Knowledge that will change your world

Preparing data for upload to XCMSonline

Stephen Barnes, PhD

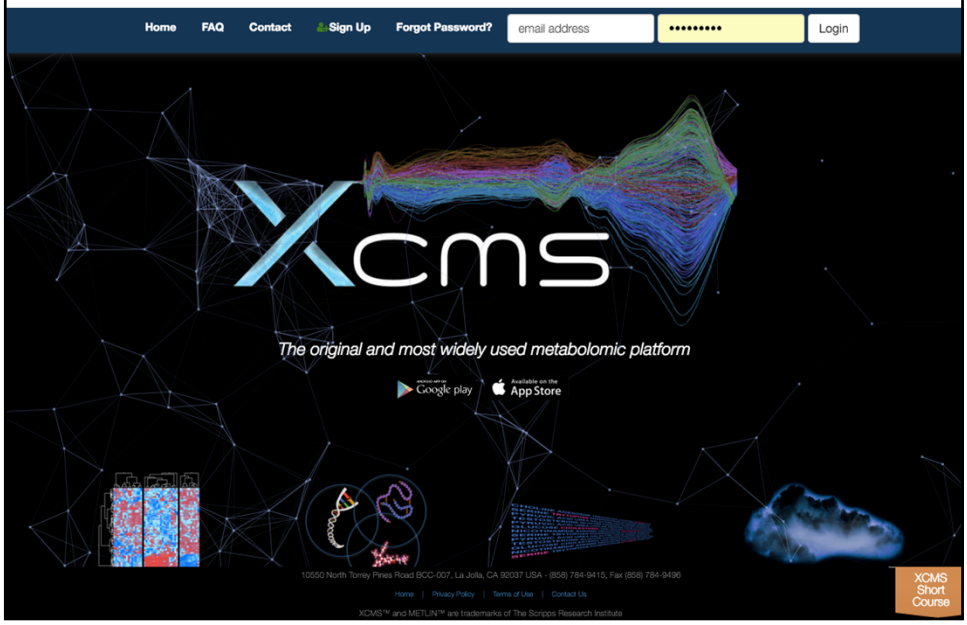
University of Alabama at Birmingham

sbarnes@uab.edu

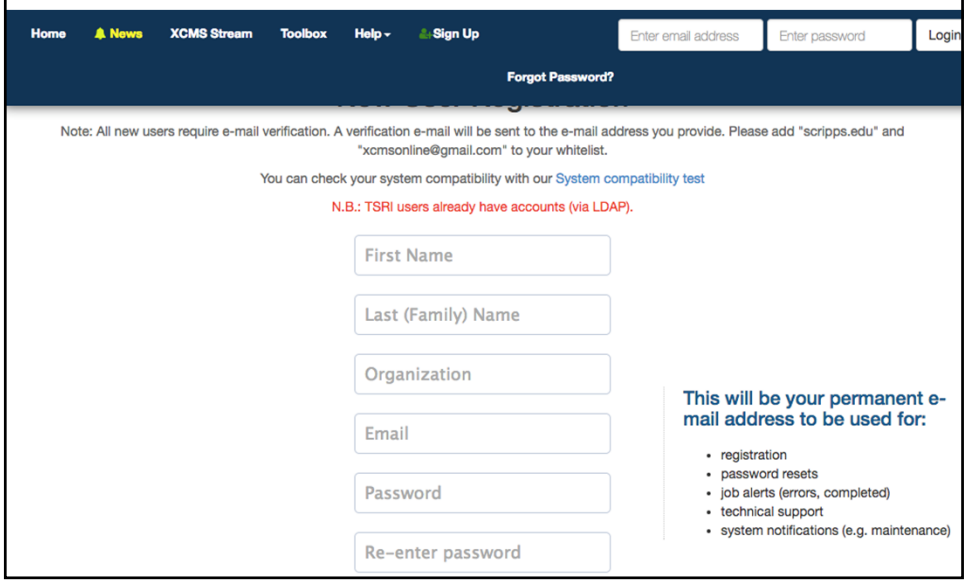
Synopsis

- **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**
  - **On February 1, Paul Benton from Scripps Research Institute (via Skype) will describe and show you how the online software XCMS works**

