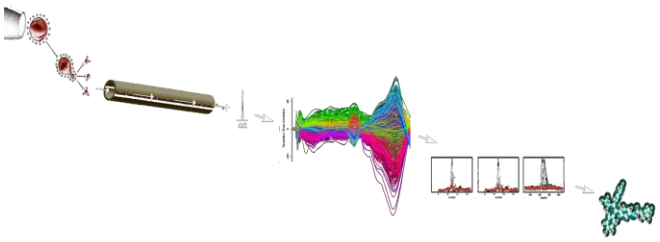
XCMS Online:
Using and understanding



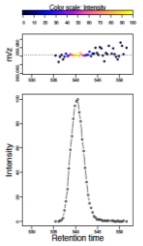
H. Paul Benton PhD
The Siuzdak Laboratory - The Scripps Research Institute

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

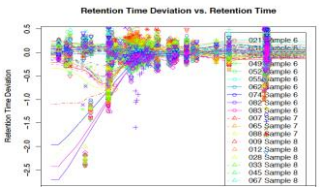
Overview of XCMS

Peak Picking



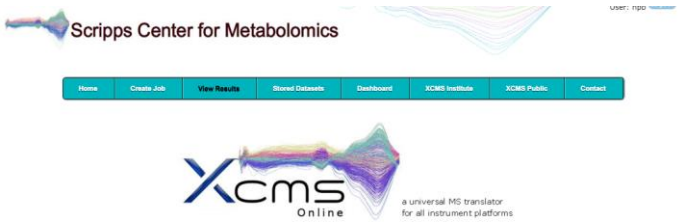
Grouping similar peaks across replicates

Retention time alignment



Statistical analysis of Peaks Between classes

XCMS Online



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:

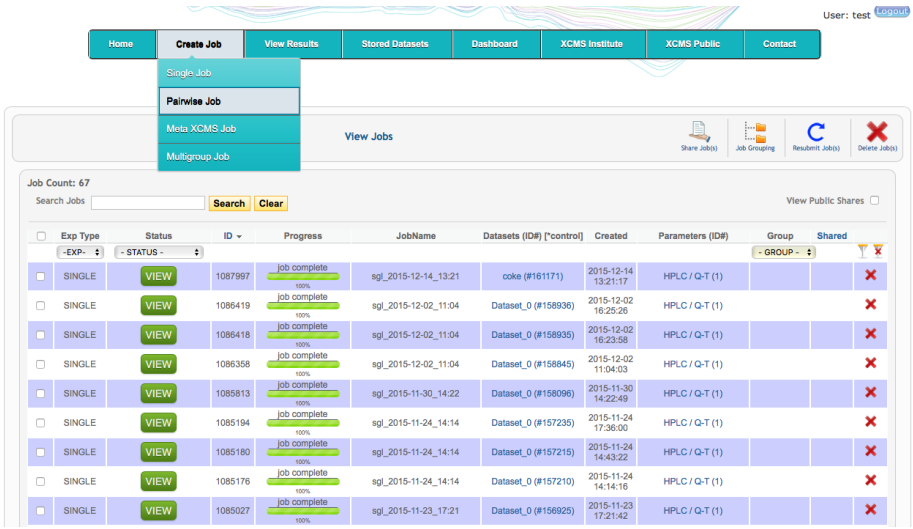
1) uploading the datasets (See File Formats for more information)
2) selecting the parameters and submitting the job.

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

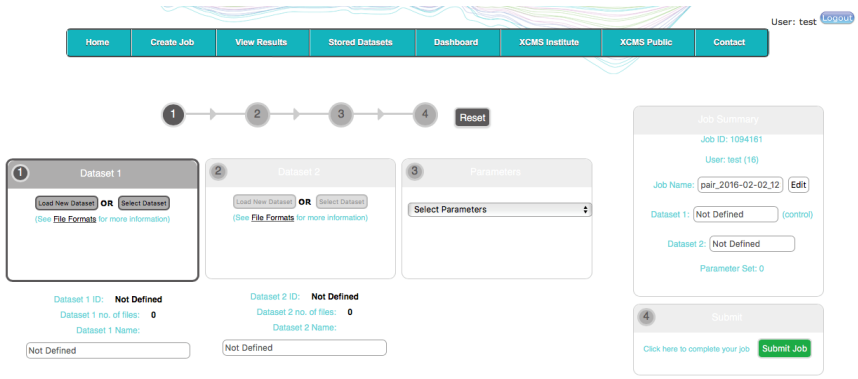
Please click "Create Job" to upload your datasets now or download the user manual for detailed instructions.

