
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 data from a small study
- Then we'll download the XCMS output file
 - Prepare files for MetaboAnalyst

Logon to XCMS with your name and password

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

Available on the AppStore

10550 North Torrey Pines Road BCC-007, La Jolla, CA 92037 USA - (858) 794-9415, Fax (858) 794-9496

Home | Privacy Policy | Terms of Use | Contact Us | Reset Password

Viewing the datasets

Home MRM Databases Create Job View Results XCMS Public XCMS Institute Stored Datasets Account Toolbox Help Logout [empty1877]

Show 15 rows Share Resubmit Job Grouping Delete Search...

JobID	Status	Progress	JobName	Datasets / Sources	Created	Parameters (ID#)	Group Filter	Share	Delete
1243745	job complete		P_2018-07-11_17:27	Urine_Itra (#542035) Urine_Noni (#542036)	2018-07-11 17:27:05	21944			

Click on this button

Viewing the data



Submit Date	Finish Date	Total Aligned Features	Parameter ID#	Log	Shared	Download Results
2018-07-11 17:28:48	2018-07-11 18:26:39	4141	LW_TripToF_Negativ (21944)	View Log	NOT SHARED	hash: 83d6e0ad5182d789492a5e51eec4439f

The finished job has the following notes:

2018-07-11 18:07:23 : i-HeatMap data prep, memory requires limiting to top 1000 features <0.0231597 p-values

The download will take several minutes. However, the folder is on your thumb drive

Next steps

- **Opening the .zip file you've downloaded (or the folder on your thumbdrive)**
- **Open the folder and create the Excel file from the XCMSdiffreport.xlsx file**
- **Preparing for MetaboAnalyst**

▶ boxplot	Today, 6:04 PM	--	Folder
CloudPlot-svg.svg	Today, 6:10 PM	2.6 MB	SVG Document
CloudPlot.pdf	Today, 6:09 PM	893 KB	PDF Document
CloudPlot.png	Today, 6:09 PM	129 KB	PNG image
▶ EIC	Today, 6:07 PM	--	Folder
MDS.pdf	Today, 6:07 PM	5 KB	PDF Document
MDS.png	Today, 6:07 PM	20 KB	PNG image
▶ ms2_spectra	Today, 6:22 PM	↑ 2.9 MB	Folder
▶ mummichog	Today, 6:23 PM	↑ 20 KB	Folder
MVstats_ScalingPlot_1243745.pdf	Today, 6:07 PM	105 KB	PDF Document
PCA-diagnostics.pdf	Today, 6:07 PM	5 KB	PDF Document
PCA-diagnostics.png	Today, 6:07 PM	5 KB	PNG image
PCA-loadings-all.pdf	Today, 6:07 PM	37 KB	PDF Document
PCA-loadings-all.png	Today, 6:07 PM	55 KB	PNG image
PCA.pdf	Today, 6:07 PM	5 KB	PDF Document
PCA.png	Today, 6:07 PM	20 KB	PNG image
result.tsv	Today, 6:22 PM	5.9 MB	Plain Text
Rplots.pdf	Today, 6:22 PM	20 KB	PDF Document
rtcor.pdf	Today, 5:44 PM	179 KB	PDF Document
rtcor.png	Today, 5:44 PM	161 KB	PNG image
TICs_rtcor.pdf	Today, 5:46 PM	189 KB	PDF Document
TICs_rtcor.png	Today, 5:46 PM	69 KB	PNG image
TICs.pdf	Today, 5:39 PM	189 KB	PDF Document
TICs.png	Today, 5:39 PM	68 KB	PNG image
XCMS-diffreport-MultiClass.xlsx	Today, 6:07 PM	2.2 MB	Micros...(xlsx)
XCMS.annotated.dif...iated_NegMode.tsv	Today, 6:09 PM	2.7 MB	Plain Text
XCMS.diffreport.Ur...iated_NegMode.tsv	Today, 6:22 PM	2.6 MB	Plain Text
xcmsMS2_location.pdf	Today, 6:22 PM	763 KB	PDF Document
xcmsMS2_location.png	Today, 6:22 PM	32 KB	PNG image
xcmsMS2-results_1243745.tsv	Today, 6:22 PM	16 KB	Plain Text
XCMSOnline_log.txt	Today, 6:11 PM	3 KB	Plain Text