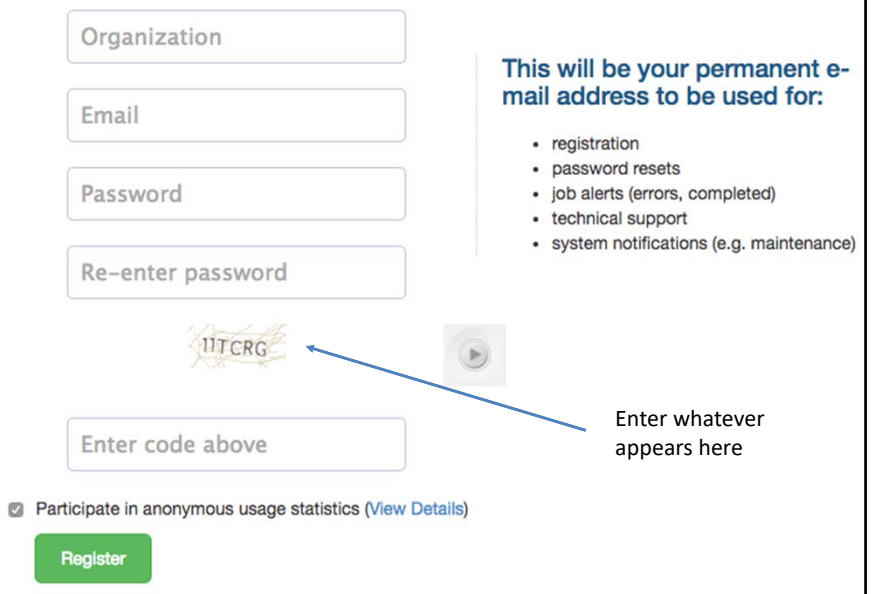
Go to https://xcmsonline.scripps.edu/



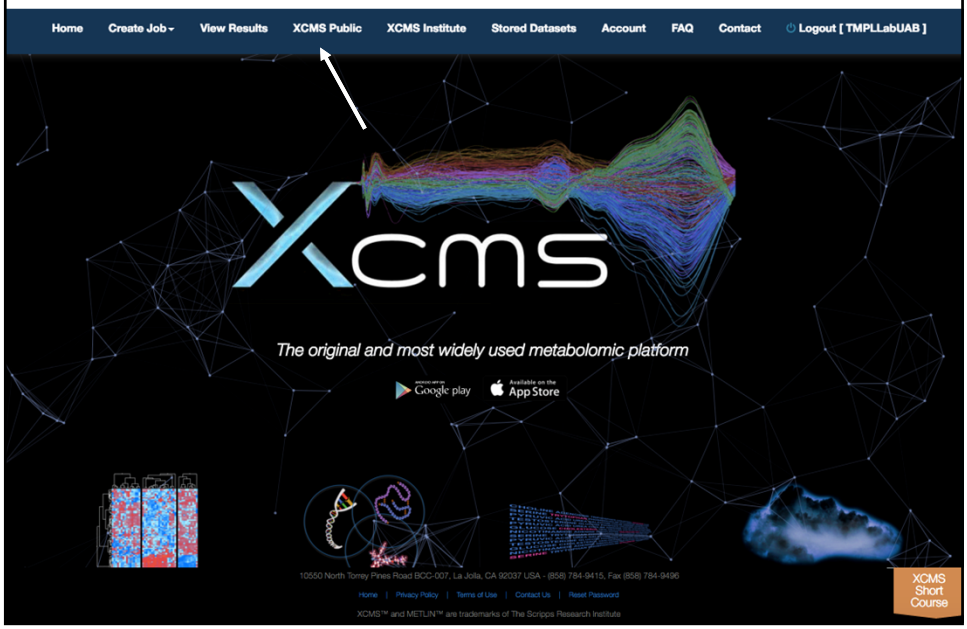
Signing up for XCMS



Enter the code below password



Starting page for XCMSonline

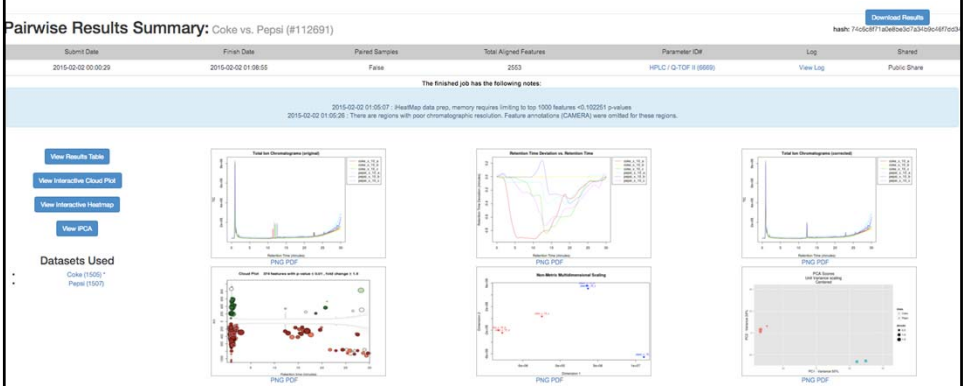


View Public Job Shares

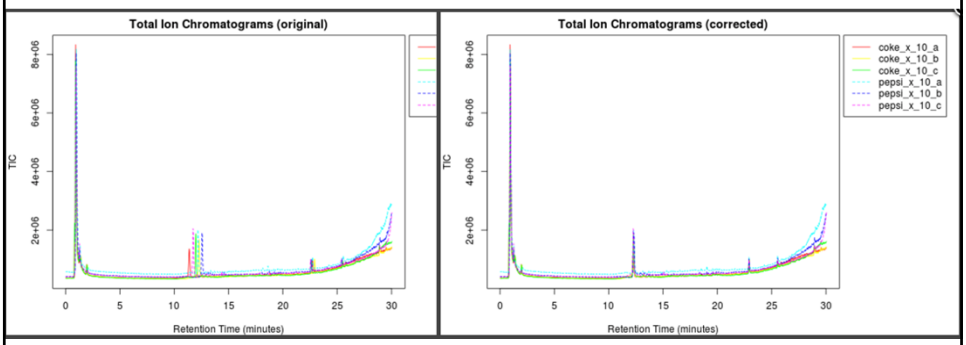
Show 10 entries Search:

Status	ID	Job Name	Author(s)	Publication	Notes
VIEW	112691	Coke vs. Pepsi	Bill Webb		demo data set II: soda
VIEW	112801	FAAH: WT vs KO	Alan Saghatelian	Saghatelian et al. Assignment of Endogenous Substrates to Enzymes by Global Metabolite Profiling, *Biochemistry*, 2004, 43 (45), pp 14332-14339	spinal cords of 6 wild-type mice and 6 fatty acid amide hydrolase (FAAH) knockout mice
VIEW	109981	SHAM vs. TNT	Gary Patti	Patti et al. Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin , *Nature Chemical Biology* 8, 232-234 (2012)	comparison of the TNT ipsilateral dorsal horn to the sham ipsilateral dorsal horn
VIEW	112692	Stone vs. Ale Smith	Bill Webb		demo data set I: beer
VIEW	126997	URA3 vs VNG2094G	Ralf Tautenhahn	Tautenhahn et al. metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data, Analytical Chemistry 2011 83 (3), 696-700	*Halobacterium salinarum* knockout strains