Job types in XCMS Online

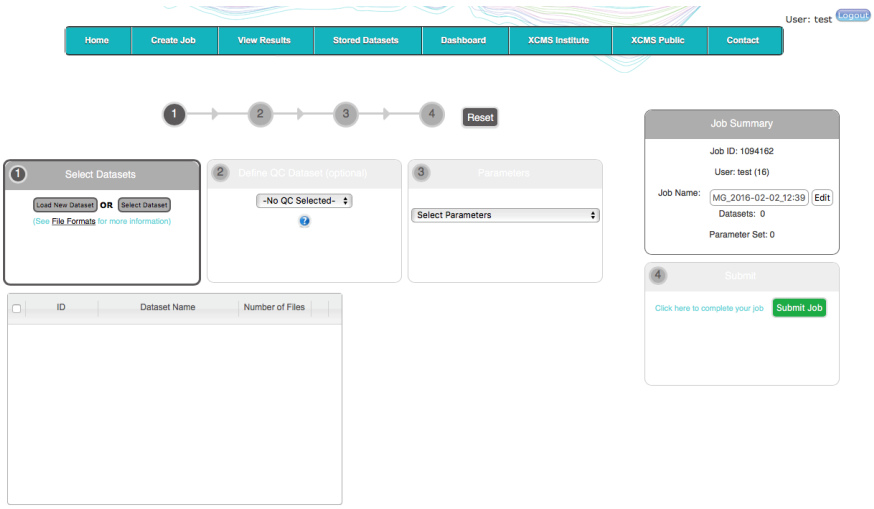
- Single Class Jobs – Either single file or single dataset (class or unknown classes)
- Pairwise analysis Jobs – Two class comparison, works best with KO –vs- WT type experiments
- Multigroup/class Jobs – Multiple classes including Quality control samples. Great for Time series jobs or multiple knockouts.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



Pairwise analysis



Multigroup job



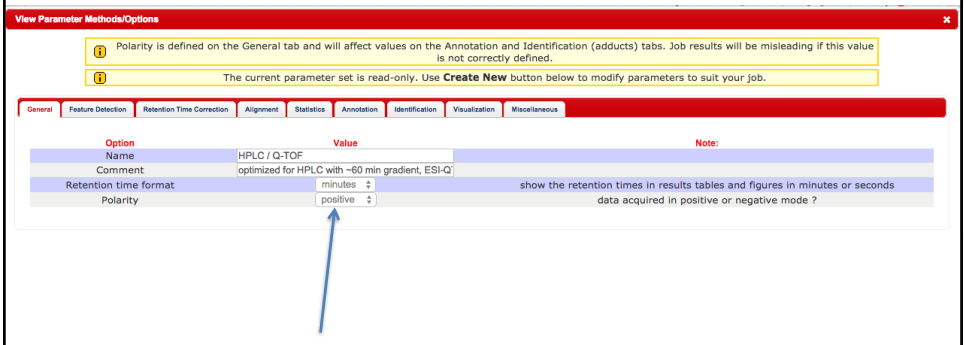
Lets look at XCMS Online in action

- 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 – bioinformticist

One thing to note

- Choose your polarity correctly!!



Peak detection choice

Peak Picking

matchedFilter

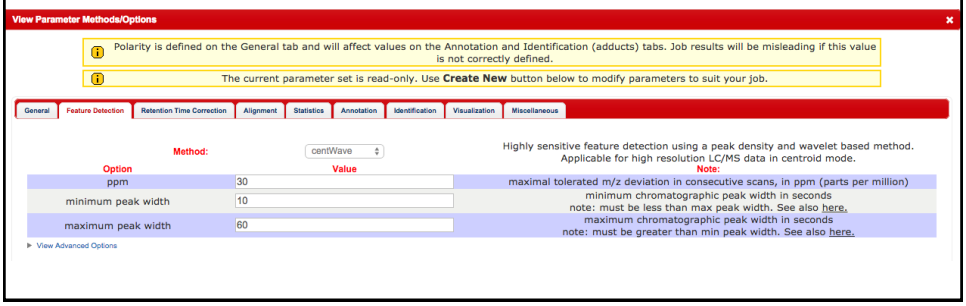
- Profile Data
- Low resolution data
- Original algorithm

