

# BMDx Reference Manual

## with Sample Data Analysis

# Contents

<b>About BMDx.....</b>	<b>2</b>
<b>Data Description.....</b>	<b>3</b>
<b>Installation and execution .....</b>	<b>4</b>
<b>Workflow Interface.....</b>	<b>4</b>
<b>Input Description.....</b>	<b>4</b>
<b>Phenotype Specification.....</b>	<b>5</b>
Load Pheotype .....	5
Select Phenotype File.....	5
Phenotype Preview .....	6
Configure Variable R Format.....	6
Specify Sample ID, Dose and Time Point Variables.....	6
Import Phenotype .....	6
Phenotype View.....	6
<b>Expression matrix .....</b>	<b>7</b>
Load Expression Matrix.....	7
<b>Gene Filtering .....</b>	<b>8</b>
<b>Model fitting and BMD computation .....</b>	<b>9</b>
<b>Parameter Selection .....</b>	<b>9</b>
BMD value filtering based on the doses.....	9
BMRF selection .....	9
Assumption of constant variance model .....	10
Model selection.....	10
<b>Results investigation.....</b>	<b>12</b>
<b>Fit of the model.....</b>	<b>13</b>
<b>Comparing results between time points.....</b>	<b>13</b>
<b>Compare Experiments.....</b>	<b>17</b>
<b>Functional Enrichment.....</b>	<b>18</b>
<b>Enrichment parameters .....</b>	<b>18</b>
<b>Enrichment results.....</b>	<b>20</b>
Heatmap.....	22
Cluster Bubble Plot .....	23
Mean BMD for Time Point.....	23
Gene BMD in Pathway .....	24
Pathways Table.....	25
Heatmap Genes .....	26
<b>Model descriptions.....</b>	<b>27</b>
<b>References.....</b>	<b>28</b>

## 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 26 of this document with model descriptions included.

In this manual, we provide a detailed step-by-step guide to using BMDx and explanation of the analysis workflow (see Figure 1). 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.

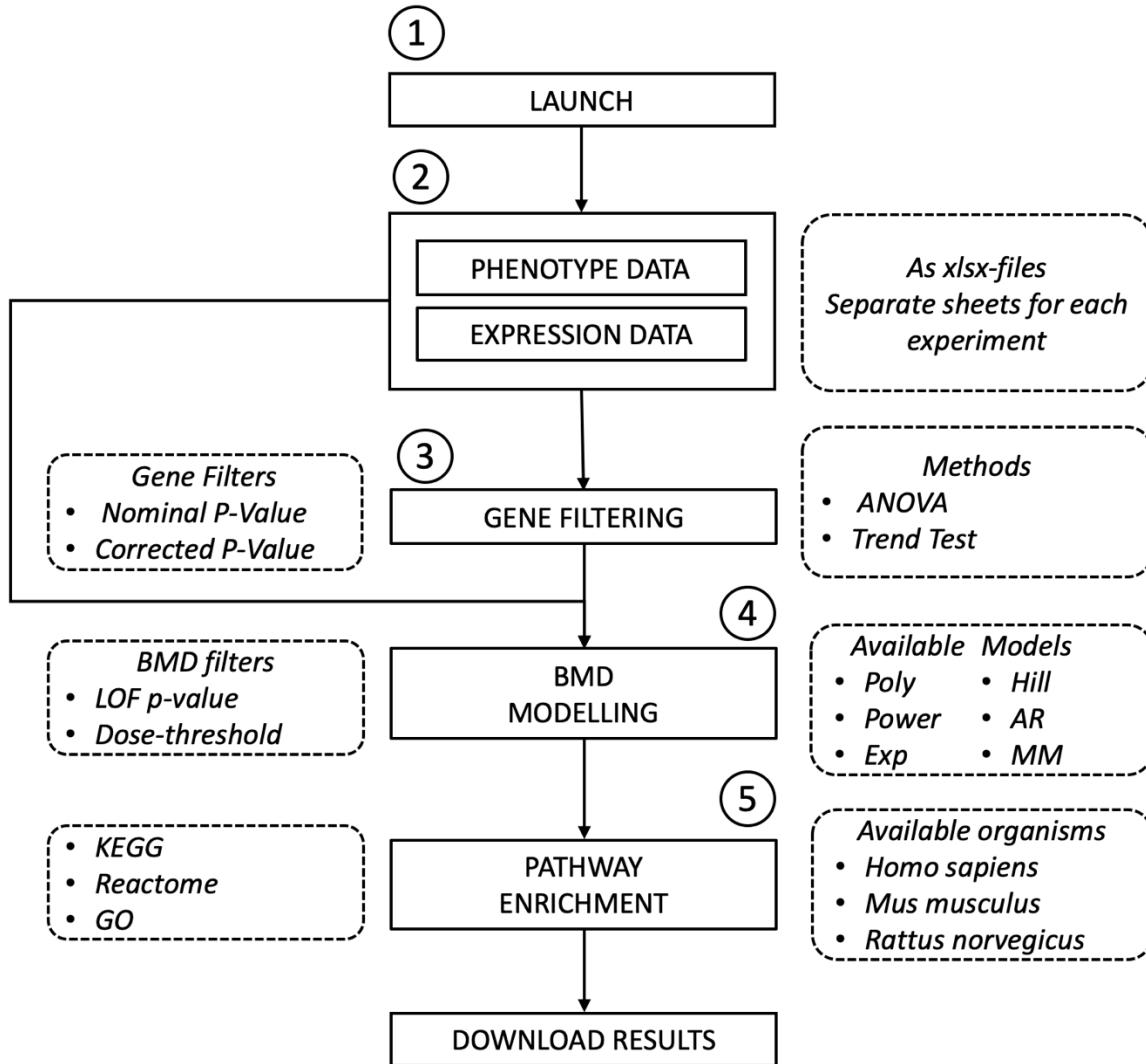


Figure 1: BMDx workflow. Schematic representation of the BMDx software workflow. The steps are represented by rectangular boxes with sharp edges, while the steps parameters are represented with rectangular boxes with rounded edges. Analytical steps are numerically coded with circular labels from 1 to 5.

## 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 the 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.

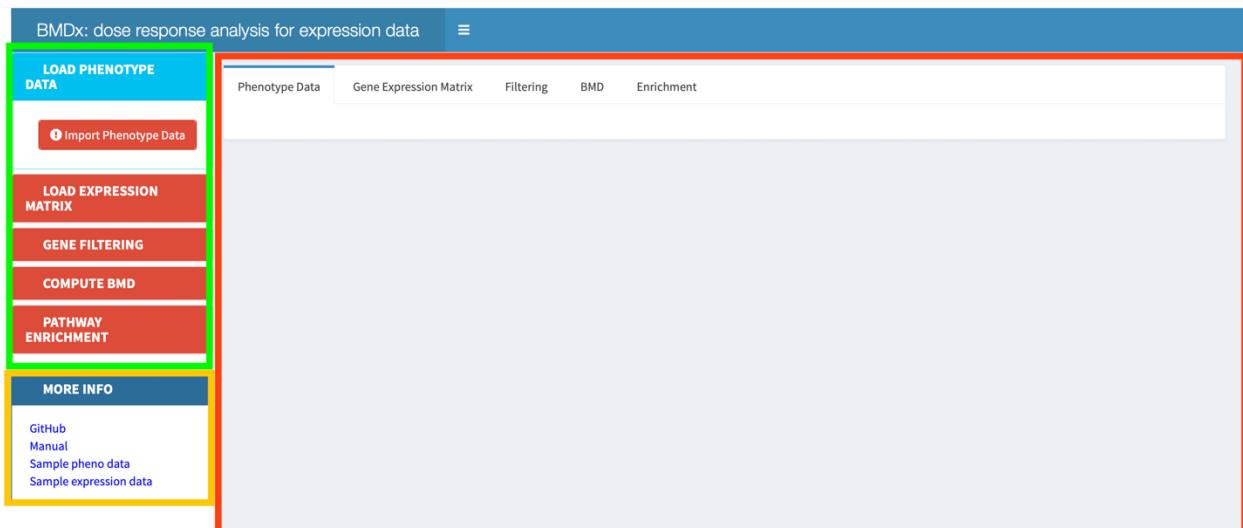
Data used as an example in this document are available on GitHub (<https://github.com/Greco-Lab/BMDx>).

