Knowledge that will change your world

Preparing data for upload to XCMSonline

Stephen Barnes, PhD

University of Alabama at Birmingham

sbarnes@uab.edu

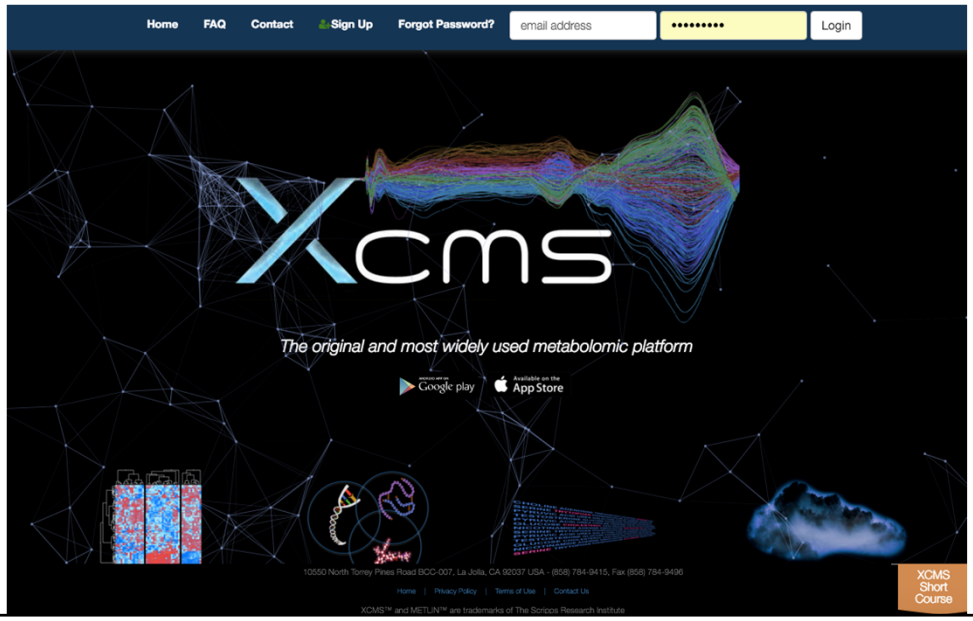
1

Synopsis

- **LC-MS (and GC-MS) analysis generates a lot of data and requires alignment of individual data sets before statistical analysis can be performed**
- **We will discuss**
  - **Uploading data sets**
  - **Alignment principles**

2

Go to https://xcmsonline.scripps.edu/



3

Starting page for XCMSonline

4

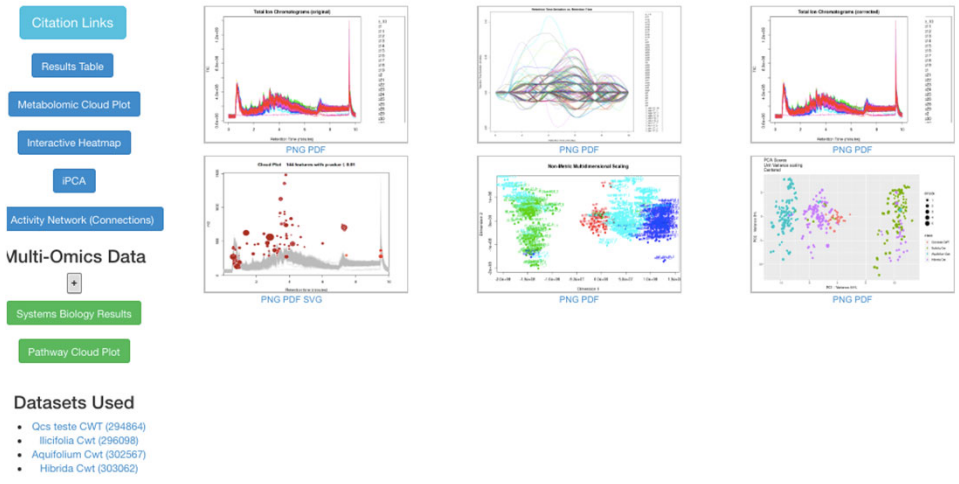
View Public Shares page 3

Status	ID	Job Name	Biosource	Author(s)	Publication	Notes
VIEW	1270779	Asthma vs COPD (uncorrected mzML files)	HUMAN			
VIEW	1271289	Asthma vs COPD (corrected mzML files)	HUMAN			
VIEW	1197171	pair_2018-01-18ï¼é¡¿ç²åŠ çç±é¡¿æ²â€"ç"Ÿç‰©©çç±	HUMAN			
VIEW	1275031	SGL_2018-12-02_23:14	HUMAN			
VIEW	1276381	P_2018-12-07_02:59	HUMAN			
VIEW	1181343	sazonalidade_fibra	HUMAN			
VIEW	1155786	agua	HUMAN			
VIEW	1155771	luz	HUMAN			
VIEW	1152766	temperature	HUMAN			
VIEW	1251319	Untargeted analysis of medicinal species: Maytenus ilicifolia, Maytenus aquifolium and their hybrid	HUMAN			