centWave

- Centroid data
- High resolution data
- Separately published algorithm

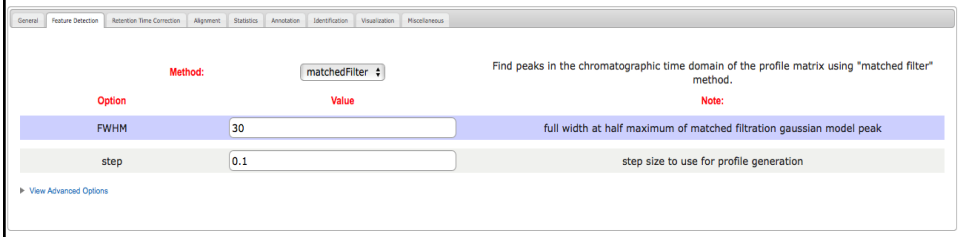
CentWave paramaters

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

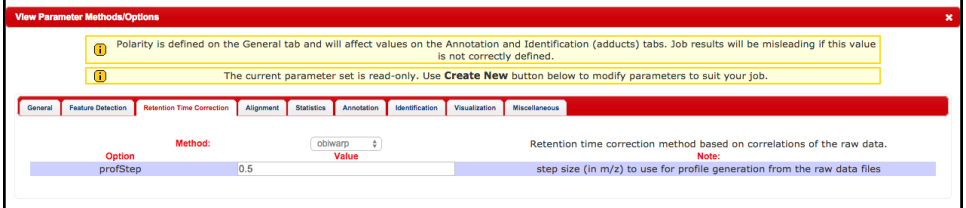


MatchedFilter paramaters

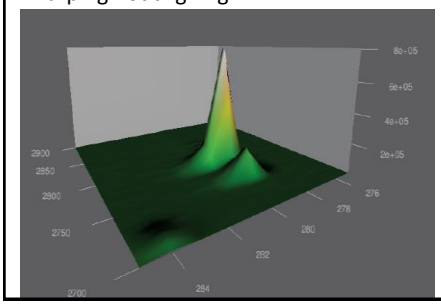
- FWHM=Peakwidth = How wide is your peak
- Step = bin size (in m/z) larger smoother smaller finer (long processing time)

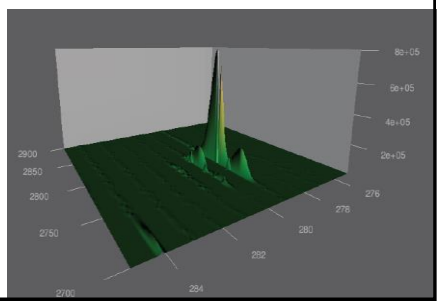


Retention time alignment

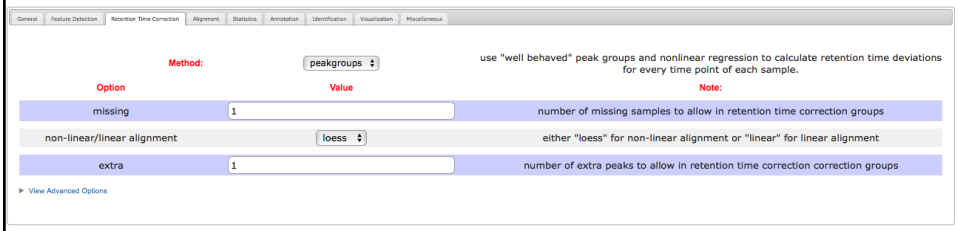


Obiwarp –
A Digital signal processing algorithm. Very good for high drift alignment. Fits data as if each LC-MS 3D landscape was play dough to squeeze these together. Technically this is warping not aligning

