Beeromics: From QC to IDs of Differentially Expressed Compounds in Craft Beer

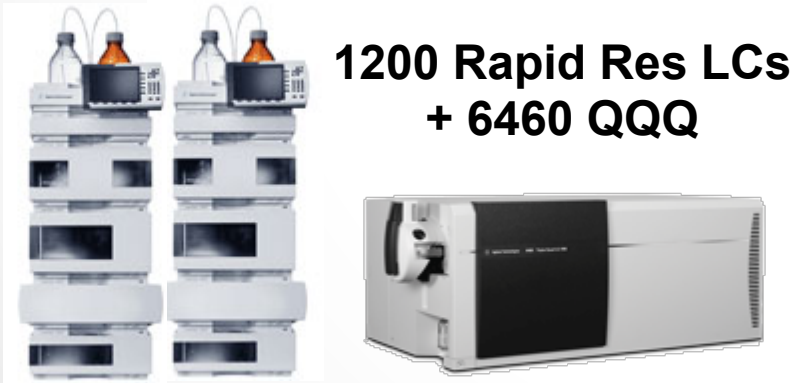
Christine A. Hughey
Jenny Phung, Chelsey M. McMinn
Department of Chemistry & Biochemistry, James Madison University, Harrisonburg, VA 22807

LC/MS Undergraduate Research Center at JMU

1290 UHPLC + 6224 TOF

1290 UHPLC + 6530 q-TOF

1200 Rapid Res LCs + 6460 QQQ



NSF MRI, Sept 2009

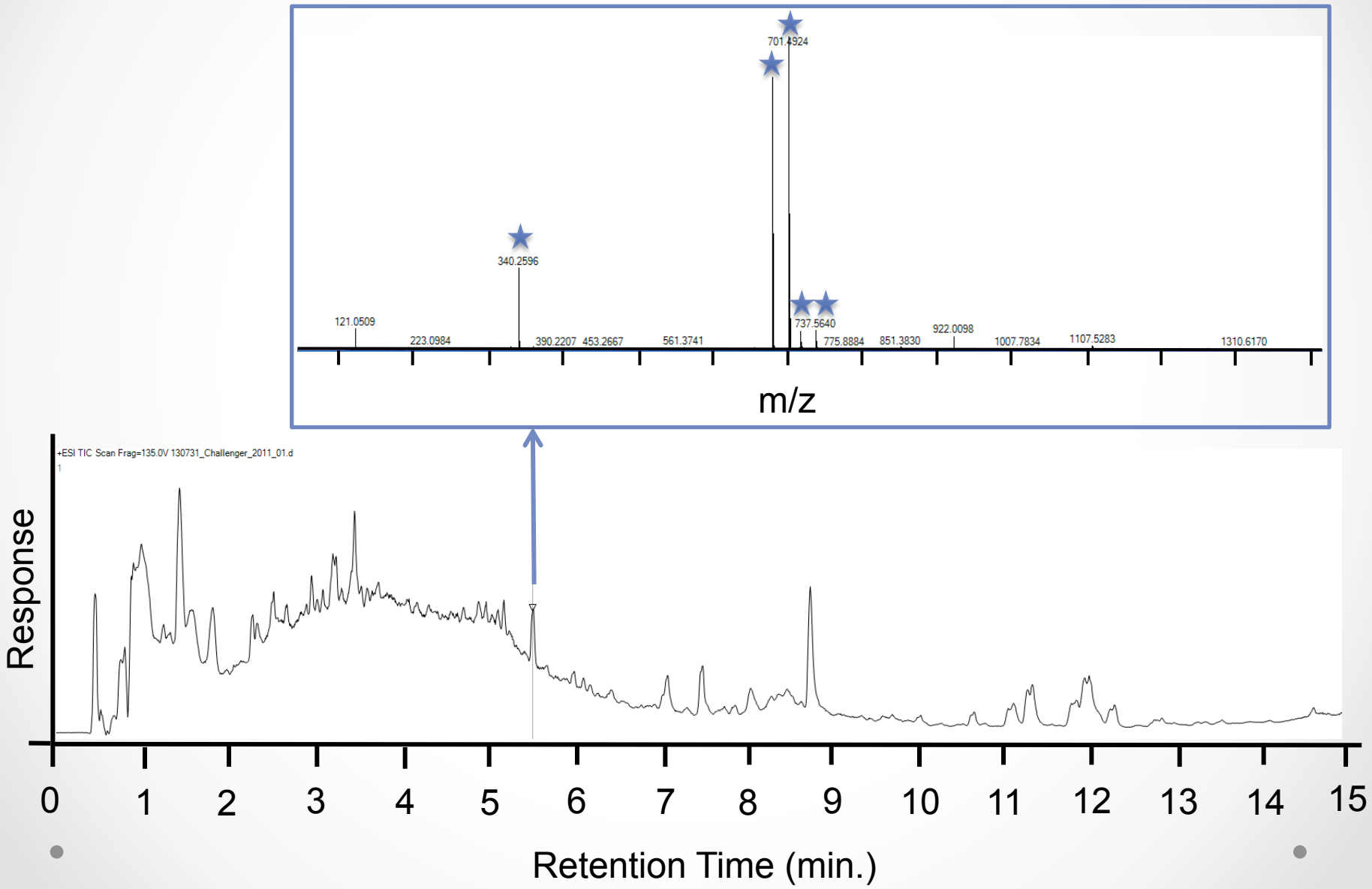
NSF MRI, January 2010

Beer-omics Objectives

- **Utilize stringent quality controls in order to continuously monitor data quality**
  - Six replicates of the QC beer were obtained on each data collection day to monitor retention time reproducibility, mass accuracy and instrument response
  - Normalized changes in instrument response with a spiked internal standard
  - Randomized samples
- **Conduct differential analysis of beer samples (e.g., hop used, brewery, beer type, production batches)**
- **Demonstrate the challenge of identifying differentially expressed compounds**
- **Use beeromics to teach metabolomic techniques in Instrumental Analysis to undergraduates.**

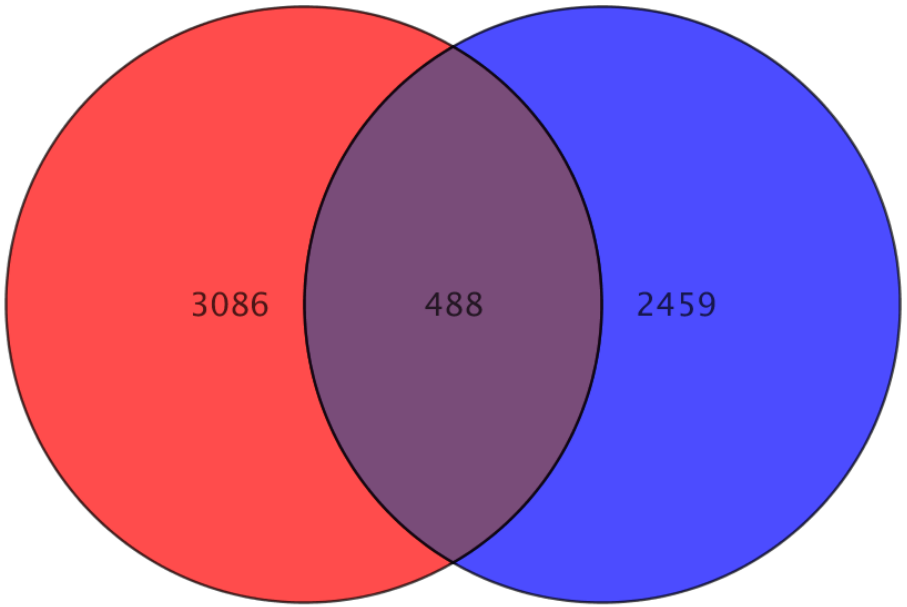
Fingerprinting by Molecular Features

Molecular feature = unique mass & retention time



QC Beer: Positive vs. Negative Ion ESI

Negative ion ESI
3574 Compounds



Positive ion ESI
2947 Compounds

6033 unique compounds in one beer sample

Workflow

Beer

Sonicate, filter, aliquot into cryovials, freeze at -80 °C

1290 UHPLC + 6530 q-TOF

Positive & negative ion ESI

Randomized X3

Molecular fingerprinting w/ Agilent's Mass Profiler Professional software

Targeted Quantitation

Part 1:
Quality Control Measures

How do you know that the differences observed are due to the samples and not variation in instrument performance?

