



MetaboAnalyst 6.0

-- a unified platform for metabolomics data processing,
analysis and interpretation

Dose Response Analysis

Module Overview



The module offers dose response analysis to quantify relationships between the concentration of a chemical and its effects on metabolomics profiles

- ✓ Providing necessary processing, normalization and differential expression analysis to select promising features;
- ✓ Perform curve fitting on those selected features against a suite of linear and non-linear models (currently 10 methods);
- ✓ For each feature, computing its benchmark doses (BMD) based on the selected model



Content

- ❑ 1. Introduction;
- ❑ 2. Choose the Module;
- ❑ 3. Data Preparation;
- ❑ 4. Dose-response Analysis;
 - 4.1 Data upload
 - 4.2 Data integrity check
 - 4.3 Data filtering and normalization
 - 4.4 Identification of features associated with dose
 - 4.5 Curve fitting
 - 4.5 Feature-level BMDs
- ❑ 5. Download Results.

1. Introduction

Background

- Dose-response analysis is commonly used in toxicology and pharmacology for understanding how varying concentrations of a chemical can impact a biological system.
- It plays a pivotal role in risk assessment of chemical exposures.
- A key output of dose-response analysis is the benchmark dose (BMD), the dose at which a chemical would cause a predetermined change in a physiological response.

Data Formats

- Both targeted and untargeted metabolomics data are accepted
- Dose-response experiment design includes a control group (dose = 0) and at least three different dose groups, typically with the same number of replicates in each group.
- The data required for processing should be a csv file, which has been correctly formatted.

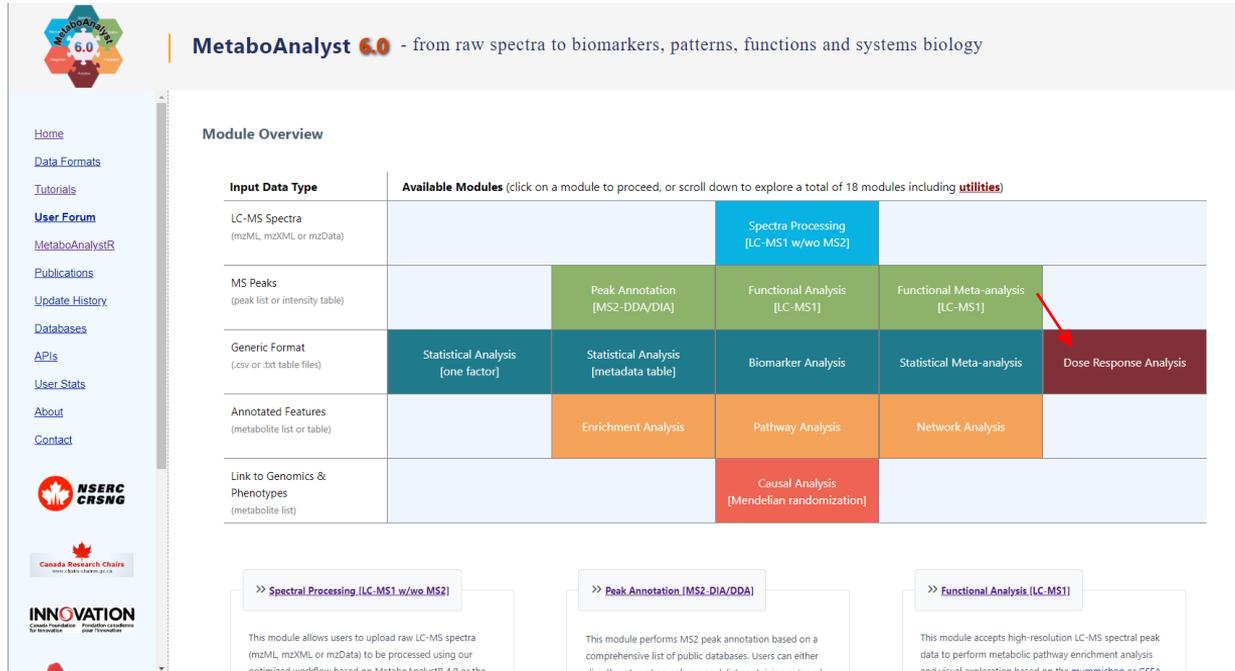
Expected Results

This module provides user comprehensive results on feature level dose response analysis:

- i. Dose response models and associated BMD summary table;
- ii. Visualizing the fitted dose response curves of individual features;
- iii. Estimated metabolomic-level point of departure (mPOD)

2. Choose the Module

Go to MetaboAnalyst (<https://www.metaboanalyst.ca>), and select the module



MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)				
LC-MS Spectra (mzML, mzXML or mzData)			Spectra Processing [LC-MS1 w/wo MS2]		
MS Peaks (peak list or intensity table)		Peak Annotation [MS2-DDA/DIA]	Functional Analysis [LC-MS1]	Functional Meta-analysis [LC-MS1]	
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadate table]	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis
Annotated Features (metabolite list or table)		Enrichment Analysis	Pathway Analysis	Network Analysis	
Link to Genomics & Phenotypes (metabolite list)			Causal Analysis [Mendelian randomization]		

>> Spectra Processing [LC-MS1 w/wo MS2]

This module allows users to upload raw LC-MS spectra (mzML, mzXML or mzData) to be processed using our optimized workflow based on MetaboAnalyst 4.0 or the

>> Peak Annotation [MS2-DIA/DIA]

This module performs MS2 peak annotation based on a comprehensive list of public databases. Users can either directly enter a functional peak list containing m/z and

>> Functional Analysis [LC-MS1]

This module accepts high-resolution LC-MS spectral peak data to perform metabolic pathway enrichment analysis and visual evaluation based on the **Metabolic** or **CCSA**

Navigation Sidebar: Home, Data Formats, Tutorials, User Forum, MetaboAnalystR, Publications, Update History, Databases, APIs, User Stats, About, Contact.

Logos: NSERC CRSNG, Canada Research Chairs, INNOVATION.

3. Data preparation

It is noted that this data table can be formatted as transposed with name and dose values located in the 1st and 2nd column, respectively.

Sample Name (1st row)

Dose values of different sample (2nd row)