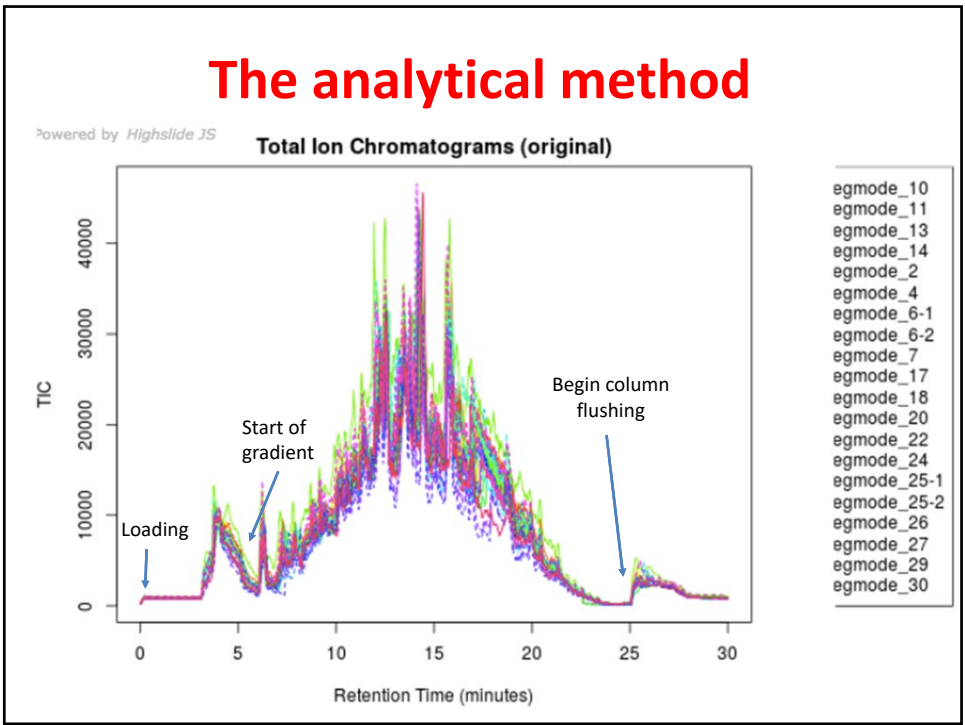
Double click on this file

The Excel DiffReport from XCMS

1	name	fold	log2fold	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
2	M242T16	2.528709	-1.3384	-12.2959	4.67E-10	9.34E-07	DOWN	242.1388	242.1332	242.1396	15.51123	15.2187	15.54245
3	M307T14	1.97078	-0.97877	-15.4337	3.56E-09	2.51E-06	DOWN	307.1746	307.1729	307.176	14.04443	13.6414	14.15875
4	M289T11	1.592988	-0.67174	-12.3448	4.47E-09	2.65E-06	DOWN	289.154	289.1486	289.1608	10.70822	10.59758	11.90052
5	M412T12	1.651227	0.723538	10.18327	6.76E-09	2.86E-06	UP	421.1384	421.1364	421.1413	11.82238	11.59022	11.85238
6	M283T19	2.708026	-1.43724	-13.2267	7.21E-09	2.89E-06	DOWN	283.1883	283.1783	283.191	19.284	19.03933	19.33933
7	M466T13	1.79271	-0.84214	-9.53068	2.42E-08	7.1E-06	DOWN	466.2056	466.2011	466.2118	12.55863	12.25908	12.6492
8	M477T9_1	1.869136	-0.90237	-9.56723	3.26E-08	8.43E-06	DOWN	477.1013	477.0978	477.105	9.0531	8.93815	9.113367
9	M429T11	1.888684	0.917382	9.118729	3.64E-08	8.95E-06	UP	429.0733	429.0725	429.0748	11.22607	11.12948	11.28857
10	M457T12	1.943456	-0.95862	-10.4882	4.7E-08	1.01E-05	DOWN	457.1473	457.145	457.1491	11.93708	11.66887	12.05877
11	M240T15	2.320588	-1.21449	-9.07403	5.37E-08	1.07E-05	DOWN	240.1208	240.1173	240.122	15.40782	15.10635	15.42427
12	M259T11	1.850303	0.887762	9.61166	7.35E-08	1.26E-05	UP	259.0095	259.0054	259.0142	11.20622	11.18288	11.2804
13	M273T16	1.647969	-0.72069	-8.63945	8.16E-08	1.32E-05	DOWN	273.1701	273.1684	273.1713	16.0643	15.43132	16.13505
14	M330T17	3.377789	1.756079	9.338606	9.48E-08	1.4E-05	UP	330.0753	330.0728	330.0806	17.48317	17.10233	18.21267
15	M384T12	1.623582	-0.69918	-9.25674	1.11E-07	1.49E-05	DOWN	384.1271	384.1214	384.1345	11.74763	11.6512	11.88625
16	M430T11	1.607448	0.684772	8.756176	1.12E-07	1.49E-05	UP	430.0819	430.0789	430.0868	11.19486	10.99428	11.26268
17	M278T15	1.863182	-0.89777	-8.45928	1.38E-07	1.63E-05	DOWN	278.0922	278.0892	278.0998	14.69891	14.5221	15.97455
18	M267T16	1.8504	-0.88784	-8.38267	1.39E-07	1.64E-05	DOWN	267.0937	267.0906	267.1037	16.10808	15.78458	16.21468
19	M429T11	2.004833	1.003482	8.370111	1.83E-07	1.92E-05	UP	429.0881	429.0841	429.0973	11.02918	10.9845	11.12535
20	M329T17	3.562235	1.832783	8.863357	1.93E-07	1.97E-05	UP	329.0716	329.0695	329.075	17.48567	17.10233	18.21267
21	M311T15	2.338164	-1.22538	-8.46809	2.02E-07	2.02E-05	DOWN	311.1459	311.1419	311.1525	15.35683	15.10635	15.40063
22	M413T20	2.091061	-1.06424	-8.30362	2.35E-07	2.23E-05	DOWN	413.2563	413.2519	413.2589	20.4105	20.23867	20.50383
23	M262T11	1.787116	0.837633	9.093879	3.11E-07	2.63E-05	UP	262.019	262.0147	262.0274	10.69167	10.62415	10.76145
24	M284T19	1.911828	-0.93495	-7.80717	3.78E-07	2.92E-05	DOWN	284.1903	284.1873	284.1932	19.2765	19.03933	20.51617
25	M176T7	3.601911	-1.84876	-10.7248	4.01E-07	3.01E-05	DOWN	176.0282	176.0265	176.0308	7.282517	7.20535	7.33075
26	M606T17	2.540086	-1.34488	-7.72598	4.02E-07	3.01E-05	DOWN	606.2257	606.2204	606.2294	17.45333	16.88033	17.982
27	M209T12	2.697048	-1.43138	-10.1982	4.27E-07	3.1E-05	DOWN	209.0496	209.0455	209.0517	11.85969	11.66887	12.03247
28	M695T20	4.637973	-2.21349	-9.17267	4.83E-07	3.28E-05	DOWN	695.4644	695.4607	695.4733	19.9125	19.71033	19.96983
29	M244T17	2.147928	-1.10295	-7.68792	4.98E-07	3.33E-05	DOWN	244.1572	244.1556	244.1593	16.51004	16.25235	16.72118
30	M756T11	1.492111	0.577355	7.583692	5.21E-07	3.4E-05	UP	756.2135	756.2046	756.2168	10.93169	10.79963	10.99377
31	M228T21	21.60493	-4.43329	-14.3026	5.27E-07	3.41E-05	DOWN	228.1609	228.1587	228.1689	20.62717	20.389	20.85083
32	M260T11	1.993798	0.99552	8.827898	5.3E-07	3.42E-05	UP	260.0013	259.9955	260.01	11.25297	9.95175	11.39432

