

BMDx Reference Manual

with Sample Data Analysis

Contents

About BMDx.....	2
Data Description.....	2
Installation and execution	3
Workflow Interface.....	3
Input Description.....	3
Phenotype Specification.....	4
Load Pheotype	4
Select Phenotype File.....	4
Phenotype Preview	4
Configure Variable R Format.....	5
Specify Sample ID, Dose and Time Point Variables.....	5
Import Phenotype	5
Phenotype View.....	5
Expression matrix	6
Load Expression Matrix.....	6
Anova Filtering.....	6
Model fitting and BMD computation	8
Parameter Selection	8
BMR	8
Model selection.....	8
Results investigation.....	10
Fit of the model.....	11
Comparing results between time points.....	11
Compare Experiments.....	15
Functional Enrichment.....	16
Enrichment parameters	16
Enrichment results.....	18
Heatmap.....	19
Cluster Bubble Plot	20
Mean BMD for Time Point.....	20
Gene BMD in Pathway	21
Pathways Table.....	22
Heatmap Genes.....	23
Model descriptions.....	24
References.....	25

About BMDx

BMDx is an R-Shiny application created for easy benchmark dose (BMD) analysis on omics data across multiple time points and experiments. The tool guides the user through multiple steps starting from an analysis of variance, through BMD computing all the way to a functional enrichment of the dose-dependent genes. BMDx not only allows the user to compare the results between multiple time points, but also multiple experiments at once. Results along the way are visualised as several types of plots and the output can be downloaded as Excel files at multiple steps of the analysis.

The benchmark dose (BMD) is the dose or concentration of a substance that corresponds to a specified level of response above or below that observed in a control or background population. The specified level of response within this definition is referred to as the benchmark response (BMR), while the statistical lower confidence bound of the BMD (referred to as BMDL) and the statistical upper confidence bound on the BMD (BMDU) have been typically used by regulatory agencies to set safe levels of exposure.

BMD modelling involves fitting the experimental data, in this case, the gene expression values, to a selection of mathematical models, such as linear, second- or third-degree polynomial, an exponential model, hill model, asymptotic regression model, and Michaelis-Menten model. The best model is selected by using a goodness of fit criteria, such as the Akaike information and the goodness-of-fit p-value. A predefined response level of interest, the BMR, is identified and the optimal model is used to predict the corresponding dose (BMD) (Abraham et al. 2012). Moreover, the European Food Safety Authority (EFSA) suggest reporting both the lower and upper 95% confidence limit on the BMD called BMDL and BMDU respectively (EFSA Scientific Committee et al. 2017). The selection of models available in BMDx are presented on page 24 of this document with model descriptions included.

In this manual, we provide a detailed step-by-step guide to using BMDx. As a result of the analysis, the user will retrieve the results of the analysis of variance, lists of dose-dependent genes with BMD, BMDL, BMDU and IC50/EC50 values, the model fitted for each gene and its corresponding lack-of-fit p-value. Moreover, the results of the functional enrichment can be downloaded providing a comprehensive view of the dose-dependent genes in the experiment.

Data Description

To demonstrate the use and effectiveness of our tool, we analysed gene expression data obtained from the Open TG-GATEs database (Igarashi et al. 2015). Out of the 170 compounds available in the database, we selected gene expression data from the liver of rats exposed to either Omeprazole or Pirinixic acid (WY-14643). Omeprazole is a commonly used proton-pump inhibitor used to treat gastroesophageal reflux disease, while Pirinixic acid is a peroxisome proliferator linked to liver carcinogenesis (Woods et al. 2007). Both datasets include 48 samples as three doses (100, 300 and 1000 µg for Omeprazole and 10, 30 and 100 µg for Pirinixic acid) and their corresponding controls at four time points (4, 8, 15 and 29 days) were included in the experiment as triplicates.

Raw data were imported into R v. 3.4 by using the justRMA function from the Bioconductor utilities (Irizarry et al. 2003) to annotate the probes to Ensembl genes (rat2302rnensgcdf v. 22.0.0

annotation file obtained from <http://brainarray.mbnl.med.umich.edu/>) and to quantile normalise the data. The experimental batch effect due to technical variables was estimated and removed using the ComBat algorithm implemented in the sva package (Leek et al. 2014). Linear models followed by eBayes pairwise comparisons (Ritchie et al. 2015) were performed to compute the log fold-change for each gene in all of the drug–control pairs. Genes with fold change $> |1.5|$ and p-value < 0.05 were determined differentially expressed and used in this analysis. Finally, the Ensembl gene names were converted to official GeneSymbols.

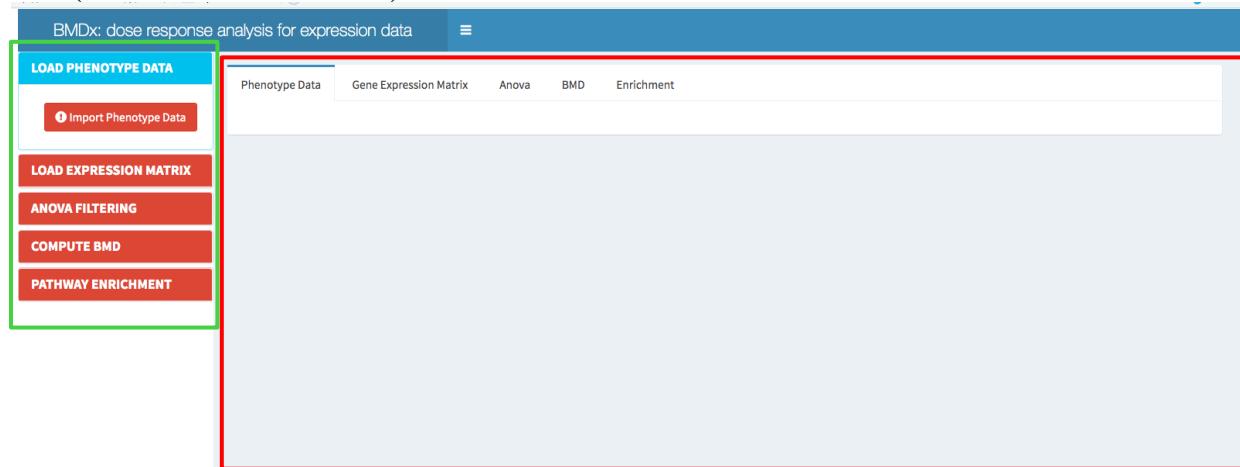
Data used as an example in this document are available on GitHub (<https://github.com/Greco-Lab/BMDx>).

Installation and execution

Instructions on how to install BMDx and its dependencies and how to launch the BMDx tool are available online at <https://github.com/Greco-Lab/BMDx>.

Workflow Interface

The workflow interface layout has a sidebar with input controls to configure and execute various steps (marked with green outline) and the output of the steps are visualized from the main display area (marked with red outline).



Input Description

BMDx takes as an input a phenotype file and an expression matrix, both provided as an Excel spreadsheet (xlsx). **If multiple experiments are included, both files must contain separate sheets for each experiment in corresponding orders.** Specific instructions for the file structures are provided below and example files are available on GitHub (<https://github.com/Greco-Lab/BMDx>).

Phenotype Specification

The phenotype file is an Excel file containing separate sheets for each experiment. Each sheet contains information about the samples used in the specific experiment. In particular, the BMDx tool requires the spreadsheets to have at least three columns that specify the following characteristics: 1) Unique sample IDs (here BARCODE) corresponding to the column names in the expression matrix, 2) the dose and 3) the time points (here SACRIFICE_PERIOD) included in the experiment. **Each sheet must have the columns (sample ID, dose and time point) in the same positions.**

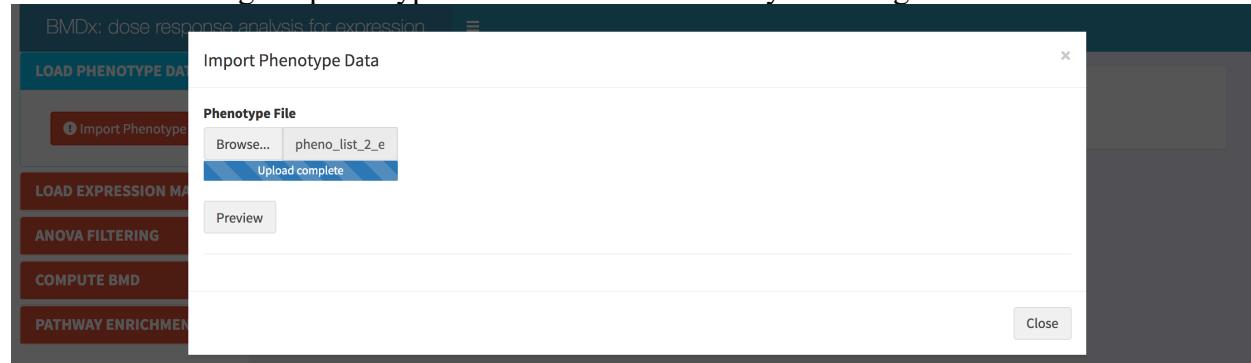
	BARCODE	DOSE	DOSE_LEVEL	SACRIFICE_PERIOD
1				
2	12800	ID_003017698023	0	Control
3	12801	ID_003017698024	0	Control
4	12802	ID_003017698025	0	Control
5	12803	ID_003017699005	0	Control
6	12804	ID_003017699006	0	Control
7	12805	ID_003017699007	0	Control
8	12806	ID_003017667006	0	Control
9	12807	ID_003017667007	0	Control
10	12808	ID_003017667008	0	Control
11	12809	ID_003017667018	0	Control
12	12810	ID_003017667019	0	Control
13	12811	ID_003017667020	0	Control
14	12812	ID_003017698026	100	Low
15	12813	ID_003017698027	100	Low
16	12814	ID_003017698028	100	Low
17	12815	ID_003017699008	100	Low
18	12816	ID_003017699009	100	Low

Load Pheotype

A popup window containing controls to configure the phenotype file import is launched by clicking *Import Phenotype Data* on the sidebar.

