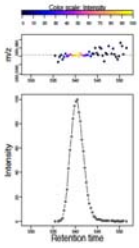
XCMS Online
Metabolomics in the Clouds

H. Paul Benton
The Siuzdak Lab @ The Scripps Research Institute La Jolla, CA

What is XCMS Online ?

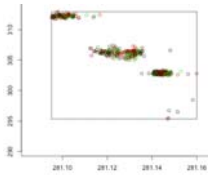
- The complete xcms system in a graphical web based system
- XCMS started in 2005
- First concept of XCMS Online was 2009 but production was started late 2011 (public mid 2012)
  - Approx 13,500 users
  - Average 50 jobs a day
  - 75 Users logging in per day

Overview of XCMS

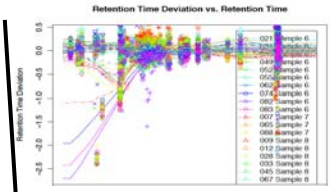


Peak Picking

Grouping similar peaks across replicates



Retention time alignment



Statistical analysis of Peaks Between classes

Peak detection choice

Peak Picking

matchedFilter

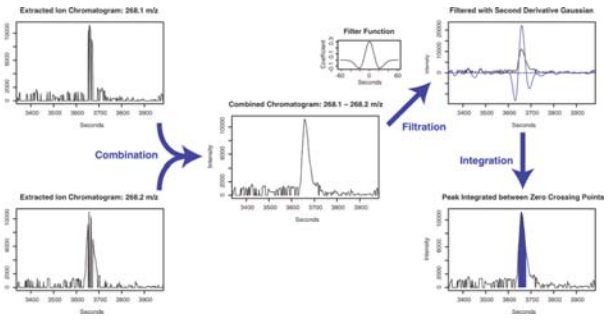
- Profile Data
- Low resolution data
- Original algorithm