## 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). Below this section, links to the GitHub page of the tool as well as the manual and sample data are provided (marked with orange outline). The output of the steps is 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.**

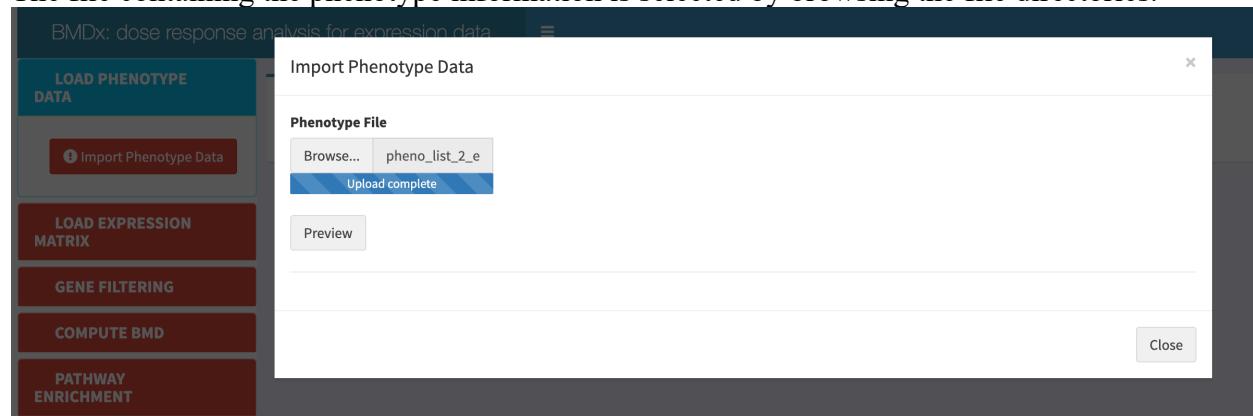
	BARCODE	DOSE	DOSE_LEVEL	SACRIFICE_PERIOD
1				
2	12800 ID_003017698023	0	Control	4
3	12801 ID_003017698024	0	Control	4
4	12802 ID_003017698025	0	Control	4
5	12803 ID_003017699005	0	Control	8
6	12804 ID_003017699006	0	Control	8
7	12805 ID_003017699007	0	Control	8
8	12806 ID_003017667006	0	Control	15
9	12807 ID_003017667007	0	Control	15
10	12808 ID_003017667008	0	Control	15
11	12809 ID_003017667018	0	Control	29
12	12810 ID_003017667019	0	Control	29
13	12811 ID_003017667020	0	Control	29
14	12812 ID_003017698026	100	Low	4
15	12813 ID_003017698027	100	Low	4
16	12814 ID_003017698028	100	Low	4
17	12815 ID_003017699008	100	Low	8
18	12816 ID_003017699009	100	Low	8

## Load Phenotype

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' dialog box. At the top, it displays 'Phenotype File' with a 'Browse...' button, the 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 contains a table of data with columns: Variable, Type, Class, Sample1, Sample2, Sample3, Sample4, Sample5, and Sa. The rows are: BARCODE [1] (factor, character), DOSE [2] (vector, numeric), DOSE\_LEVEL [3] (factor, character), and SACRIFICE\_PERIOD [4] (vector, numeric). An arrow points to the 'Type' column header with the text 'Double-click to change type'. Below the table, there are three dropdown menus: 'Sample ID Variable' (Variable 1), 'Dose Variable' (Variable 2), and 'Time Point Variable' (Variable 4). A modal window is open over these dropdowns, listing 'Variable 1', 'Variable 2' (which is highlighted in grey), 'Variable 3', and 'Variable 4'. On the right side of this modal is a blue 'Import' button with a cloud icon, and on the far right is a 'Close' button.

## 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.

BMDx: dose response analysis for expression data

**LOAD PHENOTYPE DATA**

**LOAD EXPRESSION MATRIX**

**GENE FILTERING**

**COMPUTE BMD**

**PATHWAY ENRICHMENT**

**MORE INFO**

- [GitHub](#)
- [Manual](#)
- [Sample pheno data](#)
- [Sample expression data](#)

Phenotype Data		Gene Expression Matrix	Filtering	BMD	Enrichment
Show 10 entries					
					Search: <input type="text"/>
BARCODE		DOSE	DOSE_LEVEL	SACRIFICE_PERIOD	
12800	ID_003017698023	0	Control	4	
12801	ID_003017698024	0	Control	4	
12802	ID_003017698025	0	Control	4	
12803	ID_003017699005	0	Control	8	
12804	ID_003017699006	0	Control	8	
12805	ID_003017699007	0	Control	8	
12806	ID_003017667006	0	Control	15	
12807	ID_003017667007	0	Control	15	
12808	ID_003017667008	0	Control	15	
12809	ID_003017667018	0	Control	29	

Showing 1 to 10 of 48 entries

Previous 1 2 3 4 5 Next

## 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> (link also provided in the graphical interface).

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_003017667070
2	Gad1	6.41439102304723	6.4652837902453	6.15713257780585	6.38079246591397	6.4828549058064	6.56574785793586	6.68232330197958	6.80677729775191
3	Cbln1	6.49548074605588	6.2474614500316	6.27445075342446	6.41063178671925	6.40066644578383	6.49783392846416	6.41888260148178	6.52749758647252
4	Stap1	6.08725181707615	7.2392577452614	6.22159431386313	6.16469548011423	5.849072064269376	3.95905820546415	8.3780943848238	6.432777081306126
5	Hebp1	8.16244677988902	7.98135071582213	7.30078550598878	7.376645571455227	7.567176202452617	8.7981448716476	6.95798232367227	6.76357210565463
6	Tmcc2	6.30191849868735	6.08084669634946	6.12173437902059	6.226592404203876	6.15916599240876	6.4432544609901	6.0295206830075	6.2231234714791
7	Nuak2	6.8608188928228	6.70585968433675	6.84342702783655	6.7861884544676	6.928968581636846	6.84461448501822	6.52149337211356	6.29071007301364
8	Klhd8a	6.70153179562915	6.49409288378762	6.6551999897453	6.64152966932325	6.63796441320896	6.65921658975372	6.26284294953698	6.59582961239123
9	RGD1304622	6.15895311078875	6.76185279246275	6.11118369359443	6.43082784238366	6.1305415327246	6.40917172170263	6.52607578979618	6.4080867252917
10	Slc26a1	7.78552753621991	7.05396358247276	7.30007101657088	6.989630364197457	7.313415419748147	7.0817820321819	6.72938943982973	6.54395271550508
11	Cd82	6.56495105605652	6.44335816340407	6.33658735068232	6.504151793283716	6.50158016568822	6.56574785793586	6.11668555050335	5.96920847996047
12	Gak	6.67361640520244	6.42020620132176	6.30647778993558	5.855682825268856	6.277107300601904109285	6.56578620662586	6.104302678688	
13	Crp	7.88078857737421	7.66580053158294	7.70000860901788	7.3690812077646	7.80118607311785	8.10446795117725	7.37923434832878	6.90886740631087
14	Abhd8	6.59852776436937	6.26406219642696	5.94937772417718	6.49174877829386	6.21556731267876	6.268902060724315	6.9958289056345	6.2287378268976
15	Fcrl6	6.34498541593716	6.57499470445596	6.35004675732525	6.43657614607422	6.255360540043656	6.70757134232425	6.50190075428585	6.60951214467473
16	Cpxl1	6.57155895561965	6.226489878541626	6.15541626596436	6.199614319772326	6.33638370282556	6.24614056498135	6.495695165343316	6.245733326468416

◀ ▶ omeprazole WY-14643 +

## 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.

## Gene Filtering

As the first step of the analysis, the genes can be filtered according to two different criteria that the user can select:

1. ANOVA: the genes that do not show variability across different doses are identified by performing an ANOVA test for each gene and removed from the analysis.
2. Trend Test: the genes that do not show a monotonical trend with respect to doses are identified by trend test for each gene and removed from the analysis.

Filtering is performed by click on the preferred method on the side bar. To run the 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. Furthermore, the number of cores used for the analysis can be selected.