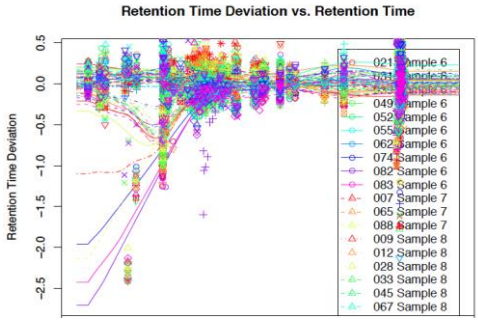




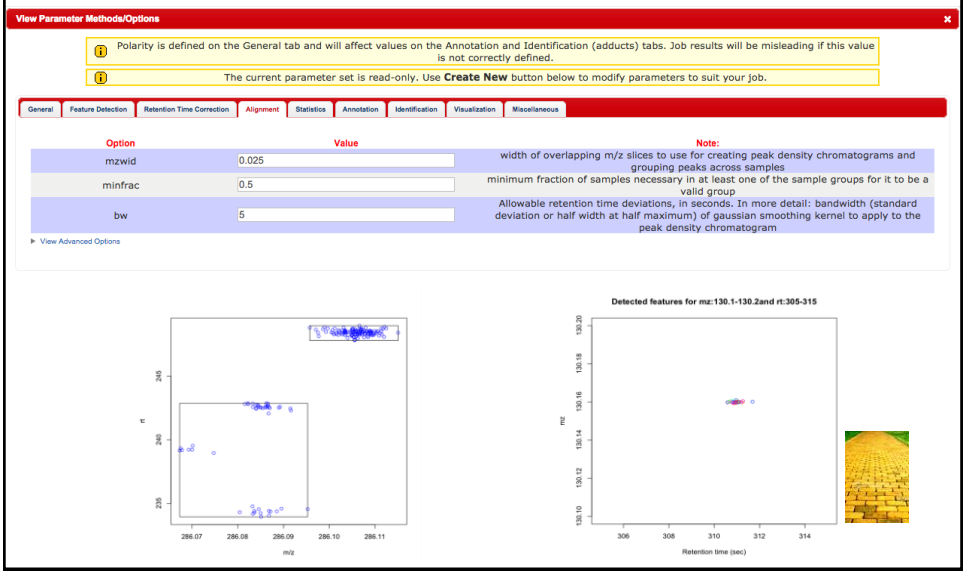
Retention time alignment



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.

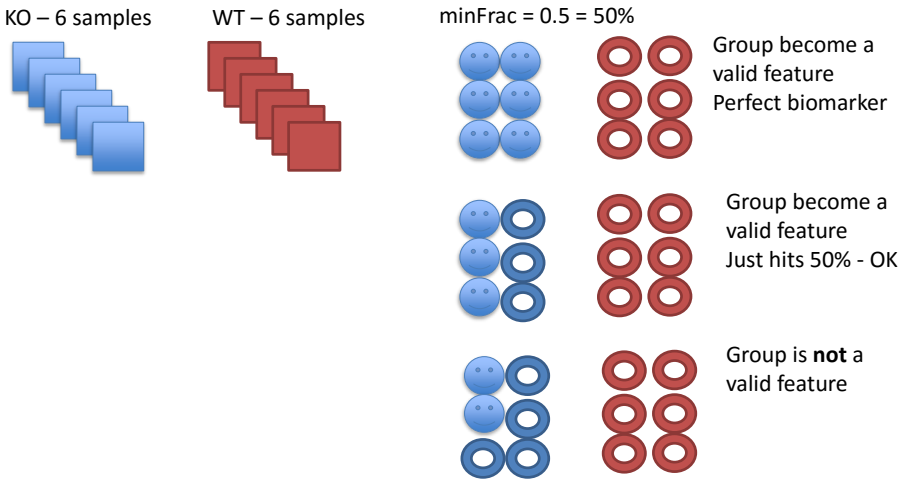


Grouping

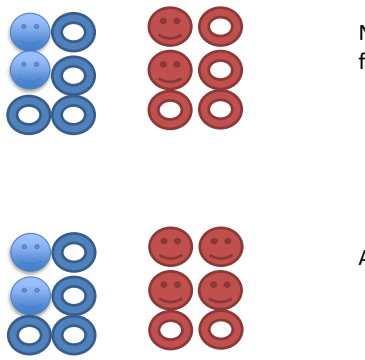


MinFrac !

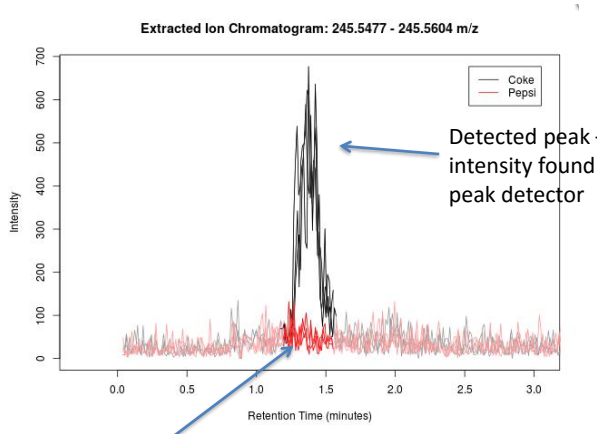
- More questions on minfrac than any other!



