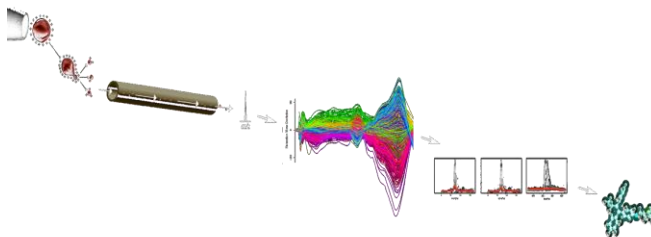


# XCMS Online: Using and understanding



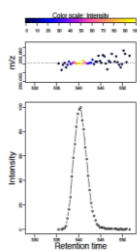
H. Paul Benton PhD  
The Siuzdak Laboratory - The Scripps Research Institute

## What we'll cover

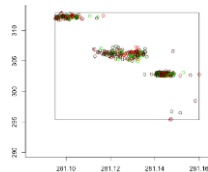
- Brief reminder of XCMS terms and concepts
- XCMS Online – This is the visual part of xcms
  - Using the system
  - Outputs
    - Using Npeaks
    - Statistical evaluations – via univariate and multivariate

# Overview of XCMS

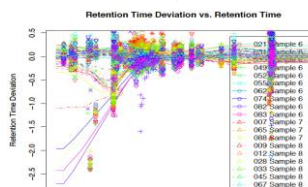
## Peak Picking



Grouping similar peaks  
across replicates




Retention time  
alignment




Statistical  
analysis of Peaks  
Between classes

# XCMS Online


 Scripps Center for Metabolomics
 user: nipo

[Home](#)
[Create Job](#)
[View Results](#)
[Stored Datasets](#)
[Dashboard](#)
[XCMS Institute](#)
[XCMS Public](#)
[Contact](#)


 XCMS Online  
 a universal MS translator  
 for all instrument platforms

Welcome to XCMS Online.

XCMS Online analysis provides the same high-quality metabolomic analysis that you are used to with XCMS but allows users to process isotope enriched data. Predefined parameter settings for different instruments (e.g. QTOF, Orbitrap, etc.) are available as well as options for customization. Results can be viewed online in an interactive, customizable table showing statistics, chromatograms, and putative METLIN identities. All results and images are available for download as zip files.

You will need a user account to use the system however user accounts are free and only require a valid e-mail address to register. All you need to do is to create a job which involves the two steps:

- 1) uploading the datasets (See [File Formats](#) for more information)
- 2) selecting the parameters and submitting the job.

After the processing is complete, you will be notified via e-mail that your results are ready for review.

Please click "Create Job" to upload your datasets now or download the user manual for detailed instructions.

## Job types in XCMS Online

- Single Class Jobs – Either single file or single dataset (class or unknown classes)
- Pairwise analysis Jobs – Two class comparison, works best with KO –vs- WT type experiments
- Multigroup/class Jobs – Multiple classes including Quality control samples. Great for Time series jobs or multiple knockouts.
- meta-XCMS- Finds the overlap between many pairwise jobs. (Must have same control samples)

The screenshot displays the XCMS Online web interface. At the top, there is a navigation menu with the following items: Home, Create Job, View Results, Stored Datasets, Dashboard, XCMS Institute, XCMS Public, and Contact. Below the menu, there is a dropdown menu for 'Create Job' with options: Single Job, Pairwise Job, Meta XCMS Job, and Multigroup Job. To the right of the dropdown, there is a 'View Jobs' button and a set of icons for 'Share Jobs', 'Job Grouping', 'Resubmit Jobs', and 'Delete Jobs'. The main content area shows a table of job results with the following columns: Exp Type, Status, ID, Progress, JobName, Datasets (ID#) (\*control), Created, Parameters (ID#), Group, and Shared. The table contains 10 rows of job data, all with a status of 'VIEW' and a progress of '100%'. The jobs are all of type 'SINGLE' and have various IDs and job names.

Job Count: 67  
Search Jobs:

