DRomics tool tutorial

Welcome in the DRomics tool tutorial. This tool aims to build dose-response curves from OMICs dataset, classify the molecular items (e.g probes, metabolites) based on their typology, and derive benchmark dose. In the available version, DRomics supports microarray data (in log scale), RNAseq data (in raw counts) or metabolomics data (in log scale). In addition, a new option allows also the management of apical data. Thus, DRomics supports the linkage of biological responses over different levels of responses, from the molecular to the apical one, using a common philosophy and a common language.

Whatever the chosen response level, the main steps of the workflow remain identical. In this tutorial, the functionalities of the application are illustrated with a microarray dataset that encompasses the fluorescence value of 1000 probes for 6 concentrations in 5 replicates, obtained via single-channel microarray analysis, and data being previously log2-transformed.

For the other response levels, example datasets for the other types of data are also available for download in the app:

- The RNAseq dataset was published by Zhou et al. 2017 (in Toxicological sciences, 160, 95-110) and corresponds to the study of the effect on kidney transcriptomes of tetrachloroethylene. It contains the raw counts of reads at 5 concentrations in 2 or 3 replicates, obtained via Illumina sequencing.

- The metabolomics dataset contains the intensity value of 109 metabolites for 6 concentrations in 3 replicates (plus 6 controls), obtained via GC-MS, and were previously pre-treated as follows. Missing values were retrieved using the half minimum value (i.e. half of the minimum value found for a metabolite across all samples), then the intensity values were log10-transformed and finally scaled to the total intensity (i.e. normalization by sum of signals in each sample).

- The apical dataset (anchoring continuous data) was published in Larras et al. 2020 (in Journal of Hazardous Material, 397, 122727) and contains photosynthesis and growth response of the microalgae Scenedesmus vacuolatus to the biocide triclosan. It consists in a range of 9 concentrations and various numbers of replicates, depending on the concentration.

Before to start:

Please consider that the dataset must present the samples in columns and the molecular items in rows. The name of the columns (first row) must be the concentration value and the name of the rows (first column) must be the name/code of each items (e.g. contig_1, contig_2, … contig_n). The required format is ".txt" with field separators as space or tab. As the DRomics application runs on R software, please avoid spaces and special characters by naming the items.

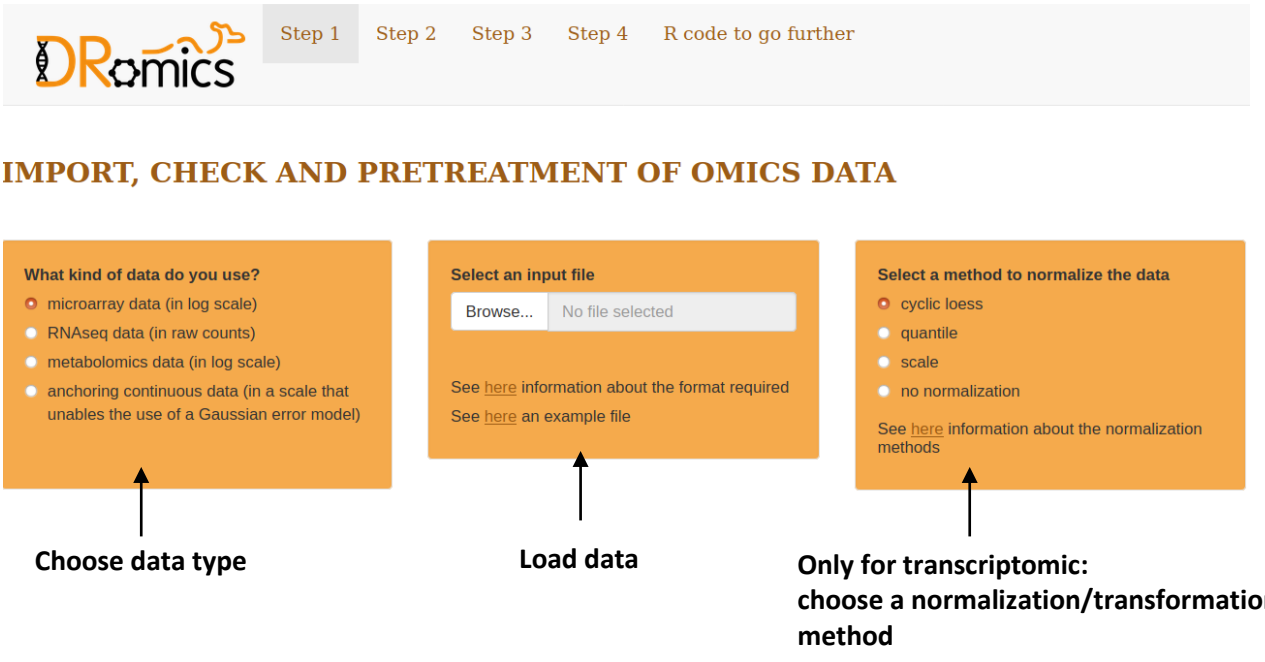
The DRomics tool contains 4 main steps, which take place on 4 different and specific tabs.

