



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/>

Home FAQ Contact Sign Up Forgot Password? email address ***** Login

XCMS

The original and most widely used metabolomic platform

Available on the Google play App Store

10550 North Torrey Pines Road BCC-007, La Jolla, CA 92037 USA - (858) 784-9415, Fax (858) 784-9436

Home Privacy Policy Terms of Use Contact Us

XCMS™ and METLIN™ are trademarks of The Scripps Research Institute

XCMS Short Course

Signing up for XCMS

Home News XCMS Stream Toolbox Help Sign Up Enter email address Enter password Login

Forgot Password?

Note: All new users require e-mail verification. A verification e-mail will be sent to the e-mail address you provide. Please add "scripps.edu" and "xcmsonline@gmail.com" to your whitelist.

You can check your system compatibility with our [System compatibility test](#)

N.B.: TSRI users already have accounts (via LDAP).

First Name

Last (Family) Name

Organization

Email


Password

Re-enter password

This will be your permanent e-mail address to be used for:

- registration
- password resets
- job alerts (errors, completed)
- technical support
- system notifications (e.g. maintenance)

Enter the code below password



This will be your permanent e-mail address to be used for:

- registration
- password resets
- job alerts (errors, completed)
- technical support
- system notifications (e.g. maintenance)

Enter whatever appears here

Participate in anonymous usage statistics ([View Details](#))

[Register](#)


Starting page for XCMSonline


Home
Create Job -
View Results
XCMS Public
XCMS Institute
Stored Datasets
Account
FAQ
Contact
Logout [TMPLLabUAB]

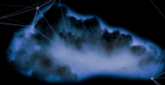


XCMS

The original and most widely used metabolomic platform





10550 North Torrey Pines Road BCC-007, La Jolla, CA 92037 USA - (858) 784-9415, Fax (858) 784-9496
[Home](#) | [Privacy Policy](#) | [Terms of Use](#) | [Contact Us](#) | [Reset Password](#)


XCMS™ and METLIN™ are trademarks of The Scripps Research Institute

XCMS
Short
Course


View Public Job Shares

Show 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. <i>Assignment of Endogenous Substrates to Enzymes by Global Metabolite Profiling, Biochemistry</i> , 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. <i>Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin, Nature Chemical Biology</i> 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. <i>metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data, Analytical Chemistry</i> 2011 83 (3), 696-700	<i>Halobacterium salinarum</i> knockout strains



Results: Coke vs Pepsi



Pairwise Results Summary: Coke vs. Pepsi (#112691) [Download Results](#)

Submit Date	Finish Date	Paired Samples	Total Aligned Features	Parameter ID#	Log	Shared
2015-02-02 00:00:29	2015-02-02 01:08:55	False	2553	HPLC / Q-TOF II (9569)	View Log	Public Share

The finished job has the following notes:
 2015-02-02 01:08:57 : insufficient data prep. memory requires limiting to top 1000 features <= 132261 p-values
 2015-02-02 01:05:26 : There are regions with poor chromatographic resolution. Feature annotations (CAMERA) were omitted for these regions.

View Results Table

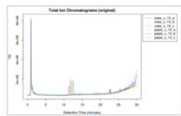
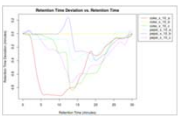
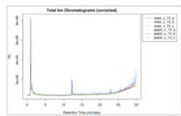
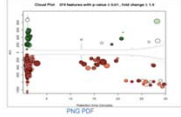


[View Interactive Cloud PPI](#)

[View Interactive Heatmap](#)

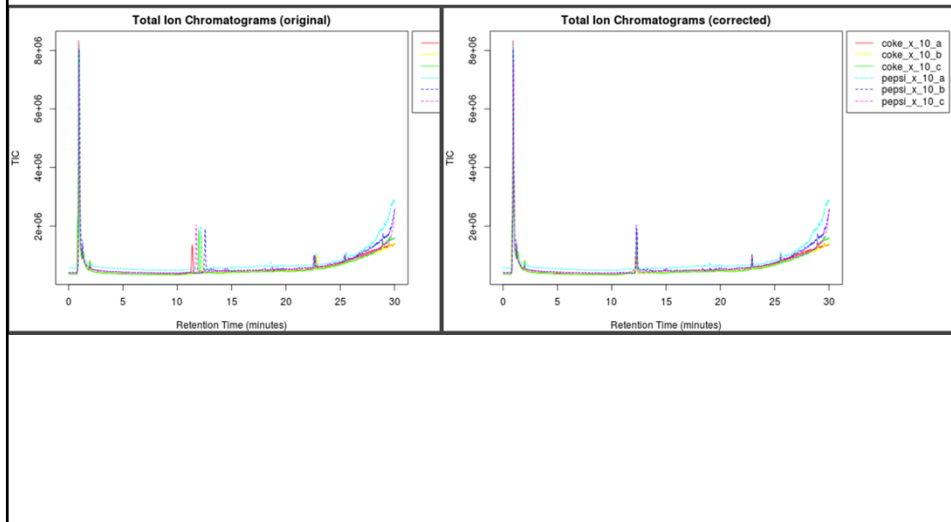
[View PCA](#)

Datasets Used

- Coke (1500)
- Pepsi (1500)