Exp Type	Status	ID	Progress	JobName	Datasets (ID#) (*control)	Created	Parameters (ID#)	Group	Shared
SINGLE	VIEW	1087997	job complete 100%	sgl_2015-12-14_13:21	coke (#161171)	2015-12-14 13:21:17	HPLC / Q-T (1)		
SINGLE	VIEW	1086419	job complete 100%	sgl_2015-12-02_11:04	Dataset_0 (#158936)	2015-12-02 16:29:26	HPLC / Q-T (1)		
SINGLE	VIEW	1086418	job complete 100%	sgl_2015-12-02_11:04	Dataset_0 (#158935)	2015-12-02 16:23:56	HPLC / Q-T (1)		
SINGLE	VIEW	1086358	job complete 100%	sgl_2015-12-02_11:04	Dataset_0 (#158845)	2015-12-02 11:04:03	HPLC / Q-T (1)		
SINGLE	VIEW	1085813	job complete 100%	sgl_2015-11-30_14:22	Dataset_0 (#158096)	2015-11-30 14:22:49	HPLC / Q-T (1)		
SINGLE	VIEW	1085194	job complete 100%	sgl_2015-11-24_14:14	Dataset_0 (#157235)	2015-11-24 17:36:00	HPLC / Q-T (1)		
SINGLE	VIEW	1085180	job complete 100%	sgl_2015-11-24_14:14	Dataset_0 (#157215)	2015-11-24 14:43:22	HPLC / Q-T (1)		
SINGLE	VIEW	1085176	job complete 100%	sgl_2015-11-24_14:14	Dataset_0 (#157210)	2015-11-24 14:14:16	HPLC / Q-T (1)		
SINGLE	VIEW	1085027	job complete 100%	sgl_2015-11-23_17:21	Dataset_0 (#156925)	2015-11-23 17:21:42	HPLC / Q-T (1)		

# Pairwise analysis

User: test [Logout](#)

Home Create Job View Results Stored Datasets Dashboard XCMS Institute XCMS Public Contact

1 → 2 → 3 → 4 [Reset](#)

**1 Dataset 1**

[Load New Dataset](#) OR [Select Dataset](#)  
(See [File Formats](#) for more information)

Dataset 1 ID: **Not Defined**  
Dataset 1 no. of files: 0  
Dataset 1 Name:

**2 Dataset 2**

[Load New Dataset](#) OR [Select Dataset](#)  
(See [File Formats](#) for more information)

Dataset 2 ID: **Not Defined**  
Dataset 2 no. of files: 0  
Dataset 2 Name:

**3 Parameters**

Select Parameters:

**Job Summary**

Job ID: 1094161  
User: test (16)

Job Name:  [Edit](#)

Dataset 1:  (control)

Dataset 2:

Parameter Set: 0

**4 Submit**

[Click here to complete your job](#) [Submit Job](#)

# Multigroup job

User: test [Logout](#)

Home Create Job View Results Stored Datasets Dashboard XCMS Institute XCMS Public Contact

1 → 2 → 3 → 4 [Reset](#)

**1 Select Datasets**

[Load New Dataset](#) OR [Select Dataset](#)  
(See [File Formats](#) for more information)

<input type="checkbox"/>	ID	Dataset Name	Number of Files

**2 Define QC Dataset (optional)**

-No QC Selected-   
[?](#)

**3 Parameters**

Select Parameters:

**Job Summary**

Job ID: 1094162  
User: test (16)

Job Name:  [Edit](#)

Datasets: 0  
Parameter Set: 0

**4 Submit**

[Click here to complete your job](#) [Submit Job](#)

## Lets look at XCMS Online in action

- We will run a simple pairwise job
  - Upload two datasets
  - Select a basic parameter set and start the job.
- Set our parameters and launched a job
  - Looking at the parameters and what they mean.
  - Junk in, junk out. – Biologist
  - Good data in, bad parameter selection, junk out – bioinformtictist

## One thing to note

- Choose your polarity correctly!!

The screenshot shows the 'View Parameter Methods/Options' window in XCMS Online. At the top, there are two yellow warning boxes: the first states 'Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.'; the second states 'The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.' Below these are tabs for 'General', 'Feature Detection', 'Retention Time Correction', 'Alignment', 'Statistics', 'Annotation', 'Identification', 'Visualization', and 'Miscellaneous'. The 'General' tab is active, showing a table of parameters:

Option	Value	Note:
Name	HPLC / Q-TOF	
Comment	optimized for HPLC with ~60 min gradient, ESI-Q	
Retention time format	minutes ↕	show the retention times in results tables and figures in minutes or seconds
Polarity	positive ↕	data acquired in positive or negative mode ?

A blue arrow points to the 'positive' dropdown menu in the 'Polarity' row.

## Peak detection choice

### Peak Picking

#### matchedFilter

- Profile Data
- Low resolution data
- Original algorithm

#### centWave

- Centroid data
- High resolution data
- Separately published algorithm

## CentWave parameters

- Peakwidth = How wide is your peak – from a minimum to a maximum in seconds
- Ppm = how much does the peak vary across scans

View Parameter Methods/Options

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous

Method: centWave

Highly sensitive feature detection using a peak density and wavelet based method. Applicable for high resolution LC/MS data in centroid mode.

Option	Value	Note:
ppm	30	maximal tolerated m/z deviation in consecutive scans, in ppm (parts per million)
minimum peak width	10	minimum chromatographic peak width in seconds note: must be less than max peak width. See also <a href="#">here</a> .
maximum peak width	60	maximum chromatographic peak width in seconds note: must be greater than min peak width. See also <a href="#">here</a> .