Select Phenotype File

The file containing the phenotype information is selected by browsing the file directories.



Phenotype Preview

Preview of the phenotype file displays the columns from the first sheet in the phenotype file as variables. Each variable has an associated R class character, numeric, or integer and data representation type as factor or vector. Number of samples and variables are reported as text labels above the preview.

Configure Variable R Format

The user can change the default data representation type by double-clicking on the representative cell and selecting the alternative option (factor or vector).

Specify Sample ID, Dose and Time Point Variables

These variables are specified by the corresponding variable index from the phenotype preview.

The screenshot shows the 'Import Phenotype Data' window. At the top, there is a 'Phenotype File' section with a 'Browse...' button, a file path 'pheno_list_2_e', and a blue 'Upload complete' button. Below this, a 'Preview' button is shown next to 'Samples: 48' and 'Variables: 4'. The main area displays a table of variables:

Variable	Type	Class	Sample1	Sample2	Sample3	Sample4	Sample5	Sa
BARCODE [1]	factor	character	ID_003017698023	ID_003017698024	ID_003017698025	ID_003017699005	ID_003017699006	ID_003
DOSE [2]	vector	numeric	0	0	0	0	0	0
DOSE_LEVEL [3]	factor	character	Control	Control	Control	Control	Control	Contro
SACRIFICE_PERIOD [4]	vector	numeric	4	4	4	8	8	8

An arrow points to the 'DOSE' row with the text 'Double-click to change type'. Below the table, there are three dropdown menus: 'Sample ID Variable' (set to 'Variable 1'), 'Dose Variable' (set to 'Variable 2'), and 'Time Point Variable' (set to 'Variable 4'). A modal dialog box is open over these dropdowns, containing a list of variables: 'Variable 1', 'Variable 2', 'Variable 3', and 'Variable 4'. The 'Import' button is visible at the bottom right of the dialog, and a 'Close' button is at the bottom right of the main window.

Import Phenotype

Finally, click on *Import* on the right bottom corner of the graphical window to import the configured phenotype file.

Phenotype View

The first sheet of the imported phenotype file is displayed in the main display area in the main *Phenotype Data* tab.

Expression matrix

The expression matrix is an Excel file with a separate sheet for each experiment. Sample columns are named with unique sample IDs (green outline) that match the sample IDs provided in the phenotype file. Gene names (orange outline) must be provided in the first column of each spreadsheet, and each following column specifies the expression values for those genes in each individual sample. The order of the sheets must match the order of the sheets in the phenotype file. Sample data can be found at <https://github.com/Greco-Lab/BMDx>.

A	B	C	D	E	F	G	H	I	J
1	ID_003017698023	ID_003017698024	ID_003017698025	ID_003017699005	ID_003017699006	ID_003017699007	ID_003017667006	ID_003017667007	ID_0030176670
2	Gad1	6.41439102304723	6.46532837902453	6.15713257780585	6.38079246591397	6.4828549058064	6.56574785793586	6.68232330197956	6.806777297751916
3	Cbln1	6.49548074605586	6.24746145003168	6.27445075342446	6.41063178671925	6.40066644578383	6.49783392846416	6.4188260148178	6.52749758647256
4	Stear1	6.087251817076115	6.72392577452614	6.221594313863136	6.164695480114235	5.84907206426937	6.395905820546415	6.83780943848238	6.432777081306126
5	Hepb1	8.16244677988902	7.98135071582217	7.30078550598876	7.376645571455227	7.567176202452617	7.87981448716476	6.957982323672276	6.76357210565463
6	Tmcc2	6.30191849868735	6.08084669634946	6.12173437902055	6.226592404203876	6.15916599240876	6.443255446099016	6.06295206830075	6.223123474147916
7	Nuak2	6.8608188928228	6.70585968433676	6.84342702783655	6.7861884544676	6.928968581636846	6.844614485018226	6.52149337211356	6.290710073013646
8	Klhdc8a	6.7015179562915	6.49409288378762	6.655199987453	6.641529669323256	6.63796441320896	6.659216589753726	6.26284294953696	6.59582961239123
9	RGD1304622	6.15895311078875	6.76185279246275	6.111183693594436	6.43082784238366	6.13054145327246	6.40971712170263	6.52607578979618	6.4080867252917
10	Slc26a1	7.785527536219917	7.05396358247276	7.30007101657085	6.98963036419745	7.31341541974814	7.08178203218196	6.72938943982973	6.54395271550506
11	Cd82	6.564951056056526	6.44335816340407	6.336587350682326	6.504151793283716	6.501580165688226	6.56574785793586	6.1168555050355	5.969208479960475
12	Gak	6.673616405202446	6.42020620132176	6.306477789935585	6.855682825268856	6.277174073006016	6.39823165499895	5.834019401092825	5.65078620662586
13	Crp	7.880788577374217	7.665800531582947	7.70000806901788	7.36908120777646	7.80118607311785	8.10446795117725	7.37923434832878	6.908867406310877
14	Abhd8	6.598527764369376	6.26406219642696	5.949377724171716	6.491748777829386	6.21556731267876	6.268902060724315	5.99958289056345	6.2287378268976
15	Fcrf6	6.34498541593716	6.5749947045596	6.35004675732526	6.436576146074226	6.2553605040043656	6.70757134232426	6.501900754285856	6.609512144674736
16	Cpxl1	6.57155895561965	6.226489878541626	6.15541626596436	6.199614319772326	6.33638370282556	6.246140564981356	6.4956951653343316	6.24573326468416.37357962944

Load Expression Matrix

The expression matrix file is imported similarly by clicking on *Import Expression Matrix* on the left side panel of the graphical interface. Expression matrix is viewed in the *Gene Expression Matrix* tab.

Anova Filtering

As the first step of the analysis, the genes that show variability across different doses are identified by performing an ANOVA test for each gene. Clicking on *Anova Filtering* launches a graphical window that allows for the specification of the parameters used for ANOVA. Alternatively, this step can be skipped.

To run the ANOVA filtering, time points included in the analysis are specified (“All” set as default). When multiple experiments are included in the analysis, all time points are included automatically and no less can be selected. P-value for the analysis can be specified between nominal and FDR corrected, and the p-value threshold can be set from the drop menu.

	ID	ID_	Search:					
9007	ID_003017667006	ID_						
793586	6.68232330197958	6.						
284641	6.41888260148178	6.						
054641	5.83780943846238	6.						
Hebp1	8.16244677988902	7.98135071582213	7.30078550598878	7.37664557145522	7.56717620245261	7.87981448716478	6.95798232367227	6.
Tmcc2	6.30191849868739	6.08084669634948	6.12173437902059	6.22659240420387	6.15916599240876	6.4432554609901	6.06295206830075	6.
Nuak2	6.8608188928228	6.70585968433679	6.84342702783659	6.7861884544676	6.92896858163684	6.84461448501822	6.52149337211356	6.
Klhdc8a	6.70153179562915	6.49409288378762	6.6551999897453	6.64152966932325	6.63796441320896	6.65921658975372	6.26284294953698	6.
RGD1304622	6.15895311078878	5.76185279246279	6.11118369359443	6.43082784238366	6.13054145327246	6.40971712170263	6.52607578979618	6.
Sic26a1	7.78552753621991	7.05396358247276	7.30007101657089	6.98963036419745	7.31341541974814	7.08178203218198	6.72938943982973	6.
Cd82	6.56495105605652	6.44335816340407	6.33658735068232	6.50415179328371	6.50158016568822	6.56574785793586	6.11668555050335	5.