This file has 4,141 lines – we will apply filters

**Preparing for analysis using
MetaboAnalyst**
<http://www.metaboanalyst.ca>



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

1	name	fold	log2fold	tstat	pvalue	qvalue	updown	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
2	M382T3_2	1.10807587	-0.1480567	-0.6514369	0.52300613	0.32851916	DOWN	382.417881	382.410159	382.421705	3.355375	3.3039	3.46276667
3	M416T3_1	1.09698108	-0.1335386	-0.7365674	0.47134416	0.30887372	DOWN	415.921851	415.916163	415.925036	3.36696667	3.3419	3.41575
4	M278T3	1.14518424	-0.1955797	-1.2700993	0.22360818	0.20129774	DOWN	277.937048	277.934124	277.938803	3.36846667	3.33978333	3.53883333
5	M196T3	1.00761229	-0.0109406	-0.1501481	0.8824778	0.45185506	DOWN	195.975461	195.973824	195.977207	3.37051667	3.33371667	3.53883333
6	M291T3	1.04372581	-0.0617428	-0.5992451	0.55661622	0.34208063	DOWN	290.944565	290.941391	290.947134	3.3723	3.3039	3.52818333
7	M316T3_1	1.19640718	-0.2582743	-1.5743605	0.13286303	0.15120673	DOWN	315.958426	315.955793	315.959914	3.37610833	3.3039	3.53883333
8	M424T3_3	1.14344344	-0.193385	-0.8067947	0.43120585	0.29257244	DOWN	424.402421	424.397687	424.407527	3.37628333	3.30986667	3.46276667
9	M318T3	1.25759926	-0.3306723	-1.591533	0.13113076	0.15005393	DOWN	318.431409	318.426713	318.43482	3.3763	3.33978333	3.4045
10	M383T3	1.15555992	-0.2085921	-0.8884291	0.38610798	0.27484227	DOWN	383.422304	383.415585	383.429135	3.3763	3.3039	3.46276667
11	M328T3_2	1.29571992	-0.3737539	-1.6918605	0.11037136	0.13545078	DOWN	328.431917	328.426196	328.433131	3.376425	3.3039	3.53883333
1008	M523T5	1.42590787	-0.5118808	-3.6101868	0.00200806	0.00801609	DOWN	523.028927	523.024055	523.03205	4.96486667	4.89351667	4.99355
1009	M679T5	1.18324279	-0.2427461	-0.8945175	0.3845001	0.27421457	DOWN	678.975962	678.968302	678.981822	4.9785	4.29158333	5.16933333
1010	M353T5	1.60633317	0.68377116	1.45777926	0.16346511	0.17006607	UP	352.83134	352.825997	352.834288	4.97886667	3.89966667	5.15938333
1011	M313T5	1.82928308	0.87127834	2.02985162	0.05823016	0.08973891	UP	330.845838	330.844053	330.84922	4.98518333	4.854	5.15938333
1012	M289T5	1.59217176	-0.670996	-1.5791323	0.14671	0.1601398	DOWN	288.986526	288.98533	288.987094	5.0106	4.773	5.21298333
1013	M342T5	1.10869088	0.14885718	0.45276009	0.65692023	0.38028251	UP	341.8982	341.896811	341.903903	5.01165833	4.85351667	6.72663333
1014	M166T5	1.4573565	-0.5433538	-2.1428487	0.06126448	0.09298497	DOWN	166.017914	166.016844	166.018698	5.02245	4.7442	5.92781667
1015	M311T5	1.52824308	-0.611874	-1.3465362	0.20737575	0.19323868	DOWN	310.968073	310.967044	310.968962	5.02675833	4.93378333	5.25665
1016	M210T5	1.13862966	-0.1872986	-1.5286463	0.15272211	0.16381192	DOWN	210.008057	210.006795	210.010572	5.03755	4.9526	6.64006667
1017	M739T5	1.59996207	-0.6780377	-2.734453	0.0173826	0.03802625	DOWN	738.992841	738.986366	738.996953	5.06338333	4.3172	5.32201667

Keep metabolites eluting between 5.00 and 25.00 minutes

The metabolites eluting before 5 min did not bind to the column, so we'll delete them since their elution times are not reproducible.