QC Beer: Reproducibility

Target compound	Monoisotopic Neutral Mass	Negative ESI			
		Interday RT Std. dev.[a] (s)	Interday Mass Error (ppm)	Interday Abs. Area (% RSD)	Intraday Abs. Area (% RSD)
Catechin	290.0790	0.9	2.65	18.8%	4.1%
Rutin	610.1534	0.7	1.16	5.8%	2.6%
Quercetin-3-*O*-glucoside	464.0955	0.7	-0.11	9.3%	2.6%
Kaempferol-3-*O*-rutinoside	594.1584	0.8	1.40	7.7%	2.2%
Isorhamnetin-3-*O*-rutinoside	624.1690	0.7	7.71	26.0%	13.0%
Kampferol-3-*O*-glucoside	448.1006	0.8	1.50	10.3%	2.9%
Kaempferol	286.0477	1.0	2.22	14.4%	3.6%
Isorhamnetin-3-*O*-glucoside	478.1111	--	--	--	--
Guanosine[c]	283.0917	--	--	--	--
Naphthoic Acid (IS)[d]	172.0524	1.3	2.32	21.4%	4.4%
Caffeine (IS)[d]	194.0804	--	--	--	--
Avg.		**0.9±0.2**	**2.36±2.33**	**14.2±7.2%**	**4.4±3.6%**
QC sample[e]				**10.5%**	**2.2±1.3%**

[a] Same column.
[b] Column was changed during data collection.
[c] Non-retained peak.
[d] Internal standard (IS).
[e] Average summed response for MFs extracted from QC samples (n=54)

QC Beer: Reproducibility

Raw Data → **Daily Targeted Quantitation**

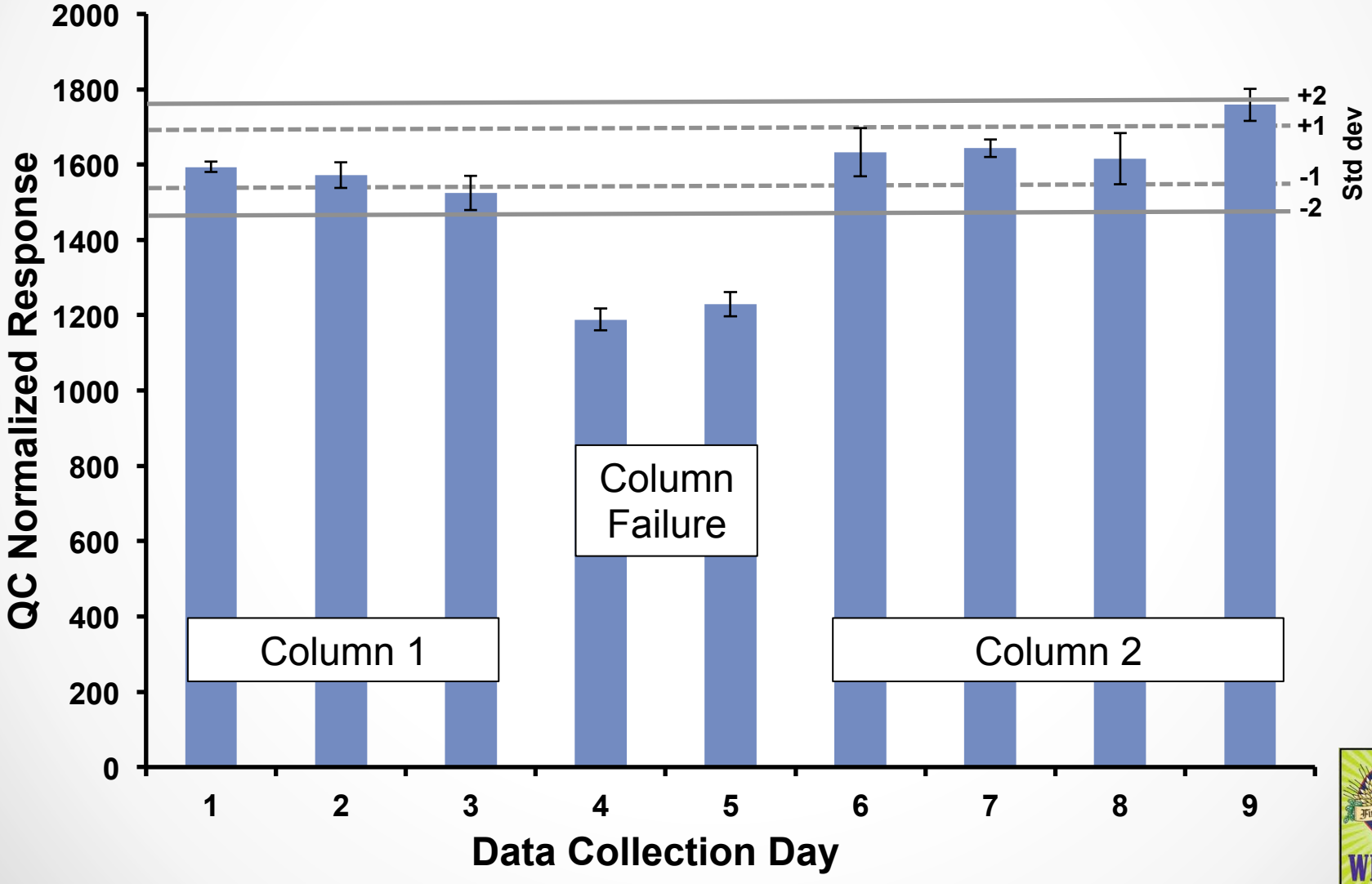
Target compound	Monoisotopic Neutral Mass	Positive ESI			
		Interday RT Std. dev.[b] (s)	Interday Mass Error (ppm)	Interday Abs. Area (% RSD)	Intraday Abs. Area (% RSD)
Catechin	290.0790	3.6	-2.93	8.6%	1.8%
Rutin	610.1534	--	--	--	--
Quercetin-3-*O*-glucoside	464.0955	2.6	0.36	6.0%	2.6%
Kaempferol-3-*O*-rutinoside	594.1584	2.2	0.13	7.5%	1.9%
Isorhamnetin-3-*O*-rutinoside	624.1690	--	--	--	--
Kampferol-3-*O*-glucoside	448.1006	--	--	--	--
Kaempferol	286.0477	--	--	--	--
Isorhamnetin-3-*O*-glucoside	478.1111	2.6	-1.33	9.3%	4.1%
Guanosine[c]	283.0917	0.9	-1.01	28.4%	11.6%
Naphthoic Acid (IS)[d]	172.0524	--	--	--	--
Caffeine (IS)[d]	194.0804	2.3	-8.21	10.0%	1.3%
Avg.		**2.4±0.9**	**-2.17±3.16**	**11.7±8.3**	**3.9±3.8%**
QC sample[e]				**8.0%**	**2.2±2.1%**

[a] Same column.
[b] Column was changed during data collection.
[c] Non-retained peak.
[d] Internal standard (IS).
[e] Average summed response for MFs extracted from QC samples (n=54)