BMDx: dose response analysis for expression data

LOAD PHENOTYPE DATA  
LOAD EXPRESSION MATRIX  
GENE FILTERING

Anova Filtering  
Trend Filtering  
Skip Filtering

Filter Genes by Anova

Time Points: All  
Pvalue:  FDR  Nominal  
Anova PValue Th: 0.05  
Number of cores: 3  
Run Anova

ID	ID_003017667006	ID_00						
Hepb1	8.16244677988902	7.98135071582213	7.30078550598878	7.37664557145522	7.56717620245261	7.87981448716478	6.95798232367227	6.76
Tmcc2	6.30191849868739	6.08084669634948	6.12173437902059	6.22659240420387	6.15916599240876	6.44325544609901	6.06295206830075	6.22

BMDx: dose response analysis for expression data

LOAD PHENOTYPE DATA  
LOAD EXPRESSION MATRIX  
GENE FILTERING

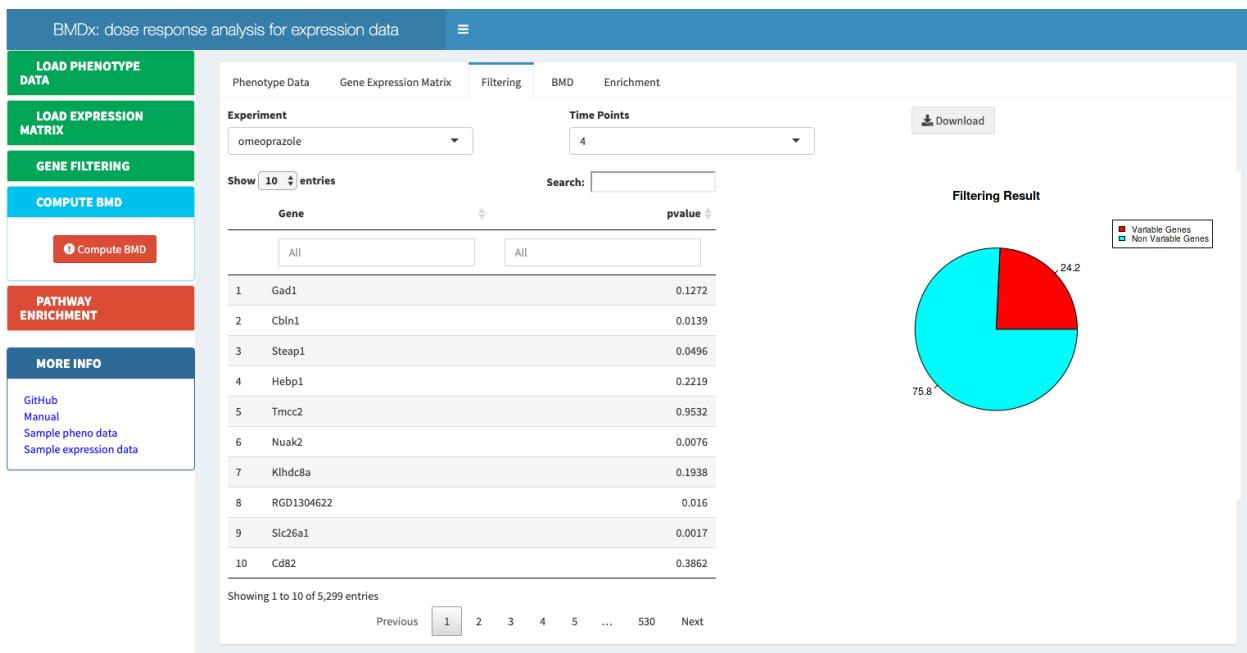
Anova Filtering  
Trend Filtering  
Skip Filtering

Filter Genes by Trend Test

Time Points: All  
Pvalue:  FDR  Nominal  
Trend PValue Th: 0.05  
Number of cores: 3  
Run Trend Test

ID	ID_003017667006	ID_00						
Hepb1	8.16244677988902	7.98135071582213	7.30078550598878	7.37664557145522	7.56717620245261	7.87981448716478	6.95798232367227	6.76
Tmcc2	6.30191849868739	6.08084669634948	6.12173437902059	6.22659240420387	6.15916599240876	6.44325544609901	6.06295206830075	6.22

Once the gene filtering step has been run, the results are displayed in the *Filtering* tab. In particular, a table with the ANOVA/Trend Test 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 filtering test for each time point. The results of the filtering test can be downloaded as one Excel file with multiple sheets, one for every time point at each experiment by clicking *Download*.



## 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 and lower limit for the estimation of the minimum and maximum BMD allowed with respect to minimum and maximum doses. Moreover, an assumption of constant variance must be specified (see below).

### BMD value filtering based on the doses

The two parameters *lowest/highest dose filter* are used to discard models where the predicted BMD value is x% lower than the lowest dose (lowest dose filter) or x% higher than the highest dose (highest dose filter). If 0 is selected, the exact minimum and maximum doses are used as limits. The value of x is selected from the drop menu for each parameter.

### BMRF selection

As a default, the BMRF is set to 1.349. As described in Thomas et al. (2007), a BMR factor 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.

## Assumption of constant variance model

If the tool is ran under the assumption of constant variance model, the BMRF is multiplied by the standard deviation of all the dose groups. Otherwise, the BMRF is multiplied by the standard deviation of the controls. (NTP, 2018)

## 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 non-statistically relevant fitting (lack-of-fit p-value  $< 0.1$ ) and predicted BMD value outside the selected range 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 26 of this file.