1. **TAB STEP 1: IMPORT, CHECK AND IF NEEDED NORMALIZATION and/or TRANSFORMATION OF OMICS DATA**

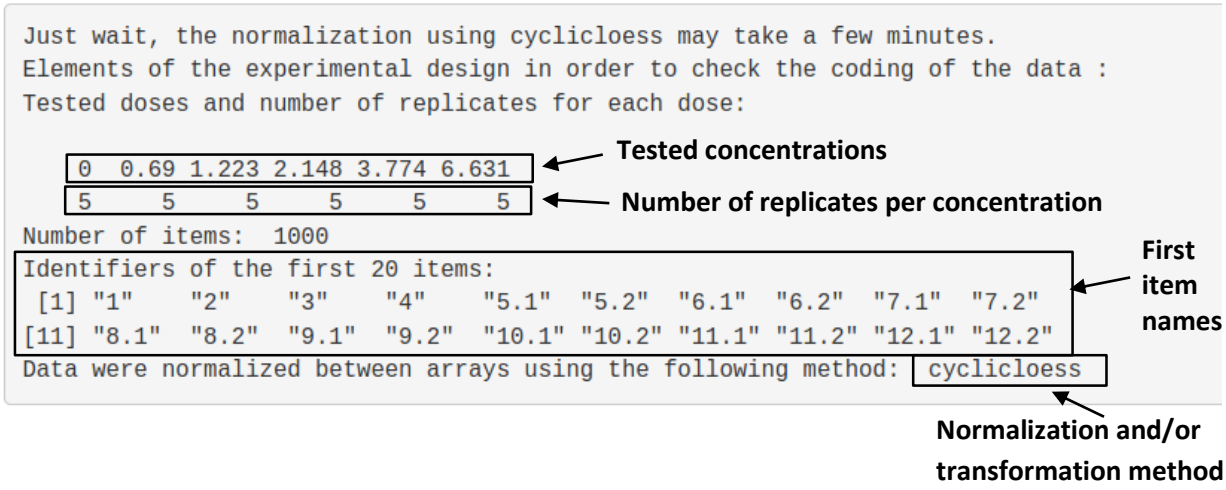
The first step consists to :

- Import the dataset,
- Check automatically the format and if the format is good,
- Offer the possibility to proceed to the normalization/transformation of the dataset (except for metabolomics and apical data). For metabolomics data, no normalization/transformation is provided by DRomics as this step have to be performed by the user prior to its loading in the app (see our recommendations in the app).

First the user must choose between **microarray**, **RNAseq, metabolomics or apical** (anchoring continuous) data. Accordingly, different options for normalization/transformation are displayed. For more information about the data requirements and the suggested methods of normalization/transformation, please check the dedicated help sections (links named "See here" available in the online app). For RNAseq data, be aware that counts are automatically rounded to ensure compatibility of counts from Kallisto or Salmon with the tool as it deals only with integer values.