Showing 21 to 30 of 31 entries

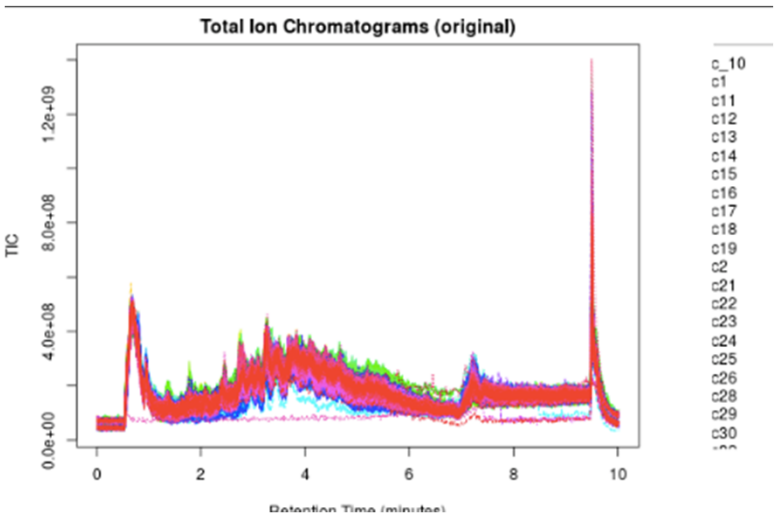
5

First look at the results of the analysis



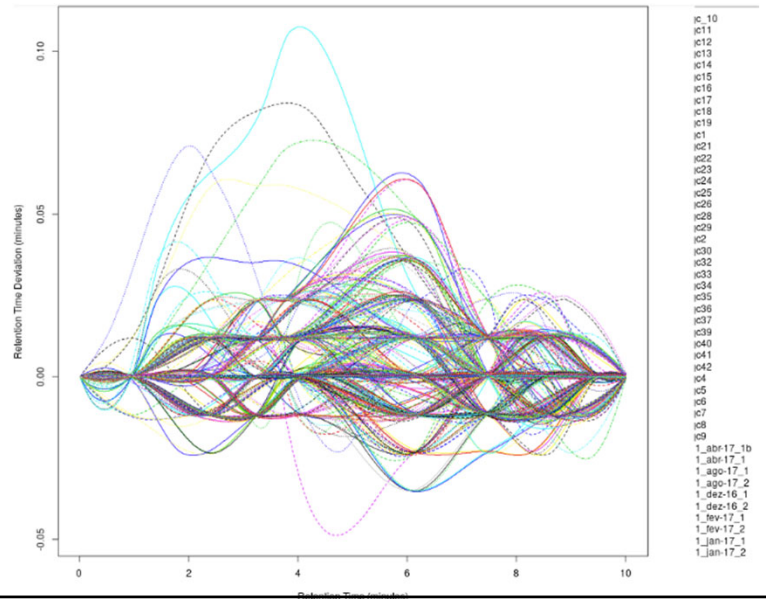
6

Overlay of all the TICs for each sample



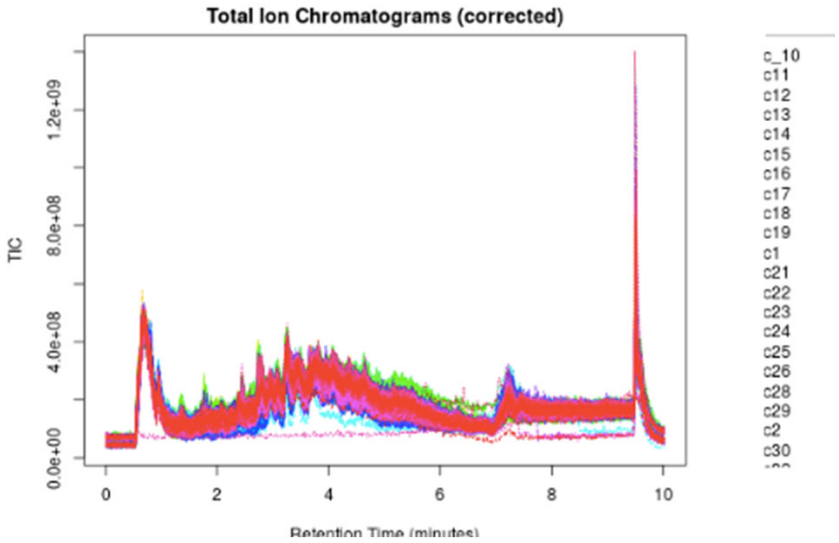
7

Retention time variation of peaks



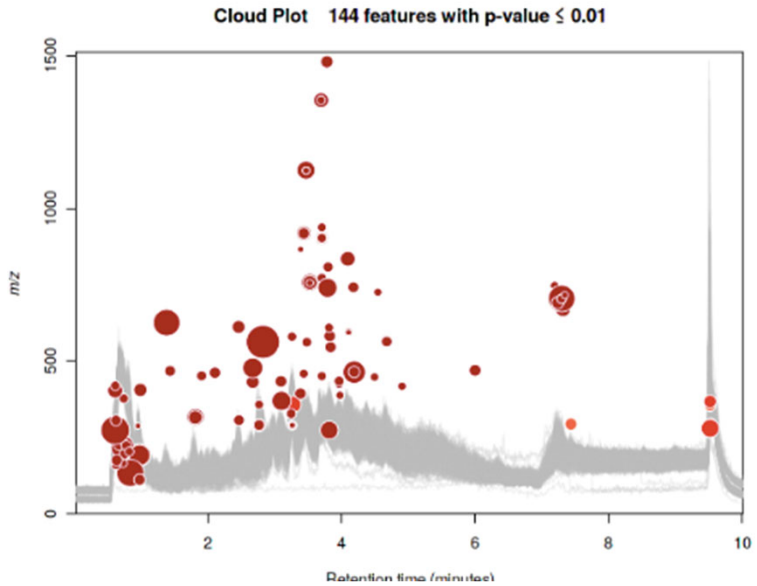
8

Corrected retention time



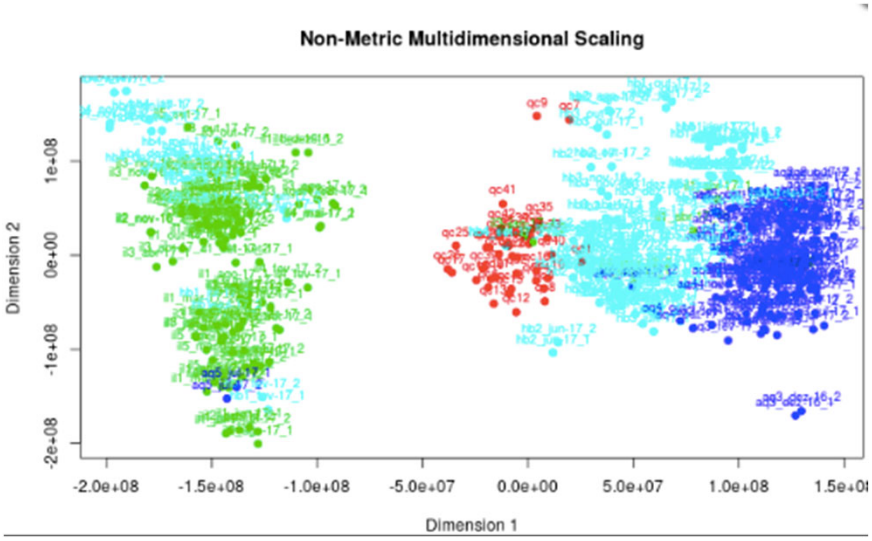
9

Cloud plot of the data



10

Scaling plot show plant differences



11

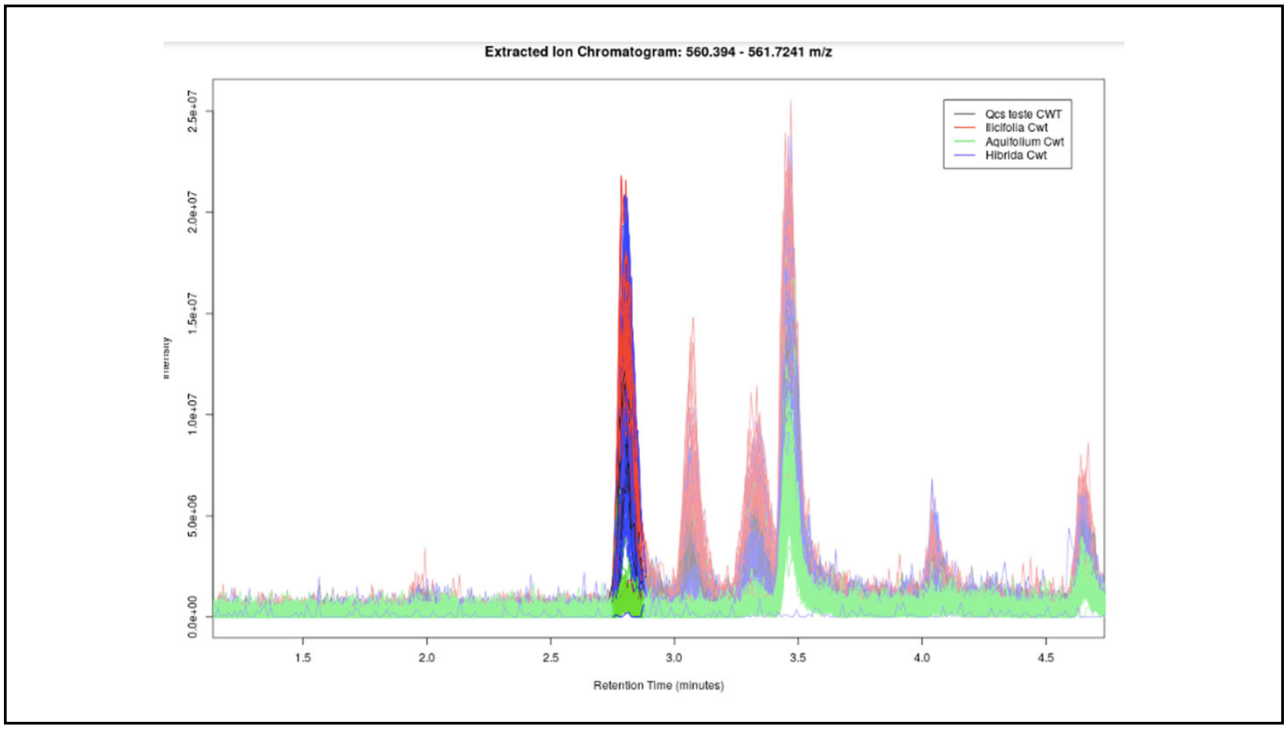
Results – most significant ions

JOB#1251319 : UNTARGETED ANALYSIS OF MEDICINAL SPECIES: MAYTENUS ILICIFOLIA, MAYTENUS AQUIFOLIU

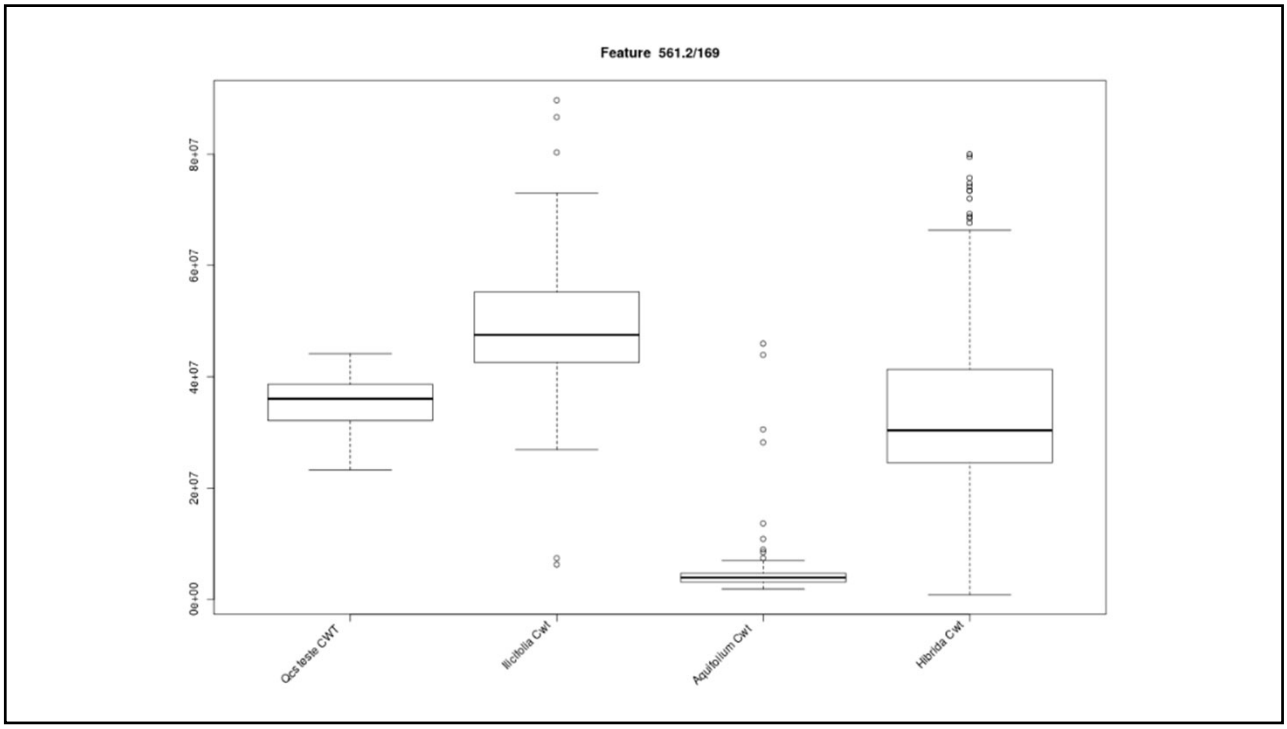
Columns | Hide isotopic peaks | Page 1 of 2 | 100

featureidx	pvalue	qvalue	CV	mzmed	rtmed	maxint
1	6.03291e-56	9.36519e-56	0.149	545.2159	3.83	8,510,464
2	8.82338e-56	9.36519e-56	0.143	273.1025	3.81	6,872,576
3	5.84188e-54	3.29595e-54	0.142	562.1447	2.81	12,175,360
4	6.61697e-54	3.51164e-54	0.135	561.1901	2.81	21,801,984
5	5.62525e-53	2.38827e-53	0.159	563.1780	4.67	10,763,264
6	1.98850e-52	7.03534e-53	0.106	561.1982	3.47	23,953,408
7	7.51842e-52	1.82616e-52	0.129	834.2289	4.09	9,391,104
8	7.53218e-52	1.82807e-52	0.137	833.3131	4.09	13,384,704
9	8.04366e-52	1.89724e-52	0.104	739.2925	3.78	29,903,872
10	2.90585e-51	6.16858e-52	0.111	740.2017	3.78	16,901,120
11	6.20722e-51	1.19789e-51	0.194	715.4036	7.34	4,780,288
12	4.37856e-50	7.62488e-51	0.107	562.1453	3.47	11,867,136
13	1.11941e-49	1.65610e-50	0.222	1,123.4251	3.46	26,100,736
14	1.31254e-49	1.86243e-50	0.169	564.1507	4.67	5,573,376
15	1.32503e-49	1.87519e-50	0.129	741.0460	3.78	8,337,408
16	1.83722e-49	2.43754e-50	0.183	834.4048	4.09	8,310,272
17	3.37369e-49	4.10225e-50	0.208	1,124.2938	3.46	10,603,520
18	3.94484e-49	4.65230e-50	0.219	1,123.6221	3.46	11,350,016

12

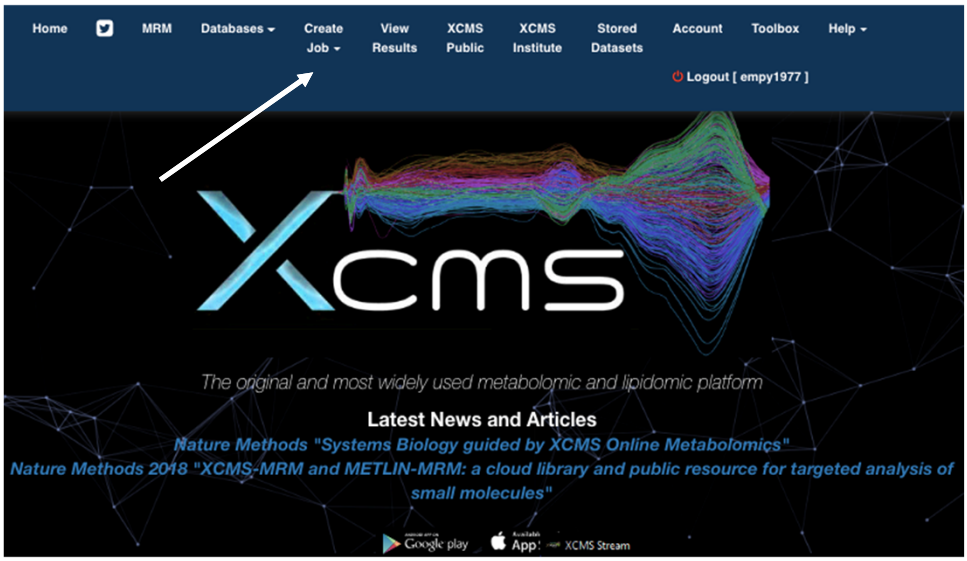

Extracted Ion Chromatogram: 560.394 - 561.7241 m/z
Qcs teste CWT
Ilicifolia Cwt
Aquifolium Cwt
Hibrida Cwt
Intensity
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
1.5
2.0
2.5
3.0
3.5
4.0
4.5
Retention Time (minutes)

13


Feature 561.2/169
0e+00
2e+07
4e+07
6e+07
8e+07
Qcs teste CWT
Ilicifolia Cwt
Aquifolium Cwt
Hibrida Cwt