Once the ANOVA tests have been run, the results are displayed in the *Anova* tab. In particular, a table with the ANOVA p-value for every gene will be displayed for every time point in each experiment separately. The experiment under inspection can be changed from *Experiment* drop menu and the specific time point for which data is shown can be changed from the drop menu *Time Points*. Furthermore, a pie chart will show the percentage of genes surviving the ANOVA test for each time point. The results of the ANOVA test can be downloaded as one Excel file with multiple sheets, one for every time point at each experiment by clicking *Download*.

	ID	ID_	Search:					
9007	ID_003017667006	ID_						
793586	6.68232330197958	6.						
284641	6.41888260148178	6.						
054641	5.83780943846238	6.						
Hebp1	8.16244677988902	7.98135071582213	7.30078550598878	7.37664557145522	7.56717620245261	7.87981448716478	6.95798232367227	6.
Tmcc2	6.30191849868739	6.08084669634948	6.12173437902059	6.22659240420387	6.15916599240876	6.4432554609901	6.06295206830075	6.
Nuak2	6.8608188928228	6.70585968433679	6.84342702783659	6.7861884544676	6.92896858163684	6.84461448501822	6.52149337211356	6.
Klhdc8a	6.70153179562915	6.49409288378762	6.6551999897453	6.64152966932325	6.63796441320896	6.65921658975372	6.26284294953698	6.
RGD1304622	6.15895311078878	5.76185279246279	6.11118369359443	6.43082784238366	6.13054145327246	6.40971712170263	6.52607578979618	6.
Sic26a1	7.78552753621991	7.05396358247276	7.30007101657089	6.98963036419745	7.31341541974814	7.08178203218198	6.72938943982973	6.
Cd82	6.56495105605652	6.44335816340407	6.33658735068232	6.50415179328371	6.50158016568822	6.56574785793586	6.11668555050335	5.

Model fitting and BMD computation

Parameter Selection

Clicking on *Compute BMD* on the side panel launches a graphical window for the selection of parameters and models for BMD analysis. The user selects the models to be fitted, the response level, the lack-of-fit p-value threshold and an upper limit for the estimation of the dose (maximum dose).

BMR

As a default, the BMR is set to 1.349 multiplied by the standard deviation of the controls. As described in Thomas et al. (2007), a BMR of 1.349 is the amount required to shift the mean transcriptional response of the control distribution such that the treated distribution contains 11% in a single tail, i.e., a 10% increase over the assume background rate of response. The 10% value for the shift in the tail area of the distribution is standard for BMD analysis.

Model selection

For each gene, a list of models is computed, and for each fitting, a lack-of-fit p-value is provided. Models with a non-statistically relevant fitting (lack-of-fit p-value < 0.1) and predicted BMD value higher than the maximum dose are removed from the analyses. The optimal model is identified as the one with the minimum Akaike Information Criterion (AIC) value.

Select the BMD analysis setting section allows for the selection of models to be used for the analysis. The models can be selected from predefined sets (*All*, *Regulatory*, *Degree of Freedom*, *Custom*) or selected manually. *Regulatory* contains models used by the regulatory agencies, *Degree of Freedom* includes models with a degree of freedom smaller than $n_d - 1$, where n_d is the number of doses, while *Custom* allows for manual selection by clicking the models one at a time. Moreover, in the lower part of the compute BMD Value window, the user will find a description of the models available in the tool. Model descriptions can also be found on page 24 of this file.

Compute BMD Value

x

Maximum dose

1000

Lack-of-fit PValue Th:

0.1

Response Level

1.349

Select the BMD analysis setting

Custom

Models available

Linear

Quadratic

Cubic

Power2

Power3

Power4

Exponential

Hill05

Hill1

Hill2

Hill3

Hill4

Hill5

AR.2

AR.3

MM.2

MM.3

Click on the box to select and deselect a model

 Run BMD

Models Description

Linear Model

Polynomial Model (Quadratic/Cubic)

Power Model

Exponential Model

Hill Model

Asymptotic Regression

Michaelis-Menten Model

 Close

Descriptions of the models are viewed by clicking the name of the model.

Models Description

Linear Model

Polynomial Model (Quadratic/Cubic)

The formula for the polynomial model is

$$f(\text{dose}) = \beta_0 + \beta_1 \text{dose} + \beta_2 \text{dose}^2 + \dots + \beta_n \text{dose}^n$$

Here n is the degree of the polynomial. The user can choose between

$$n = 2, 3$$

Power Model

Exponential Model

Hill Model

Asymptotic Regression

Michaelis-Menten Model

Results investigation

The results of the BMD analysis for each experiment can be explored on the *BMD* tab one time point at a time or different aspects of the results between time points can be visualised under *Compare TP* tab. Additionally, an UpSet plot representing the intersections between different doses and experiment can be visualised under *Compare Experiments* tab. The experiment under inspection can be changed from the *Experiment* drop menu.

On a gene level, BMD, BMDL and BMDU are calculated, as well as the IC50/EC50 value. The table also shows whether the expression of the gene is increasing or decreasing with dose, the optimal model and the lack-of-fit p-value of that model. Results can be downloaded as a single excel file with one sheet for each time point at each experiment by clicking *Download*.

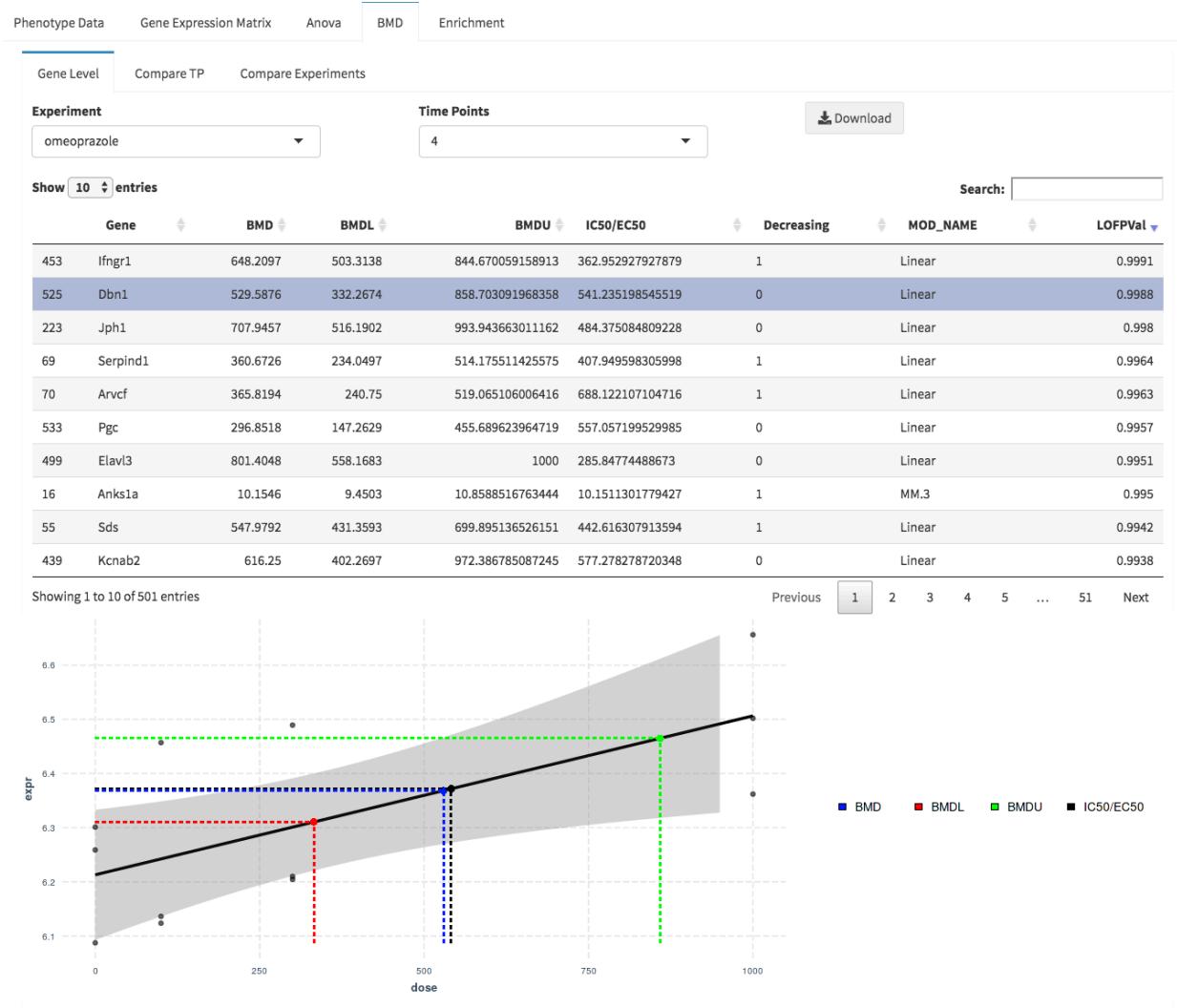
BMDx: dose response analysis for expression data									
LOAD PHENOTYPE DATA		LOAD EXPRESSION MATRIX		ANOVA		BMD		Enrichment	
COMPUTE BMD									
<input checked="" type="checkbox"/> Compute BMD									
PATHWAY ENRICHMENT									
Phenotype Data	Gene Expression Matrix	Anova	BMD	Enrichment					
Gene Level	Compare TP	Compare Experiments							
Experiment	omeoprazole		Time Points	4			<input type="button" value="Download"/>		
	omeoprazole	WY-14643					Search:		
			BMDL	BMDU	IC50/EC50	Decreasing	MOD_NAME	LOFPVal	
	All	All	All	All	All	All	All	All	
2	Nuak2	302.0141	112.6771	509.263747864883	589.238417444391	1	Linear	0.4792	
3	RGD1304622	526.8831	332.1938	849.323758607641	216.079843073897	0	Linear	0.4403	
4	Crp	644.8165	391.6501		1000	235.78320957154	1	Linear	0.7846
6	Fgf20	436.4087	201.8726	792.75777884359	550.840899134218	0	Linear	0.6282	
7	Ly86	848.5848	610.6169		1000	105.966695985808	1	Linear	0.628