View Advanced Options

## MatchedFilter parameters

- FWHM=Peakwidth = How wide is your peak
- Step = bin size (in m/z) larger smoother smaller finer (long processing time)

Option	Value	Note:
FWHM	30	full width at half maximum of matched filtration gaussian model peak
step	0.1	step size to use for profile generation

Method: matchedFilter

Note: Find peaks in the chromatographic time domain of the profile matrix using "matched filter" method.

View Advanced Options

## Retention time alignment

View Parameter Methods/Options

ⓘ Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

ⓘ The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

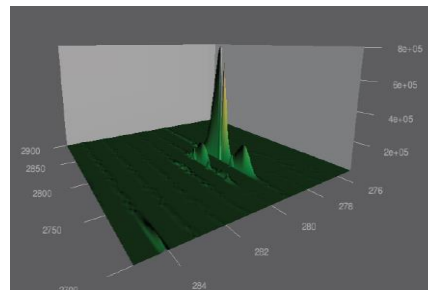
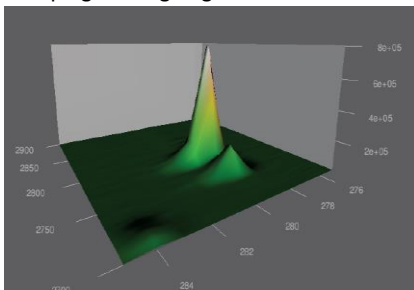
Option	Value	Note:
profStep	0.5	step size (in m/z) to use for profile generation from the raw data files

Method: obiwarp

Note: Retention time correction method based on correlations of the raw data.

### Obiwarp –

A Digital signal processing algorithm. Very good for high drift alignment. Fits data as if each LC-MS 3D landscape was play dough to squeeze these together. Technically this is warping not aligning



# Retention time alignment

General | Feature Detection | Retention Time Correction | Alignment | Statistics | Annotation | Identification | Visualization | Miscellaneous

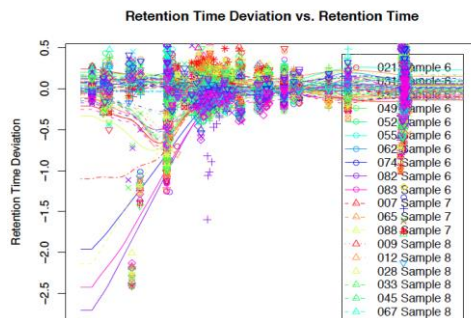
Method: **peakgroups** use "well behaved" peak groups and nonlinear regression to calculate retention time deviations for every time point of each sample.

Note:

Option	Value	Note
missing	1	number of missing samples to allow in retention time correction groups
non-linear/linear alignment	loess	either "loess" for non-linear alignment or "linear" for linear alignment
extra	1	number of extra peaks to allow in retention time correction groups

► View Advanced Options

Loess – this is a regression model to fit the data to using the residuals to correct/align the features. Relies on anchors distributed across the RT.



# Grouping

View Parameter Methods/Options

ⓘ Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

ⓘ The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General | Feature Detection | Retention Time Correction | Alignment | Statistics | Annotation | Identification | Visualization | Miscellaneous

Note:

Option	Value	Note
mzwid	0.025	width of overlapping m/z slices to use for creating peak density chromatograms and grouping peaks across samples
minfrac	0.5	minimum fraction of samples necessary in at least one of the sample groups for it to be a valid group
bw	5	Allowable retention time deviations, in seconds. In more detail: bandwidth (standard deviation or half width at half maximum) of gaussian smoothing kernel to apply to the peak density chromatogram

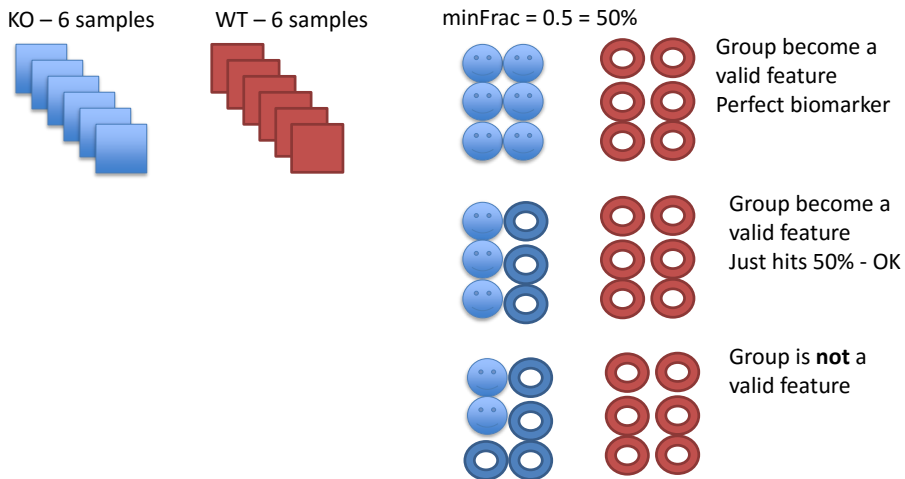
► View Advanced Options

Detected features for m/z:130.1-130.2and rt:305-315

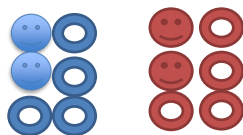


## MinFrac !

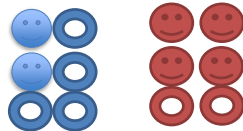
- More questions on minfrac than any other!



