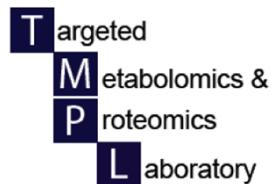




## Working with metabolomics data

Steve Barnes, PhD



## Examining output on XCMS

- First, we'll logon to XCMS and inspect the non-irradiated diet vs irradiated diet
- We'll download the XCMS output file
  - Make a Volcano plot
  - Prepare files for MetaboAnalyst



## Viewing the datasets

	PAIR	<b>VIEW</b>	1052098	job complete 100%	MCF10AT_NTVec vs KD1_NegMode	MCF10AT_NT (#89883) MCF10AT_KD (#89884)	2015-03-19 16:22:46	NanoLc Neg (10374)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1052085	job complete 100%	MCF10AT_NTVec vs KD1_PosMode	MCF10AT_NT (#89683) MCF10AT_KD (#89684)	2015-03-19 14:10:30	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051960	job complete 100%	PAIR_2015-03-18_17:24	Shan_Cuts (#15973) GSE_Cuts_#115883	2015-03-18 17:24:58	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051941	job complete 100%	SUM159VMP_VS_NF2c16_N egMode	SUM159_VMP (#89938) SUM159_NF2 (#89955)	2015-03-18 13:24:52	NanoLc Neg (10374)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051935	job complete 100%	SUM159VMP_VS_NF2c16_Po sMode	SUM159_VMP (#89688) SUM159_NF2 (#89689)	2015-03-18 12:07:33	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051415	job complete 100%	Grubbs_urine_pos_mmchg	Grubbs_Uri (#107301) Grubbs_Uri (#107315)	2015-03-12 19:11:09	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051379	job complete 100%	Aman_Set_3 Comparison	Aman_Set_3 (#115309) Aman_Set_3- (#115234)	2015-03-12 13:21:25	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051373	job complete 100%	Grubbs_urine_neg_mmchg	Grubbs_Uri (#107534) Grubbs_Uri (#107626)	2015-03-12 12:19:17	NanoLC5600 (10377)		X
□	PAIR	<b>VIEW</b>	1051324	job complete 100%	Grubbs_diet_neg_mmchg	Grubbs_Non (#108897) Grubbs_Irr (#108890)	2015-03-12 05:55:12	NanoLc Neg (10374)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1051271	job complete 100%	Grubbs_diet_pos_mmchg	Grubbs_Non (#109003) Grubbs_Irr (#108990)	2015-03-11 22:05:58	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1050983	job complete 100%	Sera Positive Mode	McLean_Har (#114812) McLean_Har (#114794)	2015-03-09 22:30:49	nanoLc_560 (9920)	Shared [Stop sharing]	X
□	PAIR	<b>VIEW</b>	1047227	job complete 100%	Grubbs_Diet_PosMode	Grubbs_Irr (#108990) Grubbs_Non (#109003)	2015-02-05 17:51:27	nanoLc_560 (9920)	Shared [Stop sharing]	X

## Viewing the data

[Download Results](#)

**Pairwise Results Summary: Grubbs\_diet\_neg\_mmchg (#1051324)**

hash: 835ca72929343f7e0d113ce314f7990

Submit Date	Finish Date	Paired Samples	Total Aligned Features	Parameter ID#	Log	Shared

[View Results Table](#)  
 [View Interactive Cloud Plot](#)  
 [View Interactive Heatmap](#)  
 [View PCA](#)  
 [Connections](#)

**Datasets Used**

- Grubbs\_NonIrrad (108897) \*
- Grubbs\_Irradiat (108890)

Total Ion Chromatograms (original)

TIC

Retention Time (minutes)

PNG PDF

Retention Time Deviation vs. Retention Time

Retention Time Deviation minutes

Retention Time (minutes)

PNG PDF

Total Ion Chromatograms (corrected)

TIC

Retention Time (minutes)

PNG PDF

Cloud Plot: 217 features with  $p\text{-value} \leq 0.01$ , fold change  $\geq 1.5$

TIC

Retention Time (minutes)

