THE UNIVERSITY OF ALABAMA AT BIRMINGHAM
Knowledge that will change your world

Working with metabolomics data

Stephen Barnes, PhD
Professor of Pharmacology & Toxicology
sbarnes@uab.edu

Targeted Metabolomics & Proteomics Laboratory

Examining output on XCMS

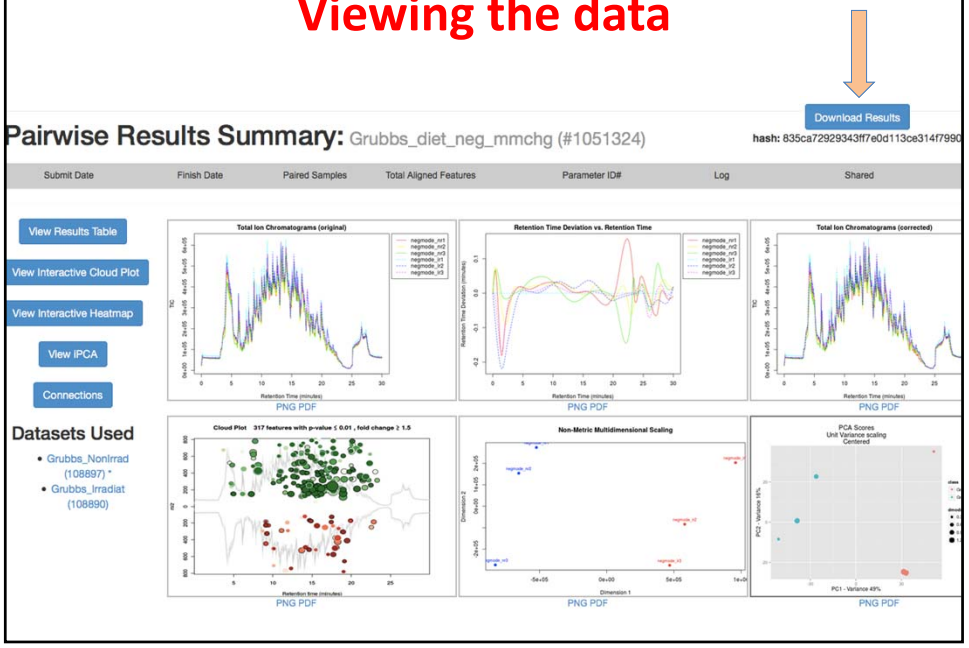
- We'll logon to XCMS and inspect the non-irradiated diet vs irradiated diet data
- Then we'll download the XCMS output file
  - Prepare files for MetaboAnalyst

Logon to XCMS with your name and password

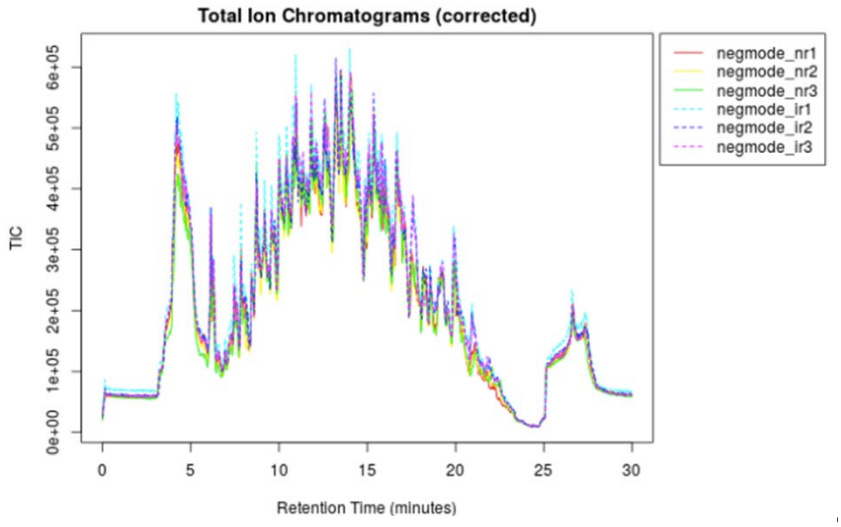
Viewing the datasets

	PAIR	VIEW	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]	✖
	PAIR	VIEW	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]	✖
	PAIR	VIEW	1051960	job complete 100%	PAIR_2015-03-18_17:24	Sham_Cutts (#115973) GSE_Cutts_ (#115883)	2015-03-18 17:24:58	nanoLC_560 (9920)			✖
	PAIR	VIEW	1051941	job complete 100%	SUM159VMP_VS_NF2c16_NegMode	SUM159_VMP (#89938) SUM159_NF2 (#89955)	2015-03-18 13:24:52	NanoLc Neg (10374)		Shared [Stop sharing]	✖
	PAIR	VIEW	1051935	job complete 100%	SUM159VMP_VS_NF2c16_PosMode	SUM159_VMP (#89688) SUM159_NF2 (#89689)	2015-03-18 12:07:33	nanoLC_560 (9920)		Shared [Stop sharing]	✖
	PAIR	VIEW	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]	✖
	PAIR	VIEW	1051379	job complete 100%	Aman Set 3 Comparison	Aman_Set3_ (#115309) Aman_Set3- (#115234)	2015-03-12 13:21:25	nanoLC_560 (9920)		Shared [Stop sharing]	✖
	PAIR	VIEW	1051373	job complete 100%	Grubbs_urine_neg_mmchg	Grubbs_Uri (#107534) Grubbs_Uri (#107626)	2015-03-12 12:19:17	NanoLC5600 (10377)			✖
	PAIR	VIEW	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]	✖
	PAIR	VIEW	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]	✖
	PAIR	VIEW	1050983	job complete 100%	Sera Positive Mode	McLean_Har (#114812) McLean_Har (#114794)	2015-03-09 22:30:49	nanoLC_560 (9920)			✖
	PAIR	VIEW	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]	✖