## minFrac test

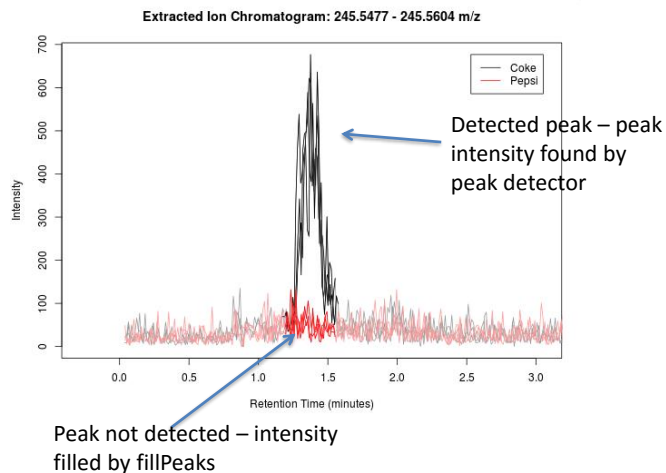


Not a valid feature



A valid feature

# Peak Filling



# Statistics !! Yea !!

View Parameter Methods/Options

**i** Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

**i** The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous

Option	Value	Note:
Statistical test	ANOVA (parametric)	Statistical test method: Welch t-test (unequal variances) or Wilcoxon Rank Sum test
Perform paired test		The selected statistical test is performed as a paired test. The sample pairs need to be specified.
Perform post-hoc analysis	True	Perform post-hoc analysis [multigroup only]
p-value threshold (highly significant features)	0.01	Features with a p-value less than this threshold are considered highly significant. Some statistical figures (e.g. Mirror plot) are generated using only the dysregulated features according to this threshold.
fold change threshold (highly significant features)	1.5	Features with a fold change greater than this threshold are considered highly significant. Some statistical figures (e.g. Mirror plot) are generated using only the dysregulated features according to this threshold.
p-value threshold (significant features)	0.01	Features with a p-value less than this threshold are not considered significant and are omitted from some calculations to save time and space. EIC's, annotations and database ID's are not generated for features with p-values above this threshold.
View Advanced Options		
value	into	intensity values to be used for the diffreport. If value="into", integrated peak intensities are used. If value="maxo", maximum peak intensities are used.
Normalization	None	Normalize the intensity values by either probabilistic quotient or cyclic loess normalization.

# Adduct selection

**View Parameter Methods/Options**

**i** Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

**i** The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous

Option	Value	Note
ppm	10	tolerance for database search
adducts	[M+H] <sup>+</sup> [M+NH <sub>4</sub> ] <sup>+</sup> [M+Na] <sup>+</sup> [M+H <sub>2</sub> O] <sup>+</sup> [M+H <sub>2</sub> H <sub>2</sub> O] <sup>+</sup> [M+K] <sup>+</sup> [M+ACN+H] <sup>+</sup> [M+ACN+Na] <sup>+</sup> [M+2Na-H] <sup>+</sup> [M+2H] <sub>2</sub> <sup>2+</sup>	adducts to be considered for database search

**Multigroup Results Summary: MG\_2015-10-30\_15:52 (#1081196)** [Download Results](#)  
 hash: 6aedaf75ca27d0cce721e40682fc5d77

Submit Date	Finish Date	Total Aligned Features	Parameter ID#	Log	Shared
2015-10-30 15:54:26	2015-10-30 16:30:24	2181	UPLC / TripleTOF pos (769)	<a href="#">View Log</a>	NOT SHARED

**WARNINGS:**

2015-10-30 16:28:50 : iHeatMap data prep\_memory requires limiting to top 1000 features <1.97616e-10 p-values

[View Results Table](#)  
[View Interactive Cloud Plot](#)  
[View Interactive Heatmap](#)  
[View iPCA](#)  
[Connections](#)

**Datasets Used**

- servier\_SUDS4 (126053)
- servier\_SUDS5 (126052)
- servier\_SUDS3 (126051)
- servier\_SUDS2 (126050)
- servier\_SUDS1 (126044)