Showing 1 to 5 of 501 entries

Previous 1 2 3 4 5 ... 101 Next

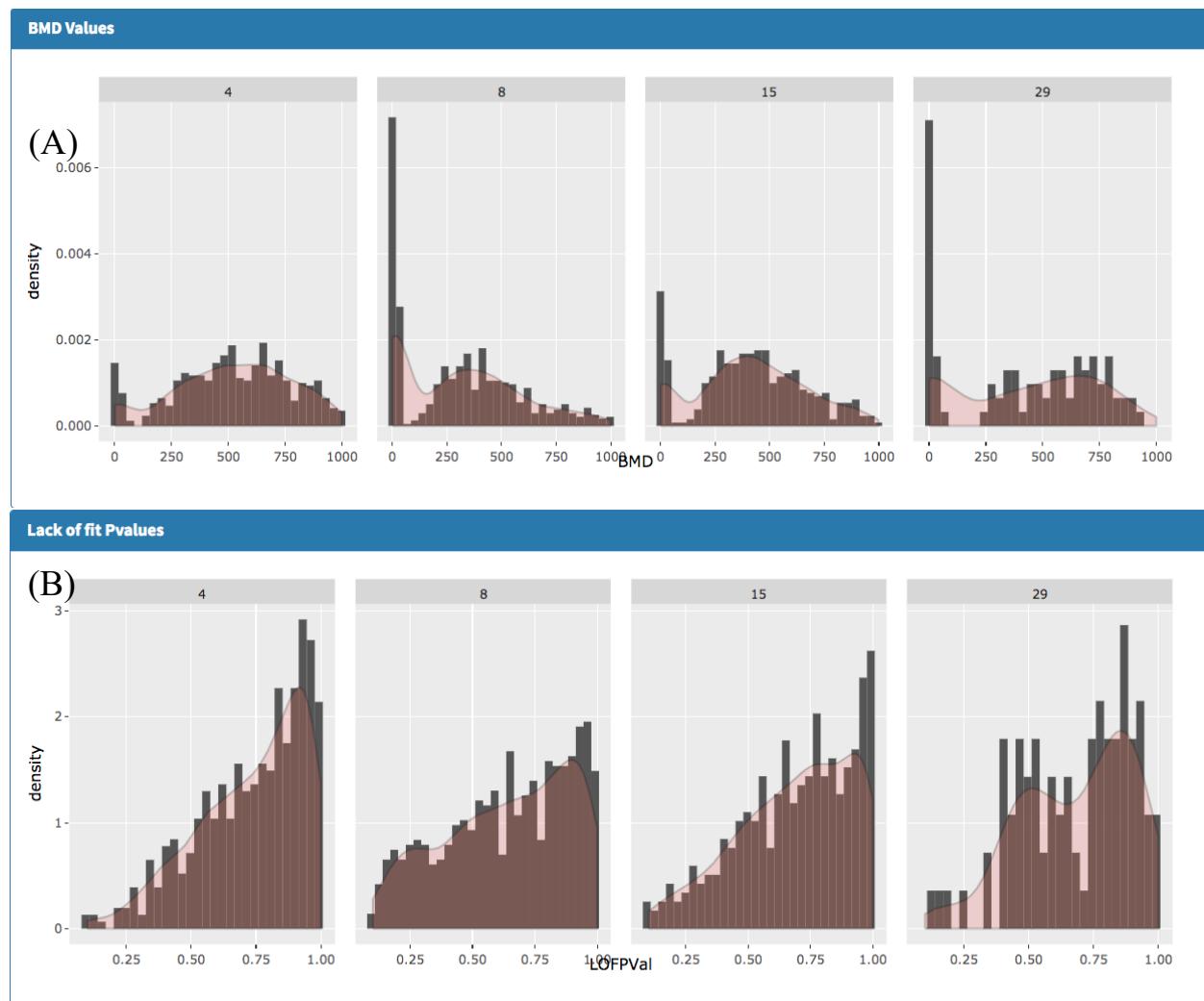
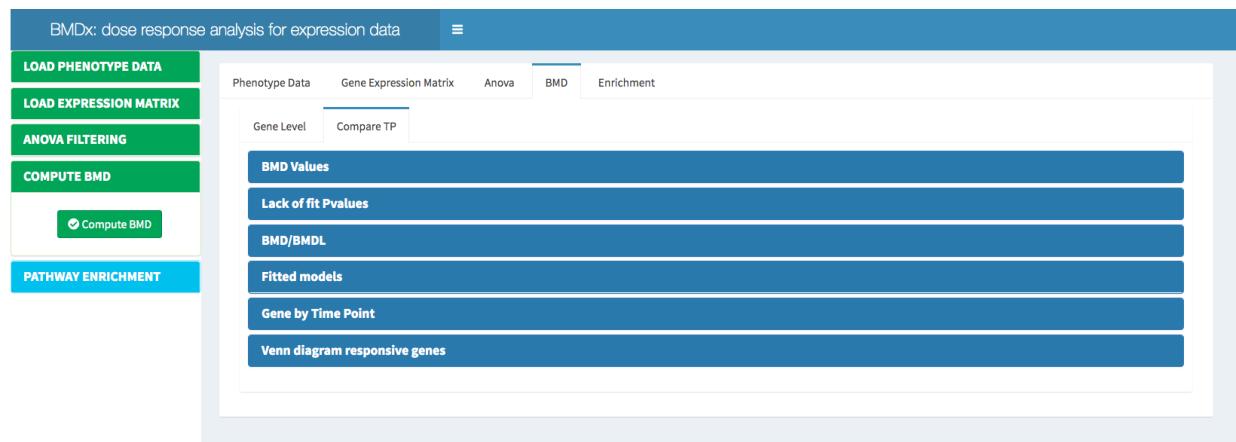
Fit of the model

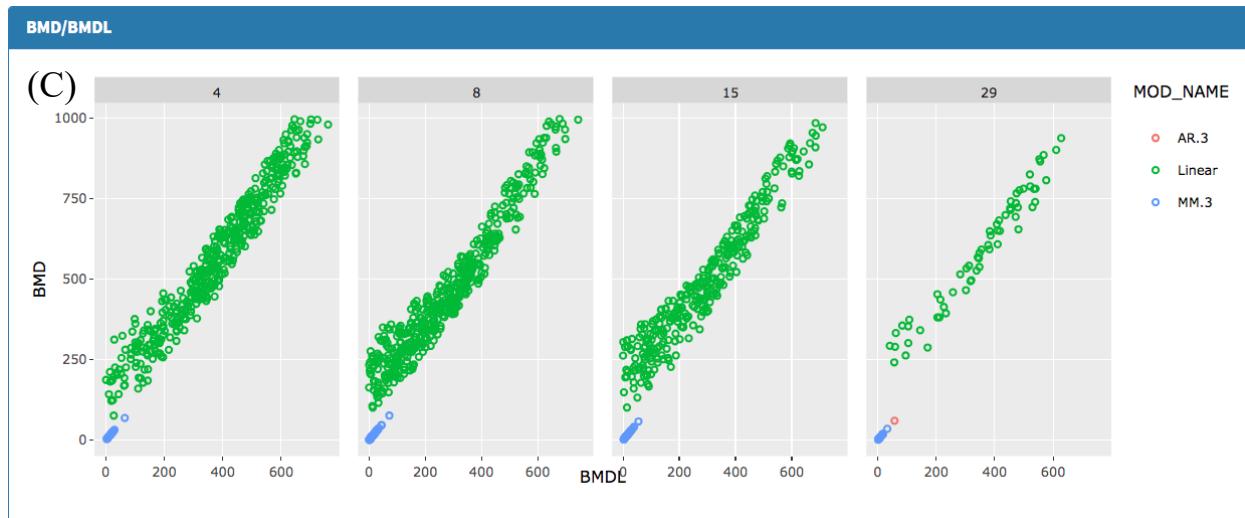
The fitting of the model can be visualised below the table by clicking on the row of the gene. Calculated values are shown in the figure with specific colours: red indicates the value for the BMDL, blue for BMD while black marks the BMDU value. IC50/EC50 is marked with black.