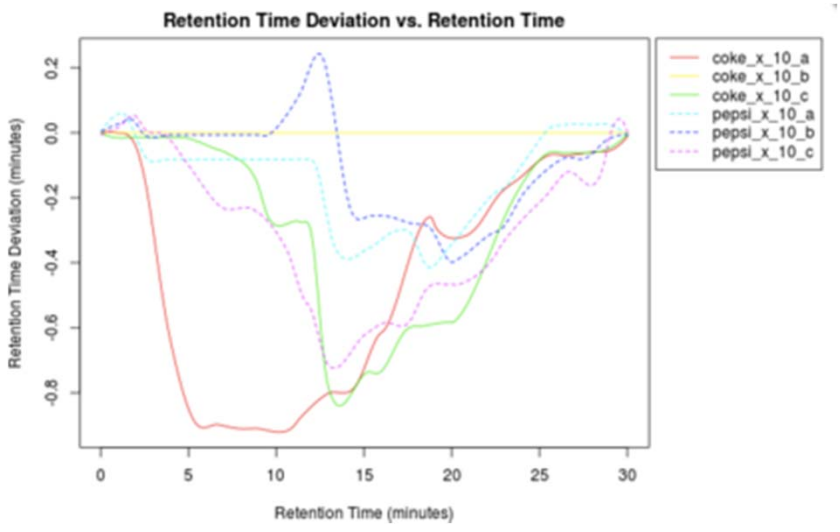
Results: Coke vs Pepsi



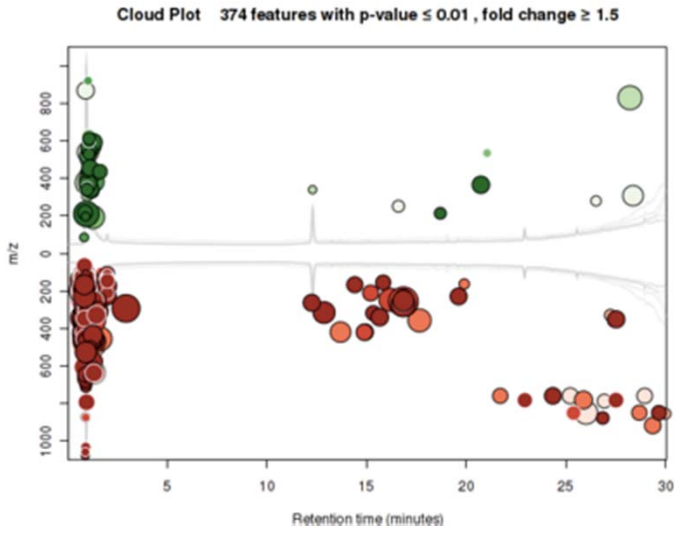
Rt correction - Coke vs Pepsi



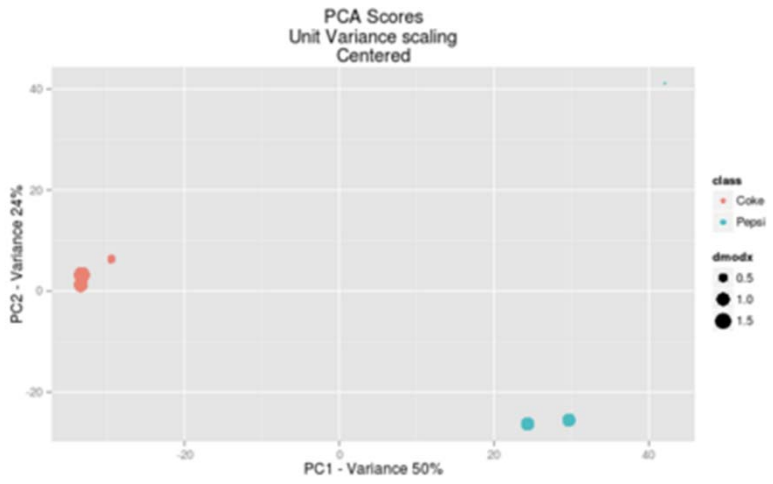
Variation in retention time



Cloud plot – Coke vs Pepsi



Principal components analysis plot



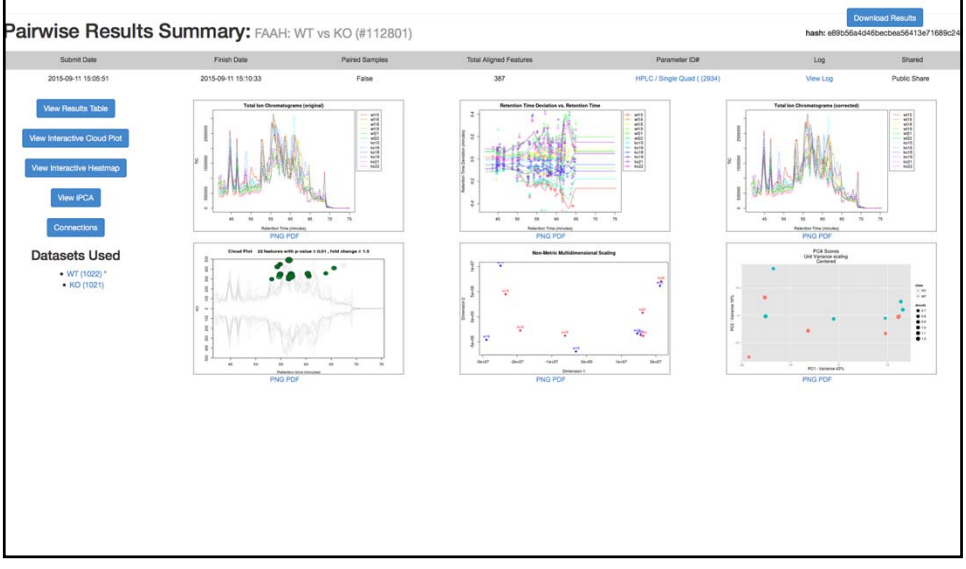
Let's look at the FAAH – WT vs KO data

View Public Job Shares

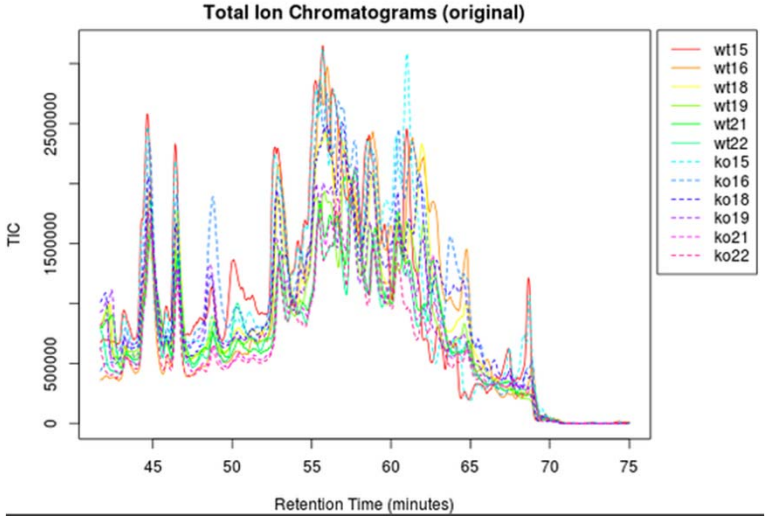
Show 10 entries Search:

Status	ID	Job Name	Author(s)	Publication	Notes
VIEW	112691	Coke vs. Pepsi	Bill Webb		demo data set II: soda
VIEW	112801	FAAH: WT vs KO	Alan Saghatelian	Saghatelian et al. Assignment of Endogenous Substrates to Enzymes by Global Metabolite Profiling, *Biochemistry*, 2004, 43 (45), pp 14332-14339	spinal cords of 6 wild-type mice and 6 fatty acid amide hydrolase (FAAH) knockout mice
VIEW	109981	SHAM vs. TNT	Gary Patti	Patti et al. Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin , *Nature Chemical Biology* 8, 232-234 (2012)	comparison of the TNT ipsilateral dorsal horn to the sham ipsilateral dorsal horn
VIEW	112692	Stone vs. Ale Smith	Bill Webb		demo data set I: beer
VIEW	126997	URA3 vs VNG2094G	Ralf Tautenhahn	Tautenhahn et al. metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data, Analytical Chemistry 2011 83 (3), 696-700	*Halobacterium salinarum* knockout strains
VIEW	126996	URA3 vs VNG1816G	Ralf Tautenhahn	Tautenhahn et al. metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data, Analytical Chemistry 2011 83 (3), 696-700	*Halobacterium salinarum* knockout strains

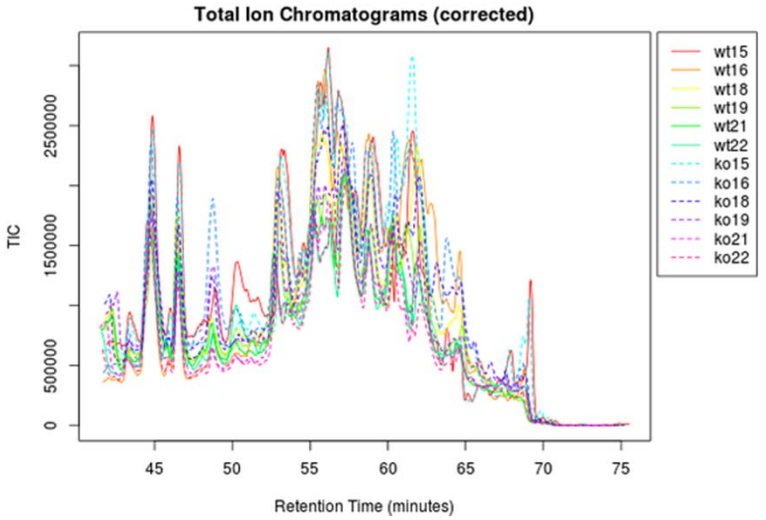
Overall look at the results page



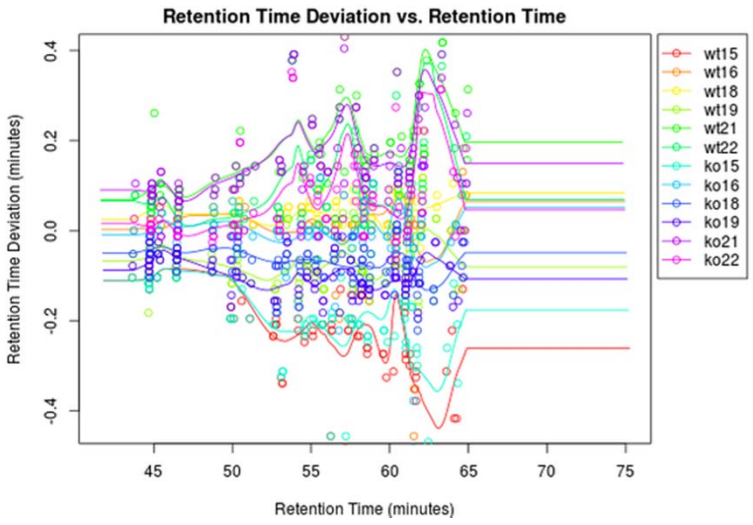
Overlay of ion chromatograms (original data)