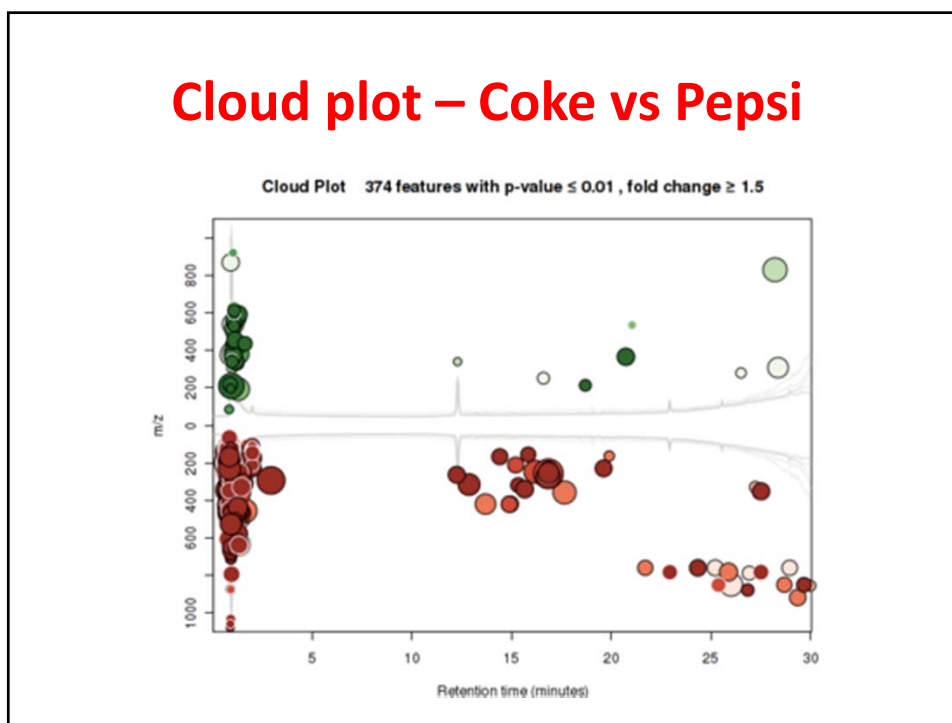
Rt correction - Coke vs Pepsi



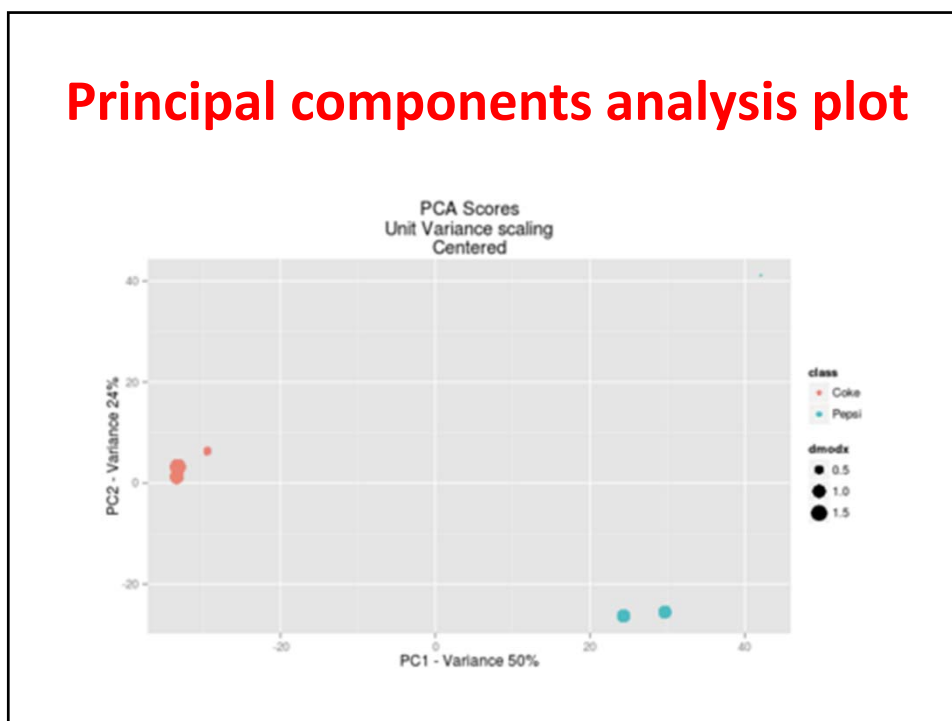
Variation in retention time



Cloud plot – Coke vs Pepsi



Principal components analysis plot



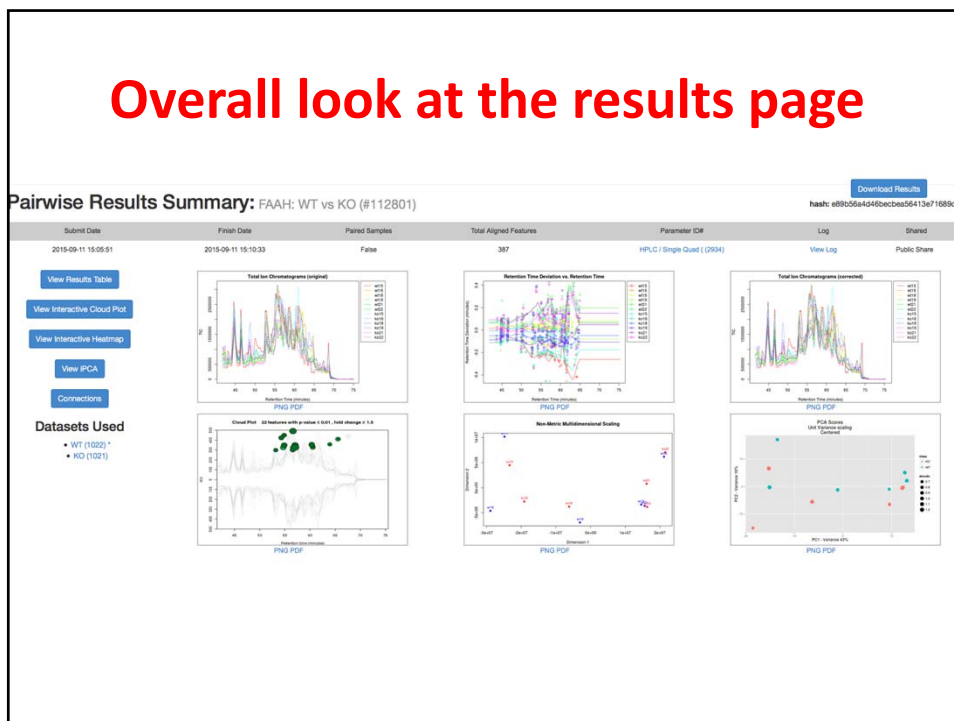
Let's look at the FAAH – WT vs KO data

View Public Job Shares

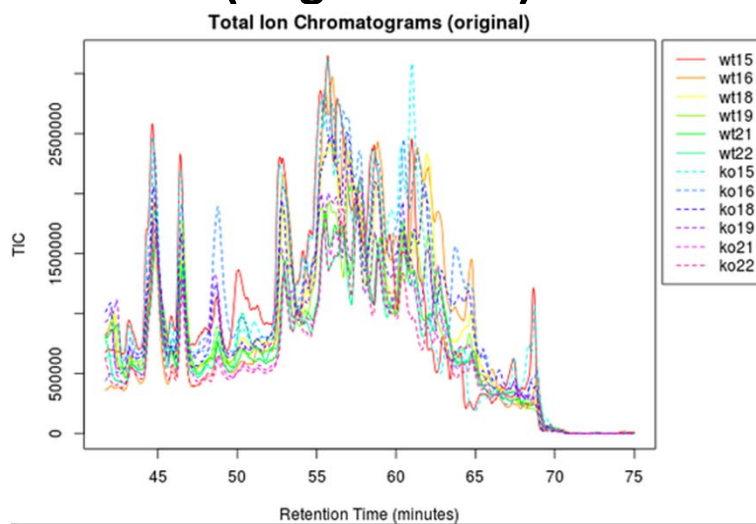
Show entries Search:

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. <i>Assignment of Endogenous Substrates to Enzymes by Global Metabolite Profiling</i> , <i>Biochemistry</i> , 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. <i>Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin</i> , <i>Nature Chemical Biology</i> 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. <i>metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data</i> , <i>Analytical Chemistry</i> 2011 83 (3), 696-700	<i>Halobacterium salinarum</i> knockout strains
VIEW	126996	URA3 vs VNG1816G	Ralf Tautenhahn	Tautenhahn et al. <i>metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data</i> , <i>Analytical Chemistry</i> 2011 83 (3), 696-700	<i>Halobacterium salinarum</i> knockout strains