Deleting the data after 25 min

053	M659T23	1.01180592	0.01693259	0.06450169	0.94950001	0.47004184	UP	659.468962	659.467123	659.47141	23.3793333	23.3268333	23.4183333
054	M207T23	1.04539194	0.06404394	0.20763841	0.83823608	0.43915008	UP	207.175944	207.174418	207.176666	23.3806667	23.3333333	23.43
055	M251T23	1.1005572	-0.1382341	-0.4605757	0.65181528	0.3784457	DOWN	251.165258	251.163882	251.166213	23.3866667	22.435	23.43
056	M252T23	1.0149763	-0.021446	-0.0683995	0.94640846	0.46922952	DOWN	252.168782	252.167581	252.169755	23.3870833	22.6441667	23.43
057	M254T23	1.12325615	0.16768696	0.54957168	0.59048454	0.35549567	UP	254.184412	254.183078	254.185446	23.4018333	23.3268333	23.845
058	M761T24	1.18231086	0.2416094	0.81563559	0.42651259	0.29061301	UP	760.967262	760.965743	760.973119	23.5133333	23.4328333	23.6413333
059	M253T24	1.19546604	-0.2575731	-0.6316044	0.53791025	0.33442912	DOWN	253.181282	253.179677	253.182315	23.594	23.3268333	23.845
060	M515T24	1.07172664	-0.099937	-0.4007647	0.69499212	0.39364613	DOWN	514.974427	514.973417	514.978549	23.6265833	23.5881667	23.6593333
061	M315T24	1.17278182	-0.2299346	-0.6702726	0.51333776	0.32490743	DOWN	315.253856	315.253261	315.254868	23.6908333	23.6618333	24.0635
062	M661T25	1.17591472	0.23378343	0.94922166	0.35800405	0.26351847	UP	660.973411	660.972146	660.979547	24.5469167	24.482	24.5633333
063	M721T25	1.09245844	0.1275784	0.74897419	0.46565698	0.3066147	UP	720.598428	720.594134	720.604095	25.13075	24.8815	25.2216667
064	M720T25	1.10052318	0.13818953	0.78393982	0.44490135	0.29825116	UP	719.594182	719.592451	719.600278	25.1415833	24.8815	25.2216667
065	M489T25	1.27908494	0.35511207	1.97603101	0.072836	0.10451063	UP	489.262599	489.261366	489.264447	25.1519167	25.109	25.1925
066	M487T25	1.07876778	0.10938433	0.5902681	0.56439073	0.34520922	UP	487.265789	487.264808	487.269012	25.1521667	24.8815	25.2216667
067	M488T25	1.23625936	0.30598144	1.50674058	0.15688468	0.16628065	UP	488.268031	488.267019	488.270078	25.1581667	25.109	25.1925
068	M439T25	1.1813542	0.24044159	1.2262862	0.24016976	0.20949731	UP	439.285422	439.284361	439.287183	25.1785833	25.1461667	25.219
069	M437T25	1.09742755	0.13412569	0.71315999	0.48697541	0.31497265	UP	437.283785	437.282935	437.28603	25.1839167	24.9025	25.229
070	M438T25	1.06723814	0.09388213	0.49734775	0.62725904	0.36945576	UP	438.286575	438.285347	438.288841	25.1909167	24.9231667	25.2471667
071	M412T25	1.08085405	0.11217173	0.67905168	0.50973398	0.32357453	UP	412.287351	412.286455	412.290711	25.1925833	24.9231667	25.278333
072	M411T25	1.15161477	0.20365819	1.21744429	0.24302508	0.21085981	UP	411.284596	411.283833	411.286953	25.1926667	24.9231667	25.2471667
073	M505T25	1.14425831	0.19441277	1.23963072	0.24411474	0.21137595	UP	505.270938	505.269394	505.272294	25.2751667	25.2395	25.313
074	M367T25	1.22030142	-0.2872375	-0.6119488	0.54890808	0.33894909	DOWN	367.158113	367.157109	367.159847	25.28625	25.0043333	25.442
075	M233T25	1.50897058	-0.5935647	-1.65361	0.1205499	0.14272841	DOWN	233.031095	233.030966	233.031878	25.2885	25.2193333	25.442
076	M496T25	1.32051893	0.40110499	1.24363358	0.22994563	0.20449667	UP	496.27235	496.269788	496.275712	25.292	24.9856667	26.3073333

Delete the lines shaded in yellow (and those below them in the file)

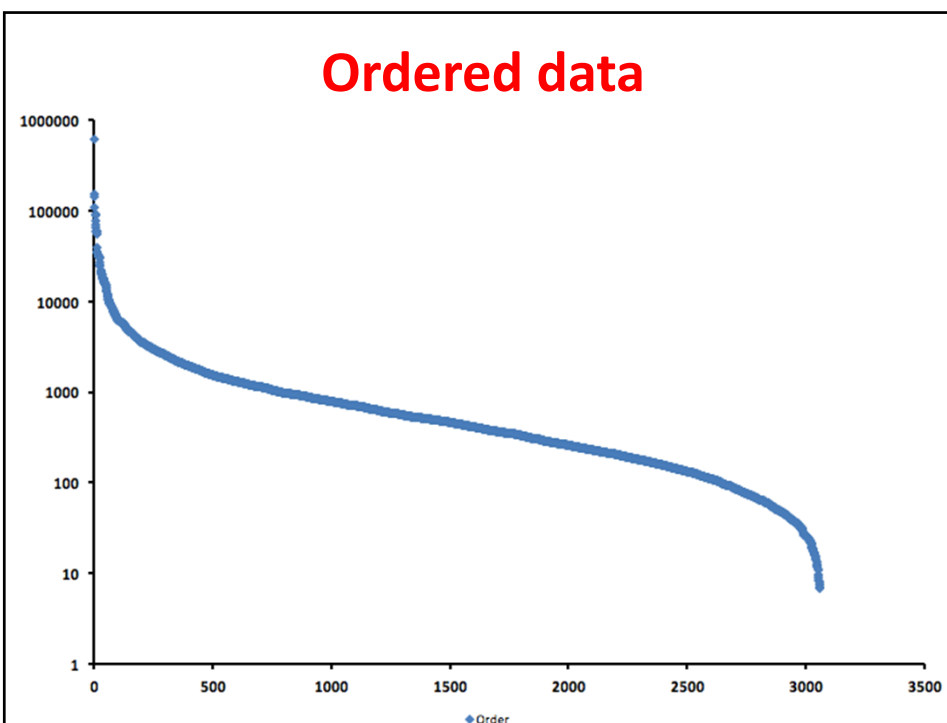
We now have 3061 features

Now order them according to peak areas

- The goal here is to consider removing low intensity noisy peaks that contribute little to the overall separation of groups
- Copy the data to a new sheet and order them according to area – use mean 1