Overlay of ion chromatograms (aligned data)



Alignment corrections for FAAH data

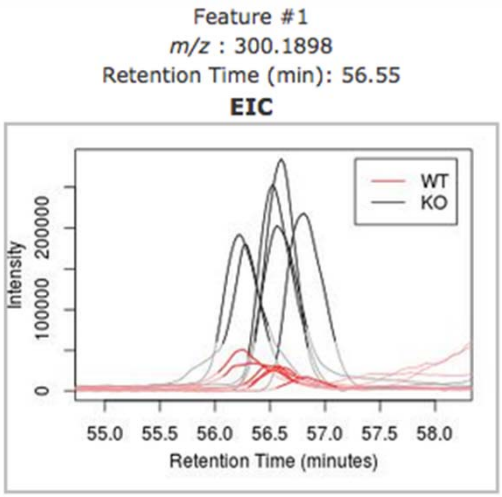


Viewing the data table for FAAH

Columns | Hide isotopic peaks | Page 1 of 4 | 100 | View 1 - 100 of 3

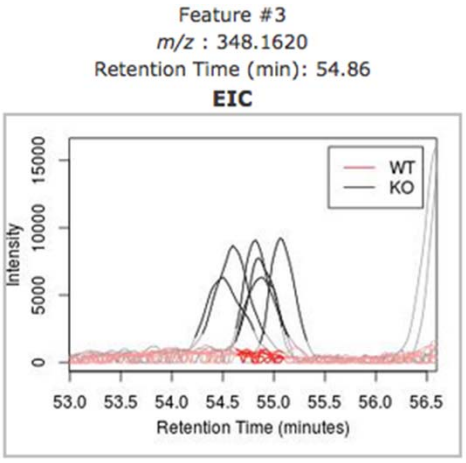
featureidx	fold	pvalue	updown	mzmed	rtmed	maxint	dataset1_n	dataset2_n	isotopes	adducts	peakgrou	usernotes
1	5.7	5.02634e-8	UP	300.1898	56.55	284,800	820,662	4,672,516	[9][M]+	[M+H]+ 29	20	
2	6.1	5.67701e-8	UP	301.1879	56.55	58,696	159,994	976,249	[9][M+1]+		20	
3	10.3	4.32371e-6	UP	348.1620	54.86	9,212	15,224	156,406			12	
4	22.3	8.17994e-6	UP	491.2000	56.61	16,104	16,106	359,085			119	
5	6.3	0.00003	UP	423.1499	54.29	11,119	32,573	204,905			127	
6	39.1	0.00009	UP	327.1989	57.05	50,504	22,165	866,093		[M+H+NH3	78	
7	5.5	0.00010	UP	411.2339	65.61	22,480	61,619	339,046	[32][M+1]+		66	
8	15.5	0.00013	UP	326.2000	56.95	232,896	237,009	3,684,173			3	
9	6.9	0.00016	UP	410.2658	65.57	71,648	152,557	1,053,935	[32][M]+		66	
10	3.7	0.00030	UP	298.1508	53.19	10,300	49,258	181,621			129	
11	4.9	0.00049	UP	449.1321	54.85	6,725	27,598	135,061			12	
12	9.2	0.00049	UP	354.2000	60.34	62,056	84,819	784,338	[21][M]+		56	
13	9.9	0.00104	UP	330.1910	60.54	9,569	13,542	134,230	[15][M+2]+		18	
14	17.3	0.00129	UP	348.1431	57.02	21,528	18,592	322,025		[M+K]+ 30	78	
15	10.3	0.00165	UP	328.2000	60.54	320,704	412,651	4,248,974	[15][M]+	[M+H]+ 32	18	
16	5.7	0.00165	UP	357.2062	63.92	26,896	60,755	344,290	[22][M+1]+		49	
17	11.2	0.00179	UP	329.2000	60.54	75,544	87,457	983,361	[15][M+1]+		18	

Viewing feature #1



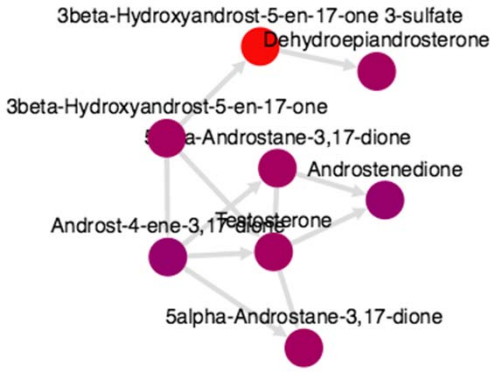
retention variation is greater than 1 min

Viewing feature #3



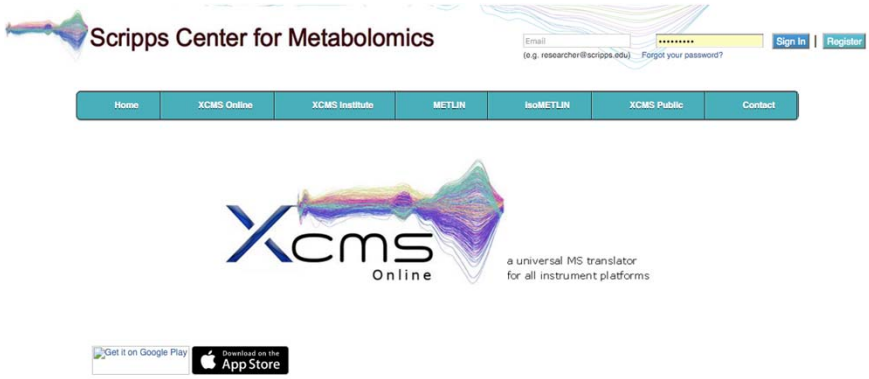
Example of where the values in the WT data are zero

"Connections"



Connections is an online version of mummichog

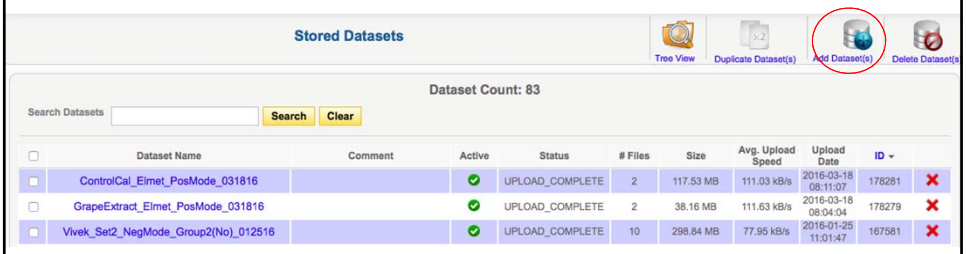
Uploading Experiments for Metabolomics Analyses using XCMS Online Applications