PNG PDF

Non-Metric Multidimensional Scaling

Dimension 2

Dimension 1

PNG PDF

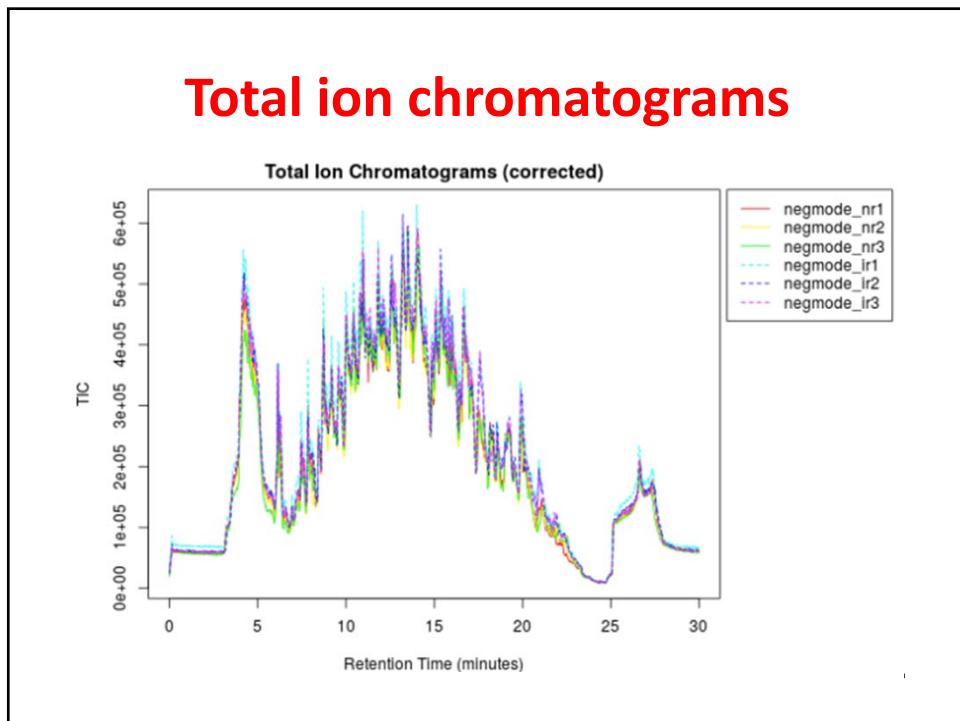
PCA Scores

Unit Variance scaling Centered

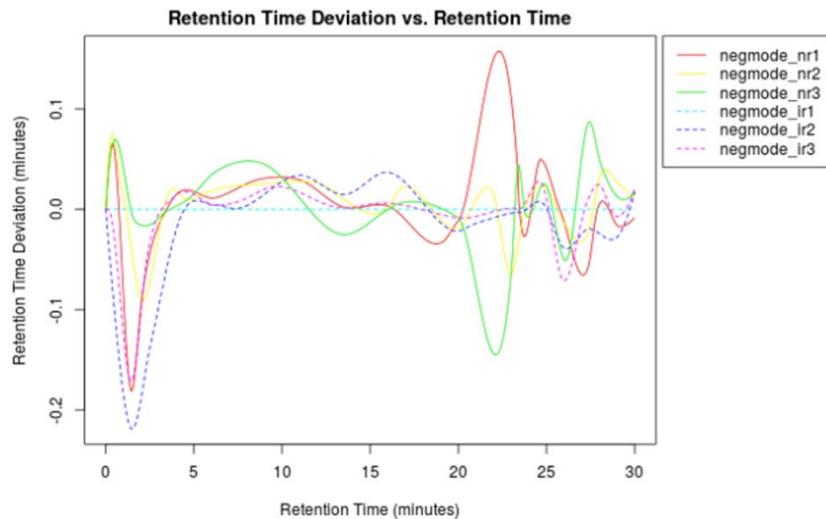
PC2 - Variance 15%

PC1 - Variance 49%

PNG PDF



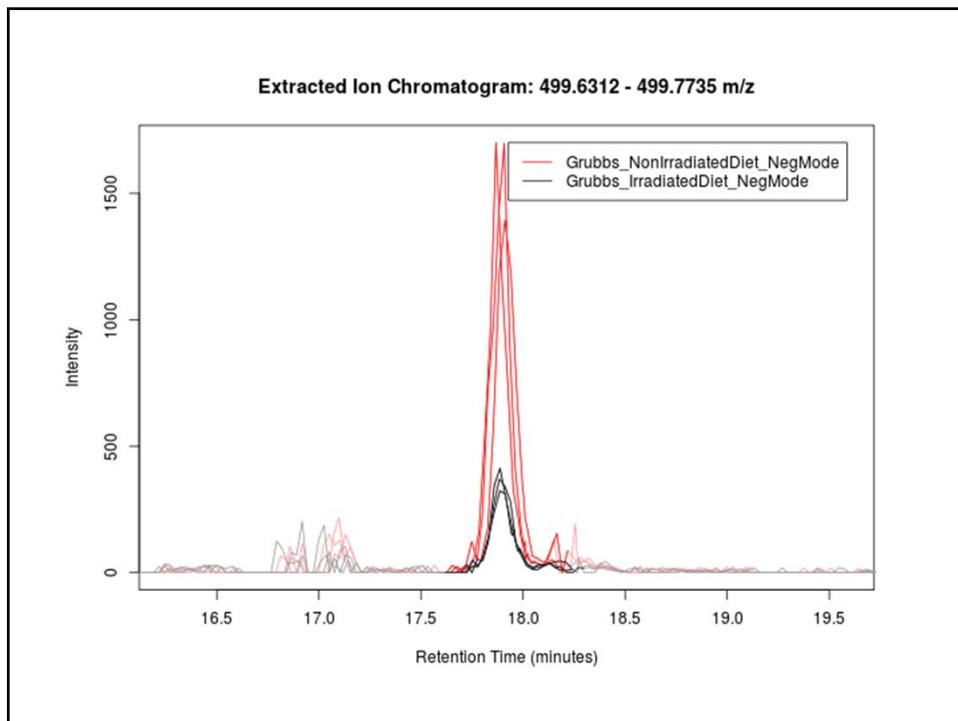
## High reproducibility of retention times