Feature intensities or responses. The features can be metabolic features or compounds from metabolomics.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	name	092816_RPLC_F	092816_RPLC_POS_50uM3											
2	Dose	0	0	0	0	10	10	200	200	200	200	50	50	0
3	M450T1918	203995.3683	217563.605	206360.6475	49882.62177	52102.0091	36055.98929	17470.06027	20520.12066	13559.56834	32409.96322	27836.25517	31369.52025	0
4	M580T1785	204.9859732	202.8237309	205.9723909	0	0	0	5839.753433	6394.975863	449.602063	1220.134486	879.8721626	0	0
5	M444T1420	29957.16125	29417.82161	30549.57536	5098.684784	2738.782886	3142.322513	3534.510789	2074.46172	1949.228619	3583.74107	2152.024487	3910.198024	0
6	M800T1627	10575.86497	10706.35279	9986.178482	351.509963	549.2744133	324.2051702	525.5504516	0	1478.653961	252.9781982	0	517.8135026	0
7	M308T1287	889.3524437	1091.860312	641.4121682	332.341983	3565.742856	3215.614653	32194.58278	31819.46037	34294.52224	4691.3703	4713.492538	8338.712113	0
8	M389T1183	117296.2056	133721.1368	113912.7402	22334.43181	3357.710391	18709.08761	13409.74373	2576.006445	11826.19853	7608.643396	4972.328562	5549.895598	0
9	M387T1511	150364.5402	126302.0331	136395.3748	14270.10268	10222.55545	28123.90686	20607.16905	19444.7336	52866.21135	3671.167955	16204.65804	22909.92259	0
10	M190T1223	9138.171288	9815.61261	9655.415637	5786.151917	6507.735962	5950.810303	11254.9024	8348.940084	8534.317535	8109.195221	8135.810486	11047.65009	0
11	M232T388	19061465.42	16490090.83	17446301.05	7727540.716	5891152.241	5709789.886	51959685.13	5932085.813	4535381.245	6338732.865	6475621.299	4414371.818	0
12	M565T1750	1003.198981	339.3318538	488.1213742	3697.674377	4239.20054	3512.164735	100336.4438	92970.48834	90220.45298	22806.02251	21969.01257	18739.81747	0
13	M462T1193	5563.977581	5382.015211	5405.183925	3470.20892	3482.408521	3790.520871	4827.91632	3515.38382	4189.957398	5130.212695	2352.086731	4439.427197	0
14	M452T1746	190686.929	215278.4454	178645.5976	62224.4245	30028.18797	53217.19754	18833.59055	18597.40397	11090.91621	26187.05963	25058.90052	17828.65056	0
15	M260T698	331091.2717	366432.321	343185.6728	44331.19415	64350.08857	104085.9975	33037.93377	70971.49592	64702.78681	73346.69713	49456.81272	54889.41357	0
16	M341T1626	424445.1727	365450.4113	335328.3321	122126.9206	95856.69608	58613.6839	102653.4532	96703.99084	92262.19953	98887.66677	96018.63663	95368.88071	0
17	M343T1288	3019.144429	4356.646656	3618.879196	10690.64772	9238.541061	9312.61418	418700.2064	481260.0438	343238.2183	68340.67153	46418.23047	90616.38382	0
18	M705T1892	2878.027802	377.1384973	303.8461908	17950.30013	19347.83247	15197.24066	27661.73861	32279.3809	13233.7861	47580.40577	45807.27924	32130.30244	0
19	M387T51_2	8556.400767	9677.98932	9517.436563	3394.704202	3553.410391	3991.398228	4037.522973	4385.57024	3544.127228	12908.05068	2778.56977	4557.05195	0
20	M444T1576	207302.1336	203098.1947	171184.3617	33104.63601	24336.61429	11907.24843	22771.87396	23089.3201	12191.42322	11642.10544	21079.44431	8056.315441	0
21	M310T1288	4430.953105	5779.200848	4513.942339	11136.95004	9573.622561	10549.26244	87175.26532	94768.51664	68685.55002	29203.52311	22266.32695	20721.05407	0
22	M400T1703	163351.8001	156146.3366	132747.0125	33587.57236	15716.20143	27522.42721	18993.13521	27731.49887	22402.65529	35711.44345	23292.04017	29170.81541	0
23	M440T1810	243920.8999	254981.6229	258965.8579	47362.18168	4442.885808	35102.23311	26925.76894	26022.89148	15273.0098	26326.44813	30986.33317	23952.58071	0
24	M158T951	17639.72294	17608.00915	14077.67288	5546.114102	3328.215983	2275.329763	22448.9819	18194.46	14341.4574	16671.26928	16004.73966	14866.10829	0
25	M479T634	1402.726791	2488.595686	1775.597764	5132.739076	4880.473455	5838.991287	3752.600671	3447.639308	4051.065815	3508.146751	3848.700346	4298.524914	0
26	M801T1627	4639.983871	5731.236819	5756.832029	515.7837901	0	329.4845928	0	250.4057955	271.8714018	581.4590297	204.707656	0	0
27	M847T617_1	5811.849485	6168.633802	6735.179887	3615.728367	3964.664256	3027.758753	7206.93015	5707.254481	6089.05068	9563.067588	3647.954126	12761.06029	0
28	M451T1915	49492.57354	60622.94533	60461.73865	12810.68054	17047.88246	10582.75831	12387.35161	9179.382343	5714.845133	11296.06038	12626.55554	9186.650098	0
29	M931T1065	23362.11643	26409.20213	25252.11723	17633.58758	16848.99484	14953.07643	23769.55259	19205.58039	18241.65102	14768.24595	20030.46819	13797.8082	0
30	M198T745	201662.6442	204155.7142	207840.2078	167425.022	176840.0131	176084.2509	209636.0573	182760.9882	145357.7997	187671.9424	186492.6251	181294.1752	0
31	M371T1301	197021.0325	233458.4916	257888.7775	69213.29601	43327.5088	34178.42046	43034.80947	46802.0047	28312.71441	17801.81957	39573.7525	425603.27148	0
32	M435T1777	7203.383322	3580.267458	17407.228	68145.30004	60084.40634	49694.88459	94468.90044	16383.35597	61949.17847	13027.42092	64380.58974	62978.96722	0
33	M1707T45_2	22957.39003	21640.16169	23705.0306	17342.70086	17033.45302	15222.88931	23644.44526	21795.47393	14442.76924	21287.64035	19183.02961	18815.77927	0
34	M233T398	2101305.007	2540890.938	2492390.743	979341.7046	746908.937	808830.5112	724367.661	802539.3374	447634.3992	677624.8805	773190.0964	685191.2547	0
35	M170T1339	113349.7512	114809.4462	112140.7973	97161.76538	92845.90878	91395.24534	102067.2047	99602.10678	84654.17018	91176.01153	99770.2062	99742.56184	0
36	M450T603	29236.31648	24759.12208	34031.77227	53971.75213	58768.22185	57202.07828	57158.03834	42529.97602	84518.27194	62659.08083	47793.7869	99995.55272	0
37	M719T1420	3087.803497	3798.13979	3512.531245	1477.076739	1535.252312	1990.748262	2868.5357	7205.186493	21692.09023	4628.497288	25755.03495	1327.780263	0
38	M568T1286	2069.366317	1785.620443	2080.050886	2997.087927	3363.669377	3482.552735	12022.18239	15446.76546	8174.165619	5691.812061	3309.97418	3112.458552	0
39	M185T890	15137.43575	15845.64485	15456.91678	12639.04632	13868.54642	13176.86502	21491.97794	25222.73668	13269.18854	18692.71656	14547.91079	14875.89837	0