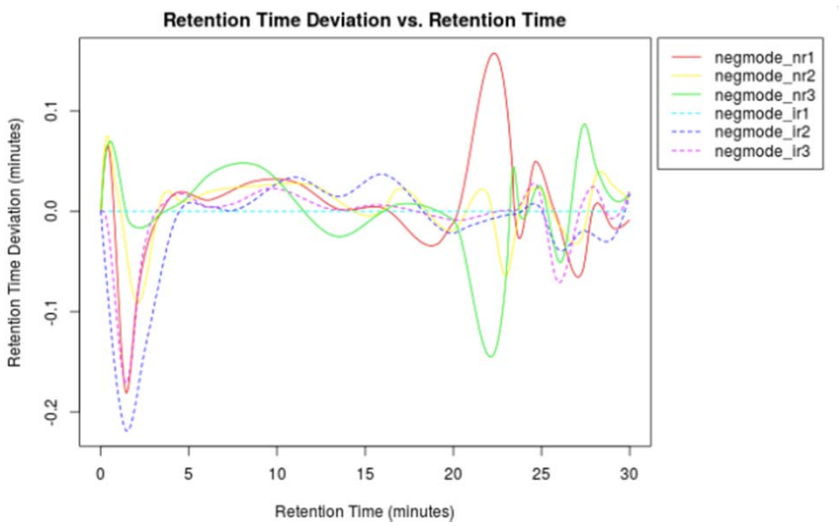
Viewing the data



Total ion chromatograms

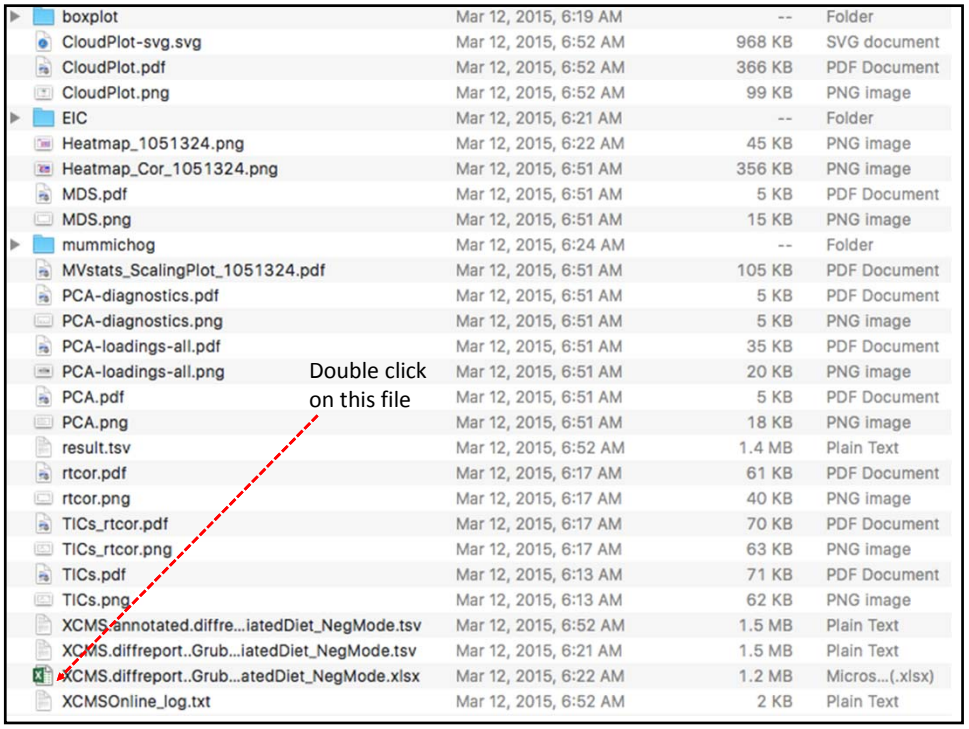


High reproducibility of retention times



Next steps

- Opening the .zip file you've downloaded (or the folder on your thumbdrive)
- Creating the Excel file from the XCMSdiffreport.xlsx file
- 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.diffre...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...atedDiet_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

