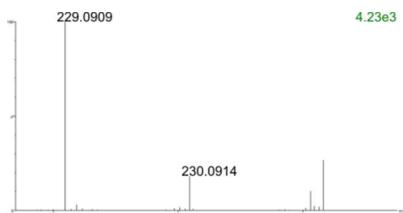


Figure S7. Intact spectrum of the isotopic envelope for thymidine.

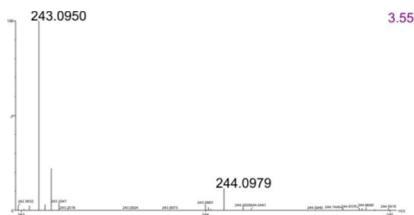


Figure S8. Retention time and drift time selected isotope envelope for estradiol.

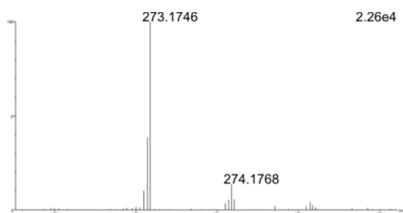


Figure S9. Retention time selected isotope envelope for glutathione.

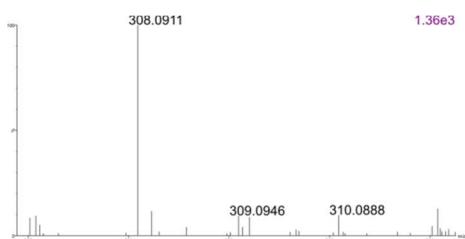


Figure S10. Retention time selected isotope envelope for 3-indole carboxylic acid glucuronide. We observe the loss of glucuronic acid (-176 Da) to generate a 162 m/z ion in the fragmentation spectrum.

Supplemental Information

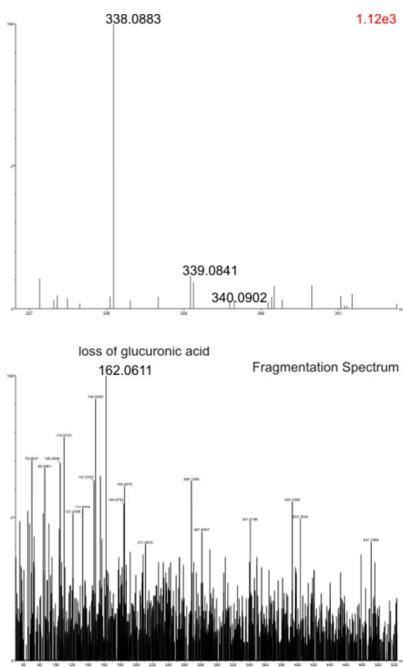


Figure S11. Retention time selected isotope envelope for 5-hydroxy-6-methoxyindole glucuronide. We observe the loss of glucuronic acid (-176 Da) to generate a 164 m/z ion in the fragmentation spectrum.

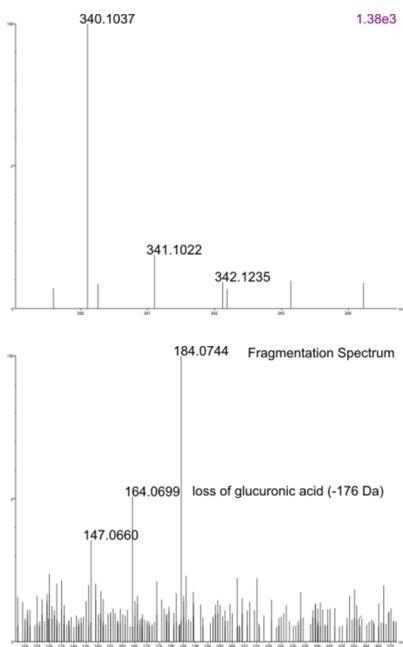


Figure S12. Retention time selected isotope envelope for dityrosine.

Supplemental Information

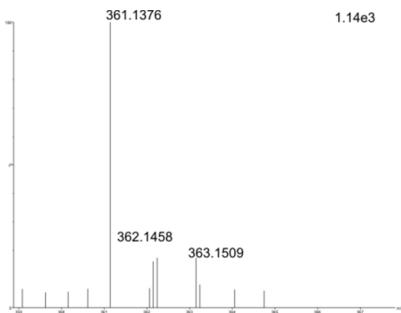


Figure S13. Retention time selected isotope envelope for LPE (20:5).

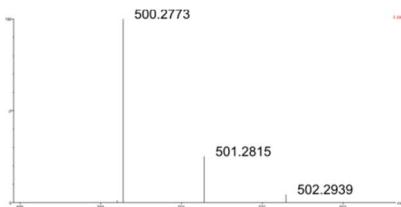


Figure S14. Retention time selected isotope envelope for LPC (18:2). A 184 head group fragment is also observed in the fragmentation spectrum below.