4.1 Dose-response data upload

Please upload your data (.csv or .txt)

Dose-response analysis was developed by the field of toxicology to identify the concentration at which a biological assay responds to chemical exposure. Dose-response experimental designs typically include a control group (dose = 0) and at least three different dose groups, typically with the same number of replicates in each group. To perform dose-response analysis using metabolomics data, three basic steps are involved:

1. Identify potential features of interest showing a relationship with dose;
2. Perform dose-response curve fitting against a suite of linear and non-linear models;
3. Determine the concentration at which the fitted curve departs from the values in the control group (i.e. benchmark dose or BMD).

Please refer to [Thomas et al. 2013](#) and [Yao et al. 2020](#) for more background information.

Data Type: Concentrations Spectral bins Peak intensities

Format:

Data File:

Try our example data

Data	Description
Dataset	Test example data to study dose-response effect using LC-MS untargeted metabolomics. BT549 breast cancer cells were treated with etomoxir concentrations that spanned 10-200 μM . Doses: 0, 10, 50, and 200 μM with three replicates at each dose (details).

At least three different dose groups with same number of replicates are required

4.2 Data integrity check

Data Integrity Check:

- Checking sample names - spaces will be replaced with underscore, and special characters will be removed;
- Checking the class labels - at least three replicates are required in each class.
- The data (except class labels) must not contain non-numeric values.
- If the samples are paired, the pair labels must conform to the specified format.
- The presence of missing values or features with constant values (i.e. all zeros).

Data processing information:

Checking data content ...passed.

Samples are in columns and features in rows.

The uploaded file is in comma separated values (.csv) format.

The uploaded data file contains 12 (samples) by 26920 (peaks(mz/rt)) data matrix.

Samples are not paired.

4 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, missing values will be replaced by 1/5 of min positive values of their corresponding variables

Click the **Proceed** button if you accept the default practice;

Or click the **Missing Values** button to use other methods.

Edit Groups

Missing Values

▶ Proceed

4.3 Data filtering and normalization

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 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); 2) 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); and 3) variables of **very small values** (close to baseline or detection limit) - these variables can be detected using mean or median.