14

Go back to home page and create a job

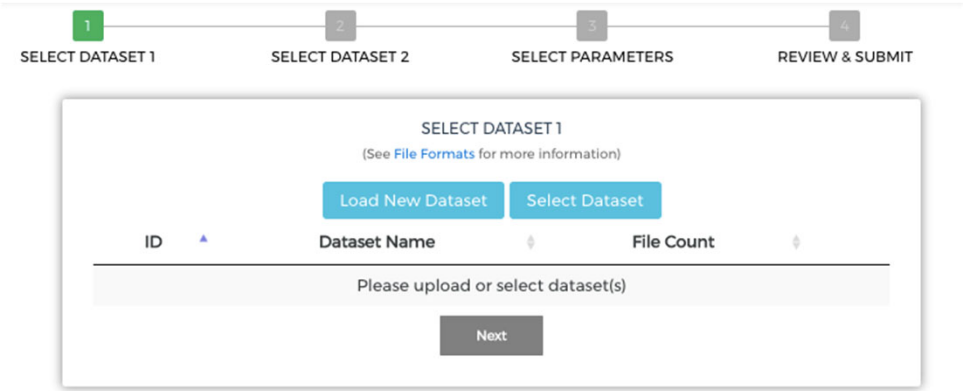


15

Select pairwise

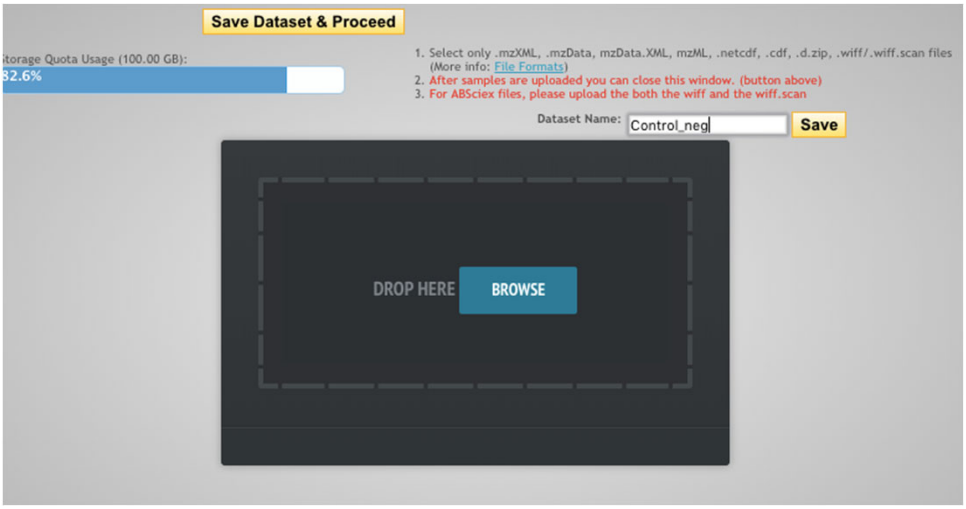
16

Let's load the control dataset



17

Name the control_neg data set



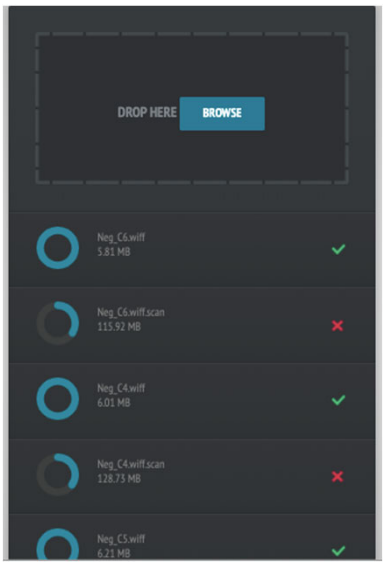
18

Locate the control files

Name	Size	Kind
Neg_C4.abf	624 MB	Document
Neg_C4.mzXML	1.6 GB	TextEdit
Neg_C4.wiff	6 MB	TextEdit
Neg_C4.wiff.scan	128.7 MB	Document
Neg_C5.abf	588.3 MB	Document
Neg_C5.mzXML	1.52 GB	TextEdit
Neg_C5.wiff	6.2 MB	TextEdit
Neg_C5.wiff.scan	121 MB	Document
Neg_C6.abf	557.5 MB	Document
Neg_C6.mzXML	1.44 GB	TextEdit
Neg_C6.wiff	5.8 MB	TextEdit
Neg_C6.wiff.scan	115.9 MB	Document

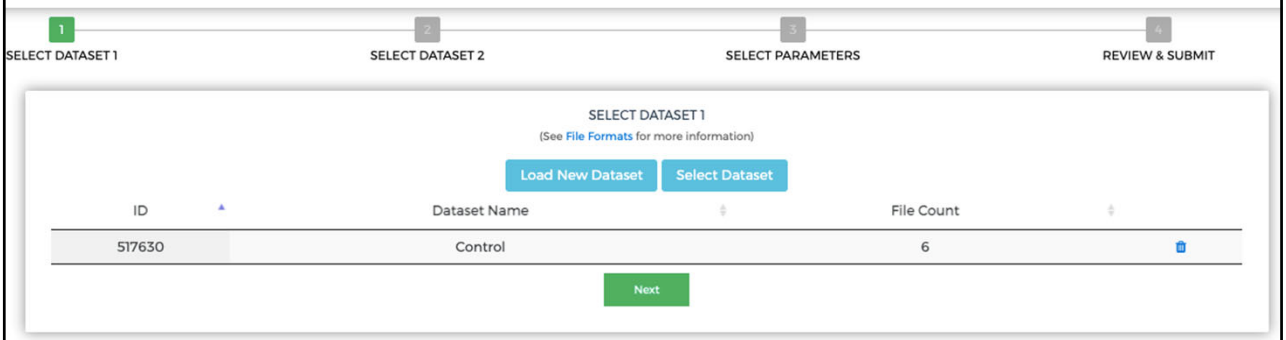
19

Showing progress of uploading



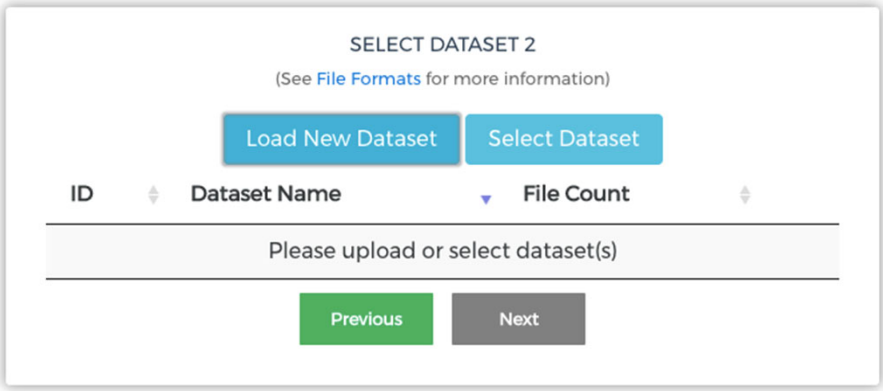
20

Moving on to the next dataset



21

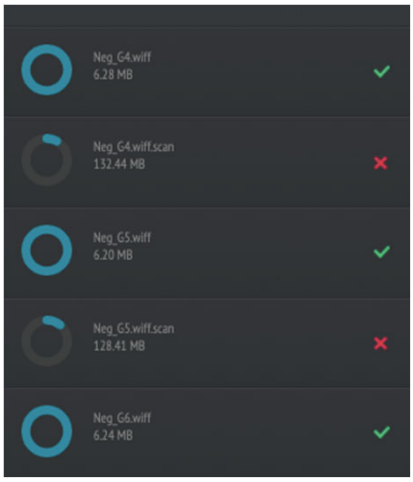
Load the second set of data



22

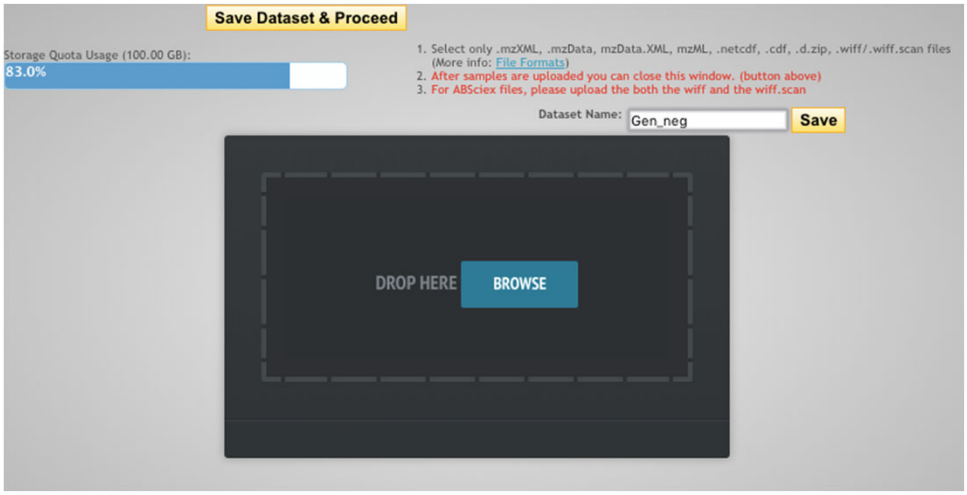
And the genistein group

Neg_G4.wiff	6.3 MB	TextEdit
Neg_G4.wiff.scan	132.4 MB	Document
Neg_G5.abf	621.9 MB	Document
Neg_G5.mzXML	1.6 GB	TextEdit
Neg_G5.wiff	6.2 MB	TextEdit
Neg_G5.wiff.scan	128.4 MB	Document
Neg_G6.abf	562.7 MB	Document
Neg_G6.mzXML	1.48 GB	TextEdit
Neg_G6.wiff	6.2 MB	TextEdit
Neg_G6.wiff.scan	116.9 MB	Document