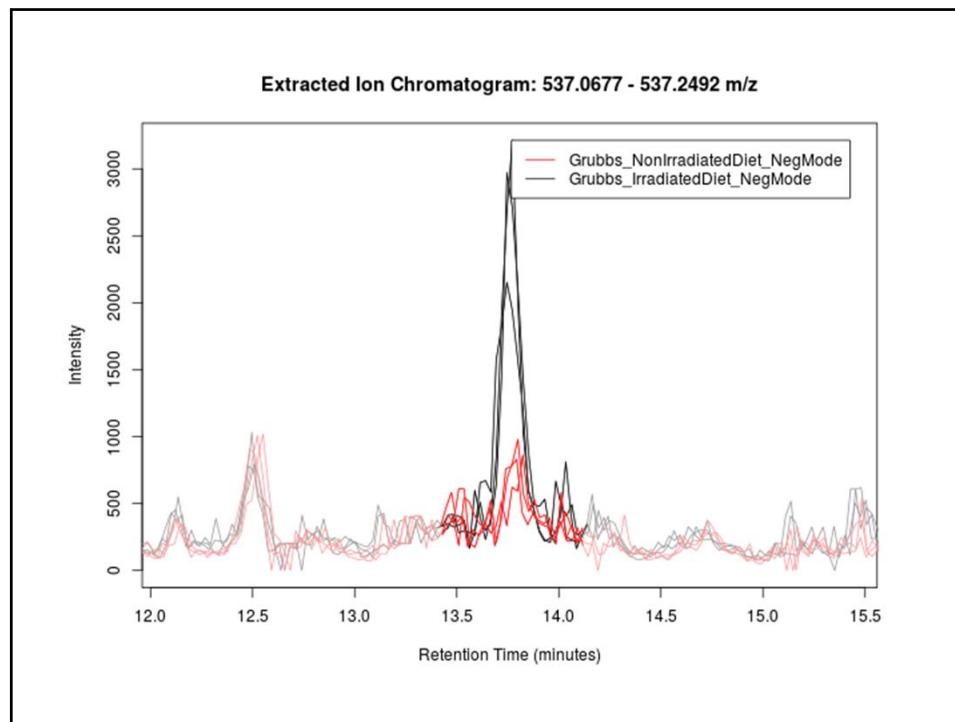


## Synopsis

- **Opening the .zip file**
- **Creating the Excel file from result.tsv file**
- **Why a Volcano plot?**
  - Making a volcano plot
  - Selecting ions that reach significance and fold-change cutoffs
- **Preparing for MetaboAnalyst**

►	boxplot	Mar 12, 2015, 6:19 AM	--	Folder
►	CloudPlot-svg.svg	Mar 12, 2015, 6:52 AM	968 KB	SVG document
►	CloudPlot.pdf	Mar 12, 2015, 6:52 AM	366 KB	PDF Document
►	CloudPlot.png	Mar 12, 2015, 6:52 AM	99 KB	PNG image
►	EIC	Mar 12, 2015, 6:21 AM	--	Folder
►	Heatmap_1051324.png	Mar 12, 2015, 6:22 AM	45 KB	PNG image
►	Heatmap_Cor_1051324.png	Mar 12, 2015, 6:51 AM	356 KB	PNG image
►	MDS.pdf	Mar 12, 2015, 6:51 AM	5 KB	PDF Document
►	MDS.png	Mar 12, 2015, 6:51 AM	15 KB	PNG image
►	mummichog	Mar 12, 2015, 6:24 AM	--	Folder
►	MVstats_ScalingPlot_1051324.pdf	Mar 12, 2015, 6:51 AM	105 KB	PDF Document
►	PCA-diagnostics.pdf	Mar 12, 2015, 6:51 AM	5 KB	PDF Document
►	PCA-diagnostics.png	Mar 12, 2015, 6:51 AM	5 KB	PNG image
►	PCA-loadings-all.pdf	Mar 12, 2015, 6:51 AM	35 KB	PDF Document
►	PCA-loadings-all.png	Mar 12, 2015, 6:51 AM	20 KB	PNG image
►	PCA.pdf	Mar 12, 2015, 6:51 AM	5 KB	PDF Document
►	PCA.png	Mar 12, 2015, 6:51 AM	18 KB	PNG image
►	result.tsv	Mar 12, 2015, 6:52 AM	1.4 MB	Plain Text
►	rtcor.pdf	Mar 12, 2015, 6:17 AM	61 KB	PDF Document
►	rtcor.png	Mar 12, 2015, 6:17 AM	40 KB	PNG image
►	TICs_rtcor.pdf	Mar 12, 2015, 6:17 AM	70 KB	PDF Document
►	TICs_rtcor.png	Mar 12, 2015, 6:17 AM	63 KB	PNG image
►	TICs.pdf	Mar 12, 2015, 6:13 AM	71 KB	PDF Document
►	TICs.png	Mar 12, 2015, 6:13 AM	62 KB	PNG image
►	XCMS.annotated.differ...iatedDiet_NegMode.tsv	Mar 12, 2015, 6:52 AM	1.5 MB	Plain Text
►	XCMS.diffreport..Grub...iatedDiet_NegMode.tsv	Mar 12, 2015, 6:21 AM	1.5 MB	Plain Text
►	XCMS.diffreport..Grub...iatedDiet_NegMode.xlsx	Mar 12, 2015, 6:22 AM	1.2 MB	Micros...(xlsx)
►	XCMSOnline_log.txt	Mar 12, 2015, 6:52 AM	2 KB	Plain Text