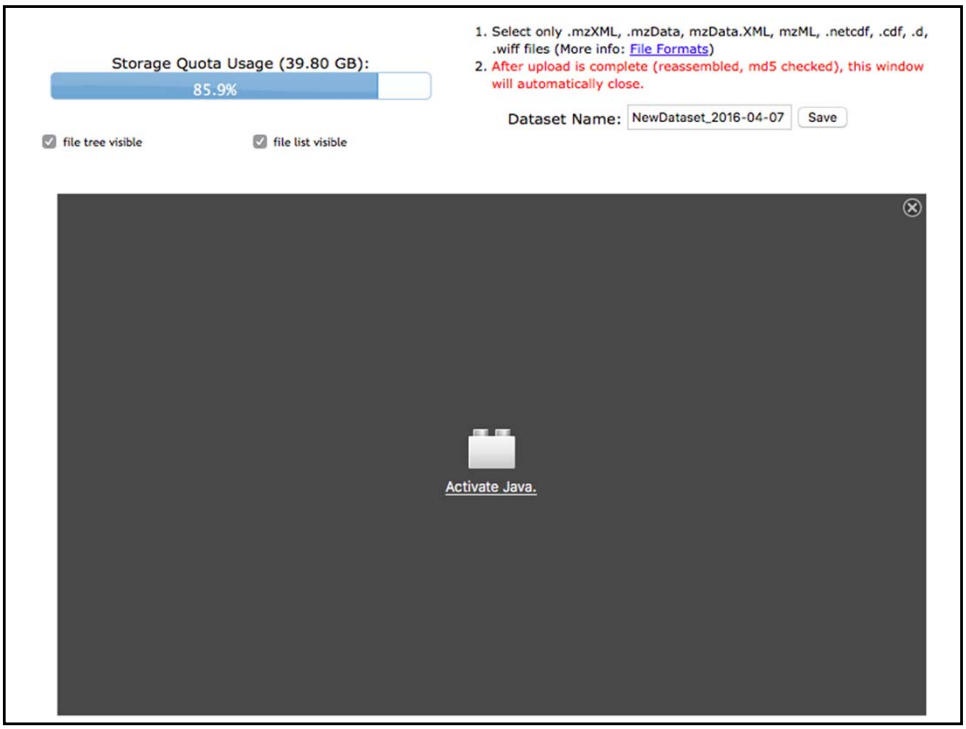
https://xcmsonline.scripps.edu/index.php

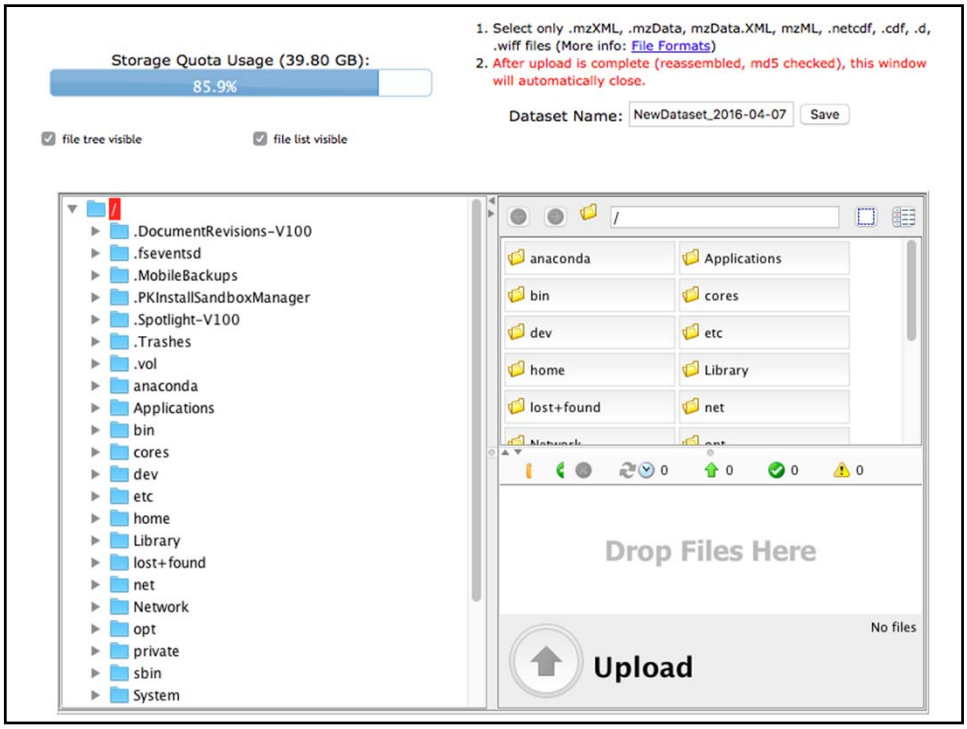
Uploading Experiments for Metabolomics Analyses using XCMS Online Applications continued…

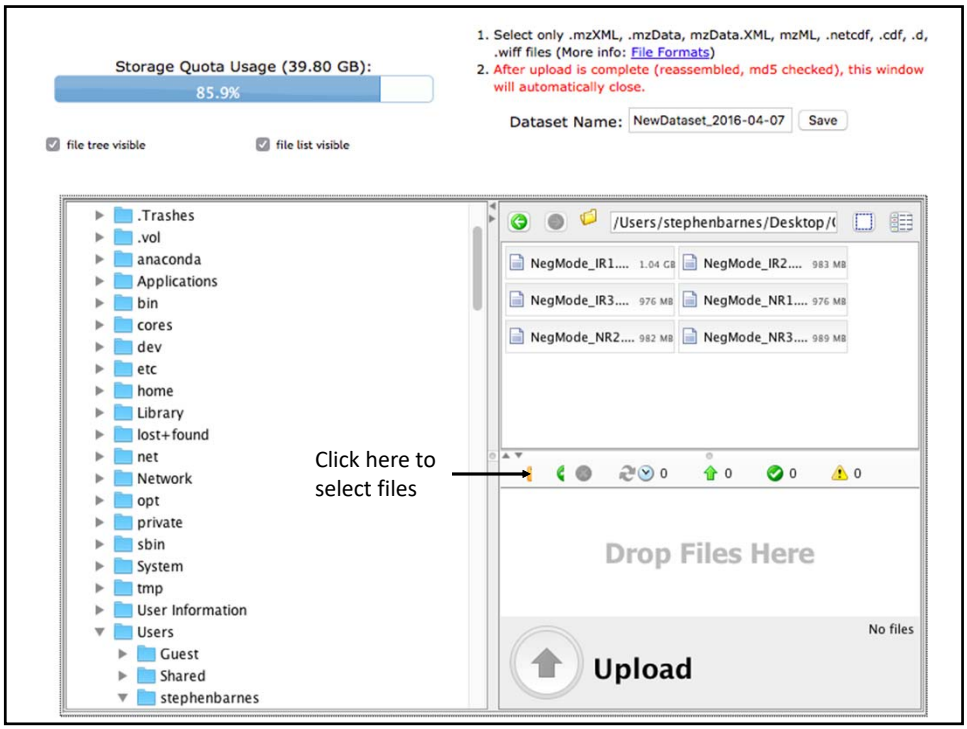
Once you have completed your LC-MS/MS analyses of samples in your particular experiment, files can be uploaded directly for most instrumentation types.