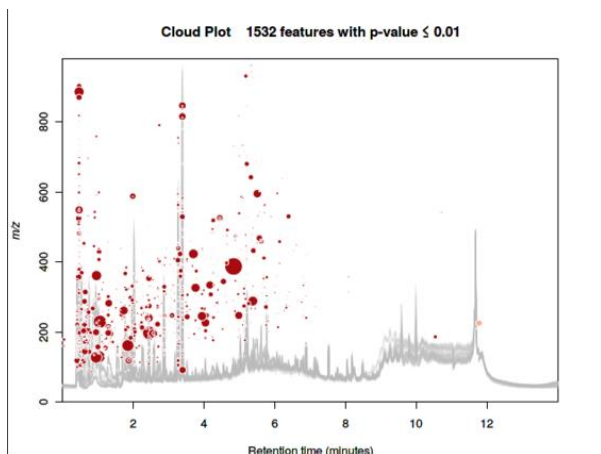







Image not available

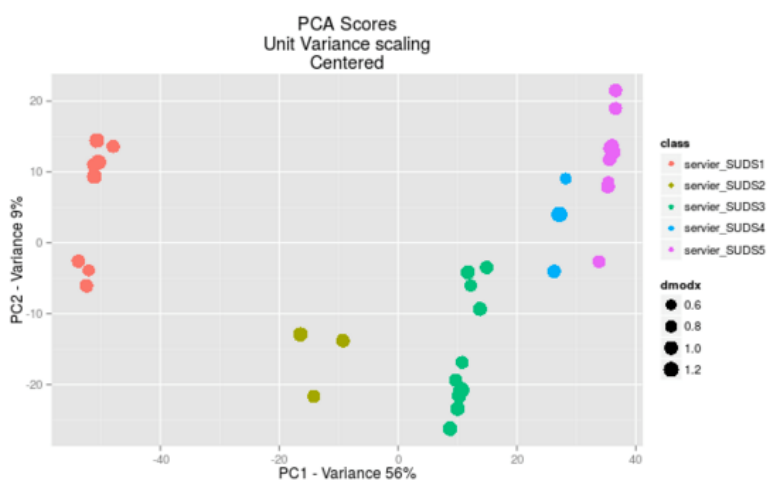
## Cloud plot

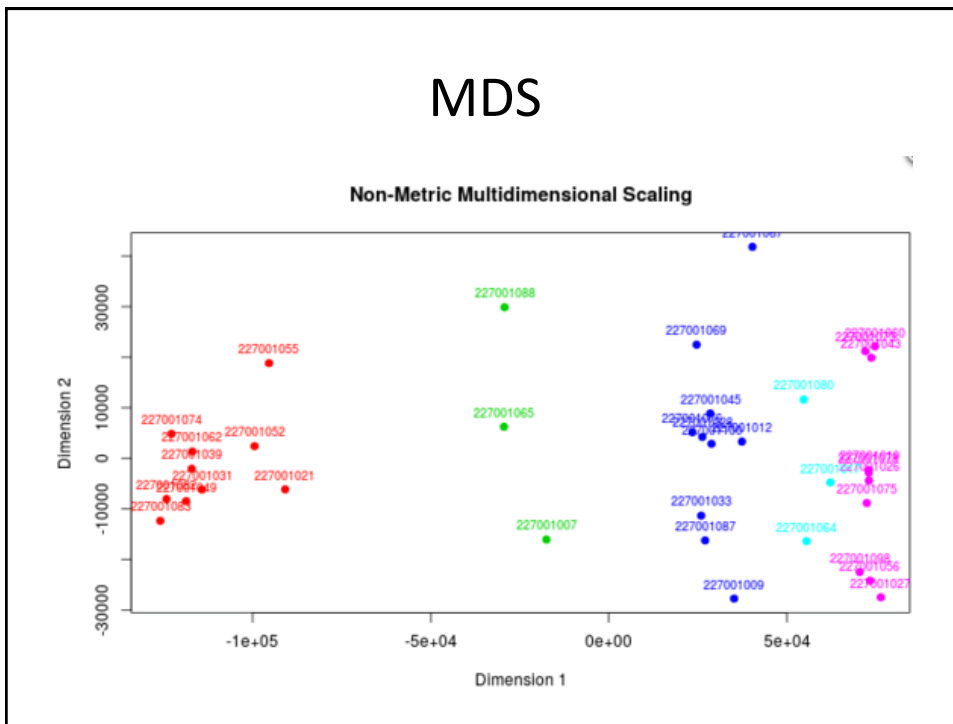


Size = fold change  
Colour = significance (lower p-value)