The Excel DiffReport from XCMS

	name	fold	log2fold	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
1	M500T18_	3.177186	-1.66775	-69.9353	4.25E-07	0.000591	DOWN	499.7199	499.7144	499.721	17.9395	17.8955	17.94383
2	M251T15	2.538102	1.34375	57.295	7.07E-07	0.000591	UP	251.0044	251.0023	251.0067	14.64618	14.62393	14.66083
3	M417T18	2.457688	1.297302	55.09394	1.17E-06	0.000651	UP	417.2121	417.2091	417.2133	17.85967	17.82967	17.8865
4	M199T19	2.493736	1.318309	40.96574	2.92E-06	0.001223	UP	199.1334	199.1321	199.1345	18.99725	18.97583	19.02717
5	M537T14_	1.778737	0.830853	31.72212	1.16E-05	0.003885	UP	537.1257	537.1252	537.1263	13.75913	13.74583	13.77243
6	M325T11	1.695199	0.761455	49.07387	1.61E-05	0.003885	UP	325.0931	325.0922	325.0951	11.21367	11.20732	11.25245
7	M144T16	2.012298	1.008844	33.20396	1.69E-05	0.003885	UP	144.0457	144.0453	144.0465	16.20212	16.19118	16.21943
8	M357T13_	2.594144	1.375258	32.88002	1.92E-05	0.003885	UP	357.0822	357.082	357.0837	12.66457	12.65622	12.68867
9	M426T8	6.841469	2.774306	22.7141	2.37E-05	0.003885	UP	426.0323	426.0321	426.0345	8.13925	8.13335	8.157967
10	M347T13	1.995841	0.996997	22.19083	2.44E-05	0.003885	UP	347.168	347.1634	347.172	13.05983	13.00283	13.11525
11	M591T15_	2.88852	1.530331	31.66855	2.55E-05	0.003885	UP	591.1369	591.135	591.1397	14.92045	14.89628	14.96255
12	M630T16	2.300034	1.201655	23.19857	3E-05	0.004174	UP	629.784	629.7792	629.7876	16.32718	16.27417	16.34668
13	M438T12_	1.608298	0.685535	24.90868	3.44E-05	0.004174	UP	438.1711	438.1709	438.1714	12.30179	12.27012	12.33347
14	M425T8	4.383821	2.132189	20.21984	3.77E-05	0.004174	UP	425.045	425.0435	425.0457	8.13335	8.112117	8.157967
15	M151T15	2.067456	1.047856	20.92318	4.47E-05	0.004174	UP	151.0399	151.0366	151.0405	14.53876	14.51367	14.5699
16	M506T15	1.435551	0.521604	20.18947	4.83E-05	0.004174	UP	506.1093	506.1066	506.1145	15.21824	15.18848	15.25067
17	M789T15	2.016565	1.0119	18.58276	5.06E-05	0.004174	UP	789.2271	789.2264	789.2278	14.96682	14.95903	14.9746
18	M315T16	3.633561	1.861384	25.13321	5.14E-05	0.004174	UP	315.143	315.1425	315.1436	16.35723	16.35473	16.38143
19	M641T16_	3.070745	1.618589	24.93903	5.26E-05	0.004174	UP	640.7743	640.774	640.7759	16.35723	16.35473	16.38143
20	M582T16_	1.484257	0.569741	18.25259	5.37E-05	0.004174	UP	582.278	582.2752	582.2823	15.50497	15.47808	15.51025

This file has 3,906 lines – we need to apply filters

Content of the Excel file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		name	fold	log2fold	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
1		M500T18_	3.177186	1.66775	-69.9353	4.25E-07	0.000591	DOWN	499.7199	499.7144	499.721	17.9395	17.8955	17.94383
2		M251T15	2.538102	1.34375	57.295	7.07E-07	0.000591	UP	251.0044	251.0023	251.0067	14.64618	14.62393	14.66083
3		M417T18	2.457688	1.297302	55.09394	1.17E-06	0.000651	UP	417.2121	417.2091	417.2133	17.85967	17.82967	17.8865
4		M199T19	2.493736	1.318309	40.96574	2.92E-06	0.001223	UP	199.1334	199.1321	199.1345	18.99725	18.97583	19.02717
5		M537T14_	1.778737	0.830853	31.72212	1.16E-05	0.003885	UP	537.1257	537.1252	537.1263	13.75913	13.74583	13.77243
6		M325T11	1.695199	0.761455	49.07387	1.61E-05	0.003885	UP	325.0931	325.0922	325.0951	11.21367	11.20732	11.25245
7		M144T16	2.012298	1.008844	33.20396	1.69E-05	0.003885	UP	144.0457	144.0453	144.0465	16.20212	16.19118	16.21943
8		M357T13_	2.594144	1.375258	32.88002	1.92E-05	0.003885	UP	357.0822	357.082	357.0837	12.66457	12.65622	12.68867
9		M426T8	6.841469	2.774306	22.7141	2.37E-05	0.003885	UP	426.0323	426.0321	426.0345	8.13925	8.13335	8.157967
10		M347T13	1.995841	0.996997	22.19083	2.44E-05	0.003885	UP	347.168	347.1634	347.172	13.05983	13.00283	13.11525
11		M591T15_	2.88852	1.530331	31.66855	2.55E-05	0.003885	UP	591.1369	591.135	591.1397	14.92045	14.89628	14.96255
12		M630T16	2.300034	1.201655	23.19857	3E-05	0.004174	UP	629.784	629.7792	629.7876	16.32718	16.27417	16.34668
13		M438T12_	1.608298	0.685535	24.90868	3.44E-05	0.004174	UP	438.1711	438.1709	438.1714	12.30179	12.27012	12.33347
14		M425T8	4.383821	2.132189	20.21984	3.77E-05	0.004174	UP	425.045	425.0435	425.0457	8.13335	8.112117	8.157967
15		M151T15	2.067456	1.047856	20.92318	4.47E-05	0.004174	UP	151.0399	151.0366	151.0405	14.53876	14.51367	14.5699
16		M506T15	1.435551	0.521604	20.18947	4.83E-05	0.004174	UP	506.1093	506.1066	506.1145	15.21824	15.18848	15.25067
17		M789T15	2.016565	1.0119	18.58276	5.06E-05	0.004174	UP	789.2271	789.2264	789.2278	14.96682	14.95903	14.9746
18		M315T16	3.633561	1.861384	25.13321	5.14E-05	0.004174	UP	315.143	315.1425	315.1436	16.35723	16.35473	16.38143
19		M641T16_	3.070745	1.618589	24.93903	5.26E-05	0.004174	UP	640.7743	640.774	640.7759	16.35723	16.35473	16.38143
20		M582T16_	1.484257	0.569741	18.25259	5.37E-05	0.004174	UP	582.278	582.2752	582.2823	15.50497	15.47808	15.51025