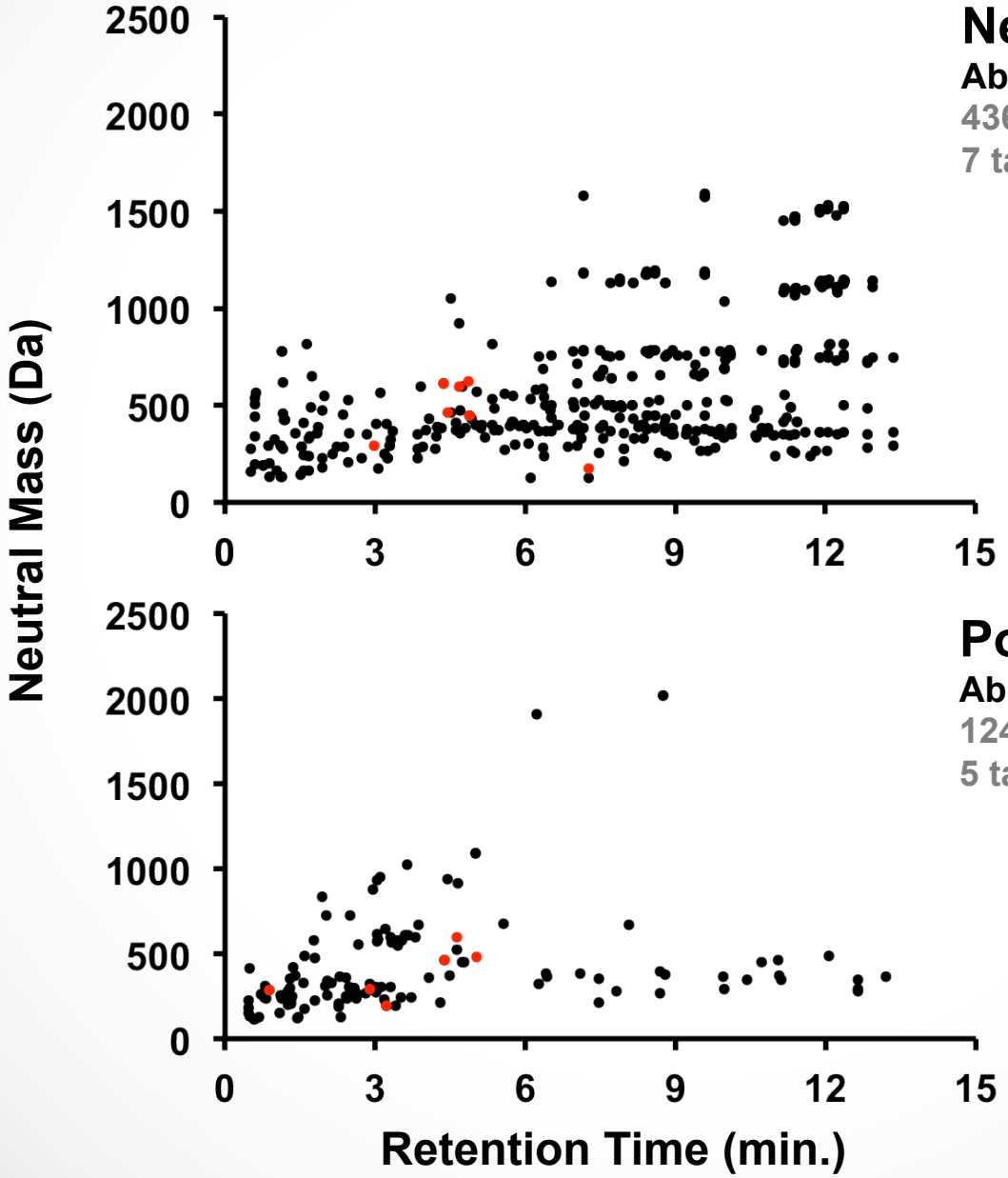
QC Beer: Reproducibility

Collected over 1.5 weeks

$$\text{Normalized Response} = \frac{\text{Summed Response of Standards}}{\text{Avg Response of Int Standard}}$$



Target Compounds vs. QC Beer



Negative Ion ESI

Abundance > 1e6

436 molecular features (**black**)

7 target compounds (**red**)

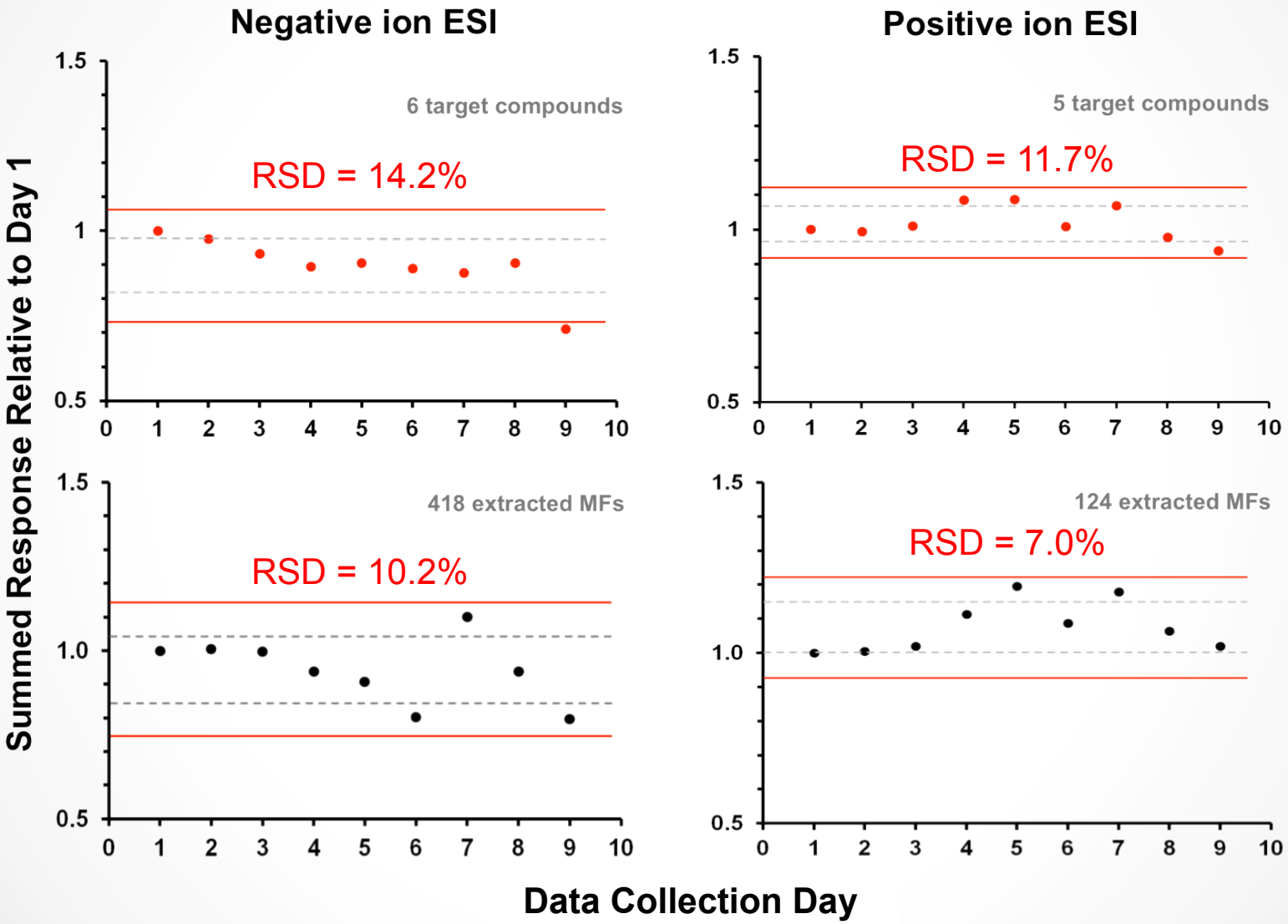
Positive Ion ESI

Abundance > 5e5

124 molecular features (**black**)