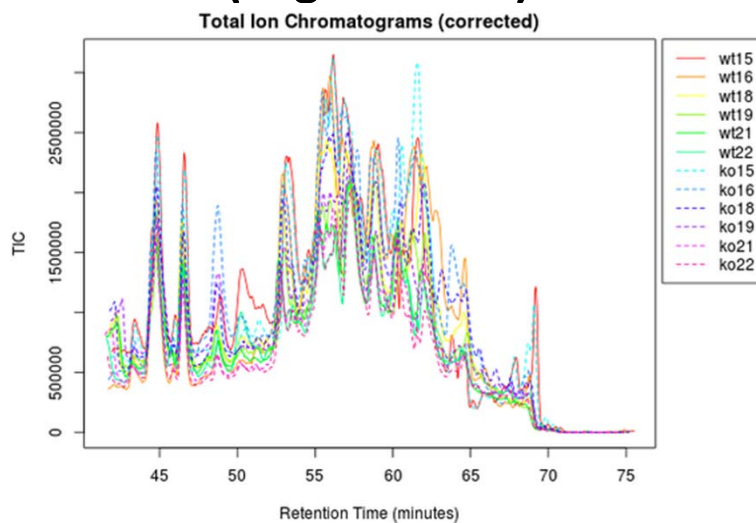
Overall look at the results page



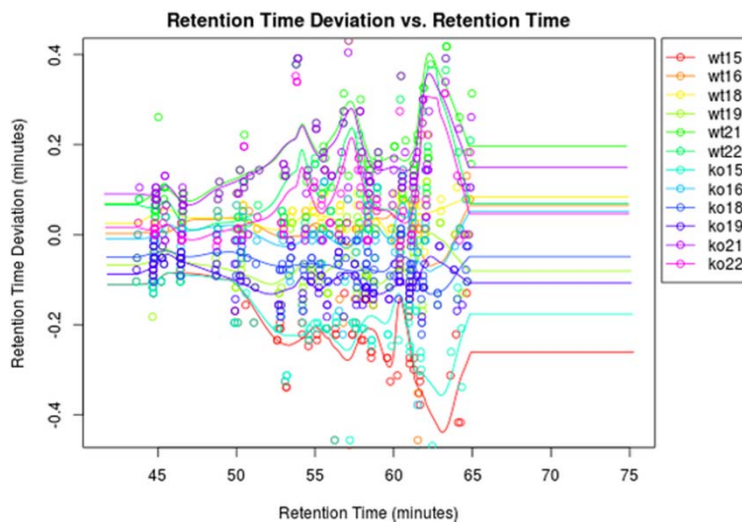
Overlay of ion chromatograms (original data)



Overlay of ion chromatograms (aligned data)



Alignment corrections for FAAH data

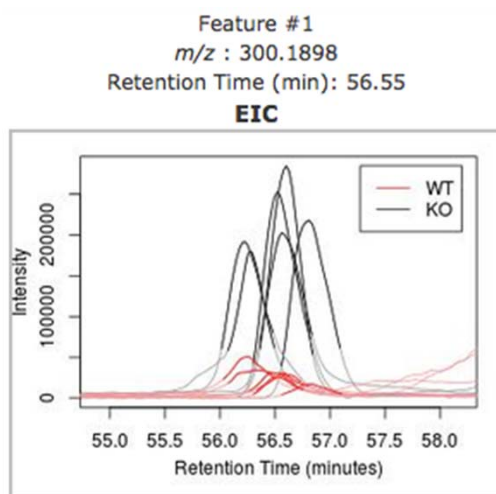


Viewing the data table for FAAH

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

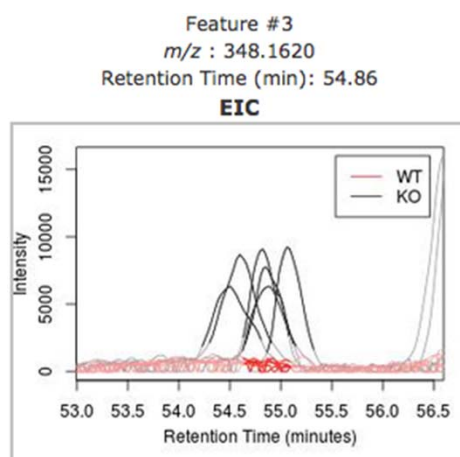
featureidx	fold	pvalue	updown	mzmed	rtmed	maxint	dataset1_n	dataset2_n	isotopes	adducts	peakgroi	usernotes
1	5.7	5.02634e-8	UP	300.1898	56.55	284,800	820,662	4,672,516	[9][M]+	[M+H]+ 29	20	
2	6.1	5.67701e-8	UP	301.1879	56.55	58,696	159,994	976,249	[9][M+1]+		12	
3	10.3	4.32371e-8	UP	348.1620	54.86	9,212	15,224	156,406			119	
4	22.3	8.17994e-8	UP	491.2000	56.61	16,104	16,106	359,085			127	
5	6.3	0.00003	UP	423.1499	54.29	11,119	32,573	204,905				
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



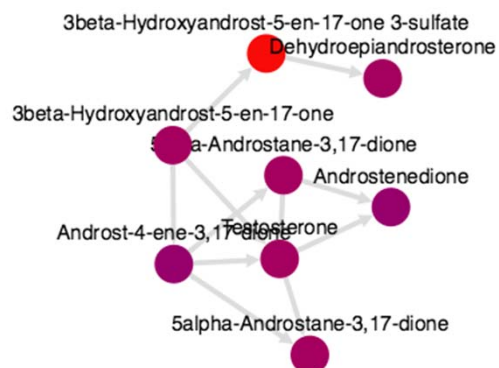
retention variation is greater than 1 min

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



XCMS Online is a metabolomic data technology based out of the Center for Metabolomics at the Scripps Research Institute. XCMS was originally developed in 2005 and has evolved to a cloud-based platform. XCMS Online provides high-quality metabolomic analysis in an intuitive format that allows users to easily upload LC/MS metabolomic data that can then be processed with a few simple mouse clicks. You will need a user account to use the system, user accounts are free and only require a valid e-mail address to register. Please go to these links to download the user manual, view demo dataset, contact us for additional space, or view options for your own onsite XCMS.

<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.

Stored Datasets

Tree View Duplicate Dataset(s) **Add Dataset(s)** Delete Dataset(s)

Dataset Count: 83

Search Datasets

<input type="checkbox"/>	Dataset Name	Comment	Active	Status	# Files	Size	Avg. Upload Speed	Upload Date	ID	
<input type="checkbox"/>	ControlCal_Elmet_PosMode_031816		<input checked="" type="checkbox"/>	UPLOAD_COMPLETE	2	117.53 MB	111.03 kB/s	2016-03-18 08:11:07	178281	✗
<input type="checkbox"/>	GrapeExtract_Elmet_PosMode_031816		<input checked="" type="checkbox"/>	UPLOAD_COMPLETE	2	38.16 MB	111.63 kB/s	2016-03-16 08:04:04	178279	✗
<input type="checkbox"/>	Vivek_Set2_NegMode_Group2(No)_012516		<input checked="" type="checkbox"/>	UPLOAD_COMPLETE	10	298.84 MB	77.95 kB/s	2018-01-25 11:01:47	167581	✗

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

Dataset Name:

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.