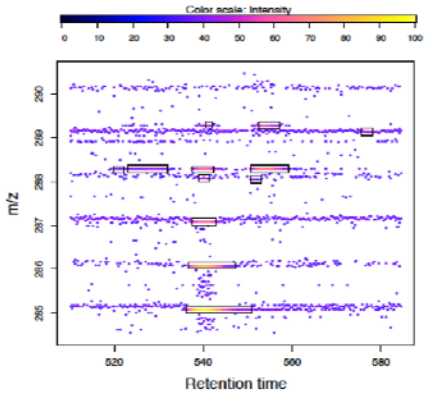
centWave

- Centroid data
- High resolution data
- Separately published algorithm

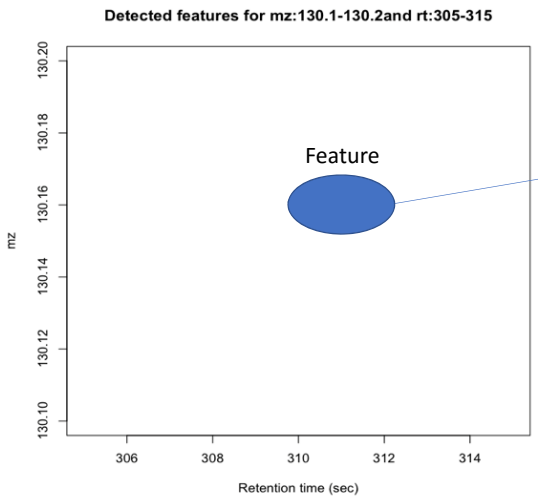
Peak detection choice

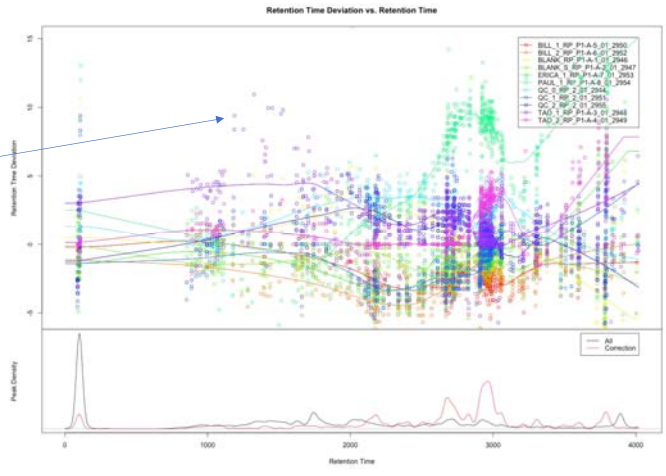
Peak Picking

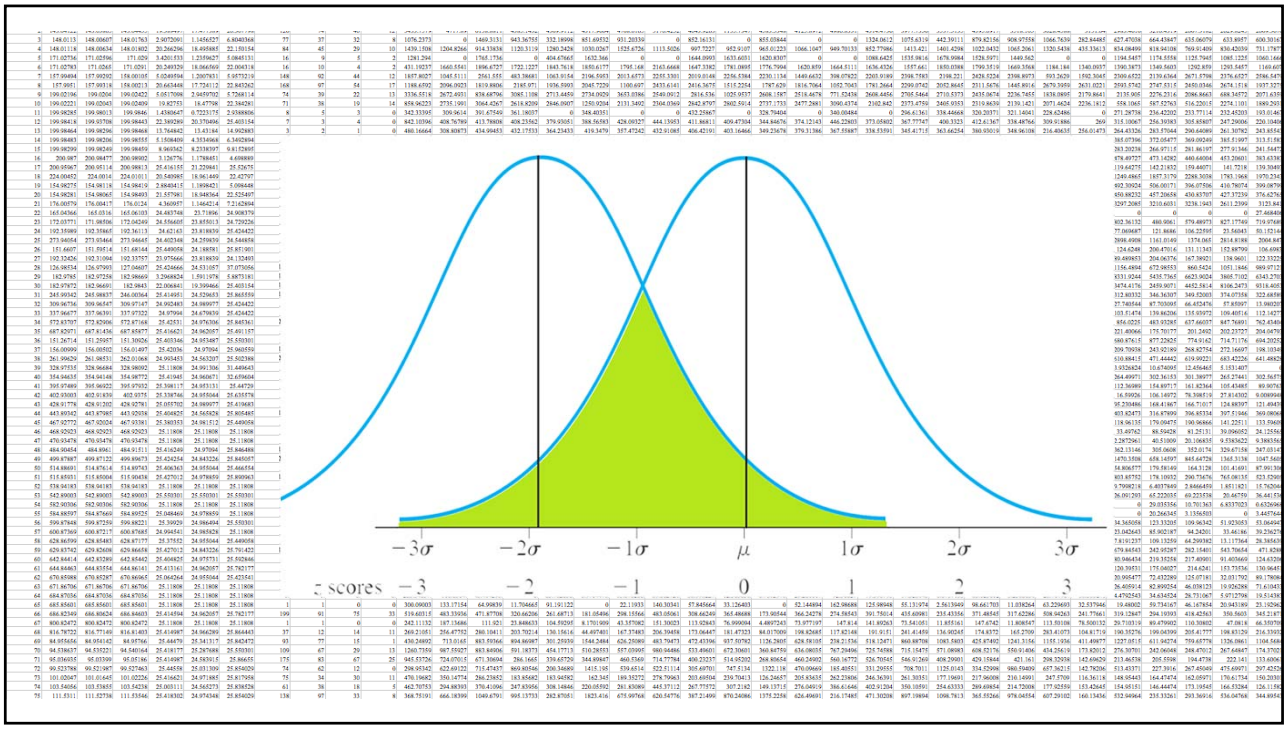




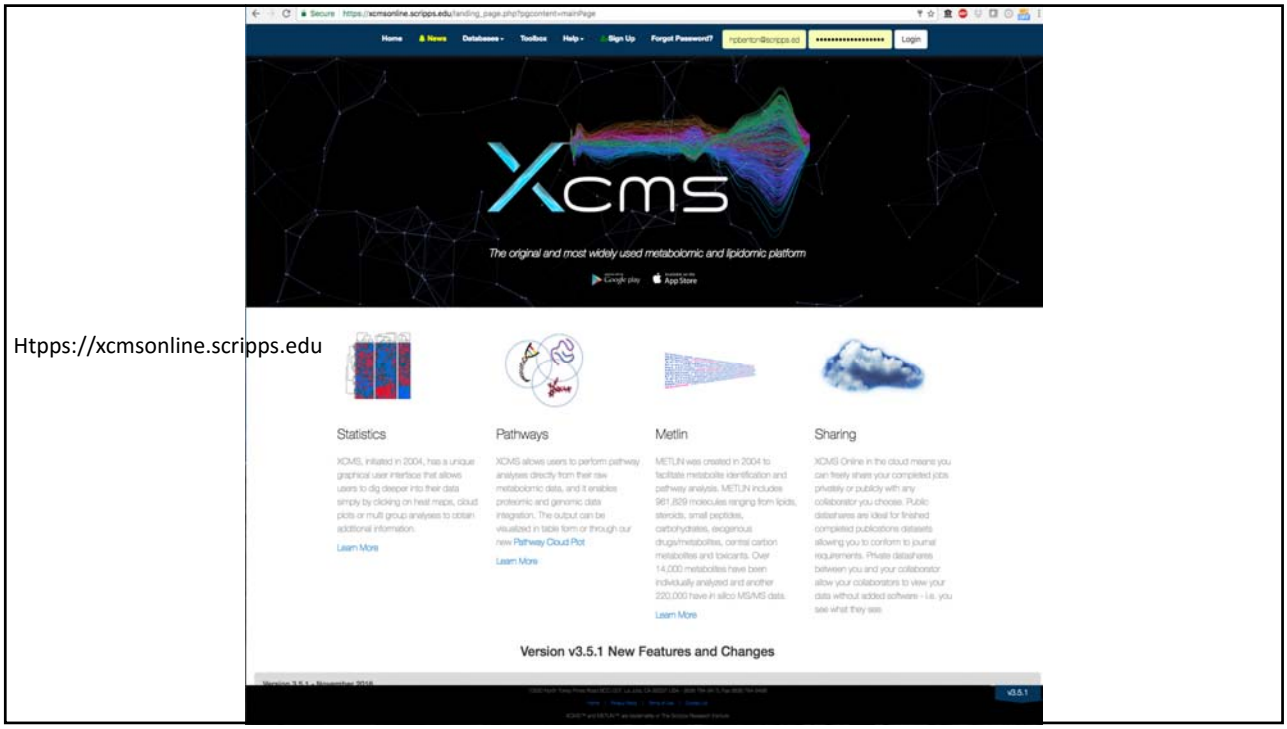