5 target compounds (**red**)

Target Compounds vs. QC Beer



Data Analysis

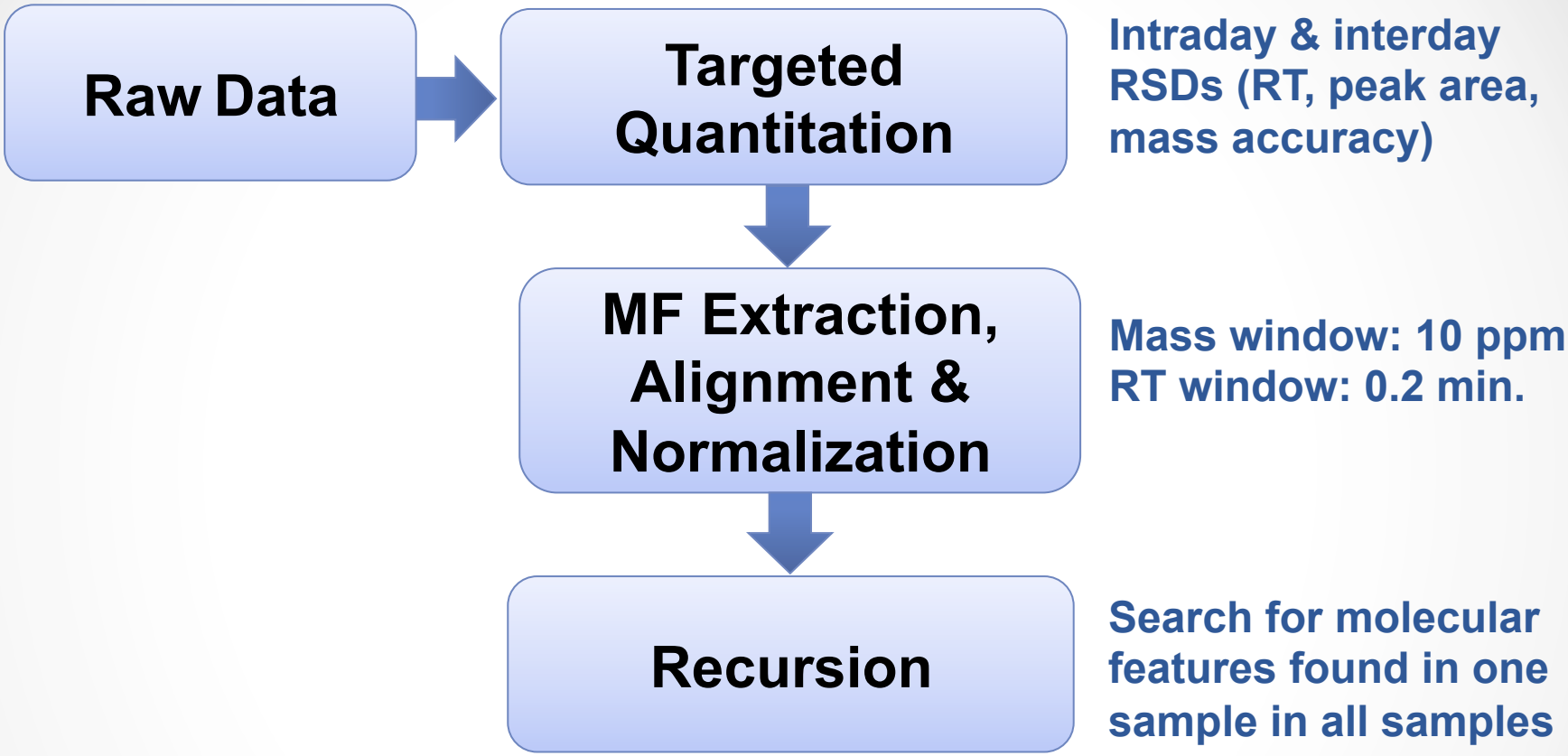
Raw Data → **Targeted Quantitation**

Intraday & interday RSDs (RT, peak area, mass accuracy)

Targeted Quantitation → **MF Extraction, Alignment & Normalization**

Mass window: 10 ppm
RT window: 0.2 min.

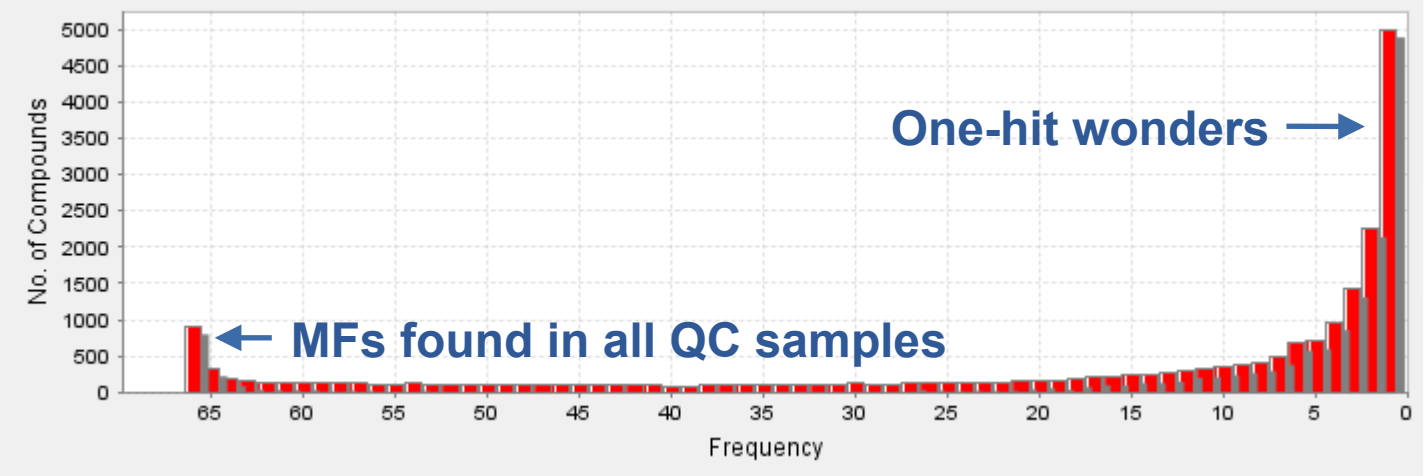
Data Analysis



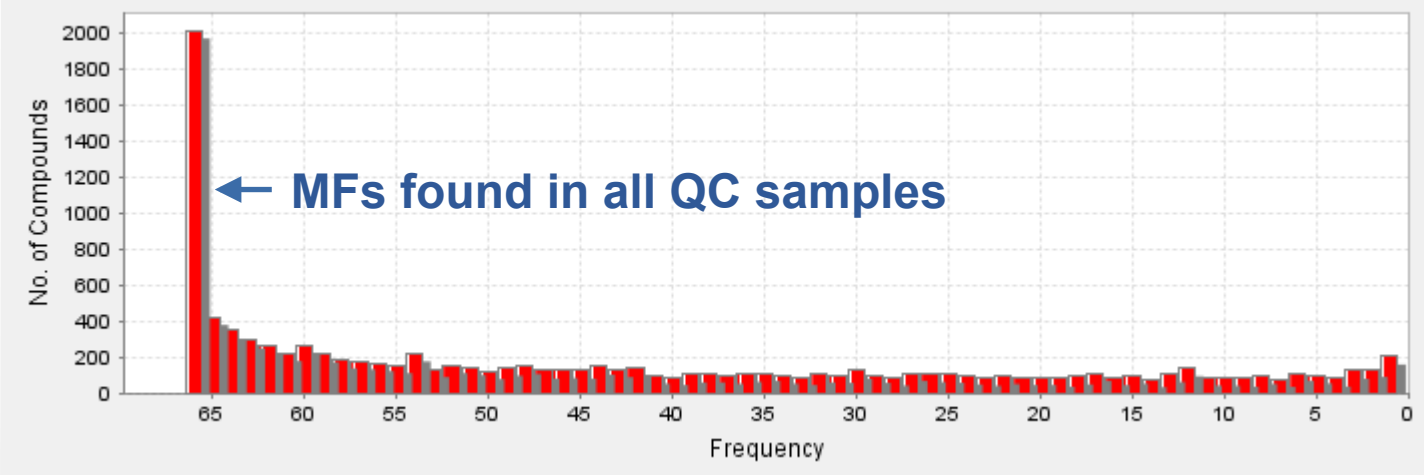
QC Beer: Recursion

Negative ion ESI

Before Recursion:



After Recursion:



Part 2:
Untargeted Differential Analysis

Identical malts and yeast was used in all the Mikkeller beers.
Malt: 67% Pilsner, 11% Cara-Crystal, 11% Munich II, 11% Flaked Oats
Yeast: American Ale (Wyeast 1056/WLP099)
6.9% ABV
~100 IBUs
--Reported by Mikkel (the brewer) on the Beer Brewing Network.com

Mikkeller's Single Hop India Pale Ales (IPAs)

24 Single Hop Beers
3 hops were used in two separate runs (2010 & 2011)
1 Quality Control = Green Flash West Coast IPA

2010	2011	2011
• Amarillo	• Amarillo	• Mt. Hood
• Cascade	• Bravo	• Nugget
• Centennial	• Centennial	• Palisade
• Chinook	• Challenger	• Simcoe
• Nelson Sauvin	• Cluster	• Super Galena
• Nugget	• Columbus	• Tettnanger
• Simcoe	• East Kent	• Willamette
• Tomahawk	• Magnum	
• Warrior		

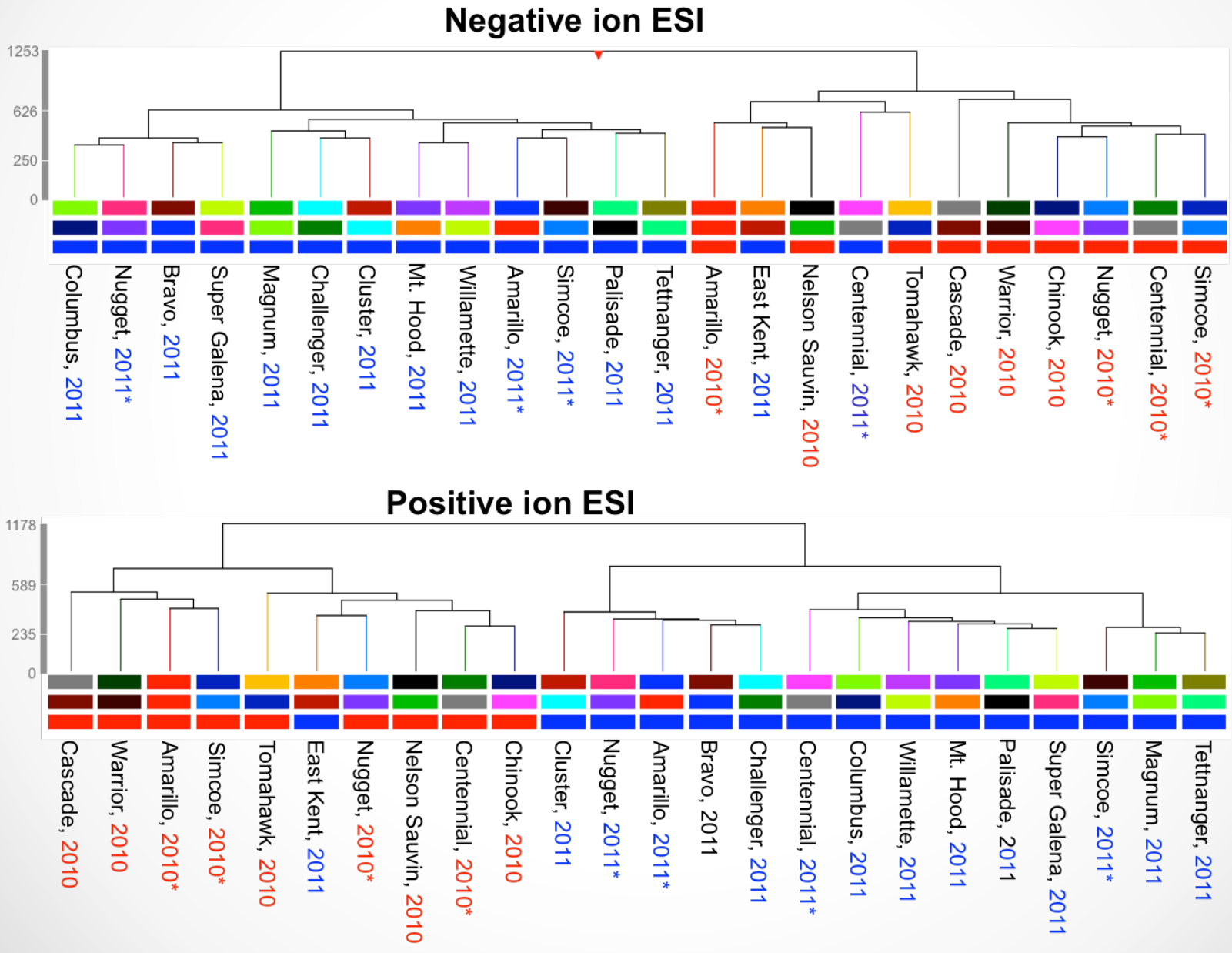
Differential Analysis: Mikkeller IPAs

Fold changes (FC) observed for 2010 vs. 2011 Mikkeller beers with a moderated t-test using Benjamini-Hochberg multiple testing correction.

	Negative ion ESI			Positive ion ESI		
	All	P < 0.05	P < 0.01	All	P < 0.05	P < 0.01
FC All	4415	2666	1936	3452	2002	1529
FC > 1.1	4291	2666	1936	3253	2002	1529
FC > 1.5	3926	2586	1909	2908	1943	1507
FC > 2.0	3542	2466	1801	2647	1845	1418
FC > 3.0	3119	2359	1719	2291	1764	1349
By chance		133	19		100	15