minFrac test



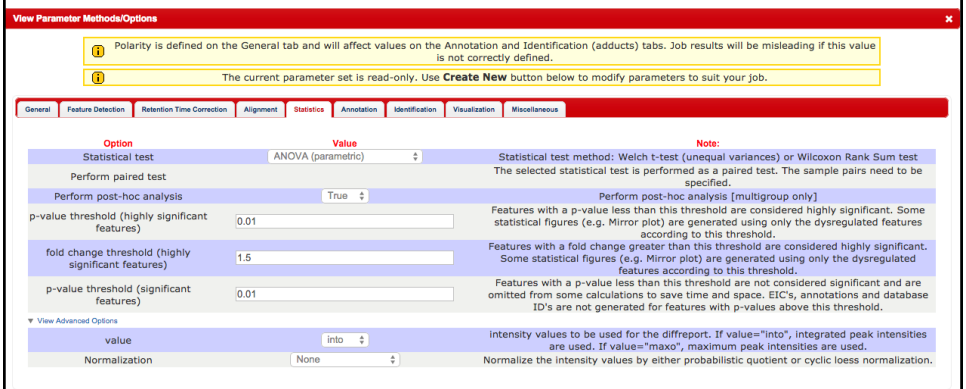
Peak Filling



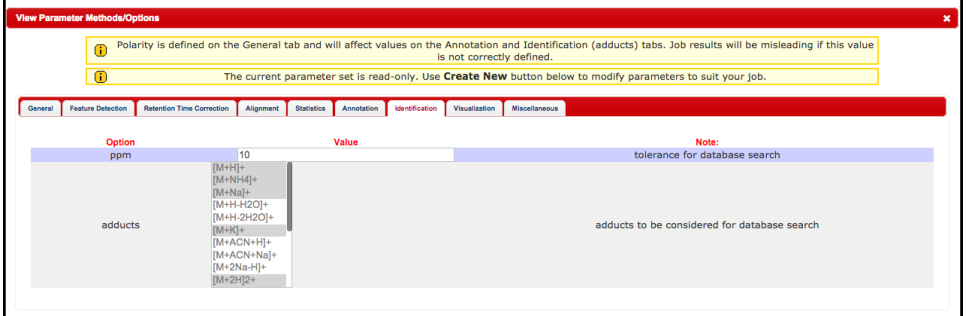
Detected peak – peak intensity found by peak detector

Peak not detected – intensity filled by fillPeaks

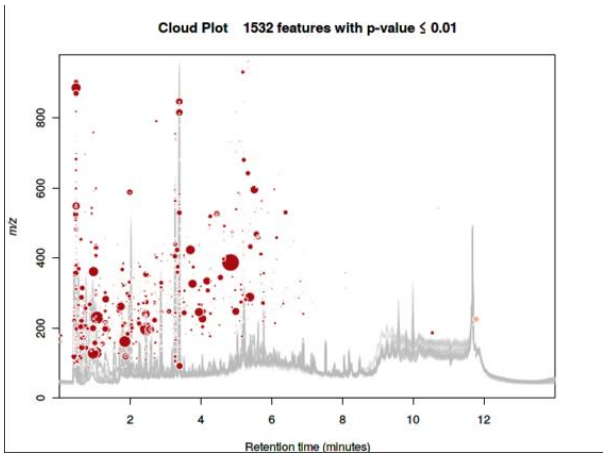
Statistics !! Yea !!