For data filtering based on the last two categories, the default parameters follow the empirical rules: 1) Less than 250 variables: 5% will be filtered; 2) Between 250 - 500 variables: 10% will be filtered; 3) Between 500 - 1000 variables: 25% will be filtered; and 4) Over 1000 variables: 40% will be filtered. You can turn off data filtering by dragging the slider to adjust the percentage to filter out to be 0, when your data contain less than 5000 features (or 2500 for power analysis) to control computing time on our server.

Reliability filter:	<input type="checkbox"/> Filtering features based on technical repeatability QC samples	RSDs greater than: <input type="range" value="25"/> 25%
Variance filter:	<input checked="" type="radio"/> Interquartile range (IQR) <input type="radio"/> Standard deviation (SD) <input type="radio"/> Median absolute deviation (MAD) <input type="radio"/> Relative standard deviation (RSD = SD/mean) <input type="radio"/> Non-parametric relative standard deviation (MAD/median)	Percentage to filter out: <input type="range" value="40"/> 40%
Abundance filter:	<input checked="" type="radio"/> Mean intensity value <input type="radio"/> Median intensity value	Percentage to filter out: <input type="range" value="0"/> 0%

Normalization Overview:

The normalization procedures are grouped into three categories. You can use one or combine them to achieve better results.

- Sample normalization is for general-purpose adjustment for systematic differences among samples;
- Data transformation applies a mathematical transformation on individual values themselves. A simple mathematical approach is used to deal with negative values in log
- Data scaling adjusts each variable/feature by a scaling factor computed based on the dispersion of the variable.

Sample normalization

- None
- Sample-specific normalization (i.e. weight, volume) [Specify](#)
- Normalization by sum
- Normalization by median
- Normalization by a reference sample (PQN) [Specify](#)
- Normalization by a pooled sample from group (group PQN) [Specify](#)
- Normalization by reference feature [Specify](#)
- Quantile normalization (suggested only for > 1000 features)