## Compute BMD Value

### Response Level

Assumption of Constant Variance

### Lack-of-fit PValue Th:

### Lowest dose filter:

### Highest dose filter:

### Number of cores:

### Select the BMD analysis setting

#### 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

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

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

$$n = 2, 3$$

**Power Model**

**Exponential Model**

**Hill Model**

**Asymptotic Regression**

**Michaelis-Menten Model**

## Results investigation

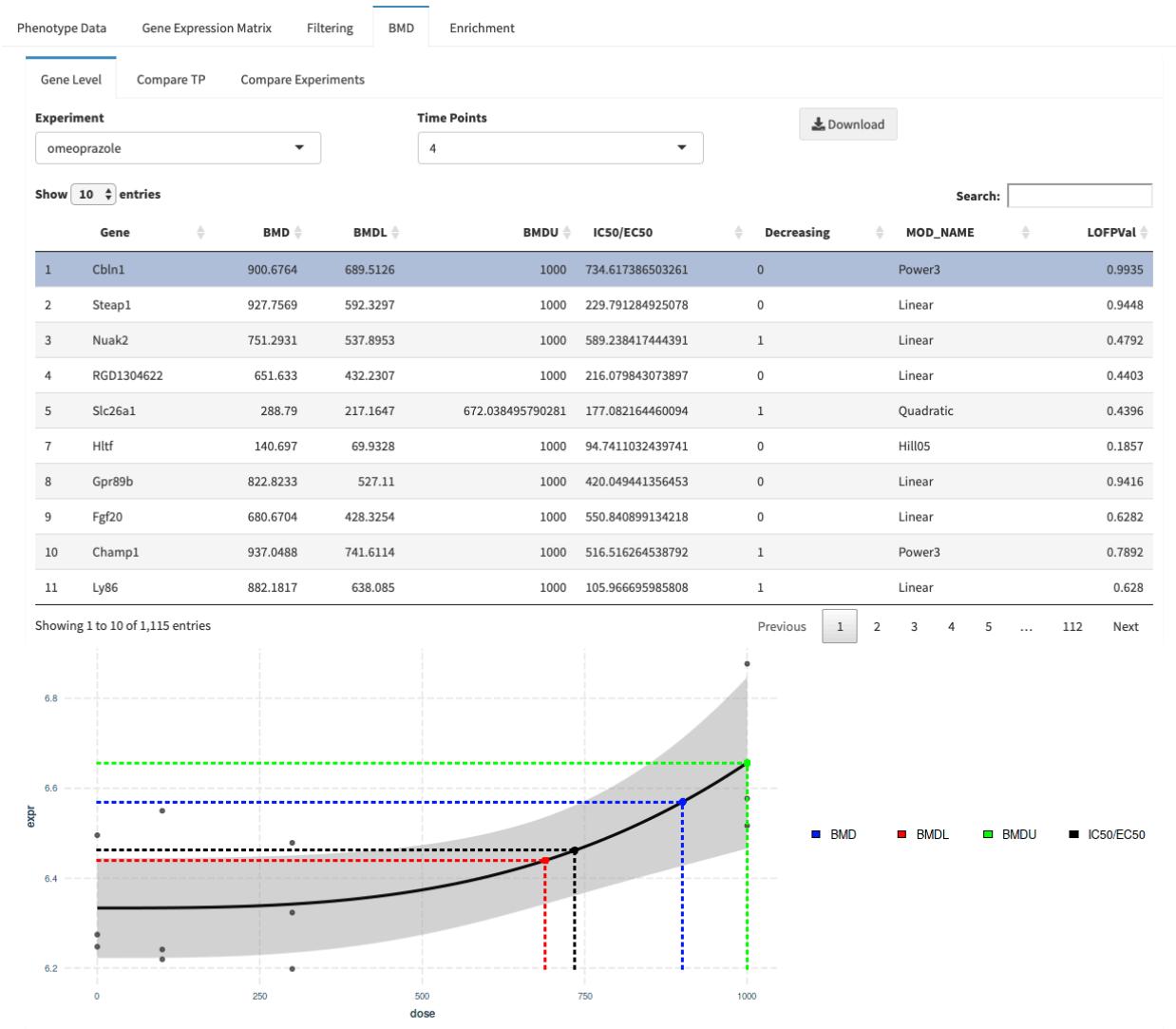
The results of the BMD analysis for each experiment can be explored on the *BMD* tab one time point at a time or different aspects of the results between time points can be visualised under *Compare TP* tab. Additionally, an UpSet plot representing the intersections between different doses and experiment 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		BMD																		
LOAD EXPRESSION MATRIX		Gene Expression Matrix		Filtering		BMD														
GENE FILTERING																				
COMPUTE BMD																				
PATHWAY ENRICHMENT																				
<b>Enrichment</b>																				
MORE INFO																				
<a href="#">GitHub</a> <a href="#">Manual</a> <a href="#">Sample pheno data</a> <a href="#">Sample expression data</a>																				
		Phenotype Data Gene Expression Matrix Filtering <b>BMD</b> Enrichment																		
		<b>Gene Level</b> Compare TP Compare Experiments																		
		<b>Experiment</b> omeoprazole		<b>Time Points</b> 4		<a href="#">Download</a>														
						<input type="text"/>														

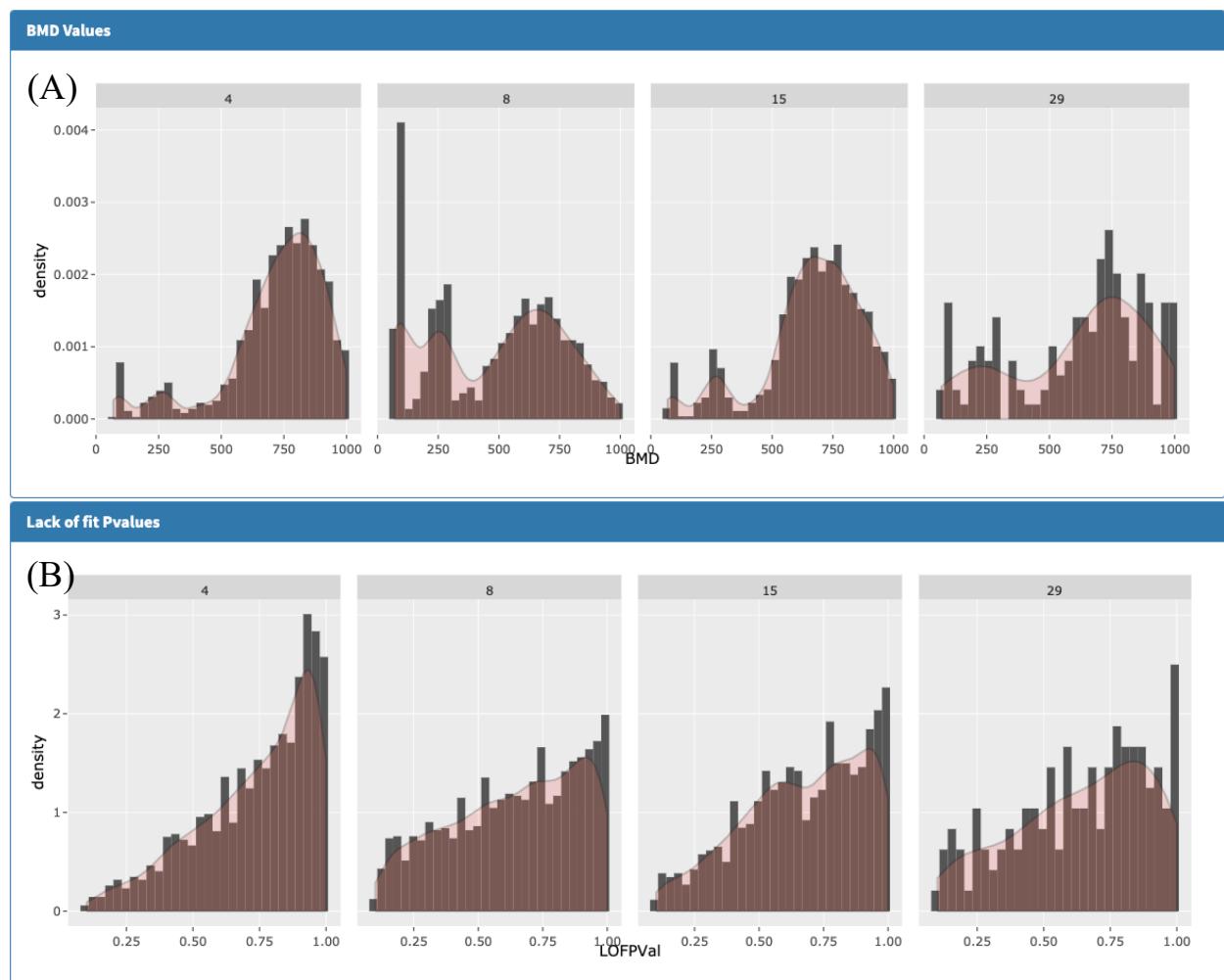
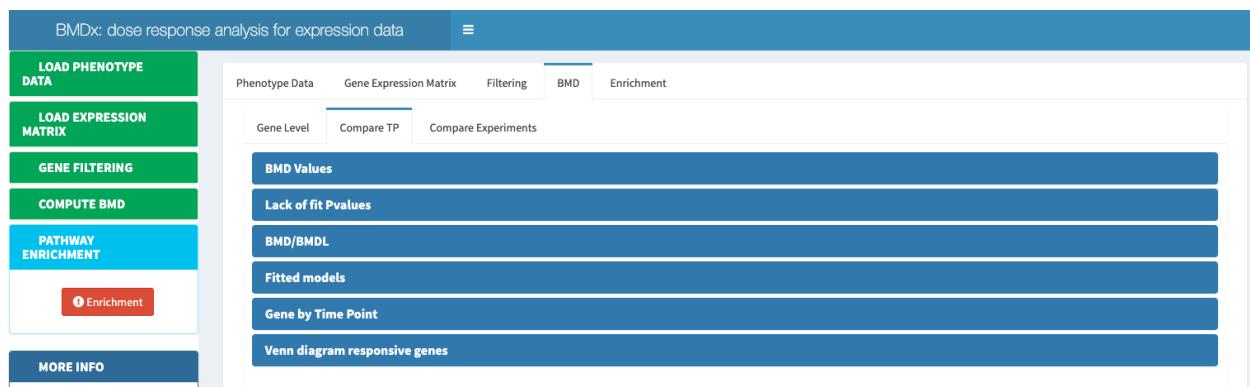
## Fit of the model