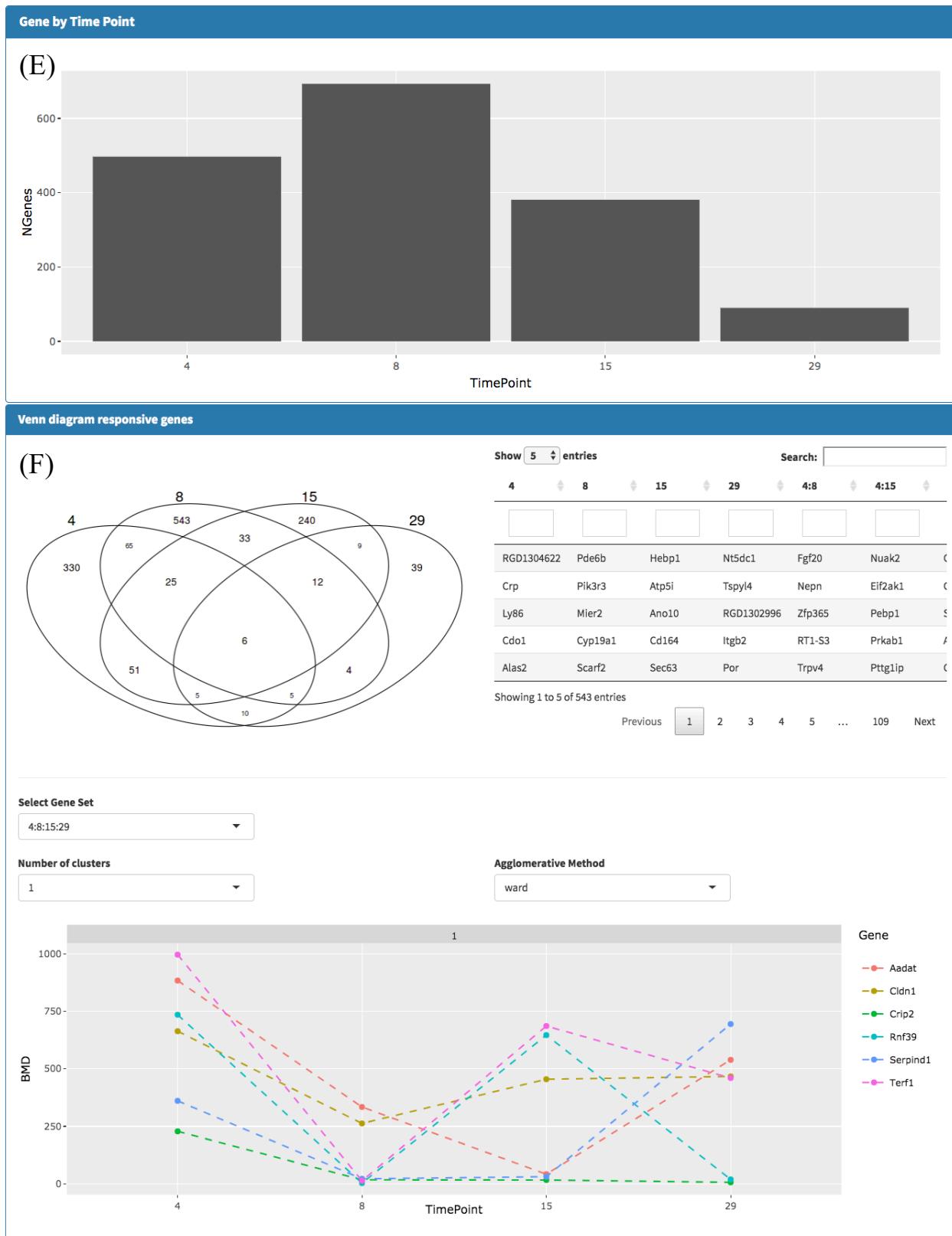


Comparing results between time points

When multiple time points are included in the analysis, the results between them can be visualised on the *Compare TP* tab. BMDx allows for the visualisation of the density of the BMD values (A) as well the lack-of-fit p-values (B). The BMD values obtained at each time point with each of the optimal models are plotted (C), the proportion of the models at each time point are visualised (D). The number of dose-dependent genes at each time point are shown as bar plots for easy comparison (E), and finally, a Venn diagram of the responsive genes is shown and the gene lists at all of the intersections can be explored (F).