[Activate Java.](#)

Storage Quota Usage (39.80 GB): 85.9%

file tree visible file list visible

Dataset Name:

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.

Storage Quota Usage (39.80 GB): 85.9%

file tree visible file list visible

Dataset Name:

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.

Storage Quota Usage (39.80 GB): 85.9%

file tree visible file list visible

Dataset Name:

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.

File list:

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

Upload progress: 3 files (2.41 GB)

Upload

Storage Quota Usage (39.80 GB): 85.9%

file tree visible file list visible

Dataset Name:

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.

File list:

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

Upload progress: 3 files (2.17 GB)

Upload 0.5% completed... 03:10:06

Uploading Experiments for Metabolomics Analyses using XCMS Online Applications continued...

Storage Quota Usage (30.52 GB): 61.7%

Upgrade Java Version

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_2015-02-01 Save

Desktop Libreries Computer Analyst Data (138.26.149.10) DataBackup OriginalData DataBackup2 Incoming_Files (138.26.149.114) OS (C:) DATAPART1 (D:) @RCYCLE.82N LpidsData2011 Metabolomics Grubbs_Clinton 012215 201415 Piazza_SerY Shevde-Samant_Lalita Stal_Matthew mdown6d.tmp ProteoPilot WIFF FILES System Volume Information Users DVD RW Drive (E:) Network Client Data Ceele SWATH Folder 071314 Public SWATH Folder 08081414

D:\Metabolomics\Grubbs_Clinton\01

NeghNode_10.wiff 4.38 MB NeghNode_10.wl... 702.9 KB
 NeghNode_11.wiff 4.43 MB NeghNode_11.wl... 703.9 KB
 NeghNode_13.wiff 4.19 MB NeghNode_13.wl... 704.4 KB
 NeghNode_14.wiff 4.79 MB NeghNode_14.wl... 913.9 KB
 NeghNode_17.wiff 4.80 MB NeghNode_17.wl... 713.1 KB
 NeghNode_18.wiff 4.82 MB NeghNode_18.wl... 813.9 KB

2 files (38.2 MB)

Upload

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...

Storage Quota Usage (30.52 GB): 61.7%

Upgrade Java Version

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_2015-02-01 Save

Desktop Libreries Computer Analyst Data (138.26.149.10) DataBackup OriginalData DataBackup2 Incoming_Files (138.26.149.114) OS (C:) DATAPART1 (D:) @RCYCLE.82N LpidsData2011 Metabolomics Grubbs_Clinton 012215 201415 Piazza_SerY Shevde-Samant_Lalita Stal_Matthew mdown6d.tmp ProteoPilot WIFF FILES System Volume Information Users DVD RW Drive (E:) Network Client Data Ceele SWATH Folder 071314 Public SWATH Folder 08081414

D:\Metabolomics\Grubbs_Clinton\01

NeghNode_10.wiff 4.38 MB NeghNode_10.wl... 702.9 KB
 NeghNode_11.wiff 4.43 MB NeghNode_11.wl... 703.9 KB
 NeghNode_13.wiff 4.19 MB NeghNode_13.wl... 704.4 KB
 NeghNode_14.wiff 4.79 MB NeghNode_14.wl... 913.9 KB
 NeghNode_17.wiff 4.80 MB NeghNode_17.wl... 713.1 KB
 NeghNode_18.wiff 4.82 MB NeghNode_18.wl... 813.9 KB

2 files (38.2 MB)

Upload

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

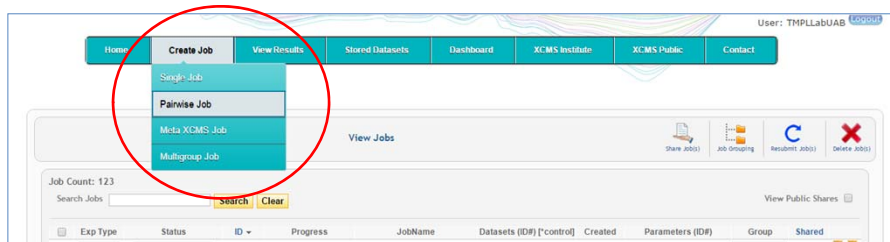
Since you will be comparing a minimum of two data sets, you will repeat the process again with the additional group(s) you wish to evaluate.