Grouping and Retention time correction





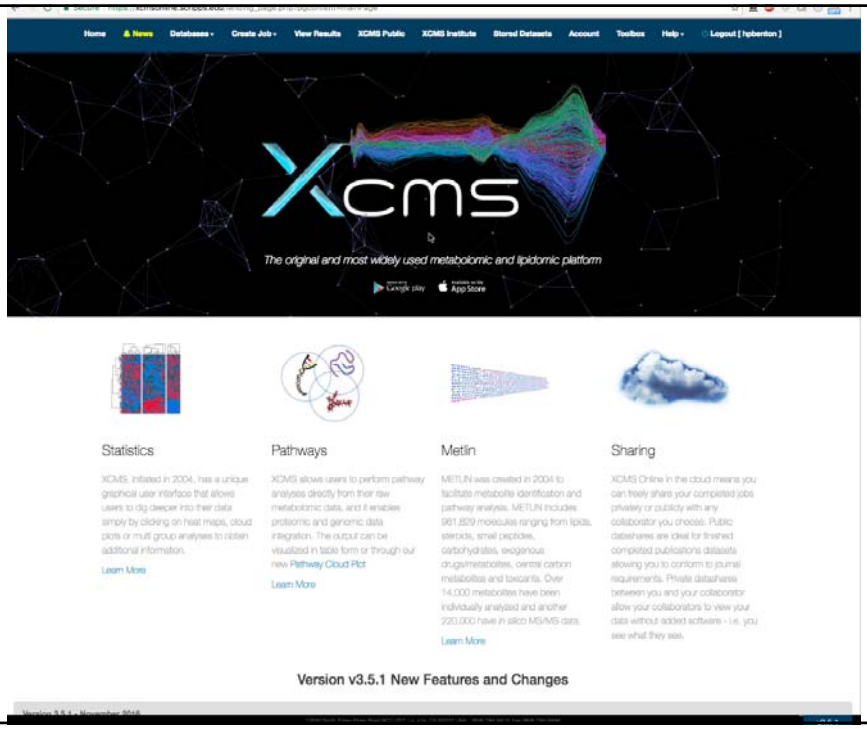


Htpps://xcmsonline.scripps.edu

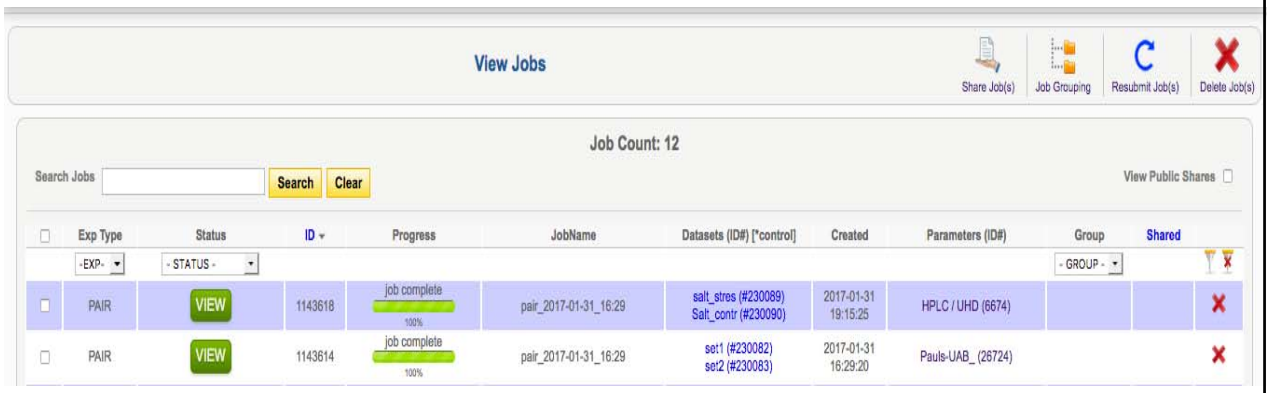


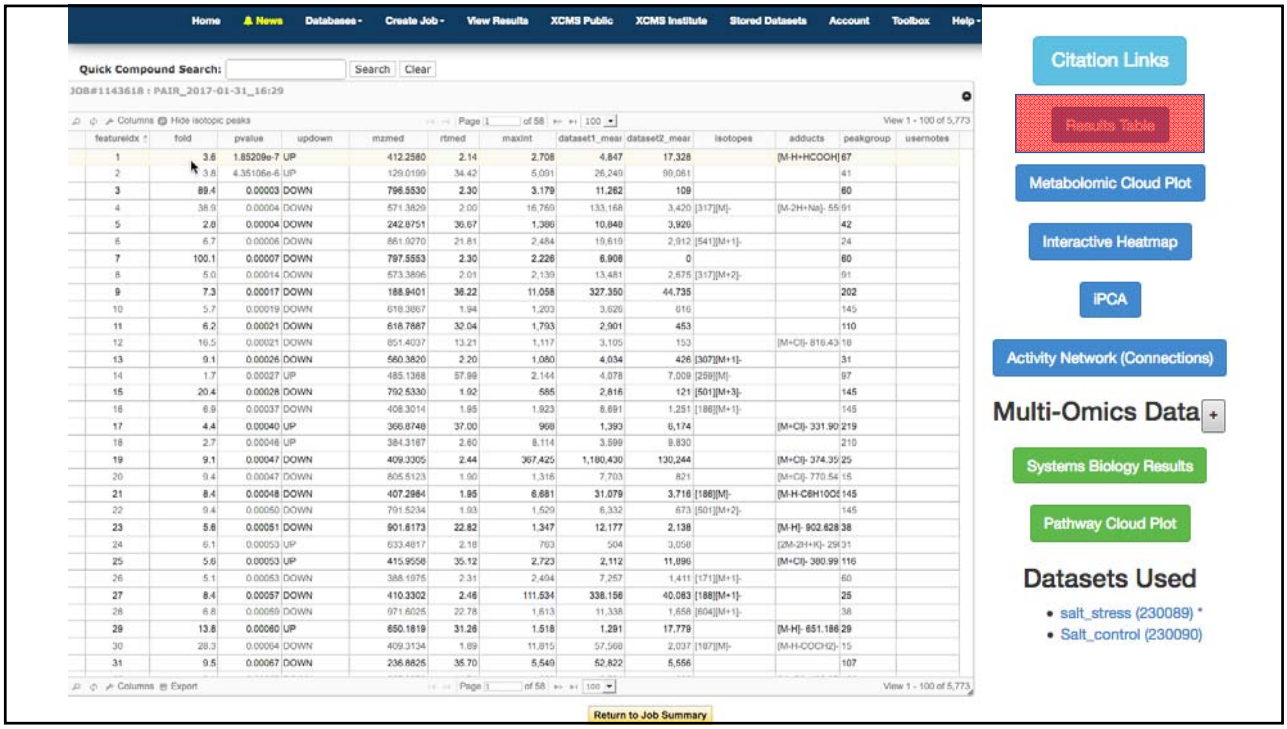
Job types in XCMS Online

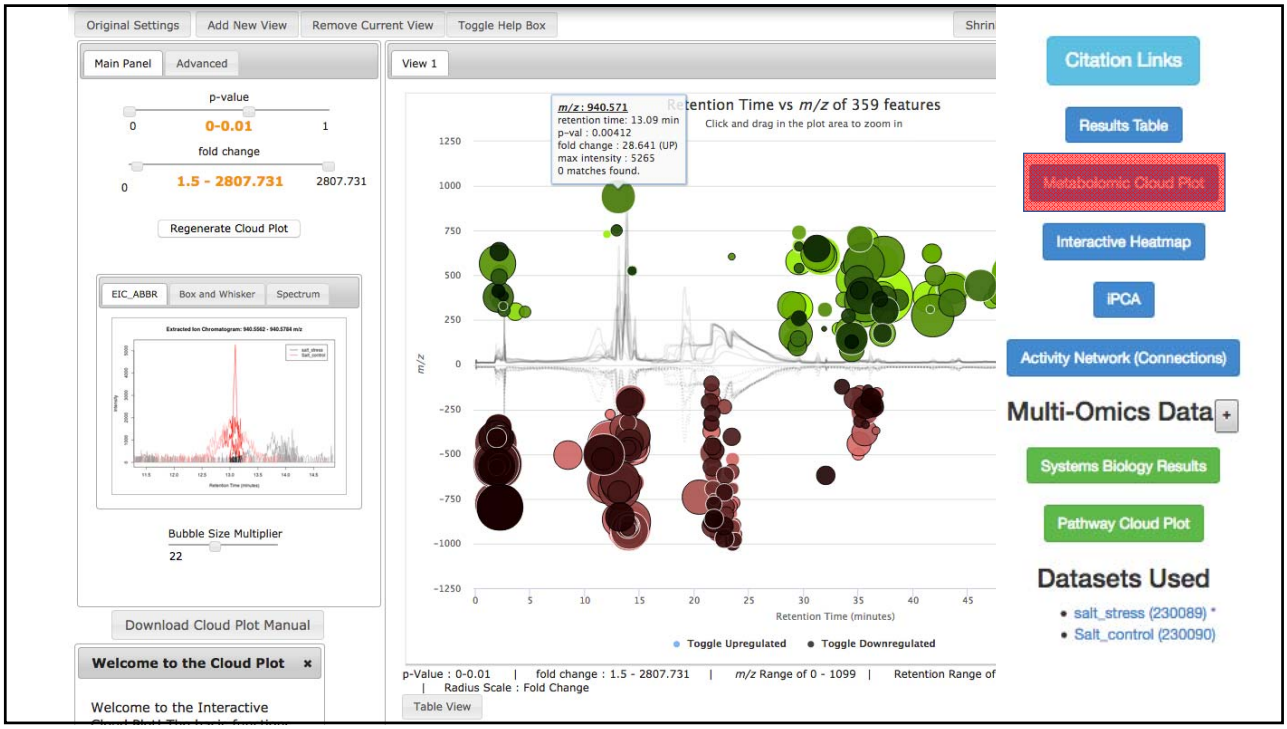
- Single Class Jobs – Either single file or single dataset (class or unknown classes)
- Pairwise analysis Jobs – Two class comparison, works best with KO –vs- WT type experiments
- Multigroup/class Jobs – Multiple classes including Quality control samples. Great for Time series jobs or multiple knockouts.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