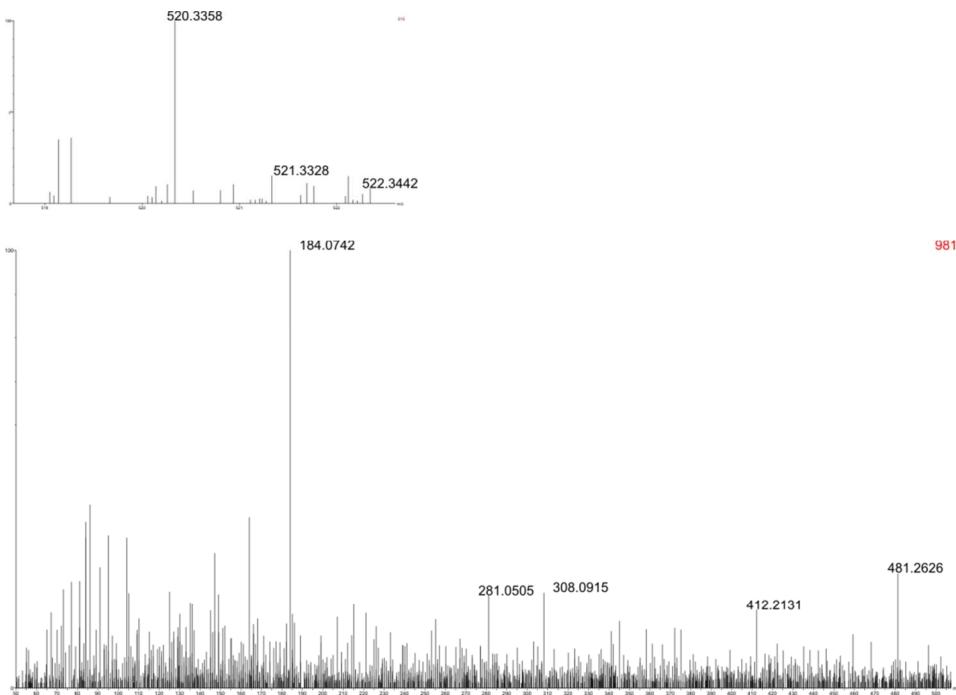


Figure S15. Representative structure for bilirubin oxidation product.

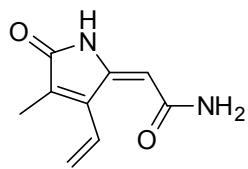


Figure S16. Retention time and drift time selected isotope envelope for 3-hydroxykynurenone.

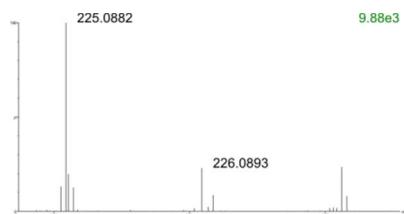


Figure S17. Retention time selected spectrum for glycocholic acid and in-source fragments.

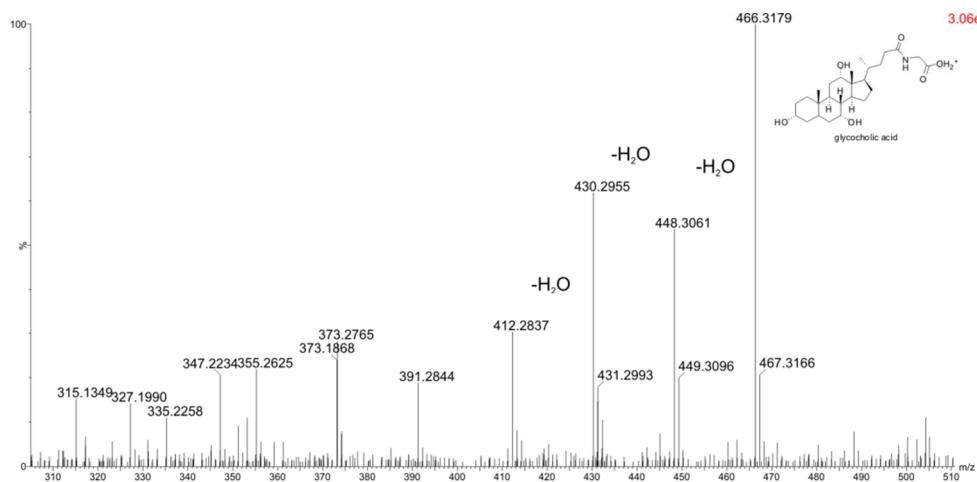


Figure S18. Retention time selected isotope envelope for 4,6-dihydroxyquinoline.

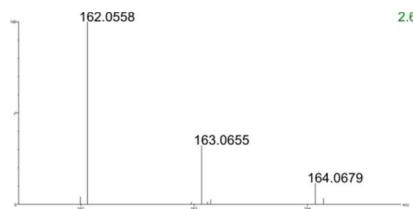


Figure S19. Retention and drift time selected isotope envelope for 3-carboxy-4-methyl-5-pentyl-2-furanpropanoic acid.

Supplemental Information

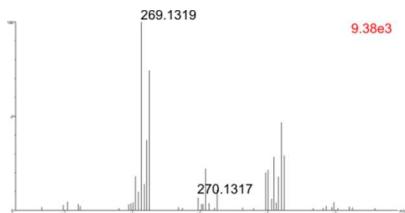


Figure S20. Retention and drift time selected isotope envelope for 6-keto-prostaglandin F1a.

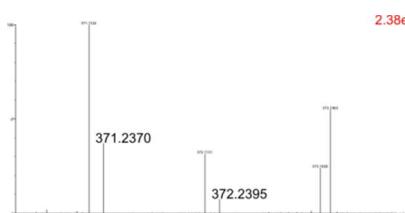


Figure S21. Retention time selected isotope envelope displaying doubly charged unknown analyte.

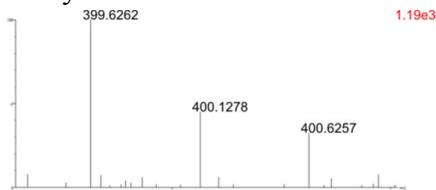


Figure S22. Retention time selected intact and fragmentation spectra for LPC(20:5). Desaturation locations are representative and not exact.