Double click  
on this file

►	boxplot	Mar 12, 2015, 6:19 AM	--	Folder
	CloudPlot.svg	Mar 12, 2015, 6:52 AM	968 KB	SVG document
	CloudPlot.pdf	Mar 12, 2015, 6:52 AM	366 KB	PDF Document
	CloudPlot.png	Mar 12, 2015, 6:52 AM	99 KB	PNG image
►	EIC	Mar 12, 2015, 6:21 AM	--	Folder
	Heatmap_1051324.png	Mar 12, 2015, 6:22 AM	45 KB	PNG image
	Heatmap_Cor_1051324.png	Mar 12, 2015, 6:51 AM	356 KB	PNG image
	MDS.pdf	Mar 12, 2015, 6:51 AM	5 KB	PDF Document
	MDS.png	Mar 12, 2015, 6:51 AM	15 KB	PNG image
►	mummichog	Mar 12, 2015, 6:24 AM	--	Folder
	MVstats_ScalingPlot_1051324.pdf	Mar 12, 2015, 6:51 AM	105 KB	PDF Document
	PCA-diagnostics.pdf	Mar 12, 2015, 6:51 AM	5 KB	PDF Document
	PCA-diagnostics.png	Mar 12, 2015, 6:51 AM	5 KB	PNG image
	PCA-loadings-all.pdf	Mar 12, 2015, 6:51 AM	35 KB	PDF Document
	PCA-loadings-all.png	Mar 12, 2015, 6:51 AM	20 KB	PNG image
	PCA.pdf	Mar 12, 2015, 6:51 AM	5 KB	PDF Document
	PCA.png	Mar 12, 2015, 6:51 AM	18 KB	PNG image
	result.tsv	Mar 12, 2015, 6:52 AM	1.4 MB	Plain Text
	rtcor.pdf	Mar 12, 2015, 6:17 AM	61 KB	PDF Document
	rtcor.png	Mar 12, 2015, 6:17 AM	40 KB	PNG image
	TICs_rtcor.pdf	Mar 12, 2015, 6:17 AM	70 KB	PDF Document
	TICs_rtcor.png	Mar 12, 2015, 6:17 AM	63 KB	PNG image
	TICs.pdf	Mar 12, 2015, 6:13 AM	71 KB	PDF Document
	TICs.png	Mar 12, 2015, 6:13 AM	62 KB	PNG image
	XCMS_annotated.differenceDiet_NegMode.tsv	Mar 12, 2015, 6:52 AM	1.5 MB	Plain Text
	XCMS_diffreport_Grubbs_diet_NegMode.tsv	Mar 12, 2015, 6:21 AM	1.5 MB	Plain Text
	XCMS_diffreport_Grubbs_diet_NegMode.xlsx	Mar 12, 2015, 6:22 AM	1.2 MB	Micros...(.xlsx)
	XCMSonline_log.txt	Mar 12, 2015, 6:52 AM	2 KB	Plain Text