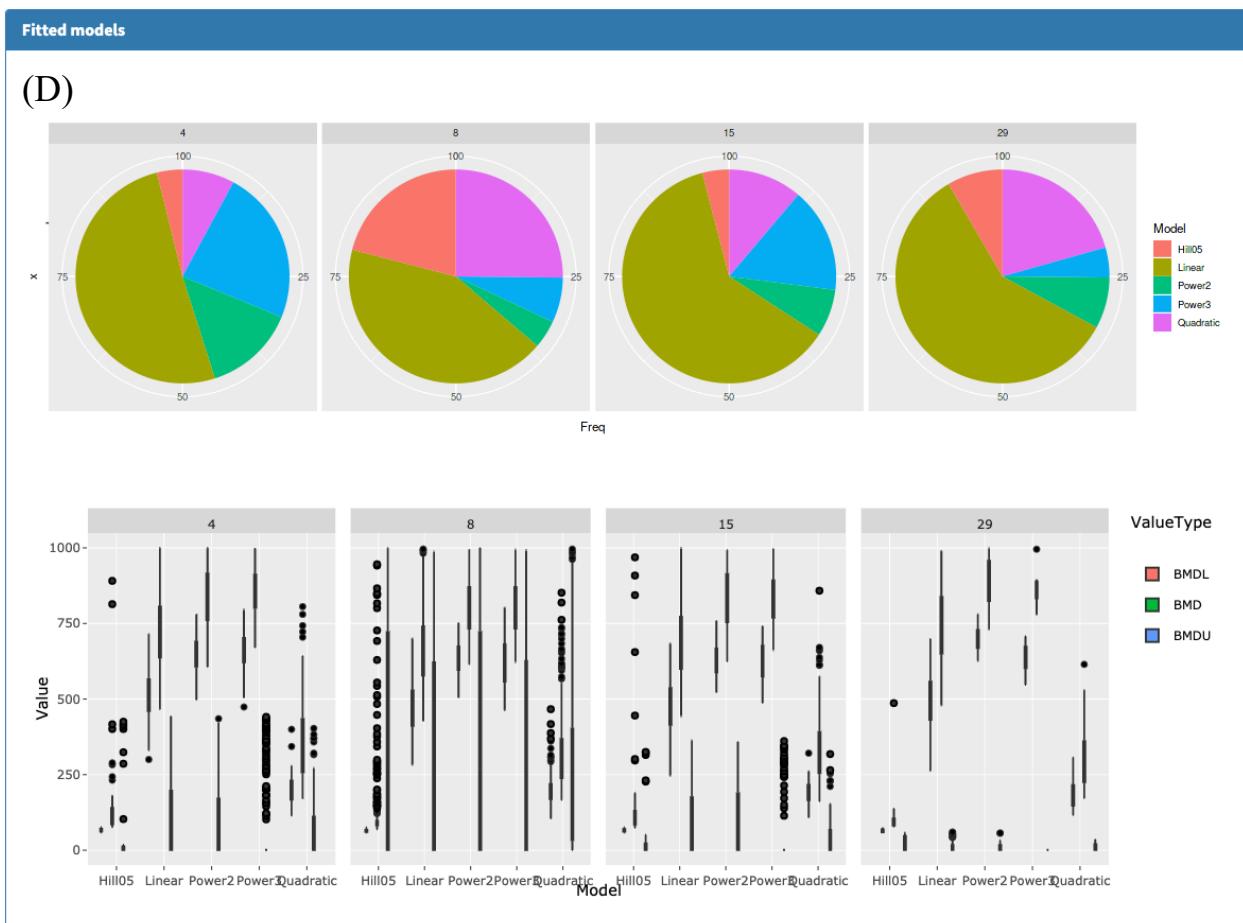
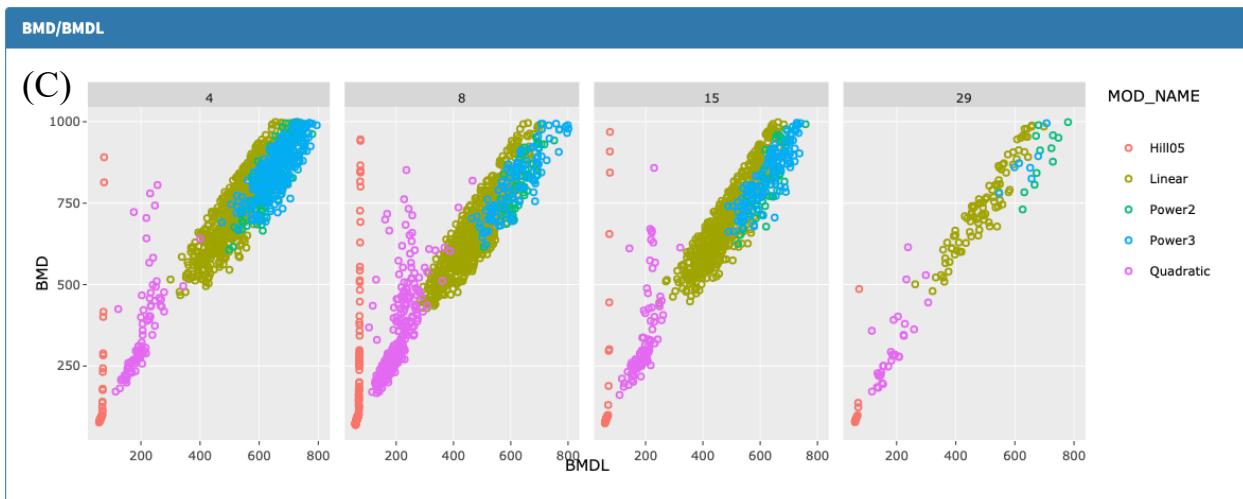
The fitting of the model can be visualised below the table by clicking on the row of the gene. Calculated values are shown in the figure with specific colours: red indicates the value for the BMDL, blue for BMD while black marks the BMDU value. IC50/EC50 is marked with black.