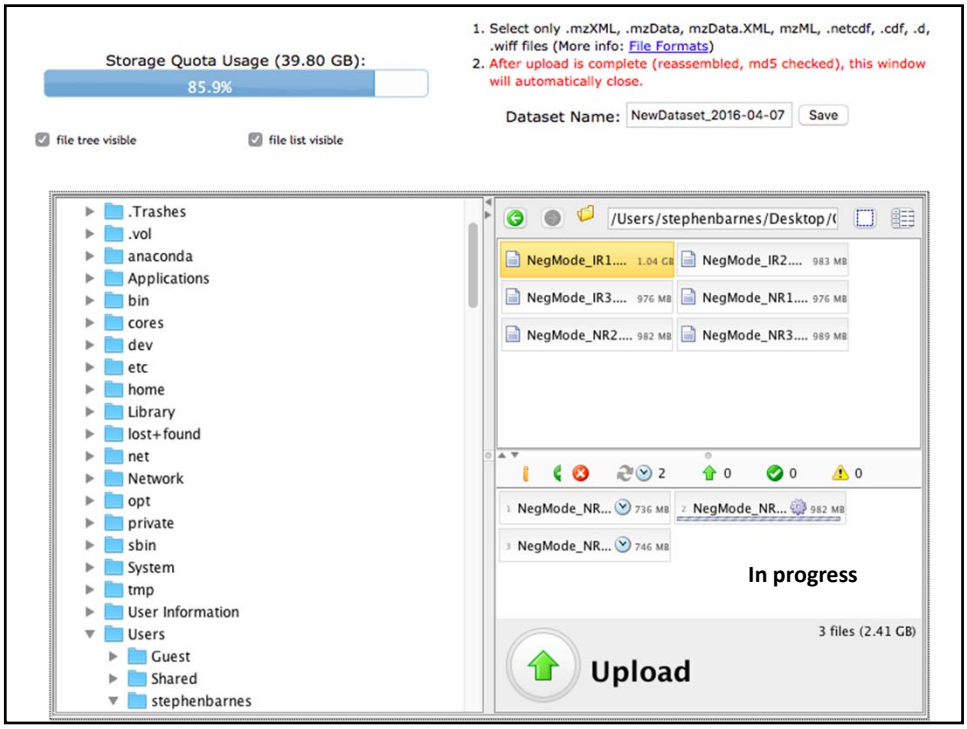


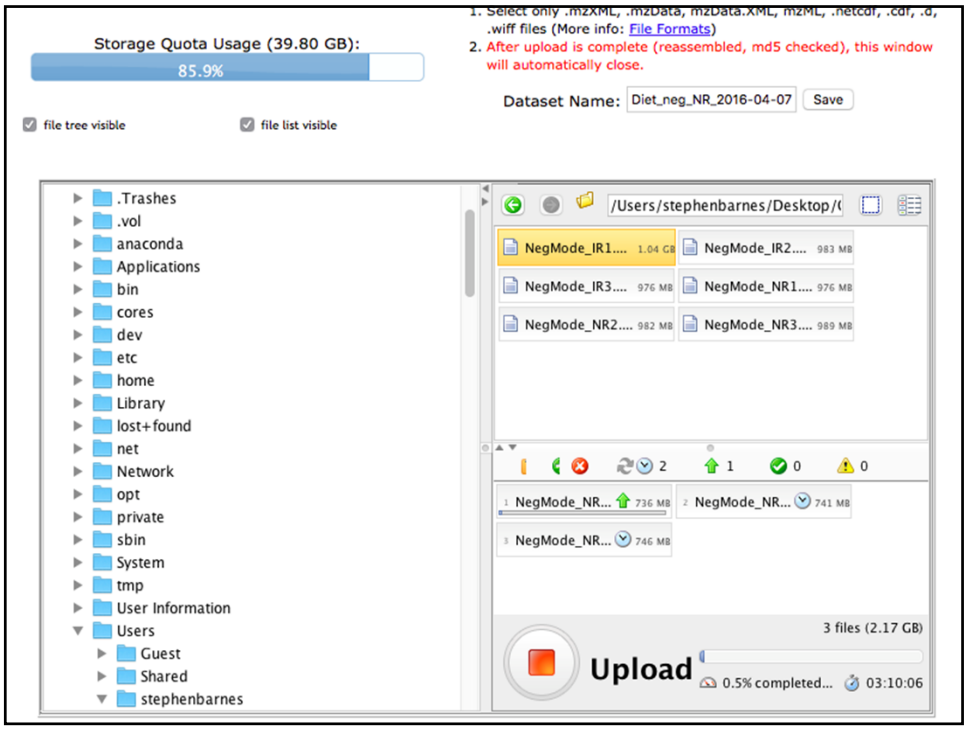
Select the option "*Stored Datasets*," on the XCMS toolbar to access datasets already stored, or to upload new datasets for future analyses.



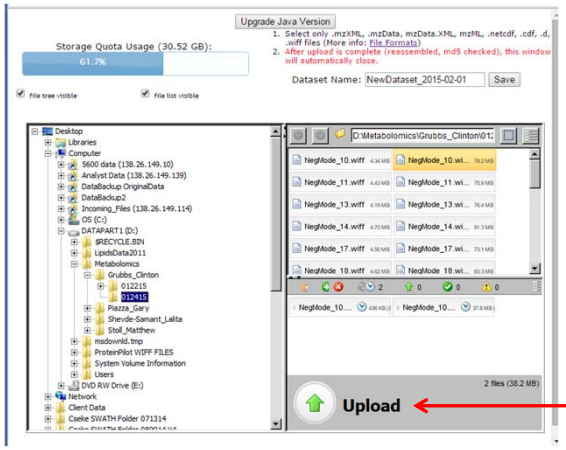

Storage Quota Usage (39.80 GB):
85.9%
file tree visible
file list visible
1. Select only .mzXML, .mzData, mzData.XML, mzML, .netcdf, .cdf, .d, .wiff files (More info: File Formats)
2. After upload is complete (reassembled, md5 checked), this window will automatically close.
Dataset Name: NewDataset_2016-04-07
Save
/
.DocumentRevisions-V100
.fseventsd
.MobileBackups
.PKInstallSandboxManager
.Spotlight-V100
.Trashes
.vol
anaconda
Applications
bin
cores
dev
etc
home
Library
lost+found
net
Network
opt
private
sbin
System
anaconda
Applications
bin
cores
dev
etc
home
Library
lost+found
net
Drop Files Here
No files
Upload


Storage Quota Usage (39.80 GB):
85.9%
file tree visible
file list visible
1. Select only .mzXML, .mzData, mzData.XML, mzML, .netcdf, .cdf, .d, .wiff files (More info: File Formats)
2. After upload is complete (reassembled, md5 checked), this window will automatically close.
Dataset Name: NewDataset_2016-04-07
Save
.Trashes
.vol
anaconda
Applications
bin
cores
dev
etc
home
Library
lost+found
net
Network
opt
private
sbin
System
tmp
User Information
Users
Guest
Shared
stephenbarnes
/Users/stephenbarnes/Desktop/(
NegMode_IR1.... 1.04 GB
NegMode_IR2.... 983 MB
NegMode_IR3.... 976 MB
NegMode_NR1.... 976 MB
NegMode_NR2.... 982 MB
NegMode_NR3.... 989 MB
Click here to select files
Drop Files Here
No files
Upload


Storage Quota Usage (39.80 GB):
85.9%
file tree visible
file list visible
1. Select only .mzXML, .mzData, mzData.XML, mzML, .netcdf, .cdf, .d, .wiff files (More info: File Formats)
2. After upload is complete (reassembled, md5 checked), this window will automatically close.
Dataset Name: NewDataset_2016-04-07
Save
.Trashes
.vol
anaconda
Applications
bin
cores
dev
etc
home
Library
lost+found
net
Network
opt
private
sbin
System
tmp
User Information
Users
Guest
Shared
stephenbarnes
/Users/stephenbarnes/Desktop/(
NegMode_IR1.... 1.04 GB
NegMode_IR2.... 983 MB
NegMode_IR3.... 976 MB
NegMode_NR1.... 976 MB
NegMode_NR2.... 982 MB
NegMode_NR3.... 989 MB
In progress
3 files (2.41 GB)
Upload