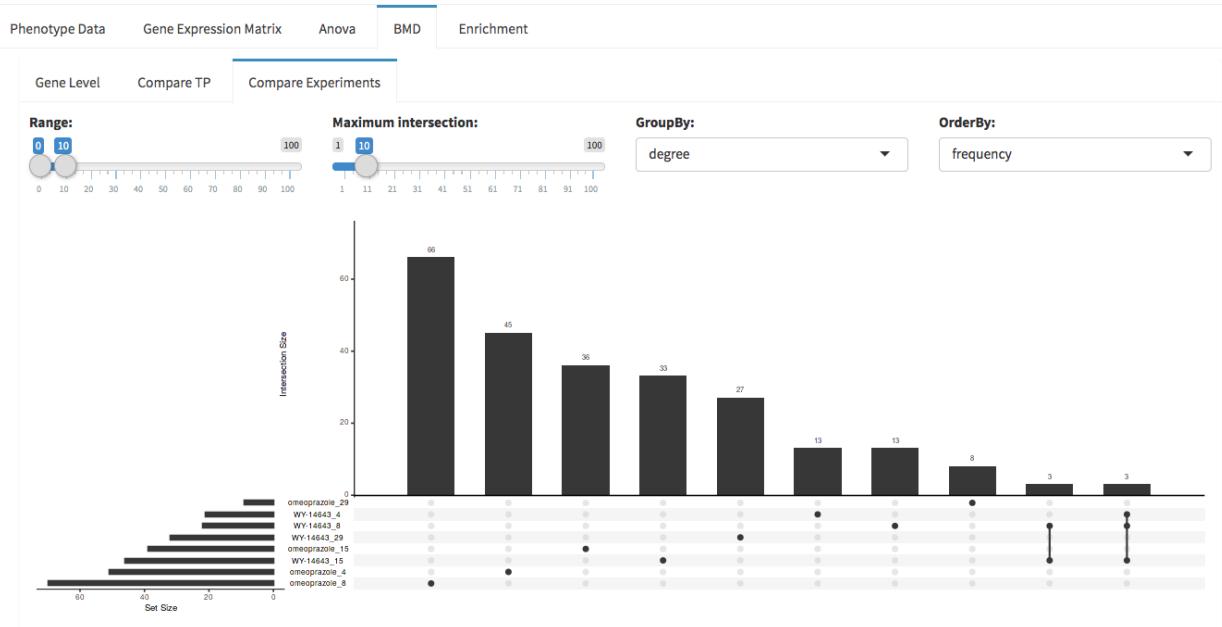






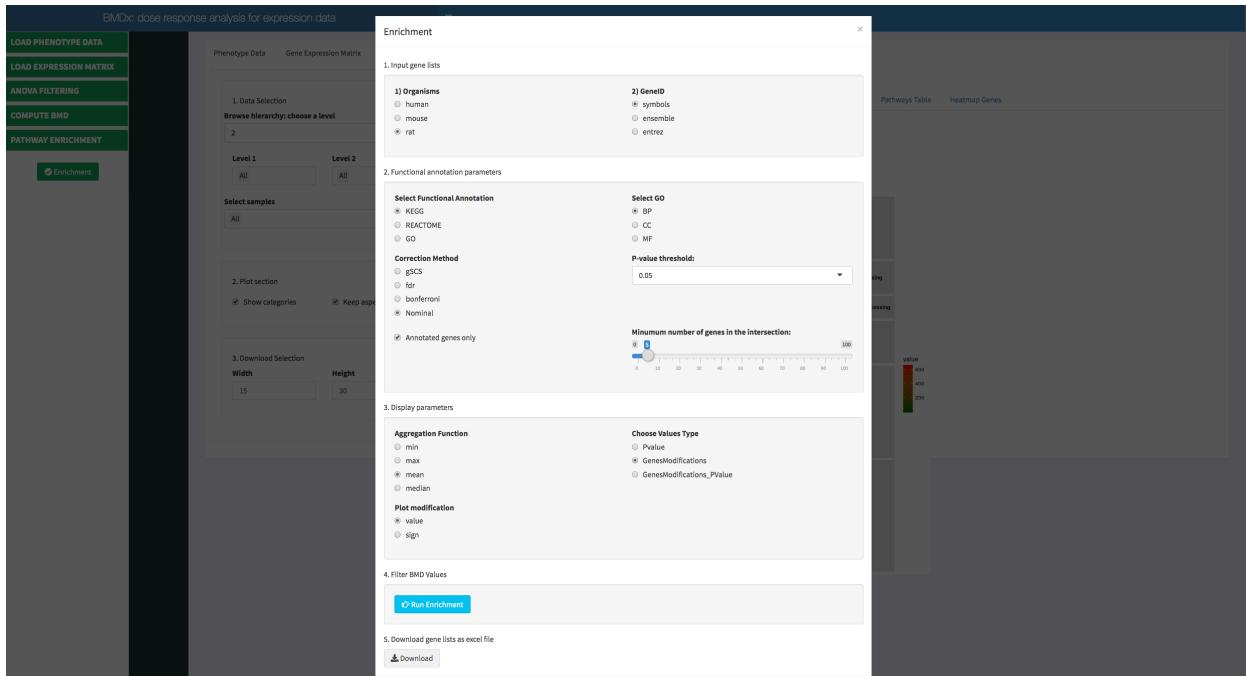
Compare Experiments

An UpSet plot can be viewed to represent the intersecting genes between time points and experiments.



Functional Enrichment

Finally, the results of the BMD analysis can be explored in the form of a pathway enrichment analysis by launching a graphical window by clicking on *Enrichment* on the side panel. For detailed information on the enrichment tool, please refer to Scala et al. (2019).



Enrichment parameters

The enrichment analysis supports human, mouse and rat genes expressed in official gene symbols or Ensemble or Entrez gene IDs. The right parameters are selected from the options provided in the first part of the graphical window (*1. Input gene lists*).

1. Input gene lists

This screenshot shows the 'Input gene lists' section of the enrichment configuration window. It contains two main groups of parameters:

- 1) Organisms:** Radio buttons for 'human', 'mouse', and 'rat'. The 'rat' button is selected.
- 2) GeneID:** Radio buttons for 'symbols', 'ensemble', and 'entrez'. The 'symbols' button is selected.

Functional annotation can be selected between KEGG and Reactome Pathways or GO terms. For GO terms, specify BP for Biological Pathways, CC for Cellular Components or MF Molecular Functions. P-value threshold for the enrichment is selected from the drop menu and the correction method for the p-value can be selected from several methods (gSCS, FDR, Bonferroni).

2. Functional annotation parameters

Select Functional Annotation

KEGG
 REACTOME
 GO

Correction Method

gSCS
 fdr
 bonferroni
 Nominal

Annotated genes only

Select GO

BP
 CC
 MF

P-value threshold:

0.01

Minimum number of genes in the intersection:

0 5 100

In the bottom of the window, the user can specify the display parameters used for the plotting of the enrichment map. *Aggregation function* (min, max, mean, median) specifies the function to be used when all the genes annotated to the same pathway are aggregated, while *Plot modification* specifies whether the enrichment map is plotted in chromatic scale or in one colour. *Choose values type* determines if the values plotted in the map are the p-values of the enrichment, the genes modifications (i.e. the BMD value) or a combination of the two.

3. Display parameters

Aggregation Function

min
 max
 mean
 median

Plot modification

value
 sign

Choose Values Type

Pvalue
 GenesModifications
 GenesModifications_PValue

To run the enrichment with selected parameters, *Run Enrichment* is clicked. If the user wants to change the parameters or enrichment type later, the window is launched again from the side panel.

4. Filter BMD Values

 Run Enrichment

The input for the enrichment tool can be downloaded by clicking the *download* button on the bottom of the graphical window. The file contains multiple sheets, each representing the genes and their BMD values at each time point of the included experiments.

5. Download gene lists as excel file



Enrichment results

After running the enrichment, the results can be explored on the *Enrichment* tab. Before visualizing the map, the user must specify the hierarchy level on which the results are shown and click on *Plot Map* to open the following view.

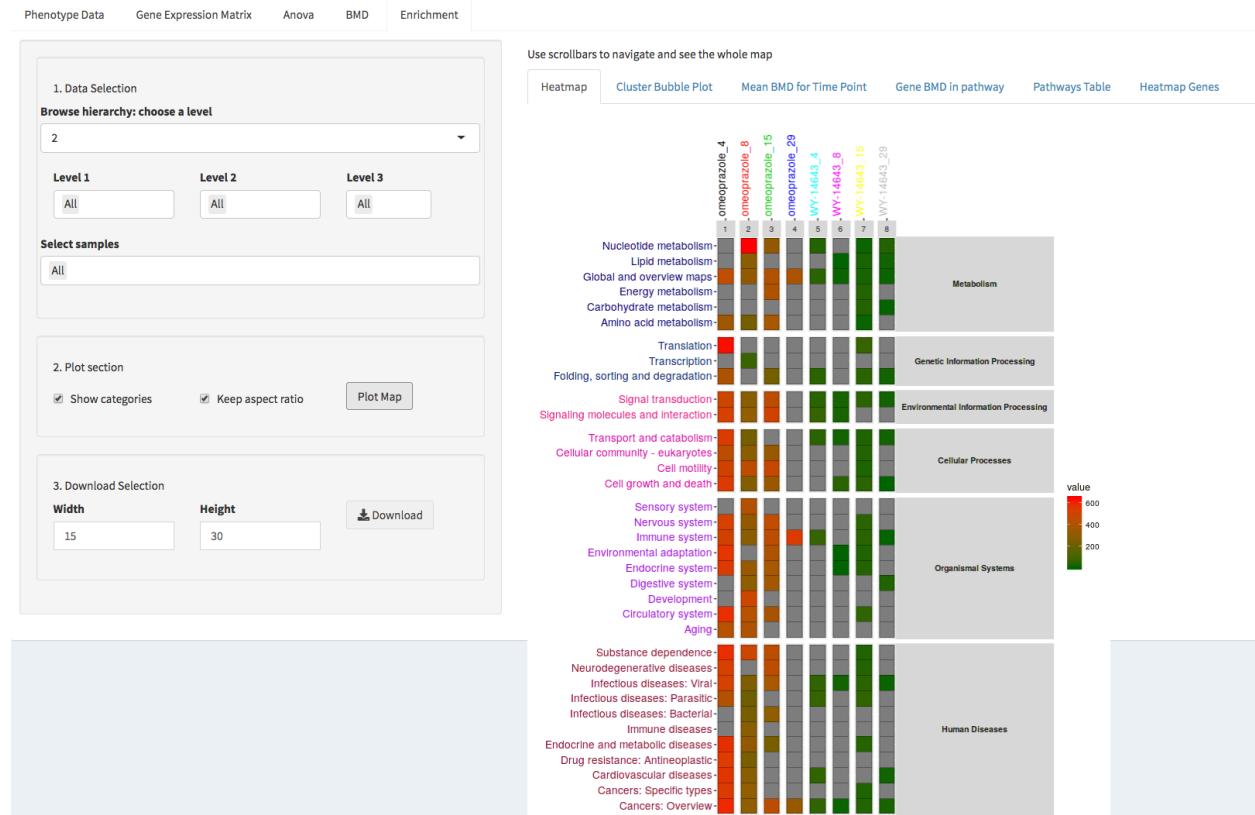
The screenshot shows a software interface for enrichment analysis. At the top, there are tabs: Phenotype Data, Gene Expression Matrix, Anova, BMD, and Enrichment (which is selected). Below the tabs, there are three main sections:

- 1. Data Selection:** A dropdown menu labeled "Browse hierarchy: choose a level" set to "1". Below it are three "Level" buttons: Level 1 (All), Level 2 (All), and Level 3 (All). A "Select samples" dropdown is set to "All".
- 2. Plot section:** Two checkboxes: "Show categories" (checked) and "Keep aspect ratio" (checked). A "Plot Map" button is located to the right.
- 3. Download Selection:** Input fields for "Width" (15) and "Height" (30). A "Download" button is located to the right.

At the top right, there is a message: "Use scrollbars to navigate and see the whole map". Below it is a horizontal bar with five buttons: Heatmap, Cluster Bubble Plot, Mean BMD for Time Point, Gene BMD in pathway, and Pathways Table. The "Heatmap" button is highlighted. Below the bar, the text "No data to plot" is displayed.