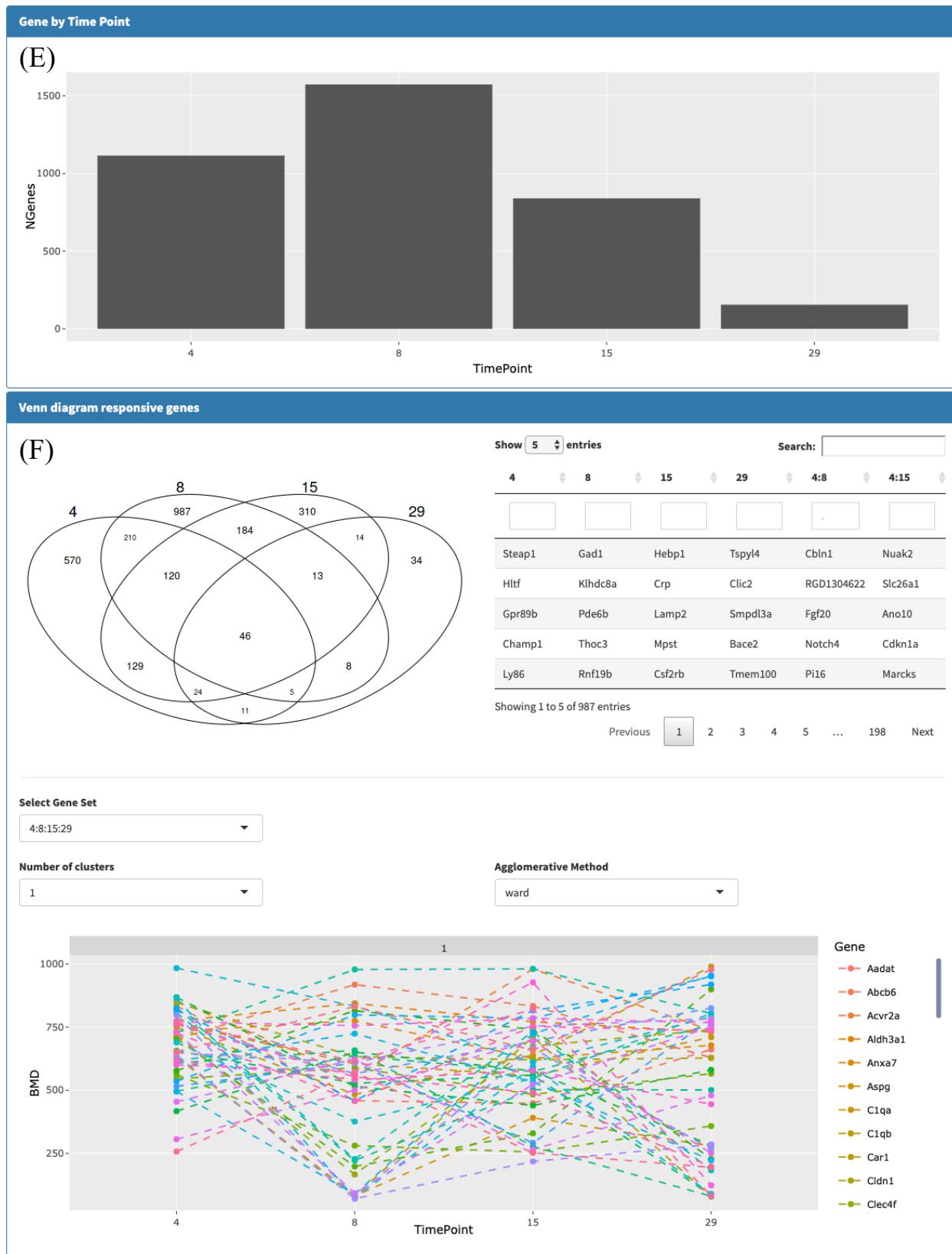


## Comparing results between time points

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). Plots C, D, and F are interactive, and the plotted features can be deselected and selected by clicking the feature (e.g. the gene names in the lower part of the plot F).

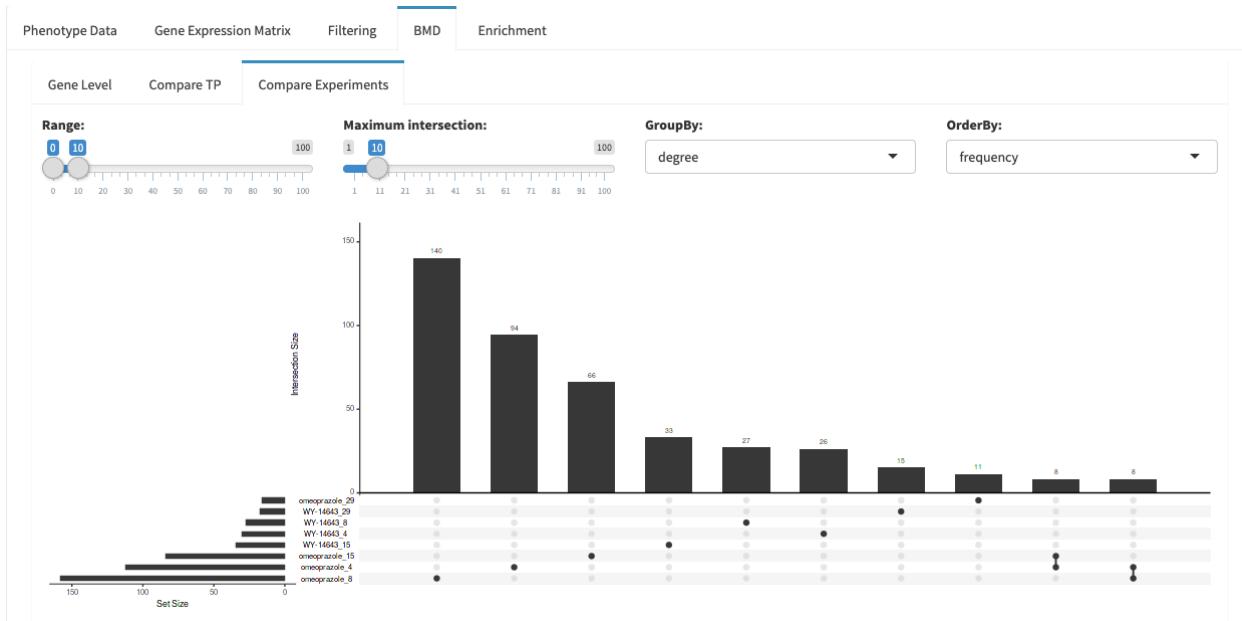






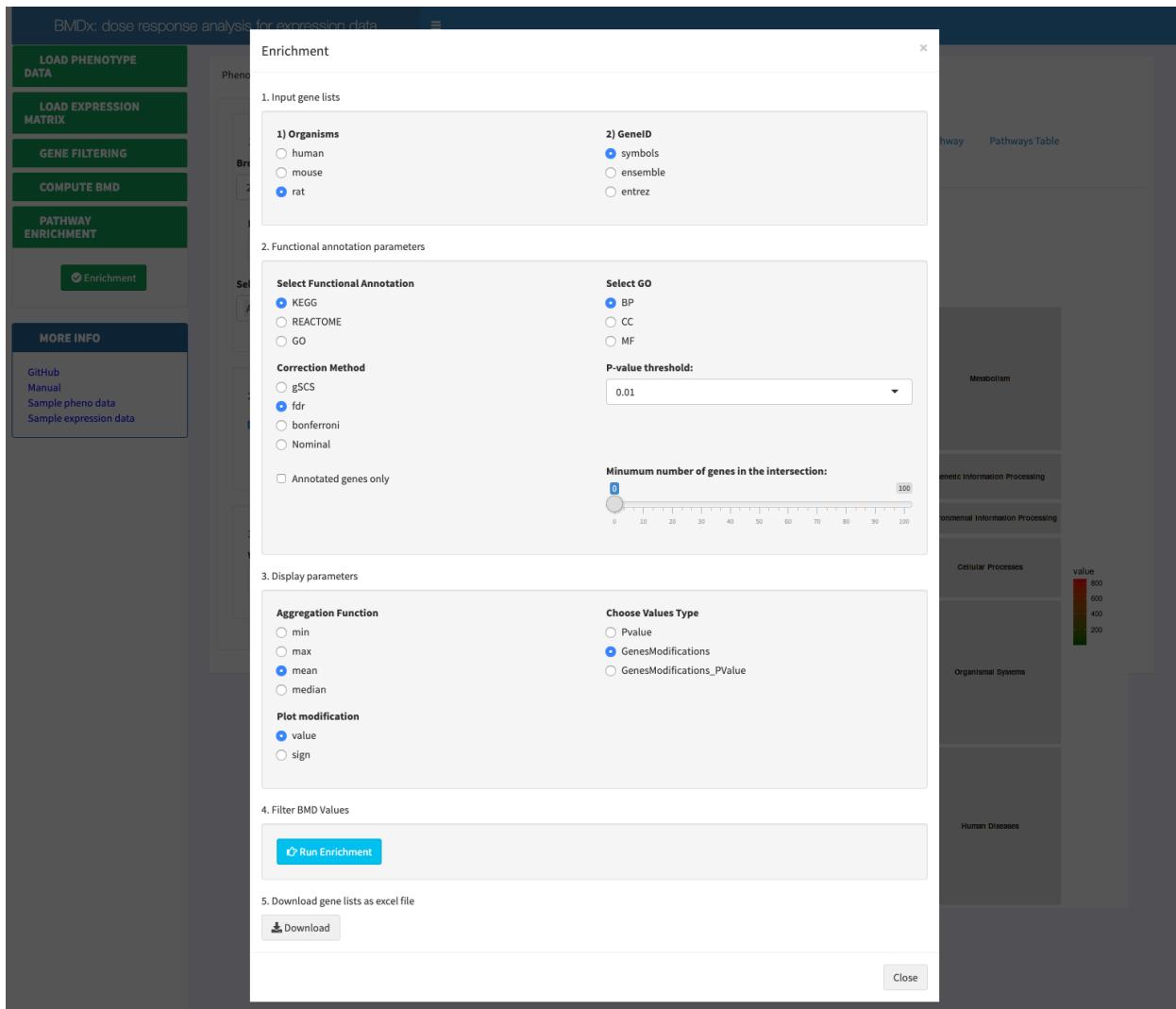
## Compare Experiments

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

<b>1) Organisms</b> <input type="radio"/> human <input type="radio"/> mouse <input checked="" type="radio"/> rat	<b>2) GenelD</b> <input checked="" type="radio"/> symbols <input type="radio"/> ensemble <input type="radio"/> entrez
---	--