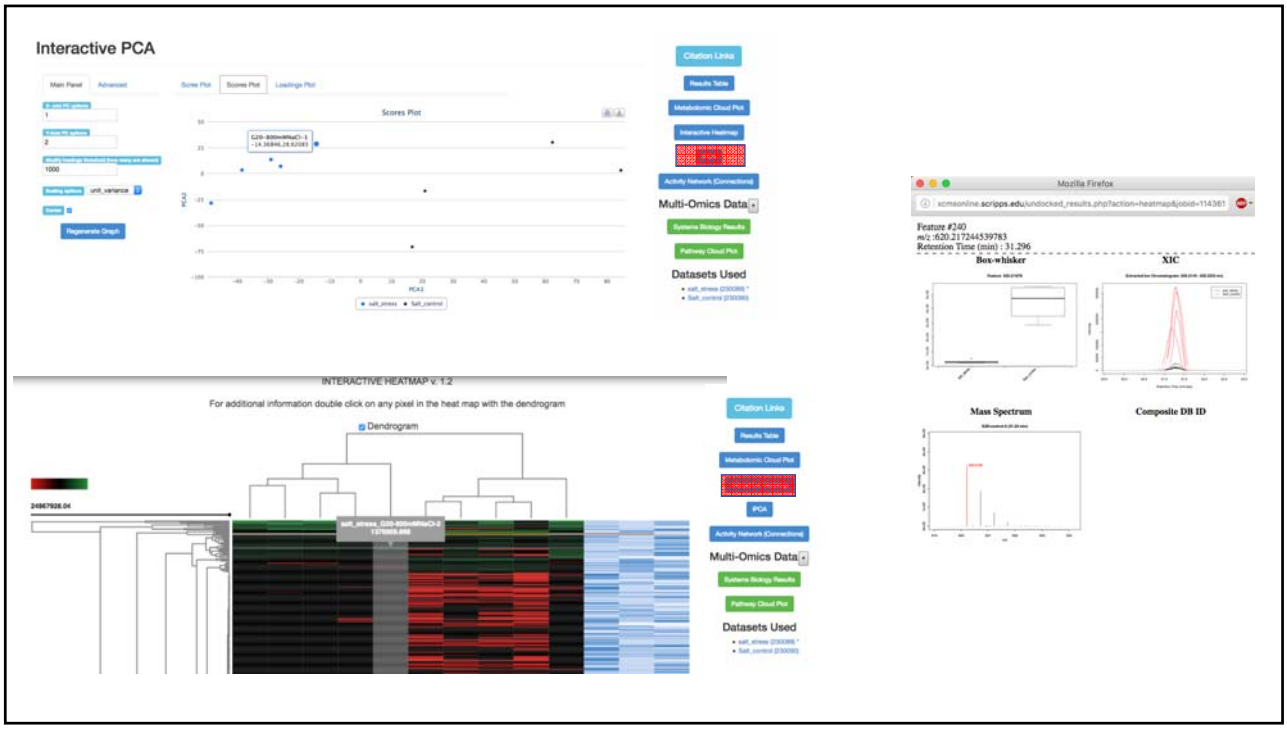


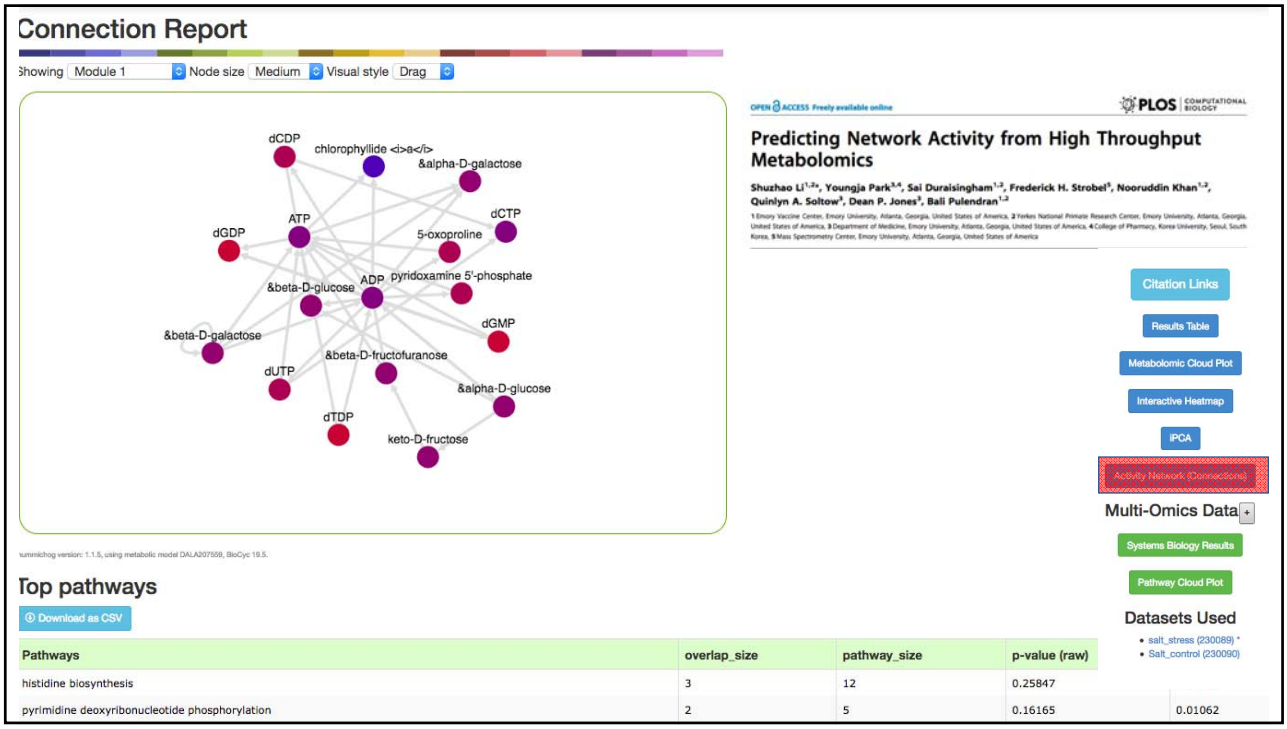
View Results section




Home
News
Databases
Create Job
View Results
XCMS Public
XCMS Institute
Stored Datasets
Account
Toolbox
Help
Quick Compound Search:
Search
Clear
Citation Links
Results Table
Metabolomic Cloud Plot
Interactive Heatmap
iPCA
Activity Network (Connections)
Multi-Omics Data
Systems Biology Results
Pathway Cloud Plot
Datasets Used
salt_stress (230089) *
Salt_control (230090)
Return to Job Summary


Original Settings
Add New View
Remove Current View
Toggle Help Box
Main Panel
Advanced
p-value
0
0-0.01
1
fold change
0
1.5 - 2807.731
2807.731
Regenerate Cloud Plot
EIC_ABBR
Box and Whisker
Spectrum
Bubble Size Multiplier
22
Download Cloud Plot Manual
Welcome to the Cloud Plot
Welcome to the Interactive
View 1
Retention Time vs m/z of 359 features
Click and drag in the plot area to zoom in
m/z : 940.571
retention time: 13.09 min
p-val : 0.00412
fold change : 28.641 (UP)
max intensity : 5265
0 matches found.
m/z
Retention Time (minutes)
Toggle Upregulated
Toggle Downregulated
p-Value : 0-0.01 | fold change : 1.5 - 2807.731 | m/z Range of 0 - 1099 | Retention Range of
Radius Scale : Fold Change
Table View
Citation Links
Results Table
Metabolomic Cloud Plot
Interactive Heatmap
iPCA
Activity Network (Connections)
Multi-Omics Data
Systems Biology Results
Pathway Cloud Plot
Datasets Used
salt_stress (230089) *
Salt_control (230090)


Interactive PCA
Scores Plot
Regenerate Graph
Citation Links
Results Table
Metabolomic Cloud Plot
Interactive Heatmap
Activity Network (Connections)
Multi-Omics Data
Systems Biology Results
Pathway Cloud Plot
Datasets Used
INTERACTIVE HEATMAP v. 1.2
For additional information double click on any pixel in the heat map with the dendrogram
Dendrogram
PCA
Mozilla Firefox
Box-whisker
XIC
Mass Spectrum
Composite DB ID


Connection Report
Showing Module 1 Node size Medium Visual style Drag
dCDP
chlorophyllide a
&alpha-D-galactose
dCTP
dGDP
ATP
5-oxoproline
&beta-D-glucose
ADP
pyridoxamine 5'-phosphate
dGMP
&beta-D-galactose
&beta-D-fructofuranose
dUTP
&alpha-D-glucose
dTDP
keto-D-fructose
Predicting Network Activity from High Throughput Metabolomics
Citation Links
Results Table
Metabolomic Cloud Plot
Interactive Heatmap
PCA
Activity Network (Connections)
Multi-Omics Data
Systems Biology Results
Pathway Cloud Plot
Datasets Used
salt_stress (230089)
Salt_control (230090)
Top pathways
Download as CSV
Pathways
overlap_size
pathway_size
p-value (raw)
histidine biosynthesis
3
12
0.25847
pyrimidine deoxyribonucleotide phosphorylation
2
5
0.16165
0.01062