Black or white ring = METLIN hits

## Static PCA





**Quick Compound Search:**

JOB#1081196 : MG\_2015-10-30\_15-52

Columns:  Hide isotopic peaks

Page 1 of 22

featureidx	pvalue	qvalue	CV	mzmed	rtmed	maxint	isotopes	adducts	peakgro;usemotet		
2	1.60983e-44	5.66866e-42	0.000	437.9477	1.04	601			34		
3	6.51410e-43	1.09802e-40	0.000	184.0087	1.22	313	[M+Na] <sup>+</sup>	161.019	120		
4	7.48726e-43	1.09802e-40	0.000	112.8482	0.50	1,225			67		
5	7.52990e-43	1.09802e-40	0.000	248.1033	1.30	205			40		
6	4.72689e-42	5.74400e-40	0.000	393.0908	0.97	1,287	[127]M <sup>+</sup>		10		
7	6.10534e-42	6.35920e-40	0.000	337.0258	0.94	907	[101]M <sup>+</sup>	[M+H] <sup>+</sup>	336.019	45	
8	1.06735e-41	9.72702e-40	0.000	530.1975	0.74	840	[181]M <sup>+</sup>	[M+H] <sup>+</sup>	529.192	29	
9	2.91179e-41	2.35699e-39	0.000	484.7745	0.49	294		[M+H+K] <sup>2+</sup>	309.657	...	
10	9.26239e-41	6.26167e-39	0.000	104.0060	0.90	246			9	...	
11	9.44697e-41	6.26167e-39	0.000	615.0380	1.04	477		[M+K] <sup>+</sup>	576.072	34	...
12	3.50489e-40	2.12952e-38	0.000	391.0436	4.29	552		[M+Na+K] <sup>2+</sup>	240.202	...	
13	4.47747e-40	2.67789e-38	0.000	152.9906	1.23	720		[M+H-C4H8] <sup>+</sup>	208.189	...	
14	6.44701e-40	3.35754e-38	0.000	114.8458	0.50	624		[M+K] <sup>+</sup>	75.8825	67	...
15	2.84877e-39	1.38128e-37	0.000	346.0878	1.34	240		[M+Na+K] <sup>2+</sup>	210.252	...	
16	3.03117e-39	1.38128e-37	0.000	552.7725	0.48	274		[M+Na+HCOOH] <sup>+</sup>	57	...	
17	7.85705e-39	3.29523e-37	0.000	291.1741	1.57	171			11	...	
18	8.13519e-39	3.29523e-37	0.000	136.9962	1.04	1,481		[M+H] <sup>+</sup>	136.989	34	...
19	1.74891e-38	6.71128e-37	0.000	374.0508	4.26	579		[M+Na] <sup>+</sup>	351.062	206	...
20	2.59815e-38	9.47162e-37	0.000	791.2372	2.75	279			64	...	
21	4.30002e-38	1.49294e-36	0.000	241.9881	0.73	1,602	[47]M <sup>+</sup>	[M+H] <sup>+</sup>	240.962	22	...
22	7.45201e-38	2.46968e-36	0.000	314.9984	4.29	614			202	...	
23	8.73503e-38	2.76903e-36	0.000	407.0122	1.04	3,052	[134]M <sup>+</sup>	[M+Na] <sup>+</sup>	364.022	34	...
24	9.72529e-38	2.95448e-36	0.000	361.0274	0.97	1,180	[117]M <sup>+</sup>	[M+H] <sup>+</sup>	360.023	10	...
25	1.04673e-37	3.05270e-36	0.000	157.0531	1.57	705		[M+Na+HCOOH] <sup>+</sup>	11	...	
26	2.23744e-37	6.27434e-36	0.000	394.0945	0.97	221	[127]M <sup>+</sup>		10	...	
27	2.85440e-37	7.70799e-36	0.000	140.0289	1.01	612		[M+K] <sup>+</sup>	101.066	[M23	...
28	3.03984e-37	7.91559e-36	0.000	282.0830	1.31	548			40	...	
29	3.30610e-37	8.31206e-36	0.000	133.9860	0.90	591			9	...	
30	3.52604e-37	8.56952e-36	0.000	134.0551	1.57	290			11	...	
31	4.94694e-37	1.16350e-35	0.000	591.0363	0.97	302			45	...	
32	5.13232e-37	1.16938e-35	0.000	182.0310	1.38	2,944	[22]M <sup>+</sup>	[M+H] <sup>+</sup>	181.024	55	...

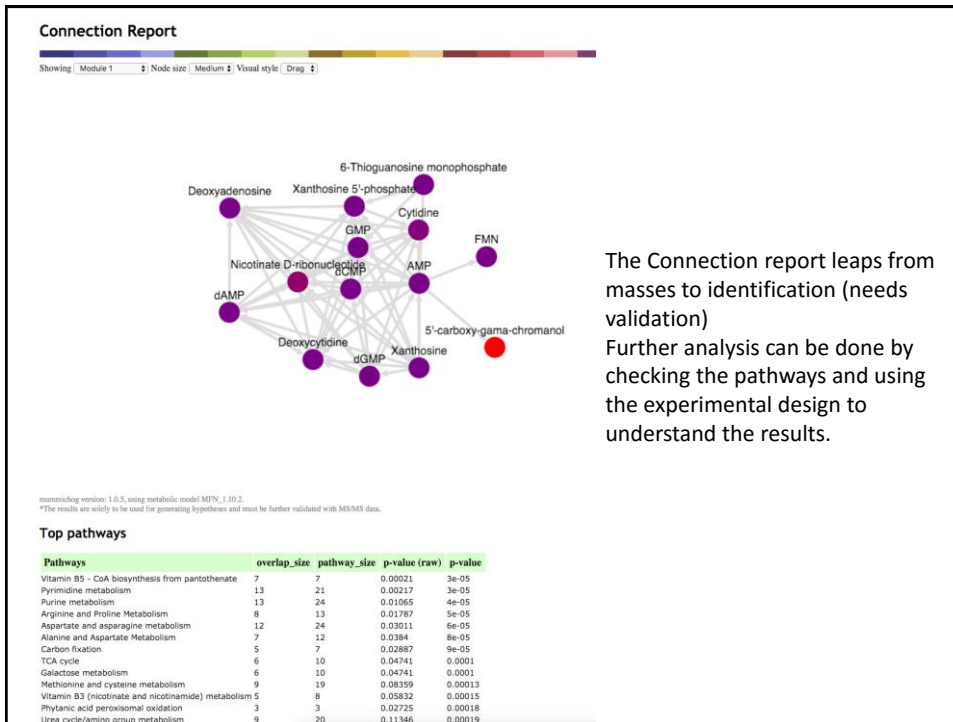
**Feature #8**  
m/z : 530.1975  
Retention Time (min): 0.74