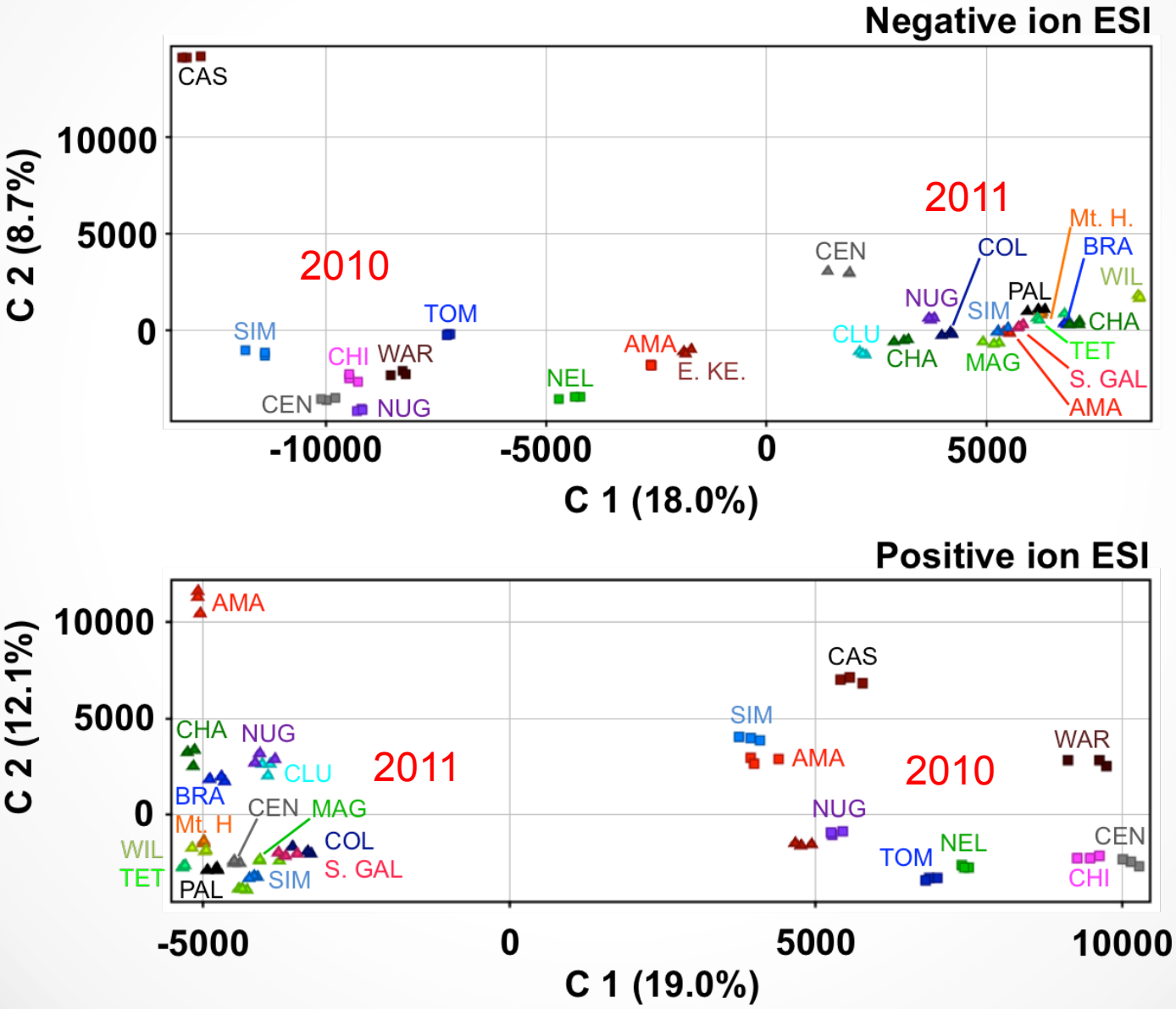
Differential Analysis: Mikkeller IPAs

Abundance filtered to retain ≥ 75th percentile



Differential Analysis: Mikkeller IPAs

Abundance filtered to retain ≥ 75th percentile



Differential Analysis: Mikkeller IPAs

Abundance filtered to retain ≥ 75th percentile

Tukey HSD post hoc test results for 2011 Mikkeller beers analyzed in negative ion mode ($p<0.05$). Gray shaded boxes indicate the number (and %) of differentially expressed molecular features. White boxes correspond to the number (and %) of features common to each pair. Retained features (2400) exhibited an abundance >75th percentile and passed ANOVA filtering ($p<0.05$).

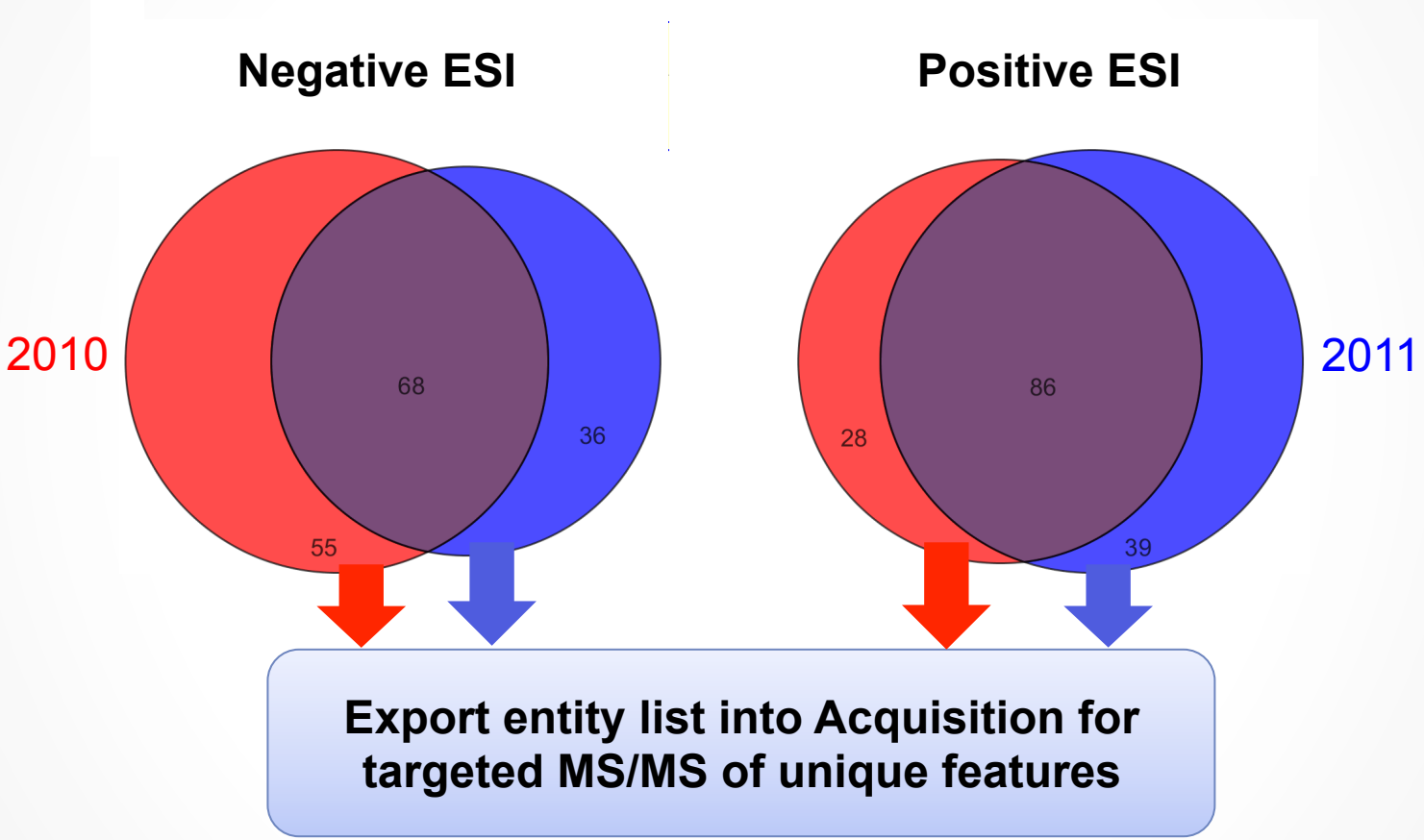
	COL	MAG	WIL	TET	S. GAL	CLU	PAL	Mt. H	CHA	CEN	BRA	SIM	E. KE	AMA	NUG
COL	2400 (100%)	609 (25%)	1099 (46%)	991 (41%)	710 (30%)	741 (31%)	828 (35%)	638 (27%)	733 (31%)	1182 (49%)	895 (37%)	695 (29%)	1015 (42%)	658 (27%)	301 (13%)
MAG	1791 (75%)	2400 (100%)	1221 (51%)	924 (39%)	888 (37%)	787 (33%)	893 (37%)	901 (38%)	800 (33%)	1178 (49%)	1058 (44%)	803 (33%)	923 (38%)	700 (29%)	727 (30%)
WIL	1301 (54%)	1179 (49%)	2400 (100%)	1079 (45%)	874 (36%)	1177 (49%)	950 (40%)	935 (39%)	956 (40%)	1338 (56%)	689 (29%)	1063 (44%)	1264 (53%)	1105 (46%)	1054 (44%)
TET	1409 (59%)	1476 (62%)	1321 (55%)	2400 (100%)	1022 (43%)	987 (41%)	890 (37%)	869 (36%)	857 (36%)	1082 (45%)	1106 (46%)	922 (38%)	1188 (50%)	860 (36%)	1009 (42%)
S. GAL	1690 (70%)	1512 (63%)	1526 (64%)	1378 (57%)	2400 (100%)	959 (40%)	685 (29%)	892 (37%)	733 (31%)	1250 (52%)	644 (27%)	816 (34%)	1118 (47%)	907 (38%)	681 (28%)