The step 1 provides a summary of the imported dataset as demonstrated below:

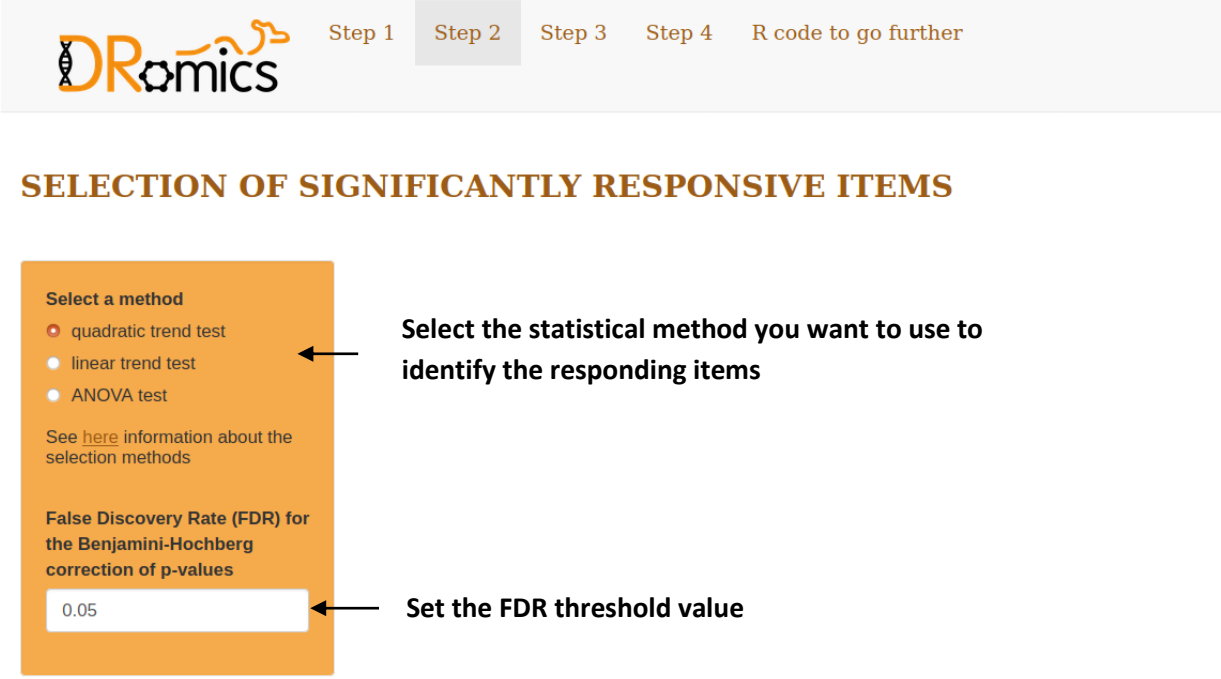


When a normalization/transformation is applied, a graph appears which compares the distribution of the values of each molecular item (y axis) in each sample (x axis) before and after normalization/transformation to visually check the normalization/transformation effect on the data.

Then, click on the tab “Step 2”.

2. TAB STEP 2: SELECTION OF SIGNIFICANTLY RESPONDING ITEMS

The second step aims to identify the significantly responding molecular items to the dose gradient. These items will then be selected to proceed to the next steps. For that, three different methods can be used: the quadratic trend test, the linear trend test and the ANOVA test. We recommend the use of the quadratic trend test for a typically dose-response design. Please check the help section to choose the most appropriate one for your design. The tool also offers the possibility to control the False Discovery Rate (FDR).



The step 2 provides the number of selected items (the responsive ones according to the chosen statistical test) and the names of the 20 most responsive ones (identified as the ones with the lowest p-values obtained from the selection test) are finally presented in a box.

```
Number of selected items using a quadratic trend test with an FDR of  0.05 :  318
Identifiers of the first 20 most responsive items:
 [1] "384.2" "383.1" "383.2" "384.1" "301.1" "363.1" "300.2" "364.2" "364.1"
[10] "363.2" "301.2" "300.1" "351.1" "350.2" "239.1" "240.1" "240.2" "370"
[19] "15"    "350.1"
```

Then, click on the tab “Step 3”.