maxint	mean1	sd1	mean2	sd2	negmode_10	negmode_11	negmode_13	negmode_14	negmode_2	negmode_4
15008.1527	601797.17	144720.17	497680.693	219965.734	821794.728	650977.198	660933.137	440569.142	564353.675	755078.492
16730.3759	153488.74	67141.314	183645.421	69078.7808	162453.407	134950.382	151040.698	66083.519	91546.2719	91176.9451
10820.8805	142047.45	42934.1149	198896.119	74700.423	135407.213	107101.991	105232.816	181577.719	151158.579	232762.607
15179.2705	108539.97	17285.1085	137980.601	30377.5875	99100.8848	99783.8582	112668.372	91457.9271	108463.28	151026.523
11360.866	91559.60	6993.80089	80436.0471	30430.0404	90632.8043	89983.6966	88442.8012	91406.4257	103053.328	102518.322
8084.88798	87163.63	8229.05989	75234.1003	26445.6315	87724.1019	88653.6399	82891.8057	87535.3794	95860.6586	102758.16
6119.97233	76925.84	6945.32236	74711.5991	9906.61673	67135.6574	78489.7469	79827.6388	65140.575	73400.53	79491.7569
8560.70886	69881.97	14846.6996	62431.3273	12942.7815	77269.7591	86597.7649	69867.7853	59864.7896	58950.0764	95148.3325
7629.28711	65577.96	7094.29101	63189.6175	9805.53973	57574.2466	59491.3356	59213.3684	70053.876	76633.8709	70361.3727
4252.72769	58849.06	14496.6082	47548.0774	20475.0389	88617.9607	58569.9905	56348.1406	47566.2551	65967.7454	67755.2354
7231.88032	58641.41	11670.9394	44350.1635	18487.1829	70206.8219	74711.2105	54598.0888	49253.8964	47320.3297	74430.0839
6791.26071	58411.60	8632.44208	56600.7057	18552.0483	49691.3394	59123.5391	50813.8479	52297.9495	54750.616	71808.5454
5843.15012	55206.68	23546.9128	55016.3662	19242.7978	39218.8327	84189.7972	32466.568	49234.0402	35447.3245	102316.524
7596.79983	38377.50	6985.23579	38472.2636	10708.6089	34153.219	38471.0341	41764.2297	23991.9902	33109.8596	43301.5115

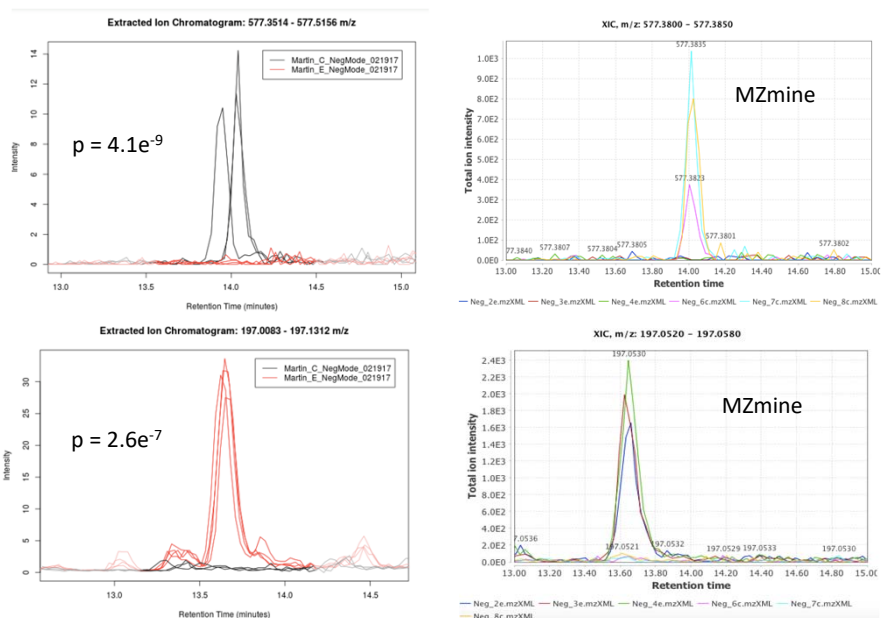
Ordered data



Rationale for data clean up

- It's a good idea to inspect the data with MZmine or XCMS to see where the peaks become ragged
 - It's also important to write down where the cutoff is for a given analysis
- **MetaboAnalyst can accept 4000 features without deletion**
 - If there are more than 4000 features, then an interquartile range filtering is applied (deletes the top and bottom 25% of features)