CLU	1659 (69%)	1613 (67%)	1223 (51%)	1413 (59%)	1441 (60%)	2400 (100%)	1147 (48%)	805 (34%)	868 (36%)	1253 (52%)	1061 (44%)	1065 (44%)	1130 (47%)	903 (38%)	792 (33%)
PAL	1572 (66%)	1507 (63%)	1450 (60%)	1510 (63%)	1715 (71%)	1253 (52%)	2400 (100%)	661 (28%)	591 (25%)	1160 (48%)	999 (42%)	601 (25%)	1028 (43%)	766 (32%)	758 (32%)
MT. H	1762 (73%)	1499 (62%)	1465 (61%)	1531 (64%)	1508 (63%)	1595 (66%)	1739 (72%)	2400 (100%)	597 (25%)	1206 (50%)	1113 (46%)	759 (32%)	1077 (45%)	691 (29%)	631 (26%)
CHA	1667 (69%)	1600 (67%)	1444 (60%)	1543 (64%)	1667 (69%)	1532 (64%)	1809 (75%)	1803 (75%)	2400 (100%)	1210 (50%)	863 (36%)	566 (24%)	921 (38%)	678 (28%)	699 (29%)
CEN	1218 (51%)	1222 (51%)	1062 (44%)	1318 (55%)	1150 (48%)	1147 (48%)	1240 (52%)	1194 (50%)	1190 (50%)	2400 (100%)	1301 (54%)	1186 (49%)	1391 (58%)	1180 (49%)	1160 (48%)
BRA	1505 (63%)	1342 (56%)	1711 (71%)	1294 (54%)	1756 (73%)	1339 (56%)	1401 (58%)	1287 (54%)	1537 (64%)	1099 (46%)	2400 (100%)	989 (41%)	1161 (48%)	1011 (42%)	945 (39%)
SIM	1705 (71%)	1597 (67%)	1337 (56%)	1478 (62%)	1584 (66%)	1335 (56%)	1799 (75%)	1641 (68%)	1834 (76%)	1214 (51%)	1411 (59%)	2400 (100%)	1005 (42%)	591 (25%)	682 (28%)
E. KE	1385 (58%)	1477 (62%)	1136 (47%)	1212 (51%)	1282 (53%)	1270 (53%)	1372 (57%)	1323 (55%)	1476 (62%)	1009 (42%)	1239 (52%)	1395 (58%)	2400 (100%)	958 (40%)	1022 (43%)
AMA	1742 (73%)	1700 (71%)	1295 (54%)	1540 (64%)	1493 (62%)	1497 (62%)	1634 (68%)	1709 (71%)	1722 (72%)	1220 (51%)	1389 (58%)	1809 (75%)	1442 (60%)	2400 (100%)	634 (26%)
NUG	2099[1] (87%)	1673 (70%)	1346 (56%)	1391 (58%)	1719 (72%)	1608 (67%)	1642 (68%)	1769 (74%)	1701 (71%)	1240 (52%)	1455 (61%)	1718 (72%)	1378 (57%)	1766 (74%)	2400 (100%)