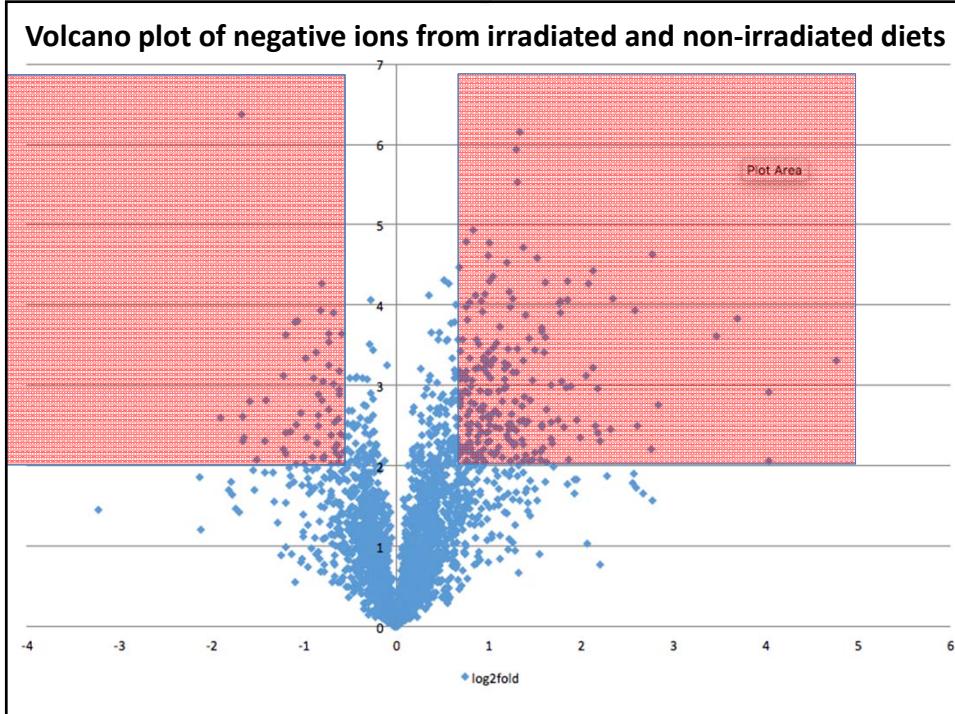




## Making a volcano plot

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
			name	fold	log2fold	-log P-value	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
1	M500T18_1	3.17718625	-1.6677497	6.37194281	-69.935309	4.2468E-07	0.00059116	DOWN	499.71988	499.714381	499.721035	17.9395	17.8955	17.9438333		
96	M165T15	1.66018323	-0.7313425	3.24367484	-36.133032	0.00057059	0.00994236	DOWN	165.055502	165.054308	165.056684	14.6846667	14.6478333	14.71805		
47	M574T18	1.23905717	-1.096975	3.78771429	-29.64074	0.00016304	0.00369183	DOWN	574.266065	574.263028	574.272462	17.7115	17.6941667	17.7368333		
58	M517T15_2	1.75943722	-0.815114	3.93600909	-27.522969	0.00011588	0.00501453	DOWN	517.194812	517.191996	517.198037	14.6337667	14.59155	14.6643167		
42	M574T17_2	1.59552933	-0.6704351	3.89904915	-23.354148	0.00012617	0.00502501	DOWN	573.751404	573.745446	573.752077	17.0508333	17.04955	17.1253333		
69	M567T12_1	1.21578562	-0.2818889	3.50811164	-22.910965	0.00031038	0.00742882	DOWN	567.214555	567.212376	567.21642	12.4962667	12.4584667	12.5235833		
206	M504T10	2.04744309	-1.0338234	2.65836288	-18.639823	0.00219602	0.01783223	DOWN	504.138121	504.136349	504.14049	10.1241	10.1203667	10.1441333		
21	M175T15_2	1.75024009	-0.8075528	4.26057988	-18.65345	5.4881105	0.00417379	DOWN	175.097496	175.09698	175.098239	14.8694333	14.866333	14.8797		
90	M529T12	1.20815162	-0.2728015	4.06091097	-18.56302	8.6914-05	0.0046548	DOWN	529.139978	529.191284	529.195264	11.8747	11.861933	11.9029833		
177	M231T16_2	3.00517465	-1.5874488	2.79422291	-16.956079	0.00160612	0.01517888	DOWN	231.15918	231.158231	231.16084	15.6376667	15.5836667	15.63795		
46	M493T18_2	2.09585981	-1.0675422	3.80434853	-16.138394	0.00015691	0.00569183	DOWN	493.237441	493.235949	493.237575	18.1235	18.1161667	18.1653333		
56	M287T12	1.49870741	-0.5837188	3.63982195	-14.759869	0.00022918	0.00684582	DOWN	287.079116	287.0773836	287.083574	12.4962667	12.40445	12.55065		
124	M565T12_1	1.40852493	-0.4941851	3.08759847	-14.151015	0.00081734	0.01094051	DOWN	565.215654	565.215406	565.217814	12.4962667	12.4853	12.5235833		
67	M175T15_1	1.653289893	-0.7253476	3.54719875	-14.110635	0.00028386	0.0070821	DOWN	175.076333	175.075715	175.077592	14.8742	14.8672667	14.8829167		
57	M490T14_1	2.28133744	-1.1898799	3.62228199	-13.866683	0.00023863	0.00700292	DOWN	490.13636	490.134832	490.13757	14.1682333	14.1656833	14.1689833		
170	M490T14_2	2.64201512	-1.4016387	2.82058063	-13.640201	0.00151154	0.01487326	DOWN	490.150671	490.14958	490.151411	14.2480333	14.2477833	14.26975		
55	M187T17	1.65717691	-0.7287276	3.64858184	-13.139663	0.0002246	0.00683111	DOWN	187.097276	187.096754	187.098391	17.455833	17.4535	17.4841667		
327	M587T12	1.21922912	-0.2859693	2.27459764	-12.965168	0.00531377	0.02715163	DOWN	587.234503	587.233346	587.23698	12.2213833	12.1999667	12.2525333		
244	M289T11	1.803374	-0.8506986	2.28411244	-12.840128	0.00519861	0.02682321	DOWN	289.082598	289.076863	289.086635	11.1201167	11.10205	11.1440333		
74	M517T11	1.18992599	-0.2508718	3.44500584	-12.785979	0.00035892	0.00808092	DOWN	317.082663	317.080868	317.083426	11.208525	11.1920667	11.2338		

Highlight columns D and E and create a chart



## Selecting cut-offs

- We should now sort the data in the Excel file by p-values (from smallest to largest)
- Copy the metabolite ions that have p-values <0.01 into a new sheet
- Now resort these data by fold-change
  - these can be increased (>1.5) or decreased (<0.667)

## Selecting all ions with p < 0.01 and making a new sheet

1	M500T18_1	3.17718625	-1.6677497	6.37194281	-69.935309	4.2468E-07	0.00059116 DOWN	499.71989	499.714381	499.721035	17.9395	17.8955	17.9438333
2	M251T15	2.5381022	1.34375016	6.15069974	57.2949952	7.0681E-07	0.00059116 UP	251.004448	251.00233	251.006679	14.646175	14.6239333	14.6608333
3	M417T18	2.45768759	1.29730154	5.93242839	55.093938	1.6683E-07	0.00065146 UP	417.212131	417.209144	417.213252	17.8296667	17.8865	
4	M199T19	2.49373589	1.31830866	5.53403200	40.9657391	2.9239E-07	0.00122277 UP	199.133402	199.132069	199.134536	18.99725	18.8758333	19.0271667
5	M537T14_2	1.77873662	0.83085291	4.93437072	31.7221237	1.1631E-05	0.00388536 UP	537.125737	537.125207	537.126264	13.7591333	13.7458333	13.7724333
6	M325T11	1.69519907	0.7614547	4.79236518	49.0738693	1.613E-05	0.00388536 UP	325.093116	325.092202	325.095083	11.2136667	11.2073167	11.25245
7	M144T16	2.01229805	1.00884401	4.77150826	33.2039583	1.6924E-05	0.00388536 UP	144.045677	144.045279	144.046454	16.2021167	16.1911833	16.2194333
8	M357T13_2	2.59414359	1.37525834	4.7177729	32.8800228	1.9153E-05	0.00388536 UP	357.082227	357.081993	357.083674	12.6645667	12.6562167	12.6886667
9	M426T8	6.84146926	2.77430619	4.6257974	22.7140999	2.367E-05	0.00388536 UP	426.03226	426.032122	426.034541	8.13925	8.13335	8.15796667
10	M347T13	1.99584111	0.99699687	4.61237735	22.1908286	2.4413E-05	0.00388536 UP	347.167955	347.163411	347.172006	13.059825	13.0026333	13.115325