Adduct selection



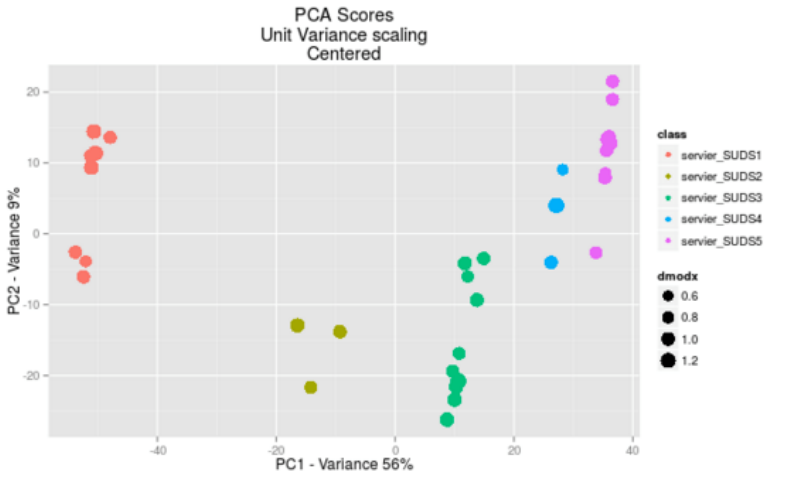
Cloud plot



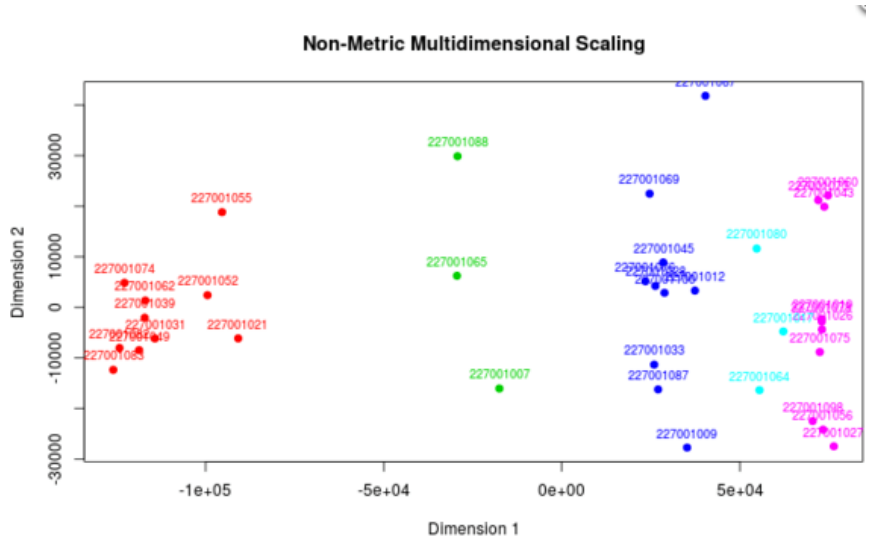
Size = fold change
Colour = significance (lower p-value)

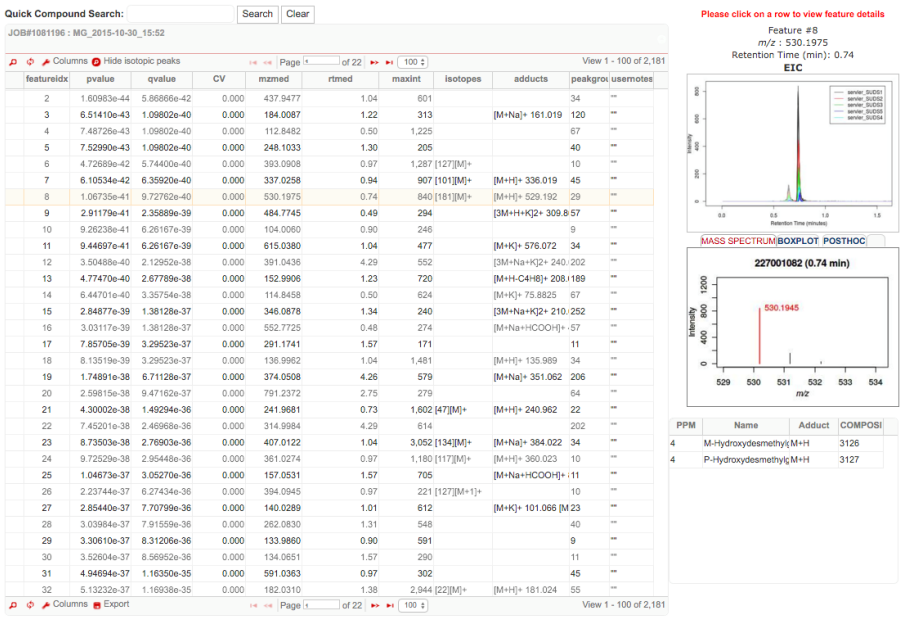
Black or white ring = METLIN hits

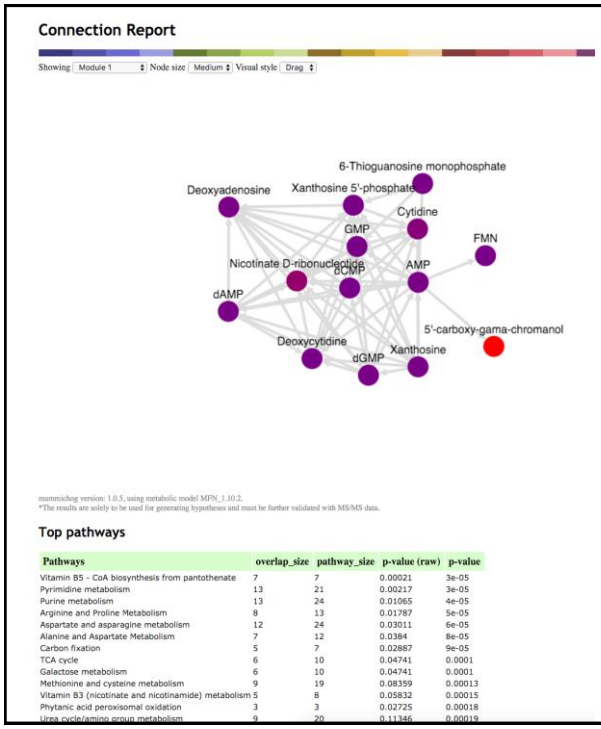
Static PCA



MDS


Non-Metric Multidimensional Scaling
Dimension 2
Dimension 1
30000
10000
0
-10000
-30000
-1e+05
-5e+04
0e+00
5e+04
227001088
227001055
227001069
227001080
227001045
227001065
227001074
227001062
227001052
227001039
227001031
227001021
227001033
227001007
227001087
227001064
227001075
227001009