Data inspection – top two most significant



Color code the sample groups

Group 1

Group 2

negmode_10	negmode_11	negmode_13	negmode_14	negmode_2	negmode_4	negmode_5.1	negmode_5.2	negmode_7	negmode_17	negmode_18	negmode_20	negmode_22	negmode_24	negmode_25	negmode_25
852.119488	1243.08521	647.932053	324.9314	394.42747	2105.07516	565.671457	641.905087	592.007544	870.650469	270.899344	485.057062	361.853756	384.142891	437.620461	604.94302
104.996374	140.590488	169.340261	102.878777	154.292668	141.235655	407.383202	274.1862	98.2076553	171.60449	121.575677	123.284284	203.790919	92.4655586	316.226633	246.566966
1099.16397	1344.89411	1305.51922	942.914781	1324.40996	1819.03985	3237.325	1461.27599	1684.54192	1278.86074	1036.28613	1245.73604	912.285837	1296.92013	154.56853	853.822313
562.947757	963.82433	410.192487	256.774362	326.378239	1517.68661	275.792707	359.549746	487.854071	857.468068	212.022295	400.991731	281.568085	343.113475	293.33973	329.131841
1084.30719	962.174955	1681.02854	1403.33248	1737.24274	998.258879	1300.37237	1361.2516	1165.30264	1153.58701	1402.11135	1201.25436	951.567464	1104.04516	913.909381	1045.10272
49.2257086	64.7693604	39.9646756	32.2400866	31.595901	71.2375558	84.2832746	38.1043982	49.7597242	50.2019243	23.9972597	40.5781925	21.4918339	40.6859892	18.7989535	25.5633721
969.736874	1030.10476	749.842751	641.370233	967.66456	698.319653	602.642245	670.062518	958.323955	669.720648	640.810727	641.257912	560.881442	550.214787	576.823843	624.754846
339.370105	544.752776	272.110721	153.054607	249.359144	785.146795	213.217557	205.77184	284.974231	507.876499	146.19873	230.332879	208.187225	288.877474	282.828193	295.148477
156.499148	262.2181	125.036519	75.754915	111.985436	324.458763	63.2751122	95.022856	157.801236	274.197512	63.9981546	109.577415	62.835451	90.3001304	67.590866	84.1164385
103.781657	106.486408	104.64113	67.336095	90.3564081	90.8842581	69.36347	112.985438	88.87384	153.799887	71.7035406	109.13913	89.4392615	79.5734308	78.4605563	112.947692
131.715145	151.862531	144.039309	107.030102	122.55632	134.610781	106.076956	144.773905	129.562523	192.768636	91.684432	165.469355	127.245835	125.740147	97.6472344	154.237222
306.07823	606.740788	255.617415	499.419788	393.410669	821.29467	206.980596	350.187229	505.955211	586.179637	283.422577	631.071152	250.128521	606.818855	195.498229	239.013568
2557.60796	1335.16259	185.01598	2769.90277	1213.95428	1168.73386	1106.46674	1152.27093	1683.43232	564.560277	1419.03587	1529.5187	551.663284	1585.00071	389.089353	429.294534
52.0863984	57.6949787	55.9711788	56.4995285	59.5053627	55.4164828	48.589301	53.9095237	57.818445	72.8647952	48.8944099	57.0664373	46.1939666	48.2911065	45.8493237	52.3483183
112.212582	156.335882	121.150209	118.179434	132.794225	140.485483	88.601618	107.719528	132.579814	180.928319	91.5540767	178.7462657	100.585527	108.128792	74.3929328	102.807158
2562.53671	1504.80131	1376.35322	2839.32699	1168.97534	1241.19675	940.020648	1255.44856	1821.18439	604.161238	1391.37416	1898.86642	524.634849	1790.61157	367.860478	494.087811
756.590374	455.620636	398.156767	813.840719	335.888515	352.152885	266.546008	363.234737	499.42688	166.002244	394.174039	559.177791	148.269018	485.523508	106.453878	147.763495
250.082286	247.431845	246.33363	303.513812	315.357595	254.209814	227.016079	232.896891	223.999137	345.033237	239.601193	306.457668	233.942531	218.796126	205.296672	242.639558

Creating .csv files for each sample

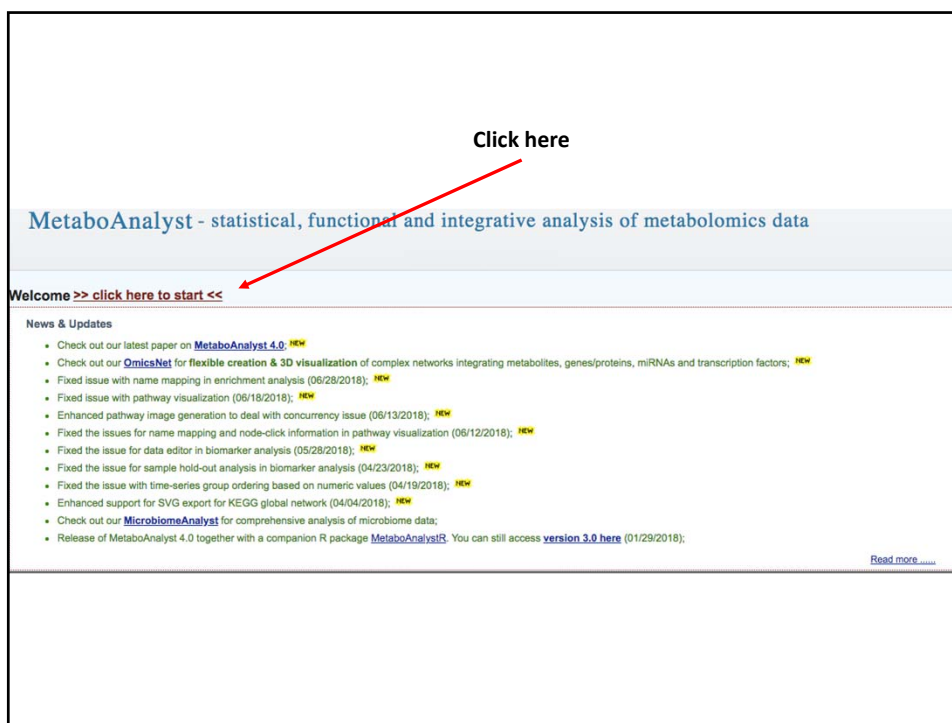
mzmed	rtmed	negmode_10
288.986526	5.0106	852.119488
341.8982	5.01165833	104.996374
166.017914	5.02245	1099.16397
310.968073	5.02675833	562.947757
210.008057	5.03755	1084.30719
738.992841	5.06338333	49.2257086
75.0101505	5.08216667	969.736874
326.94425	5.147175	339.370105
348.924709	5.17636667	156.499148
224.910626	5.672625	103.781657
208.937141	5.67875	131.715145
347.07202	5.72104167	306.07823
711.219381	5.73613333	2557.60796
126.944387	5.76803333	52.0865984
312.941373	5.77435833	112.212582
755.209423	5.78261667	2562.53671
756.213262	5.786375	756.590374
216.938459	5.83116667	250.082286