Preparing for analysis using MetaboAnalyst

http://www.metaboanalyst.ca

Make a copy of the sheet onto new sheet and sort the data by retention time

	name	fold	log2fold	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
1208	M209T3	1.369143	0.453273	2.709012	0.055919	0.077433	UP	208.9602	208.9598	208.9611	3.28215	3.265233	3.3416
378	M323T3	1.434411	0.520459	5.353103	0.006801	0.030027	UP	322.938	322.9359	322.9407	3.32005	3.295083	3.361817
549	M271T3	1.186928	0.247232	5.44988	0.01248	0.038025	UP	270.9417	270.9411	270.9425	3.32005	3.295083	3.361817
895	M387T3	1.394198	0.479436	4.649027	0.031911	0.059621	UP	386.9358	386.9344	386.9368	3.32005	3.295083	3.361817
1689	M272T3	1.088879	0.122844	2.524051	0.111581	0.110419	UP	271.941	271.9398	271.9434	3.326167	3.29085	3.361817
892	M325T3	1.375868	0.460342	3.79855	0.031801	0.059554	UP	324.9303	324.929	324.9401	3.32705	3.30445	3.361817
1378	M391T3	1.312616	0.392445	3.507716	0.071514	0.086812	UP	390.9101	390.9064	390.9131	3.32705	3.30445	3.367883
1623	M291T3	1.256324	-0.32921	-2.50694	0.102207	0.105294	DOWN	290.9445	290.9436	290.9464	3.327583	3.295083	3.361817
749	M324T3	1.275929	0.351548	3.759227	0.022331	0.049864	UP	323.9378	323.9374	323.9399	3.331108	3.298017	3.361817
2099	M154T3	1.037093	0.052546	1.655429	0.189385	0.150928	UP	154.0059	154.0023	154.0117	3.331492	3.30095	3.361817

Scroll down to Rt 5.00 min

1845	M387T5	1.154483	0.207247	2.022386	0.135817	0.123138	UP	387.114	387.113	387.1145	4.978417	4.97	4.99685
2204	M388T5	1.139738	0.188702	1.768345	0.215874	0.163842	UP	388.1168	388.115	388.1176	4.978417	4.970667	4.99615
1315	M404T5	1.345221	-0.42784	-2.9703	0.065723	0.083602	DOWN	404.1036	404.1014	404.1047	4.983425	4.945983	5.010417
1128	M564T5	1.522219	0.606175	3.224519	0.049554	0.073486	UP	564.0012	563.9964	564.0092	4.984925	4.917883	5.0261
2862	M772T5_2	1.106141	-0.14554	-0.84381	0.44698	0.26116	DOWN	772.2342	772.227	772.2367	4.992317	4.97	5.037
3327	M643T5_1	1.122221	-0.16636	-0.46213	0.67351	0.338631	DOWN	642.6763	642.6715	642.6794	4.992317	4.930767	4.99685
1060	M419T5	1.160736	-0.21504	-3.76575	0.042847	0.067616	DOWN	419.1222	419.1178	419.1251	4.998867	4.97	5.037
3479	M418T5	1.009913	0.01423	0.343979	0.752961	0.362038	UP	418.1259	418.1251	418.1274	4.999242	4.99615	5.037
805	M534T5	2.270077	1.182742	4.124107	0.02608	0.054177	UP	534.0049	533.9957	534.0074	4.999617	4.99685	5.09565
2042	M768T5	1.190096	-0.25108	-1.66652	0.174057	0.142584	DOWN	768.2396	768.2371	768.241	4.999617	4.992317	5.022167
1207	M446T5	1.377175	0.461712	2.687478	0.055893	0.077433	UP	446.1187	446.1181	446.1194	5.003833	4.970667	5.037

Delete all these records

Keep metabolites eluting between 5.00 and 25.00 minutes

Areas of aligned metabolites by sample

negmode_nr1	negmode_nr2	negmode_nr3	negmode_ir1	negmode_ir2	negmode_ir3
9846	9677	9931	3168	3131	2971
28534	27967	29683	71934	72858	73952
6340	6069	6014	15216	15225	14838
18534	19605	20066	49283	47408	48456
17847	16952	16751	31039	30484	30171
23162	22952	22637	38605	38521	39420
26142	25163	24680	52038	51045	49821
12999	12148	11915	31502	31596	33047
893	1152	1618	8147	8816	8098
10415	11219	10089	20918	20640	21755
26263	26389	23874	74889	71120	75040
3051	3046	3467	7160	7542	7298
15492	16557	16286	26297	25938	25503
3088	2348	3495	13268	13551	12332
7302	8259	8242	16460	16811	15942
23595	22994	24105	33963	34434	33088
10301	9069	9498	20043	18719	19453
6312	6126	7336	24880	22887	24081
3677	3235	2891	10144	10218	9742
13805	14693	14417	21725	20777	21197

non-irradiated diet irradiated diet

Now order them according to peak areas

- The goal here is to remove noisy peaks that contribute little to the overall separation of groups
  - It's a good idea to inspect the data with mzmine or xcms to see where the peaks become ragged
  - It's important to write down where the cutoff is for a given analysis
  - If there are >2,000 features left, then I artifically set a 2,000 features limit
    - Why? Metaboanalyst restricts the number of features it uses, but there are other stats programs that use bigger matrices