480	M611T16_2	1.17035124	-0.2269416	2.00908774	5.360212	0.00979292	0.03401502 DOWN	611.278173	611.27616	611.284794	15.8765083	15.8224333	15.9544333
482	M711T18	1.26827471	-0.3428673	2.00794425	-7.6797212	0.00981874	0.03401502 DOWN	711.31911	711.31715	711.32138	18.05925	17.9893333	18.1333333
483	M576T17_2	1.13201205	0.17889832	2.00693443	5.45812005	0.0098416	0.03401502 UP	576.245602	576.244742	576.249542	17.081	16.9721667	17.2375
484	M115T16	1.41760411	0.5034547	2.00604191	4.72201644	0.00986184	0.03401502 UP	115.004417	115.004193	115.006234	15.5579167	15.5327833	15.56165
485	M342T9	1.47951495	0.56512427	2.00572431	6.37522916	0.00986906	0.03401502 UP	342.090705	342.087875	342.09202	9.3306	9.3087333	9.37555
486	M197T10	1.26666403	0.34103592	2.00307979	6.82066584	0.00992933	0.03401502 UP	197.025861	197.023886	197.027836	9.5619667	9.5571167	9.56681667
487	M201T18	1.31936279	0.39984131	2.00283674	4.70021934	0.00993486	0.03401502 UP	201.105998	201.103323	201.107267	17.6158333	17.5908333	17.6258333
488	M162T11	1.09119523	0.12590925	2.0027729	6.75086139	0.00993633	0.03401502 UP	162.053951	162.052952	162.054783	11.4542	11.4454167	11.4963667
489	M131T14_1	1.31291812	-0.3927769	2.00218947	-5.5314033	0.00994971	0.03401502 DOWN	131.050967	131.048953	131.056393	13.7197	13.6934667	13.7197667
490	M687T15_2	1.30430893	0.38328562	2.001569	5.3380585	0.0099639	0.03401502 UP	686.755271	686.753349	686.757192	15.1120417	15.1079667	15.1161167



## Selecting those with fold-change decrease >1.5 and new page

name	fold	log2fold	-log P-value	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
M377T18	3.72924023	-1.8988817	2.59661412	-9.493137	0.00253155	0.01916149	DOWN	377.280273	377.279809	377.280485	18.2675	18.2525	18.299
M500T18_1	3.17718625	-1.6677497	6.37194281	-69.935309	4.2468e-07	0.00059116	DOWN	499.71989	499.714381	499.721035	17.9395	17.8955	17.9438333
M424T18	3.15396982	-1.6571689	2.30821987	-8.465829	0.0049179	0.02644019	DOWN	424.286919	424.285903	424.288206	18.2798333	18.2675	18.299
M493T18_1	3.14854207	-1.6546839	2.6148464	-12.339822	0.00242747	0.01871241	DOWN	492.732075	492.731096	492.733053	17.9241667	17.908	17.9403333
M251T17	3.12450783	-1.643629	2.34919635	-11.996006	0.00447511	0.02464783	DOWN	251.059425	251.059206	251.060194	16.6583333	16.6248	16.6707833
M231T16_2	3.00517465	-1.5874488	2.79422291	-16.956079	0.00160612	0.01517888	DOWN	231.15918	231.158231	231.160844	15.6376667	15.5836667	15.63795
M768T12	2.84825797	-1.5100798	2.07827637	-5.0543868	0.00835071	0.03256133	DOWN	767.829662	767.826195	767.833128	11.6209083	11.60575	11.6360667
M525T14_2	2.6848561	-1.4248448	2.30517791	-6.0601844	0.00495247	0.02646754	DOWN	525.284116	525.27453	525.287438	13.6901417	13.6671667	13.7188
M490T14_2	2.64201512	-1.4016387	2.82058063	-13.640201	0.00151154	0.01487326	DOWN	490.150671	490.14958	490.151411	14.2480333	14.2477833	14.26975