Top metabolite predictions

Download as CSV

Compound_ID	name(input_mz)	evidence_score(match_form)	MW(mz_diff)
DCDP	dCDP	5	387.0233
	386.017400	M-H[-]	0.001400
2-D-THREO-HYDROXY-3-CARBOXY-ISOCAPROATE	(2*R*,3*S*)-3-isopropylmalate	9	176.0685
	196.036600	M+Na-2H[-]	0.000700
	175.061400	M-H[-]	0.000200
	235.082800	M+CH3COO[-]	0.001000
	157.051200	M-H2O-H[-]	0.000600
5-OXOPROLINE	5-oxoproline	8	129.0426
	149.009700	M+Na-2H[-]	-0.000300
	128.035900	M-H[-]	0.000600
	164.011900	M+Cl[-]	0.000400
	128.035800	M-H[-]	0.000500
CHLOROPHYLLIDE-A	chlorophyllide *a*	6	614.238
	613.230100	M-H[-]	-0.000600
	659.238500	M+HCOO[-]	0.002900
CPD-1130	3-ethylmalate	6	162.0528
	161.046300	M-H[-]	0.000800
	197.021200	M+Cl[-]	-0.000500
UDP-MANNACA	UDP-*N*-acetyl-α-D-mannosaminouronate	8	621.0608

- Adding MS/MS checks to confirm the predicted metabolites.
  - This will result in an increased/decreased score for the pathway

Systems Biology Analysis Results

Show 25 rows | VALUES <> PERCENT | TSV | PDF | Print | Search:

Pathway	Overlapping genes	All genes*	Overlapping proteins	All proteins*	Overlapping putative metabolites[1]	All metabolites[2]*	p-values
histidine biosynthesis	0	0	0	0	3	12	1.0e-2
pyrimidine deoxyribonucleotide phosphorylation	0	0	0	0	2	5	1.1e-2
UDP-*N*-acetyl-α-D-mannosaminouronate biosynthesis	0	0	0	0	2	5	1.1e-2
NAD salvage pathway I	0	0	0	0	2	5	1.1e-2
pyrimidine deoxyribonucleotides *de novo* biosynthesis III	0	0	0	0	3	13	1.3e-2
NAD biosynthesis from 2-amino-3-carboxymuconate semialdehyde	0	0	0	0	2	6	1.5e-2
ethanol degradation II	0	0	0	0	2	6	1.5e-2

Citation Links

Results Table

Metabolomic Cloud Plot

Interactive Heatmap

iPCA

Activity Network (Connections)