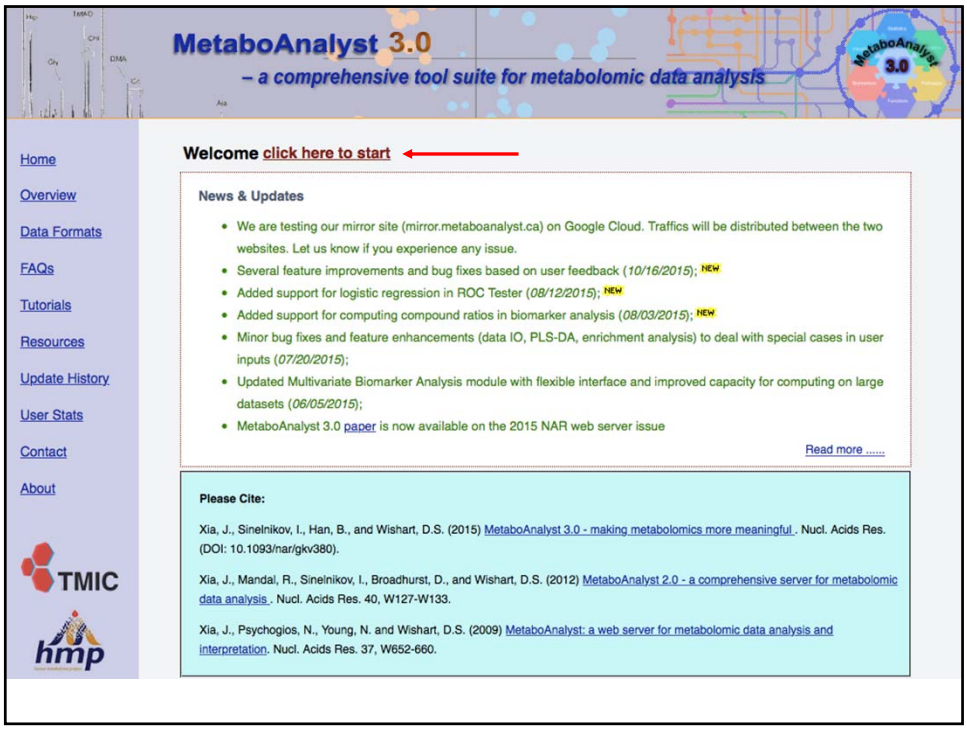
Creating .csv files for each sample

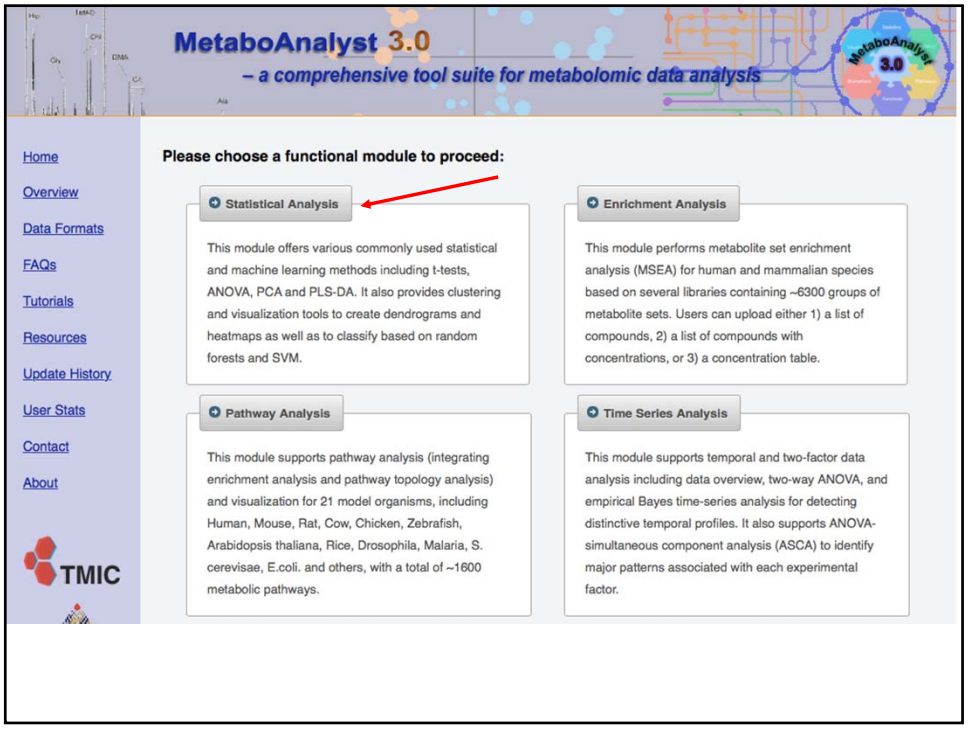
	mzmed	rtmed	negmode_nr1
2	499.71989	17.9395	9846
3	251.004448	14.646175	28534
4	417.212131	17.8596667	6340
5	199.133402	18.99725	18534
6	537.125737	13.7591333	17847
7	325.093116	11.2136667	23162
8	144.045677	16.2021167	26142
9	357.082227	12.6645667	12999
10	426.03226	8.13925	893
11	347.167955	13.059825	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