3. TAB STEP 3: MODEL SELECTION AND DOSE-RESPONSE MODELLING

The third tab consists in the selection of the best-fit model for each previously selected item (those which respond significantly according to the chosen test and the FDR chosen value) and their respective dose-response curve building. For information about the best model selection please check the dedicated help section. This step can take from minutes to about one hour.

To start this step, click on “Fit”.

Click on “Fit” to start the procedure (model fitting and best model selection)

Once the step is done, a new button appears on the right of the “Fit” button, offering you the possibility to download the dose-response curve with the fitted model for all of the items previously selected.

A summary of the analysis appears in a box and 3 options appear below the box in order to obtain different graph :

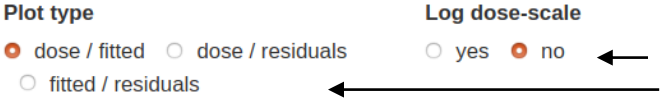
```
47 dose-response curves out of  318  previously selected were removed
        because no model could be fitted reliably.
Distribution of the chosen models among the  271  fitted dose-response curves :

            Hill           linear      exponential     Gauss-probit
               4               67               69              104
log-Gauss-probit
              27
Distribution of the typology of the  271  fitted dose-response curves :

       H.inc         H.dec         L.inc         L.dec  E.inc.convex
           1             3            46            21            21
E.dec.concave E.inc.concave  E.dec.convex          GP.U       GP.bell
          25            17             6            56            48
       lGP.U      lGP.bell
          15            12
```

Number of items for which no model adequately fit the data

Number of items per model (e.g. the Hill model best fitted the response of 4 items)

Number of items per typology (e.g. 46 items were modelled by a linear increasing curve)

For more information about the full meaning of each typology name, please check the dedicated help section.



Different graphical possibilities offered to the user to check the dose-response curves as well as the residuals.

At this step, DRomics offers the possibility to display on the online interface (directly under the previously presented box) different plots related to the model fitting :

- dose/fitted: The dose-response curve with the model fit in red, the value of each replicate as white dots and the mean value as black dots
- dose/residuals: showing residuals as a function of the dose
- fitted/residuals: showing residuals as a function of the fitted values

The two last residual plots are useful to check the error model: Gaussian distribution of residuals with a common variance.

Then, click on the tab “Step 4”.

4. TAB STEP 4: COMPUTATION OF BENCHMARK DOSES FOR RESPONSIVE ITEMS

The last step aims to derive a Benchmark Dose (BMD) according to the EFSA report (Hardy et al., 2017). Two kinds of BMD can be derived by the tool but we strongly recommend the use of the first one (see in Larras et al., 2018):

- BMD-zSD: the BMD based on z times the standard deviation of the data along the whole curve
- BMD-xfold: the BMD based on a x-fold change of the signal compared to the control

For more details about how are calculated both kind of BMDs, please check the dedicated help section. For both BMDs, the value of z (factor multiplying the standard deviation value) and x (fold change value compared to the control) are modifiable by the user. A value of 1 and 10 (for z and x, respectively) are proposed by default.