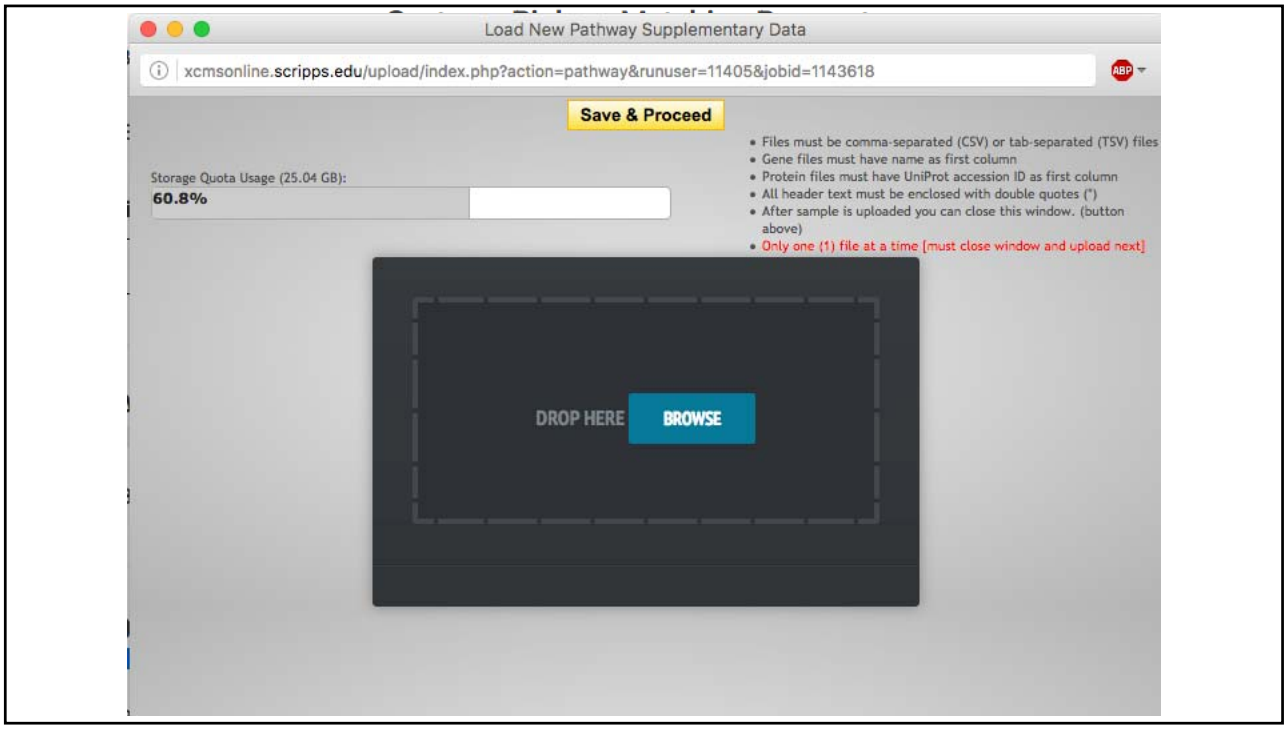
Multi-Omics Data

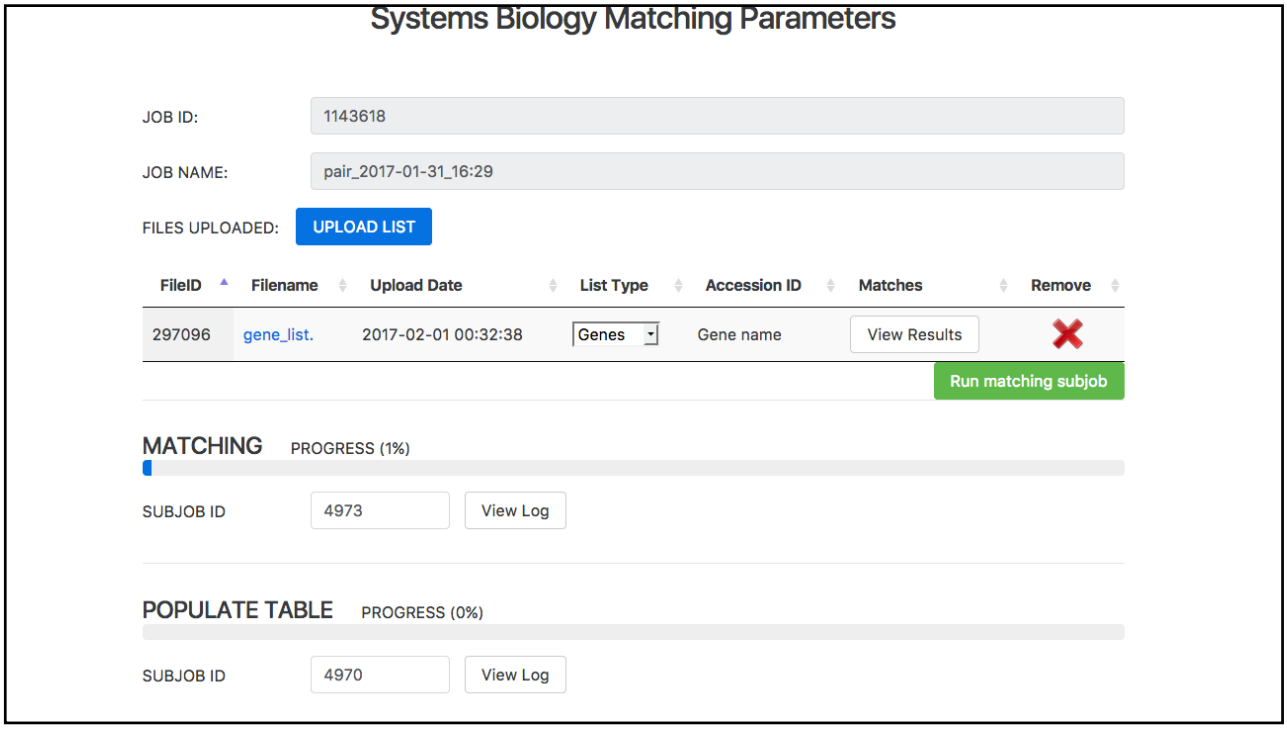
Systems Biology Results

Pathway Cloud Plot

Datasets Used

- salt_stress (230089) *
- Salt_control (230090)


Load New Pathway Supplementary Data
xcmsonline.scripps.edu/upload/index.php?action=pathway&runuser=11405&jobid=1143618
Save & Proceed
Storage Quota Usage (25.04 GB):
60.8%
Files must be comma-separated (CSV) or tab-separated (TSV) files
Gene files must have name as first column
Protein files must have UniProt accession ID as first column
All header text must be enclosed with double quotes (")
After sample is uploaded you can close this window. (button above)
Only one (1) file at a time [must close window and upload next]
DROP HERE
BROWSE

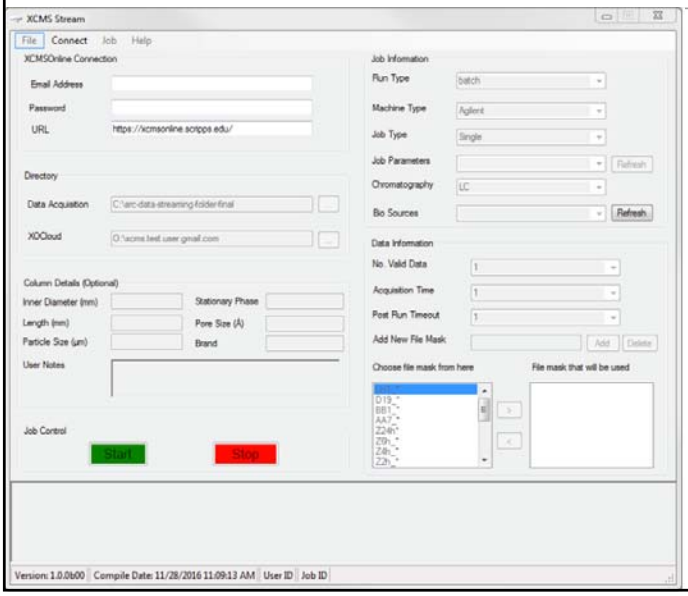

Systems Biology Matching Parameters
JOB ID:
1143618
JOB NAME:
pair_2017-01-31_16:29
FILES UPLOADED:
UPLOAD LIST
FileID
Filename
Upload Date
List Type
Accession ID
Matches
Remove
297096
gene_list.
2017-02-01 00:32:38
Genes
Gene name
View Results
Run matching subjob
MATCHING
PROGRESS (1%)
SUBJOB ID
4973
View Log
POPULATE TABLE
PROGRESS (0%)
SUBJOB ID
4970
View Log

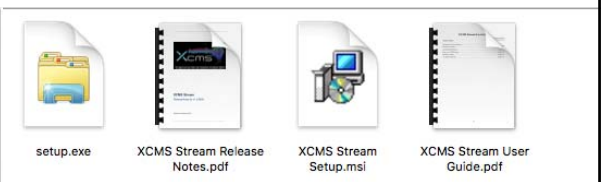
System Biology

Taking a step back

- We can automatically get to System level data quickly.
  - Where are the other dead time events?
- Sharing is caring : The future is sharing your data with others
- How can we make the data even better?