Storage Quota Usage (39.80 GB):
85.9%
file tree visible
file list visible
.wiff files (More info: File Formats)
2. After upload is complete (reassembled, md5 checked), this window will automatically close.
Dataset Name: Diet_neg_NR_2016-04-07
Save
.Trashes
.vol
anaconda
Applications
bin
cores
dev
etc
home
Library
lost+found
net
Network
opt
private
sbin
System
tmp
User Information
Users
Guest
Shared
stephenbarnes
/Users/stephenbarnes/Desktop/(
NegMode_IR1.... 1.04 GB
NegMode_IR2.... 983 MB
NegMode_IR3.... 976 MB
NegMode_NR1.... 976 MB
NegMode_NR2.... 982 MB
NegMode_NR3.... 989 MB
3 files (2.17 GB)
Upload
0.5% completed...
03:10:06

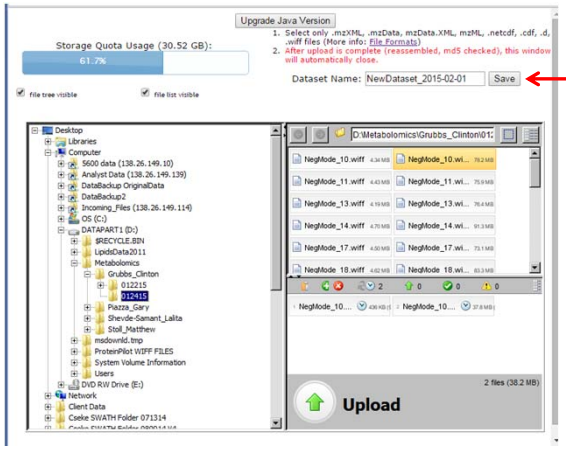
Uploading Experiments for Metabolomics Analyses using XCMS Online Applications continued…



Activate the "Upload" button to begin the process. Depending on the size and number of the files, this process can take up several hours.

If any errors occur in the uploading process, the Exclamation point icon will have a number associated with the amount of files that have issues.

Uploading Experiments for Metabolomics Analyses using XCMS Online Applications continued…



After the upload is complete, give a descriptive name to your data set by typing in the *"DataSet Name"* box and clicking Save.

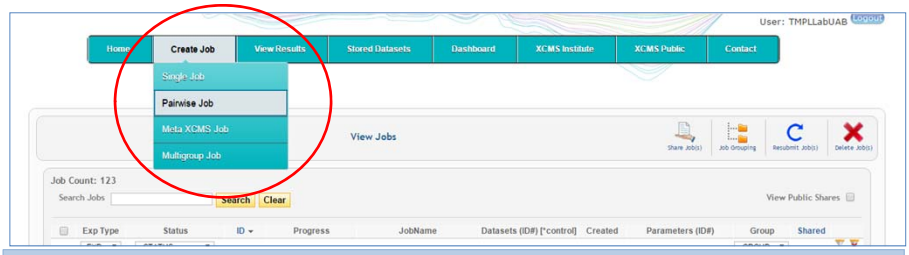
Since you will be comparing a minimum of two data sets, you will repeat the process again with the additional group(s) you wish to evaluate.

Tips for naming files for upload to XCMS Online

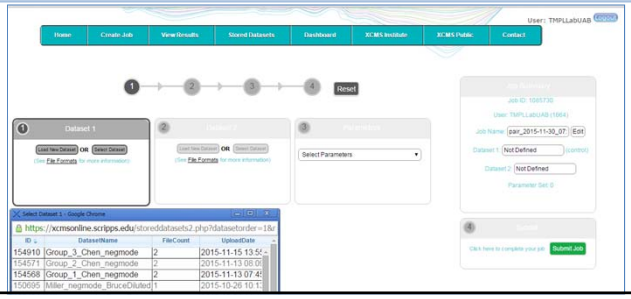
- Before uploading, it is a good idea to create separate file folders on your hard drive to better organize your data into the groups you want to examine.
- Add enough description to discriminate between different samples and sample set names.
- Adding the ionization mode in the name is preferable, i.e. 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: Grubbs_UrineGroup2_NegMode

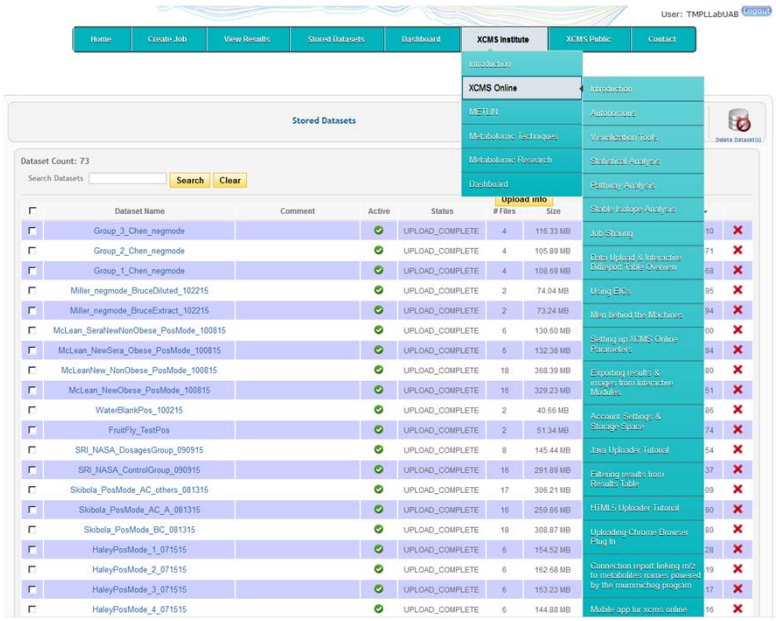
Starting a new XCMS database search…



Select the type of "Job" or test for your newly uploaded data sets. From here, you can select the files you wish to evaluate.