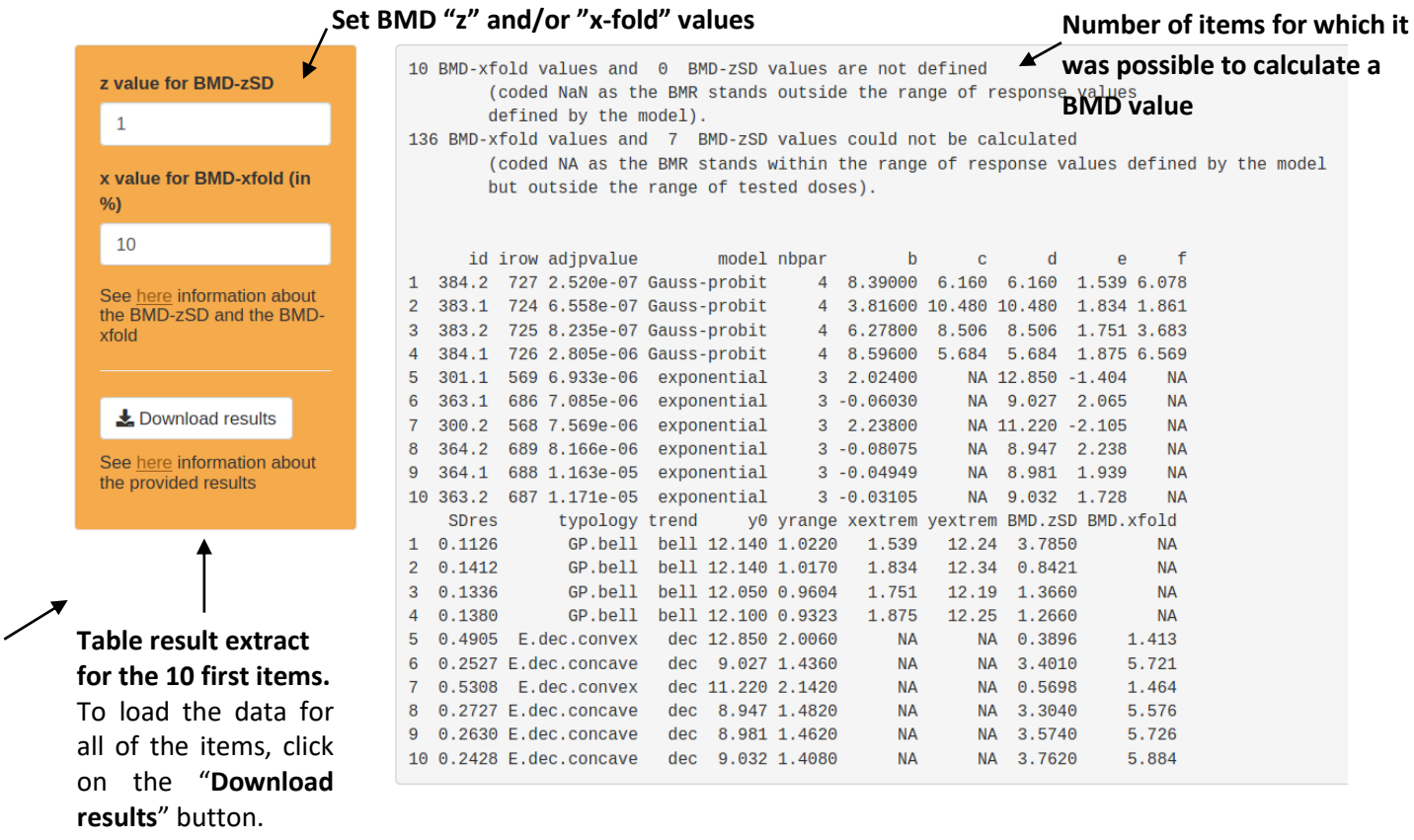


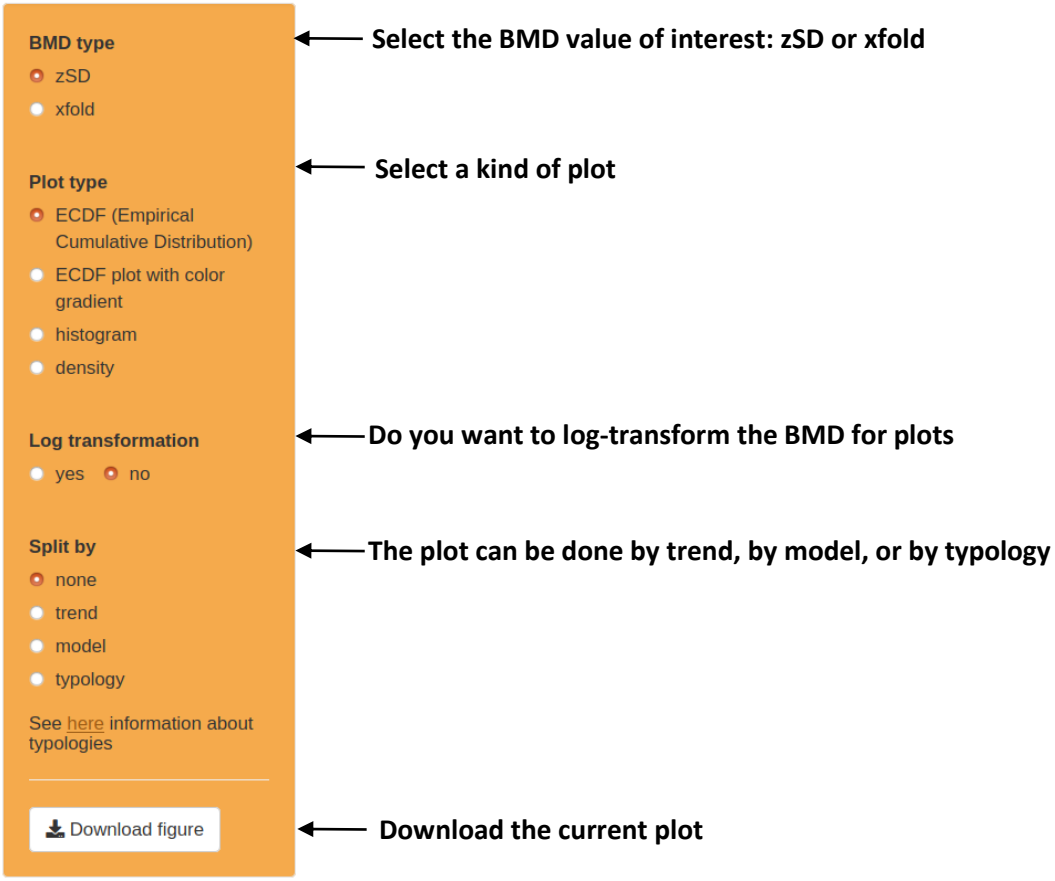
Table result extract for the 10 first items. To load the data for all of the items, click on the "**Download results**" button.

This step provides two kinds of results.

First, a visualization of the 10 first rows of a dataframe summarizing the properties of the curves, and the BMD values with following columns:

- **id**: The name of each item
- **irow**: Their line number in the initial dataset
- **adjpvalue**: The adjusted p-value resulting from the selection test (step2)
- **model**: The best model associated to the item (step3)
- **npar**: the number of parameters of the model
- **b, c, d, e, f**: the value of the parameters of the model
- **SDres**:the residual standard deviation of the best model
- **typology**: the typology associated to the item
- **trend**: the main trend of the response
- **y0: the** theoretical value at the control
- **yrange:** the theoretical y range for x within the range of tested doses
- **xextrem:** the x value at which the extremum is reached (for U and bell-shaped curves)
- **yextrem:** the y value at which the extremum is reached (for U and bell-shaped curves)
- **BMD.zSD**: the value of the BMD.SD for the selected z value
- **BMD.xfold**: the value of the BMD.xfold for the selected x value

Second, the BMD value results are also provided as a graphical output which can be modulated according the user whishes:



REFERENCES

Larras F, Billoir E, Baillard V, Siberchicot A, Scholz S, Wubet T, Tarkka M, Schmitt-Jansen M and Delignette-Muller ML (2018). DRomics: a turnkey tool to support the use of the dose-response framework for omics data in ecological risk assessment. Environmental Science & Technology. https://doi.org/10.1021/acs.est.8b04752 (or on HAL : https://hal.archives-ouvertes.fr/hal-02309919).

EFSA Scientific Committee, Hardy, A., Benford, D., Halldorsson, T., Jeger, M. J., Knutsen, K. H., ... & Ockleford, C. (2017). Update: use of the benchmark dose approach in risk assessment. *EFSA Journal*, *15*(1), e04658. https://doi.org/10.2903/j.efsa.2017.4658 .