[1]Nugget and Columbus share the most features in common (87%) - purple

[2]East Kent and Centennial have the least features in common (42%) - blue

Differential Analysis: Mikkeller IPAs

Abundance filtered to retain ≥ 75th percentile



Part 3:
ID of Unique Molecular Features

Identification of Unique Molecular Features

Targeted MS/MS

↓

METLIN Database & MS/MS Library Search

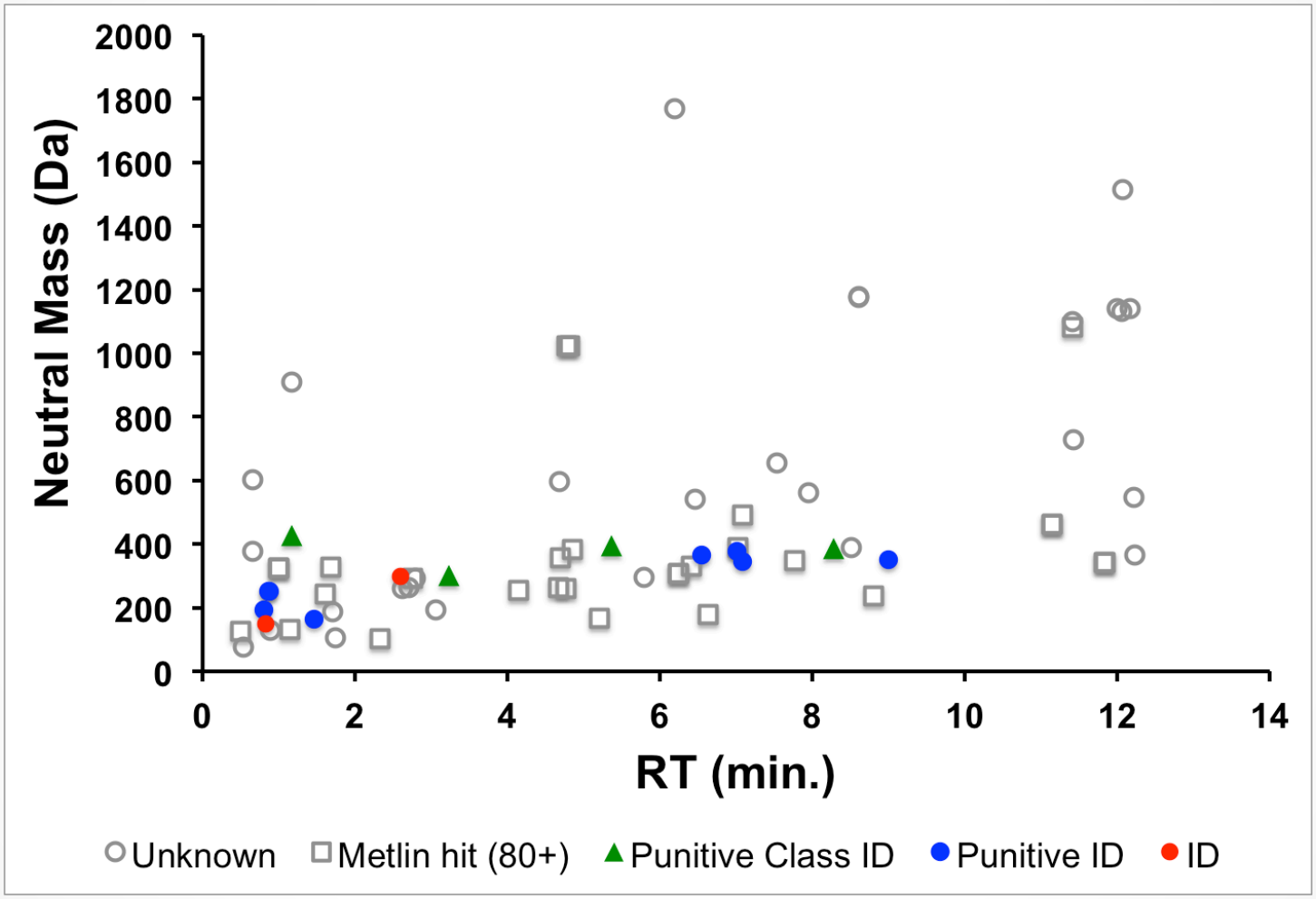
↓

Molecular Structure Correlator + ChemSpider

↔

Identification of Unique Molecular Features

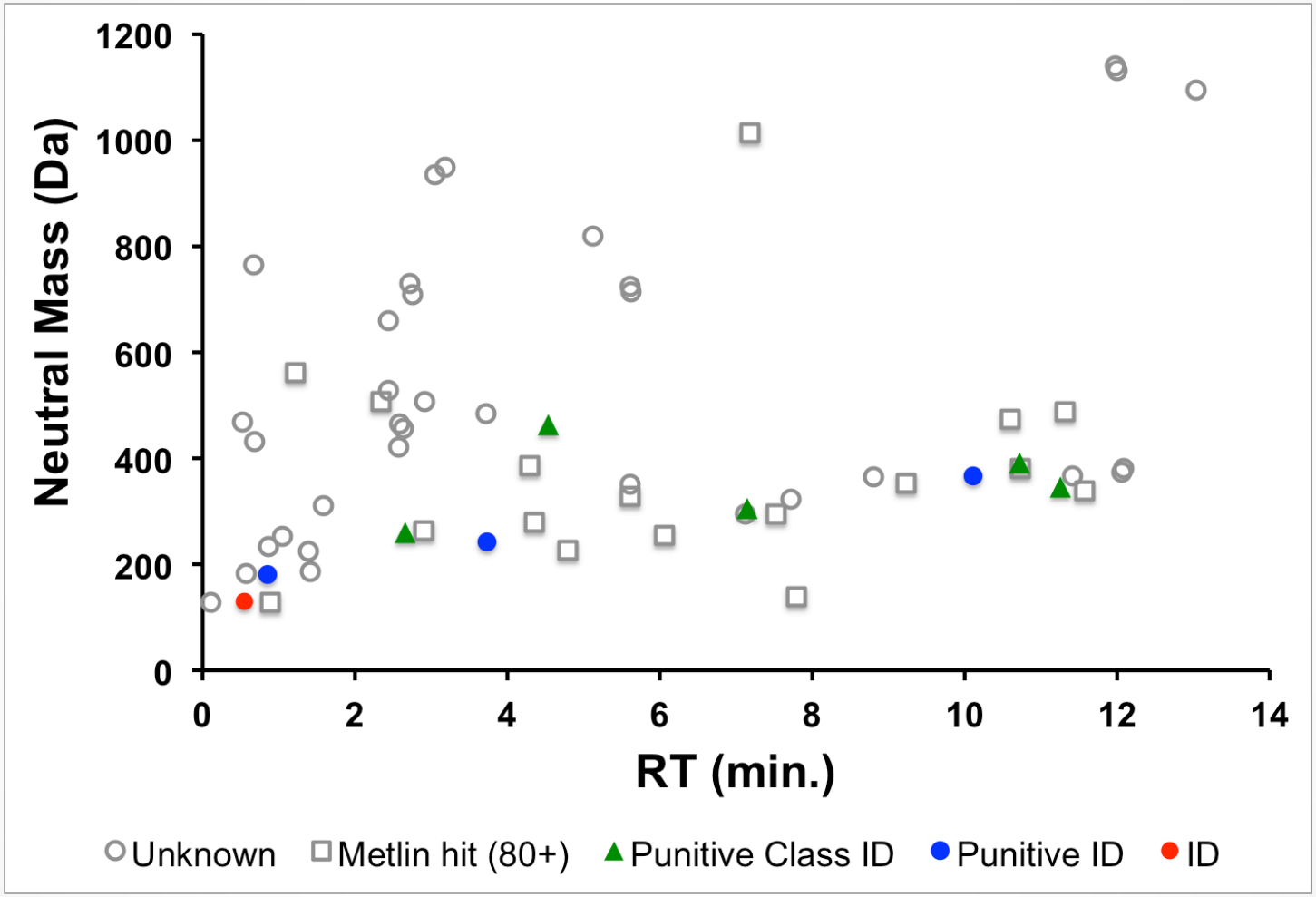
MFs unique to 2010 (negative & positive ESI combined)



- ID Level 1 = Matched to authentic standard
- ID Level 2 = Punitive ID – High-scoring MS/MS library match
- ID Level 3 = Punitive class (di- or tri-peptide)
- ID Level 4 = Unknown

Identification of Unique Molecular Features

MFs unique to 2011 (negative & positive ESI combined)



- ID Level 1 = Matched to authentic standard
- ID Level 2 = Punitive ID – High-scoring MS/MS library match
- ID Level 3 = Punitive class (di- or tri-peptide)
- ID Level 4 = Unknown

Identification of Unique Molecular Features

MFs unique to 2010 (negative & positive ESI combined)

Punative ID	ID level	RT (min.)	Neutral mass (Da)	Polarity	Year	DB Score	Library Score	MSC Score	MSC % Wt	# Metlin DB hits
citric acid	2	0.802	192.0257	neg	2010	92.4	61.9	91.7	94.3	12
methionine	1	0.826	149.0497	pos	2010	89.1	62.3	73.8	87.0	3
5'-deoxyadenosine	4[a]	0.864	251.1021	pos	2010	55.2	100	72.9	86.5	5
phenyalanine	4[a]	1.462	165.0776	pos	2010	85.1	98.8	73.5	90.8	19
5-methylthioadenosine	1	2.590	297.0881	pos	2010	90.5	98.2	84.2	90.4	1
20-carboxy-LTB4	2	6.546	366.2037	neg	2010	98.5	90.9	28.8	39.6	14
(iso)humulinone	2	7.007	378.2038	neg	2010	97.0	NA	88.9	98.6	23
(-)-11-nor-caroxy-9-THC	4[a]	7.079	344.1972	pos	2010	71.8	88.1	80.4	91.1	18
cohumulone	4[a]	7.765	348.1930	pos	2010	83.2	NA	74.5	91.7	23
19(R)-hydroxy-PGB2	2	8.992	350.2093	neg	2010	94.1	83.6	52.2	60.2	34
humulinic acid	2	9.509	266.1513	neg	2010	98.7	NA	82.4	97.1	28
agmatine	1	0.550	130.1212	pos	2011	84.8	NA	89.6	91.7	1
tyrosine	4[a]	0.854	181.0722	pos	2011	86.6	94.7	77.0	98.1	3
20-carboxy-LTB4	2	10.111	366.2043	neg	2011	99.5	86.7	7.86	10.3	13
4-deoxyhumulone	2	11.259	346.2137	neg	2011	82.44	NA	71.2	87.3	31

[a]RT did not match authentic standard. ID level of "2" would have been assigned if standard had not been compared.

ID Level 1 = Matched to authentic standard
ID Level 2 = Punitive ID – High-scoring MS/MS library match
ID Level 3 = Punitive class (di- or tri-peptide)
ID Level 4 = Unknown

Proposed minimum reporting standards for chemical analysis. *Metabolomics.* **2007**, *3*, 211-221.
Metabolomic profiling of beer reveals effect of temperature on non-volatile small molecules during short-term storage. **2012**, *135*, 1284-1289.

Conclusions

- **Utilized stringent quality controls in order to continuously monitor data quality**
  - QC standard was used to determine the interday RSDs and determine that differences observed in the samples were due to the samples themselves and not instrument response
  - Randomization of data sets further insures trends observed are real
- **Conducted differential analysis of beer samples**
  - Observed differences due to year of production
  - Beers were compositionally similar so rigorous QC was needed to identify statistically significant differences
- **Positive identification of unique features is hard—even with MS/MS and powerful metabolomic tools.**

Acknowledgements

Agilent Technologies:
Nathan Miller
Mike Scott
Andy Gieschen
Rick Reisdorph (MPP Training)
Steve Fischer

Funding:
NSF CHE-1046630
NSF CHE-0958973
NSF REU-1062629
Research Corp. Department Development
Grant 7957