**EIC**

**MASS SPECTRUM/BOXPLOT POSTHOC:**

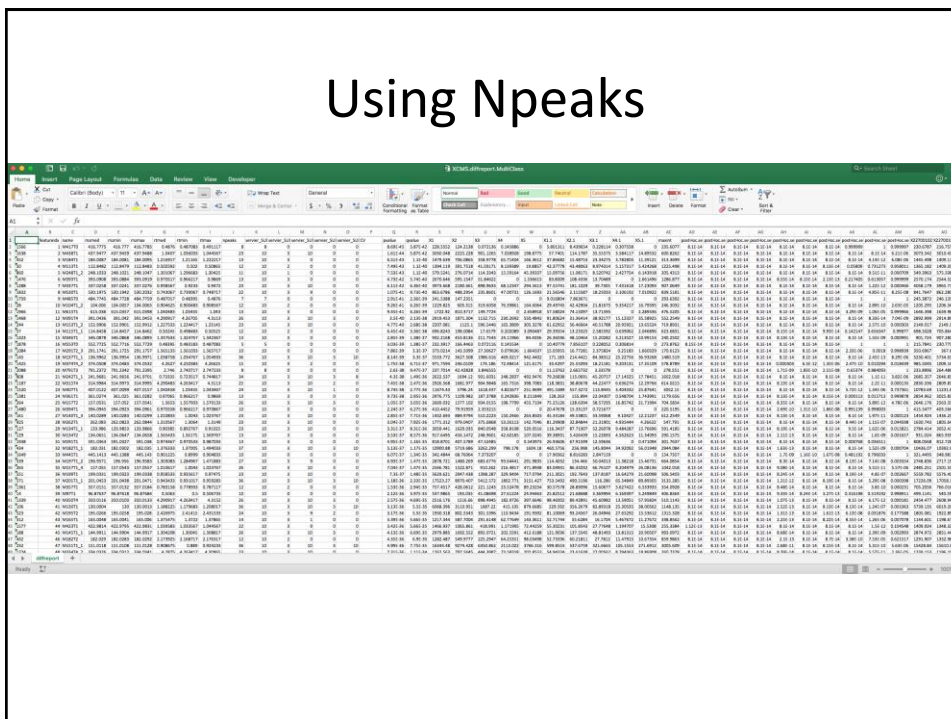
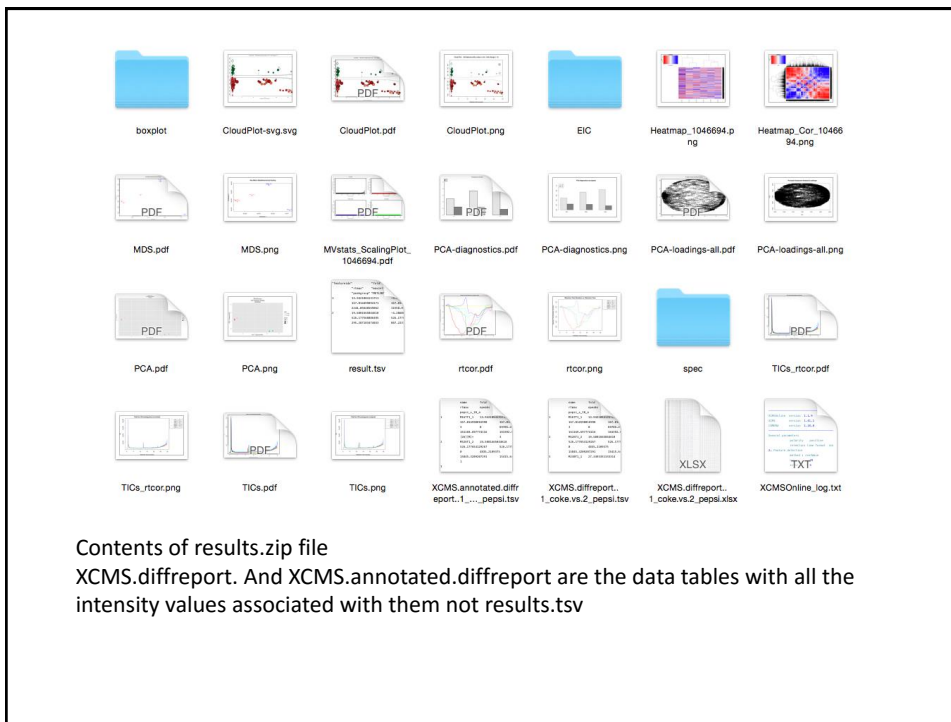
**227001082 (0.74 min)**