Steaming – upload to XCMS Online

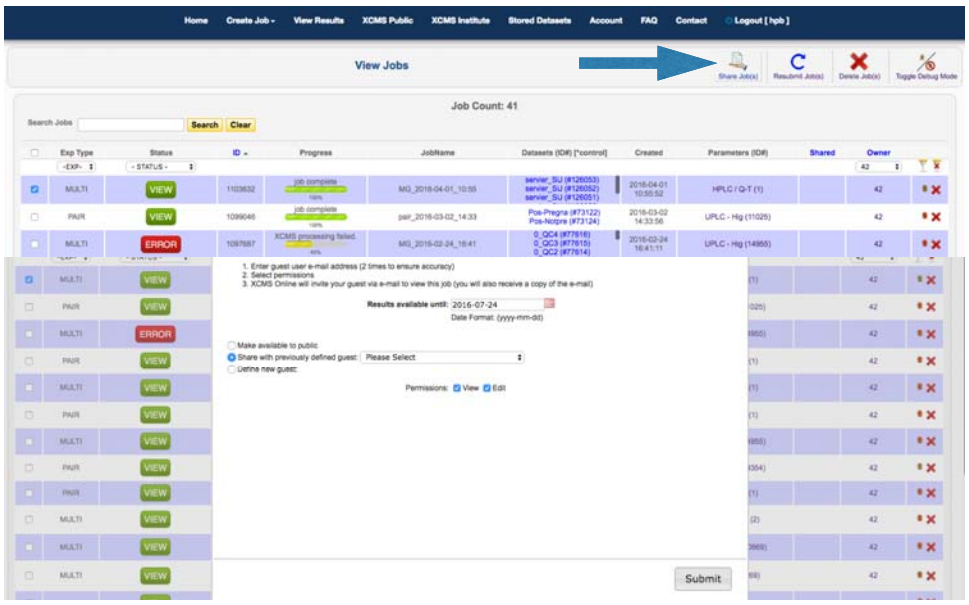




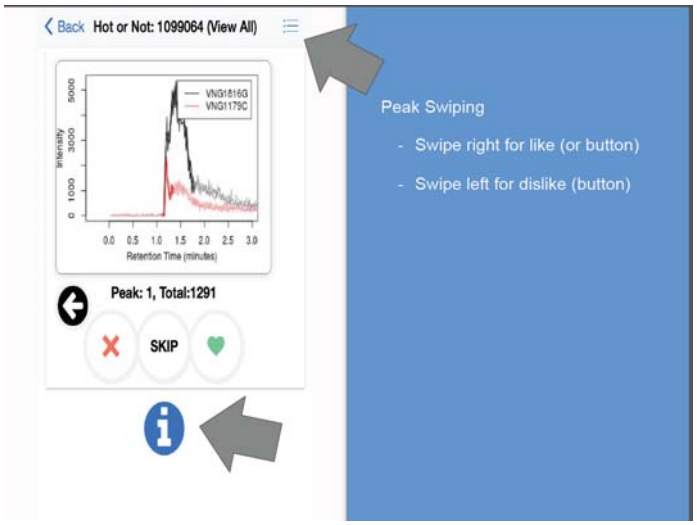
XCMS Online Stream uploads data and starts the jobs on XCMS Online automatically. This highly reduces the dead time between instrument acquisition and can also be used in offline mode.

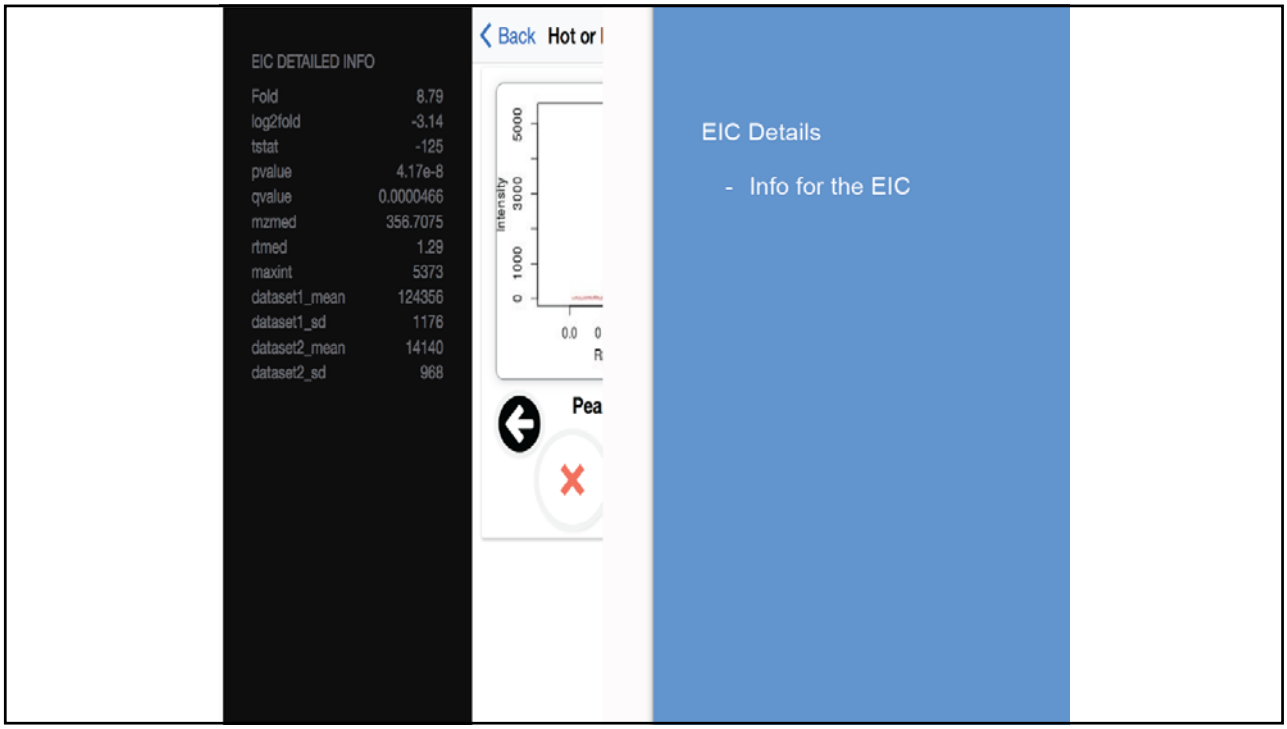
http://xcmsonline.scripps.edu/xcmsonline/XCMS_Stream-v1.0.0b00.zip

Job Sharing



EIC hot or not





XCMS Developments

- Speed: Faster jobs and XCMS Online that can handle multi 100Gb's data files.
- Sub-jobs : Allows for custom PCA, PLS-DA, Baysian Stats on selected features, custom normalisation per user
- System Biology workflow : Further developments including signaling networks
- Targeted MS studies using both full scan MS1 and triple Quad data (MRM).

Thank you ! Questions ?

Prof. Gary Siuzdak

The Scripps Research Institute

Aries Aisporna

Duane Rinehart