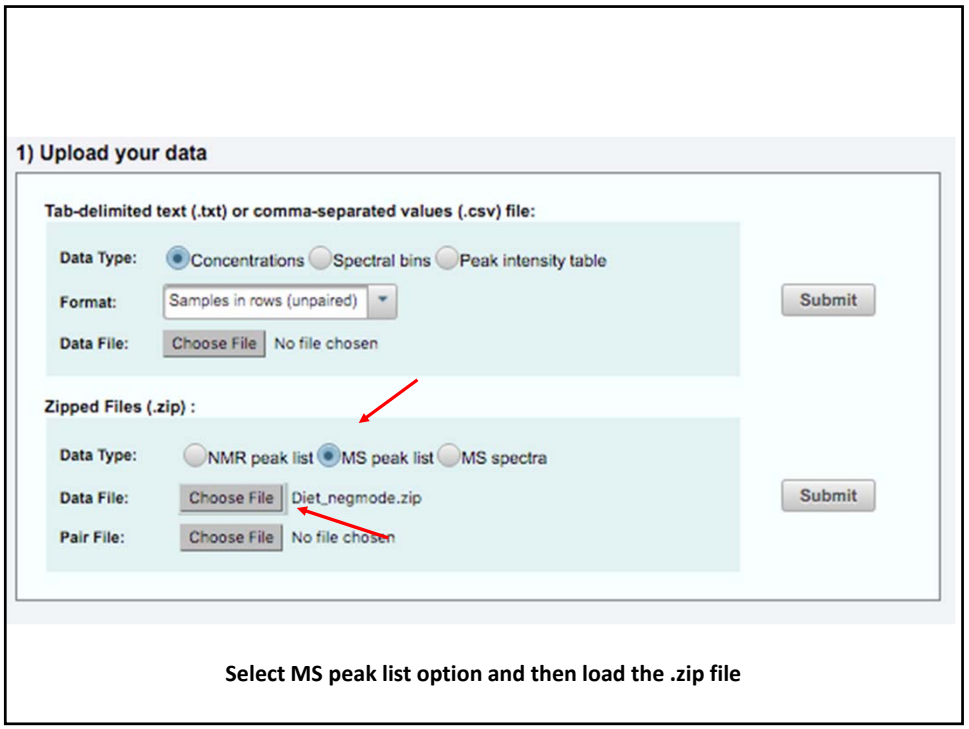
- Copy the median *m/z* and median Rt values into a new Excel file. Then copy the column of areas from the first sample in Group_1. Save as an Excel .csv file.
  - Note that the file name must not have spaces – use an underscore instead of a space.
- Leave the file open and replace the yellow column 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

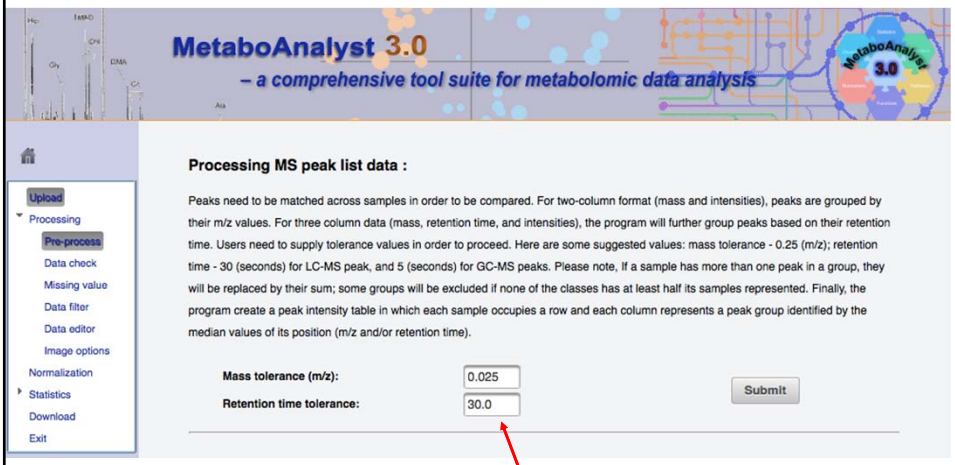
- Put each of the .csv files for group_1 samples into a folder named “Group_1”.
- Put each of the .csv files for group_2 samples into a folder named “Group_2”.
- Click on Group_1 and Group_2 folders and combine to form a .zip file.
  - Rename the .zip file as [your_name].zip
- You’re now ready to submit it to MetaboAnalyst
  - http://www.metaboanalyst.ca