Tips for naming files for upload to XCMS Online

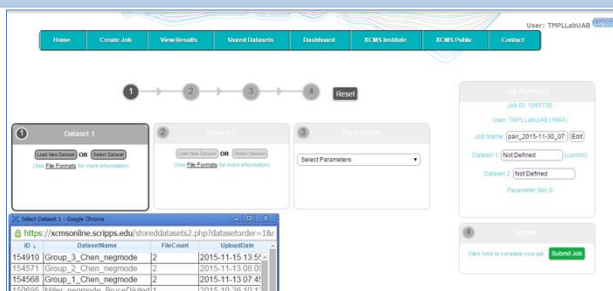
- Before uploading, it is a good idea to create separate file folders on your hard drive to better organize your data into the groups you want to examine.
- Add enough description to discriminate between different samples and sample set names.
- Adding the ionization mode in the name is preferable, i.e.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

User: TEMPLABLAB 100000

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

Stored Datasets

Dataset Count: 71

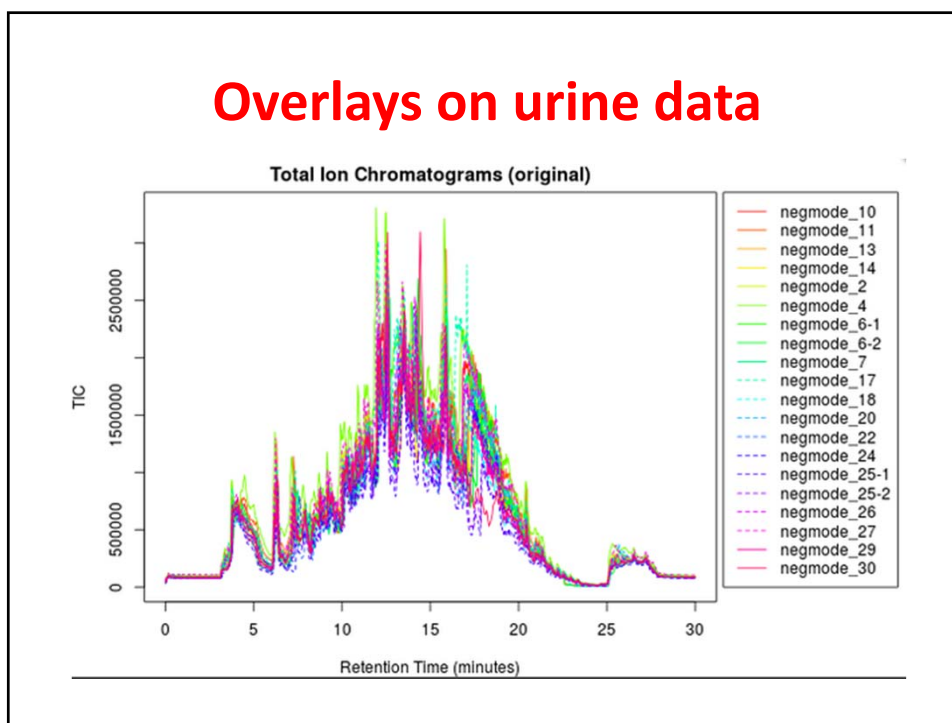
Search Datasets:

<input type="checkbox"/>	Dataset Name	Comment	Active	Status	Upload info	# Files	Size	Stable History Analysis	
<input type="checkbox"/>	Group_3_Chem_negmode		✓	UPLOAD_COMPLETE		4	116.33 MB	Job Streaming	10 ✗
<input type="checkbox"/>	Group_2_Chem_negmode		✓	UPLOAD_COMPLETE		4	105.89 MB	Data Upload & Inspection	71 ✗
<input type="checkbox"/>	Group_1_Chem_negmode		✓	UPLOAD_COMPLETE		4	109.69 MB	DMSPop: Table Overview	68 ✗
<input type="checkbox"/>	Miller_negmode_BruceDiluted_102215		✓	UPLOAD_COMPLETE		2	74.04 MB	Using EICs	65 ✗
<input type="checkbox"/>	Miller_negmode_BruceExtract_102215		✓	UPLOAD_COMPLETE		2	73.24 MB	How to find the Mixtures	64 ✗
<input type="checkbox"/>	McLean_SeraNewNonObese_PosMode_100815		✓	UPLOAD_COMPLETE		6	130.60 MB	Setting up XCMS Online	60 ✗
<input type="checkbox"/>	McLean_NewSera_Obese_PosMode_100815		✓	UPLOAD_COMPLETE		6	132.38 MB	Pharmacokinetics	64 ✗
<input type="checkbox"/>	McLeanNew_NonObese_PosMode_100815		✓	UPLOAD_COMPLETE		18	368.39 MB	Exporting results & images from Interaction Modules	60 ✗
<input type="checkbox"/>	McLean_NewObese_PosMode_100815		✓	UPLOAD_COMPLETE		16	329.23 MB		51 ✗
<input type="checkbox"/>	WaterBlankPos_100215		✓	UPLOAD_COMPLETE		2	40.66 MB	Account Settings & Storage Space	86 ✗
<input type="checkbox"/>	Fruffly_TestPos		✓	UPLOAD_COMPLETE		2	51.34 MB		74 ✗
<input type="checkbox"/>	SRJ_NASA_DosagesGroup_090915		✓	UPLOAD_COMPLETE		8	145.44 MB	Local Uploads: Tutorial	54 ✗
<input type="checkbox"/>	SRJ_NASA_ControlGroup_090915		✓	UPLOAD_COMPLETE		16	291.89 MB	Exporting results from Results Editor	37 ✗
<input type="checkbox"/>	Sibola_PosMode_AC_others_081315		✓	UPLOAD_COMPLETE		17	308.21 MB		69 ✗
<input type="checkbox"/>	Sibola_PosMode_AC_A_081315		✓	UPLOAD_COMPLETE		16	299.66 MB	HTML5 Uploader Tutorial	60 ✗
<input type="checkbox"/>	Sibola_PosMode_BC_081315		✓	UPLOAD_COMPLETE		16	308.87 MB	Uploading Chrome Browser Plugin	60 ✗
<input type="checkbox"/>	HaleyPosMode_1_071515		✓	UPLOAD_COMPLETE		6	154.62 MB		28 ✗
<input type="checkbox"/>	HaleyPosMode_2_071515		✓	UPLOAD_COMPLETE		6	162.68 MB	Connection report linking 102 to installation status powered by the instrumenting program	19 ✗
<input type="checkbox"/>	HaleyPosMode_3_071515		✓	UPLOAD_COMPLETE		6	153.23 MB		17 ✗
<input type="checkbox"/>	HaleyPosMode_4_071515		✓	UPLOAD_COMPLETE		6	144.88 MB	Mobile app for xcms online	15 ✗