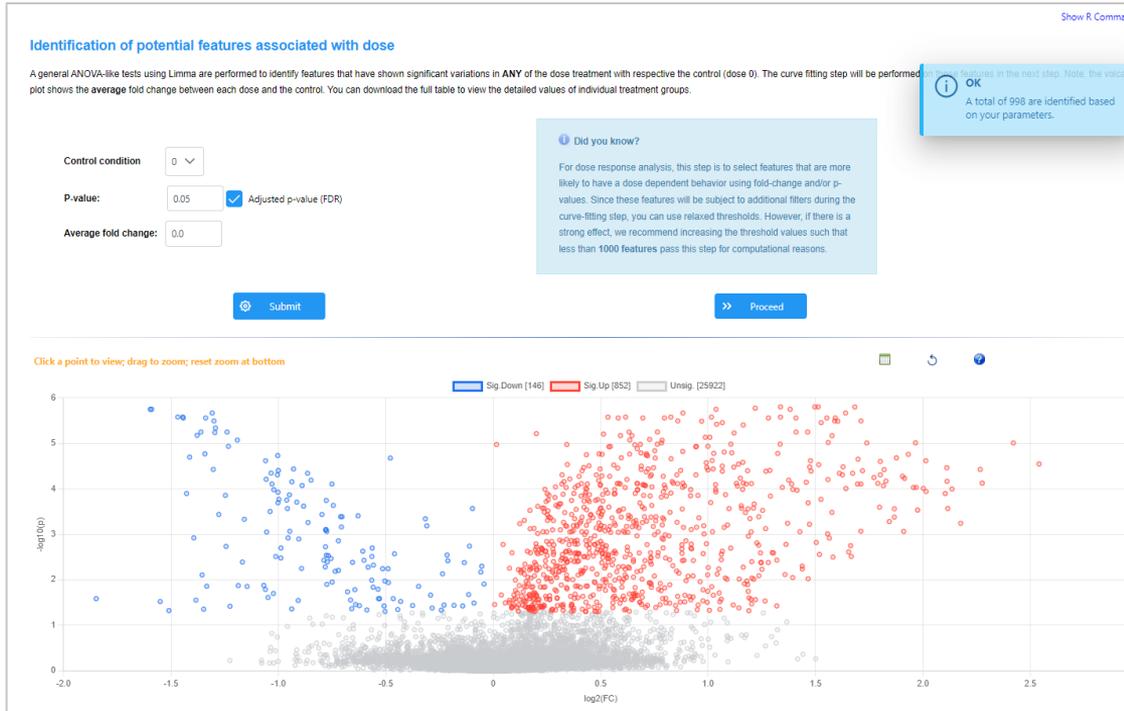
Data transformation

- None
- Log transformation (base 10)
- Square root transformation (square root of data values)
- Cube root transformation (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 the standard deviation of each variable)
- Range scaling (mean-centered and divided by the range of each variable)

4.4 Identification of features associated with dose



4.5 Curve fitting

Computing BMD requires a mathematical model describing the dose response curve – a process called curve fitting

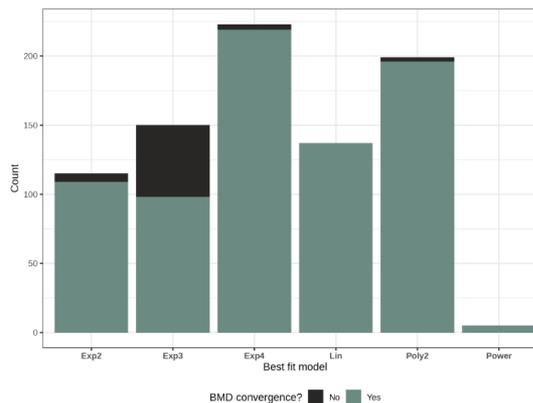
Perform curve fitting to calculate feature-level benchmark doses (BMDs)

To calculate compound-level BMDs, up to 10 statistical models are fit to the expression of each compound. Any model fits with a poor fit are filtered out, and then the best fitting model is chosen based on AIC. The selected fit is used to compute the BMD. We recommend selecting all statistical models except for Poly3 and Poly4, which should only be used if you expect a non-monotonic response. These higher order polynomials should be used with caution since they sometimes have unpredictable behavior, especially for dose-response experiments with a log-scale dosing scheme. **Note**, this process is computational intensive, a maximum of **1000 features** can be used for curve fitting.

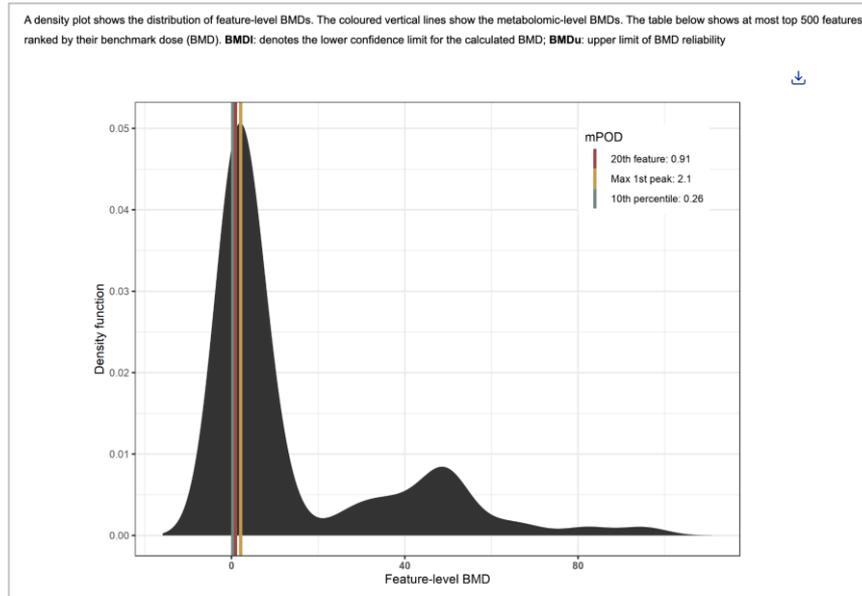
Fit models	<input checked="" type="checkbox"/> Exp2	<input checked="" type="checkbox"/> Exp3	<input checked="" type="checkbox"/> Exp4	<input type="checkbox"/> Exp5	<input checked="" type="checkbox"/> Linear
	<input checked="" type="checkbox"/> Poly2	<input type="checkbox"/> Poly3	<input type="checkbox"/> Poly4	<input type="checkbox"/> Hill	<input checked="" type="checkbox"/> Power
Calculate BMDs	Lack-of-fit p-value:	<input type="text" value="0.10"/>	<input <="" td="" type="button" value="?"/>		
	BMR factor:	<input type="text" value="1.00"/>	<input <="" td="" type="button" value="?"/>		
	Control abundance:	<input type="text" value="Mean of control samples"/>	<input <="" td="" type="button" value="?"/>		