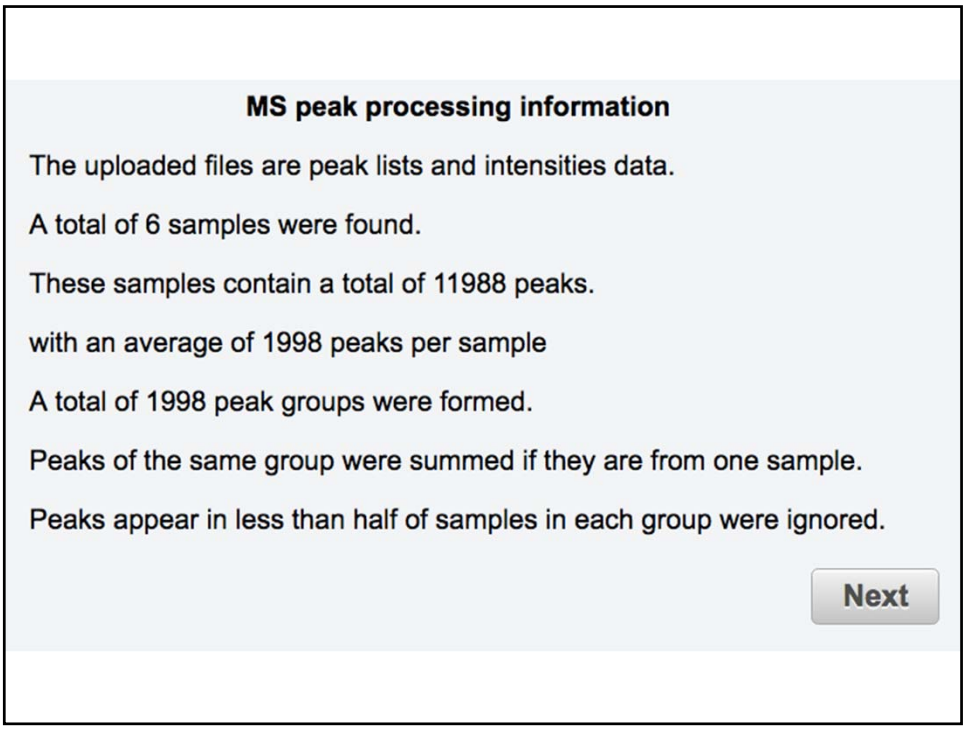


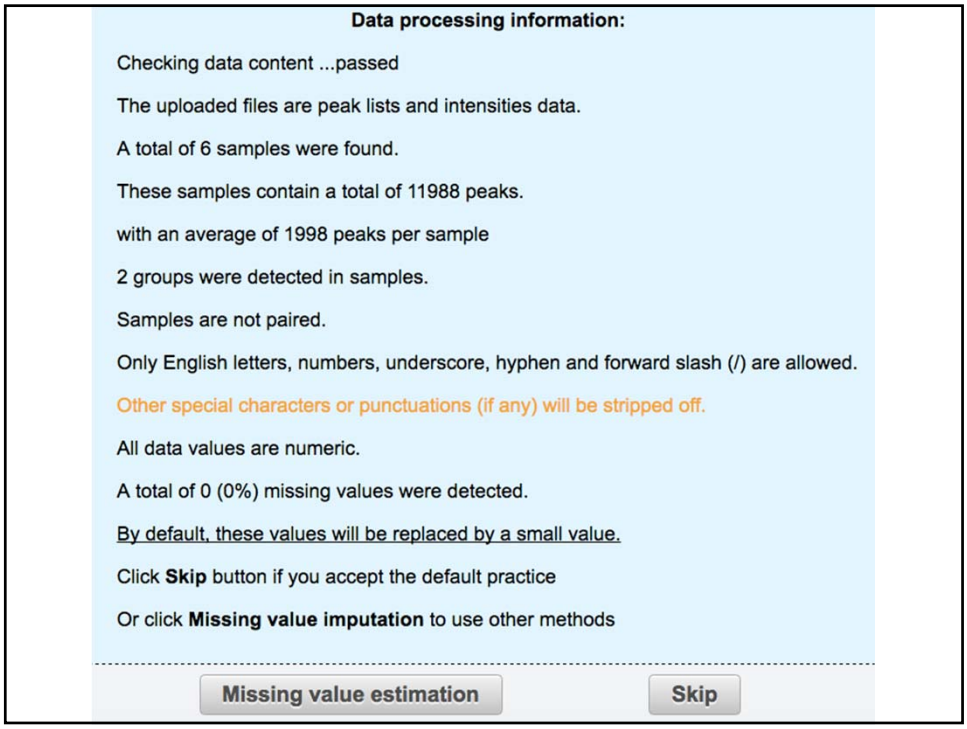


Select MS peak list option and then load the .zip file



reset these to 0.001 and 0.01, respectively





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 11988 peaks.

with an average of 1998 peaks per sample

2 groups were detected in samples.

Samples are not paired.

Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed.

Other special characters or punctuations (if any) will be stripped off.

All data values are numeric.

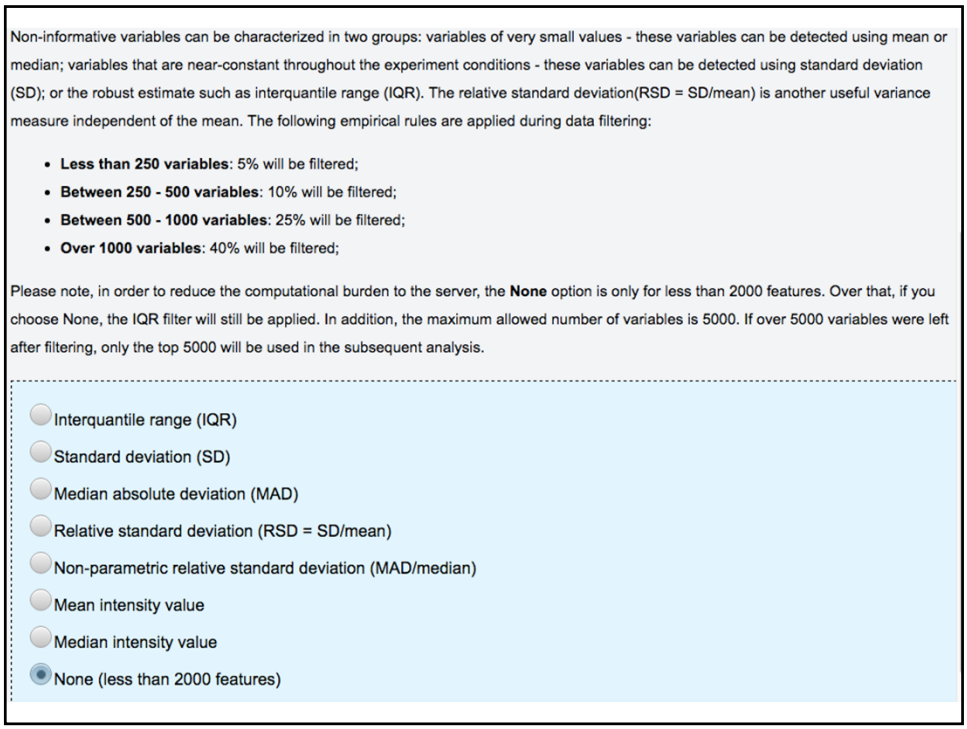
A total of 0 (0%) missing values were detected.