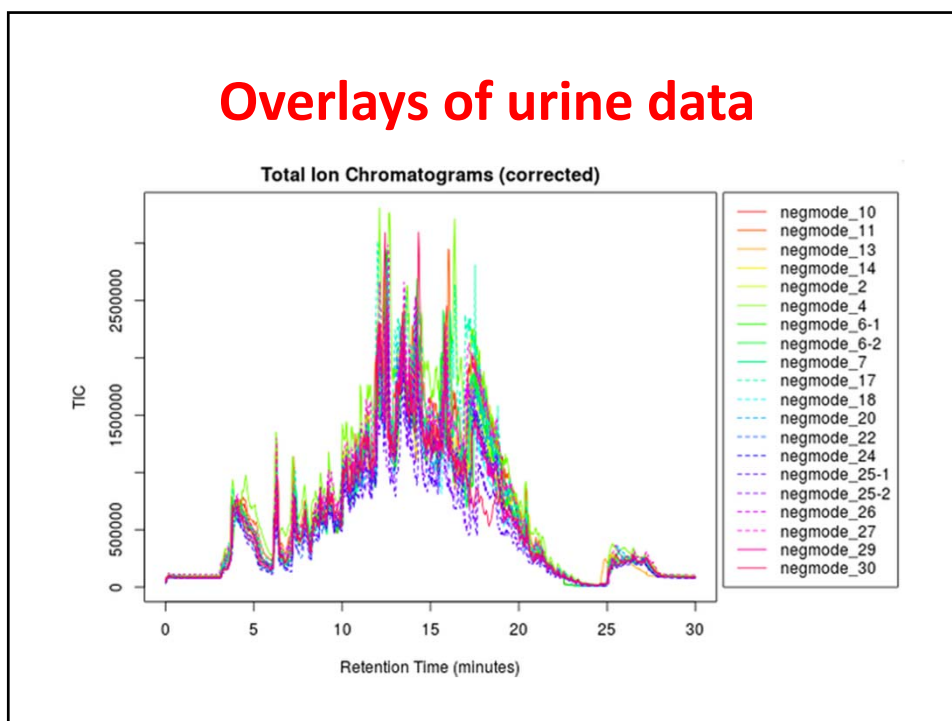
What does a LC-MS data set consist of?

