DRomics tool tutorial

Last update: 21/06/2019

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 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.

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).

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 the items (e.g. probes, metabolites). The required format is ".txt". As the DRomics application runs on R software, please avoid spaces and special characters by naming the items.

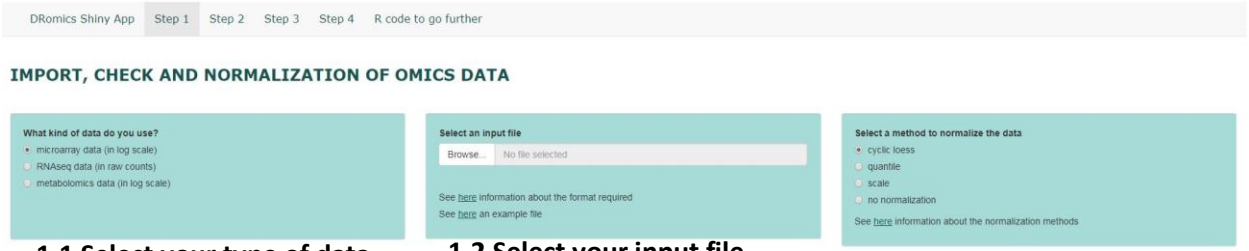
The DRomics tool contains 4 main steps, which take place on 4 different and specific tabs.

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, and depending of the type of data, proceed to the normalization/transformation of the dataset. First the user has to choose between microarray data, RNAseq data or metabolomics 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 See here available in the online app).

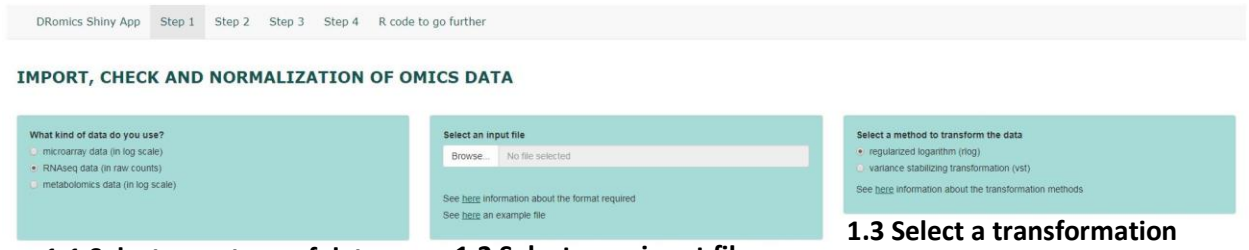
- **How to proceed?**

For **microarray data**:



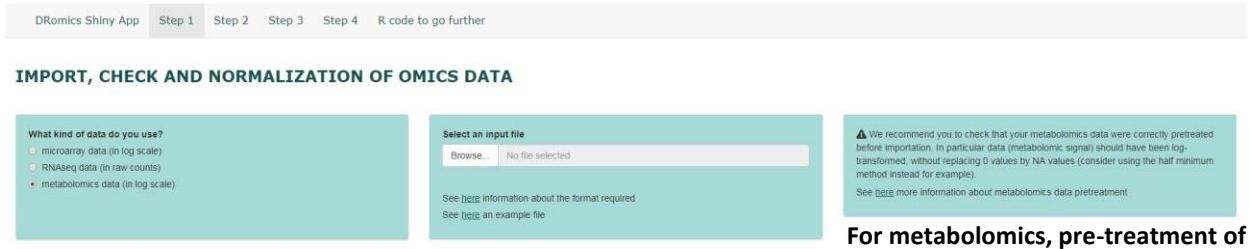
1.1 Select your type of data **1.2 Select your input file** **1.3 Select a normalization method**

For **RNAseq data**:



1.1 Select your type of data **1.2 Select your input file** **1.3 Select a transformation method**

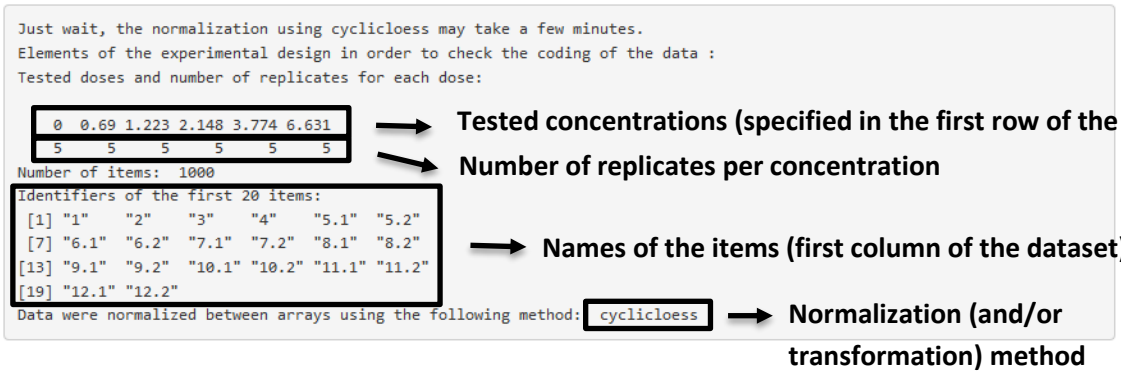
For **metabolomics data**:



1.1 Select your type of data **1.2 Select your input file** **For metabolomics, pre-treatment of data must be done before importation**

- **Results of step 1**

The summary of the imported dataset is then provided as below.



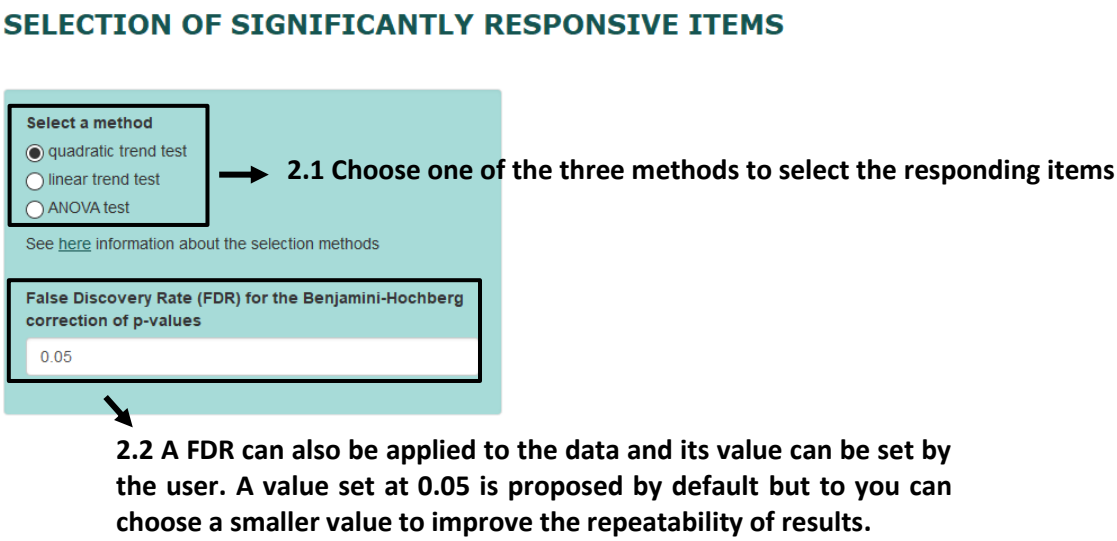
When a normalization/transformation is applied, a graph appears which compares the distribution of the values of each molecular item (y axis) in each sample (x axis) before and after normalization/transformation to visually check the normalization/transformation effect on the data.

Then, click on the tab “Step 2”.

2. TAB STEP 2: SELECTION OF SIGNIFICANTLY RESPONDING ITEMS

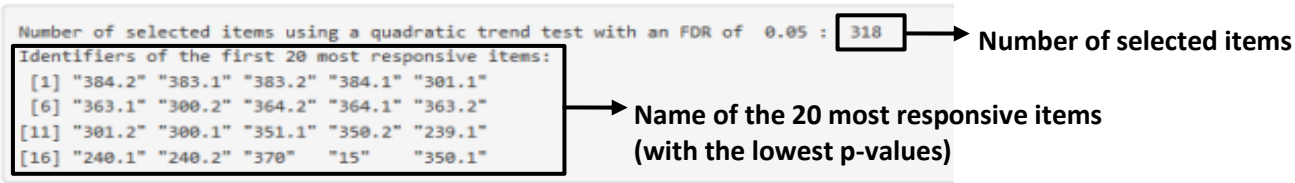
The second step aims to identify the significantly responding molecular items to the dose gradient. These items will then be selected to proceed to the next steps. For that, three different methods can be used: the quadratic trend test, the linear trend test and the ANOVA test. We recommend the use of 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).

- **How to proceed?**



- **Results of step 2**

The number of selected items (the responsive ones) 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.