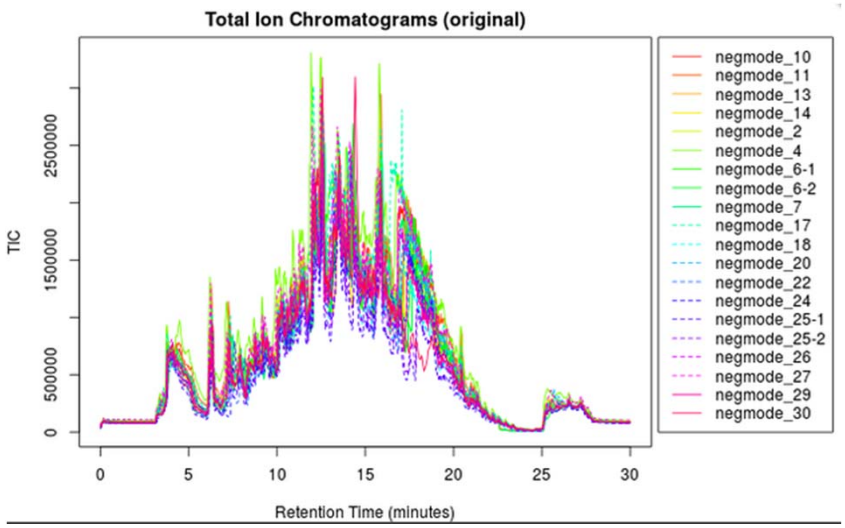
XCMS Online Tutorial Videos



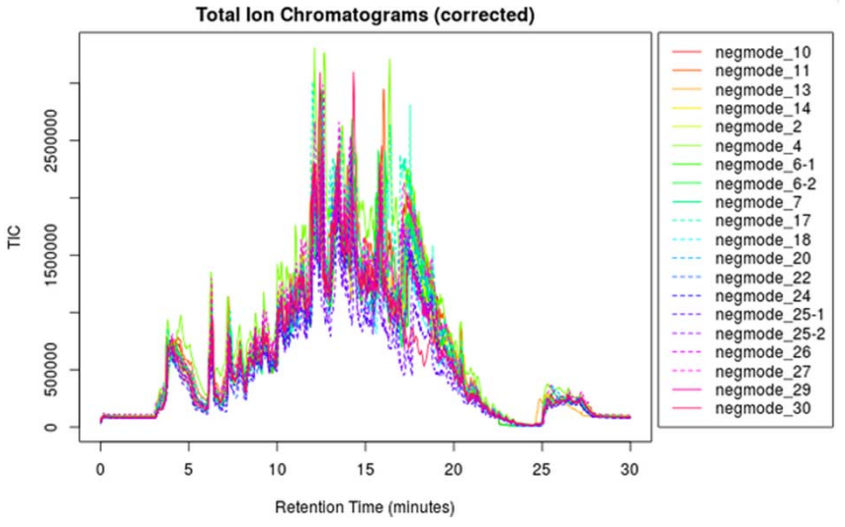
What does a LC-MS data set consist of?

- **A Q-TOF instrument during the LC run, for example, acquires data on a 2-second duty cycle**
  - **0-100 msec**
    - **High resolution/mass accuracy MS spectrum**
  - **100-2000 msec**
    - **A succession of selected MSMS spectra**
    - **If each MSMS spectrum is collected for 100 msec, then 19 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 90 sec**

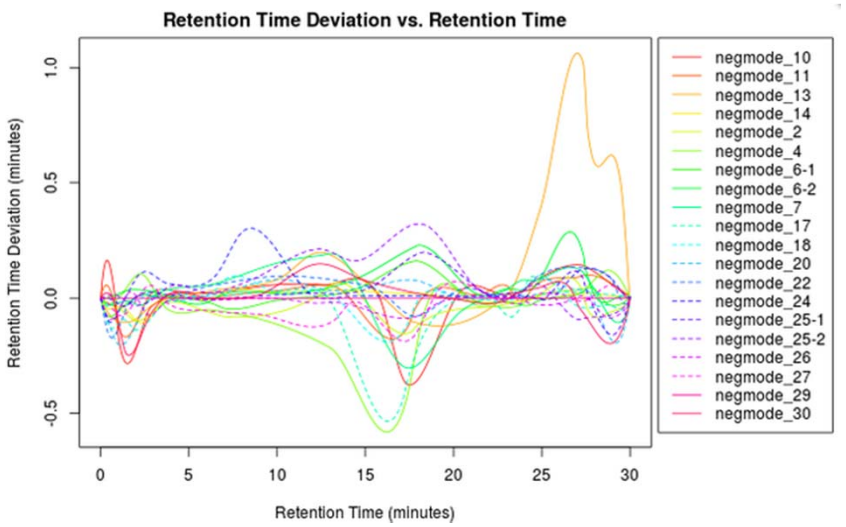
Overlays on urine data



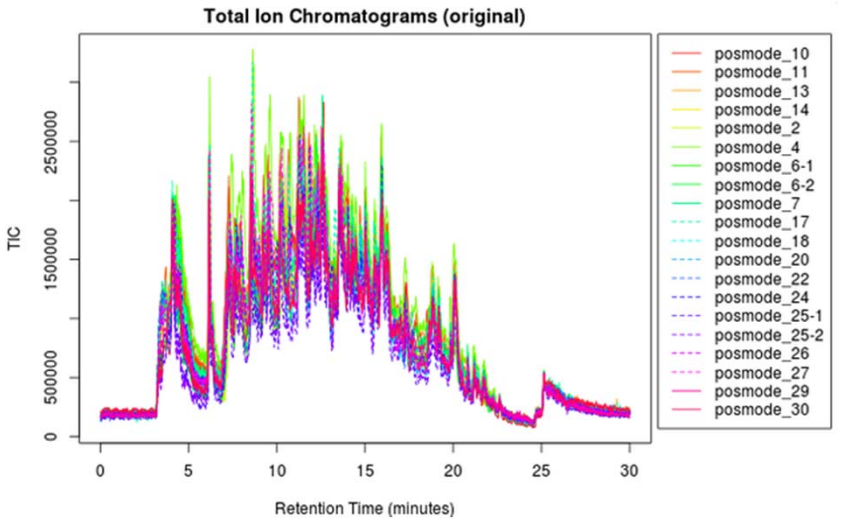
Overlays of urine data



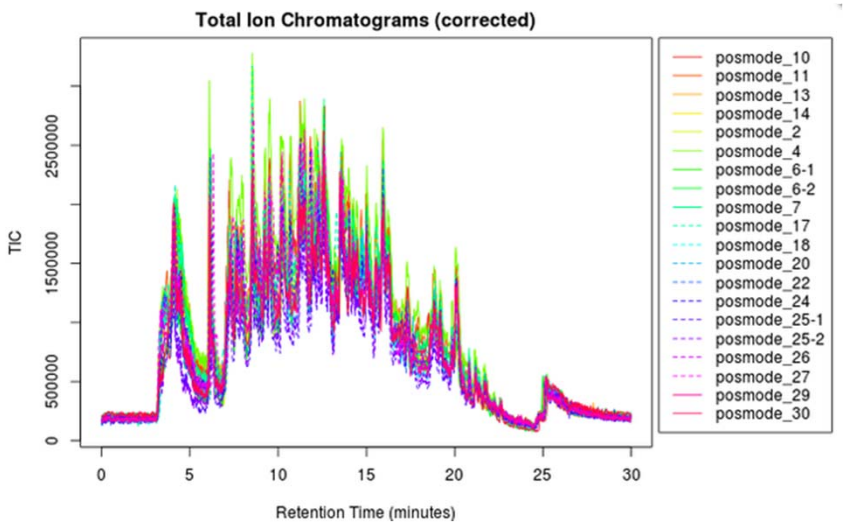
Retention time variation for urines



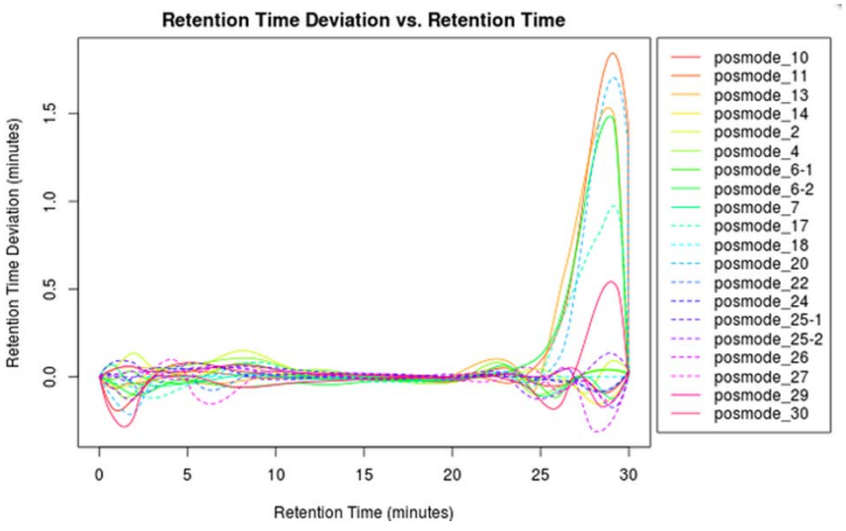
Overlays of urine data



Overlays of urine data



Retention time variation for urines



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