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

<b>Select Functional Annotation</b> <input checked="" type="radio"/> KEGG <input type="radio"/> REACTOME <input type="radio"/> GO	<b>Select GO</b> <input checked="" type="radio"/> BP <input type="radio"/> CC <input type="radio"/> MF
<b>Correction Method</b> <input type="radio"/> gSCS <input checked="" type="radio"/> fdr <input type="radio"/> bonferroni <input type="radio"/> Nominal	<b>P-value threshold:</b> 0.01
<input type="checkbox"/> Annotated genes only	<b>Minimum number of genes in the intersection:</b> 

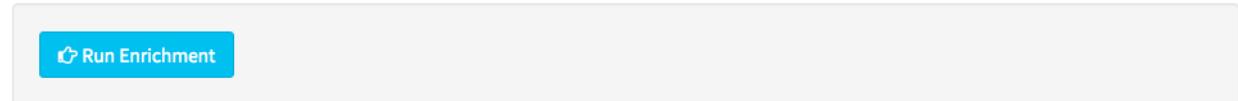
In the bottom of the window, the user can specify the display parameters used for the plotting of the enrichment map. *Aggregation function* (min, max, mean, median) specifies the function to be used when all the genes annotated to the same pathway are aggregated, while *Plot modification* specifies 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

<b>Aggregation Function</b> <input type="radio"/> min <input type="radio"/> max <input checked="" type="radio"/> mean <input type="radio"/> median	<b>Choose Values Type</b> <input type="radio"/> Pvalue <input checked="" type="radio"/> GenesModifications <input type="radio"/> GenesModifications_PValue
<b>Plot modification</b> <input checked="" type="radio"/> value <input type="radio"/> sign	

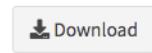
To run the enrichment with selected parameters, *Run Enrichment* is clicked. If the user wants to change the parameters or enrichment type later, the window is launched again from the side panel.

4. Filter BMD Values



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

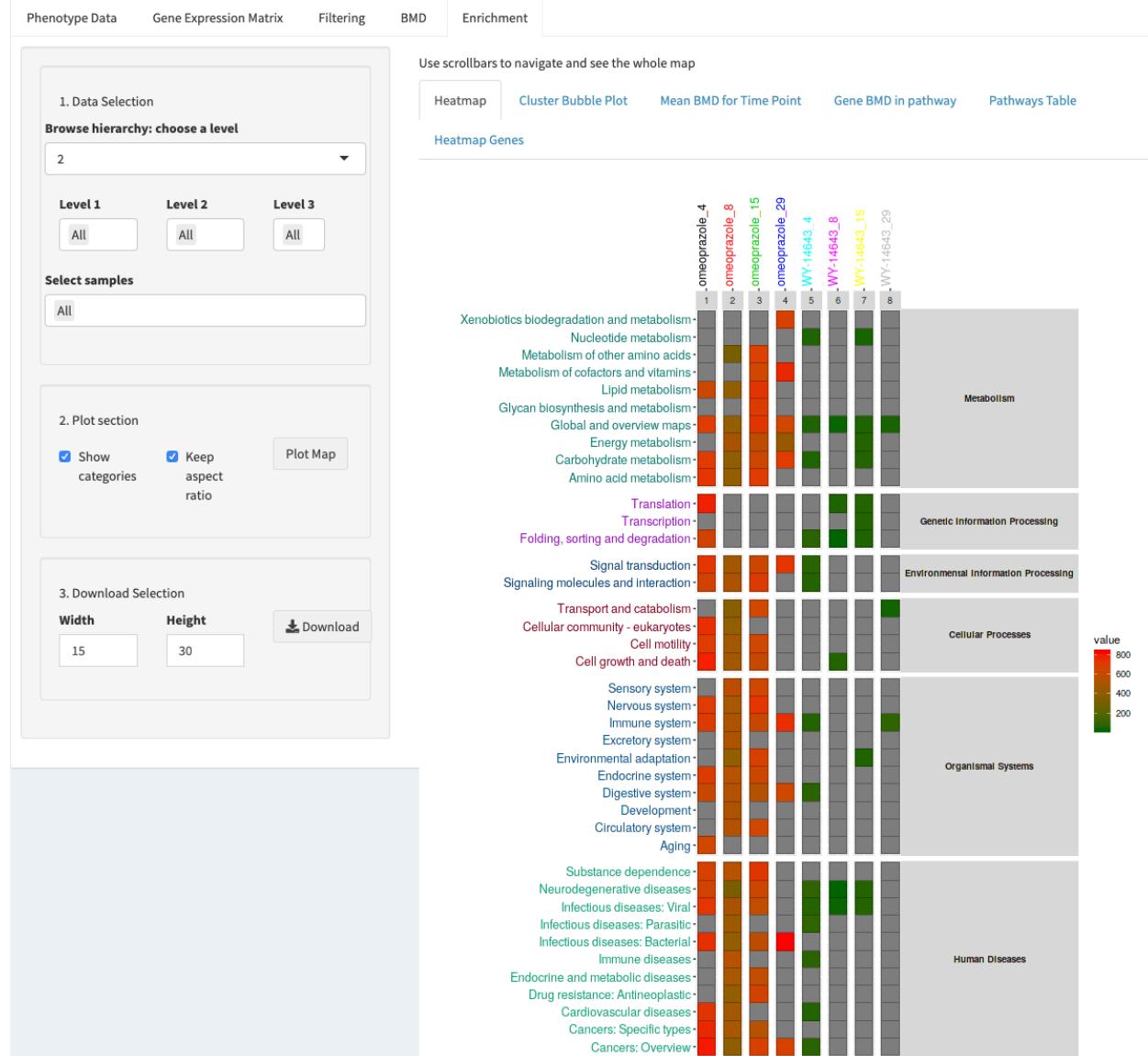
After running the enrichment, the results can be explored on the *Enrichment* tab. Before visualizing the map, the user must specify the hierarchy level on which the results are shown and click on *Plot Map* to open the following view.

A screenshot of a complex web-based application interface titled "Enrichment". The top navigation bar includes tabs for "Phenotype Data", "Gene Expression Matrix", "Filtering", "BMD", and "Enrichment", with "Enrichment" being the active tab. The main content area is divided into several sections: 1. Data Selection, 2. Plot section, and 3. Download Selection. In the "Data Selection" section, there is a dropdown menu for "Browse hierarchy: choose a level" set to "1", and three "All" buttons under "Level 1", "Level 2", and "Level 3". A "Select samples" section shows a single "All" button. In the "Plot section", there are two checked checkboxes: "Show categories" and "Keep aspect ratio", and a "Plot Map" button. In the "Download Selection" section, there are input fields for "Width" (15) and "Height" (30), and a "Download" button. Above the "Data Selection" section, a note says "Use scrollbars to navigate and see the whole map". To the right of the "Data Selection" section, there is a horizontal menu with five options: "Heatmap" (selected), "Cluster Bubble Plot", "Mean BMD for Time Point", "Gene BMD in pathway", and "Pathways Table". Below the "Heatmap" option, there is a link "Heatmap Genes" and a message "No data to plot".