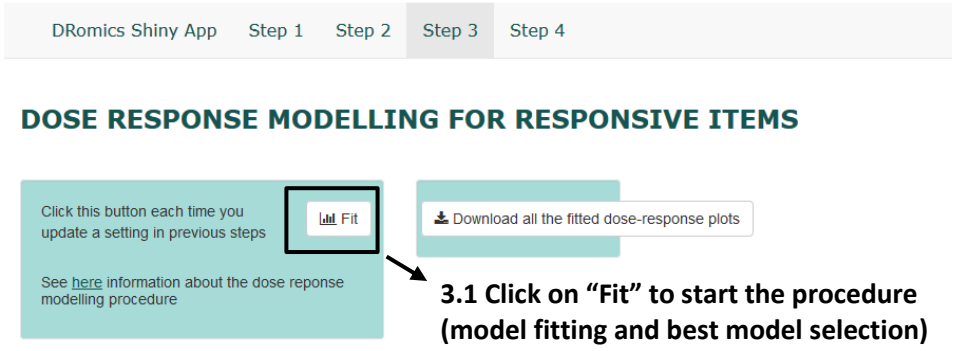


Then, click on the tab “Step 3”.

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.

- **How to proceed?**



3.1 Click on "Fit" to start the procedure (model fitting and best model selection)

- **Results of step 3**

First, a summary of the analysis is provided in a box:

```
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
```

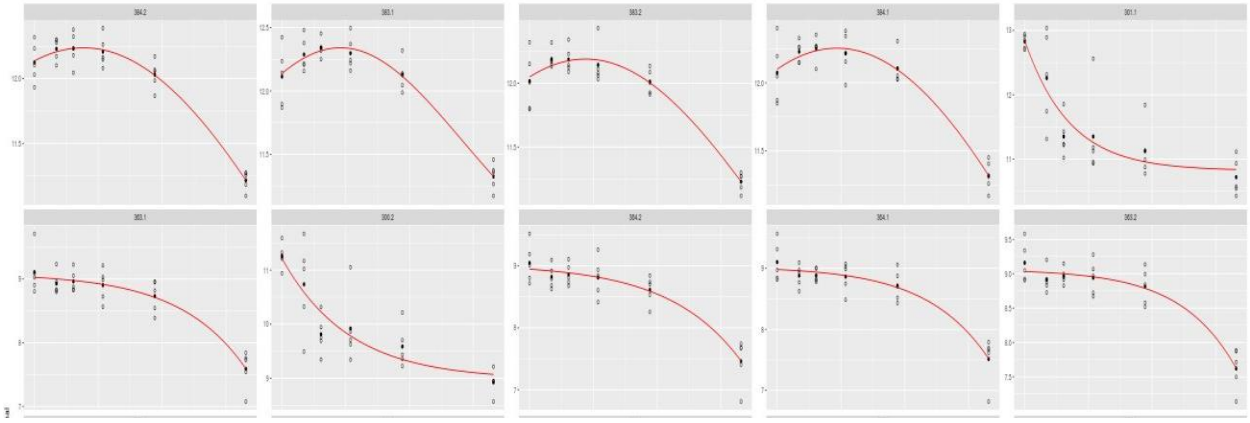
Number of items for which no model adequately fit the data

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.

The graphical presentation of the dose-response curve of the first items is also presented. Those Graphs present the value of each replicate (white dots) per sample as well as the mean value (black dots). The x axis presents the concentration values and the y axis the signal values.



Finally, the user can export the totality of the best-fit dose-response curves by clicking on “Download all the fitted dose-response plots”.

Then, click on the tab “Step 4”.

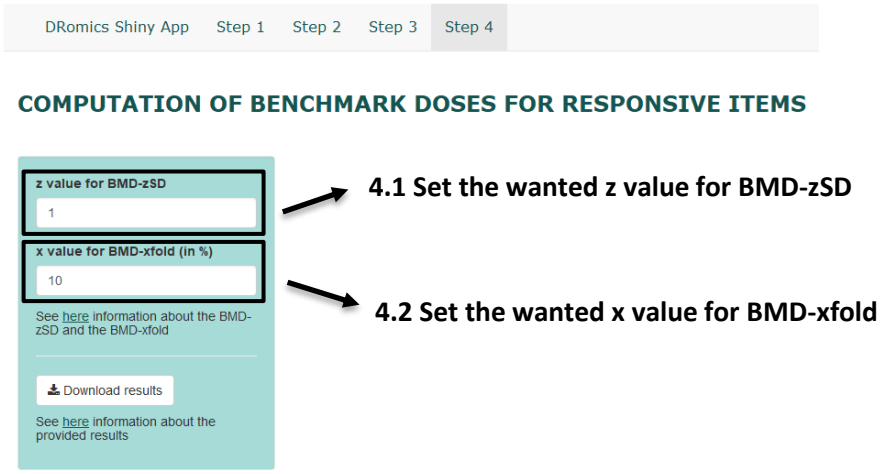
4. TAB STEP 4: COMPUTATION OF BENCHMARK DOSES FOR RESPONSIVE ITEMS

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:

- 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.