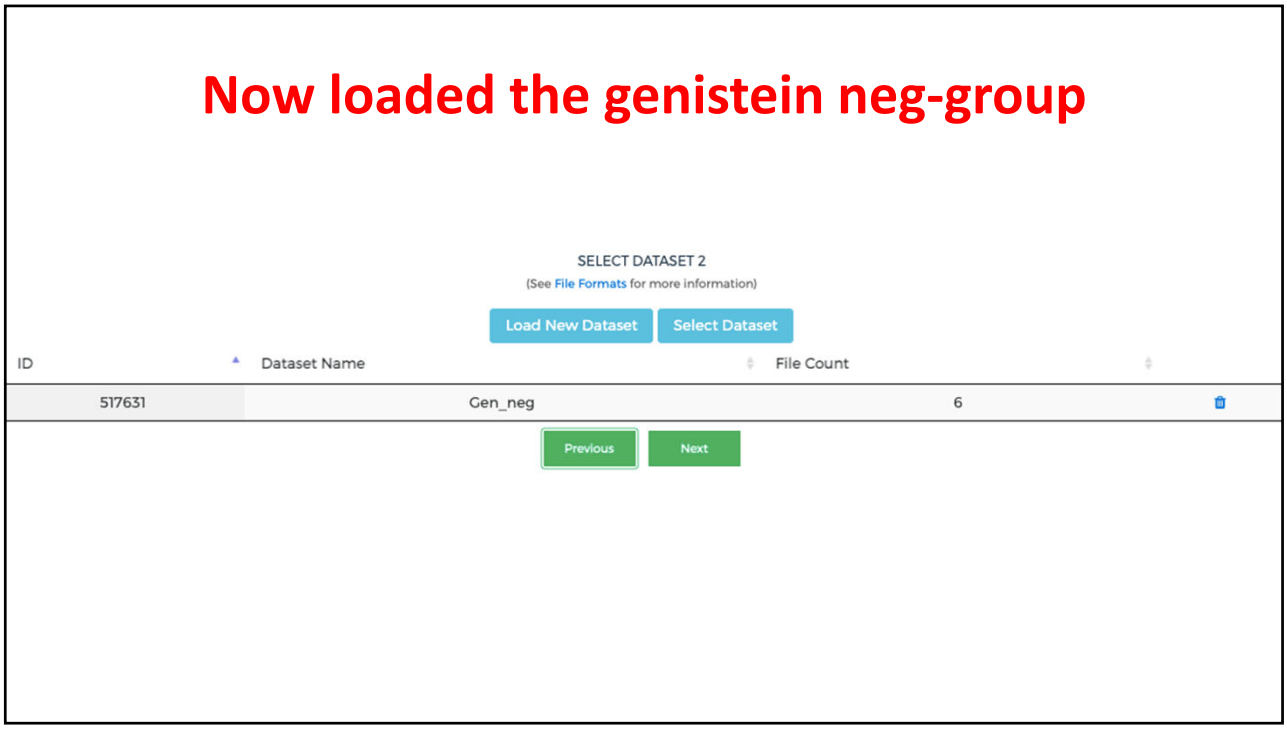
23

Name the Gen_neg data set



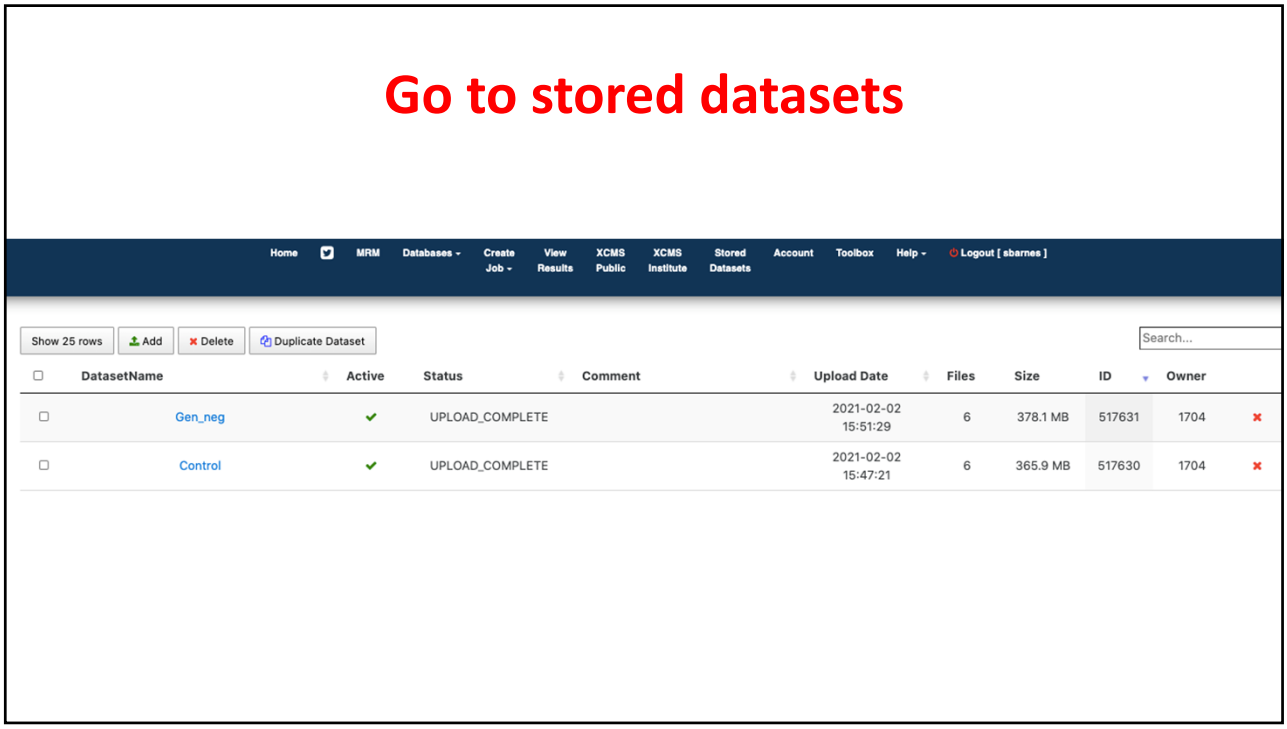
24

Now loaded the genistein neg-group

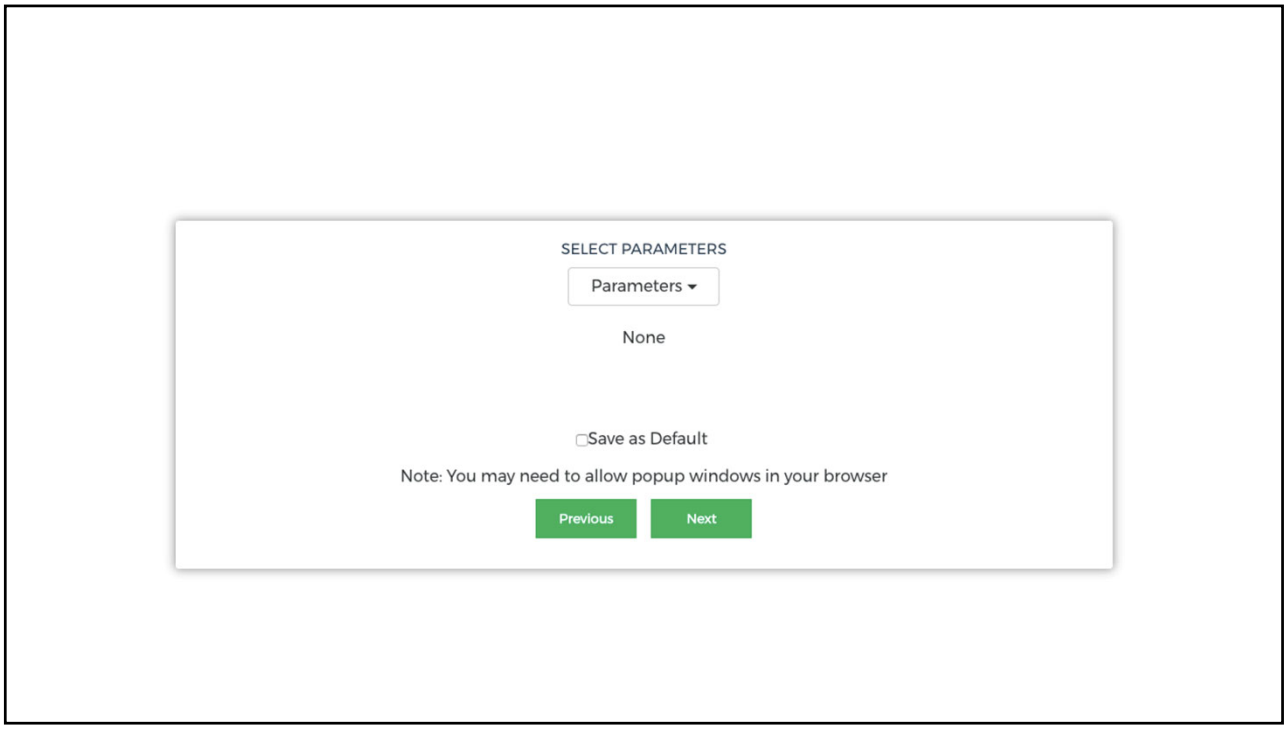


25

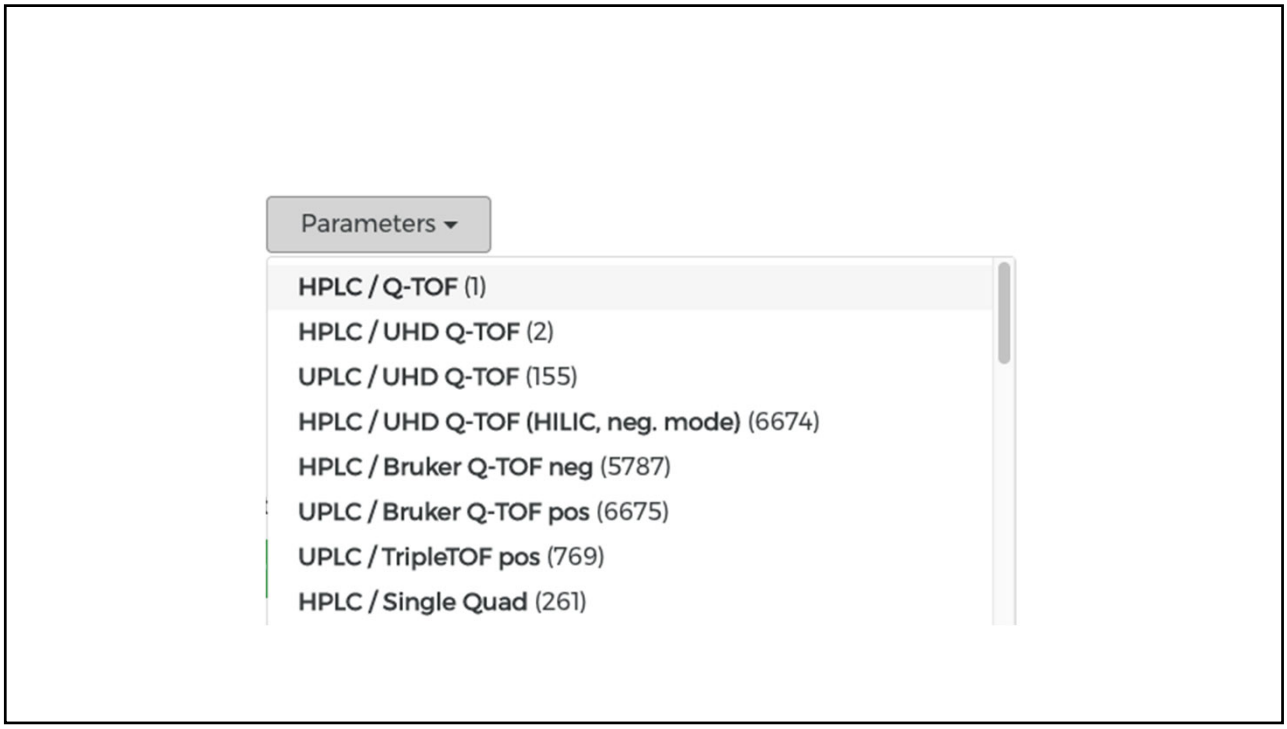
Go to stored datasets



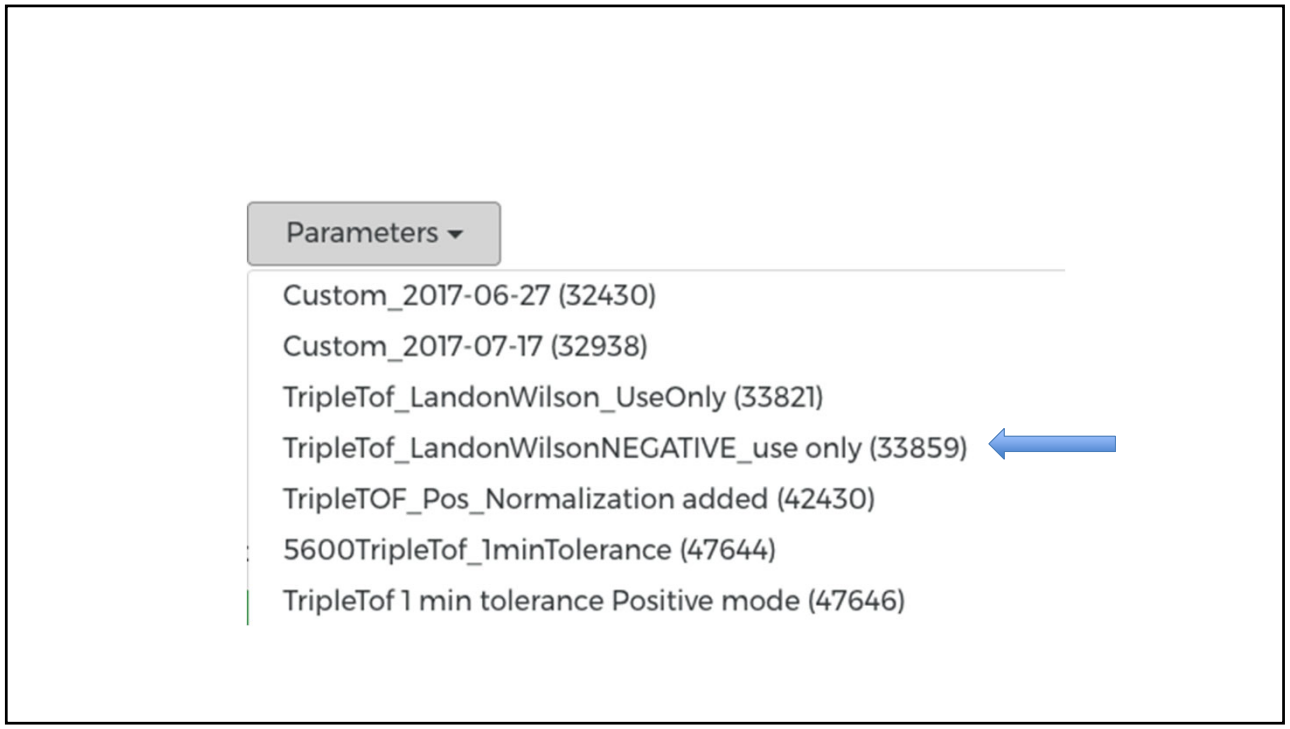
26


SELECT PARAMETERS
Parameters
None
Save as Default
Note: You may need to allow popup windows in your browser
Previous
Next

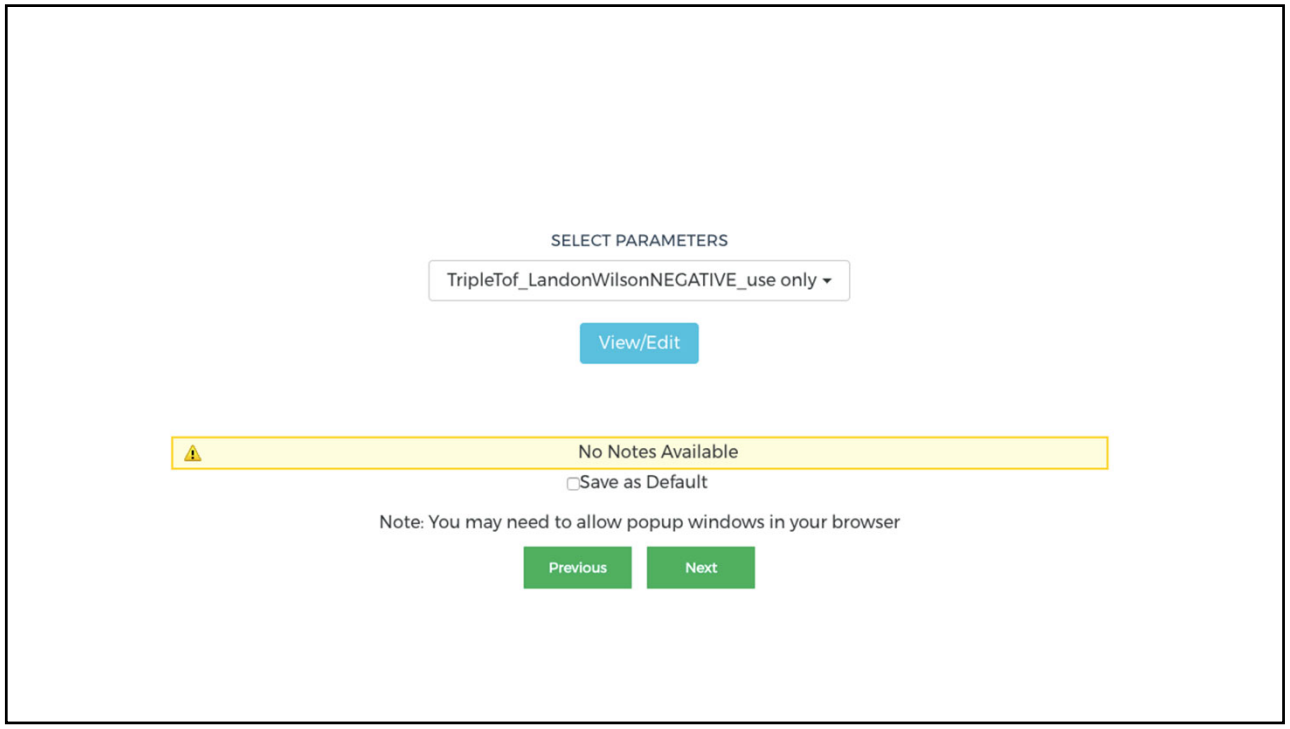
27


Parameters
HPLC / Q-TOF (1)
HPLC / UHD Q-TOF (2)
UPLC / UHD Q-TOF (155)
HPLC / UHD Q-TOF (HILIC, neg. mode) (6674)
HPLC / Bruker Q-TOF neg (5787)