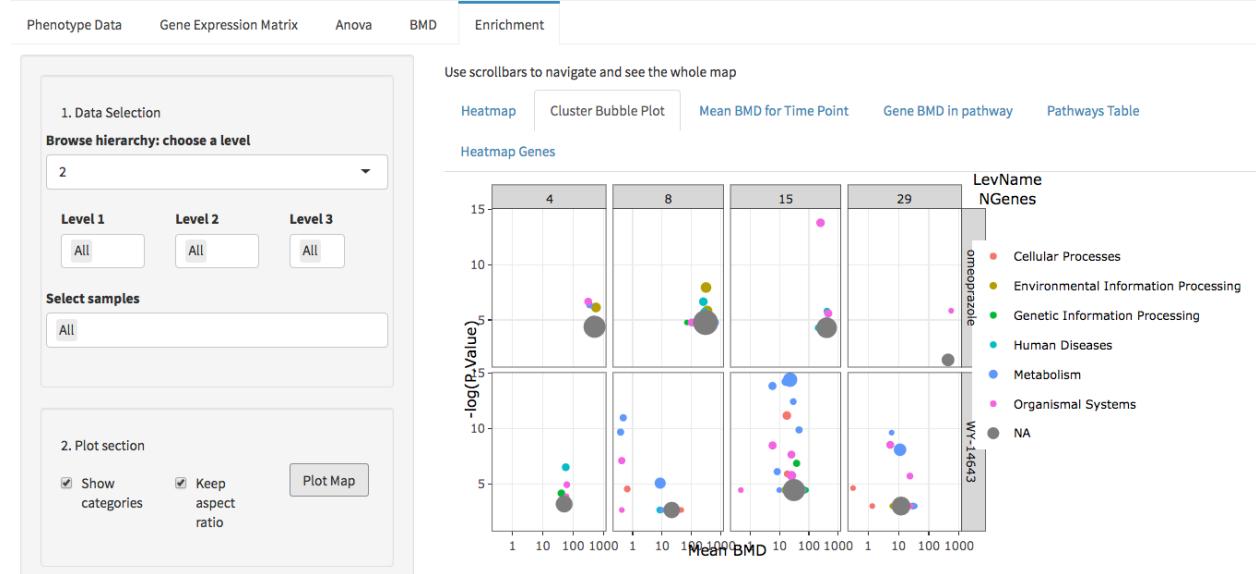
Heatmap

The enrichment heatmap shows each time point in each experiment as a separate column. Pathways are shown in the rows, and coloured boxes indicate enrichment of the pathway at the specific condition. When plotted values are shown in chromatic scale, the colour of the box changes according to the value. For example, in the figure below, the colour indicates the mean BMD value of the genes contributing to the enrichment of the pathway, showing the difference in the BMD values between the two experiments.



Cluster Bubble Plot

The functional enrichment can be visualised in the form of a bubble plot. Slots consisting of each time point in each experiment show bubbles characterising the size of the enrichment. The bigger the bubble, the more terms are included in that category.



Mean BMD for Time Point

The mean BMD values for each time point are shown as stacked bar plots for each experiment separately.



Gene BMD in Pathway

Gene BMD in a Pathway tab shows all pathways with their enrichment p-values. Selecting a row of the table plots a graph under the table with all the genes in the pathway, their BMD values as well as the lower and upper confidence bound BMD. Note! Deselect the row before selecting the next row to only include the genes in the desired pathway.

1. Data Selection

Browse hierarchy: choose a level

2

Level 1	Level 2	Level 3
All	All	All

Select samples

All

Use scrollbars to navigate and see the whole map

[Heatmap](#) [Cluster Bubble Plot](#) [Mean BMD for Time Point](#) **Gene BMD in pathway** [Pathways Table](#)

[Heatmap Genes](#)

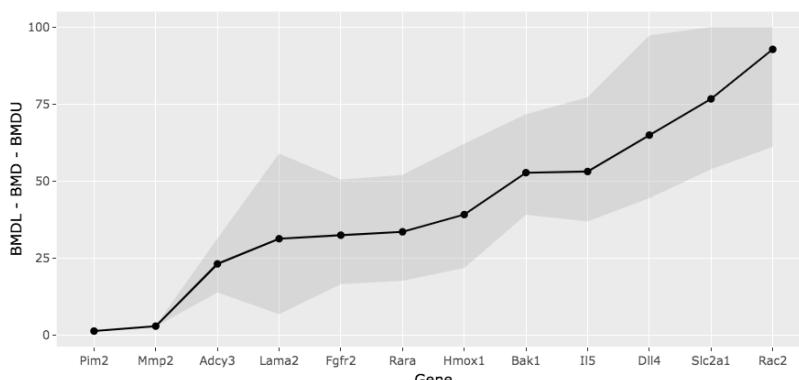
Time Points

WY-14643_4

Show 10 entries Search:

	annID	pValue	pValueAdj	Description
	All	All	All	All
1	00000	0.0344827586206897	0.0344827586206897	KEGG pathways
2	04060	0.0344827586206897	0.0344827586206897	Cytokine-cytokine receptor interaction
3	05166	0.0344827586206897	0.0344827586206897	Human T-cell leukemia virus 1 infection
4	04664	0.0344827586206897	0.0344827586206897	Fc epsilon RI signaling pathway
5	04151	0.0344827586206897	0.0344827586206897	PI3K-Akt signaling pathway
6	05202	0.0344827586206897	0.0344827586206897	Transcriptional misregulation in cancer
7	01100	0.0344827586206897	0.0344827586206897	Metabolic pathways
8	05203	0.0344827586206897	0.0344827586206897	Viral carcinogenesis
9	05200	0.0344827586206897	0.0344827586206897	Pathways in cancer
10	04015	0.0344827586206897	0.0344827586206897	Rap1 signalling pathway

Showing 1 to 10 of 29 entries Previous **1** 2 3 Next



Pathways Table

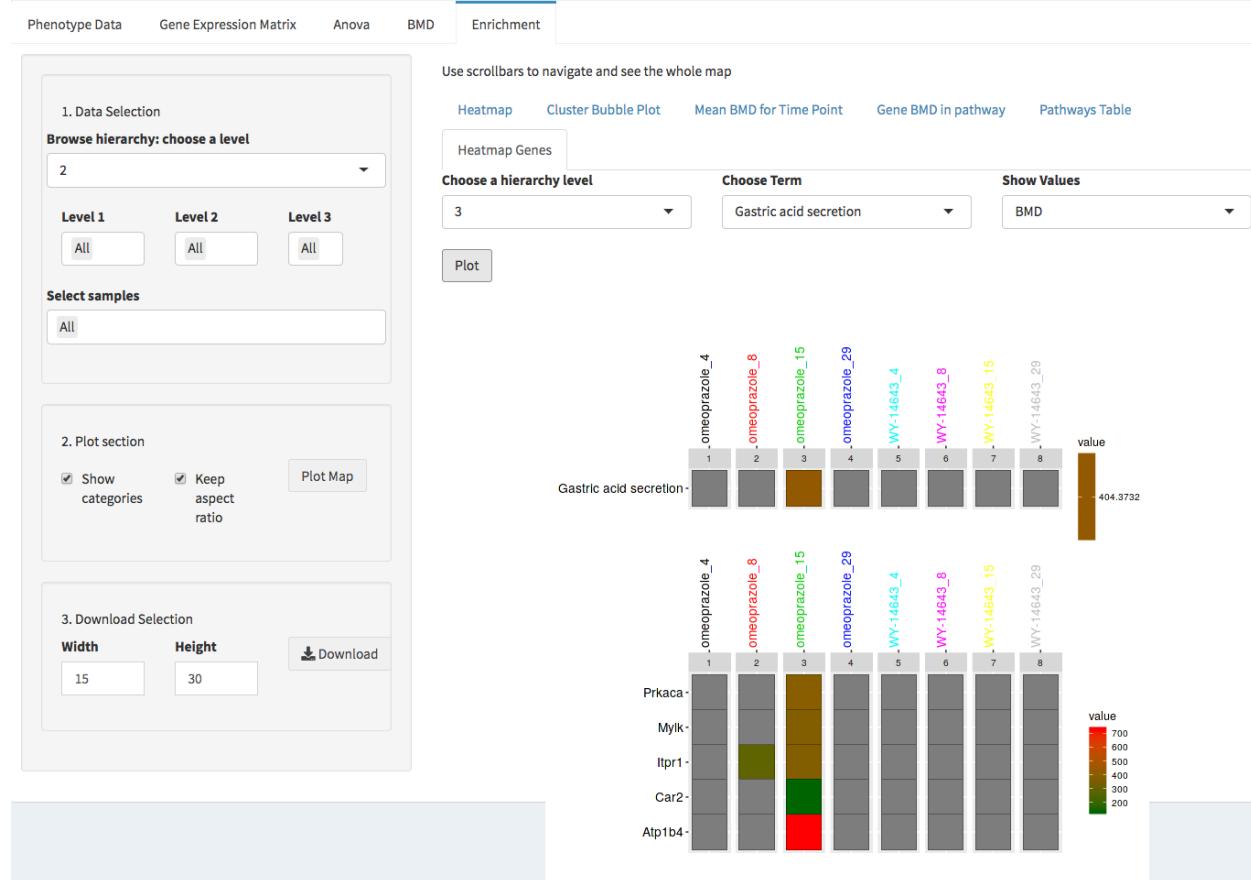
The genes mapped to each term are shown as a table in *Pathways table* tab. Time point for which data is shown can be changed from the *time point* drop menu. The tables can be downloaded as a single Excel file by clicking *download*.