- 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>

Click here



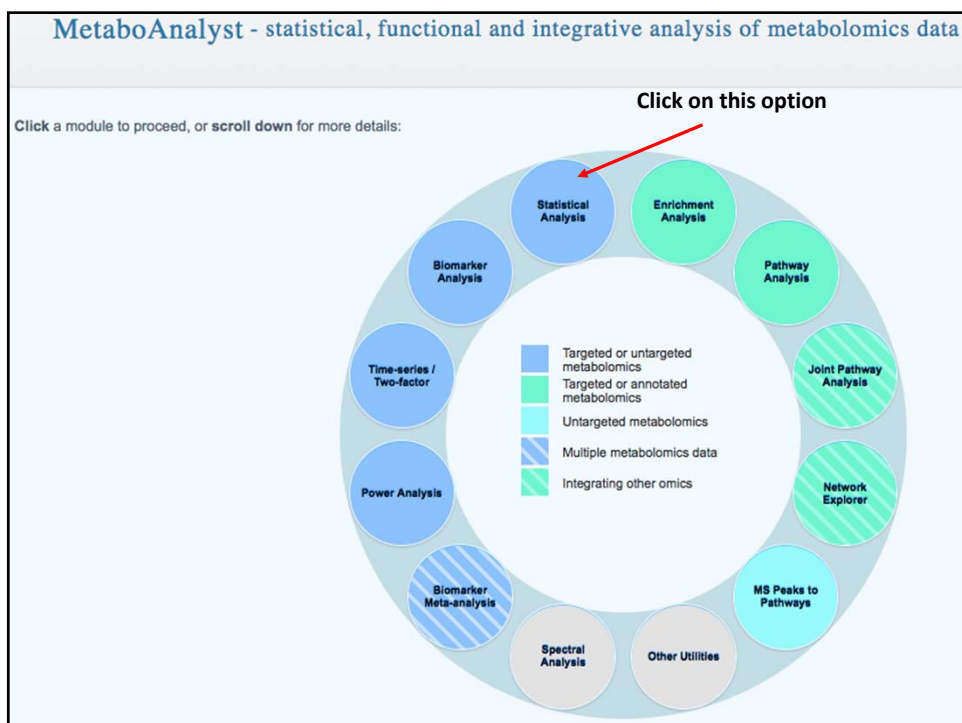
MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Welcome >> [click here to start](#) <<

News & Updates

- Check out our latest paper on [MetaboAnalyst 4.0](#); **NEW**
- Check out our [OmicsNet](#) for flexible creation & 3D visualization of complex networks integrating metabolites, genes/proteins, miRNAs and transcription factors; **NEW**
- Fixed issue with name mapping in enrichment analysis (06/28/2018); **NEW**
- Fixed issue with pathway visualization (06/18/2018); **NEW**
- Enhanced pathway image generation to deal with concurrency issue (06/13/2018); **NEW**
- Fixed the issues for name mapping and node-click information in pathway visualization (06/12/2018); **NEW**
- Fixed the issue for data editor in biomarker analysis (05/28/2018); **NEW**
- Fixed the issue for sample hold-out analysis in biomarker analysis (04/23/2018); **NEW**
- Fixed the issue with time-series group ordering based on numeric values (04/19/2018); **NEW**
- Enhanced support for SVG export for KEGG global network (04/04/2018); **NEW**
- Check out our [MicrobiomeAnalyst](#) for comprehensive analysis of microbiome data;
- Release of MetaboAnalyst 4.0 together with a companion R package [MetaboAnalystR](#). You can still access [version 3.0 here](#) (01/29/2018);

[Read more.....](#)



Uploading data to Metaboanalyst

Tab-delimited text (.txt) or comma-separated values (.csv) file:

Data Type: Concentrations Spectral bins Peak intensity table

Format:

Data File: No file chosen

Zipped Files (.zip):

Data Type: NMR peak list MS peak list MS spectra

Data File: Workshop_neg.zip

Pair File: No file chosen

Select MS peak list option and then load the .zip file
Submit the job to MetaboAnalyst

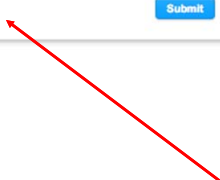
MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

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):

Retention time tolerance:



reset these to 0.001 and 0.005, respectively

MS peak processing information

The uploaded files are peak lists and intensities data.