name	fold	log2fold	-log P-value	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
M419T11_1	1.17810989	-0.2364741	2.41792855	-7.9513274	0.00382007	0.02265429	DOWN	419.035272	419.033926	419.03625	11.3799333	11.3681333	11.3881667
M611T16_2	1.17035124	-0.2269416	2.00908774	-3.3602012	0.00979292	0.03401502	DOWN	611.278173	611.27616	611.284794	15.8765083	15.8224333	15.9544333
M565T15	1.16533313	-0.2219799	2.23155622	-6.7354962	0.00586737	0.02853127	DOWN	565.120094	565.119191	565.121328	14.6141	14.59155	14.6643167
M535T14	1.16407717	-0.2191867	2.26682334	-6.1061829	0.00540975	0.02737525	DOWN	535.110263	535.106754	535.111276	13.5184917	13.4834667	13.5622667
M221T19	1.15855517	-0.2123267	2.15628902	-5.6641662	0.00697768	0.03031699	DOWN	221.084846	221.08127	221.085397	19.0063333	18.9436667	19.0306667
M119T14	1.1541478	-0.206828	2.25839361	-9.9584077	0.00551577	0.02758935	DOWN	119.03942	119.033334	119.042616	14.0117333	14.0084	14.0381833
M621T15	1.14909666	-0.2005002	2.13061062	-5.1113482	0.00740269	0.03095745	DOWN	621.206427	621.205274	621.20758	15.193658	15.1884833	15.1988333
M607T16	1.11674864	-0.1593045	2.01321577	-5.863233	0.00970028	0.03401502	DOWN	607.130977	607.129251	607.131787	15.59016833	15.87635	15.9026667
M517T12_1	1.08641081	-0.1195697	2.06759187	-5.3916682	0.00855871	0.03280741	DOWN	517.213031	517.208377	517.213509	12.1466167	12.1357167	12.1986333
M451T16	1.06454484	-0.0902367	3.25022678	-10.111341	0.00056205	0.00994236	DOWN	451.211411	451.208912	451.217074	16.034225	16.0070167	16.05845

Preparing for analysis using  
MetaboAnalyst

<http://www.metaboanalyst.ca>



## Areas of aligned metabolites by sample

negmode_nr1	negmode_nr2	negmode_nr3	negmode_ir1	negmode_ir2	negmode_ir3
3320701	3211285	3388235	3247804	3308733	3156426
1650544	1652507	1785534	1908024	1746573	1767097
1502083	1537549	1456609	1602389	1497667	1546834
1348351	1396739	1399435	1538662	1387977	1475049
942173	839275	942798	1519979	1335058	1409076
1476742	1443824	1417457	1401542	1343338	1348272
1389740	1326330	1311725	1399942	1301020	1295945
1395260	1463727	1132647	1307379	1276028	1309956
469492	464640	466149	1044107	964298	945524
963917	957988	927037	903994	904583	916594
846739	810984	854938	919809	799096	822651
751388	734458	704322	833628	740626	724893
757382	706826	785876	744625	749518	742854
685156	711925	691722	750467	674733	675713
804585	760221	750478	726812	609044	647972
717398	709880	739525	670560	613870	640035
513949	585959	582984	543410	498409	586219
22783	17915	17732	557699	528582	514971
501645	507853	464691	515378	475918	482951
477094	482742	490705	495190	477892	475722

non-irradiated diet

irradiated diet

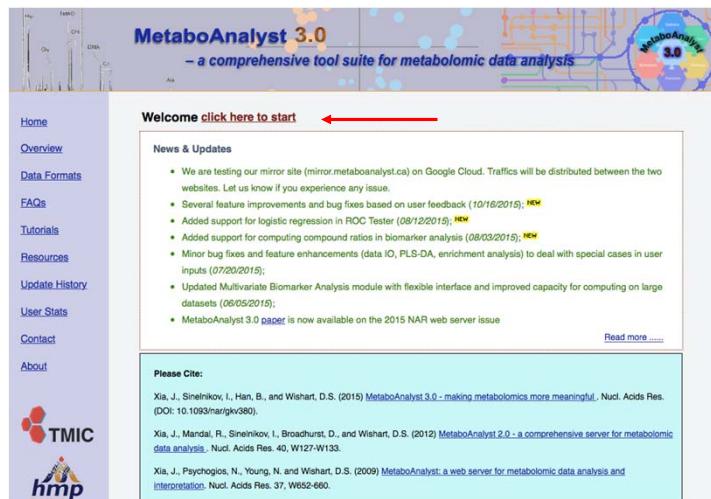
## Creating .csv files for each sample

1	mzmed	rtmed	negmode_nr1
2	499.71989	17.9395	9846
3	251.004448	14.646175	28534
4	417.212131	17.8596667	6340
5	199.133402	18.99725	18534
6	537.125737	13.7591333	17847
7	325.093116	11.2136667	23162
8	144.045677	16.2021167	26142
9	357.082227	12.6645667	12999
10	426.03226	8.13925	893
11	347.167955	13.059825	10415
12	591.136922	14.92045	26263
13	629.783983	16.3271833	3051
14	438.171148	12.3017917	15492
15	425.045022	8.13335	3088
16	151.039947	14.5387583	7302
17	506.109327	15.2182417	23595
18	789.227085	14.9668167	10301
19	315.142993	16.3572333	6312
20	640.774334	16.3572333	3677
21	582.278039	15.5049667	13805

- Copy the median *m/z* and median Rt values into a new Excel file. Then copy the column of areas from the first sample in Group\_1. Save as an Excel .csv file.
  - Note that the file name must not have spaces – use an underscore instead of a space.
- Leave the file open and replace the yellow column with the areas from the next Group\_1 sample. Save as a second .csv file.
- Continue until all Group\_1 and Group\_2 samples have a corresponding .csv file.