The screenshot shows a user interface for pathway analysis. At the top, there are tabs: Phenotype Data, Gene Expression Matrix, Anova, BMD, and Enrichment (which is selected). Below the tabs are three sections: 1. Data Selection, 2. Plot section, and 3. Download Selection. The Data Selection section includes a dropdown for 'Browse hierarchy: choose a level' (set to 2), dropdowns for Level 1, Level 2, and Level 3 (all set to All), and a 'Select samples' dropdown also set to All. The Plot section has checkboxes for 'Show categories' (checked) and 'Keep aspect ratio' (checked), and a 'Plot Map' button. The Download Selection section has input fields for 'Width' (15) and 'Height' (30), and a 'Download' button. To the right, there is a large table titled 'Pathways Table'. The table header includes columns for 'Description', 'annID', and 'gID', with an 'All' button below it. The table body contains 10 rows of pathway information, each with a unique ID, name, annID, and associated genes. Row 1: KEGG pathways, 00000, FGF20, CDO1, ALAS2, NOTCH4, PRF1, MARCKS, IPMK, JMJD1C, ACACB, RT1-S3, RPP21, RT1-N3, FABP7, NCOR2, EIF2AK1, VPS37B, PRKAB1, GLP1R, TRPV4, COL18A1, PPP1CC, CARD6, HPD, ALDH2, SD. Row 2: Hepatitis C, 05160, EIF2AK1, CLDN1, SOCS3, CLDN6, SOS2, CLDN9. Row 3: Adrenergic signaling in cardiomyocytes, 04261, PPP1CC, ATF2, CACNG4, CACNG5, PRKACA, CREB3L2, CREM. Row 4: Epstein-Barr virus infection, 05169, RT1-S3, RT1-N3, NCOR2, EIF2AK1, ATF2, PRKACA, HSPB2, TBPL1. Row 5: Fluid shear stress and atherosclerosis, 05418, TRPV4, NFE2L2, SDC2, ACVR2A, RAC2, ASS1, VCAM1, SUMO1, MMP2. Row 6: Cholinergic synapse, 04725, CACNA1B, PRKACA, GNA11, CHRNA4, CREB3L2, CHRNA3. Row 7: Prostate cancer, 05215, PDGFRA, IGF1, SOS2, CREB3L2, PDGFRB. Row 8: Estrogen signaling pathway, 04915, ATF2, SOS2, PRKACA, KRT12, CREB3L2, SP1, MMP2. Row 9: Cellular senescence, 04218, RT1-S3, RT1-N3, TRPV4, PPP1CC, MAPKAPK2, HUS1, EIF4EBP1, TGFBR2, VDAC2. Row 10: AGE-RAGE signaling pathway in diabetic complications, 04933, COL3A1, CYBB, TGFBR2, NOX4, FN1, VCAM1, MMP2. At the bottom of the table, it says 'Showing 1 to 10 of 93 entries' and has navigation buttons for Previous, 1, 2, 3, 4, 5, ..., 10, Next.

Description	annID	gID
KEGG pathways	00000	FGF20, CDO1, ALAS2, NOTCH4, PRF1, MARCKS, IPMK, JMJD1C, ACACB, RT1-S3, RPP21, RT1-N3, FABP7, NCOR2, EIF2AK1, VPS37B, PRKAB1, GLP1R, TRPV4, COL18A1, PPP1CC, CARD6, HPD, ALDH2, SD
Hepatitis C	05160	EIF2AK1, CLDN1, SOCS3, CLDN6, SOS2, CLDN9
Adrenergic signaling in cardiomyocytes	04261	PPP1CC, ATF2, CACNG4, CACNG5, PRKACA, CREB3L2, CREM
Epstein-Barr virus infection	05169	RT1-S3, RT1-N3, NCOR2, EIF2AK1, ATF2, PRKACA, HSPB2, TBPL1
Fluid shear stress and atherosclerosis	05418	TRPV4, NFE2L2, SDC2, ACVR2A, RAC2, ASS1, VCAM1, SUMO1, MMP2
Cholinergic synapse	04725	CACNA1B, PRKACA, GNA11, CHRNA4, CREB3L2, CHRNA3
Prostate cancer	05215	PDGFRA, IGF1, SOS2, CREB3L2, PDGFRB
Estrogen signaling pathway	04915	ATF2, SOS2, PRKACA, KRT12, CREB3L2, SP1, MMP2
Cellular senescence	04218	RT1-S3, RT1-N3, TRPV4, PPP1CC, MAPKAPK2, HUS1, EIF4EBP1, TGFBR2, VDAC2
AGE-RAGE signaling pathway in diabetic complications	04933	COL3A1, CYBB, TGFBR2, NOX4, FN1, VCAM1, MMP2

Heatmap Genes

Finally, the genes in different pathways can be explored in the form of a heatmap. The drop menus allow for the specification of the hierarchy level, the terms belonging to that level and specification of values shown on the heatmap. The heatmap then shows the genes mapped to the selected term on rows with the experiments and time points as columns, and coloured boxes indicating the value specified in the *Show values* drop menu.



Model descriptions

The models available for the evaluation of the BMD are:

Linear Model:

$$f(dose) = \beta_0 + \beta_1 dose$$

Polynomial model:

$$f(dose) = \beta_0 + \beta_1 dose + \beta_2 dose^2 + \dots + \beta_n dose^n$$

Here n is the degree of the polynomial. The user can choose between $n = 2, 3$

Power model:

$$f(dose) = \beta_0 + (dose)^\delta$$

The user can choose between $\delta = 2, 3, 4$

Exponential model:

$$f(dose) = \beta_0 + expr(dose)$$

Hill model:

$$f(dose) = \beta_0 + \frac{dose^n}{Kd + dose^n}$$

The user can choose between $n = 0.5, 1, 2, 3, 4, 5$, while $Kd = 10$

Asymptotic regression model:

$$f(dose) = c + (d - c) \times (1 - expr(-dose/e))$$

The parameter c is the lower limit (at $x=0$), the parameter d is the upper limit and the parameter $e > 0$ is determining the steepness of the increase of dose.

The AR.3 model is the one depending from c, d and e parameters. The AR.2 model depends only on d and e parameters, while c is set to zero

Michaelis-Menten Model:

The model is defined by the three-parameter model (MM.3) function

$$f(dose, (c, d, e)) = c + \frac{d-c}{1+(e/dose)}$$

It is increasing as a function of the dose, attaining the lower limit at dose 0 ($x=0$) and the upper limit d for infinitely large doses. The parameter e corresponds to the dose yielding a response halfway between c and d.

The common two-parameter Michaelis-Menten model (MM.2) is obtained by setting c equal to 0.

References

- Abraham, K., Mielke, H. and Lampen, A. 2012. Hazard characterization of 3-MCPD using benchmark dose modeling: Factors influencing the outcome. *European Journal of Lipid Science and Technology* 114(10), pp. 1225–1226.
- EFSA Scientific Committee, Hardy, A., Benford, D., et al. 2017. Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA Journal* 15(8).
- Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., & Yamada, H. 2015. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic acids research*, 43(Database issue), D921–D927.
- Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28(6), pp. 882–883.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7), p. e47.
- Scala, G., Serra, A., Marwah, V.S., Saarimäki, L.A. and Greco, D. 2019. FunMappOne: a tool to hierarchically organize and visually navigate functional gene annotations in multiple experiments. *BMC Bioinformatics* 20(1), p. 79.
- Thomas, R.S., Allen, B.C., Nong, A., et al. 2007. A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. *Toxicological Sciences* 98(1), pp. 240–248.
- Woods, CG., Burns, A.M., Bradford, B.U., Ross, P.K., Kosyk., O., Swenberg, J.A., Cunningham, M.L. and Rusyn, I. 2007. WY-14,643 Induced Cell Proliferation and Oxidative Stress in Mouse Liver are Independent of NADPH Oxidase. *Toxicological Sciences*, 98(2), 366.