UPLC / Bruker Q-TOF pos (6675)
UPLC / TripleTOF pos (769)
HPLC / Single Quad (261)

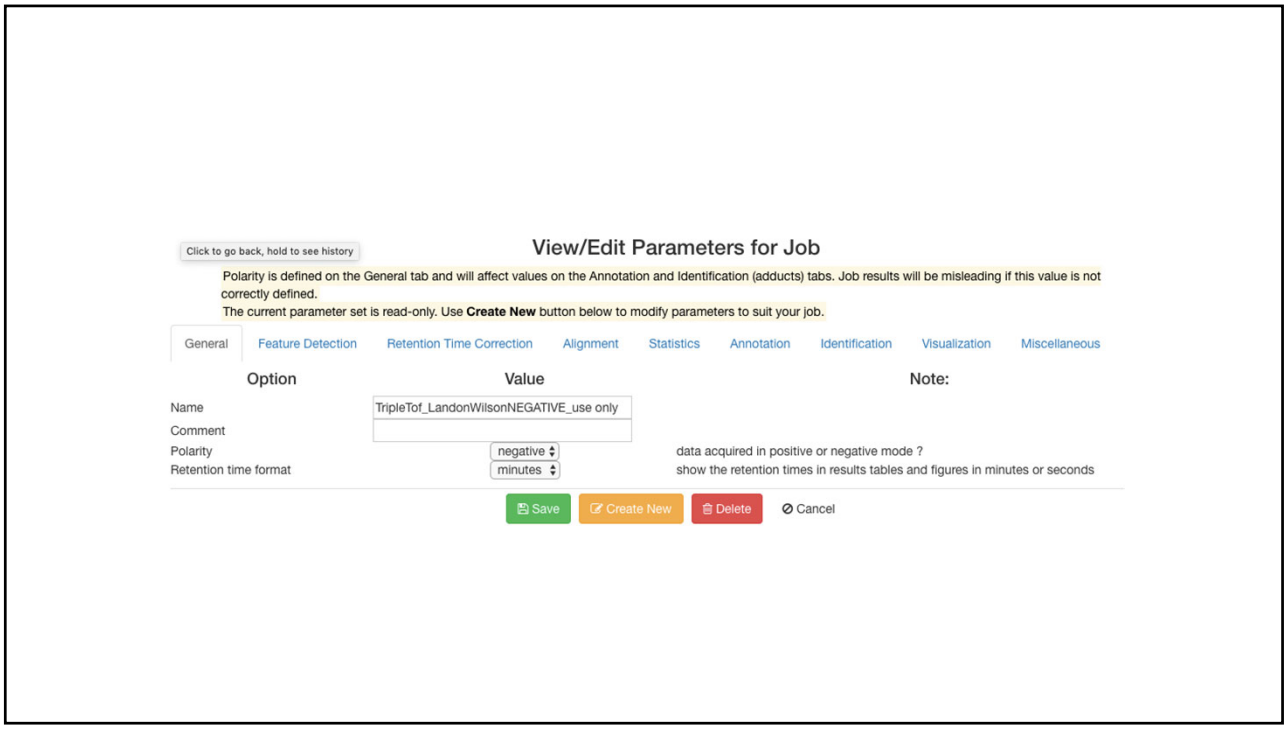
28


Parameters
Custom_2017-06-27 (32430)
Custom_2017-07-17 (32938)
TripleTof_LandonWilson_UseOnly (33821)
TripleTof_LandonWilsonNEGATIVE_use only (33859)
TripleTOF_Pos_Normalization added (42430)
5600TripleTof_1minTolerance (47644)
TripleTof 1 min tolerance Positive mode (47646)

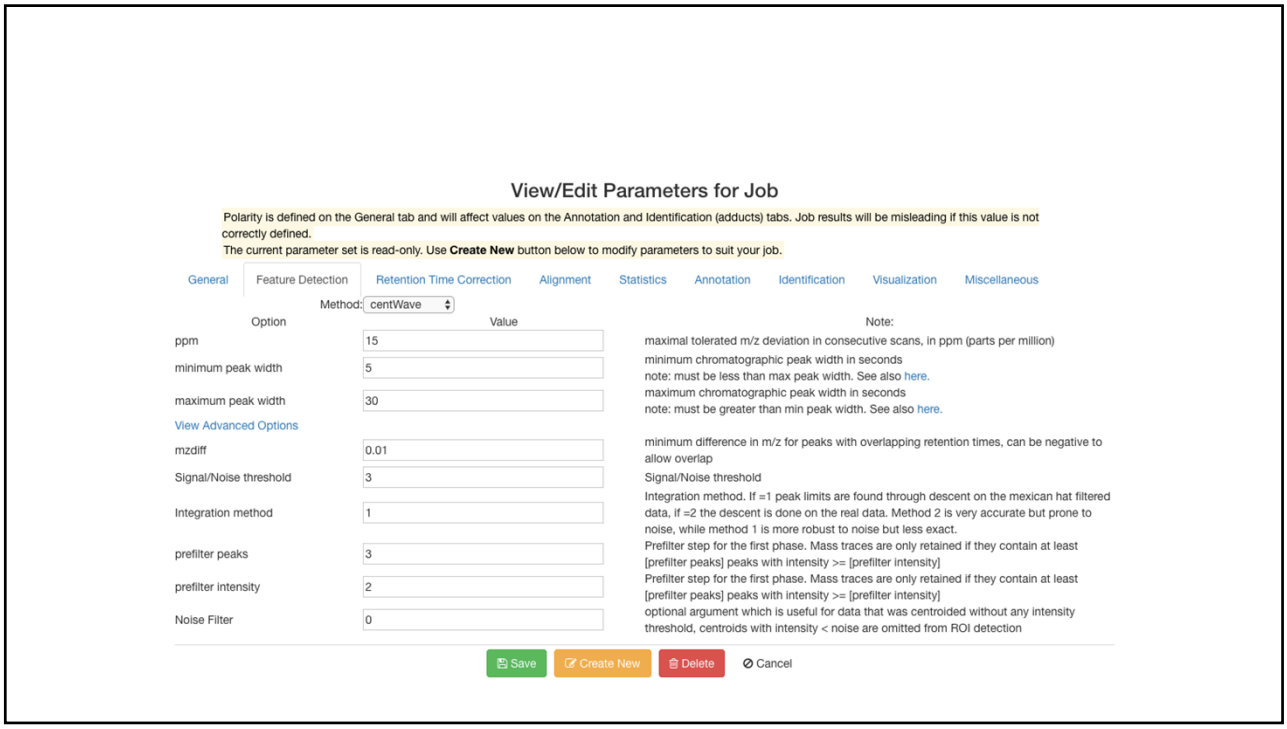
29


SELECT PARAMETERS
TripleTof_LandonWilsonNEGATIVE_use only
View/Edit
No Notes Available
Save as Default
Note: You may need to allow popup windows in your browser
Previous
Next