- **How to proceed**



- **Results of step 4**

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:

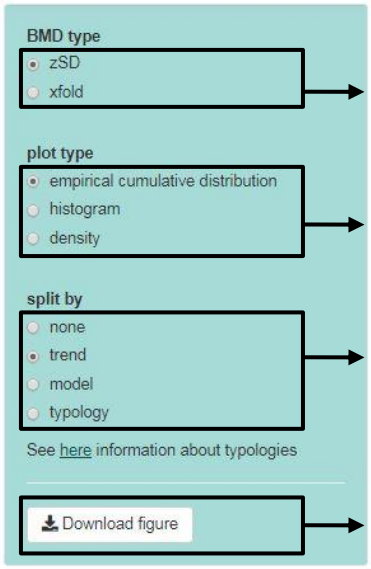
- **id**: The name of each item
- **irow**: Their line number in the initial dataset
- **adjpvalue**: The p-value resulting from the selection test (step2)
- **model**: The best model associated to the item (step3)
- **npar**: the number of parameters of the model
- **b, c, d, e, f**: the value of the 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:** theoretical value at the control
- **yrange:** theoretical y range for x within the range of tested doses
- **xextrem:** x value at which the extremum is reached (for U and bell-shaped curves)
- **yextrem:** 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

```
10 BMD-xfold values and  0  BMD-zSD values are not defined
       (coded NaN as the BMR stands outside the range of response values
       defined by the model).
136 BMD-xfold values and  7  BMD-zSD values could not be calculated
       (coded NA as the BMR stands within the range of response values defined by the model
       but outside the range of tested doses).


      id irow adjpvalue        model nbpar        b      c      d      e     f
1  384.2  727 2.520e-07 Gauss-probit     4  8.39000  6.160  6.160  1.539 6.078
2  383.1  724 6.558e-07 Gauss-probit     4  3.81600 10.480 10.480  1.834 1.861
3  383.2  725 8.235e-07 Gauss-probit     4  6.27800  8.506  8.506  1.751 3.683
4  384.1  726 2.805e-06 Gauss-probit     4  8.59600  5.684  5.684  1.875 6.569
5  301.1  569 6.933e-06  exponential     3  2.02400     NA 12.850 -1.404    NA
6  363.1  686 7.085e-06  exponential     3 -0.06030     NA  9.027  2.065    NA
7  300.2  568 7.569e-06  exponential     3  2.23800     NA 11.220 -2.105    NA
8  364.2  689 8.166e-06  exponential     3 -0.08075     NA  8.947  2.238    NA
9  364.1  688 1.163e-05  exponential     3 -0.04949     NA  8.981  1.939    NA
10 363.2  687 1.171e-05  exponential     3 -0.03105     NA  9.032  1.728    NA
    SDres      typology trend     y0 yrange xextrem yextrem BMD.zSD BMD.xfold
1  0.1126       GP.bell  bell 12.140 1.0220   1.539   12.24  3.7850        NA
2  0.1412       GP.bell  bell 12.140 1.0170   1.834   12.34  0.8421        NA
3  0.1336       GP.bell  bell 12.050 0.9604   1.751   12.19  1.3660        NA
4  0.1380       GP.bell  bell 12.100 0.9323   1.875   12.25  1.2660        NA
5  0.4905  E.dec.convex   dec 12.850 2.0060      NA      NA  0.3896     1.413
6  0.2527 E.dec.concave   dec  9.027 1.4360      NA      NA  3.4010     5.721
7  0.5308  E.dec.convex   dec 11.220 2.1420      NA      NA  0.5698     1.464
8  0.2727 E.dec.concave   dec  8.947 1.4820      NA      NA  3.3040     5.576
9  0.2630 E.dec.concave   dec  8.981 1.4620      NA      NA  3.5740     5.726
10 0.2428 E.dec.concave   dec  9.032 1.4080      NA      NA  3.7620     5.884
```

To load these data for all of the items for which a dose-response curve was build, click on the “Download results” button.

Second, the BMD value results are also restituted *via* a graphical output which can be modulated according the user whishes:



4.3 Select the BMD value of interest (zSD or xfold)

4.4 select a kind of plot

4.5 The plot can be done by trend, by model, or by typology

4.6 Download the current plot

Example of output for **plot type** set to empirical cumulated distribution and **split by** trend