## 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.



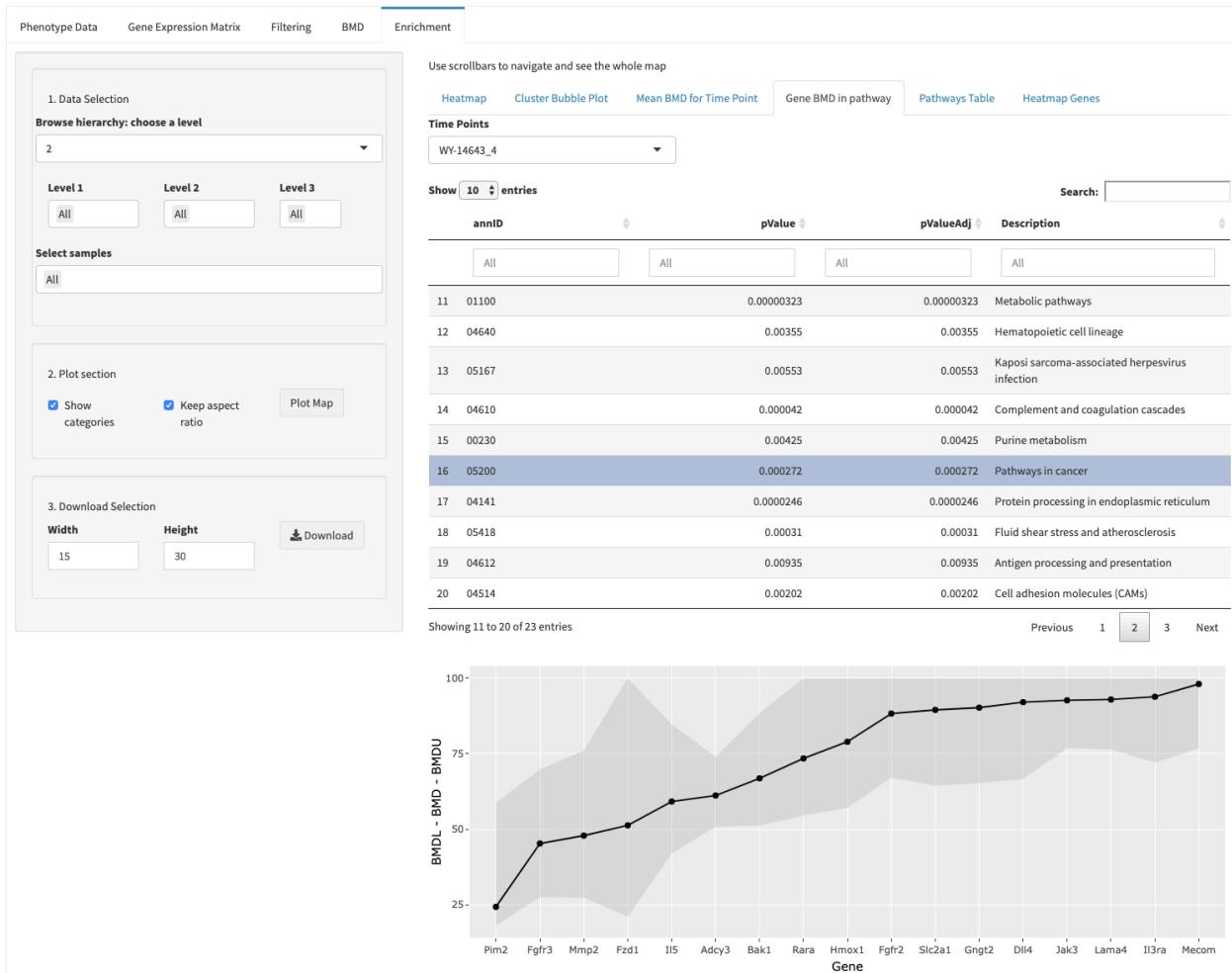
## Mean BMD for Time Point

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.



## 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*.

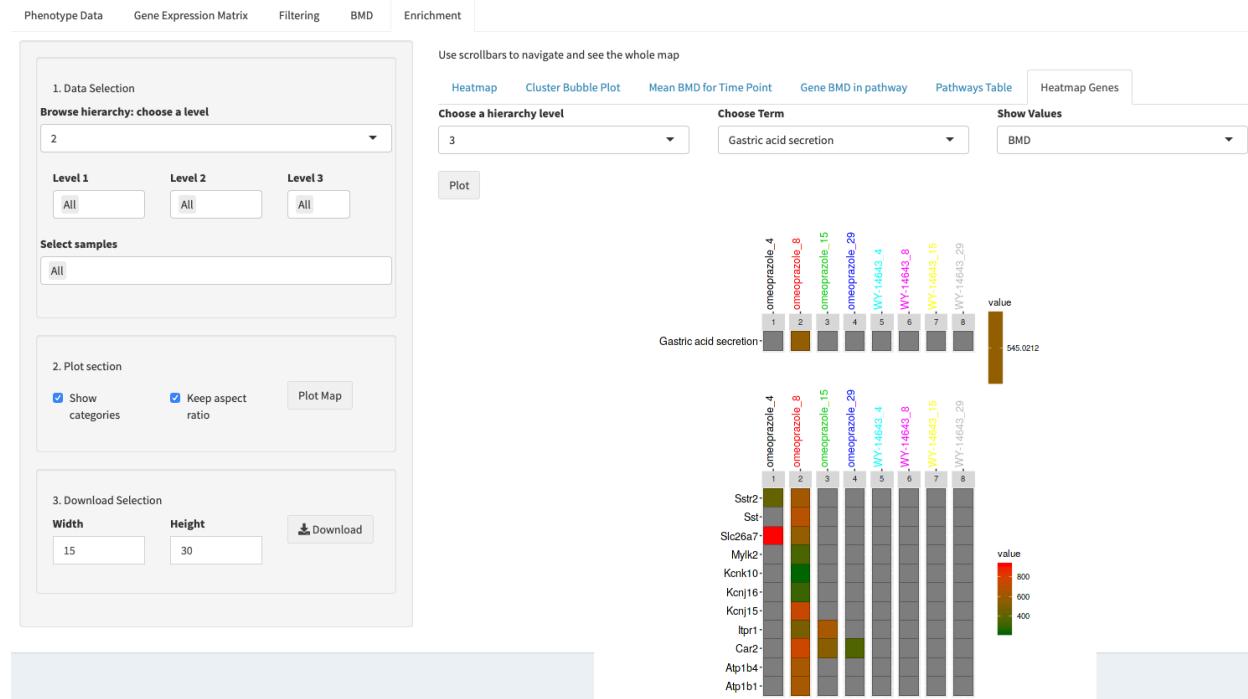
The screenshot shows the 'Pathways Table' tab selected in a top navigation bar. The interface includes sections for 'Data Selection', 'Plot section', and 'Download Selection'. A central table displays pathway information with columns for Description, annID, and gID. The table lists 10 pathways, each with a detailed description and associated gene identifiers.

	Description	annID	gID
1	KEGG pathways	00000	STEAP1,FGF20,CDO1,ALAS2,GATM,PLPP2,CCL7,PRODH1,NOTCH4,PSMB9,CDKN1A,PRF1,MARCKS,AMD1,HDAC2,IPMK,JMD1C,PTPN3,MRP518B,GJA1,FABP7,RBMX,SLC44A4,RAN,MCOLN1,NCOR2,EIF2B1,EIF2AK1,VPS37B,KL,PRKAB1,GLP1R,MSI1,ABCG1,COQ5,TRF
2	Pertussis	05133	ITGB2,C4BPA,C4BPB,SERPING1,CASP3,C1S,C1QB,C1QA,IRF8
3	Insulin signaling pathway	04910	ACACB,PRKAB1,PPP1CC,SOC5S,SOS2,PRKACA,PYGL,PRKCI,FLOT2,TSC2,TSC1,EIF4EBP1,FBP2
4	RNA transport	03013	RPP21,RAN,EIF2B1,NUP54,GEMIN5,ELAC2,PNN,EIF2B4,NUP88,EIF2B2,NUP107,STRAP,CASC3,NUP205,PRMT5,EIF4EBP1,EIF3F,SUM
5	Ascorbate and aldarate metabolism	00053	ALDH2,UGT2A1,UGT2B1,UDGH,RGN,ALDH1B1
6	Focal adhesion	04510	COL6A2,PPP1CC,PDGFRA,ITGA3,IGF1,PAK3,SOS2,BIRC3,HGF,FLNC,RAC2,VTN,BIRC2,COL9A2,FN1,PDGFRB
7	Amphetamine addiction	05031	PPP1CC,ATF2,MAOA,GRIN2C,DDC,PRKACA,SLC18A2,CREB3L2,TH
8	Prion diseases	05020	PRKACA,C8B,C8A,CCL5,C1QB,C1QA
9	Cysteine and methionine metabolism	00270	CDO1,AMD1,SDS,GCLC,APIP,MAT1A,GOT2
10	Synaptic vesicle cycle	04721	NAPA,ATP6V1A,CACNA1B,SYT1,ATP6V1G1,SLC18A2,ATP6V1D,SLC17A6,TCIRG1,AP2A2

Showing 1 to 10 of 53 entries

## 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.
- National Toxicology Program. 2018. NTP Research Report on National Toxicology Program Approach To Genomic Dose-Response Modeling.
- 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.