A total of 20 samples were found.

These samples contain a total of 61220 peaks,

with an average of 3061 peaks per sample

A total of 3061 peak groups were formed.

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

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

Data processing information:

Checking data content ...passed

The uploaded files are peak lists and intensities data.

A total of 20 samples were found.

These samples contain a total of 61220 peaks,
with an average of 3061 peaks per sample

Samples are not paired.

2 groups were detected in samples.

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](#)

Data Filtering:

The purpose of the data filtering is to identify and remove variables that are unlikely to be of use when modeling the data. No phenotype information are used in the filtering process, so the result can be used with any downstream analysis. This step is strongly recommended for untargeted metabolomics datasets (i.e. spectral binning data, peak lists) with large number of variables, many of them are from baseline noises. Filtering can usually improve the results. For details, please refer to the paper by [Hackstadt, et al.](#)

Non-informative variables can be characterized in three groups: 1) variables of **very small values** (close to baseline or detection limit) - these variables can be detected using mean or median; 2) variables that are **near-constant values** throughout the experiment conditions (housekeeping or homeostasis) - these variables can be detected using standard deviation (SD); or the robust estimate such as interquartile range (IQR); and 3) variables that show **low repeatability** - this can be measured using QC samples using the relative standard deviation (RSD = SD/mean). Features with high percent RSD should be removed from the subsequent analysis (the suggested threshold is 20% for LC-MS and 30% for GC-MS). For data filtering based on the first two categories, 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 4000 features. Over that, if you choose None, the IQR filter will still be applied. In addition, the maximum allowed number of variables is 8000. If over 8000 variables were left after filtering, only the top 8000 will be used in the subsequent analysis.

Filtering features if their RSDs are > 25 % in QC samples

- None (less than 5000 features)
- 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

Submit

Proceed

Data normalization

Sample normalization

- None
- Sample-specific normalization (i.e. weight, volume) [Specify](#)
- Normalization by sum
- Normalization by median
- Normalization by reference sample (PQN) [Specify](#)
- Normalization by a pooled sample from group [Specify](#)
- Normalization by reference feature [Specify](#)
- Quantile normalization

Normalization by sum is used for urine data

Data transformation and scaling

Data transformation

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

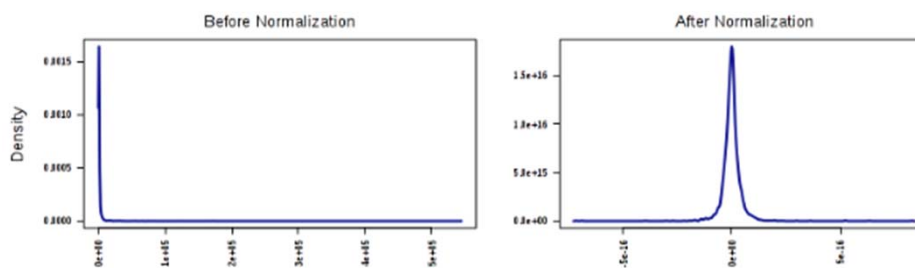
Data scaling

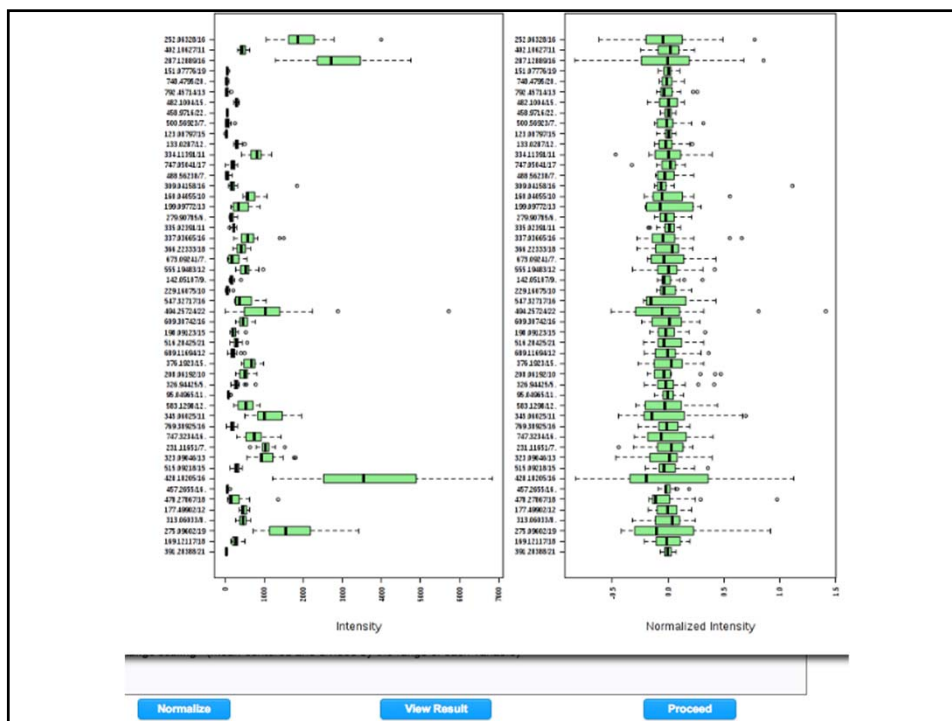
- None
- Mean centering (mean-centered only)
- Auto scaling (mean-centered and divided by the standard deviation of each variable)
- Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)
- Range scaling (mean-centered and divided by the range of each variable)

Normalize

View Result

Effect of normalization





Effect of normalization and scaling

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

Questions?