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

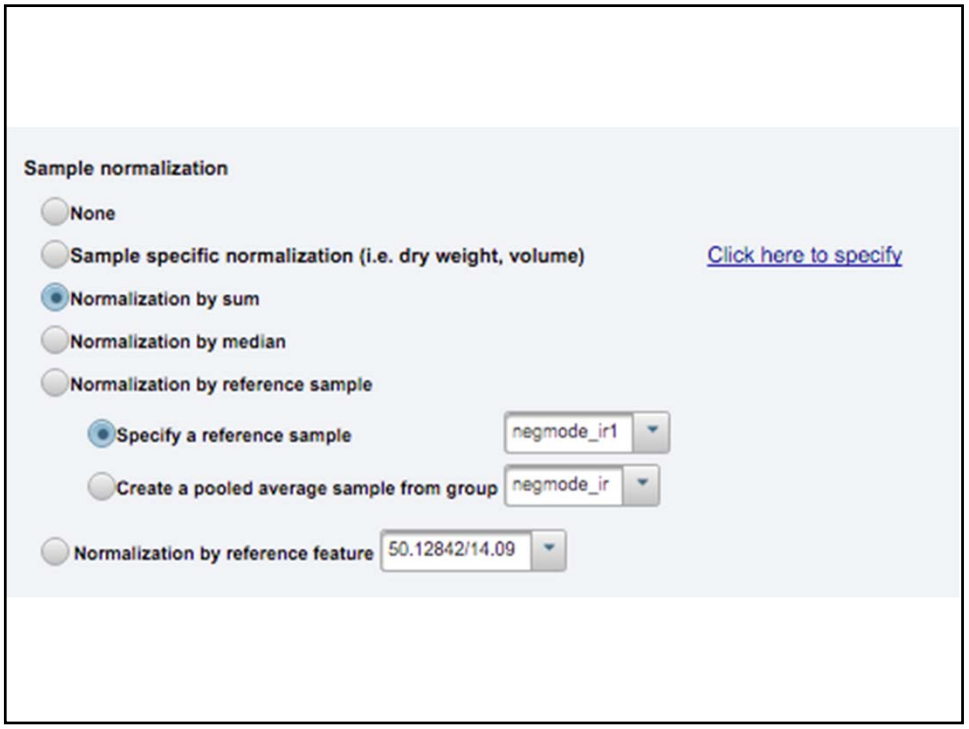


Non-informative variables can be characterized in two groups: variables of very small values - these variables can be detected using mean or median; variables that are near-constant throughout the experiment conditions - these variables can be detected using standard deviation (SD); or the robust estimate such as interquantile 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.

- ○ Interquantile 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)



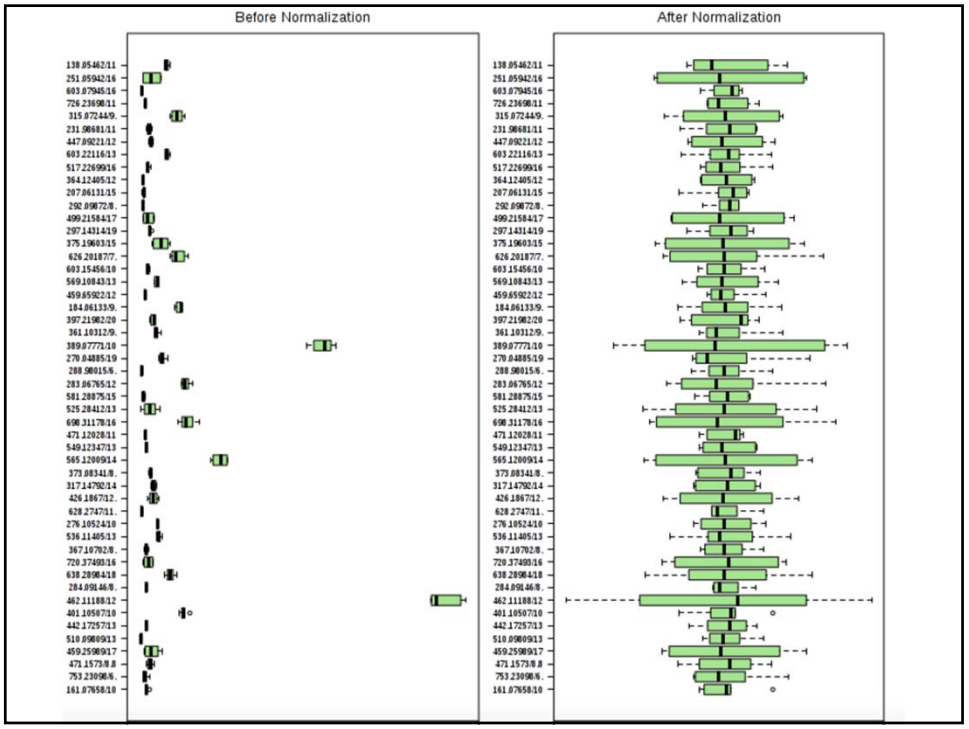
Data options before stats analysis

Data transformation

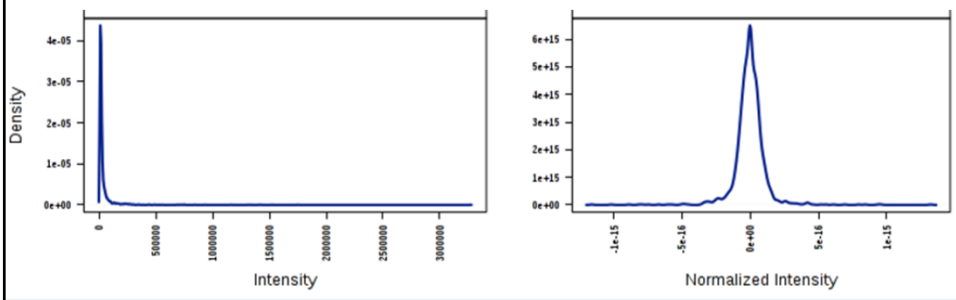
- None
- Log transformation (generalized logarithm transformation or glog)
- Cube root transformation (take cube root of data values)

Data scaling

- None
- Auto scaling (mean-centered and divided by the standard deviation of each variable)
- Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)
- Range scaling (mean-centered and divided by the range of each variable)



Effect of normalization and scaling



Submit

You're now ready to use the statistical packages of Metaboanalyst