Quick Compound Search:
Search
Clear
JOB#1081196 : MG_2015-10-30_15:52
Columns
Hide isotopic peaks
Page 1 of 22
100
View 1 - 100 of 2,181
featureidx | pvalue | qvalue | CV | mzmed | rtmed | maxint | isotopes | adducts | peakgroup | usernotes
2 | 1.60983e-44 | 5.86866e-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]+ 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]+ | | 10 |
7 | 6.10534e-42 | 6.35920e-40 | 0.000 | 337.0258 | 0.94 | 907 | [101][M]+ | [M+H]+ 336.019 | 45 |
8 | 1.06735e-41 | 9.72762e-40 | 0.000 | 530.1975 | 0.74 | 840 | [181][M]+ | [M+H]+ 529.192 | 29 |
9 | 2.91179e-41 | 2.35889e-39 | 0.000 | 484.7745 | 0.49 | 294 | | [3M+H+K]2+ 309.8 | 57 |
10 | 9.26238e-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]+ 576.072 | 34 |
12 | 3.50488e-40 | 2.12952e-38 | 0.000 | 391.0436 | 4.29 | 552 | | [3M+Na+K]2+ 240. | 202 |
13 | 4.77470e-40 | 2.67789e-38 | 0.000 | 152.9906 | 1.23 | 720 | | [M+H-C4H8]+ 208. | 189 |
14 | 6.44701e-40 | 3.35754e-38 | 0.000 | 114.8458 | 0.50 | 624 | | [M+K]+ 75.8825 | 67 |
15 | 2.84877e-39 | 1.38128e-37 | 0.000 | 346.0878 | 1.34 | 240 | | [3M+Na+K]2+ 210. | 252 |
16 | 3.03117e-39 | 1.38128e-37 | 0.000 | 552.7725 | 0.48 | 274 | | [M+Na+HCOOH]+ | 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]+ 135.989 | 34 |
19 | 1.74891e-38 | 6.71128e-37 | 0.000 | 374.0508 | 4.26 | 579 | | [M+Na]+ 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 | 241.9681 | 0.73 | 1,602 | [47][M]+ | [M+H]+ 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]+ | [M+Na]+ 384.022 | 34 |
24 | 9.72529e-38 | 2.95448e-36 | 0.000 | 361.0274 | 0.97 | 1,180 | [117][M]+ | [M+H]+ 360.023 | 10 |
25 | 1.04673e-37 | 3.05270e-36 | 0.000 | 157.0531 | 1.57 | 705 | | [M+Na+HCOOH]+ | 11 |
26 | 2.23744e-37 | 6.27434e-36 | 0.000 | 394.0945 | 0.97 | 221 | [127][M+1]+ | | 10 |
27 | 2.85440e-37 | 7.70799e-36 | 0.000 | 140.0289 | 1.01 | 612 | | [M+K]+ 101.066 [M | 23 |
28 | 3.03984e-37 | 7.91559e-36 | 0.000 | 262.0830 | 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 | 134.0651 | 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]+ | [M+H]+ 181.024 | 55 |
Columns
Export
Page 1 of 22
100
View 1 - 100 of 2,181
Please click on a row to view feature details
Feature #8
m/z : 530.1975
Retention Time (min): 0.74
EIC
MASS SPECTRUM
BOXPLOT
POSTHOC
227001082 (0.74 min)
530.1945
Intensity
m/z
PPM | Name | Adduct | COMPOSI
4 | M-Hydroxydesmethyl(| M+H | 3126
4 | P-Hydroxydesmethyl(| M+H | 3127