PPM	Name	Adduct	COMPOSI
4	M-Hydroxydesmethyl(M-H		3126
4	P-Hydroxydesmethyl(M-H		3127



## Results.zip download file

- This has all of the plots and information from the processed job.
  - Static PCA
  - Static heat map
  - Static cloud plots
  - Scaling plot – Good for looking at scaling for PCA (trend implicates heteroscedastic noise)



## Using Npeaks

D	E	F	G	H	I	J	K	L	M	N	O
mzmed	mzmin	mzmax	rtmed	rtmin	rtmax	npeaks	servier_SUI	servier_SUI	servier_SUI	servier_SUI	servier_SUI
416.7775	416.777	416.7785	0.4876	0.487083	0.491117	8	8	0	0	0	0
437.9477	437.9459	437.9488	1.0407	1.036033	1.044567	23	10	3	10	0	0
184.0087	184.0081	184.0095	1.216917	1.21165	1.222217	13	10	3	0	0	0
112.8482	112.8479	112.8483	0.502592	0.502	0.50865	12	10	2	0	0	0
248.1033	248.1021	248.1047	1.301067	1.296683	1.30425	11	10	1	0	0	0
393.0908	393.0884	393.0919	0.970517	0.966217	0.9869	13	10	3	0	0	0
337.0258	337.0241	337.0276	0.938567	0.9233	0.9472	23	10	3	10	0	0
530.1975	530.1942	530.2032	0.743067	0.739067	0.748717	22	10	3	9	0	0
484.7745	484.7728	484.7759	0.487017	0.48395	0.4876	7	7	0	0	0	0
104.006	104.0057	104.0065	0.904625	0.900483	0.908567	12	10	2	0	0	0
615.038	615.0357	615.0398	1.040483	1.03435	1.043	13	10	3	0	0	0
391.0436	391.042	391.0453	4.290917	4.26705	4.3113	26	10	3	10	3	0
152.9906	152.9901	152.9912	1.227533	1.224417	1.23145	23	10	3	10	0	0
114.8458	114.8457	114.8462	0.50245	0.498483	0.50525	12	10	2	0	0	0
346.0878	346.0868	346.0893	1.337633	1.324767	1.342367	13	10	3	0	0	0
552.7725	552.7716	552.7729	0.48395	0.483183	0.487083	5	5	0	0	0	0
291.1741	291.1725	291.1757	1.565133	1.561033	1.565717	10	10	0	0	0	0
136.9962	136.9954	136.9971	1.038758	1.034767	1.054933	36	10	3	10	3	10
374.0508	374.0483	374.0532	4.2627	4.250383	4.26625	15	10	3	2	0	0
791.2372	791.2342	791.2395	2.746	2.740717	2.747233	8	8	0	0	0	0
241.9681	241.9656	241.9701	0.73335	0.723517	0.744817	34	10	3	10	3	8
314.9984	314.9973	314.9995	4.290483	4.263417	4.3113	25	10	3	10	2	0
407.0122	407.0099	407.0157	1.040458	1.03435	1.043467	24	10	3	10	1	0
361.0274	361.025	361.0282	0.97065	0.966217	0.9869	13	10	3	0	0	0
157.0531	157.052	157.0541	1.5653	1.557933	1.570133	26	10	3	10	3	0
394.0945	394.0923	394.0961	0.970558	0.966217	0.970867	10	10	0	0	0	0
140.0289	140.0283	140.0299	1.010833	1.0043	1.023767	23	10	3	10	0	0
262.083	262.0823	262.0844	1.310567	1.3064	1.3149	23	10	3	10	0	0
133.986	133.9853	133.9866	0.90385	0.892767	0.91025	23	10	3	10	0	0

## Using N peaks

- Npeaks is a very valuable column- informing you of which classes had that feature and how robust the feature was.
- With Excel you can easily find features that are 100% robust across all classes
  - Or find perfect biomarkers (present – absent)



Thank you 😊



Prof. Gary Siuzdak

Questions?



Duane Rinehart

Comments?



Dr. Bill Webb

Thoughts?

## OBI-WARP METHOD

### 1) Alignment by OBI-Warp