Features with fitted models: **829** Features with BMDs: **764**

This bar plot shows how many times each model was found to be the best-fitting model for a compound. Each bar shows the number of model fits that have a BMD (blue) or not (dark grey). In most cases, the reason that a model fit does not have a BMD is that the feature abundance never exceeds the standard deviation of the control. You can view the detailed curve fitting results of each feature in the next page.



4.6 From feature-level BMDs to metabolomics-POD



Using BMD distribution to estimate mPOD

Result Table

(1 of 34) << < 1 2 3 4 5 6 7 8 9 10 > >> 15

ID ↑↓	P-val ↑↓	BMDl ↑↓	BMD ↓↓	BMDu ↑↓	Model name ↑↓	View
M401T1625	0.42	5.4E-7	1.1E-6	1.9E-5	Exp3	View
M741T1899	0.71	3.0E-4	3.3E-4	0.01	Exp3	View
M341T1288_1	0.7	0.0012	0.0068	0.027	Exp3	View
M167T1288_2	0.13	0.0019	0.0098	0.032	Exp3	View
M718T1936	0.81	0.0069	0.033	0.22	Exp3	View
M342T1288	0.15	0.011	0.048	0.13	Exp3	View
M723T1900	0.66	0.0081	0.058	0.32	Exp3	View
M153T1288_1	0.43	0.015	0.066	0.18	Exp3	View
M169T1288	0.17	0.014	0.068	0.2	Exp3	View
M342T1287	0.73	0.016	0.083	0.27	Exp3	View
M310T1288	0.49	0.021	0.099	0.26	Exp3	View

Detailed BMD summary

5. Download results

Download Results & Start New Journey

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button. Finally, you can continue to explore other compatible modules using the **Start New Journey** tab.

The screenshot shows a web interface with two tabs: 'Results Download' (active) and 'Start New Journey'. Below the tabs is a table of downloadable files. At the bottom of the table area is a blue 'Logout' button.

Download.zip	data_processed.csv
Rhistory.R	data_original.csv
dr_barplot_0_dpi72.png	dr_histogram_0_dpi72.png
M676T1927_0_summary_dpi72.png	snorm_0_dpi72.png
curvefit_detailed_table.csv	limma_gig_features.csv
M676T1927_1_summary_dpi72.png	data_normalized.csv
raw_dataview.csv	norm_0_dpi72.png
limma_restable.csv	dose_response_limma_all.csv
M675T1927_4_summary_dpi72.png	M675T1927_3_summary_dpi72.png
M676T1927_2_summary_dpi72.png	

Logout

All results can be downloaded here.



In summary

If you have any questions, please read/post into OmicsForum (www.omicsforum.ca)

Or contact us:

[zhiqiang.pang\[at\]xialab.ca](mailto:zhiqiang.pang@xialab.ca)

[jeff.xia\[at\]xialab.ca](mailto:jeff.xia@xialab.ca)

- Dose response data require special design – at least three doses and three replicates per dose are required
- Three basic steps are involved in dose response analysis require
 1. Identify potential features of interest showing a relationship with dose - this is achieved using regular differential expression analysis;
 2. Perform dose-response curve fitting against a suite of linear and non-linear models;
 3. Determine the concentration at which the fitted curve departs from the values in the control group (i.e. benchmark dose or BMD).
- The process can be computing intensive when there are a large number of features (i.e. LC-MS peaks from untargeted metabolomics)