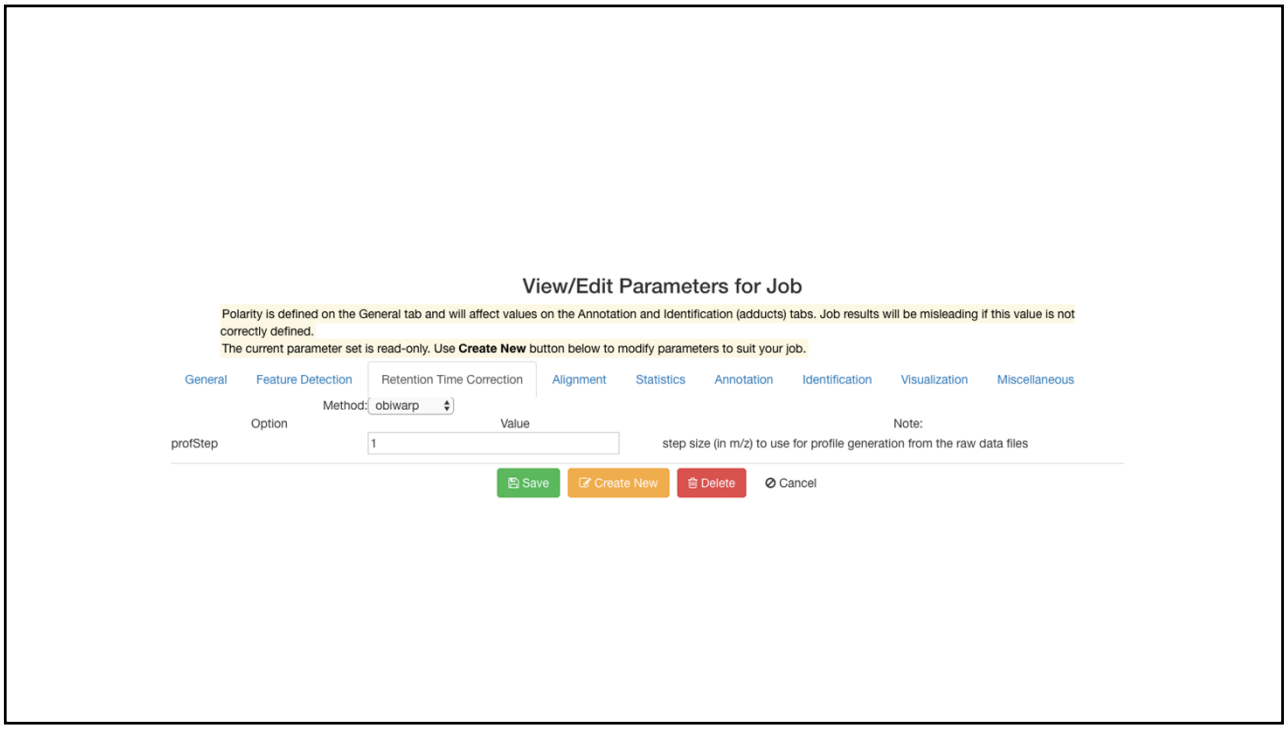
30


Click to go back, hold to see history
View/Edit Parameters for Job
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
Option
Value
Note:
Name
TripleTof_LandonWilsonNEGATIVE_use only
Comment
Polarity
negative
data acquired in positive or negative mode ?
Retention time format
minutes
show the retention times in results tables and figures in minutes or seconds
Save
Create New
Delete
Cancel

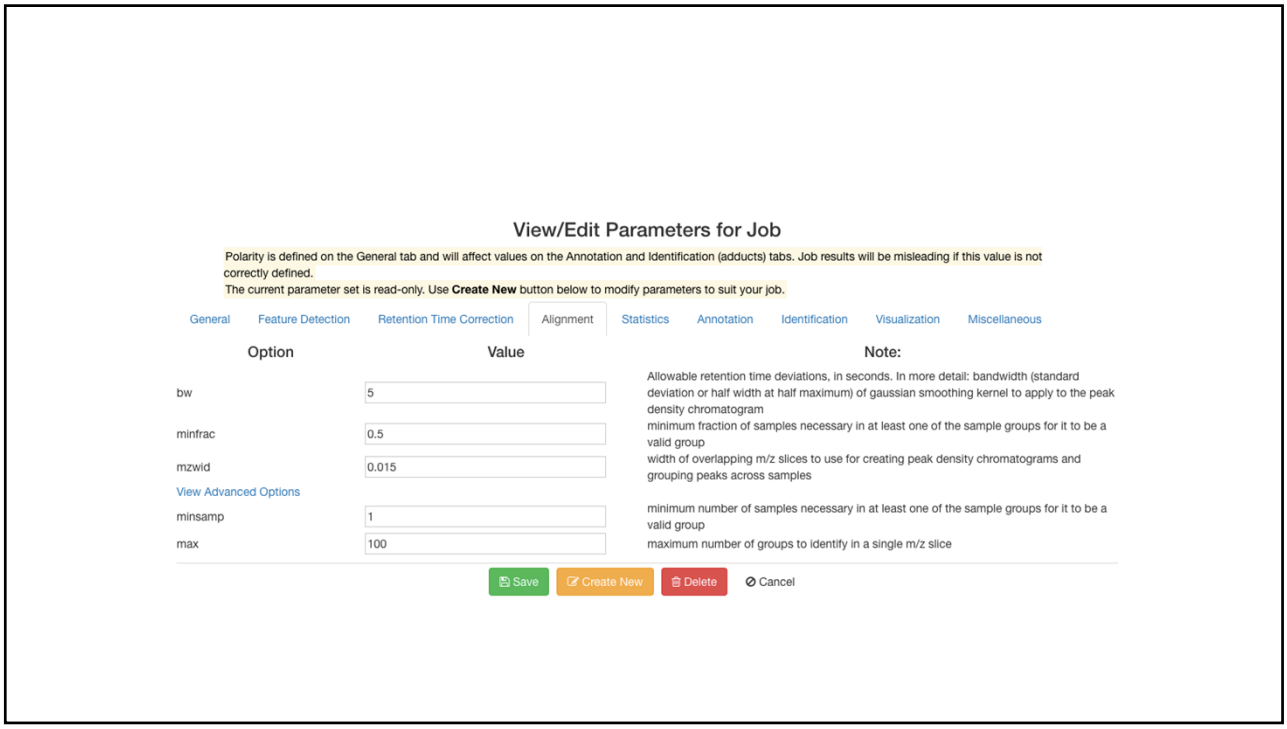
31


View/Edit Parameters for Job
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
Option
Value
Note:
ppm
15
maximal tolerated m/z deviation in consecutive scans, in ppm (parts per million)
minimum peak width
5
minimum chromatographic peak width in seconds
note: must be less than max peak width. See also here.
maximum peak width
30
maximum chromatographic peak width in seconds
note: must be greater than min peak width. See also here.
View Advanced Options
mzdiff
0.01
minimum difference in m/z for peaks with overlapping retention times, can be negative to allow overlap
Signal/Noise threshold
3
Signal/Noise threshold
Integration method
1
Integration method. If =1 peak limits are found through descent on the mexican hat filtered data, if =2 the descent is done on the real data. Method 2 is very accurate but prone to noise, while method 1 is more robust to noise but less exact.
prefilter peaks
3
Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]
prefilter intensity
2
Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]
Noise Filter
0
optional argument which is useful for data that was centroided without any intensity threshold, centroids with intensity < noise are omitted from ROI detection
Save
Create New
Delete
Cancel

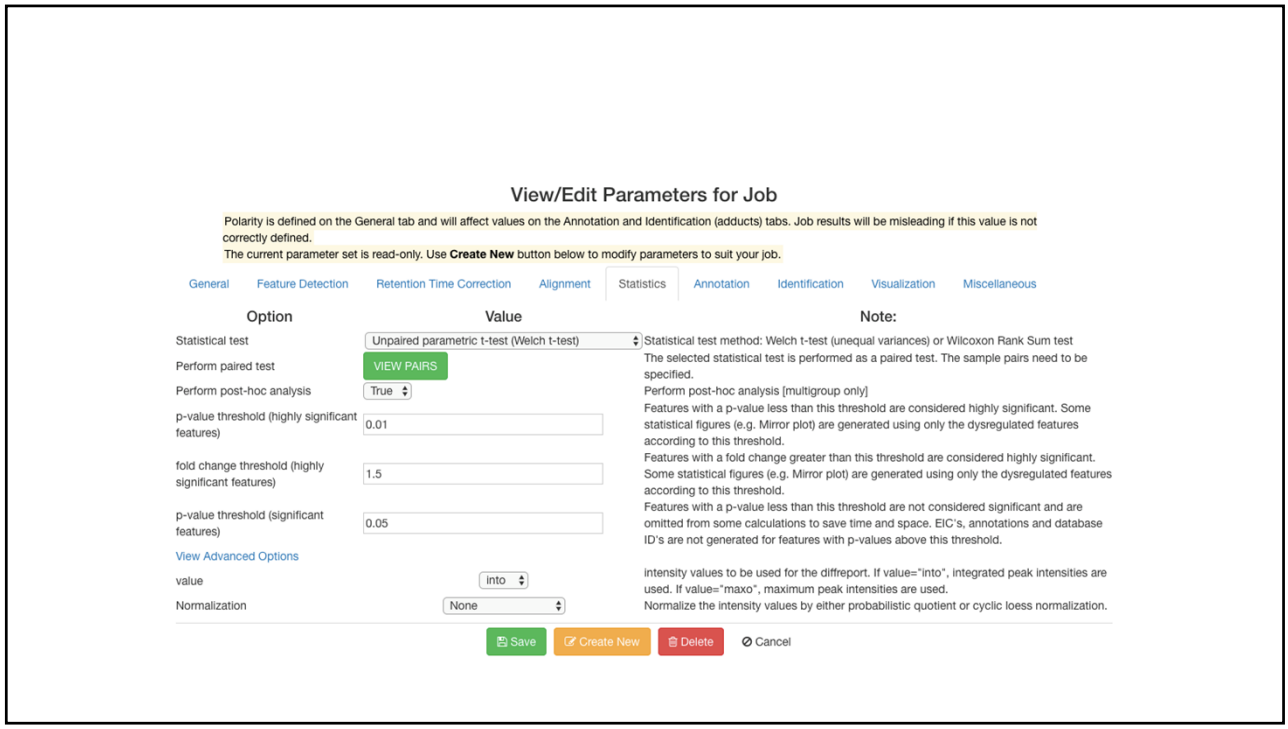
32


View/Edit Parameters for Job
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: obiwarp
Option
Value
Note:
profStep
1
step size (in m/z) to use for profile generation from the raw data files
Save
Create New
Delete
Cancel