## Preparing a .zip file

- Put each of the .csv files for group\_1 samples into a folder named “Group\_1”.
- Put each of the .csv files for group\_2 samples into a folder named “Group\_2”.
- Click on Group\_1 and Group\_2 folders and combine to form a .zip file.
  - Rename the .zip file as [your\_name].zip
- You’re now ready to submit it to MetaboAnalyst
- <http://www.metaboanalyst.ca>



The screenshot shows the MetaboAnalyst 3.0 homepage. At the top, there is a banner with chemical structures and the text "MetaboAnalyst 3.0 - a comprehensive tool suite for metabolomic data analysis". Below the banner, a sidebar on the left contains links to Home, Overview, Data Formats, FAQs, Tutorials, Resources, Update History, User Stats, Contact, and About. The TMIC logo is also present. The main content area is titled "Please choose a functional module to proceed:" and lists four options:

- Statistical Analysis** (selected): This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA and PLS-DA. It also provides clustering and visualization tools to create dendograms and heatmaps as well as to classify based on random forests and SVM.
- Enrichment Analysis**: This module performs metabolite set enrichment analysis (MSEA) for human and mammalian species based on several libraries containing ~6300 groups of metabolite sets. Users can upload either 1) a list of compounds, 2) a list of compounds with concentrations, or 3) a concentration table.
- Pathway Analysis**: This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 21 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, S. cerevisiae, E.coli, and others, with a total of ~1600 metabolic pathways.
- Time Series Analysis**: This module supports temporal and two-factor data analysis including data overview, two-way ANOVA, and empirical Bayes time-series analysis for detecting distinctive temporal profiles. It also supports ANOVA-simultaneous component analysis (ASCA) to identify major patterns associated with each experimental factor.

The screenshot shows the "1) Upload your data" step of the MetaboAnalyst workflow. It has two main sections:

- Tab-delimited text (.txt) or comma-separated values (.csv) file:**
  - Data Type:  Concentrations  Spectral bins  Peak intensity table
  - Format: Samples in rows (unpaired)
  - Data File: Choose File No file chosen
  - Submit button
- Zipped Files (.zip) :**
  - Data Type:  NMR peak list  MS peak list  MS spectra
  - Data File: Choose File Diet\_negmode.zip
  - Pair File: Choose File No file chosen
  - Submit button

A red arrow points to the "MS peak list" radio button in the "Zipped Files" section. Another red arrow points to the ".zip" file "Diet\_negmode.zip" in the same section.

**Select MS peak list option and then load the .zip file**



**Processing MS peak list data :**

Peaks need to be matched across samples in order to be compared. For two-column format (mass and intensities), peaks are grouped by their m/z values. For three column data (mass, retention time, and intensities), the program will further group peaks based on their retention time. Users need to supply tolerance values in order to proceed. Here are some suggested values: mass tolerance - 0.25 (m/z); retention time - 30 (seconds) for LC-MS peak, and 5 (seconds) for GC-MS peaks. Please note, If a sample has more than one peak in a group, they will be replaced by their sum; some groups will be excluded if none of the classes has at least half its samples represented. Finally, the program create a peak intensity table in which each sample occupies a row and each column represents a peak group identified by the median values of its position (m/z and/or retention time).

Mass tolerance (m/z):	<input type="text" value="0.001"/>
Retention time tolerance:	<input type="text" value=".01"/>

**Submit**

**MS peak processing information**

The uploaded files are peak lists and intensities data.

A total of 6 samples were found.

These samples contain a total of 18522 peaks.

with an average of 3087 peaks per sample.

A total of 3087 peak groups were formed.

Peaks of the same group were summed if they are from one sample.

Peaks appear in less than half of samples in each group were ignored.

**Next**

**Data Integrity Check:**

1. Checking the class labels - at least three replicates are required in each class.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. The presence of missing values or features with constant values (i.e. all zeros)

**Data processing information:**

Checking data content ...passed

The uploaded files are peak lists and intensities data.

A total of 6 samples were found.

These samples contain a total of 18522 peaks.

with an average of 3087 peaks per sample

2 groups were detected in samples.

Samples are not paired.

All data values are numeric.

A total of 0 (0%) missing values were detected.

By default, these values will be replaced by a small value.

Click **Skip** button if you accept the default practice

Or click **Missing value imputation** to use other methods

[Missing value estimation](#)    [Skip](#)

**Note that XCMSonline filled in peaks**

Non-informative variables can be characterized in two groups: variables of very small values - these variables can be detected using mean or median; variables that are near-constant throughout the experiment conditions - these variables can be detected using standard deviation (SD); or the robust estimate such as interquartile range (IQR). The relative standard deviation(RSD = SD/mean) is another useful variance measure independent of the mean. The following empirical rules are applied during data filtering:

- **Less than 250 variables:** 5% will be filtered;
- **Between 250 - 500 variables:** 10% will be filtered;
- **Between 500 - 1000 variables:** 25% will be filtered;
- **Over 1000 variables:** 40% will be filtered;

Please note, in order to reduce the computational burden to the server, the **None** option is only for less than 2000 features. Over that, if you choose **None**, the IQR filter will still be applied. In addition, the maximum allowed number of variables is 5000. If over 5000 variables were left after filtering, only the top 5000 will be used in the subsequent analysis.

Interquartile range (IQR)  
 Standard deviation (SD)  
 Median absolute deviation (MAD)  
 Relative standard deviation (RSD = SD/mean)  
 Non-parametric relative standard deviation (MAD/median)  
 Mean intensity value  
 Median intensity value  
 None (less than 2000 features)

[Process](#)