- 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

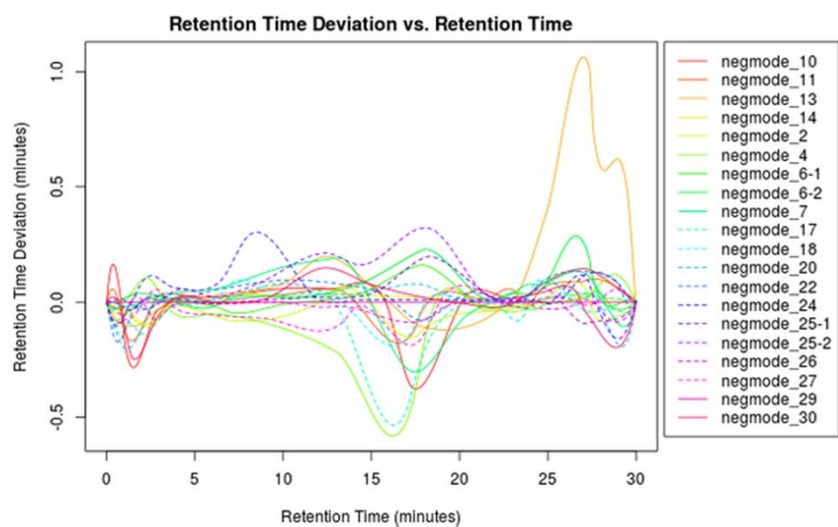
Overlays on urine data



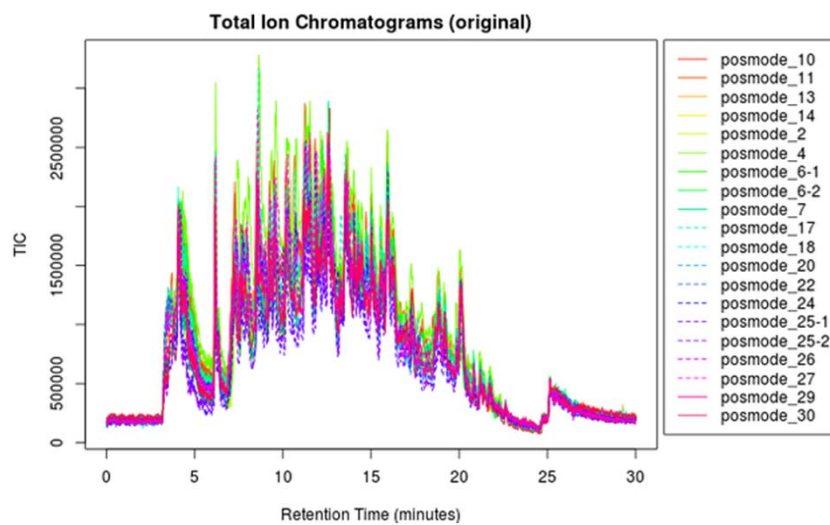
Overlays of urine data



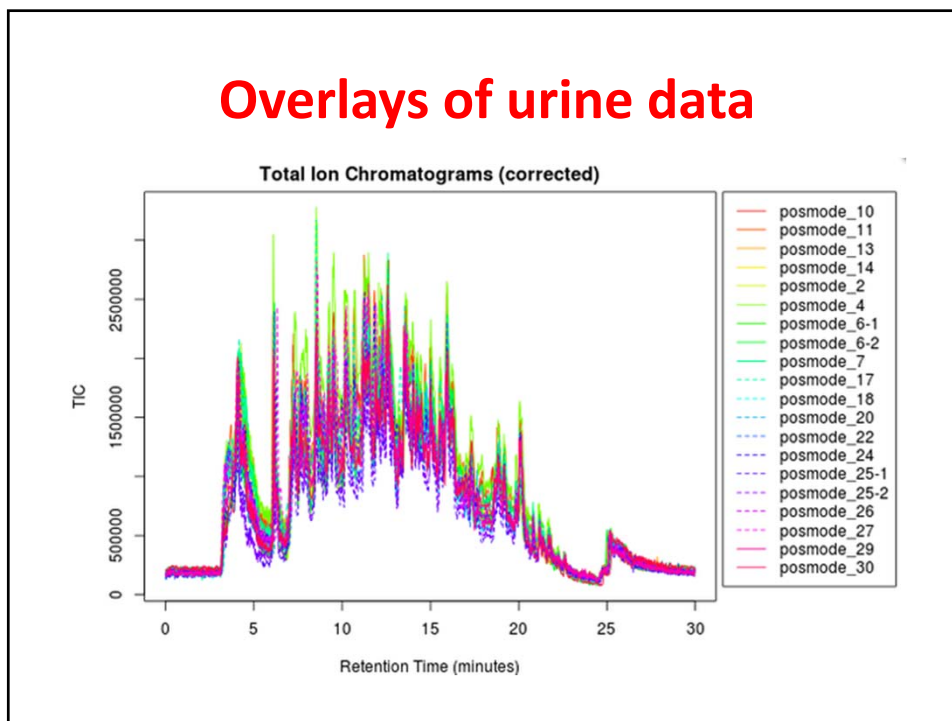
Retention time variation for urines



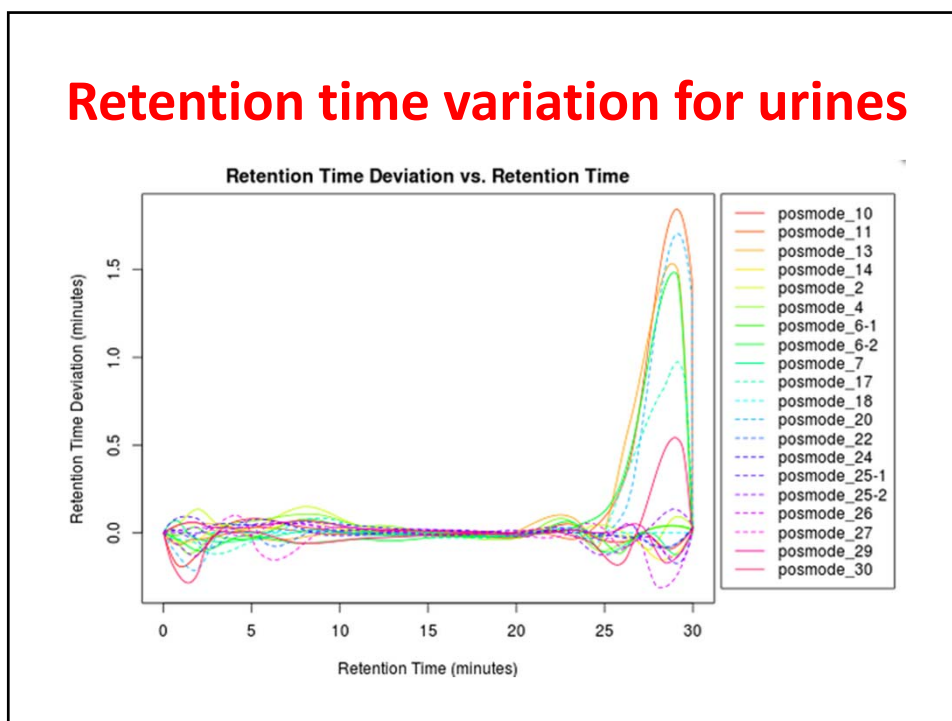
Overlays of urine data



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**