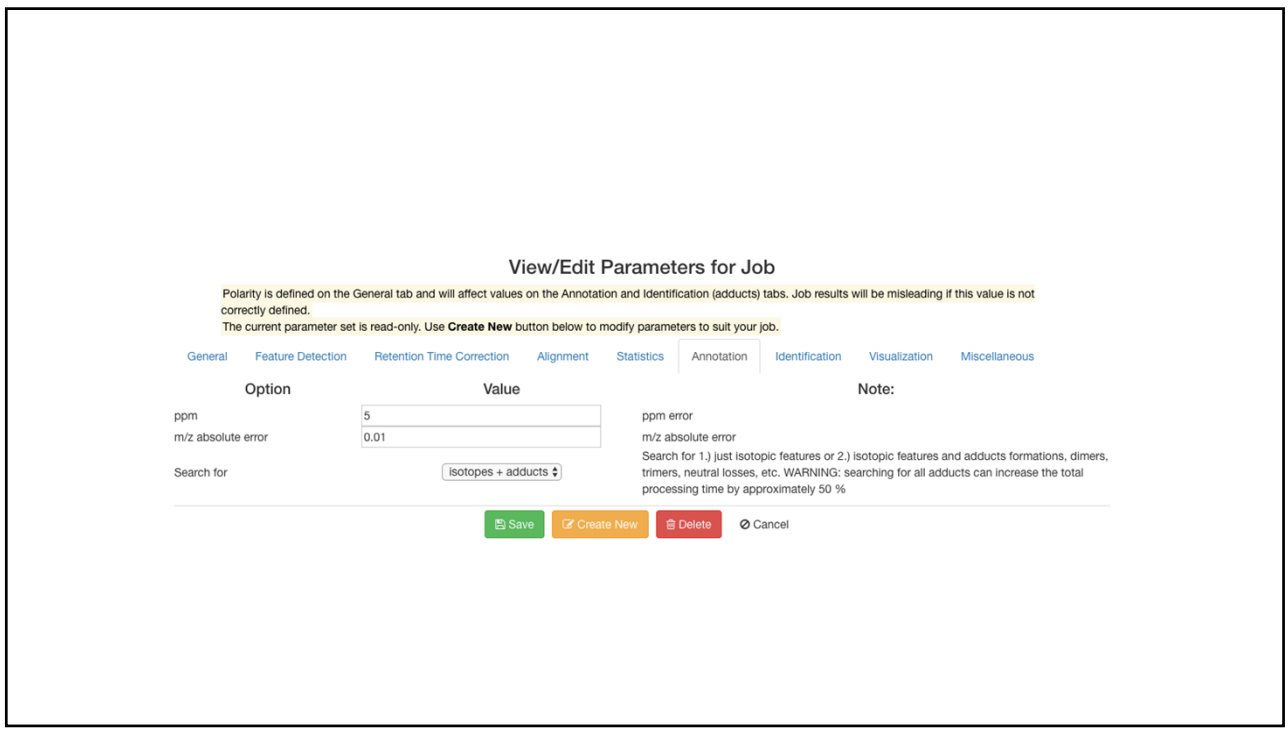
33


View/Edit Parameters for Job
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
Option
Value
Note:
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
minfrac
0.5
minimum fraction of samples necessary in at least one of the sample groups for it to be a valid group
mzwid
0.015
width of overlapping m/z slices to use for creating peak density chromatograms and grouping peaks across samples
View Advanced Options
minsamp
1
minimum number of samples necessary in at least one of the sample groups for it to be a valid group
max
100
maximum number of groups to identify in a single m/z slice
Save
Create New
Delete
Cancel

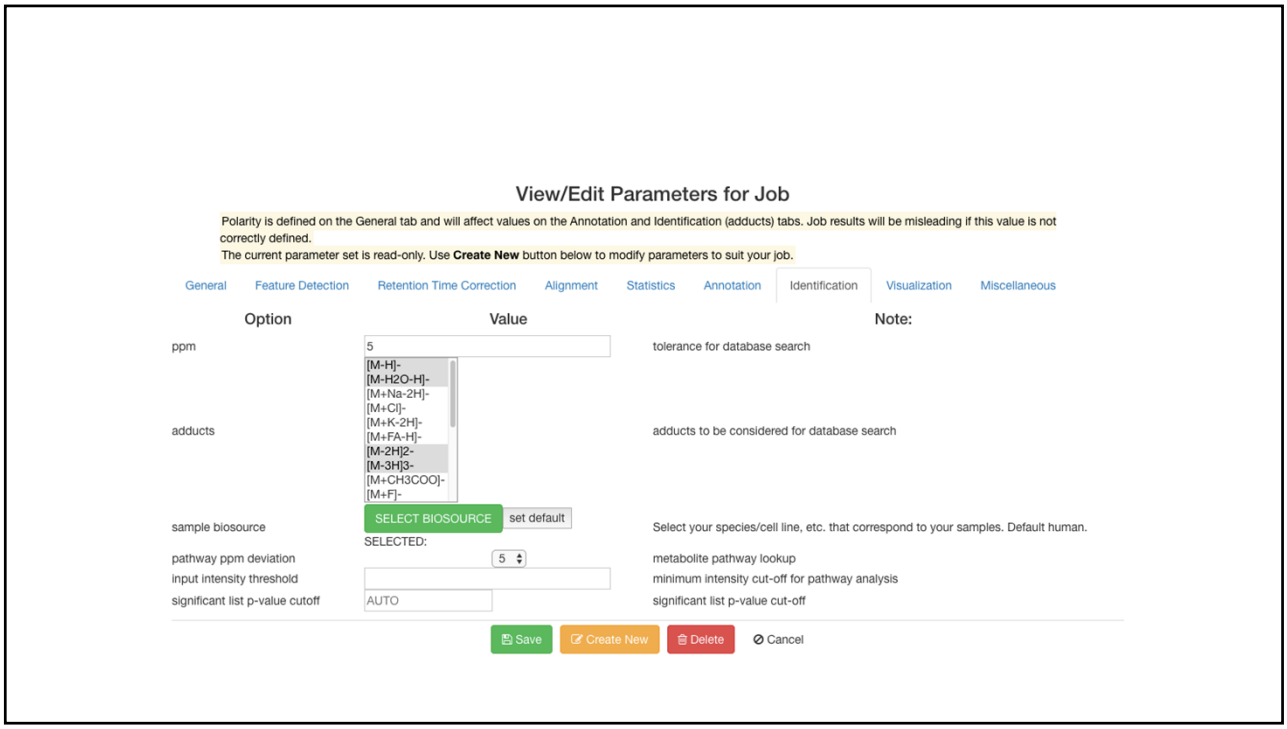
34


View/Edit Parameters for Job
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
Option
Value
Note:
Statistical test
Unpaired parametric t-test (Welch t-test)
Statistical test method: Welch t-test (unequal variances) or Wilcoxon Rank Sum test
Perform paired test
VIEW PAIRS
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)
0.05
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.
Save
Create New
Delete
Cancel

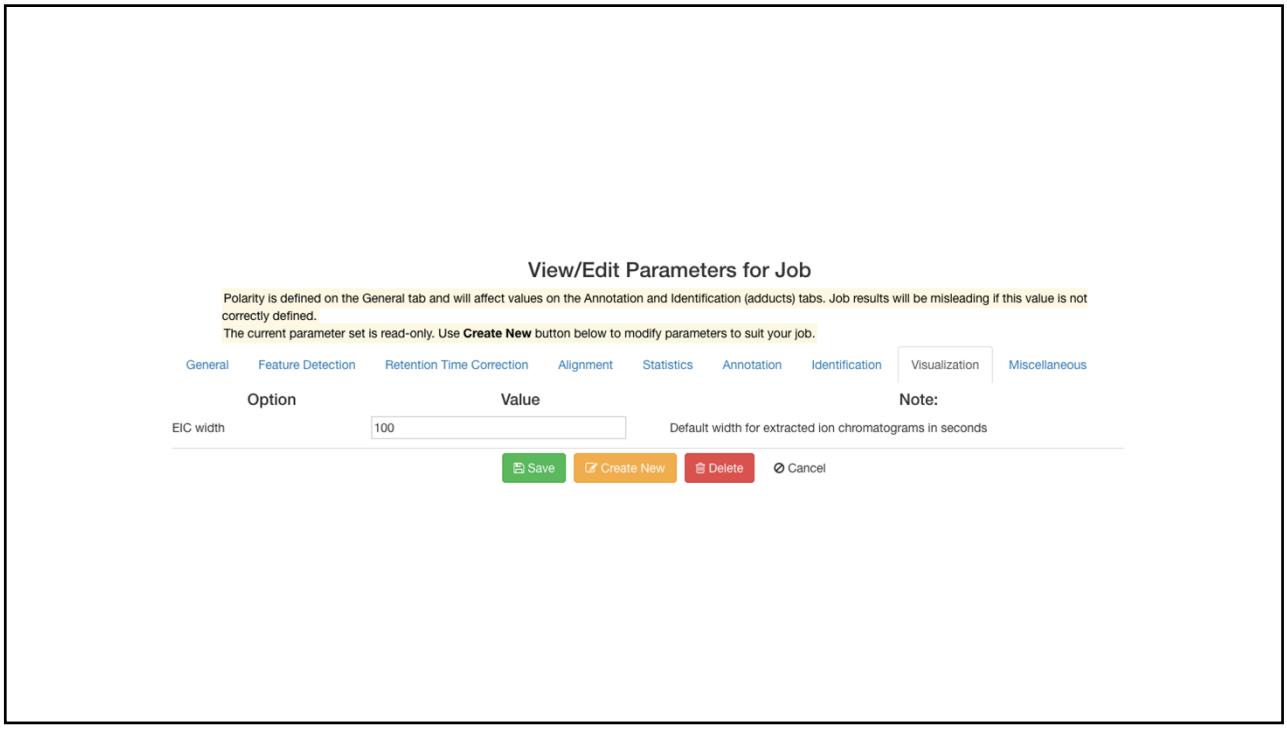
35


View/Edit Parameters for Job
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
Option
Value
Note:
ppm
5
ppm error
m/z absolute error
0.01
m/z absolute error
Search for
isotopes + adducts
Search for 1.) just isotopic features or 2.) isotopic features and adducts formations, dimers, trimers, neutral losses, etc. WARNING: searching for all adducts can increase the total processing time by approximately 50 %
Save
Create New
Delete
Cancel