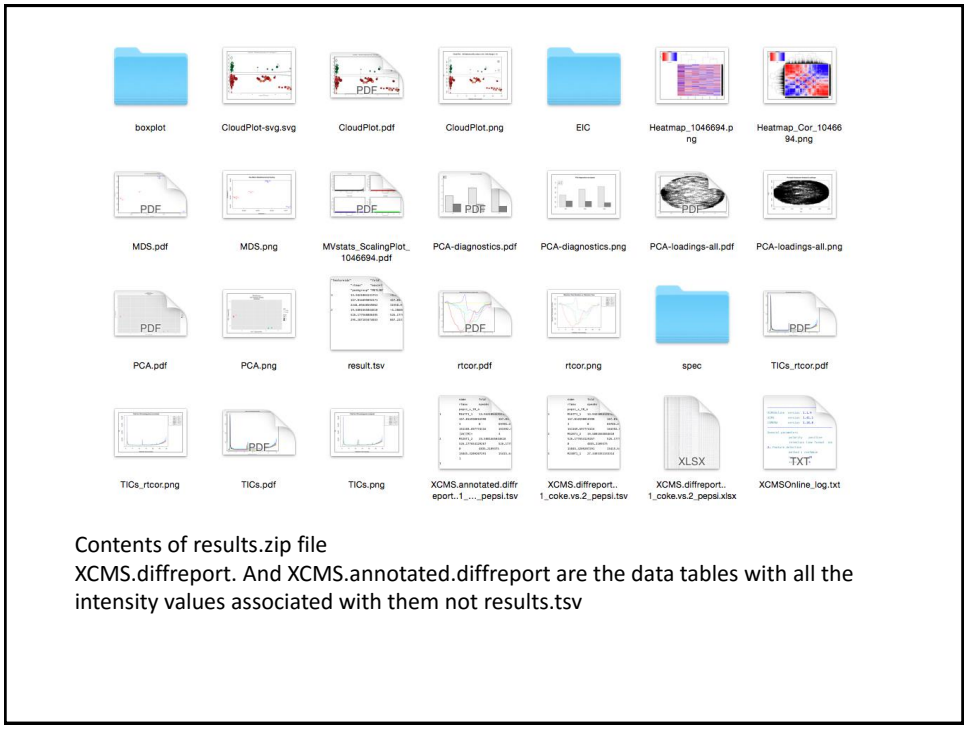


Pathways	overlap_size	pathway_size	p-value (raw)	p-value
Vitamin B5 - CoA biosynthesis from pantothenate	7	7	0.00021	3e-05
Pyrimidine metabolism	13	21	0.00217	3e-05
Purine metabolism	13	24	0.01065	4e-05
Arginine and Proline Metabolism	8	13	0.01787	5e-05
Aspartate and asparagine metabolism	12	24	0.03011	6e-05
Alanine and Aspartate Metabolism	7	12	0.0384	8e-05
Carbon fixation	5	7	0.02887	9e-05
TCA cycle	6	10	0.04741	0.0001
Galactose metabolism	6	10	0.04741	0.0001
Methionine and cysteine metabolism	9	19	0.08359	0.00013
Vitamin B3 (nicotinate and nicotinamide) metabolism	5	8	0.05832	0.00015
Phytanic acid peroxisomal oxidation	3	3	0.02725	0.00018
Urea cycle/amino group metabolism	9	20	0.11346	0.00019

The Connection report leaps from masses to identification (needs validation)
Further analysis can be done by checking the pathways and using the experimental design to understand the results.

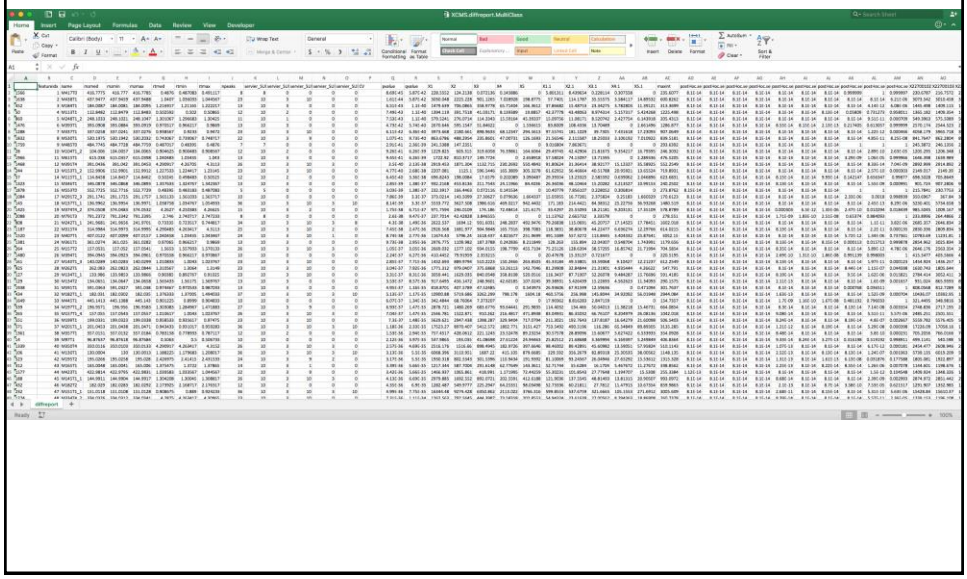
Results.zip download file

- 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)



Contents of results.zip file
XCMS.diffreport. And XCMS.annotated.diffreport are the data tables with all the intensity values associated with them not results.tsv

Using Npeaks



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