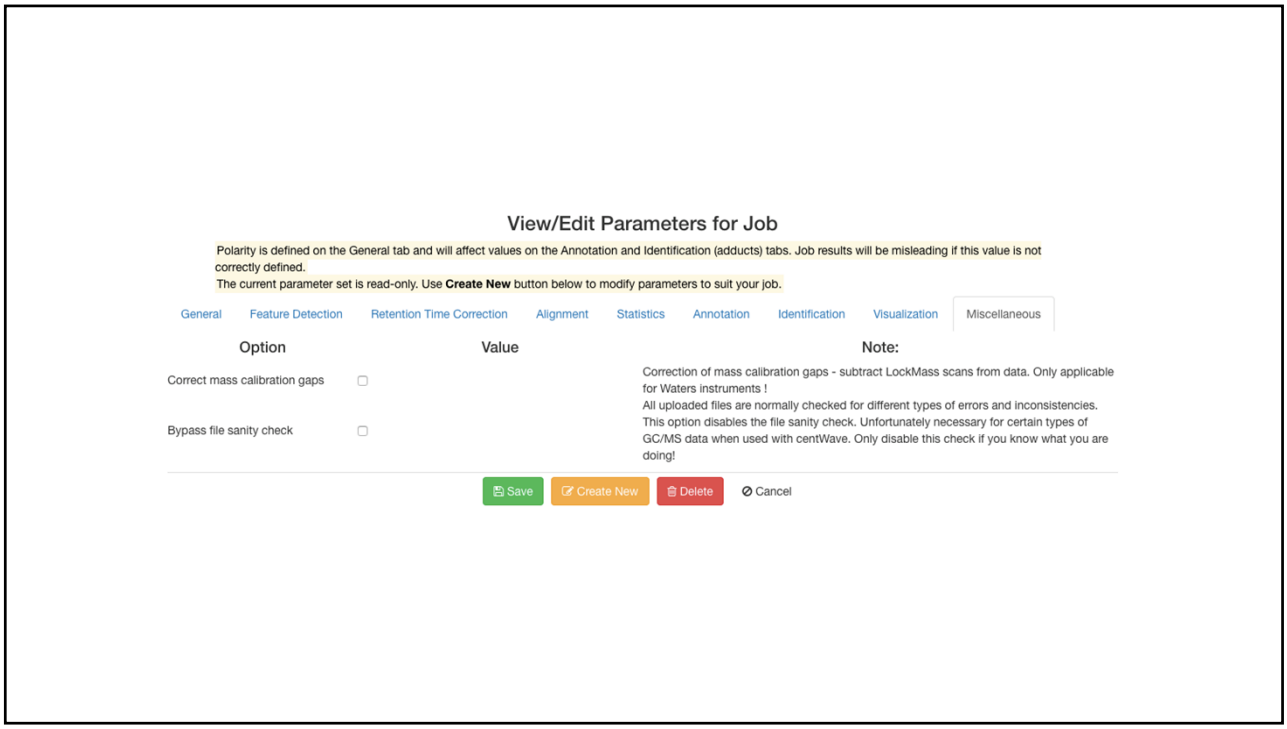
36


View/Edit Parameters for Job
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
Option
Value
Note:
ppm
5
tolerance for database search
[M-H]-
[M-H2O-H]-
[M+Na-2H]-
[M+Cl]-
[M+K-2H]-
[M+FA-H]-
[M-2H]2-
[M-3H]3-
[M+CH3COO]-
[M+F]-
adducts
adducts to be considered for database search
SELECT BIOSOURCE
set default
sample biosource
SELECTED:
Select your species/cell line, etc. that correspond to your samples. Default human.
pathway ppm deviation
5
metabolite pathway lookup
input intensity threshold
minimum intensity cut-off for pathway analysis
significant list p-value cutoff
AUTO
significant list p-value cut-off
Save
Create New
Delete
Cancel

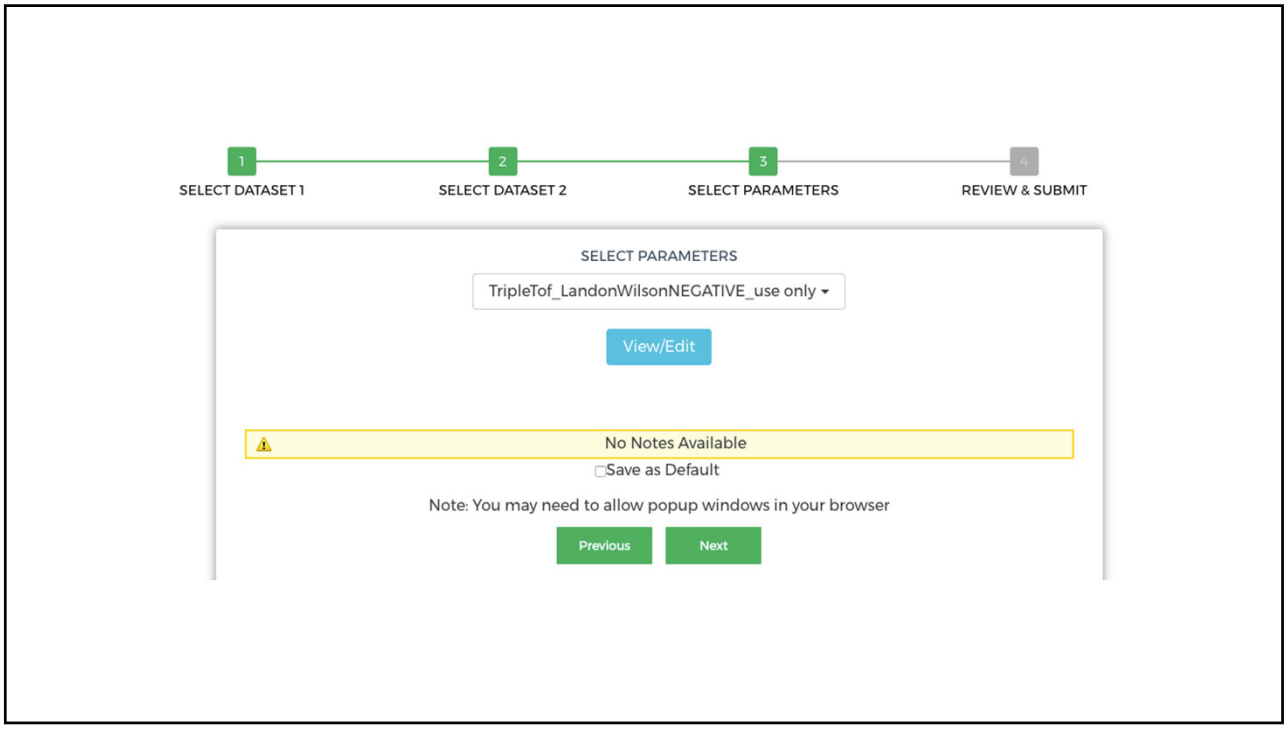
37


View/Edit Parameters for Job
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
Option
Value
Note:
EIC width
100
Default width for extracted ion chromatograms in seconds
Save
Create New
Delete
Cancel

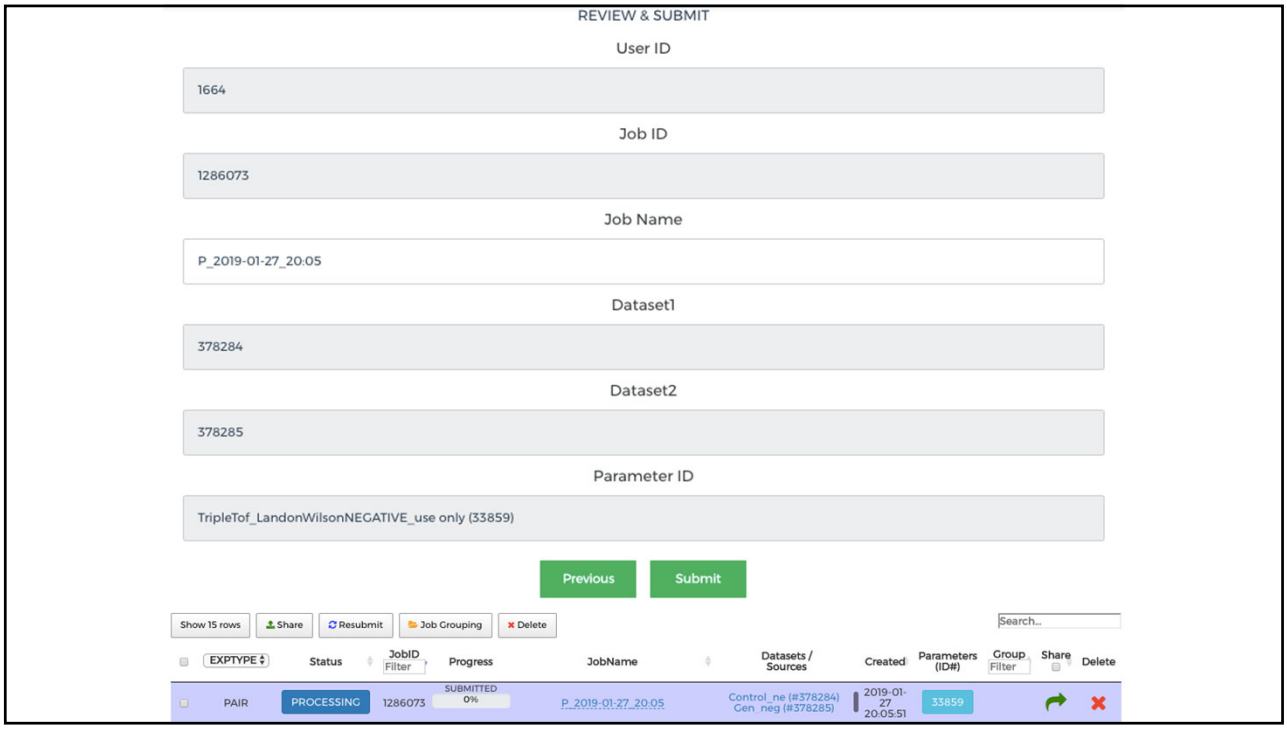
38


View/Edit Parameters for Job
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
Option
Value
Note:
Correct mass calibration gaps
Correction of mass calibration gaps - subtract LockMass scans from data. Only applicable for Waters instruments !
Bypass file sanity check
All uploaded files are normally checked for different types of errors and inconsistencies. This option disables the file sanity check. Unfortunately necessary for certain types of GC/MS data when used with centWave. Only disable this check if you know what you are doing!
Save
Create New
Delete
Cancel

39


1
SELECT DATASET 1
2
SELECT DATASET 2
3
SELECT PARAMETERS
4
REVIEW & SUBMIT
SELECT PARAMETERS
TripleTof_LandonWilsonNEGATIVE_use only
View/Edit
No Notes Available
Save as Default
Note: You may need to allow popup windows in your browser
Previous
Next

40



41

Tips for naming files for upload to XCMS Online

- **Before uploading, it is a good idea to create separate file folders on your hard drive to better organize your data into the groups you want to examine.**
- **Add enough description to discriminate between different samples and sample set names.**
- **Adding the ionization mode in the name is preferable, i.e. PosMode or NegMode.**
- **Eliminate open spaces in the data file name by using "_" (underscore) notation. Open spaces can cause upload errors in XCMS Online.**

Example of DataSet Name: Control_Neg

42

What does a LC-MS data set consist of?

- A Q-TOF instrument during the LC run, for example, acquires data on a 0.5 second duty cycle
  - 0-100 msec
    - High resolution/mass accuracy MS spectrum
  - 100-500 msec
    - A succession of selected MSMS spectra
    - If each MSMS spectrum is collected for 50 msec, then 8 precursor ions can be selected in the duty cycle
    - The precursor ions are selected from the MS spectrum observed in the current duty cycle
    - Once an ion has been selected for MSMS it can be placed on a "don't observe" list for say 30 sec

43

What ions are observed in LC-MS data?

- Ions coming from the biological system being studied
- Ions from compounds introduced into the extract during storage and extraction
- Ions from the solvent used for the chromatography
- Ions from the column material
- Ions from the previous sample that was run

44